Resource

SOURCE
DATA

TRANSPARENT
PROCESS

OPEN
ACCESS

THE
EMBO
JOURNAL

Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*

Erik Holmqvist[1], Patrick R Wright[2], Lei Li[1], Thorsten Bischler[1], Lars Barquist[1], Richard Reinhardt[3], Rolf Backofen[2,4,*] & Jörg Vogel[1,**]

Abstract

The molecular roles of many RNA-binding proteins in bacterial post-transcriptional gene regulation are not well understood. Approaches combining *in vivo* UV crosslinking with RNA deep sequencing (CLIP-seq) have begun to revolutionize the transcriptome-wide mapping of eukaryotic RNA-binding protein target sites. We have applied CLIP-seq to chart the target landscape of two major bacterial post-transcriptional regulators, Hfq and CsrA, in the model pathogen *Salmonella* Typhimurium. By detecting binding sites at single-nucleotide resolution, we identify RNA preferences and structural constraints of Hfq and CsrA during their interactions with hundreds of cellular transcripts. This reveals 3′-located Rho-independent terminators as a universal motif involved in Hfq–RNA interactions. Additionally, Hfq preferentially binds 5′ to sRNA-target sites in mRNAs, and 3′ to seed sequences in sRNAs, reflecting a simple logic in how Hfq facilitates sRNA–mRNA interactions. Importantly, global knowledge of Hfq sites significantly improves sRNA-target predictions. CsrA binds AUGGA sequences in apical loops and targets many *Salmonella* virulence mRNAs. Overall, our generic CLIP-seq approach will bring new insights into post-transcriptional gene regulation by RNA-binding proteins in diverse bacterial species.

Introduction

The fate of RNA molecules in the cell is largely determined at the post-transcriptional level by RNA–protein interactions. RNA-binding proteins (RBPs) are responsible for essential traits such as RNA stability, structure, translatability, export, and localization. Recent screens in human cells have suggested that the number of proteins with RNA-binding properties may be vastly underestimated (Baltz *et al*, 2012; Castello *et al*, 2012; Kramer *et al*, 2014), prompting new systematic searches for RBPs in many eukaryotic systems (Ascano *et al*, 2013). By comparison, our knowledge of the scope and binding preferences of prokaryotic RBPs is lagging behind eukaryotic systems, and new approaches are needed to fully elucidate the roles of RBPs in post-transcriptional control in bacterial pathogens (Barquist & Vogel, 2015). That is, although the structural details of the interactions of many positively and negatively acting proteins with DNA have been established, the paucity of understanding regarding RBPs has been holding back the field of bacterial gene regulation.

Salmonella enterica serovar Typhimurium is a widely studied food-borne bacterial pathogen that invades and replicates in many different eukaryotic host cells. Over the past decade, *Salmonella* has become a bacterial model organism to study post-transcriptional regulation by small regulatory RNAs (sRNAs) and two associated RBPs, Hfq and CsrA (Vogel, 2009; Hébrard *et al*, 2012; Westermann *et al*, 2016). Transcriptomic and RNA co-immunoprecipitation (coIP) analyses have suggested that Hfq and CsrA play global roles in the regulation of *Salmonella* virulence genes (Lawhon *et al*, 2003; Sittka *et al*, 2008; Ansong *et al*, 2009), but precisely how and where these proteins bind cellular transcripts *in vivo* remains to be fully understood.

Hfq is a widely conserved bacterial RBP of the Sm family of proteins which have a ring-like multimeric quaternary structure (Wilusz & Wilusz, 2005). In the Gram-negative bacteria *Salmonella* and *Escherichia coli*, coIP studies have predicted interactions of Hfq

with hundreds of sRNAs and an excess of one thousand mRNAs (Chao et al, 2012; Zhang et al, 2013; Bilusic et al, 2014). By helping sRNAs to regulate target mRNAs, Hfq modulates a variety of physiological traits including phosphosugar detoxification (Rice et al, 2012; Papenfort et al, 2013), catabolite repression (Beisel et al, 2012), envelope stress (Figueroa-Bossi et al, 2006; Gogol et al, 2011; Guo et al, 2014; Chao & Vogel, 2016), metal homeostasis (Desnoyers & Masse, 2012; Coornaert et al, 2013), biofilm formation (Holmqvist et al, 2010; Jørgensen et al, 2012; Mika et al, 2012; Thomason et al, 2012), motility (De Lay & Gottesman, 2012), and virulence (Sittka et al, 2007; Koo et al, 2011; Westermann et al, 2016). In pathogenic Vibrio species, Hfq and sRNAs regulate similarly complex traits, for example, quorum sensing or biofilm formation (Feng et al, 2015; Papenfort et al, 2015).

Mechanistically, Hfq promotes sRNA–mRNA annealing by increasing the rate of duplex formation (Møller et al, 2002; Zhang et al, 2002; Lease & Woodson, 2004; Link et al, 2009; Fender et al, 2010), while at the same time protecting sRNAs from the activity of cellular ribonucleases (Vogel & Luisi, 2011). In addition, Hfq may recruit auxiliary protein factors such as RNase E to promote the decay of target mRNAs (Morita & Aiba, 2011; Bandyra et al, 2012).

Structural studies of Salmonella Hfq confirmed the homo-hexameric ring model (Sauer & Weichenrieder, 2011). The two faces of the ring, denoted proximal and distal, both bind RNA, but show affinity for different RNA sequences: the proximal face tends to target single-stranded U-rich sequences, whereas the distal face interacts with single-stranded A-rich sequences (Schumacher et al, 2002; Mikulecky et al, 2004; Link et al, 2009). More recently, the rim of the Hfq hexamer has emerged as a third RNA-binding surface which interacts with UA-rich RNA and promotes inter-molecular RNA annealing (Updegrove & Wartell, 2011; Sauer et al, 2012; Panja et al, 2013; Dimastrogiovanni et al, 2014). Whereas most of these findings stem from studying Hfq interactions with selected model substrates in vitro, details of transcriptome-wide Hfq binding within RNA in vivo emerged only recently through a crosslinking-based study in pathogenic E. coli (Tree et al, 2014). However, while this study captured many known Hfq targets, it generally failed to observe Hfq binding to sRNA 3′ ends, thus contrasting with the emerging mechanistic model from recent biochemical and structural studies whereby Hfq is loaded onto sRNAs via their 3′ located poly(U) stretch (Otaka et al, 2011; Sauer & Weichenrieder, 2011; Ishikawa et al, 2012; Dimastrogiovanni et al, 2014).

CsrA, initially identified as a regulator of carbon storage and glycogen biosynthesis in E. coli (Romeo et al, 1993), belongs to the large CsrA/Rsm family of RBPs that influence physiology and virulence in numerous pathogenic and non-pathogenic bacteria (Lenz et al, 2005; Brencic & Lory, 2009; Heroven et al, 2012; Romeo et al, 2013; Vakulskas et al, 2015). CsrA/Rsm proteins primarily affect translation of mRNAs by binding to 5′ untranslated regions (UTRs). A wealth of genetic, biochemical, and structural data shows that these proteins generally recognize GGA motifs in apical loops of RNA secondary structures (Dubey et al, 2005; Duss et al, 2014a). Other reported mechanisms of CsrA activity in the cell include promotion of Rho-dependent transcription termination, or mRNA stabilization by masking of RNase E cleavage sites (Yakhnin et al, 2013; Figueroa-Bossi et al, 2014). CsrA may also govern a large post-transcriptional regulon, as inferred from transcriptomic and

RNA co-purification data in Salmonella and E. coli, respectively (Lawhon et al, 2003; Edwards et al, 2011).

The CsrA/Rsm proteins are themselves regulated by sRNAs such as CsrB and RsmZ, which contain multiple GGA sites that titrate the protein away from mRNA targets (Liu et al, 1997; Weilbacher et al, 2003; Valverde et al, 2004). Structural studies of one CsrA-like protein revealed a sequential and cooperative assembly of the protein on antagonistic sRNAs (Duss et al, 2014b). Antagonists of CsrA activity also include the Hfq-dependent sRNA McaS in E. coli (Holmqvist & Vogel, 2013; Jørgensen et al, 2013) and a sponge-like mRNA in Salmonella (Sterzenbach et al, 2013). Again, despite the strong interest in these proteins, the global binding preferences of CsrA/Rsm in vivo remain unknown.

Approaches combining in vivo crosslinking and RNA deep sequencing have been increasingly used to globally map the cellular RNA ligands and binding sites of eukaryotic RBPs in vivo (Darnell, 2010; König et al, 2011; Ascano et al, 2012). Such methods are now widely used in cell culture, tissues, and even whole animals. The purification of RNA–protein complexes after in vivo crosslinking by ultraviolet (UV) light offers several advantages over traditional coIP. Firstly, the UV-induced covalent bonds between protein and RNA survive denaturing conditions, facilitating stringent purification protocols. Secondly, crosslinking enables trimming by ribonucleases to yield protein-protected RNA fragments, pinpointing binding regions with unprecedented resolution. Thirdly, the attachment of a crosslinked peptide to a purified RNA fragment often causes mutations during reverse transcription which identify direct RNA–protein contacts at single-nucleotide resolution (Zhang & Darnell, 2011).

Here, we have employed UV crosslinking of RNA–protein complexes in living bacterial cells, followed by stringent purification and sequencing of crosslinked RNA, to detect transcriptome-wide binding sites of Hfq and CsrA in Salmonella. As well as confirming known binding sites at nucleotide resolution, our study identifies a plethora of new sites that reveal the specificities of Hfq and CsrA interactions with their RNA ligands. Our contact maps for Hfq inter-acting sRNAs and their target mRNAs support a model for Hfq as a mediator of RNA duplex formation and provide new insight into improving sRNA-target prediction. The discovery of CsrA-binding sites in mRNAs shows that CsrA is a direct regulator of Salmonella virulence genes.

Results

Selective enrichment of crosslinked RNA ligands

To comprehensively analyze direct targets of RBPs in vivo, we established a CLIP-seq protocol for purification of crosslinked RNA–protein complexes from bacterial cells irradiated with UV light (Fig 1A). Salmonella strain SL1344 expressing chromosomally FLAG-tagged Hfq was cultured in LB medium to an OD_{600} of 2.0. One half of this culture was then irradiated with UV light while the other half was left untreated. This growth condition activates the invasion genes of Salmonella, that is it enabled us to also capture potential Hfq interactions with virulence-associated transcripts. Hfq–RNA complexes were immunoprecipitated in cell lysates with a mono-clonal anti-FLAG antibody followed by several stringent washes.
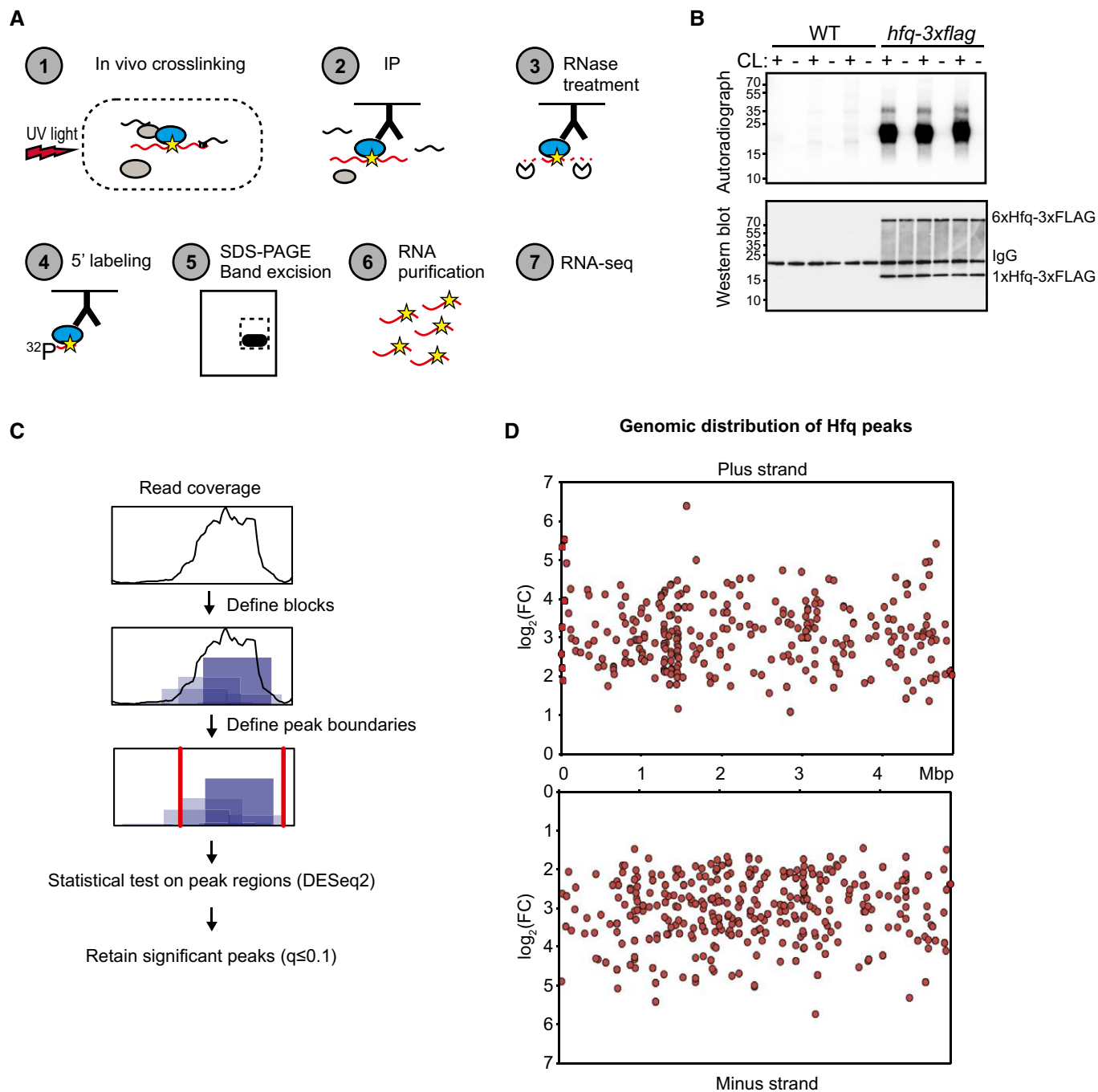
Figure 1. CLIP-seq of Hfq-3xFLAG in *Salmonella*.

A Schematic representation of the CLIP-seq protocol for bacterial RBPs that was established and used in this study. UV: ultraviolet.

B Detection of crosslinked, immunoprecipitated, and radioactively labeled RNA–protein complexes after separation on denaturing SDS–polyacrylamide gels and transfer to nitrocellulose membranes. Radioactive signals were detected by phosphorimaging (top). Detection of Hfq-3xFLAG proteins by Western blot using an anti-FLAG antibody served as a control for successful immunoprecipitation (bottom). CL: crosslinking.

C Schematic representation of binding site determination (peak calling).

D Fold change (*y*-axis) and genomic position (*x*-axis) of Hfq peaks. Mbp: mega basepair.

After on-bead RNase treatment, dephosphorylation, and radioactive labeling of RNA 5′ ends, the complexes were eluted, separated by denaturing SDS–PAGE, and transferred to a membrane. UV irradiation itself did not interfere with protein recovery (as judged by Western blot), but a strong radioactive signal corresponding to bound labeled RNA was detected only in tagged and crosslinked samples, indicating that unspecific RNA–protein interactions were successfully depleted (Fig 1B). RNA–protein complexes from

crosslinked and control samples were extracted from the membrane and treated with proteinase to yield RNA ligands for analysis by Illumina sequencing. The number of sequencing reads obtained for each cDNA library is given in Appendix Fig S1. To avoid biases introduced during library amplification, reads originating from potential PCR duplicates were removed for all downstream analyses.

A very important step in the analysis of CLIP-seq data is peak calling, which is used to differentiate between specific und unspecific binding. Here, two major problems in standard CLIP-seq protocols may confound peak calling approaches. Firstly, in contrast to traditional RNA immunoprecipitation and sequencing (RIP-seq), where comparison to a non-tagged strain or the omission of the antibody serves to control for background noise, CLIP-seq approaches usually lack a standardized negative control. Secondly, in contrast to chromatin immunoprecipitation and DNA sequencing (ChIP-seq), transcript abundance impacts read coverage independent of the affinity of the RBP for a given target. Standard peak callers such as Piranha (Uren et al, 2012) assume the majority of sites to be noise, so the sum of all sites can be used to fit a background model. However, this assumption is problematic if the RBP is a ubiquitous binder and the genome size is rather small. Both criteria apply in our case. To overcome these problems, we developed a specific peak calling algorithm able to identify Hfq-binding sites throughout the *Salmonella* transcriptome. The algorithm first divides consecutive reads into blocks and then merges overlapping blocks into peaks (Fig 1C). Subsequently, based on three biological replicates and three control replicates, each peak was tested for significant enrichment in the crosslinked samples versus the non-crosslinked samples using DESeq2 (Love et al, 2014). This strategy identified 640 significant ($q \leq 0.1$) Hfq peaks (Table EV1) which are distributed across the *Salmonella* transcriptome (Fig 1D).

As a significant advantage of CLIP-seq over simple coIP, crosslinking-induced mutations narrow RNA–protein contacts down to individual nucleotides (Zhang & Darnell, 2011). Thus, we compared the nature of read mutations that (i) occurred in both mate pairs for each read (to discriminate from sequencing errors), (ii) were exclusively present in libraries from crosslinked cultures, and (iii) overlapped with Hfq peaks (Table EV2). T to C mutations were by far the most common crosslink-specific mutation (Fig 2A), and more than half of the Hfq peaks (347/640) contained at least one crosslink-specific mutation. To provide a better display of peak density, the *Salmonella* chromosome was divided into bins of 2×10^4 basepairs. Plotting peak numbers per bin identified certain chromosomal regions in which the density of Hfq peaks is unusually high (Fig 2B). Interestingly, transcripts from the two major pathogenicity islands, SPI-1 and SPI-2, attract the highest Hfq peak

density, supporting the crucial role of Hfq in *Salmonella* virulence (Sittka et al, 2007). Dividing the Hfq peaks into different RNA classes shows that the majority map to sRNAs and mRNAs, the two RNA classes previously known to be targets of Hfq (Fig 2C). In summary, combining CLIP-seq with a new peak calling algorithm and identification of crosslinking-induced mutations provides the basis for a detailed investigation of Hfq–RNA interactions.

Hfq binding in mRNAs

To analyze the general distribution of the 551 Hfq-binding sites detected in mRNAs, we performed a meta-gene analysis of Hfq peaks with respect to mRNA start and stop codons (for polycistronic mRNAs, only the start codon of the first cistron and the stop codon of the last cistron was used). The greatest peak densities were found in 5′UTRs and 3′UTRs (Fig 2D) and confirmed—on the level of individual transcripts—previously predicted Hfq activity, for example, in the 5′UTR of *chiP* mRNA which is a target of ChiX sRNA (Figueroa-Bossi et al, 2009), or the 3′UTR of *hilD* mRNA encoding a virulence regulator (Lopez-Garrido et al, 2014) (Fig 2E and F).

To test whether Hfq recognizes disparate sequences in different parts of mRNAs, we divided the mRNA peaks into those that map to 5′UTRs, CDSs, or 3′UTRs. Using the MEME algorithm (Bailey et al, 2015), only the combined 3′UTRs yielded a significant consensus motif (Fig 2G). This motif strongly resembles Rho-independent transcription terminators present at the 3′ end of many bacterial transcripts, namely GC-rich hairpins followed by single-stranded uridine tails (Wilson & von Hippel, 1995). Indeed, we found a strong enrichment of Hfq 3′UTR peaks at predicted Rho-independent terminators that were specific to mRNAs (Fig 2H; all sRNA terminators were excluded from this analysis). Moreover, CMfinder analysis (Yao et al, 2006) on the Hfq 3′UTR peaks resulted in a motif comprising a hairpin structure followed by a U-rich sequence, strongly resembling a Rho-independent terminator (Fig EV1), suggesting that Hfq binds to mRNA 3′ ends.

Hfq binding in sRNAs

We next compared our crosslinking data to Hfq-binding sites in well-investigated sRNAs. For example, SgrS was proposed to contain an Hfq-binding module consisting of two distinct binding sites: the poly(U) sequence of the Rho-independent terminator at the very 3′ end of SgrS, and an internal hairpin preceded by a U-rich sequence (Ishikawa et al, 2012). In accordance with this, we detected two Hfq peaks within SgrS that mapped to the previously reported binding sites (Fig 3A and B). In addition, the only

Figure 2. Genomic distribution of Hfq-binding sites.

A   Percentage of the occurrence of the indicated mutations among all crosslink-specific mutations found within Hfq peaks.

B   Hfq peak distribution along the *Salmonella* chromosome divided in bins of 2×10^4 basepairs each. The genomic positions of the pathogenicity islands SPI-1 and SPI-2 are indicated. Mbp: mega basepair.

C   Distribution of Hfq peaks among the indicated RNA classes. Numbers in parentheses give the number of called peaks that overlapped with annotations belonging to the respective RNA class.

D   Global peak density distribution (meta-gene analysis) around start and stop codons. For this analysis, only those start and stop codons were used that are flanked by a 5′UTR or 3′UTR, respectively. Vertical dashed lines indicate the position of start and stop codons, respectively.

E, F   Read coverage at the *chiP* (E) and *hilD* (F) loci in libraries from crosslinked and non-crosslinked samples. Exp: experiment, CL: crosslinking

G   Consensus motif generated by MEME using sequences of Hfq peaks mapping to mRNA 3′UTRs.

H   Meta-gene analysis of peak distribution around genomic positions of predicted Rho-independent terminators.
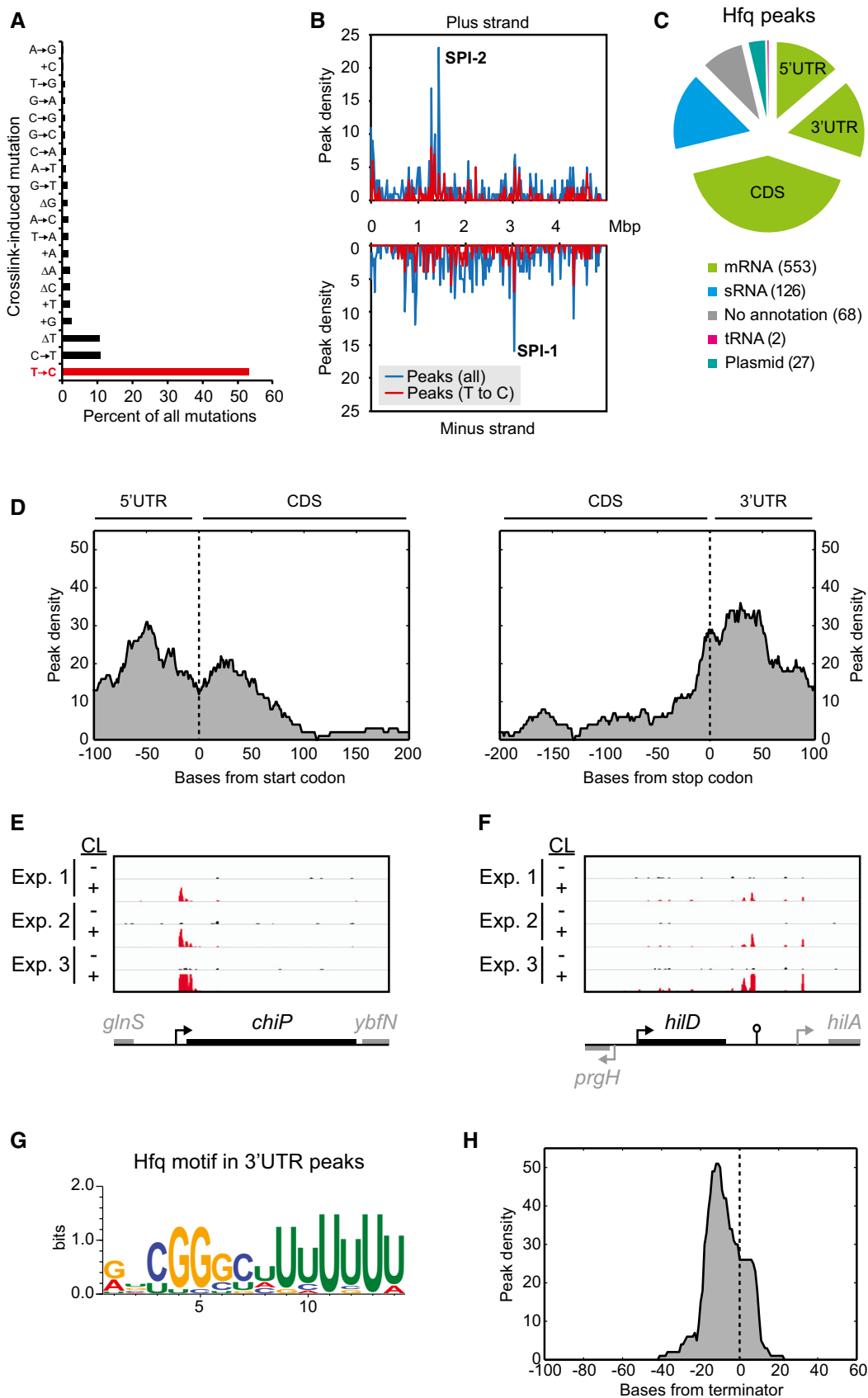
Figure 2.

crosslink-induced mutations detected in SgrS occur within the above-described U-rich sequences (Fig 3B). Likewise, we compared our crosslinking data with the interactions observed in a co-crystal of *Salmonella* Hfq and the sRNA RydC (Dimastrogiovanni *et al*, 2014). The X-ray crystallization data suggest Hfq interacts with four regions on RydC: the proximal site of Hfq interacts with the U-rich 3′ end of RydC; the rim of Hfq interacts with U23/U24, U46/U47, and the RydC 5′ end (Dimastrogiovanni *et al*, 2014). Out of the eight positions in RydC with crosslinking-induced mutations, seven perfectly fit with the crystal structure (Fig 3D). Mutations were found in the 5′ end of RydC, at positions U23, U24, U46, U47, and in the RydC 3′ end (Fig 3D). Taken together, these examples demonstrate that our crosslinking experiments faithfully capture Hfq–RNA interactions at single-nucleotide resolution, in excellent agreement with published work.

The distribution of Hfq peaks over all sRNA sequences suggests that Hfq may interact with different regions in different sRNAs; however, there is a strong bias for Hfq binding toward sRNA 3′ ends (Fig 3E). As for the 3′UTR-binding motif (Fig 2G), the consensus motif found using MEME in peaks mapping to within sRNAs resembles the 3′ region of a Rho-independent terminator (Fig 3F). Following the demonstration of Hfq interactions with 3′ portions of a few sRNAs (Sauer & Weichenrieder, 2011; Ishikawa *et al*, 2012), our screen provides the first global analysis to suggest that Hfq interacts with the 3′ end of many sRNAs detected under the growth condition studied. Taken together, Rho-independent terminators constitute a general Hfq-binding motif shared by mRNAs and sRNAs.

Hfq binding in sRNA–mRNA pairs

A key function of Hfq is to facilitate sRNA–mRNA duplex formation (Møller *et al*, 2002; Zhang *et al*, 2002; Kawamoto *et al*, 2006; Fender *et al*, 2010); this activity seems to require Hfq binding in mRNAs proximal to the site of sRNA pairing, as suggested by studies of *rpoS* mRNA which is regulated by multiple sRNAs (Soper *et al*, 2011). The simultaneous binding of both the sRNA and cognate mRNA by an Hfq hexamer may then accelerate RNA duplex formation at the rim of the protein (Panja *et al*, 2013). To understand where Hfq needs to bind within its ligand to facilitate RNA duplex formation, we performed a meta-gene analysis of Hfq peaks that mapped close to seed pairing regions in known sRNA–mRNA target pairs. In mRNAs, Hfq peaks were significantly more likely to occur 5′ of the respective sRNA interaction site ($P < 0.05$, two-tailed sign test, $n = 17$) (Fig 4A). By contrast, Hfq peaks in sRNAs were found significantly more often 3′ of sRNA seed sequences ($P < 10^{-4}$, two-tailed sign test, $n = 24$) (Fig 4A). This result supports a model whereby Hfq is sandwiched between the mRNA and sRNA of a cognate pair prior to RNA duplex formation (Fig 4B).

The presence of an Hfq site close to an sRNA site in an mRNA improves target regulation (Beisel *et al*, 2012). Therefore, we asked whether our Hfq-binding data could increase the success of sRNA-target predictions. To this end, the top 20 mRNA targets predicted by the CopraRNA algorithm (Wright *et al*, 2013) for each of 17 selected sRNAs were intersected with the list of crosslinked mRNAs, giving 48 predicted mRNA targets with at least one Hfq peak (Fig 4C, Table EV3). Strikingly, inclusion of the Hfq peaks increased the fraction of true positives from 15% to 40% ($P < 10^{-5}$, Fisher's exact test) (Fig 4C).

For experimental validation, we selected the *mglB* mRNA as a new candidate target of Spot42 sRNA. Recognition would occur by a previously established seed sequence within Spot42 (Beisel & Storz, 2011) at a conserved site downstream of the Hfq peak in *mglB* (Figs 4D and EV2). Of note, the levels of MglB, a CRP-cAMP-activated galactose ABC transporter (Zheng *et al*, 2004), are increased in Hfq-deficient cells, predicting that Spot42 represses the *mglB* mRNA in an Hfq-dependent manner (Fig EV2; Sittka *et al*, 2007; Beisel & Storz, 2011). In agreement with this prediction, deletion of *spf* (encoding Spot42) resulted in elevated levels of the *mglB* mRNA (Fig 4E). Reciprocally, we observed a 10-fold repression of this target after pulse-expression of Spot42 (Fig 4F). Spot42 repressed a constitutively transcribed translational *mglB-gfp* fusion, but not a *lacZ-gfp* control, confirming that the regulation occurs at the post-transcriptional level (Fig 4G). To test whether the observed regulation indeed relies on the predicted basepairing, we introduced disruptive mutations in the *mglB-gfp* and Spot42 plasmids (Fig 4H). Deletion of *spf* on the chromosome leads to increased expression of wild-type *mglB-gfp* but not of the mutant *mglB*-gfp* construct (Fig 4H). Likewise, while wild-type Spot42 repressed *mglB-gfp* but not *mglB*-gfp*, the Spot42* mutant repressed *mglB*-gfp* but not *mglB-gfp* (Fig 4H), strongly indicating that the observed regulation indeed relies on basepairing between Spot42 and the *mglB* mRNA, as predicted. In conclusion, these results indicate that knowing which mRNAs are bound by Hfq can dramatically improve the prediction of sRNA targets.

Transcriptome-wide mapping of CsrA-binding sites

Following the successful identification of Hfq-binding sites, we applied our CLIP-seq protocol to CsrA, an RBP that recognizes transcripts very differently compared to Hfq. CsrA has affinity for GGA sequences present in loop regions of hairpins in mRNA 5′UTRs and in a few sRNAs (Vakulskas *et al*, 2015). A *Salmonella* strain carrying a chromosomal *csrA::3xflag* allele was subjected to the same crosslinking and immunoprecipitation strategy described above. As with Hfq, radioactively labeled CsrA-RNA complexes were detected only in crosslinked samples (Fig EV3). Plotting all CsrA peaks obtained from three biological replicates along the *Salmonella* transcriptome revealed a strong enrichment within CsrB and CsrC; almost 40% of reads from all peaks map to these sRNA antagonists of CsrA (Fig 5A and Table EV4), consistent with them being the major cellular ligands of CsrA (Romeo *et al*, 2013). The *glgC* mRNA, the first transcript shown to be directly regulated by CsrA in *E. coli* (Liu *et al*, 1995; Baker *et al*, 2002), was also highly recovered in our experiments (0.5% of reads, Fig 5A and Table EV4).

The CsrB RNA carries multiple hairpins with GGA sequences which serve as high-affinity-binding sites for CsrA. Intriguingly, the read distribution within CsrB is not uniform. Regions with high read densities are separated by low-read regions (Fig 5B). Aligning the CsrA reads on the predicted secondary structure of CsrB, we find that read coverage is highest in the hairpin structures, indicating that these are indeed preferentially bound by CsrA (Fig 5B). Some hairpins show higher coverage than others, perhaps reflecting a hierarchy in CsrA capture by CsrB similar to the proposed step-wise sequestration of the homologous RsmE protein by RsmZ RNA in *Pseudomonas* (Duss *et al*, 2014b). Regarding CsrA mRNA interactions, reads from the *glgC* transcript almost perfectly overlapped
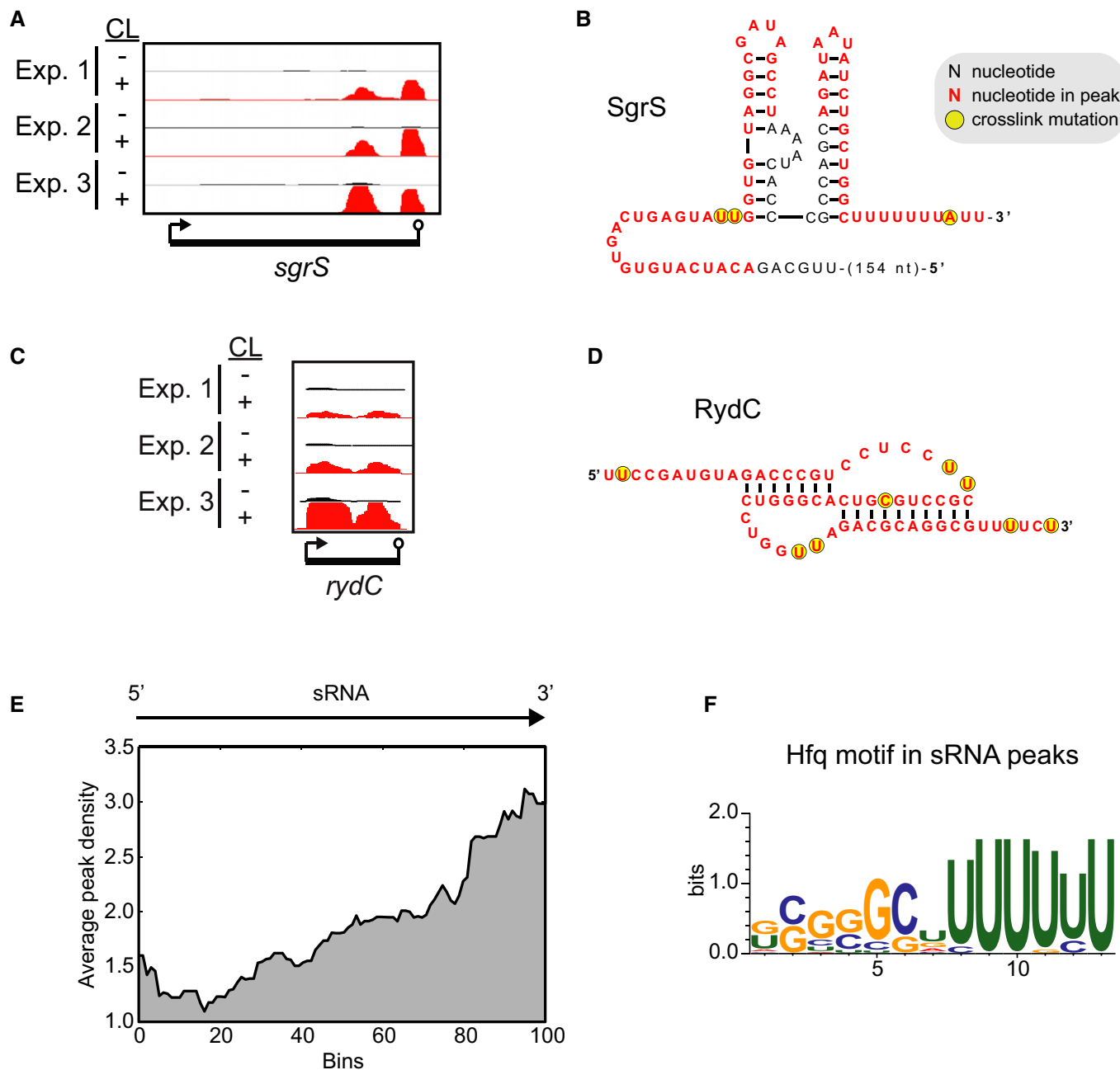
Figure 3. Hfq binding in *Salmonella* sRNAs.

A   Read coverage in libraries from crosslinked and non-crosslinked samples at the *sgrS* locus. CL: crosslinking

B   Predicted secondary structure of the sRNA SgrS. Nucleotides corresponding to a Hfq peak and positions of crosslink-induced mutations are color coded as highlighted in the legend.

C   Read coverage in libraries from crosslinked and non-crosslinked samples at the *rydC* locus. CL: crosslinking.

D   Predicted pseudoknot structure of the sRNA RydC. Nucleotides corresponding to an Hfq peak and positions of crosslink-induced mutations are color coded as highlighted in (B).

E   Meta-gene analysis of the peak distribution along *Salmonella* sRNAs. Length normalization was achieved through proportional binning according to the different lengths of the sRNA sequences.

F   Consensus motif generated by MEME using sequences of peaks mapping to sRNAs as input.

with a GGA-containing hairpin structure in the *glgC* leader (Fig 5C), which was previously defined as the element through which CsrA exercises translational repression in *E. coli* (Baker *et al*, 2002). The detection of CsrA peaks in these two well-documented targets of CsrA suggests that our method readily captures *bona fide* CsrA-binding sites (Fig 5A–C).
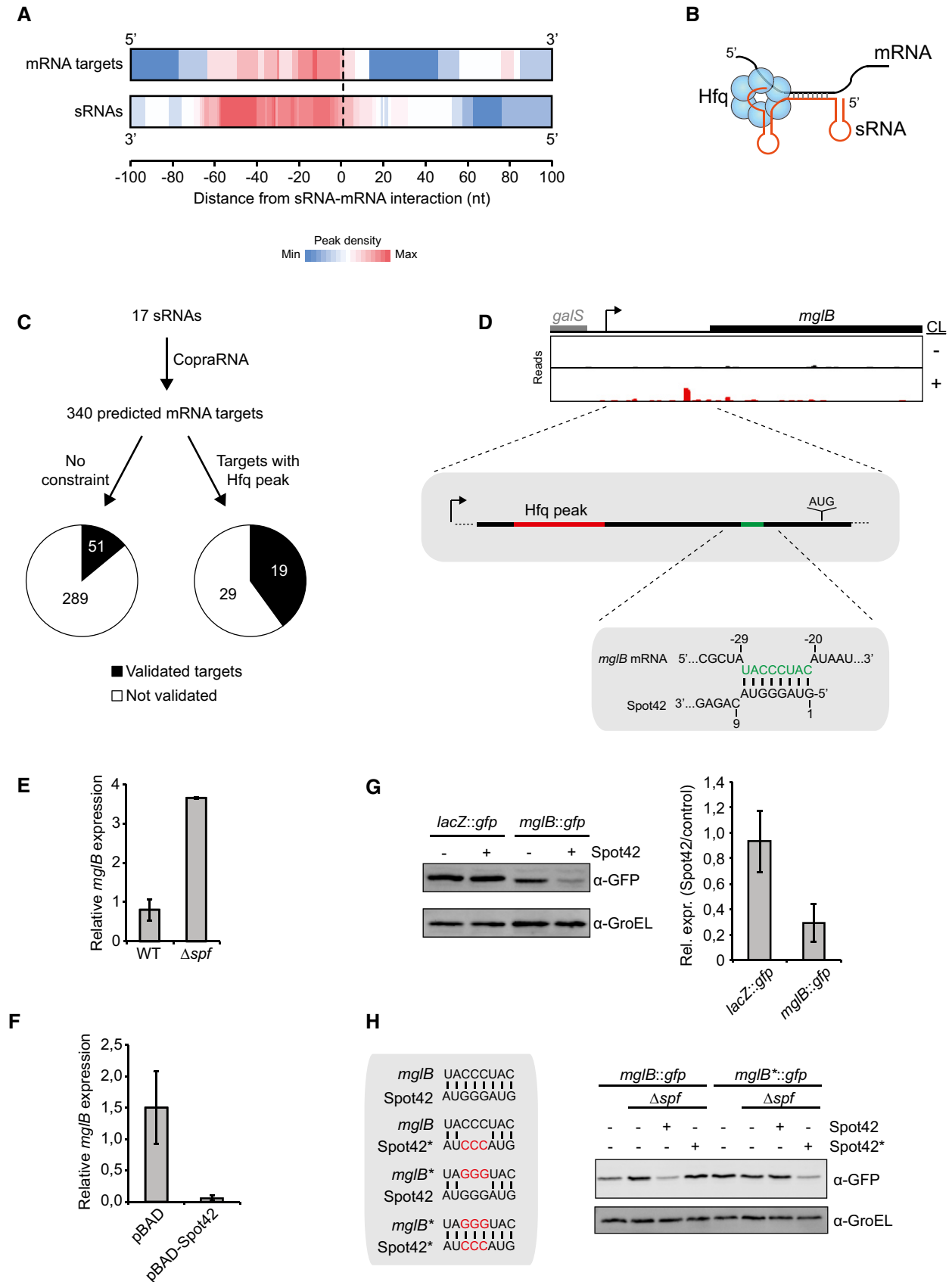
Figure 4.

◀

CsrA consensus motif

We called a total of 467 CsrA peaks, most of which map to within mRNAs (Fig 6A and Table EV4). Meta-gene analysis showed an enrichment of peaks in 5′UTRs compared to CDSs and 3′UTRs, with the strongest enrichment of peaks close to start codons, consistent with CsrA being a regulator of translation initiation (Fig 6B).

High-affinity CsrA–RNA interactions are defined by both RNA sequence and structure (Romeo *et al*, 2013). Interrogation of the CsrA peaks showed that each contained at least one minimal GGA triplet and more than half of them an ANGGA sequence (Fig 6C). Searching all peak regions using the MEME algorithm, we established [A/C]UGGA as the CsrA recognition motif in *Salmonella* (Fig 6D).

Similar to Hfq, we observed that crosslinking of CsrA to RNA frequently causes mutations during reverse transcription. T to C transitions were most prominent (Fig 6E, Table EV5), and these were most often found immediately upstream of a GGA motif (Fig 6E). To analyze the structural context of CsrA-binding sites, we performed CMfinder analysis on all CsrA peaks (Yao *et al*, 2006). Two of the resulting motifs, the one with the highest rank score and the one detected in the most peak sequences (Fig 6F left and right, respectively), consist of stem-loops with a GGA sequence present in the loop regions. Thus, our CLIP analysis confirms the preference for CsrA to interact with AUGGA sequences present in apical loops of hairpin structures. These are the first global data to prove the previous biochemical and genetical studies of individual CsrA ligands, which increasingly suggested ANGGA as a general recognition motif in a variety of bacterial species (Valverde *et al*, 2004; Dubey *et al*, 2005; Majdalani *et al*, 2005; Mercante *et al*, 2006; Babitzke *et al*, 2009; Lapouge *et al*, 2013).

CsrA regulates *Salmonella* virulence genes

Binding of CsrA to target mRNAs typically results in reduced mRNA translation and/or stability (Romeo *et al*, 2013). Since the vast majority of the CsrA sites detected here were previously unknown, we wondered whether they were functional in terms of CsrA-mediated gene regulation. One primary genomic area of CsrA peak density was the invasion gene island SPI-1; likewise, a KEGG pathway analysis suggested enrichment of CsrA peaks in mRNAs

encoding *Salmonella* virulence proteins (Fig 7A and B). Our crosslinking data (Table EV4) not only support the previously proposed direct regulation of *hilD* mRNA (encoding a SPI-1 transcription factor) by CsrA (Martinez *et al*, 2011), but also predict CsrA to target dozens of additional virulence-associated mRNAs from both *Salmonella*'s pathogenicity islands and the core genome (Appendix Fig S2).

To test whether the presence of CsrA peaks correlates with CsrA-mediated gene regulation, we constructed translational *gfp*-fusion reporters (Corcoran *et al*, 2012) to several virulence-associated ORFs from the core genome (*sopD2*) or the SPI-1 locus (*sic-sip* and *prg* operons). GFP fusion plasmids were transformed into Δ*csrB*Δ*csrC* cells harboring either a plasmid expressing CsrB, or an empty control plasmid, reasoning that CsrB-mediated titration of CsrA will translate into GFP reporter regulation. This strategy was chosen to circumvent the genetic instability observed in *csrA* deletion strains (Altier *et al*, 2000). While co-expression of CsrB had no effect on a *lacZ-gfp* control plasmid (pXG10-SF), it caused a strong derepression of a *glgC-gfp* fusion chosen as positive control (Fig EV4), arguing that this experimental setup faithfully monitors CsrA-mediated regulation.

SopD2 is an effector protein that promotes *Salmonella* replication inside macrophages (Figueira *et al*, 2013), and CLIP-seq data identified several CsrA peaks in the *sopD2* 5′UTR and CDS (Fig 7C). Western blot analysis showed that *sopD2-gfp* expression is repressed when CsrA activity is increased as a result of deletion of *csrB* and *csrC* (Fig 7D). This is reversed by complementing the double sRNA deletion strain with *csrB* on a plasmid (Fig 7D). A CsrA peak in the 5′UTR of *sopD2* overlaps with a predicted RNA hairpin structure with two GGA motifs in the loop (Fig 7E). A *sopD2-gfp* fusion in which both GGA motifs were each replaced by CCU totally abolished the regulation, strongly indicating that CsrA directly represses the production of SopD2 (Fig 7E). In further support of this, overexpression of CsrB upregulates the synthesis of endogenous SopD2 in wild-type *Salmonella* (Fig EV5).

The *prgHIJK-orgA* operon encodes components of the SPI-1 type III secretion system needed for host cell invasion, and CsrA peaks were detected in its four-first cistrons (Fig 7F). Western blot analysis with translational fusions encompassing cistron junctions with the downstream cistron being fused to *gfp* showed that translation of *prgI* and *prgJ* is activated upon CsrB overexpression, whereas
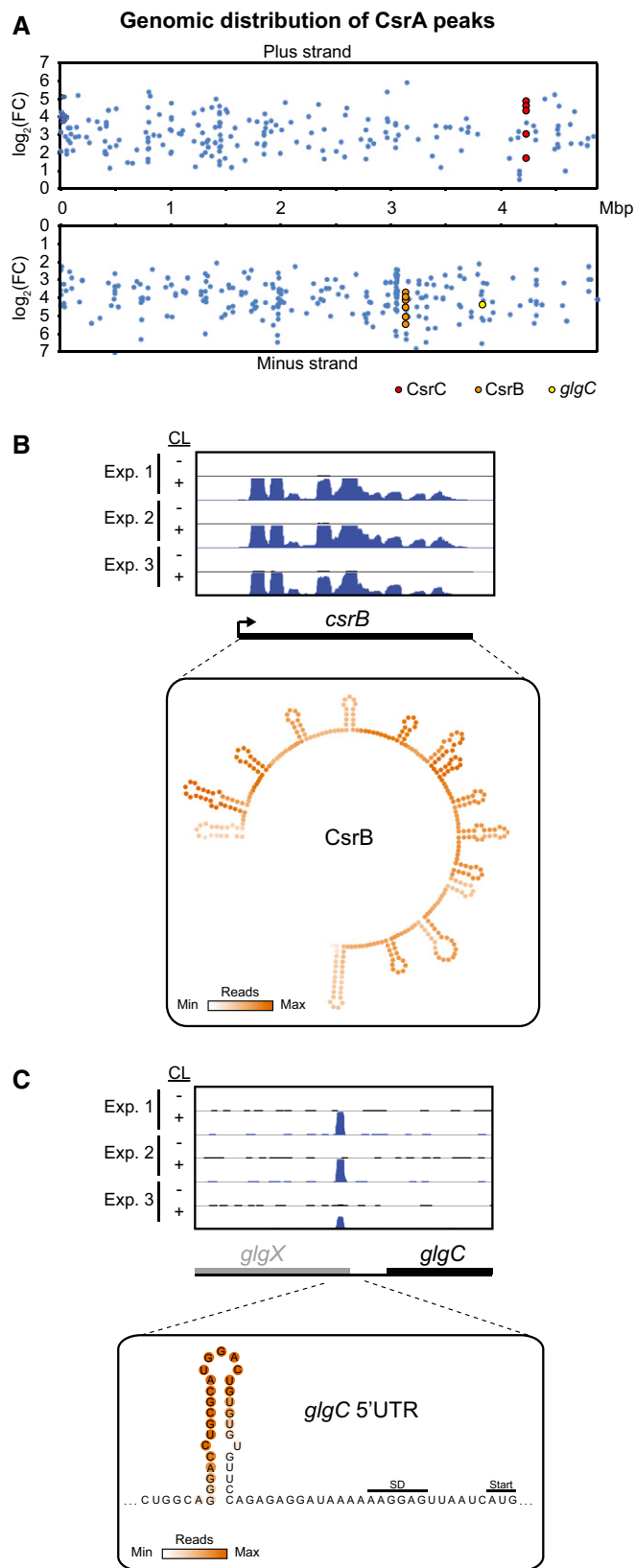
A Genomic distribution of CsrA peaks

SipC, SipD, and SipA), and a putative acyl carrier protein (IacP), and CsrA peaks are distributed across this operon (Fig 7H). Of the four fusions cloned from this operon, three (*sicA, sipC, and sipA*) were clearly upregulated upon CsrB overexpression, indicating that expression from the respective cistrons is repressed by CsrA (Fig 7I). In conclusion, the results shown in Fig 7 strongly indicate that CsrA peaks indeed mark mRNAs that are under direct control of CsrA and suggest that direct regulation of virulence functions by CsrA includes many more mRNAs than previously known.

Discussion

Historically, molecular biologists have focused on the interactions between individual proteins with target nucleic acids *in vitro*, but this approach does not scale well and fails to account for the complexity observed in transcriptional networks. Post-genomic approaches can now potentially provide the global data required to understand post-transcriptional gene regulation in bacteria (Barquist & Vogel, 2015). Specifically, *in vivo* crosslinking methods can determine protein-binding sites within RNA at high resolution and permit stringent purification that diminishes non-specific contamination. Nevertheless, these CLIP-seq approaches have been associated with considerable background noise that, if left uncorrected, increased the identification of false positive interactions (Friedersdorf & Keene, 2014). Here, we have sequenced libraries prepared from both UV crosslinked and non-crosslinked bacterial cultures to control for background RNA, yielding a high-confidence transcriptome-wide map of the binding sites of the two global RNA-binding proteins Hfq and CsrA.

We have shown that Hfq selectively and primarily crosslinks to *Salmonella* mRNAs and sRNAs (Fig 2), in accordance with our previous Hfq coIP results (Sittka *et al*, 2008; Chao *et al*, 2012). More importantly, while relatively few Hfq–sRNA interactions have been studied in biochemical or structural detail, we can faithfully reproduce such results with single-nucleotide resolution in our crosslinking experiment, as shown in Fig 3 for the model sRNAs RydC and SgrS (Ishikawa *et al*, 2012; Dimastrogiovanni *et al*, 2014). Global analysis revealed that Hfq peaks in mRNAs are enriched in 5'UTRs and 3'UTRs as compared to CDS regions (Fig 2), consistent with a role for Hfq in both sRNA-dependent regulation at mRNA 5' regions and 3' end-dependent processes. Analysis of Hfq peak density over the *Salmonella* transcriptome revealed strong enrichment in transcripts expressed from the major pathogenicity islands SPI-1 and SPI-2 (Fig 2B). This may in part be explained by the higher content of A and U residues in these transcripts compared

prgK is not affected (Fig 7G). Of note, the major peaks are located in *prgI* and *prgJ* (Fig 7F). Similarly, the *sicA-sipBCDA-iacP* operon encodes a protein chaperone (SicA), four effector proteins (SipB,
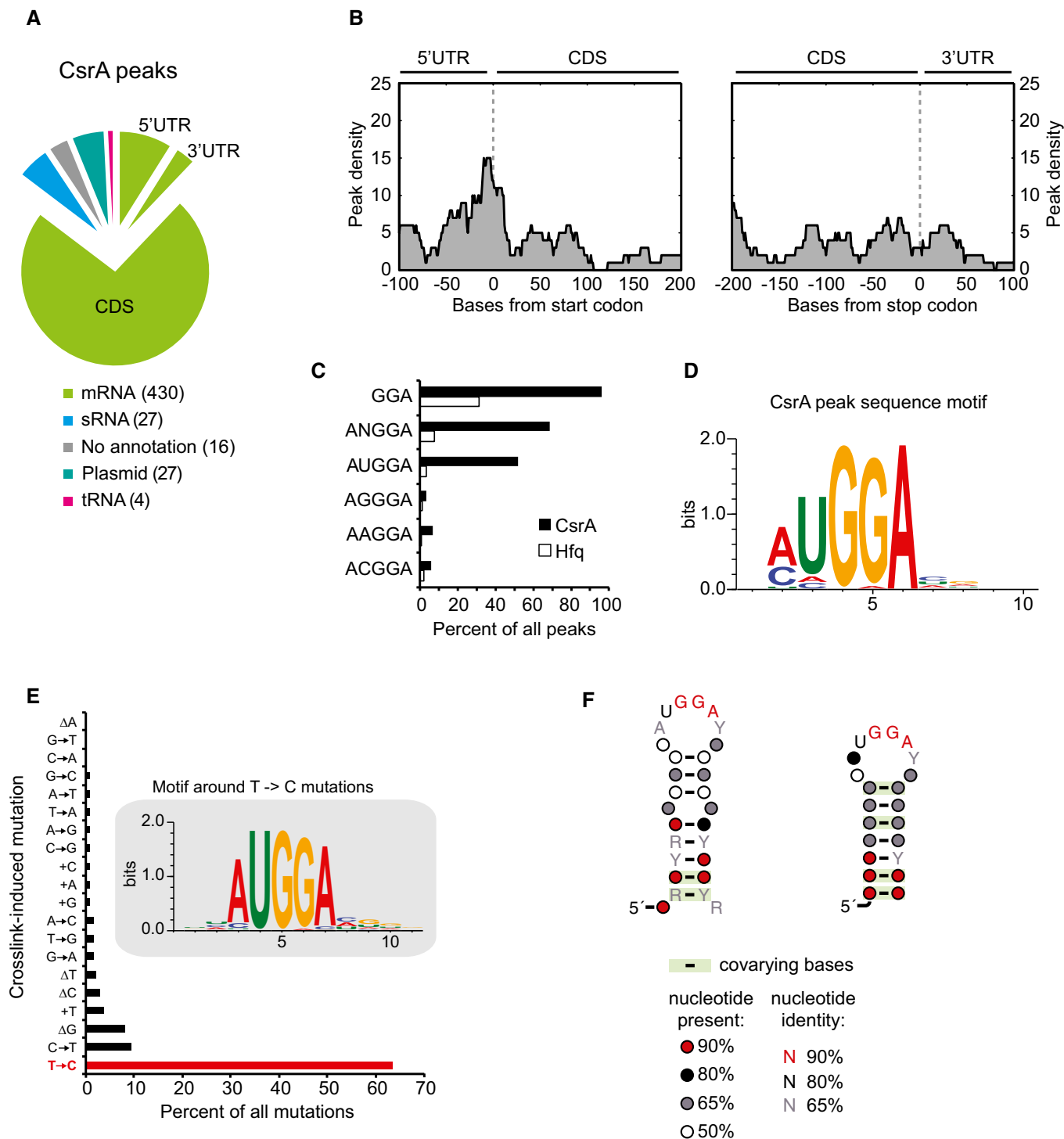
Figure 6. Sequence and structure analysis of CsrA-binding sites.

A   Distribution of CsrA peaks among the indicated RNA classes. Numbers in parenthesis represent the number of called peaks that were mapped within annotations belonging to the respective RNA class.

B   Meta-gene analysis of CsrA peaks around start and stop codons. For this analysis, only those start and stop codons were used that are flanked by a 5′UTR or 3′UTR, respectively.

C   Percentage of peaks that contain the indicated sequences.

D   Consensus motif generated by MEME based on all CsrA peak sequences.

E   Percentage of the occurrence of the indicated mutations among all crosslink-specific mutations found within CsrA peaks. The inset shows the consensus motif generated with MEME using sequences flanking a crosslink-specific T to C mutation as input.

F   Consensus motifs generated by CMfinder based on all CsrA peaks.

Figure 7. CsrA plays a major role in the regulation of *Salmonella* virulence genes.

A   CsrA peak density distribution along the *Salmonella* chromosome in bins of 2×10^4 basepairs. The genomic positions of *Salmonella* pathogenicity islands SPI-1 and SPI-2 are indicated.

B   KEGG pathways that were found significantly enriched among gene annotations to which CsrA peaks were mapped. Pathways that are related to *Salmonella* pathogenicity are highlighted in red.

C   Read coverage from CsrA CLIP-seq at the *sopD2* locus. Light blue bars represent called peaks.

D   Western blot analysis of SopD2-GFP expression from a translational *sopD2-gfp* fusion on a plasmid in the indicated strain backgrounds. Plus sign indicates the presence of plasmid pCsrB. Minus sign indicates the presence of the control vector pJV300. SopD2-GFP signals were detected with an anti-GFP antibody. Expression of GroEL served as a loading control and was detected with an anti-GroEL antibody.

E   Predicted secondary structure of the *sopD2* 5'UTR. Peak position, GGA motifs, and introduced mutations are indicated. GFP fluorescence measurements from the wild-type *sopD2-gfp* fusion or a 2xCCU mutant upon *csrBcsrC* deletion and CsrB complementation. Means and error bars representing standard deviations are based on three independent experiments.

F   Read coverage at the *prgHIJK-orgAB* locus from a CsrA CLIP-seq experiment.

G   Western blot analysis of the expression from the indicated plasmid-borne translational GFP fusions in the presence of plasmids pCsrB (plus signs) or pJV300 (minus signs).

H   Read coverage at the *sicA-sipBCDA-iacP* locus from a CsrA CLIP-seq experiment.

I   Western blot analysis of the expression from the indicated plasmid-borne translational GFP fusions in the presence of plasmids pCsrB (plus signs) or pJV300 (minus signs).

Source data are available online for this figure.

to those expressed from the core genome (Hensel, 2004). Comprehensive analysis of sRNA peaks revealed a strong enrichment of Hfq binding at 3′ ends (Fig 3). The highly enriched consensus motifs found in peak sequences from either mRNA 3′UTRs or sRNAs, respectively, both resemble the 3′ region of Rho-independent terminators (Figs 2, 3 and EV1) and were indeed found in 3′UTRs of mRNAs predicted to transcriptionally terminate in a Rho-independent manner (Fig 2).

The strong evidence for Hfq binding to 3′ ends in mRNAs and sRNAs presented here agrees with previous reports on individual Hfq ligands. Hfq protects RNA from 3′ to 5′ exonuclease activity by binding to, and stimulating the addition of, non-templated poly(A) sequences to RNA 3′ ends by poly(A) polymerase PAPI (Hajnsdorf & Regnier, 2000; Le Derout *et al*, 2003). The sRNA SgrS strongly depends on Hfq binding at its 3′ poly(U) tail for both stability and target regulation (Otaka *et al*, 2011), and the destabilization of SgrS in the absence of Hfq is dependent on the exonuclease PNPase (Andrade *et al*, 2012).

That Hfq binds so commonly to mRNA 3′ ends may be very relevant for sRNA evolution. Cloning or RNA-seq-based studies have identified many sRNAs derived from mRNA 3′UTRs (Vogel *et al*, 2003; Kawano *et al*, 2005; Sittka *et al*, 2008; Chao *et al*, 2012). Whether these sRNAs are produced from internal promoters or by endonucleolytic cleavage of the parental mRNA, they often possess a Rho-independent terminator shared with the mRNA expressed from the same locus (Miyakoshi *et al*, 2015b). Several 3′ UTR-derived sRNAs have been shown to be functional, for example DapZ (Chao *et al*, 2012), MicL (Guo *et al*, 2014), or SroC (Miyakoshi *et al*, 2015a), suggesting that mRNA 3′UTRs may serve as evolutionary birthplaces for sRNAs (Miyakoshi *et al*, 2015b; Updegrove *et al*, 2015). This extends to other types of regulatory transcripts such as recently discovered sRNA sponges that are made from the 3′ end of tRNA precursors (Lalaouna *et al*, 2015).

A key finding from our analysis of the crosslinking data is that we were able to locate Hfq-binding sites in relation to sRNA–mRNA interaction sites (Fig 4). Our observation of preferential binding of Hfq to 5′ of the sRNA interaction site in an mRNA target, and 3′ of the seed sequence in the recognizing sRNA, supports a model whereby Hfq brings the two RNAs together to facilitate RNA duplexing. We used this global information on Hfq binding to substantially

improve sRNA-target predictions (Fig 4), illustrating how global RNA–protein interaction maps can foster a better understanding of post-transcriptional networks and discovering the *mglB* mRNA as a target for the sRNA Spot42 (Fig 4). MglB is a transporter of the non-preferred carbon source galactose, and its expression is activated by CRP–cAMP (Zheng *et al*, 2004). Thus, the regulation of *mglB* by Spot42 fits with a proposed model in which Spot42 and CRP form a feed-forward loop to reduce leaky expression of proteins during carbon foraging (Fig EV2; Beisel & Storz, 2011).

The fact that Hfq binds RNA on three distinct faces of the hexamer, each with a different sequence preference, produces a challenge for CLIP-seq methods in that ligation of sequencing adapters to RBP-bound RNA, as well as UV irradiation, may introduce biases in binding site detection. This may explain why our Hfq CLIP-seq data contrast with a recent crosslinking study of Hfq in *E. coli* (Tree *et al*, 2014). This latter study identified neither the 3′-located terminator-like consensus motif nor an enrichment of Hfq-binding sites in sRNA 3′ ends. Instead, the authors concluded that Hfq binding occurs in the seed sequences located in the middle or at the 5′ end of sRNAs. These differences can be explained by differences in the protocols: 3′ adapter ligation to RNA in complex with Hfq (Tree *et al*, 2014) versus adapter ligation after the RNA fragments are released from Hfq (this study). As RNA 3′ ends may not be accessible to ligation when bound to the proximal side of Hfq, adapter ligation to purified RNA as performed here may be the preferred strategy for CLIP approaches when studying proteins that target RNA 3′ ends.

In addition, Tree *et al* (2014) reported a general ARN motif in Hfq crosslink regions, which seemed consistent with structural data on the interaction between the distal face of Hfq and A-rich sequences (Link *et al*, 2009), and the involvement of mRNA located ARN sequences in sRNA-dependent regulation (Salim & Feig, 2010; Beisel *et al*, 2012; Salim *et al*, 2012; Peng *et al*, 2014). Reviewing our CLIP-seq data, on the one hand, almost all (38/39) Hfq peaks in mRNAs known to be targeted by sRNAs (including *rpoS*, *ompA*, *ompC*, *cfa*, and *mglB*) contain at least one ARN motif (Table EV1). On the other hand, we only detected Hfq peaks in 30% of the previously described sRNA targets (Table EV1) (Wright *et al*, 2013), and we did not observe a significant enrichment of ARN motifs among the mRNA peak sequences compared to randomly selected
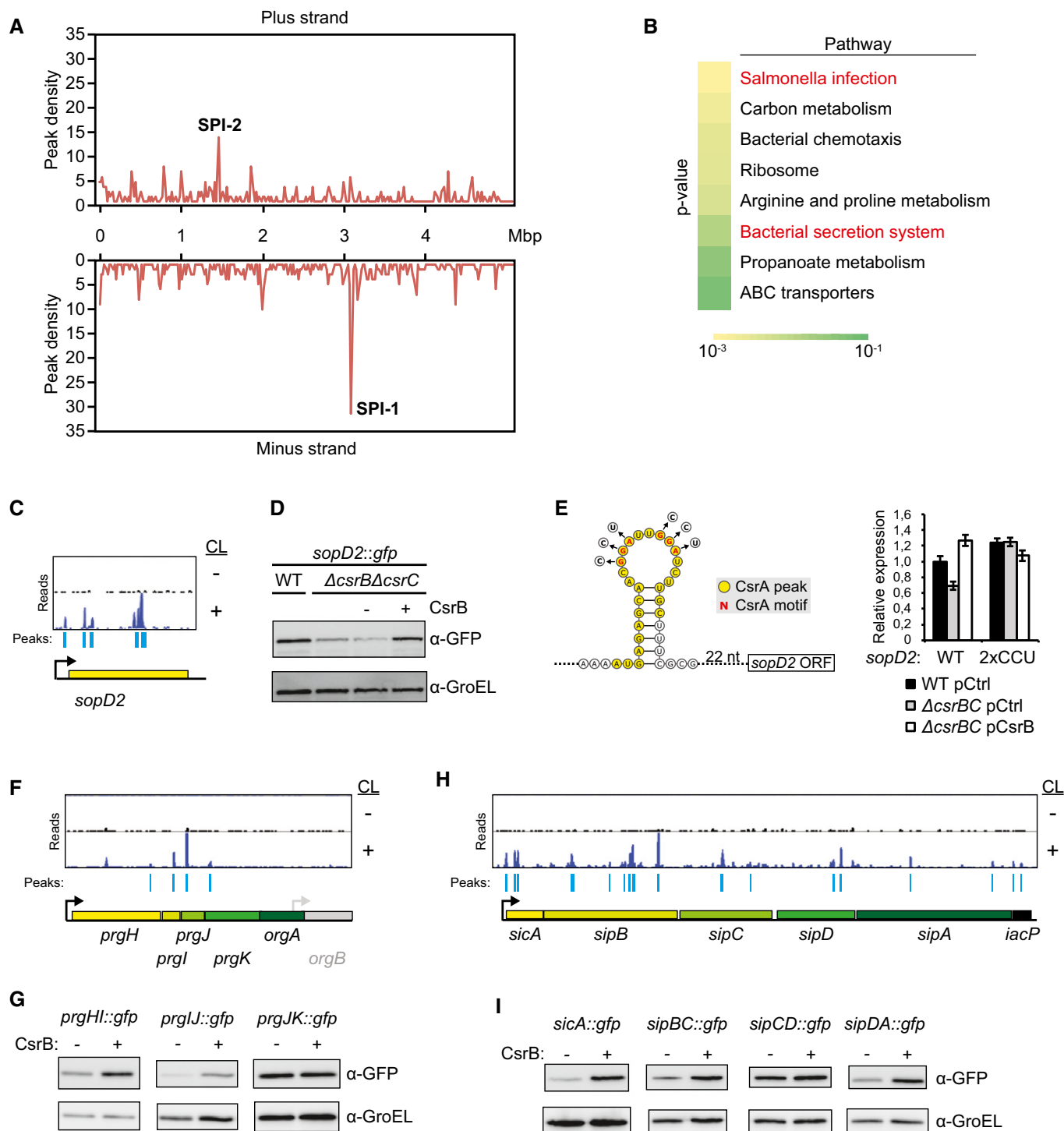
Figure 7.

sequences. One explanation for this discrepancy may be that uridines are more prone to crosslink than other nucleosides (Sugimoto *et al*, 2012); this bias together with the above-discussed adaptor ligation issues may explain why we preferentially detect binding of Hfq at 3′-located U-rich sequences, while the different adapter ligation strategy forced preferential detection of A-rich sequences in the previous *E. coli* study (Tree *et al*, 2014).

Moreover, the canonical view that sRNAs generally interact with the proximal side of Hfq and mRNA targets with the distal side has already been challenged: a recent study showed that some sRNAs use ARN sequences to interact with the distal side of Hfq, whereas their cognate targets harbor 5′UTR-located UA-rich rim-binding sequences (Schu *et al*, 2015). In support of this finding, we find crosslinking mutations in an ARN sequence in the sRNA ChiX and in a UA-rich

sequence in the cognate target mRNA *chiP* (*ybfM*) (Table EV2). Taken together, we propose that mapping of the *in vivo* binding events at each of the three Hfq interaction faces, applying CLIP-seq to mutant Hfq proteins, should be undertaken to further test the current model of distinct "sRNA" and "mRNA" binding faces of Hfq.

These issues with Hfq notwithstanding, the successful application of our crosslinking protocol to CsrA, an RBP with very different targets and recognition mode to Hfq, strongly supports the general applicability of our crosslinking protocol. In contrast to Hfq-binding regions, the vast majority of the detected CsrA-binding sites contain the crucial GGA motif for CsrA–RNA interactions (Figs 5 and 6; Vakulskas *et al*, 2015). CsrA is known to regulate virulence gene expression in *Salmonella*, and a direct interaction between CsrA and *hilD* mRNA, encoding a transcriptional activator of SPI-1, has been described (Martinez *et al*, 2011). In addition to binding *hilD* mRNA, our crosslinking data suggests that CsrA binds to a plethora of virulence-associated mRNAs (Appendix Fig S2). The regulatory potential of newly discovered CsrA-binding sites in virulence-associated mRNAs was confirmed using GFP reporters (Fig 7), consistent with previous reports showing that the levels of some of these mRNAs depend on the intracellular CsrA concentration (Altier *et al*, 2000; Lawhon *et al*, 2003). Even though our validation of CsrA targets is far from comprehensive, it already expands the number of *Salmonella* virulence mRNAs that are post-transcriptionally regulated by CsrA sixfold. Based on our findings, it is likely that more virulence mRNAs are directly regulated by CsrA.

In *Escherichia coli*, the Hfq-dependent McaS sRNA was recently reported to titrate CsrA, suggesting that sRNAs other than CsrB and CsrC may be functional CsrA interaction partners (Jørgensen *et al*, 2013). Interestingly, we also detected binding sites for CsrA in sixteen sRNAs in addition to CsrB and CsrC (Fig 6 and Table EV4), although the read coverage of these additional sRNAs was far below that of CsrB and CsrC. The majority of these sRNAs (14 of 16) carry between one and six GGA motifs, and many of the corresponding peak sequences (12 of 16) fold into hairpins with GGA sequences in the loops (Appendix Fig S3), suggesting that they possess bona fide CsrA-binding sites. Apart from a few well-characterized Hfq-binding sRNAs, of which only one (SdsR) harbors GGA motifs, the majority of the sRNAs that crosslinked to CsrA are uncharacterized. Comparative expression analysis revealed that several of these sRNAs (STnc1890, STnc2080, STnc1210, STnc1480, PinT, and SdsR) are induced in late stationary phase, a growth condition in which CsrB and CsrC are repressed (Kröger *et al*, 2013). This suggests that these six sRNAs may compete with CsrB and CsrC under specific conditions. Future studies will be required to determine whether or not these sRNAs are functional CsrA antagonists, or perhaps are regulated by CsrA.

Bacteria express a plethora of regulatory RBPs for which no global binding site information is available. Examples of these include proteins with RNA-binding domains found in cold-shock proteins (the Csp family of proteins) and proteins such as ProQ that possess a FinO-like RNA-binding domain (Phadtare *et al*, 1999; Mark Glover *et al*, 2015). We believe that our procedure for global mapping of the Hfq and CsrA interactomes with cellular RNA will lay the foundations for future studies of other important bacterial RBPs and may also rapidly identify proteins with putative RNA-binding potential. Such studies should be a major future direction in the study of post-transcriptional phenomena in bacteria and will shed light on this shadowy area of gene regulation.

Materials and Methods

Oligodeoxyribonucleotides

DNA oligonucleotides are listed in Appendix Table S1.

Bacterial strains and plasmids

All experiments were performed with *Salmonella enterica* serovar Typhimurium strain SL1344 or derivatives thereof as listed in Appendix Table S2. All plasmids used in this study are listed in Appendix Table S3. Construction of strains and plasmids is described in Appendix Supplementary Methods. The addition of a FLAG-tag to Hfq or CsrA affected neither bacterial growth nor regulation of known Hfq or CsrA targets, indicating that the tag did not compromise protein function (Appendix Fig S4).

UV crosslinking, immunoprecipitation, and RNA purification

For each biological replicate, 200 ml bacterial culture was grown until an OD_{600} of 2.0. Half of the culture was directly placed in a 22×22 cm plastic tray and irradiated with UV-C light at 800 mJ/cm^2. Cells were pelleted in 50 ml fractions by centrifugation for 40 min at 6,000 *g* and 4°C, resuspended in 800 µl NP-T buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 0.05% Tween, pH 8.0) and mixed with 1 ml glass beads (0.1 mm radius). Cells were lysed by shaking at 30 Hz for 10 min and centrifuged for 15 min at 16,000 *g* and 4°C. Cell lysates were transferred to new tubes and centrifuged for 15 min at 16,000 *g* and 4°C. The cleared lysates were mixed with one volume of NP-T buffer with 8 M urea, incubated for 5 min at 65°C in a thermomixer with shaking at 900 rpm and diluted 10× in ice-cold NP-T buffer. Anti-FLAG magnetic beads (Sigma) were washed three times in NP-T buffer (30 µl 50% bead suspension was used for a lysate from 100 ml bacterial culture), added to the lysate, and the mixture was rotated for one hour at 4°C. Beads were collected by centrifugation at 800 *g*, resuspended in 1 ml NP-T buffer, transferred to new tubes, and washed 2× with high-salt buffer (50 mM NaH$_2$PO$_4$, 1 M NaCl, 0.05% Tween, pH 8.0) and 2× with NP-T buffer. Beads were resuspended in 100 µl NP-T buffer containing 1 mM MgCl$_2$ and 2.5 U benzonase nuclease (Sigma) and incubated for 10 min at 37°C in a thermomixer with shaking at 800 rpm, followed by a 2-min incubation on ice. After one wash with high-salt buffer and two washes with CIP buffer (100 mM NaCl, 50 mM Tris–HCl pH 7.4, 10 mM MgCl$_2$), the beads were resuspended in 100 µl CIP buffer with 10 units of calf intestinal alkaline phosphatase (NEB) and incubated for 30 min at 37°C in a thermomixer with shaking at 800 rpm. After one wash with high-salt buffer and two washes with PNK buffer (50 mM Tris–HCl pH 7.4, 10 mM MgCl$_2$, 0.1 mM spermidine), one-tenth of the beads was removed for subsequent Western blot analysis. The remaining beads were resuspended in 100 µl PNK buffer with 10 U of T4 polynucleotide kinase and 10 µCi γ-^{32}P-ATP and incubated for 30 min at 37°C. After three washes with NP-T buffer, the beads were resuspended in 20 µl Protein Loading buffer (0.3 M Tris–HCl pH 6.8, 0.05% bromophenol blue, 10% glycerol, 7% DTT) and incubated for 3 min at 95°C. The magnetic beads were collected on a magnetic separator, and the supernatant was loaded and separated on a 15% SDS–polyacrylamide gel. RNA–protein complexes were transferred

to a nitrocellulose membrane, the protein marker was highlighted with a radioactively labeled marker pen and exposed to a phosphor screen for 30 min. The autoradiogram was used as a template to cut out the labeled RNA–protein complexes from the membrane. Each membrane piece was further cut into smaller pieces, which were incubated for 30 min in a thermomixer at 37°C with shaking at 1,000 rpm in 400 μl PK solution [50 mM Tris–HCl pH 7.4, 75 mM NaCl, 6 mM EDTA, 1% SDS, 10 U of SUPERaseIN (Life Technologies) and 1 mg/ml proteinase K (ThermoScientific)] whereafter 100 μl 9 M urea was added and the incubation was continued for additional 30 min. About 450 μl of the PK solution/urea was mixed with 450 μl phenol:chloroform:isoamyl alcohol in a phase-lock tube and incubated for 5 min in a thermomixer at 30°C with shaking at 1,000 rpm followed by centrifugation for 12 min at 16,000 *g* and 4°C. The aqueous phase was precipitated with 3 volumes of ice-cold ethanol, 1/10 volume of 3 M NaOAc pH 5.2, and 1 μl of GlycoBlue (Life Technologies) in LoBind tubes (Eppendorf). The precipitate was pelleted by centrifugation (30 min, 16,000 *g*, 4°C), washed with 80% ethanol, centrifuged again (15 min, 16,000 *g*, 4°C), dried 2 min at room temperature, and resuspended in 10 μl sterile water.

cDNA library preparation

To enable sequencing on Illumina instruments, libraries were prepared using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (#E7300, New England Biolabs) according to the manufacturer's instructions. About 2.5 μl purified RNA (or sterile water as negative control) was mixed with 0.5 μl 3′ SR Adaptor (diluted 1:10) and 0.5 μl nuclease-free water, incubated for 2 min at 70°C and chilled on ice. After addition of 5 μl 3′ ligation reaction buffer and 1.5 μl 3′ ligation enzyme mix, the samples were incubated for 60 min at 25°C. About 0.25 μl SR RT primer and 2.5 μl nuclease-free water were added followed by incubation for 5 min at 75°C, 15 min at 37°C, and 15 min at 25°C. For ligation of the 5′ adaptor, the sample was mixed with 0.5 μl 5′ SR adaptor (denatured, diluted 1:10), 0.5 μl 10× ligation reaction buffer, and 1.24 μl ligation enzyme mix and incubated for 60 min at 25°C. cDNA synthesis was carried out by the addition of 4 μl first strand synthesis reaction buffer, 0.5 μl murine RNase inhibitor, and 0.5 μl Protoscript reverse transcriptase and incubation at 50°C for 60 min. The reverse transcription activity was inhibited by a 15-min incubation at 70°C. The cDNA was amplified by PCR by mixing 10 μl cDNA sample with 25 μl 2× LongAmp Taq PCR master mix, 1.25 μl SR primer and 17.5 μl nuclease-free water in a thermal cycler with the following program: 30 s at 94°C, 18 rounds of (15 s at 94°C, 30 s at 62°C, and 15 s at 70°C). The PCRs were purified on columns (QIAGEN), eluted in 10 μl sterile water, and loaded on 6% polyacrylamide gels with 7 M urea together with a 50 bp DNA size marker (ThermoScientific). Gels were stained with SYBRGold (Life Technologies), and fragments between 140 and 250 bp were excised from the gels. Elution of DNA fragments was performed in 500 μl DNA elution buffer (NEB) at 16°C overnight in a thermomixer at 1,000 rpm followed by EtOH precipitation. Pellets were resuspended in 10 μl sterile water. About 2 μl gel-purified DNA was mixed with 25 μl 2× LongAmp Taq PCR master mix, 2 μl each of primer JVO-11007 and JVO-11008 (10 μM), and 19 μl sterile water and amplified using the following program: 30 s at 94°C, 6 rounds of (15 s at 94°C, 30 s at 60°C, and 15 s at 65°C). PCRs were purified on columns (QIAGEN) and eluted in 15 μl sterile water.

Sequencing

High-throughput sequencing was performed at vertis Biotechnologie AG, Freising, Germany. Twelve cDNA libraries were pooled on an Illumina NextSeq 500 mid-output flow cell and sequenced in paired-end mode (2 × 75 cycles). Raw sequencing reads in FASTQ format and coverage files normalized by DESeq2 size factors are available via Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under accession number GSE74425.

Processing of sequence reads and mapping

To assure high sequence quality, read 1 (R1) and read 2 (R2) files containing the Illumina paired-end reads in FASTQ format were trimmed independently from each other with a Phred score cutoff of 20 by the program fastq_quality_trimmer from FASTX toolkit version 0.0.13. In the same step, after quality trimming NEB, R1 and R2 3′-adapters (R1: AGATCGGAAGAGCACACGTCTGAACTCCAGTC AC, R2: GATCGTCGGACTGTAGAACTCTGAACGTGTAGATCTCGGT GGTCGCCGTATCATT) were trimmed using Cutadapt version 1.7.1 (Martin, 2011) and reads without any remaining bases were discarded. Afterward, reads without a mate in the complementary read file were excluded using cmpfastq (http://compbio.brc.iop.kcl.ac.uk/software/cmpfastq.php). In order to remove putative PCR duplicates, paired-end reads were collapsed using FastUniq (Xu *et al*, 2012). Subsequently, a size filtering step was applied in which read pairs with at least one read shorter than 12 nt or longer than 25 nt were eliminated. The collections of remaining reads were mapped to the *Salmonella* Typhimurium SL1344 chromosome (NCBI Acc.-No: NC_016810.1) and plasmid (NCBI Acc.-No: NC_017718.1, NC_017719.1, NC_017720.1) reference sequences using the RNA-seq pipeline READemption version 0.3.5 (Förstner *et al*, 2014) and segemehl version 0.2.0 (Hoffmann *et al*, 2014) with an accuracy cutoff of 80%. From the results, only reads mapping uniquely to one genomic position were considered for all subsequent analysis. Pearson correlations between all libraries were calculated on nucleotide read coverage (Appendix Fig S5).

Coverage plots representing the numbers of mapped reads per nt were generated for each replicon and strand to facilitate data visualization in a genome browser. Each resulting cDNA coverage graph was normalized using the DESeq2 (Love *et al*, 2014) size factors calculated during peak calling.

For all analyses related to annotated genomic features such as CDSs, tRNAs, and rRNAs, gene annotations from NCBI were used. We defined *ad hoc* transcriptional units (TUs) based on NCBI CDS annotations, transcription start site (TSS) annotations from Kröger *et al* (2013) and Rho-independent terminator predictions by RNIE (Gardner *et al*, 2011). Briefly, TUs were defined as starting on annotated primary TSSes and ending either with a predicted Rho-independent terminator or in the presence of an intergenic gap greater than 500 nt on the coding strand. In the absence of an upstream TSS, an arbitrary 100 nt 5′UTR was added upstream of the first CDS in the TU, and similarly in the absence of a terminator, an arbitrary 100 nt 3′UTR was added. In the event of a predicted primary TSS within an intergenic gap of less than 500 nt on the coding strand, the TU was ended 100 nt downstream of the preceding CDS, or at the end of the preceding CDS if the predicted primary TSS was less than 100 nt downstream. We defined 5′UTRs as the

regions from the start of each predicted TU to the position upstream of the first CDS in the TU and 3′UTRs as the regions from one nt downstream of the last CDS in the TU to the end of the TU. sRNA annotations are based on Perkins *et al* (2009), Chinni *et al* (2010), Kröger *et al* (2013), and KU Förstner and J Vogel (unpublished data).

Peak calling

Peak calling was performed as a two-step process. In the first step, we defined peak regions using the blockbuster algorithm for defining discrete blocks of overlapping reads (Langenberger *et al*, 2009) across all crosslinked libraries for each RNA-binding protein investigated. Mapped and collapsed reads were filtered to only contain properly paired reads. The resulting BAM files were converted to BED format using BEDTools (v2.17.0) (Quinlan & Hall, 2010). These BED files were concatenated for all crosslinked libraries. Subsequently, each read pair in the concatenated BED file was merged into a single unit representing the sequenced RNA fragment. Only fragments ≤ 25 nt and ≥ 12 nt were retained for further analysis. The resulting BED file was reformatted to satisfy the blockbuster input specifications. Blockbuster uses a greedy approach based on a Gaussian smoothing of read profiles to identify clusters of overlapping read blocks. For this procedure, we required blocks to contain at least 10 reads (i.e., the minBlockHeight option was set to 10) and clusters had to be separated by at least one base (i.e., the distance parameter was set to 1). This procedure resulted in a large set of clusters consisting of overlapping blocks of reads. We then iteratively decomposed each cluster of overlapping blocks into peaks, taking into consideration the local frequency of read counts within the cluster. We first selected the block with the highest read count from the cluster under consideration. All blocks that overlapped with this block were removed from the cluster, and a peak was defined using these overlapping blocks. This procedure, of selecting the next largest block, was repeated in the reduced cluster until no more blocks were left that contained greater than 1% of the total cluster read count (see Appendix Supplementary Methods for a formalized description of this procedure).

In the second step of our peak calling analysis, we applied DESeq2 (v1.2.10) (Love *et al*, 2014) to test each peak for a reproducible relative read count enrichment in triplicate crosslinked libraries compared to non-crosslinked controls. Reads per peak were counted using HTSeq-count (v 0.6.1p1) (Anders *et al*, 2015) for all libraries with the mode option set to "union", the order option set to "name" and the stranded option set to "yes". DESeq2 was then run with default options in R. We considered peaks genuine if they had a normalized average expression of ≥ 10 in the crosslinked libraries and a statistically significant enrichment in crosslinked libraries compared to non-crosslinked controls, defined as a false discovery rate (FDR) corrected *P*-value of 0.1 or less.

CopraRNA–Hfq peaks overlap

CopraRNA (Wright *et al*, 2013, 2014) target predictions were performed for all sRNAs from the benchmark dataset of (Wright *et al*, 2013) that had an associated Hfq peak in our data (that is, all except RyhB). Two hundred nucleotides upstream and 100 nucleotides downstream of annotated start codons were specified

as potential target regions. The top 20 CopraRNA predictions for each sRNA candidate were subsequently intersected with mRNA candidates that show an Hfq peak in our data. To test for enrichment of known targets in the intersected lists, the number of known targets in the unfiltered top 20 CopraRNA predictions and the number of known targets in the lists resulting from the intersection were compared. The benchmark dataset (Wright *et al*, 2013) was considered as a reference for verified targets and was extended with the interactions between Spot42-glpF (Beisel *et al*, 2012), OxyS-cspC (Tjaden *et al*, 2006), and RybB-STM1530 (Wright, 2012). The unfiltered list of top 20 predictions for 17 individual target predictions contains 51 verified targets in a total list of length 340. The filtered list has a length of 48 and contains 19 verified targets. The interaction between Spot42–mglB discovered in this study was not used for enrichment analysis. A one-sided Fisher's exact test was employed to test for enrichment of known targets in the filtered list relative to the unfiltered list. The test was performed in R statistics using the Fisher's exact test function with the "alternative" parameter set to "greater". For this, we considered that 19 candidates are Hfq bound and verified, 29 candidates are Hfq bound and not verified, 32 candidates are not Hfq bound and verified and 260 candidates are not Hfq bound and not verified. Based on these numbers, the test matrix is given as matrix(c(19,32,29,260), nrow = 2, ncol = 2) in R notation. For the sake of simplicity, we considered targets verified in *E. coli* also to be targets in *Salmonella*. Even though this may not hold true for every single target, this is unlikely to change the principle findings of this analysis.

Analysis of crosslink-specific mutations

For the detection of crosslinking-induced mutation sites from the CLIP-seq data, only uniquely mapped, paired-end reads were considered and used for mutation calling using samtools (v 0.1.19). To reduce bias caused by sequencing errors, we required the mutated sites to be present in both paired reads. A python script adapted from the PIPE-CLIP package (Chen *et al*, 2014) was applied to identify sites significantly enriched in mutations in each library. The number of mutations at each position was modeled as the result of a Bernoulli process with p equal to the observed mutation rate across all positions. Positions were counted as significantly enriched in mutations if the probability of a mutation count greater than or equal to that observed at the position was less than 0.01 under the implied binomial distribution. The final requirement for a site to be considered enriched for crosslinking-induced mutations was that it had to be present in at least two of the libraries from the crosslinked samples and absent in all of the libraries from non-crosslinked samples.

Global analysis of binding regions

The peak density was calculated by counting the number of peaks along the specified annotation features, which included start codons in single-cistron mRNAs and in the first cistron in multigene operons, stop codons in single-cistron mRNAs and in the last cistron in operons, sRNAs, and predicted Rho-independent terminators. These features were retrieved from the extended *Salmonella* Typhimurium SL1344 annotation described above.

Analysis of sequence and structure motifs

The sequences of peaks or sequences 10 nucleotides upstream and downstream of crosslinking mutation sites were used for sequence motif identification using MEME (Bailey *et al*, 2015) with one base shift allowed while the remaining parameters were set at default values. To verify the specificity of the peak motifs found in Hfq peaks from 3′UTRs or sRNAs, the following analysis was performed: for each annotation feature with an Hfq peak, a sequence of the same length as the Hfq peak mapping to that feature but randomly positioned within the feature was extracted. This procedure was repeated ten times. The resulting sequences were used as input for MEME.

To search for the presence of a structural motif, CMfinder 0.2.1 (Yao *et al*, 2006) was run on sequences from peak regions extended by additional 10 nt upstream and downstream, using default parameters except for allowing a minimum single stem loop candidate length of 20 nt. The top-ranked motif incorporated 396 sequences while the motif detected most frequently was found in 416 of the 467 sequences. Both motifs were visualized using R2R (Weinberg & Breaker, 2011) and are depicted in Fig 6F.

Analysis of Hfq peaks in known sRNA–mRNA pairs

Distributions of Hfq peaks in sRNAs and mRNAs with validated basepair interaction sites (Wright *et al*, 2013) were calculated and visualized as a heat map using Excel. The interactions used were restricted to those mRNAs where an Hfq peak was detected within 100 nt on either side of a validated sRNA interaction site.

Pathway analysis

Pathway information was retrieved from the KEGG database (Kanehisa & Goto, 2000), the *Salmonella* SL1344 genome annotation (Kröger *et al*, 2012), and a selection of regulons curated from literature sources. Pathway enrichment analysis was performed using Fisher's exact test, and *P*-values were corrected for multiple testing using the Benjamini–Hochberg method.

Western blot

To analyze immunoprecipitated material in the CLIP experiments, one-tenth of the magnetic beads from each sample was resuspended in 10 μl protein loading buffer and heated 4 min at 95°C. The magnetic beads were collected on a magnetic separator, and the supernatant was loaded and separated on a 15% SDS–polyacrylamide gel followed by transfer of proteins to a nitrocellulose membrane. To detect FLAG-tagged proteins, the membrane was blocked in TBS-T with 5% milk powder, washed in TBS-T for 10 min, incubated for 1 h with anti-FLAG antibody (Sigma) diluted 1:1,000 dilution in TBS-T with 3% BSA, washed in TBS-T for 10 min, incubated for 1 h with anti-mouse-HRP antibody (ThermoScientific) diluted 1:10,000 dilution in TBS-T with 3% BSA, and finally washed in TBS-T two times for 10 min before adding the ECL substrate and taking captions with a CCD camera (ImageQuant, GE Healthcare). To analyze the expression of GFP fusion proteins, bacterial cultures were harvested at an OD_{600} of 1.0, and cell pellets were boiled in protein loading buffer and separated on 12% SDS–polyacrylamide gels. Proteins were transferred to PVDF membranes and GFP signals were detected as described above but using an anti-GFP antibody (Roche) followed by HRP-coupled anti-mouse antibody (ThermoScientific).

qRT–PCR

Total RNA was extracted using hot phenol, and contaminating DNA was removed by DNase I treatment. qRT–PCRs were carried out using the RNA-to-Ct 1-step kit (ThermoFisher) with 50 ng of RNA per reaction. Relative gene expression was calculated using the $\Delta\Delta C_t$ method (Livak & Schmittgen, 2001) by normalization to the *rfaH* mRNA.

Expanded View for this article is available online.

Author contributions

JV and EH designed the experiments. EH performed the experiments. PRW, TB, and RB implemented the peak calling strategy. LL implemented the calling of crosslink mutations and performed the motif analysis together with TB. PRW, LL, TB, LB, and EH carried out additional bioinformatical analyses. RR carried out sequencing and data analysis. EH and JV wrote the manuscript, which was discussed, modified, and improved by all authors. JV and RB supervised the project.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Altier C, Suyemoto M, Lawhon SD (2000) Regulation of Salmonella enterica serovar typhimurium invasion genes by csrA. *Infect Immun* 68: 6790 – 6797

Anders S, Pyl PT, Huber W (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166 – 169

Andrade JM, Pobre V, Matos AM, Arraiano CM (2012) The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *RNA* 18: 844 – 855

Ansong C, Yoon H, Porwollik S, Mottaz-Brewer H, Petritis BO, Jaitly N, Adkins JN, McClelland M, Heffron F, Smith RD (2009) Global systems-level analysis of Hfq and SmpB deletion mutants in Salmonella: implications for virulence and global protein translation. *PLoS ONE* 4: e4809

Ascano M, Hafner M, Cekan P, Gerstberger S, Tuschl T (2012) Identification of RNA-protein interaction networks using PAR-CLIP. *Wiley Interdiscip Rev RNA* 3: 159 – 177

Ascano M, Gerstberger S, Tuschl T (2013) Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. *Curr Opin Genet Dev* 23: 20 – 28

Babitzke P, Baker CS, Romeo T (2009) Regulation of translation initiation by RNA binding proteins. *Annu Rev Microbiol* 63: 27−44

Bailey TL, Johnson J, Grant CE, Noble WS (2015) The MEME Suite. *Nucleic Acids Res* 43: W39−49

Baker CS, Morozov I, Suzuki K, Romeo T, Babitzke P (2002) CsrA regulates glycogen biosynthesis by preventing translation of glgC in *Escherichia coli*. *Mol Microbiol* 44: 1599−1610

Baltz AG, Munschauer M, Schwanhausser B, Vasile A, Murakawa Y, Schueler M, Youngs N, Penfold-Brown D, Drew K, Milek M, Wyler E, Bonneau R, Selbach M, Dieterich C, Landthaler M (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell* 46: 674−690

Bandyra KJ, Said N, Pfeiffer V, Gorna MW, Vogel J, Luisi BF (2012) The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol Cell* 47: 943−953

Barquist L, Vogel J (2015) Accelerating discovery and functional analysis of small RNAs with new technologies. *Annu Rev Genet* 49: 367−394

Beisel CL, 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

Beisel CL, Updegrove TB, Janson BJ, Storz G (2012) Multiple factors dictate target selection by Hfq-binding small RNAs. *EMBO J* 31: 1961−1974

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

Brencic A, Lory S (2009) Determination of the regulon and identification of novel mRNA targets of Pseudomonas aeruginosa RsmA. *Mol Microbiol* 72: 612−632

Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, Krijgsveld J, Hentze MW (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* 149: 1393−1406

Chao Y, Papenfort K, Reinhardt R, Sharma CM, 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

Chao Y, 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

Chen B, Yun J, Kim MS, Mendell JT, Xie Y (2014) PIPE-CLIP: a comprehensive online tool for CLIP-seq data analysis. *Genome Biol* 15: R18

Chinni SV, Raabe CA, Zakaria R, Randau G, Hoe CH, Zemann A, Brosius J, Tang TH, Rozhdestvensky TS (2010) Experimental identification and characterization of 97 novel npcRNA candidates in Salmonella enterica serovar Typhi. *Nucleic Acids Res* 38: 5893−5908

Coornaert A, Chiaruttini C, Springer M, Guillier M (2013) Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS Genet* 9: e1003156

Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, Vogel J (2012) Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol Microbiol* 84: 428−445

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

De Lay N, Gottesman S (2012) A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. *Mol Microbiol* 86: 524−538

Desnoyers G, Masse E (2012) Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes Dev* 26: 726−739

Dimastrogiovanni D, Fröhlich KS, Bandyra KJ, Bruce HA, Hohensee S, Vogel J, Luisi BF (2014) Recognition of the small regulatory RNA RydC by the bacterial Hfq protein. *ELife* 3: e05375

Dubey AK, Baker CS, Romeo T, Babitzke P (2005) RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* 11: 1579−1587

Duss O, Michel E, Diarra Dit Konte N, Schubert M, Allain FH (2014a) Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucleic Acids Res* 42: 5332−5346

Duss O, Michel E, Yulikov M, Schubert M, Jeschke G, Allain FH (2014b) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* 509: 588−592

Edwards AN, Patterson-Fortin LM, Vakulskas CA, Mercante JW, Potrykus K, Vinella D, Camacho MI, Fields JA, Thompson SA, Georgellis D, Cashel M, Babitzke P, Romeo T (2011) Circuitry linking the Csr and stringent response global regulatory systems. *Mol Microbiol* 80: 1561−1580

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

Feng L, Rutherford ST, Papenfort K, Bagert JD, van Kessel JC, Tirrell DA, Wingreen NS, Bassler BL (2015) A qrr noncoding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell* 160: 228−240

Figueira R, Watson KG, Holden DW, Helaine S (2013) Identification of salmonella pathogenicity island-2 type III secretion system effectors involved in intramacrophage replication of S. enterica serovar typhimurium: implications for rational vaccine design. *MBio* 4: e00065

Figueroa-Bossi N, Lemire S, Maloriol D, Balbontin R, Casadesus J, Bossi L (2006) Loss of Hfq activates the sigmaE-dependent envelope stress response in Salmonella enterica. *Mol Microbiol* 62: 838−852

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

Figueroa-Bossi N, Schwartz A, Guillemardet B, D'Heygere F, Bossi L, Boudvillain M (2014) RNA remodeling by bacterial global regulator CsrA promotes Rho-dependent transcription termination. *Genes Dev* 28: 1239−1251

Förstner KU, Vogel J, Sharma CM (2014) READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics* 30: 3421−3423

Friedersdorf MB, Keene JD (2014) Advancing the functional utility of PAR-CLIP by quantifying background binding to mRNAs and lncRNAs. *Genome Biol* 15: R2

Gardner PP, Barquist L, Bateman A, Nawrocki EP, Weinberg Z (2011) RNIE: genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res* 39: 5845−5852

Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA (2011) Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc Natl Acad Sci USA* 108: 12875−12880

Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G (2014) MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev* 28: 1620−1634

Hajnsdorf E, Regnier P (2000) Host factor Hfq of *Escherichia coli* stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proc Natl Acad Sci USA* 97: 1501−1505

Hébrard M, Kröger C, Srikumar S, Colgan A, Händler K, Hinton JC (2012) sRNAs and the virulence of Salmonella enterica serovar Typhimurium. *RNA Biol* 9: 437−445

Hensel M (2004) Evolution of pathogenicity islands of Salmonella enterica. *Int J Med Microbiol* 294: 95–102

Heroven AK, Bohme K, Dersch P (2012) The Csr/Rsm system of Yersinia and related pathogens: a post-transcriptional strategy for managing virulence. *RNA Biol* 9: 379–391

Hoffmann S, Otto C, Doose G, Tanzer A, Langenberger D, Christ S, Kunz M, Holdt LM, Teupser D, Hackermuller J, Stadler PF (2014) A multi-split mapping algorithm for circular RNA, splicing, trans-splicing and fusion detection. *Genome Biol* 15: R34

Holmqvist E, Reimegård J, Sterk M, Grantcharova N, Römling U, Wagner EGH (2010) Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO J* 29: 1840–1850

Holmqvist E, Vogel J (2013) A small RNA serving both the Hfq and CsrA regulons. *Genes Dev* 27: 1073–1078

Ishikawa H, Otaka H, Maki K, Morita T, Aiba H (2012) The functional Hfq-binding module of bacterial sRNAs consists of a double or single hairpin preceded by a U-rich sequence and followed by a 3′ poly(U) tail. *RNA* 18: 1062–1074

Jørgensen MG, Nielsen JS, Boysen A, Franch T, Møller-Jensen J, Valentin-Hansen P (2012) Small regulatory RNAs control the multi-cellular adhesive lifestyle of *Escherichia coli*. *Mol Microbiol* 84: 36–50

Jørgensen MG, Thomason MK, Havelund J, Valentin-Hansen P, Storz G (2013) Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev* 27: 1132–1145

Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27–30

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059–3066

Kawamoto H, Koide Y, Morita T, Aiba H (2006) Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol* 61: 1013–1022

Kawano M, Reynolds AA, Miranda-Rios J, Storz G (2005) Detection of 5′- and 3′-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res* 33: 1040–1050

König J, Zarnack K, Luscombe NM, Ule J (2011) Protein-RNA interactions: new genomic technologies and perspectives. *Nat Rev Genet* 13: 77–83

Koo JT, Alleyne TM, Schiano CA, Jafari N, Lathem WW (2011) Global discovery of small RNAs in Yersinia pseudotuberculosis identifies Yersinia-specific small, noncoding RNAs required for virulence. *Proc Natl Acad Sci USA* 108: E709–E717

Kramer K, Sachsenberg T, Beckmann BM, Qamar S, Boon KL, Hentze MW, Kohlbacher O, Urlaub H (2014) Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nat Methods* 11: 1064–1070

Kröger C, Dillon SC, Cameron AD, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hebrard M, Handler K, Colgan A, Leekitcharoenphon P, Langridge GC, Lohan AJ, Loftus B, Lucchini S, Ussery DW, Dorman CJ, Thomson NR, Vogel J et al (2012) The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. *Proc Natl Acad Sci USA* 109: E1277–1286

Kröger C, Colgan A, Srikumar S, Handler K, Sivasankaran SK, Hammarlof DL, Canals R, Grissom JE, Conway T, Hokamp K, Hinton JC (2013) An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium. *Cell Host Microbe* 14: 683–695

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

Langenberger D, Bermudez-Santana C, Hertel J, Hoffmann S, Khaitovich P, Stadler PF (2009) Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics* 25: 2298–2301

Lapouge K, Perozzo R, Iwaszkiewicz J, Bertelli C, Zoete V, Michielin O, Scapozza L, Haas D (2013) RNA pentaloop structures as effective targets of regulators belonging to the RsmA/CsrA protein family. *RNA Biol* 10: 1031–1041

Lawhon SD, Frye JG, Suyemoto M, Porwollik S, McClelland M, Altier C (2003) Global regulation by CsrA in Salmonella typhimurium. *Mol Microbiol* 48: 1633–1645

Le Derout J, Folichon M, Briani F, Deho G, Regnier P, Hajnsdorf E (2003) Hfq affects the length and the frequency of short oligo(A) tails at the 3′ end of *Escherichia coli* rpsO mRNAs. *Nucleic Acids Res* 31: 4017–4023

Lease RA, Woodson SA (2004) Cycling of the Sm-like protein Hfq on the DsrA small regulatory RNA. *J Mol Biol* 344: 1211–1223

Lenz DH, Miller MB, Zhu J, Kulkarni RV, Bassler BL (2005) CsrA and three redundant small RNAs regulate quorum sensing in Vibrio cholerae. *Mol Microbiol* 58: 1186–1202

Link TM, Valentin-Hansen P, Brennan RG (2009) Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA. *Proc Natl Acad Sci USA* 106: 19292–19297

Liu MY, Yang H, Romeo T (1995) The product of the pleiotropic *Escherichia coli* gene csrA modulates glycogen biosynthesis via effects on mRNA stability. *J Bacteriol* 177: 2663–2672

Liu MY, Gui G, Wei B, Preston JF III, Oakford L, Yuksel U, Giedroc DP, Romeo T (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem* 272: 17502–17510

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402–408

Lopez-Garrido J, Puerta-Fernandez E, Casadesus J (2014) A eukaryotic-like 3′ untranslated region in Salmonella enterica hilD mRNA. *Nucleic Acids Res* 42: 5894–5906

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550

Majdalani N, Vanderpool CK, Gottesman S (2005) Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol* 40: 93–113

Mark Glover JN, Chaulk SG, Edwards RA, Arthur D, Lu J, Frost LS (2015) The FinO family of bacterial RNA chaperones. *Plasmid* 78: 79–87

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* 17: 10–12

Martinez LC, Yakhnin H, Camacho MI, Georgellis D, Babitzke P, Puente JL, Bustamante VH (2011) Integration of a complex regulatory cascade involving the SirA/BarA and Csr global regulatory systems that controls expression of the Salmonella SPI-1 and SPI-2 virulence regulons through HilD. *Mol Microbiol* 80: 1637–1656

Mercante J, Suzuki K, Cheng X, Babitzke P, Romeo T (2006) Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance. *J Biol Chem* 281: 31832–31842

Mika F, Busse S, Possling A, Berkholz J, Tschowri N, Sommerfeldt N, Pruteanu M, Hengge R (2012) Targeting of csgD by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in *Escherichia coli*. *Mol Microbiol* 84: 51–65

Mikulecky PJ, Kaw MK, Brescia CC, Takach JC, Sledjeski DD, Feig AL (2004) *Escherichia coli* Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs. *Nat Struct Mol Biol* 11: 1206–1214

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

Miyakoshi M, Chao Y, Vogel J (2015b) Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Curr Opin Microbiol* 24: 132–139

Møller T, Franch T, Hojrup P, Keene DR, Bachinger HP, Brennan RG, Valentin-Hansen P (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* 9: 23–30

Morita T, Aiba H (2011) RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes Dev* 25: 294–298

Otaka H, Ishikawa H, Morita T, 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

Panja S, Schu DJ, Woodson SA (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange. *Nucleic Acids Res* 41: 7536–7546

Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J (2013) Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* 153: 426–437

Papenfort K, Förstner KU, Cong JP, Sharma CM, Bassler BL (2015) Differential RNA-seq of Vibrio cholerae identifies the VqmR small RNA as a regulator of biofilm formation. *Proc Natl Acad Sci USA* 112: E766–775

Peng Y, Soper TJ, Woodson SA (2014) Positional effects of AAN motifs in rpoS regulation by sRNAs and Hfq. *J Mol Biol* 426: 275–285

Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR, Dougan G (2009) A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi. *PLoS Genet* 5: e1000569

Phadtare S, Alsina J, Inouye M (1999) Cold-shock response and cold-shock proteins. *Curr Opin Microbiol* 2: 175–180

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842

Rice JB, Balasubramanian D, Vanderpool CK (2012) Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mRNA. *Proc Natl Acad Sci USA* 109: E2691–2698

Romeo T, Gong M, Liu MY, Brun-Zinkernagel AM (1993) Identification and molecular characterization of csrA, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol* 175: 4744–4755

Romeo T, Vakulskas CA, Babitzke P (2013) Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ Microbiol* 15: 313–324

Salim NN, Feig AL (2010) An upstream Hfq binding site in the fhlA mRNA leader region facilitates the OxyS-fhlA interaction. *PLoS ONE* 5: e13028

Salim NN, Faner MA, Philip JA, Feig AL (2012) Requirement of upstream Hfq-binding (ARN)x elements in glmS and the Hfq C-terminal region for GlmS upregulation by sRNAs GlmZ and GlmY. *Nucleic Acids Res* 40: 8021–8032

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

Sauer E, Schmidt S, Weichenrieder O (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition. *Proc Natl Acad Sci USA* 109: 9396–9401

Schu DJ, Zhang A, Gottesman S, Storz G (2015) Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. *EMBO J* 34: 2557–2573

Schumacher MA, Pearson RF, Möller T, Valentin-Hansen P, Brennan RG (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein. *EMBO J* 21: 3546–3556

Sittka A, Pfeiffer V, Tedin K, Vogel J (2007) The RNA chaperone Hfq is essential for the virulence of Salmonella typhimurium. *Mol Microbiol* 63: 193–217

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

Soper TJ, Doxzen K, Woodson SA (2011) Major role for mRNA binding and restructuring in sRNA recruitment by Hfq. *RNA* 17: 1544–1550

Sterzenbach T, Nguyen KT, Nuccio SP, Winter MG, Vakulskas CA, Clegg S, Romeo T, Bäumler AJ (2013) A novel CsrA titration mechanism regulates fimbrial gene expression in Salmonella typhimurium. *EMBO J* 32: 2872–2883

Sugimoto Y, König J, Hussain S, Zupan B, Curk T, Frye M, Ule J (2012) Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions. *Genome Biol* 13: R67

Thomason MK, Fontaine F, De Lay N, Storz G (2012) A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol Microbiol* 84: 17–35

Tjaden B, Goodwin SS, Opdyke JA, Guillier M, Fu DX, Gottesman S, Storz G (2006) Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res* 34: 2791–2802

Tree JJ, Granneman S, McAteer SP, Tollervey D, Gally DL (2014) Identification of bacteriophage-encoded anti-sRNAs in pathogenic *Escherichia coli*. *Mol Cell* 55: 199–213

Updegrove TB, Wartell RM (2011) The influence of *Escherichia coli* Hfq mutations on RNA binding and sRNA*mRNA duplex formation in rpoS riboregulation. *Biochim Biophys Acta* 1809: 532–540

Updegrove TB, Shabalina SA, Storz G (2015) How do base-pairing small RNAs evolve? *FEMS Microbiol Rev* 39: 379–391

Uren PJ, Bahrami-Samani E, Burns SC, Qiao M, Karginov FV, Hodges E, Hannon GJ, Sanford JR, Penalva LO, Smith AD (2012) Site identification in high-throughput RNA-protein interaction data. *Bioinformatics* 28: 3013–3020

Vakulskas CA, Potts AH, Babitzke P, Ahmer BM, Romeo T (2015) Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiol Mol Biol Rev* 79: 193–224

Valverde C, Lindell M, Wagner EG, Haas D (2004) A repeated GGA motif is critical for the activity and stability of the riboregulator RsmY of Pseudomonas fluorescens. *J Biol Chem* 279: 25066–25074

Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jäger JG, Hüttenhofer A, Wagner EG (2003) RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* 31: 6435–6443

Vogel J (2009) A rough guide to the non-coding RNA world of Salmonella. *Mol Microbiol* 71: 1–11

Vogel J, Luisi BF (2011) Hfq and its constellation of RNA. *Nat Rev Microbiol* 9: 578–589

Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, Baker CS, Georgellis D, Babitzke P, Romeo T (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol* 48: 657–670

Weinberg Z, Breaker RR (2011) R2R–software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* 12: 3

Westermann AJ, Förstner KU, Amman F, Barquist L, Chao Y, Schulte LN, Müller L, Reinhardt R, Stadler PF, Vogel J (2016) Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* 529: 496–501

Wilson KS, von Hippel PH (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc Natl Acad Sci USA* 92: 8793–8797

Wilusz CJ, Wilusz J (2005) Eukaryotic Lsm proteins: lessons from bacteria. *Nat Struct Mol Biol* 12: 1031 – 1036

Wright PR (2012) hIntaRNA – Comparative prediction of sRNA targets in prokaryotes. Diploma, Albert Ludwig University Freiburg, Freiburg, Germany

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

Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR, Backofen R (2014) CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res* 42: W119 – 123

Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S (2012) FastUniq: a fast *de novo* duplicates removal tool for paired short reads. *PLoS ONE* 7: e52249

Yakhnin AV, Baker CS, Vakulskas CA, Yakhnin H, Berezin I, Romeo T, Babitzke P (2013) CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. *Mol Microbiol* 87: 851 – 866

Yao Z, Weinberg Z, Ruzzo WL (2006) CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* 22: 445 – 452

Zhang A, Wassarman KM, Ortega J, Steven AC, Storz G (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs. *Mol Cell* 9: 11 – 22

Zhang C, Darnell RB (2011) Mapping *in vivo* protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat Biotechnol* 29: 607 – 614

Zhang A, Schu DJ, Tjaden BC, Storz G, Gottesman S (2013) Mutations in interaction surfaces differentially impact *E. coli* Hfq association with small RNAs and their mRNA targets. *J Mol Biol* 425: 3678 – 3697

Zheng D, Constantinidou C, Hobman JL, Minchin SD (2004) Identification of the CRP regulon using *in vitro* and *in vivo* transcriptional profiling. *Nucleic Acids Res* 32: 5874 – 5893

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406 – 3415