Abstract

Investigating RNA Population Dynamics Through Nucleotide Recoding:
Extending the TimeLapse-seq Toolkit to 6-Thioguanosine

Lea Kiefer

2020

The eukaryotic transcriptome is under constant flux between transcription, processing, and decay. As most biological processes occur under non-steady state conditions (e.g. during development or in reaction to an external stimulus), multiple snapshots of the transcriptome would be required to identify RNAs under complex kinetics. We and others have developed novel nucleotide chemistry capable of capturing transcriptional dynamics in addition to relative RNA abundance in an RNA-sequencing (RNA-seq) experiment. Recoding the hydrogen bonding pattern of the metabolic label 4-thiouridine (s^4U), a method termed TimeLapse-seq, enables us to distinguish sequencing reads of new RNA from reads of pre-existing RNA through the presence of T-to-C mutations. At its initial development, TimeLapse-seq provided genome-wide information assuming steady state kinetics and limited information about transcriptional induction. Additionally, at the time I started my thesis research the most commonly used metabolic labels to capture RNA dynamics through enrichment and the only label used in nucleotide recoding were analogues of uridine. The aim of my thesis, therefore, was to push the development of TimeLapse-seq to assess complex and non-steady state RNA dynamics through the introduction of a second convertible nucleoside, 6-thioguanosine (s^6G).

I extended the applications of TimeLapse-seq through multiple avenues. First, I developed the metabolic label s^6G as a convertible nucleoside capable of providing temporal information on the transcriptome genome-wide, as TimeLapse-seq with s^6G. This is, to the best of my knowledge, the first report of using a guanosine analogue to detect RNA dynamics through mutational analysis. The conversion of s^6G to an

analogue of adenosine is achieved under identical chemical conditions used to recode s^4U, therefore opening up the opportunity of using both labels within the context of a single experiment. Secondly, I improved existing protocols to extend our ability to capture nascent transcription in the context of both tissue culture and primary cells. Lastly, I engaged in developing statistical models to gain estimates of kinetic parameters and the change thereof in the context of collaboration projects. Overall, I contributed to, improved and established technologies to capture the dynamic processes underlying and regulating the cellular transcriptome.

Investigating RNA Population Dynamics Through Nucleotide Recoding: Extending the TimeLapse-seq Toolkit to 6-Thioguanosine

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

by
Lea Kiefer

Dissertation Director: Matthew D. Simon

December 2020

Contents

Acknowledgement vi

1 Introduction **1**

1.1 Cellular RNA Dynamics 1

1.2 Biochemical Enrichment to Study RNA Dynamics 5

1.3 Enrichment-Free Methods to Study RNA Dynamics 8

1.3.1 TimeLapse-seq Development and Considerations 12

1.3.2 Controls and Data Analysis 15

1.3.3 Capturing the Transient Transcriptome by Combining Biochemical Enrichment and Nucleotide Recoding 16

1.4 Determining RNA Dynamics under Non-Steady State Conditions . . 17

1.4.1 Technological Advancement - s^6G as a Recodable Nucleoside? 19

1.4.2 Background on s^6G in Technologies and the Cellular Context . 19

2 Development of Recodable Nucleosides for TimeLapse-seq **21**

2.1 Author contributions 21

2.2 Summary 22

2.3 Characterization of the s^4U Recoding Reaction 22

2.4 Assessing s^4U and s^6G Incorporation into RNA in Cells 25

2.5 Characterization of s^6G as a Recodable Nuc-leoside 30

2.6 Conclusion 37

3 Revealing RNA Population Dynamics with 6-Thioguanosine 38

3.1 Author contributions 38

3.2 Summary 38

3.3 Assessing s^6G Recoding Efficiency and Ortho-gonality in Cellular RNA 39

3.4 Capturing RNA Dynamics Genome-wide with s^6G Recoding 43

3.5 Discussion 54

4 Capturing the Transient Transcriptome with TT-TimeLapse-seq 55

4.1 Summary 55

4.2 Investigating Active Transcription After Heat Shock in Drosophila S2 Cells 56

4.3 Capturing Cell-Type Specific Transcriptional Signatures Using UPRT/TT-TimeLapse-seq in Mouse Tissues 59

4.4 Detecting Transient Transcripts in the Mouse Olfactory Epithelium . 62

4.5 Conclusion and Ongoing Efforts 66

5 Bayesian Modeling of TimeLapse-seq Data 67

5.1 Summary 67

5.2 TimeLapse-seq Data Processing and Modeling 68

5.3 Modeling the Transcriptional Response to VSV Infection 71

5.4 Conclusion 79

6 Conclusions and Ongoing Work 80

6.1 Summary 80

6.2 Integrating s^4U and s^6G 81

6.3 Discussion 83

6.4 Ongoing and Future Directions 83

7 Methods and Data Analysis 85

7.1 RNA Dot Blot - Assessing s^4U/s^6G Incorporation 85
7.2 DNA Dot Blot - Assessing s^6G Incorporation 87
7.3 s^6G DNA Restriction Digest Assay 89
7.4 s^6G TimeLapse LC-MS Assay 89
7.5 NMR 90
7.5.1 Reaction of 4-Thiouracil with TimeLapse Chemistry 90
7.5.2 Reaction of 6-Thioguanine with TimeLapse Chemistry 90
7.6 TimeLapse-seq with s^6G 90
7.7 TT-TimeLapse-seq 91

List of Figures

1 The complexity of determining the source of an increase in RNA-seq signal 4

2 TimeLapse-seq recoding of the hydrogen bonding pattern of s^4U . . . 10

3 Workflow of a s^4U TimeLapse-seq experiment 11

4 s^4U TimeLapse-seq captures RNA dynamics 14

5 TimeLapse chemistry causes clean conversion of 4-thiouracil 24

6 Measuring s^6G and 6-TG incorporation into cellular RNA by dot blot 27

7 Incorporation kinetics of s^4U, s^6G and 6-TG by dot blot 28

8 6-TG does not affect cell viability over the course of a TimeLapse-seq experiment 29

9 6-thioguanine is recoded to a 2-amino-adenine analogue under TimeLapse conditions 32

10 s^6G is efficiently recoded to A* under TimeLapse conditions 33

11 s^6dG is recoded to dA* under TimeLapse conditions in the context of dsDNA 34

12 Screen for optimal s^6G recoding conditions by restriction digest assay 35

13 s^6G and TimeLapse chemistry-dependent increase in G-to-A mutations shows orthogonality of the reaction 41

14 s^6G and TimeLapse chemistry-dependent increase in G-to-A mutations in sequencing reads of *ACTB* 42
15 s^6G TimeLapse-seq approach 46
16 6-TG treatment and TimeLapse chemistry do not affect RNA-seq output 47
17 TimeLapse-seq with s^6G leads to increases in G-to-A mutations . . . 48
18 s^6G captures transcriptional dynamics genome-wide qualitatively and quantitatively 49
19 Reproducible estimates of fraction of new reads determined by statistical modeling of TimeLapse-seq data 50
20 RNA half-lives determined by s^6G TimeLapse-seq track with s^4U-based estimates 51
21 s^6G TimeLapse-seq captures transcriptional dynamics for the uridine-poor transcript *CBX4* 53
22 Heat shock causes drastic transcriptional shutdown of all except for heat shock response genes in *Drosophila* S2 cells by TT-TimeLapse-seq 58
23 Capturing cell type-specific transcripts by TU-tagging TT-TimeLapse-seq 61
24 Transient intronic and antisense RNA in mouse olfactory epithilium by TT-TimeLapse 64
25 Mixed Poisson model of TimeLapse-seq reads to determine the fraction of newly-made RNA 69
26 Experimental design of a TimeLapse-seq time course pre- and post-VSV infection 73
27 Model estimates of the T-to-C mutation rates and kinetic RNA parameters 76
28 Ranked changes in k_{syn} between viral infection and control shows many known innate immunity transcripts 78

29 Metabolic labels interact with each other and canonical nuclear bases causing lower RNA incorporation 82

Acknowledgement

I am grateful for being given the opportunity to do what I love, continuously be allowed to learn and grow, be surrounded by people who keep me thinking outside of the box and be given the security and support to live a good life here in the United States.

I am thankful for my high school chemistry teacher, Joseph Lösch, who besides not knowing any English helped me write my applications, invented a role of school counselor, filled it with our Biology teacher and supported me in my desire to apply to chemistry programs in the U.S.. He really sparked my curiosity about the sciences and encouraged me to pursue a career in the sciences from day one. I am grateful for Dr. Joseph Provost, who took me on as his first research student after his move to the University of San Diego. I can say that I have learned all my protein biochemistry directly from him. As an advocate for undergraduate research, he made sure my presentations were well-rehearsed, that I got to go to national conferences and that my grad school applications were handed in on time. It makes all the difference when you have someone knowledgeable about the application process continuously encourage you to give your best even when some programs explicitly stated that international students shouldn't even apply. And finally, I would like to thank Matt for his contagious excitement about science at large and the possibilities to study molecular processes that lie within chemical biology. I have learned a lot not only about how to do careful science but also how to communicate science both in an effective and an aesthetically pleasing way. And I would like to thank my committee members, Karla Neugebauer and David Schatz for their insightful feedback and for supporting me throughout my graduate school career.

I would like to thank the current and former Simon lab members, who made going to work every day something to look forward to. Specifically, I would like to thank Jeremy for answering endless questions and guiding me through the beginning stages

of my project; Tyler for his mentorship during my rotation and for always providing his extensive knowledge on science in everyday life; Alec for his positivity and endless discussions about the inconsistencies of the Harry Potter books; Meg for constant encouragement and our shared love for hiking and camping adventures; Michelle for always lending an ear and for her thoughtfulness; Will for always being supportive in and outside of the lab; Martin for lots of helpful conversations and for providing another European perspective in the lab; Josh for lots of insightful discussions about science and entertaining texts during lab zooms; Leah for endless discussions about what is happening in the world outside of science (also sorry for causing you nightmares that forced you to read the Harry Potter books); Isaac for helping me think more deeply about my project and helping to communicate our statistical analyses. I would like to thank Paula Maher-Rivera for making our lab and institute run smoothly and for scheduling all my committee meetings for me. Thank you also to everyone who helped proofread and edit this thesis.

Finally, I would like to thank all my friends and family who made graduate school a memorable time. Thank you to all the friends I made in New Haven, especially to my former roommate Maddie, former singing companion Haya, Nandan and Wei for their friendship and support and Martha and Garrett for all the shared experiences. I would like to also thank Neal and his family, Bernie, Ravi, Tara, Dan and Ruby for being my home away from home in the U.S. together with our feline family Polli, Ori, Persephone and the late Johnny. And finally I would like to thank my best friend Svenja for sticking with me for so long, Steffi and the other girls from Edith Stein and my family, Anne, Georg, Jan, Eva, Robbie and Giga, Elli, Julia, Woa, Klaus and Bernie. I am very lucky to have you.

Chapter 1

Introduction

1.1 Cellular RNA Dynamics

Living cells are governed by the dynamic processes that regulate gene expression. RNA, in its duality between holding information content and exhibiting direct functional roles, similarly exists in a state of constant flux. RNA is continuously being synthesized, processed, and degraded with a wide range of synthesis rates [1]. While RNA transcription and processing, namely the co-transcriptional assembly of the spliceosome leading to removal of intronic sequences [2] as well as post-transcriptional modifications [3], happen on the order of minutes, RNA stability spans a much wider range from a few minutes for, for instance, antisense to many hours for housekeeping messenger RNAs [4]. The balance between RNA transcription and decay leads to a fine-tuned regulation of gene expression. Any disruption in the rates of transcription or decay therefore lead to imbalances of RNA levels, which can contribute to disease phenotypes. For instance, a recent study found that RNA destabilization rather than changes in RNA synthesis lead to lower RNA levels available for translation into ribosomal and mitochondrial proteins, a hallmark of amyotrophic lateral sclerosis (ALS) and other neurodegenerative diseases [5]. Conversely, mutations driving changes

in RNA synthesis, for instance through overactivation, lead to high levels of mRNA available for translation into growth-driving transcription factors, which often drive cancer progression [6]. These examples illustrate that the transcriptome is not a static entity [7] and that loss of regulation of the dynamic processes involved in maintaining RNA levels can lead to global changes in cellular function.

Under steady state conditions, the relative abundance of a given RNA is dependent on its synthesis and degradation rates.

$$\text{DNA} \xrightarrow{k_\text{s}} \text{RNA} \xrightarrow{k_\text{d}} \text{degradation product}$$

The change of the steady state RNA concentration over time can therefore be described as the rate of synthesis minus the degradation rate [8].

$$\frac{d[\text{RNA}]}{dt} = k_\text{s} - k_\text{d}[\text{RNA}]$$

By definition, under steady state conditions, the RNA concentration is not changing ($\frac{d[\text{RNA}]}{dt} = 0$), leading to a direct relationship between steady state levels of RNA and the ratio of its synthesis rate to decay rate constant.

$$[\text{RNA}]_\text{ss} = \frac{k_\text{s}}{k_\text{d}}$$

While the result of the dynamic processes of transcription and decay, namely relative RNA abundance, can be easily captured using RNA-seq, the parameters of these processes are more elusive. In order to estimate decay rates, newly-made RNA needs to be distinguishable from pre-existing RNA. This becomes apparent when examining the relationship of the degradation rate constant with the fraction of new RNA (θ), and the time frame of interest, t, assuming simple exponential decay.

$$k_{\mathrm{d}} = \frac{-\ln(1-\theta)}{t}$$

As newly-made and pre-existing RNA are inherently chemically indistinguishable, telling the two subgroups apart presents a challenge. The lack of temporal information in regular RNA-seq experiments becomes especially obvious when considering the underlying processes driving a change in RNA-seq signal. A graphic representation of the difficulty of discerning which kinetic parameter is driving the observed change in RNA-seq signal is illustrated in Figure 1A. Shown is a simple representation of how both an increase in the synthesis rate or a decrease in the degradation rate could equally result in an increase in RNA-seq signal. If, however, we were able to distinguish the newly synthesized RNA from the old RNA by, for instance coloring all newly-made RNA in red, contribution of either rate would more easily become apparent (Figure 1B). More importantly, from a quantitative perspective, the fraction of new RNA (θ) determined from this approach provides information necessary to determine the transcript-specific degradation rate constant (assuming steady state) as outlined in the equation above, as well as RNA half-lives ($hl = \frac{ln(2)}{k_d}$). As outlined in this section, determining kinetic parameters is important for identifying the molecular mechanism underlying a change in RNA abundance and to determine these parameters one needs to be able to distinguish newly-made from pre-existing RNA. In the next section I will outline various strategies that have been developed and employed over the past decades in order to address this need.

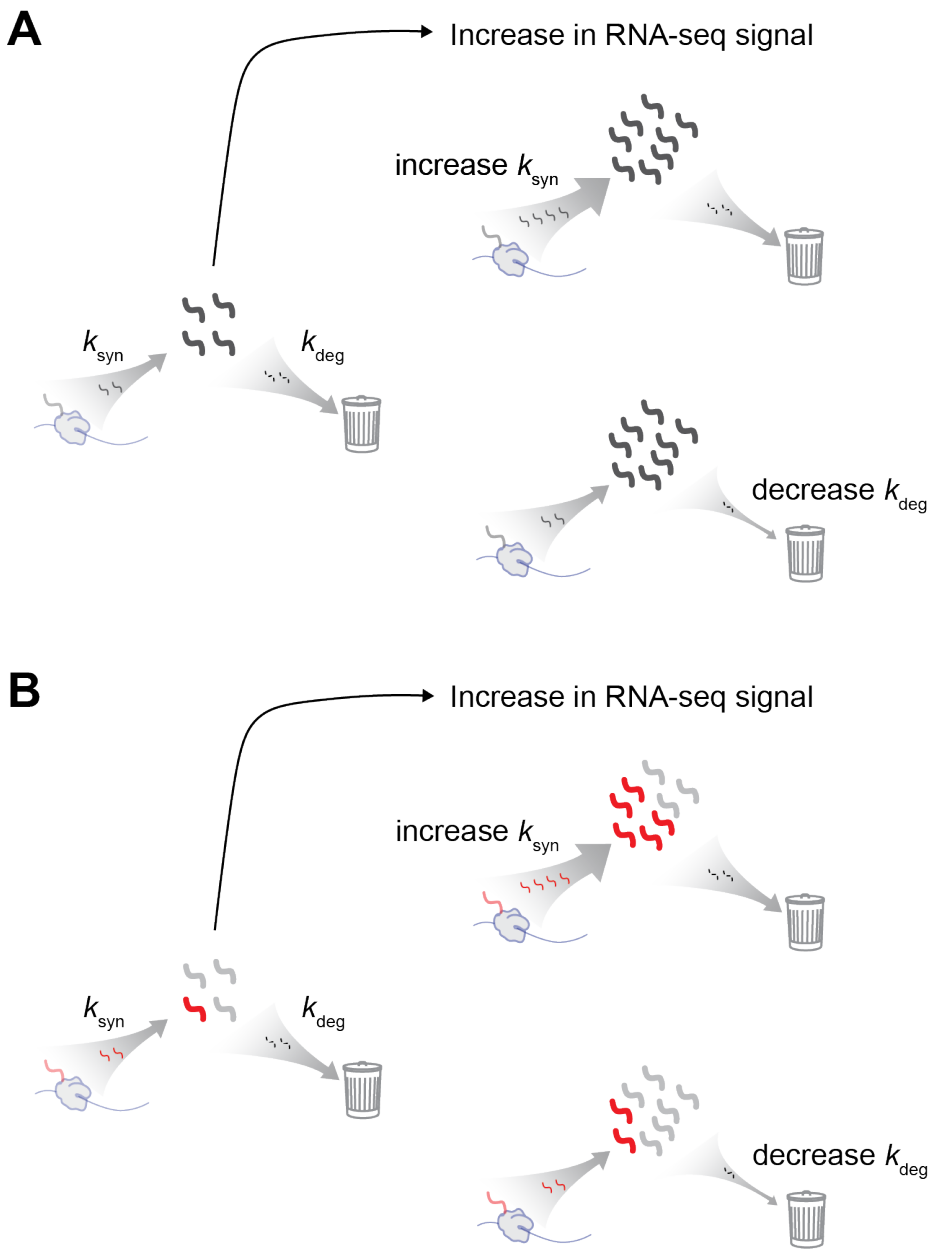

Figure 1: The complexity of determining the source of an increase in RNA-seq signal

A schematic outlining the difficulty of observing the changes in the underlying dynamic processes driving a change in RNA-seq signal. (A) An increase in RNA-seq signal (wavy lines) can either be driven by an increase in the synthesis rate (top, blob represents RNA polymerase II) or a decrease in the degradation rate (bottom, trash bin signifies RNA degradation). A change in either process could result in the same overall increase in RNA-seq signal making it difficult to identify the rate

driving the change. (B) Same as in A, with the addition of indicating newly made RNA (red wavy lines). An increase in the synthesis rate (top) leads to a higher proportion of new (red) RNA in the total pool, while a decrease in the degradation rate (bottom) leads to an accumulation of old (grey) RNA making the two processes distinguishable. This concept is the basis for identifying RNA population dynamics by using biochemical enrichment and nucleotide recoding techniques to distinguish newly-made from pre-existing RNA.

1.2 Biochemical Enrichment to Study RNA Dynamics

The same steady state levels of RNA can be achieved through multiple pathways of regulation. While RNA-seq is a powerful tool to determine relative levels of RNA, it is not well suited to identify which changes in dynamic processes led to the observed transcriptional signature. As outlined in the previous section, information on RNA transcription and decay rates could be gained if one could distinguish newly-made from pre-existing transcripts.

While this section will focus on studying RNA dynamics through biochemical enrichment of newly-made RNA, similar approaches to estimate RNA dynamics, without metabolic labeling, have been developed based on the abundance of precursor-mRNA (pre-mRNA) to mRNA. These approaches use a combination of mathematical modeling and measured profiles of pre-mRNA to mRNA abundance from RNA-seq [1, 9] to obtain transcript-specific estimates of RNA production, processing and degradation. These computational approaches highlight that intronic reads in RNA-seq can be used as a measure to detect changes in transcriptional activity [10] and have even been employed to capture transcriptional dynamics in single cells (RNA velocity [11]). The limitation of these techniques lies in the presence of introns in the sequencing data sets,

which is governed by the overall abundance of the transcript, the sequencing depth of the experiment and the timescale of splicing; the exact rates of splicing of individual introns *in vivo* still remain controversial [12]. So while these computational tools are instrumental to determine RNA dynamics from classic RNA-seq experiments, the relatively low proportion of pre-mRNA to RNA may prove limiting for fast-splicing or low abundance RNAs.

Instead of relying on pre-mRNA, metabolic labeling strategies can extend over larger timescales and can be applied to a wider range of RNA species. The field initially studied population dynamics through incorporation of radio-labeled nucleosides [13, 14, 15]; however, it wasn't until the use of non-canonical nucleosides that biochemical separation of new from old RNA became possible (reviewed in [16]). The most commonly used RNA metabolic labels are analogues of uridine, specifically 5-bromouridine (BrU) [17, 18], 5-ethynyluridine (EU) [19, 20] and 4-thiouridine (s^4U) [4, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32]. Each of the non-canonical nucleosides are readily taken up by cells from the media, phosphorylated and incorporated into RNA molecules that are actively being transcribed, without significant perturbation in most cell systems.

It has been reported that s^4U has the capacity to influence RNA tertiary interactions due to the changed preference of s^4U to base pair with G over A [33]. While this effect on RNA tertiary structure has been shown to cause negligible impacts on reactivity in the context of one functional RNA [33], it is important to assess the impact of s^4U on the biological system of interest. For instance, it has been reported that s^4U causes nucleolar stress in some cell lines [34] and previous work by former graduate student Jeremy Schofield indicates that s^4U impacts HeLa cell survival. On the other hand, the most commonly used human tissue culture cell lines, e.g. 293T or K562, show small if any impact on the total transcriptome even after hours of s^4U labeling as read out by viability and comparative RNA-seq studies [31, 35, 36, 37, 38]. It is noteworthy

that labeling mammalian cells with s^4U leads to low levels of substitution of s^4U for canonical uridine (4-5%) thereby minimizing the risk for causing RNA structural changes which may lead to impaired function. As longer labeling times increase the potential for cellular toxicity it is advisable to keep the final s^4U concentration low (tens of μM). In conclusion, it is important to maintain cautious when using s^4U labeling in a new cell-line or organism (refer to this published summary of previously used cell lines, s^4U labeling times and concentrations [39]), and it is advisable to compare total RNA-seq signal between treated and control samples to rule out any s^4U-induced changes.

The aforementioned non-canonical nucleosides each provide a biochemical "handle" allowing for the enrichment of newly-made RNA away from the pool of pre-existing RNA. While BrU-containing RNA is enriched through immunoprecipitation, EU and s^4U provide orthogonal chemical groups capable of forming a covalent bond to the enrichment molecule or resin. The ethynyl group of EU can be linked to biotin-azide through "click" chemistry in an essentially irreversible reaction, allowing for RNA enrichment through streptavidin [19]. The sulfur at position four of s^4U can be covalently linked to an activated disulfide compound (e.g. HPDP-biotin [22], MTS-biotin [31]) through the reversible formation of a disulfide bond. The reversible nature of this enrichment strategy allowed for further technological advancements in studying RNA dynamics covered in detail in Chapter 4. Using these non-canonical nucleosides newly-made RNA, namely RNA transcribed during the course of metabolic labeling, can then be biochemically enriched from total RNA post cell lysis. In the case of s^4U, tens of experiments have been performed over the years with labeling times ranging from minutes to several hours in a variety of cell systems. Through RT-qPCR and later also through next generation sequencing, RNA half-lives were determined for short and long-lived RNAs genome-wide (reviewed in [16, 39]).

As with all biochemical enrichment protocols, background contamination and

cross-experimental normalization can pose significant problems. Even though non-canonical nucleosides like EU and s^4U enable covalent linkage for enrichment, which allows for stringent wash conditions, nonspecific RNA background is still observed (up to 30% [40]). Background contamination from pre-existing RNA can lead to significant overestimation of, for instance, splicing rates of highly abundant transcripts (overestimation of spliced to unspliced RNA). Different research groups have employed several strategies to minimize background contamination including various biochemical steps [41, 25, 42] as well as pre-shearing the RNA [43]. However, background contamination remains a factor when enriching for newly-made RNA and therefore limits the accuracy of the determined kinetic parameters. Similarly, biochemical enrichment experiments require additional reference points to be able to compare across replicates and different experiments. While RNA-seq experiments can be normalized to each other based on read coverage and sequencing depth [44, 45], biochemically enriched samples often use the addition of exogenous or synthetic spike-in RNA. However, these approaches have been shown to lead to variable results and have recently been shown to be outperformed by normalization to endogenous intron counts [46]. In summary, the need to normalize across enrichment samples together with RNA background contamination create a source of extra uncertainty in the estimated kinetic parameters.

1.3 Enrichment-Free Methods to Study RNA Dynamics

These obstacles sparked the development of enrichment-free RNA-seq experiments based on the principle of nucleotide recoding to separate new from old RNA. In late 2017, we and others reported a novel approach to capture global transcriptional dynamics by differentiating newly-made from pre-existing RNA without the need for biochemical enrichment (TimeLapse-seq [36], SLAM-seq [37]). The premise of these

techniques is the orthogonal reactivity of s^4U compared to canonical nucleotides. While both technologies use the metabolic label s^4U, different chemical reactions are employed to detect s^4U through a mutation. The work flow of both technologies, the resulting data and analysis is similar yielding comparable information about RNA population dynamics. The development and considerations of TimeLapse-seq will be discussed in more detail in the next section. Briefly, nucleotides are "read" by the cellular machinery through their unique hydrogen bonding patterns. As an analogue of uridine, s^4U shares the same Watson-Crick face hydrogen bonding pattern, which allows it to be invisible to the cell in most systems and therefore be substituted in for uridine during RNA synthesis. In order to detect locations of s^4U incorporation into RNA through next generation sequencing, two different approaches can be employed. The hydrogen bonding pattern of s^4U can be disrupted trough alkylation (SLAM-seq [37]) leading to a mutation in the reverse transcription step. Our research group developed an alternative approach, namely chemical conditions that lead to the recoding of the s^4U hydrogen bonding pattern to resemble that of cytidine (Figure 2). During reverse transcription and subsequent PCR as part of the library preparation, this U-to-C transition is carried through and ultimately leads to a T-to-C mutation at the location of the s^4U in the sequencing read. As a qualitative assessment, the sequencing reads can then be crudely divided into T-to-C containing and non-T-to-C containing reads to differentiate reads stemming from newly-made and pre-existing RNA, respectively. More sophisticated analyses developed by our lab will be discussed in Chapter 5 and a schematic of the s^4U TimeLapse-seq workflow can be found in Figure 3. This approach not only circumvents the issue of background contamination, it also provides internal normalization since the standard RNA-seq analyses and normalization strategies can be applied to these experiments. Overall, nucleotide recoding of s^4U allows one to determine kinetic parameters, such as synthesis and degradation rates, from a classic RNA-seq experiment without the need for biochemical enrichment of RNA.

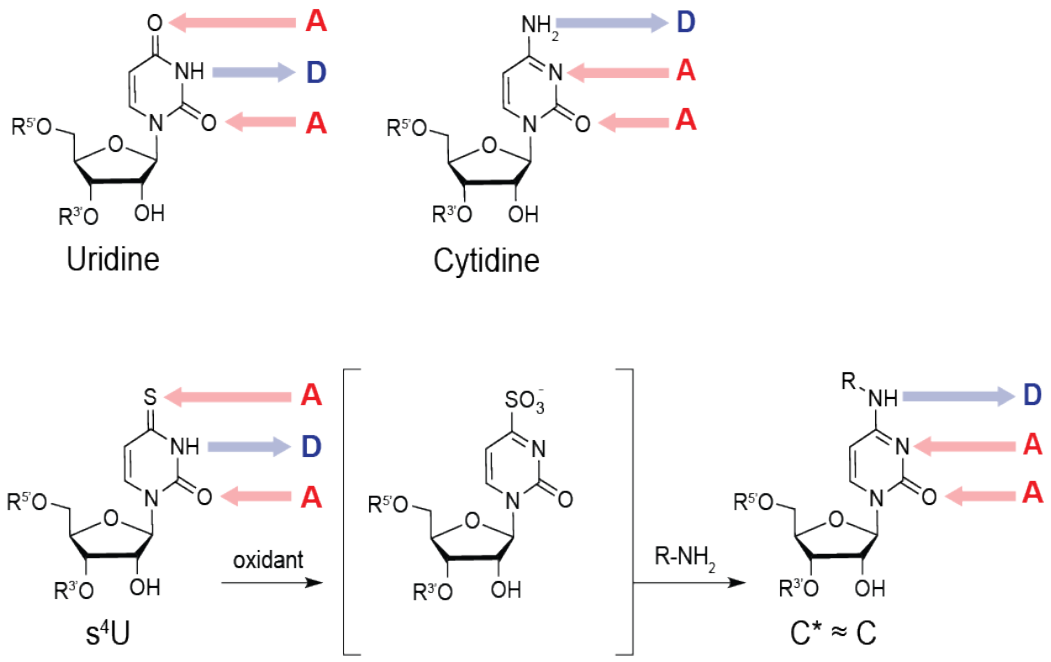

Figure 2: TimeLapse-seq recoding of the hydrogen bonding pattern of s^4U

Depicted are the chemical structures of uridine and cytidine with their corresponding donor (D) and acceptor (A) hydrogen bonding patterns (top). The reaction scheme of TimeLapse-seq recoding of the hydrogen bonding pattern of s^4U to an analogue of cytidine with the predicted intermediate state and reaction conditions (bottom). The reaction is performed for 1h at 45°C and does not compromise the integrity of the RNA.

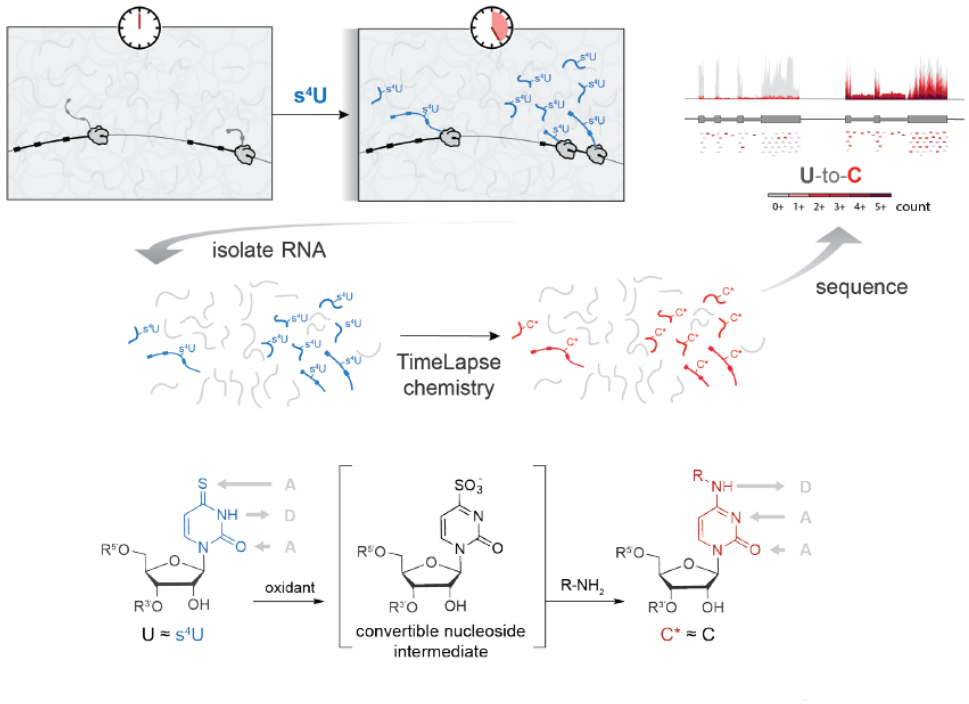

Figure 3: Workflow of a s^{4}U TimeLapse-seq experiment

Cells are treated with s^{4}U for a given amount of time, after which total RNA is extracted and subjected to TimeLapse chemistry which leads to the recoding of the hydrogen bonding pattern of uridine into that of cytidine. Subsequent reverse transcription and next generation sequencing preserves the U-to-C recoding and is identified as a T-to-C mutation in the sequencing reads compared to the reference genome. Sequencing reads with increasing number of T-to-C mutations are shown in increasingly darker shades of red. This figure was adapted from [36].

1.3.1 TimeLapse-seq Development and Considerations

As outlined in the preceding paragraph, the underlying principle of TimeLapse-seq, namely the recoding of s^4U, was motivated by the challenges of RNA enrichment experiments. In an effort to move away from biochemical enrichment, Jeremy Schofield consulted the nucleotide chemistry literature. He discovered that chemical conditions to react s^4U into an analogue of cytidine had previously been identified [47]. However, the reactions were not performed under conditions that would ensure RNA integrity. He therefore proposed RNA-friendly reagents to recode the hydrogen bonding pattern of s^4U into the pattern of cytidine (see Figure 2) and started to screen for optimal conversion conditions using a variety of chemical and biochemical assays. A list of requirements were put together to ensure the applicability of s^4U recoding to accurately infer the fraction of newly-made RNA through sequencing: The conversion of s^4U had to be orthogonal, meaning that s^4U would be recoded while all canonical nucleotides stay unreacted. The conversion reaction had to be efficient, reacting the majority of s^4U to a cytidine analogue in order to achieve sufficient signal. And finally, the reaction had to be clean, leading to the cytidine analogue as the sole product without formation of any byproducts. Through simulations of sequencing data, a chemical conversion efficiency of 50% or greater was determined to be sufficient to confidently identify changes in new RNA by TimeLapse-seq [36]. To address all of these requirements, Jeremy and I employed several chemical and biochemical assays to characterize the efficiency and product specificity of the reaction, the results of which I will detail in Chapter 2. Optimal reaction conditions to transform s^4U were identified: s^4U is reacted to trifluoroethyl-substituted cytidine (from here out referred to as C*) by oxidative-nucleophilic-aromatic substitution [36].

The output of TimeLapse-seq data conveys both qualitative and quantitative information about RNA dynamics. Even when just considering the raw sequencing data, RNA dynamics become easily apparent: By coloring sequencing reads containing

increasing numbers of T-to-C mutations in increasingly darker shades of red, the proportion of reads stemming from new and pre-existing RNAs are easily distinguishable (red new, grey old). This approach readily shows whether a particular RNA is short-lived with a high turnover rate, meaning the majority of RNA present in the cell at the time of cell harvest was transcribed during the course of the experiment, (Figure 4, top *FOSL1*) or stable for a longer period of time (Figure 4, bottom *XIST*). Importantly, we can go beyond the raw data of the sequencing reads and use statistical modeling to quantify the fraction of reads stemming from newly-made RNA (from here on referred to as the fraction new), from which we can get statistically robust estimates of transcript-specific synthesis and degradation rates with their corresponding uncertainty. The statistical modeling of TimeLapse-seq data will be covered in greater depth in Chapter 5. To conclude, TimeLapse-seq captures RNA dynamics of short to long-lived mRNAs genome-wide.

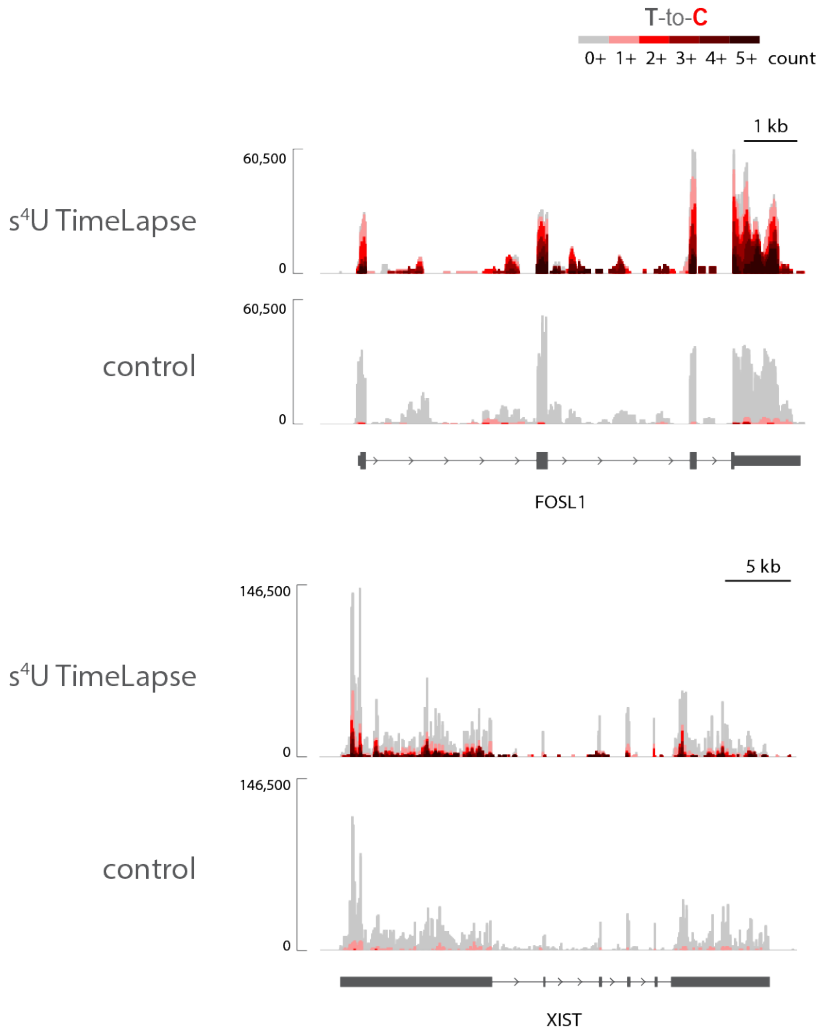

Figure 4: s^{4}U TimeLapse-seq captures RNA dynamics

Genome browser tracks of a TimeLapse-seq experiment performed in 293T cells (4h 100 μM s^{4}U , RNase-free water for control). Shown are both the output of TimeLapse-seq and the control RNA-seq experiment. TimeLapse-seq reads (red) are overlaid on top of total RNA-seq reads (grey) with reads containing increasing numbers of T-to-C mutations in increasingly darker shades of red. Shown are reads aligning to the fast turnover transcript *FOSL1* (top tracks) and the long-lived transcript *XIST* (bottom tracks). The T-to-C mutations detected in the control RNA-seq experiment are likely due to PCR and sequencing errors and are independent of TimeLapse treatment. I performed the experiments presented here.

1.3.2 Controls and Data Analysis

In this subsection I will cover the control experiments and decisions made previously in the lab when establishing s^4U TimeLapse-seq [36] as these considerations equally apply to my development of s^6G TimeLapse-seq covered in Chapters 2 and 3. After we had established that s^4U can act as a recodable nucleoside (more details provided in Chapter 2), Jeremy performed genome-wide TimeLapse-seq experiments of s^4U-treated cells. He showed that while s^4U alone leads to low levels of T-to-C mutations in the sequencing reads, with TimeLapse chemistry, a mutation rate of around 4-5% per uridine in mammalian systems can be achieved. The amine 2,2,2-trifluoroethylamine (TFEA) and the oxidant, $NaIO_4$, (TimeLapse chemistry) alone did not induce significant mutations over background for any nucleotide [36]. The oxidant $NaIO_4$ was chosen due to its common use in RNA biology, e.g. for 3'end labeling through oxidizing the 3'-end vicinal diols with minimal effects on other functional groups of the RNA [48]. Total transcript count between RNA-seq and TimeLapse-seq were found to be highly correlated indicating that neither the treatment with s^4U nor the subjection of extracted RNA to TimeLapse chemistry compromised the information gained from traditional RNA-seq experiments. The low level of T-to-C mutations detected in control experiments is independent of TimeLapse chemistry treatment and likely due to PCR and sequencing errors. Background mutations comprising of single nucleotide polymorphisms (SNPs) or over-represented nucleotides were determined from control RNA-seq experiments, and the flagged nucleotide positions were filtered from the TimeLapse-seq data. Data obtained from s^4U TimeLapse-seq reveals transcriptional dynamics genome-wide and captures RNA half-lives consistent with literature [4]. The development of this new technology aimed to study transcriptional dynamics without the need for biochemical enrichment of newly-synthesized RNA. TimeLapse-seq allows us to determine kinetic parameters from an internally normalized RNA-seq experiment, without having to adjust for background contamination and with the advantage that,

compared to enrichment approaches, very little total RNA input amount is required (tens of μg, ng, respectively).

1.3.3 Capturing the Transient Transcriptome by Combining Biochemical Enrichment and Nucleotide Recoding

While mRNA half-lives span from on the order of tens of minutes to hours, intronic, antisense and enhancer RNAs are turned over within minutes. To capture these transient transcripts a s^4U-pulse-based enrichment approach termed transient transcriptome sequencing (TT-seq) can be applied [30]. Cells are subjected to five minutes of s^4U and labeled RNA is extracted and enriched as described earlier. Unlike earlier versions of short s^4U treatments, TT-seq also includes a fragmentation step to enrich only fragments of RNA that are labeled. Former graduate student Erin Duffy applied the TT-seq approach with improved capturing chemistry (MTS-biotin [31]) to increase the specificity of the enrichment; however, the improved capture of s^4U-labeled RNA was associated with about 15-20% of background [36], similar to the original TT-seq protocol, motivating the addition of TimeLapse nucleotide conversion to identify bona fide new reads.

Due to the reversibility of the s^4U-based enrichment reaction (through reduction of the disulfide bond), the Simon lab saw the possibility of integrating the enrichment approach with TimeLapse-seq. This approach, termed TT-TimeLapse-seq, not only captures the transient transcriptome, but also enables estimation of background contamination in the enrichment. The presence of T-to-C mutations in the sequencing reads allows us to distinguish signal from noise leading to more accurate estimates of transient RNA dynamics. Since the initial proof of principle experiment performed by two former graduate students, Jeremy Schofield and Erin Duffy [36], other Simon lab members and I have optimized and streamlined the approach to allow for broader applications including primary cell samples of limited quantity. Collaboration projects

on capturing the dynamics of transient RNA species using TT-TimeLapse-seq that I was involved in, as well as ongoing work joined with other Simon lab members is covered in Chapter 4. To conclude, very short-lived transcriptional events can be accurately captured with a combination of reversible s^4U disulfide chemistry and nucleotide recoding.

1.4 Determining RNA Dynamics under Non-Steady State Conditions

The technologies outlined so far have been instrumental in studying RNA population dynamics at steady state. However, most important biological processes happen under non-steady state conditions, for instance during development or in reaction to an external stimulus. This section will outline published and ongoing work aimed at capturing RNA dynamics under complex kinetics.

Under non-steady state conditions the assumptions made to estimate kinetic parameters, such as the degradation rate constant, do not hold. Additionally, while some experimental setups are clearly governed by non-steady state conditions (e.g. transcriptional response to viral infection covered in Chapter 5) other situations are more nuanced. For instance, even under presumed steady state conditions, a particular RNA might exist in several subgroups within the cell. This could, on the one hand, be achieved by partitioning into membrane-bound or membrane-less organelles. Each of the partitions could have a different half-life, possibly resulting from the protective environment of the organelle. On the other hand, this particular RNA isoform might be under regulatory control by several different degradation pathways, such as miRNA or nonsense mediated decay degradation. However, when estimating half-lives from a TimeLapse-seq experiment, one is left with an averaged value for the specific RNA isoform which may not accurately reflect the complex kinetics. With the current

technologies, these complex regulatory mechanisms governing RNA abundance may easily be missed.

Furthermore, when estimating kinetic parameters it is assumed that these parameters are static, meaning that a given RNA has a fixed degradation rate regardless of its age. By modeling simple exponential decay and comparing the results to the observed RNA abundance, Deneke et al. were able to identify several RNAs in yeast that exhibit degradation rates which change with the age of the RNA [49]. This observation highlights the fact that the simple assumption of constant kinetic parameters even under cellular homeostasis can lead to false conclusions about RNA availability for protein synthesis. As described in their manuscript, Deneke et al. argue that parameters estimated from a single time point cannot accurately capture these complex kinetics. Instead, obtaining a snapshot of the transcriptome at several time points becomes necessary to gain accurate estimates.

The field has employed several approaches to address this need, one of which is using several RNA-seq experiments with staggered labeling times [40]. While this approach has been successfully used to, for instance, calculate elongation rates [32], this poses a new hurdle we initially eliminated with TimeLapse-seq: the issue of normalization between experiments when performing biochemical enrichment. It is evident that in order to assess complex non-steady state RNA dynamics, TimeLapse-seq and related approaches require the use of multiple experiments capturing different time points. These approaches would be simplified by the development of a second RNA metabolic label, which could be employed in a pulse-chase setup to capture multiple time points within one experiment. Challenges with establishing another RNA metabolic label include the need of a non-canonical nucleoside other than a uridine analogue that can not only be incorporated into cellular RNA, but also be recoded using existing nucleotide conversion technologies. In the following I will outline my aims to establish the metabolic label s^6G as a recodable nucleoside, and I will demonstrate that s^6G

TimeLapse-seq captured RNA dynamics genome-wide.

1.4.1 Technological Advancement - s^6G as a Recodable Nucleoside?

In order to capture multiple time points in the life of the cellular transcriptome in a single RNA-seq experiment, a second recodable nucleotide is needed. During the development of TimeLapse-seq, I explored the possibility of extending the recoding of s^4U to another RNA metabolic label.

While the most commonly used non-canonical nucleosides used to study RNA are derivatives of uridine, nucleotide conversion could hypothetically be extended to purines as well. Having access to a non-canonical nucleoside other than a uridine analogue may be beneficial and more appropriate in some experimental contexts, e.g. when studying the role of pseudouridines [50], U-tailing [51] or when determining half-lives of uridine-poor RNAs. Based on similar chemical properties, I investigated whether the hydrogen bonding pattern of s^6G can be recoded under TimeLapse conditions and whether recoding of s^6G can be used to capture RNA dynamics. I will describe the characterization of s^6G as a recodable nucleoside in Chapter 2 and my development of TimeLapse-seq with s^6G in Chapter 3. The ability to label cells with s^4U and s^6G and recode both non-canonical nucleosides under the same chemical conditions provides the opportunity to capture multiple time points within one TimeLapse-seq experiment. I will cover my ongoing work on "dual-color" TimeLapse-seq in Chapter 6.

1.4.2 Background on s^6G in Technologies and the Cellular Context

While the most widely used metabolic labels to study RNA population dynamics today are uridine analogues, early studies of artificial nucleosides suggested both s^4U

and s^6G can be incorporated into the RNA *in vivo* [21, 52]. Additionally, s^6G has also occasionally been employed in photocrosslinking experiments (PAR-CLIP [53]) and RNA structural studies [54]. More recent work suggests that while incorporation of s^6G into RNA *in vivo* results in changes in the thermodynamics, the presence of s^6G only results in minimal effects on RNA structure [55]. I therefore reasoned that RNA-friendly oxidative-nucleophilic-aromatic substitution, that we previously developed for s^4U, could be extended to recode s^6G to 2-aminoadenosine analogues. I hypothesized that the TimeLapse reaction conditions ($NaIO_4$ and TFEA) could also convert s^6G into an N6-substituted analogue of adenosine (explored in Chapter 2).

Unlike s^4U, whose natural counterpart uridine is phosphorylated to enter the nucleotide-triphosphate pool, guanosine as a purine is not thought to exist as an enzymatic substrate for phosphorylation [56]. In contrast, the purine *de novo* biosynthesis pathway involves the construction of the nuclear base guanine on the scaffold of 5-phosphoribose. Only upon completion of guanosine-monophosphate are additional phosphate groups added to result in GTP. The second major pathway of providing purine triposphates is through the purine biosynthesis *salvage* pathway. This pathway is thought to primarily use the nuclear base guanine as a substrate to react with phosphoribosyl disphosphate (PRPP) to yield GMP. This posed the question of whether treating cells with s^6G would result in similar incorporation levels as s^4U or whether we would have to take the *salvage* pathway route with the nuclear base analogue 6-thioguanine (6-TG). The next chapter describes the characterization of nucleotide recoding, with both my contribution to the establishment of s^4U recoding and my development of the metabolic label s^6G as a recodable nucleoside.

Chapter 2

Development of Recodable Nucleosides for TimeLapse-seq

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. Methods*, **15**, 221-225. doi: 10.1038/NMETH.4582

and

Kiefer, L., Schofield, J.A., Simon, M.D. (2018) Expanding the Nucleoside Recoding Toolkit: Revealing RNA Population Dynamics with 6-Thioguanosine. *J. Am. Chem. Soc.*, 140, 14567-14570. doi: 10.1021/jacs.8b08554

2.1 Author contributions

Jeremy Schofield and Matthew Simon designed the experiments for the development and application of s^4U TimeLapse-seq. Jeremy Schofield conducted the majority of the experiments. I performed characterization experiments (NMR) on the nucleoside recoding TimeLapse reaction of 4-thiouracil and I was working closely with Jeremy Schofield on the chemical and biochemical characterization assays. I conducted and

analyzed all experiments involving s^6G and 6-TG.

2.2 Summary

RNA-seq is a powerful tool to study the relative abundance of RNA in a biological sample genome-wide. However, as an endpoint assay, most temporal aspects of the transcriptome are lost. As outlined in Chapter 1, synthesis and decay rates under steady state conditions can be determined if newly-made RNA is distinguishable from pre-existing RNA. Under non-steady state conditions, multiple snapshots of the transcriptome need to be captured requiring the development of a second recodable nucleoside compatible with TimeLapse chemistry. This chapter outlines the development and considerations of recodable nucleosides for enrichment-free separation of new from old RNA (TimeLapse-seq). In order to identify reads from metabolically labeled RNA through nucleoside recoding, the chemical reaction must be orthogonal, efficient, and clean. This chapter contains my contributions to the development of s^4U TimeLapse-seq through the characterization of chemistry of s^4U recoding and my development of s^6G as a recodable nucleoside.

2.3 Characterization of the s^4U Recoding Reaction

As introduced in Chapter 1, a s^4U to C* reaction efficiency of around 50% is sufficient to ensure confident identification of changes in new RNA. This requires that the reaction from s^4U to C* must be efficient and clean, meaning that a large fraction of s^4U is reacted to C* and that the reaction mainly yields C* without byproducts that could lead to non-specific or undetectable mutations. To address the latter, I employed ^{1}H-NMR, which detects the local magnetic field of protons providing

structural information for all molecular species present in the NMR tube. This approach, therefore, not only allows us to investigate whether the expected product is formed, but also provides information about purity and any additional products formed by the reaction. I decided to analyse the conversion reaction on the nuclear base of s^4U, 4-thiouracil (4-TU), due to fewer protons and therefore a simpler spectrum. I compared the NMR spectra for 4-TU alone to 4-TU subjected to TimeLapse reagents TFEA and $NaIO_4$ (for details on the method refer to Chapter 7). Surprisingly, I found a clean and complete conversion from starting material, 4-TU, to the expected product, 2,2,2-triluoroethyl-substituted cytosine within the course of 4h. Aside from leftover amine, no additional peaks were observed, suggesting a clean reaction without the formation of side products (Figure 5). The findings also suggest that the reaction of 4-TU to a substituted cytosine proceeds efficiently over the course of 4h and under the given reaction conditions (in DMSO). Since the chemical reactivity of 4-TU could differ from that of s^4U and since the reaction was not done under standard RNA handling conditions (aqueous), Jeremy went on to identify products of the reaction of s^4U under TimeLapse chemistry conditions by liquid chromotography mass spectrometry (LC-MS) and developed a biochemical assay to monitor the reaction in *in vitro* transcribed RNA [36] (assays described in more detail in the following sections). To conclude, through careful chemical and biochemical characterization, Jeremy and I showed that under TimeLapse reaction conditions s^4U is converted to C* in an efficient and clean reaction.

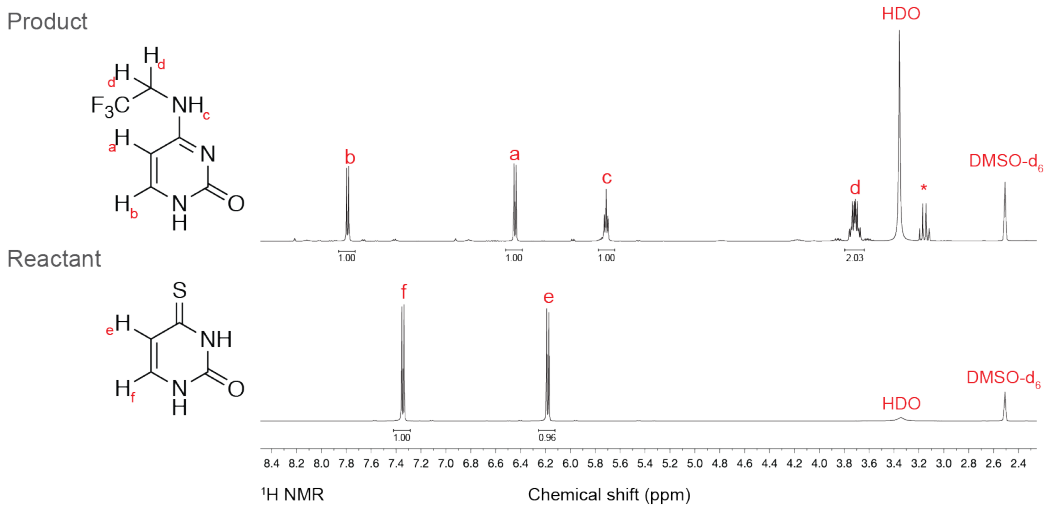

Figure 5: TimeLapse chemistry causes clean conversion of 4-thiouracil

^{1}H-NMR spectra of 4-thiouracil with (top) and without (bottom) 4h of TimeLapse chemistry (2,2,2-trifluoroethylamine, sodium periodate, 45°C) with integrations displayed below peaks. Peaks are assigned to the structures displayed (left). Peaks corresponding to amide protons are shifted downfield and are not shown in the above view. The solvent peak is indicated by DMSO-d_6 and the water peak by HDO. *Multiplet assigned to excess amine (2,2,2-trifluoroethylamine). I performed and analyzed the experiments presented here.

2.4 Assessing s^4U and s^6G Incorporation into RNA in Cells

Before testing whether s^6G can be efficiently and cleanly recoded to an analogue of adenosine, I established the extent of s^6G incorporation into cellular RNA. As s^4U has been used for decades to label cellular RNA in various cell systems (reviewed in [39]), I decided to use the metabolic label as a positive control of sufficient RNA incorporation. As described in Chapter 1, there is precedent for using s^6G to metabolically label RNA *in vivo* [21, 53]; however, the molecular mechanisms and therefore optimal form of label (nuclear base or nucleoside) remained to be determined. To address the question of incorporation efficiency and ideal delivery form, I developed an activated disulfide fluorescence labeling assay (adapted from MTS-biotin enrichment chemistry [31]). Briefly, cells are subjected to the metabolic label by addition to the cell media for a given amount of time (2h in this case), and 5 μg of extracted cellular RNA is reacted with activated disulfide conjugated to a fluorophore (MTS-TAMRA). Fluorescence is then measured and compared to that of control unlabeled RNA (Figure 6A). I found that under the same treatment conditions (concentration and time), s^6G and 6-TG are incorporated into cellular RNA to similar levels (fluorescence over background), albeit lower than s^4U (Figure 6B). Quantifying the signal over background resulted in a approximately 2.5-fold lower incorporation of either s^6G and 6-TG compared to s^4U (Figure 6C). I speculate that as guanosine is the largest of the four nuclear bases, the substitution of the oxygen at position 6 with the larger sulfur atom could lead to steric hindrance in the active site of polymerase II, which would result in lower incorporation levels of s^6G. Alternatively, s^6G may be less bioavailable, less efficiently incorporated into the pool of nucleotide triphosphates or the electronics of s^6G might negatively influence its incorporation rate. Similarly, it is unclear why incorporation levels of s^6G and 6-TG are similar. This observation suggests that s^6G could be an acceptable

substrate for phosphorylation to s^6GTP through the purine *salvage* pathway.

To assess the kinetics of label availability for incorporation into RNA, I repeated the assay as a time series and analyzed incorporation by relative fluorescence. I found that s^4U can be detected in RNA within ten minutes of labeling, with the signal increasing with labeling time (Figure 7A). I found similar kinetics of incorporation for both s^6G and 6-TG, albeit overall lower than s^4U, with nevertheless detectable incorporation after ten minutes and increasing incorporation over time (Figure 7B). I screened a range of s^6G and 6-TG concentrations and found no significant difference between 10 μM - 1 mM (data not shown). Additionally, I screened for cell viability of 6-TG under these concentrations and found no significant effect during the time frame of experimental treatment (Figure 8 top). To conclude, s^6G is incorporated into RNA in cells and there is no obvious advantage to using its nuclear base 6-TG. I determined that using 100 μM 6-TG or s^6G seems ideal for RNA labeling in the tissue culture cell lines used for all the following experiments and that the metabolic labeling does not affect cell survival.

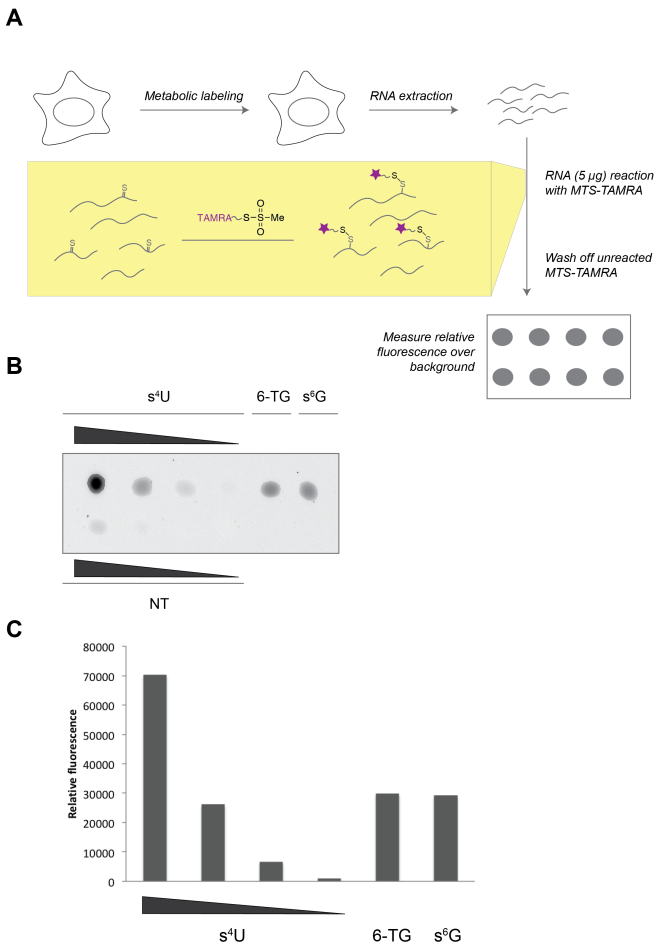

Figure 6: Measuring s^6G and 6-TG incorporation into cellular RNA by dot blot

(A) 293T cells were treated with 200 μM of s^4U, s^6G, 6-TG or water for 2 h. RNA was extracted as described above. RNA (5 μg) was reacted with MTS-TAMRA. Fluorescence intensities were determined using a GE Healthcare Typhoon FLA 9500. Depicted are dilution series of s^4U (1:4) and no treatment control (1:4), including undiluted RNA from 6-TG and s^6G treated cells (B) including quantification of the fluorescent signal (C) indicating a 2.5- fold lower incorporation of both 6-TG and s^6G compared to s^4U under the same concentration and time. This observation agrees with sequencing results presented later on indicating an average mutation rate of 4% per U in s^4U TimeLapse-seq and 1.5% per G in s^6G TimeLapse-seq (2.7 fold difference). I performed the experiments presented here.

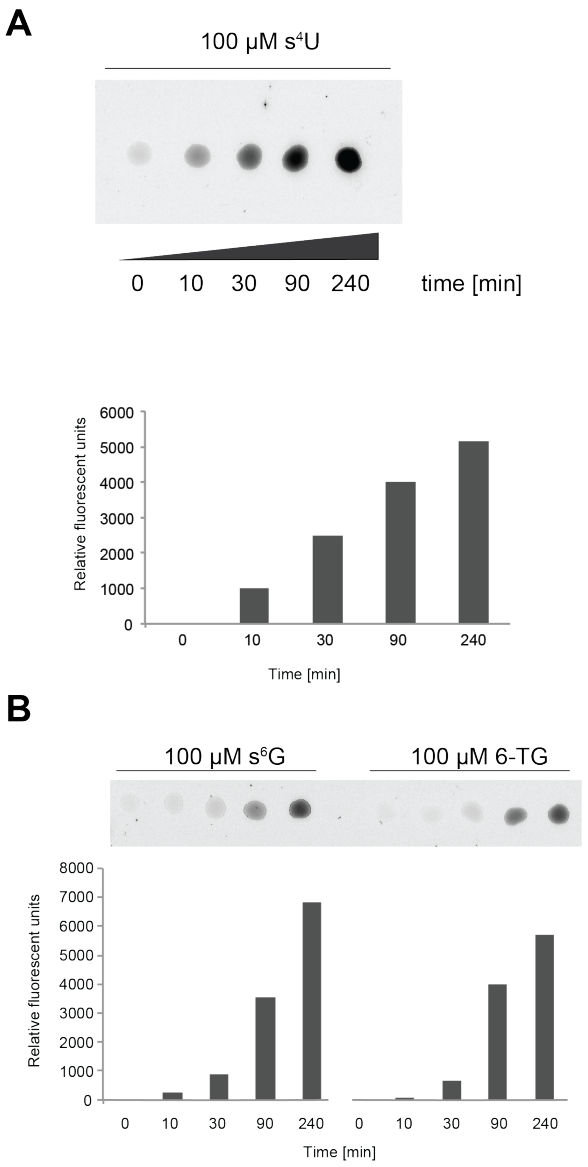

Figure 7: Incorporation kinetics of s^4U, s^6G and 6-TG by dot blot

Cells were treated with either s^4U, s^6G or 6-TG for the indicated time period. Extracted RNA was reacted with MTS-TAMRA and fluorescent signal was quantified (assay procedure see Figure 6). I performed the experiments presented here.

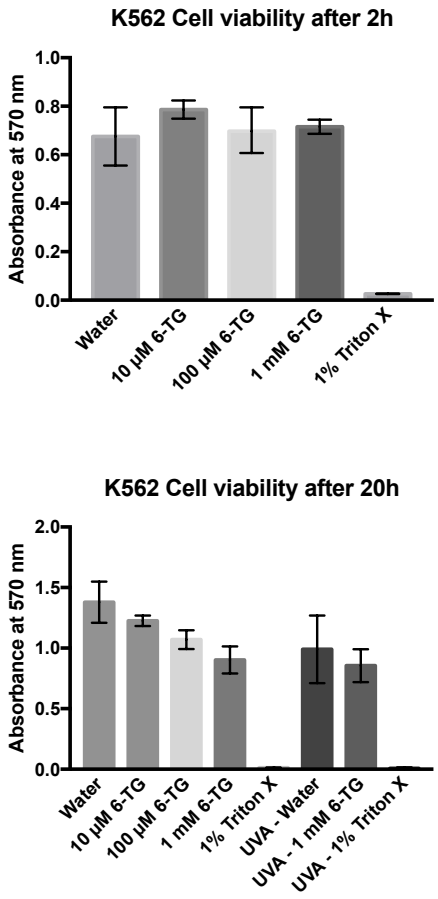

Figure 8: 6-TG does not affect cell viability over the course of a TimeLapse-seq experiment

K562 cells were treated with increasing amounts of 6-TG, water and 1% Triton X, as a negative control. Additionally, the cells were treated with UVA radiation as 6-TG is thought to be photoreactive under these conditions. Cell viability was assessed using the ATCC MTT Cell Proliferation Assay kit following manufacturer's instructions. I performed the experiments presented here.

2.5 Characterization of s^6G as a Recodable Nucleoside

In order to address whether s^6G can be recoded into an analogue of adenosine and whether I can get the reaction to proceed efficiently and cleanly, I first used ^{1}H-NMR analysis. To identify the product of s^6G subjected to TimeLapse chemistry (TFEA, $NaIO_4$), I again resorted to the nuclear base to limit the number of protons. Excitingly, the reagents led to a clean conversion of 6-TG to *N*6-trifluoroethyl substituted 2,6-diaminopurine, the predicted product (Figure 9). The data provided evidence that TimeLapse chemistry not only recodes 6-TG, but also gives an efficient and clean conversion within one hour under the given reaction conditions.

To assess whether recoding is possible under standard RNA handling conditions, I first examined the chemistry using just the nucleoside and conducted a time course experiment surveying reactants and products by LC-MS. In LC-MS, molecules are separated on a solid phase column based on their hydrophobicity and subsequently analyzed by mass spectrometry. Figure 10 shows the extracted ion chromatograms using the exact mass of the starting material and the predicted product. Since $NaIO_4$ treatment leads to the oxidation of 3'-end vicinal diols, I decided to use the deoxynucleoside of s^6G instead. I found that 6-thio-2'-deoxyguanosine (s^6dG) nucleoside was consumed within 5 min and the corresponding 2-aminoadenosine analogue (hereafter referred to as dA*) was produced within 1h (Figure 10). Furthermore, to address whether recoding proceeds with s^6G in the context of a biomolecule I modified a previously established restriction digest assay using a double stranded DNA substrate containing s^6dG to screen for conditions able to convert s^6dG to dA*. Due to documented challenges incorporating s^6G into RNA using prokaryotic RNA polymerases for *in vitro* transcription [57] I adapted the assay for double stranded DNA. A single s^6dG is incorporated into one strand of a double stranded DNA oligonucleotide. The s^6dG

is positioned at a site that creates an endonuclease restriction site upon successful recoding to dA* and PCR amplification. The resulting conversion is read out by the extent of restriction digestion of the oligonucleotide, with successful cutting as a proxy for successful conversion of s^6dG to dA*. The majority ($\sim$ 59%) of the nucleotide s^6dG was recoded to dA* in the context of a DNA duplex using ammonia in TimeLapse chemistry (Figure 11). Similar results were obtained using different amines, including TFEA (Figure 12). In summary, I was able to show that s^6G is reacted to A* using TFEA and $NaIO_4$ in an efficient and clean reaction without byproducts and in the context of a nucleic acid polymer. Additionally, the conversion efficiency of 59% in the context of DNA (Figure 11) suggests high enough efficiency to detect changes in the new RNA population as determined through simulation previously [36].

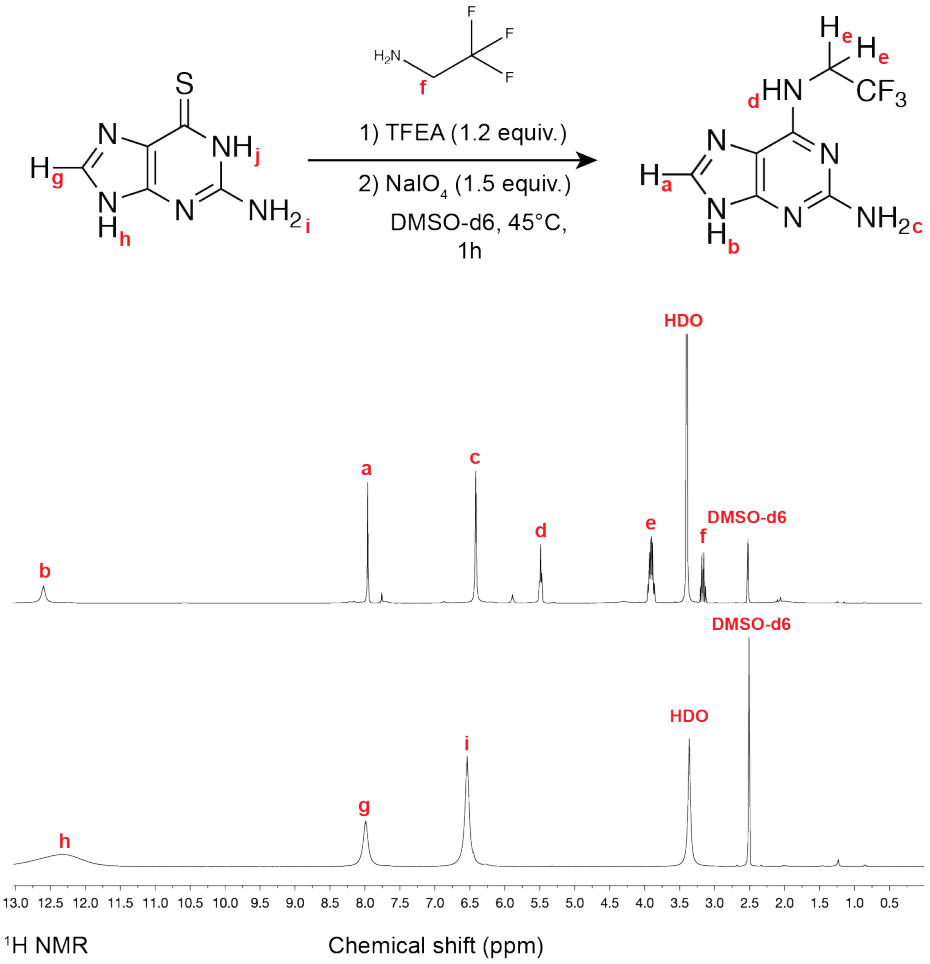

Figure 9: 6-thioguanine is recoded to a 2-amino-adenine analogue under TimeLapse conditions

6-TG was dissolved in deuterated DMSO and analyzed by ^{1}H-NMR with (top) and without (bottom) TimeLapse chemistry. I performed and analyzed the experiments presented here.

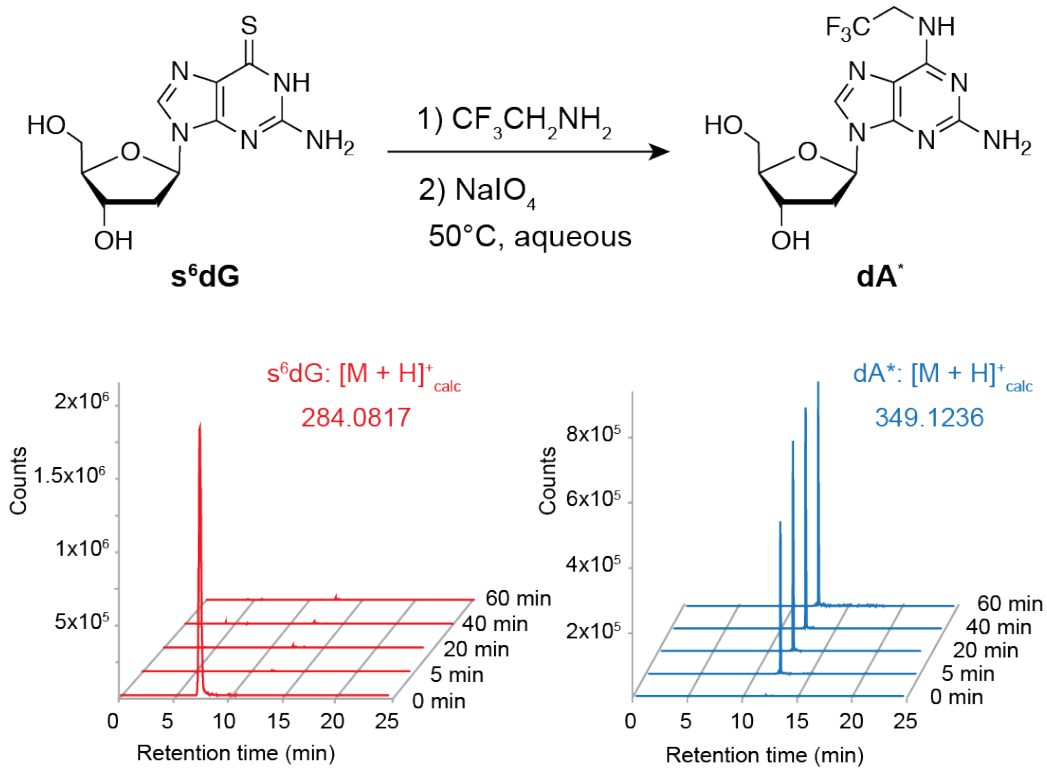

Figure 10: s^6G is efficiently recoded to A* under TimeLapse conditions

Extracted ion chromatograms corresponding to the masses of 6-thio-2'-deoxyguanosine and 2-amino-6-(2,2,2-trifluoroethyl)-amino-2'-deoxyadenosine. I performed the experiments presented here.

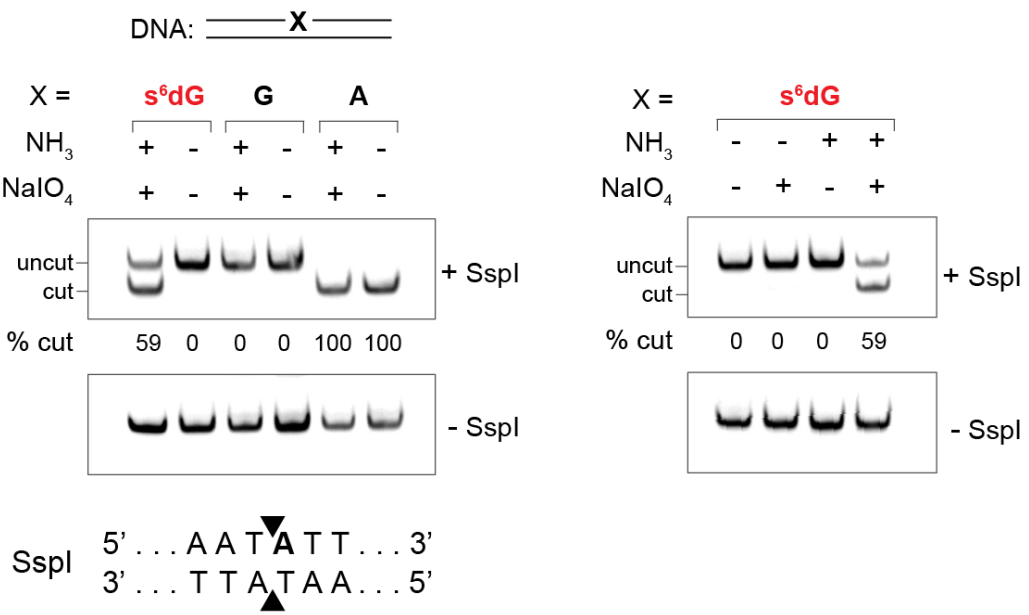

Figure 11: s^6dG is recoded to dA* under TimeLapse conditions in the context of dsDNA

Restriction digest assay of a DNA duplex containing a single 6-thio-2'-deoxyguanosine, treated with 600 mM NH_3 and 10 mM $NaIO_4$ for 1 h at 45 °C and subsequent digestion by SspI restriction enzyme (AAT<u>A</u>TT) and analysis by Native-PAGE. I performed the experiments presented here.

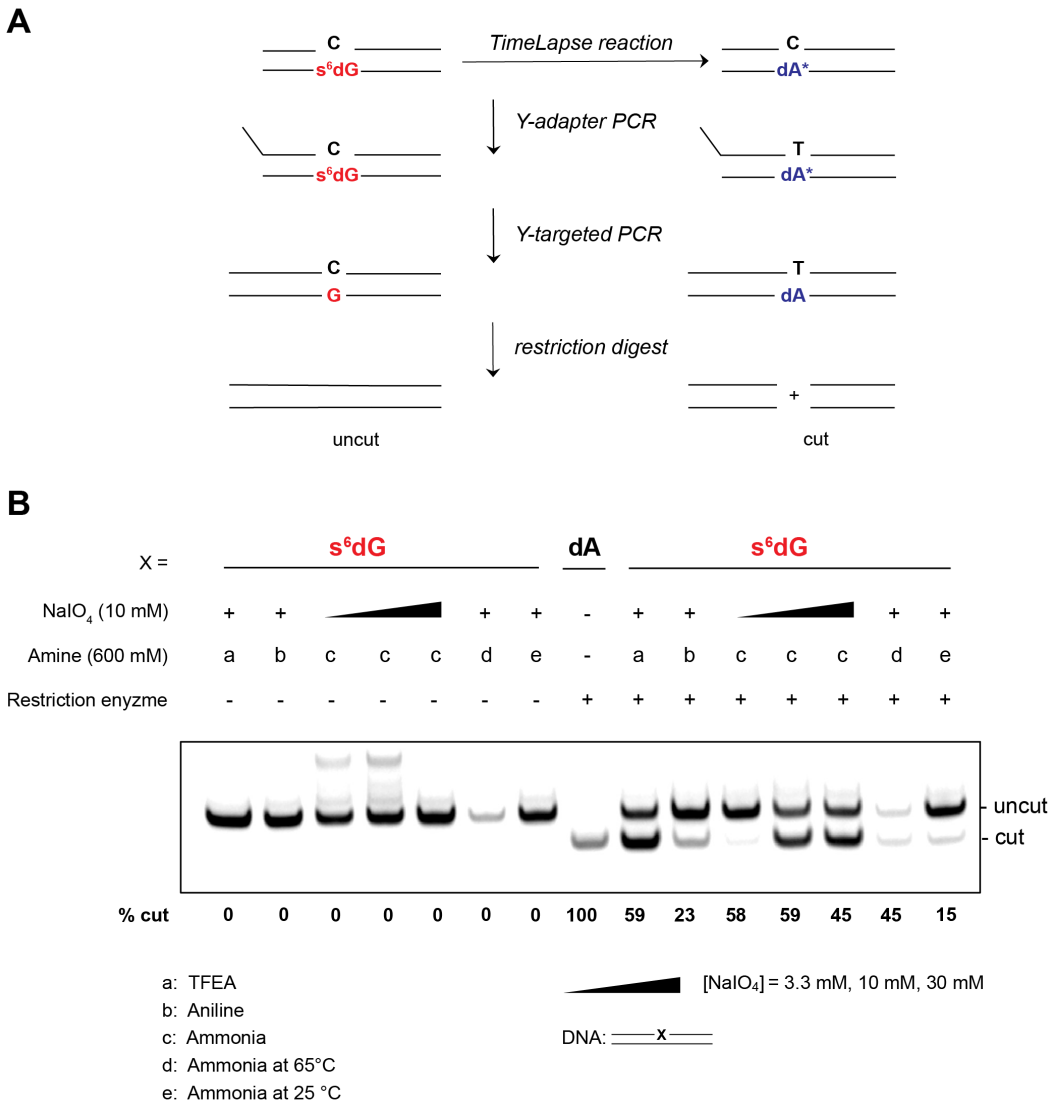

Figure 12: Screen for optimal s^6G recoding conditions by restriction digest assay

(A) Scheme of the assay in which a single s^6dG is incorporated into double stranded DNA by substituting s^6dGTP for dGTP. The DNA duplexes are subjected to TimeLapse chemistry (amine, oxidant) under the specified time and temperature conditions. Using an adapter-specific fluorescent oligo, the conversion of s^6dG is read out by incorporation of a T (successful conversion) or a C (unsuccessful conversion) during the following PCR step, creating a restriction site for SspI restriction enzyme, or not, respectively. (B) Conditions screen for TimeLapse chemistry to convert s^6dG to dA* in the context of a DNA duplex. All TimeLapse reactions were run at 45°C

(unless indicated differently) for 1 h. Subsequent restriction digest reactions were run at 37°C for 1 h. Percent cut determined by ImageJ is reported below the gel. A positive control DNA duplex containing dA is used to indicate the position of the restriction digested products. I performed the experiments presented here.

2.6 Conclusion

In order to determine RNA half-lives in an RNA-seq experiment using nucleotide recoding, the conversion reaction needs to be efficient and clean. In this chapter I highlighted my contribution to characterizing the reaction of s^4U to trifluoroethyl-substituted cytidine and the reaction of s^6G to trifluoroethyl-substituted-2-amino-adenosine using a variety of chemical and biochemical assays. Both reactions are efficient and do not produce any byproducts and are therefore suitable for the use of detecting RNA dynamics through sequencing. The next chapter will go into detail on how RNA dynamics are inferred using the newly developed recodable nucleoside s^6G.

Chapter 3

Revealing RNA Population Dynamics with 6-Thioguanosine

This chapter contains excerpts from:

Kiefer, L., Schofield, J.A., Simon, M.D. (2018) Expanding the Nucleoside Recoding Toolkit: Revealing RNA Population Dynamics with 6-Thioguanosine. *J. Am. Chem. Soc.*, 140, 14567-14570. doi: 10.1021/jacs.8b08554

3.1 Author contributions

I performed all the experiments presented in this chapter with assistance on some of the characterization experiments from Jeremy Schofield. I performed the bioinformatic analyses with assistance from the lab.

3.2 Summary

As outlined in Chapter 1, the synthesis, processing and decay of RNA form various points of regulation that ensure controlled gene expression. In order to understand the underlying molecular mechanisms at play when observing a change in RNA abundance,

we need to be able to determine which kinetic parameter or parameters caused the change. In the previous two chapters I outlined different strategies that have been employed to capture RNA dynamics, including biochemical enrichment of newly-made RNA and, more recently, nucleotide recoding of s^4U. As discussed in Chapter 1, the most commonly used RNA metabolic labels are analogues of uridine and so far, all nucleotide conversion sequencing techniques have been limited to pyrimidines (T-to-C, SLAM-seq [37], TimeLapse-seq [36]; C-to-T bisulfite-seq [58]). Additionally many important biological processes occur under non-steady state conditions, which necessitates capturing multiple snap shots of the transcriptome in order to determine accurate estimates of RNA half-lives. Given all of these considerations, I sought out to extend nucleotide recoding with TimeLapse-seq to the guanosine analogue, 6-thioguanosine (s^6G). In the previous chapter I described the development and characterization of s^6G as a recodable nucleoside. Throughout this chapter, I will offer a thorough analysis of s^6G recoding as read out by sequencing of cellular RNA and the estimation of RNA half-lives using s^6G TimeLapse-seq.

3.3 Assessing s^6G Recoding Efficiency and Orthogonality in Cellular RNA

In the previous chapter, I was able to show that s^6G is incorporated into RNA in cells and that TimeLapse chemistry leads to the reaction of s^6G to A* on the nucleoside level. The important next question was whether this reaction would indeed lead to a recoding of the hydrogen bonding pattern detectable by sequencing. In order to address this question I took a targeted sequencing approach by PCR amplifying the transcript *ACTB* following TimeLapse chemistry. The cells were grown for 1 h to allow time for incorporation of s^6G into newly synthesized RNA. I did not observe significant toxicity even after a 2 h treatment, consistent with previous reports for s^6G [53] (Figure

8). Total RNA was then isolated and subjected to TimeLapse chemistry, followed by targeted reverse transcription (*ACTB* mRNA) and next generation sequencing. Sequencing reads were mapped to the target transcript, and the mutations of each nucleotide to adenosine were counted. I found that s^6G is incorporated into newly transcribed RNA and converted into A* as inferred from the increase in G-to-A mutations at all G nucleotides that were analyzed (Figure 13B). This conversion is 6-TG treatment and TimeLapse chemistry-dependent (Figure 14), and orthogonal with only low levels of additional G-to-T mutations detected (Figure 13A), which are most likely induced through G oxidation to 8-oxo-G as observed previously [59]. Preliminary data collected by the lab indicates that, while other approaches can be used to detect s^4U directly through sequencing using an alkylation reaction (SLAM-seq [37]), using this approach for s^6G yields insufficient G-to-A mutations to infer RNA population dynamics. This is, to the best of my knowledge, the first evidence that s^6G recoding can be detected by sequencing in the context of cellular RNA.

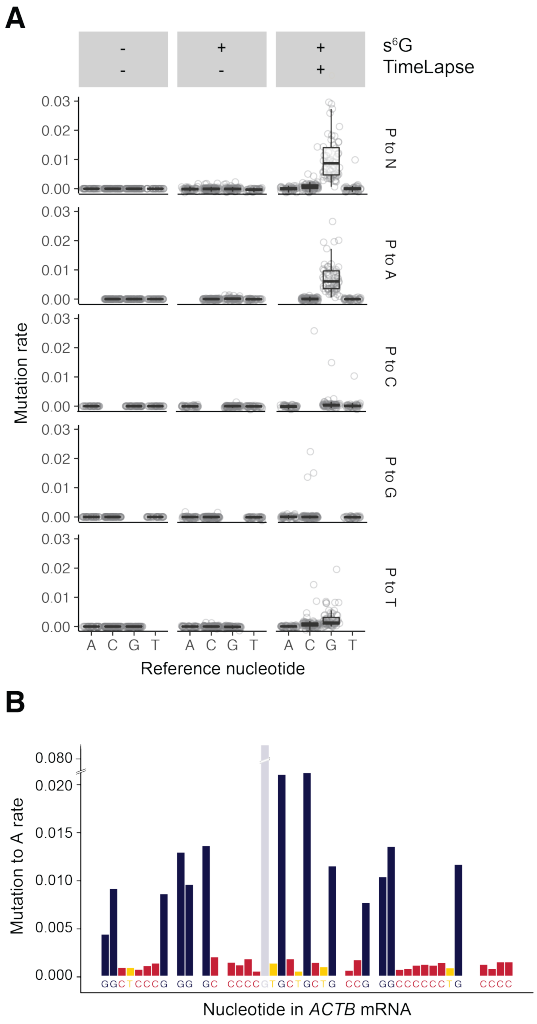

Figure 13: s^6G and TimeLapse chemistry-dependent increase in G-to-A mutations shows orthogonality of the reaction

K562 cells were treated with 500 μM 6-TG and total RNA was extracted and subjected to TimeLapse chemistry. (A) Plotted are the background subtracted mutation rates of C, G and T to each of the four nucleotides (PtoA, PtoC, PtoG, PtoT) and to any of the four nucleotides (PtoN). (B) Bar plot of a region of *ACTB* mRNA indicating the mutation rate to A for each nucleotide (not shown are positions of A). The light gray bar represents a site of a likely SNP that was identified due to its high G-to-A mutation rate in control samples. I performed the experiments presented here and conducted the data analysis with assistance from the lab.

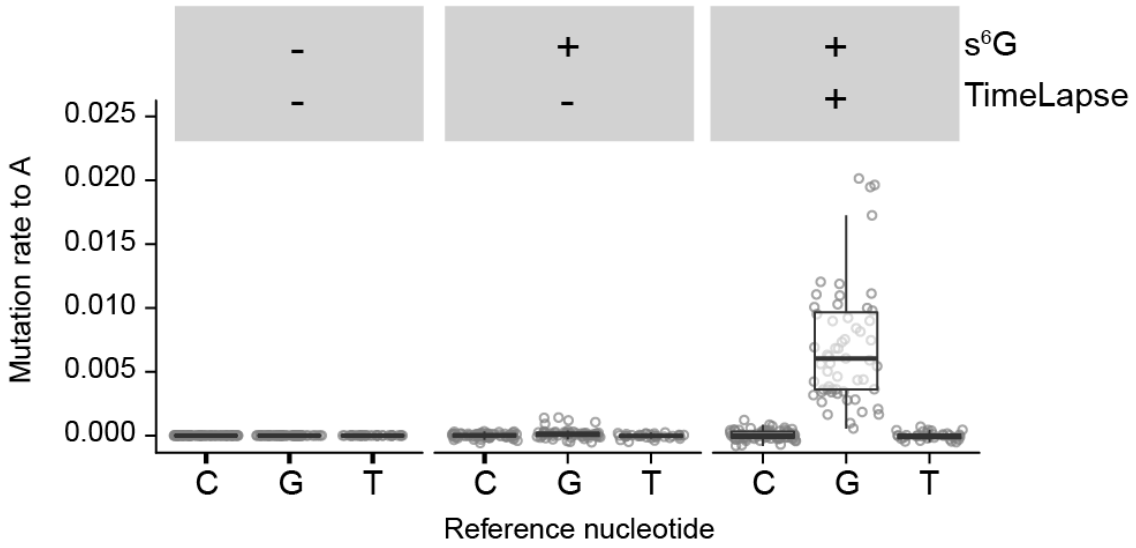

Figure 14: s^6G and TimeLapse chemistry-dependent increase in G-to-A mutations in sequencing reads of *ACTB*

K562 cells were treated with 500 μM 6-TG and total RNA was extracted and subjected to TimeLapse chemistry (600 mM TFEA, 10 mM $NaIO_4$ 1h at 45°C). Plotted is the quantification of N to A transitions in *ACTB* mRNA PCR amplicons. I performed the experiments presented here and conducted the data analysis with assistance from the lab.

3.4 Capturing RNA Dynamics Genome-wide with s^6G Recoding

Finally, to address whether s^6G can capture transcriptional dynamics genome-wide, I performed TimeLapse-seq on s^6G metabolically labeled cellular RNA. The experimental approach is outlined in the schematic in Figure 15. I labeled K562 cells for 4 h, a labeling time optimized for studying the half-lives of mRNAs [38]. Next, I extracted total RNA, treated the RNA with TimeLapse chemistry and subjected it to sequencing. Once the reads were mapped to the transcriptome, I tested whether the 6-TG treatment substantially impacted RNA levels. Expression analysis revealed that the RNA levels from 6-TG-treated cells were highly correlated with the levels from untreated cells, indicating that the process of metabolic labeling with 6-TG does not substantially impact the transcriptome in the time frame of the experiment (Figure 16, s^6G vs. untreated Pearson's $r \geq 0.97$; TimeLapse-treated vs. untreated RNA Pearson's $r \geq 0.98$). Qualitative analysis of the sequencing reads revealed a notable increase in G-to-A mutations compared to control and also revealed the single molecule nature of this data (Figure 17, zoom in on the location of G-to-A mutations in single sequencing reads). Just from the raw sequencing reads, it was apparent that transcriptional dynamics were captured genome-wide, for instance when examining individual transcripts (Figure 18); transcripts with short half-lives such as *JUN* had higher numbers of reads with G-to-A mutations (higher number of blue to grey reads) than did stable transcripts such as *GAPDH*. As with targeted sequencing, this increase in the G-to-A mutation rate was only found in RNA from cells that had been treated with 6-TG and not in control experiments (Figure 17, bottom track). The increase in G-to-A mutations also tracked with transcript half-lives (as determined using s^4U TimeLapse-seq). Transcripts that have the shortest half-lives demonstrate the greatest increase in G-to-A mutations (Wilcox test, $p < 10^{-15}$) and even transcripts with long half-lives had a significant

increase in G-to-A mutations upon 6-TG treatment (Wilcox test, $p < 10^{-15}$) (Figure 20A). The comparison to half-lives estimated by s^4U TimeLapse-seq validates that the information gained from G-to-A mutations is consistent with existing half-life estimates. s^6G TimeLapse-seq therefore provides qualitative information of RNA turnover genome-wide.

Next I harnessed the single molecule nature of TimeLapse-seq to ask quantitative questions about RNA half-lives. Each read from a TimeLapse-seq experiment reports on the mutational content of a single molecule of RNA that is either new (labeled) or pre-existing (unlabeled). We previously developed a statistical analysis of nucleotide recoding data using a binomial distribution to model the read distribution [36]. A more detailed description of the statistical modeling as well as updated analysis are discussed in Chapter 5. Based on the number of G-to-A mutations in each read and accounting for new reads lost during handling, I calculated the fractions of new RNA that were produced during the treatment for over 4000 transcripts. These fractions of new reads were reproducible across replicates (Figure 19, Pearson's $r = 0.92$). Assuming simple exponential kinetics, I estimated RNA half-lives using these fractions and these correlated well with results using s^4U TimeLapse (Figure 20B). I found that known fast turnover transcripts such as transcription factors (e.g. *JUN*) have significantly shorter half-lives than *SMG5* or slow turnover transcripts such as *GAPDH* (Figure 18, $t_{1/2}$ estimates above tracks), consistent with previous reports [36]. Notably, using TimeLapse-seq with 6-TG allowed me to estimate the half-lives of uridine-poor transcripts such as *CBX4*, whose reads have on average 10 uridine nucleotides, but 60 guanosine nucleotides (Figure 21). While low uridine content will impact the certainty in the estimates of kinetic parameters, it is unlikely to have a significant impact on the median value of the estimate. In the case of *CBX4* and other regions of the genome with low U contents, such as many intronic regions, s^6G can be employed to increase our confidence in the kinetic parameters determined. In

summary, s^6G recoding captures transcriptional dynamics genome-wide in both a qualitative and quantitative manner.

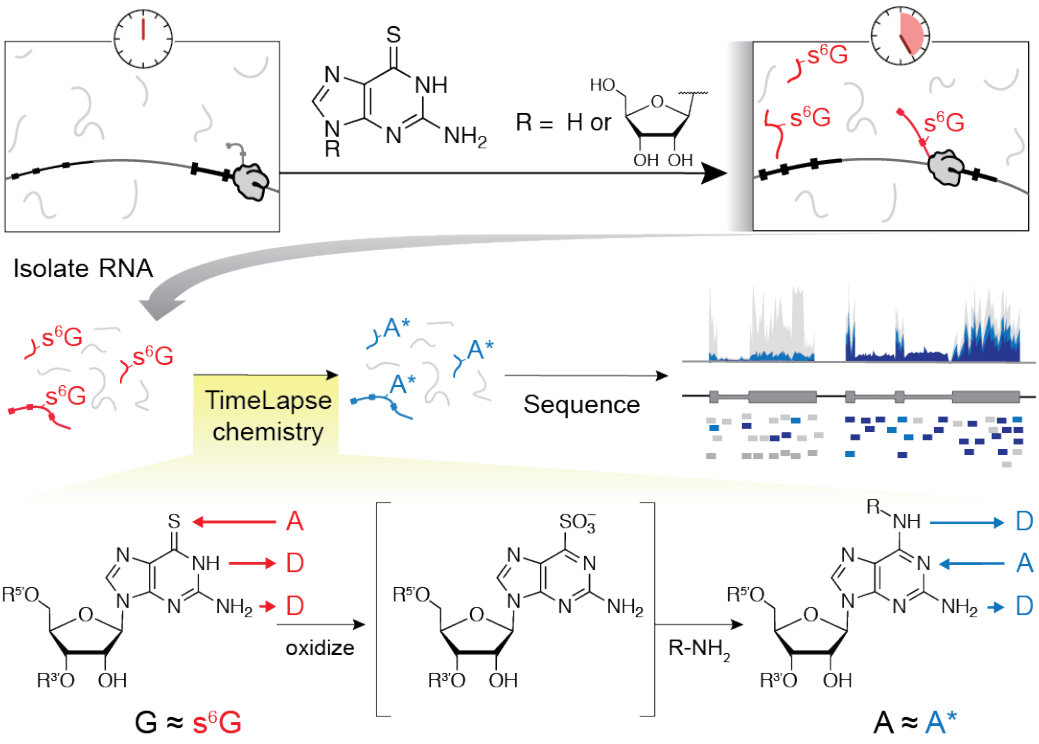

Figure 15: s^6G TimeLapse-seq approach

Cells are metabolically labeled with s^6G or 6-TG, total RNA is isolated and subjected to TimeLapse chemistry to induce nucleotide conversion. Total RNA is subsequently reverse transcribed and sequenced. Sequencing reads with increasing numbers of G-to-A mutations are shaded in increasingly darker shades of blue, while total RNA-seq reads are shown in grey.

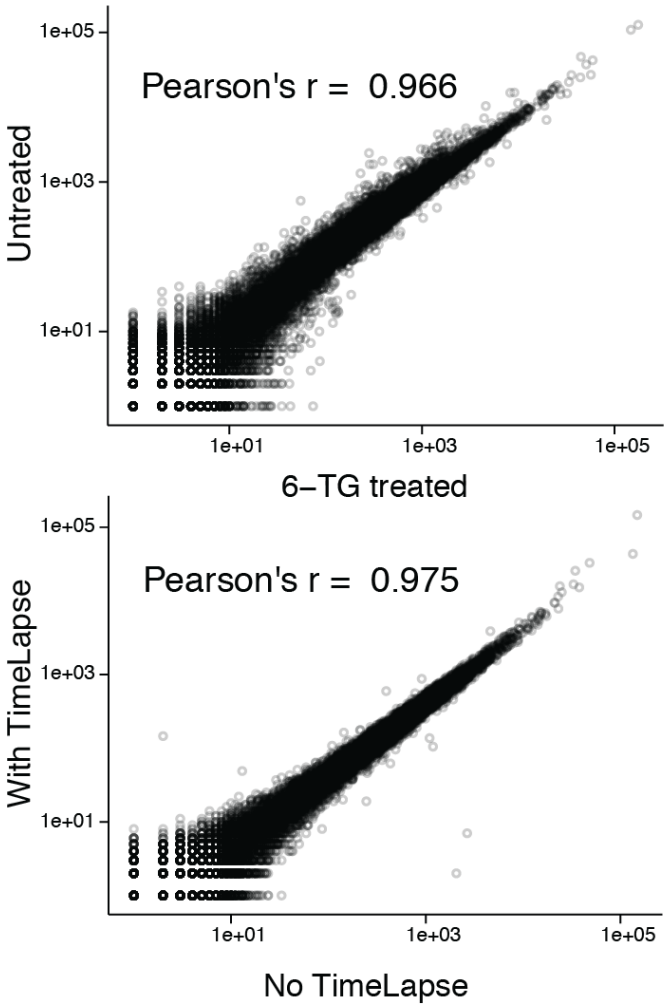

Figure 16: 6-TG treatment and TimeLapse chemistry do not affect RNA-seq output

Correlation plots of RNA-seq profile comparing 6-TG versus untreated cells (top) and TimeLapse chemistry treated and untreated RNA (bottom) with Pearson's correlation coefficient reported on each plot. I performed the experiments presented here and conducted the data analysis with assistance from the lab.

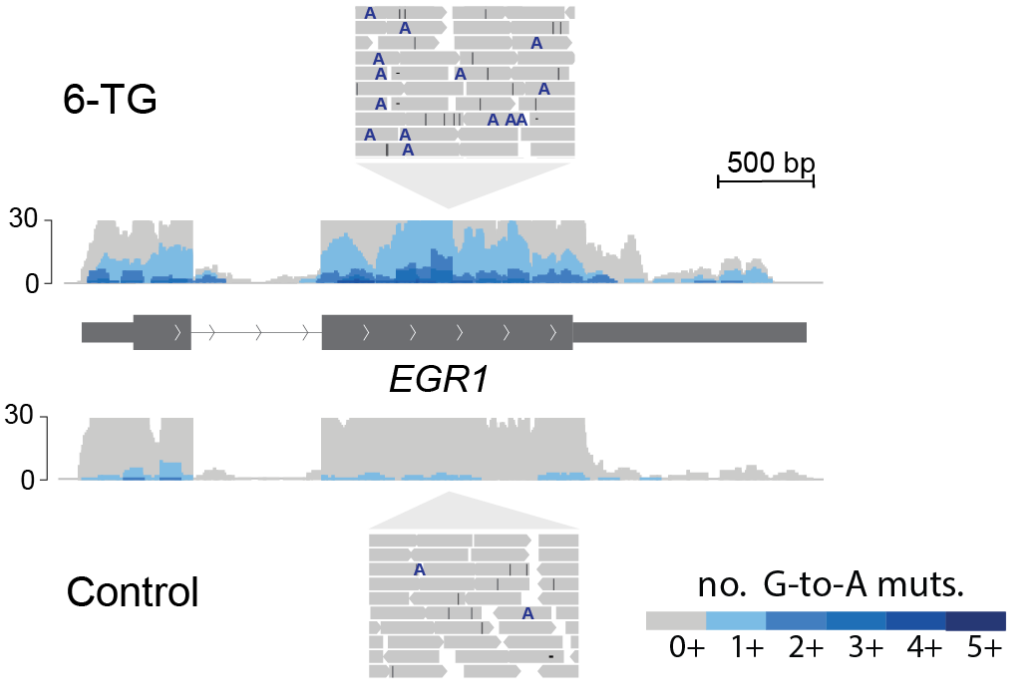

Figure 17: TimeLapse-seq with s^6G leads to increases in G-to-A mutations

Metabolic labeling of cellular RNAs leads to G-to-A mutations at sites of s^6G in sequencing reads shown for a region of the *EGR1* transcript.

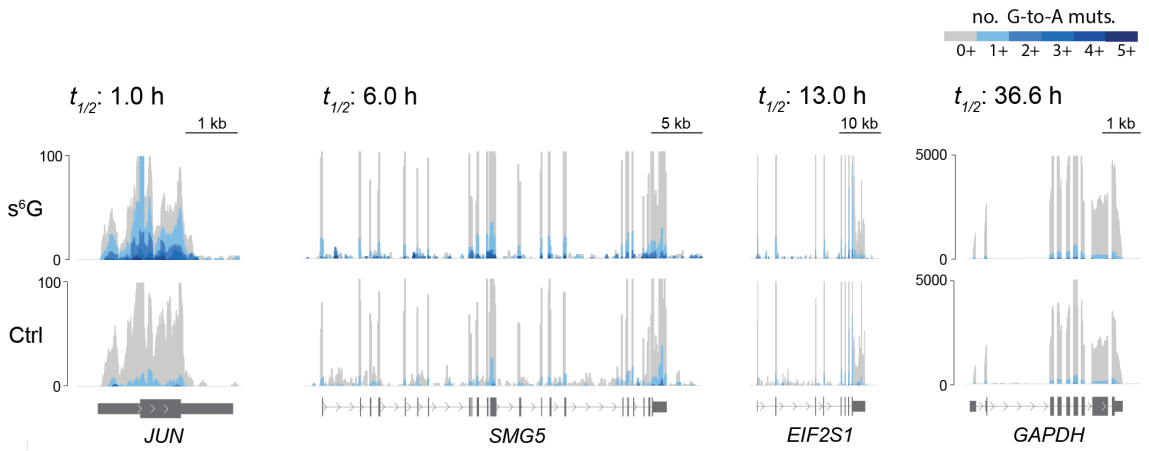

Figure 18: s^6G captures transcriptional dynamics genome-wide qualitatively and quantitatively

Genome browser tracks for representative fast (*JUN*), intermediate (*SMG5* and *EIF2S1*), and slow (*GAPDH*) turnover transcripts colored by the cumulative number of G-to-A mutations. Gray tracks represent all RNA-seq reads, with blue tracks representing the profile when only reads with the indicated number of G-to-A mutations are considered. Estimated half-lives determined through statistical modeling of the mutational data using a mixed Binomial model are indicated above each track. I performed the experiments presented here and conducted the data analysis with assistance from the lab.

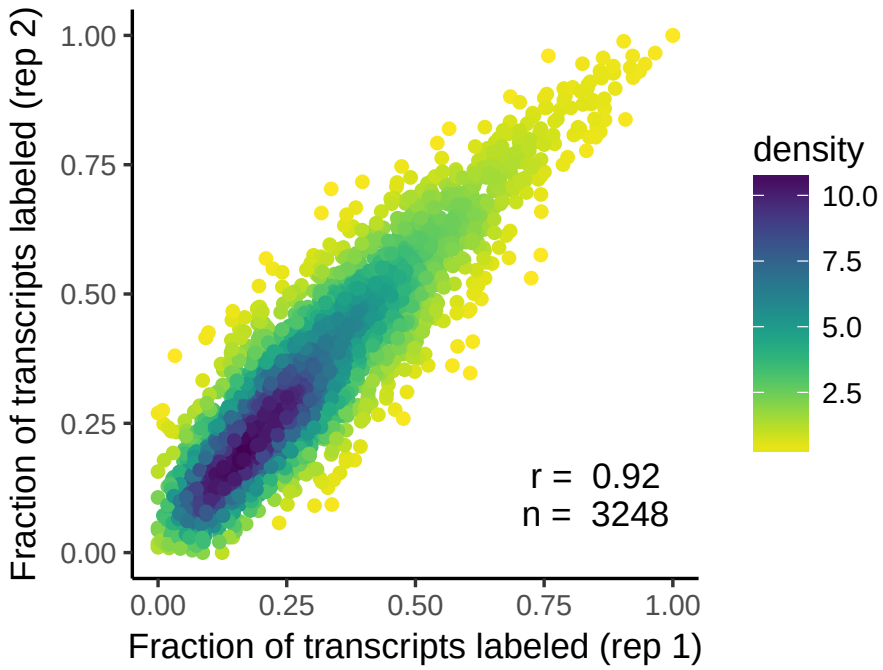

Figure 19: Reproducible estimates of fraction of new reads determined by statistical modeling of TimeLapse-seq data

The correlation between the statistically modeled fraction of new transcripts in each of the two biologically independent replicates of s^6G TimeLapse-seq. Indicated are the number of transcripts considered and the Pearson's r-value of the correlation.

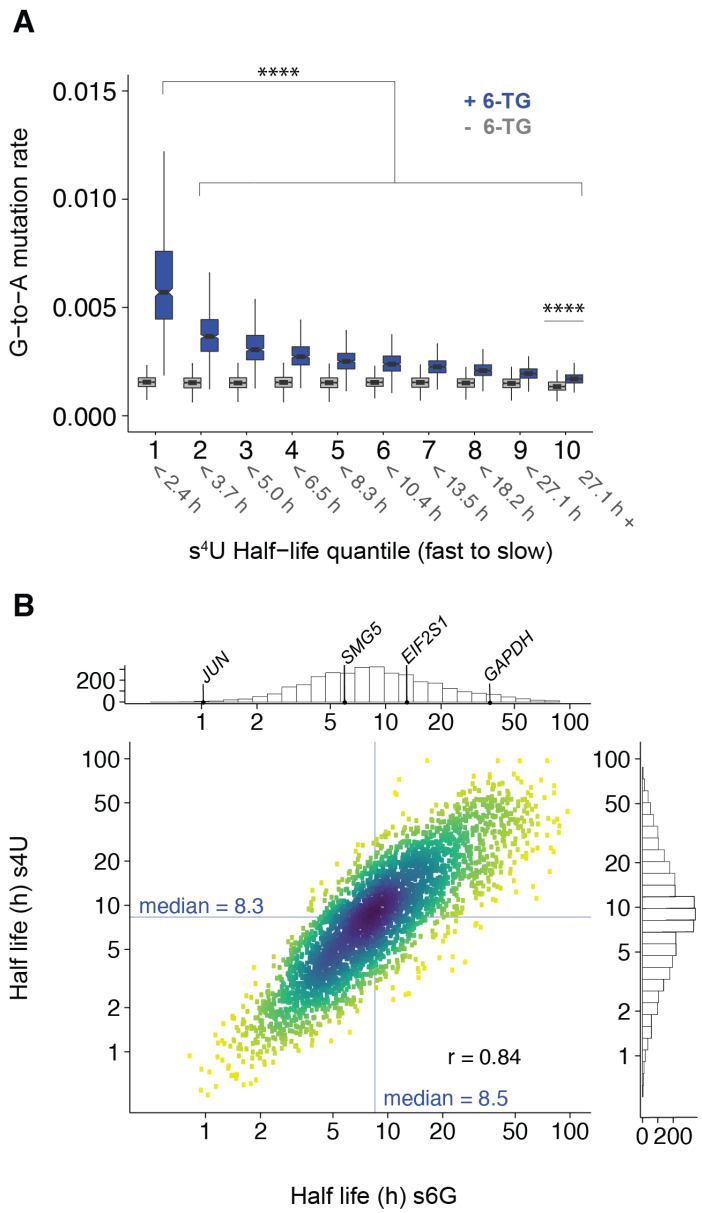

Figure 20: RNA half-lives determined by s^{6}G TimeLapse-seq track with s^{4}U-based estimates

(A) The distribution of average G-to-A mutation rate for each transcript separated by half-life quantile (calculated from validated s^{4}U TimeLapse-seq, 1 = high turnover, 10 = low turnover with half-lives indicated in grey, compared between transcripts from cells treated with 6-TG with identically treated RNA from untreated cells). **** $p < 0.0001$ based on a two sided Wilcox rank sum test. (B) Correlation plot comparing transcript half-lives (log10 transformed) calculated using s^{6}G TimeLapse-seq and s^{4}U

TimeLapse-seq. Histograms summarize the distribution of half-lives with the example transcripts indicated. The density of points is indicated by color (yellow, low; blue, high). I performed the experiments presented here and conducted the data analysis with assistance from the lab.

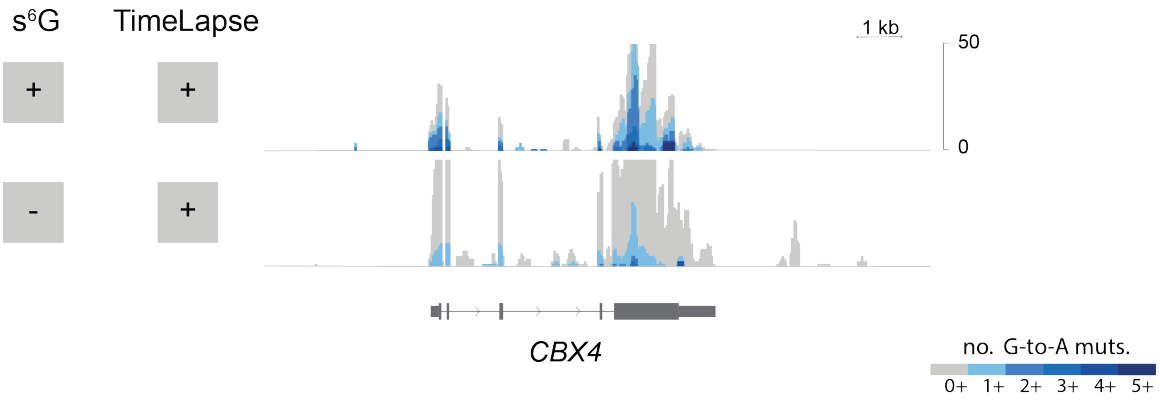

Figure 21: s^{6}G TimeLapse-seq captures transcriptional dynamics for the uridine-poor transcript *CBX4*

Genome browser tracks displaying the coverage of the *CBX4* transcript. Shown are RNA-seq tracks of s^{6}G treated sample (top) and untreated control RNA (bottom) subjected to TimeLapse chemistry (bottom). Reads with increasing number of G-to-A mutations are shown in increasingly darker shades of blue. Both tracks are shown on the same scale. I performed the experiments presented here and conducted the data analysis with assistance from the lab.

3.5 Discussion

These results demonstrate that s^6G can be used to monitor transcriptome-wide RNA population dynamics. Specifically, TimeLapse chemistry can be extended beyond s^4U and can be applied to recode s^6G to cause specific G-to-A mutations in sequencing experiments. While the lower incorporation rates of s^6G [60] lead to lower mutation rates induced by s^6G compared with those induced by s^4U (s^6G 1.5%, s^4U 4.5%), this rate is well above background (0.15% G-to-A mutations in TimeLapse-treated or untreated samples without s^6G) and allows analysis of the fraction of each transcript that is new. The slight increase in G-to-T mutations observed in Figure 13A, which could stem from low levels of G oxidation to 8-oxo-G [59], does not inhibit our ability to capture RNA population dynamics as the levels are low and do not inhibit detection of G-to-A mutations. Half-lives calculated from s^6G TimeLapse-seq correlate well with those determined using s^4U TimeLapse-seq ($r = 0.84$, Figure 20B). Similar to s^4U TimeLapse-seq, the chemical treatment and metabolic labeling with 6-TG preserve the information of traditional RNA-seq (Figure 16) while providing insight into RNA population dynamics.

The nucleotide recoding chemistry of s^6G adds a new technique to the larger set of techniques that use mutations to study nucleic acids, including the analysis of epigenetic modifications through bisulfite sequencing [58], RNA structure [61, 62, 63], RNA-protein interactions [53, 64] and posttranscriptional modifications [65]. The impact of chemical approaches to reveal nucleic acids biology through mutational analysis continues to increase as sequencing technologies continue to evolve. TimeLapse-seq with s^6G expands the power of nucleotide recoding by providing the first determination of RNA half lives using an analogue of guanosine and introduces the potential to analyze multiple time points using different metabolic labels in a single RNA-seq experiment.

Chapter 4

Capturing the Transient Transcriptome with TT-TimeLapse-seq

This chapter contains excerpts from:

Machyna, M., **Kiefer, L.**, Simon, M.D. (2020) Enhanced Nucleotide Chemistry and Toehold Nanotechnology Reveals lncRNA Spreading on Chromatin. *Nat. Struct. Mol. Biol.* **27**, 297-304. doi: 10.1038/s41594-020-0390-z

and

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. Methods*, **15**, 221-225. doi: 10.1038/NMETH.4582

4.1 Summary

While the half-lives of mRNAs range from tens of minutes to several hours, many important regulatory RNA species like anti-sense RNA, enhancer RNA and intronic sequences are degraded within minutes. Similarly, when trying to get a short snapshot

of active transcription, capturing RNA made over the time span of tens of minutes to hours does not provide the necessary resolution. In such cases, methods like TT-seq (discussed in Chapter 1) which utilizes five minute pulses of s^4U become extremely powerful. As outlined in Chapter 1, while TT-seq allows us to capture the transient transcriptome, determining accurate estimates for transcriptional or splicing rates are limited by signal-to-noise. Technological development in our lab driven by former graduate students Jeremy Schofield and Erin Duffy combined our existing s^4U enrichment approach [31] with nucleotide conversion TimeLapse-seq [36] to identify sequencing reads from bona fide nascent RNA (TT-TimeLapse-seq). In this chapter, I will highlight three of my collaborations detecting active transcription in tissue culture, capturing cell-type specific transcription in different mouse tissues and visualising the nascent transcriptome in mouse olfactory neurons.

4.2 Investigating Active Transcription After Heat Shock in Drosophila S2 Cells

Martin Machyna, a current postdoctoral fellow in the Simon lab, devised a novel strategy to capture RNA on chromatin using enhanced nucleotide chemistry and dual toehold technology. He discovered that heat shock leads to changes in roX2 lncRNA spreading on chromatin in *Drosophila* S2 cells [66]. To investigate whether transcription is influencing the changes in roX2 spreading upon heat shock, I captured the nascent transcriptome by performing TT-TimeLapse-seq on heat shocked and control cells. Briefly, Martin subjected S2 cells to heat shock (37°C for 1h, control cells were or kept at 27°C). In the last 8 minutes of the heat shock or control treatment, he subjected the cells to 1 mM s^4U , followed by cell harvest and RNA extraction. Using 35 μg of total RNA input, I used MTS-biotin to biotinylate the s^4U-labeled RNA, followed by streptavidin enrichment. I then eluted the enriched RNA by reducing

the disulfide bond, rendering the s^4U available for subsequent conversion to C* by TimeLapse chemistry. The data show wide-spread dramatic transcriptional shut down upon heat shock with robust activation of known heat shock responsive genes [66] (Figure 22, heat shock genes highlighted in yellow). We reproducibly observe the majority of the reads to originate from bona fide new RNA, with some high copy number transcripts displaying pre-existing RNA background contamination based on the absence of T-to-C mutations (Figure 22 grey reads). Even though transcriptional shut down due to heat shock in *Drosophila* cells has been reported previously [67], it was nevertheless striking to see the complete absence of actively transcribing RNA genome-wide, with the exception of heat shock responsive genes (Figure 22 top). In summary, while Martin found that transcriptional changes do not influence roX2 spreading on chromatin, this project highlights the power of TT-TimeLapse-seq to capture an almost instantaneous snap shot of transcription.

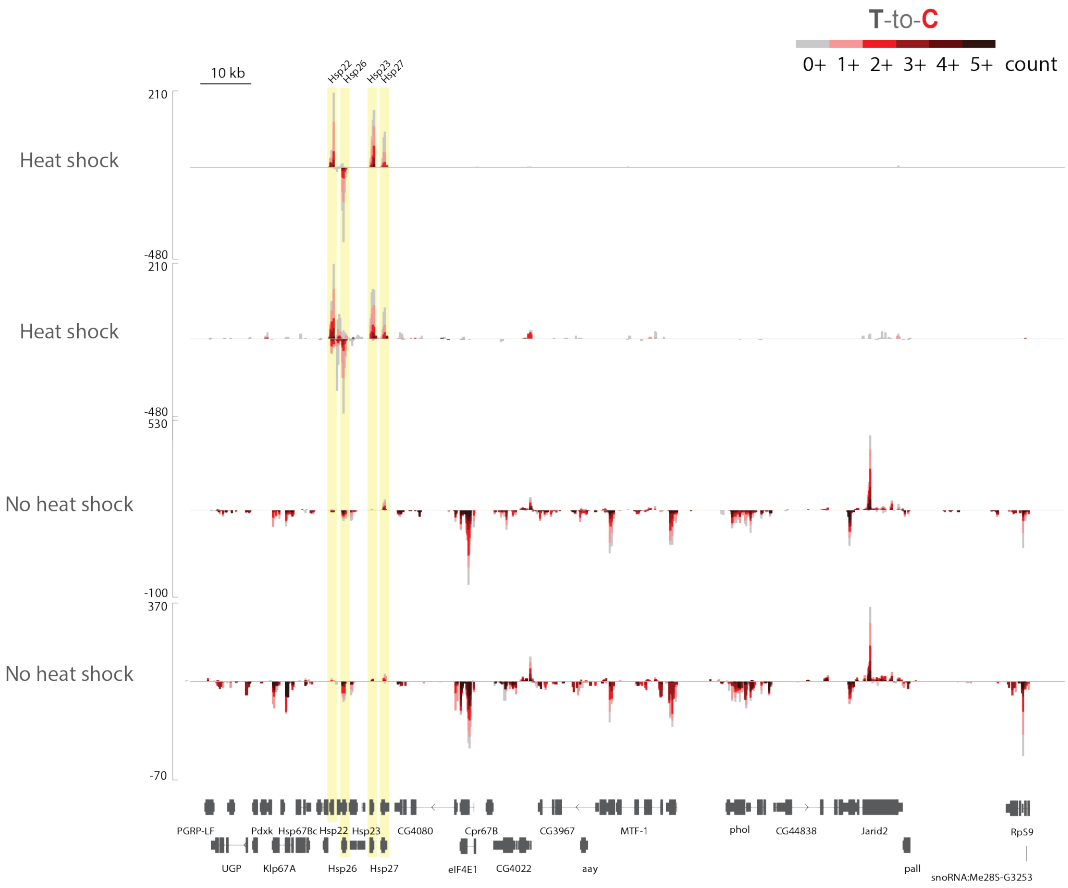

Figure 22: Heat shock causes drastic transcriptional shutdown of all except for heat shock response genes in *Drosophila* S2 cells by TT-TimeLapse-seq

Zoomed out view of a 200 kb region on *Drosophila* S2 chromosome 3L highlighting the complete shutdown of regular transcription (bottom) upon 1h of heat shock (top). Highlighted in yellow are heat shock genes (f.r.t.l.) *Hsp22*, *Hsp26*, *Hsp23* and *Hsp27*. Sequencing reads with increasing numbers of T-to-C mutations are shown in increasingly darker shades of red. Cell treatment and RNA extraction was performed by Martin Machyna. Nascent RNA enrichment and TimeLapse-seq was performed by me.

4.3 Capturing Cell-Type Specific Transcriptional Signatures Using UPRT/TT-TimeLapse-seq in Mouse Tissues

In this project done in collaboration with Fengyun Sun in Oliver Rando's lab at the University of Massachussetts Medical School, we aimed to capture mouse tissue specific transcription using TU-tagging. This technology developed in 2009 uses cell type-specific expression of the *Toxoplasma gondii* salvage pathway enzyme uracil phosphoribosyltransferase (UPRT) causing local s^4U labeling of RNA upon 4-thiouracil injection. Only in cells expressing cre-inducible UPRT can 4-thiouracil be reacted with PRPP to yield 4-thio-UMP. This is a powerful technique to get tissue specific expression profiles in whole organisms. The Rando lab has engineered several mouse lines with UPRT under the control of a liver or sertoli cell-specific promoter, respectively. We decided to perform TT-TimeLapse-seq on extracted liver and testes, to enrich for s^4U-labeled RNA and assess the level of background noise with the addition of s^4U recoding. Excitingly, the data showed enrichment of cell type-specific transcripts (Figure 23, liver-specific *Amdhd1*, sertoli-specific *Sox9*) and only background noise for solvent injected control tissues (data not shown). It is interesting to note that in contrast to capturing transient transcription (e.g. 8 min s^4U pulse as in section 4.2) the captured RNA in this case represents primarily mature RNA as s^4U labeling is done over several hours post 4-thiouracil injection. The signal is therefore dominated by mature spliced RNA, which is highlighted in the sequencing tracks, which show high read accumulation over annotated exons. It is exciting that TU-tagging is compatible with TT-TimeLapse-seq as this opens up the opportunity to, for instance, track RNA transfer from one UPRT expressing tissue into another tissue without s^4U labeling capability. To sum up, TU-tagging with s^4U -enrichment allowed us to capture the tissue-specific transcriptome with nucleotide recoding forming an important addition

to differentiate real signal from background noise.

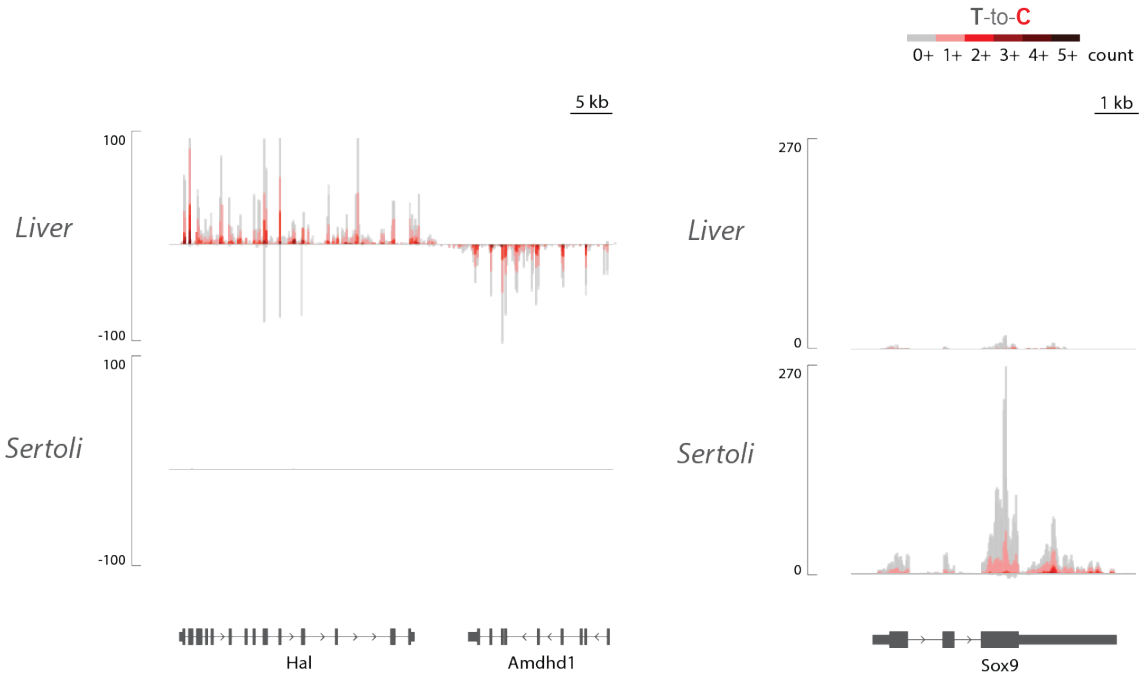

Figure 23: Capturing cell type-specific transcripts by TU-tagging TT-TimeLapse-seq

Mouse tissue from mice expressing UPRT under the control of tissue-specific promoters, injected with either 4-thiouracil or solvent control, was treated to extract total RNA. MTS-biotin enrichment was performed on total extracted RNA with subsequent nucleotide conversion by TimeLapse-seq. Displayed are sequencing tracks for three transcripts with different expression profiles for both liver and sertoli cells. The mouse work, injection and tissue extraction was performed by Fengyun. I extracted RNA, performed TT-TimeLapse-seq and conducted the data analysis.

4.4 Detecting Transient Transcripts in the Mouse Olfactory Epithelium

This project aimed at capturing the nascent transcriptome as well as the transient RNA species transcribed from the protocadherin (Pcdh) cluster in mouse olfactory sensory neurons. The experiments were performed in collaboration with Sandy Rajkumar in Daniele Canzio's lab at UCSF. The Pcdh cluster is a region on mouse chromosome 18 that produces unique cell-specific cell surface proteins through stochastic promoter activation of variable exons. The result is a cell-specific mRNA containing the stochastically chosen variable exon spliced to a set of constant exons. The translation product of this mRNA is a cell surface protein with a common intracellular domain and a cell-specific extracellular domain, which is considered to constitute a unique cellular barcode. The regulation of the variable exon choice, particularly in the Pcdhα cluster, is still not clearly understood. Recently, antisense transcription from the variable exons has been linked to the regulation of long range chromatin contacts which are involved in variable exon choice [68]. The current model suggests that a combination of stochastic variable exon promoter choice as well as alternative splicing lead to RNA isoforms unique to each individual cell. The Pcdhα RNA isoforms, which are limited to neuronal cells, are only transcribed to low levels and the variable exon antisense RNA is extremely short lived, making the cluster especially difficult to study. While the chromatin architecture during Pcdhα promoter choice has been studied using the differentiating mouse olfactory epithelium (OE) as an *in vivo* model of promoter choice, the only high resolution data of the antisense RNA had been collected from human SK-N-SH tissue culture cells. To test whether we could detect transient RNA species such as antisense, intronic and enhancer RNA in primary mouse OE and specifically whether we could identify the lowly abundant *Pcdh*α sense and antisense RNA, we employed TT-TimeLapse-seq. Sandy and I treated extracted mouse OE

with s^4U for 20 minutes and we subsequently enriched for s^4U-labeled RNA followed by TimeLapse to capture the newly-made transcriptome. TimeLapse-seq captures the primarily mature transcriptome, with majority of the signal over the exons, while TT-TimeLapse leads to increased coverage of the intronic sequences (Figure 24A). Excitingly, we also detected antisense RNA for several of the variable exons (*Pcdhα2*, *Pcdhα5*, Figure 24B). Indicating that the regulatory mechanism involving antisense RNA detected in human SK-N-SH tissue culture cells could also regulate variable promoter choice in the *Pcdhα* in the mouse OE. Finally, we were able to detect the presence of a variety of transient RNA species, such as wide-spread intronic sequences and antisense RNA genome-wide (Figure 24C).

In conclusion, Sandy and I were able to visualize the nascent transcriptome in primary mouse olfactory epithilium and provide insights into the extent of intronic and antisense transcription in the Pcdhα cluster. As a future outlook, the Canzio lab aims to couple TT-TimeLapse-seq experiments of the mouse OE harboring various genetic modifications to understand the regulatory role of antisense transcription and splicing in this region.

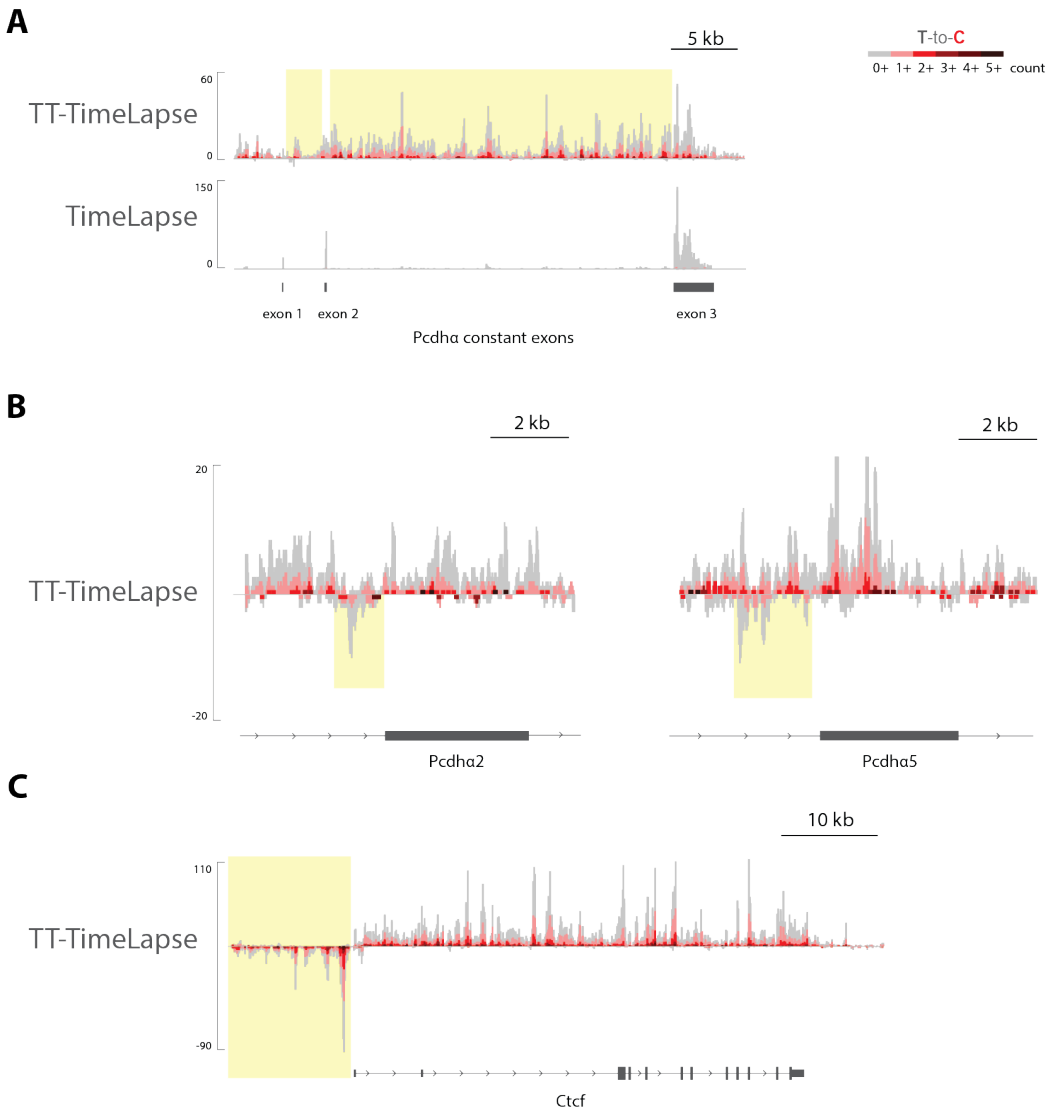

Figure 24: Transient intronic and antisense RNA in mouse olfactory epithilium by TT-TimeLapse

(A) Zoom in on the three constant exons of *Pcdhα* with tracks for enriched and nucleotide recoded samples on top and TimeLapse-treated total RNA-seq tracks on the bottom. Highlighted in yellow are the reads mapping to the intronic regions captured in the top s^4U-enriched and TimeLapse-treated samples.

(B) Genome browser tracks of reads aligning to two of the twelve variable exons of the murine *Pcdhα* cluster including reads mapping to the intronic regions flanking both exons, *Pcdhα2* and *Pcdhα5*. Highlighted in yellow are the reads mapping to the

antisense RNAs originating from the indicated variable exon promoters.

(C) Reads mapping to *Ctcf* with antisense RNA transcription highlighted in yellow and extensive intronic read coverage. Reads with increasing numbers of T-to-C mutations are shown in increasingly darker shades of red. The experiment was performed by Sandy and me.

4.5 Conclusion and Ongoing Efforts

In this chapter, I highlighted several collaborations in which I employed TT-TimeLapse-seq to provide insights into cell type-specific or transient RNA-species. I optimized the original protocol devised by Jeremy Schofield and Erin Duffy [36] (an updated protocol can be found in Chapter 7). I piloted ongoing improvement efforts based on shortcomings and hurdles identified during the above mentioned collaborations. These efforts include work done to improve signal to noise of the enrichment, as well as efforts to decrease the input RNA requirement to allow for processing of precious samples such as primary tissues which yield relatively little RNA. These efforts are ongoing and are conducted with current graduate students Leah Connor and Josh Zimmer.

Chapter 5

Bayesian Modeling of TimeLapse-seq Data

5.1 Summary

In a perfect world, we would be able to distinguish TimeLapse sequencing reads from metabolically labeled and unlabeled RNA by the presence and absence of T-to-C for s^4U or G-to-A mutations for s^6G. However, due to sequencing errors and similar artifacts, there are background mutations that occur in the absence of metabolic labels or TimeLapse chemistry treatment. Also, it is often the case that an RNA that was made in the presence of the metabolic label may not display any mutations by chance, especially if the read length is too short or the incorporation rate is low. For these reasons, we need to apply statistical modeling to infer the fraction of sequencing reads that represent new RNAs. In this chapter I will detail the statistical model used to determine RNA half-lives from s^4U and s^6G TimeLapse-seq data and introduce an updated unpublished model. I will highlight what can be learned from this approach by using a s^4U TimeLapse-seq time course data set collected in collaboration between Alice Lu (Iwasaki lab) and Jeremy Schofield. All the analyses presented in this chapter

are my own. The model was written by current graduate student Josh Zimmer, with minor modifications by myself. The collaboration is ongoing and unpublished.

5.2 TimeLapse-seq Data Processing and Modeling

Similarly to biochemical enrichment experiments outlined in Chapter 1, the ultimate objective of statistically modeling TimeLapse-seq data is to infer the fraction of new RNA. From this fraction new, we can determine the transcript-specific first-order degradation rate constant, the synthesis rate and RNA half-life assuming steady state kinetics. To determine fraction new, we take advantage of the single molecule nature of TimeLapse-seq by classifying each sequencing read by the number of passing T-to-C mutations found in the read. To get there, sequencing data from a TimeLapse-seq experiment are filtered and aligned as normally done, with some modifications to ensure detection of recoding mutations and filtering of background noise as described previously [36, 69]. Briefly, sequencing data is aligned to the reference genome and reads containing T-to-C mutations that pass background mutation and quality control filtering are called and counted for each transcript. In principle, all reads containing T-to-C mutations should have come from newly-made RNA, meaning RNA transcribed during the course of the experiment, with all unmutated reads originating from preexisting RNA. However, given the length of sequencing reads (e.g. for paired-end 75, about 150 nucleotides), assuming equal distribution of all four bases, and given a T-to-C mutation rate of 4% one would expect on average about 1.5 mutations in every read. In reality, a distribution of reads containing 0 all the way to more than 5 mutations per read is observed. Since reads stemming from new RNA with no mutations are indistinguishable from reads stemming from preexisting RNA, more sophisticated analyses are needed to estimate the fraction new.

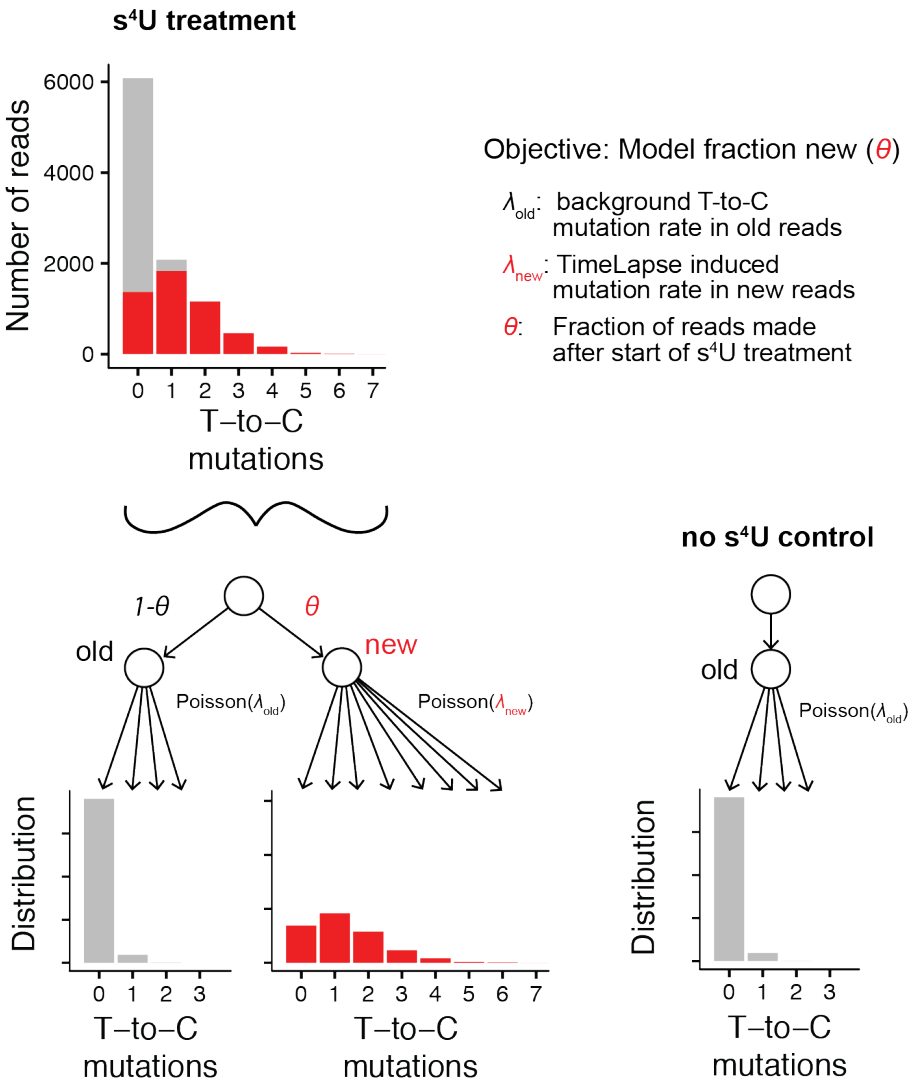

Figure 25: Mixed Poisson model of TimeLapse-seq reads to determine the fraction of newly-made RNA

Modeling the distribution of observed reads grouped by the number of T-to-C mutations as originating from a mixture of two Poisson distributions. The objective is to determine the transcript specific value θ, the fraction of sequencing reads from RNA transcribed during the course of the s^4U labeling. The scheme was adapted from [39]

There are two distributions that we use to model our reads: a binomial model and a Poisson model, the latter of which is shown in Figure 25. In either case fraction new (θ) is determined by a mixture modeling approach using a hierarchical model to pool information about mutation rates across transcripts and account for overdispersion of the data. While the binomial model (not shown) more accurately models the distribution of T-to-C mutations in the reads, the Poisson model is a reasonable approximation that is more computationally efficient and therefore will be the focus of my discussion here. The distribution of mutations in each read can be modeled as a mixture of two Poisson distributions. For a given sequencing read the two Poisson distributions describe the probabilities of observing the number of T-to-C mutations seen in the read, assuming that the read originates from either the old or the new population of RNA. The Poisson distribution of reads from pre-existing RNA (Figure 25, grey), which should not contain s^4U induced T-to-C mutations is defined by the background mutations rate (λ_{old}, estimated < 0.1 %). The reads stemming from s^4U-labeled newly-made RNA (Figure 25, red) are Poisson distributed with the experimental T-to-C mutations rate (λ_{new}, ~ 4.5 %). What we observe is a mixture of the two processes (Figure 25 top) and the point of the statistical model is to determine the contribution of each process to the observed data, with the contribution being θ, the fraction new (Figure 25 bottom left). From the input data which is a matrix of transcript-specific observations of sequencing reads with increasing numbers of T-to-C mutations, the model then estimates the mutation rate (λ_{new}), the background mutations rate (λ_{old}) and the fraction of reads made during the course of the experiment (θ). The background mutation rate can be determined from a no labeling control experiment (Figure 25 bottom right). To retrieve the parameters, we use Stan modeling language, which uses Bayesian statistical inference with MCMC sampling [70]. This approach provides posterior distributions of each parameter, from which the median value is determined together with its corresponding uncertainty (e.g.

80% credible interval) given the data. Using the fraction new value, transcript-specific parameters like RNA half-lives, can be determined as generated quantities during the modeling, which allows for propagation of the associated uncertainties. To sum up, statistical modeling of TimeLapse data allows us to get transcript-specific estimates of kinetic parameters, including our certainty in them.

5.3 Modeling the Transcriptional Response to VSV Infection

VSV is a negative-sense RNA virus, which when introduced to mouse embryonic fibroblasts (MEFs) activates the innate immune response [71]. The innate immune response is the immediate, non-specific response to a pathogen involving white blood cells and epithelial cells. In contrast, the adaptive immunity takes a week or more to be established; however, it yields high specificity and forms a long term memory. Alice Lu from Akiko Iwasaki's lab and Jeremy Schofield collected a time course data set to study the transcriptional response to VSV infection in MEFs. The experimental set up is outlined in Figure 26, with 1h s^4U RNA labeling pre, 2.5h post and 6h post viral infection. In this section I will outline the work I have done to obtain estimates of kinetic parameters using TimeLapse-seq data.

The sequencing data was processed as outlined above and the following model was set up to estimate the T-to-C mutation and background rate as well as fraction new from which estimates of the apparent degradation rate, synthesis rate, RNA half-life and contribution of k_{deg} (a measure of whether a change in k_{deg} or k_{syn} is driving the change in RNA abundance) were generated. The rates of synthesis and degradation are apparent rates and only estimated, since the data was collected under non-steady state conditions, but parameters were estimated using the equations outlined in Chapter 1 assuming steady state. The formalized model is outlined below.

$$TC_i \sim \begin{cases} \theta_{[j,l]}\ \text{Poisson}(\lambda_{n[j,l]}) + (1-\theta_{[j,l]})\ \text{Poisson}(\lambda_{o[j]}) & \text{if read is experimental} \\ \text{Poisson}(\lambda_{o[j]}) & \text{if unfed control} \end{cases}$$

$$\text{logit}(\theta_{[j,l]}) \sim \text{Normal}(0, 1.5)\ \text{(prior for } \theta)$$

$$\log(\lambda_{n[j]}) = \mu_{\log\lambda_n} + z_{n[j]}\sigma_n$$

$$\log(\lambda_{o[j]}) = \mu_{\log\lambda_o} + z_{o[j]}\sigma_o$$

$$\mu_{\log\lambda_n} \sim \text{Normal}(-1, 1)$$

$$\mu_{\log\lambda_o} \sim \text{Normal}(-3, 1)$$

$$z_{n[j]} \sim \text{Normal}(0, 1)$$

$$z_{o[j]} \sim \text{Normal}(0, 1)$$

$$\sigma_n \sim \text{Cauchy}(0, 3)$$

$$\sigma_o \sim \text{Cauchy}(0, 1)$$

$$\text{for i} = 1\ldots\text{n entries} \quad \text{for j} = 1\ldots\text{n transcripts} \quad \text{for l} = 1\ldots\text{n treatments}$$

The data, sequencing reads with a given number of T-to-C mutations for each entry i, are distributed as a mixture of Poisson distributions with contribution of each distribution given by the fraction new, θ. The fraction new is determined for each transcript j and each sample condition l. The two Poisson distributions are described by the mutations rate (λ_{n}) and the background rate (λ_{o}), with λ_{n} being determined per transcript j and each sample condition l and the background rate per transcript j. The mutation and background rate is modeled hierarchically, with the mean log mutation ($\mu_{\log\lambda_n}$) and log mean background rate ($\mu_{\log\lambda_o}$) being normally distributed with the indicated priors, allowing the transcript specific mutation rate to vary but be informed by the mean rates. This allows for pooling information across transcripts and thereby better model fit to the data.

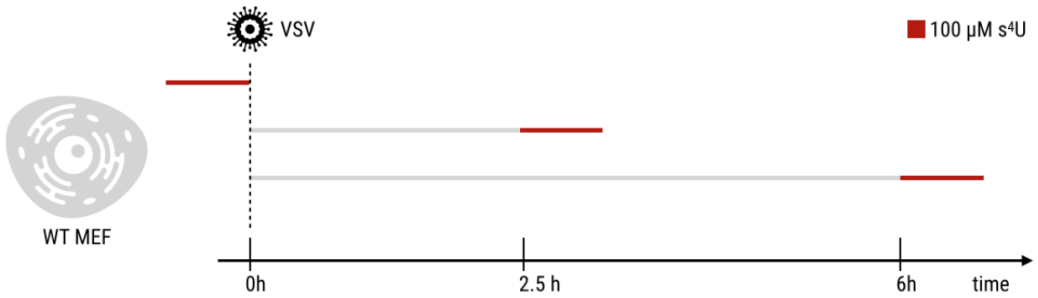

Figure 26: Experimental design of a TimeLapse-seq time course pre- and post-VSV infection

Schematic outlining the experimental designed aimed at capturing the transcriptional response in primary mouse embryonic fibroblasts to Vesicular Stromatitis Virus infection. Biological replicates of 1h s^4U-labeled (red bar) cells at three time points pre- and post-infection were collected as depicted: a zero hour control set without viral infection, a set of 2.5h post-infection and a set of 6h post-infection samples.

The degradation rate constant k_{deg} was determined from the fraction new (θ) as outlined in the introduction Chapter 1. The k_{syn} was calculated from the observed sequencing reads per transcript times the transcript's k_{deg}. And finally, the contribution of k_{deg} parameter was estimated by subtracting the absolute fold change k_{syn} between two conditions from the absolute fold change in k_{deg}.

$$\text{contrib}_{k_{\text{deg}}} = |\log_2\text{FC}(k_{\text{deg}})| - |\log_2\text{FC}(k_{\text{syn}})|$$

The model was run on all transcripts with a read cut off of 150 reads per transcript and quality control checks were performed as suggested [70]. The background and experimental T-to-C mutation rate estimates were in the range of estimates from previous experiments [36] and were high confidence estimates, as illustrated by the tight credible interval (80% credible interval, red bar hidden behind median value, black cross, see Figure 27A). By analyzing the contribution of k_{deg}, I found that the majority of changes post viral infection are driven by changes in the synthesis rate (Figure 27B). This means that the presence of the virus causes both transcriptional upregulation of innate immunity associated genes and transcriptional shut down, rather than increased RNA degradation or stabilization. I therefore ranked transcripts by their $\log_2$ fold change in k_{syn} between the 2.5h time point post viral infection compared to the 0h control. Figure 28 shows a ranked list of transcripts with top to bottom decreasing fold changes in synthesis rate upon viral infection. Excitingly, many of the upregulated transcripts are known or associated with viral infection and the innate immune response (Figure 28, red dots). Importantly, while I am presenting the median estimate value, the 80% credible intervals ((Figure 28, grey bars) give an indication of how confident one can be in the estimates. While no significant changes were observed for the degradation rates between the 2.5h and 6h time point, the synthesis rate of some transcript showed significant changes (data not shown). I speculate that the system has not reached steady state by 2.5h and therefore do not

expect the synthesis rate to stay constant between the two time points. To conclude, I was able to show that the cellular response to VSV infection is primarily driven by changes to the apparent synthesis rate, specifically for innate immune related transcripts, while post-transcriptional stabilization or destabilization is less observed.

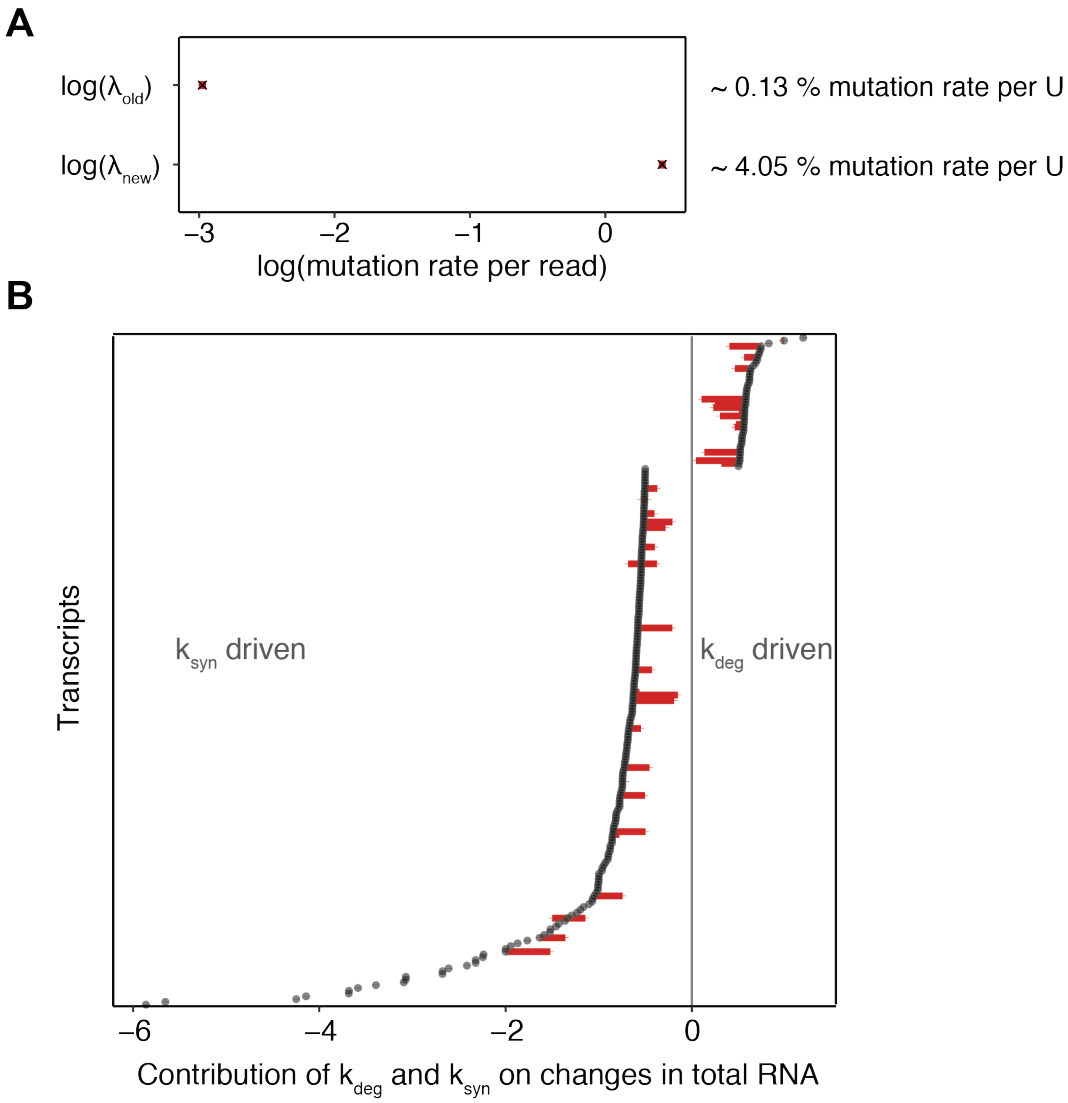

Figure 27: Model estimates of the T-to-C mutation rates and kinetic RNA parameters

(A) Plotted are the posterior distributions of mutation rate estimates with the median (black cross) and the 80% credible interval (red bar). As the mutation rates are estimated per read, an approximated rate per U was calculated assuming paired-end 150 reads, with on average 37.5 Us. This results to a background mutation rate of 0.05 and an experimental rate of 1.5 mutations per read. (B) The contribution of changes in k_{deg} and k_{syn} 2.5h post infection versus control on the observed changes in total RNA calculated as described above, with their median estimates (black dots) and corresponding uncertainties (80% credible interval red bar). Transcripts with

a median value between 0.5 and -0.5 were filtered, as well as transcripts with high uncertainty.

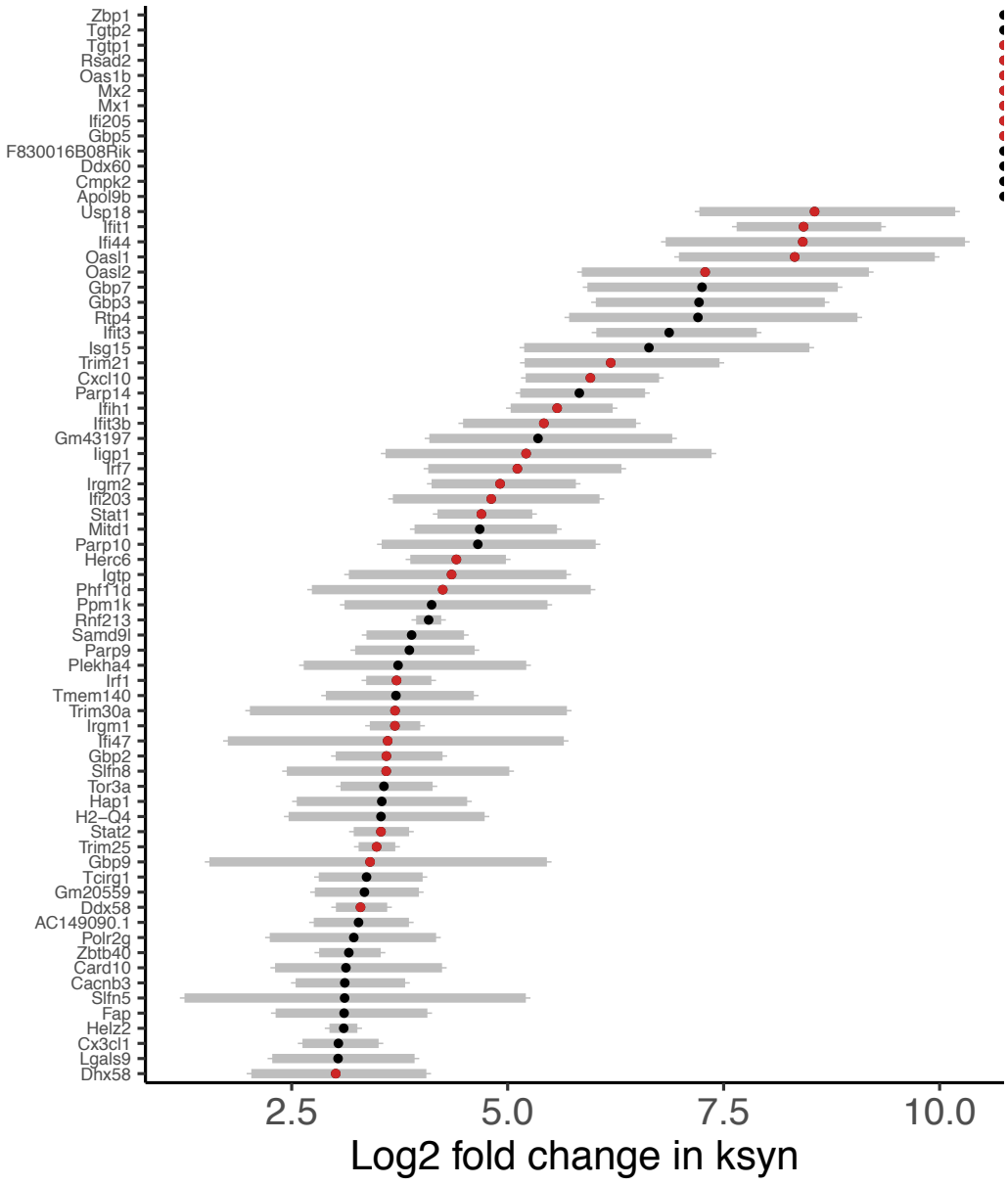

Figure 28: Ranked changes in k_{syn} between viral infection and control shows many known innate immunity transcripts

Plotted are the posterior distributions of the $\log_2(k_{\text{syn}}$ 2.5h / k_{syn} 0h) estimates with the median (black dot) and the 80% credible interval (grey bar). The top 13 transcripts are not expressed at the 0h time point leading to a $\log_2$FC of infinity. Known innate immunity associated transcripts are colored in red.

5.4 Conclusion

In this section I outlined the power of combining TimeLapse-seq with statistical modeling of the experimental parameters to estimate kinetic parameters. As outlined in Chapter 1, a change in RNA-seq signal can be driven by either a change in synthesis or degradation rate. Here, I presented a striking example of wide-scale changes in synthesis rate leading to an increase in RNA-seq signal determined by modeling of TimeLapse-seq data. Another example highlighting an increase in RNA-seq due to post-transcriptional stabilization has recently been published as a collaboration between Jeremy Schofield and Vicky Luo in the Slavoff lab [72].

With increasing length of sequencing reads as sequencing technologies advance, separating out the populations of reads stemming from newly-made and pre-existing RNA will become simpler as the distributions of reads will move further apart. However, in the meantime and also moving forward this kind of analysis will enable us to gain quantitative estimates of parameters describing RNA dynamics and their changes genome-wide.

Chapter 6

Conclusions and Ongoing Work

6.1 Summary

The ability to detect reads stemming from newly made RNA by the presence of T-to-C or G-to-A mutations in TimeLapse-seq with s^4U and s^6G, respectively, enables us to estimate synthesis and decay rates genome wide assuming simple exponential decay and steady state conditions. To detect RNA species underlying more complex kinetics, for instance RNA whose degradation cannot be described using simple exponential kinetics, or to model kinetic parameters under non-steady state conditions, multiple time points would need to be collected [49]. While there are ways to capture the transcriptome, for instance by collecting several independent experiments with different time points, the ability to recode s^4U and s^6G under identical chemical conditions opens the opportunity to collect multiple time points within one internally normalized RNA-seq experiment. Here I outline my ongoing work together with current graduate student Isaac Vock to explore complex synthesis and decay kinetics by integrating single and dual-color TimeLapse-seq.

6.2 Integrating s^4U and s^6G

An important aspect of combining s^4U and s^6G into the same experiment is compatibility of the metabolic labels. While exploring different conditions to improve s^6G incorporation into cellular RNA, I made an interesting observation. I was wondering whether an imbalance in purine nucleosides or nucleotides could influence labeling efficiency and therefore treated 293T cells with 6-TG (for which I observed comparable incorporation levels as s^6G [69]) or a combination of 6-TG and adenine. I employed a dot blot assay to monitor incorporation levels of metabolic labels into cellular RNA as described in Chapter 2. Briefly, human 293T cells were subjected to metabolic labels or combinations thereof for the indicated number of hours. Total RNA was extracted, DNase treated and reacted with activated disulfide conjugated to a fluorophore. After removal of unreacted fluorophore, incorporation levels of s^4U and s^6G (or 6-TG) were determined by relative fluorescence levels over control RNA (detailed protocol in Methods Chapter 7). I found that while incorporation levels of 6-TG increased over time (see Figure 7), the presence of adenine partially inhibited 6-TG labeling of RNA (Figure 29A). Interestingly, treating cells with a combination of 6-TG and uridine only had a minor effect on RNA labeling.

Based on these findings I set out to investigate whether s^4U and 6-TG inhibit each other's capacity of RNA labeling. Comparing each trial to a separately treated mixture of labeled RNA as a positive control (Figure 29B, last sample), I checked different labeling schemes. Firstly, I found that a 4h pre-treatment of 6-TG leads to lower incorporation of s^4U. To get a better estimate of the fluorescence stemming from s^4U incorporation and to see if this interaction also exists with canonical nuclear bases, I repeated the experiment with guanine and observed the same trend. Surprisingly, pre-treatment with 6-TG followed by treatment with s^4U and the other two nuclear bases (cytosine and adenine) led to even lower levels of RNA labeling. Importantly, none of the conditions achieved the same extend of labeling as the positive control.

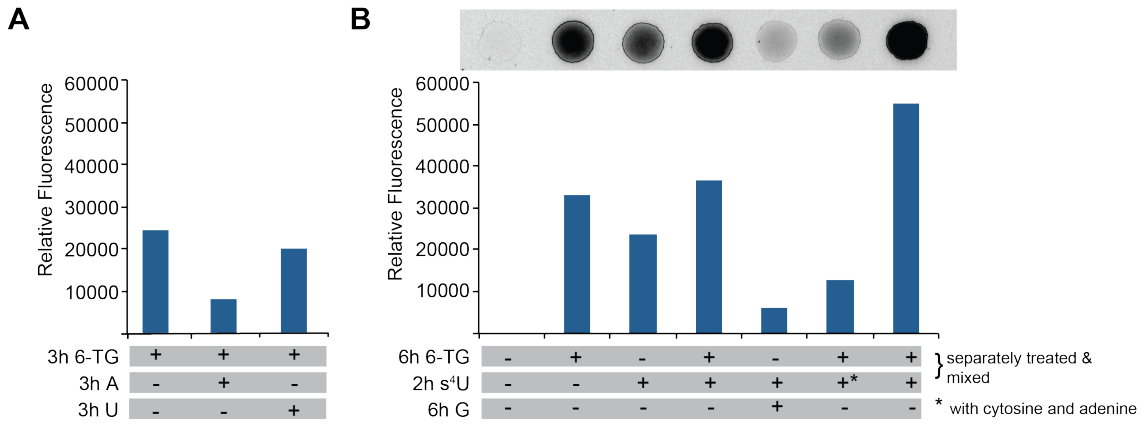

Figure 29: Metabolic labels interact with each other and canonical nuclear bases causing lower RNA incorporation

Cells were labeled with 100 μM of each canonical and non-canonical nuclear base or nucleoside for the indicated number of hours. 5 μg of extracted RNA was labeled with 500 ng of MTS-TAMRA. Fluorescence was determined relative to unlabeled control RNA.

6.3 Discussion

Given the partial inhibition of s^4U by either 6-TG and guanine pre-treatment and the minor effect of uridine on 6-TG labeling, I propose an experimental setup of s^4U pre-treatment followed by a shorter pulse of 6-TG (or s^6G). I have some preliminary data showing that incorporation of both metabolic labels is achieved with this approach. Ultimately, a sequencing experiment has to be performed to assess whether information on RNA dynamics can be gained with the extent of labeling. Interestingly, the Rieder lab employed their version of nucleoside recoding (TUC-seq [73]) to a dual labeling approach using s^6G followed by a washout and subsequent s^4U labeling and detected both G-to-A and T-to-C mutations [74]. As the paper only contains qualitative measures of RNA dynamics, it remains to be seen whether this labeling approach or a long s^4U treatment followed by a short s^6G pulse achieves high enough incorporation to allow for robust statistical modeling to estimate kinetic parameters.

6.4 Ongoing and Future Directions

Together with fellow Simon lab graduate student Isaac Vock, I set up a variety of metabolic labeling experiments to assess the quality of dually-labeled TimeLapse-seq data. Isaac meanwhile developed a statistical model in order to estimate degradation rates from dual-color data. Additionally, he came up with a strategy to detect RNA undergoing complex, non-single-exponential decay, for instance due to transcript subpopulations with different degradation rate constants. His thesis project will focus on identifying complex transcriptional kinetics using both single and dual-color TimeLapse-seq. Finally, the ultimate application of dual-color TimeLapse-seq would be in a single-cell RNA-seq experiment. While other nucleotide recoding technologies have been applied to single cell [75], recoding with two labels has not been done yet. Through unpublished work presented at a conference Jeremy Schofield attended, we

know that single cell TimeLapse-seq with s^4U is possible, indicating that a dual-color TimeLapse-seq approach is at least theoretically attainable.

Overall, nucleotide recoding is a powerful method for identifying changes in the transcriptome, enabling mechanistic studies. While other metabolic labels exist to capture the newly-made transcriptome, I have, to the best of my knowledge, developed the first purine recodable RNA metabolic label. As coding regions have an, on average, higher GC content compared to intronic regions, any biases a uridine analogue could induce when studying intronic sequences or the dynamics of splicing could be identified and overcome by employing s^6G recoding. Finally, the compatibility of the recoding conditions of s^4U and s^6G allows for the use of both RNA metabolic labels within the same experiment. Being able to detect multiple time points within one experiment opens the opportunity to increase the temporal resolution when studying a wide range of RNA half lives, to identify RNAs underlying complex kinetics, and to study RNA population dynamics under non-steady state conditions. .

Chapter 7

Methods and Data Analysis

This chapter contains excerpts from:

Kiefer, L., Schofield, J.A., Simon, M.D. (2018) expanding the Nucleoside Recoding Toolkit: Revealing RNA Population Dynamics with 6-Thioguanosine. *J. Am. Chem. Soc.*, **140**, 14567-14570. doi: 10.1021/jacs.8b08554

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. Methods*, **15**, 221-225. doi: 10.1038/NMETH.4582

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C. D. and Kitchen, R. R., Gerstein, M. B. and Simon, M. D. (2015) Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. *Molecular Cell*, **59(5)**, 858–866. doi.org/10.1016/j.molcel.2015.07.023

7.1 RNA Dot Blot - Assessing s^4U/s^6G Incorporation

Labeling and RNA extraction

1. Follow the most current TimeLapse-seq protocol for metabolic labeling and cell harvesting

2. Extract RNA and deplete genomic DNA accordingly

MTS-TAMRA reaction

1. Use 10% MTS-TAMRA to RNA by weight

2. Set up reaction by bringing RNA to final volume of 38 μl using DEPC-treated water (dot blots using 5 μg of RNA are recommended, down to 3.5 μg has been successful)

3. Add 1 μl each of 1M HEPES pH 7.4 (final concentration 20 mM) and 0.05M EDTA (final concentration 1 mM)

4. Add appropriate amount of MTS-TAMRA from 2.5 mg/ml (dry DMF) stock brought up to 10 μl final volume with DMF to reaction (50 μl final reaction volume)

5. React for 1h at RT with motion (in dark)

Reaction cleanup

1. In a well-ventilated are add 50 μl of DEPC-treated water and add 100 μl 24:1 chloroform:isoamyl alcohol

2. Transfer to a heavy phase lock tube (prep by spinning for 1 min at 12,000 $\times$ g) and shake for 15 sec, let sit for 2 min in dark

3. Spin at 12,000 $\times$ g for 5 min - transfer aqueous phase into a new 1.7 ml tube and add 350 μl of RLT (from RNeasy Mini Kit), mix well by pipetting up and down 10 times

4. Add 250 μl of 100% ethanol and mix well by pipetting up and down 10 times

5. Transfer solution into an RNeasy column and spin at 12,000 × g for 15 sec and discard flowthrough

6. Add 500 μl of RPE (from RNeasy Mini Kit) to the column, spin for 15 sec as before and discard flowthrough

7. Add 500 μl of freshly made 80% ethanol, spin for 2 min at 12,000 × g and discard flowthrough

8. Switch column into new collection tube and spin at max for 1 min to dry column

9. Add 50 μl of DEPC-treated water and elute by spinning for 1 min at 13,000 × g

10. Image by pipetting 5 μl onto the glass of the Typhoon and using settings for TAMRA

7.2 DNA Dot Blot - Assessing s^6G Incorporation

Labeling and DNA extraction

1. Treat cells for anywhere between 10 min to several hours with metabolic label of choice (added volume 10 μl)

2. At end of labeling period rinse cells with ice-cold PBS

3. Scrape cells in 500 μl PBS using a cell scraper and collect into DNA LoBind tubes

4. Pellet (300 × g, 3 min), LoBind tubes can be inserted into 15 ml falcon tubes w/out caps

5. Remove PBS, lyse cells with 300 μl gDNA lysis buffer (100 mM Tris pH 8, 50 mM EDTA, 1% SDS, 1:40 Proteinase K (20 mg/ml stock, add fresh)

6. Incubate at 37°C for 1h with gentle shaking (dark)

7. Add equal volume Phenol:Chloroform:Isoamyl Alcohol (25:24:1 v/v) and shake - spin at maximum speed for 5 min, transfer aqueous layer into new LoBind tube (DNA will appear stringy when removing aqueous layer, try to avoid carrying over phenol)

8. Add 1:10 vol 3M NaOAc, 2.5 vol Ethanol, shake and incubate in -80°C for 10 min - spin at max for 5 min, decant and wash with 70% Ethanol

9. Decant and dry for 10 min at RT

10. Resuspend DNA in 80 μl DEPC-treated water at RT overnight

Digestion and RNA removal

1. Add 10 μl Cutsmart, 9 μl HindIII HF, 1 μl RNase A (1mg/ml) and incubate at 37°C for 1h

2. Add 1:10 vol 3M NaOAc and 2.5 vol Ethanol (follow ethanol precip instructions above)

3. Redissolve in 50 μl DEPC-treated water (30 min RT w/shaking)

4. Measure conc by nanodrop

MTS-TAMRA reaction and cleanup

1. Set up MTS-reaction (1h, RT on wheel) (use 10% MTS-TAMRA to DNA by weight)

2. Add 50 μl DEPC, 100 μl Phenol:Chloroform:Isoamyl Alcohol (25:24:1 v/v) followed by ethanol precipitation

3. Redissolve in 50 μl DEPC-treated water (30 min RT w/ shaking)

4. Image on the Typhoon (5 μl on the glass) using TAMRA setting

5. Check concentration by nanodrop and 260/280 ($\sim$ 1.8 DNA, 2.0 RNA)

7.3 s^6G DNA Restriction Digest Assay

A DNA duplex containing a single s^6dG was enzymatically synthesized from a template DNA strand using Taq polymerase and 6-thio-2'-deoxyguanosine-5'-triphosphate instead of dGTP. The resulting product was purified by ethanol precipitation. The s^6dG-containing DNA was used to screen for optimal TimeLapse chemistry conditions by adding the amine solution to the duplex, followed by the oxidant and subsequent incubation at a set temperature value for a set duration of time. The products were isolated from the reaction mixture using AMPure XP beads following manufacturer's instructions. An adapter was added through a single PCR step followed by amplification with a Cy5-labeled primer. Only the PCR products templated from the s^6dG-containing strand were amplified and fluorescently labeled. The amplified PCR product was then incubated with SspI HF restriction enzyme following manufacturer's instructions (1h, 37°C). The fluorescent products of the restriction digest were visualized by Native PAGE and scanned on a GE Healthcare Typhoon FLA 9500 (Chemical Biology Institute, Yale). The extent of cutting (as a proxy for the extent of nucleotide conversion) was determined using ImageJ.

7.4 s^6G TimeLapse LC-MS Assay

TFEA (600 mM) was added to a solution of 6-thio-2'-deoxyguanosine (350 μM). $NaIO_4$ (10 mM) was dissolved in DEPC-treated water and added to the reaction mixture. After each time point (5 min, 20 min, 40 min, 60 min) at 50°C in the dark, the reaction was analyzed by reverse-phase LC-MS with a Hypersil Gold column

(Thermo, 3 μm, 160 $\times$ 2.1 mm) using chromatography conditions described previously [36]. Masses were collected using positive ion mode, extracted ions were identified using Agilent MassHunter software.

7.5 NMR

All NMR analysis was performed on an Agilent DD2 400 MHz spectrometer (Yale West Campus Analytical Core) with an Agilent OneNMR probe. ^{1}H-NMR spectra were processed using the MestReNova software.

7.5.1 Reaction of 4-Thiouracil with TimeLapse Chemistry

4-thiouracil (4.3 mg, 1 equiv) was dissolved in DMSO-d_6, and TFEA (3.4 μl 12 M, 1.3 equiv) was added to the solution. After mixing, a solution of $NaIO_4$ in DMSO-d_6 (12.3 mg, 0.2 M, 1.7 equiv) was added to the nucleobase and amine solution and incubated at 45°C for 4h.

7.5.2 Reaction of 6-Thioguanine with TimeLapse Chemistry

6-thioguanine (10.7 mg, 1 equiv.) was dissolved in DMSO-d_6 and TFEA (6.38 μl 12 M, 1.2 equiv.) was added to the solution. After mixing, a solution of $NaIO_4$ in DMSO-d_6 (12.3 mg, 0.2 M, 1.5 equiv) was added and the reaction was allowed to proceed at 45°C for 1h.

7.6 TimeLapse-seq with s^6G

This protocol was adapted from Jeremy Schofield's work on TimeLapse-seq [36]. The experimental procedure is identical, with the only change being that the s^6G stock has to be dissolved in DMSO instead of DEPC-treated water.

7.7 TT-TimeLapse-seq

This protocol was adapted from Erin Duffy's original MTS-biotin protocol [**Duffy2015TracDuffyHt**
Jeremy Schofield and Erin Duffy's initial TT-TimeLapse-seq experiments [36].

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 (5 min, 1 mM final concentration).

   Note: 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 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 300 × g for 3 min. Carefully aspirate PBS from cell pellet.
6. Resuspend pellet in 1 mL Trizol by gently pipetting up and down ∼10 times.
7. Trizol samples can be stored overnight at -80°C, or kept on ice for RNA isolation.

s^4U RNA isolation

1. Add 200 μl chloroform to the 1 ml Trizol samples. Shake the tubes vigorously for 15 sec and let sit for 2 min in the dark.
2. Centrifuge the tubes for 5 min at 12,000 $\times g, 4$°C. Transfer aqueous phase (∼ 500 μl) to new DNA LoBind tubes with 1 μl RNase-free glycogen (20 μg).

3. To each aqueous phase from step 2, add 500 μl of 100% isopropanol with 1 mM DTT final concentration (*make 10 ml isopropanol + 10 μl of 1 M DTT master mix using freshly dissolved DTT). Mix by pipetting up and down 10 times. Incubate samples at room temperature for 10 min.

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

4. Centrifuge samples 20 min at 20,000 $\times g, 4°C$. Carefully remove the isopropanol from the RNA/glycogen pellet.

5. Add 1 ml of room temperature 75% ethanol to the pellet, vortex quickly and centrifuge 3 min at 12,000 $\times g, 4°C$.

6. Remove the ethanol completely from the RNA/glycogen pellet. To do so, first remove the majority of the ethanol with a P1000 pipet tip, then spin the tubes again on a countertop microcentrifuge. Use a gel-loading/10 μl tip to remove the remaining ethanol. Let the pellet air-dry for 2 min by leaving the tube open under a Kimwipe cover in the dark. Be careful not to overdry, which will result in loss of RNA.

7. Resuspend each pellet in 40 μl of RNase-free (i.e., DEPC-treated water). Measure the RNA concentrations using a Nanodrop spectrophotometer.

   Generally we retrieve >50 μg of RNA after this step, from a 80-9010 cm plate of adhesive cells (this includes RNA as well as any contamination from DNA and free nucleotides that were carried through the Trizol extraction and precipitation).

8. Treat isolated RNA at 200 ng/μl with 1 μl Turbo DNAse per 10 μg of RNA and Turbo DNAse buffer in a PCR tube (e.g. 20 μg of RNA, 2 μl of DNAse, 10 μl 10$\times$ Turbo DNAse buffer and nuclease free water to a total volume of 100 μl).

9. Incubate at 37°C for 30 min.

10. Add 1× RNAClean beads (e.g. for a 100 μl reaction add 100 μl beads) to each sample.

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

11. Flick and invert or lightly vortex samples to mix. Incubate samples for 8 minutes.

12. Place samples on magnetic rack and allow beads to collect until solution is clear (∼5 min)

13. Remove supernatant and wash 2× 200 μl freshly prepared 80% ethanol

14. Remove ethanol and briefly spin PCR tubes to collect residual ethanol. Remove residual with 10 μl pipet tip.

15. Leave PCR tubes open on magnetic rack and allow to dry (∼5 min).

    Note: 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. Be careful not to overdry the beads.

16. Add 80 μl nuclease-free water to dried beads and flick tubes until completely resuspended. Incubate in dark for 5 min at RT.

    Note: 50 μl elution also works; adjust the 2× RNA shearing buffer.

17. Briefly spin tubes to collect sample at bottom and place on magnetic rack until solution is clear (∼1-2 min).

18. Transfer eluted RNA in water to clean PCR strip.

19. Assess RNA concentration by nanodrop.

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

1. Take 75 μg (limit of RNeasy Mini Kit column) or 80 μl of RNA samples into shearing (adjust volume to 80 μl with DEPC-treated water if necessary). Add 80 μl 2× RNA fragmentation buffer and place sample at 94°C for exactly 3 min 30 s.

2. Quickly spin the RNA sample on a countertop microcentrifuge to bring the solution to the bottom of the tube, and immediately place on ice.

3. Add 40 μl of 250 mM EDTA (final concentration: 50 mM EDTA) to each sample, mix by vortexing, and incubate on ice for 2 min.

Modified RNeasy MinElute cleanup

1. Add 700 μl buffer RLT (from RNeasy Mini Kit) and mix well by pipetting up and down 10 times. Then add 500 μl 100% ethanol to each RNA sample, and mix well by pipetting up and down 10 times. Apply these samples to RNeasy columns in two steps of about 700 μl each.

2. Microcentrifuge columns 15 sec at 12,000 $\times g$, 4°C. Discard the flowthrough.

3. To each column add 800 μl RPE buffer (from RNeasy Mini Kit) supplemented with 35 μl of 1% 2-mercaptoethanol (final concentration: 10 mM 2-mercaptoethanol) (*make a master mix of 5 ml RPE + 3.5 μl 2-mercaptoethanol).

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

4. Centrifuge the samples 15 sec at 12,000 $\times g$, 4°C. Discard the flowthrough.

5. Add 800 μl freshly prepared 80% ethanol and centrifuge the samples for 2 min at 12,000 $\times g$, 4°C. Discard the flowthrough.

6. Switch the columns to new 2-ml collection tubes (from the kit). Microcentrifuge the samples 5 min at maximum speed, 4°C.

7. Transfer column to a 1.7-ml microcentrifuge tube, and add 38 μl RNase-free (DEPC-treated) water. Incubate the column with DEPC-treated water at 65°C for 5 min before the centrifuging the samples 1 min at $\geq$12,000 $\times g$, 4°C. Apply an additional 39 μl of DEPC-treated water to the columns, incubate for 1 min at 65°C then centrifuge as before. Use 1 μl of the combined elution to assess concentration by nanodrop.

Block streptavidin beads - Part I

1. Aliquot 10 μl Dynabeads MyOne Streptavidin C1 beads per sample plus 10% (e.g. for six samples, aliquot 66 μl beads) into one 1.7-ml microcentrifuge tube. Place the tube in a magnetic rack for 2 min and remove supernatant with a pipet.

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

3. Wash beads twice with 1 ml of high-salt wash buffer (recipe under Reagents and Solutions), mixing and capturing the beads as described before.

4. Add 600 μl of high-salt buffer with 1.2 μl of glycogen stock (recipe under Reagents and Solutions) to the beads, resuspend the bead mixture completely by pipet, and incubate for 1 hr at room temperature with slight agitation (e.g. on a rotation device).

Biotinylation of s4U-RNA with the activated disulfide methane thiosulfonate (MTS) biotin

1. To biotinylate the RNA, mix the following reagents in a 1.7-ml microcentrifuge tube: 50 μg of RNA 2 μl 1 M HEPES, pH 7.4 (final concentration, 20 mM HEPES, pH 7.4) 2 μl 0.05 M disodium EDTA (final concentration, 1 mM EDTA) Nuclease-free water to 80 μl total volume.

2. Based on the input of RNA, use 10% by weight of MTSEA-biotin-XX and add DMF (20% DMF in final reaction) to 20 μl per sample (e.g. input of 50 μg of RNA requires 5 μg of MTSEA-biotin-XX, meaning 2.5 μl of the 2 mg/ml MTSEA-biotin-XX in DMF stock solution and add 17.5 μl DMF). Make a master mix of required MTSEA-biotin-XX from the 2 mg/ml stock solution (see recipe in Reagents and Solutions) plus DMF to 20 μl per sample + 10% (e.g. 6 samples, make MM for 6.6).

3. Add 20 μl of the MTSEA-biotin-XX in DMF master mix to each sample and mix well by pipetting up and down 10 times.

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

Remove unreacted MTS-biotin from RNA samples

1. Add the reactions to pre-spun phase-lock tubes (spin tubes at 12,000 $\times$ g for 1 min prior to usage) and add 100 μl 24:1 chloroform:isoamyl alcohol to each tube.

2. Shake the samples vigorously for 15 sec (careful, phase lock tubes can open) and let sit 2 min.

3. Centrifuge the samples for 5 min at 12,000 $\times g, 4°$C. For each sample, transfer the aqueous phase (100 μl) to a new labeled 1.7-ml microcentrifuge tube.

4. To each aqueous phase add 350 μl buffer RLT (from RNeasy Mini Kit) and mix by pipetting up and down 10 times. Then add 250 μl 100% ethanol, mix well by pipetting up and down 10 times and apply to an RNeasy column from the kit.

5. Centrifuge the columns for 15 sec at 12,000 $\times g$, 4°C. Discard the flowthrough.

6. Add 500 μl buffer RPE (from RNeasy Mini Kit, no 2-mercaptoethanol here!) to each column and centrifuge the columns for 15 sec at 12,000 $\times g$, 4°C. Discard the flowthrough.

7. Add 500 μl freshly prepared 80% ethanol to each column and centrifuge the columns for 2 min at 12,000 $\times g$, 4°C. Discard the flowthrough.

8. Switch the columns to new 2-ml collection tubes (from RNeasy Mini Kit). Centrifuge samples 1 min at maximum speed, 4°C.

9. Transfer the columns to new microcentrifuge tubes. Add 50 μl RNase-free water to the center of each column. Elute the RNA by centrifuging these samples for 1 min at $\geq$12,000 $\times g$, 4°C. Proceed immediately to isolation of the biotinylated s4U-RNA.

Isolate s^{4}U -containing transcripts with streptavidin beads

1. In order to compare enrichment to input material, remove part of the reaction now and store on ice for later use (e.g. for 1% input remove 0.5 μl of the 50 μl reaction).

2. Add 5 μl of high-salt wash buffer to the 50 μl RNA MTSEA-biotin-XX reactions and mix well (add 5 μl of high-salt regardless of whether input was removed). Add each sample to the beads from Block streptavidin beads. Cover with foil and incubate the samples at room temperature for 15 min on a rotator at 30 rpm (or similar) to ensure the beads are mixing during the incubation.

3. Place the tubes in a 96-well magnetic rack for 2 min and remove supernatant with a pipet. Save the supernatant on ice as "flowthrough".

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

5. Add 100 μl of TE buffer (see recipe in Reagents and Solutions) and incubate at 55°C for 15 min in a thermal cycler. During the incubation preheat 200 μl TE buffer per sample to 55°C.

6. After 15 min capture the beads on the magnetic rack as before and remove the buffer. Wash beads twice with 100 μl of pre-warmed 55°C TE buffer as in step 4.

Option A: Elution for screening and qPCR

1. Elute in 7 μl of elution buffer (see recipe in Reagents and Solutions), mix the beads by inversion, wrap the tubes in foil, and incubate them at room temperature in the dark for 15 min with rotation. Capture the beads on the magnetic device as before and capture the elution into a new PCR tube.

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

2. To the elution ($\sim$7 μl) add 2 μl of Invitrogen Vilo Mastermix and 1 μl of superscript III (Thermo Fisher), mix well and incubate for
   10 min at 25°C
   60 min at 42°C
   5 min at 85°C

to reverse transcribe the eluted RNA. To compare to input, add elution buffer to the saved input material up to 7 μl (eg. 5 μl input + 2 μl elution buffer) and add RT mix as with the elution above.

3. Take the reverse transcription (RT) reaction (input and elution separately) and dilute it with DEPC-treated water according to the number of primers to be tested (# of primers $\times$ 5 μl RT $\times$ 2 for duplicate wells + 10%), and set up a qPCR plate with the following:

   5 μl of 4 μM primer (FW & RV mix)

   5 μl of RT reaction

   10 μl iTaq Universal SYBRgreen supermix (bio-rad)

   Mix by pipetting up and down 8 times and analyze in qPCR machine using the following settings:

   Repeat the following a total of 45$\times$:

   95°C for 3 min

   95°C for 10 s

   60°C for 30 s

   Plateread

   95°C for 10 s

   55°C for 31 s

   Repeat the following a total of 81$\times$:

   55°C for 5 s

   + 0.5°C/cycle

   Ramp 0.5°C/s

Plateread

Option B: Elution for library prep

1. Add 25 μl freshly made elution buffer to the beads, mix the beads by inversion, wrap the tubes in foil, and incubate them at room temperature in the dark for 15 min with rotation.

   This elution buffer will reduce the disulfide bond that formed between biotin and 4- thiouridine, thereby eluting the s4U-RNA and leaving biotin bound to the streptavidin beads.

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

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

4. Take out 1% of the elution (e.g. 2.5 μl if elution is 25 μl) and perform RT and qPCR as described in Elution for screening and qPCR steps 2 and 3. Add elution buffer to the saved input material up to the final elution volume (e.g. if input is 5 μl and elution is 25 μl, add 20 μl of elution buffer to the input material to reduce the MTS-biotin off the RNA). Take out 1% input and precede as with the 1% elution to assess success of enrichment over input. Save the rest of the elution and input material for the library preparation.

5. To clean up the RNA from the elution and input samples add 1× RNAClean beads and follow RNA isolation steps 11.-18.. Elute in 20 μl DEPC-treated

water into new PCR tube. (Optional: Use 1 μl to assess concentration and RNA integrity on the bioanalyzer).

TimeLapse chemistry

1. To the 20 μl of eluted RNA add:

   0.835 μl 3 M sodium acetate pH 5.2,

   0.2 μl 500 mM EDTA,

   1.365 μl DEPC-treated water,

   1.3 μl Trifluoroethylamine (TFEA)

   Note: 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.

2. Close PCR tubes and flick tubes to mix well, spin down, then add freshly prepared

   1.3 μl of 200 mM sodium periodate (in DEPC-treated water).

3. Close PCR tubes and flick tubes to mix well and spin down.

4. Incubate in PCR cycler at 45°C for 1h.

5. Add 25 μl RNAClean beads (1

   $\times$

   volume) and purify RNA as described above. Elute in 10 μl DEPC-treated water and transfer into a fresh PCR strip.

Reducing treatment

1. Prepare reducing master mix (2 μl per sample), for 100 μl use

   58 μl of DEPC-treated water

   10 μl of 1M DTT

10 μl of 1 M Tris pH 7.4

2 μl of 0.5 M EDTA

20 μl of 5 M NaCl

2. Add 2 μl reducing master mix to each sample, flick tubes to mix, spin down and incubate in thermocycler at 37°C for 30 min.

3. Add 20 μl RNAClean beads (1× volume) and purify RNA as described above.

4. Elute in minimal volume of DEPC-treated water consistent with library prep input volume restriction (e.g. Takara pico input mammalian v2 kit max. input volume is 8 μl, so elute in 8μl to increase amount of RNA input into library preparation).

REAGENTS AND SOLUTIONS

Use nuclease-free (DEPC-treated) water in all recipes and protocol steps unless stated otherwise.

Bead blocking buffer

High-salt wash buffer (see recipe)

40 ng/μl glycogen (add from 20 μg/μl stock solution)

Elution buffer, 1×

100 mM dithiothreitol (DTT), freshly dissolved in DEPC-treated water

20 mM HEPES, pH 7.4

1 mM EDTA

100 mM NaCl

0.05% (v/v) Tween-20

Prepare solution fresh each experiment

(e.g. for 100 μl use 75.5 μl DEPC, 10 μl 1 M DTT, 2 μl 1 M HEPES, 2 μl 50 mM

EDTA, 10 μl 1 M NaCl, 0.5 μl 10% Tween 20)

High-salt wash buffer

100 mM Tris Cl, pH 7.4

10 mM disodium EDTA

1 M NaCl

0.05% (v/v) Tween 20

Store at room temperature up to 6 months

(e.g. for 50 ml use 5 ml 1M Tris pH 7.4, 1 ml 0.5 M EDTA pH 8, 10 ml 5M NaCl, 250 μl 10% Tween-20, 33.75 ml DEPC)

RNA fragmentation buffer, 2×

150 mM Tris Cl, pH 8.3

225 mM KCl

9 mM $MgCl_2$

Store at room temperature up to 6 months

(e.g. for 10 ml use 1.5 ml 1 M Tris Cl pH 8.3, 750 μl 3 M KCl, 90 μl 1 M $MgCl_2$, 7.66 ml MilliQ)

MTS-biotin stock solution in DMF

Dilute solid MTSEA-biotin-XX (mol. wt. 607.7 g/mol) in dry DMF to 2 mg/ml. The MTSEA-biotin-XX stocks are stable at $-20°$C for at least 3 months.

TE Buffer

10 mM Tris pH 7.4

1 mM disodium EDTA

(e.g. for 50 ml use 500 μl 1 M Tris pH 7.4, 100 μl of 0.5 M EDTA, 49.4 ml MilliQ)

Bibliography

[1] A. Zeisel, W. J. Köstler, N. Molotski, J. M. Tsai, R. Krauthgamer, J. Jacob-Hirsch, G. Rechavi, Y. Soen, S. Jung, Y. Yarden, and E. Domany, "Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli," *Molecular Systems Biology*, vol. 7, 2011. DOI: 10.1038/msb.2011.62.

[2] L. Herzel, D. S. Ottoz, T. Alpert, and K. M. Neugebauer, "Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function," *Nature Reviews Molecular Cell Biology*, vol. 18, no. 10, pp. 637–650, Oct. 2017. DOI: 10.1038/nrm.2017.63.

[3] B. S. Zhao, I. A. Roundtree, and C. He, "Post-transcriptional gene regulation by mRNA modifications," *Nature Reviews Molecular Cell Biology*, vol. 18, no. 1, pp. 31–42, Dec. 2016. DOI: 10.1038/nrm.2016.132.

[4] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, "Global quantification of mammalian gene expression control," *Nature*, vol. 473, no. 7347, pp. 337–342, May 2011. DOI: 10.1038/nature10098.

[5] E. M. Tank, C. Figueroa-Romero, L. M. Hinder, K. Bedi, H. C. Archbold, X. Li, K. Weskamp, N. Safren, X. Paez-Colasante, C. Pacut, S. Thumma, M. T. Paulsen, K. Guo, J. Hur, M. Ljungman, E. L. Feldman, and S. J. Barmada, "Abnormal RNA stability in amyotrophic lateral sclerosis," *Nature Communications*, vol. 9, no. 1, pp. 1–16, 2018. DOI: 10.1038/s41467-018-05049-z.

[6] A. Baluapuri, E. Wolf, and M. Eilers, "Target gene-independent functions of MYC oncoproteins," *Nature Reviews Molecular Cell Biology*, vol. 21, no. 5, pp. 255–267, May 2020. DOI: 10.1038/s41580-020-0215-2.

[7] G. Orphanides and D. Reinberg, "A unified theory of gene expression," *Cell*, vol. 108, no. 4, pp. 439–451, Feb. 2002. DOI: 10.1016/S0092-8674(02)00655-4.

[8] L. Gorini and W. K. Maas, "The potential for the formation of a biosynthetic enzyme in Escherichia coli," *BBA - Biochimica et Biophysica Acta*, vol. 25, no. C, pp. 208–209, Jan. 1957, ISSN: 00063002. DOI: 10.1016/0006-3002(57)90450-X.

[9] J. M. Gray, D. A. Harmin, S. A. Boswell, N. Cloonan, T. E. Mullen, J. J. Ling, N. Miller, S. Kuersten, Y. C. Ma, S. A. McCarroll, S. M. Grimmond, and M. Springer, "SnapShot-Seq: A method for extracting genome-wide, in Vivo mRNA dynamics from a single total RNA sample," *PLoS ONE*, vol. 9, no. 2, Feb. 2014. DOI: 10.1371/journal.pone.0089673.

[10] D. Gaidatzis, L. Burger, M. Florescu, and M. B. Stadler, "Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation," *Nature Biotechnology*, vol. 33, no. 7, pp. 722–729, Jul. 2015. DOI: 10.1038/nbt.3269.

[11] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, and P. V. Kharchenko, "RNA velocity of single cells," *Nature*, vol. 560, no. 7719, pp. 494–498, Aug. 2018. DOI: 10.1038/s41586-018-0414-6.

[12] K. M. Neugebauer, "Nascent RNA and the coordination of splicing with transcription," *Cold Spring Harbor Perspectives in Biology*, vol. 11, no. 8, a032227, Aug. 2019. DOI: 10.1101/cshperspect.a032227.

[13] L. E. Hokin and M. R. Hokin, "The incorporation of 32P into the nucleotides of ribonucleic acid in pancreas slices during enzyme synthesis and secretion," *BBA - Biochimica et Biophysica Acta*, vol. 13, no. C, pp. 401–412, Jan. 1954, ISSN: 00063002. DOI: 10.1016/0006-3002(54)90347-9.

[14] J. E. Logan, F. C. Heagy, and R. J. Rossiter, "Phosphorus metabolism of the adrenal gland: effect of hypophysectomy and administration of ACTH on the incorporation of radioactive phosphate into the RNA nucleotides," *Endocrinology*, vol. 56, no. 4, pp. 455–460, Apr. 1955. DOI: 10.1210/endo-56-4-455.

[15] M. Muramatsu, J. L. Hodnett, and H. Busch, "Studies on the "independence" of nucleolar ribonucleic acid synthesis," *BBA Specialized Section on Nucleic Acids and Related Subjects*, vol. 91, no. 4, pp. 592–597, Dec. 1964. DOI: 10.1016/0926-6550(64)90006-4.

[16] H. Tani and N. Akimitsu, "Genome-wide technology for determining RNA stability in mammalian cells: Historical perspective and recent advantages based on modified nucleotide labeling," *RNA Biology*, vol. 9, no. 10, pp. 1233–1238, 2012. DOI: 10.4161/rna.22036.

[17] L. J. Core, J. J. Waterfall, and J. T. Lis, "Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters," *Science*, vol. 322, no. 5909, pp. 1845–1848, Dec. 2008. DOI: 10.1126/science.1162228.

[18] H. Tani, R. Mizutani, K. A. Salam, K. Tano, K. Ijiri, A. Wakamatsu, T. Isogai, Y. Suzuki, and N. Akimitsu, "Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals," *Genome Research*, vol. 22, no. 5, pp. 947–956, May 2012. DOI: 10.1101/gr.130559.111.

[19] C. Y. Jao and A. Salic, "Exploring RNA transcription and turnover in vivo by using click chemistry," *Proceedings of the National Academy of Sciences of the*

United States of America, vol. 105, no. 41, pp. 15 779–15 784, Oct. 2008. DOI: 10.1073/pnas.0808480105.

[20] M. Meryet-Figuiere, B. Alaei-Mahabadi, M. M. Ali, S. Mitra, S. Subhash, G. K. Pandey, E. Larsson, and C. Kanduri, "Temporal separation of replication and transcription during S-phase progression," *Cell Cycle*, vol. 13, no. 20, pp. 3241–3248, Oct. 2014. DOI: 10.4161/15384101.2014.953876.

[21] W. T. Melvin, H. B. Milne, A. A. Slater, H. J. Allen, and H. M. Keir, "Incorporation of 6-Thioguanosine and 4-Thiouridine into RNA Application to Isolation of Newly Synthesised RNA by Affinity Chromatography," *Eur. J. Biochem*, vol. 92, pp. 373–379, 1978.

[22] M. D. Cleary, C. D. Meiering, E. Jan, R. Guymon, and J. C. Boothroyd, "Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay," *Nature Biotechnology*, vol. 23, no. 2, pp. 232–237, Feb. 2005. DOI: 10.1038/nbt1061.

[23] L. Dölken, Z. Ruzsics, B. Rädle, C. C. Friedel, R. Zimmer, J. Mages, R. Hoffmann, P. Dickinson, T. Forster, P. Ghazal, and U. H. Koszinowski, "High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay," *RNA*, vol. 14, no. 9, pp. 1959–1972, Sep. 2008. DOI: 10.1261/rna.1136108.

[24] C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D. E. Martin, A. Tresch, and P. Cramer, "Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast," *Molecular Systems Biology*, vol. 7, p. 458, 2011. DOI: 10.1038/msb.2010.112.

[25] A. Swiatkowska, W. Wlotzka, A. Tuck, J. D. Barrass, J. D. Beggs, and D. Tollervey, "Kinetic analysis of pre-ribosome structure in vivo," *RNA*, vol. 18, no. 12, pp. 2187–2200, Dec. 2012. DOI: 10.1261/rna.034751.112.

[26] S. E. Munchel, R. K. Shultzaberger, N. Takizawa, and K. Weis, "Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay," *Molecular Biology of the Cell*, vol. 22, no. 15, pp. 2787–2795, Aug. 2011. DOI: 10.1091/mbc.E11-01-0028.

[27] B. Neymotin, R. Athanasiadou, and D. Gresham, "Determination of in vivo RNA kinetics using RATE-seq," *RNA*, vol. 20, no. 10, pp. 1645–1652, Oct. 2014. DOI: 10.1261/rna.045104.114.

[28] L. Windhager, T. Bonfert, K. Burger, Z. Ruzsics, S. Krebs, S. Kaufmann, G. Malterer, A. L'Hernault, M. Schilhabel, S. Schreiber, P. Rosenstiel, R. Zimmer, D. Eick, C. C. Friedel, and L. Dölken, "Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution," *Genome Research*, vol. 22, no. 10, pp. 2031–2042, Oct. 2012. DOI: 10.1101/gr.131847.111.

[29] J. D. Barrass, J. E. Reid, Y. Huang, R. D. Hector, G. Sanguinetti, J. D. Beggs, and S. Granneman, "Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling," *Genome Biology*, vol. 16, no. 1, p. 282, Dec. 2015. DOI: 10.1186/s13059-015-0848-1.

[30] B. Schwalb, M. Michel, B. Zacher, K. F. Hauf, C. Demel, A. Tresch, J. Gagneur, and P. Cramer, "TT-seq maps the human transient transcriptome," *Science*, vol. 352, no. 6290, pp. 1225–1228, Jun. 2016. DOI: 10.1126/science.aad9841.

[31] E. E. Duffy, M. Rutenberg-Schoenberg, C. D. Stark, R. R. Kitchen, M. B. Gerstein, and M. D. Simon, "Tracking Distinct RNA Populations Using Efficient

and Reversible Covalent Chemistry," *Molecular Cell*, vol. 59, no. 5, pp. 858–866, 2015. DOI: 10.1016/j.molcel.2015.07.023.

[32] E. E. Duffy, D. Canzio, T. Maniatis, and M. D. Simon, "Solid phase chemistry to covalently and reversibly capture thiolated RNA," *Nucleic Acids Research*, vol. 46, no. 14, pp. 6996–7005, 2018. DOI: 10.1093/nar/gky556.

[33] S. M. Testa, M. D. Disney, D. H. Turner, and R. Kierzek, "Thermodynamics of RNA-RNA duplexes with 2- or 4-thiouridines: Implications for antisense design and targeting a group I intron," *Biochemistry*, vol. 38, no. 50, pp. 16 655–16 662, Dec. 1999. DOI: 10.1021/bi991187d.

[34] K. Burger, B. Mühl, M. Kellner, M. Rohrmoser, A. Gruber-Eber, L. Windhager, C. C. Friedel, L. Dölken, and D. Eick, "4-Thiouridine inhibits rRNA synthesis and causes a nucleolar stress response," *RNA Biology*, vol. 10, no. 10, pp. 1623–1630, 2013, ISSN: 15558584. DOI: 10.4161/rna.26214. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24025460/.

[35] L. H. Gregersen, M. Schueler, M. Munschauer, G. Mastrobuoni, W. Chen, S. Kempa, C. Dieterich, and M. Landthaler, "MOV10 Is a 5' to 3' RNA Helicase Contributing to UPF1 mRNA Target Degradation by Translocation along 3' UTRs," *Molecular Cell*, vol. 54, no. 4, pp. 573–585, May 2014, ISSN: 10974164. DOI: 10.1016/j.molcel.2014.03.017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24726324/.

[36] J. A. Schofield, E. E. Duffy, L. Kiefer, M. C. Sullivan, and M. D. Simon, "TimeLapse-seq: Adding a temporal dimension to RNA sequencing through nucleoside recoding," *Nature Methods*, vol. 15, no. 3, pp. 221–225, 2018. DOI: 10.1038/nmeth.4582.

[37] V. A. Herzog, B. Reichholf, T. Neumann, P. Rescheneder, P. Bhat, T. R. Burkard, W. Wlotzka, A. Von Haeseler, J. Zuber, and S. L. Ameres, "Thiol-

linked alkylation of RNA to assess expression dynamics," *Nature Methods*, vol. 14, no. 12, pp. 1198–1204, 2017. DOI: 10.1038/nmeth.4435.

[38] J. Russo, A. M. Heck, J. Wilusz, and C. J. Wilusz, "Metabolic labeling and recovery of nascent RNA to accurately quantify mRNA stability," *Methods*, vol. 120, pp. 39–48, 2017. DOI: 10.1016/j.ymeth.2017.02.003.

[39] E. E. Duffy, J. A. Schofield, and M. D. Simon, "Gaining insight into transcriptome-wide RNA population dynamics through the chemistry of 4-thiouridine," *Wiley Interdisciplinary Reviews: RNA*, vol. 10, no. 1, 2019. DOI: 10.1002/wrna.1513.

[40] M. Rabani, J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit, and A. Regev, "Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells," *Nature Biotechnology*, vol. 29, no. 5, pp. 436–442, 2011. DOI: 10.1038/nbt.1861.

[41] A. A. Pai, T. Henriques, K. McCue, A. Burkholder, K. Adelman, and C. B. Burge, "The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture," *eLife*, vol. 6, Dec. 2017, ISSN: 2050084X. DOI: 10.7554/eLife.32537.

[42] H. B. Li, J. Tong, S. Zhu, P. J. Batista, E. E. Duffy, J. Zhao, W. Bailis, G. Cao, L. Kroehling, Y. Chen, G. Wang, J. P. Broughton, Y. G. Chen, Y. Kluger, M. D. Simon, H. Y. Chang, Z. Yin, and R. A. Flavell, "M 6 A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways," *Nature*, vol. 548, no. 7667, pp. 338–342, Aug. 2017. DOI: 10.1038/nature23450.

[43] E. E. Duffy and M. D. Simon, "Enriching s4 U-RNA Using Methane Thiosulfonate (MTS) Chemistry," *Current protocols in chemical biology*, vol. 8, no. 4, pp. 234–250, Dec. 2016, ISSN: 21604762. DOI: 10.1002/cpch.12.

[44] M. D. Robinson and A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biology*, vol. 11, no. 3, R25, Mar. 2010. DOI: 10.1186/gb-2010-11-3-r25.

[45] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nature Methods*, vol. 5, no. 7, pp. 621–628, Jul. 2008. DOI: 10.1038/nmeth.1226.

[46] A. Lugowski, B. Nicholson, and O. S. Rissland, "DRUID: a pipeline for transcriptome-wide measurements of mRNA stability," *RNA*, no. 24, pp. 623–632, 2018. DOI: 10.1261/rna.

[47] M. Yano and H. Hayatsu, "Permanganate oxidation of 4-thiouracil derivatives. Isolation and properties of 1-substituted 2-pyrimidone 4-sulfonates," *BBA Section Nucleic Acids And Protein Synthesis*, vol. 199, no. 2, pp. 303–315, Feb. 1970, ISSN: 00052787. DOI: 10.1016/0005-2787(70)90073-0.

[48] Q. Dai, S. Moshitch-Moshkovitz, D. Han, N. Kol, N. Amariglio, G. Rechavi, D. Dominissini, and C. He, "Nm-seq maps 2-O-methylation sites in human mRNA with base precision," *Nature Methods*, vol. 14, no. 7, pp. 695–698, Jun. 2017. DOI: 10.1038/nmeth.4294.

[49] C. Deneke, R. Lipowsky, and A. Valleriani, "Complex Degradation Processes Lead to Non-Exponential Decay Patterns and Age-Dependent Decay Rates of Messenger RNA," *PLoS ONE*, vol. 8, no. 2, 2013. DOI: 10.1371/journal.pone.0055442.

[50] W. V. Gilbert, T. A. Bell, and C. Schaening, "Messenger RNA modifications: Form, distribution, and function," *Science*, vol. 352, no. 6292, 2016. DOI: 10.1126/science.aad8711.

[51] M. Lee, B. Kim, and V. N. Kim, "Emerging roles of RNA modification: M6A and U-Tail," *Cell*, vol. 158, no. 5, pp. 980–987, 2014. DOI: 10.1016/j.cell.2014.08.005.

[52] T. Woodford, R. Schlegel, and A. B. Pardee, "Selective Isolation of Newly Synthesized Mammalian mRNA after in Vivo Labeling with 4-Thiouridine or 6-Thioguanosine," *Analytical Biochemistry*, vol. 171, pp. 166–172, 1988. DOI: 10.1016/0003-2697(88)90138-8.

[53] M. Hafner, M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano, A. C. Jungkamp, M. Munschauer, A. Ulrich, G. S. Wardle, S. Dewell, M. Zavolan, and T. Tuschl, "Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP," *Cell*, no. 1, pp. 129–141, 2010. DOI: 10.1016/j.cell.2010.03.009.

[54] S. Basu, R. P. Rambo, J. Strauss-Soukup, J. H. Cate, A. R. Ferré-Damaré, S. A. Strobel, and J. A. Doudna, "A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor," *Nat. Struct. Biol.*, vol. 5, no. 11, pp. 986–992, 1998. DOI: 10.1038/2960.

[55] M. Gładysz, W. Andrałojć, T. Czapik, Z. Gdaniec, and R. Kierzek, "Thermodynamic and structural contributions of the 6-thioguanosine residue to helical properties of RNA," *Scientific Reports*, vol. 9, no. 1, pp. 1–8, Dec. 2019, ISSN: 20452322. DOI: 10.1038/s41598-019-40715-2. [Online]. Available: https://doi.org/10.1038/s41598-019-40715-2.

[56] A. Kumari, "Pyrimidine de novo Synthesis," in *Sweet Biochemistry*, Elsevier, ch. 20, pp. 101–103. DOI: 10.1016/B978-0-12-814453-4.00020-0.

[57] S. Basu and S. A. Strobel, "Biochemical detection of monovalent metal ion binding sites within RNA," *Methods*, vol. 23, no. 3, pp. 264–275, 2001. DOI: 10.1006/meth.2000.1137.

[58] M. Frommer, L. E. Mcdonald, D. S. Millar, C. M. Collist, F. Wattt, G. W. Griggt, P. L. Molloyt, and C. L. Paul, "A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands," *Genetics*, vol. 89, pp. 1827–1831, 1992. DOI: 10.1073/pnas.89.5.1827.

[59] "Schofield-etal_2018,"

[60] A. G. Baltz, M. Munschauer, B. Schwanhäusser, A. Vasile, Y. Murakawa, M. Schueler, N. Youngs, D. Penfold-Brown, K. Drew, M. Milek, E. Wyler, R. Bonneau, M. Selbach, C. Dieterich, and M. Landthaler, "The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts," *Molecular Cell*, vol. 46, no. 5, pp. 674–690, 2012. DOI: 10.1016/j.molcel.2012.05.021.

[61] N. A. Siegfried, S. Busan, G. M. Rice, J. A. Nelson, and K. M. Weeks, "RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP)," *Nature methods*, vol. 11, no. 9, pp. 959–965, 2014. DOI: 10.1038/nmeth.3029.

[62] M. Zubradt, P. Gupta, S. Persad, A. M. Lambowitz, J. S. Weissman, and S. Rouskin, "DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo," *Nature Methods*, vol. 14, no. 1, pp. 75–82, 2016. DOI: 10.1038/nmeth.4057.

[63] A. N. Sexton, P. Y. Wang, M. Rutenberg-Schoenberg, and M. D. Simon, "Interpreting Reverse Transcriptase Termination and Mutation Events for Greater Insight into the Chemical Probing of RNA," *Biochemistry*, vol. 56, no. 35, pp. 4713–4721, 2017. DOI: 10.1021/acs.biochem.7b00323.

[64] A. C. McMahon, R. Rahman, H. Jin, J. L. Shen, A. Fieldsend, W. Luo, and M. Rosbash, "TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins," *Cell*, vol. 165, no. 3, pp. 742–753, 2016. DOI: 10.1016/j.cell.2016.03.007.

[65] X. Shu, Q. Dai, T. Wu, I. R. Bothwell, Y. Yue, Z. Zhang, J. Cao, Q. Fei, M. Luo, C. He, and J. Liu, "N6-Allyladenosine: A New Small Molecule for RNA Labeling Identified by Mutation Assay," *Journal of the American Chemical Society*, vol. 139, no. 48, pp. 17 213–17 216, 2017. DOI: 10.1021/jacs.7b06837.

[66] M. Machyna, L. Kiefer, and M. D. Simon, "Enhanced nucleotide chemistry and toehold nanotechnology reveals lncRNA spreading on chromatin," *Nature Structural and Molecular Biology*, vol. 27, no. 3, pp. 297–304, Mar. 2020, ISSN: 15459985. DOI: 10.1038/s41594-020-0390-z.

[67] T. O'Brien and J. T. Lis, "Rapid changes in Drosophila transcription after an instantaneous heat shock.," *Molecular and Cellular Biology*, vol. 13, no. 6, pp. 3456–3463, Jun. 1993. DOI: 10.1128/mcb.13.6.3456.

[68] D. Canzio, C. L. Nwakeze, A. Horta, S. M. Rajkumar, E. L. Coffey, E. E. Duffy, R. Duffié, K. Monahan, S. O'Keeffe, M. D. Simon, S. Lomvardas, and T. Maniatis, "Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin α Promoter Choice," *Cell*, vol. 177, no. 3, pp. 639–653, Apr. 2019, ISSN: 10974172. DOI: 10.1016/j.cell.2019.03.008.

[69] L. Kiefer, J. A. Schofield, and M. D. Simon, "Expanding the Nucleoside Recoding Toolkit: Revealing RNA Population Dynamics with 6-Thioguanosine," *Journal of the American Chemical Society*, vol. 140, no. 44, pp. 14 567–14 570, Nov. 2018, ISSN: 15205126. DOI: 10.1021/jacs.8b08554.

[70] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A probabilistic programming language," *Journal of Statistical Software*, vol. 76, no. 1, pp. 1–32, Jan. 2017. DOI: 10.18637/jss.v076.i01.

[71] D. Baltzis, L.-K. Qu, S. Papadopoulou, J. D. Blais, J. C. Bell, N. Sonenberg, and A. E. Koromilas, "Resistance to Vesicular Stomatitis Virus Infection Requires a

Functional Cross Talk between the Eukaryotic Translation Initiation Factor 2α Kinases PERK and PKR," *Journal of Virology*, vol. 78, no. 23, pp. 12 747–12 761, Dec. 2004. DOI: 10.1128/jvi.78.23.12747-12761.2004.

[72] Y. Luo, J. A. Schofield, M. D. Simon, and S. A. Slavoff, "Global Profiling of Cellular Substrates of Human Dcp2," *Biochemistry*, May 2020. DOI: 10.1021/acs.biochem.0c00069.

[73] C. Riml, T. Amort, D. Rieder, C. Gasser, A. Lusser, and R. Micura, "Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing," *Angewandte Chemie*, vol. 129, no. 43, pp. 13 664–13 668, 2017. DOI: 10.1002/ange.201707465.

[74] C. Gasser, I. Delazer, E. Neuner, K. Pascher, K. Brillet, S. Klotz, L. Trixl, M. Himmelstoß, E. Ennifar, D. Rieder, A. Lusser, and R. Micura, "Thioguanosine Conversion Enables mRNA-Lifetime Evaluation by RNA Sequencing Using Double Metabolic Labeling (TUC-seq DUAL)," *Angewandte Chemie - International Edition*, vol. 59, no. 17, pp. 6881–6886, Apr. 2020. DOI: 10.1002/anie.201916272.

[75] F. Erhard, M. A. Baptista, T. Krammer, T. Hennig, M. Lange, P. Arampatzi, C. S. Jürges, F. J. Theis, A. E. Saliba, and L. Dölken, "scSLAM-seq reveals core features of transcription dynamics in single cells," *Nature*, vol. 571, no. 7765, pp. 419–423, Jul. 2019. DOI: 10.1038/s41586-019-1369-y.