Improved RNA stability estimation through Bayesian modeling reveals most bacterial transcripts have sub-minute half-lives

Laura Jenniches[1], Charlotte Michaux[2], Sarah Reichardt[1], Jörg Vogel[1,2,3], Alexander J. Westermann[1,2], Lars Barquist[1,3,*]

[1] Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz Centre for Infection Research (HZI), Würzburg, Germany
[2] University of Würzburg, Institute of Molecular Infection Biology (IMIB), Würzburg, Germany
[3] University of Würzburg, Faculty of Medicine, Würzburg, Germany

*Correspondence to LB: lars.barquist@helmholtz-hiri.de

Abstract (250/250 words)

RNA decay is a crucial mechanism for regulating gene expression in response to environmental stresses. In bacteria, RNA-binding proteins (RBPs) are known to be involved in post-transcriptional regulation, but their global impact on RNA half-lives has not been extensively studied. To shed light on the role of the major RBPs ProQ and CspC/E in maintaining RNA stability, we performed RNA sequencing of *Salmonella enterica* over a time course following treatment with the transcription initiation inhibitor rifampicin (RIF-seq) in the presence and absence of these RBPs. We develop a hierarchical Bayesian model that corrects for confounding factors in rifampicin RNA stability assays and enables us to identify differentially decaying transcripts transcriptome-wide. Our analysis revealed that the median RNA half-life in *Salmonella* in early stationary phase is less than 1 minute, a third of previous estimates. We found that over half of the 500 most long-lived transcripts are bound by at least one major RBP, suggesting a general role for RBPs in shaping the transcriptome. Integrating differential stability estimates with CLIP-seq revealed that approximately 30% of transcripts with ProQ binding sites and more than 40% with CspC/E binding sites in coding or 3' untranslated regions decay differentially in the absence of the respective RBP. Analysis of differentially destabilized transcripts identified a role for both proteins in the control of respiration, and for ProQ in the oxidative stress response. Our findings provide new insights into post-transcriptional regulation by ProQ and CspC/E, and the importance of RBPs in regulating gene expression.

Significance Statement (117/120 words)

Together with transcription and translation, RNA decay is one of the major processes governing protein production. Here, we have developed a new statistical approach that corrects for confounding effects when estimating RNA decay rates from RNA-seq in bacteria. Our more accurate decay rate estimates indicate that bacterial transcripts have half-lives about three times shorter than previously thought. This approach allowed us to measure the effects of RNA-binding proteins (RBPs) on decay rates, identifying large cohorts of transcripts with changes in stability following RBP deletion and conditions where post-transcriptional regulation affects survival. Our method should lead to a reevaluation of RNA stability estimates across diverse bacteria and new insights into the role of RBPs in shaping the transcriptome.

Introduction

Rapid adaptation of the proteome to environmental conditions is essential for the survival of microorganisms. RNA degradation is an important post-transcriptional process directly influencing protein abundance. The lifetime of bacterial RNA ranges from seconds to an hour (1) and depends on numerous factors, including transcript identity, genotype and growth condition (2). RNA-binding proteins (RBPs) in bacteria include structural components of the ribosome and global post-transcriptional regulators such as Hfq (3, 4) and CsrA (5) which play key roles in modulating translation and RNA stability in concert with a network of small RNAs (sRNAs) (6, 7). Beyond these model RBPs, recent years have seen the discovery of a menagerie of bacterial RBPs that bind hundreds or even thousands of transcripts (8–10), though their functions in shaping the transcriptome remain unclear.

In *Salmonella enterica* serovar Typhimurium (henceforth *Salmonella)*, these global RBPs include the FinO-domain containing protein ProQ and the cold-shock proteins CspC and CspE. ProQ has been shown to bind hundreds of mRNAs and sRNAs (11–13), affecting important biological processes including expression of virulence factors (14) and formation of antibiotic persisters (15). CspC and CspE have been shown to play partially redundant roles in virulence, affecting survival in mice, motility, biofilm formation, and survival of bile stress (16, 17). The molecular details of how these RBPs affect phenotype are not clear, although at least some of the effects of ProQ and CspC/E are mediated by the direct modulation of mRNA stability. For instance, CspC/E have been shown to stabilize the mRNA of the bacteriolytic lipoprotein EcnB by blocking digestion by the endonuclease RNase E (16). ProQ on the other hand appears to preferentially bind 3′ UTRs where in a few cases it has been shown to protect transcripts from exonuclease activity (12, 18).

While these results provide hints at the mechanisms by which RBPs regulate target gene expression, in the absence of transcriptome-wide differential RNA stability measurements it remains unclear how common regulation through stability modulation is. A classical approach to study RNA stability is to halt transcription with the transcription initiation inhibitor rifampicin (19) and monitor RNA decay over time. This approach has been scaled to the whole transcriptome by combining it with microarrays (20, 21) and high-throughput sequencing (22). However, the presence of non-linear effects in the resulting time-course data makes inference of differences in decay rates between experimental conditions difficult.

RNA-seq analysis tools such as limma (23), edgeR (24), and DEseq (25) solve the problem of accurately estimating dispersion in experiments with many measurements but few

replicates through an empirical Bayes approach (26). In empirical Bayes, information is pooled across transcripts under the assumption that transcripts with similar concentrations will exhibit similar biological and technical variation across samples, leading to more robust dispersion estimates. However, these tools are currently limited to linear models. Recent progress in the optimization of sampling methods has made the development of fully Bayesian hierarchical models increasingly efficient and accessible. In particular, the Stan probabilistic programming language (27) separates model description from sampler implementation, allowing easy development and testing of complex hierarchical models. This provides a powerful framework for developing analysis methods for sequencing data that can accommodate complex experimental techniques.

Here, we investigate the effects of ProQ and CspC/E on RNA stability across the entire transcriptome, starting from a fully Bayesian analysis of rifampicin treatment followed by RNA sequencing (RIF-seq). During model development, we discovered that accounting for confounding factors in stability assays conducted after rifampicin treatment dramatically affects the inferred half-life, leading us to substantially revise estimates for average mRNA half-life in *Salmonella* to less than 1 minute, compared to previous estimates in the range of 2 to 7 minutes in the closely related species *E. coli* (20–22). We develop a hypothesis testing procedure for determining differential decay rates that allows us to identify hundreds of gene transcripts destabilized in the absence of ProQ and CspC/E. We combine our differential stability estimates with other high-throughput datasets available for *Salmonella* to further characterize RBP interactions, identifying a role for ProQ in survival of oxidative stress and for CspC/E in the control of respiration. We additionally find a substantial population of long-lived transcripts that depend on RBPs for their stability, illustrating the importance of RBPs in shaping the bacterial transcriptome. Beyond its utility in investigating RBP interactions, our improved approach to determining transcript half-life suggests that RNA stability in bacteria has generally been overestimated and will need to be reassessed in other bacterial species.

Results

A progressive Bayesian analysis revises RNA half-lives

To determine transcriptome-wide half-lives under an infection-relevant condition, we applied RIF-seq to *Salmonella* at early stationary phase (ESP) where host invasion genes are expressed (28). Our RIF-seq workflow for data production and analysis is illustrated in **Figure 1A**: wild-type and isogenic RBP deletion strains were treated with rifampicin, and cellular RNA

samples were collected over time to capture RNA decay dynamics. We collected data from eight time points following rifampicin treatment in three (Δ*cspC/E*), six (Δ*proQ*), or nine (wild-type) replicates (see Methods). We included ERCC RNA spike-ins (29) for normalization between samples. Additionally, we developed a center-mean normalization technique to remove batch effects between replicate samples (**Figure S2;** *Methods*). Subsequently, we fitted a Bayesian statistical model to the normalized data using Hamiltonian Monte Carlo with Stan (27).

We employed a progressive Bayesian workflow to arrive at our final model (**Figure 1B**). An advantage of Bayesian analysis is that it allows the modeler to formalize their beliefs about the data generating process and provides a variety of tools for model comparison and selection. In the case of RIF-seq data, the simplest expectation would be that RNA concentrations would exhibit a linear decay on a semilog scale, which could be fit by a simple linear model with gene- and condition-dependent decay rate β. While some of our observations met this expectation (**Figure S3A**), the vast majority of transcripts exhibited more complex dynamics that prevent accurate extraction of decay rates with a linear model (**Figure 1C; S3A**), leading to large unexplained variation at late time points (**Figure S3K**). To account for this, we introduced additional parameters that capture confounding effects in the data. The first confounding effect is a gene-dependent delay parameter γ, which captures the delay commonly observed in RIF-seq data before decay initiates (**Figure 1B&C**, in green). As has been previously described, this is due to ongoing transcription from RNA polymerase already bound to DNA, which rifampicin does not block (22, 30). The ongoing transcription compensates for decay, manifesting as a delayed decay. To support the relationship between ongoing transcription and the delay parameter, we performed an analysis of elongation times on 60 base sub-genic windows, finding a clear association between the estimated elongation time and distance to annotated transcription start sites (**Figure S3B**). We used this association to infer transcription rates from our data set (see **Figure S3C**, Methods) finding a median transcription rate of 22.2 nt/s (**Figure S3D**), comparable to previous estimates in *E. coli* (22).

The second confounding effect we corrected for was an apparent gene- and condition-dependent baseline RNA concentration π beyond which no further decay was observed (**Figure 1B&C**, orange). We were initially concerned that this effect may be an artifact of the pseudocount we used to avoid dividing by zero in our calculations; however, inspection of a number of decay curves illustrated that the observed baseline was generally well above the detection threshold (**Figure S3E**, see Methods). We also verified that the half-life of a transcript is generally constant along an operon (**Figure S3G**). In agreement with previous work (31) we found a small number of stable subregions which generally corresponded to known sRNAs, but

since this was not a general feature of transcripts we excluded this as a source of the observed baseline. To confirm that the baseline is not a result of our sequencing protocol, we used independent northern blot quantifications from a rifampicin treatment time course including late time points from a previous study (11). These quantifications reproduced the observed baseline effect (**Figure 1D**), illustrating that this is a general feature of rifampicin RNA stability assays. For wild-type *Salmonella*, we find that a median of 2.6% of the initial transcript concentration appears resistant to decay (**Figure S3H**), and that the exponential decay regime ends at different timepoints for different transcripts and genotypes (**Figure S6AB&C**). Whether this fraction is truly resistant to degradation or just degrades much slower than the rest of the transcript population is unclear. However, the median fraction of baseline RNA increases to 5.7% when *proQ* is overexpressed (**Figure S3H**), suggesting that nonspecific RBP-RNA interactions may play a role in degradation resistance.

To compare the models with and without these two confounding factors, we calculated the difference in the expected log pointwise predictive density (ELPD), a measure of the expected predictive accuracy of a model on out-of-sample data, using Pareto-smoothed importance sampling approximate leave-one-out cross-validation (PSIS-LOO, see Methods) (32). Comparing the difference in ELPD between a simple linear model, the piecewise linear model correcting only for extension time, and the full model (henceforth *log-normal* model) including the baseline correction showed a clear preference for the log-normal model, particularly at late timepoints (**Figure S3I&J**). Additionally, examination of the fitted variance unexplained by our decay model, σ_g, illustrated the log-normal model captured the behavior of late timepoints better than the piecewise model (**Figure S3K-M**). Correcting for confounding factors has major implications for transcriptome-wide estimates of decay rates: while the linear and piece-wise linear models produced median half-life estimates of 2.7 and 2.5 minutes, respectively, our final log-normal model estimates a median half-life of 0.9 minutes (**Figure 1E**).

To investigate whether transcripts encoding proteins involved in different cellular functions systematically differ in their stability, we calculated average half-lives across clusters of orthologous groups (COG) categories (33) (**Figure 1F; S5**). In agreement with previous work (20, 21), transcripts for genes involved in energy production and carbohydrate metabolism tended to be longer lived. We also found many sRNAs to have longer than average half-lives. Among the least stable transcripts were those coding for genes involved in cell division (e.g. *ftsZ*) and DNA replication (e.g. *dnaA*, *dnaN*), suggesting tight control of their cognate proteins. Taken together, accurately modeling RNA-decay curves led to drastically reduced

transcriptome-wide half-life estimates and allowed us to relate transcript stability to gene function.

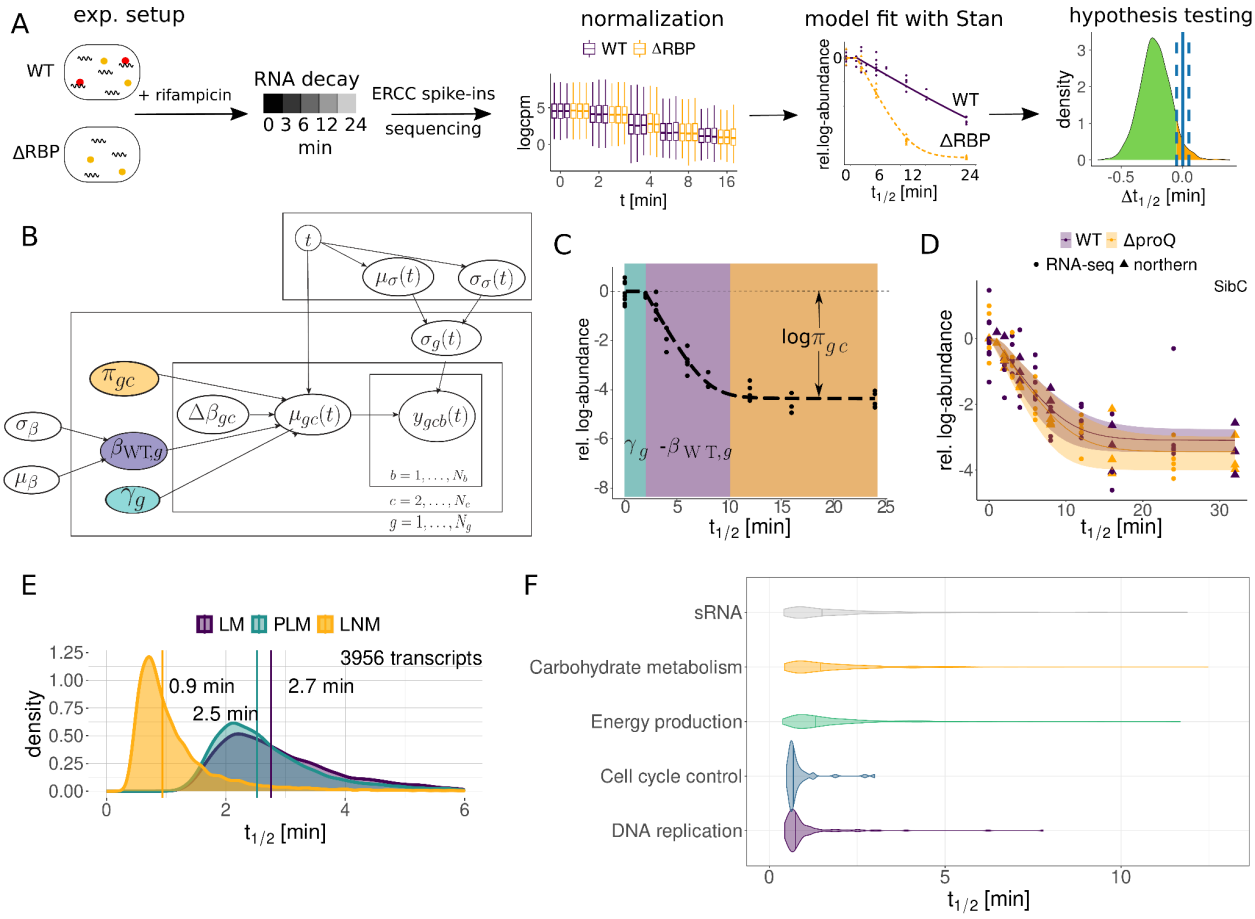


Figure 1. Pipeline and model description

(A) RIF-seq workflow: WT and ΔRBP strains are treated with rifampicin, cells are harvested at various time points and subjected to RNA-seq. Read counts are normalized before the extraction of biologically relevant parameters with a Bayesian model. Significant differences between strains are identified with Bayesian p values. (B) A plate diagram of the Bayesian models in this study. The layers indicate which indices and variables the parameters depend on. The LM is parametrized by the WT decay rate $\beta_{WT,g}$ (purple). In the PLM, the gene-wise elongation time γ_g (green) is added. The LNM adds a baseline parameter π_{gc} (orange) which corresponds to the fraction of residual RNA ($\pi_{gc} \in [0,0.2]$). The WT decay rate β_{WT} is a gene-wise parameter that is modeled hierarchically and depends on the hyperparameters μ_β and σ_β. The difference in decay rate $\Delta\beta_{gc}$ depends on the strain or condition c. The scale parameter $\sigma_g(t)$ captures variation by our decay model and depends on the time-dependent hyperparameters $\mu_\sigma(t)$ and $\sigma_\sigma(t)$. (C) Representative example of a decay curve in the LNM, illustrating regimes dominated by the different model parameters. The period of transcription elongation γ is marked in green, the exponential decay with decay rate β in purple and the constant regime governed by the fraction of baseline RNA π in orange. (D) Comparison of RNA-seq and model fit with independent northern blot quantifications for SibC (11). (E) Hyperpriors and median of transcriptome-wide WT half-lives in the three

Bayesian models. (F) Half-life distributions from the log-normal model for transcripts in selected COG categories.

Steady-state abundance does not reflect changes in transcript half-life upon RBP deletion

To study the influence of the RBPs ProQ and CspC/E on transcript stability, we applied the log-normal model to our RIF-seq data for the *proQ* and *cspC/E* deletion strains, as well as a ProQ overexpression strain (ProQ++; see Methods). To prioritize transcripts with changes in stability, we developed a hypothesis testing procedure based on examination of the posterior distribution of the change in decay rate from the wild-type (**Figure S4A**) and estimated statistical significance by calculating Bayesian p-values (**Figure S4B**). Since Bayesian p-value distributions require calibration (34, 35), we used simulation studies to estimate the false discovery rate (FDR) (**Figure S4C-E**, see Methods). To evaluate our differential stability estimates, we examined known targets of ProQ and CspC/E (**Figure 2A-C,S6A**). For deletion of ProQ we were able to confirm destabilization of the *cspD*, *cspE*, and *ompD* transcripts (**Figure 2A**), while the *cspC* transcript was hyperstabilized in the ProQ++ background (**Figure 2B**) in agreement with previous northern analysis (12). Similarly, we found the *ecnB* transcript destabilized following *cspC/E* deletion (**Figure 2C**) as previously reported (16).

For both RBP deletions, we identify hundreds of transcripts with changes in stability at an FDR of 0.1 (**Figure 2D&E**). Deletion of *cspC/E*, whose role in maintaining transcript stability is less well explored, led to strong destabilization of a large cohort of transcripts (727), while only stabilizing 164 (**Figure 2F**). Curiously, we identified more transcripts which were significantly stabilized (413) than destabilized (276) following *proQ* deletion (**Figure 2G**), which was unexpected as prior studies have focused on ProQ's stabilizing effect (11, 12). Nevertheless, stabilized transcripts tended to have much smaller changes in half-life, with a median change of 0.3 minutes (**Figure S6D**), compared to destabilized transcripts whose half-lives changed by 0.7 minutes on average.

A striking feature of our analysis of both strains was that changes in transcript half-life are not clearly related to changes in steady-state abundance upon RBP deletion (**Figure 2D-G**). In the *proQ* deletion strain, less than 10% of destabilized transcripts showed a statistically significant decrease in steady-state abundance. While this number was higher for the *cspC/E* deletion (~18%), it was still only a minor fraction of the total number of destabilized transcripts. This might be explained by altered activity of other regulatory proteins. Deletion of either RBP

led to perturbation of the stability of transcripts encoding major regulatory proteins including the anti-sigma factor Rsd, the transcription termination factors Rho and NusA, the alternative sigma factor RpoS, the nucleoid-associated HupA/B, and the cAMP receptor protein CRP (**Figure S6B**). For HupA/B and RpoS, we also observed reduced mRNA abundance in RBP deletion strains (**Figure S14**). Hence, loss of ProQ or CspC/E likely has complex, and in some cases indirect, effects on the global transcriptome. This suggests caution should be taken when deducing direct regulatory interactions from differential expression analysis of RBP-deletion mutants.

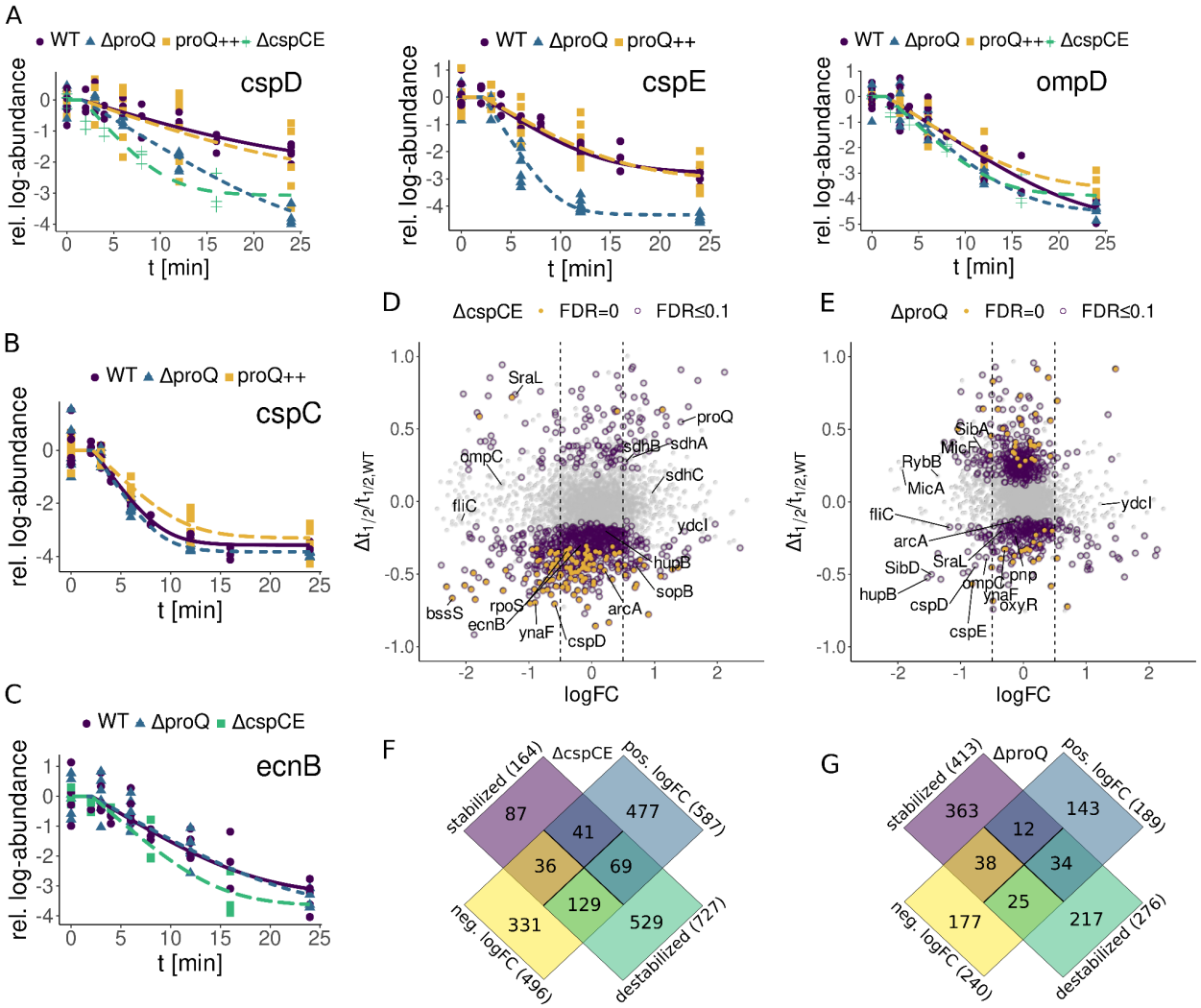


Figure 2. Differential analysis of transcript stability in the absence of CspC/CspE/ProQ
(A-C) Decay curves of known ProQ (*cspD*, *cspE*, *ompD*, *cspC*) or CspC/E (*ecnB*) targets. (D) Relative difference in half-life vs. steady state log-fold changes between the Δ*cspCE* and the WT strain. (E) Relative difference in half-life vs. steady state log-fold changes between the Δ*proQ* and the WT strain. (F) Overlap between stability changes and steady-state log-fold changes in the Δ*cspCE* strains. (G) Overlap between stability changes and steady-state log-fold changes in the Δ*proQ* strain.

Integrating high-throughput datasets identifies cohorts of mRNAs subject to known RBP regulatory mechanisms

The location of an RBP binding site within a transcript is often a key determinant of the mechanism of RBP regulation. To investigate the potential mechanisms underlying the stabilization activity of ProQ and CspC/E, we integrated our differential stability estimates with UV crosslinking and immunoprecipitation followed by RNA sequencing (CLIP-seq), which can localize RBP binding sites within a transcript. For ProQ, we reanalyzed an existing CLIP-seq dataset (12), identifying 833 peaks indicative of binding (see Methods). We produced new CLIP-seq datasets for both CspC and CspE and identified 1155 CspC and 861 CspE peaks, spread across 571 and 462 target transcripts, respectively (**Figure 3A&B;S7A-C**). In total, 717 transcripts are bound by at least one CSP, with 430 CspC peaks directly overlapping with a CspE peak (**Figure 3B**) supporting the previously reported partial redundancy between these proteins (16) and similar observations in *E. coli* (36). We saw especially dense clusters of CspC/E peaks in transcripts encoding genes involved in the TCA cycle, flagellar proteins, and proteins involved in host invasion associated with the Salmonella pathogenicity island 1 (SPI-1) type three secretion system (**Figure 3A**).

We next examined the distribution of RBP binding sites across target transcripts, beginning with ProQ. As previously reported (12), ProQ binds predominantly at the end of coding sequences, with half of detected binding sites within 100 nucleotides of the stop codon (**Figure 3D;S7D**). Amongst those genes with 3′ binding sites, we identified 86 that were significantly destabilized upon *proQ* deletion (**Table S3**). Besides the known interaction of ProQ with the the *cspE* mRNA, these include transcripts encoding the SPI-1 effectors SopD and SopE2, involved in host cell invasion (37), and OxyR, a transcription factor involved in the oxidative stress response (38). The location of these binding sites suggests that ProQ may protect the 3′ ends of a large cohort of transcript from exoribonucleases attack, as previously shown for individual model transcripts (12, 18).

In contrast to ProQ, CspC and CspE binding sites were spread across coding (CDS) regions with only slight enrichment in the vicinity of the start and stop codons (**Figure 3C;S7D**). We identified 177 transcripts with a CspC and/or CspE binding site in the CDS or 5'UTR that were destabilized upon *cspC/E* deletion (**Table S4**). These included the *ecnB* transcript (**Figure S10A**), which has previously been shown to bind CspC and CspE *in vitro* and to be protected from RNase E by CspC/E *in vivo* (16). To further investigate the role of the CSPs in protection

from RNase E cleavage, we combined our stability and CLIP-seq data with a published dataset mapping RNase E cleavage sites (39). We saw an enrichment of RNase E cleavage sites within CspC/E CLIP-seq peaks (410/2059 compared to a median of 331/2059 across 100 simulations, $p \approx 0$, see *Methods*), but the majority of CspC/E binding sites did not directly occlude known RNase E cleavage sites. Furthermore, the presence of an RNase E cleavage site within a peak did not appear to influence differential decay rates upon *cspC/E* deletion (**Figure S7E**). This suggests that rather than directly protecting cleavage sites, CspC/E may interfere with RNase E scanning (40). This is further supported by the fact that destabilized transcripts have a median of two CspC/E binding sites, while ligands without stability changes have a median of one binding site (**Figure S7F**), suggesting multiple CspC/E proteins must bind to create an obstruction of sufficient size to interfere with RNase E scanning (41).

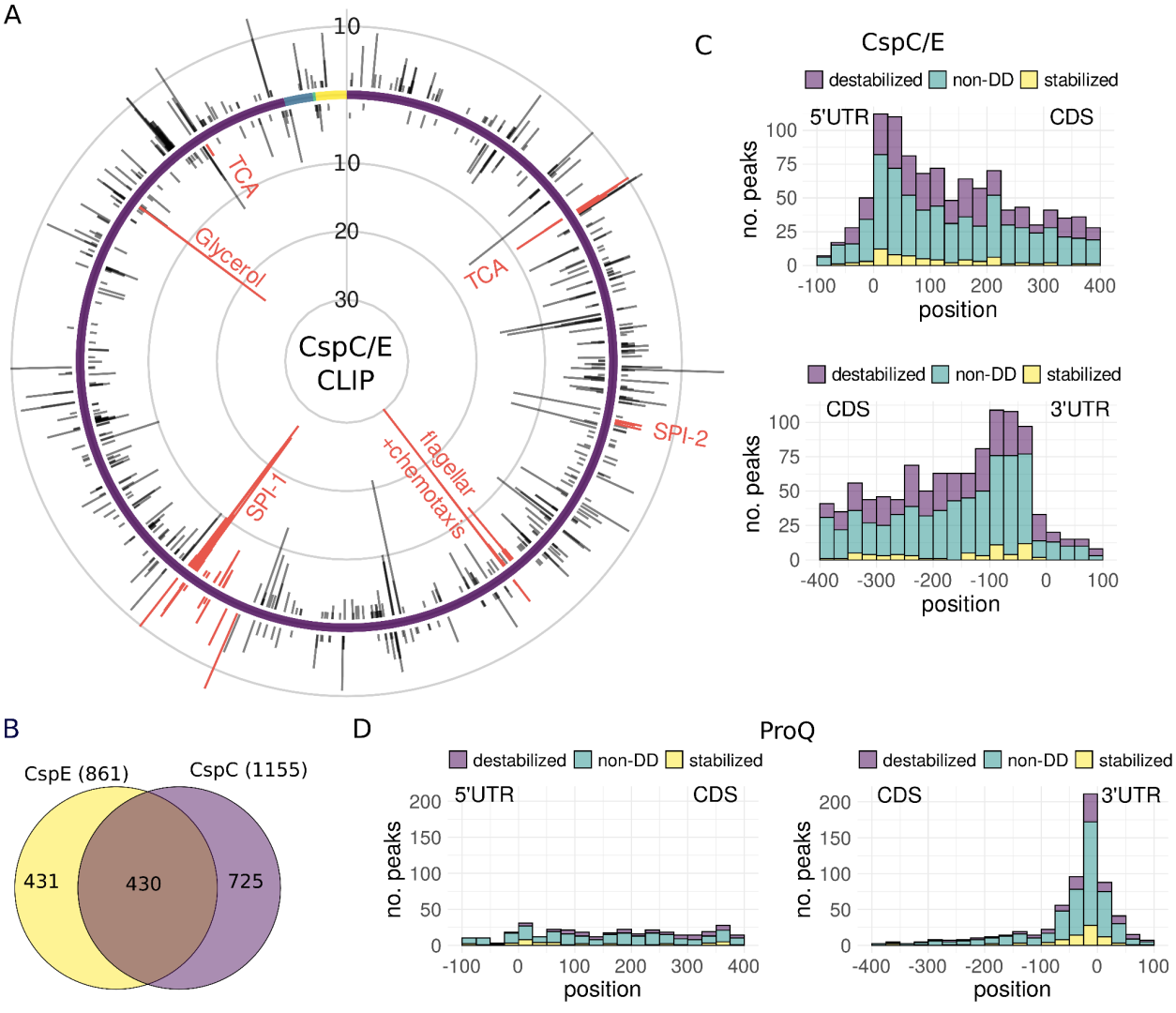


Figure 3. CspC/E CLIP-seq, comparison with RIF-seq results.
CspC/E CLIP-seq analysis: (A) Number of CspC/E binding sites binned by genomic position for the positive (outer) and negative (inner) strand. The chromosome is indicated in purple and the three plasmids in blue, green and yellow. (B) Venn diagram of binding sites, with shared targets defined by an overlap by at least 12 bases between CspE and CspC sites. (C-D) Metagene plot of transcripts bound by the respective RBP, ordered by position of CLIP-seq peak relative to the start/stop codon. Target sequences are colored by the effect of RBP deletion on stability: destabilized (purple), stabilized (yellow), or no differential decay (non-DD, green).

Long-lived transcripts rely on global RBPs for their stability

To further examine the global impact of RBPs in shaping the transcriptome, we investigated the relationship between our estimated mRNA half-lives and RBP binding as determined by CLIP-seq for four major Salmonella RBPs at early stationary phase: ProQ (12), CspC/E (this study), Hfq and CsrA (42). After sorting transcripts by stability, we saw a clear association between half-life and RBP binding, with over half of the 500 most stable transcripts ($t_{1/2}$ > 2.5 min) bound by at least one RBP (**Figure 4A**). While the probability of detecting a CLIP-seq peak increases with transcript abundance, there is no correlation between transcript abundance and stability (**Figure S3F**) suggesting the relationship between RBP binding and stability is unlikely to be an artifact of our measurements. Long-lived transcripts are also more likely to be destabilized upon RBP deletion than shorter-lived ones, regardless of RBP-binding. Of the 500 most stable transcripts, 32% are significantly destabilized in the absence of ProQ and 51% in the absence of CspC/E. Investigating the relationship between transcript half-life and differential half-life upon RBP deletion revealed large changes in median half-life for stable transcripts (**Figure 4B**), indicating that long-lived transcripts are not only bound by RBPs but also rely on them for their stability.

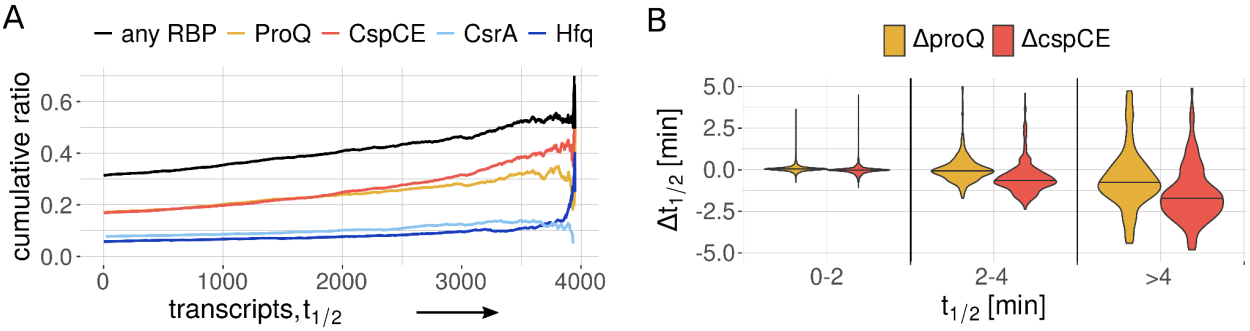


Figure 4. Global effect of RBP binding on transcript stability
(A) Cumulative ratio of transcripts bound by RBPs. The transcripts were ordered by half-life and the fraction of transcripts bound by RBPs was calculated starting from the most long-lived transcript. (B)

Change in half-life in the Δ*proQ* and Δ*cspCE* strain for transcripts with a half-life of 0-2 min, 2-4 min, and greater than 4 min.

RBPs play overlapping and complementary roles in infection-relevant pathways

To investigate the physiological consequences of RBP deletion, we identified pathways enriched in differentially stabilized and differentially expressed transcripts in the *proQ* and *cspC/E* deletion strains with the GSEA algorithm (43) (**Figure 5A & S8A**). Surprisingly, we found a large overlap in enriched gene sets in both deletion backgrounds. On the level of stability this included responses to extracellular stimulus and oxidative stress, flagellar assembly, and metabolite transport and utilization pathways including the phosphotransferase system and glyoxylate and dicarboxylate metabolism. Several of these gene sets were also enriched in differentially expressed transcripts, though the directions of the changes were often inconsistent with the observed effects on stability. For instance, genes involved in flagellar assembly were expressed at lower levels in both deletion strains despite their transcripts being stabilized (**Figure S8B & S12**). Some pathways, such as aerobic and anaerobic respiration, showed consistent changes in expression levels across both strains despite no clear shared enrichment on the level of stability.

The large overlap in pathways affected at the level of stability and expression between the two RBP deletion strains led us to investigate the relationship between ProQ- and CspC/E-mediated regulation. We examined transcripts significantly differentially expressed in both strains, finding a strong correlation between the steady-state log fold-changes (**Figure 5B**, r=0.89). The slope of a line fitted to these changes indicated stronger average changes in the *cspC/E* deletion; this was particularly clear for genes involved in flagellar assembly and chemotaxis which exhibited an ~2-fold lower expression in the Δ*cspC/E* background compared to Δ*proQ*. The lack of a similarly strong correlation for changes in transcript stability (**Figure S8C & D**) suggested that some of the similarities in changes in steady-state mRNA abundance between the two deletion strains may be due to indirect regulation, with that in the Δ*proQ* background possibly mediated by changes in CspE expression.

Given the strong changes we observed in mRNA abundance for genes involved in aerobic and anaerobic respiration (**Figure 5A & S13**), we investigated the regulon of the aerobic respiratory control response regulator ArcA whose transcript was destabilized in both RBP

deletion strains (**Figure S9A**) but not differentially expressed. We found a strong rank-correlation among genes in the ArcA regulon between differential expression previously observed in a Δ*arcA* strain (44) in both RBP deletion strains (**Figure 5C; S9B & C**), with a stronger correlation in the Δ*cspC/E* background (ρ=0.62 vs. 0.50) in keeping with the stronger destabilization of the *arcA* mRNA in this strain (**Figure S9A**). This indicates that some of the shared changes in steady state mRNA abundance observed in both deletion strains may be the result of shared regulation of ArcA expression, and further that transcript destabilization may be sufficient to affect protein abundance in the absence of clear changes in transcript abundance.

Despite the large overlaps in mRNA stability and abundance changes, there were a number of changes specific to each RBP, though these were often in the same pathways. For instance, deletion of each RBP affected the stability of a discrete set of secreted effectors involved in host cell invasion (**Figure S9D**). Another such change was in the oxidative stress response pathway, where we saw a stronger enrichment for destabilized transcripts in the Δ*proQ* strain (**Figure S11**). Transcripts destabilized by *proQ* deletion included those encoding for the oxidative stress regulator OxyR (**Figure S9E & F**), the superoxide dismutase SodB, the catalase-peroxidase KatG, and the DNA protection during starvation protein Dps; however, few of these transcripts showed significant differences in mRNA abundance.

To test if destabilization was predictive of phenotype, we exposed the Δ*proQ* and *proQ*++ strains to varying concentrations of hydrogen peroxide, including a Δ*oxyR/S* strain as a control. After exposure to 1.5mM H_2O_2 we saw a survival defect for both Δ*proQ* and *proQ*++ strains intermediate between wild-type survival and that of the Δ*oxyR/S* strain (**Figure S9G**). However, this defect was concentration dependent: exposure to 2mM H_2O_2 led to a severe survival defect in the Δ*proQ* strain that could be complemented by *proQ* overexpression (**Figure 5D**), while Δ*oxyR/S* behaved similarly to wild-type. This indicates that the Δ*proQ* survival defect after exposure to high concentrations of H_2O_2 is independent of any effects ProQ has on the stability of the *oxyR* transcript and likely depends on the effects of ProQ on other transcripts involved in the oxidative stress response. The sensitivity of the Δ*proQ* strain to oxidative stress also shows that changes in transcript stability can be predictive of RBP deletion phenotype, even without corresponding changes in transcript abundance under standard growth conditions.

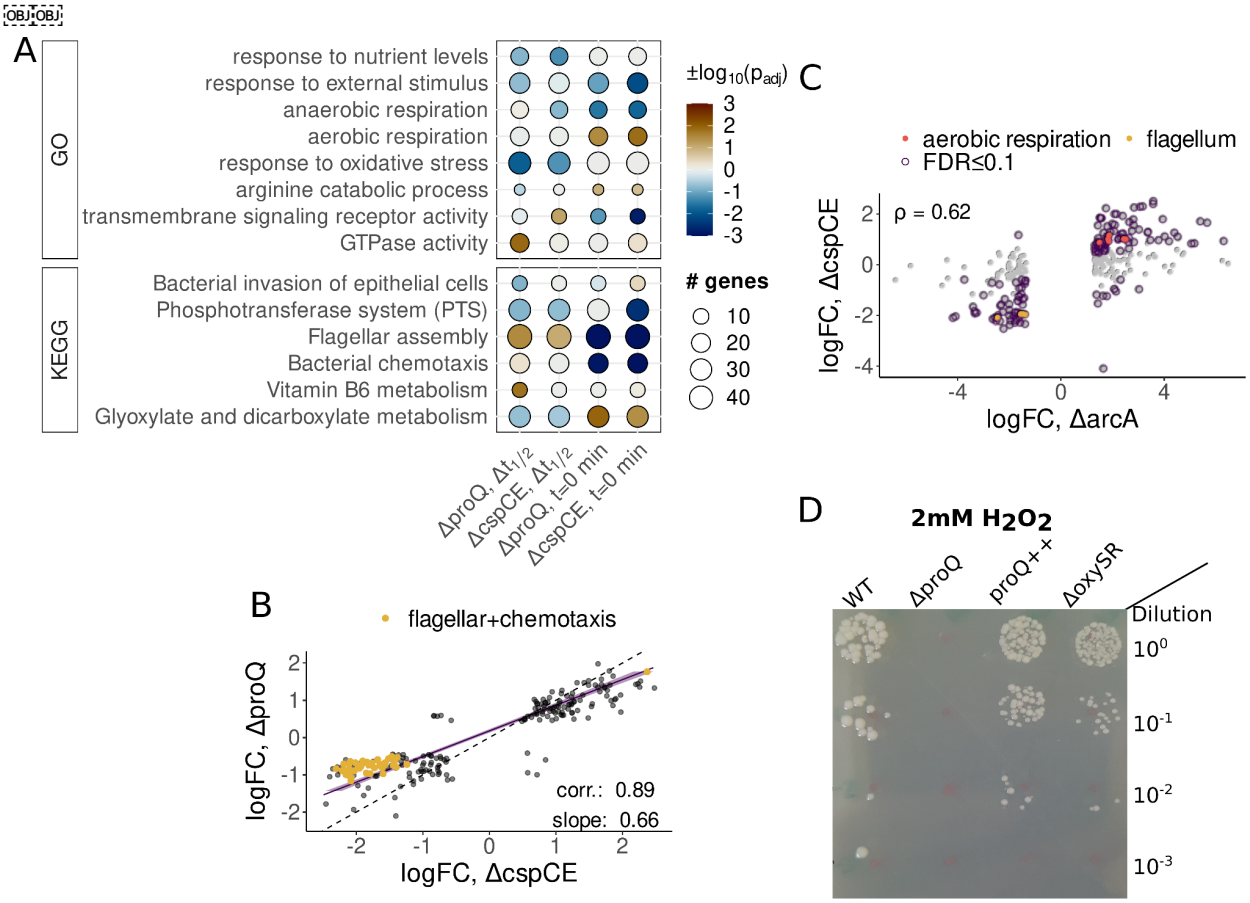


Figure 5. Integrative analysis of RBP binding and transcript stability

(A) Comparative analysis of pathways enriched in cspCE and proQ RIF-seq data. Pathways enriched in transcripts destabilized or with negative steady-state log-fold changes ($t = 0$ min) upon RBP deletion are marked blue. Pathways enriched in stabilized transcripts or positive log-fold changes are marked brown. (B) Genetic features with significant log-fold changes in both the *proQ* and the *cspC/E* deletion mutant. (C) log-fold changes in the *cspC/E* deletion mutant vs. an arcA deletion mutant (44). (D) Exposure of various Salmonella strains to 2mM of hydrogen peroxide.

Discussion

Together with transcription and translation, mRNA degradation is one of the fundamental processes controlling protein production in the cell. Rapid turnover of mRNAs underlies the ability of bacteria to rapidly adapt to new conditions: as protein production is constrained by the translational capacity of the available ribosome pool (45), clearance of transcripts encoding for unneeded proteins is essential to change the composition of the proteome. Previous work based on RNA-seq and microarray analysis of rifampicin time course data in *E. coli* and *Salmonella* has reported average mRNA half-lives in the range of 2 to 7 minutes (20–22, 46–48). The most similar prior RNA-seq study to our own reported an average half-life of 3.1 minutes across ~1200 transcripts in *E. coli* grown to stationary phase (22), over three times our

estimated average decay rate of 0.9 minutes in *Salmonella* at ESP. This discrepancy appears to originate from not accounting for the baseline stable RNA concentration, leading to a systematic underestimation of the decay rate. Interestingly, our estimates are in the range of those derived from classic experiments that pulse radiolabeled bulk RNA and determined average mRNA half-life to be ~0.7 minutes in exponentially growing *E. coli* (49), far shorter than any other subsequent estimates based on high-throughput approaches. This rapid decay may in part underlie the ability of bacteria to rapidly adapt their transcriptomes, as constant transcription would be required to maintain mRNA concentrations. As mRNA half-lives have primarily been determined by rifampicin treatment followed by sequencing or microarray analysis in those bacteria where transcriptome-wide measurements are available (50), our results suggest that mRNA stability has likely been widely overestimated and that a general reevaluation of bacterial transcript stability is in order.

Our hierarchical Bayesian analysis of RIF-seq data provides a principled framework for the analysis of RNA turnover, including the determination of differential decay rates after deletion of an RBP of interest. The flexibility of Bayesian analysis allowed us to account for nonlinearities due to confounding factors like transcription elongation after rifampicin addition and RNA baseline concentration, removing substantial biases in our determination of decay rates. Despite our best efforts, it is likely that there are still some limitations to our analysis. For instance, our control for false discovery rates means that we have likely missed some genuine instances of differential decay. Our simulations suggest an 85% sensitivity for the most precisely measured transcripts, but this falls to ~30% when considering the whole transcriptome (**Figure S4D**). Other limitations may be due to uncontrollable effects in the data. For example, some *Salmonella* promoters have previously been shown to respond specifically to subinhibitory rifampicin (51), which could introduce some bias to decay rate estimates for affected transcripts should similar effects occur with the rifampicin concentrations used here. Manual inspection of our decay curves suggests this is unlikely to be a widespread problem in our data. Similarly, if RBP deletion leads to modulation of expression of cellular RNases, our individual differential decay rates may not be reflective of the differential decay induced by simple ablation of an RBP binding site. Such a bias dependent on the cellular context of rifampicin treatment has previously been observed for the sRNA RyhB whose stability critically depends on the presence of its target mRNAs (52).

Regardless of potential biases, high-throughput methods provide at least one major advantage over classical molecular approaches to RBP characterization: numbers. Where previously a small handful of transcripts were known to be stabilized by 3′ binding of ProQ in

Salmonella, we find 86 candidates. Similarly, we expand the number of transcripts known to be stabilized by CspC/E from two to a predicted cohort of 177. By combining CLIP-seq (12) and RNase E cleavage profiling (39) with our differential stability data, we have defined cohorts of transcripts likely subject to particular modes of RBP regulation. Depending on the binding site within a transcript, up to 44% (CspC/E, CDS) and 32% (ProQ, close to the stop codon) of direct RBP targets showed altered stability upon deletion of the respective RBP. However, in both cases transcripts stabilized by RBP binding are outnumbered by those apparently bound, but unaffected at the level of stability, raising numerous questions about RBP function. How are transcripts stabilized by RBPs differentiated from those that are not? Do RBP interactions that do not affect stability perform other functions in the cell? Our analysis suggests CspC/E may protect some transcripts from RNase E through a roadblock mechanism (40); CspC/E targets may additionally or alternatively be regulated at the level of translation (53) or antitermination (54) through the manipulation of mRNA secondary structure as has been shown for the targets of other CSPs. Alternative roles of ProQ remain to be well defined, but it has been shown to play a role in gene regulation by sRNAs (14, 55). By defining and partially characterizing RBP targets, our data provides a starting point for the molecular investigations needed to further define the functions of CspC/E and ProQ.

The degree to which post-transcriptional regulation shapes the bacterial proteome has long been controversial. Recent work has suggested that, on average, protein concentrations are primarily determined by promoter on rates with post-transcriptional regulation playing only a minor role (56). Here in contrast, we have shown that deletion of bacterial RBPs thought to act primarily at the post-transcriptional level leads to large changes in both RNA stability and steady-state transcript concentration, and strong phenotypes have been observed for RBP deletion in a variety of conditions (5, 16, 57). How can these findings be reconciled? Our data provides at least two potential answers. First, as suggested by the effects of *proQ* and *cspC/E* deletion on the ArcA regulon (**Figure 5C**), modulation of stability or translation of single transcriptional regulators may ultimately cause phenotypic changes by indirectly affecting the promoter on rates of a large cohort of transcripts. The lack of correlation we observe between changes in steady-state RNA levels and differential stability (**Figure 2D&E**) indicates that such indirect effects are widespread. Secondly, our analysis shows that the majority of RNA half-lives are concentrated at less than 1 minute (**Figure 1E**), and it is indeed difficult to understand how further destabilization through post-transcriptional regulation could have strong effects on translation. However, the half-life distribution is long tailed, with ~500 transcripts having half-lives of greater than 2.5 minutes and being preferentially bound by RBPs (**Figure 4A**). The

stability of this population of transcripts is strongly affected by RBP deletion (**Figure 4B**), further suggesting they may be the major targets of post-transcriptional regulation.

An accumulating body of work suggests that the post-transcriptional regulatory networks scaffolded by RBPs are interconnected. At least two Hfq-dependent sRNAs also serve as sponges for CsrA (58, 59), and RNA-RNA interactome studies have observed a substantial fraction of shared targets between Hfq and ProQ (13). Regulatory interactions between cold shock proteins (CSP) have long been observed, with deletion of particular CSPs leading to the induction of others (16, 60), presumably through undescribed feedback mechanisms. The *cspE* mRNA has previously been used as a model for understanding the molecular mechanism of ProQ protection of 3′ ends (12); our results suggest some fraction of the change in steady-state transcript levels observed in the *proQ* deletion strain may be the result of indirect regulation through CspE (**Figure 5B**). Additionally, both RBPs affect the stability of mRNAs in similar pathways (**Figure 5A**), though often by targeting different transcripts, as for the SPI-1 effectors (**Figure S8B**). We also find effects for both strains on the stability of the CsrA-sponging sRNA CsrB, with *proQ* and *cspC/E* deletion having opposite effects on half-life (**Figure S10B & C**), adding a further potential connection between RBP regulatory networks. Our reanalysis of publicly available CLIP-seq data suggests that a substantial number of mRNAs are targeted by two or more RBPs (**Figure S10D**). What this apparently dense interconnection between RBP-mediated regulatory networks means for the cell, and how RBP activity is coordinated to maintain homeostasis in diverse environmental conditions, is an open question that will likely take significant conceptual advances to answer.

Data Availability

Data deposition: All sequencing data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (SuperSeries no. GSE234010). Transcript annotations and source code for the Stan models have been made available at https://github.com/BarquistLab/RIF-seq_repo

Acknowledgements

We thank Joel Belasco, Erik Holmqvist, Susan Gottesman and Anke Sparmann for insightful comments on the manuscript, and Alexandre Smirnov for providing northern blot quantifications from (11). This project was funded in part by the Bavarian State Ministry for Science and the Arts through the research network bayresq.net (to LB, JV).

Methods

Media and Growth Conditions

For all experiments in this study, broth cultures were grown from single colonies overnight at 37 °C in LB medium (5 g/L of yeast extract, 5 g/L of NaCl, and 10 g/L of Tryptone/Peptone ex casein; Roth). Subsequently, cultures were diluted 1:100 in fresh medium, and further grown at 37°C with shaking at 220 rpm to an OD600 of 2.0 (early stationary phase (ESP), a SPI-1 inducing condition (28)).

Bacterial Strains and Plasmids

Salmonella enterica serovar Typhimurium strain SL1344 (strain JVS-1574 (61)) is considered wild-type (WT). The generation of *proQ* and *cspC/E* deletion strains by lambda red homologous recombination (62) has been previously described (11, 16). For the *proQ*++ strain, a strain containing plasmid pZE12-ProQ was used as previously described (11, 14). The complete lists of bacterial strains, plasmids, oligos and antibodies used in this study are provided in Table S8-11.

Rifampicin assay protocol for Sequencing

Wild-type (WT), ΔRBP and RBP++ strains were grown until an OD_{600} of 2.0 in three (WT, *cspCE*) or six (WT, *proQ*, *proQ*++) replicates. The cultures were treated with 500µl/ml of rifampicin (stock solution 50mg/ml resuspended in DMSO). Samples were taken before ($t = 0$ min) and after 3, 6, 12, 24 min (proQ) or 2, 4, 8 and 16 min (cspCE) of rifampicin treatment. 2ml were collected for each sample, immediately mixed with 20% vol. stop mix (95% ethanol, 5% phenol) and snap frozen in liquid nitrogen.

Subsequently, the samples were thawed on ice and centrifuged for 20 min at 4500 rpm. Half of the resuspension of each sample was then used to perform hot phenol extraction. Bacterial pellets were resuspended in 600 µl of 0.5 mg/ml of lysozyme in TE buffer pH 8 and transferred into a 2 ml Eppendorf tube. 60 µl of 10% w/v SDS was then added and the samples were mixed by inversion. Tubes were placed at 64℃ for 1-2 min until clearance of the solution, then 66 µl of 3M sodium acetate solution at pH 5.2 was added and tubes were mixed by inversion. 750 µl of phenol (Roti-Aqua phenol #A980.3) was then added to each tube, mixed by inversion and incubated for 6 min at 64℃. Tubes were then placed on ice to cool and spun for 15 min at 13 000 rpm, 4℃. The resulting aqueous layer was transferred in a 2 ml PLG tube (5PRIME) where 750 µl of chloroform (Roth, #Y015.2) was added. After mixing by inversion, the tubes were spun for 15 min at 13 000 rpm, 4℃. The obtained aqueous layer was then collected and precipitated in a 30:1 mix of 100% ethanol: 3M sodium acetate pH 6.5 at -20℃ for at least 2 hr. After centrifugation for 30 min, 13 000 rpm, 4℃, the pellet was washed with 70% ethanol and the air-dried pellet was resuspended in nuclease-free water. Total RNA was measured by nanodrop, and integrity was checked on TBE agarose gel. 40 µg of RNA in 39.5 µl of nuclease free water were then subjected to DNAse I treatment. Total RNA was denatured for 5 min at 65℃ and put back on ice. 5 µl of DNase I (Fermentas), 5 µl of DNase I buffer (Fermentas) and 0.5 µl of Superase In (Thermo Fisher Scientific) were added to the denatured RNA and incubated at 37°C for 30 min. After incubation, 100 µl of nuclease free water was added and each reaction was placed in a PLG tube containing 150 µl of PCI. Tubes were centrifuged for 15 min at 4℃, 13 000 rpm. The aqueous phases were collected and precipitated in 30:1 Ethanol/sodium acetate mix at -20℃ for at least 2 hr. total RNA. DNase treated pellets were collected by centrifugation (30 min, 4℃, 13 000 rpm) and after 70% ethanol wash, were resuspended in 25 µl nuclease free water. Prior to rRNA depletion and cDNA library preparation, 2.5 µl of 1/10 ERCC spike-ins was added to each sample.

RNA-seq libraries were prepared by Vertis AG (Freising-Weihenstephan, Germany). Ribosomal RNA was depleted using the Ribo-Zero bacterial rRNA Removal Kit (Illumina). RNA was polyadenylated with poly(A) polymerase, 5′-triphosphates were removed with tobacco acid pyrophosphatase followed by ligation of a 5′-adapter. First-strand cDNA synthesis was performed with an oligo(dT) barcoded adapter primer and the M-MVL reverse transcriptase. The resulting cDNA was PCR-amplified with a high fidelity DNA polymerase. cDNA was purified with the Agencourt AMPure XP kit (Beckman Coulter Genomics) and sequenced on an Illumina HiSeq2000. Replicate 2 of the 24 minute time point for WT, Δ*proQ*, and *proQ*++ was excluded from subsequent analysis, as rRNA depletion failed.

Processing of Sequence Reads and Mapping RIF-seq

The 75 nt RNA-seq reads were demultiplex and quality control of each sample was performed with fastQC. Afterwards, Illumina adapters were removed with Cutadapt v4.1, and STAR (63) was used to align the reads to the SL1344 genome (NCBI accessions: FQ312003.1, HE654724.1, HE654725.1 and HE654726.1). For all analyses related to annotated genomic features such as CDSs, tRNAs, and rRNAs, gene annotations from NCBI were used. We use the same definition of transcriptional units as (42) which is based on the NCBI CDS annotations, transcription start site annotations (64), and Rho-independent terminator prediction with RNIE (65). sRNA annotations are based on (11). The ERCC92.fa sequence file for the quantification of the spike-in was obtained from ThermoScientific. For quantification, htseq-count with default options was used for counting reads aligning to CDS, sRNA and ERCC spike-ins, while the 60 base sub-genic windows were counted with the option *--nonunique all* to ensure that overlapping reads are assigned to all overlapping segments. For the 60 base windows, reads were quantified separately for the positive and the negative strand.

Read Count Normalization of RIF-seq Data with ERCC spike-ins

The normalization factor for each sample s was determined using the mean M-value across 30 detected ERCC spike-ins, as there were no apparent outliers (**Fig. S2B&C**). Only transcripts with more than 10 counts-per-million (cpm) before normalization in at least three samples in the ProQ assay were retained for further analysis.

Normalized counts-per-million (cpm) were obtained by adding a pseudo count and then dividing the read counts $Y_{gcr}(t)$ by the respective library size N_s and normalization factor $n_{f,s}$ of the sample

$$\mathrm{cpm}_{gcr,s} = \frac{(Y_{gcr,s} + 0.5) \cdot 10^6}{N_s n_{f,s} + 1}$$

The Stan models were applied to the natural logarithm of the normalized cpm values $y_{gcr}(t) \equiv ln\,(cpm_{gcr}(t))$.

Removal of batch effects: center-mean normalization

Following spike-in normalization, we observed some clustering by replicate rather than condition within time point groups (**Figure S2D&I**). To account for these batch effects, we developed a center-mean (CM) normalization procedure, which can be applied after a primary normalization, e.g. with spike-ins, and compensates for small variations in the amount of spike-ins added to the individual samples. After the normalization with spike-ins, we calculated a gene-wise mean log-

count $\underline{y_{gc}}(t)$ for every condition and every time-point (see **Figure S2E** for $t = 0$ min). This value was subtracted from the observed value in every sample

$$y_{0,gcr}(t) = y_{gcr}(t) - y_{gc}(t)$$

For every sample s (uniquely defined by condition c, time t, replicate r), we calculated the mean

$$\ln(n_{b,s}) = \frac{1}{N_g}\sum_g y_{0,gcr}(t)\,,$$

where $n_{b,s}$ is an additional normalization constant. The batch-corrected cpm values are then given by

$$\mathrm{cpm}_{gcr,s} = \frac{(Y_{gcr,s} + 0.5)\cdot 10^6}{N_s n_{f,s} n_{b,s} + 1}\,.$$

PCA confirmed that the samples separated well by time point and genotype after the CM normalization (**Figure S2F&L**), and boxplots showed an improved alignment of median logcpm values (**Figure S2G,H,J,K**). Before fitting the decay curves with the Bayesian models, we subtracted the mean log-count at $t = 0$ min

$$\tilde{y}_{gcr}(t) = y_{gcr}(t) - y_{gc}(t = 0\mathrm{min})\,.$$

Calculation of detection limit

In order to regularize zero counts, we have added a pseudo count of 0.5. The library sizes vary in size around 10 million reads and the normalization constants around 1. This results in an estimated minimum log-count of

$$y_{\min} = \ln\left(\frac{0.5\cdot 10^6}{N_s n_{f,s} n_{b,s} + 1}\right) \sim \log\left(\frac{0.5}{10}\right)\,.$$

After subtracting the mean log-count at $t = 0$ min, we can calculate the detection limit for gene g in condition c as

$$\tilde{y}_{\mathrm{lim},gc} = y_{\min} - y_{gc}(t = 0\mathrm{min})\,.$$

This corresponds to the minimal possible value of the log relative expression.

Differential gene expression analysis

Log-fold changes were calculated using glmQLFit from edgeR (24) with a cutoff of 0.25 on the log-fold changes. Since batch effects present after TMM normalization (**Fig. S1 A,B&E**), samples were additionally normalized using RUVg (66) (**Fig. S1C&G**). We selected the 800 least varying genes between the ΔRBP and the WT strain. Since the differences between the *proQ*++ and the WT strain were larger, we only took the 600 least varying genes between these two conditions. The intersection between these sets is 37 genes which we used as negative control (**Fig. S1F**). The number of factors of unwanted variation k was set to 6. After RUVg normalization, the samples clustered by strain (**Fig. S1C&H**). We selected differentially expressed genes at an FDR of 0.1 (**Fig. S1I-K, Table S2**).

Extraction of RNA half-lives from RIF-seq data

We compared three statistical models, summarized in **Figure 1B**. All models assume that the normalized log counts follow a normal distribution around a condition and gene-dependent mean

$$\tilde{y}_{gcr}(t) \sim \mathcal{N}\left(\mu_{gc}(t), \sigma_g(t)\right).$$

The variance σ_g^2 is not condition-dependent. The mean μ_{gc} is parameterized differently in the three statistical models:

Linear model (LM): $\mu_{gc}(t) = -\beta_{gc} \cdot t$

Piecewise linear model (PLM): $\mu_{gc}(t) = -\Theta(t-\gamma_g)\beta_{gc} \cdot (t-\gamma_g)$

Log-normal model (LNM): $\mu_{gc}(t) = \ln\left[\pi_{gc} + (1-\pi_{gc})\exp\left(-\Theta(t-\gamma_g)\beta_{gc} \cdot (t-\gamma_g)\right)\right]$

where Θ is the Heaviside step function which is 0 for negative arguments and 1 otherwise. The baseline parameter π_{gc} introduced in the LNM corresponds to the fraction of stable RNA for gene g in condition c as compared to steady-state levels at $t = 0$ min. A hierarchical prior is used for both the WT decay rate and the standard deviation which leads to variance shrinking and reduces the effect of outliers. For other parameters (e.g. difference in decay rate), broad priors were chosen to minimize their influence on posterior estimates. Priors were defined as follows:

WT decay rate $\beta_{g,\mathrm{WT}} \sim \mathcal{N}\left(\mu_\beta, \sigma_\beta\right)$

Mutant decay rate $\Delta\beta_{gc} = \beta_{gc} - \beta_{g,\mathrm{WT}} \sim \mathcal{N}\left(0, 0.2\right)$

Standard deviation $\sigma_g(t) \sim \mathcal{N}_{\geq 0}\left(\mu_\sigma, \sigma_\sigma\right)$

Baseline parameter $\pi_{gc} \sim \mathcal{N}\left(0, 0.25\right)$, $\pi_{gc} \in [0, 0.2]$

Hyperparameters (Cauchy/normal distribution) $\mu_\beta, \mu_\sigma, \sigma_\beta \sim \mathcal{C}_{\geq 0}\left(0, 1\right)$, $\sigma_\sigma \sim \mathcal{N}_{\geq 0}\left(0.3, 0.3\right)$

Elongation time (Cauchy distribution) $\gamma_g \sim \mathcal{C}\left(0, 2\right)$, $\gamma_g \in [0, 12]$

For $\pi_{gc} = 0$, the LNM is equivalent to the PLM, which converts to the LM as $\gamma_{gc} \to 0$. The statistical models are fitted to the RIF-seq data using the probabilistic programming language Stan (v.2.30.1) (27) with two chains and 1000 MCMC samples each (`method=sample num_samples=1000 num_warmup=1000 adapt delta=0.95 algorithm=hmc engine=nuts max_depth=15`). The statistical model was applied to all four strains (WT, Δ*proQ*, *proQ*++, Δ*cspCE*) at once. The reported parameters (decay rate, half-life, transcription elongation time) correspond to the median of the 2000 MCMC samples. The median of the transcriptome-wide half-lives corresponds to the median of the 2000 MCMC samples of the hyperparameter μ_β. In addition to the 2nd replicate of the time point taken at 24 minutes for the *proQ* experiments, the 1st replicate of the 4 min time point of the Δ*cspCE* mutant was removed from this part of the analysis because it clustered together with the 0 min time point (**Figure S1I**) which strongly influenced differences in decay rate in the Δ*cspCE* mutant.

Model comparison using leave-one-out cross validation

For a quantitative comparison of the linear model (LM), the piecewise linear model (PLM) and the log-normal model (LNM), we estimated the out-of-sample predictive accuracy using leave-one-out cross validation (LOO-CV) with *Pareto-smoothed importance sampling* (PSIS) (32). The pointwise log-likelihood `log_lik` was computed in the `generated quantities` block in Stan during MCMC sampling. We used the `loo()` function from the *loo* R package (version 2.5.1), which computes the expected log-pointwise predictive density (ELPD) using PSIS.

Calculation of transcription velocities

To calculate transcription velocities, we took advantage of ongoing transcription of RNA polymerase already bound to DNA in the RIF-seq data. We split the genome into 60 base subgenic windows, and extracted the corresponding elongation times and decay rates using the log-normal model. We split the dataset into five subsets before running the MCMC sampler (1 chain: `method=sample num_samples=1000 num_warmup=1000 adapt delta=0.95 algorithm=hmc engine=nuts max_depth=15`). Subsequently, we verified that the hierarchical parameters agreed well between the five subsets. The resulting transcription elongation times γ were combined with operon annotations taken from (42). We fitted a linear model with y-intercept a_g and slope b_g to the elongation times of operons or individual transcripts as shown for the mra/fts operons in **Figure S3C,** using the inverse of the 68% credible intervals of γ_g as obtained from the MCMC samples as weights (example: Fig. 2E). The transcription velocity v_g is given by the ratio of the window size ($s_{seg} = 60$ nt) and the slope b_g. Its error was calculated via error propagation $\Delta v_g = \Delta b_g/(s_{seg} \cdot b_g^2)$. We obtain 772 operons with at least 7 nonzero segments which fulfill the quality criterium $\Delta v_g/v_g < 0.75$.

Calculation of Bayesian p-values

The half-lives were calculated from the decay rates $t_{1/2,gc} = ln\,(2)/\beta_{gc}$. In order to calculate Bayesian p-values, we tested against the null hypothesis that the difference in half-life $\Delta t_{1/2,c} = t_{1/2,c} - t_{1/2,WT}$ is compatible with zero. There is a limit as to how precisely we measured the WT half-lives. We determined the minimum of the 90% credible intervals of the WT half-lives (~0.05). Assuming that we cannot measure a difference in half-life with higher precision than the WT half-lives, we selected the interval [-0.05, 0.05] as the null hypothesis. The p-value p_{gc} for gene g in condition c corresponding to the difference in decay rate $\Delta t_{1/2,gc}$ is given by the fraction $|S_0|/|S|$ of MCMC samples $S = \{s_1, \ldots, s_{2000}\}$ that agrees with the null hypothesis (**Figure S4A**):

For $\Delta t_{1/2,gc} > 0$, the samples $S_0 = \{s \in S | s \leq 0.05\}$ agree with the null hypothesis.

For $\Delta t_{1/2,gc} < 0$, the samples $S_0 = \{s \in S | s \geq -0.05\}$ agree with the null hypothesis.

We compared the distribution of p-values to the distribution of p-values under the null hypothesis which was obtained by bootstrapping from the distribution of MCMC samples of the WT half-lives and calculating the corresponding p-values (**Figure S4B**).

Calibration of posterior predictive p-values

In order to assign a false-discovery rate (FDR) to the p-values, we simulated a dataset with 4000 transcripts, 3 conditions (*WT*, c_1, c_2) with time points 0, 3, 6, 12, 24 and 2 conditions (*WT*, c_3) with time points 0, 2, 4, 8, 16. We drew samples from the following distributions (which we extracted from fitting the LNM to the two RIF-seq data sets), using the definition of the LNM as given above:

Relative log-counts $y_{gcr}(t) \sim \mathcal{N}\,(\mu_{gc}(t), \sigma_g(t))$

WT decay rate $\beta_{g,\mathrm{WT}} \sim \mathcal{N}\,(0.75, 0.3)$

Elongation time $\gamma_g \sim \mathcal{C}_{\geq 0, \leq 12}\,(0, 2)$

Standard deviation of log-counts $\sigma_g \sim (0.35, 0.25)$

Difference in decay rate $\Delta\beta_{gc} = \beta_{gc} - \beta_{g,\mathrm{WT}} \sim \mathcal{N}(0, 0.08)$

Baseline parameter $\pi_{gc} \sim \mathcal{N}\,(0, 0.05)$

Mean of relative log-counts

$$\mu_{gc}(t) = \ln\left[\pi_{gc} + (1-\pi_{gc})\exp\left(-\Theta(t-\gamma_g)\beta_{gc}\cdot(t-\gamma_g)\right)\right]$$

Then, we fitted the log-normal model to this dataset. Simulated absolute differences in half-life below 0.05 ($|\Delta t_{1/2,gc}| \leq 0.05$) were assumed to agree with the null hypothesis. The Pearson correlation of 0.86 between simulated and fitted differences in half-life were obtained using the `weightedCorr` function from the *wCorr* package in R with the inverse of the size of the 90% credible intervals of the fitted half-lives as weights (**Figure S4C**). We calculated the posterior predictive p-values for the fitted differences in half-life and varied the p-value cutoff between 0 and 1 with step size 0.01. The corresponding FDR is given by the fraction of transcripts whose simulated difference in half-life agrees with the null hypothesis and the total number of transcripts with a p-value below the cutoff. Subsequently, we fitted a LOESS curve in R (`span=0.2`) to determine the FDR corresponding to any p-value cutoff (**Figure S4E, Table S1**) A p-value of 0.082 corresponds to a FDR of 0.1 which we used as a cutoff for our analysis of differentially decaying transcripts. In addition to controlling the FDR, we verified that at an FDR of 0.1, the log-normal model identifies differentially decaying transcripts with a low simulated standard deviation on log-counts σ_g with high sensitivity (**Figure S4D**). For this, we selected five cutoffs on standard deviation (0.05, 0.1, 0.2, 0.4, 1) and calculated the false positive rate (FPR) and sensitivity for all transcripts below the cutoff.

Ratio between differential gene expression and stability in Δ*proQ* vs. Δ*cspCE*

The ratio between the differences in RNA half-life in the Δ*proQ* and Δ*cspCE* mutant strain (as compared to WT) was calculated by selecting only transcripts with significant stability changes in the same direction in the two RBP deletion strains (**Figure S9C**). Similarly, we selected only genes with significant $\log_2$-fold changes in the same direction in both RBP deletion strains (**Figure 4F**).

Hydrogen peroxide exposure

Bacterial cultures of all strains (*Salmonella* WT, Δ*proQ*, *proQ*++ and Δ*oxyRS*) were grown overnight. All strains contain the pJV300 plasmid. 10 mL of culture were inoculated 1:1000 in LB and grown at 37°C for 5 h. The cultures were then diluted 1:100 in 10 mL of LB and incubated for 2 h at 37°C with 1.5 mM or 2 mM of H_2O_2. Controls were incubated without H_2O_2. Viability of the cells was assessed by spotting 5 µL of a dilution series ($10^0, \ldots, 10^{-6}$) on LB agar plates which were then incubated overnight at 37°C.

RNA secondary structure

The ViennaRNA web server with default settings was used to obtain the secondary structures of RNA sequences (67). Forna (68) was used for visualization.

Geneset enrichment analysis

In order to identify pathways with transcripts either stabilized or destabilized in the absence of ProQ or CspCE, the genes in the analysis were ranked according to the quantity $-sgn\left(\Delta t_{\frac{1}{2}}\right) log_{10}(p + 10^{-4})$. For pathway analysis of log fold-changes, we used the quantity $-sgn(logFC)log_{10}(p)$ for ranking. We created a gene set database combining the terms for the

strain SL1344 from the eggnog database (69), QuickGo (70) and KEGG (71). We used the r package GSEA 1.2 (43) to calculate the enrichment scores and the corresponding adjusted p values. gsea.type was set to 'preranked' and shuffling.type to 'gene.labels'. Gene sets with sizes between 3 and 50 genes were analyzed. For the CLIP-seq data, we performed a hypergeometric test with the R stats function fisher.test. The FDR corrected p value was obtained using the Benjamini Hochberg procedure. For the hypergeometric test, the significance cutoff on the CLIP-seq data was chosen as ($p_{adj} \leq 0.1$) for CsrA, Hfq, and ProQ. For CspC/E, the value was reduced to $p_{adj} \leq 0.01$ to obtain a comparable number of interaction partners.

Correlation between changes in half-lives or transcript abundance in different mutant backgrounds and its significance

To compare the changes in half-life or transcript abundance in the Δ*proQ* and Δ*cspCE* strains, we used all transcripts with an FDR ≤ 0.1 in both deletion mutants as compared to WT. The Pearson correlation was calculated using R's *cor* function. To calculate the significance of the correlation, we permuted gene labels of the results for the Δ*cspCE* strain and applied the same criteria to obtain the correlations. This was repeated 1000 times. The fraction of permutations with a correlation larger than in the actual comparison between the Δ*proQ* and Δ*cspCE* strains was used as an estimate for the significance of the observed correlation (**Figure S8D**).

For the comparison between changes in transcript abundance in the Δ*arcA* strain and the two RBP deletion strains (always compared to WT), all transcripts with differential gene expression in the Δ*arcA* strains were used (44). To map locus tags from the *S*. Typhimurium LT2 genome (NCBI Accessions: AE006468.2, AE006471.2) to the SL1344 genome, we used proteinortho (v6.0.33) (72). The significance of the correlation between the published difference in transcript abundance and the differences in the RBP deletion strains were estimated by permuting gene labels 1000 times in the Δ*arcA* strain and calculating the corresponding correlations (**Figure S9C**).

RNase E cleavage sites in random sequences

To estimate expected overlap between CspC/E CLIP-seq peaks and RNase E cleavage sites, we generated 100 random peaks of the same length within the same transcript as the actual CspC/E CLIP-seq peak, then tested how many of these random peaks overlapped with RNase E cleavage sites. Across the 100 simulations, this resulted in a mean of 331 of overlapping binding sites compared to 410 overlapping sites in the CspC/E CLIP-seq peaks. None of the 100 simulated sets yielded a value as high or higher than 410 overlapping binding sites, resulting in a p value of about 0.

Cumulative ratio of transcripts bound by RBPs

To visualize the relationship between the fraction of transcripts bound by RBPs and transcript half-life (**Figure 4A**), we calculated the cumulative ratio R_{cum} of transcripts bound by one of the RBPs in this study (ProQ, CspC/E, CsrA, Hfq). For every RBP, we divide the set of transcripts T into transcripts bound and not bound by the respective RBP, i.e.

$$T = T_b \cup T_{ub}.$$

Subsequently, each set is ordered by half-life. Starting with the least stable transcript, the cumulative ratio of the ith transcript is given by

$$R_{cum,i} = \frac{|\{T_b | t_{1/2} \leq t_{1/2,i}\}|}{|\{T | t_{1/2} \leq t_{1/2,i}\}|}.$$

UV Crosslinking, Immunoprecipitation, and RNA Purification

CspC/E CLIP-seq data sets were generated with the same protocol as (12, 42). In short, 400mL of bacterial culture was grown to an OD_{600} of 2.0 in three biological replicates. One half of the culture was irradiated with UV-C light at 800 mJ to induce RBP crosslinking. Cells were centrifuged and resuspended in lysis buffer, mixed with 1 ml glass beads and shaken for 10 minutes. Anti-FLAG magnetic beads were added to the lysate before rotating it for 1 hour at 4℃. The beads were collected by centrifugation, resuspended and subjected to multiple washing steps. Finally, the magnetic beads were collected on a magnetic separator and the supernatant was loaded and separated on a 15% SDS-polyacrylamide gel followed by transfer to a nitrocellulose membrane. The protein size marker was highlighted with a radioactively labeled marker pen, and the membrane was exposed to a phosphor screen for 30 min. The regions of the membrane containing radioactive signal were cut out, and the same regions were selected from the control samples (**Figure S7A**). The membrane pieces were cut into smaller pieces and incubated 1hr at 37℃ with shaking at 1000 rpm in a total volume of 400 µl of PK solution (200 µl of 2xPK buffer - 100mM Tris-Hcl pH 7.9; 10mM EDTA; 1% SDS - ;20 µl of Proteinase K (Fermentas, 20 mg/ml); 1 µl of SuperaseIN (Termo Fischer Scientist) completed with nuclease-free water up to 400 µl). After incubation, 100 µl of the PK solution containing 9M Urea was added to each tube and incubated for an additional 1hr at 37℃, 1000 rpm. For RNA extraction, phase-lock tubes (5PRIME) were used to mix 450 µl of Phenol:Chloroform:Isoamyl alcohol 25:24:1 (PCI; Roth) with the supernatant from proteinase K treated samples (around 450 µl). Phase lock tubes were incubated 5 min at 30℃ under agitation (1000rpm) and spined 15 min at 4℃, 13 000rpm. The aqueous phase was collected and precipitated using a 30:1 mix of 100% ethanol/3M Sodium Acetate pH 5.2 at -20℃ for at least 2hr. After 30min centrifugation at 4℃, 13 000rpm, the RNA pellets were washed with 70% ethanol and finally resuspended in 10 µl of nuclease-free water.

CLIP-seq cDNA Library Preparation and Sequencing

cDNA libraries were prepared using the NEBnext Multiplex Small RNA library kit (#E7300) according to the manufacturer's recommendation. Briefly, for the 3' SR adaptor ligation step, 2.5 µl of RNA sample extracted from CLIP elution was mixed with 1 µl of 3'SR adaptor, diluted 1:10 in nuclease-free water), incubate in thermal cycler 2min at 70℃. While on ice, a mix of 5 µl of 3'

ligation reaction buffer and 1.5 µl 3' ligation enzyme mix was added, and the samples were incubated for 1 hr at 25℃. For the RT primer hybridization, 2.75 µl of a 1:10 diluted SR RT primer was added to the samples following an incubation of 5 min at 75℃, 15 min at 37℃ and 15 min at 25℃. During incubation period, 0.5 µl of a 1:20 5' adaptor was incubated separately for 2 min at 70℃. This denatured 5' adaptor was used for the 5' SR adaptor ligation step where it was added to the samples with 0.5 µl of 10X 5' ligation reaction buffer and 1.25 µl of 5' ligation enzyme mix. The samples were then incubated for 1 hr at 25℃. For the final step, reverse transcription, to each sample was added 4 µl of first strand synthesis reaction buffer, 0.5 µl of murine RNase inhibitor and 0.5 µl of M-MuLV reverse transcriptase. The samples were incubated for 1 hr at 50℃ and the RT enzyme later on inactivated at 70℃ for 15 min. For the cDNA amplification, 10 µl of each cDNA library was mixed with 25 µl of LongAmp Taq 2x Master mix, 1.2 µl of SR primer, 12.5 µl of nuclease free water and 1.2 µl of index primer (one different for each library). Amplification conditions applied were the following: 94℃ for 30 sec; 18 cycles of 94℃/15sec; 62℃/30sec; 70℃/15sec and a final step of 70℃ for 5 min. After amplification, samples were loaded of TBE gels and bands from amplification between 130 to 200 bp were selected by gel extraction. DNA was eluted from crushed gel pieces with 500 µl of DNA elution buffer after 2 hr incubation at RT. After collection of the supernatant using corning costar spin-X centrifuge tube filters, precipitation mix was added, and samples were placed at 80℃ for 1 hr. After centrifugation and washing steps, dried pellets were resuspended in nuclease free water. Size, quantity, and absence of primers dimers were checked by bioanalyzer before sequencing. High-throughput sequencing was performed by Vertis. The libraries were pooled on an Illumina Nextseq500 platform and sequencing done for single end 1x150 bp.

Processing of Sequence Reads and Mapping CLIP-seq

The CspC/E and ProQ CLIP-seq data was analyzed following the procedure described in (42) with a few alterations. First, putative PCR duplicates were removed using FastUniq v1.1 (73). The read pairs were trimmed together using Cutadapt v4.1 (74) and reads with fewer than 12 remaining bases were discarded. Additionally, we performed quality trimming with a minimum phred score of 20. Read pairs longer than 25 nt were eliminated for peak calling. The remaining reads were mapped to the Salmonella Typhimurium SL1344 chromosome (NCBI Acc.-No: NC_016810.1) and plasmid (NCBI Acc.-No: NC_017718.1, NC_017719.1, NC_017720.1) reference sequences using segemehl version 0.3.4 (75) with an accuracy cutoff of 80%. Only uniquely mapping reads were considered for all subsequent analysis.For quantification of peak regions, no upper limit was imposed on read length. Reads were aligned to the Salmonella

Typhimurium SL1344 chromosome and plasmids using STAR (63).

CLIP-seq Peak Calling

Segemehl read alignments were converted from BAM to BED format using BEDTools v2.17.0 and reformatted to satisfy blockbuster's input requirements. Subsequently, peaks were defined by applying blockbuster v0.0.1.1 (-minBlockHeight 10 -distance 1). This resulted in a large set of clusters with overlapping blocks of reads. In clusters with only one block the peak region was defined by the position of the block. In clusters with multiple blocks, peaks were chosen iteratively. First, the block with the highest count was selected and a peak region was defined by joining together all blocks which overlapped by at least 50% with this block. Then, all reads overlapping with this block were removed. This procedure was repeated until the largest block contained less than 1% of the reads in the corresponding cluster. A formalized description of this algorithm is given in (42). The peaks were exported to gff format and htseq-count v2.0.2 with default parameters was used to count the uniquely mapped reads in the STAR alignments.

Differential peak abundance analysis of CLIP-seq Data

DEseq2 (25) was used to identify peaks with differential abundance in the cross-linked vs. the non-cross-linked libraries. Log-fold changes were shrunk using apeglm (76). We required a log-fold change of at least 1. For ProQ, we chose the same adjusted p-value cutoff as (42) chose for CsrA and Hfq (padj<0.1, **Figure S7B**). For the CSPs, the adjusted p-value cutoff was reduced to 0.01 to obtain a comparable number of peaks (**Figure S7C**).

Supplementary Material

Table S1: Half-lives obtained by fitting the LNM to the RIF-seq data set.

Table S2: Steady-state log-fold changes of the RIF-seq data set from edgeR.

Table S3: Genetic features with ProQ-binding sites (CLIP-seq) in the 3'UTR or within 100 bases of the stop codon which are destabilized upon *proQ* deletion.

Table S4: Genetic features with CspC/E-binding sites (CLIP-seq) in the CDS or 5'UTR which are destabilized upon *cspC/E* deletion.

Table S5: Significant ProQ peaks obtained by re-analyzing the ProQ CLIP-seq data set (12).

Table S6: Significant CspC CLIP-seq peaks.

Table S7: Significant CspE CLIP-seq peaks.

Table S8: Bacterial strains used in this study.

Table S9: Plasmids used in this study.

Table S10: Oligos used in this study.

Table S11: Antibodies used in this study.

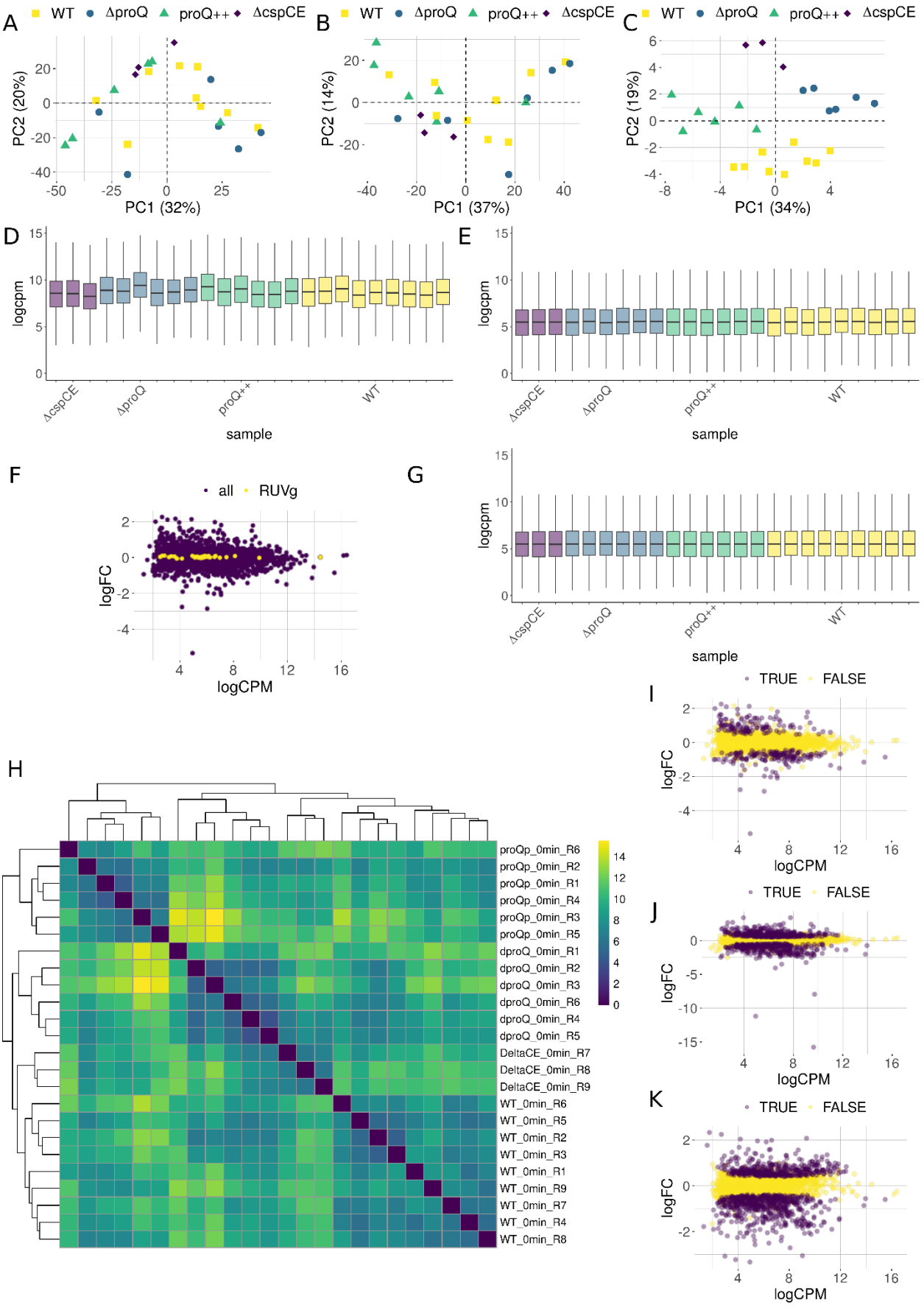


Figure S1. Differential expression (DE) analysis at t=0 min

(A) PCA plot after normalizing by library size. (B) PCA plot after TMM normalization. (C) PCA plot after running RUVg. (D) logcpm values after normalizing by library size. (E) logcpm values after TMM normalization. (F) MA plot Δ*proQ* vs. WT after TMM normalization. Genes used in RUVg are marked in yellow. (G) logcpm values after running RUVg. (H) Samples clustered by euclidean distance after running RUVg. (I-K) MA plots for Δ*proQ*, *proQ*++, and Δ*cspCE* (top to bottom) vs. WT after normalization with RUVg. Significantly DE genes (FDR < 0.05) are highlighted.

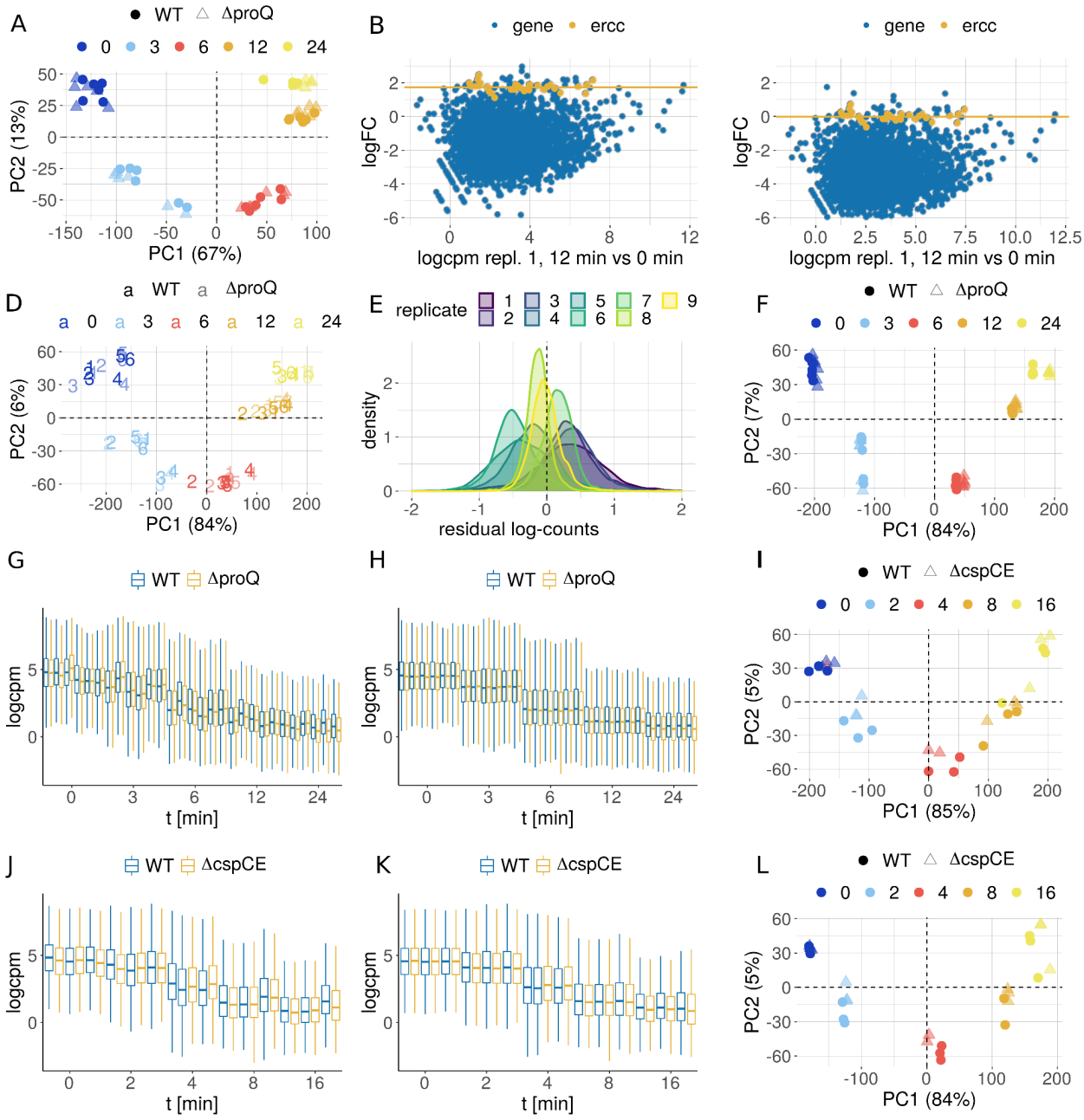


Figure S2. Normalization

(A) PCA plot of the raw proQ RIF-seq data. (B) Representative MA plot before TMM normalization with ERCC spike-ins. (C) Representative MA plot after TMM normalization with ERCC spike-ins. (D) PCA plot after normalization with ERCC spike-ins. (E) Illustration of the CM normalization for WT at $t = 0$ min: After subtracting the condition-wise mean, the offset of the mean from 0 provides an additional normalization constant. (F) PCA plot after center-mean (CM) normalization. (G) Relative log-expression (RLE) of WT and Δ*proQ* libraries after TMM normalization. (H) RLE of WT and Δ*proQ* libraries after CM normalization. (I) PCA plot after TMM normalization with ERCC spike-ins. (J) RLE of WT and Δ*cspCE* libraries after TMM normalization. (K) RLE of WT and Δ*cspCE* libraries after CM normalization. (L) PCA plot after center-mean (CM) normalization.

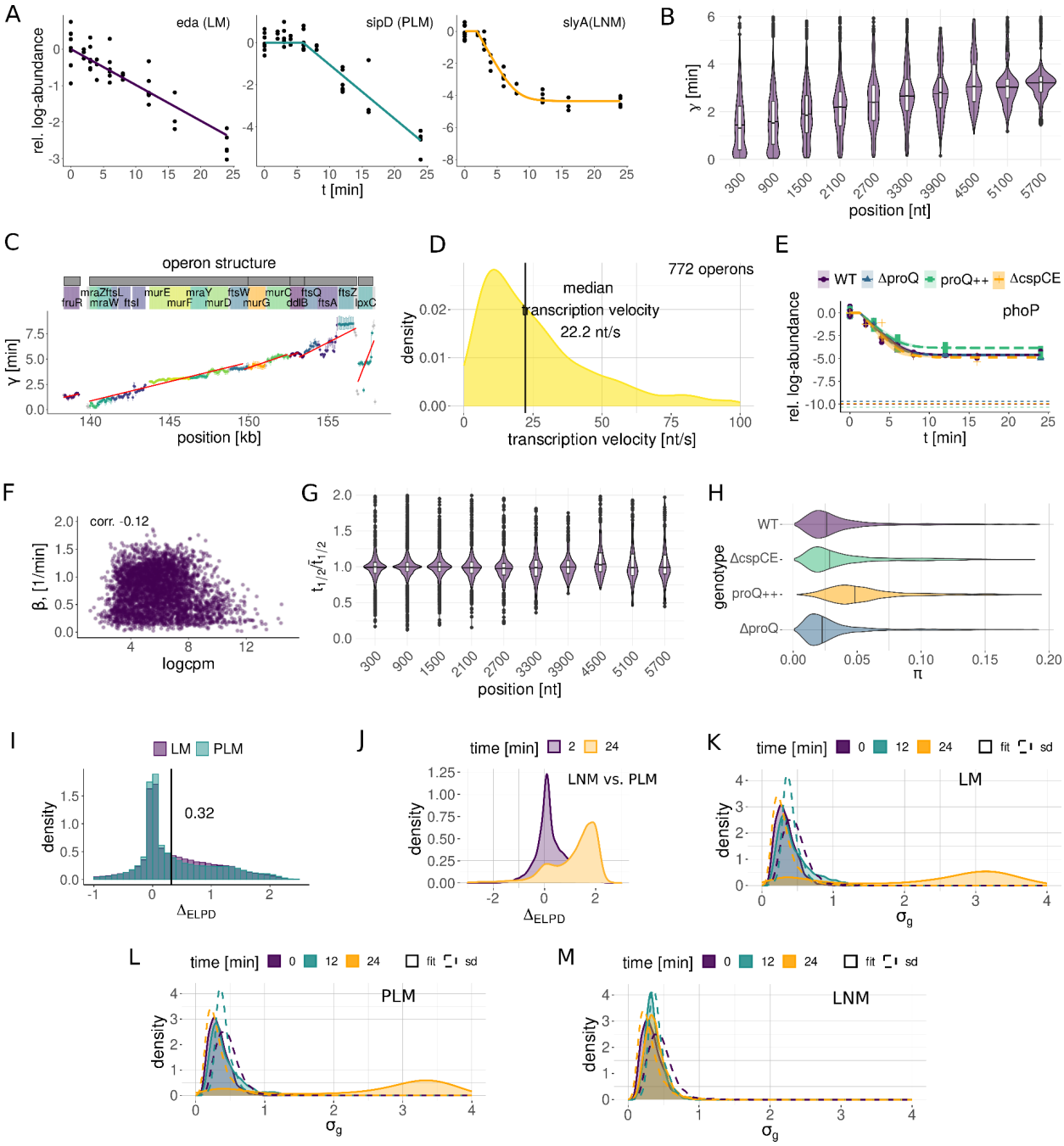


Figure S3. Model development, global transcript stability and transcription rate

(A) Representative WT decay curves exhibiting the dynamics correctly described by either the LM, PLM or LNM. (B) Genome-wide association of elongation time γ with the annotated primary transcription start sites. (C) An example of how the transcription rate was extracted from the elongation times of the 60 base windows including annotations and operon structure based on (64), where primary TSSs are indicated by black lines. (D) Distribution of genome-wide transcription rates as extracted from the 60 base windows. (E) Comparison of decay curves and detection limit due to adding a pseudocount for the *phoP* transcript. (F) Scatter plot of gene expression and decay rate. (G Half-lives of the 60 base windows scaled to the gene-average relative to the start of the CDS. (H) Distribution of genome-wide stable baseline fraction π

ordered by genotype. (I) Difference between ELPD of the LNM and the LM/PLM, respectively. Positive values favor the LNM. (J) Difference in ELPD between the LNM and the PLM at 24 min. Positive values favor the LNM. (K) Fitted unexplained variation $\sigma_g(t)$ in the LM compared to calculated standard deviation. (L) Fitted unexplained variation $\sigma_g(t)$ in the PLM compared to calculated standard deviation. (M) Fitted unexplained variation $\sigma_g(t)$ in the LNM compared to calculated standard deviation.

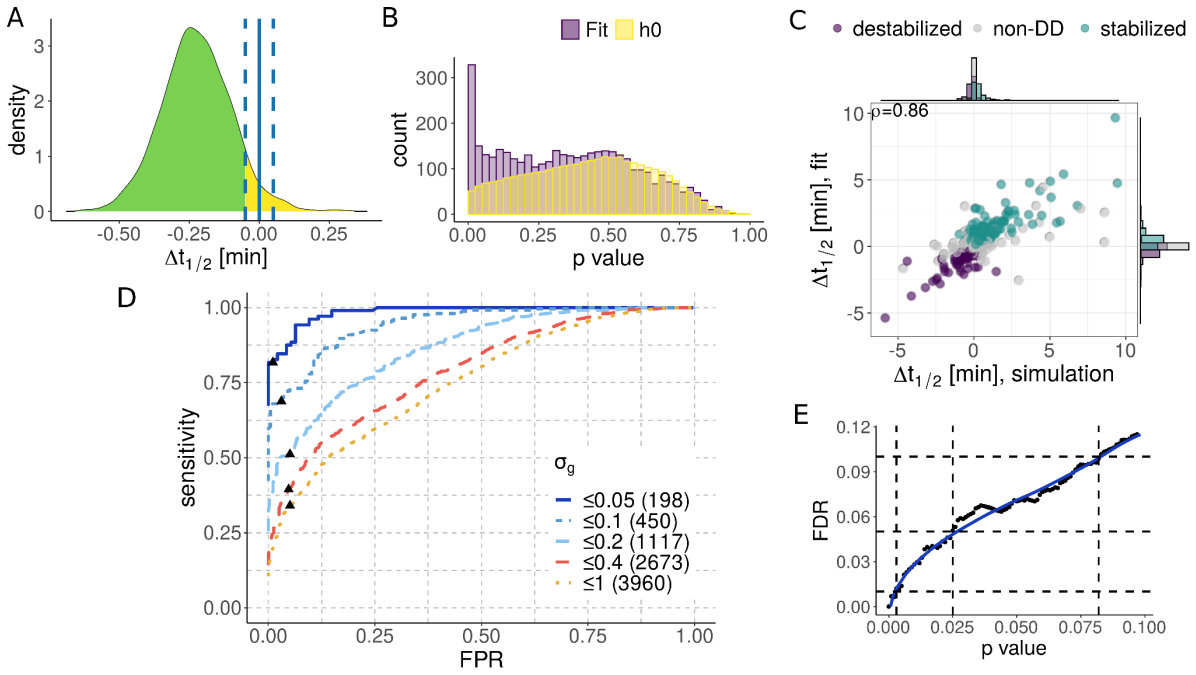


Figure S4. Posterior predictive p values und false discovery rate (FDR)
(A) Posterior distribution of the difference in half-life for a transcript in the RBP deletion strain vs. the WT. Under the null hypothesis, MCMC samples should fall within an interval around zero (blue, dashed lines). The Bayesian p value is given by the fraction of samples that overlap with the null hypothesis (yellow). (B) p value distribution for differential stability data for the Δ*proQ* strain compared to the distribution under the null hypothesis. (C) Correlation between simulated and fitted differences in half-life (Pearson $\rho = 0.86$). (D) ROC curves obtained from the simulated data for different simulated standard deviations of relative log-counts. An FDR of 0.1 is marked with a black triangle. The number of transcripts which pass the cutoff is indicated in parentheses. At an FDR of 0.1, differentially decaying transcripts with a simulated standard deviation below 0.05 are identified with a sensitivity of 0.82. (E) We determine the FDR at a given p value cutoff from the simulated data (black dots) and fit a LOESS curve to it to map p values to FDR (blue line). More details on the calibration of p values can be found in the *Methods*.

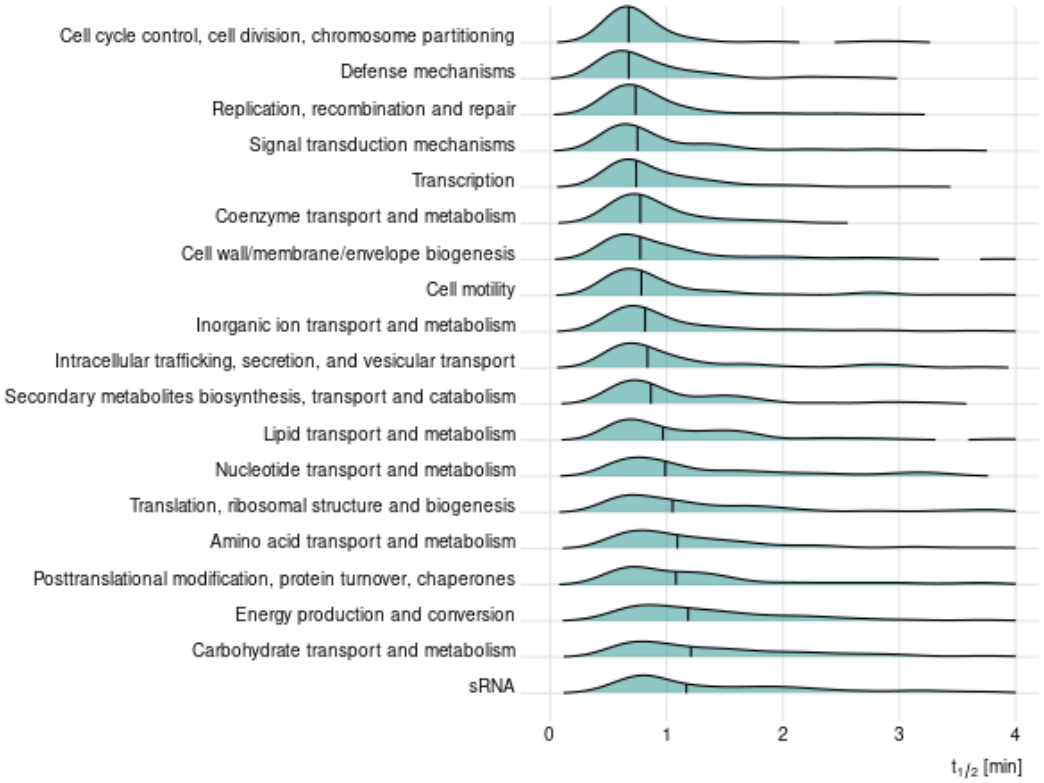


Figure S5. Median half-lives for transcripts classified by COG category. Related to figure 1F.

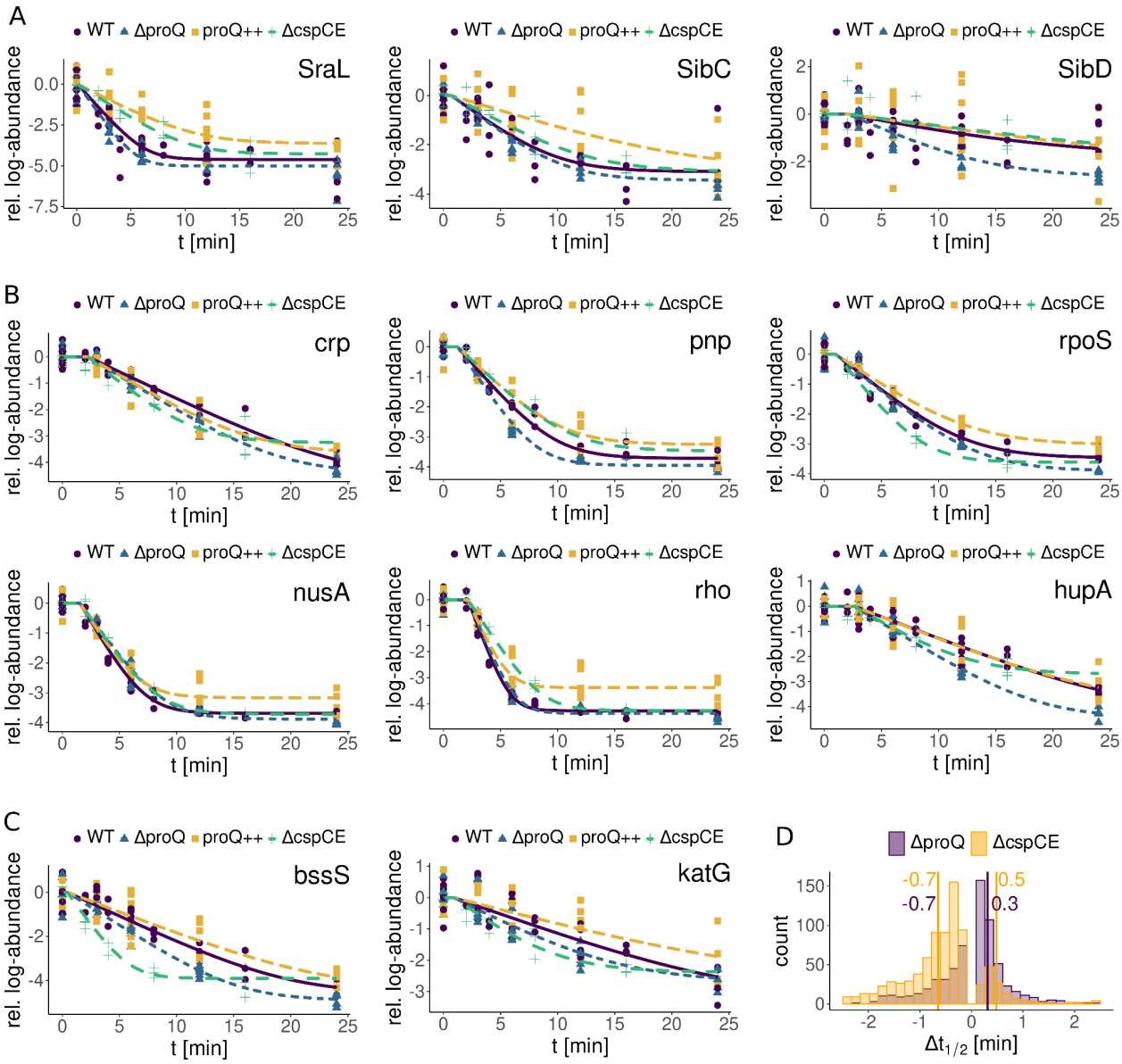


Figure S6. Representative RIF-seq results

(A) Stability changes for known ProQ targets. (B) Stability changes for the exoribonuclease PNPase and global transcriptional regulators. (C) Transcripts with large stability changes upon deletion of *proQ* and *cspCE*. (D) Distribution of significant stability changes in the two RBP deletion mutants.

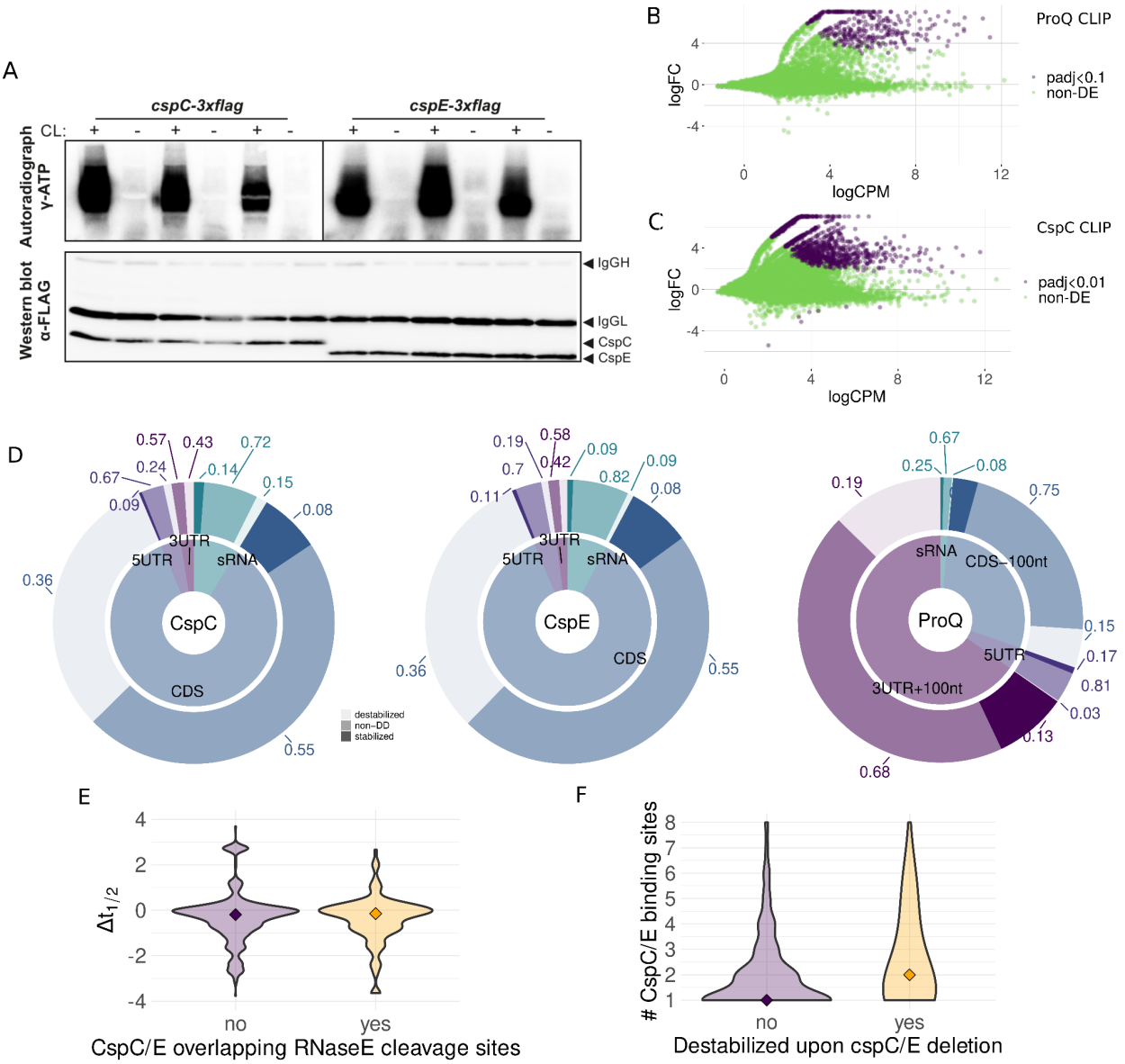


Figure S7. CLIP-seq results and comparison to RIF-seq

(A) Autoradiograph of radioactively labeled RNA fragments covalently bound by CspC/E after UV cross-linking (CL), immunoprecipitation, gel electrophoresis, and membrane transfer. The presence of the RBPs was verified by western blotting. (B,C) MA plots of the CLIP-seq analysis. (D) Fraction of transcripts bound by CspC/E/ProQ which are (de-)stabilized upon *cspC/E* deletion or which do not decay differentially (non-DD). Peaks which overlap with the CDS and the UTR have been assigned to the CDS only. For ProQ, the region within 100 nt of the stop codon has been analyzed jointly with the 3'UTR. (E) Changes in transcript stability of CspC/E targets categorized by whether or not the binding site overlaps a known RNase E cleavage site (39). (F) Number of CspC/E binding sites per transcripts categorized by whether or not the transcript is destabilized upon *cspC/E* deletion.

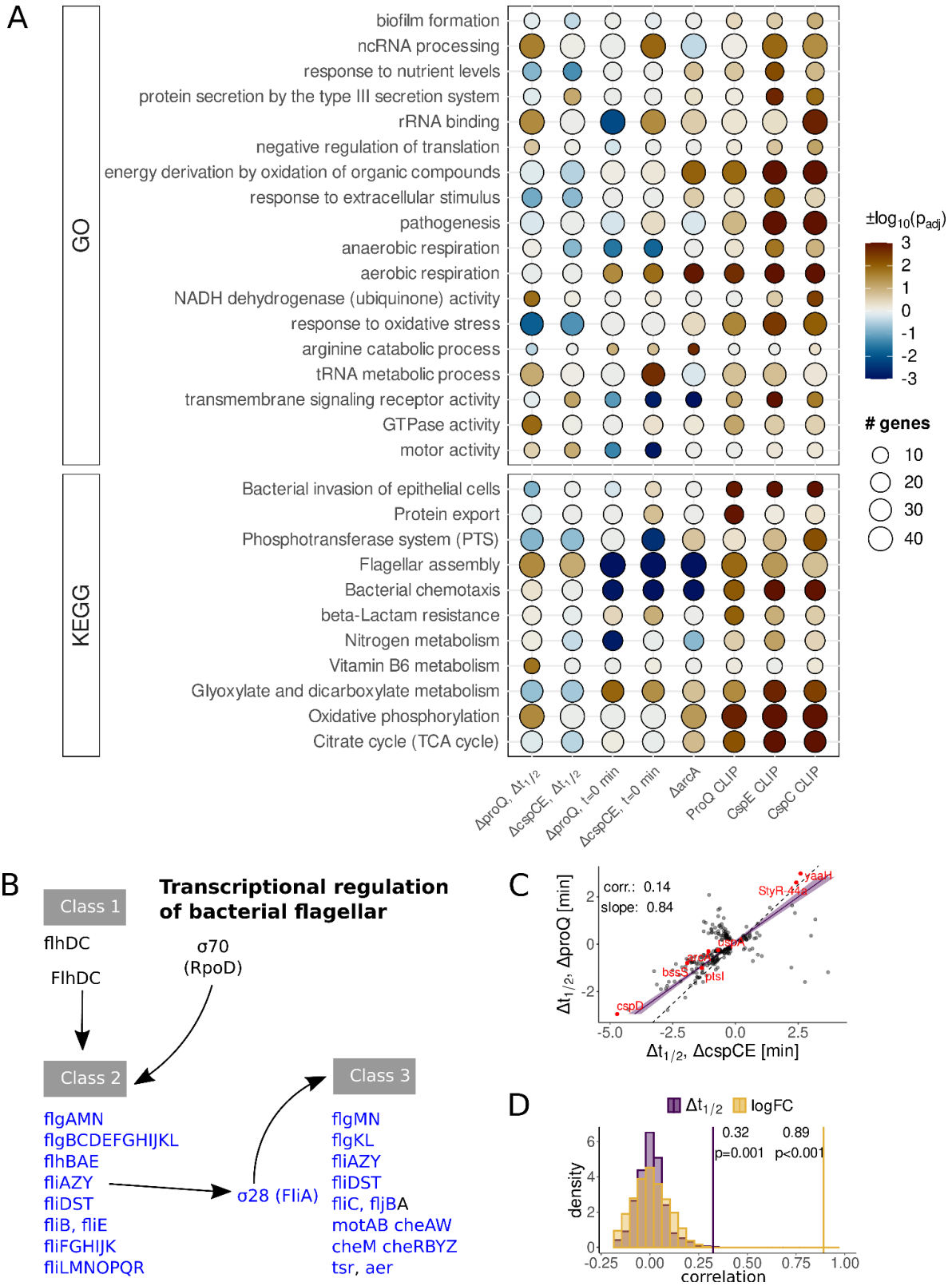


Figure S8. Pathway analyses

(A) Summary of pathway analyses. (B) Transcriptional changes of flagellar genes. Genes highlighted in blue are downregulated upon both *proQ* and *cspC/E* deletion (see **Figure S12** for more details). (C)

Correlation between differences in half-life in the Δ*proQ* and Δ*cspCE* strains. Only transcripts significantly different in both deletion mutants were considered. The linear regression includes only transcripts with changes in the same direction. (D) Significance of correlation between abundance and stability changes in the Δ*proQ* and Δ*cspCE* strains.

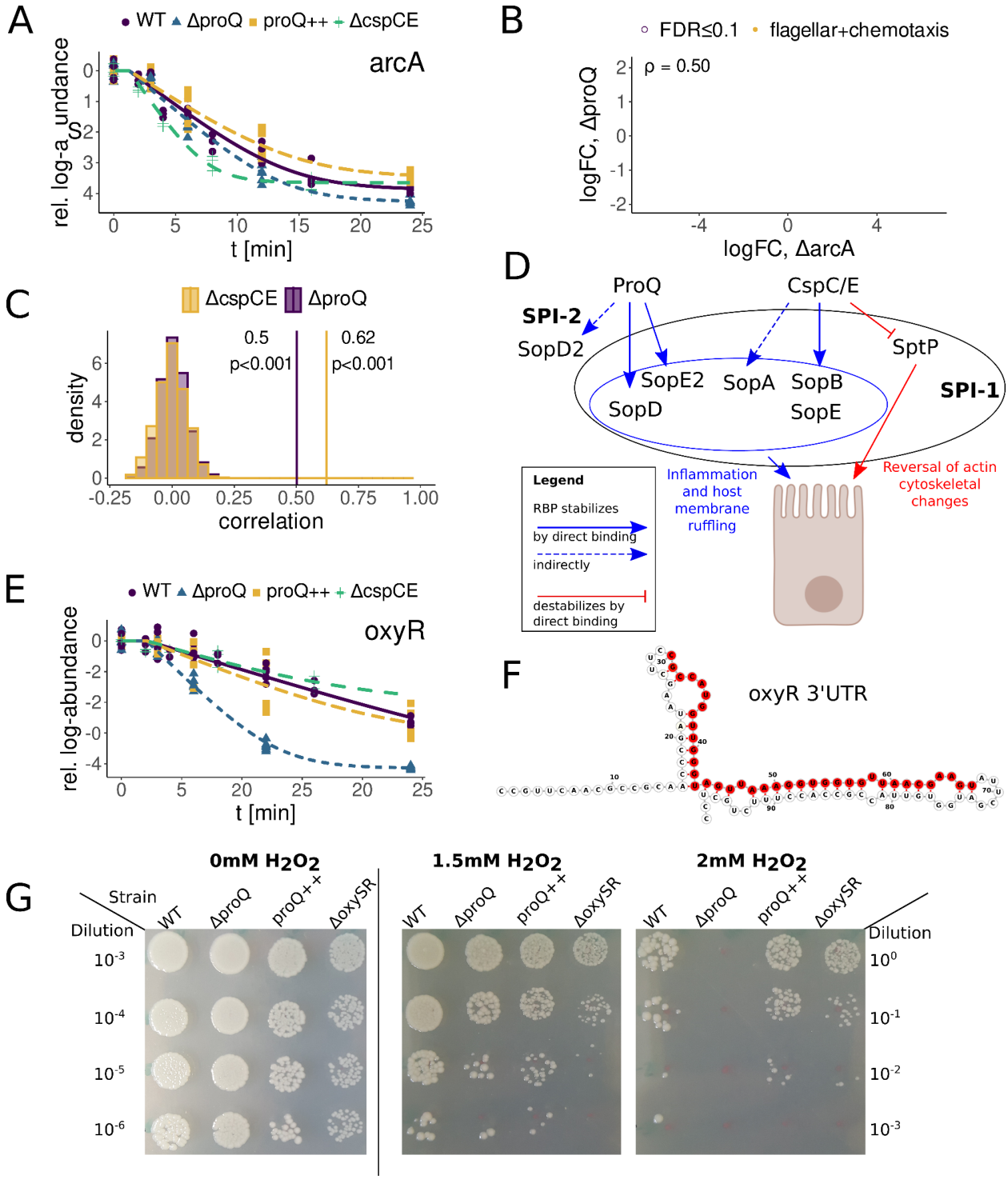


Figure S9. Integrative analysis of RBP binding and transcript stability

(A) Decay curve of significantly destabilized transcript of the *arcA* transcript. (B) log-fold changes in the Δ*proQ* vs. a Δ*arcA* strain (44). (C) Significance of correlation between abundance changes in the Δ*proQ*/Δ*cspCE* and the Δ*arcA* strains. (D) Regulation of SPI-1/2 effectors by ProQ and CspC/E. The picture of the epithelial cell was taken from *BioRender.com*. (E) Decay curve of significantly destabilized transcript of *oxyR*. (F) ProQ CLIP-seq peak in the 3'UTR of *oxyR* identified by re-analyzing (12), FDR=0.047. (G) Exposure of various *Salmonella* strains to varying levels of hydrogen peroxide.

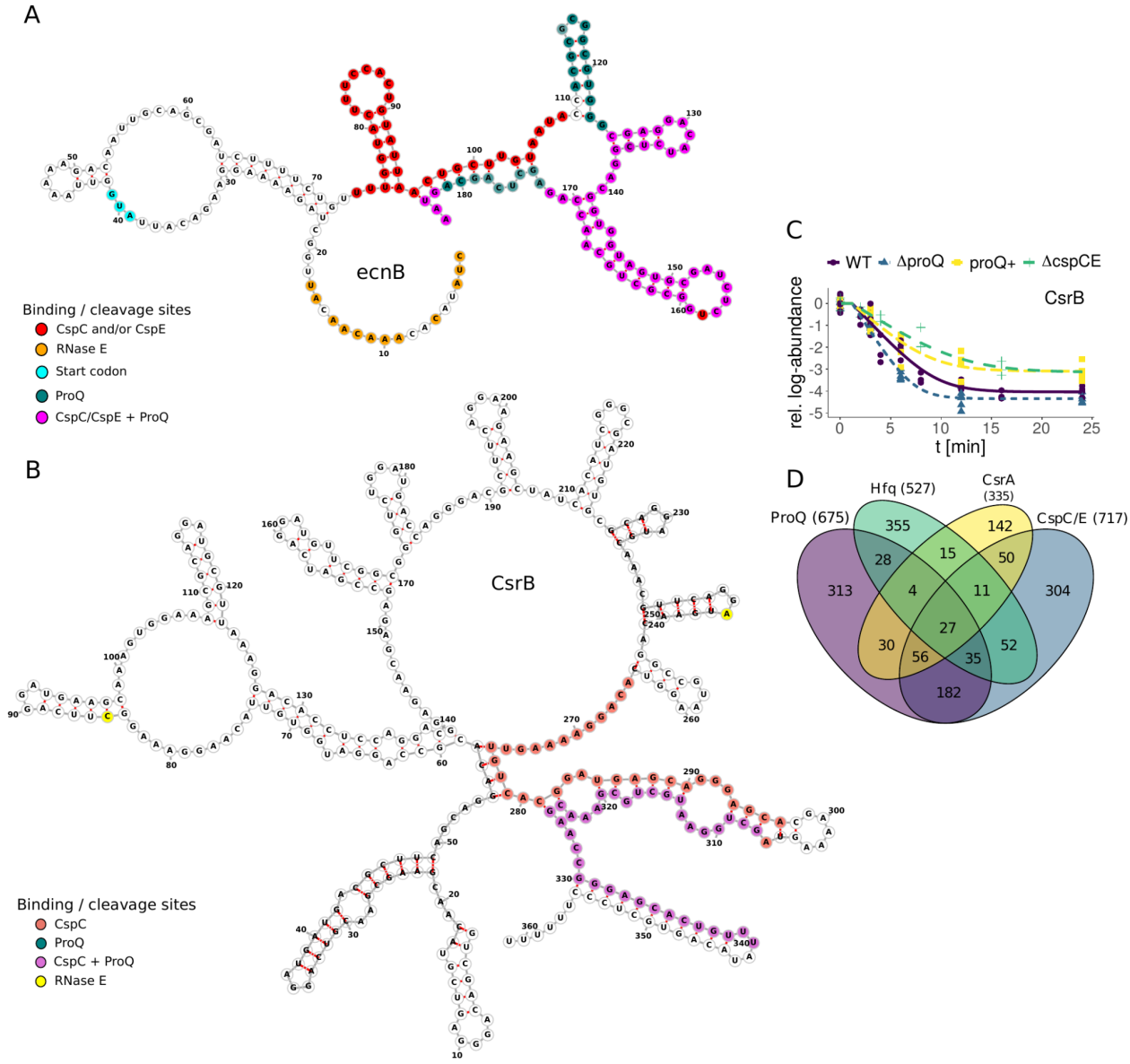
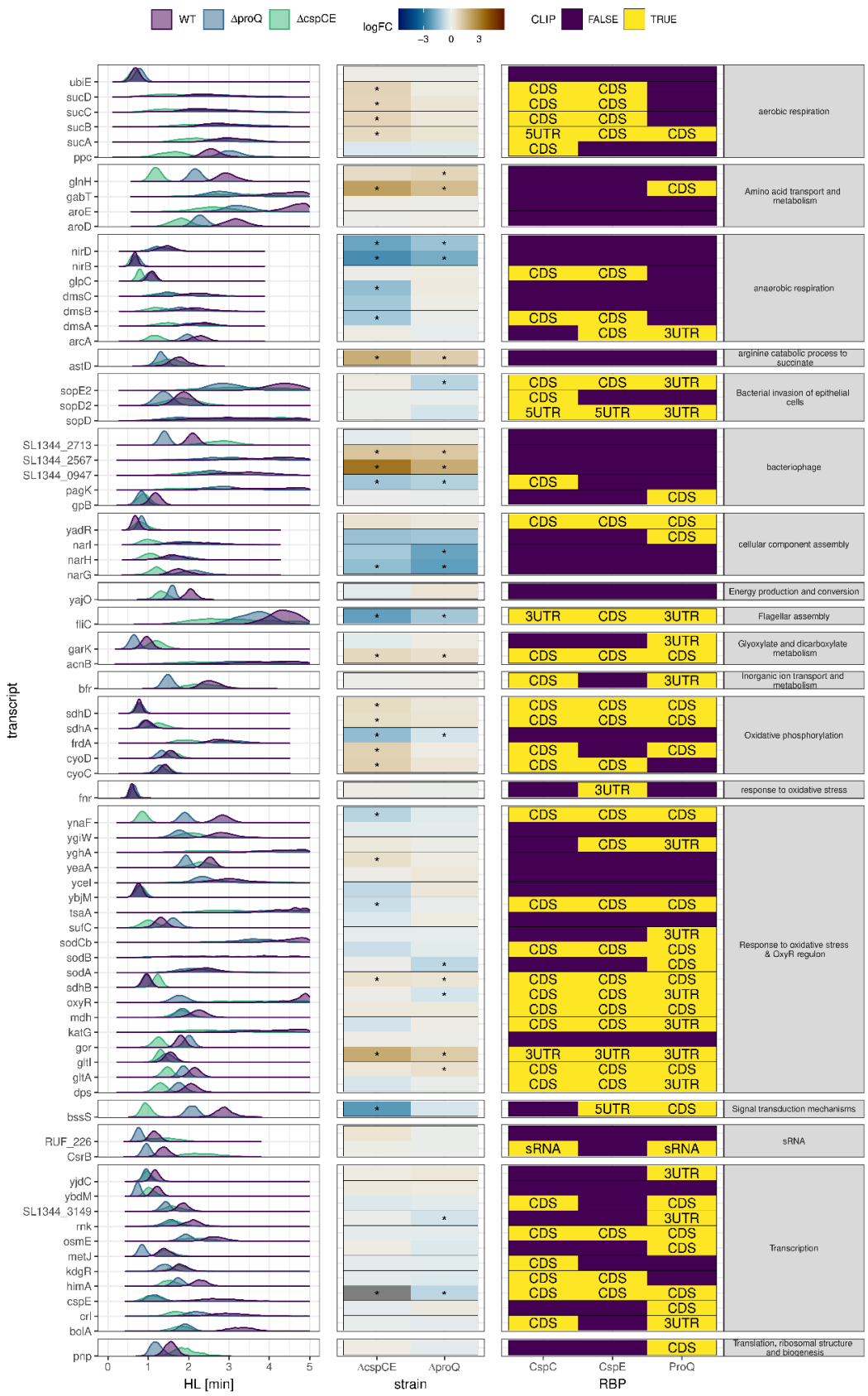


Figure S10. Integrative analysis of RNA secondary structures

(A) Secondary structure of the ProQ/CspC/E-bound mRNA of the bacteriolytic lipoprotein EcnB including RNase E cleavage sites (39). (B) Secondary structure of the ProQ/CspC/E-bound sRNA CsrB including RNase E cleavage sites (39). (C) Normalized data for CsrB, including the fitted decay curves. (D) Overlap in interaction partners between various CLIP-seq data sets with major RBPs.



1057

Figure S11. Top destabilized transcripts in the absence of ProQ and oxidative stress response

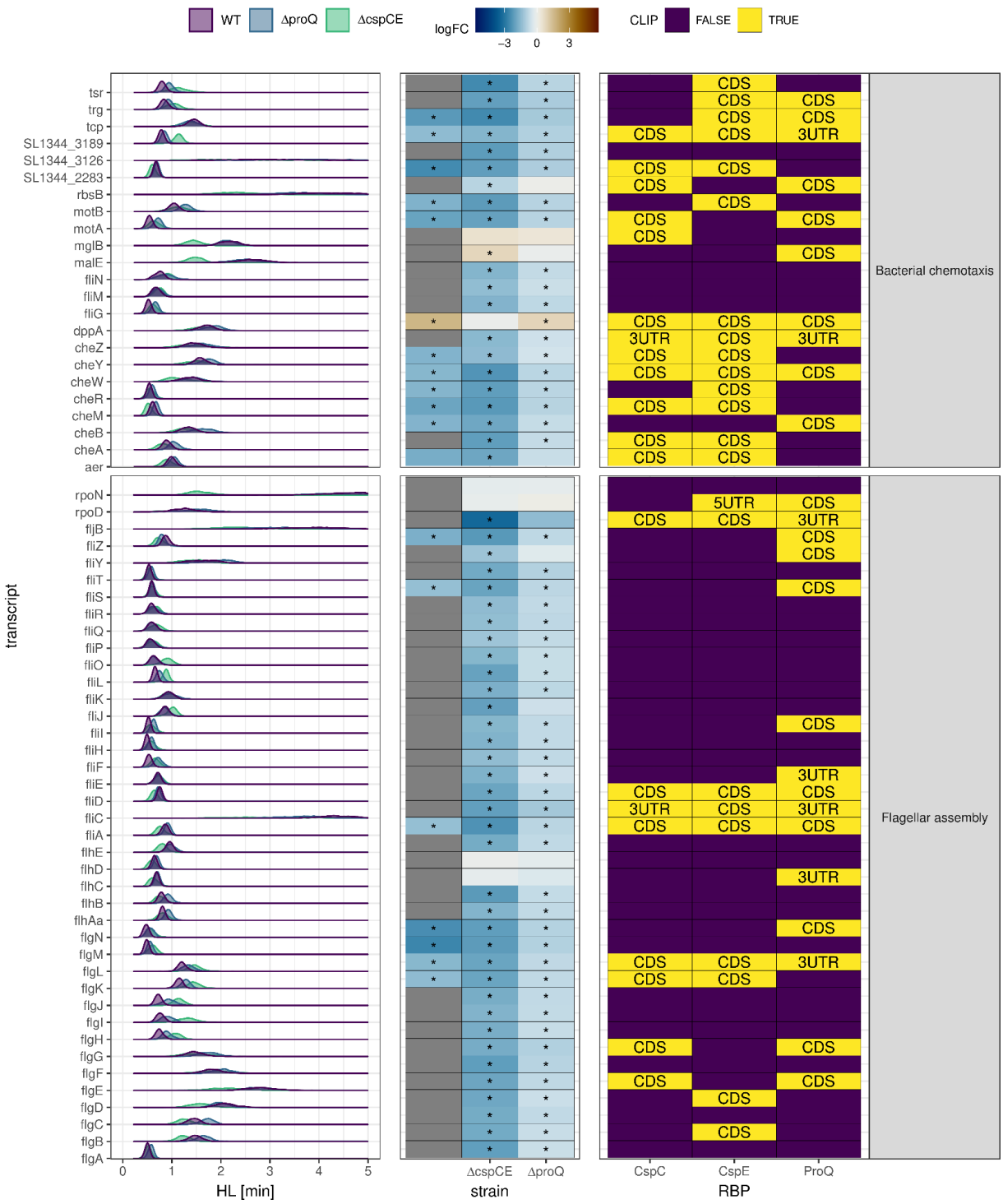


Figure S12. Flagellar genes with negative log-fold change in *proQ* and *cspC/E* deletion strains

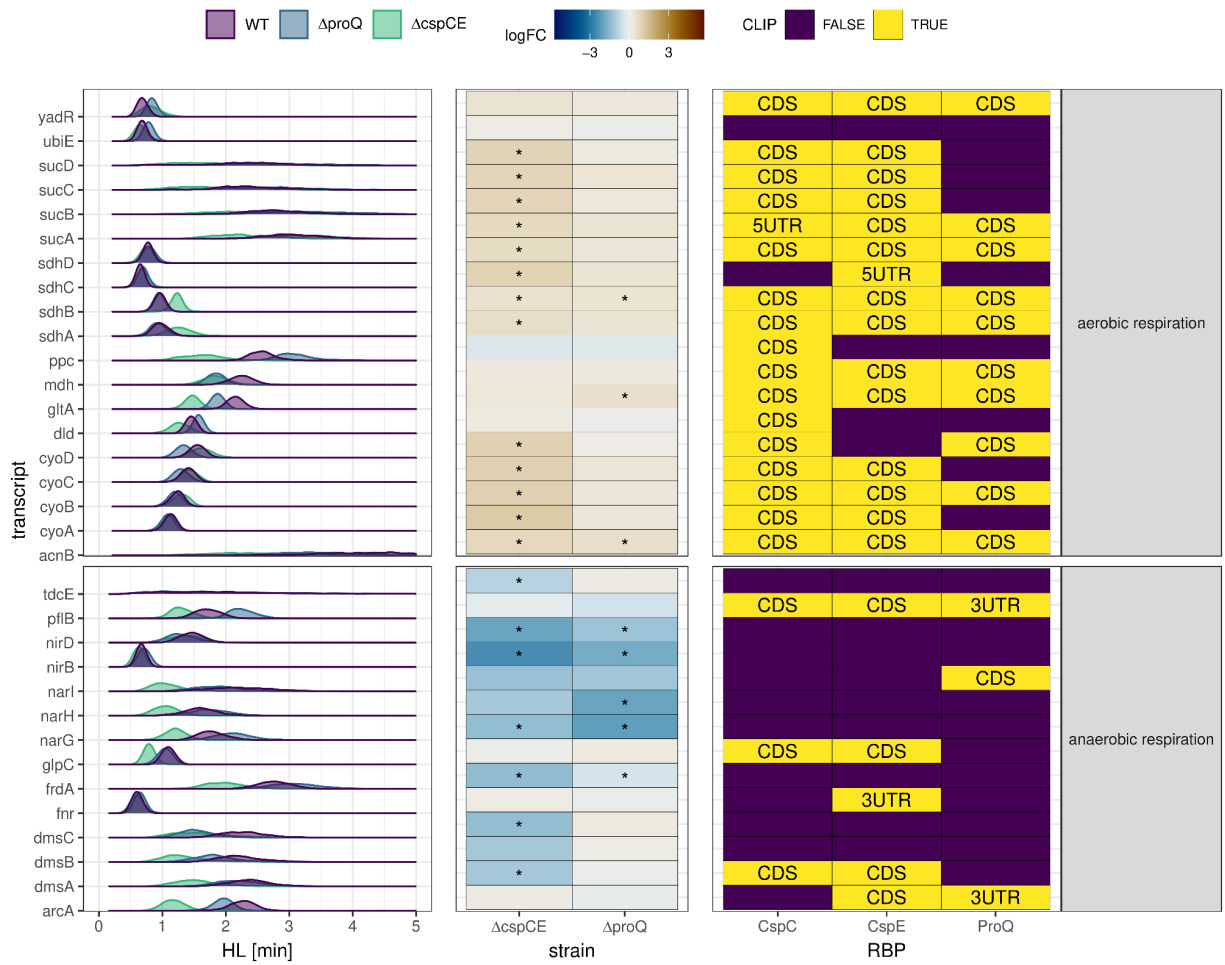


Figure S13. Genes involved in (an)aerobic respiration

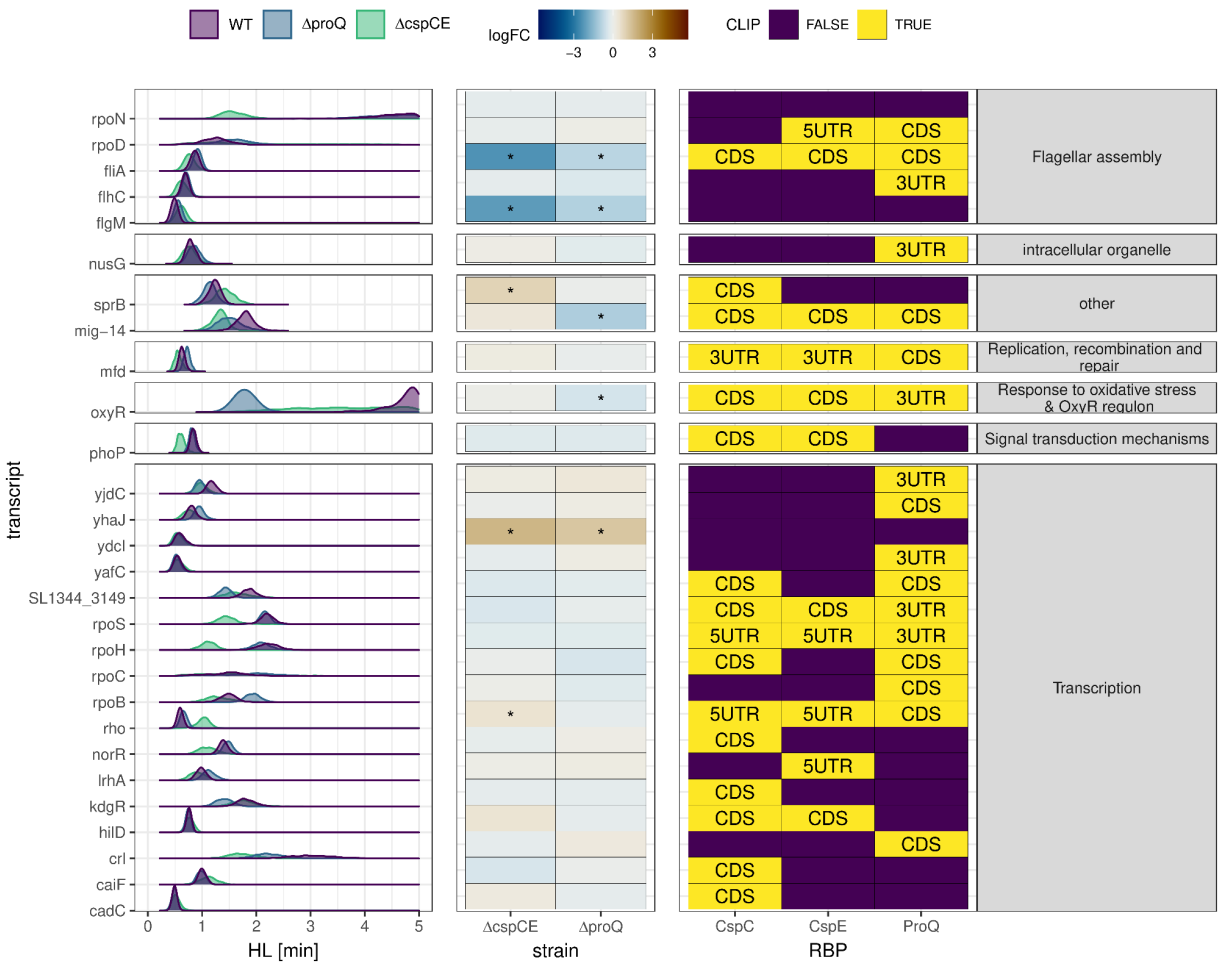


Figure S14. Transcription factors, subunits of RNA polymerase complex

References

1. J. G. Belasco, All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nat. Rev. Mol. Cell Biol.* **11**, 467–478 (2010).
2. M. P. Hui, P. L. Foley, J. G. Belasco, Messenger RNA degradation in bacterial cells. *Annu. Rev. Genet.* **48**, 537–559 (2014).
3. T. B. Updegrove, A. Zhang, G. Storz, Hfq: the flexible RNA matchmaker. *Curr. Opin. Microbiol.* **30**, 133–138 (2016).
4. K. Kavita, F. de Mets, S. Gottesman, New aspects of RNA-based regulation by Hfq and its partner sRNAs. *Curr. Opin. Microbiol.* **42**, 53–61 (2017).
5. T. Romeo, P. Babitzke, Global Regulation by CsrA and Its RNA Antagonists. *Microbiol Spectr* **6** (2018).
6. E. Holmqvist, J. Vogel, RNA-binding proteins in bacteria. *Nat. Rev. Microbiol.* (2018) https:/doi.org/10.1038/s41579-018-0049-5.
7. T. S. Stenum, E. Holmqvist, CsrA enters Hfq's territory: regulation of a base-pairing small RNA. *Mol. Microbiol.* (2021) https:/doi.org/10.1111/mmi.14785.
8. T. S. Stenum, *et al.*, RNA interactome capture in Escherichia coli globally identifies RNA-binding proteins. *Nucleic Acids Res.* (2023) https:/doi.org/10.1093/nar/gkad216.
9. L.-C. Chu, *et al.*, The RNA-bound proteome of MRSA reveals post-transcriptional roles for helix-turn-helix DNA-binding and Rossmann-fold proteins. *Nat. Commun.* **13**, 2883 (2022).
10. E. C. Urdaneta, *et al.*, Purification of cross-linked RNA-protein complexes by phenol-toluol extraction. *Nat. Commun.* **10**, 990 (2019).
11. A. Smirnov, *et al.*, Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11591–11596 (2016).
12. E. Holmqvist, L. Li, T. Bischler, L. Barquist, J. Vogel, Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3' Ends. *Mol. Cell* **70**, 971–982.e6 (2018).
13. S. Melamed, P. P. Adams, A. Zhang, H. Zhang, G. Storz, RNA-RNA Interactomes of ProQ and Hfq Reveal Overlapping and Competing Roles. *Mol. Cell* **77**, 411–425.e7 (2020).
14. A. J. Westermann, *et al.*, The Major RNA-Binding Protein ProQ Impacts Virulence Gene Expression in Salmonella enterica Serovar Typhimurium. *MBio* **10**, e02504–18 (2019).
15. A. Rizvanovic, *et al.*, The RNA-Binding Protein ProQ Promotes Antibiotic Persistence in Salmonella. *MBio*, e0289122 (2022).
16. C. Michaux, *et al.*, RNA target profiles direct the discovery of virulence functions for the cold-shock proteins CspC and CspE. *Proc. Natl. Acad. Sci. U. S. A.*, 201620772 (2017).
17. S. Ray, R. Da Costa, S. Thakur, D. Nandi, Salmonella Typhimurium encoded cold shock protein E is essential for motility and biofilm formation. *Microbiology* **166**, 460–473 (2020).
18. S. Bauriedl, *et al.*, The minimal meningococcal ProQ protein has an intrinsic capacity for structure-based global RNA recognition. *Nat. Commun.* **11**, 2823 (2020).
19. E. A. Campbell, *et al.*, Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* **104**, 901–912 (2001).
20. J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, S. N. Cohen, 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 (2002).
21. D. W. Selinger, R. M. Saxena, K. J. Cheung, G. M. Church, C. Rosenow, Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. *Genome Res.* **13**, 216–223 (2003).
22. H. Chen, K. Shiroguchi, H. Ge, X. S. Xie, Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Mol. Syst. Biol.* **11**, 781 (2015).
23. M. E. Ritchie, *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
24. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
25. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
26. I. Lönnstedt, T. Speed, REPLICATED MICROARRAY DATA. *Stat. Sin.* **12**, 31–46 (2002).

27. B. Carpenter, *et al.*, Stan: A probabilistic programming language. *J. Stat. Softw.* **76** (2017).

28. C. A. Lee, S. Falkow, The ability of Salmonella to enter mammalian cells is affected by bacterial growth state. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4304–4308 (1990).

29. L. Jiang, *et al.*, Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551 (2011).

30. R. D. Mosteller, C. Yanofsky, Transcription of the tryptophan operon in Escherichia coli: rifampicin as an inhibitor of initiation. *J. Mol. Biol.* **48**, 525–531 (1970).

31. D. Dar, R. Sorek, Bacterial Noncoding RNAs Excised from within Protein-Coding Transcripts. *MBio* **9** (2018).

32. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).

33. M. Y. Galperin, *et al.*, COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **49**, D274–D281 (2021).

34. A. Vehtari, J. Ojanen, A survey of Bayesian predictive methods for model assessment, selection and comparison. *Stat. Surv.* **6**, 142–228 (2012).

35. A. Gelman, Two simple examples for understanding posterior p-values whose distributions are far from uniform. *EJSS* **7**, 2595–2602 (2013).

36. Y. Yair, *et al.*, Cellular RNA Targets of Cold Shock Proteins CspC and CspE and Their Importance for Serum Resistance in Septicemic Escherichia coli. *mSystems* **7**, e0008622 (2022).

37. M. Raffatellu, *et al.*, SipA, SopA, SopB, SopD, and SopE2 contribute to Salmonella enterica serotype typhimurium invasion of epithelial cells. *Infect. Immun.* **73**, 146–154 (2005).

38. J. A. Imlay, Transcription Factors That Defend Bacteria Against Reactive Oxygen Species. *Annu. Rev. Microbiol.* **69**, 93–108 (2015).

39. Y. Chao, *et al.*, In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Mol. Cell* **65**, 39–51 (2017).

40. J. Richards, J. G. Belasco, Obstacles to Scanning by RNase E Govern Bacterial mRNA Lifetimes by Hindering Access to Distal Cleavage Sites. *Mol. Cell* (2019) https:/doi.org/10.1016/j.molcel.2019.01.044.

41. J. Richards, J. G. Belasco, Graded impact of obstacle size on scanning by RNase E. *Nucleic Acids Res.* **51**, 1364–1374 (2023).

42. E. Holmqvist, *et al.*, Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *EMBO J.* **35**, 991–1011 (2016).

43. A. Subramanian, *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

44. M. R. Evans, *et al.*, Analysis of the ArcA regulon in anaerobically grown Salmonella enterica sv. Typhimurium. *BMC Microbiol.* **11**, 58 (2011).

45. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1102 (2010).

46. T. Esquerré, *et al.*, Dual role of transcription and transcript stability in the regulation of gene expression in Escherichia coli cells cultured on glucose at different growth rates. *Nucleic Acids Res.* **42**, 2460–2472 (2014).

47. A. H. Potts, Y. Guo, B. M. M. Ahmer, T. Romeo, Role of CsrA in stress responses and metabolism important for Salmonella virulence revealed by integrated transcriptomics. *PLoS One* **14**, e0211430 (2019).

48. Morin Manon, Enjalbert Brice, Ropers Delphine, Girbal Laurence, Cocaign-Bousquet Muriel, Genomewide Stabilization of mRNA during a “Feast-to-Famine” Growth Transition in Escherichia coli. *mSphere* **5**, e00276–20 (2020).

49. E. Baracchini, H. Bremer, Determination of synthesis rate and lifetime of bacterial mRNAs. *Anal. Biochem.* **167**, 245–260 (1987).

50. D. A. Vargas-Blanco, S. S. Shell, Regulation of mRNA Stability During Bacterial Stress Responses. *Front. Microbiol.* **11**, 2111 (2020).

51. G. Yim, F. de la Cruz, G. B. Spiegelman, J. Davies, Transcription modulation of Salmonella enterica serovar Typhimurium promoters by sub-MIC levels of rifampin. *J. Bacteriol.* **188**, 7988–7991 (2006).

52. E. Massé, F. E. Escorcia, S. Gottesman, Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes Dev.* **17**, 2374–2383 (2003).

53. Y. Zhang, *et al.*, A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation. *Mol. Cell* (2018) https:/doi.org/10.1016/j.molcel.2018.02.035.

54. W. Bae, B. Xia, M. Inouye, K. Severinov, Escherichia coli CspA-family RNA chaperones are transcription antiterminators. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7784–7789 (2000).

55. A. Smirnov, C. Wang, L. L. Drewry, J. Vogel, Molecular mechanism of mRNA repression intransby a ProQ-dependent small RNA. *EMBO J.* **36**, 1029–1045 (2017).

56. R. Balakrishnan, *et al.*, Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria. *Science* **378**, eabk2066 (2022).

57. Y. Chao, J. Vogel, The role of Hfq in bacterial pathogens. *Curr. Opin. Microbiol.* **13**, 24–33 (2010).

58. M. G. Jørgensen, M. K. Thomason, J. Havelund, P. Valentin-Hansen, G. Storz, Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev.* **27**, 1132–1145 (2013).

59. Y.-J. Lai, *et al.*, CsrA regulation via binding to the base-pairing small RNA Spot 42. *Mol. Microbiol.* **117**, 32–53 (2022).

60. B. Xia, H. Ke, M. Inouye, Acquirement of cold sensitivity by quadruple deletion of the cspA family and its suppression by PNPase S1 domain in Escherichia coli. *Mol. Microbiol.* **40**, 179–188 (2001).

61. B. A. Stocker, S. K. Hoiseth, B. P. Smith, Aromatic-dependent “Salmonella sp.” as live vaccine in mice and calves. *Dev. Biol. Stand.* **53**, 47–54 (1983).

62. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).

63. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

64. C. Kröger, *et al.*, An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* **14**, 683–695 (2013).

65. P. P. Gardner, L. Barquist, A. Bateman, E. P. Nawrocki, Z. Weinberg, RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res.* **39**, 5845–5852 (2011).

66. D. Risso, J. Ngai, T. P. Speed, S. Dudoit, Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).

67. A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neuböck, I. L. Hofacker, The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–4 (2008).

68. P. Kerpedjiev, S. Hammer, I. L. Hofacker, Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**, 3377–3379 (2015).

69. J. Huerta-Cepas, *et al.*, eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).

70. D. Binns, *et al.*, QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).

71. M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–62 (2016).

72. M. Lechner, *et al.*, Orthology detection combining clustering and synteny for very large datasets. *PLoS One* **9**, e105015 (2014).

73. H. Xu, *et al.*, FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One* **7**, e52249 (2012).

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

75. S. Hoffmann, *et al.*, A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol.* **15**, R34 (2014).

76. A. Zhu, J. G. Ibrahim, M. I. Love, Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2018).