The Kinetics of RNA Flow Across Subcellular Compartments

Citation

Permanent link

Terms of Use

Share Your Story

HARVARD UNIVERSITY
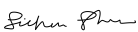**Graduate School of Arts and Sciences**

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Medical Sciences

in the subject of Biological and Biomedical Sciences

have examined a dissertation entitled

THE KINETICS OF RNA FLOW ACROSS SUBCELLULAR COMPARTMENTS
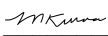
presented by Brendan Michael Smalec

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: _____

Typed Name: Dr. Sichen Shao

Signature: _____

Typed Name: Dr. Stephen Buratowski

Signature: _____

Typed Name: Dr. Mitzi Kuroda

Signature: _____

Typed Name: Dr. David Bartel

Date: September 02, 2022

THE KINETICS OF RNA FLOW ACROSS

SUBCELLULAR COMPARTMENTS

A dissertation presented
by

Brendan Michael Smalec

to
The Division of Medical Sciences
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Genetics and Genomics

Harvard University
Cambridge, Massachusetts
September 2022

Dissertation advisor: L. Stirling Churchman   Brendan Michael Smalec

THE KINETICS OF RNA FLOW ACROSS SUBCELLULAR COMPARTMENTS

ABSTRACT

Throughout their lifetimes, eukaryotic RNAs traverse across subcellular compartments. The rates of RNA flow across the cell are impacted by regulatory processes, including splicing, export, and ribosome loading. In turn, these processes impact the stability and fate of the transcripts upon which they act. RNA flow rates ultimately determine the dynamic pool of transcripts available for translation. However, current genome-wide techniques that measure RNA turnover quantify the time between synthesis and degradation but lack spatial information about where transcripts reside in the cell during this time, obscuring the subcellular kinetics of RNA flow.

To measure rates of RNA flow genome-wide, I first developed subcellular TimeLapse sequencing, a method that tracks the age of RNAs as they shuttle across subcellular compartments. I used this technique to quantify RNA half-lives in mouse NIH-3T3 and human K562 cells at subcellular resolution by fitting a kinetic model to these data in a Bayesian framework. Using this approach, I measured the rates at which transcripts are released from chromatin, exported out of the nucleus, loaded onto polysomes, and degraded in both the nucleus and cytoplasm. All RNA flow rates displayed substantial variability genome-wide, and transcripts from genes with related functions flowed across subcellular compartments with similar kinetics.

I then identified candidate determinants of RNA flow by analyzing the variability in kinetics between different transcripts. The targets of RNA binding proteins and transcription factors experienced distinct RNA flow rates. By verifying these associations, I report that two

RNA binding proteins, DDX3X and PABPC4, regulate the nuclear export of their target transcripts. I next measured RNA poly(A) tail lengths with subcellular resolution and show that transcripts residing on chromatin for longer times had extended tails, while the reverse trend was observed for mRNAs in the cytoplasm. Finally, additional genetic and molecular features that underlie RNA flow rates were identified using a machine learning model. Collectively, this work characterizes the life cycle of mammalian mRNAs, revealing the many lives of RNA transcripts and the molecular features underlying their fates.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I cannot possibly express the full extent of gratitude that I owe to others for their support in my training and help with this project. Without such guidance, this thesis could have never been imagined, much less written. Still, I would be seriously remiss to not highlight the contributions of several extraordinary individuals.

First, I owe much of who I am as a scientist today to my graduate advisor, Stirling Churchman. Even before joining the lab, Stirling made it clear that she took my ideas and suggestions seriously, instilling within me a sense of confidence and excitement that has carried me through the highs and lows of graduate school. Stirling gave me the freedom to define my own projects and develop my own sense of scientific intuition, all while being there to support and point me in the right direction when the results didn't pan out as expected. I'm a more thoughtful, rigorous, and resilient scientist because of her mentorship, and it is an honor to have been part of the talented community she has fostered.

The Churchman lab has felt like my home for the past six years, and I'm extremely lucky to have overlapped with such a brilliant group of scientists and amazing people. First, I would be nowhere today without the help of Heather Drexler and Blake Tye, who both challenged me and provided valuable advice about all things as I began graduate school. I sincerely miss working with you both but owe much of my independence to your mentorship. Similarly, Kate Lachance helped me get a running start with the computational analyses that predated what is included in this thesis. More importantly, I could always count on her encouragement and support inside and outside the lab. I'm also so thankful to have been able to work with Danya Martell during my rotation in the lab. Danya has provided a fresh perspective that has driven my project forward countless times. I'm grateful to Erik McShane for the many coffee-fueled conversations about everything related to our projects, from which I have learned so much. Erik's passion for science has left me feeling more motivated after each conversation, and this project would not be the

same without him. In the past year, I have grown so much under the mentorship of Robert Ietswaart. Robert's rigor and discipline have made me a much stronger scientist, and I'm lucky to have been able to collaborate with him. I can't imagine what this project would look like today without his input.

My time in the Churchman lab would have been a completely different experience without Karine Choquet, who has filled many shoes as a mentor, a collaborator, and a friend. I'm incredibly grateful for her guidance in each of these roles. Stefan Isaac has a remarkable talent for being able to make me feel better after any conversation, without fail. I can't think of a more considerate and thoughtful colleague than him. Iliana Soto can somehow make any experiment work, and I'm grateful for the time she spent training me on different techniques and for sharing with me some of her endless wisdom. Likewise, Nick Kramer has answered countless questions and shared so much practical advice that helped drive this project to completion. Katja Hansen exudes contagious enthusiasm for science, and I'm thankful for her steadfast optimism. My fellow graduate students in the lab–Aaron Aker, Tommy Tullius, Jake Bridgers, Hope Merens, and Gyan Prakash–have provided a strong sense of community that I can count on for anything, from giving helpful experimental advice, to lending empathetic ears, to being available for much needed coffee breaks or beer hours.

I would also like to thank my undergraduate advisors. Rachel O'Neill welcomed me into her lab as a freshman at UCONN, and that opportunity has changed my life. At the time I had no intention of continuing academic research past my undergrad years, but Rachel's enthusiasm and passion wore off and eventually convinced me to continue on to graduate school. I am also grateful to Brent Graveley for giving me a chance to work in his lab during my time at UCONN on a project that, remarkably, would end up serving as the basis for some of the validation experiments described in this thesis. At Harvard, the members of my dissertation advisory committee–Bob Kingston, Mike Springer, and Susan Shao–provided invaluable scientific feedback across all stages of this project, in addition to helpful advice on how to navigate

graduate school. I'm also thankful to the other faculty in the Program in Genetics and Genomics, including Fred Winston and Scott Kennedy, for their encouragement throughout the years.

I am so grateful to have found friends in graduate school outside the lab who have kept me grounded, sane, and energized. In addition to being a brilliant scientist and an extraordinary collaborator on this project, Emma West is the most selfless person I know and an amazing friend. Nikita Sturrock has been the first to volunteer to help countless times. I'm glad that fate continues to keep us working together since UCONN, and only wish that we had become friends sooner. I'm thankful to Blake Chancellor and Francois Thelot, both of who were two of my earliest friends in graduate school and who I can always count on to distract me from experiments and results when I need them. Additionally, I am honored to have been able to lead the graduate consulting club with two exceptional individuals, Sandrine Degryse and Audrey Sporrij. Working with both of you was incredibly rewarding, yet somehow it never felt like work. Lastly, I am fortunate to have kept in touch with many of my graduate student mentors from UCONN: Sarah Trusiak, Zaq Duda, and Julianna Crivello. You three were my role models for years and have served as my longest-standing support network in the lab; I'm lucky to have remained in your orbit.

Lastly, I'm incredibly fortunate to have the backing of an entire support system completely unrelated to science. Thank you to Davana Bolton, Steven Merola, Asahi Hoque, and Josh Skydel–your friendships predate high school and I can't imagine how different of a person I would be today without each of you. I am forever indebted to my entire family, especially my parents, Paula and Michael Smalec, my sister, Lindsay Smalec, and cousin, Morgan Dickson–you have all been my loudest cheerleaders in every one of my pursuits. And finally, thank you to my fiancee, Sarah Warack, whose support has been unwavering since before I could pipette. I can't begin to list the many ways you've helped me throughout the years, but know that I am beyond grateful for every single one.

If I have seen further it is by standing
on the shoulders of Giants.

–Isaac Newton

CHAPTER 1 –

INTRODUCTION

1.1 INTRODUCTION TO RNA FLOW

The central dogma posits that genetic information is passed from DNA to protein by means of a transient RNA intermediate (Crick, 1958). In particular, messenger RNA (mRNA) transcripts serve as templates for protein-coding genes. In eukaryotic cells, mRNAs are transcribed from DNA within the nucleus on chromatin. Before they can be translated into protein by ribosomes, mRNAs must flow across subcellular compartments, including being released from chromatin into the nucleoplasm and then exported from the nucleus. Once in the cytoplasm, ribosomes are loaded onto transcripts. Finally, mRNAs are translated and ultimately degraded. The shuttling of mRNAs between these subcellular compartments is controlled by many layers of regulatory mechanisms. Therefore, not all mRNAs undergo this stereotypical trajectory through the cell from chromatin to ribosomes, and those that are destined to flow across compartments do not necessarily do so at the same rates. Here I will summarize the regulatory steps that contribute to the different rates and trajectories at which RNAs flow across eukaryotic cells, how these processes impact cell physiology, and the methods that currently exist to quantify RNA turnover with ranging throughputs and at various resolutions.

1.2 RNA FLOW IN THE NUCLEUS

On chromatin, mRNAs are synthesized by RNA polymerase II (pol II). The residence time of mRNAs on chromatin is inherently linked to the kinetics of transcription. The amount of time required to transcribe a gene is a function of gene length and the speed of pol II elongation. A plethora of factors, including RNA processing factors, chromatin modifiers, and transcription elongation factors, contribute to the kinetics of pol II in yeast (Couvillion et al., 2022; Mason and Struhl, 2005), many of which are conserved in mammals. Pol II elongation speed varies up to 5-fold between genes in mammalian cells (Danko et al., 2013), and genes encoding peptides with similar functions, such as those related to the actin cytoskeleton, tend to have similar

elongation rates (Veloso et al., 2014). On a molecular basis, fast transcription rates have been associated with multiple histone modifications (e.g., H3K79me2, H4K20me1), and slower rates with CpG methylation and exon density (Fuchs et al., 2014; Jonkers et al., 2014; Veloso et al., 2014). One element of transcription dynamics that likely contributes to differential elongation speeds is pol II pausing, which is pervasive across gene bodies and influenced by nucleosome positioning and transcription factor binding (Churchman and Weissman, 2011; Mayer et al., 2015; Nojima et al., 2015; Weber et al., 2014). In all, the time required to transcribe an RNA can vary widely between genes and contributes to the kinetics of transcript residence on chromatin.

The elongation rate of pol II not only impacts the kinetics of gene expression but has consequences on the functions and fates of the resulting transcripts. Transcription is physically and mechanistically coupled with mRNA processing, including capping, splicing, and polyadenylation (Buratowski, 2009). Fast elongation generally favors exon skipping, while slow elongation enhances exon inclusion (Howe et al., 2003; Kadener et al., 2001, 2002; de la Mata et al., 2003). This observation is explained by the fact that alternatively spliced exons tend to have weaker splice sites (Zheng et al., 2005); their identification by the spliceosome occurs slower and thus more frequently when elongation is proceeding more slowly, providing additional time for this recognition to take place. Indeed, introns flanking alternative exons are spliced slower than those neighboring constitutive exons (Drexler et al., 2020). Likewise, pol II pausing at exon boundaries is associated with the inclusion of the exon in the final transcript (Mayer et al., 2015). Pol II elongation rates affect not only splicing decisions but also dictate where termination occurs along the gene body (Gromak et al., 2006; Hazelbaker et al., 2013; Plant Kathryn E. et al., 2005). For example, slower pol II elongation in *Drosophila* resulting from a mutation in a polymerase subunit gene promotes the production of *polo* isoforms created using an upstream poly(A) site relative to transcripts made from a downstream site; these shorter mRNAs are not translated as efficiently as the alternative isoforms, resulting in an

abdomen formation defect and lethality (Pinto et al., 2011). Thus, transcription dynamics impact mRNA fate and ultimately regulate the pool of steady-state transcripts.

Independent of transcription, splicing further impacts the time that transcripts reside on chromatin. While splicing can occur both co- and posttranscriptionally, most splicing occurs on chromatin (Choquet et al., 2022). The time required to excise introns from pre-mRNAs varies significantly, ranging from seconds to tens of minutes (Drexler et al., 2020; Martin et al., 2013; Pai et al., 2017; Reimer et al., 2021; Wachutka et al., 2019; Wan et al., 2021). Thus, splicing kinetics impacts the residence times of RNA on chromatin and in the nucleus; furthermore, splicing is a regulator of nascent RNA fate for some mRNAs, dictating when transcripts exit the nucleus. The splicing of chromatin-associated transcripts with retained introns in response to stimuli dictates their release from chromatin (Pandya-Jones et al. 2013; Yeom et al. 2021) and nuclear export (Mauger et al. 2016; Ninomiya et al. 2011). As mentioned, splicing is partially influenced by pol II speed, and the expression of different isoforms from a gene can impact cell fitness. For example, UV-induced DNA damage results in splicing changes of *BCL2L1,* resulting in the production of the pro-apoptotic transcript, rather than the anti-apoptotic isoform transcribed from the same locus. This change in relative isoform abundance is a result of the reduction in pol II elongation speed and drives apoptosis (Muñoz et al. 2009).

At the end of the transcription cycle, termination is mechanistically linked to 3' end processing. Upon the emergence of a poly(A) signal in pre-mRNA, the cleavage and polyadenylation complex (CPA) promotes pol II pausing and termination and stimulates 3' end processing; the cleavage and polyadenylation specificity factor cleaves the pre-mRNA downstream of the poly(A) signal, and the resulting 3' end is polyadenylated by poly(A) polymerase (Proudfoot, 2016). Poly(A) tails are added to an initial length of approximately 200-250 nucleotides in mammalian cells (Darnell et al., 1971; Edmonds et al., 1971) and are quickly bound by nuclear poly(A) binding protein PABPN1 (Wahle, 1991). The addition of a poly(A) tail onto pre-mRNA is necessary for mRNA export from the nucleus, with few exceptions

(Palazzo and Lee, 2018). The most notable exception to this rule is replication-dependent histone genes, which contain unique termination sequences, form a stem-loop at the 3' end, and bind stem-loop binding protein (SLBP) and U7 snRNA (Dominski and Marzluff, 2007).

Splicing, polyadenylation, nuclear export, and nuclear degradation are linked, as perturbations to one pathway have effects on the others, thereby also affecting RNA flow. Splicing is mechanistically coupled with nuclear export, as many export factors are deposited onto mRNA by the spliceosome in its wake (Cheng et al., 2006; Masuda et al., 2005). Consistent with this, reporter constructs that do not require splicing are not exported as efficiently as those with multiple exons (Ryu and Mertz, 1989; Valencia et al., 2008), and deletions in splicing sites result in increased nuclear transcript abundance (Long et al., 1995). Inhibition of splicing also results in transcript accumulation on chromatin (Custódio et al., 1999). In addition, defects in export feedback to 3' end processing. For example, mutations in several export factors in *S. cerevisiae* result in an accumulation of hyperadenylated mRNA on chromatin at the site of transcription (Hilleren and Parker, 2001; Jensen et al., 2001), and a similar phenotype was reported in human cells (Qu et al., 2009). These hyperadenylated transcripts are candidates for nuclear degradation, as several pathways that shuttle mRNAs to the nuclear exosome are known to act on molecules with elongated poly(A) tails (Bresson and Conrad, 2013; Bresson et al., 2015; Meola et al., 2016). These examples highlight the interconnectedness of different steps of gene expression in the nucleus.

Nuclear degradation serves as a means of RNA surveillance, ensuring the quality of the pool of cytoplasmic mRNAs that are translated by the ribosome. Transcripts with defective caps or those that are improperly spliced are degraded by XRN2 (Davidson et al., 2012). The nuclear exosome also degrades transcripts with defects in splicing and polyadenylation (Bresson et al., 2015; Pendleton et al., 2018). Additionally, nuclear degradation serves a regulatory function by acting as a posttranscriptional regulator of cytoplasmic mRNA abundance. For example, in *S. cerevisiae,* the use of alternative transcription termination sites of *RPL9B* results in the

production of either a stable, polyadenylated mRNA or one that is immediately degraded, the latter being predominantly produced when RPL9B protein is in excess (Gudipati et al., 2012). Furthermore, many ribosomal protein genes decrease in nuclear and cytoplasmic transcript abundance during TORC1 inactivation in a nuclear exosome-dependent manner (Bergkessel et al., 2011). Thus, nuclear RNA degradation serves many functions and plays a role in posttranscriptional gene regulation.

By contrast, some loci are tethered to nuclear pore complexes to ensure rapid export of transcripts. This concept, originally proposed as the "gene gating" hypothesis, describes that all RNAs of a transcriptionally-active locus would be rapidly exported through one or a few individual nuclear pores if the DNA was located within close physical proximity (Blobel, 1985). Several examples of this model have been discovered. For example, the *hsp-16.2* locus in *C. elegans* is localized near the nuclear envelope under basal conditions. Upon heat shock induction, the locus becomes anchored at the nuclear pore complex, a process that is dependent on both the presence of heat shock element sequences in the gene promoter and active pol II transcription (Rohner et al., 2013). Similar phenomenon have been observed in human cells, albeit in other contexts. For instance, the *Myc* oncogene becomes physically tethered to the nuclear pore in colon cancer cells. This process results in the increased cytoplasmic abundance of *Myc* transcripts relative to normal colon epithelial cells due to both protection from nuclear mRNA degradation and rapid export (Scholz et al., 2019). In summary, interactions between the transcription and export machineries regulate nuclear and cytoplasmic mRNA transcript abundances, which in turn, can have implications in translation.

1.3 RNA FLOW IN THE CYTOPLASM

Intriguingly, elements related to transcription and mRNA processing can impact mRNA fate and flow in the cytoplasm. For example, sequences present in the promoters of genes

induced in response to glucose starvation in *S. cerevisiae* dictate their cytoplasmic localization and translation levels (Zid and O'Shea, 2014). This implicates the role of transcription factors not only in regulating RNA flow on chromatin but also following nuclear export. Other *cis* elements also play a role in cytoplasmic RNA fate, as different transcript isoforms from the same gene exhibit distinct subcellular localization (Whittaker et al., 1999). For instance, an analysis of hundreds of isoforms with distinct localization patterns in neurons found that transcripts with more distal alternative last exons were enriched in neurites (neuronal projections) as compared to the soma (cell bodies) (Taliaferro et al., 2016). RNA binding proteins Mbnl1 and Mbnl2 were enriched in these exons, and depletion of these two RBPs resulted in the mislocalization of several key neuronal transcripts (Taliaferro et al., 2016). Thus, *cis*-elements that play a role in transcription and splicing affect mRNA fate in the cytoplasm, presumably via their interaction with factors in *trans.*

Once in the cytoplasm, translation initiation factors bind to the 5' end of transcripts and assemble ribosomes; however, this process does not occur with uniform kinetics across all mRNAs. For example, mRNAs with long and structured 5' UTRs require RNA helicases to resolve these structures prior to translation initiation (Lai et al., 2008; Leppek et al., 2018; Parsyan et al., 2009; Pisareva et al., 2008; Soto-Rifo et al., 2012). Furthermore, many mRNAs are not equally distributed throughout the cytoplasm (Benoit Bouvrette et al., 2018), and some require localization within the cytoplasm before translation initiation takes place (Das et al., 2021). For example, translation of *ACTB* at the cell periphery is necessary for fibroblast motility (Kislauskis et al., 1997; Shestakova et al., 2001). This transcript is bound by zipcode binding protein 1 (ZBP1), which transports the mRNA and prevents translation until *ACTB* is localized to the leading edge of the cell membrane (Hüttelmaier et al., 2005; Oleynikov and Singer, 2003; Ross et al., 1997). In this case, the delayed loading of ribosomes onto mRNAs is necessary for proper cell growth and is regulated in a transcript-specific manner.

After nuclear export, poly(A) tails are deadenylated over time (Eisen et al., 2020a; Sheiness and Darnell, 1973). While the length of an mRNA's poly(A) tail in the cytoplasm has been linked to transcript stability and translational efficiency, these relationships depend on the developmental context of the cell. Poly(A) tail length is strongly correlated with translational efficiency in oocytes and early embryonic stages, but this relationship was not observed at any point after gastrulation (Subtelny et al., 2014). These different relationships are at least partially explained by PABPC abundance being lower in early development, favoring its binding to long tails; when in excess, PABPC binds to tails of all sizes (Xiang and Bartel, 2021). Tail length, PABPC binding, and cytoplasmic stability also share a complex relationship. In conditions where PABPC is limited, transcripts with shorter tails are less likely to be bound and protected from the terminal uridylation machinery, which promotes decay (Chang et al., 2014; Xiang and Bartel, 2021). This explains why under post-embryonic conditions with excess PABPC, mRNAs with shorter tails are not necessarily quickly turned over (Lima et al., 2017). Yet, terminal uridylation activity is also regulated throughout development (Chang et al., 2018; Xiang and Bartel, 2021). Ultimately, poly(A) tails are deadenylated to a critically short length, triggering decapping and mRNA degradation (Passmore and Coller, 2022).

Cytoplasmic degradation is further promoted by the activities of microRNAs. In mammalian cells, most microRNAs stimulate the decay of their target mRNAs via the recruitment of deadenylation factors, resulting in poly(A) tail shortening (Bartel, 2018). Moreover, microRNAs further promote the rapid decay of short-tailed mRNAs (Eisen et al., 2020b). Most mature microRNAs have extremely long half-lives (>10 hours, and up to several days) (Duffy et al., 2015; Gantier et al., 2011; Reichholf et al., 2019) and are therefore thought to function in finely tuning gene expression (Bartel, 2018). However, some microRNAs are less stable (Krol et al., 2010; Rissland et al., 2011), allowing for them to serve as transient regulators in response to stimuli. For example, the abundance of a cluster of microRNAs in retinal cells was found to quickly increase in response to light exposure, and conversely, decrease during

8

dark adaptation (Krol et al., 2010). One target of such microRNAs is *Slc1a1,* encoding a voltage-dependent glutamate transporter, the protein product of which increases in photoreceptor cells in the dark and is hypothesized to play a role in depleting glutamate from the visual synapse (Krol et al., 2010). Thus, microRNAs play a role in the cytoplasmic stability of mammalian mRNAs, and consequently, impact the protein output of target transcripts.

In addition to microRNA targeting sites, several sequence elements present on transcripts impact their stability in the cytoplasm. AU-rich elements (AREs) are most commonly found in 3' UTRs and result in mRNA destabilization (Chen and Shyu, 1995), but can also confer stability, depending on the factor that binds the motif (Otsuka et al., 2019). Moreover, 5' terminal oligopyrimidine (5' TOP) motifs are present at the very beginning of transcripts and regulate translation in a posttranscriptional manner. These motifs are present on all human ribosomal protein genes as well as other translation factors and allow for the rapid inhibition of translation when cells are shifted to conditions that require a redistribution of energy and resources away from protein synthesis, such as amino acid starvation (Cockman et al., 2020). Finally, independent of specific motifs, the frequency of different codons present along transcripts has also been linked to mRNA stability. Codon optimality, related to the abundance and charge of the complementary tRNA molecules, is a determinant of cytoplasmic mRNA turnover (Mishima and Tomari, 2016; Presnyak et al., 2015; Wu et al., 2019). In all, many *cis*-factors present on mRNA molecules influence cytoplasmic stability.

Each mRNA contains its own set of features that dictate which regulatory events will act upon the transcript. These features will often also impact the kinetics of each regulatory interaction, which in turn, can affect the rate at which transcripts mature and flow through subcellular compartments. Furthermore, many steps in gene expression are coupled, and the interconnectedness of all these processes from transcription to degradation highlights the complexity of regulation present in eukaryotic cells.

1.4 METHODS TO QUANTIFY RNA TURNOVER AND TRACK RNA FLOW

The kinetics of RNA turnover has been extensively studied across biological systems for decades. These studies have principally relied on the use of two techniques to measure the turnover of constitutively expressed transcripts. First, metabolic labeling employs the use of modified nucleotides containing trackable moieties. Upon addition to cells, these nucleotides are uptaken and incorporated into nascent RNA molecules by polymerases, and the abundance of labeled RNA is measured over time (Figure 1.1A). The second technique is transcriptional inhibition, where RNA polymerases are pharmacologically inhibited, halting the synthesis of all newly transcribed RNA, and the abundance of preexisting RNA is measured over time (Figure 1.1B). Both techniques allow for measurements of RNA half-lives, defined as the time required for 50% of RNA molecules to be turned over in the cell.

Metabolic labeling in HeLa cells with radiolabeled nucleotides revealed mRNA half-lives of several hours (Puckett et al., 1975; Scherrer and Darnell, 1962; Singer and Penman, 1973). These studies provided the first insight into RNA turnover rates in mammalian cells, but the measurements lacked gene-specific resolution. Later, gene-specific mRNA half-lives were measured using this technique with the addition of gene-specific probes, revealing ten-fold variability in turnover between different genes and cellular conditions (Guyette et al., 1979; Harpold et al., 1981; Krowczynska et al., 1985; Levy and Aviv, 1976). Likewise, transcriptional arrest with α-amanitin was used to monitor the turnover of individual genes (Kren et al., 1993; Trembley et al., 1994). Shortly thereafter, transcriptional inhibition with actinomycin D or flavopiridol coupled with microarray hybridization was used to estimate mRNA half-lives genome-wide, showing that genes with related functions have similar rates of turnover (Frevel et al., 2003; Lam et al., 2001; Raghavan et al., 2002; Yang et al., 2003). However, these drugs have been shown to interfere with mRNA decay, resulting in artificially long half-lives (Schwartz, 1973; Shyu et al., 1989; Singer and Penman, 1972). Indeed, direct comparisons of
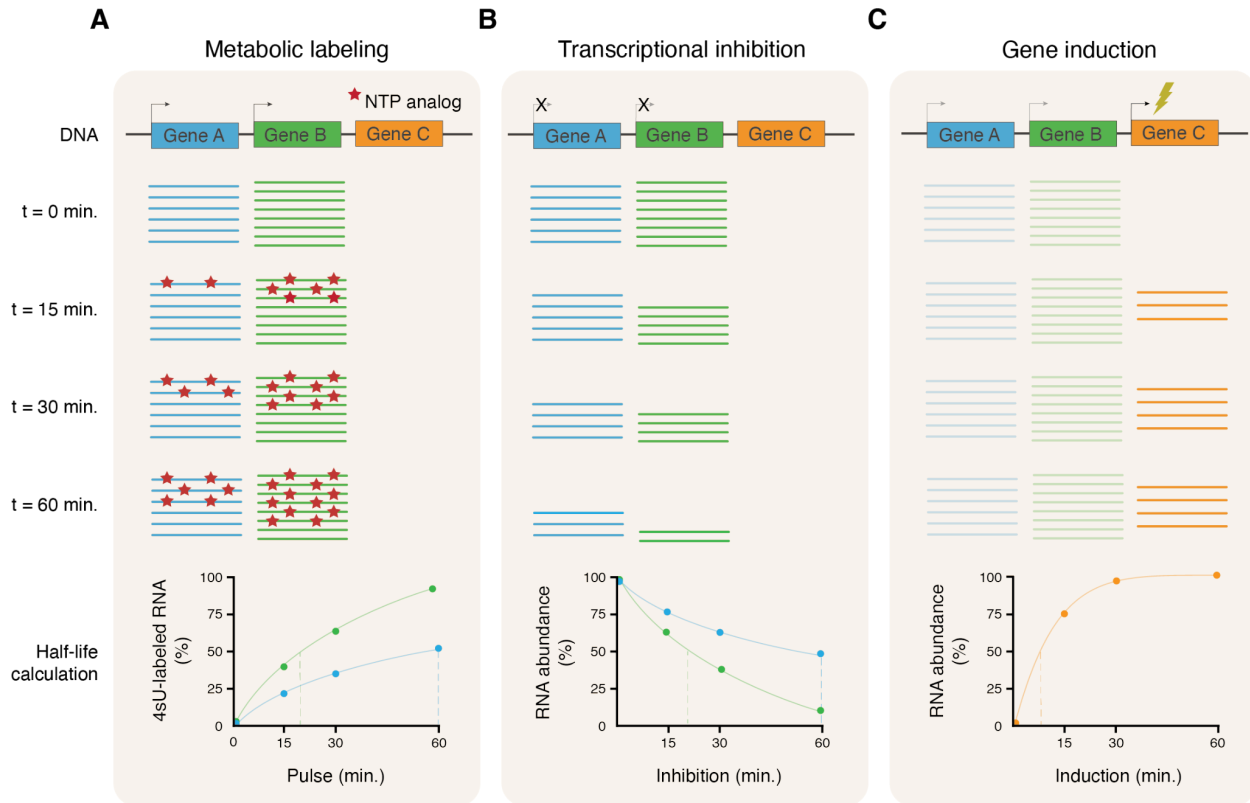
Figure 1.1: Techniques commonly used to measure RNA turnover.
A. Metabolic labeling employs the use of NTP analogs that are uptaken by cells and incorporated into RNA molecules synthesized during the labeling pulse. The proportion of labeled to unlabeled transcripts per gene is tracked over time to determine the RNA half-life (indicated with dotted lines). Colors are consistent for a given gene between DNA (top), RNA (middle), and the simulated data (bottom). **B.** Transcriptional inhibition relies on drugs that halt the synthesis of new RNAs. The abundance of all preexisting transcripts is monitored over time during treatment to determine the RNA half-life (indicated with dotted lines). Colors are consistent with (A). **C.** Genes not expressed are induced by perturbing cells and the abundance of newly synthesized transcripts (e.g., Gene C) encoding such genes are monitored over time to determine RNA half-life (indicated with the dotted line). Colors are consistent with (A).

gene-specific RNA half-lives generated from metabolic labeling and transcriptional inhibition

experiments show high correlations between methods, but longer absolute measurements

estimated in cells exposed to drug treatments (Dölken et al., 2008; Eisen et al., 2020a; Herzog

et al., 2017).

Over time, different modified nucleotides were developed for the use of metabolic

labeling; in particular, 4-thiouridine (4sU) has proved useful. The thiol group of the 4sU molecule

exhibits specific reactivity with mercurated cellulose or agarose, thus allowing for the affinity

purification of 4sU-labeled RNAs (Johnson et al., 1991; Melvin et al., 1978) and subsequent

quantification of RNA half-lives (Johnson et al., 1991; Kenzelmann et al., 2007). The thiol group can also be biotinylated, allowing half-lives to be measured genome-wide by quantifying 4sU-labeled RNAs by microarray (Cleary et al., 2005; Dölken et al., 2008; Friedel et al., 2009) or RNA sequencing (Munchel et al., 2011; Rabani et al., 2011; Schwanhäusser et al., 2011). Most recently, the thiol reactivity has been exploited to cause reverse transcriptase to misincorporate guanine, rather than adenine, into the elongating cDNA upon reading the 4sU molecule in RNA. This results in T>C sequencing mismatches upon alignment of RNA-seq reads, allowing the simultaneous quantifications of both 4sU-labeled and the total pool of RNA, and is the basis of "nucleotide conversion" protocols such as TimeLapse-seq (Schofield et al., 2018), SLAM-seq (Herzog et al., 2017), and TUC-seq (Riml et al., 2017). These methods allow genome-wide quantification of mRNA half-lives from a single sequencing library.

While all these methods were originally used to measure RNA turnover at whole cell resolution, they have also been modified to examine subcellular transcript half-lives. Classical experiments combining transcriptional inhibition, radionucleotide labeling, and biochemical fractionation showed differential mRNA half-lives in the nucleus and cytoplasm (Attardi et al., 1966; Penman et al., 1963; Soeiro et al., 1966) and tracked the lag in mRNA movement from the nucleus to the cytoplasm and onto polysomes (Penman et al., 1968). While lacking gene-specific information, these studies provided the first evidence that RNAs flow across compartments with a measurable delay, on the scale of tens of minutes to hours. 4sU labeling has been combined with cellular fractionation and RNA-seq to study rates of nuclear export and cytoplasmic degradation in *Drosophila* (Chen and van Steensel, 2017). This study revealed differences in export rates between genes, which are partially dictated by 3' UTR length, and implicated many RNA binding proteins in regulating cytoplasmic mRNA stability. Recently, the kinetics of polysome loading was examined in mouse cells by combining 4sU labeling with ribosome profiling (Schott et al., 2021), allowing for the measurement of the delay between cytoplasm entry of mRNAs and their association with the ribosome (~20 minutes in

12

macrophages, ~35 minutes in mESCs), and complete ribosome loading (an additional ~20-35 minutes). Thus, the combination of 4sU labeling and various biochemical fractionations is a promising technique to analyze RNA turnover and flow, although more comprehensive measurements of the RNA lifespan have yet to be generated.

A third classical technique to measure RNA turnover relies on inducing the expression of transcripts and monitoring the abundance of those RNAs over time (Figure 1.1C). For example, immediate early genes, such as *c-Fos,* can be induced by the addition of serum to cells growing in culture and generally have short half-lives of less than 60 minutes (Greenberg and Ziff, 1984; Lau and Nathans, 1985; Shyu et al., 1989). This principle can be extrapolated to measure the turnover of any individual transcript by cloning the gene downstream of an inducible promoter (Helms and Rottman, 1990). However, this technique removes the gene from its endogenous genomic context and yields half-lives in cellular conditions where the RNA may not normally be expressed, thus altering both the *cis* and *trans* factors that may impact turnover. Gene induction has also been used to quantify RNA flow with higher throughput by combination with RNA-seq. For example, the nuclear mRNA half-lives for 282 genes in mouse macrophages activated in response to LPS were estimated to fall between ~5-90 minutes (median 19.5 minutes) (Battich et al., 2015; Bhatt et al., 2012). RNA processing rates were measured in a similar manner for genes induced by this same perturbation (Rabani et al., 2014). While lacking precise spatial information, this nonetheless provides insight into the rates of RNA maturation at finer resolution than whole-cell turnover. In all, gene induction can be used to measure both RNA turnover and flow, yet it remains unclear whether the kinetics observed for these transcripts are generalizable to all genes and normal growth conditions.

Finally, microscopy can serve as an orthogonal technique to track mRNA flow across compartments. Live imaging of reporter constructs in mammalian cells has revealed time lags of tens of minutes between transcription initiation and cytoplasmic appearance of transcripts (Hoek et al., 2019; Mor et al., 2010; Shav-Tal et al., 2004), and that mRNAs can undergo translation

within minutes of nuclear export (Halstead et al., 2015; Hoek et al., 2019). However, these rely on the use of reporter constructs which may not be processed with the same kinetics as endogenous genes, and by design, cannot be used to assay gene-to-gene variability. Alternatively, the abundance and subcellular location of endogenous transcripts can be detected by RNA FISH. This data has been combined with mathematical modeling to estimate nuclear and cytoplasmic RNA half-lives of a handful of genes (n=9) in mouse tissue, revealing greater than 100-fold variability in compartment-specific half-lives between genes (Bahar Halpern et al., 2015). Likewise, rates of transcription initiation, elongation, chromatin release, intron processing and degradation, and cytoplasmic degradation was measured for a single endogenous gene, *FLC*, in *Arabidopsis* (Ietswaart et al., 2017; Wu et al., 2016). Thus, microscopy-based approaches can precisely quantify RNA flow but are limited by low throughput.

In all, many techniques exist which can be used to quantify RNA turnover and track transcripts through the cell with various throughputs and resolutions. However, there is a need for a method that comprehensively and quantitatively analyzes RNA flow from synthesis to degradation for all endogenous genes in a single assay in the absence of cellular perturbations.

1.5 OUTLINE OF THIS THESIS

In this thesis, I introduce the concept of examining "RNA flow," defined as the rates at which RNAs traverse across subcellular compartments. In Chapter 2, I present a technique, subcellular TimeLapse-seq, and a framework for RNA kinetic modeling that can be used to quantify RNA flow and estimate subcellular half-lives in mammalian cells. I also present genome-wide trends observed after applying this method to a human and mouse cell line. In Chapter 3, I dissect the variability observed across all genes to pinpoint what genetic and molecular features may underlie RNA flow rates. I conclude by highlighting the major outstanding questions related to RNA flow in Chapter 4.

CHAPTER 2 –

MEASUREMENTS OF RNA FLOW

This chapter is adapted from a manuscript currently under review entitled "Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle" and will be published at a later point in a peer-reviewed journal. I wrote the manuscript with input from Stirling Churchman and Robert Ietswaart. Emma West helped design, perform, and analyze the RNA FISH experiments. Robert Ietswaart performed the modeling, performed the clustering and GO enrichment, and helped with the pipeline to analyze subcellular TimeLapse-seq. Erik McShane helped with the pipeline to analyze subcellular TimeLapse-seq and with the design of the NanoStrings assay. I designed, performed, and analyzed all other experiments with input from Stirling Churchman.

2.1 INTRODUCTION

The life cycles of mRNAs are dynamic and diverse. Thousands of mRNAs are produced per minute in a typical mammalian cell (Carter et al., 2005; Schofield et al., 2018). Before they can be translated, mRNAs must flow across subcellular compartments, including release from chromatin and export from the nucleus. In the cytoplasm, ribosomes are loaded onto mRNAs, and transcripts are ultimately degraded. These transitions between compartments are controlled by numerous regulatory mechanisms. Accordingly, not all mRNAs are destined for this stereotypical trajectory, and those that are, do not flow across the cell at the same rates. In total, RNA half-lives vary greater than 100-fold between different protein-coding transcripts (Dölken et al., 2008; Friedel et al., 2009; Herzog et al., 2017; Rabani et al., 2011; Schofield et al., 2018; Schwanhäusser et al., 2011). However, half-lives measured at the whole-cell level only measure the time between synthesis and decay, obscuring the dynamics of mRNA transitions across subcellular compartments.

Around 50 years ago, metabolic labeling in mammalian cells with radiolabeled nucleotide precursors shed light on bulk RNA flow and metabolism, but these experiments could not resolve transcript-specific behaviors (Darnell et al., 1973). Since then, several methods have been developed to assay the rates of RNA flow for one or a few transcripts. For example, single-molecule microscopy approaches track reporter RNAs throughout mammalian cells (Halstead et al., 2015; Hoek et al., 2019; Mor et al., 2010; Shav-Tal et al., 2004). Endogenous

RNAs have been studied using single-molecule RNA FISH combined with mathematical modeling, yielding nuclear and cytoplasmic RNA half-lives of a handful of genes in mouse tissue (Bahar Halpern et al., 2015) and allowing for the quantification of the entire life cycle of an individual transcript in *Arabidopsis* (Ietswaart et al., 2017; Wu et al., 2016). To determine the rates at which RNAs flow across compartments with higher throughput, induced genes have been monitored over time and modeled to estimate subcellular turnover (Battich et al., 2015; Bhatt et al., 2012; Rabani et al., 2014), yet the kinetics observed for these transcripts may not extend to all genes or to cellular contexts that do not involve gene induction. Recently, metabolic labeling has been used to globally study specific stages of RNA lifespans (Berry et al., 2022; Chen and van Steensel, 2017; Schott et al., 2021), but these studies have not comprehensively characterized the entire life cycle of an mRNA across multiple compartments in mammalian cells, and none have investigated degradation in the nucleus.

Here, we quantify the rates of mRNA flow across mammalian cells genome-wide. We start by introducing subcellular TimeLapse-seq, a method that measures RNA turnover with subcellular resolution, and couple this technique with kinetic modeling to estimate the rates at which mRNAs flow across subcellular compartments. We measured RNA half-lives on chromatin, in the nucleus, and in the cytoplasm, and additionally measured nuclear export and polysome loading rates for all expressed genes in mouse NIH-3T3 and human K562 cells. Strikingly, for ~5–10% of genes, mRNA flow was predicted to involve substantial nuclear degradation. We found that RNA flow rates varied widely (sometimes >100-fold) between different genes and subcellular compartments. Our results demonstrate that functionally related genes undergo similar rates of RNA flow, implicating its role in gene regulation.

2.2 RESULTS

2.2.1 Subcellular TimeLapse-seq measures the fraction of newly synthesized RNA across subcellular compartments

We developed subcellular TimeLapse-seq to quantify RNA turnover genome-wide with subcellular resolution. This technique combines metabolic labeling with the biochemical purification of distinct RNA populations from different cellular compartments (Figure 2.1A). We first pulse labeled human K562 and mouse NIH-3T3 cells with 4-thiouridine (4sU) for 0, 15, 30, 60, and 120 minutes. As high levels of 4sU concentrations broadly impact gene expression (Burger et al., 2013), we minimized the concentration and duration of 4sU exposure. For each cell type analyzed, we confirmed that the addition of 4sU did not affect mRNA subcellular localization (Figure S2.1A-B) and led to minimal disruption of gene expression levels within each subcellular compartment (Figure S2.1C). Following each 4sU pulse, we biochemically purified five populations of RNA corresponding to distinct subcellular compartments. We isolated chromatin-associated, nuclear, and cytoplasmic RNA, confirming the purity of the fractions by western blot analysis (Figure 2.1B) (Mayer and Churchman, 2017). We also isolated polysome-bound RNA from the puromycin-sensitive fractions of sucrose gradients (Figure S2.1D). In addition, we collected total cellular RNA for comparison. We then quantified the amount of newly synthesized RNA within each of the five populations at each time point by performing TimeLapse-seq, a nucleotide conversion protocol that detects 4sU-labeled RNA within a mixture of labeled and unlabeled RNAs by the presence of 4sU-induced T>C sequencing mismatches (Figure 2.1A) (Schofield et al., 2018).

As expected, we observed an increase in the proportion of reads containing T>C mismatches with longer pulse times when examining the data corresponding to individual compartments for single genes (e.g., *Tfam,* Figure S2.2A). Additionally, we saw an increase in the frequency of T>C mismatches on reads containing these mismatches, indicating a higher
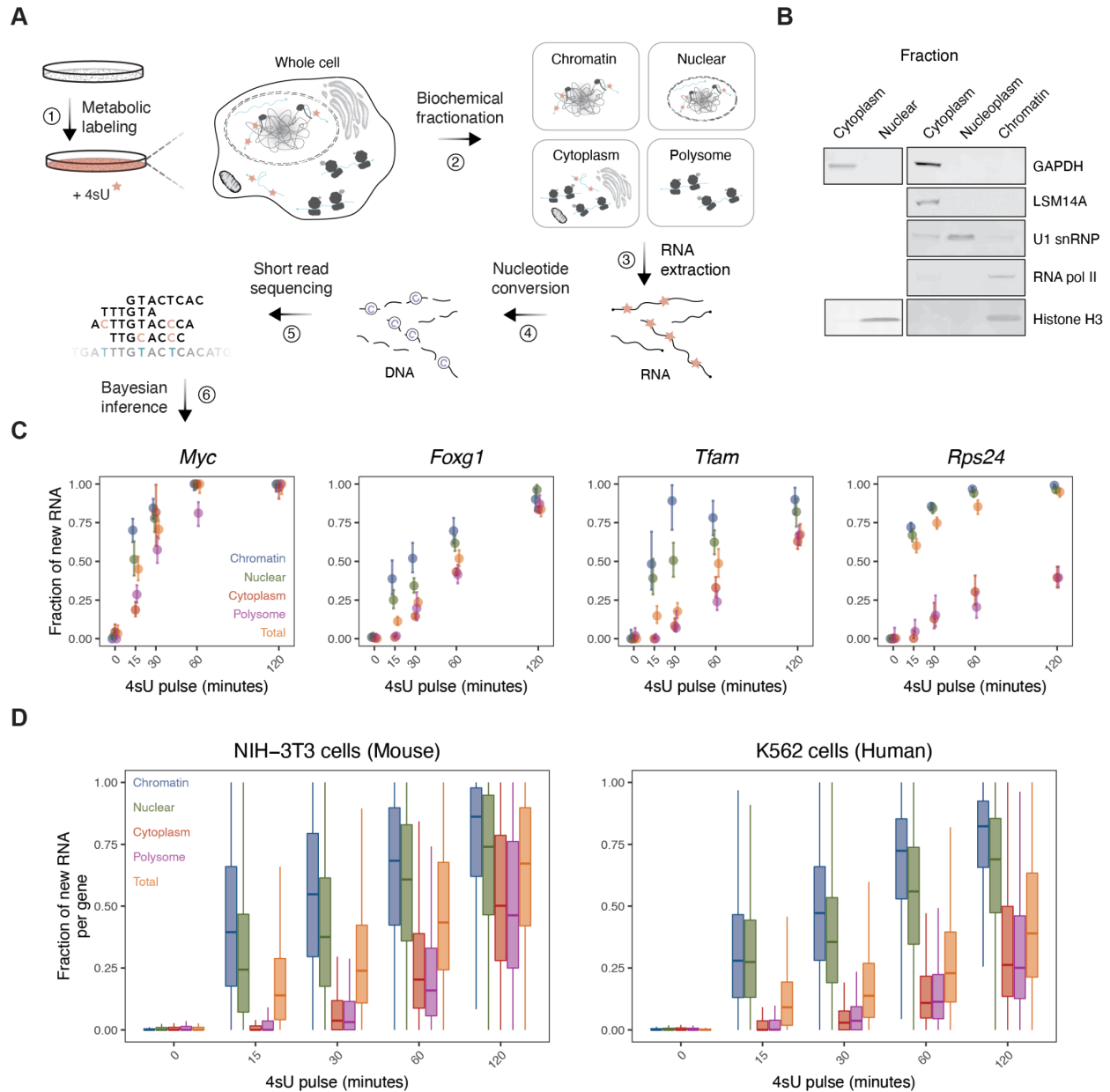
Figure 2.1: Subcellular TimeLapse-seq combines metabolic labeling and biochemical fractionation to quantify the fraction of new RNA across compartments.
A. Schematic representing subcellular TimeLapse-seq. Cells were pulse-labeled with 4-thiouridine (4sU) and biochemically fractionated into subcellular compartments. TimeLapse-seq libraries were prepared from each sample and the fraction of newly synthesized RNA per gene was estimated (see Fig. S2.2C and Methods for details). **B.** Purification of distinct subcellular compartments by biochemical fractionation. The fractionation efficiency was monitored by western blotting for marker proteins: GAPDH and LSM14A, cytoplasmic proteins; U1 snRNA, nucleoplasmic protein; histone H3 and RNA pol II, chromatin proteins. **C.** Subcellular TimeLapse-seq data for example genes *(Myc, Foxg1, Rps24)* in mouse NIH-3T3 cells. Dots represent the fraction of new RNA MAP values for one replicate, while vertical lines represent the 95% credible intervals (CIs). **D.** Genome-wide subcellular TimeLapse-seq data for all protein-coding genes in human K562 and mouse NIH-3T3 cells. Fraction of new RNA MAP values for each gene are shown for one replicate.

19

rate of 4sU incorporation with longer pulse times (Figure S2.2A). This was true when looking across all reads and all compartments (Figure S2.2B). Previous applications of nucleotide conversion approaches, such as TimeLapse-seq and SLAM-seq, have used labeling times of at least 2 hours, our longest pulse, and/or at least double the 4sU concentrations we used (Erhard et al., 2019; Herzog et al., 2017; Schofield et al., 2018). Therefore, the lower 4sU incorporation rates inherent to our minimal labeling conditions necessitated a new approach to estimate T>C conversion rates that are required for inferring the fraction of newly synthesized RNA for each gene. We developed a binomial mixture model and estimated upper and lower T>C conversion rates present in newly synthesized RNAs, as well as a background T>C conversion rate present in preexisting RNAs, for each time point and compartment (Figure S2.2C). We then input these conversion rates into GRAND-SLAM (Jürges et al., 2018) and combined the outputs to quantify the posterior distribution on the fraction of newly synthesized RNA per gene within each sample (Figure S2.2C).

We sought to validate our analysis approach due to the low 4sU incorporation rates and the implementation of a new analysis pipeline. In parallel with subcellular TimeLapse-seq, we performed a NanoStrings-based assay to quantify the fraction of newly synthesized RNA present in nuclear, cytoplasm, and total RNA populations for select genes (Figure S2.3A). This technique relies on biotinylation and bead selection to separate 4sU-labeled and unlabeled mRNA molecules and probes that directly hybridize with mRNA. Thus, this assay does not rely on accurate computational predictions of T>C conversions or sequencing. We observed similar fractions of new RNA by these two techniques across all compartments for genes with fast and slow turnover (Figure S2.3B-C), validating the robustness of our analysis pipeline across a range of mismatch rates.

After determining the fraction of newly synthesized RNA across all samples, we analyzed the data genome-wide. Within each compartment, we observed an increase in the proportion of new RNA with increasing pulse durations (Figure 2.1C-D). Furthermore, at each

20

time point, we see delays in the fraction of new RNA across chromatin, nuclear, cytoplasm, and polysome fractions (Figure 2.1C-D). This finding was true across all genes, even those with relatively fast turnover, such as Myc (Figure 2.1C). This result confirmed that our assay has proper time resolution to detect RNA flow across subcellular compartments.

2.2.2 Kinetic modeling of RNA flow across subcellular compartments

To estimate the rates at which RNAs flow across subcellular compartments for each gene, we fit a kinetic model consisting of a system of ordinary differential equations to our subcellular TimeLapse-seq data (Figure 2.2A). By coupling this model to the Bayesian inference framework of GRAND-SLAM (Figure S2.4A), we estimated the Bayesian posterior probability distribution of each flow rate per gene (Figure 2.2B-D, Table S2.1). Using the posterior mean rate (k), half-lives are then calculated as $t_{1/2} = \frac{ln(2)}{k}$. We determined the half-lives of RNAs on chromatin ("chromatin half-lives"), in the nucleus ("nuclear half-lives"), and in the cytoplasm ("cytoplasm half-lives"). We estimated rates of nuclear export (yielding "nuclear export half-lives") and the rates at which polysomes are loaded onto RNAs after entry into the cytoplasm (yielding "untranslated cytoplasm half-lives"). Finally, we estimated the rates at which RNAs are turned over at the whole-cell level (yielding "whole-cell half-lives"), using only the total RNA data. The maximum (*maximum a posteriori,* MAP) of the posterior indicates the most likely timescale value given the observed subcellular TimeLapse-seq data (Figure S2.4A). The 95% credible interval (CI) is a measure of uncertainty that indicates the range of timescales consistent with the data (Figure S2.4A).

This initial model fit the data for most genes well (e.g. *Myc* and *Foxg1*, Figure 2.2B, Table S2.1). The model predicts that whole-cell half-lives represent the sum of the nuclear and
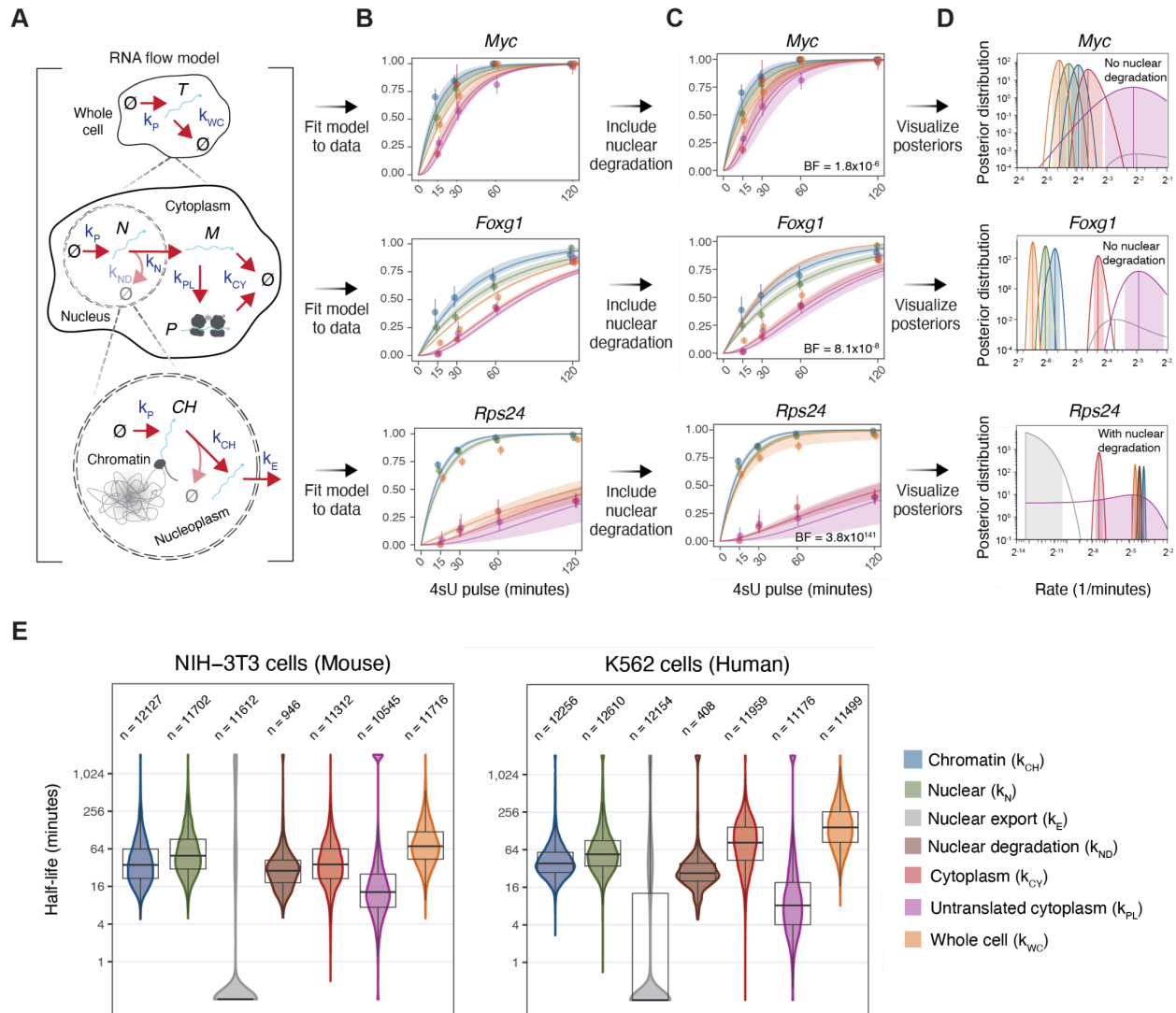
Figure 2.2: Kinetic modeling of RNA flow rates across subcellular compartments.
A. Schematic of RNA flow model. Flow rates: k_P = production rate, k_{WC} = turnover of total RNA (at whole-cell resolution), k_N = turnover of nuclear RNA, k_{ND} = nuclear degradation rate, k_{PL} = polysome loading rate, k_{CY} = turnover of cytoplasm RNA, k_{CH} = turnover of chromatin RNA, k_E = nuclear export rate. RNA populations: T = total RNA, N = nuclear RNA, M = non-polysome-bound cytoplasmic RNA, P = polysome-bound cytoplasmic RNA, CH = chromatin RNA. **B.** RNA flow model fit to subcellular TimeLapse-seq data for example genes *(Myc, Foxg1, Rps24)* in mouse NIH-3T3 cells in the absence of a nuclear degradation parameter **C.** RNA flow model fit to subcellular TimeLapse-seq data for the example genes shown in (B) with a nuclear degradation parameter. The Bayes factor (BF) for each gene is indicated. **D.** Posterior distributions for each RNA flow rate for the example genes shown in (B-C). Nuclear degradation is only shown if including this parameter improves the fit of the data (Bayes factor >=100). **E.** Genome-wide subcellular half-lives for all protein-coding genes in mouse NIH-3T3 and human K562 cells. Nuclear degradation rates are only included for genes best explained by this model. Mean half-lives are shown and the number of genes noted.

cytoplasmic half-lives, because it did not include nuclear degradation of mRNA (Schmid and

Jensen, 2018) (Figure 2.2A). However, in the presence of substantial nuclear RNA degradation,

this relation will no longer hold true, as many RNAs never exist in the cytoplasm and whole-cell data will no longer be fit by the model. We noticed this trend for a small group of genes in both cell lines, including *Rps24* in NIH-3T3 (Figure 2.2B). Therefore, we extended our model by including a nuclear RNA degradation rate. To determine if adding this parameter yields a better fit for each gene, we calculated a Bayes factor (Kass and Raftery, 1995), namely, the ratio of likelihoods between the nuclear degradation model (alternative hypothesis) and the model with no nuclear degradation (null hypothesis). In the absence of nuclear degradation, we were not able to predict the total RNA data for *Rps24* (Figure 2.2B), whereas a model with nuclear degradation was successful (Figure 2.2C, Bayes factor=10^{141}).

The chromatin, nuclear, nuclear export, nuclear degradation, cytoplasm, and whole-cell half-lives showed a large variation between genes (Figure 2.2E), generally with small CIs (Table S2.1), indicating the subcellular TimeLapse-seq data determined these timescales with high certainty. We observed a strong correspondence with previously reported whole-cell half-lives estimated using TimeLapse-seq in K562 (Figure S2.4B, Pearson correlation r=0.65). Despite the variability between genes, we see strong correlations between biological replicates for these half-lives (Figure S2.4C-D, Pearson correlation r>0.75). Most genes exhibit a relatively larger CI for the untranslated cytoplasm half-life compared to other flow rates, and the mean timescales for this rate also exhibited a larger variation between biological replicates (Figure S2.4C-D, r=0.44 in K562 and r=0.45 in NIH-3T3). Nevertheless, the CIs reproduce between replicates (overlap for 86% genes in NIH-3T3, 79% genes in K562) and do not exceed the variation between genes.

To assess the robustness of our RNA flow rates across different modeling approaches, we compared our Bayesian MAP values to the more commonly used least squares estimates (Figure S2.5A-B). The correspondence between the timescales of the two models was strong for the chromatin, nuclear, cytoplasmic and whole-cell half-lives (Figure S2.5A-B, r>0.67), indicating our estimates for these populations are robust to modeling approaches. For the

untranslated cytoplasm half-lives, correspondence was weaker (Figure S2.5A-B, r=0.42 in K562, r=0.17 in NIH-3T3) due to the inability of the least squares model to take the measurement uncertainties into account. We conclude that our Bayesian model is more robust and informative than the least squares model, as the former provides uncertainty measures in the form of CIs. Overall, we conclude that our approach robustly quantifies the flow of RNAs across subcellular compartments for endogenously expressed genes in mouse and human cells.

2.2.3 Wide gene-to-gene variability in RNA flow rates

In both cell lines, RNA flow rates varied considerably between genes (Figure 2.2E). For 90% of genes, chromatin half-lives ranged from 19 to 120 minutes in K562 and from 14 to 200 minutes in NIH-3T3, with medians around 40 minutes. Nuclear RNA half-lives in both cell lines were often somewhat longer than, but highly correlated with, chromatin half-lives (r>=0.75, Figure S2.3A-B), with median nuclear half-lives around 45 minutes that ranged between 20 and greater than 260 minutes for 90% of genes. As expected from the similarity between chromatin and nuclear half-lives for most genes, we measured very rapid nuclear export rates. In both cell lines, we measure nuclear export half-lives of <1 minute for more than half of all genes, with >70% of all genes having nuclear export half-lives of <10 minutes. We conclude that for most genes, a majority of mRNAs within the nucleus are associated with chromatin, as the chromatin half-life is >50% of the nuclear half-life for >75% of genes in both cell lines (Figure 2.3C). Still, we occasionally observed slow nuclear export and estimated longer export half-lives (>30 minutes) for at least 10% of genes in both K562 and NIH-3T3.

In the cytoplasm, mRNAs were more stable in K562 cells with a median half-life of 78 minutes, compared to 36 minutes in NIH-3T3. In both cell lines, cytoplasm half-lives varied greatly, ranging from 10 to 140 minutes for 90% of genes in NIH-3T3 and from 20 to 260
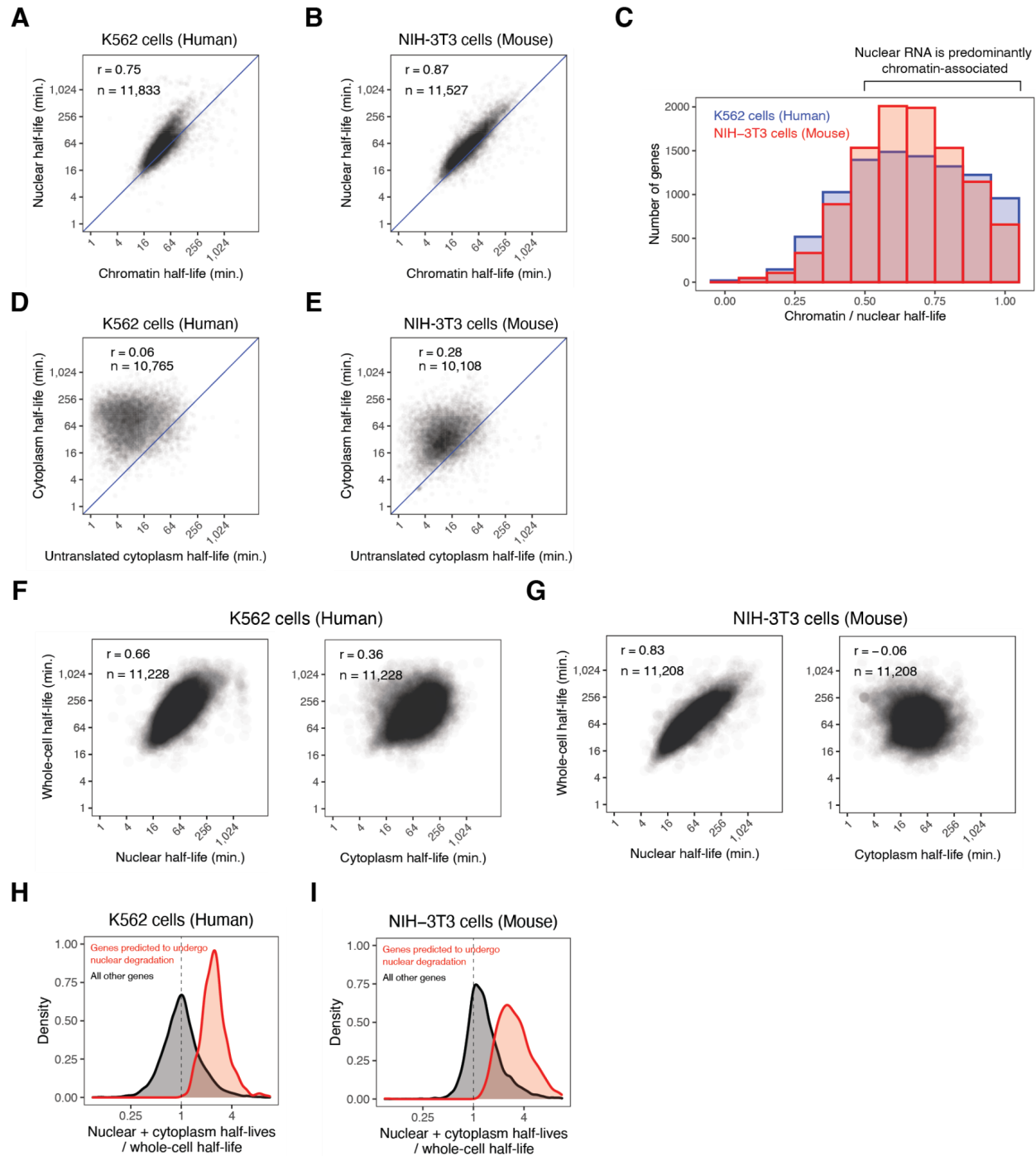
Figure 2.3: RNA flow rates show consistent genome-wide patterns.
A. Correlation between chromatin and nuclear RNA half-lives in human K562 cells. Mean half-lives
for each rate are compared with the Pearson correlation and number of genes noted. **B.** Same as (A) in
mouse NIH-3T3 cells. **C.** Distribution of the ratio of chromatin to nuclear half-lives, showing that nuclear
RNA is predominantly chromatin-associated (ratio>0.5) for a majority of genes. **D.** Correlation between
untranslated cytoplasm and cytoplasm half-lives, showing that these rates are not related in human K562
cells. Mean half-lives of each rate are compared with the Peason correlation and number of genes noted.
E. Same as (D) in mouse NIH-3T3 cells. **F.** Comparison of nuclear half-lives or cytoplasm half-lives to
whole-cell half-lives in human K562 cells. Mean nuclear half-lives (left) or mean cytoplasmic half-lives

minutes in K562 cells. Notably, differences in cytoplasm mRNA stability did not correspond with

differences in rates of polysome loading, which occurs relatively quickly after nuclear export. In

both cell lines, the median untranslated cytoplasm half-life was less than 15 minutes, with more

than 75% of genes having half-lives less than 30 minutes. Untranslated cytoplasm half-lives and

cytoplasm half-lives were uncorrelated in K562 (r=0.06) and weakly correlated in NIH-3T3

(r=0.28) (Figure 2.3D-E). Thus, loading of RNAs onto polysomes is not strongly coupled with

cytoplasmic RNA turnover.

Consistent with RNAs being more stable in the cytoplasm of K562 cells, we observed

slower turnover of mRNAs at whole cell resolution in K562 cells compared to NIH-3T3 cells,

with median whole-cell half-lives of 140 and 71 minutes, respectively. This is consistent with a

previous study showing faster mRNA turnover in mouse embryonic fibroblasts as compared to

K562 cells (Schofield et al., 2018). Thus, we see that while mRNAs are released from

chromatin, exported from the nucleus, and initially loaded onto polysomes with similar kinetics

between K562 and NIH-3T3, mRNAs are more stable in the cytoplasm of K562, resulting in

overall slower rates of turnover in this cell line. As a result, we see that on average, mRNAs for

most genes spend longer in the nucleus than the cytoplasm in NIH-3T3 cells, while the opposite

is true in K562. Interestingly, in both cell lines, we observe a stronger correlation between

nuclear residence and whole-cell residence (r = 0.66 in K562, r = 0.83 in NIH-3T3) than

cytoplasm residence and whole-cell residence (r = 0.36 in K562, r = -0.06 in NIH-3T3) (Figure

2.3F-G). Thus, even when mRNAs are more stable in cytoplasm than in the nucleus, nuclear

residence is better at predicting overall turnover rates than cytoplasm mRNA stability. As

expected, for most genes, we are able to predict the whole-cell half-life by adding the nuclear

and cytoplasm half-lives (Figure 2.3H-I). However, we see that the predicted whole-cell half-life is much longer than the observed whole-cell half-life for genes with nuclear degradation, as would be expected if many RNAs never reside in the cytoplasm (Figure 2.3H-I).

2.2.4 Genes encoding transcripts predicted to undergo nuclear degradation (PUNDs) are conserved across mouse and human cells

The Bayes factors for 9% (n= 946/11,109) of genes in NIH-3T3 cells and 4% (n= 408/10,971) of genes in K562 cells reproducibly exceeded 100 (Figure 2.4A, S2.6A, Table S2.2), indicating that the model with nuclear degradation is at least 100 times more likely to explain the subcellular TimeLapse-seq data than the model without nuclear degradation, thus providing decisive evidence (Kass and Raftery, 1995) that transcripts of these genes undergo nuclear degradation. Hereafter, we refer to these genes as PUNDs (predicted to undergo nuclear degradation). The model predicts that the large majority (>85%) of transcripts produced by PUNDs are degraded in the nucleus.

We performed Gene Ontology (GO) enrichment analysis on our PUND gene lists (Table S2.3) and found many common enriched terms between K562 and NIH-3T3 PUNDs, including those pertaining to the ribosome, RNA splicing, and nuclear mRNA export (Figure 2.4A, S2.6A). Ribosomal protein genes were highly enriched in PUNDs in both cell lines, including 24 individual ribosomal protein gene homologs. In total, we found 133 homologous genes that were shared between both cell lines, a significant overlap (Fisher's exact test: $p=2.3\times10^{-14}$) (Figure S2.6B). Given the common functions of PUNDs, as well as the overlap of individual PUNDs between cell lines, we conclude that nuclear degradation is a conserved regulatory feature that acts on select transcripts.
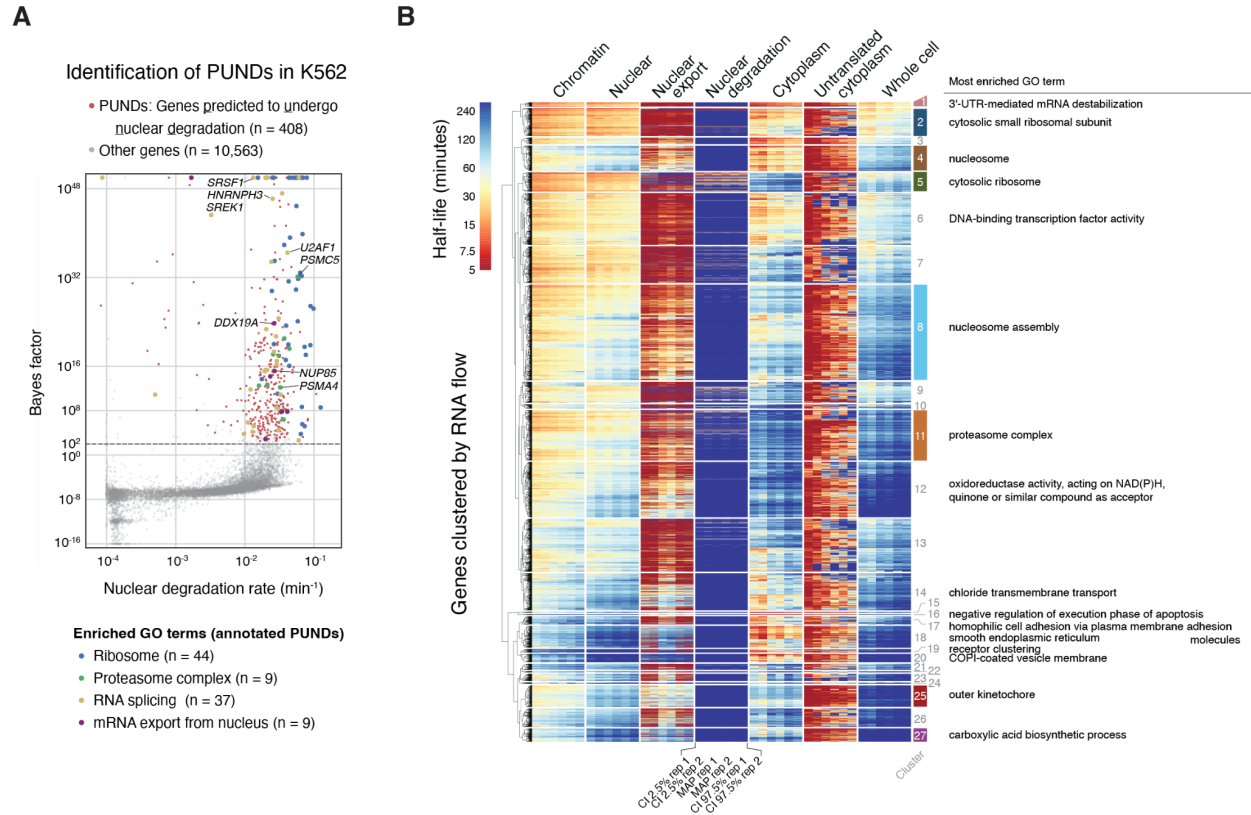
Figure 2.4: Genes with common functions experience similar RNA flow patterns.
A. Genes with a Bayes factor >100 in both replicates were labeled as those predicted to undergo nuclear degradation (PUNDs, shown in red) in human K562 cells. PUNDs are additionally colored by their associated enriched GO terms. **B.** Hierarchical clustering of human genes according to their RNA flow rates (MAPs and 95% credible intervals). The most enriched GO annotations for each cluster are displayed on the right (full results in Table S2.3).

2.2.5 Functionally related genes exhibit similar RNA flow across subcellular compartments

Genes with related functions tend to be co-regulated such that their mRNAs have similar rates of whole-cell turnover (Dölken et al., 2008; Friedel et al., 2009; Herzog et al., 2017; Rabani et al., 2011; Schofield et al., 2018; Schwanhäusser et al., 2011). To determine whether subcellular RNA flow rates may likewise serve a regulatory role, we first performed hierarchical clustering on all genes in human K562 cells based on their subcellular half-lives (Figure 2.4B). By allowing for sufficient granularity, we identified 27 total clusters ranging from the order of 10 to 10^3 genes with reproducibly distinct transcript kinetics (Figure 2.4B). We then performed GO enrichment analysis on each cluster and found that a majority have several enriched terms

28
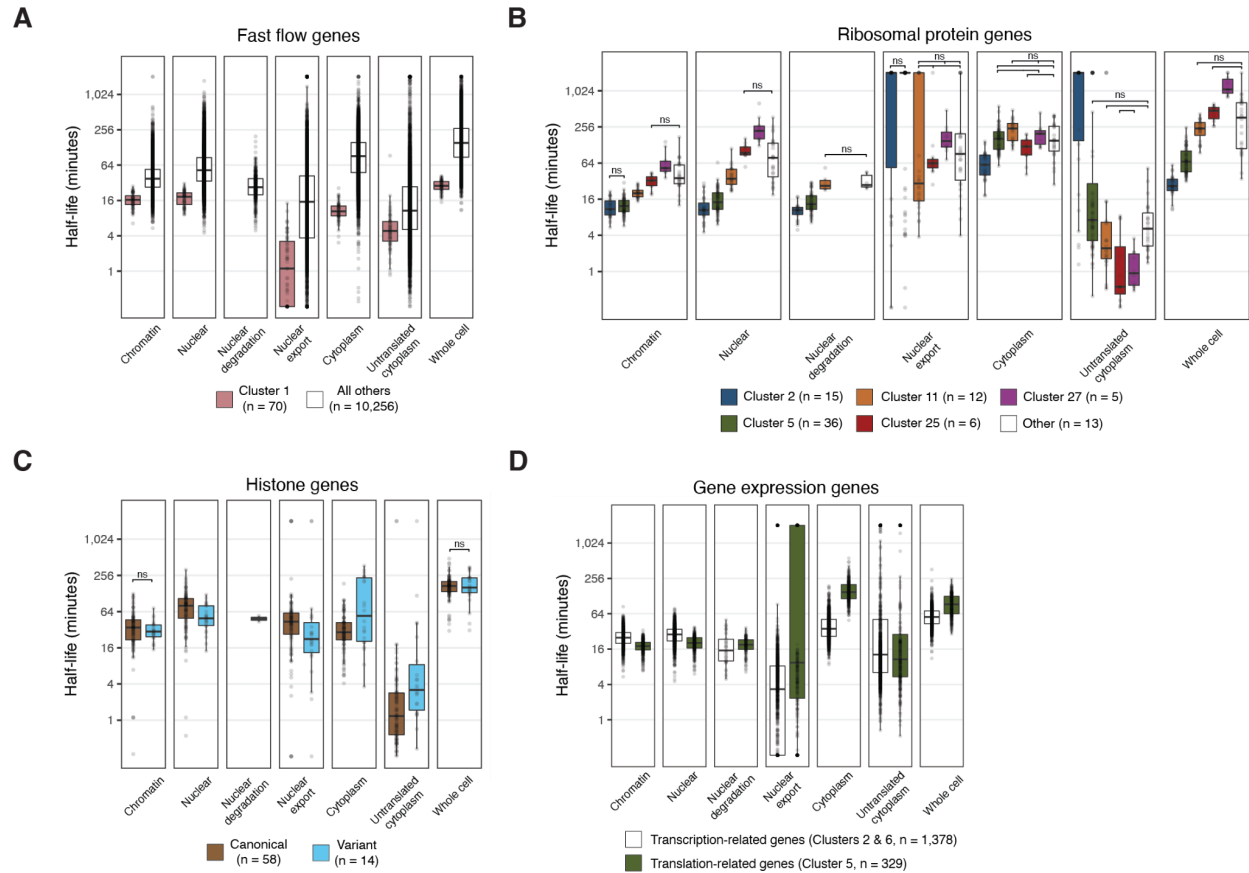
Figure 2.5: Genes with similar rates of RNA flow across the cell are functionally related.
A. Fast flow genes, i.e. those in cluster 1 in (Fig. 2.4B), were enriched for functions related to intracellular signaling and response to stimuli. All comparisons were statistically significant (p<0.0001, Wilcoxon test). **B.** Half-lives of ribosomal protein genes (RPGs) from each cluster where they were enriched (clusters 2, 5, 11, 25, and 27 in Fig. 2.4B). The number of RPGs within each cluster is noted. Non-significant comparisons are indicated as "ns" and all other comparisons were statistically significant (p<0.05, Wilcoxon test). **C.** Half-lives of histone genes. Canonical, replication-dependent histone genes were enriched in cluster 4, while non-canonical, replication-independent histone genes (including histone variants) were enriched in cluster 8 (Fig. 2.4B). The number of histone genes of each type is noted, and significance was noted as in (B). **D.** Half-lives of clusters 2 and 6 in Fig. 2.4B, containing genes related to transcription, and cluster 5, containing genes related to translation. The number of genes in each group is noted, and significance was noted as in (B).

(Figure 2.4B, Table S2.3). Cluster 8, which contains the most genes (n=1,662), represents

"canonical" RNA flow: median whole-cell half-lives of 150 min., long cytoplasmic and relatively

short chromatin and nuclear residence, fast polysome loading, and no evidence for nuclear

degradation (Figure 2B). However, the vast majority of genes (n=8,662) were divided into

smaller groups that deviated from these canonical kinetics (Figure 2.4B).

Genes involved in signal transduction and response to stimuli (e.g., *MYC, JUN,* and

CXCL2) were overrepresented in a cluster of 70 genes with "fast flow" kinetics across all

compartments (median whole-cell half-life 29 min.) (Figure 2.5A); however, even these genes

spent more than half of their life cycle on chromatin (median half-life 17 min.). We corroborated

this result independent of our clustering analysis by performing gene set enrichment analysis

(GSEA) (Liberzon et al., 2015; Subramanian et al., 2005) of all genes ranked by each

subcellular half-life and found that many hallmark gene sets involved in signaling pathways had

significantly faster RNA flow across all compartments (e.g., TNF-alpha signaling, Figure S2.7A,

Table S2.4). Yet, some signaling genes, such as those involved in IL6-dependent STAT5

signaling, only had short half-lives in the cytoplasm with fast rates of polysome loading but were

indistinguishable from other genes in other compartments (Figure S2.7B). Thus, mRNAs

encoding genes with signaling and sensing functions tended to have short half-lives across all

or most compartments.

Ribosomal protein genes (RPGs) were enriched in only five clusters (Figure 2.4B, Table

S2.2), rather than being uniformly distributed throughout all clusters (χ^2 test p<10^{-16}). Although all

five of these clusters shared long cytoplasm half-lives, consistent with previous reports (Eisen et

al., 2020; Herzog et al., 2017; Munchel et al., 2011), other RNA flow rates differed (Figure 2.5B).

A majority of RPGs were PUNDs with short half-lives on chromatin, with either slow (cluster 2)

or average (cluster 5) polysome loading kinetics. On the other hand, the remaining RPGs

exhibited slower and more canonical RNA flow without nuclear degradation and fast polysome

loading kinetics (clusters 11, 25, and 27). Thus, although not all RPGs exhibited the same rates

of RNA flow, we nonetheless observed two distinct patterns of RNA flow that they tended to

follow.

Histone genes were primarily enriched in two separate clusters (Figure 2.4B, clusters 4

and 8). Closer inspection revealed that the first group contained mostly canonical,

replication-dependent histone genes, and the second contained variant histone genes (Figure

2.5C, Table S2.2). Whole-cell half-lives did not differ between these two groups; however, canonical histones had ~2-fold longer nuclear (median 79 min.) than cytoplasm (median 28 min.) half-lives (Figure 2.5C). Additionally, canonical histones were loaded onto polysomes very quickly, often within just a few minutes, consistent with a previous study (Schott et al., 2021). Thus, canonical histones experience relatively unique RNA flow, whereas variant histones behave more like the average protein-coding transcript.

Genes involved in gene expression clustered into several distinct groups (Figure 2.4B). Most of the enriched GO terms in clusters 2 and 6 are related to transcription, including terms pertaining to transcription factors, RNA metabolism, and RNA polymerase II activity (Table S2.3). These genes included several mediator subunits, splicing factors, chromatin remodelers, and transcription factors (Table S2.1). By contrast, genes in cluster 5 are enriched for functions related to cytoplasmic translation, peptide biosynthesis, and ribosomal subunit biogenesis (Table S2.3), and include many translation factors (Table S2.1). Overall, genes involved in transcription had shorter whole-cell half-lives than genes involved in translation, consistent with multiple reports (Herzog et al., 2017; Schofield et al., 2018; Schwanhäusser et al., 2011; Yang et al., 2003) (Figure 2.5D). Surprisingly, despite their shorter whole-cell half-lives, genes involved in transcription had longer chromatin and nuclear half-lives and shorter cytoplasm half-lives compared to genes involved in translation (Figure 2.5D). Thus, genes involved in transcription and translation are enriched in clusters representing opposite RNA flow patterns.

Finally, GSEA revealed several patterns of RNA flow that were not easily detected by clustering and GO enrichment. Genes activated during the unfolded protein response (UPR) had whole-cell half-lives indistinguishable from all other genes (Figure S2.7C). However, UPR genes exhibited quick release from chromatin, yet "normal" nuclear half-lives, resulting in slow nuclear export. These transcripts then had "normal" cytoplasm half-lives but short untranslated cytoplasm half-lives (Figure S2.7C). Separately, genes encoding proteins involved in oxidative phosphorylation (not including mitochondrial-encoded genes) and genes targeted by the MYC

transcription factor exhibited similar trends. Both of these groups of genes had short chromatin half-lives but "normal" nuclear half-lives, resulting in long nuclear export half-lives, similar to UPR genes (Figure S2.7D-E). However, these groups then had long cytoplasm half-lives, resulting in long whole-cell half-lives, but with quick polysome loading rates (Figure S2.7D-E). Lastly, genes defining the epithelial-mesenchymal transition underwent RNA flow that was largely opposite from those described already, despite having whole-cell half-lives indistinguishable from all other genes. These transcripts had long chromatin and nuclear half-lives, short cytoplasm half-lives, but were slowly loaded onto polysomes (Figure S2.7F). In all, we uncover multiple lines of evidence that genes with transcripts encoding functionally related proteins flow across the cell at similar rates.

2.3 DISCUSSION

Here we have shown that our analysis of RNA flow, based on subcellular TimeLapse-seq and kinetic modeling, is capable of characterizing the many possible life cycles of mammalian transcripts across the cell and yields subcellular RNA half-lives genome-wide. We observed that the variability in RNA turnover previously observed at whole-cell resolution persisted across all subcellular compartments (Figure 2.2E). Transcripts with the same whole-cell half-lives flowed throughout the cell at different rates (Figure 2.3F-G), highlighting the variability in regulation between different genes. Yet, transcripts with similar rates of RNA flow across the cell tended to encode genes with similar functions. In both cell types analyzed, we identified many mRNAs that spent equal or more time on chromatin than in the cytoplasm, even among transcripts with the fastest whole-cell turnover (Figure 2.5A). These findings underscore the importance of examining RNA turnover at subcellular resolution and highlight the many spatiotemporal life cycles of mammalian transcripts.

Subcellular TimeLapse-seq relies on the biochemical purification of subcellular compartments, an operation with inherent limitations. Endoplasmic reticulum (ER) can partially co-sediment with the nuclear fraction in similar cell fractionation methods (Bhatt et al., 2012; Yeom and Damianov, 2017), and if this is occurring in our system, it may explain several minor observations. Transcripts enriched for functions related to the cell membrane, smooth ER, and COPI-coated vesicles, which are translated at the ER membrane (Fazal et al., 2019; Jan et al., 2014), were enriched in clusters with very long nuclear half-lives (clusters 17-20, n=703 genes, Figure 2.4B). We hypothesize that the nuclear fraction contains a mix of both newly synthesized and preexisting RNAs for these select genes, resulting in artificially long half-lives. Additionally, we are uncertain where various granules and condensates sediment, given their wide range of biochemical properties (Ditlev et al., 2018). Notably, however, P-bodies (monitored by marker protein LSM14A) sedimented in the cytoplasm (Figure 2.1B) and did not colocalize with

polysomes (Figure S2.1D). In sum, we believe that our biochemical purification accurately assigns the subcellular localizations of most transcripts.

We identified hundreds of genes that are predicted to undergo substantial nuclear degradation in both cell types analyzed where most of the transcripts encoding these genes are never exported from the nucleus. Genes with lower fractions of nuclear degraded transcripts would not be distinguished reliably through our conservative Bayes Factor analysis, so it is likely that more genes exhibit a significant, albeit lesser, degree of nuclear RNA degradation. Indeed, reduction of the Bayes factor criterion from a value of 100 to 10 identifies an additional 98 PUNDs in NIH-3T3 cells and 50 PUNDs in K562 cells. Taking into account the high production rates of many PUNDs, we estimate that at least 20% of all human and mouse protein-coding transcripts are degraded in the nucleus. Given that many PUNDs encode for splicing and nuclear RNA degradation factors, we speculate that nuclear degradation plays a role in nuclear RNA homeostasis (Berry and Pelkmans, 2022; Berry et al., 2022). Overall, we believe nuclear transcript degradation to be pervasive and likely to serve a regulatory function, which is an exciting direction for future work.

CHAPTER 3 –

FEATURES ASSOCIATED WITH RNA FLOW

This chapter is adapted from a manuscript currently under review entitled "Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle" and will be published at a later point in a peer-reviewed journal. I wrote the manuscript with input from Stirling Churchman and Robert Ietswaart. Karine Choquet sequenced several of the direct RNA nanopore sequencing libraries and analyzed all direct RNA nanopore sequencing libraries to generate poly(A) tail length measurements per read. I collected the features for the LASSO model and Robert Ietswaart developed and optimized the LASSO model. I designed, performed, and analyzed all other experiments with input from Stirling Churchman.

3.1 INTRODUCTION

A plethora of factors influence how long mRNAs reside within different subcellular compartments. On chromatin, mRNAs are synthesized by RNA polymerase II and undergo extensive processing, including splicing and polyadenylation. The time required to excise introns from pre-mRNAs varies significantly, ranging from seconds to tens of minutes (Drexler et al., 2020; Martin et al., 2013; Pai et al., 2017; Rabani et al., 2014; Reimer et al., 2021; Wachutka et al., 2019; Wan et al., 2021). In some cases, splicing regulates the nuclear dynamics of nascent RNA (Mauger et al., 2016; Ninomiya et al., 2011; Pandya-Jones et al., 2013; Yeom et al., 2021). Indeed, some mRNAs remain associated with chromatin well after transcription has completed (Bahar Halpern et al., 2015; Brody et al., 2011). In the nucleus, transcripts are subjected to either degradation or export. Nuclear degradation targets improperly processed mRNAs (Bresson et al., 2015; Davidson et al., 2012; Meola et al., 2016; Pendleton et al., 2018) but also serves additional regulatory roles for specific transcripts (Gudipati et al., 2012). By contrast, some loci are tethered near the nuclear pore complex to promote the rapid export of transcripts (Blobel, 1985; Rohner et al., 2013; Scholz et al., 2019).

mRNAs that are not degraded in the nucleus are exported into the cytoplasm. Here, they encounter translation initiation factors that ultimately recruit the ribosomal subunits to begin translation. Ribosomes are loaded onto transcripts with different kinetics. For example, mRNAs with long and structured 5' UTRs require RNA helicases to resolve these structures prior to

translation initiation (Lai et al., 2008; Leppek et al., 2018; Parsyan et al., 2009; Pisareva et al., 2008; Soto-Rifo et al., 2012). Furthermore, many mRNAs are not equally distributed throughout the cytoplasm (Benoit Bouvrette et al., 2018), and some require localization within the cytoplasm before translation initiation takes place (Das et al., 2021). Additionally, some mRNAs are sequestered into nuclear granules prior to nuclear export (Prasanth et al., 2005), as well as cytoplasmic granules, either before or after translation. Finally, poly(A) tails are deadenylated and mRNAs undergo degradation, a process that can be accelerated by the activity of microRNAs binding to transcripts (Bartel, 2018; Eisen et al., 2020a, 2020b; Passmore and Coller, 2022). Each of these processes vary in duration across genes and impact the subcellular fates of transcripts, either directly or indirectly through feedback loops.

Here, we analyze the gene-to-gene variability in RNA flow rates reported in mouse NIH-3T3 and human K562 cells (Chapter 2) to classify many features that are associated with subcellular half-lives. The targets of many RNA binding proteins (RBPs) exhibit different RNA flow rates compared to other genes. We validated that the differences in RNA flow observed for the targets of DDX3X and PABPC4 are sensitive to the expression of these RBPs. Measurement of poly(A) tails with subcellular resolution revealed that tail lengths reflect subcellular RNA half-lives. Finally, we identified the strongest genetic and molecular features that are predicted to determine RNA flow through machine learning, including transcription factors and sequence elements. Collectively, our findings uncover many potential regulators of rates of RNA flow across the cell.

3.2 RESULTS

3.2.1 RNA flow rates are associated with RNA binding proteins

mRNAs exist as part of ribonucleoprotein (RNP) complexes that consist of many RNA binding proteins (RBPs). These proteins are thought to protect mRNAs from degradation and regulate their progression throughout the cell. We hypothesized that RBPs play a role in regulating rates of RNA flow across subcellular compartments. To determine which RBPs may regulate the RNA flow of their targets, we analyzed ENCODE eCLIP datasets for 120 RBPs in K562 cells and classified the mRNAs targets of each RBP (Figure 3.1A) (Van Nostrand et al., 2020). We then identified all RBPs whose target mRNAs had significantly fast or slow subcellular half-lives relative to non-target mRNAs across any compartment (Figure 3.1B, Figure S3.1A-E, Table S3.1). We identified 43 RBPs whose targets had significantly fast or slow whole-cell turnover. Concurrently, we identified significant differences in the subcellular half-lives for the chromatin, nuclear, or cytoplasm fractions for nearly all (38/43) of these factors (Figure 3.1B), allowing us to pinpoint exactly where within the cell each RBP is associated with the stability of their target RNAs. Additionally, we identified significant differences in subcellular half-lives for the targets of an additional 37 RBPs which did not show differences at the whole-cell level. A majority (25/37) of these only exhibited significant differences in chromatin and/or nuclear half-lives (Figure 3.1B), highlighting the advantage of examining RNA flow at subcellular resolution.

Our analysis confirmed previous findings regarding several RBPs. We detected long whole-cell half-lives for IGF2BP1 and IGF2BP2 (Figure 3.1B, Figure S3.1E), which stabilize target mRNAs (Huang et al., 2018). The targets of these RBPs, in turn, had long nuclear half-lives (Figure 3.1B, Figure S3.1B), consistent with their ability to bind to mRNAs cotranscriptionally (Oleynikov and Singer, 2003; Pan et al., 2007). In addition, splicing factors presumably reside longer on pre-mRNAs of transcripts that experience slower splicing, leading

A

B
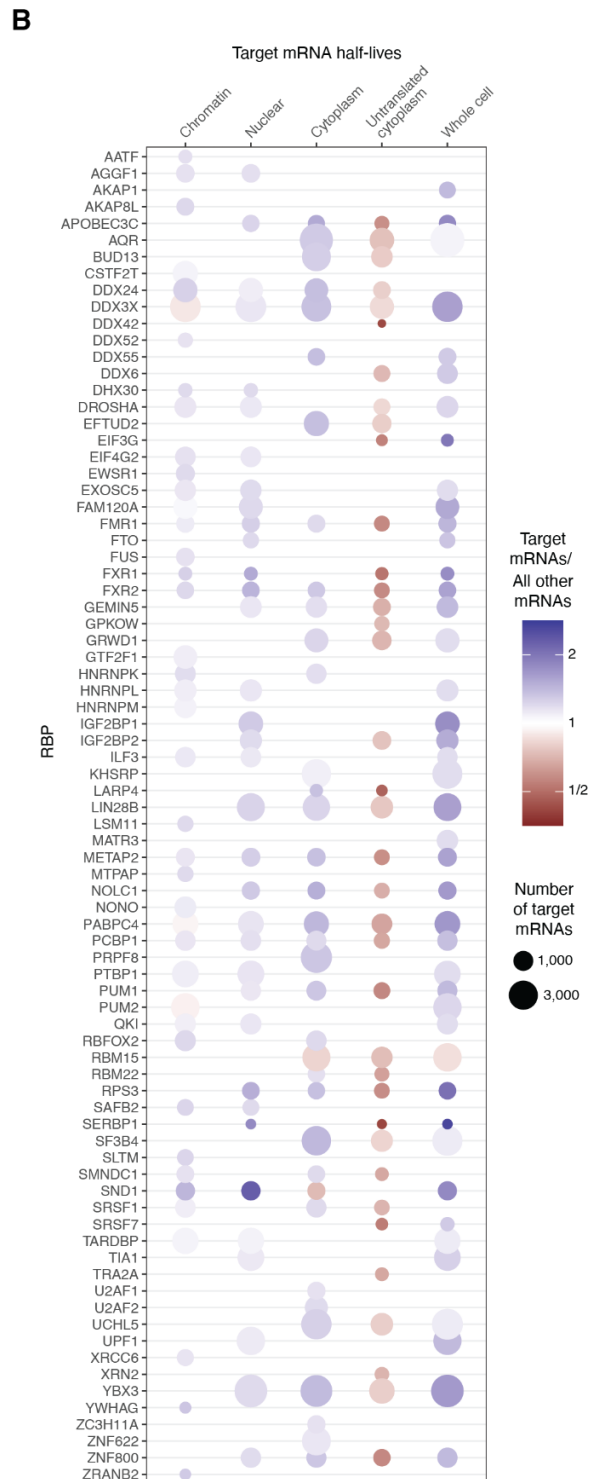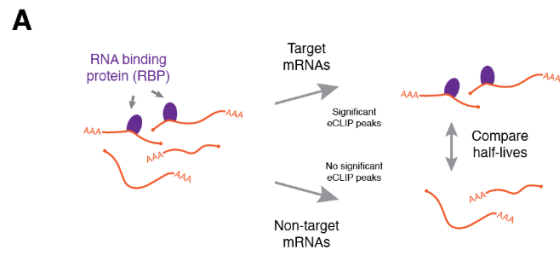
Figure 3.1: The targets of many RNA binding proteins (RBPs) exhibit distinctive RNA flow rates across the cell.
A. Schematic for RBP analysis. The mRNA binding targets of 120 RBPs in K562 were determined by identifying genes with significant eCLIP peaks published in (Van Nostrand et al., 2020). **B.** All RBPs with targets that exhibited significantly fast or slow half-lives for target RNAs compared to non-target RNAs in both biological replicates (adjusted p<0.01, Wilcoxon test, Bonferroni multiple testing correction) across any RNA flow rate. The size of the dot indicates the number of target mRNAs with measured half-lives within each compartment and the color reflects the difference (red, faster; blue, slower) of the median target over non-target half-lives.

to their enrichment in eCLIP peaks for these RBPs. Consistent with this, we identified several splicing factors whose targets had long chromatin half-lives (e.g., SRSF1, RBFOX2, QKI, PTBP1, FUS, and EWSR1; Figure 3.1B, Figure S3.1A).

3.2.2 DDX3X and PABPC4 regulate nuclear export of target mRNAs

We next investigated the roles of RBPs in dictating RNA flow. We began by focusing on DDX3X, an RNA helicase with many roles in RNA metabolism (Kanai et al., 2004; Lai et al., 2008; Samir et al., 2019; Shih et al., 2012; Soto-Rifo et al., 2012; Yedavalli et al., 2004). DDX3X targets

exhibited short chromatin half-lives and long nuclear half-lives, indicating slow export from the

nucleus (Figure 3.2A, Figure S3.1A-B). To analyze the role of DDX3X in determining the flow

rates of its target mRNAs, we used shRNAs to deplete DDX3X (Figure 3.2B) and performed

subcellular TimeLapse-seq. Upon depletion, nuclear RNA turnover of DDX3X target mRNAs

was no longer slower than non-targets (Figure 3.2C). Furthermore, DDX3X targets had long

cytoplasm and whole-cell half-lives in wild-type cells (Figure 3.1B, Figures S3.1C,E), consistent
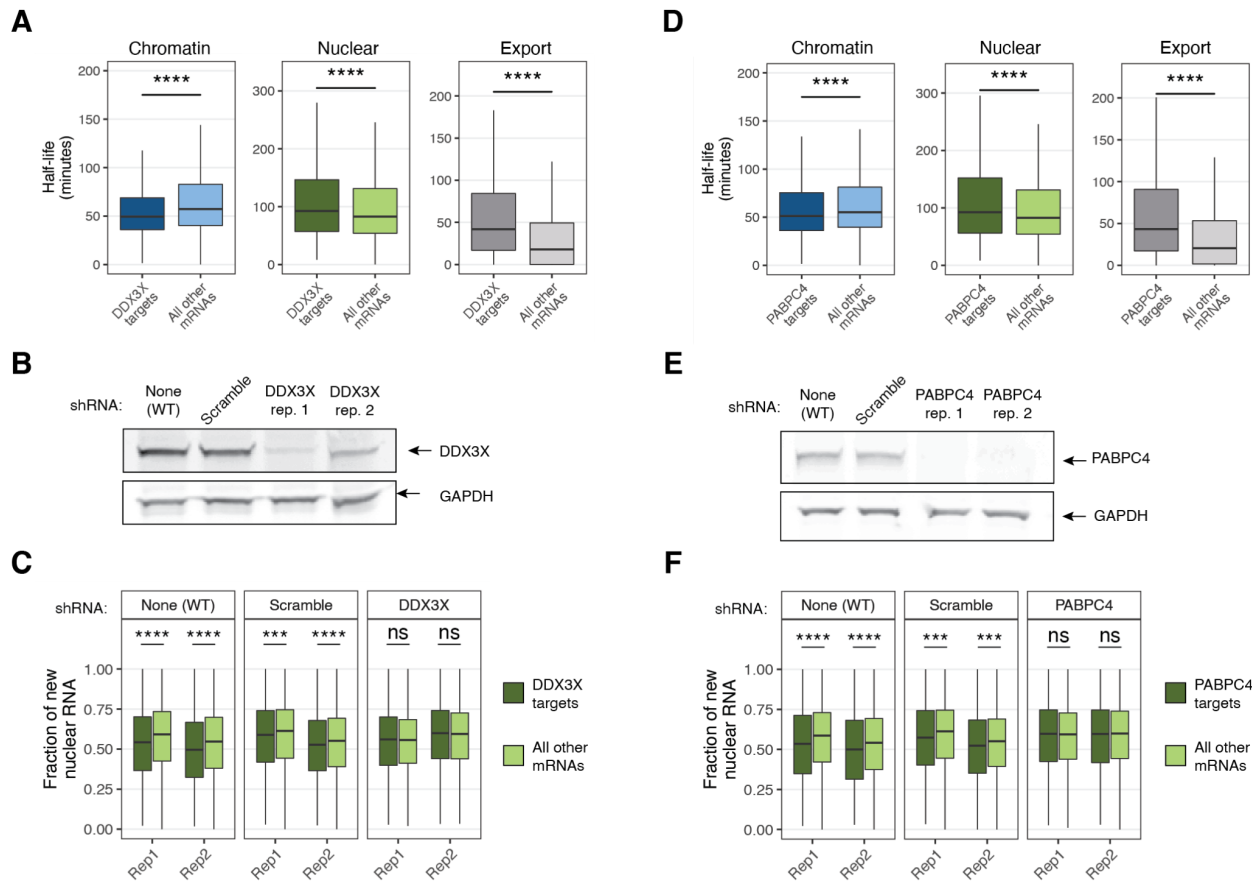
Figure 3.2: DDX3X and PABPC4 depletion affect nuclear half-lives of target mRNAs.
A. Half-lives of DDX3X mRNA targets and non-targets in K562. The chromatin, nuclear, and nuclear export half-lives of targets are compared to non-target mRNAs (****: $p<0.0001$, ***: $p<0.001$, "ns:" not significant, Wilcoxon test). **B.** Confirmation of DDX3X protein knockdown in K562. Cells were transduced with lentivirus containing plasmids expressing a scrambled shRNA sequence or one targeting DDX3X and knockdown efficiency was monitored by western blotting in samples collected for subcellular TimeLapse-seq. Wild-type (non-transduced) cells were included as a control. **C.** Fraction of new nuclear RNA measured by subcellular TimeLapse-seq of DDX3X target mRNAs compared to all other mRNAs in wild-type cells, cells expressing a DDX3X-targeting shRNA, and cells expressing a scrambled shRNA. Significance was noted as in (A). Two biological replicates ("rep") are shown. **D.** Same as (A) for PABPC4. **E.** Same as (B) for PABPC4. **F.** Same as (C) for PABPC4.

with the association of DDX3X with several cytoplasmic RNA granules (Kanai et al., 2004; Samir et al., 2019; Shih et al., 2012). However, RNA turnover on chromatin, in the cytoplasm, and at the whole-cell level were not affected in the DDX3X knockdown (Figure S3.2A). We conclude that DDX3X regulates RNA flow at the step of nuclear export.

We observed similar RNA flow rates for targets of cytoplasmic poly(A) binding protein PABPC4, which had long cytoplasm and whole-cell half-lives (Figure 3.1B, Figures S3.1C,E), consistent with the function of this RBP in stabilizing transcripts with short poly(A) tails containing AU-rich motifs (Kini et al., 2014). However, PABPC4 targets had shorter chromatin half-lives and longer nuclear half-lives than other transcripts (Figure 3.2D). As before, we used shRNAs to deplete PABPC4 in K562 cells (Figure 3.2E) and performed subcellular TimeLapse-seq. Surprisingly, we detected no differences in the chromatin, cytoplasm, or whole-cell half-lives of target mRNAs upon PABPC4 depletion (Figure S3.2B). However, the target mRNAs no longer had longer nuclear half-lives (Figure 3.2F), implicating PABPC4 in nuclear export. Based on these findings, we conclude that although both PABPC4 and DDX3X bind to their mRNA targets throughout the cell, their depletion only affects RNA flow at the step of nuclear export. These observations highlight the ability of subcellular TimeLapse-seq to study the role of RBPs within different subcellular compartments and demonstrate that RBPs regulate RNA flow.

3.2.3 mRNA poly(A) tail lengths are dynamic across and within subcellular compartments

Poly(A) tails are connected to mRNA metabolism and tail lengths have long been associated with RNA half-life and translational efficiency (Passmore and Coller, 2022). However, how much poly(A) tail lengths vary across subcellular compartments for each gene remains largely unknown. We next measured the lengths of poly(A) tails of RNAs residing within
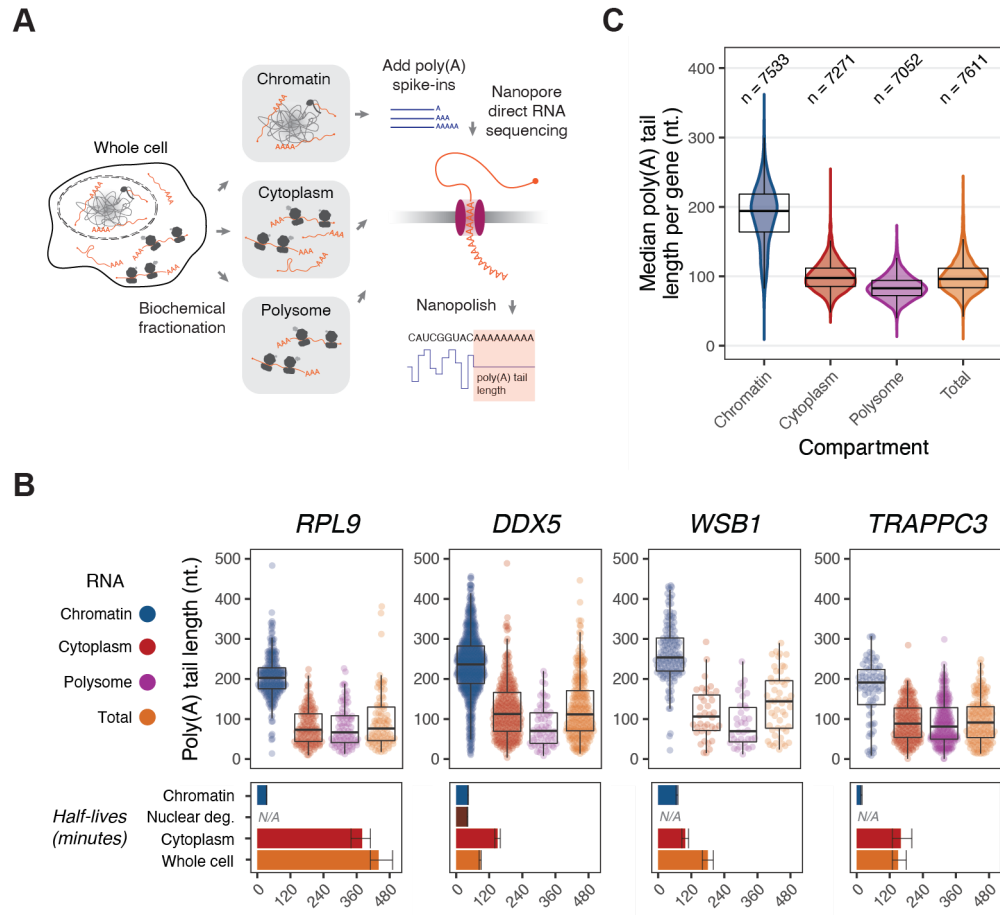
Figure 3.3: Poly(A) tail lengths across subcellular compartments.
A. Schematic for poly(A) tail length analysis. Chromatin, cytoplasm, polysome, and total RNA from K562 cells were directly sequenced by nanopores. The poly(A) tail length on each RNA was estimated using nanopolish-polya (Workman et al., 2019), and synthetic RNA spike-ins were used to normalize poly(A) tail length across sequencing runs. **B.** Poly(A) tail lengths per RNA across compartments for example genes *(RPL9, DDX5, WSB1, TRAPPC3)*. Each dot represents an individual RNA. The mean chromatin, nuclear degradation, cytoplasm, and whole-cell half-lives for one replicate are indicated below each gene, with the error bars representing credible intervals. **C.** Distribution of median poly(A) tail lengths for each gene covered by >=10 reads in each sample. The number of genes analyzed in each compartment is noted.

subcellular compartments to investigate the relationship between poly(A) tails and RNA

half-lives with subcellular resolution. We measured the tail length of each mRNA in chromatin,

cytoplasm, and polysome fractions, as well as in total cellular RNA, using nanopore direct RNA

sequencing (Figure 3.3A) (Workman et al., 2019). We confirmed that the 3'-end of >80% of

sequenced RNAs mapped to annotated poly(A) sites (Figure S3.3A). To control for technical

variations between nanopore sequencing runs, we included a set of 6 spike-in RNAs. Each of

the 6 synthetic RNAs had a different poly(A) tail length, ranging from 10-100 adenines. Using these spike-in RNAs, we calculated a poly(A) tail length size factor for each sample and normalized poly(A) tail lengths from each endogenous RNA by this size factor (Figure S3.3B-C). Within each fraction, the median tail length per gene correlated between the two biological replicates (Figure S3.3D). We see the highest correlation between replicates for chromatin poly(A) tail length, likely reflecting the fact that tails are initially added to roughly the same length to all nascent mRNAs, compared to tails being more variable in other compartments.

We observed that poly(A) tail lengths gradually shortened as RNAs flowed through the cell, as expected (Figure S3.3C). These trends could be observed at the single gene level, albeit with variability across genes (Figure 3.3B, Table S3.2). For example, we observe a median chromatin poly(A) tail lengths per gene of less than 200nt (e.g., *TRAPPC3,* median 192nt) to greater than 250nt (e.g., *WSB1,* median 254nt) (Figure 3.3B). Globally, the longest poly(A) tails were on chromatin with most genes having a median tail between 160 and 215nt (Figure 3.3C). We measured shorter polyA tail lengths in the cytoplasm, where most genes had median tail lengths of 85–110nt (Figure 3.3C). As expected, the cytoplasmic tail lengths closely resembled those in total RNA (Figure 3.3C), consistent with most polyadenylated mRNAs being cytoplasmic in K562 cells (Figure 2.2E). In general, the distribution of tail lengths in polysome RNA was slightly shorter than in cytoplasmic RNA (Figure 3.3C). Although we successfully detected artificial tails of 10nt and 20nt present on spike-in RNAs in all libraries (Figure S3.3B), we rarely detected polyA tails shorter than 25-30nt in any compartment (Figure S3.3C), consistent with poly(A) tails shorter than the footprint of a single poly(A) binding protein complex being unbound and thus unstable (Baer and Kornberg, 1980).

3.2.4 Poly(A) tail length increases with chromatin residence and decreases with cytoplasm residence

To determine the relationship between poly(A) tail lengths and RNA flow across the cell, we compared the median tail length per gene in each compartment to the respective subcellular half-life. Median chromatin RNA tail lengths are the shortest for genes with chromatin half-lives of less than 15 minutes, and chromatin tail lengths gradually increase as half-lives increase, until approximately 40 minutes (Figure 3.4A). Assuming that these differences do not arise from poly(A) polymerase adding tails of different lengths to transcripts encoding different genes, these discrepancies in chromatin RNA tail lengths could reflect continued adenylation occurring over time after the initial poly(A) tail is added, or it could be explained by the deadenylation of transcripts destined for quick release from chromatin. To discern these two possibilities, we analyzed the poly(A) tail lengths of incompletely spliced chromatin reads, reflecting pre-RNAs that have recently completed transcription but are still undergoing RNA processing. We see that poly(A) tails are shortest for reads that are completely unspliced and that tail length increases as more introns are excised from the pre-mRNA (Figure 3.4B). Therefore, we conclude that poly(A) tails are extended with increased time spent on chromatin, which is also associated with continued splicing.

We observed that median cytoplasm RNA poly(A) tail length decreased with longer cytoplasm half-lives in both K562 cells and NIH-3T3 cells (Figure S3.4A-B), and a negative correlation between median tail length in total RNA and whole-cell residence (Figure 3.4C), consistent with previous reports (Lima et al., 2017; Subtelny et al., 2014). We observed the same trend on a per-molecule basis. We saw that incompletely spliced cytoplasm RNAs, likely representing a relatively young and unstable population of RNAs that have been recently exported from the nucleus and will soon undergo nonsense-mediated decay, have longer poly(A) tails than completely spliced reads (Figure S3.4C). mRNAs are gradually deadenylated
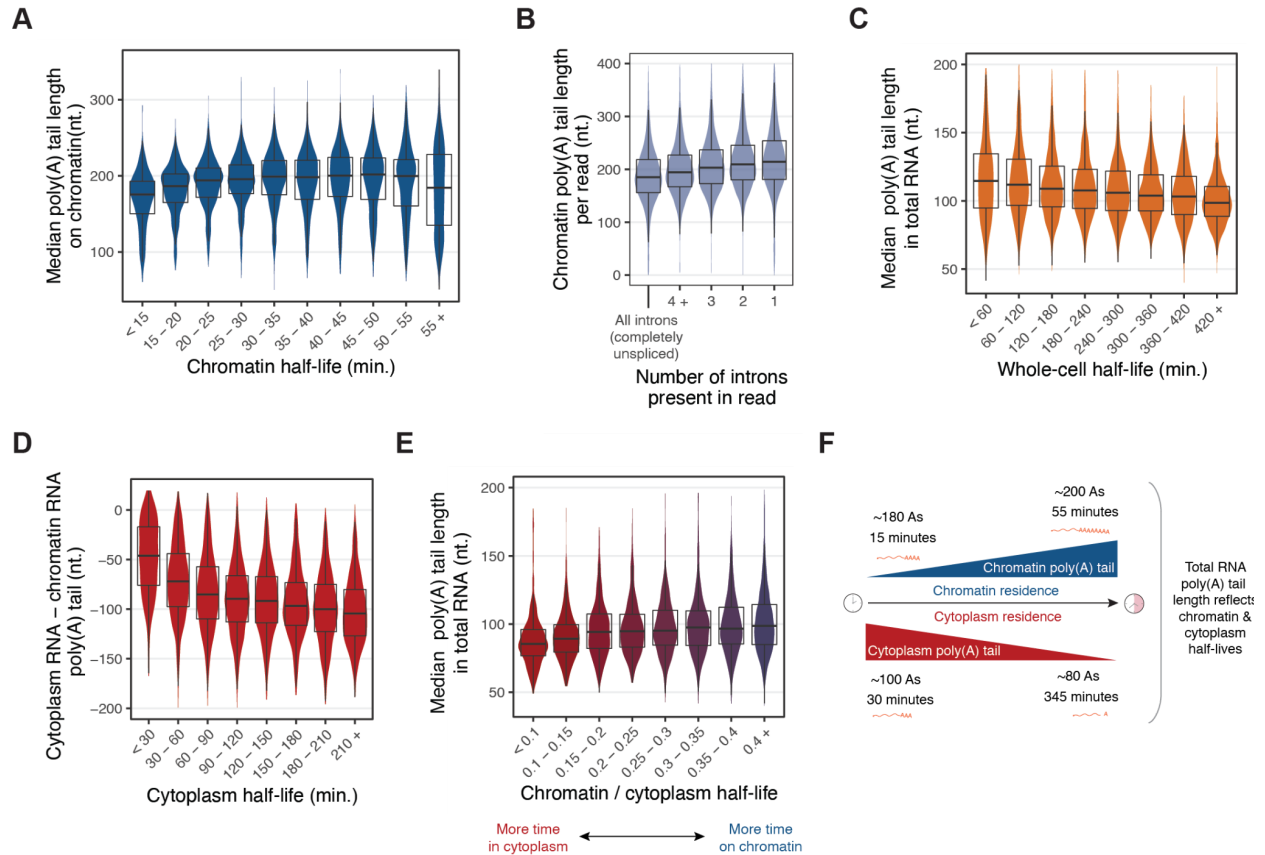
Figure 3.4: Subcellular compartment-specific poly(A) tail lengths reflect RNA flow rates.
A. Median poly(A) tail length for genes covered by >=10 reads in chromatin RNA libraries as a function of chromatin RNA half-life. **B.** Poly(A) tail lengths on chromatin as a function of splicing status. Poly(A) tail lengths were analyzed for all chromatin RNA reads with incomplete splicing (containing at least one unexcised intron), binned by the number of introns present. **C.** Median total RNA poly(A) tail length for genes covered by >=10 reads in total RNA libraries as a function of whole-cell half-life. **D.** The difference between the median chromatin poly(A) tail length and the median cytoplasm poly(A) tail length for all genes covered by >=10 reads in each library, as a function of their cytoplasm half-life. **E.** Median poly(A) tail length in total RNA as a function of the relative amount of time spent on chromatin and in the cytoplasm for each transcript, defined as the ratio of chromatin half-life to cytoplasm half-life. Genes covered by >=10 reads in total RNA libraries were included. **F.** Model of compartment-specific poly(A) tail lengths with respect to subcellular half-lives. Poly(A) tails of chromatin-associated RNAs lengthen with increased chromatin residence, while cytoplasmic poly(A) tails shorten with increased cytoplasmic residence.

in the cytoplasm, and it has been proposed that the tails of transcripts with longer half-lives are

"pruned" to a certain length, at which point they remain relatively stable (Lima et al., 2017). In

contrast, transcripts with faster turnover undergo more rapid deadenylation and degradation,

without such stabilization at shorter tail lengths (Eisen et al., 2020a). Our data are consistent

with these findings; the change in median tail lengths between the chromatin and cytoplasm

fractions of mRNAs with short cytoplasm half-lives (<30 minutes) differed by ~50nt, whereas

mRNAs with long cytoplasm half-lives (>345 minutes) differed by >100nt (Figure 3.4D).

Based on our findings that chromatin poly(A) tail length increases with chromatin

residence and cytoplasm poly(A) tail length decreases with cytoplasm residence, total RNA tail

length must be reflective of the relative time spent by transcripts in both compartments. mRNAs

that spent relatively more time on chromatin than in the cytoplasm had longer poly(A) tails in

total RNA, whereas mRNAs that spent relatively more time in the cytoplasm had shorter tails in

total RNA (Figure 3.4E). For example, *TRAPPC3* and *WSB1* had similar whole-cell half-lives,

but *TRAPPC3* spent ~7x more time in the cytoplasm than on chromatin, whereas *WSB1* mRNA

spent nearly the same amount of time on chromatin and in the cytoplasm (Figure 3.3B).

Consequently, *TRAPPC3* had a much shorter median poly(A) tail in total RNA (median 89nt)

than *WSB1* (median 142nt) (Figure 3.3B). Thus, median poly(A) tail lengths in total RNA are not

only reflective of whole-cell residence times (Figure 3.4C), but also of the relative times spent on

chromatin and in the cytoplasm (Figure 3.4E-F).

3.2.5 PUND transcripts are spliceosome targets that exhibit distinct RNA flow, splicing, and poly(A) tail phenotypes

Given the patterns of RBP binding and poly(A) tail lengths across compartments, we

wondered whether PUNDs behaved uniquely in any respect. To explore this possibility, we

began by comparing the half-lives in each compartment between PUNDs and all other

transcripts. In human K562 cells, PUND genes had faster turnover on chromatin, in the nucleus,

and at the whole-cell level than genes without evidence of nuclear degradation (Figure 3.5A).

Notably, PUND genes had longer cytoplasm half-lives than other mRNAs (Figure 3.5A),

indicating that the transcripts from PUNDs that do get exported are more stable in the cytoplasm

compared to transcripts of other genes. Thus, the high nuclear degradation of PUND transcripts
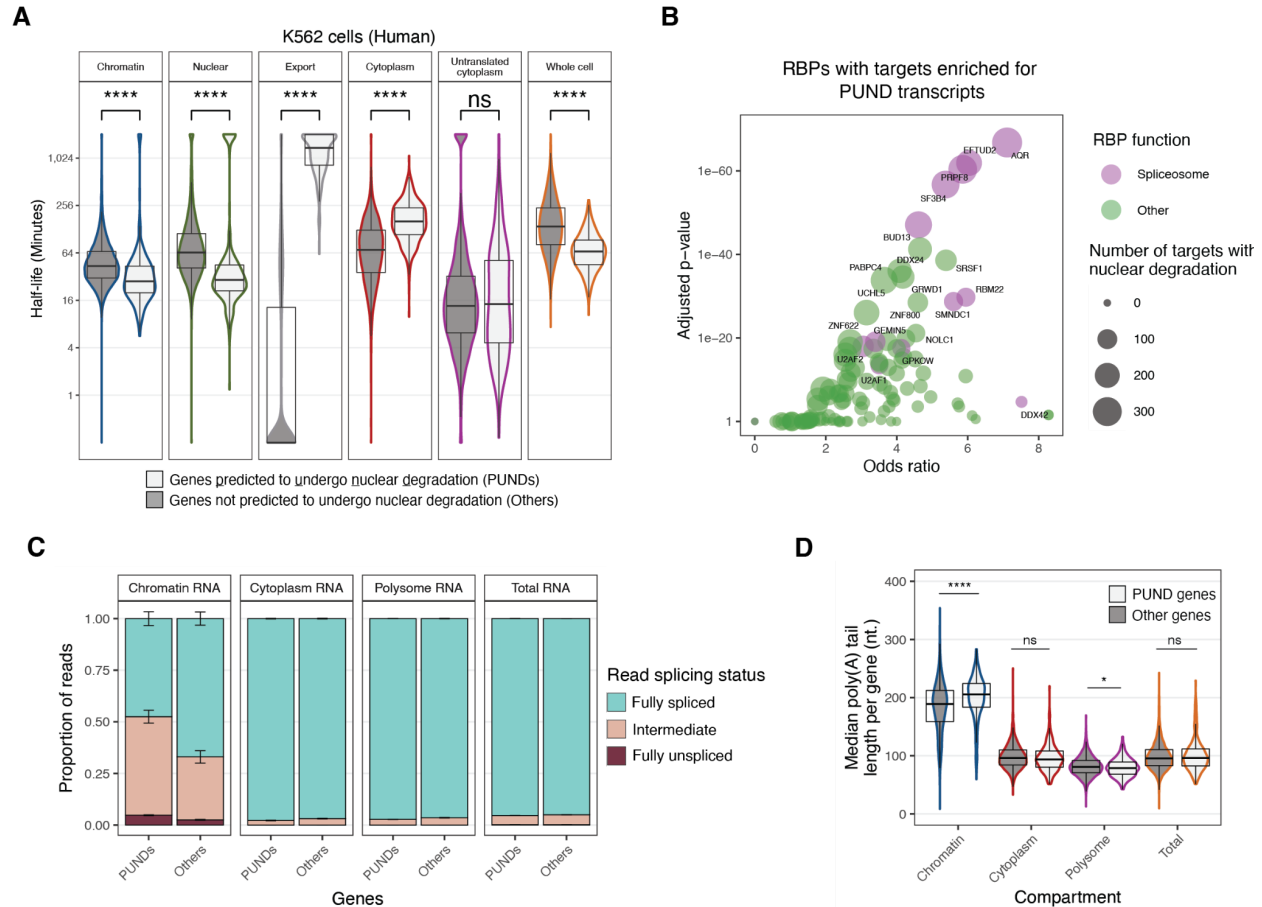
Figure 3.5: Genes predicted to undergo nuclear degradation (PUNDs) exhibit unique phenotypes related to RNA flow, splicing, and poly(A) tail lengths.
A. Half-lives of all PUND genes (n=408) compared to all other genes in human K562 cells
(****: p<0.0001, *: p<0.05, "ns:" not significant, Wilcoxon test). **B.** Volcano plot representing the odds ratio and adjusted p-value obtained from a Fisher's exact test comparing the target and non-target mRNAs for each RBP analyzed in Figure 3 (n=120) to determine which RBPs have enriched PUND targets. RBPs are colored by function as defined by (Van Nostrand et al., 2020), with "other" representing any function that is not "spliceosome." **C.** Splicing levels of PUND and other transcripts in nanopore direct RNA sequencing data across subcellular compartments and in total RNA. Error bars show standard error over two biological replicates. **D.** Median poly(A) tail length of PUND genes relative to others across subcellular compartments. The median poly(A) tail length was calculated for each gene covered by >= 10 reads in each sample. Tail lengths were compared between PUND and other genes, and significance was noted as in (A).

is not a trivial reflection of the breakdown of inherently unstable transcripts. We did not detect

any significant differences in the untranslated cytoplasm half-lives (Figure 3.5A). When

comparing PUNDs to other transcripts in NIH-3T3 cells, we observed the same trends across all

flow rates as K562, with the exception that PUND transcripts have a slight delay in polysome

loading (p=0.013) (Figure S3.5A).

We next wondered about the role of RNA binding proteins in possibly regulating nuclear degradation. For each RBP, we compared the number of PUND and other genes amongst the target and non-target mRNAs. Indeed, we found 68 RBPs to have targets that are significantly enriched for PUNDs (Figure 3.5B, Table S3.1). Remarkably, five of the proteins most significantly enriched were components of the spliceosome: AQR, EFTUD2, PRPF8, SF3B4, and BUD13 (Figure 3.5B). We also detected splicing factor SRSF1 among the most significant RBPs (Figure 3.5B). In light of this finding, we sought to determine whether PUND transcripts exhibited either slow or fast splicing compared to other mRNAs. Analyses of our direct RNA sequencing libraries revealed PUND transcripts had more incompletely spliced mRNAs on chromatin than other transcripts, but overall splicing levels did not differ in other compartments (Figure 3.5C). Finally, we investigated whether PUND transcripts exhibited any differences in poly(A) tail length relative to other transcripts, and found that they had longer tails on chromatin (median PUND tail length= 210nt, median other= 193nt, $p=1.40\times10^{-16}$) (Figure 3.5D). This trend was true across all chromatin half-lives (Figure S3.5B). This was surprising given that PUND transcripts generally reside on chromatin for less time (Figure 3.5A), and we observed that transcripts with short chromatin half-lives generally have short poly(A) tails (Figure 3.4A,F). PUNDs had slightly shorter tails on polysomes than other mRNAs (median PUND tail length=80nt, median other=83nt, p=0.013), but tail lengths did not differ in cytoplasm or total RNA. We conclude that PUND transcripts have more incomplete splicing and longer poly(A) tails on chromatin, but are largely indistinguishable from other transcripts in the cytoplasm, on polysomes, and in total RNA.

3.2.6 Machine learning model identifies molecular features that explain RNA flow rates

Finally, we sought to identify genetic and molecular features that collectively explain the variability in RNA flow rates. To this end, we developed a LASSO regression model (Figure

3.6A, Figure S3.6A) that identifies sparse relevant features through L^1 regularization (Hastie et al., 2001). The 10× cross-validation and unseen test set performances of our model varied between the subcellular compartments (Figure 3.6B), ranging from R^2 = 0.4 for cytoplasmic turnover to R^2 = 0.13 for polysome loading rates, likely due to the larger uncertainty for these estimates (Figure 2.2D, Figure S2.5A). Our model performed as well as the best published whole-cell models when we trained it on our whole-cell rates or on published "ensemble" values (Figure S3.6B) (Agarwal and Kelley, 2022; Blumberg et al., 2021; Chan et al., 2018; Chen and van Steensel, 2017; Cheng et al., 2017; Sharova et al., 2009; Spies et al., 2013; Yang et al., 2003).
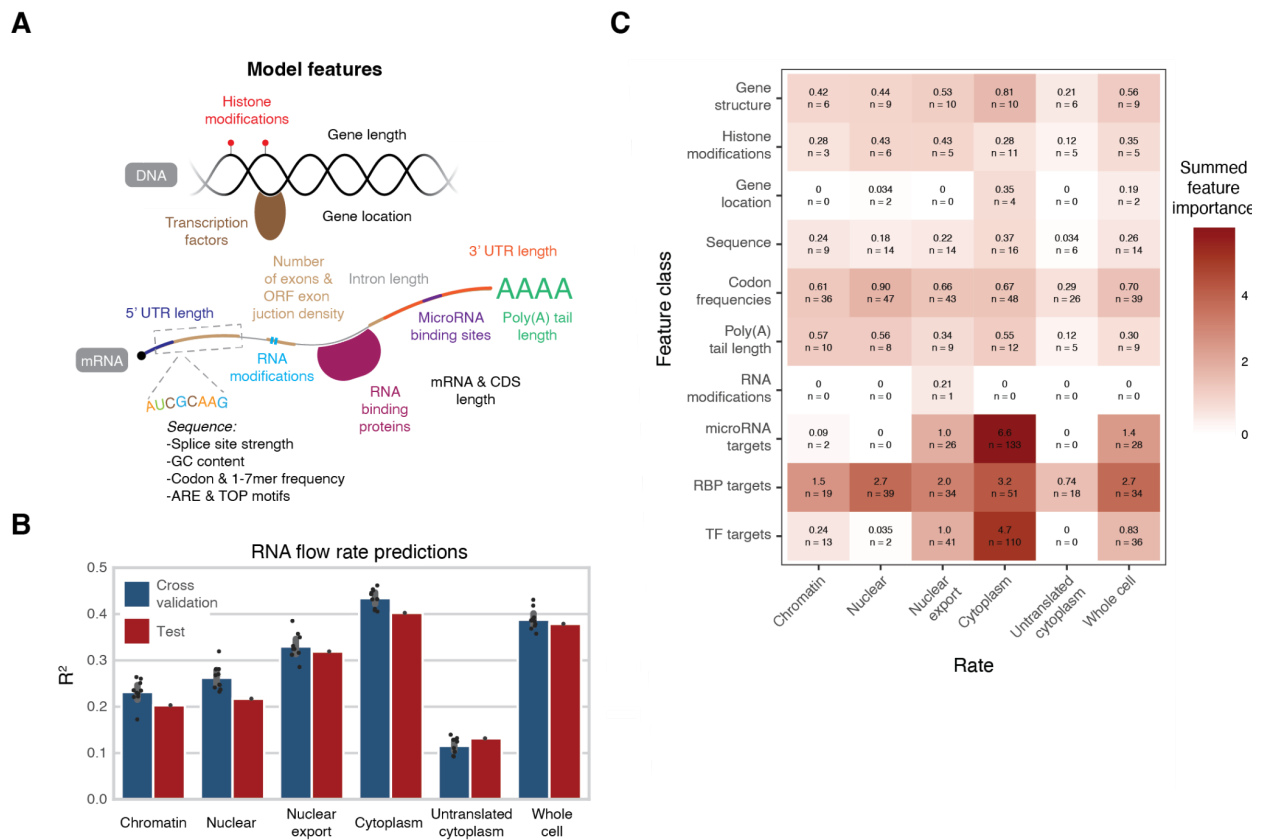
Figure 3.6: LASSO regression model identifies genetic and molecular features across multiple classes that predict RNA flow rates.
A. Schematic representing the genetic and molecular features for each gene included in LASSO model (full list: Table S3.3). **B.** 10x cross-validation and test set performances of LASSO models predicting subcellular rates. **C.** Class feature importances of LASSO models. The importance of each class represents the summed importance values across all included individual features.

Across all RNA flow rates, basic gene structure, histone modifications, sequence features, codon frequencies, RBP target sets, and compartment-specific poly(A) tail lengths were the feature classes that provided the most information (Figure 3.6C, Table S3.3). We also observed subcellular-specific relevant classes, such as microRNAs promoting cytoplasmic turnover and transcription factors targets associating with nuclear export and cytoplasmic turnover (Figure 3.6C). Sequence determinants and codon frequencies were collectively relevant as feature classes (Figure 3.6C), albeit the effect sizes of individual features were generally small and compartment-specific (Figure S3.7B, Table S3.3). For example, transcripts containing a 5' terminal oligo-pyrimidine (5' TOP) motif (Cockman et al., 2020) had faster RNA flow dynamics, including polysome loading rates, but were also more stable in the cytoplasm (Table S3.3).

Next, we investigated the strongest individual feature contributions across compartments (Figure 3.7A, Figure S3.7A). Both intron length and the number of exons were top predictors of chromatin half-lives (Figure 3.7A). Accordingly, chromatin and nuclear half-lives increased with gene length in both cell lines, likely due to transcription elongation time, albeit with large variability between genes (Figure 3.7B, Figure S3.7C, Spearman correlations for K562: chromatin=0.16, nuclear=0.19, nuclear; NIH-3T3: chromatin=0.20, nuclear=0.15). Interestingly, genes with 2 exons had shorter chromatin and nuclear half-lives in both cell lines than genes containing only 1 exon (i.e., genes that do not require splicing); chromatin and nuclear half-lives then gradually increased with the number of exons per gene (Figure S3.8A-D). Gene GC content, histone H3K36 tri-methylation, and SND1 binding all positively influenced chromatin, nuclear, and nuclear export rates (Figure 3.7A). Indeed, genes with high GC content have longer chromatin half-lives than genes with low GC content (Figure S3.9A,D). Yet, genes with high GC content are also significantly shorter than those with low GC content (Figure S3.9B-C). Thus, transcripts encoding high GC content genes would otherwise be predicted to reside on chromatin for less time, so simply comparing chromatin half-lives between these two groups
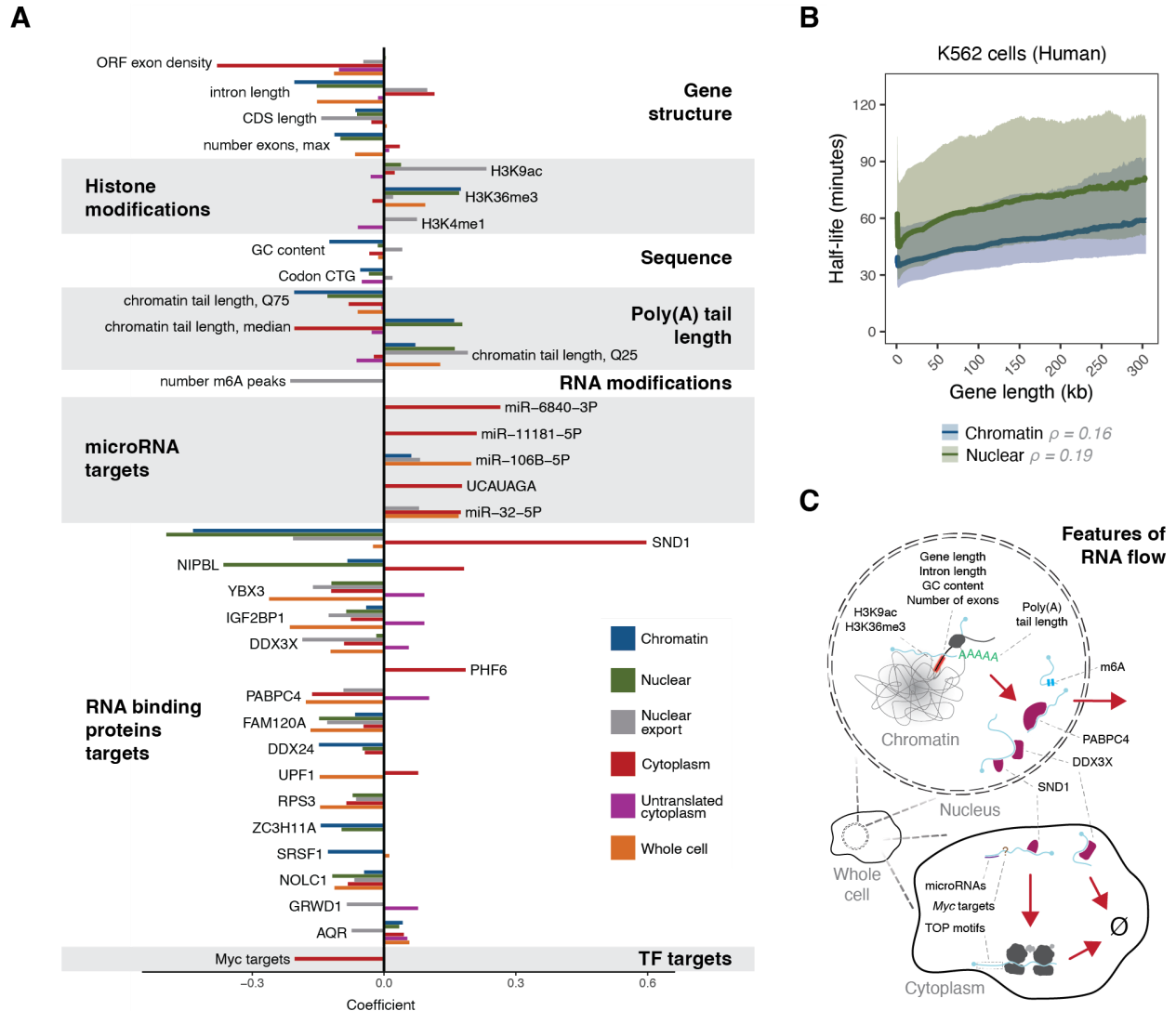
Figure 3.7: Individual features predicted to influence RNA flow rates.
A. Top individual features associated with RNA flow. The top 10 features with highest importance of each flow rate were identified and their correlation with any flow rate was shown. Individual features were grouped by feature family. **B.** Continuous averages of chromatin and nuclear half-lives as a function of gene length in human K562 cells. Solid lines represent median half-lives and shaded ribbons represent the third quartile (top) and first quartile (bottom) of half-lives. **C.** Schematic depicting relevant features related to RNA flow rates. Features are included in the compartments where they associate with RNA flow rates (top: chromatin, nuclear, and nuclear export half-lives; bottom: cytoplasm and untranslated cytoplasm half-lives).

underestimates the effect of GC content on chromatin turnover. Consistent with our RBP

analysis (Figure 3.2A-F), DDX3X and PABPC4 were predicted to regulate nuclear export (Figure

3.7A). Notably, the RNA modification N6-methyladenosine (m6A) was also strongly predictive of slower nuclear export rates, but not any other RNA flow rate (Figure 3.7A).

3.2.7 Transcription factor target genes produce transcripts that display similar RNA flow patterns

Our LASSO model predicted that genes targeted by the MYC transcription factor (TF) had long cytoplasm half-lives (Figure 3.7A), as previously shown by GSEA (Figure S2.7E). However, other TFs with fewer target genes, and thus a smaller global contribution, might have remained unidentified in our model. To more sensitively investigate whether more TFs were associated with the RNA flow rates of their target genes, we systematically identified TFs for which the half-lives of targets differed from non-targets (Figure S3.7D), similar to our RBP target analysis (Figure 3.1A-B). Remarkably, 193 TFs were associated with altered RNA flow rates (Figure S3.7D). For instance, targets of the stress-response TFs ATF4 and ATF5 had shorter chromatin half-lives (Figure S3.7D). Our LASSO model and TF analysis revealed that many uncharacterized Zinc-finger proteins were also associated with altered RNA flow rates (Figure S3.7D, Table S3.4). We conclude that many features, underlying a diverse set of regulatory mechanisms, are responsible for subcellular transcriptome dynamics (Figure 3.7C).

3.3 DISCUSSION

Here we analyze subcellular RNA half-lives and identify many genetic and molecule features that underlie the RNA flow rates of different transcripts. Some features that are known to contribute to RNA turnover at the whole-cell level, such as microRNA targeting, were shown to only affect flow rates within one subcellular compartment where the feature acts (i.e., the cytoplasm). On the other hand, other features were unexpectedly also found to only affect RNA stability within specific compartments, such as m6A methylation associating with nuclear export, and transcription factor targets being associated with cytoplasmic stability. Futhermore, additional features were unexpectedly predicted to affect flow rates across multiple compartments, such as codon frequency. Our model unveils many potential determinants of RNA flow, and identifies candidate features for future studies that can further investigate their role in dictating subcellular RNA stability.

We present significant associations between rates of RNA flow and many RBPs across multiple compartments, and further identify new and unexpected functions for DDX3X and PABPC4 in regulating nuclear export. DDX3X interacts with multiple nuclear export machineries (Lai et al., 2008; Yedavalli et al., 2004) and it can form puncta at nuclear pores (Merz et al., 2007). Given our finding that DDX3X increases the nuclear half-lives of its targets, we postulate that DDX3X may be involved with physically tethering target mRNAs at the nuclear envelope after they are released from chromatin. Furthermore, we report the first nuclear role for PABPC4, which also prolongs nuclear export of target mRNAs. Both PABPC4, as well as PABPC1, have been shown to shuttle between the nucleus and the cytoplasm (Afonina et al., 1998; Burgess et al., 2011). Furthermore, PABPC1 is known to bind pre-mRNAs in the nucleus (Hosoda et al., 2006), and our data suggests that PABPC4 may do the same. All of these data complement the finding that mRNAs are generally exported into the cytoplasm quickly following

release from chromatin (Figure 2.2E), but that this step can be regulated by DDX3X and PABPC4, and likely other RBPs.

Our comprehensive analysis of poly(A) tail dynamics across subcellular compartments shows that tail length distributions differ between genes not only in the cytoplasm, but also on chromatin. On a per-molecule basis, we used the splicing status across long reads as a proxy for relative age of the RNA and saw that tail length increases as transcripts spend more time on chromatin (Figure 3.4B) and decreases with more time in the cytoplasm (Figure S3.4C). This explains some of the variability between genes. Those with longer chromatin half-lives generally contain longer tails on chromatin (Figure 3.4A), while conversely, we see that mRNAs with longer cytoplasmic half-lives have shorter cytoplasmic tails at steady state (Figure S3.4A-B). These results shed new light on the relationship between poly(A) tail length and RNA stability. Although total RNA tail length inversely correlated with whole-cell half-lives in K562 (Figure 3.4C), this simple relationship does not capture the dynamics across all compartments as cytoplasm half-lives are generally longer than chromatin half-lives (Figure 2.2E), obscuring the contribution of chromatin RNA poly(A) tails. Indeed, we observed that total RNA tail lengths reflect the ratio of time spent on chromatin and in the cytoplasm for each gene (Figure 3.4E-F). Thus, we have furthered our understanding of the links between poly(A) tail lengths and mRNA stability to a subcellular resolution, and its full dissection will be enabled by the methodologies introduced here in future studies.

CHAPTER 4 –

DISCUSSION & FUTURE DIRECTIONS

4.1 DISCUSSION

Decades of research have shown that RNA turnover rates vary widely between genes across all biological systems, and that differences in transcript half-lives serve an integral role in regulating gene expression. Here we show that the variability in half-lives measured at whole cell resolution persists across subcellular compartments. These differences between transcripts are the result of a myriad of features that define each gene and mRNA, resulting in RNA flow rates that are similar across transcripts encoding functionally related proteins. Additionally, we identify nuclear RNA degradation as a pervasive process that acts substantially on mRNAs, likely limiting the cytoplasmic abundance of select transcripts in a manner that can be quickly altered by the cell in a posttranscriptional manner. RNA flow represents the cumulative impact of multiple layers of gene regulation that act to control the subcellular fates and trajectories of transcripts. Analysis of RNA flow throughout development and in disease systems is expected to help further elucidate how regulatory programs control cell fate and multicellular phenotypes through subcellular transcript dynamics.

4.2 OUTSTANDING QUESTIONS RELATED TO RNA FLOW

1. *What are the RNA flow rates of non-coding transcripts?*

    We limit our analysis in this study to nuclear-encoded, protein-coding mRNAs, yet the data we present could be analyzed to quantify flow rates across other RNA species. For example, the small nuclear ribonucleoprotein particles (snRNPs) that comprise the spliceosome contain non-coding transcripts referred to as small nuclear RNAs (snRNAs). These snRNAs are transcribed and processed on chromatin but then exported into the cytoplasm for additional processing before being imported back into the nucleus, where the mature snRNP functions (Matera and Wang, 2014). Measuring

the flow rates of snRNAs, as well as other non-coding RNAs that shuttle across compartments (e.g., rRNAs and tRNAs), would be an intriguing direction for future research. These analyses may require modifications to the analysis pipeline and kinetic model, depending on the RNA species.

2. *How much does RNA flow vary between different biological systems?*

Median mRNA half-lives at whole-cell resolution vary significantly between different eukaryotic model organisms, ranging from 18 minutes in *S. cerevisiae* (Munchel et al., 2011) to 140 minutes in human K562 cells (Figure 2.2E). Moreover, genome architecture (e.g., gene length, number of exons per gene) varies between organisms, as do cell size and other physical features. How RNA flow rates scale across these different factors remains to be elucidated.

3. *What other genetic and molecular features influence RNA flow rates?*

We have begun to unravel the complex network of features that influence subcellular RNA stability and identified a total of 519 individual attributes that associate with flow rates (Table S3.3). Of course, our LASSO model is not comprehensive of all elements present in the cell. One feature family that remains relatively uninvestigated is RNA modifications. We show a relationship between m6A methylation and nuclear export (Figure 3.7A), but more than 100 RNA modifications have been reported, with implications across all stages of RNA metabolism (Zhao et al., 2017). If, and how, each of these marks present in the "epitranscriptome" influences RNA flow rates remains to be seen. Another largely unexplored characteristic of genes is their physical proximity to other loci within the nucleus. Active genes, along with pol II and transcription factors, are present in phase-separated "hubs" within the nucleus (Palacio and Taatjes, 2022). These

hubs may promote the co-expression of genes by concentrating *trans*-regulatory factors; in turn, the transcripts produced from these spatially related genomic regions may flow across the cell at similar rates.

While we examine the relationship between poly(A) tails and RNA flow, several questions still remain. We show that in human K562 cells, poly(A) tail lengths reflect subcellular half-lives in a compartment-specific manner (Figure 3.4F), but it is unclear whether perturbing tail lengths would influence RNA flow. In HeLa cells, depletion of PABPC resulted in longer steady-state poly(A) tail lengths and shorter whole-cell half-lives (Xiang and Bartel, 2021), both phenotypes presumably being restricted to the cytoplasm. However, the effect of altered poly(A) tail lengths on overall turnover and RNA flow rates in the "normal" context where PABPC is in excess remains to be determined (Görlach et al., 1994). Furthermore, mRNAs can be modified at the very 3' end of the poly(A), including being uridylated and guanylated (Chang et al., 2014), and the effect of these marks on RNA flow would be intriguing to explore.

4. *How much do RNA flow rates vary between cellular conditions, and what genes are affected?*

We report RNA flow rates for cells grown under "normal" laboratory conditions, but mammalian cells are constantly responding to external stimuli, and at times, adapting to stressful conditions. How overall RNA flow rates might vary between conditions is vital to understanding the role of RNA flow in gene regulation. Genes are up- and down-regulated in response to different stimuli and conditions, and it is intriguing to speculate that genes necessary for cell survival under specific contexts might be prioritized for quick chromatin release, nuclear export, and polysome loading. This has shown to be true for specific transcripts (e.g., heat shock genes) under specific

conditions (e.g., heat shock), but RNA flow has not been examined genome-wide (Zhou et al., 2015). Interestingly, many genes involved in signaling and sensing exhibit fast flow rates across all compartments under normal conditions (Figure 2.5A), indicating that they may be "primed" to respond quickly; yet, other genes with similar functions may only flow quickly across compartments when activated (Figure S2.7B).

Mammalian cells also exist in different developmental stages and cell types, and examining the changes in RNA flow between these different contexts may help elucidate how cell identity is determined. Additionally, measuring subcellular half-lives in asymmetric cells requiring complex and transcript-specific RNA localization, such as neurons, is an exciting avenue for future research.

5. *What are the effects of perturbing individual steps of gene expression on RNA flow across the entire cell?*

Eukaryotic cells use sophisticated feedback mechanisms to maintain transcriptome integrity. For example, inhibiting transcription results in delayed RNA degradation (Chen et al., 2008), and vice versa (Hartenian and Glaunsinger, 2019). The chemical inhibition of translation has also been shown to slow RNA decay (Roy and Jacobson, 2013). Thus, RNA levels are often buffered by compensatory changes when synthesis, translation, or decay factors are impacted. The levels of buffering, as well as any genes that are differentially affected, could be quantified by measuring RNA flow under these circumstances. Inhibition of other steps directly related to RNA flow, such as nuclear export, would also be informative. Additionally, it was recently reported that treatment with spliceosome inhibitor pladienolide B resulted in longer poly(A) tails on unspliced, nuclear RNA (Alles et al., 2021), and it is intriguing to speculate that this may be due to the prolonged retention of transcripts on chromatin.

6. *How is nuclear RNA degradation regulated?*

Our analysis of RNA flow suggests that nuclear degradation serves not only as a means of quality control but as a mechanism that regulates the subcellular abundance of select transcripts. While we indeed observed that transcripts predicted to undergo nuclear degradation have delayed or impaired splicing (Figure 3.5B-C), the magnitude of nuclear degradation (predicted to act on a majority of transcripts) would suggest that this process is not only responding to sporadic mistakes in RNA processing. Others have speculated that, in addition to clearing improperly processed mRNAs, nuclear degradation acts on transcripts aberrantly retained in the nucleus for a prolonged period (Schmid and Jensen, 2018). Our data is inconsistent with this hypothesis, as we report that genes predicted to undergo nuclear degradation on average have shorter chromatin and nuclear half-lives compared to others (Figure 3.5A). These transcripts are likely recruited to the nuclear exosome in a PPD/PAXT-dependent manner, as these pathways are known to degrade hyperadenylated mRNAs (Bresson and Conrad, 2013; Meola et al., 2016), but exactly what identifies them as substrates initially remains unknown. Moreover, how the magnitude of nuclear degradation changes under different conditions, as well as which transcripts are acted upon, is an exciting line of future inquiry.

7. *Can perturbing RNA flow rates drive disease?*

Changes in RNA flow rates have the potential to alter the subcellular abundance of transcripts, and consequently, the translational output of each mRNA. Importantly, this can occur without any changes in whole-cell half-life, and thus may serve as a means of "upregulating" a gene without altering RNA stability or abundance at whole-cell resolution. For instance, in the context of cancer, an oncogenic event may cause faster

nuclear export and/or cytoplasmic stabilization of an oncogene, resulting in longer residence of the transcript in polysomes and increased protein output. Indeed, aberrant nuclear export is a common characteristic of many cancers (Borden, 2020).

One such example related to RNA flow has already been shown. The cytoplasmic translation initiation factor eIF4E also functions in CRM1-dependent mRNA nuclear export, and overexpression of a mutant allele not capable of translation initiation results in an increased cytoplasmic to nuclear mRNA abundance of specific target transcripts, along with higher protein levels, without a change in RNA half-life (Culjkovic et al., 2005). Export targets of eIF4E include *Cyclin D1, E1, A2,* and *B1*, as well as *Myc* and *Mdm2* (Culjkovic et al., 2006), and overexpression of export-competent, but not export-incompetent, eIF4E is sufficient to transform cells (Culjkovic-Kraljacic et al., 2012). This system is an attractive application for studying changes in RNA flow during oncogenesis, and it has the potential to identify all mRNAs exported through this pathway. Future studies may uncover unappreciated roles for RNA flow in driving disease in other contexts.

APPENDIX 1 –

SUPPLEMENTAL MATERIALS RELATED TO CHAPTER 2

Some materials in this appendix are adapted from a manuscript currently under review entitled "Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle" and will be published at a later point in a peer-reviewed journal. Author contributions are the same as Chapter 2.

FIGURES
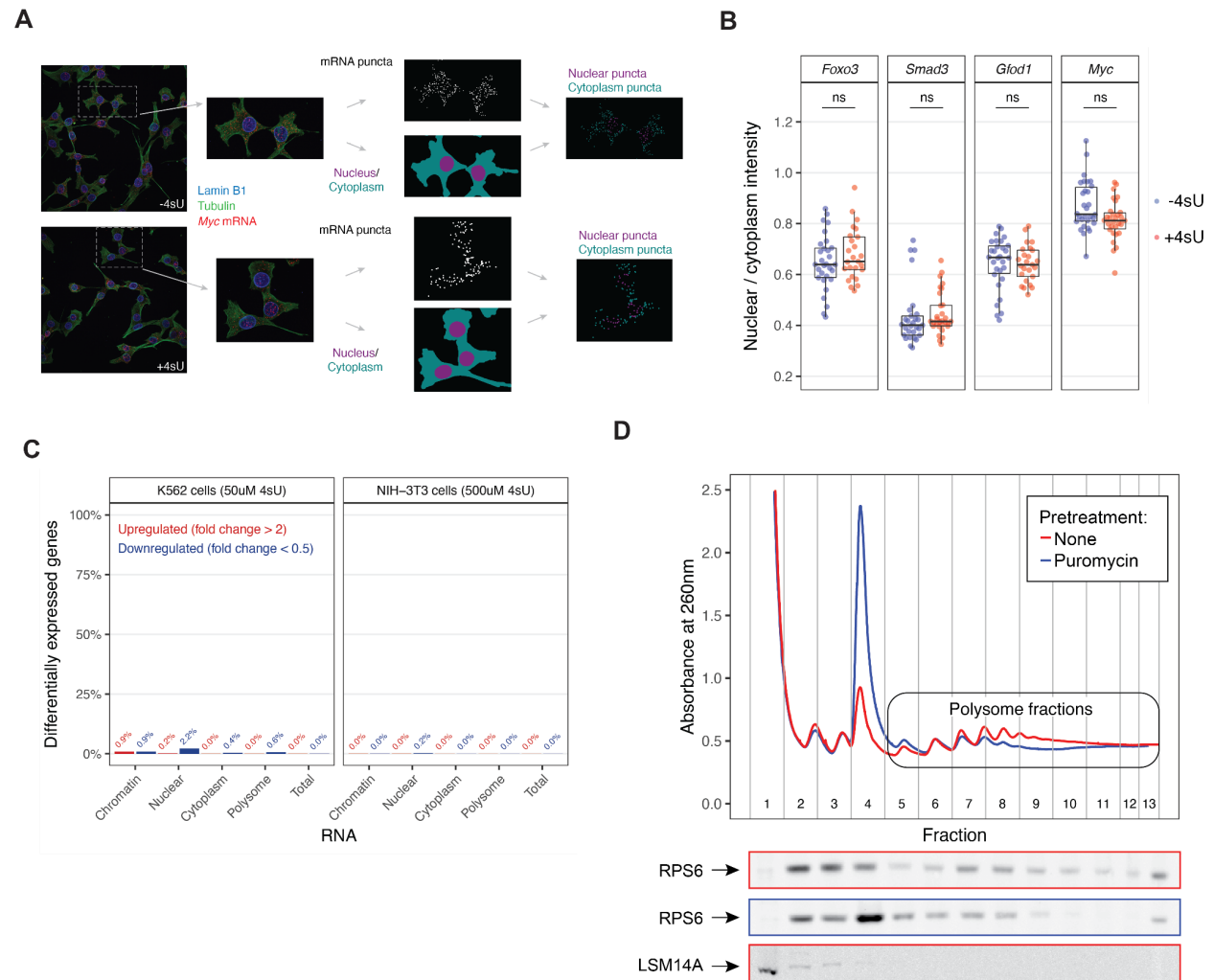
Figure S2.1: Optimization of biochemical fractionation and 4-thiouridine labeling conditions used in subcellular TimeLapse-seq.
A. SABER-FISH was performed in mouse NIH-3T3 cells to visualize *Myc* mRNA according to (Kishi et al., 2019) following 2 hours of 500uM 4sU treatment relative to no 4sU treatment. In addition to hybridizing *Myc* mRNA probes, concurrent immunohistological staining of Lamin B1 and alpha-Tubulin was performed to stain the nucleus and cytoplasm, respectively. mRNA puncta were identified according to (West et al., 2022), and nuclear and cytoplasmic regions within the image were segmented (see Methods for details). Using these defined regions, mRNA puncta were identified as nuclear or cytoplasmic and the puncta intensity within each compartment was summed over all cells in each field of view. A total of 25-30

Figure S2.1 (continued).
fields of view were analyzed for all genes. **B.** Summary of data for all genes *(Myc, Foxo3, Smad3, Gfod1)* analyzed according to (A). A t-test was performed to compare the differences in intensities between 4sU-treated and control cells ("ns:" non-significant). **C.** Number of differentially expressed genes across subcellular compartments in cells following 2 hours of 4sU pulse-labeling relative to no labeling. Differentially expressed genes were defined as those with fold change>2 or <0.5 with an adjusted p-value of <0.01 when comparing RNA-seq read counts to the unlabeled samples for each compartment using DESeq2 (Love et al., 2014). **D.** Purification of actively translating ribosomes by sucrose density gradient ultracentrifugation. Polysome profiling traces of K562 cell lysate were measured following no drug treatment (red) or 1 hour of 100ug/ml puromycin treatment (blue). Each fraction was also analyzed by western blotting for a ribosomal protein (RPS6) and a P-body component (LSM14A). Fractions 5+ were pooled and used to isolate polysome-associated RNA.
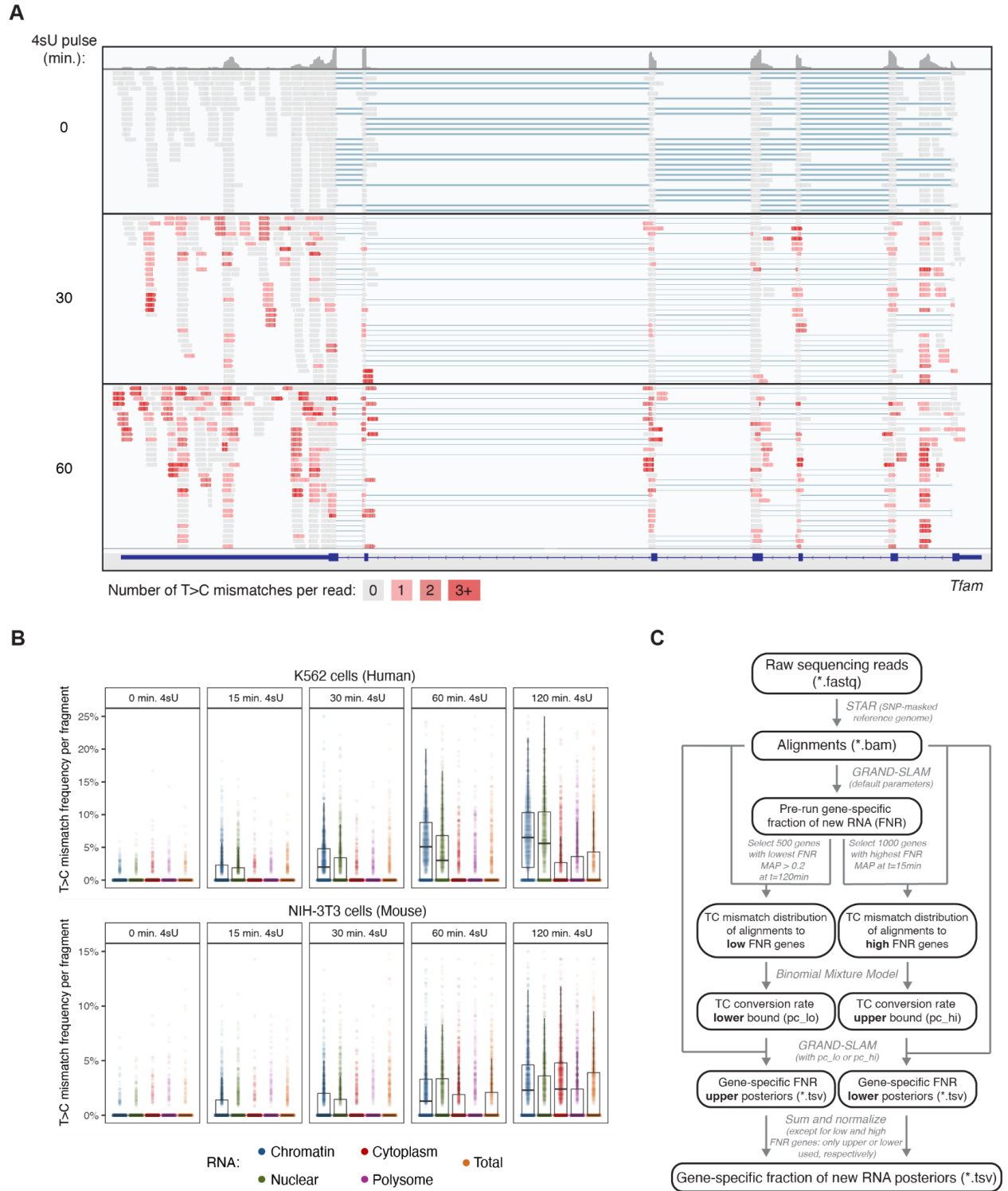
Figure S2.2: Nucleotide conversion analysis estimates the fraction of new RNA in a compartment- and time point-specific manner.
A. Example subcellular TimeLapse-seq reads in mouse NIH-3T3 cells for nuclear RNA aligning to *Tfam* following different 4sU pulse durations. Reads are colored by the number of T>C mismatches. **B.** Frequency of T>C mismatches in RNA-seq reads relative to 4sU pulse durations. The frequency of mismatches is calculated for each read as the number of T>C mismatches over the total number of Ts per

Figure S2.2 (continued).
fragment (using both forward and reverse reads). A total of 1,000 reads are analyzed for each sample. Dots represent individual fragments. **C.** Computational analysis pipeline for subcellular TimeLapse-seq data (see Methods for details).
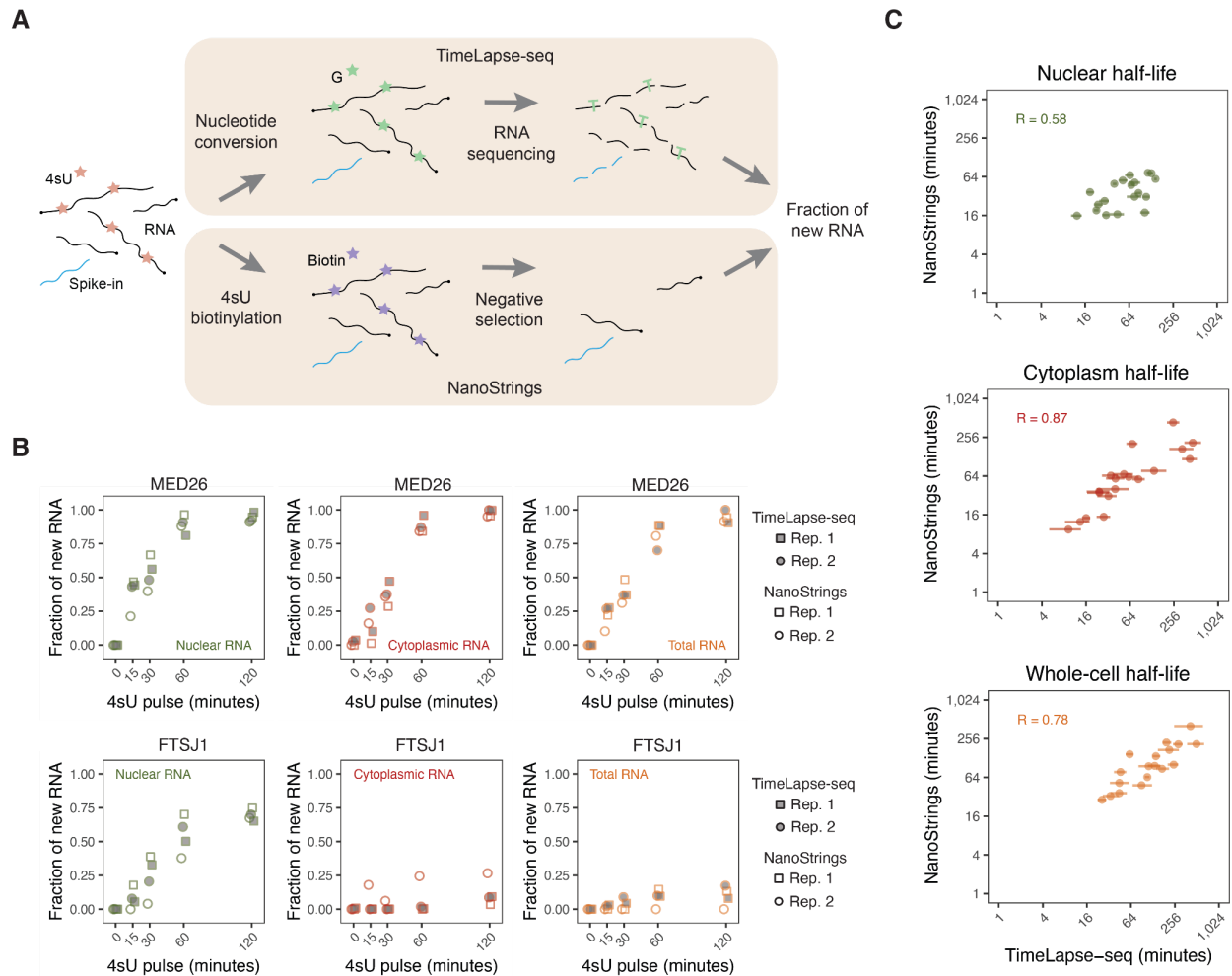
Figure S2.3: NanoStrings assay validates the fraction of new RNA estimated by subcellular TimeLapse-seq.
A. Schematic for NanoStrings-based and subcellular TimeLapse-seq estimation of fraction of new RNA. Top: TimeLapse-seq analysis uses an oxidative nucleophilic-aromatic substitution reaction (Schofield et al., 2018) to recode the 4-thiouridine (4sU) molecules as cytosines, resulting in the incorporation of a guanine nucleotide during the reverse transcription step of library preparation. These are subsequently converted into cytosines during PCR amplification and identified computationally as T (genome) to C (sequencing read) mismatches during alignment. The fraction of new RNA per gene is estimated from sequencing reads using a modified GRAND-SLAM pipeline (see Methods for more details). Bottom: NanoStrings-based analysis starts with the covalent biotinylationation of 4sU molecules, followed by the removal of 4sU-labeled RNAs by incubating the sample with streptavidin beads and retaining the supernatant (unbound RNAs). The number of RNAs per gene in the remaining sample (unlabeled RNAs) is determined by hybridization with NanoStrings probes. The fraction of 4sU-labeled is determined by normalizing the RNA counts to an unlabeled spike-in RNA and comparing to no 4sU control (see Methods for more detail). **B.** Fraction of new RNA within nuclear (left), cytoplasmic (middle), and total RNA (right) compartments as measured by subcellular TimeLapse-seq and NanoStrings for two example genes *(MED26* and *FTSJ1)* in human K562 cells. Two biological replicates ("rep") are shown for each approach. **C.** Correlation of fraction of new RNA between NanoStrings and subcellular TimeLapse-seq for all genes and compartments. The data shown in (B) is summarized by calculating a nuclear, cytoplasm, and whole-cell half-life for each gene included in the NanoStrings panel (n=20) with the fraction of new RNA from each technique. Half-lives are calculated from NanoStrings-based fraction of new RNA values with least squares estimates (see Methods for more detail). Each dot represents one

67

Figure S2.3 (continued).
gene. Mean half-lives between replicates for each technique are plotted with the Pearson correlation. The horizontal error bars represent the 95% credible intervals from the RNA kinetic flow model from subcellular TimeLapse-seq.

A

Kinetic models

- Whole cell turnover k_{WC}: $\frac{dT(t)}{dt} = k_p - k_{WC}T(t)$

- Chromatin turnover k_{CH}: $\frac{dCH(t)}{dt} = k_p - k_{CH}CH(t)$

- Nuclear turnover k_N: $\frac{dN(t)}{dt} = k_p - k_N N(t)$

- Cytoplasmic turnover k_{CY}: $\frac{dCY(t)}{dt} = k_N N(t) - k_{CY}CY(t)$

- Polysome loading k_{PL}: $CY(t) =: M(t) + P(t) \Rightarrow$
$\frac{dM(t)}{dt} = k_N N(t) - (k_{CY} + k_{PL})M(t)$
$\frac{dP(t)}{dt} = k_{PL}M(t) - k_{CY}P(t)$

- Nuclear export $k_E^{-1} := k_N^{-1} - k_{CH}^{-1}$

Nuclear degradation model as above, except:

- Nuclear export k_E: $T(t) = N(t) + CY(t) \Rightarrow$
$\frac{dT(t)}{dt} = k_p - (k_N - k_E)N(t) - k_{CY}CY(t)$

- Nuclear degradation $k_{ND} := k_N - k_E$

Analytically solve ODEs

Model fraction of new RNA for each compartment X
$\Lambda_X(\vec{k}, t)$

Pipeline described in Figure S2.2C

Posterior (probability density function) on the fraction of new RNA (π) for gene g in compartment X at time t
$\mathbb{P}^g_{X,t}(\pi)$

Change of variables to rate vector domain

Posterior (probability density function) on the rate vector ($\vec{k}$) for gene g derived from compartment X
$\mathbb{P}^g_X(\vec{k}) = \prod_{i=1}^4 \mathbb{P}^g_{X,t_i}(\pi = \Lambda_X(\vec{k}, t_i))|\det \mathbf{D}\vec{\Lambda}(\vec{k})|$

Bayes factor calculation using nuclear, cytoplasm and whole cell compartments

Univariate cases: chromatin, nuclear and whole cell compartments

Bayes Factor > 100? *Yes*

Rate summary statistics
$\mathbf{MAP}(k_X^g) = \underset{k}{\arg\max}\ \mathbb{P}^g_X(k)$

$\mathbf{Mean}(k_X^g) = \int_{\text{prior domain}} \kappa \mathbb{P}^g_X(\kappa)d\kappa$

$\mathbf{95\% CI}(k_X^g) = [k_{lo}, k_{hi}]$ such that $\begin{cases} \mathbb{P}^g_X(k \le k_{lo}) = 2.5\% \\ \mathbb{P}^g_X(k \le k_{hi}) = 97.5\% \end{cases}$

No

Nuclear export posterior (non-PUNDs)
$\mathbb{P}^g_E(k) = \int_{\text{prior domain}} \mathbb{P}^g_{nuc}\left(\frac{kz}{k+z}\right)\mathbb{P}^g_{chr}(z)dz$

Whole cell compartment and numerical integration over nuclear and cytoplasmic turnover 95%CIs

Nuclear export posterior (PUNDs)

Nuclear degradation posterior (PUNDs)
$\mathbb{P}^g_{ND}(k) = \int_{k+k_{E\,CI\,lo}}^{k+k_{E\,CI\,hi}} P_{nuc}(z)dz$

Cytoplasm compartment and numerical integration over nuclear 95%CI

Cytoplasmic turnover posterior

Polysome loading posterior

Polysome compartment and numerical integration over nuclear and cytoplasmic turnover 95%CIs

B

K562 whole-cell half-life

r = 0.65
n = 5,059

Mean, Schofield et al., 2018 (min.)

Mean, this study (min.)

C

K562 cells (Human)

Chromatin — r = 0.85, n = 11,610
Nuclear — r = 0.88, n = 12,014
Nuclear export — r = 0.85, n = 11,347

Nuclear degradation — r = 0.78, n = 384
Cytoplasm — r = 0.81, n = 11,657
Untranslated cytoplasm — r = 0.44, n = 9,852

Whole cell — r = 0.87, n = 11,103

Half-life rep. 2 (min.)
Half-life rep. 1 (min.)

D

NIH-3T3 cells (Mouse)

Chromatin — r = 0.90, n = 11,689
Nuclear — r = 0.89, n = 11,458
Nuclear export — r = 0.78, n = 11,996

Nuclear degradation — r = 0.93, n = 934
Cytoplasm — r = 0.76, n = 11,181
Untranslated cytoplasm — r = 0.45, n = 9,194

Whole cell — r = 0.88, n = 11,499

Half-life rep. 2 (min.)
Half-life rep. 1 (min.)

Figure S2.4: RNA flow can be modeled using a series of ordinary differential equations.
A. Schematic of RNA kinetic modeling (see Methods for more detail). **B.** Correlation between whole-cell

69

Figure S2.4 (continued).
(total RNA) half-lives measured in K562 cells in this study and in a previous study. The mean total RNA half-life between replicates, calculated using TimeLapse-seq following a single 4sU pulse (Schofield et al., 2018), was compared to the mean whole-cell half-lives between replicates in this study. Pearson's correlation is shown and the number of genes is noted. Each dot represents one gene. **C.** Reproducibility of RNA flow rates across biological replicates in human K562 cells. For each flow rate, the mean half-life is compared between replicates. Pearson's correlation is shown and the number of genes is noted. Each dot represents one gene. **D.** Same as (C) for mouse NIH-3T3 cells.

A

K562 cells (Human)

B

NIH-3T3 cells (Mouse)
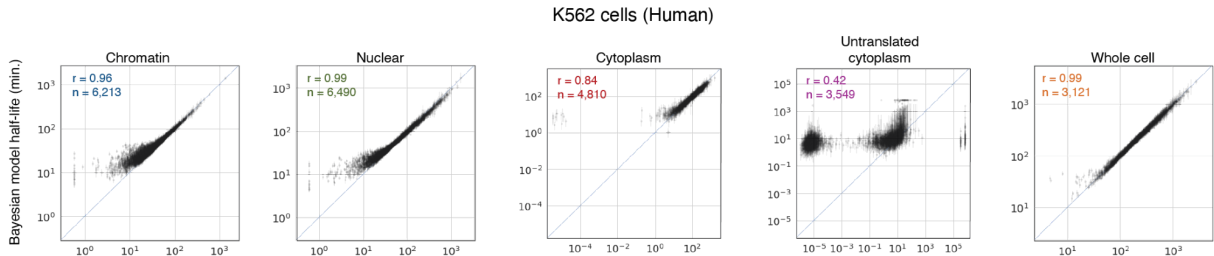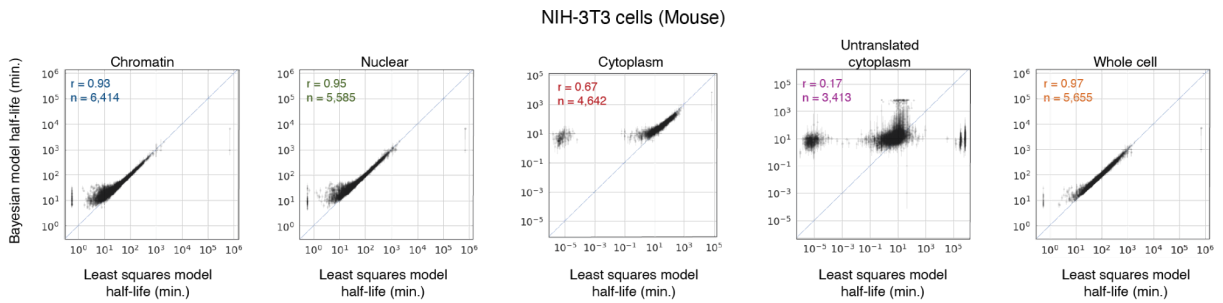
Figure S2.5: Bayesian model of RNA flow outperforms the least squares model.
A. Comparison between the Bayesian model and a least squares model of RNA flow rates in human K562 cells. The Bayesian MAP half-life for each subcellular compartment for one replicate is compared with the least squares estimate. Pearson's correlation is calculated and the number of genes is noted. Each dot represents one gene. Error bars indicate 95% credible interval of Bayesian half-lives. **B.** Same as (A) in mouse NIH-3T3 cells.

71

A

Identification of PUNDs in NIH-3T3

- PUNDs: Genes predicted to undergo nuclear degradation (n = 946)
- Other genes (n = 11,109)

Y-axis: Bayes factor (10^{48}, 10^{32}, 10^{16}, 10^{8}, 10^{2}, 10^{0}, 10^{-8}, 10^{-16})

X-axis: Nuclear degradation rate (min^{-1}) (10^{-4}, 10^{-3}, 10^{-2}, 10^{-1})

Labeled genes: Hnrnph1, Srst2, U2surp, Srek, Thoc1, Smg5, Nup85, Dkc1
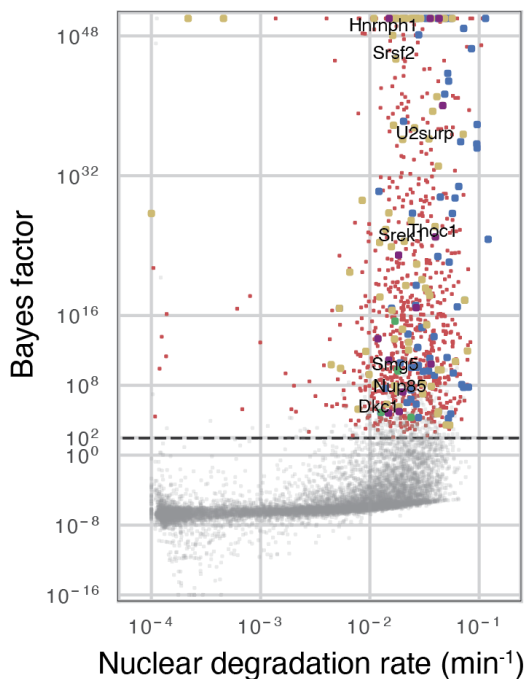
Enriched GO terms (annotated PUNDs):
- Ribosome
- Telomerase holoenzyme complex
- RNA splicing
- mRNA export from nucleus

B

NIH-3T3 PUNDs — 813 genes

133 genes
$p=2.3 \times 10^{-14}$

K562 PUNDs — 275 genes

Common genes include:
- 24 RPGs
- 7 HNRNPs
- 4 SRSFs
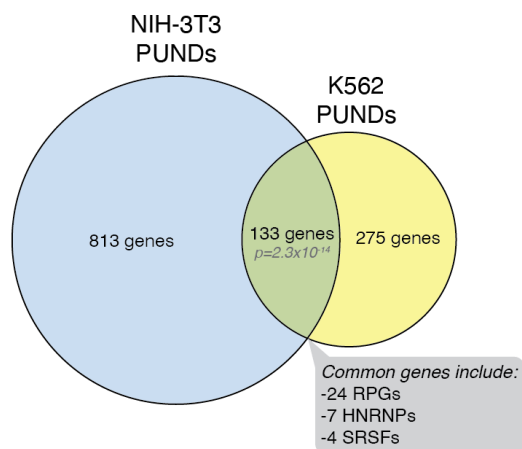
Figure S2.6: Genes predicted to undergo nuclear degradation are conserved between cell lines.
A. Identification of PUND genes in mouse NIH-3T3 cells. The Bayes factor and nuclear degradation rate are shown for each gene as in Figure 2.4A. **B.** Venn diagram showing the number of unique and common PUND genes between mouse NIH-3T3 and human K562 cells. A Fisher's exact test was performed to test the significance of the number of common genes. This common gene list included ribosomal protein genes (RPGs), heterogeneous nuclear ribonucleoproteins (hnRNPs), and SR splicing factors (SRSFs).
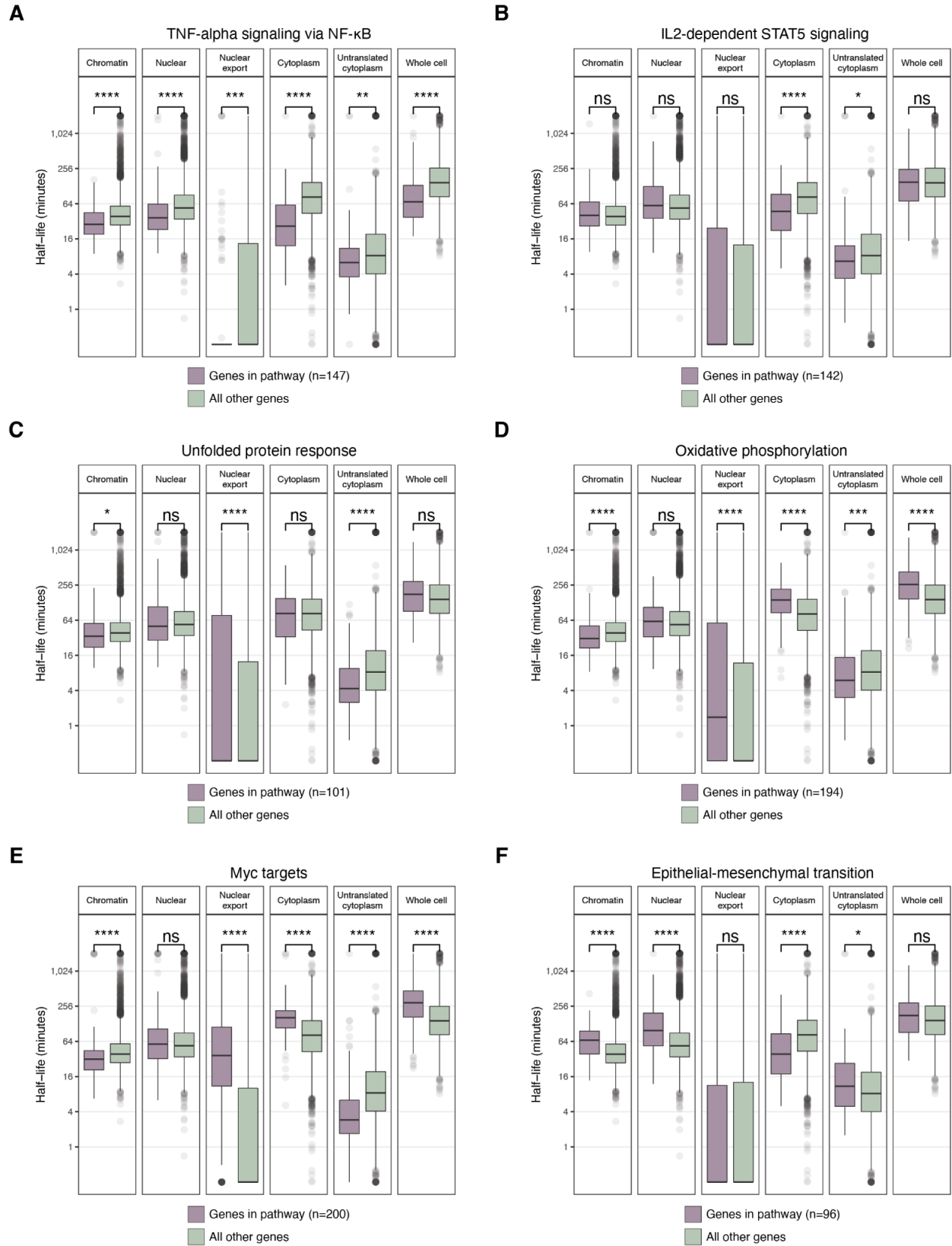
Figure S2.7: Functionally related genes exhibit similar RNA flow.

Figure S2.7 (continued).
A. RNA flow rates in human K562 cells for genes involved in TNF-alpha signaling via NF-κB, as defined by GSEA MSigDB (Subramanian et al., 2005). Subcellular half-lives were compared between groups using a Wilcoxon test (****: p<0.0001, ***: p<0.001, **: p<0.01, *: p<0.05, "ns": non-significant). **B.** Same as (A) for genes involved in IL2-dependent STAT5 signaling. **C.** Same as (A) for genes involved in the unfolded protein response. **D.** Same as (A) for genes involved in oxidative phosphorylation. **E.** Same as (A) for genes that are targets of the MYC transcription factor. **F.** Same as (A) for genes involved in the epithelial-mesenchymal transition.

TABLES

All tables are available online.

Table S2.1: RNA flow rates of protein-coding genes in human K562 and mouse NIH-3T3 cells.

Table S2.2: Gene lists generated and used in this study.

Table S2.3: Gene ontology (GO) enrichment results in human K562 and mouse NIH-3T3 cells.

Table S2.4: Gene set enrichment analyses (GSEA) of each RNA flow rate with genes ranked by subcellular half-lives in human K562 cells.

APPENDIX 2 –

SUPPLEMENTAL MATERIALS RELATED TO CHAPTER 3
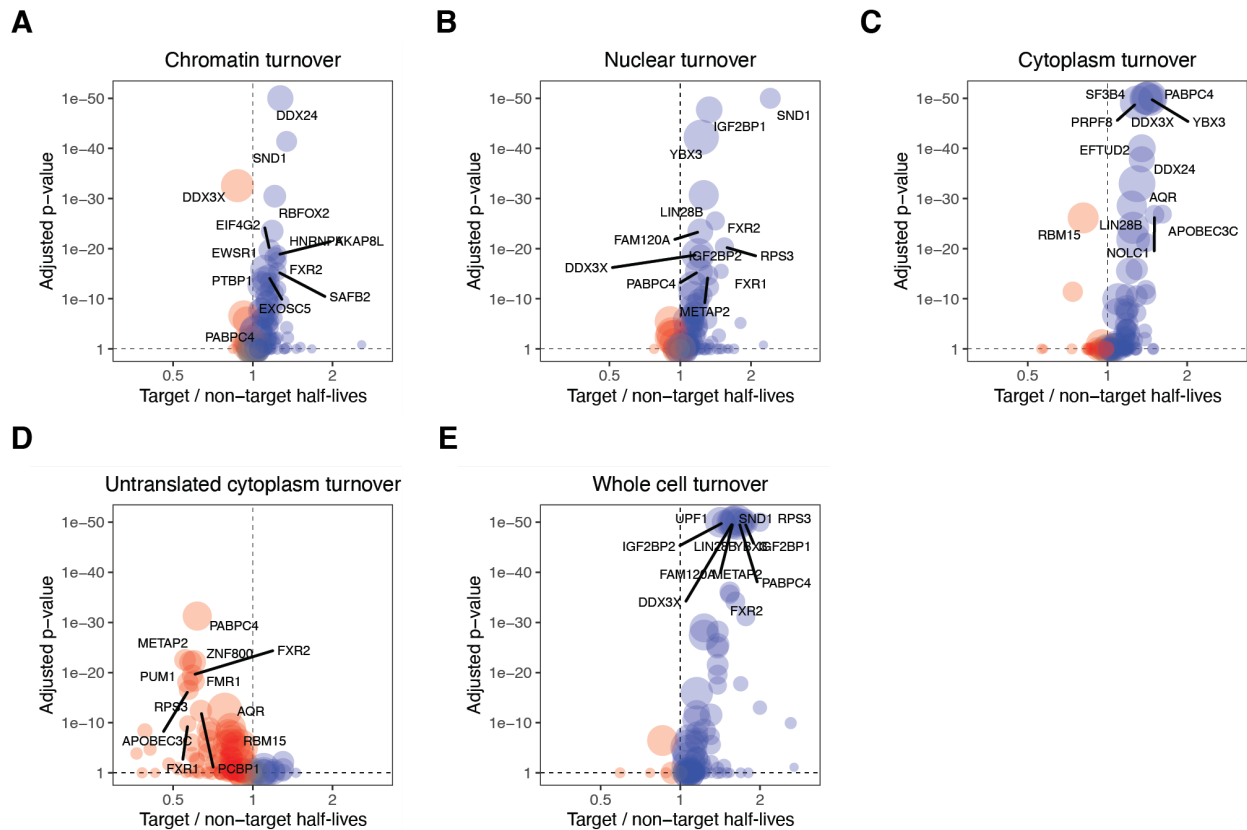
FIGURES

Figure S3.1: RNA binding proteins significantly associated with RNA flow rates.
A. Comparison of chromatin RNA half-lives between target and non-target mRNAs of each RBP according to Fig. 3.1A. The median chromatin half-life of the targets was compared to the median chromatin half-life of the non-targets with a Wilcoxon test (x-axis) with RBPs associated with slower target turnover in blue and RBPs with faster target turnover in red. The adjusted p-value following a Bonferroni correction for each RBP is indicated on the y-axis, and the size of the dot for each RBP indicates the number of target mRNAs. **B.** Same as (A) for nuclear half-lives. **C.** Same as (A) for cytoplasm half-lives. **D.** Same as (A) for untranslated cytoplasm half-lives. **E.** Same as (A) for whole-cell half-lives.

A. Fraction of new RNA in chromatin (left), cytoplasm (middle), and whole-cell (right) following DDX3X knockdown (see Fig. 3.2B) for target and non-target mRNAs. The fraction of new RNA MAP values were compared between targets and non-targets with a Wilcoxon test (****: p<0.0001, *: p<0.05). **B.** Same as (A) following PABPC4 knockdown (see Fig. 3.2E).
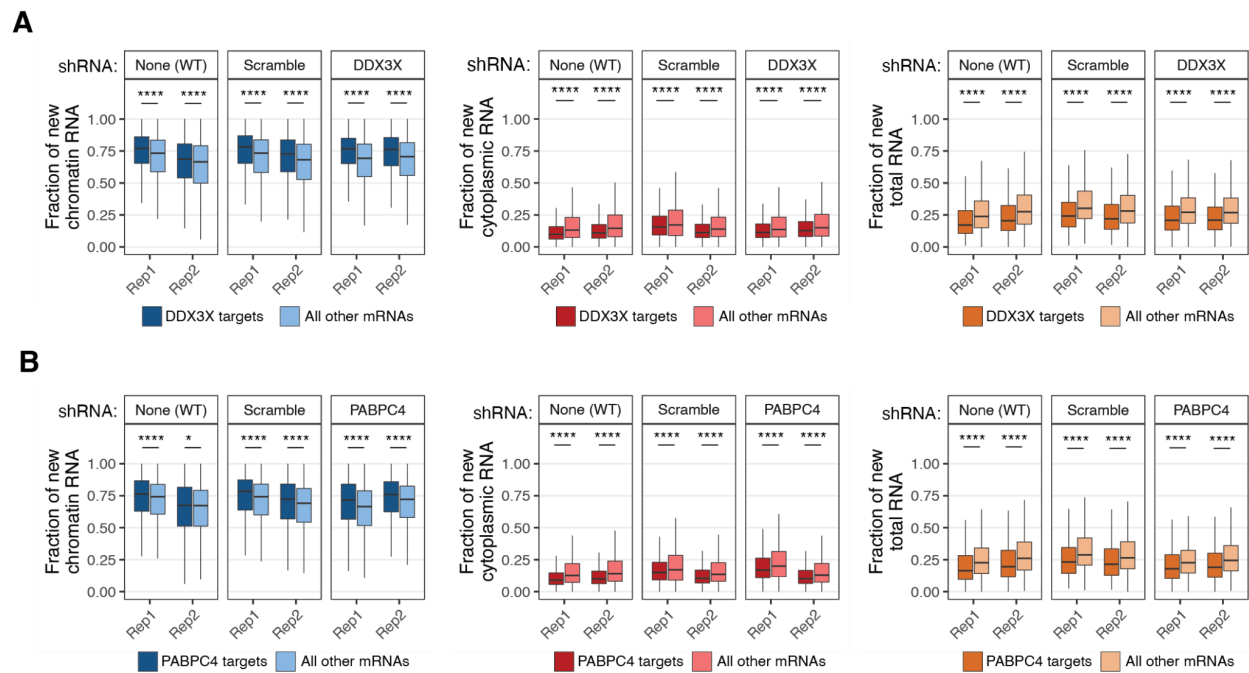
Figure S3.2: Subcellular TimeLapse-seq data for RBP depletion experiments in unaffected compartments.
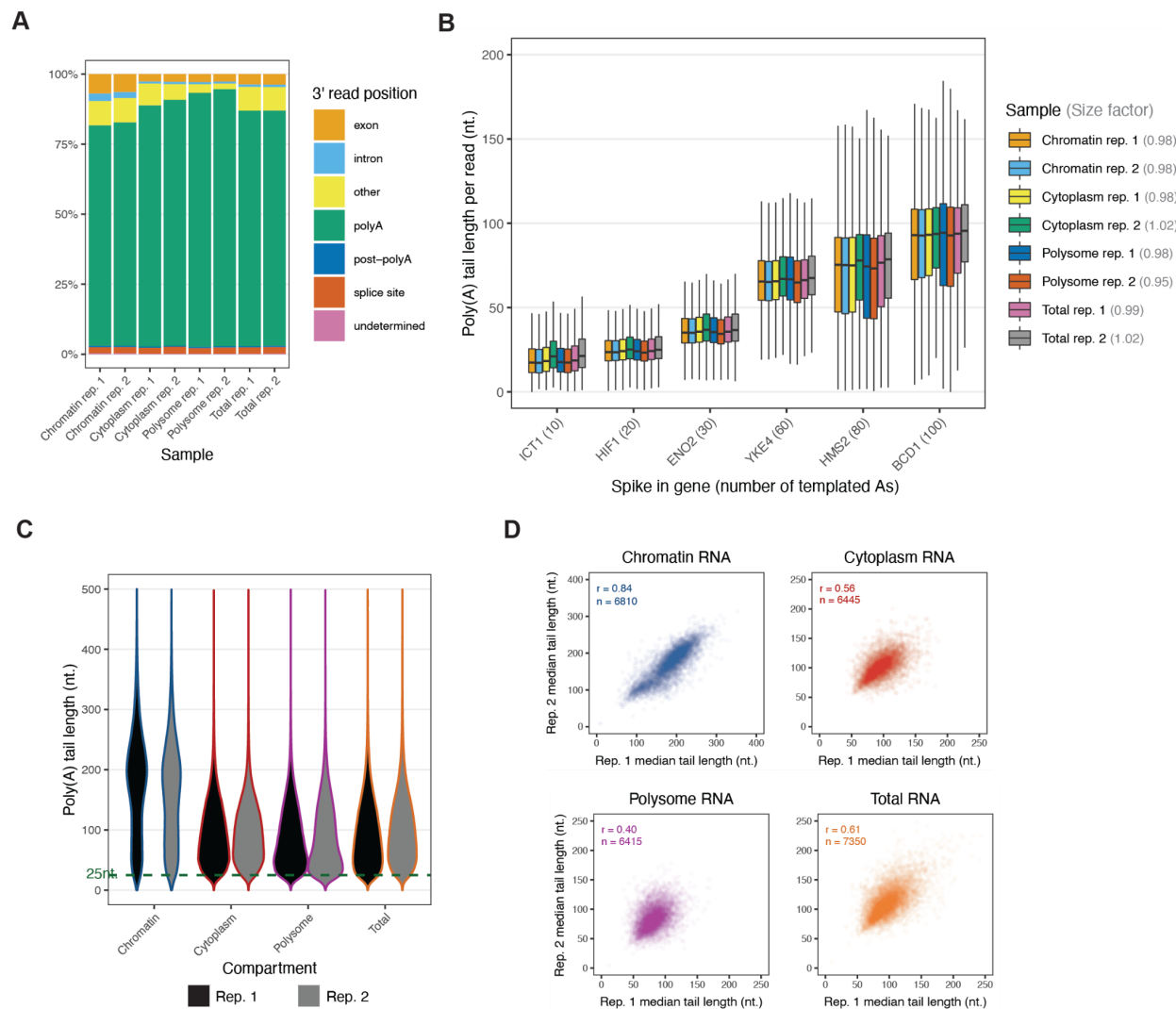
Figure S3.3: Measurement of poly(A) tail lengths across subcellular compartments.
A. Distribution of 3' ends of poly(A)-selected RNA direct sequencing reads. The genomic region corresponding to the 3' end of reads across all samples was determined according to (Drexler et al., 2020). **B.** Distribution of poly(A) tail lengths for synthetic spike-in RNAs across nanopore sequencing runs. Six transcripts (shown on the x-axis) from *S. cerevisiae* with templated poly(A) tails ranging from 10 to 100 nucleotides were transcribed in vitro and added to each sample prior to nanopore library preparation. The median poly(A) tail length of each spike-in transcript within each sample was then used to calculate a poly(A) tail length size factor for each sample (noted in gray). Raw poly(A) tail lengths for each read were normalized to this size factor (see Methods for more details). **C.** Distribution of poly(A) tail lengths per read, normalized to the synthetic spike-ins, across all samples and replicates. **D.** Correlation of median compartment-specific poly(A) tail lengths between biological replicates. Median tail lengths are calculated for all genes containing >=10 reads for each compartment, with the Pearson correlation between biological replicates and the number of total genes noted. Each dot represents one gene.
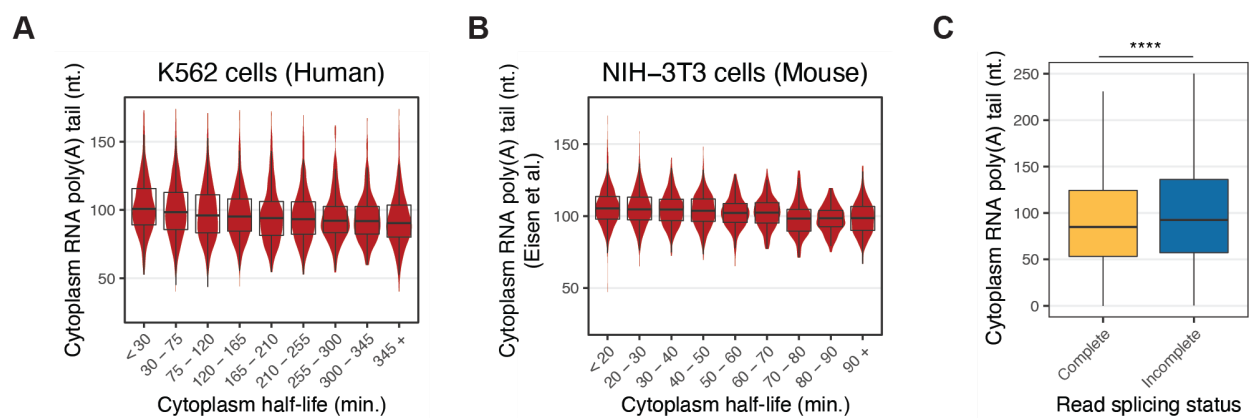
Figure S3.4: Cytoplasmic poly(A) tail length relates to cytoplasm RNA residence time.
A. Distribution of cytoplasm poly(A) tail lengths as a function of cytoplasm half-lives in human K562 cells. The median cytoplasm RNA poly(A) tail length is shown for all genes containing >=10 reads.
B. Distribution of cytoplasm poly(A) tail lengths as a function of cytoplasm half-lives in mouse NIH-3T3 cells. The mean steady state poly(A) tail length in cell line 1 (Eisen et al., 2020a) was compared to the cytoplasm half-lives (this study). **C.** Cytoplasm poly(A) tail lengths compared to read splicing status of cytoplasm RNAs. The splicing status of each cytoplasm RNA read was determined according to (Drexler et al. 2020) for each read spanning at least two introns, where "complete" means that all introns are excised and "incomplete" means that some introns are present. The distribution of poly(A) tail lengths between completely spliced and incompletely spliced RNAs was compared using a t-test (****: p<0.0001).
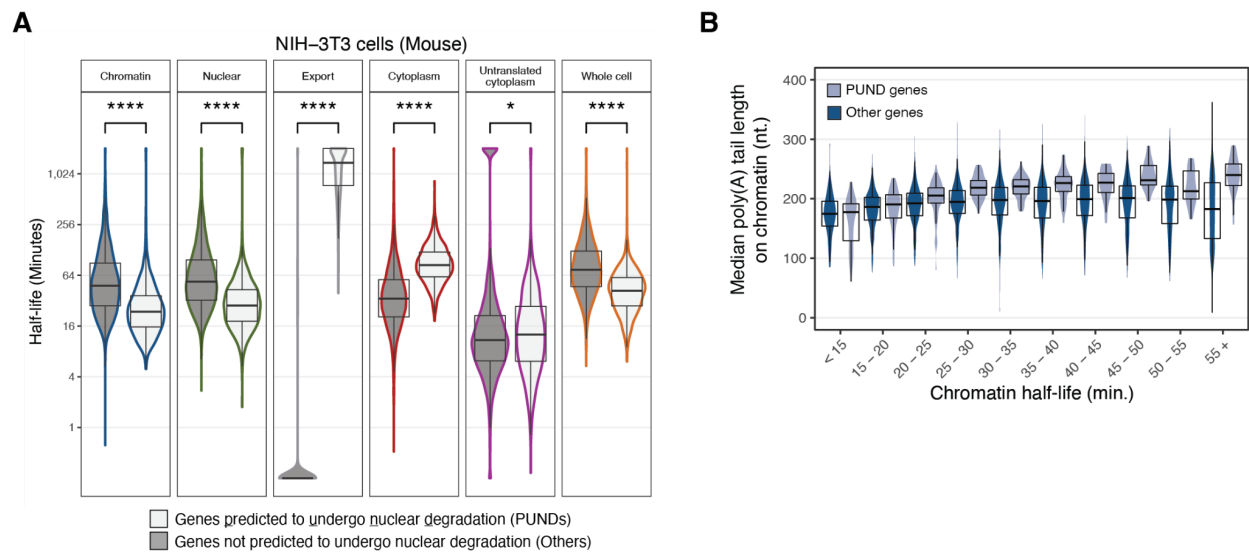
Figure S3.5: PUND phenotypes.
A. Comparison of half-lives of all PUND genes (n=946) and all other genes not predicted to undergo nuclear degradation across all RNA flow rates in NIH-3T3 using a Wilcoxon test (****: p<0.0001, *: p<0.05). **B.** Distribution of median chromatin poly(A) tail lengths for PUND genes relative to all other genes as a function of chromatin half-life. The median poly(A) tail length was calculated for all genes containing >=10 reads. Within each bin, the lighter blue represents PUND genes while the darker blue represents other genes.

A.

Genome-wide RNA flow rates for compartment X for 2 biological replicates

Feature classes
- Gene structure
- Histone modifications
- Gene location
- Gene / transcript sequences
- Codon frequencies
- Poly (A) tail length
- RNA modifications
- miRNA targets
- RBP targets
- TF targets (source: scRNAseq)
- TF and miRNA targets (source: MSigDB)

90-10 Training - test set split, same for replicates

Test set: rates for 10% of genes

Training/CV set: rates for 90% of genes

Bio rep 1 rates

Bio rep 2 rates

Round 1 LASSO 10x CV

Features from one class

Features with β > 0 at α-optimized CV performance, reproducible across both replicate runs

Union over all feature classes

Round 2 LASSO 10x CV

Features with β > 0 at α-optimized CV performance

LASSO model:
Loss = $(1 / (2 * N)) * ||y - \beta X||_2^2 + \alpha ||\beta||_1$
X = feature matrix
y = Z-score(log(rates))
N = number of rates in training set

Model performance on test set

B.

Whole cell half life predictions

R^2

Cross validation
Test

| Rates: | This study | Agarwal & Kelley | Agarwal & Kelley |
| Model: | This study | This study | Agarwal & Kelley |

Figure S3.6: LASSO model design and performance across whole-cell datasets.
A. Schematic of LASSO feature selection and model training, 10x cross-validation, and testing (see Methods for more details). **B.** Comparison of our LASSO 10x cross-validation and test performance with alternative whole-cell turnover rate estimates and models.
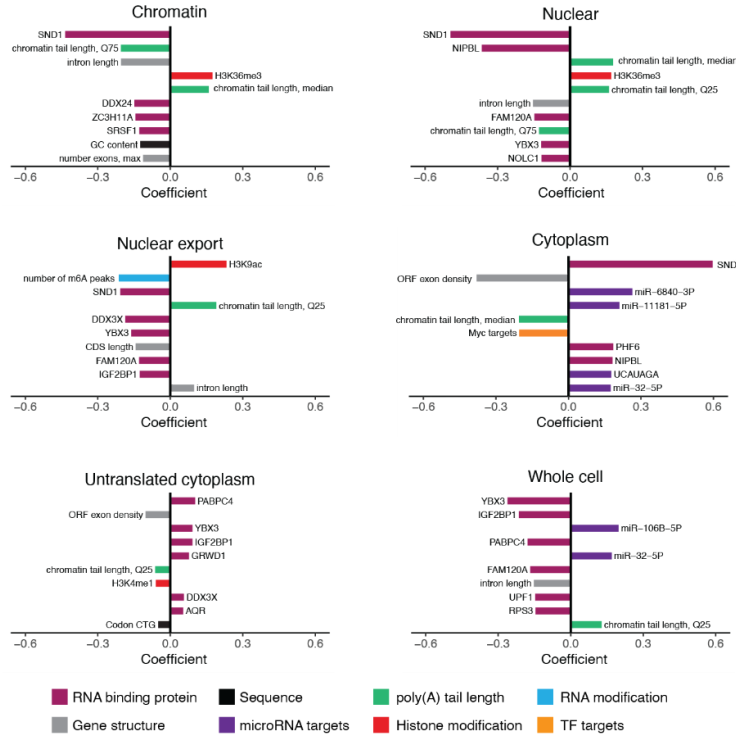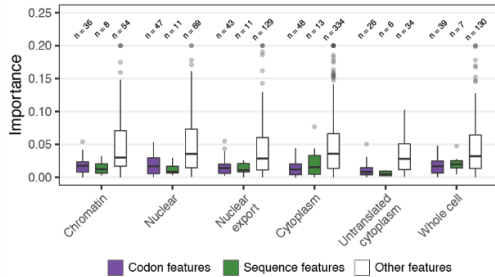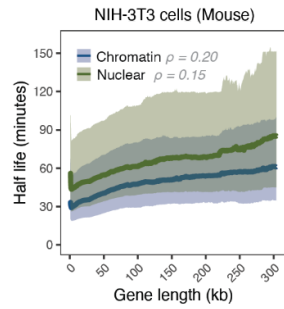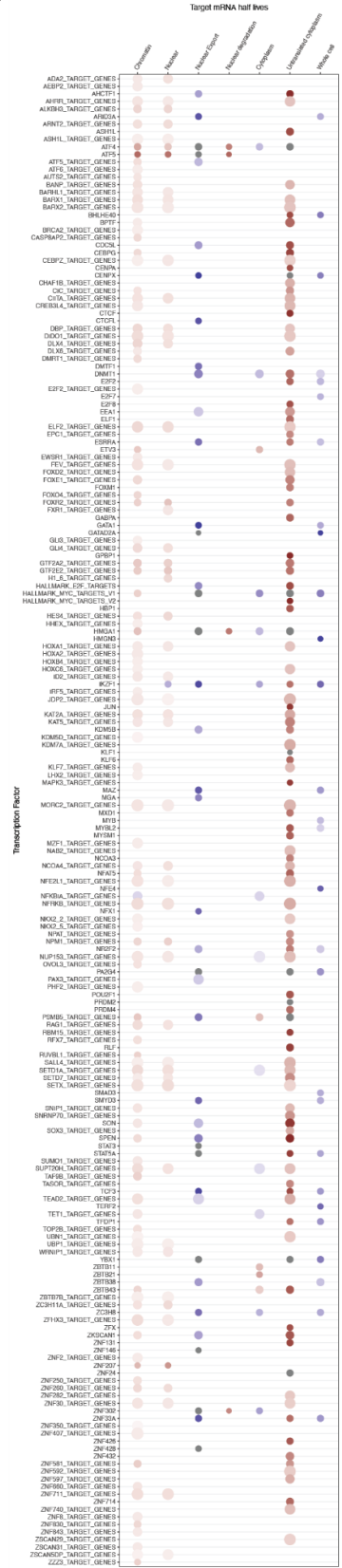
Figure S3.7: LASSO model predictions.
A. The 10 individual features with the highest importance for each RNA flow rate. Features are colored by family. **B.** The importance of individual sequence and codon features compared to individual features of other families. The number of relevant individual features is noted. **C.** Continuous averages of chromatin and nuclear half-lives as a function of gene length in mouse NIH-3T3 cells. Gene length was defined as the median genomic length of all transcripts per gene. Solid lines represent median half-lives and shaded ribbons represent the third quartile (top) and first quartile (bottom) of half-lives. **D.** All transcription factors with targets that exhibited significantly fast or slow half-lives for target RNAs compared to non-target RNAs in both biological replicates (adjusted $p<0.01$, Wilcoxon test, Bonferroni multiple testing correction) across any RNA flow rate. The size of the dot indicates the number of target mRNAs with measured half-lives within each compartment and the color reflects the difference (red, faster; blue, slower) of the median target over non-target half-lives.

Figure is displayed on previous page.
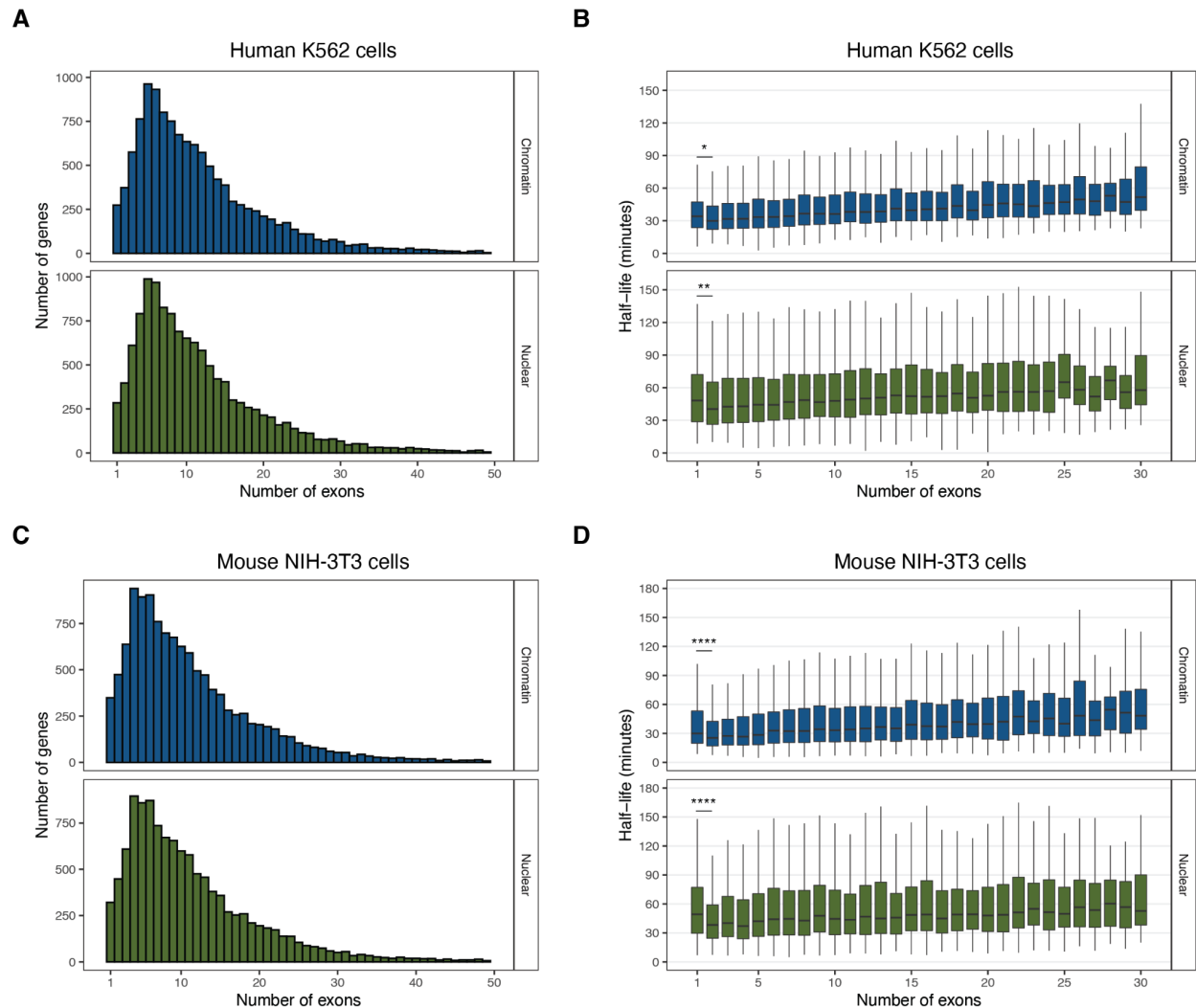
Figure S3.8: The number of exons per gene exhibits complex relationships with chromatin and nuclear RNA half-lives.
A. Histogram depicting the number of genes containing different numbers of exons in human. Genes were only included on the top (blue) if they had a measured chromatin RNA half-life and on the bottom (green) if they had a measured nuclear RNA half-life in human K562 cells. The maximum number of exons was considered if a gene contained multiple annotated isoforms. **B.** Human K562 subcellular half-lives as a function of the number of exons per gene. Chromatin RNA half-lives are plotted on the top (blue) and nuclear RNA half-lives are plotted on the bottom (green). The half-lives were compared between genes containing 1 and 2 exons using a Wilcoxon test (****: $p<0.0001$, **: $p<0.01$, *: $p<0.05$). **C.** Same as (A) in mouse NIH-3T3 cells. **D.** Same as (B) in mouse NIH-3T3 cells.
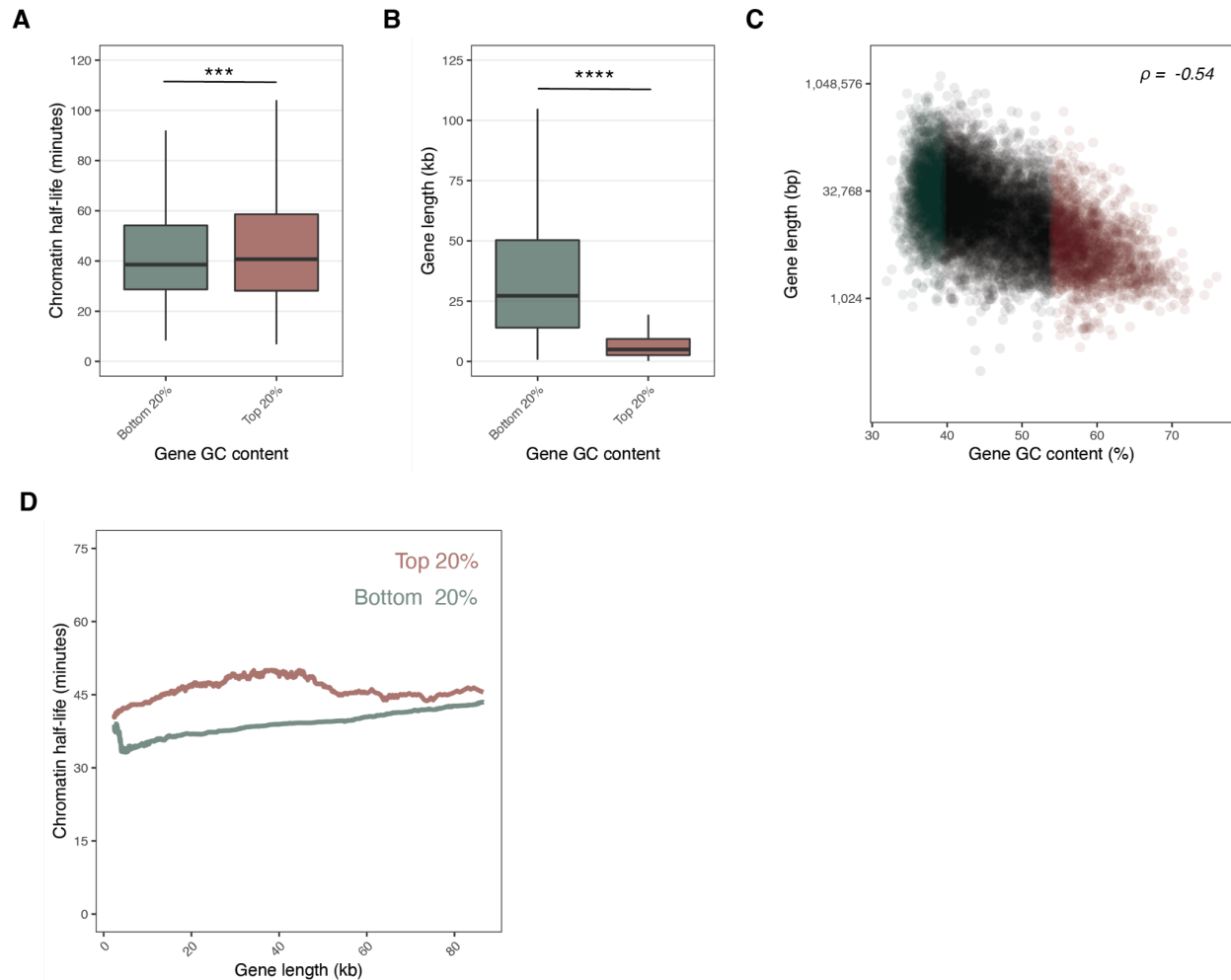
Figure S3.9: Higher GC content is associated with longer chromatin RNA half-lives.
A. Chromatin RNA half-lives as a function of gene GC content. Genes with chromatin RNA half-lives in human K562 cells were binned by gene GC content, with the lowest 20% of genes shown in teal and the top 20% of genes in salmon. Half-lives were compared using a Wilcoxon test (***: p<0.001). **B.** Gene length as a function of gene GC content. The lengths were compared between the same groups of genes analyzed in (A) using a Wilcoxon test (****: p<0.0001). **C.** Scatter plot including all genes comparing GC content and gene length. The median GC content was considered after calculating the GC content of all isoforms for genes with multiple transcripts. Genes are colored as in (A) and (B), with black dots representing genes not included in the bottom and top 20% groups. Spearman's rho is noted. **D.** Continuous averages of chromatin RNA half-lives as a function of gene length both for the 20% of genes with the highest and lowest GC content.

TABLES

All tables are available online.

Table S3.1: RNA binding proteins significantly associated with RNA flow rates in human K562 cells.

Table S3.2: Normalized median subcellular and whole-cell poly(A) tail lengths of protein-coding genes in human K562 cells.

Table S3.3: LASSO model features and performances.

Table S3.4: Transcription factors significantly associated with RNA flow rates in human K562 cells.

Table S3.5: Primers, antibodies, and plasmids used in this study.

APPENDIX 3 –

METHODS FOR CHAPTER 2

This appendix contains the materials and methods related to the results described in Chapter 2. Author contributions are the same as Chapter 2.

CELL CULTURE

NIH-3T3 cells (ATCC CRL-1658) were maintained at 37°C and 5% CO_2 in DMEM (ThermoFisher 11995073) with 10% cosmic calf serum (Cytiva SH30087.03), 100 U/mL penicillin, and 100 ug/mL streptomycin (ThermoFisher 15140122). K562 cells (ATCC CCL-243) were maintained at 37°C and 5% CO_2 in RPMI (ThermoFisher 11875119) with 10% FBS (Corning 35015CV), and 100 U/mL penicillin, and 100 ug/mL streptomycin (ThermoFisher 15140122). HEK-293T cells (ATCC CRL-3216) were maintained at 37°C and 5% CO_2 in DMEM (ThermoFisher 11995073) with 10% FBS (Corning 35015CV), 100 U/mL penicillin,and 100 ug/ mL streptomycin (ThermoFisher 15140122).

4sU LABELING

Pulse-labeling was performed with 4-thiouridine (4sU, Sigma T4509) resuspended in conditioned cell media. Labeling was performed in NIH-3T3 with cells at 40% confluency (approximately 8×10^6 cells in a 15cm plate) and a final 4sU concentration of 500uM. Labeling was performed in K562 with $4\text{-}5 \times 10^5$ cells/mL at a final 4sU concentration of 50uM. At the beginning of each labeling period, cells were removed from the incubator, 4sU was added directly to the existing cell media, and cells were returned to the incubator for the remainder of the pulse. At the end of the pulse, K562 cells were pelleted at 500xg for 2 minutes. The supernatant (cell media) was discarded for both NIH-3T3 and K562 and cells were immediately placed on ice and fractionated or lysed in 500uL RIPA buffer (ThermoFisher 89900) to collect total RNA.

89

SABER-FISH AND DATA ANALYSIS

mRNA transcripts corresponding to *Foxo3, Smad3, Gfod1,* and *Myc* were detected in NIH-3T3 cells by smRNA-FISH according to (Kishi et al., 2019). Briefly, 50-80 probes were designed per gene using PaintSHOP (Hershberg et al., 2021) with additional sequences used for SABER appended to the 3' end in a gene-specific manner (Table S3.5). All probes corresponding to the same gene were pooled at 10uM in 1x TE buffer (10 mM Tris pH 8.0, 0.1 mM EDTA). Probes were synthesized by first preparing a mix containing 10uL of 5uM hairpin oligo, 10uL of 10x PBS, 10uL of 100mM $MgSO_4$, 5uL of dNTP mix (containing 6mM of each A,C,T), 10uL of 1uM Clean.g oligo, 0.5uL of BST enzyme (McLab, BPL-300), and 44.5uL water (see Table S3.5 for oligo sequences). The reaction mix was incubated at 37°C for 15 minutes and then 10uL of 10uM pooled probes were added. Probes were then concatemerized by incubating at 37°C for 60 minutes before enzyme inactivation at 80°C for 20 minutes. Probes were purified using the MinElute PCR Purification kit (Qiagen 28004) and quantified by Nanodrop (ssDNA setting).

Cells were grown in 8-well poly-L-lysine coated chamber slides (ibidi 80826), labeled with 500uM 4sU for 2 hours, and fixed with 4% PFA in 1x PBS for 10 minutes. Unlabeled cells were also included as controls. Slides were washed 3x 5 minutes in 1x PBST and then incubated in 1x Whyb solution (2x SSC, 1% Tween-20, and 40% deionized formamide) for at least 1 hour at 43°C. 1ug of concatemerized probes were then incubated on slides in 1x Hyb solution (2x SSC 1% Tween-20, 40% deionized formamide, 10% dextran sulfate) for at least 16 hours at 43°C (total volume per ibidi chamber well of 150uL). Slides were wrapped in parafilm and placed in a humidifying chamber within the oven to prevent evaporation. After probe hybridization, each well was washed 2x 30 minutes with 1x Whyb solution (pre-warmed to 43°C), followed by 2x 5 minutes in 2x SSC, 0.1% Tween-20 (pre-warmed to 43°C). Slides were then moved to room temperature and washed 2x 1 minute with 1x PBST (1x PBS, 0.1%

Tween). For fluorescent detection of the mRNAs, slides were then transferred to an oven set at 37°C and pre-warmed for 10 minutes. Fluorescent imager oligos were then incubated with the sample at 37°C for 10 minutes, each at 0.2 uM concentration in Imager Hyb (1x PBS, 0.2% Tween). Each well was then washed 3x 5 minutes in 1x PBST at 37°C, before the sample was brought to room temperature.

Samples were then blocked for 1 hour at room temperature in Blocking Solution (1x PBST, 10% Molecular-grade BSA (ThermoFisher AM2616)). Antibodies to detect Lamin B1 and Tubulin were applied in Blocking Solution and incubated for 2 hours at room temperature (see Table S3.5 for antibody details). Samples were washed 3x 5 minutes with 1x PBST before secondary antibody incubation, which were applied for 1 hour at room temperature in Blocking Solution (see Table S3.5 for antibody details). Before imaging, samples were washed 3x 5 minutes with 1x PBST and all samples were imaged in 1x PBST.

Images were acquired and puncta were detected according to (West et al., 2022). Briefly, slides were imaged using a Nikon Ti-2 spinning disk inverted microscope with a 40x objective at the Microscopy Resources on the North Quad (MicRoN) core at Harvard Medical School. Images were acquired as multipoint, multichannel images and data were saved and exported as .nd2 files. Images were then split by channel and position and used to generate maximum projections across the z-stacks and a top-hat background subtraction was applied. The nuclear regions within each position were identified by creating a mask from the Lamin B1 signal, and the cytoplasmic regions were similarly identified using Tubulin signal after subtracting the nuclear mask. Finally, mRNA puncta were identified using a Laplacian of Gaussian filter and were called as nuclear or cytoplasmic based on overlap with the masks.

CELL FRACTIONATION

Chromatin, nuclear and cytoplasm RNA: Cells were fractionated as per (Martell et al., 2021). Briefly, cells were lysed in 400uL cytoplasm lysis buffer (10mM Tris-HCl pH 7.0, 150mM NaCl, 0.15% NP-40) and incubated on ice for 5 minutes. Lysate was then layered on top of 500uL sucrose buffer (25% sucrose, 10mM Tris-HCl, 150mM NaCl) and centrifuged at 13,000 RPM at 4°C for 10 minutes to pellet nuclei. The top (cytoplasm) fraction was isolated. Nuclei were resuspended in 800uL nuclei wash buffer (1x PBS with 1mM EDTA, 0.1% Triton-X) and centrifuged at 3,500 RPM at 4°C for 1 minute. To isolate the nuclear fraction, the washed nuclei were resuspended in 500uL RIPA buffer. To isolate the chromatin fraction, the washed nuclei were resuspended in 200uL glycerol buffer (50% glycerol, 20mM Tris-HCl pH 8.0, 75mM NaCl, 0.5mM EDTA, 0.85mM DTT). After resuspension, 200uL nuclear lysis buffer (20mM HEPES pH 7.5, 300mM NaCl, 1M urea, 0.2mM EDTA, 1mM DTT, 1% NP-40) was added and lysates were incubated on ice for 2 minutes before centrifuging at 14,000 RPM at 4°C for 2 minutes. The supernatant was discarded and chromatin pellets were resuspended in 100uL chromatin resuspension solution. The final volume of the chromatin fraction was brought to 250uL with RIPA buffer.

Polysome RNA: Cells were lysed in 500uL polysome lysis buffer (25mM HEPES pH 7.5, 5mM $MgCl_2$, 0.1M KCl, 2mM DTT, 1% Triton-X, 0.1mg/mL cycloheximide) and incubated on ice for 5 minutes. The lysate was centrifuged at 13,000 RPM at 4°C for 10 minutes to pellet nuclei. The supernatant was then loaded on top of a 12mL 10-50% sucrose gradient (25mM HEPES pH 7.5, 5mM $MgCl_2$, 0.1M KCl, 2mM DTT, 0.1mg/mL cycloheximide) and spun in an ultracentrifuge at 35,000 RPM at 4°C for 2 hours. Gradients were fractionated into 13 samples and the RNA absorbance throughout the gradient was monitored with a BioComp 153 gradient station ip (BioComp Instruments, Fredericton, New Brunswick), using a FC-2 Triax flow cell with software v1.53A (BioComp), and fractionated with a Gilson FC203B fraction collector. Lysate

from the puromycin-sensitive fractions (Figure S2.1D) was then pooled as the polysome fraction.

RNA EXTRACTION

RNA extraction was performed using Trizol LS according to the manufacturer's protocol, except with the addition of DTT at a final concentration of 0.2mM DTT in the isopropanol. For polysome samples, isolated RNA was precipitated using standard ethanol precipitation to reduce the volume of RNA after the Trizol extraction. RNA was quantified using a Nanodrop 2000 (ThermoFisher).

WESTERN BLOTTING

Samples were mixed at 1:1 volume with 2x Laemmli buffer (4% SDS, 20% glycerol, 0.2M DTT, 0.1M Tris-HCl pH 7.0, 0.02% bromophenol blue), denatured at 95°C for 5 minutes, and kept on ice. Samples were loaded onto a 4-12% Bis-Tris gel (Invitrogen NP0321BOX) in 1x MOPs buffer and run at 160V for 1 hour. The gel was transferred to a nitrocellulose membrane using the wet transfer method in 1x transfer buffer (25mM Tris base, 192mM glycine, 20% methanol) at 400mA for 75 minutes at 4°C. The membrane was blocked in 1x blocking buffer (5% non-fat milk powder in 1x TBST) for at least 60 minutes. Primary antibodies were diluted according to (Table S3.5) in 1x blocking buffer and incubated with membranes for at least 16 hours at 4°C. Membranes were washed 4x 5 minutes with 1x TBST, incubated for 1 hour at 25°C with secondary antibodies (Table S3.5), washed again 4x 5 minutes with 1x TBST, and imaged using a Li-Cor Odyssey.

TIMELAPSE-SEQ CHEMISTRY AND LIBRARY PREPARATION

Samples were prepared for sequencing according to (Schofield et al., 2018). Briefly, 2.5ug RNA was treated with 0.1M sodium acetate pH 5.2, 4mM EDTA, 5.2% 2,2,2-trifluoroethylamine, and 10mM sodium periodate at 45°C for 1 hour. RNA was then cleaned using an equal volume of RNAClean XP beads (Beckman Coulter A63987) by washing twice with 80% ethanol. The cleaned RNA was then treated with 10mM Tris-HCl pH 7.5, 10mM DTT, 100mM NaCl, and 1mM EDTA at 37°C for 30 minutes. The RNA clean up with an equal volume of RNAClean XP beads was repeated and RNA was quantified by Nanodrop. Library preparation was performed using the SMARTer Stranded Total RNA HI Mammalian kit (Takara 634873) with 0.5-1ug of RNA and samples were sequenced on the NovaSeq (Illumina, San Diego, CA) by the Bauer Core Facility at Harvard University.

QUANTIFICATION OF NEWLY SYNTHESIZED RNA FROM SUBCELLULAR TIMELAPSE-SEQ DATA

Reads were filtered for quality and adaptor sequences were trimmed using cutadapt v2.5 (Martin, 2011). The first 3nt were trimmed from the 5' of read1, and the last 3nt were trimmed from the 3' of read2, corresponding to the 3nts added by the strand-switching oligo during the reverse transcription step in the library preparation. In order to minimize the background mismatch rate, SNP-masked genomes were prepared starting with hg38 and mm10 using non-4sU total RNA TimeLapse-seq reads from K562 and NIH-3T3, respectively (Figure S2.2C). To prepare the SNP-masked genomes, reads were first mapped to the reference genome with STAR v2.7.3a (Dobin et al., 2013) using parameters --outFilterMultimapNmax 100 --outFilterMismatchNoverLmax 0.09 --outFilterMismatchNmax 15 --outFilterMatchNminOverLread 0.66 --outFilterScoreMinOverLread 0.66 --outFilterMultimapScoreRange 0 --outFilterMismatchNoverReadLmax 1. Variants were then

called with BCFtools mpileup (Li, 2011) and call using two bam files as input. The resulting

variant call file (VCF) was then split into a file with INDEL records only and a file without INDEL

records (substitutions only). The "no INDEL" VCF was further split by frequency of substitution:

loci covered by >= 5 reads and with a variant frequency >75% to a single alternate base were

assigned the alternate base; loci with variants with an ambiguous alternate base were masked

by "N" assignment. The reference FASTA was modified for these non-INDEL substitutions using

GATK FastaAlternateReferenceMaker (McKenna et al., 2010). Finally, rf2m

(https://github.com/LaboratorioBioinformatica/rf2m) was used with the INDEL-only VCF file to

further modify the FASTA genome reference as well as the corresponding GTF annotation file.

The trimmed, filtered reads were aligned to the appropriate SNP-masked genome using

STAR v2.7.0a using the following parameters: --outFilterMismatchNmax 15

--outFilterMismatchNoverReadLmax 0.09 --outFilterScoreMinOverLread 0.66

--outFilterMatchNminOverLread 0.66 --alignEndsType Local --readStrand Forward

--outSAMattributes NM MD NH. Reads that were not mapped in proper pairs, non-primary and

supplementary alignments, and reads aligning to the mitochondrial genome were all discarded

using samtools v1.9 (Li et al., 2009). Samples were then grouped by compartment and replicate

and converted into .cit files using GRAND-SLAM v2.0.5d (Jürges et al., 2018). Samples were

processed through GRAND-SLAM twice, once with the -no4sUpattern option specified and a

second time without this parameter. This ensured that the background T>C mismatch rate (p_E)

was calculated using the -4sU sample during the first run and then the data for the -4sU sample

was outputted during the second run.

The *a priori* unknown 4sU-induced T>C conversion rate (p_C) increased with the cellular

4sU concentration throughout the pulse durations (Figure S2.2B). This resulted in T>C

mismatch distributions that differed between genes according to their rates of turnover, such that

the assumption of a single global p_C for all genes, as in GRAND-SLAM (Jürges et al., 2018),

was no longer sufficient. We therefore estimated upper and lower bounds on the gene-specific

fractions of new RNA, for each sample as follows. The default GRAND-SLAM output was analyzed to select the 1,000 genes with the fastest turnover, i.e. the 1,000 protein-coding genes with the highest MAP values at the lowest non-0 4sU time point, or the 500 genes with the slowest turnover, i.e. the 500 protein-coding genes with the lowest MAP values > 0.2 at the longest 4sU time point, within each compartment and replicate. Genes with high background T>C mismatches, i.e. MAP in the unlabeled sample >0.05, were excluded from consideration. For each sample, all reads aligning to either the fast or slow turnover set of genes were then analyzed to determine the number of T>C mismatches and total number of T nts across each fragment (considering both read1 and read2 in each pair) using custom scripts. Two dimensional distributions were generated containing the number of fragments with n Ts and k T>C mismatches, in each of the fast and slow turnover gene group.

To quantify the upper and lower bounds for the 4sU-induced and background T>C conversion rates, we first developed a binomial mixture model for the background conversions with 2 T>C conversion rates (p_{E1} and p_{E2}) and a global fraction parameter (π_E) of the two populations, with $Binom$ a binomial distribution:

$$P_{BG}(k, n) = \pi_E Binom(k, n, p_{E1}) + (1-\pi_E)Binom(k, n, p_{E2})$$. The three parameters in this model were fitted to the above T>C distributions using linear regression (Python v3.7.4, package lmfit v1.0.2, function minimize). We found that our binomial mixture error model better fitted the T>C conversions of the untreated samples (Akaike Information Criterium (AIC): p = 1.0), when compared to the established approach of a binomial error model with a single background T>C conversion (AIC: p < 1e-16) (Jürges et al., 2018). Because p_{E2} turned out to be of similar magnitude (~2%) as the 4sU-induced T>C conversion rate (p_C, see below), this relatively small second background population (($1-\pi_E$) ~ 2-3%) was essential to include in the model in order to then accurately estimate the 4sU-induced T>C conversion rate (p_C) and global fraction of newly synthesized RNA (π_C). The 4sU sample T>C distributions were then modeled

96

as the 4sU-induced T>C conversions + background population:

$$P(k, n) \ = \ \pi_C \, Binom(k, n, p_C) \ + \ (1 - \pi_C)P_{BG}(k, n)$$

The above fitting procedure was applied to the T>C mismatch distributions of both the fast and slow turnover genes, which generated p_{C_HI} and p_{C_LO} , our respective upper and lower bound estimates on p_C. GRAND-SLAM was run a final two times, once each using p_{C_HI} and p_{C_LO}. For each bound $X \in \{HI, \, LO\}$, this resulted in a gene-specific fraction of new RNA (π) posterior distribution: $P_X^g(\pi) \ = \ Beta(\pi \, ; \alpha_X^g, \beta_X^g)$, a beta distribution characterized with parameters α_X^g and β_X^g. The final gene-specific posterior $P^g(\pi)$ is then as follows:

For the 1,000 genes used to calculate p_{C_HI}: $P^g(\pi) \ = \ P_{HI}^g(\pi)$.

For the 500 genes used to calculate p_{C_LO} (and those with slower turnover): $P^g(\pi) \ = \ P_{LO}^g(\pi)$.

For the other genes, incorporating the uncertainty over p_C, the normalized sum over both bound posteriors: $P^g(\pi) \ = \ \frac{1}{2}\left(P_{LO}^g(\pi) \ + \ P_{HI}^g(\pi)\right)$.

NANOSTRINGS QUANTIFICATION OF NEW RNA

RNA was denatured at 65°C for 5 minutes and immediately incubated on ice for 2 minutes. The biotinylation reaction was performed with 15ug of denatured RNA in 78ul water, 10ul of MTSEA biotin-XX (Biotium 90066) at 0.25mg/mL in dimethylformamide, and 10uL of 10x buffer (100mM Tris pH 7.5, 10mM EDTA) on a thermoblock at 24°C at 800 RPM for 30 minutes. The biotinylated RNA was purified using Phase Lock tubes (QuantaBio 2302830) using standard chloroform/isoamyl alcohol and ethanol precipitation and quantified by Nanodrop. To synthesize the spike-in RNA, ERCC-00048 DNA with an upstream T7 promoter was cloned into pUC19 and PCR amplified (primers in Table S3.5) with Phusion polymerase (New England Biolabs M0530S) using the following cycling conditions: 98°C for 30 seconds, then 35 cycles of

98°C for 10 seconds, 61°C for 15 seconds, and 72°C for 30 seconds, followed by a final extension of 72°C for 2 minutes. The PCR product was cleaned using the Monarch® PCR clean up kit (New England Biolabs T1030S) according to the manufacturer's protocol, quantified by Nanodrop, and used as the template for *in vitro* transcription reaction (New England Biolabs E2040S) performed according to the manufacturer's protocol. RNA was purified using standard ethanol precipitation and quantified by Nanodrop.

3ug of biotinylated RNA and 60pg of *in vitro* transcribed spike-in RNA (ERCC-00048) in 200ul water was mixed with 100ul of beads from the uMACS Streptavidin kit (Miltenyi Biotec 130-074-101) on a thermoblock at 24°C at 800 RPM for 15 minutes. The columns were washed once with 900uL of wash buffer (100mM Tris pH 7.5, 10mM EDTA, 1M NaCl, 0.1% Tween-20). The RNA/bead mixture was then passed through the washed columns twice and the flow-through RNA was collected, purified using the miRNeasy Nano kit (Qiagen 217084) according to (Schwalb et al., 2016), and quantified by Nanodrop. 150ng of RNA was hybridized with gene-specific DNA probes (Table S3.5), XT Tagset-24 capture and target probes (NanoString Technologies, Seattle, WA), and hybridization buffer (NanoString Technologies, Seattle, WA) according to the manufacturer's protocol at 67°C for at least 16 hours before being loaded onto a nCounter Sprint Cartridge and quantified using the nCounter SPRINT Profiler (NanoString Technologies, Seattle, WA) at the Boston Children's Hospital Molecular Genetics Core. The fraction of new RNA was calculated at each time point according to the following equation:

$$Fraction\ new(t)\ = \frac{(Gene\ counts_{t=0}\ /\ spike.in\ counts_{t=0}) - (Gene\ counts_{t}\ /\ spike.in\ counts_{t})}{(Gene\ counts_{t=0}\ /\ spike.in\ counts_{t=0})}$$

Any negative fraction of new RNA values were replaced with 0s. Note that *JUN* and *SHOX2* were included in the probe set but not analyzed due to low RNA counts (within the range of the included manufacturer's negative controls) across most time points. Non-coding transcripts *MALAT1* and *COX1* were also included in the probe set but not analyzed for these experiments.

KINETIC MODELING OF RNA FLOW

The kinetic model (Figure 2.2A), defined by a system of coupled ordinary differential equations (ODEs, Figure S2.4A), describes the average time evolution of the variables, i.e. the 4sU-labeled RNA levels in their respective subcellular compartments at 4sU-pulse time t : T, whole-cell (total); CH, chromatin; N, nucleus; CY, cytoplasm; P, polysome; M, (mature) untranslated cytoplasm. In addition to the RNA flow rates (vector $\vec{k}$, Figure S2.4A), the ODEs also contain the *a priori* unknown RNA production rate k_P. All rates are considered unaffected by 4sU treatment, a model assumption for which we provide evidence (Figure S2.1B-C). To solve these ODEs analytically (Adams and Essex, 2021), we set all labeled RNA levels to zero at t=0, i.e. before any 4sU pulsing (boundary conditions). Next, the integrating factor method was used to obtain solutions for the fractions that only depend on a single RNA flow rate: T(k_{WC}, k_P, t), CH(k_{CH}, k_P, t) and N(k_N, k_P, t). Inserting these expressions then enabled solving the coupled ODEs of CY($\vec{k}$, k_P, t), M($\vec{k}$, k_P, t) and P($\vec{k}$, k_P, t), again using an integrating factor. Since the observed quantities are fraction of new RNA, rather than RNA levels we derived the model fraction of new RNA for each compartment X: $\Lambda_X(\vec{k}, t) = \frac{X(t)}{X_{ss}}$, with X_{ss} levels of all (labeled and unlabeled) compartment RNA, which equals the steady state solution to the ODE, i.e. when $\frac{dX(t)}{dt} = 0$. Both X_{ss} and $X(t)$ are linear in k_P, so $\Lambda_X(\vec{k}, t)$ no longer depends on k_P. Full expressions of the $\Lambda_X(\vec{k}, t)$ solutions for all compartments are available at

https://github.com/churchmanlab/rna_flow/new_total_ratio_jit.py .

As described in section *Quantification of newly synthesized RNA from subcellular TimeLapse-seq data* (Figure S2.2C), the gene-specific (g) fraction of new RNA (π) Posterior

probability density function, $P^g_{X,t}(\pi)$, was experimentally estimated for 4 different 4sU-pulse times (t = 15, 30, 60 and 120 minutes) for each compartment X, through GRAND-SLAM's Bayesian inference framework (Jürges et al., 2018). To analytically derive the (in some cases multivariate) posterior on the RNA flow rates (Adams and Essex, 2021), we equated our model $\Lambda_X(\vec{k};t)$ to π, multiplied the posteriors from all (independent) timepoints, and applied the calculus of multivariate change of variables from π to $\vec{k}$ (see resulting expression in Figure S2.4A).

For the cases with a univariate RNA flow rate posterior (compartment X = T, CH, or N), this immediately provided the posterior distribution on $\vec{k} = k_X$. Summary statistics (MAP, Mean and 95% CIs) of the posterior were then determined as follows. Through numerical evaluation of the posterior on a grid (N=1,000 data points) over the prior rate domain [10^{-4}, 10^4] (unit: min^{-1}), the MAP (Python package numpy v1.16.5, function: argmax) and 95% CIs were determined (see expressions in Figure S2.4A). The mean rate (expression in Figure S3.4A) was obtained through numerical integration (Python package: scipy v1.6.2, function integrate.quad and dblquad). To speed up the numerical integration calculations, integrand functions were coded with "just in time" compilation (Python packages: numba v0.53.1, numba-scipy v0.3.0, function jit). The mean is preferred when using a single number as the RNA flow rate estimate, because it considers all of the posterior distribution. Summary statistics in the results section were reported as the mean between both biological replicates unless noted otherwise. The MAP and 95% CI together provide a more fine-grained characterization.

For the multivariate posteriors (compartment: CY, P), we marginalized the posterior by integrating over the 95% CIs of the already determined upstream RNA flow rate(s) (Figure S2.4A). For example P_{CY} depends on $\vec{k} = [k_N, k_{CY}]$, and the k_N 95% CI was already obtained

through the univariate procedure described above, so k_N was integrated out, resulting in the

posterior on k_{CY}. Then, posterior summary statistics were calculated as described above.

Lastly, the nuclear export rate k_E was determined (Figure S2.4A). Deterministically, the

time duration of nuclear export equals the difference between the nuclear and chromatin

residence times: $k_E^{-1} = k_N^{-1} - k_{CH}^{-1}$. Using its probabilistic analog (Adams and Essex, 2021), the

export posterior is then obtained by numerical integration (as described above) over the

convolution of the nuclear and chromatin posteriors (see resulting expression in Figure S2.4A).

Deterministically, the whole-cell RNA levels equal the sum of nuclear and cytoplasmic

RNA: T(t) = N(t) + CY(t) (Figure S2.4A). In absence of nuclear RNA degradation, the whole-cell

half-life thus equals the sum of nuclear and cytoplasmic half-lives (Figure 2.3H-I). In presence of

nuclear degradation (Table S2.2, see the next section for the PUND identification procedure),

this simple relation between half-lives no longer holds (Figure 2.3H-I).

Furthermore, for PUNDs the above nuclear export rate posterior is not appropriate,

because that calculation assumes no nuclear degradation is occurring. Deterministically, when

including a nuclear degradation rate in the model (Figure 2.2A), the nuclear turnover rate

becomes the sum of nuclear degradation and export rates (Figure S2.4A): $k_N = k_{ND} + k_E^{PUND}$.

This minimal model extension does not consider the chromatin compartment explicitly. However,

in the limit that nuclear export after release from chromatin is much faster than the other rates,

the above expression becomes an exact equation where k_E^{PUND} represents the rate of release

from chromatin (plus nuclear export) if not nuclear degraded, whilst k_{ND} equals the nuclear

degradation rate, which can occur on chromatin and in the nucleoplasm. Indeed, this limit

assumption is consistent with our observations, because for PUNDs the observed chromatin

and nuclear turnover rate distributions are similar (Figure 3.5A). To estimate k_E^{PUND}, we first

derived the deterministic nuclear degradation model whole-cell new RNA levels $T^{PUND}(\vec{k},\ k_p, t)$

in terms of $\vec{k} = [k_N, k_{CY}, k_E^{PUND}]$ (Figure S2.4A), and used our approach above to now derive the

whole-cell multivariate posterior over $\vec{k}$, followed by marginalization over k_N and k_{CY}, to obtain

the k_E^{PUND} posterior. Summary statistics were then determined as described above. Lastly, given

the nuclear posteriors over k_N and k_E^{PUND}, the posterior over $k_{ND} = k_N - k_E^{PUND}$ was then derived

(see expression in Figure S2.4A) and calculated through numerical integration (as described

above).

Least squares estimation (LSE) of the RNA flow rates acted as a simple deterministic

comparison model for the above described Bayesian probabilistic model. LSE was performed by

fitting the model $\Lambda_X(\vec{k}, t)$ to the MAP(π) timecourse values (scipy.optimize.least_squares,

arguments: bounds=[10^{-6},∞], gtol=1e-14, ftol=1e-14, loss='linear'). For the multivariate cases,

the stepwise estimation approach was used again, as described above. For example, for the

cytoplasm, the LSE nuclear turnover rate estimate $\hat{k}_N$ was inserted to enable fitting of $\hat{k}_{CY}$. The

LSE model suffers from the limitation that the uncertainty on the π estimates and upstream RNA

flow rates is not taken into account. For the polysome compartment, this meant that no

reproducible k_{PL} estimates could be obtained with the LSE model, in contrast to our Bayesian

model (Figure S2.5). Because of the same drawback, the LSE, but not the Baysian model, also

predicted a number of biologically unlikely fast rate values (Figure S2.5). Besides these

differences, we generally observed a strong correspondence between our Bayesian MAP and

LSE rate estimates (Figure S2.5). In addition to these "best fit" estimates, provided by both

models, only the Bayesian model provides a full posterior distribution over the rate domain

(Figure 2.2D), and thus also a 95% CI (Figure S2.5), which indicates the range of rate values

consistent with the subcellular Timelapse-seq data.

Since the NanoStrings approach provides single fraction of new RNA values, as opposed to a posterior distribution, NanoStrings RNA flow rate estimation was performed with the LSE model, as described above.

BAYES FACTOR MODEL COMPARISON TO IDENTIFY NUCLEAR RNA DEGRADATION

To perform formal Bayesian model comparison, we calculated the Bayes factor K (Kass and Raftery, 1995), i.e. the ratio of likelihoods of the kinetic model that includes a nuclear degradation rate (alternative hypothesis M_1) over the simpler "nuclear residence" model with no nuclear degradation (null hypothesis M_0): $K = \frac{P(D|M_1)}{P(D|M_0)}$. D indicates the data, in this case subcellular Timelapse-seq used to distinguish the two models: the timeseries of nuclear, cytoplasmic and whole-cell fraction of new RNA posteriors (Figure S2.2C), as described in the above sections. The likelihood for the nuclear residence model for gene g is then:

$$P^g(D|M_0) = \frac{1}{\left(k_{hi}-k_{lo}\right)^2} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \prod_{i=1}^{4} P^g_{N,t_i}(\pi = \Lambda_N(k_N, t_i)) P^g_{CY,t_i}(\pi = \Lambda_{CY}(k_N, k_{CY}, t_i)) \times$$

$$P^g_{T,t_i}(\pi = \Lambda_T(k_N, k_{CY}, t_i)) dk_N dk_{CY}$$

And the equivalent for the nuclear degradation model:

$$P^g(D|M_1) = \frac{1}{\left(k_{hi}-k_{lo}\right)^3} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \prod_{i=1}^{4} P^g_{N,t_i}(\pi = \Lambda_N(k_N, t_i)) P^g_{CY,t_i}(\pi = \Lambda_{CY}(k_N, k_{CY}, t_i)) \times$$

$$P^g_{T,t_i}(\pi = \Lambda_T^{PUND}(k_N, k_{CY}, k_E^{PUND}, t_i)) dk_N dk_{CY} dk_E^{PUND}$$

$k_{lo}=10^{-4}$ and $k_{hi}=10^4$ min^{-1} indicate the prior rate domain bounds. Note that, although the nuclear degradation rate is not explicitly present in the above equation, it is still included in this model since by definition $k_{ND} := k_N - k_E^{PUND}$. We calculated these integrals numerically (as described in the previous section).

When $K > 100$, it is considered "decisive" evidence in favor of the alternative model, and "strong" evidence if K ranges from 10 to 100 (Kass and Raftery, 1995). In our case, genes with transcripts predicted to undergo nuclear RNA degradation (PUNDs) are defined as having $K > 100$ for both biological replicates (Table S2.2). Bayes factors and model likelihoods for all genes and replicates are included in Table S2.1.

Lastly, for each PUND gene, we estimated f_{ND}, the average fraction of transcripts that are nuclear degraded as opposed to exported: $f_{ND} = \frac{k_{ND}}{k_{ND} + k_E^{PUND}}$, using the mean nuclear degradation (k_{ND}) and export (k_E^{PUND}) rates as described in the previous section. With these fractions, we then estimated the total cellular fraction of nuclear degraded protein-coding transcripts as: $f_{ND}^{cell} = \frac{\sum\limits_{PUND\ genes\ g} k_p^g f_{ND}^g}{\sum\limits_{all\ genes\ g} k_p^g}$. k_p^g indicates the gene-specific steady state mRNA production rate (unit: RPKM min^{-1}), as estimated from the chromatin compartment RNA levels (units: RPKM) and our chromatin turnover rates: $k_p^g = RPKM_{CH}^g \times k_{CH}^g$. We also confirmed our conclusions were consistent when using RNA production rate estimates derived from our whole-cell TimeLapse-seq data. Notably, whole-cell estimates suffer from the drawback that whole-cell turnover rates are a convolution of nuclear and cytoplasmic degradation rates in the case of PUNDs, which then underestimate the resulting production rate estimates. Using the chromatin compartment data resolves this bias.

HIERARCHICAL CLUSTERING OF GENES BY RNA FLOW RATES

Hierarchical gene clustering (scipy.hierarchy.linkage, arguments: metric ='seuclidean', method='complete', optimal_ordering=False) was performed on log-transformed half-lives, i.e. $ln(t_{1/2}) = ln\left(\frac{ln(2)}{k}\right)$, with k the MAP and 95% credible interval (CI) endpoints of the turnover

rate posteriors, for two biological replicates and all the subcellular compartments: chromatin, nucleus, cytoplasm, polysome and whole-cell. Genes with missing half-live values were excluded. Other RNA flow rates, i.e. nuclear export and degradation, are derived from these subcellular compartment posteriors, and therefore not used for clustering, but nevertheless included in the heatmap for visualization (Figure 2.4B). For the polysome compartment, only the 95% CIs were used, given that their MAP values were less reproducible across biological replicates (Figure S2.4C-D). To identify gene clusters (scipy.hierarchy.fcluster, arguments: criterion='maxclust'), we allowed sufficient granularity through a total cluster number of 50. To ensure the robustness of our findings (Figure 2.4B), the following downstream analyses were also repeated with a total cluster number of 35 and 100. Next, we filtered out all clusters that comprised of genes with irreproducible patterns, i.e. if the median of the half-lives within a cluster differed at least 4 fold between biological replicates for any of the compartments. This resulted in 27 clusters with reproducible patterns, comprising 10,326 genes. Lastly, we reordered the genes from short to long half-lives, whilst respecting the hierarchical clustering (R version 4.1.1, package pheatmap v1.0.12, function: reorder, arguments: agglo.FUN = mean), after which the heatmap with clustered half-lives was visualized (package pheatmap, function: pheatmap, Figure 2.4B).

FUNCTIONAL ANALYSIS OF GENES CLUSTERED BY RNA FLOW RATES

Gene Ontology enrichment analysis of each human gene cluster was performed in Python (package: GOAtools v1.1.6, object: GOEnrichmentStudyNS, arguments: propagate_counts = True, alpha = 0.05/27, to correct for the multiple testing over all 27 clusters, methods = ['fdr_bh'], Benjamini Hochberg multiple testing correction over the GO terms, and gene_universe = all human genes with any RNA flow rates) (Klopfenstein et al., 2018). For each cluster, the most enriched GO term, i.e. with highest odds ratio (Figure 2.4B),

105

and all significant GO terms are listed (Table S2.3). GO enrichment analysis of the human and mouse PUND genes was performed as for the gene clusters, except with alpha=0.05, and for mouse PUNDs with mouse GO annotations and mouse-specific gene_universe (Figure 2.4A, S3.5A, Table S2.3).

Gene set enrichment analysis (GSEA) (Subramanian et al., 2005) was performed with MSigDB (v7.5.1) gene sets h.all, c5.all, c3.all and c2.all (Liberzon et al., 2015), and parameter settings as in (Ietswaart et al., 2021). Briefly, GSEAPreranked (v4.2.2) was run with ranked, Z-score normalized, log-transformed mean half-lives of each human RNA flow rate type as .rnk input file.

APPENDIX 4 –

METHODS FOR CHAPTER 3

This appendix contains the materials and methods related to the results described in Chapter 3. Author contributions are the same as Chapter 3.

CELL CULTURE

See Appendix 3.

4sU LABELING

See Appendix 3.

CELL FRACTIONATION

See Appendix 3.

RNA EXTRACTION

See Appendix 3.

WESTERN BLOTTING

See Appendix 3.

TIMELAPSE-SEQ CHEMISTRY AND LIBRARY PREPARATION

See Appendix 3.

QUANTIFICATION OF NEWLY SYNTHESIZED RNA FROM SUBCELLULAR TIMELAPSE-SEQ DATA

See Appendix 3.

RNA BINDING PROTEIN ASSOCIATIONS WITH RNA FLOW RATES

K562 eCLIP data for a total of 120 RBPs from (Van Nostrand et al., 2020) was analyzed. The following RNA binding proteins were excluded from the following analyses due to extremely low number of target genes: SLBP, SBDS, UTP3, SUPV3L1, WDR3, PUS1, GNL3, and RPS11. To identify RBPs with significant associations with RNA flow, each replicate of the K562 RNA flow rates were analyzed independently. For each RBP and subcellular half-life, genes were identified as "targets" if the gene contained at least one eCLIP peak with significant enrichment over input. The half-lives of target genes were compared to the half-lives of "non-target" genes (those lacking any significant eCLIP peaks) using a Wilcoxon test with Bonferroni multiple testing correction. The target/non-target half-life was quantified by dividing the median target half-life / median non-target half-life. To identify RBPs containing targets significantly enriched for PUND genes, the targets and non-targets were defined as above. Enrichment was quantified by performing a Fisher's exact test with Bonferroni multiple testing correction.

shRNA KNOCKDOWN OF DDX3X AND PABPC4

K562 knockdown lines were generated according to (Sundararaman et al., 2016) with slight modifications. Plasmid DNA was purified from pLKO.1 backbone vectors expressing shRNAs targeting DDX3X (Horizon Discovery, TRCN0000000003), PABPC4 (Horizon Discovery, TRCN0000074658), and a scrambled control (Addgene 1864). In parallel, psPAX2 (Addgene 12260) and pMD2.G (Addgene 12259) lentiviral plasmid DNA was purified. All plasmid DNA was quantified by Nanodrop. HEK-293T cells growing in 6-well plates at 50% confluency were transfected with 500ng of shRNA-expressing plasmid, 500ng of psPAX2 plasmid, 50ng of pMD2.G plasmid, and 3.1ul FuGENE HD transfection reagent (Promega E2311) in a total volume of 100ul with Opti-MEM I media (ThermoFisher 31985062). Media was discarded after 24 hours and lentiviral-containing media was collected at 48 and 72 hours after

transfection (replacing media every 24 hours) and stored at -80°C. Lentiviral transduction was performed by combining 2×10^6 K562 cells in 1.75mL media, 1.25mL thawed lentiviral-containing media, and 24ug polybrene (Sigma TR-1003-G). Cells were centrifuged at 1,000 RCF at 33°C for 2 hours, the supernatant was discarded, and replaced with 3mL of K562 media. After 24 hours, 3ug/mL puromycin (Sigma P9620) was added and cells were maintained in the presence of the antibiotic at $0.2-1.0 \times 10^6$ cells/mL for 4 days. Knockdowns were confirmed by western blotting analyses.

POLY(A) SELECTION AND DIRECT RNA SEQUENCING

Poly(A)+ selected from 15-30ug of RNA using the Dynabeads mRNA purification kit (ThermoFisher 61006) according to the manufacturer's protocol and quantified by Nanodrop. Synthesis of yeast spike-in RNAs was modeled after the protocol described in https://www.ebi.ac.uk/ena/browser/view/PRJEB28423?show=reads for the *S. cerevisiae ENO2* gene. Briefly, six different *S. cerevisiae* genes (*BDC1, ICT1, HIF1, ENO2, YKE4, HMS2)* were amplified from their genomic locus using HiFi Hotstart DNA polymerase (KAPA) (primers in Table S3.5) in a total volume of 100 uL using the following cycling conditions: 3 minutes at 95°C, then 30 cycles of 15 seconds at 95°C, 15 seconds at 62°C, 2 minutes at 72°C. The PCR amplicons were purified using 1X volume RNA Clean XP beads and eluted in 33 uL water. A second round of PCR was performed with nested primers, wherein the forward primer encodes a T7 RNA polymerase promoter site and the reverse primers have either 10, 15, 30, 60, 80, or 100 thymidines on the 5' end (primers in Table S3.5) using the following cycling conditions: 3 minutes at 95°C, then 18 cycles of 15 seconds at 95°C, 15 seconds at 62°C, 2 minutes at 72°C. The PCR amplicons were purified using 1X volume RNA Clean XP beads and eluted in 33 uL water. In vitro transcription was performed using 500ng of DNA template and the MEGAScript™ T7 Transcription kit (ThermoFisher AM1334) according to the manufacturer's instructions. RNA

was cleaned up with the MEGAClear™ Transcription Clean-up kit (ThermoFisher AM1908) according to the manufacturer's instructions, the concentration was measured by Nanodrop, and the size of the transcripts was verified by TapeStation (Agilent). The six transcripts were pooled at an equimolar concentration (10 picomoles each). 400-700ng of poly(A)+ RNA was combined with 5% spike-in RNA and used to generate direct RNA sequencing libraries with the SQK-RNA002 kit (Oxford Nanopore Technologies) according to the manufacturer's protocol, except for the ligation of the reverse transcription adapter (RTA), which was incubated for 15 minutes instead of 10 minutes. Samples were sequenced on a MinION device (Oxford Nanopore Technologies) with FLO-MIN106D flow cells for up to 72 hours with live basecalling using MinKNOW.

DIRECT RNA SEQUENCING DATA ANALYSIS

All reads with a base calling threshold >7 were converted into DNA sequences by substituting U to T bases. Reads were aligned to the reference human genome (ENSEMBL GRCh38, release-86) concatenated with the six yeast spike-in sequences using minimap2 (version 2.10-r764-dirty) (Li, 2011) with parameters -ax splice -uf -k14. Poly(A) tail lengths were estimated using nanopolish v0.13.3 (Workman et al., 2019). Raw signal fast5 files were indexed with nanopolish index and poly(A) tail lengths were calculated with nanopolish polya using default parameters. Reads with the quality control flag "PASS" and with estimated tail lengths greater than 0 were used in subsequent analyses. To map aligned reads to annotated genes from ENSEMBL GRCh38 (release-86), we used bedtools intersect (Quinlan and Hall, 2010) with options -s -F 0.5 -wo -a $ensembl_bed_file -b $bam, requiring that at least half of the read map to a given gene. For subsequent poly(A) tail length analyses, we filtered for protein-coding genes with at least 10 mapped reads in each sample.

For normalization of poly(A) tail lengths to the spike-ins, we used a median of ratios strategy modeled after the size factor calculation for differential gene expression in DESeq (Love et al., 2014). Poly(A) tail lengths from reads mapping to the yeast spike-in sequences were extracted. For each spike-in, the median poly(A) tail length was calculated in each sample and the geometric mean of medians across samples was computed. The ratio of the median poly(A) tail length per sample over the geometric mean was calculated. Finally, the size factor was defined as the median of ratios across the six spike-ins in each sample. Poly(A) tail lengths from endogenous genes were divided by this size factor for each read, yielding the normalized poly(A) tail length. Of note, the size factors ranged between 0.95 and 1.02 (Figure S3.3B), indicating low technical variability between sequencing runs.

Analysis of RNA 3' ends (deriving from the 50 ends of sequenced reads) was performed as described in (Drexler et al., 2020, 2021). Briefly, "Poly(A)" sites are defined as regions within 50 nucleotides of the end coordinate of annotated protein-coding genes or RNA-PET annotations from cytoplasm and chromatin fractions in K562 ENCODE data (ENCODE Project Consortium, 2012). Determining the splicing status of introns and reads was performed as described in (Drexler et al., 2020, 2021). Code for the analysis of RNA 3' ends and determining the splicing status of introns and reads are available at https://github.com/churchmanlab/nano-COP.

LASSO MACHINE LEARNING MODEL FOR RNA FLOW RATE DETERMINANTS

The objective was to develop a machine learning model that explains a gene-specific RNA flow rate value in terms of that gene's molecular and genetic features (Figure S3.6A). Given the large number of input features (70145, Table S3.3), LASSO regression was chosen as a model because it is a linear model, $y = \beta X$, with L1 regularization, which ensures sparse feature selection (Hastie et al., 2001): $Loss = \frac{1}{2N}||y - \beta X||_2^2 + \alpha||\beta||_1$ (Python, package

scikit-learn v1.0.1, function linear_model.Lasso, arguments: fit_intercept=False, random_state=42, selection='random', max_iter and alpha as specified below). RNA flow rates with genome-wide coverage, i.e. chromatin, nuclear, cytoplasmic, untranslated cytoplasm, and whole-cell turnover and nuclear export, were log-transformed, followed by Z-score normalization, resulting in the model dependent variable (y). Rates from different genes constitute independent data points used for model training (N). The input features originated from various sources and were grouped into classes based on their biological type (Table S3.3). Features classes gene location, histone modifications, microRNA targets, RBP targets, and TF targets correspond to gene sets, i.e. categorical features that were one hot encoded into the LASSO feature matrix (X). All other features were quantitative, and therefore Z-score normalized to facilitate regularization.

To learn the relevant features, and their effect sizes (coefficient vector β), that best explain the rate variation across the genome, we took a two step learning approach. We split the rates into a 90% training set and a 10% test set, with an identical split between biological replicates. Next, we performed the following round 1 LASSO for each feature class separately (Figure S3.6A): (1) Using the features from an individual class, we performed 10x cross validation (CV) twice, once on the training rates from each biological replicate, for a range of values of hyperparameter α : [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}] if the number of features < 1000, or else: [10^{-3}, 10^{-2}, 10^{-1}], and max_iter_ = 2e4. (2) The optimal (round 1) α was then identified, such that the 10x CV R^2 distribution, joined over both replicate runs, was significantly larger than zero in a one-way t-test ($p < 0.05$, scipy.stats.ttest_1samp, arguments: popmean=0, nan_policy='omit', alternative='greater') and larger than any previously selected α in a two-way t-test ($p < 0.05$, scipy.stats.ttest_rel, arguments: nan_policy='omit', alternative='greater'), similar to a selection approach by (Agarwal and Kelley, 2022). If the average performance did not exceed 0 for any α, none of the features were selected for LASSO round 2 from that particular class. (3) Given the

113

optimal α values for each replicate, any individual feature i with a model coefficient $\beta_i > 0$ in both replicate runs, were thus reproducible and (round 1) relevant and thus selected for round 2 learning.

Next, round 2 LASSO was performed: (1) Reproducible round 1 relevant features from all classes were merged into one feature matrix. (2) The training set rates from both biological replicate were joined into the same 10x CV data split to increase the amount of training data. (3) 10x CV was then performed with a fine-grained range for α: [10^{-4}, 3.3×10^{-4}, 6.6×10^{-4}, 10^{-3}, 3.3×10^{-3}, 6.6×10^{-3}, 10^{-2}, 3.3×10^{-2}, 6.6×10^{-2}, 10^{-1}] and max_iter_ = 5e4. (4) The final optimal α was then identified by finding the maximal average 10x CV R^2, such that the average 10x CV prediction R^2 did not exceed the average 10x CV training R^2 by more than 10% (Figure 3.6B, S9B), to avoid overfitting and robust identification of relevant features. (5) Given the optimal α value, any feature i with a (round 2) model coefficient $\beta_i > 0$ was considered a relevant feature (Figure 3.6C, 3.7A, S3.7A, Table S3.3). (6) Lastly, we tested the trained round 2 LASSO model by determining its performance on the 10% unseen test set (Figure 3.6B, S3.6B).

For the "consensus" whole-cell half-lives (Agarwal and Kelley, 2022), the 10x CV and testing was performed as in round 2 LASSO, described above, with reproducible round 1 relevant features from our whole-cell turnover rates (Figure S3.6B).

Continuous averaging plots (Figure 3.7B, S3.7C), were generated as in (Ietswaart et al., 2017), with minor modifications. First, genes were ranked $g = 1 .. N$ according to their gene length from short to long, where N is the total number of genes. This was followed by calculation of the "continuous" averages, $< L_k >$ with $k = 1 .. (2N - 1)$, over these ranked gene subpopulations of their gene length L, i.e. the independent variable: $< L_k > := \frac{1}{k} \sum_{g=1}^{k} L_g$ and

$< L_{N+k} > := \frac{1}{N-k} \sum_{g=1+k}^{N} L_g$. Next, the corresponding continuous median, and Q25 and Q75

(shaded error bands) of the half-lives, i.e. dependent variables, were calculated over the same gene subpopulations. The shortest 1% and longest 1% of genes were excluded from this analysis.

TRANSCRIPTION FACTOR ASSOCIATIONS WITH RNA FLOW RATES

The same transcription factor target gene sets were used as the input features in our LASSO model (Table S3.3). The statistical analysis was performed as for the RBP targets as described above in *RNA binding protein associations with RNA flow rates.*

REFERENCES

Adams, R.A., and Essex, C. (2021). Calculus: A Complete Course (Pearson Education (US)).

Afonina, E., Stauber, R., and Pavlakis, G.N. (1998). The human poly(A)-binding protein 1 shuttles between the nucleus and the cytoplasm. J. Biol. Chem. *273*, 13015–13021.

Agarwal, V., and Kelley, D. (2022). The genetic and biochemical determinants of mRNA degradation rates in mammals.

Alles, J., Legnini, I., Pacelli, M., and Rajewsky, N. (2021). Rapid nuclear deadenylation of mammalian messenger RNA.

Attardi, G., Parnas, H., Hwang, M.I., and Attardi, B. (1966). Giant-size rapidly labeled nuclear ribonucleic acid and cytoplasmic messenger ribonucleic acid in immature duck erythrocytes. J. Mol. Biol. *20*, 145–182.

Baer, B.W., and Kornberg, R.D. (1980). Repeating structure of cytoplasmic poly(A)-ribonucleoprotein. Proc. Natl. Acad. Sci. U. S. A. *77*, 1890–1892.

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. (2015). Nuclear Retention of mRNA in Mammalian Tissues. Cell Rep. *13*, 2653–2662.

Bartel, D.P. (2018). Metazoan MicroRNAs. Cell *173*, 20–51.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of Transcript Variability in Single Mammalian Cells. Cell *163*, 1596–1610.

Benoit Bouvrette, L.P., Cody, N.A.L., Bergalet, J., Lefebvre, F.A., Diot, C., Wang, X., Blanchette, M., and Lécuyer, E. (2018). CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in Drosophila and human cells. RNA *24*, 98–113.

Bergkessel, M., Whitworth, G.B., and Guthrie, C. (2011). Diverse environmental stresses elicit distinct responses at the level of pre-mRNA processing in yeast. RNA *17*, 1461–1478.

Berry, S., and Pelkmans, L. (2022). Mechanisms of cellular mRNA transcript homeostasis. Trends Cell Biol. *32*, 655–668.

Berry, S., Müller, M., Rai, A., and Pelkmans, L. (2022). Feedback from nuclear RNA on transcription promotes robust RNA concentration homeostasis in human cells. Cell Syst *13*, 454–470.

Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell *150*, 279–290.

Blobel, G. (1985). Gene gating: a hypothesis. Proc. Natl. Acad. Sci. U. S. A. *82*, 8527–8529.

Blumberg, A., Zhao, Y., Huang, Y.-F., Dukler, N., Rice, E.J., Chivu, A.G., Krumholz, K., Danko, C.G., and Siepel, A. (2021). Characterizing RNA stability genome-wide through combined

analysis of PRO-seq and RNA-seq data. BMC Biol. *19*, 30.

Borden, K.L.B. (2020). The Nuclear Pore Complex and mRNA Export in Cancer. Cancers *13*, 42.

Bresson, S.M., and Conrad, N.K. (2013). The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. PLoS Genet. *9*, e1003893.

Bresson, S.M., Hunter, O.V., Hunter, A.C., and Conrad, N.K. (2015). Canonical Poly(A) Polymerase Activity Promotes the Decay of a Wide Variety of Mammalian Nuclear RNAs. PLoS Genet. *11*, e1005610.

Brody, Y., Neufeld, N., Bieberstein, N., Causse, S.Z., Böhnlein, E.-M., Neugebauer, K.M., Darzacq, X., and Shav-Tal, Y. (2011). The in vivo kinetics of RNA polymerase II elongation during co-transcriptional splicing. PLoS Biol. *9*, e1000573.

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Mol. Cell *36*, 541–546.

Burger, K., Mühl, B., Kellner, M., Rohrmoser, M., Gruber-Eber, A., Windhager, L., Friedel, C.C., Dölken, L., and Eick, D. (2013). 4-thiouridine inhibits rRNA synthesis and causes a nucleolar stress response. RNA Biol. *10*, 1623–1630.

Burgess, H.M., Richardson, W.A., Anderson, R.C., Salaun, C., Graham, S.V., and Gray, N.K. (2011). Nuclear relocalisation of cytoplasmic poly(A)-binding proteins PABP1 and PABP4 in response to UV irradiation reveals mRNA-dependent export of metazoan PABPs. J. Cell Sci. *124*, 3344–3355.

Carter, M.G., Sharov, A.A., VanBuren, V., Dudekula, D.B., Carmack, C.E., Nelson, C., and Ko, M.S.H. (2005). Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. Genome Biol. *6*, R61.

Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. Elife *7*, e32536.

Chang, H., Lim, J., Ha, M., and Kim, V.N. (2014). TAIL-seq: genome-wide determination of poly(A) tail length and 3' end modifications. Mol. Cell *53*, 1044–1052.

Chang, H., Yeo, J., Kim, J.-G., Kim, H., Lim, J., Lee, M., Kim, H.H., Ohk, J., Jeon, H.-Y., Lee, H., et al. (2018). Terminal Uridylyltransferases Execute Programmed Clearance of Maternal Transcriptome in Vertebrate Embryos. Mol. Cell *70*, 72–82.e7.

Chen, C.Y., and Shyu, A.B. (1995). AU-rich elements: characterization and importance in mRNA degradation. Trends Biochem. Sci. *20*, 465–470.

Chen, T., and van Steensel, B. (2017). Comprehensive analysis of nucleocytoplasmic dynamics of mRNA in Drosophila cells. PLoS Genet. *13*, e1006929.

Chen, C.-Y.A., Ezzeddine, N., and Shyu, A.-B. (2008). Messenger RNA half-life measurements in mammalian cells. Methods Enzymol. *448*, 335–357.

Cheng, H., Dufu, K., Lee, C.-S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5' end of mRNA. Cell *127*, 1389–1400.

Cheng, J., Maier, K.C., Avsec, Ž., Rus, P., and Gagneur, J. (2017). Cis regulatory elements explain most of the mRNA stability variation across genes in yeast. RNA *23*, 1648–1659.

Choquet, K., Koenigs, A., Dülk, S.-L., Smalec, B.M., Rouskin, S., and Churchman, L.S. (2022). Pre-mRNA splicing order is predetermined and maintains splicing fidelity across multi-intronic transcripts.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368–373.

Cleary, M.D., Meiering, C.D., Jan, E., Guymon, R., and Boothroyd, J.C. (2005). Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. Nat. Biotechnol. *23*, 232–237.

Cockman, E., Anderson, P., and Ivanov, P. (2020). TOP mRNPs: Molecular Mechanisms and Principles of Regulation. Biomolecules *10*, 969.

Couvillion, M., Harlen, K.M., Lachance, K.C., Trotta, K.L., Smith, E., Brion, C., Smalec, B.M., and Churchman, L.S. (2022). Transcription elongation is finely tuned by dozens of regulatory factors. Elife *11*, e78944.

Crick, F.H. (1958). On protein synthesis. Symp. Soc. Exp. Biol. *12*, 138–163.

Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M., and Borden, K.L.B. (2005). eIF4E promotes nuclear export of cyclin D1 mRNAs via an element in the 3'UTR. J. Cell Biol. *169*, 245–256.

Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M., and Borden, K.L.B. (2006). eIF4E is a central node of an RNA regulon that governs cellular proliferation. J. Cell Biol. *175*, 415–426.

Culjkovic-Kraljacic, B., Baguet, A., Volpon, L., Amri, A., and Borden, K.L.B. (2012). The oncogene eIF4E reprograms the nuclear pore complex to promote mRNA export and oncogenic transformation. Cell Rep. *2*, 207–215.

Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H.S., Grosveld, F., and Antoniou, M. (1999). Inefficient processing impairs release of RNA from the site of transcription. EMBO J. *18*, 2855–2866.

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

Darnell, J.E., Philipson, L., Wall, R., and Adesnik, M. (1971). Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. Science *174*, 507–510.

Darnell, J.E., Jelinek, W.R., and Molloy, G.R. (1973). Biogenesis of mRNA: genetic regulation in mammalian cells. Science *181*, 1215–1221.

Das, S., Vera, M., Gandin, V., Singer, R.H., and Tutucci, E. (2021). Intracellular mRNA transport and localized translation. Nat. Rev. Mol. Cell Biol. *22*, 483–504.

Davidson, L., Kerr, A., and West, S. (2012). Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. EMBO J. *31*, 2566–2578.

Ditlev, J.A., Case, L.B., and Rosen, M.K. (2018). Who's In and Who's Out-Compositional Control of Biomolecular Condensates. J. Mol. Biol. *430*, 4666–4684.

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

Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., et al. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959–1972.

Dominski, Z., and Marzluff, W.F. (2007). Formation of the 3' end of histone mRNA: getting closer to the end. Gene *396*, 373–390.

Drexler, H.L., Choquet, K., and Churchman, L.S. (2020). Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Mol. Cell *77*, 985–998.e8.

Drexler, H.L., Choquet, K., Merens, H.E., Tang, P.S., Simpson, J.T., and Churchman, L.S. (2021). Revealing nascent RNA processing dynamics with nano-COP. Nat. Protoc. *16*, 1343–1375.

Duffy, E.E., Rutenberg-Schoenberg, M., Stark, C.D., Kitchen, R.R., Gerstein, M.B., and Simon, M.D. (2015). Tracking Distinct RNA Populations Using Efficient and Reversible Covalent Chemistry. Mol. Cell *59*, 858–866.

Edmonds, M., Vaughan, M.H., Jr, and Nakazato, H. (1971). Polyadenylic acid sequences in the heterogeneous nuclear RNA and rapidly-labeled polyribosomal RNA of HeLa cells: possible evidence for a precursor relationship. Proc. Natl. Acad. Sci. U. S. A. *68*, 1336–1340.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., Lin, K.S., McGeary, S.E., Gupta, S., and Bartel, D.P. (2020a). The Dynamics of Cytoplasmic mRNA Metabolism. Mol. Cell *77*, 786–799.e10.

Eisen, T.J., Eichhorn, S.W., Subtelny, A.O., and Bartel, D.P. (2020b). MicroRNAs Cause Accelerated Decay of Short-Tailed Target mRNAs. Mol. Cell *77*, 775–785.e8.

Erhard, F., Baptista, M.A.P., Krammer, T., Hennig, T., Lange, M., Arampatzi, P., Jürges, C.S., Theis, F.J., Saliba, A.-E., and Dölken, L. (2019). scSLAM-seq reveals core features of

transcription dynamics in single cells. Nature *571*, 419–423.

Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of Subcellular RNA Localization Revealed by APEX-Seq. Cell *178*, 473–490.e26.

Frevel, M.A.E., Bakheet, T., Silva, A.M., Hissong, J.G., Khabar, K.S.A., and Williams, B.R.G. (2003). p38 Mitogen-activated protein kinase-dependent and -independent signaling of mRNA stability of AU-rich element-containing transcripts. Mol. Cell. Biol. *23*, 425–436.

Friedel, C.C., Dölken, L., Ruzsics, Z., Koszinowski, U.H., and Zimmer, R. (2009). Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. Nucleic Acids Res. *37*, e115.

Fuchs, G., Voichek, Y., Benjamin, S., Gilad, S., Amit, I., and Oren, M. (2014). 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. Genome Biol. *15*, R69.

Gantier, M.P., McCoy, C.E., Rusinova, I., Saulep, D., Wang, D., Xu, D., Irving, A.T., Behlke, M.A., Hertzog, P.J., Mackay, F., et al. (2011). Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. Nucleic Acids Res. *39*, 5692–5703.

Görlach, M., Burd, C.G., and Dreyfuss, G. (1994). The mRNA poly(A)-binding protein: localization, abundance, and RNA-binding specificity. Exp. Cell Res. *211*, 400–407.

Greenberg, M.E., and Ziff, E.B. (1984). Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. Nature *311*, 433–438.

Gromak, N., West, S., and Proudfoot, N.J. (2006). Pause sites promote transcriptional termination of mammalian RNA polymerase II. Mol. Cell. Biol. *26*, 3986–3996.

Gudipati, R.K., Neil, H., Feuerbach, F., Malabat, C., and Jacquier, A. (2012). The yeast RPL9B gene is regulated by modulation between two modes of transcription termination. EMBO J. *31*, 2427–2437.

Guyette, W.A., Matusik, R.J., and Rosen, J.M. (1979). Prolactin-mediated transcriptional and post-transcriptional control of casein gene expression. Cell *17*, 1013–1023.

Halstead, J.M., Lionnet, T., Wilbertz, J.H., Wippich, F., Ephrussi, A., Singer, R.H., and Chao, J.A. (2015). Translation. An RNA biosensor for imaging the first round of translation from single cells to living animals. Science *347*, 1367–1671.

Harpold, M.M., Wilson, M.C., and Darnell, J.E., Jr (1981). Chinese hamster polyadenylated messenger ribonucleic acid: relationship to non-polyadenylated sequences and relative conservation during messenger ribonucleic acid processing. Mol. Cell. Biol. *1*, 188–198.

Hartenian, E., and Glaunsinger, B.A. (2019). Feedback to the central dogma: cytoplasmic mRNA decay and transcription are interdependent processes. Crit. Rev. Biochem. Mol. Biol. *54*,

385–398.

Hastie, T., Friedman, J., and Tibshirani, R. (2001). The Elements of Statistical Learning (Springer New York).

Hazelbaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. Mol. Cell *49*, 55–66.

Helms, S.R., and Rottman, F.M. (1990). Characterization of an inducible promoter system to investigate decay of stable mRNA molecules. Nucleic Acids Res. *18*, 255–259.

Hershberg, E.A., Camplisson, C.K., Close, J.L., Attar, S., Chern, R., Liu, Y., Akilesh, S., Nicovich, P.R., and Beliveau, B.J. (2021). PaintSHOP enables the interactive design of transcriptome- and genome-scale oligonucleotide FISH experiments. Nat. Methods *18*, 937–944.

Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods *14*, 1198–1204.

Hilleren, P., and Parker, R. (2001). Defects in the mRNA export factors Rat7p, Gle1p, Mex67p, and Rat8p cause hyperadenylation during 3'-end formation of nascent transcripts. RNA *7*, 753–764.

Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H., Sonneveld, S., Verhagen, B.M.P., Boersma, S., Vermeulen, M., and Tanenbaum, M.E. (2019). Single-Molecule Imaging Uncovers Rules Governing Nonsense-Mediated mRNA Decay. Mol. Cell *75*, 324–339.e11.

Hosoda, N., Lejeune, F., and Maquat, L.E. (2006). Evidence that poly(A) binding protein C1 binds nuclear pre-mRNA poly(A) tails. Mol. Cell. Biol. *26*, 3085–3097.

Howe, K.J., Kane, C.M., and Ares, M., Jr (2003). Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. RNA *9*, 993–1006.

Huang, H., Weng, H., Sun, W., Qin, X., Shi, H., Wu, H., Zhao, B.S., Mesquita, A., Liu, C., Yuan, C.L., et al. (2018). Recognition of RNA N6-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. Nat. Cell Biol. *20*, 285–295.

Hüttelmaier, S., Zenklusen, D., Lederer, M., Dictenberg, J., Lorenz, M., Meng, X., Bassell, G.J., Condeelis, J., and Singer, R.H. (2005). Spatial regulation of beta-actin translation by Src-dependent phosphorylation of ZBP1. Nature *438*, 512–515.

Ietswaart, R., Rosa, S., Wu, Z., Dean, C., and Howard, M. (2017). Cell-Size-Dependent Transcription of FLC and Its Antisense Long Non-coding RNA COOLAIR Explain Cell-to-Cell Expression Variation. Cell Syst *4*, 622–635.e9.

Ietswaart, R., Gyori, B.M., Bachman, J.A., Sorger, P.K., and Churchman, L.S. (2021). GeneWalk identifies relevant gene functions for a biological context using network representation learning. Genome Biol. *22*, 55.

Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science *346*, 1257521.

Jensen, T.H., Patricio, K., McCarthy, T., and Rosbash, M. (2001). A block to mRNA nuclear export in S. cerevisiae leads to hyperadenylation of transcripts that accumulate at the site of transcription. Mol. Cell *7*, 887–898.

Johnson, T.R., Rudin, S.D., Blossey, B.K., Ilan, J., and Ilan, J. (1991). Newly synthesized RNA: simultaneous measurement in intact cells of transcription rates and RNA stability of insulin-like growth factor I, actin, and albumin in growth hormone-stimulated hepatocytes. Proc. Natl. Acad. Sci. U. S. A. *88*, 5287–5291.

Jonkers, I., Kwak, H., and Lis, J.T. (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife *3*, e02407.

Jürges, C., Dölken, L., and Erhard, F. (2018). Dissecting newly transcribed and old RNA using GRAND-SLAM. Bioinformatics *34*, i218–i226.

Kadener, S., Cramer, P., Nogués, G., Cazalla, D., de la Mata, M., Fededa, J.P., Werbajh, S.E., Srebrow, A., and Kornblihtt, A.R. (2001). Antagonistic effects of T-Ag and VP16 reveal a role for RNA pol II elongation on alternative splicing. EMBO J. *20*, 5759–5768.

Kadener, S., Fededa, J.P., Rosbash, M., and Kornblihtt, A.R. (2002). Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. Proc. Natl. Acad. Sci. U. S. A. *99*, 8185–8190.

Kanai, Y., Dohmae, N., and Hirokawa, N. (2004). Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuron *43*, 513–525.

Kass, R.E., and Raftery, A.E. (1995). Bayes Factors. J. Am. Stat. Assoc. *90*, 773–795.

Kenzelmann, M., Maertens, S., Hergenhahn, M., Kueffer, S., Hotz-Wagenblatt, A., Li, L., Wang, S., Ittrich, C., Lemberger, T., Arribas, R., et al. (2007). Microarray analysis of newly synthesized RNA in cells and animals. Proc. Natl. Acad. Sci. U. S. A. *104*, 6164–6169.

Kini, H.K., Kong, J., and Liebhaber, S.A. (2014). Cytoplasmic poly(A) binding protein C4 serves a critical role in erythroid differentiation. Mol. Cell. Biol. *34*, 1300–1309.

Kishi, J.Y., Lapan, S.W., Beliveau, B.J., West, E.R., Zhu, A., Sasaki, H.M., Saka, S.K., Wang, Y., Cepko, C.L., and Yin, P. (2019). SABER amplifies FISH: enhanced multiplexed imaging of RNA and DNA in cells and tissues. Nat. Methods *16*, 533–544.

Kislauskis, E.H., Zhu, X., and Singer, R.H. (1997). beta-Actin messenger RNA localization and protein synthesis augment cell motility. J. Cell Biol. *136*, 1263–1270.

Klopfenstein, D.V., Zhang, L., Pedersen, B.S., Ramírez, F., Warwick Vesztrocy, A., Naldi, A., Mungall, C.J., Yunes, J.M., Botvinnik, O., Weigel, M., et al. (2018). GOATOOLS: A Python library for Gene Ontology analyses. Sci. Rep. *8*, 10872.

Kren, B.T., Kumar, N.M., Wang, S.Q., Gilula, N.B., and Steer, C.J. (1993). Differential regulation of multiple gap junction transcripts and proteins during rat liver regeneration. J. Cell Biol. *123*, 707–718.

Krol, J., Busskamp, V., Markiewicz, I., Stadler, M.B., Ribi, S., Richter, J., Duebel, J., Bicker, S., Fehling, H.J., Schübeler, D., et al. (2010). Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. Cell *141*, 618–631.

Krowczynska, A., Yenofsky, R., and Brawerman, G. (1985). Regulation of messenger RNA stability in mouse erythroleukemia cells. J. Mol. Biol. *181*, 231–239.

Lai, M.-C., Lee, Y.-H.W., and Tarn, W.-Y. (2008). The DEAD-box RNA helicase DDX3 associates with export messenger ribonucleoproteins as well as tip-associated protein and participates in translational control. Mol. Biol. Cell *19*, 3847–3858.

Lam, L.T., Pickeral, O.K., Peng, A.C., Rosenwald, A., Hurt, E.M., Giltnane, J.M., Averett, L.M., Zhao, H., Davis, R.E., Sathyamoorthy, M., et al. (2001). Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. Genome Biol. *2*, RESEARCH0041.

Lau, L.F., and Nathans, D. (1985). Identification of a set of genes expressed during the G0/G1 transition of cultured mouse cells. EMBO J. *4*, 3145–3151.

Leppek, K., Das, R., and Barna, M. (2018). Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. Nat. Rev. Mol. Cell Biol. *19*, 158–174.

Levy, S., and Aviv, H. (1976). Quantitation of labeled globin messenger RNA by hybridization with excess complementary DNA covalently bound to cellulose. Biochemistry *15*, 1844–1847.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics *27*, 2987–2993.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst *1*, 417–425.

Lima, S.A., Chipman, L.B., Nicholson, A.L., Chen, Y.-H., Yee, B.A., Yeo, G.W., Coller, J., and Pasquinelli, A.E. (2017). Short poly(A) tails are a conserved feature of highly expressed genes. Nat. Struct. Mol. Biol. *24*, 1057–1063.

Long, R.M., Elliott, D.J., Stutz, F., Rosbash, M., and Singer, R.H. (1995). Spatial consequences of defective processing of specific yeast mRNAs revealed by fluorescent in situ hybridization. RNA *1*, 1071–1078.

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

Martell, D.J., Ietswaart, R., Smalec, B.M., and Churchman, L.S. (2021). Profiling metazoan transcription genome-wide with nucleotide resolution using NET-seq (native elongating transcript sequencing). Protocols.io https://doi.org/10.17504/protocols.io.bpymmpu6.

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

Martin, R.M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. (2013). Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. Cell Rep. *4*, 1144–1155.

Mason, P.B., and Struhl, K. (2005). Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. Mol. Cell *17*, 831–840.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N., and Reed, R. (2005). Recruitment of the human TREX complex to mRNA during splicing. Genes Dev. *19*, 1512–1517.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol. Cell *12*, 525–532.

Matera, A.G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. *15*, 108–121.

Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity. Neuron *92*, 1266–1278.

Mayer, A., and Churchman, L.S. (2017). A Detailed Protocol for Subcellular RNA Sequencing (subRNA-seq). Curr. Protoc. Mol. Biol. *120*, 4.29.1–4.29.18.

Mayer, A., di Iulio, J., Maleri, S., Eser, U., Vierstra, J., Reynolds, A., Sandstrom, R., Stamatoyannopoulos, J.A., and Churchman, L.S. (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. Cell *161*, 541–554.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

Melvin, W.T., Milne, H.B., Slater, A.A., Allen, H.J., and Keir, H.M. (1978). Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. Eur. J. Biochem. *92*, 373–379.

Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., et al. (2016). Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. Mol. Cell *64*, 520–533.

Merz, C., Urlaub, H., Will, C.L., and Lührmann, R. (2007). Protein composition of human mRNPs spliced in vitro and differential requirements for mRNP protein recruitment. RNA *13*, 116–128.

Mishima, Y., and Tomari, Y. (2016). Codon Usage and 3' UTR Length Determine Maternal mRNA Stability in Zebrafish. Mol. Cell *61*, 874–885.

Mor, A., Suliman, S., Ben-Yishay, R., Yunger, S., Brody, Y., and Shav-Tal, Y. (2010). Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells. Nat. Cell Biol. *12*, 543–552.

Munchel, S.E., Shultzaberger, R.K., Takizawa, N., and Weis, K. (2011). Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay. Mol. Biol. Cell *22*, 2787–2795.

Ninomiya, K., Kataoka, N., and Hagiwara, M. (2011). Stress-responsive maturation of Clk1/4 pre-mRNAs promotes phosphorylation of SR splicing factor. J. Cell Biol. *195*, 27–40.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell *161*, 526–540.

Oleynikov, Y., and Singer, R.H. (2003). Real-time visualization of ZBP1 association with beta-actin mRNA during transcription and localization. Curr. Biol. *13*, 199–207.

Otsuka, H., Fukao, A., Funakami, Y., Duncan, K.E., and Fujiwara, T. (2019). Emerging Evidence of Translational Control by AU-Rich Element-Binding Proteins. Front. Genet. *10*, 332.

Pai, A.A., Henriques, T., McCue, K., Burkholder, A., Adelman, K., and Burge, C.B. (2017). The kinetics of pre-mRNA splicing in the Drosophila genome and the influence of gene architecture. Elife *6*, e32537.

Palacio, M., and Taatjes, D.J. (2022). Merging Established Mechanisms with New Insights: Condensates, Hubs, and the Regulation of RNA Polymerase II Transcription. J. Mol. Biol. *434*, 167216.

Palazzo, A.F., and Lee, E.S. (2018). Sequence Determinants for Nuclear Retention and Cytoplasmic Export of mRNAs and lncRNAs. Front. Genet. *9*, 440.

Pan, F., Hüttelmaier, S., Singer, R.H., and Gu, W. (2007). ZBP2 facilitates binding of ZBP1 to beta-actin mRNA during transcription. Mol. Cell. Biol. *27*, 8340–8351.

Pandya-Jones, A., Bhatt, D.M., Lin, C.-H., Tong, A.-J., Smale, S.T., and Black, D.L. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid

A-induced gene expression. RNA *19*, 811–827.

Parsyan, A., Shahbazian, D., Martineau, Y., Petroulakis, E., Alain, T., Larsson, O., Mathonnet, G., Tettweiler, G., Hellen, C.U., Pestova, T.V., et al. (2009). The helicase protein DHX29 promotes translation initiation, cell proliferation, and tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. *106*, 22217–22222.

Passmore, L.A., and Coller, J. (2022). Roles of mRNA poly(A) tails in regulation of eukaryotic gene expression. Nat. Rev. Mol. Cell Biol. *23*, 93–106.

Pendleton, K.E., Park, S.-K., Hunter, O.V., Bresson, S.M., and Conrad, N.K. (2018). Balance between MAT2A intron detention and splicing is determined cotranscriptionally. RNA *24*, 778–786.

Penman, S., Scherrer, K., Becker, Y., and Darnell, J.E. (1963). Polyribosomes in normal and poliovirus-infected HeLa cells and their relationship to messenger-RNA. Proc. Natl. Acad. Sci. U. S. A. *49*, 654–662.

Penman, S., Vesco, C., and Penman, M. (1968). Localization and kinetics of formation of nuclear heterodisperse RNA, cytoplasmic heterodisperse RNA and polyribosome-associated messenger RNA in HeLa cells. J. Mol. Biol. *34*, 49–60.

Pinto, P.A.B., Henriques, T., Freitas, M.O., Martins, T., Domingues, R.G., Wyrzykowska, P.S., Coelho, P.A., Carmo, A.M., Sunkel, C.E., Proudfoot, N.J., et al. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. EMBO J. *30*, 2431–2444.

Pisareva, V.P., Pisarev, A.V., Komar, A.A., Hellen, C.U.T., and Pestova, T.V. (2008). Translation initiation on mammalian mRNAs with structured 5'UTRs requires DExH-box protein DHX29. Cell *135*, 1237–1250.

Plant Kathryn E., Dye Michael J., Lafaille Celina, and Proudfoot Nick J. (2005). Strong Polyadenylation and Weak Pausing Combine To Cause Efficient Termination of Transcription in the Human Gγ-Globin Gene. Mol. Cell. Biol. *25*, 3276–3285.

Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. Cell *123*, 249–263.

Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon optimality is a major determinant of mRNA stability. Cell *160*, 1111–1124.

Proudfoot, N.J. (2016). Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science *352*, aad9926.

Puckett, L., Chambers, S., and Darnell, J.E. (1975). Short-lived messenger RNA in HeLa cells and its impace on the kinetics of accumulation of cytoplasmic polyadenylate. Proc. Natl. Acad. Sci. U. S. A. *72*, 389–393.

Qu, X., Lykke-Andersen, S., Nasser, T., Saguez, C., Bertrand, E., Jensen, T.H., and Moore, C. (2009). Assembly of an export-competent mRNP is needed for efficient release of the 3'-end processing complex after polyadenylation. Mol. Cell. Biol. *29*, 5327–5338.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., et al. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat. Biotechnol. *29*, 436–442.

Rabani, M., Raychowdhury, R., Jovanovic, M., Rooney, M., Stumpo, D.J., Pauli, A., Hacohen, N., Schier, A.F., Blackshear, P.J., Friedman, N., et al. (2014). High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. Cell *159*, 1698–1710.

Raghavan, A., Ogilvie, R.L., Reilly, C., Abelson, M.L., Raghavan, S., Vasdewani, J., Krathwohl, M., and Bohjanen, P.R. (2002). Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. Nucleic Acids Res. *30*, 5529–5538.

Reichholf, B., Herzog, V.A., Fasching, N., Manzenreither, R.A., Sowemimo, I., and Ameres, S.L. (2019). Time-Resolved Small RNA Sequencing Unravels the Molecular Principles of MicroRNA Homeostasis. Mol. Cell *75*, 756–768.e7.

Reimer, K.A., Mimoso, C.A., Adelman, K., and Neugebauer, K.M. (2021). Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis. Mol. Cell *81*, 998–1012.e7.

Riml, C., Amort, T., Rieder, D., Gasser, C., Lusser, A., and Micura, R. (2017). Osmium-Mediated Transformation of 4-Thiouridine to Cytidine as Key To Study RNA Dynamics by Sequencing. Angew. Chem. Int. Ed Engl. *56*, 13479–13483.

Rissland, O.S., Hong, S.-J., and Bartel, D.P. (2011). MicroRNA destabilization enables dynamic regulation of the miR-16 family in response to cell-cycle changes. Mol. Cell *43*, 993–1004.

Rohner, S., Kalck, V., Wang, X., Ikegami, K., Lieb, J.D., Gasser, S.M., and Meister, P. (2013). Promoter- and RNA polymerase II-dependent hsp-16 gene association with nuclear pores in Caenorhabditis elegans. J. Cell Biol. *200*, 589–604.

Ross, A.F., Oleynikov, Y., Kislauskis, E.H., Taneja, K.L., and Singer, R.H. (1997). Characterization of a beta-actin mRNA zipcode-binding protein. Mol. Cell. Biol. *17*, 2158–2165.

Roy, B., and Jacobson, A. (2013). The intimate relationships of mRNA decay and translation. Trends Genet. *29*, 691–699.

Ryu, W.S., and Mertz, J.E. (1989). Simian virus 40 late transcripts lacking excisable intervening sequences are defective in both stability in the nucleus and transport to the cytoplasm. J. Virol. *63*, 4386–4394.

Samir, P., Kesavardhana, S., Patmore, D.M., Gingras, S., Malireddi, R.K.S., Karki, R., Guy, C.S., Briard, B., Place, D.E., Bhattacharya, A., et al. (2019). DDX3X acts as a live-or-die checkpoint in stressed cells by regulating NLRP3 inflammasome. Nature *573*, 590–594.

Scherrer, K., and Darnell, J.E. (1962). Sedimentation characteristics of rapidly labelled RNA from HeLa cells. Biochem. Biophys. Res. Commun. *7*, 486–490.

Schmid, M., and Jensen, T.H. (2018). Controlling nuclear RNA levels. Nat. Rev. Genet. *19*, 518–529.

Schofield, J.A., Duffy, E.E., Kiefer, L., Sullivan, M.C., and Simon, M.D. (2018). TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. Nat. Methods *15*, 221–225.

Scholz, B.A., Sumida, N., de Lima, C.D.M., Chachoua, I., Martino, M., Tzelepis, I., Nikoshkov, A., Zhao, H., Mehmood, R., Sifakis, E.G., et al. (2019). WNT signaling and AHCTF1 promote oncogenic MYC expression through super-enhancer-mediated gene gating. Nat. Genet. *51*, 1723–1731.

Schott, J., Reiter, S., Lindner, D., Grosser, J., Bruer, M., Shenoy, A., Geiger, T., Mathes, A., Dobreva, G., and Stoecklin, G. (2021). Nascent Ribo-Seq measures ribosomal loading time and reveals kinetic impact on ribosome density. Nat. Methods *18*, 1068–1074.

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

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Schwartz, R.J. (1973). Control of glutamine synthetase synthesis in the embryonic chick neural retina. A caution in the use of actinomycin D. J. Biol. Chem. *248*, 6426–6435.

Sharova, L.V., Sharov, A. a., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S.H. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res. *16*, 45–58.

Shav-Tal, Y., Darzacq, X., Shenoy, S.M., Fusco, D., Janicki, S.M., Spector, D.L., and Singer, R.H. (2004). Dynamics of single mRNPs in nuclei of living cells. Science *304*, 1797–1800.

Sheiness, D., and Darnell, J.E. (1973). Polyadenylic acid segment in mRNA becomes shorter with age. Nat. New Biol. *241*, 265–268.

Shestakova, E.A., Singer, R.H., and Condeelis, J. (2001). The physiological significance of beta-actin mRNA localization in determining cell polarity and directional motility. Proc. Natl. Acad. Sci. U. S. A. *98*, 7045–7050.

Shih, J.-W., Wang, W.-T., Tsai, T.-Y., Kuo, C.-Y., Li, H.-K., and Wu Lee, Y.-H. (2012). Critical

roles of RNA helicase DDX3 and its interactions with eIF4E/PABP1 in stress granule assembly and stress response. Biochem. J *441*, 119–129.

Shyu, A.B., Greenberg, M.E., and Belasco, J.G. (1989). The c-fos transcript is targeted for rapid decay by two distinct mRNA degradation pathways. Genes Dev. *3*, 60–72.

Singer, R.H., and Penman, S. (1972). Stability of HeLa cell mRNA in actinomycin. Nature *240*, 100–102.

Singer, R.H., and Penman, S. (1973). Messenger RNA in HeLa cells: kinetics of formation and decay. J. Mol. Biol. *78*, 321–334.

Soeiro, R., Birnboim, H.C., and Darnell, J.E. (1966). Rapidly labeled HeLa cell nuclear RNA. II. Base composition and cellular localization of a heterogeneous RNA fraction. J. Mol. Biol. *19*, 362–372.

Soto-Rifo, R., Rubilar, P.S., Limousin, T., de Breyne, S., Décimo, D., and Ohlmann, T. (2012). DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs. EMBO J. *31*, 3745–3756.

Spies, N., Burge, C.B., and Bartel, D.P. (2013). 3′ UTR-isoform choice has limited influence on the stability and translational efficiency of most mRNAs in mouse fibroblasts. Genome Research *23*, 2078–2090.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.

Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. Nature *508*, 66–71.

Sundararaman, B., Zhan, L., Blue, S.M., Stanton, R., Elkins, K., Olson, S., Wei, X., Van Nostrand, E.L., Pratt, G.A., Huelga, S.C., et al. (2016). Resources for the Comprehensive Discovery of Functional RNA Elements. Mol. Cell *61*, 903–913.

Taliaferro, J.M., Vidaki, M., Oliveira, R., Olson, S., Zhan, L., Saxena, T., Wang, E.T., Graveley, B.R., Gertler, F.B., Swanson, M.S., et al. (2016). Distal Alternative Last Exons Localize mRNAs to Neural Projections. Mol. Cell *61*, 821–833.

Trembley, J.H., Kren, B.T., and Steer, C.J. (1994). Posttranscriptional regulation of cyclin B messenger RNA expression in the regenerating rat liver. Cell Growth Differ. *5*, 99–108.

Valencia, P., Dias, A.P., and Reed, R. (2008). Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. *105*, 3386–3391.

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Chen, J.-Y., Cody, N.A.L., Dominguez, D., et al. (2020). A large-scale binding and functional map of human

RNA-binding proteins. Nature *583*, 711–719.

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

Wachutka, L., Caizzi, L., Gagneur, J., and Cramer, P. (2019). Global donor and acceptor splicing site kinetics in human cells. Elife *8*, e45056.

Wahle, E. (1991). A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. Cell *66*, 759–768.

Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro, G., Chow, C.C., Hafner, M., et al. (2021). Dynamic Imaging of Nascent RNA Reveals General Principles of Transcription Dynamics And Stochastic Splice Site Selection. Cell *184*, 2878–2895.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol. Cell *53*, 819–830.

West, E.R., Lapan, S.W., Lee, C., Kajderowicz, K.M., Li, X., and Cepko, C.L. (2022). Spatiotemporal patterns of neuronal subtype genesis suggest hierarchical development of retinal diversity. Cell Rep. *38*, 110191.

Whittaker, K.L., Ding, D., Fisher, W.W., and Lipshitz, H.D. (1999). Different 3' untranslated regions target alternatively processed hu-li tai shao (hts) transcripts to distinct cytoplasmic locations during Drosophila oogenesis. J. Cell Sci. *112*, 3385–3398.

Workman, R.E., Tang, A.D., Tang, P.S., Jain, M., Tyson, J.R., Razaghi, R., Zuzarte, P.C., Gilpatrick, T., Payne, A., Quick, J., et al. (2019). Nanopore native RNA sequencing of a human poly(A) transcriptome. Nat. Methods *16*, 1297–1305.

Wu, Q., Medina, S.G., Kushawah, G., DeVore, M.L., Castellano, L.A., Hand, J.M., Wright, M., and Bazzini, A.A. (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. Elife *8*, e45396.

Wu, Z., Ietswaart, R., Liu, F., Yang, H., Howard, M., and Dean, C. (2016). Quantitative regulation of FLC via coordinated transcriptional initiation and elongation. Proc. Natl. Acad. Sci. U. S. A. *113*, 218–223.

Xiang, K., and Bartel, D.P. (2021). The molecular basis of coupling between poly(A)-tail length and translational efficiency. Elife *10*, e66493.

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

Yedavalli, V.S.R.K., Neuveut, C., Chi, Y.-H., Kleiman, L., and Jeang, K.-T. (2004). Requirement

of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. Cell *119*, 381–392.

Yeom, K.-H., and Damianov, A. (2017). Methods for Extraction of RNA, Proteins, or Protein Complexes from Subcellular Compartments of Eukaryotic Cells. Methods Mol. Biol. *1648*, 155–167.

Yeom, K.-H., Pan, Z., Lin, C.-H., Lim, H.Y., Xiao, W., Xing, Y., and Black, D.L. (2021). Tracking pre-mRNA maturation across subcellular compartments identifies developmental gene regulation through intron retention and nuclear anchoring. Genome Res. *31*, 1106–1119.

Zhao, B.S., Roundtree, I.A., and He, C. (2017). Post-transcriptional gene regulation by mRNA modifications. Nat. Rev. Mol. Cell Biol. *18*, 31–42.

Zheng, C.L., Fu, X.-D., and Gribskov, M. (2005). Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. RNA *11*, 1777–1787.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R., and Qian, S.-B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594.

Zid, B.M., and O'Shea, E.K. (2014). Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature *514*, 117–121.