

TaRTLEt: Transcriptionally-active Riboswitch Tracer Leveraging Edge deTection

Sachit Kshatriya and Sarah C. Bagby

Department of Biology, Case Western Reserve University, Cleveland, OH, United States of America

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 this 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 the 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.

Submitted 15 November 2024

Accepted 13 April 2025

Published 26 May 2025

Corresponding author

Sarah C. Bagby, scb126@case.edu

Academic editor

Aleksandra Galitsyna

Additional Information and
Declarations can be found on
page 18

DOI 10.7717/peerj.19418

© Copyright

2025 Kshatriya and Bagby

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Microbiology, Molecular Biology

Keywords Riboswitch, Transcriptomics, Edge-detection, Mechanism

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 & Suess, 2014](#)), to potential targets for antimicrobials and novel therapeutics ([Blount & Breaker, 2006](#); [Ellinger et al., 2023](#); [Giarimoglou et al., 2022](#)). Yet our understanding of natural riboswitch function remains patchy, with high-confidence computational identification of candidate riboswitch sequences and ligands ([Nawrocki, 2014](#); [Nawrocki & Eddy, 2013a](#); [Chang et al., 2009](#); [Eddy & Durbin, 1994](#); [Yao et al., 2007](#); [Stav et al., 2019](#)) far outpacing our understanding of their activity *in vivo*.

This gap is widened by the breadth of riboswitches' regulatory repertoire. Biochemical methods such as transcript fragment length and ribosome binding assays ([Winkler et al., 2003](#); [Nou & Kadner, 2000](#); [Welz & 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. But this broad pattern includes riboswitches that affect any of several different levels of gene expression. Most straightforwardly, riboswitch ligand binding can sequester or expose conventional expression control logic elements like intrinsic terminators/antiterminators ([Mironov et al., 2002](#); [Mandal & Breaker, 2004](#)) or Rho-binding sites ([Hollands et al., 2012](#)) to alter mRNA production, or translation start sites to alter protein synthesis ([Breaker, 2018](#)). However, riboswitch mechanisms can also go beyond interaction with direct control architectures. Thiamine pyrophosphate (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 transcript degradation by catalyzing upstream mRNA cleavage ([Collins et al., 2007](#); [Klein & 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](#)), among numerous other examples documented to date ([Bédard, Hien & Lafontaine, 2020](#); [Ariza-Mateos, Nuthanakanti & Serganov, 2021](#)).

To understand the role riboswitches play in a given microbe's physiology, we need to understand at what regulatory level each acts. Regulatory mechanisms have been established for at least one member of the riboswitch families binding roughly a dozen ligands, so that in principle we might expect to infer the mechanism used by other members of the same families. But even in the small set of families with ≥ 2 biochemically characterized members, we find cases where different members of a family act by different mechanisms ([Barrick & Breaker, 2007](#)). Thus, robust identification of riboswitch ligands is insufficient to predict regulatory mechanism even for comparatively well-studied riboswitch families, while for many more families, no representative riboswitch has yet been characterized *in vitro*. *In silico* methods seeking to fill this gap have largely relied on DNA-level signals, using 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 ([Barrick & Breaker, 2007](#); [Sun & 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 applies high-throughput computational approaches to RNA-seq data to determine which of the known riboswitches identified in a (meta)omic dataset show evidence of regulating gene expression by altering transcription termination efficiency. 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 at riboswitch loci in paired-end RNA-seq datasets; and, further, that this pattern could be robustly identified using computational approaches developed for edge detection (Canny, 1986). Importantly, this pattern should support distinguishing riboswitch-mediated transcription termination from the broader set of changes in gene expression that could be evoked by other transcriptional regulators (e.g., transcription factor proteins) acting at these loci. We show that the occurrence of this pattern at riboswitch loci identified in publicly available transcriptomic datasets is in good agreement with existing *in vitro* characterization of riboswitch regulatory modes. By extracting information from the traces of riboswitch activity that are left in the RNA pool of cells actively regulating gene expression, our approach complements and extends earlier methods based on inference from DNA sequence. Our high-throughput approach can identify signatures of riboswitch-mediated transcription termination in existing or new (meta)transcriptomic data, greatly expanding not only the set of riboswitch loci with data-driven predictions of regulatory mode but also our understanding of the range of growth conditions under which each transcription-terminating riboswitch is active.

MATERIALS & METHODS

Portions of this text were previously published as part of a preprint (<https://doi.org/10.1101/2024.10.15.618519>).

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, Soutourina & Martin-Verstraete, 2014; Rosinski-Chupin et al., 2019; Adams et al., 2017; Yu, Vogel & Förstner, 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 (Figs. 1A, 1B). 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 (Fig. 1A top). By contrast, in the ON state, RNA polymerase should

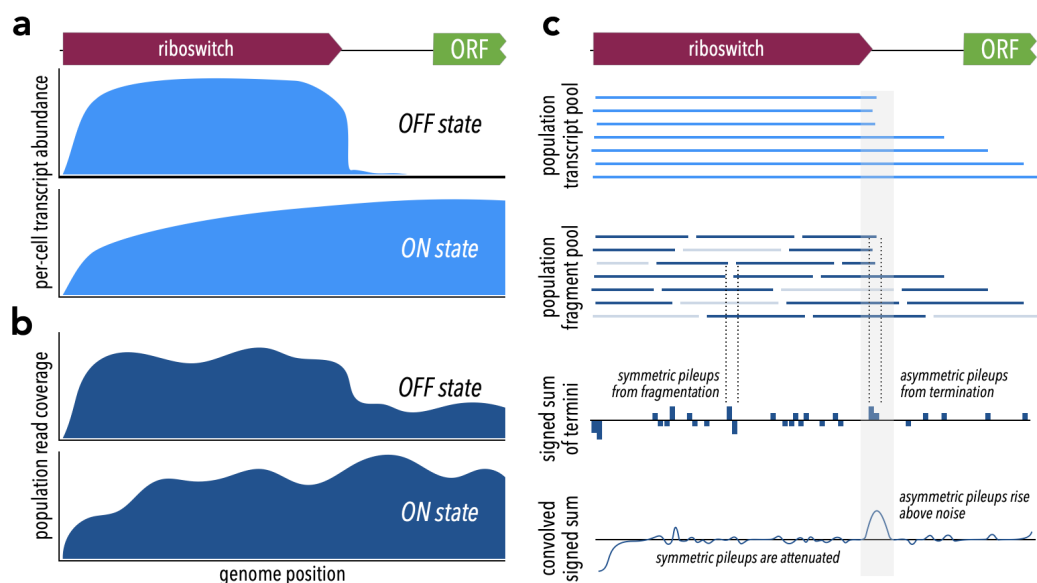


Figure 1 Theoretical framework for TaRTLET's approach. (A) At riboswitches that act by modulating transcription termination, a single cell's transcript abundance profile differs sharply between the riboswitch OFF and ON states. (B) When RNA is extracted from a population of cells for sequencing, coverage profiles are made noisier by a combination of biological (e.g., subpopulations in ON states alongside others in OFF states) and technological factors (e.g., transcript fragmentation, sequencing of a subset of the fragment pool). (C) Weighted convolution on the signed sum of fragment termini attenuates the noise signal, making transcription termination signals readily detectable even when the population-level read coverage does not fall to zero.

Full-size [DOI: 10.7717/peerj.19418/fig-1](https://doi.org/10.7717/peerj.19418/fig-1)

read through, producing transcripts that extend well into the downstream gene (Fig. 1A bottom). A riboswitch that produces these two patterns under different conditions exhibits behavior consistent with riboswitch-mediated conditional regulation of transcription termination.

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, 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](#); [Albawi, Mohammed & Al-Zawi, 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 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. 1C](#)), then calculate a weighted average of surrounding sums for each position. 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 & 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 two-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

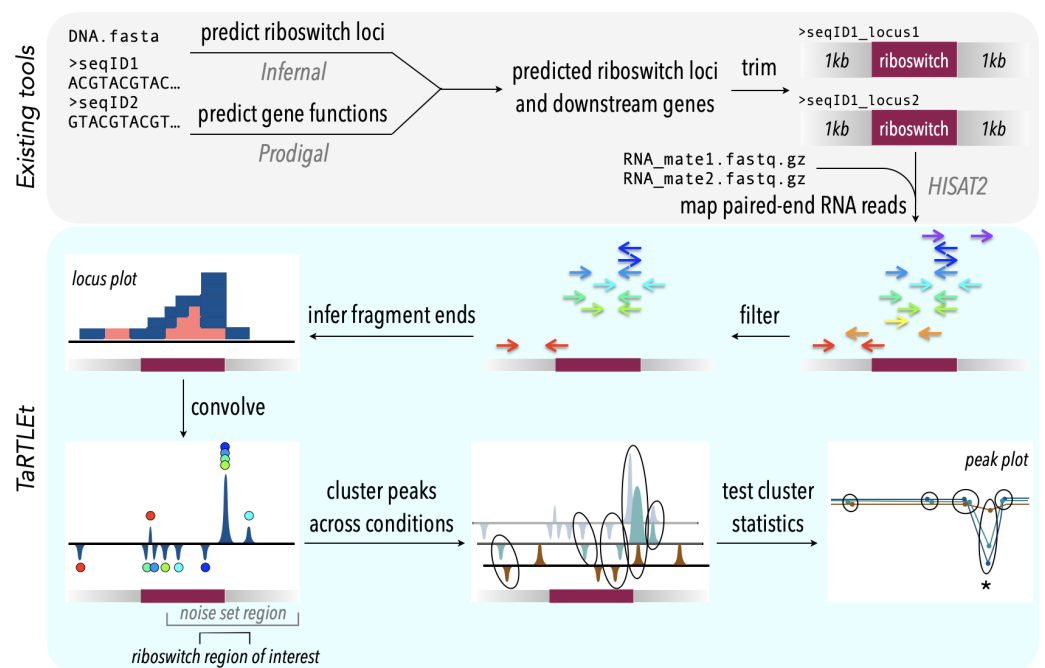


Figure 2 TaRTLEt pipeline for detection of transcription-terminating riboswitches in paired-end RNA-seq data. Existing tools (top) are used to identify known riboswitch families in (meta)genomic data, predict downstream open reading frames, and map RNA-seq reads to a reduced dataset of riboswitch loci and their downstream regions. TaRTLEt then filters mapped reads by orientation and analyzes the distribution of fragment termini, generating a signed sum of termini at each position and applying a weighted convolution to attenuate the symmetric signals arising from mRNA fragmentation before sequencing. Peaks in the convolution are clustered across conditions by genome position; then TaRTLEt analyzes changes in fractional read coverage across each cluster to identify sites of condition-dependent transcription termination. Locus plots are generated for each riboswitch locus under each experimental condition represented in the RNA-seq data, to support per-locus per-condition calling of riboswitch ON or OFF state; peak plots capture all experimental conditions, to support per-locus calling of riboswitch regulatory mode.

Full-size [DOI: 10.7717/peerj.19418/fig-2](https://doi.org/10.7717/peerj.19418/fig-2)

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 $0.5 \times \text{riboswitch size}$ both upstream and downstream of the riboswitch 3' end and examine the fractional coverage change (that is, $(\text{change in coverage across peak}) / (\text{average coverage depth across entire riboswitch})$) 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 (Fig. 2). Any cluster that captures both ON and OFF states of a transcription-terminating riboswitch should exhibit fractional coverage change with a mean that is significantly more negative and a variance that is significantly larger than baseline (Figs. 2 and 3).

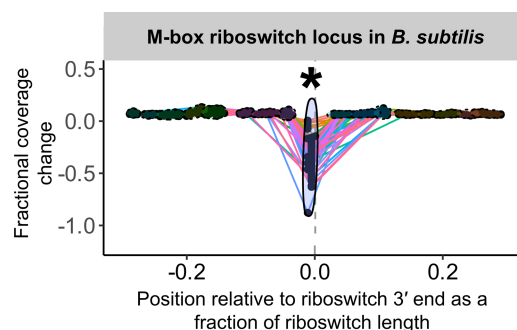


Figure 3 Sample peak plot output of TaRTLEt, showing evidence of conditional transcription termination at a riboswitch 3' end. For each riboswitch locus, the locus plot arising from each experimental condition is convolved and fragment terminus peaks are detected. Each point shows the fractional coverage change across one such peak. Colored lines connect peaks detected under the same experimental conditions. 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. Peaks are clustered across experimental conditions based on x-position; ellipses mark peak clusters. All detected peaks are clustered, but only clusters containing at least one transcription-termination peak are viable candidates. Significance testing of the mean and variance of each such cluster's fractional coverage drop identifies clusters consistent with conditional transcription termination by the riboswitch locus (asterisk).

Full-size [DOI: 10.7717/peerj.19418/fig-3](https://doi.org/10.7717/peerj.19418/fig-3)

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 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>). The pipeline is summarized conceptually in Fig. 2 and in more detail in Fig. S2.

Identification of riboswitches and open reading frames

Input sequences in FASTA format are processed through Infernal v1.1.5 (<http://eddylab.org/infernal>) (Nawrocki & Eddy, 2013b) and Prodigal v2.6.3 (<https://github.com/hyattprodigal>) (Hyatt et al., 2010). For Infernal, each input sequence is passed to cmscan using the --cut_ga --rfam --nohmonly --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 open reading frames

(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 1,000 nt upstream of each identified riboswitch locus to 1,000 nt downstream. (For metagenomic inputs, where riboswitch loci may lie less than 1,000 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 1,000 nt to preserve base quality ([Tan et al., 2019](#)). Consequently, the generated reference neighborhoods are typically (96.61% of riboswitch-aligned reads in our validation dataset) large enough to capture both a riboswitch-aligned read and its paired-end mate. All references generated by TaRTLEt are oriented 5' → 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, Cadle-Davidson & Osier, 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 ~150 bp, comparable in scale to individual short reads. In the FR orientation, the left-most aligned base (looking 5' → 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; the validation presented here uses the default value of 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 one-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 candidate transcription-termination peaks individually to determine which rise above background. Working with per-locus, per-condition convolution plots, TaRTLEt compares each positive peak within a user-defined region to a two-dimensional multivariate normal (MVN) distribution fitted to the widths and amplitudes of each other peak in the region (the “noise set”). 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. (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.) The likelihood of drawing a given peak from the MVN is used as a pseudo *p*-value, with a significance threshold of < 0.05 . Peaks that pass this test and exceed a user-configurable coverage change threshold (by default, a drop of $\geq 20\%$ of the mean coverage across the locus) are called as transcription termination events.

To facilitate biological interpretation of riboswitch activity, TaRTLEt then asks, for each condition and each locus, whether any passing transcription-termination peaks were observed. This within-condition check does not in itself offer evidence for riboswitch regulation by transcription termination. Instead, it (1) feeds into the cluster-level calls described below; (2) supports manual curation of results; and (3) allows the user to ask, among the transcription-terminating riboswitches identified below, what set of experimental conditions evoke the ON state *vs.* the OFF state.

Peak clustering and statistics

To detect differential transcription termination, TaRTLEt compares data from all treatment conditions at each riboswitch locus. All peaks that fall within the riboswitch region of interest (user-configurable; by default, within $\pm 0.5 \times$ 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. The default (user-configurable) cophenetic distance threshold of 0.04 was chosen through trial and error (Fig. S3).

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 60 comparisons have been made. If the superset is exhausted before 60 comparisons are made, all peaks are replaced and the clusters are randomly resampled to reach 60 comparisons. TaRTLEt returns the median of the resulting 60 p -values for each peak cluster.

Finally, TaRTLEt combines peak- and cluster-level tests to assess significance. To be called as showing evidence of condition-dependent transcription termination, a riboswitch locus must have at least one cluster that (i) includes at least one significant transcription-termination peak (*i.e.*, distinct from the noise-set MVN at $p < 0.05$ and coverage drop greater than or equal to the user-defined threshold; see previous section); (ii) passes the Mann–Whitney U-test for mean fractional coverage change; and (iii) passes the Levene test for increased variance.

Description of outputs

TaRTLEt outputs data on two layers of abstraction: single-condition convolution peaks, and cross-condition peak clusters. At each level, TaRTLEt generates both tabular output on each feature and plots of each riboswitch locus. For each candidate peak in the convolution, the single-condition tabular output (`peak_log.csv`) captures the riboswitch locus, the RNA-seq dataset, the location and shape of the peak, the “noise set” of comparator peaks, TaRTLEt’s pass/fail call for that peak as a candidate transcription termination event, and the reason for that call. Single-condition locus plots are sorted into `pass/` and `fail/` subdirectories, where `pass/` contains locus plots for all conditions in which any peak, regardless of location within the locus, was called as a transcription termination event. A sample locus plot of the molybdenum cofactor (MOCO) riboswitch from *E. coli* is shown in [Fig. 4](#); for a sample peak log, see [Table S1](#).

At the level of cross-condition cluster comparisons, TaRTLEt records each cluster’s mean and variance test statistics in `cluster_stats.csv`, then uses this file in conjunction with `peak_log.csv` to generate a summary peak plot that allows easy inspection of condition-dependent modulation of transcription termination efficiency near the 3’ ends of all riboswitches in the dataset. [Figure 5](#) shows a sample summary peak plot. The user may simply scan the summary peak plot to identify loci in which a cluster was marked significant (*), optionally using the locus plots to identify particular experimental conditions that evoked a change in riboswitch transcription termination efficiency (at loci identified as modulating transcription termination) or to spot-check peak call quality.

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.

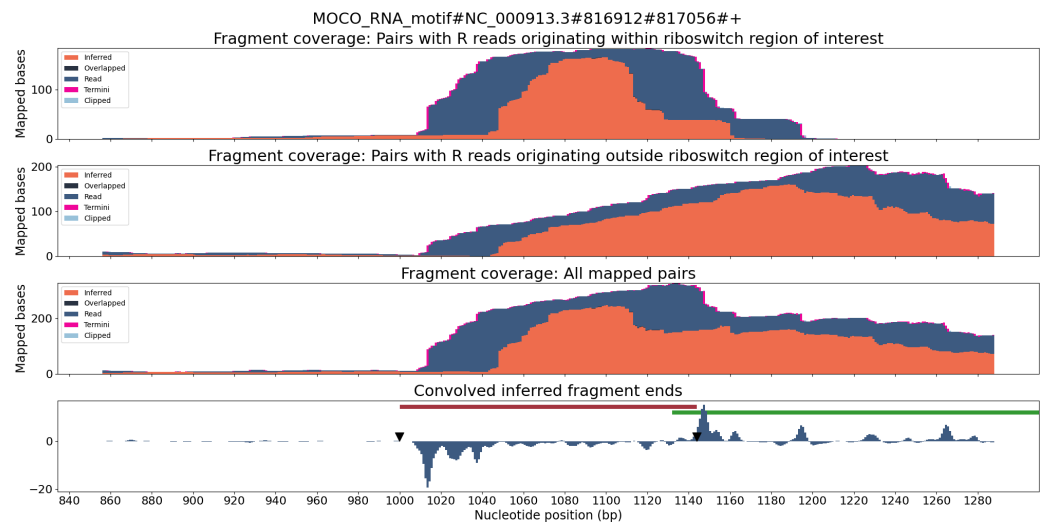


Figure 4 Locus plot of the MOCO riboswitch from *E. coli* for RNA-seq sample SRR7154624. The top panel collects fragment coverage for read pairs with R-oriented reads starting in the riboswitch region of interest ($\pm 0.5 \times$ riboswitch size). The second panel collects fragment coverage for read 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.

Full-size [DOI: 10.7717/peerj.19418/fig-4](https://doi.org/10.7717/peerj.19418/fig-4)

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 29 genomes were successfully processed through TaRTLEt. This collection of microbes is phylogenetically diverse, with good representation of both Gram-positive and Gram-negative bacteria (Fig. 6A), and includes 21 riboswitch loci that have previously been experimentally characterized.

While this low count underscores the need for high-throughput methods to supplement *in vitro* approaches, it also presents a practical impediment to robust determination of error rates. In light of this limitation, TaRTLEt uses conservative analytical and statistical approaches and thresholds, as well as making intermediate analysis steps available to the user for inspection. Our 29-genome validation dataset included 405 riboswitch loci and a total of 655 RNA-seq readsets. Manual curation of TaRTLEt's locus plots led us to exclude one genome (*P. fluorescens*, 10 loci, 40 readsets) as poorly mapped. Among the remaining 28 genomes and 395 loci, TaRTLEt identified 115,130 coverage-change features as potential transcription-termination peaks; comparison to transcriptomic noise allowed us to reject all but 3,437 candidates. In cross-condition significance testing of peak clusters, 1,765 of these candidates lay within the 104 clusters at 95 riboswitch loci that passed our mean and variance test criteria. TaRTLEt calls riboswitch loci as showing evidence of condition-dependent transcription termination based on this latter, most conservative,

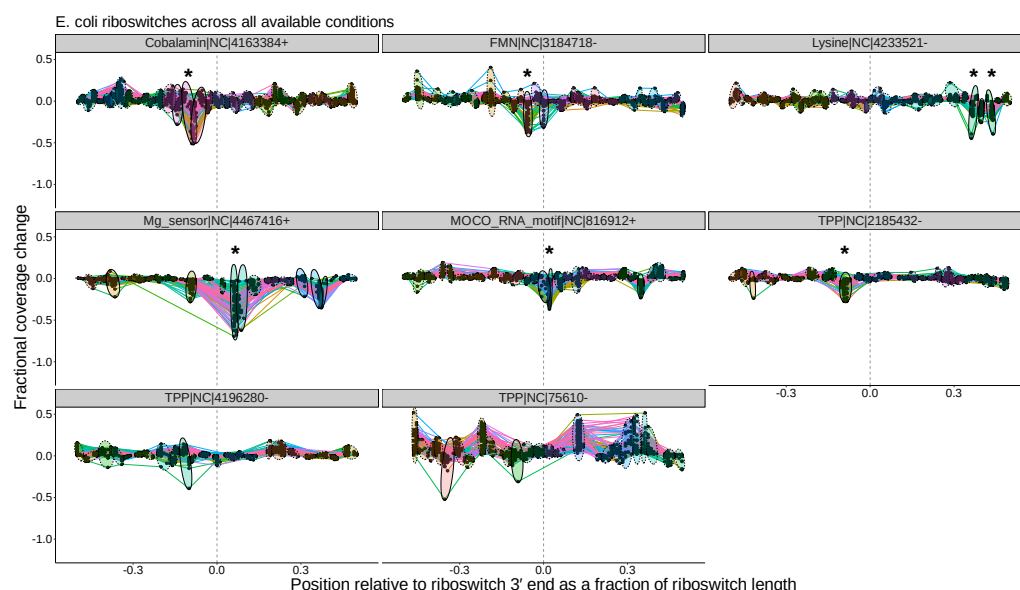


Figure 5 Peak plots (see Fig. 3) for the eight *E. coli* riboswitch loci identified by Infernal in the *E. coli* genome. Peak clusters are grouped by colored ellipses; clusters with at least one significant transcription-termination peak are bounded by solid ellipses, and those with no significant peaks are bounded by dashed ellipses. Asterisks (*) mark peak clusters that (i) contain at least one significant transcription-termination peak, (ii) have significantly lower mean ($p < 0.05$, Mann–Whitney one-tailed U-test) than the set of all other peak clusters, and (iii) have significantly higher variance ($p < 0.05$, Levene’s test) when compared to randomly sampled sets of the same number of peaks.

Full-size [DOI: 10.7717/peerj.19418/fig-5](https://doi.org/10.7717/peerj.19418/fig-5)

most data-rich set, while the within-condition single-peak calls (and supporting data) are made available to the user to support manual curation.

Overall, our results are in good agreement with literature findings (Fig. 6B). TaRTLEt’s peak plots (Fig. 5, Figs. S4, S7) showed clear condition-dependent transcription termination at 10 of the 19 loci known to use this regulatory mode: in *B. anthracis*, the upstream tandem TPP riboswitch (Welz & Breaker, 2007); in *B. subtilis*, the *lysC* lysine, *mgtE* M-box, *xpt* purine, and *samT* and *metK* SAM riboswitches (Sudarsan et al., 2003; Dann et al., 2007; Mandal & Breaker, 2004; Winkler et al., 2003); and in *E. coli* the *btuB* cobalamin, *ribB* FMN, *lysC* lysine, and *thiM* TPP riboswitches (Nou & Kadner, 1998; Hollands et al., 2012; Sedlyarova et al., 2016; Bastet et al., 2017). In addition, TaRTLEt detected no indication (Figs. S4, S6) of transcription-termination activity at either of the two negative-control loci experimentally characterized as acting strictly through mechanisms other than transcription termination (the *B. subtilis* *glmS* and *Enterococcus faecalis* S_{MK} riboswitches (Collins et al., 2007; Fuchs, Grundy & Henkin, 2007; Smith et al., 2010)).

We also examined the validation dataset’s performance across the pipeline to identify the range of situations that can give rise to inconclusive or negative results, as observed at the nine remaining loci in our dataset with previous experimental characterization. First, TaRTLEt cannot provide insight into riboswitch loci with read coverage too low to support statistical analysis. We observed this outcome in 29 of the 395 riboswitch loci investigated

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

Microbe	GEO accession	Citation
<i>Acinetobacter baumannii</i>	GSE183334	Unpublished
<i>Aliivibrio fischeri</i> ES114	GSE237189	Vander Griend et al. (2024)
<i>Apilactobacillus kunkeei</i>	GSE205998	Seeger et al. (2023)
<i>Bacillus anthracis</i>	GSE152356	Corsi et al. (2021)
<i>Bacillus subtilis</i> str. 168	GSE219221	Guo & Herman (2023)
<i>Bacillus subtilis</i> str. 168	GSE226559	Unpublished
<i>Bacteroides fragilis</i>	GSE220692	Fiebig et al. (2024)
<i>Bacteroides thetaiotaomicron</i>	GSE129572	Briliūtė et al. (2019)
<i>Bacteroides thetaiotaomicron</i>	GSE169260	Lewis & Gui (2023)
<i>Bacteroides xylanisolvens</i>	GSE74379	Despres et al. (2016)
<i>Burkholderia pseudomallei</i>	GSE152295	Avican et al. (2021)
<i>Caulobacter vibrioides</i>	GSE241057	McLaughlin, Fiebig & Crosson (2023)
<i>Clostridioides difficile</i>	GSE173804	Weiss et al. (2021)
<i>Desulfovibrio vulgaris</i> str. Hildenborough	GSE101911	Chen et al. (2019)
<i>Desulfovibrio vulgaris</i> str. Hildenborough	GSE78834	Gao et al. (2016)
<i>Enterococcus faecalis</i>	GSE152295	Avican et al. (2021)
<i>Escherichia coli</i>	GSE114358	Guzmán et al. (2019)
<i>Escherichia coli</i>	GSE122211	Sastry et al. (2019)
<i>Escherichia coli</i>	GSE122295	Sastry et al. (2019)
<i>Escherichia coli</i>	GSE122296	Sastry et al. (2019)
<i>Escherichia coli</i>	GSE122779	Anand et al. (2019)
<i>Eubacterium limosum</i>	GSE149269	Jeong et al. (2020)
<i>Klebsiella pneumoniae</i>	GSE229867	Liu et al. (2025)
<i>Mycobacterium tuberculosis</i> str. H37Rv	GSE218354	Martini et al. (2023)
<i>Mycolicibacterium smegmatis</i> MC2 155	GSE232901	Grigorov et al. (2023)
<i>Neisseria gonorrhoeae</i>	GSE191020	Ray et al. (2022)
<i>Piscirickettsia salmonis</i>	GSE235725	Carril et al. (2023)
<i>Pseudomonas chlororaphis</i> subsp. Aureofaciens 30-84	GSE61200	Wang et al. (2016)
<i>Pseudomonas fluorescens</i>	GSE200822	Hasnain et al. (2023)
<i>Salmonella enterica</i> sv. Typhimurium	GSE203342	Kant et al. (2023)
<i>Sinorhizobium meliloti</i>	GSE151705	Fagorzi et al. (2021)
<i>Staphylococcus epidermidis</i>	GSE152295	Avican et al. (2021)
<i>Streptococcus sanguinis</i> SK36	GSE174672	Puccio et al. (2022)
<i>Streptomyces coelicolor</i>	GSE234439	Li et al. (2023)
<i>Synechococcus elongatus</i> PCC7942	GSE227397	Lu et al. (2023)
<i>Synechocystis</i> sp. PCC 6803	GSE132275	Unpublished
<i>Xanthomonas albilineans</i>	GSE229478	Wimmer et al. (2024)
<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC 10331	GSE89651	Park et al. (2022)

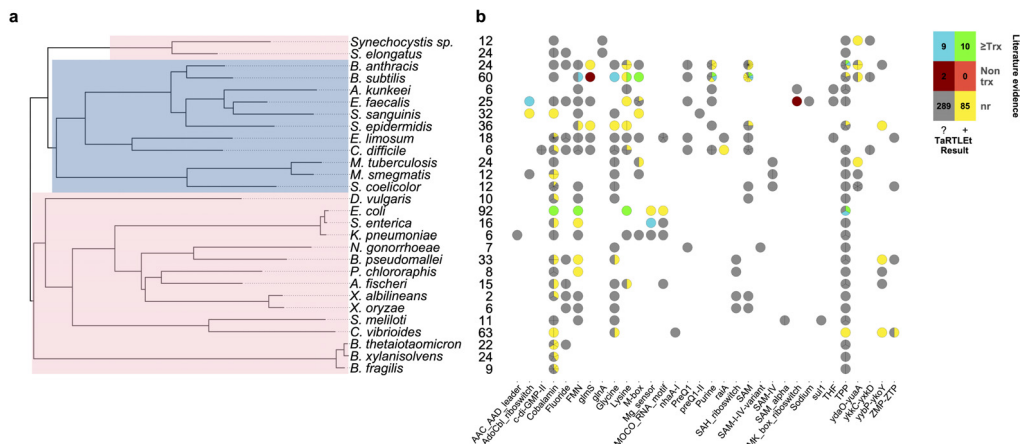


Figure 6 Compilation of TaRTLEt results for a collection of commonly studied bacteria, using RNA-seq data obtained from public GEO repositories. (A) Phylogenetic relationships between bacteria studied with TaRTLEt. The tree is based on GTDB (*Parks et al., 2022; Rinke et al., 2021; Parks et al., 2020; Parks et al., 2018*) bacterial phylogenies (release 214.1) and built using the ggtree R package (*Yu et al., 2017*). Clades highlighted blue and pink represent Gram-positive and Gram-negative taxa respectively. (B) Evidence for condition-dependent transcription termination for every riboswitch representative, as found in the literature and by TaRTLEt. Counts at left give the number of experimental conditions for which paired-end RNA-seq data was available; a wider range of conditions increases the likelihood that the dataset will capture condition-dependent transcription termination. At each locus, evidence reported in the literature may support regulation by transcription termination, potentially among other mechanisms (“≥Trx”), or by non-transcription-termination mechanisms only (“Non trx”), or the locus may not have been characterized *in vitro* (“nr”); and TaRTLEt may be inconclusive (“?”) or provide positive evidence (“+”) for regulation by transcription termination. Where a species encodes multiple riboswitches of a given family, plotting symbols are divided into wedges for the different loci. Counts in the color key are tallies of riboswitch loci in each group.

Full-size DOI: [10.7717/peerj.19418/fig-6](https://doi.org/10.7717/peerj.19418/fig-6)

here, leaving 366 loci tractable for further investigation. Second, TaRTLEt might detect no candidate transcription-termination events in the search space at a riboswitch locus (132 of 366 tractable loci); this could happen either because there is no terminator present in the region of interest, or because under the conditions tested the terminator is always inactive. In our validation set, this group included two loci previously characterized as acting *via* transcription termination, the *B. subtilis metI* SAM riboswitch (Fig. S4) and the *E. faecalis pduQ* AdoCbl riboswitch (Fig. S6) (Winkler *et al.*, 2003; DebRoy *et al.*, 2014).

Third, a locus with transcription-termination peaks might have no clusters that pass 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, 20 of the 234 loci with passing peaks failed this test. This group includes two riboswitch loci (the *E. coli thiC* TPP and *B. subtilis pbuE* purine riboswitches) that have previously been shown to modulate transcription ([Winkler, Nahvi & Breaker, 2002](#); [Chauvier et al., 2017](#); [Mandal & Breaker, 2004](#)) but show no mean fold coverage drop here, suggesting that the range of conditions examined was too narrow to capture the OFF state for these riboswitches ([Fig. 5](#)). Interestingly, the purine ligand does vary enough across the

conditions tested in *B. subtilis* to cause a detectable switch in transcription termination at the *xpt* purine riboswitch (Fig. 6B, Fig. S4), suggesting that these two *B. subtilis* purine riboswitches are responsive to different ligand concentrations.

Finally, among the clusters at a locus that both include transcription-termination peaks and pass the mean test, none may pass the variance test (119 of the remaining 214 loci). A cluster might fail the variance test either if 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, five loci previously found to mediate transcription termination fell into this group: the *B. anthracis* *tenA* TPP (Fig. S5), *B. subtilis* *ribD* FMN, *gcvT* glycine and *yoaD* SAM (Fig. S4), and *S. enterica* *mgtA* Mg^{2+} sensor riboswitches (Fig. S7) (Welz & Breaker, 2007; Wickiser et al., 2005; Mandal et al., 2004; Winkler et al., 2003; Cromie et al., 2006). Two of these, the *B. subtilis* *ribD* FMN riboswitch (Fig. S4, panel FMN|NC|2431617-) and the *S. enterica* *mgtA* Mg^{2+} sensor riboswitch (Fig. S7, panel Mg_sensor|4699497+), appear to fail the variance test by being consistently OFF under the conditions tested; two more (*B. subtilis* *gcvT* glycine and *yoaD* SAM; Fig. S4, panels Glycine|NC|2549606- and SAM|NC|2025251-) are clearly consistent with conditional transcription termination but fall just shy of significance. 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. However, where regulatory mechanism is already established by other means, the signal TaRTLEt exploits may still show riboswitch regulatory state even when the conditions sampled capture only ON or only OFF.

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 nearly all of the genomes in our validation set. *In vitro* 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, Nahvi & Breaker, 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 & Breaker, 2007). Of the 57 TPP riboswitches in the validation dataset, only six (two from *B. anthracis* and one each from *B. subtilis*, *S. epidermidis*, *E. coli*, and *C. vibrioides*) TPP representatives show transcriptional activity under the conditions tested. Intriguingly, the peak plot for *B. anthracis* showed transcriptional activity by the upstream but not the downstream *tenA* TPP riboswitch (Fig. S5, panels TPP|752099+ and TPP|752272+), suggesting that the two tandem riboswitches are responsive under different conditions *in vivo*.

A second very widespread riboswitch class binds cobalamin, with at least one example in all but four genomes in our validation dataset. Previous research has predicted that cobalamin riboswitches in Gram-positive taxa will typically work at the transcriptional level, while those in Gram-negative taxa will regulate translation (Lennon & Batey, 2022). Contra this expectation, we see clear evidence of transcriptionally-active cobalamin

riboswitches in the majority of Gram-negative taxa we investigated (Fig. 6; 15 Gram-negative taxa with cobalamin riboswitch loci, of which ten have at least one positive TaRTLEt result). 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 regulate other aspects of mRNA production, stability, or translation, or regulate transcription termination at a different range of ligand concentrations than this dataset captures. 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 not functionally. Our results indicate that this MOCO riboswitch modulates transcription termination. Across the 32 riboswitch families represented in our validation dataset, we were able to identify previous *in vitro* characterization for members of only 12 families, leaving 20 families completely uncharacterized. We find clear evidence for riboswitch-mediated transcription termination at 12 loci across five of these understudied families (MOCO, *raiA*, *ydaO-yuaA*, *yybP-ykoY*, and ZMP-ZTP families; Fig. 6). That is, TaRTLEt-enabled opportunistic analysis of data originally gathered for other purposes expands by nearly 50% the range of riboswitch families for which we have data-driven insights into regulatory mechanism.

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 eight species in our validation dataset for which the RNA-seq collection included ≤ 9 conditions, TaRTLEt could detect evidence for transcriptional regulation at any riboswitch in only four (Fig. 6B; compare to positive results in 19 of 21 species with data on ≥ 10 conditions). Importantly, our approach supports inference from RNA-seq data collected across multiple different experiments, such that marginal additional data collection for the less well-studied 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 regulatory modes a single riboswitch uses, our high-throughput approach can ask which of any number of riboswitch loci appear to 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*.

ACKNOWLEDGEMENTS

This work is a contribution of the EMERGE Biology Integration Institute. We thank the members of EMERGE-BII, the VirSoil collaboration, and the Bagby lab for helpful discussions. This work used an allocation of computing time from the Ohio Supercomputer Center.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Sarah C. Bagby and Sachit Kshatriya were supported by the Department of Energy Office of Science award number DE-SC0023307 and by the NSF Biology Integration Institutes Program award number 2022070. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
The Department of Energy Office of Science: DE-SC0023307.
The NSF Biology Integration Institutes Program: 2022070.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Sachit Kshatriya conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

- Sarah C. Bagby conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, secured funding, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

All validation data is public; validation data harvesting, processing, and analysis code is available at Github and Zenodo:

- <https://github.com/lepton-7/tartlet-pub>,
- Sachit Kshatriya. (2025). lepton-7/tartlet: v1.0.0: Initial Release (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.14866620>
- Sachit Kshatriya, & Sarah C. Bagby. (2025). lepton-7/tartlet-pub: DOI Release (v1.1). Zenodo. <https://doi.org/10.5281/zenodo.14867191>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.19418#supplemental-information>.

REFERENCES

- Adams PP, Flores Avile C, Popitsch N, Bilusic I, Schroeder R, Lybecker M, Jewett MW. 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 DOI 10.1093/nar/gkw1180.
- Albawi S, Mohammed TA, Al-Zawi S. 2017. Understanding of a convolutional neural network. In: *2017 international conference on engineering and technology (ICET)*. Piscataway: IEEE, 1–6 DOI 10.1109/ICEngTechnol.2017.8308186.
- Amadei F, Reichenbach M, Gallo S, Sigel RK. 2023. The structural features of the ligand-free *moaA* riboswitch and its ion-dependent folding. *Journal of Inorganic Biochemistry* 242:112153 DOI 10.1016/j.jinorgbio.2023.112153.
- Anand A, Olson CA, Yang L, Sastry AV, Catoi E, Choudhary KS, Phaneuf PV, Sandberg TE, Xu S, Hefner Y, Szubin R, Feist AM, Palsson BO. 2019. Pseudogene repair driven by selection pressure applied in experimental evolution. *Nature Microbiology* 4(3):386–389 DOI 10.1038/s41564-018-0340-2.
- Ariza-Mateos A, Nuthanakanti A, Serganov A. 2021. Riboswitch mechanisms: new tricks for an old dog. *Biochemistry* 86:962–975 DOI 10.1134/S0006297921080071.
- Avican K, Aldahdooh J, Togninalli M, Mahmud AF, Tang J, Borgwardt KM, Rhen M, Fällman M. 2021. RNA atlas of human bacterial pathogens uncovers stress dynamics linked to infection. *Nature Communications* 12(1):3282 DOI 10.1038/s41467-021-23588-w.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Seroa N, Davis S, Soboleva A. 2012. NCBI GEO: archive for

- functional genomics data sets—update. *Nucleic Acids Research* **41**(D1):D991–D995 DOI [10.1093/nar/gks1193](https://doi.org/10.1093/nar/gks1193).
- Barrick JE, Breaker RR. 2007.** The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biology* **8**:R239 DOI [10.1186/gb-2007-8-11-r239](https://doi.org/10.1186/gb-2007-8-11-r239).
- Bastet L, Chauvier A, Singh N, Lussier A, Lamontagne A-M, Prévost K, Massé E, Wade JT, Lafontaine DA. 2017.** Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic Acids Research* **45**(12):7474–7486 DOI [10.1093/nar/gkx434](https://doi.org/10.1093/nar/gkx434).
- Bédard A-SV, Hien ED, Lafontaine DA. 2020.** Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochimica Et Biophysica Acta (BBA)—Gene Regulatory Mechanisms* **1863**(3):194501 DOI [10.1016/j.bbagr.2020.194501](https://doi.org/10.1016/j.bbagr.2020.194501).
- Blount KF, Breaker RR. 2006.** Riboswitches as antibacterial drug targets. *Nature Biotechnology* **24**(12):1558–1564 DOI [10.1038/nbt1268](https://doi.org/10.1038/nbt1268).
- Bonfield JK, Marshall J, Danecek P, Li H, Ohan V, Whitwham A, Keane T, Davies RM. 2021.** HTSlib: C library for reading/writing high-throughput sequencing data. *Gigascience* **10**(2):giab007 DOI [10.1093/gigascience/giab007](https://doi.org/10.1093/gigascience/giab007).
- Boussebayle A, Torka D, Ollivaud S, Braun J, Bofill-Bosch C, Dombrowski M, Groher F, Hamacher K, 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 DOI [10.1093/nar/gkz216](https://doi.org/10.1093/nar/gkz216).
- Breaker RR. 2018.** Riboswitches and translation control. *Cold Spring Harbor Perspectives in Biology* **10**(11):a032797 DOI [10.1101/cshperspect.a032797](https://doi.org/10.1101/cshperspect.a032797).
- Briliūtė J, Urbanowicz PA, Luis AS, Baslé A, Paterson N, Rebello O, Hendel J, Ndeh DA, Lowe EC, Martens EC, Spencer DIR, Bolam DN, Crouch LI. 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 DOI [10.1038/s41564-019-0466-x](https://doi.org/10.1038/s41564-019-0466-x).
- 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, Huang L. 2024.** Ribocentre-switch: a database of riboswitches. *Nucleic Acids Research* **52**(D1):D265–D272 DOI [10.1093/nar/gkad891](https://doi.org/10.1093/nar/gkad891).
- Canny J. 1986.** A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-8**(6):679–698 DOI [10.1109/TPAMI.1986.4767851](https://doi.org/10.1109/TPAMI.1986.4767851).
- Caron M-P, Bastet L, Lussier A, Simoneau-Roy M, Massé E, Lafontaine DA. 2012.** Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America* **109**(50):E3444–E3453 DOI [10.1073/pnas.1214024109](https://doi.org/10.1073/pnas.1214024109).
- Carril G, Winther-Larsen HC, Løvoll M, 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**:1253577 DOI [10.3389/fcimb.2023.1253577](https://doi.org/10.3389/fcimb.2023.1253577).

- Chang T-H, Huang H-D, Wu L-C, Yeh C-T, Liu B-J, Horng J-T. 2009. Computational identification of riboswitches based on RNA conserved functional sequences and conformations. *RNA* 15(7):1426–1430 DOI 10.1261/rna.1623809.
- Chauvier A, Picard-Jean F, Berger-Dancause J-C, Bastet L, Naghdi MR, Dubé A, Turcotte P, Perreault J, Lafontaine DA. 2017. Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nature Communications* 8(1):13892 DOI 10.1038/ncomms13892.
- Cheah MT, Wachter A, Sudarsan N, Breaker RR. 2007. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* 447(7143):497–500 DOI 10.1038/nature05769.
- Chen Z, Gao S-h, Jin M, Sun S, Lu J, Yang P, Bond PL, Yuan Z, 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 DOI 10.1016/j.envint.2019.01.058.
- Collins JA, Irnov I, Baker S, Winkler WC. 2007. Mechanism of mRNA destabilization by the *glmS* ribozyme. *Genes & Development* 21(24):3356–3368 DOI 10.1101/gad.1605307.
- Corsi ID, Dutta S, Van Hoof A, Koehler TM. 2021. AtxA-controlled small RNAs of *Bacillus anthracis* virulence plasmid pXO1 regulate gene expression *in trans*. *Frontiers in Microbiology* 11:610036 DOI 10.3389/fmicb.2020.610036.
- Cromie MJ, Shi Y, Latifi T, Groisman EA. 2006. An RNA sensor for intracellular Mg²⁺. *Cell* 125(1):71–84 DOI 10.1016/j.cell.2006.01.043.
- Dann CE, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC. 2007. Structure and mechanism of a metal-sensing regulatory RNA. *Cell* 130(5):878–892 DOI 10.1016/j.cell.2007.06.051.
- DebRoy S, Gebbie M, Ramesh A, Goodson JR, Cruz MR, Van Hoof A, Winkler WC, Garsin DA. 2014. A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. *Science* 345(6199):937–940 DOI 10.1126/science.1255091.
- Despres J, Forano E, Lepercq P, Comtet-Marre S, Jubelin G, Yeoman CJ, Miller MEB, Fields CJ, Terrapon N, Le Bourvellec C, Renard CMGC, Henrissat B, White BA, Mosoni P. 2016. Unraveling the pectinolytic function of *Bacteroides xylanisolvens* using a RNA-seq approach and mutagenesis. *BMC Genomics* 17:147 DOI 10.1186/s12864-016-2472-1.
- Eddy SR, Durbin R. 1994. RNA sequence analysis using covariance models. *Nucleic Acids Research* 22(11):2079–2088 DOI 10.1093/nar/22.11.2079.
- Ellinger E, Chauvier A, Romero RA, Liu Y, Ray S, Walter NG. 2023. Riboswitches as therapeutic targets: promise of a new era of antibiotics. *Expert Opinion on Therapeutic Targets* 27(6):433–445 DOI 10.1080/14728222.2023.2230363.
- Fagorzi C, Bacci G, Huang R, Cangioli L, Checcucci A, Fini M, Perrin E, Natali C, Diczio GC, Mengoni A. 2021. Nonadditive transcriptomic signatures of genotype-by-genotype interactions during the initiation of plant-rhizobium symbiosis. *mSystems* 6(1):e00974-20 DOI 10.1128/msystems.00974-20.

- Fiebig A, Schnizlein MK, Pena-Rivera S, Trigodet F, Dubey AA, Hennessy MK, Basu A, Pott S, Dalal S, Rubin D, Sogin ML, Eren AM, Chang EB, Crosson S. 2024. Bile acid fitness determinants of a *Bacteroides fragilis* isolate from a human pouchitis patient. *mBio* 15(1):e02830-23 DOI 10.1128/mbio.02830-23.
- Fuchs RT, Grundy FJ, Henkin TM. 2007. S-adenosylmethionine directly inhibits binding of 30S ribosomal subunits to the S_{MK} box translational riboswitch RNA. *Proceedings of the National Academy of Sciences of the United States of America* 104(12):4876–4880 DOI 10.1073/pnas.0609956104.
- Gao S-H, Ho JY, Fan L, Richardson DJ, Yuan Z, Bond PL. 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 DOI 10.1128/AEM.01655-16.
- Giarimoglou N, Kouvela A, Maniatis A, Papakyriakou A, Zhang J, Stamatopoulou V, Stathopoulos C. 2022. A riboswitch-driven era of new antibacterials. *Antibiotics* 11(9):1243 DOI 10.3390/antibiotics11091243.
- Gilbert SD, Stoddard CD, Wise SJ, Batey RT. 2006. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *Journal of Molecular Biology* 359(3):754–768 DOI 10.1016/j.jmb.2006.04.003.
- Gong S, Wang Y, Wang Z, Zhang W. 2017. Co-transcriptional folding and regulation mechanisms of riboswitches. *Molecules* 22(7):1169 DOI 10.3390/molecules22071169.
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Research* 31(1):439–441 DOI 10.1093/nar/gkg006.
- Grigorov AS, Skvortsova YV, Bychenko OS, Aseev LV, Koledinskaya LS, Boni IV, Azhikina TL. 2023. Dynamic transcriptional landscape of *Mycobacterium smegmatis* under cold stress. *International Journal of Molecular Sciences* 24(16):12706 DOI 10.3390/ijms241612706.
- Groher F, Suess B. 2014. Synthetic riboswitches—a tool comes of age. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms* 1839(10):964–973 DOI 10.1016/j.bbagrm.2014.05.005.
- Guo T, Herman JK. 2023. Magnesium modulates *Bacillus subtilis* cell division frequency. *Journal of Bacteriology* 205(1):e00375-22 DOI 10.1128/jb.00375-22.
- Gusarov I, Nudler E. 1999. The mechanism of intrinsic transcription termination. *Molecular Cell* 3(4):495–504 DOI 10.1016/S1097-2765(00)80477-3.
- Guzmán GI, Sandberg TE, LaCroix RA, Nyerges Á, Papp H, de Raad M, King ZA, Hefner Y, Northen TR, Notebaart RA, Pál C, Pálsson BO, Papp B, Feist AM. 2019. Enzyme promiscuity shapes evolutionary innovation and optimization. *Molecular Systems Biology* 15:e8462 DOI 10.15252/msb.20188462.
- Hanson S, Berthelot K, Fink B, McCarthy JE, Suess B. 2003. Tetracycline-aptamer-mediated translational regulation in yeast. *Molecular Microbiology* 49(6):1627–1637 DOI 10.1046/j.1365-2958.2003.03656.x.
- Hasnain A, Balakrishnan S, Joshy DM, Smith J, Haase SB, Yeung E. 2023. Learning perturbation-inducible cell states from observability analysis of transcriptome dynamics. *Nature Communications* 14:3148 DOI 10.1038/s41467-023-37897-9.

- Hollands K, Proshkin S, Sklyarova S, Epshtein V, Mironov A, Nudler E, Groisman EA. 2012. Riboswitch control of Rho-dependent transcription termination. *Proceedings of the National Academy of Sciences of the United States of America* 109(14):5376–5381 DOI 10.1073/pnas.1112211109.
- Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119 DOI 10.1186/1471-2105-11-119.
- Jeong J, Kim J-Y, Park B, Choi I-G, Chang IS. 2020. Genetic engineering system for syngas-utilizing acetogen, *Eubacterium limosum* KIST612. *Bioresource Technology Reports* 11:100452 DOI 10.1016/j.biteb.2020.100452.
- Kant S, Till JKA, Liu L, Margolis A, Uppalapati S, Kim J-S, 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 DOI 10.1371/journal.pbio.3002051.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37(8):907–915 DOI 10.1038/s41587-019-0201-4.
- Klein DJ, Ferré-D'Amaré AR. 2006. Structural basis of *glmS* ribozyme activation by glucosamine-6-phosphate. *Science* 313(5794):1752–1756 DOI 10.1126/science.1129666.
- Lennon SR, Batey RT. 2022. Regulation of gene expression through effector-dependent conformational switching by cobalamin riboswitches. *Journal of Molecular Biology* 434(18):167585 DOI 10.1016/j.jmb.2022.167585.
- Lewis JP, Gui Q. 2023. Iron deficiency modulates metabolic landscape of *Bacteroidetes* promoting its resilience during inflammation. *Microbiology Spectrum* 11(4):e04733-22 DOI 10.1128/spectrum.04733-22.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079 DOI 10.1093/bioinformatics/btp352.
- Li C, Urem M, Du C, Zhang L, Van Wezel GP. 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 DOI 10.1128/aem.01674-23.
- Liu Y, Niu Z, Wu R, Yang D, Chen J, Liu G, Zhao J. 2025. Transcriptomic analysis reveals pathways underlying the multi-antibiotic resistance of *Klebsiella pneumoniae*. *IET Systems Biology* 19(1):e12112 DOI 10.1049/syb2.12112.
- Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser C, Cossart P, Johansson J. 2009. A *trans*-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell* 139(4):770–779 DOI 10.1016/j.cell.2009.08.046.
- Lu K-J, Chang C-W, Wang C-H, Chen FY, Huang IY, Huang P-H, Yang C-H, Wu H-Y, Wu W-J, Hsu K-C, Ho M-C, Tsai M-D, Liao JC. 2023. An ATP-sensitive

- phosphoketolase regulates carbon fixation in cyanobacteria. *Nature Metabolism* 5(7):1111–1126 DOI 10.1038/s42255-023-00831-w.
- Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR. 2003. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* 113(5):577–586 DOI 10.1016/S0092-8674(03)00391-X.
- Mandal M, Breaker RR. 2004. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature Structural & Molecular Biology* 11:29–35 DOI 10.1038/nsmb710.
- Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306(5694):275–279 DOI 10.1126/science.1100829.
- Martini BA, Grigorov AS, Skvortsova YV, Bychenko OS, Salina EG, Azhikina TL. 2023. Small RNA MTS1338 configures a stress resistance signature in *Mycobacterium tuberculosis*. *International Journal of Molecular Sciences* 24(9):7928 DOI 10.3390/ijms24097928.
- McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. 2017. Riboswitch diversity and distribution. *RNA* 23(7):995–1011 DOI 10.1261/rna.061234.117.
- McLaughlin M, Fiebig A, Crosson S. 2023. XRE transcription factors conserved in *Caulobacter* and ϕ CbK modulate adhesin development and phage production. *PLOS Genetics* 19(11):e1011048 DOI 10.1371/journal.pgen.1011048.
- Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E. 2002. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 111(5):747–756 DOI 10.1016/S0092-8674(02)01134-0.
- Musich R, Cadle-Davidson L, Osier MV. 2021. Comparison of short-read sequence aligners indicates strengths and weaknesses for biologists to consider. *Frontiers in Plant Science* 12:657240 DOI 10.3389/fpls.2021.657240.
- Nawrocki EP. 2014. Annotating functional RNAs in genomes using Infernal. In: Gorodkin J, Ruzzo WL, eds. *RNA sequence, structure, and function: computational and bioinformatic methods. Methods in Molecular Biology*, vol. 1097. Totowa, NJ: Humana Press, 163–197.
- Nawrocki EP, Eddy SR. 2013a. Computational identification of functional RNA homologs in metagenomic data. *RNA Biology* 10(7):1170–1179 DOI 10.4161/rna.25038.
- Nawrocki EP, Eddy SR. 2013b. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29(22):2933–2935 DOI 10.1093/bioinformatics/btt509.
- Nou X, Kadner RJ. 1998. Coupled changes in translation and transcription during cobalamin-dependent regulation of *btuB* expression in *Escherichia coli*. *Journal of Bacteriology* 180(24):6719–6728 DOI 10.1128/jb.180.24.6719-6728.1998.
- Nou X, Kadner RJ. 2000. Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proceedings of the National Academy of Sciences of the United States of America* 97(13):7190–7195 DOI 10.1073/pnas.130013897.
- Park J, Jung H, Mannaa M, Lee SY, Lee H-H, Kim N, Han G, Park D-S, Lee S-W, Lee S-W, Seo Y-S. 2022. Genome-guided comparative *in planta* transcriptome analyses

- p for identifying cross-species common virulence factors in bacterial phytopathogens.
- Frontiers in Plant Science*
- 13**
- :1030720 DOI
- [10.3389/fpls.2022.1030720](https://doi.org/10.3389/fpls.2022.1030720)
- .
- Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020.** A complete domain-to-species taxonomy for bacteria and archaea. *Nature Biotechnology* **38**(9):1079–1086 DOI [10.1038/s41587-020-0501-8](https://doi.org/10.1038/s41587-020-0501-8).
- Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, 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 DOI [10.1093/nar/gkab776](https://doi.org/10.1093/nar/gkab776).
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarszewski A, Chaumeil P-A, Hugenholtz P. 2018.** A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology* **36**(10):996–1004 DOI [10.1038/nbt.4229](https://doi.org/10.1038/nbt.4229).
- Puccio T, An S-S, Schultz AC, Lizarraga CA, Bryant AS, Culp DJ, Burne RA, 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 DOI [10.1111/mmi.14854](https://doi.org/10.1111/mmi.14854).
- Ray JC, Smirnov A, Maurakis SA, Harrison SA, Ke E, Chazin WJ, Cornelissen CN, Criss AK. 2022.** Adherence enables *Neisseria gonorrhoeae* to overcome zinc limitation imposed by nutritional immunity proteins. *Infection and Immunity* **90**(3):e00009–22 DOI [10.1128/iai.00009-22](https://doi.org/10.1128/iai.00009-22).
- Rinke C, Chuvochina M, Mussig AJ, Chaumeil P-A, Davin AA, Waite DW, Whitman WB, Parks DH, Hugenholtz P. 2021.** A standardized archaeal taxonomy for the genome taxonomy database. *Nature Microbiology* **6**(7):946–959 DOI [10.1038/s41564-021-00918-8](https://doi.org/10.1038/s41564-021-00918-8).
- Roberts JW. 2019.** Mechanisms of bacterial transcription termination. *Journal of Molecular Biology* **431**(20):4030–4039 DOI [10.1016/j.jmb.2019.04.003](https://doi.org/10.1016/j.jmb.2019.04.003).
- Rosinski-Chupin I, Sauvage E, Fouet A, Poyart C, Glaser P. 2019.** Conserved and specific features of *Streptococcus pyogenes* and *Streptococcus agalactiae* transcriptional landscapes. *BMC Genomics* **20**:236 DOI [10.1186/s12864-019-5613-5](https://doi.org/10.1186/s12864-019-5613-5).
- Rosinski-Chupin I, Soutourina O, Martin-Verstraete I. 2014.** Riboswitch discovery by combining RNA-seq and genome-wide identification of transcriptional start sites. In: Burke-Aguero DH, ed. *Methods in Enzymology: Riboswitch Discovery, Structure and Function*. Vol. 549. Cambridge: Academic Press, 3–27.
- Sastry AV, Gao Y, Szubin R, Hefner Y, Xu S, Kim D, Choudhary KS, Yang L, King ZA, Palsson BO. 2019.** The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nature Communications* **10**:5536 DOI [10.1038/s41467-019-13483-w](https://doi.org/10.1038/s41467-019-13483-w).
- Sedlyarova N, Shamovsky I, Bharati BK, Epshtein V, Chen J, Gottesman S, Schroeder R, Nudler E. 2016.** sRNA-mediated control of transcription termination in *E. coli*. *Cell* **167**(1):111–121 DOI [10.1016/j.cell.2016.09.004](https://doi.org/10.1016/j.cell.2016.09.004).

- Seeger C, Dyrhage K, Näslund K, Andersson SG. 2023. *Apilactobacillus kunkeei* releases RNA-associated membrane vesicles and proteinaceous nanoparticles. *MicroLife* 4:uqad037 DOI 10.1093/femsm/luqad037.
- Smith AM, Fuchs RT, Grundy FJ, Henkin TM. 2010. The SAM-responsive S_{MK} box is a reversible riboswitch. *Molecular Microbiology* 78(6):1393–1402 DOI 10.1111/j.1365-2958.2010.07410.x.
- Stav S, Atilho RM, Mirihana Arachchilage G, Nguyen G, Higgs G, Breaker RR. 2019. Genome-wide discovery of structured noncoding RNAs in bacteria. *BMC Microbiology* 19:66 DOI 10.1186/s12866-019-1433-7.
- Sudarsan N, Cohen-Chalamish S, Nakamura S, Emilsson GM, Breaker RR. 2005. Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chemistry & Biology* 12(12):1325–1335 DOI 10.1016/j.chembiol.2005.10.007.
- Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. 2003. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & Development* 17(21):2688–2697 DOI 10.1101/gad.1140003.
- Sun EI, Rodionov DA. 2014. Computational analysis of riboswitch-based regulation. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms* 1839(10):900–907 DOI 10.1016/j.bbagr.2014.02.011.
- Tan G, Opitz L, Schlapbach R, Rehauer H. 2019. Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports* 9:2856 DOI 10.1038/s41598-019-39076-7.
- Vander Griend JA, Isenberg RY, Kotla KR, Mandel MJ. 2024. Transcriptional pathways across colony biofilm models in the symbiont *Vibrio fischeri*. *mSystems* 9(1):e00815-23 DOI 10.1128/msystems.00815-23.
- Wachsmuth M, Findeiß S, Weissheimer N, Stadler PF, Mörl M. 2013. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research* 41(4):2541–2551 DOI 10.1093/nar/gks1330.
- Wang D, Yu JM, Dorosky RJ, Pierson III LS, Pierson EA. 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 DOI 10.1371/journal.pone.0148003.
- Weiss A, Lopez CA, Beavers WN, Rodriguez J, Skaar EP. 2021. *Clostridioides difficile* strain-dependent and strain-independent adaptations to a microaerobic environment. *Microbial Genomics* 7(12):000738 DOI 10.1099/mgen.0.000738.
- Welz R, Breaker RR. 2007. Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA* 13(4):573–582 DOI 10.1261/rna.407707.
- Wickiser JK, Winkler WC, Breaker RR, Crothers DM. 2005. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular Cell* 18(1):49–60 DOI 10.1016/j.molcel.2005.02.032.
- Wimmer F, Englert F, Wandera KG, Alkhnbashi OS, Collins SP, Backofen R, Beisel CL. 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 DOI [10.1093/nar/gkad1097](https://doi.org/10.1093/nar/gkad1097).
- Winkler W, Nahvi A, Breaker RR. 2002.** Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**(6910):952–956 DOI [10.1038/nature01145](https://doi.org/10.1038/nature01145).
- Winkler WC, Nahvi A, Sudarsan N, Barrick JE, Breaker RR. 2003.** An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Structural & Molecular Biology* **10**(9):701–707 DOI [10.1038/nsb967](https://doi.org/10.1038/nsb967).
- Yao Z, Barrick J, Weinberg Z, Neph S, Breaker R, Tompa M, Ruzzo WL. 2007.** A computational pipeline for high-throughput discovery of *cis*-regulatory noncoding RNA in prokaryotes. *PLOS Computational Biology* **3**(7):e126 DOI [10.1371/journal.pcbi.0030126](https://doi.org/10.1371/journal.pcbi.0030126).
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-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 DOI [10.1111/2041-210X.12628](https://doi.org/10.1111/2041-210X.12628).
- Yu S-H, Vogel J, Förstner KU. 2018.** ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *Gigascience* **7**(9):giy096 DOI [10.1093/gigascience/giy096](https://doi.org/10.1093/gigascience/giy096).