NAR Genomics and Bioinformatics, 2024, **6**, lqae035
https://doi.org/10.1093/nargab/lqae035
Advance access publication date: 17 April 2024
Standard Article

OXFORD

ChimericFragments: computation, analysis and visualization of global RNA networks

Malte Siemers [1,2], **Anne Lippegaus** [1] and **Kai Papenfort** [1,2,*]

[1]Friedrich Schiller University, Institute of Microbiology, 07745 Jena, Germany
[2]Microverse Cluster, Friedrich Schiller University Jena, 07743 Jena, Germany

[*]To whom correspondence should be addressed. Tel: +49 3641 949 311; Email: kai.papenfort@uni-jena.de

Abstract

RNA–RNA interactions are a key feature of post-transcriptional gene regulation in all domains of life. While ever more experimental protocols are being developed to study RNA duplex formation on a genome-wide scale, computational methods for the analysis and interpretation of the underlying data are lagging behind. Here, we present ChimericFragments, an analysis framework for RNA-seq experiments that produce chimeric RNA molecules. ChimericFragments implements a novel statistical method based on the complementarity of the base-pairing RNAs around their ligation site and provides an interactive graph-based visualization for data exploration and interpretation. ChimericFragments detects true RNA–RNA interactions with high precision and is compatible with several widely used experimental procedures such as RIL-seq, LIGR-seq or CLASH. We further demonstrate that ChimericFragments enables the systematic detection of novel RNA regulators and RNA–target pairs with crucial roles in microbial physiology and virulence. ChimericFragments is written in *Julia* and available at: https://github.com/maltesie/ChimericFragments.

Introduction

Base-pairing of two complementary RNA sequences is a fundamental principle of gene expression control in many, if not all, organisms. In eukaryotes, numerous different classes of non-coding RNAs have been described (e.g. microRNAs, long non-coding RNAs, and circular RNAs) controlling transcription, translation, or both (1,2). In bacteria, the majority of base-pairing regulators classify as small RNAs (sRNAs), which are ∼50–250 nucleotides in length and typically act together with RNA chaperones to recognize target transcripts through RNA duplex formation (3). The consequences associated with successful base-pairing range from translation inhibition and transcript degradation to translation activation and increased protein synthesis (4,5).

Comparative genomics and global transcriptome analysis have uncovered thousands of non-coding RNAs with usually unknown regulatory functions (6). To close this gap, various experimental protocols have been developed to capture RNA–RNA interactions at a transcriptome-wide scale (7–10). These tools typically rely on proximity-based ligation of two RNAs followed by high-throughput sequencing of the chimeric RNA molecules, allowing to infer regulatory interactions for annotated, as well as newly identified transcripts. For example, LIGR-seq (LIGation of interacting RNA followed by high-throughput sequencing) led to the discovery of small nucleolar (sno)RNAs interacting with mRNAs in human cells (11), whereas RIL-seq (RNA interaction by ligation and sequencing) revealed novel RNA–RNA interactions and sRNA regulators in bacteria (12,13). Other key technologies for global RNA interactome analysis are CLASH (14), SPLASH (15), GRIL-seq (16) and PARIS (17), all of which enable the genome-wide annotation of RNA–RNA pairs.

Although the above mentioned protocols differ in their experimental design, they all produce chimeric sequencing reads that can be analyzed through various bioinformatic tools. Previously, each method came with its own computational pipeline to detect and quantify RNA duplex formation, however, several new tools now provide a generic platform for data analysis. For example, the RNA$_{\mathrm{NUE}}$ (18) platform relies on the segemehl mapping tool (19) for split read alignment, groups similar split alignments into clusters, and annotates these clusters with overlapping genome features. The resulting interactions are statistically evaluated based on their frequency and the number of complementary bases as well as the hybridization energy that is computed for every read. Similarly, ChiRA (20) enables RNA duplex detection with a two-pass mapping strategy using bwa-mem (21), which aligns long segments with high accuracy in a first run and deals with multi-mapped short segments in a second run. ChiRA offers static plots summarizing all results and base-pairing predictions for each read, however, both tools do not allow interactive data visualization.

Here, we introduce ChimericFragments, a computational platform for the analysis and interpretation of RNA–RNA interaction datasets starting from raw sequencing files (Figure 1). Our platform enables rapid computation of RNA–RNA pairs, RNA duplex prediction, and a graph-based, interactive visualization of the results. ChimericFragments employs a new algorithm based on the complementarity of chimeric fragments around the ligation site, which boosts the identification of *bona fide* RNA duplexes. When applied to a published dataset, ChimericFragments allowed the discovery a novel sRNA that controls virulence gene expression in the major human pathogen, *Vibrio cholerae*. ChimericFragments

Received: January 2, 2024. **Revised:** March 8, 2024. **Editorial Decision:** March 27, 2024. **Accepted:** March 28, 2024

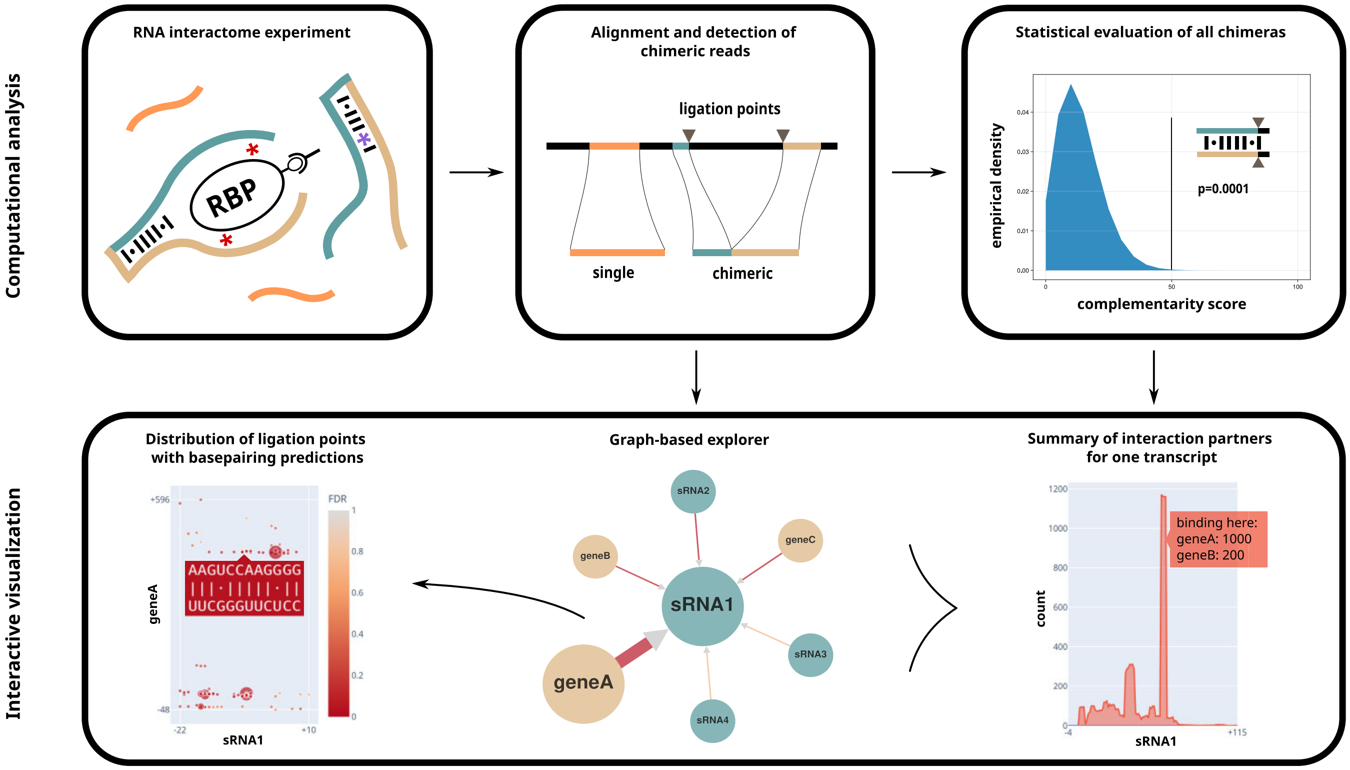

Figure 1. Graphical summary of the computational (top) and the visual component (bottom) of ChimericFragments. Different experimental procedures (e.g. RIL-seq, CLASH, LIGR-seq) produce chimeric RNA molecules and sequence reads. Alignments of those reads are collected and sorted according to the position on the read they belong to (top left). Alignments are classified to be single or chimeric and ligation points are saved (top middle). For each ligation point, the complementarity between the corresponding fragments is computed. A null model is computed for pairs of fragments from random positions on the genome and used to assign a *p*-value to ligated fragments (top right). The global RNA–RNA network can be explored in an interactive graph-based visualization with color-coded annotation types and complementarity strength. Edge thickness and node size relate to the number of represented reads (bottom middle). Each ligation event can be inspected together with a base-pairing prediction (bottom left). All base-pairing predictions for a selected transcript are visualized together and for each position, all partners are highlighted interactively (bottom right).

is implemented in *Julia* and available at: https://github.com/maltesie/ChimericFragments.

Materials and methods

Complete genome annotation

ChimericFragments tags each alignment with an annotation. To capture all chimeras in a dataset, it relies on a fully annotated genome for both forward and reverse strands. The quality of the results depends on the quality of the annotation, so it is recommended to supply ChimericFragments with an accurate annotation of non-coding RNAs and 5′ and 3′ UTRs. The automatic annotation of the genome requires a set of coding sequences (CDS) for which the regions up- and downstream are extended up to a maximum length or until another annotation is reached. The remaining regions of the genome will then be annotated for each strand as IGRs named by their flanking genes.

Reads preprocessing

As a first step in the analysis, ChimericFragments uses fastp to preprocess the raw sequencing reads (22). fastp removes adapters from the reads and read ends with low quality get trimmed and short reads discarded.

Table 1. Synthetic datasets to benchmark bwa-mem2 for chimeric alignments

Fragment length	Sequencing error
15	1
30	1
45	1
Uniform random, 15–45	Uniform random, 0 or 1

Chimeric alignments with bwa-mem2

ChimericFragments uses bwa-mem2 (21) to map reads to the genome. Aligned fragments with alternative alignments (SAM format XA tag not set) are discarded. To benchmark bwa-mem2 in its ability to detect chimeric alignments, several synthetic datasets were generated from and aligned to a random sequence of 5 million nucleotides length with equal probability across the bases. All of them are comprised of chimeric sequences of different length with a mutation at a random position or without. The datasets are summarized in Table 1. To investigate the effect of different TPRs and FPRs on the analysis results of ChimericFragments (Supplementary Figure S5d, e), we created a library of 1 million chimeric reads randomly sampled with lengths from 15 to 40 and a probability of 0.5 of a randomly incorporated nucleotide substitution from 200 pairs of interacting regions in the genome of *V. cholerae* and applied ChimericFragments to it.

Table 2. Classification criteria for reads

Class	Criterion
single	single alignment (same location for PE)
chimeric	split alignment with at least 2 parts
self-chimeric	split alignment with parts mapping to same annotation
multi-chimeric	split alignment with at least 3 parts

Annotation of alignments

Every aligned fragment is annotated uniquely. To efficiently find the annotation with the largest overlap, a binary interval tree is constructed to find m overlaps in a set of n annotations in $O(m + \log n)$. The annotation overlapping the most with the alignment is used to tag it with a name and a type and in cases of tied overlaps, the annotation with the lowest left coordinate on the genome is chosen.

Sorting, merging and classification of alignments

All alignments are sorted first with respect to the read they come from and second to the order of their origin on the read sequence from 5′ to 3′. This procedure results in an ordered set of alignments necessary to detect ligation events between adjacent fragments. If paired end (PE) reads are analyzed, fragments from read1 and read2 are merged if their alignment intervals on the reference are within a specified distance in the correct order and if the same is true for all fragments that follow towards the 3′ end. The resulting set of merged alignments is then classified into being of single or chimeric origin and if classified as a chimera, further checked to be self-chimeric or multi- chimeric according to the criteria listed in Table 2. Reads which do not fall in any of the listed classes, e.g. if only one of two PE reads can be mapped, are discarded.

Ligation points

Each chimeric read is further analyzed and checked for ligation events. A ligation event is defined as a split alignment of two adjacent chimeric fragments on the same read with a user-specified maximal distance of a few nucleotides. The coordinates of the nucleotide closest to the partner fragment are defined as the ligation points of those fragments (Figure 1, top). Ligation points always sit on the edge of an aligned fragment on the side pointing towards the ligated partner. To identify ligation points, first, the read length chosen in the experiment must be sufficiently high to cover at least two fragments and second, the parameters for mapping the reads to the genome will define a lower boundary for the detection of fragments on the reads.

Computation of complementarity

ChimericFragments computes the complementarity of two ligated fragments as a local alignment using the Smith-Waterman (SW) algorithm with a substitution matrix. Instead of representing mutation probabilities with the scores, we designed a matrix that roughly represents binding affinity between two nucleotides (G-C: 5, A-U: 4, G-U: 0, other pairs: −7, open gap: −8, extend gap: −3). The parameterization of the matrix can be chosen by the user. Predictions are made for two intervals of specified length around each ligation point of a ligated pair of fragments. To evaluate each base-pairing prediction, we assign a score s to it. For this, we take the complementarity score s_c from the SW algorithm and subtract it by a weighted shift score s_s:

$$s = s_c - w_s s_s \quad (1)$$

The shift score s_s is defined as the absolute difference of the ends of the complementarity regions found in the two sequences with respect to the observed intervals. This score can be used to select for base-pairing predictions which end at similar positions with respect to ligation points. The weight w_s can be defined by the user. It is set to 1 by default to facilitate a moderate influence of shifted complementarity regions on the total score.

Statistical evaluation

ChimericFragments implements two statistical evaluations of chimeras. First, we construct a null model by sampling complementarity scores s_i between random patches on the genome of specified length. The empirical cumulative density function ECDF of those scores s is computed and a *p*-value can be attributed to predictions:

$$p = 1 - ECDF\,(s) \quad (2)$$

In general, multiple ligation events are found for a given pair of annotations. To summarize the events, we assess the distribution of the *p*-values for each pair and combine them by first computing the FDR with the method of Benjamini and Hochberg (23). ChimericFragments offers multiple methods to combine the *p*-values from ligation points of the same interaction. Either by taking the minimum of the FDR values of all sampled ligation points for one interaction as a combined p-value to represent the probability of error when considering the strongest complementarity found, or by using the methods proposed by Fisher (24) or Stouffer (25), which both test against the null hypothesis of a uniform distribution of *p*-values in the null model they are derived from. The FDR reported in the results tables is based on all combined *p*-values and is used to select interactions shown in the graph. To assess the possibility of detecting RNA–RNA interactions merely because both partners are close to each other or the chaperone protein of interest by chance, each pair gets assigned a statistical significance using Fisher's exact test according to (26). The Fisher exact test queries a hypergeometric distribution to evaluate a contingency table of the two binary variables RNA1 and RNA2, which are true if the respective part of a chimeric pair or a single alignment map to the region of the corresponding annotation, and false otherwise. This behavior is adjustable in the configuration file to exclude single alignments from the background or ignore the order in which the fragments are present on a read.

Comparison of ChimericFragments with RNA_{NUE}

We collected 93 published and experimentally verified RNA–RNA interactions (56 for *E. coli* and 37 for *V. cholerae* (27,28). The complementary regions computed by ChimericFragments in two published RIL-seq experiments (13,27) were compared to this reference set and for each validated interaction, the best overlap is reported. RNA_{NUE} was run with default parameters (params.cfg on https://github.com/ChristopherAdelmann/RNAnue/) in paired-end mode. The resulting set of interacting regions was compared to our reference set and the best overlap was reported.

Visualization of RNA–RNA networks

ChimericFragments visualizes the network of interactions as a graph. The nodes in the graph correspond to annotations on the genome, and an edge represents all chimeras associated with the two annotations connected by the edge (Figure 1, bottom). Edges are directed and the direction of the edge represents the order, in which the fragments are found on their reads. An edge pointing from annotation A to annotation B indicates the fragment mapping to A was found further upstream on the read(s) than the fragment mapping to B. In our graphical representation of the network, the size of the nodes and edges correspond to the share of reads within the selected set of interactions. ChimericFragments is implemented as a web application based on Dash and Dash Bio (29) to support fast drawing and smooth repositioning of hundreds of nodes.

Node positioning in graph drawings

To draw the RNA–RNA interaction network without any prior positional information on the nodes of the graph, we use a technique called stress majorization (30) and choose the parameterization of this algorithm according to (31). Stress majorization is applied to each connected component in the graph and the components are then placed with a simple guillotine bin packing algorithm (32) that fills a rectangle with all connected components while minimizing empty space between them. Together, those two techniques lead to reproducible graph drawings, which are similar for graphs with similar sets of nodes and edges.

Plotting ligation points

Upon clicking on an edge in the graph, an interactive scatter plot of the ligation points in the coordinate frame of the annotations (Figure 1 bottom left, Supplementary Figure S1b) is displayed. In this frame, +1 refers to the first nucleotide in the annotation, counting from −1 down in the upstream direction and from +1 up in the downstream direction. For merged annotations containing a CDS, +1 corresponds to the first nucleotide in the CDS. Each dot in the scatter plot shows the corresponding region of complementarity when hovered upon and size and color of the dot correlate with the number of supporting reads and the FDR associated with the prediction. An FDR-cutoff can be set which applies to FDR values computed from the p-values of the ligation points of currently selected interactions.

Plotting aggregated complementarity

For every node in the graph, all predictions between the node and all of its partners with an FDR below a specified cut-off in the graph are aggregated in a summary plot (Figure 1 bottom right, Supplementary Figure S1c). To compute this summary, for every position in the annotation all ligation points with overlapping regions of complementarity from all interactions in the current selection are summed up. Hovering upon the plotted data shows for every position in the corresponding annotation all partners with a complementary site at this position.

Bacterial strains and growth conditions

All strains used in this study are listed in Supplementary Table S3. *V. cholerae* and *E. coli* strains were grown aerobically in LB or AKI medium at 37°C. Where appropriate, antibiotics were used at following concentrations: 20 μg/ml chloramphenicol and 50 μg/ml kanamycin.

Plasmid construction

All plasmids and DNA oligonucleotide sequences used in this study are listed in Supplementary Tables S2 and S4, respectively. GFP fusions were cloned as described previously (33) using previously determined transcriptional start sites (34,35). Inserts were amplified from *V. cholerae* genomic DNA with the respective oligonucleotide combinations indicated and cloned into linearized pXG10 vectors (KPO-1702/-1703) via Gibson assembly (36); pSM002 (KPO-5251/-5252), pSM003 (KPO-5247/-5248), pJR004 (KPO-2460/-2462), pJG020 (KPO-8756/-8757) and pAL069 (KPO-9358/- 9359). For pNP040 (KPO-1832/-1833) and pNP045 (KPO-1838/-1839), pXG10 and respective inserts were digested with NsiI and NheI and ligated. The constitutive sRNA expression plasmid pAL062 (KPO-9233/-9234) was constructed by PCR amplification of the respective sRNA from *V. cholerae* genomic DNA and cloned into linearized pEVS143 vector (37) (KPO-0092/-1397) via Gibson assembly. Site directed mutagenesis of pJR006 using KPO-9373 and -9374 resulted in pAL077.

Fluorescence measurements

To validate interactions captured by RIL-seq, GFP fluorescence measurements were performed as described previously (38) with *E. coli* Top10 cells cultivated overnight in LB medium. Cells were washed and resuspended in PBS and relative fluorescence was measured with a Spark 10 M plate reader (Tecan). Control strains not expressing fluorescent proteins were used to subtract background fluorescence.

RNA isolation and northern blot analysis

Total RNA sample preparation and Northern blot analyses were performed as previously described (39). Membranes were hybridized in Roti-Quick buffer (Carl Roth) with [32P]-labelled DNA oligonucleotides at 42°C. Signals were visualized using a Typhoon Phosphorimager (GE Healthcare). Oligonucleotides for Northern blot analyses are listed in Supplementary Table S4.

Western blot analysis

Total protein sample preparation and western blot analyses of 3XFLAG-tagged fusions were performed as previously described (40). 3XFLAG-tagged fusions were detected using a mouse anti-FLAG antibody (Sigma; F1804). RNAPα served as loading control and was detected using rabbit anti-RNAPα antibody (BioLegend; WP003). Signals were visualized using a Fusion FX EDGE imager and quantified with BIO-1D software (Vilber Lourmat).

Results

ChimericFragments combines computational analysis and interactive visualization

Experimental workflows such as RIL-seq (13), GRIL-seq (16), CLASH (14), SPLASH (15) and PARIS (17) generate chimeric sequencing reads that are analyzed through tailored bioinformatics pipelines (Table 3). The resulting data are typically presented as large tables providing information on the identity, frequency and statistical significance of the detected RNA–

Table 3. Comparison of bioinformatics analyses of various RNA-interactome protocols

Protocol	Type of interactome	Summary of bioinformatics analysis	Improvements offered by ChimericFragments
RIL-seq (13)	Global RBP-licensed interactome	Analysis with the RILseq package:1. Separate mapping of paired-end reads to detect chimeric reads.2. Detect interacting regions of the genome using Fisher exact test.3. Compare to total RNA and compute hybridization energy between full length transcripts.	simplicity, complementarity-based statistics, interactive browser, accuracy of mapping
GRIL-seq (16)	Enrichment of targets of single sRNA	1. Select reads containing sRNA of interest.2. Generate coverage of the ligated targets.3. Compare with DGEs under overexpression of the sRNA, and compute base-pairing predictions with IntaRNA.	simplicity, interactive browser
SPLASH (15)	Global interactome	1. Map with bwa-mem, remove splice junctions, and detect intra- and inter-molecular interactions.	simplicity, interactive browser, complementarity- and frequency-based statistics
CLASH (14)	Global RBP-licensed interactome	Analysis with the hyb package:1. Mapping with several mappers possible.2. Detection of chimeras based on several filtering criteria.	simplicity, interactive browser, complementarity- and frequency-based statistics
PARIS (17)	Global interactome	Analysis with a very complicated set of scripts and commands:1. Map with STAR and detect intra- and inter-molecular interactions.2. Generate static plots per interaction showing dot-bracket-encoded base-pairing.	simplicity, interactive browser, accessibility of results
LIGR-seq (11)	Global interactome	Analysis with the Aligater package:1. Map with bowtie, remove splice junctions, and detect chimeras.2. Test for significance with a binomial model based on relative abundance of mapping hits per transcript annotation.	interactive browser, complementarity-based statistics

RNA interactions. However, this type of tabular presentation of the results fails to address the complex network structure associated with global RNA interactome studies and does not provide information on the positions of the relevant RNA duplexes (41). ChimericFragments closes both of these gaps employing six main features (summarized in Figure 1) and is compatible with a wide range of experiments producing chimeric RNA sequences.

The analysis of each dataset is split into three main categories: configuration, computation and visualization. The configuration is set by defining the parameters in the template configuration file and all parameters are outlined in the supplied configuration template (default_config.jl). The computational part follows several main steps:(i) preprocessing, (ii) generation of a complete genome annotation, (iii) split read alignment using bwa-mem2 (21,42), (iv) sorting, merging and classification of the produced alignments and (v) statistical evaluation of all interactions (Figure 1, top). For the latter, ChimericFragments captures ligation sites between two fragments and computes the complementarity around them using a parameterized local alignment procedure, which is optimized for complementary pairs and penalizes gaps. The resulting complementarity score is compared to a random model to generate *p*-values (Figure 1, top right). ChimericFragments requires a single configuration file together with two scripts: one for the computational analysis and one to start the web application, which hosts the interactive graphical browser and allows sharing of experimental results online.

The browser uses a graph-based visualization displaying the annotated regions of the genome as nodes and the aggregate of all chimeras mapping to the same two nodes as edges (Figure 1, bottom middle). Additional interactive plots enable the prediction of RNA duplexes formation for every chimeric sequence displaying the frequency of the interaction, as well the position of RNA duplex formation relative to the annotation of the genes involved (Figure 1, bottom left). Finally, the aggregate of all detected ligation sites is shown for each interacting transcript, allowing for the identification of preferred base-pairing sequences in regulatory RNAs and their targets (Figure 1, bottom right). The visualization is implemented as a web application and can be used to share experimental results in the local network or over the internet. A detailed description of the control elements and the multiple data visualization modes in the graphical interface is provided in Supplementary Figures S1--S4.

Optimized mapping parameters increase the number of detected chimeras

A key step in the detection and analysis of base-pairing interactions from global RNA interactome studies is the mapping of chimeric sequencing reads to specific positions in the genome. Bwa-mem2 is an architecture-aware implementation of the bwa-mem algorithm, which can handle split alignments (21,42). We decided to use bwa-mem2 over other comparable algorithms, such as bowtie2 or STAR, based on former studies and due to its computational efficiency and precision (18,43,44).

To determine the suitability of bwa-mem2 for our approach, we set up several benchmarks for the aligner. Specifically, we quantified the sensitivity and precision of bwa-mem2 with respect to the size of the sequencing read and putative sequencing errors. We generated several synthetic sequence libraries of different fragment lengths (15, 25 and 40 nucleotides) with sequencing errors (Table 1). The libraries were each aligned with different values for two limiting parameters, i.e. seed length and minimum alignment score. bwa-mem2 aligns sequences of various lengths without sequencing error to the correct position with a precision of ≥99% (45). A single sequencing error per read resulted in a dependency of the TPR and the FPR on the mapping parameters, as well as the length of the aligned sequence (Supplementary Figure S5a). We applied the same evaluation to another synthetic dataset of mixed length and an error probability of 0.5 per sequence (Supplementary Figure S5b, c). For this dataset, up to 82% of all chimeras were correctly aligned (highlighted in orange) with a FPR of 0.0119, while other combinations of seed length and minimum alignment score recovered 79%, 74% and 67% of the chimeras with FPRs of 0.054, 0.0003 and 0.0002, respectively (shown in red, violet and green). These findings highlight the importance of the alignment parameters for the detection of chimeric fragments, as well as a trade-off towards false discovery, which comes with less stringent parameters.

We next tested the impact of the selected TPRs and FPRs on the analysis output. To this end, we generated a synthetic dataset of 200 pairs of interacting regions and sampled one million chimeras. We then analyzed this dataset using default parameters, only varying the seed length and the minimum alignment score. For the TPRs, the number of recovered chimeras per interacting pair almost perfectly matched the TPR found before (compare Supplementary Figure S5b and d). For the FPRs, we detected a high number of false chimeras, however, the vast majority came with drastically lower counts than the true chimeras (Supplementary Figure S5e). Few systematically false chimeras occurred in duplicated regions of the genome, which hindered bwa-mem2 to unequivocally map these reads. This behavior will lead to falsely detected interactions between repeated regions in the genome, such as multiple copies of ribosomal or tRNA genes. When annotated accordingly, these false hits are easy to remove by the user. Alternatively, alignments matching to repeated regions can be excluded from the analysis in the configuration of ChimericFragments. If repeats are masked properly, bwa-mem2 will not map to the masked regions, also circumventing this problem.

To test our findings from the synthetic datasets with experimental data, we applied ChimericFragments to a previously published RIL-seq experiment containing two independent biological replicates (27). Since no ground truth is available for these experiments, we used the correlation between the number of detected chimeric reads per interaction in the replicates as an indicator for the frequency of random misalignment. The Pearson correlation coefficient of the chimeric in the two replicates was consistently high (≥0.9, Supplementary Figure S5f). To get a more sensitive measure of our results, we also computed the rank correlation and found it to be very strong (≥0.9) for the top 10% fraction of the dataset, and decreasing when more interactions with lower read counts were added (Supplementary Figure S5g). All following analyses of experimental data were performed with a seed length of 12 and a minimum alignment score of 17, as this combination showed the best compromise between the TPR and the FPR in our synthetic benchmarks and was comparable to stricter parameters applied for the experimental data.

ChimericFragments uses bwa-mem2 to compute chimeric alignments which have unique coordinates in the genome. In contrast, ChiRA relies on the output-identical but slower bwa-mem and additionally considers alignments to multiple very similar locations in the genome. This can result in higher mapping rates for fragments shorter than 15 nucleotides (20). $\mathrm{RNA_{NUE}}$ computes alignments with segemehl and the tool's higher precision should result in slightly increased detection rates for true chimeras and lower rates of false chimeras at the expense of the required computational resources.

Ligation points serve as indicators for stable RNA duplex formation

The above mentioned tools for global RNA interactome studies (*e.g.* RIL-seq, LIGR-seq, and CLASH) rely on the ligation of two proximal RNA molecules, resulting in the generation of chimeric sequencing reads (7,46). The general interpretation associated with the detection of a chimeric read is that the detected sequences base-paired, however, it remains unclear if these events describe spurious interactions, or stable RNA duplex formation. Previous work has addressed this problem by calculating the statistical significance of an interaction based on its frequency (18,20,26,47) and initial attempts have been made to also consider additional parameters such as sequence complementarity and hybridization energy (18,20), however, a statistical evaluation of those measures has not been performed.

To close this gap, ChimericFragments collects information on the two positions closest to the ligation site in a chimeric sequence, which we call the *ligation point* (Figure 2A). For each chimeric read with a ligation point, the complementarity of the two fragments around the ligation point is computed. A complementarity score gets assigned to every chimera and its significance is evaluated by comparison to a model computed from the complementarity scores of randomly selected pairs of fixed sequences length from the genome. The distributions from the random model and the selected RIL-seq experiment overlap, however, clearly differ from each other (Figure 2B). We next filtered our dataset based on the FDR assigned to each interaction and computed the correlation between the two biological replicates in the RIL-seq dataset. We observed a strong correlation (≥0.9) among the replicates for interactions with a FDR of 0.05 and decreasing correlation for less stringent FDR-cutoffs (Supplementary Figure S5h), which resembled our previous analysis (Supplementary Figure S5f).

To further understand the effect of filtering interactions based on their complementarity score, we analyzed the distribution of the complementary regions in their respective sequences. In the random model, the ends of the interacting sequences are distributed symmetrically (Figure 2C). Using the significance of the complementarity score as a separation criterion (FDR < 0.25), the total distribution of predicted RNA duplexes around ligation points splits into two populations (Figure 2D), with the non-significant part closely resembling the distribution in the random model (Figure 2E). The dis-

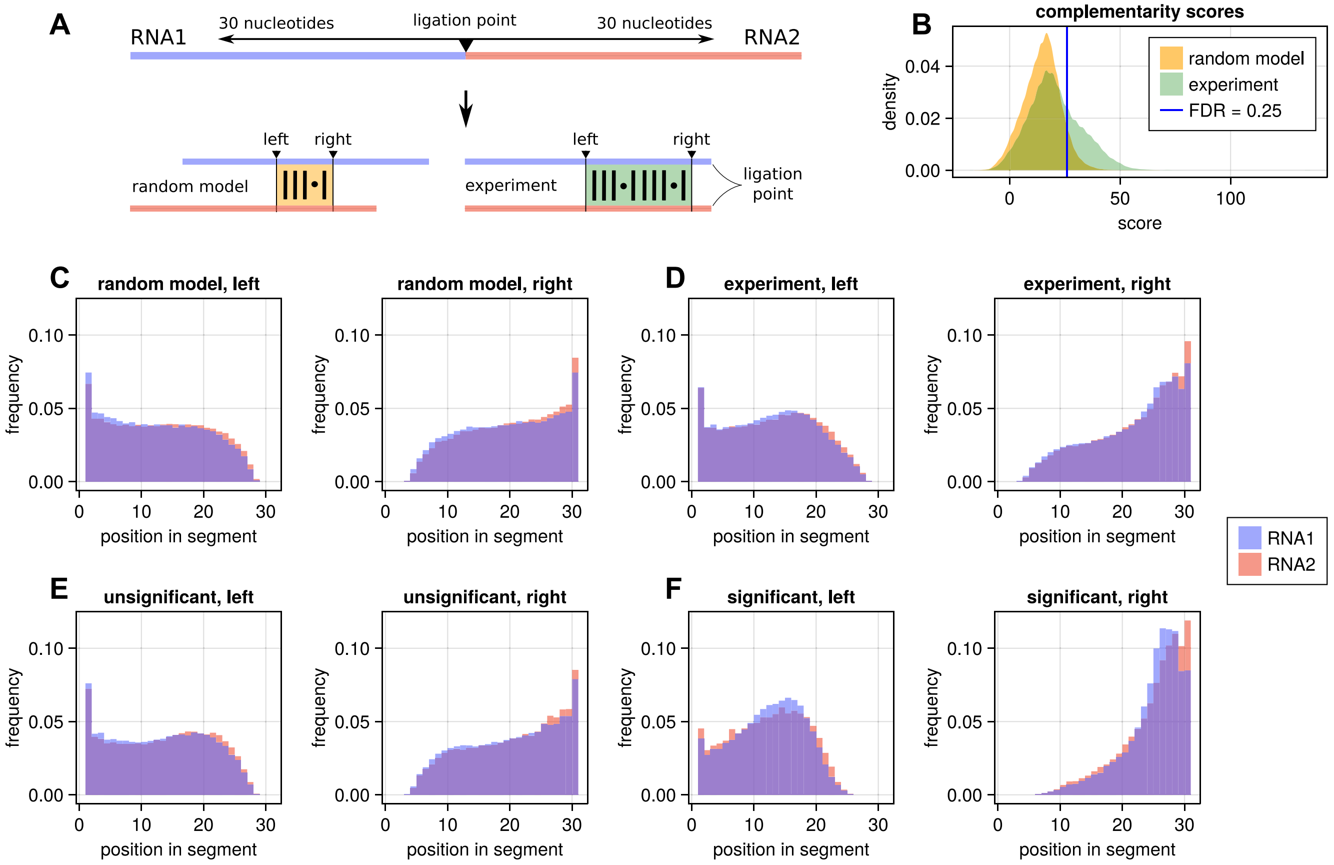

Figure 2 Statistics of base-pairing predictions. (**A**) Schematics of expected base-pairing predictions between random (left, orange) and experimentally derived (right, green) sequence pairs (**B**) Densities of score distributions for complementarity between random sequences (orange) and sequences around ligation points (green) from a RIL-seq experiment, with a fixed length of 30 nucleotides. (**C**) Alignment ends histogram for alignments of random sequences of length 30. (**D**) Alignment ends histogram for all alignments in a RIL-seq experiment. (**E**) Alignment ends histogram for unsignificant (FDR > 0.25) alignments from panel D. (**F**) Alignment ends histogram for significant (FDR ≤ 0.25) alignments from panel D. RNA1 and RNA2 correspond to the first and second partner from 5′ to 3′ in a chimeric RNA.

tribution of the significant interactions shows a strong preference for RNA duplex formation close to the ligation point (position 30, Figure 2F, right) and the left ends of complementarity regions follow a left-tailed distribution with a peak around position 15 (Figure 2F, left). The distribution of the length of the complementarity regions also supports ranking of a dataset according to the significance derived from the random model (Supplementary Figure S6a). The lengths of all complementary regions (Supplementary Figure S6b) split into non-significant (Supplementary Figure S6c) and significant (Supplementary Figure S6d) fractions, with the distribution of non-significant regions closely resembling the random model (compare Supplementary Figure S6a and S6c). In the fraction of significant regions, the average length is 13.6, compared to 8.0 in the random model. On average, experimentally supported interaction sites in our benchmarking dataset are 12.7 base-pairs long, further supporting our proposed model (Supplementary Table S1). Lowering the FDR cut-off to values <0.25 excludes shorter complementary regions and leads to the loss of several experimentally validated interactions such as between VqmR and *aphA* (48). Therefore, we used this level of significance for all following analyses. In contrast, comparison of the significance values from our statistical evaluation and the commonly used Fisher exact test showed no association between the two methods (Supplementary Figure S7a, b). The presented data indicate that the ligation point of a chimeric sequence is a powerful indicator for the identification of stable RNA duplexes in global RNA interactome studies.

To benchmark our approach, we applied ChimericFragments to seven published datasets, involving six different organisms and three different experimental pipelines, i.e. *V. cholerae* (RIL-seq, (27)), *Escherichia coli* (*E. coli*, RIL-seq and CLASH, (13,49)), enteropathogenic *E. coli* (RIL-seq, (50)), *Pseudomonas aeruginosa* (RIL-seq, (51)), *Salmonella enterica* (RIL-seq, (52)) and *Bacillus subtilis* (LIGR-seq, (53)). In all cases, ChimericFragments recovered significantly more interactions than initially reported in the respective studies (Supplementary Figure S7c–i). Of note, although our results are difficult to compare to the previously reported interactions due to differences in the statistical methods to filter the datasets, we discovered ligation points for the majority of interactions in all studies. Except for the LIGR-seq dataset, ChimericFragments revealed more interactions with significant complementarity around the ligation points when compared to the respective initial studies.

We also analyzed two previously published RIL-seq datasets (13,27) using the ChimericFragments pipeline. These datasets were collected in two different model organisms (*E. coli* and *V. cholerae*), allowing us to compare our results with 93 published and experimentally verified RNA–RNA interactions (56 for *E. coli* and 37 for *V. cholerae* (27,54)). We detected relevant chimeras for 55 and 35 of these interactions, respectively (Supplementary Table S1). In *E. coli*, ChimericFragments successfully predicted the reported interaction in 40 cases (∼73%) and similar numbers were obtained in *V. cholerae* (22/35, ∼63%). For comparison, RNAnue detected 19/55 (∼35%) interactions in *E. coli* and 15/35 (∼43%) interactions in *V. cholerae* (Supplementary Table S1). Of note, ChimericFragments also comes with an improved runtime requiring 10 s/Mio. reads, whereas $\mathrm{RNA_{NUE}}$ required 1290 s/Mio. reads (Supplementary Figure S8a).

The computation of complementary regions done by ChimericFragments is parameterized to match with the experimentally confirmed predictions of interaction sites listed in Supplementary Table S1. We expect the regions detected by our approach to be similar to thermodynamically informed predictions. ChimericFragments computes complementary regions between sequences of a fixed length taken from the genome, thus avoiding sequencing errors or short fragments to interfere with the process. This behavior can result in more detected interaction sites compared to $\mathrm{RNA_{NUE}}$'s hybridization of read sequences, as long as the interacting regions are close to the ligation site between the RNA fragments.

ChimericFragments reveals hidden sRNA–target mRNA pairs

To evaluate the ability of ChimericFragments to predict undetected RNA duplexes in global RNA interactome studies, we reanalyzed published RIL-seq datasets derived from *V. cholerae* to search for unknown interaction with high complementarity scores (27). Specifically, we allowed read counts as low as 3 to apply our complementarity-based test of significance and investigated the complementarity in low frequency interactions. When compared to the previous analysis (using a minimal cut-off of 20 reads per interaction and a Fisher exact test FDR < 0.05), our optimized mapping parameters increased the number of RNA pairs by ∼12% (3580). Omitting the Fisher exact test resulted in 8976 interactions (∼2.8-fold increase), whereas of 32890 RNA pairs with read counts of ≥ 3, 12917 (∼39.3%, ∼4.1-fold increase when compared to our previous analysis) came with significant base-pairing predictions with an FDR ≤ 0.25 (Figure 3A).

To investigate the effect of discarding unsignificant interactions according to the Fisher exact test (FDR values >0.05), we selected 10 putative new targets of the well-studied Spot 42 sRNA (55), and tested their regulation using an *in vivo* post-transcriptional reporter assay (33) (Figure 3B). Eight of these targets also showed significant RNA duplex formation and for all these targets we confirmed a regulatory effect >25% (Figure 3C, dotted line). In contrast, the two remaining targets did not display regulation by Spot 42.

ChimericFragments provides insights into the mechanisms of sRNA-mediated gene regulation

Bacterial sRNAs frequently employ multiple base-pairing sequences to interact with target mRNAs, which adds to their function as global regulators of gene expression (4,56). However, the identification of base-pairing sequence elements in a given sRNA is typically not straight-forward based on conservation analysis alone (28,57,58). ChimericFragments addresses this problem as it computes RNA duplexes for every chimeric RNA pair with a ligation point, which are visualized in a summary plot (Figure 1, bottom right). Single peaks indicate one base-pairing sequence in the sRNA (Supplementary Figure S8b), or the target (Supplementary Figure S8c), whereas multiple peaks predict more than one base-pairing sequence (Figure 4A). Using this strategy, we discovered 20 sRNAs with a single base-pairing sites in *V. cholerae* and 35 that contained two or more sites (Supplementary Figure S8d–f).

The summary plots also classify the results depending on the position of a fragment in a sequencing read relative to its partner (RNA1 precedes RNA2; Figure 2A). These data can inform a potential mode of regulation as targets preferentially occupy the first position, whereas regulators (e.g. sRNAs) are frequently found in the second position (13,59). Indeed, when analyzed by our ChimericFragments pipeline, we confirmed that Hfq-binding sRNAs were more frequently recovered as RNA2, when compared to their target mRNAs (Supplementary Figure S8g).

To test if this information would allow us to discover new sRNA targets, we focused on the FarS sRNA, which was previously shown to use a single base-pairing site to control two related fatty acid degradation genes (60). In contrast, our data suggested FarS regulates additional genes using two base-pairing sites (Figure 4A). To validate these predictions, we focused on two target mRNAs with high frequency exclusive to each region: *vc1043* (encoding a fatty acid transporter (61)), which interacts with a novel base-pairing sequence in FarS and *vca0848* (encoding a GGDEF family protein (62)), employing the previously reported base-pairing site (Figure 4B, C). Post-transcriptional reporter assays revealed that *vc1043* and *vca0848* are both repressed by FarS and introduction of single nucleotide mutations (G34C and G54C; Figure 4C) confirmed that base-pairing is specific to the predicted base-pairing site (Figure 4D). Taken together, our analyses show that ChimericFragments generates testable hypothesis that enable a better understanding of the molecular mechanisms underlying post-transcriptional gene regulation.

Identification and characterization of novel regulatory RNAs using ChimericFragments

Regulatory RNAs and target mRNAs have distinct properties in global RNA networks. Whereas regulatory RNAs often interact with hundreds of targets, mRNAs mostly interact with one or few sRNAs, but not with other mRNAs (3,4,12,63). We used this difference in the local network structure to search for undiscovered sRNAs in intergenic regions (IGRs) of the genome. To this end, we computed the network of all interactions between CDSs and IGRs, revealing two IGRs pairing with numerous putative target mRNAs (Supplementary Figure S9). We further analyzed the IGR with the highest number of targets (located between the *vc0715* and *vc0719* genes; Figure 5A) and found that most targets shared a predicted binding site within the *vc0715*::*vc0719* IGR (Figure 5B). To support our hypothesis for a regulatory function of the *vc0715*::*vc0719* IGR, we inspected published transcriptome datasets for a potential sRNA transcript (34,35). Indeed, these analyses revealed a ∼100 nt long transcript, which we named NetX (network derived RNA), and we validated its expression by Northern blot analysis (Figure 5C).

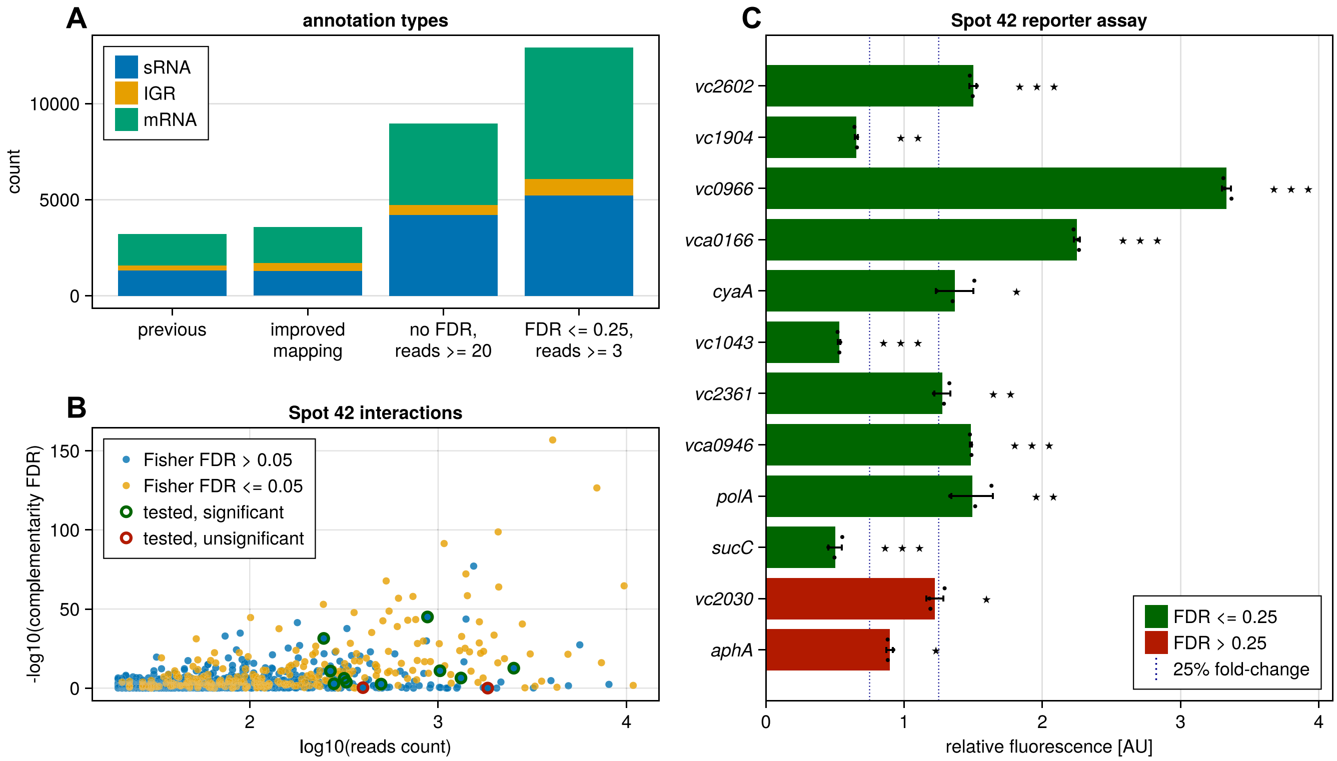

Figure 3 Improved characterization of known regulators. (**A**) Comparison of amounts of interactions recovered in the RIL-seq dataset with our previously published analysis and ChimericFragments. (**B**) Significance of base-pairing predictions for all interactions with more than 20 reads. Colors indicate interactions detected with our previous analysis (yellow) and additional interactions detected only with ChimericFragments (blue). Interactions picked for validation are highlighted with circles. Green circles indicate significant (FDR $\leq$ 0.25) base-pairing predictions and red circles no significant base-pairing predictions (FDR > 0.25). (**C**) Translational GFP reporter fusions were cotransformed with a constitutive Spot 42 expression plasmid or an empty control plasmid in *E. coli* Top10 cells and GFP production was measured. Green bars reflect significant (FDR $\leq$ 0.25) and red bars unsignificant (FDR > 0.25) base-pairing predictions. Bars show the mean of the measurements in three independent biological replicates. For each target, all measurements were divided by the mean of the control measurements and error bars are equal to the respective propagated uncertainty. Significance (unpaired *t*-test) of the difference towards the control samples is indicated by stars: *$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$.

We next focused on the regulatory role of NetX. According to ChimericFragments the most abundant target mRNA of NetX is *aphA*, encoding a key regulator of quorum sensing, biofilm formation, competence, and virulence in *V. cholerae* (48,64–66). Our analysis predicted strong base-pairing between NetX and *aphA* (Figure 5D) and Western blot analysis showed that over-expression of NetX resulted in reduced AphA protein levels in all stages of growth (Figure 5E). Given the documented role of AphA in virulence gene expression and pathogenesis of *V. cholerae* (67), we extended our analysis and monitored the effect of NetX expression on cholera toxin (CtxAB) production. In line with our previous data, NetX strongly reduced CtxAB levels (Figure 5F). Taken together, our results show that ChimericFragments allows the detection of novel RNA regulators and supports the hypothesis-driven research into their regulatory roles in the cell.

Discussion

Base-pairing between two RNA molecules often depends on RNA chaperones such as Hfq and ProQ in bacteria, or SM-like proteins from eukaryotic and archaeal organisms (68–70). Mutation of their respective genes typically impairs RNA duplex formation and in the case of Hfq has been associated with pleiotropic phenotypic alterations, including defects in virulence gene expression in pathogenic bacteria (71). Therefore, studying the molecular processes underlying RNA chaperone activity and global RNA–RNA interactions patterns is not only an important aspect of fundamental research, but also has implications for medicine and public health (72).

The past few years have brought a revolution in our understanding of how RNA–RNA interactions form at a global level due to the development of various new sequencing-based technologies (7,9,10). In contrast to previous approaches, which frequently relied on the identification of individual RNA duplexes and/or the characterization of single RNA regulators, these technologies have paved the way to simultaneously analyze the interactomes of dozens to hundreds of regulatory RNAs and thousands of RNA–RNA pairs (7). However, computational pipelines addressing the complexity of these datasets are scarce and it is often unclear how the detected interactions translate into functionally important discoveries (6).

ChimericFragments offers a computational framework that can help to overcome these limitations (Figure 1). Specifically, the integrated graphical interface allows visualization of global RNA-interactomes, which can provide important information on the relevance of individual regulators or RNA–RNA interactions in the network. Previous work has shown that cellular RNAs constantly compete for interaction with RNA-binding proteins, shaping the biophysical and biochemical parameters driving post-transcriptional gene regulation

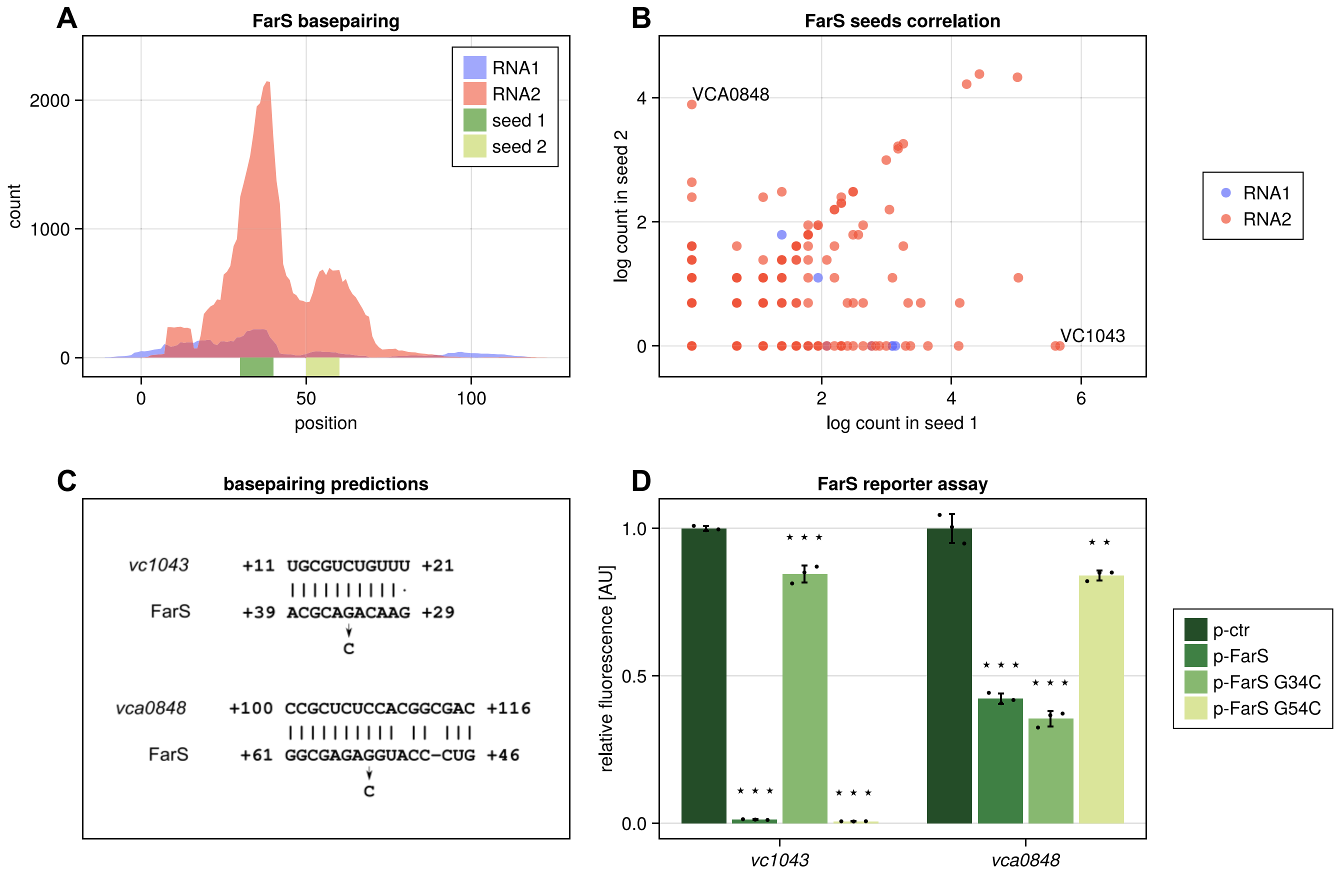

Figure 4 Improved characterization of known regulators. (**A**) Two seed regions in FarS (seed 1, dark green; seed 2, light green) detected by aggregation of all base-pairing predictions of all its targets (FDR ≤ 0.25). (**B**) Counts of reads transformed by natural logarithm supporting complementarity in seed1 or seed2 in FarS for all predicted targets (FDR ≤ 0.25, 1 pseudocount added). (**C**) Base-pairing prediction between FarS seed 1 and *vc1043* (top) and FarS seed 2 and *vca0848* (bottom) with bases for nucleotide exchange marked. (**D**) Translational GFP reporter fusions for *vc1043* and *vca0848* together with an empty control plasmid (p-ctr) or FarS expression plasmids (p-FarS, G34C and G54C) were cotransformed in *E. coli* Top10 and GFP production was measured. Bars show the mean of the measurements in three independent biological replicates and error bars represent the respective standard deviation. For both targets, all measurements were divided by the mean of the control measurements. Mutations G34C and G54C correspond to the seed regions marked with the respective color in panel A. Significance (unpaired *t*- test) of the difference towards the control samples is indicated by stars: $^{**}P \leq 0.01$, $^{***}P \leq 0.001$

(73,74). Thus, to understand the regulatory principles underlying RNA network performance, it is crucial to determine the structure of the network and key regulatory players involved (75). Of note, ChimericFragments also enables the comparison of two or more network states and the study of regulatory features driving RNA network dynamics.

The evolution of regulatory RNAs is highly dynamic and, when compared to their protein counterparts, only poorly understood (56,76–78). Regulatory RNAs can be expressed from IGRs, as well as from the 5′- and 3′UTRs of mRNAs, and the CDS (79). In addition, RNA regulators can also originate from stable transcripts, such as tRNAs (80,81). Therefore, the identification of base-pairing regulators from the pool of all cellular transcripts can be difficult based on standard transcriptome data. ChimericFragments allows the discovery of base-pairing regulators based on the number and quality of the interactions (Figure 5). We showcased this feature of ChimericFragments through the identification and characterization of NetX, which we demonstrate is a previously unknown regulator of virulence gene expression in *V. cholerae*.

Our approach also revealed a second new sRNA regulator, named NetY, which is expressed from the IGR between the *vcr069* and *vc1803* genes (Supplementary Figure S10a). NetY accumulates as a ∼80 nucleotide long sRNA (Supplementary Figure S10b) and base-pairs, like NetX, with various transcripts (Supplementary Figure S10c). However, in contrast to NetX (Figure 5A), the majority of NetY's interaction partners are other non-coding RNAs, suggesting that this sRNA might act as an RNA sponge. RNA sponges base-pair with and inhibit the activity of non-coding regulators and are ubiquitous in prokaryotic and eukaryotic systems (82,83). Further investigations towards the mechanism underlying NetY-mediated base-pairing supported its role as a sponge RNA as the vast majority of chimeras contained NetY at the first position of the sequencing read (indicated in blue; Supplementary Figure S10c), which is a hallmark of target mRNAs and sponge RNAs (13,26,59). In contrast, analogous analyses focusing on FarS and NetX revealed that their corresponding transcripts are typically found in the second position of the sequencing reads (indicated in red; Figures 4A and 5B), suggesting their primary function is to regulate other transcripts. In the case of FarS, we also discovered that the sRNA contains two base-pairing sequences to interact with target transcripts and we also dis-

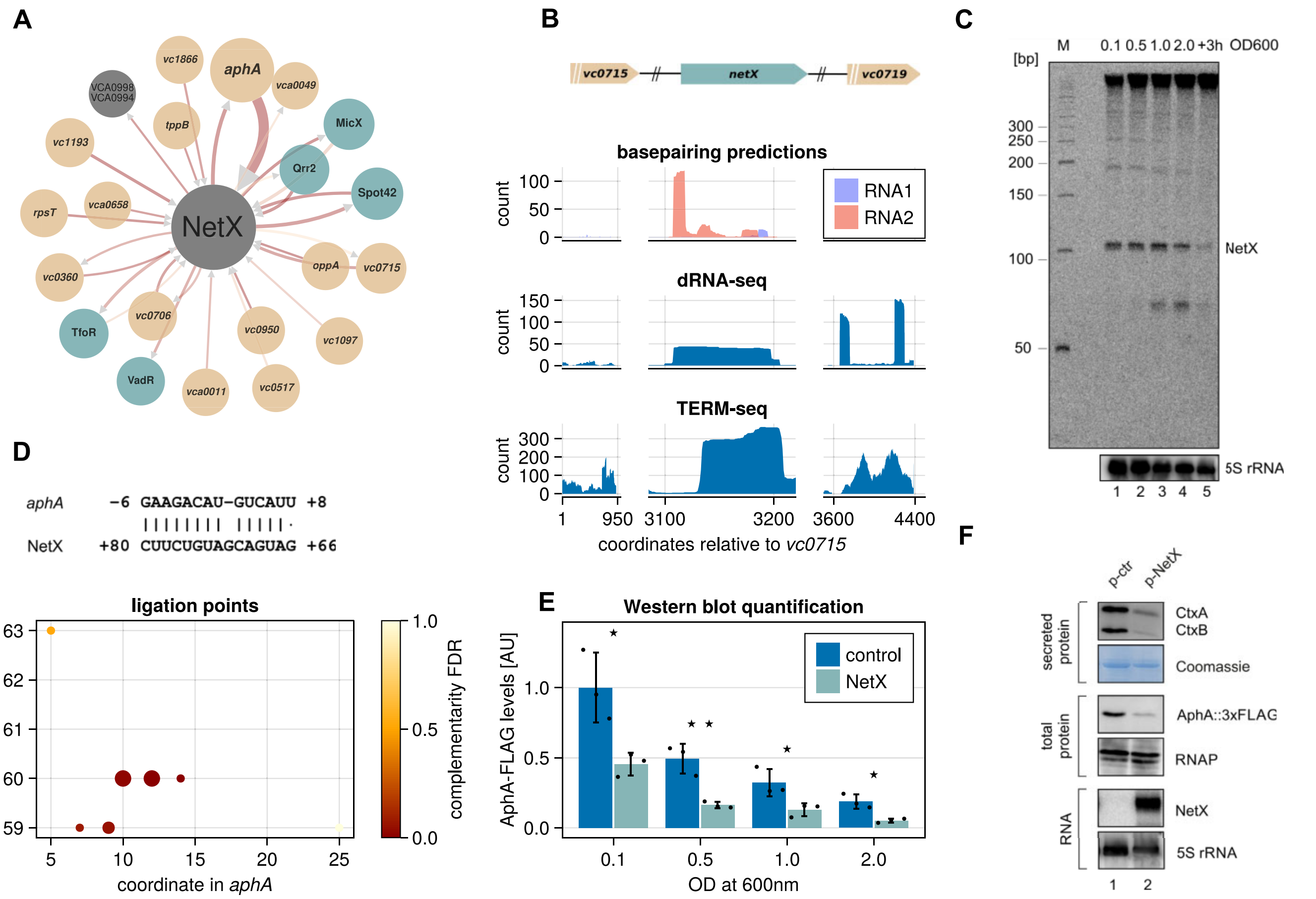

Figure 5 Identification and characterization of the NetX sRNA. (**A**) Graph of all interactions of NetX with more than 3 reads captured by RIL-seq. (**B**) Genomic context of the newly characterized sRNA NetX with aggregation of base-pairing predictions for NetX and flanking genes together with coverage from a dRNA-seq experiment and a TERM-seq experiment at the same location. (**C**) Northern blot detecting a transcript made at the region shown in panel B. RNA samples from *V. cholerae* wild-type cells were collected at various stages of growth. 5S ribosomal RNA served as a loading control. (**D**) ChimericFragments plot of ligation points (bottom) between NetX and the interaction partner *aphA* with a selected base-pairing prediction shared by all ligation points in the lower left corner of the plot (top). (**E**) Quantification of Western blots comparing protein levels of AphA between WT and overexpression of NetX. Protein samples from *V. cholerae* wild-type cells carrying a chromosomal 3XFLAG in the *aphA* gene were collected at various stages of growth. Western Blot analysis was performed to measure AphA levels. Bars show the mean of three independent biological replicates and error bars are equal to the respective standard deviation. Significance (unpaired t-test) of the difference towards the control samples is indicated by stars: $*P \leq 0.05$, $**P \leq 0.01$ (**F**) *V. cholerae* wild-type cells carrying a chromosomal 3XFLAG-tag in the *aphA* gene were cultivated in AKI medium. Secreted protein, total protein and RNA samples were collected, and RNA and protein samples were monitored respectively by Northern and Western blot analysis. Coomassie staining, RNAP and 5S ribosomal RNA served as a loading control for Western and Northern blots, respectively.

covered multiple base-pairing regions in various other sRNAs (Supplementary Figure S8e, f).

These analyses also identified RNA regulators that carry signatures of both categories, i.e. they seem to act as sponges when base-pairing with one set of targets, while in other interactions they likely function as the regulator. Of note, the base-pairing sequence involved in these interactions can be either overlapping (*e.g.* see VSsrna24; Supplementary Figure S10d) or occupy separate segments of the sRNA (e.g. see GcvB; Supplementary Figure S10e). The latter case could indicate a switch in the regulatory function of an sRNA depending on the use of a specific base-pairing sequence, which has not been previously observed. Again, these results highlight the strength of ChimericFragments in generating data-driven hypotheses that can be tested experimentally.

Finally, we designed ChimericFragments to be compatible with various experimental setups that have been used to detect RNA–RNA in bacteria (e.g. RIL-seq, CLASH, and LIGR-seq). ChimericFragments can also analyze data from eukaryotic organisms, as we demonstrate for a CLASH experiment from *Saccharomyces cerevisiae* (Supplementary Figure S10f; (84)). However, we note that larger genome sequences together with the relatively small size of eukaryotic microRNAs, siRNAs, and piRNAs will reduce the number of uniquely mapping sequencing reads in our pipeline, which complicates downstream analysis. Therefore, a refinement of the mapping strategy would be required to adjust ChimericFragments to these alternative datasets.

Data availability

All datasets analyzed in this study are published and available online. Sequencing data of RNA interactome studies are available under the following accession codes: RIL-seq *E. coli* (ArrayExpress, E-MTAB-3910), RIL-seq *V. cholerae* (GEO, GSE198671), RIL-seq EPEC (ArrayExpress, E-MTAB-8806), RIL-seq *S. enterica* (GEO, GSE163336), RIL-seq *P. aeruginosa* (GEO, GSE216135), CLASH *E. coli* (GEO, GSE123050), LIGR-seq *B. subtilis* (ArrayExpress, E-MTAB-8490) and CLASH *S. cerevisiae* (GEO, GSE114680). Term-seq and dRNA-seq sequencing data can be found under the GEO accession codes 'GSE144478' and 'GSE62084', respectively. The code to reproduce all analyses done in this study is available in Github (https://github.com/maltesie/ChimericFragmentsFigures) and Zenodo (ChimericFragments, https://doi.org/10.5281/zenodo.10664038; ChimericFragments Figures, https://doi.org/10.5281/zenodo.10890087). A running instance of a ChimericFragments visualization of the RIL-seq dataset from *V. cholerae* is available at https://vch-interactome.uni-jena.de.

Supplementary data

Supplementary Data are available at NARGAB Online.

Acknowledgements

We thank Andreas Starick and Yvonne Greiser for excellent technical support. We thank Björn Voss (University of Stuttgart) for comments on the manuscript and all members of the Papenfort lab for insightful discussions and suggestions.

Funding

DFG [EXC2051 - Project-ID 390713860]; Vallee Foundation; European Research Council [ArtRNA, CoG-101088027].

Conflict of interest statement

None declared.

References

1. Morris,K.V. and Mattick,J.S. (2014) The rise of regulatory RNA. *Nat. Rev. Genet.*, **15**, 423–437.
2. Marx,V. (2022) How noncoding RNAs began to leave the junkyard. *Nat. Methods*, **19**, 1167–1170.
3. Hor,J., Gorski,S.A. and Vogel,J. (2018) Bacterial RNA biology on a genome scale. *Mol. Cell*, **70**, 785–799.
4. Papenfort,K. and Melamed,S. (2023) Small RNAs, large networks: posttranscriptional regulons in gram-negative bacteria. *Annu. Rev. Microbiol.*, **77**, 23–43.
5. Deogharia,M. and Gurha,P. (2022) The "guiding" principles of noncoding RNA function. *Wiley Interdiscip. Rev. RNA*, **13**, e1704.
6. Guo,J.K. and Guttman,M. (2022) Regulatory non-coding RNAs: everything is possible, but what is important?*Nat. Methods*, **19**, 1156–1159.
7. Singh,S., Shyamal,S. and Panda,A.C. (2022) Detecting RNA–RNA interactome. *Wiley Interdiscip. Rev. RNA*, **13**, e1715.
8. Schonberger,B., Schaal,C., Schafer,R. and Voss,B. (2018) RNA interactomics: recent advances and remaining challenges. *F1000Res*, **7**, 1824.
9. Esteban-Serna,S., McCaughan,H. and Granneman,S. (2023) Advantages and limitations of UV cross-linking analysis of protein-RNA interactomes in microbes. *Mol. Microbiol.*, **120**, 477–489.
10. Melamed,S. (2020) New sequencing methodologies reveal interplay between multiple RNA-binding proteins and their RNAs. *Curr. Genet.*, **66**, 713–717.
11. Sharma,E., Sterne-Weiler,T., O'Hanlon,D. and Blencowe,B.J. (2016) Global mapping of Human RNA–RNA interactions. *Mol. Cell*, **62**, 618–626.
12. Bar,A., Argaman,L., Altuvia,Y. and Margalit,H. (2021) Prediction of novel bacterial small RNAs from RIL-seq RNA–RNA interaction data. *Front. Microbiol.*, **12**, 635070.
13. Melamed,S., Peer,A., Faigenbaum-Romm,R., Gatt,Y.E., Reiss,N., Bar,A., Altuvia,Y., Argaman,L. and Margalit,H. (2016) Global mapping of small RNA–target interactions in bacteria. *Mol. Cell*, **63**, 884–897.
14. Kudla,G., Granneman,S., Hahn,D., Beggs,J.D. and Tollervey,D. (2011) Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10010–10015.
15. Aw,J.G., Shen,Y., Wilm,A., Sun,M., Lim,X.N., Boon,K.L., Tapsin,S., Chan,Y.S., Tan,C.P., Sim,A.Y., *et al.* (2016) In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell*, **62**, 603–617.
16. Han,K., Tjaden,B. and Lory,S. (2016) GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. *Nat. Microbiol.*, **2**, 16239.
17. Lu,Z., Zhang,Q.C., Lee,B., Flynn,R.A., Smith,M.A., Robinson,J.T., Davidovich,C., Gooding,A.R., Goodrich,K.J., Mattick,J.S., *et al.* (2016) RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell*, **165**, 1267–1279.
18. Schafer,R.A. and Voss,B. (2021) RNAnue: efficient data analysis for RNA–RNA interactomics. *Nucleic Acids Res.*, **49**, 5493–5501.
19. Hoffmann,S., Otto,C., Kurtz,S., Sharma,C.M., Khaitovich,P., Vogel,J., Stadler,P.F. and Hackermuller,J. (2009) Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.*, **5**, e1000502.
20. Videm,P., Kumar,A., Zharkov,O., Gruning,B.A. and Backofen,R. (2021) ChiRA: an integrated framework for chimeric read analysis from RNA–RNA interactome and RNA structurome data. *Gigascience*, **10**, giaa158.
21. Vasimuddin,M., Misra,S., Li,H. and Aluru,S. (2019) Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 314–324.
22. Chen,S., Zhou,Y., Chen,Y. and Gu,J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.
23. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B*, **57**, 289–300.
24. Brown,M.B. (1975) 400: a method for combining non-independent, one-sided tests of significance. *Biometrics*, **31**, 987–992.
25. Demerath,N.J. (1949) The American Soldier: volume I, adjustment during Army life. By S. A. Stouffer, E. A. Suchman, L. C. DeVinney, S. A. Star, R. M. Williams, Jr. Volume II, combat and its aftermath. By S. A. Stouffer, A. A. Lumsdaine, M. H. Lumsdaine, R. M. Williams, Jr., M. B. Smith, I. L. Janis, S. A. Star, L. S. Cottrell, Jr. Princeton, New Jersey: princeton University Press, 1949. Vol. I, 599 pp., vol. II, 675 pp.. *Soc. Forces*, **28**, 87–90.
26. Melamed,S., Faigenbaum-Romm,R., Peer,A., Reiss,N., Shechter,O., Bar,A., Altuvia,Y., Argaman,L. and Margalit,H. (2018) Mapping the small RNA interactome in bacteria using RIL-seq. *Nat. Protoc.*, **13**, 1–33.
27. Huber,M., Lippegaus,A., Melamed,S., Siemers,M., Wucher,B.R., Hoyos,M., Nadell,C., Storz,G. and Papenfort,K. (2022) An RNA sponge controls quorum sensing dynamics and biofilm formation in Vibrio cholerae. *Nat. Commun.*, **13**, 7585.
28. Peer,A. and Margalit,H. (2014) Evolutionary patterns of Escherichia coli small RNAs and their regulatory interactions. *RNA*, **20**, 994–1003.
29. Hossain,S., Calloway,C., Lippa,D., Niederhut,D. and Shupe,D. (2019) Visualization of bioinformatics data with Dash Bio. *SciPy*, https://doi.org/10.25080/Majora-7ddc1dd1-012.
30. Gansner,E.R., Koren,Y. and North,S. (2005) Graph drawing by stress majorization. In: *International Symposium on Graph Drawing*. pp. 239–250.

31. Kamada,T. and Kawai,S. (1989) An algorithm for drawing general undirected graphs. *Inform. Process. Lett.*, **31**, 7–15.
32. Amossen,R.R. and Pisinger,D. (2010) Multi-dimensional bin packing problems with guillotine constraints. *Comput. Oper. Res.*, **37**, 1999–2006.
33. Corcoran,C.P., Podkaminski,D., Papenfort,K., Urban,J.H., Hinton,J.C. and Vogel,J. (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol. Microbiol.*, **84**, 428–445.
34. Papenfort,K., Forstner,K.U., Cong,J.P., Sharma,C.M. and Bassler,B.L. (2015) Differential RNA-seq of *vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E766–E775.
35. Hoyos,M., Huber,M., Forstner,K.U. and Papenfort,K. (2020) Gene autoregulation by 3’ UTR-derived bacterial small RNAs. *eLife*, **9**, e58836.
36. Gibson,D.G. (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.*, **37**, 6984–6990.
37. Dunn,A.K., Millikan,D.S., Adin,D.M., Bose,J.L. and Stabb,E.V. (2006) New rfp- and pES213-derived tools for analyzing symbiotic Vibrio fischeri reveal patterns of infection and lux expression in situ. *Appl. Environ. Microb.*, **72**, 802–810.
38. Urban,J.H., Papenfort,K., Thomsen,J., Schmitz,R.A. and Vogel,J. (2007) A conserved small RNA promotes discoordinate expression of the glmUS operon mRNA to activate GlmS synthesis. *J. Mol. Biol.*, **373**, 521–528.
39. Papenfort,K., Pfeiffer,V., Mika,F., Lucchini,S., Hinton,J.C. and Vogel,J. (2006) SigmaE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. *Mol. Microbiol.*, **62**, 1674–1688.
40. Papenfort,K., Silpe,J.E., Schramma,K.R., Cong,J.-P., Seyedsayamdost,M.R. and Bassler,B.L. (2017) A vibrio cholerae autoinducer–receptor pair that controls biofilm formation. *Nat. Chem. Biol.*, **13**, 551.
41. Kang,J., Tang,Q., He,J., Li,L., Yang,N., Yu,S., Wang,M., Zhang,Y., Lin,J., Cui,T., *et al.* (2022) RNAInter v4.0: RNA interactome repository with redefined confidence scoring system and improved accessibility. *Nucleic Acids Res.*, **50**, D326–D332.
42. Li,H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv doi: https://arxiv.org/abs/1303.3997, 26 May 2013, preprint: not peer reviewed.
43. Musich,R., Cadle-Davidson,L. and Osier,M.V. (2021) Comparison of short-read sequence aligners indicates strengths and weaknesses for biologists to consider. *Front. Plant Sci.*, **12**, 657240.
44. Harrath,Y., Mahjoub,A., AbuBakr,F. and Azhar,M. (2019) Comparative evaluation of short read alignment tools for next generation DNA sequencing. In: *2019 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*. pp. 1–6.
45. Thankaswamy-Kosalai,S., Sen,P. and Nookaew,I. (2017) Evaluation and assessment of read-mapping by multiple next-generation sequencing aligners based on genome-wide characteristics. *Genomics*, **109**, 186–191.
46. Silverman,A. and Melamed,S. (2023) Biological insights from RIL-seq in bacteria. arXiv doi: https://arxiv.org/abs/2309.11399, 20 September 2023, preprint: not peer reviewed.
47. Travis,A.J., Moody,J., Helwak,A., Tollervey,D. and Kudla,G. (2014) Hyb: a bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods*, **65**, 263–273.
48. Herzog,R., Peschek,N., Frohlich,K.S., Schumacher,K. and Papenfort,K. (2019) Three autoinducer molecules act in concert to control virulence gene expression in Vibrio cholerae. *Nucleic Acids Res.*, **47**, 3171–3183.
49. Iosub,I.A., van Nues,R.W., McKellar,S.W., Nieken,K.J., Marchioretto,M., Sy,B., Tree,J.J., Viero,G. and Granneman,S. (2020) Hfq CLASH uncovers sRNA–target interaction networks linked to nutrient availability adaptation. *eLife*, **9**, e54655.
50. Pearl Mizrahi,S., Elbaz,N., Argaman,L., Altuvia,Y., Katsowich,N., Socol,Y., Bar,A., Rosenshine,I. and Margalit,H. (2021) The impact of hfq-mediated sRNA-mRNA interactome on the virulence of enteropathogenic Escherichia coli. *Sci. Adv.*, **7**, eabi8228.
51. Gebhardt,M.J., Farland,E.A., Basu,P., Macareno,K., Melamed,S. and Dove,S.L. (2023) Hfq-licensed RNA–RNA interactome in Pseudomonas aeruginosa reveals a keystone sRNA. *Proc. Natl. Acad. Sci. U.S.A.*, **120**, e2218407120.
52. Matera,G., Altuvia,Y., Gerovac,M., El Mouali,Y., Margalit,H. and Vogel,J. (2022) Global RNA interactome of Salmonella discovers a 5’ UTR sponge for the MicF small RNA that connects membrane permeability to transport capacity. *Mol. Cell*, **82**, 629–644.
53. Durand,S., Callan-Sidat,A., McKeown,J., Li,S., Kostova,G., Hernandez-Fernaud,J.R., Alam,M.T., Millard,A., Allouche,D., Constantinidou,C., *et al.* (2021) Identification of an RNA sponge that controls the RoxS riboregulator of central metabolism in Bacillus subtilis. *Nucleic Acids Res.*, **49**, 6399–6419.
54. Peer,A. and Margalit,H. (2011) Accessibility and evolutionary conservation mark bacterial small-rna target-binding regions. *J. Bacteriol.*, **193**, 1690–1701.
55. Beisel,C.L., Updegrove,T.B., Janson,B.J. and Storz,G. (2012) Multiple factors dictate target selection by hfq-binding small RNAs. *EMBO J.*, **31**, 1961–1974.
56. Updegrove,T.B., Shabalina,S.A. and Storz,G. (2015) How do base-pairing small RNAs evolve?*FEMS Microbiol. Rev.*, **39**, 379–391.
57. King,A.M., Vanderpool,C.K. and Degnan,P.H. (2019) sRNA target prediction organizing tool (SPOT) integrates computational and experimental data to facilitate functional characterization of bacterial small RNAs. *mSphere*, **4**, e00561-18.
58. Tello,M., Avalos,F. and Orellana,O. (2018) Codon usage and modular interactions between messenger RNA coding regions and small RNAs in Escherichia coli. *Bmc Genomics (Electronic Resource)*, **19**, 657.
59. Melamed,S., Adams,P.P., Zhang,A., Zhang,H. and Storz,G. (2020) RNA–RNA interactomes of ProQ and hfq reveal overlapping and competing roles. *Mol. Cell*, **77**, 411–425.
60. Huber,M., Frohlich,K.S., Radmer,J. and Papenfort,K. (2020) Switching fatty acid metabolism by an RNA-controlled feed forward loop. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 8044–8054.
61. Turgeson,A., Morley,L., Giles,D. and Harris,B. (2022) Simulated docking predicts putative channels for the transport of long-chain fatty acids in Vibrio cholerae. *Biomolecules*, **12**, 1269.
62. Alyaqoub,F.S., Aldhamen,Y.A., Koestler,B.J., Bruger,E.L., Seregin,S.S., Pereira-Hicks,C., Godbehere,S., Waters,C.M. and Amalfitano,A. (2016) In Vivo synthesis of cyclic-di-GMP using a recombinant Adenovirus preferentially improves adaptive immune responses against extracellular antigens. *J. Immunol.*, **196**, 1741–1752.
63. Nitzan,M., Rehani,R. and Margalit,H. (2017) Integration of bacterial small RNAs in regulatory networks. *Annu. Rev. Biophys.*, **46**, 131–148.
64. Yang,M., Frey,E.M., Liu,Z., Bishar,R. and Zhu,J. (2010) The virulence transcriptional activator AphA enhances biofilm formation by vibrio cholerae by activating expression of the biofilm regulator VpsT. *Infect. Immun.*, **78**, 697–703.
65. Rutherford,S.T., van Kessel,J.C., Shao,Y. and Bassler,B.L. (2011) AphA and LuxR/HapR reciprocally control quorum sensing in vibrios. *Genes Dev.*, **25**, 397–408.
66. Haycocks,J.R.J., Warren,G.Z.L., Walker,L.M., Chlebek,J.L., Dalia,T.N., Dalia,A.B. and Grainger,D.C. (2019) The quorum sensing transcription factor AphA directly regulates natural competence in Vibrio cholerae. *PLoS Genet.*, **15**, e1008362.
67. Lu,R., Osei-Adjei,G., Huang,X. and Zhang,Y. (2018) Role and regulation of the orphan AphA protein of quorum sensing in pathogenic Vibrios. *Future Microbiol*, **13**, 383–391.

68. Mura,C., Randolph,P.S., Patterson,J. and Cozen,A.E. (2013) Archaeal and eukaryotic homologs of Hfq: a structural and evolutionary perspective on Sm function. *RNA Biol*, **10**, 636–651.
69. Holmqvist,E. and Vogel,J. (2018) RNA-binding proteins in bacteria. *Nat. Rev. Micro.*, **16**, 601–615.
70. Santiago-Frangos,A. and Woodson,S.A. (2018) Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdiscip. Rev. RNA*, **9**, e1475.
71. Djapgne,L. and Oglesby,A.G. (2021) Impacts of small RNAs and their chaperones on bacterial pathogenicity. *Front. Cell. Infect. Microbiol.*, **11**, 604511.
72. Sparmann,A. and Vogel,J. (2023) RNA-based medicine: from molecular mechanisms to therapy. *EMBO J.*, **42**, e114760.
73. Woodson,S.A., Panja,S. and Santiago-Frangos,A. (2018) Proteins that chaperone RNA regulation. *Microbiol. Spectr.*, **6**, 385–397.
74. Wagner,E.G. (2013) Cycling of RNAs on Hfq. *RNA Biol*, **10**, 619–626.
75. Strobel,E.J., Watters,K.E., Loughrey,D. and Lucks,J.B. (2016) RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. *Curr. Opin. Biotechnol.*, **39**, 182–191.
76. Mattick,J.S., Amaral,P.P., Carninci,P., Carpenter,S., Chang,H.Y., Chen,L.L., Chen,R., Dean,C., Dinger,M.E., Fitzgerald,K.A., *et al.* (2023) Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat. Rev. Mol. Cell Biol.*, **24**, 430–447.
77. Gray,E.C., Beringer,D.M. and Meyer,M.M. (2020) Siblings or doppelgangers? Deciphering the evolution of structured cis-regulatory RNAs beyond homology. *Biochem. Soc. Trans.*, **48**, 1941–1951.
78. Tanzer,A. and Stadler,P.F. (2006) Evolution of MicroRNAs. *MicroRNA Protoc.*, **342**, 335–350.
79. Adams,P.P. and Storz,G. (2020) Prevalence of small base-pairing RNAs derived from diverse genomic loci. *Biochim. Biophys. Acta Gene Regul. Mech.*, **1863**, 194524.
80. Kuhle,B., Chen,Q. and Schimmel,P. (2023) tRNA renovatio: rebirth through fragmentation. *Mol. Cell*, **83**, 3953–3971.
81. Lalaouna,D., Carrier,M.C., Semsey,S., Brouard,J.S., Wang,J., Wade,J.T. and Masse,E. (2015) A 3’ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. *Mol. Cell*, **58**, 393–405.
82. Thomson,D.W. and Dinger,M.E. (2016) Endogenous microRNA sponges: evidence and controversy. *Nat. Rev. Genet.*, **17**, 272–283.
83. Bossi,L. and Figueroa-Bossi,N. (2016) Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. *Nat. Rev. Micro.*, **14**, 775–784.
84. Dudnakova,T., Dunn-Davies,H., Peters,R. and Tollervey,D. (2018) Mapping targets for small nucleolar RNAs in yeast. *Wellcome Open Res*, **3**, 120.

Received: January 2, 2024. **Revised:** March 8, 2024. **Editorial Decision:** March 27, 2024. **Accepted:** March 28, 2024