Title

Investigating the role of DDX3X in regulating mRNA export

Author

Xu, Albert

Publication Date

2023

Investigating the role of DDX3X in regulating mRNA export

by
Albert Xu

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Aimee Kao
745CF57D40DB494...
Aimee Kao
Chair

DocuSigned by:
Stephen Floor
3709C1064F93461...
Stephen Floor

DocuSigned by:
Martin Kampmann
03BA4012CC044AF...
Martin Kampmann

Committee Members

Dedications and Acknowledgements

I would like to acknowledge and sincerely thank all of those who have helped me along this long and still unfinished path. In the laboratory, I would like to thank first my PI, Stephen Floor. I am forever grateful for his support, his encouragement, and not just being a great scientific mentor but being a genuine and caring person as well. I would also like to thank everyone in the Floor lab for nurturing a truly kind, helpful, genuine, and drama-free environment that allowed us to focus on the science and our careers. Special shout-out to my graduate student cohort and fellow soon-to-be PhDs, Ziad Jowhar, José Liboy Lugo, and Jesslyn Park. Ziad has been not just an amazing collaborator but also an amazing friend, and I’m so grateful to have had a fellow MSTP student in my lab and bay to share the good times and the bad. Along with José and Jesslyn, I am grateful for the 3 of you for your continuous support, for always pushing each other to do the best science but also to graduate on time, and I’m so happy that we’re all going to graduate together (or at least within 6 months of each other). Also a big shout-out to the other graduate students in the lab: our senior graduate student and now post-doc Kevin Wilkins and our junior graduate students Margaret Gadek and Abby Hein for nurturing a fun and supportive environment. Also, huge thanks to Yizhu Lin for being a superb post-doc mentor in all things computational. I’ve learned so much from you over the years and for that I am forever grateful.

I would also like to thank my thesis committee members, Aimee Kao and Martin Kampmann for their constant support, mentorship, and guidance. I could not have asked for better committee members. I would also like to thank my past mentors – Kunxin Luo and Qingwei Zhu from UC Berkeley for introducing me to science and for their continuous support, Jonathan Weissman and Britt Adamson for an amazing year as a technician where I learned so much and all the great friends and mentors I met in the Weissman lab, and Luke Gilbert and Becky Fu for their continued mentorship in all things CRISPR even 4 long years after my lab rotation.

Finally, I would like to thank my family – my amazing and supportive wife Leslie and our amazing and supportive dog Poppy. I would also like to thank my parents as well as my in-laws for their continuous support. I am happy to tell them there's finally an answer to the question "When are you graduating?". No words can express how grateful I am for their support, and I'm so lucky to such a loving and supportive family.

Contributions

This dissertation includes contributions from Brendan M. Smalec, Robert Ietswaart, Karine Choquet, Erik McShane, Ziad Mohamoud Jowhar, Emma R. West, Becky Xu Hua Fu, Luke Gilbert, Stephen N. Floor, L. Stirling Churchman.

Abstract

"Investigating the role of DDX3X in regulating mRNA export"

Albert Xu

Mutations in *DDX3X* are associated with autism spectrum disorder, brain malformations, and epilepsy, and account for up to 3% of cases of females with unexplained intellectual disability. However, little is currently known about the molecular mechanism linking *DDX3X* mutations to neurodevelopmental disease. *DDX3X* encodes an RNA helicase of the DEAD-box protein family and has been implicated in many aspects of RNA metabolism, yet we still lack a mechanistic understanding of *DDX3X*'s role in these processes, as well as how patient mutations in *DDX3X* affect RNA metabolism.

To address this gap in knowledge of understanding of how *DDX3X* mutations perturb cellular function and contribute to neurodevelopmental disease, this study investigated the mechanism of how *DDX3X* regulates its target transcripts at the level of mRNA nuclear export, and how pathogenic mutations in *DDX3X* alter these processes. In addition to DDX3X's role in translation, *DDX3X* has been implicated in nuclear mRNA processing, and our preliminary data show that the genes in the mRNA export pathway are genetic interaction partners of *DDX3X*. However, we still do not understand what the function of *DDX3X* is in the nucleus or mRNA export. This study aimed to gain a greater understanding the precise mechanism of how *DDX3X* regulates RNA metabolism both at the level of translation initiation as well as mRNA export. This knowledge is critical for advancing our knowledge of RNA metabolism, as well as understanding and developing treatments for patients with *DDX3X* syndrome.

Table of Contents

Introduction..1

Background..3

Results..7

Discussion 30

Materials and Methods...37

Figures...77

Tables 110

References 148

List of Figures

Figure 1: Subcellular TimeLapse-seq and kinetic modeling estimate genome-wide RNA flow rates. 77

Figure 2: PUND transcripts are degraded by the nuclear exosome and transcripts with similar RNA flow are functionally related. 79

Figure 3: The targets of many RNA binding proteins (RBPs) exhibit distinctive RNA flow rates across the cell. 81

Figure 4: Genetic interaction mapping to link DDX3X depletion to mRNA export defects. 83

Figure 5: Subcellular compartment-specific poly(A) tail lengths reflect RNA flow rates 86

Figure 6: Genes predicted to undergo nuclear degradation (PUNDs) exhibit unique phenotypes related to RNA flow, splicing, and poly(A) tail lengths. 88

Figure 7: LASSO regression model identifies most relevant genetic and molecular features that predict RNA flow rates. 89

Supplemental Figure 1: Optimization of biochemical fractionation and 4-thiouridine labeling conditions used in subcellular TimeLapse-seq, related to Figure 1. 91

Supplemental Figure 2: Nucleotide conversion analysis estimates the fraction of new RNA in a compartment- and time point-specific manner, related to Figure 1. 93

Supplemental Figure 3: RNA flow can be modeled using a series of ordinary differential equations in a Bayesian framework, related to Figure 1 95

Supplemental Figure 4: RNA flow rates show consistent genome-wide patterns between cell lines, related to Figure 1. 97

Supplemental Figure 5: PUND stabilization upon nuclear exosome depletion and functionally related genes exhibit similar RNA flow, related to Figure 2. 99

Supplemental Figure 6: RNA binding proteins are associated with RNA flow rates, related to Figure 3. 101

Supplemental Figure 7: Genetic interaction mapping to link DDX3X depletion to mRNA export defects, related to Figure 4. 103

Supplemental Figure 8: Poly(A) tail lengths are related to RNA flow rates, related to Figure 5. 105

Supplemental Figure 9: PUND phenotypes, related to Figure 6. 107

Supplemental Figure 10: LASSO model predictions, related to Figure 7. 108

List of Tables

Table 1: Primers and Protospacers 110

Table 2: Reagents 145

INTRODUCTION

This thesis work was aimed to understand the molecular mechanism linking *DDX3X* mutations to neurodevelopmental disease. *DDX3X* encodes an RNA helicase of the DEAD-box protein family and has been implicated in many aspects of RNA metabolism, yet we still lack a mechanistic understanding of *DDX3X*'s role in these processes, as well as how patient mutations in *DDX3X* affect RNA metabolism. During the course of this work, we and the Churchman lab at Harvard Medical School independently came to the same conclusion: that DDX3X was in fact important in controlling nuclear export, which was in contrast to its well characterized role in controlling translation initiation. Given our complementary findings, we decided to combine their work developing a method for measuring RNA turnover with my thesis project using unbiased genetic screens into understanding the molecular role of DDX3X into a combined manuscript currently being prepared for publication.

Dissecting the myriad regulatory mechanisms controlling eukaryotic transcripts from production to degradation requires quantitative measurements of mRNA flow across the cell. We developed subcellular TimeLapse-seq to measure the rates at which RNAs are released from chromatin, exported from the nucleus, loaded onto polysomes, and degraded within the nucleus and cytoplasm. These rates varied substantially, yet transcripts from genes with related functions or targeted by the same transcription factors and RNA binding proteins flowed across subcellular compartments with similar kinetics. Verifying these associations uncovered roles for DDX3X and PABPC4 in nuclear export. For 10% of transcribed genes, most transcripts were degraded within the nucleus by the exosome, while the remaining molecules were exported and persisted with stable lifespans. Transcripts residing on chromatin for longer had extended poly(A) tails, whereas the reverse was observed for cytoplasmic mRNAs. Finally, a machine learning model identified additional molecular features that predict the diverse life cycles of mammalian mRNAs.

BACKGROUND

The life cycles of mRNAs are dynamic and diverse. Thousands of mRNAs are produced per minute in a typical mammalian cell [1,2]. 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. Thus, RNA flow impacts cell function by determining the dynamic pool of mRNAs available for translation.

On chromatin, mRNAs are synthesized by RNA polymerase II and undergo extensive processing, including splicing and polyadenylation. The time required to splice introns from pre-mRNAs varies significantly, ranging from seconds to tens of minutes [3–9]. In some cases, splicing regulates the nuclear dynamics of nascent RNA [10–13]. In the nucleus, transcripts are subjected to either degradation or export. Nuclear degradation targets improperly processed mRNAs [14–17] but also serves additional regulatory roles for specific transcripts [18]. By contrast, some loci are tethered near the nuclear pore complex to promote rapid export of transcripts [19–21].

In the cytoplasm, ribosomes are loaded onto transcripts with different kinetics, partially influenced by 5'UTR length and structure [22–26], and promoter elements [27]. Finally, mRNAs undergo degradation, a process that can be driven by both poly(A) tail deadenylation and targeting by microRNAs [28–30]. Each of these processes vary in duration across genes and impact the subcellular fates of transcripts, either directly or indirectly through feedback loops. In total, RNA half-lives vary greater than 100-fold between different protein-coding transcripts [2,31–35]. 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.

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 [36]. 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 [37–40]. 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 [41] and allowing for the quantification of the entire life cycle of an individual transcript in *Arabidopsis* [42,43]. To determine the rates at which RNAs flow across compartments with higher throughput, induced genes following a stimulus have been monitored over time and modeled to estimate subcellular turnover [8,44–46], 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 [47–50], but have not comprehensively

characterized the entire life cycle of an mRNA across multiple compartments in mammalian cells, including 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 ~10% of genes, mRNA flow was predicted to involve substantial nuclear degradation. We found that RNA flow rates varied widely (>100-fold) between different genes and subcellular compartments. Our results demonstrate that functionally related genes undergo similar rates of RNA flow. The targets of many RNA binding proteins (RBPs) exhibit different RNA flow rates compared to other genes. We found for DDX3X and PABPC4 that these differences dissipated upon RBP depletion. Further investigation revealed a strong role for DDX3X in regulating export of its binding targets and genetic interactions between nuclear mRNA export factors and DDX3X essentiality. 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 provide a comprehensive characterization of the dynamics of an mRNA throughout its life cycle within a mammalian cell.

RESULTS

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

To quantity RNA turnover genome-wide with subcellular resolution, we developed subcellular Timelapse-seq, a method that combined metabolic labeling with the biochemical purification of distinct RNA populations from different cellular compartments (**Figure 1A**). To generate the samples for TimeLapse-seq, we pulse labeled human K562 and mouse NIH-3T3 cells with 4-thiouridine (4sU) for 0, 15, 30, 60, and 120 minutes. Because high concentrations of 4sU can broadly impact gene expression [51], we minimized the concentration and duration of 4sU exposure and confirmed that the addition of 4sU did not affect mRNA subcellular localization (**Figure S1A-B**) and led to minimal disruption of gene expression levels within each subcellular compartment (**Figure S1C**). Following each 4sU pulse, we biochemically purified chromatin-associated, nuclear, and cytoplasmic RNA (**Figure 1B**) [52], as well as polysome-bound RNA (**Figure S1D**). We also collected total cellular RNA. We then quantified the amount of newly synthesized RNA in each sample by performing TimeLapse-seq, a nucleotide conversion protocol that detects 4sU-labeled RNA within a mixture of labeled and unlabeled RNAs based on the presence of 4sU-induced T>C sequencing mismatches (**Figure 1A**) [2].

Previous applications of nucleotide conversion approaches have used labeling conditions that achieved high 4sU incorporation rates, aiding the identification of labeled

RNAs [2,33,53,54]. Therefore, the lower 4sU incorporation rates inherent to our minimal labeling conditions necessitated a new approach to estimate the fraction of newly synthesized RNA (**Figure S2A-B**). We developed a binomial mixture model to estimate upper and lower bounds on the T>C conversion rates for each 4sU pulse time and compartment, which were robust across read counts (**Figure S2C**). We then inputted these conversion rates into GRAND-SLAM [55] and combined the outputs to quantify the posterior distribution on the fraction of newly synthesized RNA per gene within each sample (**Figure S2B**). We validated our approach with a NanoStrings-based assay that does not rely on sequencing and predictions of T>C conversions (**Figure S2D**). With this technique, we observed fractions of new nuclear, cytoplasm, and total RNA similar to those estimated by subcellular TimeLapse-seq for select genes with both fast and slow turnover (**Figure S2E-F**), validating the robustness of our analysis pipeline across a range of mismatch rates.

Genome-wide, we observed an increase in the proportion of new RNA with increasing pulse durations within each compartment (**Figure 1C-D**). Furthermore, at each time point, we saw delays in the fraction of new RNA across chromatin, nuclear, cytoplasm, and polysome fractions (**Figure 1C-D**), even for genes with relatively fast turnover (e.g., *Myc*, **Figure 1C**). This result confirmed that our assay has the time resolution suitable for detecting RNA flow across subcellular compartments.

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 1E**). By coupling this model to the Bayesian inference framework of GRAND-SLAM (**Figure S3A**), we estimated the Bayesian posterior probability distribution of each flow rate per gene (**Figure 1F-G**). Using the posterior mean rate (*k*), half-lives are then calculated as t1/2= 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, corresponding to how quickly RNAs exit the nucleus after release from chromatin into the nucleoplasm (yielding "nuclear export half-lives"), and the rates at which polysomes are loaded onto RNAs while in the cytoplasm (yielding "untranslated cytoplasm half-lives"). All these subcellular RNA flow rates reflect their net flow, averaged over possibly varying kinetics of subpopulations within their respective compartments [56–59]. 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 (**Figure 1E**).

This initial model fit the data for most genes well (e.g. *Myc* and *Foxg1*, **Figure 1F**). The model predicts that whole-cell half-lives represent the sum of the nuclear and cytoplasmic half-lives, because it did not include nuclear degradation of mRNA [60] (**Figure 1E**). 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 S3B**). 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 [61], 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 S3B**), whereas a model with nuclear degradation was successful (**Figure 1F**, Bayes factor=10^{197}).

For genes modeled well without nuclear degradation, the chromatin half-life indicates how long it takes for transcripts to be released from chromatin. However, for genes best explained by the nuclear degradation model, the chromatin residence represents the averaged time before either nuclear degradation or release from chromatin into the nucleoplasm. To distinguish the half-lives of RNA on chromatin for each of these transcript fates, we estimated a separate chromatin release rate specific to the population of transcripts destined to be exported for genes modeled with nuclear degradation.

Chromatin, chromatin release, nuclear, nuclear export, nuclear degradation, cytoplasm, and whole-cell half-lives were strongly correlated between biological replicates, with small 95% credible intervals (CIs) (**Figure S3D,E**, Pearson correlation r>=0.59), and correspondence with previously reported whole-cell half-lives (**Figure S3C**, Spearman correlation r>=0.45). However, most genes had larger 95% CIs for their untranslated cytoplasm half-lives than for other flow rates, and the half-lives also exhibited greater variation between biological replicates (**Figure S3D,E**, r=0.44 in K562 and r=0.47 in NIH-3T3). Nevertheless, the 95% CIs reproduced between replicates (overlap for 87% genes in NIH-3T3, 83% genes in K562) and did not exceed the variation between genes. To assess the robustness of our RNA flow rates across modeling approaches, we compared our Bayesian distributions to least-squares estimates. Although chromatin, nuclear, cytoplasm, and whole-cell half-lives were similar (r>=0.67), the least squares model failed to estimate untranslated cytoplasm half-lives (r=0.42 in K562, r=0.17 in NIH-3T3) due to its inability to take measurement uncertainties into account (**Figure S3E**). We conclude that our Bayesian approach is capable of robustly quantifying the flow of RNAs across subcellular compartments for endogenously expressed genes in mouse and human cells.

Wide gene-to-gene variability in RNA flow rates

In both cell lines, RNA flow rates varied considerably between genes (**Figure 1H**). For 90% of genes, chromatin half-lives ranged from 20 to 100 minutes in K562 and from 15 to 160 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>=0.57, **Figure S4A-B**), with median nuclear half-lives around 50 minutes that ranged between 20 and 200 minutes for 90% of genes. As expected from the similarity between chromatin and nuclear half-lives for most genes, about 50% of all genes had nuclear export half-lives of less than 30 minutes. Thus, most mRNAs within the nucleus are associated with chromatin (**Figure S4C**). Notably, the nuclear export calculation is restricted by our time lapse resolution of 15 mins (**Figure 1C**). If we estimate nuclear export through comparing nuclear and chromatin half-lives, we similarly observe that nuclear export can occur faster than our time resolution (**Figure S4D**).

In the cytoplasm, mRNAs were more stable in K562 cells with a median half-life of 71 minutes, compared to 36 minutes in NIH-3T3. Thus, turnover of mRNAs at whole-cell resolution was also slower in K562 cells than in NIH-3T3 cells, with median whole-cell half-lives of 140 and 70 minutes, respectively. In both cell lines, cytoplasm half-lives varied greatly, ranging from 13 to 110 minutes for 90% of genes in NIH-3T3 and from 20 to 220 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=0.05) and weakly correlated in NIH-3T3 (r=0.29) (**Figure S4E-F**). Thus, loading of RNAs onto polysomes is not strongly coupled with cytoplasmic RNA turnover. On average, mRNAs for most genes spent longer in the nucleus than the cytoplasm in

NIH-3T3 cells. Nonetheless, in both cell lines, nuclear and whole-cell half-lives (r=0.66 in K562, r=0.83 in NIH-3T3) were more strongly correlated than cytoplasm and whole-cell half-lives (r=0.4 in K562, r=-0.02 in NIH-3T3) (**Figure S4G-H**).

As expected, we accurately estimated the whole-cell half-life by adding the nuclear and cytoplasm half-lives for most genes, but could not do so for genes predicted to undergo nuclear RNA degradation (**Figure S4I-J**). We saw similar predictability when estimating the nuclear half-life by summing the chromatin and export half-lives (**Figure S4K-L**). Intriguingly, there was only a weak correlation between chromatin and chromatin release half-lives for genes predicted to undergo nuclear degradation, suggesting that the residence times of transcripts on chromatin destined for degradation and export are largely independent (**Figure S4M-N**).

A majority of transcripts from genes predicted to undergo nuclear degradation (PUNDs) are degraded by the nuclear exosome

The Bayes factors for 12% (n= 1350/11,127) of genes in NIH-3T3 cells and 8% (n= 936/11,021) of genes in K562 cells reproducibly exceeded 100 (**Figure 2A, S5A**), suggesting 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 [61]. Hereafter, we refer to these genes as PUNDs (predicted to undergo nuclear degradation). The model predicts that the large majority (>68%) of transcripts produced by PUNDs are degraded in the nucleus without being exported into the cytoplasm. As expected, we observed a higher nuclear to cytoplasm abundance for these genes (**Figure S5B**).

We performed Gene Ontology (GO) enrichment analysis on our PUND gene lists 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 2A, S5A**). Ribosomal protein genes were highly enriched in PUNDs in both cell lines, including 24 individual ribosomal protein gene homologs. In total, we found 278 homologous genes that were shared between both cell lines, a significant overlap (Fisher's exact test: $p<10^{-5}$) (**Figure S5C**). 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.

To confirm that most PUND transcripts are degraded in the nucleus, we independently depleted DIS3 and EXOSC10, two catalytic components of the nuclear exosome [60], and performed subcellular TimeLapse-seq in K562 cells (**Figure S5D-E**). Additionally, we also depleted PABPN1 and ZFC3H1, two cofactors in the poly(A) exosome targeting connection (PAXT) required for targeting transcripts to the nuclear exosome through the PABPN1 and Poly(A) Polymerase-mediated Decay (PPD) pathway [14,17,62] (**Figure S5F-G**). In all cases, we observed a significantly stronger reduction in nuclear turnover for PUNDs as compared to all other genes (**Figure 2B, S5H**). Consistently, this was also true for whole-cell turnover (**Figure S5I**). Importantly, we observed a strong correlation between nuclear transcript stabilization in all depletions and the predicted magnitude of nuclear degradation for PUND genes (**Figure 2C**), validating our model-predicted nuclear degradation half-lives. Interestingly, we observed a similar, albeit weaker, correlation for all genes that may indicate a global role for nuclear degradation in RNA flow across a large continuum of magnitudes (**Figure S5J**). We concluded that our model predicts the extent to which the nuclear exosome acts across transcripts from all genes and that PUNDs are those that undergo relatively high magnitudes of nuclear degradation.

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 [2,31–35]. 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 2D**). By allowing for sufficient granularity, we identified 32 total clusters ranging from the order of 10 to 10^3 genes with reproducibly distinct transcript kinetics (**Figure 2D**). We then performed GO enrichment analysis on each cluster and found that a large majority have several enriched terms (**Figure 2D**). Cluster 6 contains the most genes (n=2,205), many of which exhibit “canonical” RNA flow: median whole-cell half-lives of 84 minutes, long cytoplasmic and relatively short chromatin and nuclear residence, fast polysome loading, and no evidence for nuclear degradation (**Figure 2D**). However, the vast majority of genes (n=8,178) were divided into smaller groups that deviated from these canonical kinetics (**Figure 2D**). Genes involved in signal transduction and response to stimuli (e.g., *MYC, JUN,* and *CXCL2*) were overrepresented in a cluster of 71 genes with “fast flow” kinetics across all compartments (median whole-cell half-life 26 minutes) (**Figure 2E**); however, even these genes spent more than half of their life cycle on chromatin (median half-life 15 minutes). We corroborated this result independently from our clustering analysis by performing gene set enrichment analysis (GSEA) [63,64] 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 S5K**). Thus, mRNAs with short half-lives across all compartments were enriched for signaling and sensing functions.

Ribosomal protein genes (RPGs) were enriched in only four clusters (**Figure 2D, S5L**), rather than being uniformly distributed throughout all clusters (χ^2 test $p<10^{-16}$). Although all four of these clusters shared long cytoplasm half-lives, consistent with previous reports [33,65,66], other RNA flow rates differed. A majority of RPGs were PUNDs with short half-lives on chromatin, with either slow (cluster 2) or average (cluster 3) 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 15, 25). 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 2F**, clusters 7 and 11). Closer inspection revealed that the first group contained mostly variant histone genes, and the second contained canonical, replication-dependent histone genes (**Figure 2F**). Whole-cell half-lives did not differ between these two groups; however, canonical histones had ~2-fold longer nuclear (median 68 minutes) than cytoplasm (median 30 minutes) half-lives (**Figure 2F**). Additionally, canonical histones were loaded onto polysomes very quickly, often within just a few minutes, consistent with a previous study [49]. Thus, canonical histones experience relatively unique RNA flow, whereas variant histones behave more like the average protein-coding transcript.

Finally, genes involved in gene expression clustered into several distinct groups (**Figure S5M**). Most of the enriched GO terms in clusters 4-6 were related to transcription, including terms pertaining to transcription factors, RNA metabolism, and RNA polymerase II activity. These genes included several mediator subunits, splicing factors, chromatin remodelers, and transcription factors. By contrast, genes in clusters 2, 3, and 25 were enriched for functions related to cytoplasmic translation, peptide biosynthesis, and ribosomal subunit biogenesis, and include many translation factors. Genes involved in transcription had longer chromatin and nuclear half-lives and shorter cytoplasm half-lives compared to genes involved in translation (**Figure S5M**). Thus, genes involved in transcription and translation are enriched in clusters representing opposite RNA flow patterns.

RNA flow rates are associated with RNA binding proteins

mRNAs exist as part of ribonucleoprotein complexes containing many RBPs, which control RNA metabolism and are likely regulators of RNA flow. To determine which

RBPs correspond with specific RNA flow rates, we analyzed ENCODE eCLIP datasets for 120 RBPs in K562 cells [67], classified the mRNAs targets of each RBP, and identified all RBPs with target mRNAs that had significantly shorter or longer subcellular half-lives than non-targets (**Figure 3A-B, Figure S6A-E**). We identified significant differences in the subcellular half-lives for nearly all (43/44) of the RBPs whose targets show overall slower or faster whole-cell half-lives than other genes (**Figure 3B**), allowing us to pinpoint exactly where in the cell each RBP may regulate RNA flow. Additionally, we identified significant differences in subcellular half-lives for the targets of an additional 37 RBPs that did not show differences at the whole-cell level, and about half (18/37) of these RBPs only exhibited significant differences in chromatin and/or nuclear half-lives (**Figure 3B**).

DDX3X and PABPC4 regulate nuclear export of target mRNAs

Intriguingly, we identified several RBPs that were associated with opposite trends across compartments, such as those with targets that exhibited faster nuclear half-lives and slower cytoplasm half-lives when compared to all other genes. We reasoned that these RBPs may play multifaceted roles in gene expression and further dissected their roles in regulating RNA flow rates. We began by focusing on DDX3X, an RNA helicase with many roles in RNA metabolism [22,26,68–74]. DDX3X targets exhibited short chromatin half-lives and long nuclear half-lives, indicating slow export from the nucleus, despite the role of this RBP in the nucleus being largely undefined (**Figure 3C, Figure S6A-B**). To analyze the role of DDX3X in determining RNA rates of its target mRNAs in the nucleus, we used shRNAs to deplete DDX3X (**Figure S6F**) and performed subcellular TimeLapse-seq. Upon acute depletion, nuclear RNA turnover of DDX3X target mRNAs was no longer slower than non-targets (**Figure 3D**). Furthermore, DDX3X targets had long cytoplasm and whole-cell half-lives in wild-type cells (**Figure 3B, Figure S6C,E**), consistent with the association of DDX3X with several cytoplasmic RNA granules [69,70,73]. However, RNA turnover on chromatin, in the cytoplasm, and at the whole-cell level were not affected in the DDX3X knockdown (**Figure S6H).** 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 3B, Figure S6C,E**), consistent with the function of this RBP in stabilizing transcripts with short poly(A) tails containing AU-rich motifs [75]. However, despite the fact that PABPC4 has no known roles in the nucleus, targets of this RBP had shorter chromatin half-lives and longer nuclear half-lives than other transcripts (**Figure 3E**). As before, we used shRNAs to deplete PABPC4 in K562 cells (**Figure S6G**) 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 S6I**). However, the target mRNAs no longer had longer nuclear half-lives (**Figure 3F**), 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.

A genetic interaction screen links DDX3X and mRNA export to proliferation

To investigate the mechanistic basis of DDX3X essentiality in relation to its functions across subcellular compartments in an unbiased manner, we conducted genome-wide CRISPR interference (CRISPRi) screens [76] to define genetic interaction networks of DDX3X. We engineered an isogenic pair of K562 mammalian cell lines that expressed the CRISPRi silencing machinery under a doxycycline-inducible promoter. One cell line constitutively expressed a single guide RNA (sgRNA) targeting DDX3X and a control cell line constitutively expressed a widely used negative control sgRNA targeting the yeast Gal4 promoter (**Figure 4A**). We then conducted parallel genome-wide CRISPRi screens in replicate targeting all known protein-coding open reading frames and measured the changes in sgRNA abundance over 7 doublings for each cell line (**Figure 4B, S7D**). We determined that longer growth periods led to selection for cells that escaped knockdown of DDX3X, likely a result of selection imposed by the role of DDX3X in proliferation [77,78], while shorter growth periods might reduce genetic effect sizes. Genes that had either synergistic or buffering growth defects in combination with DDX3X knock-down were classified as genetic interactions (GIs), which identified both known and novel functional relationships with DDX3X (**Figure 4C-D**). We expected that depletion of genes that act in conjunction with DDX3X would have a similar effect of growth either alone or in combination with DDX3X knockdown (buffering interaction)

(**Figure 4D**). Gene ontology analysis on all buffering genes revealed enrichment for genes involved in mRNA export from nucleus and mRNA RNP complex nuclear export (**Figure 4E-F**). Closer examination of specific genes that have this buffering interaction included the nuclear pore protein complex member *NUP160,* the mRNA export factor *NXF1* [71,79,80], and scaffold protein *CNOT1* of the CCR4-NOT deadenylase complex (**Figure 4C**). Individual cloning and validation of these genes in an array-format flow-based competition assay confirmed their buffering activity with DDX3X depletion (**Figure S7A-C**). Our validation efforts indicated this screen format is more sensitive to buffering than synthetic interactions due to the impact of sgDDX3X on cell growth.

One potential mechanism for how these genetic interactions arise is through DDX3X's known role as a translational regulator [74,81,82]. To test this hypothesis, we performed ribosome profiling in K562 cells with doxycycline-inducible CRISPRi-mediated DDX3X depletion (**Figure 4G**). We first validated that previously described genes *(ODC1, NT5DC2, DVL2)* whose translation was dependent on DDX3X in other human cell lines had DDX3X-dependent translation in the K562 CRISPRi line (**Figure S7E-F**). We identified DDX3X translation targets on the whole-genome DDX3X GI map (**Figure 4H-I**). Surprisingly, we found no evidence that DDX3X's role in regulating the translation of specific genes is linked to DDX3X's buffering or synthetic genetic interactions (**Figure 4I**). Very few genes with significantly decreased translation levels after DDX3X depletion were buffering or synthetic interactors (**Figure 4H-I, S7I**) and the genes that had the strongest buffering interactions showed negligible translation changes (**Figure**

S7G-K). The mRNA export genes that showed buffering genetic interactions with DDX3X depletion also did not change at the translational level after DDX3X depletion, suggesting that at least at the single-gene level these buffering interactions are not directly related to DDX3X's role in regulating translation (**Figure 4H**).

To test whether buffering GIs are influencing nuclear export, we performed CRISPRi depletion and subcellular RNA-seq. We found that individual knockdown of DDX3X using CRISPRi led to a decrease in the nuclear abundance of DDX3X-binding mRNAs compared to all other genes (**Figure 4J**), similar to what we observed when DDX3X was depleted by shRNA (**Figure S7L**). Differences in the effect size of the export phenotype is likely due to a difference in experimental setup and the acute nature of the shRNA knockdown versus propagation of CRISPRi cell lines targeting DDX3X. Individual knockdown of the GI buffering interactors CNOT1 and NUP160 reversed this phenotype, leading to an increase in DDX3X-target mRNA nuclear abundance in the CNOT1 knockdown and a less substantial decrease in nuclear abundance of DDX3X-binding mRNAs for the NUP160 knockdown (**Figure 4K, S7M,O**). This increased nuclear abundance of DDX3X-binding mRNAs by the buffering genetic interactors is consistent with the GI map buffering phenotype we see at the level of cell growth.

To determine whether the enzymatic activity of DDX3X is required for its role in nuclear mRNA export, we conducted rescue experiments by knocking down endogenous DDX3X using CRISPRi and overexpressing either wild-type or inactive mutants of

DDX3X that have been previously shown to impair its helicase activity [83–85] (**Figure 4L**). We found all five tested inactive DDX3X mutants restored the nuclear abundance of DDX3X-binding mRNAs to the same level as wild-type DDX3X, despite inactivating DDX3X by distinct mechanisms (**Figure 4M, S7N,P**). Our results indicate that DDX3X's role in nuclear export is independent of its helicase activity. Altogether, these data corroborate our subcellular TimeLapse-seq data that identified a difference in export rates for DDX3X binding targets, illustrating the power of the approach for identifying unappreciated roles for RNA binding proteins.

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

Poly(A) tail lengths are connected to mRNA metabolism [30], so we next analyzed the relationship between poly(A) tail length and RNA flow rates. 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 5A**) [86]. We confirmed that the 3'-end of >80% of sequenced RNAs mapped to annotated poly(A) sites (**Figure S8A**). To control for technical variations between sequencing runs, we included a set of six spike-in RNAs, each with a different poly(A) tail length. We calculated a poly(A) tail length size factor for each sample and used it to normalize endogenous RNAs poly(A) tail lengths (**Figure S8B-C**, methods). Within each fraction, the median tail length per gene correlated well between the two biological replicates (**Figure S8D**).

We observed that poly(A) tail lengths gradually shortened as RNAs flowed through the cell (Figure S8C), as expected [66,87]. These trends could be observed at the single gene level, albeit with variability across genes (**Figure 5B**). For example, median chromatin poly(A) tail lengths ranged from below 200nt to above 250nt (e.g., *TRAPPC3* and *WSB1*, Figure 4B). Globally, the longest poly(A) tails were on chromatin. We measured shorter polyA tail lengths in the cytoplasm, where most genes had median tail lengths of 85–110nt (**Figure 5C**). Total RNA tail lengths closely resembled those of cytoplasm RNA (**Figure 5C**). In general, the distribution of tail lengths in polysome RNA were slightly shorter than in cytoplasmic RNA (**Figure 5C**).

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. On chromatin, median poly(A) tail lengths were the shortest for genes with chromatin half-lives of less than 15 minutes, and chromatin tail lengths gradually increased with half-life until ~40 minutes (**Figure 5D**). Across all genes, chromatin poly(A) tails were shortest on completely unspliced transcripts, and tails increased in length on intermediately spliced transcripts as more introns were removed from the pre-mRNA (**Figure 5E**). This observation indicated that poly(A) tails grow as transcripts spend more time on chromatin, a period 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 S8E-F**), and a negative correlation between median tail length in total RNA and whole-cell residence (**Figure 5F**), consistent with previous reports [88,89].

Because cytoplasmic mRNAs are deadenylated over time, it has been hypothesized that more stable transcripts undergo more deadenylation prior to degradation [89], which has been difficult to directly test without compartment-specific poly(A) tail and half-life measurements. Our data are consistent with this hypothesis; the change in 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 5G**).

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 5H**). 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 5B**). Consequently, *TRAPPC3* had a much shorter median poly(A) tail in total RNA (median 89nt) than *WSB1* (median 142nt) (**Figure 5B**). Thus, median poly(A) tail lengths in total RNA are not only reflective of whole-cell residence times (**Figure 5F**), but also of the relative times spent on chromatin and in the cytoplasm (**Figure 5H-I**).

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 6A, S9A**). Notably, PUND genes had longer cytoplasm half-lives than other mRNAs (**Figure 6A, S9A**), 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 is not a trivial reflection of the breakdown of inherently unstable transcripts.

We found 97 RBPs with target genes significantly enriched for PUNDs (**Figure 6B**). Remarkably, four of the proteins most significantly enriched were components of the spliceosome: PRPF8, SF3B4, EFTUD2, and BUD13 (**Figure 6B**). We also detected splicing factor SRSF1 among the most significant RBPs, along with several other spliceosome components (**Figure 6B**). 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 genes had more incompletely spliced mRNAs on chromatin than other genes, but overall splicing levels did not differ meaningfully in other compartments (**Figure 6C, S9B**). 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= 204nt, median other= 193nt, $p=1.40\times10^{-16}$) (**Figure 6D, S9C**). This was surprising given that PUND transcripts generally reside on chromatin for less time (**Figure 6A**), and we observed that transcripts with short chromatin half-lives generally have short poly(A) tails (**Figure 5D,I**). PUNDs had slightly shorter tails in the cytoplasm (median PUND tail length=96nt, median other=98nt, $p=0.018$) and on polysomes (median PUND tail length=81nt, median other=83nt, $p=0.046$). 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.

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 (**Figure 7A**). To this end, we developed a LASSO regression model (**Figure 7A, S10A**) that identifies sparse relevant features through L^1 regularization [90]. The 10× cross-validation and unseen test set performances of our model varied between the subcellular compartments (**Figure 7B**), ranging from $R^2 = 0.4$ for nuclear export and cytoplasmic turnover to $R^2 = 0.15$ for polysome loading rates, likely due to the larger uncertainty for these estimates (**Figure 1G, S3D**). 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 S10B**) [48,91–97].

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 7C, S10C**). We also observed subcellular-specific relevant classes, such as predicted microRNAs promoting cytoplasmic turnover, gene location predicting nuclear half-lives, and transcription factors targets associating with cytoplasmic turnover (**Figure 7C**). The classes vary widely in the number of quantitative and qualitative input features, which may contribute to the class summed feature importance and explained variation (**Figure 7C, S10C**). Indeed, sequence determinants and codon frequencies were collectively relevant as feature classes (**Figure 7C**), albeit the effect sizes of individual features were generally small and compartment-specific (**Figure S10C-D**). For example, transcripts containing a 5' terminal oligo-pyrimidine (TOP) motif [98] had faster RNA flow dynamics, including polysome loading rates (**Figure S10C**).

Next, we investigated the strongest individual feature contributions across compartments (**Figure 7D**). Intron length, GC content, and the number of exons were top predictors of chromatin half-lives (**Figure 7D**). Accordingly, chromatin and nuclear half-lives increased with gene length in both cell lines, likely due to transcription elongation time, albeit with a large variability between genes (**Figure 7E, S10E**, Spearman correlations for K562: chromatin=0.15, nuclear=0.19, nuclear; NIH-3T3: chromatin=0.20, nuclear=0.16). Gene GC content and SND1 binding both positively influenced chromatin, nuclear, and nuclear export rates (**Figure 7D**). Consistent with our RBP analysis (**Figure 3C-F**), DDX3X and PABPC4 were predicted to regulate nuclear export.

Our LASSO model predicted that genes targeted by the MYC transcription factor (TF) had long cytoplasm half-lives (**Figure 7D**), and we confirmed this finding orthogonally by GSEA. With rare features, for instance TFs with few target genes, their individual feature importance only pertains to their target genes. However, since LASSO selects for global feature relevance and rare features generally make a smaller global contribution, it is more likely that rare features 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 S10F**), similar to our RBP target analysis (**Figure 3A-B**). Remarkably, 193 TFs were associated with altered RNA flow rates (**Figure S10F**). For instance, targets of the stress-response TFs ATF4 and ATF5 had shorter chromatin half-lives (**Figure S10F**). Our LASSO model and TF analysis revealed that many uncharacterized Zinc-finger proteins were also associated with altered RNA flow rates (**Figure S10F**). We conclude that many features, underlying a diverse set of regulatory mechanisms, are predictive of RNA flow rates across the cell (**Figure 7F**).

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 1H**). Transcripts with the same whole-cell half-lives flowed throughout the cell at different rates (**Figure 2D, S4G-H**), highlighting the variability in regulation between different genes. 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 2E**). Our data suggest that many mRNAs remain associated with chromatin for an extended period of time after transcription has completed, during which fully transcribed mRNAs continue to undergo splicing and polyadenylation (**Figure 5E**).

We identified around a thousand protein-coding genes that 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 (**Figure S5J**). Taking into account the high production rates of many PUNDs, we estimate that at least 23% of all human and mouse protein-coding transcripts are degraded in the nucleus. PUND transcripts have unique features, such as incomplete splicing, long poly(A) tails, and association with splicing factors and other RBPs (**Figure 6**). Indeed, we find that PUNDs are degraded through the nuclear exosome, PAXT complex and PPD pathways (**Figure 2B-C, S5H-J**), suggesting that recognition of PUND poly(A) tails contributes to their degradation, as found previously for a limited number of non-coding RNAs and reporter transcripts [14,62].

Given the functions of PUND transcripts in splicing and RNA processing, nuclear mRNA degradation may play a role in autoregulation of RNA metabolism. Many PUNDs are splicing factors, and chromatin PUND transcripts are less spliced on average (**Figure 6C, S9B**). A stimulus may instigate the complete splicing of PUND pre-mRNAs, resulting in their nuclear export, rather than nuclear degradation. When translated, this would result in higher protein levels, perhaps resulting in more efficient splicing genome-wide of the RNAs that are transcribed in response to the stimulus, in turn resulting in more efficient export and translation. Furthermore, in a hypothetical negative feedback mechanism, increased export of PUNDs encoding for nuclear RNA degradation factors, such as EXOSC1/2/9, might result in subsequent increased nuclear RNA degradation. Thus, PUNDs, especially those involved in RNA processing, might act as “sensors” in controlling nuclear RNA homeostasis, a regulatory process recently reported [99,100]. Overall, we believe nuclear transcript degradation to be pervasive and likely to serve a regulatory function, which is an exciting direction for future work.

We observed significant associations between the rates of RNA flow and many RBPs. We report the first nuclear role, in mRNA export, for PABPC4, which has been shown to shuttle between the nucleus and cytoplasm [101]. We also identified new functions for DDX3X in regulating nuclear export. DDX3X is an essential gene in yeast, cultured human cells, and animals, but the genetic basis of its essentiality has been unclear [102]. DDX3X also shuttles between the nucleus and the cytoplasm [72]. In the cytoplasm, DDX3X is understood as a translation regulator, but the role of DDX3X in the nucleus has been less clear [74,81,82]. DDX3X physically associates with the mRNA export factor NXF1 and has been implicated in the HIV-1 viral RNA export from the nucleus [71,79,80]. Surprisingly, we find that limiting nuclear mRNA export buffers against loss of DDX3X in terms of cell proliferation, which provides a mechanistic link to its essentiality. Thus, genetic interactions that buffer DDX3X loss do so, at least in part, by increasing the cytoplasmic-to-nuclear ratio of DDX3X binding targets, indicating that cytoplasmic abundance of DDX3X targets, is important in DDX3X's impact on proliferation.

Human genetic data in DDX3X syndrome and DDX3X-associated cancers indicate selection and associated phenotypic differences between DDX3X depletion or missense variants [85,103–108]. Here, we find that the helicase activity is not required for DDX3X's role in nuclear export, leaving open the possibility that DDX3X may act as a scaffold protein in this process. Since DDX3X helicase mutants do affect translation [74,109], these results further indicate that DDX3X's role in export is likely distinct from its role in translation, highlighting how an RBP can separately control multiple points in the RNA life cycle.

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. 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 4F**), this simple relationship does not capture the dynamics across all compartments as cytoplasm half-lives are generally longer than chromatin half-lives (**Figure 1H**), 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 4H-I**). 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.

We also report distinct patterns of RNA flow for the targets of many transcription factors (**Figure S10F**). While the possible mechanisms that underlie these associations remain to be elucidated, it was recently reported that at least half of the transcription factors in human cells may bind RNA [110]. Thus, a direct role of transcription factors in regulating RNA flow is intriguing. Finally, our LASSO model identified many features to be predictive of RNA flow, serving as a roadmap for unraveling the regulatory control of the RNA life cycle.

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.

Limitations of the study

Subcellular TimeLapse-seq relies on the biochemical purification of subcellular compartments, which has inherent limitations. Other studies using the same or similar cellular fractionation protocols have reported partial co-sedimentation of the ER with the nuclear fraction [44,111]. We believe this may also be occurring in our system and may explain several minor results. First, we see that transcripts enriched for functions related to cell membrane proteins, smooth ER, and COPII-coated vesicles, which are translated at the ER membrane [112,113], were enriched in clusters that show a combination of very long nuclear and extremely short cytoplasm half-lives (clusters 26, 29-30, n=263 genes, **Figure 2D**). We hypothesize that the nuclear fraction may contain a mix of both newly synthesized and preexisting RNAs for these genes, resulting in an artificially long half-life for this compartment. Conversely, the cytoplasm may be enriched for relatively new transcripts that have not yet been loaded onto polysomes and translated at the ER, resulting in an artificially short half-life.

Given the additional purification steps involved in purifying chromatin from the nucleus, we hypothesize that mature transcripts are less likely to co-purify with this compartment. Still, we observe a few genes with extremely long chromatin half-lives with relatively short poly(A) tails (**Figure 5D**), which deviates from the observed trend that chromatin poly(A) tail length increases with chromatin half-lives. We therefore hypothesize that some cytoplasmic mRNAs from these few highly abundant genes may be contaminating the chromatin fraction, yielding artificially long chromatin half-lives (60+ minutes) and short median poly(A) tails (<180nt). Finally, given the variability in the biochemical properties of different granules and condensates [114], we are uncertain where all of these may sediment. However, we note that P-bodies (monitored by marker protein LSM14A) sediment in the cytoplasm (**Figure 1B**) and do not colocalize with polysomes (**Figure S1D**). In all, we believe that our biochemical purification fractionates most transcripts accurately based on their subcellular localization.

Our study did not observe a direct connection between DDX3X-dependent translation and DDX3X genetic interactions, which could be attributed to the aggregate effect of DDX3X on translation on multiple genes that would not be captured in the individual CRISPRi experiments. We observed a differential effect on nuclear export correlating to varying levels of DDX3X depletion, suggesting a dose-dependent relationship that is a promising avenue for future research. Finally, our analysis found a strong effect of proliferation when DDX3X and nuclear export factors were coordinately depleted in

K562 cells, and the impact of DDX3X on translation and nuclear export on proliferation in other cell types will be important future work.

MATERIALS AND METHODS

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), and 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 $4\text{-}5\times10^5$ 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.

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 [115]. Briefly, 50-80 probes were designed per gene using PaintSHOP [116] with additional sequences used for SABER appended to the 3' end in a gene-specific manner. 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. 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. 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. 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 [117]. 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 [118]. 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 S1D**) 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 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, washed again 4x 5 minutes with 1x TBST, and imaged using a Li-Cor Odyssey.

For experiments in **Figure 4/Supplemental Figure 7**, cells were lysed in RIPA Lysis and Extraction Buffer (Thermo Fisher cat: 8990) supplemented with Halt Protease

Inhibitor Cocktail (Thermo Fisher cat: 78429) for 30 min on ice, then centrifuged at max speed at 4°C for 15 min. Supernatant was combined with 4X SDS loading buffer, boiled at 100°C for 10 min, then loaded onto a 4-20% Mini-Protean TGX Precast gel (Bio-Rad cat: 4561095) and transferred to a Trans-Blot Turbo Mini 0.2 μm PVDF membrane (Bio-Rad cat: 1704156). Membranes were blocked in 5% milk in PBST for 30 min, incubated in antibody in 1x PBST overnight at 4°C, washed 3x for 5-10 minutes with 1x PBST, incubated for 1-2 hours at room temperature with secondary antibodies, washed again 3x for 5-10minutes with 1x PBST, and imaged using a Li-Cor Odyssey.

TimeLapse-seq chemistry and library preparation

Samples were prepared for sequencing according to [2]. 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.

RNA sequencing of nuclear wash following subcellular fractionation

RNA was sequenced from four biological replicates of approximately 5 million K562 cells each following fractionation. K562 cells were fractionated as described above ("Chromatin, nuclear and cytoplasm RNA") into chromatin and nuclear fractions with the following modifications. After nuclei were washed in nuclear wash buffer (1x PBS with 1mM EDTA, 0.1% Triton-X) and spun down, the supernatant ("nuclear wash fraction", ~700 μL) was collected and moved to a separate tube. The nuclear pellet was then resuspended in 300 μL RIPA.

A yeast RNA spike-in mix was generated by mixing six in vitro transcribed transcripts at set ratios. 10 ng of the spike-in mix was added to the total volume of each fraction (cytoplasmic, nuclear, nuclear wash). RNA was then extracted from 250 μL of the cytoplasmic and nuclear fractions and 650 μL of the nuclear wash fraction as described above ("RNA extraction"), and fractionation efficiencies were confirmed by western blotting. RNA sequencing libraries were prepared with the SMARTer Stranded Total RNA HI Mammalian kit (Takara 634873) with 1 μg of RNA as input. Libraries were sequenced on a NextSeq 500 (Illumina, San Diego, CA) using 2 x 75 bp reads.

Quantification of newly synthesized RNA from subcellular TimeLapse-seq data

Reads were filtered for quality and adaptor sequences were trimmed using cutadapt v2.5 [119]. 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 S2B**). To prepare the SNP-masked genomes, reads were first mapped to the reference genome with STAR v2.7.3a [120] 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 [121] 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 [122]. 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 [123]. Samples were then grouped by compartment and replicate and converted into .cit files using GRAND-SLAM v2.0.5d [55]. 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 S2A**). 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 [55], 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-π_E)$Binom(k,n,p_{E2})$. The three parameters in this model were fitted to the above T>C distributions using least squares regression (Python v3.7.4, package lmfit v1.0.2, function minimize; arguments: method='least_squares'; p_{E1} initial value=0.0001, min=0, max=1; p_{E2} initial value=0.00001, min=0, max=1, π_E initial value=0.5, min=0, max=1). As a control, we also performed this fitting procedure with the established approach of a binomial error model with a single background T>C conversion p_{E1} [55]. 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 binomial error model (AIC: $p < 1e-16$). 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-π_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). p_{E2} most likely arises from read alignment errors given our increased mismatch tolerance STAR alignment parameters described above: spliced reads that end near exon junctions could align as an unspliced read that extend into introns with mismatches, rather than spanning to the next exon. Additionally, while we generated a SNP-masked genome, we cannot rule out that additional SNPs may be present. The 4sU sample T>C distributions were then modeled as the 4sU-induced T>C conversions + background population: $P(k,n) = \pi_C \, Binom(k,n,p_C) + (1-\pi_C)P_{BG}(k,n)$

To fit the parameters p_C and π_C to the 4sU sample T>C distributions, we used a grid of initial values, $p_{C,init}$ in [0.001,0.01] and $\pi_{C,init}$ in [0.05, 0.5], and applied a similar least squares regression fitting procedure as described above (lmfit v1.0.2, function minimize; arguments: p_C initial value=$p_{C,init}$, min=0, max=1; π_C initial value=$\pi_{C,init}$, min=0, max=1), and retained the fit values with the lowest χ^2 goodness-of-fit value across the grid as our parameter fits for p_C and π_C. This procedure was applied to the T>C mismatch distributions of both the fast and slow turnover genes to generate p_{C_HI} and p_{C_LO}, our respective upper and lower bound estimates on p_C. In very few cases the slow turnover genes did not contain sufficient TC conversions above the background distribution in a particular compartment and timepoint to estimate p_{C_LO} from the data directly. This scenario occurred when the above fitting procedure resulted in ${\pi_C}^{LO}$ <= 5% and a p_{C_LO} that was logically inconsistent as a lower bound: either because it was lower than the p_{E1} and p_{E2} values, higher than p_{C_HI}, or higher than p_{C_LO} at a later timepoint within the same or upstream compartments. In this case, we set the p_{C_LO} to a logically consistent lower bound value: min(p_{C_HI}, p_{E2}, p_{C_LO}) from the same or upstream compartments and/or the same or earlier timepoints.

Next, 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 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 [124], and quantified by Nanodrop.

150ng of RNA was hybridized with gene-specific DNA probes, 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 1E**), defined by a system of coupled ordinary differential equations (ODEs, **Figure S3A**), describes the average time evolution of the variables, i.e. the 4sU-labeled RNA levels in their respective subcellular compartments and kinetic rates at 4sU-pulse time *t* : T, whole-cell (total); CH, chromatin; N, nucleus; N_E, nuclear export; CY, cytoplasm; P, polysome; M, mature untranslated cytoplasm. In addition to the RNA flow rates (vector $\vec{k}$, **Figure S3A**), 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 S1B-C**). To solve these ODEs analytically [125], 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 N_E($\vec{k}, k_P$, *t*) 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 in Supplemental Methods file 1.

As described in section *Quantification of newly synthesized RNA from subcellular TimeLapse-seq data* (**Figure S2A**)*,* the gene-specific (g) fraction of new RNA (π) Posterior probability density function, $P_{X,t}^{g}(\pi)$, 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 [55]. To analytically derive the (in some cases multivariate) posterior on the RNA flow rates [125], 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 S3A**).

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 S3A**). The mean rate (expression in **Figure S3A**) 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 (for CY, P, N_E), we marginalized the posterior by integrating over the 95% CIs of the already determined upstream RNA flow rate(s) (**Figure S3A**). 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.

Deterministically, the whole-cell RNA levels equal the sum of nuclear and cytoplasmic RNA: T(t) = N(t) + CY(t) (**Figure S3A**). In absence of nuclear RNA degradation, the whole-cell half-life thus equals the sum of nuclear and cytoplasmic half-lives (**Figure S4I-J**). In presence of nuclear degradation (see the next section for the PUND identification procedure), this simple relation between half-lives no longer holds (**Figure S4I-J**).

Deterministically, the time duration of nuclear export equals the difference between the nuclear and chromatin residence times: $k_E^{-1} = k_N^{-1} - k_{CH}^{-1}$. Indeed we see this relation globally (**Figure S4K-L**), albeit with a minor bias, possibly due to a slight overestimation of k_E^{-1}. To investigate nuclear export further with a method alternative to the above described ODE model (**Figure S4D**), we used the above equation's probabilistic analog [125] to determine the export posterior by numerical integration (as described above) over the convolution of the nuclear and chromatin posteriors: $P_E^g(k) = \int_{prior\ domain} \quad P_N^g(\frac{kz}{k+z})P_{CH}^g(z)dz$.

The above described models assume no nuclear RNA degradation is occurring. Therefore, for PUNDs, we extended the ODE model by including nuclear degradation, which can occur on chromatin and in the nucleoplasm (**Figure 1E, S3A**). The objective then became to estimate the posteriors on the nuclear degradation rate k_{ND}, chromatin release rate k_R , and nuclear export rate k_E for PUNDs. Solving the ODEs for T, N, CH and CY compartments again with an integrating factor method, resulted in fraction of new RNA solutions that were multivariate in $\vec{k} = [k_R, k_E \ , k_{ND}, k_{CY}]$. Therefore, to achieve the (marginal) posteriors on an individual rate, we multiplied the posteriors for these compartments, applied the change of variables, and numerically integrated (Python package: scipy, function integrate.nquad) over the 95% CIs for k_{CY} and the prior domains of the remaining two out of three unknown rates (**Figure S3A**). With these resulting three posteriors on respectively k_R, k_E and k_{ND} , the MAP and 95% CIs for were calculated as above, but on a grid with N=100 data points on the prior domain to limit the compute time. Calculating the means was computationally prohibitive so instead the MAP values we used as a substitute mean estimate. To mitigate numerical underflow and/or overflow issues during numerical integration, we added a constant factor to the (unnormalized) joint distributions where needed to stabilize. This procedure does not alter the resulting summary statistics, but enabled the rate calculation for more genes.

In summary, our posterior rate distribution with the 95% CI as the main metric of uncertainty, relies on three parts: 1) the GRAND-SLAM Bayesian model (Jürges et al. 2018) to quantify the fraction of new RNA posterior at individual timepoints. The prior is a uniform distribution, a model assumption that is the standard practice of the field when assuming no prior knowledge; 2) Our modifications to estimate the fraction of new RNA posterior to capture any residual uncertainty that GRAND-SLAM did not consider, such as the time delay for 4sU to be incorporated into nascent RNA after introduction to the media, described in methods section *Quantification of newly synthesized RNA from subcellular TimeLapse-seq data*; 3) Our kinetic model through a system of ODEs (**Figure S3A**). Here, we take a minimal modeling approach to explain the data, which assumes the least possible prior knowledge. In this part, we formally propagate upstream uncertainty to our credible intervals in a mathematically rigorous manner. So, our prior strengths remain as minimal and weak as possible.

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 S3E**). Because of the same drawback, the LSE, but not the Baysian model, also predicted a number of biologically unlikely fast rate values (**Figure S3E**). Besides these differences, we generally observed a strong correspondence between our Bayesian MAP and LSE rate estimates (**Figure S3E**). 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 1G**), and thus also a 95% CI (**Figure S3E**), which indicates the range of rate values consistent with the subcellular Timelapse-seq data.

Since the NanoString 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.

The model prediction curve of the PUND fold stabilization in the nuclear exosome mutants (**Figure 2C, S5J**) is derived as follows. In this minimal model, the assumptions are: 1) in the scramble treated samples, nuclear degradation occurs happens with the k_{ND} MAP value estimated from the WT samples; 2) in the mutants, nuclear degradation is disrupted so that $k_{ND} = 0$. 3) Nuclear RNA flow through chromatin release and nuclear export into the cytoplasm, modeled jointly as an effective rate k^* as described in the next section, is assumed to be unaltered between scramble and KD conditions. Then, after the experimental 4sU pulse time of 1h: $MAP_{scr} = 1 - exp(-(k_{ND,WT} + k^*) \times 1h)$, and $MAP_{KD} = 1 - exp(-k^* \times 1h)$. The model predicted fold stabilization curve then equals $\frac{1-MAP_{KD}}{1-MAP_{scr}} = 2^{1h\,/\,t_{\frac{1}{2}ND,WT}}$.

Bayes Factor model comparison to identify nuclear RNA degradation

To perform formal Bayesian model comparison, we calculated the Bayes factor K [61], i.e. the ratio of likelihoods of the kinetic model that includes a nuclear degradation rate (alternative hypothesis *M₁*) over the simpler “nuclear residence” model with no nuclear degradation (null hypothesis *M₀*): $K = \frac{P(D|M_1)}{P(D|M_0)}$. *D* indicates the data, in this case subcellular Timelapse-seq used to distinguish the two models: the timeseries of chromatin, nuclear, cytoplasmic and whole-cell fraction of new RNA posteriors (**Figure S2B**), as described in the above sections.

$$P^g(D|M_0) = \frac{1}{(k_{hi} - k_{lo})^2} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \prod_{i=1}^{4} P^g_{CH,t_i}(\pi = \Lambda_{CH}(k_R, t_i)) P^g_{N,t_i}(\pi$$

$$= \Lambda_{N_E}(k_R, k_E, t_i)) \times$$

$$P^{g}_{CY,t_i}(\pi = \Lambda_{CY}(k_R, k_E, k_{CY}, t_i))\ P^{g}_{T,t_i}(\pi = \Lambda_T(k_R, k_E, k_{CY}, t_i))dk_R dk_E dk_{CY}$$

And the equivalent for the nuclear degradation model:

$$P^{g}(D|M_1) = \frac{1}{(k_{hi} - k_{lo})^3} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \prod_{i=1}^{4} \ P^{g}_{CH,t_i}(\pi$$

$$= \Lambda_{CH}(k_R + k_{ND}, t_i))P^{g}_{N,t_i}(\pi = \Lambda_{N_E}(k_R, k_E, k_{ND}, t_i)) \times$$

$$P^{g}_{CY,t_i}(\pi = \Lambda_{CY}(k_R, k_E, k_{ND}, k_{CY}, t_i))\ P^{g}_{T,t_i}(\pi = \Lambda_T(k_R, k_E, k_{ND}, k_{CY}, t_i))dk_R dk_E dk_{CY} dk_{ND}$$

k_{lo}=10^{-4} and k_{hi}=10^{4} min^{-1} indicate the prior rate domain bounds. We calculated these integrals numerically (as described in the previous section).

For some genes, the above described 4 compartment Bayes Factor could not be determined using the equations above for both biological replicates, for instance due to the lack of chromatin compartment data. For these cases we considered a simpler model of nuclear degradation, the 3 compartment Bayes Factor, without using the chromatin compartment:

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

$$= \Lambda_{CY}(k_N, k_{CY}, t_i)) \ \times$$

$$P^{g}_{T,t_i}(\pi = \Lambda^{3}_{T}(k_N, k_{CY}, t_i))dk_N dk_{CY}$$

and

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

$$= \Lambda_{CY}(k_N, k_{CY}, t_i)) \times$$

$$P^{g}_{T,t_i}(\pi = \Lambda^{3}_{T}(k_N, k_{CY}, k^{*}\ \ , t_i))dk_N dk_{CY} dk^{*}$$

With $\Lambda_T^3 = 1 - \Psi k_{CY}(k_N - k_{CY} - k^*)exp(-k_N\, t) - \Psi\, k^*\, k_N\, exp(-k_{CY}t)$, with $\Psi = \frac{1}{(k_N - k_{CY}) * (k_{CY} + k^*)}$. Here, k^* indicates the flow of RNA from the nucleus into the cytoplasm, but importantly, it is not to be confused with the nuclear export rate. 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 $k_{ND} := k_N - k^*$. For the nuclear residence model: $k^* = k_N \Leftrightarrow k_{ND} = 0$. The drawback of this model is that k_R and k_E are not present as variables, so this model cannot be used for their rate posterior estimation. Nevertheless, it is suitable, albeit less powered by using 3 instead of 4 compartment data, to determine whether nuclear degradation better explains the 3 compartment data.

When $K > 100$, it is considered “decisive” evidence in favor of the alternative model, and “strong” evidence if K ranges from 10 to 100 [61]. In our case, genes with transcripts predicted to undergo nuclear RNA degradation (PUNDs) are defined as having $K > 100$ for both biological replicates.

Lastly, for each PUND gene, we estimated f_{ND}, the average fraction of transcripts that are nuclear degraded as opposed to exported: $f_{ND} = \frac{k_{ND}}{k_{ND}+k_R} \frac{k_R + k_{ND}+k_E}{k_{ND}+k_E}$, using the mean nuclear degradation (k_{ND}), chromatin release (k_R), and export (k_E) rates as described in the previous section. With these fractions, we then estimated the total cellular fraction of nuclear degraded protein-coding transcripts as: $f_{ND}^{cell} = \frac{\sum_{PUND\ genes\ g} \quad k_p^g f_{ND}^g}{\sum_{all\ genes\ g} \quad k_p^g}$. 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. chromatin release, 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 2D**). For the polysome compartment, only the 95% CIs were used, given that their MAP values were less reproducible across biological replicates (**Figure S3D**). To identify gene clusters (scipy.hierarchy.fcluster, arguments: criterion='maxclust'), we allowed sufficient granularity through a total cluster number of 70. To ensure the robustness of our findings (**Figure 2D**), 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 less than 5 genes and genes with irreproducible patterns, i.e. if the median of the half-lives within a cluster differed at least 10 fold between biological replicates for any of the compartments. This resulted in 32 clusters with reproducible patterns, comprising 10,376 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 2D**).

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 = 0.05/32, to correct for the multiple testing over all 32 clusters, methods = ['fdr_bh'], Benjamini Hochberg multiple testing correction over the GO terms, and gene_universe = all human genes with any RNA flow rates) [126]. For each cluster, the most enriched GO term, i.e. with highest odds ratio (Figure 2D), and all significant GO terms are listed. 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 2A, S5A**).

Gene set enrichment analysis (GSEA) [63] was performed with MSigDB (v7.5.1) gene sets h.all, c5.all, c3.all and c2.all [64], and parameter settings as in [127]. 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.

ShinyGO 0.76.2 (http://bioinformatics.sdstate.edu/go/) [128] was used for calculating GO enrichment or DDX3X buffering genes (**Figure 4F**).

RNA binding protein associations with RNA flow rates

K562 eCLIP data from the ENCODE consortium for a total of 120 RBPs from [67] (eCLIP peaks previously identified according to was analyzed. The following RBPs 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 assess whether this statistical approach has any systematic biases, we performed simulations where random target gene sets were sampled with the same size distributions as the original target gene sets. Then, the same statistical approach was performed as described above. This sampling procedure was repeated 100 times.

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, PABPC4, DIS3, EXOSC10, PABPN1, and ZFC3H1

K562 knockdown lines were generated according to [129] with slight modifications. To generate constitutive knockdown cell lines, 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. To generate inducible knockdown cell lines, plasmid DNA was purified from pTRIPZ backbone vectors with shRNA targeting DIS3 (Horizon Discovery, V2THS_96258), EXOSC10 (Horizon Discovery, V2THS_275659), PABPN1 (Horizon Discovery, V2THS_41638), ZFC3H1 (Horizon Discovery, V2THS_35985), and a scrambled control (Horizon Discovery, RHS4743) under a doxycycline-inducible promoter. 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 $0.2\text{-}1.0 \times 10^6$ cells/mL for 4 days. For the inducible knockdown lines (DIS3, EXOSC10, PABPN1, and ZFC3H1), 1ug/mL of doxycycline was added to cell media beginning 48 prior to cell harvest, and media was refreshed with new doxycycline every 24 hours. All 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 (*BCD1, ICT1, HIF1, ENO2, YKE4, HMS2)* were amplified from their genomic locus using HiFi Hotstart DNA polymerase (KAPA) 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 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) [121] with parameters -ax splice -uf -k14. Poly(A) tail lengths were estimated using nanopolish v0.13.3 [86]. 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 [130] 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 [131]. 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 S7B**), 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 [7,132]. 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 [7,132]. 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 S9A**). Given the large number of input features (70145), LASSO regression was chosen as a model because it is a linear model, $y = \beta X$, with L1 regularization, which ensures sparse feature selection [90]: $Loss = \frac{1}{2N}||y - \beta X||_2^2 + \alpha||\beta||_1$ (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. Features classes gene location, histone modifications, predicted 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. Z-score normalization of the categorical variables did not improve the model performance (Fig 7B) and was therefore not further considered.

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 S10A**): (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 [97]. 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 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 7B, S10B**), 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 7C-D**, **S10C-D**). (6) Lastly, we tested the trained round 2 LASSO model by determining its performance on the 10% unseen test set (**Figure 7B, S10B**).

For the "consensus" whole-cell half-lives [97], 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 S10B**).

Continuous averaging plots (**Figure 7E, S10E**), were generated as in [42], 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: $< L_k >:= \frac{1}{k}\sum_{g=1}^{k} \quad L_g$ and $< L_{N+k} >:= \frac{1}{N-k}\sum_{g=1+k}^{N} \quad L_g$. Next, the corresponding continuous median, and Q25 and Q75 (shaded error bands) of the hal- 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 TF target gene sets were used as the input features in our LASSO model [133]. The statistical analysis was performed as for the RBP targets as described above in *RNA binding protein associations with RNA flow rates.*

DDX3X knockdown and DDX3X mutant cell line generation

The doxycycline-inducible and constitutive K562 CRISPRi cell lines were generated in previous work [134]. The dox-inducible cell line was infected with lentiviral particles encoding sgRNAs targeting DDX3X and the control yeast Gal4 promoter as well as a mCherry reporter at a MOI of 0.3-0.5, then sorted using a Sony SH800 Cell Sorter for mCherry expression. Cell lines were sorted twice for purity before expansion and freezing down of stocks. The constitutive K562 CRISPRi cell line was infected with lentiviral particles from the DDX3X transgene overexpression plasmids at a high MOI of >1.0, then sorted for high mCherry (top 10-20% of population) on a Sony SH800 twice for purity before expansion and freezing down of stocks.

Cloning sgRNA and DDX3X mutant overexpression plasmids

sgRNAs were ordered as oligos (IDT), annealed, then ligated into BstXI/BlpI-digested pAX71 (hU6-sgRNA-EEF1A-GFP-modified-BstXI) expression cassette which has no puromycin cassette and has a modified BstXI restriction site that will not be amplified during library prep or pAX9 (pCRISPRia-v2-mU6-sgRNA-EEF1A-Puro-T2A-BFP) from [76]. DDX3X transgene overexpression plasmids were subcloned into the backbone of pAX49 (JKNp64-pHR_UCOE_SFFV_dCas9-HA-XTEN-VPR-P2A-P2A-BFP, gift from James Nunez), with the final insert being: UCOE-SFFV-DDX3-[WT/Mutant]-3xflag-2xP2A-mCherry. Point mutations were generated by site-directed mutagenesis with NEBuilder HiFi (New England Biolabs cat: E2621L).

Genome-scale CRISPRi screening

Genome-scale CRISPRi screens were conducted based off of [76]. We used the hCRISPRi v2 library with 5 sgRNAs/gene targeting 18,905 genes, 104,535 total targeting sgRNAs, and 1,895 negative control sgRNAs. The hCRISPRi v2 library and the plasmids were propagated using MegaX DH10B T1R Electrocomp Cells (Invitrogen cat: C640003) and electroporated under the following conditions: 25 μF, 200 ohm, 2.5kV, for 0.2 cm cuvette before recovering with SOC media for 1 hr at 37C and culturing in 500mL LB media overnight at 37C. Lentivirus was prepared in HEK293T cell lines using TransIT-LT1 Transfection Reagent (Mirus Bio cat: MIR 2306) and filtered using a 0.45μm PES filter before freezing at -80C. For the genome-wide screen, 50 million cells were infected with 14mL lentivirus containing the hCRISPRi v2 library for each condition. At 3 days post infection, a MOI of 0.5 was assessed via flow cytometry and cells were placed under puromycin (3ug/ml) selection. After 4 days of puromycin selection, cells were spun out of puromycin and allowed to recover for 3 days. At this point, initial sgRNA abundance ("T0 timepoint") was assessed through the aliquot and freezing down of 100 million cells per replicate and replicates were split into their respective flasks and maintained independently. Doxycycline was also added at a final concentration of 50ng/ml and doubling and cell viability was tracked throughout the course of the screen. The sgGal4 replicates were collected after 11 days of doxycycline induction (7 doublings) and the sgDDX3 replicates were collected after 14 days of doxyclicine induction (7 doublings) (100 million cells were collected per replicate). All replicates were maintained between 50-100 million cells at all times to maintain library diversity.

Genome-scale CRISPRi high-throughput sequencing sample preparation

Sample preparation for high-throughput sequencing was conducted as described in [76]. Briefly, the NucleoSpin Blood XL kit (Machery-Nagel cat: 740950.50) was used for DNA extraction. DNA was digested with MfeI-HF (New England Biolabs cat: R3589L) to size enrich the sgRNA-containing fragments of genomic DNA before running all DNA on a

0.8% 1X TAE agarose gel. DNA between the sizes of 400bp-650bp were excised and gel purified using a NucleoSpin Gel Cleanup kit (Machery-Nagel cat: 740609.250). PCR enrichment of the sgRNA cassette and addition of Illumina sample indexes, and sequencing adapters, was conducted with the NEBNext Ultra II Q5 Master Mix (New England Biolabs cat: M0544L) and then gel purified. Libraries were quantified with a Qubit 3 Fluorometer and an Agilent 2100 Bioanalyzer (Agilent cat: 5067-4626) and sequenced on an Illumina HiSeq4000 with a custom sequencing primer. Processing of screen data was performed as previously described using the ScreenProcessing package [76]. Briefly, sequencing reads were aligned to the CRISPRi v2 library top-5 sgRNA protospacer sequences and counted. A negative control gene population was generated and Mann-Whitney p-values for each gene targeted was calculated. Genetic interaction scores were calculated by fitting the data to a best-fit quadratic curve, then calculating the distance of each point to the curve [135].

Flow-based competition assay

Two cell lines were used: the sgGal4 containing K562 line and the sgDDX3 containing K562 line. Each cell line was infected with either a 2nd sgRNA targeting Gal4 or a guide targeting a gene to be validated at desired MOI of 0.4-0.6. 3 days after infection, the percentage of cells expressing the sgRNA (assessed by a BFP reporter in the sgRNA cassette) was evaluated by on a BD LSRFortessa flow cytometer. Then doxycycline was added at 50ng/ml and time points were taken at days 7, 11, and 14 after the addition of doxycycline. This mimics the buffering/synthetic genetic interaction that was conducted in the screen. Log2 enrichment/depletion was calculated by dividing the percentage of cells expressing the guide by the day 0 time point percentage and taking the logarithm of that ratio.

Ribosome profiling

Ribosome profiling was performed as previously described in [74]. Cells were treated for 5 min with 100 µg/ml cycloheximide to trap ribosomes before 20 million cells for each replicate were harvested and 10% of the cytosolic lysate was saved for RNA sequencing. RNA was isolated with a Direct-zol RNA Miniprep kit (Zymo Research R2050) and RNAseq libraries were prepared as per manufacturer's instructions using the NEBNext rRNA Depletion Kit V2 (NEB cat: E7400) and the NEBnext Ultra II Directional RNA Library Prep Kit (NEB cat: E7760). Libraries were quantified with a Qubit 3 Fluorometer and an Agilent 2100 Bioanalyzer (Agilent cat: 5067-4626) and sequenced on an Illumina Hiseq 4000. Ribosome profiling data was high quality: the mean number of uniquely coding-sequence mapping reads was 7,557,798 per replicate (range: 5,205,290 – 11,887,131 reads) and the average overall mapping rate after collapsing of PCR duplicates and filtering out repeat RNAs was 14.5% (range: 10.8% - 16.69% mapping).

Ribosome profiling data analysis

Raw sequencing reads were converted from .fastq to .fasta using fastx-toolkit (v0.0.13), adapter sequences were trimmed using cutadapt (v3.4) [119], and reads were collapsed by UMI using fastx-toolkit. Bowtie2 (v2.4.1) [136] was used to filter out repeat RNAs (rRNA, snoRNA, miRNA). A custom Perl script was used to trim reads a second time and bbduk was used to directly remove contaminating reads mapping to the sgRNA. STAR aligner [120] (v2.7.5a, GrCH38, Gencode v25) was used to align reads. RSubread [137] was used to generate counts and the DESeq2 [131] and Riborex [138] packages were used for differential expression and differential translation analysis.

Subcellular fractionation for RNA-seq

Constitutive CRISPRi K562 cells were infected with lentivirus of sgRNA expressing plasmids targeting GI map genes at a MOI of 0.3-0.5. Constitutive CRISPRi K562 cells overexpressing DDX3X mutants were infected with lentivirus of sgRNA targeting DDX3X at a MOI of 0.3-0.5. After 2 days, cells were selected to purity with puromycin (3ug/ml) for 3 days, recovered in non-puromycin media for 2 days and then harvested for subcellular fractionation. Subcellular fractionation was conducted as described previously in the “Cell fractionation” methods section. Proteinase K (Thermo Fisher cat: EO0491) was added to the nuclear fraction at 1:100, and incubated at 50°C for 10 min prior to RNA extraction. RNA was isolated from nuclear and cytoplasmic fractions with the Quick-RNA Microprep kit (Zymo cat: R1051). RNAseq libraries were prepared as per manufacturer’s instructions using the NEBNext rRNA Depletion Kit V2 (NEB cat: E7400) and the NEBnext Ultra II Directional RNA Library Prep Kit (NEB cat: E7760). Libraries were quantified with a Qubit 3 Fluorometer and an Agilent 2100 Bioanalyzer (Agilent cat: 5067-4626) and sequenced on an Illumina Novaseq 6000.

Subcellular Fractionation Analysis

Raw sequencing reads were aligned and gene counts were generated with STAR Aligner (v2.7.5a, ENSEMBL GrCh38, NCBI RefSeq GCF_000001405.40). RPKMs below 20 were filtered out for the final analysis.

FIGURES

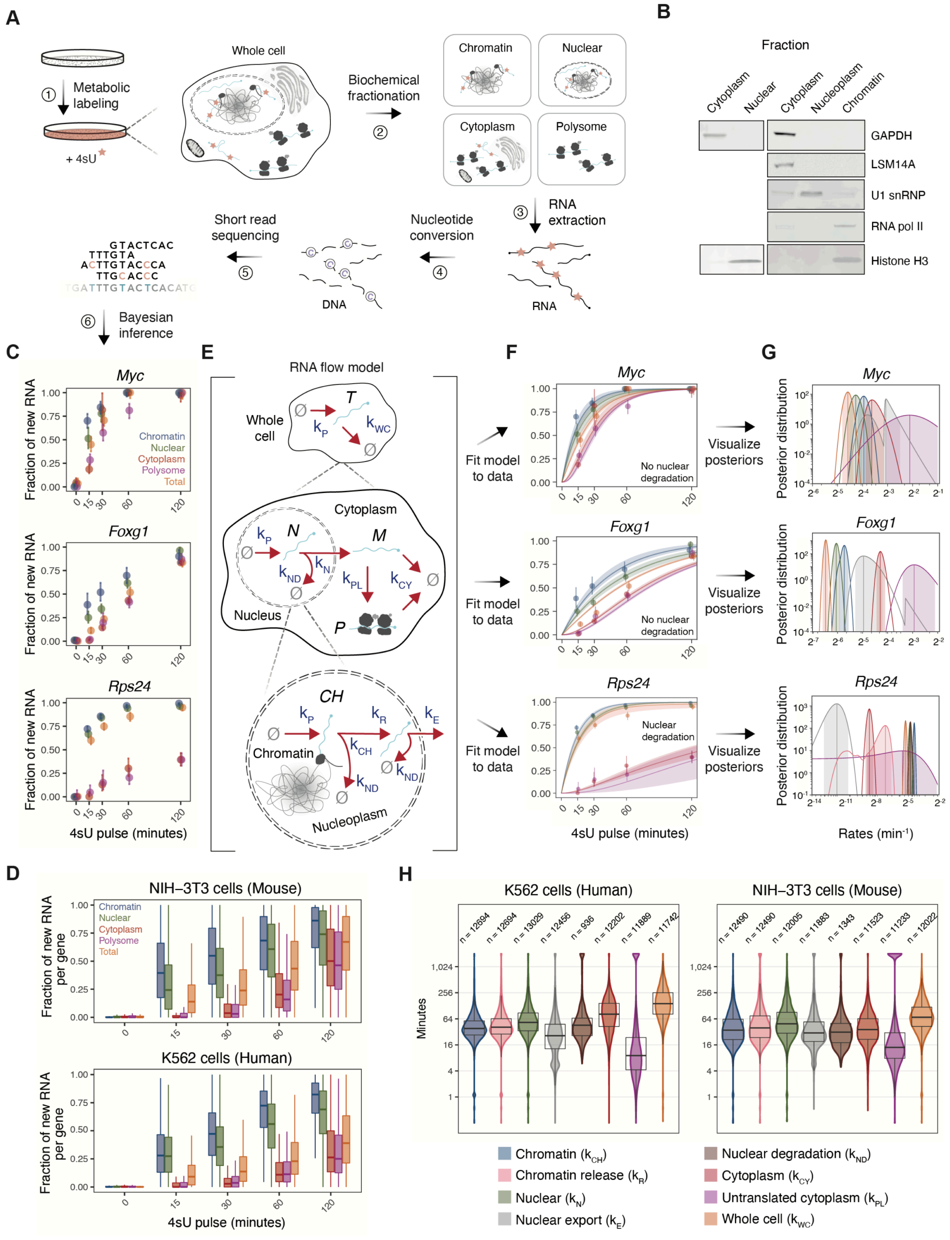


Figure 1: Subcellular TimeLapse-seq and kinetic modeling estimate genome-wide RNA flow rates.

(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. S2B and Methods for details).

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(B) Western blot of subcellular marker proteins: GAPDH and LSM14A, cytoplasmic proteins; U1snRNA, 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.

(E) Schematic of the RNA flow model (see Fig. S3A and Methods for details).

(F) RNA flow model fit to subcellular TimeLapse-seq data for the example genes shown in (C). The dark lines represent the RNA flow rate MAPs while the ribbons show the 95% CIs. Colors are consistent with the RNA populations in (C).

(G) Posterior distributions for each RNA flow rate modeled in (F) with the MAPs represented with vertical lines and 95% CIs in shading. Colors are consistent with the rates shown below in (H).

(H) 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.

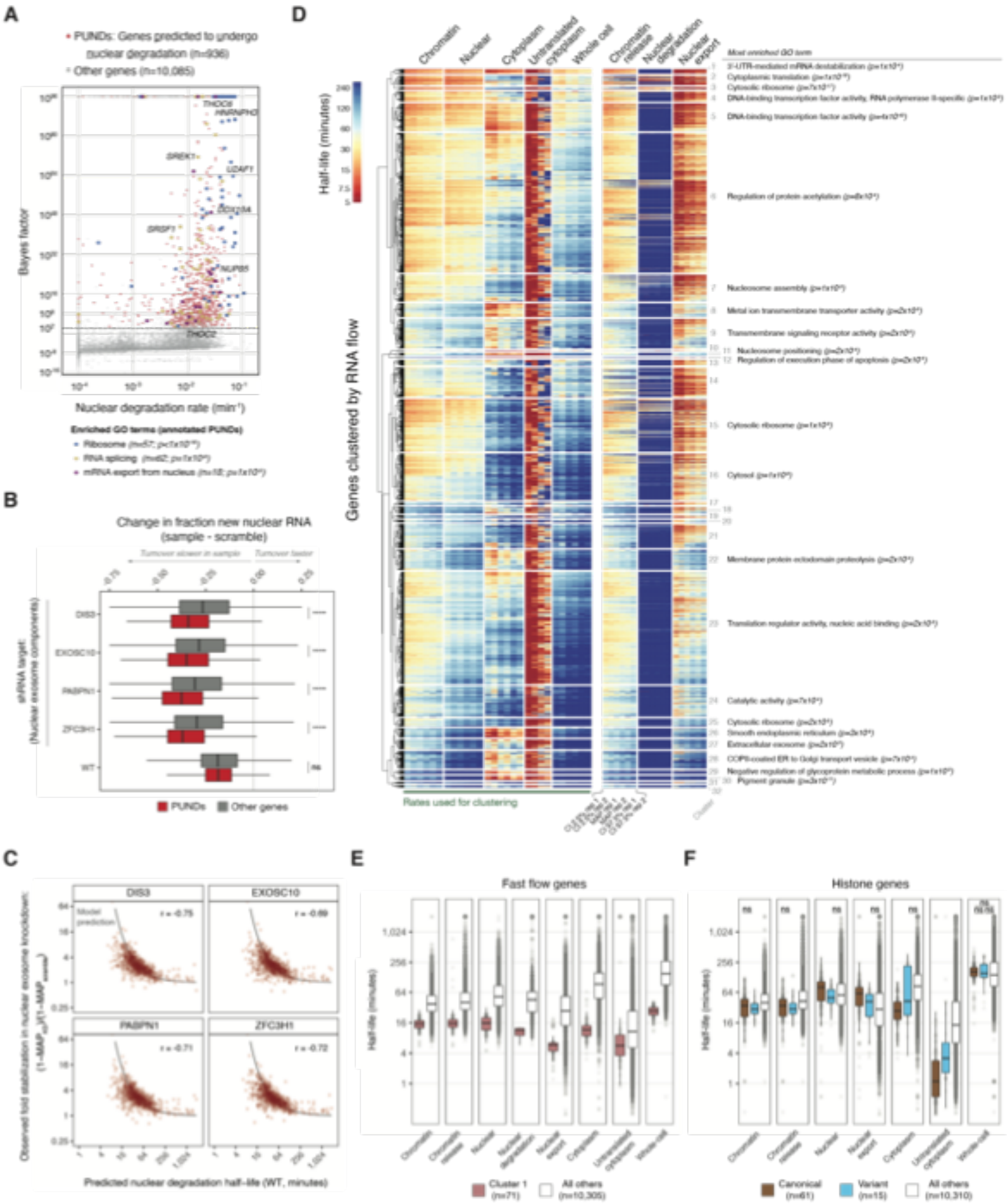


Figure 2: PUND transcripts are degraded by the nuclear exosome and transcripts with similar RNA flow are functionally related.

(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. Genes with Bayes factor>10^{96} are capped at 10^{96}.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(B) The difference in fraction of new nuclear RNA upon depletion of four different components (DIS3, EXOSC10, PABPN1, ZFC3H1) of the nuclear RNA degradation pathway in K562 compared to a non-targeting control shRNA for PUNDs and all other genes. Comparison of wild-type K562 to the non-targeting shRNA is included as a control. Data for one replicate is shown (****: $p<0.0001$, Wilcoxon test).

(C) The stabilization in fraction of new nuclear RNA observed in the four nuclear RNA degradation pathway depletions compared to the nuclear degradation half-life predicted in wild-type cells for PUNDs. The line indicates the stabilization predicted in the depletions and the Spearman correlation is shown.

(D) Hierarchical clustering of human genes according to their RNA flow half-lives (MAPs and 95% credible intervals). Genes were clustered using half-lives that were directly measured from the observed compartments (left columns). The most enriched GO annotations for each cluster are displayed on the right.

(E) Fast flow genes, i.e. those in cluster 1 in (D), were enriched for functions related to intracellular signaling and response to stimuli. Nuclear degradation half-lives are only shown for genes best explained by this model. All comparisons were statistically significant ($p<0.0001$, Wilcoxon test).

(F) Half-lives of histone genes. Canonical, replication-dependent histone genes were enriched in cluster 11, while non-canonical, replication-independent histone genes (including histone variants) were enriched in cluster 7. The number of histone genes of each type is noted, and significance was noted as in (E) unless otherwise specified (“ns:” not significant).

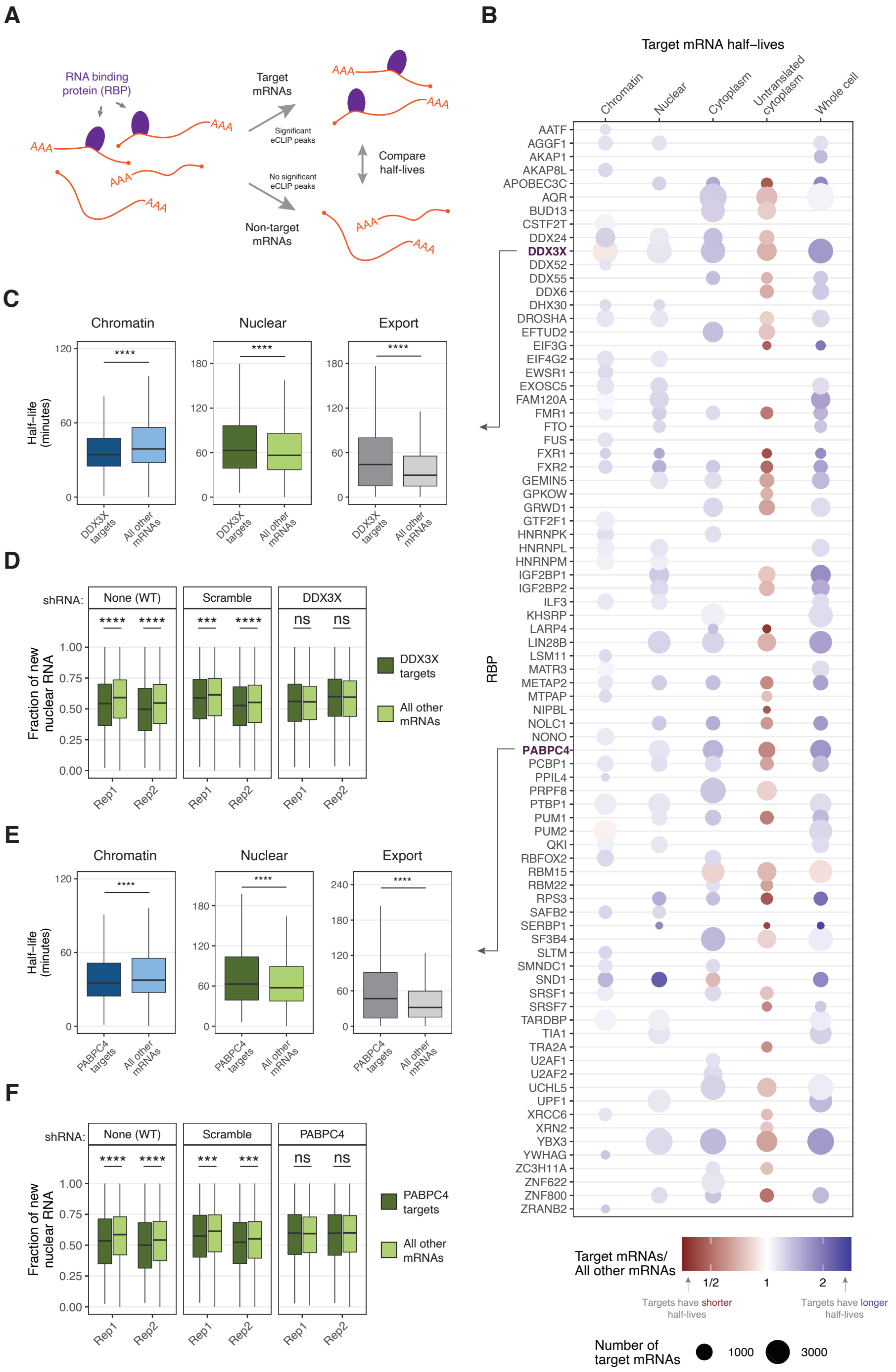


Figure 3: The targets of many RNA binding proteins (RBPs) exhibit distinctive RNA flow rates across the cell.

(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).

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(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 that was directly measured from the observed compartments. 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.

(C) Half-lives of DDX3X mRNA targets and non-targets. 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).

(D) 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 (C). Two biological replicates (“rep”) are shown.

(E) Same as (C) for PABPC4 target genes.

(F) Same as (D) for PABPC4.

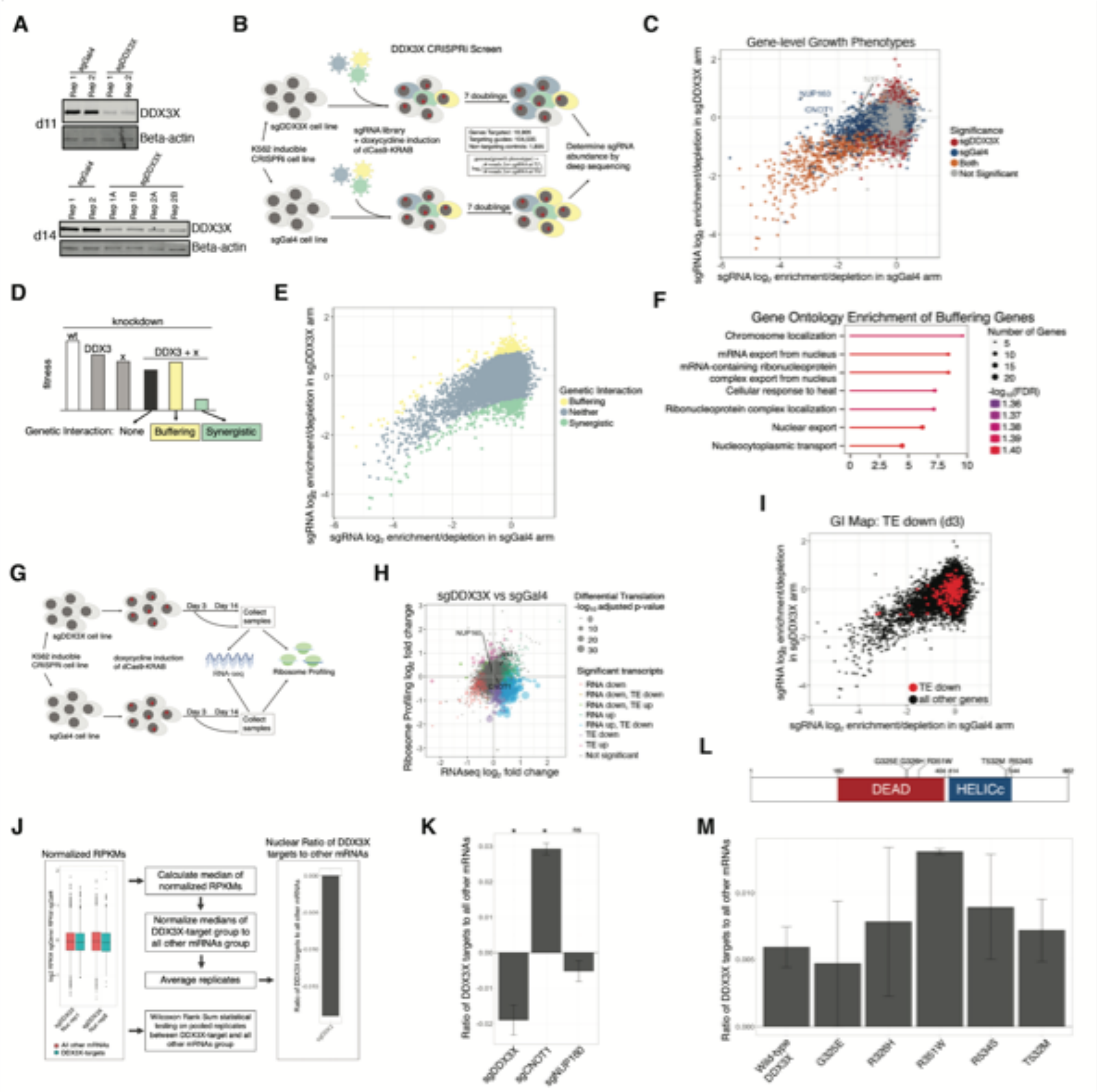


Figure 4: Genetic interaction mapping to link DDX3X depletion to mRNA export defects.

(A) Western blot showing DDX3X levels during the screen. Samples were collected on day 11 (d11) post-induction for sgGal4 and day 14 (d14) for sgDDX3.

(B) Schematic of the growth-based CRISPRi screen. A doxycycline-inducible CRISPRi K562 cell line was infected with a sgRNA targeting either a control yeast promoter Gal4 or DDX3X. The cells were infected with a genome-scale sgRNA library, puromycin selected for cells containing a library element, treated with doxycycline to induce dCas9, and collected after 7 cell doublings. The difference in sgRNA abundances was determined by high-throughput sequencing. Selected buffering hits *NUP160, CNOT1, NXF1* are labeled on the genetic interaction (GI) map.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(C) Results from the GI screen for factors affecting the growth of control or DDX3X knockdown cells. Each point represents the average phenotype of the top 3 guides for each gene. Genes are colored by whether the average phenotype is considered significant in each arm of the screen.

(D) Schematic showing expected growth phenotypes and potential biological pathways from the individual knockdown of DDX3X and hypothetical gene X alone or in combination.

(E) A GI score was calculated for each gene by fitting a quadratic curve to the results of the GI map and calculating the distance from the curve for each gene. A cutoff value of -0.8 and +0.8 was used to categorize buffering or synthetic genes.

(F) Gene ontology enrichment conducted on the buffering genes group in (E).

(G) Schematic of ribosome profiling experiment. The sgDDX3X and sgGal4 cell lines described in (B) were collected for RNA-seq and ribosome profiling after 3 and 14 days of doxycycline treatment.

(H) Results of the ribosome profiling experiment described in (G). Genes are colored by significance. Ribosome profiling counts represent translational efficiency. Selected buffering hits *NUP160, CNOT1, NXF1* from (B) were overlaid on the ribosome profiling results.

(I) Genes with significantly decreased translational changes from ribosome profiling shown in (H) overlaid on the GI map as in (C).

(J) Overview of how the ratio of DDX3 targets to all other mRNAs was calculated. Genes were split into two groups: DDX3-binders and all other genes. The reads per kilobase of exon per million reads mapped (RPKM) for each gene in the knock-down sample was normalized to the RPKM of each gene in the nontargeting sample, after excluding lowly expressed transcripts below RPKM of 20. For statistical testing, Wilcoxon Rank Sum statistical testing was performed on the DDX3X-target and the all-other-mRNAs distributions after averaging replicates. Then the median value of the log2-transformed normalized RPKMs was calculated. The median values were normalized between the DDX3X-target mRNAs and all other mRNAs group by dividing the difference in medians of these two groups by the median value of the all-other mRNAs group. This last ratio is averaged between replicates.

(K) DDX3X and selected buffering hits were depleted and the nuclear and cytoplasmic RNA fractions were sequenced. Data processing and statistical testing (*: p-value < 0.05, ns: not significant) as described in (J). Individual replicates are plotted as points.

(L) Overview of DDX3X domain structure and where point mutations fall in the core helicase DEAD and HELICc RecA-like domains.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(M) Endogenous DDX3X was depleted by CRISPRi and rescued by either overexpression of wild-type DDX3X (DDX3X-wt) or mutant DDX3X, followed by sequencing of the nuclear and cytoplasmic RNA fractions. Data plotted as in (K). The DDX3X-target and the all-other-mRNA distributions were not statistically significant in any of the conditions.

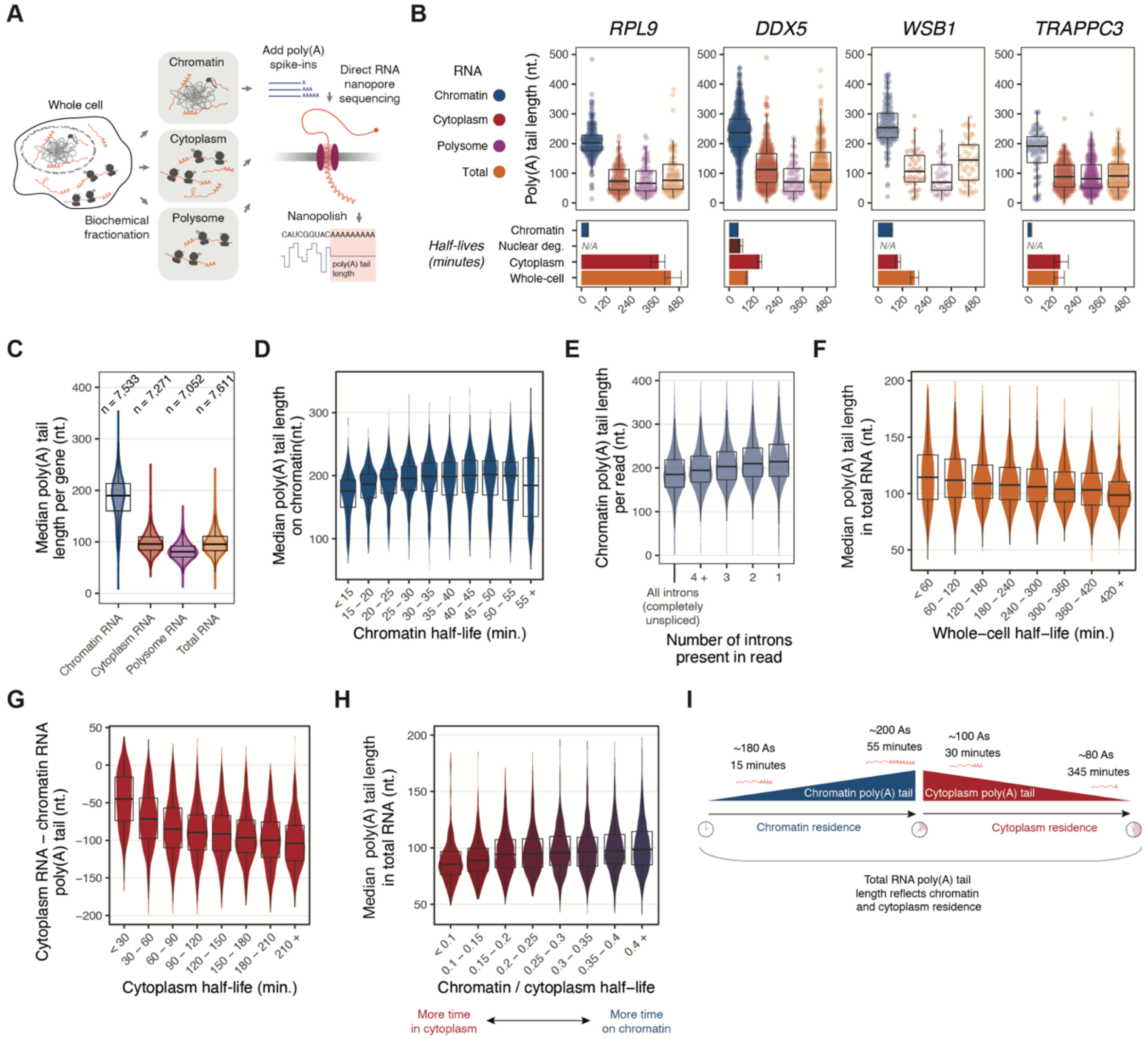


Figure 5: Subcellular compartment-specific poly(A) tail lengths reflect RNA flow rates.

(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.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(D)Median poly(A) tail length for genes covered by >=10 reads in chromatin RNA libraries as a function of chromatin RNA half-life.

(E) 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.

(F) 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.

(G) 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.

(H) 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.

(I) 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.

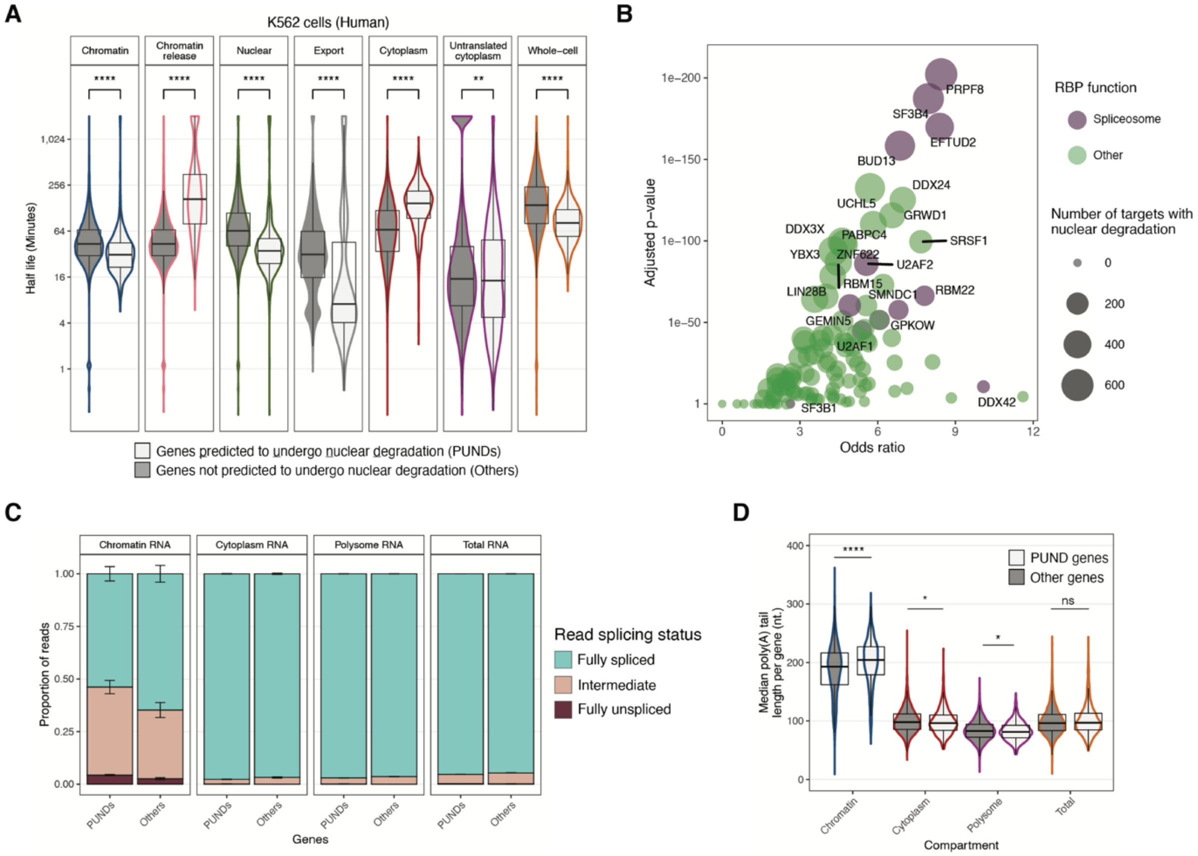


Figure 6: Genes predicted to undergo nuclear degradation (PUNDs) exhibit unique phenotypes related to RNA flow, splicing, and poly(A) tail lengths.

(A) Half-lives of all PUND genes (n=936) compared to all other genes in human K562 cells (****: $p<0.0001$, **: $p<0.01$, *: $p<0.05$, “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. All comparisons were significant ($p<0.0001$, chi-square contingency test)

(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).

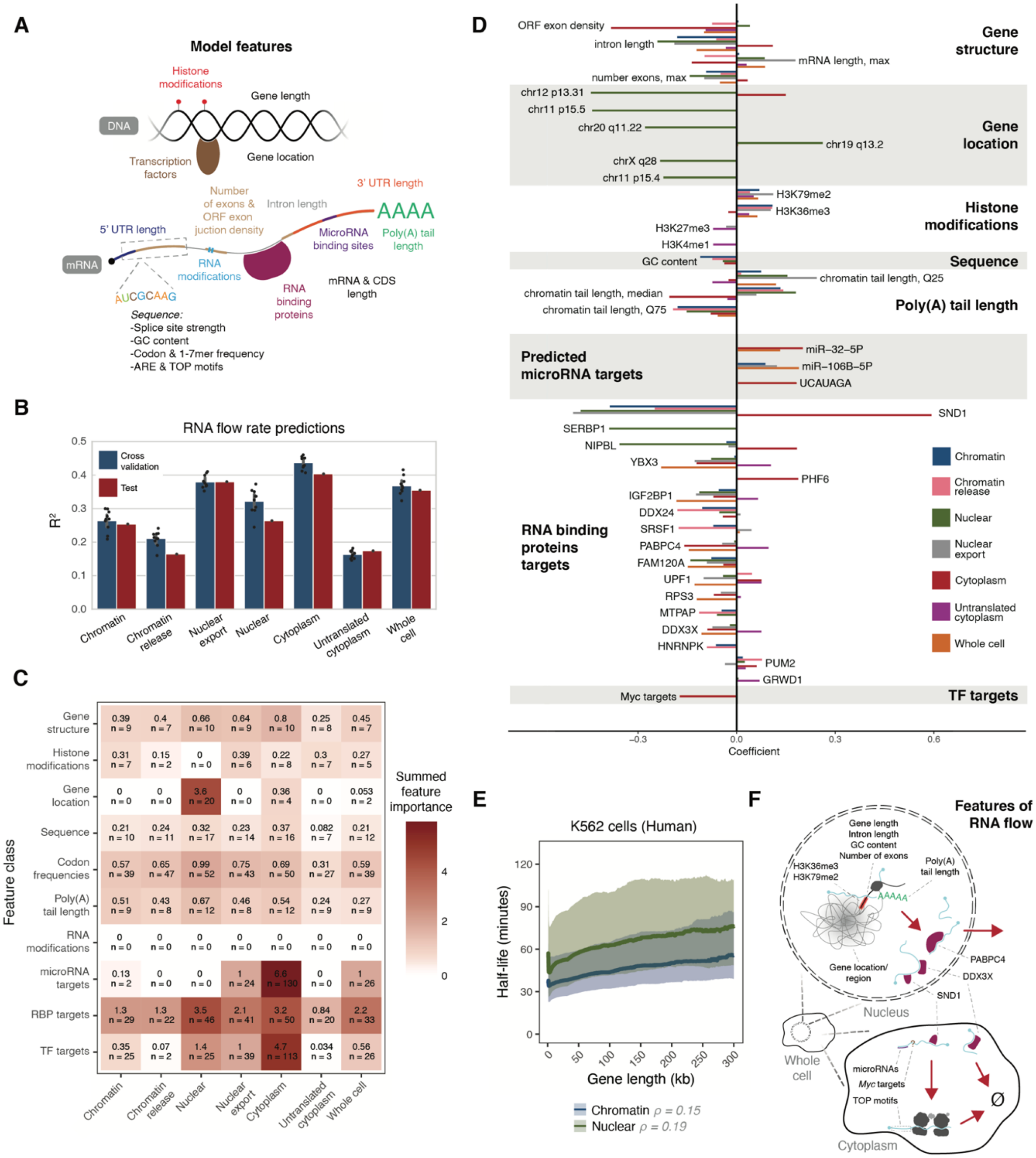


Figure 7: LASSO regression model identifies most relevant genetic and molecular features that predict RNA flow rates.

(A) Schematic representing the genetic and molecular features for each gene included in LASSO model.

(B) 10x cross-validation and test set performances of LASSO models predicting subcellular rates.

(Figure caption continued on the next page.)

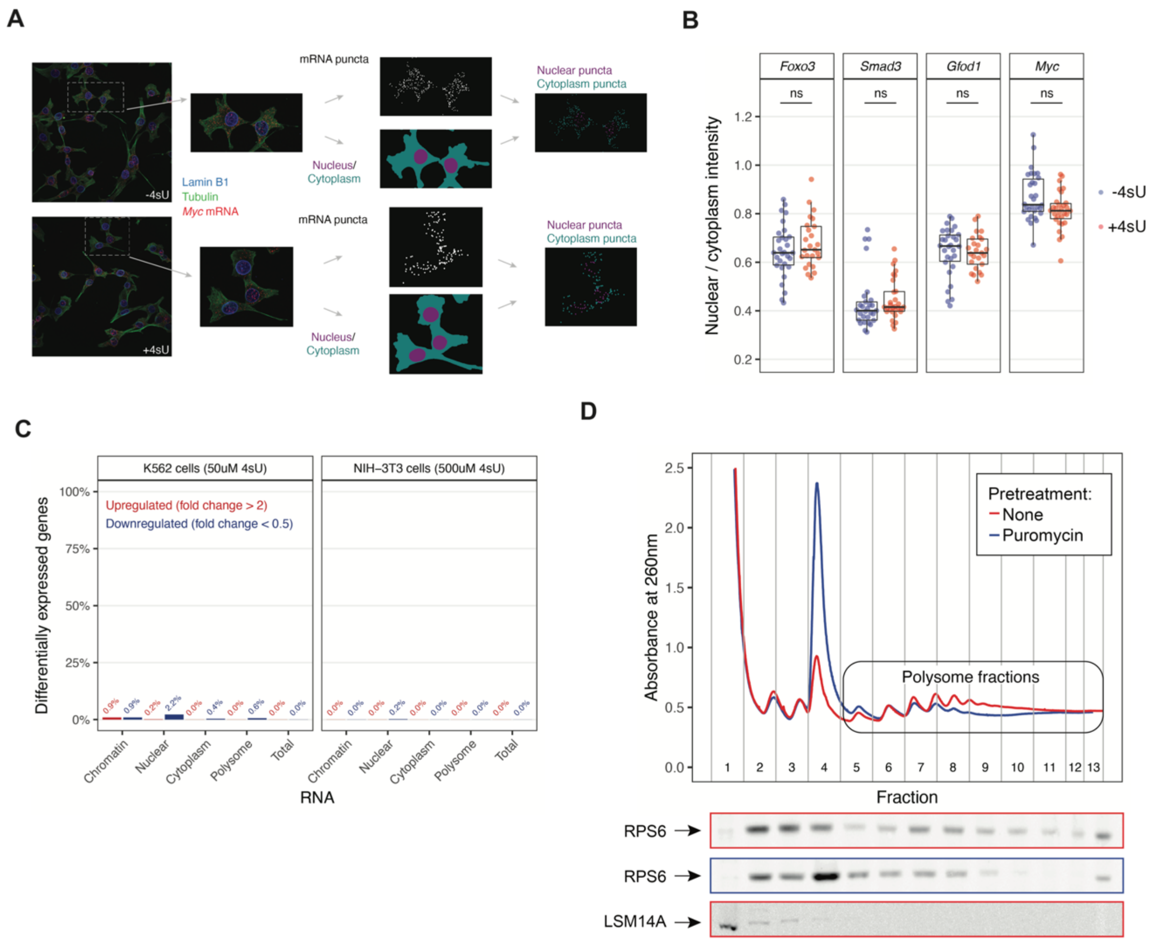
(Figure caption continued from the previous page.)

(C) Class feature importances of LASSO models. The feature importance for each class was summed across the importance of each individual feature within that class, with the number of individual features noted.

(D) 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 as in (C).

(E) 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.

(F) 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).



Supplemental Figure 1: Optimization of biochemical fractionation and 4-thiouridine labeling conditions used in subcellular TimeLapse-seq, related to Figure 1.

(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 fields of view were analyzed for all genes.

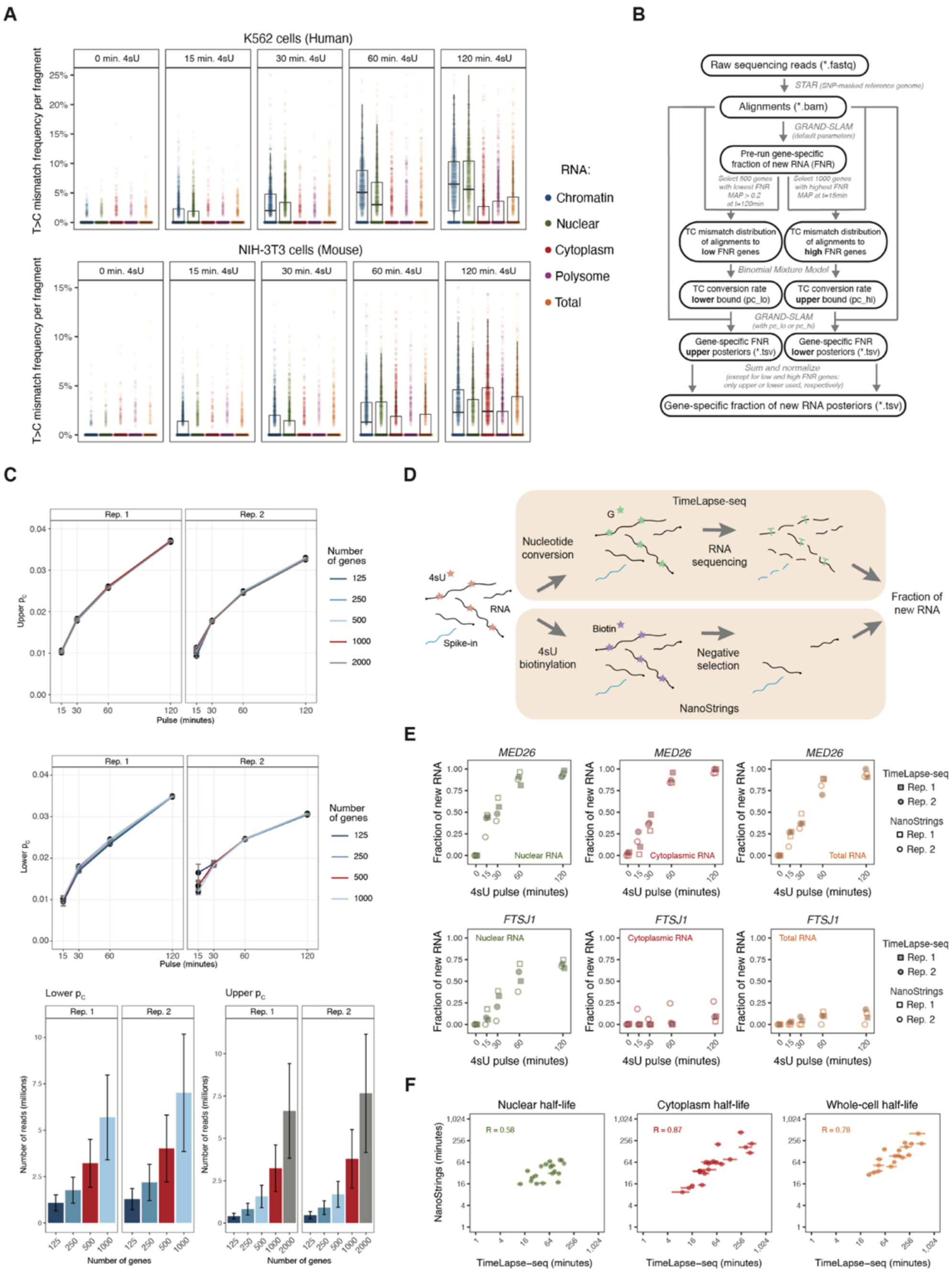
(B) Summary of data for all genes *(Myc, Foxo3, Smad3, Gfod1)* analyzed according to (A). A t-test was performed to compare the differences in intensities between 4sU-treated and control cells ("ns:" non-significant).

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(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.



Supplemental Figure 2: Nucleotide conversion analysis estimates the fraction of new RNA in a compartment- and time point-specific manner, related to Figure 1. (A) 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 fragment (using both forward and reverse reads). A total of 1,000 reads were analyzed for each sample. Dots represent individual fragments.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

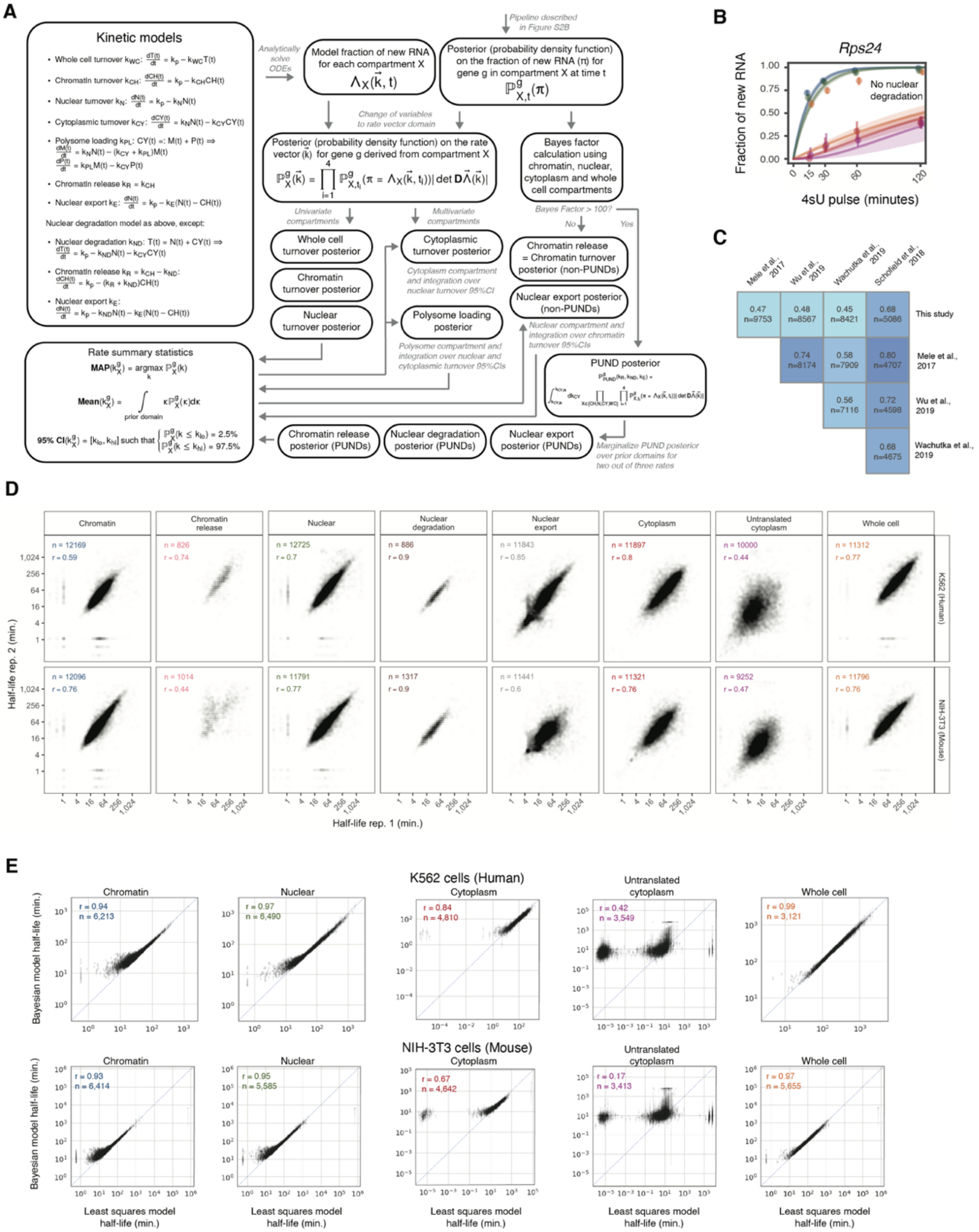
(B) Computational analysis pipeline for subcellular TimeLapse-seq data (see Methods for details).

(C) The upper and lower p_C values were calculated for NIH-3T3 total RNA libraries using reads that aligned to a varying number of genes. Top: upper and lower p_C values. Error bars represent the standard deviation between biological replicates. Bottom: Number of sequencing reads corresponding to the number of genes. Error bars represent the standard deviation between biological replicates.

(D) 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 as per (B). Bottom: NanoStrings-based analysis starts with the covalent biotinylation 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 NanoString 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).

(E) 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.

(F) Correlation of fraction of new RNA between NanoStrings and subcellular TimeLapse-seq for all genes and compartments. The data shown in (D) 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 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.



	Mele et al., 2017	Wu et al., 2019	Wachutka et al., 2019	Schofield et al., 2018
This study	0.47 n=9753	0.48 n=8567	0.45 n=8421	0.68 n=5086
Mele et al., 2017		0.74 n=8174	0.58 n=7909	0.80 n=4707
Wu et al., 2019			0.56 n=7116	0.72 n=4598
Wachutka et al., 2019				0.68 n=4675

Supplemental Figure 3: RNA flow can be modeled using a series of ordinary differential equations in a Bayesian framework, related to Figure 1.
(A) Schematic of RNA kinetic modeling (see Methods for more detail).

(Figure caption continued on the next page.)

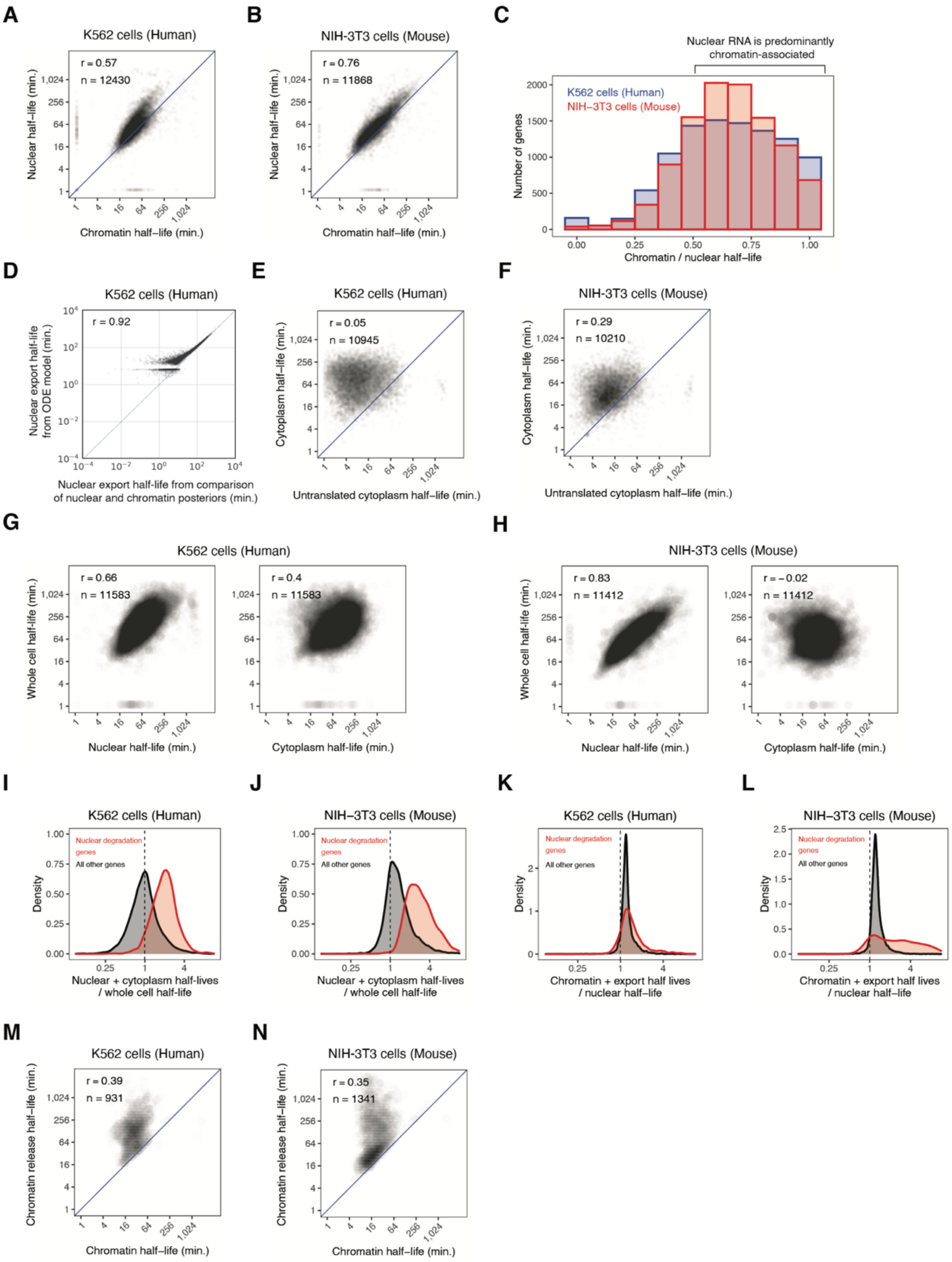
(Figure caption continued from the previous page.)

(B) *Rps24* model fits in the absence of nuclear degradation, related to **Figure 1F** (subcellular TimeLapse-seq in mouse NIH-3T3). Colors are consistent with **Figure 1F**.

(C) Correlation between whole-cell (total RNA) half-lives measured in K562 cells in this study and previous studies. The mean whole-cell half-life between replicates in this study was used for comparison. Pearson’s correlation is shown and the number of genes is noted.

(D) Reproducibility of RNA flow rates across biological replicates. For each flow rate, the mean half-life is compared where they could be determined for both replicates. Nuclear degradation and chromatin release half-lives are only shown for genes best explained by this model; export rates are shown separately for genes modeled with and without nuclear degradation. Pearson's correlation is shown and the number of genes is noted. Each dot represents one gene. CI values refer to the fraction of genes that have overlapping 95% credible intervals of the RNA flow rate between both replicates, and thus reflect an additional measure of reproducibility between replicates.

(E) Comparison between the Bayesian model and a least squares model. The Bayesian MAP half-life for each subcellular compartment for one replicate is compared with the least squares estimate. The least squares model fails for compartments that depend on multiple rates, in particular untranslated cytoplasm, unlike the Bayesian model. 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.



Supplemental Figure 4: RNA flow rates show consistent genome-wide patterns between cell lines, related to Figure 1.

(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. Each dot represents one gene.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(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. Data for both human K562 and mouse NIH-3T3 cells is shown as a histogram.

(D) Correlation between the nuclear export half-lives as calculated by comparing the nuclear and chromatin posteriors versus an ODE model (Fig. S3A) in K562 cells. Spearman correlation is noted.

(E) 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. Each dot represents one gene.

(F) Same as (E) in mouse NIH-3T3 cells.

(G) 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 (right) are compared to mean whole-cell half-lives with the Pearson correlation and number of genes shown. Each dot represents one gene.

(H) Same as (G) in mouse NIH-3T3 cells.

(I) Density distribution of the “predicted” whole-cell half-life (the sum of the nuclear and cytoplasm half-lives) divided by the observed whole-cell half-life for each gene in human K562 cells. Genes with model fits without nuclear degradation are shown in gray and genes with model fits including nuclear degradation are shown in red.

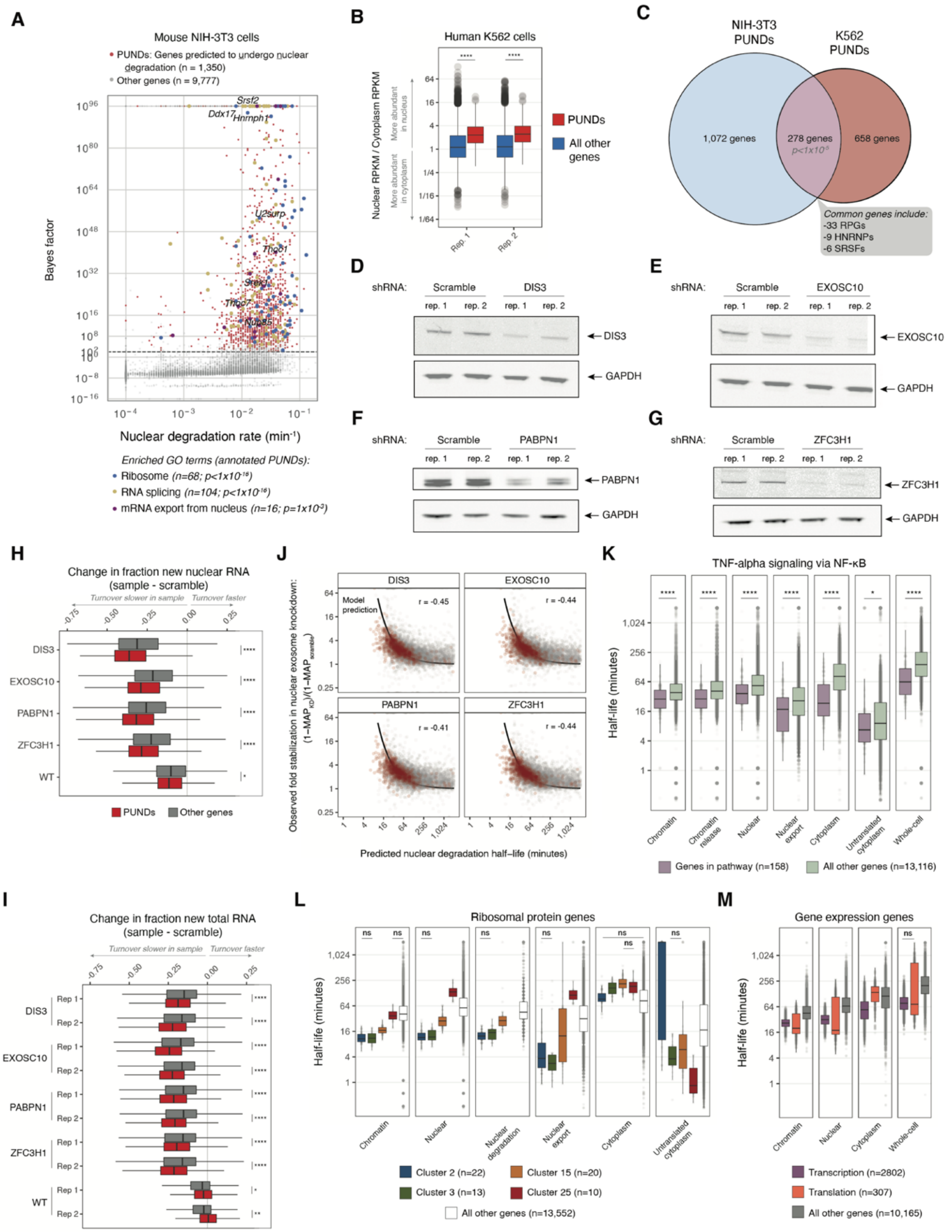
(J) Same as (I) in mouse NIH-3T3 cells.

(K) Density distribution of the “predicted” nuclear half-life (the sum of the chromatin and export half-lives) divided by the observed nuclear half-life for each gene in human K562 cells. Colors are consistent with (I).

(L) Same as (K) in mouse NIH-3T3 cells.

(M) Correlation between chromatin and chromatin release half-lives for genes modeled with nuclear degradation in human K56 cells. Mean half-lives for each rate are compared with the Spearman correlation and number of genes noted. Each dot represents one gene.

(N) Same as (M) in mouse NIH-3T3 cells.



Supplemental Figure 5: PUND stabilization upon nuclear exosome depletion and functionally related genes exhibit similar RNA flow, related to Figure 2.
(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 2A.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(B) Comparison of relative nuclear to cytoplasm RNA abundance between PUNDs and all other genes (****: p<0.0001, Wilcoxon test).

(C) 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).

(D) Confirmation of DIS3 protein knockdown in K562. Cells were transduced with lentivirus-containing plasmids expressing a scrambled shRNA sequence or one targeting DIS3, expression of the shRNA was induced with doxycycline treatment for 48 hours, and knockdown efficiency was monitored by western blotting in samples collected for subcellular TimeLapse-seq.

(E) Same as (D) for EXOSC10.

(F) Same as (D) for PABPN1.

(G) Same as (D) for ZFC3H1.

(H) Same as **Figure 2B** for the second biological replicate (****: p<0.0001, ***: p<0.001, **: p<0.01, *: p<0.05, Wilcoxon test).

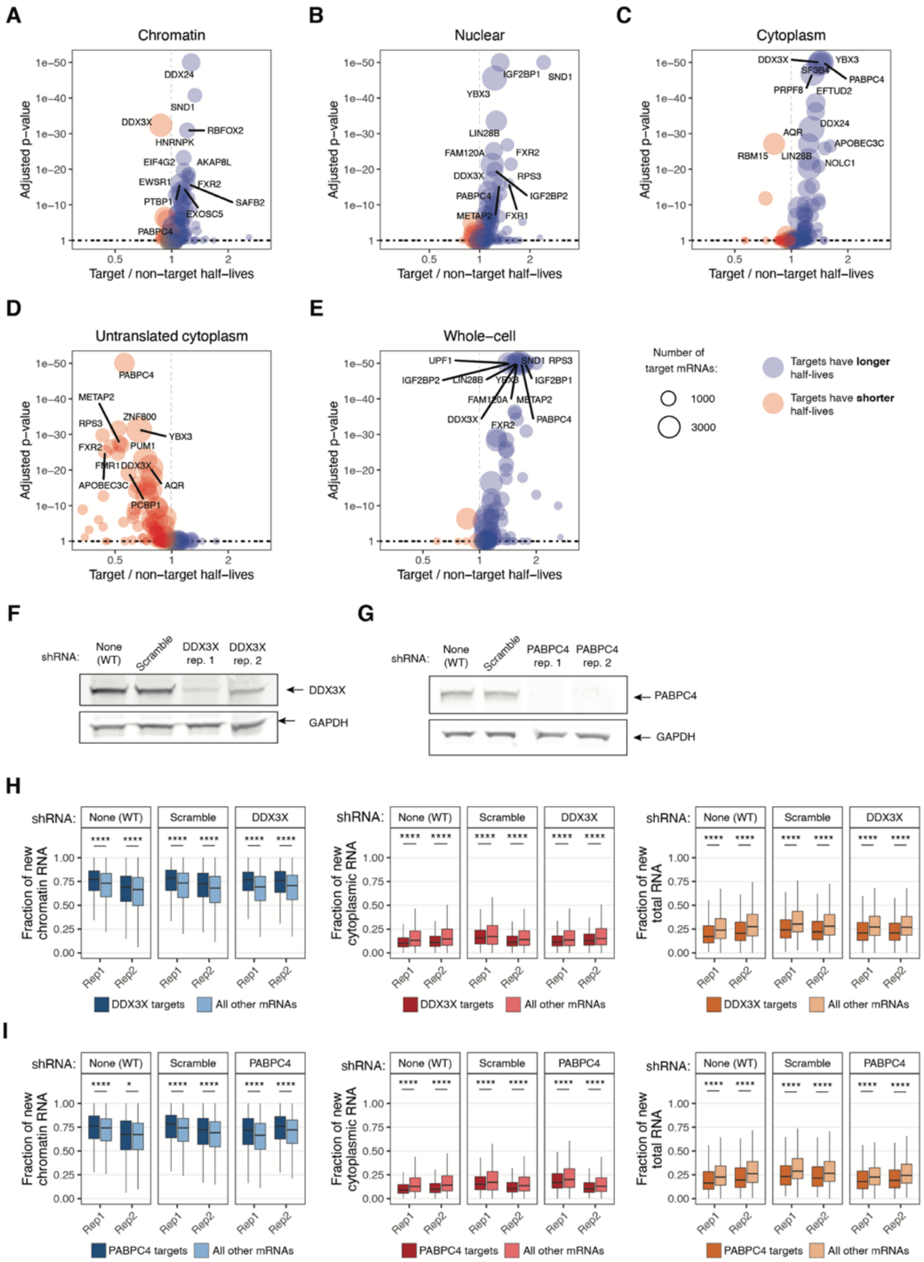
(I) Same as **Figure 2B** and (H) for total RNA.

(J) Same as **Figure 2C** for all genes. PUNDs are shown in red and all other genes are shown in gray.

(K) RNA flow rates 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 as in (H).

(L) Half-lives of ribosomal protein genes (RPGs) from each cluster where they were enriched (clusters 2, 3, 15, and 25 in **Figure 2D**). The number of RPGs within each cluster is noted. Nuclear degradation half-lives are only shown for genes best explained by this model. Non-significant comparisons are indicated as "ns" and all other comparisons were statistically significant (p<0.05, Wilcoxon test).

(M) Half-lives of clusters 4, 5, and 6, containing genes related to transcription, and clusters 2,3, and 25, containing genes related to translation. The number of genes in each group is noted, and significance was noted as in (L).



Supplemental Figure 6: RNA binding proteins are associated with RNA flow rates, related to Figure 3.

(A) Comparison of chromatin RNA half-lives between target and non-target mRNAs of each RBP according to Fig. 3A. 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.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(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.

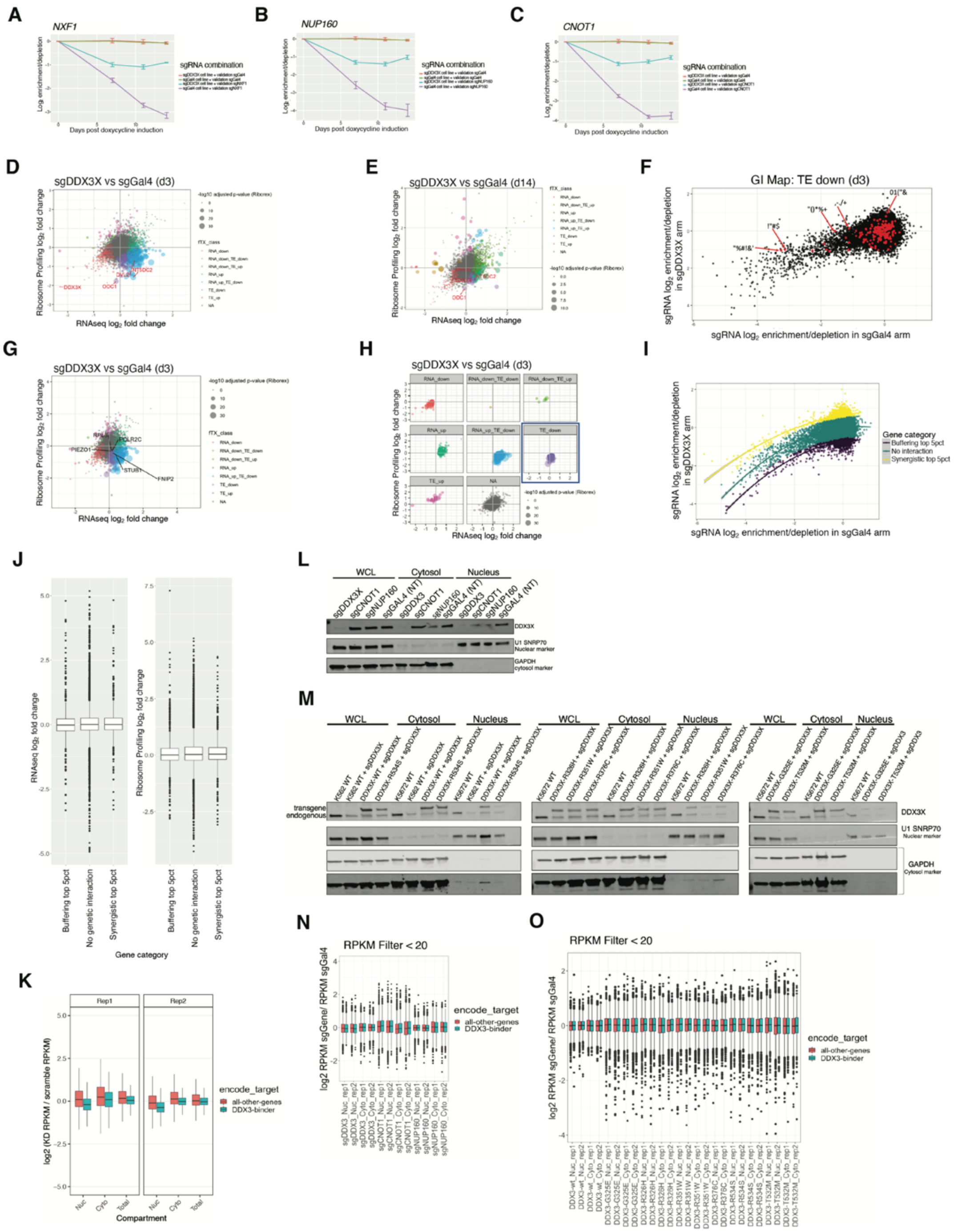
(F) The proportion of eCLIP peaks that fall within the first exon of target transcripts out of all targets for each RBP. RBPs are sorted from highest to lowest proportion.

(G) Confirmation of DDX3X and PABPC4 protein knockdown in K562. Cells were transduced with lentivirus-containing plasmids expressing a scrambled shRNA sequence or one targeting the RBP of interest 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.

(H) Fraction of new RNA in chromatin, cytoplasm, and whole-cell following DDX3X knockdown for target and non-target mRNAs. Fraction of new RNA MAP values were compared between targets and non-targets with a Wilcoxon test (****: $p<0.0001$, *: $p<0.05$).

(I) The proportion of eCLIP peaks that fall within the last exon of target transcripts out of all targets for each RBP. RBPs are sorted from highest to lowest proportion.

(J) Same as (H) for PABPC4.



Supplemental Figure 7: Genetic interaction mapping to link DDX3X depletion to mRNA export defects, related to Figure 4.

(A) Competition assay between cells expressing a sgRNA targeting Gal4 alone, Gal4 and NXF1, DDX3X alone, and DDX3X and NXF1. Enrichment or depletion was calculated by comparing the percentage of cells expressing the guide RNA at each time point versus day 0.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(B) Same as (A) for NUP160.

(C) Same as (A) for CNOT1.

(D) Replicate plots for the CRISPR screen. Gamma phenotype values are calculated as described in **Figure 4B**. Non-targeting sgRNAs are overlaid in gray. sgGal4 screen replicates had a Pearson correlation coefficient of r=0.66 at the individual sgRNA level and r=0.81 when the top 3 sgRNA phenotypes for each gene were averaged. sgDDX3X screen replicates had a Pearson correlation coefficient of r=0.25 at the individual sgRNA level and r=0.36 when the top 3 sgRNA phenotypes for each gene were averaged.

(E) Ribosome profiling results at 3 days post-infection with the sgDDX3X guide. Colors indicate genes that with significant changes in RNA abundance and/or translation efficiency. Overlaid are genes previously described to have DDX3X-dependent translation.

(F) Same as (E) at 14 days post-infection.

(G) Gene with significantly decreased translational changes in (D-E) overlaid on the GI map with those in the buffering range annotated.

(H) Same as (E) with genes from (G) annotated.

(I) Ribosome profiling results from **Figure 4H** plotted by significance category.

(J) A quadratic curve was fitted to the results of the GI map in (**Figure 4E**) and the distance from the curve for each gene was calculated. A distance-from-curve cutoff value of the top 5% and bottom 5% was used to categorize buffering or synthetic genes.

(K) The change in RNA abundance and translational efficiency for the categories identified in (J).

(L) The change in nuclear, cytoplasmic, and total RNA abundance for DDX3X targets and all other genes following constitutive DDX3X depletion (as shown in **Figure S6G**).

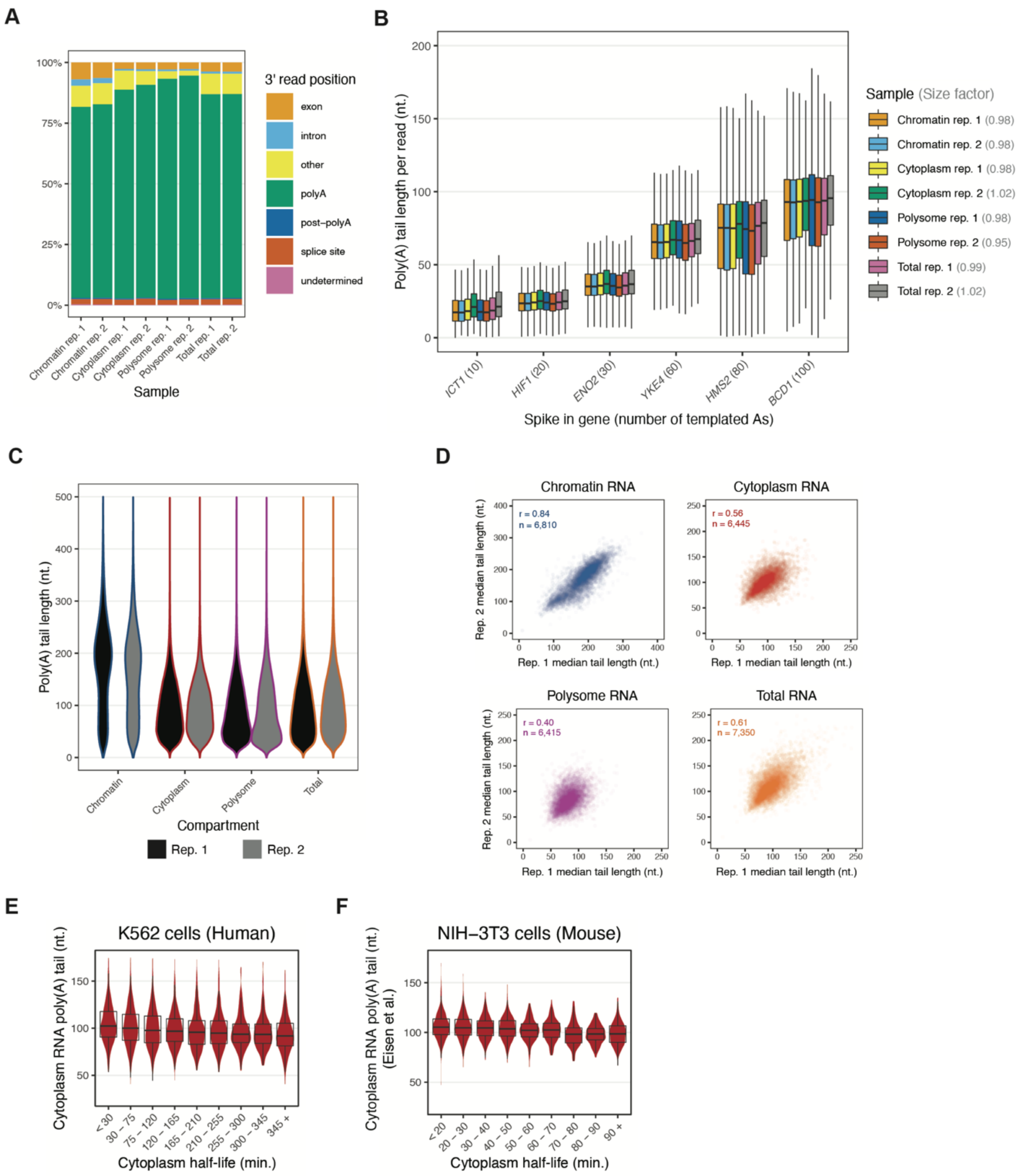
(M) The change in relative nuclear to cytoplasmic RNA abundance for DDX3X targets and all other genes following constitutive DDX3X depletion (as shown in **Figure S6G**) (****: $p<0.0001$, Wilcoxon test).

(N) Western blot from the individual sgRNA knock-down validating subcellular fractionation. U1 SNRP70 was used as a nuclear marker and GAPDH was used as a cytoplasm marker. WCL indicates whole-cell lysate. 1 representative replicate of 2 replicates is shown.

(O) Same as (L) for the DDX3X rescue.

(P) The difference in subcellular abundances for DDX3X targets and all other genes. The RPKMs were normalized to a non-targeting sgRNA sample.

(Q) Same as (N) for the DDX3X rescue. RPKMs were normalized to an uninfected sample.



Supplemental Figure 8: Poly(A) tail lengths are related to RNA flow rates, related to Figure 5.

(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).

(Figure caption continued on the previous page.)

(Figure caption continued from the previous page.)

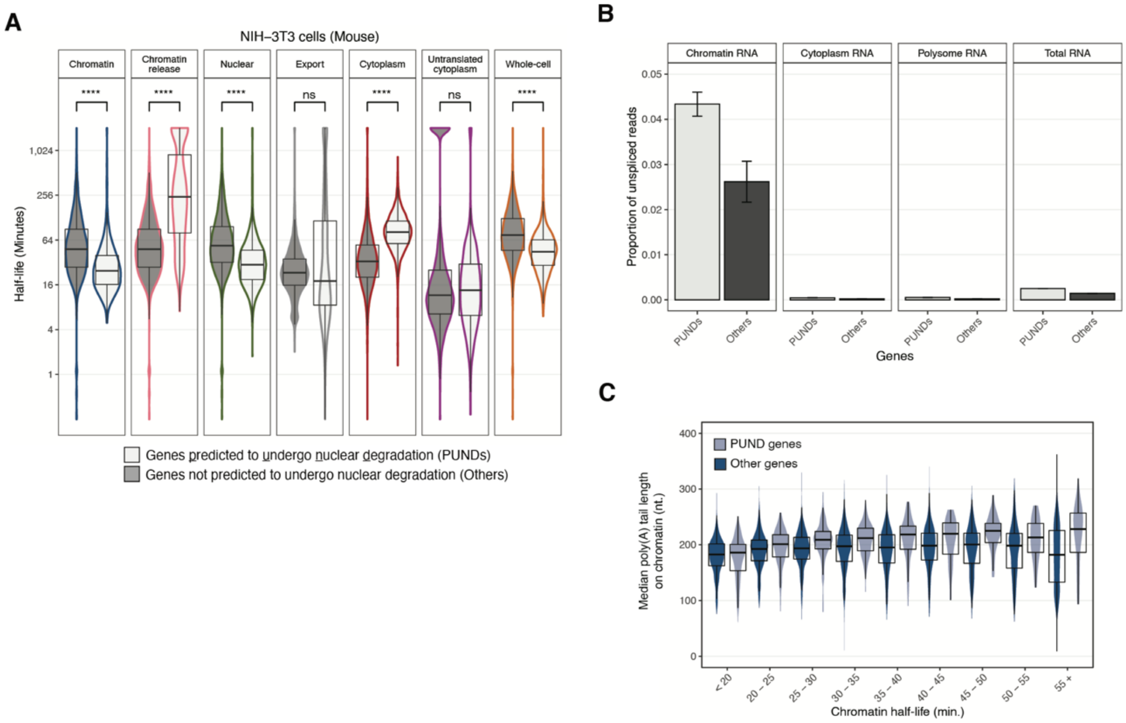
(B) Distribution of poly(A) tail lengths for synthetic spike-in RNAs across nanopore sequencing runs. Six transcripts (shown on the x-axis) from *S. cerevisiae* with templated poly(A) tails ranging from 10 to 100 nucleotides were transcribed in vitro and added to each sample prior to nanopore library 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.

(E) 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.

(F) 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).

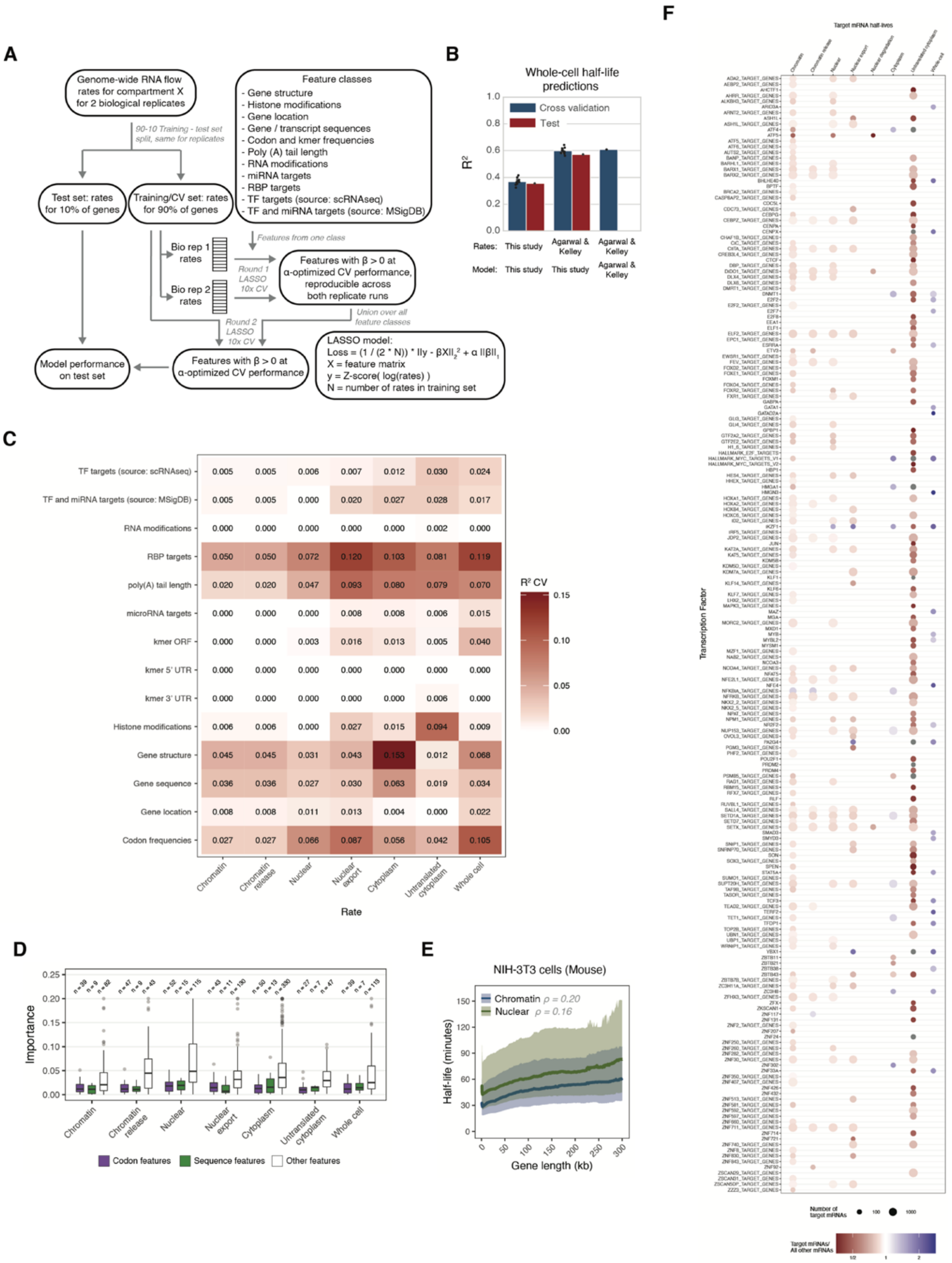


Supplemental Figure 9: PUND phenotypes, related to Figure 6.

(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, “ns:” not significant).

(B) The proportion of unspliced reads in each compartment for PUNDs and all other genes. All comparisons were significant (p<0.0001, Fisher’s exact test).

(C) 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.



	Chromatin	Chromatin release	Nuclear	Nuclear export	Cytoplasm	Untranslated cytoplasm	Whole cell
TF targets (source: scRNAseq)	0.005	0.005	0.006	0.007	0.012	0.030	0.024
TF and miRNA targets (source: MSigDB)	0.005	0.005	0.000	0.020	0.027	0.028	0.017
RNA modifications	0.000	0.000	0.000	0.000	0.000	0.002	0.000
RBP targets	0.050	0.050	0.072	0.120	0.103	0.081	0.119
poly(A) tail length	0.020	0.020	0.047	0.093	0.080	0.079	0.070
microRNA targets	0.000	0.000	0.000	0.008	0.008	0.006	0.015
kmer ORF	0.000	0.000	0.003	0.016	0.013	0.005	0.040
kmer 5' UTR	0.000	0.000	0.000	0.000	0.000	0.000	0.000
kmer 3' UTR	0.000	0.000	0.000	0.000	0.000	0.006	0.000
Histone modifications	0.006	0.006	0.000	0.027	0.015	0.094	0.009
Gene structure	0.045	0.045	0.031	0.043	0.153	0.012	0.068
Gene sequence	0.036	0.036	0.027	0.030	0.063	0.019	0.034
Gene location	0.008	0.008	0.011	0.013	0.004	0.000	0.022
Codon frequencies	0.027	0.027	0.066	0.087	0.056	0.042	0.105

Supplemental Figure 10: LASSO model predictions, related to Figure 7.
(A) Schematic of LASSO feature selection and model training, 10x cross-validation (CV) and testing.

(Figure caption continued on the next page.)

(Figure caption continued from the previous page.)

(B) Comparison of our LASSO 10x CV and test performance with alternative whole-cell turnover rate estimate and models.

(C) Heatmap showing the CV coefficient of variation (R^2) of round 1 LASSO feature selection per class, as described in (A), across all rates.

(D) The importance of individual sequence and codon features compared to individual features of other families.

(E) Continuous averages of chromatin and nuclear half-lives as a function of gene length in mouse NIH-3T3 cells (see Methods). 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.

(F) 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 (see Methods for details).

TABLES

Table 1: Primers and Protospacers

Experiment	Identifier	Sequence (5' - 3')
Poly(A) spike-ins	BCD1_F1	GCCATCAGATTGTGTTTGTTAGTCGCTATGGCGGTGTTGTGTGGTGTATG
	BCD1_R1	GCTTACGGTTCACTACTCACGACGATGTCATGCAGTGAGGAAATCCATGG
	ICT1_F1	GCCATCAGATTGTGTTTGTTAGTCGCTATGTGGACAAACACTTTCAAATG
	ICT1_R1	GCTTACGGTTCACTACTCACGACGATGTTACTTTGACAGGAACGAGACTA
	HIF1_F1	GCCATCAGATTGTGTTTGTTAGTCGCTATGAAACTAAGGGCAGAAGACGT
	HIF1_R1	GCTTACGGTTCACTACTCACGACGATGTCAATGCCTTCTAGGCTTCTTCT
	YKE4_F1	GCCATCAGATTGTGTTTGTTAGTCGCTATGAAGGCGTCGCACATTTGCTC
	YKE4_R1	GCTTACGGTTCACTACTCACGACGATGTCAATGCTCATCCATGAGCGCCA
	HMS2_F1	GCCATCAGATTGTGTTTGTTAGTCGCTATGGATGCAACATCGAGGATGGA
	HMS2_R1	GCTTACGGTTCACTACTCACGACGATGTCACGTTCGAAGATGTTTGGAAA

Experiment	**Identifier**	**Sequence (5' - 3')**
	ENO2_F1	GCCATCAGATTGTGTTTGTTAGTCGCTATGGCTGTCTCTAAAGTTTACG
	ENO2_R1	GCTTACGGTTCACTACTCACGACGATGTTACAACTTGTCACCGTGGTGG
	T7_F	TAATACGACTCACTATAGGGAGAGCCATCAGATTGTGTTTGTTAGTCGCT
	polyT_10_R	TTTTTTTTTTGCTTACGGTTCACTACTCACGACGATG
	polyT_15_R	TTTTTTTTTTTTTTTGCTTACGGTTCACTACTCACGACGATG
	polyT_30_R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTTACGGTTCACTACTCACGACGATG
	polyT_60_R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTTACGGTTCACTACTCACGACGATG
	polyT_80_R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCTTACGGTTCACTACTCACGACGATG
	polyT_100_R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT

Experiment	Identifier	Sequence (5' - 3')
		TTTTTGCTTACGGTTCACTACTCACGACGATG
NanoStrings spike-in	ERCC48_F	TAATACGACTCACTATAGGGAGA
	ERCC48_R	CAGGTCACTTTTCTGTCGTAGC
NanoStrings probes	FTSJ1_probeA	CCGGCCTCGGAATATCTTGGCCACAAAGCAGCCCCCCTCAAGACCTAAGCGACAGCGTGACCTTGTTTCA
	RPS6KA3_probeA	AACTTTCAGTGTGGCCTTCTTCAATACCTTCATGGCATAAAGCTGCCTAGCATCCTCTTCTTTTCTTGGTGTTGAGAAGATGCTC
	MYLK3_probeA	CCTCCGAGAGCCCTTCAAAGGTGTCAGCATCAAAATCCCAGCTACAGTTTCACAATTCTGCGGGTTAGCAGGAAGGTTAGGGAAC
	MCM8_probeA	GCTGGTGGGGAATGGGATCTATTGTTTCTCCAGGAACCACCTTTAGTCTTCTGTTGAGATTATTGAGCTTCATCATGACCAGAAG
	UHMK1_probeA	CCTCCAAAACATCTCGGGCACAATGCTGTATCATCCACATGGAACAACCCCAAAGACGCCTATCTTCCAGTTTGATCGGGAAACT
	DDX50_probeA	TGCTGGCCTCTGAGACCAATGACACTGGATGGCCAAATGTTCCACAGCGAA

Experiment	Identifier	Sequence (5' - 3')
		CCTAACTCCTCGCTACATTCCTATTGTTTTC
	ABHD5_probeA	AGCATAGACAGGTCTGTTGGTGCAAAGATCTCCAAAATTCAGTGCCCAGACCAATTTGGTTTTACTCCCCTCGATTATGCGGAGT
	ZSCAN22_probeA	GAGTGACACAGTCAGATCTAAGTAATATCAAGAGTCCCCTGGAGAGTGATCTTTCGGGTTATATCTATCATTTACTTGACACCCT
	mt-co1_probeA	GCGGAGGTGAAATATGCTCGTGTGTCTACGTCTATTCCTACTGTAAATATCAACAGCCACTTTTTTTCCAAATTTTGCAAGAGCC
	RBM15_probeA	GTCACCCTGCAACAGATGCATGTTGGAAGGAAAGTTGCTGTTCTTCAGTACACCGTGTGGACGGCAACTCAGAGATAACGCATAT
	FOXO3_probeA	AGAGATTCAAATGCTCCTTATTCTTTAAGTGGGGTTCACTAGGTAGAGGCCCTGGAGTTTATGTATTGCCAACGAGTTTGTCTTT
	SOX5_probeA	CACTTCCATCAGAATCTCCACTCAGATTGAAATCCATCATTGCATGGCTACAGATAAGGTTGTTATTGTGGAGGATGTTACTACA
	SHOX2_probeA	GGGAAGAGGGCCGGCTCCCGAGGTCTCAAAGGGGTAACGGAGAAGCTTC

Experiment	Identifier	Sequence (5' - 3')
		CTTCCTGTGTTCCAGCTACAAACTTAGAAAC
	UPF2_probeA	GGTGTGAATATATCAATTCCAGGCCCATGTTCTTCTGGTGTTGGTTTGTCCATAAAATTGGTTTTGCCTTTCAGCAATTCAACTT
	MYC_probeA	CGCTCCAAGACGTTGTGTGTTCGCCTCTTGACATTCTCCTCGGTGCTGGTCAAGACTTGCATGAGGACCCGCAAATTCCT
	BRCA2_probeA	TTTGGCCGGAGTAAGCTGACAAAAACCGCGCCGGTCACAAATCTGTCCCCCTTTCGTTGGGACGCTTGAAGCGCAAGTAGAAAAC
	HP1BP3_probeA	CTCTGAACTTATGTCTGGTTCTGGCTTTTCTTCCTCCCAGCAGACCTGCAATATCAAAGTTATAAGCGCGT
	TOB1_probeA	GAGAAGTACGTGCAACCTTGTTGCTACGGCCACTATTCTTCATTTTGGTACCTGCCAATGCACTCGATCTTGTCATTTTTTTGCG
	MED26_probeA	CCCTCGCCCCACAGTACAGCATAAAACCAGTAGCACCCACAATAACTTTTCAAACTGGAGAGAGAAGTGAAGACGATTTAACCCA
	JUN_probeA	ATAAGATTTGCAGTTCGGACTATACTGCCGACCTGGCTGGCTGGCTGCGA

Experiment	Identifier	Sequence (5' - 3')
		TTGCTGCATTCCGCTCA ACGCTTGAGGAAGTA
	CHD8_probeA	GACCTGAATGATTCTGC AATTCTTTTGTCTTGCCA TTTTCAGCTGGGCTGAG GCTGTTAAAGCTGTAGC AACTCTTCCACGA
	CDCA2_probeA	CAGGAACAGGTGACTG CTCAAGCAGCAGGGAA CTGAGACCACTGAAGTC TCTAGGACGCAAATCAC TTGAAGAAGTGAAAGCG AG
	ERCC-00048_probeA	GCTTACGAAGCAGGCGT ACGGGTTGACCATGTCA CTATATGTCGTCCGTTC CACGCGATGACGTTCGT CAAGAGTCGCATAATCT
	MALAT1_probeA	CCTGAAAGTGCTCACAA GGCAAATCGCCATGGAA AGCGAGTTCAAGTGGCC ATTTGGAATGATGTGTA CTGGGAATAAGACGACG
	FTSJ1_probeB	CGAAAGCCATGACCTCC GATCACTCGAGAAGAAG ACCTGCAGCTGGCTGTA GAGGAGCGTCACATC
	RPS6KA3_probeB	CGAAAGCCATGACCTCC GATCACTCTTAACCTCT ACCAAGATATCACGTTC CATTTTTGTCCGAACTC GGTCTCG
	MYLK3_probeB	CGAAAGCCATGACCTCC GATCACTCTCTGCAGCT CTTCTCTTTGACCAGCA

Experiment	Identifier	Sequence (5' - 3')
		ACCGGGAAACAAAGTCCTTGGCCT
	MCM8_probeB	CGAAAGCCATGACCTCCGATCACTCTAGCCTTGGGTACACATACTGCCGAGCATAGCCAATGTACTTTCTCAATA
	UHMK1_probeB	CGAAAGCCATGACCTCCGATCACTCTGGTTTGAGGTCCGCATGGACATAGCCCTCATGATGAAGAAAAGCAAGGG
	DDX50_probeB	CGAAAGCCATGACCTCCGATCACTCATAGCCCTCCCTTCAGACCCACTGTAGACTTGAAGGACATCTCCAATAAC
	ABHD5_probeB	CGAAAGCCATGACCTCCGATCACTCTCACTGTCAAACCTGGGTCTACTACTTCGTCCAAAACCCAATAGGTCAAA
	ZSCAN22_probeB	CGAAAGCCATGACCTCCGATCACTCTGTTCAGGAAGCTGTGCTGTAGAAGATGGGTTGGAGAGCTTTCCACCTTG
	mt-co1_probeB	CGAAAGCCATGACCTCCGATCACTCCGAGTCAGCTAAATACTTTGACGCCGGTGGGGATAGCGATGATTATGGTA
	RBM15_probeB	CGAAAGCCATGACCTCCGATCACTCACTTTGCCTCCAGTTGAACCCTCCAC

Experiment	**Identifier**	**Sequence (5' - 3')**
		AAGAAGACTACTAGCCACTTGGAG
	FOXO3_probeB	CGAAAGCCATGACCTCCGATCACTCCAACATGGAGGACTTCTTTTTGACTGCTTTATTCTTCATGGCCTTTTCCA
	SOX5_probeB	CGAAAGCCATGACCTCCGATCACTCTACCACGCCCTCGGGATTCCCTATAAATTCTTGACTCTGAGACTCCAG
	SHOX2_probeB	CGAAAGCCATGACCTCCGATCACTCTCCGAGTCCAAGATGCGATAGGGGACGAGGGATGGTCAGTGAGGC
	UPF2_probeB	CGAAAGCCATGACCTCCGATCACTCCTTCATCTTCCCATATACCACCTTCCAAGTCATATTCTCCAGGTTTACCA
	MYC_probeB	CGAAAGCCATGACCTCCGATCACTCTCTGGTCACGCAGGGCAAAAAAGCTCCGTTTTAGCTCGTTCCTCCTCTGG
	BRCA2_probeB	CGAAAGCCATGACCTCCGATCACTCTATTCCTCCAATGCTTGGTAAATAAGTCCGCTCCAGAGGTGCAGTTCTTT
	HP1BP3_probeB	CGAAAGCCATGACCTCCGATCACTCGTGGAGTTTCATTCTCTTGTTCTTCTACAGTGGAGACAGATTC

Experiment	Identifier	Sequence (5' - 3')
	TOB1_probeB	CGAAAGCCATGACCTCCGATCACTCGATGGCTTTCTGCTTCAAGAGGTCATTCACATTCAAGCCGAGGTTGATGG
	MED26_probeB	CGAAAGCCATGACCTCCGATCACTCGAATCTAAAGGAATGGTTATGAGTGTTTCCAGCGTACAGACAGGTCTTCC
	JUN_probeB	CGAAAGCCATGACCTCCGATCACTCGGGAGCCACAGGCGCTAGCTCTGGGCAGTTAGAGAGAAGGTGAAAAGAAA
	CHD8_probeB	CGAAAGCCATGACCTCCGATCACTCTTGTGACTTTACTTTTTTTCCTTTGCGTCCACGAGGCACAGGGATAGATA
	CDCA2_probeB	CGAAAGCCATGACCTCCGATCACTCGTTTTCAAGATTCTCCCCCTTGTCATCAAAATCTGGTTGAGGTAATGGCT
	ERCC-00048_probeB	CGAAAGCCATGACCTCCGATCACTCAATAGTAAGCACAAACGACATGATTCGCCAGCCTCTTCAAAGGACCGTCG
	MALAT1_probeB	CGAAAGCCATGACCTCCGATCACTCCCCACGGCCCGCACGGAAATTTTTCTACCGTTTTTCAGCTTCCAGGCTCT

Experiment	**Identifier**	**Sequence (5' - 3')**
SABER: Probe synthesis	Clean.g	CCCCGAAAGTGGCCTCGGGCCTTTTGGCCCGAGGCCACTTTCG
	Hairpin_25 (used for Smad3)	ACCAATAATAGGGCCTTTTGGCCCTATTATTGGTTATTATTGG/3InvdT/
	Hairpin_27 (used for Myc and Foxo3)	ACATCATCATGGGCCTTTTGGCCCATGATGATGTATGATGATG/3InvdT/
	Hairpin_28 (used for Gfod1)	ACAACTTAACGGGCCTTTTGGCCCGTTAAGTTGTGTTAAGTTG/3InvdT/
SABER: Myc probes	Mus_Myc_FISH_1	CATAGGATGGAGAGCAGAGCCCGGAGCGGCTTTCATCATCAT
	Mus_Myc_FISH_2	CAATGGGCAAAGTTTCCCAGCGGCGGCGAGTTTCATCATCAT
	Mus_Myc_FISH_3	AGGAGGGGAGCTGAGTGAGGCGAGTCGGACTTTCATCATCAT
	Mus_Myc_FISH_4	TTCCTCCGTGTCTGAGGACCTGGGGCTGGATTTCATCATCAT
	Mus_Myc_FISH_5	GCGCTTCAGCTCGTTCCTCCTCTGACGTTCTTTCATCATCAT
	Mus_Myc_FISH_6	GCACTTGCGGTTGTTGCTGATCTGCTTCAGGTTTCATCATCAT
	Mus_Myc_FISH_7	CAAGACGTTGTGTGTCCGCCTCTTGTCGTTTTTCATCATCAT

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Myc_FISH_8	TACAGGCTGGAGGTGGAGCAGACGCTGTGCTTTCATCATCAT
	Mus_Myc_FISH_9	TCGAGTTTGTGTTTCAACTGTTCTCGTCGTTTCCTCATTTCATCATCAT
	Mus_Myc_FISH_10	GCCCAAAGGAAATCCAGCCTTCAAACAGCTCGTTTCATCATCAT
	Mus_Myc_FISH_11	CCTGGGGAGTCCTGTCCTGGCTCGCAGATTTTTCATCATCAT
	Mus_Myc_FISH_12	CACGGCTCTTCCAACCGTCCGCTCACTCCCTTTCATCATCAT
	Mus_Myc_FISH_13	TTGGGTCAGTCGCAGGGTTGGGGAGAGTGGTTTCATCATCAT
	Mus_Myc_FISH_14	CTTGAATGGACAGGATGTAGGCGGTGGCTTTTTTTCATCATCAT
	Mus_Myc_FISH_15	TCGCAGATGAAATAGGGCTGTACGGAGTCGTAGTTTCATCATCAT
	Mus_Myc_FISH_16	AGGAGCTCTTTTCAGGAGAGCTGATCGCGGTTTCATCATCAT
	Mus_Myc_FISH_17	CGGTCATCATCTCCAGCTGATCGGCGGTGGTTTCATCATCAT
	Mus_Myc_FISH_18	GACCAGTGGGCTGTGCGGAGGTTTGCTGTGTTTCATCATCAT

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Myc_FISH_19	CTGGTGAGTGGAGACGTGGCACCTCTTGAGTTTCATCATCAT
	Mus_Myc_FISH_20	GCTCTGCTGTTGCTGGTGATAGAAATTCTCTTCCTCGTTTCATCATCAT
	Mus_Myc_FISH_21	CCTCTTCTCCACAGACACCACATCAATTTCTTCCTCATTTCATCATCAT
	Mus_Myc_FISH_22	GTGGAATCGGACGAGGTACAGGATTTGGGCTTTCATCATCAT
	Mus_Myc_FISH_23	AGTTAGGTCAGTTTATGCACCAGAGTTTCGAAGCTGTTTTCATCATCAT
	Mus_Myc_FISH_24	CCGCTCCACATACAGTCCTGGATGATGATGTTCTTTTCATCATCAT
	Mus_Myc_FISH_25	ACCCTGCCACTGTCCAACTTGGCCCTCTTGTTTCATCATCAT
	Mus_Myc_FISH_26	GAAGCTCTGGTTCACCATGTCTCCTCCAAGTAACTTTTCATCATCAT
	Mus_Myc_FISH_27	CCCTTGGGGAGAAGGACGTAGCGACCGCAATTTCATCATCAT
	Mus_Myc_FISH_28	GTAAGTTCCAGTGAGAAGTGTCTGCCCGCTGTTTCATCATCAT
	Mus_Myc_FISH_29	ACCACTGAGGGGTCAATGCACTCGGACGCGTTTCATCATCAT

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Myc_FISH_30	GAGCTGCTGTCGTTGAGCGGGTAGGGAAAGTTTCATCATCAT
	Mus_Myc_FISH_31	AGAAGTTGCCACCGCCGCCGTCATCGTCTTTTTCATCATCAT
	Mus_Myc_FISH_32	GCGCATCAGTTCTGTCAGAAGGAACCGTTCTTTTCATCATCAT
	Mus_Myc_FISH_33	TCGAGGTCATAGTTCCTGTTGGTGAAGTTCACGTTTTTCATCATCAT
	Mus_Myc_FISH_34	GCAGAGGCAGAGAACACTGTCCCCAAATAGACAAAATTTCATCATCAT
	Mus_Myc_FISH_35	CAGAGCTGCCTTCTTAGGTCGCCCCGGAGCTTTCATCATCAT
	Mus_Myc_FISH_36	GATGAGCCCGACTCCGACCTCTTGGCAGGGTTTCATCATCAT
	Mus_Myc_FISH_37	AGCAGCGAGTCCGAGGAAGGAGAGAAGGCCTTTCATCATCAT
	Mus_Myc_FISH_38	TCCGAGACCAGCTTGGCAGCGGCTGAGAAATTTCATCATCAT
	Mus_Myc_FISH_39	AATTCAGGGATCTGGTCACGCAGGGCAAAATTTCATCATCAT
	Mus_Myc_FISH_40	CCTCATGCAGCACTAGGGGCTCAGGGCTGGTTTCATCATCAT

Experiment	Identifier	Sequence (5' - 3')
	Mus_Myc_FISH_41	GCAGGCTCGGAGGCAAAGCCCCTCTCACTCTTTCATCATCAT
	Mus_Myc_FISH_42	CTTCCAGATATCCTCACTGGGCGCGGGCGGTTTCATCATCAT
	Mus_Myc_FISH_43	CTGTCTCTCGCTGGAATTACTACAGCGAGTCAGAAAATTTCATCATCAT
	Mus_Myc_FISH_44	CCAAGGTTGTGAGGTTAGGCTTTGAGCATGCATTTTATTTCATCATCAT
	Mus_Myc_FISH_45	GTGCTGTCTTTGCGCGCAGCCTGGTAGGAGTTTCATCATCAT
	Mus_Myc_FISH_46	TGGCTGAAGCTTACAGTCCCAAAGCCCCAGTTTCATCATCAT
	Mus_Myc_FISH_47	CCTTACTCTCACGAGAGATTCCAGCTCCTCCTCGTTTCATCATCAT
	Mus_Myc_FISH_48	GTCCTTTTCAGAGGTGAGCTTGTGCTCGTCTGTTTCATCATCAT
	Mus_Myc_FISH_49	TGATGAAGGTCTCGTCGTCAGGATCGCAGATTTTCATCATCAT
	Mus_Myc_FISH_50	GCAGCTGGATAGTCCTTCCTTGTGGAGGGGTTTCATCATCAT
SABER: Smad3 probes	Mus_Smad3_FISH_1	AGAATTTTCTCCTACAGCATTCTCCAGCAACTGGCCTTTCCAATAATA

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Smad3_FISH_2	GGAAATATGCCACAGGCTCCTGCAGGAGGGTTTCCAATAATA
	Mus_Smad3_FISH_3	TTCTTCCCCTCTGTAGCAGGACAGGACTTCCTTTCCAATAATA
	Mus_Smad3_FISH_4	TCATTGATGCAACTTGTCTCCATAAACCAACACAGCCTTTCCAATAATA
	Mus_Smad3_FISH_5	TGATTACACGTACCCAAATGACCTGTGTTCTCCATCATTTCCAATAATA
	Mus_Smad3_FISH_6	GATCCTTCCAAGGAAGAATGGGTTCTCCCCAGCTTTCCAATAATA
	Mus_Smad3_FISH_7	GGGGAGTACCTATGTAGGCCATCCAGTTTGACATTTCCAATAATA
	Mus_Smad3_FISH_8	GGTCTGTGCCCTTTCAGAGAACCTGAGGGGTTTCCAATAATA
	Mus_Smad3_FISH_9	ATGGGAAATCCATGCAGAAACATGGGCTCGTTTTCCAATAATA
	Mus_Smad3_FISH_10	GTGTCTCAAGGAGGATCAGCATGCCCAGCTTTTCCAATAATA
	Mus_Smad3_FISH_11	ACCATAGCTAGGTAATCCTACATCCCAGGGTGAAGATTTCCAATAATA
	Mus_Smad3_FISH_12	CGTGTGTCTGAAGCAGCAGTCCACAGACCATTTCCAATAATA

Experiment	Identifier	Sequence (5' - 3')
	Mus_Smad3_FISH_13	GGGGCATCAGAAGATTTATCACATTGGAGGGTACACTTTTCCAATAATA
	Mus_Smad3_FISH_14	TTTACACTCGTTTTCTCTGTAATTGCCGCTGGGAACTTTCCAATAATA
	Mus_Smad3_FISH_15	AATAGACTGTCCCTAGCCACCTGGAGGATAGCTTTTCCAATAATA
	Mus_Smad3_FISH_16	TGACAGAGCGCTCTGCACCAAGCCTGCAATTTTCCAATAATA
	Mus_Smad3_FISH_17	AGTACCAGTTGCAGGTGGCGATCCTGTGTATTTCCAATAATA
	Mus_Smad3_FISH_18	ATAACAGAGGACCATGAATGCTTCCCTAGGAGCCTTTCCAATAATA
	Mus_Smad3_FISH_19	GGCAGAAGCGCTCCGAGTTGGAGGGGTCAGTTTCCAATAATA
	Mus_Smad3_FISH_20	AGTGCTCATTACACAGTGTCACAGTTTGCTGTGGTTTCCAATAATA
	Mus_Smad3_FISH_21	TCTCAGGTACGGCACAATGTTCACGCCCATTTTCCAATAATA
	Mus_Smad3_FISH_22	TTTTAATAGCCCTTTTCATTTGCCCATGTCAGCCTGGTTTCCAATAATA
	Mus_Smad3_FISH_23	CGAAGCTCATACGGATGGTGCACATGCGCGTTTCCAATAATA

Experiment	Identifier	Sequence (5' - 3')
	Mus_Smad3_FISH_24	TCAGCTGGTAGACAGCCTCAAAGCCCTGGTTTTCCAATAATA
	Mus_Smad3_FISH_25	CCATTGTGGTTTTCCAGCCAACATACGTACCCATTTCCAATAATA
	Mus_Smad3_FISH_26	TGGTGGGATCTTGCAGACAGTGGCCGGGTGTTTCCAATAATA
	Mus_Smad3_FISH_27	CACAAAGACAGGAGGAGCGAGCTCTAACCAGGTTTCCAATAATA
	Mus_Smad3_FISH_28	GAGATGCATTCATTCGGTGTGGTGCCACCTTTTCCAATAATA
	Mus_Smad3_FISH_29	CTTTTCACAACAGAAGCAGCAGCCAAGAAGCGTTTCCAATAATA
	Mus_Smad3_FISH_30	ACTCAGAAATGGCAGCTTGGGGCTAAGAATGAAGTTTTCCAATAATA
	Mus_Smad3_FISH_31	TAGGAACTGTACATCACCTATACGCTGACCTTCAGCGTTTCCAATAATA
	Mus_Smad3_FISH_32	GCCTTCTCCAGCTCGTCCAACTGCCCCGTCTTTCCAATAATA
	Mus_Smad3_FISH_33	TCTACCCTCCGGCGTGTTCCCGTCTCTCTGTTTCCAATAATA
	Mus_Smad3_FISH_34	TGCTGTTCATACGAAAAGGAGAACTGATTTGGGAGGGTTTCCAATAATA

Experiment	Identifier	Sequence (5' - 3')
	Mus_Smad3_FISH_35	GCCCTTCTTCCAACCCAGCAGGCGCTTCACTTTCCAATAATA
	Mus_Smad3_FISH_36	AGAGGCCCTGCTCTCCAGTCTAGGCTTTCTTTTCCAATAATA
	Mus_Smad3_FISH_37	AAACCTCAGGGCAAGAAGCAGTGGCAGGGTTTTCCAATAATA
	Mus_Smad3_FISH_38	TGGAGAAAGGCACTTTCCTTTTGGACTGTGACATCCTTTCCAATAATA
	Mus_Smad3_FISH_39	CGTGGCTGGGGCTACCTGAACCAGCAGGAATTTCCAATAATA
	Mus_Smad3_FISH_40	AGCCATCACTTGGCCTTCAGACAGGGGCACTTTCCAATAATA
	Mus_Smad3_FISH_41	GTGGGCACCCAGGGCTGGAACAGCACATGTATTTCCAATAATA
	Mus_Smad3_FISH_42	TGCTGCTTCCTGTGAACACTCTGAGACTGATGTTTTCCAATAATA
	Mus_Smad3_FISH_43	AGTTGAGTGGAATCGGTGTTCTTTGACAAGACCAAGTTTTCCAATAATA
	Mus_Smad3_FISH_44	TGGATGGAGTTCTCTTCCAAGGTCATTTTCCCCATTTCCAATAATA
	Mus_Smad3_FISH_45	ACAGCAGAGCGTTGGTGTGTTCTGACTCAGTTTCCAATAATA

Experiment	Identifier	Sequence (5' - 3')
	Mus_Smad3_FISH_46	GTGTCTCTAAGACACACTGGAACAGCGGATGCTTTTCCAATAATA
	Mus_Smad3_FISH_47	CGGGGAACCCATCTGGGTGAGGACCTTGTCTTTCCAATAATA
	Mus_Smad3_FISH_48	TTCAGAATTGTATCAACGGAGACTCACACAGGCATCTTTTCCAATAATA
	Mus_Smad3_FISH_49	TACGAAAGCGGCAGAATGCTAAACATCAGTTCTGTGTTTCCAATAATA
	Mus_Smad3_FISH_50	CCTCCTGGTGAATTGGCTTCTATAACTCTCAGCTTCCTTTCCAATAATA
	Mus_Smad3_FISH_51	TGACAGACTGAGCTAGGAGGGCAGCAAATTCCTTTCCAATAATA
	Mus_Smad3_FISH_52	GACCCACATCCTGGTGTGTGGCTTTCTACCTTTCCAATAATA
	Mus_Smad3_FISH_53	CCAGCCATAGCGCTGGTTGCAGTTGGGAGATTTCCAATAATA
	Mus_Smad3_FISH_54	CCCCGTCTGCAATGCCACATCTTCATGGTTTTTCCAATAATA
	Mus_Smad3_FISH_55	GCAAAGACCTCCCCTCCGATGTAGTAGAGCCTTTCCAATAATA
	Mus_Smad3_FISH_56	GCATTCCGGTTGACATTGGACAGTAGGCCCTTTCCAATAATA

Experiment	Identifier	Sequence (5' - 3')
	Mus_Smad3_FISH_57	GCACTTGGTGTTCACGTTCTGCGTGGTGATTTTCCAATAATA
	Mus_Smad3_FISH_58	TGAAGCCATCTACTGTCATGGATGGCTGTGAGGTTTCCAATAATA
	Mus_Smad3_FISH_59	CGTGGAATGTCTCCCCAACTCGCTGGTTCATTTCCAATAATA
	Mus_Smad3_FISH_60	CTTCTCGCACCACTTCTCCTCCTGCCCGTTTTTCCAATAATA
	Mus_Smad3_FISH_61	CTGCGTCCATGCTGTGGTTCATCTGGTGGTTTTCCAATAATA
	Mus_Smad3_FISH_62	CACTGGTTTCTCCATCTTCACTCAGGTAGCCAGGTTTCCAATAATA
	Mus_Smad3_FISH_63	CCGGTTCCCTCTACTTTCCAGTTGCAGCAGTTTCCAATAATA
	Mus_Smad3_FISH_64	AGGGAAGTTAGTGTTCTCGGGAATGGAATGGCTTTTCCAATAATA
	Mus_Smad3_FISH_65	GGATCTCGGTGTGGCGTGGCACCAACACTGTTTCCAATAATA
	Mus_Smad3_FISH_66	TATACATCAGGGTTGTGGTGCCAGCTCCCATTTCCAATAATA
	Mus_Smad3_FISH_67	CCATGGCCCGTAATTCATGGTGGCTGTGCATTTCCAATAATA

Experiment	Identifier	Sequence (5' - 3')
	Mus_Smad3_FISH_68	CGGGCCATCGCCACAGGCGGCAGTAGATAATTTCCAATAATA
	Mus_Smad3_FISH_69	CACCCAGGACACATACTCCACGAGTGCCCTTTTCCAATAATA
	Mus_Smad3_FISH_70	AGTAGGGAGGGCCTGTGGTGCATCAGACATTTTCCAATAATA
	Mus_Smad3_FISH_71	GCCAACTTCTCAGGAGTGCTGAAGACAAAGGCTTTCCAATAATA
	Mus_Smad3_FISH_72	GTCAGGCCACACCCACTGCTGCACTGTCATTTTCCAATAATA
	Mus_Smad3_FISH_73	CTTGTTGGCCCAGTCTACTCTAATGCTTGTGGATCTTTTCCAATAATA
	Mus_Smad3_FISH_74	TTCTTGAGCTTCTTCACCAAGCTCTTGACCGCTTTCCAATAATA
	Mus_Smad3_FISH_75	TGGTGTGACCTTTCCTAGACTCCAAGCCTTTGATTTCCAATAATA
	Mus_Smad3_FISH_76	AAGCCACTGCAAGGGTCCATTCAGGTGTAGTTTCCAATAATA
	Mus_Smad3_FISH_77	AATAACATGAAGGGGTGTACTCCCACTGTGTGCCTTTCCAATAATA
	Mus_Smad3_FISH_78	TCCAGAAGCCGCCATCTGCTCTGAGCAGTGTTTCCAATAATA

Experiment	Identifier	Sequence (5' - 3')
	Mus_Smad3_FISH_79	GTGACAAGAACCTCAGCCTCCTAAACAAGAGTCCATTTCCAATAATA
	Mus_Smad3_FISH_80	ATGTAGCAGACACAGCTGTTCATAAATCCATCCCTGGTTTCCAATAATA
SABER: Foxo3 probes	Mus_Foxo3_FISH_1	CTTCTTCTTGGCTGCCCGGCCTCGGCTCTTTTTCATCATCAT
	Mus_Foxo3_FISH_2	CTGTGCAGGGACAGGTTGTGCCGGATGGAGTTTCATCATCAT
	Mus_Foxo3_FISH_3	GGTGTACTTGTTGCTGTTGTCCATGGAGACCGTTTCATCATCAT
	Mus_Foxo3_FISH_4	CCCCATCGGGGTTGATGATCCACCAAGAGCTTTCATCATCAT
	Mus_Foxo3_FISH_5	TCTTGCCCGTGCCTTCATTCTGAACGCGCATTTCATCATCAT
	Mus_Foxo3_FISH_6	CTGCCAGGCCACTTGGAGAGCTGGGAAGGATTTCATCATCAT
	Mus_Foxo3_FISH_7	GCTCGAACTCTGGGTCCAGCTCCACTTCGATTTCATCATCAT
	Mus_Foxo3_FISH_8	CTTCAGGAACGAGGCGGGAGGGCGTACGATTTTCATCATCAT
	Mus_Foxo3_FISH_9	CGCCGCCTTCCACATTCCTCCTCCTCCCTTTTTCATCATCAT

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Foxo3_FISH_10	GGCGGCCGCGGCAGCAGCACAAAGTTATAGTTTCATCATCAT
	Mus_Foxo3_FISH_11	GCCCTTATCCTTGAAGTAGGGCACACAGCGTTTCATCATCAT
	Mus_Foxo3_FISH_12	CACCATCCACTCGTAGATCTGGGACAAAGTGAGCTTTCATCATCAT
	Mus_Foxo3_FISH_13	GGAGCTCTCGATGGCGCGGGTGATCAGGTCTTTCATCATCAT
	Mus_Foxo3_FISH_14	CAGGCATTCCGCCGCGAGGAGCATTTCCTTTTTCATCATCAT
	Mus_Foxo3_FISH_15	CTCCTGGAGCCAGCAGCATGGCCGAATCCTTTTCATCATCAT
	Mus_Foxo3_FISH_16	GCCGCCGTCCTCGTCGTCTTCATCGTCGTCTTTCATCATCAT
	Mus_Foxo3_FISH_17	CTCGGGGATCATGGAGTCTGCGGCCGTCTCTTTCATCATCAT
	Mus_Foxo3_FISH_18	GCCACGTACAGGAGCGTGGCCGACTCTGTGTTTCATCATCAT
	Mus_Foxo3_FISH_19	CTTGGAGTGTCTGGTTGCCGTAGTGTGACACTTTCATCATCAT
	Mus_Foxo3_FISH_20	GGAAGAGAAGGTGGCTGGTCTGTTCTCCTGGTTTCATCATCAT

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Foxo3_FISH_21	ATAGTCTGCATGGGTGACTGACGCAAGGAGTTTTCATCATCAT
	Mus_Foxo3_FISH_22	TCAGAGACGAAGGTCCAAACACGGTACTGTTGATTTCATCATCAT
	Mus_Foxo3_FISH_23	GAGCCCAGGCCGGAGCTCTTGGCGGTATATTTTCATCATCAT
	Mus_Foxo3_FISH_24	CTGCGATGGCGGGAGCGCGATGTTATCCAGTTTCATCATCAT
	Mus_Foxo3_FISH_25	CAGGTCGTCCATGAGGTTCTCGGCCAGCCCTTTCATCATCAT
	Mus_Foxo3_FISH_26	ATCATTCAGATTCATGGTGCCGGCCATGTCCTTTCATCATCAT
	Mus_Foxo3_FISH_27	CGCGGAAGCTCCACAGTACACGGCTTGCTCTTTCATCATCAT
	Mus_Foxo3_FISH_28	GAGGGCGACAGGCTGGCAGAGCTGCTGTACTTTCATCATCAT
	Mus_Foxo3_FISH_29	GGGGACAGGGGTCCATCATCATCCTGGACGTTTCATCATCAT
	Mus_Foxo3_FISH_30	CCGCTCACGGTGCTGGCGTTGGAATTGGTGTTTCATCATCAT
	Mus_Foxo3_FISH_31	CGAGCGGAAGTCGGTCCACGCATCCAGCTCTTTCATCATCAT

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Foxo3_FISH_32	GCTCAAAAGCTTGCGATGCTCAAGGCCAGATTTCATCATCAT
	Mus_Foxo3_FISH_33	ACACCTTCCTTCTGCTTTTAAGTGTGCTAGGGAAAGGTTTCATCATCAT
	Mus_Foxo3_FISH_34	ACTGCAGGTTACTGTGTGTAGAAAACTCAGAGGGTCTTTCATCATCAT
	Mus_Foxo3_FISH_35	TAAACCATCATGATAGTAAAACACGACGGAGTCGCCATTTCATCATCAT
	Mus_Foxo3_FISH_36	CTGATGGGTGCTGTCCACGCTGGCAGGTCATTTCATCATCAT
	Mus_Foxo3_FISH_37	TGCATATGTTATAGAGTTCTGTTCCACGGGTAAGGGCTTTCATCATCAT
	Mus_Foxo3_FISH_38	TTCACCCAGGGCCAGGGCTGCTAACAGTCTTTTCATCATCAT
	Mus_Foxo3_FISH_39	CTGCTGGGTTAGGGAAGCGGCGTGGGAGTCTTTCATCATCAT
	Mus_Foxo3_FISH_40	TCAAAGGTGTCAAGCTGTAAACGGATCACTGTCCATTTCATCATCAT
	Mus_Foxo3_FISH_41	CTTGCTGAGAGCAGATTTGGCAAAGGGTTTTCTCTTTTCATCATCAT
	Mus_Foxo3_FISH_42	CCTTTCCTCAGTGATCCTTCAGCCTGGTACCCTTTCATCATCAT

Experiment	Identifier	Sequence (5' - 3')
	Mus_Foxo3_FISH_43	AGCTTTGAGATGAGGCCTGCTTAGCACCAGTTTCATCATCAT
	Mus_Foxo3_FISH_44	TGAAGTTCCCCACATTCAAACCAACAACGTTCTGTTTTCATCATCAT
	Mus_Foxo3_FISH_45	GTGGAGATGAGGGAGTCAAAGTTAAAATCCAACCCGTTTTCATCATCAT
	Mus_Foxo3_FISH_46	CAGCATCCATGAGTTCACTACGGATGATGGACTCCTTTCATCATCAT
	Mus_Foxo3_FISH_47	ATGTCACATTCCAAGCTCCCATTGAACATGTCCAGTTTCATCATCAT
	Mus_Foxo3_FISH_48	GTCCAAGTCACTGGGGAACTTATCGTGGCCTTTCATCATCAT
	Mus_Foxo3_FISH_49	CATGACGGGAAGGTTTGCACTAGCTGAATACAGTGTTTCATCATCAT
	Mus_Foxo3_FISH_50	AGGAGCCTGAGAGAGAGTCCGAGAGGGTTTTTTCATCATCAT
	Mus_Foxo3_FISH_51	GCATAGACTGGCTGGCGGGAGACTGCTGCTTTTCATCATCAT
	Mus_Foxo3_FISH_52	GGTGTTTGGCTGAGCCAAGGCTGCTGGAGTTTTCATCATCAT
	Mus_Foxo3_FISH_53	CACTCAAGCCCATGTTGCTGACAGAATTTGACAAGGTTTCATCATCAT

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Foxo3_FISH_54	CCCTGGGTTTGGTGCTGGTGGTGGAGCAAGTTTCATCATCAT
	Mus_Foxo3_FISH_55	TTCTGATTGACCAAACTCCCCTGGGTAGGCTTTTCATCATCAT
	Mus_Foxo3_FISH_56	GGGCAGCAAAGGACATCATTGGATCGTTGCTTTCATCATCAT
	Mus_Foxo3_FISH_57	GAAGCATCACGTTCCGGCGGGCATTCTGGGTTTCATCATCAT
	Mus_Foxo3_FISH_58	CGGACACGGCGGTGCTAGCCTGAGACATCATTTCATCATCAT
	Mus_Foxo3_FISH_59	AGGGGTCCGACTGGGTCATCATGACGTCGCTTTCATCATCAT
	Mus_Foxo3_FISH_60	TGTGGCTGAGTGAGTCTGAAGCAAGCAGGTTTTCATCATCAT
SABER: Gfod1 probes	Mus_Gfod1_FISH_1	TTGGACAGCCTAGCCCTTGTGGTGGCTGCATTTCAACTTAAC
	Mus_Gfod1_FISH_2	AGGGACAGGGCAGAATAAAGGAGATCAGATCAGATCATTTCAACTTAAC
	Mus_Gfod1_FISH_3	TAGGCTTTTGCTTCTTCACCTGGGATGTCACGGTTTCAACTTAAC
	Mus_Gfod1_FISH_4	AATAAGCCACTGCATTTCTTAGAAGCTATGGGGCTCATTTCAACTTAAC

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Gfod1_FISH_5	TGCACTAAGCGTCAAGCGTGTCCACTCGAATTTCAACTTAAC
	Mus_Gfod1_FISH_6	ACACACAGTGACAGGGGACTCTATGGCCCTTTTCAACTTAAC
	Mus_Gfod1_FISH_7	GACAATTTCCAGGTCCAGGCAAGGTTTGAACGCTTTCAACTTAAC
	Mus_Gfod1_FISH_8	ATGACAGACTATAAAAGCACCAGGCAGCAAACAAAGCTTTCAACTTAAC
	Mus_Gfod1_FISH_9	CGTAGTTGCTAGAAGGTTCTCCCTCCCCTAAACCTTTCAACTTAAC
	Mus_Gfod1_FISH_10	AAACTCTGAGGAAACCACATCACTGCAGAGAAGTTCTTTTCAACTTAAC
	Mus_Gfod1_FISH_11	AGAAGCCCCAGCTAAATGCAGAGCGTAACCTTTCAACTTAAC
	Mus_Gfod1_FISH_12	CACCTGCTGGACGCTGACCGGAGTGAGCTATTTCAACTTAAC
	Mus_Gfod1_FISH_13	GCAGAATACAAGTGCTGGGGATGGCTCTGGTTTCAACTTAAC
	Mus_Gfod1_FISH_14	AGCCATACTTCAGACTTCTGACACCTCCCTTGCTTTCAACTTAAC
	Mus_Gfod1_FISH_15	TAGGATTGTCTGACAGGTACTTAGTTCCCCAGTGTCATTTCAACTTAAC

Experiment	Identifier	Sequence (5' - 3')
	Mus_Gfod1_FISH_16	TCCTGTCCTGCAAGATCCGGGGTTTTGGTGTTTCAACTTAAC
	Mus_Gfod1_FISH_17	GGCAAAAGAAATCACCAGTGAATGGAGCCAGAGTTTTCAACTTAAC
	Mus_Gfod1_FISH_18	TTCTTATCACTCTGTTCTCTGCTGGCGTAAACAAGGGTTTCAACTTAAC
	Mus_Gfod1_FISH_19	GCTGCAGCAGACTTCAGGGAACTCAGTAGTGCTTTCAACTTAAC
	Mus_Gfod1_FISH_20	AATGCGTCAGTCGCCTGGAAGAGAGCCAGATTTCAACTTAAC
	Mus_Gfod1_FISH_21	GGCTTCAGTGGCCCTCTTAAGGGCTCCAGCTTTCAACTTAAC
	Mus_Gfod1_FISH_22	AGATAAGGACCCTGATGCAGGGGAGAAAAGGCTTTCAACTTAAC
	Mus_Gfod1_FISH_23	CGGTCATGATGGCGATGTTCTGCCACTCTCTTTCAACTTAAC
	Mus_Gfod1_FISH_24	CAGTCTGGCTGGACCTCTTGATGGTGTCTACCTTTCAACTTAAC
	Mus_Gfod1_FISH_25	ACGCAAAGAGCGTACAGGCAGTCGTCGAAGTTTCAACTTAAC
	Mus_Gfod1_FISH_26	CGTCCCACGTGCGCCGGTCATCTTGATCCTTTTCAACTTAAC

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Gfod1_FISH_27	GAAGGCCTGGCGCACTGCCTGCATCATCTTTTTCAACTTAAC
	Mus_Gfod1_FISH_28	CGCAGGTAGGGCGAGGGGATGTCGCTGAATTTTCAACTTAAC
	Mus_Gfod1_FISH_29	GCCTTCTCAGGCAGCAGGGAGTTGCTCACGTTTCAACTTAAC
	Mus_Gfod1_FISH_30	GATGTAGTGTCCTGAAGCAGCAGCTCCTGCTTTCAACTTAAC
	Mus_Gfod1_FISH_31	TCTGGGGCGCTATTGCGCTGTCCATAGAGGTTTCAACTTAAC
	Mus_Gfod1_FISH_32	CGGAACCCACCACGGTCACGTCCTGCTTGATTTCAACTTAAC
	Mus_Gfod1_FISH_33	ACTCTCCAGGCACATTGAAGTTCAGGGTAACGGTTTCAACTTAAC
	Mus_Gfod1_FISH_34	TGCAGCACACTCCACCCTCCAGCACCATCTTTTCAACTTAAC
	Mus_Gfod1_FISH_35	GGAAGGTGCAGAAGTCATCACTGGTGATCTGACGTTTCAACTTAAC
	Mus_Gfod1_FISH_36	GATGCCCTTGATGTGGTCCGTCTGCTTGACTTTCAACTTAAC
	Mus_Gfod1_FISH_37	AAAGGTCTTGAGCAGCCCGTGGACCTTGACTTTCAACTTAAC

Experiment	Identifier	Sequence (5' - 3')
	Mus_Gfod1_FISH_38	AGCCTTCTGGCCCGTGAGGAAGGTGAGCAGTTTCAACTTAAC
	Mus_Gfod1_FISH_39	GTCAATGATGTAGGTGCCCACGGAGTGCAGTTTCAACTTAAC
	Mus_Gfod1_FISH_40	CCACCACCCATCAGGTCGTCGCAGCTCCAGTTTCAACTTAAC
	Mus_Gfod1_FISH_41	TTGTATTTCTTGCCCAGCAGACTACCGCTGTGTTTCAACTTAAC
	Mus_Gfod1_FISH_42	CAGCAGCTCGCCCACGTAGCCCTCCTCAATTTTCAACTTAAC
	Mus_Gfod1_FISH_43	CAGCTGCTTCATGCGTACGAAGGCTGGGAGTTTCAACTTAAC
	Mus_Gfod1_FISH_44	GAAGCGCAGCACGTTGCCCATGATGCTCATTTTCAACTTAAC
	Mus_Gfod1_FISH_45	GAGCTTCGGGTAGTAGTGGGCCGCTGACATTTTCAACTTAAC
	Mus_Gfod1_FISH_46	GCGGCGTGGCGGTACGGTCACAGATGACATTTTTCAACTTAAC
	Mus_Gfod1_FISH_47	CGGCGGCAGGTTGATGCACACCAAGTCCACTTTCAACTTAAC
	Mus_Gfod1_FISH_48	ATCCTGGTGCAGGAGCACCTCGTCGATGCGTTTCAACTTAAC

Experiment	Identifier	Sequence (5' - 3')
	Mus_Gfod1_FISH_49	GCTGGTGTAGAAAGGAACGCTCATCTCCTTGGCTTTCAACTTAAC
	Mus_Gfod1_FISH_50	CACAGCGCCTTCACCGCGAAGCCCTCATCTTTTTTCAACTTAAC
	Mus_Gfod1_FISH_51	CAGCAGCGGGATGATGACTCGGGCTGTGAGTTTCAACTTAAC
	Mus_Gfod1_FISH_52	CCGGTTCTGCTGCTGCCGGGATCTCCAGTCTTTCAACTTAAC
	Mus_Gfod1_FISH_53	CCATGGCCGGCTCATCCCGGCAGAATCTCTTTTCAACTTAAC
	Mus_Gfod1_FISH_54	GCTTCACCACCCTCTGTGGGCTGCGGATCTTTTCAACTTAAC
	Mus_Gfod1_FISH_55	TTTTCCCCAAGTCCTTGGTAACGAATGGAACCCATTTCAACTTAAC
	Mus_Gfod1_FISH_56	GAGGCCACCCGGAGGCCGACTAAGGATGAGTTTCAACTTAAC
	Mus_Gfod1_FISH_57	GCTCGTAGTGTACCAGGCGCCGATCCCGAGTTTCAACTTAAC
	Mus_Gfod1_FISH_58	CGGCGCCAGGCTCCTCATTCTGCTGCCATATTTTCAACTTAAC
	Mus_Gfod1_FISH_59	CACAGGCTAGTCAGAGTGAAAGTTCACATCCAGCATTTCAACTTAAC

Experiment	**Identifier**	**Sequence (5' - 3')**
	Mus_Gfod1_FISH_60	ATAGGACAATAGCCTATGAGTAGAGTGGGCCTTGCGTTTCAACTTAAC
	Mus_Gfod1_FISH_61	TTCTTCTCAGGGTGAATACCTGGCTCTCTGTCCCTTTCAACTTAAC
	Mus_Gfod1_FISH_62	AACACATCTGTAAAATACGGGTCTTCCTGTGACAGCTTTTCAACTTAAC
	Mus_Gfod1_FISH_63	CAGCGAATACTGTCCGGTGCAGTGCAGGTCTTTCAACTTAAC
	Mus_Gfod1_FISH_64	TGAGCTCATCAGCTAGGCAGCTTAATGGCCTTTTCAACTTAAC
	Mus_Gfod1_FISH_65	GGATGAGCCTCCCTTCCCTGTGTGGTAGCCTTTCAACTTAAC
	Mus_Gfod1_FISH_66	AAAATGGGCTCCCCAACCCTGAATCTAAAGAAGTGGTTTCAACTTAAC
	Mus_Gfod1_FISH_67	GTGGCCATGGCCATTGTGTTCCTCCCTTGGTTTCAACTTAAC
	Mus_Gfod1_FISH_68	GTGTCTTCTTCCTATGCCCCACAACCAAGTCCTTTCAACTTAAC
	Mus_Gfod1_FISH_69	AAGACAGCCAATGATGCTTAGACAGTCTCCAGCGTTTCAACTTAAC
	Mus_Gfod1_FISH_70	CCGAGGGCTGCTGAGTCACTTAAACATCAACATGTTTCAACTTAAC

Experiment	Identifier	Sequence (5' - 3')
	Mus_Gfod1_FISH_71	TAAAAGCAGCTAAGTGACATTTCCATCTGGGAAGGGGTTTCAACTTAAC
	Mus_Gfod1_FISH_72	AAAACTACATTTTCTGCAGCTGTAACTGCCAACCTGCTTTCAACTTAAC
	Mus_Gfod1_FISH_73	TGGCATCCTGCTCAACTGTGGCTGTTGTGGTTTCAACTTAAC
	Mus_Gfod1_FISH_74	ACCCCAGGCCACGCCTCTTATACTGTTGCTTTTCAACTTAAC
	Mus_Gfod1_FISH_75	CCCCTACAGGGCTCCATGAGGAATGCTTTCTTTCAACTTAAC
	Mus_Gfod1_FISH_76	CCGGCCAGAACAACAAGTACAGACAGCGATTTTCAACTTAAC
	Mus_Gfod1_FISH_77	TTGCTCACTCATTGTGTGGCTTTGGGCATGTTTTCAACTTAAC
	Mus_Gfod1_FISH_78	AAGATGTATCTACAAGACTGTGTTCGGTGGGTGAGCTTTCAACTTAAC
	Mus_Gfod1_FISH_79	CGGGTGGCACTGTACCTTGAGTGCTGACGATTTCAACTTAAC
	Mus_Gfod1_FISH_80	GACCTGACCCAGAGATGTGCTGCCAGGAAATTTCAACTTAAC
SABER: Fluorescent oligos	Hairpin_25_633	/5ATTO633N/ttTATTATTGGTTATTATTGGT/3InvdT/

Experiment	**Identifier**	**Sequence (5' - 3')**
	Hairpin_27_550	/5ATTO550N/ttATGATGATGTATGATGATGT/3InvdT/
	Hairpin_28_633	/5ATTO633N/ttGTTAAGTTGTGTTAAGTTGT/3InvdT/
GI-map validation: Protospacer sequences	DDX3X	GACCGCGAAGGCCCTCTCAC
	Gal4	GAACGACTAGTTAGGCGTGTA
	CNOT1	GGCGGAGTTAACCGAAGAGG
	NUP160	GTCCACGAAGCTCCGTTCCA
	NXF1	GTCCTGAGCGCTTGGGAGTT

Table 2: Reagents

Type	Reagent	Supplier	Identifier	Working Dilution
Antibody	PABPC4 antibody	Bethyl Laboratories	A301-466A	1:2,000
	DDX3X antibody	Bethyl Laboratories	A300-474A	1:2,000
	DIS3 antibody	Bethyl Laboratories	A303-764A-T	1:2,000
	EXOSC10 antibody	Bethyl Laboratories	A303-989A-T	1:2,000
	PABPN1 antibody	MBL International	RN023PW	1:2,000
	ZFC3H1 antibody	Bethyl Laboratories	A301-456A	1:2,000
	GAPDH antibody	Thermo	MA5-15738	1:10,000
	RNA pol II antibody	Active Motif	61986	1:1,000
	Histone H3 antibody	Abcam	ab1791	1:5,000
	U1 snRNP 70 antibody	Santa Cruz Biotechnology	sc-390899	1:1,000
	LSM14A antibody	Santa Cruz Biotechnology	sc-398552	1:1,000
	RPS6 antibody	Proteintech	14823-1-AP	1:1,000
	Lamin B1 antibody	Abcam	ab229025	1:100
	alpha-Tubulin antibody	Abcam	ab7291	1:200
	anti-rabbit IRDye 800CW	Licor	926-32211	1:10,000
	anti-mouse IRDye 680RD	Licor	926-68070	1:10,000
	anti-Rat Cyanine5 antibody	Invitrogen	A10525	1:10,000
	Donkey anti-Mouse 405	Jackson Immunoresearch	715-475-150	1:500
	Donkey anti-Mouse 488	Jackson Immunoresearch	715-545-150	1:500
	Donkey anti-Rabbit 405	Jackson Immunoresearch	711-475-152	1:500
	Donkey anti-Rabbit 488	Jackson Immunoresearch	711-545-152	1:500

Type	Reagent	Supplier	Identifier	Working Dilution
	DDX3X antibody	custom made by Genemed Synthesis using peptide ENALGLDQQFAGLDLNSSDNQS		1:5,000
	beta-actin	Santa Cruz Biotechnology	sc-47778-AF680	1:10,000
	GADPH	GeneTex	GTX627408	1:5,000
	anti-mouse IgG 800	Licor	926-32210	1:20,000
	anti-rabbit IgG 680	Licor	926-68073	1:20,000
	anti-mouse IgG 680	Licor	926-68070	1:20,000
Plasmid	DDX3X shRNA plasmid	Horizon Discovery	TRCN0000000003	NA
	PABPC4 shRNA plasmid	Horizon Discovery	TRCN0000074658	NA
	scramble shRNA plasmid (constitutive expression)	Addgene	1864	NA
	DIS3 shRNA plasmid	Horizon Discovery	V2THS_96258	NA
	EXOSC10 shRNA plasmid	Horizon Discovery	V2THS_275659	NA
	PABPN1 shRNA plasmid	Horizon Discovery	V2THS_41638	NA
	ZFC3H1 shRNA plasmid	Horizon Discovery	V2THS_35985	NA
	scramble shRNA plasmid (dox inducible)	Horizon Discovery	RHS4743	NA
	Packaging plasmid psPAX2	Addgene	12260	NA
	Envelope plasmid pMD2.G	Addgene	12259	NA
	lentiviral plasmid: hU6-sgRNA-EEF1A-GFP-modified-BstXI	gift from Luke Gilbert	pAX71	NA
	lentiviral plasmid: pCRISPRia-v2-mU6-sgRNA-EEF1A-Puro-T2A-BFP	Horlbeck et al. 2016 PMID: 27661255	pAX9 / Addgene #84832	NA
	lentiviral plasmid: UCOE-SFFV-	this study	pAX55	NA

Type	Reagent	Supplier	Identifier	Working Dilution
	DDX3_wt-3xflag-2xP2A-mCherry			
	lentiviral plasmid: UCOE-SFFV-DDX3_R534S-3xflag-2xP2A-mCherry	this study	pAX66	NA
	lentiviral plasmid: UCOE-SFFV-DDX3_R326H-3xflag-2xP2A-mCherry	this study	pAX67	NA
	lentiviral plasmid: UCOE-SFFV-DDX3_R351W-3xflag-2xP2A-mCherry	this study	pAX68	NA
	lentiviral plasmid: UCOE-SFFV-DDX3_R376C-3xflag-2xP2A-mCherry	this study	pAX69	NA
	lentiviral plasmid: UCOE-SFFV-DDX3_G325E-3xflag-2xP2A-mCherry	this study	pAX70	NA
	lentiviral plasmid: UCOE-SFFV-DDX3_T532M-3xflag-2xP2A-mCherry	this study	pAX105	NA

REFERENCES

1. 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.

2. 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.

3. 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*. 10.7554/eLife.32537.

4. Wachutka, L., Caizzi, L., Gagneur, J., and Cramer, P. (2019). Global donor and acceptor splicing site kinetics in human cells. Elife *8*. 10.7554/eLife.45056.

5. 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.

6. 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.

7. 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.

8. 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.

9. 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.e20.

10. 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.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

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

17. 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.

18. 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.

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

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

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

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

30. 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.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

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

37. 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.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

43. 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.

44. 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.

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

46. Lefaudeux, D., Sen, S., Jiang, K., and Hoffmann, A. (2022). Kinetics of mRNA nuclear export regulate innate immune response gene expression. Nat. Commun. *13*, 1–16.

47. 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. 10.1016/j.cels.2022.04.005.

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

49. Schott, J., Reitter, 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.

50. Ren, J., Zhou, H., Zeng, H., Wang, C.K., Huang, J., Qiu, X., Sui, X., Li, Q., Wu, X., Lin, Z., et al. (2023). Spatiotemporally resolved transcriptomics reveals the subcellular RNA kinetic landscape. Nat. Methods, 1–11.

51. 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.

52. 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.

53. 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.

54. 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.

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

56. 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.

57. Hubstenberger, A., Courel, M., Bénard, M., Souquere, S., Ernoult-Lange, M., Chouaib, R., Yi, Z., Morlot, J.-B., Munier, A., Fradet, M., et al. (2017). P-Body Purification Reveals the Condensation of Repressed mRNA Regulons. Mol. Cell *68*, 144–157.e5.

58. 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*. 10.7554/eLife.45396.

59. Buschauer, R., Matsuo, Y., Sugiyama, T., Chen, Y.-H., Alhusaini, N., Sweet, T., Ikeuchi, K., Cheng, J., Matsuki, Y., Nobuta, R., et al. (2020). The Ccr4-Not complex monitors the translating ribosome for codon optimality. Science *368*. 10.1126/science.aay6912.

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

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

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

63. 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.

64. 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.

65. 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.

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

67. 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.

68. 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.

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

70. 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.

71. Fröhlich, A., Rojas-Araya, B., Pereira-Montecinos, C., Dellarossa, A., Toro-Ascuy, D., Prades-Pérez, Y., García-de-Gracia, F., Garcés-Alday, A., Rubilar, P.S., Valiente-Echeverría, F., et al. (5 2016). DEAD-box RNA helicase DDX3 connects CRM1-dependent nuclear export and translation of the HIV-1 unspliced mRNA through its N-terminal domain. Biochim. Biophys. Acta *1859*, 719–730.

72. Brennan, R., Haap-Hoff, A., Gu, L., Gautier, V., Long, A., and Schröder, M. (9 2018). Investigating nucleo-cytoplasmic shuttling of the human DEAD-box helicase DDX3. Eur. J. Cell Biol. *97*, 501–511.

73. 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.

74. Calviello, L., Venkataramanan, S., Rogowski, K.J., Wyler, E., Wilkins, K., Tejura, M., Thai, B., Krol, J., Filipowicz, W., Landthaler, M., et al. (2021). DDX3 depletion represses translation of mRNAs with complex 5′ UTRs. Nucleic Acids Res. *49*, 5336–5350.

75. 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.

76. Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields, A.P., Park, C.Y., Corn, J.E., Kampmann, M., et al. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. Elife *5*, e19760.

77. Wang, T., Birsoy, K., Hughes, N.W., Krupczak, K.M., Post, Y., Wei, J.J., Lander, E.S., and Sabatini, D.M. (2015). Identification and characterization of essential genes in the human genome. Science *350*, 1096–1101.

78. Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N., Stukalov, A., et al. (2015). Gene essentiality and synthetic lethality in haploid human cells. Science *350*, 1092–1096.

79. Lai, M.-C., Lee, Y.-H.W., and Tarn, W.-Y. (9 2008). The DEAD-Box RNA Helicase DDX3 Associates with Export Messenger Ribonucleoproteins as well asTip-associated Protein and Participates in Translational Control. Mol. Biol. Cell *19*, 3847–3858.

80. 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.

81. 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.

82. Oh, S., Flynn, R.A., Floor, S.N., Purzner, J., Martin, L., Do, B.T., Schubert, S., Vaka, D., Morrissy, S., Li, Y., et al. (2016). Medulloblastoma-associated DDX3 variant selectively alters the translational response to stress. Oncotarget *7*, 28169–28182.

83. Epling, L.B., Grace, C.R., Lowe, B.R., Partridge, J.F., and Enemark, E.J. (2015). Cancer-associated mutants of RNA helicase DDX3X are defective in RNA-stimulated ATP hydrolysis. J. Mol. Biol. *427*, 1779–1796.

84. Floor, S.N., Condon, K.J., Sharma, D., Jankowsky, E., and Doudna, J.A. (2016). Autoinhibitory Interdomain Interactions and Subfamily-specific Extensions Redefine the Catalytic Core of the Human DEAD-box Protein DDX3. J. Biol. Chem. *291*, 2412–2421.

85. Lennox, A.L., Hoye, M.L., Jiang, R., Johnson-Kerner, B.L., Suit, L.A., Venkataramanan, S., Sheehan, C.J., Alsina, F.C., Fregeau, B., Aldinger, K.A., et al. (2020). Pathogenic DDX3X Mutations Impair RNA Metabolism and Neurogenesis during Fetal Cortical Development. Neuron *106*, 404–420.e8.

86. 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.

87. Alles, J., Legnini, I., Pacelli, M., and Rajewsky, N. (2023). Rapid nuclear deadenylation of mammalian messenger RNA. iScience *26*, 105878.

88. 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.

89. 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.

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

91. 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.

92. 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. 10.1101/gr.156919.113.

93. 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. 10.1261/rna.062224.117.

94. 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*. 10.7554/eLife.32536.

95. 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.

96. 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.

97. Agarwal, V., and Kelley, D.R. (2022). The genetic and biochemical determinants of mRNA degradation rates in mammals. Genome Biol. *23*, 245.

98. Cockman, E., Anderson, P., and Ivanov, P. (2020). TOP mRNPs: Molecular Mechanisms and Principles of Regulation. Biomolecules *10*. 10.3390/biom10070969.

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

100. 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.e15.

101. 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.

102. Sharma, D., and Jankowsky, E. (2014). The Ded1/DDX3 subfamily of DEAD-box RNA helicases. Crit. Rev. Biochem. Mol. Biol. *49*, 343–360.

103. Snijders Blok, L., Madsen, E., Juusola, J., Gilissen, C., Baralle, D., Reijnders, M.R.F., Venselaar, H., Helsmoortel, C., Cho, M.T., Hoischen, A., et al. (2015). Mutations in DDX3X Are a Common Cause of Unexplained Intellectual Disability with Gender-Specific Effects on Wnt Signaling. Am. J. Hum. Genet. *97*, 343–352.

104. Scala, M., Torella, A., Severino, M., Morana, G., Castello, R., Accogli, A., Verrico, A., Vari, M.S., Cappuccio, G., Pinelli, M., et al. (8 2019). Three de novo DDX3X variants associated with distinctive brain developmental abnormalities and brain tumor in intellectually disabled females. Eur. J. Hum. Genet. *27*, 1254–1259.

105. Pugh, T.J., Weeraratne, S.D., Archer, T.C., Pomeranz Krummel, D.A., Auclair, D., Bochicchio, J., Carneiro, M.O., Carter, S.L., Cibulskis, K., Erlich, R.L., et al. (2012). Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature *488*, 106–110.

106. Robinson, G., Parker, M., Kranenburg, T.A., Lu, C., Chen, X., Ding, L., Phoenix, T.N., Hedlund, E., Wei, L., Zhu, X., et al. (2012). Novel mutations target distinct subgroups of medulloblastoma. Nature *488*, 43–48.

107. Jiang, L., Gu, Z.-H., Yan, Z.-X., Zhao, X., Xie, Y.-Y., Zhang, Z.-G., Pan, C.-M., Hu, Y., Cai, C.-P., Dong, Y., et al. (9 2015). Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. Nat. Genet. *47*, 1061–1066.

108. Gong, C., Krupka, J.A., Gao, J., Grigoropoulos, N.F., Giotopoulos, G., Asby, R., Screen, M., Usheva, Z., Cucco, F., Barrans, S., et al. (2021). Sequential inverse dysregulation of the RNA helicases DDX3X and DDX3Y facilitates MYC-driven lymphomagenesis. Mol. Cell *81*, 4059–4075.e11.

109. Valentin-Vega, Y.A., Wang, Y.-D., Parker, M., Patmore, D.M., Kanagaraj, A., Moore, J., Rusch, M., Finkelstein, D., Ellison, D.W., Gilbertson, R.J., et al. (2016). Cancer-

associated DDX3X mutations drive stress granule assembly and impair global translation. Sci. Rep. *6*, 25996.

110. Oksuz, O., Henninger, J.E., Warneford-Thomson, R., Zheng, M.M., Erb, H., Vancura, A., Overholt, K.J., Hawken, S.W., Banani, S.F., Lauman, R., et al. (2023). Transcription factors interact with RNA to regulate genes. Mol. Cell *83*, 2449–2463.e13.

111. 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.

112. Jan, C.H., Williams, C.C., and Weissman, J.S. (2014). Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling. Science. 10.1126/science.1257521.

113. 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.

114. 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.

115. 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.

116. 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.

117. 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.

118. 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. 10.17504/protocols.io.bpymmpu6.

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

120. 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.

121. 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.

122. 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.

123. 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.

124. 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.

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

126. 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.

127. 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.

128. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics *36*, 2628–2629.

129. 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.

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

131. 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.

132. 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.

133. Gupta, A., Martin-Rufino, J.D., Jones, T.R., Subramanian, V., Qiu, X., Grody, E.I., Bloemendal, A., Weng, C., Niu, S.-Y., Min, K.H., et al. (2022). Inferring gene regulation from stochastic transcriptional variation across single cells at steady state. Proc. Natl. Acad. Sci. U. S. A. *119*, e2207392119.

134. Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

135. Hickey, K.L., Dickson, K., Cogan, J.Z., Replogle, J.M., Schoof, M., D'Orazio, K.N., Sinha, N.K., Hussmann, J.A., Jost, M., Frost, A., et al. (2020). GIGYF2 and 4EHP

Inhibit Translation Initiation of Defective Messenger RNAs to Assist Ribosome-Associated Quality Control. Mol. Cell *79*, 950–962.e6.

136. Langmead, B., and Salzberg, S.L. (4 2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

137. Liao, Y., Smyth, G.K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. *47*, e47.

138. Li, W., Wang, W., Uren, P.J., Penalva, L.O.F., and Smith, A.D. (2017). Riborex: fast and flexible identification of differential translation from Ribo-seq data. Bioinformatics *33*, 1735–1737.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

Albert Xu

5A43598081C04FB... Author Signature

11/20/2023

Date