RNAs are well-suited to act as cellular sensors that detect and respond to metabolite changes in the environment, due to their ability to fold into complex structures. Here, we introduce a genome-wide strategy called PARCEL that experimentally identifies RNA aptamers in vitro, in a high-throughput manner. By applying PARCEL to a collection of prokaryotic and eukaryotic organisms, we have revealed 58 new RNA aptamers to three key metabolites, greatly expanding the list of natural RNA aptamers. The newly identified RNA aptamers exhibit significant sequence conservation, are highly structured and show an unexpected prevalence in coding regions. We identified a prokaryotic precursor tmRNA that binds vitamin B2 (FMN) to facilitate its maturation, as well as eukaryotic mRNAs that bind and respond to FMN, suggesting FMN as the second RNA-binding ligand to affect eukaryotic expression. PARCEL results show that RNA-based sensing and gene regulation is more widespread than previously appreciated in different organisms.
Microrganisms are constantly sensing their environment for changes in temperature, pH, metabolites, and nutrients so as to regulate their gene expression programs to best adapt to different signals for growth, survival, and virulence. As such, comprehensive mapping of regulatory networks in microbes under different environmental conditions is crucial to understanding their biology. While many of these regulators have been extensively studied at the protein and DNA levels, RNA’s role as direct sensors and responders remains relatively under-explored. One class of cellular RNA sensors, known as riboswitches, can recognize and respond to specific metabolites by altering gene expression. Upon binding to their ligands, riboswitches undergo conformational changes that result in the regulation of gene expression through diverse means, such as changes in transcription, translation and decay, informing the host of its environmental conditions. The modularity of riboswitches also allows them to be transplantable to different systems, broadening their use as biological sensors.

While in vitro selection methods, such as SELEX, have been applied with variable success to generate new synthetic RNA aptamers, the ability to comprehensively identify natural RNA aptamers from transcriptomes would expand our toolbox for synthetic biology and deepen our understanding of RNA-based gene regulation in vivo. Currently, potential natural RNA aptamers are mostly identified through computational determination of sequence and structural homology to known riboswitches. However, as RNA can adopt different folds for binding to the same ligand and organisms can diverge greatly in sequence content, computational means of searching for riboswitches through sequence homology have limited scope, and strategies that allow direct experimental detection are highly desirable. One recent strategy, Term-seq, utilizes high-throughput sequencing to detect differential transcription termination events in bacteria under different conditions. However, complementary strategies for detecting riboswitches that act through other mechanisms (such as translation inhibition), as well as riboswitches that bind to metabolites whose intracellular concentrations are not easily altered, need to be developed.

Here, we report an in vitro method for experimentally identifying RNA aptamers in transcriptomes by detecting ligand-induced RNA structural changes using high-throughput sequencing (Fig. 1a). This method allows us to rapidly screen through transcriptomes to identify natural RNA aptamers toward almost any ligand of choice. Specifically, we extract total RNA from organisms grown under different conditions and probe their structures in the presence or absence of different metabolites by using a double-stranded nuclease, RNase V1, which recognizes and cleaves at base-paired regions in RNAs. The different cleavage sites, in the presence and absence of metabolites, are cloned into cDNA libraries for deep sequencing. Subtle differences between these two libraries point to the few true ligand-specific structure changing RNA elements in the genome and we developed a sensitive and robust computational analysis pipeline to identify these (Methods). This experimental and computational approach, termed Parallel Analysis of RNA Conformations Exposed to Ligand binding (PARCEL), revealed the breadth of RNA-ligand interactions in prokaryotic and eukaryotic transcriptomes, identifying many new natural RNA aptamers in the process.

**Results**

**Development of PARCEL.** To establish PARCEL, we systematically tested different structure-probing strategies to determine the approach that best captures ligand-induced structural changes genome-wide, allowing for a simplified, cost-effective workflow without multiple probing assays. Strategies using double-strand or single-strand specific nucleases (RNase V1 and S1 nuclease, respectively), as well as in-line probing which probes nucleotide flexibility, were tested for their abilities to detect structural changes using high-throughput sequencing (Fig. 1b, Supplementary Fig. 1, 2, 3). The known thiamine pyrophosphate (TPP) and S-adenosylmethionine (SAM) riboswitches were used as positive controls and other RNA sequences not known to bind TPP or SAM were used as negative controls in this experiment. As expected, the known riboswitches showed large structural changes upon ligand binding (Fig. 1b, Supplementary Fig. 1, 2), while the negative controls did not (Supplementary Fig. 3), indicating that the structure changes captured by nuclease digestion followed by sequencing are highly specific. These structural differences can be less pronounced in libraries prepared using in-line probing (Fig. 1b, Supplementary Fig. 2). This is likely due to the noise introduced by the additional 5′ phosphorylation step that is used in in-line probing library preparation. As degraded cellular RNAs are also phosphorylated and cloned into the library, it is difficult to distinguish in-line probed fragments from degradation fragments. Among the nucleases, we observed a higher degree of correlation between two biological replicates of RNase V1 (R = 0.99) versus S1 nuclease libraries (R = 0.66, Supplementary Fig. 4a), a key feature for differential analysis. The structural changes observed using the nucleases could also be reproduced using low-throughput RNA footprinting and mapped to the secondary structure of the TPP riboswitch (Supplementary Fig. 1b-d). Correspondingly, we selected RNase V1 as the probing strategy of choice in all PARCEL experiments to identify natural RNA aptamers genome-wide.

As *Bacillus subtilis* and *Pseudomonas aeruginosa* are bacteria for which many riboswitches are known, we performed structure probing in the presence and absence of key metabolites known to interact with RNAs in their transcriptomes. To maximize our chances of finding RNA aptamers which may only be expressed under specific conditions, we grew the bacteria in rich or minimal media to exponential or stationary phases (Methods). We then extracted total RNA from the pooled bacteria, performed ribosomal depletion to enrich for mRNAs, and did structure probing, followed by deep sequencing. We had two biological replicates for each experiment and obtained more than 7 million reads per replicate (Supplementary Table 1). RNA aptamers that bind specifically to one ligand should not recognize other unrelated ligands. As such, we developed a novel computational pipeline to identify contiguous positions of structural change (to increase signal-to-noise ratio) that show statistically different numbers of reads that indicate base pairing in one metabolite condition but not in others (Supplementary Fig. 4b, Methods). This approach aggregates signals of variation in each base pair across conditions to define ligand-responsive regions using dynamic programming, and combines this with the computation of a BLAST-like E-value to identify RNA aptamers with statistical confidence (Methods).

**PARCEL finds known riboswitches in *B. subtilis* and *P. aeruginosa*.** To evaluate PARCEL, we first determined if we could identify known riboswitches that have been previously reported in the literature. We identified 17 out of 20 known riboswitches that interact with key metabolites (TPP, FMN, and SAM) in *B. subtilis* and *P. aeruginosa*, including 4/5 known TPP riboswitches, 2/2 FMN riboswitches, 9/11 SAM riboswitches in *B. subtilis*, as well as 1/1 TPP and 1/1 FMN riboswitches in *P. aeruginosa*, highlighting the high sensitivity of the method (85%; Fig. 1c, Supplementary Fig. 5a-d). Furthermore, pair-wise analysis of control libraries that were generated using the same metabolite did not identify any candidate RNA aptamers, indicating that the approach is highly specific to the presence of metabolites.
We noted that undetected known riboswitches have low numbers of RNase V1 and S1 nuclease reads along their regulons, indicating that they are either poorly expressed or present in nuclease-inaccessible regions (Supplementary Fig. 5f, g).

Besides detecting structural changes upon ligand binding, we also determined whether PARCEL can detect RNA-ligand interactions quantitatively. As TPP riboswitches bind most strongly to TPP, followed by thiamine and oxythiamine, we tested the sensitivity of PARCEL in detecting RNA structural changes due to differences in ligand binding affinities. Indeed, we observed the strongest structural change in TPP riboswitches upon binding to TPP, followed by thiamine and then oxythiamine (Fig. 1d, Supplementary Fig. 6a). We also treated B. subtilis...
newly identified upstream of their operons, we observed that 28% and 47% of the induced structure changes identified by PARCEL were found in coding regions, respectively (Fig. 2b, Supplementary Table 2). RNA footprinting validated four out of six ligand-induced structure changes identified by PARCEL (Supplementary Fig. 7a-d). While known riboswitches are mostly located upstream of their operons, we observed that 28% and 47% of the newly identified B. subtilis and P. aeruginosa RNA aptamers are found in coding regions, respectively (Fig. 2b, Supplementary Fig. 8). Two of our validated RNA aptamers fall in coding regions (Supplementary Fig. 7a, c), indicating that these coding regions exhibit real structural changes in the presence of TPP.

To study the structural properties of these RNA aptamers, we utilized the program Alifoldz to calculate folding energies across their RNA orthologs in different bacterial species (Methods). The newly identified elements exhibit lower folding energies than their dinucleotide shuffled controls, indicating that they are more structured (Fig. 2c). As functionally important RNAs frequently show evolutionary constraints in their sequences, we calculated the nucleotide substitution rate of our new RNA aptamers, as compared to other synonymous regions and UTRs. Similar to known riboswitches, the novel RNA aptamers show a significant reduction in nucleotide substitution rate (Fig. 2d, e), suggesting that they are evolutionarily conserved and likely to be functional.

We observed that one of our prokaryotic metabolite-sensitive regions encodes a small non-coding RNA, specifically a tmRNA (Fig. 3a). The B. subtilis tmRNA is transcribed from two promoters to produce both long and short precursor RNAs, which are then processed into the mature tmRNA. To understand the functional role of metabolite sensing in tmRNA, we cloned and in vitro transcribed all three isoforms of tmRNA, then performed structure probing in the presence and absence of FMN in the dark, to avoid FMN-induced photocleavage of RNAs. Interestingly, we observed FMN-induced structure changes in both precursor forms of tmRNA, but not in the mature form (Fig. 3b, c, d, Supplementary Fig. 7d), highlighting that it is the precursor tmRNA that responds to FMN. To determine whether PARCEL finds new eukaryotic FMN aptamers in Candida albicans. To date, only riboswitches that bind to TPP have been found in eukaryotes, and they regulate splicing and 3' UTR usage. Identifying new eukaryotic riboswitches is important for broadening our understanding of eukaryotic gene regulation. To maximize our chances of finding eukaryotic riboswitches, we screened the fungal pathogen, C. albicans, using a pool of metabolites that correspond to highly abundant classes of riboswitches in bacteria, including FMN, SAM, glycine, lysine, and vitamin B12 (AdoCbl). PARCEL identified 23 new RNA aptamers that exhibited structural changes in the presence of the metabolite pool (Supplementary Table 3). 87% of the new C. albicans RNA aptamers reside in coding regions (Fig. 4a, Supplementary Fig. 8), in contrast to known riboswitches and new prokaryotic aptamers, indicating that they may have different functions from classical riboswitches. We validated seven out of nine PARCEL-identified structural changes by performing in-line probing of these novel RNA aptamers in the presence of the pooled metabolites (Supplementary Fig. 9–11), all of which fall in the coding regions of these genes, confirming that the PARCEL-detected structural changes are real. Similar to their prokaryotic counterparts, the eukaryotic RNA aptamers were found to be significantly more structured compared to dinucleotide shuffled controls, suggesting that structure is likely to be important for their function (Fig. 4b). As many of the new eukaryotic RNA aptamers are located in highly conserved coding regions, we observed an expected increase in conservation of these elements as compared to UTRs, but not a further reduced nucleotide substitution rate compared to other coding sequences (Fig. 4c).
**Fig. 2** PARCEL identifies new RNA aptamers in bacterial species. **a** PARCEL identifies a total of 52 RNA aptamers in *B. subtilis* and *P. aeruginosa*. Black and white bars indicate the numbers of known riboswitches and novel aptamers that are identified in our study, respectively. **b** Distribution of known riboswitches and new RNA aptamers along the 5′ UTR, CDS, and 3′ UTR regions for *B. subtilis* and *P. aeruginosa*, showing that a substantial proportion of RNA aptamers are located in the 3′ UTR and CDS regions. **c** Comparison of score distribution of Alifoldz12 for RNA aptamers vs. shuffled counterparts. The upper, middle, and lower bounds of the boxplot represent the 75, 50, and 25th percentile of the values, respectively. A negative score indicates a stable, conserved consensus structure. p-value was obtained using the non-parametric Kolmogorov–Smirnov test. **d, e** Comparison of the nucleotide substitution rate (number of substitutions per base-pair) for new RNA aptamers in coding region (KrCDS), new RNA aptamers in UTR (KrUTR), 3′ UTR (K3UTR), 5′ UTR (K5UTR), synonymous sites (Ks), and non-synonymous sites (Ka). The upper, middle, and lower bounds of the boxplot represent the 75, 50, and 25th percentile of the values, respectively. To calculate nucleotide substitutions, *B. subtilis* 168 was compared to *B. subtilis* subsp. *spizizenii* W23 (d), and *P. aeruginosa* PAO1 was compared to *P. aeruginosa* PA7 (e). Note that Krknown denotes the substitution rate of known riboswitches in *B. subtilis* (15 in total) as annotated in the RegPrecise database11. Krknown was not calculated in *P. aeruginosa* as there are too few known TPP and FMN riboswitches. p-values were calculated using the non-parametric Kolmogorov–Smirnov test.
Precursor tmRNA can act as an RNA sensor for FMN. a Schematic of the *B. subtilis* tmRNA genomic locus (top) and predicted secondary structures of the long precursor tmRNA, short precursor tmRNA, and mature tmRNA, using the RNAfold program20 (bottom). b RNA footprinting analysis of the long precursor tmRNA, using a SHAPE-like chemical (NAI), in the presence (lane 4) and absence (lane 3) of 100 μM FMN. Also shown are A ladder (lane 1) and unmodified RNA (lane 2). The red bar indicates bases that become more single-stranded in the presence of FMN. c Predicted secondary structure of the tmRNA long precursor using the RNAfold program20. The red bases correspond to the positions marked by the red bar in b. d Average footprinting analysis (n = 3, SAFA) of mature (top), short precursor (middle), and long precursor tmRNA (bottom), in the presence (red) and absence (black) of 100 μM FMN, in the dark. The beige box indicate the region of increased flexibility in the precursor tmRNAs in the presence of FMN. The stars indicate bases that show statistically significant changes with FMN (p ≤ 0.05, Student t-test). e qPCR analysis of the mRNA expression level of precursor tmRNA and mature tmRNA, across six biological replicates, after addition of 100 μM of riboflavin to the growth media of *B. subtilis*. Fold changes are normalized to the negative control Veg gene. The known *B. subtilis* FMN riboswitch is used as the positive control. p-values were calculated by Student’s t-test, the error bars indicate standard deviation of the replicates.

of FMN could alter gene expression changes of RPS31 or ATP1 in vivo, we integrated FLAG-tagged *C. albicans* RPS31 or ATP1 in a *S. cerevisiae* FMN1 synthase deletion mutant (*fmn1Δ*) as *S. cerevisiae* is known to take up exogenous FMN, unlike *C. albicans*. Both transcript and protein levels of FLAG-tagged *C. albicans* RPS31 and ATP1 were measured, following growth of the integrated strains under varying FMN concentrations. We found that while transcript levels of FLAG-tagged *C. albicans* RPS31 and ATP1 in *fmn1Δ* did not change with increasing FMN concentrations (Supplementary Fig. 10c, 11b), RPS31 and ATP1 protein levels were decreased and increased, respectively (Fig. 4e, Supplementary Fig. 10d-f, 11c), suggesting that FMN sensing could have a regulatory effect on gene expression in vivo, and that
**Fig. 4** PARCEL identifies new RNA aptamers in *Candida albicans*. a Pie chart of the number of *C. albicans* RNA aptamers that are located in 5' UTR, CDS, and 3' UTR. The majority of *C. albicans* RNA aptamers are found in CDSs. b Comparison of the distribution of Alifoldz scores for RNA aptamers vs. shuffled counterpart. The upper, middle, and lower bounds of the boxplot represent the 75th, 50th, and 25th percentile of the values, respectively. A negative score indicates a stable, conserved consensus structure. The *p*-value was obtained using the non-parametric Kolmogorov-Smirnov test. c Nucleotide substitution rates, calculated as the number of substitutions per base-pair, for RNA aptamers (K_r), 3' UTR (K_3UTR), 5' UTR (K_5UTR), synonymous sites (K_s), and non-synonymous sites (K_a). The upper, middle, and lower bounds of the boxplot represent the 75th, 50th, and 25th percentile of the values, respectively. *C. albicans* SC5314 was compared to *Candida dubliniensis* for the calculation. The *p*-value was obtained using the non-parametric Kolmogorov-Smirnov test. d Gel analysis of RPS31 mRNA using in-line probing (left) and RNase V1 (right) in the presence (lane 3) and absence (lane 2) of 100 μM FMN. The A ladder (A, lane 1) is also shown. The black arrows indicate positions along the RNA that changed in the presence of FMN. e A representative Western blot showing RPS31::FLAG (top) and loading (bottom) protein levels in RPS31::FLAG knock-in strains with WT (left) and *fmn1Δ* (right) backgrounds cultured at different FMN concentrations (mM). Using *t*-test (*n* = 8), significant *p*-values of 0.009 and 0.01 (for 2.5 and 5.0 mM against 10.0 mM, respectively) were determined for *fmn1Δ*, but not WT (*p*-values of 0.2 and 0.5). f Gel analysis of RPS31 mRNA using in-line probing in the presence of 20, 100, or 500 μM of FMN, FAD or riboflavin. In-line probing of RNA in the absence of metabolite (H_2O, lane 2) and A ladder (A, lane 1) are also shown. g SAFA analysis of WT RPS31 (top) and codon-optimized RPS31 (bottom) in the presence (red line) and absence (black line) of 100 μM FMN. The beige box indicates the region of structural change in WT RPS31 when it interacts with FMN. This structural change is absent in the codon-optimized RPS31. h A representative Western blot showing codon-optimized RPS31::FLAG (top) and loading (bottom) protein levels in codon-optimized RPS31::FLAG knock-in strains with WT (left) and *fmn1Δ* (right) backgrounds cultured at different FMN concentrations (mM). Using *t*-test (*n* = 3), calculated *p*-values for 2.5 and 5.0 mM were insignificant for both *fmn1Δ* (both 0.7) and WT (0.9 and 0.09).
these genes could represent the first known eukaryotic FMN riboswitches.

To further understand ligand binding specificities and affinities of this putative FMN riboswitch, we performed detailed structural studies on the RPS31 transcript. We observed that RPS31 RNA binds specifically to FMN and does not respond to structurally similar analogs, such as riboflavin and FAD (Fig. 4f, Supplementary Fig. 11d). Integrating double-stranded (RNase V1), single-stranded (S1 nuclease) structure probing and in-line probing information along the length of the RPS31 transcript into the RNAfold prediction software showed that RPS31 RNA consists of seven stems around a central loop, and that FMN binding results in extensive structural rearrangements (Supplementary Fig. 12a). The FMN-bound RPS31 aptamer consists of six stems around the central FMN bound loop, and appears to resemble the prokaryotic FMN riboswitch structure36. This attests to the structural plasticity of RNA molecules, whereby different sequences can be utilized to form similar structures for cellular function.

To further test whether the FMN-induced change in RPS31 protein levels is mediated post-transcriptionally, we designed a codon-optimized version of C. albicans RPS31 (by changing nucleotides of synonymous bases) to disrupt RPS31 RNA structure without altering its protein sequence. As expected, codon-optimized RPS31 mRNA is structurally different from wildtype RPS31 mRNA, and does not show structure changes in the presence and absence of FMN (Fig. 4g). Codon-optimized RPS31 maintained similar mRNA and protein levels at varying concentrations of FMN (Fig. 4h, Supplementary Fig. 12b-d), supporting the hypothesis that the binding of FMN to native C. albicans RPS31 RNA results in post-transcriptional regulation of RPS31 protein levels.

**Discussion**

In summary, we have developed a new strategy named PARCEL that experimentally identifies RNA aptamers transcriptome-wide by detecting ligand-induced structure changes. As PARCEL allows us to rapidly screen through transcriptomes to identify RNA aptamers, we identified a total of 58 novel candidate RNA aptamers in two prokaryotic and one eukaryotic species, including a second class of putative eukaryotic riboswitches. Unlike known riboswitches described in the literature, the newly identified aptamers reside in both UTRs and coding sequences, and are not necessarily linked to the biosynthetic pathways of their respective ligands. Further characterization of three new RNA aptamers showed that they could be riboswitches as FMN-sensing induces RNA structural changes and regulates transcript levels of tRNA in B. subtilis, and protein levels of RPS31 and ATP1 in C. albicans. As PARCEL can be readily applied to any transcriptome and ligand, we believe that further application of PARCEL to diverse organisms will result in the identification of many novel natural RNA aptamers in the near future, providing new building blocks for biological sensing and deepening our understanding of RNA-based gene regulation in vivo.

**Methods**

**Bacterial and yeast cultures.** P. aeruginosa PA01 and B. subtilis 168, were grown in LB or minimal media to log (OD600 = 0.6–0.8) or stationary phases (OD600 > 2). Total RNA from P. aeruginosa was extracted by Trizol reagent (Thermo Fisher Scientific). Total RNA from B. subtilis was extracted by first incubating B. subtilis in 4 mg per mL of lysozyme for 15 min before using Trizol LS reagent. Ribosomal depleted RNA, Ribo(−) RNA, was obtained using Ribo-Zero rRNA Removal Kit (Epitinger) according to manufacturer's instructions. S. cerevisiae S288C was grown in YPD to exponential phase (OD600 = 0.6–0.8), C. albicans strain SC5314 was grown in YPD or GMM (yeast nitrogen base without amino acids and with 2% glucose) to exponential (OD600 = 0.6–0.8) or stationary phases (OD600 > 2). Total RNA from S. cerevisiae or C. albicans was extracted using a slightly modified protocol that uses hot acid phenol. Poly(A) + RNA was obtained by using the Poly(A) Purist MAG kit (Thermo Fisher Scientific) according to manufacturer’s instructions. Poly(A)+ or Riboc(−) RNA were then structure probed in the presence and absence of metabolites.

**fmn1Δ mutant was created by replacement of FMN1 in BY4741 strain with KanMX using homologous recombination35. RPS31 from C. albicans, with a C-terminal FLAG-tag (GTTACAAGGCGAGCATGAAAG), was integrated together with URA3 at the ura3Δ site to generate the RPS31::FLAG knock-in strain.**

**RNA footprinting analysis.** Cleavage and modification sites along structure probed RNA were identified using primer extension. Briefly, a primer located −30−50 bases downstream of the structure probed region was labeled with γ32P ATP using T4 PNK kinase. The labeled primer was then purified using a 15% TBE urea PAGE gel. The gel was visualized using phosphorimager gel documentation. The RNA was eluted in 12 µl of nuclease free water and concentrated to 2 µl using a vacuum centrifuge. The RNA was ligated to 5′ and 3′ adapters from NEBNext Multiplex Small RNA Library Prep Set for Illumina using T4 RNA ligase (T4 RNA ligase buffer, 1 mM ATP, 10% PEG, 10% DMSO) at 16 °C overnight. After the 5′ adapter ligated RNAs were then purified through a 6% TBE Urea gel and size selected for 50−200 bases. The RNA was then ligated to 3′ adapter, reverse transcribed, and PCR amplified using the NEBNext Multiplex Small RNA Library Prep Set (New England Biolabs) for Illumina using manufacturer's instructions. Eighteen cycles of PCR amplification were typically performed for each library.

**RNA structure models.** RNA secondary structure predictions were generated using RNA folding software with tRNA as constraints, using the program RNAfold35 with default parameters.

**qPCR and Western blotting for wildtype and codon-optimized RPS31.** The RPS31::FLAG and ATP1::FLAG strains were inactivated from single colonies into 2 mL SC-ura media and grown overnight at 30 °C, with shaking. Cells were harvested after 4-6 h of growth at 30 °C, with shaking (when OD600 reaches 0.4). The cultures were split for RNA extraction and Western blotting, pelleted, and washed once with PBS. The resultant cell pellets were frozen and stored at −80 °C.

**RNA extraction and qPCR.** RNA was extracted from frozen yeast pellets using the hot acidic phenol method and treated with Turbo™ DNase (ThermoFisher Scientific). We made cDNA using the Transcriptor First Strand cDNA Synthesis Kit (Roche) and qPCR was performed using SYBR Green Master Mix (Roche) on a Light Cycler (Roche). Primers used are listed as below. The RPS31 and ATP1 primers are specific for the knock-in C. albicans RPS31 and ATP1, and do not amplify the endogenous S. cerevisiae RPS31 and ATP1. Normalized fold changes were calculated by normalizing against actin (ACT1) and the respective strain cultured at 10 mM FMN.
Western blotting. The frozen cell pellet was resuspended in 80 μl of lysis buffer (50 mM Tris pH 7.4, 4% SDS) with proteinase inhibitor (one tablet of a complete Mini, EDTA-free [Roche] in 2.5 ml of lysis buffer). An equal volume of glass beads (425–600 μm, Sigma-Aldrich) was added and cells were lysed in a Mini Beadbeater-96 (Biostep Products) for two cycles of 15 s with a two minute interval. Cell lysates were centrifuged and supernatants were run on 4–20% Mini-Protein TGX Stain-Free protein gels (Bio-Rad). To determine the relative levels of total protein (loading), the gels were first imaged on a ChemiDoc MP System (Bio-Rad) and analyzed with ImageJ22. Corresponding FLAG M2 primary antibody (1:4000 in PBS at 4 °C overnight, Sigma-Aldrich, F3165) and GAPDH (1:4000 in PBS at 4 °C overnight, Sigma-Aldrich, G9666) were used for protein (loading), the gels were next imaged on a ChemiDoc MP System (Bio-Rad) and analyzed with ImageJ22. Uncropped images of blots (in main text and supplementary information) were then integrated them into their corresponding transcriptomes before alignment by bowtie223.

Read mapping. Short reads from PARCEL libraries (Illumina HiSeq, 50 bp, single-end) in P. aeruginosa PA01, B. subtilis 168, and E. coli K12 were aligned to their corresponding reference genomes downloaded from NCBI using the short-read aligner bowtie2 (parameters: –k 1–local). In the case of S. cerevisiae S288C and C. albicans SC5314, we extracted the UTR annotation from Bruno et al.24,25 and integrated them into their corresponding transcriptomes before alignment by bowtie2. In both cases, only uniquely mapped reads were used for subsequent analysis.

Identification of RNA aptamers. For each position along the genome or transcriptome, we counted the number of reads whose first mapped base was one base downstream of the inspected position. Higher counts suggest greater accessibility to V1 nuclease, and are more likely to be associated with a double-stranded conformation. In all expressed transcripts, positions with zero count could either be associated with a single-stranded conformation, or come from a heavily folded region that is inaccessible to V1 nuclease.

Since RNA structural changes should typically span across multiple bases, we looked for regions that exhibit differential V1 counts to increase the sensitivity/specificity of detecting RNA aptamers. We first evaluated the significance of differential V1 counts at each nucleotide position using the edgeR package22, where we compared samples treated by one specific metabolite (e.g., TPP) to samples from all other conditions. We focused on positions that were generally accessible to the V1 nuclease by applying a minimum abundance threshold (average counts per sample per position, c > 1) and then computed a score, s_i, for each passed position i based on edgeR-generated p-values (pval_i): s_i = ln(pval_i) – ln(pval)* (in effect, giving negative scores for p-values > 0.1).

The higher the score, s, the more likely that differential V1 cutting was observed at that specific position. Accordingly, positions that failed to pass the abundance threshold were assigned a penalty score of – 10. We then looked for segments of contiguous positions (e.g., a segment from position m to n) with the highest aggregate score S = \sum s_i by applying the Kadane algorithm (maximal subarray problem). We then determined the significance of these high-scoring segments based on Karlin–Altschul statistics, similar to the approach used in BLAST27.

As described by Karlin and Altschul27, the expected value (E-value) of high-scoring segments with an aggregate score of at least S is given by the formula:

\[ E = Ke^{-\lambda S} \]  

where E is the E-value, K is a constant, λ is the expected value of high-scoring segments based on Karlin–Altschul statistics, similar to the approach used in BLAST27.

Therefore, we examined the extreme value distribution of the aggregate score S to estimate the two parameters required, i.e., K and λ. Specifically, λ can be calculated from the formula: \[ \sum p_i e^{\lambda S} = 1 \], where p_i is the corresponding probability of the scores. As s_i = ln(0.1) – ln(pval_i) and pval_i approximately follows a uniform distribution (U(0,1)) due to the assumption that the majority of nucleotide positions do not undergo any structural changes, the equation \[ \sum p_i e^{\lambda S} = \text{I}_c \] can be translated to:

\[ pval_i \approx e^{\lambda S} \]  

This can be solved to \[ \lambda = -\frac{\ln(0.1)}{S} \] and \[ \lambda = 0.862871 \]. The parameter K is bounded between \[ K = C_1^2 + C_2^2 \] and \[ K = C_1^2 + C_2^2 \]. Since δ is the smallest span of s, K is bounded between \[ \min K \] and \[ \max K \] and C is defined by the formula:

\[ C = \frac{1}{\delta^2} \sum p_i e^{\lambda S} \]  

Here, S_i is the random variable representing the sum of k independently chosen s_i, i.e., \[ S_i = \sum s_i = k \ln(p_i) + \sum (\ln(1-p_i)) \]. Since pval_i ~ U(0,1), \[ S_i \approx \ln(p_i) \] approximately follows the gamma distribution i.e., \[ \Gamma(1, k) = \frac{k}{\delta} \]. Let \[ X_k = \sum s_i - \ln(p_i) \], then it can be derived that:

\[ E(S_k) = \frac{1}{\delta} \int_{s=0} \frac{e^{\lambda S} s^k}{(k-1)!} ds \]  

and

\[ \text{Prob}(S_k \geq 0) = 1 - \frac{1}{\delta} \int_{s=0} e^{-\lambda S} s^k ds \]  

Taken together, C can be solved to take the value of 0.0809635, and the upper and lower bounds of K, i.e., K = 0.0809635. We then calculated E for high-scoring segments by applying equation (1). Segments that pass the E threshold of 10 were considered as candidate RNA aptamers that undergo metabolite-responsive conformational changes. Under all conditions, we report candidate regions that have positions with absolute fold-change > 2, relative abundance greater than median + standard deviation for the transcript (abundance-filter; to avoid segments with lower accessibility) and low bonferroni-corrected p-value (<0.10; to avoid segments with no strongly changing position).

Distribution of RNA aptamers across operons and transcripts. We evaluated the distribution of RNA aptamers across operons in bacteria, and along transcripts in fungi (including a 500 bp window on either side when UTRs were not specified). We plotted the histogram of all RNA aptamer positions, with operons in bacteria and coding regions in fungi being scaled to 1 kbp. There are cases where the same position can be considered as belonging to multiple classes and in such cases, we preferentially assigned positions to the 5′ UTR, then to the operon or CDS, and lastly to the 3′ UTR.

Sequence conservation of RNA aptamers. We estimated the sequence conservation of identified RNA aptamers by measuring nucleotide substitution rate of these regions to their blastn-identified orthologous sequences. If the identified aptamer regions were shorter than 200 bases in length, we extended them on both sides to a maximum of 200 bases. As highly divergent and highly similar sequences would result in an unreliable estimate of nucleotide substitution rate28, we chose fairly divergent, and yet not too divergent species (median Ks ranges from 0.1 to 0.4) for this analysis. The orthologous riboswitches, 3′ UTR, 5′ UTR, and protein coding regions were identified using blastn, for non-coding, or genblastG29, for coding sequences, in other species, respectively. To identify orthologous noncoding sequences in other organisms with high sensitivity, we changed the default blastn parameters as follows: “- e 1e-5 -word_size 7 –gapopen 2 –gapextend 1” 30. We aligned the noncoding sequences using MUSCLE2, and the coding sequences using MACASE2, to construct the multiple sequence alignment. The nucleotide substitution rate of riboswitches, 3′ UTR and 5′ UTR were calculated using Kimura’s 2-parameter method31. The synonymous and non-synonymous substitution rates was calculated using Kascalculate32 with the LPB method.

Calculating the degree of pairedness for RNA aptamers and controls. We searched for orthologous sequences of RNA aptamers identified in B. subtilis, P. aeruginosa, and C. albicans across the Bacillus, Pseudomonas, and Candida genus using blastn (with parameters: “- e 1e-5 -word_size 7 –gapopen 2 –gapextend 1”). The species that were used in each genus are: B. subtilis XF-1, B. subtilis BSn5, B. malacitensis CR-95, B. natto BEST195, B. licheniformis DSM13, B. subtilis subsp.
spitzewii W23, and B. cereus ATCC-14579 for Bacillus; P. aeruginosa PAO1, P. aeruginosa PA7, P. mendocina 1287_FMEM, P. knuckensis B13, P. oryzihabitans USDA-ARS-USMARC-56511, P. pseudodiglucogenes KF707, P. stutzeri A1501, P. stutzeri A1501, P. mendocina ymp, P. entomophaga L48, P. putida F1, P. fluorescens SBW25, and P. syringae pv. tomato str. DC3000 for Pseudomonas; C. albicans SC5314, C. albicans WO-1, and C. dubliniensis for Candida. We then built multiple species alignments for each RNA aptamer region using MUSCLE. We used the program Alifold to calculate the energy and structural stability of the consensus structure. For each RNA aptamer alignment, a shuffled alignment was generated as a control using the "shuffle.pl" script from the Alifold package.

Data availability All relevant data are available from the authors upon request. Data has been deposited under GEO accession number GSE106133.

Received: 21 December 2017 Accepted: 5 March 2018
Published online: 29 March 2018

References
1. Gasch, A. P. et al. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11, 4241–4257 (2000).
2. Breaker, R. R. Riboswitches and the RNA world. Cold Spring Harb. Perspect. Biol. 4, a003566 (2012).
3. Conrad, R. C., Baskerville, S. & Ellington, A. D. In vitro selection methodologies for transcriptional regulation in bacteria. Methods Mol. Biol. 419, 53–68 (2008).
4. Winkler, W., Navi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419, 952–956 (2002).
5. Winkler, W. C., Navi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. Nat. Struct. Biol. 10, 701–707 (2003).
6. Novichkov, P. S. et al. RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genomics 14, 745–2164-14-745 (2013).
7. Washietl, S. & Hofacker, I. L. Consensus folding of aligned sequences as a new reference to the detection of functional RNAs by comparative genomics. J. Mol. Biol. 342, 19–30 (2004).
8. Li, S. & Breaker, R. R. Eukaryotic TTP riboswitch regulation of alternative splicing involving long-distance base pairing. Nucleic Acids Res. 41, 3022–3031 (2013).
9. Wachter, A. et al. Riboswitch control of gene expression in plants by splicing and alternative 3’ end processing of mRNAs. Plant Cell 19, 3437–3450 (2007).
10. Shaul, T. S. et al. Potential anti-infective targets in pathogenic yeasts: structure and function of 3,4-dihydroxy-2-butanone 4-phosphate synthase of Candida albicans. J. Mol. Biol. 341, 1085–1096 (2004).
11. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. Proc. Natl Acad. Sci. USA 99, 15908–15913 (2002).
12. Collart, M. A. & Oliviero, S. Preparation of yeast RNA. Curr. Protoc. Mol. Biol. Chapter 13, Unit13.12 (2001).
13. Guldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res. 24, 2519–2524 (1996).
14. Dar, D. et al. Term-seq reveals abundant ribo-regulation of antibiotics in Escherichia coli. Mol. Divers. 17, 359–368 (2013).
15. Echt, S. et al. Potential anti-infective targets in pathogenic yeasts: structure and function of 3,4-dihydroxy-2-butanone 4-phosphate synthase of Candida albicans. J. Mol. Biol. 341, 1085–1096 (2004).
16. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. Proc. Natl Acad. Sci. USA 99, 15908–15913 (2002).
17. Collart, M. A. & Oliviero, S. Preparation of yeast RNA. Curr. Protoc. Mol. Biol. Chapter 13, Unit13.12 (2001).
18. Gasch, A. P. et al. Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11, 4241–4257 (2000).
19. Breaker, R. R. Riboswitches and the RNA world. Cold Spring Harb. Perspect. Biol. 4, a003566 (2012).
20. Conrad, R. C., Baskerville, S. & Ellington, A. D. In vitro selection methodologies for transcriptional regulation in bacteria. Methods Mol. Biol. 419, 53–68 (2008).
21. Winkler, W., Navi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419, 952–956 (2002).
22. Winkler, W. C., Navi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. Nat. Struct. Biol. 10, 701–707 (2003).
23. Novichkov, P. S. et al. RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. BMC Genomics 14, 745–2164-14-745 (2013).
24. Washietl, S. & Hofacker, I. L. Consensus folding of aligned sequences as a new reference to the detection of functional RNAs by comparative genomics. J. Mol. Biol. 342, 19–30 (2004).
25. Li, S. & Breaker, R. R. Eukaryotic TTP riboswitch regulation of alternative splicing involving long-distance base pairing. Nucleic Acids Res. 41, 3022–3031 (2013).
26. Wachter, A. et al. Riboswitch control of gene expression in plants by splicing and alternative 3’ end processing of mRNAs. Plant Cell 19, 3437–3450 (2007).
27. Shaul, T. S. et al. Potential anti-infective targets in pathogenic yeasts: structure and properties of 3,4-dihydroxy-2-butanone 4-phosphate synthase of Candida albicans. J. Mol. Biol. 341, 1085–1096 (2004).
28. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. Proc. Natl Acad. Sci. USA 99, 15908–15913 (2002).
29. Collart, M. A. & Oliviero, S. Preparation of yeast RNA. Curr. Protoc. Mol. Biol. Chapter 13, Unit13.12 (2001).
30. Guldener, U., Heck, S., Fielder, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res. 24, 2519–2524 (1996).
31. Das, R., Laederach, A., Pearlman, S. M., Herschlag, D. & Altman, R. B. SAFA: semi-automated footprinting and sequence analysis software for high-throughput quantification of nucleic acid footprinting experiments. RNA 11, 344–354 (2005).
32. Lorenz, R. et al. ViennaRNA Package 2.0. Algorithms Mol. Biol. 6, 26-7188-6-26 (2011).
33. Posch, A., Kohn, J., Oh, K., Hammond, M. & Liu, N. V3 stain-free workflow for a practical, convenient, and reliable total protein loading control in western blotting. J. Vis. Exp. 82, 50948 (2013).
34. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012).II
35. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

© The Author(s) 2018

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018