Abstract

Measuring Transcriptional Dynamics Using MTS Chemistry

Erin Duffy

2018

RNA is continuously transcribed and degraded in a tightly regulated and transcript-specific manner. The dynamics of different RNA populations can be studied by targeted incorporation of non-canonical nucleosides such as 4-thiouridine (s^4U) into newly transcribed RNA, followed by enrichment with activated disulfides. I found that the commonly used activated disulfide, HPDP-biotin, is inefficient, leading to low yields and bias toward longer RNAs. I then demonstrated that the activated disulfide methane thiosulfonate (MTS) biotin reacted much more efficiently, which drastically improved the yield in s^4U enrichment studies and alleviated the length bias previously associated with s^4U metabolic labeling experiments. MTS chemistry enabled the first study of microRNA (miRNA) turnover in proliferating cells with flux through the miRNA pathway, which revealed several fast-turnover miRNAs that were previously thought to be stable. This chemistry can be used to improve methods that use s^4U metabolic labeling, including the study of *Mettl3*-dependent RNA stability upon IL-7 stimulation in naïve T-cells, and enrichment methods can be combined with mutational mapping (TimeLapse-seq) to filter non-s^4U contamination for additional sensitivity.

The improved efficiency of MTS chemistry allowed the development of MTS resin, a one-step, completely covalent method to capture s^4U-RNAs from small populations of cells. This solid-phase chemistry expanded the utility of s^4U metabolic labeling experiments from cell culture to primary cells and enabled the first study of RNAPII elongation rates in mouse cortical neurons. Tissue-specific RNA dynamics, particularly in primary tissues, are not well characterized due to the high scale required previously for these experiments, so MTS resin should allow a better understanding of RNA turnover in many tissue types.

Finally, I sought to use MTS chemistry to identify the acute transcriptional response to corticosteroid stimulation in A549 epithelial cells. Corticosteroids bind the glucocorticoid receptor (GR) and cause a wide variety of physiological effects including regulation

of glucose synthesis and suppression of the immune response. Despite decades of research into the mechanism of GR activity, the immediate early targets of GR activation are poorly characterized due to the long treatment times (1-4 hours) required to observe transcriptional changes by traditional RNA-seq. This area is an ideal application of MTS chemistry because efficient enrichment of s^4U-RNA is required following very short incubation of cells with s^4U and corticosteroids. I observed novel induction kinetics of GR-responsive genes and enhancers in A549 cells. These studies lend important insights into the field of RNA dynamics and offer a novel toolkit of methods that opens the door for future study of RNA turnover in complex systems.

Measuring Transcriptional Dynamics Using MTS Chemistry

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Erin Duffy

Dissertation Director: Dr. Matthew Simon

May 2018

Contents

List of Figures

List of Tables

Acknowledgements

For the research I am very grateful to my research collaborators, who have contributed greatly to my work. First, I would like to thank Dr. Daniele Canzio, whom I have collaborated with throughout my PhD and who has given countless hours to scientific discussion and advice. To Dr. Mitch Omar, Juliana Shaw, Dr. Xiao Xiao, and Dr. Tony Koleske, thank you for being amazing colleagues and collaborators through the last phase of my PhD and for inspiring my interest in neurobiology. To Dr. Robert Kitchen and Dr. Mark Gerstein, Dr. Huabing Li and Dr. Richard Flavell; Dr. Rachel Duffie and Dr. Stavros Lomvardas; Dr. Paulina Pawlica, Dr. Ari Landon and Dr. Joan Steitz; Dr. Jennifer Yang and Dr. Sherman Weissman; Dr. Xinguo Chen and Dr. Sandy Wolin; and Dr. Sei Won Lee and Dr. Charles de la Cruz; your projects using MTS chemistry have not only exposed me to a fascinating variety of biology and technology, you have also pushed the boundaries of my technologies and helped improve so many aspects of the protocols. To my thesis committee, Dr. Joan Steitz and Dr. Ron Breaker, you have constantly pushed me to do more and be curious. Your guidance throughout my PhD has been invaluable, particularly at the points in my project when I doubted myself the most.

For everything else To my advisor, Dr. Matt Simon, I can't thank you enough for accepting me as the first graduate student in your lab and for constantly pushing me to be the best scientist I can be. All of the success I have achieved in graduate school was possible because of opportunities that you created for me, and I hope to pass on all of the lessons that you taught me to my own students some day. To members of the Simon lab, both past and present, you are the greatest group of people to work with every day, and your individual expertise has been invaluable to every aspect of my projects. Alec, thank you for

always giving me perspective on my science and reminding me to not take my experiments personally, even when I'm very bad at following that advice. Martin, your zen about science is inspirational (or "pretty good" as you would say) and I value all of the feedback that you gave me in our subgroup meetings. Tyler, my first rotation student, I'm so glad you joined the lab and brought so much fun and adventure with you. Will, a.k.a. "Mr. President," thank you for teaching me all of the synthesis I know and for being a wonderful friend and office-mate. I'm confident you will keep the lab from burning down when I'm gone. Jeremy, my second rotation student, thanks for making this s^4U project into a dynamic duo and pushing the methodology farther than I could have imagined. Lea, my outfit twin, I know I'm leaving the MTS pulldowns in very capable hands and I'm so excited to see all the amazing things you'll do with RNA and chromatin. Meaghan, I am constantly amazed by your fearlessness and I'm so proud of how much you've grown in this lab. I will take your dance lessons, including the "funky butt," wherever I go. Catherine, my extraordinary undergrad, it has been a pleasure to watch you grow into a curious and capable scientist and to continue to follow your amazing experiences at Stanford. Elizabeth, my inquisitive high schooler, you are just beginning to dive into the amazing world of science and I know you will be successful no matter where you go. To Peter and Maddie, I'm so glad I got to know each of you during your time in the Simon lab, and I can't wait to follow all of your amazing accomplishments to come. And to members of the Strobel and Slavoff labs next door, thank you for being a constant source of inspiration, scientific support, and overall good times.

To my informal mentors in the MB&B department, especially Mark Solomon, Susan Baserga, Tony Koleske, Sarah Slavoff, Mark Hochstrasser, and Karla Neugebauer, thank you for taking the time to offer words of support, advice, and encouragement. You may not have realized it at the time, but those moments meant the world to me. To the BBSB 1st year class of 2012, you are all amazing and I couldn't ask for a better cohort with which to survive Macro and Methods and Logic. I wish you all the best in your future endeavors and I can't wait to see what amazing things we all do next. To my Irish dance teachers, Patty K. Lenihan, Meghan Lenihan and Erin Collins, as well as the entire Lenihan School, you have become my second family through graduate school. I don't think I would have achieved all

I did in graduate school without dance as an escape to clear my head and come back to my experiments with fresh eyes. To my friends outside Yale, thank you for supporting me through this seemingly crazy journey and listening to my rants about science, whether they were comprehensible or not.

To my parents, Michael and Maureen Duffy, and my sister Meghan, I love you so much. Thank you for supporting me through every failed experiment and celebrating every victory with me. I couldn't have become Dr. Duffy 2.0 without your constant love and encouragement. To my amazing fiancé Michael Lacy, I love you and I'm so glad that this program brought us together as BBSB first year students. I'm not sure what the future holds for us as Dr. and Dr., but I can't wait to share it with you.

Chapter 1

Introduction

1.1 Dynamics of RNA populations

RNA steady-state levels are determined by the tight regulation of RNA synthesis by polymerases and degradation by nucleases. Eukaryotes can modulate RNA equilibrium by altering transcription or decay, thus enabling a rapid and energy-efficient response to environmental stimuli. This dynamic regulatory network can give the same RNA steady-state levels via multiple pathways. For example, a cell can increase RNA levels by increasing transcription or decreasing degradation. Both of these mechanisms are shown to occur upon both lipopolysaccharide (LPS) stimulation in mouse dendritic cells and 4-hydroxytamoxifen (OHT) stimulation in mouse fibroblasts [104, 18], suggesting that cells regulate RNA levels by altering both transcription and degradation in a transcript-specific manner. RNA turnover can be stimulus-specific as well as tissue-specific, highlighting a need to understand the contribution of RNA synthesis and decay, which I will hereafter refer to as RNA dynamics, in a variety of metabolic contexts. This introduction will give an overview of the steps in RNA metabolism and their importance for RNA function, as well as the methods used to understand each step of RNA dynamics.

1.2 RNA synthesis

1.2.1 TFs and the regulation of eukaryotic transcription initiation

TFs bind specific sites on DNA, namely the core promoter, promoter-proximal region, or distal enhancers, thereby activating or repressing the assembly of the RNA polymerase II (RNAPII) initiation complex and modulating RNA synthesis. These DNA-TF interactions can be constitutive or regulated by environmental stimuli, including other TFs. A classic example of an inducible TF is the glucocorticoid receptor (GR), which is typically sequestered in the cytoplasm by chaperones but is able to bind the stress hormone corticosteroids. This in turn causes a conformational change in the protein that releases GR from its chaperone complex and allows GR to translocate to the nucleus where it alters transcription and modulates tissue-specific responses such as the regulation of glucose synthesis and the suppression of inflammation [141, 125, 109].

GR and other TFs have been studied for decades, and chromatin immunoprecipitation (ChIP) studies have mapped the precise binding events for many TFs in many cell types [11]. These studies reveal a complex network TF binding in which a single TF may bind genome-wide to many thousands of sites. However, not all binding events lead to a change in transcription, and the same TF may activate transcription at one binding site and repress transcription in another. Therefore, a highly sensitive approach that monitors RNA at GR-responsive genes would provide insight into the mechanism of GR acivity.

Typically, transcriptional effects of TFs are measured by TF stimulation or depletion, followed by genome-wide RNA sequencing (RNA-seq). Transcriptional response is measured in the presence and absence of TF perturbation and differential RNA expression (typically greater than 1.5–fold or less than -1.5–fold change) is classified as a TF-responsive gene. However, many inducible TFs affect transcription in a matter of minutes, while it may take several hours to see a 1.5-fold change in RNA expression due to the presence of pre-existing RNA before stimulation. This is an even more pronounced problem when TF binding causes gene repression; once RNA synthesis is repressed, the pre-existing RNA must be degraded before changes can be observed by RNA-seq, and the average half-life in mammalian cells is several hours (see below, RNA stability). To complicate matters further, TF induction

may initiate several waves of transcription, with the immediate early genes being the direct consequence of TF binding, but other changes in gene expression changes may arise minutes to hours after a stimulus due to a cascade in transcriptional regulation (i.e. the immediate early gene produces a protein or ncRNA that in turn modulates the expression of different genes). Therefore, temporal resolution is a key factor in understanding the mechanism of TF responses.

One study used nascent RNA detection by global run-on and sequencing (GRO-seq) to study the immediate transcriptional effect of estrogen (E2) stimulation of the estrogen receptor (ER) in breast cancer cells [42]. GRO-seq is an adaptation of classical nuclear run-on assays in which nuclei are isolated from cells, endogenous nucleotides are removed to inhibit transcription and treated with sarkosyl to prevent initiation of new transcription [12]. Nucleoside triphosphates are added back including an additional ribonucleotide analog 5-bromouridine 5'-triphosphate (BrUTP) to allow transcriptionally-engaged polymerases to resume elongation and BrU-label all RNA transcribed during the run-on step. BrU-containing RNA is enriched via immunoprecipitation with anti-BrdU antibodies and sequenced using high-throughput sequencing. GRO-seq, and the later developed PRO-seq that offers nucleotide resolution [63], are advantageous for studying immediate transcriptional responses to TF stimulation because nascent RNAs offer a direct measure of transcriptional changes, and GRO-seq filters out pre-existing RNAs that complicate RNA-seq measurements. This methodology allowed the identification of primary transcriptional targets of E2 stimulation after short treatments (10 and 40 min) rather than the many hours required previously (3-24h) [42]. These data revealed a rapid and transient response to E2 stimulation in breast cancer cells that was not observed previously and is likely a common phenomenon for inducible TFs.

1.2.2 Enhancer RNAs mark active enhancers

Synthesis of mRNAs is driven by TF binding at enhancer regions, which enables looping of enhancers to promoters that stabilizes the transcription initiation complex and promotes RNAPII escape from the promoter [69, 70]. To understand the regulation of RNA synthesis, a crucial component is understanding the activation and repression of enhancers upon

environmental stimulation.

Enhancers are characterized by high levels of the chromatin mark H3K4me1, but low levels of H3K4me3, which denotes gene promoters [47]. Based on this chromatin signature, as well as binding of the general coactivator CBP, Kim *et al.* discovered that thousands of enhancers can recruit RNAPII and transcribe enhancer RNAs (eRNAs) upon neuronal activation [60], a phenomenon that has been confirmed in many tissues and species as an important feature of transcriptional regulation [61]. Additionally, expression of eRNAs is thought to be a good proxy for functional enhancer activity because increases in eRNA expression correlate with increases in mRNA expression from proximal target genes [60].

Enhancer RNAs are difficult to detect in RNA-seq experiments because they are rapidly degraded by the exosome ($t_{1/2}$ ~5 min) and as a result their steady-state levels are very low [120, 105]. Therefore, methods that enrich for transient RNAs are advantageous to detect eRNAs and to understand their regulation in response to environmental stimuli. Indeed, enrichment-based techniques such as GRO-Seq, native elongation transcript sequencing (NET-seq), and transient transcriptome sequencing (TT-seq) have become more common for identifying enhancer RNAs in mammalian cells. [42, 65, 79, 120].

Hah *et al.* used GRO-seq to study the transcriptional response to E2 in breast cancer cells and discovered the upregulation of hundreds of E2-regulated eRNAs that were highly enriched for estrogen receptor α (ERα) binding at enhancers [42]. Subsequently, Li *et al.* demonstrated that these ERα-bound enhancers stabilize enhancer-promoter looping, in part by interacting with cohesin [73]. Conversely, Lam *et al.* found that Rev-Erb nuclear receptors bind enhancers and inhibit the transcription of eRNAs as well as nearby mRNAs [65]. These studies and others highlight the importance of identifying eRNA expression, including potential functional roles for eRNAs, and demonstrate the utility of enrichment-based techniques to study eRNA regulation.

Recently, Schwalb *et al.* adapted short (5 min) metabolic labeling with 4-thiouridine (s^4U) that does not require nuclear isolation and captures enhancer RNAs and other unstable RNA species, known as transient transcriptome sequencing (TT-seq) [120]. Unlike BrUTP, cells rapidly take s^4U up and incorporate it into newly transcribed RNA without the need for nuclear isolation and can be enriched using activated disulfides conjugated to

biotin rather than antibody-based enrichment. TT-seq was later used to study immediate response genes following LPS stimulation in T-cells and revealed that the earliest changes in RNA synthesis between paired enhancers and promoters is simultaneous (within 5 min), contrary to previous reports that observe transcriptional changes many hours after stimulation [56]. Additional studies on the stimulus-dependent kinetics of eRNA transcription should aid our understanding of the role of enhancers and eRNA expression in global RNA synthesis.

1.2.3 RNAPII pausing and elongation are dynamically regulated

Once transcription is initiated by the concerted effort of TF binding, enhancer activity, and transcription initiation complex assembly, transcriptionally-engaged RNA polymerase II (RNAPII) pauses after transcribing 20-65 nt downstream of the TSS [91, 12] before the polymerase is released by P-TEFb and other factors and productive elongation begins [52]. Cells dynamically regulate the rate of RNA synthesis by altering the rate of RNAPII initiation and elongation in response to signaling [14], and RNAPII elongation rates have been shown to vary more than four-fold across transcripts in mammalian cell culture [15, 30, 95, 133]. In addition, rates vary within genes, and productive elongation has been shown to increase from ∼0.5 kb/min to 2-5 kb/min after ∼15 kb [53]. These rates are correlated with processes that regulate gene expression involving co-transcriptional splicing, termination, and RNA stability [118]. Faster RNAPII elongation rates are negatively correlated with splicing efficiency [87], and slow elongation rates can favor alternative exon skipping [24]. In addition, faster elongation rates correlate with epigenetic features such as increased H3K79me2 and H4K20me1 density and DNA methylation, suggesting that gene-specific rates are regulated epigenetically [139]. Hazelbaker *et al.* demonstrated that increased RNAPII elongation rate increases the frequency of read-through transcription at specific terminators in yeast [45], and Gromak *et al.* demonstrate that RNAPII pausing proximal to the the poly(A) cleavage site promotes transcriptional termination in mammalian cells [39]. These results illustrate the regulatory role that variation in RNAPII elongation rates among genes and within a single gene play in gene expression.

Sites of RNAPII pausing have been observed by ChIP-seq, GRO-seq, and PRO-seq

[2, 52], but the measurement of RNAPII elongation rates requires quantifying the distance that the polymerase travels as a function of elapsed time. This is achieved by induction of signaling pathways or synchronization of polymerase with small molecules such as DRB 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB). Upon washout of DRB, non-canonical nucleosides such as BrU and s^4U are added to metabolically label the newly transcribed RNA, and cells are harvested at multiple times following washout. Because RNAPII is synchronized at the pause site during DRB treatment, the distance RNAPII travels directly measures the elongation rate, and rates can be measured genome-wide [139, 30]. RNAPII elongation rates calculated by this method are unable to detect variations in elongation within genes, and only relatively long genes can be used. Regardless, genome-wide DRB-seq methods offer improvements over previous methods that use RT-PCR [124] or fluorescent labeling [15] to measure elongation rates of only a handful of genes at a time. Understanding the mechanism for dynamic regulation of polymerase elongation in a variety of genes and cell types will help give a more complete picture of how RNA synthesis is regulated.

1.3 RNA degradation

For understanding how RNA metabolism modulates cellular RNA steady-state levels, transcriptional regulation is only part of the story. RNA degradation is critical to maintain RNA homeostasis, respond to environmental stimuli, and eliminate nonfunctional or defective transcripts [36]. The majority of mRNAs decay by digestion of the polyA tail, followed by 5' decapping and degradation by exonucleases. Degradation commonly occurs in the 3' to 5' direction by a complex of enzymes called the exosome, although the exonuclease Xrn1 degrades mRNA 5' to 3' (reviewed by [143]). This RNA degradation pathway, while nearly universal for mRNAs, is highly regulated and transcript-specific, giving vastly different RNA half-lives. Much research has been devoted to understanding the RNA motifs that lead to differential RNA stability, which includes AU-rich elements (AREs) in the 3'UTR that recruit ARE binding proteins and facilitate mRNA decay via interaction with the exosome [89]. Conversely, RNA binding proteins such as HuR and polyA binding protein (PABP) have been shown to block access of the exosome and Ccr4, respectively, to mRNAs

and increase their stability [27, 138].

Initial studies suggested that the average half-life of mRNA in mammalian cells is several hours [123, 43] and rapidly-degraded RNAs ($t_{1/2}$ <1h) may be functionally related [103, 43]. More recently, genome-wide microarray and high-throughput sequencing studies have confirmed that the median mRNA half-life in mammalian cells is several hours [147, 132, 21] while a small percentage of transcripts (∼250 mRNAs, or 5%) have very rapid half-lives ($t_{1/2}$ <2h) in many cell lines [147].

1.3.1 mRNA stability is tied to function

Not surprisingly, the coordinated regulation of mRNA half-life has profound consequences on protein levels. This connection between function and RNA half-life has been demonstrated using functional assignment of human and yeast mRNAs based on gene ontology (GO) enrichment [147]. mRNAs for genes involved in transcription regulation, e.g. TFs, tend to be more rapidly degraded than those associated with biosynthesis, which decay slower [147]. In addition, transcripts involved in the same signaling pathway or in a multi-protein complex also tend to decay at similar rates [142]. Modeling studies have also shown that transcripts that are induced rapidly are degraded rapidly, which is especially important for induced genes because they need to be rapidly up- and down-regulated in response to environmental stimuli. Therefore, RNA half-life has been shown to be critical for function; those for transcriptional regulatory factors and genes that must be rapidly induced have a short half-life, whereas transcripts responsible for biosynthesis, whose levels must remain relatively constant for the cell to remain in homeostasis, have much longer half-lives and therefore their steady-state levels tend to fluctuate less [147].

1.3.2 Noncoding RNAs with extreme stabilities

Although much work has been devoted to understanding the relationship between mRNA stability and protein function, the last decade has seen significant developments in the understanding of noncoding RNAs (ncRNAs) and their functions in a variety of physiological contexts. Interestingly, some ncRNAs have been shown to have even more extreme half-lives compared to mRNAs that are related to their function (or potentially lack of function

in some cases). For instance, eRNAs can have extremely short half-lives ($t_{1/2}$ ~5 min), which is hypothesized to limit any eRNA function to be in cis, as these molecules would be degraded before they could diffuse away from their transcription site and act in trans [74]. Conversely, microRNAs (miRNAs) are ~22nt transcripts that regulate mRNA levels and are highly stable, with half-lives on the order of hours to days [40]. This stability is gained when miRNAs form a complex with the Ago2 protein, which allows miRNAs to recognize target RNAs and mediate their translational repression or degradation. Interestingly, more recent reports demonstrate that miRNAs can be cleaved in the presence of their target RNA, a phenomenon known as target-mediated miRNA degradation that enables the rapid turnover of specific miRNAs [115]. However, in most cases, Ago2 protects miRNA ends from degradation by exonucleases, leading to their unusually high stability [4]. The longest reported half-life belongs to TERC, the RNA component of the telomerase RNP, which is estimated to be 3-4 weeks in cancer and stem cells [149], although the mechanism for this exceptionally high stability remains unknown as the protein subunits of telomerase cannot entirely account for this phenomenon [97].

1.3.3 Methods to study RNA stability

Historically, genome-wide RNA stability was most commonly measured by monitoring RNA levels after transcriptional shutoff with the small molecules actinomycin D (ActD), 5-6-dichloro-1-D-ribofuranosyl-benzimidazole (DRB) or α-amanitin (α-Am) [64, 8]. These drugs block global transcription, and therefore rates of RNA degradation can be measured by quantifying total RNA over time following transcriptional arrest. While these methods have been extremely powerful for understanding principles of RNA stability, global transcriptional shutoff has profound physiological effects that can confound measurements of RNA stability. For instance, some transcripts can be rapidly stabilized following ActD or DRB treatment [122, 29]. In addition, highly stable transcripts such as miRNAs are not sufficiently degraded before cells die due to prolonged drug exposure, so the precise half-lives of these transcripts can only be estimated [40]. Therefore, RNA stabilities measured by transcriptional shutoff must now be complemented with techniques that allow continuous transcription and cellular proliferation.

More recently, metabolic labeling experiments have become a common way to measure RNA stability as they minimally perturb cells. 5-bromouridine (BrU) and 4-thiouridine (s^4U) are the most common nucleosides for metabolic labeling when measuring RNA stability as they allow active RNA transcription and processing, unlike 5-ethynyluridine (EU) which is toxic to cells after prolonged exposure [131]. Both BrU and s^4U can be used in a pulse-chase strategy, where newly synthesized transcripts are labeled with either nucleoside for several hours, followed by chase with excess uridine to rapidly diminish the incorporation of BrU or s^4U. Total RNA is extracted over time following the addition of the uridine chase, and metabolically labeled RNA is enriched and sequenced. The percentage of labeled transcript relative to steady-state abundance is interpreted identically to the percentage of transcript that remains following transcriptional shutoff. However, metabolically labeled cells are able to actively transcribe and proliferate during the chase, allowing a more physiologically accurate measurement of RNA stability.

1.4 Tissue specificity of RNA dynamics

Differences in RNA levels between tissues represent different chromatin state and cellular metabolism, as well as tissue-specific TF activity. While TF activity is known to be crucial to tissue identity, the mechanism by which TFs achieve tissue specificity is still not well understood. Recent high-throughput studies suggest that the combinatorial activity of TFs, rather than TF expression levels, modulate tissue-specific TF activity [126]. In addition, non-canonical DNA interactions driven by cellular context are thought to comprise the majority of tissue-specific transcriptional processes [126]. In addition, post-transcriptional modifications have been shown to regulate RNA degradation in a tissue-specific manner [72]. Therefore, while measuring the steady-state levels of RNAs in different tissues is important to understand tissue specificity in gene expression, RNA metabolic labeling experiments can provide additional information about the relative contribution of RNA synthesis and degradation to tissue-specific gene expression.

1.5 Overview

The broad goal of my doctoral research was to develop metabolic labeling methods using more efficient chemistry to capture 4-thiouridine-labeled RNAs (s^4U-RNAs) and to apply these methods to understand RNA dynamics in both cell culture and primary cells. My first aim was to establish methane thiosulfonate (MTS) as a more efficient activated disulfide for the capture of s^4U-RNAs and to apply this chemistry to the study of genome-wide microRNA stability in HEK293T cells with the bioinformatic assistance of Michael Rutenberg-Schoenberg, Robert Kitchen, and Mark Gerstein. The details of this chemistry and the results are described in Chapter II, and a detailed protocol is included in Appendix A. I demonstrated that MTS chemistry improves yield and reduces biases that were previously associated with s^4U metabolic labeling, and this chemistry enabled the study of m^6A-mediated RNA stability in a class of IL-7 regulated transcripts in mouse T-cells as a collaboration with the Flavell lab, as well as the extension of TT-seq in combination with nucleoside conversion chemistry (TT-TimeLapse-seq) to filter out non-s^4U background as a collaboration with Jeremy Schofield in the Simon lab. These results, as well as a description of the many types of experiments available to interface with MTS chemistry, are described in Chapter III. In Chapter IV, I describe the development of a one-step MTS resin for the purification of s^4U-RNA from small numbers of cells and applied this resin to the first measurement of RNAPII elongation rates in mouse cortical neurons as a collaboration with the Maniatis lab. A detailed protocol is included in Appendix B. Finally, in Chapter V, I apply MTS chemistry to study transcription immediately following GR stimulation with dexamethasone (dex) in A549 cells, as well as preliminary data in collaboration with the Koleske lab that these dynamics could be studied in mouse hippocampus slices. Collectively, these results provide a diverse set of insights into the regulation of RNA synthesis and stability in a variety of cell types.

Chapter 2

Tracking distinct RNA populations using efficient and reversible covalent chemistry

This chapter is an excerpt from:

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., Simon, M.D. (2015) Tracking distinct RNA populations using efficient and reversible covalent chemistry. *Mol. Cell* 59(5):858-66. doi: 10.1016/j.molcel.2015.07.023.

2.1 Author Contributions

I performed all experiments, with the assistance of Catherine Stark in Figure 2.2A. Michael Rutenberg-Schoenberg performed all bioinformatic analysis with discussions and input from myself, Matthew Simon, and Mark Gerstein. Robert Kitchen performed bioinformatic analysis of miRNA sequencing with Michael Rutenberg-Schoenberg.

2.2 Summary

I describe a chemical method to label and purify 4-thiouridine (s^4U)-containing RNA. Methanethiolsulfonate (MTS) reagents form disulfide bonds with s^4U more efficiently than

the commonly used HPDP-biotin, leading to higher yields and less biased enrichment. This increase in efficiency allowed the use of s^4U-labeling to study global microRNA (miRNA) turnover in proliferating cultured human cells without perturbing global miRNA levels or the miRNA processing machinery. This improved chemistry will enhance methods that depend on tracking different populations of RNA, such as 4-thiouridine-tagging to study tissue-specific transcription and dynamic transcriptome analysis (DTA) to study RNA turnover.

2.3 Introduction

RNA is continuously transcribed and degraded in a tightly regulated and transcript-specific manner. The dynamics of different RNA populations can be studied by targeted incorporation of non-canonical nucleosides. These nucleosides can provide a chemical handle for labeling and enriching RNA subpopulations. The labeling of RNA employs 5-bromouridine (5-BrU; [132]), 5-ethynyluridine (5-EU; [49]), and 4-thiouridine (TU or s^4U; [9, 86], which provide different vehicles for antibody detection, cycloaddition reactions, and thiol-specific reactivity, respectively. 4-thiouridine holds the advantage that labeling is covalent, unlike the antibody detection of 5-BrU, and also that the disulfide bond is reversible, unlike the click chemistry used to label 5-EU (reviewed in [131]).

Methods to enrich s^4U-incorporated RNA (s^4U-RNA) initially relied on organomercurial affinity matrices [82], but the use of s^4U in metabolic labeling expanded after HPDP-biotin, a 2-pyridylthio-activated disulfide of biotin, was developed as a practical means to biotinylate s^4U-RNA using reversible disulfide chemistry, followed by enrichment using a streptavidin matrix [9, 20]. The s^4U-RNAs can be eluted by reduction of the disulfide linkage and subsequently analyzed by microarray, qPCR, or deep sequencing. This modified protocol sparked a surge in techniques that use s^4U metabolic labeling. For example, half-lives of specific RNAs can be measured using s^4U metabolic labeling by quantifying the ratio of pre-existing (flow through) to newly transcribed (elution) RNA [20]. This approach has been extended to genome-wide analysis using high-throughput sequencing (s^4U-seq; [104]). Combining s^4U metabolic labeling with dynamic kinetic modeling has led to the development of dynamic transcriptome analysis (DTA; [85]), and comparative dynamic transcriptome analysis

(cDTA) when using *S. pombe* standards for normalization, which allows the determination of absolute rates of mRNA synthesis and decay [130]. Reversible transcriptional inhibition has been combined with s^4U metabolic labeling to measure transcriptional elongation rates [30]. Recently, s^4U metabolic labeling has been used with approach-to-equilibrium kinetics to determine absolute RNA degradation and synthesis rates based on multiple time points after s^4U labeling (RATE-seq; [92]). In addition to these methods for analyzing RNA turnover, the enrichment of s^4U-RNA can also be used to determine cell-type-specific transcription (4-thiouridine tagging), which is particularly helpful for analyzing the transcriptomes of cell types that are difficult to isolate by dissection or dissociation methods [86].

As the efficient chemical modification of s^4U is central to all of these techniques, I tested the reactivity of s^4U with HPDP-biotin. Here I report that the reaction and corresponding enrichment of s^4U-RNA with HPDP are inefficient. Therefore, I developed and validated chemistry using activated disulfides to label and enrich s^4U-RNA. This chemistry increases labeling yields and decreases enrichment bias. Due to the increased efficiency of this chemistry, I was able to extend s^4U-metabolic labeling to the study of microRNAs (miRNAs), providing insight into miRNA turnover in proliferating cells without inhibition of miRNA processing pathways. These studies expand the utility of s^4U in metabolic labeling applications and provide the foundation for clearer insight into cellular RNA dynamics through the improvement of all the methods listed above.

2.4 Design

Chemistry to enrich s^4U-RNA should satisfy several considerations. First, the chemistry should be efficient, leading to high yields of labeled s^4U residues. To maintain the advantages of reversible covalent chemistry, I focused on activated disulfide reagents, which allow reductive release after enrichment. This labeling chemistry should be rapid, minimizing time required for purification and decreasing RNA degradation during handling. Finally, the chemistry needs to be specific for s^4U and should not react with RNA that lacks thiol groups. These improvements would lead to a more robust protocol for s^4U-RNA isolation.

Additionally, optimized chemistry could allow the extension of labeling to small RNAs including miRNAs. Smaller RNAs are expected to be particularly sensitive to the efficiency of s^4U labeling, as they tend to have fewer uridine residues and therefore have lower probability of successful labeling. To develop chemistry that meets the above criteria, I first used simple chemical systems to determine the reactivity of activated disulfides. I studied the specificity of labeling chemistry using synthetic RNA with and without s^4U. I used metabolic labeling experiments together with RNA sequencing (RNA-seq) to test the application of this chemistry in the context of complex RNA samples. Finally, I evaluated the use of this chemistry to study miRNA turnover, revealing fast- and slow-turnover miRNAs in proliferating cells without perturbing miRNA processing pathways.

2.5 Results

2.5.1 Optimizing labeling chemistry using free nucleosides

To examine the reactivity of s^4U-RNA with HPDP-biotin, I first studied the labeling of the s^4U nucleoside using liquid chromatography coupled to mass spectrometry (LC-MS; Figures 2.1A and B). I found biotinylation of the s^4U nucleoside with HPDP-biotin to be inefficient when using buffer conditions that are commonly used in the retrieval of s^4U-RNA [38]. This inefficiency stems from the forward and reverse disulfide exchange reactions (Figure 2.1A). Any disulfide formed with the electron-poor pyrimidine ring of s^4U results in a more activated product, therefore favoring the reverse rather than the forward labeling reaction. For this reason, it is not surprising that HPDP-biotin is an inefficient reagent for disulfide exchange with s^4U. Improving this chemistry would expand the utility of s^4U, improve the sensitivity of s^4U labeling, and reduce bias in s^4U-RNA enrichment.

Of the numerous activating chemistries used to make asymmetric disulfides [50, 57], thiosulfates and alkylthiosulfonates are particularly attractive (Figure 2.1C). I found that, in sharp contrast to the slow and inefficient reaction with HPDP-biotin, methylthiosulfonate-activated biotin (MTS-biotin) reacts efficiently with s^4U, leading to >95% conversion to the mixed disulfide within just 5 minutes (Figure 2.1D). I validated this difference in s^4U reactivity between MTS reagents and 2-pyridylthio-activated disulfides using NMR (Figures

2.1E). While only a minority of s^4U reacted using 2-pyridylthio chemistry (<20%), MTS chemistry led to >95% conversion of s^4U to the mixed disulfide.

2.5.2 Extending MTS labeling chemistry to s^4U-RNA

This MTS chemistry was used to specifically fluorescently label s^4U-RNA in the context of cell extracts (Figure 2.2A). Furthermore, the use of MTS-biotin leads to superior biochemical enrichment of s^4U-RNA in comparison to HPDP-biotin (compare flow through to eluent in Figures 2.1F and G) or thiosulfate-biotin (TS-biotin, Figures 2.2D). Importantly, MTS and HPDP chemistries are specific for enrichment of s^4U, as no significant enrichment of RNA without s^4U occurred in either case (Figures 2.1F and G). I therefore conclude that MTS chemistry provides a specific and highly efficient means of detecting and biochemically purifying s^4U-RNA.

I next tested the efficacy of MTS biotin as a reagent to examine newly transcribed RNA in HEK293T cells (Figure 2.3A). Cells were treated with s^4U-supplemented media and reacted the isolated RNA with either HPDP-biotin (as described previously by [38]) or MTS-biotin. Biotinylated RNA was enriched, and the resulting RNA was analyzed by RNA-seq in collaboration with Michael Rutenberg-Schoenberg. To compare the RNA-seq reads across experiments, I used a normalization approach developed by Sun et al. [130] in which the same amount of RNA from *S. pombe* is added to each sample prior to constructing the library for RNA-seq. Consistent with my prior analysis, compared to HPDP-biotin, the use of MTS-biotin led to significantly greater normalized coverage of the human transcriptome (Figures 2.3B and C, 2.4B). This enrichment was reproducible across biological replicates (Pearson's $r = 0.92$, Figure 2.4A) and was validated by qPCR (Figure 2.4C). To test the specificity of MTS chemistry, Rutenberg-Schoenberg examined MTS-biotin-treated RNA from cells that had not been treated with s^4U and found substantially fewer normalized reads than with either HPDP-biotin or MTS-biotin-enriched s^4U-RNA (Figures 2.3B and 2.3C). The result from this control experiment validated the specificity of MTS-biotin for metabolically labeled s^4U-RNA.

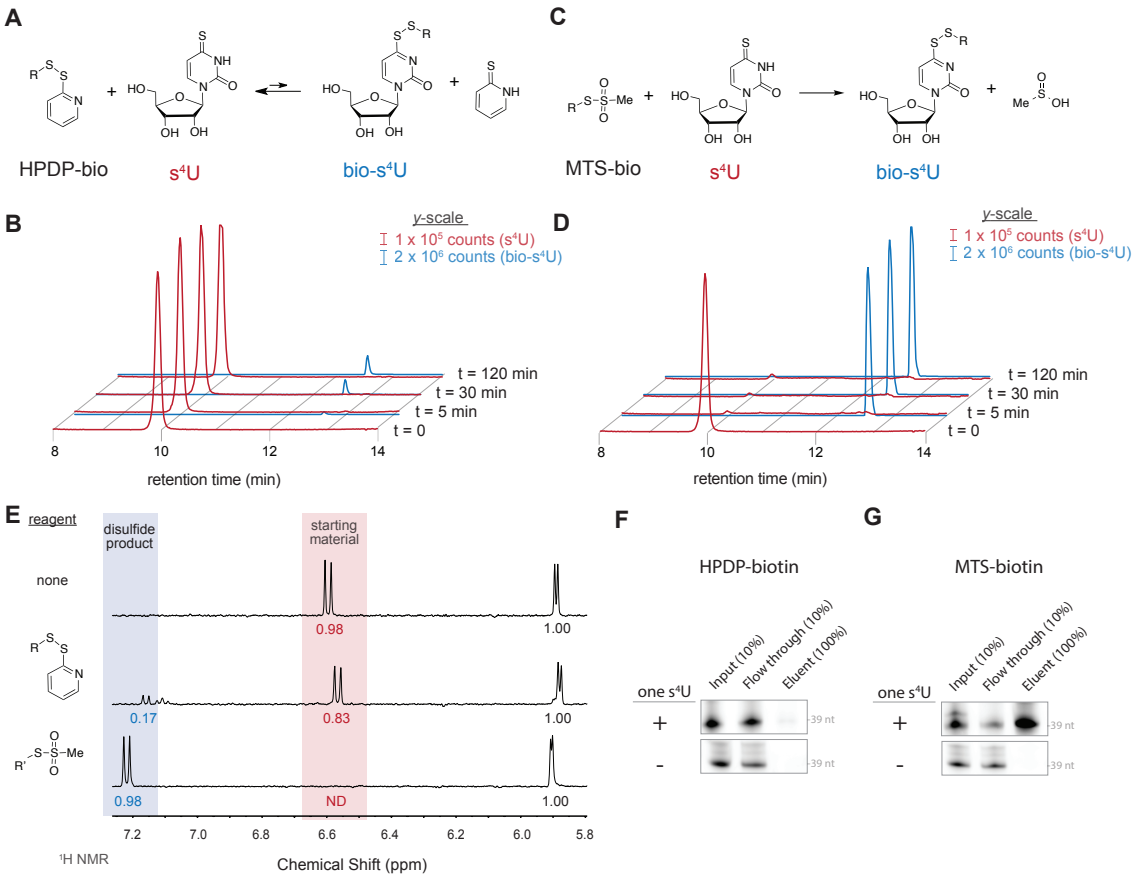

Figure 2.1: **Efficient formation of disulfides with s^{4}U via MTS chemistry.** (A) s^{4}U disulfide exchange with HPDP-biotin. (B) LC-MS extracted ion chromatograms of s^{4}U (red) and biotin-s^{4}U (blue) for HPDP-biotin at the indicated reaction times. (C) s^{4}U disulfide exchange with MTS-biotin. (D) LC-MS chromatograms as in (B). (E) Downfield 1H NMR spectra of (top) s^{4}U alone, (center) s^{4}U reacted with 3-[2-Pyridyldithio]propionyl hydrazide (PDPH), an HPDP-like disulfide, and (bottom) methyl-MTS. Peaks for the starting material (red shading) and products (blue shading) were integrated and normalized to the sum of the anomeric protons of s^{4}U and its products (5.9 ppm). (F and G) Enrichment of a singly thiolated 39-nt RNA by HPDP-biotin (F) or MTS-biotin (G). Fluorescently labeled 39-nt RNAs with or without a single s^{4}U were biotinylated with the indicated reagent and enriched on streptavidin beads, followed by urea-PAGE and fluorescence imaging.

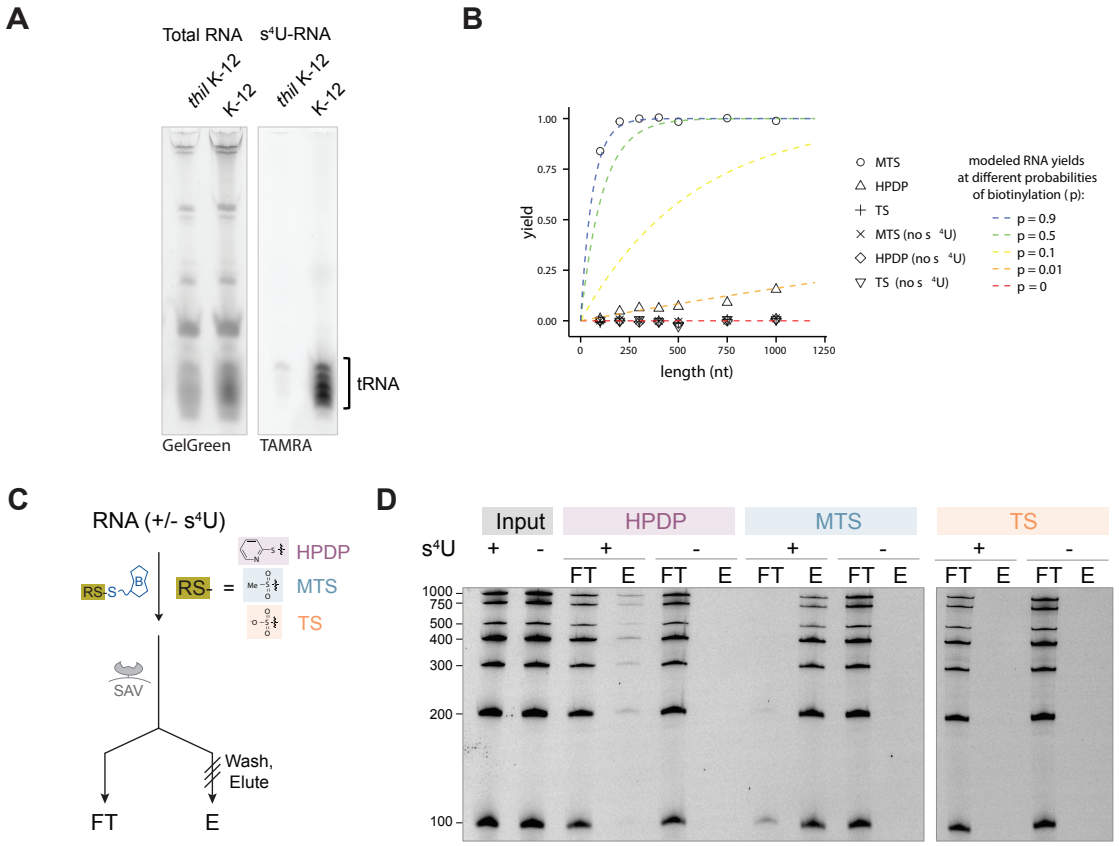

Figure 2.2: **Reactivity of activated disulfides with s^{4}U and *in vitro* modulation of bias in MTS- and HPDP-biotin enrichments.** (A) RNA from *E. coli* K-12 cells was reacted with MTS-TAMRA fluorescent dye and visualized on a 5% urea-PAGE gel. K-12 cells express ThiI, an enzyme that selectively modifies U8 of tRNA to s^{4}U8 [88]. RNA from a *δthiI* knockout shows little TAMRA signal (traces of unmethylated 2-thiouridine on tRNA can still react), whereas a strong TAMRA signal is present in the K-12 cells only in tRNA. Total RNA was stained with GelGreen. (B) Comparison between the yields observed in (D) and expected enrichment using models that assume different biotinylation efficiencies. In all cases modeled lines assume ratio of $s^4U/U_{total} = 0.075$ to determine the expected yield given different biotinylation efficiencies (y_{bio}) based on the equation:

$$yield_{RNA} = \sum_{j=0}^{N_i} [1 - (1 - y_{bio})^j] p(U_i = j)$$

In comparison to the models results, empirical yields using the band intensities from (D) were plotted based on transcript length. (C) Schematic of *in vitro* enrichment of s^{4}U-RNA using an RNA ladder. An RNA ladder was *in vitro* transcribed with Cy5-CTP and with or without added s^{4}UTP. s^{4}U-RNAs were enriched by reacting with disulfide-activated biotin derivatives using either HPDP, MTS, or thiolsulfonate (TS, an alternative disulfide activated biotin reagent) chemistry. (D) Input, flow-through, and elution RNAs were analyzed by urea-PAGE and visualized by Cy5 fluorescence. Band intensities were quantified using ImageJ.

2.5.3 Alleviating length bias using MTS-biotin

Next, the distributions of enriched RNAs using MTS- and HPDP-biotin were compared. Purification of s^4U-RNA using HPDP- biotin is reported to bias enrichment toward longer RNAs that tend to contain increasing numbers of uridines, hereafter referred to as length bias [86, 85]. This bias was confirmed in my study (Figure 2D). While this bias can be partially mitigated statistically [86, 85]), more fruitful biochemical enrichment is clearly preferable, especially when examining overlapping transcript models of different sizes (e.g., spliced and unspliced, see Supplemental Information). To examine how MTS chemistry impacted the length bias in comparison with other activated disulfides, I used an *in vitro* transcribed RNA ladder with and without s^4U to test the relative yields of RNAs with different lengths. This analysis confirmed the presence of a length bias, and agrees well with modeling results (Figure 2.2B), demonstrating how MTS chemistry largely alleviates length bias in RNA turnover experiments. Indeed, RNA-seq data analysis reveals that MTS-biotin is less prone to length bias compared to HPDP-biotin (Figure 2.3D). For example, long transcripts like MALAT1 (8.7 kb) are isolated by HPDP-biotin and MTS-biotin with approximately equal efficiency, whereas shorter transcripts like SCYL1 and LTBP3 (2.3 kb and 3.4 kb, respectively, when fully spliced) are found at much greater levels in the MTS-biotin pull down (Figure 2.3E).

2.5.4 Studying miRNA turnover using MTS chemistry

Given the substantial increase in s^4U-RNA yields observed when using MTS chemistry, I hypothesized that this chemistry could extend s^4U metabolic labeling to the study of miRNAs. The dynamics of miRNA biogenesis and degradation have gained interest because disruption of miRNA homeostasis is implicated in many diseases, particularly for miRNAs that regulate progression through the cell cycle [7]. Generally, miRNA turnover has been investigated by blocking transcription or by inhibiting miRNA processing, followed by analysis of miRNA stability [4, 33, 40]. These approaches have demonstrated that while many miRNAs remain stable for tens of hours, there are also some miRNAs that turn over much more quickly (e.g., miR- 222). Extending these studies using metabolic labeling would allow

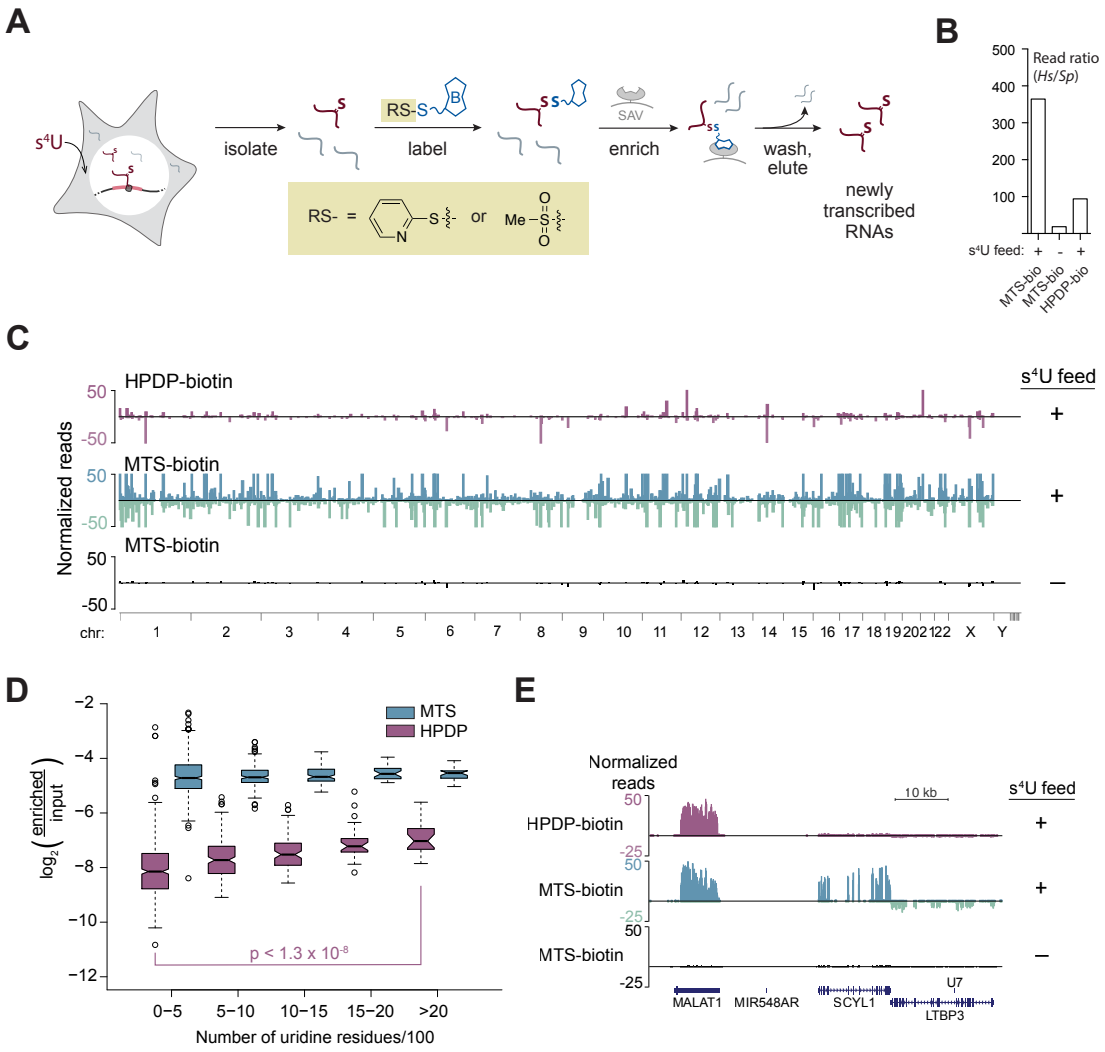

Figure 2.3: **MTS-biotin affords higher specific yields and lower length bias of s^{4}U-RNA** (A) Schematic of s^{4}U metabolic labeling. HEK293T cells were treated with s^{4}U (700 μM) for 1 hr, followed by total RNA extraction, biotinylation with either HPDP- or MTS-biotin, and enrichment on streptavidin-coated magnetic resin. (B) Total reads for each RNA-seq sample that mapped to the *H. sapiens* genome, divided by total number of reads that mapped to the *S. pombe* genome. (C) Whole-genome alignments of eluted samples from HPDP- or MTS-biotin enrichments. y axis indicates number of reads normalized by *S. pombe* spike-ins (see Experimental Procedures). Forward and reverse strand reads are represented as positive and negative values on the y axis, respectively. To compare coverage between samples on the same y axis scale, in some cases, read coverage exceeds the y axis upper limit in MTS-biotin (127 cases) and HPDP-biotin (4 cases). Chromosomes are indicated below the mapped reads. (D) Box plot of transcripts recovered by MTS-biotin and HPDP-biotin binned by transcript length. Blue, MTS-biotin; purple, HPDP-biotin. (E) Examples of genes enriched by HPDP- and MTS-biotin, along with a no s^{4}U-feed control. MALAT1 (8.7 kb), SCYL1 (2.3 kb cDNA), and LTBP3 (3.4 kb cDNA) gene architectures displayed below.

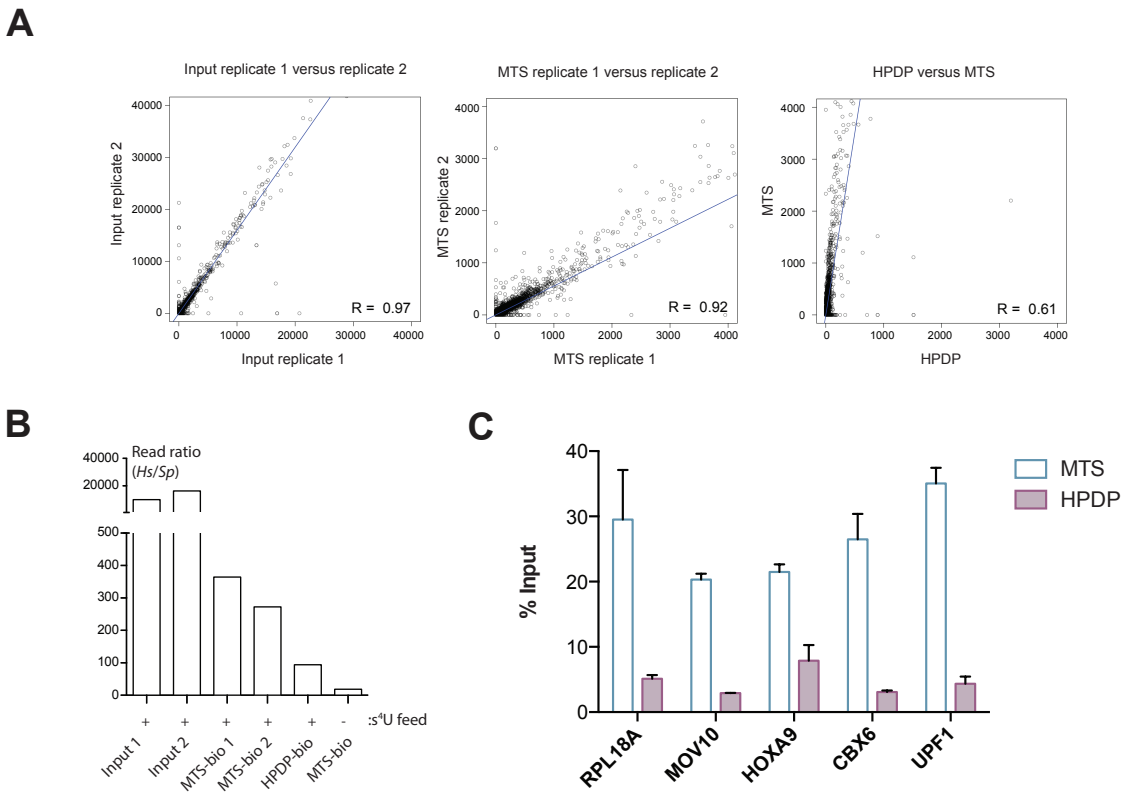

Figure 2.4: **Reproducibility of MTS-biotin enrichment** (A) Scatter plots and Pearson correlations of normalized FPKM values for *H. sapiens* transcript isoforms. Plots show Input 1 vs. Input 2 (left), MTS-biotin replicate 1 vs. HPDP-biotin (center), and MTS-biotin replicate 1 vs. MTS- biotin replicate 2 (right). (B) Total reads for each RNA-seq sample that mapped to the *H. sapiens* genome, normalized by total number of reads that mapped to the *S. pombe* genome, as in Figure 2.3D. (E) Samples enriched by MTS- or HPDP-biotin from RNA-seq submission were analyzed by qPCR using gene-specific primers for RPL18A, MOV10, HOXA9, CBX6, and UPF1 with two replicates. C_t values from qPCR were used to calculate percent input using the equation:

$$\frac{1}{2^{Ct_{sample} - Ct_{input}}} \tag{2.2}$$

where the input is the average of two replicates. Error bars indicate the mean of two technical replicates +/- SEM.

the analysis of native miRNA levels in a proliferating system (unlike those studies using transcriptional block) without perturbing miRNA biogenesis or global miRNA levels (unlike studies where miRNA processing is blocked).

To investigate rates of global miRNA turnover, HEK293T cells were treated with s^4U for a range of times (Figure 2.5A) and enriched s^4U-miRNAs using MTS chemistry, followed by deep sequencing. The miRNA bioinformatic analysis was carried out in collaboration with Michael Rutenberg-Schoenberg, Robert Kitchen, and Mark Gerstein. To test whether s^4U perturbs miRNA steady-state levels, miRNA levels were examined in cells with and without s^4U treatment for 22 days, and high correlations in miRNA levels were found (Pearson's $r = 0.99$, Figure 2.6A), demonstrating that s^4U incorporation has minimal impact on miRNA levels. These findings are consistent with previous accounts that s^4U causes minimal perturbation of longer transcripts [38, 41] and my own data with longer RNAs (Figure 2.6B). Next, s^4U-miRNAs were evaluated at different times after initiating s^4U treatment. MiRNA levels were reproducibly enriched from replicate samples (Figures 2.5C). Furthermore, miRNA levels in neighboring time points were most similar to each other, and those enriched at later time points (1 day, 3 days, and 6 days) approached the levels observed at steady state (22 days). As expected, the steady-state miRNA levels most closely resembled the input miRNA levels (Figure 2.5C).

To determine which miRNAs turned over most quickly, the relative distribution of enriched miRNAs was analyzed at early time points (20 min) versus steady state (6 days or greater; Figure 2.5D). Many RNAs were identified whose relative enrichment was significantly different from steady state at early time points and these miRNAs displayed a consistent trend across time (Figure 2.5E). Fast-turnover miRNAs are expected to be over-represented relative to the population in early time points, and slow-turnover miRNAs are expected to be under-represented (Figure 2.5B). To evaluate this expectation, the analysis took advantage of established properties of miRNA processing (reviewed in [115, 145]. During miRNA biogenesis, one of the two strands from the duplex precursor generally degrades rapidly (referred to here as the miR-star), while the other strand is incorporated into the RNA-induced silencing complex (RISC) and exhibits higher stability. Therefore, the miR-star sequences are expected be over-represented at early time points, and this hypothesis

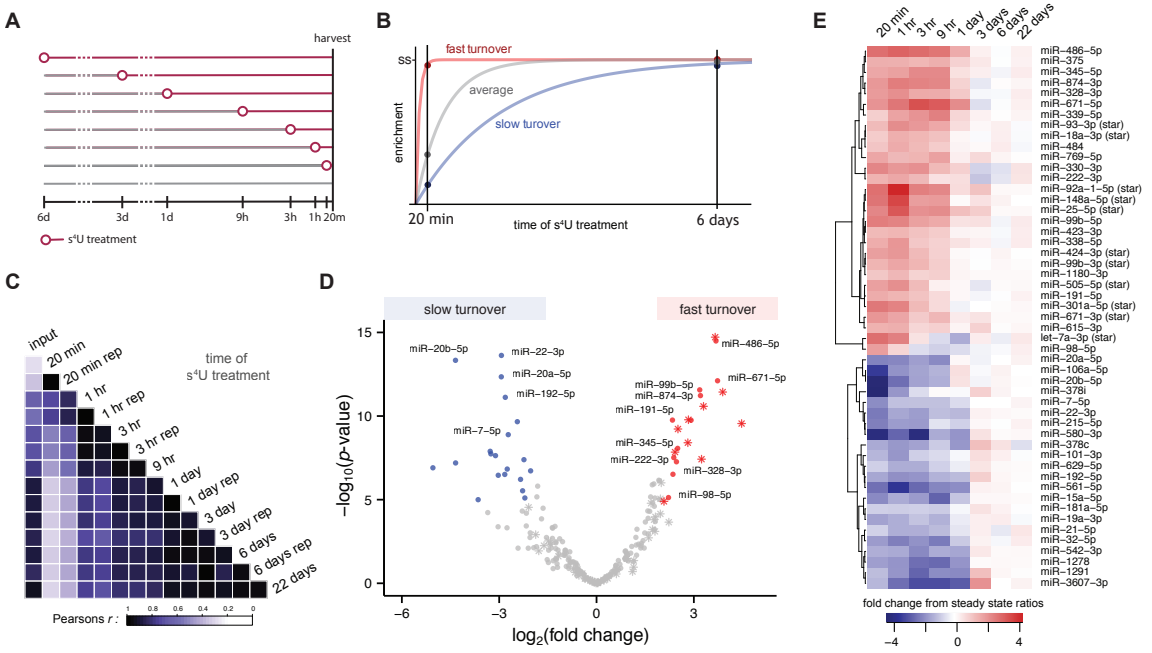

Figure 2.5: **MTS chemistry reveals fast- and slow-turnover miRNAs in miRNA RATE-seq experiments** (A) Schematic of s^4U treatments used in miRNA RATE-seq. (B) Cartoon of anticipated behavior of fast-turnover and slow-turnover miRNAs in comparison to average. Fast-turnover miRNAs are expected to be over- represented in the early time points, whereas slow-turnover miRNAs are depleted, relative to steady state (ss). (C) Heatmap depicting correlation coefficients (Pearson's r) between miRNA levels at different times after s^4U treatment. Replicate samples are indicated by (rep). (D) Volcano plot depicting results from a comparative analysis of miRNAs that are significantly enriched or depleted in early time points (20 min, 1 hr) relative to steady-state levels (6 and 22 days). Fast-turnover miRNAs (fold difference early time points from steady state >4; P <2x10^{-5}; Bonferroni family-wise error rate <0.005) are colored red; slow-turnover miRNAs (fold difference early time points from steady state <0.25; P <2x10^{-5}; Bonferroni family-wise error rate <0.005) are shown in blue. Stars indicate miRNAs defined as miRNA-stars (see Experimental Procedures); the others are indicated with circles. (E) Heatmap indicating normalized miRNA enrichment relative to steady-state level at each time point in RATE-seq for the fast- and slow-turnover miRNAs in (C). For clarity of presentation, the most significant fast-turnover miRNA in this analysis (miR-4521, $\log_2$(fold change) = 10.8; P = 2.9x10^{-40}) has been omitted from (C) and (D) due to values exceeding the indicated scales.

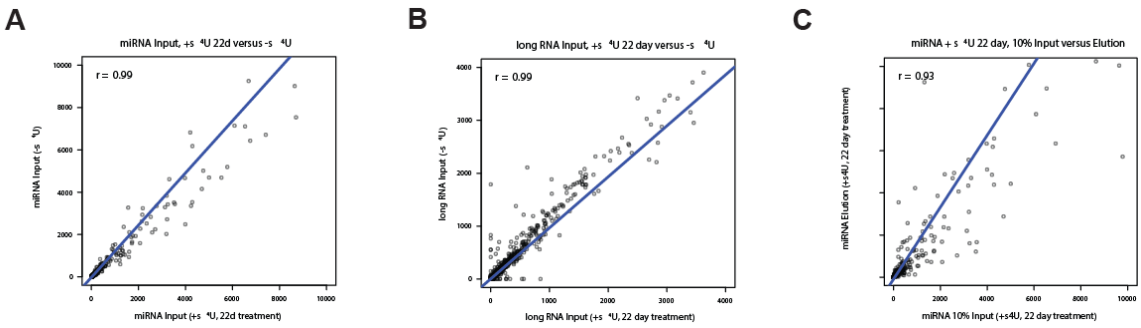

Figure 2.6: **Scatter plots and Pearson correlations of RNA-seq quantifications of *H. sapiens* miRNA transcripts.** (A) Reads from miRNA isolated from cells with no s^4U treatment compared to reads from total miRNA from cells after 22 days of s^4U treatment. (B) Analysis of long RNAs from the same cells as in (A). (C) Analysis of miRNA isolated from 22 day s^4U treatment (10% input) vs. MTS-biotin enriched miRNA from 22 days of s^4U treatment.

was verified: of the 52 significantly enriched and depleted miRNAs (FDR $<5x10^5$), about one-third of the fast-turnover miRNAs were miR-star sequences (11/30), while none of the stable miRNAs (0/22) were annotated as miR-star sequences. The fast-turnover miRNAs identified include miRNAs that agree with previous results using transcriptional blockade (e.g., miR-222; [40]. Other miRNAs were found to be slow turnover (e.g., miR-7), and many of these are also in agreement with past studies [4, 40]. In general, these results using metabolic labeling of miRNAs agree well with results from analysis of degradation after blocking miRNA production [4, 40]. There are exceptions, however, such as miR-98-5p and miR- 191-5p, which were identified as fast-turnover miRNAs in this analysis (Figures 2.5D and E; for a full list of fast-turnover non-star miRNAs, see Table 3.1), yet upon transcriptional blockade these miRNAs are stable [4, 40]. While these results may be due to tissue or cell line differences, it is more likely the faster turnover observed for miR-98-5p and miR-191-5p is due to the cell-cycle regulation of these miRNAs [101, 135]. Turnover in response to progression through the cell cycle is masked when using transcriptional inhibition, but this turnover is evident using a metabolic labeling approach to study miRNA dynamics in dividing cells, underscoring one of the advantages of this improved chemistry.

microRNA	% turnover after 8h (Bail *et al.*, 2010)	% turnover after 4h (Guo *et al.*, 2015)	% turnover after 12h (Guo *et al.*, 2015)
hsa-miR-191-5p	20	23	-16
hsa-miR-671-5p	N/A	-2	39
hsa-miR-1180	N/A	23	7
hsa-miR-328	N/A	-9	5
hsa-miR-769	N/A	31	1
hsa-miR-486	N/A	11	-4
hsa-miR-345	N/A	-5	-14
hsa-miR-484	N/A	-39	-23
hsa-miR-99b	N/A	3	-27
hsa-miR-874	N/A	9	-28
hsa-miR-615	N/A	-15	-32

Table 2.1: Fast-turnover miRNAs that are stable under transcriptional blockade.

2.6 Discussion

Together, these results demonstrate that MTS-biotin is a specific reagent that can be used to efficiently label and enrich s^4U-RNA with higher yields and less bias than the commonly used HPDP-biotin. The dramatic improvement over existing s^4U biotinylation protocols renders MTS chemistry useful for studying dynamics of free nucleosides (Figures 2.1B, D, and E), synthetic RNAs (Figures 2.1F and G), *E. coli* extracts (Figure 2.2A), and s^4U-RNA in metabolic labeling experiments (Figure 2.3). In RNA-turnover experiments, for example, the superior MTS chemistry alleviates transcript length bias, decreases the amount of starting material required, and may allow for the use of lower doses of s^4U to avoid potential toxicities that some have observed [5], but not others [38, 41], when metabolically labeling cells. The utility of this MTS chemistry was demonstrated using miRNA RATE-seq, which allowed the identification of fast- and slow-turnover miRNAs in proliferating cells with flux through the miRNA pathway (Figure 2.5). This advance provides the foundation for more detailed kinetic analyses of miRNA processing and turnover. More generally, applying the chemistry described herein should provide a superior means to gain insights into RNA dynamics in diverse biological systems.

2.6.1 Limitations

This manuscript describes improved capture of s^4U-RNA, but the enrichment will only be successful when the RNA contains sufficient levels of s^4U. In metabolic labeling experiments, incorporation of s^4U into RNA can be controlled by the concentration of s^4U during cell treatment and the time of s^4U exposure. Insufficient s^4U incorporation leads to low yields and will also favor enrichment of longer transcripts that have more uridine residues (and therefore a greater probability of s^4U incorporation).

Chapter 3

Applications of MTS chemistry to improve s^4U metabolic labeling methods

3.1 Summary

Metabolic labeling with s^4U has been used for decades [149, 9, 20, 104], and numerous experimental and computational methods have been developed to adapt s^4U metabolic labeling to the study of many facets of RNA dynamics. For all enrichment-based methods, the MTS chemistry described in Chapter 3 can be applied to improve yield and reduce bias. Here, I discuss in more detail some existing methods using s^4U metabolic labeling that can incorporate MTS chemistry, as well as two examples of MTS chemistry in published work.

3.2 Introduction

3.2.1 4sU-seq

4sU-seq measures rates of synthesis, processing, and degradation genome-wide by using a short metabolic labeling period (10 min) at varying times following cellular stimulation ([104], Figure 2.1A). s^4U-RNA is enriched and sequenced along with total RNA, and the

relative contribution of s^4U-RNA to total is used. RNA steady-state levels represent the equilibrium contribution of RNA synthesis and degradation, whereas newly transcribed RNA contains only RNA synthesized during the s^4U labeling period and therefore represents a "local integration" of average transcription and degradation [104]. During a short (10 min) s^4U labeling period, very little s^4U-RNA is degraded, with the notable exception of highly unstable transcripts, e.g. enhancer RNAs and aberrant transcription. In addition, by performing s^4U metabolic labeling at varying times following cellular stimulation, the transcriptional response can be monitored with high temporal resolution. For instance, Rabani et al. found that many transcripts that are induced upon LPS stimulation are transiently induced, and the peak synthesis rate is observed much earlier in 4sU-seq data compared to RNA-seq, likely due to the delayed observation of changes in steady-state levels upon transcriptional induction [104].

To model synthesis, processing, and degradation rates, Rabani et al. developed a computational approach that infers degradation rates based on steady-state RNA levels from total RNA-seq and newly synthesized RNA levels from 4sU-seq. The simplest model ("constant degradation") assumes that no degradation of s^4U-RNA occurs during the metabolic labeling period, whereas the "varying degradation" model assumes that degradation occurs during labeling. Transcripts that significantly reject the constant degradation model are highly transient, so the degradation rate significantly contributes to changes in steady-state levels. Processing rates can be estimated by comparing the ratio of intronic reads to spliced reads that contain exon-exon junctions. These methods were later refined and formalized in a computational pipeline called DRiLL [105], although users have been unable to use the published code in practice [17] and have developed similar pipelines (e.g. INSPeCT) that are more user-friendly [17].

However, several considerations should be incorporated when using 4sU-seq. Primarily, the enrichment uses HPDP-biotin to enrich s^4U-RNA, but the inefficient reactivity of this activated disulfide toward s^4U significantly biases RNA capture toward longer RNAs ([85] and Figure 2.3), particularly introns. If the enrichment is biased to over-estimate the contribution of unspliced, intron-containing RNA, this would consequently alter the estimation of processing rates. Therefore, the addition of MTS chemistry to 4sU-seq should improve

understanding of the relative contribution of synthesis and degradation in shaping cellular response to stimuli.

3.2.2 RATE-seq and cDTA

While 4sU-seq is ideal to study RNA dynamics of cellular response to stimuli, it can also be desirable to understand RNA dynamics during cell homeostasis. For this purpose, RATE-seq measures RNA synthesis and degradation genome-wide using approach-to-equilibrium kinetics [92]. Cells are labeled in the same growth conditions with s^4U for several time points (Figure 2.5A) and s^4U-RNA is enriched, followed by the addition of multiple exogenous spike-ins to normalize between different amounts of s^4U-RNA after sequencing. The amount of time required for a given transcript to reach steady-state equilibrium reflects the relative contribution of synthesis and degradation (Figure 2.5B), and can be modeled by the following equation:

$$Y(t) = Y_{eq}(1 - e^{-(\alpha_{RNA}+\alpha_{growth})(t-t_d)}) \tag{3.1}$$

where Y(t) is the amount of transcript at time t, Y_{eq} is the abundance of labeled transcript at steady state, α_{RNA} is the transcript's degradation rate constant, α_{growth} is the growth rate constant of the culture, t is the time after addition of label, and t_d is a time delay between the addition of label and the time at which labeled transcripts can be detected. This approach is based on a similar method, comparative Dynamic Transcriptome Analysis (cDTA, [130], an updated version of DTA [85]), in which rates of mRNA synthesis and degradation are estimated by determining the ratio of s^4U-RNA to total RNA normalized to an exogenous spike-in of *S. pombe* RNA. However, cDTA relies on enrichment of s^4U-RNA from a single time point, which may not accurately capture kinetic parameters.

Either method, RATE-seq or cDTA, can be improved with the use of MTS chemistry. Efficient capture of s^4U-RNA is essential to the accurate estimation of synthesis and degradation rates, particularly when comparing enriched RNA levels to total RNA; under-estimation of s^4U-RNA levels will systematically decrease synthesis and increase degradation rate estimates. Therefore, I assert that efficient chemistry is essential for these methods.

While in principle, normalization with exogenous spike-ins, either synthetic or from *S. pombe*, are an effective way to normalize different s^4U-RNA levels from different times of metabolic labeling, in practice I find that the normalization is variable. Specifically, small variability at early RATE-seq time points can significantly influence synthesis and degradation rates in an enrichment-independent manner. For this reason, I hesitate to use RATE-seq or cDTA for quantitative measurement of RNA synthesis and degradation rates, although relative stabilities are easily inferred and should not be affected by normalization to the same degree as absolute quantitation (Figure 2.5).

3.2.3 TT-seq

Transient RNAs, including many noncoding transcripts such as enhancer, antisense, and promoter-associated RNAs, are particularly difficult to detect in RNA-seq experiments because they have very short half lives ($t_{1/2}$ of minutes), which leads to very low steady-state levels. To study these RNAs, very short s^4U metabolic labeling can be used to enrich these transient RNAs from the steady-state pool. Transient transcriptome sequencing (TT-seq) involves 5 min s^4U labeling followed by RNA fragmentation and enrichment to isolate the short 3' region of the nascent transcript that is labeled during the labeling period, thus avoiding 5' bias from long pre-existing regions that is associated with 4sU-seq [104]. TT-seq uniformly maps the human transient transcriptome and enriches enhancer RNAs and other transient species with similar efficiency to the more commonly used GRO-seq and PRO-seq [12, 63]. Because s^4U is readily uptaken into cells without the need for nuclear isolation, TT-seq avoids some of the technical challenges and potential biases associated with GRO-seq and PRO-seq. TT-seq can also be used to estimate RNA half-lives in a similar manner to cDTA, although the same caveats with enrichment efficiency apply to TT-seq.

Although the fragmentation step in TT-seq avoids some of the length bias issues associated with 4sU-seq and other metabolic labeling experiments (described in [85] and shown for HPDP-biotin in Figure 2.5D), TT-seq also requires 50-500 μg total RNA to enrich enough s^4U-RNA for high-throughput sequencing. Therefore, more efficient MTS chemistry would improve yield and decrease the scale required for these TT-seq experiments, making them

more accessible for cell types that are difficult to culture at high scale.

3.2.4 TimeLapse-seq

TimeLapse-seq is a new s^4U metabolic labeling method that does not require enrichment of s^4U-RNAs [119]. Rather, Schofield *et al.* developed oxidative-nucleophilic-aromatic-substitution chemistry that recodes the hydrogen bonding pattern of s^4U to match the hydrogen bonding pattern of cytosine, yielding apparent U-t-C mutations that mark new transcripts upon sequencing. TimeLapse-seq is a single molecule approach that is adaptable to many applications, and reveals RNA dynamics and induced differential expression concealed in traditional RNA-seq. In addition, these results are internally normalized, as both pre-existing and new transcripts are present in the same library. Therefore, TimeLapse-seq represents the new gold standard for quantitative RNA half-life measurements. However, TimeLapse-seq, like RNA-seq, does not easily capture transient RNAs due to their low abundance, therefore an initial enrichment step with MTS chemistry followed by TimeLapse nucleoside conversion is necessary to detect transient RNAs following short metabolic labeling with s^4U.

3.3 Results

3.3.1 MTS chemistry enables the study of m^6A-dependent changes in mRNA stability in naïve mouse T-cells upon IL-7 stimulation

This section is an excerpt from:

Li, H.*, Tong, Y.*, Zhu, S.*, Batista, P.J., **Duffy, E.E.**, Zhao, J, Bailis, W., Cao, G., Kroehling, L., Chen, Y., Wang, G., Chen, Y.G., Broughton, J.P., Kluger, Y., Simon, M.D., Chang, H.Y., Yin, Z. and Flavell, R. (2017) m^6A mRNA methylation controls T cell homeostasis by targeting IL-7/STAT/SOCS pathway. *Nature* 548(7667):338-342. doi: 10.1038/nature23450.

Author Contributions

As a collaboration with Dr. Richard Flavell at Yale University, I sought to apply MTS-based s^4U enrichment to study N^6-methyladenosine (m^6A)-dependent changes in mRNA stability upon IL7 stimulation in naïve mouse T-cells. Huabing Li metabolically labeled WT and *Mettl3-KO* naïve mouse T-cells following IL7 stimulation, after which I purified total RNA from these samples, enriched s^4U-RNA, and prepared high-throughput sequencing libraries from the enriched and total RNA. I analyzed the sequencing data with the help of Matthew Simon.

Results

The most common and abundant messenger RNA modification is m^6A, and its levels are modulated by writers', erasers' and readers' of this mark. *In vitro* data have shown that m^6A influences all fundamental aspects of mRNA metabolism, mainly mRNA stability, to determine stem cell fates. However, its *in vivo* physiological function in mammals and adult mammalian cells is still unknown. Here, members of the Flavell lab showed that the deletion of m^6A writer' protein *Mettl3* in mouse T cells disrupts T cell homeostasis and differentiation. In a lymphopaenic mouse adoptive transfer model, naïve *Mettl3*-deficient T cells failed to undergo homeostatic expansion and remained in the naïve state for up to 12 weeks, thereby preventing colitis. Consistent with these observations, the mRNAs of the Suppressor of cytokine signaling (SOCS) family genes encoding the STAT signaling inhibitory proteins *SOCS1*, *SOCS3* and *CISH* were marked by m^6A, exhibited slower mRNA decay and showed increased mRNAs and levels of protein expression in *Mettl3*-deficient naïve T cells. This increased SOCS family activity consequently inhibited IL-7-mediated STAT5 activation and T cell homeostatic proliferation and differentiation. I used s^4U metabolic labeling at continuous intervals following IL-7 stimulation in WT and *Mettl3*-KO cells in collaboration with Huabing Li (Figure 3.1A) and demonstrated that RNA levels are regulated by m^6A-dependent degradation, rather than changes in synthesis. This experiment was essential to understand the inducible degradation of *Socs* mRNAs in response to IL-7 signaling to reprogram naïve T cells for proliferation and differentiation. This study elucidates the first *in*

vivo biological role of m^6A modification in T-cell-mediated pathogenesis and reveals a novel mechanism of T cell homeostasis and signal-dependent induction of mRNA degradation.

Steady-state levels of mRNA are tightly regulated by both transcription and degradation [131]. While the abundance of most transcripts is mainly controlled by transcription rate, it has been shown that the mRNA levels of a minority of genes (∼17%) are significantly regulated by mRNA degradation rates, notably immediate-early inducible genes [104, 105]. *Socs* genes are well-known immediate-early genes induced upon IL-7 stimulation [150, 99], thus members of the Flavell lab hypothesized that m^6A specifically targets signal-dependent immediate-early genes for degradation. Interestingly, upregulated genes (including *Socs1* and *Socs3*) in *Mettl3*-KO naïve T cells were found to be significantly enriched in the degradation-controlled group of genes from LPS-stimulated dendritic cells (chi square test, P <0.0001, Figure 3.1B). In addition, using a transcriptional inhibition RNA decay assay, mRNAs of all three *Socs* genes were found to be degraded faster upon IL-7 stimulation, as early as 10 min after IL-7 stimulation, compared to control treatment in wild-type cells, whereas the accelerated mRNA degradation upon IL-7 stimulation was abrogated in *Mettl3*-KO naïve T cells (Figure 3.1C). To extend these observations genome-wide and estimate the rates of synthesis and degradation, I conducted a time-course 4sU-seq with IL-7 induction and found a cluster of 34 transcripts including *Cish*, *Socs1* and *Socs3* that were increased in *Mettl3*-KO relative to wild-type cells and show similar kinetics of induction (Figure 3.1A and B). My analysis also confirmed that the estimated degradation rates were lower in *Mettl3*-KO cells for *Socs* transcripts after IL-7 induction (Figure 3.1C and D). These data allowed the Flavell lab to conclude that m^6A targets a group of immediate-early inducible genes including *Socs1*, *Socs3* and *Cish* for rapid mRNA degradation upon IL-7 stimulation, allowing IL-7/JAKs signaling to activate the downstream target STAT5, to initiate the re-programming of naïve T cells for differentiation and proliferation.

3.3.2 MTS chemistry can be combined with nucleoside recoding to filter contaminating RNAs after enrichment

This section is an expert from: Schofield, J.A. **Duffy, E.E.**, Kiefer, L., Sullivan, M.C. and Simon, M.D. (2018) TimeLapse-seq: Adding a temporal dimension to RNA sequencing

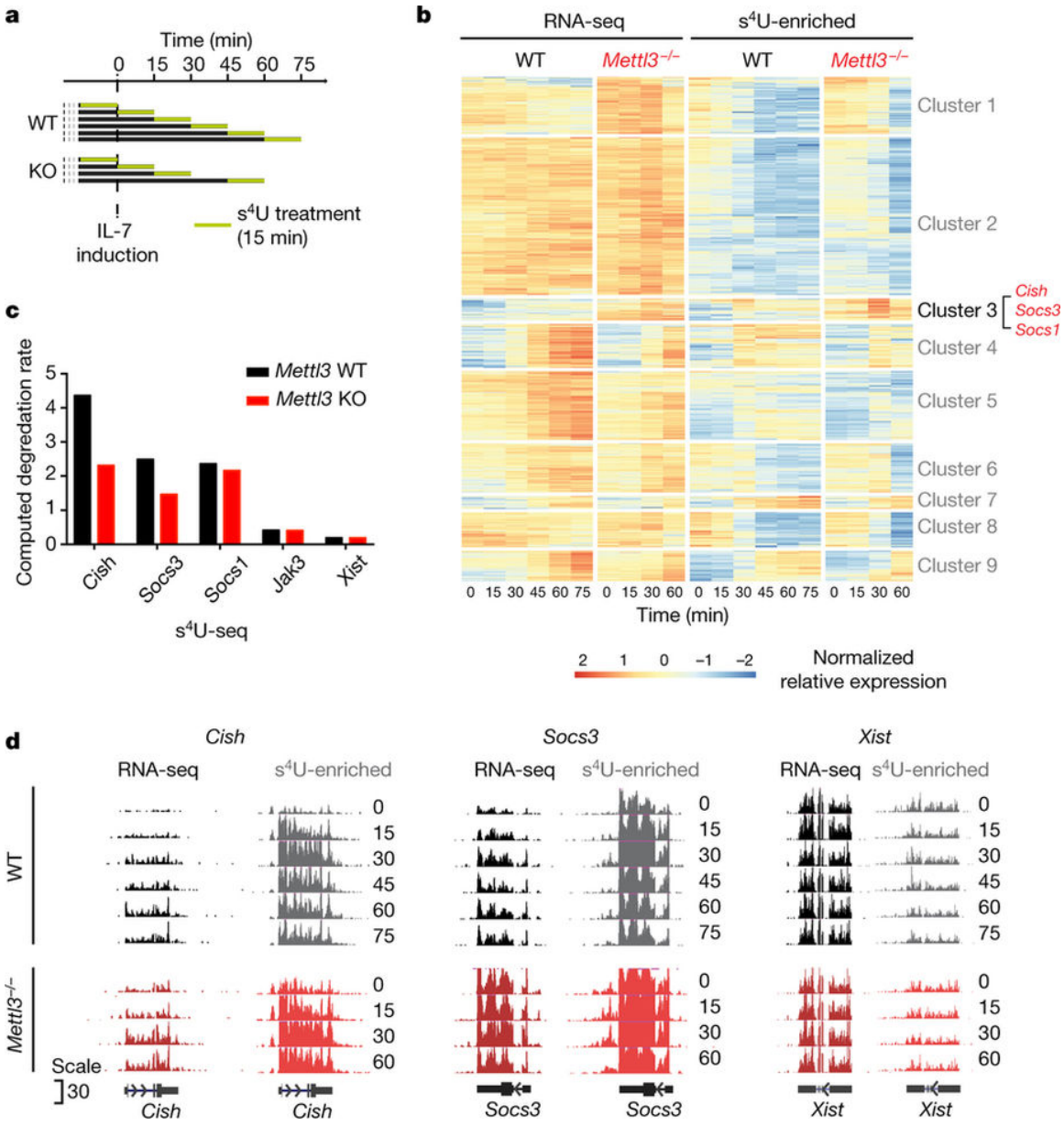

Figure 3.1: **m^{6}A specifically targets a group of immediate-early genes for degradation upon IL-7 stimulation.** (a) 4sU-seq experiment overview. (b) Heatmap showing the results of clustering that normalizes transcript expression levels with significant changes after IL-7 induction and differences between wild-type and *Mettl3-KO*. Cluster 3 contains 34 transcripts with similar expression profiles including *Cish*, *Socs3* and *Socs1*. (c) Computed RNA degradation rates from s^{4}U-seq data. (d) Read density for total RNA and s^{4}U-enriched RNA at the indicated genes for wild-type and *Mettl3-KO* samples after IL-7 stimulation.

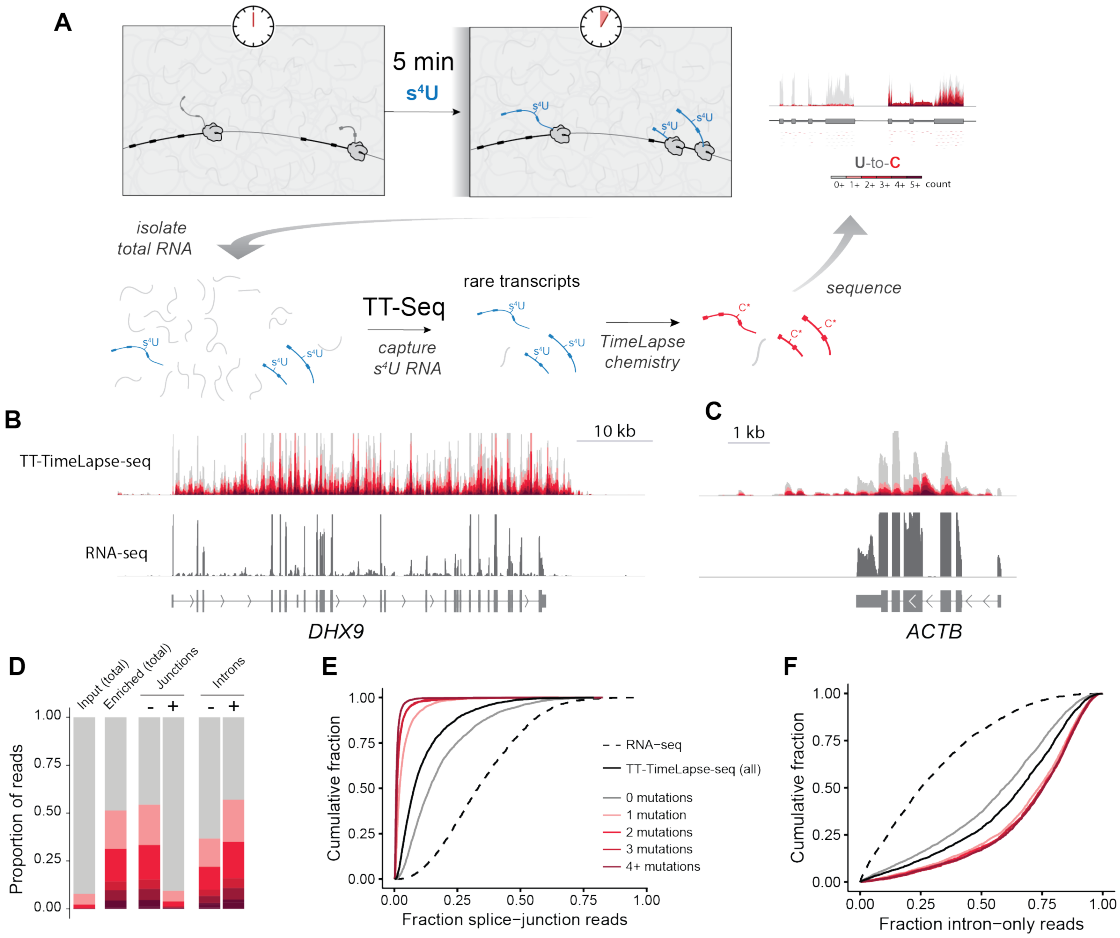

Figure 3.2: **TT-TimeLapse-seq captures transient RNA dynamics.** (A) Scheme of TimeLapse chemistry in conjunction with TT-seq to identify bona fide new transcripts in transient populations of RNA. (B) TT-TimeLapse-seq and RNA-seq tracks depicting coverage from all reads (gray) or reads with increasing numbers of T-to-C mutations (see scale) for *DHX9*. (C) TT-TimeLapse-seq track of *ACTB*. (D)Barplot of the distribution of T-to-C mutations in RNA-seq input, TT-TimeLapse-seq total, or TT-TimeLapse-seq reads filtered for intron or splice-junction content. (E) Cumulative distribution plot of reads containing splice-junctions in RNA-seq, and TT-TimeLapse-seq (black, total; gray-red, separated by mutation count). (F) Cumulative distribution plot of intron-only reads in RNA-seq and TT-TimeLapse-seq with the same scale as in E.

through nucleoside recoding. *Nat. Methods* doi: 10.1038/nmeth.4582.

Author Contributions

In this manuscript, Jeremy Schofield developed novel chemistry which uses oxidative-nucleophilic-aromatic substitution to convert s^4Us into cytidine analogs, yielding apparent U-to-C mutations that mark new transcripts upon sequencing. I hypothesized that this chemistry might interface well with s^4U enrichment to filter nonspecific background that is inherent to all enrichment-based methods. I metabolically labeled K562 cells with s^4U, enriched s^4U-RNA with MTS chemistry followed by TimeLapse chemistry performed by Schofield, and finally I prepared high-throughput libraries from enriched and total RNA. Matthew Simon analyzed the data for U-to-C mutations.

Results

Very transient RNA species, such as reads beyond the poly-A termination signal in a gene body, provide insight into transcriptome dynamics but are generally too rare to be observed at high levels by RNA-seq. While these dynamics can be studied through biochemical enrichment of very recently made RNAs after short (5 min) s^4U treatments (TT-seq [120]), biochemically enriched s^4U-RNA always contains contaminating reads from unlabeled RNAs (estimated to be up to 30% in some experiments [105], particularly in highly abundant RNAs like *ACTB*). This contaminating background can limit analyses; for example, abundant spliced transcripts observed in RNA enriched after short s^4U pulses has been interpreted as fast splicing [90], but these results could also be explained by contaminating background (e.g., from fully spliced mature RNAs). To test if TimeLapse chemistry could be used in conjunction with transient transcriptome sequencing (TT-seq) to distinguish bona fide new RNAs from contaminating background, K562 cells were labeled for 5 min with s^4U, and biochemical enrichment was performed as in TT-seq [120], except with more efficient MTS chemistry to biotinylate the s^4U-RNA [23] (Figure 3.2A). After enrichment and prior to sequencing, enriched RNA samples were treated with TimeLapse chemistry.

As expected, transient RNA species were enriched for introns (two-sample Kolmogorov-Smirnov test, $P<10^{-15}$, Figure 3.2B-D) but depleted for splice junctions ($P<10^{-15}$). Both

enrichment of introns and depletion of splice junctions were slightly greater than previously observed [120], likely due to the efficiency of MTS-chemistry. Even with only 5 min of s^4U treatment, the majority of the biochemically enriched reads contained TimeLapse-induced mutations (Figure 3.2B). Mutation-containing reads represented a subpopulation that was further enriched for introns, and depleted for splice junctions (Figure 3.2E and F). This suggests that mutated reads effectively capture the profile of new RNAs, while the reads without mutations represent a subpopulation that is contaminated by unlabeled reads. An estimated 15-20% of total TT-seq reads arise from contaminating RNA (estimate from splice-junction content: 17-20%; from intronic content: 18-20%, Table 3.1), similar to estimates from previous s^4U experiments [105]. Reads without mutations were enriched for contaminating reads (estimate from splice junctions, 33-39%; estimate from introns, 35-40%), while reads containing mutations are depleted in contamination. For reads with a single mutation, contaminating reads make up <5% of the signal; for reads with two mutations, the contamination is <1%. Taken together, RNA contamination contributes to the signal at the level of RNA-seq, but TimeLapse chemistry-induced mutations can be used to discriminate between signal from new RNAs and contaminating reads. These results demonstrate transcripts including *ACTB* (Figure 3.2C) are not highly spliced on this timescale (5 min), and highlight how TimeLapse chemistry can provide an extra specificity filter when analyzing rare, transient RNAs.

	Est. from junction reads			**Est. from intron-only reads**		
mutations	replicate 1	replicate 2	merge	replicate 1	replicate 2	merge
total	0.2001	0.1214	0.1854	0.1992	0.1787	0.1887
0 mutations	0.3927	0.3343	0.3627	0.3977	0.3525	0.3744
1 mutation	0.0408	0.0330	0.0368	0.0299	0.0271	0.0284
2 mutations	0.0102	0.0077	0.0089	0.0037	0.0054	0.0045

Table 3.1: Proportions of contaminating reads for TT-TimeLapse-seq replicates by mutation count per read estimated through splice-junction or intron analysis

3.4 Discussion

Many s^4U metabolic labeling techniques have been published that use HPDP-biotin as the activated disulfide, and for all of these experiments, MTS chemistry offers an improvement

in efficiency that at worst allows the use of less input RNA, and at best alleviates length bias and enables experiments at a scale that was not previously possible. Depending on the available amount of total RNA and the biological question, s^4U-RNA enrichment with MTS-biotin may be appropriate, or if internal normalization is required and s^4U labeling time is sufficient (<1h), TimeLapse-seq without enrichment may be ideal due to the possibility of internal normalization. Here I present two applications of MTS chemistry, 4sU-seq in naïve mouse T-cells and TT-TimeLapse-seq, which benefitted from improved chemistry. I also describe several additional methods that use activated disulfide enrichment in which MTS chemistry would offer improvements. Indeed, several recent protocols have been published that utilize MTS chemistry for improved enrichment efficiency [116, 136, 1].

Chapter 4

Solid phase chemistry to capture RNA population dynamics in primary mouse neurons

This chapter is an excerpt from: **Duffy, E.E.**, Canzio, D., Maniatis, T. and Simon, M.D. Solid phase chemistry captures metabolically labeled RNAs and reveals RNAPII activity in mouse cortical neurons. (2018) (*under review*).

4.1 Author contributions

I performed all experiments and bioinformatic analyses with input from Matthew Simon. Daniele Canzio cultured mouse cortical neurons, performed s^4U metabolic labeling, and purified total RNA from these cells.

4.2 Summary

Here, I describe an approach to enrich newly transcribed RNAs from primary mouse neurons using 4-thiouridine (s^4U) metabolic labeling and solid phase chemistry. This one-step enrichment procedure captures s^4U-RNA by using highly efficient methane thiosulfonate (MTS) chemistry in an immobilized format. Like solution based methods, this solid-phase

enrichment can distinguish mature RNAs (mRNA) with differential stability, and can be used to reveal transient RNAs such as enhancer RNAs (eRNAs) and primary microRNAs (pri-miRNAs) from short metabolic labeling. Most importantly, the efficiency of this solid-phase chemistry made possible the first large scale measurements of RNA polymerase II (RNAPII) elongation rates in mouse cortical neurons. Thus, my approach lays the foundation to study regulation of RNA metabolism in specific tissue contexts as a mean to better understand gene expression *in vivo*.

4.3 Introduction

Tissue-specific regulation of steady-state RNA levels is achieved through precise control of RNA synthesis, processing, and degradation. These dynamics are critical for control of global transcript levels and for the cell to respond to environmental stimuli in a rapid and energy-efficient manner. Cells are able to regulate the rate of RNA synthesis by altering the rate of RNA polymerase II (RNAPII) initiation and elongation in response to signaling[14]. RNAPII elongation rates have been shown to vary more than four-fold across transcripts in mammalian cell culture [15, 30, 95, 133], and these rates are correlated with processes that regulate gene expression involving co-transcriptional splicing, termination, and RNA stability [118]. Factors including chromatin context (such as the presence of H3K9me3 in the gene body) are also shown to modulate elongation rates [117, 139]. In addition, cell-type specific chromatin state, trans-acting factors and cellular metabolism suggest elongation rates may differ between tissues [96, 126, 129], but these rates have rarely been studied in primary cells or tissues, primarily because the scale of metabolic labeling experiments used to determine these factors is limited by the scale required for biochemical enrichment (ca. 50 μg of input RNA is generally required for these protocols [20, 34]). I sought to develop more efficient biochemical capture of metabolically labeled RNA to understand RNA metabolism in primary cells, including mouse cortical neurons.

Metabolic labeling allows identification of newly synthesized RNAs by treating cells with a non-canonical nucleoside that get incorporated during RNA synthesis. To determine which transcripts are new, total RNA is isolated and new, labeled RNAs are enriched. The

three most widely used RNA metabolic labels are 5-bromouridine [BrU], 5-ethinyluridine [5-EU], and 4-thiouridine [s^4U]). Depending on the nucleoside, labeled RNAs are can be captured with a specific antibody (BrU) [132], or biotinylated using either click chemistry (5-EU) [49], or thiol-specific reactivity (s^4U) [9, 20, 23], respectively. Enriched RNAs can be analyzed by RNA-seq, qPCR, or microarray. Protocols based on s^4U are advantageous because s^4U is rapidly incorporated into the cellular NTP pool in living cells without the need for cell lysis and nuclear isolation. In addition, s^4U shows minimal perturbation to cellular physiology at low concentrations even after long treatment, and allows both covalent and reversible capture [131]. Therefore, s^4U metabolic labeling can be employed to study the stability of both rare, transient RNAs [120, 144] as well as stable transcripts [149] in the same experiment [116].

Transcriptional elongation can be studied using 4sUDRB-seq, which combines transcriptional inhibition with s^4U labeling [30]. Specifically, cells are treated with 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) to block the release of paused RNAPII into productive elongation, until ongoing RNAPII transcription has completed. After DRB is removed from the media, the wave of new transcriptional elongation can be followed by s^4U metabolic labeling of the new RNAs [30]. This approach has been used to calculate rates of transcription genome-wide in HeLa cells [30], but existing methods to capture new RNA require prohibitively large amounts input RNA material to study new transcription in mouse cortical neurons. I therefore sought to use 4sUDRB-seq with an improved method to enrich s^4U-RNA to calculate elongation rates in mouse cortical neurons where the availability of input RNA is limited.

Recently, I published, along with Michael Rutenberg-Schoenberg, Catherine Stark, Robert Kitchen, Mark Gerstein, and Matthew Simon, more efficient chemistry for the enrichment of s^4U-RNA using the activated disulfide methane thiosulfonate (MTS), which improves yield and decreases biases in metabolic labeling experiments [23, 72]. I identified two additional potential improvements that would expand the scope of s^4U-RNA enrichment: decrease the number of steps in the protocol by directly conjugating the MTS activated disulfide to a solid support, and use higher stringency rinses in the context of a fully covalent system to decrease the amount of contaminating RNA without sacrificing

yield. I synthesized MTS-resin and confirmed it could be used to distinguish fast- and slow-turnover RNAs in a metabolic labeling experiment. The MTS resin also performed well in transient transcriptome sequencing (TT-seq) experiments where only very low levels of RNA are labeled with s^4U [120]. Finally, in collaboration with Daniele Canzio and Tom Maniatis, I used this resin to measure RNAPII elongation rates in mouse cortical neurons using 4sUDRB-seq. These results reveal transcription elongation rates in primary neurons and highlight the diverse applications for small-scale s^4U-RNA purification in the study of RNA dynamics and metabolism.

4.4 Results

4.4.1 Resin synthesis and characterization

Developing solid-phase chemistry to capture s^4U-RNA requires the use of a highly activated disulfide as the loading of the resin limits the extent to which mass-action can be used to drive the chemical reaction. Previously, I demonstrated that methane thiosulfonate (MTS) activated disulfides react more efficiently with s^4U than the commonly used HPDP-biotin, and that this improved chemistry enables more sensitive applications of s^4U metabolic labeling [23]. I hypothesized that coupling the MTS moiety directly to magnetic sepharose beads should decrease the loss that occurs during the multiple steps of a biotin/streptavidin purification and increase fold enrichment of s^4U-RNA by enabling higher stringency rinses. After s^4U-RNAs are eluted by reducing the disulfide bond, these RNAs could be analyzed by microarray, qPCR, or RNA-seq. To test this approach, methane thiosulfonate (MTS) was coupled to a solid support (Figure 4.1) by reacting methane thiosulfonate ethylamine (MTSEA) with commercially available N-hydroxysuccinimide (NHS) activated magnetic beads using a modification of standard protocols (See Methods and Data Analysis). These beads are shown to be capable of capturing s^4U nucleosides as well as metabolically labeled s^4U-RNA (Figure 4.1B, C).

To determine the binding capacity of MTS resin, beads were incubated with a vast molar excess of s^4U nucleoside to saturate the available sites on the resin. After rinsing and elution, the amount of the captured nucleoside was quantified. Binding was found to

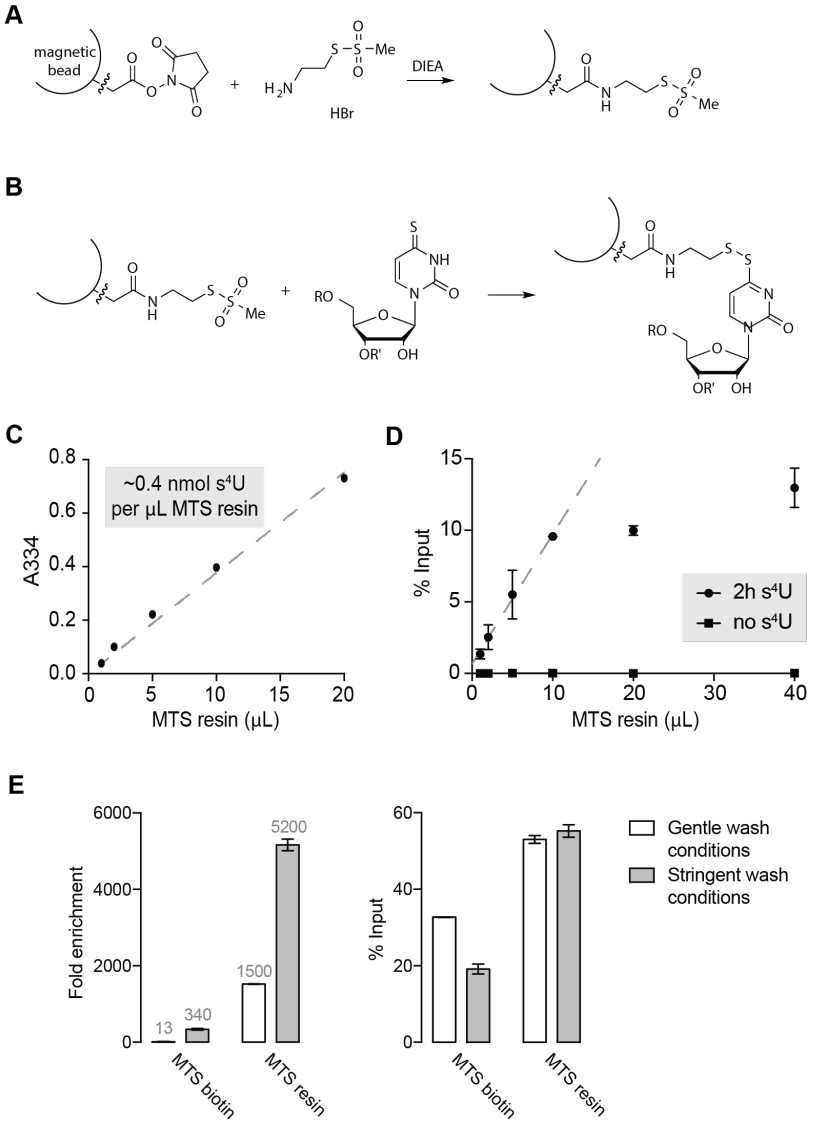

Figure 4.1: **MTS resin covalently and reversibly enriches s^{4}U-RNA** (A) Scheme of MTS resin synthesis. (B) Scheme of MTS resin reactivity with s^{4}U. (C) Quantification of MTS resin loading capacity with free s^{4}U nucleoside. Molar excess of s^{4}U was enriched with increasing volumes of MTS resin, and eluted nucleoside concentration was measured by A334 absorbance. Binding capacity was calculated by linear regression of absorbance and converted to nmol s^{4}U based on a standard curve. (D) Quantification of MTS resin loading capacity for total RNA. Total RNA from K562 cells was metabolically labeled with 1 mM s^{4}U for 2 h (or no labeling as a -s^{4}U control) and reacted with increasing volumes of MTS resin. Enriched RNA was quantified by RT-qPCR using primers for *CDKN1B* RNA. Percent input was quantified relative to a 10% input RNA sample. (E) Comparing enrichment by MTS biotin and MTS resin methods. Metabolically labeled total RNA from HEK293T cells was spiked with non-labeled RNA from *S. pombe* as an internal non-s^{4}U control and enriched using MTS biotin [22] or MTS resin using gentle or stringent wash conditions (described in Methods and Data Analysis). Fold enrichment (left panel) was calculated as s^{4}U-RNA transcript enriched from the human sample (%Input$_{Hs}$) divided by signal from the background sample (%Input$_{Sp}$). Percent input (right panel) was calculated as in (D). Error bars in (C, D and E) represent SD of two technical replicates.

be linear with increasing quantity of resin, and the resin can capture 0.4 nmol s^4U per μL MTS resin slurry (Figure 1C). While this loading capacity provides an upper bound of the amount of s^4U-RNA that can be captured, other factors such as site accessibility could provide practical limits to the resin loading in a biological sample. Therefore, metabolically labeled RNA from K562 cells (2 h, 1 mM s^4U) was used to test the relationship between amounts of MTS resin slurry used and the amount of s^4U-RNA retrieved (Figure 4.1D). An increase in s^4U-RNA enrichment was observed with increasing amounts of MTS resin (plateauing at 10 μL resin per μg of input RNA). To measure nonspecific background, exogenous unlabeled RNA was added from *S. pombe* (1 ng unlabeled standard per 1 μg input labeled RNA), and this background did not significantly increase even at the highest amounts of MTS resin tested.

Next I compared the sensitivity of this small-scale s^4U-RNA enrichment (1 μg input RNA) on MTS resin to previously published protocols using MTS-biotin [23, 22] and tested the effect of stringent rinses on both purifications (Figure 4.1E). The retrieval of a metabolically labeled mRNA (*CDKN1B*) was measured in comparison with an unlabeled RNA from *S. pombe*. While stringent rinses improve the fold enrichment in MTS biotin experiments (13-fold enrichment with standard rinses, 340-fold with stringent rinses), the enrichment is significantly increased with MTS resin (1500-fold enrichment with standard rinses, 5200-fold with stringent rinses). In addition, while the stringent rinses decrease s^4U-RNA yield in MTS biotin purification (33% input with standard rinses, 19% input with stringent rinses), a significant difference in yield is not detected with MTS resin enrichment stringent rinses (53% input with standard rinses, 55% input with stringent rinses). These results demonstrate that high levels of enrichment can be achieved when using an entirely covalent purification made possible with the MTS-resin.

4.4.2 4-thiouridine pulse-chase labeling (s^4U Chase-Seq)

After MTS resin was synthesized and shown to specifically capture s^4U-RNA, I next tested if MTS resin could be used to distinguish fast turnover RNA populations from slower turnover populations, a common application of s^4U metabolic labeling experiments [9, 20, 85, 104]. A pulse-chase experiment similar to Yi *et al.* [149] was chosen, where the cells are exposed

to a s^4U pulse (1 mM, 2 h) followed by a chase phase where the cells were treated with 20 mM uridine (Figure 2A). Those RNAs that were labeled with s^4U at 30 min were compared with the population of s^4U-RNA that remained after 18 h to detect long-lived RNAs. A short 2 h pulse was chosen that leads to only small amounts of s^4U-RNA, which provides a stringent test of the resin. To optimize this s^4U-chase labeling, longer treatments could be used, as s^4U can be present in cells for several days without affecting RNA dynamics due to its low toxicity [9, 149]. To ensure this experiment could be performed on a scale compatible with cells from primary tissues, RNA from only $2x10^6$ cells was used for the enrichment. This low quantity of input RNA (1 μg of total RNA) contrasts with the higher quantities (generally >25 μg) used in previous experiments using s^4U [28, 149, 151].

High-throughput sequencing was used to analyze s^4U-RNA from both the 30 min and 18 h chase that had been enriched with MTS-resin. Counts of the aligned reads revealed consistent enrichment across biological replicates (triplicate, Pearson's $r = 0.87$ - 0.97, Figure 4.3A). The relative stability of each transcript could be inferred based on the relative levels of different RNAs in the 0.5 h versus 18 h s^4U populations (Figure 4.2B, C). Fast-turnover mRNAs were expected to be under-represented relative to the population in the 18 h chase and slow-turnover mRNAs were expected to be over-represented. Previously identified fast- and slow-turnover transcripts [29, 147] showed expected s^4U profiles in genome browser tracks (Figure 4.2B, 4.3B). To determine which RNAs displayed relatively fast or slow turnover compared to the rest of the transcriptome, I performed differential expression analysis on the s^4U-RNAs enriched after a short chase (0.5 h) versus a long chase (18 h) with uridine (Figure 4.2C). Using standard differential expression analysis pipelines (DESeq2, P_{adj} ¡ 0.05), over a thousand RNAs were identified as enriched in each category (1757 RNAs as fast turnover; 1571 RNAs as slow turnover). Because mRNA stability has been tied to protein function [29, 147], transcripts were analyzed by GO-enrichment analysis (Figure 4.3D), and enriched GO-terms for fast- and slow-turnover transcripts were consistent with previous GO-enrichment analysis of mRNA half-life data [147]. For example, transcripts encoding TFs and histone mRNAs are enriched in the list of fast turnover RNAs [100, 147], whereas mRNAs for proteins involved in biosynthetic processes were enriched in the slow turnover RNAs. In addition, the $\log_2$ fold difference between s^4U-RNA in the 0.5 h and 18

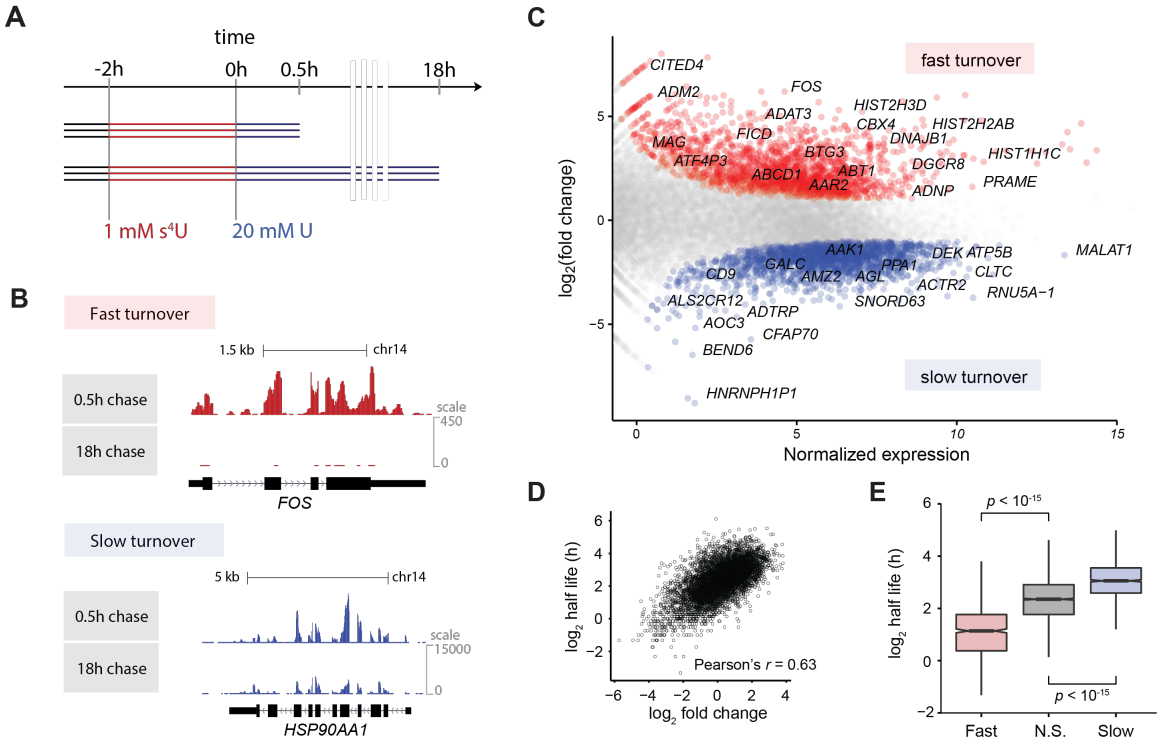

Figure 4.2: **s^{4}U Chase-seq identifies fast- and slow-turnover RNAs in K562 cells**. (A) Schematic of s^{4}U Chase-seq. K562 cells were metabolically labeled with 1 mM s^{4}U for 2 h, followed by a 20 mM uridine chase for 0.5 h or 18 h. (B) Example genome browser view of s^{4}U-RNA enriched after 0.5 h and 18 h uridine chase. *FOS* and *HSP90AA1* were identified as fast and slow turnover, respectively, by differential expression analysis. Reads from biological triplicates were summed for display. (C) Scatterplot of fold change versus normalized expression based on comparative analysis of RNAs that are significantly enriched or depleted in early time points (0.5 h chase) relative to late time points (18 h chase). Fast-turnover RNAs (fold difference >2; P <2x10^{-5}) are colored red; slow-turnover RNAs (fold difference <0.5; P <2x10^{-5}) are shown in blue. (D) Scatterplot of log$_2$ fold change from (C) versus log$_2$ RNA half-life from Friedel *et al.* [29]. (E) Box and whisker plot of log$_2$ RNA half-life from Friedel *et al.* for RNAs binned as fast turnover, slow turnover, or not significant in (C). RNAs were filtered for normalized expression >2 (logCPM).

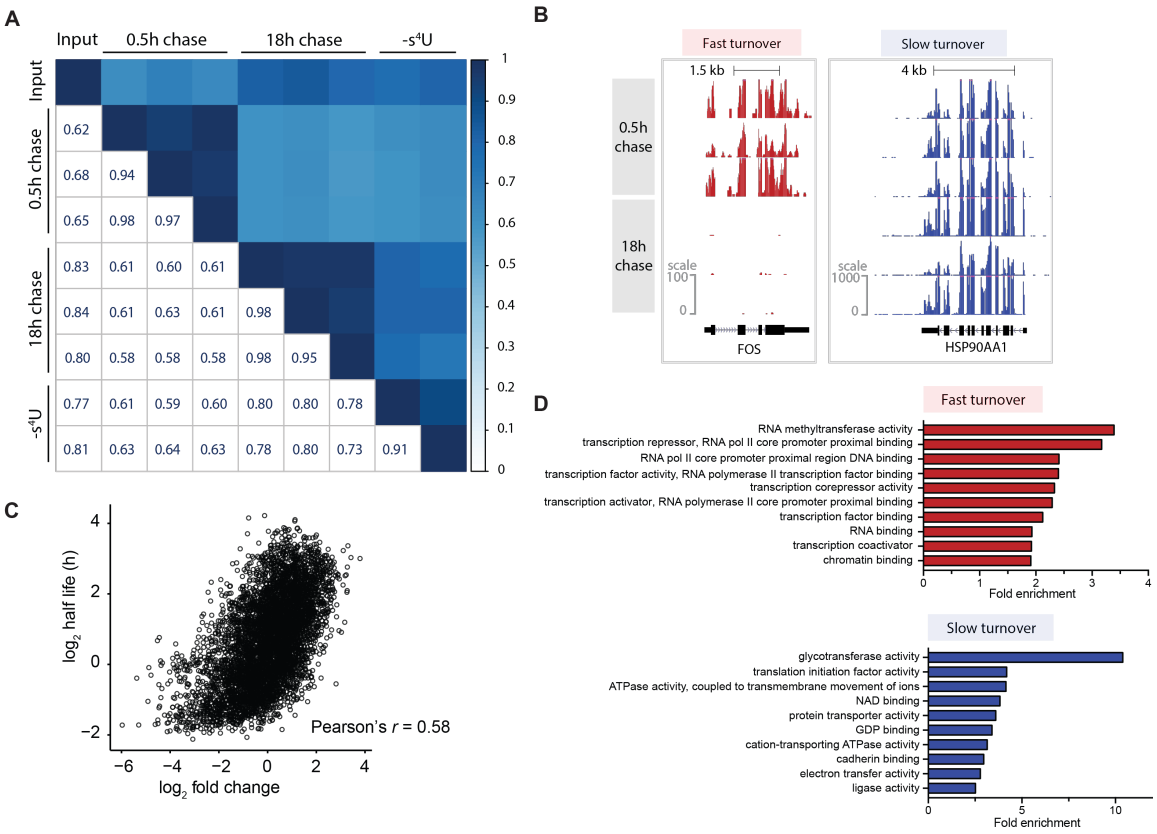

Figure 4.3: **Fast- and slow-turnover RNAs correlate well and are enriched for biological functions.** (A) Heatmap of pairwise correlation coefficients between biological triplicates of s^4U Chase-seq, as well as RNA-seq and RNA enriched from K562 cells without s^4U treatment (-s^4U) as a negative control. Pearson's r correlations are shown. (B) Example genome browser view of three biological replicate samples of s^4U-RNA enriched after 0.5 h and 18 h uridine chase. *FOS* and *HSP90AA1* were identified as fast and slow turnover, respectively, by differential expression analysis. (C) Scatterplot of $\log_2$ fold change from (C) versus $\log_2$ RNA half-life from Schofield *et al.* [119]. (D) Significantly enriched biological functions for fast- and slow-turnover RNAs identified using PANTHER.

h chase correlates with genome-wide RNA half-lives calculated by Friedel *et al.* (Pearson's r = 0.63, Figure 4.2D) and Schofield *et al.* (Pearson's r = 0.58, Figure 4.3C) [29, 119]. Slow-turnover transcripts had a significantly longer half-life on average compared to transcripts that were not identified as significantly fast- or slow-turnover (P <2.2x10^{-16}); conversely, fast-turnover transcripts had a significantly shorter half-life on average (P <2.2x10^{-16}; Figure 4.2E). These results demonstrate that MTS resin can be used to purify s^4U-RNA from small numbers of cells and can distinguish fast and slow turnover transcripts.

4.4.3 MTS resin-based transient transcriptome sequencing (MTS-TT-seq)

For applications such as measuring transcription elongation rates in primary cells, a s^4U-RNA capture method must perform well at small scale and also have the sensitivity to capture rare, newly made RNAs from the higher concentrations of pre-existing RNA pool. For this reason, MTS resin was tested in the context of a transient transcriptome sequencing (TT-seq) experiment (Figure 4.4A), which uses very short s^4U labeling (5 min, 1 mM) followed by RNA fragmentation and enrichment. This approach captures rare RNAs including introns, enhancer RNAs (eRNAs), and pri-miRNAs [120]. Using MTS-resin, this experiment was performed using only 2.5 μg of input RNA, substantially below (<20-fold) the scale that is generally used for these experiments [84, 120]. Enriched RNA samples, as well as input RNA, were analyzed by high-throughput sequencing and reads were mapped to the human genome. Enrichment was consistent across biological replicates (Pearson's $r = 0.99$, Figure 4.5A). As expected if the MTS resin successfully enriched the transient RNA population, the MTS-TT-seq samples correlate better with each other than with -s^4U controls or input, which suggests that the resin is capturing new transcripts over nonspecific background (Figure 4.4B). Consistent with this, MTS-TT-seq samples correlate well with previously published TT-seq data (Pearson's $r = 0.89$, Figure 4.5C). Notably, MTS-TT-seq was found to enrich transient RNA species including pre-mRNA (as revealed by intronic RNA enrichment), eRNAs, and primary microRNAs (pri-miRNAs) relative to input (Figure 4.4C, D), consistent with previous data [120]. Overall, I determine that MTS resin is able to enrich rare RNA populations from small numbers of cells with similar sensitivity compared to existing methods at higher scale.

Unlike other methods to analyze transient RNA species such as GRO-seq and PRO-seq [13, 12], s^4U-based methods do not require purification of nuclei. On the other hand, despite the biochemical challenge of these protocols, GRO-seq and PRO-seq have been more extensively validated than analogous s^4U enrichment techniques. Therefore, I wondered whether data from MTS-TT-seq could be analyzed using analysis pipelines originally developed for these more validated approaches. To test this, a recently developed bioinformatics pipeline

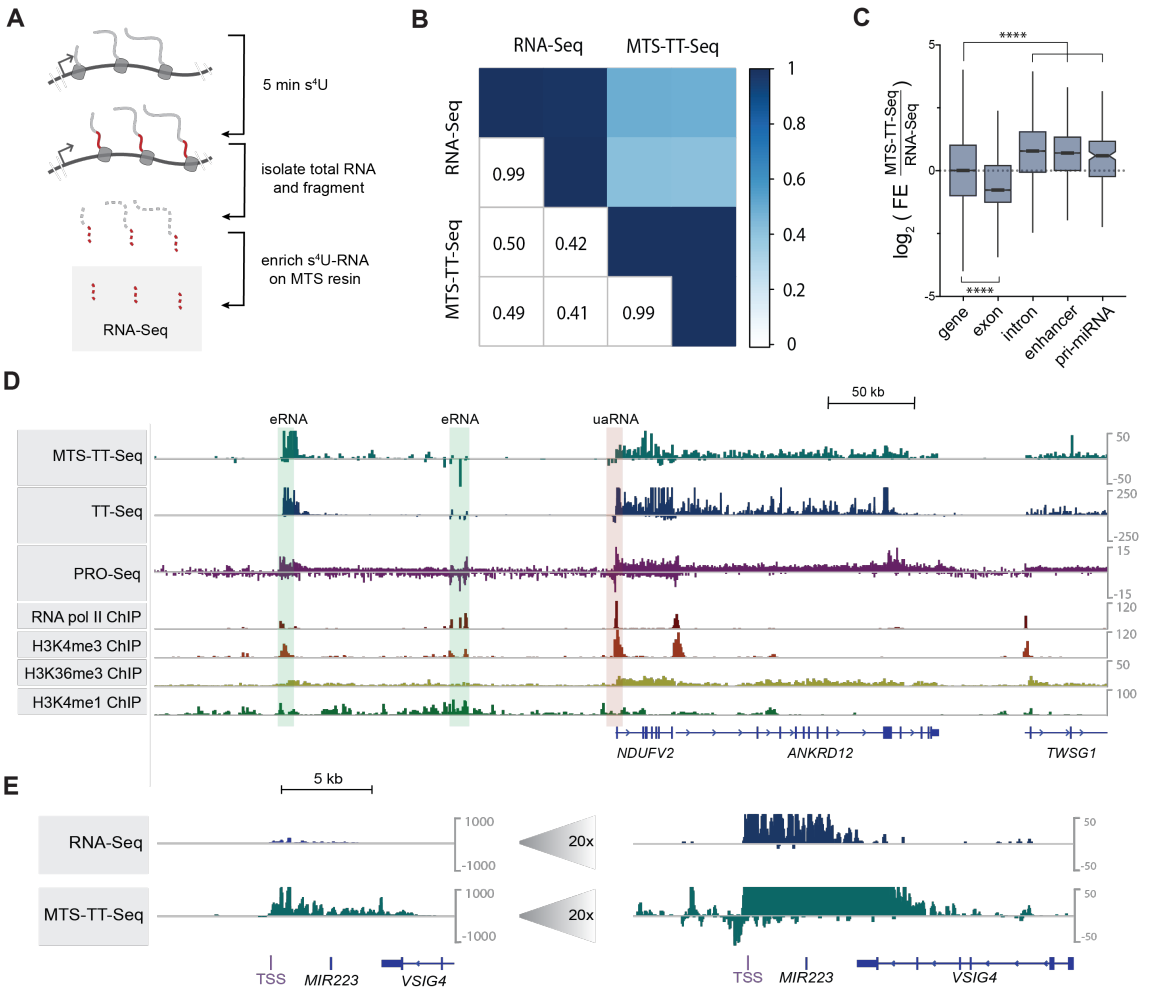

Figure 4.4: **MTS-TT-seq captures unstable RNAs.** (A) Scheme of TT-seq enrichment using MTS resin. K562 cells were metabolically labeled with 1 mM s^4U for 5 min, followed by total RNA isolation and fragmentation to ~200 nt. RNA fragments containing s^4U are enriched on MTS resin and analyzed by RNA-seq (B) Heatmap of pairwise correlation coefficients between MTS-TT-seq (Pearson's $r = 0.99$) and RNA-seq (Pearson's $r = 0.99$) biological replicates. (C) Box plot of $\log_2$ fold enrichment of unstable RNA populations in MTS-TT-seq compared to RNA-seq. **** = P <0.0001. Pseudo-count of 1 was added to all input counts. (D) Example genome browser view showing enrichment of transient species such as introns, antisense transcription (orange box) and enhancer RNAs (green boxes) [chr18:8,836,895-9,370,470] in MTS-TT-seq, TT-seq [120], and GRO-Seq [13], as well as ChIP data from RNAPII, H3K4me3, H3K36me3, and H3K4me1 histone marks [11]. Reads above the x-axis indicate transcription on the forward strand, whereas reads below the x-axis indicate transcription on the reverse strand. (E) Example transcription start site (TSS) of mir223 identified using MTS-TT-seq and the mirSTP pipeline. The left panel shows normalized signal in RNA-seq and MTS-TT-seq data, while the right panel (zoomed in y-axis) shows bidirectional transcription at the TSS in MTS-TT-seq.

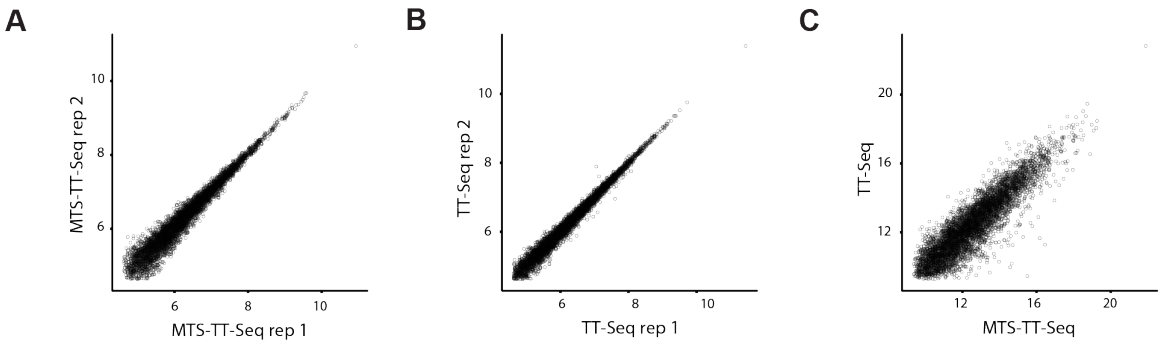

Figure 4.5: **MTS-TT-seq correlates with published TT-seq data** Log-log plot of gene-level quantification in MTS-TT-seq (Pearson's $r = 0.98$) and TT-seq ([120], Pearson's $r = 0.99$) biological replicates, as well as summed replicates of TT-seq and MTS-TT-seq (Pearson's $r = 0.89$). TT-seq data from Schwalb *et al.* were down-sampled to equalize read depth between TT-seq and MTS-TT-seq experiments and common genes ($\log_2$FC <1) in RNA-seq between TT-seq and MTS-TT-seq experiments are displayed.

for GRO-seq and PRO-seq data called mirSTP was used to identify the active transcription start sites of intergenic miRNAs [76]. mirSTP uses divergent sharp peaks around transcription start sites and continuous coverage over active transcription regions to identify intergenic miRNA TSSs. Analyzing my MTS-TT-seq data with this pipeline, 89 miRNA TSSs were identified in K562 cells (Figure 4.4D, complete list of miRNA TSSs in Table 4.1). Of the 89 TSSs discovered by TT-seq, 50 overlapped with the 90 TSSs annotated my mirSTP using GRO-seq data, which is consistent with the proportion of overlap observed between mirSTP and Hua *et al.*, which identifies miRNA TSSs using H3K4me3 and DNase I hypersensitive sites [48, 76]. This demonstrates that GRO-seq bioinformatics pipelines can be successfully used to analyze MTS-TT-seq data. In addition, the smaller scale afforded by MTS resin for TT-seq analysis led us to conclude that this approach will allow the study of transcriptional dynamics analyses in a wider variety of cell types and tissues than was previously feasible.

miRNA	Chr	TSS	Strand	Score_{plus}	Score_{minus}	Pvalue_{gb}	Num_{5k}
hsa-mir-223	X	65236298	+	288.3	71.4	0	NA
hsa-mir-3118-1	1	142652221	+	9.3	6.2	0	228
hsa-mir-3118-3	1	143433376	-	52.7	0.5	0	300
hsa-mir-3180-4	16	15346562	-	10.4	1.5	0	33
hsa-mir-3668	6	140460393	+	1.7	8.2	0	30
hsa-mir-3688-1	4	160107919	-	8.5	2.8	0	50
hsa-mir-3688-2	4	160025218	+	10.0	0.6	0	70

miRNA	Chr	TSS	Strand	Score_{plus}	Score_{minus}	Pvalue_{gb}	Num_{5k}
hsa-mir-4435-1	2	87904295	+	24.2	7.9	0	206
hsa-mir-4465	6	140474813	+	2.1	2.2	0	3
hsa-mir-4509-2	15	28649196	+	10.1	7.5	0	93
hsa-mir-4678	10	89208992	+	3.0	5.9	0	30
hsa-mir-550b-2	7	32796448	-	11.8	0.7	0	64
hsa-mir-6074	12	66516545	-	0.6	0.9	0	NA
hsa-mir-6723	1	569210	-	367.4	842.4	0	NA
hsa-mir-3675	1	17193649	-	1.3	1.9	3.79E-291	255
hsa-mir-4509-3	15	28751703	-	15.3	23.0	1.87E-270	67
hsa-mir-30b	8	135842248	-	5.4	0.6	1.68E-260	51
hsa-mir-6503	11	59992959	-	6.8	3.9	3.46E-236	75
hsa-mir-4652	7	93234007	+	2.7	3.1	3.13E-213	3
hsa-mir-30d	8	135842254	-	5.3	0.5	4.95E-213	54
hsa-mir-1273c	6	155158553	+	3.9	1.5	1.81E-210	66
hsa-mir-4477a	9	68425913	-	9.3	5.2	2.78E-179	53
hsa-mir-4771-1	2	87379801	+	6.0	1.3	1.26E-148	4
hsa-mir-548a-2	6	135540856	+	1.5	14.2	1.07E-144	9
hsa-mir-3680-2	16	29623765	-	2.7	2.9	3.61E-138	69
hsa-mir-1299	9	69014314	-	12.8	6.5	1.30E-133	34
hsa-mir-3680-1	16	21522955	-	3.7	2.8	7.69E-88	98
hsa-mir-29c	1	208003852	-	1.0	6.1	1.95E-82	16
hsa-mir-4509-1	15	22690983	-	8.8	9.7	2.27E-80	27
hsa-mir-4454	4	164037041	-	7.0	0.3	3.05E-80	9
hsa-mir-3916	1	247373344	-	1.9	3.1	4.75E-79	77
hsa-mir-29b-2	1	208003843	-	1.1	6.0	1.43E-76	12
hsa-mir-4660	8	8891964	+	4.2	1.4	1.36E-64	23
hsa-mir-5093	16	85394487	-	4.4	2.5	1.06E-58	11
hsa-mir-10a	17	46661665	-	4.5	0.8	2.17E-44	98
hsa-mir-3975	18	33162420	+	1.7	2.6	1.29E-43	52
hsa-mir-6511a-3	16	16449340	+	1.2	10.0	1.45E-40	33
hsa-mir-573	4	24524070	-	7.0	2.2	1.83E-33	NA
hsa-mir-3671	1	65530013	-	2.6	5.6	3.49E-32	46
hsa-mir-4720	16	81415488	+	2.2	1.0	4.68E-32	NA
hsa-mir-331	12	95697161	+	0.7	1.8	7.71E-31	NA
hsa-mir-101-1	1	65530012	-	2.6	5.6	8.81E-30	50
hsa-mir-3685	12	95701201	+	1.6	2.0	9.99E-27	NA
hsa-mir-563	3	15904938	+	0.7	4.1	1.28E-26	52
hsa-mir-130b	22	22005433	+	5.2	2.0	2.02E-25	NA
hsa-mir-182	7	129422818	-	1.4	1.4	5.20E-23	20
hsa-mir-5089	17	45019568	+	1.5	5.4	7.15E-21	7
hsa-mir-1304	11	93468005	-	5.7	4.7	1.04E-18	NA
hsa-mir-550a-2	7	32770580	+	4.6	1.1	1.15E-18	NA
hsa-mir-4477b	9	68410808	+	1.0	5.0	4.91E-18	NA
hsa-mir-3150a	8	96071459	+	0.4	2.8	8.41E-18	12
hsa-mir-23a	19	13952556	-	1.3	3.4	3.39E-17	35
hsa-mir-27a	19	13952549	-	1.4	3.3	8.27E-17	33
hsa-mir-6089-1	X	2514980	+	1.0	1.0	4.72E-16	22

miRNA	Chr	TSS	Strand	Score_{plus}	Score_{minus}	Pvalue_{gb}	Num_{5k}
hsa-mir-6089-2	Y	2464980	+	1.0	1.0	4.72E-16	22
hsa-mir-301b	22	22005433	+	5.2	2.0	4.49E-15	NA
hsa-mir-3180-1	16	15000459	+	1.2	0.2	4.47E-14	NA
hsa-mir-4712	15	50651018	+	1.2	2.7	9.75E-14	NA
hsa-mir-4445	3	109260349	+	0.7	2.3	5.23E-13	4
hsa-mir-183	7	129422810	-	1.5	1.3	1.32E-12	23
hsa-mir-96	7	129422847	-	1.5	1.3	3.28E-12	23
hsa-mir-505	X	139009542	-	0.7	1.5	4.45E-12	NA
hsa-mir-3939	6	167412470	-	3.1	0.6	2.33E-11	NA
hsa-mir-3945	4	185776352	-	0.3	1.2	3.58E-11	29
hsa-mir-5684	19	12892749	+	0.2	0.6	1.99E-10	NA
hsa-mir-4666a	1	228647714	+	0.8	0.8	3.07E-10	NA
hsa-mir-4734	17	36860190	-	2.0	0.2	3.14E-09	NA
hsa-mir-5692a-1	7	97590279	+	1.2	0.2	6.11E-09	NA
hsa-mir-3683	7	7112790	-	0.9	0.1	7.98E-09	21
hsa-mir-337	14	101330935	+	0.6	0.5	9.13E-08	19
hsa-mir-422a	15	64186264	-	0.8	1.0	1.08E-07	3
hsa-mir-548a-3	8	105498142	-	0.9	3.7	1.28E-07	NA
hsa-mir-548ay	3	32555860	-	5.8	0.5	1.36E-07	17
hsa-mir-3670-1	16	14999269	+	1.1	0.1	2.29E-07	NA
hsa-mir-3670-2	16	16397923	+	1.1	0.1	2.29E-07	NA
hsa-mir-665	14	101330935	+	0.6	0.5	2.84E-07	19
hsa-mir-3677	16	2319008	+	0.3	0.6	1.41E-06	NA
hsa-mir-4786	2	240892617	-	0.6	0.3	1.42E-06	8
hsa-mir-196a-1	17	46719077	-	0.2	0.7	1.92E-06	7
hsa-mir-1324	3	75674701	+	0.5	2.2	3.02E-06	NA
hsa-mir-5692c-1	5	135154799	-	0.2	0.8	3.60E-06	4
hsa-mir-197	1	110139699	+	0.4	3.3	4.44E-06	NA
hsa-mir-3679	2	134883136	+	2.2	0.9	4.51E-06	NA
hsa-mir-4757	2	19546191	+	1.6	2.2	1.66E-05	NA
hsa-mir-4277	5	1714365	-	0.8	2.7	1.87E-05	15
hsa-mir-4496	12	109028275	+	2.1	1.6	4.25E-05	NA
hsa-mir-3179-2	16	16389583	+	0.6	0.5	8.03E-05	NA
hsa-mir-1827	12	100574206	+	0.7	0.1	8.21E-05	13
hsa-mir-940	16	2319078	+	0.2	0.5	9.29E-05	NA

Table 4.1: Annotated miRNA transcription start sites, log likelihood score to estimate the sharp peaks at the sense and antisense strand (Score_{plus} and Score_{minus}, respectively), and the minimum number of reads among the continuous 5kb window (Num_{5k}).

4.4.4 RNAPII elongation rates in mouse cortical neurons

Having established that MTS resin is compatible with s^4U-RNA capture from small numbers of cells, the resin was applied in the context of a 4sUDRB-seq experiment to study

Figure 4

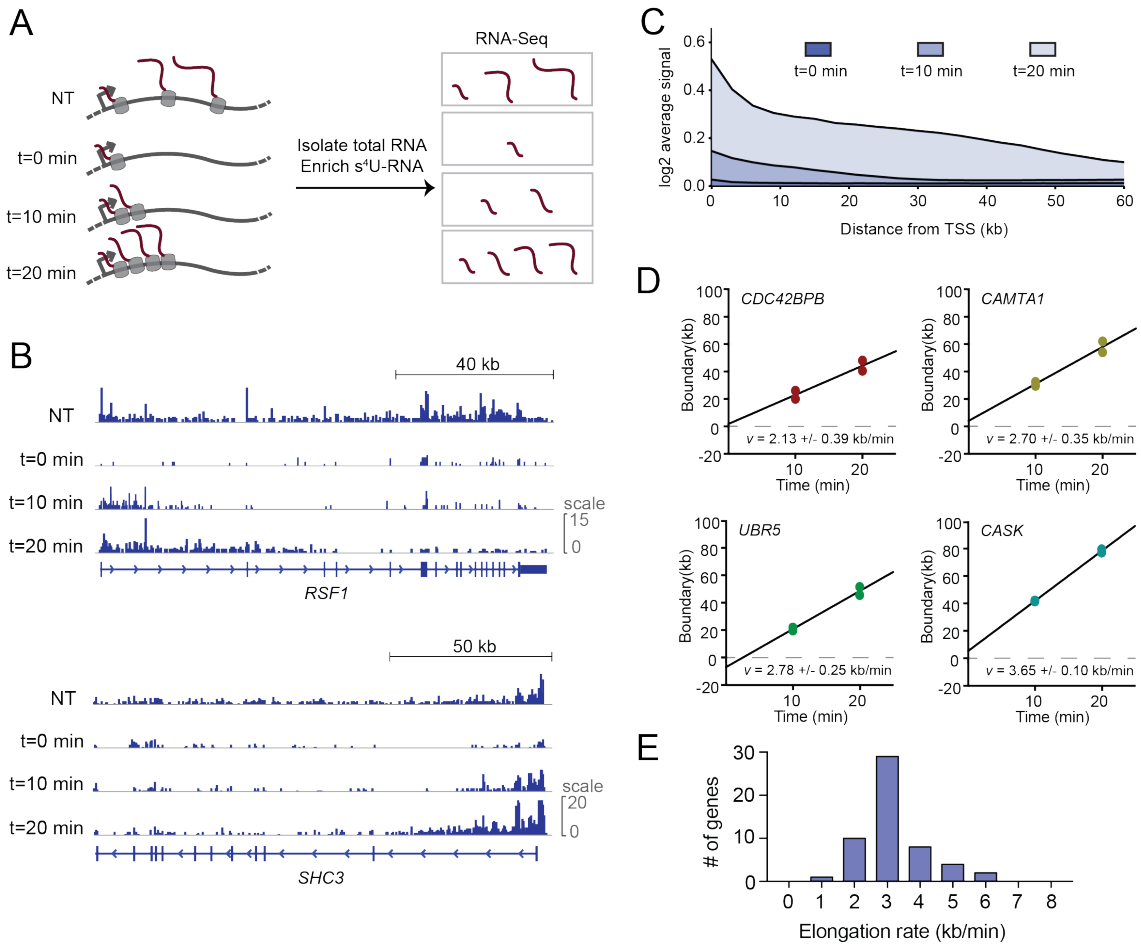

Figure 4.6: **MTS resin reveals RNAPII elongation rates in mouse cortical neurons.** (A) Scheme of 4sUDRB-Seq using MTS resin. NT: no DRB treatment (30 min s^4U-Seq); t=0: no DRB release. S^4U is added to cells for 30 min, but DRB is not removed from cells; t=10, t=20: s^4U-RNA enrichment following 10 or 20 min after DRB release. Total RNA is purified from cells and s^4U-RNA is enriched with MTS resin, followed by high-throughput sequencing. (B) 4sUDRB-Seq enriched RNA from two representative genes, *RSF1* and *SHC3*. Arrows mark the direction of transcription. (C) Average distribution of reads in all genes longer than 50 kb from s^4U-RNA enriched after 0, 10 or 20 min of DRB removal from two biological replicates. (D) Linear fit of transcriptional boundaries. The slope (V) represents the elongation rate in kb/min, with confidence intervals as indicated. (E) Distribution of RNAPII elongation rates calculated from 55 genes in cortical neurons.

Figure S4

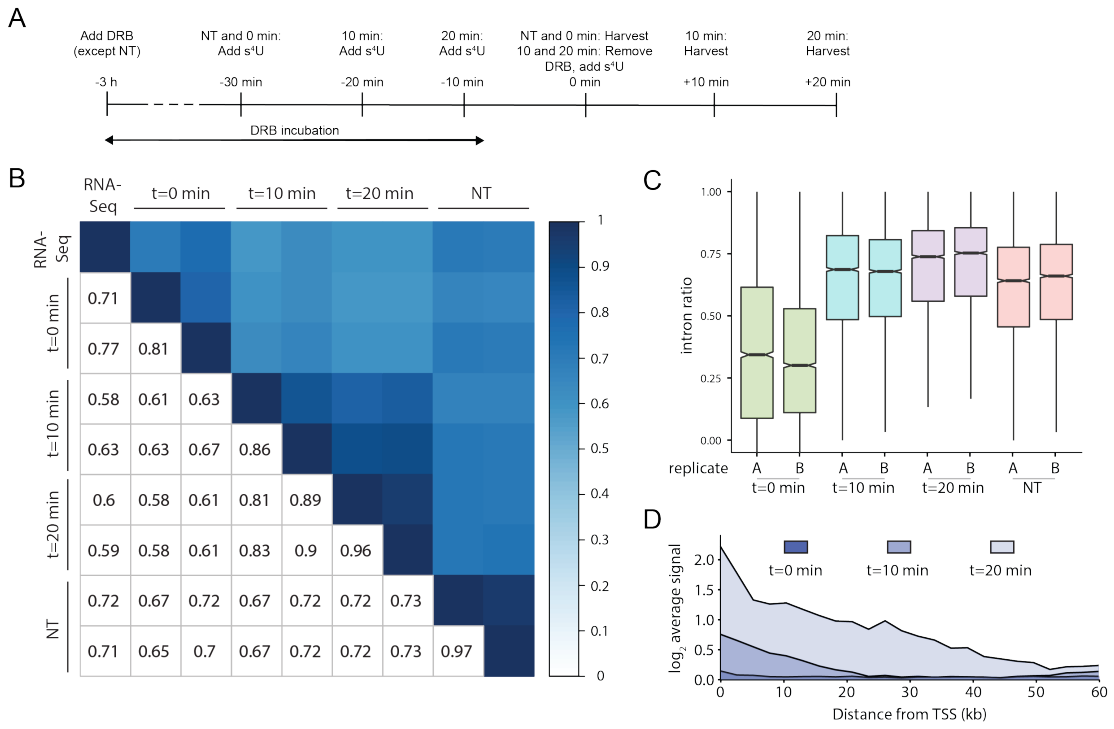

Figure 4.7: **4sUDRB-seq replicates correlate and enrich for nascent RNAs.** (A) Experimental design for 4sUDRB-seq. (B) Heatmap of pairwise correlation coefficients between biological replicates of 4sUDRB-Seq and RNA-seq. Pearson's r correlations are shown. (C) Box plot of the ratio of intronic/total reads per gene in 4sUDRB-Seq samples. (D) Average distribution of reads over the genes in Table 4.2 from s^4U-RNA enriched after 0, 10 or 20 min of DRB removal from two biological replicates.

the transcriptional dynamics in primary neurons in collaboration with Daniele Canzio and Tom Maniatis. Canzio treated ∼2x10^6 neurons with DRB to synchronize RNAPII at the promoter-proximal pause site. Next, s^4U was added to the media (before DRB release to allow s^4UTP to build up in the cellular pool) and the DRB was removed to allow the polymerase to elongate and incorporate s^4U into the newly synthesized RNA. Total RNA from cells harvested after 0, 10, or 20 min of DRB release was purified, and I enriched s^4U-RNA with MTS resin followed by high-throughput sequencing (Figure 4.6A, 4.7A).

Read counts mapping to transcripts were consistent between biological replicates (Pearson's r = 0.81 0.97). As I would expect if 4sUDRB-seq is capturing the transcriptional wave of newly synthesized RNA, t=10 min samples correlated better to t=20 min (Pearson's r = 0.81 - 0.90) than to RNA-seq (Pearson's r = 0.58 - 0.63), t=0 min (Pearson's r = 0.58 - 0.67) or no treatment controls (Pearson's r = 0.67-0.72, Figure 4.7B). In addition, the t=10 min, t=20 min, and no treatment samples contained a higher proportion of reads mapping to introns compared to the t=0 min samples, consistent with the expectation that 4sUDRB-seq enriches for nascent RNA (Figure 4.7C).

A transcriptional wave is observed in the t=10 min and t=20 min samples in which the nascent RNA reads in the t=20 min sample extended much beyond the t=10 min sample relative to the TSS (Figure 4.6B, C). These profiles are similar to those demonstrated in HeLa cells, despite using 20-fold less RNA for enrichment (Figure 4.6B, C). Using criteria established by Fuchs *et al.*, I limited the analysis to the most abundant isoform of transcripts >50 kb in length where transcriptional boundary algorithms (see Methods and Data Analysis) gave convergent values for both biological replicates in the t=10 and t=20 min samples. In addition, I filtered these elongation rates for an x-intercept between -2.5 min and 10 min, which is proportional to the longer times used in this experiment (10 and 20 min, compared to 4 and 8 min in [30]). Based on these conservative criteria, I was able to calculate the elongation rate for 55 genes in cortical neurons (Figure 4.6D, E, Table 4.2). The s^4U-RNAs that mapped to these regions show a similar transcriptional wave compared to the profile for all genes >50 kb (compare Figure 4.7D and Figure 4.6C). From these data, the mean transcription elongation rate in mouse cortical neurons for these transcripts was found to be 3.17 kb/min, which is similar to the mean transcription elongation rate that

Fuchs *et al.* calculated in HeLa cells (3.6 kb/min). In addition, the RNAPII elongation rate varied more than 3-fold between genes (1.5 to >6 kb/min), which is consistent with previous findings in cell culture [15, 30, 95, 133].

Gene	Chr	Start	End	Strand	Elongation rate (kb/min)
Ank2	3	126921611	127411652	-	3.575
App	16	84954439	85173707	-	2.725
Arid2	15	96287517	96404992	+	2.275
Atp2b2	5	122453512	122502225	-	1.775
Bach2	4	32238803	32586108	+	1.375
Bcl9	3	97203661	97297917	-	2.700
Bptf	11	107033080	107132127	-	6.250
Brd4	17	32196273	32284722	-	2.875
Camta1	4	150917321	151861876	-	2.700
Capzb	4	139192898	139291818	+	3.275
Cask	X	13517079	13851367	-	3.650
Ccdc88c	12	100912699	101028983	-	2.350
Cdc42bpb	12	111292971	111377718	-	2.125
Daam1	12	71831077	71992367	+	3.550
Dennd4c	4	86748554	86850603	+	3.425
Dip2c	13	9276527	9668928	+	2.900
Dlg2	7	91090705	92449247	+	1.700
Elmo1	13	20090595	20606528	+	3.325
Eri3	4	117550364	117674297	+	2.975
Exoc4	6	33249084	33973979	+	2.825
Ext1	15	53064037	53346159	-	3.000
Fnbp1l	3	122538718	122619715	-	3.275
Gm45062	6	8259449	8597480	+	2.850
Hdac9	12	34047579	34917095	-	4.075
Insr	8	3122060	3279617	-	4.875
Jakmip2	18	43531407	43687773	-	3.175
Kcnq1ot1	7	143212154	143296549	-	2.950
Kdm4b	17	56326061	56402870	+	3.225
Lrp8	4	107802260	107876840	+	3.725
Mark3	12	111574509	111656227	+	3.550
Myt1l	12	29528383	29923213	+	2.575
Nbea	3	55625194	56183701	-	3.675
Ncor1	11	62316425	62458541	-	2.625
Nedd4l	18	64887755	65217826	+	2.350
Nrcam	12	44328884	44601964	+	2.050
Ntng1	3	109780039	110144011	-	1.950
Phtf1	3	103968109	104024598	+	2.575
Pias2	18	77065207	77155708	+	3.225
Ppp2r2b	18	42637431	43059471	-	5.475
Prkar2b	12	31958475	32061296	-	5.125
Rsf1	7	97579888	97692778	+	5.075

Gene	Chr	Start	End	Strand	Elongation rate (kb/min)
Shc3	13	51431040	51567084	-	2.325
Smarca2	19	26605049	26778322	+	3.000
Snx27	3	94497543	94582716	-	3.250
Soga3	10	29143838	29199630	+	2.500
Ssbp3	4	106910700	107049694	+	2.750
Tmem178b	6	106910700	107049694	-	2.650
Trim37	11	87127076	87220683	+	4.400
Ttc3	16	94370617	94522117	+	2.250
Ubr5	15	37967327	38078853	-	2.775
Usp24	4	106316212	106441322	+	2.950
Usp34	11	23306894	23490560	+	6.325
Wnk1	6	119923968	120038672	-	3.175
Zfp704	3	9427019	9610085	-	2.950

Table 4.2: Genes with calculated RNAPII elongation rates in kb/min and 50% confidence intervals (+/- kb/min).

4.5 Discussion

Here I present solid-phase chemistry with MTS conjugated to beads for the capture of s^4U metabolically labeled RNA. This one-step enrichment reduces handling loss that is prevalent in other s^4U enrichment protocols, thereby enabling 20-50-fold less input RNA. This resin can capture s^4U-RNA from small numbers of cells and identify fast- and slow-turnover transcripts in K562 cells using s^4U Chase-seq, and the results correlate well with published RNA half-lives [29, 119]. In addition, MTS resin captures transient RNA species including introns, eRNAs, and pri-miRNAs in K562 cells in the context of a TT-seq experiment, but with much less input material. These data correlate well with published data at higher scale and the bioinformatic pipeline mirSTP, as well as other pipelines originally developed for GRO-seq and PRO-seq data, can be applied to MTS-TT-seq to identify intergenic miRNA TSSs.

Using MTS resin, the first analysis of RNAPII elongation rates in cortical neurons was carried out in collaboration with Daniele Canzio and Tom Maniatis. Despite the differences in metabolic state of primary neurons compared to rapidly dividing HeLa cells, their mean transcription rates are remarkably similar (3.1 in neurons versus 3.6 kb/min, Figure 4.6D). I found that these rates vary more than 3-fold across the set of genes analyzed, which is

consistent with rates in cell culture [15, 30, 95, 133]. The differences in these rates is likely to be important for co-transcriptional processes like alternative splicing and RNA degradation [52]. Given the extensive differential processing of RNA in neurons in particular [32], the ability to measure RNAPII elongation rates in mouse cortical neurons lays the foundation for understanding the relationship between the control of elongation and the generation of different neuronal RNA isoforms.

Finally, the MTS resin is an ideal substrate for studying RNA dynamics in samples where input RNA is limiting. I suggest MTS resin is a particularly useful option for smaller-scale purification of s^4U-RNA (e.g., primary tissues or microdissections). This methodology expands the existing metabolic labeling toolkit to holistically study transient and stable RNAs in the same experimental setup, thereby offering a glimpse into the complex transcriptional network of a variety of cell types such as primary neuronal cultures.

Chapter 5

Metabolic labeling with s^4U reveals transcriptional kinetics of glucocorticoid stimulation

5.1 Author contributions

I performed all experiments and bioinformatic analyses in A549 cells with input from Matthew Simon. Xiao Xiao and Juliana Shaw provided mouse hippocampal slices and input on mouse hippocampus experimental design with Tony Koleske. I performed metabolic labeling, total RNA purification, s^4U-RNA enrichment and analysis in mouse hippocampal slices.

5.2 Summary

The classical transcription factor GR is the target of many biochemical and pharmacological studies due to its clinical relevance in suppressing inflammation and mediating stress responses. While the kinetics of GR binding to DNA have been well-characterized, a powerful complement to these studies would be to measure acute local changes in transcription to study GR function. To capture the transcriptional kinetics of GR activation, 4-thiouridine metabolic labeling was used to capture transient RNAs (TT-seq) following a timecourse

of treatment with dex, a GR-specific corticosteroid, in A549 epithelial cells. I hypothesized that short s^4U labeling allows the capture of active transcription, which would allow the detection of unstable RNAs, including eRNAs, with high temporal resolution. Indeed, several hundred dex-regulated transcripts were identified and clustered based on their temporal profiles. Additionally, using standard RNA-seq pipelines I was able to identify four dex-responsive intergenic eRNAs. These eRNAs were enriched for direct GR binding at enhancers, rapid transcriptional response, and stable transcription after 27 h of dex treatment. Finally, I discuss future analyses and experiments to potentially increase the number of dex-responsive transient RNAs identified in A549 cells and extend these experiments to primary tissues including the hippocampus, where GR activity in response to the native GC corticosterone is critical for learning and memory. These data reveal heterogeneity in the kinetics of transcription activation and repression in response to dex, including a subset of immediate early genes that tend to bind GR closer to their promoters. In addition, this experimental strategy offers promise for future studies in a variety of primary tissues to understand the diverse and complex functions of GR.

5.3 Introduction

Glucocorticoids (GCs) are steroid hormones that bind GR and cause a wide variety of physiological effects including regulation of glucose synthesis and suppression of the immune response [93]. GR is typically sequestered in the cytoplasm by the *HSP90* chaperone complex until it binds GC, whereby conformational changes in the protein release it from its chaperones and allow GR to translocate to the nucleus [102]. Once in the nucleus, GR binds DNA at >10,000 sites, but this widespread binding alters the expression of only a few hundred genes. This is because GR recruits additional coregulators that either directly or indirectly influence RNAPII recruitment or transcriptional activation [16]. In addition, not all GR binding events lead to a change in local transcription, generating a complex picture of how GR regulates gene expression that is difficult to interpret based on DNA binding alone. Ultimately, the combination of factors including GR isoform, DNA sequence, chromatin state, and cooperativity with coregulators determines which genes are regulated

and whether they are activated or repressed [81]. Therefore, while the kinetics of GR binding to DNA are well-characterized [35], a powerful complement to these studies would be to measure acute local changes in transcription to study GR function. Microarray and RNA-seq studies [141, 125, 109] have enabled the identification of up- and down-regulated genes and, in combination with GR chromatin immunoprecipitation (ChIP), have identified promoter-proximal and -distal binding events that correspond to GR activation and repression. Notably, expression of *PER1*, a TF involved in regulating circadian rhythms, was found to be robustly induced by dex, along with many of genes involved in cell differentiation, development, and morphogenesis, suggesting a role for GCs in early development [109, 107]. While these experiments were useful to identify the global transcriptional outcome of dex signalling, they do not capture the dynamics of the immediate transcriptional response due to contamination from pre-existing RNA. Moreover, RNA seek experiments do a poor job of capturing transient RNAs, such as enhancer RNAs (eRNAs) and antisense transcripts, that may be critical to the regulation of RNA expression and have been shown to be dependent on TF binding in other systems [42]. Therefore, I sought to measure the transcriptional kinetics of GC stimulation in A549 epithelial cells.

While many studies have focused on the mechanism of GR regulation at promoters [37], more recent studies have focused on the importance of enhancer-promoter looping in the transcriptional response to inducible TFs [98]. Chromatin capture methods detect enhancer-promoter interactions [46], but identifying active enhancers that are dex-responsive is critical to understanding the relationship between GR binding and transcriptional regulation. Traditionally, enhancers have been identified by the presence of H3K4me1 and H3K27ac chromatin marks, which denote enhancer regions and active enhancers, respectively [47, 106]. More recently, these enhancer regions have been found to be actively transcribed in a stimulus-dependent manner; the transcripts are referred to as enhancer RNAs (eRNAs) [60]. The relationship between histone marks and eRNAs is an interesting and under-explored area of study, so I sought to identify whether eRNAs are transcribed at enhancers following dex stimulation.

I previously demonstrated that short metabolic labeling with 4-thiouridine (s^4U) can capture newly transcribed RNAs, including transient RNAs such as pre-RNAs, eRNAs,

and pri-miRNAs (see Chapter 4), and that the activated disulfide methane thiosulfonate (MTS) conjugated to magnetic beads captures s^4U-RNA with high efficiency even from small populations of cells. I hypothesized that MTS chemistry could be used to capture newly made s^4U-RNAs following GR stimulation with the corticosteroid analog dex to separate immediate, direct transcriptional changes in the minutes after stimulation from indirect, later changes. I discovered 533 transcripts whose transcription was significantly altered during dex treatment in A549 epithelial cells, and identified distinct classes of dex-responsive genes based on their induction kinetics. In addition, I identified four eRNAs whose expression was regulated in a dex-dependent manner. This result lays the foundation for longer-term exploration to identify more dex-regulated eRNAs, as well as the extension of these experiments to primary tissues such as the mouse hippocampus.

5.4 Results

5.4.1 TT-seq captures the transcriptional response to corticosteroid stimulation in A549 cells

To investigate the transcriptional response to corticosteroids with temporal resolution, I treated A549 epithelial cells with 100 nM dex for 15 min to 27h. To capture active transcription at each time point of dex treatment, I also added s^4 during the last 15 min of dex treatment (Figure 5.1A). Metabolically labeled s^4-RNA was enriched from two biological replicates of the dex-treated A549 cells using MTS chemistry [23, 22], sequenced using Illumina high-throughput sequencing, and reads were mapped to the human genome. Reads were counted over annotated genes and counts were highly correlated between biological replicates (average Pearson's $r = 0.97$, Figure 5.1B). Counts were also enriched for reads in introns, (Figure 5.1C), which is consistent with the observation that TT-seq captures nascent RNA [120].

Previous studies have measured differential expression of dex-resposive genes using RNA-seq [109], so I investigated whether my TT-seq data were consistent with these findings. I saw a robust response for previously reported dex-resposive transcripts in my dataset (Figure 5.1D, E). *PER1* is strongly induced upon dex treatment in A549 cells [110], while *IER2*,

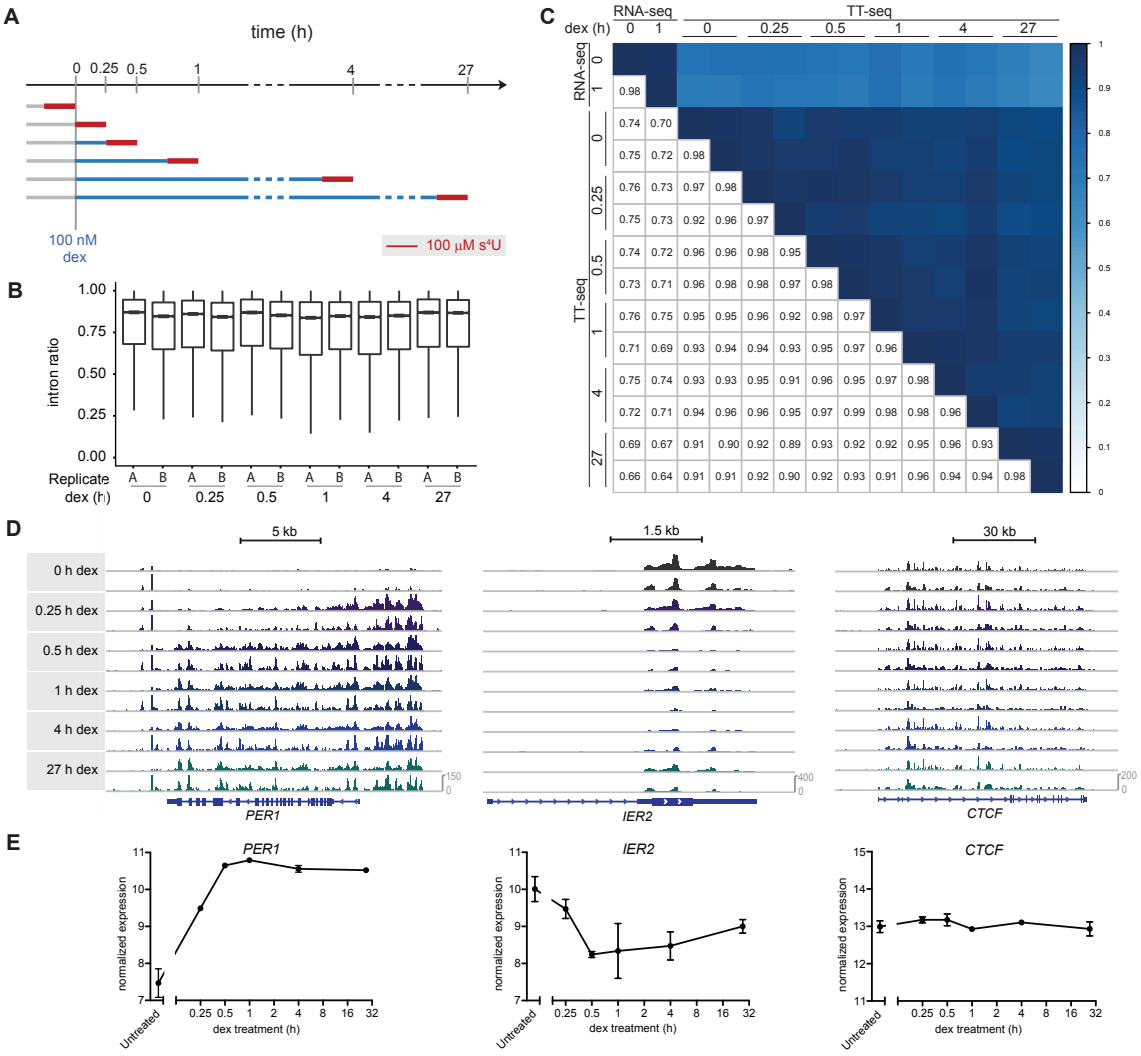

Figure 5.1: **TT-seq provides a detailed view of the dex-regulated transcriptional response in A549 cells.** (A) Scheme of timecourse TT-seq following dex treatment. A549 cells are treated with 100 nM dex for 15 min to 27h, and 1 mM s^4U is added during the last 15 min of dex treatment. Cells are harvested and s^4U-RNA is enriched using MTS biotin and streptavidin. (B) Heatmap of pairwise correlation coefficients between biological replicates of TT-Seq and RNA-seq. Pearson's r correlations are shown. (B) Box plot of the ratio of intronic/total reads per gene in TT-Seq samples. (D) Example genome browser view of s^4U-RNA enriched after dex treatment. *PER1* and *IER2* were previously identified as activated and repressed, respectively, whereas *CTCF* is unchanged upon dex treament. (E) Normalized expression of transcripts from (D). Mean and SEM of two biological replicates is displayed.

a TF involved in cellular differentiation, is downregulated upon dex treatment, consistent with previous findings by Reddy *et al.* [109]. I observed increased *PER1* expression after 30 min of dex treatment which was sustained after 27h of dex treatment. Unexpectedly, I observed a rapid decrease in *IER2* transcription with maximal repression after 30 min of dex treatment, which contrasts previous findings that transcriptional repression by GR is slower than activation [109, 108]. Following maximal repression by 30 min, *IER2* transcription returned to untreated levels after 27h of dex (Figure 5.1D, E), suggesting that dex response may be transient at some sites. I speculate that these rapid induction kinetics revealed by TT-seq but not RNA-seq are due to the ability of TT-seq to capture active transcription, rather than the steady-state profile of RNAs that contains new transcripts along with pre-existing RNA.

Next, I sought to capture the transcriptional profile of induced and repressed transcripts in A549 cells. I used pairwise differential expression analysis (edgeR), comparing normalized s^4-RNA expression between dex-treated and untreated replicates at each time of dex treatment [127]. I identified 533 dex-responsive transcripts, including 53/77 (69%) of transcripts with a >2-fold change in expression after 1h dex by RNA-seq [109]. I speculate that the higher number of dex-responsive genes in my TT-seq dataset is due to the increased sensitivity for detecting immediate transcription compared to RNA-seq, as shown by comparing up- and down-regulated transcripts following 15 min of dex treatment using RNA-seq and TT-seq. Only one dex-responsive gene was identified in these RNA-seq data, compared to the 100 dex-responsive genes identified by TT-seq on the same timescale of dex treatment (Figure 5.2A).

To distinguish different classes of temporally regulated transcripts, I clustered transcripts based on the time of peak transcription (Figure 5.2B, C), a strategy that was used previously to identify acute changes upon estrogen receptor alpha (ERα) stimulation [42]. Based on this analysis, I would expect transcripts with peak transcription after 15 or 30 min of dex (cluster 2) to be directly regulated by GR binding, whereas transcripts whose transcription peaks by ¿ 1h (clusters 4-6) to be either secondary transcriptional effects of dex stimulation or direct binding events that require additional cofactors or chromatin rearrangements for transcriptional activation. I searched for GR binding sites <10kb of

dex-responsive genes using GR ChIP data from dex-stimulated A549 cells [11]. Consistent with this hypothesis, I observe GR binding sites in clusters 2 and 3 to be closer to promoters on average compared to clusters 1, 4, 5, and 6 (Figure 5.2D).

5.4.2 Influence of dex on enhancer RNA transcriptional response

Establishing the influence of GR at the promoters of annotated genes is relatively easier than doing so for more distant enhancers, yet enhancers can modulate rapid chromatin unfolding at GR binding sites that are particularly interesting for a sustained transcriptional response that is independent of GR binding, but have only recently been explored [54]. In addition, recent evidence suggests that transcription factor competition at enhancers facilitates *trans*-repression of downstream genes [148]. However, eRNAs have not been studied in the context of GR stimulation, so I first sought to identify any transcriptionally active enhancers following dex stimulation.

I first used a conservative approach to identify only intergenic enhancers, as transcription from enhancers located within introns would be complicated by unspliced RNAs in TT-seq data. Therefore, I quantified s^4U-RNA signal over intergenic enhancer regions using chromatin feature annotations from the Encode Project in A549 cells [11]. I identified 6,922 intergenic eRNAs that overlapped with predicted enhancer sites and were expressed in at least 2/12 biological samples. I then used the same pairwise differential expression analysis (edgeR) as for identifying dex-responsive genes. Surprisingly, only four eRNAs were identified as significantly dex-responsive by pairwise differential expression (Figure 5.3A, B). The four significant transcripts displayed interesting features including sustained rather than transient transcription throughout the timescale of dex-treatment, suggesting that GR binds enhancers with higher stability than promoters of genes in cluster 2 (Figure 5.2B, C), which display a transient transcriptional response to dex stimulation.

All four dex-responsive enhancers were located within 200kb of a dex-responsive gene (Table 5.1), a strategy that was previously used to predict which enhancers potentially looped to promoters upon ERα stimulation in MCF-7 cells [73]. Of these genes, *KLF6* and *ARRB1* are involved in embryonic development, whereas *ERRFI1* is an antiproliferative factor, both functions that have been previously associated with GR stimulation [109, 10].

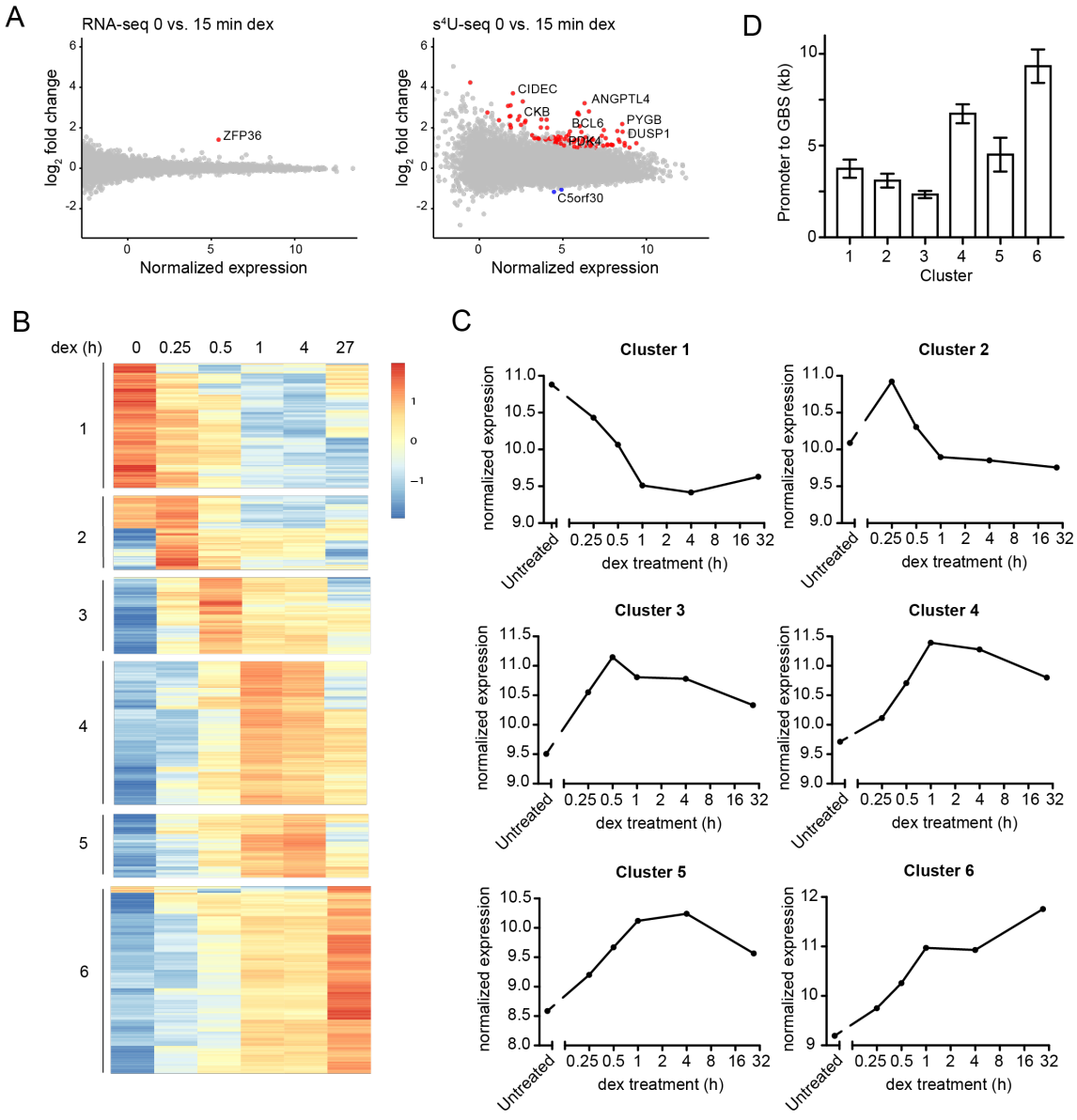

Figure 5.2: **TT-seq identifies dex-regulated transcript clusters in A549 cells.** (A) Scatterplot of fold change versus normalized expression based on comparative analysis of RNAs that are significantly enriched or depleted in samples treated with 15 min dex versus control identified by RNA-seq (left) and TT-seq (right). Upregulated genes (Fold change >2; P <2x10^{-5}) are colored red; downregulated (Fold change <0.5; P <2x10^{-5}) are shown in blue. (B) Heatmap of normalized relative expression for transcripts with significant changes after dex treatment. Clusters are based on the time of peak transcription.(C) Mean TT-seq signal for all transcripts in clusters from (C). (D) Distance in kb from GR binding site to promoter of dex-responsive genes. Mean distance in each cluster and SD is displayed.

All four reach maximal transcription after 15-30 min of dex induction and remain transcriptionally active, like their nearby enhancers, in contrast to many other genes in class 2 (Figure 5.3C).

Enhancer locus	**Gene**	**Gene locus**	**Enhancer distance to gene promoter** (kb)
chr1 8138399-8139200	*ERRFI1*	chr1 8071779-8086393	52.006
chr10 3817599-3817800	*KLF6*	chr10 3818188-3827473	9.673
chr10 3851999-3852200	*KLF6*	chr10 3818188-3827473	24.526
chr11 75084199-75084400	*ARRB1*	chr11 74971166-75062875	21.324

Table 5.1: Loci of dex-responsive eRNAs and dex-responsive genes within 200kb of each enhancer.

5.5 Discussion

Here I present the transcriptional kinetics of dex-responsive genes in A549 cells with temporal resolution. Previously, transcriptional response has been measured by RNA-seq [109, 140], but rapid transcriptional changes are difficult to observe over pre-existing RNA. Therefore, I used short 4-thiouridine metabolic labeling to capture the transient transcriptome (TT-seq) to study active transcription at several times following dex stimulation. I confirmed many previously identified dex-responsive genes and identified hundreds of novel targets and classified these RNAs based on their peak dex expression. Many transcripts peak in expression after 15 min of dex treatment before transcription levels return to untreated levels, suggesting that GR binding is transient at these sites. Other transcripts peak in expression by 4-27h, suggesting that these genes may be secondary transcriptional effects, rather than direct sites of GR binding. Consistent with this hypothesis, I observe GR binding sites in closer proximity to the promoters of early dex-induced genes (clusters 2 and 3) compared to repressed and late induced genes (clusters 1, 4, 5, and 6). Alternatively, gene regulation in clusters 1 and 4-6 may require chromatin remodeling or the recruitment of cofactors upon GR binding, highlighting the importance of integrating GR binding data with transcriptional responses. These data do not help to confidently separate direct from indirect targets, but additional experiments such as TT-seq following cycloheximide inhibition would identify transcripts that directly require GR binding, rather than translation of

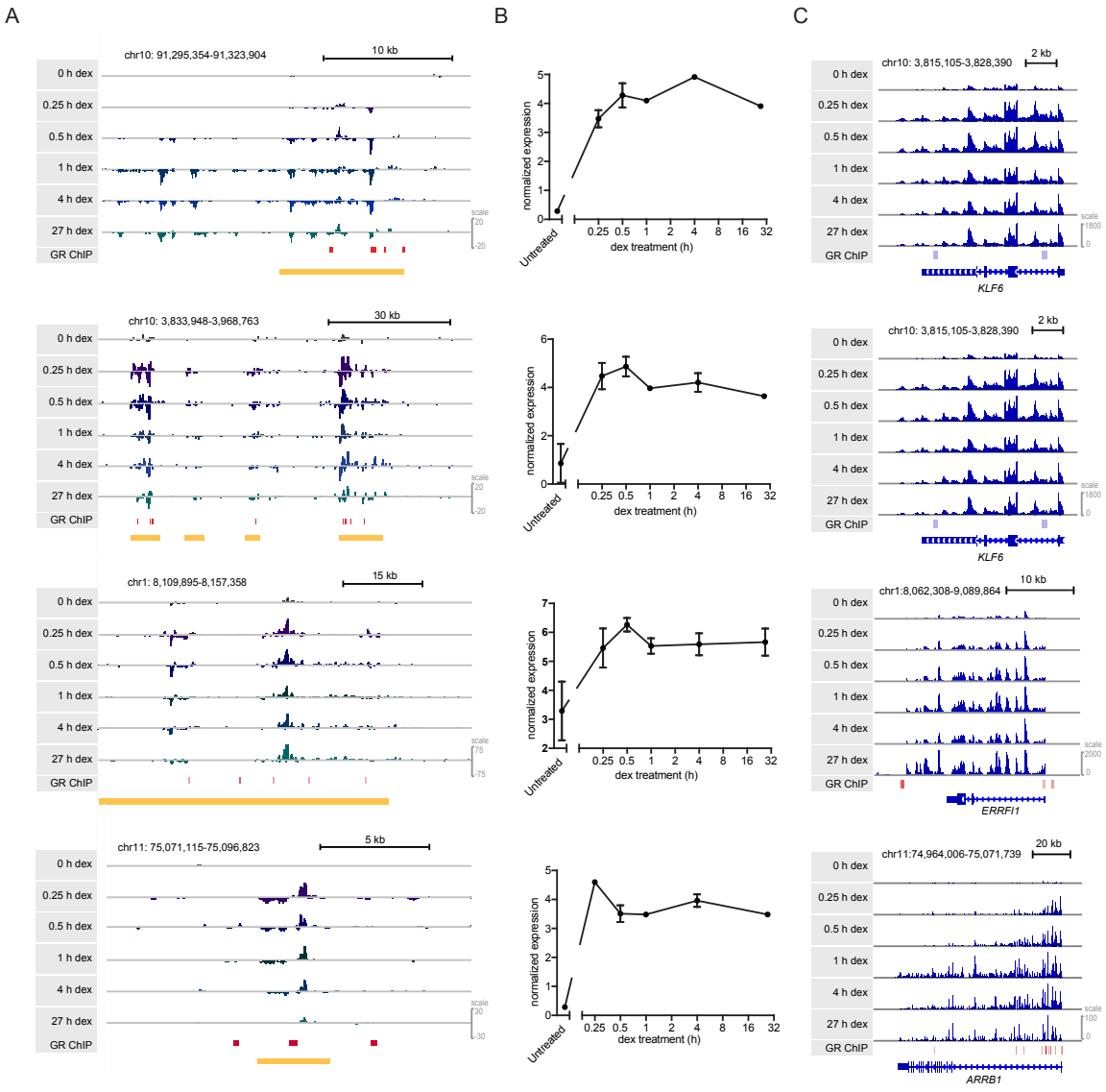

Figure 5.3: **TT-seq identifies dex-regulated enhancers in A549 cells.** (A) Genome browser views of the four dex-responsive eRNAs identified by TT-seq. Red bars indicate GR binding sites, and orange regions indicate annotated enhancers in A549 cells. (B) Normalized expression of transcripts from (A). Mean and SEM of two biological replicates are displayed. (C) Genome browser views of dex-responsive genes within 200kb of the eRNAs in (A).

a dex-responsive gene, as has been shown for other TFs [75].

Active enhancers have been shown to play a vital role in ERα-dependent transcriptional activation [73], and eRNAs are shown to be a good proxy for enhancer activity [66], although eRNAs have not been identified upon GR stimulation. Using a conservative bioinformatic approach to identify only intergenic eRNAs, I identified four significant dex-responsive eRNAs that are likely important for GR activity. These eRNAs contain multiple GR binding sites within their annotated enhancer regions and all display a similar kinetic response of rapid and sustained transcriptional induction upon dex-treatment, suggesting that GR binds rapidly and stably to these enhancers. This trend contrasts with that of many dex-responsive genes, where rapid transcriptional induction eventually returns to untreated levels after several hours of dex treatment. In addition, all four dex-responsive enhancers were located within 50kb of a dex-responsive gene that is rapidly and stably upregulated throughout dex treatment. Interestingly, all four genes with a nearby dex-responsive eRNA are also bound by GR at their promoters, which brings up questions about the mechanism for transcriptional induction: Is GR binding at enhancers or promoters, or cooperativity between GR at the two sites, responsible for rapid and sustained transcriptional activation? Collectively, these data lend new insight into the temporal transcriptional response to GC stimulation and raise questions about the mechanism for rapid and transient transcriptional activation of dex-responsive genes in contrast to the rapid yet sustained activation of dex-responsive enhancers and their nearby genes.

5.5.1 Future Directions

While my conservative bioinformatic approach allowed the identification of the first dex-responsive eRNAs, I might have expected a greater number of eRNAs to be dex-responsive due to the extensive binding of GR at enhancer regions [51] as well as studies in response to other stimuli that reveal more extensive eRNA expression [42, 60, 18]. I speculate that alternative bioinformatic pipelines, including those developed for the analysis of TT-seq data which quantify transcripts into discrete transcriptional units first, may be more effective in identifying eRNAs compared to chromatin annotation alone [120]. In addition, the short half-life ($t_{1/2}$ $\sim$5min) of eRNAs may make these transcripts difficult to detect after 15

min of s^4U metabolic labeling due to lower s^4U-RNA signal compared to more stable dex-responsive transcripts. Therefore, shorter metabolic labeling (5 min s^4U, as in Chapter 4.4.3) combined with greater sequencing depth is likely to aid in more robust identification of dex-responsive eRNAs. By identifying a greater number of dex-responsive eRNAs, I expect to find examples of eRNA repression as well as activation in A549 cells.

While the increased signal gained from treating cells with s^4U for 15 min rather than 5 min aided the identification of several hundred dex-responsive genes, the relative instability of eRNAs may have made it difficult to robustly quantify dex-responsive enhancers in the same dataset. Therefore, another possible future direction is to repeat the experiment with 5 min s^4U labeling to identify more dex-responsive eRNAs and potentially other transient RNAs such as antisense RNAs, which have recently been shown to modulate transcriptional response to TF binding [121]. In addition, more sophisticated clustering algorithms such as those described by Rabani *et al.* [104] could provide a more detailed understanding of temporal regulation in TT-seq data. To understand the relationship between enhancer and gene regulation, chromatin capture data can be used to identify which enhancers loop to which promoters. Finally, while TT-seq is an ideal method to identify transient RNAs and acute transcriptional responses without the need for nuclear isolation, abundant RNAs may contaminate the enrichment and mask transcriptional changes in stimulus-responsive genes. Therefore, an approach such as TT-TimeLapse-seq (Chapter 3.3.2) can be used to filter out contaminating RNAs and increase the specificity of TT-seq.

Additionally, tissue-specific mechanisms of GR and other inducible TFs are a subject of ongoing research. While my studies were performed in cell culture, s^4U-RNA purification with MTS resin (Chapter 4) should enable the study of GR transcriptional response in many cell types and primary tissues. I are currently studying the transcriptional response to corticosterone (*CORT*), a native GR ligand, in the mouse hippocampus. While changes in protein levels can be observed only after 11 days of *CORT* treatment in mouse hippocampus, I hypothesize that s^4U metabolic labeling can be used to identify *CORT*-dependent changes in transcription after a few hours. In this way, I can hope to understand transcriptional regulation in two very different cell types to understand tissue-specific mechanisms of GR activity.

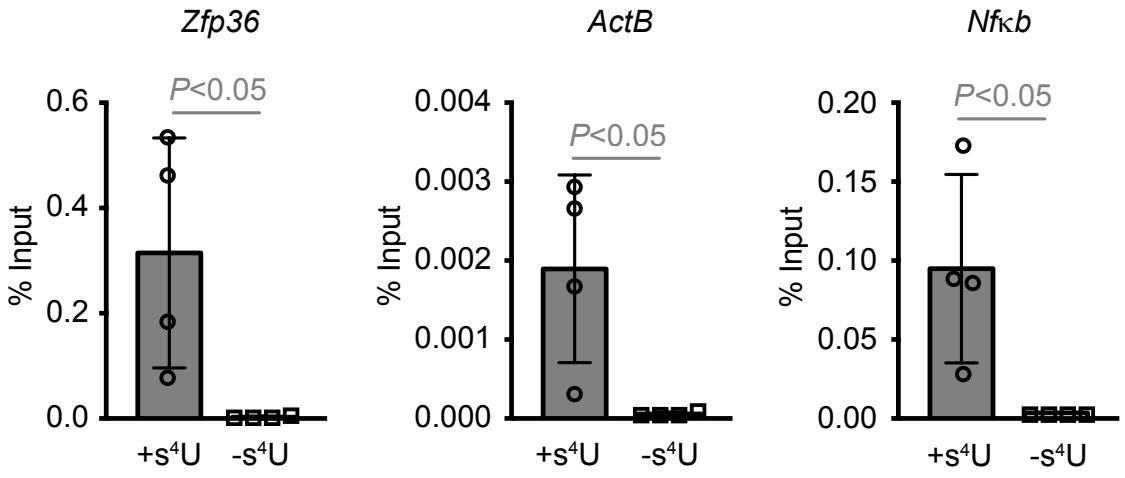

Figure 5.4: **s^{4}U is incorporated into mouse hippocampal slices.** The hippocampus was dissected from four biological replicates and slices were metabolically labeled with 1 mM s^{4}U for 2 h (or no labeling as a s^{4}U control). Total RNA was purified and s^{4}U-RNA was enriched using MTS chemistry. Enriched RNA was quantified by RT-qPCR for the *ZFP36*, *ACTB*, and *NFκB* RNAs. Percent input was quantified relative to a 10% input RNA sample. *P*-values were calculated by the paired T-test.

As a preliminary study, I soaked slices of mouse hippocampus, which were prepared by Xiao Xiao and Juliana Shaw, in media with or without 1 mM s^{4}U for 2h, purified total RNA, and enriched s^{4}U-RNA using MTS chemistry. Although the s^{4}U incorporation is lower in hippocampal slices (0.005-0.5% input) than cell culture (HEK293T cells, 1-10% input) after 2h of labeling, I see significant enrichment of s^{4}U-RNA over a non-s^{4}U control that is consistent across four biological replicates and three qPCR primer sets (Figure 5.4). Future studies may optimize s^{4}U incorporation by decreasing hippocampal slice thickness, thereby increasing surface area of the tissue. I see this as a promising strategy to study RNA dynamics in whole tissues to understand tissue-specific transcriptional regulation.

Chapter 6

Methods and Data Analysis

6.1 Methods

6.1.1 Cell Lines and s^4U Metabolic Labeling

HEK293T cells were cultured in high-glucose DMEM media supplemented with 10% (v/v) fetal bovine serum, and 1% (v/v) 2 mM L-glutamine. K562 cells were cultured in RPM1 media supplemented with 10% (v/v) fetal bovine serum, and 1% (v/v) penicillin/streptomycin. For labeling of long RNAs (Chapter 2), cultured cells at 80% confluence were treated with 700 μM s^4U for 60 min, washed with PBS, trypsinized, and harvested. Cells were resuspended in TRIzol reagent, flash frozen, and stored overnight at 80°C. Cell lysates were chloroform extracted once, and total RNA was purified by the RNeasy mini kit (QIAGEN). For miRNA labeling, cultured cells were grown for 6 days and split 1:8 on day 3. Cells were grown in the presence of 100 mM s^4U for 22 days, 6 days, 3 days, 1 day, 9h, 3h, 1h, 20 min, or in the absence of s^4U. On day 6, all cells were harvested using trypsin and resuspended in TRIzol reagent with exogenous s^4U-containing miRNAs (Dharmacon) and one exogenous non-s^4U miRNA (IDT). Samples were flash frozen and stored overnight at 80°C. Cell lysates were chloroform extracted once and total RNA purified by the miRvana miRNA isolation kit (Life Technologies).

naïve T cells were isolated from *Mettl3*-KO and wild-type mice with naïve T cell purification kits (StemCell). The cells were counted and aliquoted at 4-million per 1 mL FACS

buffer per Eppendorf tube. Each tube except t=0 received 10 μg/mL IL-7 cytokine and was incubated at 37°C. 15 min before being spun down and lysed with Trizol, the cells were labelled with 250 μM s^4U.

For TT-seq, K562 cells were treated with 500 μM s^4U for 5 min (Chapter 4) and A549 cels were treated with 100 μM s^4U for 15 min during the last 15 min of dex treatment for 0 min, 15 min, 30 min, 1 h, 4 h, or 27 h. For s^4U Chase-Seq experiments, K562 cells were treated with 1 mM s^4U for 2h. s^4U-containing media was exchanged with RPM1 media containing 20 mM uridine, and cells were harvested after 0.5 h or 18h of incubation. HEK293T cells were treated with 100 μM s^4U for 2h. Cells in TRIzol were flash frozen and stored overnight at -80°C.

Mouse cortical neurons were isolated at E18 as previously described [55]. About $2x10^6$ cells were plated per 6-well dish. Two days post isolation, cells were treated with DRB for 3 h or with DMSO for the minus DRB control. 4sUDRB experiments were performed as previously described [31] with the following changes: (1) Cells were treated with 1 mM s^4U for 30 min before being collected (2) Cells were collected 0, 10, or 20 min after DRB removal. Cells were harvested with Trizol, flash frozen and stored at -80°C.

6.1.2 Purification of total RNA

Cell lysates were chloroform extracted once and precipitated with one volume of isopropanol (supplemented with 100 μM DTT and 5-10 μg glycoblue) incubated for 10 min at room temperature and centrifuged at 20,000 x g for 20 min. The pellet was washed with an equal volume of 75% ethanol. Purified RNA was dissolved in RNase-free water to a concentration of 200 ng/μL. Contaminating DNA was digested with Turbo DNase (1 U per 10 μg RNA, 37°C, 30 min.) Samples were purified by phenol:chloroform extraction followed by a second isopropanol precipitation and resuspension in 40 μL of RNase-free water. For sheared RNA samples, 2x fragmentation buffer (40 μL, 150 mM Tris pH 8.3 [buffer made at 25°C], 225 mM KCl, 9 mM $MgCl_2$) was added to the RNA and incubated at 94°C for 4 min. Shearing was quenched by immediately placing samples on ice and adding EDTA (20 μL 250 mM EDTA pH 8.0, final concentration 50 mM). Samples were cooled on ice (2 min) and were purified by the following modification of the RNeasy MinElute Cleanup Kit (Qiagen). 350

μL of buffer RLT and 250 μL of 100% EtOH were added to the RNA sample and placed over an RNeasy column. Columns were centrifuged for 15 s at 12,000 x g, 4°C, and flow through was discarded. Columns were washed with 500 μL buffer RPE supplemented with 35 μL of 1% βME (final concentration 10 mM βME) and centrifuged as above. Columns were washed with freshly prepared 80% EtOH and centrifuged for 2 min at 12,000 x g, 4°C, and flow through was discarded. Columns were dried by centrifugation at maximum speed for 5 min at 4°C. Samples were eluted into a fresh microfuge tube with 14 μL RNase-free water and centrifuged at maximum speed for 1 min at 4°C.

6.1.3 Purification of s^{4}U-Labeled RNA with MTS biotin

Chapter 2: Biotinylation and s^{4}U-RNA enrichment with HPDP-biotin were carried out based on protocols adapted from Gregersen et al. [38] and optimized for MTS-biotin. Reactions were carried out in a total volume of 250 μL, containing 70 μg total RNA, 10 mM HEPES (pH 7.5), 1 mM EDTA, and 5 μg MTSEA biotin- XX (Biotium) or 50 μg HPDP-biotin (Pierce) freshly dissolved in DMF (final concentration of DMF = 20%). Reactions were incubated at room temperature for 2 hr (HPDP) or 30 min (MTS) in the dark. Following biotinylation, excess biotin reagents were removed by addition of 1 volume phenol:chloroform (Sigma), followed by vigorous mixing for 15 s, 2 min incubation at RT, and centrifugation in a Phase-Lock-Gel tube (5Prime) at 12,000 x g for 5 min. Supernatant was removed, and RNA was precipitated with a 1:10 volume (20 μL) of 5 M NaCl and an equal volume of isopropanol (200 μL) and centrifuged at 20,000 x g for 20 min. The pellet was washed with an equal volume of 75% ethanol. Purified RNA was dissolved in 50 μL RNase-free water and denatured at 65°C for 10 min, followed by rapid cooling on ice for 5 min. Biotinylated RNA was separated from non-labeled RNA using mMacs Streptavidin Microbeads (Miltenyi). Beads (200 μL) were added to each sample and incubated for 15 min at room temperature. In the meantime, mColumns were placed in the magnetic field of the mMacs separator and equilibrated with nucleic acid wash buffer supplied with the beads (Miltenyi). Reactions were applied to the mColumns, and flow-through was collected as the pre-existing RNA fraction. mColumns were washed twice with high-salt wash buffer (500 μL each, 100 mM Tris-HCl [pH 7.4], 10 mM EDTA, 1 M NaCl, and 0.1% Tween-20).

s^4U-RNA was eluted from mColumns with 100 μL freshly prepared 100 mM DTT followed by a second elution with an additional 100 μL 5 min later. RNA was recovered from the flow-through and eluent samples using the MinElute Spin columns (QIAGEN) according to the instructions of the manufacturer. *S. pombe* total RNA (11 ng, a generous gift from Julien Berro) was added to each sample for downstream normalization.

Chapter 3 and 5: 20 μg sheared RNA was incubated with 2 μg MTS-biotin in biotinylation buffer for 30 min. Excess biotin was removed via chloroform extraction using Phase-Lock Gel Tubes. RNA was precipitated with a 1:10 volume of 3 M NaOAc and an equal volume of isopropanol and centrifuged at 20,000 x g for 20 min. The pellet was washed with an equal volume of 75% ethanol. Purified RNA was dissolved in 50 μL RNase-free water. Biotinylated RNA was separated from non-labelled RNA using glycogen-blocked Dynabeads Streptavidin C1 Beads (Invitrogen). Beads (10 μL) were added to each sample and incubated for 15 min at room temperature, then washed three times with high salt wash buffer. To improve the stringency of the washes, an additional three washes with buffer TE (10 mM Tris pH 7.4, 1 mM EDTA) at 55°C were added to the protocol. s^4U-RNA was eluted from Dynabeads with 25 μL freshly prepared elution buffer (10 mM DTT, 100 mM NaCl, 10 mM Tris pH 7.4, 1 mM EDTA, 10 pg μL *S. pombe* total RNA and incubated for 15 min, followed by a second elution with an additional 25 μL elution buffer. Both elutions were pooled and purified by ethanol precipitation. **Chapter 3 only:** 1% input (200 ng sheared RNA) was also saved from each sample before enrichment, and 500 pg *S. pombe* total RNA was added as a normalization spike-in.

6.1.4 miRNA RATE-seq s^4U RNA Enrichment

Biotinylation and s^4U-RNA enrichment were carried out as described above (purification of s^4U-labeled RNA with MTS biotin) with the following modifications. Excess biotinylation reagent was removed using a nucleotide cleanup kit (QIAGEN). Following enrichment, RNA was concentrated by ethanol precipitation and re- suspended in 14 μL RNase-free water. After enrichment, samples were supplemented with four synthetic miRNA standards (Dharmacon; Table S3).

6.1.5 TT-TimeLapse-seq

K562 cells were grown at 37°C in RPMI containing 10% FBS and 1% P/S. At approximately 50% confluence, the media was supplemented with s^4U (1 mM). The cells were incubated at 37°C for 5 min, at which point total RNA isolation and genomic DNA depletion were performed as described above (Purification of total RNA). 50 μg of total RNA was subjected to MTS chemistry, followed by biotinylation and streptavidin enrichment as described above (purification of s^4U-labeled RNA with MTS biotin). After elution from SAV beads, enriched RNA was purified using one equivalent volume of Agencourt RNAclean XP beads according to manufacturer's instructions instead of purification by ethanol precipitation.

Isolated total RNA was added to a mixture of TFEA (600 mM), EDTA (1 mM) and sodium acetate (pH 5.2, 100 mM) in water. A solution of $NaIO_4$ (10 mM) was then added drop wise and the reaction mixture was incubated for 1h at 45°C. Potassium chloride (300 mM) and sodium acetate (pH 5.2, 300 mM) were added and the reaction mixture was allowed to stand on ice for 10 min prior to centrifugation (>10000 rpm, 30 min, 4°C) to precipitate remaining periodate. The RNA in the supernatant was then ethanol precipitated and washed three times with 75% ethanol prior to resuspension in nuclease-free water.

Chemically treated RNA was purified using 1 equivalent volume of Agencourt RNAclean XP beads according to manufacturer's instructions. Purified material was then incubated in a reducing buffer (10 mM DTT, 100 mM NaCl, 10 mM Tris pH 7.4, 1 mM EDTA) at 37°C for 30 min, followed by a second RNAclean bead purification.

6.1.6 Synthesis of MTS resin

For ten samples, 100 μL NHS magnetic sepharose (GE Healthcare) was added to 1 mL RNase-free water, mixed well, and captured on a magnetic rack, discarding supernatant. One mL ice cold 1 mM HCl was added to beads, incubated 2 min with agitation, and beads were captured on a magnetic rack, discarding supernatant. Beads were washed twice with 1 mL 1x PBS, then 100 μL MTSEA (Biotium, 10 mg/mL in PBS) plus 0.75 μL DIEA was added. Beads were incubated at room temperature for 15 min with rotation. Beads were washed twice with 1 mL 1x PBS and blocked with 1 mL blocking buffer (1x PBS, 5 μL

acetic anhydride) for 15 min with rotation. Beads were washed twice with 1 mL 1x PBS and twice with 1 mL quench buffer (1 M Tris pH 7.4, 200 mM ethanolamine). Beads were quenched with 1 mL quench buffer for 15 min with rotation. Beads were washed twice with 1 mL quench buffer and twice with 1 mL binding buffer (100 mM NaCl, 10 mM HEPES pH 7.4, 1 mM EDTA, 0.05% Tween). Beads can be stored at 4°C for several hours in binding buffer or aliquoted (10 μL/sample) and binding buffer removed to use immediately.

6.1.7 Purification of s^{4}U-Labeled RNA with MTS resin

1-5 μg RNA in 9 μL RNase-free water was mixed with 1.5 μL 10x binding buffer (1 M NaCl, 100 mM HEPES pH 7.4, 10 mM EDTA, 0.5% Tween), and 3 μL DMF, added to MTS resin (10 μL aliquot), and incubated at room temperature in the dark for 2 h with rotation. Beads were captured on a magnetic rack and supernatant removed. Beads were incubated for 5 min with the following wash buffers: once with 100 μL 1x binding buffer, twice with 100 μL high salt wash buffer (1M NaCl, 100 mM Tris pH 7.4, 10 mM EDTA, 0,05% Tween), twice with 100 μL denaturing buffer (8 M guanidinium chloride), and three times with 100 μL buffer TE (10 mM Tris pH 7.4, 1 mM EDTA) at 55°C. RNA was eluted with 10 μL elution buffer (100 mM NaCl, 10 mM Tris pH 7.4, 1 mM EDTA, 10 mM DTT, and 100 pg fragmented total RNA from *S. pombe* and incubated at room temperature in the dark for 15 min with rotation. Eluent was ethanol precipitated with 5-10 μg glycoblue, and RNA concentration was assayed by Bioanalyzer RNA 6000 Pico Kit (Agilent) according to the manufacturer's instructions.

6.1.8 Library preparation and sequencing

Chapter 2: Sequencing libraries were constructed using standard protocols by the Yale Center for Genomic Analysis (YCGA) and run on Illumina HiSeq 2500 instruments. Long RNA-seq was performed using 5 μg of RNA from input RNA, flow-through, or eluted fractions. Strand-specific library preparation was performed using poly-A-selected RNA collected from flow-through and eluted fractions. Samples were multiplexed using Illumina bar codes and sequenced using paired-end 2 x 75-nt cycles. For small RNA-seq, 10% input and RNA collected from eluted fractions were used for small RNA library preparation and

sequenced with single-end 75-nt cycles.

For all other sequencing libraries, 10 ng input, as well as enriched RNA, was prepared using the SMARTer Stranded Total RNA-seq Pico Input Mammalian Kit (Clontech) according to the manufacturer's instructions. Samples were multiplexed with Illumina TRU Seq i7 barcodes and sequencing was performed at the Yale Center for Genomic Analysis on Illumina HiSeq 2500 instruments with single-end 1 x 75 nt (Chapter 2.6) paired-end 2 x 150 nt (Chapter 2.7), or paired-end 2 x 75 nt sequencing runs (Chapter 3).

6.1.9 qPCR Assays

Input or enriched RNA was converted into cDNA with VILO reverse-transcription kit (Life Technologies). qPCR was carried out on the CFX96 real-time system (BioRad) with the iTaq Universal SYBR Green Mix. Results from all primers used (listed in Table C.1) were corrected for amplification efficiency.

6.1.10 Mass Spectrometry of s^4U Disulfide Exchange

Reactions (50 μL) contained s^4U (50 μM), buffer (20 mM HEPES [pH 7.5], 1 mM EDTA), and MTS- or HPDP-biotin (5 μM) dissolved in DMF (final concentration of DMF = 5%). Aliquots were taken at designated time points and analyzed on an Agilent 6650A Q-TOF using a reverse phase column (Thermo Scientific Hypersil GOLD 3 mm, 160 3 2.1 mm) detected by electrospray ionization (positive ion mode). Chromatography conditions were established based on Su et al. [128]. Briefly, analysis was initiated with an isocratic gradient of 100% buffer A at 0.4 mL/ min for 6 min followed by a linear gradient of 0%–50% buffer B over 6 min, 50%–75% buffer B over 2 min, then an isocratic elution at 75% buffer B (buffer A: H_2O in 0.1% [v/v] formic acid; buffer B: acetonitrile in 0.1% [v/v] formic acid).

6.1.11 NMR of s^4U Disulfide Exchange

Reactions (600 μL) were performed in D_2O containing 10 mM HEPES, s^4U (1 mg, 6.4 mM), and five equivalents of MeMTS or PDPH dissolved in DMF-δ_7 (60 μL, 10% total volume). These reactions were incubated in the dark, 2 hr for PDPH and 30 min for MeMTS. Reactions were analyzed on an Agilent DD2 400 MHz NMR with 16 scans.

6.1.12 Enrichment of Singly Thiolated RNA

Two fluorescently labeled RNAs were synthesized for s^4U enrichment: non- s^4U 39-nt RNA (EED046, for sequences see Table C.2) and one s^4U 39-nt RNA (EED047) (Dharmacon). Biotinylation reactions (50 μL total) contained RNA (1 μM), 10 mM HEPES (pH 7.5), 1 mM EDTA, and 25 μM MTS- or HPDP-biotin (dissolved in DMF at 250 μM). Reactions were incubated at room temperature in the dark for 30 min or 2 hr, respectively. Following biotinylation, excess biotinylation reagents were removed with two consecutive chloroform washes, followed by purification with a nucleotide cleanup kit (QIAGEN) according to the manufacturer's instructions. Biotinylated RNA was separated from non-labeled RNA using Dynabeads MyOne Streptavidin C1 beads (Invitrogen). Biotinylated RNA was incubated with 50 μL Dynabeads with rotation for 1 hr at room temperature in the dark. Beads were magnetically fixed and washed twice with Dynabeads high-salt wash buffer. s^4U-RNA was eluted with 100 μL of elution buffer (10 mM Tris [pH 7.4] and 100 mM DTT). Fractions were concentrated by ethanol precipitation, separated on a 12% urea-PAGE gel, and visualized by Typhoon fluorescence imager (GE).

6.1.13 Enrichment of an *in vitro* Transcribed RNA Ladder

An RNA ladder of 1001,000 nt was transcribed *in vitro* using the RNA Century Plus Marker Template and Maxiscript T7 transcription kit (Invitrogen) using Cy5-CTP at a ratio of 1:1 Cy5-CTP:CTP for downstream visualization, with the option of adding s^4UTP (TriLink Biotechnologies) at a ratio of 1:1 s^4UTP:UTP to the reaction. After the reaction, excess nucleotides were removed by an Illustra Microspin G-25 column (GE Healthcare Life Sciences) according to the manufacturer's instructions. RNA ladders were reacted with HPDP-, MTS-, or thiosulfonate-biotin (Biotium), following the protocol described above. Enriched samples were separated on a 5% urea-PAGE gel, stained with GelGreen, and visualized by Typhoon fluorescence imager (GE).

6.1.14 Enrichment of Thiolated tRNA from *E. coli*

E. coli WT and Δ*thiI* cultures were grown to mid-log phase in LB media. Strains were a generous gift from Eugene Mueller [88]. Cells were pelleted at 3,250 x g for 10 min at 4°C. Total RNA was purified by the mirVana miRNA isolation kit (Life Technologies). RNA pull downs were performed as above (Purification of s^4U-labeled RNA) and fractions separated on a 5% urea-PAGE gel, followed by visualization with GelGreen stain.

6.1.15 MTS resin binding capacity

A solution of 100 nmol s^4U nucleoside in 18 μL RNase-free water was mixed with 3 μL 10x binding buffer and 6 μL DMF (total volume 30 μL) was enriched using 1 μL, 2 μL, 5 μL, 10 μL, or 20 μL MTS resin as above. Enriched nucleoside was quantified by UV-Vis based on A334 absorbance, and resin binding capacity was calculated using a standard curve of different concentrations of s^4U.

6.1.16 MTS resin saturation

1 μg of total RNA from K562 cells (prepared as above) was combined with 1 ng total RNA from *S. pombe* and enriched on 1 μL, 2 μL, 5 μL, 10 μL, 20 μL, or 40 μL MTS resin as above. Enriched RNA and 10% input was reverse transcribed using the Superscript VILO cDNA synthesis kit (Invitrogen) and analyzed by qPCR using iTaq Universal SYBR Green Supermix (Bio-Rad) as above. Enrichment of s^4U-RNA was analyzed using the *H. sapiens* qPCR primer for *CDKN1B* total RNA [38]. Background from *S. pombe* RNA was analyzed using the *S. pombe* qPCR primer for 28S rRNA [152].

6.1.17 Comparison between MTS biotin and MTS resin enrichment

1 *mu*g of total RNA from HEK293T cells was combined with 1 ng total RNA from *S. pombe* and enriched with MTS biotin or MTS resin as above. Enriched samples were assayed by RT-qPCR as above and fold enrichment was calculated as

$$\frac{2^{C_t(Enriched_{Hs})-C_t(Input_{Hs})}}{2^{C_t(Enriched_{Sp})-C_t(Input_{Sp})}} \tag{6.1}$$

6.2 Data Analysis

6.2.1 Mapping and Quantification of s^{4}U-Seq Libraries (Chapter 2)

Sequencing reads were aligned using Tophat2 (version 2.0.12; Bowtie2 version 2.2.3), to a joint index of the *H. sapiens* and *S. pombe* genomes (hg19 and PomBase v22) and transcriptomes (GENCODE v19 and Ensembl Fungi v22; [44, 58]). Alignments and analyses were performed on the Yale High Performance Computing clusters. Following this, Rutenberg-Schoenberg used Cufflinks (version 2.2.1) [137] to quantify annotated *H. sapiens* and *S. pombe* transcripts, using only reads that were uniquely mapped (MAPQ >20) and that aligned with up to two mismatches to the reference.

6.2.2 s^{4}U-Seq Normalization

To compare transcript levels between samples, Rutenberg-Schoenberg normalized expression values to *S. pombe* spike-ins as follows:

$$FPKM_{norm} = FPKM_{raw} S_{norm} \tag{6.2}$$

where FPKM_{norm} is the normalized FPKM of a human transcript or gene, FPKM_{raw} is the original FPKM calculated for the sample of interest, and S_{norm} is the slope of the linear regression line of raw *S. pombe* gene FPKMs with the normalizing sample on the y axis and the sample of interest on the x axis. To normalize genomic coverage tracks, Rutenberg-Schoenberg used a similar scheme:

$$Coverage_{norm} = Coverage_{raw} S_{norm} \frac{R_{sample}}{R_{norm}} \tag{6.3}$$

where $Coverage_{norm}$ and $Coverage_{raw}$ are the normalized and raw read coverages at a given genomic position, and R_{sample} and R_{norm} are the numbers of unique reads in the sample of interest and the normalizing sample, respectively. The R_{sample}=R_{norm} adjustment factor reflects that raw reads are compared, instead of FPKMs. Stranded genomic coverage tracks were generated using IGVTools (version 2.3.32) [134]. For all analyses, samples were

normalized relative to the *S. pombe* spike in the HPDP-biotin sample. The 10-fold biochemical dilution of the input samples was also accounted for prior to library preparation by multiplying normalized values for these samples by ten.

6.2.3 Assessment of Length Bias in Eluted s^{4}U-Seq RNA

Because incorporation and biotinylation of s^{4}U are not perfectly efficient, especially when using HPDP-biotin, it is expected that transcripts with more uridines will be purified at rates greater than or equal to those of shorter trancripts. To assess length bias for each reagent, transcript isoforms were binned by numbers of uridines present and compared the fractions of total input RNA that were purified between bins using the Wilcoxon rank-sum test. To avoid noise from misassignment of reads between isoforms of individual genes, only the dominant isoforms of genes (>90% of total expression) in all samples were included in the analysis. Only transcripts greater than 200 nt were included, since shorter transcripts were biochemically depleted in the library preparation, and transcripts with expression levels in the bottom quartile of the input sample were removed.

6.2.4 miRNA RATE-seq Bioinformatic Analysis

To analyze my smRNA RATE-seq data, a hierarchical mapping pipeline was used combining the sRNAbench [114], Bowtie [68], and Bowtie2 tools [67]. Before mapping the reads, sequencing adapters were removed using fastx-clipper (http:// hannonlab.cshl.edu/fastx_toolkit/). Bowtie2 was used to map reads first to synthetic spikes, and then to the UniVec laboratory contaminant database (http://www.ncbi.nlm.nih.gov/tools/vecscreen/univec/) and ribosomal RNAs from the GENCODE v19 annotation [44]. These two categories of sequences are not expected to produce reads in my miRNA libraries, except by contamination or RNA degradation. The remaining unmapped reads were then mapped using sRNAbench, first to the miRBase miRNA 21 annotation [62], and then to the entire human genome (hg19). Input reads under 19 nt or with greater than one mismatch were removed from all analyses of miRNA and spike quantifications.

To perform differential expression analysis between smRNA RATE-seq time points, the

edgeR package was used (version 3.2.4, [111, 112]). Specifically, three early time points (both 20 min replicates and a deeply sequenced 1 hr time point) were compared to three late time points (two 6-day replicates and a 22-day sample). miRNA read counts and dispersions were fit to a negative binomial distribution, and differential expression was evaluated using the negative binomial exact test. To correct for multiple hypothesis testing, the Bonferroni correction was used, and a family-wise error rate of 0.005 was set to select differentially expressed miRNAs between early time points and the steady state.

6.2.5 Mapping and quantification of s^{4}U-seq libraries (Chapter 3)

Sequencing reads were aligned using STAR (version 2.4.2a) [19] to a joint index of the *M. musculus* and *S. pombe* genomes (mm10 and sp2) and transcriptomes (UCSC and Ensembl Fungi v22) [113, 58]. Alignments and analysis were performed on the Yale High Performance Computing clusters. Following alignment, HTSeq-count (version 0.6.1p1) was used to quantify annotated *M. musculus* and *S. pombe* transcripts for total RNA (-t gene) and mRNA (-t exon) [3]. Tracks normalized using the *S. pombe* reads were uploaded to the UCSC genome browser. The scale was normalized using exogenous *S. pombe* RNA added before library preparation.

Transcript abundance, synthesis and degradation rates were estimated using the INSPEcT package in R [17]. Spearman correlation between samples were visualized using the corrplot package. Transcripts with significant time-dependent changes were determined and clustered using the maSigPro package in R and heatmaps were made using the pheatmap package [94].

6.2.6 TT-Timelapse-seq alignment and mutational analysis

Reads were filtered for unique sequences using FastUniq [146], trimmed using cutadapt [78] to remove Illumina adapter sequences filtering for reads greater than 20 nt (–minimum-length=20) and aligned to the mouse GRCm38 or human GRCh38 genome and transcriptome annotations using HISAT2 [59], using default parameters and –mp 4,2. Files were further processed with Picard tools (http://broadinstitute.github.io/picard/) including FixMateInformation, SortSam and BuildBamIndex. The samtools [71] software was used to

retain only reads that aligned uniquely (flag: 83/163, 99/147), with MAPQ 2, and without insertions (because of ambiguity in mutational analysis) for further analysis.

Reads that uniquely map to the human GRCh38 version 26 (Ensembl 88) or mouse GRCm38 (p6) were identified using HTSeq-count using union mode [3]. Reads mapping to only mature isoforms or to anywhere in the gene body were determined separately and compared to identify intron-only reads. To determine the number of uridine residues inferred from each read, and the sites of T-to-C mutations, the aligned bam files were processed in R using Rsamtools (http://bioconductor.org/packages/release/bioc/html/Rsamtools.html) and the sites and numbers of mutations were determined using a custom R function (available upon request). Only mutations at positions with a base quality score of greater than 45, that were at least three nt from the end of the read were counted. Reads were excluded where there were greater than five T-to-C mutations and these mutations did not account for at least one third of the observed mutations (NM tag). Without adequate filtering, SNPs could interfere with TimeLapse analysis. To identify sites of SNPs (or RNA modifications that could be mis-identified as TimeLapse mutations), the following two strategies were used. First, T-to-C SNP sites were identified in control samples using bcftools [77] with default options and excluded these sites from my analysis. Second, locations were compiled where T-to-C mutations were high in non-s^4U treated controls and excluded these sites from analysis. Once the putative SNPs were filtered, the total number of unique mutations in each read pair was counted. To examine the distribution of reads with each minimum number of T-to-C mutations, the bam files were filtered using Picard tools. To make genome-coverage tracks, STAR aligner (inputAlignmentsFromBam mode, outWigType bedGraph) was used and the tracks were normalized using factors derived from RNA-seq analyses using values from DESeq2 (estimateSizeFactors)[77]. Tracks were converted to binary format (toTDF, IGVtools) and visualized in IGV [134].

6.2.7 Mapping and quantification of s^4U-seq libraries (Chapter 4)

Sequencing reads were aligned using STAR (version 2.4.2a) [19] to a joint index of the *H. sapiens* and *S. pombe* genomes (GRCh38 and sp2) and transcriptomes (NCBI and Ensembl Fungi v22) [58, 80]. Alignments and analysis were performed on the Yale High Performance

Computing clusters. Following alignment, HTSeq-count (version 0.6.1p1) [3] was used to quantify annotated *H. sapiens* and *S. pombe* transcripts for total RNA (-t gene) and mRNA (-t exon). Tracks normalized using the *S. pombe* reads were uploaded to the UCSC genome browser. Pearson correlations between samples were visualized using the corrplot package. GO-enrichment analysis was performed using the PANTHER classification system [83] using the statistical overrepresentation test and GO molecular function complete gene list. TT-seq data from Schwalb et al. [120] (GEO series accession GSE75792 and, respectively) were randomly downsampled for the same read depth as MTS-TT-seq data and processed as above. PRO-Seq data were downloaded from GEO (series accession GSM1480327), and ChIP data for RNAPII, H3K4me1, H3K36me3 and H3K4me1 were downloaded from ENCODE [11]. Enhancer RNAs were quantified using BEDtools coverage (50% minimum read overlap) of the Chromatin State Segmentation by HMM from ENCODE/Broad [25] and filtered for strong enhancers. MicroRNA transcription start sites were assigned using the mirSTP pipeline [76] and relaxed annotations are reported. RNAPII elongation rates were calculated using the scripts published by Fuchs et al. [30], with the following modifications: 1) Due to the lower sequencing depth in my samples, the bin size was changed from 100bp to 500bp. 2) Because the time of s^4U metabolic labeling was 10 and 20 min, rather than 4 and 8 min published by Fuchs et al., the x-intercept filtering was expanded to -2.5 min to 10 min, rather than -1 to 4 min. 3) Rates were calculated for the most abundant isoform of each transcript.

6.2.8 Mapping and quantification of s^4U-seq libraries (Chapter 5)

Sequencing reads were aligned using STAR (version 2.4.2a) [19] to the *H. sapiens* genome (GRCh38) and transcriptome (NCBI [58]). Alignments and analysis were performed on the Yale High Performance Computing clusters. Following alignment, HTSeq-count (version 0.6.1p1) [3] was used to quantify annotated transcripts for total RNA (-t gene) and mRNA (-t exon). Tracks normalized using a Variance Stabilization Transformation (DESeq2, [77]) were loaded into IGV [134]. Pearson correlations between samples were visualized using the corrplot package. GR ChIP narrowPeak tracks were downloaded from ENCODE [11]. Intergenic enhancer RNAs were quantified using GenomicRanges coverage of the Chromatin

State Segmentation by HMM from ENCODE/Broad [25].

Bibliography

[1] F. Abdul-Rahman and D. Gresham. *High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies*, pages 15–24. Sprinter New York, New York, NY, 2018.

[2] K. Adelman and J. T. Lis. Promoter-proximal pausing of RNA polymerase II: Emerging roles in metazoans. *Nature Reviews Genetics*, 13(10):720–731, 2012.

[3] S. Anders, P. T. Pyl, and W. Huber. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, 2015.

[4] S. Bail, M. Swerdel, H. Liu, X. Jiao, L. Goff, R. Hart, , and M. Kiledjian. Differential regulation of microRNA stability. *RNA*, 16:1032–1039, 2010.

[5] K. Burger, B. Muhl, M. Keliner, M. Rohrmoser, A. Gruber-Eber, L. Windhanger, C. Friedel, L. Dolken, and D. Eick. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biology*, 10(10):1623–1630, 2013.

[6] D. Burow, M. Umeh-Garcia, M. True, C. Bakhai, D. Ardell, and M. Cleary. Dynamic regulation of mRNA decay during neural development. *Neural Development*, 10:11, 2015.

[7] T. Chang and J. Mendell. MicroRNAs in vertebrate physiology and human disease. *Annual Reviews of Genomics and Human Genetics*, 8:215–239, 2007.

[8] C. Chen, N. Ezzeddine, and A. Shyu. Messenger RNA half-life measurements in mammalian cells. *Methods in Enzymology*, 448:335–357, 2008.

[9] M. Cleary, C. Meiering, E. Jan, R. Guymon, and J. Boothroyd. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nature biotechnology*, 23(2):232–237, 2005. ISSN 1087-0156. 10.1038/nbt1061.

[10] E. Colvin, H. Ma, Y. Chen, A. Hernandez, and P. Fueger. Glucocorticoid-induced suppression of β-cell proliferation is mediated by Mig6. *Endocrinology*, 154(3):1039–1046, 2013.

[11] E. P. Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489:57–74, 2012.

[12] L. J. Core, J. J. Waterfall, and J. T. Lis. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, 322(5909):1845–1848, 2008.

[13] L. J. Core, A. L. Martins, C. G. Danko, C. Waters, A. Siepel, and J. T. Lis. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics*, 46(12):1311–1320, 2014.

[14] C. G. Danko, N. Hah, X. Luo, A. L. Martins, L. J. Core, J. T. Lis, A. Siepel, and W. L. Kraus. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Molecular Cell*, 50(2):212–222, 2013.

[15] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer. *In vivo* dynamics of RNA polymerase II transcription. *Nature Structural and Molecular Biology*, 14:796–806, 2007.

[16] K. De Bosscher, W. Vanden Berghe, and G. Haegeman. Mechanisms of anti-inflammatory action and of immunosuppression by glucocorticoids: Negative interference of activated glucocorticoid receptor with transcription factors. *Journal of Neuroimmunology*, 2109:16–22, 2000.

[17] S. de Pretis, T. Kress, M. J. Morelli, G. E. M. Melloni, L. Riva, B. Amati, and M. Pelizzola. INSPEcT: a computational tool to infer mRNA synthesis, processing

and degradation dynamics from RNA- and 4sU-seq time course experiments. *Bioinformatics*, 31(17):2829–2835, 2015.

[18] S. de Pretis, T. Kress, M. J. Morelli, A. Sabo, C. Locarno, A. Verrecchia, M. Doni, S. Campaner, B. Amati, and M. Pelizzola. Integrative analysis of RNA polymerase II and transcriptional dynamics upon MYC activation. *Genome Research*, 27:1658–1664, 2017.

[19] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 2013.

[20] L. Dölken, Z. Ruzsics, B. Rädle, C. Friedel, R. Zimmer, J. Mages, R. Hoffmann, P. Dickinson, T. Forster, P. Ghazal, and U. Koszinowski. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA (New York, N.Y.)*, 14(9):1959–1972, 2008. 10.1261/rna.1136108.

[21] J. Duan, J. Shi, X. Ge, L. Dolken, W. Moy, D. He, S. Shi, A. R. Sanders, J. Ross, and P. V. Gejman. Genome-wide survey of interindividual differences of RNA stability in human lymphoblastoid cell lines. *Scientific Reports*, 3:1318, 2013.

[22] E. E. Duffy and M. D. Simon. Enriching s^4U-RNA using methane thiosulfonate (MTS) chemistry. *Current Protocols in Chemical Biology*, 8(4):234–250, 2016.

[23] E. E. Duffy, M. Rutenberg-Schoenberg, C. D. Stark, R. R. Kitchen, M. B. Gerstein, and M. D. Simon. Tracking distinct RNA populations using efficient and reversible covalent chemistry. *Molecular Cell*, 59(5):858–866, 2015.

[24] G. Dujardin, C. Lafaille, M. la Mata, L. E. Marasco, M. J. Munoz, C. Le Jossic-Corcos, L. Corcos, and A. R. Kornblihtt. How slow RNA polymerase II elongation favors alternative exon skipping. *Molecular Cell*, 54:683–690, 2014.

[25] J. Ernst, P. Kheradpour, T. Mikkelsen, N. Shoresh, L. Ward, C. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis, and B. Bernstein.

Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*, 473(7345):43–49, 2011.

[26] A. Favre, R. Bezerra, E. Hajnsdorf, Y. Lemaigre-Dubreuil, and A. Expert-Bezancon. Substitution of uridine *in vivo* by the intrinsic photoactivable probe 4-thiouridine in *Escherichia coli* RNA. Its use for *E. coli* ribosome structural analysis. *European Journal of Biochemistry*, 160(7345):441–449, 1986.

[27] L. P. Ford, J. Watson, J. D. Keene, and J. Wilutz. ELAV proteins stabilize deadenylated intermediates in a novel *in vitro* mRNA deadenylation/degradation system. *Genes and Development*, 13:188–201, 1999.

[28] L. H. Fossom, C. R. Sterling, and W. A. Tank. Regulation of tyrosine hydroxylase gene transcription rate and tyrosine hydroxylase mRNA stability by cyclic AMP and glucocorticoid. *Molecular pharmacology*, 42:898–908, 1992.

[29] C. C. Friedel, L. Dolken, Z. Ruzsics, U. Koszinowski, and R. Zimmer. Conserved principles of mammalian transcription regulation revealed by RNA half-life. *Nucleic Acids Research*, 37(17):e115, 2009.

[30] G. Fuchs, Y. Voichek, S. Benjamin, S. Gilad, I. Amit, and M. Oren. 4sUDRB-seq: measuring genome-wide transcriptional elongation rates and initiation frequencies within cells. *Genome Biology*, 15(5):R69, 2014.

[31] G. Fuchs, Y. Voichek, M. Rabani, S. Benjamin, S. Gilad, I. Amit, and M. Oren. Simultaneous measurement of genome-wide transcription elongation speeds and rates of RNA polymerase II transition into active elongation with 4sUDRB-seq. *Nature Protocols*, 10:605–618, 2015.

[32] J.-M. Gallo, P. Jin, C. A. Thorton, H. Lin, J. Robertson, I. D'Souza, and W. W. Schlaepfer. The role of RNA and RNA processing in neurodegeneration. *Journal of Neuroscience*, 25(45):10372–1375, 2005.

[33] M. Gantier, C. McCoy, I. Rusinova, D. Saulep, D. Xu, A. Irving, M. Behlke, P. Hert-

zog, F. Mackay, , and B. Williams. Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. *Nucleic Acids Research*, 39:5692–5703, 2011.

[34] L. Gay, M. R. Miller, P. B. Ventura, V. Devasthali, Z. Vue, H. L. Thompson, S. Temple, H. Zong, M. D. Cleary, S. Kryn, and C. Q. Doe. Mouse TU tagging: a chemical/genetic intersectional method for purifying cell type-specific nascent RNA. *Genes & Development*, 27:98–115, 2013.

[35] A. George, R. Schiltz, and G. Hager. Dynamic access of the glucocorticoid receptor to response elements in chromatin. *International Journal of Biochemistry and Cell Biology*, 41(1):214–224, 2009.

[36] S. Ghosh and A. Jacobson. RNA decay modulates gene expression and controls its fidelity. *Wiley Interdisciplinary Reviews RNA*, 1(3):351–361, 2011.

[37] D. K. Granner, J.-C. Wang, and K. R. Yamamoto. *Regulatory Actions of Glucocorticoid Hormones: From Organisms to Mechanisms*, volume 872, pages 3–31. Sprinter, New York, NY, 2015.

[38] L. Gregersen, M. Schueler, M. Munschauer, G. Mastrobuoni, W. Chen, S. Kempa, C. Dieterich, , and M. Landthaler. MOV10 Is a 5' to 3' RNA helicase contributing to UPF1 mRNA target degradation by translocation along 3' UTRs. *Molecular Cell*, 54:573–585, 2014.

[39] N. Gromak, S. West, and N. J. Proudfoot. Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Molecular and Cellular Biology*, 26: 3986–3996, 2006.

[40] Y. Guo, J. Liu, S. Elfenbein, Y. Ma, M. Zhong, C. Qiu, Y. Ding, , and J. Lu. Characterization of the mammalian miRNA turnover landscape. *Nucleic Acids Research*, 43:2326–2341, 2015.

[41] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. J. Ascano, A. Jungkamp, M. Munschauer, A. Ulrich, G. Wardle, S. Dewell,

Z. M., and T. T. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, 141:129–141, 2010.

[42] N. Hah, C. G. Danko, L. Core, J. J. Waterfall, A. Siepel, J. T. Lis, and W. L. Kraus. A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell*, 145(4):622–634, 2011.

[43] M. Harpold, M. Wilson, and J. Darnell Jr. Chinese hamster polyadenylated messenger ribonucleic acid: Relationship to non-polyadenylated sequences and relative conservation during messenger ribonucleic acid processing. *Molecular and Cellular Biology*, 1:188–198, 1981.

[44] J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodrigues, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigo, and T. J. Hubbard. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Research*, 22:1760–1774, 2012.

[45] D. Z. Hazelbaker, S. Marquardt, W. Wlotzka, and S. Buratowski. Kinetic competition between RNA polymerase II and Sen1-dependent transcription termination. *Molecular Cell*, 49(1):55–66, 2013.

[46] B. He, C. Chen, L. Teng, and K. Tan. Global view of enhancerpromoter interactome in human cells. *Proceedings of the National Academy of Sciences*, 111(21):E2191–E2199, 2014.

[47] N. Heintzman, R. Stuart, G. Hon, Y. Fu, C. Ching, R. Hawkins, L. Barrera, S. Van Calcar, C. Qu, K. Ching, W. Wang, Z. Weng, R. Green, G. Crawford, and B. Ren. Distinct and predictive chromatin signatures of transcriptional promoter and enhancers in the human genome. *Nature Genetics*, 39:311–318, 2007.

[48] X. Hua, L. Chen, J. Wang, J. Li, and E. Wingender. Identifying cell-specific microRNA transcriptional start sites. *Bioinformatics*, 32(16):2403–2410, 2016.

[49] C. Jao and A. Salic. Exploring RNA transcription and turnover *in vivo* by using click chemistry. *Proceedings of the National Academy of Sciences of the United States of America*, 105(41):15779–15784, 2008.

[50] G. Jeschke. Conformational dynamics and distribution of nitroxide spin labels. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 72:42–60, 2013.

[51] T. Johnson, R. Chereji, D. Stavreva, S. Morris, G. Hager, and D. Clark. Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin *in vivo*. *Nucleic Acids Research*, 46(1):203–214, 2018.

[52] I. Jonkers and J. T. Lis. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16:167–177, 2015.

[53] I. Jonkers, H. Kwak, and J. T. Lis. Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife*, 3:e02407, 2014.

[54] A. Jubb, S. Boyle, D. Hume, and W. Bickmore. Glucocorticoid receptor binding induces rapid and prolonged large-scale chromatin decompaction at multiple target loci. *Cell Reports*, 21(11):3022–3031, 2017.

[55] S. Kaech and G. Banker. Culturing hippocampal neurons. *Nature Protocols*, 1(5): 2406–2415, 2006.

[56] M. Kaikkonen, N. Spann, S. Heinz, C. Romanonski, K. Allison, J. Stender, H. Chun, D. Tough, R. Prinjha, C. Benner, and C. Glass. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular Cell*, 51:310–325, 2013.

[57] G. Kenyon and T. Bruice. Novel sulfhydryl reagents. *Methods in Enzymology*, 47: 407–430, 1978.

[58] P. Kersey, J. Allen, M. Christensen, P. Davis, L. Falin, C. Grabmueller, D. Hughes, J. Humphrey, A. Kerhornou, J. Khobova, N. Langridge, M. McDowall, U. Maheswari, G. Maslen, M. Nuhn, C. Ong, M. Paulini, H. Pedro, I. Toneva, M. Tuli, B. Walts, G. Williams, D. Wilson, K. Youens-Clark, M. Monaco, J. Stein, X. Wei, D. Ware, D. Bolser, K. Howe, E. Kulesha, D. Lawson, and D. Staines. Ensembl Genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Research*, 42:D546–D552, 2014.

[59] D. Kim, B. Langmead, and S. Salzberg. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12:357–360, 2015.

[60] T.-K. Kim, M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bitp, P. F. Worley, G. Kreiman, and M. E. Greenberg. Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465:182–187, 2010.

[61] T.-K. Kim, M. Hemberg, and J. M. Gray. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harbor Perspectives in Biology*, 7(1): a018622, 2015.

[62] A. Kozomara and S. Griffith-Jones. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, 42:D68–D73, 2014.

[63] H. Kwak, N. J. Fuda, L. J. Core, and J. T. Lis. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, 339:950–953, 2013.

[64] L. Lam, O. Pickeral, A. Peng, A. Rosenwald, E. Hurt, J. Giltnane, L. Averett, H. Zhao, R. Davis, M. Sathyamoorthy, L. Wahl, E. Harris, J. Mikovits, A. Monks, M. Hollingshead, E. Sausville, and L. Staudt. Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biology*, 2:H0041, 2001.

[65] M. T. Lam, H. Cho, H. P. Lesch, D. Gosselin, S. Heinz, Y. Tanaka-Oishi, C. Benner, M. U. Kaikkonen, A. S. Kim, M. Kosaka, C. Y. Lee, A. Watt, T. R. Grossman, M. G. Rosenfeld, R. M. Evans, and C. K. Glass. Rev-Erbs repress macrophage gene

expression by inhibiting enhancer-directed transcription. *Nature*, 498(7455):511–515, 2013.

[66] M. T. Lam, W. Li, M. G. Rosenfeld, and C. K. Glass. Enhancer RNAs and regulated transcriptional programs. *Trends in Biochemical Sciences*, 39(4):170–182, 2014.

[67] B. Langmead and S. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9:357–359, 2012.

[68] B. Langmead, C. Trapnell, M. Pop, and S. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10:R25, 2009.

[69] B. Lenhard, A. Sandelin, and P. Carninci. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nature Reviews Genetics*, 13: 233–245, 2012.

[70] M. Levine, C. Cattoglio, and R. Tjian. Looping back to leap forward: transcription enters a new era. *Cell*, 157:13–25, 2014.

[71] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and . G. P. D. P. Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 24:2078–2079, 2009.

[72] H.-B. Li, J. Tong, S. Zhu, P. J. Batista, E. E. Duffy, J. Zhao, W. Bailis, G. Cao, L. Kroehling, Y. Chen, G. Wang, J. P. Broughton, Y. G. Chen, Y. Kluger, M. D. Simon, H. Y. Chang, Z. Yin, and R. A. Flavell. m^6A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature*, 548:338–342, 2017.

[73] W. Li, D. Notani, Q. Ma, B. Tanasa, E. Nunez, A. Y. Chen, D. Merkurjev, J. Zhang, K. Ohgi, X. Song, S. Oh, H.-S. Kim, C. K. Glass, and M. G. Rosenfeld. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 498:516–520, 2013.

[74] W. Li, M. T. Lam, and D. Notani. Enhancer RNAs. *Cell Cycle*, 13(20):3151–3152, 2014.

[75] C. Lin, A. Strom, V. Vega, S. Kong, A. Yeo, J. Thomsen, W. Chan, B. Doray, D. Bangarusamy, A. Ramasamy, L. Vergara, S. Tang, A. Chong, V. Bajic, L. Miller, J. Gustafsson, and E. Liu. Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome Biology*, 5(9):R66, 2004.

[76] Q. Liu, J. Wang, Y. Zhao, C.-l. Li, K. R. Stengel, P. Acharya, G. Johnston, S. W. Hiebert, and Y. Shyr. Identification of active miRNA promoters from nuclear run-on RNA sequencing. *Nucleic Acids Research*, 45(13):e121, 2017.

[77] M. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15:550, 2014.

[78] M. Martin. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.journal*, 17:10–12, 2011.

[79] A. Mayer, J. di Iulio, S. Maleri, U. Eser, J. Vierstra, A. Reynolds, R. Sandstrom, J. A. Stamatoyannopoulos, and L. S. Churchman. Native Elongating Transcript Sequencing Reveals Human Transcriptional Activity at Nucleotide Resolution. *Cell*, 161(3):541–554, 2011.

[80] M. McDowall, M. Harris, A. Lock, K. Rutherford, D. Staines, J. Bähler, P. Kersey, S. Oliver, and V. Wood. PomBase 2015: updates to the fission yeast database. *Nucleic Acids Research*, 43:D656–661, 2015.

[81] S. H. Meijsing. *Mechanisms of glucocorticoid-regulated gene transcription*, volume 872, pages 59–81. Sprinter, New York, NY, 2015.

[82] W. T. Melvin, H. B. Milne, A. A. Slater, H. J. Allen, and H. M. Keir. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesized RNA by affinity chromatography. *European Journal of Biochemistry*, 92 (2):373–379, 1978.

[83] H. Mi, A. Muruganujan, and P. D. Thomas. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Research*, 41(D1):D377–D386, 2013.

[84] M. Michel, C. Demel, B. Zacher, B. Schwalb, S. Krebs, H. Blum, J. Gagneur, and P. Cramer. TT-seq captures enhancer landscapes immediately after T-cell stimulation. *Molecular Systems Biology*, 13:920, 2017.

[85] C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D. Martin, A. Tresch, and P. Cramer. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Molecular Systems Biology*, 4(7):458, 2011. ISSN 1744-4292.

[86] M. R. Miller, K. J. Robinson, M. D. Cleary, and C. Q. Doe. TU-tagging: cell type-specific RNA isolation from intact complex tissues. *Nature Methods*, 6:439–441, 2009.

[87] E. Moehle, H. Braberg, N. Krogan, and C. Guthrie. Adventures in time and space: splicing efficiency and RNA polymerase II elongation rate. *RNA Biology*, 11:313–319, 2014.

[88] E. Mueller, C. Buck, P. Palenchar, L. Barnhart, and J. Paulson. Identification of a gene involved in the generation of 4-thiouridine in tRNA. *Nucleic Acids Research*, 26: 2606–2610, 1998.

[89] D. Mukherjee, M. Gao, P. J. O'Connor, R. Raijmakers, G. Pruijn, C. S. Lutz, and J. Wilusz. The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *The EMBO Journal*, 21:165–174, 2017.

[90] N. Mukherjee, L. Calviello, A. Hirsekorn, S. de Pretis, M. Pelizzola, and U. Ohler. Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nature Structural and Molecular Biology*, 24:86–96, 2017.

[91] G. Muse, D. Gilchrist, S. Nechaev, R. Shah, J. Parker, S. Grissom, J. Zeitlinger, and K. Adelman. RNA polymerase is poised for activation across the genome. *Nature Genetics*, 39:1507–1511, 2007.

[92] B. Neymotin, R. Athanasiadou, and D. Gresham. Determination of *in vivo* RNA kinetics using RATE-seq. *RNA*, 20:1645–1652, 2014.

[93] I. Nikolic, M. Vujicic, T. Saksida, T. Berki, S. Stosic-Grujicic, and I. Stojanovic. The role of endogenous glucocorticoids in glucose metabolism and immune status of MIF-deficient mice. *European Journal of Pharmacology*, 714(1-3):498–506, 2013.

[94] M. J. Nueda, S. Tarazona, and A. Conesa. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics*, 30(18):2598–2602, 2014.

[95] T. O'Brien and J. T. Lis. Rapid changes in Drosophila transcription after an instantaneous heat shock. *Molecular and Cellular Biology*, 13(6):3456–3463, 1993.

[96] C.-T. Ong and V. G. Corces. Enhancer function: new insights into the regulate of tissue-specific gene expression. *Nature Reviews Genetics*, 12:283–293, 2011.

[97] M. B. Ozturk, Y. Li, and V. Tergaonkar. Current insights to regulation and role of telomerase in human diseases. *Antioxidants (Basel)*, 6(1):17, 2017.

[98] V. Paakinaho, H. Makkonen, T. Jaaskelainen, and J. J. Palvimo. Glucocorticoid receptor activates poised *FKBP51* locus through long-distance interactions. *Molecular Endocrinology*, 24(3):511–525, 2010.

[99] D. Palmer and N. Restifo. Suppressors of cytokine signaling (SOCS) in T cell differentiation, maturation, and function. *Trends in Immunology*, 12(30):592–602, 2009.

[100] S. W. Peltz and J. Ross. Autogenous regulation of histone mRNA decay by histone proteins in a cell-free system. *Molecular and Cellular Biology*, 7(12):4345–4356, 1987.

[101] D. Polioudakis, N. Abell, and V. Iyer. MiR-191 regulates primary human fibroblast proliferation and directly targets multiple oncogenes. *PLoS ONE*, 10:e0126535, 2015.

[102] W. Polioudakis, Y. Morishima, M. Murphy, and M. Harrell. *Chaperoning of glucocorticoid receptors*, volume 172, pages 111–138. Sprinter, Berlin, Heidelberg, 2006.

[103] L. Puckett, S. Chambers, and J. Darnell. Short-lived messenger RNA in HeLa cells and its impact on the kinetics of accumulation of cytoplasmic polyadenylate. *Proceedings of the National Academy of Sciences*, 72:389–391, 1975.

[104] M. Rabani, J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit, and A. Regev. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature Biotechnology*, 29(5):436–442, 2011.

[105] M. Rabani, R. Raychowdhury, M. Jovanovic, M. Rooney, D. J. Stumpo, A. Pauli, N. Hacohen, A. F. Schier, P. J. Blackshear, N. Friedman, I. Amit, and A. Regev. Determining mRNA Decay Rates Using RNA Approach to Equilibrium Sequencing (RATE-Seq). *Cell*, 159(7):1698–1710, 2014.

[106] A. Rada-Iglesias, R. Bajpai, T. Swigut, S. Brugmann, R. Flynn, and J. Wysocka. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, 470:279–283, 2011.

[107] J. Rainer, J. Lelong, D. Bindreither, C. Mantinger, C. Ploner, S. Geley, and R. Kofler. Research resource: Transcriptional response to glucocorticoids in childhood acute lymphoblastic leukemia. *Molecular Endocrinology*, 26(1):178–193, 2012.

[108] S. Ramamoorthy and J. A. Cidlowski. Ligand-induced repression of the glucocorticoid receptor gene is mediated by an NCoR1 repression complex formed by long-range chromatin interactions with intragenic glucocorticoid response elements. *Molecular and Cellular Biology*, 33(9):1711–1722, 2013.

[109] T. E. Reddy, F. Pauli, R. O. Sprouse, N. F. Neff, K. M. Newberry, M. J. Garabedian, and R. M. Myers. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Research*, 19:2163–2171, 2009.

[110] T. E. Reddy, J. Gertz, E. Crawford, Gregory, M. J. Garabedian, and R. M. Myers. The hypersensitive glucocorticoid response specifically regulates period 1 and expression of circadian genes. *Molecular and Cell Biology*, 32(18):3756–3767, 2012.

[111] M. Robinson and G. Smyth. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, 23:2881–2887, 2007.

[112] M. Robinson, D. McCarthy, and G. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26:139–140, 2010.

[113] K. Rosenbloom, J. Armstrong, G. Barber, J. Casper, H. Clawson, M. Diekhans, T. Dreszer, P. Fujita, L. Guruvadoo, M. Haeussler, R. Harte, S. Heitner, G. Hickey, A. Hinrichs, R. Hubley, D. Karolchick, K. Learned, B. Lee, C. Li, K. Miga, N. Nguyen, B. Paten, B. Raney, A. Smit, M. Speir, A. Zweig, D. Haussler, R. Kuhn, and W. Kent. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research*, 43: D670–681, 2015.

[114] A. Rueda, G. Barturen, R. Lebron, C. Gomez-Martin, A. Alganza, J. Oliver, and M. Hackenberg. sRNAbench: profiling of small RNAs and its sequence variants in single or multi-species high-throughput experiments. *Methods in Next Generation Sequencing*, 43:W467–W473, 2014.

[115] S. Rüegger and H. Groβhans. MicroRNA turnover: when, how, and why. *Trends in Biochemical Sciences*, 37:436–446, 2012.

[116] J. Russo, A. M. Heck, J. Wilusz, and C. J. Wilusz. Metabolic labeling and recovery of nascent RNA to accurately quantify mRNA stability. *Methods*, 120:39–48, 2017.

[117] V. Saint-André, E. Batsché, C. Rachez, and C. Muchardt. Histone H3 lysine 9 trimethylation and HP1gamma favor inclusion of alternative exons. *Nature Structural and Molecular Biology*, 18:337–344, 2011.

[118] T. Saldi, M. A. Cortazar, R. M. Sheridan, and D. L. Bentley. Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *Journal of Molecular Biology*, 428(12):2623–2635, 2016.

[119] J. A. Schofield, E. E. Duffy, L. Kiefer, M. C. Sullivan, and M. D. Simon. TimeLapse-

seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nature Methods*, (in press), 2018.

[120] B. Schwalb, M. Michel, B. Zacher, K. Frühauf, C. Demel, A. Tresch, J. Gagneur, and P. Cramer. TT-seq maps the human transient transcriptome. *Science*, 352(6290): 1225–1228, 2016.

[121] B. Scruggs, D. Gilchrist, S. Nechaev, G. Muse, A. Burkholder, D. Fargo, and K. Adelman. Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Molecular Cell*, 58(6):1101–1112, 2015.

[122] A. Shyu, M. Greenberg, and J. Belasco. The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. *Genes and Development*, 3: 60–72, 1989.

[123] R. Singer and S. Penman. Messenger RNA in HeLa cells: Kinetics of formation and decay. *Journal of Molecular Biology*, 78:321–334, 1973.

[124] J. Singh and R. A. Padgett. Rates of *in situ* transcription and splicing in large human genes. *Nature Structural and Molecular Biology*, 16:1128–1133, 2009.

[125] A. So, C. Chaivorapol, E. Bolton, H. Li, and K. Yamamoto. Determinants of cell- and gene-specific transcriptional regulation by the glucocorticoid receptor. *PLoS Genetics*, 3:e9410, 2007.

[126] A. R. Sonawane, J. Platig, M. Fagny, C.-Y. Chen, J. N. Paulson, C. M. Lopes-Ramos, D. L. DeMeo, J. Quackenbush, K. Glass, and M. L. Kuijer. Understanding tissue-specific gene regulation. *Cell Reports*, 21(4):1077–1088, 2017.

[127] D. Spies, P. Renz, T. Beyer, and C. Ciaudo. Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Briefings in Bioinformatics*, page bbx115, 2017.

[128] D. Su, C. Chan, C. Gu, K. Lim, Y. Chionh, M. McBee, B. Russell, I. Babu, T. Begley, and P. Dedon. Quantitative analysis of ribonucleoside modifications in tRNA by HPCL-coupled mass spectrometry. *Nature Protocols*, 9:828–841, 2014.

[129] W. Sui, C. Cao, C. W, C. J, X. W, L. P, L. Guo, and D. Y. Comparative analyses of histone H3K9 trimethylations in the heart and spleen of normal humans. *Genetics and Molecular Research*, 13(1):1697–1706, 2014.

[130] M. Sun, B. Schwalb, D. Schulz, N. Pirkl, S. Etzold, L. Larivire, K. Maier, M. Seizl, A. Tresch, and P. Cramer. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Research*, 22:1350–1359, 2012.

[131] H. Tani and N. Akimitsu. Genome-wide technology for determining RNA stability in mammalian cells: Historical perspective and recent advantages based on modified nucleotide labeling. *RNA biology*, 9(10):1233–1238, 2012. 10.4161/rna.22036.

[132] H. Tani, R. Mizutani, K. Salam, K. Tano, K. Ijiri, A. Wakamatsu, T. Isogai, Y. Suzuki, and N. Akimitsu. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome research*, 22(5):947–956, 2012. 10.1101/gr.130559.111.

[133] C. N. Tennyson, H. J. Klamut, and R. G. Worton. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nature Genetics*, 9:184–190, 1995.

[134] J. Thorvaldsdottir, H.and Robinson and J. Mesirov. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinformatics*, 14:178–192, 2013.

[135] H.-J. Ting, J. Messing, Y.-K. Sayeda, and Y.-F. Lee. Identification of microRNA-98 as a therapeutic target inhibiting prostate cancer growth and a biomarker induced by vitamin D. *Journal of Biological Chemistry*, 288:1–9, 2013.

[136] J. Tomorsky, L. DeBlander, C. G. Kentros, C. Q. Doe, and C. M. Niell. TU-tagging: A method for identifying layer-enriched neuronal genes in developing mouse visual cortex. *eNeuro*, 4(5):pi, 2017.

[137] C. Trapnell, B. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. van Baren, S. Salzberg, B. Wold, and L. Pachter. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28:511–515, 2010.

[138] M. Tucker, R. R. Staples, M. A. Valencia-Sanchez, D. Muhlrad, and R. Parker. Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in *Saccharomyces cerevisiae*. *The EMBO Journal*, 21:1427–1436, 2002.

[139] A. Veloso, K. S. Kirkconnell, B. Magnuson, B. Biewen, M. T. Paulsen, T. E. Wilson, and M. Ljungman. Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Research*, 24:896–905, 2014.

[140] C. Vockley, A. D'Ippolito, M. I.C., W. Majoros, A. Safi, L. Song, G. Crawford, and T. Reddy. Direct GR binding sites potentiate clusters of TF binding across the human genome. *Cell*, 166(5):1269–1281, 2016.

[141] J. Wang, M. Derynck, D. Nonaka, D. Khodabakhsh, C. Haqq, and K. Yamamoto. Chromatin immunoprecipitation (ChIP) scanning identifies primary glucocorticoid receptor target genes. *Proceedings of the National Academy of Sciences*, 101:1560315608, 2004.

[142] Y. Wang, C. Liu, J. Storey, R. Tibshirani, D. Herschlag, and P. Brown. Precision and functional specificity in mRNA decay. *Proceedings of the National Academy of Sciences*, 99:5860–5865, 2002.

[143] C. J. Wilutz and J. Wilutz. Bringing the role of mRNA decay in the control of gene expression into focus. *Trends in Genetics*, 20(10):491–497, 2004.

[144] L. Windhanger, T. Bonfert, K. Burger, Z. Ruzsics, S. Krebs, S. Kauffmann, G. Malterer, A. L'Hernault, M. Schilhaben, S. Schreiber, P. Rosenstiel, R. Zimmer, D. Eick, C. C. Friedel, and L. Dölken. Ultrashort and progressive 4sU-tagging reveals key

characteristics of RNA processing at nucleotide resolution. *Genome Research*, 22(10): 2031–2042, 2012.

[145] J. Winter, S. Jung, S. Keller, R. Gregory, , and S. Diederichs. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nature Cell Biology*, 11(11):228–234, 2009.

[146] H. Xu, X. Luo, J. Qian, X. Pang, J. Song, G. Qian, J. Chen, and S. Chen. FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One*, 7:e1004610, 2012.

[147] E. Yang, E. van Nimwegen, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnasco, and J. E. J. Darnell. Decay rates of human mRNAs: correlation with functional characteristics and sequence attributes. *Genome Research*, 13(8):1863–1872, 2003.

[148] F. Yang, Q. Ma, Z. Liu, W. Li, Y. Tan, C. Jin, W. Ma, Y. Hu, J. Shen, K. A. Ohgi, F. Telese, W. Liu, and M. G. Rosenfeld. Glucocorticoid receptor: MegaTrans switching mediates the repression of an ERα-regulated transcriptional program. *Molecular Cell*, 66(3):321–331.e6, 2017.

[149] X. Yi, V. M. Tesmer, I. Savre-Train, J. W. Shay, and W. E. Wright. Both transcriptional and posttranscriptional mechanisms regulate human telomerase template RNA levels. *Molecular and cellular biology*, 19(6):3989–3997, 1999.

[150] A. Yoshimura, T. Naka, and M. Kubo. SOCS proteins, cytokine signaling and immune regulation. *Nature Reviews in Immunology*, 7(6):454–465, 2007.

[151] G. M. Zeiner, M. D. Cleary, A. E. Fouts, C. D. Meiring, E. S. Mocarski, and J. C. Boothroyd. RNA analysis by biosynthetic tagging using 4-thiouracil and uracil phosphoribosyltransferase. *Methods in Molecular Biology*, 419:135–146, 2008.

[152] B. Zhang, G. Yang, Y. Chen, Y. Zhao, P. Gao, B. Liu, H. Wang, and Z.-L. Zheng. C-terminal domain (CTD) phosphatase links Rho GTPase signaling to Pol II CTD phosphorylation in Arabidopsis and yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 113(50):E8197–E8206, 2016.

Appendix A

MTS biotin enrichment protocol

This chapter is an excerpt from:

Duffy, E.E. and Simon, M.D. (2016) Enriching s^4U-RNA Using Methane Thiosulfonate (MTS) Chemistry. *Curr. Protoc. Chem. Biol.* 8(4):234-250.

A.1 Metabolic labeling of cells and isolation of total cellular RNA

The protocol describes the s^4U metabolic labeling of HEK293T cells in a 6-well plate, and includes two time points and one control, with each sample in duplicate (six total, Figure 1A). This protocol can be adapted to enrich s^4U-RNA from several different types of input RNAs from various sources. Labeling experiments with s^4U have been described in several organisms and cellular systems including bacterial (e.g., [26]), yeast (e.g., [85, 130]), insect (e.g., [6]), and mammalian (e.g., [20, 104]) cells. The experimental design can also be adapted to different time points or treatment conditions. The details of the metabolic labeling described here offer a guide to one successful labeling regime that provides strong enrichment.

1. Plate HEK293T cells in a 6-well plate at 3 x 10^5 cells/well and allow the cells to recover in RPE1 media (without s^4U) overnight.

    - Cells are generally plated at a low enough density so that they will not reach

confluence during the experiment. In this protocol, the cells are expected to reach ∼80% confluence by the end of the experiment. While I generally do not split cells during treatment, s^4U treatments have been successful through splitting and re-plating cells (including trypsin treatment).

2. Freshly dissolve solid s^4U in water to 1 M (260 mg/mL). Dilute 14 μL of this 1 M s^4U stock into 20 mL of RPE1 media (to 700 μM s^4U final) and mix by vortexing for 10 seconds. Aspirate media from cells in all six wells. For wells 1-2, add 3 mL/well of s^4U containing media. For the samples in wells 3-6, add 3 mL/well of media without s^4U.

    - While freshly dissolving s^4U solid prior to the experiment is recommended, enrichment has been observed when using stocks that were stored at -20°C for more than a week.
    - Successful experiments have been performed using a range of concentration of s^4U in the media. Because of the efficiency of MTS chemistry, lower concentrations (as low as 100 μM) of s^4U were found to provide sufficient enrichment. In other cases, s^4U can be used up to 1 mM—these higher concentrations may be advantageous for especially short treatments [30, 120]. One concern that has been raised about s^4U is its potential for toxicity [5]. I have found that at low concentrations of s^4U (100 μM), the total RNA levels in cells are very similar between treated and untreated cells, even after 22 days of treatment [23], consistent with others' findings [38, 41], suggesting that any perturbations to the steady state RNA levels caused by these s^4U treatment conditions are minimal.

3. After 1.5h of s^4U labeling, remove media from cells in wells 3-4 and add 3 mL of s^4U-containing media (described in step 2).

4. After the samples in wells 1-2 have been exposed to s^4U for 2h total (and the samples in wells 3-4 have been exposed for 30 min), remove the media from all wells and immediately add 1 mL TRIzol. Pipet up and down five times and transfer the TRIzol samples to labeled 1.7 mL microfuge tubes.

- Ideally the RNA purification will be performed immediately, but I have observed successful enrichment for RNA samples that have been frozen at -80°C for up to three months.

A.2 RNA isolation

5. Add 200 μL chloroform to each of the 1 mL TRIzol samples from step 4. Shake the tubes vigorously for 15 s and let sit for 2 min.

6. Centrifuge the tubes for 5 min at 12,000 x g, 4°C. Transfer aqueous phase ($\sim$500 μL) to new labeled tubes.

7. To each aqueous phase from step 6, add 0.5 mL of 100% isopropanol, 1 μL 100 mM DTT (final concentration: 0.1 mM DTT) and 1 μL RNase-free glycogen (5-10 μg). Incubate samples at room temperature for 10 min.

    - The s^4U-RNA is light sensitive and prone to oxidation. While these steps can be performed under standard laboratory lighting, try to minimize the time of light exposure. The DTT is included to help minimize oxidation of the s^4U.

8. Centrifuge samples at 12,000 x g for 20 min at 4°C. Carefully remove the isopropanol from the RNA/glycogen pellet.

9. Add 1mL of 75% ethanol to the pellet, shake for 10 s and centrifuge at 12,000 x g for 10 min at 4°C.

10. Remove the ethanol completely from the RNA/glycogen pellet. To do so, first remove the majority of the ethanol with a P1000, then spin the tubes again on a counter top centrifuge. Use a gel-loading tip to remove the remaining ethanol. Let the pellet air-dry for 2 min by leaving the tube open under a Kimwipe cover.

11. Resuspend each pellet in 20 μL of RNase-free water. Measure the RNA concentrations using a Nanodrop spectrophotometer. Adjust the volume to bring RNA concentration to 200 ng/μL with water.

- Generally 10-20 μg of material is retrieved from 10^6 HEK293T cells after this step, which includes RNA as well as any contamination from DNA and free nucleotides that were carried through the TRIzol extraction and precipitation.

12. Add 10x Turbo DNase buffer to a final 1x concentration. Add 1 μL Turbo DNase (2U) per 10 μg of RNA. Invert the tube several times and concentrate the liquid in the bottom by centrifugation for $\sim$5 s. Incubate the reaction at 37°C for 30 min.

13. Prepare and label six (one per sample) phase lock gel tubes by centrifugation at 10,000 x g for 30 sec according to the manufacturer's directions.

14. Adjust the volume of the RNA samples from step 12 to 315 μL with water. Add 35 μL of sodium acetate (3M stock, pH 5.5) and add samples to phase lock tubes. Then add 350 μL phenol:chloroform:isoamyl alcohol (ratio 25:24:1 and pH 7.7-8.3) to each phase lock tube. Shake the tubes vigorously for 15 s and let incubate at room temperature for 2 min. Centrifuge at 12,000 x g for 10 min at 4°C.

15. Transfer the aqueous phase into in a new tube. Repeat steps 7-10 to precipitate the RNA with isopropanol.

16. Resuspend the RNA pellet in 30 μL of RNase-free water.

17. Assay RNA quality by Nanodrop to calculate RNA concentration and also perform a UV-Vis scan for s^4U concentration.

    - Generally this protocol yields 10-12 μg of RNA from 10^6 HEK293T cells. s^4U absorbs light at 334 nm. In principle this absorbance can be used to calculate the amount of s^4U in the sample, but I have found that conventional Nanodrop is not sufficiently sensitive to allow accurate quantitation. However, this scan does provide a qualitative check to see that there is a local maximum at 334 nm.

A.3 RNA shearing (optional)

While this protocol successfully enriches s^4U-RNA with unsheared RNA samples, I have found that RNA shearing can increase yields and decrease background. As most short-read RNA-seq protocols require shearing of the RNA samples as part of a library preparation, it is convenient to perform the shearing prior to enrichment if it does not conflict with experimental design.

18. Adjust the RNA samples from step 16 to 40 μL with water. Add 40 μL 2x fragmentation buffer and place sample at 94°C for exactly 4 min.

19. Quickly spin the RNA sample on a countertop centrifuge and immediately place on ice.

20. Add 20 μL of 250 mM EDTA (final concentration: 50 mM EDTA) to each sample, mix by vortexing and incubate on ice for 2 min.

A.4 Modified RNeasy MinElute Cleanup

21. Add 350 μL buffer RLT and 250 μL 100% EtOH to each RNA sample from step 16 or 20, mix well by pipetting. Apply these samples to RNeasy columns.

22. Centrifuge the columns 15 s at 12,000 x g, 4°C. Discard the flow through.

23. To each column, add 500 μL RPE buffer supplemented with 35 μL of 1% β -me (final concentration: 10 mM β -me).

    - Note that the addition of a reducing agent in this step is important to reduce any disulfides that have formed with the s^4U.

24. Centrifuge the samples 15 s at 12,000 x g, 4°C. Discard the flow through.

25. Add 500 μL freshly prepared 80% EtOH and centrifuge the samples for 2 min at 12,000 x g, 4°C. Discard the flow through.

26. Switch the columns to new 2 mL collection tubes. Centrifuge the samples 5 min at maximum speed, 4°C.

27. Transfer column to a 1.7 mL microfuge tube, and add 14 μL RNase-free water. Centrifuge the samples 1 min at 12,000 x g, 4°C.

A.5 Biotinylation of s^4U-RNA with the activated disulfide methane thiosulfonate (MTS) bitoin

28. Dilute solid MTSEA-biotin-XX (MW 607.7 g/mol) in dry DMF to 1 mg/mL (1.64 mM) to make a concentrated stock. The MTSEA-biotin-XX stocks are stable at -20°C for at least 3 months.

29. To biotinylate the RNA samples, mix the following reagents in a 1.7 mL microfuge tube:

    - 2 to 5 μg of RNA from step 27
    - 1 μL 1 M HEPES, pH 7.4 (final concentration: 20 mM HEPES, pH 7.4)
    - 1 μL 0.5 M EDTA (final concentration: 1 mM EDTA)
    - Nuclease-free water to 40 μL total volume

30. Dilute 4 μL of the 1 mg/mL MTSEA-biotin-XX stock into 76 μL DMF (50 μg/mL, 82 μM final concentration) and mix by vortexing.

31. Add 10 μL of the MTSEA-biotin-XX solution from step 30 into each reaction from step 29 (results in the addition of 500 ng MTSEA-biotin-XX to each reaction, final concentrations 16.4 μM MTSEA-biotin-XX, and 20% DMF).

    - These conditions include excess MTSEA-biotin-XX and should be sufficient for up to 5 μg of RNA. For larger scale reactions with more RNA, I recommend increasing the volume of the reaction but retaining the concentrations of each component.

32. Cover the reactions with foil and incubate these reactions at room temperature in the dark for 30 min with rotation.

A.6 Remove unreacted MTS-biotin from RNA samples

33. Adjust the volume of each sample to 100 μL by adding 50 μL of water. Add the solutions to labeled phase lock tubes and add 100 μL chloroform:isoamyl alcohol to each tube.

34. Shake the samples vigorously for 15 s and let sit 2 min.

35. Centrifuge the samples for 5 min at 12,000 x g, 4°C. For each sample, transfer the aqueous phase (100 μL) to a new labeled 1.7 mL microfuge tube.

36. To each aqueous phase from step 35, add 350 μL buffer RLT and 250 μL 100% EtOH, mix well by pipetting, and apply to an RNeasy column.

37. Centrifuge the columns for 15 s at 12,000 x g, 4°C. Discard the flow through.

38. Add 500 μL buffer RPE to each column and centrifuge the columns for 15 s at 12,000 x g, 4°C. Discard the flow through.

39. Add 500 μL freshly prepared 80% EtOH to each column and centrifuge the columns for 2 min at 12,000 x g, 4°C. Discard the flow through.

40. Switch the columns to new 2 mL collection tubes. Centrifuge samples 1 min at maximum speed, 4°C.

41. Transfer the columns to new microfuge tubes. Add 50 μL RNase-free water to the center of each column. Elute the RNA by centrifuging these samples for 1 min at 12,000 x g, 4°C. Proceed immediately to isolation of the biotinylated s^4U-RNA.

A.7 Block streptavidin beads

This portion of the protocol can be carried out in parallel with the RNA biotinylation step, as steps 28-41 require ∼1h to complete.

42. For six samples, aliquot 66 μL Dynabeads MyOne Streptavidin C1 beads into one 1.7 mL microfuge tube. Place the tube in a magnetic rack for 2 min and remove supernatant with a pipet. Prepare 10 μL of beads per sample and include 10% extra volume.

43. Wash the beads twice by resuspending them in 500 μL nuclease-free water and mixing by pipetting up and down five times. Place the tubes in a magnetic rack for 2 min and remove supernatant with a pipet after each rinse.

44. Wash beads twice with 500 μL high salt wash buffer, mixing and capturing the beads as described in step 43.

45. Add 330 μL freshly made bead blocking buffer, resuspend the bead mixture completely by pipet, and incubate for 1h at room temperature. Place the tubes in a magnetic rack for 2 min and remove the supernatant with a pipet.

46. Wash the beads twice with 500 μL high salt wash buffer. Place the tubes in a magnetic rack for 2 min and remove the supernatant with a pipet.

47. Resuspend the beads in 660 μL of high salt wash buffer. Pipet up and down and immediately aliquot 100 μL of bead suspension into each of six PCR tubes in a PCR-strip.

48. Immediately before applying the RNA samples (step 49), capture the beads in a 96-well magnetic rack and remove the buffer by pipet.

A.8 Isolate s^4U-containing transcripts with streptavidin beads

49. Add 5 μL high salt wash buffer to the 50 μL RNA solutions from step 34 and mix well. Add each sample to beads from step 48. Cover with foil and incubate the samples at room temperature for 15 min on a rotator at 30 rpm (or similar) to ensure the beads are mixing during the incubation.

50. Place the tubes in a 96-well magnetic rack for 2 min and remove supernatant with a pipet. Save the supernatant on ice as Flow through.

51. Wash the beads three times by resuspending them in 100 μL high salt wash buffer and mix by inverting the tubes five times rapidly. Quickly centrifuge the tubes. Place the tubes in a magnetic rack for 2 min and remove the supernatant with a pipet after each rinse.

52. Resuspend beads in 100 μL TE buffer and mix by inversion as in step 51. Incubate tubes at 55°C for 15 min in a thermocycler. Place the tubes in a magnetic rack for 2 min and remove the supernatant with a pipet.

53. Wash the beads two times by resuspending them in 100 μL TE buffer (pre-heated to 55°C in a heat block) and mix by inverting the tubes five times rapidly. Quickly centrifuge the tubes. Place the tubes in a magnetic rack for 2 min and remove the supernatant with a pipet after each rinse.

54. Add 25 μL freshly made elution buffer to the beads, mix the beads by inversion as in step 51, wrap the tubes in foil, and incubate them at room temperature in the dark for 15 min with rotation as in step 49.

    - This elution buffer will reduce the disulfide bond that formed between biotin and 4-thiouridine, thereby eluting the s^4U-RNA and leaving biotin bound to the streptavidin beads.

55. Quickly centrifuge the tubes and capture the beads in a 96-well magnetic rack for 2 min. Carefully retrieve the supernatant with a pipet and save this sample as elution.

56. Add another 25 μL elution buffer and immediately place tubes in a 96-well magnetic rack for 2 min. Remove the supernatant with a pipet and combine this elution for each sample with the corresponding elution from step 55. Place samples on ice.

Appendix B

MTS resin enrichment protocol

Metabolic labeling of cells and isolation of total RNA is carried out as in Appendix A (MTS biotin enrichment protocol) to step 27.

B.1 MTS resin synthesis

MTS resin should be prepared fresh for every experiment and can be stored for several hours at 4°C.

1. For six samples, add 660 μL nuclease-free water into one 1.7 mL microfuge tube. Add 66 μL NHS magnetic sepharose beads and pipet up and down until all beads detach from inside the pipet tip. Place the tube in a magnetic rack for 30s and remove supernatant with a pipet.

    - Prepare 10 μL of beads per sample and include 10% extra volume.
    - When pipetting NHS magnetic sepharose beads for the first time, aliquot beads into 1.7 mL microfuge tubes, 100 μL per tube. Beads tend to stick to the sides of tubes, so for best results, quick spin tube, use isopropanol supernatant to wash sides of tube for any beads stuck to walls, and then resuspend beads well by pipetting.
    - NHS magnetic sepharose should be replaced every 6 months or when binding capacity decreases, see binding capacity assay below.

2. Add 660 μL ice cold 1 mM HCl to beads, vortex well and incubate at room temperature with rotation for 2 min. Place the tube in a magnetic rack for 30s and remove supernatant with a pipet.

3. Wash the beads twice by resuspending them in 660 μL 1x PBS and mixing by vortexing. Place the tubes in a magnetic rack for 30s and remove supernatant with a pipet after each rinse.

4. Add 66 μL MTSEA dissolved in 1x PBS (10 mg/mL) plus 0.5 μL DIEA, resuspend the bead mixture completely by pipet, and incubate for 15 min at room temperature with rotation. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

    - MTSEA does not need to be prepared fresh for every experiment, but it will hydrolyze over time. MTSEA stocks should be made fresh every 2 months.

5. Wash the beads twice with 660 μL PBS. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

6. Resuspend the beads 660 μL blocking buffer (1x PBS, 0.05% acetic anhydride) and resuspend the bead mixture completely by pipet. Incubate for 15 min at room temperature with rotation. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

7. Wash the beads twice with 660 μL PBS. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

8. Wash the beads twice with 660 μL quench buffer (1 M Tris pH 7.4, 200 mM ethanolamine). Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

9. Resuspend the beads 660 μL quench buffer and resuspend the bead mixture completely by pipet. Incubate for 15 min at room temperature with rotation. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

10. Wash the beads twice with 660 μL quench buffer. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

11. Wash the beads twice with 660 μL binding buffer (10 mM HEPES pH 7.4, 1 mM EDTA, 100 mM NaCl, 0.05% Tween). Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

12. Resuspend the beads in 660 μL of binding buffer. Pipet up and down and aliquot 100 μL of bead suspension into each of six PCR tubes in a PCR-strip.

    - Alternatively, beads can be stored in binding buffer at 4°C for several hours.

B.2 Test resin binding capacity (optional)

This step is recommended to test the success of MTS resin coupling before precious RNA samples are incubated with resin, as well as to assay any variation between batches of beads or decrease in coupling efficiency over time.

13. Mix the following reagents in a 1.7 mL microfuge tube:

    - 1 μL 66 mM s^4U nucleoside dissolved in water
    - 1.5 μL 10x pH 8 binding buffer (10 mM Tris pH 8, 1 mM EDTA, 100 mM NaCl)
    - 3 μL DMF

14. Add s^4U nucleoside in binding buffer to 100 μL beads from step 12 and mix well. Cover with foil and incubate the samples at room temperature for 15 min on a rotator at 30 rpm (or similar) to ensure the beads are mixing during the incubation.

15. Place the tubes in a 96-well magnetic rack for 30s and remove supernatant with a pipet.

16. Wash the beads three times by resuspending them in 100 μL high salt wash buffer and mix by inverting the tubes five times rapidly. Quickly centrifuge the tubes. Place the tubes in a magnetic rack for 2 min and remove the supernatant with a pipet after each rinse.

17. Add 10 μL freshly made elution buffer to the beads, mix the beads by inversion as in step 16, wrap the tubes in foil, and incubate them at room temperature in the dark for 15 min with rotation as in step 14.

    - This elution buffer will reduce the disulfide bond that formed between biotin and 4-thiouridine, thereby eluting the s^4U-RNA and leaving biotin bound to the streptavidin beads.

18. Quickly centrifuge the tubes and capture the beads in a 96-well magnetic rack for 30s. Carefully retrieve the supernatant with a pipet and save this sample.

19. Assay s^4U yield by Nanodrop using UV-Vis absorbance at 334 nm. Quantify s^4U concentration relative to a standard curve of s^4U.

    - Expected range of s^4U binding capacity: 300-600 pmol per μL MTS resin.

B.3 Isolate s^4U-containing transcripts with MTS resin

20. Mix the following reagents in a 1.7 mL microfuge tube:

    - 1-5 μg of RNA from step 27 (Appendix A)
    - 1.5 μL 10x binding buffer (100 mM HEPES, pH 7.4, 10 mM EDTA, 100 mM NaCl, 0.5% Tween)
    - 3 μL DMF
    - Nuclease-free water to 15 μL total volume

    Note: Reserve an additional 10% of RNA (100-500 ng) in a PCR tube and label "total RNA"

21. Add 15 μL RNA solutions from step 20 to beads from step 14 and mix well by pipetting. Cover with foil and incubate the samples at room temperature for 2h on a rotator at 30 rpm (or similar) to ensure the beads are mixing during the incubation.

22. Place the tubes in a 96-well magnetic rack for 30s and remove supernatant with a pipet.

23. Wash the beads by resuspending them in 100 μL 1x binding buffer and mix by inverting the tubes five times rapidly. Cover with foil and incubate the samples at room temperature for 5 min with rotation. Quickly centrifuge the tubes. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

24. Wash the beads two times by resuspending them in 100 μL high salt wash buffer. Mix and incubate as in step 23 after each wash. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet after each rinse.

25. Wash the beads two times by resuspending them in 100 μL denaturing buffer (8M guanidinium hydrochloride). Mix and incubate as in step 23 after each wash. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet after each rinse.

26. Wash the beads two times by resuspending them in 100 μL binding buffer. Mix and incubate as in step 23 after each wash. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet after each rinse.

27. Resuspend beads in 100 μL TE buffer and mix by inversion as in step 23. Incubate tubes at 55°C for 5 min in a thermocycler. Place the tubes in a magnetic rack for 30s and remove the supernatant with a pipet.

28. Repeat step 27 twice more.

29. Add 10 μL freshly made elution buffer to the beads, mix the beads by inversion as in step 23, wrap the tubes in foil, and incubate them at room temperature in the dark for 15 min with rotation.

    - This elution buffer will reduce the disulfide bond that formed between biotin and 4-thiouridine, thereby eluting the s^4U-RNA and leaving biotin bound to the streptavidin beads.

30. Quickly centrifuge the tubes and capture the beads in a 96-well magnetic rack for 30s. Carefully retrieve the supernatant with a pipet and save this sample on ice.

B.4 Assay s^4U-RNA yield and fold enrichment by RT-qPCR

Yield can be assayed by several methods including Qubit, nanodrop, and RNA Bioanalyzer, but I recommend RT-qPCR for assaying fold enrichment due to the high sensitivity and dynamic range of this assay.

31. Add 7 μL RNA from step 30 to a chilled PCR tube. In a separate tube, add total RNA from step 20. To each sample (enriched and input RNA), add 2 μL Superscript VILO master mix and 1 μL Superscript III Reverse Transcriptase. Mix well by inversion three times. Quick spin tubes and place in a thermocylcer for:

    - 25°C, 10 min
    - 42°C, 60 min
    - 85°C, 5 min
    - 4°C, ∞

32. Dilute cDNA with 50 μL nuclease-free water. Dilute primers (*CDKN1B*, *UPF1* and *RNA18S*) to 4 μM as follows and mix well by vortexing:

    - 500 μL nuclease-free water
    - 2 μL 100 μM forward primer
    - 2 μL 100 μM reverse primer

33. Add 10 μL iTaq Universal Sybr Green Supermix, 5 μL diluted cDNA, and 5 μL primer mix in a 96-well PCR plate. Mix by pipetting up and down 8 times, avoiding bubbles. Place in qPCR machine for:

    - 95°C, 30 sec
    - 95°C, 5 sec
    - 60°C, 30 sec
    - Plate read, go to step 2 40 times

- 4°C, ∞

34. Calculate s^4U-RNA yield using the equation

$$\%Input = 10 * 2^{C_t(Enriched) - C_t(Input)} \tag{B.1}$$

35. Calculate fold enrichment over a -s^4U control using the equation:

$$FoldEnrichment = \frac{\%Input_{+s^4U}}{\%Input_{-s^4U}} \tag{B.2}$$

Appendix C

List of primers and oligos

Target	Organism	Orientation	Sequence
RPL18A	*H. sapiens*	Forward	GGAGAGCACGCCATGAAG
		Reverse	AAGATTCGCATGCGGTAGAG
MOV10	*H. sapiens*	Forward	ACAGGTGGAGAAAATCCGTTAC
		Reverse	TCTTGGCCTTGGAATTCTTC
HOXA9	*H. sapiens*	Forward	CCCCATCGATCCCAATAA
		Reverse	CACCGCTTTTTCCGAGTG
CBX6	*H. sapiens*	Forward	GCTGAGCAAGATGGAGCTGT
		Reverse	CCCTTCATTTCACCAGGTA
UPF1	*H. sapiens*	Forward	AGATCACGGCACAGCAGA
		Reverse	GTGGCAGAAGGGTTTTCCTT
CDKN1B	*H. sapiens*	Forward	TTTGACTTGCATGAAGAGAAGC
		Reverse	AGCTGTCTCTGAAAGGGACATT
RNA28S	*S. pombe*	Forward	TGAGAAGGGATGTTGGACCTGCTT
		Reverse	ATTGCGTCAACACCACTTTCTGGC
PER1	*H. sapiens*	Forward	GATCTTTCTTCCCCTACTCCCCG
		Reverse	GGCGCTCAGAAAATGCTCAGTAG
TSC22D3	*H. sapiens*	Forward	TATTGAATTCTGGCTGGCCTGTG
		Reverse	CTTTTGAGGACAGCTTTTGGGGT

Target	Organism	Orientation	Sequence
ZFP36	*H. sapiens*	Forward	TTCATCCACAACCCTAGCGAAGA
		Reverse	AGAAGCTGATGCTCTGGCGAAG
IER3	*H. sapiens*	Forward	ACCCTCTTCAGCCATCAGGATCT
		Reverse	AACCCAGCCAAAAGGCTTCTCTT
MYC	*H. sapiens*	Forward	GCTGCTTAGACGCTGGATTT
		Reverse	TAACGTTGAGGGGCATCG
GAPDH	*H. sapiens*	Forward	GAAGATGGTGATGGGATTTC
		Reverse	GAAGGTGAAGGTCGGAGTC
Per1	*M. musculus*	Forward	TCTCACAGTTCATCTTCTGGC
		Reverse	CTGTGAGTTTGTACTCTTGCTG
Tsc22d3	*M. musculus*	Forward	CTGTTGGCCTCGACTGCTG
		Reverse	GCCGAAAGTTGCTCACGAAG
Zfp36	*M. musculus*	Forward	GGTACCCCAGGCTGGCTTT
		Reverse	ACCTGTAACCCCAGAACTTGGA
Tnc	*M. musculus*	Forward	CCAGGGTTGCCACCTATTT
		Reverse	GTCTAGAGGATCCCACTCTACTT
Myc	*M. musculus*	Forward	GCGTAGTTGTGCTGGTGAGT
		Reverse	GTACCTCGTCCGATTCCACG
Gapdh	*M. musculus*	Forward	AGGTCGGTGTGAACGGATTTG
		Reverse	GGGGTCGTTGATGGCAACA

Table C.1: List of qPCR primers in the 5' to 3' direction.

Name	Sequence
EED046	DY547-GGAACCGCCCGGA(s^4U)AGUGUCCUUGGGAAA CCAAGUCCGGGCACCA
EED047	DY647-GGAACCGCCCGGAUAGUGUCCUUGGGAAA CCAAGUCCGGGCACCA

Table C.2: List of synthetic RNAs in the 5' to 3' direction.