Abstract

TimeLapse-seq: Examining the Dynamics of the Transcriptome through 4-Thiouridine Nucleoside Recoding

Jeremy Schofield
2019

The transcriptome continuously changes through the processes of RNA synthesis and decay, but standard RNA-sequencing techniques provide only a snapshot of RNAs at a given time point. Metabolic labeling with 4-thiouridine (s^4U) with biochemical enrichment is a powerful approach to identify new populations of RNA through sequencing. The need for careful normalization and large amounts of RNA input, however, limit the utility of biochemical enrichment. Here I describe the development of TimeLapse-seq, a method to chemically recode s^4U to analogs of cytosine (C*) to identify newly transcribed RNAs through sequencing. I developed RNA-friendly oxidative nucleophilic aromatic substitution chemistry to efficiently convert s^4U to C*. TimeLapse-seq reveals RNA dynamics transcriptome-wide using orders of magnitude less input material than biochemical enrichment. I further demonstrate the use of TimeLapse-seq to reveal acute changes in RNA populations due to cellular stress, differential RNA isoform stability, and as a specificity filter for transient RNA analysis.

The development of TimeLapse-seq opened a variety of collaborative projects at Yale University to study the dynamics of RNAs. We apply TimeLapse-seq to study the stability of mRNAs with DCP2 KO to identify the substrates of the decapping complex. We adapt small RNA sequencing methods to interrogate miRNA turnover in the presence of the HSUR1 viral ncRNA. We perform TimeLapse-seq on RNAs from X-linked dystonia parkinsonism patient-derived primary cells to study the dynamics of RNAs in

the TAF1 disease locus. We also demonstrate through TimeLapse-seq applications bias general to the field s^4U-based approaches and develop best practices to reduce this bias.

Finally, I report preliminary data investigating the acute transcriptional responses to chromatin perturbation by histone deacetylase (HDAC) inhibition. I found that treatment of cells with HDAC inhibitor trichostatin-A (TSA) results in widespread changes in the transcriptome, including changes in the transcription start site (TSS) of hundreds of transcripts. I profile the dynamics of changing TSSs usage, revealing changes in chromatin within minutes of TSA treatment.

The dynamics of the transcriptome continues to be an active study, and the development of methods to profile new biological systems will open new avenues for discovery. TimeLapse-seq provides a flexible platform to study the temporal aspects of RNA sequencing over a wide range of timescales.

TimeLapse-seq: Examining the Dynamics of the Transcriptome through 4-Thiouridine Nucleoside Recoding

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

by
Jeremy Schofield

Dissertation director: Dr. Matthew Simon
May 2019

Table of Contents

List of figures

List of tables

Acknowledgements

I would like to acknowledge the many collaborators who I have worked with throughout my PhD, including: Vicky Luo, Dr. Zhenkun Na, and Dr. Sarah Slavoff; Dr. Paulina Pawlica and Dr. Joan Steitz; Dr. Anna Szekely and Dr. Sherman Weissman; Dr. Daniele Canzio and Dr. Tom Maniatis; Alice Lu and Dr. Akiko Iwasaki; Sarah Dugger and Dr. David Goldstein; Dr. Anthony Rampello and Dr. Christian Schlieker. These collaborations have opened my eyes to so many amazing biological questions and helped to push the development of TimeLapse-seq into what it is today.

I would like to thank Dr. Karla Neugebauer, Dr. Mark Hochstrasser, Dr. Michael Koelle, and Dr. Julie Park, the professors and professionals I worked with while teaching at Yale University for pushing my development at an instructor.

To my thesis advisor, Dr. Matthew Simon, for providing a great learning environment and training me to be a careful and creative scientist. I appreciate all of the time you took to help me develop ideas and for personally teaching me so much of what I know today. From day one in my rotation when you taught me how to perform solid-phase peptide synthesis, I felt that you invested in me and always pushed me to the best of my ability.

My thesis committee members Dr. Karla Neugebauer and Dr. Sarah Slavoff for not only their critical feedback of my evolving research topics but also valuable advice on career opportunities and life after grad school.

To the members of the Simon lab for the fun they brought to lab every day, and for the friendships we formed outside of lab. Trivia nights, game nights, fun shirt Fridays, laser tag, and the New Haven Road Race are among the many memories I will cherish.

To the BBSB entering class of 2014 and all of the friends in New Haven I made along the way, you made the long experience of the PhD manageable. From traveling to weddings, to concerts in New York, to adventures in Europe, it was a blast.

To Alyssa for your love and the joy you bring to my life every day. I appreciate your constant support and patience, and I truly cannot wait to start a new chapter and have new adventures with you in Seattle.

To my parents and grandmother for their endless love and support and the opportunities they have given me to get to this point, I couldn't have done all of this without you.

Chapter 1

Introduction

1.1 Cellular RNA turnover

The balance of RNAs present in a given cell provides information about the state of the cell, and RNA sequencing is a powerful tool to determine the identity and abundance of RNAs in a sample. RNA sequencing provides a static view of the transcriptome at the time of sample collection, but the transcriptome is in continuous flux to maintain proper cellular homeostasis. Steady state levels of RNAs are maintained through processes of synthesis and decay, and defects in the core components of these processes are associated with a number of pathologies [1, 2]. Different classes of RNA exhibit different dynamics of transcription and decay in a manner consistent with RNA function. Enhancer RNAs (eRNAs) are highly unstable, and although the exact function of eRNAs is debated, they are known to play a role in promoting transcription through interaction with the promoter in a temporally coordinated manner [3]. Ribosomal RNAs are highly stable, as they are required for the constant output of protein to maintain cellular homeostasis[4]. mRNAs exhibit moderate stability, but vary in half-life depending on their coding function [5]: genes coding for tightly regulated transcription factors produce relatively unstable mRNAs in mammals, whereas genes coding for components of biosynthetic pathways produce relatively stable mRNAs that often persist for longer than one cell cycle [6].

The processes of transcription and decay contribute to the abundance of RNAs at steady state, and changing RNA populations due to cellular perturbations can be achieved through modulation of either process. For example, in the cellular response to heat shock stress, transcription is globally shut down with the exception of a subset of rapidly induced genes specifically transcribed as a response to the stress [7]. Cellular contexts involving changes in RNA decay include the clearance of maternal transcripts during the maternal to zygotic transition [8] and the stimulation of nonsense mediated decay (NMD) pathway in response to mRNAs containing premature termination codons [9]. There is much active research to understand the interplay between transcription, decay, and chromatin environment, as well as the development of tools to study each of these processes.

1.2 Dynamics of transcription

RNA polymerase II (RNAPII) transcription occurs through several stages including initiation, elongation, and termination. Each stage of transcription requires the orchestration of the core transcriptional machinery as well as many regulatory factors. Transcription of mRNAs begins with initiation, which brings together RNAPII and the general transcription factors (GTFs) [10, 11]. The TATA-binding protein (TBP) is required for binding to the promoter, and is associated with a number of TATA-associated factors (TAFs) to form a scaffold referred to as TFIID for the pre-initiation complex (PIC). Other GTFs assemble with the PIC including TFIIH, which contains a kinase domain able to phosphorylate the C-terminal domain of RNAPII [12], and an ATPase domain that functions to open DNA and allow for templating of the growing

RNA polymer [13]. A large body of work has described these biochemical activities necessary for the process of initiation, but variability in gene context and the regulation of initiation is still an active area of research. For example, biochemical reconstitution of the transcriptional machinery is typically performed using a TATA promoter, but only about 20% of all promoters in eukaryotes are TATA-containing [14]. In addition, eukaryotes can utilize the SAGA complex in place of TFIID for initiation. SAGA contains many of the same protein components as TFIID, and it has been demonstrated in yeast that genes that predominantly use SAGA over TFIID tend to be stress responsive and tightly regulated [15]. More recent work has demonstrated a more general role for SAGA at all genes, including those that were previously shown to be TFIID-dominated [16]. Once the components of the PIC are assembled onto DNA, RNAPII transcribes the first 20-60 nt of nascent RNA, and for a large subset of genes, RNAPII pauses after this initial transcription [17].The paused state is maintained in part through the association the factors DSIF and NELF. At this stage, the paused RNA can be "released" into elongation through phosphorylation of NELF by the kinase PTEF-b [18], which itself can be recruited by factors including the bromodomain protein BRD4 [19]. The role of pausing is an active topic of research, but it is thought that pausing could be a quality control checkpoint to ensure proper assembly of the transcriptional and co-transcriptional components [20]. Several processes occur co-transcriptionally, and components of these processes are found to interact with RNAPII during transcription, including 5'-end capping components [21, 22] and splicing components [23]. Pausing may also be a strategy for cells to produce a poised state for a rapid transcriptional response to the changing cellular environment. For example, genes responsive to heat shock stress were

found to exhibit a high degree of pausing [24], which can be released through binding of specific transcription factors [7]. Pausing also occurs and appears to be regulated in a similar fashion at enhancers, but the proportion of enhancer RNAs that proceed into productive elongation is much lower than that of mRNAs [25].

Once nascent RNAs are released from pause, RNAPII proceeds through the gene body to generate pre-mRNA in a process called elongation. The use of the reversible small molecule inhibitor DRB has been used to block entry into elongation [26], and removal of DRB produces a wave of transcription that can provide rates of elongation using metabolic labeling [27]. Elongation rates have been estimated in cell culture and for a selection of genes in vivo, with median elongation rates of 3.7 and 2.7 Kb/min, respectively [27, 28]. Mutants of RNAPII that change the rate of elongation have been demonstrated to alter the frequency of splicing [29, 30] and the preferred site of polyadenylation [31, 32], suggesting that elongation rate is finely tuned to integrate cotranscriptional processes. The quantity of mature RNA produced by any given gene correlates with the rate of elongation as well, further highlighting the ability of cells to regulate gene expression through multiple layers of the RNA life cycle [33].

The rate of elongation also plays a vital role in the later stages of transcription in the stage of termination. After mRNA cleavage and polyadenylation, RNAPII continues to travel downstream through the gene body and must be removed from the DNA template. 5’-3’-exonuclease activity was proposed to interfere with the elongating RNAPII resulting in termination of transcription [34], and it was later discovered that the Xrn2 exonuclease is required for precise termination [35, 36]. Additionally, it was found that disruption of Xrn2 function or the rate of RNAPII elongation shifts the location of

termination [37]. These observations lend evidence to the proposed “torpedo” model of kinetic competition whereby Xrn2 progresses through nascent RNA to knock RNAPII off of chromatin. Further support for this model comes from the observation that elongation slows at the 3’ end of genes, which decreases the amount of time needed for Xrn2 to catch the elongating RNAPII [38].

1.3 RNA decay

Cellular control of the generation of RNA messages through transcriptional regulation is one strategy to maintain homeostasis, but cells also modulate gene expression post-transcriptionally. The translational output of mRNAs varies, and ribosome profiling has demonstrated that the translational efficiency of specific genes can change under stress conditions [39]. The miRNA associated RISC complex can interfere with components of translational initiation, allowing for temporal regulation of gene expression through translational repression [40]. RISC and translationally repressed RNAs can then be localized to P-bodies in the cytoplasm, where they are sequestered to reduce the expression of the gene [41]. The RISC complex is also able to coordinate the degradation of mRNAs as a means of reducing the concentration of coding transcripts in the cell [40].

miRNA-independent RNA decay requires a separate repertoire of biochemical factors that coordinate to determine the half-life of an RNA, and to ultimately degrade the RNA. Several features of the RNA itself dictate its rate of decay, including: secondary structure [42, 43]; protein binding and primary sequence in untranslated regions [44, 45]; m^6A modifications [46]; defects recognized by RNA surveillance pathways [9, 47]; and

codon optimality [48]. Depending on the cellular context, different decay pathways may be engaged to degrade an RNA [2]. These decay pathways utilize a range of molecular complexes with different biochemical activities, including deadenylation, 5'-decapping, exonuclease, and endonuclease action. The specificity of regulation by different decay pathways provides the cell with an extra layer of regulation to adapt to changing cellular conditions with subsets of different genes [5].

1.4 Regulation of transcription and chromatin

The specific RNAs transcribed and rates of their synthesis depend on the context of the cell, and can vary by cell type, cell cycle, and in response to acute changes in environment. It was observed in the early 20th century that different regions of chromosomal DNA display different staining densities, indicative of the open and repressed regions of the genome referred to as euchromatin and heterochromatin [49]. More recently, ChIP-seq analysis of a number of histone modifications has revealed patterns of modifications associated with active chromatin or repressed chromatin. The promoters of active genes tend to be enriched for H3K4me3, H3K9ac, and H3K14ac modifications, while inactive promoters tend to be depleted in acetylation and enriched for H3K9me3 and H3K27me3 [50]. H4ac is also enriched at the promoters of active genes, though its abundance shows a weaker correlation to transcriptional activity than other acetylated histones [51]. Histone modification is a dynamic process, and histone modifications display a range of timescales over which they change. Gene bodies are typically enriched for H3K36me3, a modification that is deposited co-transcriptionally to promote RNAPII elongation and to reduce cryptic initiation downstream of the promoter

region [52, 53, 54]. Metabolic labeling has demonstrated that nucleosomes at the promoter undergo turnover in about an hour [55]. In addition, quantitative mass spectrometry shows the half-lives of histones bearing acetyl marks is relatively short, typically 1-2 hours [56], whereas methylated histones tend to be more stable, with half-lives in the range of a few hours to several days [57]. Changes in histone acetylation are therefore relevant on the timescale of cellular perturbation, while changes in a broad range of epigenetic modifications are relevant on timescale of development. Enhancers and their associated chromatin environment provide a good example for the dynamic regulation of gene expression. Enhancers are regions of DNA promoting the activity of an associated promoter region, and are often required for efficient gene expression for associated genes in vivo [58]. Active enhancers are marked with H3K4me1 and H3K27ac, whereas poised enhancers contain H3K4me1 but are deacetylated [59]. Furthermore, cells undergoing changes in cell state display profound changes in enhancer activity through the regulation of histone modifications [60, 61]. Changes in chromatin signatures over shorter time scales therefore provide the ability for enhancers to fine tune the expression of genes through changing cellular contexts [62].

Histone post-translational modifications function in part to help recruit factors to physically decompact chromatin, allowing RNAPII access to genomic DNA. Recent work has allowed for finer resolution in defining active regions of chromatin through differential chromatin accessibility, using methods including MNase-seq and ATAC-seq [63, 64]. These active chromatin regions, referred to as nucleosome depleted regions (NDRs), tend to correlate with increased transcriptional initiation [65]. Nucleosomes present a boundary to efficient transcription, and nucleosome remodeler enzymes act to

loosen chromosomal DNA-histone contacts allowing for access of DNA to RNAPII [66]. The exact positioning of nucleosomes at active promoters, especially the +1 nucleosome directly following the TSS, tunes the rate of RNA polymerase pausing [67], a process known to be important in regulating the output and process of transcription. It has been demonstrated that changes in nucleosome positioning associated with H4 acetylation and Snf2 chromatin remodeling is a vital process in the establishment of cellular quiescence after glucose starvation in yeast [68]. Furthermore, changes in nucleosome occupancy over regulatory regions has been shown to occur through the process of differentiation in iPSCs, underscoring the importance of regulating the nucleosomal landscape to maintain specific patterns of gene expression [69].

1.5 Methods to study RNA synthesis and decay

A variety of methods have been used to measure the transcriptional output of cells. Transcription can be blocked through inhibition of polymerase with drugs including alpha-amanitin and actinomycin D [70, 71], and the decay of transcripts can be measured through qPCR or sequencing. Inhibiting general transcription, however, has been shown to perturb cellular physiology [72], prompting the development of strategies to purify nascent transcripts from total RNA. These purification methods to study transcription rely on biochemical enrichment to specifically isolate nascent transcripts or the transcription machinery. NET-seq utilizes an antibody against RNA polymerase to determine the position of the transcriptional machinery at the time of sample preparation, and can be used to enrich for different CTD variants to study different stages of transcription [73, 74]. Nascent transcripts can be metabolically labeled with orthogonal nucleosides and

specifically enriched using antibodies or orthogonal chemistry. GRO-seq relies on metabolic labeling of nascent RNAs with 5-bromouridine 5'-triphosphate followed by enrichment of 5-bromouridine-containing RNA with an antibody [75]. 5-ethynyl uridine can also be incorporated into newly transcribed RNAs and enriched using click chemistry [76]. 4-thiouridine is commonly used to metabolically label nascent RNAs and enriched using a variety of methodologies, including organomercury affinity chromatography and disulfide capture using HPDP-biotin and more recently MTS-biotin [77, 78, 79, 80]. Biochemical enrichment methods have provided insights into promoter-proximal pausing [75], transcriptional responses in dendritic cells from LPS stimulation [81], and transcriptome-wide RNA half-lives [79].

Although biochemical enrichment has been the standard for studying the dynamics of RNA, some biological systems remain underexplored due to a number of limitations inherent to the methods. Biochemical enrichment requires tens to hundreds of micrograms of RNA, limiting the approach to large scale cell culture or the equivalent number of cells from model organisms. The need to study small cell populations is exemplified by research on cellular heterogeneity, which contributes to the difficulty in combating drug resistance in cancerous tumors [82]. Other processes, including cellular differentiation, are also marked by a variety of cell types undergoing different transcriptional programs. In addition, many biological processes are not recapitulated in stable cell culture and require the use of primary cells, limiting the amount of material available for study. Recent advances in transcriptome sequencing technology have allowed for the study of rare cell populations and single cells at the level of total RNA

[83, 84], but studying RNA dynamics using these platforms requires inference based on observed intronic reads rather than direct measurement of nascent transcripts [85].

The analysis of biochemically enriched RNAs is similar to standard analysis of RNA-seq data, though extra considerations must be followed to properly interpret the data. I use s^4U-based disulfide enrichment is an example, but these considerations are general to all biochemical enrichment strategies. First, enriched RNA libraries need to be carefully normalized to an exogenous spike-in as total RNA levels are not measurable within the same sample. Second, all biochemical enrichment strategies are associated with some level of contamination from non-specific binding interactions, which needs to be accurately estimated and corrected for. It has been estimated that in s^4U crosslinking experiments, this background is as high as 30% [86], and the background in transient transcriptome sequencing (TT-seq [87]) with optimized MTS-biotin chemistry is estimated between 15-20% [88]. Third, inefficient enrichment can lead to biases in the enriched product; for example, HPDP-biotin is inefficient at disulfide formation with s^4U, resulting in a higher likelihood of enriching transcripts with higher relative s^4U content and leading to a length bias in sequencing [89, 80]. For all reasons outlined above, there is a need to develop tools to study RNA dynamics using small quantities of total RNA input.

1.6 Chemical biology of nucleoside recoding

Nucleoside recoding and mutation mapping have been used to study the properties of nucleic acids in a number of biological contexts. These approaches use either sequencing or gel electrophoresis as a readout of reverse transcriptase interacting

with a non-standard nucleoside to produce a mutation or a pre-mature stop. Stops and mutations are caused by either the propensity for endogenously modified nucleosides to affect the RT, or through chemical modification of bases to induce RT arrest or misincorporation [90]. For example, m^1A is found endogenously in several classes of cellular RNAs and is known to play a role in tRNA structure [91, 92], and m^1A can be generated in single stranded regions of RNA through DMS chemical modification to determine RNA secondary structure [93]. Cytosine methylation is an epigenetic modification that plays an important role in development and gene regulation [94]. Bisulfite sequencing recodes cytosine to uracil via deamination, but is unable to deaminate methylated cytosine, revealing the locations of 5-mC by resistance to conversion [95]. Adenine deamination is catalyzed by ADAR enzymes to produce inosine in RNAs, often resulting in changes in protein translation and splicing [96, 97]. Reverse transcription of inosine leads to the incorporation of cytosine, resulting in an A-to-G mutation that can be identified through transcriptome sequencing [98]. tRNA molecules contain a variety of RNA modifications vital to their structure and function, including s^4U found in bacterial tRNA. Early work from Ziff and Fresco mapped the location of s^4U in tRNA through oxidative nucleophilic aromatic substitution to create a radioactive cytidine derivative [99]. These studies demonstrate the ability for nature to utilize chemical diversity to regulate a wide number of biological systems.

Chemical biologists have also expanded the chemical diversity of nucleic acids through synthetic means. Convertible nucleosides containing functional leaving groups can be incorporated site-specifically into oligonucleotides through solid phase nucleic acid synthesis [100]. These leaving groups on convertible nucleosides are displaced by

nucleophiles, generating substituted nucleoside derivatives. The convertible nucleoside approach has been used to physically tether nucleic acids to study their physical properties and to probe the role of specific contacts between DNA and proteins [101].

1.7 Overview

Current work in the field of molecular biology focuses on the mechanistic details of a variety of dynamics processes in RNA and chromatin. The study of molecular biology is complemented by chemical biology methods able to probe features of macromolecules with specificity. I describe in this thesis the development of TimeLapse-seq, an approach to identify newly transcribed RNAs through s^4U metabolic labeling and nucleoside recoding [88]. TimeLapse-seq has allowed us to profile RNAs on a number of timescales, using orders of magnitude less material than is typically used in biochemical enrichment methods. I will describe the use of TimeLapse-seq to examine steady-state RNA turnover, acute changes in RNA transcription, differences in mRNA and miRNA stability, and transient transcription. Next, I will report work to reduce bias associated with sample preparation in s^4U-based methods. Finally, I will show preliminary work using TimeLapse-seq and chromatin profiling to examine the transcriptional consequences of histone deacetylase inhibition.

Chapter 2

TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding

This chapter contains excerpts from:

Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C., Simon, M.D. (2018) TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat. Meth.* 15:221–225. doi: https://doi.org/10.1038/nmeth.4582

2.1 Author Contributions

I performed all experiments with assistance on TT-TimeLapse-seq experiments from Erin Duffy, and with assistance on NMR characterization from Lea Kiefer. Matthew Simon performed bioinformatic analyses with assistance from myself and Meaghan Sullivan.

2.2 Summary

I describe a method (TimeLapse-seq) to chemically recode 4-thiouridine (s^4U) metabolic labels to cytosine analogs in newly transcribed RNA. Recoded metabolic labels are identified through whole transcriptome sequencing, and bioinformatic analyses model the fraction of new RNA on a gene-by-gene basis. TimeLapse-seq allows for the study of

RNA dynamics without the need for biochemical enrichment, greatly reducing the amount of RNA input necessary for analysis. TimeLapse-seq allows for the estimation of transcript half-lives, acute changes in transcription from cellular stresses, and can be used together with biochemical enrichment (TT-seq) to identify high confidence new RNAs.

2.3 Introduction

Global changes in transcription can occur on the timescale of minutes to hours in many mammalian systems, including circadian rhythms and the immune response. One can identify such changes by monitoring new transcripts that cofractionate with chromatin [102, 103] or that have not yet been spliced [104, 105]. New RNA populations can also be identified by examining sites of active RNA polymerase II through biochemical enrichment of transcripts that are being synthesized (e.g., PRO-seq [106] and NET-seq [73]) or metabolic labeling and enrichment of new transcripts (e.g., TT-seq [87] and s^4U-seq [81, 80]). These techniques require large amounts of input sample and extensive handling, and they present challenges when normalizing enrichment and estimating contamination. To capture temporal information about RNA directly in a sequencing experiment without biochemical enrichment, we developed TimeLapse-seq (Fig. 1a), a method in which cells are exposed to a noncanonical nucleoside that becomes incorporated into only new transcripts. Rather than enriching the metabolically labeled RNAs, we developed chemistry that recodes the hydrogen-bonding pattern of the uridine analog 4-thiouridine (s^4U) to match the hydrogen-bonding pattern of cytosine, thereby causing mutations in a sequencing experiment.

2.4 Design

The main considerations of chemically recoding s^4U are the efficiency of chemical conversion and the integrity of the RNA. As the incorporation rate of s^4U into new RNA has been demonstrated to be less than 10% [107, 108], highly efficient recoding is necessary to observe enough mutations in new RNAs above background. Based on literature precedent, I focused on developing a strategy of oxidative nucleophilic aromatic substitution to convert thiolated nucleosides to their amine-bearing derivatives [99, 109]. The chosen oxidant should efficiently form an oxidized intermediate with a convertible leaving group, but should not be so strong as to oxidize non-s^4U bases. The chosen amine should be nucleophilic to displace the convertible leaving group, but should have a low pKa as to be nucleophilically active at low pH to prevent RNA hydrolysis. I screened oxidant and amine combinations on a nucleoside substrate using reverse phase liquid chromatography and mass spectrometry. I then developed a restriction enzyme assay to optimize the nucleotide conversion conditions, referred to as TimeLapse conditions, in an oligonucleotide substrate. I next performed nucleoside recoding on cellular total RNA from s^4U metabolically labeled cells and analyzed transcriptome wide RNA dynamics (TimeLapse-seq) at steady state and in response to a heat shock stress. I performed nucleoside recoding after MTS-biotin enrichment of cellular RNAs treated for 5 minutes with s^4U in collaboration with Erin Duffy, and used the induced mutations as a specificity filter for bona fide new RNAs (TT-TimeLapse-seq). Finally, the development of thiol-specific nucleoside recoding

allows for the use of a second orthogonal metabolic label, which Lea Kiefer demonstrated by developing s^6G TimeLapse-seq.

2.5 Results

2.5.1 Assessment of oxidative nucleophilic aromatic substitution on the s^4U nucleoside

I first explored chemistry to convert the free nucleoside (s^4U) to cytidine derivatives (Fig 2.1) while minimizing oxidation of guanosine and using amines with low pKa values that remain deprotonated under neutral reaction conditions. I found using reverse phase liquid chromatography and mass spectrometry that treating s^4U with a number of oxidants and amines led to conversion to cytidine derivatives. In particular, treating s^4U with 2,2,2-trifluoroethylamine (TFEA) and meta-chloroperoxybenzoic acid (mCPBA) results in near-complete consumption of s^4U, producing only small amounts of

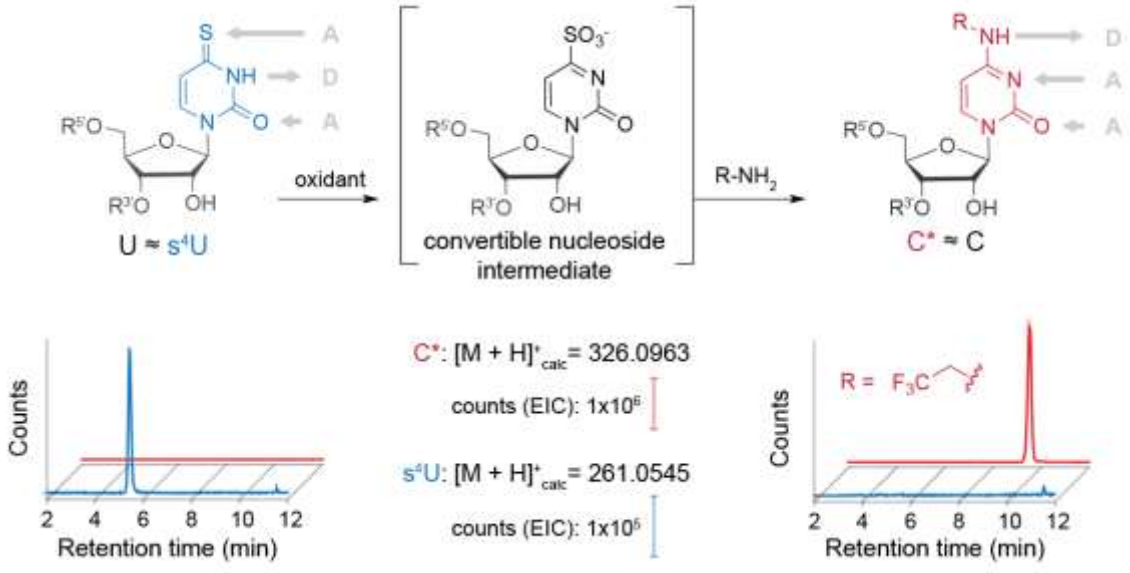

Fig 2.1. LC-MS analysis of oxidative nucleophilic aromatic substitution reaction to recode s^4U to C*. Extracted ion chromatograms for s^4U starting material (blue) and trifluoroethylated cytosine (red) demonstrate complete consumption of s^4U and appearance of product of expected mass

the hydrolysis product uridine, and mostly the desired trifluoroethylated cytidine (C*, Fig 2.1).

2.5.2 Optimization of nucleoside recoding chemistry on s^{4}U-containing RNA

To optimize the efficiency of s^{4}U nucleoside recoding in the context of an RNA macromolecule, I developed a restriction enzyme digestion assay. First, I in-vitro transcribed an RNA oligonucleotide containing a single s^{4}U incorporation site located within a NotI restriction enzyme cut site (Fig 2.2a). Only upon successful conversion of s^{4}U to a C analog, NotI will be able to digest the product resulting from reverse transcription and PCR amplification of the RNA substrate. Using this assay, I screened oxidant, amine, reagent concentration, time, and temperature to increase the efficiency of chemical recoding. I found that 10 mM sodium periodate ($NaIO_4$) and 600 mM TFEA at 45^{o}C for 1 h results in efficient (~80%) conversion (Fig 2.2b).

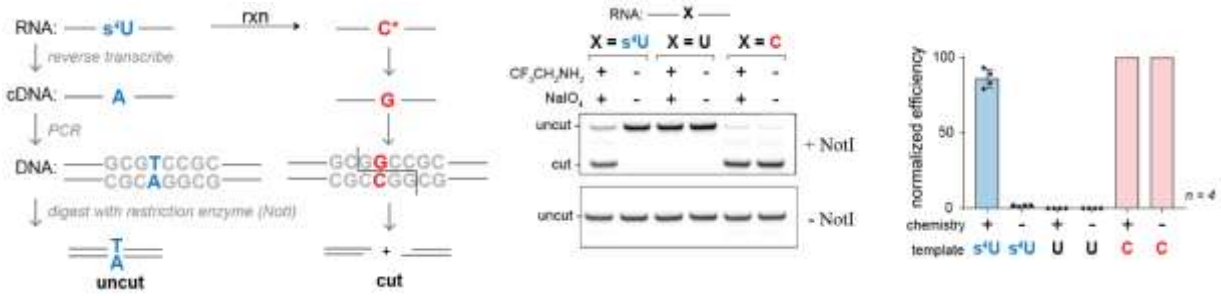

Figure 2.2. a) Scheme of restriction enzyme assay to assess conversion of s^{4}U to C* in RNA. b) Gel pictured demonstrates efficient digestion (~80%) of s^{4}U to C* in s^{4}U and chemistry-specific manner. X = U RNA target is not digested, independent of chemistry, and X = C RNA target is completely digested, independent of chemistry. Densitometry quantification is shown in barplot.

2.5.3 Targeted TimeLapse-seq of cellular RNAs

To test optimized TimeLapse chemistry conditions ($NaIO_4$ and TFEA) with cellular s^4U-RNA, I exposed mouse and human cells to a range of concentrations of s^4U. After RNA isolation and chemical treatment, I examined the apparent U-to-C conversion rates (inferred from T-to-C mutations in the cDNA, hereafter referred to as T-to-C) by targeted RT-PCR coupled to paired-end sequencing (see Table 1). I observed a notable and specific increase in T-to-C transitions in chemically treated samples (Fig 2.3), demonstrating that TimeLapse recoding is able to identify newly transcribed RNAs on a sequencing platform.

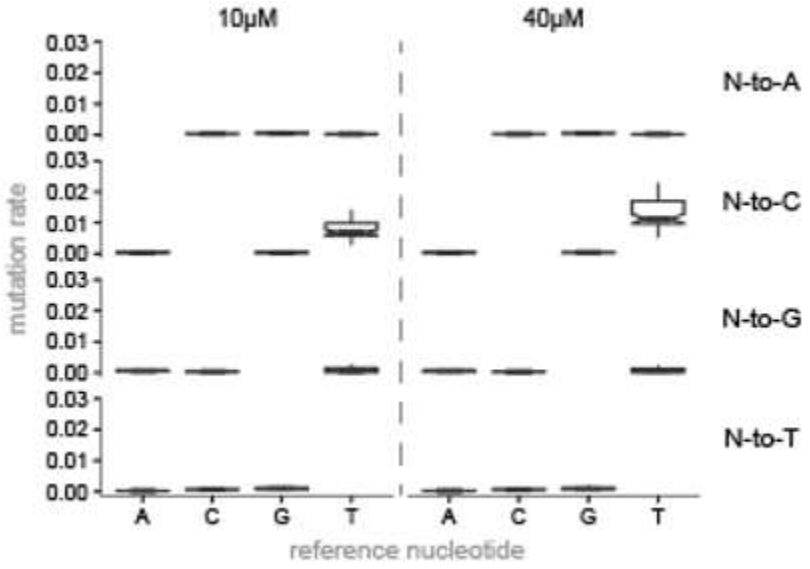

Figure 2.3: Mutation rates for each N-to-N combination for targeted TimeLapse-seq of Actb and Gapdh (700 uM s^4U for 2 h). T-to-C mutations are increased in a s^4U concentration-specific manner, while other mutation rates remain relatively low.

2.5.4 Transcriptome-wide TimeLapse-seq monitors steady state RNA dynamics

To examine the dynamics of cellular RNAs, I treated MEF cells with s^4U for 1 h and performed TimeLapse chemistry before sequencing. The total transcript counts from

each sample were highly correlated irrespective of s^4U exposure or chemical treatment (Pearson's $r \geq 0.97$), demonstrating that TimeLapse-seq retains information from a traditional RNA-seq experiment. By counting the mutations in each aligned read pair, we found a specific and reproducible increase in T-to-C mutations dependent on both metabolic labeling with s^4U and chemical treatment (Fig 2.4). Other mutation rates remained below background levels of T-to-C mutations in untreated samples (e.g., the small increase in G-to-T mutations). Additionally, the reaction was efficient even in regions of RNA secondary structure. The T-to-C mutation counts were dramatically higher in fast-turnover transcripts (e.g., Myc and Fosl2), compared to more stable transcripts (e.g., Dhx9 and Ybx1) (Fig. 2.5a). We observed an enrichment of T-to-C mutations in intronic reads (Fig. 2.5c) consistent with the fast turnover of intronic RNA.

To quantify these results, we modeled reads as arising from two populations: pre-existing RNAs (background mutation rate) and new RNAs (high T-to-C mutation rate; Fig. 2.5b; see methods). Reads from newly synthesized RNAs had an average of 2.2 mutations per read, corresponding to an ~3% mutation rate per uridine (compared to ~0.1% T-to-C mutation rates in controls and for pre-existing RNAs). From each gene, we determined the fraction of newly made transcripts ($r \geq 0.94$; 2,992 genes) and estimated transcript half-lives, which correlated with those reported previously [6]. As expected, the fast-turnover RNAs (top 10%, n = 360) were enriched for transcripts such as transcription

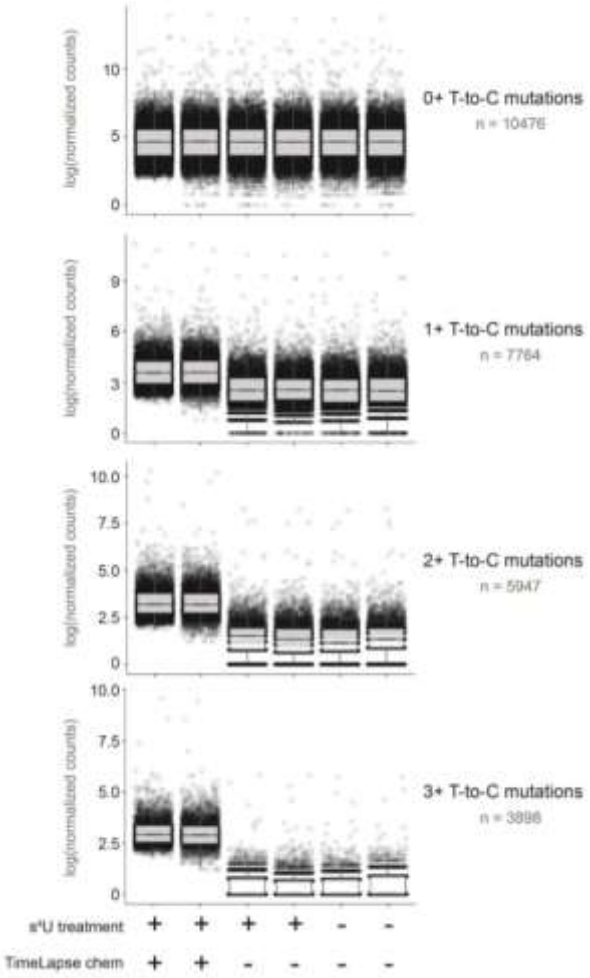

Figure 2.4: s^4U treatment and TimeLapse chemical conversion result in an increase in T-to-C mutations not apparent with s^4U alone. Read counts were normalized to the total RNA-seq analysis (top) and log transformed after adding a pseudocount of one to each transcript. Data for individual transcripts are overlaid with box and whiskers plots using default parameters from ggplot2. Transcripts were only included if at least two samples had more than 20 counts. The number of transcripts in each analysis are indicated next to each graph.

factors (Results from PANTHER analysis: DNA-templated transcription, $P < 10^{-20}$), while the slow-turnover RNAs (top 10%, n = 361) were enriched for those that are involved in translation (Results from PANTHER analysis: ribosomal biogenesis, $P < 10^{-6}$; translation, $P < 10^{-27}$).

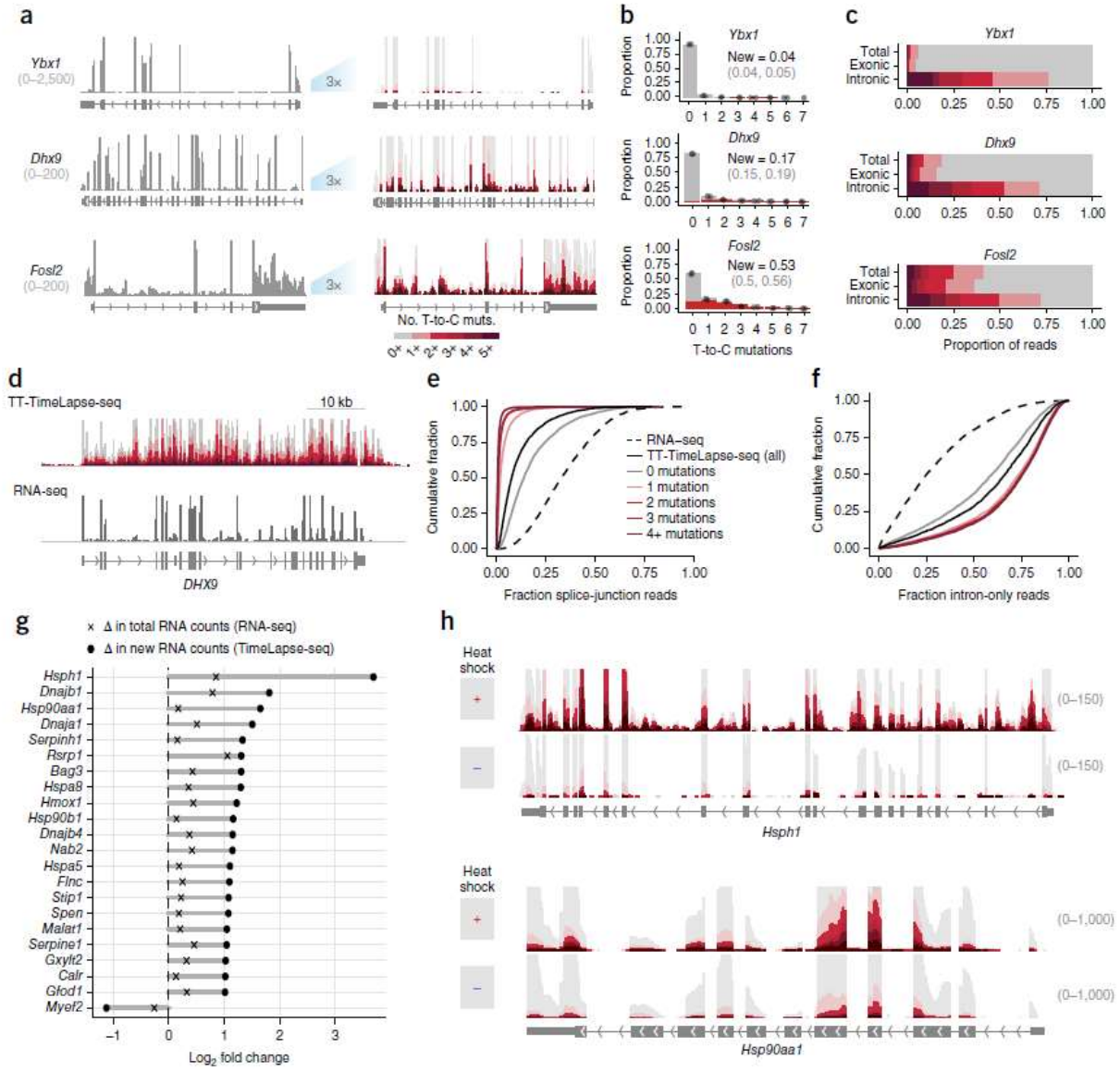

Figure 2.5. Global analysis of steady-state and transient RNA dynamics using TimeLapse-seq. (a) Left, tracks depicting coverage from all reads (gray) for transcripts with slow (Ybx1), moderate (Dhx9), or fast (Fosl2) rates of turnover. Right, tracks from reads with increasing numbers of T-to-C mutations (see scale) displaying mutational content provided by TimeLapse chemistry (right, y-axis zoom 3×). (b) Distribution of reads with each number of T-to-C mutations (points) overlaid on a model of the estimated distribution of reads from new transcripts (red) and pre-existing transcripts (gray) for Ybx1, Dhx9, and Fosl2. The estimated fraction of new reads is indicated for each plot. Light gray, 95% confidence interval. (c) Distribution of T-to-C mutations found in reads mapping to Ybx1, Dhx9, and Fosl2, separated by total, exonic, or intronic reads. (d) TT-TimeLapse-seq and RNA-seq tracks of DHX9. (e) Cumulative distribution plot of reads containing splice-junctions in RNA-seq, and TT-TimeLapse-seq. (f) Cumulative distribution plot of intron only reads in RNA-seq and TT-TimeLapse-seq with the same scale as in e. (g) Using TimeLapse-seq to distinguish new RNAs after heat shock. Log2-fold changes after heat shock in total RNA-seq counts and new RNA counts for the top RNAs identified in b as significantly changed upon heat shock (Padj < 0.01). (h) RNA-seq and TimeLapse-seq tracks of Hsph1 (top) and Hsp90aa1 (bottom) upon heat shock.

2.5.5 TimeLapse-seq identifies acute changes in RNA with heat shock stress

To test whether TimeLapse-seq could reveal induced changes in RNA populations, I subjected MEF cells to a mild heat shock (42°C, 1 h), where only modest changes in total RNA levels were apparent [110, 7, 111]. I observed induction of a few transcripts such as Hspa1b by RNA-seq, but TimeLapse-seq revealed the induction of many transcripts encoding heat shock proteins in the new transcript pool that are not apparent by RNA-seq alone (Fig. 2.5g). For example, whereas RNA-seq is less sensitive to the small absolute changes in Hsph1 and Hsp90aa1 (as they are already abundant before heat shock; RNA-seq fold-change, Hsph1 = 1.8-fold, Hsp90aa1 = 1.1-fold, DEseq2), TimeLapse-seq reveals substantial induction of both transcripts in the new transcript pool (TimeLapse-seq fold change, Hsph1 = 12.7-fold; Hsp90aa1 = 3.1-fold, DEseq2) (Fig. 2.5h). Unlike PRO-seq and NET-seq, however, which are not sensitive to changes in RNA populations after transcription has completed, TimeLapse-seq captures changes in RNA processing—I observed the induction of a new terminal exon in Rsrp1 upon heat shock as w ell as post-transcriptional down-regulation of histone mRNAs upon heat shock, neither of which would be apparent from analysis of nascent RNA (Fig. 2.6).

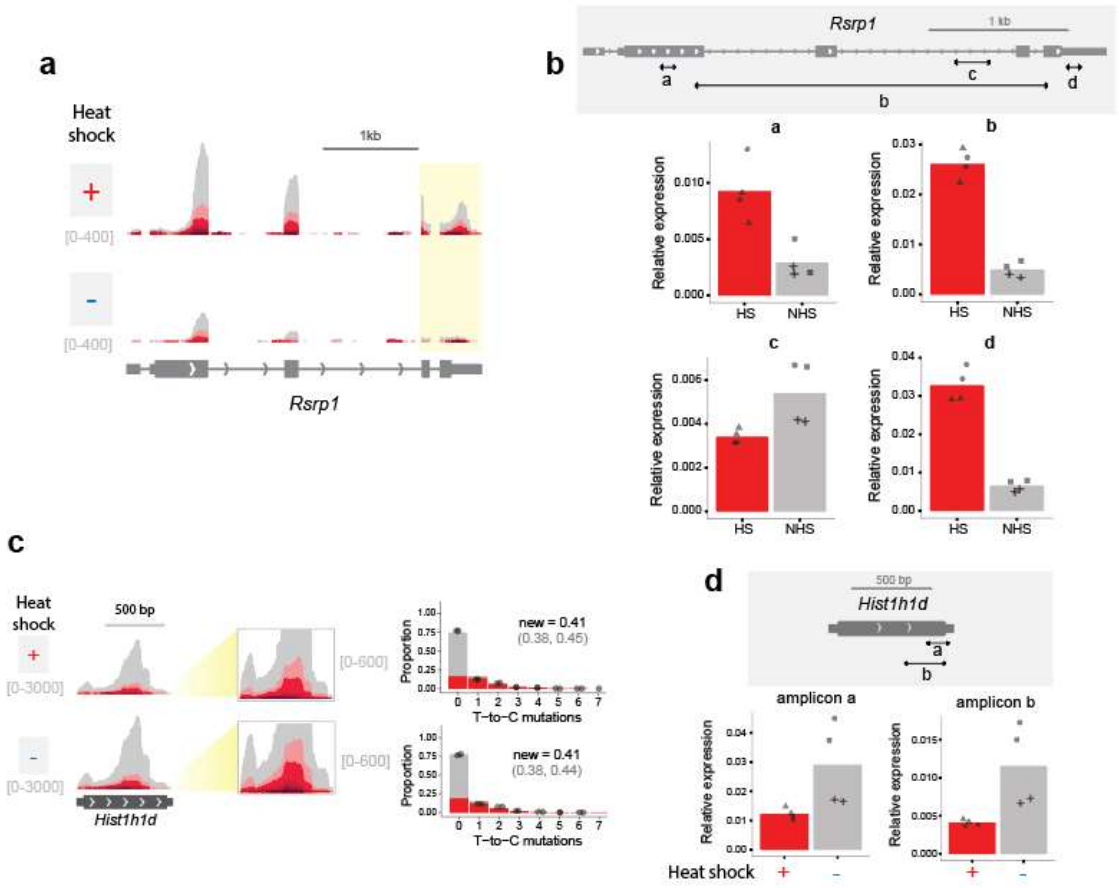

Figure 2.6. Changes in mRNA processing with heat shock stress. (a) The *Rsrp1* transcript appears to be stabilized upon heat shock, with the terminal exon displaying the highest degree of stabilization. (b) qPCR quantification of varying *Rsrp1* transcript features. (c) Browser shot and T-to-C mutation distribution of *Hist1h1d* mRNA. Total RNA levels of *Hist1h1d* mRNAs decrease, though the fraction of new RNAs during the 1h heat shock remain consistent, suggesting a post-transcriptional degradation mechanism observed previously in replicative histone mRNAs during select stages of the cell cycle. (d) qPCR analysis of *Hist1h1d*.

2.5.6 TT-TimeLapse-seq distinguishes transient RNAs from background

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 through transient transcriptome sequencing (TT-seq [87]), biochemically enriched s^4U-

RNA always contains contaminating reads from unlabeled RNAs (estimated to be up to 30% in some experiments [86]). This contaminating background can limit analyses; for example, abundant spliced transcripts observed in RNA enriched after short s^4U pulses have been interpreted as fast splicing [112], 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 TT-seq to distinguish bona fide new RNAs from contaminating background, I treated K562 cells with s^4U for 5 min, and Erin Duffy performed biochemical enrichment as in TT-seq [87], except with more efficient MTS chemistry to biotinylate the s^4U-RNA [80]. After enrichment and before sequencing, I performed TimeLapse chemistry. As expected, transient RNA species were enriched for introns (two-sample Kolmogorov–Smirnov test, $P < 10^{-15}$; Fig. 2.5d–f) but depleted for splice junctions ($P < 10^{-15}$). Both enrichment of introns and depletion of splice junctions were slightly greater than previously observed [87], which was 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 (Fig. 2.5d). Mutation-containing reads represented a subpopulation that was further enriched for introns and depleted for splice junctions (Fig. 2.5e,f). This suggests that mutated reads

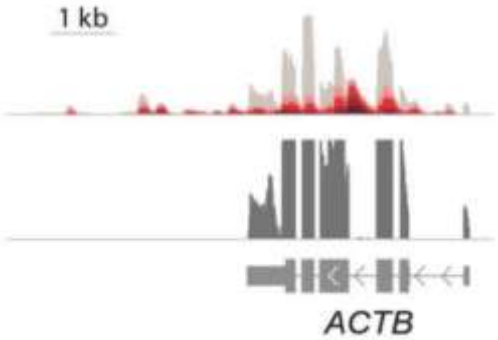

Figure 2.7. TT-TimeLapse-seq distinguishes bona fide reads from background contamination in the ACTB locus. Shown top is a 5 min TT-TimeLapse-seq track, which displays an uneven mutation distribution across the gene body. In particular, there are many more unmutated (grey) reads in exonic regions, consistent with background contamination from abundant processed mRNAs (shown in total RNA-seq track, bottom).

effectively capture the profile of new RNAs, while the reads without mutations represent a subpopulation that is contaminated by unlabeled reads. We estimated that 15–20% of total TT-seq reads arise from contaminating RNA (estimate from splice-junction content, 17–20%; from intronic content, 18–20%; see methods and Table 2), similar to estimates from previous s^4U experiments [86]. 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 (Fig 2.7) 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.

2.5.7 RNA isoform-specific dynamics with TimeLapse-seq

I applied TimeLapse-seq using treatment conditions optimized for studying mRNA turnover (4-h s^4U) [113] in a chronic myelogenous leukemia model cell line (K562). Using a binomial model applied to the observed mutation distributions (see methods 6.19), we obtained highly reproducible half-life estimates that correlated with previous observations [114] (Fig 2.8 c-e). Inspection of individual transcripts revealed reads mapping to both a shorter isoform of ASXL1 (NM_001164603), as well as a longer isoform (NM_015338) of ASXL1. The ASXL1 protein is involved in epigenetic

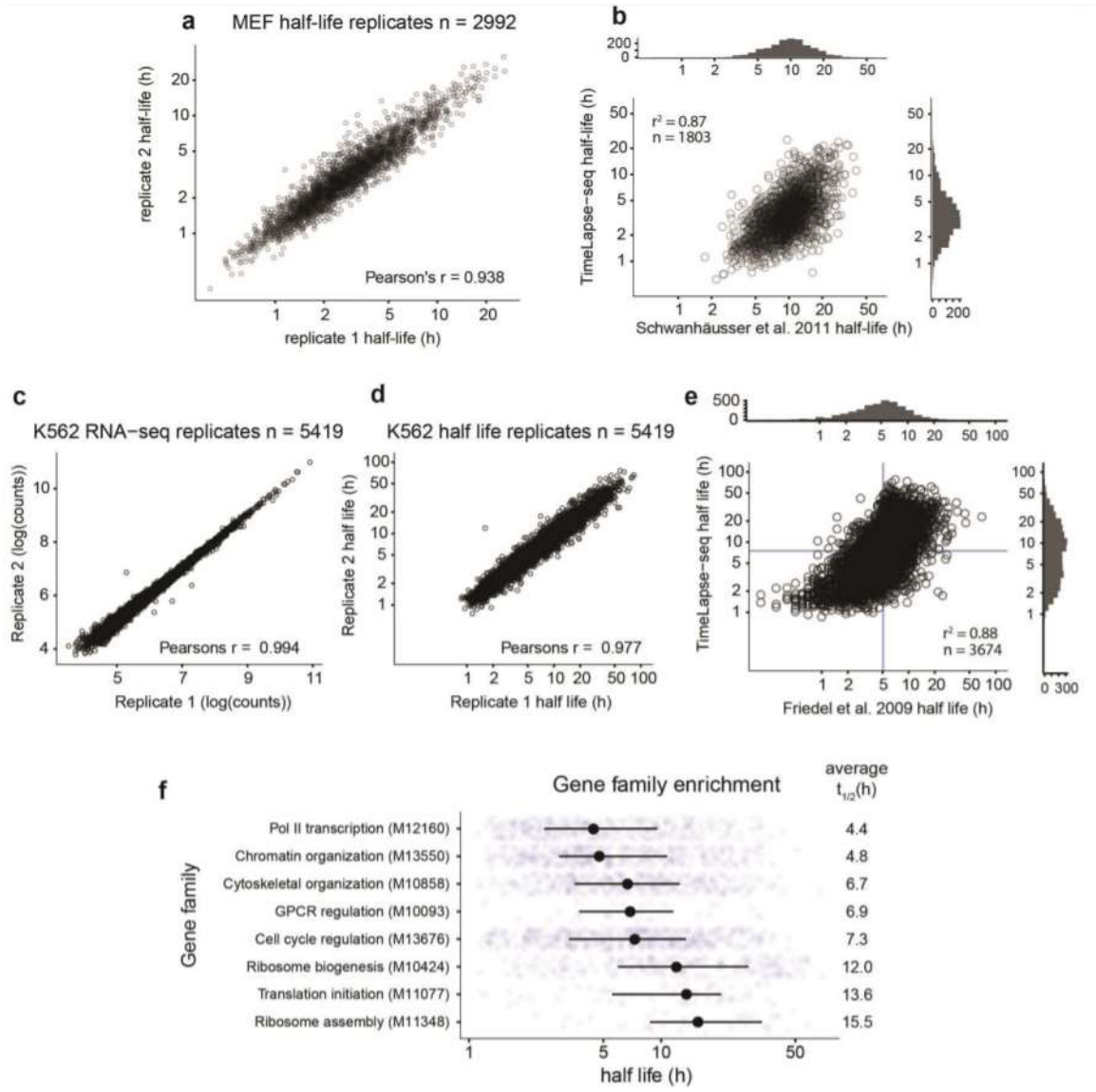

Figure 2.8: Reproducible estimates of transcript half-lives with in MEF and K562 cells. a) Correlation of estimated RNA half-lives between 1 h TimeLapse-seq replicates in MEF cells. b) Correlation with previously reported RNA half-lives in mouse 3T3 cells (Schwanhäusser et al. 2011). c-d) Correlation of replicate RNA-seq counts and estimated RNA transcript half-lives (see methods) from K562 4 h TimeLapse-seq samples. e) Correlation of estimated RNA transcript half-lives from TimeLapse-seq compared to RNA transcript half-lives derived from Friedel et al. 2009. f) Plot of estimated half-lives for K562 transcripts, filtered by indicated GO annotation term, and average half-lives for each family.

regulation of chromatin, and mutations in the longer isoform of this gene are implicated in myelodysplastic syndromes (MDSs) [115]. Analysis of the mutational content of the

individual exons from ASXL1 demonstrated that reads mapping to the longer isoform had substantially greater turnover than those mapping to the first four exons (Figs. 2.9 a-b), a conclusion supported by transcriptional inhibition (data not shown). The different stability of ASXL1 isoforms is particularly intriguing given the importance of RNA processing to many pathologies, including MDS [116].

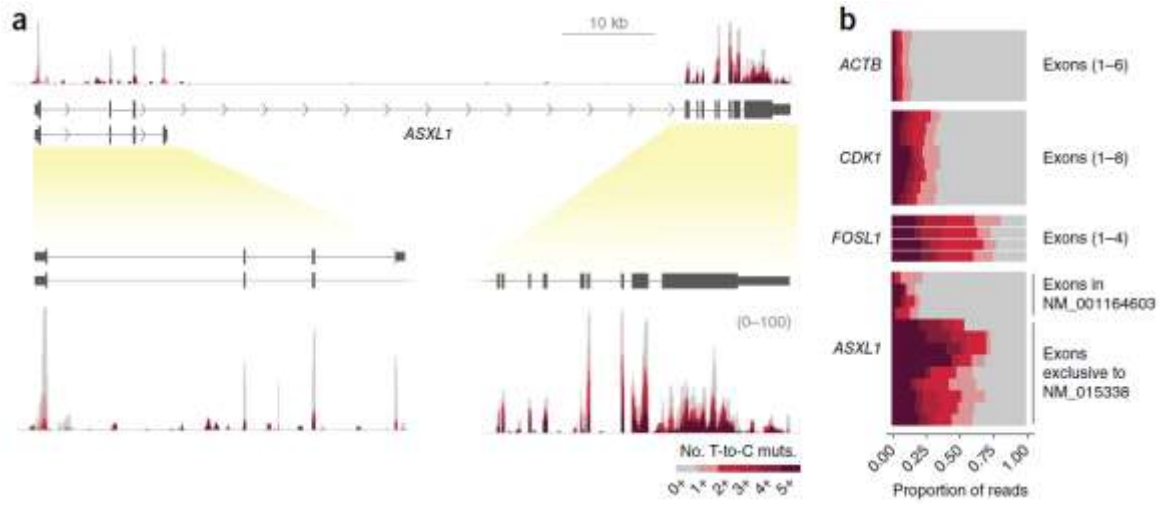

Figure 2.9. TimeLapse-seq reveals differential transcript isoform stability of the ASXL1 transcript. (a) ASXL1 tracks from TimeLapse-seq (4-h s^4U treatment) with exon-containing regions expanded (lower panel). (b) Exonic T-to-C mutation distributions for ASXL1 in comparison with three transcripts with different stabilities, ACTB, CDK1, and FOSL1.

2.6 Discussion

The oxidative nucleophilic aromatic substitution reaction optimized for TimeLapse-seq can in principle be used in combination with any existing s^4U-based method to assess RNA dynamics. We demonstrate the utility of combining TimeLapse chemistry as a bioinformatic filter for TT-seq. We have also combined TimeLapse chemistry with small RNA isolation in collaboration with the Steitz lab (see Chapter 4.2). In addition, TimeLapse chemistry is applicable to other thiol-based metabolic labels that are known to be incorporated into newly transcribed RNAs [86, 117]. Lea Kiefer found that TimeLapse chemistry efficiently converts the s^6dG nucleoside to a substituted adenosine by mass spectrometry and a restriction enzyme digestion assay [118]. Furthermore, performing

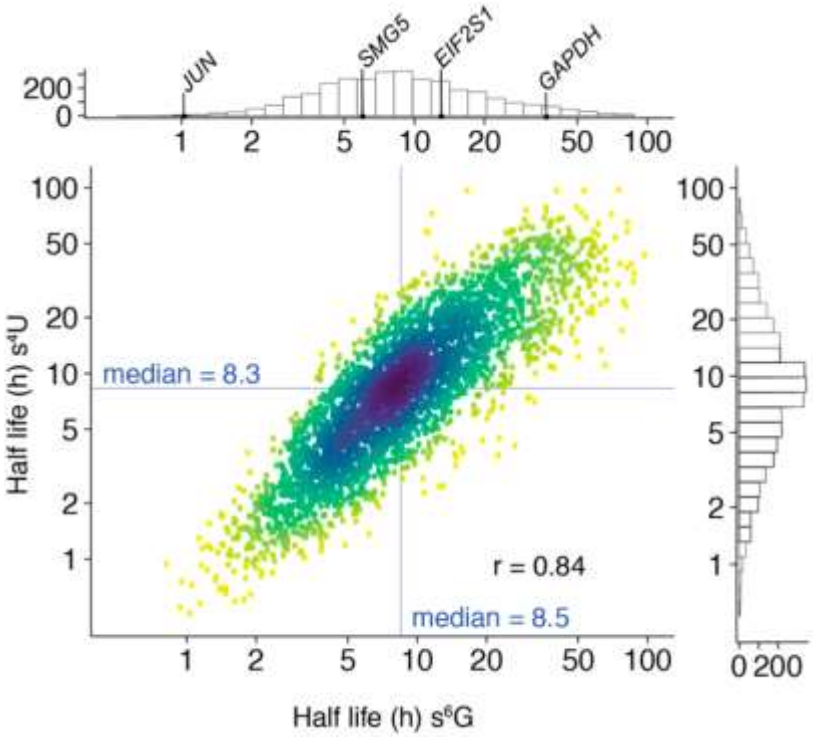

Figure 2.10 mRNA half-lives determined through s^6G recoding correlate well with those determined through s^4U recoding. Half lives were determined after 4 h s^6G treatment and binomial modeling, as in Figure 2.8, then compared to s^4U half-lives from Schofield et al. 2018. Figure reproduced from Kiefer et al. 2018.

transcriptome-wide TimeLapse-seq with a 4 h s^6G treatment in K562 cells, Lea found that estimated half-lives from our previously published s^4U and s^6G experiments correlate well (R = 0.84), with similar median half-lives (s^4U = 8.3 h, s^6G = 8.5 h, Fig. 2.10). These results demonstrate the ability to extend TimeLapse-seq into a second orthogonal metabolic label, which could be used in applications where use of s^4U is prohibitive, and to observe multiple timescales of RNA dynamics in a single sample.

The experiments reported in this chapter were performed using a pulse metabolic labeling strategy, though it has been demonstrated that cells can be metabolically labeled with 4-thiouridine for extensive periods of time followed by a chase with uridine to follow RNA decay rates. This approach has been used to estimate the decay rates of s^4U-labeled transcript using s^4U alkylation in an approach similar to TimeLapse-seq [119].

2.7 Limitations

In general, s^4U metabolic labeling has been found to be non-toxic to many cell types, though s^4U has been demonstrated to cause cellular stress in specific cell types [120]. An assessment of cellular viability (e.g. MTT colorimetric assay) should be performed before performing TimeLapse-seq in a novel system. The sensitivity of TimeLapse-seq depends on the number mutations observed, therefore the s^4U incorporation rate should be assessed (e.g. dot blot with MTS-TAMRA) and may require optimization for any given experiment. Finally, care should be taken during isolation and handling of metabolically labeled RNAs prior to TimeLapse-seq to prevent thiol-specific read loss and subsequent bias in sequencing. I describe in Chapter 3 improved handling

conditions to mitigate this read loss, but measurement of read loss and bioinformatic correction is often necessary.

Chapter 3

Considerations on the design and analysis of nucleoside recoding experiments

3.1 Author Contributions

I performed all experiments and bioinformatic analyses with assistance from Matthew Simon.

3.2 Summary

There are several considerations in properly designing, executing, and analyzing TimeLapse-seq and related s^4U-based approaches to study RNA dynamics. Among these considerations are the labeling time and s^4U concentration, handling of RNA material to prevent thiol-specific read loss, and accurate identification of new RNAs in sequencing.

3.3 Introduction

The transcriptome continually undergoes synthesis and decay to properly tune gene expression. In addition, regulated shifts in synthesis or decay rates allow cells to quickly change their expression profile, and defects in RNA stability are associated with a number of diseases. There is a growing need to assess transcriptome-wide RNA dynamics, particularly using methods that capture changes in both old and new RNA populations. Recently, a number of high-throughput approaches (TimeLapse-seq [88],

SLAM-seq [119], TUC-seq [121]) using metabolic labeling in combination with nucleoside modification have been reported to assess the dynamics of RNA. These approaches all take advantage of the orthogonal reactivity of s^4U to induce T-to-C mutations in new RNAs visible through RNA sequencing, though the method of nucleoside modification differs among these approaches. There are two principal strategies to induce mutations in thiolated nucleosides: recoding the Watson-Crick face through oxidative nucleophilic aromatic substitution, and disrupting the Watson-Crick face through alkylation.

With recent the increase in reports assessing total RNA with thiolated metabolic labels, there is a need to critically assess the robustness and biases of these methods. We have observed in our data and in literature an underrepresentation of sequencing reads aligning to fast turnover RNA features, including introns and short-lived transcripts. We expect these high turnover species to contain higher metabolic label content, indicating the bias is specific to thiolated RNAs. Without bioinformatic correction, this bias results in an underestimation of new RNAs over the labeling period. We therefore aimed to optimize the handling of thiolated total RNA to increase the robustness of metabolic labeling methodologies. We found that alterations in standard isolation methods upstream of nucleoside modification decreases thiol-specific read loss. These changes in the RNA handling protocol lead to reduced biases in transcriptome-wide metabolic labeling experiments, and we demonstrate the robustness of the protocol using TimeLapse-seq. This protocol can be widely applicable to all methods that require the isolation of s^4U or s^6G labeled RNAs prior to biochemical enrichment or nucleoside modification and sequencing.

3.4 Design

Experiments performed using s^4U metabolic labeling are prone to thiol-specific read loss. In order to screen for RNA isolation conditions that reduce s^4U-specific read loss, I designed an RT-qPCR assay to recapitulate the s^4U-specific loss without the need for time intensive and costly RNA sequencing. This dropout assay compares fast turnover transcripts, which are more likely to be lost in a s^4U-specific manner, to slow turnover transcripts containing relatively little s^4U. Using this assay, I screened through modifications to standard RNA isolation protocols to determine conditions that reduce the degree of thiol-specific read loss. To demonstrate the generality of the isolation conditions, I prepared transcriptome-wide TimeLapse-seq libraries with standard isolation and improved isolation. I then compared the fraction of intronic reads over total reads per gene, with the assumption that intronic reads will be predominantly new over a 2 h metabolic labeling period.

3.5 Results

3.5.1 Optimizing metabolically labeled RNA isolation using in-vitro dropout assay

To efficiently assess the degree of thiol-specific loss, I employed an RT-qPCR based dropout assay, as described in the methods. The dropout assay compares relative abundance of fast-turnover transcripts, which are expected to have higher thiol content and are therefore more frequently lost after metabolic labeling. Purification of s^4U-treated RNA using a standard isolation protocol results in extensive loss (referred to herein as

"dropout") of fast turnover transcripts (Fig 3.1), and this loss is specific to s^4U treatment. I tested several modifications of the isolation protocol, including addition of reducing reagents to various steps of the isolation, choice of cell lysis buffer, and method of RNA concentration. I found two steps in the isolation protocol that contributed most to the degree of dropout: cell lysis, and RNA concentration following genomic DNA depletion. Modifying these two steps in the isolation protocol (see methods) significantly reduces the degree of dropout in s^4U-treated samples (Fig 3.1). Additionally, these protocol modifications are generalizable to dropout induced by the G-based labels s^6G and 6-TG.

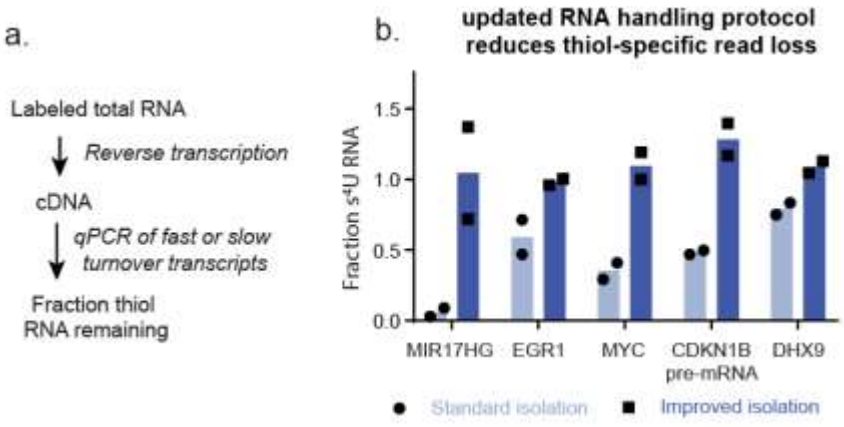

Fig 3.1. a) Scheme of RT-qPCR assay to assess thiol specific read loss. B) Comparison of standard RNA isolation protocol to improved isolation. Fast turnover transcripts incorporate a higher proportion of s^4U during labeling and are lost to a higher degree through standard isolation. Improved isolation rescues this s^4U-specific read loss.

3.5.2 Improved handling reduces bias in transcriptome-wide metabolic labeling

I prepared transcriptome wide TimeLapse-seq libraries treated with 200 uM metabolic label (s^4U, s^6G, or 6-TG) for 2 hours to compare standard isolation with our optimized isolation. We theorized that because introns are generally higher turnover species than exons, the fraction of intronic reads per annotated transcript observed in TimeLapse-seq data should be reduced in samples treated with the standard isolation. I in

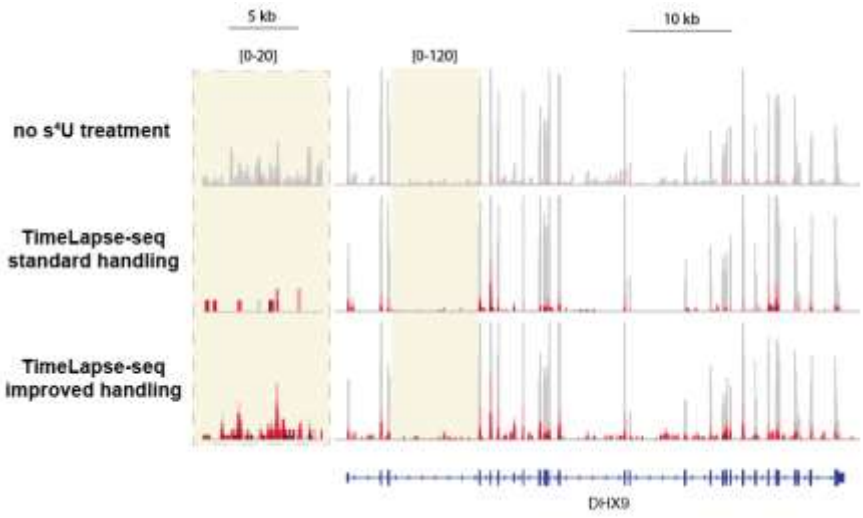

Figure 3.2: Representative browser shot of TimeLapse-seq tracks with standard isolation and improved isolation. Fast turnover transcripts are underrepresented in standard handling tracks with s^4U treatment, but are rescued in improved handling tracks. Left inset shows zoomed intronic region in yellow.

fact observe a depletion of intronic reads in metabolically labeled libraries proportional to gene intronic content using the standard isolation (Fig 3.3), evident in our data as well as s^4U-treated data obtained from literature [122]. Our modified protocol rescues the loss of high turnover introns, apparent in both TimeLapse-seq browser tracks and by the fraction of intronic reads observed per gene (Figs 3.2 & 3.3). As expected, because rescue of high turnover RNA species increases the number of mutated reads retained in sequencing, I observe a modest increase in per gene T-to-C mutation rates for s^4U labeled samples, as well as an increase in G-to-A mutation rates of s^6G and 6-TG labeled samples (standard average intronic T-to-C mutation rate = 4.3%, improved T-to-C = 5.0% T-to-C, standard G-to-A = 1.0-1.1%, improved = 1.4-1.5%). I then compared the consequences of our improved handling conditions on the analysis of fraction new of mRNAs at steady state. As expected, half-life estimates without correction using s^4U metabolic labeling and standard isolation are longer than fraction new estimates from the improved isolation

(standard mean = 7.9 h, improved mean = 6.0 h), due to an overrepresentation of pre-existing unlabeled RNAs using standard isolation.

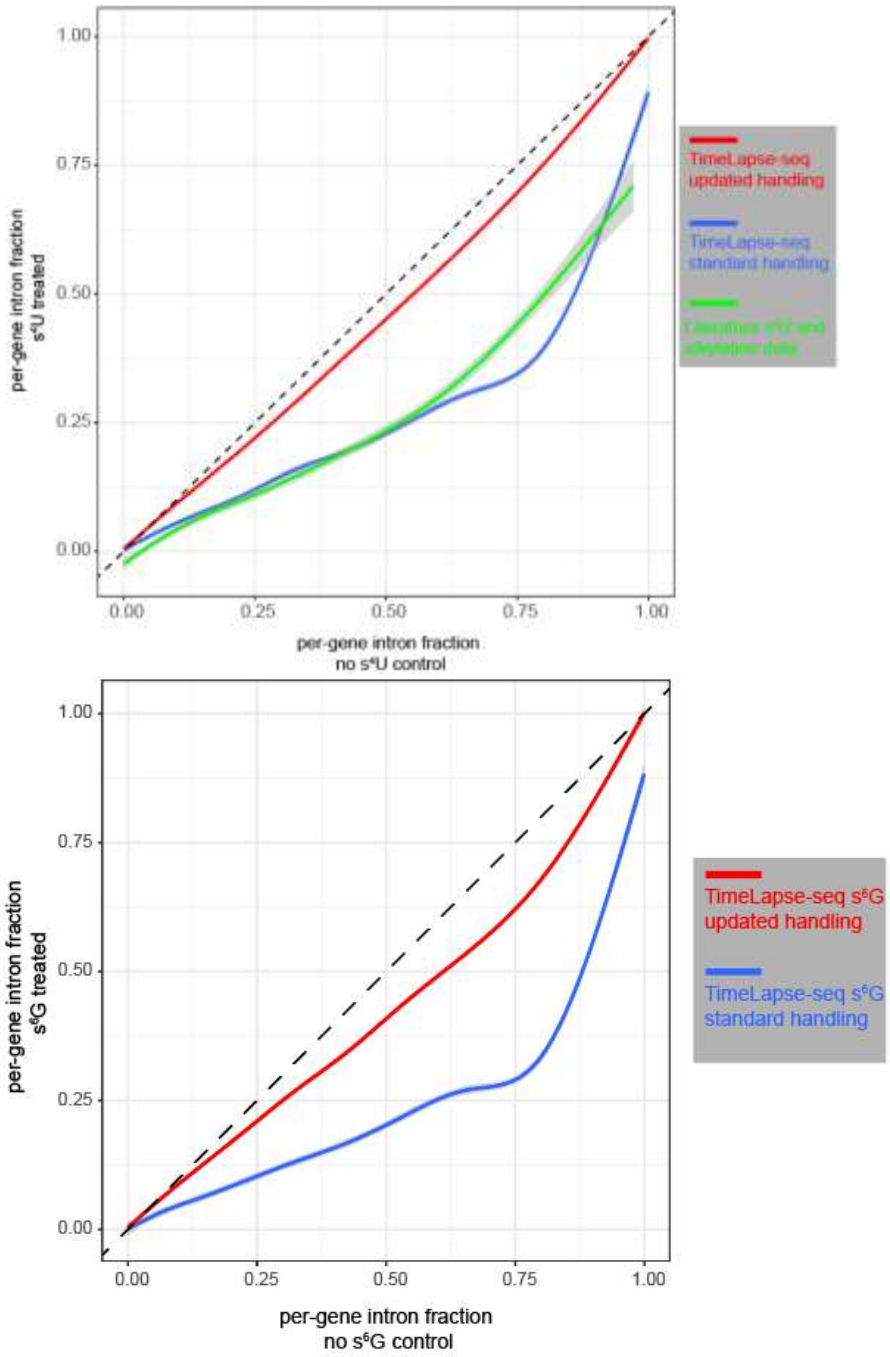

Figure 3.3. Transcriptome-wide analysis of thiol-specific read loss in standard isolation, improved isolation, and literature datasets. For each gene, the proportion of intronic reads is plotted for thiol-treated samples compared to untreated samples. A) plotted are s^4U-treated samples with standard isolation (blue), literature data (Muhar et al., green), or improved isolation (red). B) plotted are s^6G-treated samples with standard isolation (blue) or improved isolation (red). In general, there is a suppression of intronic reads in all thiolated nucleoside-treated samples, but the loss is more severe with standard isolation and in the processed literature dataset.

3.5.3 Binomial modeling approaches and bias correction for s^4U-specific read loss

We use three principal approaches to estimate the fraction new using the binomial distribution, each with its own advantages and drawbacks. A detailed description of each approach is described in the methods section, but broadly each approach takes into account the following: observed background and s^4U and chemistry-induced mutation rates; U content per read; and number of observations [88]. The first approach is a Bayesian hierarchical model and no U-turn MCMC sampling, which takes into account both global and gene-specific mutation rates to fit data. In addition, a parameter can be used for s^4U-specific read loss for experiments in which the observed s^4U-specific read loss is relatively high. This approach provides robust measures of experimental parameters but is computationally expensive. The second approach is a non-linear minimization (nlm) of the binomial distribution for observed mutations. The nlm approach generates reproducible estimates of fraction new for thousands of transcripts, increasing the scale of analysis. Nlm modeling does not fit values for a small proportion of transcripts, however. The final approach is assigning new and old reads based on a per-read mutation rate cutoff. New RNA and old RNA mutation distributions overlap to an extent dependent on s^4U-specific T-to-C mutation rate, background mutation rate, and read length. The mutation rate cutoff should be chosen to balance the sensitivity (true positive rate) and specificity (true negative rate) of assigned reads. I curated a test set of intronic reads in highly expressed transcripts as a proxy for new reads (with s^4U treatment) and old reads (without s^4U treatment). For each read, I determined the mutation fraction (T-to-C / number of Ts) and determined the sensitivity and specificity for a range of mutation rate cutoffs. I found for a paired-end 75 nt experiment an optimal

cutoff of 3.6% mutation content, with a sensitivity of 0.60 and a specificity of 0.85. Increasing read length to 150 nt increases sensitivity to 0.65 and specificity to 0.94. Comparison of nlm estimates to a simple mutation rate cutoff based on the fraction of mutations per read indicates demonstrates a high degree of correlation (Fig 3.4), indicating that the simpler cutoff approach is valid, and provides an orthogonal means of determining fraction new.

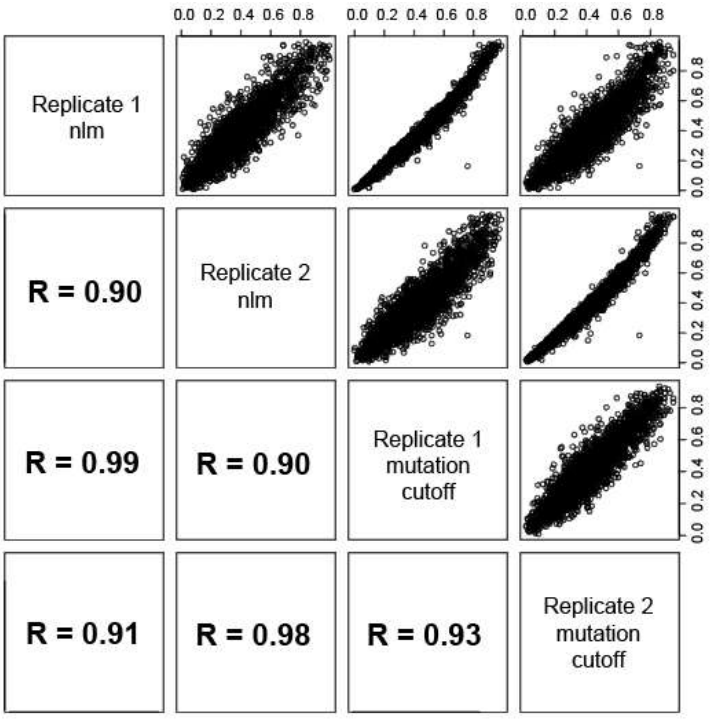

Figure 3.4: Comparison of fraction new estimates from biological replicates derived from non-linear minimization and fraction of mutations per-read cutoff. Replicates and strategies for binomial modeling display high levels of correlation.

3.6 Discussion

The utility of s^4U to examine RNA dynamics is evident by the wide range of applications by a number of research groups. Careful examination of all biochemical methods is necessary to avoid biases and to generate reproducible data. I report in this chapter an assessment of s^4U-specific read loss, which leads to an overestimation of

stable transcripts in total RNA sequencing. The analysis of s^4U-treated RNA sequencing data from literature [122] parallels our observation of thiol-specific read loss using standard handling conditions, indicating the generality of the bias. I report improved handling conditions to reduce this bias using straightforward modifications to the isolation protocol. Changes to the isolation protocol to improve handling are consistent with thiolated RNAs preferentially sticking to surfaces during lysis and concentration.

3.7 Limitations

The improved handling conditions developed in this chapter are specific to adherent cells with a Trizol-based isolation strategy. There are many isolation strategies for a diverse range of applications, and a critical examination of each strategy should be performed when necessary. RT-qPCR assays to assess read loss can be applied to any isolation strategy, provided the quantity of RNA is sufficient for the assay. Analysis of several sequencing libraries suggests that thiol-specific read loss is unavoidable to some extent, though improved isolation strategies can make this loss negligible. Bioinformatic assessment of read loss should be performed to determine whether correction is necessary.

Chapter 4

Collaborative work to profile RNA dynamics with TimeLapse-seq

4.1 Introduction

When developing TimeLapse-seq, we saw the potential for its application to a number of different types of RNAs in different biological systems. We demonstrated that TimeLapse-seq profiling can be used with nanograms of total RNA to profile mRNAs and combined with biochemical enrichment to profile transient RNAs [88]. To date, I have worked with a number of research groups to profile the dynamics of RNAs in a range of contexts, including: mRNA stability; miRNA turnover; transcriptome-wide and gene-specific analysis of transient RNAs; transcriptome-wide responses to drug treatment and viral infection; cellular localization of RNAs; effects on RNA processing with RNA binding protein mutants; and dynamics of mRNA splicing. This collaborative work has allowed us to continuously optimize s^4U-based sequencing methodologies and analyses, and opens up further avenues to investigate a variety of biological questions. In this section are selected applications of TimeLapse-seq in collaboration with research groups at Yale University.

4a Identifying changes in mRNA stability in DCP2 knockout

4a.1 Author Contributions

Experiments were designed by myself, Vicky Luo, Matt Simon and Sarah Slavoff. Experiments and data analysis were carried out by myself and Vicky Luo.

4a.2 Summary

I performed TimeLapse-seq to identify changes in mRNA stability in DCP2, MSI2, and NoBody KO cell lines. Using fold change analyses on both new and old RNA populations, I identified groups of transcripts that are more affected by changes in mRNA stability than changes in transcription. After validating stabilized transcripts using RT-qPCR, we explored the relationship between changes in transcript degradation rates, P-body localization, and m^6A modification. Our results provide evidence of the subset of transcripts that are substrates of DCP2 decapping, and support the role of DCP2 association with P-body enriched RNAS.

4a.3 Introduction

The regulation of mRNA stability is a complex process that is dependent on a number of factors. A widely studied aspect of RNA stability is the function and composition of processing bodies (P-bodies), ribonucleoprotein complexes that are composed of translationally repressed RNAs as well as RNA destabilizing protein

complexes [123]. A significant amount of work has suggested that P-bodies exhibit the properties of liquid droplets, phase separated regions that concentrate biochemical factors for increased local concentration or sequestration [124]. Liquid droplet formation requires multivalent interactions between biochemical factors, which is accomplished in P-bodies through a network of mRNAs and interacting proteins. Among the protein factors found to be enriched in these networks is YTHDF2, a reader of the m^6A RNA modification known to play a role in modulating mRNA stability [125]. P-bodies have been proposed to act as both storage sites and degradation sites for RNAs [126, 127], though recent microscopy work has suggested that mRNA degradation by the factors enriched in P-bodies does not occur within the P-bodies themselves but instead occurs in the cytoplasm outside of the phase separated body [128]. FRAP studies have demonstrated that components of translational repression and RNA decay shuttle dynamically between the P-body and the cytoplasm in at a rate inversely proportional to valency [129]. Among the factors that dynamically shuttle between P-bodies and the cytoplasm are the decapping Dcp1/Dcp2 complex, and the 5'-3' exonuclease Xrn1. 5'-3' RNA decay by the Xrn1 exonuclease require the removal of the m^7G cap structure prior to degradation, making decapping a rate limiting step in the 5'-3' decay pathway. The process of 5'-decapping has long thought to be catalyzed predominantly by Dcp2, though it was demonstrated that decapping occurs in cell types lacking DCP2 [130]. Recent studies suggest an expanded role in mRNA decapping by Dcp2 homologs Nudt16 [130] and Nudt3 [131]. Furthermore, novel components of RNA decay pathways continue to be discovered or further characterized. These factors include the RNA associated factor Msi2, which has been found to confer specificity in degradation in certain cellular

contexts [132], and has been demonstrated to bind to 3'-UTR regions of mRNA targets [133]. In addition, a novel microprotein named NoBody has been found to interact with components of mRNA decapping machinery, and perturbation of NoBody alters P-body abundance and the NMD process [134]. A number of questions remain concerning the exact role of particular components of mRNA destabilization pathways. Of particular interest is the role of DCP2 in mRNA decay, including its transcript specificity, and how DCP2 functions in the context of phase separated bodies to promote decay.

Despite consistent efforts to study P-bodies and mRNA stability, the exact substrates of DCP2 and other mRNA destabilizing factors remain difficult to determine. Transcriptional shutdown is highly perturbing to cells [72], and no single RNA-protein interaction or RNA motif has been shown to fully explain changes in RNA stability. We therefore worked with the Slavoff lab to determine the transcriptome-wide effects on mRNA stability with DCP2, MSI2, and Nobody KOs. Our analysis on perturbed stability of RNAs due to knockouts in components of RNA stability pathways relates specific subsets of transcripts to known determinants of RNA stability, including P-body enrichment and m^6A modification.

4a.4 Design

The Slavoff lab generated DCP2, MSI2, and NoBody HEK293T knockout cell lines to study the roles of these proteins on mRNA stability. I performed TimeLapse-seq on these cell lines using a 2 h s^4U labeling period to study changes in stability of mRNAs between a WT cell line and the mutant cell lines. I then developed analyses to determine rates of mRNA synthesis and decay, and used these analyses to determine candidate

transcripts that are post-transcriptionally stabilized in mutant cell lines. Changes in synthesis and decay were validated by transcriptional shutdown and RT-qPCR by Vicky Luo.

4a.5 Results

I performed s^4U metabolic labeling (2 h, 500 uM s^4U) in WT, DCP2 KO, MSI2 KO, and NoBody KO HEK293T cell lines. I performed RNA isolation, TimeLapse chemistry, and library preparation as described in Chapter 3. I estimated new and old read counts using the fraction of mutations per read cutoff. Using these counts, I performed differential expression analysis using DEseq2 to determine the log2-fold change and significance of change for new and old RNAs. I then performed rank ordering of new

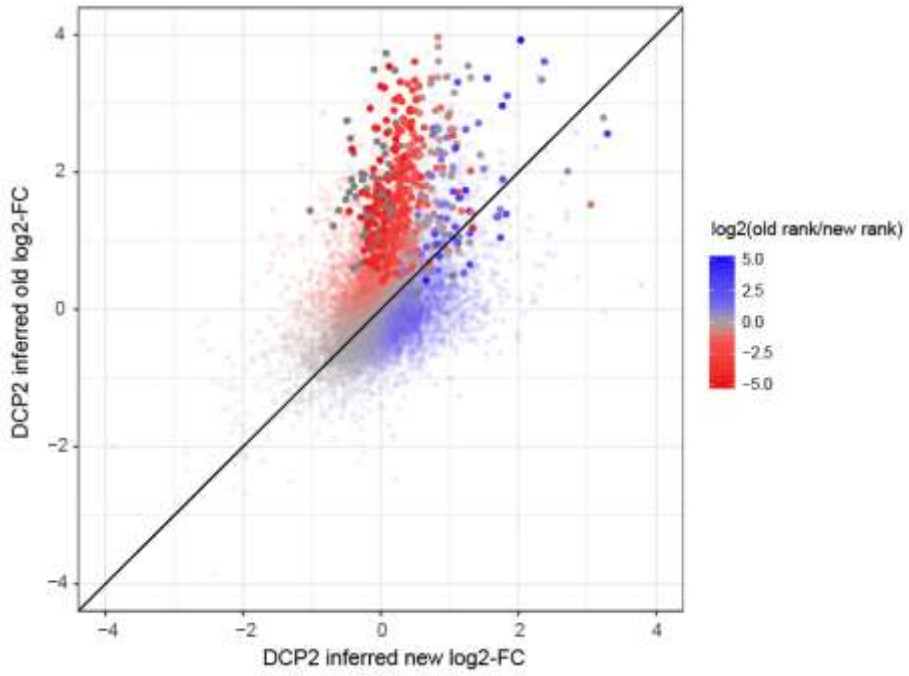

Fig 4.1.1. Plot of log2-fold change (DCP2 KO / WT) values for inferred new reads and inferred old reads from TimeLapse-seq data. Data points are colored by log2 ratio of P-value rank (lowest rank number = lowest P-value), and points in bold are the top 5% most significantly changed transcripts by old RNA-FC.

RNAs and old RNAs based on p-value, and ranks of new RNAs against ranks of old RNAs. I found that the separation of fold change analysis based on new vs. old RNAs allowed for the categorization of RNAs into distinct categories, including increasing transcription, decreasing transcription, increasing stability, and decreasing stability (Figures 4.1.1 and 4.1.2). Vicky Luo performed transcriptional shutdown time courses followed by RT-qPCR and validated several transcripts as being stabilized, changed in transcription, or unchanged by DCP2 KO.

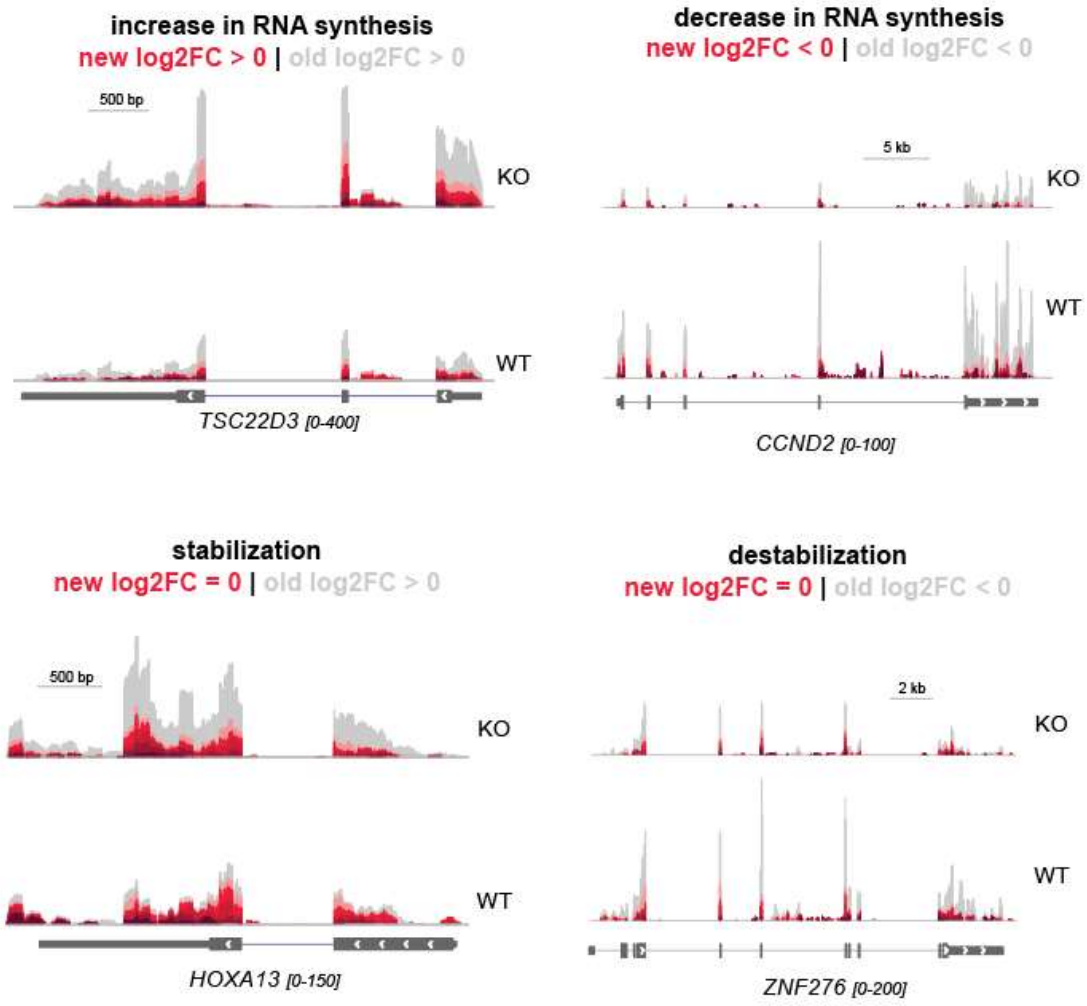

Figure 4.1.2. Selected examples of transcripts displaying diverse changes after DCP2 KO. Changes in both new RNAs (red) and old RNAs (grey) are interpreted as changes in transcription, whereas changes in old RNAs only are interpreted as changes in stability.

We then explored the relationship between RNA degradation rates and biochemical determinants of mRNA stability using literature data. I estimated transcript fraction new using a fraction of mutations per read cutoff, and determined the rates of synthesis and degradation for each transcript using the following equations:

$K_{deg} = -\log(1-F_{new})/2$

$K_{syn} = N_{total} * K_{deg}$

I then determined the fold change in degradation rate for transcripts in DCP2 and MSI2 KO compared to WT. Compared to all transcripts, we found a negative shift in degradation rate for p-body enriched transcripts (P-body enrichment > 2.5 [135]), and a positive shift in degradation rate for P-body depleted transcripts (P-body enrichment < -2.5 [135]), consistent with the role in P-bodies in the regulation of mRNA processing. We also found a negative shift in degradation rate for m^6A-enriched transcripts [136], consistent with the role of m^6A in mRNA stability. In addition, transcripts that are P-body enriched and m^6A enriched are stabilized to a greater degree than transcripts in either category alone. I performed this same analysis on MSI2 KO TimeLapse-seq data and observed the same trends in relation to P-body and m^6A enrichment, though the magnitude of shift is less than that of DCP2 KO (data not shown).

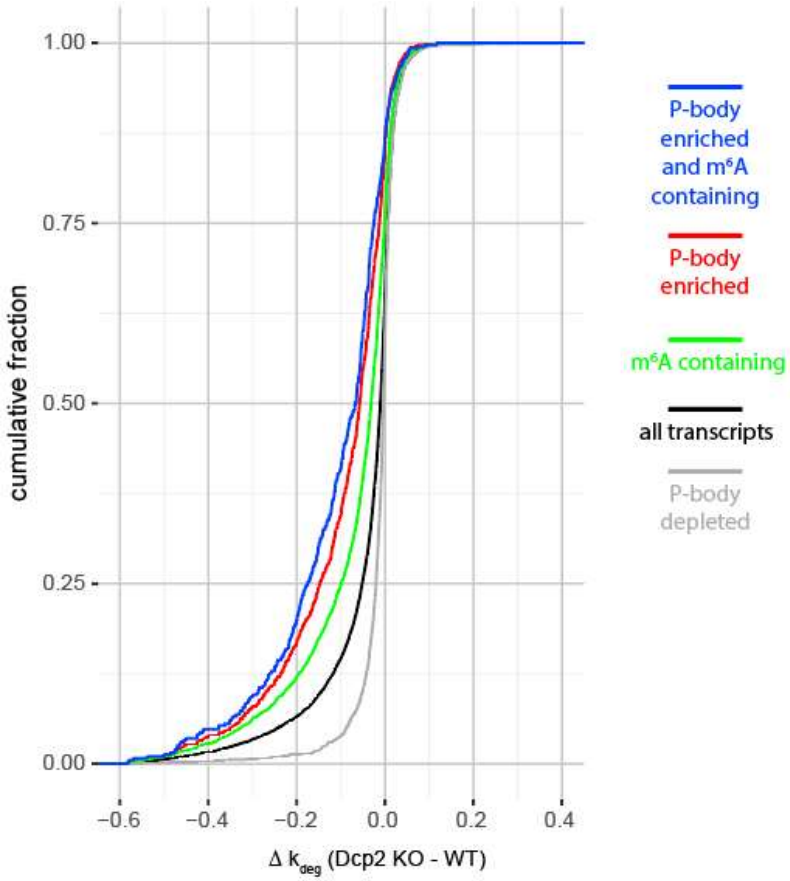

Fig 4.1.3. Cumulative distribution plot of change in degradation rate between DCP2 KO and WT transcripts. Transcripts enriched in P-bodies and m^6A modifications tend to have greater decreases in degradation rate with DCP2 KO, whereas P-body depleted transcripts display relatively little decrease in degradation rate with DCP2 KO.

4a.6 Discussion

TimeLapse-seq allows for the estimation of relative rates of synthesis and decay without the use of transcriptional shutdown. Our 2 h TimeLapse-seq data has provided a functional connection to DCP2, MSI2, and NoBody to a particular subset of transcripts. Follow-up DCP2 analyses provide further evidence of a functional link to P-body

localization of transcripts and mRNA degradation pathways. Further work on MSI2 and NoBody data will help to advance our understanding of the determinants of mRNA stability and the regulation of cellular homeostasis.

4.b Tracking changes in miRNA turnover with HSUR1 expression

4b.1 Author Contributions

Experiments were designed by myself, Paulina Pawlica, Matthew Simon, and Joan Steitz. Experiments were carried out by myself and Paulina Pawlica. miRNA data pre-processing was performed by Paulina Pawlica. Half-life modeling was performed by Matthew Simon with assistance from myself.

4b.2 Summary

I performed TimeLapse-seq on isolated miRNAs from cells treated with s^4U for 6 h and 12 h expressing the viral ncRNA HSUR1 to study the effect of HSUR1 on the stability of miR27-a. Consistent with literature, we found that HSUR1 leads to an increase in fraction new of miR27-a. We continue to build models to assess miRNA turnover and to determine changes in rates of processing steps in RNA turnover, including rates of tailing and degradation.

4b.3 Introduction

miRNAs are small RNAs that play a role in the post-transcriptional regulation of mRNA targets. miRNAs bind to AGO to form the miRNA-RISC complex, which can then bind to mRNA targets through base pair complementarity and either induce cleavage or translational repression [137]. Though their stability ranges widely, miRNAs tend to

have long half-lives relative to other RNA species [138, 139]. Much like mRNAs, miRNAs are also subject to modification post-synthesis, and these modifications can influence the miRNA's stability or targeting activity [140]. A common miRNA modification is non-templated 3'-end tailing through ribonucleotidyltransferase activity, including uridine addition through terminal uridine transferase (TUTase, [141]) activity. Tailing modification is able to take place after the miRNA has formed an association with AGO, suggesting that the turnover of the miRNA-RISC complex can be dynamically regulated through miRNA modification.

Viruses have the ability to target several aspects of the host miRNA pathway, including modulating expression of host miRNA processing factors and expressing viral miRNAs able to inhibit the host viral response [142]. Herpesvirus saimiri (HVS) is a virus that results in acute T-cell lymphoma in select species of monkeys [143]. HVS expresses a number of structured non-coding RNAs that have been shown to interact with host miR-27a, promoting accelerated miRNA degradation [144]. The HSUR1 ncRNA expressed in HVS has been demonstrated to interact with miR-27a, resulting in an increase in tailed miR-27a species. It follows that the mechanism of miR-27a destabilization could be through an increase in tailing rate, leading to a larger population of unstable miRNA species.

Given the precedence to study miRNAs using s^4U metabolic labeling [80], we worked with the Steitz lab to directly measure changes in stability of miRNA targets with HSUR1 expression. We examined several miRNA and miR* species, as well as 3'-end tailed miR-27a transcripts in order to better understand the mechanism of miRNA destabilization by HSUR1.

4b.4 Design

miRNAs tend to have longer half-lives than mRNAs, necessitating a longer metabolic labeling period. WT cells, empty-vector containing cells, and ncRNA HSUR1 expressing cells were treated with s^4U for 6 h and 12 h. miRNA libraries were prepared using small RNA-specific protocols adapted to integrate TimeLapse chemistry. We then used MCMC modeling to fit binomial models to estimate the fraction new of miRNAs and miR* transcripts, and determined the change in fraction new of different miRNA species with HSUR1 expression.

4b.5 Results

Small RNA sequencing reads were analyzed using a custom Perl script to identify T-to-C mutations and 3'-end tailing modifications. Because miRNAs are short in length, they contain very few uridines available for s^4U incorporation and chemical transformation. Despite this limitation, we observe miRNA reads containing multiple mutations, allowing us to determine miRNA and miR* fractions new. Attempts to model fraction new were difficult in longer (12 h) time points, as cellular s^4U levels diminish over the observation period, resulting in varying mutation probabilities for miRNAs. We therefore focused our analyses on the shorter 6 h time point to determine fraction new and fold change of miRNA species with HSUR1 expression. We utilized a binomial model to fit the data and determined the posterior ratio of HSUR1 to no viral expression. The data in the model consisted of reads from several abundant miRNA and miR* species with varying degrees of turnover, including miR-92, miR-27a*, miR-27a, miR20-a, miR-16. miR* species are intermediates in the generation of mature miRNA and are

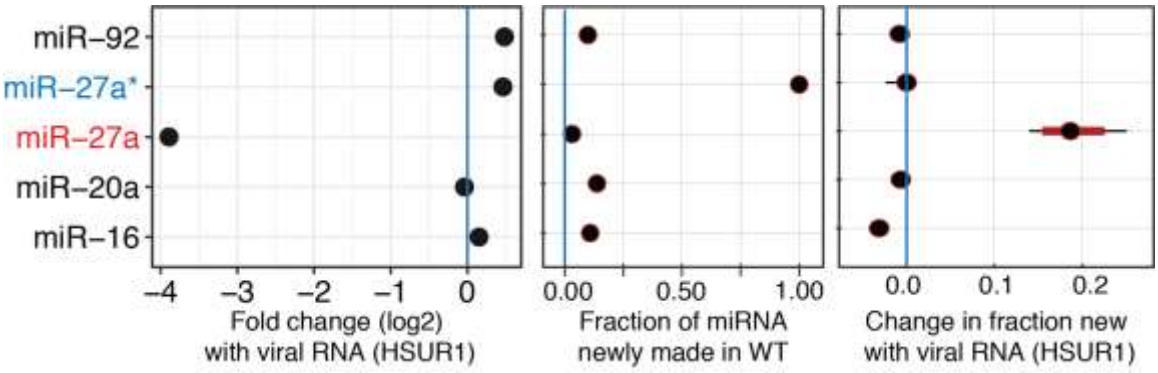

Figure 4.2.1. Effect of HSUR1 ncRNA viral expression on the dynamics of various miRNA and miR* species, estimated through binomial modeling. A) miR-27a is substantially downregulated with HSUR1, whereas other transcripts are not, consistent with previous reports (Cazalla et. al 2010). B) Estimated fractions new of slow turnover miRNAs and fast turnover miR-27a*. C) Estimated change in fraction new with HSUR1 expression. miR-27a appears less stable with HSUR1 expression.

therefore unstable relative to the 6 h labeling period and display less heterogeneity in mutation rates, as their window of synthesis is smaller. Data was incremented by viral expression, as mutation rates differed between samples in a HSUR1-dependant manner. Consistent with other reports [144], we found that with HSUR1 expression, both miR-27a and miR-27b display an increase in fraction new while other miRNAs and miR*s remained unchanged with HSUR1 expression. We subdivided miR27-a reads into different categories based on 3'-end modification, and found an increase in the number of tailed species with HSUR1 expression. We found that tailed miRNA species appear unstable, though our estimates for tailed miRNA species display a high degree of uncertainty due to the limited number of observations (data not shown).

4b.6 Discussion

We demonstrate in this section the utility of TimeLapse-seq to investigate the dynamics of miRNAs, and provide further evidence for the destabilization of miR-27a in

response to viral HSUR1 expression. Despite relatively few uridines available for mutations in small RNAs, we are able to model the fraction new of abundant miRNA and miR* species. Mechanistic insights into the specific mechanism of miRNA destabilization in relation to rates of miRNA tailing remain limited in this dataset due to high variance in estimates for rare tailed miRNA species. Further functional analyses following perturbation of tailing enzymes may be necessary to determine the exact mechanism of HSUR1-mediated miRNA destabilization.

4c Investigating transcriptional landscape in X-linked dystonia parkinsonism disease locus TAF1

4c.1 Author Contributions

Experiments were designed by myself, Anna Szekly, Matthew Simon and Sherman Weissman. Experiments were conducted by myself and Anna Szekly. Data analysis was performed by myself and Matthew Simon.

4c.2 Summary

Patients with X-linked dystonia parkinsonism frequently have repeat expansion mutations in an intronic region of the TAF1. We performed TimeLapse-seq on RNA isolated from family-matched human patient primary cells to analyze the transcriptional landscape of the disease locus. We performed analysis on transcripts originating from the TAF1 locus to explore changes in RNA processing between disease and non-disease patients. Our preliminary analysis suggests changes in intron retention in disease TAF1 locus, opening up further analyses of RNA processing in X-inked dystonia parkinsonism.

4c.3 Introduction

X-linked dystonia parkinsonism is a disease characterized by involuntary muscle contractions and bradykinesia [145]. XDP affected individuals contain mutations in the

intronic regions of the TAF1 locus that include SNPs and transposon insertions referred to as a SINE-VNTR-Alu element (SVA) [146]. XDP disease phenotypes tend to be more severe for affected individuals with increased numbers of repeats in the SVA insertion, motivating the study of the molecular biology in the TAF1 locus. Diseases associated with SVA insertions have been shown to have effects on gene expression driven by a number of molecular mechanisms, including changes in splicing and DNA methylation [147]. We established a collaboration with Sherman Weissman, who works with primary cells derived from XDP affected patients, to investigate the RNA dynamics related to SVA insertion in XDP affected patients.

4c.4 Design

Primary cells for TimeLapse-seq analysis were collected from patients with X-linked dystonia parkinsonism (from the CCXDP) and a family member without the disease as a control. Cells were from two affected patients with differing numbers of repeat insertions in the TAF1 locus, which correspond to the severity of the disease phenotype. We performed a 1 h s^4U metabolic labeling followed by TimeLapse-seq to capture a wide range of RNA dynamics. We performed transcriptome-wide TimeLapse-seq analysis and a detailed analysis on different TAF1 regions, including individual exons and introns.

4c.5 Results

TimeLapse-seq reads from XDP affected patient RNA and family matched RNA were analyzed using differential expression analysis, which revealed modest changes in

the total RNA population with SVA insertion (data not shown). Because previous work with SVA insertion diseases have been associated with changes in RNA processing, we further investigated TimeLapse-seq reads aligning specifically to the TAF1 locus. We observed a high degree of intron retention near intron 32, proximal to the site of SVA insertion (Fig 4.3.1). Recently, another group published the observation of XDP intron retention [148], underscoring the interest in RNA processing in the TAF1 locus.

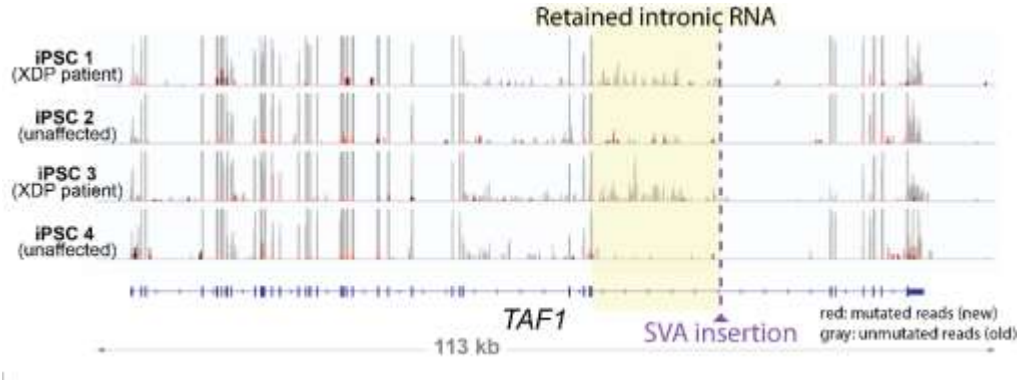

Figure 4.3.1. Browser shot of TimeLapse-seq reads aligning to the TAF1 locus. Data is derived from primary cells from XDP affected individuals and family-matched unaffected individuals (iPSC 1 and iPSC 2 are matched, and iPSC3 and iPSC 4 are matched). Intronic reads in intron 32 (highlighted in yellow), proximal to the SVA insertion site, are retained in XDP affected patient cells.

4c.6 Discussion

Retrotransposon insertions and subsequent changes in RNA processing and epigenetic regulation have been shown to contribute to the pathology of a number of diseases [149]. Interpretation of our TimeLapse-seq data from XDP affected individuals is limited due to relatively low read depth in intronic regions of the TAF1 locus. Future work will take advantage of the repertoire of RNA tools developed in the Simon lab to study a range of RNA dynamics in iPSC cells and the TAF1 locus specifically. Transient

RNA reads generating from the TAF1 locus will be analyzed by short s^4U pulses (TT-TimeLapse-seq), and changes in splicing and RNA turnover will be analyzed using deeper TimeLapse-seq libraries. Each of these experiments will be improved by the optimized handling methodology described in Chapter 3.

Chapter 5

Examining the dynamics of transcription start site switching with HDAC inhibition

5.1 Author Contributions

I performed all experiments with assistance from Josh Zimmer. I performed bioinformatic analyses with assistance from Josh Zimmer and Matthew Simon.

5.2 Summary

I performed TimeLapse-seq on 293T cells after treatment with histone deacetylase (HDAC) inhibitor trichostatin A (TSA). I observed changes in transcription start site (TSS) usage after TSA treatment for hundreds of transcripts, in which downstream start sites are upregulated and upstream start sites are repressed. We hypothesize that downstream TSS activation results in feedback to suppress upstream transcription, or that nucleosome remodeling activity functions to physically occlude the upstream TSS. To study potential mechanisms of TSS switching, I performed a detailed RNA-seq analysis, ChIP-qPCR for chromatin modifications, and TT-TimeLapse-seq to study transient transcription. I found that transcription and acetylation at switching TSSs change dramatically within minutes of TSA treatment. I am continuing work to assess the dynamics of transient transcription due to acute HDAC inhibition and the open chromatin landscape to determine the mechanism of TSS switching.

5.3 Introduction

The ability for cells to specialize and adapt to their environment is provided in large part by the dynamic regulation of chromatin. Histone modifications define regulatory features and help coordinate factors that promote or suppress transcription. In addition, histone modifications and chromatin structure change drastically through processes including the cell cycle [150], X-chromosome inactivation [151], and differentiation [152]. Histone acetylation is frequently associated with regions of active transcription, including enhancer and promoter regions [153, 154]. Histone acetylation levels are regulated by processes of acetylation by histone acetyltransferases (HATs) and deacetylation by histone deacetylases (HDACs).

HDAC inhibitors are used clinically to treat certain cancers, in part because they lead to the induction of apoptosis pathways and because HDACs are known to act on non-histone proteins (including P53 [155]) that control cellular proliferation [156]. It has been observed in cell culture that treatment with HDAC inhibitors lead to increased acetylation in the gene body away from the promoter [157]. This gene body acetylation can result in alternative splicing due to increased PolII processivity, and has been observed to induce cryptic TSS initiation far downstream of the promoter [158, 159]. These so-called treatment-induced non-annotated TSSs (TINATs) can result in truncated and misfunctional proteins. Brocks and coauthors found that TINATs originate at long terminal repeats (LTRs), sequences originating from retrotransposon insertion into host chromosomal DNA. TINATs were induced by treatment with suberanilohydroxamic acid (SAHA), a class I and II HDAC (zinc hydrolase family) inhibitor structurally similar and with overlapping specificity for HDAC enzymes to the thoroughly characterized

trichostatin-A (TSA) [160]. The median distance between the new TSS and pre-existing TSS is 9 kb, and were generally not expressed in normal tissues except for testicular and fetal thymic samples.

Profiling of the transcriptome after HDAC inhibition shows widespread changes in gene expression, including the upregulation and downregulation of thousands of transcripts within hours of treatment [161, 162, 163]. Global changes in H4ac also occur within the first 6 hours of treatment, consistent with observed timescales of epigenetic regulation for acetylated histones [161]. Given the large degree of changes to transcription and RNA processing over timescale of hours, we decided to perform TimeLapse-seq after HDAC inhibition to better understand the early changes to the transcriptome after TSA treatment.

5.4 Design

To observe the transcriptional changes due to changes in histone acetylation, I treated 293T cells with HDAC inhibitor TSA for 6 h, then performed s^4U metabolic labeling for 1 h followed by TimeLapse-seq. I determined changes in transcription through differential expression analysis for both total RNA-seq as well as inferred new reads based on T-to-C mutations. To study changes in RNA isoform usage due to TSA treatment, I performed differential exon usage analysis, and found widespread changes in transcription start site usage. Using ChIP-qPCR, I probed the levels of acetylation at pre-existing and newly established transcription start sites, and performed a time course to determine the kinetics of acetylation changes. After determining that drastic changes occur in minutes after TSA treatment, I profiled transient transcription and start site

dynamics using TT-TimeLapse-seq and Start-seq after 10 and 30 minutes of TSA treatment.

5.5 Results

5.5.1 TimeLapse-seq transcriptional changes after TSA treatment

Treatment of 293T cells with TSA for 6 h resulted in widespread changes in the transcriptome. I identified 1034 genes significantly upregulated and 780 genes significantly downregulated in total RNA (DEseq Padj < 0.05, log2-FC cutoff of 1.5). Using the mutations per read cutoff strategy as described in the methods section, I identified 963 genes significantly upregulated and 844 genes significantly downregulated in new RNA. Compared to the 1048 genes reported as "TSA responsive" in several cell types from a recently published ligation-based expression profiling dataset [162], 283 genes are found in the top ~1000 most significant differentially expressed genes in total RNA after 6 h of TSA treatment in our dataset (DEseq Padj < 10^{-30}).

5.5.2 Alternative RNA isoform usage with TSA treatment

Apart from changes in transcription, I also observed changes in RNA processing after 6 h TSA treatment. I observed instances of alternative splicing after 6 h of TSA treatment, consistent with previous work demonstrating the effect of histone acetylation on PolII processivity and splicing factor binding [158]. In particular, I observed cassette exon skipping of exon 33 in FN1 (data not shown), a gene known to utilize different isoforms dependent on cellular context [164] and shown to be TSA responsive [158]. To further study changes in splicing due to TSA treatment in our data, I performed

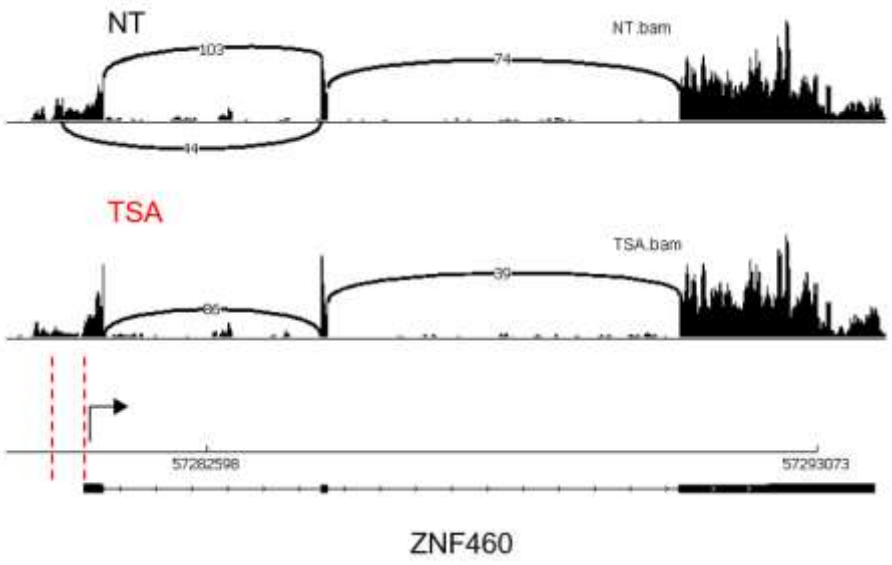

Figure 5.1: Example RNA-seq tracks and Sashimi plots of ZNF460 gene undergoing alternative TSS usage with TSA treatment. Numbers indicate observations of splice junction containing reads spanning indicated arc. Without treatment, upstream and downstream splice sites are utilized, though with TSA treatment usage of the upstream splice site is suppressed.

differential exon usage analysis using DEXSeq to identify genes with significant changes in isoform usage due to TSA treatment in our RNA-seq data. There were 2363 genes displaying alternative exon usage ($p < 0.05$), recapitulating widespread changes in RNA isoform usage described in literature. To better assess the nature of alternative splicing with TSA, I inspected RNA-seq tracks for the 265 most significantly changed genes ($p < 10^{-7}$). Unexpectedly, 71 of these genes display changes in the first exon (including ZNF460, Fig 5.1), suggesting TSA treatment causes changes in transcription start site usage. Of the 71 initially observed genes with alternative start site usage, 67 display a relative increase in downstream TSS usage. To determine the approximate positions of transcription start sites from RNA-seq data, I visually examined aligned data using the IGV genome browser. I assigned TSS locations to the nearest annotated start site from

NCBI refseq and/or locations of unambiguous read density at the 5'-end of the gene. The majority of these switching TSSs are within a range of 200-5000 nt downstream of the first TSS, with a median distance of 600 nt, shorter than the previously characterized TINATs [159]. To quantify the observed changes in total RNA, I determined the number of RNA-seq reads in the range of 0-200 nt downstream of upstream and downstream start sites, and normalized these reads to gene-specific expression changes between TSA and no treatment (Fig 5.2). I found that there was no change in normalized reads following the downstream TSS, but there was a notable decrease in normalized reads following the upstream TSS.

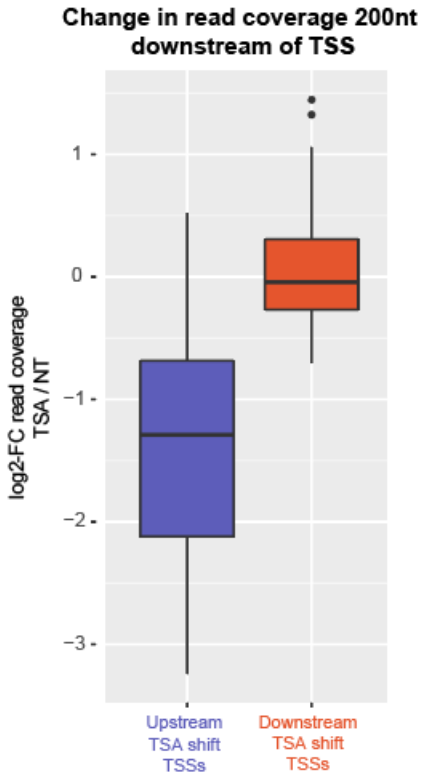

Figure 5.2: Suppression of total RNA following upstream TSSs after 6 h of TSA treatment. RNA-seq reads were counted in the region of 0-200 nt downstream of each TSS, and log2-FC with TSA treatment was calculated and normalized to the overall log2-FC of the gene.

5.5.3 Transcription start site analysis after 6 h of TSA treatment

To further characterize the changes in TSS usage with TSA treatment, Josh Zimmer performed Start-seq on samples treated with TSA for 6 h. Start-seq utilizes 5'-cap enrichment and gel electrophoresis size selection to isolate newly initiated RNA [165]. I then performed differential expression analysis on TSSs identified by TSScall. I inspected the 100 most significantly changes TSSs ($p < 10^{-10}$), several of which were identified through total RNA-seq analysis. I then used peak locations from Start-TimeLapse-seq to precisely define the location of upstream and downstream TSSs for the expanded list of switching genes (n = 158). Using the positions of TSSs as defined by TSScall and through visual inspection, I determined the number of reads aligning to upstream and downstream sites. Consistent with mRNA analysis, the Start-seq abundance at upstream TSSs appears to be suppressed (data not shown).

5.5.4 Analysis of chromatin modifications at switching TSSs

I next performed meta analyses on ChIP-seq data downloaded from literature to profile the chromatin state of the upstream and downstream TSSs. As expected, the upstream TSS displays a nucleosome depleted region flanked by chromatin marks H3K4me3 as well as acetylation. The downstream TSS also resides in a nucleosome depleted region marked with a downstream H3K4me3 peak of equal magnitude to the upstream peaks. In contrast, the upstream TSS is strongly acetylated, while the downstream TSS displays only weak acetylation. The chromatin landscape at downstream TSSs resembles that of poised enhancers, which are marked by H3K4me1 but devoid of acetylation

To determine whether TSS switching is general to all genes with similar gene structures, I created a control set of genes by pulling gene annotations from UCSC and filtered for genes with two annotated start sites separated by distances within the same range (200-5000nt) of the majority of switching genes (n = 1039). The acetyl landscape of the control upstream start sites appears similar to those of switching upstream start sites. I also compared the acetyl landscape of promoters after SAHA treatment from literature data [157]. Treatment with SAHA leads to an increase in relative acetylation directly downstream of the TSS, consistent with literature observations. Together these results suggest that the downstream start sites are poised for activation, and are activated upon HDAC inhibition and acetylation in the gene body.

To assess the changes in acetylation due to TSA at specific upstream and downstream TSSs, I next performed ChIP-qPCR in 293T lysates using a pan-Kac

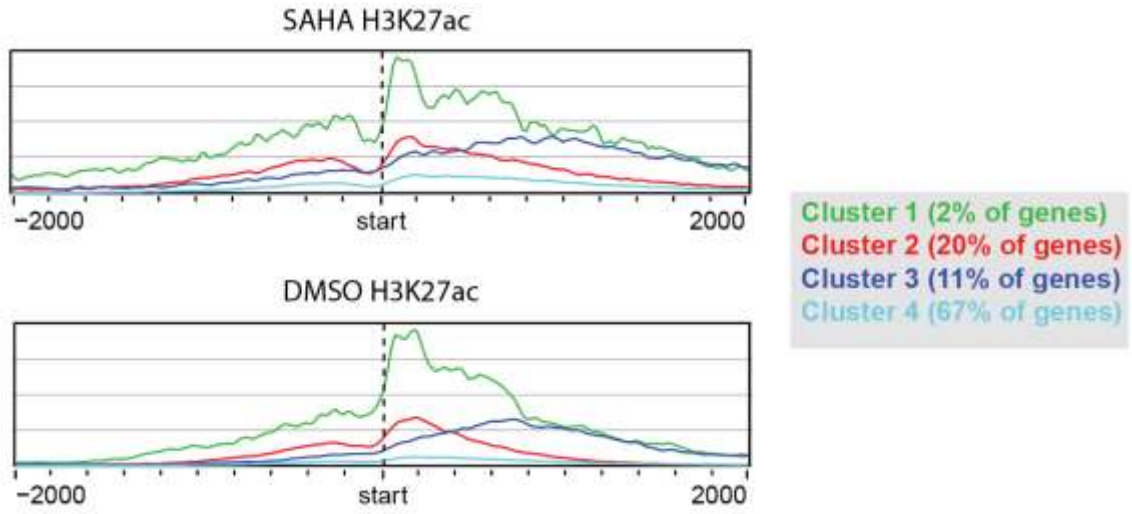

Fig 5.3: H3K27ac enrichment increases in the gene body following SAHA treatment. ChIP-seq data obtained from Greer et al. 2015. H3K27ac signal surrounding the promoter region was clustered into 4 classes based on no HDACi control (DMSO). Classes of genes with low levels of acetylation (clusters 2 and 4) display an increase in acetylation downstream of the promoter, consistent with previous observations.

antibody and an H4ac antibody, with and without 6 h TSA treatment. I designed qPCR primers against several loci that undergo TSS switching, as well as control gene-poor genomic regions that are expected to be devoid of acetylation. I found that the control intergenic regions increased in acetylation (pan-Kac mean fold change = 20.9, H4ac mean fold change = 10.6, n = 2), consistent with reports of increased acetylation in intergenic regions with HDAC inhibition [157]. The regions corresponding to the downstream TSS appear to increase in both total acetylation and H4ac (pan-Kac mean fold change = 3.1, H4ac mean fold change = 2.1, n = 5), consistent with their increase in TSS usage in TimeLapse-seq data. Acetylation appears to decrease at the upstream TSS (pan-Kac mean fold change = 0.6, H4ac mean fold change = 0.3, n = 6), suggesting not only that the downstream TSS is activated but that the upstream TSS is being suppressed with TSA treatment.

5.5.5 Profiling dynamics of TSS switching after TSA treatment

Our TimeLapse-seq data analyzes the dynamic transcriptome after 6 hours of TSA treatment. Global changes in histone acetylation are known to occur within the first 6 hours, and depending on the modification site and chromatin context, acetylation is dynamic on the timescale of minutes [166]. To determine the time scale on which the

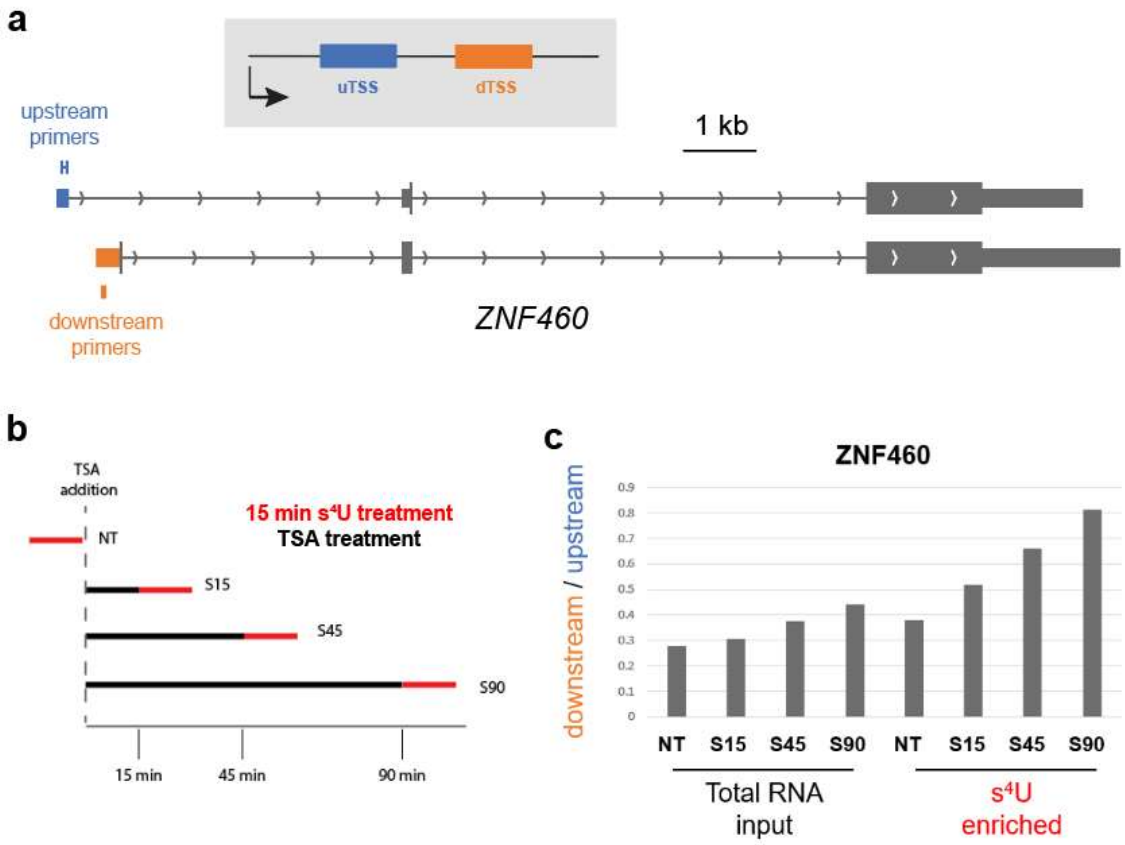

Fig 5.4: TSA switching gene ZNF460 undergoes change in relative usage within first 30 min of treatment. a) Locations of primers to assess relative abundance of ZNF460 isoforms. B) Scheme of MTS-biotin enrichment of s^4U treated cells after a time course of TSA treatment. c) Relative abundance of downstream vs upstream ZNF460 expression following TSA treatment by RT-qPCR.

TSS switching occurs, I designed qPCR primers against the upstream and downstream spliced RNA isoforms for several genes. I performed TSA treatment in 293T cells for 15 min, 45 min, and 90 min, followed by 15 min of s^4U metabolic labeling after each time point. I then performed MTS-biotin enrichment and RT-qPCR to measure the relative upstream vs. downstream isoform usage. I found that the relative change in transcription in unspliced primers for ZNF460 begins within the first 15-30 min of TSA treatment (Fig 5.4). In spliced primer sets, there is a delay in observing upstream vs. downstream

switching, occurring within the first 45-60 min of TSA treatment. These results suggest the switch in TSS occurs in an acute manner, well within the first hour of TSA treatment.

I then performed a time course of TSA treatment followed by ChIP-qPCR for H4ac. Consistent with prior reports, I observed that the level of H4ac spikes within the first 15-45 min of TSA treatment for both upstream and downstream sites (Fig 5.5) [166]. The deacetylation dynamics appear different for upstream and downstream sites, however, as the level of acetylation after 180 min of TSA treatment remains high relative to t = 0 for downstream TSSs, but returns to or drops below t=0 for upstream TSSs.

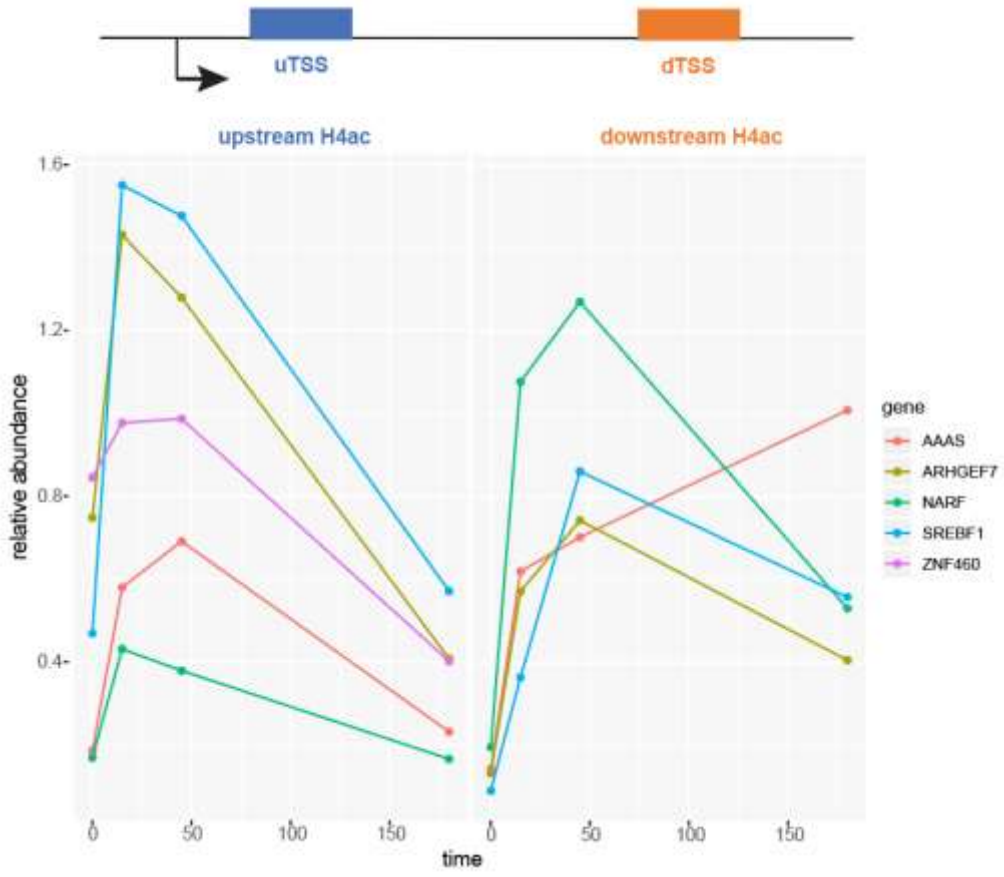

Figure 5.5: H4-acetylation exhibits highly dynamic behavior within minutes of TSA treatment. ChIP-qPCR relative abundance of pan-H4ac using primers located at upstream and downstream TSSs over a time course of TSA treatment. H4ac signal peaks for both upstream and downstream TSSs, but appears to remain higher at downstream TSSs after 180 min of treatment.

These results suggest widespread and dynamic changes to the chromatin following TSA treatment that may contribute to significant changes to chromatin structure and the regulation of activity at alternative promoters.

5.5.6 Measuring transient transcription after TSA treatment

To measure the dynamics of the transient transcriptome after HDAC inhibition, I performed TT-TimeLapse-seq with 5 min of s^4U after short (10 min and 30 min) TSA treatment. I filtered for bona fide new transcripts across the gene body by using the fraction of mutations approach as described in the methods. I input these inferred new reads into k-means clustering to assess the range of responses over time to TSA. I found that genes exhibit a wide range of acute transcriptional responses over the first 30 minutes of TSA treatment (Fig 5.6). Over 50% of TSS switching genes are in clusters 1

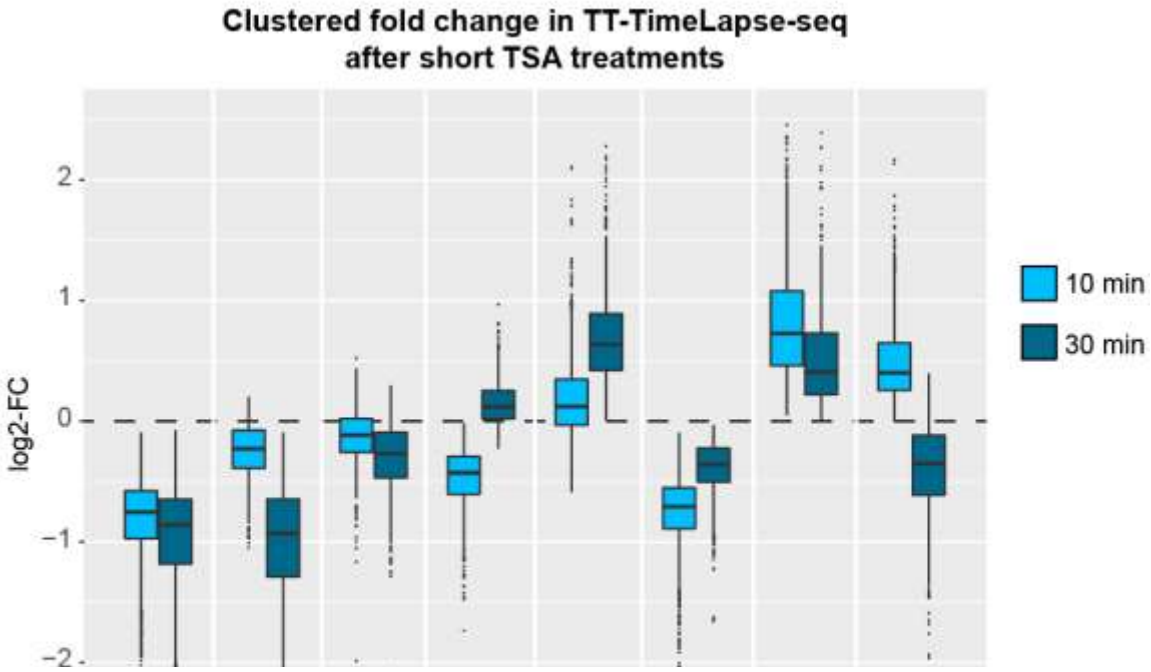

Figure 5.6: Genes display a wide range of acute transcriptional responses to short TSA treatment. Log2-FC is plotted on the Y-axis for inferred new reads from TT-TimeLapse-seq data after 10 min (light blue) or 30 min (dark blue) TSA treatment. Clusters were determined through k-means clustering (see methods).

and 2, which display the highest degree of downregulation after 30 min of TSA treatment.

5.6 Discussion

Inhibiting the biochemical activity of histone deacetylation results in profound effects on chromatin and on the dynamics of transcription. I found in particular that treatment with HDAC inhibitor TSA results hundreds of significantly differentially expressed transcripts after 6 hours. In addition, hundreds of genes are changed in their relative usage of TSSs after TSA treatment. Much of the reported work profiling transcription after HDAC inhibition is performed after 24 hours or more of drug treatment, but our results and others in literature suggest [166] that many effects on chromatin occur within minutes. The fast kinetics of changing acetyl landscape opens up a number of directions for study. My initial profiling of transient transcription in the promoter regions of switching genes does not indicate induction of antisense transcription generated from the downstream TSS, inconsistent with a model of transcriptional interference from downstream activation. Despite this inconsistency, further characterization of the transient transcriptome through early time points of TSA treatment will provide insights into the rapidly changing chromatin landscape after HDAC inhibition. In addition, profiling of the accessible chromatin landscape in early time points of TSA treatment may indicate whether occlusion of the upstream TSS is responsible for the observed repression of upstream transcription. It has been demonstrated previously in yeast that nucleosome remodeling activity at gene promoters is important to preserve the fidelity of transcription initiation and reduce cryptic initiation

[167], and that chromatin remodeling and nucleosome repositioning is important for changing transcriptional profiles in a number of contexts [68, 168]. In addition, the median distance between upstream and downstream TSSs in our switching genes is about 600 nt, a distance equivalent to approximately 4 nucleosome units. In yeast, the SER3 serine biosynthetic gene has been demonstrated to be suppressed by transcriptional interference from the tandem promoter SRG1 oriented in the sense direction 400 nt upstream [169]. Expression of SER3 in serine depleted media was found to be dependent on the repression of upstream SRG1 [170], and furthermore this repression is dependent on nucleosome remodeling activity by Spt6 to occlude the upstream TSS [171]. A TSA time course with a genome-wide accessibility assay such as ATAC-seq will provide insights into whether the suppression of upstream transcription could be due to rapid chromatin remodeling activity leading to occlusion of the upstream TSS. Finally, an investigation into gene-specific functional consequences of RNA isoform usage may provide insights into biological disease types. For example, profiling of alternative isoform usage in multiple cancer cell types has demonstrated that the expression of specific alternative RNA isoforms is associated with decreased survival rate [172]. One such example of a switching gene with differential cancer outcome is ZNF12, a transcription factor containing a KRAB box transcriptional repression domain in one of its commonly expressed isoforms. We observe in our data TSS switching in the ZNF460 gene, which also contains a KRAB box domain in one of its isoforms. Isoform switching events in our data change the sequence of 3'-UTR and/or the open reading frame, potentially changing the stability of the mRNA or function of the protein.

The data presented in this chapter present changes in TSS usage distinct from previous HDACi-induced cryptic initiation [159], and open up a number of potential mechanisms for the suppression of upstream transcription. Presented in Figure 5.7 are two potential mechanisms for TSS switching, driven predominantly by transcriptional interference or nucleosome remodeling activity. Ongoing work to assess the transient transcription and chromatin accessibility after short TSA treatments will help to refine these models and provide further insights into the regulation of chromatin.

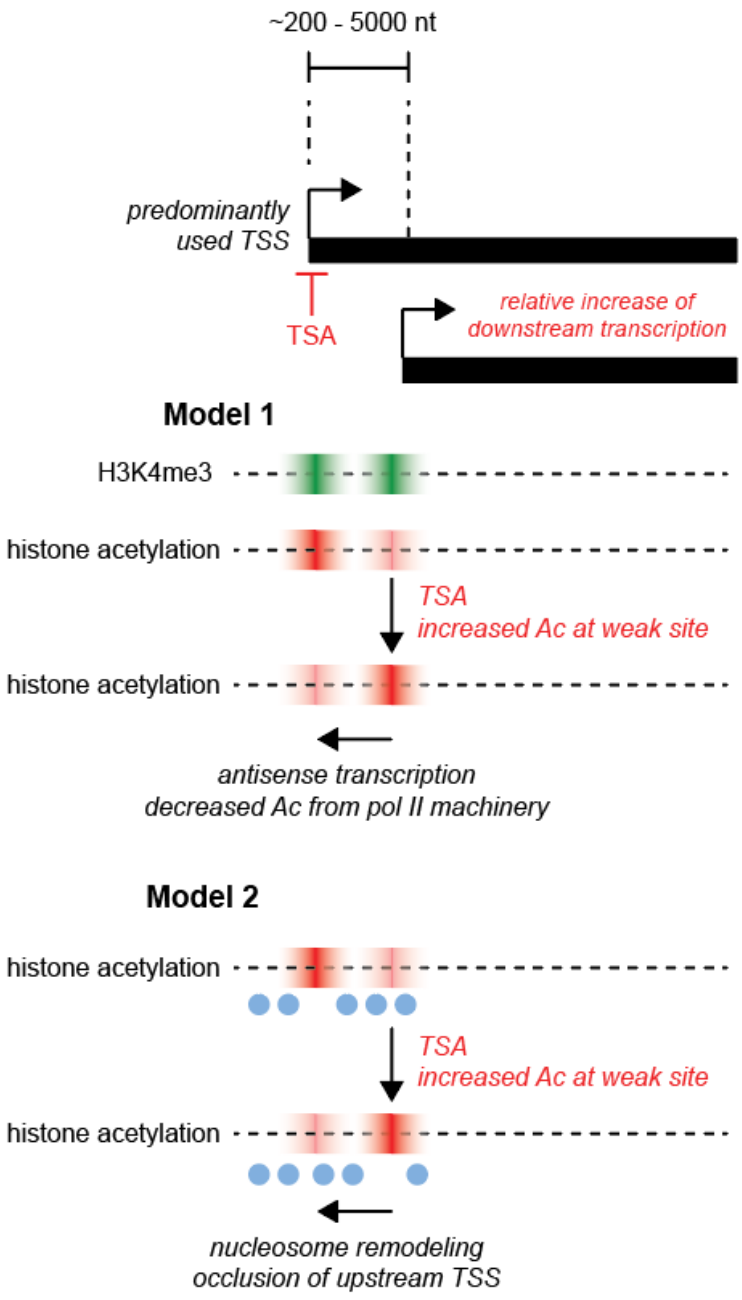

Figure 5.7: Two potential models for TSS switching with TSA treatment. 1) Increase of acetylation in gene body results in increased transcriptional activity interfering with upstream TSS. 2) Chromatin remodeling activity at downstream TSS leads to occlusion of upstream TSS by nucleosomes.

Chapter 6

Methods and analyses

6.1 LC-MS analysis of nucleosides

To a solution of s^4U (50 μM) and ammonium bicarbonate (10 mM) was added and amine (e.g.TFEA, 600 mM final). A solution of oxidant (e.g. mCPBA, 10 mM final) was added dropwise to the reaction mixture. After 1 h at 25°C, the reaction was analyzed by reverse-phase LC-MS with a Hypersil GOLD column (Thermo, 3 μm, 160 × 2.1 mm) using chromatography conditions described in Duffy et al. 2015 [80]. Masses were collected using positive ion mode, and extracted ions were identified and integrated using Agilent MassHunter software.

6.2 Nuclear magnetic resonance analysis of nucleobase chemistry

4-thiouracil (4.3 mg, 1 equiv) was dissolved in DMSO-d6, and TFEA (3.4 μl, 1.3 equiv) was added to the solution. After mixing, a solution of $NaIO_4$ in DMSO-d6 (12.3 mg, 1.7 eq) was added to the nucleobase and amine solution, and the reaction was allowed to proceed at 45 °C for 4 h. 1H NMR spectra were processed using the MestReNova software.

6.3 NotI restriction endonuclease assay

An RNA containing a single s^4U nucleotide was in vitro transcribed (IVT) from a synthetic DNA template (see Table 1) strand using T7 RNA polymerase and s^4UTP in

place of UTP for 16 h at 37°C. The reaction mixture was treated with TURBO DNase for 1h at 37°C. The RNA was purified using denaturing PAGE, and the resulting band was extracted by crushing the gel slice and soaking it in extraction buffer (1 mM EDTA, 1 mM DTT, 20 mM Tris, 300 mM NaOAc pH 5.2) at 4°C for 4 h. The supernatant was passed through a 0.45 μM syringe filter, and the RNA was ethanol precipitated and washed with 75% ethanol before resuspension in nuclease-free water.

IVT RNAs were screened for optimal TimeLapse chemistry as follows. RNA (120 ng) was added to a mixture of amine and water. A solution of oxidant was then added dropwise, and the reaction mixture was incubated at the temperature and time chosen for the particular screen. The RNA was then ethanol precipitated and washed three times with 75% ethanol before resuspension in nuclease-free water.

After chemical treatment, IVT RNA (50 ng) was reverse transcribed with SuperScript III according to the manufacturer's directions. The cDNA was PCR amplified for 30 cycles with a fluorescent forward primer, then it was amplified an additional 2 cycles using 1/5 of the previous PCR reaction material with nonlabeled primers. The amplified PCR product was then incubated with NotI HF for 1 h at 37°C. The fluorescent products were visualized using native PAGE followed by scanning with a Typhoon FLA imager, and the proportion of cut product was determined relative to a positive control (with C in the RNA instead of s^4U) using ImageJ.

6.4 Primer-extension assay

IVT RNA containing a single s^4U nucleotide (200 ng RNA) was treated with TimeLapse chemistry and purified as described above. Chemically treated IVT RNA (34

ng) was then annealed to a Cy5 5′ end-labeled primer, and reverse transcription was performed according to manufacturer's instructions using the SmartScribe First Stand cDNA Synthesis kit (15 min). The reaction was then treated with RNase H, and the fluorescent products were visualized using urea PAGE followed by scanning with a Typhoon FLA imager. Full-length and truncated RT products were quantified by densitometry using ImageJ.

6.5 Targeted TimeLapse sequencing

MEF cells were grown at 37 °C in DMEM containing 10% FBS and 1% P/S at approximately 60% confluence, and the media was replaced with media supplemented with s^4U (700 μM). After 2 h, the cells were rinsed with PBS, resuspended in TRIzol reagent, and stored overnight at −80°C. Following chloroform extraction, total RNA was ethanol precipitated including 1 mM DTT to prevent oxidation of the s^4U RNA and washed with 75% ethanol. Total RNA was resuspended and treated with TURBO DNase, then extracted with acidic phenol:chloroform:isoamyl alcohol and ethanol precipitated and washed as described above. 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 dropwise, and the reaction mixture was incubated for 1 h 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 before centrifugation (>10,000 r.p.m., 30 min, 4°C) to precipitate remaining periodate. The RNA in the supernatant was then ethanol precipitated and washed three times with 75% ethanol before resuspension in nuclease-free water. The chemically treated RNAs were

then reverse transcribed using a mixture of mouse Actb- and Gapdh-specific mRNA RT primers (see Table 1). The resulting cDNA was then amplified with Phusion polymerase using corresponding forward PCR primers to produce PCR amplicons approximately 150 nt in length. An Illumina sequencing library was constructed using the Illumina TruSeq Index adapters. Paired-end 75 bp sequencing was performed on an Illumina HiSeq 2500 instrument. Sequencing reads were trimmed to remove adaptor sequences and aligned to the mouse genome using Bowtie2 [173]. Aligned reads were parsed to identify mutations at each nucleotide position in the Actb and Gapdh mRNAs using a published software package [90]. Raw mutation probabilities were determined by dividing the number of recorded mutation events by the number of reads at that position. Mutation probabilities were normalized to appropriate control samples and filtered by read depth (only positions with depth >3,000 were included in analyses). Analyses and figure plot generation were performed in R using the tidyverse, corrplot, and multiplot packages [174, 175]. The enrichment in mutation rates was tested for significance using a two-sided Wilcoxon test. Targeted sequencing was performed in duplicate using biologically distinct samples.

Targeted TimeLapse-seq of K562 RNA was performed similarly with the following exceptions. Cells were grown at 37°C in RPMI containing 10% FBS and 1% P/S. At approximately 50% confluence, the media was supplemented with a range of s^4U concentrations (10–40 μM) for 1 h. Total RNA was isolated and chemically treated as described previously. The chemically treated RNAs were then reverse transcribed using a mixture of human MYC-specific mRNA RT primers (see Table 1). A targeted sequencing library was prepared and analyzed as described above.

6.6 Transcriptome-wide TimeLapse-seq

For experiments described in Chapter 2, MEF cells were grown at 37 °C in DMEM containing 10% FBS and 1% P/S. At approximately 60% confluence, the media was replaced and supplemented with s^4U (1 mM). The cells were incubated at 37°C for 1 h, at which point total RNA was isolated and chemically treated as described in the targeted sequencing section. For heat shock analyses, at approximately 60% confluence, the media was replaced and supplemented with s^4U (1 mM), and heat-shocked cells were incubated at 42°C for 1 h. RNA was prepared as described for the Targeted TimeLapse-seq libraries. For each sample, 10 ng of total RNA was used to construct a sequencing library using the Clontech SMARTer Stranded Total RNA-Seq kit (Pico Input) with ribosomal cDNA depletion. Paired-end 100 bp sequencing was performed on an Illumina HiSeq 4000 instrument. TimeLapse-seq was performed in duplicate using biologically distinct samples for experimental samples both with and without heat shock. Raw and processed sequencing data have been submitted to the GEO database.

For experiments performed in Chapters 4 and 5, the updated handling protocol was utilized upstream of TimeLapse chemistry. s^4U metabolic labeling times and concentrations are noted in the results section. After TimeLapse chemical treatment, RNA was isolated using an equal volume of RNAclean beads according to manufacturer's instructions, then incubated in reducing buffer (10 mM DTT, 100 mM NaCl, 10 mM Tris pH 7.4, 1 mM EDTA) for 30 min at 37°C. RNA was then again isolated using an equal volume of RNAclean beads.

6.7 Cell viability

MEF cells were grown at 37°C in DMEM containing 10% FBS and 1% P/S. Cells were plated at 10^6 cells/mL in a 96-well microtiter plate and allowed to recover overnight. Cells were then treated in triplicate with increasing concentrations of s^4U (0–1 mM) for 1 h, and the ATCC MTT Cell Proliferation Assay kit was used according to manufacturer's instructions to assess cell viability.

6.8 Standard isolation of RNA

After metabolic labeling of 293T cells in a 10 cm dish, media was aspirated and cells were rinsed once with PBS. 1 ml of Trizol was then added to the cell culture plate and the cell lysate was resuspended until soluble, transferred to a 1.5 ml Eppendorf tube and stored at -80°C. After thawing the trizol sample, 200 ul chloroform was added and the sample was shaken vigorously for 15 seconds and transferred to a heavy phase lock tube. The sample was incubated for 3 min at RT, then centrifuged at 4°C for 10 min at 12,000 x g. The aqueous phase was transferred to a fresh 1.5 ml Eppendorf tube, an equal volume of isopropanol plus 100 uM DTT was added and samples were mixed by inversion 10 times. The samples were incubated in the dark for 10 minutes, then precipitated in a centrifuge at 4°C for 20 min at 22,000 x g. The pellet was washed twice with 75% ethanol, then dried and resuspended in DEPC treated water. The sample was then treated with Turbo DNAse according to manufacturer instructions, then purified by ethanol precipitation.

6.9 Updated isolation of RNA to reduce thiol-specific read loss

This protocol follows the standard isolation of RNA with the following modifications: 1. after metabolic labeling, the cell plate was rinsed once with ice cold PBS. 5 ml ice cold PBS was added to the plate, and cells were removed using a cell scraper. The cell suspension was transferred to a 15 ml lobind conical tube on ice, and the plates were rinsed with 5 ml of ice cold PBS to collect remaining cells. The cells were pelleted at 4°C for 5 min at 300 x g, and the supernatant was aspirated. The pellet was then resuspended in 1 ml of Trizol. 2. After Turbo DNAse treatment, RNA was purified using an equal volume of RNAclean beads according to manufacturer instructions.

6.10 K562 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. 50 μg of total RNA was subjected to MTS chemistry, followed by biotinylation and streptavidin enrichment essentially as previously described [80] with the following modification: after SAV beads were washed three times with high-salt wash buffer (1 M NaCl, 100 mM Tris pH 7.4, 10 mM EDTA, 0.05% Tween), beads were incubated in TE buffer (10 mM Tris pH 7.4, 1 mM EDTA) at 55°C for 15 min, followed by two washes with prewarmed 55°C TE buffer. 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. Enriched RNA and input RNA were chemically treated as

previously described. 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. For each sample, all enriched material or 10 ng of total RNA input was used to construct a sequencing library using the Clontech SMARTer Stranded Total RNA-Seq kit (Pico Input) with ribosomal cDNA depletion. Paired-end 150 bp sequencing was performed on an Illumina HiSeq 4000 instrument. TimeLapse-seq was performed in duplicate using biologically distinct samples for experimental samples. Raw and processed sequencing data have been submitted to the GEO database.

6.11 TSA treated 293T TT-TimeLapse-seq

293T cells were grown at 37°C in DMEM containing 10% FBS and 1% P/S. Cells were then treated with Trichostatin A (400 nM) for 10 min, 30 min, or 6 h, followed by treatment with 1 mM s^4U for 5 min. RNA isolation and DNAse treatment was performed as described in the updated isolation protocol. After DNAse treatment, 293T RNA was mixed with a 4% spike-in of s^4U-treated S2 RNA for normalization. MTS enrichment, TimeLapse chemistry, and sequencing library preparation were then performed as described above.

6.12 Samples for TimeLapse-seq of K562 mRNA

K562 cells were grown as previously described. At approximately 50% confluence, the media was supplemented with s^4U (100 μM). The cells were incubated at

37°C for 4 h, at which point total RNA was isolated using the RNeasy mini kit with the following modifications: buffers RLT and RPE were supplemented with 1% final 2-mercaptoethanol (BME); an additional 80% EtOH wash was performed after the RPE step; and the column was spun at maximum speed for 5 min to dry before elution with water. The isolated RNA was then chemically treated and purified as previously described. For each sample, 10 ng of total RNA was used to construct a sequencing library using the Clontech SMARTer Stranded Total RNA-Seq kit (Pico Input) with ribosomal cDNA depletion. Paired-end 150 bp sequencing was performed on an Illumina HiSeq 4000 instrument. TimeLapse-seq was performed in duplicate using biologically distinct samples for experimental samples. Raw and processed sequencing data have been submitted to the GEO database.

6.13 Samples for TimeLapse-seq in miRNA

Cells expressing WT HSUR1 or empty-vector were treated in duplicate with s^4U for 6 h, 12 h, or untreated, then small RNAs were gel purified. 3' and 5' adapters were ligated to purified small RNAs, and the ligated products were treated with TimeLapse chemistry as described previously. Excess oxidant was then removed by addition of 300 mM KCl and 300 mM sodium acetate, followed by centrifugation at 4°C for 30 min at 20,000 x g. The supernatant was transferred to a fresh 1.5 ml LoBind Eppendorf tube, and was then ethanol precipitated and washed twice with 75% EtOH. The pellet was dried, resuspended in nuclease-free water, and SR primer was reannealed prior to reverse transcription using SmartScribe RT. The remainder of the sequencing library protocol was prepared according to the NEBNext Small RNA Library prep kit manual.

6.14 Transcriptional inhibition

K562 cells were grown as described above. At approximately 50% confluence, cells were treated in duplicate with actinomycin D (2 μg/mL final) for 30 min, 1 h, 3 h, 5 h, and 9 h, or left untreated. Total RNA isolation and genomic DNA depletion were then performed as previously described. RT was performed using the SuperScript VILO cDNA synthesis kit, and qPCR was performed using primers specific to ACTB, DHX9, and ASXL1. qPCR ct values for DHX9 and ASXL1 were then averaged and normalized to those of ACTB for each timepoint. The normalized fraction remaining was estimated for each primer pair by dividing the relative abundance of each timepoint by the relative abundance at t = 0.

6.15 Start-seq

Start-seq was performed as described in Nechaev et al. [165] using the same RNA isolated for TT-seq analyses. Start RNA peaks were called using TSSCall with default settings with a read threshold of 15. Locations of called TSSs were input as exon features into DEXSeq [176] for differential expression analysis, and genes with significant changes in TSS uasge ($P < 10^{-10}$) were inspected manually for transcripts with upstream TSS suppression.

6.16 ChIP-qPCR

30 million 293T cells were treated with trichostatin A (400 nM) for indicated amount of time ranging from 15 min to 6 h. Media was aspirated and plates were rinsed

with PBS. Cells were fixed with 1% formaldehyde and crosslinked with shaking for 5 min. Formaldehyde was quenched with 125 mM glycine with shaking for 5 min, followed by 2 rinses with cold PBS. Cells were scraped from plates in 10 mL cold PBS and pelleted in a centrifuge. Cells were lysed 4 mL in ChIP Lysis Buffer (50 mM HEPES pH 7.9, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100), incubated on ice for 10 min, and pelleted at 4^{o}C for 5 min at 4,000 RPM. Supernatant was removed and the pellet was washed through resuspension and pelleting twice in 3 mL ChIP Wash Buffer (10 mM Tris-Cl pH 8.1, 200 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0). The pellet was gently washed without resuspension twice with 1.5 mL ChIP Shearing Buffer (0.1% SDS, 1 mM EDTA, 10 mM pH 8.1). The pellet was then resuspended in 2 mL ChIP Shearing Buffer and sheared in a 1 mL Covaris tube with the following settings: 140 W, 5% duty cycle, 200 burst/cycle, 12 min. Sheared fractions were combined and were mixed with 230 ul 10% Triton X-100 and 69 ul 5 M NaCl, then spun in 1.5 mL Eppendorf tubes at max speed for 10 min. To the supernatant was added either H4ac or pan-Kac antibody, except for 100 uL reserved for input. The IP was incubated at 4°C overnight, then added to 30 ul Protein G DynaBeads pre-equilibrated with 0.5 ml ChIP Shearing Buffer and incubated at 4°C for 1.5 h. The beads were washed in 1 mL for 5 min for each of the following steps: 2x ChIP Low Salt Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Hepes-KOH pH 7.9, 150 mM NaCl), 2x High Salt Buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Hepes-KOH pH 7.9, 500 mM NaCl), 1x LiCl Buffer (100 mM Tris-HCl pH 7.5, 0.5 M LiCl, 1% NP-40, 1% Sodium deoxycholate), 1x TE buffer. Beads were resuspended in 100 ul ChIP PK buffer (20 mM HEPES pH 7.9, 1 mM EDTA, 0.5% SDS), to which 2 ul proteinase K was added

and samples were incubated at 50°C for 30 min with shaking. Beads were captured with a magnet and supernatant was collected, to which was added 3 ul 5M NaCl and 0.5 ul RNase A, and samples were incubated overnight at 65°C. 1.5 ul proteinase K was added and samples were incubated for 1 h at 50°C with shaking. DNA was purified using a Qiaquick PCR purification kit. qPCR was then performed as previously described.

6.17 Sequencing alignment and mutational analysis.

Reads were filtered for unique sequences using FastUniq [177], trimmed using cutadapt [178] to remove Illumina adaptor 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 [179], 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 [180] 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. 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 3 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 misidentified as TimeLapse mutations), we used the following two strategies. First, we identified T-to-C SNP sites in control samples using bcftools [181] with default options and excluded these sites from our analysis. Second, we compiled locations 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) [182] was used, and the tracks were normalized using factors derived from RNA-seq analyses using values from DESeq2 (estimateSizeFactors) [183]. Tracks were converted to binary format (toTDF, IGVtools) and visualized in IGV [184].

6.18 Secondary structure analysis

Aligned reads from the 4 h K562 TimeLapse-seq experiment overlapping the 5′ stem loop of 7SK were extracted using samtools. A Python script developed for analyses of chemical probing data (RTEventsCounter [90]) was used to calculate the U-to-C mutation frequency for each uridine nucleotide. These frequencies were normalized by

subtracting mutation frequencies of control samples that were not subjected to TimeLapse chemistry. The frequencies of mutations at each position were binned and mapped onto a conformational model of this region of human 7SK [185]. Each nucleotide was classified as either single stranded or basepaired. A two-sided Wilcoxon test was used to determine the significance of differences between mutation rates of the basepaired and single-stranded nucleotides.

6.19 Estimation of the fraction of new transcripts and transcript half-lives

Two different models were used to examine the mutation distribution in TimeLapse-seq data set: a simpler Poisson model (which does not take into account the uridine content of different reads) and a binomial model that does take the number of uridines into account. We obtained consistent results from both models. For the simpler Poisson model, for each sample (s_j), the distribution of T-to-C mutations (Y_i) was determined in each read, and the reads were grouped based on the transcripts to which they map. A negative control sample (no s^4U treatment) was used to estimate the background rate of read pairs containing T-to-C mutations that map to each transcript. These frequencies depended on the cell line used (MEF samples required higher s^4U treatment to obtain similar levels of mutations compared to K562 cells) as well as the sequencing experiment (different samples led to different background rates independent of chemistry or s^4U treatment). The mutation rate and fraction of new transcripts was modeled as a two-component mixture of Poisson distributions with probability mass function:

$$f(y \mid \lambda_o, \lambda_n, \theta_n) = \theta_n \,\text{Poisson}(y;\ \lambda_n) + (1-\theta_n)\,\text{Poisson}(y;\ \lambda_o)$$

where θ_n is the fraction of new transcripts, λ_O is the rate of background mutations (determined from $-s^4U$ controls), λ_n is the rate of mutations found in new transcripts, and y_i is the number of passing T-to-C mutations found in read i. Reasonable estimates of these values could be approximated by examining the mutation rates in fast turnover RNAs such as introns. To obtain more objective estimates of the global parameters λ_O and λ_n while allowing for low levels of transcript-to-transcript variability, we used a Bayesian hierarchical modeling approach using RStan software (Version 2.16.2 [186]) that uses no-U-turn Markov Chain Monte Carlo (MCMC) sampling. To estimate a global mean and s.d. for λ_O and λ_n, we used weakly informative priors (see below). We estimated gene-specific rates by drawing from the global mean and s.d., with a mixing rate with an uninformative prior ($\theta_n \sim$Uniform(0,1)) where the mixing rate (θ_n) estimates the fraction of each transcript that was new:

Global parameters:

$$\lambda_{\circ,\mu} \sim \text{Normal}(\mu = 0, \sigma = 1)$$

$$\lambda_{n,\mu} \sim \text{Normal}(\mu = 0, \sigma = 10)$$

$$\lambda_{\circ,\sigma} \sim \text{Normal}(\mu = 0, \sigma = 10)$$

$$\lambda_{n,\sigma} \sim \text{Normal}(\mu = 0, \sigma = 10)$$

$$\lambda \sim \text{Normal}(\mu = \lambda_{\circ,\mu}, \sigma = \lambda_{\circ,\sigma})$$

$$I_s = \begin{cases} 0 & s \in \text{controls} \\ 1 & \text{otherwise} \end{cases}$$

$$g \in \{1, 2, ..., n_{\text{genes}}\}$$

Priors:

$^{\lambda}n,g \sim \text{Normal}(\mu = \lambda_{n,\mu}, \sigma = \lambda_{n,\sigma})$

$^{\lambda}{}_{\circ},g \sim \text{Normal}(\mu = \lambda_{\circ,\sigma}, \sigma = \lambda_{\circ,\sigma})$

for read $i \in \{1, 2, ..., n_g\}$:

$$f_g(y_g \mid \theta_{ng}, \lambda_{\circ g}, \lambda_{ng})$$

$$= \prod_{i=1}^{n_g} (Is\theta_{ng}\,\text{Poisson}(y_i|\lambda_{ng}) + (1 - Is\theta_n)\,\text{Poisson}(y_i|\lambda_{\circ g}))$$

Attempts to model entire TimeLapse-seq data sets using this approach were computationally challenging, but we found that consistent results were obtained using 20 representative transcripts from each sample. The majority of these transcripts were chosen randomly from all reasonably expressed transcripts (>200 reads), but we included few transcripts that were hand chosen to ensure the modeling included both fast and slow turnover RNAs such as Myc and Actb. The results using 20 transcripts were consistent with results from 200 transcripts. In the case of the MEF samples shown in Figure 2.5, the λo was estimated as 0.07 mutations/read (50% credible interval 0.062-0.074), and λn was estimated as 2.3 mutations/read (2.298 mutation/read, 50% CI 2.10–2.30 for heat shock; 2.288 mutation/read, 50% CI 1.90-2.29 for untreated).

Once these global parameters were determined, they were used to estimate the fraction of new transcripts (θnew), using expectation maximization by minimizing the log likelihood using the nlm function in the MASS package in R:

$$-\sum_{i=1}^{n_g} \log(f(y;\theta_n, \lambda_\circ, \lambda_n))$$

The 95% Wald confidence interval was calculated using the Hessian (nlm option hessian = TRUE), to calculate:

$$\hat{\theta}_{\text{new}} \pm z_{0.975} \times \text{std err}\left(\hat{\theta}_{\text{new}}\right)$$

To ensure the mutations were both s^4U-treatment and TimeLapse-chemistry dependent, we only included transcripts where there was sufficient data (reads > 100 counts in at least two samples), and where the fit converged ($-0.05 < \theta_n < 1.05$; hessian > 1,000). The inferred new read counts were determined by multiplying the estimated fraction of new transcripts by the total RNA-seq transcript count. Correlations between replicates were determined using the log10-transformed counts. While the reproducibility of the data was generally high when all converged transcripts were included (Pearson's r > 0.91), filtering for transcripts with at least 75 inferred new reads provided slightly more reproducible results (n = 3,603, r = 0.934), and this filter was used for further analysis.

To account for differences in the number of uridine residues in each read pair, an alternative model was used based on the binomial distribution. Specifically, the data were modeled as mixture of two binomial distributions:

$$f(y \mid \theta_n, p_o, p_n) = \theta_n \text{Binom}(y; p_n, n_{u,i}) + (1-\theta_n)\,\text{Binom}(y; p_o, n_{u,i})$$

where p_o, p_n are the probabilities of mutation at each uridine nucleotide for old and new transcripts, and nu is the number of uridines observed for read i. To determine the global mutation rate, we used Bayesian hierarchical modeling as described above for the

Poisson model but using a mixture of binomial distributions. From this analysis, we estimate the background mutation rate (p_o) to be 0.0012 mutations/uridine (50% CI 0.00121, 0.00123) and the mutation rate for new reads (pn) to be 0.0332 mutations/uridine (50% CI 0.0329, 0.0335). In other words, ~0.1% of Us are mutated to C in pre-existing reads, and in new reads ~3% of Us are mutated to C. Using these global parameters, the distributions of individual genes were fit with nlm similarly to what is described above, except by minimizing the log likelihood of the binomial model instead:

$$-\sum_{i=1}^{n_g} \log(f(y;\theta_{n,}p_o,p_n,n_{u,i}))$$

In addition to computing the confidence interval using the hessian, we also examined the quality of the fit by plotting the observed frequency of mutations in each replicate in the TimeLapse data (gray points in distribution plots) to a simulated distribution of the expected new and old reads based on the binomial model (Fig. 2.5). Estimates of the fraction new were highly similar between those determined using the binomial model and the Poisson model.

In addition to the nlm modeling approach, we also implemented a simpler binomial approach of separating new and old reads based on the fraction of T-to-C mutations on a per-read basis. First, I filtered a non-treated and s^4U-treated dataset for reads representative of new and old RNAs: intronic reads in highly expressed genes with minimal overlapping transcript isoforms. I then calculated the fraction of mutations per read (Fmut) by dividing T-to-C mutations by the number of Ts per read (muts / Nobs). I then performed a sensitivity and specificity analysis for a range of possible values of

Fmut to determine the threshold of old vs. new reads, which is the value that maximizes sensitivity and specificity according to the following equations:

Sensitivity = [True Positive] / ([True Positive] + [False Negative])

Specificity = [True Negative] / ([True Negative] + [False Positive])

I then assigned new reads and old reads in a TimeLapse-seq experiment as those with Fmut above or below the determined threshold, respectively. The fraction new estimates determined using this Fmut cutoff approach correlate reasonably well to those determined through nlm modeling (R = 0.98)

To account for any specific loss of transcripts that might arise from biased loss of s^4U-RNA transcripts independent of TimeLapse chemistry, or TimeLapse-dependent loss due to reverse-transcription termination, we developed a means of estimating the loss of fast-turnover transcripts in the data. This correction was only used when estimating transcript half-lives after observing a modest, but statistically significant loss of reads from high turnover RNAs. To estimate the fraction of new reads missing, we used the R package nlm to fit the equation:

$$s_y N_y \left(1+\theta_n\left(\frac{x}{1-x}\right)\right)=s_o N_o$$

where s_y and s_o are scale factors that adjust for library sizes determined using DESeq2 with the total (RNA-seq) transcript counts for the experimental sample and control, respectively; N_y and N_o are the counts for each transcript; and θn is the unadjusted fraction new of each transcript. This equation was fit using transcripts where $0.8 < \theta_n$ for K562 RNA, but $0.5 < \theta_n$ in the case of MEF RNA (the shorter s^4U treatment lead to fewer transcripts with high θn, so the threshold was lowered to increase the number of

transcripts). For example, the adjustment factor determined for chemistry-induced dropout was ~5% in our MEF data (i.e., x = 0.05 in the equation above, which leads a transcript with 75% new reads to be adjusted to 79% and a transcript with 25% new reads would be adjusted to 26% new reads).

The transcript half-lives were determined using the adjusted fraction of new RNA assuming a simple exponential model of their kinetics. The half-life values were compared to similar reports and the r^2 determined using the lm function in R.

6.20 Gene ontology analysis

GO analysis from the PANTHER database (version 12.0) [187] was performed using a statistical over-representation test (default parameters) on the complete biological process annotation set using the top 10% slow or top 10% fast turnovers RNAs in our 1 h MEF TimeLapse-seq data as determined by the half-life analyses described above.

6.21 Differential expression analysis.

Differential expression analysis was performed using DESeq2. To examine the inferred differences in the new transcript pool based on TimeLapse mutations, we used the unadjusted estimates of the fraction of new RNA to infer the number of counts resulting from new transcripts as described above. As TimeLaspse-seq data are internally controlled, we used the size factors determined from total counts to scale each data set (i.e., we ran DESeq2 on the total RNA-seq data and used the sizeFactors function to scale the inferred new RNA counts to the RNA-seq-determined values) with default conditions including the Benjamini–Hochberg [188] adjusted P value (padj in text). RNA-seq

analysis was performed on all reads (i.e., reads that had zero or more T-to-C mutations) using DESeq2 with default parameters.

6.22 Differential exon usage analysis

Differential exon usage was performed using Ensemble GRCh38 exon annotations for total TimeLapse-seq. For differential TSS usage, a custom annotation was generated from the output of TSScall for Start-seq data. Called TSSs were assigned to genes using bedtools (intersect) [189], and a flattened GTF was generated with each TSS assigned as an exon feature. For each sequencing format, DEXseq was performed (testForDEU), and switching start sites were identified by examining the most significant calls of differential usage.

6.23 Estimation of contaminating reads in TT-TimeLapse-seq

Reads from TT-TimeLapse-seq were processed and analyzed as for TimeLapse-seq. Junction-containing reads were determined from the presence of "N" characters in the CIGAR string in the aligned bam file using bamtools (version 2.3, https://hcc-docs.unl.edu/display/HCCDOC/BamTools). The levels of contaminating reads were estimated by assuming the contaminating reads have the same ratios as RNA-seq data, and that reads with three or more mutations constitute the true ratio of reads. We use of reads with three or more mutations as true positives, because the probability of a read containing three or more mutations without s^4U is $<10^{-5}$. We used the fraction of intron- or junction-containing reads for the RNA-seq data (r_o), the total in the true positive population (r_{tp}), and the total for each population (r_x). In each analysis, we only

considered reads that had nonzero ratios and ratios that were less than one. The fraction of reads from contamination (c_x) was then estimated:

$$c_x = \frac{r_{tp} - r_o}{r_x - r_o}$$

For comparisons with the TT-TimeLapse-seq data presented here, the data from Schwalb et al. [87], (SRR4000390, SRR4000391 and SRR4000397) were aligned and processed using the same pipeline described for TimeLapse-seq. For this comparison, we reprocessed our TT-TimeLapse-seq data using only 75 nt of each read, and this trimming was performed on fastq files before alignment. This step was performed because the probability of a sequencing read containing a splice junction or being an intron-only read is dependent on the read length. Otherwise, all processing was handled equally between data sets.

6.24 Transcriptome-wide estimates of intronic read loss

TimeLapse-seq or literature reads [122] were aligned to the GRCh38 or GRCm38 genome and annotated transcriptome with HISAT2 and parameters described in the TimeLapse-seq alignment methods section. Intronic vs. exonic features were counted using HTSEQ and the fraction of intronic reads per gene was determined using the following equation:

$$Frac_{intronic} = (Gene - Mature) / Gene$$

6.25 Estimating rates of synthesis and decay in DCP2 KO cells

Old and new RNAs were assigned based on the fraction of mutations per read approach, with a cutoff determined through maximization of sensitivity and specificity from intronic read distributions. New and old read counts were used to calculate relative rates of synthesis and decay using the following equations:

$$kdeg = -log(1 - fnew)/t$$

$$ksynth = [RNA] * kdeg$$

6.26 Clustering analysis of TT-TimeLapse-seq data

TT-TimeLapse-seq alignment and mutation calling was performed as described previously. Inferred new reads were identified by the fraction of mutations approach as described previously. Inferred new reads were filtered for genes with a minimum of 10 reads and a maximum of 10,000 reads per sample (n = 12,499). The optimal number of clusters was determined by minimizing the sum of squares for *n* clusters, followed by k-means clustering using the Pheatmap package in R, using correlation distance for samples and Euclidean distance for genes.

Table 1: Primers used

Primers for in-vitro transcription (IVT) and restriction enzyme (RE) PCR		
IVT template	CTTGCGTTTCTCTCGTCCTTctgtcttgctgttgttcG CGACCGCcctgtGCGGTTGTGTCTTTTGTCCT GCCTATAGTGAGTCGTATTAATTTC	
IVT positive control template	CTTGCGTTTCTCTCGTCCTTctgtcttgctgttgttcG CGGCCGCcctgtGCGGTTGTGTCTTTTGTCCT GCCTATAGTGAGTCGTATTAATTTC	
IVT T7 promoter	GAAATTAATACGACTCACTATA	
RE RT and reverse primer	CTTGCGTTTCTCTCGTCCTT	
RE fluorescent forward primer	/5Cy5/AGGACAAAAGACACAACCGC	
RE forward primer	AGGACAAAAGACACAACCGC	
Primer extension RT	/5Cy5/CTTGCGTTTCTCTCGTCCTT	
Primers for targeted TimeLapse-seq		
Name	**Forward**	**Reverse**
mActB1	CTACACGACGCTCTTCC GATCTNNNNN ACTGAGCTGCGTTTTAC ACCC	CAGACGTGTGCTCTTCC GATCTTCCTGAGTCAAA AGCGCCAAAAC
mActB2	CTACACGACGCTCTTCC GATCTNNNNN AATTTCTGAATGGCCCA GGTCT	CAGACGTGTGCTCTTCC GATCTGGTGTGGCACTT TTATTGGTCTCAAGTC
mActB3	CTACACGACGCTCTTCC GATCTNNNNN ATGGTGGGAATGGGTCA GAAGG	CAGACGTGTGCTCTTCC GATCTGCCACACGCAGC TCATTGTAG
mActB4	CTACACGACGCTCTTCC GATCTNNNNN TGAAGTGTGACGTTGAC ATCCG	CAGACGTGTGCTCTTCC GATCTGAGGAGCAATGA TCTTGATCTTCATGG
mGapdh1	CTACACGACGCTCTTCC GATCTNNNNN TCCGTCGTGGATCTGAC GTG	CAGACGTGTGCTCTTCC GATCTCATCGAAGGTGG AAGAGTGGG
mGapdh2	CTACACGACGCTCTTCC GATCTNNNNN CTCTTCCACCTTCGATG CCG	CAGACGTGTGCTCTTCC GATCTGGTGGGTGGTCC AGGGTTTC
mGapdh3	CTACACGACGCTCTTCC GATCTNNNNN AGGACACTGAGCAAGA GAGGC	CAGACGTGTGCTCTTCC GATCTGTGGGTGCAGCG AACTTTATTG
mGapdh4	CTACACGACGCTCTTCC GATCTNNNNN CAGCAATGCATCCTGCA CCA	CAGACGTGTGCTCTTCC GATCTACAGCTTTCCAG AGGGGCC

hMYC1	CTACACGACGCTCTTCCGATCTNNNNNCTTGGCGGGAAAAAGAACGG	CAGACGTGTGCTCTTCCGATCTTATTCGCTCCGGATCTCCCT
hMYC2	CTACACGACGCTCTTCCGATCTNNNNNGCATCCACGAAACTTTGCCC	CAGACGTGTGCTCTTCCGATCTCCTTTCAGAGAAGCGGGTCC
hMYC3	CTACACGACGCTCTTCCGATCTNNNNNTACTGCGACGAGGAGGAGAA	CAGACGTGTGCTCTTCCGATCTCGAAGGGAGAAGGGTGTGAC
hMYC4	CTACACGACGCTCTTCCGATCTNNNNNCAGGACTGTATGTGGAGCGG	CAGACGTGTGCTCTTCCGATCTGGTACAAGCTGGAGGTGGAG
PCR Amplification primers	CAGACGTGTGCTCTTCCGATC	CTACACGACGCTCTTCCGATCT
qPCR primers for TimeLapse-seq validation		
Name	**Forward**	**Reverse**
Rsrp1 (a, heat shock)	AAGTACAGGCGCTACTCACG	GACGGCGACTTGTAGTACCT
Rsrp1 (b, heat shock)	AAGATCCAGAACCAGGTCGC	GCAGTGGCTTTGCTACGGAA
Rsrp1 (c, heat shock)	TCCTTGGACAGCTAGGGGAT	CACGAATACCCGACTCCTGT
Rsrp1 (d, heat shock)	GAGGGGTTTGTGTCCAGCAT	ACCTCAACCATGAACGTCCC
Hist1h1d (a, heat shock)	AAGAAGGCAGCAAAGAGTCCA	AGATTTTCAAAGCAGGACGCA
Hist1h1d (b, heat shock)	AAGCCTAAGAAGGCGACTGG	TCTACTTCTTGCGAGGGGCA
ACTB (normalization)	GGCATGGGTCAGAAGGATT	CACACGCAGCTCATTGTAGA
DHX9 (validation and dropout)	CCGATTCCTCCATGCGAGTT	TCTGGCCTTCTACCGAGACA
ASXL1 (Pair 1)	TCGGATGCTCCAATGACACC	CCTTCTGCCTCTATGACCTGC
ASXL1 (Pair 2)	ACCAGGCCCCTTCATCTTAAT	TCCCAAGCTTACAGCAGGTT
ASXL1 (Pair 3)	GAAGCCCCGGCTTGAAGAT	TGTGGCTTTTCGGTGTGAAC
ASXL1 (Pair 4)	ATCCTCACCGACTGATTGCC	CATGAGCCACCAAGCCCTAA
ASXL1 (Pair 5)	TGCATTGCCTGGGGATTTGA	CTCGAGATGGCACAGTCCAG
CDKN1B pre-mRNA (dropout)	GTCTTAGGTGTTCAGTGCTACC	CAGAAAGTGCATTCTCCGCC
MIR17HG (dropout)	GACTCCTGACAAAATGCAGCC	TCTGAAGTCTCAAGTGGGCAT
EGR1 (dropout)	GCCACCACCTCATACCCATC	CTGTCATGTCCGAAAGCCCT

MYC (dropout)	CATACATCCTGTCCGTCCAAG	GAGTTCCGTAGCTGTTCAAGT
Primers for ChIP and switching PCR		
Name	**Forward**	**Reverse**
ZNF460 upstream (ChIP and enriched PCR)	TGGGTGGCTATTTTAACTGCTCA	CCCGCGACTTCTCGCTT
ZNF460 downstream (ChIP and enriched PCR)	TTACGCCCCTTCTTCGCGTC	CCGGTGGGACACCTCGTT
AAAS upstream (ChIP)	GAACAACCCCAGAGAGCACAT	ATCTAGCCCGGGAACCGAG
AAAS downstream (ChIP)	TTGTCCTTCCCAGGAACACC	TGGGGACAGTATCTGGGGAG
ARHGEF7 upstream (ChIP)	CGTGGCTCATCACTCTGGG	AGAGGACCACCCCATCCTTC
ARHGEF7 downstream (ChIP)	GTGATCAGGCAAAGTCGTGC	AGAGAAAGGAGTGCGGCTG
NARF upstream (ChIP)	TCACCGACTCCCTGGTGAA	CCGGTTTCCGCTTCTCGT
NARF downstream (ChIP)	CTCTTCAAGGCGCCCCC	ACAAGTTCAGGGAGAACCGC
SREBF1 upstream (ChIP)	GCTGAAGGGTGGCTCGTC	AGGAAGGGCCGTACGAGG
SREBF1 downstream (ChIP)	TTGGAGGTAGCCCTCCACTT	GCTACGCAACTTGAGCAGGA
Intergenic control 1 (ChIP)	CATGTGCTGTCACCCTACCT	GACACATGGGGGTTTCAGGA
Intergenic control 2 (ChIP)	ACATGAGAGGGGTGTCCCTT	ATCTCTCCTTCTGGCCCTGT

Table 2. Analysis of contaminating reads in TT-TimeLapse-seq

Proportions of contaminating reads for TT-TimeLapse-seq replicates by mutation count per read estimated through splice-junction or intron analysis (see methods).

	Est. from junction reads			**Est. from intron-only reads**		
mutations	replicate 1	replicate 2	merged	replicate 1	replicate 2	merged
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	0.033	0.0368	0.0299	0.0271	0.0284
2 mutations	0.0102	0.0077	0.0089	0.0037	0.0054	0.0045

References

1. Lee, T.I. & Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251 (2013).
2. Schoenberg, D.R. & Maquat, L.E. Regulation of cytoplasmic mRNA decay. *Nature Reviews Genetics* **13**, 448-448 (2012).
3. Kim, T.K., Hemberg, M. & Gray, J.M. Enhancer RNAs: A Class of Long Noncoding RNAs Synthesized at Enhancers. *Csh Perspect Biol* **7** (2015).
4. Loeb, J.N., Howell, R.R. & Tomkins, G.M. Turnover of Ribosomal Rna in Rat Liver. *Science* **149**, 1093-& (1965).
5. Wang, Y. et al. Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci U S A* **99**, 5860-5865 (2002).
6. Schwanhausser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
7. Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G. & Lis, J.T. Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Mol Cell* **62**, 63-78 (2016).
8. Yartseva, V. & Giraldez, A.J. The Maternal-to-Zygotic Transition During Vertebrate Development: A Model for Reprogramming. *Curr Top Dev Biol* **113**, 191-232 (2015).
9. Isken, O. & Maquat, L.E. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev* **21**, 1833-1856 (2007).
10. Roeder, R.G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* **21**, 327-335 (1996).
11. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P.A. Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-561 (1989).
12. Akoulitchev, S., Makela, T.P., Weinberg, R.A. & Reinberg, D. Requirement for Tfiih Kinase-Activity in Transcription by Rna-Polymerase-Ii. *Nature* **377**, 557-560 (1995).
13. Conaway, R.C. & Conaway, J.W. An Rna Polymerase-Ii Transcription Factor Has an Associated DNA-Dependent Atpase (Datpase) Activity Strongly Stimulated by the Tata Region of Promoters. *P Natl Acad Sci USA* **86**, 7356-7360 (1989).
14. Donczew, R. & Hahn, S. Mechanistic Differences in Transcription Initiation at TATA-Less and TATA-Containing Promoters. *Mol Cell Biol* **38** (2018).
15. Huisinga, K.L. & Pugh, B.F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Mol Cell* **13**, 573-585 (2004).
16. Donczew, R.W., L.; Erijman, A.; Hahn, S. Two separate roles for the transcription coactivator SAGA and a set of genes redundantly regulated by TFIID and SAGA. *bioRxiv* (2019).
17. Adelman, K. & Lis, J.T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* **13**, 720-731 (2012).
18. Peterlin, B.M. & Price, D.H. Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23**, 297-305 (2006).
19. Yang, Z. et al. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* **19**, 535-545 (2005).

20. Henriques, T. et al. Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. *Mol Cell* **52**, 517-528 (2013).
21. Schroeder, S.C., Schwer, B., Shuman, S. & Bentley, D. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Gene Dev* **14**, 2435-2440 (2000).
22. Komarnitsky, P., Cho, E.J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Gene Dev* **14**, 2452-2460 (2000).
23. Kotovic, K.M., Lockshon, D., Boric, L. & Neugebauer, K.M. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol Cell Biol* **23**, 5768-5779 (2003).
24. Rasmussen, E.B. & Lis, J.T. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc Natl Acad Sci U S A* **90**, 7923-7927 (1993).
25. Henriques, T. et al. Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev* **32**, 26-41 (2018).
26. Sehgal, P.B., Darnell, J.E., Jr. & Tamm, I. The inhibition by DRB (5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole) of hnRNA and mRNA production in HeLa cells. *Cell* **9**, 473-480 (1976).
27. Fuchs, G. et al. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol* **15**, R69 (2014).
28. Duffy, E.E., Canzio, D., Maniatis, T. & Simon, M.D. Solid phase chemistry to covalently and reversibly capture thiolated RNA. *Nucleic Acids Res* **46**, 6996-7005 (2018).
29. de la Mata, M. et al. A slow RNA polymerase II affects alternative splicing in vivo. *Molecular Cell* **12**, 525-532 (2003).
30. Fong, N. et al. Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Gene Dev* **28**, 2663-2676 (2014).
31. Pinto, P.A. et al. RNA polymerase II kinetics in polo polyadenylation signal selection. *Embo J* **30**, 2431-2444 (2011).
32. Liu, X.C. et al. Transcription elongation rate has a tissue-specific impact on alternative cleavage and polyadenylation in Drosophila melanogaster. *Rna* **23**, 1807-1816 (2017).
33. Danko, C.G. et al. Signaling pathways differentially affect RNA polymerase II initiation, pausing, and elongation rate in cells. *Mol Cell* **50**, 212-222 (2013).
34. Connelly, S. & Manley, J.L. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev* **2**, 440-452 (1988).
35. Kim, M. et al. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* **432**, 517-522 (2004).
36. West, S., Gromak, N. & Proudfoot, N.J. Human 5 '-> 3 ' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432**, 522-525 (2004).
37. Fong, N. et al. Effects of Transcription Elongation Rate and Xrn2 Exonuclease Activity on RNA Polymerase II Termination Suggest Widespread Kinetic Competition. *Molecular Cell* **60**, 256-267 (2015).
38. Sheridan, R.M., Fong, N., D'Alessandro, A. & Bentley, D.L. Widespread Backtracking by RNA Pol II Is a Major Effector of Gene Activation, 5 ' Pause Release, Termination, and Transcription Elongation Rate. *Molecular Cell* **73**, 107-+ (2019).
39. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. & Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223 (2009).
40. Pratt, A.J. & MacRae, I.J. The RNA-induced silencing complex: a versatile gene-silencing machine. *J Biol Chem* **284**, 17897-17901 (2009).

41. Sen, G.L. & Blau, H.M. Argonaute 2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. *Nat Cell Biol* **7**, 633-636 (2005).
42. Brown, J.A., Valenstein, M.L., Yario, T.A., Tycowski, K.T. & Steitz, J.A. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MENbeta noncoding RNAs. *Proc Natl Acad Sci U S A* **109**, 19202-19207 (2012).
43. Geisberg, J.V., Moqtaderi, Z., Fan, X., Ozsolak, F. & Struhl, K. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* **156**, 812-824 (2014).
44. Wu, X.B. & Bartel, D.P. Widespread Influence of 3 '-End Structures on Mammalian mRNA Processing and Stability. *Cell* **169**, 905-+ (2017).
45. Heaton, B. et al. Analysis of chimeric mRNAs derived from the STE3 mRNA identifies multiple regions within yeast mRNAs that modulate mRNA decay. *Nucleic Acids Res* **20**, 5365-5373 (1992).
46. Wang, X. et al. N-6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-+ (2014).
47. Doma, M.K. & Parker, R. RNA quality control in eukaryotes. *Cell* **131**, 660-668 (2007).
48. Presnyak, V. et al. Codon Optimality Is a Major Determinant of mRNA Stability. *Cell* **160**, 1111-1124 (2015).
49. Passarge, E. Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *Am J Hum Genet* **31**, 106-115 (1979).
50. Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H. & Tora, L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* **13**, 424 (2012).
51. Pokholok, D.K. et al. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517-527 (2005).
52. Krogan, N.J. et al. Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II. *Mol Cell Biol* **23**, 4207-4218 (2003).
53. Kizer, K.O. et al. A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol Cell Biol* **25**, 3305-3316 (2005).
54. Neri, F. et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-77 (2017).
55. Deal, R.B., Henikoff, J.G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161-1164 (2010).
56. Zheng, Y.P., Thomas, P.M. & Kelleher, N.L. Measurement of acetylation turnover at distinct lysines in human histones identifies long-lived acetylation sites. *Nature Communications* **4** (2013).
57. Zheng, Y.P., Tipton, J.D., Thomas, P.M., Kelleher, N.L. & Sweet, S.M.M. Site-specific human histone H3 methylation stability: fast K4me3 turnover. *Proteomics* **14**, 2190-2199 (2014).
58. Maniatis, T., Goodbourn, S. & Fischer, J.A. Regulation of inducible and tissue-specific gene expression. *Science* **236**, 1237-1245 (1987).
59. Creyghton, M.P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936 (2010).
60. Lee, J.E. et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife* **2** (2013).

61. Luyten, A., Zang, C.Z., Liu, X.S. & Shivdasani, R.A. Active enhancers are delineated de novo during hematopoiesis, with limited lineage fidelity among specified primary blood cells. *Gene Dev* **28**, 1827-1839 (2014).
62. Pham, T.H. et al. Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. *Blood* **119**, e161-171 (2012).
63. Schones, D.E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-898 (2008).
64. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
65. Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D. & Lieb, J.D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* **36**, 900-905 (2004).
66. Becker, P.B. & Workman, J.L. Nucleosome remodeling and epigenetics. *Cold Spring Harb Perspect Biol* **5** (2013).
67. Jimeno-Gonzalez, S., Ceballos-Chavez, M. & Reyes, J.C. A positioned +1 nucleosome enhances promoter-proximal pausing. *Nucleic Acids Res* **43**, 3068-3078 (2015).
68. Spain, M.M.B., K.C.A.; Tsukiyama, T. SWI/SNF coordinates transcriptional activation through Rpd3-mediated histone hypoacetylation during quiescence entry. *bioRxiv* (2018).
69. West, J.A. et al. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat Commun* **5**, 4719 (2014).
70. Lindell, T.J., Weinberg, F., Morris, P.W., Roeder, R.G. & Rutter, W.J. Specific inhibition of nuclear RNA polymerase II by alpha-amanitin. *Science* **170**, 447-449 (1970).
71. Perry, R.P. & Kelley, D.E. Inhibition of RNA synthesis by actinomycin D: characteristic dose-response of different RNA species. *J Cell Physiol* **76**, 127-139 (1970).
72. Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2**, 103-108 (2011).
73. Churchman, L.S. & Weissman, J.S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368-373 (2011).
74. Nojima, T. et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**, 526-540 (2015).
75. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848 (2008).
76. Jao, C.Y. & Salic, A. Exploring RNA transcription and turnover in vivo by using click chemistry. *Proc Natl Acad Sci U S A* **105**, 15779-15784 (2008).
77. Melvin, W.T., Milne, H.B., Slater, A.A., Allen, H.J. & Keir, H.M. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. *Eur J Biochem* **92**, 373-379 (1978).
78. Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R. & Boothroyd, J.C. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol* **23**, 232-237 (2005).
79. Dolken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959-1972 (2008).
80. Duffy, E.E. et al. Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Mol Cell* **59**, 858-866 (2015).
81. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29**, 436-442 (2011).

82. Saunders, N.A. et al. Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. *Embo Mol Med* **4**, 675-684 (2012).
83. Hwang, B., Lee, J.H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* **50** (2018).
84. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res* **25**, 1491-1498 (2015).
85. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494-+ (2018).
86. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141**, 129-141 (2010).
87. Schwalb, B. et al. TT-seq maps the human transient transcriptome. *Science* **352**, 1225-1228 (2016).
88. Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C. & Simon, M.D. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods* **15**, 221-225 (2018).
89. Miller, C. et al. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7**, 458 (2011).
90. Sexton, A.N., Wang, P.Y., Rutenberg-Schoenberg, M. & Simon, M.D. Interpreting Reverse Transcriptase Termination and Mutation Events for Greater Insight into the Chemical Probing of RNA. *Biochemistry* **56**, 4713-4721 (2017).
91. Zhang, C. & Jia, G. Reversible RNA Modification N(1)-methyladenosine (m(1)A) in mRNA and tRNA. *Genomics Proteomics Bioinformatics* **16**, 155-161 (2018).
92. Hauenschild, R. et al. The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence dependent. *Nucleic Acids Res* **43**, 9950-9964 (2015).
93. Peattie, D.A. & Gilbert, W. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci U S A* **77**, 4679-4682 (1980).
94. Walsh, C.P. & Bestor, T.H. Cytosine methylation and mammalian development. *Genes Dev* **13**, 26-34 (1999).
95. Frommer, M. et al. A Genomic Sequencing Protocol That Yields a Positive Display of 5-Methylcytosine Residues in Individual DNA Strands. *P Natl Acad Sci USA* **89**, 1827-1831 (1992).
96. Rueter, S.M., Dawson, T.R. & Emeson, R.B. Regulation of alternative splicing by RNA editing. *Nature* **399**, 75-80 (1999).
97. Licht, K. et al. Inosine induces context-dependent recoding and translational stalling. *Nucleic Acids Research* **47**, 3-14 (2019).
98. Levanon, E.Y. et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* **22**, 1001-1005 (2004).
99. Ziff, E.B. & Fresco, J.R. A method for locating 4-thiouridylate in the primary structure of transfer ribonucleic acids. *Biochemistry* **8**, 3242-3248 (1969).
100. Macmillan, A.M. & Verdine, G.L. Synthesis of Functionally Tethered Oligodeoxynucleotides by the Convertible Nucleoside Approach. *J Org Chem* **55**, 5931-5933 (1990).
101. Allerson, C.R., Chen, S.L. & Verdine, G.L. A chemical method for site-specific modification of RNA: The convertible nucleoside approach. *Journal of the American Chemical Society* **119**, 7423-7433 (1997).
102. Bhatt, D.M. et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279-290 (2012).
103. Menet, J.S., Rodriguez, J., Abruzzi, K.C. & Rosbash, M. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *Elife* **1**, e00011 (2012).

104. Wada, Y. et al. A wave of nascent transcription on activated human genes. *PNAS* **106**, 18357-18361 (2009).
105. Gaidatzis, D., Burger, L., Florescu, M. & Stadler, M.B. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat Biotechnol* **33**, 722-729 (2015).
106. Kwak, H., Fuda, N.J., Core, L.J. & Lis, J.T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-953 (2013).
107. Favre, A., Moreno, G., Salet, C. & Vinzens, F. 4-Thiouridine incorporation into the RNA of monkey kidney cells (CV-1) triggers near-UV light long-term inhibition of DNA, RNA and protein synthesis. *Photochem Photobiol* **58**, 689-694 (1993).
108. Kenzelmann, M. et al. Microarray analysis of newly synthesized RNA in cells and animals. *Proc Natl Acad Sci U S A* **104**, 6164-6169 (2007).
109. Saladino, R., Mincione, E., Crestini, C. & Mezzetti, M. Transformations of thiopyrimidine and thiopurine nucleosides following oxidation with dimethyldioxirane. *Tetrahedron* **52**, 6759-6780 (1996).
110. Trinklein, N.D., Murray, J.I., Hartman, S.J., Botstein, D. & Myers, R.M. The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response. *Mol Biol Cell* **15**, 1254-1261 (2004).
111. Shalgi, R., Hurt, J.A., Lindquist, S. & Burge, C.B. Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. *Cell Rep* **7**, 1362-1370 (2014).
112. Mukherjee, N. et al. Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat Struct Mol Biol* **24**, 86-96 (2017).
113. Russo, J., Heck, A.M., Wilusz, J. & Wilusz, C.J. Metabolic labeling and recovery of nascent RNA to accurately quantify mRNA stability. *Methods* **120**, 39-48 (2017).
114. Friedel, C.C., Dolken, L., Ruzsics, Z., Koszinowski, U.H. & Zimmer, R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res* **37**, e115 (2009).
115. Gelsi-Boyer, V. et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol* **145**, 788-800 (2009).
116. Scotti, M.M. & Swanson, M.S. RNA mis-splicing in disease. *Nat Rev Genet* **17**, 19-32 (2016).
117. Basu, S. et al. A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nat Struct Biol* **5**, 986-992 (1998).
118. Kiefer, L., Schofield, J.A. & Simon, M.D. Expanding the Nucleoside Recoding Toolkit: Revealing RNA Population Dynamics with 6-Thioguanosine. *Journal of the American Chemical Society* **140**, 14567-14570 (2018).
119. Herzog, V.A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* **14**, 1198-1204 (2017).
120. Burger, K. et al. 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. *RNA Biol* **10**, 1623-1630 (2013).
121. Riml, C. et al. Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing. *Angew Chem Int Ed Engl* **56**, 13479-13483 (2017).
122. Muhar, M. et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. *Science* **360**, 800-805 (2018).
123. Parker, R. & Sheth, U. P bodies and the control of mRNA translation and degradation. *Mol Cell* **25**, 635-646 (2007).

124. Luo, Y., Na, Z. & Slavoff, S.A. P-Bodies: Composition, Properties, and Functions. *Biochemistry* **57**, 2424-2431 (2018).
125. Wang, X. et al. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117-120 (2014).
126. Sheth, U. & Parker, R. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**, 805-808 (2003).
127. Aizer, A. et al. Quantifying mRNA targeting to P-bodies in living human cells reveals their dual role in mRNA decay and storage. *J Cell Sci* **127**, 4443-4456 (2014).
128. Tutucci, E. et al. An improved MS2 system for accurate reporting of the mRNA life cycle. *Nature Methods* **15**, 81-+ (2018).
129. Xing, W.M., D.; Parker, R.; Rosen, M. A quantitative inventory of yeast P body proteins reveals principles of compositional specificity. *bioRxiv* (2018).
130. Song, M.G., Li, Y. & Kiledjian, M. Multiple mRNA Decapping Enzymes in Mammalian Cells. *Molecular Cell* **40**, 423-432 (2010).
131. Grudzien-Nogalska, E., Jiao, X.F., Song, M.G., Hart, R.P. & Kiledjian, M. Nudt3 is an mRNA decapping enzyme that modulates cell migration. *Rna* **22**, 773-781 (2016).
132. Duggimpudi, S. et al. Transcriptome-wide analysis uncovers the targets of the RNA-binding protein MSI2 and effects of MSI2's RNA-binding activity on IL-6 signaling. *J Biol Chem* **293**, 15359-15369 (2018).
133. Bennett, C.G. et al. Genome-wide analysis of Musashi-2 targets reveals novel functions in governing epithelial cell migration. *Nucleic Acids Res* **44**, 3788-3800 (2016).
134. D'Lima, N.G. et al. A human microprotein that interacts with the mRNA decapping complex. *Nature Chemical Biology* **13**, 174-180 (2017).
135. Hubstenberger, A. et al. P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. *Mol Cell* **68**, 144-157 e145 (2017).
136. Wei, J. et al. Differential m(6)A, m(6)Am, and m(1)A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm. *Mol Cell* **71**, 973-985 e975 (2018).
137. Gebert, L.F.R. & MacRae, I.J. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol* **20**, 21-37 (2019).
138. Bail, S. et al. Differential regulation of microRNA stability. *RNA* **16**, 1032-1039 (2010).
139. Guo, Y. et al. Characterization of the mammalian miRNA turnover landscape. *Nucleic Acids Res* **43**, 2326-2341 (2015).
140. Burroughs, A.M. et al. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Res* **20**, 1398-1410 (2010).
141. Modepalli, V. & Moran, Y. Evolution of miRNA Tailing by 3' Terminal Uridylyl Transferases in Metazoa. *Genome Biol Evol* **9**, 1547-1560 (2017).
142. Swaminathan, G., Martin-Garcia, J. & Navas-Martin, S. RNA viruses and microRNAs: challenging discoveries for the 21st century. *Physiol Genomics* **45**, 1035-1048 (2013).
143. Fickenscher, H.F., B. Herpesvirus saimiri. *RSTB* (2001).
144. Cazalla, D., Yario, T. & Steitz, J.A. Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science* **328**, 1563-1566 (2010).
145. Lee, L.V., Kupke, K.G., Caballar-Gonzaga, F., Hebron-Ortiz, M. & Muller, U. The phenotype of the X-linked dystonia-parkinsonism syndrome. An assessment of 42 cases in the Philippines. *Medicine (Baltimore)* **70**, 179-187 (1991).
146. Makino, S. et al. Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *Am J Hum Genet* **80**, 393-406 (2007).

147. Hancks, D.C. & Kazazian, H.H., Jr. SVA retrotransposons: Evolution and genetic instability. *Semin Cancer Biol* **20**, 234-245 (2010).
148. Aneichyk, T. et al. Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell* **172**, 897-909 e821 (2018).
149. Hancks, D.C. & Kazazian, H.H. Roles for retrotransposon insertions in human disease. *Mobile DNA-Uk* **7** (2016).
150. Koike, N. et al. Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals. *Science* **338**, 349-354 (2012).
151. Kohlmaier, A. et al. A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol* **2**, E171 (2004).
152. Dixon, J.R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336 (2015).
153. Heintzman, N.D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318 (2007).
154. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349-352 (1997).
155. Condorelli, F., Gnemmi, I., Vallario, A., Genazzani, A.A. & Canonico, P.L. Inhibitors of histone deacetylase (HDAC) restore the p53 pathway in neuroblastoma cells. *Br J Pharmacol* **153**, 657-668 (2008).
156. Li, Y. & Seto, E. HDACs and HDAC Inhibitors in Cancer Development and Therapy. *Cold Spring Harb Perspect Med* **6** (2016).
157. Greer, C.B. et al. Histone Deacetylases Positively Regulate Transcription through the Elongation Machinery. *Cell Rep* **13**, 1444-1455 (2015).
158. Hnilicova, J. et al. Histone Deacetylase Activity Modulates Alternative Splicing. *Plos One* **6** (2011).
159. Brocks, D. et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat Genet* **49**, 1052-1060 (2017).
160. Ververis, K., Hiong, A., Karagiannis, T.C. & Licciardi, P.V. Histone deacetylase inhibitors (HDACIs): multitargeted anticancer agents. *Biologics* **7**, 47-60 (2013).
161. Lauffer, B.E. et al. Histone deacetylase (HDAC) inhibitor kinetic rate constants correlate with cellular histone acetylation but not transcription and cell viability. *J Biol Chem* **288**, 26926-26943 (2013).
162. Yeakley, J.M. et al. A trichostatin A expression signature identified by TempO-Seq targeted whole transcriptome profiling. *Plos One* **12** (2017).
163. Mitsiades, C.S. et al. Transcriptional signature of histone deacetylase inhibition in multiple myeloma: biological and clinical implications. *Proc Natl Acad Sci U S A* **101**, 540-545 (2004).
164. Kornblihtt, A.R. et al. The fibronectin gene as a model for splicing and transcription studies. *Faseb Journal* **10**, 248-257 (1996).
165. Nechaev, S. et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327**, 335-338 (2010).
166. Covault, J. & Chalkley, R. The identification of distinct populations of acetylated histone. *J Biol Chem* **255**, 9110-9116 (1980).
167. Whitehouse, I., Rando, O.J., Delrow, J. & Tsukiyama, T. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**, 1031-1035 (2007).
168. Shivaswamy, S. & Iyer, V.R. Stress-dependent dynamics of global chromatin remodeling in yeast: dual role for SWI/SNF in the heat shock stress response. *Mol Cell Biol* **28**, 2221-2234 (2008).

169. Martens, J.A., Laprade, L. & Winston, F. Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* **429**, 571-574 (2004).
170. Martens, J.A., Wu, P.Y. & Winston, F. Regulation of an intergenic transcript controls adjacent gene transcription in Saccharomyces cerevisiae. *Genes Dev* **19**, 2695-2704 (2005).
171. Hainer, S.J., Pruneski, J.A., Mitchell, R.D., Monteverde, R.M. & Martens, J.A. Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev* **25**, 29-40 (2011).
172. Vitting-Seerup, K. & Sandelin, A. The Landscape of Isoform Switches in Human Cancers. *Mol Cancer Res* **15**, 1206-1220 (2017).
173. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
174. Wei, T. Package 'corrplot'. *Statistician* **56**, 316-324 (2015).
175. Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A. & Weeks, K.M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* **11**, 959-965 (2014).
176. Li, Y., Rao, X., Mattox, W.W., Amos, C.I. & Liu, B. RNA-Seq Analysis of Differential Splice Junction Usage and Intron Retentions by DEXSeq. *PLoS One* **10**, e0136653 (2015).
177. Xu, H. et al. FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One* **7**, e52249 (2012).
178. Martin, M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. *EMBnet.journal* **17**, 10-12 (2011).
179. Kim, D., Langmead, B. & Salzberg, S.L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015).
180. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
181. Danecek, P. & McCarthy, S.A. BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* **33**, 2037-2039 (2017).
182. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
183. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
184. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192 (2013).
185. Van Herreweghe, E. et al. Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. *Embo J* **26**, 3570-3580 (2007).
186. Carpenter, B.G., A.; Hoffman, M.; Lee, D.; Goodrich, B.; Betancourt, M.; Guo, J.; Li, P.; Riddel, A. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* (2017).
187. Thomas, P.D. et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13**, 2129-2141 (2003).
188. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).
189. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

Appendix A
Total RNA isolation and TimeLapse chemistry protocol

A1. s^4U treatment and cell harvesting

1. Plate and grow cells to ~60-70% confluence.
2. Supplement media with s^4U and incubate cells for a time determined based on desired application (e.g. mRNA turnover, 4h s^4U treatment; transient species, 5 min s^4U treatment).

Note 1: Final concentration of s^4U will depend on application and cell type. For example, 293T cells can be treated with 100 µM s^4U for 4 h (or longer), or 1 mM s^4U for 5 min.

Note 2: If the cells' culture media is relatively fresh (split within the last 24 h), transfer media from culture plate to falcon tube, add stock of s^4U in sterile water (e.g. 500 mM stock), invert gently to mix, and reintroduce to cell culture plate. If the culture media is relatively depleted, use pre-warmed fresh media with supplemented s^4U.

Note 3: s^4U is photosensitive, keep solutions wrapped in foil and minimize exposure of samples to light.

3. After incubation period, place cell culture plates on ice. Aspirate media from plate, gently rinse plate once with 5 ml ice cold PBS and aspirate again.
4. Add 500 µl ice cold PBS to cells. Gently scrape cells from plate using a cell scraper, and transfer cell suspension to a 1.5 ml loBind epi tube.
5. Pellet cells in a pre-chilled (4°C) centrifuge at 300xg for 3 min. Carefully aspirate PBS from cell pellet.

Note 4: If centrifuge does not accommodate 1.5 ml epi tubes, the epi tube can be placed in an empty 15ml falcon tube for centrifugation.

6. Resuspend pellet in 1 mL Trizol by gently pipetting up and down ~10 times.

Note 5: Trizol is toxic, use with care and in well ventilated area.

7. Trizol samples can be stored overnight at -80°C, or kept on ice for RNA isolation.

A.2 RNA isolation

1. Thaw Trizol samples at room temperature. Once completely thawed, keep samples at room temperature for 5 minutes in dark.
2. Add 200µl chloroform to each Trizol sample. Shake samples vigorously for 15 seconds.
3. Transfer Trizol/chloroform mix to heavy phase-lock tube.

Note 6: Prepare phase lock tubes before adding sample by spinning tubes at 12,000xg for 1 min.

4. Allow sample to stand for 2 min at RT in dark.
5. Centrifuge sample at 12,000xg for 10 min in pre-chilled (4°C) microcentrifuge.

6. Transfer 500 µl of aqueous layer to fresh loBind epi.

Note 7: 1 µl of Glycoblue coprecipitant can be added to sample at this step if anticipated yield is low.

7. Add 500 µl (1 eq) of isopropanol supplemented with 1mM final freshly prepared DTT to sample. Invert tube ~10 times or until thoroughly mixed.
8. Incubate samples 10 min at RT in dark.
9. Pellet RNA at ~20,000xg for 20 min in pre-chilled (4°C) microcentrifuge.
10. Remove supernatant from pellet, wash pellet once with freshly prepared 75% EtOH.
11. Remove EtOH and dry pellet.

Note 8: Tube can be spun after removing bulk of EtOH to collect residual liquid at bottom of tube. Carefully remove residual EtOH using narrow pipette tip (e.g. 10 µl filter tip), and sample tube can be left open covered by a kimwipe to dry. Do not overdry or pellet will be difficult to resuspend.

12. Resuspend pellet in small volume (~50 µl) of nuclease-free water and keep on ice.
13. Assess RNA concentration by nanodrop.

A.3 Genomic DNA depletion

1. Combine 20 ug of isolated RNA, 2 µl Turbo DNAse, 10 µl Turbo DNAse buffer, add nuclease free water to total volume of 100 µl in a PCR tube.
2. Incubate at 37°C for 30 min.
3. Add 100 µl RNAclean beads to each sample.

Note 9: Prior to use, vortex RNAclean beads and allow to come to room temperature for 30 minutes.

4. Flick and invert or lightly vortex samples to mix. Incubate samples for 8 minutes.
5. Place samples on magnetic rack and allow beads to collect until solution is clear (~5 min)
6. Remove supernatant and wash 2x 200 µl freshly prepared 80% EtOH
7. Remove EtOH and briefly spin PCR tubes to collect residual EtOH. Removed residual with 10 µl pipet tip.
8. Leave PCR tubes open on magnetic rack and allow to dry (~5 min).

Note 10: When beads are dry, you may see a slight crack in the pellet and the beads will appear lighter in color. Larger bead volumes (e.g. 100 µl) require longer to dry.

9. Add 20µl nuclease-free water to dried beads and flick tubes until completely resuspended. Incubate in dark for 5 min at RT.
10. Briefly spin tubes to collect sample at bottom and place on magnetic rack until solution is clear (~1-2 min).
11. Transfer eluted RNA in water to clean PCR strip.
12. Assess RNA concentration by nanodrop.

A.4 TimeLapse chemistry (25µl volume)

1. Dilute RNA (~2-5 ug) into 15µl nuclease-free water.
2. Prepare TimeLapse master mix (8.7 µl per sample, make 10% extra)

-Per sample: 1.3µl Trifluoroethylamine (TFEA), 0.84µl 3 M sodium acetate pH 5.2, 0.2µl 500 mM EDTA, 6.4 µl nuclease free water.

Note 11: TFEA is volatile, use care when pipetting to ensure adding proper volume. Pipetting TFEA up and down a few times will equilibrate the vapor pressure.

3. Freshly prepare a 200 mM solution of sodium periodate in nuclease free water.
4. To each sample, add 8.7 µl TimeLapse master mix. Flick tubes to combine well and briefly spin to collect sample at bottom of tube.
5. Add 1.3 µl sodium periodate solution to each sample and flick tubes to combine well. Briefly spin tubes to collect sample at bottom of tube.
6. Incubate in PCR cycler at 45°C for 1 h.
7. Purify RNA with an equal volume (25 µl) of RNAClean beads, repeating steps 3-10 in genomic DNA depletion section of protocol (elute with 20 µl nuclease free water).
8. Transfer 18 µl eluate to fresh PCR tube.

A.5 Reducing treatment

1. Prepare reducing master mix (2 µl per sample)

-For 100µl: 58 µl nuclease-free water, 10 µl 1 M DTT, 10 µl 1 M Tris pH 7.4, 2 µl 500 mM EDTA, 20 µl 5 M NaCl.

2. Add 2 µl reducing master mix to each sample, flick tubes to combine well. Briefly spin tubes to collect sample at bottom of tube.
3. Incubate in PCR cycler at 37°C for 30 min.
4. Purify RNA with an equal volume (20 µl) of RNAClean beads, repeating steps 3-10 in genomic DNA depletion section of protocol (elute with between 10-20 µl nuclease free water, depending on downstream application).
5. Assess RNA concentration using nanodrop or bioanalyzer.