TaRTLEt: Transcriptionally-active Riboswitch Tracer Leveraging Edge deTection

Sachit Kshatriya[1] **and Sarah C. Bagby**[1]

[1]**Case Western Reserve University Department of Biology**

Corresponding author:

Sarah C. Bagby[1]

Email address: sarah.bagby@case.edu

ABSTRACT

Structured RNAs have emerged as a major component of cellular regulatory systems, but their mechanism of action is often poorly understood. Riboswitches are structured RNAs that allosterically regulate gene expression through any of several different mechanisms. *In vitro* approaches to characterizing mechanism are costly, low-throughput, and must be repeated for each individual riboswitch locus of interest. Bioinformatic methods promise higher throughput; despite robust computational identification of riboswitches, however, computational classification of riboswitch mechanism has so far been both model-bound, relying on identification of sequence motifs known to be required for specific models of riboswitch activity, and empirically untested, with predictions far outpacing biological validation. Here, we introduce `TaRTLEt` (Transcriptionally-active Riboswitch Tracer Leveraging Edge deTection), a new high-throughput tool that recovers *in vivo* patterns of riboswitch-mediated transcription termination from paired-end RNA-seq data using edge detection methods. `TaRTLEt` successfully extracts transcription termination signals despite numerous sources of biological and technical noise. We tested the effectiveness of `TaRTLEt` on riboswitches identified from a wide range of sequenced bacterial taxa by utilizing publicly available paired-end RNA-seq readsets, finding broad agreement with previously published *in vitro* characterization results. In addition, we use `TaRTLEt` to infer the *in vivo* regulatory mechanism of uncharacterized riboswitch loci from existing public data. `TaRTLEt` is available on GitHub and can be applied to paired-end RNA-seq datasets from isolates or complex communities.

INTRODUCTION

Riboswitches are highly conserved, widespread gene regulatory elements that exploit RNA's potential for complex structure to build ligand-responsive differential regulation of gene expression directly into an mRNA. Encoded in the untranslated regions (UTRs) of genes, riboswitches are highly specific aptamers for a wide variety of small-molecule ligands. In the two decades since the discovery of the first riboswitch (Mironov et al., 2002), we have learned that riboswitches are compact, specific, and responsive to transient changes; that they often regulate critical organism functions; and that they operate via diverse and complex modes (Mandal et al., 2003; Amadei et al., 2023; Sudarsan et al., 2005). Given these properties, riboswitches have been investigated for uses ranging from synthetic sensors (Wachsmuth et al., 2013; Boussebayle et al., 2019), to logic elements in complex gene expression systems (Hanson et al., 2003; Groher and Suess, 2014), to potential targets for antimicrobials and novel therapeutics (Blount and Breaker, 2006; Ellinger et al., 2023; Giarimoglou et al., 2022). Yet our understanding of riboswitch function *in vivo* remains patchy.

Biochemical methods such as transcript fragment length and ribosome binding assays (Winkler et al., 2003; Nou and Kadner, 2000; Welz and Breaker, 2007; Hollands et al., 2012) have identified an overall

pattern of riboswitch regulation. Riboswitch-mediated allosteric regulation of expression begins with selective binding of a ligand to the non-coding aptamer domain, stabilizing the domain's conformation (Gilbert et al., 2006) and altering the structure of the expression platform. Structural changes in the expression domain can sequester or expose conventional expression control logic elements such as intrinsic terminators/antiterminators (Mironov et al., 2002; Mandal and Breaker, 2004), rho-binding sites (Hollands et al., 2012), and translation start sites (Breaker, 2018). However, riboswitch mechanisms can occasionally go beyond interaction with direct control architectures. For example, TPP riboswitches in the fungus *Neurospora crassa* alter gene expression by controlling mRNA splicing (Cheah et al., 2007), the *glmS* riboswitch-ribozyme in *Bacillus subtilis* recruits exonucleases for degradation by catalyzing upstream mRNA cleavage (Collins et al., 2007; Klein and Ferré-D'Amaré, 2006), the *Escherichia coli lysC* riboswitch both sequesters the Shine-Dalgarno (SD) sequence and exposes an RNase E binding site (Caron et al., 2012), and a SAM riboswitch in *Listeria monocytogenes* is capable of acting in *trans* on a distal target (Loh et al., 2009). Numerous other examples have been documented (Bédard et al., 2020; Ariza-Mateos et al., 2021), and the full range of riboswitch regulatory mechanisms remains unknown.

A large part of contextualizing riboswitch function within a given microbe is illuminating these modes of regulation. Advances in computational techniques have enabled high-confidence identification of candidate riboswitch sequences (Nawrocki, 2014; Nawrocki and Eddy, 2013a; Chang et al., 2009). However, because a single ligand-binding class may include riboswitches that act by different mechanisms (Barrick and Breaker, 2007), even robust identification of riboswitches and their ligands (Eddy and Durbin, 1994; Yao et al., 2007; Stav et al., 2019) is not in itself sufficient to predict riboswitch mechanism or effects in vivo. To fill this gap, in silico methods have largely relied on the computation of folding energies (Gong et al., 2017) coupled with the identification of SD sequences, U-rich terminator motifs, or rho-binding sites to determine states that could alter transcription/translation efficiency to predict mechanism (Barrick and Breaker, 2007; Sun and Rodionov, 2014). While these analytical methods can produce *a priori* predictions about riboswitch regulatory modes, they leave unexplored a more direct readout of the realized activity of riboswitches within living cells: the distribution of fragments captured in RNA-seq data.

Here, we describe the new tool `TaRTLEt` (Transcriptionally-active Riboswitch Tracer Leveraging Edge deTection), which implements a method to identify riboswitches that operate via alteration of transcription termination efficiency using (a) high-throughput computational approaches and (b) publicly available RNA-seq data. We hypothesized that different experimental conditions would, directly or indirectly, evoke changes in riboswitch ligand concentrations that might alter riboswitch regulatory state. Under this hypothesis, we predicted that, when the range of conditions tested included both above- and below-threshold ligand concentrations, transcription-attenuating riboswitches would produce a distinctive coverage pattern in paired-end RNA-seq datasets that could be robustly identified using computational approaches developed for edge detection (Canny, 1986). We show that the occurrence of this pattern in publicly available transcriptomic datasets is in good agreement with existing *in vitro* riboswitch characterization. Our approach complements and extends earlier methods based on inference from DNA sequence by extracting information from the traces of riboswitch activity that are left in the RNA pool of cells actively regulating gene expression.

MATERIALS & METHODS

Theoretical framework

Predicted transcript pools for transcriptionally active riboswitches

Riboswitches are often associated with transcription start sites (TSS), allowing transcription start site analysis to serve as a strategy for identifying novel riboswitches and regulatory RNA (Rosinski-Chupin et al., 2014, 2019; Adams et al., 2017; Yu et al., 2018). We reasoned that a transcription-terminating riboswitch in the 5′ UTR of a gene should produce different patterns of RNA-seq read coverage in the ON vs. OFF state (Figure 1). In either case, transcription should begin at the TSS and proceed through the riboswitch itself. In the OFF state, formation of the transcription-terminator structure should truncate transcripts near the riboswitch 3′ end (Figure 1 top). By contrast, in the ON state, RNA polymerase should read through, producing transcripts that extend well into the downstream gene (Figure 1 bottom). A riboswitch that produces these two patterns under different conditions exhibits behavior consistent with riboswitch-mediated conditional regulation of transcription termination.

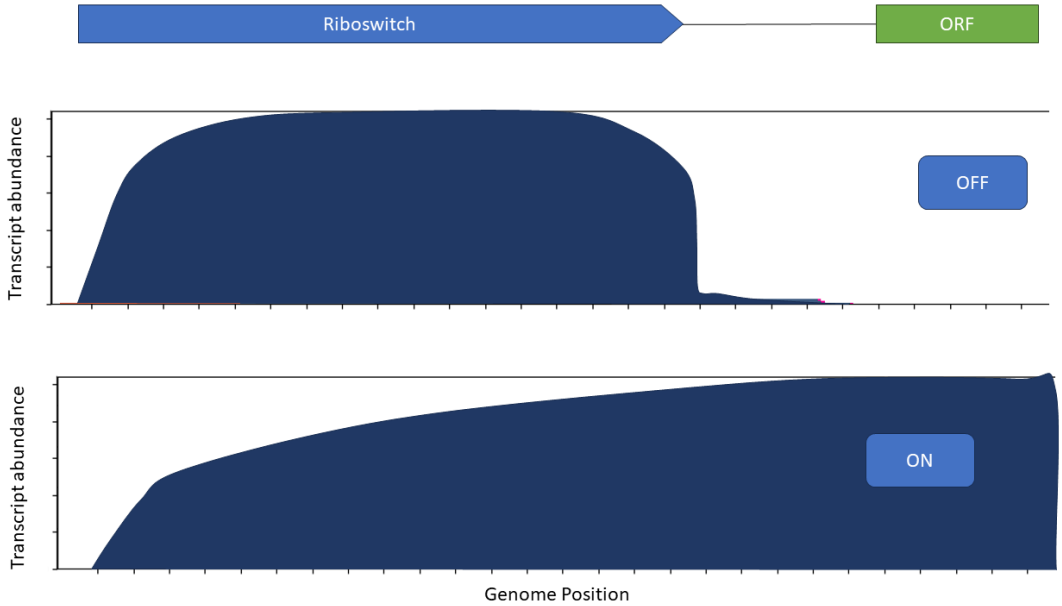

Figure 1. Conceptual coverage map for RNA-seq alignment to a riboswitch transcriptionally regulating a downstream open reading frame. In the OFF conformation, the riboswitch truncates transcripts. In the ON conformation, the riboswitch allows transcription through the start of the ORF.

Fragment distributions and terminal pileups

What relation does the coverage map of sequencing reads bear to the pool of transcripts? Each RNA-seq read represents some region of a transcript, but because transcripts are typically fragmented prior to sequencing we know only that the $5'$ end of a read must align with a fragment terminus and not whether that fragment is internal or terminal to a transcript. Fragmentation often, as in the validation dataset used here, results in a random biased distribution (Fig. S1) of fragment sizes. Importantly, however, the new $3'$ and $5'$ termini produced by each fragmentation event come in neighboring pairs. Even though the total number of fragment termini (at most 2^i arising from i fragmentations), and thus the total number of read termini, may far exceed the number of original transcript termini, we can still extract a transcript size signal by considering pileups of read termini. Because paired-end sequencing generates reads that represent both ends of each sequenced fragment, the resulting data offers a higher signal:noise ratio than single-end sequencing: there, fragmentation is more extensive *and* at most one terminus of each fragment is represented in the read set, so that comparable sequencing depth yields greater stochastic fluctuations in the set of termini captured. We therefore designed `TaRTLEt` v1.0 to restrict analysis to paired-end datasets. Future versions could be extended to analyze appropriately deep single-end datasets.

We reasoned that a convolution approach would allow us to both extract the needed transcript size signal and smooth the overall signal. Convolution is heavily used in applications like edge-detection algorithms (Canny, 1986; Maini and Aggarwal, 2009; Albawi et al., 2017), where it can act as an attenuating filter for neighboring signals. Here, we want neighboring pileups of $3'$ and $5'$ read termini to drop out, filtering out transcript-internal fragmentation out to reveal the transcript-terminal signal. Tracking $3'$ ends as $+1$ and $5'$ ends as -1, we can first compute the signed sum of read termini for each position in the reference (**Fig. 2a**), then calculate a weighted average of surrounding sums for each position (**Fig. 2b**). We call the resulting convolution's features "peaks", and we seek to identify peaks at which a pileup of $3'$ termini is not accompanied by a pileup of $5'$ termini immediately downstream.

The choice of kernel function for computation of the weighted average affects the filter's selectivity for changes in coverage. Rather than a flat uniform distribution, we chose to use kernels constructed by discretizing a Gaussian distribution, such that tuning the spread of this distribution (by the standard deviation σ) tunes the filter's selectivity for how sharply coverage must change. Because sites of transcription termination, especially intrinsic termination, are constrained (Roberts, 2019; Gusarov and

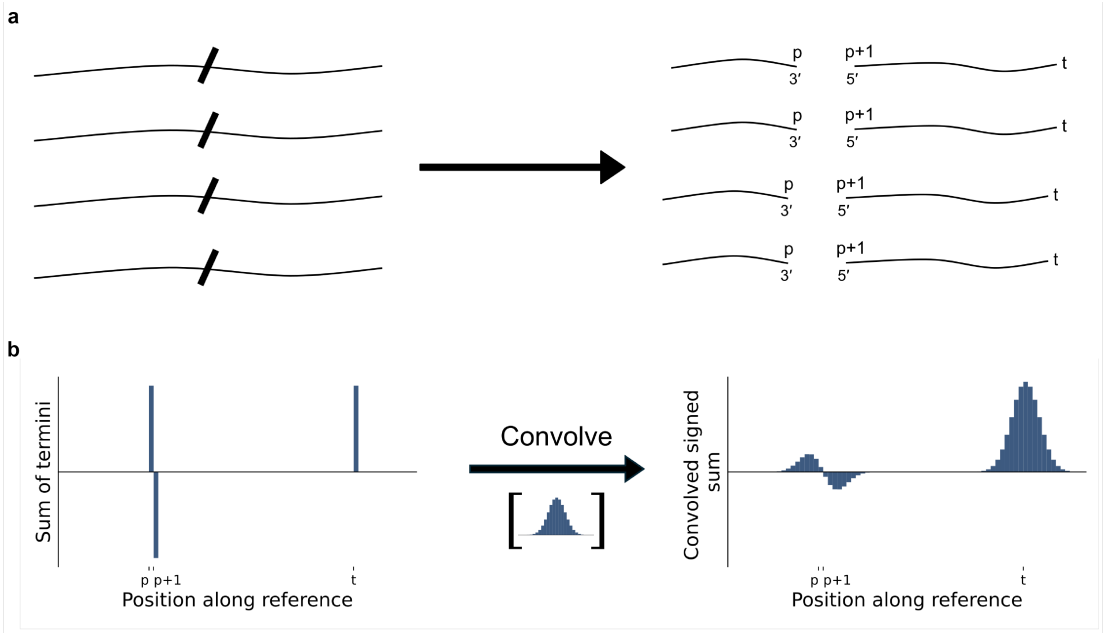

Figure 2. (a) Illustrative example showing the generation of new neighboring 3′ and 5′ termini at positions p and $p+1$ during transcript fragmentation along with the original transcript terminus at position t. (b) Application of the weighted convolution to signed pileups of new and original fragment termini resulting in the attenuation of the signal from new fragment termini.

Nudler, 1999), riboswitch-mediated transcription termination should produce a sharp drop in coverage. To match this expectation, we might use a tight convolution (small σ) kernel that attenuates peaks arising from changes over >1–2 nt. However, both biological and technical factors (e.g., thermodynamic fluctuations; processes in library preparation, sequencing, and quality control) are expected to broaden the biological signal to some degree. Through trial and error, we chose a default σ of 1.5 for kernel generation, corresponding to a filter that passes coverage changes localized within a 6-nt region.

Regardless of the number of fragmentation events, the convolution always returns non-zero peaks at positions with asymmetric pileups of 5′ or 3′ termini. These surviving peaks should represent a range of events: transcription termination should be relatively rare and should cause a large, tightly centered coverage drop, giving rise to high, narrow peaks; sporadic processes (e.g., artifacts in library prep or sequencing) should be relatively common and yield lower, broader peaks. While these differences in peak shapes should provide an opportunity to further refine the set of candidate peaks, the probability and details of the processes generating background peaks may be sequence-dependent in as yet poorly understood ways, complicating comparison to a global baseline. Instead, we reasoned that we could collect a per-locus "noise set" of peaks within a defined neighborhood scaled to riboswitch size. Fitting a 2-dimensional Gaussian noise distribution to this collection of peaks provides a basis for comparison, allowing us to robustly identify unusually sharp peaks at each locus as candidate sites of riboswitch-mediated transcription termination.

Probing fragment-end distributions across treatments

For any of these sites to reveal conditional regulation by a transcription-terminating riboswitch, the efficiency of termination (that is, the proportion of transcripts that terminate) near the riboswitch 3′ end must vary across treatments that evoke ON and OFF states. For a riboswitch that acts by transcription termination, termination efficiency should be low (low or no coverage drop) in the ON state and high (large coverage drop) in the OFF state. We bound the region of interest for this search to a range of 0.3×riboswitch size both upstream and downstream of the riboswitch 3′ end and examine the fractional coverage change (that is, (change in coverage across peak) / (average coverage depth across entire riboswitch)) at each convolution peak in the region under each treatment condition. To compare results at coincident peaks across treatment conditions, we account for the possibility that peak location might shift

slightly due to small variations in process noise by clustering peak locations across treatment conditions. Where the mean fractional coverage change is significantly stronger and the variance of the fractional coverage change is significantly larger than baseline (**Fig. 3**), we infer that the cluster captures both ON and OFF states of a transcription-terminating riboswitch.

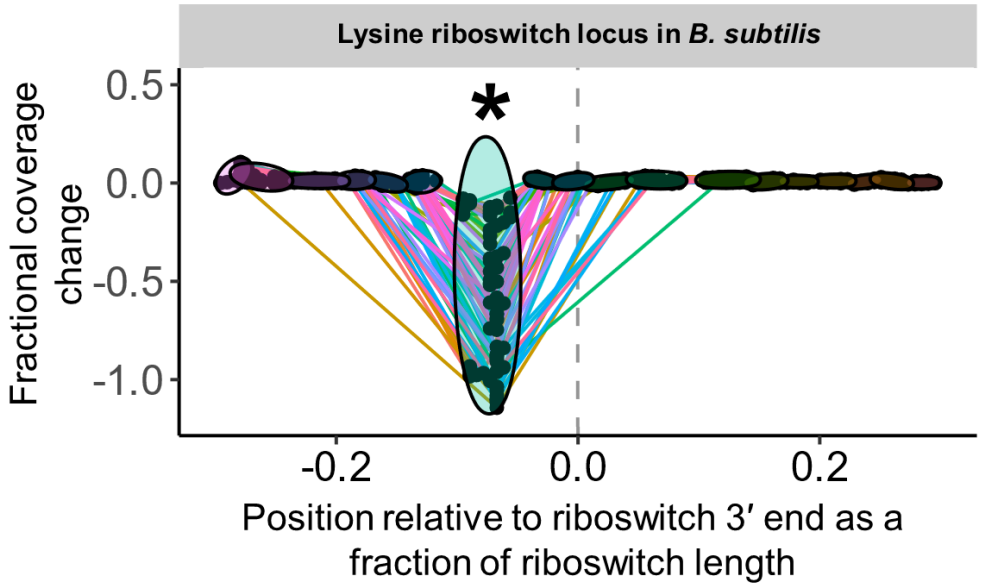

Figure 3. Evidence of conditional transcription termination at a riboswitch 3′. Points represent peaks detected under different conditions; within-condition points are connected by colored lines. Peak locations are shown relative to riboswitch size and 3′ end; that is, a peak at 0.1 on the x-axis lies 10% of the riboswitch length downstream of the riboswitch 3′ end. Ellipses bound peaks clustered together based on their x-positions. The star marks a cluster whose mean coverage drop is significantly larger and whose variance is significantly higher than other clusters'.

`TaRTLEt` Implementation

`TaRTLEt` is implemented in `python` v3.11 with a dedicated command-line interface (CLI). Its usage does not require knowledge of python and the included `conda` environment contains installations for any external CLI tools. For widespread utility, we chose to incorporate a riboswitch identification pipeline into our approach, so that inputs may be unannotated genomes/sequences; identified riboswitch loci are then submitted to the mechanistic inference pipeline. A complete walk-through of `TaRTLEt` usage and the analysis of data generated by `TaRTLEt` is available at the tool GitHub repository (`https://github.com/lepton-7/tartlet`).

Identification of riboswitches and open reading frames

Input sequences in FASTA format are processed through `Infernal` v1.1.5 (`http://eddylab.org/infernal`) (Nawrocki and Eddy, 2013b) and `Prodigal` v2.6.3 (`https://github.com/hyattpd/Prodigal`) (Hyatt et al., 2010). For `Infernal`, each input sequence is passed to `cmscan` using the `--cut_ga --rfam --nohmmonly --noali --tblout` options against a comprehensive riboswitch covariance model of families compiled from the latest Rfam collection (version 14.9 as of writing) (`https://rfam.org/`) (Griffiths-Jones et al., 2003). `cmscan` output is parsed to identify riboswitch loci for further processing. `Prodigal` is called with the default output and translation options. For each identified riboswitch, the downstream gene is identified from the `Prodigal` output and recorded for downstream plot annotations. The downstream gene is defined as the first open reading frame past the riboswitch 5′ end and on the same strand. This definition captures ORFs that overlap with part of the identified riboswitch sequence, in order to maximize the likelihood of capturing ORFs that are biologically relevant to the riboswitch.

Riboswitch reference generation

To reduce computational requirements for transcript mapping, `TaRTLEt` trims input genomic sequences back to riboswitch neighborhoods, producing reference sequences that stretch from 1000 nt upstream of each identified riboswitch locus to 1000 nt downstream. (For metagenomic inputs, where riboswitch loci may lie less than 1000 nt from contig ends, neighborhoods are permitted to be smaller or asymmetric.) We chose this neighborhood size on the basis of common practice in paired-end sequencing library preparation, where the target fragmentation size is often less than 1000 nt to preserve base quality (Tan et al., 2019). Consequently, the generated reference neighborhoods are typically (96.66% of riboswitch-aligned reads in our validation dataset) large enough to capture both a riboswitch-aligned read and its mate. All references generated by `TaRTLEt` are oriented $5' \rightarrow 3'$; references derived from (-) strand riboswitches are reverse complemented before saving to file.

Reference alignment

RNA-seq reads are aligned to the generated riboswitch reference using `HISAT2` (`http://daehwankimlab.github.io/hisat2/`) (Kim et al., 2019). `HISAT2` was chosen for its fast per-read alignment time, good parallel scaling, and acceptable alignment rates (Musich et al., 2021). `HISAT2` calls are made using the default options, but any additional valid alignment options may be passed through the `TaRTLEt` interface. The current `TaRTLEt` implementation requires that the input RNA-seq datasets be paired-end. The resulting SAM files are converted to sorted BAMs for downstream processing.

We recommend running `HISAT2` using the options `--no-unal --score-min L,0,-0.4`. The use of `--no-unal` dramatically reduces output file sizes by omitting unaligned reads. The use of `--score-min L,0,-0.4` decreases stringency to allow reads with alignment mismatches/soft-clipping to be included in the output; any alignment errors are handled during downstream processing. The validation dataset presented here was run using these parameters in addition to `-p 40` (performance tuning with multiple threads) and `-t` (wall-time logging).

Processing mapped reads

`TaRTLEt` heavily utilizes the `pysam` (`https://github.com/pysam-developers/pysam`) library (a python interface for `samtools` (Li et al., 2009) and `htslib` (Bonfield et al., 2021)) to parse BAM files and generate alignment data objects for downstream data transformation. Once mapped reads/read pairs have been collected from the sorted BAMs, they are filtered to discard pairs in invalid orientations (RF, TANDEM, FFR, RRF), which may reflect sequencing errors. Only pairs in the EQ and FR orientations, and pairs with only one mapped read, are retained as valid. The EQ ("equal") orientation denotes reads aligned in opposite directions that overlap across their entire length, as is often the case when the fragment sequenced is $\sim$150bp, comparable in scale to individual short reads. In the FR orientation, the left-most aligned base (looking $5' \rightarrow 3'$ in the reference) is either shared between the F and R oriented reads or belongs only to the F oriented read, and the right-most aligned base belongs exclusively to the R read.

Next, `TaRTLEt` generates fragment coverage pileups using the filtered set of reads for every position in the reference sequence. Single reads (with unmapped mates) are discarded during counting unless the `--allow-single-reads` flag is passed to `TaRTLEt`. `TaRTLEt` tracks five types of coverage. Two can incorporate data from single reads if included: "read coverage" counts bases directly mapped by a read, while "clipped fragment coverage" counts bases mapped by `HISAT2`'s soft-clipped read regions. The remaining three types require both mates to be mapped. "Overlapped fragment coverage" counts bases overlapped by both mates; "inferred fragment coverage" counts unmapped bases that lie between the mapped regions of the two mates; and "fragment termini coverage" counts only the mapped base that is the origin of read synthesis, i.e., looking $5' \rightarrow 3'$ in the reference, the left-most mapped base for an F read and right-most mapped base for an R read. This last category captures the ends of sequenced fragments to feed into the signed sum of termini for convolution analysis; we count the fragment terminus captured by the F read as -1 and the fragment terminus captured by the R read as $+1$. Reads that have soft-clipped regions are discarded from this step unless the `--allow-soft-clips` flag is specified. Fragment termini counts are removed from the read coverage and/or overlapped coverage counts, and read coverage counts are removed from the overlapped coverage counts, such that no mapped base is counted more than once. This ensures that counts across all coverage types at a given position sum to the read depth at that position.

Finally, to ensure sufficient coverage for further analysis, riboswitch loci are filtered based on the average read coverage across the length of the riboswitch. The coverage threshold can be adjusted with the `--min-cov-depth` option (default 50; the validation presented here uses 15).

Transformation and feature identification

Fragment termini coverage pileups are transformed using a convolution to extract information about transcription starts and stops. `TaRTLEt` generates a 51-element 1-dimensional weighted kernel by discretizing a Gaussian probability density function with $\sigma = 1.5$. The weighted kernel is then applied to the fragment termini coverage array through the convolution to generate an array of peaks corresponding to transcription events. The convolution transforms coverage at each position into convolution amplitude (arbitrary units).

`TaRTLEt` conducts simple peak detection as follows for a positive peak: a local convolution amplitude maximum is defined as the peak "summit". The peak region covers all the bases on either side of the summit at which convolution amplitude (a) is monotonically decreasing and (b) does not enter or cross a zero-region, defined as the interval $(-0.1, 0.1)$. To delineate ambiguous shoulder regions between peaks, peak bounds follow pythonic inclusive-exclusive convention where the left peak bound represents the left-most valid base according to the monotonicity and zero-crossing criteria and the right peak bound represents the first base that fails the criteria. Negative and positive amplitude peaks are considered candidate transcription start and termination events, respectively.

Next, `TaRTLEt` assesses the significance of candidate transcription termination events. Each positive peak within a user-defined region is compared to a 2-dimensional multivariate normal (MVN) distribution fitted to the widths and amplitudes of each other peak in the region (the noise set). The likelihood of drawing the peak in question from the MVN is used as a pseudo *p*-value, with a significance threshold of < 0.05. Peaks that pass this threshold are called as transcription termination events. The region that feeds into the noise set is defined relative to riboswitch length, with parameters defining the proportion of the riboswitch that should be excluded at the $5'$ end and how far past the riboswitch $3'$ end the region should stretch. Thus, our validation analysis used `--ext-prop -0.3 1.0` to define a noise set region that excludes the first 30% of the riboswitch, includes the remainder, and continues beyond the riboswitch $3'$ end to include a region equal in length to the complete riboswitch.

Peak clustering and statistics

For each riboswitch locus, `TaRTLEt` compares data from all treatment conditions to identify differential transcription termination. All peaks that fall within the riboswitch region of interest (by default, within $\pm 0.3 \times$riboswitch size of the riboswitch $3'$ end) under any condition are clustered by position to enable cross-condition comparisons. The riboswitch $3'$ end defines position 0, and peak positions are given relative to this site as a fraction of the riboswitch size. For example, a position of -0.15 would indicate that the peak is located 15% of the riboswitch size upstream of the riboswitch $3'$ end. Clustering is implemented via `fclusterdata` from the `SciPy` package and follows a cophenetic distancing algorithm with complete linkage. A fixed cophenetic distance threshold of 0.04 was chosen through trial and error.

Once peaks are assigned to clusters, cross-cluster comparisons at each locus are used to test for condition-dependent transcription termination. Each peak is characterized by the change in total fragment coverage across its bounds. The fractional coverage change is defined as the difference in total coverage between the left and right bounds of the peak expressed as a fraction of the average total read coverage between the $5'$ and $3'$ ends of the riboswitch. `TaRTLEt` performs a Mann-Whitney one-tailed U-test (`scipy.stats.mannwhitneyu`) to compare each cluster's fractional coverage change to the mean fractional coverage change of all other peaks, to test whether the cluster mean is more negative than the mean of the set of all other peaks.

Testing for differences in cluster variance must account for bias from differences in cluster sizes. `TaRTLEt` uses Levene's test (`scipy.stats.levene`) to run pairwise comparisons between the variance of a given cluster enclosing *n* peaks and that of random samples of *n* peaks chosen without replacement from all other peaks at the same riboswitch locus until the superset of peaks is exhausted or at least 30 comparisons have been made. If the superset is exhausted before 30 comparisons are made, all peaks are replaced and the clusters are randomly resampled to reach 30 comparisons. `TaRTLEt`

returns the median of the resulting 30 p-values and exports this value for each peak cluster, along with Mann-Whitney test results, to `cluster_stats.csv`.

Description of outputs

`TaRTLEt` outputs data on two layers of abstraction: locus plots of fragment coverage with the convolution output for each riboswitch locus and RNA-seq treatment, and the summary peak log. A sample locus plot of the MOCO riboswitch from *E. coli* in shown in Fig. 4. Convolution peaks from each user-defined dataset (`--dset` flag; see documentation) are collected across all riboswitches from the dataset and RNA-seq alignments and recorded in the peak log for the dataset (Table, S1). The peak log can be used to generate a summary peak plot (tutorial and `R` source used are present in the repository) that allows easy inspection of condition-dependent modulation of transcription termination efficiency near the riboswitch $3'$ end (Fig. 5). Data for the clustering of peaks and statistical testing of clusters is written out along with the peak log and informs inferences made from peak plots.

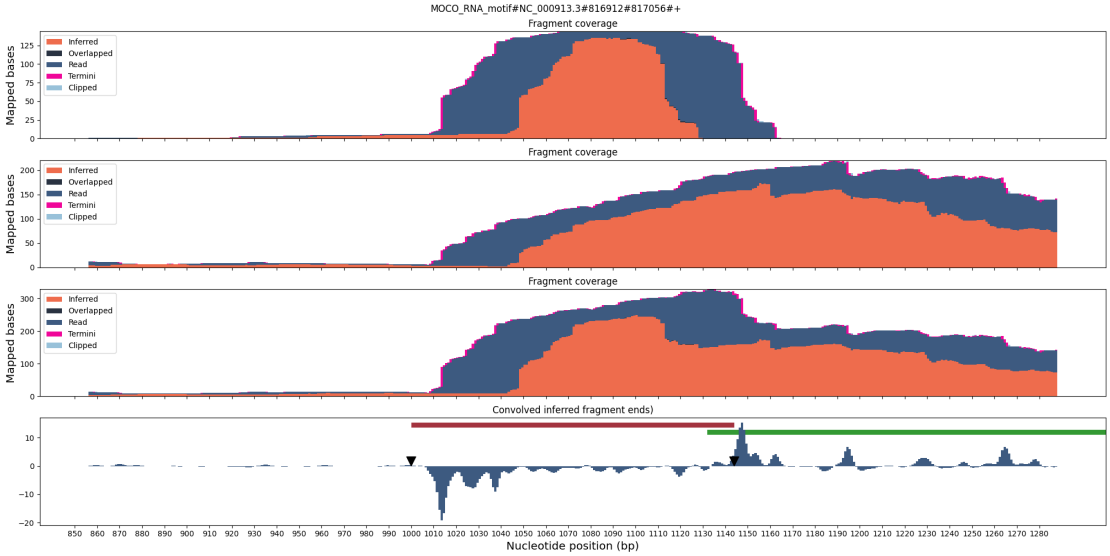

Figure 4. Locus plot of the MOCO riboswitch from *E. coli* for RNA-seq sample SRR7154624. The top panel collects fragment coverage for read pairs with R-oriented reads starting in the riboswitch region of interest ($\pm 0.3\times$riboswitch size). The second panel collects fragment coverage for read pairs with R-oriented reads starting outside the region of interest. Coverage counts in the third panel are the union of the top two panels. The bottom panel records the results from applying the Gaussian convolution to the fragment termini coverage. Vertical axes for all panels are not normalized and are scaled to fit the maximum displayed value.

Validation datasets

All the paired-end RNA-seq datasets used to test and characterize `TaRTLEt` here are presented in Table 1. These datasets were obtained from the Gene Expression Omnibus (GEO) (Barrett et al., 2012). We used the Ribocentre-switch database (Bu et al., 2024) to facilitate identification of previously characterized riboswitches in this dataset.

RESULTS

`TaRTLEt` performs well on known transcriptionally-active riboswitches

To validate this new tool, we ran `TaRTLEt` on publicly available paired-end RNA-seq datasets for 31 microbial genomes (Table 1). Of these, datasets for 28 genomes were successfully processed through `TaRTLEt`. This collection of microbes is phylogenetically diverse, spanning numerous taxa (Fig. 6a), and includes 16 riboswitches that have previously been experimentally characterized.

Our results are in good agreement with literature findings (Fig. 6b). `TaRTLEt`'s peak plots (Fig. S2, S3) showed condition-dependent termination modulation for the *samT* (*yitJ*) SAM, *xpt* purine, *mgtE*

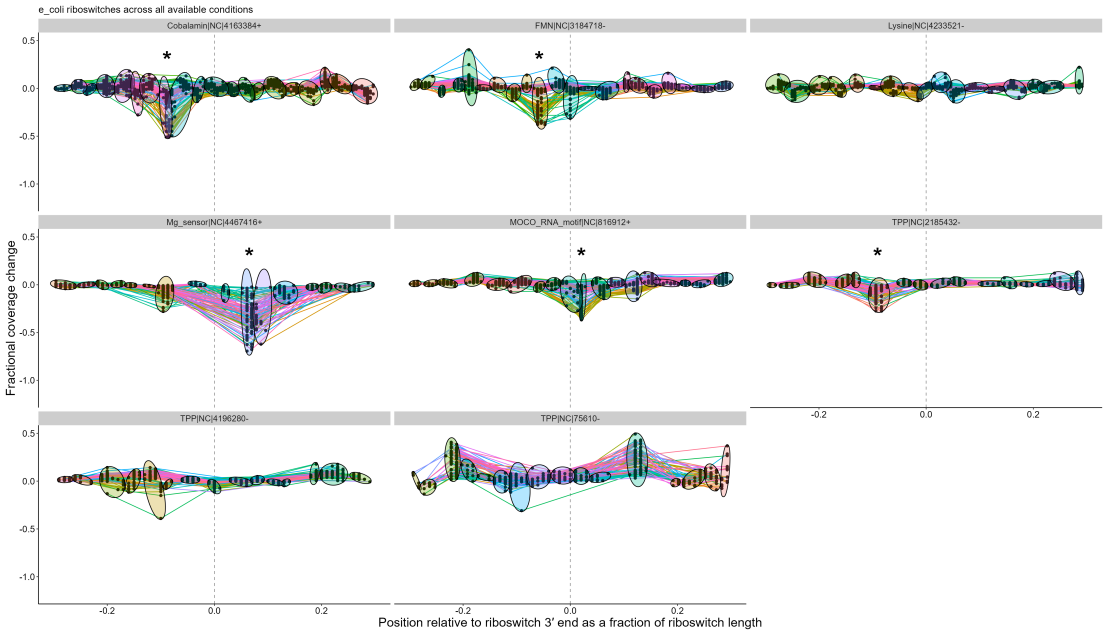

Figure 5. Peak plots (see Fig. 3) for the eight *E. coli* riboswitch loci identified by `Infernal` in the *E. coli* genome. Asterisks mark peak clusters with significantly lower mean ($p < 0.05$, Mann-Whitney one-tailed U-test) than the set of all other peak clusters and significantly ($p < 0.05$, Levene's test) higher variance when compared to randomly sampled sets of the same number of peaks.

M-box, and *lysC* lysine riboswitches in *B. subtilis* (Winkler et al., 2003; Mandal and Breaker, 2004; Dann et al., 2007; Sudarsan et al., 2003), and for the *thiM* TPP riboswitches in *E. coli* (Ontiveros-Palacios et al., 2008; Sedlyarova et al., 2016; Bastet et al., 2017). Conversely, the *B. subtilis glmS* riboswitch, the *Enterococcus faecalis* S_{MK} riboswitch, and the *E. coli lysC* lysine riboswitch have been shown to act via other mechanisms (Collins et al., 2007; Fuchs et al., 2007; Caron et al., 2012), and for these loci `TaRTLEt` finds no evidence of riboswitch regulation by transcription termination across the conditions represented in the dataset.

The validation dataset highlights the range of situations that can give rise to inconclusive or negative results. First, `TaRTLEt` cannot provide insight into riboswitch loci with read coverage too low to support statistical analysis. We observed this outcome in roughly 37 of the 395 riboswitch loci investigated here, leaving 358 loci tractable for further investigation. Second, no candidate transcription-termination events might be detected in the search space at a riboswitch locus (156 of 358); this could happen either because there is no terminator present, or because under the conditions tested the terminator is always inactive.

Third, a candidate transcription-termination event might fail the mean fold coverage drop test if the conditions examined yield consistently high rates of transcriptional read-through (whether because the conditions used sample only the ON state or because the riboswitch acts downstream of transcription). In our validation dataset, 60 of 358 failed this test. The dataset includes three riboswitch loci (the *B. subtilis gcvT* glycine, *ribB* FMN, and *pbuE* purine riboswitches) that have previously been shown to modulate transcription (Mandal et al., 2004; Hollands et al., 2012; Mandal and Breaker, 2004) but that show no mean fold coverage drop here, suggesting that the range of conditions examined was too narrow to capture a state change for these riboswitches. Interestingly, the purine ligand does vary enough across these conditions to cause a detectable switch in transcription termination at the *xpt* purine riboswitch, suggesting that these two purine riboswitches are responsive to different ligand concentrations. Similarly, the *E. coli thiC* TPP riboswitch has been shown *in vitro* to be a potential modulator of both translation and transcription (Winkler et al., 2002; Chauvier et al., 2017), but our validation dataset does not capture evidence of condition-dependent termination modulation *in vivo*.

Finally, a candidate that passes the mean test may fail the variance test (134 of 358 loci), either if

Table 1. List of GEO accessions for paired-end RNA-seq read sets used for characterizing riboswitches from each organism.

Microbe	GEO Accession	Citation
Acinetobacter baumannii	GSE183334	Unpublished
Aliivibrio fischeri ES114	GSE237189	Vander Griend et al. (2024)
Apilactobacillus kunkeei	GSE205998	Seeger et al. (2023)
Bacillus anthracis	GSE152356	Corsi et al. (2021)
Bacillus subtilis str. 168	GSE219221	Guo and Herman (2023)
Bacillus subtilis str. 168	GSE226559	Unpublished
Bacteroides fragilis	GSE220692	Fiebig et al. (2024)
Bacteroides thetaiotaomicron	GSE129572	Briliūtė et al. (2019)
Bacteroides thetaiotaomicron	GSE169260	Lewis and Gui (2023)
Bacteroides xylanisolvens	GSE74379	Despres et al. (2016)
Burkholderia pseudomallei	GSE152295	Avican et al. (2021)
Caulobacter vibrioides	GSE241057	McLaughlin et al. (2023)
Clostridioides difficile	GSE173804	Weiss et al. (2021)
Desulfovibrio vulgaris str. Hildenborough	GSE101911	Chen et al. (2019)
Desulfovibrio vulgaris str. Hildenborough	GSE78834	Gao et al. (2016)
Enterococcus faecalis	GSE152295	Avican et al. (2021)
Escherichia coli	GSE114358	Guzman et al. (2018)
Escherichia coli	GSE122211	Sastry et al. (2019)
Escherichia coli	GSE122295	Sastry et al. (2019)
Escherichia coli	GSE122296	Sastry et al. (2019)
Escherichia coli	GSE122779	Anand et al. (2019)
Eubacterium limosum	GSE149269	Jeong et al. (2020)
Klebsiella pneumoniae	GSE229867	Unpublished
Mycobacterium tuberculosis str. H37Rv	GSE218354	Martini et al. (2023)
Mycolicibacterium smegmatis MC2 155	GSE232901	Grigorov et al. (2023)
Neisseria gonorrhoeae	GSE191020	Ray et al. (2022)
Piscirickettsia salmonis	GSE235725	Carril et al. (2023)
Pseudomonas chlororaphis subsp. Aureofaciens 30-84	GSE61200	Wang et al. (2016)
Pseudomonas fluorescens	GSE200822	Hasnain et al. (2023)
Salmonella enterica sv. Typhimurium	GSE203342	Kant et al. (2023)
Sinorhizobium meliloti	GSE151705	Fagorzi et al. (2021)
Staphylococcus epidermidis	GSE152295	Avican et al. (2021)
Streptococcus sanguinis SK36	GSE174672	Puccio et al. (2022)
Streptomyces coelicolor	GSE234439	Li et al. (2023)
Synechococcus elongatus	GSE227397	Lu et al. (2023)
Synechocystis sp. PCC 6803	GSE132275	Unpublished
Xanthomonas albilineans	GSE229478	Wimmer et al. (2024)
Xanthomonas oryzae pv. Oryzae KACC 10331	GSE89651	Park et al. (2022)

the variance of other peak clusters is large (e.g., due to shallow coverage) or if the variance of the candidate cluster is small (due to a set of conditions that capture only one riboswitch regulatory state). In our validation set, we found one clear instance of a previously characterized transcription-terminating riboswitch, the *B. subtilis ribD* FMN riboswitch (Wickiser et al., 2005), that appears to be consistently OFF under the conditions tested, thus failing the variance test (Fig. S2, panel `FMN|NC|2431617-`). Thus while our mean- and variance-based calling approach can in general be applied to existing datasets, the noisiness and the scope of the datasets examined can limit robust inference.

`TaRTLEt` gives new insights into undercharacterized riboswitches

Thiamine pyrophosphate (TPP) riboswitches are the most abundant riboswitches in genome databases (McCown et al., 2017) and were identified in most of the genomes in our validation set. *In vitro*

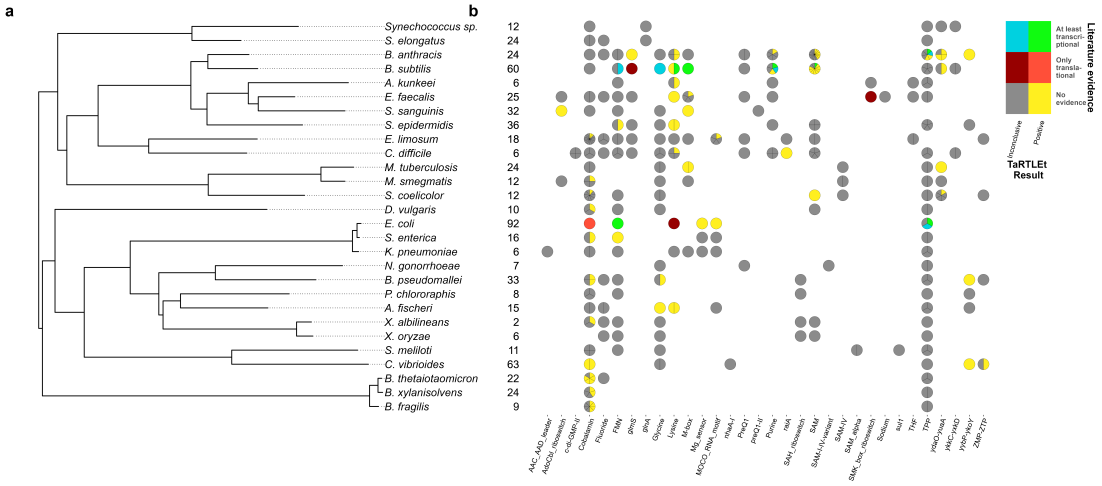

Figure 6. Compilation of `TaRTLEt` results for a collection of commonly studied bacteria, using RNA-seq data obtained from public GEO repositories. (a) Phylogenetic relationships between bacteria studied with `TaRTLEt`. The tree is based on GTDB (Parks et al., 2022; Rinke et al., 2021; Parks et al., 2020, 2018) bacterial phylogenies (release 214.1) and built using the `ggtree` R package (Yu et al., 2017). (b) Presence/absence of evidence for condition-dependent termination efficacy for every riboswitch representative.

characterization of this class has variously pointed to regulation at the level of transcription termination, translation initiation, or even both (Sedlyarova et al., 2016; Winkler et al., 2002; Chauvier et al., 2017). In one notable case, *B. anthracis* encodes two TPP riboswitches in tandem upstream of the *tenA* gene, and each has been shown *in vitro* to be able to regulate transcription termination independently (Welz and Breaker, 2007). Of the 57 TPP riboswitches in the validation dataset, only 3 (1 from *E. coli* and 2 from *B. anthracis*) TPP representatives show transcriptional activity under the conditions tested. Intriguingly, the peak plot for *B. anthracis* showed transcriptional activity by the upstream riboswitch but not by the downstream representative (Fig. S3, panels `TPP|752099+` and `TPP|752272+`), suggesting that the two tandem riboswitches are responsive under different conditions.

A second very widespread riboswitch class binds cobalamin, with at least one example in all but 4 genomes in our validation dataset. Previous research indicates that cobalamin riboswitches in Gram-positive taxa typically work at the transcriptional level, while those in Gram-negative bacteria regulate translation (Lennon and Batey, 2022). Contra this expectation, we see clear evidence of transcriptionally-active cobalamin riboswitches in nearly half the Gram-negative taxa we investigated (Fig. 6). Interestingly, however, in nearly every case, these taxa also contained one or more additional cobalamin riboswitches that showed no evidence of transcriptional regulation. These other riboswitches could either act by different mechanisms or act transcriptionally but with sensitivity to a different range of ligand concentrations. While `TaRTLEt` cannot distinguish between these possibilities with the data used here, the observation underscores the potential variability of riboswitches within a ligand class and highlights the utility of high-throughput analyses for identifying interesting candidates for experimental characterization.

`TaRTLEt` also offers insights into the less-studied riboswitch classes. For example, the molybdenum cofactor ("MOCO") riboswitch in *E. coli* regulating the *moaA* operon is well-characterized structurally (Amadei et al., 2023) but acts via an unknown mechanism. Our results indicate that this MOCO riboswitch modulates transcription termination. Our analysis of existing public data further revealed evidence for condition-dependent modulation of termination efficiency by riboswitches responsive to concentrations of glycine (2 of 23 riboswitch loci, representing 2 of 17 species), lysine (10 of 20; 7 of 10 species), M-box (5 of 13; 4 of 8 species), purine (3 of 18; 2 of 6 species), SAM (17 of 43; 3 of 9 species), cyclic di-AMP (*ydaO-yuaA*; 6 of 15; 4 of 6 species), Mn^{2+} (*yybP-ykoY*; 3 of 6; 3 of 6 species), and ZMP-ZTP (1 of 5; 1 of 4 species). That is, `TaRTLEt`-enabled opportunistic analysis of data originally gathered for other purposes offers strong predictions of molecular mechanism at >30 individual riboswitch loci.

DISCUSSION

TaRTLEt's built-in visualizations facilitate user assessment of positive-call reliability. Clusters of peaks at multiple loci under multiple conditions can be rapidly inspected in peak plots, with the underlying coverage plots available to check results at lower degrees of abstraction. Thus, TaRTLEt serves as a much-needed high-throughput tool for preliminary assessment of riboswitch mechanism and can readily identify strong candidates for *in vitro* validation.

As currently implemented, TaRTLEt can analyze paired-end RNA-seq data to detect positive evidence of condition-dependent changes in transcription termination at riboswitch loci. The absence of such evidence cannot be used to infer a different mechanism of riboswitch action, because TaRTLEt's approach depends on having a dataset that captures both ON and OFF states. Thus, among the 8 species in our validation dataset for which the RNA-seq collection included fewer than 10 conditions, TaRTLEt could detect evidence for transcriptional regulation at any riboswitch in only two (Fig. 6b). Importantly, our approach supports inference from RNA-seq data collected across multiple different experiments, such that marginal additional data collection for these species could make detection rates rise substantially.

CONCLUSIONS

We identified a transcriptomic signal that should arise from riboswitch-mediated conditional regulation of transcription termination, developed TaRTLEt to apply convolution methods to extract the signal from existing public RNA-seq data, and demonstrated that this tool supports confident identification of riboswitches that control transcription termination. Our results with this method are in good agreement with the limited number of *in vitro* characterizations available, offer new insights into previously uncharacterized riboswitches, and emphasize that riboswitches within the same ligand class, even within the same organism, can exhibit different ligand sensitivities and act by different mechanisms. Whereas detailed *in vitro* characterization can establish which of several mechanisms govern a single riboswitch, our high-throughput approach can ask which of any number of riboswitch loci act by transcription termination and broadly survey the conditions under which this activity is invoked *in vivo*. Application of TaRTLEt to microbial transcriptomic and metatranscriptomic datasets therefore offers a path forward to understanding the regulation of a wide range of biosynthetic and transport pathways in biological contexts ranging from pure culture to natural communities *in situ*.

CODE AVAILABILITY

The tartlet GitHub repository houses the tool source, while the tartlet-pub GitHub repository tracks the data collection, processing, and analysis code used for this work.

ACKNOWLEDGMENTS

This study was supported in part by the Department of Energy Office of Science award number DE-SC0023307. This research is a contribution of the EMERGE Biology Integration Institute, funded by the NSF Biology Integration Institutes Program, award number 2022070. We thank the members of EMERGE-BII, the VirSoil collaboration, and the Bagby lab for helpful discussions.

REFERENCES

Adams, P. P., Flores Avile, C., Popitsch, N., Bilusic, I., Schroeder, R., Lybecker, M., and Jewett, M. W. (2017). In vivo expression technology and 5′ end mapping of the *Borrelia burgdorferi* transcriptome identify novel RNAs expressed during mammalian infection. *Nucleic Acids Research*, 45(2):775–792.

Albawi, S., Mohammed, T. A., and Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *2017 International Conference on Engineering and Technology (ICET)*, pages 1–6. IEEE.

Amadei, F., Reichenbach, M., Gallo, S., and Sigel, R. K. (2023). The structural features of the ligand-free *moaA* riboswitch and its ion-dependent folding. *Journal of Inorganic Biochemistry*, 242:112153.

Anand, A., Olson, C. A., Yang, L., Sastry, A. V., Catoiu, E., Choudhary, K. S., Phaneuf, P. V., Sandberg, T. E., Xu, S., Hefner, Y., Szubin, R., Feist, A. M., and Palsson, B. O. (2019). Pseudogene repair driven by selection pressure applied in experimental evolution. *Nature Microbiology*, 4(3):386–389.

Ariza-Mateos, A., Nuthanakanti, A., and Serganov, A. (2021). Riboswitch mechanisms: new tricks for an old dog. *Biochemistry (Moscow)*, 86:962–975.

Avican, K., Aldahdooh, J., Togninalli, M., Mahmud, A. F., Tang, J., Borgwardt, K. M., Rhen, M., and Fällman, M. (2021). RNA atlas of human bacterial pathogens uncovers stress dynamics linked to infection. *Nature Communications*, 12(1):3282.

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., and Soboleva, A. (2012). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research*, 41(D1):D991–D995.

Barrick, J. E. and Breaker, R. R. (2007). The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biology*, 8(11):1–19.

Bastet, L., Chauvier, A., Singh, N., Lussier, A., Lamontagne, A.-M., Prévost, K., Massé, E., Wade, J. T., and Lafontaine, D. A. (2017). Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic Acids Research*, 45(12):7474–7486.

Bédard, A.-S. V., Hien, E. D., and Lafontaine, D. A. (2020). Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1863(3):194501.

Blount, K. F. and Breaker, R. R. (2006). Riboswitches as antibacterial drug targets. *Nature Biotechnology*, 24(12):1558–1564.

Bonfield, J. K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., and Davies, R. M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience*, 10(2):giab007.

Boussebayle, A., Torka, D., Ollivaud, S., Braun, J., Bofill-Bosch, C., Dombrowski, M., Groher, F., Hamacher, K., and Suess, B. (2019). Next-level riboswitch development—implementation of capture-SELEX facilitates identification of a new synthetic riboswitch. *Nucleic Acids Research*, 47(9):4883–4895.

Breaker, R. R. (2018). Riboswitches and translation control. *Cold Spring Harbor Perspectives in Biology*, 10(11):a032797.

Briliūtė, J., Urbanowicz, P. A., Luis, A. S., Baslé, A., Paterson, N., Rebello, O., Hendel, J., Ndeh, D. A., Lowe, E. C., Martens, E. C., Spencer, D. I. R., Bolam, D. N., and Crouch, L. I. (2019). Complex *N*-glycan breakdown by gut *Bacteroides* involves an extensive enzymatic apparatus encoded by multiple co-regulated genetic loci. *Nature Microbiology*, 4(9):1571–1581.

Bu, F., Lin, X., Liao, W., Lu, Z., He, Y., Luo, Y., Peng, X., Li, M., Huang, Y., Chen, X., Xiao, B., Jiang, J., Deng, J., Huang, J., Lin, T., Miao, Z., and Huang, L. (2024). Ribocentre-switch: a database of riboswitches. *Nucleic Acids Research*, 52(D1):D265–D272.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 679–698.

Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., and Lafontaine, D. A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences*, 109(50):E3444–E3453.

Carril, G., Winther-Larsen, H. C., Løvoll, M., and Sørum, H. (2023). Cohabitation of *Piscirickettsia salmonis* genogroups (LF-89 and EM-90): synergistic effect on growth dynamics. *Frontiers in Cellular and Infection Microbiology*, 13.

Chang, T.-H., Huang, H.-D., Wu, L.-C., Yeh, C.-T., Liu, B.-J., and Horng, J.-T. (2009). Computational identification of riboswitches based on RNA conserved functional sequences and conformations. *RNA*, 15(7):1426–1430.

Chauvier, A., Picard-Jean, F., Berger-Dancause, J.-C., Bastet, L., Naghdi, M. R., Dubé, A., Turcotte, P., Perreault, J., and Lafontaine, D. A. (2017). Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nature Communications*, 8(1):13892.

Cheah, M. T., Wachter, A., Sudarsan, N., and Breaker, R. R. (2007). Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, 447(7143):497–500.

Chen, Z., Gao, S.-h., Jin, M., Sun, S., Lu, J., Yang, P., Bond, P. L., Yuan, Z., and Guo, J. (2019). Physiological and transcriptomic analyses reveal CuO nanoparticle inhibition of anabolic and catabolic activities of sulfate-reducing bacterium. *Environment International*, 125:65–74.

Collins, J. A., Irnov, I., Baker, S., and Winkler, W. C. (2007). Mechanism of mRNA destabilization by the

glmS ribozyme. *Genes & Development*, 21(24):3356–3368.

Corsi, I. D., Dutta, S., Van Hoof, A., and Koehler, T. M. (2021). AtxA-controlled small RNAs of *Bacillus anthracis* virulence plasmid pXO1 regulate gene expression *in trans*. *Frontiers in Microbiology*, 11:610036.

Dann, C. E., Wakeman, C. A., Sieling, C. L., Baker, S. C., Irnov, I., and Winkler, W. C. (2007). Structure and mechanism of a metal-sensing regulatory RNA. *Cell*, 130(5):878–892.

Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Yeoman, C. J., Miller, M. E. B., Fields, C. J., Terrapon, N., Le Bourvellec, C., Renard, C. M. G. C., Henrissat, B., White, B. A., and Mosoni, P. (2016). Unraveling the pectinolytic function of *Bacteroides xylanisolvens* using a RNA-seq approach and mutagenesis. *BMC Genomics*, 17:1–14.

Eddy, S. R. and Durbin, R. (1994). Rna sequence analysis using covariance models. *Nucleic Acids Research*, 22(11):2079–2088.

Ellinger, E., Chauvier, A., Romero, R. A., Liu, Y., Ray, S., and Walter, N. G. (2023). Riboswitches as therapeutic targets: promise of a new era of antibiotics. *Expert Opinion on Therapeutic Targets*, 27(6):433–445.

Fagorzi, C., Bacci, G., Huang, R., Cangioli, L., Checcucci, A., Fini, M., Perrin, E., Natali, C., Dicenzo, G. C., and Mengoni, A. (2021). Nonadditive transcriptomic signatures of genotype-by-genotype interactions during the initiation of plant-rhizobium symbiosis. *mSystems*, 6(1):10–1128.

Fiebig, A., Schnizlein, M. K., Pena-Rivera, S., Trigodet, F., Dubey, A. A., Hennessy, M. K., Basu, A., Pott, S., Dalal, S., Rubin, D., Sogin, M. L., Eren, A. M., Chang, E. B., and Crosson, S. (2024). Bile acid fitness determinants of a *Bacteroides fragilis* isolate from a human pouchitis patient. *mBio*, 15(1):e02830–23.

Fuchs, R. T., Grundy, F. J., and Henkin, T. M. (2007). S-adenosylmethionine directly inhibits binding of 30S ribosomal subunits to the SMK box translational riboswitch RNA. *Proceedings of the National Academy of Sciences*, 104(12):4876–4880.

Gao, S.-H., Ho, J. Y., Fan, L., Richardson, D. J., Yuan, Z., and Bond, P. L. (2016). Antimicrobial effects of free nitrous acid on *Desulfovibrio vulgaris*: implications for sulfide-induced corrosion of concrete. *Applied and Environmental Microbiology*, 82(18):5563–5575.

Giarimoglou, N., Kouvela, A., Maniatis, A., Papakyriakou, A., Zhang, J., Stamatopoulou, V., and Stathopoulos, C. (2022). A riboswitch-driven era of new antibacterials. *Antibiotics*, 11(9):1243.

Gilbert, S. D., Stoddard, C. D., Wise, S. J., and Batey, R. T. (2006). Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *Journal of Molecular Biology*, 359(3):754–768.

Gong, S., Wang, Y., Wang, Z., and Zhang, W. (2017). Co-transcriptional folding and regulation mechanisms of riboswitches. *Molecules*, 22(7):1169.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003). Rfam: an rna family database. *Nucleic Acids Research*, 31(1):439–441.

Grigorov, A. S., Skvortsova, Y. V., Bychenko, O. S., Aseev, L. V., Koledinskaya, L. S., Boni, I. V., and Azhikina, T. L. (2023). Dynamic transcriptional landscape of *Mycobacterium smegmatis* under cold stress. *International Journal of Molecular Sciences*, 24(16):12706.

Groher, F. and Suess, B. (2014). Synthetic riboswitches—a tool comes of age. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1839(10):964–973.

Guo, T. and Herman, J. K. (2023). Magnesium modulates *Bacillus subtilis* cell division frequency. *Journal of Bacteriology*, 205(1):e00375–22.

Gusarov, I. and Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Molecular Cell*, 3(4):495–504.

Guzman, G. I., Sandberg, T. E., LaCroix, R. A., Nyerges, A., Papp, H., Raad, M. d., King, Z. A., Northen, T. R., Notebaart, R. A., Pál, C., Palsson, B. O., Papp, B., and Feist, A. M. (2018). Enzyme promiscuity shapes evolutionary innovation and optimization.

Hanson, S., Berthelot, K., Fink, B., McCarthy, J. E., and Suess, B. (2003). Tetracycline-aptamer-mediated translational regulation in yeast. *Molecular Microbiology*, 49(6):1627–1637.

Hasnain, A., Balakrishnan, S., Joshy, D. M., Smith, J., Haase, S. B., and Yeung, E. (2023). Learning perturbation-inducible cell states from observability analysis of transcriptome dynamics. *Nature Communications*, 14(1):3148.

Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., and Groisman, E. A.

(2012). Riboswitch control of Rho-dependent transcription termination. *Proceedings of the National Academy of Sciences*, 109(14):5376–5381.

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11:1–11.

Jeong, J., Kim, J.-Y., Park, B., Choi, I.-G., and Chang, I. S. (2020). Genetic engineering system for syngas-utilizing acetogen, *Eubacterium limosum* KIST612. *Bioresource Technology Reports*, 11:100452.

Kant, S., Till, J. K. A., Liu, L., Margolis, A., Uppalapati, S., Kim, J.-S., and Vazquez-Torres, A. (2023). Gre factors help *Salmonella* adapt to oxidative stress by improving transcription elongation and fidelity of metabolic genes. *PLoS Biology*, 21(4):e3002051.

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *Nature biotechnology*, 37(8):907–915.

Klein, D. J. and Ferré-D'Amaré, A. R. (2006). Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science*, 313(5794):1752–1756.

Lennon, S. R. and Batey, R. T. (2022). Regulation of gene expression through effector-dependent conformational switching by cobalamin riboswitches. *Journal of Molecular Biology*, 434(18):167585.

Lewis, J. P. and Gui, Q. (2023). Iron deficiency modulates metabolic landscape of *Bacteroidetes* promoting its resilience during inflammation. *Microbiology Spectrum*, 11(4):e04733–22.

Li, C., Urem, M., Du, C., Zhang, L., and van Wezel, G. P. (2023). Systems-wide analysis of the ROK-family regulatory gene *rokL6* and its role in the control of glucosamine toxicity in *Streptomyces coelicolor*. *Applied and Environmental Microbiology*, 89(12):e01674–23.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, . G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009). A *trans*-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell*, 139(4):770–779.

Lu, K.-J., Chang, C.-W., Wang, C.-H., Chen, F. Y., Huang, I. Y., Huang, P.-H., Yang, C.-H., Wu, H.-Y., Wu, W.-J., Hsu, K.-C., Ho, M.-C., Tsai, M.-D., and Liao, J. C. (2023). An ATP-sensitive phosphoketolase regulates carbon fixation in cyanobacteria. *Nature Metabolism*, 5(7):1111–1126.

Maini, R. and Aggarwal, H. (2009). Study and comparison of various image edge detection techniques. *International Journal of Image Processing (IJIP)*, 3(1):1–11.

Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C., and Breaker, R. R. (2003). Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell*, 113(5):577–586.

Mandal, M. and Breaker, R. R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Structural & Molecular Biology*, 11(1):29–35.

Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., and Breaker, R. R. (2004). A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science*, 306(5694):275–279.

Martini, B. A., Grigorov, A. S., Skvortsova, Y. V., Bychenko, O. S., Salina, E. G., and Azhikina, T. L. (2023). Small RNA MTS1338 configures a stress resistance signature in *Mycobacterium tuberculosis*. *International Journal of Molecular Sciences*, 24(9):7928.

McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E., and Breaker, R. R. (2017). Riboswitch diversity and distribution. *RNA*, 23(7):995–1011.

McLaughlin, M., Fiebig, A., and Crosson, S. (2023). XRE transcription factors conserved in *Caulobacter* and φCbK modulate adhesin development and phage production. *PloS Genetics*, 19(11):e1011048.

Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., and Nudler, E. (2002). Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, 111(5):747–756.

Musich, R., Cadle-Davidson, L., and Osier, M. V. (2021). Comparison of short-read sequence aligners indicates strengths and weaknesses for biologists to consider. *Frontiers in Plant Science*, 12:657240.

Nawrocki, E. P. (2014). *Annotating functional RNAs in genomes using Infernal*, volume 1097 of *Methods in Molecular Biology*, chapter 9, pages 163–197. Humana Press, Totowa, NJ.

Nawrocki, E. P. and Eddy, S. R. (2013a). Computational identification of functional RNA homologs in metagenomic data. *RNA Biology*, 10(7):1170–1179.

Nawrocki, E. P. and Eddy, S. R. (2013b). Infernal 1.1: 100-fold faster RNA homology searches.

Bioinformatics, 29(22):2933–2935.

Nou, X. and Kadner, R. J. (2000). Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proceedings of the National Academy of Sciences*, 97(13):7190–7195.

Ontiveros-Palacios, N., Smith, A. M., Grundy, F. J., Soberon, M., Henkin, T. M., and Miranda-Ríos, J. (2008). Molecular basis of gene regulation by the THI-box riboswitch. *Molecular Microbiology*, 67(4):793–803.

Park, J., Jung, H., Mannaa, M., Kim, N., Han, G., Lee, S.-W., Lee, S.-W., and Seo, Y.-S. (2022). Genome-guided comparative *in planta* transcriptome analyses for identifying cross-species common virulence factors in bacterial phytopathogens. *Frontiers in Plant Science*, 13:1030720.

Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., and Hugenholtz, P. (2020). A complete domain-to-species taxonomy for bacteria and archaea. *Nature biotechnology*, 38(9):1079–1086.

Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A., and Hugenholtz, P. (2022). Gtdb: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic acids research*, 50(D1):D785–D794.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., and Hugenholtz, P. (2018). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology*, 36(10):996–1004.

Puccio, T., An, S.-S., Schultz, A. C., Lizarraga, C. A., Bryant, A. S., Culp, D. J., Burne, R. A., and Kitten, T. (2022). Manganese transport by *Streptococcus sanguinis* in acidic conditions and its impact on growth in vitro and in vivo. *Molecular Microbiology*, 117(2):375–393.

Ray, J. C., Smirnov, A., Maurakis, S. A., Harrison, S. A., Ke, E., Chazin, W. J., Cornelissen, C. N., and Criss, A. K. (2022). Adherence enables *Neisseria gonorrhoeae* to overcome zinc limitation imposed by nutritional immunity proteins. *Infection and Immunity*, 90(3):e00009–22.

Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P.-A., Davín, A. A., Waite, D. W., Whitman, W. B., Parks, D. H., and Hugenholtz, P. (2021). A standardized archaeal taxonomy for the genome taxonomy database. *Nature microbiology*, 6(7):946–959.

Roberts, J. W. (2019). Mechanisms of bacterial transcription termination. *Journal of Molecular Biology*, 431(20):4030–4039.

Rosinski-Chupin, I., Sauvage, E., Fouet, A., Poyart, C., and Glaser, P. (2019). Conserved and specific features of *Streptococcus pyogenes* and *Streptococcus agalactiae* transcriptional landscapes. *BMC Genomics*, 20:1–15.

Rosinski-Chupin, I., Soutourina, O., and Martin-Verstraete, I. (2014). Riboswitch discovery by combining RNA-seq and genome-wide identification of transcriptional start sites. In *Methods in Enzymology*, volume 549, pages 3–27. Elsevier.

Sastry, A. V., Gao, Y., Szubin, R., Hefner, Y., Xu, S., Kim, D., Choudhary, K. S., Yang, L., King, Z. A., and Palsson, B. O. (2019). The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nature Communications*, 10(1):5536.

Sedlyarova, N., Shamovsky, I., Bharati, B. K., Epshtein, V., Chen, J., Gottesman, S., Schroeder, R., and Nudler, E. (2016). sRNA-mediated control of transcription termination in *E. coli*. *Cell*, 167(1):111–121.

Seeger, C., Dyrhage, K., Näslund, K., and Andersson, S. G. (2023). *Apilactobacillus kunkeei* releases RNA-associated membrane vesicles and proteinaceous nanoparticles. *Microlife*, 4:uqad037.

Stav, S., Atilho, R. M., Mirihana Arachchilage, G., Nguyen, G., Higgs, G., and Breaker, R. R. (2019). Genome-wide discovery of structured noncoding RNAs in bacteria. *BMC Microbiology*, 19(1):1–18.

Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G. M., and Breaker, R. R. (2005). Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chemistry & Biology*, 12(12):1325–1335.

Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S., and Breaker, R. R. (2003). An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & Development*, 17(21):2688–2697.

Sun, E. I. and Rodionov, D. A. (2014). Computational analysis of riboswitch-based regulation. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1839(10):900–907.

Tan, G., Opitz, L., Schlapbach, R., and Rehrauer, H. (2019). Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports*, 9(1):2856.

Vander Griend, J. A., Isenberg, R. Y., Kotla, K. R., and Mandel, M. J. (2024). Transcriptional pathways across colony biofilm models in the symbiont *Vibrio fischeri*. *mSystems*, 9(1):e00815–23.

Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P. F., and Mörl, M. (2013). *De novo* design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research*, 41(4):2541–2551.

Wang, D., Yu, J. M., Dorosky, R. J., Pierson III, L. S., and Pierson, E. A. (2016). The phenazine 2-hydroxy-phenazine-1-carboxylic acid promotes extracellular DNA release and has broad transcriptomic consequences in *Pseudomonas chlororaphis* 30–84. *PLoS One*, 11(1):e0148003.

Weiss, A., Lopez, C. A., Beavers, W. N., Rodriguez, J., and Skaar, E. P. (2021). *Clostridioides difficile* strain-dependent and strain-independent adaptations to a microaerobic environment. *Microbial Genomics*, 7(12):000738.

Welz, R. and Breaker, R. R. (2007). Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA*, 13(4):573–582.

Wickiser, J. K., Winkler, W. C., Breaker, R. R., and Crothers, D. M. (2005). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular Cell*, 18(1):49–60.

Wimmer, F., Englert, F., Wandera, K. G., Alkhnbashi, O. S., Collins, S. P., Backofen, R., and Beisel, C. L. (2024). Interrogating two extensively self-targeting type I CRISPR-Cas systems in *Xanthomonas albilineans* reveals distinct anti-CRISPR proteins that block DNA degradation. *Nucleic Acids Research*, 52(2):769–783.

Winkler, W., Nahvi, A., and Breaker, R. R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, 419(6910):952–956.

Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E., and Breaker, R. R. (2003). An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Structural & Molecular Biology*, 10(9):701–707.

Yao, Z., Barrick, J., Weinberg, Z., Neph, S., Breaker, R., Tompa, M., and Ruzzo, W. L. (2007). A computational pipeline for high-throughput discovery of *cis*-regulatory noncoding RNA in prokaryotes. *PLoS Computational Biology*, 3(7):e126.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., and Lam, T. T.-Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1):28–36.

Yu, S.-H., Vogel, J., and Förstner, K. U. (2018). ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *Gigascience*, 7(9):giy096.