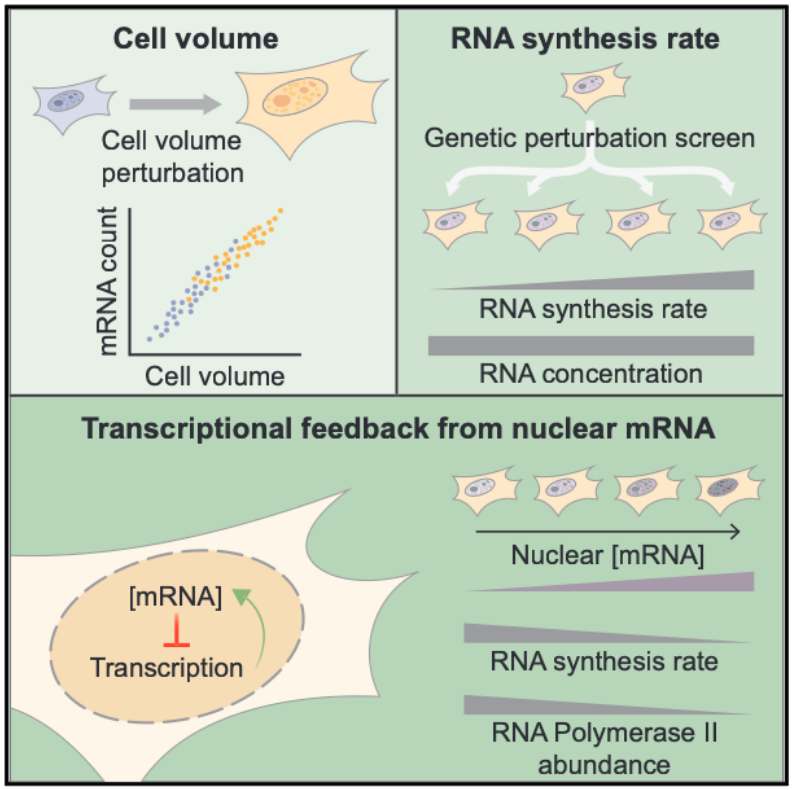
Article

Cell Systems

Feedback from nuclear RNA on transcription promotes robust RNA concentration homeostasis in human cells

Graphical abstract



Highlights

- mRNA concentration homeostasis is robust to cell size perturbation in human cells
- Image-based genome-wide screen reveals factors controlling RNA synthesis rates
- mRNA concentration homeostasis is maintained despite altered RNA synthesis rates
- Nuclear mRNA abundance negatively regulates RNA polymerase activity and abundance

Authors

Scott Berry, Micha Müller, Arpan Rai, Lucas Pelkmans

Correspondence

scott.berry@unsw.edu.au (S.B.), lucas.pelkmans@mls.uzh.ch (L.P.)

In brief

Berry et al. use RNA metabolic labeling and counting of mRNA transcripts in single human cells to reveal the robustness of mRNA concentration homeostasis to perturbation of both RNA synthesis rates and cell size. Using genetic screening and highly multiplexed immunofluorescence, they quantify changes in the RNA production and processing machinery when RNA synthesis rates are altered. This leads to a model in which nuclear mRNA negatively feeds back on the activity and abundance of the mRNA production machinery.

Berry et al., 2022, Cell Systems *13*, 454–470
June 15, 2022 © 2022 Elsevier Inc.
https://doi.org/10.1016/j.cels.2022.04.005

CellPress

CellPress

Cell Systems

Article

Feedback from nuclear RNA on transcription promotes robust RNA concentration homeostasis in human cells

Scott Berry,[1,2,*] Micha Müller,[1,3] Arpan Rai,[1] and Lucas Pelkmans[1,4,*]

[1]Department of Molecular Life Sciences, University of Zurich, Zurich, Switzerland

[2]EMBL Australia Node in Single Molecule Science, School of Medical Sciences, University of New South Wales, Sydney, NSW, Australia

[3]Present address: Oncode Institute, Hubrecht Institute-KNAW and University Medical Center Utrecht, 3584 CT Utrecht, the Netherlands

[4]Lead contact

*Correspondence: scott.berry@unsw.edu.au (S.B.), lucas.pelkmans@mls.uzh.ch (L.P.)

https://doi.org/10.1016/j.cels.2022.04.005

SUMMARY

RNA concentration homeostasis involves coordinating RNA abundance and synthesis rates with cell size. Here, we study this in human cells by combining genome-wide perturbations with quantitative single-cell measurements. Despite relative ease in perturbing RNA synthesis, we find that RNA concentrations generally remain highly constant. Perturbations that would be expected to increase nuclear mRNA levels, including those targeting nuclear mRNA degradation or export, result in downregulation of RNA synthesis. This is associated with reduced abundance of transcription-associated proteins and protein states that are normally coordinated with RNA production in single cells, including RNA polymerase II (RNA Pol II) itself. Acute perturbations, elevation of nuclear mRNA levels, and mathematical modeling indicate that mammalian cells achieve robust mRNA concentration homeostasis by the mRNA-based negative feedback on transcriptional activity in the nucleus. This ultimately acts to coordinate RNA Pol II abundance with nuclear mRNA degradation and export rates and may underpin the scaling of mRNA abundance with cell size.

INTRODUCTION

As the template for protein production, mRNA concentration is fundamental to cell physiology. While the number of transcripts of a given gene per cell is variable between cells in a population (Raj et al., 2008), much of this variability can be explained by cell size variation—with larger cells containing more transcripts (Battich et al., 2015; Ietswaart et al., 2017; Kempe et al., 2015; Padovan-Merhar et al., 2015; Sun et al., 2020; Zhurinsky et al., 2010). In fission yeast (Sun et al., 2020; Zhurinsky et al., 2010), plants (Ietswaart et al., 2017), and mammalian cells (Padovan-Merhar et al., 2015; Schmidt and Schibler, 1995), size scaling of mRNA "abundance" is achieved by size scaling of mRNA "production rates." The underlying mechanism is unclear; however, the predominant hypothesis is that cells contain a factor that is limiting for transcription whose abundance is closely coupled to cell size—leading to size scaling of transcript production from a fixed amount of DNA template (Marguerat and Bähler, 2012). In yeast, it has been suggested that RNA polymerase II (RNA Pol II) itself is a limiting factor for transcription (Sun et al., 2020; Swaffer et al., 2021); however, it remains unclear how levels of RNA Pol II are determined and maintained.

Studies in budding yeast have further shown that mRNA concentration homeostasis involves coupling of mRNA production and degradation rates, so that perturbations to one are compensated by changes to the other (referred to as "mRNA buffering") (Haimovich et al., 2013; Sun et al., 2012, 2013). However, these studies do not consider the effects of cell size changes, which often occur upon disruption of global RNA metabolism (Jorgensen et al., 2002; Maitra et al., 2018; Mena et al., 2017). Moreover, unlike in human cells and fission yeast, budding yeast increases RNA stability rather than RNA production rate with increasing cell size (Mena et al., 2017), indicating a different size-scaling mechanism. In mammalian cells, mRNA stabilization upon transcriptional inhibition has also been reported (Helenius et al., 2011; Slobodin et al., 2020), but both acceleration (Abernathy et al., 2015; Gilbertson et al., 2018) and disruption (Lee et al., 2012; Singh et al., 2019) of cytoplasmic RNA degradation have been associated with transcriptional repression. Therefore, it remains unclear to what extent RNA concentrations are robust to perturbation in mammalian cells (Hartenian and Glaunsinger, 2019).

Since both size scaling and buffering contribute to RNA concentration homeostasis, they may be part of the same mechanism, but whether this is the case and how this works is unknown. To study these phenomena together, we here combine genome-wide genetic perturbation screening with multiplexed quantitative measurements of single cells. Crucially, this allows us to account for perturbation-induced changes to cell size or cell-cycle stage at the single-cell level and also enables us to connect effects seen in perturbations with naturally varying

properties of unperturbed cells. We uncover hundreds of perturbations that affect global RNA synthesis rates in single cells, including pathways not previously implicated in transcriptional control. However, in most cases, we find that RNA concentration is not disrupted. Systems-level analysis of the genes involved together with detailed characterization of molecular phenotypes at the single-cell level suggest a model in which transcription rates are negatively regulated by nuclear mRNA concentration. We propose that the activity, and ultimately abundance, of RNA Pol II is determined by this mRNA-based feedback to enable robust mRNA concentration homeostasis in human cells.

RESULTS

Cell size perturbation leads to precise adaptation of mRNA abundance

In unperturbed human cells, mRNA abundance scales with cell size (Figure S1A) (Battich et al., 2015; Kempe et al., 2015; Padovan-Merhar et al., 2015). To investigate if mRNA concentrations are robustly maintained upon cell volume perturbation, we established a method of cell volume measurement compatible with high-throughput branched DNA (bDNA) single-molecule RNA fluorescence *in situ* hybridization (smFISH) (Figures S1B–S1H; STAR Methods) and measured cytoplasmic transcript abundance for 14 size-scaling genes (Battich et al., 2013) in populations of HeLa cells. Cell size was perturbed using siRNA-mediated knockdown of GRIP2 and SBF2 (Figure S1I), genes which are not known to affect transcription but whose knockdowns result in smaller and larger cells, respectively (Berchtold et al., 2018) ($\overline{V}_{GRIP2}$ = 1.0 pL, $\overline{V}_{Scrambled}$ = 2.3 pL, $\overline{V}_{SBF2}$ = 4.9pL). Average mRNA abundance was dramatically reduced in smaller cells (GRIP2 RNAi) and increased in larger cells (SBF2 RNAi) (Figures 1A, 1B, and S1J). However, mRNA abundance remained proportional to cell volume (Figures S1J and S1K), and volume still accounted for the majority of transcript abundance variation for most genes (Figure 1C). Moreover, mean mRNA abundance in perturbations, $\overline{n}_{Perturbed}$, was well predicted by the change in volume ($\overline{n}_{Perturbed} \approx \overline{n}_{Scrambled}\overline{V}_{Perturbed}/\overline{V}_{Scrambled}$; Figure 1D) and normalizing spot count by volume at the single-cell level often led to overlapping distributions (Figures 1B and S1M). We also measured total RNA abundance using RNA Strandbrite, a fluorescent stain specific for RNA, as well as by total RNA extraction and quantification (Figures S1N–S1P). Overall mRNA abundance was measured by FISH against polyadenylated RNA (poly(A) FISH). These different approaches all consistently showed that changes to cell volume in GRIP2 or SBF2 perturbations were associated with proportional changes in total RNA and mRNA abundance (Figures S1N–S1Q). Cell volume is therefore a dominant source of heterogeneity in mRNA abundance in cell populations, and its perturbation leads to precise adaptation of transcript abundance to maintain both global and gene-specific mRNA concentration.

RNA production rates are coupled to cell cycle and cell volume

RNA abundance is determined by its production and degradation rates. However, inferring gene-specific mRNA production rates in single cells is technically challenging and often requires assumptions (Ietswaart et al., 2017; Padovan-Merhar et al., 2015; Sun et al., 2020). Because size scaling is a transcriptome-wide phenomenon, we measured the rate of bulk RNA production in single cells *in situ* using metabolic pulse labeling with 5-ethynyl uridine (EU) (Jao and Salic, 2008; Padovan-Merhar et al., 2015; Shah et al., 2018) in combination with measurement of cell volume and immunofluorescence to assign cells to G1/S/G2 cell-cycle phases (Figures 1E, 1F, and S2A–S2J). In two cell lines (HeLa and 184A1) and also in primary human keratinocytes, EU incorporation increased close to proportionally with cell volume (Figure S2I), indicating that RNA production rates scale with cell size, in agreement with previous results in fibroblasts (Padovan-Merhar et al., 2015). However, including cell-cycle information revealed that G2- and S-phase cells showed higher EU incorporation than G1 cells of the same volume (Figures 1F and S2J). Furthermore, HeLa cells showed an additional increase during S phase (Pfeiffer and Tolmach, 1968)—above the level seen in G2. While these cell-cycle effects are interesting and suggest differences compared with yeast (Voichek et al., 2016), we here focus on cell size scaling of EU incorporation—using precise cell-cycle information to exclude changes to DNA template abundance.

EU is incorporated into all major RNA species by RNA polymerases I, II, and III (Jao and Salic, 2008). To evaluate the contribution of ribosomal RNA (rRNA) to EU incorporation, we treated cells with the RNA Pol I inhibitor CX-5461 (Drygin et al., 2011), which resulted in elimination of EU incorporation in the nucleolus—the site of rRNA transcription (Figures 1G, S2G, and S2H). To quantify this, we segmented the nucleolus and nucleoplasm (non-nucleolus) and summed EU intensity in each region separately (Figure S2L). CX-5461 reduced nucleolar EU by 89%–92% while the reduction in total nuclear EU was more modest (46%–56%) (Figures 1H and S2M). Moreover, nucleolar segmentation in untreated cells revealed that the nucleolus contributes 45% of nuclear EU incorporation. Taken together, this suggests that rRNA contributes approximately half of the EU incorporation in nascent RNA. This is less than the contribution of rRNA to total RNA abundance (80%; Wolf and Schlessinger, 1977) and is consistent with much greater stability of rRNA than mRNA (estimated average half-life 3–8 days [Gillery et al., 1995; Halle et al., 1997] and 3.5 h, respectively [Herzog et al., 2017; Tani et al., 2012]). Cell volume and cell-cycle dependence of total nuclear EU incorporation was similar between CX-5461-treated and untreated cells (Figures 1G and S2K). Moreover, in untreated cells, EU intensities in the nucleolus and nucleoplasm were highly correlated and showed similar cell volume and cell-cycle dependence (Figures 1I and 1J). Together, these data indicate that the production rates of rRNA and other cellular RNA are similarly coordinated with cell volume and cell-cycle stage.

To investigate if bulk RNA production is coordinated with cellular growth rate, we grew HeLa cells in a low concentration of the translation inhibitor cycloheximide, which reduces cell growth rate and increases cell-cycle duration (Ginzberg et al., 2018) (Figures S2N and S2O). However, we observed very little change in EU incorporation (Figure S2P). To modulate cell volume, we treated HeLa cells with the cyclin-dependent kinase (CDK) inhibitor roscovitine for 48 h (Cadart et al., 2018), which led to an approximate doubling of cell volume in each cell-cycle phase, without affecting the cell-cycle distribution (Figures S2Q and S2R). In this case, EU incorporation increased approximately

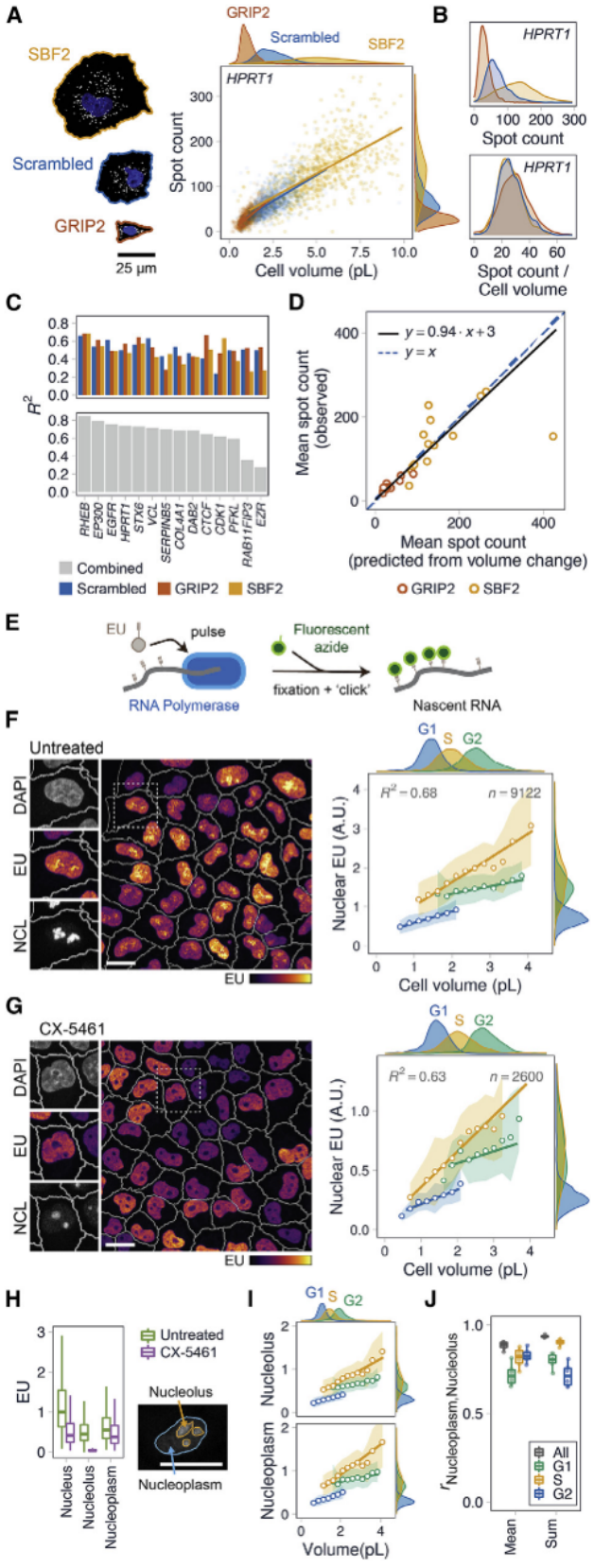


Figure 1. RNA abundance and production rates scale with cell size

(A) Number of cytoplasmic *HPRT1* transcripts detected by smFISH as a function of cell volume, in cells transfected with scrambled siRNA (n = 1,315 cells), siRNA targeting GRIP2 (n = 1,413 cells), or SBF2 (n = 834 cells). DAPI (blue) and *HPRT1* smFISH (gray) for selected example cells.

(B) Single-cell spot count distribution for *HPRT1*, normalized by cell volume in lower panel.

(C) R^2 for linear regression predicting spot count from cell volume, for the genes indicated. Either considering genetic perturbations separately (top) or combined (bottom).

(D) Mean spot count predicted from the change in cell volume in perturbations. Blue dashed line shows the expected fit given by the equation $\bar{n}_{perturbed} = \bar{n}_{scrambled}\bar{V}_{perturbed}/\bar{V}_{scrambled}$, black line shows the actual fit line obtained by robust regression.

(E) Metabolic RNA labeling using EU to measure RNA production rates.

(F) EU incorporation in untreated cells during a 30 min pulse. Nucleolin (NCL) immunofluorescence, EU, and DAPI inset. Quantification of single-cell sum nuclear EU intensity as a function of cell volume, for different cell-cycle stages. Data points show mean of cell volume bins with shaded regions showing ± s.d.. Lines fit to single-cell data using robust regression, plotted for 2nd–98th percentiles.

(G) As in (F), with 2 h CX-5461 pre-treatment.

(H) Sum EU intensity in subnuclear regions in CX-5461-treated (n = 2,600) and untreated cells (n = 9,122). Boxplots summarize single-cell values with outliers omitted for clarity. Values relative to median nuclear EU intensity of untreated cells.

(I) As in (F), for sum nucleolar and sum nucleoplasmic EU.

(J) Correlation of nucleolar and nucleoplasmic EU across cells, for either the sum or the mean of pixel values. Boxplots summarize correlations observed in individual replicate wells (n = 6). All data from HeLa cells. Scale bars, 25 µm. See also Figures S1–S3.

in proportion to cell volume changes. This indicates that RNA production rates in human cells are coordinated with cell volume rather than cellular growth rate, in agreement with studies at the cell-population level in fission yeast (Fraser and Nurse, 1979; Zhurinsky et al., 2010).

Finally, we tested whether scaling of mRNA abundance with cell size is in part also achieved by decreasing mRNA degradation rates in larger cells. To study this, we estimated mRNA degradation rates for 18 genes using smFISH after transcriptional inhibition (Figure S3). In agreement with previous work in human cells (Padovan-Merhar et al., 2015) and fission yeast (Sun et al., 2020), we found that mRNA degradation rates were not systematically lower in larger cells. This further supports a model in which size scaling of mRNA abundance is caused by size dependence of mRNA production rather than degradation.

Genetic perturbation screen reveals large changes in RNA production rates

To investigate the genetic control of RNA production rates and their coordination with cell size, we next applied this RNA metabolic labeling assay in the context of an arrayed genome-wide siRNA perturbation screen (Figure 2A) (Müller et al., 2021). The screen comprises spatially resolved multivariate measurements of ~80 million HeLa cells across 21,823 perturbations, providing a comprehensive resource to perform systems-level analyses of the regulation of RNA production as a function of cell size and cell-cycle stage. Because cellular protein content and nuclear area are both proportional to cell volume (Figure S1G) (Cantwell and Nurse, 2019; Kafri et al., 2013), we used these as measures of cell size in the screen.

Many perturbations led to increases or decreases in EU incorporation (Figure 2B), with changes often of a large magnitude (mean EU fold-change percentiles $P_1 = 0.56$ and $P_{99} = 2.2$)

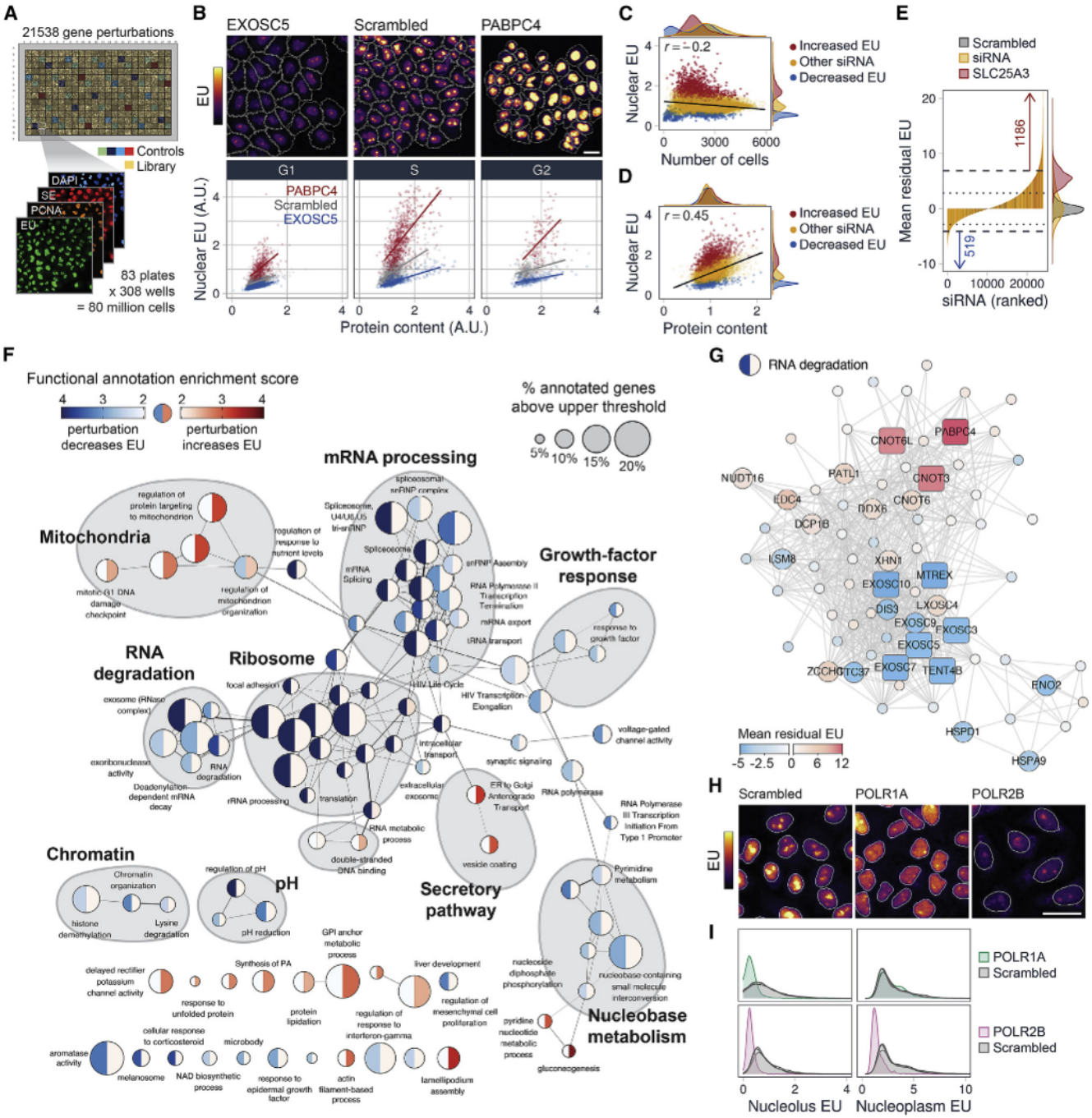


Figure 2. Genetic screen identifies regulators of RNA production

(A) Arrayed genome-wide siRNA screen.

(B) Upper: example images of EU metabolic labeling from the screen. Lower: sum nuclear EU intensity as a function of cellular protein content, for each cell-cycle phase.

(C) Perturbation-averaged sum nuclear EU as a function of number of interphase cells in the screen.

(D) Perturbation-averaged sum nuclear EU as a function of cellular protein content (G1 cells only).

(E) Mean residual EU across the screen for scrambled siRNA (n = 1,826), SLC25A3 (n = 664), and library siRNAs (n = 21,538). Gene perturbations ranked by mean residual EU. Dotted and dashed lines indicate lower and upper hit thresholds ($p_{posterior}$ = 0.5, 0.85), respectively (Müller et al., 2021).

(F) Network of functional annotations enriched in perturbations with increased or decreased mean residual EU. Functional annotation enrichment score defined in STAR Methods.

(legend continued on next page)

(Figure S4A). While some perturbations were associated with reduced viability, the general trend was weak (Figure 2C), in agreement with our finding (Figures S2O and S2P) that perturbing growth rates does not lead to characteristic changes in RNA production. In contrast, perturbation-averaged EU incorporation did correlate with cellular protein content across the screen (Figure 2D), showing that, across a genome-wide set of genetic perturbations, altered cell size typically leads to a coordinated change in RNA production. Furthermore, this was observed at the single-cell level within all perturbations (Figures S4E–S4G). To systematically identify perturbations of EU incorporation, we therefore derived a measure of EU intensity that is corrected for cell size and cell-cycle stage at the single-cell level (STAR Methods). We refer to this as "mean residual EU." It is negative for perturbations with reduced EU incorporation (n = 519 hits, 413 of which have >500 cells) and positive for those with increased EU incorporation (n = 1,186 hits, 1,183 of which have >500 cells) (Figure 2E). Mean residual EU hits are not associated with characteristic cell-cycle distribution or cell size changes (Figure S4C).

To reveal the functions of genes whose perturbation underlies altered EU incorporation, we performed rank-based enrichment analysis of functional annotations using mean residual EU (STAR Methods). Enrichment scores were higher for "down" hits than "up" hits, suggesting that reduction in EU incorporation occurs through disruption of a more focused (or better annotated) set of pathways. As expected, ribosomal biogenesis (RNA Pol I), as well as RNA-Pol-II- and Pol-III-dependent transcription were strongly associated with reduced EU incorporation (Figure 2F). However, nuclear RNA processing and splicing, chromatin organization, nuclear RNA export, and RNA degradation annotations were also strongly enriched for reduced EU incorporation. Closer examination of the genes underlying these annotations (Figures 2G, S4H, and S5) revealed, for example, that enrichment of the term "RNA degradation" for reduced EU incorporation was driven mostly by the "nuclear" rather than "cytoplasmic" RNA degradation factors (Grudzien-Nogalska and Kiledjian, 2017; Łabno et al., 2016; Schmid and Jensen, 2018; Siwaszek et al., 2014) (Figure S6A), especially the RNA exosome (a multi-component 3′ to 5′ ribonuclease; Schmid and Jensen, 2018), and nuclear exosome targeting factors such as MTREX, ZFC3H1, and TENT4B. Annotations enriched for increased EU incorporation include the secretory pathway—specifically vesicle coating and ER to Golgi anterograde transport (e.g., RAB1A, SEC16B, and TRAPPC1), and also protein lipidation, especially genes involved in GPI-anchor biosynthesis (e.g., PIGP and DPM3) (Figures S4H and S5). These previously unreported phenotypes suggest intriguing links between cell surface homeostasis and transcriptional regulation. EU incorporation increases were also observed upon perturbation of many sequence-specific DNA-binding proteins, for example, NFX1/NFXL1 and ELK1, and also histone demethylases such as KDM2B (H3K4/K36-demethylase) (Figures S4H and S5), indicating that individual transcription factors and histone modifiers can also have strong repressive effects on overall RNA synthesis.

Nucleolar and nucleoplasmic RNA production are highly coordinated in single cells (see Figure 1I). Across the genome-wide screen, we also observed a high correlation between perturbation-averaged mean EU intensities in the nucleolus and nucleoplasm (r = 0.97), with sum EU intensities slightly less well correlated (r = 0.86) (Figure S6B), due to changes in nucleolus size (Boulon et al., 2010). This suggests that perturbations typically have similar effects on ribosomal and non-ribosomal transcription. Despite this general correspondence, however, we did identify a subset of hits which specifically reduced nucleolar but not nucleoplasmic EU, many of which were related to RNA-Pol-I-dependent transcription (Figures S6C–S6J; STAR Methods). At face value, this may be unsurprising; however, we did not observe a converse nucleoplasm-specific EU reduction when targeting genes associated with RNA Pol II transcription (Figure S6K). For example, POLR1A RNAi specifically affected nucleolar EU while POLR2B RNAi affected both the nucleolus and the nucleoplasm (Figures 2H and 2I). This is consistent the role of RNA Pol II in promoting RNA-Pol-I-dependent transcription (Abraham et al., 2020; Burger et al., 2013; Caudron-Herger et al., 2016) and suggests that size scaling of RNA-Pol-I-dependent transcription may occur as a consequence of size scaling of RNA-Pol-II-dependent transcription.

RNA concentration is stable in perturbations with altered RNA production

To determine how RNA abundance is affected in conditions in which synthesis rates are perturbed, we measured mRNA abundance using poly(A) FISH and total RNA abundance using RNA Strandbrite (Figures 3A, S7A, and S7B) on a set of 436 gene perturbations. Perturbations were chosen from enriched annotations in the genome-wide screen to maintain both functional and phenotypic diversity (Müller et al., 2021). Both mRNA and total RNA abundance correlated with cellular protein content across these 436 perturbations (Figure S7C) but were much less strongly perturbed than EU intensities (Figures 3B and S7D), implying that RNA concentration homeostasis is not generally disrupted. For example, POLR2B and PABPC4 knockdown resulted in a 5.7-fold reduction or 2.6-fold increase, respectively, in EU incorporation relative to scrambled siRNA controls, but both showed a 1.1- or 1.2-fold reduction in mRNA abundance in the nucleus and cytoplasm, respectively (Figure 3A). Moreover, the variability *within* cell populations was also greater for EU incorporation than for protein, mRNA, and total RNA abundance (Figures 3C and S7E–S7G).

To systematically identify changes in RNA abundance, we first corrected for cell size and cell-cycle changes using linear regression, which led to narrower distributions of perturbation-averaged poly(A) FISH and RNA Strandbrite intensities, and greater overlap with scrambled siRNA controls (Figures S7H and S7I). This indicates that cell size and cell-cycle changes

(G) STRING protein-protein association network of genes with RNA degradation annotation. Conditions with low cell number omitted. Labeled circles for $p_{posterior} > 0.5$ and squares for $p_{posterior} > 0.85$.

(H) Images of EU incorporation in the genome-wide screen for POLR1A or POLR2B knockdown.

(I) Single-cell distributions of sum nucleoplasmic and nucleolar EU intensity (arbitrary units) for perturbations in (H). Three scrambled siRNA wells from the same plates shown for comparison. Scale bars, 25 μm. See also Figures S4–S6.

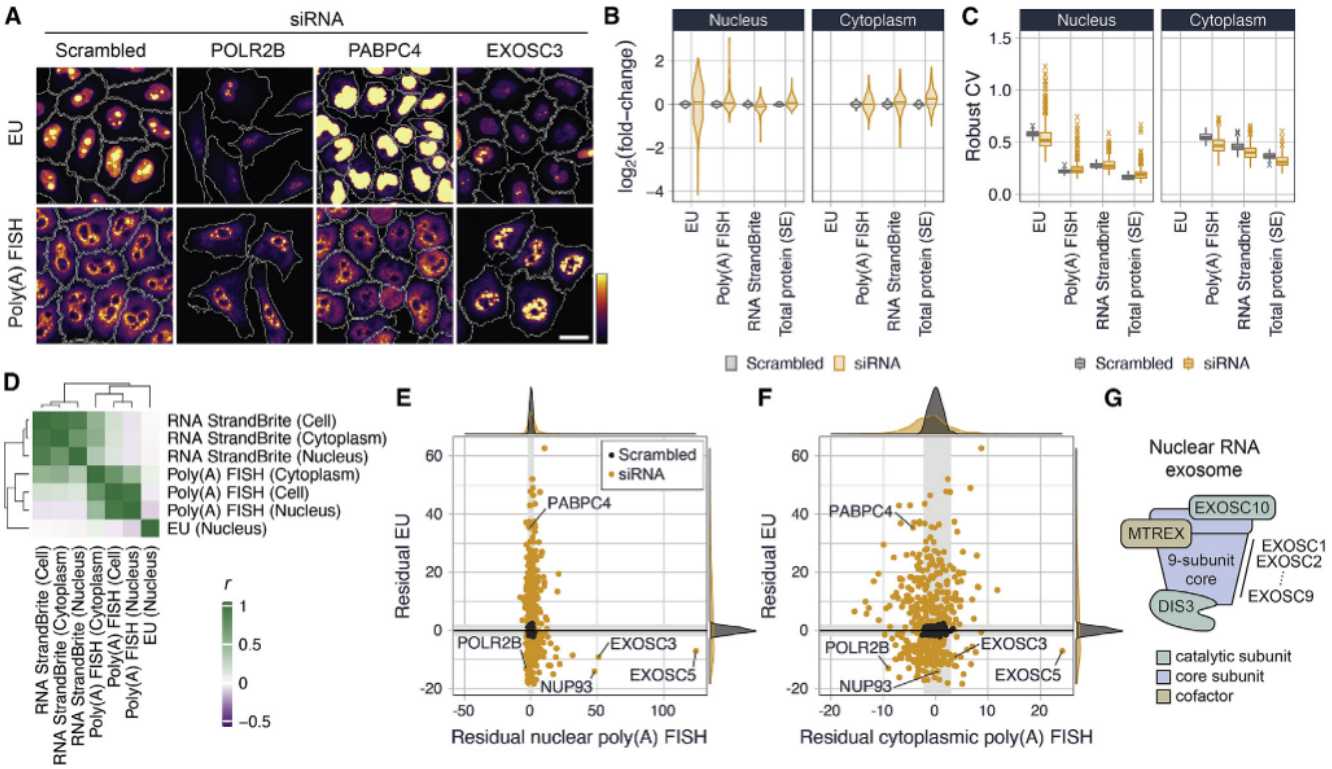


Figure 3. RNA concentration is stable in conditions with altered RNA production

(A) EU metabolic labeling and poly(A) FISH images for selected siRNA perturbations.

(B) Fold-change in perturbation-averaged mean nuclear and cytoplasmic intensities relative to scrambled siRNA controls. Scrambled siRNA (n = 80), library siRNA (n = 415). Low cell number conditions omitted.

(C) Cell-to-cell variability (robust coefficient of variation) of mean nuclear and cytoplasmic intensities within each well, for data in (B).

(D) Pairwise correlations of mean residual RNA abundance measurements and EU across perturbations (n = 379).

(E) Mean residual EU versus mean residual nuclear poly(A) FISH. Mean of two replicates shown for each condition. Gray boxes indicate 1st/99th percentiles of scrambled siRNA controls.

(F) As in (E), for cytoplasmic poly(A) FISH.

(G) Nuclear RNA exosome schematic. See also Figures S7–S9.

explain some of the differences seen in RNA abundance. We then compared these RNA abundance measurements with EU incorporation across perturbations (Figures 3D–3F, S8A, and S8B). Average total RNA and mRNA abundances were moderately correlated with one another across perturbations; however, there was a conspicuous lack of strong correlation of either with EU (Figures 3D–3F and S8B–S8D). This implies that in human cells, RNA degradation rates generally change in concert with production rates, similar to budding yeast (Haimovich et al., 2013; Sun et al., 2012). Notably, however, when we specifically analyzed nuclear mRNA abundance, we did observe a small number of perturbations that show reduced EU incorporation and strong overaccumulation of mRNA in the nucleus (Figures 3E and S8C), which we validated by RNA dot blot (Figures S8E–S8J). This disruption of mRNA concentration homeostasis occurs when targeting nuclear pore component NUP93 or core components of the nuclear RNA exosome: EXOSC3 and EXOSC5 (Fan et al., 2018; Silla et al., 2018), suggesting critical roles for nuclear RNA degradation and export in this process. In addition to nine core subunits (EXOSC1–EXOSC9), the nuclear RNA exosome (Figure 3G) contains two catalytic subunits: DIS3, which targets products of RNA Pol II transcription, and EXOSC10, which plays a role in nucleolar RNA processing (Davidson et al., 2019; Schmid and Jensen, 2018). Although perturbation of both EXOSC10 and DIS3 showed reduced EU incorporation in the genome-wide screen (Figure 2G), their knockdown did not lead to increased nuclear mRNA levels (Figure S8A), likely because of partial functional redundancy between them (Davidson et al., 2019; Fan et al., 2018; Tomecki et al., 2010).

To measure gene-specific cytoplasmic mRNA abundance, we applied bDNA smFISH to detect transcripts for nine genes across 50 genetic perturbations (Figures S9A–S9H). Perturbations included RNA exosome and nuclear pore components, transcription machinery and splicing factors, as well as diverse perturbations with increased EU incorporation. Although genetic perturbations often led to changes in gene-specific transcript abundance, changes were not in a consistent direction for a given perturbation (Figures S9C, S9D, and S9I), and transcript abundance of specific genes was typically not correlated with EU incorporation across perturbations (Figures S9J and S9K). Moreover, mRNA abundance remained highly coordinated with cellular protein content at the single-cell level (Figure S9E), indicating maintenance of gene-specific mRNA concentration.

Together, these data reveal that RNA concentrations are generally robustly maintained in conditions with altered EU incorporation, pointing to a generic coordination of RNA synthesis and degradation in human cells. From the hundreds of conditions studied, we identified only a few exceptions in which mRNA concentration homeostasis was strongly disrupted—specifically in the nucleus. These perturbations involve core components of the nuclear RNA exosome and nuclear pore (EXOSC3, EXOSC5, and NUP93), and all show strong accumulation of nuclear poly(A) FISH signal and reduced EU incorporation. Since perturbation of nuclear mRNA processing and export are also associated with reduced EU incorporation in the genome-wide screen, this points to a mechanism by which cells downregulate RNA synthesis in response to an inability to export or degrade nuclear mRNA.

Characterization of molecular changes underpinning perturbation of RNA production

The diverse set of genetic perturbations identified that affect EU incorporation provide an opportunity to investigate how cells globally regulate bulk RNA production. To determine how the abundance of RNA Pol II and other proteins involved in RNA production and processing are altered when RNA production rates change, we quantified EU incorporation together with the abundance and localization of 18 proteins or post-translational modifications (PTMs) in the same cells, using iterative indirect immunofluorescence imaging (4i) (Figures 4A and 4B) (Gut et al., 2018). The antibody panel includes markers of active promoters (H3K4me3) and heterochromatin (H3K9me3), as well as RNA Pol II (POLR2A, also known as RPB1) and its phosphorylated forms: POLR2A-S5P and POLR2A-S2P, which are markers of transcription initiation (Glover-Cutter et al., 2009) and elongation (Peterlin and Price, 2006), respectively. We also measured nuclear speckles (SON) and speckle-associated proteins involved in splicing (SNRPB2), RNA stability (PABPN1), and export (ALYREF), as well as the nucleolus (NCL) and RNA Pol I (POLR1A).

We selected 63 perturbations from the genome-wide screen, focusing on gene perturbations in which mRNA concentration homeostasis is perturbed (EXOSC3/5 and NUP93) as well as other pathways enriched in the screen, including RNA processing (U2AF1 and HNRNPK), transcription factors (NFXL1 and ELK1), chromatin modification (SETD1A and CBX3), the endomembrane system (RAB1A and TRAPPC8) and GPI-anchor biosynthesis (PIGC and PIGP). After validating 4i-based quantification of protein abundance in single cells (STAR Methods; Figures S10A and S10B), we performed duplicate 4i experiments in which we analyzed an average of 5,500 ± 2,500 cells (mean ± s.d.) per perturbation, together with over 80,000 control cells (Figures S10C–S10G). The abundance of most proteins measured, including POLR1A and POLR2A, also positively correlated with total cellular protein content in unperturbed cells (Figures S10H and S10I).

To obtain an overview of how different proteins and PTM levels change in perturbations, we used hierarchical clustering of perturbation-averaged mean intensities (which reflect concentrations), focusing on G1 cells (Figure 4C). This revealed diverse molecular phenotype profiles, which clustered into groups with characteristic EU incorporation. Notably, perturbation of four of the five RNA exosome components (DIS3, EXOSC3, EXOSC5, and MTREX—but not EXOSC10) clustered together, indicating a common cellular phenotype. This "DIS3 phenotype" was characterized by reduction in RNA Pol II concentration and other components associated with active transcription (e.g., CDK7 and H3K4me3), together with increases in ALYREF and H2B. The DIS3 phenotype differed from other conditions with reduced EU incorporation, such as the disruption of components of pre-mRNA processing machinery (e.g., HNRNPK, U2AF1, and SNRPF), which showed greater reductions in cell viability, and more extreme changes to cell morphology.

To determine which markers show concordant changes across perturbations, we took perturbation-average mean intensity values of each marker and calculated pairwise correlations between them (Figure 4D). Clustering this correlation matrix revealed 4 groups of markers. Foremost, we observed that EU clustered with all three markers of RNA Pol II, revealing that changes in EU incorporation induced by genetic perturbations are most closely related to changes in RNA Pol II concentration (Figure 4E). A second group contained markers associated with "active transcription" such as CDK7, CDK9, and H3K4me3, which also correlated well with EU. A third group containing H2B, H3K9me3, and ALYREF did not show a close relationship with EU and was anti-correlated with PABPC4, which showed distinct changes across perturbations compared with all other markers, even PABPC1 (Figure 4D). Similar to the genome-wide screen (Figure S6B), nucleolar and nucleoplasmic EU incorporation were highly correlated across perturbations (Figures S10J and S10K) and both showed similar correlations with 4i markers to those seen for whole-nucleus EU (Figure S10L). To further investigate the link between RNA Pol II abundance and EU incorporation, we corrected EU and POLR2A intensities for cell-cycle and cell size effects (STAR Methods). Residual EU and residual POLR2A were well correlated across perturbations ($r_{EU,\,POLR2A} = 0.57$; Figures 4F, S11A, and S11B), and significant changes to one were associated with changes to the other (Figures 4H and 4I). To validate this finding, we made use of MRC5 cells in which both copies of POLR2A are tagged with GFP (Steurer et al., 2018). Imaging POLR2A-GFP and EU across the same set of 63 perturbations revealed a similar positive correlation ($r_{EU,\,POLR2A-GFP} = 0.65$; Figures 4G, S11C, and S11D), confirming that RNA Pol II abundance typically changes together with transcriptional activity in perturbations.

To visualize the multidimensional character of the measured phenotypes in single cells, we embedded all 434,575 cells in a 2D uniform manifold approximation and projection (UMAP) (McInnes et al., 2018), using intensity and texture features derived from 4i together with morphology and cell crowding features (Figures 5A and 5B). Perturbations typically localized to specific regions of the UMAP; however, almost all contained a subpopulation that cannot be readily distinguished from scrambled siRNA controls (Figures 5C and S11E), likely corresponding to non-perturbed cells. Although EU was omitted when constructing the UMAP, other features in the data led to a non-random pattern of EU intensity, which is distinct from both cell cycle and nuclear area (Figures 5B and S11F). We observed two major axes of variability in the main body of the UMAP: from bottom to top, nuclei become larger, and cell-cycle

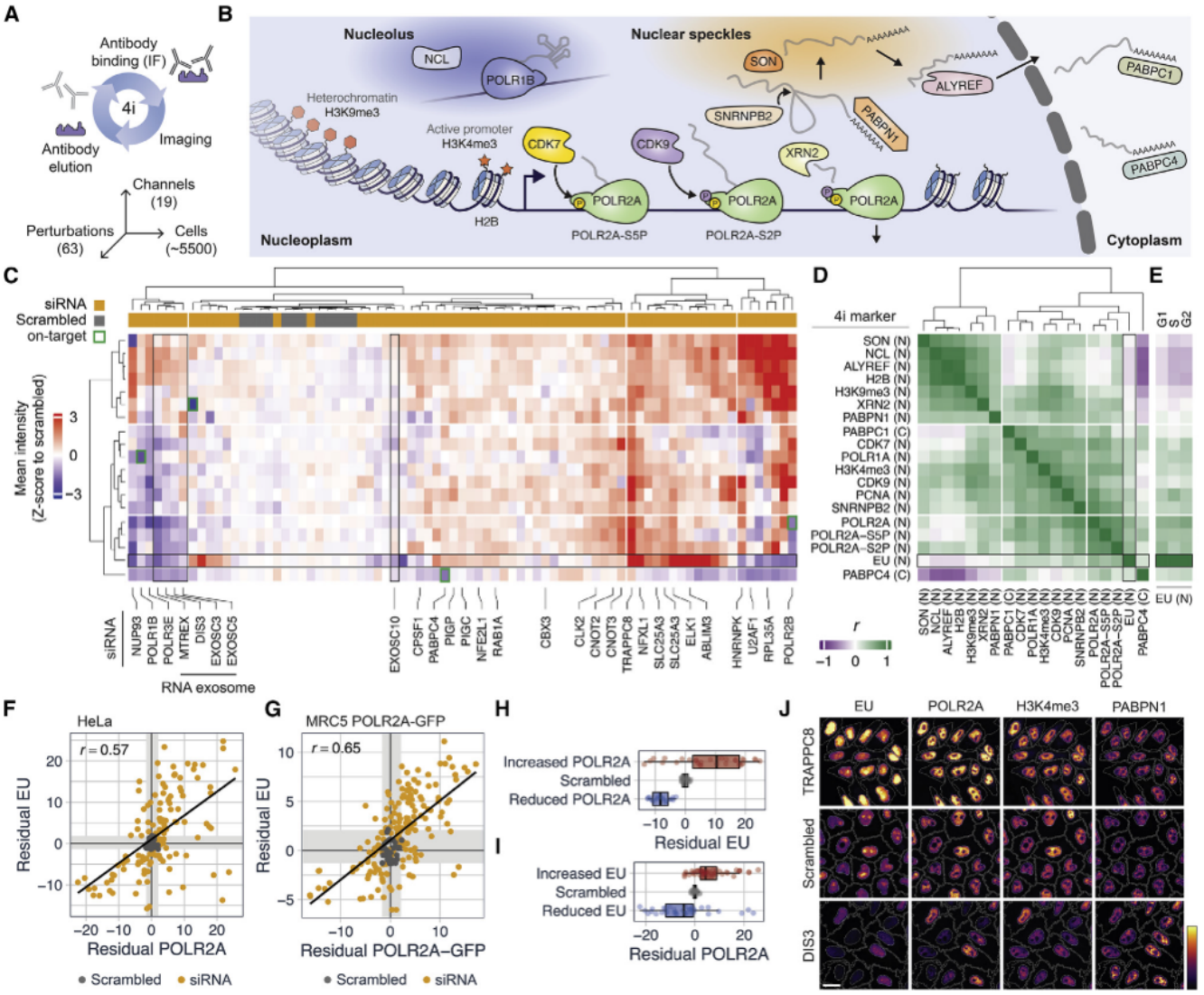


Figure 4. Highly multiplexed profiling of cellular phenotypes associated with altered RNA production

(A) 4i schematic.

(B) Diagram of proteins and PTMs measured by 4i and their roles in RNA metabolism.

(C) Hierarchical clustering of mean intensities of 4i markers across 63 siRNA perturbations. Each marker standardized to non-perturbed (scrambled siRNA) cells using the robust *Z* score before averaging replicates. G1 cells only. On-target indicates that siRNAs directly target a gene in the antibody panel. Row labels shown in (D).

N indicates nuclear mean intensity while C indicates cytoplasmic mean intensity.

(D) Pairwise correlations between perturbation-averaged 4i marker intensities (the rows of C).

(E) Correlations of perturbation-averaged 4i marker intensities with EU by cell-cycle phase.

(F) Residual mean POLR2A versus residual mean EU for each well in the 4i experiment. Pearson's correlation for non-controls inset. Gray shaded boxes represent 1st/99th percentile of scrambled siRNA controls.

(G) As in (F), for MRC5 POLR2A-GFP cells with POLR2A quantified using GFP intensity, across the same 63 perturbations (n = 3 per perturbation).

(H) 4i experiment: residual mean EU for perturbations in which POLR2A is increased/decreased compared with scrambled siRNA controls (Benjamini-Hochberg adjusted $p < 0.05$).

(I) As in (F), with POLR2A and EU reversed.

(J) Example 4i images. See also Figures S10 and S11.

progresses, while from left to right, POLR2A intensity increases together with EU intensity, and other markers of active transcription. Perturbations with increased EU incorporation, such as TRAPPC8 and NFXL1, typically occupy this region with increased intensity of active transcription markers (Figure 5C). As expected, targeting one of the proteins in the antibody panel whose intensities are used to create the UMAP (POLR1A, PABPC4, XRN2, and POLR2B), led to perturbations in which cells localized away from the main group of cells (Figure 5A). Notably, RNA exosome perturbations localized to two distinct

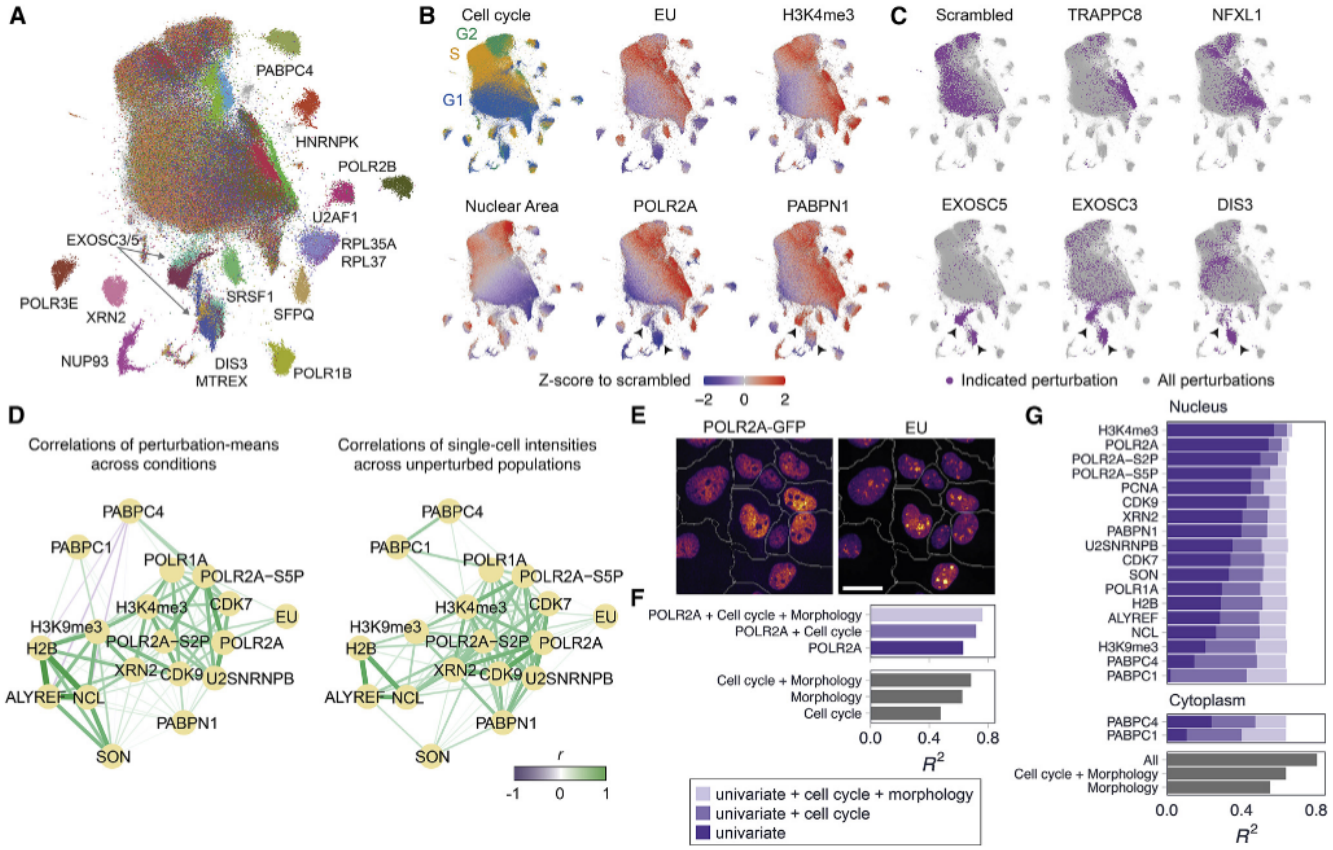


Figure 5. Coordination of transcription machinery abundance with RNA production rates at the single-cell level

(A) UMAP of 434,575 HeLa cells generated from 4i marker intensities, excluding EU. Selected perturbations colored, with remaining cells gray. Labels indicate predominant perturbations in each cluster.

(B) UMAP colored by cell cycle, nuclear area, or nuclear mean intensity of indicated marker.

(C) Distribution of cells on UMAP for selected perturbations.

(D) Pairwise correlations between 4i markers represented as a network, either calculated across perturbations or across unperturbed cell populations. G1 cells only.

(E) GFP fluorescence and EU metabolic labeling in MRC5 POLR2A-GFP cells.

(F) Coefficient of determination (R^2) of linear regression predicting sum nuclear EU in MRC5 POLR2A-GFP cells at the single-cell level. Predictors indicated on axis.

(G) R^2 of linear regression predicting sum nuclear EU in HeLa cells, with 4i markers as predictors, as indicated. Univariate models have a single predictor, while "+ Cell cycle," "+ Morphology," and "All" indicate the additional predictors included (STAR Methods). See also Figures S11 and S12.

clusters (Figure 5C), with DIS3 and MTREX primarily occupying the lower cluster, and EXOSC3 and EXOSC5 occupying both. The EXOSC3/5-specific cluster is distinguished from the other by increased PABPN1 levels (nuclear poly(A)-binding protein) and a less pronounced reduction in POLR2A (Figures S11F and S12A), which may also be related to the increases in nuclear mRNA that we saw using poly(A) FISH for EXOSC3/5, but not MTREX/DIS3 (Figures 3A and S8A).

We noticed that the distribution of unperturbed (scrambled siRNA) cells on the UMAP extended into regions typically occupied by genetically perturbed cells (Figures 5C and S11E). This indicates that there is substantial cell-to-cell variability in levels of measured proteins and PTMs even within unperturbed cell populations. To understand how this heterogeneity is related to that of RNA production rates in unperturbed controls, we calculated correlations between mean intensities of 4i markers and EU across single cells in these control populations. Several markers, including POLR2A, CDK7, and H3K4me3, were positively correlated with EU (Figure S12B), indicating that abundance of RNA Pol II and other active transcription markers in unperturbed populations are coordinated with RNA production rates in single cells. Interestingly, 4i markers that co-vary with EU in unperturbed populations are typically those that also showed concordant changes with EU when the latter is perturbed. That is, correlations $r_{x,EU}^{(\text{unperturbed cells})}$ and $r_{x,EU}^{(\text{perturbation average})}$ are themselves positively correlated across 4i markers ($r = 0.65$, $p < 10^{-7}$; Figures S12C and S12D). Moreover, network representations of these pairwise correlations calculated at the perturbation-scale and the single-cell scale were highly similar (Figure 5D). At both levels, we observed a module of active transcription markers including CDK7, H3K4me3, POLR2A, and SNRPB2 that are mutually coordinated and positively related to EU incorporation, and a second set of factors including H2B, ALYREF, and NCL are also mutually coordinated but are unrelated to EU incorporation.

To determine the relationship between RNA Pol II abundance and RNA production at the single-cell level, we combined EU metabolic labeling with cell volume and cell-cycle measurements in unperturbed MRC5 POLR2A-GFP cells. POLR2A-GFP and EU intensities were highly correlated at the single-cell level, and POLR2A-GFP showed a similar relationship as EU with both cell volume and cell cycle (Figures 5E and S12E–S12H). Furthermore, using linear regression, we found that POLR2A-GFP intensity alone explained a similar fraction of variance in EU incorporation as a "cell size" model with cell volume and protein content as predictors (mean R^2 = 0.61 for both models; Figure 5F). In the HeLa 4i dataset, regression models predicting EU at the single-cell level from each 4i marker individually had R^2 ranging from 0.02 for nuclear PABPC1 to 0.55 and 0.58 for POLR2A and H3K4me3, respectively, with latter values similar to those of a morphology-only model (R^2 = 0.55; Figure 5G). When combined with morphology and cell-cycle information, POLR2A and H3K4me3 were the only two markers which increased R^2 compared with models with morphology and cell-cycle features alone. Overall, this analysis indicates that cells coordinate the cellular abundance of RNA Pol II and other markers of active transcription with RNA production—both in unperturbed cells and when RNA production is perturbed.

Feedback regulation of RNA Pol II activity by nuclear mRNA

Perturbation of the nuclear RNA exosome was a rare example in which mRNA homeostasis is disrupted. In this case, nuclear mRNA levels increase, and this is associated with reduced EU incorporation and reduced levels of RNA Pol II and other markers of active transcription (Figures S12I and S12J). This suggests that nuclear mRNA abundance may negatively feedback on transcription to achieve mRNA concentration homeostasis. Such a feedback could act to coordinate RNA production with nuclear size, which is itself coupled to cell size (Cantwell and Nurse, 2019), and would be consistent with our genome-wide screen, which revealed that nuclear (rather than cytoplasmic) RNA degradation factors are associated with reduced EU incorporation (see Figure S6A). Moreover, *in vitro* work has shown that mRNA is indeed inhibitory for transcription (Henninger et al., 2021; Pai et al., 2014), so this may involve a direct effect.

To explore this hypothesis further, we developed a minimal mathematical model incorporating RNA-Pol-II-mediated transcription and nuclear mRNA. Our experimental observation that transcriptional activity was closely associated with RNA Pol II abundance across a functionally and phenotypically diverse set of perturbations (see Figures 4F and 4G) indicates that RNA Pol II *abundance* is often determined by transcriptional *activity* rather than the other way around. Moreover, it is well known that inhibiting transcription with α-amanitin (Lee et al., 2002; Mitsui and Sharp, 1999; Nguyen et al., 1996) or triptolide (Alekseev et al., 2017; Bensaude, 2011; Steurer et al., 2018) both lead to RNA Pol II degradation. Yet, when these compounds are combined with proteasome inhibition, RNA Pol II detaches from chromatin but remains stable (Steurer et al., 2018). These observations are consistent with RNA Pol II being targeted for degradation primarily when it is not chromatin associated. To formulate this mathematically, we adapted a kinetic model of RNA Pol II transcription derived from fluorescence recovery after photobleaching (FRAP) (Steurer et al., 2018), adding the hypothesis that only "unbound" RNA Pol II can be degraded (Figures 6A and S13A; Table S1). Next, we measured POLR2A and POLR2A-S2P levels using immunofluorescence, after treating cells with triptolide or CDK9 inhibitor AZD4573 (which prevents Ser2-phosphorylation of RNA Pol II CTD; Cidado et al., 2020) and optimized a single free parameter (RNA Pol II synthesis rate) to fit the data. The model quantitatively reproduced triptolide-induced loss of POLR2A, and AZD4573-induced reductions of POLR2A-S2P and POLR2A (Figures 6B and S13B). In contrast, an alternative model in which all RNA Pol II species were subject to degradation did not show RNA Pol II abundance changes. Using this model as a foundation, we next implemented feedback from nuclear mRNA as either a stimulation of RNA Pol II degradation, or inhibition of RNA Pol II initiation, pausing, or pause release. As expected, including this feedback made RNA Pol II levels dependent on nuclear RNA degradation and export rates (Figure 6C).

To test the model, we experimentally examined the timescale of transcriptional adaptation upon disruption of nuclear RNA degradation, using acute depletion of catalytic exosome subunits DIS3 or EXOSC10 in HCT116 cells (Davidson et al., 2019) with the auxin-induced degron (AID) system (Nishimura et al., 2009) (Figures 6D, 6E, and S14A–S14F). In both cases, auxin addition resulted in reduced EU incorporation—with the strongest effects for DIS3. EU reduction was evident from 60 min onward and decreased progressively at later time points. For DIS3, we found that EU reduction was accompanied by accumulation of nuclear mRNA over the same timescale (Figure 6E). EXOSC10 did not accumulate mRNA (Figure S14D), consistent with its direct targets being non-polyadenylated nucleolar RNA (Davidson et al., 2019). To determine whether rapid DIS3 depletion is associated with changes in RNA Pol II, we measured POLR2A, POLR2A-S5P, and POLR2A-S2P by immunofluorescence (Figures 6F and S14G). After 3.5 h of auxin treatment, we observed a ~40% reduction in EU incorporation, which was accompanied by a ~20% reduction in POLR2A-S2P, a ~10% reduction in POLR2A-S5P and a ~6% reduction in total POLR2A. DIS3 depletion therefore initially leads to reduced RNA Pol II *activity*, with reduction in RNA Pol II *abundance* occurring over a longer timescale (see Figures 4C and 4J). We next compared these experimental results with mathematical models in which the transcriptional feedback acts at different stages of the transcription cycle. When mRNA activates RNA Pol II degradation or represses initiation or pausing, the model predicts simultaneous loss of phosphorylated and total POLR2A (Figures 6G and S13C). Conversely, when mRNA represses the transition from pausing to elongation, the model predicts greater loss of POLR2A-S2P than POLR2A-S5P and total POLR2A. Although the differences between RNA Pol II phosphorylation states predicted by the model are small, this latter model is most similar to experimental observations upon DIS3 depletion, suggesting that nuclear mRNA abundance regulates transcriptional activity downstream of transcription initiation.

As a second test of the negative feedback of nuclear RNA concentration on transcription, we made use of artificial arginine-enriched mixed-charge domain (R-MCD) proteins, which were recently found to drive nuclear mRNA retention (Greig et al.,

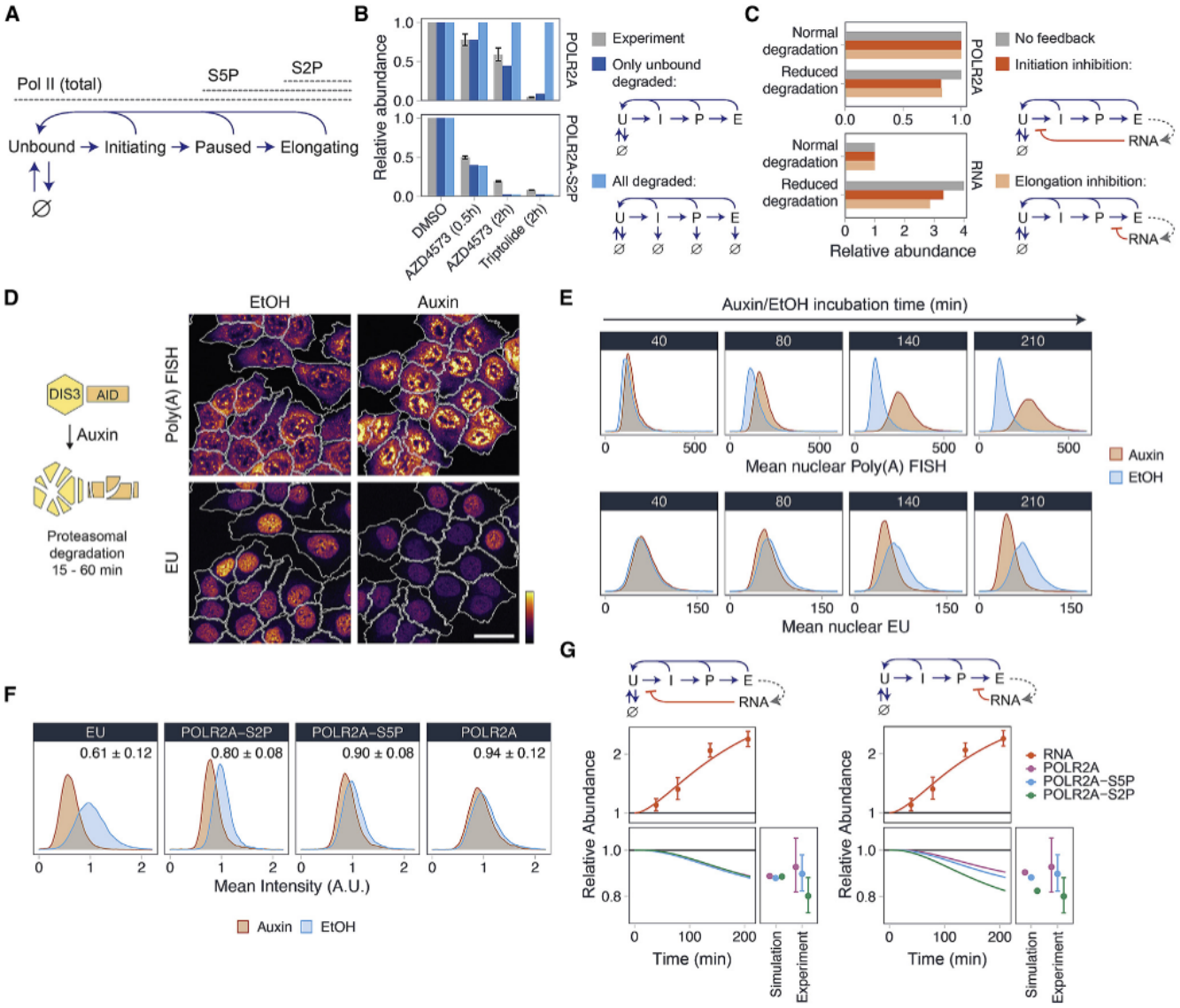


Figure 6. Feedback regulation of RNA polymerase II activity by nuclear mRNA

(A) Minimal mathematical model of RNA Pol II transcriptional states. Arrows to/from ∅ indicate RNA Pol II degradation/synthesis.

(B) Nuclear POLR2A and POLR2A-S2P levels relative to DMSO controls, measured by immunofluorescence upon chemical perturbation of transcription. Error bars represent mean ± s.d. for all pairwise comparisons of treated ($n \geq 2$) and control wells ($n = 6$). Best-fit models shown for comparison.

(C) Steady-state levels of POLR2A and RNA when nuclear RNA degradation rate is reduced (4-fold), for the models indicated in legend.

(D) Poly(A) FISH and EU incorporation (20 min pulse) in DIS3-AID cells after 140 min Auxin or EtOH.

(E) Single-cell intensity distributions of mean nuclear poly(A) FISH and nuclear EU in DIS3-AID cells after auxin/EtOH treatment for the indicated durations. EU pulse for final 20 min before fixation.

(F) Single-cell intensity distributions of mean nuclear EU and POLR2A (immunofluorescence) in DIS3-AID cells treated with Auxin or EtOH for 3.5 h. Inset: population median for auxin relative to EtOH controls (mean ± s.d., n = 6 [IF] or n = 18 [EU] wells).

(G) Model simulation of DIS3-AID depletion experiment for models including negative feedback from RNA, as indicated in schematics above. Best-fit models shown with poly(A) FISH data from (E) and RNA Pol II immunofluorescence data from (F). All error bars represent mean ± s.d. for all pairwise comparisons of treated and control wells. See also Figures S13 and S14.

2020). When expressed in cells, positively charged R-MCD$^{+0.1}$-mGFP and R-MCD$^{+0.2}$-mGFP localized to nuclear speckles, leading to dose- and charge-dependent nuclear accumulation and cytoplasmic depletion of mRNA—in line with previous data (Greig et al., 2020) (Figures 7A, 7B, S15A, and S15B). Total cellular mRNA abundance was not affected. In agreement with our model, nuclear mRNA retention was accompanied by reduced EU incorporation (Figure 7B). Moreover, EU reduction and nuclear mRNA accumulation showed similar quantitative dependence on R-MCD levels (GFP intensity). In addition, nuclear mRNA retention driven by R-MCD$^{+0.2}$-eGFP expression was associated with reduced POLR2A and POLR2A-S2P—most prominently at the highest GFP intensities (Figures 7C, 7D, and S15C–S15F).

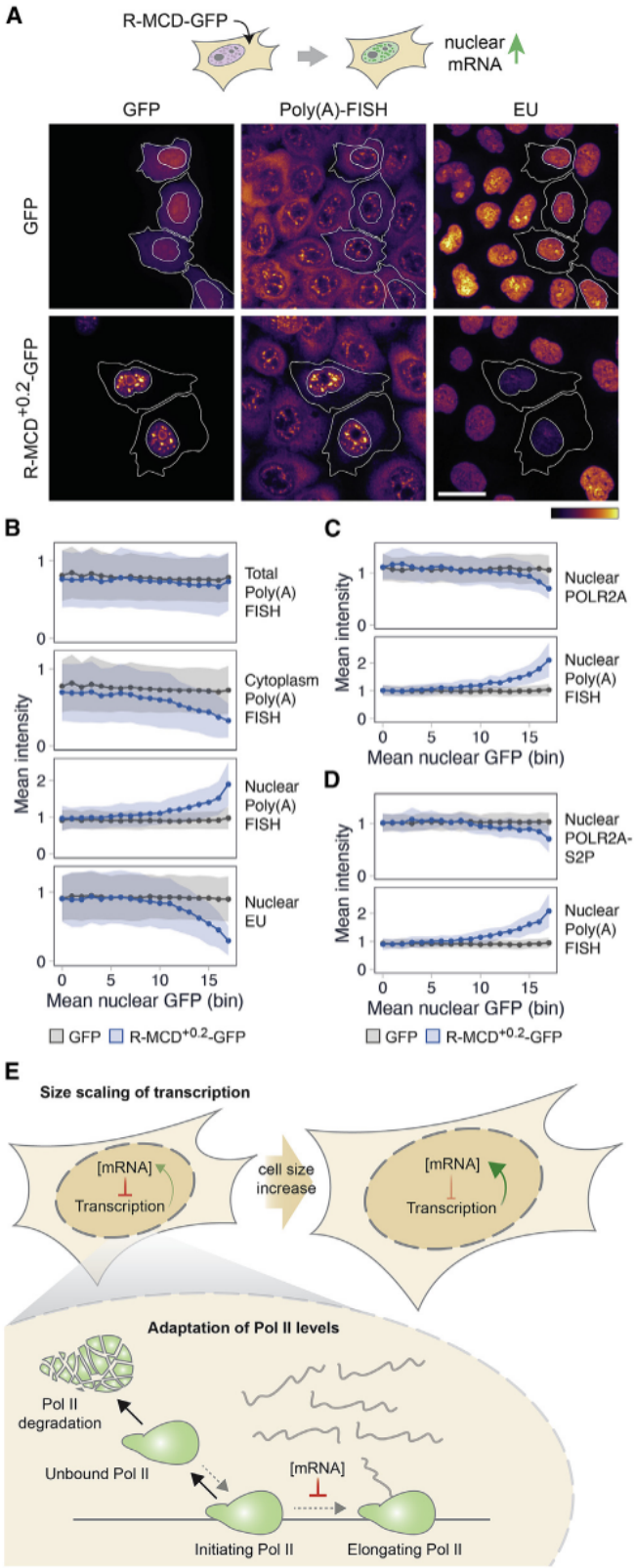


Figure 7. Nuclear mRNA retention leads to transcriptional repression

(A) GFP, poly(A) FISH and EU incorporation in cells expressing R-MCD$^{+0.2}$-GFP or GFP control. GFP-expressing cells outlined.

Thus, elevating nuclear mRNA abundance by either acutely disrupting nuclear RNA degradation or forcing nuclear mRNA retention leads to reduced RNA synthesis. Because exosome depletion initially affects transcriptional activity, with RNA Pol II abundance changes occurring over longer timescales, this suggests that RNA Pol II abundance is dictated by transcriptional activity, rather than determining it. Moreover, since R-MCD expression results in cytoplasmic depletion of mRNA, this confirms observations from the genome-wide screen that nuclear rather than cytoplasmic RNA abundance is relevant for the reduction of RNA synthesis. Further analysis of the model revealed that negative feedback from mRNA makes steady-state concentrations of RNA Pol II and nuclear mRNA less sensitive to changes in parameters such as RNA Pol II synthesis and degradation (Figures S13D and S13E) and also allows faster restoration of RNA Pol II and mRNA concentration homeostasis upon perturbation of cell volume (Figures S13F and S13G)—increasing the robustness of mRNA concentration homeostasis in fluctuating environments.

DISCUSSION

Here, we developed a technique for simultaneous measurement of cell volume, cell-cycle stage, and bulk RNA production rates in single cells, which reveals the relationships between these three variables in asynchronous unperturbed human cells. In agreement with earlier work in fission yeast (Fraser and Nurse, 1979; Zhurinsky et al., 2010), we found that cell size changes in chemical or genetic perturbations were generally associated with corresponding changes to RNA production rates (Figures 2D, S2Q, and S2R), whereas reduced cell growth rates did not generally lead to reduced RNA production rates (Figures 2C and S2N–S2P). By measuring cell cycle, cell size, and EU incorporation simultaneously in single cells, we were able to derive a perturbation-level measurement of RNA production that is "corrected" for changes to cell size and cell-cycle stage (residual mean EU). Previous studies of mRNA buffering in budding yeast were normalized by cell number rather than cell size (Haimovich et al., 2013; Sun et al., 2012, 2013) and are thus insensitive to indirect effects of perturbation-induced cell size changes on the kinetics of RNA metabolism. Given that perturbation of transcription or RNA degradation is often associated with extreme cell size changes in budding yeast (Jorgensen et al., 2002; Maitra et al., 2018), studying these two phenomena together is crucial to resolving the underlying mechanisms (Mena et al., 2017). As we show, this can be achieved by combining perturbations

(B) Mean poly(A) FISH intensity for the cell, cytoplasm, or nucleus, together with nuclear EU as a function of binned nuclear GFP intensity (GFP: n = 23,927, >715 per bin; R-MCD$^{+0.2}$-GFP: n = 6,296, >290 per bin). Points and shaded regions show mean ± s.d.

(C) As in (B), for mean nuclear POLR2A (GFP: n = 6,442, >175 per bin; R-MCD$^{+0.2}$-GFP: n = 1,225, >51 per bin).

(D) As in (B), for mean nuclear POLR2A-S2P (GFP: n = 5,167, >123 per bin; R-MCD$^{+0.2}$-GFP: n = 1,346, >61 per bin).

(E) Schematic of the proposed mechanism. Cell size increase relieves repression on transcription through transient mRNA dilution, allowing stabilization of RNA Pol II and coordinated increase in RNA synthesis rate with cell size. See also Figure S15.

with quantitative multivariate measurements of single cells at the population scale. The datasets generated here will therefore be valuable resources for further investigating the interplay between cell size and transcription rates, and how this is impacted by the diverse molecular processes that we identified as having functional roles in maintaining normal levels of cellular RNA production.

Using highly multiplexed immunofluorescence, we found that transcription rates are positively correlated with RNA Pol II abundance and other hallmarks of RNA production including H3K4me3, CDK7, and SNRPB2. However, both long-term and acute perturbation experiments show that this coordination is not because RNA Pol II abundance directly determines transcription rates but because RNA Pol II abundance is adapted to transcriptional activity. Rather, our experiments and modeling point to nuclear mRNA concentration as being the quantity under strict homeostatic regulation and reveal that this negatively impinges on transcriptional activity to ultimately determine the abundance of the transcription machinery. In contrast to the "limiting factor" model of transcriptional scaling (Lin and Amir, 2018; Marguerat and Bähler, 2012; Padovan-Merhar et al., 2015; Sun et al., 2020; Swaffer et al., 2021), this mechanism has a strong parallel with classic "feedback inhibition"—employed pervasively throughout metabolic networks to coordinate activities of biosynthetic enzymes with cellular requirements (Pardee and Reddy, 2003). Feedback inhibition through allosteric effects is more robust than control of enzyme abundance (Sander et al., 2019), which is analogous to mRNA acting primarily on transcriptional activity rather than on RNA Pol II abundance, as suggested by fitting the model to the DIS3-AID experiments. Importantly, the mRNA feedback model is not in conflict with recent findings in budding yeast that overexpression or depletion of RNA Pol II changes the amount of RNA Pol II associated with chromatin (Swaffer et al., 2021). These perturbations alter the ratio between RNA Pol II and nuclear mRNA and would override the homeostatic mechanism through which nuclear RNA controls RNA Pol II levels. To further test this model, it will be important to resolve the molecular mechanisms controlling RNA Pol II abundance, and to determine how these are impacted by nuclear mRNA concentration. Microinjection of exogenous RNA, more extensive quantitative control of nuclear and cytoplasmic RNA nuclease activity, acute perturbations of cell and nuclear volume, and live cell imaging of RNA Pol II kinetics will be useful approaches to elucidate this.

The mechanism proposed here for mammalian cells contrasts the buffering model proposed for budding yeast, which was based on the finding that deletion of cytoplasmic RNA degradation factors, especially Xrn1, results in reduced transcription (Haimovich et al., 2013; Sun et al., 2013). While we also find that cytoplasmic RNA concentration is buffered upon genetic perturbation of transcription, multiple observations indicate a crucial role for the nucleus in this process: First, the genome-wide screen reveals that perturbation of nuclear rather that cytoplasmic RNA degradation factors impact transcription (Figures 2G and S5A). Second, only upon perturbation of nuclear RNA degradation and export is homeostasis of mRNA concentration broken (Figure 3E). Third, nuclear RNA retention experiments indicate that nuclear rather than cytoplasmic RNA is relevant for the homeostatic feedback on transcription (Figures 6H and 6I). This is also consistent with reports that accelerated cytoplasmic mRNA degradation does not result in a compensatory increase in transcription in human cells (Abernathy et al., 2015; Gilbertson et al., 2018). Interestingly, recent work in fission yeast also suggests that nuclear size rather than cell size may be the quantitative determinant of mRNA size scaling (Sun et al., 2020).

The detailed mechanism by which mRNA concentration regulates RNA production remains to be characterized. A direct effect is possible. Indeed, *in vitro* transcription is inhibited by exogenous RNA. This has been proposed to occur by RNA directly interfering with RNA Pol II binding to the DNA template (Pai et al., 2014) and could involve interference with the phase separation of transcriptional condensates on chromatin (Henninger et al., 2021; Portz and Shorter, 2021). An indirect effect is also possible: with mRNA modulating the activity or localization of a limiting transcriptional regulator. Nuclear RNA concentration has been proposed to regulate condensation of nuclear RNA-binding proteins (Maharana et al., 2018), and, interestingly, nuclear speckles are enriched for transcription elongation factors, rather than transcription initiation factors (Galganski et al., 2017). It is therefore possible that increased mRNA levels in nuclear speckles, for example, as observed upon R-MCD overexpression, could result in retention of elongation factors and thereby suppress efficient RNA Pol II elongation.

The minimal model developed here explains several experimental observations and "buffers" mRNA concentration. However, in its current form, it does not achieve robust "set point" homeostasis (Briat et al., 2016; Reed et al., 2017). Further elaboration of the model requires more details on the molecular nature of the transcriptional feedback, and on how mRNA degradation and export rates are affected by mRNA abundance. For example, nuclear mRNA export in human cells appears to depend on ongoing transcription (Tokunaga et al., 2006), which could also contribute to the maintenance of nuclear mRNA levels that we observed in perturbations with strong transcriptional repression, such as POLR2B knockdown. In addition, work in budding yeast has revealed that blocking nuclear export can lead to increased degradation of newly synthesized mRNA (Tudek et al., 2018). Such mechanisms could both play key roles in set-point homeostasis of nuclear mRNA concentration, but it is currently unclear how they are interconnected. The rich perturbation datasets and simple model that emerge from our work, linking activity and abundance of the transcription machinery to nuclear mRNA concentration, provide a starting point to explore links between these global cellular controls of mRNA metabolism and how they relate to the volume of mammalian cells.

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 SUBJECT DETAILS
  - Cell lines and culture conditions

- METHOD DETAILS
  - siRNA transfection
  - Plasmid transfection
  - Chemical treatments
  - Image acquisition
  - DNA and protein staining
  - High-throughput cell volume measurement
  - Fluorescence exclusion method (FXm)
  - Single-molecule RNA fluorescent in situ hybridisation (smFISH)
  - In situ metabolic labelling of nascent RNA
  - Immunofluorescence and iterative indirect immunofluorescence imaging (4i)
  - Poly(A) RNA fluorescent in situ hybridisation (FISH)
  - Strandbrite total RNA staining
  - RNA extraction and quantification
  - Quantitative reverse transcription PCR (RT-qPCR)
  - RNA dot blot
  - Immunoblotting
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Image processing
  - Data cleaning
  - Background subtraction
  - Data normalisation
  - Cell cycle classification
  - Quantitative data analysis
  - Systems-level analysis
  - Mathematical modelling

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cels.2022.04.005.

ACKNOWLEDGMENTS

We thank Larisa Venkova and Matthieu Piel for FXm training and resources; Merve Avar, Daniel Heinzer, Marc Emmenegger, and Adriano Aguzzi for printing siRNA libraries; David Spector for the anti-SNRPB2 antibody; and all members of Lucas Pelkmans' lab for discussions and manuscript comments. AID cells were kindly provided by Steven West. POLR2A-GFP cells were kindly provided by Jurgen Marteijn. S.B. is supported by an EMBO long-term fellowship (ALTF1175-2016), an HFSP long-term fellowship (LT000238/2017-L), and the University of New South Wales. L.P. is supported by the European Research Council (ERC-2019-AdG-885579). L.P. and A.R. are supported by the Swiss National Science Foundation (SNSF grant 310030_192622). All authors are supported by the University of Zurich.

AUTHOR CONTRIBUTIONS

Conceptualization, S.B. and L.P.; methodology, S.B.; investigation, S.B., M.M., and A.R.; formal analysis, S.B. and M.M.; writing—original draft, S.B.; writing—review & editing, S.B., M.M., A.R., and L.P.; funding acquisition, S.B., A.R., and L.P.

DECLARATION OF INTERESTS

L.P. has filed a patent on the 4i technology (patent WO2019207004A1).

Received: May 10, 2021
Revised: December 13, 2021
Accepted: April 21, 2022
Published: May 24, 2022

REFERENCES

Abernathy, E., Gilbertson, S., Alla, R., and Glaunsinger, B. (2015). Viral nucleases induce an mRNA degradation-transcription feedback loop in mammalian cells. Cell Host Microbe *18*, 243–253.

Abraham, K.J., Khosraviani, N., Chan, J.N.Y., Gorthi, A., Samman, A., Zhao, D.Y., Wang, M., Bokros, M., Vidya, E., Ostrowski, L.A., et al. (2020). Nucleolar RNA polymerase II drives ribosome biogenesis. Nature *585*, 298–302.

Alekseev, S., Nagy, Z., Sandoz, J., Weiss, A., Egly, J.-M., May, N.L., and Coin, F. (2017). Transcription without XPB establishes a unified helicase-independent mechanism of promoter opening in eukaryotic gene expression. Mol. Cell *65*, 504–514.e4.

Arachchige, C.N.P.G., Prendergast, L.A., and Staudte, R.G. (2022). Robust analogs to the coefficient of variation. J. Appl. Stat. *49*, 268–290.

Battich, N., Stoeger, T., and Pelkmans, L. (2013). Image-based transcriptomics in thousands of single human cells at single-molecule resolution. Nat. Methods *10*, 1127–1133.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of transcript variability in single mammalian cells. Cell *163*, 1596–1610.

Bensaude, O. (2011). Inhibiting eukaryotic transcription. Which compound to choose? How to evaluate its activity? Transcription *2*, 103–108.

Berchtold, D., Battich, N., and Pelkmans, L. (2018). A systems-level study reveals regulators of membrane-less organelles in human cells. Mol. Cell *72*, 1035–1049.e5.

Biedermann, T., Pontiggia, L., Böttcher-Haberzeth, S., Tharakan, S., Braziulis, E., Schiestl, C., Meuli, M., and Reichmann, E. (2010). Human eccrine sweat gland cells can reconstitute a stratified epidermis. J. Invest. Dermatol. *130*, 1996–2009.

Boulon, S., Westman, B.J., Hutten, S., Boisvert, F.-M., and Lamond, A.I. (2010). The nucleolus under stress. Mol. Cell *40*, 216–227.

Briat, C., Gupta, A., and Khammash, M. (2016). Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. Cell Syst. *2*, 15–26.

Burger, K., Mühl, B., Rohrmoser, M., Coordes, B., Heidemann, M., Kellner, M., Gruber-Eber, A., Heissmeyer, V., Strässer, K., and Eick, D. (2013). Cyclin-dependent kinase 9 links RNA polymerase II transcription to processing of ribosomal RNA. J. Biol. Chem. *288*, 21173–21183.

Cadart, C., Monnier, S., Grilli, J., Sáez, P.J., Srivastava, N., Attia, R., Terriac, E., Baum, B., Cosentino-Lagomarsino, M., and Piel, M. (2018). Size control in mammalian cells involves modulation of both growth rate and cell cycle duration. Nat. Commun. *9*, 3275.

Cadart, C., Zlotek-Zlotkiewicz, E., Venkova, L., Thouvenin, O., Racine, V., Le Berre, M.L., Monnier, S., and Piel, M. (2017). Fluorescence eXclusion measurement of volume in live cells. Methods Cell Biol. *139*, 103–120.

Cantwell, H., and Nurse, P. (2019). Unravelling nuclear size control. Curr. Genet. *65*, 1281–1285.

Caudron-Herger, M., Pankert, T., and Rippe, K. (2016). Regulation of nucleolus assembly by non-coding RNA polymerase II transcripts. Nucleus *7*, 308–318.

Cidado, J., Boiko, S., Proia, T., Ferguson, D., Criscione, S.W., San Martin, M., Pop-Damkov, P., Su, N., Roamio Franklin, V.N., Sekhar Reddy Chilamakuri, C., et al. (2020). AZD4573 is a highly selective CDK9 inhibitor that suppresses MCL-1 and induces apoptosis in hematologic cancer cells. Clin. Cancer Res. *26*, 922–934.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educ. Psychol. Meas. *20*, 37–46.

Da, T.K.F., Loriot, S., and Yvinec, M. (2022). 3D alpha shapes. In CGAL user and reference manual, [5.4 Edition] (CGAL Editorial Board).

Davidson, L., Francis, L., Cordiner, R.A., Eaton, J.D., Estell, C., Macias, S., Cáceres, J.F., and West, S. (2019). Rapid depletion of DIS3, EXOSC10, or XRN2 reveals the immediate impact of exoribonucleolysis on nuclear RNA metabolism and transcriptional control. Cell Rep. *26*, 2779–2791.e5.

Drygin, D., Lin, A., Bliesath, J., Ho, C.B., O'Brien, S.E., Proffitt, C., Omori, M., Haddach, M., Schwaebe, M.K., Siddiqui-Jain, A., et al. (2011). Targeting RNA

polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. Cancer Res. *71*, 1418–1430.

Eaton, J.D., Davidson, L., Bauer, D.L.V., Natsume, T., Kanemaki, M.T., and West, S. (2018). Xrn2 accelerates termination by RNA polymerase II, which is underpinned by CPSF73 activity. Genes Dev. *32*, 127–139.

Elzhov, T.V., Mullen, K.M., Spiess, A.-N., and Bolker, B. (2016). minpack.lm: R interface to the Levenberg-Marquardt nonlinear least-squares algorithm found in MINPACK, plus support for bounds. https://CRAN.R-project.org/package=minpack.lm.

Fan, J., Kuai, B., Wang, K., Wang, L., Wang, Y., Wu, X., Chi, B., Li, G., and Cheng, H. (2018). mRNAs are sorted for export or degradation before passing through nuclear speckles. Nucleic Acids Res. *46*, 8404–8416.

Fraser, R.S., and Nurse, P. (1979). Altered patterns of ribonucleic acid synthesis during the cell cycle: a mechanism compensating for variation in gene concentration. J. Cell Sci. *35*, 25–40.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. *33*, 1–22.

Galganski, L., Urbanek, M.O., and Krzyzosiak, W.J. (2017). Nuclear speckles: molecular organization, biological function and role in disease. Nucleic Acids Res. *45*, 10350–10368.

Gilbertson, S., Federspiel, J.D., Hartenian, E., Cristea, I.M., and Glaunsinger, B. (2018). Changes in mRNA abundance drive shuttling of RNA binding proteins, linking cytoplasmic RNA degradation to transcription. eLife *7*, 243.

Gillery, P., Georges, N., Wegrowski, J., Randoux, A., and Borel, J.-P. (1995). Protein synthesis in collagen lattice-cultured fibroblasts is controlled at the ribosomal level. FEBS Lett. *357*, 287–289.

Ginzberg, M.B., Chang, N., D'Souza, H., Patel, N., Kafri, R., and Kirschner, M.W. (2018). Cell size sensing in animal cells coordinates anabolic growth rates and cell cycle progression to maintain cell size uniformity. eLife *7*, 7729.

Glover-Cutter, K., Larochelle, S., Erickson, B., Zhang, C., Shokat, K., Fisher, R.P., and Bentley, D.L. (2009). TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. Mol. Cell. Biol. *29*, 5455–5464.

Green, V.A., and Pelkmans, L. (2016). A systems survey of progressive host-cell reorganization during rotavirus infection. Cell Host Microbe *20*, 107–120.

Greenberg, E., Hochberg-Laufer, H., Blanga, S., Kinor, N., and Shav-Tal, Y. (2019). Cytoplasmic DNA can be detected by RNA fluorescence in situ hybridization. Nucleic Acids Res. *47*, e109.

Greig, J.A., Nguyen, T.A., Lee, M., Holehouse, A.S., Posey, A.E., Pappu, R.V., and Jedd, G. (2020). Arginine-enriched mixed-charge domains provide cohesion for nuclear speckle condensation. Mol. Cell *77*, 1237–1250.e4.

Grudzien-Nogalska, E., and Kiledjian, M. (2017). New insights into decapping enzymes and selective mRNA decay. Wiley Interdiscip. Rev. RNA *8*, e1379.

Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics *32*, 2847–2849.

Gut, G., Herrmann, M.D., and Pelkmans, L. (2018). Multiplexed protein maps link subcellular organization to cellular states. Science *361*, eaar7042.

Hahsler, M., Hornik, K., and Buchta, C. (2008). Getting things in order: an introduction to the R package seriation. J. Stat. Softw. *25*, 1–34.

Haimovich, G., Medina, D.A., Causse, S.Z., Garber, M., Millán-Zambrano, G., Barkai, O., Chávez, S., Pérez-Ortín, J.E., Darzacq, X., and Choder, M. (2013). Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell *153*, 1000–1011.

Halle, J.P., Müller, S., Simm, A., and Adam, G. (1997). Copy number, epigenetic state and expression of the rRNA genes in young and senescent rat embryo fibroblasts. Eur. J. Cell Biol. *74*, 281–288.

Habets, W., Hoet, M., Bringmann, P., Lührmann, R., and Venrooij, W.van (1985). Autoantibodies to ribonucleoprotein particles containing U2 small nuclear RNA. Embo J. *4*, 1545–1550.

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362.

Hartenian, E., and Glaunsinger, B.A. (2019). Feedback to the central dogma: cytoplasmic mRNA decay and transcription are interdependent processes. Crit. Rev. Biochem. Mol *54*, 1–14.

Heidemann, M., Hintermair, C., Voß, K., and Eick, D. (2013). Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochim. Biophys. Acta *1829*, 55–62.

Helenius, K., Yang, Y., Tselykh, T.V., Pessa, H.K.J., Frilander, M.J., and Mäkelä, T.P. (2011). Requirement of TFIIH kinase subunit Mat1 for RNA Pol II C-terminal domain Ser5 phosphorylation, transcription and mRNA turnover. Nucleic Acids Res. *39*, 5025–5035.

Henninger, J.E., Oksuz, O., Shrinivas, K., Sagi, I., LeRoy, G., Zheng, M.M., Andrews, J.O., Zamudio, A.V., Lazaris, C., Hannett, N.M., et al. (2021). RNA-mediated feedback control of transcriptional condensates. Cell *184*, 207–225.e24.

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.

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.

Jao, C.Y., and Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. Proc. Natl. Acad. Sci. USA *105*, 15779–15784.

Jorgensen, P., Nishikawa, J.L., Breitkreutz, B.-J., and Tyers, M. (2002). Systematic identification of pathways that couple cell growth and division in yeast. Science *297*, 395–400.

Kafri, R., Levy, J., Ginzberg, M.B., Oh, S., Lahav, G., and Kirschner, M.W. (2013). Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. Nature *494*, 480–483.

Kempe, H., Schwabe, A., Crémazy, F., Verschure, P.J., and Bruggeman, F.J. (2015). The volumes and transcript counts of single cells reveal concentration homeostasis and capture biological noise. Mol. Biol. Cell *26*, 797–804.

Kramer, B.A., and Pelkmans, L. (2019). Cellular state determines the multimodal signaling response of single cells. Biorxiv. 2019.12.18.880930.

Kim, S. (2015). ppcor: partial and semi-partial (part) correlation. https://CRAN.R-project.org/package=ppcor.

Łabno, A., Tomecki, R., and Dziembowski, A. (2016). Cytoplasmic RNA decay pathways—enzymes and mechanisms. Biochim. Biophys. Acta *1863*, 3125–3147.

Lee, J.E., Lee, J.Y., Trembly, J., Wilusz, J., Tian, B., and Wilusz, C.J. (2012). The PARN deadenylase targets a discrete set of mRNAs for decay and regulates cell motility in mouse myoblasts. PLOS Genet. *8*, e1002901.

Lee, K.-B., Wang, D., Lippard, S.J., and Sharp, P.A. (2002). Transcription-coupled and DNA damage-dependent ubiquitination of RNA polymerase II in vitro. Proc. Natl. Acad. Sci. USA *99*, 4239–4244.

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. R News *2*, 18–22.

Lin, J., and Amir, A. (2018). Homeostasis of protein and mRNA concentrations in growing cells. Nat. Commun. *9*, 4496.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods *25*, 402–408.

Maharana, S., Wang, J., Papadopoulos, D.K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J., Franzmann, T.M., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science *360*, 918–921.

Maitra, N., Anandhakumar, J., Blank, H.M., Kaplan, C.D., and Polymenis, M. (2018). Perturbations of transcription and gene expression-associated processes alter distribution of cell size values in *Saccharomyces cerevisiae*. G3 Genes Genomes Genet. *9*, 239–250.

Marguerat, S., and Bähler, J. (2012). Coordinating genome expression with cell size. Trends Genet. *28*, 560–565.

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at arXiv. https://doi.org/10.48550/arXiv.1802.03426.

Melville, J. (2021). uwot: the uniform manifold approximation and projection (UMAP) method for dimensionality reduction. https://CRAN.R-project.org/package=uwot.

Mena, A., Medina, D.A., García-Martínez, J., Begley, V., Singh, A., Chávez, S., Muñoz-Centeno, M.C., and Pérez-Ortín, J.E. (2017). Asymmetric cell division requires specific mechanisms for adjusting global transcription. Nucleic Acids Res. *45*, 12401–12412.

Meyer, D., Zeileis, A., and Hornik, K. (2020). vcd: visualizing categorical data. https://cran.r-project.org/package=vcd.

Mitsui, A., and Sharp, P.A. (1999). Ubiquitination of RNA polymerase II large subunit signaled by phosphorylation of carboxyl-terminal domain. Proc. Natl. Acad. Sci. USA *96*, 6054–6059.

Morris, J.H., Kuchinsky, A., Ferrin, T.E., and Pico, A.R. (2014). enhancedGraphics: a Cytoscape app for enhanced node graphics. F1000Res *3*, 147.

Müller, M., Avar, M., Heinzer, D., Emmenegger, M., Aguzzi, A., Pelkmans, L., and Berry, S. (2021). High content genome-wide siRNA screen to investigate the coordination of cell size and RNA production. Sci. Data *8*, 162.

Nash, J.C., and Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: optimx for R. J. Stat. Softw. *43*, 1–14.

Nguyen, V.T., Giannoni, F., Dubois, M.F., Seo, S.J., Vigneron, M., Kédinger, C., and Bensaude, O. (1996). In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. Nucleic Acids Res. *24*, 2924–2929.

Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T., and Kanemaki, M. (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. Nat. Methods *6*, 917–922.

Padovan-Merhar, O., Nair, G.P., Biaesch, A.G., Mayer, A., Scarfone, S., Foley, S.W., Wu, A.R., Churchman, L.S., Singh, A., and Raj, A. (2015). Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol. Cell *58*, 339–352.

Pai, D.A., Kaplan, C.D., Kweon, H.K., Murakami, K., Andrews, P.C., and Engelke, D.R. (2014). RNAs nonspecifically inhibit RNA polymerase II by preventing binding to the DNA template. RNA *20*, 644–655.

Pardee, A.B., and Reddy, G.P.-V. (2003). Beginnings of feedback inhibition, allostery, and multi-protein complexes. Gene *321*, 17–23.

Peterlin, B.M., and Price, D.H. (2006). Controlling the elongation phase of transcription with P-TEFb. Mol. Cell *23*, 297–305.

Pfeiffer, S.E., and Tolmach, L.J. (1968). RNA synthesis in synchronously growing populations of HeLa S3 cells. I. Rate of total RNA synthesis and its relationship to DNA synthesis. J. Cell. Physiol. *71*, 77–93.

Portz, B., and Shorter, J. (2021). Biochemical timekeeping via reentrant phase transitions. J. Mol. Biol. *433*, 166794.

R Core Team (2014). R: a language and environment for statistical computing (R Core Team).

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Rämö, P., Sacher, R., Snijder, B., Begemann, B., and Pelkmans, L. (2009). CellClassifier: supervised learning of cellular phenotypes. Bioinformatics *25*, 3028–3030.

R Core Team (2021). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). https://www.R-project.org/.

Reed, M., Best, J., Golubitsky, M., Stewart, I., and Nijhout, H.F. (2017). Analysis of homeostatic mechanisms in biochemical networks. Bull. Math. Biol. *79*, 2534–2557.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio (Boston, MA: PBC). http://www.rstudio.com/.

Sander, T., Farke, N., Diehl, C., Kuntz, M., Glatter, T., and Link, H. (2019). Allosteric feedback inhibition enables robust amino acid biosynthesis in E. coli by enforcing enzyme overabundance. Cell Syst. *8*, 66–75.e8.

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

Schmidt, E.E., and Schibler, U. (1995). Cell size regulation, a mechanism that controls cellular RNA accumulation: consequences on regulation of the ubiquitous transcription factors Oct1 and NF-Y and the liver-enriched transcription factor DBP. J. Cell Biol. *128*, 467–483.

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.-H.L., Koulena, N., Cronin, C., Karp, C., Liaw, E.J., et al. (2018). Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. Cell *174*, 363–376.e16.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

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.

Simillion, C. (2020). GeneSets.Homosapiens: gene set annotation tables. https://github.com/C3c6e6/GeneSets.

Singh, P., James, R.S., Mee, C.J., and Morozov, I.Y. (2019). mRNA levels are buffered upon knockdown of RNA decay and translation factors via adjustment of transcription rates in human HepG2 cells. RNA Biol. *16*, 1147–1155.

Siwaszek, A., Ukleja, M., and Dziembowski, A. (2014). Proteins involved in the degradation of cytoplasmic mRNA in the major eukaryotic model systems. RNA Biol. *11*, 1122–1136.

Slobodin, B., Bahat, A., Sehrawat, U., Becker-Herman, S., Zuckerman, B., Weiss, A.N., Han, R., Elkon, R., Agami, R., Ulitsky, I., et al. (2020). Transcription dynamics regulate poly(A) tails and expression of the RNA degradation machinery to balance mRNA levels. Mol. Cell *78*, 434–444.e5.

Snijder, B., Sacher, R., Rämö, P., Liberali, P., Mench, K., Wolfrum, N., Burleigh, L., Scott, C.C., Verheije, M.H., Mercer, J., et al. (2012). Single-cell analysis of population context advances RNAi screening at multiple levels. Mol. Syst. Biol. *8*, 579.

Soetaert, K., Petzoldt, T., and Setzer, R.W. (2010). Solving differential equations in R : package deSolve. J. Stat. Softw. *33*, 1–25.

Sommer, C., Straehle, C., Kothe, U., and Hamprecht, F.A. (2011). Ilastik: interactive learning and segmentation toolkit. In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro 1, pp. 230–233.

Steurer, B., Janssens, R.C., Geverts, B., Geijer, M.E., Wienholz, F., Theil, A.F., Chang, J., Dealy, S., Pothof, J., van Cappellen, W.A., et al. (2018). Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA polymerase II. Proc. Natl. Acad. Sci. USA *115*, E4368–E4376.

Stoeger, T., Battich, N., Herrmann, M.D., Yakimovich, Y., and Pelkmans, L. (2015). Computer vision for image-based transcriptomics. Methods *85*, 44–53.

Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods *18*, 100–106.

Sun, M., Schwalb, B., Pirkl, N., Maier, K.C., Schenk, A., Failmezger, H., Tresch, A., and Cramer, P. (2013). Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol. Cell *52*, 52–62.

Sun, M., Schwalb, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., and Cramer, P. (2012). Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. *22*, 1350–1359.

Sun, X.-M., Bowman, A., Priestman, M., Bertaux, F., Martinez-Segura, A., Tang, W., Whilding, C., Dormann, D., Shahrezaei, V., and Marguerat, S. (2020). Size-dependent increase in RNA polymerase II initiation rates mediates gene expression scaling with cell size. Curr. Biol. *30*, 1217–1230.e7.

Swaffer, M.P., Marinov, G.K., Zheng, H., Jones, A.W., Greenwood, J., Kundaje, A., Snijders, A.P., Greenleaf, W.J., Reyes-Lamothe, R., and Skotheim, J.M. (2021). RNA polymerase II dynamics and mRNA stability

feedback determine mRNA scaling with cell size. Preprint at bioRxiv. https://doi.org/10.1101/2021.09.20.461005.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. *47*, D607–D613.

Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res. *22*, 947–956.

Titov, D.V., Gilman, B., He, Q.-L., Bhat, S., Low, W.-K., Dang, Y., Smeaton, M., Demain, A.L., Miller, P.S., Kugel, J.F., et al. (2011). XPB, a subunit of TFIIH, is a target of the natural product triptolide. Nat. Chem. Biol. *7*, 182–188.

Todorov, V., and Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. J. Stat. Softw. *32*, 1–47.

Tokunaga, K., Shibuya, T., Ishihama, Y., Tadakuma, H., Ide, M., Yoshida, M., Funatsu, T., Ohshima, Y., and Tani, T. (2006). Nucleocytoplasmic transport of fluorescent mRNA in living mammalian cells: nuclear mRNA export is coupled to ongoing gene transcription. Genes Cells *11*, 305–317.

Tomecki, R., Kristiansen, M.S., Lykke-Andersen, S., Chlebowski, A., Larsen, K.M., Szczesny, R.J., Drazkowska, K., Pastula, A., Andersen, J.S., Stepien, P.P., et al. (2010). The human core exosome interacts with differentially localized processive RNases: hDIS3 and hDIS3L. EMBO J. *29*, 2342–2357.

Tudek, A., Schmid, M., Makaras, M., Barrass, J.D., Beggs, J.D., and Jensen, T.H. (2018). A nuclear export block triggers the decay of newly synthesized polyadenylated RNA. Cell Rep. *24*, 2457–2467.e7.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T.; scikit-image Contributors (2014). scikit-image: image processing in Python. PeerJ *2*, e453.

Vandesompele, J., Preter, K.D., Pattyn, F., Poppe, B., Roy, N.V., Paepe, A.D., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*. RESEARCH0034.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

Voichek, Y., Bar-Ziv, R., and Barkai, N. (2016). Expression homeostasis during DNA replication. Science *351*, 1087–1090.

Wang, X., Le, N., Denoth-Lippuner, A., Barral, Y., and Kroschewski, R. (2016). Asymmetric partitioning of transfected DNA during mammalian cell division. Proc. Natl. Acad. Sci. USA *113*, 7177–7182.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., et al. (2019). Welcome to the tidyverse. J. Open Source Softw. *4*, 1686.

Wolf, S.F., and Schlessinger, D. (1977). Nuclear metabolism of ribosomal RNA in growing, methionine-limited, and ethionine-treated HeLa cells. Biochemistry *16*, 2783–2791.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics *13*, 134.

Zhurinsky, J., Leonhard, K., Watt, S., Marguerat, S., Bähler, J., and Nurse, P. (2010). A coordinated global control over cellular transcription. Curr. Biol. *20*, 2010–2015.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-ALYREF	Santa Cruz Biotechnology	Cat#sc-32311; Clone#11G5; Lot#G1517; RRID: AB_626667
Mouse monoclonal anti-CDK7	Santa Cruz Biotechnology	Cat#sc-7344; Clone#C-4; Lot#L011; RRID: AB 627243
Rabbit monoclonal anti-CDK9	Cell Signaling Technology	Cat#2316; Clone#C12F7; Lot#7; RRID: AB_2291505
Chicken polyclonal anti-H2B	Abcam	Cat#ab134211; RRID: AB_2889048
Rabbit polyclonal anti-H3K4me3	Abcam	Cat#ab8580; Lot#GR3190162-1; RRID: AB_306649
Rabbit polyclonal anti-H3K9me3	Abcam	Cat#ab8898; RRID: AB_306848
Rabbit polyclonal anti-NCL (C23)	Santa Cruz Biotechnology	Cat#sc-13057; Lot#G3114; RRID: AB 2229696
Mouse monoclonal anti-PABPC1	Santa Cruz Biotechnology	Cat#sc-32318; Clone#10E10; Lot#D1913; RRID: AB_628097
Rabbit polyclonal anti-PABPC4	Atlas Antibodies	Cat#HPA027301; RRID: AB 2672208
Rabbit monoclonal anti-PABPN1	Abcam	Cat#ab75855; Clone#EP3000Y; Lot#GR3202568-6; RRID: AB 1310538
Rabbit monoclonal anti-PCNA	Cell Signaling Technology	Cat#13110; Clone#D3H8P; Lot#4; RRID: AB_2636979
Mouse monoclonal anti-POLR1A (RPA194)	Santa Cruz Biotechnology	Cat#sc-48385; Clone#C-1; RRID: AB_675814
Mouse monoclonal anti-POLR2A (RPB1)	Santa Cruz Biotechnology	Cat#sc-55492; Clone#F-12; Lot#C0119; RRID: AB_630203
Rat monoclonal anti-POLR2A-S2P	Millipore	Cat#04-1571; Clone#3E10; Lot#3169853; RRID: AB_10627998
Rat monoclonal anti-POLR2A-S5P	Millipore	Cat#04-1572; Clone#3E8; RRID:AB_10615822
Rabbit polyclonal anti-SON	Atlas Antibodies	Cat#HPA023535; RRID: AB_1857362
Mouse monoclonal anti-SNRPB2 (U2 B")	Laboratory of David Spector (Habets et al., 1985)	Clone#4G3; RRID: N/A
Rabbit polyclonal anti-XRN2	Atlas Antibodies	Cat#HPA047118; RRID:AB_2679941
Goat polyclonal anti-Mouse IgG (H&L), Alexa Fluor 488	Thermo-Fisher	Cat#A11029; RRID: AB_138404
Goat polyclonal anti-Rabbit IgG (H&L), Alexa Fluor 568	Thermo-Fisher	Cat#A11036; RRID:AB_10563566
Goat polyclonal anti-Chicken IgY (H&L), Alexa Fluor 405	Abcam	Cat#ab175674; RRID:AB_2890171
Goat polyclonal anti-Rabbit IgG (H&L), Alexa Fluor 488	Thermo-Fisher	Cat#A11034; RRID: AB_2576217
Goat polyclonal anti-Mouse IgG (H&L), Alexa Fluor 568	Thermo-Fisher	Cat#A11031; RRID: AB_144696
Goat polyclonal anti-Rat IgG (H&L), Alexa Fluor 568	Thermo-Fisher	Cat#A11077; RRID: AB_141874
Mouse monoclonal anti-mini-AID-tag	MBL International	Cat# M214-3, RRID:AB_2890014
Mouse polyclonal anti-DIS3	Abnova	Cat#H00022894-B01P; RRID:AB 10562157
Rabbit monoclonal anti-Actin	Abcam	Cat#ab179467; RRID:AB_2737344
Horse anti-Mouse IgG (H&L), HRP	Cell Signaling Technology	Cat#7076; RRID:AB_330924
Goat anti-Rabbit IgG (H&L), HRP	Bio-Rad	Cat#1721019; RRID:AB_11125143
Biological Samples		
Human: Primary Foreskin Keratinocytes	Laboratory of Ernst Reichmann	https://www.uniskingrafts.uzh.ch/en.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Alexa Fluor 488 Azide	Thermo-Fisher	Cat#A10266
Alexa Fluor 647 Azide	Thermo-Fisher	Cat#A10277
Alexa Fluor 488 NHS Ester	Thermo-Fisher	Cat#A20000
Alexa Fluor 647 NHS Ester	Thermo-Fisher	Cat#A20006
Sodium ascorbate	Sigma-Aldrich	Cat#A7631
Copper sulfate	Sigma-Aldrich	Cat#7758-98-7
5-ethynyl uridine (5-EU)	Baseclick Gmbh	Cat#BCN-003
DAPI (4',6-diamidino-2-phenylindole, dihydrochloride)	Thermo-Fisher	Cat#D1306
Dextran, Alexa Fluor 488; 10,000 MW, Anionic, Fixable	Thermo-Fisher	Cat#D22910
FluoSpheres NeutrAvidin-Labeled Microspheres, 0.04 μm, red fluorescent (580/605)	Thermo-Fisher	Cat#F8770
EZ-Link Sulfo-NHS-LC-Biotin	Thermo-Fisher	Cat#21335
Auxin (IAA)	Sigma-Aldrich	Cat#I3750
Triptolide	Adipogen Life Sciences	Cat#AG-CN2-0448
AZD4573	Selleckchem	Cat#S8719
CX-5461	Axon Medchem	Cat#2173
Roscovitine	Santa Cruz Biotechnology	Cat#sc-24002
Cycloheximide	Sigma-Aldrich	Cat#C7698
Epidermal growth factor (EGF)	Sigma-Aldrich	Cat#01-107
Insulin	Sigma-Aldrich	Cat#I1882
Hydrocortisone	Sigma-Aldrich	Cat#H0888
Cholera toxin	Sigma-Aldrich	Cat#C8052
Fibronectin	Sigma-Aldrich	Cat#F0895
Fetal bovine serum	Sigma-Aldrich	Cat#F7524
Horse serum	Gibco	Cat#16050122
DMEM, high glucose	GIbco	Cat#41965062
DMEM/F12	Gibco	Cat#11330032
OptiMEM Reduced Serum Medium	Gibco	Cat#31985070
CnT-57 media	CELLnTEC	Cat#CnT-57
Penicillin/Streptomycin	Gibco	Cat#15140122
Lipofectamine RNAimax transfection reagent	Invitrogen	Cat#13778150
Lipofectamine 2000	Invitrogen	Cat#11668030
16% Paraformaldehyde	Electron Microscopy Sciences	Cat#EMS-15710
Intercept blocking buffer	Li-Cor	Cat#927-70001
Bovine serum albumin	Abcam	Cat#ab7475
Triton X-100	Sigma-Aldrich	Cat#T8787
Rat tail collagen I	Gibco	Cat#A1048301
RNAse A	Roche	Cat#10109142001
Formamide	Ambion	Cat#AM9342
20X SSC (saline sodium citrate), RNAse-free	Sigma-Aldrich	Cat#SRE0068
Dextran sulfate sodium salt	Sigma-Aldrich	Cat#D8906
ULTRAhyb™-Oligo	Invitrogen	Cat#AM8663
SDS, 20% Solution, RNase-free	Invitrogen	Cat#AM9820
RNaseOUT™ Recombinant RNase Inhibitor	Invitrogen	Cat#10777-019
SuperScript™ III Reverse Transcriptase	Invitrogen	Cat#18080044
SYBR™ Select Master Mix	Applied Biosystems	Cat#4472903

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
Cell Navigator Live Cell RNA Imaging Kit (Strandbrite RNA Green)	AAT Bioquest	Cat#22630
ViewRNA high-content screening assay kit	Affymetrix	Cat#QVP0011
ViewRNA Type 1 signal amplification kit 488nm	Affymetrix	Cat#QVP0202
ViewRNA Type 1 signal amplification kit 550nm	Affymetrix	Cat#QVP0202
ViewRNA Type 6 signal amplification kit 650nm	Affymetrix	Cat#QVP0202
ViewRNA Type 10 signal amplification kit 488nm	Affymetrix	Cat#QVP0202
Amersham ECL Western Blotting Detection Reagent	Cytiva (Merck)	Cat#RPN2109
Chemiluminescent Nucleic Acid Detection Module Kit	Thermo Scientific	Cat#89880
Deposited Data		
Genome-wide siRNA screen images, single-cell feature values and summaries	Müller et al. (2021); This paper; Image Data Resource (IDR)	idr0093; http://doi.org/10.17867/10000157
Secondary siRNA screen (EU) single-cell feature values and summaries	This paper; Mendeley Data	https://doi.org/10.17632/3v4bkmg92x.1
Secondary siRNA screen (Poly(A)-FISH / RNA StrandBrite) single-cell feature values and summaries	This paper; Mendeley Data	https://doi.org/10.17632/yfx32prktv.1
Tertiary siRNA screen (4i) single-cell feature values and summaries	This paper; Mendeley Data	https://doi.org/10.17632/w3bhntjwp6.1
Experimental Models: Cell Lines		
Human: HeLa, cervical cancer cell line (single-cell clone)	Battich et al. (2015)	Kyoto
Human: 184A1, breast epithelial cell line (single-cell clone)	Kramer and Pelkmans (2019)	ATCC CRL-8798
Human: HCT116:TIR1	Steven West (Eaton et al., 2018)	N/A
Human: HCT116:TIR1 XRN2-AID	Steven West (Eaton et al., 2018)	N/A
Human: HCT116:TIR1 DIS3-AID	Steven West (Davidson et al., 2019)	N/A
Human: HCT116:TIR1 EXOSC10-AID	Steven West (Davidson et al., 2019)	N/A
Human: MRC5	Jurgen Marteijn (Steurer et al., 2018)	N/A
Human: MRC5 RPB1-GFP	Jurgen Marteijn (Steurer et al., 2018)	N/A
Oligonucleotides		
siRNA Silencer Select Human Genome Wide Library V4	Thermo-Fisher	Cat#4397926
siRNA targeting sequences against selected human genes; Ambion Silencer Select	Thermo-Fisher	see Deposited Data
Negative control siRNA #1; Ambion Silencer Select	Thermo-Fisher	Cat#4390843
siRNA targeting KIF11; Ambion Silencer Select	Thermo-Fisher	Cat#s7902
siRNAs targeting PIM2; Ambion Silencer Select	Thermo-Fisher	Cat#s21749, s21750, s21751
siRNAs targeting SLC25A3; Ambion Silencer Select	Thermo-Fisher	Cat#s10427, s10428, s10429
siRNAs targeting NUDT4; Ambion Silencer Select	Thermo-Fisher	Cat#s22020, s22021, s22022
siRNAs targeting SBF2; Ambion Silencer Select	Thermo-Fisher	Cat#s37818, s37819, s37820
siRNAs targeting GRIP2; Ambion Silencer Select	Thermo-Fisher	Cat#s37445, s37446, s37447
siRNAs targeting LRRK2; Ambion Silencer Select	Thermo-Fisher	Cat#s42413, s42414, s42415
siRNAs targeting EXOSC3; Ambion Silencer Select	Thermo-Fisher	Cat#s27229, s27230, s27231
siRNAs targeting EXOSC5; Ambion Silencer Select	Thermo-Fisher	Cat#s224433, s32381, s32382
Poly(A) FISH probe: dT-30-Cy5	Microsynth AG	N/A
Poly(A) FISH probe: dT-30-Atto488	Microsynth AG	N/A
Type 1 ViewRNA FISH probes: COL4A1	Thermo-Fisher	Cat#VA1-10870-VC
Type 10 ViewRNA FISH probes: CSPG4	Thermo-Fisher	Cat#VA10-11247-VC
Type 1 ViewRNA FISH probes: CTCF	Thermo-Fisher	Cat#VA1-12430-VC
Type 1 ViewRNA FISH probes: DAB2	Thermo-Fisher	Cat#VA1-12305-VC
Type 1 ViewRNA FISH probes: EGFR	Thermo-Fisher	Cat#VA1-11736-VC
Type 1 ViewRNA FISH probes: EP300	Thermo-Fisher	Cat#VA1-12301-VC

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Type 1 ViewRNA FISH probes: EZR	Thermo-Fisher	Cat#VA1-12203-VC
Type 6 ViewRNA FISH probes: GLS	Thermo-Fisher	Cat#VA6-10620-VC
Type 1 ViewRNA FISH probes: HPRT1	Thermo-Fisher	Cat#VA1-11124-VC
Type 6 ViewRNA FISH probes: HPRT1	Thermo-Fisher	Cat#VA6-13194-VC
Type 10 ViewRNA FISH probes: HPRT1	Thermo-Fisher	Cat#VA10-10498-VC
Type 10 ViewRNA FISH probes: NCOA4	Thermo-Fisher	Cat#VA10-10779-VC
Type 1 ViewRNA FISH probes: PFKL	Thermo-Fisher	Cat#VA1-12251-VC
Type 1 ViewRNA FISH probes: RAB11FIP3	Thermo-Fisher	Cat#VA1-12035-VC
Type 6 ViewRNA FISH probes: RELA	Thermo-Fisher	Cat#VA6-3168253-VC
Type 1 ViewRNA FISH probes: RHEB	Thermo-Fisher	Cat#VA1-12093-VC
Type 1 ViewRNA FISH probes: SERPINB5	Thermo-Fisher	Cat#VA1-12247-VC
Type 1 ViewRNA FISH probes: STX6	Thermo-Fisher	Cat#VA1-12085-VC
Type 10 ViewRNA FISH probes: TERF2IP	Thermo-Fisher	Cat#VA10-10772-VC
Type 1 ViewRNA FISH probes: VCL	Thermo-Fisher	Cat#VA1-12204-VC
Type 1 ViewRNA FISH probes: dapB	Thermo-Fisher	Cat#VF1-11712-VC
Type 6 ViewRNA FISH probes: dapB	Thermo-Fisher	Cat#VF6-10407-VC
Biotinylated Oligo(dT) Probe	Promega	Cat#Z5261
Primers for RT-qPCR, see Table S2	Microsynth AG	N/A
Recombinant DNA		
pmEGFP-N1	Greig et al. (2020)	N/A
pmEGFP-N1 R-MCD0.1	Greig et al. (2020)	N/A
pmEGFP-N1 R-MCD0.2	Greig et al. (2020)	N/A
Software and Algorithms		
TissueMAPS v0.6.0	N/A	https://github.com/pelkmanslab/TissueMAPS
Ilastik v1.3	Sommer et al. (2011)	https://www.ilastik.org/
Rstudio v1.2	RStudio Team (2020)	https://rstudio.com/
R v3.6.3	R Core Team (2021)	https://www.R-project.org/
Cellpose	Stringer et al. (2021)	http://www.cellpose.org/
Popcon	This paper; from (Snijder et al., 2012)	https://github.com/scottberry/popcon
Cytoscape	Shannon et al. (2003)	https://cytoscape.org/
Other		
FluoroDish, 35mm	World Precision Instruments	Cat#FD35
RTV615 Clear FDA Potting Silicone; Momentive	Techsil	Cat#MOSI17098
384-well μClear plates	Greiner Bio-One	Cat#781092
96-well μClear plates	Greiner Bio-One	Cat#655097

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Lucas Pelkmans (lucas.pelkmans@mls.uzh.ch).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Image data, single-cell level feature data and summary data for the genome-wide screen have been deposited at the Image Data Resource (IDR) under accession number idr00093, and are described in an accompanying manuscript (Müller et al., 2021). Single-cell-level features and summaries for other genetic perturbation experiments are provided at the locations listed in key resources table. Raw image datasets for other experiments are available upon reasonable request.

- Image analysis was performed using TissueMAPS, an open-source project for high-throughput image analysis developed in our group. All analysis modules, including code written for this paper, is packaged together with the main repository at https://github.com/pelkmanslab/TissueMAPS (https://doi.org/10.5281/zenodo.6468707). An example analysis pipeline description with module files containing parameter settings for the genome-wide screen is provided at IDR (idr00093). This example is a typical TissueMAPS workflow, however the exact modules used and parameter values differ depending on experiments. These can be provided on request. Code for calculating population-context features was written in python based on previous MATLAB code from our group (Snijder et al., 2012) and is available at https://github.com/scottberry/popcon (https://doi.org/10.5281/zenodo.6468575) R code for the analysis of single-cell data derived from images was developed on a per-experiment basis and is available on request.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and culture conditions

HeLa Kyoto (female) cell populations were derived from a single-cell clone and were tested for identity by karyotyping (Battich et al., 2015). HeLa cells were cultured in high glucose DMEM supplemented with 10% fetal bovine serum (FBS) and 1% GlutaMAX. Cells with low passage number (2-6) were used for all experiments.

184A1 (human female breast epithelial) cell populations were derived from a single-cell clone, and were used at low passage number (2-6) for all experiments. 184A1 cells were cultured in DMEM/F12 media supplemented with 5% horse serum, 20ng/ml epidermal growth factor, 10μg/ml insulin, 0.5μg/ml hydrocortisone, 10ng/ml cholera toxin.

Keratinocytes were donated by a healthy 2.5-year-old male (Biedermann et al., 2010), isolated and kindly provided by E. Reichmann and L. Pontiggia (University of Zurich, Zurich). Keratinocytes were cultured in CnT-57 medium (CELLnTEC) supplemented at 1:100 (v/v) with Penicillin/Streptomycin. For propagation and experiments, plastic cell culture plates were coated with rat-tail collagen I, according to manufacturer's instructions. Experimental plates were prepared 5 days after thawing cells (passage 1).

Parental MRC-5 and derived POLR2A-GFP (RPB1-eGFP) cells (Steurer et al., 2018) were cultured in DMEM/F12 supplemented with 10% FBS and 1:100 (v/v) Penicillin/Streptomycin.

Parental HCT116:TIR1 and derived EXOSC10-AID, DIS3-AID, XRN2-AID cells (Davidson et al., 2019; Eaton et al., 2018) were cultured in DMEM supplemented with 10% FCS and 1:100 (v/v) Penicillin/Streptomycin. All cells were tested for absence of mycoplasma before use and grown at 37°C, 95% humidity and 5% CO_2.

Unless otherwise specified, cells were grown and imaged in uncoated Greiner μClear plastic-bottom 96- or 384-well plates.

METHOD DETAILS

siRNA transfection

Transfection with siRNA was performed as previously described (Berchtold et al., 2018). Briefly, 900 HeLa or 1500 MRC5 POLR2A-GFP cells were plated per well in 384-well plates for reverse transfection onto a mixture of pooled siRNAs (5 nM final concentration) and Lipofectamine RNAiMAX (0.08μl per well in OptiMEM) according to manufacturer's specifications. Cells were subsequently grown for 72 hours at 37°C in a final volume of 50μL growth media, to establish efficient knockdown of the targeted genes. For the genome-wide and secondary screens, siRNAs were dispensed in a much lower volume using an acoustic dispenser, and assays were performed in a final volume of 40μL. A detailed protocol was described previously (Müller et al., 2021).

The genome-wide screen was performed using a pool of 3 siRNAs (Ambion) targeting each gene. 21538 siRNA pools were assayed across 83 384-well plates, as previously described (Müller et al., 2021). The secondary library (EU metabolic labelling, poly(A) FISH, RNA Strandbrite) employed the same siRNA pools for 436 gene perturbations. Secondary assays were performed in duplicate. The tertiary library (4i, smFISH) consisted of a set of 63 perturbations, mostly from the secondary library. Gene lists and other metadata are provided together with the experimental results at the locations listed in key resources table.

Plasmid transfection

2700 HeLa cells were plated in each well of a 384-well plate. 12 hours after seeding, 50ng plasmid was transfected per well using Lipofectamine 2000, according to manufacturer's protocol. Cells were then incubated for 2 hours at 37°C and then washed three times into fresh media to remove unbound transfection reagent/plasmid complexes. Cells were subsequently grown for 48 hours at 37°C before fixation, staining and imaging.

Chemical treatments

RNA polymerase I inhibitor, CX-5461 (Drygin et al., 2011) was dissolved in 5mN HCl at a concentration of 5mM and used at 2μM. XPB (TFIIH) inhibitor triptolide (Titov et al., 2011) was dissolved in DMSO at a concentration of 10mM and used at 2μM. CDK9 inhibitor AZD4573 (Cidado et al., 2020) was dissolved in DMSO at a concentration of 10mM and used at 0.1μM. CDK inhibitor roscovitine

was dissolved in DMSO to a final concentration of 10mM and used at 20μM. Cycloheximide was dissolved in H20 to a concentration of 40mM and used at 1μM. Auxin was dissolved in ethanol at 200mM and used at 0.5mM. Durations of chemical treatments vary, and are noted throughout the text and figure legends.

Image acquisition

Unless otherwise specified, all imaging was performed an automated spinning-disk microscope (CellVoyager 7000, Yokogawa), which is equipped with four excitation lasers (405, 488, 568, 647nm) and two Neo sCMOS cameras (Andor). For cell volume measurement and smFISH experiments, a 40X/NA0.95 air objective was used. For other experiments such as the genome-wide screen, secondary screen, poly(A) FISH, RNA Strandbrite, 4i, and R-MCD-GFP experiments, a 20X/NA0.75 objective was used. Certain examples images (e.g., Figure 1D) were acquired with a 60X/NA1.27 objective. With the exceptions of cell volume experiments and 3D smFISH, images were maximum-projected during acquisition. Images presented in the same figure for the same stain were always identically rescaled.

DNA and protein staining

Nuclear DNA was stained using 4',6-diamidino-2-phenylindole, dihydrochloride (DAPI) for 5-10 minutes at a final concentration of 0.4μg/mL in phosphate buffered saline (PBS). Total protein was stained using Alexa Fluor 488 NHS Ester or Alexa Fluor 647 NHS Ester (succinimidyl ester) for 10 minutes at a final concentration of 0.2μg/mL in 50mM carbonate-bicarbonate buffer pH 9.2.

High-throughput cell volume measurement

To measure cell volume in fixed cell populations, we attached fluorescent beads to the upper surface of cells, and also the slide surface (Figure S1B). We then determine the three-dimensional positions of these beads using spinning-disk confocal microscopy and use these to generate a computational reconstruction of the cell.

All staining and washing steps were performed either manually, or on a semi-automated liquid handling platform (BioTek) using the following method: cells were fixed with 4% PFA for 30 min and washed four times with phosphate-buffered saline (PBS). To avoid beads entering inside cells, we performed biotinylation and bead attachment before cell permeabilization. To biotinylate cell surface, EZ-Link Sulfo-NHS-LC-Biotin was freshly dissolved in PBS and added to cells at a final concentration of 0.25mg/mL for 5 min. Cells were then washed four times with PBS. To prepare beads for attachment, we diluted 40nm-diameter streptavidin-coated fluorescent beads (Fluospheres) into a buffer containing 0.5x PBS and 0.01% Triton X-100, to a concentration of 0.005% solids (1/200 from 1% stock). Beads were then dispersed by sonication in 1mL aliquots for 3 x 30 s in a Bioruptor water-bath sonicator (Agilent), and were then added to cells by adding an equal volume of bead suspension to the residual PBS in multiwell plates (final bead concentration 0.0025% solids). Bead suspensions were mixed in the wells by pipetting up and down, or brief vigorous shaking of plates and were then incubated at room temperature for 10 min before washing off unbound beads with PBS. Cells were then permeabilised before proceeding with other stains.

After additional staining, beads were imaged as a separate acquisition (excitation laser: 568nm, emission filter 590/20 nm) using a 40X/NA0.95 (air) or 60X/NA1.27 (water immersion) objective on a spinning-disk microscope (CellVoyager 7000, Yokogawa), equipped with sCMOS cameras (0.1625μm or 0.10833 μm pixel dimensions, for 40X/60X objectives, respectively). Approximately 60 confocal Z-slices were obtained per imaging site with step sizes of 0.25μm or 0.33μm. Calculation of cell volume from bead images is described in quantification and statistical analysis.

Fluorescence exclusion method (FXm)

FXm was performed as previously described (Cadart et al., 2017). Briefly, a PDMS chamber with height 21.8μm was prepared using an epoxy mould provided by L. Venkova and M. Piel (Institut Curie, Paris, France). The chamber was then attached to a 35mm FluoroDish using plasma treatment, and internal surfaces were coated with 50ug/mL human fibronectin for 1 hour at 37°C. Chambers were then incubated in cell growth media overnight, and washed once more with growth media. Cells were then trypsinised and loaded as a suspension into the FXm chamber. After allowing cells to adhere for 4 hours, media was replaced with fresh media containing 1 mg/mL Dextran-Alexa Fluor 488 (10,000 MW). FXm measurements were then performed on live cells at 37°C, 5% CO_2 using an epifluorescence microscope (VisiTIRF, Visitron) with a 20X/0.4NA objective lens (Nikon). Fluorescently labelled dextran was excited with 450-490nm (SOLA lightengine) and emission collected from 500-550nm. Data analysis was performed using custom software as previously described (Cadart et al., 2017). For comparison with cell volume reconstruction method, cells were fixed in the chamber using 4% PFA and bead attachment was performed manually, as described above. Chamber-level overviews were derived by computationally stitching images from both methods, and the same single cells were identified in both datasets by aligning these overviews.

Single-molecule RNA fluorescent in situ hybridisation (smFISH)

Single-molecule FISH was performed as previously described (Battich et al., 2013), using ViewRNA high-content screening assay kit (Affymetrix) with Type 1 (Figure 1) or Type 1, 6, and 10 (Figure S9) primary probes and signal amplification kits (Affymetrix). Samples were imaged on a spinning-disk microscope using a 40X/NA0.95 air objective with step-size of 1.0μm. Computational identification of spots was performed in 3D (Figure 1, with cell volume measurement), or in 2D from maximum-projected images (Figure S9, genetic perturbation screen), as previously described (Stoeger et al., 2015).

For smFISH experiments in the (tertiary) genetic perturbation screen (Figure S9), we focused on nine genes (*RHEB*, *STX6*, *CTCF*, *GLS*, *RELA*, *HPRT1*, *CSPG4*, *NCOA4*, *TERF2IP*). These genes are expressed throughout the cell cycle, encompass diverse biological functions, and have a range of estimated mRNA half-lives (2.3h to >24h (Tani et al., 2012)). They were also previously shown to have cytoplasmic mRNA abundance that scales with cell size (Figures 1A–1D and S1J–S1L) (Battich et al., 2015). The genes targeted by siRNA in this experiment are shown in Figure S9E. Only one of the nine genes (*HPRT1*) had mean mRNA abundance that correlated with EU incorporation across perturbations (Figures S9J and S9K). *HPRT1* was typically down-regulated in perturbation conditions with reduced EU and up-regulated in perturbations with increased EU – a trend consistent with its biological function in nucleotide metabolism.

In situ metabolic labelling of nascent RNA

Nascent RNA was visualised using metabolic labelling as previously described (Jao and Salic, 2008), with modifications. Briefly, adherent cells were cultured in complete media at 37°C, 5% CO_2 for 2-3 days. 5-ethynyl uridine (EU) was then dissolved to a concentration of 2mM in pre-warmed complete media. EU was added to cells by partially aspirating growth media and dispensing an equal volume of 2mM EU using a BioTek washer-dispenser (e.g. 30μL 2mM EU added to 30μL residual for 384-well plates: final EU concentration = 1mM). Cells were then incubated for 20 or 30 min at 37°C, 5% CO_2, before fixation with 4% PFA at room temperature for 20-30 min. After fixation, cells were permeabilised with 0.5% Triton X-100 and washed 3 times with TBS (50mM Tris pH 8.0, 150mM NaCl).

To render nascent RNA fluorescent, we prepared sufficient volume of click reaction master mix for all wells on a plate at 1.5x concentration, as follows: 75μM Alexa Fluor 488 azide or Alexa Fluor 647 azide, 3mM CuSO4, 150mM Sodium ascorbate, in TBS. Click reaction was dispensed using a BioTek washer-dispenser and incubated for 30 min at room temperature before washing cells 3x into PBS. The click reaction is detrimental to the intensity of several fluorophores including the fluorescent beads used for cell volume measurement and also GFP. To combine these with EU metabolic RNA labelling, we fixed, permeabilised and imaged either GFP or beads before click reaction. Images were subsequently aligned using the DAPI channel, which was included in both imaging rounds, using the computational procedure described previously for 4i (Gut et al., 2018).

Immunofluorescence and iterative indirect immunofluorescence imaging (4i)

For PCNA staining in non-4i immunofluorescence experiments, and for the genome-wide screen, cells were blocked with 1% BSA in PBS for 1 hour and then incubated for 2 hours with rabbit anti-PCNA antibody dissolved in 1% BSA in PBS. RNA Pol II immunofluorescence in non-4i immunofluorescence experiments was performed using Intercept blocking buffer (LI-COR Biosciences) for blocking and antibody incubations.

Immunofluorescence-based quantification in combination with *in situ* RNA metabolic labelling across multiple imaging cycles was validated using MRC5 POLR2A-GFP cells (Steurer et al., 2018). We first pulsed cells with EU, then fixed and imaged GFP and DAPI. Subsequently, we performed click reaction and POLR2A immunofluorescence and then re-imaged. Images from the two rounds were aligned to the DAPI channel, using the computational procedure described previously for 4i (Gut et al., 2018). We then quantitatively compared POLR2A immunofluorescence with GFP intensity (Figures S10A and S10B): IF and GFP were highly correlated (r = 0.92) and both showed similar partial correlation with EU ($r_{x,\mathrm{EU}|\text{ nuclear area}}$ of 0.49 and 0.46, respectively).

4i was performed as previously described (Gut et al., 2018) with two modifications: Intercept blocking buffer (LI-COR Biosciences) was used for all blocking, primary and secondary antibody incubations, and 50mM HEPES was included in imaging buffer – which was adjusted to a pH of 7.4. Before 4i experiments, all antibodies were tested for compatibility with elution buffer using the following criteria: similar staining on normal and elution-buffer treated cells, minimal residual signal after elution and re-staining with secondary antibody. After 4i, we also validated that agreement between replicates was high (Figure S10E) and that the technical variability between control (scrambled siRNA) wells was much less than the differences induced by the perturbations (Figure S10F). To ensure successful antibody elution in each cycle, we included elution controls in each imaging cycle. This consists of re-probing a test well with a secondary antibody in the imaging cycle after it was stained with primary and secondary antibody and imaged. Efficient elution was verified in all cases.

Poly(A) RNA fluorescent in situ hybridisation (FISH)

Poly(A) FISH was adapted from published smFISH protocols (Raj et al., 2008) for high-throughput liquid handling. Briefly, cells were fixed for 15 min in 4% paraformaldehyde, permeabilised in 0.2% Triton X-100 for 15 min (secondary screen) or 70% EtOH for 1 hour (HCT116 experiments). If required, click reaction was then performed before applying FISH probes. Cells were then washed into 2X saline sodium citrate buffer (SSC) containing 10% (v/v) formamide. Before hybridisation, samples were transferred to hybridisation buffer (2X SSC, 10% (v/v) formamide, 100mg/mL dextran sulfate), and pre-incubated for 1 hour at 37°C. Fluorescently labelled DNA oligonucleotide probes, dT(30)-Atto488 or dT(30)-Cy5, were purchased from Microsynth as labelled and HPLC-purified. These were diluted into the hybridisation buffer and applied to cells at a final concentration of 400nM. After 16 hours at 37°C in a rotating incubator, cells were washed into 2X SSC, 10% formamide, incubated again at 37°C for 1 hour, before washing into 2X SSC. For combining with subsequent immunofluorescence, poly(A) FISH was imaged before immunostaining. Images were aligned using the DAPI channel, which was included in both imaging rounds, using the computational procedure described previously for 4i (Gut et al., 2018).

Strandbrite total RNA staining

RNA Strandbrite was used according to manufacturer's instructions. We confirmed that RNA Strandbrite staining is specific for RNA by incubating cells with 0.5mg/mL RNAse A at 37°C for 30 min (Figure S1N). We also confirmed that RNA Strandbrite staining is proportional to RNA abundance measured by extracting and quantifying RNA from a fixed number of cells (Figures S1O and S1P). RNA Strandbrite green is spectrally distinct from poly(A) FISH when using Cy5-labelled oligo(dT) probes, so the two assays were combined. However, during protocol optimisation, we found that RNA Strandbrite staining was not stable during the poly(A) FISH protocol, showing an altered localisation pattern and reduced signal intensity after FISH. However, poly(A) FISH was not affected by prior RNA Strandbrite staining. To combine the two assays, we therefore performed RNA Strandbrite staining and imaging before poly(A) FISH. Images were aligned using the DAPI channel, which was included in both imaging rounds, using the computational procedure described previously for 4i (Gut et al., 2018).

RNA extraction and quantification

Cells were grown in 24-well plates and collected using 0.05% Trypsin-EDTA 72h after reverse transfection, using the same siRNA and RNAimax concentrations as in screening assays. Trypsinised cells were resuspended in DMEM and counted on CASY cell counter (OLS OMNI Life Sciences GmbH). Equal number of cells for all knockdown conditions and scrambled siRNA controls were pelleted for RNA extraction. Total RNA extraction was performed with Quick-RNA MicroPrep Kit (Zymo Research Europe GmbH). On-column DNase I treatment was performed to prevent genomic DNA contamination. RNA was resuspended in DNAse/RNAse-Free water and the concentration of all samples measured by UV absorbance on Nanodrop (Thermo Fisher Scientific). RNA purity was checked by measuring A260/A280 & A260/A230 > 1.8. Experiments were performed in triplicate.

Quantitative reverse transcription PCR (RT-qPCR)

50ng total RNA was used for cDNA synthesis using SuperScript™ III Reverse Transcriptase, following manufacturer's specifications. A pool of all reverse PCR primers was used for cDNA synthesis. RT-qPCR was performed in 96-well plates using SYBR™ Select Master Mix on a QuantStudio 3 Real-Time PCR System (Applied Biosystems). The following conditions were used: 50°C for 2 minutes, 95°C for 2 minutes, followed by 50 cycles at 95°C for 15 seconds and 60°C for 1 minute. 3 independent experiments were performed. All primers for RT-qPCR were designed using PrimerBLAST (Ye et al., 2012) (Table S2). They were designed so that one primer of the pair spans a splice site, so they should not amplify genomic DNA. As expected, reverse transcription reactions without reverse transcriptase enzyme did not generally lead to successful amplification in qPCR. In a few cases amplification was detected, however because it occurred more than 10 cycles after the corresponding cDNA sample, the effect on quantification was negligible.

RNA dot blot

RNA samples quantified using UV absorbance (Nanodrop) and then diluted in DNAse/RNAse-free water to achieve 5, 10 and 20 ng of total RNA in 3 μL total volume per tube. All samples were incubated at 95°C in a heat block for 3 min to disrupt secondary structures followed by immediately cooling in ice for 3 minutes. 3 μL of RNA was blotted onto the Hybond®-N+ hybridization membranes (RPN203B, Cytiva) and cross-linked using a UV crosslinker (Stratagene Stratalinker) at 120,000 microjoules/cm2. To test specificity for RNA, an RNA sample from scrambled siRNA control cells was treated with RNAse A for 5 mins at room temperature before heating to 95°C and adding RNaseOUT Recombinant RNase Inhibitor. After UV crosslinking, blots were prehybridized in ULTRAhyb™-Oligo hybridization buffer for 30 minutes at 42°C, followed by overnight incubation at 37°C with biotinylated Oligo(dT) probe in hybridization buffer. Oligo-dT probe stock (50pmol/μl) was diluted 1:10,000 in the hybridization buffer before adding to the blot. Blots were then washed thrice for 30 mins each at 42°C with wash buffer containing 2xSSC and 0.5% SDS, followed by 2 washes with 2x SSC at room temperature for 15 mins each. Signal was revealed using Chemiluminescent Nucleic Acid Detection Module Kit according to manufacturer specifications. Images were adjusted for contrast using ImageJ. Three independent experiments were performed.

Immunoblotting

HCT116-AID tagged cell lines were grown in 6 well culture dishes (TPP, Switzerland). Cells were treated with Auxin (dissolved in ethanol) or ethanol carrier control for one hour. Cells were washed thrice with ice-cold PBS before adding lysis Buffer (150 mM NaCl, 50 mM HEPES, 1% Triton X-100, 0.1% SDS, 2 mM DTT, 5 mM EDTA and 2X protease inhibitor cocktail dissolved in Milli-Q H_2O). Next, cells were lysed by scraping using cell scraper, followed by four rounds of passing through an Omnican Insulin-50 syringe (Braun, Switzerland). Lysates were incubated for 15 min on ice and then centrifuged at 21,000*g* for 15 min at 4 °C. The supernatant was collected and the proteins in the lysates were denatured by the addition of 2X loading buffer and boiling at 90 °C for 10 min. Proteins were resolved by 4-20% Mini-Protean TGX protein gel (Bio-Rad) gel and transferred onto a polyvinylidene difluoride (PVDF) membrane (Immobilon-P, 0.45 μm, Millipore) using Trans-blot transfer system (Bio-Rad). Membranes were blocked with blocking buffer comprising of 5% low-fat milk in 1× PBS-T (1× PBS with 0.1% Tween-20) for 1 h at room temperature, followed by overnight incubation with primary antibodies in blocking buffer at 4 °C, and then incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies in blocking buffer for 1 h at room temperature. Signal was revealed using chemiluminescent reagent (Merck). The following antibodies were used: AID (1:300), DIS3 (1:1000) and Actin (1:5000). Images were adjusted for contrast using ImageJ. Three independent experiments were performed.

QUANTIFICATION AND STATISTICAL ANALYSIS

Image processing

The majority of image analysis was performed in TissueMAPS, using previously developed methods (Stoeger et al., 2015). All images were corrected for illumination artefacts as previously described (Stoeger et al., 2015). In some cases, we also developed new image analysis strategies, or used pixel classification (Sommer et al., 2011) or neural networks (Stringer et al., 2021) to aid in segmentation tasks. This is indicated where applicable.

Nuclear and cell segmentation

Nuclei were segmented by adaptive thresholding of the DAPI or H2B signal. After separating clumped nuclei and excluding small objects, cells were then segmented using total protein stain (succinimidyl ester), using these nuclei as seeds for a watershed, as described previously (Stoeger et al., 2015). The analysis pipeline for the genome-wide screen is described in (Müller et al., 2021). For MRC5 cells and the 4i experiment, we used the cellpose generalist nuclear segmentation neural network (Stringer et al., 2021) without additional training because it yielded superior results for separating touching nuclei. However, we found that nuclei identified by cellpose were not always completely contained within the DAPI-positive regions of the image, or did not completely fill the nucleus. Cellpose nuclei were therefore further refined using adaptive thresholding and/or watershed expansion using the DAPI signal.

Nucleolar segmentation

The nucleolus was segmented from DAPI and Alexa Fluor 647-succinimidyl ester images using pixel classification in Ilastik (Sommer et al., 2011), as depicted in Figure S2L and described previously (Müller et al., 2021).

Cell volume measurement

To perform computational 3D cell reconstruction, cells were first segmented in two dimensions using procedures described above. We then localised beads in 3D, using an approach previously developed for smFISH (Raj et al., 2008; Stoeger et al., 2015). The brightest voxel of each segmented bead in the original image was taken as the bead centre. Next, the slide surface was estimated by fitting a plane in 3D through all beads that were detected outside of cells. The 3D positions of beads are then recalibrated relative to the slide surface, so that the list of (x,y,z) coordinates for each cell represents the height of the upper cell surface above the slide. To reconstruct the cell surface, outlier beads were excluded by fitting a 3D alpha shape (Da et al., 2022) to the list of coordinates for each cell, as well as a set of points at the 2D cell periphery, which are computationally anchored to the slide surface. The upper cell surface was then linearly interpolated using the remaining points. Code is written in python, making use of numpy (Harris et al., 2020) and scipy (Virtanen et al., 2020), and was integrated into TissueMAPS. We directly compared cell volumes estimated with this method with those obtained by fluorescence exclusion method (FXm) (Cadart et al., 2017) finding good agreement at the single-cell level (Figures S1B–S1E). We also found that cell volumes correlated well with other measures of cell size, such as nuclear and cell area, and total protein levels, visualised with succinimidyl ester staining (Figure S1F). Both protein content and nuclear area were approximately proportional to cell volume in HeLa cells (Figure S1G) and are therefore excellent proxies for cell volume in unperturbed cells. Correlations of volume with nuclear area were slightly higher for HeLa than for "flatter" 184A1 cells and keratinocytes (Figure S1H), which often form long projections.

Chromatic aberration correction – 4i

Chromatic aberrations were evident when comparing different imaging channels in the 4i experiment. To correct these, we imaged multicolor fluorescent beads and used these to fit parameters of an affine transformation aligning each channel to the 568nm channel. These channel-specific transformations were then applied to all images for the 405nm, 488nm and 647nm channels, using a custom procedure developed in python making use of scipy (Virtanen et al., 2020), numpy (Harris et al., 2020) and scikit-image (van der Walt et al., 2014).

Data cleaning

After cell segmentation, we trained supervised machine-learning models (support-vector machines) to exclude certain acquisition and segmentation artefacts using the TissueMAPS framework. This interactive procedure is similar in nature to previous software developed in the lab, CellClassifier (Rämö et al., 2009). Briefly, it involves manually selecting cells with certain properties as training data for a classifier, visualising the results of the classification, and selecting additional examples to refine the classification (Figure S2A). Throughout this work, we excluded border cells (those with pixels touching the image boundary), mitotic/apoptotic cells (identified using DAPI texture features), cells with cytoplasmic DAPI, polynucleated cells, and all cells from sites that were not in focus. In the genome-wide screen, we also trained a classifier to exclude mis-segmented cells (Müller et al., 2021).

Additional data cleaning – cell volume measurement

Cell volume measurement involves computationally fitting a plane in 3D to define the slide surface. When cells were too densely packed, this cannot be achieved accurately, preventing accurate determination of cell volume. We therefore did not compute cell volumes for these sites, which were automatically identified based on the proportion of the image surface identified as unoccupied. We also excluded cells with insufficient bead density, or those which were not fully captured in the confocal volume. These were identified and removed using manually defined thresholds during exploratory data analysis.

Additional data cleaning – smFISH

smFISH probes were dispensed into 384-well plates using automated liquid handling, as previously described (Battich et al., 2013). Inaccuracies in handling of small volumes led to a 'quadrant effect' in which the upper left well in each group of 4 wells contained

smFISH spots with weaker intensity. Spots in these wells could not be unambiguously distinguished from background, so these were excluded from further analysis. This reduced the number of perturbations studied from the target of 63 to 50.

Additional data cleaning – 4i

The 4i perturbation experiment contained 160 wells and 18 antibodies (2880 antibody/well combinations). To identify outlier wells, we compared mean intensity measurements for each well and each antibody. Where replicates differed in any stain by more than 30%, the quantitative data and underlying images were manually examined. In most cases, the discrepancy was due to cell density differences between replicate wells, and all data was retained. In nine cases, we identified that staining was uniformly lower in one replicate, likely due to errors in automated liquid-handling. These antibody/well combinations (0.3% of 2880) were excluded. For multivariate analysis (e.g. UMAP), affected wells were entirely excluded, resulting in a single replicate being used for 6 of the 66 siRNA conditions studied: CBX3, KDM4D, POLR3E, PGAP3, RAB5C, SNRPF.

Additional data cleaning – plasmid transfection experiments

In agreement with previous work, we found that plasmid DNA could be detected in the cytoplasm of transfected cells using poly(A) FISH (Greenberg et al., 2019). This was prominent at 12h and 24h after transfection, but was reduced after 48h, likely due to asymmetric partitioning during cell division (Wang et al., 2016). To identify cytoplasmic DNA foci, we used Ilastik-based pixel classification, with poly(A) FISH and DAPI as a two-channel input. We excluded all cells containing these foci.

Background subtraction

Quantitative intensity values for stains such as DAPI, succinimidyl ester, RNA Strandbrite, and poly(A) FISH were background subtracted using a constant value measured in a region outside the objects. For immunofluorescence measurements, background values were taken from a region within the cell for a well in which the primary antibody was omitted. For EU, background values were taken from a well that was not incubated with EU but was subject to click reaction. Since click reaction staining in the absence of EU showed a similar staining pattern to succinimidyl ester, we used a linear modelling approach to predict the expected background based on succinimidyl ester intensity. This slightly outperformed a constant-value background subtraction approach.

Data normalisation

In some cases, we observed variation in staining intensity for different rows of the plate. These inconsistencies arise because of the small differences in reagent volumes delivered by the automated liquid handling system, which operates row-wise. This was assessed routinely during data analysis by manual inspection of intensity distributions after background subtraction. When necessary, we multiplied all background-subtracted intensity values by a correction factor to equalize the medians of the rows. Where possible, in genetic perturbation experiments, only negative control (scrambled siRNA) wells were used for normalisation. The same correction strategy was used to correct for plate-to-plate variation in multi-plate experiments.

Additional normalisation – Poly(A) FISH

DNA probes for poly(A) FISH and also Strandbrite reagents were dispensed using a 96-well pipette head into 384-well plates. This resulted in a quadrant-specific artefact in intensity that was corrected in the same way as described for row- and plate-corrections described above.

Additional normalisation - 4i

Incubation in 4i imaging buffer (Gut et al., 2018) is detrimental to signal intensity for some antibodies. This is observed as a reduction in intensity over the time of image acquisition (12-16 hours per imaging cycle). We corrected for this artefact by fitting an aympototic regression model (SSasymp function in R) to the median signal intensity of all scrambled siRNA control cells as a function of acquisition time (Figures S10C and S10D). Fitting was achieved using the nlsLM function from the minpack.lm package (Elzhov et al., 2016), after removing outlier control wells. We then multiplied all values by a correction factor derived from this curve and used it to multiplicatively scale all data based on acquisition time. This resulted in excellent agreement between intensity values of replicate (non-control) wells, which were typically imaged 4-5 hours apart (Figure S10E).

We also observed smaller well-specific staining biases that could not be explained by time-dependent intensity reduction, but were also not clear technical outliers. These represent technical variation inherent to the experiment. Because these can lead to biases when computing correlations among populations of single cells from different wells, it was sometimes necessary to correct these. This was achieved by multiplicative centering of well medians between experimental duplicates. To ensure that this did not bias our results, we verified that correlation coefficients across the combined populations were similar to those seen within individual wells (Figure S12D). Importantly, this correction was only applied to allow pooling cells from multiple wells and was therefore not performed whenever values from individual wells were summarized and considered as replicates (for example for fold-change calculations, or correlations between well-averages).

Cell cycle classification

Mitotic cells do not incorporate EU, and were therefore not of interest for this work. We first identified and removed these as described in 'Data cleanup' (Figure S2A). G1 and G2 cells can be distinguished based on DNA content measured from DAPI staining intensity. However, identifying S-phase cells requires more information. The gold standard for identifying cells undergoing DNA replication is to use 5-ethynyl-2'-deoxyuridine (EdU) metabolic labelling, which is similar to EU labelling except that cells are pulsed with EdU instead of EU to label newly synthesized DNA. Because EU and EdU detection rely on the same click chemistry, they cannot be combined on the same sample. We therefore included several wells of EdU in our RNA metabolic labelling experiments, and used

PCNA immunofluorescence as an S-phase marker. PCNA is localised to DNA replication foci and has a characteristic punctate localisation in S-phase cells. By performing PCNA immunofluorescence on all wells, we could identify S-phase cells using EdU in the EdU-treated wells and train a classifier based on PCNA to identify S-phase cells also in the EU-treated wells (Figure S2B).

More specifically, after background subtraction and feature normalisation as described above, we trained a random forest classifier using the randomForest package in R (Liaw and Wiener, 2002), to classify S-phase cells using PCNA and DAPI texture features. We used thresholded EdU intensity as a ground truth for cells that are undergoing DNA replication (Figure S2C). Classifiers were typically 95-98% accurate (Müller et al., 2021) (Figure S2D). Remaining non-S-phase cells were classified as G1 or G2 based on DNA content (Figures S2E and S2F). In experiments with genetic perturbations, we included siRNA knockdown of *GRIP2*, *SBF2*, *NUDT4* and scrambled siRNA controls in the EdU training data to ensure robustness to cell size and morphology perturbations, as described and validated in the accompanying manuscript (Müller et al., 2021).

Quantitative data analysis

After extracting quantitative features from images, quantitative analysis for individual experiments was done in R (R Core Team, 2014), relying heavily on tidyverse packages (Wickham et al., 2019).

smFISH

Linear regression to predict spot count from cell volume was performed in R using the lm function with model $n_{spots} = a + bV$. Model fits were quantified using the coefficient of determination (R^2), calculated using $R^2 = 1 - RSS/TSS$, where the residual sum of squares, $RSS = \sum_i (y_i^{(observed)} - y_i^{(predicted)})^2$, and the total sum of squares, $TSS = \sum_i \left(y_i^{(observed)} - \frac{1}{n}\sum_j y_j^{(observed)}\right)^2$ for n observations.

Following (Padovan-Merhar et al., 2015) we quantified the fraction of predicted mRNA abundance that is volume independent ($a/(a+bV)$) or volume-dependent ($bV/(a+bV)$) for all genes measured. This is only interpretable for $a,b \geq 0$, so we omitted genes where a was significantly less than zero, and set $a = 0$ for intercepts that were slightly negative but not significantly ($p > 0.05$) different from zero.

Estimation of RNA degradation rates

Cells were grown in 384-well plates and were treated with 10μM triptolide for 0, 1.5, 2.5, 4.5h before fixation. bDNA smFISH was used to measure transcript abundance at all time points for 39 genes (*ABL1*, *ACACA*, *APP*, *BIN1*, *COL4A1*, *CTCF*, *CTSB*, *DAB2*, *EGFR*, *EIF4E*, *EP300*, *ETS2*, *FGF2*, *GTF2B*, *HPRT1*, *KIF11*, *LAMP2*, *MDM2*, *MSLN*, *MYC*, *NOTCH2*, *NUP98*, *PEX19*, *PFKL*, *PIP5K1A*, *PPARG*, *PTEN*, *PTRF*, *RAB11A*, *RAB11FIP3*, *RB1*, *RHEB*, *SERPINB5*, *SOS1*, *SRPRA*, *STX6*, *TNFRSF12A*, *UBE2C*, *VCL*). Genes for which transcript counts did not decrease by at least 40% during the 4.5h triptolide treatment (21 of 39 genes) were not considered for further analysis. To estimate mRNA degradation rates for the remaining genes, cells were separated into size bins based on protein content (Figure S3B). We then fitted a linear model to log-transformed cytoplasmic transcript abundance ($\log(\text{spot count} + 1)$) as a function of time for each gene and each cell size bin (Figures S3C and S3D). The fit was performed using the lm function in R for all bins with an average cell count > 10. Slope values from these models are an estimate of RNA degradation rates. These showed no consistent dependence on cell size (Figure S3E).

As a second approach to estimating degradation rates from this data, we adapted a previously described method (Padovan-Merhar et al., 2015). Briefly, we used linear regression to predict the the cytoplasmic transcript abundance for each gene in non-inhibited cells (Figure S3F). This was done using the lm function in R, with protein content, cell area and DNA content as predictors (mean $R^2 = 0.51$). We then applied these models to transcriptionally inhibited cells, to estimate the number of transcripts expected at the single cell level in the absence of transcriptional inhibition (Figure S3G). This leads to two values for each cell: the measured spot count, $n_{\text{measured}}(t > 0)$, and the predicted spot count before transcriptional inhibition, $n_{\text{predicted}}(t = 0)$. If transcript abundance decays as a single exponential and cell size does not change during transcriptional inhibition these values are related by the following equation:

$$n_{\text{measured}}(t) = n_{\text{predicted}}(0)e^{-\lambda t}.$$

Where λ is the RNA degradation rate. Using this equation, we calculated λ and then examined how it varies with cell size (Figure S3H).

Quantitative reverse transcription PCR (RT-qPCR)

Relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). To normalize between samples, ΔCt was computed by calculating the mean Ct of three control genes (RPL13, ACTIN, UBC) in each sample and subtracting this from the Ct value for every other gene in that sample (Vandesompele et al., 2002). $\Delta\Delta Ct$ was computed for each gene by first calculating the mean ΔCt of that gene in the three scrambled siRNA control samples and then subtracting this value from the ΔCt values of the gene in all samples. Relative expression values and 95% confidence intervals were estimated from the $n = 3$ independent experiments as $2^{\text{mean}(-\Delta\Delta Ct)}$ and $\left[2^{\text{mean}(-\Delta\Delta Ct) - \frac{1.96\ \text{sd}(-\Delta\Delta Ct)}{\sqrt{n}}}, 2^{\text{mean}(-\Delta\Delta Ct) + \frac{1.96\ \text{sd}(-\Delta\Delta Ct)}{\sqrt{n}}}\right]$, respectively.

Robust linear regression

Where noted, fit lines were obtained using robust linear regression. This was done in R using the lmrob function from the robustbase package (Todorov and Filzmoser, 2009) with lmrob.control("KS2014").

Fold-change estimates

To compute fold-changes in intensity values observed between conditions, we took the median across all cells in each well (or a subset of cells such as a certain cell-cycle stage, as specified). We then compared these well-medians of a particular experimental condition with the well-medians of the appropriate control wells by computing all pairwise ratios (for example n replicate experimental wells with k control wells gives $n \times k$ fold-change estimates). We either summarized all $n \times k$ ratios as a boxplot or reported summary statistics for these ratios such as their mean and standard deviation.

Partial correlations

Pairwise partial correlation coefficients (Figure S12H) were calculated using the pcor function in the ppcor package in R (Kim, 2015). Such partial correlations represent correlations in which the effects of all other variables on the two under consideration are first taken into account. In cases where we account for the effects of a single additional variable before calculating the correlation (e.g. nuclear area ($r_{x,\mathrm{EU}|\mathrm{nuclear\ area}}$)), partial correlations were calculated using the pcor.test function.

Multiple linear regression

Linear regression was performed with sum nuclear EU intensity as the response variable and different combinations of predictors. In cases where there were more than three predictors, we standardised all variables, performed principal component analysis (using the prcomp function in R), and kept enough principal components to capture 99.9% of the variance in the data. For 4i data, we omitted outlier antibody/well combinations, identified as described above. Multiple linear regression was performed using the 'lasso' method with 10-fold cross-validation using the glmnet R package (cv.glmnet) (Friedman et al., 2010). To fit models, we omitted each replicate well from the training data, trained a model on the remaining data, and then predicted values for the non-training well. This procedure was repeated for each replicate well, to generate predicted values for every cell. The procedure avoids testing models on data that was used to fit the model, but does not arbitrarily assign a training and test set.

'Residual'-based corrected intensities

Changes in EU incorporation or RNA abundance in genetically perturbed cells are confounded with changes in cell size, cell cycle and/or population-context. To correct for these effects, we used regression to linearly account for any confounding variables at the single cell level, and then used the residuals of this regression as a 'corrected' intensity measurement. In all cases, regression models were trained on control cells from the same experiment, so the residual measurement represents the deviation of a particular cell from that expected for a control cell. Variables that were corrected in this way are described throughout the main text, including mean nuclear EU, sum nuclear EU, sum nucleolar EU and sum nucleoplasmic EU in metabolic RNA labelling experiments; RNA Strandbrite and poly(A) FISH in the RNA abundance screen; smFISH transcript abundance counts; POLR2A and EU intensities in 4i dataset, and POLR2A-GFP in MRC5 POLR2A-GFP cells.

The particular confounding variables considered differed between experiments, depending on which additional information was available from the other cellular stains, and also the type of analysis being considered. The variables whose effects are included in the correction are identified in the text. Typically, these included some or all of the following: nuclear and cell area, cell cycle stage, total protein content (sum cell succinimidyl ester), and local cell density. We now describe in detail how these corrections are performed for a single response variable.

We first identified if there were any outlier control wells. These are important to remove before using as training data for the regression models. To identify these, we trained a model on all replicate control wells except one, and then predicted the response variable in the well omitted from model training. Quantifying this model fit using using R^2, we then used a one-sided boxplot rule on R^2 to classify wells as outliers if $R^2 < Q_1(R^2) - 1.5 \times \mathrm{IQR}(R^2)$, where IQR is the interquartile range. This specifically identifies wells that are poorly predicted by models trained on all other replicate wells. Such 'outlier' wells are retained in the dataset but were omitted for training the regression model used to correct non-control wells to 'residual'-type variables.

Regression models were trained in several different ways, depending on the number of predictors. For example, residual mean EU is derived from a model containing a single categorical predictor: cell cycle stage, so the 'residual' simply corresponds to subtracting a cell-cycle specific value from each measurement. In contrast residual sum EU for the nucleus, nucleolus and nucleoplasm contains total protein content a continuous predictor as well as cell cycle as a discrete predictor, so simple linear regression was used. For more than two predictors, we used cross-validated 'lasso' multiple linear regression using the glmnet R package (cv.glmnet) (Friedman et al., 2010). In cases with several correlated predictors we also included a PCA-based dimensionality reduction of the predictor variables (maintaining 99.5% of variance) before regression. Typically, these values were averaged on a per-well basis by taking the mean. These "mean residual" measurements were finally standardised to the mean and variance of scrambled siRNA control wells using the Z-score, to give a measurement in units of standard deviation of scrambled siRNA controls.

Genome-wide and secondary screen analysis

Quantitative data analysis of single-cell intensity measurements from the genome-wide screen and secondary screen were described previously (Müller et al., 2021). Lower and upper hit thresholds correspond to posterior probabilities of 0.5 and 0.85 for perturbations being reproducibly observed outside the 1st and 99th percentile of scrambled siRNA controls. Calculation of these thresholds was also described previously (Müller et al., 2021). 'Low cell number' in the screen corresponds to 500 interphase cells imaged and segmented after data cleanup. There were 258 conditions with less than 500 cells in the genome-wide screen (approximately 1% of gene perturbations). These were included in certain overviews (e.g. Figures 2C and 2D) but were typically excluded, as noted throughout, for example in hit scoring and functional annotation enrichment.

Hit scoring made use of two residuals-based models: either predicting 'sum EU' (with cellular protein content and cell-cycle stage as predictors) or 'mean EU' (EU intensity averaged over nuclear area, with only cell cycle stage as a predictor). The 'sum EU' model

explicitly accounts for cell size, while 'mean EU' implicitly takes cell size changes into account because nuclear size also scales with cell size (Cantwell and Nurse, 2019). In unperturbed cells, both nuclear area and protein content are proportional to cell volume (Figure S1G). These two residual-based measurements were highly correlated ($r = 0.94$), but 'mean EU' was slightly less variable between wells and screens and was therefore more sensitive, possibly because it avoids technical variability in measurement of cellular protein content (Figures S4B and S4D) (Müller et al., 2021). Mean residual sum EU for the nucleolus and nucleoplasm were calculated as described for the sum EU model. Comparison of mean residual measurements between primary and secondary screens was used to define hit thresholds in all cases, as described previously (Müller et al., 2021).

The selection of 436 siRNA perturbations for the secondary screen (RNA metabolic labelling and poly(A) FISH / RNA Strandbrite) was done in an automated manner designed to preserve both functional and phenotypic diversity of the panel, as described previously (Müller et al., 2021). Selection of genes for the final panel of 63 perturbations for detailed characterization by 4i was done manually based on enriched functional annotations and results of the secondary screen. Gene lists and other metadata are provided together with the experimental results at the locations listed in key resources table.

Estimation of transcriptional scaling in genome-wide screen

Linear regression to determine slope and intercept values for the relationship between sum nuclear EU and cellular protein content was performed for all perturbations in the genome-wide screen with at least 500 cells. This was done in R using the lmrob function from the robustbase package (Todorov and Filzmoser, 2009) with lmrob.control("KS2014"). If there were less than 30 cells in a particular cell-cycle stage, those cells were omitted. This analysis was complicated by the presence of strong overall changes to EU incorporation and incomplete phenotypic penetrance of perturbations, which together resulted in cell populations with large cell-to-cell variability in EU incorporation. Across perturbations, slopes were correlated with mean EU intensity, however they were highly variable even between replicates (Figure S4F).

Robust coefficient of variation (RCV)

Despite data clean-up, single cell data can contain spurious outlier observations, particularly in genetic perturbation experiments. To estimate the population variability, we therefore used a robust analogue of the coefficient of variation given by $RCV_M = 1.4826 \times \mathrm{MAD} / \mathrm{median}$, where $RCV_M = 1.4826 \times \mathrm{MAD} / \mathrm{median}$ is the median absolute deviation, given by $\mathrm{MAD}(x) = \mathrm{median}(|x - m|)$ where $m = \mathrm{median}(x)$ (Arachchige et al., 2022).

4i data analysis – UMAP

After background subtraction and data normalisation, we combined mean nuclear intensities and textures for all markers except EU, together with mean cytoplasmic intensities and textures of PABPC1 and PABPC4, nuclear and cell morphology measurements, and measurements of local cell density (Snijder et al., 2012) into a matrix of 436,593 cells by 718 features. We then standardised all features to scrambled siRNA control cells using the robust z-score: $x \rightarrow (x - \mathrm{median}(x_{\mathrm{scrambled}}))/(1.4826 \times \mathrm{MAD}(x_{\mathrm{scrambled}}))$ and performed dimensionality reduction via principal component analysis (PCA) using the R function prcomp, keeping 95% of the variance (first 108 principal components). We then randomly sampled 1400 cells from each siRNA perturbation condition and used this PCA-transformed data as input for umap using the R uwot package (Melville, 2021) with parameters a = 0.25 and b = 0.9. Finally, we used the umap_transform function to embed the full dataset in UMAP space.

Systems-level analysis

Functional annotation enrichment scoring

We retrieved GO biological process, KEGG pathways, and Reactome annotations for all genes assayed using the GeneSets.Homo.sapiens R package (Simillion, 2020), and removed any genes with zero annotations. Genes were then ranked using a quantitative phenotype (e.g. mean residual EU) for both 'up' and 'down' hit classes separately. These ranked lists were used to search for over-represented annotations, adapting a previously published method (Green and Pelkmans, 2016). Briefly, for each annotation, we count the number times, k that a given annotation is observed between rank 1 and n, and compute the probability that this occurs by chance, given the number of assayed genes, N, of which K have the corresponding annotation. This probability is given by the hypergeometric distribution function:

$$P(k;K,N,n) = \sum_{x=k}^{K} \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}},$$

which is implemented in R as phyper(k-1,K,N-K,n,lower.tail = FALSE). This calculation is repeated for all ranks from $n = 1,\ldots,n_t$ where n_t is the rank at which the probability threshold exceeds 0.5 (50/50 whether a hit is reproducible). The functional annotation enrichment score (FAES) is then calculated using the minimum value of this probability, $\mathrm{FAES} = -\log_{10}(\min_{n<n_t} P(k;K,N,n))$. We also note the rank $n = n_{\min}$ at which this minimum occurs.

To represent these enriched annotations as a network, we selected up and down-enriched annotations with FAES > 2, removing any annotations with less than 20 or more than 3500 genes. We then measured the pairwise similarities between annotations, using the Cohen's kappa statistic, κ, (Cohen, 1960) provided by the Kappa function in the vcd R package (Meyer et al., 2020). Annotations for which $\kappa > 0.85$, were merged into a group, keeping the highest FAES score for an annotation in that group. We then built a graph of these enriched annotations, where edges between nodes represent annotations with some overlap in the gene set ($\kappa > 0.15$). The

network was finally visualised in Cytoscape (Shannon et al., 2003) using spring-embedded layout based on the edge weights (κ), also making use of enhancedGraphics (Morris et al., 2014).

To calculate FAES for the nucleolar-specificity of annotations that were enriched for reduced EU incorporation, we calculated "nucleolar preference" of a perturbation, which we defined as the residual of "residual nucleolus sum EU" after regressing out "residual nucleoplasm sum EU" (Figure S6H). Nucleolar preference gives a quantitative measurement of the reduction in nucleolar EU compared to that expected based on nucleoplasmic EU. Nucleolar preference is positive for perturbations with stronger nucleolar EU reduction than nucleoplasmic EU reduction. We then performed rank-based enrichment analysis for nucleolar preference using all genes with $p_{posterior} > 0.5$ for either residual nucleoplasm sum EU or residual nucleolus sum EU (Figure S6I). Finally, we compared this to FAES for reduced residual nucleolus sum EU, and identified annotations that were enriched in both analyses (Figure S6J).

Gene interaction networks

After grouping enriched annotations using Cohen's κ, as described above, we obtained protein-protein association networks from the STRING database (Szklarczyk et al., 2019) for all genes in a certain annotation group. We then constructed networks in which genes are connected if their aggregated interaction score is > 0.7. Networks were visualised in Cytoscape (Shannon et al., 2003) using spring-embedded layout based on the edge weights (association score).

Hierarchical clustering

Hierarchical clustering was performed in R using the seriation package (Hahsler et al., 2008) using Ward's algorithm with optimal leaf ordering. Heatmaps were visualised with the ComplexHeatmap package (Gu et al., 2016).

Mathematical modelling

Ordinary differential equation (ODE) modelling was performed in R (R Core Team, 2014) using the deSolve (Soetaert et al., 2010) package. Parameter optimisation was done using the optimx (Nash and Varadhan, 2011) package, using least-squares minimisation on log-transformed data and predictions.

Minimal model without RNA-based feedback

Steurer et al. (2018) developed a model of the RNA Pol II transcription cycle in which RNA Pol II can exist in one of four states, which we here refer to as Unbound, Initiating, Paused, and Elongating. Their model was parameterised and validated using RNA Pol II fluorescence recovery after photobleaching (FRAP) of MRC5 POLR2A-GFP cells, using several chemical inhibitors with known mechanism of action. To adapt this model to an ordinary differential equation framework, we modelled the RNA Pol II transcription cycle using the reaction network depicted in Figure S13A. Using mass-action kinetics, this gives rise to the following system of ODEs in terms of concentration,

$$\begin{aligned}
\dot{p_u} &= k_s + k_1 p_i + k_2 p_p + k_t p_a - k_d p_u - k_i p_u \\
\dot{p_p} &= k_p p_i - (k_2 + k_a) p_p \\
\dot{p_a} &= k_a p_p - k_t p_a \\
\dot{R} &= k_t p_a - r_d R
\end{aligned}$$

where p_u, p_i, p_p, p_a are unbound, initiating, paused, active (elongating) RNA Pol II, R is RNA, and $\dot{x} = \frac{dx}{dt}$ denotes the time-derivative of x. Note that RNA is created at a rate $k_t p_a$, the same overall rate at which RNA Pol II terminates elongation, meaning that we do not consider premature termination of transcription (other than transcription aborting after pausing). We set k_t using the FRAP recovery timescale of the elongating state (1370 seconds), and then optimised the remaining RNA Pol II transition rate parameters (k_i, k_p, k_a, k_1,k_2) to obtain the observed fractions of RNA Pol II states at steady-state (7% Free, 10% Initiating, 23% Paused, 60% Elongating). A further constraint from (Steurer et al., 2018) relates to the finding that only 12.7% of polymerases that attempt initiation proceed to promoter pausing, and only 7.6% of promoter-paused polymerases continue to productive elongation. This allowed us to specify k_1 and k_2 in terms of k_p and k_a, respectively ($k_1 = 6.87 \times k_p$, $k_1 = 12.16 \times k_a$).

After setting the total concentration of RNA Pol II in the nucleus at 10^5 molecules pL^{-1} (Steurer et al., 2018), there remains a single free parameter (k_s), which sets the timescale of RNA Pol II synthesis / degradation. This parameter has units of molecules $pL^{-1}s^{-1}$, so a constant value corresponds to an absolute (molecular) rate of RNA Pol II synthesis that scales with cell volume. To extract k_s from our data, we simulated triptolide-based transcriptional inhibition by setting $k_i = 0$, and AZD4573-based CDK9 inhibition by setting $k_a = 0$, keeping all other parameters unchanged. We then optimised k_s to fit the combined data from these two experiments, with POLR2A-S2P identified as elongating RNA Pol II (Heidemann et al., 2013) (POLR2A-S2P $= p_a$). Parameters and their final fitted values are listed in Table S1. To investigate alternative models in which multiple RNA Pol II states were subject to degradation (Figures 6B and S13B), we added extra $-k_d p_x$ (for $x \in i,p,a$) terms to the relevant ODEs, and refit k_s, as described above.

Adding RNA-based feedback

Having parameterised the RNA Pol II model, we then included RNA-based feedback as either a Hill-type activation of RNA Pol II degradation $k_d \rightarrow k_d(R/K + R)$, or a Hill-type repression of k_i, k_p, k_a (with Hill coefficient $n = 1$), $k_x \rightarrow k_x(K/K + R)$ for $x \in i,p,a$. In all cases, this change required re-fitting of some of the parameters from the base model to achieve the correct fractions of RNA Pol II in each state and the same total RNA Pol II concentration (Table S1).

To fit the RNA degradation rate and feedback parameters, r_d, and K, we simulated auxin-induced degradation of DIS3 by setting,

$$r_d \rightarrow r_d(t) = r_{d(\min)} + \left(r_{d(\max)} - r_{d(\min)}\right) 2^{-t/20}$$

which simulates the exponential decrease in DIS3 abundance observed experimentally (Davidson et al., 2019). r_d does not decrease to zero because this parameter represents all nuclear mRNA degradation and export, as well as dilution due to nuclear growth. We used $r_{d(\min)} = r_{d(\max)}/4$, which assumes that DIS3 is responsible for ~3/4 of nuclear mRNA degradation, however the actual value does not qualitatively affect the differences seen between models. Assuming POLR2A $= p_u + p_i + p_p + p_a$, POLR2A-S5P $= p_p + p_a$, POLR2A-S2P $= p_a$ and mRNA $= R$, we optimised r_d, and K to fit mRNA (poly(A) FISH) and RNA Pol II (POLR2A, POLR2A-S5P, POLR2A-S2P immunofluorescence) measurements from DIS3-AID experiments. The magnitude of RNA Pol II changes in the model are governed mostly by K (representing the concentration of RNA with half-maximal effect on k_x). However, the precise value of K does not affect the qualitative differences between models in terms of the relative effects on POLR2A, POLR2A-S5P, and POLR2A-S2P.

To analyse how steady-state RNA levels are affected by changes to model parameters (parameter sensitivity analysis) in Figures S13D and S13E, we focused on the model which was most consistent with the experimental data (RNA Pol II feedback on "paused" to "elongating" transitions: $k_a \rightarrow k_a(K/K+R)$). We varied either k_s (to simulate over- or under-production of RNA Pol II) or r_d (to simulate changes to nuclear RNA degradation, export or growth rate), and then re-calculated the steady-state values of POLR2A or RNA for this new parameter set (represented relative to steady state levels obtained from the best-fit parameters). In Figure S13E, we further considered how non-linear feedback of RNA affects buffering capacity by $k_a \rightarrow k_a(K^n/K^n+R^n)$ for $n \in 1, 2, 4$. In Figures S13F and S13G, we considered how models respond dynamically to perturbation of cell volume. For example, a 20% cell volume increase was modelled by dividing each of the steady-state concentrations by 1.2. We then simulated the return to steady-state concentrations using numerical integration.