Ubiquitous mRNA decay fragments in *E. coli* redefine the functional transcriptome

Lydia Herzel [ID], Julian A. Stanley [ID], Chun-Chen Yao and Gene-Wei Li[*]

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

ABSTRACT

Bacterial mRNAs have short life cycles, in which transcription is rapidly followed by translation and degradation within seconds to minutes. The resulting diversity of mRNA molecules across different life-cycle stages impacts their functionality but has remained unresolved. Here we quantitatively map the 3' status of cellular RNAs in *Escherichia coli* during steady-state growth and report a large fraction of molecules (median>60%) that are fragments of canonical full-length mRNAs. The majority of RNA fragments are decay intermediates, whereas nascent RNAs contribute to a smaller fraction. Despite the prevalence of decay intermediates in total cellular RNA, these intermediates are underrepresented in the pool of ribosome-associated transcripts and can thus distort quantifications and differential expression analyses for the abundance of full-length, functional mRNAs. The large heterogeneity within mRNA molecules *in vivo* highlights the importance in discerning functional transcripts and provides a lens for studying the dynamic life cycle of mRNAs.

INTRODUCTION

Cellular mRNAs are present in different forms—nascent, full-length, partially degraded, and others. Whereas full-length molecules are often depicted as the canonical mRNAs in the central dogma of molecular biology, the other forms are necessary intermediates of their birth-death cycle and have potential to (mis)interact with the same pool of factors that regulate mRNA functionality, such as translation. A complete molecular census of the mRNA population is thus critical for quantifying the functional transcriptome, while also providing an opportunity to reveal the life cycle of mRNAs *in vivo*.

In bacteria, functional half-lives of mRNAs are often short, which could in principle give rise to a large heterogeneity among the transcripts that co-exist at the same time ([1–3]). After transcription initiates, mRNAs remain nascent for about a minute (considering a typical operon of 2 kilobases (kb) and a chain elongation rate of 45 nucleotides (nt)/s ([4,5])). Meanwhile, RNA endonucleases, such as RNase E in *Escherichia coli*, typically cleave mRNAs within a few minutes ([6–10]). Following endonuclease cleavage, the upstream RNA fragments are digested by exonucleases, such as the 3'-to-5' exonucleases PNPase and RNase II in *E. coli*, whereas the downstream fragments are further cleaved endonucleolytically ([11]). The timescales of these subsequent events are less characterized but can nevertheless determine the extent of decay intermediates present in the transcriptome ([12]).

Amidst the heterogeneous RNA pool, translation takes place concurrently during transcription and mRNA decay in bacteria ([13]). Whereas translation of nascent mRNAs is beneficial in *E. coli* by preventing Rho-dependent transcription termination, translation of mRNA decay intermediates may produce aberrantly truncated proteins and require rescue of ribosomes that are stalled at the end of RNA fragments ([14–16]). In the meantime, some mRNA decay intermediates are translation-incompetent but often included by RNA quantification methods as the total cellular RNA ([17]). These considerations highlight the need to quantitatively resolve the forms of mRNA molecules across the transcriptome.

Existing methodologies for quantification primarily rely on regional information and are thus blind to distinct forms of RNAs. RT-qPCR and conventional RNA-seq detect ∼100 bases at a time. Although Northern blotting could distinguish RNAs of different lengths, it only detects transcripts containing the short stretch of complementary sequence that is probed. Further, partial RNA fragments in Northern blots are often spread out over a large size range that dilutes the signal compared to full-length RNAs. Recently, several high-throughput single-molecule strategies have been employed to explore major RNA isoforms (SMRT-Cappable-seq, SEnd-seq, Oxford Nanopore technologies) ([18–20]), which could in principle be used to probe the presence of nascent or partially degraded RNAs. How-

ever, these methods involve at least one size-selection step that restricts the range of RNAs (or cDNAs) analyzed. Therefore, the full spectrum of RNA molecules present in the cell remains unmapped.

Here, we present a molecular census of RNA 3' ends across the *E. coli* transcriptome. The 3' terminal position of RNAs varies during their life cycle, and the relative representations can be quantified using methods that are not necessarily restricted to a limited size range (21,22). We find that, during steady-state growth, the majority of RNA molecules do not possess the mature 3' ends of the corresponding transcription units (TUs). Instead, the 3' termini are widely distributed across positions internal to TUs. Most RNA fragments at steady-state are decay intermediates and not nascently transcribed. By contrast, translation primarily takes place on full-length and nascent RNAs. We further demonstrate that changes in the pool of translation-competent mRNAs may not be detectable by RNA-seq as they are masked by the large fraction of decay intermediates. Together, our results indicate that the *E. coli* transcriptome consists of substantial amounts of both functional and non-functional mRNAs, which are often indiscernible by existing quantification methods.

MATERIALS AND METHODS

Strains, cell growth and harvest

Strains used in this study are listed in Supplementary Table S1. To generate fresh gene deletions in the strain background MG1655, P1 transduction from the KEIO collection was done as described previously (23,24). *Escherichia coli* cells were grown in LB, MOPS complete, or MOPS minimal media supplemented with 0.2% glucose. *B. subtilis* cells were grown in MCCG media (25). To harvest cells in exponential growth, cells were back-diluted from overnight cultures into the respective growth media to enable ~10 doublings at 37°C prior to harvesting at an OD_{600} smaller or equal to 0.3 (26). To harvest cells in early stationary growth, cells were collected shortly after growth had plateaued (harvest at $OD_{600} = 2.8$ for MOPS complete and at $OD_{600} = 2.26$ for MOPS minimal). For kasugamycin treatment, kasugamycin was added at a final concentration of 1 mg/ml to cells at $OD_{600} = 0.22$ grown in MOPS complete media. After 15 min, treated and untreated cells were harvested for total RNA analysis. In general, for total RNA analysis, 5 ml of cells were transferred into 0.55 ml of ice-cold stop solution (95:5 ethanol:phenol) in a 15 ml conical tube, quickly inverted and spun at 3220 g at 4°C for 5–10 min. The supernatant was discarded and cell pellets flash-frozen in liquid nitrogen and stored at –80°C until lysis and RNA extraction. For nascent RNA and polysome analysis, 250 ml of cells were filter harvested at 37°C and scraped into liquid nitrogen (27). The cell paste was kept at –80°C until lysis and RNA extraction.

Genome version, annotation, and definition for simple and complex transcription units

For all experiments described here, the NCBI Reference Sequence and CDS annotation for the *E. coli* genome NC_000913.2 and the *B. subtilis* genome NC_000964.3 were used.

Definition of simple transcription units. We focused our analysis and quantification of transcriptional and decay intermediates to simple transcription units (TUs). Simple TUs are transcripts with only one isoform—i.e. have only one transcription start site (giving rise to the mature 5' end) and one transcription termination or processing site (giving rise to the mature 3' end). In order to classify regions in the *E. coli* and *B. subtilis* genome as simple TUs, we utilized high-coverage Rend-seq data from *E. coli* MG1655 grown in the same conditions as for our 3' end-seq experiments (MOPS-Complete and LB, harvested at OD_{600} 0.3) and *B. subtilis* 168 grown in LB until OD_{600} 0.3 (28). By enriching for mature 5' and 3' ends of RNA, Rend-seq allows to map transcript isoform ends (peaks) with single nucleotide precision (28). We called 5' and 3' peaks as described previously (28), using a threshold Z-score of 12. We considered 5' or 3' peaks that were within ±2 nt in different datasets as 1 peak, because this corresponds to the expected peak width from these data and can have maxima that vary by ±2 nt around the center (28).

We then identified all transcripts which met our criteria for being a simple TU. First, we matched any 5' and 3' peaks that were connected by expressed regions and separated by regions that did not have sufficient read coverage (defined by median coverage). This procedure created a list of potential mRNA isoforms across the genome. Second, we added the restriction that the size of each transcription unit lies within the 2–98% quantile of all isoforms (75–20872 nt for *E. coli*). This resulted in 7635 isoforms for *E. coli* from 3617 unique 5' peaks and 1681 unique 3' peaks. Third, we excluded rRNA and tRNA regions from further analysis by keeping only non-rRNA and tRNA-overlapping isoforms using bedtools intersect (rRNA & tRNA annotation Supplementary Table S3, bedtools/2.25.0 (29)). That reduced the number of isoforms to 5676. Fourth, to obtain a candidate set of simple TUs, we overlapped the isoforms with each other using bedtools intersect and only kept the 215 isoforms that did not overlap with any other ones. Lastly, we manually inspected and curated the simple TUs in the genome browsers IGV and MochiView to ensure that no apparent peaks in Rend-seq data (that did not meet our stringent peak calling criteria) were in the vicinity (30,31). These would indicate less abundant transcript isoforms. This reduced the number of simple TUs to 142. These simple TUs were used for the 3' end analysis reported in the main text (Supplementary Table S4). In total, 163 genes are expressed in simple TUs, representing 4% of protein-coding genes in the *E. coli* genome and about 10% of expressed genes, respectively.

The same approach (omitting the curation in the genome browser) yielded 418 simple TUs in *B. subtilis* that were used for the analysis shown in Supplementary Figure S3 (Supplementary Table S5). There are likely additional simple TUs that we did not capture using our approach due to our conservative criteria for identifying simple TUs and because of the focus on exponential phase in rich media conditions.

Definition of mature 3' ends in complex TUs. To extend our analysis beyond simple TUs, we identified additional (non-simple) 3' peaks that are associated with one or multiple 5' peaks upstream in Rend-seq data. To calculate the internal 3' end density proximal to these 3' peaks, only 3' peaks at least 220-nt downstream of a preceding 5' or 3' peak were included. This yielded additional 378 3' peaks with one upstream 5' peak and 406 3' peaks with multiple upstream 5' peaks, for which 364 and 392, respectively, had sufficient read coverage (Supplementary Table S8). We note that some TUs in *E. coli* are not associated with well-defined mature 3' ends, likely due to Rho-dependent transcription termination that does not have a unique position of termination. These TUs are not included in our analysis and are likely more heterogeneous.

Definition of genomic regions encoding mRNAs. For analysis of 3' and 5' end profiles corresponding to RNase E cleavages, RNA polymerase pauses, and non-templated adenylation in any protein-coding gene, we developed a coarse definition of genomic regions that encode mRNAs. To this end, we grouped CDSs based on their distance to each other (<121 nt) and strand orientation. To include untranslated regions (UTRs) of mRNAs, we extended CDSs in each direction by 60 nt, which is the typical UTR length observed in Rend-seq (Supplementary Table S6). These genomic regions can contain multiple mRNA isoforms.

Synthesis of reference RNAs for spike-in

We used a total of five reference RNAs to spike-in to our 3' end sequencing samples for quality control purposes. The two shortest reference RNAs (28 nt and 60 nt) were ordered as PAGE-purified RNA oligos from IDT (Supplementary Table S2). The remaining three RNA spike-ins were generated by *in vitro* transcription (details below).

We generated the 120-nt reference RNA by amplifying the *S. cerevisiae* gene PDR5 from genomic DNA using oligos oLH245 and oLH238 (Supplementary Table S2). We followed the PCR with an agarose gel extraction and a column purification using the Oligo Clean & Concentrator kit (Zymo Research). The DNA template was then *in vitro* transcribed using the HiScribe T7 high yield RNA synthesis kit (NEB) in a 20 µl reaction for 10 h at 37°C. We then digested the template DNA using TurboDNase at a concentration of 0.1 U/µl in 200 µl and cleaned up the reaction using the RNA Clean & Concentrator-25 kit (Zymo Research). Finally, we performed a gel extraction from a 10% polyacrylamide gel as described in (27) and quantified the resultant RNA concentration using the fluorometric Qubit RNA BR assay.

To generate the 920-nt reference RNA, the pTXB1vector (NEB) was digested with BamHI, gel purified and *in vitro* transcribed using the HiScribe T7 high yield RNA synthesis kit (NEB) in a 20 µl reaction for 10 h at 37°C. TurboDNase digestion with a final concentration of 0.1 U/µl in 200 µl, clean-up using the RNA Clean & Concentrator-25 kit (Zymo Research), agarose gel analysis and quantification with the Qubit RNA BR assay followed.

To obtain an 1800-nt reference RNA, the FLuc Control Template from the HiScribe T7 high yield RNA synthesis kit (NEB) was *in vitro* transcribed and purified as described for the linearized pTXB1 vector.

Aliquots of each reference RNA were kept in 10 mM Tris pH 7.0 at the final concentration for pooling at –80°C (one reference RNA set for one RNA extraction contained approximately 6 ng 28-nt reference RNA, 57 ng 60-nt reference RNA, 53 ng 120-nt reference RNA, 70 ng 920-nt reference RNA, 170 ng 1800-nt reference RNA). Each pooled batch was sequenced by 3' end-sequencing without spiking it into a cell suspension to serve as an input for quality normalization (four input replicates in Figure 1C, normalized to mean Φ_{end} of these four technical replicates).

RNA purification

For all total RNA extractions from cell pellets, cell pellets were thawed for 4 min in a 4°C pre-cooled tabletop centrifuge while spinning at 3220 g. Tubes were transferred on ice and residual media was removed by pipetting.

Acid-phenol:chloroform extraction (hot phenol). 500 µl acid-phenol:chloroform, pH 4.5 (with IAA, 125:24:1, Thermo Fisher Scientific) containing 29 µl of 20% SDS was prewarmed to 65°C prior adding it to 500 µl sample. The sample solution containing Acid-Phenol:Chloroform and SDS was incubated for 5 min at 65°C in an Eppendorf Thermomixer at 1400 rpm. Subsequently the solution was transferred to ice and chilled for 5 min. After a 2 min spin at room temperature and 20000 g 90% of the aqueous layer were transferred to a fresh tube and precipitated with sodium acetate/isopropanol. The sample was further purified using the RNA Clean & Concentrator-5 kit (Zymo Research) and DNase treated subsequently (details in 'RNA cleanup' section).

RNAsnap. RNA extraction using RNAsnap was done as previously described (32) with small adjustments. The cell pellets were resuspended in 500 µl RNA extraction solution (RES, 18 mM EDTA, 0.025% SDS, 1% BME, 95% Formamide, RNA spike-ins). The RES was prepared fresh each time from stock components and reference RNA aliquots were added just before use. The RES-cell suspension was vortexed for 50 s and incubated for 7 min at 95°C. For *B. subtilis* cells, the RES-cell suspension was added to 0.2 ml cold 0.1 mm Zirconia beads, vortexed for 5 min at maximum speed and incubated for 7 min at 95°C. A 5 min spin at room temperature at 20000 g followed. 400 µl of supernatant were transferred to a fresh tube containing 1.6 ml of DEPC-water. Subsequently, this 1:4 dilution was split into 4 × 500 µl aliquots for sodium acetate / isopropanol precipitation (details in 'RNA cleanup' section). Right after precipitation the sample was further purified using the RNA Clean & Concentrator-5 kit (Zymo Research) and DNase treated subsequently.

RiboPure RNA purification kit. RNA extraction with the RiboPure RNA Purification Kit (Thermo Fisher Scientific) was done as recommended by the manufacturer.

RNeasy RNA purification kit. Cell pellets were resuspended in 100 µl of 4 mg/ml lysozyme in 10 mM Tris, pH

8.0 with reference RNA and incubated for 5 min at 37°C. RNA was extracted as described in the manual without the use of gDNA (genomic DNA) columns. Instead, the sample was DNase treated and purified as described below. rRNA was removed from 20 μg of RNA with two reactions of the MICROB*Express* Bacterial mRNA Enrichment Kit (Thermo Fisher Scientific).

Trizol extraction. Cell pellets were resuspended in 100 μl of 4 mg/ml lysozyme in 10 mM Tris, pH 8.0 with reference RNA and incubated for 5 min at 37°C. 1.5 ml Trizol was added and mixed by inversion. After 5 min incubation at room temperature, 300 μl of chloroform was added. After vigorous shaking for 15–30 s the lysate was spun for 15 min at 12000 g in a pre-cooled 4°C centrifuge. About 800 μl of the aqueous phase was transferred to a fresh tube and precipitated with sodium acetate / isopropanol. As for RNAsnap, the sample was further purified using the RNA Clean & Concentrator-5 kit (Zymo Research) and DNase treated subsequently.

RNA cleanup by precipitation and column and DNase treatment

To concentrate and cleanup RNA samples, sodium acetate–ethanol or isopropanol precipitation was carried out. 3 M sodium acetate (pH 5.4, Life Technologies) was added to make up 1/10th of the sample volume. 1–2 μl of GlycoBlue (Invitrogen) was used as co-precipitant. Depending on the sample volume either 3 volumes of ice-cold 100% Ethanol (<300 μl) or 1 volume of ice-cold 100% Isopropanol was added. After thorough mixing, samples were precipitated at –80°C for at least 30 min and spun for at least 60 min in the pre-cooled 4°C centrifuge at 18213 g. RNA pellets were washed with at least 250 μl of ice-cold 80% ethanol and spun for at least 4 min. After aspiration by pipetting RNA pellets dried at room temperature for 3–5 min and were resuspended in 10 mM Tris, pH 7.0. To clean up samples further and remove traces of short DNA oligos post DNase treatment, we purified RNA samples by RNA Clean & Concentrator-5 kit (Zymo Research). If this kit was used prior to DNase treatment, elution was done with 85 μl DEPC-water. Otherwise, samples were eluted into 10 mM Tris, pH 7.0. For DNase treatment, 5 μl 10× Turbo DNase buffer and 5 μl 2U/μl TurboDNase (Thermo Fisher Scientific) were added to 85 μl of sample and incubated for 20–30 min at 37°C. RNA was cleaned up either with the RNA Clean & Concentrator-5 kit (Zymo Research) or by Ethanol precipitation.

Protein analysis

Protein samples were run on a 4–12% SDS polyacrylamide gel followed by SyproRuby staining (Thermo Fisher Scientific) and western blotting (as described in the NuPAGE Technical Guide from Thermo Fisher Scientific). Western blots of different fractions during the RNA polymerase immunoprecipitation were performed to assay efficiency of flag-tagged RNA polymerase enrichment using the monoclonal Anti-Flag M2 antibody from mouse (F1804, Sigma).

RNA polymerase immunoprecipitation

The RNA polymerase immunoprecipitation was carried out as described in (33) with the following modification. Nascent RNAs were extracted from the magnetic bead-bound fraction by adding 1.5 ml Trizol to the sample and proceeding as described above. For comparison to total RNA, 100 μl of lysate were diluted to 250 μl in lysis buffer and RNA was extracted using Trizol. Following DNase treatment and sample cleanup, rRNA was removed using the MICROB*Express* Bacterial mRNA Enrichment Kit (Thermo Fisher Scientific) and precipitated with isopropanol. RNA levels were quantified fluorometrically with the Qubit RNA BR assay. 1.8 μg of RNA was used for 3' end-seq library preparation.

Polysome gradient centrifugation

Cell paste from 250 ml of cells, that were filter-harvested at or below OD_{600} 0.3, was cryo-lysed as done for ribosome profiling (27). 1 M sodium chloride (34), 200 U/ml SUPERaseIn RNase Inhibitor and 100 U/ml TurboDNase were included in the lysis buffer in addition to the standard composition of 20 mM Tris pH 8.0, 100 mM ammonium chloride, 0.4% Triton X-100, 0.1% NP-40, 10 mM magnesium chloride and 5 mM calcium chloride. 250 μl of lysate were loaded onto a 10–55% sucrose gradient and spun for 156 min at 151263 g with a SW41 Ti rotor at 4°C. Twelve gradient fractions were collected into disposable glass tubes for fraction collection with the same instrumentation as for ribosome profiling (27). The remaining liquid and pellet in the tube were either taken as a separate fraction (Supplementary Figure S5C) or pooled as a final fraction (Supplementary Figure S5B). Each fraction was transferred to a fresh tube and precipitated with ice-cold isopropanol as described above. The pellet was resuspended in 500 μl 10 mM Tris, pH 7.0 containing 0.5% SDS. RNA was extracted subsequently using hot phenol extraction as described above. Following DNase treatment and sample cleanup, RNA levels were quantified fluorometrically with the Qubit RNA BR assay and 3–5 μg were used for Northern blot analysis and 1.8 μg for 3' end-seq library preparation.

PCR & RT-qPCR

Reverse transcription (RT) of RNA was done using SuperScript III reverse transcriptase (Thermo Fisher Scientific). The protocol recommended for the enzyme was used in either 10 or 20 μl reactions. 100 ng of RNA were used as input. Random hexamers (Thermo Fisher Scientific, 1 μl per 20 μl) were used for priming. 'no RT primer' and 'no enzyme' controls were also included to assess random priming from residual short gDNA oligonucleotides and presence of contaminating gDNA, respectively. RNA, dNTPs and RT primer were denatured at 80°C for 3 min and immediately cooled on ice for at least 1 min. Afterwards RT buffer, 0.1 M DTT and SuperaseIn (Thermo Fisher Scientific) were added in the respective amounts. Superscript III enzyme was added last (or water for 'no enzyme' controls). Samples were incubated at room temperature for 5 min, then 55°C for 45 min. RNA was hydrolyzed at 95°C with

0.1 M NaOH (final concentration) for 5 min and neutralized with HCl. cDNA samples were diluted 1:10 for further use in qPCR. For subsequent qPCRs, Kappa qPCR reaction mix (Roche) was used. In each well of a 96-well plate we added 5 μl 2× Kappa qPCR reaction mix, 3 μl of the 1 μM primer solution and 2 μl of the 1:10 diluted cDNA solution.

We assayed each sample in combination with several controls, such as a 'no RT primer', 'no enzyme' and a 'no cDNA' control, in 2–3 technical replicates. Each run was done as recommended by the manufacturer. Only samples that had a Ct value at least 10 cycles lower than the control Ct value were used for further analysis. A smaller than 10 Ct value difference to the 'no enzyme' control would suggest residual genomic DNA that would need to be reduced by a second round of DNase digestion prior RT. TurboDNase treatment as described above removed most genomic DNA. Furthermore, small fragments of DNA that could have primed as random primers (detectable by 'no RT primer' control) were cleared away efficiently by the RNA Clean & Concentrator-5 kit (Zymo Research). Hence, a second round of DNase treatment was never necessary.

To quantify gene expression differences, the double-Ct method was employed (35). Namely, Ct values were first normalized by the expression of *cysG*, a gene whose expression level did not change across the assayed conditions and that was used in previous analyses (36). After normalizing both the target strain and the wildtype strain to *cysG* we then compared the ratios of their relative abundance to estimate fold changes in the gene of interest between strains.

For conventional PCRs, Q5 DNA polymerase (NEB) was used with recommended reaction settings and reagent concentrations. 35-40 cycle PCR reactions were carried out. PCR products were analyzed by conventional 1.5% agarose gel electrophoresis. Templates for *in vitro* transcription reactions were gel purified using the QIAquick Gel Extraction kit (Qiagen).

Northern blot

Northern blot analysis was done as described in (28). DNA oligos used as probes are given in Supplementary Table S2. For total RNA analysis, 2% agarose gels were used. 1 pmol of DNA probe was used for 5' terminal γ-^{32}P labeling followed by cleanup with ProbeQuant G-50 columns (Sigma) when assessing mRNA levels. To probe highly abundant RNAs, such as *ssrA* and *rnpB,* 0.01 pmol of DNA probe sufficed and enabled full removal of signal by stripping of transcript specific probes with three 20 min washes with boiling 0.1% SDS followed by an overnight incubation at 47°C.

Labeled membranes were exposed to a phosphor storage screen (GE Life Science) for hours to days depending on the signal strength and imaged with a 635 nm laser scanner (Typhoon FLA9500, GE Life Sciences or Amersham Typhoon) with 750 V and 100 μm pixel size. By testing several exposure times of the radioactively labeled membranes to the phosphor storage screen and testing different PMT voltages in the range from 400 to 750 V, we ensured that neither the phosphor storage screen nor the Typhoon scanner saturated (none of our tested conditions yielded sat-

uration). The Amersham Typhoon Scanner Control Software 2.0.0.6 converts the original 32-bit data into a 16-bit data format for saving as tif-file that was used for further analysis in Fiji (37). Images were cropped to only include the membrane of interest. Intensity profiles were plotted per lane with a linewidth of 50 pixels. These images and profiles are shown in Figures 1E, 3C and Supplementary Figures S1A, S5D. For all Northern blot images, we included the full range of pixel intensities. The fraction of full-length RNAs was quantified semi-automatically for Northern blots that showed uniform background signal across the full membrane and showed uniform sample and run characteristics for all lanes on the gel. The median background signal above the full-length signal was subtracted. The Northern blot shown in Figure 1E showed a background distribution that increased linearly from top to bottom. Hence, we determined the median background signal above the full-length product and at the bottom of the blot, where no RNA was present and subtracted the linearly increasing profile between these two background intensities.

Upon log-transformation of the background-subtracted intensities for all lanes on one membrane, two inflection points were defined, indicating the boundaries of full-length RNA to (i) no signal above and (ii) partial transcript signal below. The signal between these two points reflected full-length RNA signal (FL) (shading in Figure 1F). The signal below the second inflection point reflected any signal of partial transcripts (P). The ratio of FL/(FL + P) was defined as the fraction of full-length RNAs F_{FL} and compared to Φ_{end} derived from 3' end-seq (Figure 1G).

For the analysis of *ssrA* distribution across polysomes (Supplementary Figure S5B–D), Northern blots were performed for three biological wildtype replicates (representative wildtype replicate shown in Supplementary Figure S5B) and one replicate for *Δpnp* (Supplementary Figure S5C). To quantify the increased signal in polysomal fractions of *Δpnp* that is already visible by eye, we calculated the background-subtracted cumulative signal per equally sized band and adjusted for the amount (μg) of loaded RNA per lane. Individual fractions between gradients were often slightly differently aligned with respect to the migration of individual ribosomal subunits, monosomes and polysomes. Hence, we did not compare individual fractions, but grouped fractions as ribosome-free (subunits & lightest fraction(s); fraction 1, 2 and 0–2, respectively in WT and *Δpnp*) and ribosome-associated fractions (fractions >2). We summed the background- and loading-adjusted band intensities for both groups and show the relative proportions in Supplementary Figure S5D.

RNA-seq and Rend-seq

Oligos used for library preparations are listed in Supplementary Table S2. RNA-seq for expression quantification in Supplementary Figure S5G–I was executed as described in (25). Rend-seq libraries were prepared as described in (28) from 20 μg of RNA prior rRNA removal. RNA fragmentation was done for 25 s. Final libraries were analyzed for size, quality and concentration using qPCR and the Fragment analyzer at the Biomicrocenter at MIT and sequenced on Illumina HiSeq 2000 or NextSeq500 machines. RNA-

seq and Rend-seq data processing and mapping was done as described in (25,28).

3' end-sequencing library preparation

For 3' end sequencing, the Rend-seq protocol was modified in the following ways: 3' end ligation with Linker-I was done with 1.8 μg of non-fragmented total or nascent RNA, followed by RNA precipitation using isopropanol. This ligation targeted hydroxylated RNA 3' ends that are typical products of transcription and canonical RNA decay (38). Phosphorylated RNA 3' ends that may arise as a result of non-enzymatic fragmentation or RNase A contamination were not ligated using this approach (38,39). RNA was fragmented for 2 min following ligation, which was followed again by RNA precipitation using isopropanol. 32–63 nt fragmented, 3' end-ligated RNA was size selected from a 15% TBE–urea polyacrylamide gel and precipitated for reverse transcription, circularization and final PCR amplification with 6–8 cycles as described in (28).

Our approach differed from Term-seq (40) in the following ways: RNA ligation was done with truncated T4 RNA ligase II, K227Q instead of T4 RNA ligase 1 and different adaptor sequences. For cleanup between reaction steps, RNA and cDNA was precipitated or gel extracted and precipitated, whereas sample cleanup in Term-seq was done with SPRI beads. Primers and enzymes for reverse transcription and consequently buffers and reaction conditions were also different. To add PCR handles on both sides of the cDNA fragment, our protocol included an intramolecular cDNA circularization step and Term-seq relied on an intermolecular ligation of a separate DNA adaptor. By design, our samples were sequenced from the 5' end of the insert into the 3' end adaptor and Term-seq sequences from the 3' end.

3' end-seq processing and mapping

In analogy to Rend-seq, 3' end linker sequences were trimmed. To deal with non-template addition during reverse transcription, all reads were further trimmed with the fastx_trimmer of the fastxtoolkit/0.0.13 by 1 nt at their 5' end and only reads of 15 nt length or longer were kept for mapping with bowtie/1.2 using the options -v 0 -k 1 (41). For further data analysis, mapped data were transformed into bam-format and bedgraph-format using samtools/1.5 and bedtools/2.25.0 (29,42). Coverage data included only 3' terminal single nucleotide counts per read.

To identify 3' adenylated transcript ends, the unmapped reads were trimmed using cutadapt/1.16 (43). First, 3' terminal CCAs were removed with the following settings: -a "CCA\$" –minimum-length 15 -O 3 –trimmed-only. This step is necessary to avoid confusing CCA-tails, e.g. at tRNA 3' ends, with 3' end adenylation that arise through different mechanisms *in vivo* (44). Second, A-tails of any length were trimmed (-a "A{38}" –minimum-length 15 -O 1 –trimmed-only). These trimmed reads were mapped and further converted with the same settings and code as untrimmed reads above. For 3' end analysis of total RNA, mapped trimmed (i.e. A-tailed) and untrimmed 3' end coverage profiles were summed.

Reference RNA mapping

Unmapped reads after mapping 3' end-sequencing to the *E. coli* genome (described above) were mapped with the same settings and code as above, but to the five reference RNA template sequences (Supplementary Table S2). In most cases, T7 RNA polymerase adds a few non-templated nucleotides at the 3' end of the template (45). To include these 3' ends for full-length RNA analysis, unmapped reads after mapping to the reference RNA template were iteratively trimmed by 1 nt, filtered for length to only include reads of 15 nt length or longer (fastx_trimmer -Q 33 -t 1 -m 15) and remapped to the template for five times. After trimming 5 nt, only background mapping was observed and thus trimming was stopped. Trimmed mapped reads and reads that mapped to the reference RNA template prior trimming were merged and converted to bedgraph files using sam- and bedtools.

Expression, Φ_{end} and internal density calculation, and rRNA ratio calculation

Expression analysis. To quantify mRNA levels (rpkm, reads per kilobase million) per coding sequence (CDS) from Rend-seq and RNA-seq data, coverage at each base in the 3'-mapped channel (3' terminal positions of reads) was divided by the total number of mapped reads in million. This yielded reads per million (rpm). Next, we calculated the total coverage per CDS excluding 50 nt at both ends of the CDS to avoid conflating effects by end-enrichment in Rend-seq and divided by the length of the CDS in kb minus 0.1 kb, yielding rpkm. Depending on the depth of the RNA-seq dataset, we either required a minimum of 30 or 100 reads for further analysis.

Φ_{end} *and internal density calculation.* The fraction of reads that map to mature 3' ends (Φ_{end}) was calculated as the ratio of 3'-end-seq reads mapping to the mature 3' end of simple TUs (± 2 nt based on the expected possible peak width (28)) and all 3' end reads mapping to that TU (Supplementary Table S7). A minimum of 20 3' end reads per TU was required (internal and mature 3' end combined). Φ_{end} scales from 0 to 1. For a Φ_{end} of 1 all transcripts mapping to this TU have a mature 3' end. A Φ_{end} of 0 indicates that all transcripts have 3' ends mapping internal to the TU with no full-length RNA. Hence, we used this measure as intuitive way to estimate the transcriptome heterogeneity among the pool of full-length and partial transcripts. Φ_{end} is an upper bound for the fraction of full-length RNA at a given TU. As Φ_{end} depends on the precise definition of TUs, it cannot be immediately transferred to more complex TUs.

To consider other TUs as well, we compared the internal 3' end densities relative to the nearby mature 3' end signal. To do so, we calculated the internal read density (average internal 3' end count per nucleotide) in 200-nt windows 10-nt upstream of mature 3' ends that are either part of simple or complex TUs (as defined above) and divided it by the read counts at the associated mature 3' end. We required at least 20 3' end counts in the 200-nt window and at the mature 3' end combined. We $\log_{10}$-transformed the relative internal densities for visualization as cumulative distribution (Supplementary Figure 1C). Internal densities between

simple TUs and more complex TUs were largely similar (*P*-value 0.16 between simple TUs and TUs with one mature 3' end, Kolmogorov–Smirnov test, multiple testing adjusted), allowing us to conclude that the high degree of transcriptome heterogeneity is universal across TUs. For mature 3' ends that are part of TUs with multiple 5' and 3' ends we observed a slightly higher relative internal density than for simple TUs (*P*-value 0.008 between simple TUs and TUs with multiple mature 3' ends, Kolmogorov–Smirnov test, multiple testing adjusted), which indicates even higher transcriptome heterogeneity in addition to the full-length isoform diversity than in simple TUs. This is feasible considering the possibility of stable RNA decay intermediates, transcriptional pauses, rho termination, transcript leaders among others.

rRNA ratio calculation. The three very differently sized ribosomal RNAs (rRNAs) 5S (0.12kb), 16S (1.5kb) and 23S (2.9kb) are expressed at a molar stoichiometry of 1 and very stable with half-lives beyond the doubling time of *E. coli* (46,47). Hence, we reasoned that these can be used as internal size standards and their mature 3' end signal should reflect the 1:1 ratio in samples that were not depleted of rRNAs prior library preparation. That was the case for our RNAsnap and hot phenol extracted RNA samples (Figure 1B). In addition to this analysis, Φ_{end} of rRNAs were also >0.8 even for the 23S rRNA (Figure 3A).

Peak calling

In 5' end and 3' end sequencing datasets, 3' end or 5' end coverage peaks were called based on *Z*-score calculation across a window of 100 nt. We used peak calling in the analyses to identify RNA polymerase pause sites (Figure 2), putative RNase E sites (Figure 3) and mature 3' ends across all TUs (Figure 3F). Specifically, we extracted the read coverage in a region ±50 nt of any position covered by a read in the dataset and normalized by the total amount of reads in this window. We required a minimum of 15 reads in each window for *Z*-score calculation. To calculate mean and standard deviation of the background, the center position was excluded. The *Z*-score reflects the normalized coverage at the position of interest minus the background mean divided by the standard deviation. *Z*-scores >7 were considered peaks. These could be within or outside of CDSs.

To call putative RNase E sites from remapped 5' end sequencing data from (48), we compared data in the *rne-3071* ts strain (temperature-sensitive mutant of RNase E) and the wildtype strain. 5' end peaks should be absent or substantially reduced in the *rne-3071* ts strain compared to wildtype. Hence, we required that *Z*-scores in the wildtype strain were >7 and at least four times lower in the 5' end-seq data in the *rne-3071* ts strain. Lastly, we focused our analyses on sites within protein-coding transcription units. The putative RNase-dependent sites are listed in Supplementary Table S10.

As described above, these sites were derived computationally from a published study that grew the wildtype and *rne-3071* ts strain at 44°C for 45 min before harvesting at an unspecified cell density. Thus, RNA expression differences between the wildtype and *rne-3071* ts strain might be

present that originated from other sources than the mutation, limiting the calling of true RNase E sites. Hence, we refer to differentially detected peaks in our analysis as 'putative' RNase E sites, which indeed recapitulated the expected sequence preference of RNase E sites (Figure 3B) (49,50). Lastly, RNase E cleavage sites might differ at different growth temperatures, e.g., 37°C used in this study and 44°C for the comparison of the wildtype and *rne-3071* ts strain. Nevertheless, we detect a specific enrichment of 3' ends at putative RNase E sites (Figure 3B, Supplementary Figure 3B–D), suggesting that many of these sites were also RNase E cleavage sites in our experimental conditions.

Quantification of nascent RNA fraction in total RNA

For the nascent RNA 3' end data, 21355 positions had *Z*-scores above the cutoff of 7 and sufficient read coverage that were 2.1% of all positions passing the read cutoff of 15 in the window around the position of interest. We overlayed peaks with the CDS annotation using bedtools intersect and classified them according to their relative position as in- or outside of CDSs (71% inside CDSs). To visualize the nucleotide distribution around peaks internal to CDSs, the underlying sequence 10 nt upstream and 5 nt downstream around peak positions was extracted from the genome fasta sequence and visualized as sequence logo with Weblogo 3 (51).

To refine the classification of RNA polymerase elemental pause positions for the quantification of nascent RNAs in total RNA, we calculated a pause motif score across the 16 nt of interest and excluded the 25% lowest scoring sites from any further analysis. To this end, we calculated the position weight matrix from all sites in CDSs that have a peak with a *Z*-score >7. From that, we generated a log-odds matrix by comparing to the nucleotide frequencies within protein-coding genes. For each individual 16 nt peak sequence, the pause motif score was calculated by summing up log-odds matrix entries at the respective sequence position (column) and base identity (row). Peaks with pause scores <2 (25% lowest scoring peaks) did not show the sequence preference as given in Figure 2C when visualized as sequence logo and were not used for the quantification of nascent RNAs in total RNA that is described below.

The peak distribution and sequence comparison between total and nascent RNA samples suggested little nascent RNA in the pool of total RNA (Figure 2B, C). To estimate the nascent RNA fraction in total RNA, we reasoned that the amount of total RNA 3' ends at positions of RNA polymerase elemental pause sites provides an estimate of the contribution of nascent RNAs to the total transcriptome. However, not all nascent RNA 3' ends fall into these pause sites. Hence, we calculated the fraction of nascent RNA signal in pause sites f_{pause} compared to the whole transcription unit (total TU counts); the same was done for total RNA (without the adjustment for biochemical enrichment, see below). The ratio of the two fractions (total/nascent) was considered to reflect the "Nascent RNA fraction in total RNA", which is an upper bound for nascent RNA in the pool of total RNAs, because 3' ends at pause positions could also come from decaying transcripts that happen to be caught at that position.

Some 3' end signal across the whole transcription unit originated from total RNA due to the incomplete biochemical enrichment of nascent RNAs. To account for this, we multiplied the total TU counts with 0.91, which is the median of the internal end fraction for simple TUs with 30 or more 3' end reads in the nascent RNA sample. It indicated that at least 9% of 3' ends came from total RNA. The distribution of the "Nascent RNA fraction in total RNA" is given in Figure 2D and Supplementary Table S9 for simple TUs and all TUs (Supplementary Table S6, mRNA-encoding regions). The medians of these distributions were 0.14 and 0.16, respectively.

Analysis around putative RNase E sites, 3' end peaks and across CDSs

Sequence logo of putative RNase E sites. We aimed to visualize the nucleotide distribution around putative RNase E sites (Figure 3B). To this end, we extracted the underlying sequence 5 nt upstream and 3 nt downstream of peak positions from the genome fasta sequence using bedtools getfasta and visualized them as sequence logo with Weblogo 3 (29,51).

Normalized mean distributions around putative RNase E sites and nascent and total RNA 3' end peaks. For alignment of 3' and 5' end coverage relative to putative RNase E sites and RNA polymerase pauses, we focused on sites within protein-coding genes including adjacent UTRs (see definition of mRNA-encoding genomic regions above). We defined symmetrical 60 nt windows around these sites and extracted coverage information for all positions. For further analysis, only sites were considered that met our read cutoff criteria (0.5 reads per position in wildtype & *Δpnp* for putative RNase E sites in Figure 3 and Supplementary Figure S3, 0.15 reads per position in polysome gradient samples for RNA polymerase pause sites and RNase E sites in Figure 4B). To be able to consider many different sites despite differences in (local) expression for analysis, coverage around each site was normalized by the mean coverage in the 60 nt window. We refer to this as relative coverage. A value of 1 resembled the region mean, everything >1 was increased relative to the mean and everything between 0 and 1 was decreased. For visualization, the composite mean of all relative coverages was plotted for each position centered around the sites of interest. In Figure 3B and Supplementary Figure S3, the full 60 nt window is shown and in Figure 4B we show the mean signal at the respective site with error bars as standard error of the mean. For visualization as a heatmap, the relative coverage for each site was plotted ranked by the RNase E site *Z*-score (see peak calling for the data in (48)). In Supplementary Figure S3D we plotted the mean of the 3' end adenylation percentage per position for the same sites as in Figure 3B and Supplementary Figure S3B, C.

Normalized mean distributions across CDSs relative to the start codon. For alignment of 3' end coverage relative to the start of CDSs, we focused on the 816 first CDSs in our mRNA-encoding genomic regions. We added 100 nt at the start of each CDS to include 5' UTRs in the coverage analysis. We filtered CDSs based on length and read coverage

(CDS > 150 nt, mean read coverage per 100 nt 5'UTR + CDS at least 0.01 in the lowest coverage dataset (fraction 1), 538 genes). As described above for sites in a 60 nt window, we normalized coverage in each CDS by the mean coverage in the whole region. Next, we applied a running average smoothing per region within a 100 nt window. The smoothed relative coverage data were winsorized to the 95% quantile for each region. Finally, we computed the mean of the smoothed coverage at every position for –100 nt to 1900 nt across all CDSs that were at least as long as the respective position. Thus, more CDSs were included in the smaller distances to the start codon than towards later positions and the profile became noisier towards later positions (Figure 4C).

3' end adenylation of mature 3' ends and RNase E sites

To investigate the global prevalence of 3' end adenylation at mature 3' ends and putative RNase E sites internal to protein-coding genes in wildtype, *Δpnp* and *ΔpcnB*, we included all protein-coding genes in the analysis. Nucleotide positions that were within the mRNA-encoding genomic regions, but downstream of CDSs, and showed 3' end peaks with *Z*-scores >7 were considered mature 3' ends (1341 (wildtype), 2003 (*Δpnp*), 905 (*ΔpcnB*) positions). The percentage of 3' end adenylation for these sites was plotted as cumulative distribution for all three data sets in Figure 3F with a minimum of 10 reads per mature 3' end. Similar distributions were obtained for simple TUs only. We considered putative RNase E sites within mRNA-encoding genomic regions and with 10 or more 3' end reads (135 sites) for the quantification of 3' end adenylation in Figure 3F. To visualize the distribution of adenylation at and around putative RNase E sites (Supplementary Figure S3D), we included all putative RNase E sites that have at least 30 adenylated and non-adenylated reads in a 60 nt window (1423 sites). Hence, the average adenylation was given for the same sites as in Figure 3B and Supplementary Figure S3B, C.

Published data integration

Data from the following studies were remapped and integrated into the analysis described above:

- Term-seq from *E. coli* in LB at OD 0.5 (ERR2433552, ERR2433553, ERR2433554) and *B. subtilis* in LB OD 0.1-0.2 (ERR1232477, ERR1248362, ERR1248363) (22,40).
- 5' end-seq data *rne-3071* ts (GSM1405878), *rne* wild-type (GSM1405877) (48).
- Rend-seq data from *E. coli* MG1655 (GSM2500131), E. coli MG1655 *Δpnp* (GSM2971253) and *B. subtilis* 168 (GSM2500127) (28).
- Scripts used for read processing and mapping can be found on https://github.com/gwlilabmit/Herzel2022_RNAdemographics.

We used the reported mRNA half-lives obtained by τ-seq in (52) to compare to our Φ_{end} distribution (Supplementary Figure S3A). Only CDSs were considered with half-lives differing less than two-fold between τ-seq replicates.

Data analysis presented in this manuscript was done using bash scripts and software as cited above, python 2.7, MATLAB R2021b, Rstudio (Version 1.2.5042) and on the high-performance cluster LURIA maintained by the Biomicrocenter at MIT.

RESULTS

A molecular census of *in vivo* mRNA states by 3' end sequencing

To obtain a comprehensive molecular census of mRNAs *in vivo*, the methods of RNA extraction and detection must be able to capture every type of molecule with a similar efficiency. They must also minimize additional RNA fragmentation during and after cell lysis. Here, we describe our approach to meet these criteria.

To detect every type of RNA molecule throughout its life cycle, we take advantage of the ability of RNA 3' end ligation to quantitatively mark each molecule containing a 3' hydroxyl group with a preadenylated DNA adapter (Figure 1A) (53,54). Deep sequencing of ligation junctions can be carried out to determine the 3' end position, henceforth referred to as 3' end sequencing (conceptually similar to previously described approaches (21,40)). Compared to 5' end positions, sequences at RNA 3' ends allow us to distinguish transcripts with mature TU ends from those that are nascent or decay intermediates. Although the exact nature of RNA length remains undetermined, 3' end sequencing provide a partial molecular census without excluding certain size ranges, as is the case with existing full-length mapping strategies (18–20).

To minimize biases and fragmentation introduced during RNA extraction, we tested several extraction methods and found RNAsnap to have the least amount of impact (32). Several metrics were used to evaluate the quality of extracted RNAs. First, the three ribosomal RNAs (5S, 16S and 23S) spanning a size range of 120 nt to 2905 nt are expected to be in similar molar concentrations (46,47). Using 3' end sequencing, we found that both RNAsnap and hot phenol extraction reproduce the expected molar ratios among rRNAs (1.02 ± 0.22, mean $\pm$ standard deviation), whereas the commonly used TRIzol reagent strongly favors the shorter 5S rRNA as reported earlier (Figure 1B) (32). Second, we spiked-in *in vitro* synthesized (chemically synthesized or transcribed) reference RNAs during extraction to examine the degree of fragmentation after cell harvesting. As a measure for the integrity of each RNA species, we define Φ_{end} as the fraction of the corresponding 3' end sequencing reads that map to the full-length 3' end. We found that Φ_{end} for RNAsnap-extracted sample remains largely unchanged across different sizes of reference RNAs (Figure 1C). By contrast, other extraction methods yielded reduced Φ_{end} for longer reference RNAs, indicating damages or skewed size representation during extraction. Third, additional evidence based on Northern blotting and density fractionation, as presented below, confirms that the RNA molecules observed using 3' end sequencing of RNAsnap-extracted samples are faithful representations of the heterogeneity *in vivo*.

Full-length transcripts constitute a small molar fraction of total mRNA

The robust extraction and mapping methods enable us to examine the diverse RNA forms present in the *E. coli* transcriptome during exponential growth. As expected, the positions that correspond to mature 3' ends of a TU often have the most prominent signals (Figure 1D). However, a substantial amount of RNA 3' ends are also observed at many positions internal to the TU, suggesting a large transcriptome heterogeneity. To quantify this heterogeneity, we again use Φ_{end} to denote the fraction of reads that map to a transcript's mature 3' end for each transcription unit (TU). Because the RNAs that end at mature 3' ends include both full-length and some decay intermediates (Figure 1A), Φ_{end} represents an upper bound for the fraction of RNA molecules that are full-length. For a representative transcription unit encoding *ycaO*, the molar fraction of full-length RNAs is <57% ($\Phi_{end} = 0.57$, Figure 1D). Overall, the median Φ_{end} across all simple TUs (defined as those with a single promoter and a single terminator, $n = 142$, Materials and Methods) is 0.4 and highly reproducible (Figure 1G, H), indicating that full-length transcripts are a minority among total mRNA molecules and that partial transcripts, which we define as RNAs lacking at least one mature end, make up a substantial proportion of the transcriptome.

Confirming this result, Northern blot analysis against the 5' portion of simple TUs detected both full-length transcripts and partial ones of diverse lengths (Figure 1E and Supplementary Figure S1A). Similar to 3' end sequencing results, Northern blot signals from partial transcripts are dimly scattered across a wide size range, but nevertheless add up to a substantial fraction of total signals. Northern blots provide an estimate for the fraction of probe-binding RNAs that are full-length (F_{FL}). Although neither Φ_{end} or F_{FL} represents the exact fraction of full-length RNAs, they showed good correspondence among the 8 simple TUs that we tested (Figure 1F). These results indicate that the low Φ_{end} is not an artifact of the sequencing method. Additional confirmatory evidence comes from previously published transcriptomic data in *E. coli*. Term-seq is a similar method to 3' end sequencing but uses a different RNA extraction kit and library design (22). Despite these differences, it showed a similar distribution of Φ_{end} as our data (Figure 1H and Supplementary Figure S1B). Lastly, complex TUs (TUs with multiple mature 5' and/or 3' ends) showed similar ratios of internal 3' end density to mature 3' end reads as simple TUs, indicating that the high degree of transcriptome heterogeneity is general across TUs in *E. coli* (Supplementary Figure S1C). Overall, we conclude that the molar fraction of full-length RNA is small.

Majority of partial transcripts are not nascent

Next, we quantified the contribution of nascent transcripts to the surprisingly high fraction of internal 3' ends. If most partial transcripts were nascent, we would expect internal 3' ends from total RNA and nascent RNA to follow similar patterns of RNA polymerase (RNAP) pausing across the body of TUs. Nascent transcripts are associated with RNAP as a part of the biochemically stable transcription
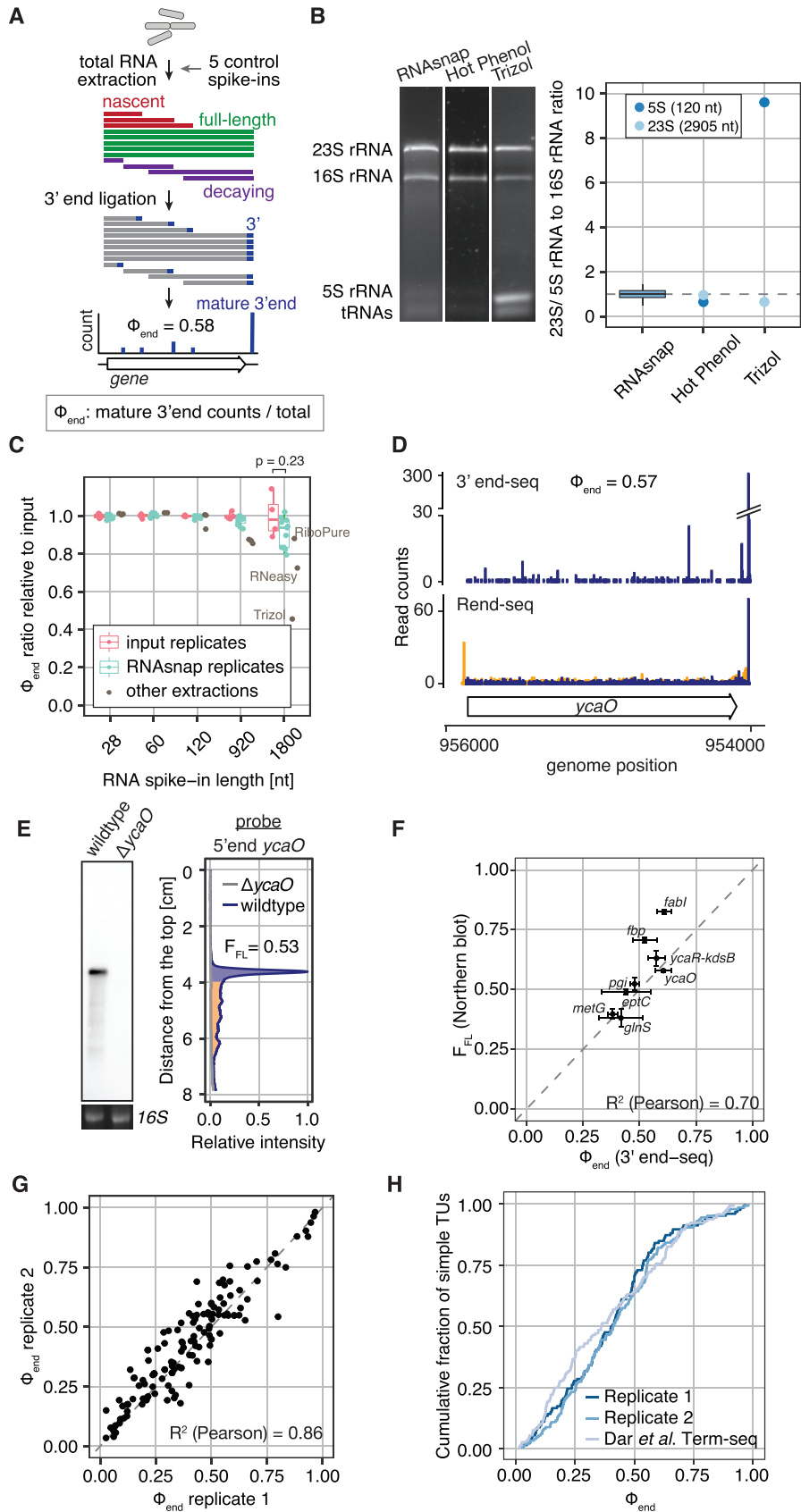
Figure 1. *In vivo* RNA 3' end profiles reveal that the majority of transcripts are not full-length in *E. coli*. (**A**) Schematic of 3' end sequencing and hypothetical data derived from the suite of nascent, decaying and full-length transcripts. The measure Φ_{end} is calculated as the fraction of 3' end sequencing reads that

elongation complex and can be purified by immunoprecipitation ([33,53]) (Figure [2]A and Supplementary Figure S2, Methods). Sequencing data for 3' ends of RNAP-associated transcripts showed strikingly different profiles from the 3' ends of total RNAs (Figure [2]B). Furthermore, the elemental pause sequence for RNAP was recapitulated in nascent RNA but not in total RNA ([33,55,56]) (Figure [2]C). These two observations indicate that most partial transcripts were not nascent.

To estimate the nascent RNA fraction for each TU in total RNA, we compared the fraction of 3' ends that map to elemental pause sites, denoted as f_{pause}, in both the total and nascent RNA samples. If pause site signals in total RNAs are entirely contributed by nascent RNAs, the nascent RNA fraction in total RNAs can be calculated as the ratio of f_{pause} between the total and nascent RNA samples. In reality, pause site signals in total RNAs may be also derived from decaying RNAs. Therefore, the ratio of f_{pause}s provides an upper bound for nascent transcripts in the total RNA pool. The distribution median was 0.15, indicating that nascent RNAs contribute at most 15% of all RNA molecules for a typical transcription unit (Figure [2]D).

Majority of partial transcripts are decay intermediates

If a small fraction of partial transcripts is nascent, the remainder might be partially degraded. In this model, RNAs with longer half-lives should correspond to a low prevalence of RNA decay intermediates and thus a high Φ_{end}. Indeed, stable noncoding RNAs showed Φ_{end} close to 1, whereas messenger RNAs of similar lengths have much lower Φ_{end} (Figure [3]A). Among mRNAs, Φ_{end} also correlates positively with mRNA half-life (Supplementary Figure S3A, R (Pearson) = 0.46). Because genome-wide measurements of RNA half-lives do not report full-length RNA stability (see below), we do not expect a perfect correlation between half-lives and Φ_{end}. These results suggest that most RNA molecules bearing internal 3' ends could be generated during the degradation process.

We next examined the location of internal 3' ends and its relationship to RNA-decay pathways. In *E. coli,* mRNA degradation primarily occurs through an initial endonuclease cleavage, followed by 3'–5' exonucleolytic activities and further rounds of endonuclease cleavage. RNase E is the main endonuclease and is essential for viability. PNPase is the main RNA exonuclease whose function can be partly compensated by other 3'–5' exonucleases, such as RNase II ([7,8]). To test if RNase E cleavages could explain the internal 3' end profiles, we compared our wildtype 3' end-seq data to previously published 5' end sequencing data for an RNase E deficient *E. coli* strain ([48]). We found that 3' end signal is enriched among 1423 putative RNase E cleavage sites that have sufficient read coverage in our data, and the enrichment is even more pronounced in the strain lacking PNPase (Figure [3]B and Supplementary Figure S3B, C). Consistently, partial transcripts accumulate upon PNPase deletion, as indicated by both 3' end sequencing and Northern blots (Figure [3]C, D). These results suggest that endonucleolytic cleavage followed by exonucleolytic digestion contributes to the ubiquitous internal 3' ends that we observed.

Further supporting the prevalence of decay intermediates *in vivo*, we found signatures of adenylated 3' ends (Figure [3]E). RNA 3' end adenylation has been identified as a decay enhancing posttranscriptional RNA modification in *E. coli* ([57]). Adenylation is mediated by the Poly(A) polymerase PcnB and facilitates 3'-5' exonucleolytic decay, especially for structured RNA ends. Indeed, our 3' end sequencing data show a substantial level of adenylation at mature TU ends (median = 11.5% in wildtype, 38.5% in Δpnp, 0.0% in $\Delta pcnB$, Figure [3]F). Internal 3' ends that map to putative RNase E cleavage sites also show adenylation in a manner that depends on PNPase and PcnB (median = 10.0% in wildtype, 27.8% in Δpnp, 0.0% in $\Delta pcnB$, Figure [3]F and Supplementary Figure S3D). These adenylated internal 3' ends likely represent transient intermediates between actions of the Poly(A) polymerase and 3'-5' exonucleases. Overall, the length of A-tails responds similarly to deletions of PNPase and PcnB: The median A-tail length is short with 2 nt in wildtype cells, consistent with what was described previously ([58]). It increases to 3 nt in cells lacking PNPase and vanishes in cells lacking PcnB (Figure [3]G and Supplementary Figure S3). Together, these lines of evidence support the model that many partial transcripts are RNA decay intermediates.

Interestingly, we found a higher overall Φ_{end} in a different bacterial species, *B. subtilis*, that possesses an additional degradation pathway via the 5'–3' exonuclease RNase J ([59]). The median Φ_{end} is 0.6 for *B. subtilis*, compared to

---

map to the 3' end of the mature transcript. Φ_{end} represents an upper bound for the fraction of full-length mRNA. (**B**) Minimization of length bias in RNA extraction. Left: Representative total RNA profiles by agarose gel electrophoresis from 3 different RNA extraction methods. Right: Molar ratio of rRNAs measured by 3' end-seq. The mean 5S or 23S to 16S rRNA 3' end ratio ± standard deviation for RNAsnap samples is 1.02 ± 0.22. The dashed line indicates the expected ratio of 1. Boxplot whiskers indicate 1.5 times the interquartile range. (**C**) Integrity of spiked-in reference RNAs. Five reference RNAs are synthesized *in vitro* ('input') and spiked-in during RNA extraction. Boxplot shows Φ_{end} of the reference RNAs (spiked-in or input) normalized by the mean Φ_{end} of the input. Several technical replicates are shown for RNAsnap and input. The p-value was calculated between "input" and "RNAsnap" replicates with the Wilcoxon rank-sum test. No statistical test was applied to compare "other extractions" to "input" as this is a heterogenous group of samples from multiple RNA extractions with different characteristics. (**D**) 3' end- and Rend-seq data for the simple transcription unit (TU) encoding *ycaO*. For 3' end-seq, 3'-mapped (blue) read counts are plotted. Each read corresponds to one RNA 3' end. For Rend-seq, 5'-mapped (orange) and 3'-mapped (blue) read counts are plotted separately to illustrate transcript boundaries. Here, a 5' or 3'-mapped read corresponds to one RNA fragment generated during library preparation. (**E**) Northern blot (left) and intensity profile (right) for the same TU as in D. F_{FL} is calculated as the ratio of full-length signal intensity (blue) over the total intensity (orange + blue). The probe hybridizes to the 5' region of the TU. 16S rRNA detected with SybrSafe staining after agarose gel electrophoresis prior to transfer is shown as loading control. (**F**) Correlation between 3' end sequencing and Northern blot analysis. Φ_{end} and F_{FL} for 8 simple TUs, ranging from 0.9-2.1kb, are plotted. Error bars reflect standard deviation of 3 and 2 replicates from Northern blotting and 3' end-seq, respectively. Full Northern blots are given in Supplementary Figure S1A. (**G**) Reproducibility of Φ_{end} by 3' end sequencing of RNAsnap-extracted RNA. Scatter plot shows Φ_{end} from simple TUs between two biological replicates. (**H**) Cumulative distributions of Φ_{end} for two biological replicates and published data from ([22]), with medians around 0.4.
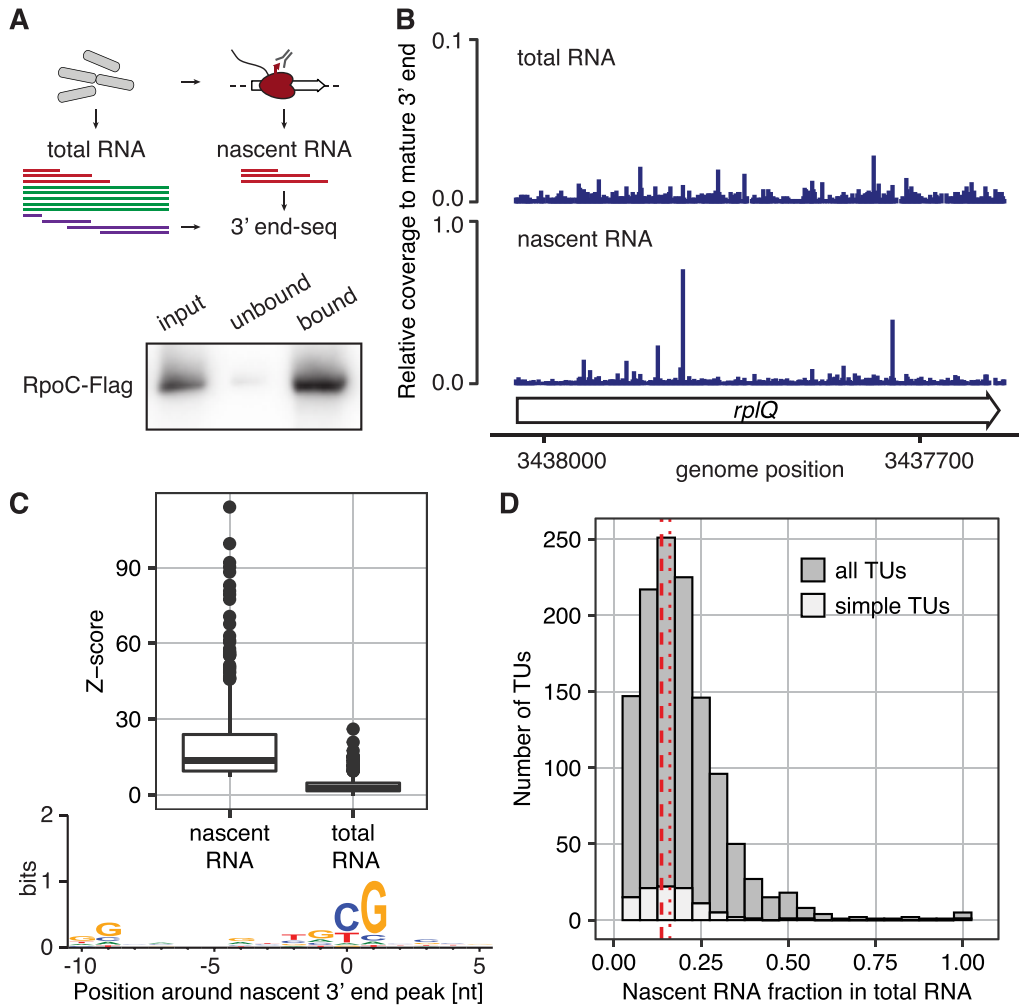
Figure 2. Nascent RNAs contribute to a small fraction of the transcriptome. (**A**) Schematic of 3' end sequencing of nascent and total RNA. Nascent RNA is enriched by immunoprecipitation of Flag-tagged RNA polymerase (RpoC-Flag) and subject to 3' end sequencing. Lower panel shows efficient isolation of RNA polymerases by western blot analysis with the Flag-antibody detecting RpoC-FLAG prominently in the total lysate (input) and eluate (bound), but less in the unbound fraction after immunoprecipitation. Full western in Supplementary Figure S2. (**B**) 3' end sequencing results for nascent and total RNA. Internal 3' end read counts are normalized to the read count at the mature 3' end of *rplQ*. (**C**) Signals at transcriptional pause sites. Transcriptional pause sites are identified as peaks in 3' end signals for nascent RNAs. Boxplot compares Z-score distributions for these sites in nascent RNA and total RNA. Sequence logo for regions around transcriptional pause sites resembles the elemental pause sequence. (**D**) Estimates of nascent RNA fraction in total RNA. Histogram shows the distribution of estimated fraction of total RNA that is nascent across different TUs. Simple TUs are a subset of all TUs and overlayed as white bars. Red lines mark distributions medians (dashed – simple TUs, dotted – all TUs).

0.4 for *E. coli* under a similar growth condition (Supplementary Figure S3I, J). This difference is consistent with the facts that mRNA decay in *B. subtilis* may not require sequential rounds of endonuclease cleavage, and that transcription elongation is much faster in *B. subtilis* (7,60).

Translation primarily occurs on a small fraction of transcripts

Among the majority of mRNA molecules that are decay intermediates, many may lack ribosome binding sites and hence are incompetent for translation. Indeed, density fractionation by sucrose gradients shows that partial transcripts are depleted in ribosome-rich fractions, as indicated by increased Φ_{end} in heavier fractions (Figure 4A). The increased Φ_{end} beyond the input levels further confirms that the limited Φ_{end} observed in total RNA is not a result of post-lysis fragmentation and that partial transcripts contribute

to a smaller fraction of translated RNAs. Consistent with the expectation that decay intermediates are incompetent for translation, we found that RNAs that end at putative RNase E cleavage sites are much more prevalent in the ribosome-free fraction than the ribosome-rich fractions (Figure 4B).

We expect that translated partial mRNAs primarily consist of nascent mRNAs. Indeed, unlike total RNAs, RNAP pause site signals are prevalent in ribosome-rich fractions (Figure 4B). Further, the 3' ends of the monosome fraction are enriched immediately downstream of the first start codons in operons, whereas 3' ends of higher polysome fractions are enriched at increasing distances from start codons (Figure 4C). Northern blot analysis confirms that partial transcripts increase in size with increasing sucrose density, giving way to full-length transcripts at heaviest density (Figure 4D and Supplementary Figure S4). Together, these re-
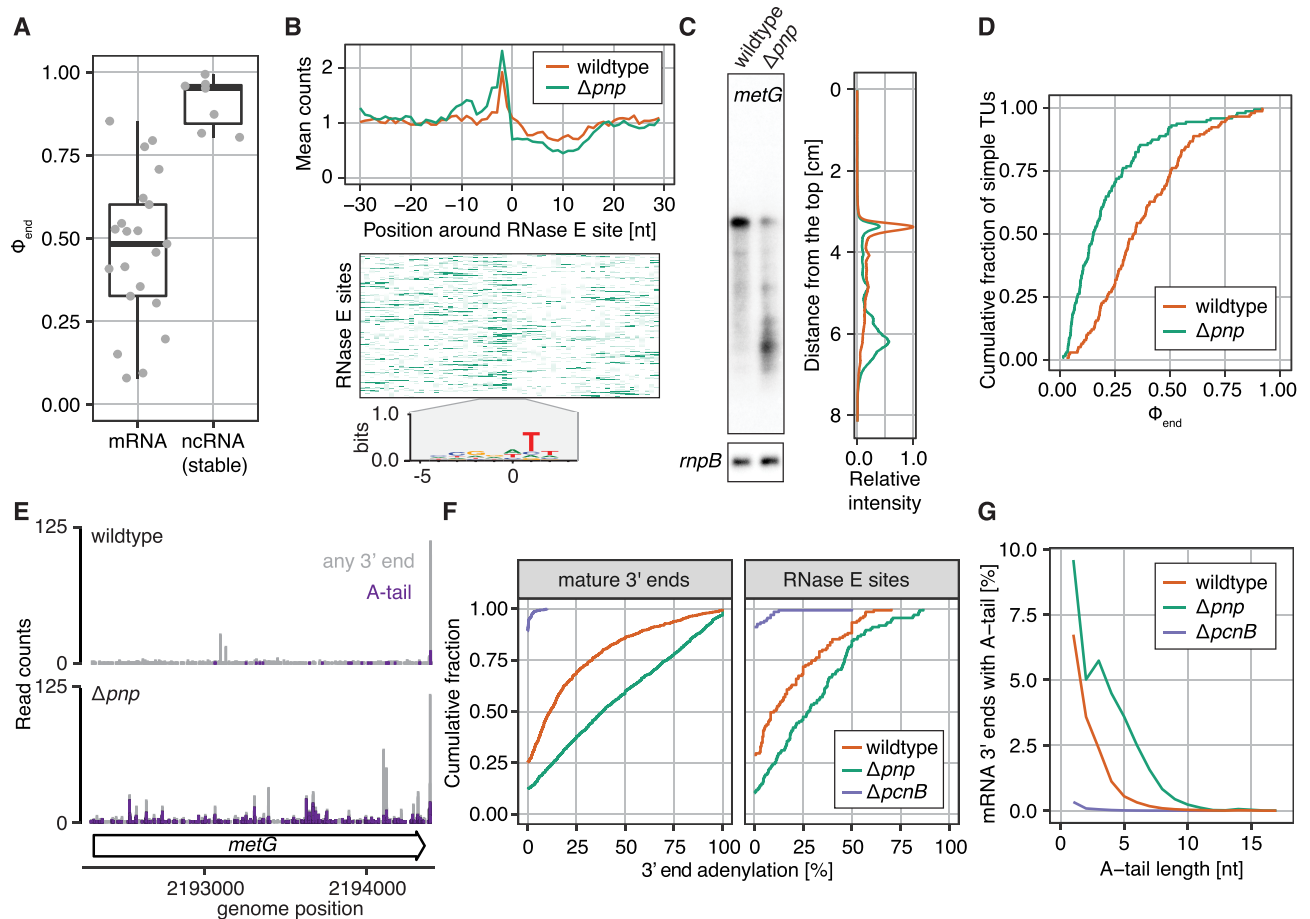
Figure 3. *In vivo* RNA decay signatures corroborate high amounts of decaying transcripts. (**A**) Distribution of Φ_{end} for size-matched mRNAs and stable noncoding RNAs (ncRNA). Boxplot shows Φ_{end} for stable ncRNAs (*ffs*, *gcvB*, *rnpB*, 5S, 16S, 23S rRNAs, *ssrA*) and mRNAs whose full-length products are in the same size range as the former. Whiskers correspond to 1.5 times the interquartile range. (**B**) Enrichment of 3' end signal at putative RNase E cleavage sites. Top: Average 3' end signals surrounding 1423 putative RNase E cleavage sites that has at least 30 reads in a 60-nt window. For each site, the read count for each position is normalized to the average read count in a symmetric 60-nt window. Line plots show the mean normalized count at each position across all sites. Middle: Normalized 3' end read counts across individual RNase E sites in for Δpnp. Each row represents data from each putative RNase E site. Color intensity represents the normalized 3' end signal across the 60-nt window. Rows are sorted by descending *Z*-score at the putative cleavage site. The 200 sites with the highest number of reads in the 60-nt window are shown. (See corresponding wildtype heatmap in Supplementary Figure S3B and full heatmaps in Supplementary Figure S3C) Bottom: Sequence logo at these putative RNase E cleavage sites. (**C**) Accumulation of partial transcripts in Δpnp. Left: Northern blot probing the 5' region of *metG* in wildtype and Δpnp shows decreased RNA integrity in cells lacking PNPase. Probe against stable ncRNA *rnpB* as loading control. Right: Intensity profiles of the Northern blot. See analyses for more TUs and replicates in Supplementary Figure S1A. (**D**) Global decrease in full-length RNAs in the *pnp* deletion (Δpnp) compared to wildtype. Cumulative distributions of Φ_{end} for Δpnp and wildtype are plotted for simple TUs. (**E**–**G**) Prevalence of short Poly(A) polymerase (PcnB)-dependent poly(A) tails. (E) 3' end profiles across *metG* in wildtype and Δpnp. Coverage shown in purple corresponds to adenylated 3' ends, grey coverage corresponds to all 3' ends. (F) Cumulative distributions of the percent of adenylation for mature 3' ends and putative RNase E sites (positions with at least 10 reads). Adenylation is abolished upon deletion of the Poly(A) polymerase ($\Delta pcnB$) and increases upon deletion of the exonuclease PNPase (Δpnp). (G) Percent of mRNA 3' ends with A-tails of different lengths (includes mature and internal 3' ends). The average A-tail length is very short and increases upon deletion of the exonuclease PNPase (Δpnp).

sults confirm that, despite a high molar fraction of decaying transcripts in the transcriptome, most translation occurs on nascent and full-length transcripts.

Quantitation of the functional transcriptome requires discerning decay intermediates

The prevalence of translation-incompetent mRNAs in the transcriptome suggests that many RNA quantification methods may not accurately measure the levels of functional mRNAs. To demonstrate this potential problem, we compared *E. coli* RNA-seq data for wildtype cells and cells

lacking PNPase. Previous studies using DNA microarray or RNA-seq, which does not distinguish functional and partial RNAs, have noted limited changes in RNA abundance and half-lives in the absence of PNPase (36,61). This result was recapitulated in data from our own lab (Figure 5A)(28). However, 3' end sequencing showed that the full-length RNA fractions are dramatically reduced, as shown earlier in Figure 3D. For example, the gene *fbp* has similar RNA-seq levels (2078 vs. 2295 rpkm in WT vs. Δpnp) but very different Φ_{end} values (0.48 versus 0.10) (Figure 5A-B). Globally, the log$_2$-fold-change in RNA-seq signal is closely centered around 0, but Φ_{end} values decrease by 3-fold on av-
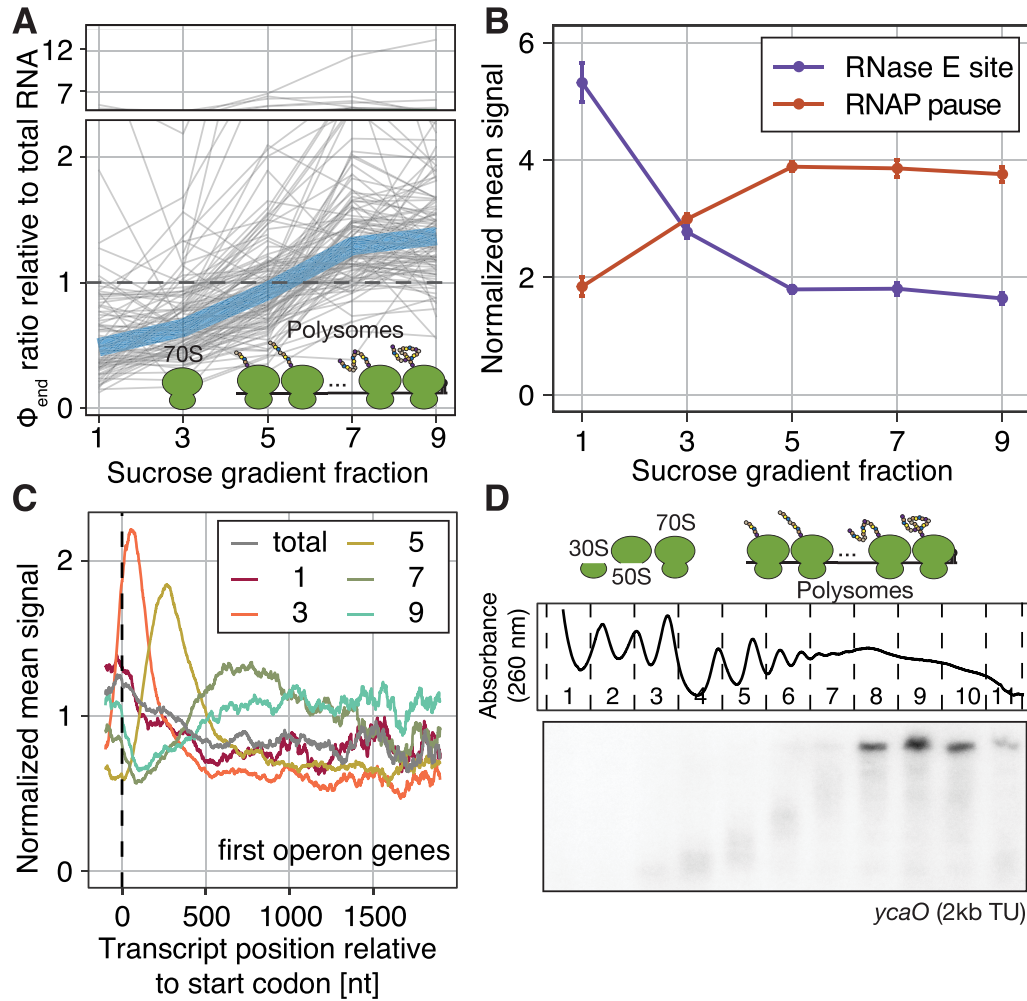
Figure 4. Enrichment of translation-competent RNAs in polysomes. (**A**) Global increase of Φ_{end} in heavy fractions of sucrose gradient. For each simple TU, the Φ_{end} in each fraction relative to the input Φ_{end} is plotted in grey. The median among all simple TUs is shown in blue, which starts below the input in lighter fractions and rises above the input in heavier fractions. The ribosome cartoon illustrates where monosomes (fraction 3) and polysomes (fractions >3) sediment across the gradient (compare absorbance profile in D). (**B**) Signatures of nascent RNA and decay intermediates across sucrose gradient. For each fraction, 3' end signal at each pause site (or putative RNase E cleavage site) is first normalized to the mean signal in a symmetrical 60 nt window, and then averaged across 396 RNA polymerase pause sites (or 460 putative RNase E sites). Error bars reflect standard error of the mean. (**C**) Enrichment of 3' end positions relative to start codons in different polysome fractions. For 538 first CDSs in operons, the average 3' end signal normalized to the CDS mean signal (Materials and Methods) is plotted with respect to distance to the start codon for individual fractions. (**D**) Size distribution for partial mRNAs in different polysome fractions. Top: absorbance profile across the sucrose gradient. Fraction numbers and the peaks that associate with 30S, 50S, 70S and polysomes are labeled. Bottom: Northern blot analysis probing the 5' region of *ycaO* across different fractions.

erage (Figure 5C). This result indicates that RNA-seq fails to report the substantially decreased levels of functional mRNA, which are accompanied with increased partial transcripts that are indistinguishable by RNA-seq.

Consistent with the global increase in partial, non-functional transcripts in the Δ*pnp* strain, we found that tmRNA, which rescues stalled ribosomes at the end of RNAs that lack stop codons (15,16), has increased association with ribosomes (Supplementary Figure S5B–E). Meanwhile, the cellular backup system mediated by the alternative ribosome rescue factor ArfA (62,63) is strongly upregulated (Supplementary Figure S5F), suggesting a pronounced need for ribosome rescue due to accumulation of decaying transcripts. These results, together with a much longer cell doubling time (Supplementary Figure S5A), cor-

roborate the substantially different functional transcriptome that was not captured by RNA-seq.

Similar to cells lacking PNPase, we found that other genetic or antibiotic perturbations also lead to changes in the functional transcriptome without co-occurring expression changes measured by RNA-seq (Figure 5D and Supplementary Figure S5G–I). In other words, RNA-seq largely underestimated the pronounced effect on RNA integrity that is visible in Φ_{end} changes. We also observed poor correlations between changes in RNA-seq quantitation and RNA integrity when cells are grown in different media or growth phase (Supplementary Figure S5J–N). In conclusion, our study shows that end-based mapping approaches are important to study RNA decay and that prevalent decay intermediates can interfere with gene expression profiling.
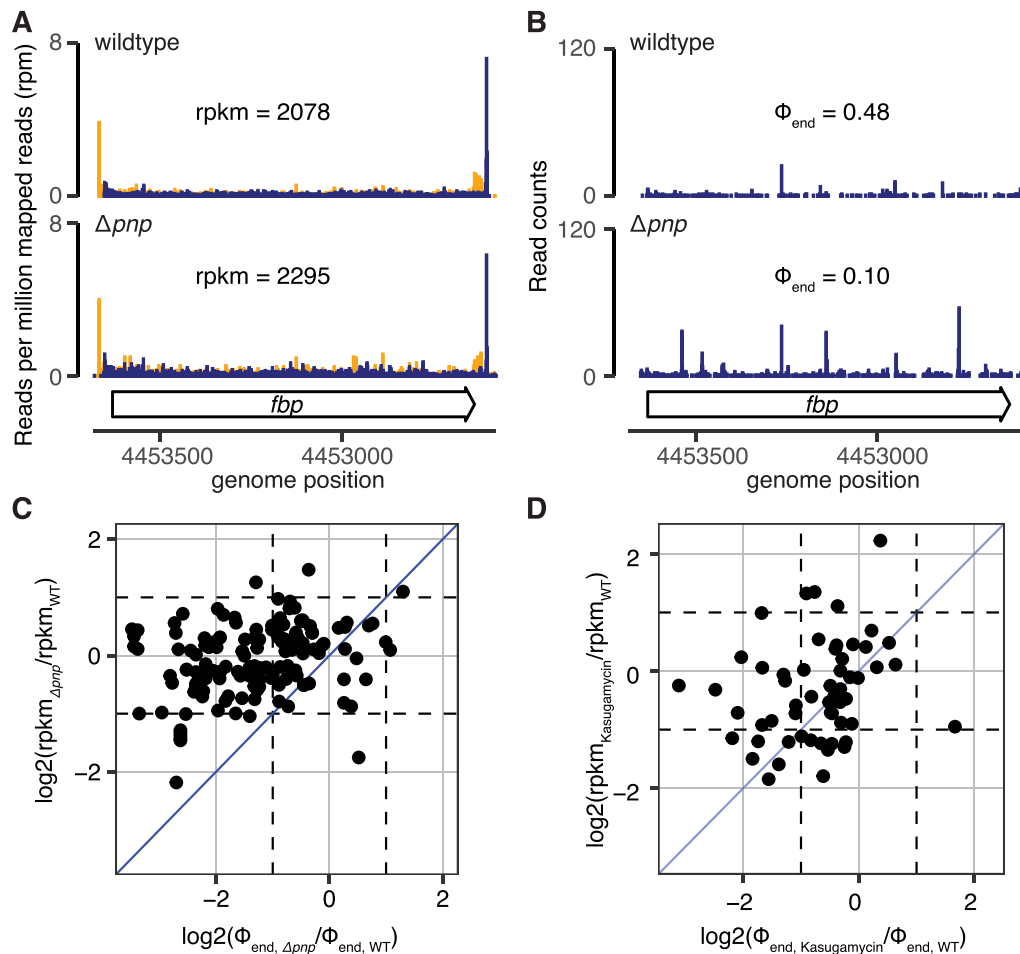
Figure 5. Transcriptome heterogeneity confounds quantification of functional mRNAs by RNA-seq. (**A**) Representative RNA-seq data in wildtype *E. coli* and Δ*pnp*. Rend-seq data show coverage across *fbp* (colored as in Figure 1D). The reads per kilobase mapped per million reads (rpkm) in each sample are calculated using 3'-mapped reads, excluding the first and last 50 nts of the gene. (**B**) Representative 3' end sequencing data for the same samples as in (A). Φ_{end} values are indicated. (C, D) Fold-changes in Φ_{end} and rpkm values for simple TUs between Δ*pnp* and wildtype (**C**), and wildtype with and without Kasugamycin treatment (**D**). Each dot represents one transcription unit. Dashed lines correspond to 2-fold changes in either Φ_{end} or rpkm, whereas blue lines mark the diagonal that would indicate the same degree of change by both metrics. In (D), cells were treated with the translation inhibitor Kasugamycin for 15 min.

DISCUSSION

We devised an approach to obtain a molecular representation of the *in vivo* transcriptome that reflects its life cycles and capabilities for translation. We find that in *E. coli*, mature full-length mRNAs are often not the dominant species. Most partial transcripts are *in vivo* RNA decay intermediates, albeit they are less prevalent in the ribosome-bound transcriptome. We showed that the presence of decay intermediates in total RNA can distort quantifications for the abundance of full-length, functional mRNAs.

By profiling individual 3' ends of cellular RNA, we obtain a single-nucleotide perspective of transient RNA decay dynamics that are possibly linked to other aspects of gene regulation. 3' ends generated by endonucleoytic RNase cleavage are unstable in *E. coli* due to the presence of 3'–5' exonucleases (6–10). Hence, we were initially surprised to pick up such cleavage signatures, but it highlights that our approach captures snapshots of *in vivo* dynamics and transient intermediates of RNA decay. Consistent with previ-

ous studies, we detected a sizeable fraction of adenylated transcripts (64,65) and found that very short tails of non-templated As are present at RNA 3' ends at steady-state (58). The length of A-tails likely reflects the balance of concomitant adenylation and deadenylation by Poly(A) polymerase PcnB and exonucleases, respectively.

Changes to transcriptome heterogeneity are conceivable in conditions where mRNA decay is perturbed, which would in turn lead to translational stress. Upon deletion of non-essential enzymes involved in mRNA decay in *E. coli* or inhibition of translation initiation (66), we detected decreases in Φ_{end}, with the strongest effect observed for the deletion of PNPase, which also showed a strong growth defect (Figure 5 and Supplementary Figure S5). We found that the increases in partial transcripts upon PNPase deletion are accompanied with accumulation of stalled ribosomes on non-stop RNAs, as indicated by the increased demand for ribosome rescue by *trans*-translation and through the alternative ribosome rescue factor *arfA* (Supplementary Figure S5). Similar responses have been detected when deplet-

ing the ribosome recycling factor RRF, overexpression of the ribosome-associated mRNA cleaving toxin RelE, and in the PNPase deletion in *B. subtilis* (67–69). In general, environmental responses and stress conditions might result in changes of transcriptome heterogeneity, e.g., due to altered mRNA decay (3), that manifest themselves in a pronounced ribosome rescue response.

The comparison between RNA-seq and 3′ end sequencing data upon PNPase deletion reveals limitations in RNA-seq to quantify functional mRNAs (Figure 5). Conventional RNA-seq does not distinguish between partial and full-length transcripts and thus reports average expression values that can arise from multiple possible combinations of partial and full-length transcripts. Many partial transcripts are longer than the RNA size cutoffs that are commonly used during purification, such as the 200-nt cutoff in many column-based RNA extraction methods (Figures 1E, 3C and Supplementary Figure S1). Thus, this population can substantially contribute to the expression estimate despite being non-functional for protein production in many cases. Similar limitations may apply to other RNA quantification methods, such as RT-qPCR which is based on amplification of ∼100-nt regions. Future improvements to full-length RNA sequencing approaches, especially mitigations to current length biases, will greatly advance our ability to quantify the functional part of the transcriptome.

The high transcriptome diversity due to RNA decay intermediates requires consideration in all biological systems. In eukaryotes, mRNAs go through even more transitions during their life cycles, including co-transcriptional splicing, cleavage and polyadenylation, nuclear export, deadenylation, decapping, and exonucleolytic decay. Although these populations can be coarsely differentiated by various enrichment methods, it remains difficult to precisely quantify the levels of translatable mRNAs. For example, RNA-seq data generated from poly(A)-selected samples have characteristic enrichment in the 3′ portion of transcripts, which could arise from RNA decay intermediates *in vivo* and *in vitro* (70). Additionally, poly(A) selection could introduce biases based on the poly(A)-tail length and the overall transcript length due to different capture conditions and affinities to oligo(dT) (70–72). On the other hand, alternative methods to remove ribosomal RNA by targeted depletion do not distinguish functional mRNAs from nascent and decay intermediates, which may arise from the widespread co-translational degradation, 3′-UTR fragments, and mis-spliced products (73–76). Together, these considerations suggest that common RNA-seq methods for eukaryotes may be similarly blinded to, or even distorted by, changes in the population structure of mRNAs when conditions change. Our study shows that the ability to resolve the exact forms of RNAs will be important for both quantifying functional mRNAs and studying the *in vivo* kinetics across their life cycles.

DATA AVAILABILITY

The sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE189181. Scripts used for read processing and mapping can be found on github.com/gwlilabmit/Herzel2022_RNAdemographics.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Kuwano,M., Ono,M., Endo,H., Hori,K., Nakamura,K., Hirota,Y. and Ohnishi,Y. (1977) Gene affecting longevity of messenger RNA: a mutant of Escherichia coli with altered mRNA stability. *Mol. Gen. Genet. MGG*, **154**, 279–285.
2. Bernstein,J.A., Khodursky,A.B., Lin,P.-H., Lin-Chao,S. and Cohen,S.N. (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 9697–9702.
3. Vargas-Blanco,D.A. and Shell,S.S. (2020) Regulation of mRNA stability during bacterial stress responses. *Front. Microbiol.*, **11**, 2111.
4. Vogel,U., Sørensen,M., Pedersen,S., Jensen,K.F. and Kilstrup,M. (1992) Decreasing transcription elongation rate in *Escherichia coli* exposed to amino acid starvation. *Mol. Microbiol.*, **6**, 2191–2200.
5. Proshkin,S., Rahmouni,A.R., Mironov,A. and Nudler,E. (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*, **328**, 504–508.
6. Cohen,S.N. and McDowall,K.J. (1997) RNase E: still a wonderfully mysterious enzyme. *Mol. Microbiol.*, **23**, 1099–1106.
7. Hui,M.P., Foley,P.L. and Belasco,J.G. (2014) Messenger RNA degradation in bacterial cells. *Annu. Rev. Genet.*, **48**, 537–559.

8. Mohanty,B.K. and Kushner,S.R. (2016) Regulation of mRNA decay in bacteria. *Annu. Rev. Microbiol.*, **70**, 25–44.

9. Bechhofer,D.H. and Deutscher,M.P. (2019) Bacterial ribonucleases and their roles in RNA metabolism. *Crit. Rev. Biochem. Mol. Biol.*, **54**, 242–300.

10. Trinquier,A., Durand,S., Braun,F. and Condon,C. (2020) Regulation of RNA processing and degradation in bacteria. *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, **1863**, 194505.

11. Donovan,W.P. and Kushner,S.R. (1986) Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. U.S.A.*, **83**, 120–124.

12. Fazal,F.M., Koslover,D.J., Luisi,B.F. and Block,S.M. (2015) Direct observation of processive exoribonuclease motion using optical tweezers. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 15101–15106.

13. Yarchuk,O., Jacques,N., Guillerez,J. and Dreyfus,M. (1992) Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *J. Mol. Biol.*, **226**, 581–596.

14. Richardson,J.P. (2002) Rho-dependent termination and ATPases in transcript termination. *Biochim. Biophys. Acta BBA - Gene Struct. Expr.*, **1577**, 251–260.

15. Keiler,K.C. (2015) Mechanisms of ribosome rescue in bacteria. *Nat. Rev. Microbiol.*, **13**, 285–297.

16. Müller,C., Crowe-McAuliffe,C. and Wilson,D.N. (2021) Ribosome rescue pathways in bacteria. *Front. Microbiol.*, **12**, 558.

17. Laalami,S., Zig,L. and Putzer,H. (2014) Initiation of mRNA decay in bacteria. *Cell. Mol. Life Sci.*, **71**, 1799–1828.

18. Yan,B., Boitano,M., Clark,T.A. and Ettwiller,L. (2018) SMRT-Cappable-seq reveals complex operon variants in bacteria. *Nat. Commun.*, **9**, 3676.

19. Ju,X., Li,D. and Liu,S. (2019) Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nat. Microbiol.*, **4**, 1907–1918.

20. Grünberger,F., Ferreira-Cerca,S. and Grohmann,D. (2022) Nanopore sequencing of RNA and cDNA molecules in Escherichia coli. *RNA*, **28**, 400–417.

21. Mondal,S., Yakhnin,A.V., Sebastian,A., Albert,I. and Babitzke,P. (2016) NusA-dependent transcription termination prevents misregulation of global gene expression. *Nat. Microbiol.*, **1**, 15007.

22. Dar,D. and Sorek,R. (2018) Extensive reshaping of bacterial operons by programmed mRNA decay. *PLoS Genet.*, **14**, e1007354.

23. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.0008.

24. Thomason,L.C., Costantino,N. and Court,D.L. (2007) E. coli genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.*, **Chapter 1**, 1.17.1–1.17.8.

25. Parker,D.J., Lalanne,J.-B., Kimura,S., Johnson,G.E., Waldor,M.K. and Li,G.-W. (2020) Growth-optimized aminoacyl-tRNA synthetase levels prevent maximal tRNA charging. *Cell Syst.*, **11**, 121–130.

26. Sezonov,G., Joseleau-Petit,D. and D'Ari,R. (2007) *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.*, **189**, 8746–8749.

27. Johnson,G.E. and Li,G.-W. (2018) Chapter ten - genome-wide quantitation of protein synthesis rates in bacteria. In: Carpousis,A.J. (ed). *Methods in Enzymology, High-Density Sequencing Applications in Microbial Molecular Genetics*. Academic Press, Vol. **612**, pp. 225–249.

28. Lalanne,J.-B., Taggart,J.C., Guo,M.S., Herzel,L., Schieler,A. and Li,G.-W. (2018) Evolutionary convergence of pathway-specific enzyme expression stoichiometry. *Cell*, **173**, 749–761.

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

30. Homann,O.R. and Johnson,A.D. (2010) MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol.*, **8**, 49.

31. Robinson,J.T., Thorvaldsdóttir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.

32. Stead,M.B., Agrawal,A., Bowden,K.E., Nasir,R., Mohanty,B.K., Meagher,R.B. and Kushner,S.R. (2012) RNA snap™: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res.*, **40**, e156.

33. Larson,M.H., Mooney,R.A., Peters,J.M., Windgassen,T., Nayak,D., Gross,C.A., Block,S.M., Greenleaf,W.J., Landick,R. and Weissman,J.S. (2014) A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science*, **344**, 1042–1047.

34. Mohammad,F., Green,R. and Buskirk,A.R. (2019) A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *eLife*, **8**, e42591.

35. Rao,X., Huang,X., Zhou,Z. and Lin,X. (2013) An improvement of the 2^(–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostat. Bioinforma. Biomath.*, **3**, 71.

36. Pobre,V. and Arraiano,C.M. (2015) Next generation sequencing analysis reveals that the ribonucleases RNase II, RNase R and PNPase affect bacterial motility and biofilm formation in E. coli. *BMC Genomics*, **16**, 72.

37. Schindelin,J., Arganda-Carreras,I., Frise,E., Kaynig,V., Longair,M., Pietzsch,T., Preibisch,S., Rueden,C., Saalfeld,S., Schmid,B. *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, **9**, 676–682.

38. Yang,W. (2011) Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.*, **44**, 1–93.

39. Herzel,L. (2015) Co-transcriptional splicing in two yeasts. Technische Universität Dresden, https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-179274.

40. Dar,D., Shamir,M., Mellin,J.R., Koutero,M., Stern-Ginossar,N., Cossart,P. and Sorek,R. (2016) Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science*, **352**, aad9822.

41. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

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

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

44. Wellner,K., Czech,A., Ignatova,Z., Betat,H. and Mörl,M. (2018) Examining tRNA 3′-ends in Escherichia coli: teamwork between CCA-adding enzyme, RNase T, and RNase R.*RNA*, **24**, 361–370.

45. Kao,C., Zheng,M. and Rüdisser,S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA*, **5**, 1268–1272.

46. Bremer,D. and Dennis,P.P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. *E. coli SalmonellaCell Mol. Biol.*, **2**, 97.

47. Dong,H., Nilsson,L. and Kurland,C.G. (1996) Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.*, **260**, 649–663.

48. Clarke,J.E., Kime,L., Romero A.,D. and McDowall,K.J. (2014) Direct entry by RNase E is a major pathway for the degradation and processing of RNA in *Escherichia coli*. *Nucleic Acids Res.*, **42**, 11733–11751.

49. McDowall,K.J., Lin-Chao,S. and Cohen,S.N. (1994) A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J. Biol. Chem.*, **269**, 10790–10796.

50. Chao,Y., Li,L., Girodat,D., Förstner,K.U., Said,N., Corcoran,C., Śmiga,M., Papenfort,K., Reinhardt,R., Wieden,H.-J. *et al.* (2017) In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. *Mol. Cell*, **65**, 39–51.

51. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.

52. Moffitt,J.R., Pandey,S., Boettiger,A.N., Wang,S. and Zhuang,X. (2016) Spatial organization shapes the turnover of a bacterial transcriptome. *eLife*, **5**, 636.

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

54. Carrillo Oesterreich,F., Herzel,L., Straube,K., Hujer,K., Howard,J. and Neugebauer,K.M. (2016) Splicing of nascent RNA coincides with intron exit from RNA polymerase II. *Cell*, **165**, 372–381.

55. Vvedenskaya,I.O., Vahedian-Movahed,H., Bird,J.G., Knoblauch,J.G., Goldman,S.R., Zhang,Y., Ebright,R.H. and Nickels,B.E. (2014) Interactions between RNA polymerase and the 'core recognition element' counteract pausing. *Science*, **344**, 1285–1289.

56. Imashimizu,M., Takahashi,H., Oshima,T., McIntosh,C., Bubunenko,M., Court,D.L. and Kashlev,M. (2015) Visualizing translocation dynamics and nascent transcript errors in paused RNA polymerases in vivo. *Genome Biol.*, **16**, 98.

57. Kushner,S.R. (2015) Polyadenylation in E. coli: a 20 year odyssey. *RNA*, **21**, 673–674.

58. Waters,S.A., McAteer,S.P., Kudla,G., Pang,I., Deshpande,N.P., Amos,T.G., Leong,K.W., Wilkins,M.R., Strugnell,R., Gally,D.L. *et al.* (2017) Small RNA interactome of pathogenic E. coli revealed through crosslinking of RNase E. *EMBO J.*, **36**, 374–387.

59. Mathy,N., Bénard,L., Pellegrini,O., Daou,R., Wen,T. and Condon,C. (2007) 5′-to-3′ Exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5′ stability of mRNA. *Cell*, **129**, 681–692.

60. Johnson,G.E., Lalanne,J.-B., Peters,M.L. and Li,G.-W. (2020) Functionally uncoupled transcription–translation in Bacillus subtilis. *Nature*, **585**, 124–128.

61. Bernstein,J.A., Lin,P.-H., Cohen,S.N. and Lin-Chao,S. (2004) Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 2758–2763.

62. Chadani,Y., Ono,K., Kutsukake,K. and Abo,T. (2011) Escherichia coli YaeJ protein mediates a novel ribosome-rescue pathway distinct from SsrA- and ArfA-mediated pathways. *Mol. Microbiol.*, **80**, 772–785.

63. Chadani,Y., Ono,K., Ozawa,S., Takahashi,Y., Takai,K., Nanamiya,H., Tozawa,Y., Kutsukake,K. and Abo,T. (2010) Ribosome rescue by *Escherichia coli* ArfA (YhdL) in the absence of trans-translation system. *Mol. Microbiol.*, **78**, 796–808.

64. Maes,A., Gracia,C., Innocenti,N., Zhang,K., Aurell,E. and Hajnsdorf,E. (2017) Landscape of RNA polyadenylation in *E. coli*. *Nucleic Acids Res.*, **45**, 2746–2756.

65. Mohanty,B.K. and Kushner,S.R. (2006) The majority of Escherichia coli mRNAs undergo post-transcriptional modification in exponentially growing cells. *Nucleic Acids Res.*, **34**, 5695–5704.

66. Schneider,E., Blundell,M. and Kennell,D. (1978) Translation and mRNA decay. *Mol. Gen. Genet. MGG*, **160**, 121–129.

67. Hwang,J.-Y. and Buskirk,A.R. (2017) A ribosome profiling study of mRNA cleavage by the endonuclease RelE. *Nucleic Acids Res.*, **45**, 327–336.

68. Saito,K., Green,R. and Buskirk,A.R. (2020) Ribosome recycling is not critical for translational coupling in Escherichia coli. *eLife*, **9**, e59974.

69. Oussenko,I.A., Abe,T., Ujiie,H., Muto,A. and Bechhofer,D.H. (2005) Participation of 3'-to-5' exoribonucleases in the turnover of Bacillus subtilis mRNA. *J. Bacteriol.*, **187**, 2758–2767.

70. Weinberg,D.E., Shah,P., Eichhorn,S.W., Hussmann,J.A., Plotkin,J.B. and Bartel,D.P. (2016) Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.*, **14**, 1787–1799.

71. Kojima,S., Sher-Chen,E.L. and Green,C.B. (2012) Circadian control of mRNA polyadenylation dynamics regulates rhythmic protein expression. *Genes Dev.*, **26**, 2724–2736.

72. Eisen,T.J., Eichhorn,S.W., Subtelny,A.O., Lin,K.S., McGeary,S.E., Gupta,S. and Bartel,D.P. (2020) The dynamics of cytoplasmic mRNA metabolism. *Mol. Cell*, **77**, 786–799.

73. Adiconis,X., Borges-Rivera,D., Satija,R., DeLuca,D.S., Busby,M.A., Berlin,A.M., Sivachenko,A., Thompson,D.A., Wysoker,A., Fennell,T. *et al.* (2013) Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods*, **10**, 623–629.

74. Pelechano,V., Wei,W. and Steinmetz,L.M. (2015) Widespread co-translational RNA decay reveals ribosome dynamics. *Cell*, **161**, 1400–1412.

75. Sudmant,P.H., Lee,H., Dominguez,D., Heiman,M. and Burge,C.B. (2018) Widespread accumulation of ribosome-associated isolated 3′ UTRs in neuronal cell populations of the aging brain. *Cell Rep.*, **25**, 2447–2456.

76. Herzel,L., Straube,K. and Neugebauer,K.M. (2018) Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.*, **28**, 1008–1019.