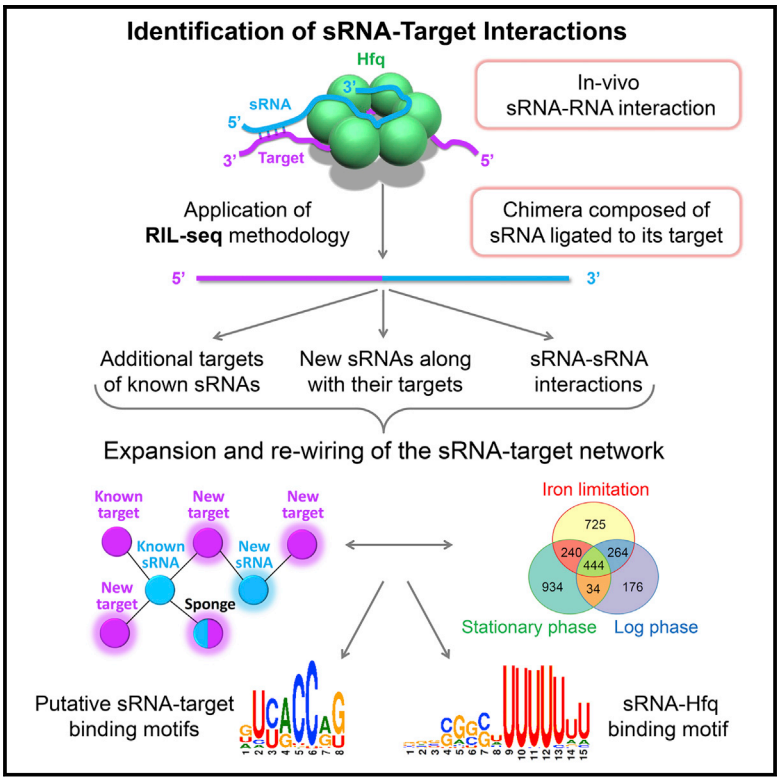
Resource

Molecular Cell

Global Mapping of Small RNA-Target Interactions in Bacteria

Graphical Abstract



Highlights

- A widely applicable method for in vivo global mapping of small RNA interactome
- Substantial re-wiring of the network upon changes in cellular conditions
- Regulatory circuits involving two regulators derived from the same transcript
- sRNAs acting in *trans* are encoded within almost every possible genomic element

Authors

Sahar Melamed, Asaf Peer, Raya Faigenbaum-Romm, ..., Yael Altuvia, Liron Argaman, Hanah Margalit

Correspondence

yaelal@md.huji.ac.il (Y.A.), liron.argaman@mail.huji.ac.il (L.A.), hanahm@ekmd.huji.ac.il (H.M.)

In Brief

Melamed et al. describe RIL-seq, an approach that can identify Hfq-bound pairs of small RNAs (sRNAs) and their targets. They apply RIL-seq to *E. coli* grown in different conditions, identifying new sRNAs and describing sRNA interactome changes upon change in conditions.

Accession Numbers

E-MTAB-3910

Melamed et al., 2016, Molecular Cell *63*, 884–897
September 1, 2016
http://dx.doi.org/10.1016/j.molcel.2016.07.026

CellPress

Molecular Cell
Resource

Global Mapping of Small RNA-Target Interactions in Bacteria

Sahar Melamed,[1,2,3] Asaf Peer,[1,3] Raya Faigenbaum-Romm,[1,3] Yair E. Gatt,[1] Niv Reiss,[1] Amir Bar,[1] Yael Altuvia,[1,*] Liron Argaman,[1,*] and Hanah Margalit[1,4,*]

[1]Department of Microbiology and Molecular Genetics, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem 9112102, Israel

[2]Present address: Division of Molecular and Cellular Biology, NICHD, National Institutes of Health, 18 Library Dr MSC 5430, Bethesda, MD 20892-5430, USA

[3]Co-first author

[4]Lead Contact

*Correspondence: yaelal@md.huji.ac.il (Y.A.), liron.argaman@mail.huji.ac.il (L.A.), hanahm@ekmd.huji.ac.il (H.M.)

http://dx.doi.org/10.1016/j.molcel.2016.07.026

SUMMARY

Small RNAs (sRNAs) associated with the RNA chaperon protein Hfq are key posttranscriptional regulators of gene expression in bacteria. Deciphering the sRNA-target interactome is an essential step toward understanding the roles of sRNAs in the cellular networks. We developed a broadly applicable methodology termed RIL-seq (RNA interaction by ligation and sequencing), which integrates experimental and computational tools for in vivo transcriptome-wide identification of interactions involving Hfq-associated sRNAs. By applying this methodology to *Escherichia coli* we discovered an extensive network of interactions involving RNA pairs showing sequence complementarity. We expand the ensemble of targets for known sRNAs, uncover additional Hfq-bound sRNAs encoded in various genomic regions along with their *trans* encoded targets, and provide insights into binding and possible cycling of RNAs on Hfq. Comparison of the sRNA interactome under various conditions has revealed changes in the sRNA repertoire as well as substantial re-wiring of the network between conditions.

INTRODUCTION

Posttranscriptional control of gene expression by small RNAs (sRNAs) is a major layer of regulation in bacteria. Most known sRNAs exert their regulatory function by base-pairing with their target RNAs, usually affecting the target translation, stability, and/or processing (Storz et al., 2011; Wagner and Romby, 2015). Base-pairing sRNAs are commonly divided into *cis*- and *trans*-encoded. While the former are transcribed antisense to their targets, the *trans* sRNAs and their targets are encoded in different genomic locations, usually exhibiting partial sequence complementarity. Base pairing between *trans*-encoded sRNAs and their targets is often mediated by the RNA chaperon protein Hfq, a ring-shaped hexamer that is abundant in Gram-negative bacteria (Wagner and Romby, 2015). A major challenge is to identify the complete network of Hfq-mediated sRNA-target interactions and to study its structure and dynamics. Yet, no methodology for large-scale direct capturing of these interactions has been published.

Previous large-scale approaches identified RNA transcripts bound on Hfq by co-immunoprecipitation (coIP) followed by microarray (Zhang et al., 2003) or RNA-seq analysis (Bilusic et al., 2014; Chao et al., 2012; Sittka et al., 2008) and were further revised by including Hfq-RNA UV crosslinking (Holmqvist et al., 2016; Tree et al., 2014). These experiments provided a global view of the RNA molecules bound to Hfq, but specific sRNA-target pairs could only be indirectly deduced by additional sequence-dependent predictive schemes.

To overcome this limitation, and inspired by the CLASH and iPAR-CLIP methodologies that were applied to detect protein-mediated miRNA-target interactions in eukaryotes (Grosswendt et al., 2014; Helwak et al., 2013), we developed RIL-seq (RNA interaction by ligation and sequencing), an experimental-computational methodology for detecting Hfq-bound sRNA-target pairs in bacteria. We applied RIL-seq to *Escherichia coli* grown under three different growth conditions and identified ~2,800 Hfq-bound RNA pairs, dominated by mRNA-sRNA pairs exhibiting sequence complementarity. Our re-discovery of known sRNA targets, as well as the computational and experimental assessments that we provide, strongly support the reliability of RIL-seq data and the applicability of this approach for capturing the sRNA interaction network. The network we uncovered includes additional targets of established sRNAs as well as additional sRNAs transcribed from various genomic regions, along with their targets. Comparison of the interactions discovered under different conditions revealed substantial re-wiring of the network, implying the role of sRNAs in regulation of global cellular processes.

RESULTS

RIL-seq Protocol Optimization

The major steps of RIL-seq are described in Figure 1 and the Experimental Procedures. The experiments were performed in *E. coli* carrying a functional tagged Hfq (*hfq-Flag*) (Morita et al.,

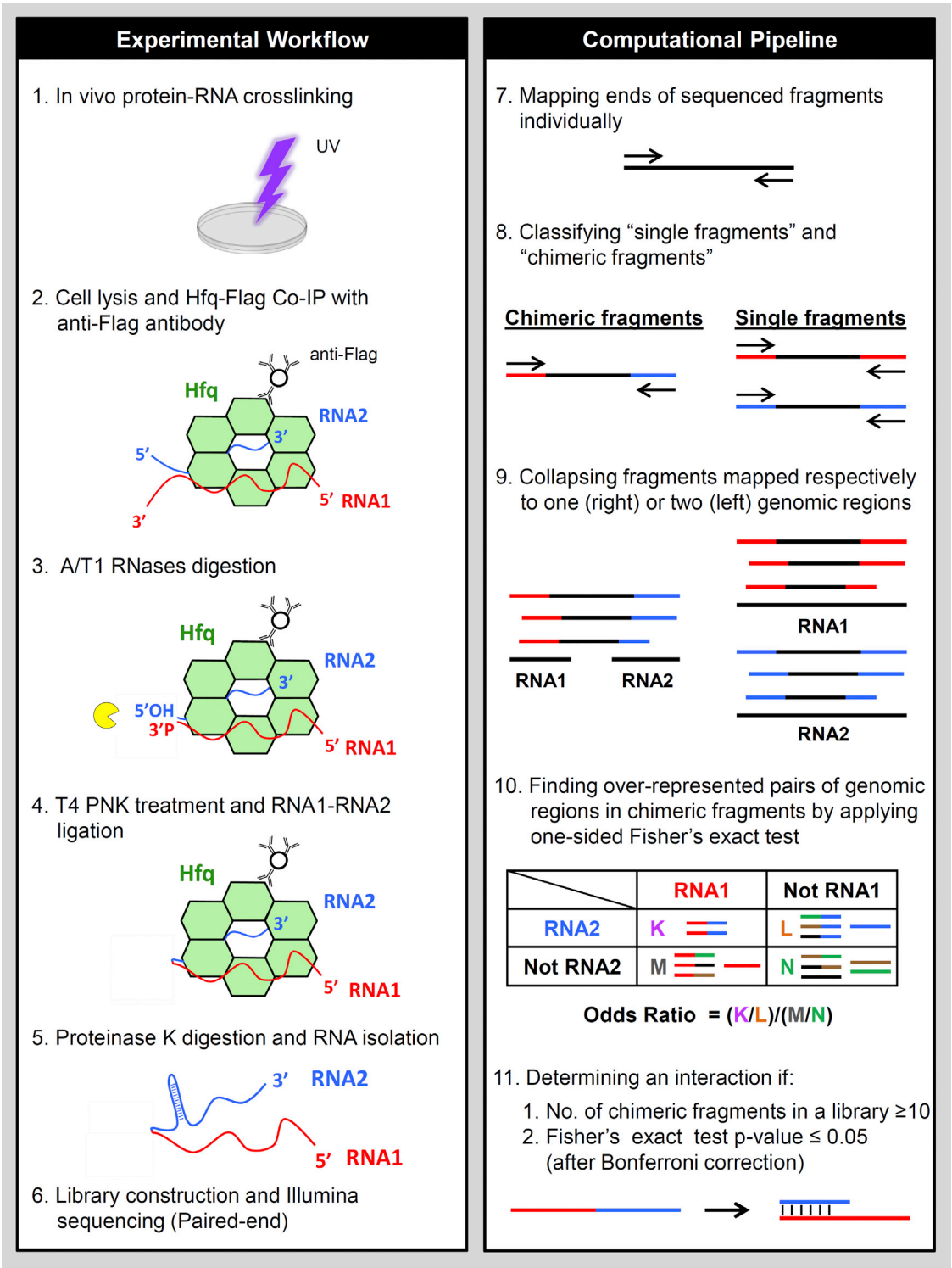


Figure 1. Overview of RIL-seq Experimental and Computational Procedures
See also Figures S1 and S2 and Table S1.

2005) and involved culturing of bacteria in desired conditions, in vivo protein-RNA crosslinking, cell lysis, coIP of Hfq and bound RNAs (Figure S1), RNA ligation, RNA isolation, paired-end sequencing, and mapping of the sequenced fragments to *E. coli* genome. Each sequenced fragment was represented by two sequences at its ends (pair mates), which were mapped

individually to the genome. This mapping revealed two types of fragments: (1) "single" fragments, where the pair mates were mapped to the same genomic location, most likely originating from a single transcript and (2) "chimeric" fragments, where the pair mates were mapped to two distinct genomic locations, supporting a putative interaction between RNAs derived from these regions. To avoid spurious interactions (e.g., resulting from the ligation of two random adjacent RNAs), we selected RNA pairs that were statistically significantly over-represented in chimeric fragments. We assessed the statistical significance by Fisher's exact test with Bonferroni correction and computed the odds ratio (considering the numbers of chimeric fragments involving the two RNAs, other fragments they are involved in, and all other fragments) (Figure 1). Only statistically significant RNA pairs supported by at least 10 chimeric fragments were regarded as putatively interacting (hereinafter, S-chimeras). While the odds ratio corresponds to the over-representation of the pair above random expectation, it does not convey the functional impact of the interaction on the target. To estimate the latter, we computed an additional measure based on the assumption that the higher the enrichment of the RNAs on Hfq, the larger the effect the interaction may have. This measure, termed normalized odds ratio, considers both the odds ratio and the enrichment of the two interacting RNAs on Hfq (Supplemental Experimental Procedures).

We tested various versions of the protocol in log-phase cells (Supplemental Experimental Procedures), six of which yielded consistent sequencing results for both the single and chimeric fragments (Table S1 and Figures S2A–S2D) and hence were considered repetitions of the same experiment. In all successive experiments, including three in stationary phase and three under iron limitation, one protocol (highlighted in Table S1) was used.

We assessed the quality of RIL-seq protocol in several ways (Supplemental Experimental Procedures). (1) By applying RIL-seq to a wild-type (WT) *E. coli* strain, where Hfq was not tagged (*hfq-WT*), we demonstrated that in all three conditions the counts of chimeric fragments as well as fragments supporting S-chimeras were substantially smaller in *hfq-WT* compared to *hfq-Flag* libraries (Table S1). (2) By mixing the *E. coli* lysate with a *Saccharomyces cerevisiae* lysate before the coIP and applying the protocol to the mixture, we verified that the vast majority of RIL-seq interactions originated in vivo (Table S1). (3) By computational analysis of short chimeric fragments where the ligation point could be located, we verified that potential biases stemming from possible tendency of the ligase to fuse specific sequences or structures were minimal (Figures S2E–S2G). Still, there might be some bias that we did not identify, mainly because only 0.02% of the sequenced chimeric fragments were short enough to enable the above analysis.

A major difference between RIL-seq and iPAR-CLIP or CLASH protocols is that RIL-seq does not employ stringent conditions to remove RNAs that were not crosslinked to the protein after the coIP, in order to ensure the Hfq hexamer integrity. Therefore, especially as the crosslinking efficiency is estimated at 1%–5% (Darnell, 2010), it is conceivable that our libraries contain RNAs that were and were not crosslinked to Hfq, questioning the necessity of crosslinking. Thus, we re-applied the protocol without crosslinking (Table S1). Since the experiment with mixed *E. coli* and yeast lysates yielded a higher fraction of in vitro interactions in the protocol without crosslinking (3.2%–3.8%) compared to the one with crosslinking (<1%) (Table S1), we concluded that the crosslinking allows better capturing of in vivo Hfq-mediated RNA interactions. Hence, all the results described below are based on the protocol with crosslinking.

RIL-seq Results Are Reproducible

We evaluated the reproducibility of the results within same-condition libraries (log phase, stationary phase, and iron limitation) (Supplemental Experimental Procedures). For both single and chimeric fragments, there was correlation in the numbers of mapped sequenced fragments in corresponding genomic windows between same-condition libraries (Figures S2A–S2C). In addition, about half of the S-chimeras were detected in at least two same-condition libraries (Figure S2D), compared to only one chimera in the *hfq-WT* libraries. Given the reproducibility of the results and to gain better statistical power, we assembled the mapped single and chimeric sequenced fragments of all libraries in each condition to a unified dataset, resulting in three datasets, to each of which we applied the statistical test. This has allowed many interactions that did not pass the threshold of 10 sequenced chimeric fragments or the statistical test in the individual libraries to be revealed, while only a relatively small number of interactions that passed these criteria in the individual libraries were lost. All further analyses were applied to the interactions in the unified datasets (Table S2).

Corroborating the Interactions Revealed by RIL-seq

RIL-seq Data Are Enriched with sRNA-mRNA Pairs

Classification of all mapped fragments by their genomic annotation revealed high fractions of single and chimeric fragments mapped to sRNAs in the *hfq-Flag* libraries, which were substantially smaller in the *hfq-WT* control libraries (Figures 2A, S3A, and S3B). Furthermore, the distribution of genomic element pairs in S-chimeras was dominated by sRNAs paired with RNAs derived from CDS and/or 5′ UTR regions (Figures 2B, S3C, and S3D).

We compiled a set of 154 experimentally verified Hfq-mediated sRNA-target interactions (hereinafter "known" interactions), involving 26 sRNAs (Table S3), out of which 25 were revealed by RIL-seq (Table S2). Compatible with the transcriptome study of Thomason et al. (2015), we detected most of the 25 sRNAs in all three conditions (Figure S3E), although their relative fractions could differ between conditions, consistent with the observations of Chao et al. (2012). Out of the 154 known interactions, 117 were previously determined under the growth conditions studied here, and we discovered 63 of them. While the remaining 37 interactions were originally determined under other specific stress conditions, we detected 23 of them, resulting in the identification of a total of 86 known interactions (Figure 3A and Table S3). Examples of re-discovered known interactions are illustrated in Figures 3B–3D for *ompA*-MicA interaction (Udekwu et al., 2005) and *lpp*-MicL interaction (Guo et al., 2014).

Objective Re-discovery of Known Binding Sites Supports the Interacting Partners

Since a given sRNA often binds different targets through the same binding site, objective identification of a common

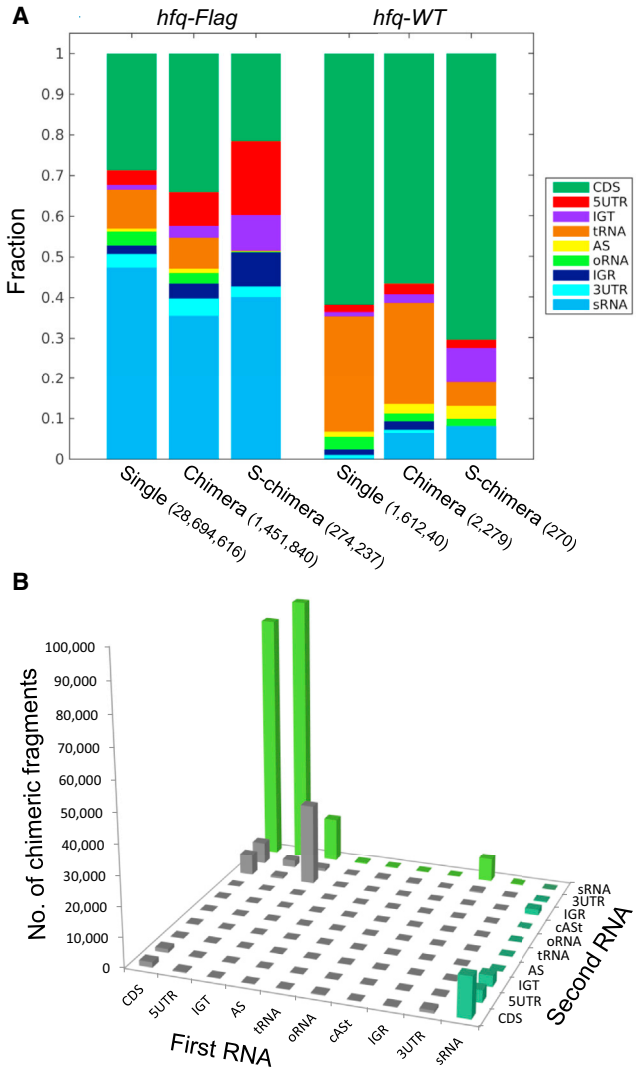


Figure 2. RIL-seq Data Are Enriched with sRNAs and mRNAs

(A) Distribution of RNAs derived from various genomic elements. Single (Single), chimeric (Chimera), and statistically significant chimeric (S-chimera) sequenced fragments from the log phase *hfq-Flag* and *hfq-WT* libraries were classified into nine major categories: 5UTR (5′ UTR), CDS, 3UTR (3′ UTR), tRNA, sRNA, oRNA (other non-coding RNAs), AS (antisense), IGR (intergenic region), and IGT (intergenic within transcript). rRNA-derived fragments were excluded. Fractions are shown (total counts are denoted in parentheses).

(B) Total number of S-chimera fragments for each combination of genomic elements. Mapped fragments were classified as in (A) with an additional subdivision of the AS category to cASt (*cis* antisense with putative *trans* target). Bars representing fragments with sRNAs as the first/second RNA in the chimera are colored in dark/light green, respectively.

See also Figure S3 and Table S2.

sequence motif in its set of RIL-seq putative targets, which is complementary to the sRNA known binding site, would support the identified targets. 20 of the known sRNAs had both at least four RIL-seq targets and a previously experimentally determined binding site (Peer and Margalit, 2011, 2014). We applied MEME (Bailey et al., 2009) to search for a common motif in each target sequence set of these sRNAs. For 18 sets we identified a common motif complementary to the sRNA. For 15 sRNAs the best common motif complemented the known sRNA binding site (Figures 4A and 4B and Table S4). Intriguingly, for some sRNAs such a motif could be identified in greater than 90% of the putative target sequences (Figure 4A and Table S4). Additionally, for MgrR, which lacks a known binding site, we identified a common motif in the target sequences that complements it, hinting at its binding site (Figures 4A and 4C). This suggests that such an analysis can identify the binding sites of newly discovered sRNAs (Figure 4D and Table S4). These results strongly support the validity of the targets identified by RIL-seq.

RIL-seq Captures True sRNA-Target Interactions

GcvB is a predominant sRNA in *E. coli*, which is highly conserved in evolution and regulates a large number of target genes, many of which encode amino acid transporters (Miyakoshi et al., 2015; Sharma et al., 2011). In *Salmonella*, GcvB base pairs with most of its targets through two binding sites, R1 and R2 (conserved also in *E. coli*), R1 being the dominant one (Sharma et al., 2011). We evaluated the capability of RIL-seq to capture true interactions, by applying it to a Δ*gcvB* strain carrying plasmids expressing WT *gcvB* or a *gcvB* where R1 was deleted, *gcvB*ΔR1. As demonstrated in Figure 4E and detailed in Table S5, targets of GcvB known to interact through R1 were revealed only with the WT *gcvB* plasmid, while targets known to interact not solely through R1 were revealed in both WT *gcvB* and *gcvB*ΔR1, indicating that the chimeras captured by RIL-seq represent true interactions. In this experiment the false negative rate was 20%, as we identified eight out of ten known GcvB interactions reported in Table S3. Following the pattern observed for known GcvB targets, we can deduce whether additional RIL-seq GcvB targets are bound mainly via R1 or via an alternative binding site (Figure 4E). This experiment not only supports the reliability of RIL-seq protocol and the revealed interactions, it also suggests that targets that share a binding site for a specific sRNA can be revealed by a "comparative RIL-seq."

Effect of the sRNA on the Expression Level of Its Bound Partners

The previous experimental and computational analyses strongly suggest that RIL-seq captures sRNA-target duplexes bound on Hfq, providing a snapshot of the sRNA interactome under the studied condition. Since functional sRNA interactions are believed to impact the transcript and/or protein expression level of the target, we tested whether the expression levels of RIL-seq bound partners are affected by the sRNA. To this end, we first used data from seven published transcriptome experiments that followed the change in gene expression upon overexpression of the sRNA (Supplemental Experimental Procedures). Indeed, for six of the seven sRNAs, the overall change in expression of the set of their bound partners was larger compared to the rest of the genes ($p \leq 0.002$ by Kolmogorov-Smirnov test; Figure S4A, right panel for each sRNA). Next, we conducted a similar analysis to test the sRNA effect on the translation of RIL-seq targets, using ribosome profiling data upon change in sRNA expression, available for RyhB (Wang et al., 2015). Indeed, there was a greater change in the ribosome occupancy of RyhB bound partners identified by RIL-seq compared to background

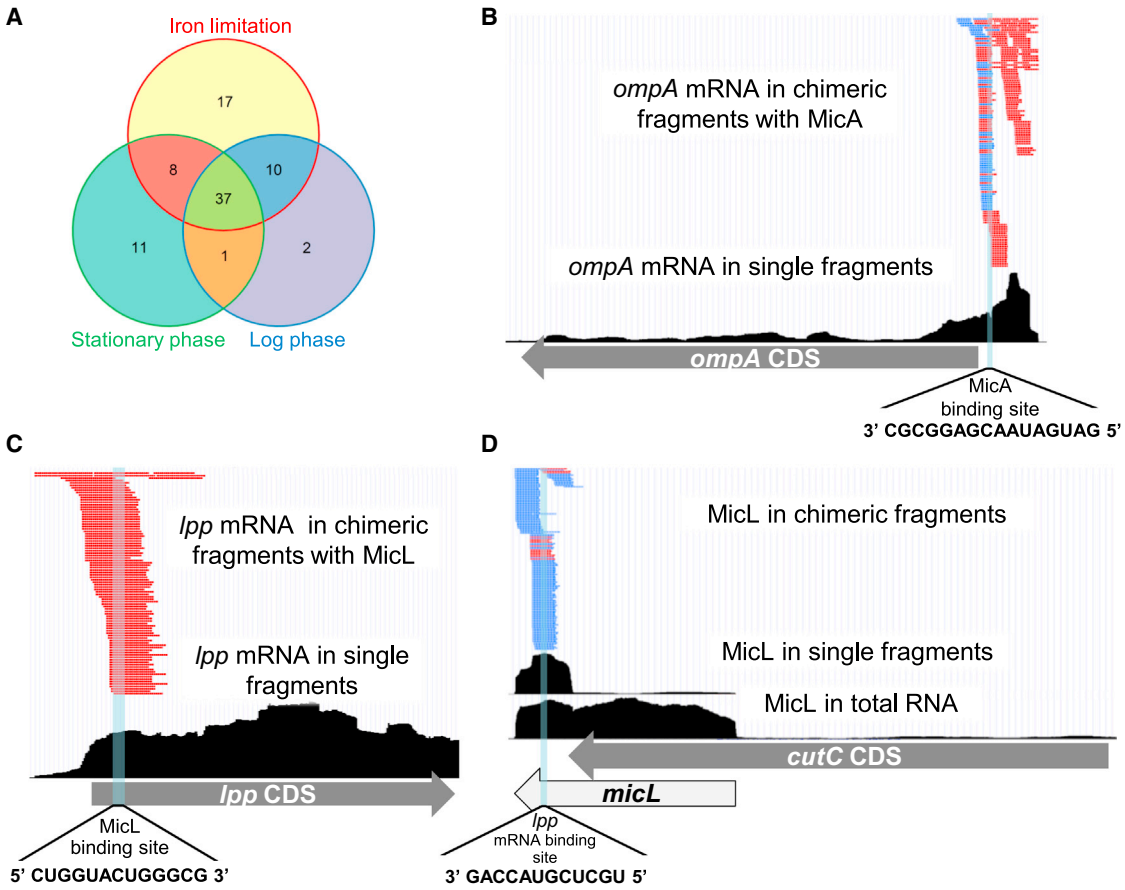


Figure 3. Known Targets Are Recovered by RIL-seq

(A) 86 out of 154 known sRNA-target pairs (~56%) were collectively identified by RIL-seq under the three tested conditions (not drawn to scale).

(B–D) Examples of identified interactions in log-phase library. (B) *ompA*-MicA interaction. Coverage of *ompA*, a known target of MicA is shown. Bottom: in single fragments (black). Top: in S-chimeras with MicA (broken lines in red or blue for *ompA* mRNA as first or second RNA in the chimera, respectively). MicA reported binding site (Udekwu et al., 2005) is highlighted. (C and D) *lpp*-MicL interaction. (C) Coverage of *lpp*, the major target of MicL, in RIL-seq library is shown. Bottom: in single fragments (black). Top: in S-chimeras with MicL (red broken lines). Colors are as in (B). The reported MicL binding site (Guo et al., 2014) is highlighted. (D) Coverage of MicL, a sRNA processed from a transcript transcribed from within the *cutC* gene, is shown. Bottom: in a total RNA library (black). Middle: in single fragments in RIL-seq library (black), showing that Hfq binds mainly the MicL processed fragment. Top: in S-chimeras (colors are as in B). The reported *lpp* mRNA binding site (Guo et al., 2014) is highlighted. Data are illustrated with UCSC genome browser (Kent et al., 2002).

See also Table S3.

genes ($p \leq 1.16 \times 10^{-8}$ by Kolmogorov-Smirnov test; Figure S4B, right panel).

A close inspection of the datasets of the transcriptome and ribosome profiling experiments suggested that each of the statistically significant results corresponded to a subset of the targets, while other targets showed relatively little changes in their transcript levels or ribosome occupancy (Figure S4, left panel for each sRNA). While it is possible that RIL-seq data include false-positive targets, the identified common motifs described above, which are complementary to the known binding sites of the sRNAs, suggest that many of these transcripts indeed form duplexes with the sRNA on Hfq. However, the impact of these interactions is not evident at the mRNA or translation level of the bound partners under the studied conditions.

To identify for each sRNA the subset of targets that not only bind but are also affected at their expression level, we computed the normalized odds ratio (see above). Using the ribosome profiling data, we verified that the normalized odds ratio values (Table S2, iron limitation data) of highly affected targets (fold change of ribosome profiling measure ≥ 2) were higher than the values of all other RyhB targets ($p \leq 8.42 \times 10^{-6}$ by Wilcoxon test). Thus, sRNA targets with relatively high normalized odds ratio values are more likely to be affected by the sRNA at their expression level compared to targets of the same sRNA with low values of this measure (Table S2).

Biological Insights from RIL-seq Data

RNA Order in the Chimeras Is Consistent with the Hfq Binding Mode of sRNAs

In the majority of chimeras involving sRNAs, the sRNA was the second RNA (RNA2) (i.e., at the 3′ end of the chimera), while the target was the first RNA (RNA1) (Figures 2B, 5A, S3C, S3D,

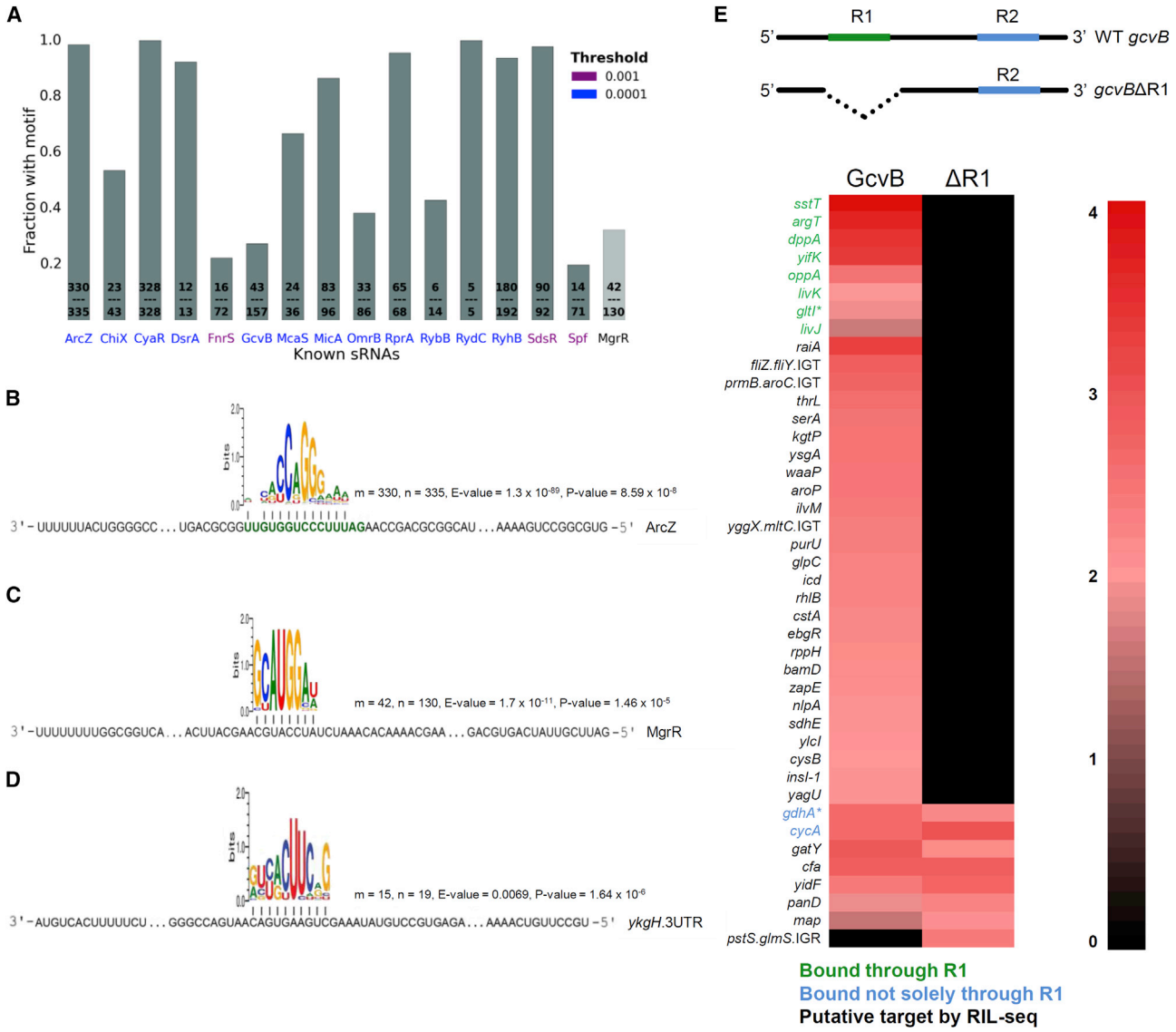


Figure 4. Computational and Experimental Analyses Support the Validity of the Interactions Revealed by RIL-seq

(A–D) Common sequence motifs in putative target sets are complementary to binding sites on the sRNA. (A) Bars represent the fractions of targets sharing a common motif that complements the known binding site of a sRNA. Numbers on the bars indicate the number of targets that shared the motif out of all RIL-seq targets of that sRNA. The color of the sRNA name presents the MAST p value threshold in which the motif matched the known binding site, when applicable. MgrR ($p \leq 5.55 \times 10^{-8}$) has no known binding site. (B–D) Common sequence motifs identified in RIL-seq target sets shown for: (B) ArcZ, (the known binding site is marked in green), (C) MgrR, showing a putative binding site, and (D) *ykgH* 3′ UTR, a putative 3′ UTR-derived sRNA, for which a putative binding site was identified. Indicated are the number of target sequences that shared the common motif (m), the number of sequences in the target set (n), the E value of MEME, and the p value of MAST. See also Table S4.

(E) RIL-seq captures true sRNA-target interactions. Upper part: a schematic representation of the plasmids expressing WT *gcvB* and *gcv*BΔR1 used in the experiment. Lower part: a heatmap presenting RIL-seq results of GcvB interactions identified in a Δ*gcvB* strain carrying WT *gcvB* or *gcvB*ΔR1 plasmid. For clarity, only known GcvB-target interactions and interactions supported by ≥100 sequenced fragments are shown. For each GcvB-target chimera, the normalized count of sequenced fragments in the respective library multiplied by 10^5 is presented (log scale). Targets of GcvB, shown in *Salmonella* to interact through R1 (green), are revealed only with the WT *gcvB*, while targets shown in *Salmonella* to interact not solely through R1 (blue) are revealed in both WT *gcvB* and *gcvB*ΔR1, implying RIL-seq captures true interactions. *Known targets in *Salmonella* not included in our compilation (Table S3).
See also Table S5.

S5A, and S5B). These findings are consistent with recent crystallographic results showing that the 3′ end of the sRNA is bound within the Hfq hexamer (Dimastrogiovanni et al., 2014), rendering it less accessible to the ligase. Furthermore, a search for a common motif in all RNA2 sequences identified a GC-rich motif followed by poly-U (Figure 5B), matching the Rho-independent

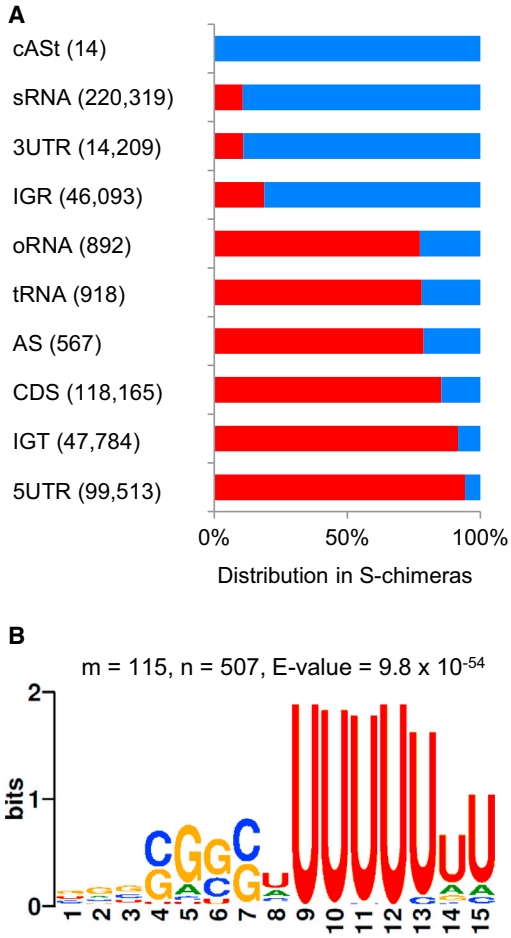


Figure 5. sRNAs Tend to be Second in the Chimeric Fragments

(A) Distribution of RNA locations as first (red) and second (blue) in S-chimera fragments for RNAs derived from various genomic elements. Mapped fragments were classified as in Figure 2A. See also Figures S5A and S5B and Table S6.

(B) The Rho-independent transcription terminator motif was revealed as a common motif for the second RNAs in S-chimeras.

See also Figure S5C. The results presented in (A) and (B) are based on S-chimera fragments in the log-phase dataset. n and m are as in Figure 4.

transcription terminator. Consistently, small-scale experiments showed that a poly-U tail at the end of sRNAs is critical for their binding to Hfq (Morita et al., 2015; Otaka et al., 2011; Sauer and Weichenrieder, 2011). Interestingly, we found that the poly-U tails of Rho-independent terminators of sRNAs are longer than those of mRNAs ($p \leq 2.7 \times 10^{-7}$ by Wilcoxon test; Figure S5C). A search for a common motif in RNA1 sequences of the chimeras did not yield such a motif. Notably, non-random order of RNAs in chimeras was also observed in miRNA-target chimeras on Argonaute (Moore et al., 2015). Altogether, it seems that RIL-seq data hold valuable information on RNA binding to other RNAs as well as to Hfq itself.

sRNA Candidates Are Encoded within Various Genomic Elements

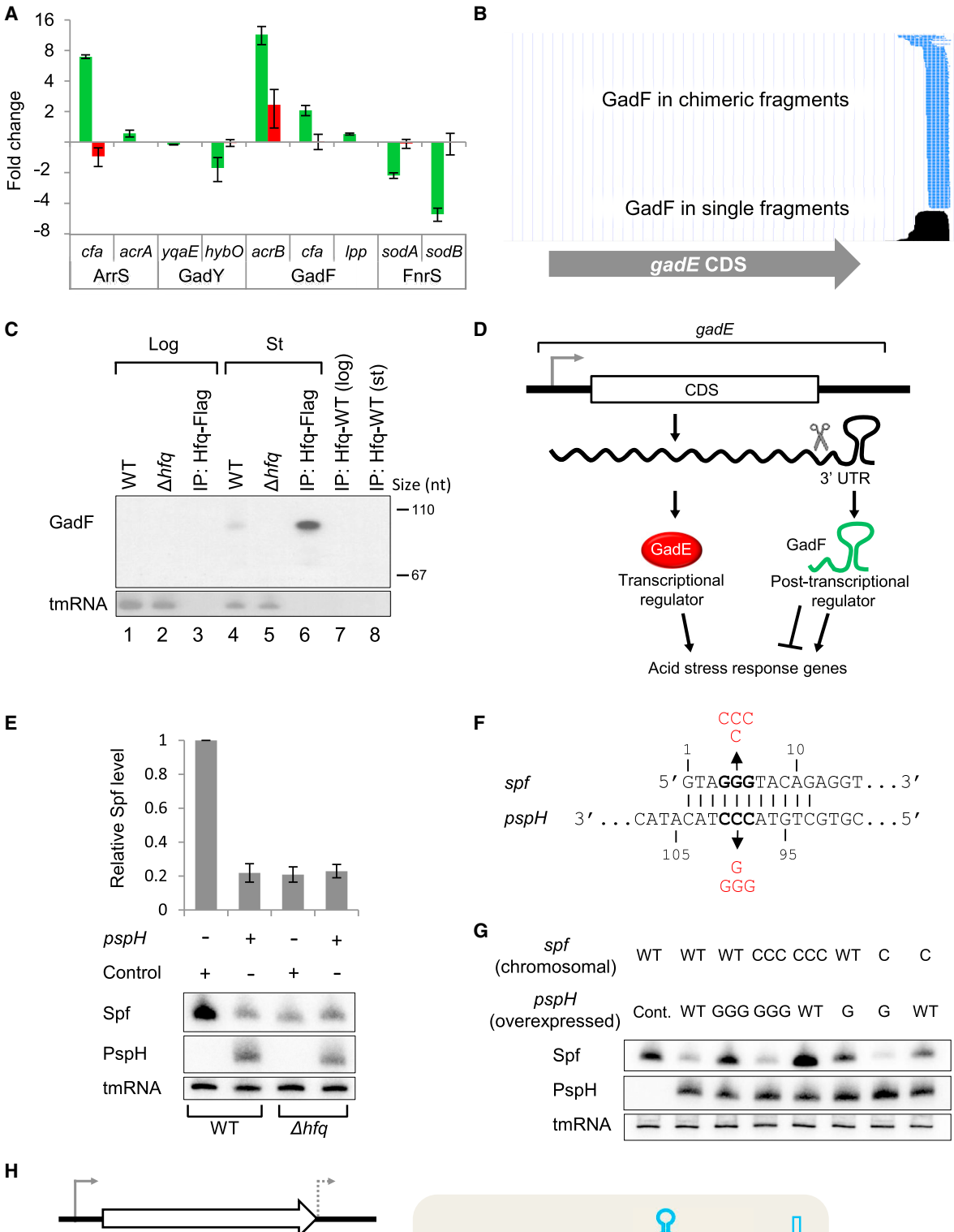
About one-third of RIL-seq interactions do not involve well-established sRNAs (Figure 2B and Table S2). We performed a rigorous analysis examining if the RNAs involved in these interactions show features of known sRNAs. These features included the location of sRNAs as second RNA in the chimeric fragment, a sequence compatible with Rho-independent terminator with a relatively long poly-U tail, as well as involvement in multiple interactions with CDSs or 5′ UTRs (Table S6). We found that RNAs encoded in various genomic elements could be characterized by their location in the chimeric fragment (Figures 5A, S5A, and S5B). tRNAs and RNAs derived from 5′ UTR, CDS, intergenic regions within transcripts and antisense regions tended to appear first in the chimeras, as observed for targets of sRNA. RNAs derived from 3′ UTR and from intergenic regions between transcripts tended to be second in the chimeras, supporting their annotation as putative sRNAs. In addition, the poly-U tails at the 3′ ends of these sRNA candidates are on average longer than those of other mRNAs ($p \leq 0.012$ by Wilcoxon test; Figure S5C). In further support of this classification, the sequences of interactors for 10 of these sRNA candidates show common sequence motifs that are complementary to the respective sRNA sequence (Figure 4D and Table S4). While most sRNA candidates originated from 3′ UTRs and intergenic regions, there were exceptions. For example, fragments overlapping the CDS of *aceK* and the 5′ UTR of *motA* were preferentially located as second in their chimeras, exhibited a long U-tail, and had relatively high numbers of interactions, mostly with CDS-derived fragments, suggesting that they may be classified as sRNAs.

cis-Antisense RNAs Can Act as trans Regulators

Whereas most antisense regions identified in chimeras exhibited characteristics of targets, two antisense RNAs, ArrS and GadY, which were previously documented to act as *cis*-antisense RNA regulators in the acid response pathway, exhibited characteristics of sRNAs (cASt in Figure 5A). GadY was shown to bind Hfq (Opdyke et al., 2004), supporting its classification by RIL-seq data as an Hfq-dependent sRNA. RIL-seq identified bound partners for GadY and ArrS, enriched with genes involved in acid stress response ($p \leq 0.004$ and $p \leq 6.4 \times 10^{-7}$, respectively, by Fisher's exact test), some of which were further supported by reporter experiments using GFP translational fusions (Corcoran et al., 2012) (Supplemental Experimental Procedures, Figure 6A, and Table S7). These findings support a model by which some sRNAs can act as both *cis* and *trans* regulators (Thomason and Storz, 2010; Mandin and Gottesman, 2010; Jäger et al., 2012; Sayed et al., 2012). Furthermore, our data suggest that the same RNA can act within a pathway as both *cis* and *trans* regulator. Interestingly, no sense-antisense S-chimeras were revealed in RIL-seq data, probably because this type of interaction does not require the involvement of Hfq.

A Regulatory Circuit Involving a Transcription Factor and a sRNA Encoded in the Same Transcript

E. coli often encounters acidic environments and hence has evolved sophisticated regulatory systems to survive under these conditions. A central transcription factor that regulates genes involved in acid stress response is GadE, which is highly expressed in stationary phase and under acidic stress (Hommais et al., 2004; Masuda and Church, 2003). We identified a putative sRNA in the 3′ UTR region of *gadE* (annotated originally in



(legend on next page)

our data as gadE.mdtE.IGT and termed here GadF), which was highly abundant among both single and chimeric fragments of stationary-phase libraries (Table S2 and Figure 6B). We further verified that GadF is an autonomous RNA by northern blot analysis (Figure 6C). Apparently, GadF is processed from *gadE* transcript, as *gadE* full-length transcripts were recognized by a probe complementary to GadF (Figure S6A), and GadF level substantially declined in RNA-seq libraries treated with 5′ phosphate-dependent exonuclease (Thomason et al., 2015). GadF also appears to be involved in the response to acidic stress, as its RIL-seq targets were enriched with genes involved in acid response ($p \leq 2.7 \times 10^{-5}$ by Fisher's exact test). Our results suggest an intriguing structure of a regulatory circuit in which a transcription factor (GadE) and a sRNA (GadF) are encoded in the same transcript and co-regulate genes in the same pathway (Figure 6D).

Two of GadF RIL-seq targets, *acrB* and *cfa*, were upregulated upon GadF overexpression, as indicated by GFP reporter assays (Figure 6A). *cfa* was also upregulated by ArrS (Figure 6A). Cfa protein plays an important role in membrane modification, conferring resistance to acidic stress (Chang and Cronan, 1999). Recently, another sRNA, RydC in *Salmonella*, was shown to upregulate *cfa* by sequestration of RNase E binding site at its 5′ UTR (Fröhlich et al., 2013). The binding sites of RydC and *cfa* are both conserved in *E. coli*, and indeed the *cfa*-RydC interaction was detected by RIL-seq. Interestingly, ArrS and GadF probably use the same activation mechanism as RydC, as their putative binding sites on *cfa* (Table S7) match the RydC binding site reported in *Salmonella*. Finally, ArrS is also known to stabilize the mRNA of the transcription factor GadE by binding to it in *cis*. It seems that there is an intricate network of sRNAs involved in regulating the acid stress response in *cis* and *trans*.

A 3′ UTR-Derived sRNA Acts as a Sponge of Spf

RIL-seq identified an interaction between a putative sRNA originating from the 3′ UTR region of *pspG* (annotated originally in our data as pspG.qor.IGR and termed here PspH) and its RIL-seq-identified sole partner, the sRNA Spf. Data from previous studies (Romero et al., 2014; Thomason et al., 2015) suggest that this sRNA could be independently transcribed from its own promoter. PspH overexpression from a plasmid led to almost 5-fold reduction in Spf level in WT *E. coli*, while in Δ*hfq* mutant the level of Spf was rather low already without PspH, and its overexpression had no effect (Figure 6E). In a complementary assay, Spf overexpression had a small effect on PspH level (Figure S6B). However, since Spf level was high at the tested condition even without plasmid expression, we cannot conclude whether the Spf-PspH interaction also affects the level of PspH. In addition, we showed by compensatory mutations that PspH directly base pairs with Spf through the known binding site of Spf (Beisel and Storz, 2011) (Figures 6F and 6G). Finally, overexpression of PspH led to upregulation of Spf targets *sthA* and *gltA* in WT cells, while no change in their expression was observed in Δ*spf* cells (Figure S6C). Our results support a role for PspH as a sponge of Spf (Figure 6H).

DISCUSSION

RIL-seq Pros and Cons

RIL-seq protocol has several prominent advantages: first, by ligating the Hfq-bound RNAs it directly captures the interacting partners without any additional computational sequence matching. Second, RIL-seq does not require manipulations of the cellular RNAs, such as over- or under-expression, and can be applied to cells without interfering with the in vivo cellular environment under study. Third, it simultaneously captures the Hfq-mediated interactions of all sRNAs expressed under a given condition. This is crucial for obtaining a realistic picture of the sRNA interactome, as due to the nature of sRNA-target interaction, targets that are co-regulated by the same RNA affect each other's binding to the sRNA and sRNAs that share targets affect each other's binding to the target (Figueroa-Bossi et al., 2009; Lalaouna et al., 2015; Miyakoshi et al., 2015; Overgaard et al., 2009). RIL-seq inherently takes these mutual effects into

Figure 6. Functional sRNAs

(A) Effect of ArrS, GadY and GadF on the GFP intensity of GFP translational fusions of targets identified by RIL-seq. The effect of FnrS on its GFP fused known targets, *sodA* and *sodB*, is shown as a positive control. The fold change between GFP fluorescence in the presence of a sRNA overexpressing plasmid and of a control plasmid is presented (WT cells, green bars; Δ*hfq* mutant, red bars). To maintain a unified scale, negative values represent the fold change of control/sRNA overexpression (reduction in GFP intensity) and positive values represent the fold change of sRNA overexpression/control (increase in GFP intensity). Error bars indicate 95% confidence intervals based on at least three independent repeats. The predicted binding sites for each pair are shown in Table S7.

(B) The coverage of GadF, a sRNA encoded in the *gadE* 3′ UTR, in RIL-seq stationary-phase library. Bottom: single fragments (black). Top: S-chimeras (broken lines in blue, indicating that GadF is found always as second RNA in the chimera).

(C) Northern blot of GadF. Total RNA from cells grown to log phase (OD_{600} = 0.3) (lanes 1, 2) and to stationary phase (6 hr) (lanes 4, 5); RNA coimmunoprecipitated from lysates of *hfq-Flag* cells (lanes 3, 6) or *hfq-WT* cells (lanes 7, 8) grown to log or stationary phase.

(D) Schematic representation of a regulatory circuit involving a transcription factor (GadE) and a sRNA (GadF), generated from the same transcript and regulating genes in the acid stress response pathway at the transcriptional and posttranscriptional level, respectively.

(E) The effect of overexpressing PspH on its target, Spf. Northern blot analysis of total RNA from WT or Δ*hfq* cells carrying either a pJV300 plasmid (Control) or a *pspH* overexpressing plasmid (*pspH*). The relative Spf level was determined from three independent replicates; Error bars are as in (A).

(F) Spf-PspH base-pairing region. Compensatory mutations in chromosomal *spf* and in plasmid-borne *pspH*, of either one base change (G or C) or three base changes (GGG or CCC), are indicated in red.

(G) The effect of compensatory mutations in *spf* and *pspH* on Spf levels. WT cells or cells carrying mutations in chromosomal *spf* were transformed with a PspH (WT or mutated) overexpressing plasmid or with a pJV300 plasmid as control (Cont.). Cells were grown to OD_{600} = 0.3, and total RNA was extracted and subjected to northern blot.

(H) Schematic description of Spf regulation by PspH. PspH acts as a sponge of Spf, leading to a decrease in Spf level and to abolishment of the negative regulation of Spf targets. In all northern blot experiments, tmRNA served as a loading control.

See also Figure S6.

account. Taken together, these properties of RIL-seq allow it to provide an in vivo snapshot of the Hfq-mediated sRNA interactome under any desired naturally occurring or artificially manipulated condition.

Since targeting by sRNAs is achieved by base pairing, there have been attempts to harness computational approaches employing binding free energy considerations to develop target prediction algorithms, such as CopraRNA (Wright et al., 2013), which also takes into account evolutionary considerations. While RIL-seq is slightly better than CopraRNA at identifying known targets in *E. coli* (53.4% versus 52.7% sensitivity and 4.7% versus 2.7% precision; see Supplemental Experimental Procedures), it has several important advantages over current computational approaches. (1) It provides a specific sRNA-target network per condition, capturing the effect of competition between RNAs and allowing network comparison between conditions, while the network established by predictions is generic, ignoring the in vivo considerations. (2) It does not require prior knowledge of the RNA sequences, which is essential for computational predictions. (3) RIL-seq is not limited to reported sRNAs and enables the detection of novel sRNAs along with their interacting partners. In fact, RIL-seq can be potentially applied to any bacteria with a sequenced genome, without previous knowledge of its sRNA repertoire, given the interactions are protein mediated. Especially, it can be a powerful tool to reveal sRNA functions in bacterial pathogenesis. Many pathogenic bacteria express Hfq-dependent sRNAs from horizontally acquired virulence regions. These sRNAs as well as their binding sites are usually less conserved, and thus it is more difficult to predict their targets computationally (Padalon-Brauch et al., 2008; Westermann et al., 2016). RIL-seq can overcome this obstacle since it captures interacting RNAs without any prior knowledge of their binding sites.

Like other large-scale approaches, RIL-seq is prone to potential technical biases that can affect the final set of reported interactions. Two major bias types can be noted. (1) Over- or under-representation of true biological interactions due to the characteristics of the involved RNA molecules. These may include, for example, distinct sequences and/or structures that may affect the ligation potential. However, our analysis revealed a negligible ligation bias (Figures S2E–S2G). (2) Capture of non-biological interactions, resulting from non-specific binding to the antibody or ligation of RNAs transiently sharing the same Hfq hexamer. There are weak traces of such interactions in the data, such as those detected for the *hfq-WT* control, or from in vitro ligation of unrelated RNAs, as observed in the mixed yeast-*E. coli* library. However, the number of such interactions is rather small and is dramatically reduced with the application of the statistical filter. This emphasizes the importance of the statistical filtering step, which substantially reduces the number of spurious interactions and enriches RIL-seq final set with true interactions. Evidently our stringent criteria filter out known interactions, implying that the network we provide is incomplete. However, this is a price worth paying for increasing the reliability of the reported interactions, which is demonstrated by the GcvB experiment and the common motif analysis (Figure 4 and Table S4). Our assessments strongly imply that the S-chimeras represent duplexes that take place on Hfq and that there is an extensive network of base-pairing RNAs on Hfq.

Different Consequences of sRNA-Target Interaction

The whole spectrum of RNA pairs in RIL-seq data can be divided into two subsets: those that were identified but did not pass the statistical filter and those that passed the statistical filter and are included in the RIL-seq final set (Table S2). Most of the RNA pairs in the first subset (including 59 known interactions) probably represent transient interactions that stem from low expression of the RNA participants under the studied conditions, or weak binding. In support of that we found for the three studied conditions a correlation of about 0.5 (Spearman correlation, $p \leq 2.2 \times 10^{-31}$) between the expression levels of interacting transcripts and their involvement in chimeric fragments, which was increased to ~0.75 ($p \leq 2.9 \times 10^{-4}$) when only the known sRNAs themselves were regarded. Also, the RNA duplexes that did not pass the statistical filter were predicted to be less stable than the other duplexes (median of ~−4.4 versus ~−8.2 kcal/mol for computed hybridization free energy (Mückstein et al., 2006); $p \leq 1.0 \times 10^{-10}$ by Wilcoxon test). Thus, the RNA pairs in the RIL-seq final dataset form relatively stable duplexes on Hfq, as also supported by the motif analysis (Figures 4A–4D). However, while these interactions are relatively stable, their impact on the target's transcript level or translation was not obvious. Some targets were affected at their transcript or translation level, while others were not (Figure S4). A similar observation was reported recently for miRNA targets revealed by the CLASH methodology (Helwak et al., 2013; Agarwal et al., 2015).

The potential effect of the interaction depends on many factors, which are not necessarily directly represented by RIL-seq data, such as the enrichment of a target mRNA on Hfq. The normalized odds ratio measure, calculated within the computational pipeline, may assist in distinguishing functional targets of a given sRNA from its binders per se under the studied condition (Table S2). The above conjectures about Hfq-mediated sRNA-target interaction and regulation are consistent with current models that suggest continuous recycling of transcripts on Hfq (Fender et al., 2010), and kinetic differences between the interactions and regulatory outcomes for different targets of a sRNA (Fei et al., 2015).

RIL-seq Substantially Expands the sRNA and Target Repertoires

We discovered a total of ~2,800 putative interactions mediated via Hfq, expanding the current sRNA interactome by an order of magnitude (Table S2). RIL-seq re-discovered ~56% of previously reported sRNA-target interactions (Table S3) and also identified recently reported less-conventional interactions, such as an interaction between two sRNAs, GcvB and SroC (Miyakoshi et al., 2015).

While the sRNAs detected in the first genome-wide surveys were predominantly independent transcripts encoded in intergenic regions (Argaman et al., 2001; Rivas et al., 2001; Wassarman et al., 2001), recent studies revealed expression of candidate sRNAs originating from various genomic regions, including antisense to coding regions (Thomason et al., 2015), 3′ UTR

(Chao et al., 2012; Chao and Vogel, 2016), 5′ UTR, and even coding regions (Bilusic et al., 2014). Consistently, ~2/3 of RIL-seq interactions involved 25 of the 26 well-characterized sRNAs (Figure S3E), while 1/3 involved other RNAs, including additional sRNAs derived from various genomic regions along with their bound partners. Whereas many of these transcripts were identified before, in transcriptome or Hfq coIP studies, their classification as sRNAs was indefinite, and their capturing with their targets by RIL-seq supports this classification and provides clues into their cellular roles. In turn, finding these putative sRNAs in previous studies supports our conclusions. Table S6 includes additional information about all RNAs in RIL-seq data, including the support of sRNA candidates by previous studies.

Some of the discovered 3′ UTR-derived sRNAs determine intriguing regulatory circuits, where two layers of regulation are encoded within the same transcript. One such regulatory module, encoded in *cpxP* transcript and defined by the protein CpxP and the sRNA CpxQ, both involved in response to cell envelope stress, was recently reported (Chao and Vogel, 2016; Grabowicz et al., 2016) and identified here as well. Another circuit discovered in our study, which seems to play a role in acid stress response (Figure 6A), is determined by the transcription factor GadE and the sRNA GadF, both encoded in *gadE* transcript (Figure 6D). In such circuits there is ultimate coordination between the two regulators.

Dynamic Changes in the Hfq-Mediated sRNA Interactome upon Change in Cellular Conditions

RIL-seq data suggest that there is substantial re-wiring of the sRNA-target interaction network between conditions, reflecting the changing transcriptome and the different functional roles of the sRNAs (Figure 7). Changes in interactions between conditions can be achieved either by change in the cellular levels of sRNAs or target mRNAs or by the expression of additional targets that attract the regulator (competition effects). Indeed, we observed both scenarios in our data. Consistent with previous coIP experiments (Chao et al., 2012), we observed a dynamic change in the Hfq-bound sRNA repertoire upon change in cellular conditions (Figure S3E). For example, under iron limitation the relative level of RyhB increased and the number of interactions involving RyhB increased accordingly (Figures 7B–7D). An example of the second scenario is provided by GcvB, for which the distribution of interactions changed dramatically between log and stationary phase (Table S2). Our data indicate that in stationary phase 70% of GcvB's interactions were with its sponge, SroC (Miyakoshi et al., 2015), lowering the number of interactions with other targets and suggesting a relief of the repression of these targets. Indeed, we observed that the expression levels of most GcvB known targets were higher in stationary phase than in log phase ($p \leq 3.38 \times 10^{-2}$ by one-sided Wilcoxon test; Table S2). Sponging interactions were also reported for ChiX by *chbBC* intergenic region (Figueroa-Bossi et al., 2009; Overgaard et al., 2009) and for RyhB and RybB by the 3′ processed region of *leuZ* tRNA (Lalaouna et al., 2015), all also identified by RIL-seq. Another sponge we identified is PspH, a sRNA overlapping the 3′ UTR of *pspG*, which interacted with Spf and led to upregulation of its targets (Figures 6E–6H). In summary, increasing and decreasing the activity of a sRNA can be achieved by changing its cellular level or by the binding of a sponge, and RIL-seq data provide hints at these mechanisms.

Outlook

The examples of sRNA-target interactions discussed here are only the tip of the iceberg of RIL-seq data. Given that ~95% of the 2,817 unique interactions revealed in our data were not reported before, the potential outcome of investigating them, as well as interactions that might be found in other growth conditions, is exciting. As RIL-seq is not restricted to a certain type of RNA, a specific RNA-binding protein, or a specific organism, it can be widely applied for finding RNA interacting pairs in general and sRNA-target interactions in particular. In many bacteria, including pathogenic or biotechnologically relevant bacteria, sRNA research is in its infancy. Applying RIL-seq to these organisms will provide a precious view of their posttranscriptional regulatory network, leading to further insights into the regulation of their physiology and/or pathogenesis.

EXPERIMENTAL PROCEDURES

Strains and Media

The *E. coli* MG1655 or TOP10 strains served as wild-type. An *E. coli* MG1655 strain carrying *hfq-Flag* (HM34) was used in the RIL-seq protocol. All other bacterial and yeast strains used in this study are listed in Table S8. Bacterial strains were routinely grown at 37°C in rich medium. Antibiotics or 2,2′-Dipyridyl (200 μM) were added where appropriate. All oligonucleotides and plasmids used in this study are listed in Table S8.

RIL-seq Protocol

The RIL-seq protocol is described here in brief and detailed in the Supplemental Experimental Procedures.

Experimental Procedure

hfq-Flag strain or *hfq-WT* strain (control) were grown to log phase, with or without the addition of the iron chelator 2,2′-Dipyridyl, or to stationary phase, and exposed to 800 mJ of 254 nm UV irradiation in order to in vivo crosslink the proteins and RNA molecules. Following mechanical lysis of the cells, Hfq with its bound RNAs was coimmunoprecipitated using anti-Flag antibody. Exposed regions of the RNAs were trimmed by RNases, the 5′ and 3′ ends of the RNAs were treated with T4 PNK, neighboring RNAs were ligated, the Hfq protein was digested, and RNA was then isolated. Sequencing libraries (RIL-seq and total RNA libraries) were constructed by RNAtag-seq protocol (Shishkin et al., 2015), with a few modifications that prevented loss of sRNA fragments. The libraries were sequenced by paired-end sequencing using Nextseq500 Sequencer (Illumina).

Computational Procedure

Raw reads were processed to remove adaptor sequences, low quality ends, and low-complexity reads. The first 25 nucleotides of each read were mapped to the genome of *E. coli* K12 MG1655 (GenBank: NC_000913.2), sorting the fragments into "single" and "chimeric" fragments.

Next, for the detection of over-represented chimeras, we divided the genome into non-overlapping 100 nt-long windows and for each window pair counted the number of mapped chimeric sequenced fragments. We applied Fisher's exact test to identify the chimeric fragments that appeared in the experiment statistically significantly more than expected at random based on the number of single and chimeric fragments mapped to each region ($p \leq 0.05$ after Bonferroni correction). The computational pipeline is implemented in Python and is available as a package called RILseq on pypi (https://pypi.python.org/pypi/RILseq). Annotation of inferred interacting RNAs was based on EcoCyc (http://ecocyc.org) version 19.0 (Keseler et al., 2013). Representation of RIL-seq single and chimeric fragments in UCSC genome browser (Kent et al., 2002) can be

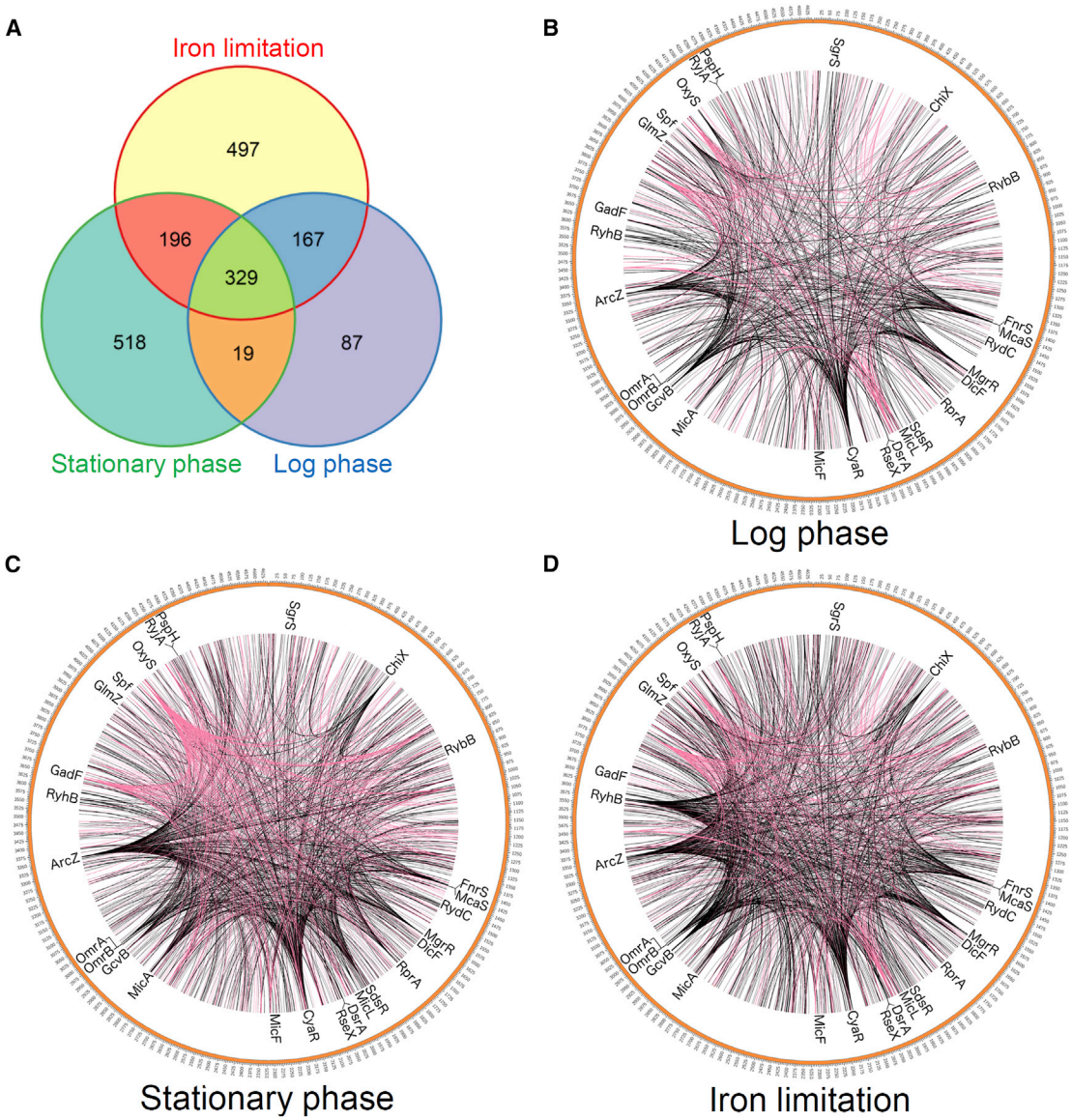


Figure 7. Substantial Re-wiring of the Posttranscriptional Regulatory Network between Conditions

(A) Numbers of unique interactions of known sRNAs with targets detected by RIL-seq under the three tested conditions. 1,813 unique interactions were collectively identified for these 25 sRNAs.

(B–D) Re-wiring of the sRNA regulatory network between conditions. RIL-seq interactions are represented on *E. coli* chromosome. Edges connect between the genomic locations of interacting RNAs, where the thickness of an edge is proportional to the normalized number of chimeras. Interactions involving known sRNAs are marked in black, and all other interactions are in pink. Re-wiring is observed for many sRNAs (note, for example, the increase in RyhB interactions under iron limitation). Network views were drawn by circos software (http://circos.ca/).

found in the following links: RIL-seq single fragments (http://microbes.ucsc.edu/cgi-bin/hgTracks?db=eschColi_K12&hgt.customText=http://www.cs.huji.ac.il/~asafp5/RILseq/RILseq_coverage.wig.gz), RIL-seq chimeric fragments (http://microbes.ucsc.edu/cgi-bin/hgTracks?db=eschColi_K12&hgt.customText=http://www.cs.huji.ac.il/~asafp5/RILseq/RILseq_interactions.bed.gz). Note that chimeras with $\geq$5 sequenced fragments are presented, but only chimeras with $\geq$10 sequenced fragments were included in all other analyses in the paper; chimera tracks are hidden by default.

GcvB Comparative RIL-seq

RIL-seq experiments were performed in log phase using *E. coli* Δ*gcvB* strain (HM54) transformed with a WT *gcvB* or a *gcvB*ΔR1 plasmid, each repeated three times. Mapping of the sequenced reads was done to the chromosome and to the relevant plasmid. Presented interactions in Figure 4E and Table S5 are those identified in both the unified log-phase libraries and in this experiment.

Classification of RNAs Revealed by RIL-seq

RNAs were classified based on their genomic annotation (see legend of Figure 2 and Supplemental Experimental Procedures). Importantly, only sRNAs with at least one validated *trans* target were regarded as sRNAs. Genomic regions encoding other annotated small RNAs were all termed other-RNA (oRNA). The fraction among the mapped single, chimeric, and S-chimera fragments was computed for each genomic element. For the assessment of

RIL-seq results, we compiled an updated dataset of 154 experimentally determined sRNA-target interactions (Table S3) based on our previous compilations (Peer and Margalit, 2011, 2014) and information in EcoCyc (http://ecocyc.org) version 19.0 (Keseler et al., 2013).

Recognition of Binding Motifs

The target sequences of each sRNA that had at least four targets (in each condition or summed over all tested conditions) were extracted and searched for a common motif by MEME (Bailey et al., 2009). To verify that an identified motif matched a complementary binding site on the respective sRNA, we searched for it in the sRNA reverse-complement sequence using MAST (Bailey and Gribskov, 1998). For more detail, see Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the sequencing data reported in this paper is ArrayExpress: E-MTAB-3910.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.07.026.

AUTHOR CONTRIBUTIONS

H.M. initiated and supervised the study; Methodology, S.M. and A.P.; Experimental Investigation, S.M., R.F.-R., and L.A.; Data Curation, A.P., Y.A., and R.F.-R.; Software Development, A.P.; Computational Analysis, A.P., Y.A., R.F.-R., Y.E.G., N.R., and A.B.; Writing – Original Draft, Review & Editing, S.M., A.P., Y.A., L.A., R.F.-R., and H.M.

ACKNOWLEDGMENTS

This study was supported by the European Research Council Advanced Grant #322920, 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 T. Kaplan for his helpful advice, R. Rehani for her contribution to the experiments, H. Aiba for the *hfq-Flag* strain, O. Pines for the *Saccharomyces cerevisiae* strain, S. Pearl-Mizrahi, O. Shechter, and M. Nitzan for technical assistance, G. Storz for work conducted by S.M. in her lab and for very helpful advice, S. Altuvia, E. Holmqvist, I. Rosenshine, and S. Ben-Yehuda for their very useful comments, and K. McDowall and J. Wade for data sharing.

Received: October 19, 2015
Revised: April 25, 2016
Accepted: July 27, 2016
Published: September 1, 2016

REFERENCES

Agarwal, V., Bell, G.W., Nam, J.W., and Bartel, D.P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. eLife *4*, 05005.

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G., Margalit, H., and Altuvia, S. (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr. Biol. *11*, 941–950.

Bailey, T.L., and Gribskov, M. (1998). Methods and statistics for combining motif match scores. J. Comput. Biol. *5*, 211–221.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Beisel, C.L., and Storz, G. (2011). The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. Mol. Cell *41*, 286–297.

Bilusic, I., Popitsch, N., Rescheneder, P., Schroeder, R., and Lybecker, M. (2014). Revisiting the coding potential of the *E. coli* genome through Hfq co-immunoprecipitation. RNA Biol. *11*, 641–654.

Chang, Y.Y., and Cronan, J.E., Jr. (1999). Membrane cyclopropane fatty acid content is a major factor in acid resistance of *Escherichia coli*. Mol. Microbiol. *33*, 249–259.

Chao, Y., and Vogel, J. (2016). A 3′ UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. Mol. Cell *61*, 352–363.

Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M., and Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J. *31*, 4005–4019.

Corcoran, C.P., Podkaminski, D., Papenfort, K., Urban, J.H., Hinton, J.C., and Vogel, J. (2012). Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. Mol. Microbiol. *84*, 428–445.

Darnell, R.B. (2010). HITS-CLIP: panoramic views of protein-RNA regulation in living cells. Wiley Interdiscip. Rev. RNA *1*, 266–286.

Dimastrogiovanni, D., Fröhlich, K.S., Bandyra, K.J., Bruce, H.A., Hohensee, S., Vogel, J., and Luisi, B.F. (2014). Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. eLife *3*, e05375.

Fei, J., Singh, D., Zhang, Q., Park, S., Balasubramanian, D., Golding, I., Vanderpool, C.K., and Ha, T. (2015). RNA biochemistry. Determination of *in vivo* target search kinetics of regulatory noncoding RNA. Science *347*, 1371–1374.

Fender, A., Elf, J., Hampel, K., Zimmermann, B., and Wagner, E.G. (2010). RNAs actively cycle on the Sm-like protein Hfq. Genes Dev. *24*, 2621–2626.

Figueroa-Bossi, N., Valentini, M., Malleret, L., Fiorini, F., and Bossi, L. (2009). Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. Genes Dev. *23*, 2004–2015.

Fröhlich, K.S., Papenfort, K., Fekete, A., and Vogel, J. (2013). A small RNA activates CFA synthase by isoform-specific mRNA stabilization. EMBO J. *32*, 2963–2979.

Grabowicz, M., Koren, D., and Silhavy, T.J. (2016). The CpxQ sRNA negatively regulates Skp to prevent mistargeting of β-barrel outer membrane proteins into the cytoplasmic membrane. MBio *7*, e00312–e00316.

Grosswendt, S., Filipchyk, A., Manzano, M., Klironomos, F., Schilling, M., Herzog, M., Gottwein, E., and Rajewsky, N. (2014). Unambiguous identification of miRNA:target site interactions by different types of ligation reactions. Mol. Cell *54*, 1042–1054.

Guo, M.S., Updegrove, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A., and Storz, G. (2014). MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. Genes Dev. *28*, 1620–1634.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D. (2013). Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell *153*, 654–665.

Holmqvist, E., Wright, P.R., Li, L., Bischler, T., Barquist, L., Reinhardt, R., Backofen, R., and Vogel, J. (2016). Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. EMBO J. *35*, 991–1011.

Hommais, F., Krin, E., Coppée, J.Y., Lacroix, C., Yeramian, E., Danchin, A., and Bertin, P. (2004). GadE (YhiE): a novel activator involved in the response to acid environment in *Escherichia coli*. Microbiology *150*, 61–72.

Jäger, D., Pernitzsch, S.R., Richter, A.S., Backofen, R., Sharma, C.M., and Schmitz, R.A. (2012). An archaeal sRNA targeting cis- and trans-encoded mRNAs via two distinct domains. Nucleic Acids Res. *40*, 10964–10979.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., et al. (2013). EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res. *41*, D605–D612.

Lalaouna, D., Carrier, M.C., Semsey, S., Brouard, J.S., Wang, J., Wade, J.T., and Massé, E. (2015). A 3′ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise. Mol. Cell *58*, 393–405.

Mandin, P., and Gottesman, S. (2010). Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J. *29*, 3094–3107.

Masuda, N., and Church, G.M. (2003). Regulatory network of acid resistance genes in *Escherichia coli*. Mol. Microbiol. *48*, 699–712.

Miyakoshi, M., Chao, Y., and Vogel, J. (2015). Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. EMBO J. *34*, 1478–1492.

Moore, M.J., Scheel, T.K.H., Luna, J.M., Park, C.Y., Fak, J.J., Nishiuchi, E., Rice, C.M., and Darnell, R.B. (2015). miRNA-target chimeras reveal miRNA 3′-end pairing as a major determinant of Argonaute target specificity. Nat. Commun. *6*, 8864.

Morita, T., Maki, K., and Aiba, H. (2005). RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes Dev. *19*, 2176–2186.

Morita, T., Ueda, M., Kubo, K., and Aiba, H. (2015). Insights into transcription termination of Hfq-binding sRNAs of *Escherichia coli* and characterization of readthrough products. RNA *21*, 1490–1501.

Mückstein, U., Tafer, H., Hackermüller, J., Bernhart, S.H., Stadler, P.F., and Hofacker, I.L. (2006). Thermodynamics of RNA-RNA binding. Bioinformatics *22*, 1177–1182.

Opdyke, J.A., Kang, J.G., and Storz, G. (2004). GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. J. Bacteriol. *186*, 6698–6705.

Otaka, H., Ishikawa, H., Morita, T., and Aiba, H. (2011). PolyU tail of rho-independent terminator of bacterial small RNAs is essential for Hfq action. Proc. Natl. Acad. Sci. USA *108*, 13059–13064.

Overgaard, M., Johansen, J., Møller-Jensen, J., and Valentin-Hansen, P. (2009). Switching off small RNA regulation with trap-mRNA. Mol. Microbiol. *73*, 790–800.

Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H., and Altuvia, S. (2008). Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. Nucleic Acids Res. *36*, 1913–1927.

Peer, A., and Margalit, H. (2011). Accessibility and evolutionary conservation mark bacterial small-rna target-binding regions. J. Bacteriol. *193*, 1690–1701.

Peer, A., and Margalit, H. (2014). Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. RNA *20*, 994–1003.

Rivas, E., Klein, R.J., Jones, T.A., and Eddy, S.R. (2001). Computational identification of noncoding RNAs in *E. coli* by comparative genomics. Curr. Biol. *11*, 1369–1373.

Romero, D.A., Hasan, A.H., Lin, Y.F., Kime, L., Ruiz-Larrabeiti, O., Urem, M., Bucca, G., Mamanova, L., Laing, E.E., van Wezel, G.P., et al. (2014). A comparison of key aspects of gene regulation in Streptomyces coelicolor and Escherichia coli using nucleotide-resolution transcription maps produced in parallel by global and differential RNA sequencing. Mol. Microbiol. *94*, 963–987.

Sauer, E., and Weichenrieder, O. (2011). Structural basis for RNA 3′-end recognition by Hfq. Proc. Natl. Acad. Sci. USA *108*, 13065–13070.

Sayed, N., Jousselin, A., and Felden, B. (2012). A cis-antisense RNA acts in trans in *Staphylococcus aureus* to control translation of a human cytolytic peptide. Nat. Struct. Mol. Biol. *19*, 105–112.

Sharma, C.M., Papenfort, K., Pernitzsch, S.R., Mollenkopf, H.J., Hinton, J.C., and Vogel, J. (2011). Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. Mol. Microbiol. *81*, 1144–1165.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., et al. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. Nat. Methods *12*, 323–325.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C., and Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genet. *4*, e1000163.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell *43*, 880–891.

Thomason, M.K., and Storz, G. (2010). Bacterial antisense RNAs: how many are there, and what are they doing? Annu. Rev. Genet. *44*, 167–188.

Thomason, M.K., Bischler, T., Eisenbart, S.K., Förstner, K.U., Zhang, A., Herbig, A., Nieselt, K., Sharma, C.M., and Storz, G. (2015). Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. J. Bacteriol. *197*, 18–28.

Tree, J.J., Granneman, S., McAteer, S.P., Tollervey, D., and Gally, D.L. (2014). Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. Mol. Cell *55*, 199–213.

Udekwu, K.I., Darfeuille, F., Vogel, J., Reimegård, J., Holmqvist, E., and Wagner, E.G. (2005). Hfq-dependent regulation of OmpA synthesis is mediated by an antisense RNA. Genes Dev. *19*, 2355–2366.

Wagner, E.G., and Romby, P. (2015). Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. Adv. Genet. *90*, 133–208.

Wang, J., Rennie, W., Liu, C., Carmack, C.S., Prévost, K., Caron, M.P., Massé, E., Ding, Y., and Wade, J.T. (2015). Identification of bacterial sRNA regulatory targets using ribosome profiling. Nucleic Acids Res. *43*, 10308–10320.

Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. *15*, 1637–1651.

Westermann, A.J., Förstner, K.U., Amman, F., Barquist, L., Chao, Y., Schulte, L.N., Müller, L., Reinhardt, R., Stadler, P.F., and Vogel, J. (2016). Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature *529*, 496–501.

Wright, P.R., Richter, A.S., Papenfort, K., Mann, M., Vogel, J., Hess, W.R., Backofen, R., and Georg, J. (2013). Comparative genomics boosts target prediction for bacterial small RNAs. Proc. Natl. Acad. Sci. USA *110*, E3487–E3496.

Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. Mol. Microbiol. *50*, 1111–1124.

Molecular Cell, Volume *63*

Supplemental Information

Global Mapping of Small RNA-Target Interactions in Bacteria

Sahar Melamed, Asaf Peer, Raya Faigenbaum-Romm, Yair E. Gatt, Niv Reiss, Amir Bar, Yael Altuvia, Liron Argaman, and Hanah Margalit

TABLE OF CONTENTS

Supplementary Figures

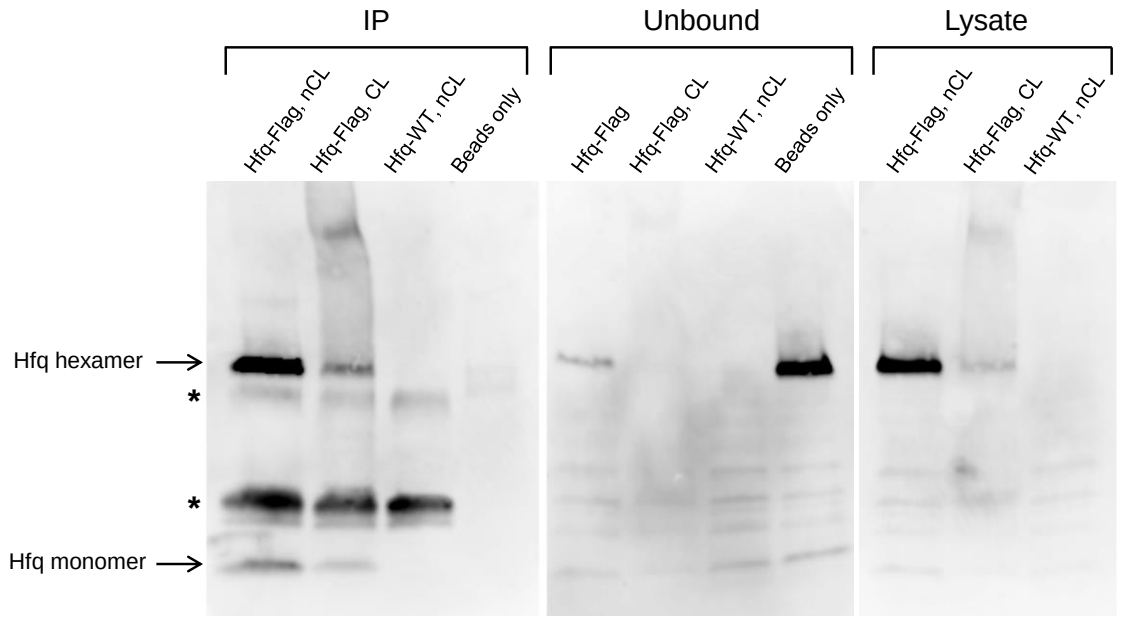


Figure S1 (related to Figure 1). IP assay of crosslinked (CL) and non-crosslinked (nCL) Hfq-Flag

hfq-Flag and *hfq-WT* strains were grown to log phase, the cells were exposed to UV irradiation in order to generate protein-RNA crosslinking or unexposed, and cell lysates were prepared. The lysates were subjected to IP assay using magnetic beads carrying M2 anti-Flag monoclonal antibody. The lysates, unbound fraction and bound fraction (IP) were analyzed by Western blotting using Anti-Flag antibody. As a control for the IP assay, we incubated *hfq-Flag* lysate, without crosslinking, with magnetic beads carrying no M2 Anti-Flag antibody (Beads only).

* Bands of M2 anti-Flag monoclonal antibody light (bottom) and heavy (middle) chains in the IP lanes.

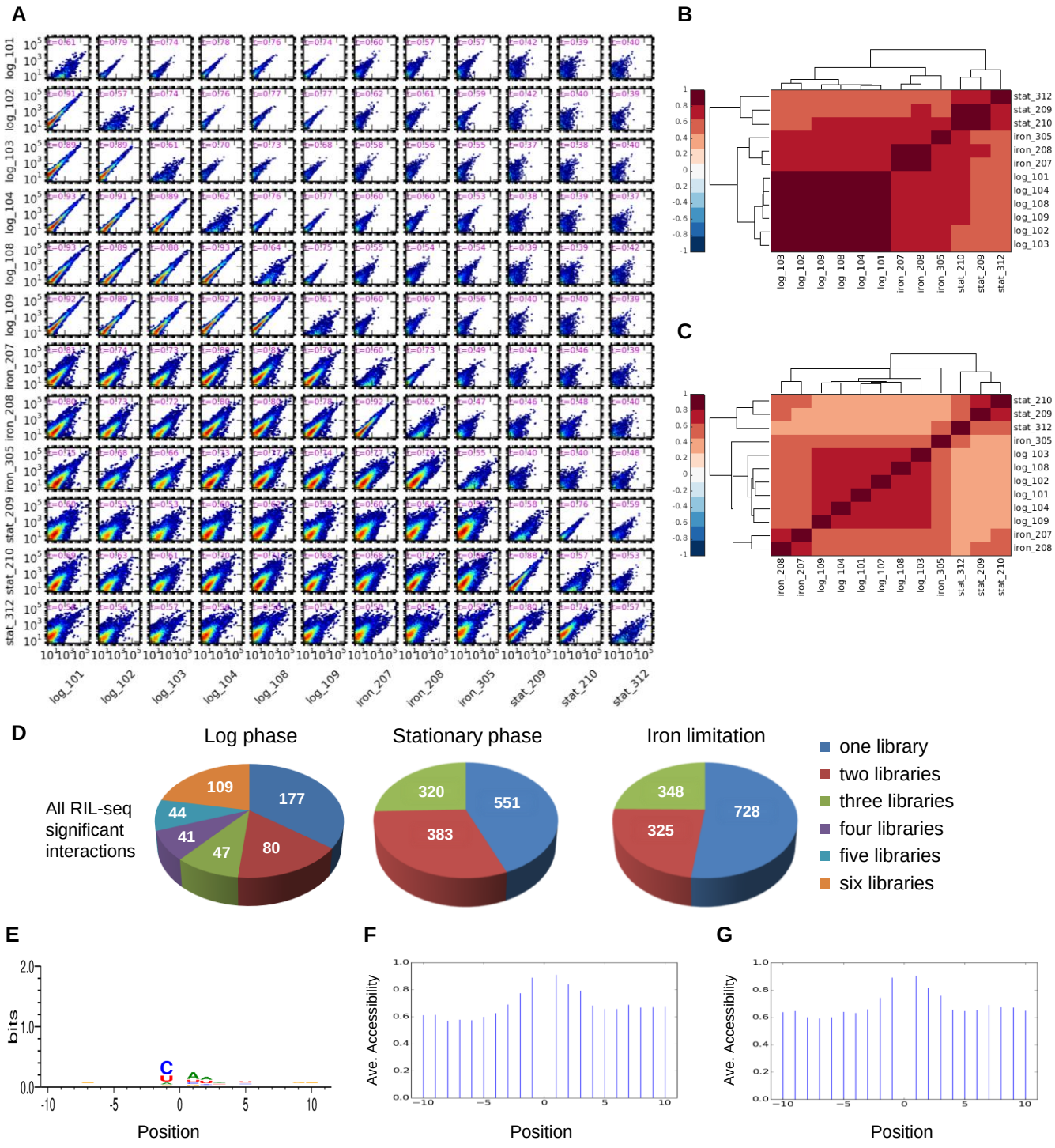


Figure S2 (related to Figure 1). Reproducibility of the data

The reproducibility of the results within same-condition libraries was evaluated at the level of single fragments, chimeric fragments and statistically significant chimeric fragments (S-chimeras). (A) The scatter plots compare the sequenced fragments between two respective libraries. Each point in a scatter plot represents the numbers of fragments

mapped to a 100 nt long region of the genome in two libraries. The intensity of the dots in each figure is scaled from blue (small number of fragments) to red (high number of fragments). Each scatter plot below the diagonal shows these results for single fragments. Each scatter plot above the diagonal shows these results for chimeric fragments. Each plot along the diagonal corresponds to a specific library and shows the number of single fragments mapped to each region of the genome vs. the number of chimeric fragments for which one side of the chimera is mapped to the same region. The Spearman correlation coefficients are reported for each scatter plot. log: logarithmic phase libraries. iron: iron limitation libraries. stat: stationary phase libraries. The name of a library includes the condition and number of library, as listed in Table S1. (B-C) Libraries were clustered by the correlation coefficients they have for the single fragment counts (B) and for the chimeric fragment counts (C) with all other libraries. As is clearly seen, same conditions libraries are in general more similar to one another than libraries of different conditions. (D) Re-discovery of statistically significant chimeras (S-chimeras) in individual libraries. The numbers of S-chimeras re-discovered in n libraries (n=1, 2, 3, 4, 5, 6 for log phase libraries; n=1, 2, 3 for stationary phase and iron limitation libraries) are shown. 48%, 56% and 48% of the S-chimeras were revealed in at least 50% of the libraries of log phase, stationary phase and iron limitation, respectively. (E-G) Assessment of ligation bias. (E) Distribution of nucleotides flanking the ligation point (position 0). A weak motif corresponding to the cleavage preference of RNase A was detected (Raines, 1998). (F) Average probability of positional accessibility around the ligation point. The probability of accessibility was computed using RNAfold (Hofacker, 2003). (G) Average probability of positional accessibility for all possible 4-mer sequences at the 5' end and 3' end of the ligated sequences, showing that on average the relative high accessibility probability at the ends is sequence independent. This plot is based on 1,471,227 sequences (5,747 sequences detected in chimeras in log phase, for which all 4^4 nucleotide combinations at the four last or first positions were tested, i.e. 5747x256 excluding a few sequences that were too short).

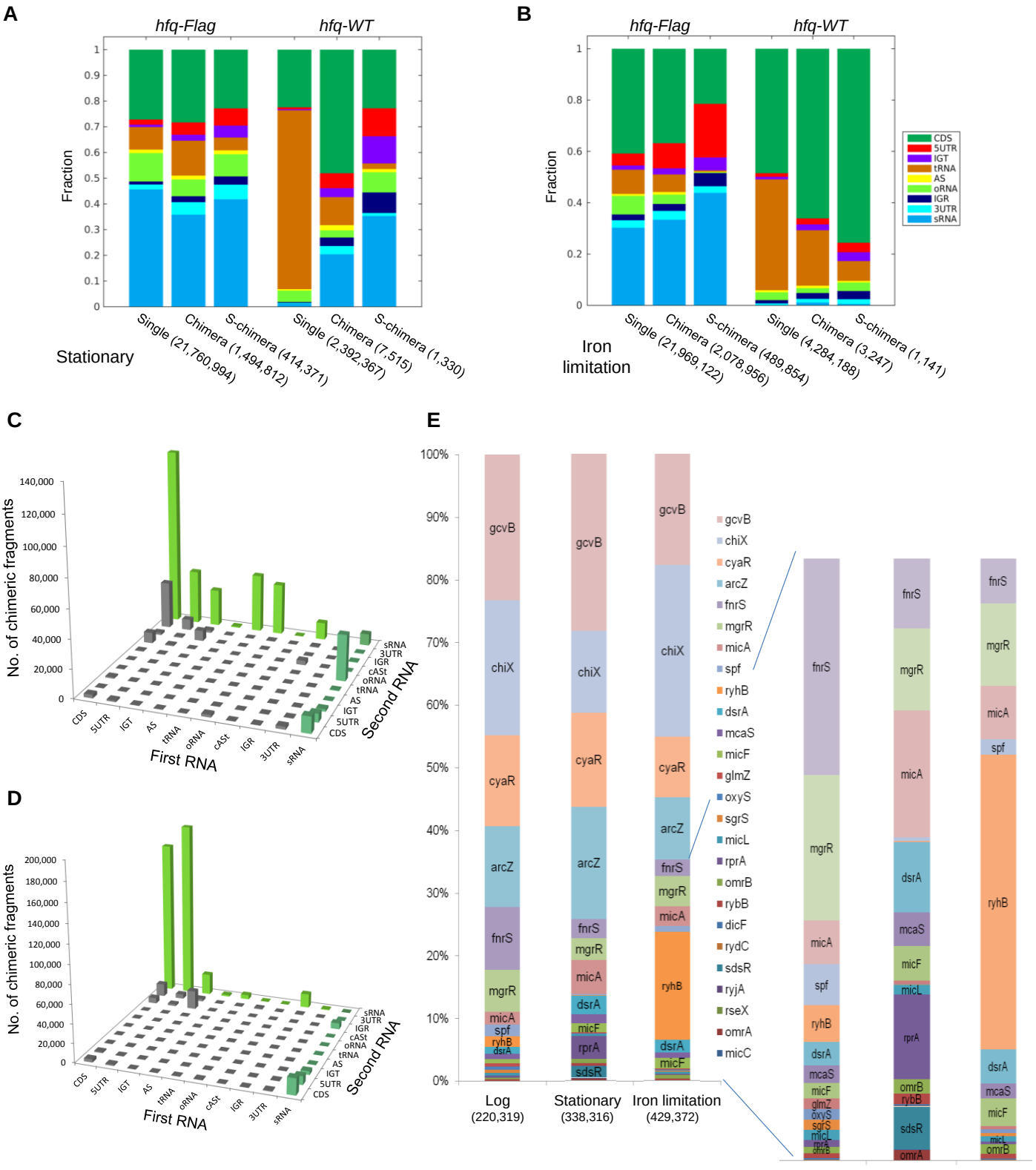

A
hfq-Flag
hfq-WT
Fraction
Stationary
Single (21,760,994)
Chimera (1,494,812)
S-chimera (414,371)
Single (2,392,367)
Chimera (7,515)
S-chimera (1,330)
B
hfq-Flag
hfq-WT
Fraction
Iron limitation
Single (21,969,122)
Chimera (2,078,956)
S-chimera (489,854)
Single (4,284,188)
Chimera (3,247)
S-chimera (1,141)
CDS
5UTR
IGT
tRNA
AS
oRNA
IGR
3UTR
sRNA
C
No. of chimeric fragments
First RNA
Second RNA
D
No. of chimeric fragments
First RNA
Second RNA
E
Log
(220,319)
Stationary
(338,316)
Iron limitation
(429,372)
gcvB
chiX
cyaR
arcZ
fnrS
mgrR
micA
spf
ryhB
dsrA
mcaS
micF
glmZ
oxyS
sgrS
micL
rprA
omrB
rybB
dicF
rydC
sdsR
ryjA
rseX
omrA
micC

Figure S3 (related to Figure 2). Statistical analysis of RNAs by genomic annotation and frequencies

(A-B). Distribution of RNAs in RIL-seq data by their genomic annotations. Single (Single), chimeric (Chimera) and statistically significant chimeric (S-chimera) fragments from the *hfq-Flag* and *hfq-WT* stationary phase (A) and iron limitation (B) libraries were classified into nine major categories based on their mapping annotation: 5UTR (5'UTR), Coding Sequence (CDS), 3UTR (3'UTR), tRNA, sRNA (sRNA with at least one known *trans* target), oRNA (Other small RNAs), AS (AntiSense), IGR (InterGenic Region) and IGT (InterGenic within Transcript). rRNA-derived fragments were filtered out. Fractions are shown (total counts are denoted in parentheses). (C-D) Total number of S-chimera fragments for each combination of genomic elements (the order of RNAs in the chimeric fragment is taken into account) for stationary phase (C) and iron limitation (D) libraries. Mapped IGR fragments were classified as in A with an additional sub-division of the AS category to cASt (*cis* AntiSence with putative *trans* target). Bars representing fragments with sRNAs as the first/second RNA in the chimera are colored in dark/light green, respectively. (E) Distribution of sRNAs. The relative fraction of each sRNA within the S-chimeras is shown in the three growth conditions. sRNAs are ordered by their occurrence frequency in the log phase. The distributions are based on the number of relevant sequenced fragments. Note that MicC has the value of 0 in all conditions as this is the one sRNA that was not captured by RIL-seq

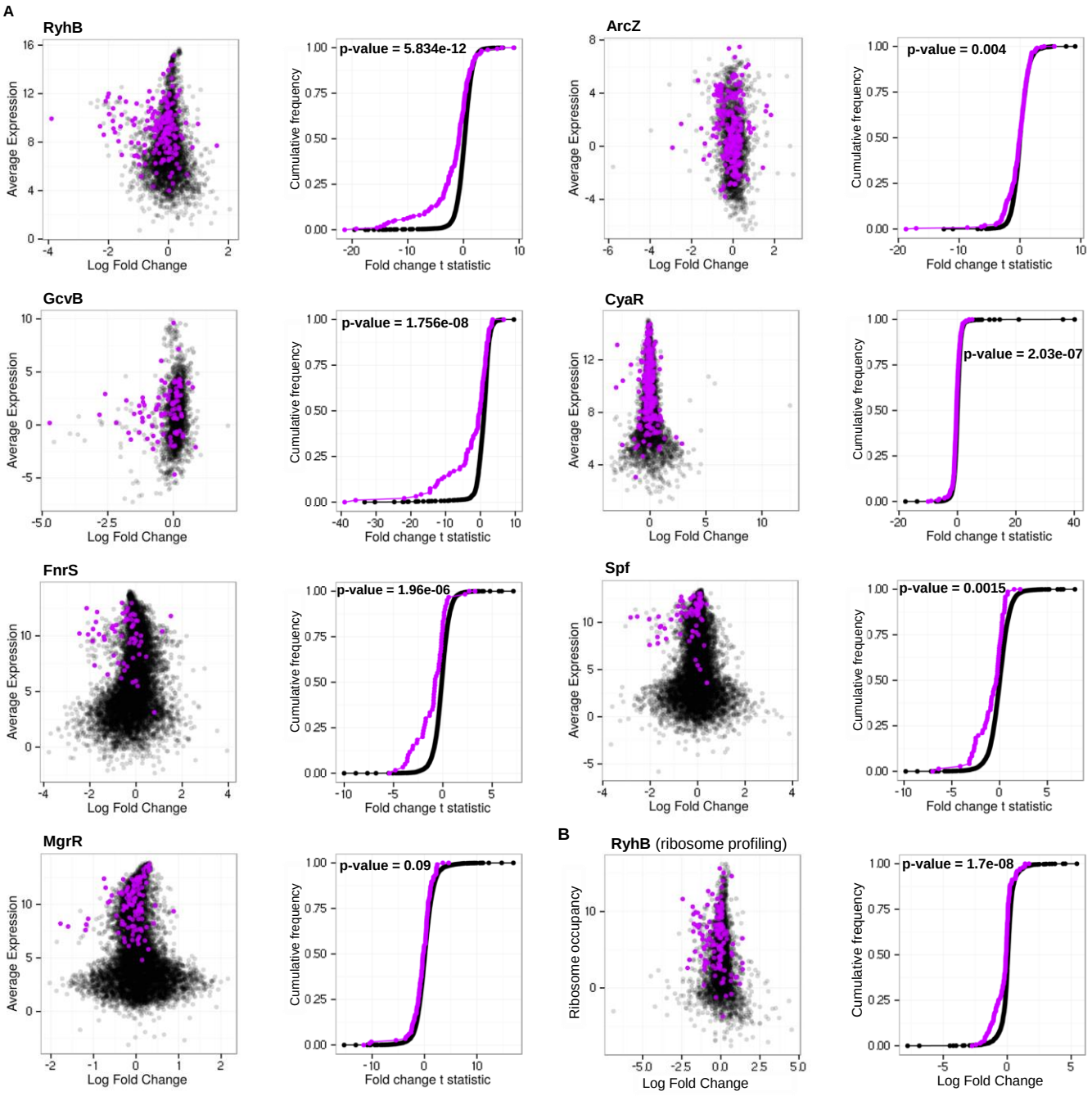


Figure S4 (related to Results in the main text). Impact of the identified interactions on gene expression levels of sRNA bound partners

(A) Microarray experiments were used to assess the impact of the interactions on changes in gene expression of the sRNA bound partners. These experiments provide gene expression data with and without a tested sRNA (induction of the tested sRNA using a plasmid vs. a null plasmid). Such data were available for the *E.coli* RNAs RyhB (Masse et al., 2005), CyaR (De Lay and Gottesman, 2009), FnrS (Durand and Storz, 2010), Spf

(Beisel and Storz, 2011) and MgrR (Moon and Gottesman, 2009). We also used data from experiments conducted in *Salmonella* for the sRNAs GcvB (Sharma et al., 2011) and ArcZ (Papenfort et al., 2009), assuming the sRNAs to have similar effects in *Salmonella* and in *E. coli*. Using these data we could test if an identified set of bound partners of a specific sRNA shows larger changes in gene expression compared to all other genes. The scatter plots display the distribution of the log2 fold change of genes vs. their average expression in the experiment. Genes identified as interacting with the sRNA by RIL-seq are marked in magenta. For each gene in each experiment there is a t-statistics of the difference in gene expression with and without the sRNA (calculated by lmFit of the limma packages). The cumulative distribution plots show the cumulative frequencies of the t-statistics values of the interacting partners in magenta and all other genes in black. The statistical significance of the difference between these two distributions is represented as a p-value using Kolmogorov-Smirnov test. For six out of seven sRNAs the change in gene expression of their targets was statistically significantly larger than that of the rest of genes. (B) The same analysis was performed using change in ribosome occupancy data under change in expression of RyhB (Wang et al., 2015), which was performed for testing changes in translation. Here, the cumulative distribution plot shows the cumulative frequencies of the log2 fold change values in ribosome occupancy of the interacting partners (in magenta) and all other genes (in black), and the statistical test was performed on the log2 fold change values. The difference is statistically significant.

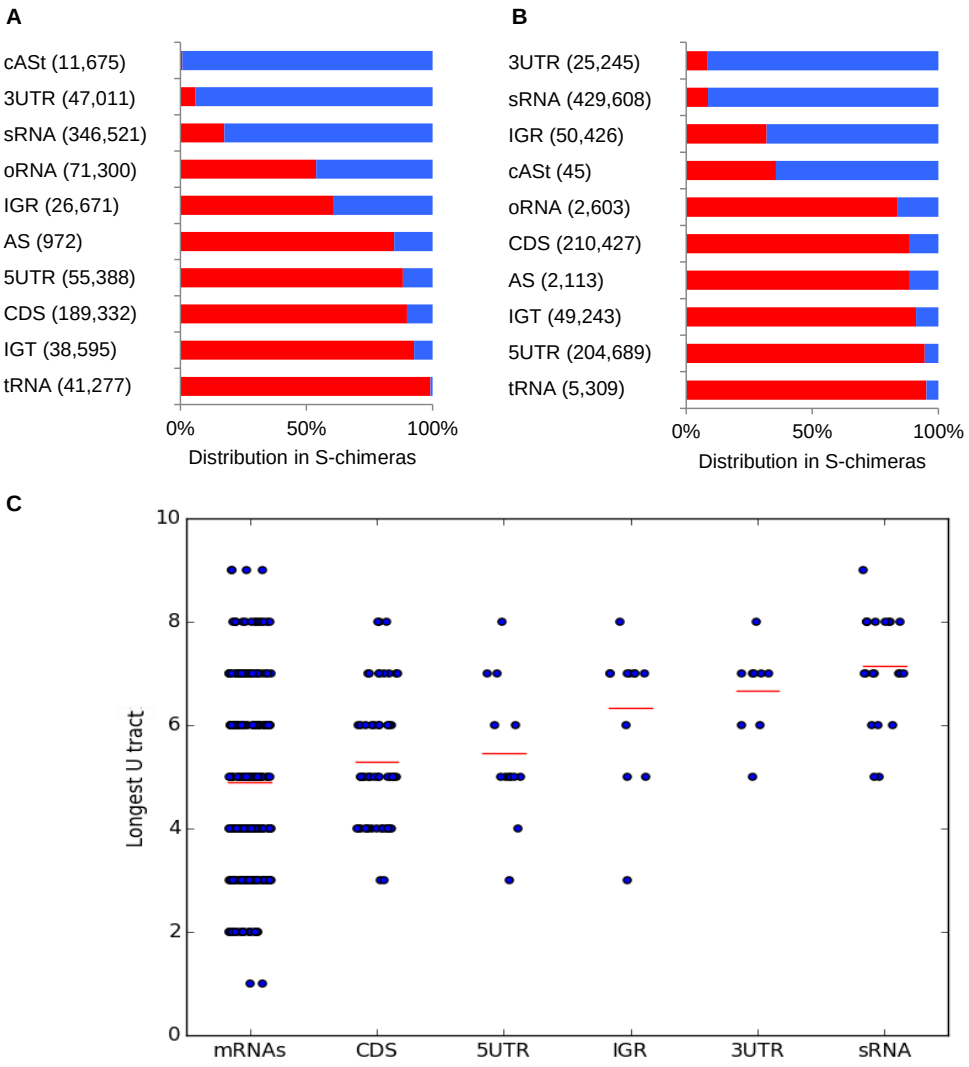


Figure S5 (related to Figure 5). Order of RNAs in the chimeric fragments and analysis of U-tract length

(A-B) Distribution of RNA locations as first (red) and second (blue) in S-chimera fragments for RNAs derived from various genomic elements in stationary phase (A) and iron limitation (B) libraries. Calculation is based on the total number of S-chimera fragments (noted in parentheses). (C) Comparison between the U-tract length of mRNAs and RNAs derived from various genomic elements (included are only RNAs with at least five interactions and mRNAs for which the Rho-independent terminator was annotated in EcoCyc (Keseler et al., 2013). The average U-tract length (indicated by a red line) of sRNAs is 7.14 nt and of putative 3'UTR- and IGR-derived sRNAs is 6.67 nt and 6.3 nt, respectively compared to 4.9 nt in annotated mRNA terminators. p-values computed applying Wilcoxon test with Bonferroni correction: 2.7e-7, 0.005, 0.012, 0.26, and 0.57, for the known sRNAs, 3’UTR, IGR, CDS, and 5’UTR against annotated mRNA terminators).

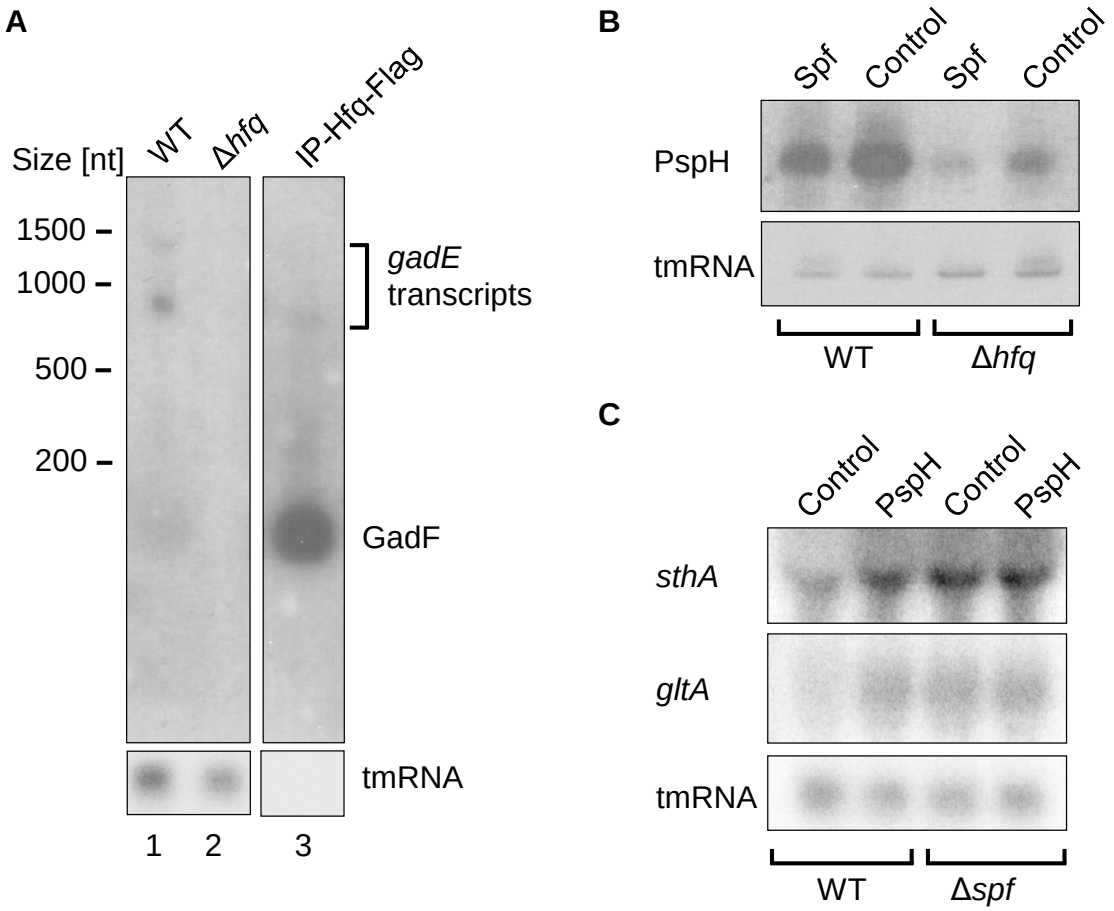


Figure S6 (related to Figure 6). GadF and PspH additional experimental data

Additional experiments testing the novel sRNAs GadF and PspH. (A) Northern blot of *gadE* RNA. Lanes 1 and 2: Total RNA from WT and Δ*hfq* cells grown to stationary phase. Lane 3: RNA co-IPed with anti-Flag antibody from lysate of *hfq-Flag* cells grown to stationary phase. The probe is complementary to GadF central region. (B) Northern blot of PspH in total RNA extracted from log phase WT strain and from a Δ*hfq* mutant, both carrying a Spf overexpressing plasmid or a control plasmid (pJV300). (C) Northern blots of *sthA* and *gltA.* Total RNA was extracted from log phase WT and Δ*spf* cells, either overexpressing PspH or carrying a control plasmid (pJV300).
tmRNA served as a loading control in all three blots.

Supplementary Tables

Table S1. Number of fragments in sequencing libraries, related to Figure 1 (Excel file)

The table describes the different libraries used in the experiment with their protocol parameters and statistics regarding the number of sequenced fragments. The total number of fragments includes the number of fragments available after quality control, adapter removal and DUST filter application. The RIL-seq computational pipeline was used to evaluate the results of each library separately and the results of the unified libraries per condition.

Table S2. RIL-seq RNA pairs identified in unified datasets, related to Figure 2 (Excel file with a different tab for each condition and a tab for a summary table)

The table includes all interactions between two RNAs (minimal number of interactions ≥ 10), which were supported by statistically significant chimeras in the unified datasets. In each condition specific tab, 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. Coordinates are based on the genome of *E. coli K12* MG1655 (NC_000913.2). The table includes data from BioCyCTM pathway/genome database under license from SRI international. **Name:** Common name of the gene (additional information is included in the EcoCyc ID column). **# of chimeric fragments:** Number of chimeras supporting the interaction. **# of libraries:** Number of individual libraries where this interaction was revealed as statistically significant. "U" denotes an interaction that was identified only in a unified library. **Odds Ratio:** (K/L)/(M/N), where K= Number of chimeric fragments of RNA1-RNA2, L=number of other fragments involving RNA2, M=number of other fragments involving RNA1, N=number of all other fragments (that do not involve RNA1 and RNA2). **Normalized Odds Ratio:** Odds Ratio multiplied by the relative enrichments of RNA1 and RNA2 on Hfq. **Fisher's exact test p-value:** p-value for observing at least this number of chimeric fragments given their background frequencies on Hfq. The Odds Ratio provides the effect size of the test. **Free energy of hybridization (kcal/mol):** Hybridization free energy values between the two interacting RNAs computed by RNAup (Muckstein et al., 2006). The sRNA and target sequences

were extracted based on the genome coordinates of the chimeras as documented in Table S2: RNA1 start was taken 20 nucleotides upstream the coordinate termed "RNA1 from" and RNA1 end was taken 20 nucleotides downstream the coordinate termed "RNA1 to". RNA2 start was taken 20 nucleotides upstream the coordinate termed "RNA2 from" and the end was taken at coordinate of "RNA2 to". **Genomic annotation:** e.g. sRNA, CDS, 3'UTR, etc. **Description:** Description of the gene product, taken from RefSeq. **RNA1 from:** Position of the first nucleotide of the most 5' chimera mapped to the first RNA. **RNA1 to:** Position of the first nucleotide of the most 3' chimera mapped to the first RNA. **RNA2 from:** Position of the last nucleotide of the most 5' chimera mapped to the second RNA. **RNA2 to:** Position of the last nucleotide of the most 3' chimera mapped to the second RNA. **Strand:** The genome strand the sequence was mapped to. **Other fragments of RNA1:** Number of fragments in which the first RNA appears as first, including single fragments. **Other fragments of RNA2:** Number of fragments in which the second RNA appears as second, including single fragments. **Total other fragments:** Number of fragments in the experiment excluding the above. **RNA1 in total RNA (# of reads):** The sum of the number of fragments of RNA1 in total RNA libraries. **RNA2 in total RNA (# of reads):** The sum of the number of fragments of RNA2 in total RNA libraries. **RNA1 IP/Total ratio:** (fraction of RNA1 fragments in RIL-seq library) / (fraction of total RNA of RNA1 in total RNA library). **RNA2 IP/Total ratio:** (fraction of RNA2 fragments in RIL-seq library) / (fraction of total RNA of RNA2 in total RNA library). **EcoCyc ID:** The accession number of the gene in EcoCyc database. When the RNA was mapped to a region outside a CDS, the name is followed by 5UTR or 3UTR in case it resides in an annotated UTR (in EcoCyc), EST5UTR or EST3UTR if the UTR is unknown and the interaction is 100 nt upstream or downstream the CDS (or shorter if these regions spanned another transcript or were more likely to be a UTR of the neighbouring transcript). Two gene names and IGR or IGT represent a binding region located between two genes in two different transcription units (IGR) or on the same transcription unit (IGT). AS stands for RNA mapped to the antisense of a gene. **Code:** A unique name assigned to all interactions involving two RNAs, (disregarding the order in chimera and regarding 5'UTR and CDS of a target as one genomic entity). Pairs are

ordered alphabetically. The table is sorted according to RNA2 name and then by the Normalized Odds Ratio.

A summary table of all interactions identified by RIL-seq in Log phase, Stationary phase and Iron limitation is presented in a separate tab. In the summary tab, each pair appears only once (disregarding the order in chimera and regarding 5'UTR and CDS of a target as one genomic entity), where all its interactions per a condition were summed. For interactions with known sRNAs or RIL-seq putative sRNAs the sRNA was placed as RNA2, otherwise RNA1 and RNA2 were ordered alphabetically. The free energy of hybridization is the minimal energy out of the identified conditions. "U" in the 'Libraries' columns denotes an interaction that was identified only when in a unified library.

Table S3. Previously reported sRNA-target interactions recovered by RIL-seq, related to Figure 3 (Excel file)

Known sRNA-target pairs detected in RIL-seq experiment in the three different growth conditions. For six targets the reads could be mapped either to the 5'UTR of the target gene X or to the CDS of the preceding gene Y. While the RIL-seq automatic annotation originally annotated it to the CDS of gene Y, it was manually assigned to gene X (commented in the table). Of note, there are two known interactions that were each revealed in the individual libraries (MicA-*ompW* in stationary phase and RybB-*fimA* in iron limitation), but were not re-discovered in the unified dataset of all interactions per condition, and therefore they are not listed in Table S3 as identified by RIL-seq.

Table S4. Common motifs identified in sRNA target sequence sets, related to Figure 4A-D (Excel file)

The table contains ’Complementarity’ and ‘Common motifs’ tabs in excel file:

’Complementarity’ tab - Common sequence motifs in RIL-seq target sets are complementary to binding sites on known and putative sRNAs. Shown for: sRNAs with known binding sites except for MgrR for which no binding site was yet reported and a putative site was revealed by this analysis. Separated by a blank row, putative additional sRNAs derived from different genomic elements are shown. Sequences with the motif (m) - The number of target sequences that shared the common motif; Total number of

targets (n) - The total number of sequences in the target set of the respective sRNA; MEME e-value - The E-value calculated by MEME (Bailey et al., 2009); MAST p-value - The p-value calculated by MAST (Bailey and Gribskov, 1998); Base-pairing of the sRNA and the common target motif - The sequence of the respective sRNA is presented in a complementary orientation below each motif, and known binding sites are marked in green. A vertical line was drawn if at least one of the motif nucleotides matched the corresponding position on the sRNA sequence Note that uhpT.3UTR was identified recently as an independent RNA (Fitzgerald et al., 2014) and that cpxP.3UTR was characterized most recently (Chao and Vogel, 2016; Grabowicz et al., 2016). Logos were created using weblogo 3.4 (Crooks et al., 2004).

'Common motifs' tab - Detailed information of the common motifs. MEME was used to test if the targets of a known sRNA or of a putative sRNA have a common sequence motif. The motif was further tested to see if it matches the reverse complement of the sRNA. Target sets where n≥4 across all conditions or in each condition separately were analysed. For each target set (in one condition or across all conditions) we ran the analyses using either the subset of targets that appeared first in the chimera, or the subset of targets that appeared second in the chimera, or using the full set of targets, appearing both first and second in the chimera. Only the best MEME motif for every condition is shown. A condition marked with * indicates that a different MEME motif was chosen for Figure 4A and for the motif visualization in the complementarity tab. The 'Matches known binding site' column shows motifs that complemented the known binding site with a MAST p-value ≤0.0001. NMH stands for "No Motif Hit" and is written when the MAST p-value is >0.01. Of note, Rho-independent terminator motifs of the sRNAs were identified as common motifs when a set of sRNAs and/or putative sRNAs interacted with the same target. Detailed explanation about the table content appears in a Legend for 'Common motifs' tab of the excel table.

Table S5. **Support of RIL-seq protocol by GcvB experiment, related to Figure 4E** (Excel file)

GcvB targets identified in both the unified log phase libraries and in libraries of Δ*gcvB* strain with WT *gcvB* or *gcvB*ΔR1 plasmid are shown. Targets of GcvB known

to interact through R1 (green) are revealed only with the WT plasmid, while targets known to interact through R2 (blue) are revealed in both WT *gcvB* and *gcvB*ΔR1, implying RIL-seq captures true interactions.

Table S6. Summary of genomic elements identified by RIL-seq, related to Figure 5A (Excel file)

The table lists all RNAs that were included in the statistically significant chimeras (listed in Table S2). Each RNA has a single entry disregarding its position in the chimera (first or second RNA). When counting the number of interacting partners of a RNA, 5'UTR and CDS of a target were regarded as one genomic entity, the growth condition it was revealed in or its specific genomic coordinate. The information displayed for each RNA component is detailed in the 'legend' sheet within the table.

Table S7. Predicted base-pairing between sRNAs and their putative targets, related to Figure 6A (Excel file)

Base-pairing of putative pairs of sRNAs and their targets that were tested using GFP fusions (Fig. 6A) were predicted using RNAup (Muckstein et al., 2006). The sRNA and target sequences were extracted based on the gemone coordinates of the chimeras as documented in Table S2: RNA1 start was taken 50 nucleotides upstream the coordinate termed "RNA1 from" and RNA1 end was taken 50 nucleotides downstream the coordinate termed "RNA1 to". RNA2 start was taken 50 nucleotides upstream the coordinate termed "RNA2 from" and the end was taken at the coordinate of "RNA2 to". The sRNA (bottom) and its targets (top) are oriented 3’->5’ and 5’->3’, respectively. The RNAup predicted paired regions are marked in red and their coordinates are presented in red in the table. The free energy values reported here differ from the ones reported in Table S2 because there the computations were done with padding of 20 nucleotides at the ends of chimeras. Coordinates are based on the genome of *E. coli K12* MG1655 (NC_000913.2).

Table S8. Strains, deoxyribonucleotides and plasmids used in this study, related to Experimental Procedures (Excel file)

Supplemental Experimental Procedures

Strains and media

The bacterial strains used in this study are listed in Table S8. The *E. coli* MG1655 or TOP10 strains served as wild-type. Strain TM615 carrying *hfq-Flag* (Morita et al., 2005) was kindly provided by H. Aiba (Nagoya University, Japan). Strain HM33 was constructed by transferring *hfq-Flag* from Strain TM615 into MG1655 using P1 transduction. Plasmid pCP20 (Cherepanov and Wackernagel, 1995) was used to eliminate the chromosomal *cat* gene of strain HM33, resulting in the formation of strain HM34. Strain HM43 was constructed by transferring an *hfq::kn* allele from strain JW4130 of the Keio collection (Baba et al., 2006) into MG1655 using P1 transduction. Strain HM54 was constructed from its parent strain HM34 by the one-step inactivation method (Datsenko and Wanner, 2000) using oligonucleotides 234/235 and pKD3 plasmid. Strain HM55 was constructed by transferring the *hfq::kn* allele of JW4130 strain into strain TOP10 by linear transformation using oligonucleotides 26/27. Strains HM56 and HM57 were constructed by the one-step inactivation method (Datsenko and Wanner, 2000) using oligonucleotides 280/281 and 338/339, respectively, and pKD3 plasmid. To create strains HM64 and HM65, PCR fragments that contained the desired mutations in *spf* and a *cat* gene downstream to *spf* were synthesized, using oligonucleotides 361/363 (HM64) or 362/363 (HM65) and pKD3 plasmid. The PCR fragments were transformed into MG1655 by linear transformation. *S. cerevisiae* BY4742 was kindly provided by O. Pines (The Hebrew university of Jerusalem, Israel). Bacterial strains were routinely grown at 37 °C in LB medium (Bertani, 2004). Ampicillin (100 µg/ml), kanamycin (45 µg/ml), chloramphenicol (25 µg/ml) or 2,2'-Dipyridyl (200 µM) were added where appropriate. The oligonucleotides used for strain constructions are listed in Table S8.

RIL-seq protocol

Experimental procedure

E .coli Strains MG1655 and HM34 were grown overnight in LB at 37 °C with shaking (200 r.p.m.), diluted 100-fold in fresh LB, and re-grown with shaking at 37 °C to

log phase (OD_{600} = 0.5), with or without the addition of the iron chelator 2,2'-Dipyridyl (200 µM, 30 min), or to stationary phase (6 hr). A culture volume corresponding to 40 OD_{600} per ml was centrifuged and the pelleted cells were washed twice and resuspended in 10 ml of ice-cold phosphate-buffered saline (PBS) (pH 7.4). Next, cells were exposed to 800 mJ of 254 nm UV irradiation (Stratalinker® UV Crosslinker 1800) on a metal block cooled to (-20) ^{o}C and then mechanically lysed in a total of 1200 µl Lysis Buffer (50 mM Sodium Phosphate, 300 mM NaCl, 0.1% IGEPAL, 10 mM Imidazole, 0.1 unit/µl Recombinant RNase inhibitor (Takara) and EDTA free protease inhibitor cocktail set III (Calbiochem) diluted 1:200, pH 8.0), using 400 µl glass beads in a Retch MM400 mixer. Cell lysates were cleared by centrifugation and incubated for 1 hr with 20 µl of magnetic beads (Novus Biologicals) in order to reduce the nonspecific binding in the next step. Thereafter the cell lysates were incubated for 1.5 hr with 20 µl of magnetic beads carrying M2 anti-Flag monoclonal antibody (3 µg; Sigma) at 4 °C. As a control, the lysate was incubated with magnetic beads without M2 anti-Flag monoclonal antibody. The lysate was removed and the beads were washed five times with 200 µl lysis buffer at 4 °C. To trim the exposed parts of the RNAs, beads were resuspended in 200 µl of lysis buffer supplied with 0.5 unit of RNase A/T1 mix (Thermo Scientific) without RNase inhibitor, and incubated for 3-15 min at 22 °C or 37 °C. RNase digestion buffer was removed and the beads were washed three times with 200 µl lysis buffer supplied with 0.1 U of SUPERase IN^{TM} RNase Inhibitor (Life technologies) at 4 ^{o}C. 5'OH ends were phosphorylated and 2'P / 3'P ends were dephopsphorylated by incubation with 40 units of T4 PNK (New England Biolabs) in 80 µl of PNK buffer (Supplied by the manufacturer) for 2 hr at 22 °C. PNK reaction mixture was removed and the beads were washed three times with 200 µl lysis buffer at 4 ^{o}C. Neighbouring RNAs were ligated overnight using 216 units of T4 RNA Ligase 1 (New England Biolabs) in 80 µl ligase buffer (Supplied by the manufacturer). Ligation mixture was removed and the beads were washed three times with 200 µl lysis buffer at 4 ^{o}C. To release the RNA from Hfq, The beads were treated with Proteinase K (100 µg; Thermo Scientific) in 300 µl Proteinase K buffer (50 mM Tris HCl [pH=7.8], 50 mM NaCl, 0.1% IGEPAL, 10 mM Imidazole, 1% SDS, 5 mM EDTA, 5 mM β-mercaptoethanol) supplied with 0.1 unit/ul RNase inhibitor for 2 hr at 55 °C. RNA was extracted according to the standard LS-TriReagent protocol (Sigma). RNA

concentration was analysed using Agilent Bioanalyzer (Agilent Technologies) and RNA-seq libraries were constructed based on RNAtag-Seq protocol (Shishkin et al., 2015) with few modifications. Briefly, RNA was subjected to fragmentation, alkaline phosphatase and DNase treatment. RNA was cleaned-up by RNA Clean and Concentrator 5 kit (Zymo) and ligated to 3' barcoded adaptors (Table S8). Next, the ligation mixtures were pooled together and the RNA was purified with RNA Clean and Concentrator 5 kit. rRNA was removed using RiboZero (Illumina), and rRNA depleted samples were purified using 2.5X RNAClean XP Beads (Beckman Coulter) and 1.5X Isopropanol. First cDNA strand was generated using SuperScript III (Life Technologies), RNA was degraded and samples were purified using 2.5X AMPure XP Beads (Beckman Coulter) and 1.5X Isopropanol. A second adapter was ligated to the cDNAs 3' end, and the cDNA was cleaned-up twice with 2.5X AMPure XP Beads and 1.5X Isopropanol. Libraries were PCR amplified by HiFi HotStart DNA polymerase (KAPA) using Illumina compatible primers and DNA was purified with 1.5X AMPure XP Beads. Libraries of total RNA were constructed likewise. The libraries were sequenced by paired-end sequencing using Nextseq500 Sequencer (Illumina). It is of note that we applied a few other variations of the experimental protocol which included: replacement of A/T1 RNases with S1 Nuclease, addition of TAP enzyme, or replacement of T4 Ligase 1 with CircLigase™ II. However, as these attempts resulted in lower quality libraries, they are not detailed above. In addition, to estimate the potential bias of *in vitro* RNA interactions on Hfq (Mili and Steitz, 2004), the *E. coli* lysate was mixed with a *Saccharomyces cerevisiae* lysate before the Co-IP. The chimeric fragments containing sequences from both *E. coli* and *S. cerevisiae* comprised <1% of all S-chimera fragments, suggesting that the majority of RIL-seq data involve in-vivo interactions..

Computational procedure

Processing of sequencing data

Raw reads were split into their library of origin using the barcode sequences in the beginning of first read. The paired reads were then processed by cutadapt (Martin, 2011) to remove adapter sequences and low quality ends (less than Q15), eliminating reads shorter than 25 nucleotides (nt) on either mate. The fragments were mapped to the

genome of *E. coli* K12 MG1655 (RefSeq accession number NC_000913.2) using bwa aln followed by bwa sampe (Li, 2013). This mapping was used in later steps of the analysis to filter inconclusive mappings. In order to detect chimeric fragments, the first 25 nucleotides of each mate were mapped to the genome. Prior to mapping to the genome, these short reads were filtered using the DUST algorithm with a threshold of 10 to remove low complexity reads. The 25 nt ends were mapped to the genome of *E. coli* using bwa aln allowing three mismatches, followed by bwa samse.

We first checked whether the two mates of a sequenced fragment were mapped to the same transcript, as annotated in EcoCyc version 19.0 (http://ecocyc.org) (Keseler et al., 2013) or within a distance of 1,000 nt, mapped concordantly or in reverse order (supporting a circular RNA). If the mates mapped concordantly, the fragment was declared as a "single" fragment. If the fragment was not declared as "single" and mapping of each mate had at most one mismatch, the fragment was declared as "chimeric". In case either read was mapped to more than one location in the genome, we analysed all combinations of mapped positions. If any combination fits the definition of "single" above, it was declared as such; otherwise, it was declared as chimeric arbitrarily, using the first mapping of each read to the genome. Fragments mapped to rRNAs, either as single fragments or chimeric fragments were removed from the analysis. Fragments mapped concordantly as paired-end reads in the previous steps and were not defined as “single” in the later step were omitted as well.

Identifying over-represented RNA pairs in chimeric fragments

We divided the genome into non-overlapping 100 nt-long windows and counted the number of sequenced fragments showing fusion between each pair of windows X and Y. Window pairs (X,Y) that appeared in more than five sequenced fragments were further analysed. We applied Fisher's exact test to the count of pair (X,Y) in sequenced fragments, testing if it appears more than expected at random based on the count of X in other chimeric fragments and in single fragments, the count of Y in other chimeric fragments and in single fragments, and the count of all other windows (single and chimeric). Since we observed that certain RNAs are found preferentially in the 5' end of the chimeric fragment (acceptor side of the ligation) or the 3' end (donor side), each pair

of windows was tested twice, testing either window as the 5' end and the 3' end of the chimeric fragments. Identification of chimeric fragments of windows (X,Y) whose count is statistically significantly above the expected at random determined X and Y RNAs as interacting via Hfq (S-chimera).

The neighbouring windows of windows (X,Y) might also appear in chimeric fragments, reinforcing the inferred interaction between X and Y. Thus, we increased the size of tested windows gradually in leaps of 100 nt up to 500 nt. All the combinations of windows of both putative interacting regions were tested (from length of 100 to 500 nt and from four windows upstream each original window to four windows downstream). The p-value of each such test was computed using the above Fisher's exact test and the combination of windows with the lowest p-value were considered the interacting regions. The process of finding interacting regions starts with the pairs of windows (X,Y) having the highest number of chimeric fragments. We applied to the p-value Bonferroni correction for multiple hypotheses (taking into account number of alternative window sizes and positions). Another correction was done afterwards by multiplying the p-value by the number of window pairs for which the count of chimeric fragments was ≥ a pre-determined threshold (10 in this study). Along with the p-value, the Odds Ratio was computed, as a measure of the effect size of the Fisher's exact test. The Odds Ratio is computed as (K/L)/(M/N), where K= Number of chimeric fragments of RNA1-RNA2, L=number of other fragments involving RNA2, M=number of other fragments involving RNA1, N=number of all other fragments (that do not involve RNA1 and RNA2). L, M, and N include chimeric and single fragments. Representation of RIL-seq single and chimeric fragments in UCSC genome browser (Kent et al., 2002) can be found in the following links: RIL-seq single fragments, RIL-seq chimeric fragments (Note: chimeras with ≥5 sequenced fragments are presented but only chimeras with ≥10 sequenced fragments were included in all other analyses in the paper. Chimera tracks are hidden by default. Network views in Figure 7 were drawn in by circos software (http://circos.ca/).

Computing the Normalized Odds Ratio

We computed a measure that 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 (IP/total of RNA1 and by IP/total of RNA2). In the computation we defined the maximal IP/total as 1 (maximal level of RNA bound to Hfq is the level of total RNA) and computed all other values of IP/total relative to it. To avoid outlier we took the lowest value in the top 1 percentile as the maximum. The top 1 percentile was computed over all windows covering the transcriptome. The normalization factor is the same for the entire data. Let the normalization factor be maxIP_T, then the Normalized Odds Ratio is computed by (Odds Ratio)*[(IP1/Total1)/maxIP_T]*[(IP2/Total2)/maxIP_T], where IP1 and Total1 are the IP values and total RNA values for RNA1 in the pair, and likewise for RNA2. The total RNA and IP values were extracted from the relevant RNA-seq data following the genome coordinates annotated for RNA pairs in RIL-seq chimeras. For the unified libraries we summed the values of the individual libraries.

Annotation of RNAs

Genome annotation was based on EcoCyc version 19.0 (http://ecocyc.org) (Keseler et al., 2013). If the 5' and 3' ends of a protein-coding gene were annotated in EcoCyc we used these annotations to determine the 5'UTR and 3'UTR (termed 5UTR and 3UTR in figures and tables). If the transcription start site or transcription termination site of a gene were unknown, the UTRs were considered the regions 100 nt upstream the ATG and downstream the stop codon (or shorter if these regions spanned another transcript or were more likely to be a UTR of the neighbouring transcript). In such cases the UTRs were termed EST3UTR/EST5UTR for estimated UTRs. A coding region within a mRNA was termed CDS. Intergenic regions were termed IGR if their boundary genes were not in the same transcription unit or IGT if the two boundary genes shared a transcript. In addition we termed regions encoding non-coding RNAs as following: tRNA, rRNA, sRNA - a sRNA that exerts its regulation via base-pairing and has at least one known *trans* target. Regions antisense to genes or to IGT were termed AS and AS_IGT respectively. If the AS was annotated and known or suspected to base-pair with its sense gene or near the sense gene it was termed cAS. The remaining genomic regions encoding non-coding RNAs were all termed other-RNA (oRNA). In several analyses we used sub-division of the AS types based on information revealed by RIL-seq. These

include ASt, which are unannotated AS with a putative *trans* target, and cASt, which are cAS with an additional putative trans target. A word of caution regarding the annotation: since the annotation is done automatically and uses rigid rules to assign each RNA a single annotation, there are inevitably some mis-annotations, especially in boundary regions and even more so for neighbouring genes on the same operon. Nevertheless, in all large-scale analyses of the paper and in all the tables and figures we kept the original automatic annotation and did not change it according to manual curation unless specifically specified.

GcvB comparative RIL-seq

RIL-seq experiments were performed in log phase as described, using *E. coli* Δ*gcvB* strain (HM54) transformed with a WT *gcvB* or a *gcvB*ΔR1 plasmid, each repeated three times, resulting in three sequencing libraries per strain. Mapping of the sequenced reads was done to the chromosome and to the relevant plasmid. Since the R1 region of *gcvB* gene was deleted, we replaced the missing nucleotides in *gcvB* in *E. coli* genome sequence with "N". Mapped reads in the three libraries per strain were unified, and the two unified data sets of WT *gcvB* or *gcvB*ΔR1 were compared. For comparing the number of corresponding interactions (GcvB with same target) between the two strains, the numbers of chimeric fragments corresponding to an interaction (Table S5) were normalized by the total number of sequenced fragments in the respective unified library.

Plasmids

Construction of sRNA over-expressing plasmids was done according to (Urban and Vogel, 2009) using the pZE12-luc as the scaffold plasmid. Construction of GFP-fusion plasmids was done essentially as described in (Urban and Vogel, 2009), using the pXG10-SF (Corcoran et al., 2012) as a backbone. Briefly, 5' regions of target genes, including the region captured in the chimeric fragments, were PCR amplified, digested with Mph1103I and NheI and cloned into pXG10-SF digested with the same restriction enzymes. Mutagenesis of pZE-PspH was done by the QuikChange Lightning Site-Directed Mutagenesis Kit (Stratagene), using oligonucleotides 350/351 and 348/349 to create pZE12-PspH$_{G100}$ and pZE12-PspH$_{GGG,}$ respectively. Deletion of RI region from

pJU-014 was done by PCR amplification with oligonucleotides 232/233, using pJU-014 as a template. The PCR product was DpnI treated and self-ligated to create pGcvBΔR1. The plasmids used in this study and the oligonucleotides used for plasmid constructions are listed in Table S8.

GFP reporter assay

The GFP reporter assay was done essentially as described previously (Corcoran et al., 2012; Urban and Vogel, 2009). Wild type (TOP10) or Δ*hfq* (HM55) cells were transformed with a target-GFP reporter plasmid and with a sRNA over-expressing plasmid. Control plasmids were a non-GFP plasmid (pXG0; Urban and Vogel, 2007) and sRNA control plasmids (pJV300 or pTP011; Urban and Vogel, 2007). Single colonies were grown overnight at 30 °C in LB supplemented with Ampicillin and chloramphenicol. The cultures were diluted 1:100 in fresh medium and grown at 30 °C to OD_{600} of 0.3. 1 ml of each culture was centrifuged and the pellet was resuspended in 300 µl of 1X PBS. Fluorescence was measured using the BD Accuri™ flow cytometer. Regulation level was calculated by subtracting the auto-fluorescence and then calculating the ratio between the fluorescence level of a strain carrying the sRNA over-expressing plasmid and a strain carrying the control plasmid. Three to six biological repeats were prepared for every sample.

Immunoprecipitation (IP) and Western blotting

Lysate of cells grown to log or stationary phase were prepared as described for the RIL-seq procedure, and incubated for 1.5 hr with magnetic beads (Novus Biologicals) carrying M2 anti-Flag monoclonal antibody (3 µg; Sigma) in Lysis buffer at 4 °C. As a control, we incubated *hfq-Flag* lysate, without crosslinking, with magnetic beads carrying no M2 Anti-Flag antibody. The lysate was removed and the beads were washed five times with 200 µl lysis buffer at 4 °C. Proteins were eluted with Elution buffer (50 mM Sodium Phosphate, 300 mM NaCl, 300 mM Imidazole, 0.1 unit/ul Recombinant RNase inhibitor (Takara) and EDTA free protease inhibitor cocktail set III (Calbiochem) diluted 1:200, pH 8.0). Samples were either subjected to RNA extraction for *gadE* and *gadF* Northern blot or heated at 95 °C for 5 min and subjected to a 4%-20% polyacrylamide

SDS gel electrophoresis followed by electrotransfer onto a nitrocellulose membrane (Pall). The membrane was probed with M2 anti-Flag monoclonal antibody (Sigma) and then with anti-Mouse secondary antibodies (Jackson ImmunoResearch). Signals were visualized by the ECL system (Amersham).

Northern analysis

Overnight cultures were diluted 1/100 in fresh medium and grown to logarithmic phase (OD_{600}=0.3) or to stationary phase (6 hr growth). Cells were centrifuged at 4 °C and resuspended in 50 µl 10 mM Tris-HCl (pH 7.5) containing 1 mM EDTA. Lysosyme was added to 0.5 mg/ml, and samples were subjected to three freeze-thaw cycles. RNA was extracted using TRI Reagent (Sigma). For analyses of *sthA* and *gltA*, RNA samples (15 µg) were denatured for 10 min at 65 °C in RNA loading buffer containing 40% formamide and 1.3 M formaldehyde, and then separated on 1.4% agarose gel containing 1.17 M formaldehyde in 20 mM MOPS, 2 mM Na-Citrate and 1 mM EDTA pH 8.0. RNA was transferred to Zeta-Probe membrane (Bio Rad) by capillary transfer in 10X SSC (1.5 M NaCl and 150 mM Na-citrate). For *sthA* probing, a specific [^{32}P] end labelled probe (386; (Beisel and Storz, 2011)) was used, while for *gltA* probing a specific riboprobe was used. For synthesis of *gltA* riboprobe, a template containing the T7 promoter upstream to *gltA* antisense were PCR-amplified using oligonucleotides 390/391. *In vitro* RNA transcription of the riboprobe was done essentially as described (Argaman and Altuvia, 2000), using 600 ng of the PCR template, 0.5 mM of each CTP, GTP and UTP, 20 µM ATP and 30 µCi [^{32}P]αATP. For the analysis of GadF, Spf and PspH, equal amounts of RNA samples (10-30 µg) were denatured for 10 min at 65 °C in loading buffer containing 65% formamide, separated on 7 M urea/6% polyacrylamide gels in 44.5 M Tris-base, 44.5 M Boric acid and 2 mM EDTA pH 8.0, and transferred to Zeta-Probe membrane by electroblotting. The membranes were hybridized with specific [^{32}P] end labelled DNA probes. The probes sequences are listed in Table S8.

Computational analysis

Assessing bias in ligation

To test potential ligation biases in RIL-seq data we searched for chimeric fragments in which the ligation point can be deduced from the sequence. These sequences were aligned using the ligation point as an anchor and the nucleotide distribution, as well as the probability of positional accessibility in the regions flanking the ligation point were computed. RNAfold –p was used for the latter (Hofacker, 2003). To assess the sequence dependency of the accessibility probability at the ends of RNA sequences, we generated many combination of four nucleotides at the four positions at the 5' and 3' of the ligated sequences and folded each such sequence as well. This analysis was based on 1,471,227 sequences (5,747 sequences detected in chimeras in log phase, for which all 4^4 nucleotide combinations at the four last or first positions were tested, i.e. 5747x256 excluding a few sequences that were too short).

Reproducibility of the Data

Spearman correlation of mapped fragment counts between libraries

For each sequenced library the sequenced fragments were mapped to 100 nt-long windows along the genome, as described above. Thus, for each library each window was assigned the count of single fragments mapped to it, and the Spearman correlation coefficient of window counts could then be computed between each pair of libraries. Next, for each library we counted the number of times each pair of windows appeared in chimeric fragments (regardless of the order of RNAs in the chimera), and computed the Spearman correlation between libraries. Another correlation was computed for each library between the number of single fragments mapped to each window and the number of chimeric fragments for which one side was mapped to the window.

Clustering the libraries

For each library we generated a vector with the correlation coefficients it had for the single fragment counts with all other libraries. We clustered the libraries by these vectors of correlation coefficients (Matlab clustergram). Likewise, for each library we

generated a vector with the correlation coefficients it had for the chimeric fragment counts with all other libraries and clustered the libraries by these vectors.

Assessing the reproducibility of statistically significant chimeras

For each library we listed the chimeras that passed the statistical significance (S-chimeras). We counted in how many libraries a chimera corresponding to RNA pair X-Y (disregarding the order in the chimera) passed the statistical significance. RNAs mapped to either the 5'UTR or CDS of a gene were considered equivalently.

Analysis of genomic elements encoding RNAs revealed by RIL-seq

For each genomic element (e.g. CDS, 3'UTR) we computed its fraction among the mapped single and chimeric fragments, as well as among the statistically significant chimeric fragments. The fraction of genomic element X among chimeric fragments (RNA1, RNA2) was computed as follows: (total number of fragments mapped to genomic region of type X, considering both RNA1 and RNA2)/ (2N), where N is the number of total chimeric fragments. In case of single fragments, the reads from the two paired ends are mapped to the same locus, and hence, obviously, to the same genomic region. For consistency, the computation is done as above.

Detection of known targets in RIL-seq chimeras

We compiled an updated dataset of 154 experimentally determined sRNA-target interactions based on our previous compilations (Peer and Margalit, 2011, 2014) and current information in EcoCyc version 19.0 0 (http://ecocyc.org) (Keseler et al., 2013). Exceptionally, *mutS*-SdsR interaction, which was revealed only *in vitro*, was included (because it was the only reported target of SdsR). We considered a known sRNA-target pair as detected by RIL-seq if there was at least one statistically significant inferred interaction that included the sRNA and target, where mapping to the target was either at its CDS, 5'UTR or part of a an operon transcript annotated in the RIL-seq as IGT.

Recognition of binding motifs

Motifs were searched in each sRNA target sequence set by MEME (Bailey et al., 2009), using the "zero or one per sequence" option and allowing motif width to range from 6 to 15 nucleotides. The target sequence sets were defined by different parameters: experimental condition (log phase, stationary phase, iron limitations and all conditions together), location within the chimeric fragments (first in the chimera, second and the chimera and regardless of chimera position, named both). Sets were defined by all combinations of these parameters. The search for a common motif was applied to each sRNA target set that included at least four targets in at least one of the experimental conditions. The sequences to which the search was applied were extracted using the mapping of the chimeric fragments supporting the interaction and overlapping fragments were concatenated. These sequences were then padded by 50 nt from each side. To verify that an identified motif matches a complementary binding site on the respective sRNA, we searched for it in the reverse-complement sequence of the sRNA using MAST (Bailey and Gribskov, 1998). We submitted to MAST the motifs found by MEME with an E-value $\leq$10 and ran it with "sep", so that the search will be done only on the submitted strand. Sites identified by MAST with a p-value threshold of 0.01 or less were considered as hits, and further analyzed. While we usually selected for further analysis the best motif based on the MEME e-value and MAST p-value, in some cases we applied heuristic considerations. For example, if the best motif did not overlap the known binding site of the sRNA and there was a second best motif that overlapped it, it was selected. The MEME e-value, MAST p-value, MAST score, and the quality of match between the sRNA and target were also considered for the selection of motifs for further analysis.

In addition, we ran the analysis automatically for all other RNAs included in RIL-seq data, given they had at least four interactors in one of the conditions. This enabled the support of putative sRNAs derived from various genomic elements, by identifying common motifs in the sequences of their bound partners, and showing they are complementary to the sequence of the putative sRNA. For running MAST for these RNAs, which do not have annotated boundaries, we determined their sequences as following: for CDS we used the actual sequence, for 5'UTR we used -100, +50 nucleotides around the AUG, and for 3'UTR, -50, +100 nucleotides around the stop

codon. For other cases we determined the region for the analysis based on the mapped sequenced fragments. Of note, since the analysis was applied automatically to all RNAs in the data, in some cases mRNAs that interact with several sRNAs were submitted, and, trivially, the sRNA terminator sequence was revealed as a common motif. The Logos presented in this study were created using weblogo 3.4 (Crooks et al., 2004).

Testing the impact of sRNA-target interaction using gene expression and ribosome profiling data

The raw data of six experiments used in this analysis (RyhB (Masse et al., 2005), ArcZ (Papenfort et al., 2009), FnrS (Durand and Storz, 2010), MgrR (Moon and Gottesman, 2009), Spf (Beisel and Storz, 2011) and GcvB (Sharma et al., 2011)) were obtained from NCBI's Gene Expression Omnibus (GEO). These experiments provide gene expression data with and without a tested sRNA (using a deletion strain of the tested sRNA strain overexpressing it from a plasmid vs. the deletion strain expressing a null plasmid). The limma Bioconductor R package (Ritchie et al., 2015) with the GEOquery package (Davis and Meltzer, 2007) were used to assess the log2 fold change in expression of all genes measured in each of the experiments. The raw data for the CyaR experiment (De Lay and Gottesman, 2009) were not available in GEO, but we could use a supplementary table of that paper, including the average normalized intensities for all genes. In these analyses we included RIL-seq targets that were included also in the transcriptome studies. Thus, targets in 3'UTR regions for which no expression data were available were excluded. Candidate targets in intergenic and antisense regions were not considered. For each experiment we computed the t-statistics of the fold-change following over-expression of the sRNA (compared to a null plasmid) using limma lmFit function. We tested by Kolmogorov-Smirnov test if these t-statistics differ statistically significantly between the group of putative RIL-seq targets and the rest of the genes (background).

The raw data from the ribosome profiling experiment under over-expression of RyhB were obtained from the authors of the paper (Wang et al., 2015). Here we tested by Kolmogorov-Smirnov test if the log2 fold change in ribosome profiling between WT

condition and over-expression of RyhB differ statistically significantly between the group of putative RIL-seq targets and the rest of the genes (background).

Comparing identification of known targets by RIL-seq to predictions by CopraRNA

For this comparison we used 17 sRNAs for which CopraRNA [http://rna.informatik.uni-freiburg.de] (Wright et al., 2013) provides predictions in their website, and the sRNA was included in RIL-seq data of statistically significant interactions. For these 17 sRNAs there are 146 known targets (Table S3). In total, CopraRNA predicted 2761 targets for these sRNAs ($p \leq 0.05$), out of which 77 were known targets, and RIL-seq identified 1631 targets, out of which 78 were known targets.

Supplementary References

Argaman, L., and Altuvia, S. (2000). fhlA repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. J Mol Biol *300*, 1101-1112.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol *2*, 2006 0008.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res *37*, W202-208.

Bailey, T.L., and Gribskov, M. (1998). Methods and statistics for combining motif match scores. J Comput Biol *5*, 211-221.

Beisel, C.L., and Storz, G. (2011). The base-pairing RNA spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. Mol Cell *41*, 286-297.

Bertani, G. (2004). Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol *186*, 595-600.

Chao, Y., and Vogel, J. (2016). A 3' UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. Mol Cell *61*, 352-363.

Cherepanov, P.P., and Wackernagel, W. (1995). Gene disruption in: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene *158*, 9-14.

Corcoran, C.P., Podkaminski, D., Papenfort, K., Urban, J.H., Hinton, J.C., and Vogel, J. (2012). Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. Mol Microbiol *84*, 428-445.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome Research *14*, 1188-1190.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A *97*, 6640-6645.

Davis, S., and Meltzer, P.S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics *23*, 1846-1847.

De Lay, N., and Gottesman, S. (2009). The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. J Bacteriol *191*, 461-476.

Durand, S., and Storz, G. (2010). Reprogramming of anaerobic metabolism by the FnrS small RNA. Mol Microbiol *75*, 1215-1231.

Fitzgerald, D.M., Bonocora, R.P., and Wade, J.T. (2014). Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. PLoS Genet *10*, e1004649.

Grabowicz, M., Koren, D., and Silhavy, T.J. (2016). The CpxQ sRNA negatively regulates Skp to prevent mistargeting of β-barrel outer membrane proteins into the cytoplasmic membrane. mBio *7*.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res *31*, 3429-3431.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res *12*, 996-1006.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., *et*

al. (2013). EcoCyc: fusing model organism databases with systems biology. Nucleic Acids Res *41*, D605-612.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In ArXiv e-prints, pp. 3997.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal *17*, http://journal.embnet.org/index.php/embnetjournal/article/view/200.

Masse, E., Vanderpool, C.K., and Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in *Escherichia coli*. J Bacteriol *187*, 6962-6971.

Mili, S., and Steitz, J.A. (2004). Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. RNA *10*, 1692-1694.

Moon, K., and Gottesman, S. (2009). A PhoQ/P-regulated small RNA regulates sensitivity of *Escherichia coli* to antimicrobial peptides. Mol Microbiol *74*, 1314-1330.

Morita, T., Maki, K., and Aiba, H. (2005). RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. Genes Dev *19*, 2176-2186.

Muckstein, U., Tafer, H., Hackermuller, J., Bernhart, S.H., Stadler, P.F., and Hofacker, I.L. (2006). Thermodynamics of RNA-RNA binding. Bioinformatics *22*, 1177-1182.

Papenfort, K., Said, N., Welsink, T., Lucchini, S., Hinton, J.C., and Vogel, J. (2009). Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. Mol Microbiol *74*, 139-158.

Peer, A., and Margalit, H. (2011). Accessibility and evolutionary conservation mark bacterial small-rna target-binding regions. J Bacteriol *193*, 1690-1701.

Peer, A., and Margalit, H. (2014). Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. RNA *20*, 994-1003.

Raines, R.T. (1998). Ribonuclease A. Chemical Reviews *98*, 1045-1066.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res *43*, e47.

Sharma, C.M., Papenfort, K., Pernitzsch, S.R., Mollenkopf, H.J., Hinton, J.C., and Vogel, J. (2011). Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. Mol Microbiol *81*, 1144-1165.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., *et al.* (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. Nat Methods *12*, 323-325.

Urban, J.H., and Vogel, J. (2007). Translational control and target recognition by *Escherichia coli* small RNAs in vivo. Nucleic Acids Res *35*, 1018-1037.

Urban, J.H., and Vogel, J. (2009). A green fluorescent protein (GFP)-based plasmid system to study post-transcriptional control of gene expression in vivo. Methods Mol Biol *540*, 301-319.

Wang, J., Rennie, W., Liu, C., Carmack, C.S., Prevost, K., Caron, M.P., Masse, E., Ding, Y., and Wade, J.T. (2015). Identification of bacterial sRNA regulatory targets using ribosome profiling. Nucleic Acids Res *43*, 10308-10320.

Wright, P.R., Richter, A.S., Papenfort, K., Mann, M., Vogel, J., Hess, W.R., Backofen, R., and Georg, J. (2013). Comparative genomics boosts target prediction for bacterial small RNAs. Proc Natl Acad Sci U S A *110*, E3487-3496.