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

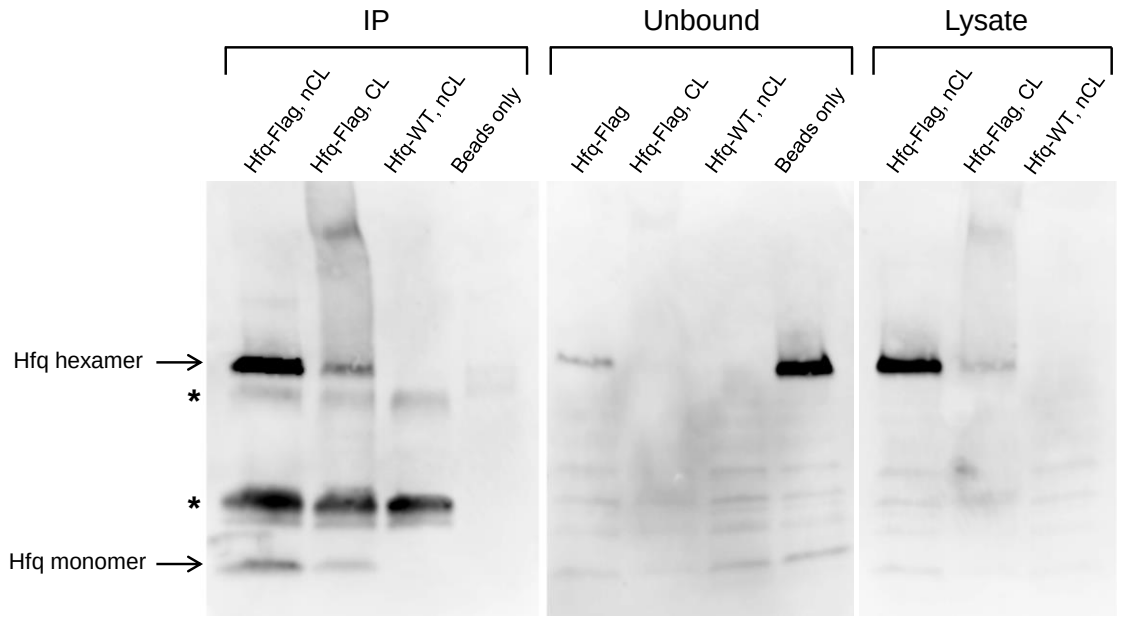

Figure S1 (related to Figure 1). IP assay of crosslinked (CL) and non-crosslinked (nCL) Hfq-Flag

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.

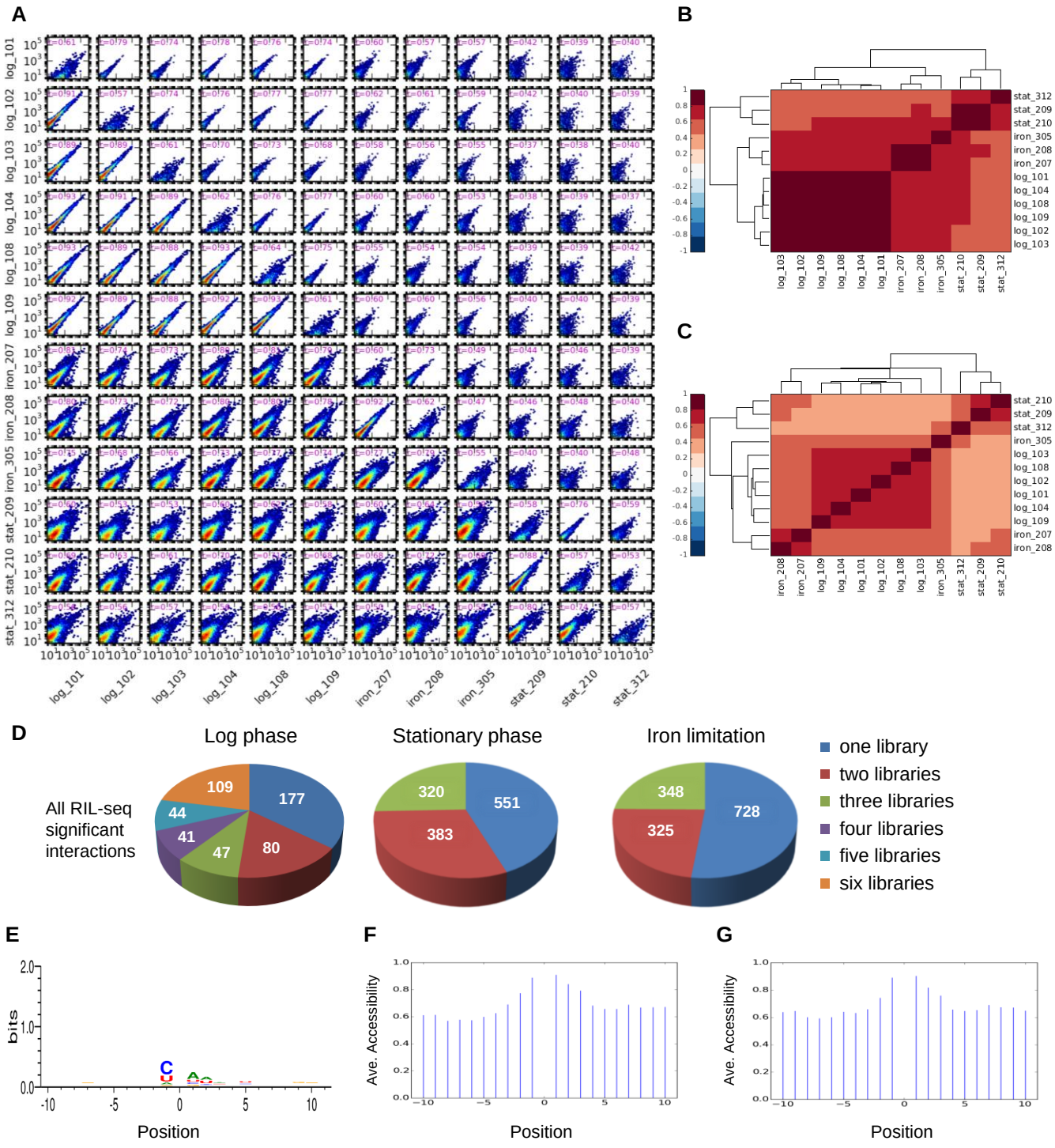

Figure S2 (related to Figure 1). Reproducibility of the data

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).

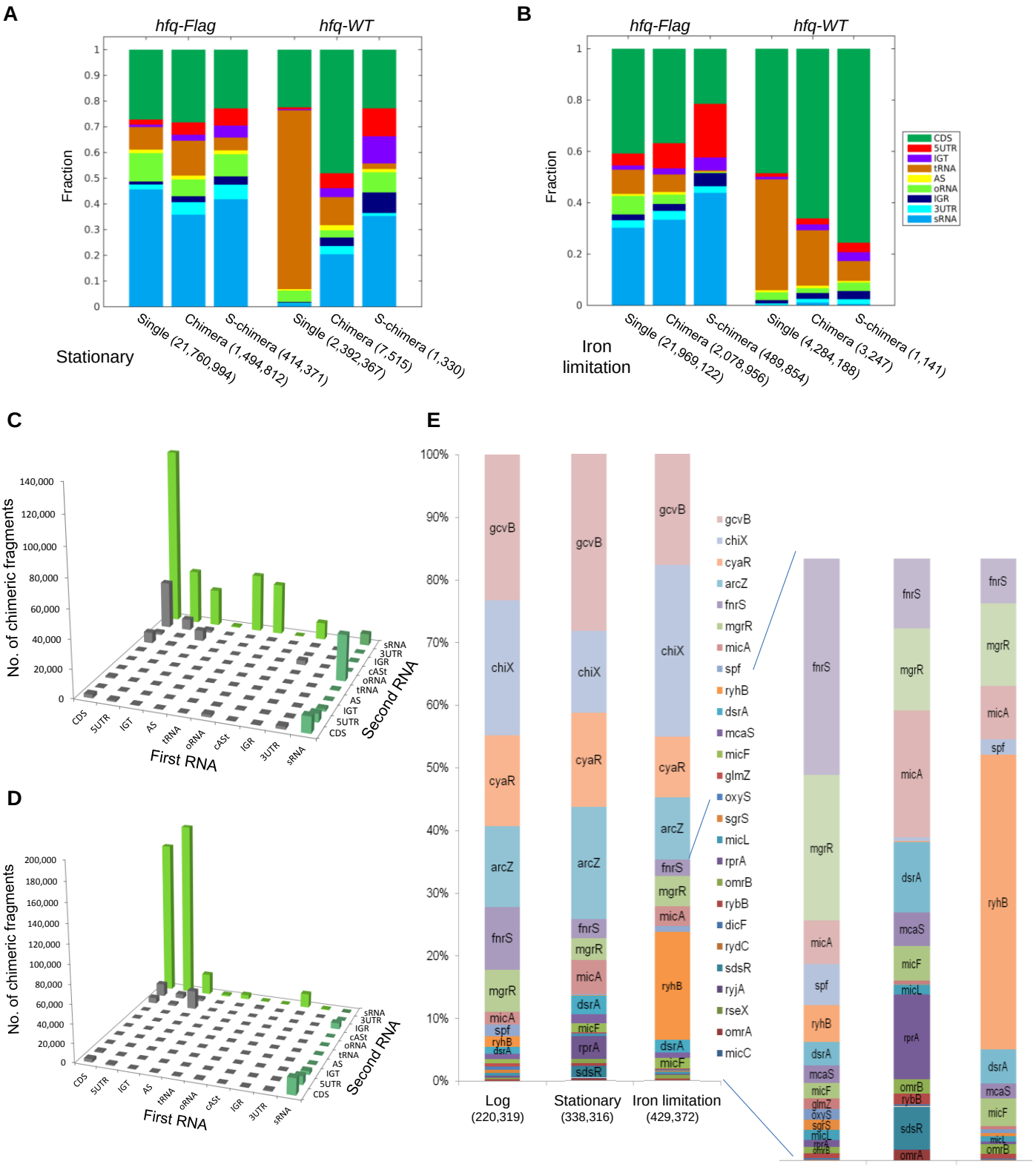
A
hfq-Flag
hfq-WT
Fraction
Stationary
Single (21,760,994)
Chimera (1,494,812)
S-chimera (414,371)
Single (2,392,367)
Chimera (7,515)
S-chimera (1,330)
B
hfq-Flag
hfq-WT
Fraction
Iron limitation
Single (21,969,122)
Chimera (2,078,956)
S-chimera (489,854)
Single (4,284,188)
Chimera (3,247)
S-chimera (1,141)
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

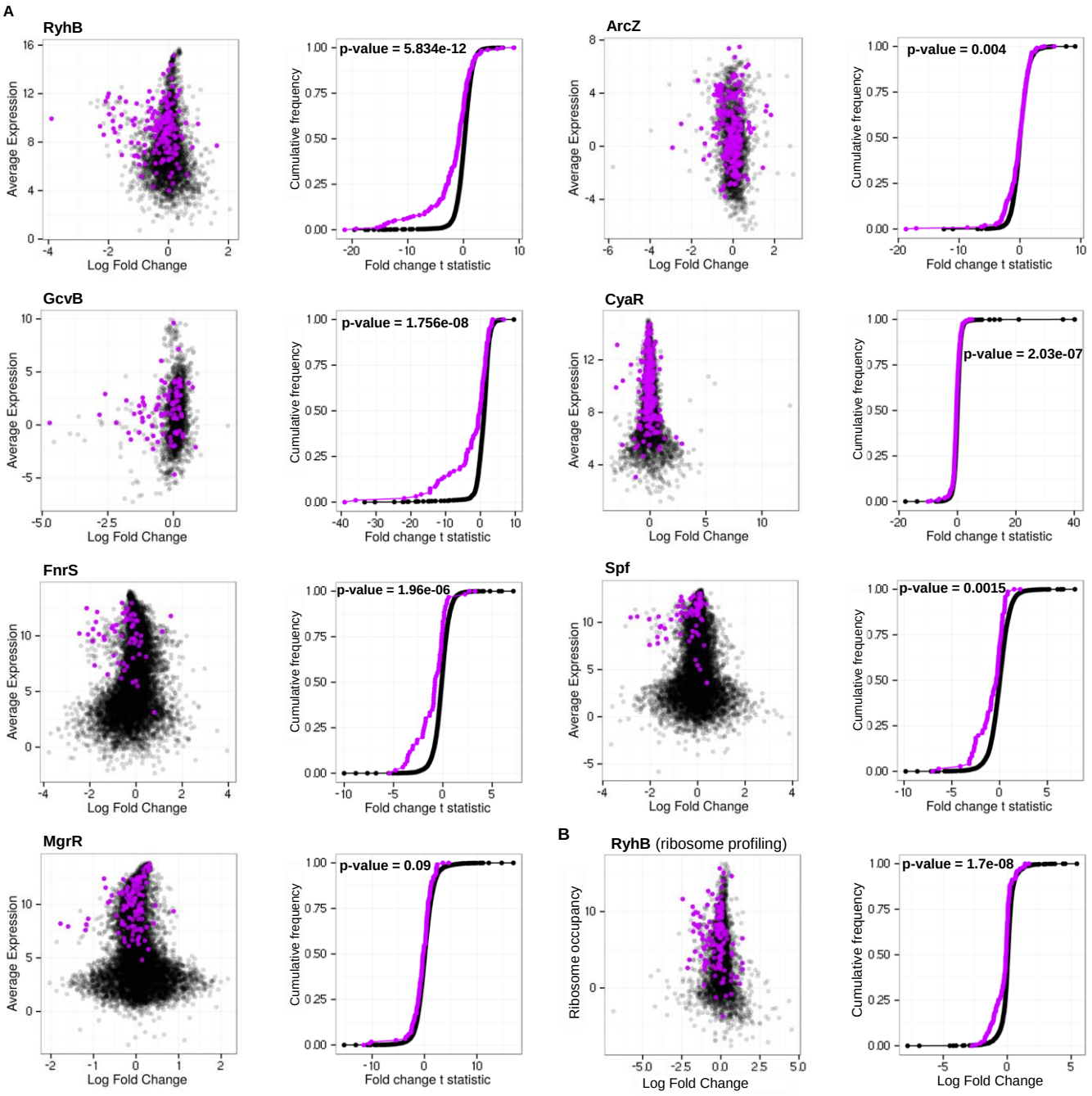

Figure S4 (related to Results in the main text). Impact of the identified interactions on gene expression levels of sRNA bound partners

(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.

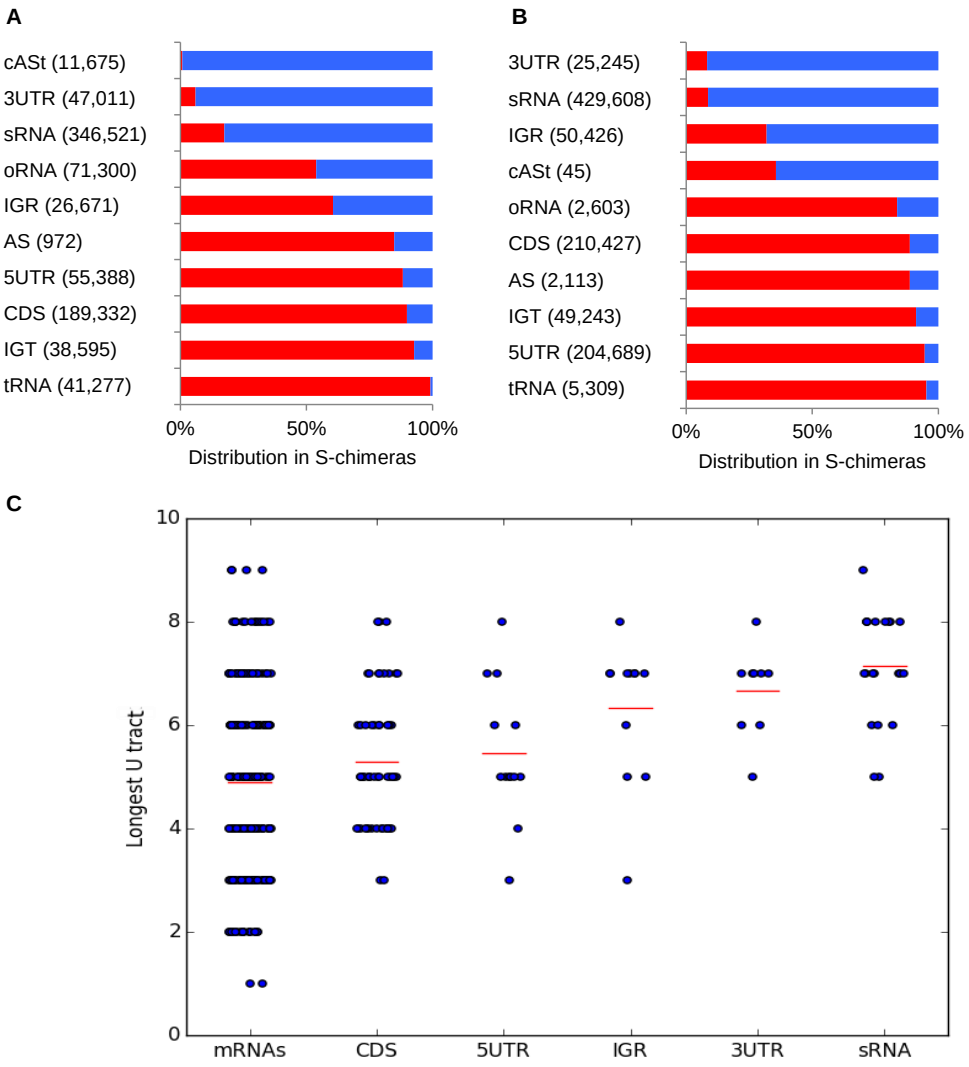

Figure S5 (related to Figure 5). Order of RNAs in the chimeric fragments and analysis of U-tract length

(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).

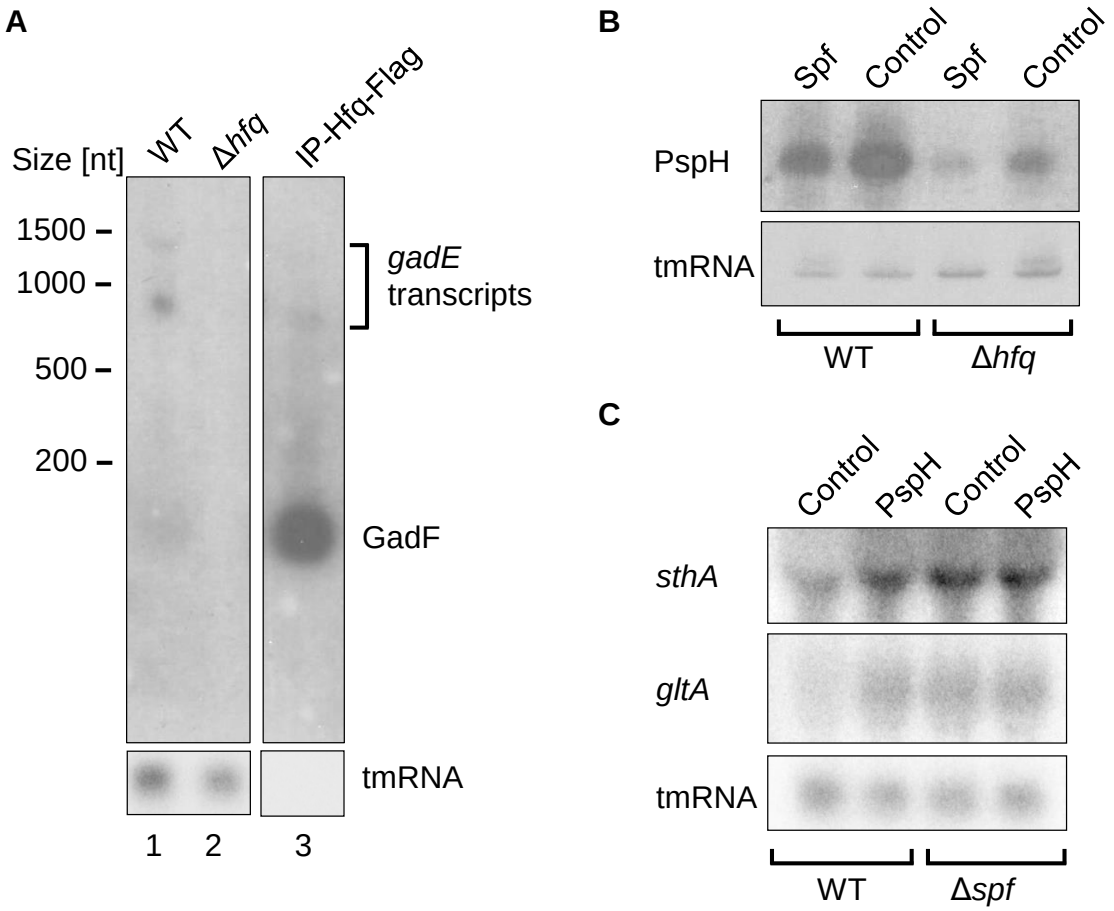

Figure S6 (related to Figure 6). GadF and PspH additional experimental data

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.