Supplemental methods:

Overlap of pausing/terminator/start sites and statistics

Sites from other *E. coli* strains than BW25113 were mapped to the BW25113 genome by perfect string matching. 3' sites (term-seq) and 5' sites (dRNA-seq) were from(1) and pausing sites from (2, 3). We additionally used the experimentally verified terminator list from RegulonDB (RegulonDB 10.10, 02/28/2022) (4). Pausing sites from the Larson data within a range of 50 nt were clustered to one representative site. For the terminators we allowed an overlap of +/- 2bins, i.e +/-100 nt to account for the relative coarse 'rifi' site detection. Due to the high number of sites in the Larson pausing set the overlap for pausing sites was set to +/- 1bins, i.e +/-50 nt. The p-values for the overlaps being higher than expected by chance were calculated with the exact binomial test. Based on the estimate, that roughly the whole genome is transcribed either in forward or reverse direction, we used a total transcriptome size of 4,631,470 nt. The transcriptome size was divided by the allowed overlap window size into n windows. The assumed random probability for hitting a site is the number of sites in the reference divided by the number of windows $P_{random} = \frac{\#_{sites}}{n}$. The quasi-proportional Venn diagrams were done with the nVennR R-package (5).

Evaluation of pausing site detection sensitivity and specificity

Stochastic simulation was performed for the transcription of a 2000 nt transcript in triplicates with three synthesis rates (0.2, 0.4 and 0.8 initiations/s), two different transcription velocities (15 and 30 nt/s) and three half lifes (1, 2 and ~3.33 min) corresponding to the decay constants (0.003465736, 0.01155 and 0.00578) and 4 different pausing off-rates (0.02, 0.03, 0.05, 0.1) corresponding to the average pausing life times of (50, 33.3, 20 and 10s) for a pausing site at position 1000. The pausing start probability was 1. This leads to in total 72 different parameter combinations. The number of transcripts was counted every 50nt for the timepoints 0, 1, 2, 3, 4, 6, 8, 10, 15 and 20 min after rifampicin addition, as for the Dar & Sorek dataset. Another simulation was done with a higher time resolution (0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.2, 2.5, 2.8, 3, 3.2, 3.5, 3.8, 4, 4.2, 4.5, 4.8, 5, 5.2, 5.5, 5.8, 6, 7, 8, 9, 10, 15, 20 min). The data (n=3), with or without added poisson noise, were fitted with the co-transcriptional decay model to estimate the delay for each position. The delay values were then segmented by the 'rifi' fragmentation method with 73 different split and outlier penalty combinations (split penalty from 0.5 to 4, outlier penalty from 0.3 to 3.2). Each simulation and fitting cycle was repeated 50 times leading to 72x50 = 3600 segmentations with a simulated pausing for each penalty pair. A correct split needs to be in a window of +/- 2 bins and the p-value needs to be $p \leq 0.01$. There could not be more than 1 correct split for each segmentation. First the number of true positive splits (TP) for each parameter set (50 replications) was counted and divided by the number of the actual splits, i.e $TP + FN = 50$ to get the true positive rate $TPR_{parset} = \frac{TP}{50}$, yielding in

total 72 parameter dependent TPR_{parset} for each penalty set. These TPR_{parset} were averaged to get the final penalty-set dependent TPR_{penset}. To calculate the final penalty-set dependent false positive rate FPR_{penset}, again the 72 parameter dependent FPR_{parset} were averaged. The $FPR_{parset} = \frac{FP}{FP + TN}$ was calculated from the FPs, i.e. the total number of splits minus the true positives, and the true negatives (TN). For each parameter set we did 50 replications and each replication contains exactly one true positive split, thus there is no replication without a split and the number of true negatives is 50. The 72 FPR_{penset} and FPR_{penset} were used for the ROC-like curves in Fig. 5.

Velocity sensitivity and specificity

To test the accuracy of the velocity estimate a 1000nt transcript with constant elongation rate was simulated as described above. In total 36 parameter combinations (decay constants: 0.003465736, 0.01155 and 0.00578 1/s; initiation rates: 0.2, 0.4 and 0.8 initiations/s) including 4 different elongation rates (15, 30, 45 and 60 nt/s) were used. The simulations for each parameter set were repeated 50 times and the resulting data segmented with the penalties used for the Dar & Sorek data (split penalty 3.5, outlier penalty 2.5). The velocities of the resulting segments were compared with the input velocities.

In order to investigate the segmentation performance upon a velocity change the simulation was done for a 1000nt RNA with a transcription velocity change at 500nt. The parameters were as above besides that all 12 possible speed changes were considered, which leads to 108 different parameter sets. The calculation of TPR and FPR based on different segmentation penalties was done as described above.

Internal start sites

Simulation was done as above with 108 different parameter sets (half-life: 1, 2; velocity: 15 or 30 nt/s; synthesis rate 1 + 2: 0.2, 0.4 or 0.8 molecules/s; iTSS position: 500, 1000 or 2000nt), with a data point every 100 nt instead of every 50nt. The calculation of TPR and FPR based on different segmentation penalties was done as described above.

Supplementary Dataset 1 (.zip):

Contains 5 Excel files with the main rifi results for *E. coli* BW25113, *E. coli* MG1655, *Klebsiella aerogenes* KCTC2190, *Synechococcus* PCC 7002 and *Synechocystis* PCC 6803. Each file contains 3 tabs.

Segments summary - tab1: This tab contains all “intensity segments” of the rifi analysis. Due to the hierarchical segmentation approach (see methods), the intensity segments are the final segment type, corresponding to a transcript segment that contains bins/probes of equal behavior, i.e. of equal intensity, stability (half-life) and a steadily linearly increasing delay.

Events - tab2: This tab contains all events investigated by rifi with their p-values and further information. Potential events appear at the border of two consecutive segments in the same transcriptional unit (TU). Pausing sites (ps), internal start sites (iTSS_I) and elongation rate changes (velocity) are estimated based on the delay segments. Processing -or stabilization sites leading to a differential operon decay and different stabilities/half-lives of two consecutive segments (HL-event) are estimated from half-life segments. Different intensities (Int_events) are estimated between intensity segments and the synthesis rate based events for termination (Termination) and internal start sites (iTSS_II) are calculated from half-life segments and intensity segments.

Rst-Ti - tab3: All instances of rifampicin sensitive termination and potential transcriptional interference.

Supplementary Dataset 2 (.zip):

Full genome visualizations of the ‘rifi’ results of E. coli BW25113, E. coli MG1655, Klebsiella aerogenes KCTC2190, Synechococcus PCC7002 and Synechocystis PCC6803. Each page shows a 10kb segment of both strands with their respective annotations. The ‘rifi’ TU estimates are shown as purple bars with their respective TU-ID. The first tracks (closest to the annotation) show the delay of the onset of the decay for the individual probes. The delay should be linearly increasing for

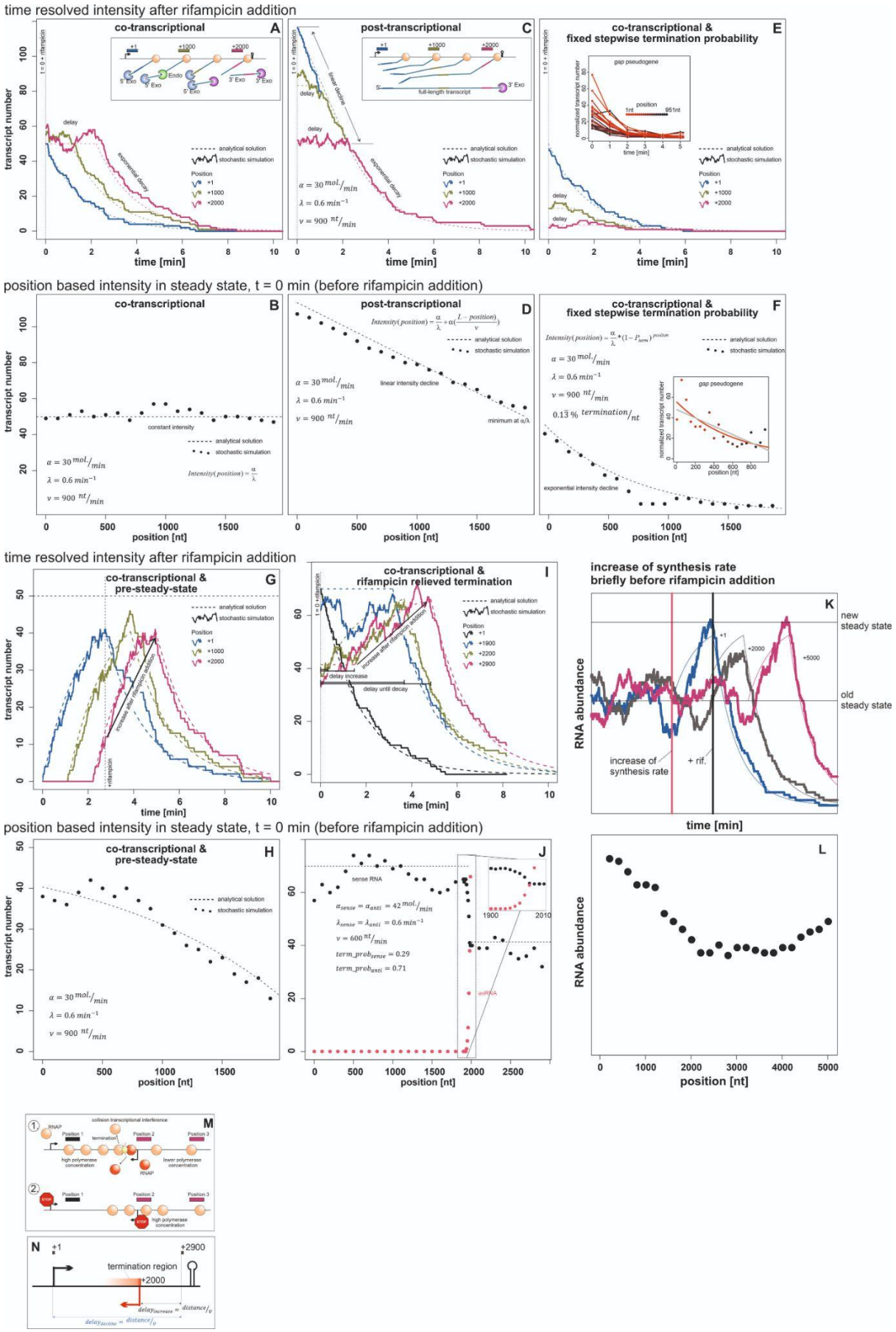
continuous transcripts, which are clustered by dynamic programming and indicated by matching colors and a trendline. A sudden delay increase between two segments indicates a transcription polymerase pausing site (PS), while a sudden decrease indicates a new (internal) transcriptional start site (iTSS). The slope of the delay segment allows us to estimate the speed of the RNA Polymerase. Changes in the velocity are indicated by a "V". The second tracks show the fitted half-life of the probes and the clustered half-life segments. If two segments within the same transcriptional unit have different half-life (HL) a processing/stabilization site for one or the other segment can be assumed. Difference in stability is indicated by “HL”. The third tracks show the intensity and the intensity segments at time point 0 ($mRNA_{t=0}$ or before Rifampicin addition). If two segments within the same transcriptional unit have different intensities, the fold-change (FC) could be either due to a partial termination (Ter) or a new transcriptional start site (NS). Significant events are assigned with an '*'.

Supplementary Dataset 3 (Pdf):

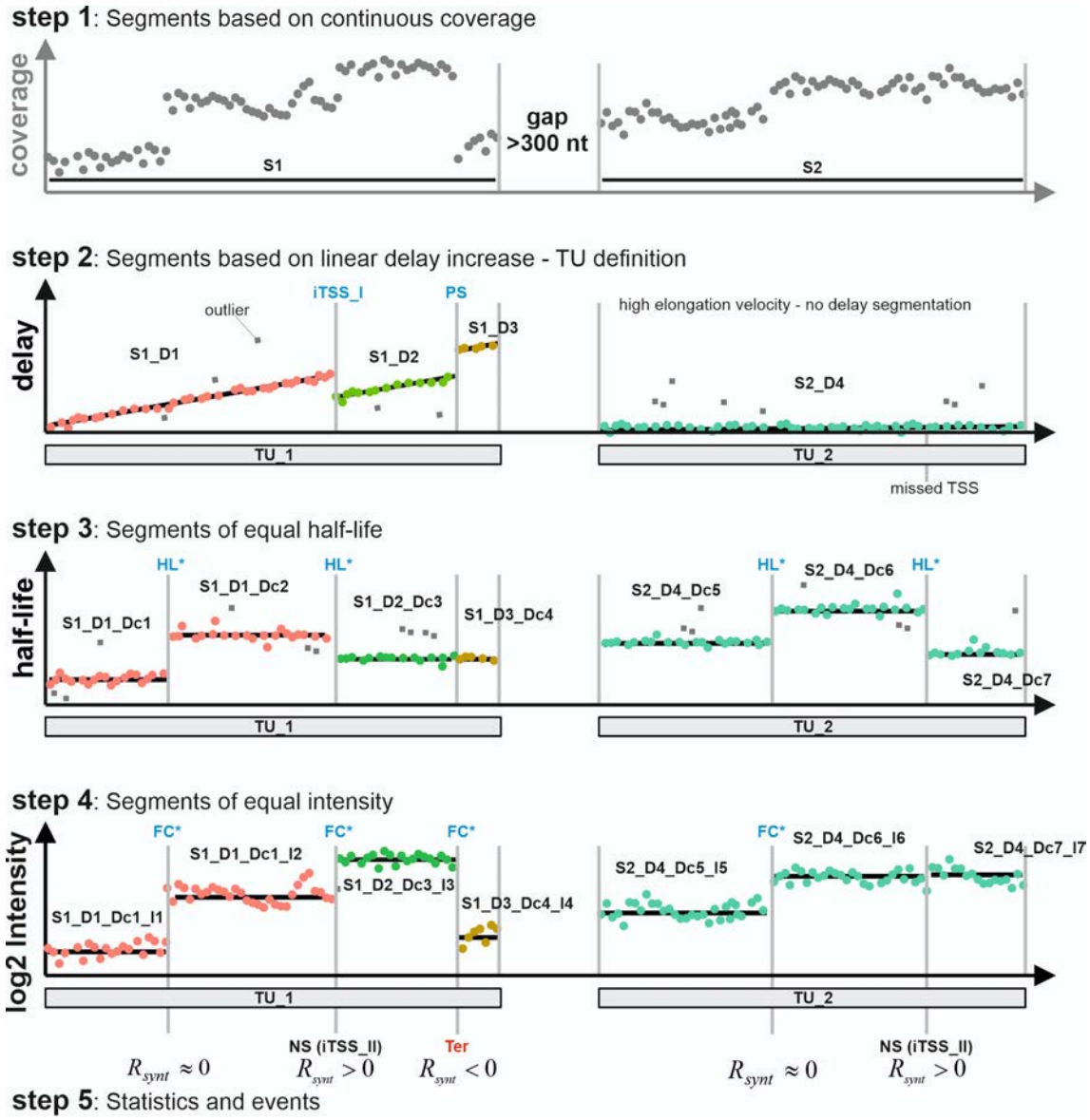
Visualization from selected E. coli RST instances (compare Table 1). The black vertical lines indicate the estimated RST start site. The red horizontal lines represent reported asRNA transcription (Thomasson et al. 2014(6), Supplementary Table 3, Gundy et al. 2022(7)) or sRNA binding sites (*galK*/spot42(8)*, yebK*/*pykA* DicF(9) and *rbsC*/*rbsK* RybB(10))

Supplementary Table 1 (.xlsx)

All sites of differential operon decay (p-value ≤ 0.05, |log2 FC| ≥ 1 min) in E. coli BW25113 detected by ‘rifi’ based on the Dar & Sorek raw data (11). The 31 of 49 recovered sites from Dar & Sorek (11) are indicated.

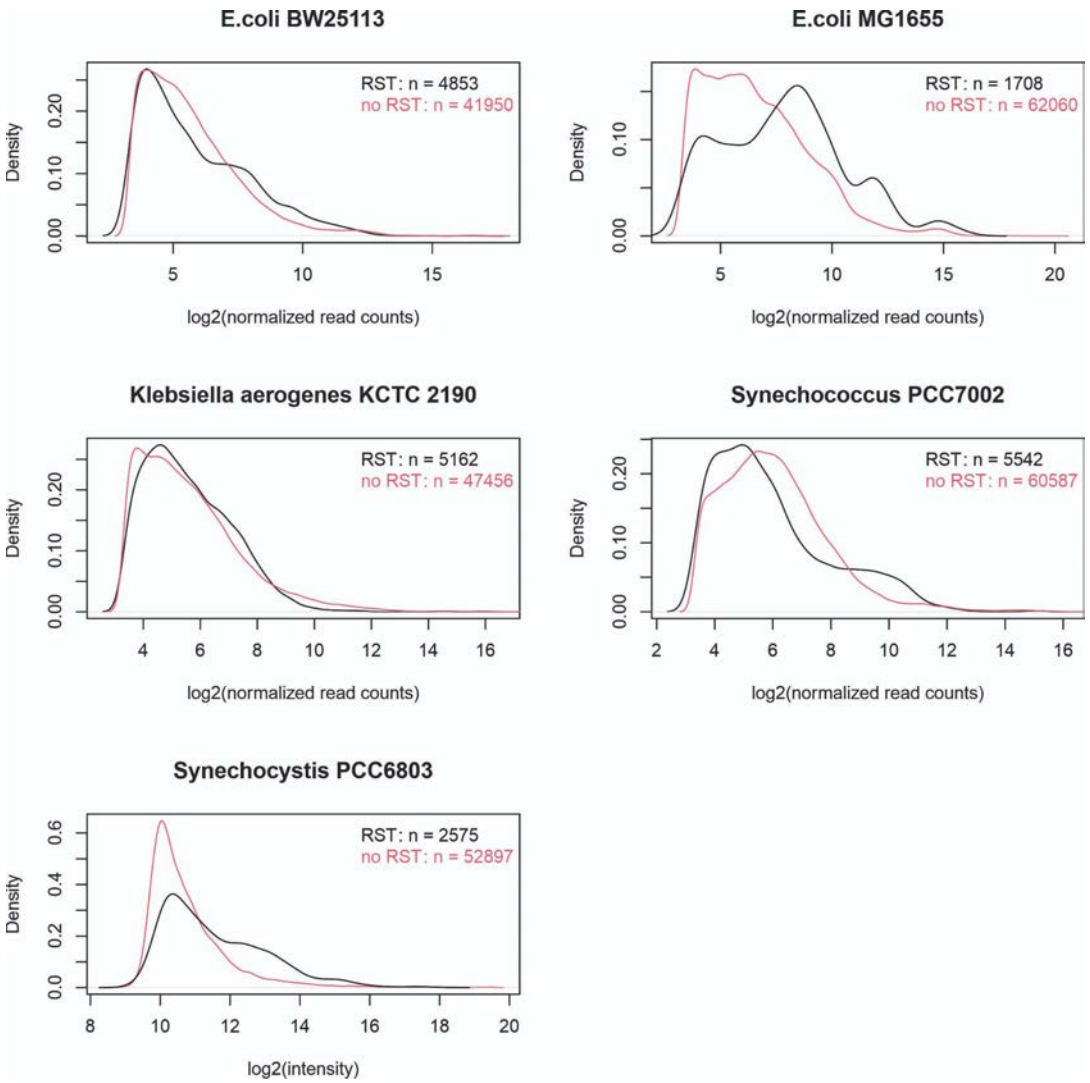

time resolved intensity after rifampicin addition
co-transcriptional
A
post-transcriptional
C
co-transcriptional &
fixed stepwise termination probability
E
transcript number
time [min]
delay
exponential decay
linear decline
analytical solution
stochastic simulation
Position
+1
+1000
+2000
gap pseudogene
position
1nt
951nt
normalized transcript number
full-length transcript
position based intensity in steady state, t = 0 min (before rifampicin addition)
B
D
F
constant intensity
linear intensity decline
minimum at α/λ
exponential intensity decline
position [nt]
time resolved intensity after rifampicin addition
co-transcriptional &
pre-steady-state
G
co-transcriptional &
rifampicin relieved termination
I
increase after rifampicin addition
delay increase
delay until decay
+1900
+2200
+2900
increase of synthesis rate
briefly before rifampicin addition
K
new steady state
old steady state
RNA abundance
increase of synthesis rate
+ rif.
position based intensity in steady state, t = 0 min (before rifampicin addition)
H
J
L
sense RNA
asRNA
collision transcriptional interference
RNAP
termination
high polymerase concentration
lower polymerase concentration
STOP
M
N
termination region
+2900
+2000
delay_increase = distance/v
delay_decline = distance/v

Supplementary Figure 1: **Comparison of the co- and post-transcriptional decay modes**. For all cases we assume that the transcript concentration was in its steady state at the time of rifampicin addition. A) Simulated decay curves for the co-transcriptional case at different positions within a 2000 nt transcript (solid lines). The analytical solutions for each decay curve are given as broken lines. The delay increases with the distance to the transcriptional start site. The transcription parameters for the simulation are given in panel b. The scheme in the inlay shows that no transcript position is overrepresented due to the co-transcriptional degradation by different nucleases. B) Transcript numbers before rifampicin addition (t = 0 min) as a function of the distance to the start site for the co-transcriptional case. The simulated values are given as dots and the analytical solution is given as a broken line. The transcript number is roughly the same (α/λ) for all positions. C) Simulated decay curves for the post-transcriptional case and different transcript positions (solid lines). The analytical solutions are given as broken lines. All parameters are the same as for the co-transcriptional case. The delay is again the time until the last polymerase has passed the respective position. The delay is followed by a linear decline, which resembles the time window needed from the last polymerase, from passing the position to reach the end of the transcript. The last phase is the exponential decay. Interestingly, the curves of all positions fall together for the linear and exponential decline. The scheme in the inlay shows that, if decay is only possible by 3' exonucleases after the transcription of the full-length transcript, the concentration of 5' positions is higher than the concentration of the 3' end. D) Transcript numbers before rifampicin addition (t = 0 min) as a function of the distance to the start site, the post-transcriptional case. The simulated values are given as dots and the analytical solution as a broken line. The concentration is linearly dependent on the distance from the start. The very 3' end has the same steady-state concentration as in the co-transcriptional case (α/λ). The formula for the linear relationship between transcript concentration and position is given in the panel (L = total transcript length). The higher the synthesis rate (α) and the lower the elongation velocity (v) the higher the concentration at the 5' end and the steeper the linear decline. E) Simulation for the co-transcriptional case with the assumption that there is a fixed probability for a transcription termination after each elongation step. Also this case leads to position-dependent concentration decline in the steady state (F), but in this case there is no linear decline part and the decay curves of different positions do not fall together. The inlay shows the decay curves for the bins from the *gap* pseudogene (Dar & Sorek dataset) which resemble this scenario. All parameters beside the termination rate (P_{term}) are the same as in panels A and C. G-N) **RNA abundance-increase AFTER rifampicin addition - Rifampicin-sensitive transcription termination**. G) Example simulation for a gene which has not yet reached steady-state expression at the time of rifampicin addition. The RNA abundance change is shown against the time before/after rifampicin addition for three different positions 3' of the transcriptional start site (TSS). The parameters used for the analytical solution and the simulation are given in panel h. H) RNA abundance at time point 0 before rifampicin addition against the distance from the TSS. I) Simulation for the RNA abundance changes after rifampicin addition in case of a partial internal transcription termination by the collision mechanism via an asRNA. A scheme of the positions of sense and asRNA genes is given in N). Position +1 and +1900 are not affected by the asRNA and show the typical (delayed) exponential decline. Positions +2200 and +2900 show a reduced RNA abundance before rifampicin addition. When the synthesis at the antisense promoter is stopped, queued polymerases lead to a temporary (delayed) increase of RNA abundance. In the shown scenario roughly 30% of the sense RNA polymerases are terminated. The parameters used for the analytical solution and the simulation are given in panel J. We assume that the termination probability upon collision is higher for the non-translated asRNA ($term\_prob_{anti}$=0.71) than for the mRNA ($term\_prob_{sensei}$=0.29). Promoter exclusion and sitting duck interference are not considered. The analytical solution (broken line) assumes a 29% termination exactly at position +2000 and does not consider the small area of a gradually increasing termination rate shortly in front of the asRNA TSS (panel j inlay). K) Example simulation for a gene which has reached a (old) steady-state expression but 3 min before rifampicin addition the synthesis rate was increased by 75% (0.4 mol/s to 0.7 mol/s, $\lambda = 0.01$, $v = 15\, nt/s$). The pattern is very similar to the RST case in panel I, however the abundance change at t=0, shown in panel L, is gradual compared to the sharp drop in the RST case.M) Scheme for RST by the transcriptional interference collision mechanism due to a *cis*-asRNA.

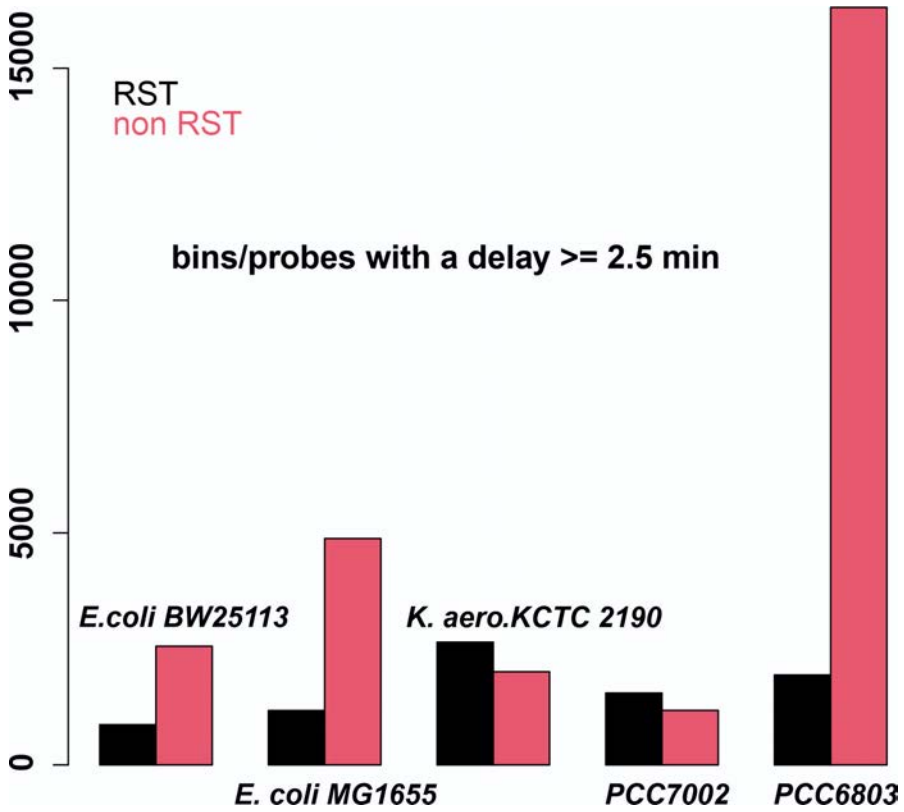


Supplementary Fig. 2: 'rifi' workflow (idealized example). Step 1: All regions with continuous coverage, i.e. regions with no coverage gap greater than a given threshold (default: >300nt), are taken as initial segments. This step dramatically reduces the runtime for the subsequent dynamic programming steps. Each segment is numbered progressively (S1 - Sn; n = number of segments). Step 2: The delay is dependent on the distance of a given position in the transcript from the transcriptional start site. Given a constant elongation rate and no internal transcriptional start site and no pausing site, the delay should increase linearly from the transcriptional start site to the end of the transcript. In this step the best set of linear fits for all bins/probes within the same coverage segment is searched by dynamic programming with the respective penalties for splits and outliers (see methods for more details). A split either represents a new TSS, a pausing site (PS) or a change of transcription velocity. The delay segments are again named progressively and inherit the respective Segment name as prefix (e.g. S1_D1, for the first delay segment or S2_D4 for the fourth delay segment). Outliers that are excluded from the segment are indicated as gray dots. Within a transcriptional unit, i.e. a transcript resulting from the continuous elongation of RNA polymerase from a TSS to the final terminator, the delay should also increase continuously. We use the delay segments for a TU estimate. A new TU is triggered if the delay of a segment starts close to 0 or if the drop in the delay between two consecutive segments is high enough. A minor delay-decrease is taken as internal start sites (iTSS_I, see. Fig. 5A). Note that the elongation velocity within coverage segment S2 is high, which leads to delays of ~0 for all bins of the segment. Here the segmentation based on the delay fails and a start was missed (missed TSS).

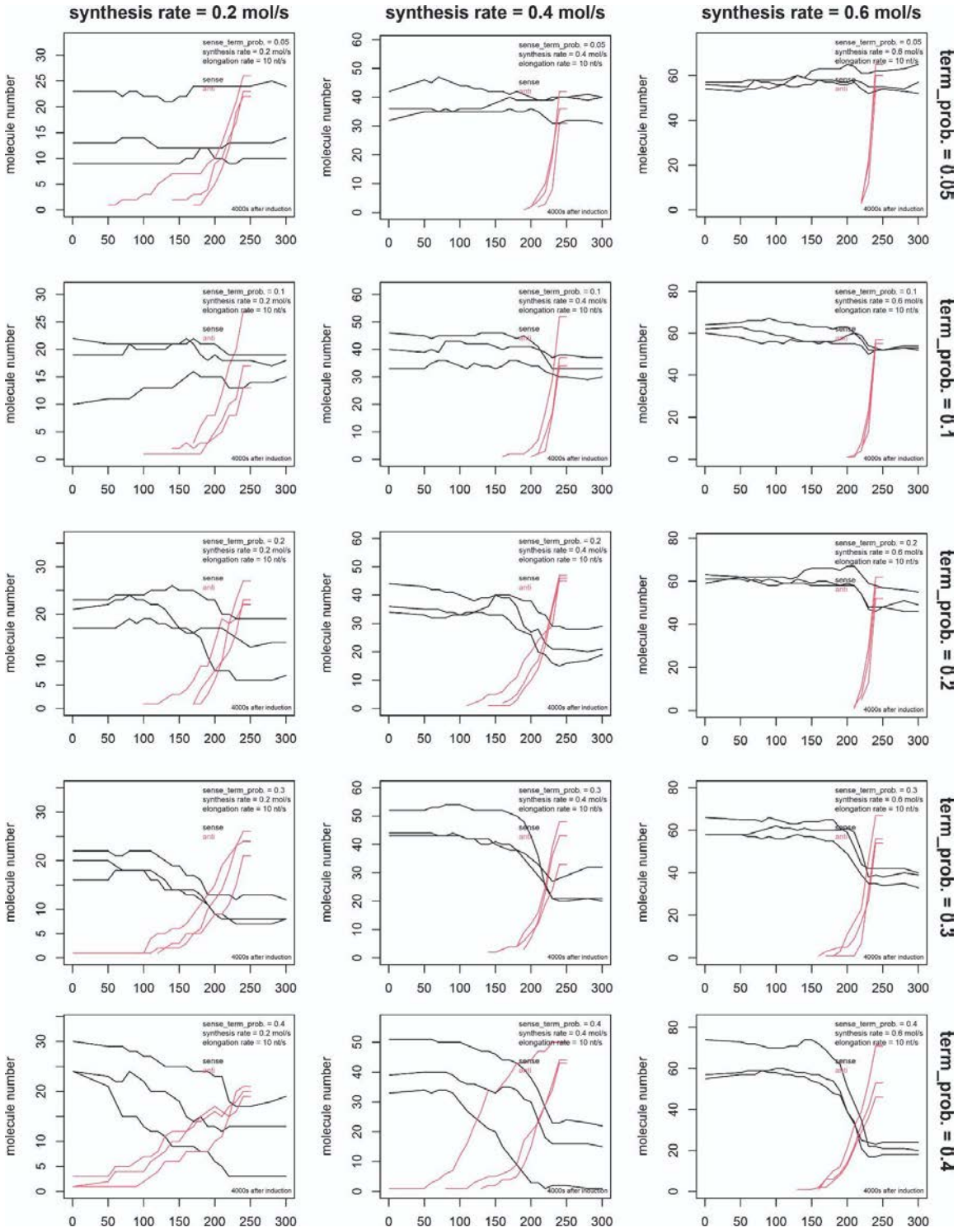
Step 3: Each delay segment is splitted into the best set of half-life segments. The segments are scored by the deviation of the individual bin/probe-wise half-lifes to the mean half-life of the segment. The segments are named progressively and inherit the names of the respective delay segment as prefixes (e.g. S1_D1_Dc3, third half-life (decay - Dc) segment). Step 4: Each half-life segment is splitted into the best set of RNA abundance segments based on the log2 abundance of each bin/probe. Again the splits are optimized to get the lowest deviation of the individual bin/probe-wise intensities to the mean abundance of the segment. The segments are named progressively and inherit the names of the respective delay segment as prefixes (e.g. S2_D4_Dc5_I5). Step 5: 'rifi' investigates if the consecutive segments before and after a split are significantly different from each other. Only splits within the same estimated TU are analyzed. The p-values for all splits are available in the final output. Delay: Consecutive delay segments are analyzed via ANCOVA for differences in the slope (velocity change) or via a two-sided student's t-test for pausing sites (PS) or internal start sites (iTSS_I). Half-life: Consecutive segments are tested via a two-sided student's t-test. Significant splits ($p \leq 0.05$) are indicated by HL* and indicate transcript fragments with different half-lives within the same transcriptional unit. RNA abundance: Consecutive segments are tested via a two-sided student's t-test. Significant splits ($p \leq 0.05$) are indicated by FC*. Synthesis rate based events: RNA abundance changes between two transcript segments in the same TU can be due to stability differences, differences in the synthesis rate (partial termination or internal start site) or a combination of both. We calculate the log2 foldchange of the synthesis rates (R_{synt}) between consecutive segments (see methods). $R_{synt} \approx 0$ indicates no change in the synthesis rate, i.e. the abundance change is only due to different transcript stabilities. $R_{synt} > 0$ indicates a new start (NS, iTSS_II). In the example the missed start in S2 is detected by the R_{synt} measure. $R_{synt} < 0$ indicates a partial termination (Ter). The significance of the R_{synt} value is tested via MANOVA. For further details see methods.



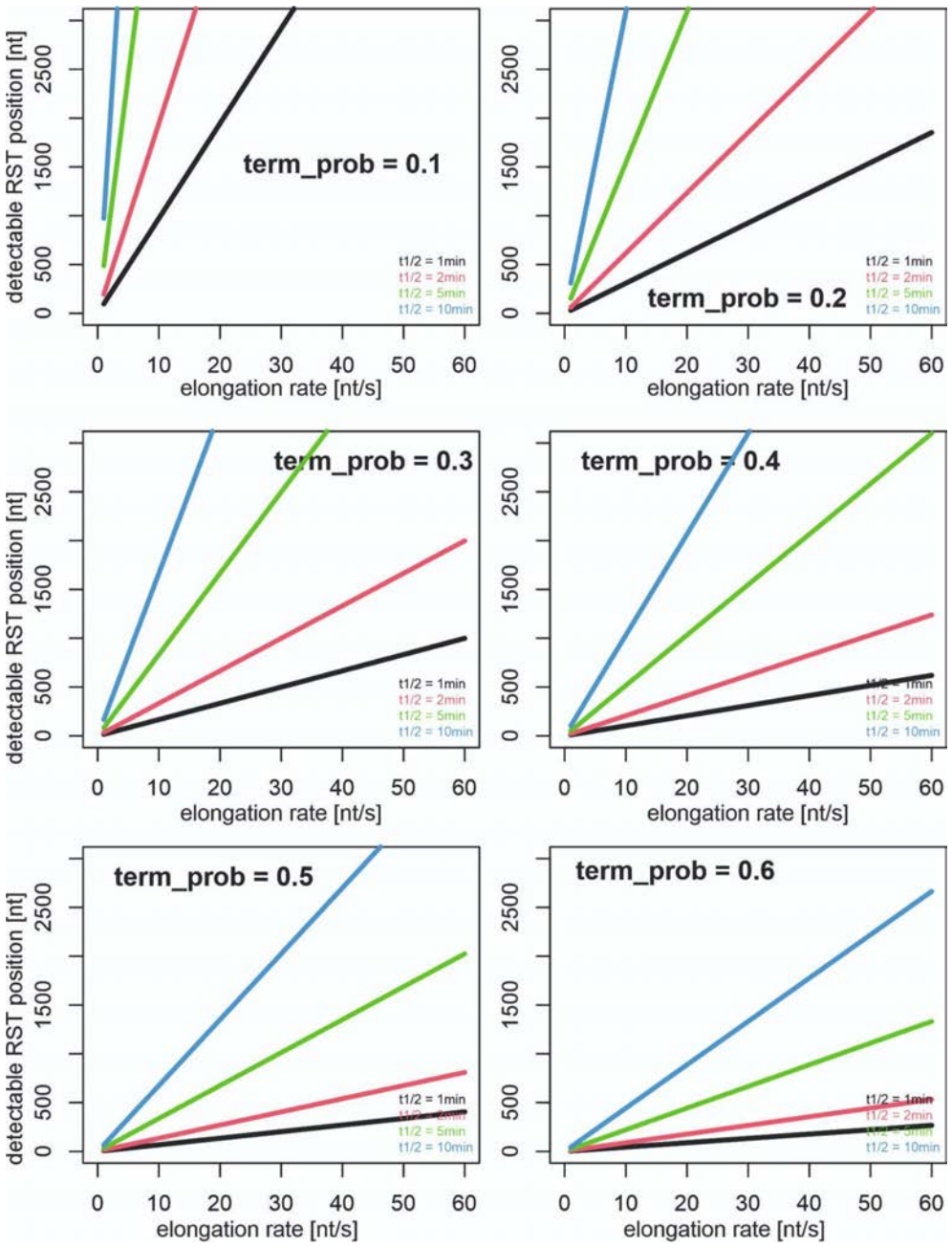
Supplementary Fig. 3: Read count and intensity distributions for RST and non-RST bins/probes in the investigated datasets.



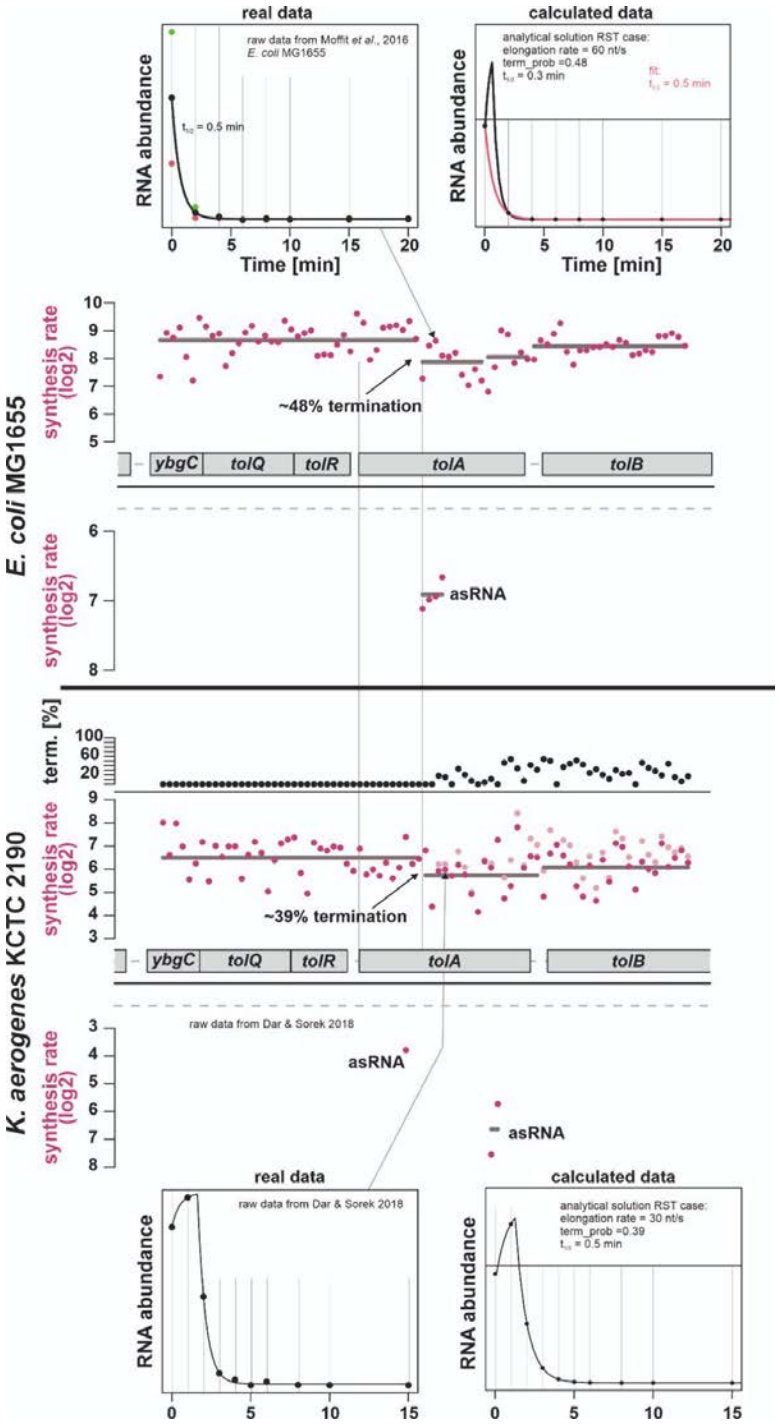
Supplementary Fig. 4: Number of RST and non RST bins/probes with a delay >= 2.5 min.



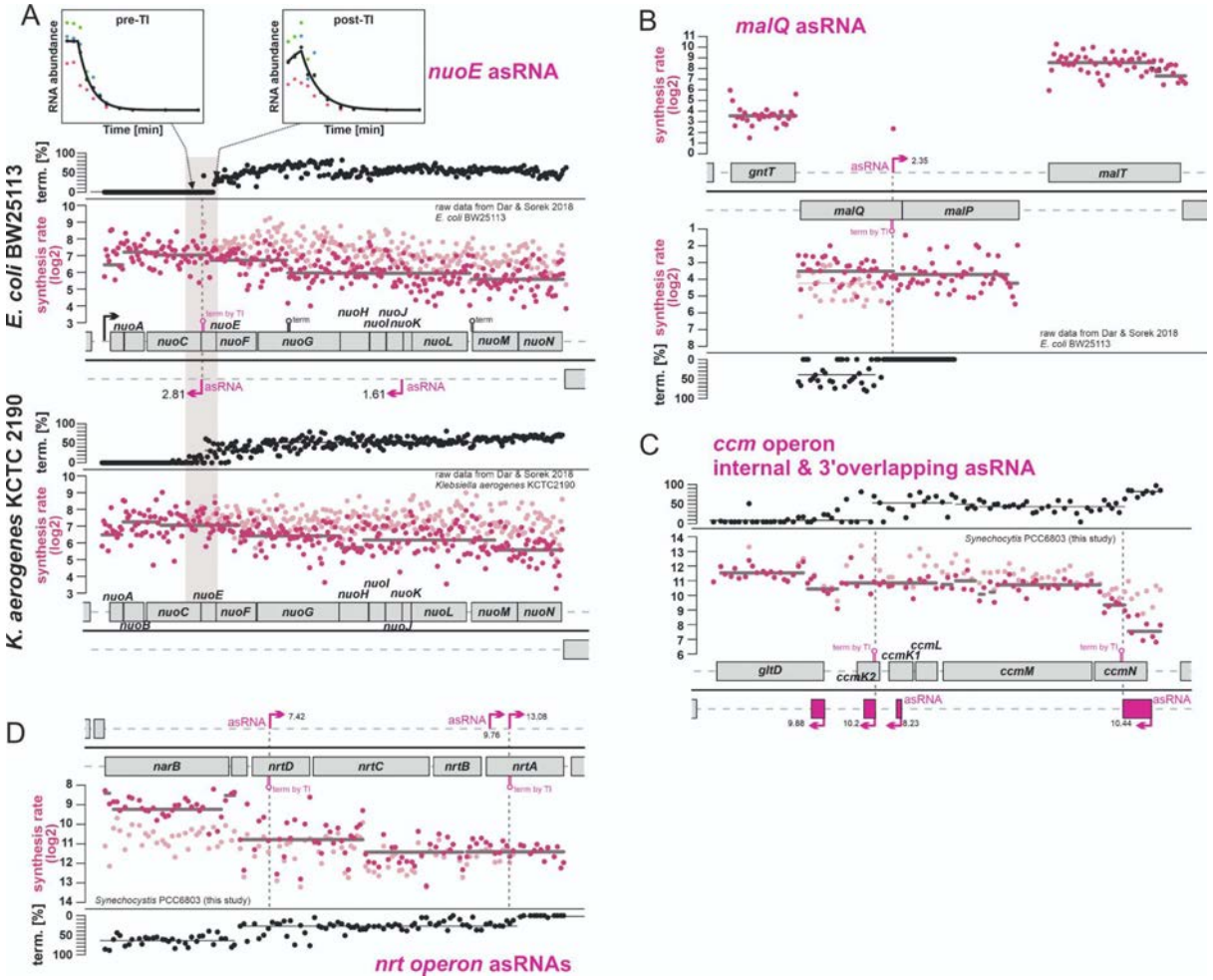
Supplementary Fig. 5: Simulation (n=3) for sense mRNA and asRNA lengths for different sense RNA termination probabilities (asRNA term. prob. = 1- sense RNA term. prob.) and synthesis rates at a fixed degradation constant of 0.01 1/s and a fixed elongation rate of 10 nt/s. The asRNA TSS is at position 250 relative to the sense RNA TSS. The length resolution is 10 nt.



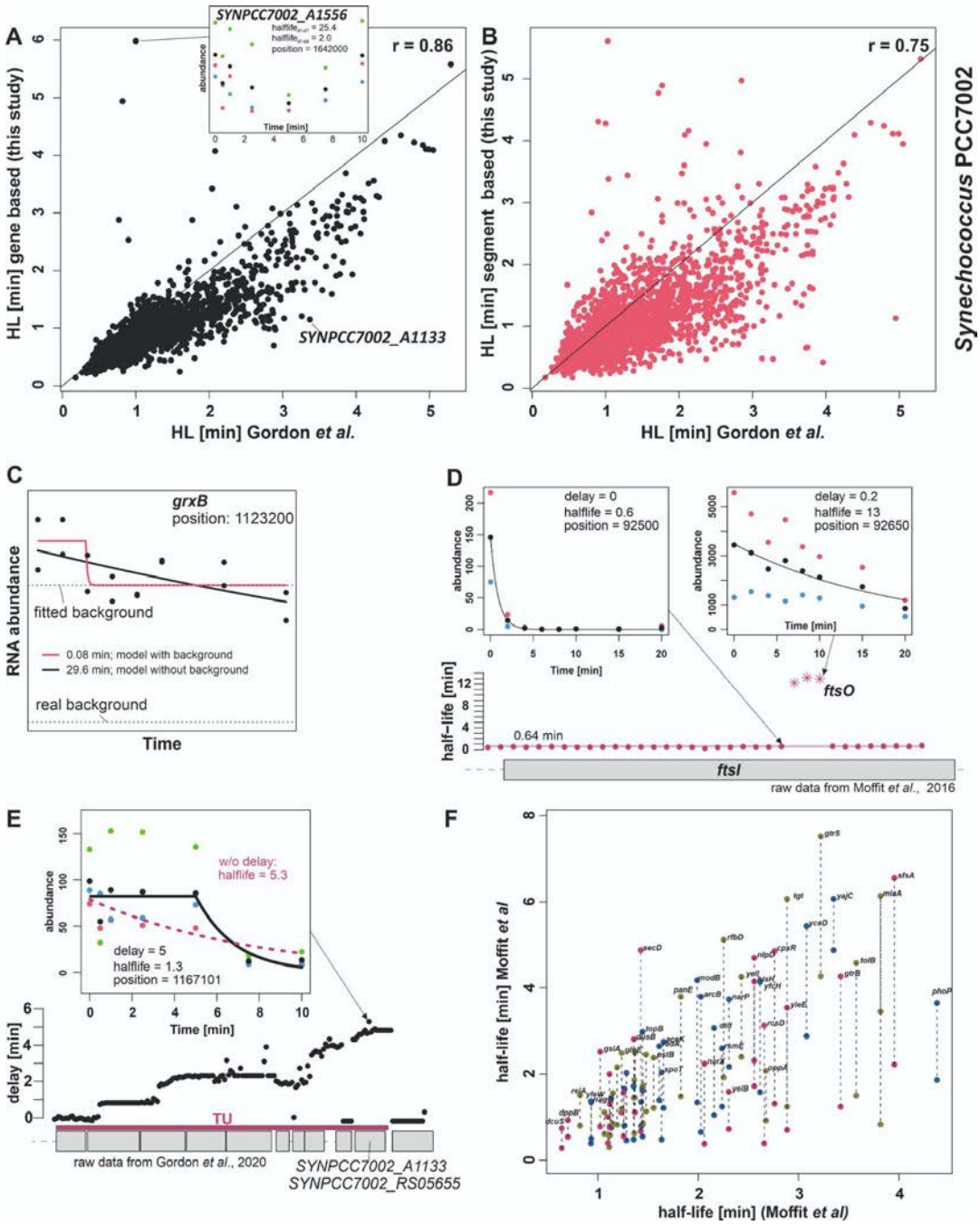
Supplementary Fig. 6: Detectability of RST instances. Minimal termination site distance from the transcriptional start site to reach an RNA abundance increase of ≥ 7.5% dependent on termination rate, elongation rate and RNA stability.



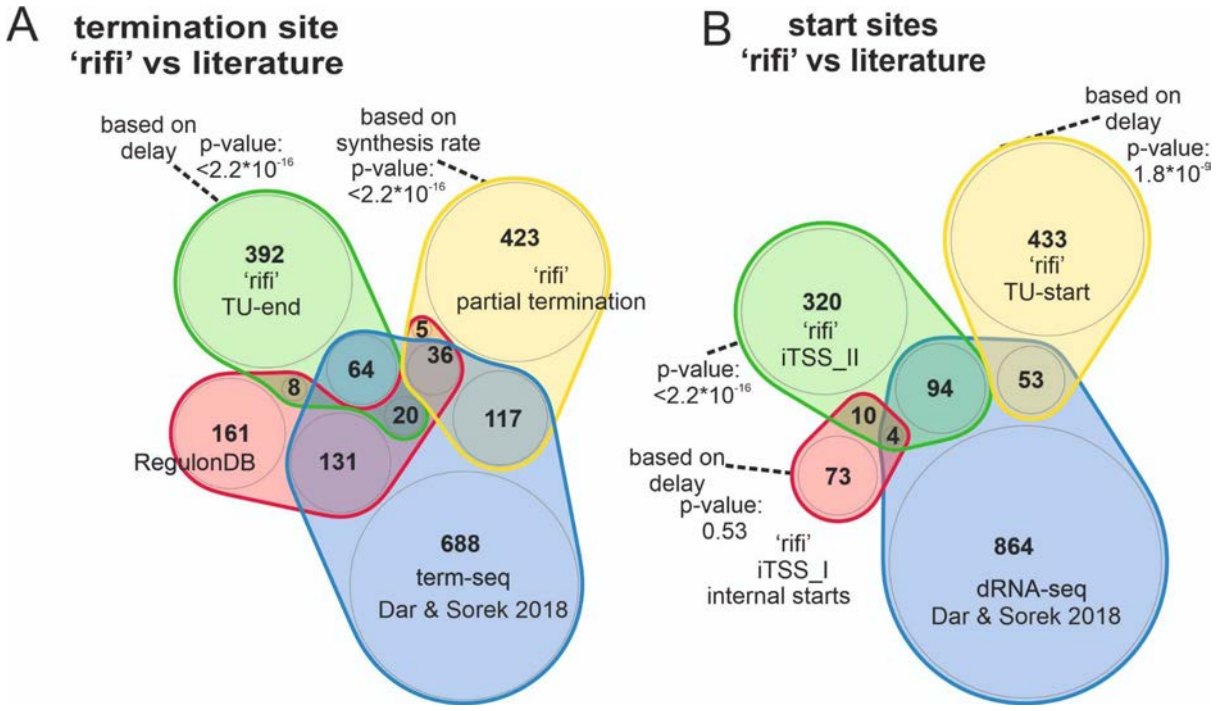
Supplementary Fig. 7: Example for a probable RST instance in *E.coli* MG1655, that is not detectable due to a relatively low time resolution. Upper part: *E. coli tol*-operon with an asRNA in the middle of *tolA*. Exactly at the position of the asRNA we detected 48% drop in the synthesis rate. However, we did not detect a post rifampicin RNA abundance increase as expected for RST. The time course for an example bin is given at the top left (the sampled timepoints 0, 2, 4, 6, 8,10, 15, 20 min are indicated by lines and dots). We calculated the expected time course for the RST case with the fitted information (top right). At the fitted high elongation rate and low half-life the post rifampicin RNA abundance increase would not be visible at the sampled time points. Note that the fitted half-life would be slightly too high (0.5 vs 0.3 min) if there would be a non detected RST in reality. Below, the same operon is displayed for *K. aerogenes*. We see an asRNA and a 39% termination at the same position in *tolA*. Here, a post rifampicin RNA abundance increase is detectable, because the 1 min time point is available.



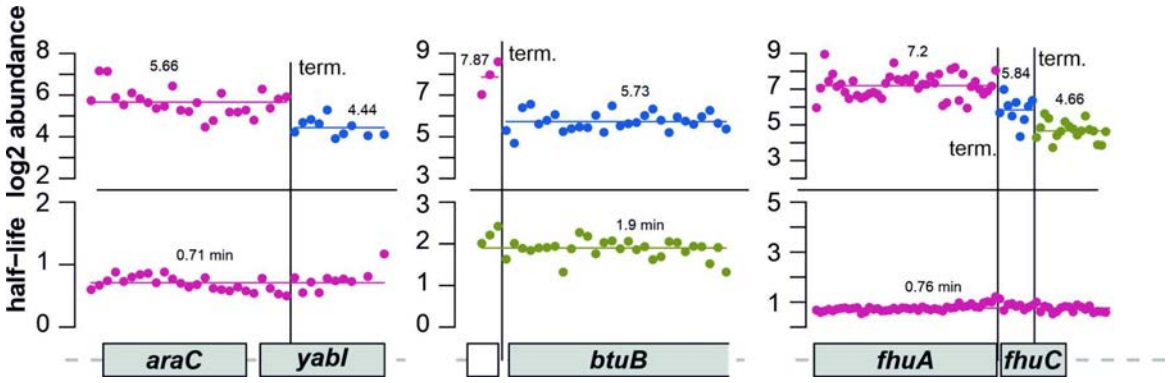
Supplementary Figure 8: Examples for RST from different organisms that are presumably caused by transcriptional interference due to *cis*-asRNA transcription. 'rifi' was used with high sensitivity parameters for RST detection. The log2 synthesis rate (dark pink dots) is displayed as a function of the genomic coordinates. The gray line indicates the mean log2 synthesis rate for a clustered segment. The light pink dots show the calculated synthesis rate without RST. The black dots in the lane above or below show the estimated termination rate in % as a function of the genomic coordinates. For selected pre and post RST bins the fitted decay curves and the raw data are displayed. The pink arrows show the antisense signals with the respective calculated synthesis rate beneath.



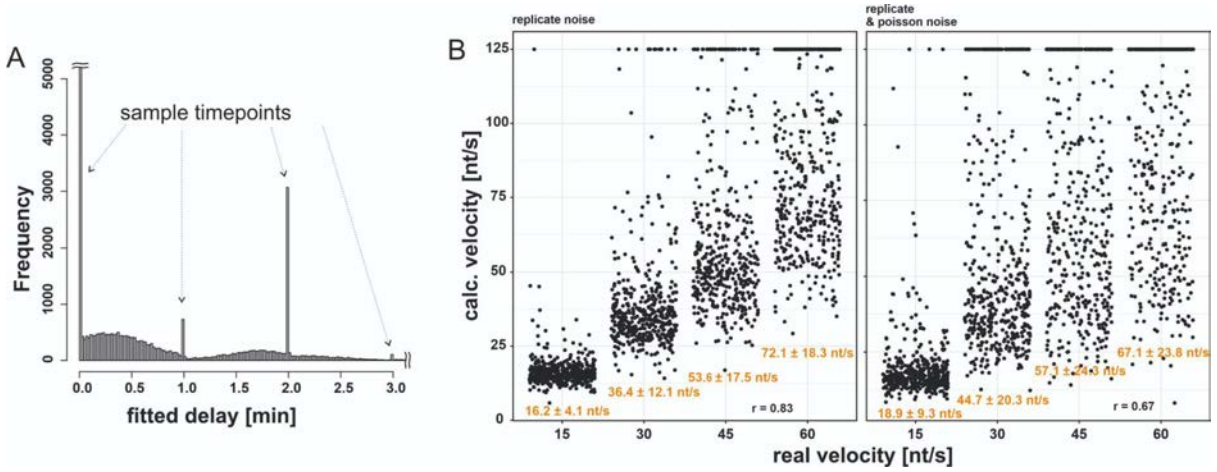
Supplementary Figure 9: Comparison of stability estimations with examples for deviations. A) Scatterplot comparing the original *Synechococcus* PCC 7002(12) data with our gene-based results and B) with the fragment-based results. A) Inlay: Example for difficult input data. After the rifampicin addition the RNA abundance initially declines, but after the 4th data point the abundance increases again which leads to a U-shaped pattern. Fitting all 7 data points results in a half-life of 25.4 min while a fit based on the first 4 data points results in a half-life of 2.0 min. In the original paper(12) later times were excluded if they increased the residuals of the fit. 'rifi' allows the use of custom filter functions to deal with that data. C) The effect for a fit with a model with and without the consideration of a general background for a bin within the *grxB* gene. D) Here the stability for the *ftsI* gene and the internal FtsO sRNA is shown. Plotted data are based on Moffit *et al*.(13). E) Example for a long TU in *Synechococcus* PCC 7002. Not considering the ~5 min delay for a bin in the *SYNPCC7002_A1133* transcript will lead to an artificial higher half-life of 5.3 min (broken pink line: no-delay fit; black line: delay adjusted fit). F) Genes covered by two or more transcripts of different half-lifes, for which the gene annotation based half-life calculation leads to wrong results.



Supplementary Figure 10: Comparison of rifi events with other experimental methods. rifi detects start sites and terminators based on changes in the synthesis rate (iTSS_II, (partial) termination) or changes in the delay (start_of_TU, iTSS_I and end_of_TU). The delay based estimates are discussed in a later section. A) Comparison of 'rifi' terminator predictions based on the synthesis rate (all events with p-value ≤ 0.05 & >15% termination rate) or based on the delay based TU-definition with term-seq(1) and RegulonDB terminators(4). The p-values for the overlap with the term-seq sites better than expected by chance were calculated with the exact binomial test. B) Comparison of 'rifi' transcriptional start site predictions based on the synthesis rate (iTSS_II,p-value ≤ 0.05), based on the delay based TU-definition (major starts) and delay based minor internal starts (iTSS_I, p-value ≤ 0.05) with a dRNAseq dataset (1). Start sites are estimated based on synthesis rate changes between two segments (iTSS_II), too. Only 14 of the delay based and synthesis rate TSS predictions overlap. That is mostly due to the lower sensitivity within the delay based method. If the elongation rate is fast, the time resolution of the data is not good enough and there is often no detectable delay change at a start. Nevertheless, an increased synthesis rate still indicates a start (example in Supplementary Fig. 2). In addition to the synthesis rate based terminator estimate, each end of an TU also resembles a predicted termination site. We compared the "end of TU" terminators with the term-seq data and the RegulonDB dataset as for the synthesis rate based terminator estimates (Supplementary Fig. 4A). Here, 84/484 sites overlap with term-seq terminators ($p < 2.2 * 10\text{-}16$) and 28/484 with the RegulonDB set ($p = 7.6 * 10\text{-}9$).

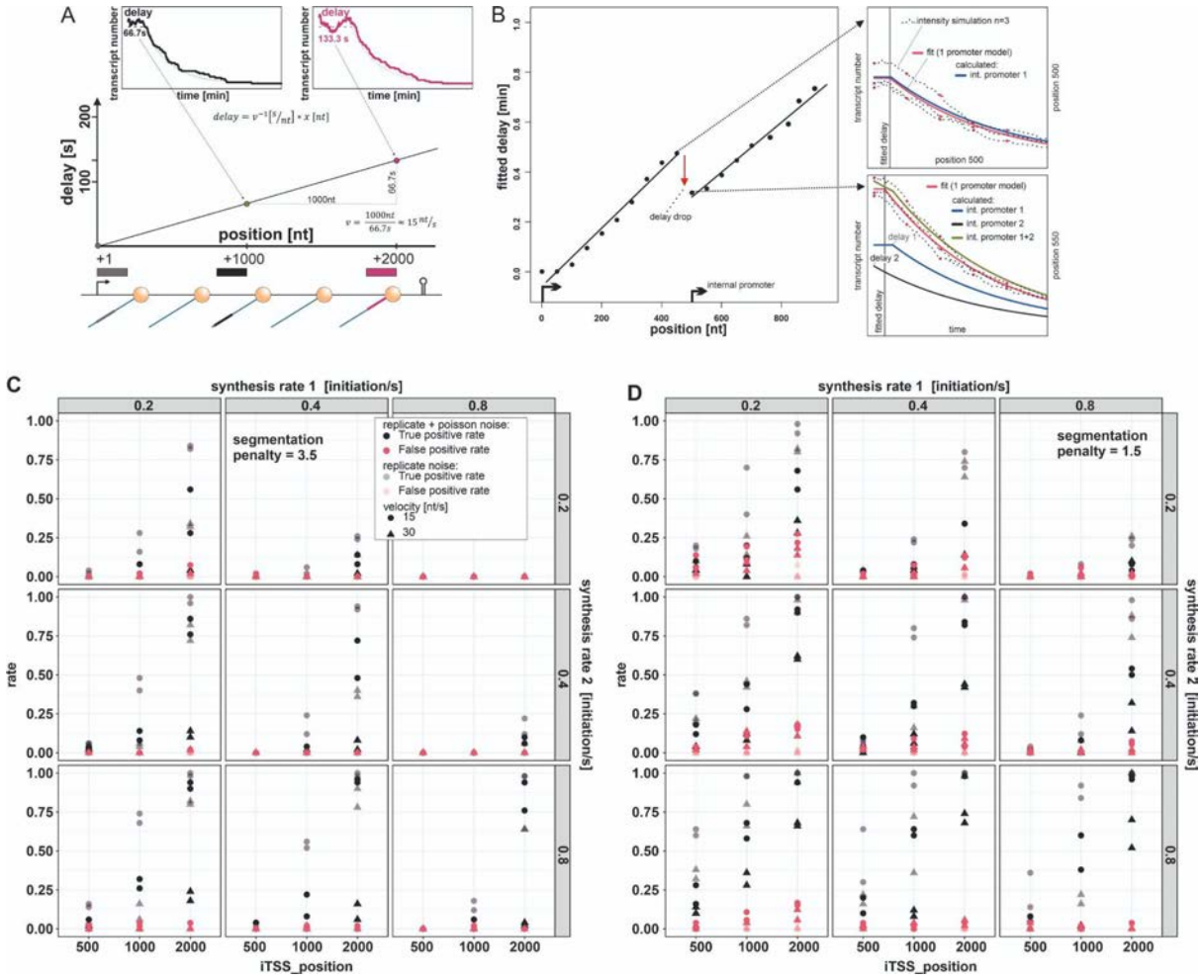


Supplementary Figure 11: Examples for high confidence terminators detected by rifi but not by term-seq (11). The termination sites are indicated by vertical lines.



Supplementary Figure 12: A) Histogram of the fitted delays for the *E. coli* BW25113 (11) dataset. Sampling timepoints are over-represented. The x -and y-axes are cut and not all data points are represented. B) A simulation was done to assess the general accuracy of the elongation rate estimate. Data are based on the simulations with 36 parameter combinations (half-life: 1, 2 or 3.33 min; velocity: 15, 30, 45 or 60 nt/s; synthesis rate: 0.2, 0.4 or 0.8 molecules/s) and a continuous transcription without a parameter change. The jitter plots show the real input velocity used for the simulation versus the calculated velocity based on the fits. The Pearson correlation r-value and the mean fitted velocities with their standard deviation are given in the plots. At the top the number of fragments with a velocity higher than 120 nt/s and the total number of fitted segments are given. The fragments with a biologically unrealistic velocity higher than 120 nt/s were excluded from the mean, sd and correlation calculation

To test the ability to detect velocity changes during transcription we used a 1000nt transcript with a velocity change position at 500nt (for parameters see legend Fig. 5). A simulation example is given in Fig. 5C. The method works in general, but for the chosen transcript length and speed change position in the simulation, no segmentations were done with the segmentation penalty of 3.5 (Fig. 5C) indicating an underestimation of velocity change points in the real life E. coli dataset. It needs to be considered that the segmentation for longer segments >500nt will be better.

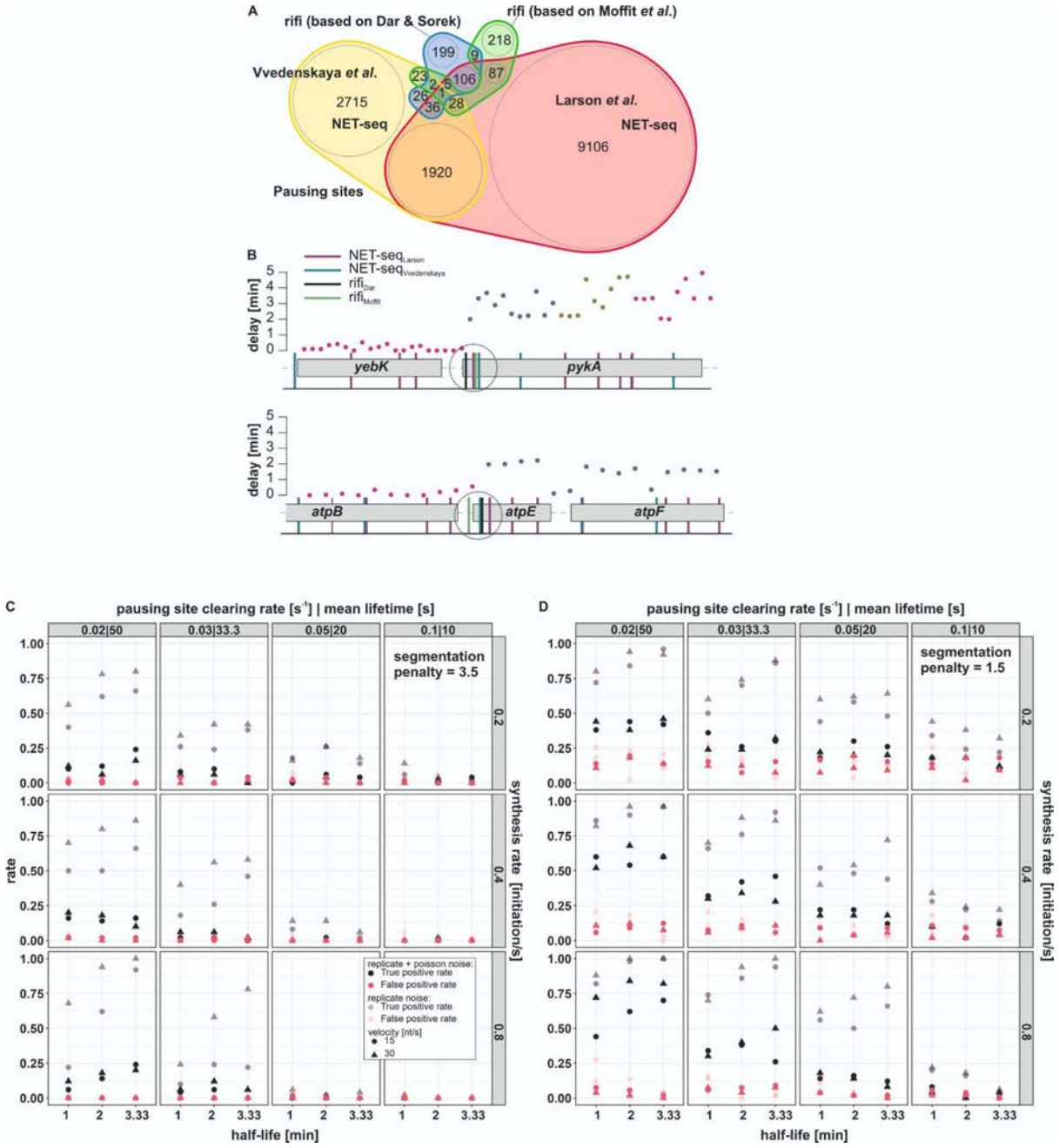


Supplementary Figure 13: A) Idealized example for a 2000 nt transcript with a constant transcription velocity over the whole transcript. The delay is linearly dependent on the distance from the transcriptional start site and the slope is the inverse of the velocity. The inlays show the decay curves of simulated data at position 1000 nt and 2000 nt with the fitted delay. B) Fitted delays based on stochastic simulation of an internal promoter. The inlays show the simulation data (dotted black lines, n=3) and data points taken for the fit (pink dots), the delayed exponential decay based on the fit with the one promoter model (pink line) and the analytical curves for the transcript starting at promoter 1 (blue), promoter 2 (black) and combined for both transcripts (green). The upper right panel shows position 500 directly before the second promoter and the lower right panel position 550, 50 nucleotides upstream the second promoter. Velocity and decay constants for both transcripts were 15 nt/s and 0.3 1/s. The synthesis rate at promoter one and two were 0.4 molecules/s and 0.3 molecules/s, respectively. C+D) In order to estimate the accuracy and sensitivity of the iTSS detection we used simulated data with different initial and secondary synthesis rates, decay constants and synthesis velocities. For each of the 108 parameter combinations (half-life: 1, 2; velocity: 15 or 30 nt/s; synthesis rate 1 +2: 0.2, 0.4 or 0.8 molecules/s; iTSS position: 500, 1000 or 2000nt) we run 50 independent simulations. The true positive rate (black symbols with and gray symbols without technical noise) is calculated based on the correctly identified pausing sites. C) A segmentation penalty of 3.5 and an outlier penalty of 2.5 was used. D) A segmentation penalty of 1.5 and an outlier penalty of 1.1 was used.

The complicated case of internal start sites and overlapping transcription

Many attempts to map the whole transcriptome of bacterial organisms have revealed that transcription does not exclusively start at the 5’ end of operons or single gene transcripts. Actually, there are plenty of internal transcriptional start sites within operons and genes (14–16). This leads to overlapping transcription processes and transcripts initiated from different starting sites, which violates the simple delayed-co-transcriptional decay model. Fitting those cases with the simple model can lead to an “artificial” delay (Supplementary Fig 13B), which lies between the delay of the two independent transcription processes. In theory, the model could be adapted to two or more overlapping transcripts (see Methods). However, as displayed in Supplementary Fig 13B (lower right panel) the fit with the wrong simple model and the solution for the two-promoter model are very close. Together with the relatively coarse sampling points, the subtle differences in the delay will not result in major differences in the residuals. Furthermore, the two-promoter model introduces up to three additional free parameters (a second synthesis rate, delay, and decay constant each). In reality a decision between models for unknown data would require accounting for these additional parameters using e.g. the Bayesian information criterion (17). Together with the inherent noise in the data it appears unlikely that the individual parameters of two or even more overlapping transcripts are detectable. In reality, we only use the one-promoter model and a combination of the real individual transcript parameters is fitted. The simple model can not estimate the strength (synthesis rate) of an internal start site, but the drop in the delay between pre and post internal starting site positions is in principle detectable. The performance to detect internal transcriptional start sites (iTSS) was estimated based on simulated

data with 108 parameter combinations. The subsequent delay fragment detection was done with and without the addition of poisson noise to account for technical sequencing noise. The sensitivity and specificity is dependent on the stringency of the dynamical programming segmentation and hence the segmentation penalty. The ROC-like curves in (Fig. 5A) shows that the method works in general. Every aspect that increases the duration of the delay drop, i.e. a low transcription velocity, a higher distance from the previous start and a relatively higher second synthesis rate, improves the detection of a new start (Supplementary Fig. 13CD). We tried to distinguish between weaker internal starts that do not start a new transcriptional unit (TU) and stronger starts that do trigger a new TU (compare Fig. 5D) via our segmentation strategy. The ‘rifi’ internal start site detection (iTSS_I) based on the delay has only an overlap of 4 sites with the dRNA-seq starts (1), which is not better than random ($p = 0.53$). In contrast, the stronger delay drops used for the delay based TU-start have a significant overlap with the dRNA-seq data ($p = 1.8*10^{-9}$) (Supplementary Fig. 10B).



Supplementary Figure 14: A) Comparison of *E. coli* pausing sites detected by rifi and by NET-seq. Due to the high number of closely spaced sites in the Larson *et al*. dataset, sites within the range of 50 nt were clustered prior to the comparison. B) Examples for pausing sites detected in all 4 datasets. The fitted delay (Dar & Sorek data) is given as a function of the genome position. Delay data points within a clustered delay fragment share the same color. The estimated pausing sites are given as upright lines (purple: NET-seq$_{Larson}$, blue: NET-seq$_{VVedenskaya}$, black: rifi$_{Dar}$, green: rifi$_{Moffit}$). The common pausing site is indicated by a broken circle. C+D) In order to estimate the accuracy and sensitivity of the pausing site detection we used simulated data with different pausing site off rates, synthesis rates, decay constants and synthesis velocities. For each of the 72 parameter combinations (half-life: 1, 2 or 3.33 min; velocity: 15 or 30 nt/s; synthesis rate: 0.2, 0.4 or 0.8 molecules/s; pausing site off rate: 0.02, 0.03, 0.05 or 0.1 1/s) we run 50 independent simulations. The true positive rate (black symbols with and gray symbols without technical noise) is calculated based on the correctly identified pausing sites. C) A segmentation penalty of 3.5 and an outlier penalty of 2.5 was used. D) A segmentation penalty of 1.5 and an outlier penalty of 1 was used.

Pausing sites:

Native-elongating-transcript sequencing (NET-seq), a method to globally detect transcriptional pausing sites, showed that transcriptional pausing is a frequent phenomenon(2, 3). Using rifampicin data, pauses are theoretically detectable as jumps in the delay (Fig. 5B), if at least some polymerases continue with the transcription after the pausing. However, if the pausing leads to an efficient transcriptional termination, as in the well described attenuator pausing sites in e.g. the *trp* operon leader in *E. coli*(18), the pausing signal will not be visible in rifampicin data. The pausing duration is usually in the range of <10 seconds, but longer than 25 seconds are also reported(19). Compared to these timescales the timely resolution of typical rifampicin data, in the range of minutes, is rather coarse. In order to assess the method, we compared the sites of two independent *E. coli* rifampicin datasets(1, 13) and two NET-seq datasets(2, 3). The pausing site overlap for the two rifampicin datasets is, with 17, rather small, but significantly better than random (p = 2.6e-8). For the NET-seq data we used the Larson dataset as reference because of the highest number of sites reported. Using an overlap window of +/- 1bin (i.e. +/- 50nt) the probability to randomly hit a NET-seq$_{Larson}$ site is ~24%. The actual overlaps with all other datasets are significantly higher than randomly expected (Supplementary Fig. 9a). Roughly ~39% of the rifi-sites$_{Dar}$ (p = 6.7e-10) and ~32% of the rifi-sites$_{Moffit}$ (p = 2.2e-4) overlap with sites in NET-seq$_{Larson}$, which is not far away from the ~42% overlap of NET-seq$_{Vvedenskaya}$ with NET-seq$_{Larson}$ (p < 2.2e-16). Examples for sites detected in all datasets are atpE and pykA (Supplementary Fig. 9B). In general, even the two NET-seq studies have substantial site variations >50%. Despite this relatively low overlap, both NET-seq methods revealed a similar pausing site consensus motif $G_{-10}Y_{-1}G_{+1}$ (2, 3). Together, this indicates that pausing sites might be very variable

and condition dependent. There are two major differences between the *E. coli* rifampicin datasets used in this study, the growth temperatures of 32°C (13) vs 37°C (1) and the two different *E. coli* strains. Unfortunately the resolution of the bin-based rifi data is not high enough to do a meaningful motif based on the detected sites. The known motif is very common and exits ~ $1.2 * 10^5$ times in the forward strand of the *E. coli* genome. 84% of the $rifi_{Dar}$ sites contain the motif, considering a window of +/- 25nt around an identified pausing. However, from 10^6 randomly sampled 50nt stretches only 75% contained the motif, which makes the 84% motif appearance in $rifi_{Dar}$ significantly more frequent than random ($p = 2.7 * 10^{-5}$). In conclusion, it is hard to verify the soundness of the rifi pausing site detection based on other experimental data. Thus, we additionally used simulated transcription data with known pausing sites and 72 parameter combinations including 4 different mean pausing lifetimes (10, 20, 33.3 and 50s). The ROC-like curve in Fig. 5B shows the effect of the penalties and that the pausing site detection by 'rifi' works in general. However, the quality depends on the noise level and the timely resolution of the data. A simulation with more sampling time points resulted in a strongly improved accuracy and sensitivity (Fig. 5B). Furthermore, a high synthesis rate and a longer mean lifetime of the pause improves pausing site detection (Supplementary Fig. 9CD). Two independent rifampicin datasets from *E. coli* BW25113 (1) and *E. coli* MG1655 (13) share only 17 pausing sites. The overlap of the rifi results with the results of the established NET-seq method is still relatively low, but also significantly higher than random. Interestingly, the overlap of two independent NET-seq datasets is below 50%, too, which suggests a general high pausing site variability. Simulations suggest that the rifi pausing site detection method is actually working. In any case rifi will rather detect pausing with long lifetimes such as backtracked- or hairpin- stabilized

pausing sites instead of elemental pausing sites (20) and an experimental verification of individual rifi pausing sites would be of interest.

References:

1. Dar,D. and Sorek,R. (2018) Extensive reshaping of bacterial operons by programmed mRNA decay. *PLOS Genet.*, **14**, e1007354.
2. 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.
3. 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.
4. Santos-Zavaleta,A., Salgado,H., Gama-Castro,S., Sánchez-Pérez,M., Gómez-Romero,L., Ledezma-Tejeida,D., García-Sotelo,J.S., Alquicira-Hernández,K., Muñiz-Rascado,L.J., Peña-Loredo,P., *et al.* (2019) RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. *Nucleic Acids Res.*, **47**, D212–D220.
5. Pérez-Silva,J.G., Araujo-Voces,M. and Quesada,V. (2018) nVenn: generalized, quasi-proportional Venn and Euler diagrams. *Bioinformatics*, **34**, 2322–2324.
6. Thomason,M.K., Bischler,T., Eisenbart,S.K., Förstner,K.U., Zhang,A., Herbig,A., Nieselt,K., Sharma,C.M. and Storz,G. (2015) Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J. Bacteriol.*, **197**, 18–28.
7. Van Gundy,T., Martin,E., Bono,J., Hatton,O. and Lybecker,M.C. (2022) An Antisense RNA Fine-Tunes Gene Expression of the Type II MazEF Toxin-Antitoxin System. *mBio*, **13**, e03443-21.
8. Wang,X., Ji,S.C., Jeon,H.J., Lee,Y. and Lim,H.M. (2015) Two-level inhibition of galK expression by Spot 42: Degradation of mRNA mK2 and enhanced transcription termination before the galK gene. *Proc. Natl. Acad. Sci.*, **112**, 7581–7586.
9. Balasubramanian,D., Ragunathan,P.T., Fei,J. and Vanderpool,C.K. (2016) A Prophage-Encoded Small RNA Controls Metabolism and Cell Division in Escherichia coli. *mSystems*, **1**, e00021-15.
10. Melamed,S., Adams,P.P., Zhang,A., Zhang,H. and Storz,G. (2020) RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles. *Mol. Cell*, **77**, 411-425.e7.
11. Dar,D. and Sorek,R. (2018) Extensive reshaping of bacterial operons by programmed mRNA decay. *PLOS Genet.*, **14**, e1007354.
12. Gordon,G.C., Cameron,J.C., Gupta,S.T.P., Engstrom,M.D., Reed,J.L. and Pfleger,B.F. (2020) Genome-Wide Analysis of RNA Decay in the Cyanobacterium Synechococcus sp. Strain PCC 7002. *mSystems*, **5**.
13. 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**, e13065.
14. Mitschke,J., Georg,J., Scholz,I., Sharma,C.M., Dienst,D., Bantscheff,J., Voß,B., Steglich,C., Wilde,A., Vogel,J., *et al.* (2011) An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci.*, **108**, 2124–9.
15. Sharma,C.M., Hoffmann,S., Darfeuille,F., Reignier,J., Findeisz,S., Sittka,A., Chabas,S., Reiche,K., Hackermuller,J., Reinhardt,R., *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, **464**, 250–255.
16. Thomason,M.K., Bischler,T., Eisenbart,S.K., Förstner,K.U., Zhang,A., Herbig,A., Nieselt,K., Sharma,C.M. and Storz,G. (2014) Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in Escherichia coli. *J. Bacteriol.*, 10.1128/JB.02096-14.
17. Schwarz,G. (1978) Estimating the Dimension of a Model. *Ann. Stat.*, **6**, 461–464.
18. Turnbough,C.L. (2019) Regulation of Bacterial Gene Expression by Transcription Attenuation. *Microbiol. Mol. Biol. Rev. MMBR*, **83**, e00019-19.
19. Neuman,K.C., Abbondanzieri,E.A., Landick,R., Gelles,J. and Block,S.M. (2003) Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking. *Cell*, **115**, 437–447.
20. Kang,J.Y., Mishanina,T.V., Landick,R. and Darst,S.A. (2019) Mechanisms of Transcriptional Pausing in Bacteria. *J. Mol. Biol.*, **431**, 4007–4029.