TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding

Jeremy A Schofield\textsuperscript{1,2}, Erin E Duffy\textsuperscript{1,2}, Lea Kiefer\textsuperscript{1,2}, Meaghan C Sullivan\textsuperscript{1,2} & Matthew D Simon\textsuperscript{1,2}

RNA sequencing (RNA-seq) offers a snapshot of cellular RNA populations, but not temporal information about the sequenced RNA. Here we report TimeLapse-seq, which uses oxidative-nucleophilic-aromatic substitution to convert 4-thiouridine into cytidine analogs, yielding apparent U-to-C mutations that mark new transcripts upon sequencing. TimeLapse-seq is a single-molecule approach that is adaptable to many applications and reveals RNA dynamics and induced differential expression concealed in traditional RNA-seq.

Global changes in transcription can occur on the timescale of minutes to hours in many mammalian systems, including circadian rhythms and the immune response. One can identify such changes by monitoring new transcripts that cofractionate with chromatin\textsuperscript{1,2} or that have not yet been splice\textsuperscript{3,4}. New RNA populations can also be identified by examining sites of active RNA polymerase II through biochemical enrichment of transcripts that are being synthesized (e.g., PRO-seq\textsuperscript{5} and NET-seq\textsuperscript{6}) or metabolic labeling and enrichment of new transcripts (e.g., TT-seq\textsuperscript{7} and s\textsuperscript{4}U-seq\textsuperscript{8,9}). These techniques require large amounts of input sample and extensive handling, and they present challenges when normalizing enrichment and estimating contamination.

To capture temporal information about RNA directly in a sequencing experiment without biochemical enrichment, we developed TimeLapse-seq (Fig. 1a), a method in which cells are exposed to a noncanonical nucleoside that becomes incorporated into only new transcripts.

RESULTS
Development of TimeLapse chemistry

Rather than enriching the metabolically labeled RNAs, we developed chemistry that recodes the hydrogen-bonding pattern of the uridine analog 4-thiouridine (s\textsuperscript{4}U) to match the hydrogen-bonding pattern of cytosine, thereby causing mutations in a sequencing experiment. This strategy is similar to bisulfite sequencing, which uses chemically induced mutations to recode nucleobase hydrogen bonding to provide insight into DNA methylation. In our strategy, the recoded nucleosides distinguish which RNAs were transcribed during the time of the experiment. TimeLapse-seq results are internally normalized, as both pre-existing and new transcripts are present in the same library. These mutations reveal which RNAs were recently synthesized by the cell, thereby capturing the rich dynamics of the transcriptome.

To develop TimeLapse-seq we focused on s\textsuperscript{4}U because of its utility in RNA metabolic labeling experiments\textsuperscript{10,11} and the orthogonal reactivity of its thione relative to other functional groups found in RNA. The s\textsuperscript{4}U base itself leads to low levels of U-to-C transitions upon reverse transcription\textsuperscript{12}, but does so at levels too low to robustly identify new transcripts. While recent applications of s\textsuperscript{4}U have focused on the thione as a nucleophile\textsuperscript{8}, or for UV cross-linking\textsuperscript{11,13}, we were inspired by less explored reactivity—transforming s\textsuperscript{4}U using oxidative-nucleophilic-aromatic substitution\textsuperscript{14}. We reasoned that oxidation of s\textsuperscript{4}U would transform it into a convertible nucleoside, providing an intermediate that could be converted into an analog of cytosine by aminolysis (Fig. 1a). The s\textsuperscript{4}U base retains uridine’s Watson–Crick hydrogen-bonding pattern, and while other chemical conditions used to modify s\textsuperscript{4}U (e.g., alkylation) change the base’s hydrogen bonding pattern, they do not recode the base to match C’s native hydrogen-bonding pattern. While not widely explored, the oxidative reactivity of s\textsuperscript{4}U has precedent in UV-cross-linking studies, where sites of s\textsuperscript{4}U-protein crosslinks are enriched for T-to-C mutations, or in mapping the locations of s\textsuperscript{4}U bases in Escherichia coli tRNA\textsuperscript{11,15}. If conducted before an RNA-seq analysis, this reaction could reveal sites of s\textsuperscript{4}U incorporation through T-to-C mutations stably introduced in the cDNA.

We explored chemistry to convert the free nucleoside (s\textsuperscript{4}U) to cytidine derivatives (Fig. 1a and Supplementary Fig. 1) while minimizing oxidation of guanosine (Supplementary Fig. 2) and using amines with low pK\textsubscript{a} values that remain deprotonated under neutral reaction conditions. We found that treating s\textsuperscript{4}U with 2,2,2-trifluoroethylamine (TFEA) and meta-chloroperoxybenzoic acid (mCPBA) results in near-complete consumption of s\textsuperscript{4}U, producing only small amounts of the hydrolysis product uridine, and mostly the desired trifluoroethylated cytidine (C\textsuperscript{\*}, Supplementary Fig. 3a). Similar conditions were successful in the context of an oligoribonucleotide. Optimization of the nucleophile, oxidant, temperature, and time through a restriction-enzyme-digestion assay (see Online Methods, Supplementary methods).

\textsuperscript{1}Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, Connecticut, USA. \textsuperscript{2}Chemical Biology Institute, Yale University, West Haven, Connecticut, USA. Correspondence should be addressed to M.D.S. (matthew.simon@yale.edu).

Received 15 May 2017; accepted 14 December 2017; published online 22 January 2018; DOI:10.1038/nmeth.4582
Supplementary Fig. 4a–c, and Supplementary Table 1) led us to the combination of TFEA and sodium periodate (NaIO₄; Fig. 1b). These reagents cause clean transformation of 4-thiouracil to N4-trifluoroethylcytosine by ¹H NMR (Supplementary Fig. 3b). When RNA with a single s¹U was subjected to these conditions (45 °C, 1 h), reverse transcriptase could efficiently transcribe the product, and the majority of the resulting DNA (~80%) had the desired T-to-C mutation (Supplementary Figs. 3c,d and 4d). NaIO₄ is an oxidant commonly used in RNA biology to oxidize the 3’-end vicinal diol of RNAs with minimal effects on other functional groups, even through multiple rounds of oxidation¹⁶. To test NaIO₄ and TFEA with cellular s¹U-RNA, we exposed mouse and human cells to a range of concentrations of s¹U. After RNA isolation and chemical treatment, we examined the apparent U-to-C conversion rates (inferred from T-to-C mutations in the cDNA, hereafter referred to as T-to-C) by targeted RT-PCR coupled to paired-end sequencing (Supplementary Table 1). We observed a notable and specific increase in T-to-C transitions in chemically treated samples (Supplementary Fig. 5).

Monitoring global steady-state RNA turnover

To examine the dynamics of cellular RNAs, we treated MEF cells with s¹U for 1 h (where no s¹U toxicity was observed; Supplementary Fig. 6a–c) and performed TimeLapse chemistry before sequencing. The total transcript counts from each sample were highly correlated irrespective of s¹U exposure or chemical treatment (Pearson’s r ≥ 0.97, Supplementary Fig. 6d), demonstrating that TimeLapse-seq retains information from a traditional RNA-seq experiment. By counting the mutations in each aligned read pair, we found a specific and reproducible increase in T-to-C mutations dependent on both metabolic labeling with s¹U and chemical treatment (Supplementary Figs. 7 and 8). Other mutation rates remained below background levels of T-to-C mutations in untreated samples (e.g., the small increase in G-to-T mutations, Supplementary Fig. 2c,d). Additionally, the reaction was efficient even in regions of RNA secondary structure (Supplementary Fig. 9). The T-to-C mutation counts were dramatically higher in fast-turnover transcripts (e.g., Myc and Fosl2), compared to more stable transcripts (e.g., Dlx9 and Ybx1) (Fig. 2a,b and Supplementary Fig. 10). We observed an enrichment of T-to-C mutations in intronic reads (Fig. 2c) consistent with the fast turnover of intronic RNA. To quantify these results, we modeled reads as arising from two populations: pre-existing RNAs (background mutation rate) and new RNAs (high T-to-C mutation rate; Fig. 2b; see Online Methods). Reads from newly synthesized RNAs had an average of 2.2 mutations per read, corresponding to an ~3% mutation rate per uridine (compared to ~0.1% T-to-C mutation rates in controls and for pre-existing RNAs). From each gene, we determined the fraction of newly made transcripts (r ≥ 0.94; 2,992 genes; Supplementary Table 2) and estimated transcript half-lives, which correlated with those reported previously¹⁷ (Supplementary Fig. 11). As expected, the fast-turnover RNAs (top 10%, n = 360) were enriched for transcripts such as transcription factors (Results from PANTHER analysis: DNA-templated transcription, P < 10⁻²⁰), while the slow-turnover RNAs (top 10%, n = 361) were enriched for those that are involved in translation (Results from PANTHER analysis: ribosomal biogenesis, P < 10⁻⁶; translation, P < 10⁻⁷). Estimates of the fraction of newly synthesized RNA were particularly robust when the new transcripts represented ~200 reads in the experiment (Supplementary Fig. 12 and Supplementary Note).

Distinguishing transient transcripts from contaminating reads

Very transient RNA species, such as reads beyond the poly-A termination signal in a gene body, provide insight into transcriptome dynamics but are generally too rare to be observed at high levels by RNA-seq. While these dynamics can be studied through biochemical enrichment of very recently made RNAs after short (5 min) s¹U treatments through transient transcriptome sequencing (TT-seq⁷), biochemically enriched s¹U-RNA always contains contaminating reads from unlabeled RNAs (estimated to be up to 30% in some experiments¹²). This contaminating background can limit analyses; for example, abundant spliced transcripts observed in RNA enriched after short s¹U pulses have been interpreted as fast splicing¹⁸, but these results could also be explained by contaminating background (e.g., from fully spliced mature RNAs). To test if TimeLapse chemistry could be used in conjunction with TT-seq to distinguish bona fide new RNAs from contaminating background, K562 cells were labeled for 5 min with s¹U, and...
biochemical enrichment was performed as in TT-seq\(^7\), except with more efficient MTS chemistry to biotinylate the s\(^4\)U-RNA\(^9\) (Supplementary Fig. 13a). After enrichment and before sequencing, we performed TimeLapse chemistry. As expected, transient RNA species were enriched for introns (two-sample Kolmogorov–Smirnov test, \(P < 10^{-15}\), Fig. 2d–f and Supplementary Fig. 13) but depleted for splice junctions (\(P < 10^{-15}\)). Both enrichment of introns and depletion of splice junctions were slightly greater than previously observed\(^7\) (Supplementary Fig. 13c, d), which was likely due to the efficiency of MTS chemistry. Even with only 5 min of s\(^4\)U treatment, the majority of the biochemically enriched reads contained TimeLapse-induced mutations (Fig. 2d).

**Figure 2** | Global analysis of steady-state and transient RNA dynamics using TimeLapse-seq. (a) Left, tracks depicting coverage from all reads (gray) for transcripts with slow (Ybx1), moderate (Dhx9), or fast (Fosl2) rates of turnover. Right, tracks from reads with increasing numbers of T-to-C mutations (see scale) displaying mutational content provided by TimeLapse chemistry (right, y-axis zoom 3x). (b) Distribution of reads with each number of T-to-C mutations (points) overlaid on a model of the estimated distribution of reads from new transcripts (red) and pre-existing transcripts (gray) for Ybx1, Dhx9, and Fosl2. The estimated fraction of new reads is indicated for each plot. Light gray, 95% confidence interval. (c) Distribution of T-to-C mutations found in reads mapping to Ybx1, Dhx9, and Fosl2, separated by total, exonic, or intronic reads. (d) TT-TimeLapse-seq and RNA-seq tracks of DHX9. (e) Cumulative distribution plot of reads containing splice-junctions in RNA-seq, and TT-TimeLapse-seq. (f) Cumulative distribution plot of intron-only reads in RNA-seq and TT-TimeLapse-seq with the same scale as in e. (g) Using TimeLapse-seq to distinguish new RNAs after heat shock. Log\(_2\) fold changes after heat shock in total RNA-seq counts and new RNA counts for the top RNAs identified in b as significantly changed upon heat shock (\(P_{adj} < 0.01\)). (h) RNA-seq and TimeLapse-seq tracks of Hsp11 (top) and Hsp90aa1 (bottom) upon heat shock.
Figure 3 | TimeLapse-seq reveals differential transcript isoform stability of the ASXL1 transcript. (a) ASXL1 tracks from TimeLapse-seq (4-h s\textsuperscript{3}U treatment) with exon-containing regions expanded (lower panel). (b) Exonic T-to-C mutation distributions for ASXL1 in comparison with three transcripts with different stabilities, ACTB, CDK1, and FOSL1.

(Fig. 2e,f and Supplementary Fig. 13c,d). This suggests that mutated reads effectively capture the profile of new RNAs, while the reads without mutations represent a subpopulation that is contaminated by unlabelled reads. We estimated that 15–20% of total TT-seq reads arise from contaminating RNA (estimate from splice-junction content, 17–20%; from intronic content, 18–20%; see Online Methods), similar to estimates from previous s\textsuperscript{3}U experiments\textsuperscript{12}. Reads without mutations were enriched for contaminating reads (estimate from splice junctions, 33–39%; estimate from introns, 35–40%), while reads containing mutations are depleted in contamination. For reads with a single mutation, contaminating reads make up <5% of the signal; for reads with two mutations, the contamination is <1%. Taken together, RNA contamination contributes to the signal at the level of RNA-seq, but TimeLapse-chemistry-induced mutations can be used to discriminate between signal from new RNAs and contaminating reads. These results demonstrate transcripts including ACTB (Supplementary Fig. 13b) are not highly spliced on this timescale (5 min) and highlight how TimeLapse chemistry can provide an extra specificity filter when analyzing rare, transient RNAs.

Identifying acute transcriptional changes with TimeLapse-seq
To test whether TimeLapse-seq could reveal induced changes in RNA populations, we subjected MEF cells to a mild heat shock (42 °C, 1 h), where only modest changes in total RNA levels were apparent\textsuperscript{19–21}. We observed induction of a few transcripts such as Hspa1b by RNA-seq (Supplementary Fig. 14a), but TimeLapse-seq revealed the induction of many transcripts encoding heat shock proteins in the new transcript pool that are not apparent by RNA-seq alone (Fig. 2g and Supplementary Fig. 14). For example, whereas RNA-seq is less sensitive to the small absolute changes in Hsp\textit{h}1 and Hsp90\textit{a}a1 (as they are already abundant before heat shock; RNA-seq fold-change, H\textit{h}sp1 = 1.8-fold, Hsp\textit{90}aa1 = 1.1-fold, DEseq2), TimeLapse-seq reveals substantial induction of both transcripts in the new transcript pool (TimeLapse-seq fold change, H\textit{h}sp1 = 12.7-fold, Hsp\textit{90}aa1 = 3.1-fold, DEseq2) (Fig. 2h). Unlike PRO-seq and NET-seq, however, which are not sensitive to changes in RNA populations after transcription has completed, TimeLapse-seq captures changes in RNA processing—we observed the induction of a new terminal exon in Rsr\textit{p}1 upon heat shock (Supplementary Fig. 14c,d) as well as post-transcriptional down-regulation of histone mRNAs upon heat shock (Supplementary Fig. 14f,g), neither of which would be apparent from analysis of nascent RNA.

Discovering isoform-specific transcript dynamics
We applied TimeLapse-seq using treatment conditions optimized for studying mRNA turnover (4-h s\textsuperscript{3}U)\textsuperscript{22} in a chronic myelogenous leukemia model cell line (K562). We obtained highly reproducible half-life estimates that correlated with previous observations\textsuperscript{23} (Supplementary Fig. 11). Inspection of individual transcripts revealed reads mapping to both a shorter isoform of ASXL1 (NM_001164603), as well as a longer isoform (NM_015338) of ASXL1. The ASXL1 protein is involved in epigenetic regulation of chromatin, and mutations in the longer isoform of this gene are implicated in myelodysplastic syndromes (MDSs)\textsuperscript{24}. Analysis of the mutational content of the individual exons from ASXL1 demonstrated that reads mapping to the longer isoform had substantially greater turnover than those mapping to the first four exons (Figs. 3a,b), a conclusion supported by transcriptional inhibition (Supplementary Fig. 15b,c). The different stability of ASXL1 isoforms is particularly intriguing given the importance of RNA processing to many pathologies, including MDS\textsuperscript{25}.

DISCUSSION
TimeLapse-seq is a single-molecule approach to monitor transcriptome dynamics. The method reveals different rates of RNA turnover, changes in RNA processing, and acute changes in the transcriptome that are not apparent using standard RNA-seq. TimeLapse-seq is conceptually similar to SLAM-seq\textsuperscript{26}, a method that was reported during the revision of this manuscript. Whereas SLAM-seq uses alkylation of the s\textsuperscript{3}U thione to induce mutations, TimeLapse-seq instead uses an oxidative nucleophilic aromatic substitution reaction to fully recode the hydrogen bonding pattern of s\textsuperscript{3}U to match the native pattern of cytidine. TimeLapse-seq allowed us to conduct, to our knowledge, the first enrichment-free
analysis RNA dynamics transcriptome wide, covering entire transcripts including their intrinsic sequences. TimeLapse-seq is broadly applicable to any application compatible with metabolic labeling (e.g., TT-seq). This approach provides a flexible platform to investigate dynamic biological systems.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Steitz, A. Schepartz, D. Söll, D. Canzio, and the Simon Lab for insightful comments; and we thank Y. Wang and A. Sexton for assistance and scripts used in mutational analysis of targeted sequencing data. This work was supported by the NIH NIGMS T32GM007223 (J.A.S. and E.E.D.); NSF Graduate Research Fellowship (E.E.D.); NIH New Innovator Award DP2 HD083992-01 (M.D.S.), and a Searle scholarship (M.D.S.).

AUTHOR CONTRIBUTIONS

J.A.S. and M.D.S. designed experiments. J.A.S., M.C.S., and M.D.S. performed computational analyses of experiments. J.A.S., E.E.D., and L.K. carried out experiments. J.A.S., M.C.S., and M.D.S. performed computational analyses of data. J.A.S. and M.D.S. wrote the manuscript with assistance from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Bhatt, D.M. et al. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. Cell 150, 279–290 (2012).
2. Menet, J.S., Rodriguez, J., Abruzzi, K.C. & Rosbash, M. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. elife 1, e00011 (2012).
3. Wada, Y. et al. A wave of nascent transcription on activated human genes. Proc. Natl. Acad. Sci. USA 106, 18357–18361 (2009).
4. Gaidatzis, D., Burger, L., Florescu, M. & Stadler, M.B. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat. Biotechnol. 33, 722–729 (2015).
5. Kwak, H., Fuda, N.J., Core, L.J. & Lis, J.T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science 339, 950–953 (2013).
6. Churchman, L.S. & Weissman, J.S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature 469, 368–373 (2011).
7. Schwab, B. et al. TT-seq maps the human transient transcriptome. Science 352, 1225–1228 (2016).
8. Rabani, M. et al. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nat. Biotechnol. 29, 436–442 (2011).
9. Duffy, E.E. et al. Tracking distinct RNA populations using efficient and reversible covalent chemistry. Mol. Cell 59, 858–866 (2015).
10. Melvin, W.T., Milne, H.B., Slater, A.A., Allen, H.J. & Keir, H.M. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. Eur. J. Biochem. 92, 373–379 (1978).
11. Rabani, M. et al. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. Cell 159, 1698–1710 (2014).
12. Hafner, M. et al. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141, 129–141 (2010).
13. Mishima, Y. & Steitz, J.A. Site-specific crosslinking of 4-thioridine-modified human tRNA(3Lys) to reverse transcriptase from human immunodeficiency virus type I. EMBO J. 16, 2679–2687 (1995).
14. Yano, M. & Hayatsu, H. Permanganate oxidation of 4-thiouracil derivatives. Isolation and properties of 1-substituted 2-pyrimidone 4-sulfonates. Biochim. Biophys. Acta 199, 303–315 (1970).
15. Ziff, E.B. & Fresco, J.R. A method for locating 4-thioridylate in the primary structure of transfer ribonucleic acids. Biochemistry 8, 3242–3248 (1969).
16. Dai, Q. et al. Nm-seq maps 2’-O-methylation sites in human mRNA with base precision. Nat. Methods 14, 695–698 (2017).
17. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. Nature 473, 337–342 (2011).
18. Mukherjee, N. et al. Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat. Struct. Mol. Biol. 24, 86–96 (2017).
19. Trinklein, N.D., Murray, J.L., Hartman, S.J., Botstein, D. & Myers, R.M. The role of heat shock transcription factor 1 in the genome-wide regulation of the mammalian heat shock response. Mol. Biol. Cell 15, 1254–1261 (2004).
20. Mahat, D.B., Salamanca, H.H., Duarte, F.M., Danko, C.G. & Lis, J.T. Mammalian heat shock response and mechanisms underlying its genome-wide transcriptional regulation. Mol. Cell 62, 63–78 (2016).
21. Shalgi, R., Hurt, J.A., Lindquist, S. & Burge, C.B. Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. Cell Rep. 7, 1362–1370 (2014).
22. Russo, J., Heck, A.M., Wilusz, J. & Wilusz, C.J. Metabolic labeling and recovery of nascent RNA to accurately quantify mRNA stability. Methods 120, 39–48 (2017).
23. Friedel, C., Dölken, L., Ruzzicis, Z., Koszinowski, U.H. & Zimmer, R. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. Nucleic Acids Res. 37, e115 (2009).
24. Getz-Gesi-Boyer, V. et al. Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br. J. Haematol. 145, 788–800 (2009).
25. Scotti, M.M. & Swanson, M.S. RNA mis-splicing in disease. Nat. Rev. Genet. 17, 19–32 (2016).
26. Herzog, V.A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods 14, 1198–1204 (2017).
ONLINE METHODS
A step-by-step protocol is available as a Supplementary Protocol and an open resource in Protocol Exchange.²⁷

Materials. All commercially available materials were purchased from the indicated suppliers and used without further purification. 4-thiouridine (s⁴U) and meta-chloroperoxybenzoic acid (mCPBA) were purchased from Alfa Aesar (Haverhill, Massachusetts). 4-thiouridine-5’-triphosphate (s⁴UTP) was purchased from TriLink BioTechnologies (San Diego, California). 2,2,2-trifluoroethylamine (TFEA), sodium acetate, EDTA, Tris hydrochloride, acrylamide/bis-acrylamide 30% solution, phenol:chloroform:isoamyl alcohol (25:24:1), and actinomycin D were purchased from Sigma Aldrich (St. Louis, Missouri). Sodium periodate (NaIO₄) and ammonium bicarbonate were purchased from Acros Organics (Geel, Belgium). Methane thiosulfonate biotin-XX (MTSEA-biotin-XX) was purchased from Biotium. Dynabeads MyOne Streptavidin C1 beads were purchased from Thermo Fisher Scientific. Agencourt RNA Clean XP beads were purchased from Beckman Coulter (Brea, California). Phusion HF PCR master mix and Dithiothreitol (DTT) were purchased from Thermo Fisher Scientific (Waltham, Massachusetts). Phosphate buffered saline (PBS) was purchased from AmericanBio (Natick, Massachusetts). Dulbecco’s Modified Eagle Medium (DMEM), fetal bovine serum (FBS), Trizol reagent, TURBO DNase, and SuperScript III Reverse Transcriptase were purchased from Life Technologies (Carlsbad, California). KAPA Taq Ready Mix was purchased from Kapa Biosystems Inc. (Wilmington, Massachusetts). DMSO-d₆, penicillin-streptomycin (P/S), and 33 mm 0.45 μm PDVF syringe filters were purchased from EMD Millipore (Billerica, Massachusetts). ATCC MTT Cell Proliferation Assay kit was purchased from American Type Culture Collection (Manassas, Virginia). NotI HF restriction enzyme was purchased from New England Biolabs (Ipswich, Massachusetts). SMARTer Stranded cDNA Synthesis Kit (Pico Input) was purchased from Takara Bio (California). Hypersil Gold 3 μm, 160 x 2.1 mm column was purchased from Thermo Fisher Scientific (Waltham, Massachusetts). K562 cells were a gift from the Strobel Lab, Yale Department of Molecular Biophysics and Biochemistry.

Instrumentation. LC-MS measurements were carried out on an Agilent 6550A Q-TOF (Yale West Campus Analytical Core). NMR spectroscopy was performed on an Agilent DD2 400 MHz spectrometer with an Agilent OneNMR probe. Analysis of fluorescent RNAs was carried out on a GE Healthcare Typhoon FLA 9500. Sequencing was performed on illumina HiSeq 2500 and illumina HiSeq 4000 instruments at the Yale Center for Genome Analysis (YCGA).

LC-MS analysis of nucleosides. To a solution of s⁴U (50 μM) and ammonium bicarbonate (10 mM) was added TFEA (600 mM). mCPBA (10 mM) was dissolved in ethanol and added dropwise to the reaction mixture. After 1 h at 25 °C, the reaction was analyzed by reverse-phase LC-MS with a Hypersil GOLD column (Thermo, 3 μm, 160 x 2.1 mm) using chromatography conditions described previously.⁹ Masses were collected using positive ion mode, and extracted ions were identified and integrated using Agilent MassHunter software.

Nuclear magnetic resonance analysis of nucleobase chemistry. 4-thiouracil (4.3 mg, 1 equiv.) was dissolved in DMSO-d₆ and TFEA (3.4 μl, 1.3 equiv.) was added to the solution. After mixing, a solution of NaIO₄ in DMSO-d₆ (12.3 mg, 1.7 eq.) was added to the nucleobase and amine solution, and the reaction was allowed to proceed at 45 °C for 4 h. ¹H NMR spectra were processed using the MestReNova software.

NotI restriction endonuclease assay. An RNA containing a single s⁴U nucleotide was in vitro transcribed (IVT) from a synthetic DNA template (see Supplementary Table 1) strand using T7 RNA polymerase and s⁴UTP in place of UTP for 16 h at 37 °C. The reaction mixture was treated with TURBO DNase for 1 h at 37 °C. The RNA was purified using denaturing PAGE, and the resulting band was extracted by crushing the gel slice and soaking it in extraction buffer (1 mM EDTA, 1 mM DTT, 20 mM Tris, 300 mM NaOAc pH 5.2) at 4 °C for 4 h. The supernatant was passed through a 0.45 μm syringe filter, and the RNA was ethanol precipitated and washed with 75% ethanol before resuspension in nuclease-free water.

IVT RNAs were screened for optimal TimeLapse chemistry as follows. RNA (120 ng) was added to a mixture of amine and water. A solution of oxidant was then added dropwise, and the reaction mixture was incubated at the temperature and time indicated (see Supplementary Fig. 4). The RNA was then ethanol precipitated and washed three times with 75% ethanol before resuspension in nuclease-free water.

After chemical treatment, IVT RNA (50 ng) was reverse transcribed with SuperScript III according to the manufacturer’s directions. The cDNA was PCR amplified for 30 cycles with a fluorescent forward primer, then it was amplified an additional 2 cycles using 1/5 of the previous PCR reaction material with non-labeled primers. The amplified PCR product was then incubated with NotI HF for 1 h at 37 °C. The fluorescent products were visualized using native PAGE followed by scanning with a Typhoon FLA imager, and the proportion of cut product was determined relative to a positive control (with C in the RNA instead of s⁴U) using ImageJ.

Primer-extension assay. IVT RNA containing a single s⁴U nucleotide (200 ng RNA) was treated with TimeLapse chemistry and purified as described above. Chemically treated IVT RNA (34 ng) was then annealed to a Cy5 5’ end-labeled primer, and reverse transcription was performed according to manufacturer’s instructions using the SmartScribe First Stand cDNA Synthesis kit (15 min). The reaction was then treated with RNase H, and the fluorescent products were visualized using nuclease PAGE followed by scanning with a Typhoon FLA imager. Full-length and truncated RT products were quantified by densitometry using ImageJ.

Targeted TimeLapse sequencing. MEF cells were grown at 37 °C in DMEM containing 10% FBS and 1% P/S at approximately 60% confluence, and the media was replaced with media supplemented with s⁴U (700 μM). After 2 h, the cells were rinsed with PBS, suspended in TRIzol reagent, and stored overnight at −80 °C. Following chloroform extraction, total RNA was ethanol precipitated.
including 1 mM DTT to prevent oxidation of the s^4U RNA and washed with 75% ethanol. Total RNA was resuspended and treated with TURBO DNase, then extracted with acidic phenol:chloroform:isoamyl alcohol and ethanol precipitated and washed as described above. Isolated total RNA was added to a mixture of TFEA (600 mM), EDTA (1 mM) and sodium acetate (pH 5.2, 100 mM) in water. A solution of NaOCl (10 mM) was then added dropwise, and the reaction mixture was incubated for 1 h at 45 °C. Potassium chloride (300 mM) and sodium acetate (pH 5.2, 300 mM) were added, and the reaction mixture was allowed to stand on ice for 10 min before centrifugation (>10,000 r.p.m., 30 min, 4 °C) to precipitate remaining periodate. The RNA in the supernatant was then ethanol precipitated and washed three times with 75% ethanol before resuspension in nuclease-free water. The chemically treated RNAs were then reverse transcribed using a mixture of mouse Actb- and Gapdh-specific mRNA RT primers (see Supplementary Table 1). The resulting cDNA was then amplified with Phusion polymerase using corresponding forward PCR primers to produce PCR amplicons approximately 150 nt in length. An Illumina sequencing library was constructed using the Illumina TruSeq Index adapters. Paired-end 75 bp sequencing was performed on an Illumina HiSeq 2500 instrument. Sequencing reads were trimmed to remove adaptor sequences and aligned to the mouse genome using Bowtie2 (ref. 28). Aligned reads were parsed to identify mutations at each nucleotide position in the Actb and Gapdh mRNAs using a published software package29. Raw mutation probabilities were determined by dividing the number of recorded mutation events by the number of reads at that position. Mutation probabilities were normalized to appropriate control samples and filtered by read depth (only positions with depth >3,000 were included in analyses). Analyses and figure plot generation were performed in R using the tidyverse, corplot, and multiplot packages30,31. The enrichment in mutation rates was tested for significance using a two-sided Wilcoxon test. Targeted sequencing was performed in duplicate using biologically distinct samples.

Targeted TimeLapse-seq of K562 RNA was performed similarly with the following exceptions. Cells were grown at 37 °C in RPMI containing 10% FBS and 1% P/S. At approximately 50% confluence, a solution of NaOCl (10 mM) was added to 10^n cells/mL in a 96-well microtiter plate and allowed to recover overnight. Cells were then treated in triplicate with increasing concentrations of s^4U (0–1 mM) for 1 h. Total RNA was isolated and chemically treated as described previously. The chemically treated RNAs were then reverse transcribed using a mixture of human MYC-specific mRNA RT primers (see Supplementary Table 1). A targeted sequencing library was prepared and analyzed as described above.

Cell viability. MEF cells were grown at 37 °C in DMEM containing 10% FBS and 1% P/S. Cells were plated at 10^6 cells/mL in a 96-well microtiter plate and allowed to recover overnight. Cells were then treated in triplicate with increasing concentrations of s^4U (0–1 mM) for 1 h, and the ATCC MTT Cell Proliferation Assay kit was used according to manufacturer’s instructions to assess cell viability.

Transcriptome-wide TimeLapse-seq. MEF cells were grown at 37 °C in DMEM containing 10% FBS and 1% P/S. At approximately 60% confluence, the media was replaced and supplemented with s^4U (1 mM). The cells were incubated at 37 °C for 1 h, at which point total RNA was isolated and chemically treated as described in the targeted sequencing section. For heat shock analyses, at approximately 60% confluence, the media was replaced and supplemented with s^4U (1 mM), and heat-shocked cells were incubated at 42 °C for 1 h. RNA was prepared as described for the Targeted TimeLapse-seq libraries. For each sample, 10 ng of total RNA was used to construct a sequencing library using the Clontech SMARTer Stranded Total RNA-Seq kit (Pico Input) with ribosomal cDNA depletion. Paired-end 100 bp sequencing was performed on an Illumina HiSeq 4000 instrument. TimeLapse-seq was performed in duplicate using biologically distinct samples for experimental samples both with and without heat shock. Raw and processed sequencing data have been submitted to the GEO database.

TT-TimeLapse-seq. K562 cells were grown at 37 °C in RPMI containing 10% FBS and 1% P/S. At approximately 50% confluence, the media was supplemented with s^4U (1 mM). The cells were incubated at 37 °C for 5 min, at which point total RNA isolation and genomic DNA depletion were performed as described above. 50 µg of total RNA was subjected to MTS chemistry, followed by biotinylation and streptavidin enrichment essentially as previously described3 with the following modification: after SAV beads were washed three times with high-salt wash buffer (1 M NaCl, 100 mM Tris pH 7.4, 10 mM EDTA, 0.05% Tween), beads were incubated in TE buffer (10 mM Tris pH 7.4, 1 mM EDTA) at 55 °C for 15 min, followed by two washes with prewarmed 55 °C TE buffer. After elution from SAV beads, enriched RNA was purified using one equivalent volume of Agencourt RNAClean XP beads according to manufacturer’s instructions instead of purification by ethanol precipitation. Enriched RNA and input RNA were chemically treated as previously described. Chemically treated RNA was purified using 1 equivalent volume of Agencourt RNAClean XP beads according to manufacturer’s instructions. Purified material was then incubated in a reducing buffer (10 mM DTT, 100 mM NaCl, 10 mM Tris pH 7.4, 1 mM EDTA) at 37 °C for 30 min, followed by a second RNAClean bead purification. For each sample, all enriched material or 10 ng of total RNA input was used to construct a sequencing library using the Clontech SMARTer Stranded Total RNA-Seq kit (Pico Input) with ribosomal cDNA depletion. Paired-end 150 bp sequencing was performed on an Illumina HiSeq 4000 instrument. TimeLapse-seq was performed in duplicate using biologically distinct samples for experimental samples. Raw and processed sequencing data have been submitted to the GEO database.

Samples for TimeLapse-seq analysis of K562 mRNA. K562 cells were grown as previously described. At approximately 50% confluence, the media was supplemented with s^4U (100 µM). The cells were incubated at 37 °C for 4 h, at which point total RNA was isolated using the RNAeasy mini kit with the following modifications: buffers RLT and RPE were supplemented with 1% final 2-mercaptoethanol (BME); an additional 80% EtOH wash was performed after the RPE step; and the column was spun at maximum speed for 5 min to dry before elution with water. The isolated RNA was then chemically treated and purified as previously described. For each sample, 10 ng of total RNA was used to construct a sequencing library using the Clontech SMARTer Stranded Total RNA-Seq kit (Pico Input) with ribosomal cDNA.

doi:10.1038/nmeth.4582
K562 cells were grown as described above. At approximately 50% confluence, cells were treated in duplicate with actinomycin D (2 μg/mL final) for 30 min, 1 h, 3 h, 5 h, and 9 h, or left untreated. Total RNA isolation and genomic DNA depletion were then performed as previously described. RT was performed using the SuperScript VILO cDNA synthesis kit, and qPCR was performed using primers specific to ACTB, DHX9, and ASXL1. qPCR cv values for DHX9 and ASXL1 were then averaged and normalized to those of ACTB for each timepoint. The normalized fraction remaining was estimated for each primer pair by dividing the relative abundance of each timepoint by the relative abundance at t = 0.

Sequencing alignment and mutation analysis. Reads were filtered for unique sequences using FastUniq32, trimmed using cutadapt33 to remove Illumina adaptor sequences filtering for reads greater than 20 nt (−minimum-length = 20) and aligned to the mouse GRCm38 or human GRCh38 genome and transcriptome annotations using HISAT234, using default parameters and–mp 4,2. Files were further processed with Picard tools (http://broadinstitute.github.io/picard/) including FixMateInformation, SortSam, and BuildBamIndex. The samtools35 software was used to retain only reads that aligned uniquely (flag: 83/163, 99/147), with MAPQ ≥ 2, and without insertions (because of ambiguity in mutational analysis) for further analysis.

Reads that uniquely map to the human GRCh38 version 26 (Ensembl 88) or mouse GRCm38 (p6) were identified using HTSeq-count using union mode36. Reads mapping to only mature isoforms or to anywhere in the gene body were determined separately and compared to identify intron-only reads. To determine the number of uridine residues inferred from each read, and the sites of T-to-C mutations, the aligned bam files were processed in R using Rsamtools (http://bioconductor.org/packages/release/bioc/html/Rsamtools.html), and the sites and numbers of mutations were determined using a custom R function (available upon request). Only mutations at positions with a base quality score of greater than 45 that were at least 3 nt from the end of the read were counted. Reads were excluded where there were greater than five T-to-C mutations, and these mutations did not account for at least one-third of the observed mutations (NM tag). Without adequate filtering, SNPs could interfere with TimeLapse analysis. To identify sites of SNPs (or RNA modifications that could be misidentified as TimeLapse mutations), we used the following two strategies. First, we identified T-to-C SNP sites in control samples using bcftools37 with default options and excluded these sites from our analysis. Second, we compiled locations where T-to-C mutations were high in non-s4U-treated controls and excluded these sites from analysis. Once the putative SNPs were filtered, the total number of unique mutations in each read pair was counted. To examine the distribution of reads with each minimum number of T-to-C mutations, the bam files were filtered using Picard tools. To make genome-coverage tracks, STAR aligner (inputAlignmentsFromBam mode, outWigType bedGraph) was used, and the tracks were normalized using factors derived from RNA-seq analyses using values from DESeq2 (estimateSizeFactors)38. Tracks were converted to binary format (toTDF, IGVtools) and visualized in IGV39.

Transcriptional inhibition. K562 cells were grown as described above. At approximately 50% confluence, cells were treated in duplicate with actinomycin D (2 μg/mL final) for 30 min, 1 h, 3 h, 5 h, and 9 h, or left untreated. Total RNA isolation and genomic DNA depletion were then performed as previously described. RT was performed using the SuperScript VILO cDNA synthesis kit, and qPCR was performed using primers specific to ACTB, DHX9, and ASXL1. qPCR cv values for DHX9 and ASXL1 were then averaged and normalized to those of ACTB for each timepoint. The normalized fraction remaining was estimated for each primer pair by dividing the relative abundance of each timepoint by the relative abundance at t = 0.

Sequencing alignment and mutation analysis. Reads were filtered for unique sequences using FastUniq32, trimmed using cutadapt33 to remove Illumina adaptor sequences filtering for reads greater than 20 nt (−minimum-length = 20) and aligned to the mouse GRCm38 or human GRCh38 genome and transcriptome annotations using HISAT234, using default parameters and–mp 4,2. Files were further processed with Picard tools (http://broadinstitute.github.io/picard/) including FixMateInformation, SortSam, and BuildBamIndex. The samtools35 software was used to retain only reads that aligned uniquely (flag: 83/163, 99/147), with MAPQ ≥ 2, and without insertions (because of ambiguity in mutational analysis) for further analysis.

Reads that uniquely map to the human GRCh38 version 26 (Ensembl 88) or mouse GRCm38 (p6) were identified using HTSeq-count using union mode36. Reads mapping to only mature isoforms or to anywhere in the gene body were determined separately and compared to identify intron-only reads. To determine the number of uridine residues inferred from each read, and the sites of T-to-C mutations, the aligned bam files were processed in R using Rsamtools (http://bioconductor.org/packages/release/bioc/html/Rsamtools.html), and the sites and numbers of mutations were determined using a custom R function (available upon request). Only mutations at positions with a base quality score of greater than 45 that were at least 3 nt from the end of the read were counted. Reads were excluded where there were greater than five T-to-C mutations, and these mutations did not account for at least one-third of the observed mutations (NM tag). Without adequate filtering, SNPs could interfere with TimeLapse analysis. To identify sites of SNPs (or RNA modifications that could be misidentified as TimeLapse mutations), we used the following two strategies. First, we identified T-to-C SNP sites in control samples using bcftools37 with default options and excluded these sites from our analysis. Second, we compiled locations where T-to-C mutations were high in non-s4U-treated controls and excluded these sites from analysis. Once the putative SNPs were filtered, the total number of unique mutations in each read pair was counted. To examine the distribution of reads with each minimum number of T-to-C mutations, the bam files were filtered using Picard tools. To make genome-coverage tracks, STAR aligner (inputAlignmentsFromBam mode, outWigType bedGraph) was used, and the tracks were normalized using factors derived from RNA-seq analyses using values from DESeq2 (estimateSizeFactors)38. Tracks were converted to binary format (toTDF, IGVtools) and visualized in IGV39.

Secondary structure analysis. Aligned reads from the 4h K562 TimeLapse-seq experiment overlapping the 5’ stem loop of 7SK were extracted using samtools. A Python script developed for analyses of chemical probing data (RTEeventsCounter42) was used to calculate the U-to-C mutation frequency for each uridine nucleotide. These frequencies were normalized by subtracting mutation frequencies of control samples that were not subjected to TimeLapse chemistry. The frequencies of mutations at each position were binned and mapped onto a conformational model of this region of human 7SK37. Each nucleotide was classified as either single stranded or basepaired. A two-sided Wilcoxon test was used to determine the significance of differences between mutation rates of the basepaired and single-stranded nucleotides.

Estimation of the fraction of new transcripts and transcript half-lives. Two different models were used to examine the mutation distribution in TimeLapse-seq data set: a simpler Poisson model (which does not take into account the uridine content of different reads) and a binomial model that does take the number of uridines into account. We obtained consistent results from both models. For the simpler Poisson model, for each sample (s), the distribution of T-to-C mutations (Yi) was determined in each read, and the reads were grouped based on the transcripts to which they map. A negative control sample (no s4U treatment) was used to estimate the background rate of read pairs containing T-to-C mutations that map to each transcript. These frequencies depended on the cell line used (MEF samples required higher s4U treatment to obtain similar levels of mutations compared to K562 cells) as well as the sequencing experiment (different samples led to different background rates independent of chemistry or s4U treatment). See Supplementary Note. The mutation rate and fraction of new transcripts was modeled as a two-component mixture of Poisson distributions with probability mass function:

\[
f(y | \lambda_1, \lambda_2, \theta_n) = \theta_n \text{Poisson}(y; \lambda_2) + (1 - \theta_n) \text{Poisson}(y; \lambda_1)
\]

where \(\theta_n\) is the fraction of new transcripts, \(\lambda_2\) is the rate of background mutations (determined from −s4U controls), \(\lambda_n\) is the rate of mutations found in new transcripts, and \(y_i\) is the number of passing T-to-C mutations found in read \(i\). Reasonable estimates of these values could be approximated by examining the mutation rates in fast turnover RNAs such as introns. To obtain more objective estimates of the global parameters, \(\lambda_2\) and \(\lambda_n\) while allowing for low levels of transcript-to-transcript variability, we used a Bayesian hierarchical modeling approach using RStan software (Version 2.16.2 (ref. 40)) that uses no–U-turn Markov Chain Monte Carlo (MCMC) sampling. To estimate a global mean and s.d. for \(\lambda_2\) and \(\lambda_n\), we used weakly informative priors (see below). We estimated gene-specific rates by drawing from the global mean and s.d., with a mixing rate with an uninformative
prior ($\theta_n \sim \text{Uniform}(0,1)$) where the mixing rate ($\theta_n$) estimates the fraction of each transcript that was new:

Global parameters:

- $\lambda_{n,\text{old}} \sim \text{Normal}(\mu = 0, \sigma = 1)
- \lambda_{n,\text{new}} \sim \text{Normal}(\mu = 0, \sigma = 10)
- \lambda_{n,\text{old}} \sim \text{Normal}(\mu = 0, \sigma = 10)
- \lambda_{n,\text{new}} \sim \text{Normal}(\mu = 0, \sigma = 10)
- \lambda_n \sim \text{Normal}(\mu = \lambda_{n,\text{old}}, \sigma = \lambda_{n,\text{new}})$

$$I_g = \begin{cases} 0 & \text{if controls} \\ 1 & \text{otherwise} \end{cases}$$

$g \in \{1, 2, \ldots, n_{\text{genes}}\}$

Priors:

- $\lambda_{n,g} \sim \text{Normal}(\mu = \lambda_{n,\text{old}}, \sigma = \lambda_{n,\text{old}})$
- $\lambda_{o,g} \sim \text{Normal}(\mu = \lambda_{n,\text{new}}, \sigma = \lambda_{n,\text{new}})$

for $i \in \{1, 2, \ldots, n_g\}$:

$$f_g(y_i | \theta_{ng}, \lambda_{og}, \lambda_{ng}) = \prod_{i=1}^{n_g} \left( IS\theta_{ng}\text{Poisson}(y_i | \lambda_{ng}) + \left[ 1 - IS\theta_n \right] \text{Poisson}(y_i | \lambda_{og}) \right)$$

Attempts to model entire TimeLapse-seq data sets using this approach were computationally challenging, but we found that consistent results were obtained using 20 representative transcripts from each sample. The majority of these transcripts were chosen randomly from all reasonably expressed transcripts (>$200$ reads), but we included few transcripts that were hand chosen to ensure the modeling included both fast and slow turnover RNAs such as Myc and Actb. The results using 20 transcripts were consistent with results from 200 transcripts. In the case of the MEF samples shown in Figure 2, the $\lambda_n$ was estimated as 0.07 mutations/read (50% credible interval 0.062-0.074), and $\lambda_n$ was estimated as 2.3 mutations/read (2.298 mutation/read, 50% CI 2.10-2.30 for heat shock; 2.288 mutation/read, 50% CI 1.90-2.29 for untreated).

Once these global parameters were determined, they were used to estimate the fraction of new transcripts ($\theta_{\text{new}}$), using expectation maximization by minimizing the log likelihood using the nlm function in the MASS package in R:

$$- \sum_{i=1}^{n_g} \log(f(y_i | \theta_n, \lambda_o, \lambda_n))$$

The 95% Wald confidence interval was calculated using the Hessian (nlm option hessian = TRUE), to calculate:

$$\hat{\theta}_{\text{new}} \pm z_{0.975} \times \text{std err}(\hat{\theta}_{\text{new}})$$

To ensure the mutations were both s^4U-treatment and TimeLapse-chemistry dependent, we only included transcripts where there was sufficient data (reads $> 100$ counts in at least two samples), and where the fit converged ($-0.05 < \theta_n < 0.05$; hessian $> 1,000$). The inferred new read counts were determined by multiplying the estimated fraction of new transcripts by the total RNA-seq transcript count. Correlations between replicates were determined using the log_{10}-transformed counts (Supplementary Fig. 8). While the reproducibility of the data was generally high when all converged transcripts were included (Pearson's $r > 0.91$), filtering for transcripts with at least 75 inferred new reads provided slightly more reproducible results ($n = 3,603, r = 0.934$), and this filter was used for further analysis.

To account for differences in the number of uridine residues in each read pair, an alternative model was used based on the binomial distribution. Specifically, the data were modeled as mixture of two binomial distributions:

$$f(y | \theta_n, p_0, p_n) = \theta_n \text{Binom}(y; p_n, n_{\text{old}}) + (1 - \theta_n) \text{Binom}(y; p_0, n_{\text{new}})$$

where $p_0$, $p_n$ are the probabilities of mutation at each uridine nucleotide for old and new transcripts, and $n_{\text{old}}$ is the number of uridines observed for read $i$. To determine the global mutation rate, we used Bayesian hierarchical modeling as described above for the Poisson model but using a mixture of binomial distributions. From this analysis, we estimate the background mutation rate ($p_0$) to be 0.0012 mutations/uridine (50% CI 0.00121, 0.00123) and the mutation rate for new reads ($p_n$) to be 0.0332 mutations/uridine (50% CI 0.0329, 0.0335). In other words, ~0.1% of Us are mutated to C in pre-existing reads, and in new reads ~3% of Us are mutated to C. Using these global parameters, the distributions of individual genes were fit with nlm similarly to what is described above, except by minimizing the log likelihood of the binomial model instead:

$$- \sum_{i=1}^{n_g} \log(f(y_i | \theta_n, p_0, p_n, n_{\text{old}}))$$

In addition to computing the confidence interval using the hessian, we also examined the quality of the fit by plotting the observed frequency of mutations in each replicate in the TimeLapse data (gray points in distribution plots) to a simulated distribution of the expected new and old reads based on the binomial model (Fig. 2b and Supplementary Fig. 16). Estimates of the fraction new were highly similar between those determined using the binomial model and the Poisson model.

To account for any specific loss of transcripts that might arise from biased loss of s^4U-RNA transcripts independent of TimeLapse chemistry, or TimeLapse-depended loss due to reverse-transcription termination, we developed a means of estimating the loss of fast-turnover transcripts in the data. This correction was only used when estimating transcript half-lives after observing a modest, but statistically significant loss of reads from high turnover RNAs (see Supplementary Fig. 3d). To estimate the fraction of new reads missing, we used the R package nlm to fit the equation:

$$s_y N_y \left( 1 + \theta_n \left( \frac{x}{1-x} \right) \right) = s_o N_o$$

where $s_y$ and $s_o$ are scale factors that adjust for library sizes determined using DESeq2 with the total (RNA-seq) transcript counts for the experimental sample and control, respectively; $N_y$ and $N_o$ are the counts for each transcript; and $\theta_n$ is the unadjusted fraction new of each transcript. This equation was fit using transcripts where $0.8 < \theta_n$ for K562 RNA, but $0.5 < \theta_n$ in the case of MEF RNA (the shorter s^4U treatment lead to fewer transcripts with high $\theta_n$, so the threshold was lowered to increase the number of transcripts).
In the case of the comparison shown in Supplementary Figure 3d, the adjustment factor determined for chemistry-induced dropout was ~5% (i.e., \( x = 0.05 \) in the equation above, which leads a transcript with 75% new reads to be adjusted to 79% and a transcript with 25% new reads would be adjusted to 26% new reads).

The transcript half-lives were determined using the adjusted fraction of new RNA assuming a simple exponential model of their kinetics. The half-life values were compared to similar reports and the \( r^2 \) determined using the lm function in R.

**Gene ontology analysis.** GO analysis from the PANTHER database (version 12.0) was performed using a statistical over-representation test (default parameters) on the complete biological process annotation set using the top 10% slow or top 10% fast turnovers RNAs in our 1 h MEF TimeLapse-seq data as determined by the half-life analyses described above.

**Differential expression analysis.** Differential expression analysis was performed using DESeq2. To examine the inferred differences in the new transcript pool based on TimeLapse mutations, we used the unadjusted estimates of the fraction of new RNA to infer the number of counts resulting from new transcripts as described above. As TimeLapse-seq data are internally controlled, we used the size factors determined from total counts to scale each data set (i.e., we ran DESeq2 on the total RNA-seq data and used the sizeFactors function to scale the inferred new RNA counts to the RNA-seq-determined values) with default conditions including the Benjamini–Hochberg \( P \) value \( (p_{\text{adj}} \) in text). RNA-seq analysis was performed on all reads (i.e., reads that had zero or more T-to-C mutations) using DESeq2 with default parameters.

**Estimation of contaminating reads in TT-TimeLapse-seq.** Reads from TT-TimeLapse-seq were processed and analyzed as for TimeLapse-seq. Junction-containing reads were determined from the presence of “N” characters in the CIGAR string in the aligned bam file using bamtools (version 2.3, https://hcc-docs.unl.edu/display/HCCDOC/BamTools). The levels of contaminating reads were estimated by assuming the contaminating reads have the same ratios as RNA-seq data, and that reads with three or more mutations constitute the true ratio of reads. We use of reads with three or more mutations as true positives, because the probability of a read containing three or more mutations without \( s^t\) is \(<10^{-5} \). We used the fraction of intron- or junction-containing reads for the RNA-seq data \( (r_0) \), the total in the true positive population \( (r_p) \), and the total for each population \( (r_x) \). In each analysis, we only considered reads that had nonzero ratios and ratios that were less than one. The fraction of reads from contamination \( (c_x) \) was then estimated:

\[
c_x = \frac{r_p - r_0}{r_x - r_0}
\]

For comparisons with the TT-TimeLapse-seq data presented here, the data from Schwalb et al \( ^7 \), \( (\text{SRR4000390, SRR4000391 and SRR4000397}) \) were aligned and processed using the same pipeline described for TimeLapse-seq. For this comparison, we reprocessed our TT-TimeLapse-seq data using only 75 nt of each read, and this trimming was performed on fastq files before alignment. This step was performed because the probability of a sequencing read containing a splice junction or being an intron-only read is dependent on the read length. Otherwise, all processing was handled equally between data sets.

**Code availability.** All software and parameters used are described above, and custom scripts and functions are available upon request.

**Life Sciences Reporting Summary.** Further information on experimental design is available in the Life Sciences Reporting Summary.

**Data availability.** Data are available in the Gene Expression Omnibus (GEO) under accession number GSE95854.

27. Schofield et al. Protocol in Protocol Exchange DOI 10.1038/protx.2018.004.
28. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
29. Sexton, A.N., Wang, P.Y., Rutenberg-Schoenberg, M. & Simon, M.D. Interpreting reverse transcriptase termination and mutation events for greater insight into the chemical probing of RNA. Biochemistry 56, 4713–4721 (2017).
30. Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A. & Weeks, K.M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat. Methods 11, 959–965 (2014).
31. Wei, T. & Simko, V. RPackage “corplot”: visualization of a correlation matrix (Version 0.84) (2017).
32. Xu, H. et al. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS One 7, e52249 (2012).
33. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBOnet.journal 17, 10–12 (2011).
34. Kim, D., Langmead, B. & Salzberg, S.L. HISAT2: a fast splice aligner with low memory requirements. Nat. Methods 12, 357–360 (2015).
35. Li, H. et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009).
36. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
37. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
38. Robinson, J.T. et al. Integrative genomics viewer. Nat. Biotechnol. 29, 24–26 (2011).
39. Van Herreweghe, E. et al. Dynamic remodelling of human 7SK snRNP controls the nuclear level of active P-TEFb. EMBO J. 26, 3570–3580 (2007).
40. Carpenter, B. et al. STAN: a probabilistic programming language. J. Stat. Softw. 76, 1–32 (2017).
41. Thomas, P.D. et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13, 2129–2141 (2003).
42. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J. R. Stat. Soc. B Stat. Methodol. 57, 289–300 (1995).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

1. Sample size
   Describe how sample size was determined.
   All targeted sequencing samples were prepared in biological duplicates with the exception of 10μM and 40μM K562 targeted samples (Supplementary Fig. S6e). All TimeLapse-seq samples for both MEF and K562 cells and TT-TimeLapse-seq samples were performed as biological duplicates as is noted in the online methods. No sample size calculation was performed.

2. Data exclusions
   Describe any data exclusions.
   Cases where data were excluded are described in the text and methods. In the targeted sequencing section, data points were excluded if they did not meet a depth threshold of 3000 reads. The criteria for exclusion of reads and mutations in genome wide TimeLapse-seq data are described in the text and methods. Reads were filtered for unique sequences before alignment to either the mouse GRCm38 or human GRCh38 genome. Non-unique aligned reads were removed. Insertion-containing reads were not considered in mutational analyses. Sites of mutations were only considered if their base quality was 45 or above. Reads were removed if they contained greater than five T-to-C mutations and these mutations did not account for at least one third of the observed mutations (NMtag). Mutation data points were also removed if their base position was identified as a SNP. When estimating fraction new, transcripts were only included if at least two samples had more than 100 counts.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   The efficiency of oxidative-nucleophilic-aromatic-substitution was found to be reproducible in multiple screens assayed by our restriction endonuclease assay. All TimeLapse-seq and TT-TimeLapse-seq samples were performed as biological duplicates and the correlation for each duplicate analysis is presented.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   Randomization was not relevant to our study.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   Blinding was not relevant to our study. After upstream treatment of cells and isolated RNA, all samples were handled and analyzed with the same protocols and pipelines.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

Nature Methods: doi:10.1038/nmeth.4582
6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

| n/a | Confirmed |
|-----|-----------|

- The exact sample size \((n)\) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.).
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
- A statement indicating how many times each experiment was replicated.
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section).
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons.
- The test results (e.g. \(p\) values) given as exact values whenever possible and with confidence intervals noted.
- A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range).
- Clearly defined error bars.

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study. All parameters used to analyze data are described in methods section.

Custom scripts implementing these calculations are available upon request as is noted in the revised online methods.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

- No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

- No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

- MEF cell line was described in Yildirim et al. 2012, Nat. Struct. Mol. Biol.
  K562 cell line was a generous gift of the Slavoff lab, Yale University Department of Chemistry, New Haven, CT, 06511, USA.

b. Describe the method of cell line authentication used.

- No further authentication beyond what is described in Yildirim et al. 2012, Nat. Struct. Mol. Biol. was performed.

c. Report whether the cell lines were tested for mycoplasma contamination.

- Cell lines were not tested for mycoplasma contamination.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

- No commonly misidentified cell lines were used.
Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
   Provide details on animals and/or animal-derived materials used in the study.
   No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants
   Describe the covariate-relevant population characteristics of the human research participants.
   Study did not involve human research participants.