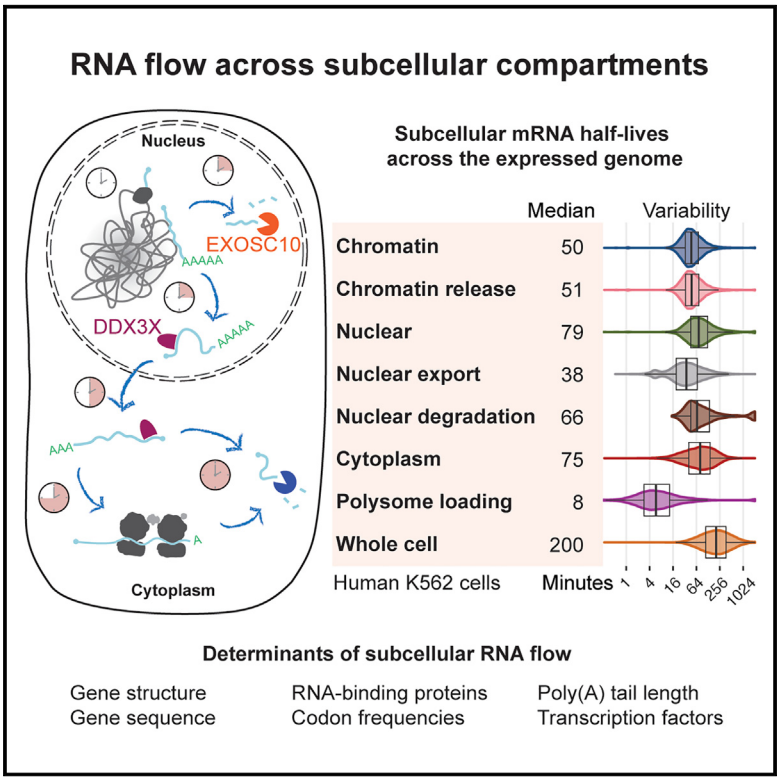
Molecular Cell

Resource

Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle

Graphical abstract



Highlights

- Subcellular mRNA flow rates determined by labeling, cell fractionation, and modeling
- DDX3X binding to target mRNAs is linked to slower nuclear export
- Abundant nuclear mRNA isoforms with retained introns are degraded by EXOSC10
- Genome-wide mRNA life cycles associate with gene function and many molecular features

Authors

Robert Ietswaart, Brendan M. Smalec, Albert Xu, ..., Luke Gilbert, Stephen N. Floor, L. Stirling Churchman

Correspondence

robert.ietswaart@cellforma.com (R.I.), stephen@floorlab.org (S.N.F.), churchman@genetics.med.harvard.edu (L.S.C.)

In brief

Ietswaart et al.'s analysis of mRNA flow kinetics across subcellular compartments uncovered many mammalian transcript life cycle determinants genome wide, including a link between the RNA-binding protein DDX3X and nuclear export. For hundreds of genes, mRNA isoforms with retained introns undergo frequent nuclear degradation, highlighting a potential mechanism of gene regulation.

Ietswaart et al., 2024, Molecular Cell *84*, 1–20
July 25, 2024
https://doi.org/10.1016/j.molcel.2024.06.008

CellPress

Molecular Cell

CellPress

Resource

Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle

Robert Ietswaart,[1,6,7,*] Brendan M. Smalec,[1,7] Albert Xu,[2,7] Karine Choquet,[1] Erik McShane,[1] Ziad Mohamoud Jowhar,[2] Chantal K. Guegler,[1] Autum R. Baxter-Koenigs,[1] Emma R. West,[1] Becky Xu Hua Fu,[3] Luke Gilbert,[3,4,5] Stephen N. Floor,[2,4,*] and L. Stirling Churchman[1,8,*]

[1]Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Cell and Tissue Biology, University of California, San Francisco, San Francisco, CA 94143, USA
[3]Arc Institute, Palo Alto, CA 94305, USA
[4]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA
[5]Department of Urology, University of California, San Francisco, San Francisco, CA 94518, USA
[6]Present address: Cellforma, Cambridge, MA 02139, USA
[7]These authors contributed equally
[8]Lead contact
*Correspondence: robert.ietswaart@cellforma.com (R.I.), stephen@floorlab.org (S.N.F.), churchman@genetics.med.harvard.edu (L.S.C.)
https://doi.org/10.1016/j.molcel.2024.06.008

SUMMARY

Dissecting the regulatory mechanisms controlling mammalian 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 in human and mouse cells. 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 a link between DDX3X and nuclear export. For hundreds of RNA metabolism genes, most transcripts with retained introns were degraded by the nuclear exosome, while the remaining molecules were exported with stable cytoplasmic lifespans. Transcripts residing on chromatin for longer had extended poly(A) tails, whereas the reverse was observed for cytoplasmic mRNAs. Finally, machine learning identified molecular features that predicted the diverse life cycles of mRNAs.

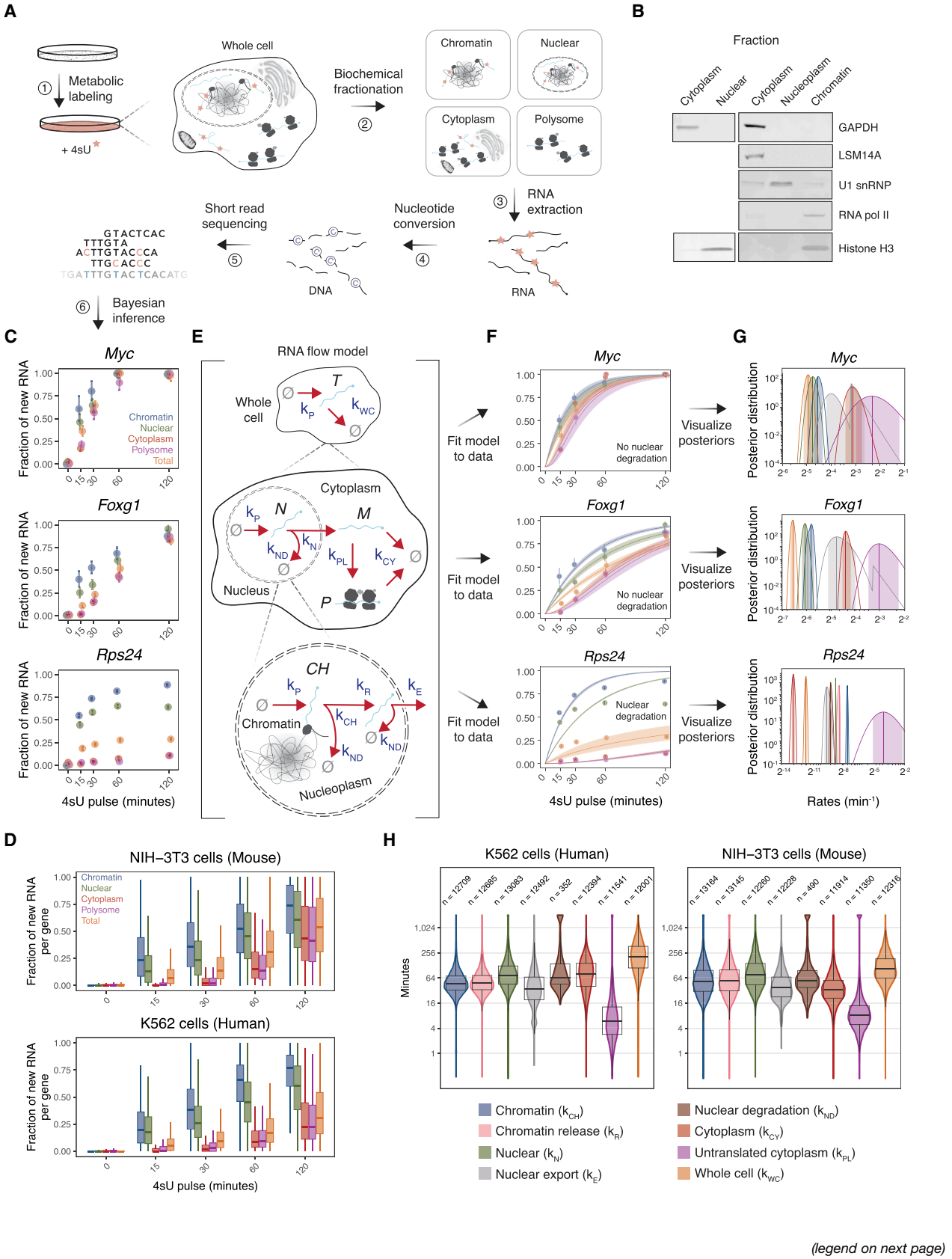
INTRODUCTION

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 (RNA Pol II) and undergo extensive processing, including splicing and polyadenylation.[3–9] In the nucleus, transcripts are subjected to either degradation or export. Nuclear degradation targets improperly processed mRNAs[10–13] but also serves additional regulatory roles for specific transcripts.[14] In the cytoplasm, ribosomes are loaded onto transcripts, influenced by 5′ UTR length and structure[15–19] and promoter elements.[20] Finally, mRNAs undergo degradation, a process that can be driven by both poly(A) tail deadenylation and targeting by microRNAs.[21–23] Each of these processes varies in duration across genes. In total, RNA half-lives vary greater than 100-fold between different protein-coding transcripts.[2,24–28] However, half-lives measured at the whole-cell level only represent the time between synthesis and decay, obscuring the dynamics of mRNA transitions across subcellular compartments.

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.[29–32] Endogenous RNAs have been studied using single-molecule RNA fluorescence *in situ* hybridization (FISH) combined with mathematical modeling, yielding nuclear and cytoplasmic RNA half-lives of a handful of genes in mouse tissue[33] and allowing for the quantification of the entire life cycle

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008



(legend on next page)

of an individual transcript in *Arabidopsis*.[34,35] Induced genes following a stimulus have been monitored to estimate subcellular turnover at higher throughput,[8,36–38] yet these kinetics may not extend to all genes or to cellular contexts that do not involve gene induction. Metabolic labeling has been used to globally study specific stages of RNA lifespans[39–42] but has not comprehensively characterized the entire life cycle of an mRNA across multiple compartments in mammalian cells, including degradation in the nucleus.

Here, we quantified the rates of mRNA flow across mammalian cells genome wide. We developed 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. For ~3%–4% of expressed genes, mRNA flow involved substantial nuclear degradation. RNA flow rates varied widely (>100-fold) between different genes and subcellular compartments, and functionally related genes undergo similar rates of RNA flow. The targets of many RNA-binding proteins (RBPs) exhibited different RNA flow rates compared with other genes. For DDX3X, these differences dissipated upon RBP depletion. Further investigation revealed a link between DDX3X and slow 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 mRNA half-lives. Finally, we identified the strongest genetic and molecular features that are predicted to determine RNA flow through machine learning, including transcription factors (TFs) and sequence elements. Our findings provide a comprehensive characterization of the dynamics of mRNAs throughout their cellular life cycles.

RESULTS

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

To quantify 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 min. Because high concentrations of 4sU can broadly impact gene expression,[43] we minimized the concentration and duration of 4sU exposure and confirmed that the addition of 4sU did not affect mRNA subcellular localization (Figures S1A and S1B) 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 (Figures 1B and S1D)[44] as well as polysome-bound RNA (Figure S1E). For comparison, we also collected total (whole cell) RNA.

Cellular fractionation was monitored by western blot analysis and RNA sequencing (RNA-seq) (Figures 1B, S1D, S1F, and S1G). Nuclear matrix and nuclear speckles have been reported to partly co-sediment with the chromatin.[45,46] Consistently, we found that nuclear speckle component SRSF7 was located in both the chromatin and nucleoplasm fractions (Figure S1D). To confirm effective isolation of compartment-specific RNA with minimal loss, we harvested RNA from the nuclear wash step and found 12- and 7-fold less mRNA in the wash compared with respectively the nuclear and cytoplasmic fractions (Figures S1F and S1G). Only 67 (0.6%) genes had levels in the nuclear wash comparable (<2-fold different) with their cytoplasmic levels. In conclusion, our fractionation led to robust separation of cellular compartments.

We then quantified the amount of newly synthesized RNA in each 4sU-labeled 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 studies used labeling conditions that achieved high 4sU incorporation rates, aiding the identification of labeled RNAs.[2,26,47,48] The lower 4sU incorporation rates inherent to our minimal labeling conditions necessitated a new computational approach to estimate the fraction of newly synthesized RNA (Figures S1H–S1J). We validated our approach with a NanoStrings-based assay that does not rely on sequencing and estimates of T>C conversions (Figures S1K–S1M).

Genome wide, we observed an increase in the proportion of new RNA with increasing pulse durations within each

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

(A) Schematic representing subcellular TimeLapse-seq.

(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 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 maximum a posteriori probability (MAP) values for each gene are shown for one replicate.

(E) Schematic of the RNA flow model.

(F) RNA flow model fit to subcellular TimeLapse-seq data for the example genes shown in (C). The dark lines represent the fits using RNA flow rate MAP values, while the ribbons show fits using the rate 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 as shown below in (H).

(H) Genome-wide subcellular half-lives for all protein-coding genes in mouse NIH-3T3 and human K562 cells. Mean half-lives are shown and the number of genes noted.

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

compartment (Figures 1C and 1D). Furthermore, at each time point, we saw delays in the fraction of new RNA across chromatin, nuclear, cytoplasm, and polysome fractions (Figures 1C and 1D), even for genes with relatively fast turnover (e.g., *Myc*, Figure 1C), confirming that our assay had 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 S2A), we estimated the Bayesian posterior probability distribution of each flow rate per gene (Figures 1F and 1G; Table S1). Using the posterior mean rate (k), half-lives are then calculated as $t_{1/2} = \frac{ln(2)}{k}$. We determined the half-lives of RNAs on chromatin ("chromatin half-lives"), in the nucleus ("nuclear half-lives"), and in the cytoplasm ("cytoplasm half-lives"). We estimated rates of nuclear export, 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.[49–52] 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; Table S1). 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[53] (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 cells (Figure S2B). Therefore, we extended our model by including a nuclear RNA degradation rate. To determine whether adding this parameter yields a better fit for each gene, we calculated a Bayes factor,[54] 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 S2B), whereas a model with nuclear degradation was successful (Figure 1F, Bayes factor $> 10^{16}$).

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 (Figure S2A). Similarly, for this population, nuclear turnover is a convolution of export and nuclear degradation, so the export rate was also estimated separately.

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) (Figures S2C and S2D; Table S1, Pearson correlations r = 0.69–0.9, CI overlaps = 44%–99%), and corresponded well with previously reported whole-cell half-lives (Figure S2C, Spearman correlation ρ = 0.59–0.89). For the genes best explained by the nuclear degradation model, nuclear export rates exhibited larger biological replicate variation (Figure S2D, r = 0.58–0.78), but with reproducible CI overlap (63%–86%). Lastly, as the uncertainty in the nuclear and cytoplasmic half-life estimates propagates to our estimates for the untranslated cytoplasm half-lives, these latter estimates exhibited greater variation between biological replicates (Figure S2D, r = 0.53–0.55). Nevertheless, the 95% CIs for the untranslated cytoplasm half-lives globally reproduced between replicates (overlaps = 85%–97%). Compared with our Bayesian distributions, least-squares regression failed to estimate rates late in the RNA life cycles, such as the polysome loading (r = 0.26–0.46), due to its inability to take measurement uncertainties into account (Figure S2E).

Finally, we compared our flow-rate-derived nuclear-to-cytoplasmic transcript abundance ratios with high-throughput single-molecule RNA FISH data (seqFISH+) in NIH-3T3 cells.[55] Although seqFISH+ was not intended for subcellular RNA quantification, we found that flow rate ratios showed a monotonic correlation with seqFISH+ nuclear-to-cytoplasmic RNA count ratios for all genes (Figure S3A). Moreover, we observed strong correspondence with nuclear-to-cytoplasmic RNA abundance ratios from spike-in normalized subcellular RNA-seq in K562 cells (Figure S3B). Together, we conclude that our flow rate measurements accurately reflect the timing of the eukaryotic RNA life cycle.

Wide gene-to-gene variability in RNA flow rates

In both cell lines, RNA flow rates varied considerably between genes (Figure 1H; Table S1). Nuclear RNA half-lives (median, 90% range: K562 79, 26–400 min, 3T3 76, 23–380 min) highly correlated with chromatin half-lives (median, 90% range: K562 50, 21–170 min, 3T3 62, 17–320 min, Figures S3C and S3D). Consequently, the nuclear export half-lives were (median, 90% range) in K562: 38, 5–210 min, and 3T3: 30, 9–130 min. Thus, most mRNAs within the nucleus are associated with chromatin (Figure S3E). By comparing nuclear and chromatin half-lives, we estimated that nuclear export can occur faster than our time resolution of 15 min (Figure S3F). In the cytoplasm, mRNAs were more stable in human K562 cells (median, 90% range: 75, 14–270 min) compared with mouse 3T3 cells (median, 90% range: 33, 9–130 min). Whole-cell turnover was also slower in K562 cells (median, 90% range: 200, 49–750 min) than in 3T3 cells (median, 90% range: 110, 33–430 min). Polysome loading occurs quickly after nuclear export (median, 90% range: K562 8, 1–82 min, 3T3 8, 2–50 min) and did not correlate with cytoplasmic turnover (Figures S3G and S3H). On average, mRNAs spent longer in the nucleus than the cytoplasm in NIH-3T3 cells. Moreover, in both cell lines, nuclear half-lives correlated more strongly than

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

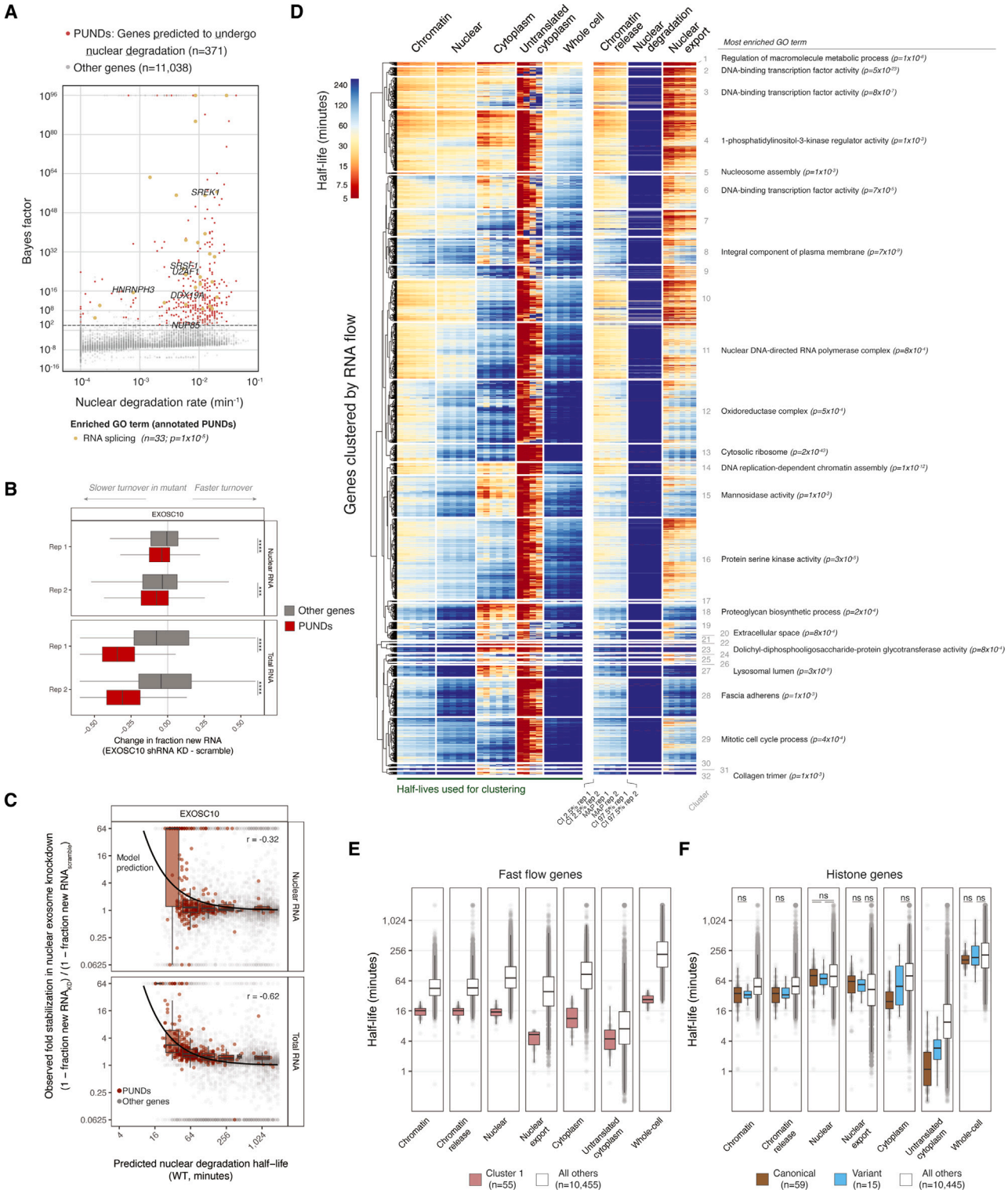


Figure 2. PUND transcripts are degraded by the nuclear exosome, and transcripts with similar RNA flow are functionally related

(A) Genes with a Bayes factor >100 in both replicates were labeled as those predicted to undergo nuclear degradation (PUNDs, 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}.

(B) The difference in fraction of new nuclear RNA upon 48 h doxycycline-induced shRNA depletion of EXOSC10 in K562 compared with a non-targeting control shRNA after a 60-min 4sU pulse. (****$p < 0.0001$, ***$p < 0.001$, Wilcoxon rank-sum test.)

(legend continued on next page)

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

cytoplasmic half-lives with whole-cell half-lives (Figures S3I and S3J). We accurately predicted the whole-cell half-life by adding the nuclear and cytoplasm half-lives for most genes, but not for genes predicted to undergo nuclear RNA degradation (Figures S3K and S3L). Similarly, summing the chromatin and export half-lives predicted the nuclear half-lives (Figures S3M and S3N). Chromatin and chromatin release half-lives correlated weakly for genes best explained by the nuclear degradation model, suggesting that nuclear degradation and export timescales are largely independent (Figures S3O and S3P).

Transcripts predicted to undergo nuclear degradation from genes with functions in RNA metabolism are degraded by the nuclear exosome

We identified genes whose transcripts are predicted to undergo nuclear degradation (PUNDs) as those genes whose Bayes factors reproducibly exceeded 100, indicating that the model with nuclear degradation is at least 100 times more likely to explain the subcellular TimeLapse-seq data than the model without nuclear degradation.[54] We found that 4% (n = 509/11,667) of expressed genes in NIH-3T3 cells and 3% (n = 371/11,409) of expressed genes in K562 cells met this criterion (Figures 2A and S4A; Table S2). The model predicts that the majority (71% in NIH-3T3 and 52% in K562) of transcripts produced by all PUNDs are degraded in the nucleus without being exported into the cytoplasm. We confirmed a higher nuclear to cytoplasm RNA abundance for PUNDs (Figure S4B). PUND transcripts were not preferentially removed during the nuclear wash steps (Figures S3A and S3B). For 38 and 165 genes in NIH-3T3 and K562 cells, respectively, neither model fit the subcellular TimeLapse-seq data (Table S1, Bayes factor = 0/0), including 49 out of 87 expressed human ribosomal protein genes (Fisher's exact test, odds ratio = 131, $p < 10^{-16}$). The undetermined genes overlapped with the small fraction of 4sU-sensitive genes (Figure S1C, Fisher's exact test, odds ratio = 2.6, $p = 4 \times 10^{-6}$), explaining why neither model provided a good fit. Nevertheless, our PUND identification method was conclusive for the vast majority (99%) of the expressed genes. Gene Ontology (GO) enrichment analysis (Table S2) resulted in common enriched functions between K562 and NIH-3T3 PUNDs, including RNA splicing (Figures 2A and S4A) and a significant overlap between homologous genes identified as PUNDs in both cell lines (Figure S4C). We conclude that nuclear degradation is a conserved regulatory feature that acts on select transcripts.

To confirm that PUND transcripts are degraded in the nucleus, we depleted subunits of the nuclear exosome (Figures S4D–S4G). Upon EXOSC10 depletion, we observed a reproducible reduction in nuclear turnover for PUNDs as compared with all other genes (Figures 2B and S4H) and PUND transcripts were stabilized in total RNA (Figures 2B and S4I). Importantly, we observed a correspondence between nuclear and whole-cell transcript stabilization in the EXOSC10 depletion and the predicted magnitude of nuclear degradation for PUNDs (Figures 2C and S4J), validating our model-predicted nuclear degradation half-lives. Thus, our model predicted the impact of the nuclear exosome on PUND transcripts, which undergo high levels 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.[2,24–28] To determine whether subcellular RNA flow rates may likewise serve a regulatory role, we performed hierarchical clustering on all genes in human K562 cells based on their subcellular half-lives (Figure 2D). We identified 32 total clusters with reproducibly distinct transcript kinetics (Figure 2D). GO enrichment analysis on each cluster determined that a large majority had functional enrichment (Figure 2D; Table S2). Cluster 16 contains the most genes (n = 1,297), which exhibit "canonical" RNA flow: median whole-cell half-lives of 210 min, 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 = 9,213) were divided into smaller groups that deviated from these canonical kinetics (Figure 2D). For example, genes involved in signal transduction and response to stimuli (e.g., *MYC*) were overrepresented in a cluster of 55 genes with "fast flow" kinetics across all compartments (Figure 2E). Consistently, gene set enrichment analysis (GSEA)[56,57] determined that signaling gene sets had faster RNA flow across all compartments (e.g., tumor necrosis factor alpha [TNF-α] signaling, Figure S4K; Table S2).

DNA-replication-dependent histone genes also clustered together (Figure 2D, cluster 14; Table S2). Comparison between canonical and variant histone genes indicated that canonical histones experience relatively unique cytoplasmic RNA flow, whereas variant histones behave more like the average protein-coding transcript (Figure 2F). Genes involved in gene expression clustered into several distinct groups, representing opposite RNA flow patterns (Figure S4L). Clusters 2–4 and 6 were enriched for transcription-related genes (Table S2), with shorter chromatin, nuclear, cytoplasm, and whole-cell half-lives (Figure S4L). By contrast, genes in clusters 13 and 28 were enriched for functions related to cytoplasmic translation (Table S2).

RNA flow rates are associated with RNA-binding proteins

mRNAs exist as part of ribonucleoprotein complexes containing many RBPs. To determine whether RBPs corresponded with

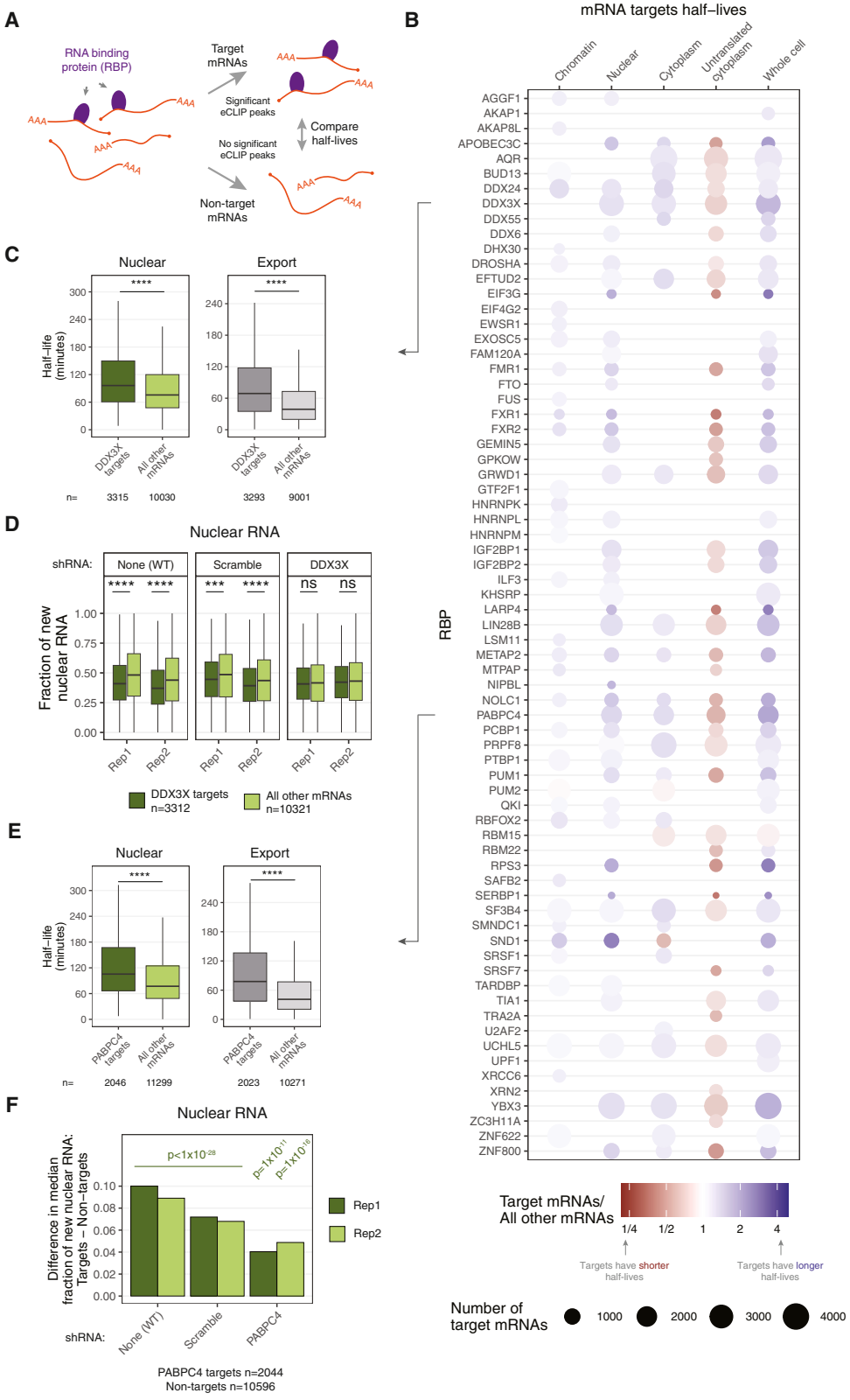
(C) The stabilizations in fraction of new nuclear and total RNA observed in the EXOSC10 depletion compared with the nuclear degradation half-life predicted in wild-type cells for PUNDs (red, with Spearman correlation) and other genes (gray). The line indicates the stabilization predicted in the depletion for PUNDs based on their nuclear degradation rate.

(D) Hierarchical clustering of human genes according to their RNA flow half-lives that were directly measured from the observed compartments (left columns). The most enriched GO annotations for each cluster are indicated (Table S2).

(E) Fast flow genes, i.e., those in cluster 1 in (D). All comparisons were statistically significant ($p < 0.0001$, Wilcoxon rank-sum test).

(F) Comparison of subcellular half-lives of histone genes. The number of genes of each type and significance is noted as in (E), unless otherwise specified (ns, not significant).

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008



(legend on next page)

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

specific RNA flow rates, we analyzed ENCODE enhanced cross-linking and immunoprecipitation (eCLIP) datasets for 120 RBPs in K562 cells[58] and identified all RBPs with target mRNAs that had significantly shorter or longer subcellular half-lives than non-targets (Figures 3A, 3B, and S5A–S5E; Table S3). Our approach was robust to randomized target sampling, which led to no significant associations. Although these associations do not imply a causal relation and there may be more RBP target mRNAs than those identified with eCLIP, it indicates where in the cell each RBP may regulate RNA flow. We identified significant differences in the subcellular half-lives for nearly all (45/47) of the RBPs whose targets show overall slower or faster whole-cell half-lives than other genes (Figure 3B). An additional 24 RBPs did not show differences at the whole-cell level, but most (19/24) exhibited significant differences in a specific compartment (Figure 3B).

DDX3X and PABPC4 link to slow 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 with 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 focused on DDX3X, an RNA helicase with many roles in RNA metabolism.[15,19,59–65] Consistent with reports that DD3X3 binds selectively to 5′ UTRs,[65] DDX3X eCLIP data identified 3,314 target transcripts with binding primarily occurring within the first exon (Figure S5F; Table S3). 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 (Figures 3C, S5A, and S5B). Furthermore, DDX3X targets had long cytoplasm and whole-cell half-lives (Figures 3B, S5C, and S5E), consistent with the association of DDX3X with several cytoplasmic RNA granules.[60,61,64] Upon DDX3X depletion, nuclear RNA turnover of DDX3X targets was no longer slower than turnover of non-targets (Figures 3D and S5G). RNA turnover on chromatin and at the whole-cell level was not affected (Figure S5H), while the difference in cytoplasm turnover between DDX3X targets and all other transcripts was subdued (Figure S5H). These data indicate 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 (Figures 3B, S6C, and S6E), consistent with the function of this RBP in stabilizing transcripts with short poly(A) tails containing AU-rich motifs.[66] Most eCLIP peaks were located in the last exon of target transcripts (Figure S5I). PABPC4 targets also had longer nuclear half-lives than other transcripts (Figure 3E). Upon PABPC4 depletion, the difference in nuclear half-lives of the targets was less (Figure 3F), without differences in the chromatin, cytoplasm, or whole-cell half-lives of target mRNAs (Figures S5G and S5J). We conclude that although both PABPC4 and DDX3X bind to their mRNA targets throughout the cell, their depletion, directly or indirectly, 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 orthogonally investigate DDX3X essentiality in relation to its functions across subcellular compartments, we conducted genome-wide CRISPR interference (CRISPRi) screens[67] to define genetic interaction networks of DDX3X. Using a K562 cell line constitutively expressing a single guide RNA (sgRNA) targeting DDX3X and a negative control cell line targeting the yeast Gal4 promoter (Figure 4A), we targeted all known protein-coding genes using CRISPRi frames (Figures 4B and S6A–S6D; Table S4). Genes that had either synergistic or buffering growth defects in combination with DDX3X knockdown were classified as genetic interactions (GIs), which identified both known and novel functional relationships with DDX3X (Figures 4C and 4D). GO analysis on all buffering genes revealed enrichment for genes involved in mRNA export from the nucleus and mRNA RNP complex nuclear export (Figures 4E and 4F). Closer examination of specific genes that have this buffering interaction included the nuclear pore protein complex member *NUP160*, the mRNA export factor *NXF1*,[15,59,62] and scaffold protein *CNOT1* of the CCR4-NOT deadenylase complex (Figure 4C). Individual cloning and validation of these genes confirmed their buffering activity with DDX3X depletion (Figures S6A–S6C). Importantly, using ribosome profiling, we identified DDX3X translation targets and found no evidence that DDX3X's role in regulating the translation of specific genes is linked to DDX3X's buffering or synthetic GIs (Figures 4G–4I and S6E–S6K). 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 with all other genes (Figure 4J), similar to what we observed when DDX3X was depleted by short hairpin RNA (shRNA) (Figures S6L and S6M). Individual knockdown of the

Figure 3. The targets of many RBPs exhibit distinctive RNA flow rates across human K562 cells

(A) Schematic for RBP analysis. The mRNA-binding targets of 120 RBPs in K562 were determined by identifying genes with significant eCLIP peaks.

(B) All RBPs with targets that exhibited significantly fast or slow RNA flow half-lives for target mRNAs compared with non-target mRNAs in both biological replicates (adjusted $p < 0.01$, Wilcoxon rank-sum test, Bonferroni multiple testing correction) in K562 cells. The number of target genes for each RBP is listed in Table S3.

(C) Nuclear and nuclear export half-lives of DDX3X mRNA targets and non-target mRNAs (****$p < 0.0001$, ***$p < 0.001$, ns, not significant, Wilcoxon rank-sum test).

(D) Fraction of new nuclear RNA of DDX3X target mRNAs compared with all other mRNAs in wild-type cells, cells expressing a constitutive DDX3X-targeting shRNA, and cells expressing a scrambled shRNA after a 60-min 4sU pulse. Significance as in (C).

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

(F) Difference in median fraction of new nuclear RNA of PABPC4 targets compared with all other genes in wild-type cells, cells expressing a constitutive PABPC4-targeting shRNA, and cells expressing a scrambled shRNA after a 60-min 4sU pulse.

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

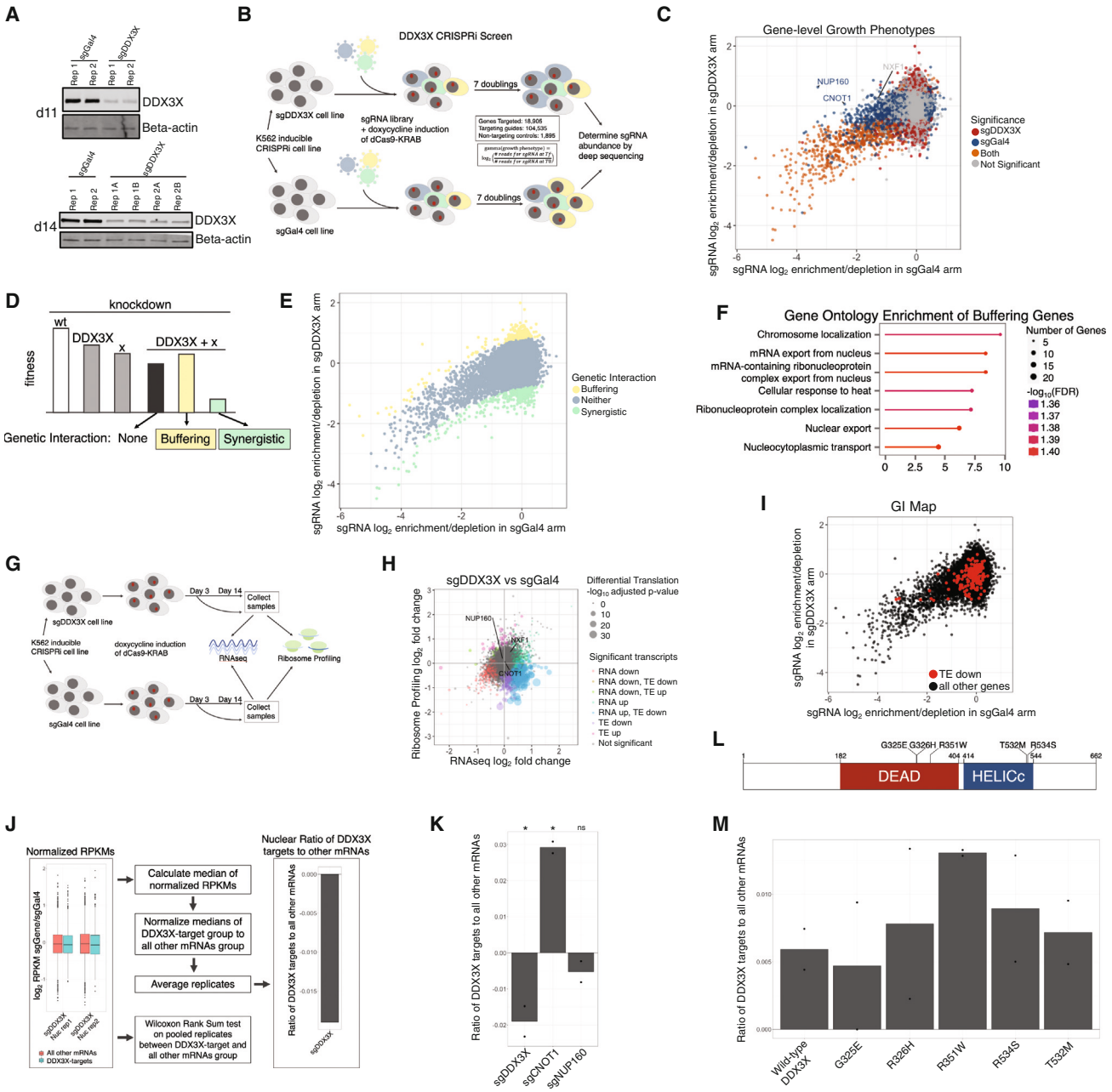


Figure 4. Genetic interaction mapping to link DDX3X depletion to mRNA export defects

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

(B) Schematic of the growth-based CRISPRi screen.

(C) Results from the GI screen for factors affecting the growth of control or DDX3X knockdown cells. Colored indicates significance in each screen arm.

(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 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.

(H) Ribosome profiling counts represent translational efficiency. Genes are colored by significance.

(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.

(legend continued on next page)

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

GI buffering interactors CNOT1 and NUP160 reversed this export 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 (Figures 4K, S6N, S6P, and S6Q). This reversal of the export phenotype of DDX3X-binding mRNAs by the buffering genetic interactors is consistent with the buffering phenotype we see at the level of cell growth from the GI map, although the cell growth defect may be due to more than the mRNA export defect alone. These phenotypes could be rescued with the expression of wild-type DDX3X or any of five inactive mutants of DDX3X that have been previously shown to impair its helicase activity[68–70] (Figures 4L, 4M, S6O, and S6Q), indicating that DDX3X's role in nuclear export is independent of its helicase activity. 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 RBPs.

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

Poly(A) tail lengths are connected to mRNA metabolism,[23,71,72] so we measured the tail length of each mRNA in chromatin, cytoplasm, polysome fractions, and total RNA using nanopore direct RNA-seq (Figures 5A and S7A–S7D; Table S5).[73] The longest poly(A) tails were on chromatin, with shorter polyA tail lengths in the cytoplasm, on polysomes, and in total RNA (Figures 5B, 5C, and S7A–S7D; Table S5). Total RNA tail lengths closely resembled those of cytoplasm RNA (Figure 5C). In general, the tail lengths in polysomal RNA were slightly shorter than in cytoplasmic RNA (Figure 5C).

Comparison of median tail lengths to subcellular half-lives revealed that, on chromatin, median poly(A) tail lengths were the shortest for genes with chromatin half-lives of less than 15 min, which gradually increased with half-life until ~50 min (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 (Figures S7E and S7F), and there was a negative correlation between median tail length in total RNA and whole-cell half-lives (Figure 5F), as previously observed.[74,75] It has been hypothesized that more stable transcripts undergo more deadenylation prior to degradation.[75] Indeed, the change in tail lengths between the chromatin and cytoplasm fractions of mRNAs with short cytoplasm half-lives (<30 min) differed by ~50 nt, whereas mRNAs with long cytoplasm half-lives (>240 min) differed by >100 nt (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 ~5× 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 89 nt) than *WSB1* (median 142 nt) (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 (Figures 5H and 5I).

PUNDs express transcripts with retained introns that undergo nuclear degradation by the exosome

In order to elucidate PUND regulation, we examined the RBPs targeting PUND transcripts. Remarkably, five of the 67 proteins most significantly enriched for PUNDs were components of the spliceosome: AQR, PRPF8, SF3B4, BUD13, and EFTUD2 (Figure 6A; Table S3). Analyses of our long-read direct RNA-seq libraries revealed that PUND genes had more incompletely spliced mRNAs on chromatin than other genes, but overall splicing levels did not differ meaningfully in other compartments (Figures 6B and S7G). PUND transcripts had longer tails on chromatin (median PUND tail length = 213 nt, median other = 193 nt, $p < 1 \times 10^{-16}$) (Figures 6C and S7H). Consistently, we also found that PUNDs were enriched for genes that express transcripts with detained introns (Fisher's exact test, odds ratio = 8, $p < 10^{-5}$) or nuclear retained introns (RIs) that are not sensitive to nonsense-mediated decay.[76]

To identify which isoforms of PUND transcripts are degraded in the nucleus by EXOSC10, we analyzed the nuclear turnover of: (1) exons of all protein-coding isoforms (PC isoforms), (2) isoforms containing RIs and any exons (RI isoforms), (3) RIs, (4) other introns (OIs) not annotated as RI from all isoforms, and (5) unspliced intron-exon junctions (UJs) from all isoforms (Figure 6D). Global nuclear degradation of all transcripts was assumed to be unaltered, consistent with previous reports.[77–79] Upon EXOSC10 depletion, stabilization of PUNDs was strongest for their RI isoforms and RIs, while OIs and UJs remained unaffected (Figure 6D) in a manner that corresponded with the nuclear degradation rates (Figure S7I). Thus, EXOSC10 degrades nuclear PUND transcripts with RIs.

Finally, PUND genes had faster turnover on chromatin, in the nucleus, and at the whole-cell level than genes without evidence of nuclear degradation (Figures 6E and S7J). PUND genes had

(K) DDX3X and selected buffering hits were depleted, and the nuclear and cytoplasmic RNA fractions were sequenced. Statistical testing (*$p < 0.05$, 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.

(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 as in (K). The DDX3X-target and the all-other-mRNA distributions were not statistically significantly different in any of the conditions.

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

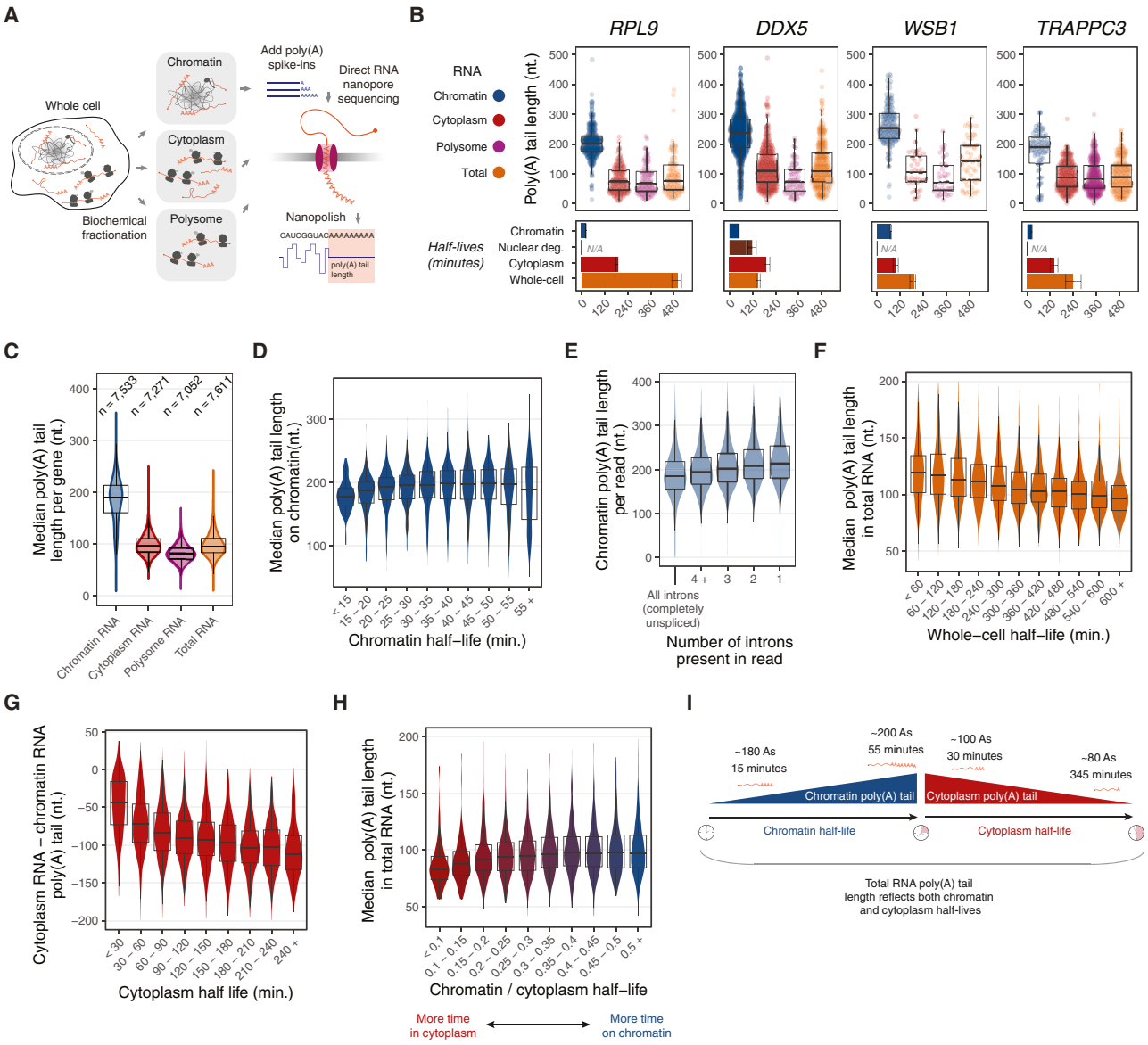


Figure 5. Subcellular compartment-specific poly(A) tail lengths reflect RNA flow rates

(A) Schematic for poly(A) tail length analysis.

(B) Poly(A) tail lengths per RNA across compartments for example genes (*RPL9*, *DDX5*, *WSB1*, and *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 95% CIs.

(C) Distribution of median poly(A) tail lengths for each gene. *n*, number of genes analyzed.

(D) Median poly(A) tail length for genes in chromatin RNA, as a function of chromatin RNA half-life.

(E) Poly(A) tail lengths on chromatin as a function of splicing status.

(F) Median total RNA poly(A) tail length for genes in total RNA 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, as a function of their cytoplasm half-life.

(H) Median poly(A) tail length in total RNA as a function of the ratio of chromatin half-life to cytoplasm half-life.

(I) Model of compartment-specific poly(A) tail lengths with respect to subcellular half-lives.

longer chromatin release and cytoplasm half-lives than other mRNAs (Figures 6E and S7J). Altogether, our data indicate that PUNDs express many transcripts that are degraded on chromatin and enriched for RIs. The PUND transcripts that are exported from the nucleus are fully spliced with long half-lives in the cytoplasm.

Machine learning model identifies molecular features that explain RNA flow rates

To identify genetic and molecular features that collectively explain the variability in RNA flow rates (Figure 7A; Table S6), we developed a LASSO regression model (Figures 7A

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

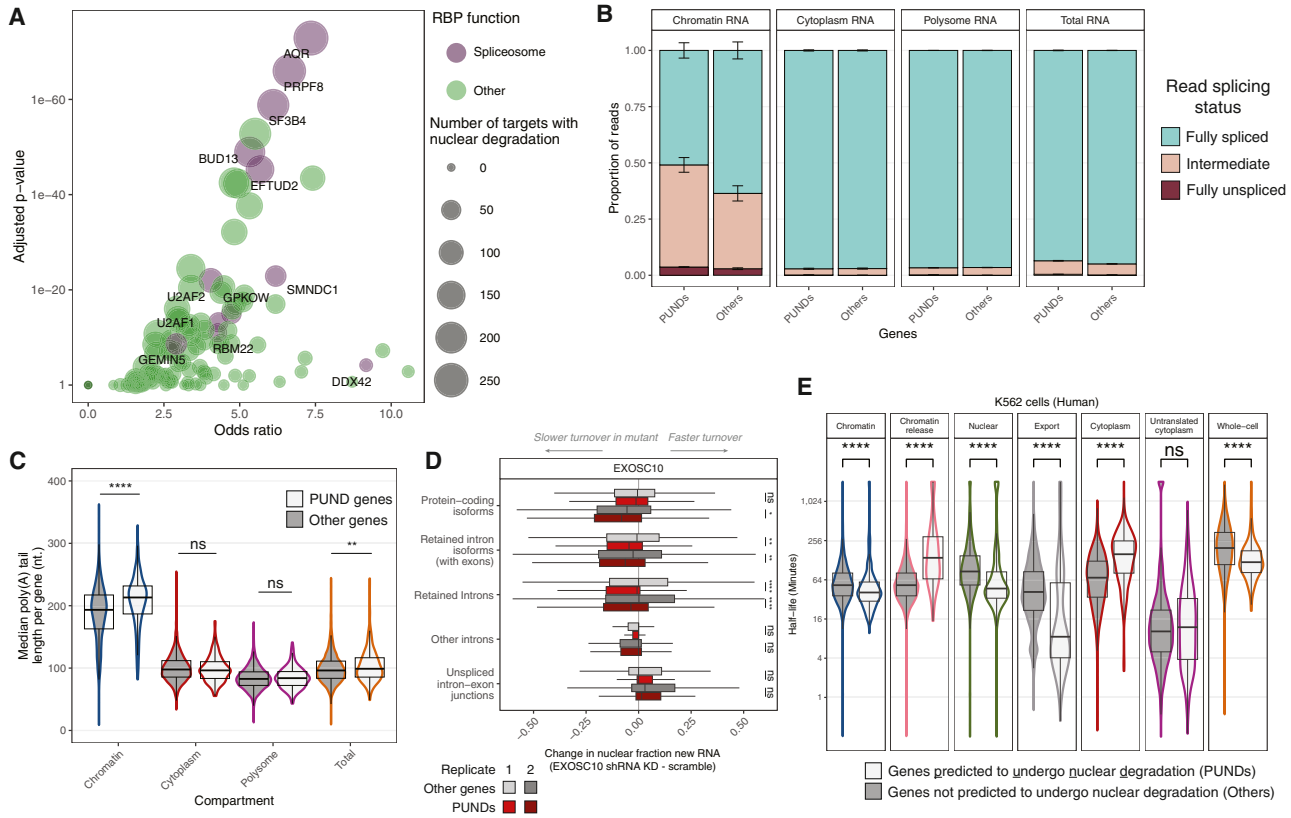


Figure 6. PUNDs exhibit unique phenotypes related to RNA flow, splicing, and poly(A) tail lengths

(A) 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.

(B) Splicing levels of PUND and other transcripts in nanopore long-read direct RNA sequencing data across subcellular compartments and in total RNA. Error bars show standard error over two biological replicates. All comparisons were significant (chi-square contingency test, $p < 10^{-16}$).

(C) Median poly(A) tail length of PUND genes relative to other genes across subcellular compartments. (****$p < 0.0001$, **$p < 0.01$, ns, not significant, Wilcoxon rank-sum test.)

(D) The difference in fraction of new nuclear RNA for transcript isoform types upon EXOSC10 depletion relative to a scramble control for PUNDs and other genes with specifications noted in Figure 2B, and significances as in (C).

(E) Half-lives of all PUND genes ($n = 371$) compared with all other genes in human K562 cells and significances as in (C).

and S7K) with cross-validation and unseen test set performances varying between the subcellular compartments (Figure 7B), ranging from $R^2 = 0.5$ for nuclear export and cytoplasmic turnover to $R^2 = 0.2$ for polysome loading rates, likely due to the larger uncertainty for these estimates (Figures 1G and S2D). Furthermore, 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 S7L).[40,80–86]

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 (Figures 7C and S7M; Table S6). We also observed subcellular-specific relevant classes, such as predicted microRNAs promoting cytoplasmic turnover, gene location predicting nuclear half-lives, and TF targets associating with cytoplasmic turnover (Figure 7C). The classes vary widely in the number of quantitative and qualitative input features (Table S6). Sequence determinants and codon frequencies were collectively relevant as feature classes (Figure 7C), with small and compartment-specific effect sizes of individual features (Figures S7M and S7N; Table S6). For example, transcripts containing a 5′-terminal oligo-pyrimidine (TOP) motif[87] had faster RNA flow dynamics, including polysome loading rates (Figure S7M; Table S6).

Next, we investigated the strongest individual feature contributions (Figure 7D). Intron length 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 (Figures 7E and S7O). Gene GC content and SND1 binding both negatively influenced chromatin turnover, nuclear turnover, and nuclear export rates (Figure 7D). Consistent with our RBP analysis (Figures 3C–3F), DDX3X and PABPC4 were predicted to slow down nuclear turnover.

RNA flow rates are associated with TFs

Our LASSO model indicated that several TFs, including MYC, are predictive for cytoplasm half-lives (Figure 7D; Table S6). We

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

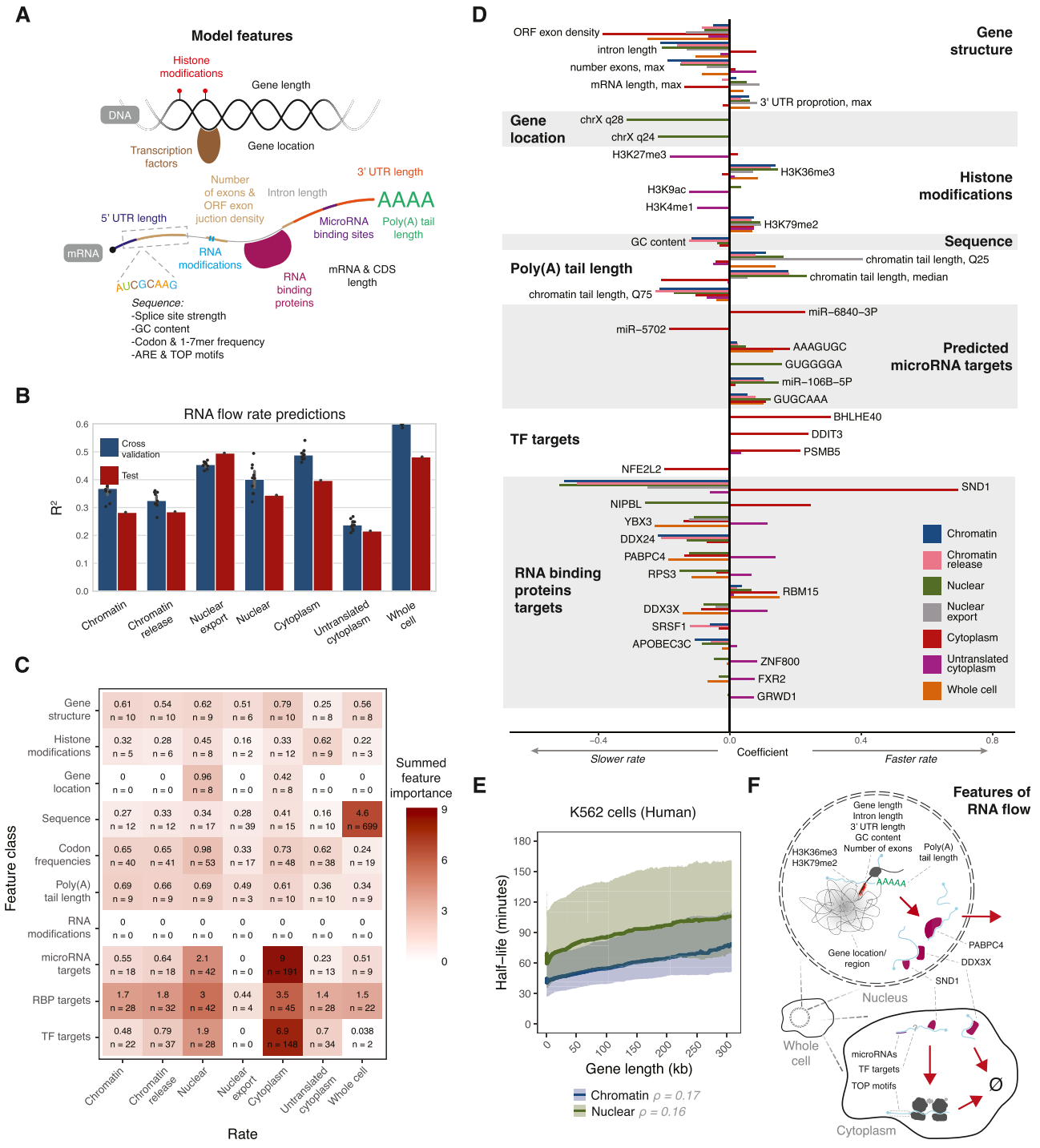


Figure 7. LASSO regression model identifies most relevant genetic and molecular features that predict RNA flow rates in human K562 cells

(A) Schematic representing the genetic and molecular features included in LASSO model (Table S6).

(B) 10× cross-validation and test set performances of LASSO models predicting RNA flow rates.

(C) Class feature importance of LASSO models, representing the sum over the importance of each individual feature within that class, with the number (*n*) of features.

(legend continued on next page)

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

confirmed this finding on MYC orthogonally by GSEA (Table S2). LASSO selects for global feature relevance and, therefore, rare features that generally make a smaller global contribution—for instance, TFs with few target genes—might have remained unidentified. To more sensitively investigate whether more TFs were associated with the RNA flow rates of their target genes, we identified TFs for which the half-lives of targets differed from non-targets (Figure S7P), similar to our RBP analysis (Figures 3A and 3B). 201 TFs were associated with altered RNA flow rates (Figure S7P; Table S3). Targets of the stress-response TF ATF5 had shorter chromatin half-lives (Figure S7P). Our LASSO model and TF analysis also revealed RNA flow associations with many uncharacterized zinc-finger proteins (Figure S7P; Table S3).

Lastly, we observed global trends for RBP and TF target half-lives dependent on the compartment (Figures 3B and S7P). Targets of RBPs exhibited slower chromatin and nuclear RNA flow kinetics (Figure 3B), in contrast to the faster chromatin kinetics of TF targets (Figure S7P). Nuclear export and cytoplasmic turnover was slower for TF targets and polysome entry faster for both TF and RBP targets. Randomized sampling of TF targets did not provide significant associations. In conclusion, hundreds of features underlying diverse regulatory mechanisms are predictive of RNA flow rates across the RNA life cycle (Figure 7F; Table S6).

DISCUSSION

Our analysis of RNA flow, based on subcellular TimeLapse-seq and kinetic modeling, has characterized the many possible life cycles of mammalian transcripts across the cell and yields subcellular RNA half-lives genome wide. We observed a large variability in RNA turnover across all subcellular compartments. We identified many mRNAs that spent equal or more time on chromatin than in the cytoplasm. 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.

We quantified and validated nuclear mRNA degradation kinetics by the nuclear exosome of several hundred protein-coding genes (PUNDs), which express RNA isoforms with detained introns, extending the repertoire of exosome targets.[10,13,77–79,88–93] 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. 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. In a hypothetical negative feedback mechanism, increased export of PUNDs encoding for nuclear RNA degradation factors, 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.[39,94,95] Overall, nuclear transcript degradation is likely to serve a regulatory function, an exciting direction for future work.

We observed associations between RNA flow rates and many RBPs. We report the first link between nuclear mRNA export and PABPC4, which has been shown to shuttle between the nucleus and cytoplasm.[96] We identified a connection between DDX3X and nuclear export. Consistently, mRNA association kinetics for DDX3X and PABPC4 are similar to those of canonical export factors DDX39B/UAP56 and ALYREF while dissociation is slower, as expected, given their additional cytoplasmic roles.[97] Shuttling between the nucleus and cytoplasm, DDX3X is essential in yeast, cultured human cells, and animals, but the genetics of its essentiality has remained unclear.[63,98] DDX3X is understood as a cytoplasmic translation regulator, but DDX3X's nuclear role has remained unclear.[19,65,99] We found that limiting nuclear mRNA export buffers against loss of DDX3X in terms of cell proliferation. Thus, GIs 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 genetics of DDX3X syndrome and DDX3X-associated cancers indicate selection and associated phenotypic differences between DDX3X depletion or missense variants.[70,100–105] 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. Because DDX3X helicase mutants affect translation,[65,106] DDX3X's role in export is likely distinct from its role in translation, highlighting how an RBP could control multiple points in the RNA life cycle.

Although total RNA poly(A) tail length inversely correlated with whole-cell half-lives in K562, this simple relationship does not capture the dynamics across all compartments. We observed that total RNA tail lengths reflect the ratio of time spent on chromatin and in the cytoplasm for each gene. RNA flow represents the cumulative impact of multiple layers of gene regulation that act to control the subcellular fates and trajectories of transcripts. Investigating RNA flow throughout development and in disease systems could 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 endoplasmic reticulum (ER) with the nuclear fraction.[36,107] We observed that transcripts annotated with GO term ER were enriched in clusters that show a combination of very long nuclear and export and extremely short cytoplasm half-lives (clusters

(D) Top 10 individual features associated with RNA flow 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.

15, 18, 23, and 27; n = 228 genes, Figure 2D). We therefore expect that the nuclear fraction contains a mix of both newly synthesized and preexisting RNAs, resulting in artificially long nuclear and export half-lives for these genes. Conversely, the cytoplasm may be enriched for relatively new transcripts that have not yet been loaded onto polysomes and translated at the ER. Importantly, we found that the nuclear wash step does not add a bias to the RNA flow rate estimations (Figures S1F and S1G).

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. Finally, given the variability in the biochemical properties of different granules and condensates,[108] we are uncertain where all of these may sediment. Nevertheless, nuclear speckle component SRSF7 co-sediments with the chromatin and nucleoplasm fractions (Figure S1D), and P-bodies (marker protein LSM14A) sediment in the cytoplasm (Figure 1B) and do not colocalize with polysomes (Figure S1E). In all, our biochemical purification fractionates most transcripts accurately based on their subcellular localization.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell culture
  - shRNA knockdown of DDX3X, PABPC4, DIS3, EXOSC10, PABPN1, and ZFC3H1 cell line generation
  - CRISPRi DDX3X knockdown and DDX3X mutant cell line generation
- METHOD DETAILS
  - 4sU labeling
  - Cell fractionation
  - RNA extraction
  - TimeLapse-seq chemistry and library preparation
  - Western blotting
  - SABER-FISH
  - NanoStrings quantification of new RNA
  - Subcellular TimeLapse-seq in RBP and exosome shRNA knockdown cell lines
  - Poly(A) selection and direct RNA sequencing
  - Cloning sgRNA and DDX3X mutant overexpression plasmids
  - Genome-scale CRISPRi screening
  - Genome-scale CRISPRi high-throughput sequencing sample preparation
  - Flow-based competition assay
  - Ribosome profiling
  - Cell fractionation and RNA sequencing of CRISPRi lines
  - RNA sequencing of nuclear wash following cell fractionation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Quantification of newly synthesized RNA from subcellular TimeLapse-seq data
  - Kinetic modeling of RNA flow
  - Bayes factor model comparison to identify nuclear RNA degradation
  - Hierarchical clustering of genes by RNA flow rates
  - Functional analysis of genes clustered by RNA flow rates
  - RNA-binding protein associations with RNA flow rates
  - Direct RNA sequencing data analysis
  - LASSO machine learning model for RNA flow rate determinants
  - Transcription factor associations with RNA flow rates
  - Ribosome profiling data analysis
  - Subcellular RNA sequencing of CRISPRi lines data analysis
  - Nuclear wash RNA sequencing data analysis
  - SeqFISH+ comparison analysis
  - Transcript isoform turnover analysis in nuclear exosome mutants

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.06.008.

ACKNOWLEDGMENTS

We thank A. Aker, G. Prakash, I. Patop, R.S. Isaac, H. Merens, A. Sarfatis, C. Cepko, and members of the Churchman lab for helpful discussions; N. Kramer for advice regarding lentiviral transductions; M. Couvillion for creating the SNP-masked genomes; J. Diego Martin Rufino for providing TF target gene sets; J. Liboy and J. Park for technical assistance; D. Martell, J. Bridgers, and I. Patop for critical reading of the manuscript; Gene Yeo lab for sharing eCLIP data; and M. Florio and L. Chan for help with sequencing. This work was supported by National Institutes of Health (NIH) grants R01HG007173 and R21HG011682 (L.S.C.), R01NS120667 (S.N.F.), National Science Foundation Graduate Research Fellowship DGE-1745303 (B.M.S.), NIH T32GM007748 (R.I.), NIH F30HD110250 (A.X.), UCSF Moritz-Heyman Discovery Fellowship (Z.M.J.), and postdoctoral fellowships from the Fonds de Recherche du Québec Santé and the Canadian Institutes of Health Research (K.C.). C.K.G. is a Merck fellow of the Jane Coffin Childs Fund for Medical Research. S.N.F. is a Pew scholar in the biomedical sciences, supported by The Pew Charitable Trusts.

AUTHOR CONTRIBUTIONS

R.I., B.M.S., and L.S.C. conceived and designed all experiments, unless specified otherwise. R.I. conceived and developed the mathematical modeling and machine learning frameworks. E.M. conceived and designed the NanoString assay. A.X., Z.M.J., B.X.H.F., L.G., and S.N.F. conceived and designed the DDX3X CRISPRi screen and its validation experiments. B.M.S., A.X., K.C., E.M., Z.M.J., C.K.G., A.R.B.-K., and E.R.W. performed experiments. R.I., B.M.S., A.X., K.C., E.M., E.R.W., and B.X.H.F. performed data analyses. R.I., B.M.S., and L.S.C. wrote the manuscript, with input from A.X., S.N.F., and other authors.

DECLARATION OF INTERESTS

R.I. is a founder, board member, and/or shareholder of Cellforma, unrelated to the present work.

Received: September 23, 2022
Revised: May 15, 2024
Accepted: June 11, 2024
Published: July 3, 2024

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. https://doi.org/10.1186/gb-2005-6-7-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. https://doi.org/10.1038/nmeth.4582.
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*, e32537. https://doi.org/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*, e45056. https://doi.org/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. https://doi.org/10.1016/j.celrep.2013.08.013.
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. https://doi.org/10.1016/j.molcel.2020.12.018.
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. https://doi.org/10.1016/j.molcel.2019.11.017.
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. https://doi.org/10.1016/j.cell.2014.11.015.
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. https://doi.org/10.1016/j.cell.2021.04.012.
10. 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. https://doi.org/10.1371/journal.pgen.1005610.
11. 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. https://doi.org/10.1261/rna.064899.117.
12. Davidson, L., Kerr, A., and West, S. (2012). Co-transcriptional degradation of aberrant pre-mRNA by Xrn2. EMBO J. *31*, 2566–2578. https://doi.org/10.1038/emboj.2012.101.
13. 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. https://doi.org/10.1016/j.molcel.2016.09.025.
14. 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. https://doi.org/10.1038/emboj.2012.81.
15. 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. https://doi.org/10.1091/mbc.e07-12-1264.
16. 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. https://doi.org/10.1016/j.cell.2008.10.037.
17. 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. USA *106*, 22217–22222. https://doi.org/10.1073/pnas.0909773106.
18. 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. https://doi.org/10.1038/nrm.2017.103.
19. 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. https://doi.org/10.1038/emboj.2012.220.
20. 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. https://doi.org/10.1038/nature13578.
21. 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. https://doi.org/10.1016/j.molcel.2019.12.004.
22. Bartel, D.P. (2018). Metazoan microRNAs. Cell *173*, 20–51. https://doi.org/10.1016/j.cell.2018.03.006.
23. 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. https://doi.org/10.1038/s41580-021-00417-y.
24. 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. https://doi.org/10.1038/nature10098.
25. 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. https://doi.org/10.1093/nar/gkp542.
26. 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. https://doi.org/10.1038/nmeth.4435.
27. 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. https://doi.org/10.1261/rna.1136108.
28. 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. https://doi.org/10.1038/nbt.1861.
29. 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. https://doi.org/10.1126/science.1099754.
30. 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. https://doi.org/10.1038/ncb2056.
31. 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. https://doi.org/10.1016/j.molcel.2019.05.008.
32. Halstead, J.M., Lionnet, T., Wilbertz, J.H., Wippich, F., Ephrussi, A., Singer, R.H., and Chao, J.A. (2015). An RNA biosensor for imaging the first round of translation from single cells to living animals. Science *347*, 1367–1671. https://doi.org/10.1126/science.aaa3380.
33. 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. https://doi.org/10.1016/j.celrep.2015.11.036.
34. 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. https://doi.org/10.1016/j.cels.2017.05.010.

35. 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. USA *113*, 218–223. https://doi.org/10.1073/pnas.1518369112.
36. 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. https://doi.org/10.1016/j.cell.2012.05.043.
37. Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of transcript variability in single mammalian cells. Cell *163*, 1596–1610. https://doi.org/10.1016/j.cell.2015.11.018.
38. 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.
39. 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. https://doi.org/10.1016/j.cels.2022.04.005.
40. Chen, T., and van Steensel, B. (2017). Comprehensive analysis of nucleocytoplasmic dynamics of mRNA in Drosophila cells. PLoS Genet. *13*, e1006929. https://doi.org/10.1371/journal.pgen.1006929.
41. 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. https://doi.org/10.1038/s41592-021-01250-z.
42. 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 *20*, 695–705. https://doi.org/10.1038/s41592-023-01829-8.
43. 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. https://doi.org/10.4161/rna.26214.
44. 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. https://doi.org/10.1002/cpmb.44.
45. Jackson, D.A., Hassan, A.B., Errington, R.J., and Cook, P.R. (1993). Visualization of focal sites of transcription within human nuclei. EMBO J. *12*, 1059–1065. https://doi.org/10.1002/j.1460-2075.1993.tb05747.x.
46. Reyes, J.C., Muchardt, C., and Yaniv, M. (1997). Components of the human SWI/SNF complex are enriched in active chromatin and are associated with the nuclear matrix. J. Cell Biol. *137*, 263–274. https://doi.org/10.1083/jcb.137.2.263.
47. 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. https://doi.org/10.1002/anie.201707465.
48. 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. https://doi.org/10.1038/s41586-019-1369-y.
49. 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. https://doi.org/10.1016/j.cell.2015.02.029.
50. 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. https://doi.org/10.1016/j.molcel.2017.09.003.
51. Wu, Q., Medina, S.G., Kushawah, G., DeVore, M.L., Castellano, L.A., Hand, J.M., Wright, M., and Bazzini, A.A. (2019). Translation affects mRNA stability in a codon-dependent manner in human cells. eLife *8*, e45396. https://doi.org/10.7554/eLife.45396.
52. 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*, eaay6912. https://doi.org/10.1126/science.aay6912.
53. Schmid, M., and Jensen, T.H. (2018). Controlling nuclear RNA levels. Nat. Rev. Genet. *19*, 518–529. https://doi.org/10.1038/s41576-018-0013-2.
54. Kass, R.E., and Raftery, A.E. (1995). Bayes factors. J. Am. Stat. Assoc. *90*, 773–795. https://doi.org/10.1080/01621459.1995.10476572.
55. Eng, C.L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G.-C., et al. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. Nature *568*, 235–239. https://doi.org/10.1038/s41586-019-1049-y.
56. 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. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.
57. 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. https://doi.org/10.1016/j.cels.2015.12.004.
58. 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. https://doi.org/10.1038/s41586-020-2077-3.
59. 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. https://doi.org/10.1016/j.cell.2004.09.029.
60. Kanai, Y., Dohmae, N., and Hirokawa, N. (2004). Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuron *43*, 513–525. https://doi.org/10.1016/j.neuron.2004.07.022.
61. 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. https://doi.org/10.1042/BJ20110739.
62. 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. (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. https://doi.org/10.1016/j.bbagrm.2016.03.009.
63. Brennan, R., Haap-Hoff, A., Gu, L., Gautier, V., Long, A., and Schröder, M. (2018). Investigating nucleo-cytoplasmic shuttling of the human DEAD-box helicase DDX3. Eur. J. Cell Biol. *97*, 501–511. https://doi.org/10.1016/j.ejcb.2018.08.001.
64. 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. https://doi.org/10.1038/s41586-019-1551-2.
65. 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. https://doi.org/10.1093/nar/gkab287.

66. 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. https://doi.org/10.1128/MCB.01683-13.

67. 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. https://doi.org/10.7554/eLife.19760.

68. 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. https://doi.org/10.1016/j.jmb.2015.02.015.

69. 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. https://doi.org/10.1074/jbc.M115.700625.

70. 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. https://doi.org/10.1016/j.neuron.2020.01.042.

71. Alles, J., Legnini, I., Pacelli, M., and Rajewsky, N. (2023). Rapid nuclear deadenylation of mammalian messenger RNA. iScience *26*, 105878. https://doi.org/10.1016/j.isci.2022.105878.

72. 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. https://doi.org/10.1016/j.molcel.2019.12.005.

73. 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. https://doi.org/10.1038/s41592-019-0617-2.

74. 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. https://doi.org/10.1038/nature13007.

75. 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. https://doi.org/10.1038/nsmb.3499.

76. Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. Genes Dev. *29*, 63–80. https://doi.org/10.1101/gad.247361.114.

77. Gockert, M., Schmid, M., Jakobsen, L., Jens, M., Andersen, J.S., and Jensen, T.H. (2022). Rapid factor depletion highlights intricacies of nucleoplasmic RNA degradation. Nucleic Acids Res. *50*, 1583–1600. https://doi.org/10.1093/nar/gkac001.

78. Torre, D., Fstkchyan, Y.S., Ho, J.S.Y., Cheon, Y., Patel, R.S., Degrace, E.J., Mzoughi, S., Schwarz, M., Mohammed, K., Seo, J.-S., et al. (2023). Nuclear RNA catabolism controls endogenous retroviruses, gene expression asymmetry, and dedifferentiation. Mol. Cell 83, 4255–4271.e9. https://doi.org/10.1016/j.molcel.2023.10.036.

79. Fan, J., Kuai, B., Wu, G., Wu, X., Chi, B., Wang, L., Wang, K., Shi, Z., Zhang, H., Chen, S., et al. (2017). Exosome cofactor hMTR4 competes with export adaptor ALYREF to ensure balanced nuclear RNA pools for degradation and export. EMBO J. *36*, 2870–2886. https://doi.org/10.15252/embj.201696139.

80. 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. https://doi.org/10.1093/dnares/dsn030.

81. 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 Res. *23*, 2078–2090. https://doi.org/10.1101/gr.156919.113.

82. 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. https://doi.org/10.1261/rna.062224.117.

83. Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. eLife *7*, e32536. https://doi.org/10.7554/eLife.32536.

84. 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. https://doi.org/10.1186/s12915-021-00949-x.

85. 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. https://doi.org/10.1101/gr.1272403.

86. Agarwal, V., and Kelley, D.R. (2022). The genetic and biochemical determinants of mRNA degradation rates in mammals. Genome Biol. *23*, 245. https://doi.org/10.1186/s13059-022-02811-x.

87. Cockman, E., Anderson, P., and Ivanov, P. (2020). TOP mRNPs: molecular mechanisms and principles of regulation. Biomolecules *10*, 969. https://doi.org/10.3390/biom10070969.

88. Beaulieu, Y.B., Kleinman, C.L., Landry-Voyer, A.-M., Majewski, J., and Bachand, F. (2012). Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. PLoS Genet. *8*, e1003078. https://doi.org/10.1371/journal.pgen.1003078.

89. Bresson, S.M., and Conrad, N.K. (2013). The human nuclear poly(a)-binding protein promotes RNA hyperadenylation and decay. PLoS Genet. *9*, e1003893. https://doi.org/10.1371/journal.pgen.1003893.

90. Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J.J., Yates, J.R., 3rd, Tian, B., and Manley, J.L. (2017). An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. Genes Dev. *31*, 1257–1271. https://doi.org/10.1101/gad.302604.117.

91. Silla, T., Karadoulama, E., Mąkosa, D., Lubas, M., and Jensen, T.H. (2018). The RNA exosome adaptor ZFC3H1 functionally competes with nuclear export activity to retain target transcripts. Cell Rep. *23*, 2199–2210. https://doi.org/10.1016/j.celrep.2018.04.061.

92. Wu, M., Karadoulama, E., Lloret-Llinares, M., Rouviere, J.O., Vaagensø, C.S., Moravec, M., Li, B., Wang, J., Wu, G., Gockert, M., et al. (2020). The RNA exosome shapes the expression of key protein-coding genes. Nucleic Acids Res. *48*, 8509–8528. https://doi.org/10.1093/nar/gkaa594.

93. Wang, Y., Fan, J., Wang, J., Zhu, Y., Xu, L., Tong, D., and Cheng, H. (2021). ZFC3H1 prevents RNA trafficking into nuclear speckles through condensation. Nucleic Acids Res. *49*, 10630–10643. https://doi.org/10.1093/nar/gkab774.

94. Berry, S., and Pelkmans, L. (2022). Mechanisms of cellular mRNA transcript homeostasis. Trends Cell Biol. *32*, 655–668. https://doi.org/10.1016/j.tcb.2022.05.003.

95. Han, X., Xing, L., Hong, Y., Zhang, X., Hao, B., Lu, J.Y., Huang, M., Wang, Z., Ma, S., Zhan, G., et al. (2024). Nuclear RNA homeostasis promotes systems-level coordination of cell fate and senescence. Cell Stem Cell *31*, 694–716.e11. https://doi.org/10.1016/j.stem.2024.03.015.

96. 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. https://doi.org/10.1242/jcs.087692.

97. Choi, Y., Um, B., Na, Y., Kim, J., Kim, J.-S., and Kim, V.N. (2024). Time-resolved profiling of RNA binding proteins throughout the mRNA life cycle. Mol. Cell *84*, 1764–1782.e10. https://doi.org/10.1016/j.molcel.2024.03.012.
98. Sharma, D., and Jankowsky, E. (2014). The Ded1/DDX3 subfamily of DEAD-box RNA helicases. Crit. Rev. Biochem. Mol. Biol. *49*, 343–360. https://doi.org/10.3109/10409238.2014.931339.
99. 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. https://doi.org/10.18632/oncotarget.8612.
100. 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. https://doi.org/10.1016/j.ajhg.2015.07.004.
101. Scala, M., Torella, A., Severino, M., Morana, G., Castello, R., Accogli, A., Verrico, A., Vari, M.S., Cappuccio, G., Pinelli, M., et al. (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. https://doi.org/10.1038/s41431-019-0392-7.
102. 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. https://doi.org/10.1038/nature11329.
103. 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. https://doi.org/10.1038/nature11213.
104. 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. (2015). Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. Nat. Genet. *47*, 1061–1066. https://doi.org/10.1038/ng.3358.
105. 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. https://doi.org/10.1016/j.molcel.2021.07.041.
106. 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. https://doi.org/10.1038/srep25996.
107. 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. https://doi.org/10.1007/978-1-4939-7204-3_12.
108. 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. https://doi.org/10.1016/j.jmb.2018.08.003.
109. Melé, M., Mattioli, K., Mallard, W., Shechner, D.M., Gerhardinger, C., and Rinn, J.L. (2017). Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. Genome Res. *27*, 27–37. https://doi.org/10.1101/gr.214205.116.
110. Ietswaart, R. (2024). RNA flow additional data sets. https://doi.org/10.5281/zenodo.11356463.
111. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.
112. 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. https://doi.org/10.1093/bioinformatics/bts635.
113. 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. https://doi.org/10.1093/bioinformatics/btr509.
114. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.
115. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
116. Jürges, C., Dölken, L., and Erhard, F. (2018). Dissecting newly transcribed and old RNA using GRAND-SLAM. Bioinformatics *34*, i218–i226. https://doi.org/10.1093/bioinformatics/bty256.
117. Ge, S.X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics *36*, 2628–2629. https://doi.org/10.1093/bioinformatics/btz931.
118. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
119. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
120. 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. https://doi.org/10.1093/nar/gkz114.
121. 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. https://doi.org/10.1186/s13059-014-0550-8.
122. 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. https://doi.org/10.1093/bioinformatics/btx047.
123. 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. https://doi.org/10.1016/j.molcel.2016.02.012.
124. 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. https://doi.org/10.1016/j.cell.2014.09.029.
125. Martell, D.J., Ietswaart, R., Smalec, B.M., and Churchman, L.S. (2021). Profiling metazoan transcription genome-wide with nucleotide resolution using NET-seq (native elongating transcript sequencing). protocols.io. https://doi.org/10.17504/protocols.io.bpymmpu6.
126. 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. https://doi.org/10.1038/s41592-019-0404-0.
127. 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. https://doi.org/10.1038/s41592-021-01187-3.

128. 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. https://doi.org/10.1016/j.celrep.2021.110191.
129. 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. https://doi.org/10.1126/science.aad9841.
130. 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. https://doi.org/10.1126/science.aac7041.
131. 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. https://doi.org/10.1126/science.aac7557.
132. 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. https://doi.org/10.1016/j.molcel.2020.07.007.
133. McShane, E., Couvillion, M., Ietswaart, R., Prakash, G., Smalec, B.M., Soto, I., Baxter-Koenigs, A.R., Choquet, K., and Churchman, L.S. (2024). A kinetic dichotomy between mitochondrial and nuclear gene expression processes. Mol. Cell *84*, 1541–1555.e11. https://doi.org/10.1016/j.molcel.2024.02.028.
134. Zhang, Y., Park, C., Bennett, C., Thornton, M., and Kim, D. (2021). HISAT-3N: a rapid and accurate three-nucleotide sequence aligner. Preprint at bioRxiv. https://doi.org/10.1101/2020.12.15.422906.
135. Zimmer, J.T., Vock, I.W., Schofield, J.A., Kiefer, L., Moon, M.H., and Simon, M.D. (2023). Improving the study of RNA dynamics through advances in RNA-seq with metabolic labeling and nucleotide-recoding chemistry. Preprint at bioRxiv. https://doi.org/10.1101/2023.05.24.542133.
136. Berg, K., Lodha, M., Delazer, I., Bartosik, K., Garcia, Y.C., Hennig, T., Wolf, E., Dölken, L., Lusser, A., Prusty, B.K., et al. (2024). Correcting 4sU induced quantification bias in nucleotide conversion RNA-seq data. Nucleic Acids Res. *52*, e35. https://doi.org/10.1093/nar/gkae120.
137. Adams, R.A., and Essex, C. (2021). Calculus: A Complete Course (Pearson Education).
138. 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. https://doi.org/10.1038/s41598-018-28948-z.
139. 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. https://doi.org/10.1186/s13059-021-02264-8.
140. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514. https://doi.org/10.1038/nmeth.3810.
141. 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. https://doi.org/10.1038/s41596-020-00469-y.
142. Hastie, T., Friedman, J., and Tibshirani, R. (2001). The Elements of Statistical Learning (Springer).
143. 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. USA *119*, e2207392119. https://doi.org/10.1073/pnas.2207392119.
144. Martell, D.J., Merens, H.E., Caulier, A., Fiorini, C., Ulirsch, J.C., Ietswaart, R., Choquet, K., Graziadei, G., Brancaleoni, V., Cappellini, M.D., et al. (2023). RNA polymerase II pausing temporally coordinates cell cycle progression and erythroid differentiation. Dev. Cell *58*, 2112–2127.e4. https://doi.org/10.1016/j.devcel.2023.07.018.

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-PABPC4 antibody	Bethyl Laboratories	Cat# A301-466A; RRID: AB_999661
Rabbit polyclonal anti-DDX3X antibody	Bethyl Laboratories	Cat# A300-474A; RRID: AB_451009
Rabbit polyclonal anti-DIS3 antibody	Bethyl Laboratories	Cat# A303-764A-T; RRID: AB_11204773
Rabbit polyclonal anti-EXOSC10 antibody	Bethyl Laboratories	Cat# A303-989A-T; RRID: AB_2620338
Rabbit polyclonal anti-PABPN1 antibody	MBL International	Cat# RN023PW; RRID: AB_10597725
Rabbit polyclonal anti-ZFC3H1 antibody	Bethyl Laboratories	Cat# A301-456A; RRID: AB_960958
Mouse monoclonal anti-GAPDH antibody	Thermo	Cat# MA5-15738; RRID: AB_10977387
Rat monoclonal anti-RNA pol II antibody	Active Motif	Cat# 61986; RRID: AB_2687451
Rabbit polyclonal anti-Histone H3 antibody	Abcam	Cat# ab1791; RRID: AB_302613
Mouse monoclonal anti-U1 snRNP 70 antibody	Santa Cruz Biotechnology	Cat# sc-390899; RRID: AB_2801569
Mouse monoclonal anti-LSM14A antibody	Santa Cruz Biotechnology	Cat# sc-398552; RRID: AB_3099527
Rabbit polyclonal anti-RPS6 antibody	Proteintech	Cat# 14823-1-AP; RRID: AB_2181025
Rabbit monoclonal anti-Lamin B1 antibody	Abcam	Cat# ab229025; RRID: AB_3083735
Mouse monoclonal anti-alpha-Tubulin antibody	Abcam	Cat# ab7291; RRID: AB_2241126
Goat polyclonal anti-rabbit IRDye 800CW antibody	Licor	Cat# 926-32211; RRID: AB_621843
Goat polyclonal anti-mouse IRDye 680RD antibody	Licor	Cat# 926-68070; RRID: AB_10956588
Goat polyclonal anti-Rat Cyanine5 antibody	Invitrogen	Cat# A10525; RRID: AB_2534034
Donkey polyclonal anti-Mouse 405 antibody	Jackson Immunoresearch	Cat# 715-475-150; RRID: AB_2340839
Donkey polyclonal anti-Mouse 488 antibody	Jackson Immunoresearch	Cat# 715-545-150; RRID: AB_2340846
Donkey polyclonal anti-Rabbit 405 antibody	Jackson Immunoresearch	Cat# 711-475-152; RRID: AB_2340616
Donkey polyclonal anti-Rabbit 488 antibody	Jackson Immunoresearch	Cat# 711-545-152; RRID: AB_2313584
DDX3X antibody (Figures 4 and S7 only)	Custom made by Genemed Synthesis using peptide ENALGLDQQFAGLDLNSSDNQS	N/A
Mouse monoclonal anti-beta-actin antibody (Figures 4 and S7 only)	Santa Cruz Biotechnology	Cat# sc-47778-AF680; RRID: AB_626632
Mouse monoclonal anti-GADPH antibody (Figures 4 and S7 only)	GeneTex	Cat# GTX627408; RRID: AB_11174761
Goat polyclonal anti-mouse IgG 800 antibody (Figures 4 and S7 only)	Licor	Cat# 926-32210; RRID: AB_621842
Donkey polyclonal anti-rabbit IgG 680 antibody (Figures 4 and S7 only)	Licor	Cat# 926-68073; RRID: AB_10954442
Rabbit polyclonal anti-SFRS7/ SRSF7 antibody	Abcam	Cat# ab138022
Chemicals, peptides, and recombinant proteins		
2,2,2-trifluoroethylamine (TFEA)	Sigma-Aldrich	Cat# 269042
Sodium periodate	Sigma-Aldrich	Cat# 311448
Critical commercial assays		
SMARTer Stranded Total RNA HI Mammalian kit	Takara	Cat# 634873

(*Continued on next page*)

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
uMACS Streptavidin kit	Miltenyi Biotec	Cat# 130-074-101
Direct RNA sequencing kit	Oxford Nanopore Technologies	Cat# SQK-RNA002
Deposited data		
ENCODE K562 eCLIP data	Van Nostrand et al.[58]	https://www.encodeproject.org/
K562 transcript half-life data	Wachutka et al.,[4] Table S3	N/A
K562 transcript half-life data	Schofield et al.,[2] Table S2	N/A
K562 transcript half-life data	Wu et al.,[51] source data 1 (Figure 1)	N/A
K562 transcript half-life data	Mele et al.,[109] Table S10	N/A
Whole-cell half-life model and "consensus" half-lives	Agarwal and Kelley[86]	https://doi.org/10.5281/zenodo.6326409
Raw and analyzed subcellular TimeLapse-seq data	This study	GEO: GSE207924
Raw and analyzed direct RNA sequencing data	This study	GEO: GSE208225
Raw and analyzed NanoStrings data	This study	GEO: GSE207925
CRISPRi GI map data	This study	GEO: GSE230018
Ribosome profiling data	This study	GEO: GSE230029
DDX3X subcellular fractionation RNA sequencing data	This study	GEO: GSE231290
Nuclear Wash subcellular fractionation RNA sequencing data	This study	GEO: GSE264556
Additional data	This study	Zenodo[110]; https://www.doi.org/10.5281/zenodo.11356463
Experimental models: Cell lines		
Mouse: NIH-3T3 cells	ATCC	Cat# CRL-1658
Human: K562 cells	ATCC	Cat# CCL-243
Human: HEK-293T cells	ATCC	Cat# CRL-3216
Oligonucleotides		
All primers and protospacers	See Table S7 for all sequences	N/A
Recombinant DNA		
TRC Lentiviral Human DDX3X shRNA plasmid	Horizon	Cat# TRCN0000000003
TRC Lentiviral Human PABPC4 shRNA plasmid	Horizon	Cat# TRCN0000074658
TRC Lentiviral Human scramble shRNA plasmid	Addgene	Cat# 1864
TRIPZ Inducible Lentiviral Human DIS3 shRNA plasmid	Horizon	Cat# V2THS_96258
TRIPZ Inducible Lentiviral Human EXOSC10 shRNA plasmid	Horizon	Cat# V2THS_275659
TRIPZ Inducible Lentiviral Human PABPN1 shRNA plasmid	Horizon	Cat# V2THS_41638
TRIPZ Inducible Lentiviral Human ZFC3H1 shRNA plasmid	Horizon	Cat# V2THS_35985
TRIPZ Inducible Lentiviral Non-silencing shRNA control plasmid	Horizon	Cat# RHS4743
Lentiviral packaging psPAX2 plasmid	Addgene	Cat# 12260
Lentiviral envelope pMD2.G plasmid	Addgene	Cat# 12259
Lentiviral plasmid: hU6-sgRNA-EEF1A-GFP-modified-BstXI	Gift from Luke Gilbert	pAX71

(*Continued on next page*)

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Lentiviral plasmid: pCRISPRia-v2-mU6-sgRNA-EEF1A-Puro-T2A-BFP	Addgene	Cat# 84832
Lentiviral plasmid: UCOE-SFFV-DDX3_wt-3xflag-2xP2A-mCherry	This study	pAX55
Lentiviral plasmid: UCOE-SFFV-DDX3_R534S-3xflag-2xP2A-mCherry	This study	pAX66
Lentiviral plasmid: UCOE-SFFV-DDX3_R326H-3xflag-2xP2A-mCherry	This study	pAX67
Lentiviral plasmid: UCOE-SFFV-DDX3_R351W-3xflag-2xP2A-mCherry	This study	pAX68
Lentiviral plasmid: UCOE-SFFV-DDX3_R376C-3xflag-2xP2A-mCherry	This study	pAX69
Lentiviral plasmid: UCOE-SFFV-DDX3_G325E-3xflag-2xP2A-mCherry	This study	pAX70
Lentiviral plasmid: UCOE-SFFV-DDX3_T532M-3xflag-2xP2A-mCherry	This study	pAX105
Software and algorithms		
cutadapt v2.5	Martin[111]	https://cutadapt.readthedocs.io/en/v2.5/
STAR (v2.7.3a, v2.7.0a)	Dobin et al.[112]	https://github.com/alexdobin/STAR
BCFtools	Li[113]	https://github.com/samtools/bcftools
GATK	McKenna et al.[114]	https://github.com/broadinstitute/gatk
rf2m	N/A	https://github.com/LaboratorioBioinformatica/rf2m
samtools v1.9	Li et al.[115]	https://github.com/samtools/samtools
GRAND-SLAM v2.0.5d	Jürges et al.[116]	https://github.com/erhard-lab/gedi/wiki/GRAND-SLAM
GSEA	Subramanian et al.[56]	https://www.gsea-msigdb.org/gsea/index.jsp
MSigDB v7.5.1	Liberzon et al.[57]	https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/MSigDB_v7.5.1_Release_Notes
ShinyGO 0.76.2	Ge et al.[117]	http://bioinformatics.sdstate.edu/go76/
minimap2 v2.10-r764-dirty	Li[113]	https://github.com/lh3/minimap2
nanopolish v0.13.3	Workman et al.[73]	https://github.com/jts/nanopolish
bedtools	Quinlan and Hall[118]	https://github.com/arq5x/bedtools2
Bowtie2 v2.4.1	Langmead and Salzberg[119]	https://github.com/BenLangmead/bowtie2
RSubread	Liao et al.[120]	https://bioconductor.org/packages/release/bioc/html/Rsubread.html
DESeq2	Love et al.[121]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Riborex	Li et al.[122]	https://github.com/smithlabcode/riborex

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, L. Stirling Churchman (churchman@genetics.med.harvard.edu).

Materials availability

All unique reagents generated in this study will be available from the lead contact upon request.

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

Data and code availability

- Raw and processed data are available as the date of publication from the Gene Expression Omnibus at accession numbers GEO: GSE207924 (Subcellular TimeLapse-seq data), GEO: GSE208225 (Direct RNA sequencing data), GEO: GSE207925 (NanoString data), GEO: GSE230018 (CRISPRi GI map data), GEO: GSE230029 (Ribosome Profiling data), GEO: GSE231290 (DD3X subcellular RNA-sequencing data), GEO: GSE264556 (Nuclear Wash subcellular RNA-sequencing data), and on Zenodo[110] at https://www.doi.org/10.5281/zenodo.11356463 (additional data).
- Code for analysis of all data are available at http://github.com/churchmanlab/rna_flow.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

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).

shRNA knockdown of DDX3X, PABPC4, DIS3, EXOSC10, PABPN1, and ZFC3H1 cell line generation

K562 knockdown lines were generated according to Sundararaman et al.[123] 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 $0.2\text{-}1.0\times10^6$ 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 hours prior to cell harvest, and media was refreshed with new doxycycline every 24 hours. All knockdowns were confirmed by western blotting analyses.

CRISPRi DDX3X knockdown and DDX3X mutant cell line generation

The doxycycline-inducible and constitutive K562 CRISPRi cell lines were generated in previous work.[124] 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.

METHOD DETAILS

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.

Cell fractionation

Chromatin, nuclear and cytoplasm RNA

Cells were fractionated as per Martell et al.[125] 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).

TimeLapse-seq chemistry and library preparation

Samples were prepared for sequencing according to Schofield et al.[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, which includes rRNA depletion, and samples were sequenced on the NovaSeq (Illumina, San Diego, CA) by the Bauer Core Facility at Harvard University.

Western blotting

Samples were mixed at 1:1 volume with 2x Laemmli buffer (4% SDS, 20% glycerol, 0.2M DTT, 0.1M Tris-HCl pH 7.0, 0.02% bromophenol blue), denatured at 95°C for 5 minutes, and kept on ice. Samples were loaded onto a 4-12% Bis-Tris gel (Invitrogen NP0321BOX) in 1x MOPs buffer and run at 160V for 1 hour. The gel was transferred to a nitrocellulose membrane using the wet transfer method in 1x transfer buffer (25mM Tris base, 192mM glycine, 20% methanol) at 400mA for 75 minutes at 4°C. The membrane was blocked in 1x blocking buffer (5% non-fat milk powder in 1x TBST) for at least 60 minutes. Primary antibodies were diluted according to (Table S7) in 1x blocking buffer and incubated with membranes for at least 16 hours at 4°C. Membranes were washed 4x 5 minutes with 1x TBST, incubated for 1 hour at 25°C with secondary antibodies (Table S7), washed again 4x 5 minutes with 1x TBST, and imaged using a Li-Cor Odyssey.

For experiments in Figures 4 and S6, 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.

SABER-FISH

mRNA transcripts corresponding to *Foxo3, Smad3, Gfod1*, and *Myc* were detected in NIH-3T3 cells by smRNA-FISH according to Kishi et al.[126] Briefly, 50–80 probes were designed per gene using PaintSHOP[127] with additional sequences used for SABER appended to the 3' end in a gene-specific manner (Table S7). All probes corresponding to the same gene were pooled at 10uM in 1x TE buffer (10 mM Tris pH 8.0, 0.1 mM EDTA). Probes were synthesized by first preparing a mix containing 10uL of 5uM hairpin oligo, 10uL of 10x PBS, 10uL of 100mM $MgSO_4$, 5uL of dNTP mix (containing 6mM of each A,C,T), 10uL of 1uM Clean.g oligo, 0.5uL of BST enzyme (McLab, BPL-300), and 44.5uL water (see Table S7 for oligo sequences). The reaction mix was incubated at 37°C for 15 minutes and then 10uL of 10uM pooled probes were added. Probes were then concatemerized by incubating at 37°C for 60 minutes before enzyme inactivation at 80°C for 20 minutes. Probes were purified using the MinElute PCR Purification kit (Qiagen 28004) and quantified by Nanodrop (ssDNA setting).

Cells were grown in 8-well poly-L-lysine coated chamber slides (ibidi 80826), labeled with 500uM 4sU for 2 hours, and fixed with 4% PFA in 1x PBS for 10 minutes. Unlabeled cells were also included as controls. Slides were washed 3x 5 minutes in 1x PBST and then incubated in 1x Whyb solution (2x SSC, 1% Tween-20, and 40% deionized formamide) for at least 1 hour at 43°C. 1ug of concatemerized probes were then incubated on slides in 1x Hyb solution (2x SSC 1% Tween-20, 40% deionized formamide, 10% dextran sulfate) for at least 16 hours at 43°C (total volume per ibidi chamber well of 150uL). Slides were wrapped in parafilm and placed in a humidifying chamber within the oven to prevent evaporation. After probe hybridization, each well was washed 2x 30 minutes with 1x Whyb solution (pre-warmed to 43°C), followed by 2x 5 minutes in 2x SSC, 0.1% Tween-20 (pre-warmed to 43°C). Slides were then moved to room temperature and washed 2x 1 minute with 1x PBST (1x PBS, 0.1% Tween). For fluorescent detection of the mRNAs, slides were then transferred to an oven set at 37°C and pre-warmed for 10 minutes. Fluorescent imager oligos were then incubated with the sample at 37°C for 10 minutes, each at 0.2 uM concentration in Imager Hyb (1x PBS, 0.2% Tween). Each well was then washed 3x 5 minutes in 1x PBST at 37°C, before the sample was brought to room temperature.

Samples were then blocked for 1 hour at room temperature in Blocking Solution (1x PBST, 10% Molecular-grade BSA (ThermoFisher AM2616)). Antibodies to detect Lamin B1 and Tubulin were applied in Blocking Solution and incubated for 2 hours at room temperature (see Table S7 for antibody details). Samples were washed 3x 5 minutes with 1x PBST before secondary antibody incubation, which were applied for 1 hour at room temperature in Blocking Solution (see Table S7 for antibody details). Before imaging, samples were washed 3x 5 minutes with 1x PBST and all samples were imaged in 1x PBST.

Images were acquired and puncta were detected according to West et al.[128] 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.

NanoStrings quantification of new RNA

RNA was denatured at 65°C for 5 minutes and immediately incubated on ice for 2 minutes. The biotinylation reaction was performed with 15ug of denatured RNA in 78ul water, 10ul of MTSEA biotin-XX (Biotium 90066) at 0.25mg/mL in dimethylformamide, and 10uL of 10x buffer (100mM Tris pH 7.5, 10mM EDTA) on a thermoblock at 24°C at 800 RPM for 30 minutes. The biotinylated RNA was purified using Phase Lock tubes (QuantaBio 2302830) using standard chloroform/isoamyl alcohol and ethanol precipitation and quantified by Nanodrop. To synthesize the spike-in RNA, ERCC-00048 DNA with an upstream T7 promoter was cloned into pUC19 and PCR amplified (primers in Table S7) with Phusion polymerase (New England Biolabs M0530S) using the following cycling conditions: 98°C for 30 seconds, then 35 cycles of 98°C for 10 seconds, 61°C for 15 seconds, and 72°C for 30 seconds, followed by a final extension of 72°C for 2 minutes. The PCR product was cleaned using the Monarch PCR clean up kit (New England Biolabs T1030S) according to the manufacturer's protocol, quantified by Nanodrop, and used as the template for *in vitro* transcription reaction (New England Biolabs E2040S) performed according to the manufacturer's protocol. RNA was purified using standard ethanol precipitation and quantified by Nanodrop.

3ug of biotinylated RNA and 60pg of *in vitro* transcribed spike-in RNA (ERCC-00048) in 200ul water was mixed with 100ul of beads from the uMACS Streptavidin kit (Miltenyi Biotec 130-074-101) on a thermoblock at 24°C at 800 RPM for 15 minutes. The columns were washed once with 900uL of wash buffer (100mM Tris pH 7.5, 10mM EDTA, 1M NaCl, 0.1% Tween-20). The RNA/bead mixture was then passed through the washed columns twice and the flow-through RNA was collected, purified using the miRNeasy Nano kit (Qiagen 217084) according to Schwalb et al.,[129] and quantified by Nanodrop. 150ng of RNA was hybridized with gene-specific DNA probes (Table S7), 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.

Subcellular TimeLapse-seq in RBP and exosome shRNA knockdown cell lines

For all knockdowns (DDX3X, PABPC4, DIS3, EXOSC10, ZFC3H1, PABPN1), K562 cells were labeled for 60 minutes with 50uM 4sU (see "4sU labeling" section above for additional details). For inducible knockdowns (DIS3, EXOSC10, ZFC3H1, PABPN1), 4sU labeling was started after 47 hours of beginning the doxycycline treatment, such that cells were exposed to doxycycline fo 48 hours and 4sU for 1 hour at the time of harvest. Cells were harvested and fractionated according to the same protocol as WT cells (see "cell fractionation" section above for additional details).

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) (primers in Table S7) in a total volume of 100 uL using the following cycling conditions: 3 minutes at 95°C, then 30 cycles of 15 seconds at 95°C, 15 seconds at 62°C, 2 minutes at 72°C. The PCR amplicons were purified using 1X volume RNA Clean XP beads and eluted in 33 uL water. A second round of PCR was performed with nested primers, wherein the forward primer encodes a T7 RNA polymerase promoter site and the reverse primers have either 10, 15, 30, 60, 80, or 100 thymidines on the 5' end (primers in Table S7) 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.

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 Horlbeck et al.[67] 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 on.[67] 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

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

were collected after 14 days of doxycycline induction (7 doublings) (100 million cells were collected per replicate). 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,[130,131] while shorter growth periods might reduce genetic effect sizes. 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 Horlbeck et al.[67] 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.[67] 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.[132]

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. Log_2 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 Calviello et al.[65] 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.

Cell fractionation and RNA sequencing of CRISPRi lines

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" 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.

RNA sequencing of nuclear wash following cell fractionation

RNA was sequenced from four biological replicates of approximately 5 million K562 cells each following fractionation. K562 cells were fractionated as described above in the "cell fractionation" section into cytoplasm 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. Total (whole cell) RNA samples from 5 M K562 cells were also collected in parallel.

A yeast RNA spike-in mix was generated by mixing six *in vitro* transcribed transcripts at set ratios (sequences and ratios in Table S7). 10 ng of the spike-in mix was added to the total volume of each fraction (total, 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. Extracted RNA was treated with DNase I (NEB) according to the manufacturer's protocol to remove genomic DNA. Following DNase I treatment, RNA was purified using the RNA Clean & Concentrator-5 kit (Zymo Research). RNA sequencing libraries were then prepared with the SMARTer Stranded Total

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

CellPress

RNA HI Mammalian kit (Takara 634873) with 1 μg of RNA as input in all cases except nuclear wash samples prepared without rRNA depletion (see below). Nuclear wash libraries were generated with and without rRNA depletion. For libraries without rRNA depletion, 100 ng total RNA was fragmented for two minutes at 94 °C with the NEBNext Magnesium RNA Fragmentation Module (NEB) according to the manufacturer's protocol to generate fragments ~200-350 nucleotides in length and isolated via ethanol precipitation. Fragmented RNA samples were then input into the SMARTer Stranded Total RNA HI Mammalian kit at the First-Strand cDNA Synthesis step (Protocol B) with the following modification. 13.5 μL of fragmented RNA was mixed with 2 μL 10X Total RNA First-Strand Buffer (TakaraBio). Tubes were incubated at 72 °C in a preheated, hot-lid thermal cycler for 3 minutes to allow for reverse transcription primer annealing without further fragmentation of RNA. The libraries were then prepared according to the manufacturer's protocol from Step 3 of the First-Strand cDNA Synthesis onward. All libraries were sequenced on a NextSeq 500 (Illumina, San Diego, CA) using 2 x 75 bp paired-end reads.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of newly synthesized RNA from subcellular TimeLapse-seq data

Reads were filtered for quality and adaptor sequences were trimmed using cutadapt v2.5.[111] 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 for K562 and NIH-3T3 cell lines were prepared as described in McShane et al.[133] and are available on Zenodo.[110] In brief, we used reads from subcellular Timelapse-seq samples without 4sU treatment from the total, chromatin and polysome compartments. Reads were first aligned to the reference hg38 genome with STAR using parameters –outFilterMultimapNmax 100 –outFilterMismatchNoverLmax 0.09 –outFilterMismatchNmax 15 –outFilterMatchNminOverLread 0.66 –outFilterScoreMinOverLread 0.66 –outFilterMultimapScoreRange 0 –seedSearchStartLmax 5. Variants were then called with BCFtools 'mpileup' and 'call' using at least four 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 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. Variants covered by < 5 reads or with low quality (Phred-scaled quality score < 20) are ignored and the reference kept as-is. The reference FASTA was modified for these non-INDEL substitutions using GATK FastaAlternateReferenceMaker. Finally, rf2m was used with the INDEL-only VCF file to further modify the FASTA genome reference as well as the corresponding GTF annotation file. A modified version of one of the rf2m scripts (genome_creator_corrected.pl) is required for accurate GTF update. All scripts to generate the SNP-masked genomes can be found here: https://github.com/churchmanlab/mitoRNA_kinetics/tree/main/PipelineForHalfLifeDetermination/CustomGenomeGeneration.

The trimmed, filtered reads were aligned to the appropriate SNP-masked genome using STAR v2.7.0a (Figure S1I), 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.[115] Samples were then grouped by compartment and replicate and converted into.cit files using GRAND-SLAM v2.0.5d.[116] 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 S1H). 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,[116] 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:

$P_{BG}(k,n) = \pi_E\, Binom(k,n,p_{E1}) + (1\text{-}\pi_E)Binom(k,n,p_{E2})$. 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}.[116] 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 ($(1-\pi_E) \sim$ 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. It remains a possibility that our two-component binomial mixture modeling of unlabeled data could be simplified to a binomial model when using a 3 nucleotide alignment strategy,[134] as proposed in reports that came out during the review process of this study.[135,136] 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 ${\pi_C}^{LO}$ <= 2% and a p_{C_LO} that was logically inconsistent as a lower bound: either because it was 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_{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: $P^g_X(\pi) = Beta(\pi;\alpha^g_X,\beta^g_X)$, a beta distribution characterized with parameters α^g_X and β^g_X. The final gene-specific posterior $P^g(\pi)$ is then as follows:

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

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

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

Kinetic modeling of RNA flow

The kinetic model (Figure 1E), defined by a system of coupled ordinary differential equations (ODEs, Figure S2A; Ietswaart[110]), 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 S2A), 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 (Figures S1B and S1C). To solve these ODEs analytically,[137] 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 on Zenodo.[110]

As described in section "quantification of newly synthesized RNA from subcellular TimeLapse-seq data" (Figure S1I), the gene-specific (g) fraction of new RNA (π) Posterior probability density function, $P^g_{X,t}(\pi)$, was experimentally estimated for 4 different 4sU-pulse times (t = 15, 30, 60 and 120 minutes) for each compartment *X*, through GRAND-SLAM's Bayesian inference framework.[116] To analytically derive the (in some cases multivariate) posterior on the RNA flow rates,[137] 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 S2A).

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 S2A). The mean rate (expression in Figure S2A) 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

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

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 S2A). For example, P_{CY} depends on $\overrightarrow{k} = [k_N, k_{CY}]$, 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 S2A). In absence of nuclear RNA degradation, the whole-cell half-life thus equals the sum of nuclear and cytoplasmic half-lives (Figures S3K and S3L). In presence of nuclear degradation (Table S2, see the next section for the PUND identification procedure), this simple relation between half-lives no longer holds (Figures S3K and S3L).

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 (Figures S3M and S3N), 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 S3F), we used the above equation's probabilistic analog[137] to determine the export posterior by numerical integration (as described above) over the convolution of the nuclear and chromatin posteriors: $P_E^g(k) = \int_{prior\ domain} P_N^g\left(\frac{kz}{k+z}\right) P_{CH}^g(z)dz$.

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 (Figures 1E and S2A).[110] 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,[110] resulted in fraction of new RNA solutions that were multivariate in $\overrightarrow{k} = [k_R, k_E, k_{ND}, k_{CY}]$. 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 S2A). 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 these are not reported. 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[116] 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 section "quantification of newly synthesized RNA from subcellular TimeLapse-seq data"; 3) Our kinetic model through a system of ODEs (Figure S2A). 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 $\Lambda_X(\overrightarrow{k}, t)$ to the MAP(π) timecourse values (scipy.optimize.least_squares, arguments: bounds=[$10^{-6}, \infty$], 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 $\widehat{k_N}$ was inserted to enable fitting of $\widehat{k_{CY}}$. 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 S2E). Because of the same drawback, the LSE, but not the Baysian model, also predicted a number of biologically unlikely fast rate values (Figure S2E). Besides these differences, we generally observed a strong correspondence between our Bayesian MAP and LSE rate estimates (Figure S2E). 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 S2E), 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 (Figures 2C, S4J, and S7I) 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,[54] i.e. the ratio of likelihoods of the kinetic model that includes a nuclear degradation rate (alternative hypothesis M_1) over the simpler "nuclear residence" model with no nuclear degradation (null hypothesis M_0): $K = \frac{P(D|M_1)}{P(D|M_0)}$. D indicates the data, in this case subcellular Timelapse-seq used to distinguish the two models: the timeseries of chromatin, nuclear, cytoplasmic and whole-cell fraction of new RNA posteriors (Figure S1I), as described in the above sections.

$$P^g(D|M_0) = \frac{1}{(k_{hi} - k_{lo})^3} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \prod_{i=1}^{4} P^g_{CH,t_i}(\pi = \Lambda_{CH}(k_R, t_i)) P^g_{N,t_i}(\pi = \Lambda_{N_E}(k_R, k_E, t_i)) \times$$

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

And the equivalent for the nuclear degradation model:

$$P^g(D|M_1) = \frac{1}{(k_{hi} - k_{lo})^4} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \int_{k_{lo}}^{k_{hi}} \prod_{i=1}^{4} P^g_{CH,t_i}(\pi = \Lambda_{CH}(k_R + k_{ND}, t_i)) P^g_{N,t_i}(\pi = \Lambda_{N_E}(k_R, k_E, k_{ND}, t_i)) \times$$

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

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 $\Lambda^3_T = 1 - \Psi k_{CY}(k_N - k_{CY} - k^*)exp(-k_N t) - \Psi k^* k_N exp(-k_{CY} t)$, 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.[54] In our case, genes with transcripts predicted to undergo nuclear RNA degradation (PUNDs) are defined as having $K > 100$ for both biological replicates (Table S2). Bayes factors (both 3 and 4 compartments) and model likelihoods for all genes and replicates are included in Table S1.

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

CellPress

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: $f_{ND}^{cell} = \frac{\sum_{PUND\ genes\ g} k_p^g f_{ND}^g}{\sum_{all\ genes\ g} k_p^g}$. k_p^g indicates the gene-specific steady state mRNA production rate (unit: RPKM min^{-1}), as estimated from the chromatin compartment RNA levels (units: RPKM) and our chromatin turnover rates: $k_p^g = RPKM_{CH}^g \times k_{CH}^g$. We also confirmed our conclusions were consistent when using RNA production rate estimates derived from our whole-cell TimeLapse-seq data. Notably, whole-cell estimates suffer from the drawback that whole-cell turnover rates are a convolution of nuclear and cytoplasmic degradation rates in the case of PUNDs, which then underestimate the resulting production rate estimates. Using the chromatin compartment data resolves this bias.

Hierarchical clustering of genes by RNA flow rates

Hierarchical gene clustering (scipy.hierarchy.linkage, arguments: metric ='seuclidean', method='complete', optimal_ordering= False) was performed on log-transformed half-lives, i.e. $ln(t_{1/2}) = ln\left(\frac{ln(2)}{k}\right)$, with k the MAP and 95% credible interval (CI) endpoints of the turnover rate posteriors, for two biological replicates and all the subcellular compartments: chromatin, nucleus, cytoplasm, polysome and whole-cell. Genes with missing half-live values were excluded. Other RNA flow rates, i.e. 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 S2D). 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).[138] For each cluster, the most enriched GO term, i.e. with highest odds ratio (Figure 2D), and all significant GO terms are listed (Table S2). 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 (Figures 2A and S4A; Table S2).

Gene set enrichment analysis (GSEA)[56] was performed with MSigDB (v7.5.1) gene sets h.all, c5.all, c3.all and c2.all,[57] and parameter settings as in Ietswaart et al.[139] 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 (Table S2).

ShinyGO 0.76.2 (http://bioinformatics.sdstate.edu/go/)[117] 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 Van Nostrand et al.[58] (eCLIP peaks previously identified according to Van Nostrand et al.[140]) 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 rank-sum 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 (Table S3). 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.

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

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)[113] with parameters -ax splice -uf -k14. Poly(A) tail lengths were estimated using nanopolish v0.13.3.[73] 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[118] 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.[121] 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 Drexler et al.[7,141] Briefly, ''Poly(A)'' sites are defined as regions within 50 nucleotides of the end coordinate of annotated protein-coding genes or RNA-PET annotations from cytoplasm and chromatin fractions in K562 ENCODE data (ENCODE Project Consortium, 2012). Determining the splicing status of introns and reads was performed as described in Drexler et al.[7,141] 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 S7K). Given the large number of input features (70145, Table S6), LASSO regression was chosen as a model because it is a linear model, $y = \beta X$, with L1 regularization, which ensures sparse feature selection[142]: $Loss = \frac{1}{2N}\|y - \beta X\|_2^2 + \alpha\|\beta\|_1$ (Python, package scikit-learn v1.0.1, function linear_model.Lasso, arguments: fit_intercept=False, random_state=42, selection='random', max_iter and alpha as specified below). RNA flow rates with genome-wide coverage, i.e. chromatin, nuclear, cytoplasmic, untranslated cytoplasm, and whole-cell turnover and nuclear export, were log-transformed, followed by Z-score normalization, resulting in the model dependent variable (y). Rates from different genes constitute independent data points used for model training (N). The input features originated from various sources and were grouped into classes based on their biological type (Table S6). 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 (Figure 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 S7K): (1) Using the features from an individual class, we performed 10x cross validation (CV) twice, once on the training rates from each biological replicate, for a range of values of hyperparameter α: [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}] if the number of features < 1000, or else: [10^{-3}, 10^{-2}, 10^{-1}], and max_iter_ = 2e4. (2) The optimal (round 1) α was then identified, such that the 10x CV R^2 distribution, joined over both replicate runs, was significantly larger than zero in a one-way t-test ($p < 0.05$, scipy.stats.ttest_1samp, arguments: popmean=0, nan_policy='omit', alternative='greater') and larger than any previously selected α in a two-way t-test ($p < 0.05$, scipy.stats.ttest_rel, arguments: nan_policy='omit', alternative='greater'), similar to a selection approach by Agarwal and Kelley.[86] 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}, $3.3x10^{-4}$, $6.6x10^{-4}$, 10^{-3}, $3.3x10^{-3}$, $6.6x10^{-3}$, 10^{-2}, $3.3x10^{-2}$, $6.6x10^{-2}$, 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% (Figures 7B and S7L), 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 (Figures 7C, 7D, S7M, and S7N; Table S6). (6) Lastly, we tested the trained round 2 lasso model by determining its performance on the 10% unseen test set (Figures 7B and S7L).

For the ''consensus'' whole-cell half-lives,[86] 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 S7L).

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

Continuous averaging plots (Figures 7E and S7O), were generated as in,[34] with minor modifications. First, genes were ranked $g = 1..N$ according to their gene length from short to long, where N is the total number of genes. This was followed by calculation of the "continuous" averages, $<L_k>$ with $k = 1..(2N-1)$, over these ranked gene subpopulations of their gene length L, i.e. the independent variable: $<L_k> := \frac{1}{k}\sum_{g=1}^{k} L_g$ and $<L_{N+k}> := \frac{1}{N-k}\sum_{g=1+k}^{N} L_g$. Next, the corresponding continuous median, and Q25 and Q75 (shaded error bands) of the 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 (Table S6).[143] The statistical analysis was performed as for the RBP targets as described above in "RNA-binding protein associations with RNA flow rates."

Ribosome profiling data analysis

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).

Raw sequencing reads were converted from.fastq to.fasta using fastx-toolkit (v0.0.13), adapter sequences were trimmed using cutadapt (v3.4),[111] and reads were collapsed by UMI using fastx-toolkit. Bowtie2 (v2.4.1)[119] 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[112] (v2.7.5a, GrCH38, Gencode v25) was used to align reads. RSubread[120] was used to generate counts and the DESeq2[121] and Riborex[122] packages were used for differential expression and differential translation analysis.

Subcellular RNA sequencing of CRISPRi lines data 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.

Nuclear wash RNA sequencing data analysis

Raw sequencing.bcl files were converted to.fastq with bcl2fastq (v2.20.0.422, default parameters). Illumina adapter sequences were trimmed and filtered for quality using cutadapt (v2.5, run 1 parameters: -e.2 -m 20 -u 3; run 2 parameters: -q 20,0 –nextseq-trim=20 -m 20). From the yeast spike-in sequences (Table S7) we generated custom fasta and gtf files and concatenated these to the human fasta (hg38) and gtf files (GRCh38.86). Reads were then aligned with STAR (v2.7.0a; default alignment parameters) to this combined human/yeast spike-in genome. Gene counts were determined using featureCounts (subread v1.6.2, parameters: -g gene_id -t gene -s 1 -d 20). Differential expression was performed with DEseq2 as previously described with minor modifications.[139,144] Yeast spike-in counts were used to determine the DEseq sizeFactors of all samples. Differential expression (estimateDispersions, fitType='local'; and nbinomWaldTest, betaPrior=FALSE, design = ~condition) was then performed on all samples with conditions corresponding to the assayed compartments: whole cell, cytoplasm, nuclear, nuclear wash, and nuclear wash without rRNA depletion. The resulting DEseq2 statistics (Log2FoldChange and FDR adjusted p-value) using contrasts nuclear wash versus nuclear, nuclear wash versus cytoplasm, and cytoplasm versus nuclear were then used to compare RNA expression between these compartments (Figures S1E, S1F, and S3B).

SeqFISH+ comparison analysis

SeqFISH+ spot location and DAPI signal field of view.tif images of mouse NIH-3T3 cells as published in Eng et al.[55] were loaded into Matlab (v9.14.0.2306882, function: tiffreadVolume) and analyzed as follows. RNA spots at locations where DAPI signal was above a threshold value of 380 were classified as nuclear, and cytoplasmic otherwise. Given a gene, for all field of view images, all nuclear and cytoplasmic spots corresponding to that gene's transcript were summed over to generate nuclear and cytoplasmic RNA counts over the whole population of cells (202 cells, after 74 cells were removed that had no detectable RNAs). The ratio of these values is then the seqFISH+ nuclear to cytoplasmic RNA ratio (NCR). This procedure is repeated for each gene.

To compare the seqFISH+ NCR values to their RNA flow model equivalents (Figure S3A), we first calculated the model-predicted steady state (ss) levels of nuclear and cytoplasmic RNA, respectively, based on the most granular model that includes RNA production, chromatin release, nuclear degradation (for PUNDs only, set to zero otherwise), nuclear export, and cytoplasmic degradation rates: $RNA^{ss}_{Nuc} = RNA^{ss}_{Chromatin} + RNA^{ss}_{Nucleoplasm} = \frac{k_P(k_R+k_E+k_{ND})}{(k_R+k_{ND})(k_E+k_{ND})}$ and $RNA^{ss}_{Cyto} = \frac{k_P k_R k_E}{k_{CY}(k_R+k_{ND})(k_E+k_{ND})}$. The RNA flow model NCR then equates to: $RNA\ flow\ NCR = \frac{RNA^{ss}_{Nuc}}{RNA^{ss}_{Cyto}} = \frac{k_{CY}(k_R+k_E+k_{ND})}{k_R k_E}$.

Transcript isoform turnover analysis in nuclear exosome mutants

To analyzed the turnover in the nuclear exosome mutants, GRAND-SLAM v2.0.5d was used as described above in section "quantification of newly synthesized RNA from subcellular TimeLapse-seq data" with the following modifications. By default, it calculates

Please cite this article in press as: Ietswaart et al., Genome-wide quantification of RNA flow across subcellular compartments reveals determinants of the mammalian transcript life cycle, Molecular Cell (2024), https://doi.org/10.1016/j.molcel.2024.06.008

the fraction of new RNA from read alignments that fully overlap with feature types listed as "exon" as in the input gtf. In the gtf, other annotated transcript isoforms, such as RI isoforms, are also listed with feature "exon", but contain an additional transcript_biotype field to specify them. Effectively, GRAND-SLAM does not distinguish between protein coding exons or other transcript isoforms. It also does not consider reads that (partly) overlap with introns. This default setting was used for all our analyses, except for the transcript isoform analysis (Figures 6D and S7I). For PC isoforms, a custom gtf was generated from the SNP-masked (default) gtf by filtering for the string 'transcript_biotype "protein_coding"' using awk. For RI isoforms, the default gtf was filtered for 'transcript_biotype "retained_intron"'. For RIs, the RI isoform's gtf were filtered further by excluding all "exon" features that were also present in any isoform without annotation as transcript_biotype "retained_intron". For OIs and UJs, a custom gtf was generated where the feature type "gene" was converted to "exon", to ensure consideration of OI and UJ read alignments by GRAND-SLAM. In these cases we additionally filtered the input.bam files as follows. For OIs, first a sorted introns.bed file was generated containing exclusively intronic regions that have no overlap with any default gtf "exon" feature (bedtools complement -i). Then alignments overlapping completely with introns (bedtools v2.27.1, intersect -a -b introns.bed -f 1.0 -wa) followed by exclusion of non-primary alignments (samtools v1.9, view -F 0x8) were selected as OIs. For UJs, an exons.gtf, i.e. the default gtf filtered for "exon" features only (awk), and introns.gtf (R v4.1.3, package: GenomicFeatures, function intronicParts, linked.to.single.gene.only=TRUE) were generated. To filter for the UJ reads, first unspliced reads, i.e. without an N present in CIGAR string were selected (awk), followed by filtering for not fully within any exon and not fully within any intron (bedtools intersect -b exons.gtf -f 1.0 -v -wa | bedtools intersect -b introns.gtf -f 1.0 -v -wa). The resulting OI and UJ alignment.bam files were then filtered for primary alignments (samtools view -F 0x8), sorted (samtools sort) and indexed (samtools index). Each custom gtf was then used together with the SNP-masked.fasta to generate a.oml index (gedi -e IndexGenome -nobowtie -nokallisto -nostar -p -D), which served as input to GRAND-SLAM. P_C estimates were obtained as follows, because the global 4sU incorporation was too low for our conventional method to estimate reliable p_C values. The ratio of pc in scramble to pc value in a mutant was fitted by assuming no global changes in exons turnover, i.e. by setting the default GRAND-SLAM output median fraction new RNA equal in scramble and control. The absolute scramble p_C values were fitted through the assumption that median fraction new RNA over all expressed genes in scramble matched the WT K562 values 0.4, i.e. no global exon turnover differences. No distinction between top and bottom p_C's was needed as 60min 4sU pulses in nuclear and total RNA had generated similar top/bottom p_C for K562 WT samples. GRAND-SLAM was then run twice as described above: first a prerun (gedi -e Slam -allGenes -full -mode All -snpConv 0.2 -strandness Sense -trim5p 10 -trim3p 5 -progress -plot -D, and in the presence of an unlabeled sample, -no4sUpattern UL). The GRAND-SLAM generated resulting pc estimates in *rates.tsv was then replaced with the internal estimated pc values, followed by the second run (same gedi command without -no4sUpattern UL). For all cases in the isoform analysis, a.bamlist containing a list input.bam files was used as -reads input. The output mean fraction new RNA values were then used to compare the fraction new differences between exosome mutants and scrambled controls (Figures 2B, 2C, 6D, S4H–S4J, and S7I).