Mapping the small RNA interactome in bacteria using RIL-seq

Sahar Melamed[1–3], Raya Faigenbaum-Romm[1,3], Asaf Peer[1–3], Niv Reiss[1], Omer Shechter[1], Amir Bar[1], Yael Altuvia[1], Liron Argaman[1] & Hanah Margalit[1]

[1]Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel. [2]Present addresses: Division of Molecular and Cellular Biology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA (S.M.); Evogene Ltd., Rehovot, Israel (A.P.). [3]These authors contributed equally to this work. Correspondence should be addressed to Y.A. (yaelal@ekmd.huji.ac.il), L.A. (liron.argaman@mail.huji.ac.il) or H.M. (hanahm@ekmd.huji.ac.il).

Published online 7 December 2017; doi:10.1038/nprot.2017.115

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Small RNAs (sRNAs) are major post-transcriptional regulators of gene expression in bacteria. To enable transcriptome-wide mapping of bacterial sRNA–target pairs, we developed RIL-seq (RNA interaction by ligation and sequencing). RIL-seq is an experimental–computational methodology for capturing sRNA–target interactions *in vivo* that takes advantage of the mutual binding of the sRNA and target RNA molecules to the RNA chaperone protein Hfq. The experimental part of the protocol involves co-immunoprecipitation of Hfq and bound RNAs, ligation of RNAs, library preparation and sequencing. The computational pipeline maps the sequenced fragments to the genome, reveals chimeric fragments (fragments comprising two ligated independent fragments) and determines statistically significant overrepresented chimeric fragments as interacting RNAs. The statistical filter is aimed at reducing the number of spurious interactions resulting from ligation of random neighboring RNA fragments, thus increasing the reliability of the determined sRNA–target pairs. A major advantage of RIL-seq is that it does not require overexpression of sRNAs; instead, it simultaneously captures the *in vivo* targets of all sRNAs in the native state of the cell. Application of RIL-seq to bacteria grown under different conditions provides distinctive snapshots of the sRNA interactome and sheds light on the dynamics and rewiring of the post-transcriptional regulatory network. As RIL-seq needs no prior information about the sRNA and target sequences, it can identify novel sRNAs, along with their targets. It can be adapted to detect protein-mediated RNA–RNA interactions in any bacterium with a sequenced genome. The experimental part of the RIL-seq protocol takes 7–9 d and the computational analysis takes ~2 d.

INTRODUCTION

Noncoding RNAs have been acknowledged as major post-transcriptional regulators of gene expression in all kingdoms of life. In many organisms, there are as many regulatory RNAs as transcription factors, suggesting that in both pro- and eukaryotes the layer of post-transcriptional regulation is as important as transcriptional regulation. Despite the growing interest in this regulation layer and the fact that the number of sequenced non-coding RNAs has been constantly increasing, the identification of their targets has lagged behind and is far from complete. The major group of regulatory RNAs in bacteria comprises small RNAs (sRNAs), 50- to 400-nucleotide-long RNA molecules that base-pair with their target RNAs, affecting their translation and/or stability. Most of the currently known sRNA–target interactions involve short and incomplete base-pairing[1]. In Gram-negative bacteria, many of these interactions are mediated by the RNA chaperone Hfq, a ring-shaped hexameric protein that binds both the sRNA and its target, and promotes their pairing[2]. Recently, members of the FinO/ProQ domain protein family were also shown to bind sRNAs and to mediate gene silencing in bacteria[3].

The development of high-throughput techniques, first micro-arrays and then RNA-seq, paved the way for transcriptome-wide approaches for identifying sRNA targets. These techniques enabled global identification of transcripts that change their expression level upon overexpression or deletion of a specific sRNA, but they cannot distinguish direct from indirect targets. Immunoprecipitation (IP) of Hfq, with or without a preceding cross-linking step, followed by microarray[4] or RNA-seq analysis[5–7], led to the identification of a large number of Hfq-bound transcripts, many of which were noncoding RNAs and mRNAs. However, sRNA–target interacting pairs could only be indirectly deduced in these studies by additional sequence-dependent predictive algorithms. Recently developed methodologies for identification of Argonaute-bound miRNA–target pairs in eukaryotes overcame this limitation by adding a ligation step that connects the miRNA and its target RNA by a covalent bond[8–10]. Inspired by such techniques, we developed RIL-seq, an experimental–computational methodology for direct identification of *in vivo* Hfq-mediated RNA–RNA interactions[11]. Applying RIL-seq to *Escherichia coli* at different growth conditions allowed the determination of an extensive and dynamic network of RNA–RNA interactions, which rewires in accord with growth conditions. RIL-seq data expanded the *E. coli* sRNA interactome by ~20 fold to ~2,800 interactions and demonstrated that sRNAs are involved in almost every aspect of bacterial life, further emphasizing the importance of global mapping of bacterial sRNA–target interactions.

Overview of the procedure

RIL-seq consists of an experimental part (Steps 1–125), in which Hfq-bound RNAs are ligated, isolated and sequenced, and a computational part (Steps 126–135), in which the sequenced fragments are mapped to the genome and chimeric fragments that represent putative RNA–RNA interacting pairs are identified (**Fig. 1**). The experimental part of RIL-seq takes advantage of the mutual binding of the sRNA and its target to the Hfq chaperone, where their base-pairing takes place. Bacteria carrying a Flag-tagged Hfq, shown previously to be functional[12], are grown under desired conditions and are UV-irradiated to cross-link the protein and bound RNAs (Steps 1–9). The cells are then lysed and Hfq is

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

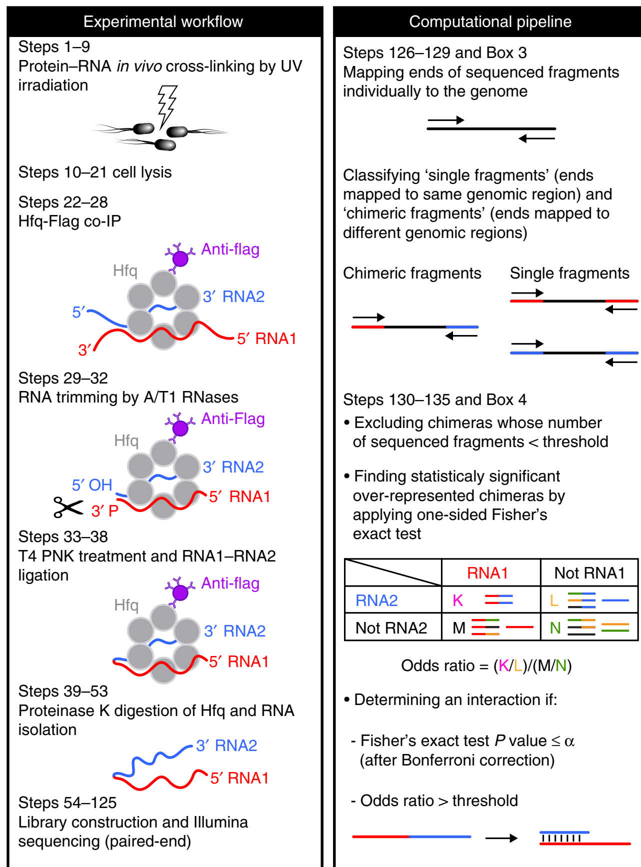

Figure 1 | Overview of RIL-seq experimental and computational procedures. Image adapted with permission from ref. 11, Elsevier.

immunoprecipitated along with its bound RNAs (Steps 10–28). The exposed regions of bound RNAs are trimmed using RNases, neighboring RNAs are ligated and the RNAs are isolated and used for cDNA library construction (Steps 29–124). The library is paired-end sequenced (Step 125), and the resulting sequencing reads are then computationally analyzed. Each sequenced fragment is represented by two sequences at its ends (pair mates), which are mapped individually to the genome. Pair mates that are mapped to the same genomic location most likely originated from a single transcript and are denoted as 'single' fragments. Pair mates that are mapped to two distinct genomic locations represent interacting RNAs that were ligated, and are denoted as 'chimeric' fragments (Steps 126–129). A statistical analysis is then applied to select only those chimeric fragments that are statistically significantly overrepresented in the data. This final analysis is expected to filter out chimeras generated by ligation of random adjacent RNAs and to include in the final subset of reported chimeric fragments those that are highly likely to represent true RNA–RNA interactions (Steps 130–135).

RIL-seq advantages and applications

RIL-seq has several prominent advantages: (i) its experimental part does not require radiolabeling of the RNA, SDS-PAGE and transfer of the radiolabeled RNA onto a membrane, as is the case with other similar methods[8–10]. Instead, the RNA is isolated after the ligation. (ii) As the sRNA is being ligated to its Hfq-bound targets, RIL-seq directly captures the base-paired RNAs, eliminating the need for computational prediction of the interacting pairs. (iii) RIL-seq does not require an artificial overexpression of the sRNA under study and can be applied to the native cells, avoiding biases that can stem from saturation of the Hfq molecules by the overexpressed sRNA. (iv) The interactions are captured simultaneously for all sRNAs expressed under a given condition, providing a representative picture of the sRNA interactome at the tested growth condition while inherently taking into account the mutual effects between co-regulated targets. (v) RIL-seq does not require prior knowledge of either the sRNA or the target sequences, and thus it can discover new sRNAs along with their targets. Taken together, RIL-seq can provide an *in vivo* snapshot of the Hfq-mediated sRNA interactome under any condition of interest. It can be applied to any cultured bacterium with a sequenced genome, without previous knowledge of its sRNA repertoire, given that the interactions are protein-mediated. A natural extension of the original application of RIL-seq, which was conducted in *E. coli* K12 MG1655, would be to apply it to other bacteria in which sRNA–target interactions occur while the interacting RNAs are bound to Hfq, such as in *Yersinia* spp.[13]. Another straightforward application of RIL-seq would be with other RNA-binding proteins involving base-paired RNA regions, such as RNAse III (ref. 14).

In parallel to RIL-seq, a few other RNA-seq-based methods were recently applied to bacteria for the identification of sRNA targets. The CLASH technique was applied to pathogenic *E. coli* to map the sRNA–target pairs that are bound to RNase E (ref. 15), as this endoribonuclease is often recruited by sRNAs to accelerate the target RNA processing. RNase E-CLASH resulted in the identification of a comparable number of sRNA–target pairs, and it can be used as a complementary method to RIL-seq. Notably, unlike CLASH, which includes a stringent purification protocol, RIL-seq is carried out under native conditions and therefore may be advantageous for mapping RNA interactions mediated by proteins that bind RNA only as multimers, such as Hfq that functions in the form of a hexamer[2]. Two other methods that were applied to bacteria are MAPS[16] and GRIL-seq[17]. However, unlike RIL-seq or CLASH, these two methods enable the identification of targets of one sRNA at a time and require overexpression of the studied sRNA.

Limitations of RIL-seq

RIL-seq is designed to capture interactions mediated by a protein, and therefore it cannot capture protein-independent interactions. In addition, a single RIL-seq run will identify only interactions mediated by the chosen tagged protein. This limitation, however, can in principle be addressed by using strains carrying other tagged RNA-binding proteins. RIL-seq requires a reference genome and is thus limited to organisms with sequenced genomes. Finally, as is often the case with high-throughput methodologies, there are some potential technical issues that might bias the results. Such biases may be due to, for example, specific sequences that are preferred for ligation or from ligation of close RNAs that are not regulator–target pairs. However, we demonstrated that for Hfq in *E. coli*, such possible biases are negligible[11].

Experimental design

Initial preparation step: protein tagging. An *hfq-Flag* strain can be constructed by P1 transduction from an *E. coli* K12 *hfq-Flag*

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 1 | Verification of Hfq-Flag IP ● TIMING 3–4 d

Before performing the full experiment, we recommend testing the efficiency of the first and most basic part of the protocol, the IP of the Hfq-Flag. This can be done by carrying out a preparatory experiment as detailed below:

1. Grow the following three cultures as indicated in Step 1 in the PROCEDURE:
   - Culture of an *hfq-Flag* strain, which will be the 'experimental sample'.
   - Culture of an *hfq-Flag* strain, which will be used as the 'No Ab control'.
   - Culture of an *hfq-WT* strain, which will be used as the '*hfq-WT* control'.
2. Proceed to Steps 2–20. At the end of Step 20, transfer 24 µl of each lysate to a new tube, and add 8 µl of 4× Laemmli buffer. Mark the tubes as 'total protein' and store them at −20 °C for a western blot (WB) analysis.
3. Proceed to Step 22.
4. Proceed to Step 23. The tube 'No Ab control' should be left without the anti-Flag M2 antibody.
5. Proceed to Steps 24 and 25.
6. Place the tubes on a magnetic rack. After the solution becomes clear, transfer 24 µl of the lysate to new tubes and add 8 µl of 4× Laemmli buffer to each tube. Mark the tubes as 'unbound protein' and store them at −20 °C for WB analysis.
7. Discard the remaining lysate and remove the tubes from the magnetic rack.
8. Proceed to Step 28.
9. Add 15 µl of 4× Laemmli buffer to the 'No Ab control', 'experimental sample' and '*hfq-WT* control' tubes containing the magnetic beads.
10. Heat the total protein samples taken at step 2 of **Box 1**, the unbound protein samples taken at step 6 of **Box 1** and the 'No Ab control' and 'experimental sample' tubes from step 7 of **Box 1** to 55 °C for 5 min. Run the samples on a 12.5% SDS-PAGE resolving gel, and perform a Western blot analysis using an anti-Flag antibody, following standard procedures.

Anticipated results of the SDS-PAGE analysis: in the 'experimental sample' lane, a main band of ~70 kDa, representing the hexameric Hfq-Flag, is expected. A weaker band of ~14 kDa, representing Hfq-Flag monomer, sometimes appears. These bands should be absent from the 'No Ab control' and '*hfq-WT* control' lanes. The two bands are expected to appear in the total protein samples of both the 'experimental sample' and the 'No Ab control', and to be absent from the total protein of the '*hfq-WT* control'. The lanes of the unbound sample should show weak or no bands, except for the 'No Ab control', in which the bands should look as intense as in the total protein samples. Although an SDS-PAGE analysis is usually expected to show proteins in their denatured form and therefore show only monomers of Hfq, the Flag-tagged Hfq runs in the gel mainly as a hexamer. See **Supplementary Figure 1** (reprinted with permission from ref. 11, Elsevier).

? TROUBLESHOOTING

strain (Strain TM615 (ref. 12)) into a desired *E. coli* strain, or by epitope tagging as described by Uzzau *et al.*[18]. To apply RIL-seq to other RNA-binding proteins, one should verify that the tagged protein is stable and functional (this was done by Morita *et al.*[12] for Hfq-Flag). In addition, the tagged protein should be tested for efficient antibody binding (**Box 1**; **Supplementary Fig. 1**).

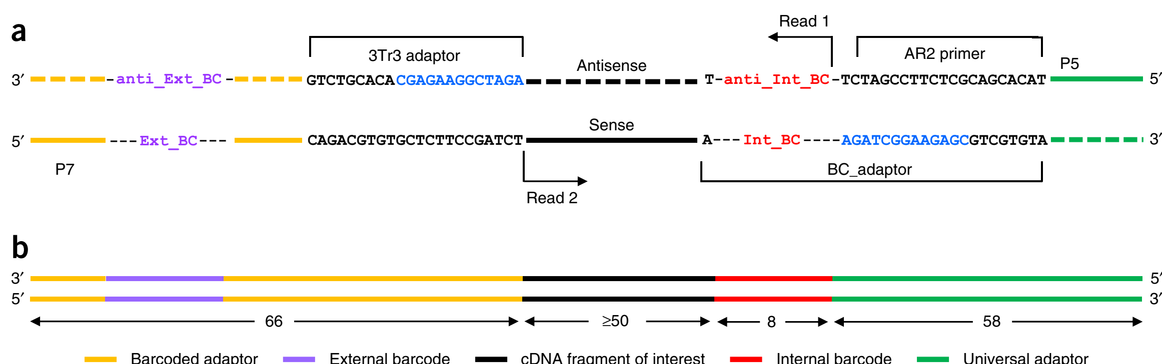

Figure 2 | The configuration of DNA fragments in the sequencing library. (**a**) Schematic representation of the library of DNA fragments after the amplification steps (Steps 109 and 119). Solid and dashed lines of the same color represent corresponding sense and antisense sequences (black: cDNA; green and orange: the P5 and P7 Illumina sequences, respectively). The sequences of the Illumina directly ligated 3′ end adaptors are shown with the common part marked in blue. The internal (Int_BC) and external (Ext_BC) barcodes and their antisense sequences (anti_Int_BC, anti_Ext_BC) are shown in red and purple, respectively. The barcoded (BC) adaptor, AR2 primer and 3Tr3 adaptor sequences are marked by black brackets. The start positions of Read 1 and Read 2 are designated with black arrows. The 5′ end of Read 1 consists of the reverse complement of the internal barcode, followed by T, followed by the reverse complement of the 3′ end of the original transcript. Its 3′ end may contain remnants of the Illumina adaptor. The 5′ end of Read 2 corresponds to the 5′ end of the original transcript. Its 3′ end may contain the internal barcode preceded by A and, rarely, part of the Illumina 3′ adaptor. The Illumina 3′ adaptors are most likely to be found as part of the read when the sequenced fragment is relatively short. Typically, the 5′ Illumina adaptors are not expected to be part of the sequenced reads. The barcodes and the 3′ adaptors are removed at the processing step before mapping (**Box 2**). (**b**) Schematic representation of the length (bp) of different parts within the DNA fragment.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 2 | Generation of the *.bam and *.bam.bai files ● TIMING variable

Here, we describe the pipeline applied for the preprocessing and mapping tailored to the sequence reads generated as described in this article. It will not necessarily apply to sequencing reads generated by alternative library protocols. For more details and additional options, see the relevant software documentation.

1. Generate a local directory for the initial input FASTQ files

```
mkdir <data-dir>
```

Download and store the FASTQ files of your sequenced libraries in `<data-dir>`
It is recommended to keep your files in .gz format to save disk space

2. (Optional, but recommended) Check the sequencing quality:
in `<data-dir>`

```
fastqc *.fastq.gz
```

At the end of this step, you should have in `<data-dir>`, for each of the original FASTQ files, an HTML file with the suffix '`.fastqc.html`' and a compressed directory of FastQC results with the suffix '`.fastqc.zip`'.

3. Open the HTML files in a web browser and check the various sequencing parameters. In particular, check whether the quality of both reads (pair mates) is good. At this processing stage, when the RNAtag-seq protocol[19] is used, Read1 sequences are expected to have the anti-barcode sequence at their 5′ end. Typically, there is a drop in quality at position 9 of Read1. The nucleotide at this position is mostly T (originating from the first adaptor). In addition, substrings of the adaptors (**Table 1** and **Fig. 2**) can be observed for both Read1 and Read2 (see below).

4. Split to individual libraries and trim the anti-barcodes (if needed).
There are many demultiplexing tools; cutadapt is one example. Here, we use an alternative custom script available for download from GitHub: https://github.com/asafpr/RNAseq_scripts/blob/master/index_splitter.py and https://github.com/asafpr/RNAseq_scripts/blob/master/FastSeqIO.py
To run this script, you must generate a barcode table `<barcode-table-name>`
For libraries generated with the RNAtag-seq protocol[19], the barcode table should be in the following format: first field—library name; second field—empty; third field—empty; fourth field—(anti-barcode)T—The anti-barcode is the reverse complement of the internal barcode (shown in bold in **Table 1**) attached to the adaptor and it should be followed by T at the 3′ end.
▲ **CRITICAL STEP** Each line should have exactly four tab-delimited fields; the file should not have empty lines.

5. Run index_splitter.py

```
index_splitter.py --files1 <A comma separated list of files of the first pair mate (the files with the anti-barcode at the 5'end of read)> --files2 <A comma separated list of files of the second pair mate in the same order as files1> -b <barcode-table-name> -m
```

`-m`: one mismatch is allowed (default: no mismatches are allowed).
Note that your input files can be in .gz format.
▲ **CRITICAL STEP** Make sure that index_splitter.py is saved in the `<data-dir>`, or, alternatively, give its full path.
▲ **CRITICAL STEP** When you run with `-m`, all internal barcodes should differ in at least 2 nt.
At the end of this step, you will have in `<data-dir>` two new FASTQ files for each barcoded library, in which the barcodes themselves are trimmed. Note that we require that positions 1–8 of Read1 match the anti-barcode sequences. Consequently, reads that do not match any of the anti-barcodes starting at their first position will be excluded, along with their pair mates.

6. Remove low-quality reads and trim the adaptors:
We use cutadapt to remove Illumina adaptors from the 3′ends of the reads (**Fig. 2**). Cutadapt can be used with multiple parameters. Consult the cutadapt website for the full parameter description. Note: Rarely, sequences of the Illumina adaptors can be found at the 5′ end of the reads; however, as this is associated with imperfect sequencing, we do not trim 5′ adaptors. Consequently, such reads will not be mapped.
Run cutadapt—For 3′ adaptor trimming, apply the following rows (exemplified for a single library) to each of your libraries.
in `<data-dir>`

```
cutadapt -m 25 -q 15 -n 5 -e 0.15 -o <cutadapt_output_of_libraryA_trimmed_1.fastq.gz> -p <cutadapt_output_of_libraryA_trimmed_2.fastq.gz> -a AGATCGGAAGAGC -A AGATCGGAAGAGC -A {A[barcode-sense-sequence]} <libraryA_1.fastq> <libraryA_2.fastq>
```

It is important to note that (i) the string of the barcode sequence you used in Step 59 for indexing the library is preceded by 'A' (**Fig. 2**). The reverse complement of the corresponding sequence in the barcode table you created for the index_splitter.py run will generate the required sequence, i.e., A+barcode-sense. (ii) It is recommended that the minimal size of a read that is defined by the -m parameter be identical to the –l <LENGTH> used in map_chimeric_fragments.py (see below). (iii) It should be verified that the output and input file names are properly listed in the command line. Note that the input files are the last two arguments in the command line. (iv) For the generation of the compressed file, the output file names should end with '.gz'. The below command lines assume that the files are compressed. Two output files (for Read1 and Read2) will be generated for each run:

```
cutadapt_output_of_libraryA_trimmed_1.fastq.gz
cutadapt_output_of_libraryA_trimmed_2.fastq.gz
```

With the above run parameters, the two output files will include only rows for which both reads (pair mates; Read1 and Read2) passed the cutadapt filtering.

(continued)

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 2 | Generation of the *.bam and *.bam.bai files ● TIMING variable (continued)

It is recommended to monitor the number and size of the cutadapt-processed reads. By default, a report will be included in the standard output. You can redirect this output to a file for future reference.

7. Optional (recommended) step: repeat the FastQC quality check for the two final processed files:

```
fastqc *trimmed_*fastq.gz
```

▲ **CRITICAL STEP** If trimmed properly, adaptors and barcodes are not expected to be detected at this stage. If reported in the FastQC files, further processing is needed.

8. Mapping

We use the BWA algorithm[20] for mapping. This mapping is wrapped in a script called map_single_fragments.py, distributed in the RILseq package. For the parameter options type `map_single_fragments.py -h`

Apply the mapping script map_single_fragments.py to each of the trimmed files

```
  foreach file (*_trimmed_1.fastq.gz)
  foreach? set second_file = 'echo $file | sed 's/_1\.fastq/_2\.fastq/''
  foreach? map_single_fragments.py <dir-of-reference-genome/my-organism-genome.fa> -r -1
$file -2 $second_file
  foreach? end
```

▲ **CRITICAL STEP** Make sure that `$file` and `$second_file` are the first and second mate FASTQ files and that their suffixes are `_1.fastq.gz` and `_2.fastq.gz`, respectively. You can use the echo command to see the value of the `$file` and `$second_file` variables

▲ **CRITICAL STEP** For downstream processing, the BAM files should include all reads, even the unmapped ones. Do not use parameters that filter out unmapped reads.

▲ **CRITICAL STEP** If BWA or SAMtools software is not globally installed, you should define the specific path/command of these tools using the map_single_fragments.py parameters `--samtools_cmd` and/or `--bwa_exec`.

▲ **CRITICAL STEP** When the RNAtag-seq protocol[19] is used, the mapped coordinates of the BWA output files correspond to the reverse complement of the original transcript.

At the end of this step, you should have one `.bam` and one `.bam.bai` file for each library.

? TROUBLESHOOTING

Choosing growth conditions. Many sRNAs, as well as many of their targets, are expressed in response to a certain environmental stress or under distinct growth conditions. RIL-seq can be easily applied under various growth conditions. Applying RIL-seq to *E. coli* grown under three different growth conditions (log phase, stationary phase and iron limitation) revealed that different growth conditions result in different sRNA–target networks[11]. Although there are interactions that are revealed in all tested growth conditions, others are condition-dependent. Thus, it is advisable to perform the experiment under more than one growth condition, expanding the repertoire of potential sRNA–target interactions and capturing condition-specific known and novel sRNAs along with their targets.

Replicates and number of samples. We recommend that, for each tested growth condition, RIL-seq be carried out in three biological replicates, each starting from a single colony. This way, the resulting chimeric fragments can be compared and the reproducibility of the results can be tested. The chimeras from the replicates can later be unified to increase the detection power of poorly expressed interacting RNAs. Twelve RIL-seq samples can conveniently be prepared in parallel.

Controls. A RIL-seq experiment includes two types of controls: (i) An obligatory control for each RIL-seq experiment should include a sample of a bacterial strain having a nontagged Hfq (*hfq-WT*), which will be subjected to the full procedure. This control will test the nonspecific binding of RNA to the magnetic beads and to the antibody. (ii) Recommended controls for first-time application of RIL-seq to a different bacterium or a different flagged protein are as follows: (ii.i) A control for the Hfq IP experiment, testing the nonspecific binding of Hfq to the magnetic beads in the absence of an antibody. (ii.ii) A control for nonspecific binding of Hfq to the anti-Flag antibody and to the magnetic beads used for the IP. This control is obtained by applying RIL-seq to a bacterial strain having a nontagged Hfq (*hfq-WT*). Controls (ii.i) and (ii.ii) will be carried out only until the end of the IP (Step 28) and will be tested by a western blot as detailed in **Box 1**. (ii.iii) A control for verifying that the chimeric reads represent interactions that occurred *in vivo* on Hfq. The bacterial lysate is mixed with a yeast lysate before Hfq IP. If interactions occur on Hfq after the cells were lysed, one would expect to find mixed yeast–bacteria chimeras. Performing this control experiment, we found that mixed yeast–bacteria chimeric reads consisted of less than 1% of the total number of statistically significant chimeras[11], confirming that the vast majority of the statistically significant chimeras are formed *in vivo*. Although repeating this verification for Hfq is unnecessary, it is recommended to perform this control experiment when applying RIL-seq to other RNA-binding proteins.

Library construction. RIL-seq libraries can be constructed by various protocols, provided that short RNAs are represented. We recommend using the RNAtag-seq protocol[19] with slight modifications, enabling the capturing of short RNAs as described in Steps 54–125. **Figure 2** illustrates the configuration of the DNA fragments in a library.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

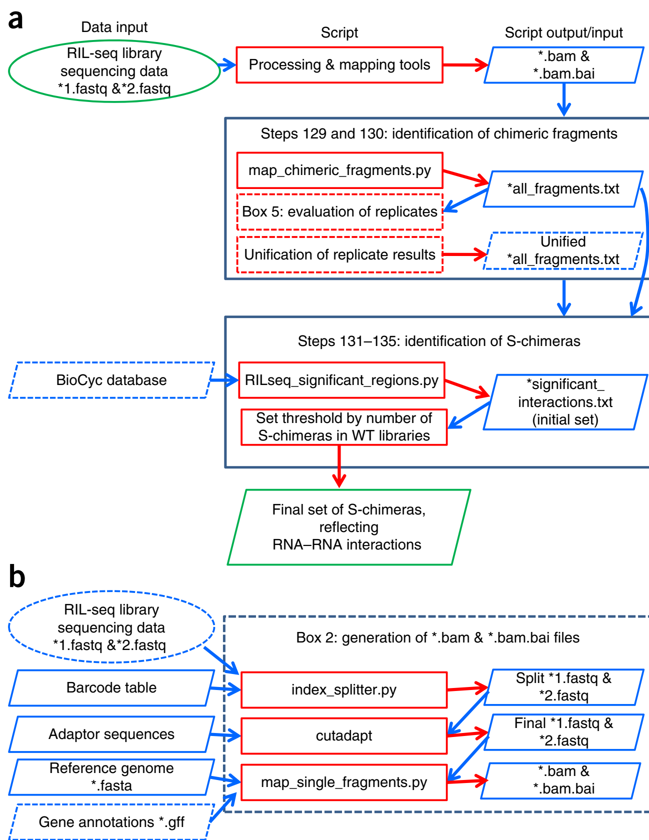

Figure 3 | Flow chart of the computational analysis. (**a**, top) Processing and mapping of the raw reads. (**a**, bottom) The core of the RIL-seq computational analysis that generates the final set of S-chimeras. (**b**) The processing and mapping pipeline tailored to the current experimental protocol. Ovals represent start points. The scripts (rectangles) are in the center of the flow chart, data input files (parallelograms) are shown to the left of the script and script output/input files (parallelograms) are to the right of the script. Red and blue frames designate script and input/output files, respectively. Dashed frames designate optional scripts and optional input/output files. Blue and red arrows represent input and output files, respectively. Green frames designate the primary input and output files of the RIL-seq pipeline.

Sequencing depth. On the basis of our experience with *E. coli* K12 MG1655, ~10 million reads per *hfq-Flag* library are needed to get a sufficient number of mapped chimeric fragments (several hundred thousand), which enables the identification of hundreds to a few thousand statistically significant chimeric fragments.

Computational part (RILseq package): design and considerations. The computational analysis comprises two major parts. The first involves processing of the raw reads (demultiplexing, if needed; adaptor and low-quality sequence removal), followed by initial mapping of the reads. This part can be carried out by various computational tools; the detailed pipeline tailored to the current experimental protocol is described in **Box 2**. The second part, which is the core of the RIL-seq computational pipeline, comprises two major steps: (i) identification of chimeric fragments, and (ii) statistical filtering of the chimeric fragments, which enables the selection of those that are statistically significantly overrepresented in the data and are more likely to represent true RNA–RNA interactions. A flow chart of the computational analysis is shown in **Figure 3**.

Computational part: identification of chimeric fragments. The major challenge in this part is to apply a mapping approach that is flexible enough to enable the detection of sequenced fragments originating from two different genomic locations and is stringent enough to reject false chimeric fragments. This is achieved mainly by adding a second mapping step in which shorter subsequences of the reads (of size 25 nucleotides in the current protocol), corresponding to the ends of the sequenced fragments, are mapped individually to the genome by the BWA algorithm[20]. If these two sequences are mapped to the same transcript, the sequenced fragment is denoted as 'single', whereas mapping to two different genomic locations identifies the fragment as a 'chimeric fragment'. In the case of ambiguities, priority is given to single-transcript mapping. The mapping-related parameters can all be changed and fine-tuned to the specific experimental protocol and organism of interest (**Box 3**). To remove low-complexity reads and poly-G reads, an extra filtering step is applied in the current computational pipeline, before the remapping. The script outputs a list of sequenced fragments, each classified as single or chimeric.

Computational part: finding statistically significant chimeras. The major challenge at this step is to filter out from the initial set of chimeric fragments those that are likely to result from sporadic, nonspecific ligation of RNAs, and retain only chimeric fragments that are more likely to represent true RNA–RNA interactions. To this end, we first define the two components comprising each chimera, RNA1 and RNA2. We then count the number of chimeric fragments involving both RNA1 and RNA2, the number of all other fragments (single and chimeric) involving only RNA1 or only RNA2, and the number of all fragments (single and chimeric) that do not involve either RNA1 or RNA2. We apply a one-sided Fisher's exact test to these numbers, estimating whether the chimeras comprised of RNA1–RNA2 appear statistically significantly more than would be expected if they were to occur randomly. Chimeric fragments that pass the statistical test are reported as statistically significant chimeras (S-chimeras). To keep this process as general as possible, all calculations are executed at the level of genomic coordinates, independent of specific annotation (**Box 4**). Typically, the reported list includes a large number of S-chimeras (ranging from hundreds to a few thousands), representing interacting RNA pairs. Although the RILseq package does not support global ranking of all the potential sRNA–target pairs, it enables heuristic ranking of all the targets of a specific sRNA at the specific tested condition (**Box 4**).

Of note, we do not remove PCR duplicates. In general, RNA duplicates are not easily distinguished from biologically meaningful duplicates that imply higher abundance of a specific RNA sequence. This is especially true for RIL-seq, as there is a cleavage step of the RNA bound to Hfq, which is expected to generate cleaved fragments with similar ends.

MATERIALS

REAGENTS

▲ CRITICAL All the reagents used in Steps 10–124 should be RNase-free.

- *E. coli* strain MG1655 (S. Altuvia, The Hebrew University of Jerusalem)
- *E. coli* strain MG1655 *hfq-Flag*[11]
- Sodium chloride (NaCl; J.T. Baker, cat. no. 3624)
- Yeast extract (BD, cat. no. 212750)

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 3 | Tips on map_chimeric_fragments.py parameters

Here, we describe the main parameters and the rationale of the specific values selected for the analysis of the sequencing data generated as described in this protocol. Note that changing the parameter values might result in slightly different classification of some of the sequenced fragments as single or chimeric. For all parameters, unless specifically specified otherwise, we used the default value. The full list of parameters with their default values can be obtained by

`map_chimeric_fragments.py -h`

A major consideration in the parameter selection is reducing misclassification of single fragments as chimeric. We prefer to misclassify some chimeric fragments as single and not the other way around. To achieve this, we adhere to the following guidelines:

1. Fragments that were classified as chimeras by map_chimeric_fragments.py but were mapped to a single transcript in the initial mapping (input files from Step 127) are considered single.
2. To be reported as a chimeric fragment or single fragment, each mate can have at most one mismatch (`--allowed_mismatches`). However, if two pair mates were mapped to different genomic locations but there is an alternative mapping to a single transcript, even with as many as three mismatches per mate (`--max_mismatches`), it will not be considered a chimera and will be excluded from the analysis.
3. We assume that pair mates that are mapped on the same operon are likely to originate from a single transcript, even if they are mapped to different genes. Thus, we use the genomic transcriptome coordinates to classify the sequenced fragments. Note: In some cases, this could lead to misclassification of RNA interactors derived from the same operon as single fragments.
4. Fragments determined as circular are excluded by default (can be changed with `--keep_circular`).

Parameters related to additional sequence filtering

Low-complexity reads and reads with poly-G can cause mapping errors. Poly-G reads appear in the NextSeq sequencing results (or other two-dye SBS sequencers) because G designates no signal. map_chimeric_fragments.py enables the removal of such reads.

`--dust_thr DUST_THR` **(default:** `10`**)**—Filters out low-complexity reads, excluding fragments that do not pass the threshold of `DUST_THR`. To skip this filtering, set `DUST_THR` to 0.

`--maxG MAXG` **(default:** `0.8`**)**—Reads with a fraction of G nucleotides higher than the `MAXG` value are excluded. When using other sequencing technologies, set to 1.

Mapping parameters

`-l LENGTH` **(default:** `25`**)**—When using paired-end mapping, it is highly likely that one of the mates of a true chimera will be composed of sequences decoded in two different genomic regions, spanning the ligation point, and thus will not be mapped properly. The longer the sequence, the higher the probability of a mapping failure. To overcome this, it is recommended to map shorter reads of size `LENGTH` starting at the read first position. In the current analysis, we mapped the first 25 nt of each read. Note: as we consider shorter reads, it is recommended to perform a very strict mapping, allowing only a small number of mismatches. It is recommended that `LENGTH` be compatible with the minimal read size defined at the preprocessing stage (e.g., in cutadapt see **Box 2**).

▲ **CRITICAL** A `LENGTH` value that is too small can result in a high number of false-positive chimeric fragments.

BWA[20] and SAMtools[21] are used for mapping and indexing, respectively. For a detailed description of the BWA[20] and SAMtools[21] run parameters, see their original documentation. Run parameters can be passed to the map_chimeric_fragments.py script by:

`--params_aln PARAMS_ALN` **(default:** `-N -M 0 -t 8`**)**—Pass additional parameters to `bwa aln`.

`--samse_params SAMSE_PARAMS` **(default:** `-n 1000`**)**—Pass additional parameters to `bwa samse`. Importantly, `bwa samse` and not `sampe` is used, as Read1 and Read2 are mapped independently to the genome and not as paired-end reads.

`-r, --reverse_complement` **(default:** `False`**)**—The value used in the current protocol is `True`. It means that Read1 is considered to be on the reverse complement of the original transcript. This parameter must be applied if the RNAtag-seq protocol[19] or another similar protocol is used.

▲ **CRITICAL** Define the direction of the Read1 and Read2 sequences relative to the original transcript in your protocol and use this parameter accordingly.

`--max_mismatches MAX_MISMATCHES` **(default:** `3`**)**—The initial number of mismatches allowed between the read and the genome when running the BWA mapping (passed to the `bwa aln -n` parameter). This should typically be larger than `--allowed_mismatches` (see below), thus allowing a slightly less stringent preliminary mapping, which is used to eliminate chimeric fragments with alternative less stringent single-transcript mapping.

`--allowed_mismatches ALLOWED_MISMATCHES` **(default:** `1`**)**—The final number of mismatches allowed between the read and the genome. Reads with less than `MAX_MISMATCHES` but more than `ALLOWED_MISMATCHES` will be ignored.

`--keep_circular` **(default:** `False`**)**—Removes fragments that are likely circular RNAs (i.e., are mapped to different strands in close proximity and in opposite order). To keep these fragments, use `--keep_circular`.

`-s DISTANCE` **(default:** `1000`**)**—By default, the read coordinates (not their annotation) are used for the single/chimera classification. If Read1 and Read2 of a sequenced fragment map to the same strand but their distance is larger than `DISTANCE`, they are considered chimeric fragments. It is recommended to use the average or median of the organism's transcript lengths.

`-t TRANSCRIPTS` **(default:** `None`**)**—A GFF file of genes/transcripts (see http://www.ensembl.org/info/website/upload/gff.html for the full GFF format description). If this parameter is used, the script uses both `DISTANCE` and the annotation information for single/chimeric classification. When in conflict `DISTANCE` is overridden, i.e., sequenced fragments will be defined as single if they reside on the same transcript, even if their distance is > `DISTANCE`. It is very useful for screening ribosomal RNAs, which often reside

(continued)

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 3 | Tips on map_chimeric_fragments.py parameters (continued)

in very long transcripts. The feature (column 3) and attributes (column 9) in the input GFF file can be defined by `-f FEATURE` **(default:** `exon`**)** and `-i IDENTIFIER` **(default:** `gene_id`**)**, respectively.

▲ **CRITICAL** Downloadable GFF files often have multiple feature types (e.g., exon, CDS, gene). Furthermore, the same genomic region can be assigned two different features. The script considers only one feature and one identifier. Entries that do not exactly match the defined `FEATURE` and `IDENTIFIER` parameters will be ignored. Thus, it is critical to verify (and change if needed) that all relevant entries have the proper `FEATURE` and `IDENTIFIER`. Note: The default 'exon' value is often assigned in bacteria to tRNAs and rRNAs and not to all genes.

- Tryptone (BD, cat. no. 211705)
- Agar (BD, cat. no. 214010)
- EDTA (J.T. Baker, cat. no. 8993-01)
- Na_2HPO_4 (J.T. Baker, cat. no. 3827-01)
- NaH_2PO_4 (Mallinckrodt Chemicals, cat. no. 7892-04)
- Ultra-pure water, RNase- and DNase-free (Biological Industries, cat. no. 01-866-1A)
- PBS 10×, pH 7.4 (Biological Industries, cat. no. 02-023-5A)
- Glass beads, 0.1-mm diameter (Biospec, cat. no. 11079101)
- IGEPAL (Sigma-Aldrich, cat. no. I8896)
- Imidazole (Sigma-Aldrich, cat. no. I5513)
- Trizma base (Sigma-Aldrich, cat. no. T1503)
- SDS, 20% (wt/vol) solution (Sigma-Aldrich, cat. no. 05030)
- NaCl, 5 M solution (Sigma-Aldrich, cat. no. 71386)
- β-Mercaptoethanol (Sigma-Aldrich, cat. no. M6250) ! **CAUTION** This reagent is highly toxic upon inhalation, skin contact and eye contact. Use a fume hood, a lab coat, eye protection and gloves. Dispose of it as toxic waste.
- Recombinant RNase inhibitor (Takara, cat. no. 2313A)
- SUPERase IN RNase inhibitor (Ambion, cat. no. AM2696)
- Protease Inhibitor Cocktail Set III, EDTA-free (Calbiochem, cat. no. 539134)
- Protein A/G magnetic beads (Pierce, cat. no. 88803)
- M2 anti-Flag mouse monoclonal antibodies (Sigma-Aldrich, cat. no. F1804)
- RNase A/T1 mix (Thermo Fisher Scientific, cat. no. EN0551)
- T4 polynucleotide kinase, including PNK buffer 10× (New England BioLabs, cat. no. M0201)
- DMSO (Sigma-Aldrich, cat. no. D8418)
- T4 RNA ligase, high concentration, including 10× T4 RNA Ligase Buffer, ATP (100 mM) and PEG 8000 (50% (wt/vol)) (New England BioLabs, cat. no. M0437M)
- Proteinase K (Thermo Fisher Scientific, cat. no. EO0491)
- TriReagent (Sigma-Aldrich, cat. no. T9424) ! **CAUTION** TriReagent is highly toxic upon inhalation, skin contact and eye contact. Use a fume hood, a lab coat, eye protection and gloves. Dispose of it as toxic waste.
- TriReagent LS (Sigma-Aldrich, cat. no. T3934) ! **CAUTION** TriReagent LS is highly toxic upon inhalation, skin contact and eye contact. Use a fume hood, a lab coat, eye protection and gloves. Dispose of it as toxic waste.
- Chloroform (MP Biomedicals, cat. no. 02194002) ! **CAUTION** Chloroform is highly toxic upon inhalation, skin contact and eye contact. Use a fume hood, a lab coat, eye protection and gloves. Dispose of it as toxic waste.
- Isopropanol (J.T. Baker, cat. no. 9080-01)
- Ethanol (J.T. Baker, cat. no. 8025)
- GlycoBlue (Thermo Fisher Scientific, cat. no. AM9516)
- Bioanalyzer RNA 6000 Pico kit (Agilent Technologies, cat. no. 5067-1513)
- Bioanalyzer RNA 6000 Nano kit (Agilent Technologies, cat. no. 5067-1511)
- FastAP, including FastAP buffer X10 (Thermo Fisher Scientific, cat. no. EF0654)
- TURBO DNase (Thermo Fisher Scientific, cat. no. AM2238)
- RNA Clean & Concentrator-5 kit (Zymo Research, cat. no. R1016)
- RLT buffer (Qiagen, cat. no. 79216)
- Ribo-Zero kit for bacteria (Illumina, cat no. MRZGN126)
- SuperScript III first strand kit (Invitrogen, cat. no. 18080-051)
- RNAClean XP (Beckman Coulter, cat. no. A63987)
- AMPure XP (Beckman Coulter, cat. no. A63881)
- Glacial acetic acid (MP Biomedicals, cat. no. 2300000) ! **CAUTION** It is very hazardous in the case of skin contact, eye contact and inhalation. Use a fume hood, a lab coat, eye protection and gloves.
- Sodium hydroxide (J.T. Baker, cat. no. 0348) ! **CAUTION** It is very hazardous in the case of skin contact, eye contact, ingestion and inhalation. Use a fume hood, a lab coat, eye protection and gloves.
- HIFI HotStart RM (Kapa Biosystems, cat. no. KK5101)
- Qubit dsDNA HS Assay Kit (Invitrogen, cat. no. Q32854)
- High-sensitivity D1000 ScreenTape (Agilent Technologies, cat. no. 5067-5584)
- High-sensitivity D1000 reagents (Agilent Technologies, cat. no. 5067-5585)
- Barcoded adaptors and additional Illumina adaptors (**Table 1**)
- 4× Laemmli sample buffer (Bio-Rad, cat. no. 1610747)

EQUIPMENT

- Water bath shaker (New Brunswick Scientific, model no. C76)
- Spectrophotometer (Amersham, Biosciences, model no. Ultraspec 10)
- Refrigerated centrifuge for conical tubes (Thermo Fischer Scientific, model no. Megafuge 16R)
- Spin-down centrifuge (GeneReach Biotechnology, model cubee)
- UV Cross-linker with UV bulbs, $\lambda = 254$ nm (Stratagene, model no. Stratalinker 1800)
- Metal block (a part of a heating block; e.g., Thermo Scientific, cat. no. 2069Q)
- Dry block (Thermo Fischer Scientific, cat. no. 88870005)
- Mixer mill (Retch, model no. MM400)
- Mixer mill tube adaptors (Retch, cat. no. 22.008.0008)
- Magnetic rack (MagneSphere; Promega, cat. no. Z5342)
- Rotator for tubes (ELMI, model no. RM-2L)
- Fume hood
- Thermocycler (Sensoquest Lab-Cycler)
- Agilent Bioanalyzer (Agilent Technologies)
- TapeStation Instrument (Agilent Technologies)
- Qubit fluorometer (Thermo Fischer Scientific, cat. no. Q33216)
- TapeStation loading tips (Agilent Technologies, cat. no. 5067-5153)
- TapeStation tubes and caps (Agilent Technologies, cat. nos. 401428, 401425)
- Sequencing system (Illumina, model no. NextSeq 500). Other Illumina sequencing platforms can be used, provided that the yield will be at least 10 million paired-end reads, 40 nt each, per library.
- Cold (4 °C) room
- Filter tips (Sorenson, cat. nos. 30340, 34000, 14220)
- LoBind 1.5-ml tubes (Eppendorf, cat. no. 022431021) or Maxymum Recovery 1.7-ml tubes (Axygen, cat. no. MCT-175-L-C)
- Maxymum Recovery 0.2-ml tubes (Axygen, cat. no. PCR-02-L-C)
- Conical 50-ml tubes (Miniplast, cat. no. 835-050-21-111)
- Disposable pipettes
- Pipette aid (Drumond, model Pipet-Aid XP)
- Pipette set (Finnpipette, F2 series)

Software

▲ **CRITICAL** Download and install all the software listed below (refer to the software websites for instructions). The current computational protocol used the software versions described below. Other versions might not work or might generate unpredictable results.

- FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; tested on v0.11.2)
- cutadapt (https://cutadapt.readthedocs.org/; tested on v1.7.1)
- BWA (http://bio-bwa.sourceforge.net/; tested on v0.7.15-r1140)
- SAMtools (http://www.htslib.org/; tested on v1.3.1)
- Python (https://www.python.org/downloads; v2.7)
- pysam (https://github.com/pysam-developers/pysam; v0.8.3)
- numpy and scipy (https://www.scipy.org/; v1.11.1rc1 and v0.17.1, respectively)

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 4 | Tips on RILseq_significant_regions.py parameters

Here, we describe the main parameters used in the run of RILseq_significant_regions.py. In the current protocol, we used the default values for all parameters, unless specifically specified otherwise. The full list of parameters with their default values can be obtained by `RILseq_significant_regions.py -h`.

`-g GENOME` **(default:** `None`**)**—Insert the genome reference file. This parameter is obligatory when you do not have a BioCyc data directory. In the current analysis, we did not use this parameter, as the genome was retrieved from the BioCyc data stored in the local directory (`<ecocyc-data>`).

`--ribozero` **(default:** `False`**)**—Although the experimental procedures include rRNA removal, you will usually still have sequenced fragments mapped to rRNA. When this parameter has the value `True`, rRNAs are excluded from the analysis. In the current analysis, we set it to `True` by running with `--ribozero`.

▲ **CRITICAL** This parameter can be used only when the BioCyc data directory is given. If such data are not available, it is recommended to define the rRNA coordinates in the tested organism and extract the fragments that correspond to these regions from the map_chimeric_fragments.py output.

The core parameters for the detection of S-chimeras

The genome is divided into windows of size `--seglen SEGLEN` **(default:** `100`**)**. The number of sequenced fragments that are composed of the pair of windows *x* and *y* is counted. As the order of the fused fragments might be important, each pair of windows is tested twice (once for *x* as the 5′ end of the chimeric fragments and once for *y* as the 5′ end). Window pairs with count ≤`--min_int MIN_INT` **(default:** `5`**)** are excluded from further analysis. We then apply one-sided Fisher's exact test to estimate whether the count of the sequence fragments of the window pair (*x*,*y*) is higher than that expected at random, considering the count of *x* and *y* in all other fragments, and the number of all the other fragments (i.e., chimeric and single fragments that do not involve either windows *x* or *y*; see **Figure 1** for Fisher's exact test and odds ratio definitions). Chimeric fragments of windows (*x*,*y*) with odds_ratio ≤`--min_odds_ratio MIN_ODDS_RATIO` **(default:** `1.0`**)** are excluded. The inferred interaction between *x* and *y* is reinforced if the neighboring windows of windows (*x*,*y*) also appear in chimeric fragments. Thus, the size of tested windows is gradually increased in leaps of `SEGLEN` nt up to `SEGLEN × --maxsegs MAXSEGS` **(default:** `5`**)** nt. We start this process with the pair of `SEGLEN`-long windows (*x*,*y*) having the highest number of chimeric fragments in a range of five windows. We then try to extend the initial windows by adding more windows either upstream or downstream of each of the regions *x* and *y*. We limit the length of each region to `SEGLEN × MAXSEGS`. The tested regions are from `MAXSEGS` – 1 windows upstream each original window to `MAXSEGS` – 1 windows downstream. The *P* value of each combination is computed using the above Fisher's exact test, and the combination of windows with the lowest *P* value determines the interacting regions. This pair of interacting regions consists of either the original window pair we started with, or an extended region pair that obtained a better *P* value. Bonferroni correction for multiple hypotheses is applied to the *P* value, taking into account the number of alternative window sizes and positions. At the end of this process, an additional correction is applied by multiplying the *P* value by the final number of region pairs. Reported are interactions with *P* value (after correction) below a certain threshold `--max_pv MAX_PV` **(default:** `0.05`**)**.

Parameters related to the heuristic scores that can be used to rank targets of a specific sRNA

The first score represents the free energy gain by the base-pairing between the interacting RNAs, which is obtained by RNAup[22]. To get the score information for each chimeric fragment, use the following parameters:

`--run_RNAup` **(default:** `False`**)**—Runs RNAup to compute the hybridization free energy of the two RNAs in a chimeric fragment and report it in the final table in a column titled 'Free energy of hybridization'. By default, each fragment in the chimera is padded on both sides, with its flanking genomic nucleotides spanning a length of`--pad_seqs PAD_SEQS` **(default:** `50`**)**. The default RNAup run command is: `RNAup -Xp -w 25 -b -o`. This can be changed by the `--RNAup_cmd RNAUP_CMD` parameter. Statistical evaluation of the hybridization free energy can be computed using `-s SHUFFLES` **(default:** `0`**)**, which shuffles the first sequence `SHUFFLES` times and computes the empirical *P* value of the hybridization free energy, reported in the table in a column titled 'empirical p-value of free energy'. Note that adding the RNAup calculation will increase the run time, especially if the `-s` option is used.

The second score, termed normalized odds ratio (NOR), takes into account the odds ratio (reflecting the tendency of the RNAs to co-appear on Hfq beyond random expectation), and the enrichment of each of the RNAs in the pair on Hfq (see ref. 11 for more details). The NOR score calculation takes into account total RNA in the cell for each gene. To calculate the NOR scores, add the total RNA-seq BAM files using the `--total_RNA` option:

`--total_RNA TOTAL_RNA` **(default:** `None`**)**—Computes the NOR score based on total RNA-seq libraries, and reports it in the final table in a column titled 'Normalized Odds Ratio (NOR)'. Multiple BAM files can be given as a comma-separated list. With RNAtag-seq[19]-like approaches, in which the libraries are the reverse-complement of the transcribed RNA, add the `--total_reverse` **(default:** `False`**)** parameter as well.

▲ **CRITICAL** The total RNA-seq library should be generated under the same conditions as the RIL-seq *hfq-Flag* libraries. Although a single total library would suffice, it is recommended (as we did for this protocol) to build for each of the RIL-seq *hfq-Flag* libraries a total RNA-seq library from the same sample, using the total RNA extracted in Step 26.

Each of the above scores can be used for heuristic ranking of all the targets of a specific sRNA. Importantly, they cannot be applied for global ranking of all the targets. Highly ranked targets (lower RNAup free energy or high NOR values) may be more likely to show an sRNA-mediated change in expression level[11].

(continued)

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 4 | Tips on RILseq_significant_regions.py parameters (continued)

Parameters related to incorporation of annotation information
You can add various types of annotation that will be presented in the final reported table.
The annotation information is derived from various possible sources:
`--bc_dir BC_DIR` **(default:** `None`**)**—The RILseq_significant_regions.py algorithm defines and reports regions based on genomic coordinates, disregarding annotations. Thus, the primary identifier of an S-chimera is its coordinates. If available, annotation is added to the primary identifier. To assign the annotations, download the relevant files from the BioCyc database (see **Box 5**), and store it in `BC_DIR`. If not given, only genomic coordinates will be reported for each region.
`--BC_chrlist BC_CHRLIST BAM_NAME,BIOCYC-NAME` **(default:** `chr,COLI-K12`**)** - When using BioCyc data, if the chromosome/plasmid names in the BioCyc files and in the BAM files differ, you must add a comma-separated conversion list, matching the BAM file names to the BioCyc names. For example, for *E. coli* K12, the BAM file name is 'chr' and the BioCyc name in the EcoCyc database is 'COLI-K12', hence the chosen default value is `chr,COLI-K12`.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

- Biopython (http://biopython.org/; v1.67)
- RNAup (from the ViennaRNA Package; https://www.tbi.univie.ac.at/RNA/index.html; will only work with v.1.x; not compatible with newer versions)
- RILseq (https://github.com/asafpr/RILseq; v0.55)

REAGENT SETUP

▲ CRITICAL All buffers should be made with ultrapure water.

1 M Sodium phosphate buffer, pH 8.0 Mix 6.8 ml of sterile 1 M NaH_2PO_4 and 93.2 ml of sterile 1 M Na_2HPO_4 in a sterile bottle. This buffer can be stored at room temperature (20–24 °C) for a year or more.

Tris HCl 1 M, pH 7.8 Dissolve 121.14 g of Trizma base in 900 ml of water. Adjust the pH to 7.8 with HCl, adjust the volume to 1 liter with water and autoclave the solution. This buffer can be stored at room temperature for a year or more.

EDTA 0.5 M, pH 8.0 Dissolve 146.12 g of EDTA in 900 ml of water. Adjust the pH to 8.0 with NaOH, adjust the volume to 1 liter with water and autoclave the solution. This buffer can be stored at room temperature for a year or more.

Salt solution for wash buffer Mix sodium phosphate and NaCl to a final concentration of 50 mM sodium phosphate and 300 mM NaCl. This buffer is used for the preparation of both wash buffer and RNase buffer. The buffer can be kept at 4 °C for several weeks.

1× Low TE Mix Tris and EDTA to a final concentration of 10 mM Tris and 0.1M EDTA. The buffer can be stored at room temperature for a year or more.

EQUIPMENT SETUP

Hardware setup The software used in the current protocol was tested on a 64-bit machine running GNU/Linux 4.2.8, Debian 9 ('Stretch'). For an efficient run, at least 16 GB RAM is needed. The required disk space and CPU are highly dependent on the size of the sequencing data. For one *E. coli* hfq-Flag library with initial FASTQ files in the size of 2.5 GB, an additional ~5 GB was used for storage and it took ~40 min to run the basic RILseq analysis. On machines with multithreading support, the RILseq run time for multiple libraries will be shorter compared with the sum over the individual library run times. The required disk space will be additive.

Software setup for bioinformatics analysis For detailed instructions, see **Box 5**.

PROCEDURE

Cross-linking and bacteria freezing ● TIMING 24 h

▲ CRITICAL In the following steps, the growth conditions (growth phase, medium and temperature, as well as different stress conditions) can be varied, as explained in the Experimental design section. Hereinafter, we describe the protocol for an experiment with *E. coli* grown to log phase in LB medium at 37 °C with shaking. Additional tested growth conditions are described in Melamed *et al.*[11].

1| Grow 2 ml of overnight cultures of *E. coli hfq-Flag* and *hfq-WT* strains, starting from single colonies, in LB medium at 37 °C with shaking (200 r.p.m.). The next day, dilute the overnight cultures 100-fold in fresh LB medium and grow them at 37 °C with shaking (200 r.p.m.) until they reach an OD_{600nm} value of 0.5. The culture volume to be grown is calculated according to the equation 'culture volume = 40/OD_{600nm} value' (e.g., 80 ml for cells at OD_{600nm} value of 0.5).

2| Transfer the culture to 50-ml tubes and centrifuge the tubes at 4,500*g* for 10–15 min at 4 °C. Note that you may need more than one 50-ml tube per sample. Combine the tubes of the same sample in the next washing step (Step 3).

3| Discard the supernatants, resuspend the pelleted cells in 20 ml of ice-cold 1× PBS and centrifuge at 4,500*g* for 10 min at 4 °C. Repeat the PBS wash and resuspend the pelleted cells in 10 ml of ice-cold PBS.
▲ CRITICAL STEP From this step and until the cells are being frozen in liquid nitrogen (Step 9), samples must be kept on ice at all times.

4| Spread the 10-ml suspension of one sample on a Petri dish and immediately place the plate on a metal block cooled to –20 °C.

5| Expose the cells to 80,000 µJ/cm² of 254-nm UV irradiation, at a distance of 3 cm from the bulbs, using a Stratalinker 1800 UV cross-linker.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

TABLE 1 | Oligonucleotides used for library construction.

Name	Use	Oligo sequence 5′–>3′(*)	Anti-barcode
3Tr3	Ligation to the cDNA 3′ end	P-*AGATCGGAAGAGC*ACACGTCTG-ddC	
AR2	Reverse transcription primer	TACACGACGCTCTTCCGAT	
BC1_adaptor	Ligation to the RNA 3′ end	P-A**ACATTATT***AGATCGGAAGAGC*GTCGTGTA-ddC	AATAATGT
BC2_adaptor	Ligation to the RNA 3′ end	P-A**AGTGTTG***AGATCGGAAGAGC*GTCGTGTA-ddC	CAACACTT
BC3_adaptor	Ligation to the RNA 3′ end	P-A**AGAATTAT***AGATCGGAAGAGC*GTCGTGTA-ddC	ATAATTCT
BC4_adaptor	Ligation to the RNA 3′ end	P-A**ATATGGAC***AGATCGGAAGAGC*GTCGTGTA-ddC	GTCCATAT
BC5_adaptor	Ligation to the RNA 3′ end	P-A**ATCACTTG***AGATCGGAAGAGC*GTCGTGTA-ddC	CAAGTGAT
BC6_adaptor	Ligation to the RNA 3′ end	P-A**CCAAGTCG***AGATCGGAAGAGC*GTCGTGTA-ddC	CGACTTGG
BC7_adaptor	Ligation to the RNA 3′ end	P-A**CAACTCGC***AGATCGGAAGAGC*GTCGTGTA	GCGAGTTG
BC8_adaptor	Ligation to the RNA 3′ end	P-A**CCCGTCTT***AGATCGGAAGAGC*GTCGTGTA-ddC	AAGACGGG
BC9_adaptor	Ligation to the RNA 3′ end	P-A**CCCTACAG***AGATCGGAAGAGC*GTCGTGTA-ddC	CTGTAGGG
BC10_adaptor	Ligation to the RNA 3′ end	P-A**CCCTCGGC***AGATCGGAAGAGC*GTCGTGTA-ddC	GCCGAGGG
BC11_adaptor	Ligation to the RNA 3′ end	P-A**CCGGTACC***AGATCGGAAGAGC*GTCGTGTA-ddC	GGTACCGG
BC12_adaptor	Ligation to the RNA 3′ end	P-A**CGGAGGGC***AGATCGGAAGAGC*GTCGTGTA-ddC	GCCCTCCG
BC13_adaptor	Ligation to the RNA 3′ end	P-A**CTCGGTAC***AGATCGGAAGAGC*GTCGTGTA-ddC	GTACCGAG
BC14_adaptor	Ligation to the RNA 3′ end	P-A**CGGCACTT***AGATCGGAAGAGC*GTCGTGTA-ddC	AAGTGCCG
BC15_adaptor	Ligation to the RNA 3′ end	P-A**CTCTAACT***AGATCGGAAGAGC*GTCGTGTA-ddC	AGTTAGAG
BC16_adaptor	Ligation to the RNA 3′ end	P-A**CTGGATCG***AGATCGGAAGAGC*GTCGTGTA-ddC	CGATCCAG
BC17_adaptor	Ligation to the RNA 3′ end	P-A**GCAGCCAC***AGATCGGAAGAGC*GTCGTGTA-ddC	GTGGCTGC
BC18_adaptor	Ligation to the RNA 3′ end	P-A**GAGATTGT***AGATCGGAAGAGC*GTCGTGTA-ddC	ACAATCTC
BC19_adaptor	Ligation to the RNA 3′ end	P-A**GAGCCATC***AGATCGGAAGAGC*GTCGTGTA-ddC	GATGGCTC
BC20_adaptor	Ligation to the RNA 3′ end	P-A**GTAACTGC***AGATCGGAAGAGC*GTCGTGTA-ddC	GCAGTTAC
P5_Enr	PCR enrichment	5′AATGATACGGCGACCACCGAGATCTACACTCTTT CCCTACACGACGCTCTTCCGATCT 3′	
P7_BC1_Enr	PCR enrichment	5′CAAGCAGAAGACGGCATACGAGATTCGTGTGCGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′	GCACACGA
P7_BC2_Enr	PCR enrichment	5′CAAGCAGAAGACGGCATACGAGATTCGCCAGAGT GACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3′	TCTGGCGA

*Internal barcodes are in bold. External barcodes are underlined. The common part in 3Tr3 and BC adaptors are underlined and in italics. 'P-' denotes 5′ phosphorylation. '-ddC' denotes dideoxy-C at the 3′ end.

Box 5 | Software setup for bioinformatics analysis ● TIMING variable

This box provides information about downloading and installing the RILseq package and preparations for bioinformatics analysis in the PROCEDURE section.

Downloading and installing RILseq package

The RILseq package, as well as additional scripts used in this protocol, is available through the RILseq repository on GitHub (https://github.com/asafpr/RILseq) and can be directly installed through pypi. To install using pypi, run '`pip install RILseq`'. Using pip, the package will be installed in the default directory configured in pip.

▲ **CRITICAL** The RILseq software directory should be either added to your PATH variable, or alternatively all related programs should be run with their full paths.

(continued)

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 5 | Software setup for bioinformatics analysis ● TIMING variable (continued)

Additional site packages are installed under the subdirectory <lib/pythonX.Y/site-packages/RILseq/>.

Reference genome and genome annotation

In the study of Melamed *et al.*[11], we mapped the reads to the genome of *E. coli* K12 MG1655 (NC_000913.2). The reference genome, the annotation file and the relevant files from the EcoCyc database[23] that match this genome, as well as the new *E. coli* K12 MG1655 reference genome NC_000913.3 and its related relevant files, can be downloaded from the RILseq GitHub repository (https://github.com/asafpr/RILseq; files are in the 'data' subdirectory).

▲ **CRITICAL** Download the reference genome in FASTA format and index the file.

1. Generate a directory for the reference genome

```
mkdir <dir-of-reference-genome>
cd <dir-of-reference-genome>
```

2. Save in <dir-of-reference-genome> a FASTA format file of the studied organism <my-organism-genome.fa>.

Format example:

```
>chr
AAACCCCC...
>plasmid1
GGGGCCCCC...
```

▲ **CRITICAL** The names of the genomic elements (e.g., chromosome, plasmid) should be consistent with the mapping in the BAM files and in the input annotation files (if used).

3. Generate indexed files of the genome in <dir-of-reference-genome>:

```
bwa index <my-organism-genome.fa>
```

At the end of this step, you should have in your <dir-of-reference-genome> the following files: my-organism-genome.fa, my-organism-genome.fa.amb, my-organism-genome.fa.ann, my-organism-genome.fa.bwt, my-organism-genome.fa.pac and my-organism-genome.fa.sa.

(Optional) If available, download or generate a genome annotation file in a GFF format.

An example of a line from a GFF file that is used in RILseq:

```
chr EcoCyc exon 190 255 . + . gene_id "EG11277"; transcript_id "EG11277";
```

Note that the GFF file is tab-delimited. The value '.' is assigned to fields that are not required for the RILseq run. For more details on the GFF format, see http://www.ensembl.org/info/website/upload/gff.html.

▲ **CRITICAL** The name of the genomic chromosome/plasmid (column 1 in the GFF file) and the coordinates (column 4) must match the reference genome.

▲ **CRITICAL** The coordinates should be 1-based, i.e., the first position of the genome is 1 and not 0.

▲ **CRITICAL** By default, the subsequent program looks for the word 'exon' in the feature column (third) and 'gene_id' in the attribute column (ninth). These default words can be changed throughout execution. Importantly, only one word can be chosen for the feature field and one word for the attribute field, and all relevant entries should be assigned by these words (**Box 3**).

BioCyc database—(optional)

Transcript annotation information is available for many bacteria in the BioCyc database collection[24] (https://biocyc.org/). In the study by Melamed *et al.*[11], we used files from EcoCyc—the BioCyc database for *E. coli* K12 MG1655[23]. It is possible to use information from other sources, provided that the input files to the scripts comply with the BioCyc format.

Generate a directory for the BioCyc data of your organism. e.g., <ecocyc-data> for the *E. coli* K12 MG1655 data EcoCyc.

```
mkdir <ecocyc-data>
```

Store the following EcoCyc-related files in this directory.

```
genes.dat
promoters.dat
terminators.dat
```

*.fsa (non-redundant files of the genomic element sequences)

If promoter and terminator information is missing, you should retain empty files by the same names.

The files in this directory can be used in the run of RILseq_significant_regions.py (Step 132). They can also be used to generate BioCyc-based GFF files that are either based solely on the genes.dat file or alternatively use additional transcript information from the promoter and terminator data as well. See the RILseq GitHub documentation (https://github.com/asafpr/RILseq).

The relevant files for *E. coli* K12 MG1655 (NC_000913.2 and NC_000913.3) can be downloaded from the RILseq github/data/E_coli_K12/ver2 and ver3 directories, respectively.

▲ **CRITICAL** It is critical that all annotation and sequence files relate to the same reference genome.

Working environment—The command lines throughout the article assume, unless specifically specified otherwise, that all the files needed for the analysis are in a single directory <work-dir>. It is also assumed that the user uses tcsh as the UNIX shell and is working in <work-dir>.

Generate the working directory. Open a shell window and type:

```
tcsh
mkdir <work-dir>
cd <work-dir>
```

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

6| Transfer the irradiated cells to a cold 50-ml tube and immediately place it on ice.
▲ **CRITICAL STEP** Keep the treated and pretreated samples on ice while treating the other samples.

7| Centrifuge the cell suspensions at 4,500*g* for 10 min at 4 °C. Resuspend the cells in 1 ml of ice-cold PBS and transfer the suspensions to precooled 2-ml Eppendorf tubes.

8| Centrifuge the cell suspensions at 17,000*g* for 3 min at 4 °C and discard the supernatants.

9| Freeze the pellets using liquid nitrogen and store them at –80 °C.
■ **PAUSE POINT** Frozen pellets can be kept at –80 °C for at least 1 week.

Lysate preparation ● TIMING 1–2 h
▲ **CRITICAL** Work on ice and use cold (4 °C) wash buffer during the lysate preparation.

10| Prepare fresh wash buffer without recombinant RNase inhibitor.

Reagent	Volume (μl) per reaction	Final concentration
Salts solution for wash buffer[a]	4,920	
IGEPAL 100% (wt/vol)	5	0.1%
Imidazole, 1 M	50	10 mM
Protease inhibitor cocktail	25	
Buffer total volume	5,000	

[a]See Reagent setup.

▲ **CRITICAL STEP** It may take a few minutes for the IGEPAL to dissolve; make sure that it is fully dissolved (by vortexing) before starting to work with the buffer.

11| Withdraw 0.5 ml of the wash buffer without recombinant RNase inhibitor per sample and keep it on ice for Step 29. To the remaining buffer (4.5 ml per sample), add 11.25 μl of recombinant RNase inhibitor (40 U/μl) for a final concentration of 0.1 U/μl. This is the wash buffer. Keep the buffers on ice.

12| For each sample, prepare a new 2-ml Eppendorf tube filled with 0.1-mm diameter glass beads up to the 400-μl mark. Place the tubes on ice.

13| Thaw the frozen pelleted cells from Step 9 on ice, and resuspend the cells with 800 μl of cold wash buffer. Transfer the resuspended cells to the 2-ml tubes prepared in Step 12.

14| Place the tubes in precooled (–20 °C) Retch tube adaptors, place the adaptors in the Retch MM400 mixer mill and grind the cells at a frequency of 30/s for 5 min.

15| Remove the tube adaptors from the mixer mill, take the tubes out of the adaptors and place them on ice for 2 min. Place the tube adaptors at –20 °C for 2 min.

16| Repeat the grinding as described in Step 14.

17| Centrifuge the tubes at 17,000*g* for 2 min at 4 °C. Meanwhile, place the tube adaptors at –20 °C. Transfer the supernatants to new Eppendorf tubes. This is the bacterial lysate.
▲ **CRITICAL STEP** You can leave a small fraction of the supernatant in the grinding tube, to avoid transferring the glass beads.

18| Add 400 μl of cold wash buffer to the grinding tubes containing the glass beads and repeat the grinding as described in Step 14.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

19| Centrifuge the tubes at 17,000*g* for 2 min at 4 °C and transfer the supernatants to the tubes already containing the lysate from Step 17.

20| Centrifuge the collected lysate at 17,000*g* for 15 min at 4 °C. Transfer the lysates to new Eppendorf tubes. The total volume of the lysate should be ~800–900 µl.
▲ **CRITICAL STEP** The lysate should be kept on ice and used within a short time on the same day in the following IP procedure.

21| Transfer 75 µl of the lysate of each sample from Step 20 to a new tube for total RNA extraction (Step 26) and keep the tubes on ice.

Lysate pre-clearing ● TIMING 1.5 h

22| Mix the protein A/G magnetic bead solution by inversion or gentle vortexing. For each sample, aliquot 20 µl of beads in a new tube. Add 200 µl of cold wash buffer, mix by tapping the tubes and place the tubes on a magnetic rack. After the solution becomes clear, discard the buffer while the tubes are still on the rack. Remove the tubes from the magnetic rack. Add 750 µl of cell lysate from Step 20 to the washed magnetic beads and rotate the mixture for 60 min at 4 °C. In the meantime, proceed with Step 23.
▲ **CRITICAL STEP** Ensure that all the beads adhere to the magnet and that the solution is clear before removing the buffer.

Co-immunoprecipitation using anti-Flag antibody ● TIMING 3 h

▲ **CRITICAL** If possible, Steps 23–28 should be performed in a cold room. Moving the samples from 4 °C to room temperature may reduce the efficiency of the protocol.
▲ **CRITICAL** From this step on, it is highly recommended to use LoBind or Maxymum Recovery tubes in order to decrease RNA loss.

23| For each sample, aliquot 20 µl of well-mixed protein A/G magnetic beads in a new tube. Add 200 µl of cold wash buffer, mix by tapping the tubes and place the tubes on a magnetic rack. After the solution becomes clear, discard the buffer while the tubes are still on the rack, without touching the beads. Add 200 µl of cold wash buffer and 3 µl of anti-Flag M2 antibody to the washed beads. Rotate the mixture for 30 min at 4 °C to bind the anti-Flag M2 antibody to the protein A/G magnetic beads.

24| Spin down the tubes from Step 23 and place them on a magnetic rack. After the solution becomes clear, discard the supernatant containing unbound antibody while the tubes are on the magnetic rack. Remove the tubes from the rack, add 200 µl of cold wash buffer, mix by tapping the tubes and place the tubes on the magnetic rack. After the solution becomes clear, discard the buffer while the tubes are on the magnetic rack. Remove the tubes from the magnetic rack and place them on ice.

25| Spin down the tubes from Step 22, which containing protein A/G beads and lysate, and place them on a magnetic rack. After the solution becomes clear, transfer the cleared lysate to the tubes prepared in Step 24. Rotate the tubes for 90 min at 4 °C to bind the Flag-tagged Hfq to the anti-Flag antibodies on the magnetic beads. In the meantime, proceed with Step 26.

26| Extract the total RNA taken in Step 21 according to the standard protocol of TriReagent extraction and store it at –80 °C for up to 5 years. Alternatively, once the total RNA samples are homogenized with the TriReagent, they may be stored at –80 °C for up to 1 month.
! CAUTION TriReagent is toxic and should be handled in a fume hood, working with appropriate protective equipment.

27| Spin down the tubes from Step 25 at 1,500*g* for 10 s at room temperature and place them on a magnetic rack. After the solution becomes clear, discard the lysate and remove the tubes from the magnetic rack.

28| Wash the beads: add 200 µl of cold wash buffer and rotate the tubes for 10 min at 4 °C. Spin down the tubes and place them on a magnetic rack. After the solution becomes clear, discard the buffer. Repeat the washes four additional times. In the meantime, prepare the buffers described in Steps 29 and 30.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Trimming RNA ends ● TIMING 40 min

29| Prewarm the wash buffer without recombinant RNase inhibitor that was withdrawn in Step 11 (500 µl per sample) to room temperature. Dilute the RNase A/T1 mixture 1/20 in water. Mix 1 µl of diluted RNase A/T1 with 499 µl of prewarmed wash buffer without recombinant RNase inhibitor to prepare 500 µl of RNase digestion buffer per sample. This volume includes 20 µl extra, in order to have sufficient amount for all the samples.
▲ CRITICAL STEP When pipetting the RNase A/T1 mixture for the preparation of the 1/20 dilution, make sure not to have an additional amount of enzyme outside the tip, as this would increase the RNase concentration.

30| Prepare 650 µl of SUPERase IN Wash per sample by adding 3.25 µl of SUPERase IN RNase inhibitor to 647 µl of wash buffer, resulting in a final concentration of 0.1 U/µl. Keep the buffer on ice. SUPERase IN RNase inhibitor also inhibits RNase T1, which is not inhibited by the recombinant RNase inhibitor.

31| To trim the exposed parts of the RNAs, add 480 µl of RNase digestion buffer prepared in Step 29 to the tubes from Step 28. Incubate the tubes at 22 °C with gentle agitation for 7 min.
▲ CRITICAL STEP If the magnetic beads precipitate during the incubation time, you can tap the tubes a few times or vortex very gently for 2–3 s.
▲ CRITICAL STEP Note that extended incubation of the samples with the RNase A/T1 mixture may result in overdigestion of the RNA.

32| Place the samples briefly on ice to slow the reaction, spin down the tubes at 1,500*g* for 10 s at room temperature and then place them on a magnetic rack. After the solution becomes clear, remove the RNase digestion buffer. Add 200 µl of cold SUPERase IN Wash prepared in Step 30 to each tube, rotate the tubes for 5 min at 4 °C and then spin down the tubes and place them on the magnetic rack. After the solution becomes clear, discard the buffer. Repeat the wash two additional times. In the last wash, do not discard the buffer and store the tubes on ice until Step 34.
▲ CRITICAL STEP The washes in Step 32 should be performed in a cold room.

5′OH end phosphorylation and 2′P/3′P end dephosphorylation ● TIMING 2.5 h

▲ CRITICAL In Steps 33–35, the 5′OH and 2′P/3′P RNA ends that were formed by the nuclease treatment (Steps 29–32) are modified by T4 polynucleotide kinase (PNK), which is capable of both phosphorylating the 5′ OH end and dephosphorylating the 2′P/3′P ends of the RNA. Thus, PNK generates fragments with 5′P and 3′OH ends that can be subsequently ligated (Step 36).

33| Prepare the following PNK reaction mix (80 µl per sample; *x* samples+1):

Reagent	Volume (µl) per reaction	Final concentration
Nuclease-free water	65.2	
PNK buffer, 10×	8	1×
ATP (100 mM)[a]	0.8	1 mM
Recombinant RNase inhibitor (40 U/µl)	2	1 U/µl
T4 polynucleotide kinase (10,000 U/ml)	4	0.5 U/µl
Mix total volume	80	

[a]Included in the T4 RNA Ligase kit.

34| Place the tubes from Step 32 on a magnetic rack and, after the solution becomes clear, discard the wash buffer. Remove the tubes from the magnetic rack and add 80 µl of PNK reaction mix to each tube. Incubate the tubes for 2 h at 22 °C with a gentle agitation.
▲ CRITICAL STEP If the magnetic beads precipitate during the incubation time, you can tap the tubes a few times or vortex them very gently for 2–3 s once every 30 min.

35| Place the tubes on a magnetic rack and, after the solution becomes clear, remove the PNK reaction mixture and add 200 µl of cold wash buffer to each tube. Rotate the tubes for 5 min at 4 °C. Place the tubes on the magnetic rack and,

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

after the solution becomes clear, discard the buffer. Repeat the washes with cold wash buffer two additional times. To keep the beads from drying out, retain the wash buffer from the last wash until the ligation mixture is ready in the next step and remove it before the addition of the ligation mixture.

Ligation of Neighboring RNAs ● TIMING overnight

36| Set up the following ligation mix (80 μl per sample; *x* samples + 2) at room temperature:

Reagent	Volume (μl) per reaction	Final concentration
10× T4 RNA ligase buffer	8	
DMSO (100% (vol/vol))	7.2	
ATP (100 mM)	0.8	1 mM
PEG 8000 (50% (wt/vol))	32	
Recombinant RNase inhibitor (40 U/μl)	1.2	0.6 U/μl
T4 RNA ligase 1 (30,000 U/ml)	7.2	2.7 U/μl
Nuclease-free water	23.6	
Mix total volume	80	

▲ **CRITICAL STEP** Set up the mix at room temperature to prevent DMSO precipitation. Pipette the PEG very slowly for accurate aspiration, as it is very viscous. Mix well by tapping the tube, as the mixture is very viscous, and spin down.

37| Add 80 μl of ligation mixture to each tube. Mix well by tapping the tube, and if necessary spin down.

38| Incubate the tubes at 22 °C overnight.

Protein digestion ● TIMING 2.5 h

39| Prepare fresh wash buffer (the amounts per single sample are detailed):

Reagent	Volume (μl) per reaction	Final concentration
Salts solution for wash buffer	638	
IGEPAL 100% (wt/vol)	0.65	0.1%
Imidazole 1 M	6.5	10 mM
Recombinant RNase inhibitor (40 U/μl)	1.625	0.1 U/μl
Protease inhibitor cocktail	3.25	
Buffer total volume	650	

▲ **CRITICAL STEP** Keep the buffer on ice.

40| Place the tubes from Step 38 in the magnetic rack.
▲ **CRITICAL STEP** It may take a few minutes for the magnetic beads to pellet because of the viscosity of the solution.

41| After the solution becomes clear, discard the ligation mix and add 200 μl of cold wash buffer prepared in Step 39 to each tube. Rotate the tubes for 5 min at 4 °C. Spin down the tubes and place them on the magnetic rack. After the solution becomes clear, discard the buffer. Repeat the washes with cold wash buffer two additional times.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

42| Prepare fresh proteinase K reaction mix (300 μl per sample; *x* samples +1) by mixing the following components at room temperature:

Reagent	Volume (μl) per reaction	Final concentration
Nuclease-free water	255	
Tris-HCl, pH 7.8 (1 M)	15	50 mM
NaCl (5 M)	3	50 mM
IGEPAL (100% (wt/vol))	0.3	0.1%
Imidazole (1 M)	3	10 mM
SDS (20% (wt/vol))	15	1%
EDTA pH 8.0 (0.5 M)	3	5 mM
β-mercaptoethanol (14.3 M)	0.1	5 mM
Recombinant RNase inhibitor (40 U/μl)	1.3	0.1 U/μl
Proteinase K (20 mg/ml)	5	
Mix total volume	300	

! CAUTION β-mercaptoethanol is toxic. The stock should be handled in a fume hood, while working with appropriate protective equipment.

43| Add 300 μl of proteinase K reaction mix to each tube. Mix well and incubate the samples for 2 h at 55 °C with gentle agitation.
▲ CRITICAL STEP If the magnetic beads precipitate during the incubation, you can flick the tubes a few times or vortex them very gently for 2–3 s once every 30 min.

RNA extraction according to the standard TriReagent LS protocol ● TIMING 14 h
▲ CRITICAL From this step and until the end of the experimental procedure (Step 125), use filter tips only.

44| Add 0.9 ml of TriReagent LS (prewarmed to room temperature) to the tubes containing both the beads and the proteinase K buffer. Resuspend thoroughly by pipetting to homogenization, and incubate the tubes for 5 min at room temperature.
! CAUTION TriReagent LS is toxic and should be handled in a fume hood, while working with appropriate protective equipment.

45| Add 200 μl of chloroform, mix by inversion of the tubes for 15 s and then incubate them for 10 min at room temperature.
! CAUTION Chloroform is toxic and should be handled in a fume hood, while working with appropriate protective equipment.

46| Centrifuge the tubes at 17,000*g* for 10 min at 4 °C and transfer the upper phase (~600–700 μl) to a new Eppendorf tube.
! CAUTION The lower phase contains phenol and should be disposed of as toxic waste.
▲ CRITICAL STEP Be careful not to transfer the lower phase to the new tube.

47| Add 500 μl of isopropanol, mix thoroughly by inversion of the tubes and incubate them for 10 min at room temperature.

48| Spin down the tubes, add 1.5 μl of GlycoBlue and mix by pipetting. Incubate the tubes overnight at –20 °C.
▲ CRITICAL STEP As the RNA amount at this step is very small, the addition of GlycoBlue is critical to precipitating the RNA.
■ PAUSE POINT Samples can be stored at –20 °C for up to 1 month before proceeding to the next step.

49| Centrifuge the tubes at 17,000*g* for 15 min at 4 °C and discard the supernatant.
? TROUBLESHOOTING

50| Wash the pellet by the addition 1 ml of freshly made 75% (vol/vol) ethanol, followed by centrifugation at 17,000*g* for 5 min at 4 °C and discarding of the supernatant. Repeat this washing step once. After discarding the supernatant from

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

the second wash, spin down the tubes again and discard the remaining supernatant. Leave the tubes open for 10 min to dry the pellet.
▲ CRITICAL STEP When performing the washes, be careful not to lose the RNA pellet.

51| Resuspend the pellet in 20 μl of nuclease-free water by tapping the tube. If the resuspension is difficult, warm the tube at 37 °C for a few minutes and flick the tube. After the resuspension, place the RNA immediately on ice.
▲ CRITICAL STEP It is important not to resuspend the pellet by pipetting, as the amount of the RNA is very small and the pellet might stick to the tip.

52| Analyze the total RNA samples (diluted 1/10) from Step 26 using a Bioanalyzer RNA 6000 Nano kit and the RNA samples (not diluted) from Step 51 using a Bioanalyzer RNA 6000 Pico kit to assess RNA quantity and size distribution. Typically, the concentrations obtained for the total RNA samples and for the samples from Step 51 range between 1,000 and 5,000 pg/μls and between 500 and 1,500 ng/μl, respectively. Representative RNA profiles are shown in **Figure 4a–d**. For more details, see 'Anticipated Results'.
? TROUBLESHOOTING

53| Transfer 15 μl of each RNA solution from Step 51 to PCR tubes and add 1 μl of recombinant RNase inhibitor (40 U/μl).
■ PAUSE POINT Continue with RNA-seq library construction or store the samples at –80 °C for up to 1 month.

Fragmentation and DNase–FastAP combined treatment ● TIMING 45 min
▲ CRITICAL RIL-seq libraries are constructed using the RNAtag-Seq protocol[19] with a few modifications to allow capture of short RNA fragments. The library construction process includes a step of random RNA fragmentation. This step is required to obtain full coverage of long RNAs, as RNA that was extracted in prior steps (Steps 44–51) includes fragments of >800 nt (**Fig. 4b**), which will not be represented in the library unless they are fragmented.

54| Add 4 μl of 10× FastAP buffer to the RNA tubes from Step 53, mix well and incubate the tubes in a preheated thermal cycler for 1.5 min at 92 °C.

55| Place the samples on ice and prepare DNase–phosphatase reaction mix (20 μl per sample; *x* samples + 1):

Reagent	Volume (μl) per reaction	Final concentration (volume = 40 μl)
Recombinant RNase inhibitor (40 U/μl)	1	1 U/μl
TURBO DNase (2 U/μl)	4	0.2 U/μl
FastAP (1 U/μl)	10	0.25 U/μl
Nuclease-free water	5	
Mix total volume	20	

56| Add 20 μl of the mixture to each tube containing fragmented RNA, mix well and then incubate the tubes in a thermal cycler for 30 min at 37 °C.

RNA cleanup using RNA Clean & Concentrator-5 kit ● TIMING 1 h
57| Increase the reaction volume to 80 μl by adding 40 μl of nuclease-free water and transfer the mixture to an Eppendorf tube. Clean up the RNA according to the kit standard protocol for recovering RNA species of ≥17 nt. In the final step, elute the RNA with 8 μl of nuclease-free water.

58| As a quality control, use an Agilent Bioanalyzer RNA 6000 Pico kit to analyze random samples and check their fragmentation profile. Representative RNA profiles are shown in **Figure 4e**,**f**; for more details see 'Anticipated Results'.

Ligation of a 3′ adaptor (RNA/DNA) ● TIMING 2.5 h
59| Make a list of the barcoded adaptors and the corresponding samples that will be ligated. The barcoded adaptors are listed in **Table 1**. In a PCR tube, mix 5 μl of dephosphorylated RNA from Step 57 with 1 μl of the respective barcoded adaptor (100 μM). Heat the tube at 70 °C for 2 min and place it on ice.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

60| Set up the following ligation mix (14 μl per sample; *x* samples + 2):

Reagent	Volume (μl) per reaction	Final concentration (volume = 20 μl)
T4 RNA ligase buffer, 10×	2	
DMSO (100% (vol/vol))	1.8	
ATP (100 mM)	0.2	1 mM
PEG 8000 (50% (wt/vol))	8	
Recombinant RNase inhibitor (40 U/μl)	0.3	0.6 U/μl
T4 RNA ligase 1 (30,000 U/ml)	1.7	2.55 U/μl
Mix total volume	14	

▲ **CRITICAL STEP** Additional barcoded adaptors can be found at Shishkin *et al.*[19].
▲ **CRITICAL STEP** Set up the mixture at room temperature to prevent DMSO precipitation. Pipette very slowly for accurate aspiration of PEG, as it is very viscous. Mix well by tapping the tube, as the solution is very viscous, and spin down.

61| Add 14 μl of ligation mix to each tube containing 6 μl of denatured RNA + adaptor. Mix well by tapping the tubes, as the solution is very viscous, and incubate the tubes at 22 °C for 2 h.

Pooling the libraries and RNA cleanup using RNA Clean & Concentrator-5 kit ● TIMING 30 min

62| Add 60 μl of RLT buffer to each sample to inhibit ligase activity and mix the solution well.

63| Pool the samples together in one tube and clean up the RNA using the Clean & Concentrator-5 kit. Use the kit standard protocol for recovering RNA species of ≥17 nt. In the final step, elute twice into the same tube with 15 μl of nuclease-free water. Store the samples on ice for Step 70.
▲ **CRITICAL STEP** If you have more than eight samples, load the mixture of pooled samples, binding buffer and ethanol onto two columns and elute twice with 7.5 μl of nuclease-free water. Combine the eluates in one tube.

rRNA removal using a Ribo-Zero kit ● TIMING 45 min

64| Pipette 225 μl of well-mixed Ribo-Zero Magnetic Beads into a 1.5-ml RNase-free Eppendorf tube.
▲ **CRITICAL STEP** Mix the magnetic beads well by pipetting or gentle vortexing. Do not vortex at high speed.

65| Place the tube on the magnetic rack for at least 1 min. When the solution appears clear, discard the supernatant.

66| Remove the tube from the magnetic rack and add 225 μl of nuclease-free water. Mix well and place the tube on the magnetic rack for at least 1 min. When the solution appears clear, discard the supernatant.

67| Repeat the washing step once more, as described in Step 66.

68| Remove the tube from the magnetic rack. Add 65 μl of Magnetic Bead Resuspension Solution to the tube. Mix well by pipetting.

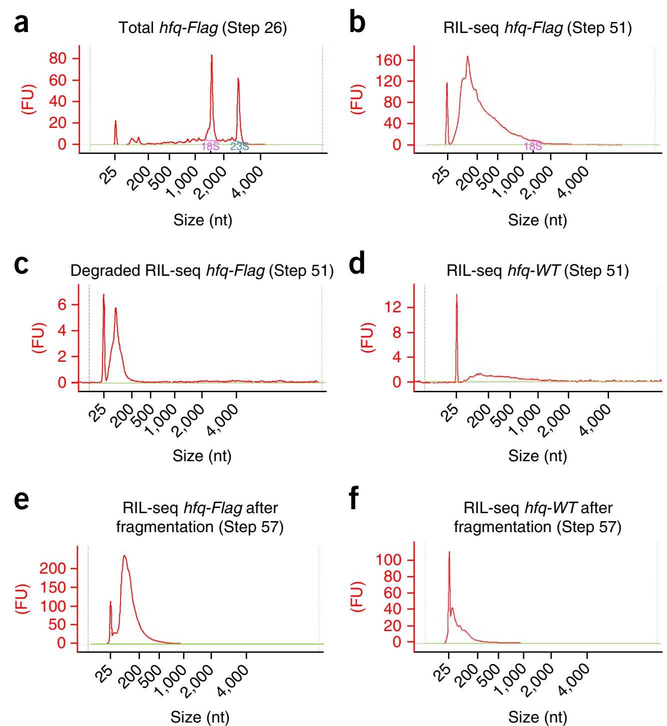

Figure 4 | Typical Bioanalyzer electropherograms showing the RNA profile at different steps of the protocol. (**a**) A Total RNA sample extracted from a lysate of the *hfq-Flag* strain. (**b**–**d**) RNA samples extracted at Step 51. (**b**) RNA of an *hfq-Flag* strain (estimated concentration: 4,242 pg/μl). (**c**) RNA of an *hfq-Flag* strain that was degraded (estimated concentration: 269 pg/μl). (**d**) RNA of an *hfq-WT* strain (estimated concentration: 54 pg/μl). (**e**,**f**) RNA samples after fragmentation at Step 57. (**e**) RNA of an *hfq-Flag* strain. (**f**) RNA of an *hfq-WT* strain. Analyses were performed using the Bioanalyzer RNA 6000 Nano kit (**a**) or the RNA 6000 Pico kit (**b**–**f**). FU, fluorescence units.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

69| Add 1 µl of RiboGuard RNase Inhibitor (supplied with the Ribo-Zero kit) to the bead suspension and mix by tapping the tube. Store the bead suspension at room temperature, until using it in Step 72.

70| In a new tube, prepare the rRNA removal mix:

Reagent	Volume (µl) per reaction
Pooled RNA from Step 63	28
RiboZero reaction buffer, 10×	4
rRNA Removal solution	8
Mix total volume	40

71| Incubate the tube for 10 min at 68 °C, and then for 5 min at room temperature.

72| Add the treated RNA from Step 71 to the tube containing the washed Ribo-Zero Magnetic Beads from Step 69 and, without changing the pipette tip, immediately mix by pipetting at least 10 times.

73| Vortex the tube immediately at medium speed for 10 s, and place the tube at room temperature for 5 min.

74| Vortex at medium speed and then incubate the tube at 50 °C for 5 min.

75| Remove the tube from 50 °C and immediately place it on a magnetic rack for at least 1 min.

76| Carefully collect the supernatant (85–90 µl) containing the rRNA-depleted RNA and transfer it into a new tube. Place the tube on ice.
■ **PAUSE POINT** Store the rRNA-depleted RNA at –80 °C for up to 1 month or proceed to the next step.

rRNA-depleted RNA purification ● TIMING 45 min

77| Purification of rRNA-depleted samples is done by using RNAClean XP Beads. First, gently vortex the RNAClean XP Beads until they are well dispersed. Measure the exact volume of the rRNA-depleted RNA solution. Add a 2.5 reaction volume (2.5×) of well-mixed RNAClean XP Beads and a 1.5 reaction volume (1.5×) of isopropanol to the rRNA-depleted RNA solution. Mix thoroughly by pipetting the entire volume 15 times and incubate the solution at room temperature for 15 min.
▲ **CRITICAL STEP** Adjust the RNAClean XP Beads and isopropanol volumes according to the rRNA-depleted RNA solution volume. For example, if the rRNA-depleted RNA solution volume is 90 µl, use 225 µl (2.5×) of RNAClean XP Beads and 135 µl (1.5×) of isopropanol.
▲ **CRITICAL STEP** The addition of isopropanol in the indicated ratio is critical to keeping short RNAs (<70 nt) in the RNA pool. The isopropanol and RNAClean XP bead ratios were both optimized in order to remove nonligated adaptors while saving short RNA fragments.

78| Place the tube(s) on the magnetic rack at room temperature for at least 5 min, until the solution appears clear.

79| Discard the supernatant.
▲ **CRITICAL STEP** At this step, the RNA is bound to the beads. Be careful not to pipette out the beads.

80| While the tube is still on the magnetic rack, add 500 µl of freshly prepared 80% (vol/vol) ethanol to the tube. Incubate the tube at room temperature for at least 30 s without removing the tube from the magnetic rack. Discard the supernatant without touching the beads. Repeat the ethanol wash one more time, and discard all the remaining ethanol droplets.
▲ **CRITICAL STEP** It is important to avoid removing the tube from the magnetic rack during the ethanol washes, in order to maintain the RNA binding to the beads.

81| Leave the tube open on the magnetic rack and allow the beads to air-dry at room temperature for 15 min.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

82| Remove the tube from the magnetic rack, add 12 µl of nuclease-free water and resuspend the beads by pipetting 10 times.

83| Incubate the tube at room temperature for 2 min and then place the tube back onto the magnetic rack. After the solution appears clear, collect 11 µl of the RNA-containing supernatant and transfer it to a new tube. Avoid carryover of magnetic particles. Store the solution on ice. At this step, the RNA is eluted in water.
■ **PAUSE POINT** Store the sample at –80 °C for up to 1 month or proceed to the next step.

First-strand cDNA synthesis using the SuperScript III First-Strand Synthesis system ● TIMING 1.25 h
84| Add 1 µl of 50 µM AR2 primer (**Table 1**) to the eluted samples from Step 83 and mix well.

85| Heat the mixture to 70 °C for 2 min and immediately place it on ice.

86| Set up the following reverse transcription mix on ice:

Reagent	Volume (µl) per reaction	Final concentration (volume = 25 µl)
dNTP mix (10 mM)	1.25	0.5 mM
10× RT buffer	2.5	
$MgCl_2$ (25 mM)	5	5 mM
DTT (0.1 M)	2.5	0.01 M
RNaseOUT (40 U/µl)	0.5	0.8 U/µl
SuperScript III RT (200 U/µl)	1.25	10 U/µl
Mix total volume	13	

87| Add 13 µl of reverse transcription mix to the mixture from Step 85. Mix well and incubate the mixture at 50 °C for 55 min.
■ **PAUSE POINT** Store the sample at –20 °C for up to 1 month or proceed to the next step.

RNA degradation ● TIMING 20 min
88| Add 2.5 µl of 1N NaOH to the tube from Step 87 and incubate it at 70 °C for 12 min.

89| Add 5 µl of freshly diluted 0.5 M acetic acid and mix well.

cDNA cleanup ● TIMING 45 min
90| Add 7.5 µl of nuclease-free water for a final volume of 40 µl and transfer it to a new Eppendorf tube.

91| Add 1.5× (60 µl) isopropanol and 2.5× (100 µl) AMPure XP beads, mix by pipetting 15 times and incubate the tube at room temperature for 15 min.

92| Place the tube on a magnetic rack for ~5 min until the solution is clear and discard the supernatant.

93| Wash the beads with 200 µl of freshly prepared 80% (vol/vol) ethanol without removing the tube from the magnetic rack. Incubate the tube for 30 s and discard the supernatant. Repeat the wash once more, and discard all the remaining ethanol droplets.

94| Leave the tube open on the magnetic rack and allow the beads to air-dry at room temperature for 10 min.

95| Remove the tube from the magnetic rack, resuspend the beads in 5 µl of nuclease-free water and incubate the tube for 2 min at room temperature.
▲ **CRITICAL STEP** Keep the cDNA in the tube with the beads for the following ligation step.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Ligation of a second adaptor at the cDNA 3′ end (ssDNA/ssDNA) ● TIMING overnight

96| Add 2 μl of 40 μM 3Tr3 adaptor (**Table 1**) to the cDNA+beads solution, incubate the tube at 75 °C for 3 min and place it on ice.

97| Set up the following ligation mix at room temperature:

Reagent	Volume (μl) per reaction	Final concentration (volume = 20 μl)
T4 RNA ligase buffer, 10×	2	
DMSO (100% (vol/vol))	0.8	
ATP (100 mM)	0.2	1 mM
PEG 8000 (50% (wt/vol))	8.5	
T4 RNA ligase 1 (30,000 U/ml)	1.5	2.25 U/μl
Mix total volume	13	

▲ **CRITICAL STEP** Set up the mix at room temperature to prevent DMSO precipitation. Pipette very slowly for accurate aspiration of PEG, as it is very viscous. Mix well by tapping the tube, as the mixture is very viscous, and spin down.

98| Add 13 μl of ligation mix to the tube from Step 96. Mix well by tapping the tubes, as the solution is very viscous. Spin down and incubate the tube overnight at 22 °C.

Cleanup of cDNA ● TIMING 45 min

99| Increase the volume to 40 μl by adding 20 μl of nuclease-free water. Add 1.5× (60 μl) isopropanol and 2.5× (100 μl) AMPure XP beads, mix by pipetting 15 times and incubate the tube at room temperature for 15 min.

100| Place the tube on a magnetic rack for ~5 min until the solution is clear and discard the supernatant.

101| Wash the beads with 200 μl of freshly prepared 80% (vol/vol) ethanol without removing the tube from the magnetic rack. Incubate the tube for 30 s and discard the supernatant. Repeat the wash once more and discard all the remaining ethanol droplets. Leave the tube open to air-dry for 10 min.

102| Resuspend the dried beads in 25 μl of nuclease-free water. Incubate the tube for 2 min at room temperature and then place the tube on the magnetic rack. After the solution appears clear, transfer the supernatant containing the eluted cDNA to a new Eppendorf tube.

■ **PAUSE POINT** Store the sample at –20 °C for up to 1 month or proceed to the next step.

2nd Cleanup of cDNA to remove the remaining adaptors ● TIMING 45 min

103| Add 1.5× (37.5 μl) isopropanol and 2.5× (62.5 μl) AMPure XP beads, mix by pipetting 15 times and incubate the tube at room temperature for 15 min.

104| Place the tube on a magnetic rack for ~5 min until the solution is clear and discard the supernatant.

105| Wash the beads with 200 μl of freshly prepared 80% (vol/vol) ethanol without removing the tube from the magnetic rack. Incubate the tube for 30 s and discard the supernatant. Repeat the wash once more and discard all the remaining ethanol droplets. Leave the tube open to air-dry for 5 min.

106| Resuspend the dried beads in 25 μl of nuclease-free water. Incubate the tube for 2 min at room temperature and then place the tube on the magnetic rack. After the solution appears clear, transfer the supernatant containing the eluted cDNA to a new Eppendorf tube.

■ **PAUSE POINT** Store the sample at –20 °C for up to 5 years or proceed to the next step.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

PCR enrichment test to determine the final number of PCR cycles ● TIMING 45 min

107| Prepare the following PCR mix:

Reagent	Volume (μl) per reaction
Nuclease-free water	3.5
P5_Enr primer (12.5 μM)	2
P7_BC1_Enr primer (12.5 μM)	2
cDNA	5
KAPA HIFI HotStart ReadyMix, 2×	12.5
Mix total volume	25

▲ **CRITICAL STEP** To increase multiplexing, it is possible to use P7 enrichment primers that carry different indexes. An additional primer is included in **Table 1** (see ref. 19 for more information).

108| Prepare 8-μl aliquots of PCR mix in three PCR tubes.

109| Amplify the cDNA using 9/12/15 cycles, with the following PCR program:

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	3 min	1
Denaturation	95	30 s	
Annealing	55	30 s	9/12/15
Extension	72	30 s	
Final extension	72	10 min	1
Hold	4	∞	

■ **PAUSE POINT** Store the amplified libraries at –20 °C for up to 1 month or proceed to the next step.

PCR cleanup ● TIMING 1.5 h

110| Add nuclease-free water to a final volume of 25 μl. Add 1.5× AMPure XP beads, mix by pipetting 15 times and incubate the tube at room temperature for 15 min.

111| Place the tube on a magnetic rack for ~5 min until solution is clear and discard the supernatant.

112| Wash the beads with 200 μl of freshly prepared 80% (vol/vol) ethanol without removing the tube from the magnetic rack. Incubate the tube for 30 s and discard the supernatant. Repeat the wash once more and discard all the remaining ethanol droplets. Leave the tube open to air-dry for 5 min.

113| Resuspend the dried beads in 10 μl of 1× low TE. Incubate the tube for 2 min at room temperature and then place the tube on the magnetic rack. After the solution appears clear, transfer the supernatant containing the eluted PCR fragments to a new Eppendorf tube.

114| Analyze 1 μl of each PCR product, using the Qubit dsDNA HS Assay Kit according to the manufacturer's instructions. Estimate the PCR product concentration (nanograms per microliter). The typical concentration range is 2–70 ng/μl.
? TROUBLESHOOTING

115| Analyze 1 μl of each PCR product using an Agilent TapeStation and High-Sensitivity D1000 ScreenTape. Estimate the average size of the PCR products. A representative profile is shown in **Figure 5a**; for more details, see 'Anticipated Results'.
▲ **CRITICAL STEP** The minimal library fragment size is 185 bp (insert size is ~50 bp). Smaller fragments represent incomplete amplification products, primer dimers or fragments that will yield very short reads. Therefore, if the libraries contain products smaller than 185 bp, the sample should be cleaned up once more with 1× AMPure XP beads, as described in Steps 110–113, and the Qubit and TapeStation analyses should be redone.
? TROUBLESHOOTING

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

116| On the basis of the analyses done in Steps 114 and 115, calculate the molar concentration of the PCR product according to the equation below and estimate the minimal number of cycles that would provide sufficient material for sequencing. Calculation equation:

$$\text{Concentration}(\text{nM}) = \frac{\text{ng}/\mu\text{l} \times 10^6}{650\,\text{Da} \times \text{Average library product size (bp)}}$$

(The average weight of a DNA base pair (sodium salt) is 650 Da). For example, nine cycles of amplification gave a concentration of 428 ng/ml = 0.428 ng/µl (Qubit), and the mean peak size was 185 bp (TapeStation). The molar concentration is

$$\frac{0.428 \times 10^6}{650 \times 185} = 3.56\,\text{nM}$$

This concentration was sufficient for sequencing in the NextSeq machine, and therefore the final number of PCR cycles in Step 117 was set to 9.

■ **PAUSE POINT** Store the amplified libraries at –20 °C for up to 5 years or proceed to the next step.

PCR amplification of the libraries ● TIMING 1 h

117| Choose the optimal PCR cycle number based on the analysis described in Steps 114–116.

▲ **CRITICAL STEP** Choose the minimal number of PCR cycles in order to decrease PCR duplications.

118| Set up the following PCR mix:

Reagent	Volume (µl) per reaction
Nuclease-free water	7
P5_Enr primer (12.5 µM)	4
P7_BC1_Enr primer (12.5 µM)	4
cDNA	10
KAPA HIFI HotStart ReadyMix, 2×	25
Mix total volume	50

119| Mix well and amplify the cDNA using the chosen number of cycles (*x*) with the following PCR program:

Step	Temperature (°C)	Time	Number of cycles
Initial denaturation	95	3 min	1
Denaturation	95	30 s	
Annealing	55	30 s	X
Extension	72	30 s	
Final extension	72	10 min	1
Hold	4	∞	

■ **PAUSE POINT** Store the sample at –20 °C for up to 5 years or proceed to the next step.

Library cleanup ● TIMING 1.5 h

120| Add 1.5× (75 µl) AMPure XP beads, mix by pipetting 15 times and incubate the tube at room temperature for 15 min.

121| Place the tube on a magnetic rack for ~5 min until the solution is clear and discard the supernatant.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

122| Wash the beads with 200 µl of freshly prepared 80% (vol/vol) ethanol without removing the tube from the magnetic rack. Incubate the tube for 30 s and discard the supernatant. Repeat the wash once more and discard all the remaining ethanol droplets. Leave the tube open to air-dry for 10 min.

123| Resuspend the dried beads in 25 µl of 1× low TE. Incubate the tube for 2 min at room temperature and then place the tube on the magnetic rack. After the solution appears clear, transfer the supernatant containing the eluted PCR fragments to a new Eppendorf tube.

124| Analyze the amplified cDNA using Agilent TapeStation and Qubit as described in Steps 114–116, and calculate the molar concentrations of the pooled libraries. The typical range of the molar concentrations is 1–15 nM.
▲ **CRITICAL STEP** If the libraries contain products smaller than 185 bp, the sample should be cleaned up once more with 1× AMPure XP beads.
? TROUBLESHOOTING
■ **PAUSE POINT** The amplified libraries can be stored at −20 °C for up to 5 years.

Sequencing of libraries ● TIMING 12 h
125| Sequence the libraries on a NextSeq or HiSeq sequencing machine, using paired-end settings with at least 40 cycles per read.

Preliminary steps for running the RILseq software ● TIMING variable
▲ **CRITICAL** The RILseq software requires the mapped read files *.bam and *.bam.bai for each library. In addition, a FASTA format file of the reference genome is required. A GFF format gene annotation file and/or BioCyc related annotation database are optional (**Box 5**). RILseq analysis can be applied to data generated by various library-generation protocols and by various mapping tools. The computational pipeline for the generation of the *.ba and *.bam.bai files tailored to the library protocol presented above is described in **Box 2**.
▲ **CRITICAL** In the following, files and directories are indicated by <>.

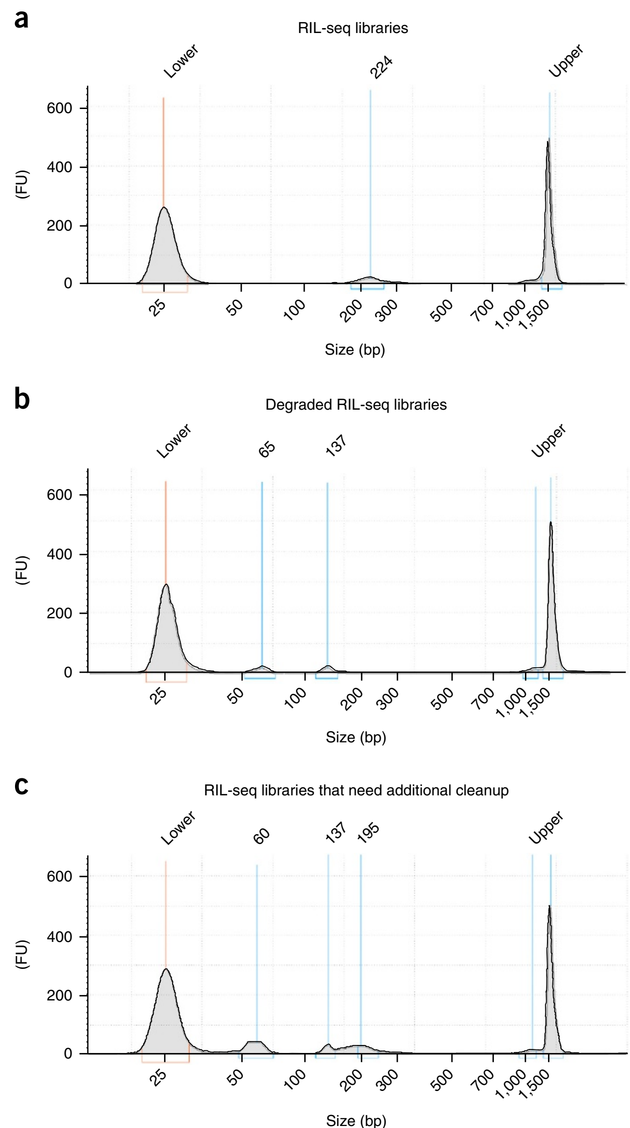

Figure 5 | TapeStation electropherograms. (**a**) RIL-seq libraries with a peak size of 224 bp. (**b**) PCR amplification products of a sample in which the RNA was degraded during library preparation. Only peaks of primers (65 bp) and primer dimers (137 bp) are present. (**c**) RIL-seq libraries that need additional cleanup. Peaks of 60 and 137 bp are present in addition to the library peak of 195 bp. 'Lower' and 'Upper' peaks are size markers of 25 and 1,500 bp, respectively. FU, fluorescence units.

126| If not done yet (**Box 5**), open a tcsh shell and generate a working directory. Open a shell window and type:

```
tcsh

mkdir <work-dir>

cd <work-dir>
```

127| Copy all the relevant *.bam and *.bam.bai input files to this directory.

128| (Optional) If a GFF file `transcript_annotation_file.gff` is available, copy it to `<dir-of-reference-genome>`.
▲ **CRITICAL STEP** The annotation file coordinates should follow the reference genome.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Identification of chimeric fragments ● TIMING variable

129| Run the map_chimeric_fragments.py script to generate a list of sequenced fragments with their mapping positions and their classification as single or chimeric fragments. The run parameters listed below are tailored to the sequencing data generated as described in this protocol. Tips and comments on parameter selection are described in **Box 3**.

Apply the map_chimeric_fragments.py script to each of the *.bam files in `<work-dir>`

```
foreach library (*.bam)

foreach? map_chimeric_fragments.py -r -t <transcript_annotation_file.gff> <dir-of-reference-genome/my_organism_genome.fa> $library >! {$library}_all_fragments.txt &

foreach? end
```

The `-t <transcript_annotation_file.gff>` parameter is optional. When an annotation file of transcript boundaries is given, sequenced fragments whose pair mates mapped to the same annotated transcript are considered 'single'. If an annotation file is not used, consider raising the value of the `-s` parameter (**Box 3**) from its default value of 1000 to the expected size of a transcript.

At the end of this step, you should have in `<work-dir>` a tab-delimited file with the suffix all_fragments.txt for each library that includes the information for each mapped fragment in the following format:

Fields 1–3 (chromosome, coordinate, strand)—the genomic location where the 5′ end of the sequenced fragment mapped

Fields 4–6 (chromosome, coordinate, strand)—the genomic location where the 3′ end of the sequenced fragment mapped

Field 7—original read name (for both reads of a fragment)

Field 8—classification of the sequenced fragment: single—when the pair mates of a sequenced fragment map to the same transcript; chimera—when the pair mates map to two different transcripts

Example output lines:

```
chr 2701420 -    chr    2701233 -    NS500183:127:H7WVWBGXX:3:13606:20305:12283    single

chr 2443052 -    chr    3613274 +    NS500183:127:H7WVWBGXX:4:21412:20728:1492    chimera
```

Additional files related to the independent mapping of the 25-nt-long ends of Read1 and Read2 are generated and stored in a directory within `<work-dir>` named `<remapped_data>`: *_ends_1.fastq, *_ends_1.bam, *_ends_1.bam.bai, *_ends_2.fastq, *_ends_2.bam, *_ends_2.bam.bai

Note that when additional filtering is applied (see `--dust_thr` and `--maxG` in **Box 3**), it is applied independently to the Read1 and Read2 files. Consequently, the FASTQ and BAM files can have different numbers of reads. The *all_fragments.txt files, however, will report only fragments for which both reads (pair mates) passed the additional filtering and both were properly mapped.

▲ **CRITICAL STEP** The reference genome should be identical to the one used for the generation of the BAM files in Step 127.

▲ **CRITICAL STEP** For the subsequent analysis, map_chimeric_fragments.py should output the original transcript coordinates and not its reverse, (see discussion of the -r parameter in **Box 3**).

▲ **CRITICAL STEP** The chimeric fragments annotated as chimeras at this stage must be evaluated statistically (next step hereinafter) to generate a reliable set of putative RNA–RNA interactions.

▲ **CRITICAL STEP** The '&' at the end of the command line should be applied only with multicore-processor or multiple -processor platforms.

▲ **CRITICAL STEP** When using technologies other than two-dye sequencing-by-synthesis sequencers, the default value of the --maxG parameter must be changed (**Box 3**).

? TROUBLESHOOTING

Identification of statistically significant chimeras (S-chimeras) ● TIMING variable

130| Test the reproducibility of replicates and determine whether replicate libraries can be unified.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Box 6 | Evaluation of replicate reproducibility ● TIMING variable

When RIL-seq is run in replicate experiments, it is recommended to test the reproducibility of the results from libraries of replicate experiments before application of the RILseq_significant_regions.py script. High reproducibility enables library unification, whereas low reproducibility might suggest exclusion of specific replicate(s). The reproducibility of the data is evaluated by comparing the numbers of corresponding mapped sequenced fragments between each pair of libraries. In practice, the genome is divided into 100-nt-long regions, the number of mapped sequence fragments for each window is registered for each library and a correlation coefficient is computed. The correlation analysis is applied separately to single- and chimeric-sequence fragments. We noticed that the correlation between the mapped single fragments is higher than the correlation between the chimeric fragments[11]. To be considered reproducible, the correlation of replicates should be >0.4.

1. Generate the files with all chimeric fragments (including nonsignificant ones)

```
foreach fragment_file (*_all_fragments.txt)
foreach? RILseq_significant_regions.py --bc_dir <ecocyc-data> --BC_chrlist  chr,COLI-K12
--ribozero --min_odds_ratio 1 --all_interactions $fragment_file  >! {$fragment_file}_all_
interactions.txt &
foreach? end
```

At the end of this step, you will have an output filewith the suffix `_all_interactions.txt` for each individual library file that lists all chimeric fragments, including the nonsignificant ones. Note that the information in the output file (i.e., the number and numbering of the presented columns) depends on the selected run parameters.

2. Generate the 'single' files

```
foreach fragment_file (*_all_fragments.txt)
foreach? RILseq_significant_regions.py --bc_dir <ecocyc-data> --BC_chrlist chr,COLI-K12
--ribozero --min_odds_ratio 1 --all_interactions --only_singles $fragment_file >! {$fragment_
file}_singles.txt &
foreach? end
```

▲ **CRITICAL STEP** If you do not have an annotation database, exclude `--bc_dir <ecocyc-data>` and `--BC_chrlist chr,COLI-K12` from the two above commands. If the run is not on *E.coli* K-12, you need to use the `--BC_chrlist` parameter with the respective chromosome/plasmid names in the BAM and BioCyc files. Excluding the annotation directory will not allow the use of `--ribozero` (**Box 4**)

▲ **CRITICAL STEP** You should run the RILseq_significant_regions.py (steps 1 and 2 above) with the exact same parameters used in Step 132. In particular, if you used `--ribozero` in Step 132, apply this parameter here as well.

▲ **CRITICAL STEP** The '&' at the end of the command lines should only be applied with multicore-processor or multiple-processor platforms.

At the end of this step, you will have an output file with the suffix `_singles.txt` for each individual library file, listing single fragments. Note that the information in the output file (i.e. the number and numbering of the presented columns) depends on the selected run parameters.

NOTE: Although the output file lists single fragments, its format is identical to the chimera output file listing both RNA1 and RNA2. When annotation information is available, typically RNA1 and RNA2 would have identical annotations. Occasionally, however, when reads span multiple annotations (e.g., CDS and 3′ UTR), the names might differ and include adjacent genomic entities.

3. Generate the input file for the correlation analysis

Generate a space-delimited file `<file-for-correlation>` in which the columns are the name of the library, the `all_interactions.txt` file name and the `_singles.txt` file name. For simplicity of representation, it is recommended to use short names in the first column. Note that *hfq-WT* libraries will typically have too few interactions for this analysis and therefore it is highly recommended not to include them in the table, as they will generate run errors.

Examples:

Example for `<file-for-correlation>`, the input file to the script plot_regions_interactions.py for the correlation analysis. The file is tab-delimited and contains a separate line for each individual library `<file-for-correlation>`:

```
libA-rep1       libraryA-replicate1_all_interactions.txt    libraryA-replicate1_single.txt
libB-rep2       libraryB-replicate2_all_interactions.txt    libraryB-replicate2_single.txt
```

4. Run the script:

```
plot_regions_interactions.py <file-for-correlation> <output_name>
```

At the end of this step, you will have a figure file with the suffix `_scatter.tif` showing the scatter plots that compare the sequenced fragments between two pairs of libraries, as well as the correlation coefficients. An additional heatmap file, with the suffix `_heatmap.tif`, is also generated. Details regarding these output figures are in the supplementary material of Melamed *et al.*[11].

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

It is highly recommended to test the reproducibility of the library replicates (**Box 6**). Low reproducibility might suggest that specific replicates should be excluded. Libraries showing reproducible results can be unified. Unified data can increase the detection power of poorly expressed interacting RNAs.

131| Generate a unified file of the results of replicate experiments.

Assume that we have three replicates with acceptable reproducibility (library_X1, library_X2, library_X3):

```
cat <libraryX1_all_fragments.txt> <libraryX2_all_fragments.txt> <libraryX3_all_
fragments.txt> > <unified-X1-X2-X3_cond1_all_fragments.txt>
```

▲ **CRITICAL STEP** Unification can be applied only to libraries that were found to be reproducible in Step 130.
▲ **CRITICAL STEP** Make sure that the unified file name has the `_all_fragments.txt` suffix or the hereinafter foreach command will not be applied to the files.

132| Apply RILseq_significant_regions.py to the files generated in Step 129 and to the unified file from Step 131.

If you do not have an annotation database, exclude `--bc_dir <ecocyc-data>` and `--BC_chrlist chr,COLI-K12` from the command below. Excluding the annotation directory will not allow the use of `--ribozero` (**Box 4**).

```
  foreach fragment_file (*_all_fragments.txt)

  foreach? RILseq_significant_regions.py --bc_dir <ecocyc-data> --BC_chrlist
chr,COLI-K12 --ribozero --min_odds_ratio 1 $fragment_file >! {$fragment_
file}_significant_interactions.txt &

 foreach? end
```

At the end of this step, you should have a tab-delimited file with the suffix `_significant_interactions.txt` for each of your input files, which lists the statistically significant chimeric fragments (S-chimeras, example in **Supplementary Table 1**).
▲ **CRITICAL STEP** The run parameters listed in the above RILseq_significant_regions.py command line are compatible with the sequencing data generated as described in this protocol. Tips and comments on parameter selection are described in **Box 4**. Note that the information in the output file (i.e., the number and numbering of the presented columns) depends on the selected run parameters.

133| Extract the final set of putative RNA–RNA pairs. The number of reported S-chimeras depends to a large extent on the minimal number of sequenced fragments (MIN_INT) required to declare an S-chimera. Usually, deeper sequencing results in a higher number of reported S-chimeras, which may require further filtering. We noticed that the number of S-chimeras detected for an *hfq-WT* library can be used for rough estimation of a more stringent additional threshold, above which a chimera will be determined as an S-chimera. It is recommended to set this threshold to a value (predetermined limit (L)) that determines the number of reported S-chimeras that should not be exceeded in the *hfq-WT* individual libraries. Thus, the additional threshold can be set to the number of sequenced fragments that guarantees that L will not be exceeded in the *hfq-WT* library.
▲ **CRITICAL STEP** The same threshold should be set to all the libraries sequenced in the same batch and preferably to all the libraries in a RIL-seq run.
▲ **CRITICAL STEP** The ‘&’ at the end of the command line should be applied only with multicore-processor or multiple-processor platforms.

134| Count the number of S-chimeras in a file

```
wc -l *significant_interactions.txt
```

This will calculate the number of lines in each file, including the header line. Thus, you need to subtract 1 to get the actual number of interactions.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
Box 1	No band of ~70 kDa in the "experimental sample" and/or a strong band of ~70 kDa in the unbound fraction of the *hfq-Flag* sample	The anti-FLAG antibody or the protein A/G magnetic beads are inactive	Try another batch of anti-FLAG antibody and/or protein A/G magnetic beads
		The buffer conditions were not suitable	Prepare new wash buffer. Aspirate the IGEPAL accurately
49	There is no blue pellet	The RNA precipitation did not work properly	Add an additional 1.5 µl of GlycoBlue, incubate the tube at –80 °C for 1 h and centrifuge again. If a blue pellet forms, continue to the next step
52	The total RNA profile does not look as shown in **Figure 4a**	RNA degradation has occurred during the lysate preparation or the extraction process	Wear gloves and change them frequently. Work only with RNase-free reagents, tips and tubes. Replace your reagents, as they might be contaminated with RNases. Use cold solutions, tubes and tube holders for Retch mixer mill. When possible, keep the bacteria on ice before and during lysate preparation
	The Bioanalyzer profile of the RNA from the *hfq-Flag* sample does not show any RNA peak	The RNA pellet was lost during ethanol washes at Step 50	Make sure that you see a blue pellet at Step 49 and during the ethanol washes at Step 50
	The Bioanalyzer profile of the RNA from the *hfq-Flag* sample does not look as shown in **Figure 4b** but appears as in **Figure 4c**	RNA degradation has occurred during the lysate preparation or the extraction process	Make sure you work on ice when needed, wear gloves and change them frequently, and work only with RNase-free reagents, tips and tubes. To find the step at which RNA degradation occurred, perform a diagnostic experiment: start the RIL-seq protocol with four *hfq-Flag* samples, and at the end of each part (end of Steps 28, 32, 35 and 41), stop one of the samples by proteinase K treatment and RNA extraction as described in Steps 42–51, while continuing with the rest of the samples to the next steps. Analyze the resulting RNA samples using a Bioanalyzer RNA 6,000 Pico kit. If RNA degradation is observed in one of the steps, it might be caused by RNase contamination of one of the reagents used. Replace all the reagents used in this step. If degradation is observed after the RNase treatment (Step 32), it might result from an RNase concentration that is higher than needed, or overincubation with the RNases. To avoid such a problem, make the 1/20 dilution of RNases accurately, touching the RNases solution only with the tip end, in order not to cover the external part of the tip with the viscous RNases solution. Perform the treatment of RNases for exactly 7 minutes for each sample
	The profile of the RNA from the *hfq-WT* sample is similar to the profile of RNA from the *hfq-Flag* sample (example in **Figure 4b**)	Unknown	This phenomenon may sometimes occur. Continue with the procedure. A typical *hfq-WT* sample will contain a percentage of chimeras that is <1% of the total chimeras[11]

(continued)

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

TABLE 2 | Troubleshooting table (continued).

Step	Problem	Possible reason	Solution
114, 115	Qubit analysis shows that the concentration of all three samples is similar and does not increase with the PCR cycle number. TapeStation analysis shows a very small peak of the desired size for all three samples	The cDNA yield at the end of Step 107 was very low	Repeat the PCR-enrichment test, increasing the number of cycles
115	The TapeStation analysis shows only DNA peaks that are <185 bp. See example in **Figure 5b**	RNA degradation during the library construction	Repeat the experiment and take 1-µl samples at the end of Steps 63 and 83 in the course of library construction. Check these samples using a Bioanalyzer RNA 6000 Pico kit in order to verify that the RNA was not degraded
115, 124	The TapeStation analysis shows peaks of ~70 and ~130 bp, in addition to the desired >185-bp peak (example in **Figure 5c**)	Primers and primer dimers were not removed during the library cleanup (Steps 110–113 or 120–123)	Repeat the cleanup once or twice, using AMPure XP beads at a ratio of 1×, and verify by TapeStation analysis that the undesired peaks disappear. If the ~70- and ~130-bp peaks still appear, repeat the cleanup with AMPure XP beads using a ratio of 0.85×. If successful in the cleanup of the small fragments, repeat the Qubit measurement, as the cleanup process results in a decreased concentration
Box 2	A large fraction of Read1 and Read2 pair mates are mapped to the forward and reverse strand of the same genomic location (you can use SAMtools to view the *.bam reads)	'map_single_fragments.py' was run with two identical input files	Verify that '$file' is the first mate file '*_1.fastq' and '$second_file' is the second mate file '*_2.fastq', and re-run map_single_fragments.py
	A low percentage of total mapped fragments (<50% of total fragments) in the initial mapping	There is a high fraction of chimeric fragments that cannot be detected at this stage	Continue to Step 129. If the fraction of both mapped single and chimeric fragments is now at least 50%, then continue to the next step. Otherwise see other possible solutions below
		BAM files are incompatible with the reference genome	Check that the proper reference genome was used
		The RNA was degraded at one of the steps	Run the FastQC and check overrepresented adaptors or reads enriched with poly-Ns and/or poly-Gs. If a large fraction of the reads contains such sequences, see the solution recommendations for Step 52
129	The number of chimeric fragments is very high, whereas the number of single fragments is very low	The input BAM file was generated with two identical paired-end input files	See troubleshooting comments for **Box 2**
	A negligible number of chimeric fragments were found in the *hfq-Flag* sample	A problem in the mapping or in the depth of the sequencing	Check the percentage of mapped fragments; it should be >50% of the processed sequenced fragments. If it is less, see troubleshooting comments for **Box 2**. Based on our experience with *E.coli* MG1655, you need approximately 10 million reads per *hfq-Flag* library to get a sufficient number of mapped chimeric fragments (in the hundred thousands)
			Consider repeating the sequencing in order to obtain deeper coverage
		Inefficient ligation at Steps 36–38	Consider repeating the experiment, using a new ligation kit

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

135| If the number of S chimeras in each *hfq-WT* library ≤ L, filtering is not required; you have the final set of S-chimeras. If the number in at least one *hfq-WT* library > L, select the *hfq-WT* library file with the maximum number of S-chimeras and sort the file by the table column entitled 'interactions' from the highest to the lowest number. Set the new threshold for determining a chimeric fragment as an S-chimera as the number in the 'interactions' fields above which you have only L S-chimeras.

To apply this threshold to the *hfq-Flag* libraries, sort each library by the 'interactions' field, as described above, and exclude all rows in which the number of interactions is lower than the new threshold.
▲ **CRITICAL STEP** Following the results for *hfq-WT* libraries in Melamed *et al.*[11], we suggest setting L to 50.

? TROUBLESHOOTING
Troubleshooting advice can be found in **Table 2**.

● **TIMING**
Steps 1–9, cross-linking and bacteria freezing: 24 h
Steps 10–21, lysate preparation: 1–2 h
Step 22, lysate preclearing: 1.5 h
Steps 23–28, co-immunoprecipitation using anti-Flag antibody: 3 h
Steps 29–32, trimming of RNA ends: 40 min
Steps 33–35, 5′OH end phosphorylation and 2′P/3′P end dephosphorylation: 2.5 h
Steps 36–38, ligation of neighboring RNAs: overnight
Steps 39–43, protein digestion: 2.5 h
Steps 44–53, RNA extraction according to the standard TriReagent LS protocol: 14 h
Steps 54–56, fragmentation and DNase–FastAP combined treatment: 45 min
Steps 57 and 58, RNA cleanup using RNA Clean & Concentrator-5 kit: 1 h
Steps 59–61, ligation of a 3′ adaptor (RNA/DNA): 2.5 h
Steps 62 and 63, pooling the libraries and RNA cleanup using RNA Clean & Concentrator-5 kit: 30 min
Steps 64–76, rRNA removal using RiboZero kit: 45 min
Steps 77–83, rRNA-depleted RNA purification: 45 min
Steps 84–87, first-strand cDNA synthesis using the SuperScript III First-Strand Synthesis System: 1.25 h
Steps 88 and 89, RNA degradation: 20 min
Steps 90–95, cDNA cleanup: 45 min
Steps 96–98, ligation of a second adaptor at the cDNA 3′ end (ssDNA/ssDNA): overnight
Steps 99–102, cleanup of cDNA: 45 min
Steps 103–106, 2nd cleanup of cDNA to remove the remaining adaptors: 45 min
Steps 107–109, PCR enrichment test to determine the final number of PCR cycles: 45 min
Steps 110–116, PCR cleanup: 1.5 h
Steps 117–119, PCR amplification of the libraries: 1 h
Steps 120–124, library cleanup: 1.5 h
Step 125, sequencing of libraries: 12 h
Steps 126–128, preliminary steps for the application of RIL-seq software: variable
Step 129, identification of chimeric fragments: variable
Steps 130–135, identification of statistically significant chimeras (S-chimeras): variable
Box 1, verification of Hfq-Flag IP: 3–4 d
Box 2, generation of the *.bam and *.bam.bai files: variable
Box 5, software setup for bioinformatics analysis: variable
Box 6, evaluation of replicate reproducibility: variable

ANTICIPATED RESULTS

Step 52

This step provides a profile of the isolated RNA that will be used to construct RIL-seq cDNA libraries. **Figure 4b** shows a typical Bioanalyzer profile of the RNA obtained from an *hfq-Flag* strain. Note that the size of the RNA molecules ranges between 25 and 2,000 nt. An RNA-size curve in which all the RNA molecules are smaller than 200 nt indicates that the sample was degraded, as exemplified in **Figure 4c**. The degraded RNA cannot be further used for library construction, and the troubleshooting advice for Step 52 can be helpful in determining the cause of the degradation. **Figure 4d** shows a typical

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

Bioanalyzer analysis of the RNA obtained from the *hfq-WT* strain. Note that the RNA size distribution in the *hfq-WT* sample is similar to that in the *hfq-Flag* sample (**Fig. 4b**), but the height of the peak and the concentration of the RNA are much lower (height: 2 vs. 160 fluorescence units (FU), concentration: 54 vs. 4,242 pg/μl of RNA from *hfq-WT* vs. *hfq-Flag*, respectively). In some RIL-seq experiments, however, the profile of RNA from the *hfq-WT* strain is very similar to the profiles of RNA from an *hfq-Flag* strain (as in **Fig. 4b**). The reason is unknown, but the RNA sample can be further used for library construction. **Figure 4a** shows the Bioanalyzer analysis of total RNA from the *hfq-Flag* strain, extracted in Step 26 from a lysate not subjected to the IP step. The typical ribosomal RNA peaks of 5S, 16S and 23S are the main peaks, with a slight background of molecules of other sizes, indicating that a large fraction of the RNA is kept intact before performing the IP.

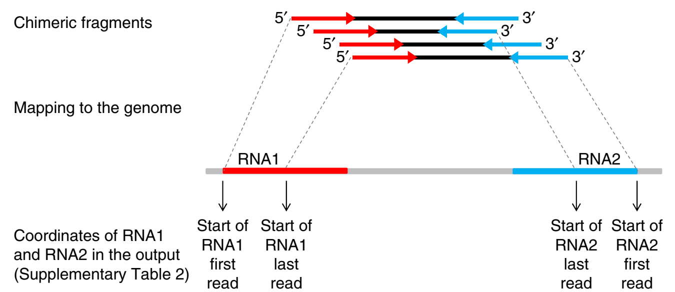

Figure 6 | Schematic representation of the S-chimera coordinates reported in the output file (**Supplementary Table 1**). A reported S-chimera is typically derived from an ensemble of chimeric fragments, depicted in the figure. The two reads (pair mates) of each chimeric fragment are shown in red and blue, corresponding to RNA1 and RNA2, respectively. Arrows indicate the sequencing direction in the paired-end sequencing. The reported coordinates designated for RNA1 are the first nucleotide of the first (most 5′) read and the first nucleotide of the last (most 3′) read. The reported coordinates designated for RNA2 are the first nucleotide of the last (most 5′) and first nucleotide of the first (most 3′) read. For simplicity the genome is represented by a single line, with the two transcripts spanning from left to right (red) and from right to left (blue)

Step 58

This step provides a profile of the fragmented RNAs. **Figure 4e**,**f** show typical profiles of fragmented RNA from *hfq-Flag* and *hfq-WT*, respectively. As expected, we do see a lower concentration in the *hfq-WT* sample as compared with that in the *hfq-Flag* sample.

Steps 115 and 124

This step can help in evaluating the success of the library construction and the size distribution of the DNA molecules. **Figure 5a** shows a TapeStation electropherogram of RIL-seq libraries. Note that the peak's size is ~220 bp. The minimal library fragment size that is required is 185 bp (the insert size is ~50 bp), but often smaller fragments appear and must be removed by an additional AMPure XP cleanup, as shown in **Figure 5b,c**. A complete absence of the expected library fragment peak (**Fig. 5b**) indicates that the RNA was degraded. See troubleshooting advice for Step 115.

Steps 126–135

Assessment of the RIL-seq data can be obtained by computing the number of expected sequenced fragments at various stages of the computational analysis. **Supplementary Table 2** displays this information for an example run. The analysis was done using *E. coli* K-12 MG1655 genome version NC_000913.3. The full scope of input and output files for this example run is available at http://margalit.huji.ac.il/RILseq/index.php.

Step 129

The total number of mapped fragments is the sum of the number of fragments that were mapped as single fragments and the number of fragments that were mapped as chimeric fragments. In the RIL-seq data from Melamed *et al.*[11], the percentage of mapped fragments out of the total number of processed fragments (i.e., following the application of cutadapt, dust and polyG removal procedures) ranged between 50 and 70%. The percentage of ribosomal RNA out of the total mapped fragments in the *hfq-Flag* libraries was less than 2%. The percentage of the mapped chimeras out of the total mapped fragments ranged in the *hfq-Flag* libraries between 4 and 12%, and it was usually less than 1% in the *hfq-WT* libraries[11].

Step 135

The S-chimeras are reported in a table generated by the RILseq_significant_regions.py script. Each row in the table represents an S-chimera, composed of an ensemble of chimeric fragments linking two RNA transcripts termed RNA1 and RNA2 (**Fig. 6**). A pair of RNAs might appear more than once if it involves multiple interacting regions or if it appears in the chimera once as RNA1–RNA2 and once as RNA2–RNA1. The core table (highlighted in gray in **Supplementary Table 1**) includes the genomic position of each RNA and the Fisher's exact test related information. Computed features that can be used for heuristic ranking of the targets of a specific sRNA are displayed if the corresponding run parameters were applied (highlighted in light orange in **Supplementary Table 1** and detailed in **Box 4**). If RILseq_signifnicant_regions.py was run with BioCyc annotation information (highlighted in green in **Supplementary Table 1**), the S-chimeras are annotated as follows:

- BioCyc ID/ Gene name—the RNA is within the CDS.
- 5UTR/EST5UTR/3UTR/EST3UTR—the RNA is mapped to a region outside a CDS. The name is followed by '5UTR' or '3UTR' if it resided in a BioCyc-annotated UTR, or by 'EST5UTR' or 'EST3UTR' if the UTR is unknown and the RNA mapped to a region of

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

100 nt upstream or downstream of the CDS, respectively (or a shorter distance, if these regions spanned another transcript or were more likely to be a UTR of the neighboring transcript).

- IGT and IGR preceded by two gene names—the RNA is mapped to the region between the two genes on the same transcription unit (IGT) or between two different transcription units (IGR). The IGR region is defined in a strand-independent manner (i.e., closest two genes). Thus, fragments mapped to an IGR region will have the same name even if the orientation defined by their mapping differs. This fragment orientation, however, can be retrieved from the strand column.
- AS and AS_IGT—the RNA is mapped to the antisense of a gene or of an IGT region, respectively.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS This study was supported by European Research Council Advanced Grant 322920, the I-CORE Programs of the Planning and Budgeting Committee and The Israel Science Foundation (grants 1796/12 and 41/1), and the Israel Science Foundation, administered by the Israeli Academy for Sciences and Humanities (grant 1411/13). We thank Y. Gatt, S. Pearl-Mizrahi, A. Zhang and H. Zhang for their useful comments.

AUTHOR CONTRIBUTIONS H.M. guided the study. S.M. and A.P. developed the methodology; S.M., R.F.-R. and L.A. carried out the experiments.; A.P., Y.A. and R.F.-R. curated the data; A.P. developed the software; A.P., Y.A., R.F.-R., N.R. and A.B. carried out the computational analyses; N.R., O.S. and A.B. updated and tested the software. S.M., A.P., Y.A., L.A., R.F.-R. and H.M. wrote the paper.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Wagner, E.G. & Romby, P. Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv. Genet.* **90**, 133–208 (2015).
2. Vogel, J. & Luisi, B.F. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 (2011).
3. Olejniczak, M. & Storz, G. ProQ/FinO-domain proteins: another ubiquitous family of RNA matchmakers? *Mol. Microbiol.* **104**, 905–915 (2017).
4. Zhang, A. *et al.* Global analysis of small RNA and mRNA targets of Hfq. *Mol. Microbiol.* **50**, 1111–1124 (2003).
5. Sittka, A. *et al.* Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet.* **4**, e1000163 (2008).
6. Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M. & Vogel, J. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).
7. Bilusic, I., Popitsch, N., Rescheneder, P., Schroeder, R. & Lybecker, M. Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. *RNA Biol.* **11**, 641–654 (2014).
8. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).
9. Grosswendt, S. *et al.* Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. *Mol. Cell* **54**, 1042–1054 (2014).
10. Moore, M.J. *et al.* miRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. *Nat. Commun.* **6**, 8864 (2015).
11. Melamed, S. *et al.* Global mapping of small RNA-target interactions in bacteria. *Mol. Cell* **63**, 884–897 (2016).
12. Morita, T., Maki, K. & Aiba, H. RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.* **19**, 2176–2186 (2005).
13. Schiano, C.A. & Lathem, W.W. Post-transcriptional regulation of gene expression in *Yersinia* species. *Front. Cell Infect. Microbiol.* **2**, 129 (2012).
14. Lioliou, E. *et al. In vivo* mapping of RNA-RNA interactions in *Staphylococcus aureus* using the endoribonuclease III. *Methods* **63**, 135–143 (2013).
15. Waters, S.A. *et al.* Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *EMBO J.* **36**, 374–387 (2017).
16. Lalaouna, D. *et al.* 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 (2015).
17. Han, K., Tjaden, B. & Lory, S. GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by *in vivo* proximity ligation. *Nat. Microbiol.* **2**, 16239 (2016).
18. Uzzau, S., Figueroa-Bossi, N., Rubino, S. & Bossi, L. Epitope tagging of chromosomal genes in *Salmonella*. *Proc. Natl. Acad. Sci. USA* **98**, 15264–15269 (2001).
19. Shishkin, A.A *et al.* Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat. Methods* **12**, 323–325 (2015).
20. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv e-prints* **1303**, 3997 (2013).
21. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
22. Muckstein, U. *et al.* Thermodynamics of RNA-RNA binding. *Bioinformatics* **22**, 1177–1182 (2006).
23. Keseler, I.m. *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**, D605–612 (2013).
24. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **40**, D742–D753 (2012).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.