Hfq CLASH uncovers sRNA-target interaction networks involved in adaptation to nutrient availability

Ira A. Iosub[1], Marta Marchioretto[2], Brandon Sy[4], Stuart McKellar[1], Karen J. Nieken[3], Rob W. van Nues[3], Jai J. Tree[4], Gabriella Viero[2] and Sander Granneman[1*]

Affiliations:

[1]Centre for Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3BF, UK.

[2]Institute of Biophysics, CNR Unit at Trento, Italy.

[3]Institute of Cell Biology, University of Edinburgh, Edinburgh EH9 3FF, UK.

[4]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney 2052, NSW, Australia.

*To whom correspondence should be addressed:
Sander Granneman
e-mail: sgrannem@ed.ac.uk
Tel: +44 131 6519082

Abstract

By shaping gene expression profiles, small RNAs (sRNAs) enable bacteria to very efficiently adapt to constant changes in their environment. To better understand how *Escherichia coli* acclimatizes to changes in nutrient availability, we performed UV cross-linking, ligation and sequencing of hybrids (CLASH) to uncover sRNA-target interactions. Strikingly, we uncovered hundreds of novel Hfq-mediated sRNA-target interactions at specific growth stages, involving many novel 3'UTR-derived sRNAs and a plethora of sRNA-sRNA interactions. We discovered sRNA-target interaction networks that play a role in adaptation to changes in nutrient availability. We characterized a novel 3'UTR-derived sRNA (MdoR), which is part of a regulatory cascade that enhances maltose uptake by (a) inactivating repressive pathways that block the accumulation of maltose transporters and (b) by reducing the flux of general porins to the outer membrane. Our work provides striking examples of how bacteria utilize sRNAs to integrate multiple regulatory pathways to enhance nutrient stress adaptation.

Microorganisms are renowned for their ability to adapt to environmental changes by rapidly rewiring their gene expression program. These responses are mediated through integrated transcriptional and post-transcriptional networks. Control at the transcriptional level dictates which genes are expressed (Balleza et al., 2009; Martínez-Antonio et al., 2008) and is well-characterised in *Escherichia coli*. Post-transcriptional regulation is key for controlling adaptive responses. By using riboregulators and RNA-binding proteins (RBPs), cells can efficiently integrate multiple pathways and incorporate additional signals into regulatory circuits. *E. coli* employs many post-transcriptional regulators, including small regulatory RNAs (sRNAs (Waters and Storz, 2009)), *cis*-acting RNAs (Kortmann and Narberhaus, 2012), and RNA binding proteins (RBPs) (Holmqvist and Vogel, 2018). The sRNAs are the largest class of bacterial regulators, which work in tandem with RBPs to regulate their RNA targets (Storz et al., 2011; Waters and Storz, 2009). The base-pairing interactions are often mediated by RNA chaperones such as Hfq and ProQ, which help to anneal or stabilize the sRNA and sRNA-target duplex (Smirnov et al., 2017, 2016; Updegrove et al., 2016). Small RNAs can repress or stimulate translation and transcription, as well as control mRNA stability (Sedlyarova et al., 2016; Updegrove et al., 2016; Vogel and Luisi, 2011; Waters and Storz, 2009).

During growth in rich media, *E. coli* are exposed to continuously changing conditions, such as fluctuations in nutrient availability, pH and osmolarity. Consequently, *E. coli* elicit complex responses that result in physiological and behavioural changes such as envelope composition remodelling, quorum sensing, nutrient scavenging, swarming and biofilm formation. Even subtle changes in the growth conditions can trigger rapid adaptive responses. Accordingly, each stage of the growth curve is characterised by different physiological states driven by activation of different transcriptional and post-transcriptional networks. Moreover, growth phase dependency of virulence and pathogenic behaviour has been demonstrated in both Gram-positive and Gram-negative bacteria. In some cases a particular growth stage is non-permissive for the induction of virulence (Mäder et al., 2016; Mouali et al., 2018). Although the exponential and stationary phases have been characterised in detail(Navarro Llorens et al., 2010; Pletnev et al., 2015), little is known about the transition between these two phases. During this transition, the cell population starts to scavenge alternative carbon sources, which requires rapid remodelling of their transcriptome (Baev et al., 2006a, 2006b; Sezonov et al., 2007).

To understand sRNA-mediated adaptive responses, detailed knowledge of the underlying post-transcriptional circuits is required. In *E. coli*, hundreds of sRNAs have been discovered, but only a fraction of these have been characterised. A key step to unravel the roles of sRNAs in regulating adaptive responses is to identify their targets. To tackle this globally, high-throughput methods have been developed that have uncovered a plethora of

77 sRNA-target interactions, many more than previously anticipated (Han et al., 2016; Hör et al.,
78 2018; Hör and Vogel, 2017; Lalaouna et al., 2015; Melamed et al., 2016; Waters et al., 2016).
79 To uncover sRNA-target RNA interaction dynamics that take place during the entry
80 into stationary phase, we applied UV cross-linking, ligation and sequencing of hybrids
81 (CLASH) (Helwak et al., 2013; Kudla et al., 2011) on *E. coli*. First, we demonstrate that the
82 highly stringent purification steps make CLASH a robust method for direct mapping of Hfq-
83 mediated sRNA-target interactions in *E. coli*. This enabled us to significantly expand on the
84 sRNA-target interaction repertoire found by RNase E CLASH (Waters et al., 2016) and RIL-
85 seq (Melamed et al., 2016), and we show that Hfq CLASH can generate very reliable results.
86 Using CLASH we identified many potentially novel 3'UTR-derived sRNAs, confirming that this
87 class of sRNAs (Chao et al., 2012, 2017; Chao and Vogel, 2016; Miyakoshi et al., 2015a) is
88 highly prevalent.
89 Next, we focussed our analyses on interactions that were specifically recovered during
90 the transition phase where we identified a surprisingly large number of interactions, including
91 sRNA-sRNA interactions. Our data suggests that during the transition stage, ArcZ represses
92 CyaR levels, thereby indirectly controlling genes involving nutrient uptake during the transition
93 phase. We also characterized a novel 3'UTR-derived sRNA, which we refer to as MdoR (*mal*-
94 dependent OMP repressor). Unlike the majority of bacterial sRNAs, MdoR is transiently
95 expressed during the transition phase. We demonstrate that MdoR is a degradation
96 intermediate of the *malG* 3'UTR, the last transcript of the *malEFG* polycistron that encodes
97 components of the maltose transport system. We show that MdoR directly downregulates
98 several mRNAs encoding major porins and suppresses the envelope stress response
99 controlled by σ^E. We propose that MdoR is part of a regulatory network that, during the
100 transition phase, promotes accumulation of high affinity maltose transporters in the outer
101 membrane by repressing competing pathways.
102

103 # Results

104 **Hfq CLASH in *E. coli.***

105 To unravel the post-transcriptional networks that underlie the transition between
106 exponential and stationary growth phases in *E. coli*, we performed CLASH (Helwak et al.,
107 2013; Kudla et al., 2011) using Hfq as bait (Figure 1A). To generate high quality Hfq CLASH
108 data made a number of improvements to the original protocol used for RNase E CLASH
109 (Waters et al., 2016).
110

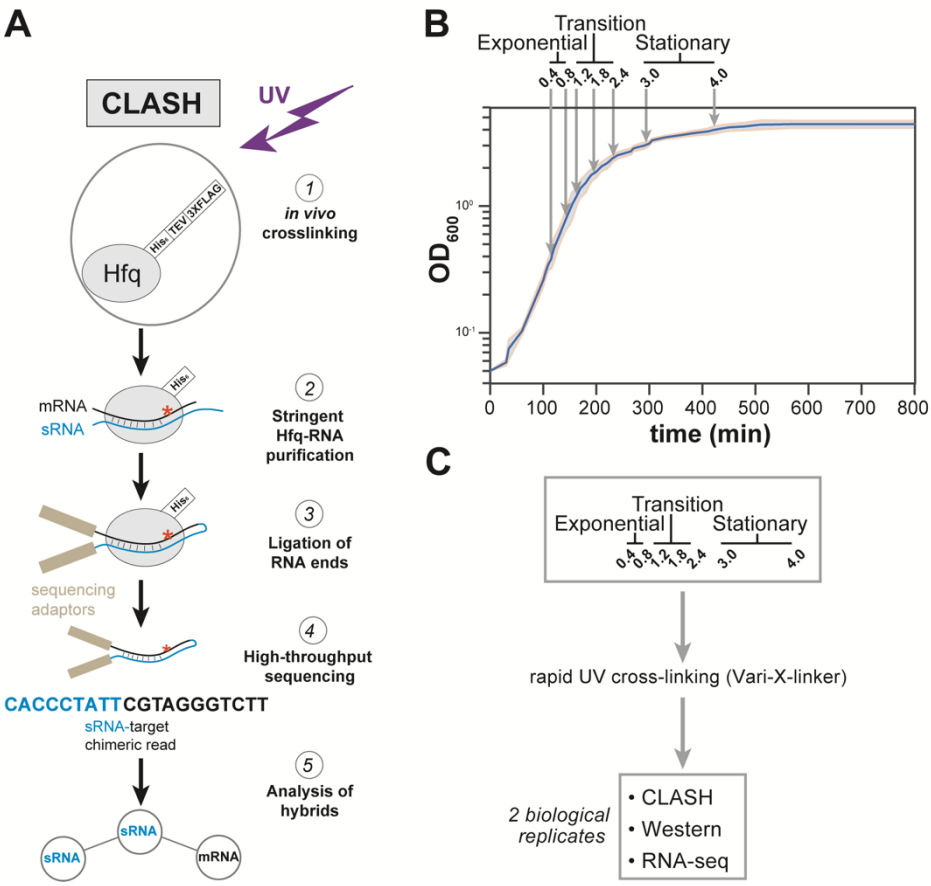


Figure 1. Hfq CLASH experiments at different growth phases in *E. coli*.
(**A**) Overview of the critical experimental steps for obtaining the Hfq CLASH data. *E. coli* cells expressing an HTF (His6-TEV-3xFLAG)-tagged Hfq(Jai J. Tree et al., 2014) were grown in LB and an equal number of cells were harvested at different optical densities (OD_{600}). Hfq binds to sRNA-target RNA duplexes, and RNA ends that are in close proximity are ligated together. After removal of the protein, cDNA libraries are prepared and sequenced. The single reads can be used to map Hfq-RNA interactions, whereas the chimeric reads can be traced to sRNA-target interactions. (**B**) A growth curve of the cultures used for the Hfq CLASH experiments, with OD_{600} at which cells were cross-linked indicated by circles, and each growth stage is indicated above the plot. The results show the mean and standard deviations of two biological replicates. Source data are provided as a Source Data file. (**C**) Cultures at the same OD_{600} cross-linked and harvested by filtration were analysed by Hfq CLASH, RNA-seq and Western blotting to detect Hfq.

Our Hfq CLASH protocol has several advantages over the related RIL-seq method (see Materials and Methods and Discussion). As negative controls, replicate CLASH experiments were performed on the untagged parental strain. When combined, the control samples had ~10 times lower number of single-mapping reads and contained only 297 unique chimeric reads, compared to the over 70.000 chimeras identified in the tagged Hfq data. This demonstrates that the purification method produced very low background levels.

131 Cell samples from seven different optical densities were subjected to Hfq CLASH.
132 Based on the growth curve analysis shown in Figure 1B, we categorized OD_{600} densities 0.4
133 and 0.8 as exponential growth phase, 1.2, 1.8, 2.4 as the transition phase from exponential to
134 stationary, and 3.0 and 4.0 as early stationary phase. To complement the CLASH data, RNA-
135 seq and Western blot analysis was performed on UV-irradiated cells to quantify steady state
136 RNA and Hfq protein levels, respectively (Figure 1C, Figure 1-figure supplement 1,
137 Supplementary Table 1). Western blot analyses revealed that Hfq levels gradually increased
138 during growth, however, when normalized to the levels of the chaperone GroEL, the increase
139 was modest (Figure 1-figure supplement 1A-B). To determine the cross-linking efficiency, Hfq-
140 RNA complexes immobilized on nickel beads were radiolabelled, resolved on NuPAGE gels
141 and detected by autoradiography. The data show that the recovery of Hfq and radioactive
142 signal was comparable at each optical density studied (Figure 1-figure supplement 1C).
143 Comparison of normalized read counts of replicate CLASH and RNA-seq experiments showed
144 that the results were highly reproducible (Figure 1-figure supplement 2).
145
146 **Hfq binds to the transcriptome in a growth-stage dependent manner.**
147 Meta-analyses of the Hfq CLASH sequencing data revealed that the distribution of Hfq
148 binding across mRNAs was very similar at each growth stage. We observed the expected Hfq
149 enrichment at the 5'UTRs and at the 3'UTRs at each growth stage (see Figure 1-figure
150 supplement 3A and 3B for examples). After identifying significantly enriched Hfq binding peaks
151 (FDR <= 0.05; see Methods for details) we used the genomic coordinates of these peaks to
152 search for Hfq binding motifs in mRNAs. The most enriched k-mer included poly-U stretches
153 (Figure 1-figure supplement 3C) that resemble the poly-U tracts characteristic to Rho-
154 independent terminators found at the end of many bacterial transcripts (Wilson and Hippel,
155 1995), and confirms the motif uncovered in CLIP-seq studies in *Salmonella* (Holmqvist et al.,
156 2016).
157
158

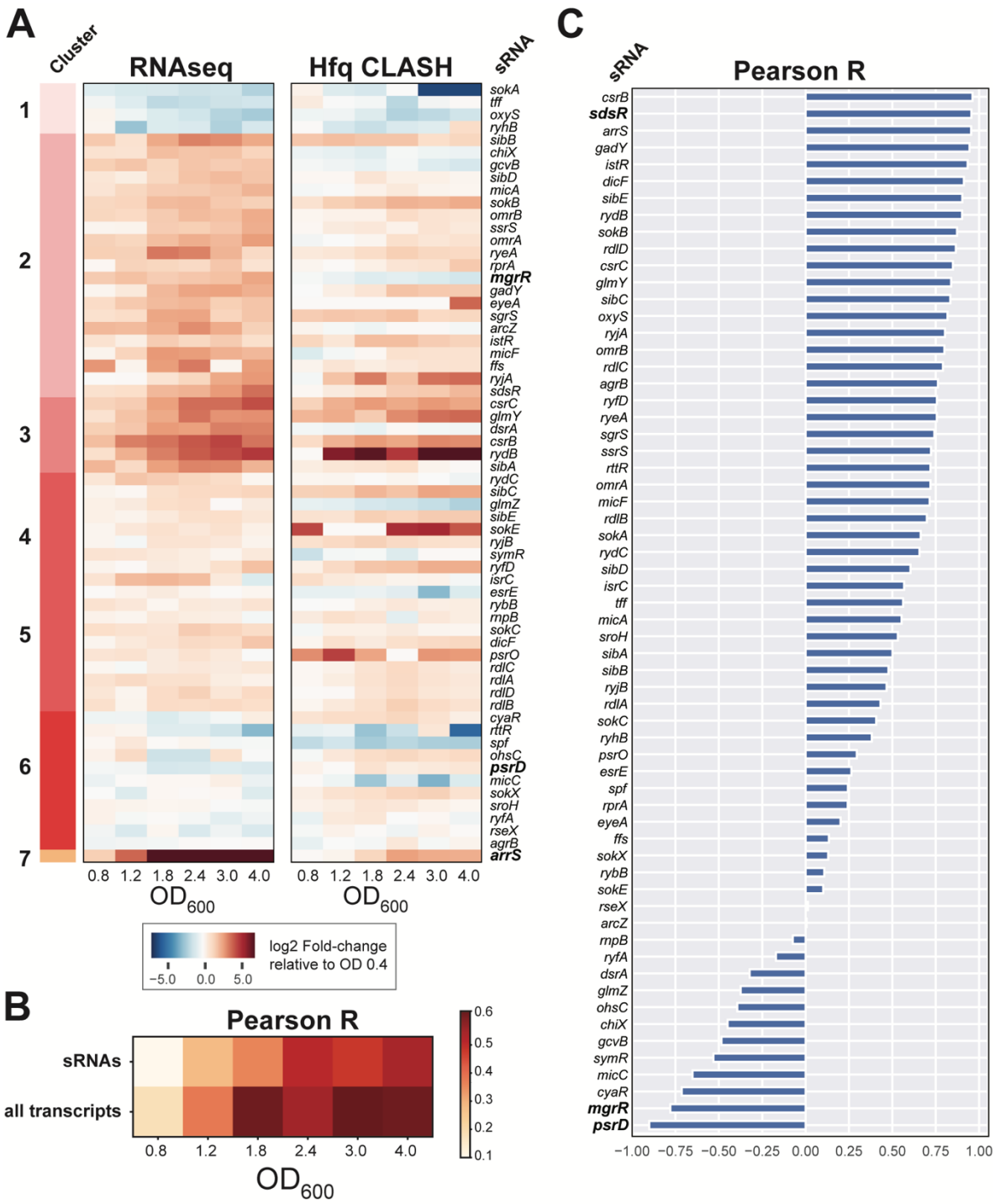
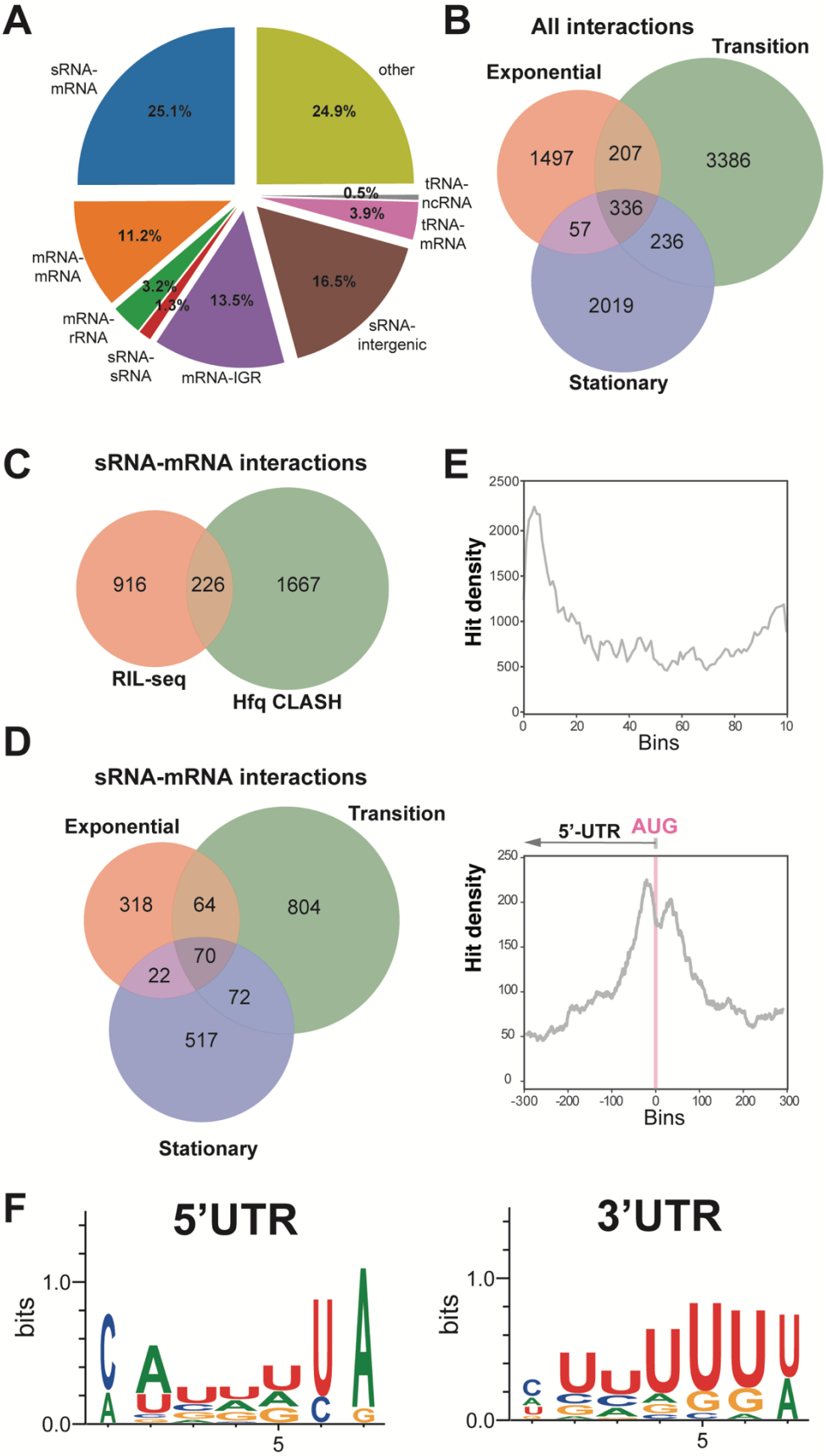


Figure 2. Hfq binding to sRNAs and sRNA steady state levels in *E. coli* do not always correlate highly.

(**A**) Heatmaps showing the changes in sRNA steady-states (Left) and Hfq cross-linking (Right) relative to OD_{600} 0.4. The clustered RNA-seq data were generated by k-means clustering using the fold-changes of normalized counts (log_2 of transcripts per million (TPM)) relative to OD_{600} 0.4. A blue shade indicates a reduction in levels compared to OD_{600} 0.4, whereas a dark orange shade indicates an increase. The cluster assignment is indicated at the left, and the growth stage is indicated as OD_{600} units below each heatmap. Note that ssrS and ffs encode the cytoplasmic RNAs 6S, a regulator of RNA polymerase, and 4.5S, the signal recognition particle RNA. EyeA is an uncharacterized sRNA mapped by(Sætrom et al., 2005). (**B**) Global correlation of between Hfq binding and steady state RNA levels

169 increases at higher cell densities. The heatmap shows the changes in Pearson R correlation between
170 Hfq binding and RNA expression (normalised as in (a)), for sRNAs (top) and all transcripts (bottom).
171 (**C**) Assessment of correlation between changes in expression and Hfq binding profiles for individual
172 sRNAs. The y-axis shows the gene name of the sRNAs and the Pearson coefficient indicating the
173 correlation between Hfq binding and steady state levels for each sRNA is shown on the x-axis.
174
175 Given the established role of Hfq in sRNA stabilization and mediating sRNA-target
176 interactions, it was logical to assume that changes in Hfq binding would also be reflected in
177 changes in sRNA steady-state levels. This would imply that the Hfq binding data would show
178 a strong correlation with the RNA-seq data. To test this, we compared the Hfq cross-linking
179 data to the RNA-seq data. K-means clustering of the normalized data revealed 7 different
180 patterns of changes in normalised read counts in the Hfq cross-linking and total RNA-seq data
181 (Figure 2A). Reminiscent of recent work performed in *Salmonella* (Chao et al., 2012), most
182 sRNAs in *E. coli* appear to be preferentially expressed when the cells reach the transition and
183 stationary phase (Figure 2A). However, much to our surprise, the Hfq cross-linking profile did
184 not always follow the same trend (Figure 2A-C, Figure 2-figure supplement 1A). Globally,
185 changes in sRNA expression did not correlate strongly with the Hfq-binding profile (and *vice*
186 *versa*) (Fig 2B-C). The correlation between changes in sRNA expression levels versus
187 changes in Hfq cross-linking was particularly poor at lower cell density (OD_{600} 0.8: r = 0.10)
188 but gradually improved as the cells approach stationary phase (OD_{600} 4.0 r =0.54; Figure 2B
189 and Figure 2-figure supplement 1A). A very similar result was obtained when comparing all
190 Hfq-bound RNAs, including mRNAs (Figure 2B and Figure 2-figure supplement 1B). Striking
191 examples are MgrR and PsrD, which showed a strong anti-correlation between Hfq CLASH
192 and RNA-seq counts (Figure 2C) In the case of PsrD, Hfq binding showed a modest increase
193 during the growth phase (Figure 2A; right heat map, cluster 6), whereas sRNA levels steadily
194 decreased (Figure 2A; left heat map; cluster 2). PsrD/SraB has also been shown to bind ProQ
195 (Smirnov et al., 2017), which may explain why its accumulation does not correlate with Hfq
196 binding. On the other side of the spectrum, SdsR showed a very high positive Pearson
197 correlation (Figure 2), suggesting that its accumulation heavily relies on Hfq binding.
198 We conclude that the dynamics of sRNA expression and binding to Hfq are not always
199 highly correlated.
200
201 **Hfq CLASH robustly detects RNA-RNA interactions.**
202 To get a complete overview of the RNA-RNA interactions captured by Hfq CLASH, we
203 merged the data from the two biological replicates of CLASH growth phase experiments
204 (Supplementary Table 2.1). Overlapping paired-end reads were merged and unique chimeric
205



206 **Figure 3. Hfq CLASH detects RNA-RNA interactions in *E. coli*.**

207 (**A**) Intermolecular transcript combinations found in interactions captured by Hfq CLASH. Combination
208 count of all uniquely annotated hybrids on genomic features. *tRNA-tRNA and rRNA-rRNA chimeras
209 originating from different coding regions were removed. (**B**) Venn diagram showing the intersection
210 between interactions from statistically filtered CLASH data from two biological replicates, recovered at
211 three main growth stages: exponential (OD_{600} 0.4 and 0.8), transition (OD_{600} 1.2, 1.8, 2.4) and early
212 stationary (OD_{600} 3.0 and 4.0). (**C**) Comparison of sRNA-mRNA interactions found in RIL-seq S-chimera
213 data and Hfq CLASH data. (**D**) Same as in (**B**) but for sRNA-mRNA interactions. (**E**) (Top) Distribution
214 of chimeras representing statistically filtered interactions, which uniquely map to mRNA genes.
215 Overlapping fragments of the two individual parts of the chimeras were collapsed into single clusters
216 followed by generation of distribution plots. Each gene was divided in 100 bins the number of clusters
217 that map to each bin (hit density; y-axis) was calculated; (Bottom) For the distribution plot around the
218 AUG, the gene length was normalized in 601 bins (x-axis) 5'-end overlap (-300) before the start of the
219 coding sequence, and 300 bins downstream AUG (+300); the bins corresponding to the start codon are
220 indicated with a pink line. (**F**) Enriched motifs in chimeras that uniquely overlap 5'UTRs and 3'UTRs;
221 the logos were drawn using the top 20 K-mers.
222
223 reads were identified using the hyb pipeline (Travis et al., 2013). To select RNA-RNA
224 interactions for further analysis, we applied a probabilistic analysis pipeline previously used
225 for the analysis of RNA-RNA interactions in human cells (Sharma et al., 2016) and adapted it
226 for the analyses of RNase E CLASH data (Waters et al., 2016). This pipeline tests the
227 likelihood that observed interactions could have formed spuriously. Strikingly, 87% of the
228 chimeric reads had a Benjamini-Hochberg adjusted p-value of 0.05 or less, indicating that it is
229 highly unlikely that these interactions were generated by random ligation of RNA molecules.
230 These analyses demonstrate the robustness of Hfq CLASH protocol. A complete overview of
231 statistically significantly enriched chimeras is provided in Supplementary Table 2.2.

232 The distribution of combinations of transcript classes found in the statistically filtered
233 chimeric reads indicates the sRNA-mRNA interactions as the most frequent recovered Hfq-
234 mediated interaction type (~25%; Figure 3A). We suspect that this number might be higher,
235 as about 16% of the chimeras contained sRNA and fragments that mapped to intergenic
236 regions (Figure 3A). Manual inspection of several of these indicated that some of the intergenic
237 sequences were located near genes for which the UTRs were unannotated or short. The vast
238 majority of these interactions were growth stage-specific (Figure 3B). Hfq CLASH identified
239 almost 2000 sRNA-mRNA interactions (Figure 3C; Supplementary Table 2). Around 20% of
240 the interactions found with RIL-seq and ~21% (27 out of 126) of the experimentally verified
241 interactions present in sRNATarbase3 (Supplementary Table 2.7) were recovered. These
242 results suggest that while the CLASH data contained known and many novel interactions, the
243 analyses clearly were not exhaustive.

244 Meta-analysis revealed that the majority of interactions were identified in the transition
245 phase (Figure 3D) and that the mRNA fragments found in chimeric reads were strongly
246 enriched in 5'UTRs peaking near the translational start codon (Figure 3E). The latter is
247 consistent with the canonical mode of translational inhibition by sRNAs (Bouvier et al., 2008)
248 and demonstrates the robustness of the data. Enrichment was also found in 3'UTRs of mRNAs
249 (Figure 3F). Motif analyses revealed a distinct sequence preference in 5'UTR and 3'UTR
250 binding sites (Figure 3G, Supplementary Table 3). The motifs enriched in the 5'UTRs chimeric
251 fragments are more consistent with Hfq binding to Shine Dalgarno-like $(ARN)_n$ sequences (Jai
252 J. Tree et al., 2014) and U-tracts, whereas the 3'UTR-containing chimera consensus motif
253 corresponds to poly-U transcription termination sites (Figure 3F and Supplementary Table 3).
254
255 **Hfq CLASH predicts sRNA-sRNA interactions as a widespread layer of post-**
256 **transcriptional regulation.**
257 We also uncovered a surprisingly large number of sRNA-sRNA interactions
258 (Supplementary Table 2.4), many of which were uniquely found in our Hfq CLASH data (Figure
259 4A). Many interactions were growth-stage specific and the sRNA-sRNA networks show
260 extensive rewiring across the exponential, transition and stationary phases (Figure 4-figure
261 supplement 1). The sRNA-sRNA network is dominated by several abundant sRNAs that
262 appear to act as hubs that have many interacting partners: ChiX, Spot42 (spf), ArcZ and GcvB.
263 In many cases the experimentally-validated sRNA seed sequences were found in the chimeric
264 reads, for both established and novel interactions. For example, the vast majority of ArcZ
265 sRNA-sRNA chimeras contained the known and well conserved seed sequence (Figure 4B,
266 Figure 4-figure supplement 2).
267 The sRNA-sRNA chimeras containing CyaR fragments were of particular interest, as
268 the sRNA is primarily expressed during the transition from late exponential to stationary phase
269 (De Lay and Gottesman, 2009). In the case of CyaR, the known seed sequence (De Lay and
270 Gottesman, 2009; Papenfort et al., 2008) as well as a conserved ~25 nt fragment in the 5'
271 region was found in chimeras (Figure 4-C; Figure 4-figure supplement 2). Similar seed
272 sequences were identified in CLASH experiments using RNase E as a bait (Waters et al.,
273 2016), suggesting that this region represents a *bona fide* interaction site. Notably, we identified
274 ArcZ-CyaR chimeras containing the seed sequence from both sRNAs (Figure 4-figure
275 supplement 2) and these were detected specifically in the transition phase (Figure 4B-C),
276 suggesting that these sRNAs could influence each other's activity. To validate these findings,
277 we used an *E. coli* plasmid-based assay that is routinely used to monitor sRNA-sRNA
278 interactions and expression of their target mRNAs (Melamed et al., 2016; Miyakoshi et al.,
279 2015b; Jai J. Tree et al., 2014). An advantage of this system is that each sRNA would be
280 uncoupled from the chromosomally encoded regulatory networks (that were thought to act

281 largely in a 1:1 stoichiometry) and to allow the specific effects of the sRNA-target RNA to be
282 assessed (Miyakoshi et al., 2015b). Importantly, these sRNAs were induced during early
283 exponential growth phase when the endogenous (processed) ArcZ and CyaR sRNAs are
284 detectable at only very low levels (Figure 4-figure supplement 3B, lanes 1, 2, 5, 7). The qPCR
285 data were subsequently normalized to the results obtained with a control scrambled sRNA to
286 calculate fold changes in expression levels. Since it is difficult to predict directly from the
287 CLASH data which sRNA in each pair acts as the decoy/sponge, we tested both directions.
288 ArcZ overexpression not only decreased the expression of its mRNA targets (*tpx*, *sdaC*) by
289 more than 50% but also that of CyaR (Figure 5D, panel I). Concomitantly, we observed a
290 substantial increase in CyaR targets *nadE* and *yqaE* (Figure 4D, panel I). CyaR
291 overexpression reduced the level of a direct mRNA target (*nadE*) by ~40% but it did not
292 significantly alter the level of ArcZ or ArcZ mRNA targets (*tpx* and *sdaC*; Figure 4D, panel II).
293 Notably, in this two-plasmid assay CyaR was not expressed at levels higher than ArcZ (Figure
294 4-figure supplement 3A, panel II). Therefore, it is possible that under the tested conditions the
295 CyaR overexpression was not sufficient to see an effect on ArcZ. We find this unlikely as
296 overexpression of CyaR also did not significantly affect endogenous ArcZ levels, which was
297 ~80-fold less abundant than CyaR in this experiment (Figure 4-figure supplement 3A, panel
298 III). The qPCR results were also confirmed by Northern blot analyses (Figure 4-figure
299 supplement 3B, lanes 1-8), which also demonstrate that ArcZ processing was not affected
300 upon CyaR overexpression. These results suggest that the regulation is unidirectional,
301 reminiscent of what has been described Qrr3 in *Vibrio harveyi* (Feng et al., 2015). We
302 conclude that ArcZ acts as a CyaR anti-sRNA and can trigger its degradation.
303 To provide additional support for direct interactions between these sRNAs, we
304 generated mutations in the seed sequences of the sRNAs analysed here (Figure 4C). We
305 found that two G to C nucleotide substitutions in ArcZ was sufficient to disrupt ArcZ
306 downregulation of CyaR (Figure 4C-D; ArcZ 70-71 + CyaR). This regulation, however, was
307 almost fully restored when complementary mutations were introduced in the CyaR region
308 (Figure 5C-D; ArcZ 70-71 + CyaR 38-39). These data also demonstrate that it is very unlikely
309 that the observed changes in CyaR levels were be the result of Hfq redistribution due to over-
310 expression of ArcZ over-expression (Moon and Gottesman, 2011; Papenfort et al., 2009), as
311 the ArcZ seed mutant stably accumulated (and therefore effectively binds Hfq), but did not
312 affect CyaR levels. Unexpectedly, the wild-type ArcZ was also able to effectively suppress the
313 CyaR seed mutant (Figure 5D; ArcZ + CyaR 38-39), possibly because ArcZ can still form
314 stable base-pairing interactions with the CyaR mutant.
315 These results, together with the CLASH data, strongly support the notion that ArcZ and
316 CyaR base-pair *in vivo*, resulting in degradation of CyaR but not *vice versa*.
317

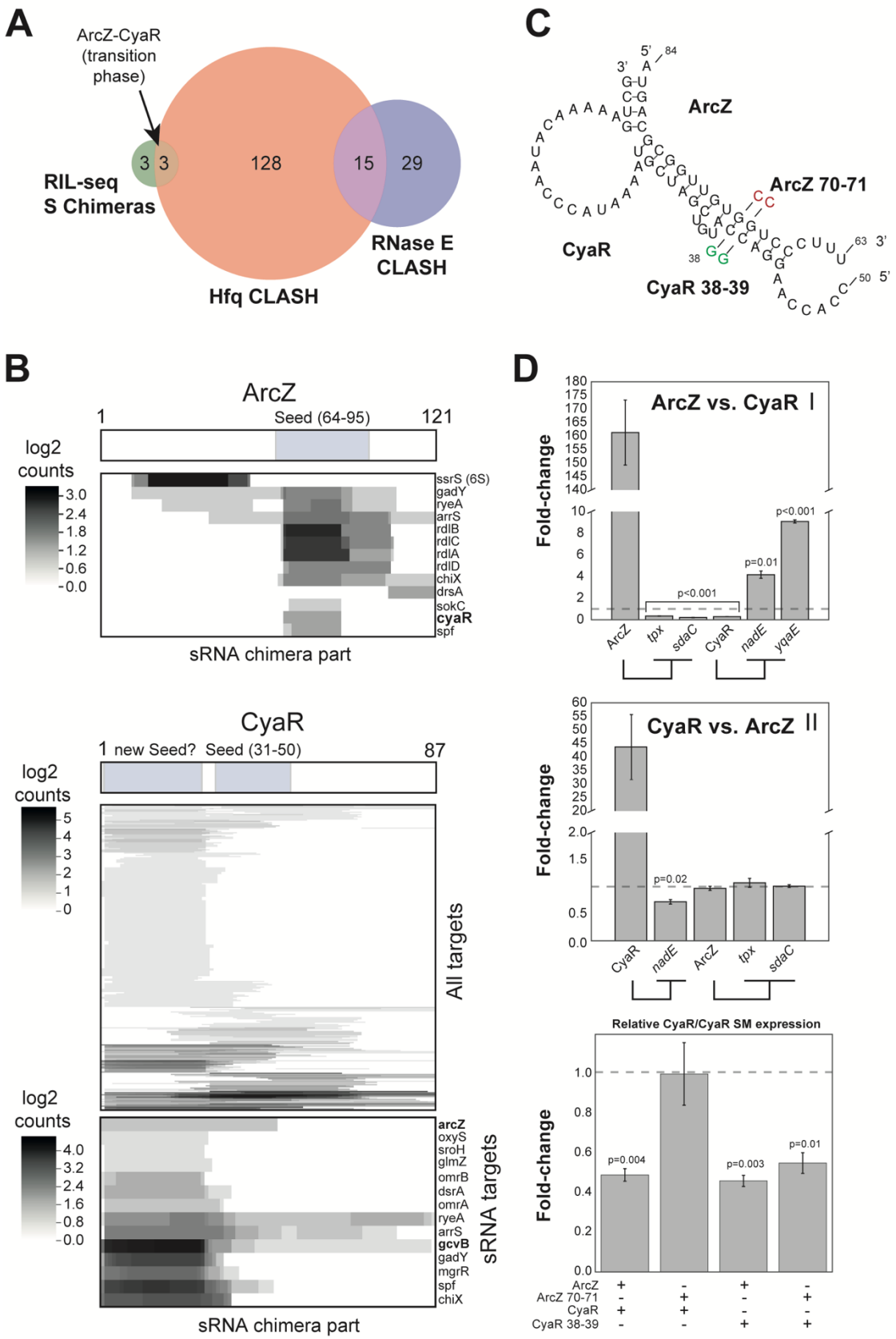
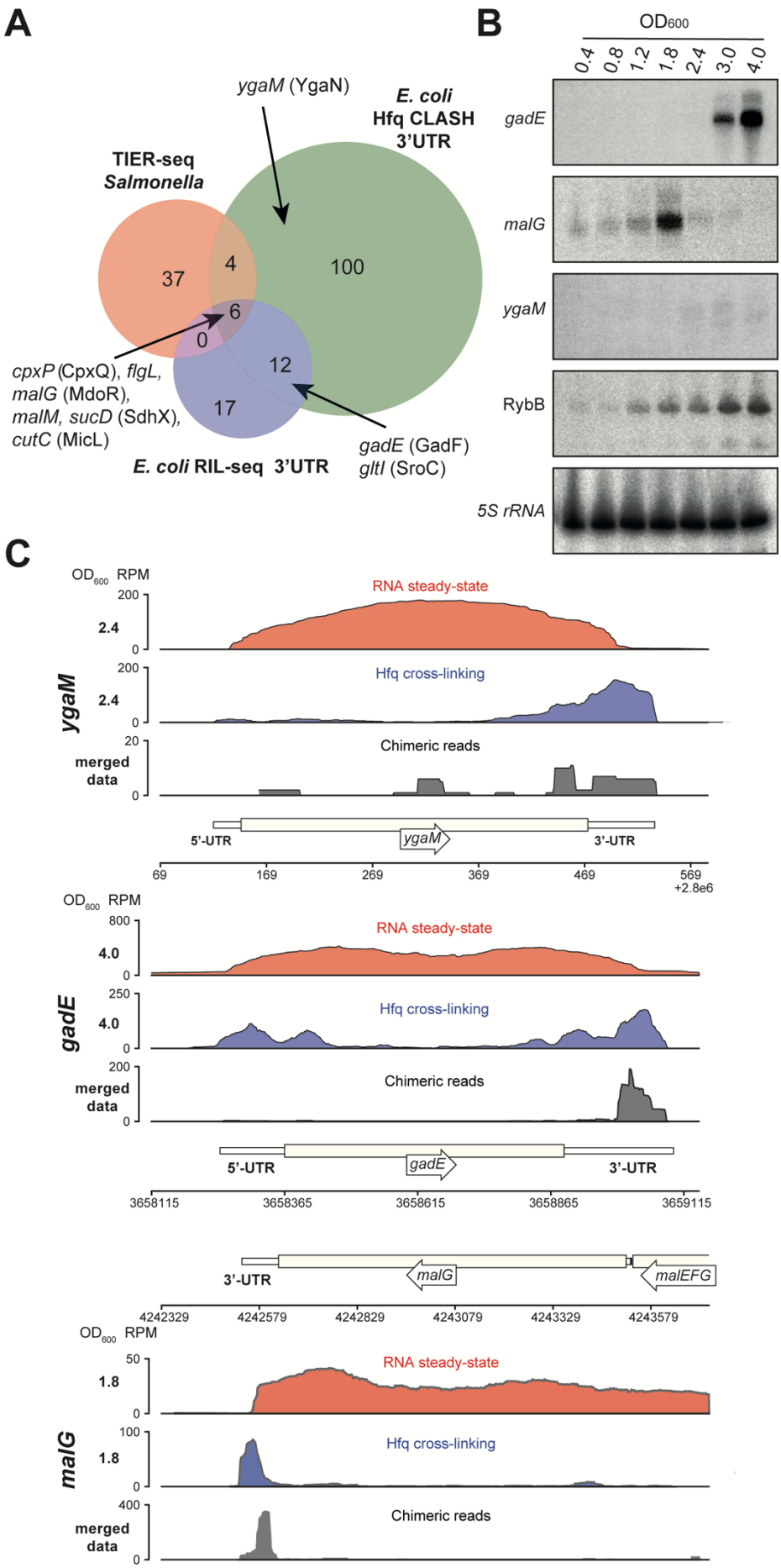


Figure 4. sRNA-RNA interactions identified by CLASH are growth-stage specific.

(**A**) Hfq CLASH uncovers sRNA-sRNA networks: comparison between statistically filtered sRNA-sRNA interactions in the Hfq CLASH data, RIL-seq S-chimeras (Melamed et al., 2016) (log and stationary)

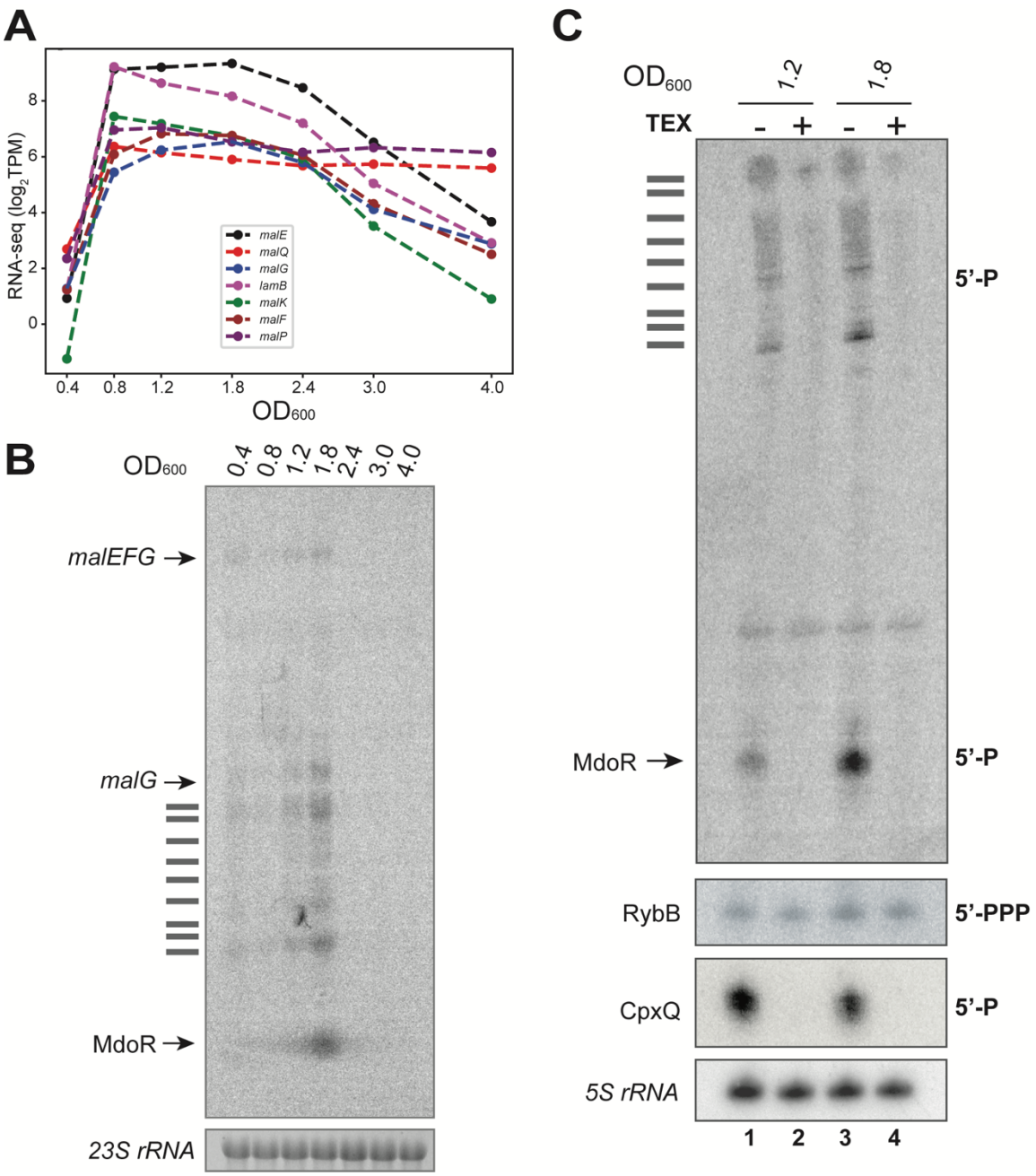
322 and RNase E CLASH (Waters et al., 2016). Only core genome sRNAs were considered. Red coloured
323 sRNA-sRNA interactions have been characterized in more detail. (**B**) Heatmaps showing the read
324 density (log_2(chimera count)) of chimeric fragments mapping to ArcZ and CyaR. (Top) ArcZ regions
325 involved in sRNA-sRNA interactions. The location of the known ArcZ seed sequence is indicated above.
326 (Bottom) CyaR heatmaps that show all CyaR chimeras and CyaR-sRNA chimeras, respectively. The
327 location of the known CyaR seed sequence, as well as a new seed, is indicated above. (**C**) Base-pairing
328 interactions predicted from the ArcZ-CyaR chimeras using RNAcofold. The nucleotide substitutions for
329 experimental validation of direct base-pairing are shown as red or green residues. (**D**) SRNA-sRNA
330 interactions coordinate nutritional stress responses. ArcZ, CyaR were overexpressed and the levels of
331 their targets were monitored by RT-qPCR. The *tpx* and *sdaC* mRNAs are ArcZ mRNA targets. The
332 *nadE* and *yqaE* mRNAs are CyaR targets. The *dppA* mRNA is a GcvB target. Experiments were
333 performed in biological and technical triplicates; Error bars indicate the standard error of the mean
334 (SEM) of the three biological replicates. The dashed horizontal line indicates the level of the
335 overexpressed scrambled RNA. (**E**) ArcZ and CyaR directly interact. The sRNAs and mutants as in (**C**)
336 were ectopically co-expressed in *E. coli* and CyaR and CyaR 38-39 levels were quantified by RT-qPCR.
337
338
339 **Hfq CLASH identifies novel sRNAs in untranslated regions**
340 Two lines of evidence from our data indicate that many more mRNAs may be
341 harbouring sRNAs in their UTRs or be involved in base-pairing among themselves. First,
342 around 11% of the intermolecular chimeras mapped to mRNA-mRNA interactions (Figure 3A).
343 Secondly, we observed extensive binding of Hfq in 3'UTRs near transcriptional terminators
344 (Figure 1-figure supplement 3A, C), indicating that like in *Salmonella*, *E. coli* 3'UTRs may
345 harbour many functional sRNAs (Chao et al., 2017). We identified 122 3'UTR-containing
346 mRNA fragments that were involved in 550 interactions. Sixty-five of these interactions were
347 also identified in the RIL-seq S-chimeras data (Melamed et al., 2016). Eighteen of the 3'UTRs
348 were found as part of chimeras in the RIL-seq data, while 10 appeared stabilised upon
349 transient inactivation of RNase E performed in *Salmonella* (TIER-seq data (Chao et al., 2017));
350 Figure 6A, Supplementary Tables 2.5 and 2.6). Out of the 550 3'UTR-mRNA chimeric reads,
351 79 were 3'UTRs fused to 5'UTRs of mRNAs, suggesting that these may represent 3'UTR-
352 derived sRNAs that base-pair with 5'UTRs of mRNAs, a region frequently targeted by sRNAs
353 (Supplementary Table 2.6). Strikingly, 223 interactions contained 3'UTR fragment of *cpxP*, 51
354 of which were also found in the RIL-Seq data (Supplementary Table 2.6). In *Salmonella cpxP*
355 harbours the CpxQ sRNA (Chao and Vogel, 2016). Our analyses greatly increased the
356 number of potential CpxQ mRNA targets and show that the vast majority of CpxQ interactions
357 take place during the transition and stationary phases (Supplementary Table 2.6).
358


A
ygaM (YgaN)
E. coli
Hfq CLASH
3'UTR
TIER-seq
Salmonella
37
4
100
6
0
12
17
cpxP (CpxQ), flgL,
malG (MdoR),
malM, sucD (SdhX),
cutC (MicL)
gadE (GadF)
gltI (SroC)
E. coli RIL-seq 3'UTR
B
OD600
0.4
0.8
1.2
1.8
2.4
3.0
4.0
gadE
malG
ygaM
RybB
5S rRNA
C
OD600 RPM
RNA steady-state
Hfq cross-linking
Chimeric reads
ygaM
2.4
merged data
5'-UTR
3'-UTR
69
169
269
369
469
569
+2.8e6
gadE
4.0
3658115
3658365
3658615
3658865
3659115
malG
malEFG
1.8
4242329
4242579
4242829
4243079
4243329
4243579

359

360 **Figure 5. Hfq CLASH uncovers novel 3'UTR-derived sRNAs.**
361 (**A**) Genes of which the 3'UTRs were found fused to mRNAs, were selected from the statistically filtered
362 CLASH data and RIL-seq S-chimera data. The RIL-seq RNA-RNA interaction set (Melamed et al., 2016)
363 S-chimeras for Log and Stationary phases of growth was filtered for the 3'UTR/EST3UTR annotations
364 on either orientation of the mRNA-mRNA pairs. Both were intersected with the set of mRNAs that were
365 predicted by TIER-seq studies (Chao et al., 2017) to harbour sRNAs that get released from 3'UTRs by
366 RNase E processing. Known (CpxQ, SdhX, MicL, GadF and SroC) and novel 3'UTR derived sRNAs
367 (MdoR, *figL* 3'UTR and YgaN) are indicated. (**B**) MdoR is transiently expressed during the transition
368 from exponential to stationary phase. RybB was probed as a sRNA positive control and 5S rRNA as
369 the loading control. See Figure 6-figure supplement 1 for full-size blots. Source data are provided as a
370 Source Data file. (**C**) Genome-browser snapshots of several regions containing candidate sRNAs for
371 optical densities at which the RNA steady-state was maximal for each candidate; the candidate names
372 and OD_{600} are indicated at the left side of the y-axes; the y-axis shows the normalized reads (RPM:
373 reads per million); red: RPM of RNA steady-states from an RNA-seq experiment, blue: Hfq cross-linking
374 from a CLASH experiment; black: unique chimeric reads found in this region.
375
376 We identified six mRNA 3'UTRs that were uncovered in all three (Hfq CLASH, RIL-seq
377 and TIER-seq) datasets (Figure 5A), suggesting these likely contain sRNAs released from
378 3'UTRs by RNase E processing. Northern blot analyses confirmed the presence of sRNAs in
379 *malG* and *gadE* 3'UTRs (Figure 5B, Figure 5-figure supplement 1A). The latter was also
380 recently experimentally confirmed in the RIL-seq data and was annotated as GadF (Melamed
381 et al., 2016). Furthermore, significant Hfq cross-linking could be detected in the 3'UTRs of
382 these transcripts (Figure 5C). In addition, we could show that the 3'UTR of *ygaM*, which was
383 found in chimeric reads in our data, also likely harbours a ~100 nt sRNA (hereafter referred to
384 as YgaN; Figure 5-figure supplement 1A) and robust Hfq cross-linking could be detected in
385 this region (Figure 5B-C).
386 To substantiate these results, we analysed RNA-seq data from a study that used
387 Terminator 5'-Phosphate Dependent Exonuclease (TEX) to map transcription start sites
388 (TSS) of coding and non-coding RNAs in *E. coli* (Thomason et al., 2015). TEX degrades
389 processed transcripts that have 5' monophosphates, but not primary transcripts with 5'
390 triphosphates. Therefore, these data enabled us to determine whether (a) a TSS was detected
391 in the 3'UTR and whether these were generated by RNase-dependent processing (TEX
392 sensitive) or originated from an independent promoter (TEX insensitive). For 47 of the 122
393 predicted 3'UTR-derived sRNAs TEX data provided strong evidence for the presence of
394 sRNAs (Figure 5-figure supplement 1B-C, Supplementary Table 2.5 and see Data and Code
395 availability). The TEX data indicate that *ygaM* has (at least) two promoters, one of which is
396 located near the 3' end of the gene that we predict is the TSS for YgaN (Figure 5-figure
397 supplement 1B). Furthermore, we speculate that YgaN is processed by RNases. This is based

398 on the observation that multiple YgaN species were detected in the Northern blot analyses
399 (Figure 5-figure supplement 1A) and the TEX data indicate that shorter YgaN RNAs are
400 sensitive to TEX treatment (Figure 5-figure supplement 1B).
401 The majority of the sRNAs we analysed are more abundant at higher cell densities
402 (including GadF, YgaN and RybB; see Figure 5B, 2A). In sharp contrast, 3' UTR *malG* sRNA
403 was expressed very transiently and peaked at an OD_{600} of 1.8 (Figure 5B). We envisage that
404 the particularly transient expression of this sRNA may be correlated with a role in the adaptive
405 responses triggered during transition from exponential to stationary phases of growth. We
406 named it MdoR (*mal*-dependent OMP repressor) and characterized it in detail.
407 The steady state levels of *malG* and *malEFG* transcripts recapitulate the same
408 expression profile: both peak at OD_{600} of 1.8 and drop to very low levels at OD_{600} 2.4 (Figure
409 6A). Additionally, we identified shorter *malG* 3'UTR-containing fragments of intermediate
410 length between *malG* and MdoR that could be degradation intermediates (Figure 6B). The
411 detection of these intermediate species suggests that the *malEFG* primary transcript is
412 undergoing serial ribonucleolytic cleavage steps.
413 MdoR is a 104 nt sRNA that contains part of the *malG* coding sequence, including the
414 stop codon and the Rho-independent terminator (Figure 5-figure supplement 1A-B). Two lines
415 of evidence suggest that MdoR is generated via endonucleolytic cleavage: RNase E cleavage
416 was detected in the 3'UTR of *malG* in the *Salmonella* TIER-seq data (Chao et al., 2017).
417 Secondly, the TEX RNA-seq data supported the existence of a short RNA in the same region
418 that has a 5' monophosphate (Figure 5-figure supplement 1B-C). We verified the MdoR TEX
419 data by Northern blot analyses (Figure 6C). Consistent with the TEX RNA-seq data, MdoR
420 could not be detected in our RNA samples treated with TEX (Figure 6C, lanes 2 and 4),
421 confirming it bears a 5' monophosphate. Like MdoR, the positive control CpxQ (Chao and
422 Vogel, 2016) was degraded in the presence of TEX (Figure 6C). In contrast, RybB, an sRNA
423 with a 5' triphosphate transcribed from an independent promoter (Johansen et al., 2006;
424 Papenfort et al., 2006) was a poor substrate for the exonuclease (Figure 6C).
425



426

427 **Figure 6. MdoR is a degradation product that emerges at the transition between exponential and**
428 **stationary phase of growth.**

429 (**A**) The *mal* regulon gene expression peaks at the transition between exponential and stationary
430 phases of growth: the plot shows averages of $\log_2$(TPM) normalized RNA steady-state levels (y-axis);
431 the x-axis indicates the cell densities (OD_{600}) at which samples were taken. (**B**) Northern blot using total
432 RNA from *E. coli* harvested at different cell densities (OD_{600}) probed with an oligo antisense to *malG*
433 3'UTR; 23 S rRNA was used as the loading control; the identity of the bands is indicated at the left of
434 the panel; horizontal bars indicate *malEFG* degradation intermediates. (**C**) MdoR is a degradation
435 product: Northern blot using total RNA from cells at indicated OD_{600} with (+; lanes 2 and 4) or without (-
436 ; lanes 1 and 3) 5'-Phosphate-Dependent Exonuclease (TEX) treatment. The sRNAs RybB (5'ppp) and
437 CpxQ (5'p) were used as negative and positive controls, respectively. The 5S rRNA is a loading control.

438 The text on the right of the blot indicates the phosphorylation state of the 5'-termini for each sRNA.
439
440 TEX treatment of the total RNA also reduced the levels of the longer intermediate
441 species as well as the full-length *malG*. These data support a mechanism by which the full-
442 length polycistronic RNA undergoes decay that is initiated at a site in the upstream *malEFG*
443 region. The distal gene *malE* is clipped off by the degradosome and selectively stabilized,
444 allowing it to be expressed at higher levels than other members of the operon (Newbury et al.,
445 1987). The *malG* 3'UTR, however, would be less susceptible to degradation as it is stabilized
446 by Hfq binding.
447
448 **MdoR directly regulates the expression of major outer membrane porins and represses**
449 **the envelope stress response pathway**
450 The MdoR chimeras frequently contained 5'UTR fragments of two mRNAs encoding
451 major porins, *ompC* and *ompA* (Figure 7A-C), which were also significantly enriched in the
452 RIL-seq data (Figure 8B). The Hfq CLASH data, however, also contained MdoR fragments
453 fused to a *cis*-encoded sRNA, OhsC (Figure 7B-C), suggesting that either MdoR controls its
454 expression by sponging/degradation, or *vice versa*). The most abundant and favourable
455 interactions of MdoR with mRNAs (*ompC* and *ompA*) appear to be utilizing roughly the same
456 region of the *malG* 3'UTR for base-pairing (Figure 7C), suggesting that the corresponding site
457 on the predicted sRNA may be a main, functional seed. The predicted interaction between
458 MdoR and *ompC* is unusually long and consists of two stems interrupted by a bulge,
459 suggesting that these two RNAs form a stable complex. Conservation analyses and *in silico*
460 target predictions (CopraRNA (Wright et al., 2014, 2013)) indicate that the seed sequence
461 predicted by CLASH is relatively well-conserved (Figure 7-figure supplement 1A), and could
462 be utilized for the regulation of multiple targets (Figure 7-figure supplement 1B).
463 To verify the MdoR CLASH data we pulse-overexpressed the sRNA from a plasmid-
464 borne arabinose inducible promoter followed by RNA sequencing (Figure 7D-E). To minimize
465 secondary changes in gene expression, cells were harvested after only 15 minutes of MdoR
466 induction. Note that induction was performed at OD_{600} = 0.4, when endogenous levels of MdoR
467 are very low. As a control we used cells harbouring an empty vector. Differential gene
468 expression analysis (DESeq2 (Love et al., 2014)), identified ~20 transcripts that were
469 significantly enriched in the control data compared to the MdoR overexpression data (Figure
470 7E-F; Supplementary Table 4). Thus, these transcripts are likely downregulated by MdoR *in*
471 *vivo*. This set of transcripts included the sigma factor *rpoE* (σ^E), which plays an important role
472 in controlling gene expression during stress responses, including envelope stress (Alba and
473 Gross, 2004; Bossi et al., 2008; De Las Peñas et al., 1997; Rhodius et al., 2006). The
474 observed reduction in mRNA levels of the anti-σ^E protein RseA can be explained by the fact

that σ^E and RseA are encoded by the same operon.

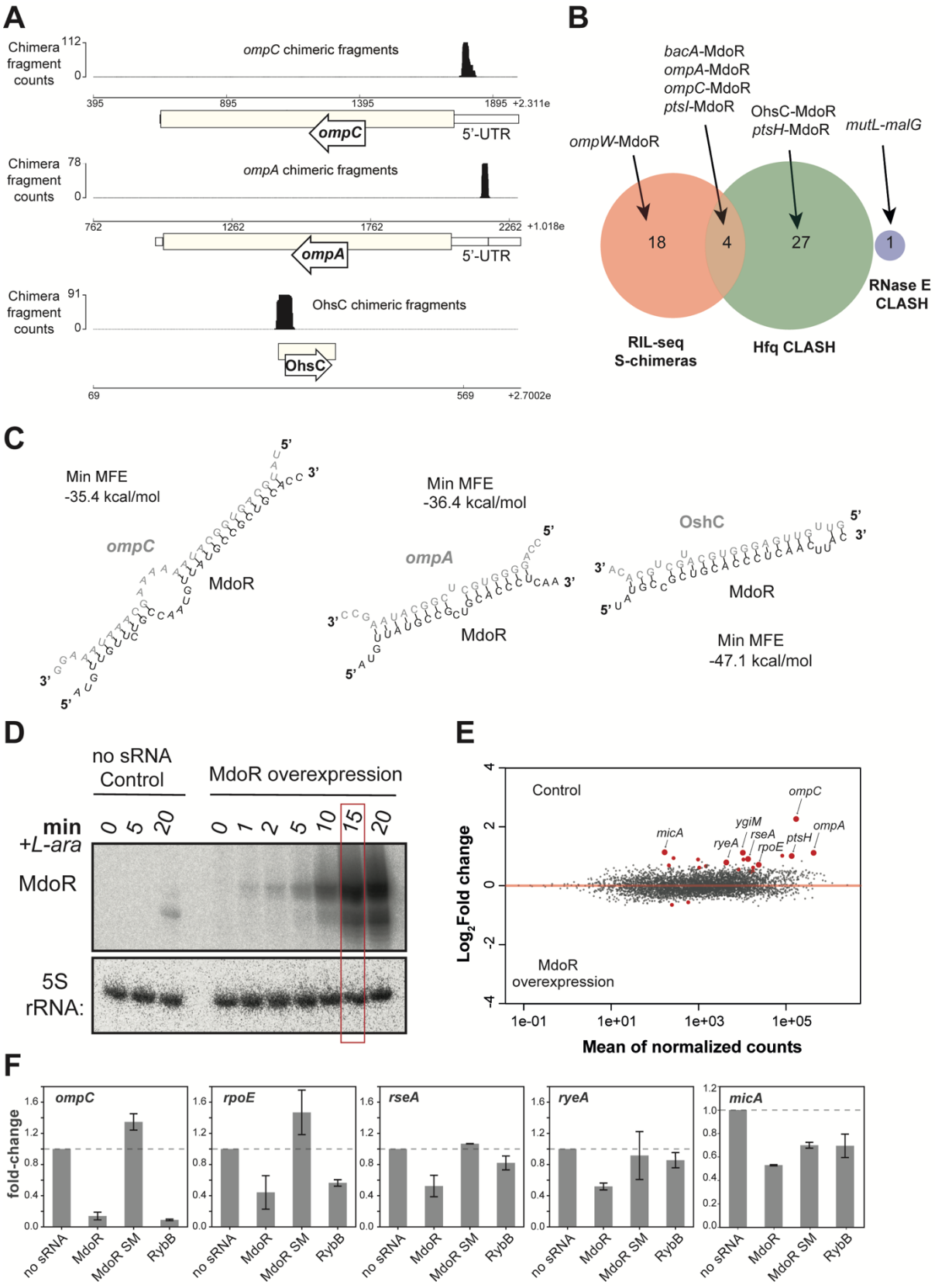


Figure 7. MdoR destabilizes *ompC* and *ompA* mRNAs and downregulates key members the σ^E regulon.

480 (**A**) The three genome browser tracks show the distribution of target mRNA and sRNA fragments
481 (*ompC, ompA* and OhsC) that were fused to MdoR fragments in chimeric reads. (**B**) MdoR-target
482 interactions found in Hfq CLASH, RIL-seq S-chimera data (log and stationary phase) and RNase E
483 CLASH data. (**C**) MdoR forms stable duplexes with the 5'UTR of porin-encoding mRNAs and the OhsC
484 sRNA. *In silico* prediction (RNAcofold (Lorenz et al., 2011)) of hybrid structures derived from the most
485 abundant MdoR chimeric reads with the indicated transcripts. The min. MFE is the minimum folding
486 energy assigned by RNAcofold. (**D**) Pulse-overexpression of MdoR using L-arabinose. The empty
487 pBAD plasmid served as a negative control. Samples were harvested 15-minutes after induction. (**E**)
488 MdoR regulates expression of various mRNAs. DESeq2 analyses were performed on RNA-seq data
489 from three biological replicates. Red points indicate differentially expressed transcripts. Transcripts with
490 a $\log_2$ fold-change > 0 were enriched in the Control data; those with a $\log_2$ fold-change < 0 were enriched
491 in the MdoR overexpression data. The annotated, enlarged red data points indicate several differentially
492 expressed transcripts discussed in the text. (**F**) The MdoR seed region is important for target regulation.
493 RT-qPCR analysis of several differentially expressed transcripts (gene names shown at the top of the
494 plots), in the presence of sRNAs. The plasmid-borne sRNAs were induced for 15 minutes using L-
495 arabinose. The sRNA names are indicated below each bar; the 'no sRNA' sample contains the empty
496 plasmid as reference for fold-change calculations; *recA* was used as the internal reference gene;
497 experiments were performed in technical triplicates; the standard error of the mean (SEM) of two
498 biological replicates are reported as error bars.
499
500 Intriguingly, MdoR overexpression also reduced the levels of sRNAs RyeA and MicA,
501 the latter of which depends on σ^E for its expression (Udekwu and Wagner, 2007). In
502 *Salmonella*, MicA downregulates LamB, a high affinity maltose/maltodextrin transporter (Bossi
503 and Figueroa-Bossi, 2007). Fragments of three mRNAs (*ompC*, *ompA* and *ptsH*) that were
504 found in MdoR chimeric reads were also differentially expressed in the MdoR overexpression
505 RNA-seq data, providing strong evidence that these are direct MdoR targets.
506 All the available data suggest that MdoR is part of a mixed coherent feed forward
507 regulatory network (FFL) that enhances the uptake of maltose/maltodextrin by maltose
508 transporters (see Discussion). By base-pairing with the 5'UTRs of the mRNAs it firstly reduces
509 the flux of the more general porins such as OmpC and OmpA to the outer membrane.
510 Secondly, we propose that MdoR stimulates the accumulation of the high-affinity maltose porin
511 LamB in the OMP by suppressing the inhibitory σ^E pathway and MicA. To further test this
512 model, we performed additional validation experiments. First, we confirmed the DESeq2
513 results for a number of the regulated genes (*ompC, rpoE*, *micA* and *ryeA*) by RT-qPCR (Figure
514 7F). We included an MdoR mutant in which seed sequence in stem 1 was changed into its
515 complementary sequence (Figure 8-figure supplement 1A). As control we also included the
516 RybB sRNA, which regulates *rpoE* and *ompC* expression (Gogol et al., 2011; Papenfort et al.,
517 2006; Thompson et al., 2007). In all cases, the MdoR SM mutations reduced the negative

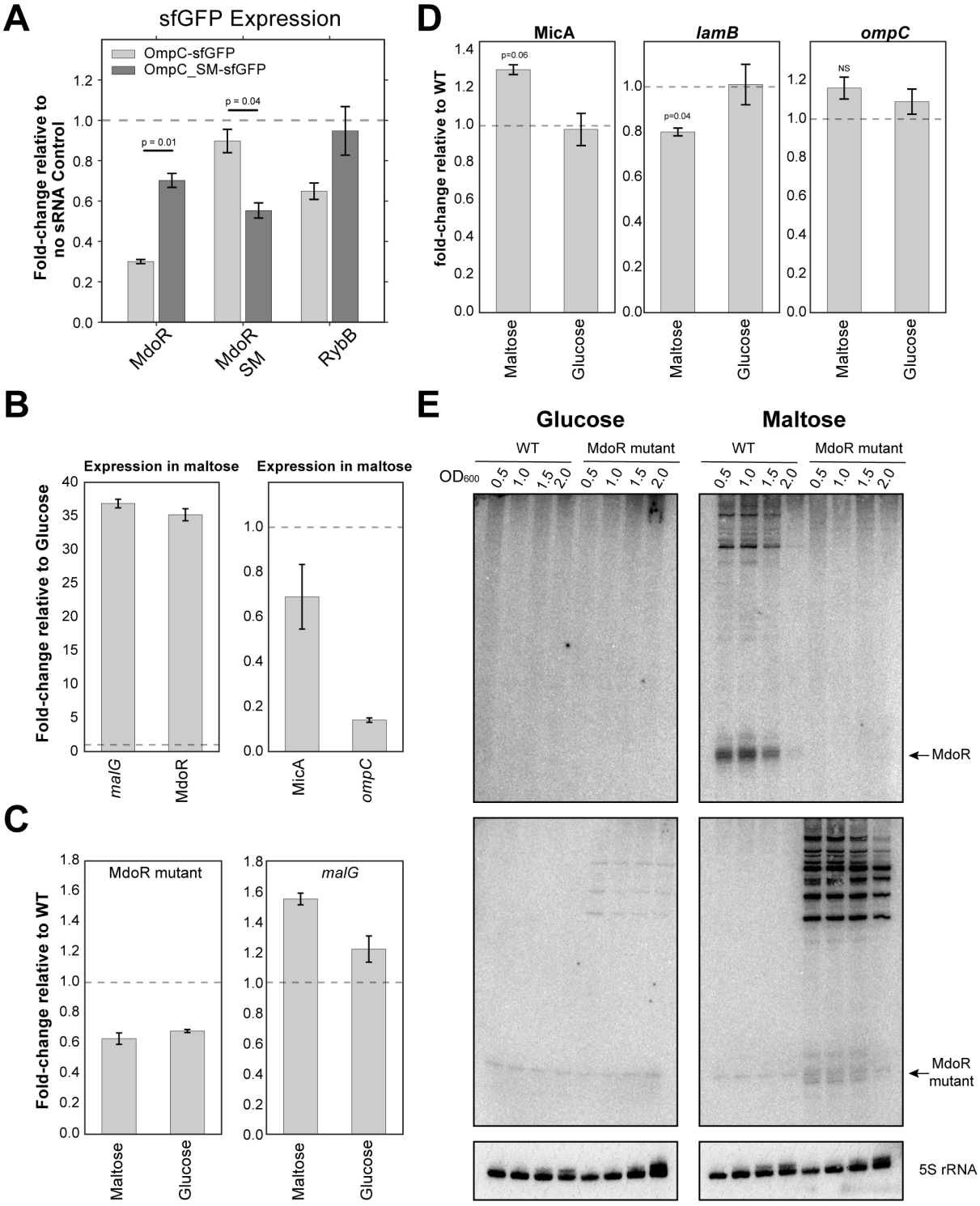
518 regulatory effect on target expression (Figure 7F).

519 To demonstrate direct target regulation *in vivo*, we employed a well-established
520 reporter system where an sRNA is co-expressed with a construct containing the mRNA target
521 region fused to the coding sequence of superfolder green fluorescent protein (sfGFP)
522 (Corcoran et al., 2012; Urban and Vogel, 2007) (Figure 8-figure supplement 1B). Fusions for
523 *ompC*, *ompA* and σ^E were constructed, but only the OmpC and OmpA-sfGFP reporters
524 produced stable fusions that could be analysed. We also included an MdoR sRNA seed
525 sequence mutant (MdoR SM) and an *ompC* mutant containing compensatory mutations
526 (OmpC SM; Figure 8-figure supplement 1A). Unfortunately, our OmpA-GFP reporter construct
527 that contained compensatory mutations in the target region did not generate a stable fusion.
528 Therefore, we were unable to use this reporter system to verify the MdoR-*ompA* interaction.
529 As positive controls we used the MicC and RybB sRNAs as they both regulate *E. coli ompC*
530 expression (Chen et al., 2004; Gogol et al., 2011). Fluorescence measurements confirmed
531 that levels of OmpC-sfGFP and OmpA-sfGFP fusions were significantly lower in cells
532 expressing MdoR (Figure 8-figure supplement 1C). Importantly, MdoR overexpression did not
533 change the expression of the GFP reporter itself (Figure 8-figure supplement 1C). Mutating
534 the MdoR seed region largely restored OmpA- and OmpC-sfGFP reporter levels to the levels
535 of the no sRNA negative control. The MdoR SM mutant was still able to partially suppress the
536 expression of the OmpC SM-sfGFP mutant, suggesting that base-pairing interactions between
537 these two mutants is less stable compared to the wild-type (Figure 8-figure supplement 1C).
538 The wild-type MdoR was also able to partially suppress the expression of the *ompC* SM
539 mutant. We suggest that the predicted base-pairing interactions between MdoR and *ompC* in
540 the second stem (Figure 8-figure supplement 1A) might be sufficient to partially suppress
541 *ompC* expression. Regardless, the data strongly imply that MdoR directly regulates *ompC*
542 expression.

543 To determine whether the changes in fluorescence signal correlate with changes in
544 reporter mRNA levels, we measured the expression of the GFP reporters by RT-qPCR. The
545 results were essentially identical to the GFP fluorescence measurements (Figure 8A);
546 Overexpression of the wild-type MdoR, but not seed mutant, reduced *ompC*-sfGFP mRNA
547 levels. The *ompC* seed mutation (SM) did not fully disrupt regulation by wild-type MdoR.
548 However, the MdoR mutant containing compensatory mutations (SM mutant) was able to
549 much better suppress the *ompC* SM mRNA levels, consistent with the idea that base-pairing
550 was (largely) restored. Next, we performed polysome profiling experiments to assess the level
551 of *ompC* translation upon overexpression of MdoR. Although MdoR overexpression did not
552 noticeably affect 70S and polysome levels (Figure 8-figure supplement 2A), we observed a
553 significant (~37%) reduction of *ompC* mRNA in the polysomal fractions, relative to the upper
554 fractions (Figure 8-figure supplement 2B). We conclude that MdoR can regulate *ompC*

555 expression at the post-transcriptional and translational level.

556



557

558

559 **Figure 8. MdoR modulates expression of key factors involved in maltose intake.**

560 (**A**) MdoR directly downregulates *ompC* mRNA through base-pairing interactions: RT-qPCR analyses

561 of the *ompC-sfGFP* and *ompC SM-sfGFP* fusions expression in the presence of MdoR. The bars

562 indicate the mean fold-change in expression relative to the no sRNA Control (horizontal dashed line).

563 Error bars indicate the standard error of the mean from two biological replicates. The significance of the
564 differences between the WT and MdoR SM was assessed with a two-tailed Student's t-test. (**B**)
565 Endogenous MdoR and *malG* expression is induced during growth on maltose, and *ompC* levels are
566 significantly lower in maltose compared to glucose. Total RNA extracted from exponentially growing
567 cells (OD_{600} 0.5) and MdoR, *malG* (Left), and MdoR direct (*ompC*) and indirect (MicA) targets was
568 quantified by RT-qPCR. The data were normalized to 5S rRNA levels. The bars indicate the mean fold-
569 change in expression relative to expression in cells growing in glucose (indicated on the plot with a
570 horizontal dashed line). Error bars indicate the standard error of the mean from two biological replicates.
571 (**C**) The mutant MdoR is ~50% less abundant than the wild-type. Cell growth and RT-qPCR analysis
572 of MdoR and *malG* expression was performed as in (b). The bars indicate the mean fold-change in
573 expression relative to the wild-type. Error bars indicate the standard error of the mean from two
574 biological replicates. (**D**) Increased MicA and decreased *lamB* levels in the MdoR seed mutant. P-values
575 were calculated using with a one-sample t-test. (**E**) MdoR seed sequences are important for RNase E
576 recruitment and MdoR biogenesis. Northern blot that compares MdoR and longer *malG*-3'UTR
577 containing fragments expression in wild-type *E. coli* and the MdoR seed mutant strain.
578

579 **MdoR enhances maltoporin expression during maltose fermentation**

580 To further substantiate these results, we next switched to a more controlled system to
581 investigate the effect of endogenous MdoR on its targets. To determine whether MdoR has a
582 role in adaptation to maltose-metabolising conditions, single overnight cultures grown in
583 glucose were split and (re)inoculated in fresh medium containing either glucose or maltose as
584 the sole carbon sources, and expression of several *mal* regulon genes and MdoR targets were
585 quantified. We show that MdoR and its parental transcript *malG* are almost undetectable
586 during growth in glucose, and highly expressed during growth in maltose (~35-fold increase,
587 Figure 8B). This is consistent with catabolite repression of the *mal* regulon by glucose, and its
588 induction my maltose (Boos and Shuman, 1998). Intriguingly, we observed that *ompC* mRNA
589 levels are overall significantly lower during growth in maltose, compared to glucose (Figure
590 8B). This suggests that porin expression is also regulated by nutrient source in *E. coli*.
591 Similarly, MicA, a repressor of LamB synthesis, has reduced levels in maltose compared to
592 glucose (Figure 8B). We next mutated the entire seed sequence of the chromosomal copy of
593 MdoR (both stem 1 and 2; Figure 8-figure supplement 1A) in the chromosomal context, to
594 completely disrupt base-pairing with *ompC* mRNA. Notably, the fully-processed mutant MdoR
595 sRNA is less abundant than the wild-type (Figure 9C) and longer (unprocessed) fragments
596 that contain upstream *malG* regions could be readily detected (Figure 9E). We speculate that
597 the mutation in the upstream MdoR sequence might have affected RNase E recruitment or
598 cleavage, impairing MdoR processing. The MdoR mutant strain also accumulated significantly
599 higher levels of MicA (Figure 8D) and less *lamB*. Levels of *ompC* mRNA levels were also

slightly elevated in the mutant in maltose relative to the parental strain, although in this case a similar increase was also observed in glucose.

Collectively, the data suggest a role for MdoR in enhancing the uptake of maltose when more favourable carbon sources become limiting.

Discussion

Microorganisms need to constantly adapt their transcriptional program to counteract changes in their environment, such as changes in temperature, cell density and nutrient availability. In bacteria, small RNAs (sRNAs) and their associated RNA-binding proteins are believed to play a key role in this process. By controlling translation and degradation rates of mRNAs upon stress imposition (Holmqvist and Wagner, 2017; Nitzan et al., 2017; Shimoni et al., 2007), they can regulate the kinetics of gene expression as well as suppress noisy signals (Beisel and Storz, 2011), enabling organisms to more efficiently adapt to environmental changes. A major challenge for bacteria is the transition from exponential growth to stationary phase, when the most favourable nutrients become limiting. To counteract this challenge, cells need to rapidly remodel their transcriptome to be able to efficiently metabolize alternative carbon sources. This transition is very dynamic and involves activation and repression of diverse metabolic pathways. However, it is unclear to what degree sRNAs contribute to this transition. The most useful piece of information would be to know what sRNAs are upregulated during this transition phase and to identify their RNA targets. This would help to uncover the regulatory networks that govern this adaptation, as well as provide a starting point for more detailed functional analyses on sRNAs predicted to play a key role in this process. For this purpose, we performed UV cross-linking, ligation and sequencing of hybrids (CLASH (Kudla et al., 2011)) to unravel the sRNA-target interactions during this transition. Using Hfq as a bait we uncovered thousands of unique sRNA-target interactions. Our data are consistent with previously published work (Melamed et al., 2016; Waters et al., 2016) but we also identified almost 1700 novel sRNA-mRNA interactions and over 100 novel sRNA-sRNA interactions. We experimentally validated several of the interactions found in our CLASH findings. We identified functional sRNA-sRNA interactions and describe a novel 3'UTR derived sRNA that plays a role in enhancing uptake of an alternative carbon source during the transition to stationary phase.

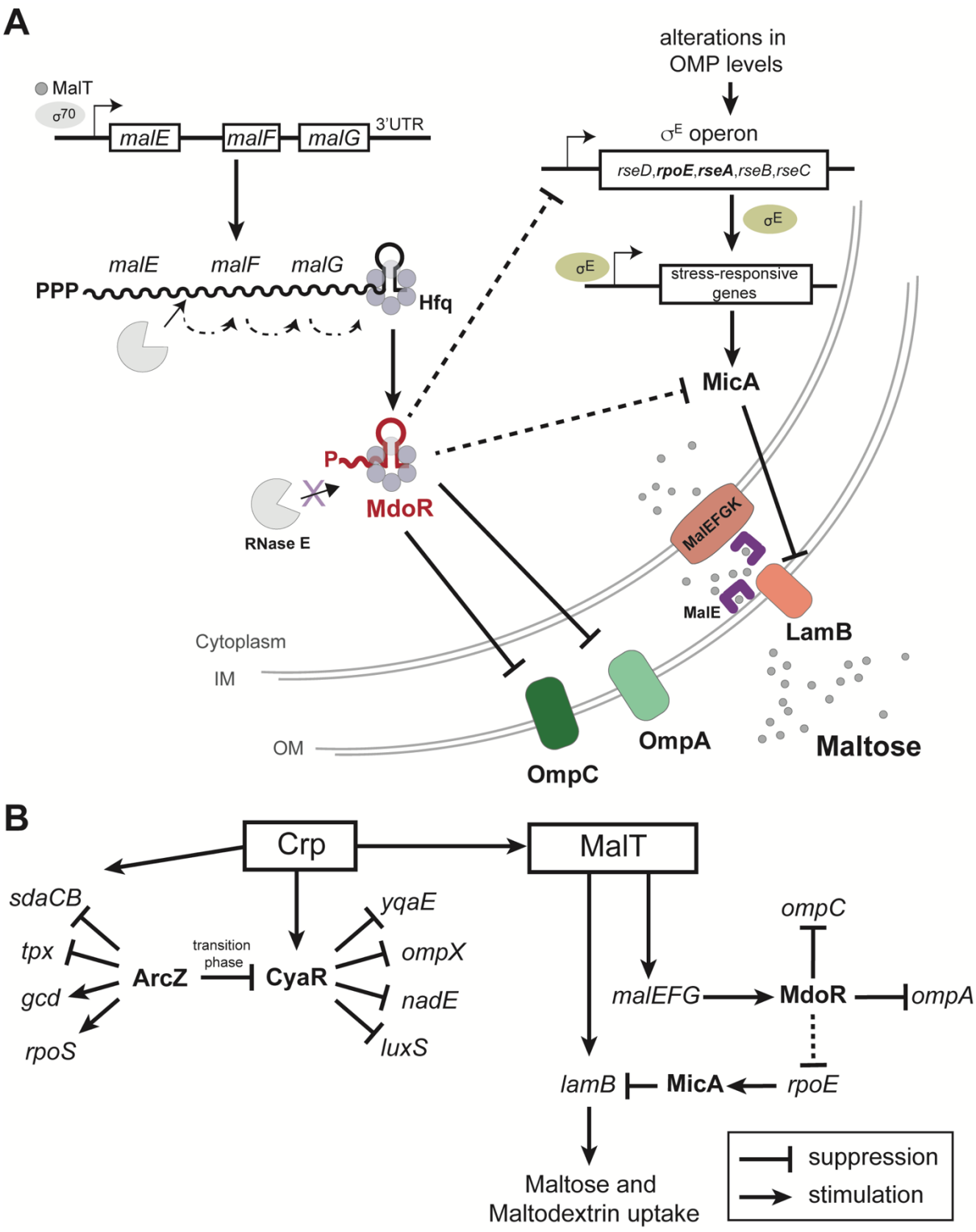


Figure 9. CLASH uncovers sRNA-target interaction networks that regulate adaptation to changes in nutrient availability.

(**A**) Model for MdoR biogenesis and its role in post-transcriptional regulation. During maltose (grey circles) utilization, MalT transcribes the *malEFG* operon that encodes components for maltose uptake. A fraction of the *malEFG* transcripts is degraded by RNase E, generating the 3'UTR-derived MdoR sRNA. MdoR base-pairs with *ompC* mRNA, leading to its decay. Additionally, MdoR downregulates expression of *ompA* and MicA that represses LamB synthesis. Therefore, as part of a feed-forward loop

644 (FFL), MdoR promotes accumulation of maltose specific porins, which facilitates maltodextrin uptake.
645 Increased production of LamB exerts additional pressure on the OM assembly machinery. This
646 extracytoplasmic stress, is suppressed by σ^E activation, that controls the expression of stress-
647 responsive genes and most of the assembly and insertion factors. But σ^E activation is not desirable in
648 these conditions for two reasons: its unbalanced activation has precarious effects on the cell, and would
649 produce high amounts of MicA, that would ultimately repress LamB production. For these reasons,
650 MdoR regulation includes a regulatory arm that mitigates excessive σ^E response through MdoR
651 repression of *rpoE* expression and prevention of σ^E activation. The latter is achieved by downregulation
652 of *omp* mRNAs, which indirectly relieves envelope stress. (**B**) sRNA-target networks regulating
653 peptide/amino acids and maltose/maltodextrin uptake. Boxed are the key transcription factors, sRNAs
654 are in bold and italicized names are mRNA targets. CRP-cAMP induces expression of CyaR, which is
655 repressed by ArcZ, specifically during the transition phase, a regulatory circuit that connects multiple
656 pathways related to the onset of stationary phase and biofilm formation to quorum sensing, cellular
657 adherence and the nutritional state of the cell. MalT transcribes *lamB* and, via *malEFG* transcription,
658 promotes MdoR accumulation. MdoR indirectly promotes LamB synthesis by repressing the opposing
659 σ^E pathway and its sRNA, MicA. Thus, MalT and MdoR jointly promote *lamB* expression, forming a
660 mixed coherent FFL.
661
662 **Hfq CLASH**
663 Our *S. cerevisiae* Cross-linking and cDNA analysis data (CRAC; (Granneman et al.,
664 2009)) showed that a percentage of the cDNAs were formed by intermolecular ligations of two
665 RNA fragments (chimeras) known to base pair *in vivo* (Kudla et al., 2011). These findings
666 prompted us to develop a refined protocol to enrich for sRNA-target chimeric reads using Hfq
667 as an obvious bait. The initial Hfq UV cross-linking data (CRAC; (Tree et al., 2014)) did not
668 yield sufficiently high numbers of chimeric reads to extract new biological insights. In line with
669 observations from other groups (Bandyra et al., 2012; Bruce et al., 2018; Morita et al., 2005),
670 it was proposed that duplexes formed by Hfq are rapidly transferred to the RNA degradosome
671 (Bandyra et al., 2012; Bruce et al., 2018; Morita et al., 2005). This can cause an extensive
672 reduction in the likelihood of capturing sRNA-target interactions with Hfq using CLASH
673 (Waters et al., 2016). However, a recent study demonstrated that Hfq can be used effectively
674 as a bait to enrich for sRNA-target duplexes under lower-stringency purification conditions
675 suggesting that sRNA-mRNA duplexes are sufficiently stable on Hfq during purification
676 (Melamed et al., 2016). This encouraged us to further optimize the CLASH method. We made
677 a number changes to the protocol that, when combined, enabled us to recover a large number
678 of sRNA-target chimeric reads (detailed in Materials and Methods). We shortened various
679 incubation steps to minimize RNA degradation and performed very long and stringent washes
680 after bead incubation steps to remove background binding of non-specific proteins and RNAs.
681 Crucially, we very carefully controlled the RNase digestion step that is used to trim the cross-

682 linked RNAs prior to making cDNA libraries, ensuring the recovery of longer chimeric RNA
683 fragments. The resulting cDNA libraries were paired-end sequenced to increase the recovery
684 of chimeric reads with high mapping scores from the raw sequencing data. These
685 modifications led to a substantial improvement in the recovery of chimeric reads (9.5%
686 compared to 0.001%. 0.71% were intermolecular chimeras).

687 Both RIL-seq and Hfq CLASH have advantages and disadvantages, but they are highly
688 complementary approaches. A major strength of CLASH, however, is that the purification
689 steps are performed under highly stringent and denaturing conditions. During the first FLAG
690 affinity purification steps the beads are extensively washed with high salt buffers and the
691 second Nickel affinity purification step is done under completely denaturing conditions (6M
692 guanidium hydrochloride). These stringent purification steps enable can significantly reduce
693 noise by strongly enriching for RNAs covalently cross-linked to the bait protein (Granneman
694 et al., 2009). Indeed, we show that Hfq CLASH can generate high quality RNA-RNA interaction
695 data with very low background: Only a few hundred chimeric reads were found in control
696 datasets, compared to the over 70.000 that co-purified with Hfq. The RIL-seq library
697 preparation protocol uses an rRNA depletion step to remove contaminating ribosomal RNA,
698 whereas for Hfq CLASH this is not necessary. Our library preparation protocol also includes
699 the use of random nucleotides in adapter sequences to remove potential PCR duplicates
700 ("collapsing") from the data.

701 The very stringent purification conditions used in CLASH could, in some cases also be
702 a disadvantage as it completely relies on UV cross-linking to isolate directly bound RNAs. In
703 cases where protein-RNA cross-linking efficiencies are low (for example proteins that only
704 recognize double-stranded RNA), RIL-seq may be a better approach as it does not completely
705 rely on UV cross-linking (Melamed et al., 2016).

706 A large number of interactions were unique to both RIL-seq and Hfq CLASH datasets,
707 which we believe can be explained by a number of technical and experimental factors. The
708 denaturing purification conditions used with CLASH completely disrupts the Hfq hexamer
709 ((Tree et al., 2014) and this work) and therefore during the adapter ligation reactions the RNA
710 ends are likely more accessible for ligation. In support of this, in the RIL-Seq data, the sRNAs
711 are mostly found in the second half of the chimeras (Melamed et al., 2016), in the Hfq CLASH
712 data we see sRNAs fragments roughly equally distributed in both sides (45% in left fragment
713 and 55% in right fragment). Indeed, it was proposed that in RIL-seq the 3' end of the sRNA is
714 buried in the hexamer and therefore not always accessible for ligation (Melamed et al., 2016).

715 For the RIL-seq experiments, the authors harvested the cells at 4ºC and they
716 resuspend the cells in ice-cold PBS prior to UV irradiation (Melamed et al., 2018, 2016), which
717 results in a cold-shock that can affect the sRNA-interactome as well as sRNA stability. We
718 cross-link actively growing cells in their growth medium and we UV irradiate our cells only for

719 several seconds using the Vari-X-linker we recently developed (van Nues et al., 2017). We
720 use filtration devices to rapidly harvest our cells (less than 30 seconds) and the filtered cells
721 are subsequently stored at -80ºC. We previously showed that filtration combined with short
722 UV cross-linking times dramatically reduces noise introduced by the activation of the DNA
723 damage response and significantly increased the recovery of short-lived RNA species (van
724 Nues et al., 2017). We speculate that many of the interactions that are unique to our Hfq
725 CLASH data represent short-lived RNA duplexes that are preferentially captured with our UV
726 cross-linking and rapid cell filtration setup.
727
728 **The 3'UTR derived sRNA MdoR functions in a mixed feed forward loop by suppressing**
729 **opposing pathways**
730 We found a large number of chimeras that represent over a hundred distinct
731 intermolecular interactions between 3'UTRs and other mRNA regions, which implicate direct
732 mRNA-mRNA communication. These interactions could have been formed by 3'UTR
733 fragments that have been processed by RNase E (or other RNases) or by an sRNA located
734 within the 3'UTR that transcribed from an internal promoter. These 3'UTR fragments are
735 primarily described as decoys or sponges for other sRNAs (Miyakoshi et al., 2015b), but could
736 act as *trans*-acting sRNAs as well (Chao and Vogel, 2016). Among the interactions in our data
737 we found 79 interactions between 3'UTRs and 5'UTRs (Supplementary Table 2.6). We
738 speculate that many of these represent novel 3'UTR-derived sRNAs that target 5'UTRs of
739 mRNAs. Reanalysis of RNA-seq data from a study that globally mapped *E. coli* transcription
740 start sites (TSS) (Thomason et al., 2015) identified 47 TSS in these 3'UTR, strongly
741 suggesting that these 3'UTRs host sRNAs. The majority of these 3'UTR fragments appear to
742 have 5' monophosphates. We also verified some of our findings by Northern blot analyses.
743 Thus, we have potentially uncovered many novel 3'UTR-derived sRNAs.
744 Some of these predictions were validated by our follow up work (discussed below) and,
745 while this work was in progress, also by others (Melamed et al., 2016). One of the 3'UTR-
746 derived sRNAs we uncovered (MdoR) was of particular interest as it is only detected during
747 the transition from late exponential to early stationary phase. A model of how MdoR
748 contributes to maltose uptake is shown in Figure 9A.
749 The genetic structure and transcriptional regulation of the *mal* regulon are well
750 understood. However, its post-transcriptional regulation has remained largely unexplored. Our
751 work uncovered new links between the maltose uptake (*mal*) regulon, envelope stress-
752 responses and membrane composition/assembly pathways. While initially cells metabolize
753 more favourable carbon sources such as glucose, these are generally rapidly depleted and
754 bacteria need to quickly switch to alternative carbon sources, such as maltose and
755 maltodextrins. During maltose utilisation, the *malEFG* operon is transcribed by the MalT

756 transcription factor and the transcript is processed by RNase E as well as other degradosome
757 components (Khemici and Carpousis, 2004). Here we show that the MdoR sRNA is a product
758 of *malEFG* processing, which is protected from further degradation by Hfq.

759 MdoR is unique in a sense that it not only a 3'UTR derived sRNA that targets multiple
760 pathways, but also because it is part of a mixed coherent feed forward loop (FFL) that we
761 predict promotes maltose uptake via LamB. Efficient uptake of maltose/maltodextrin not only
762 requires the inner membrane transporters encoded by the *malEFG* operon, but also the high-
763 affinity maltose transporter LamB (Figure 9A), which cooperates with the inner membrane
764 proteins to import these carbon sources. *lamB* is significantly upregulated when cells start to
765 utilize alternative carbon sources. A major advantage of a coherent FFL is that it can
766 accelerate or delay responses to stimuli, such as changes in carbon source availability. In this
767 FFL, the key activator is the MalT transcription factor, which induces the expression of both
768 *malEFG* and *lamB* when cells start to consume maltose (Figure 9B). Expression of new OMPs,
769 however, needs to be carefully coordinated as any changes in the protein composition of the
770 OM, such as increased levels of LamB (or accumulation of misfolded LamB in the periplasm),
771 can lead to induction of the σ^E envelope stress response caused by the reorganisation of the
772 membrane when maltodextrins are utilised (Figure 9A)(Kenyon et al., 2005). We propose that
773 *malEFG* indirectly promotes LamB accumulation via the MdoR sRNA that (a) reduces the
774 levels of general porins (OmpC and OmpA) and (b), likely indirectly, dampens the σ^E stress
775 response as a result of OMP level reduction. Dampening σ^E expression important during
776 maltose uptake as it induces the expression of MicA, which directly downregulates *lamB*
777 translation (Bossi and Figueroa-Bossi, 2007). When ectopically expressed at high levels,
778 MdoR reduces mRNA levels of *ompC* and *ompA*, which may free up the resources enabling
779 efficient accumulation of LamB in the outer membrane. At physiological levels MdoR
780 significantly reduces MicA expression and we observed a ~20% increase in *lamB* mRNA
781 levels. *OmpC* mRNA levels were not substantially affected. While these changes appear
782 modest, it is important to take into consideration the very high abundance and intrinsic stability
783 of *lamB* and *ompC* mRNAs (minute-long half-lives). It is therefore expected that even a mild
784 reduction in their mRNA levels can profoundly relieve the pressure on the OMP translation
785 and assembly pathways (Guo et al., 2014). The net outcome of *mal* regulon transcription,
786 MdoR biogenesis and regulatory activity, is increased expression of high-affinity components
787 of maltose-specific transport (MalE and LamB).

788 MdoR is reminiscent of the *Vibrio cholerae* MicX sRNA that also mapped to the 3'
789 region of *malG* and regulates levels of outer membrane proteins (Davis and Waldor, 2007).
790 However, the MicX targets are different from MdoR. Furthermore, unlike MdoR, MicX is
791 produced from an independent promoter: Conditions that change the expression of the

792 upstream Mal operon in *V. cholerae* do not affect accumulation of MicX (Davis and Waldor,
793 2007), whereas the levels of MdoR strongly correlate with *malEFG* mRNA levels.
794 Nevertheless, both studies demonstrate that an sRNA resides in the 3'UTR of the *malG*
795 transcript and pinpoint a mechanism where the expression of multiple porins is regulated by
796 sRNAs encoded within transporter mRNAs.
797
798 **CLASH reveals large sRNA-sRNA interaction networks.**

799 Our analyses unearthed an unexpectedly large number (>100) of sRNA-sRNA
800 interactions. The majority of these interactions involved known sRNA seed sequences,
801 suggesting that these could represent *bona fide* interactions that prevent sRNAs from base-
802 pairing with their targets or result in the degradation of the sRNAs. About 40 sRNA-sRNA
803 interactions involving sRNAs from the core genome were detected in the RNase E CLASH
804 dataset, about a quarter of which were also detected in our Hfq data. The relatively low overlap
805 between the datasets is not surprising, given the differences in the growth conditions
806 (virulence inducing conditions for enterohaemorrhagic *E. coli* vs growth transitions for
807 commensal *E. coli* MG1655). Moreover, many of the sRNA-sRNA interactions recovered in
808 association with RNase E are likely duplexes in the process of being degraded. Many of the
809 sRNA-sRNA interactions unique to our Hfq data may represent interactions between anti-
810 sRNAs and seed sequences of sRNAs which may not necessarily involve recruitment/activity
811 of RNase E.
812 The chimeras containing CyaR fragments were of particular interest as CyaR is
813 preferentially expressed during the transition from late exponential to stationary phase (De
814 Lay and Gottesman, 2009; Wassarman et al., 2001) and may therefore play an important role
815 in adaptation to nutrient availability. We could show that overexpression of ArcZ, which base-
816 pairs with CyaR in our CLASH data, significantly reduced CyaR levels (Figure 9B). Using seed
817 mutants and compensatory mutations, we confirmed direct interactions between ArcZ and
818 CyaR and our data suggests that this interaction specifically takes place during the transition
819 phase. Interestingly in *Salmonella*, overexpression of ArcZ showed a dramatic reduction in
820 CyaR bound to Hfq and upregulation of CyaR targets, such as *nadE* (Papenfort et al., 2009),
821 suggesting that this activity is conserved between these two Gram-negative bacteria. A similar
822 type of asymmetric regulation has also been elegantly demonstrated for the Qrr3 sRNA of
823 Vibrio cholera (Feng et al., 2015). The fate of these sRNA-sRNA duplexes may depend on the
824 position of the interaction; It was shown that if sRNA-target RNA base-pair within a stabilizing
825 5' stem, the sRNA will be preferentially degraded (Feng et al., 2015). Consistent with this,
826 folding of the chimeric reads suggests that ArcZ preferentially base-pairs with CyaR at the 5'
827 end (Figure 4), which may alter secondary structures that normally help to stabilize the sRNA.

828 The biological role of ArcZ targeting CyaR is unclear, however, a possible function could be
829 to reduce noise in CyaR expression by preventing CyaR levels from overshooting during the
830 transition phase.
831 ArcZ and CyaR target mRNAs are associated with many different processes. Thus,
832 these interactions are expected to connect multiple pathways. For example, ArcZ regulation
833 of CyaR may connect adaptation to stationary phase/biofilm development (De Lay and
834 Gottesman, 2009; Monteiro et al., 2012) to quorum sensing and cellular adherence (De Lay
835 and Gottesman, 2009). CyaR expression is controlled by the global regulator Crp. Most of the
836 genes controlled by Crp are involved in transport and/or catabolism of amino acids or sugar.
837 Interestingly, ArcZ downregulates the *sdaCB* dicistron which encodes for proteins involved in
838 serine uptake and metabolism (Papenfort et al., 2009). This operon has been shown to be
839 regulated by Crp as well, suggesting that ArcZ can counteract the activity of Crp.
840

841 ## Materials and Methods

842

843 **Bacterial strains and culture conditions**

844 An overview of the bacterial strains used in this study is provided in Supplementary
845 Table 5. The *E. coli* MG1655, TOP10 or TOP10F' strains served as parental strains. The *E.*
846 *coli* K12 strain used for CLASH experiments, MG1655 *hfq*::HTF was previously reported (Tree
847 et al., 2014). Cells were grown in Lysogeny Broth (LB) or minimal medium with supplements
848 (1xM9 salts, 2 mM $MgSO_4$,0.1 mM $CaCl_2$, 0.03 mM thiamine, 0.2% carbon-source) at 37°C
849 under aerobic conditions with shaking at 200 rpm. The media were supplemented with
850 antibiotics where required at the following concentrations: ampicillin (Sigma, UK, A9518) - 100
851 µg/ml, chloramphenicol (Corning, US, C239RI) - 25 µg/ml, kanamycin (Gibco, US, 11815-024)
852 - 50 µg/ml. Where indicated, 0.2% glucose or maltose were used. For induction of sRNA
853 expression from plasmids, 1 mM IPTG, 200 nM anhydrotetracycline hydrochloride (Sigma,
854 1035708-25MG) or 0.2% L-arabinose (Sigma, A3256) were used.
855

856 **Construction of sRNA expression plasmids**

857 For the pulse-overexpression constructs, the sRNA gene of interest was cloned at the
858 transcriptional +1 site under P*ara* control by amplifying the pBAD+1 plasmid (Supplementary
859 Table 5) by inverse PCR using Q5 DNA Polymerase (NEB). The pBAD+1 template is derived
860 from pBAD*myc*HisA (Tree et al., 2014). The sRNA genes and seed mutants (SM) were
861 synthesized as ultramers (IDT; Supplementary Table 5), which served as the forward primers,
862 as described (Tree et al., 2014). The reverse primer (oligo pBAD+1_5P_rev) bears a
863 monophosphorylated 5'-end to allow blunt-end self-ligation. The PCR reaction was digested
864 with 10U DpnI (NEB) for 1h at 37°C and purified by ethanol precipitation. The sRNA-pBAD

865 linear PCR product was circularized by self-ligation, performed as above. Ligations were
866 transformed in DH5α competent cells. Positive transformants were screened by sequencing.
867 The control plasmid pBAD+1 was constructed similarly by self-ligation of the PCR product
868 generated from oligonucleotides pBAD+1_XbaI_fwd and pBAD+1_5P_rev. Small RNA
869 overexpression constructs derived from the pZA21MCS and pZE12*luc* (Expressys) were
870 generated identically, using the indicated ultramers in Supplementary Table 5 as forward
871 primers, and oligos pZA21MCS_5P_rev and pZE12_5P_rev as reverse primers, respectively,
872 and transformed in *E. coli* TOP10F'.
873
874 **Construction of mRNA-superfolder GFP fusions**
875 Supplementary Table 5 lists all the plasmids, gene fragments and primers used for
876 cloning procedures in this work. To construct constitutively expressed, in-frame mRNA-*sfGFP*
877 fusions for the fluorescence reporter studies, the 5'UTR, start codon and first ~5 codons of
878 target genes were cloned under the control of PLtetO-1 promoter in a pXG10-SF backbone
879 as previously described (Corcoran et al., 2012; Urban and Vogel, 2007). Derivatives of the
880 target–GFP fusion plasmids harbouring seed mutations (SM) were generated using synthetic
881 mutated gene-fragments (IDT, Belgium, Supplementary Table 5). To prepare the inserts, the
882 target region of mRNA of interest was either amplified by PCR from *E. coli* genomic DNA or
883 synthesized as g-blocks (IDT, Belgium) and cloned using NheI and NsiI restriction sites.
884 Transformants were screened by restriction digest analysis and verified by Sanger sequencing
885 (Edinburgh Genomics, Edinburgh, UK).
886
887 **Hfq UV Cross-linking, Ligation and Analysis of Hybrids (Hfq-CLASH)**
888 CLASH was performed essentially as described(Waters et al., 2016), with a number
889 of modifications including changes in incubation steps, cDNA library preparation, reaction
890 volumes and UV cross-linking. *E. coli* expressing the chromosomal Hfq-HTF were grown
891 overnight in LB at 37°C with shaking (200 rpm), diluted to starter OD_{600} 0.05 in fresh LB, and
892 re-grown with shaking at 37°C in 750 ml LB. A volume of culture equivalent to 80 OD_{600} per
893 ml was removed at the following cell-densities (OD_{600}): 0.4, 0.8, 1.2, 1.8, 2.4, 3.0 and 4.0, and
894 immediately subjected to UV (254 nm) irradiation for 22 seconds (~500 mJ/cm2) in the Vari-
895 X-linker (van Nues et al., 2017) (https://www.vari-x-link.com). Cells were harvested using a
896 rapid filtration device (van Nues et al., 2017) (https://www.vari-x-link.com) onto 0.45 µM
897 nitrocellulose filters (Sigma, UK, HAWP14250) and flash-frozen on the membrane in liquid
898 nitrogen. Membranes were washed with ~15 ml ice-cold phosphate-buffered saline (PBS),
899 and cells were harvested by centrifugation. Cell pellets were lysed by bead-beating in 1
900 volume per weight TN150 buffer (50mM Tris pH 8.0, 150 mM NaCl, 0.1% NP-40, 5 mM β-
901 mercaptoethanol) in the presence of protease inhibitors (Roche, A32965), and 3 volumes 0.1

902 mm Zirconia beads (Thistle Scientific, 11079101z), by performing 5 cycles of 1 minute
903 vortexing followed by 1-minute incubation on ice. One additional volume of TN150 buffer was
904 added. To reduce the viscosity of the lysate and remove contaminating DNA the lysate was
905 incubated with RQ1 DNase I (10U/ml Promega, M6101) for 30 minutes on ice. Two-additional
906 volumes of TN150 were added and mixed with the lysates by vortexing. The lysates were
907 centrifuged for 20 minutes at 4000 rpm at 4°C and subsequently clarified by a second
908 centrifugation step at 13.4 krpm, for 20 min at 4°C. Purification of the UV cross-linked Hfq-
909 HTF-RNA complexes and cDNA library preparation was performed as described (Granneman
910 et al., 2009). Cell lysates were incubated with 50 µl of pre-equilibrated M2 anti-FLAG beads
911 (Sigma, M8823-5ML) for 1-2 hours at 4°C. The anti-FLAG beads were washed three times 10
912 minutes with 2 ml TN1000 (50 mM Tris pH 7.5, 0.1% NP-40, 1M NaCl) and three times 10
913 minutes with TN150 without protease inhibitors (50 mM Tris pH 7.5, 0.1% NP-40, 150mM
914 NaCl). For TEV cleavage, the beads were resuspended in 250 µl of TN150 buffer (without
915 protease inhibitors) and incubated with home-made GST-TEV protease at room temperature
916 for 1.5 hours. The TEV eluates were then incubated with a fresh 1:100 dilution preparation of
917 RNaceIt (RNase A and T1 mixture; Agilent, 400720) for exactly 5 minutes at 37ºC, after which
918 they were mixed with 0.4g GuHCl (6M, Sigma, G3272-100G), NaCl (300mM), and Imidazole
919 (10mM, I202-25G). Note this needs to be carefully optimized to obtain high-quality cDNA
920 libraries. The samples were then transferred to 50 µl Nickel-NTA agarose beads (Qiagen,
921 30210), equilibrated with wash buffer 1 (6 M GuHCl, 0.1% NP-40, 300 mM NaCl, 50 mM Tris
922 pH 7.8, 10 mM Imidazole, 5 mM beta-mercaptoethanol). Binding was performed at 4ºC
923 overnight with rotation. The following day, the beads were transferred to Pierce SnapCap spin
924 columns (Thermo Fisher, 69725), washed 3 times with wash buffer 1 and 3 times with 1xPNK
925 buffer (10 mM $MgCl_2$, 50mM Tris pH 7.8, 0.1% NP-40, 5 mM beta-mercaptoethanol). The
926 washes were followed by on-column TSAP incubation (Thermosensitive alkaline
927 phosphatase, Promega, M9910) treatment for 1h at 37°C with 8 U of phosphatase in 60 µl of
928 1xPNK, in the presence of 80U RNasin (Promega, N2115). The beads were washed once
929 with 500 µl wash buffer 1 and three times with 500 µl 1xPNK buffer. To add 3'-linkers (App-
930 PE – Supplementary Table 5), the Nickel-NTA beads were incubated in 80 µl 3'-linker ligation
931 mix with (1 X PNK buffer, 1 µM 3'-adapter, 10% PEG8000, 30U Truncated T4 RNA ligase 2
932 K227Q (NEB, M0351L), 60U RNasin). The samples were incubated for 4 hours at 25°C. The
933 5'-ends of bound RNAs were radiolabelled with 30U T4 PNK (NEB, M0201L) and 3µl ^{32}P-
934 γATP (1.1µCi; Perkin Elmer, NEG502Z-500) in 1xPNK buffer for 40 min at 37°C, after which
935 ATP (Roche, 11140965001) was added to a final concentration of 1mM, and the incubation
936 prolonged for another 20 min to complete 5'-end phosphorylation. The resin was washed three
937 times with 500 µl wash buffer 1 and three times with equal volume of 1xPNK buffer. For on-

938 bead 5'-linker ligation, the beads were incubated 16h at 16°C in 1xPNK buffer with 40U T4
939 RNA ligase I (NEB, M0204L), and 1 µl 100 µM L5 adapter (Supplementary Table 5), in the
940 presence of 1mM ATP and 60U RNasin. The Nickel-NTA beads were washed three times with
941 wash buffer 1 and three times with buffer 2 (50 mM Tris–HCl pH 7.8, 50 mM NaCl, 10 mM
942 imidazole, 0.1% NP-40, 5 mM β-mercaptoethanol). The protein-RNA complexes were eluted
943 in two steps in new tubes with 200 µl of elution buffer (wash buffer 2 with 250 mM imidazole).
944 The protein-RNA complexes were precipitated on ice by adding TCA (T0699-100ML) to a final
945 concentration of 20%, followed by a 20-minute centrifugation at 4°C at 13.4 krpm. Pellets were
946 washed with 800 µl acetone, and air dried for a few minutes in the hood. The protein pellet
947 was resuspended and incubated at 65°C in 20 µl 1x NuPage loading buffer (Thermo Scientific,
948 NP0007), resolved on 4–12% NuPAGE gels (Thermo Scientific, NP0323PK2) and visualised
949 by autoradiography. The cross-linked proteins-RNA were cut directly from the gel and
950 incubated with 160 µg of Proteinase K (Roche, 3115801001) in 600 µl wash buffer 2
951 supplemented with 1% SDS and 5 mM EDTA at 55°C for 2-3 hours with mixing. The RNA was
952 subsequently extracted by phenol-chloroform extraction and ethanol precipitated. The RNA
953 pellet was directly resuspended in RT buffer and was transcribed in a single reaction with the
954 SuperScript IV system (Invitrogen, 18090010) according to manufacturer's instructions using
955 the PE_reverse oligo as primer. The cDNA was purified with the DNA Clean and Concentrator
956 5 kit (Zymo Research) and eluted in 11 µl DEPC water. Half of the cDNA (5 µl) was amplified
957 by PCR using Pfu Polymerase (Promega, M7745) with the cycling conditions (95°C for 2 min;
958 20-24 cycles: 95°C for 20s, 52°C for 30s and 72°C for 1 min; final extension of 72°C for 5 min).
959 The PCR primers are listed in Supplementary Table 5. PCR products were treated with 40U
960 Exonuclease 1 (NEB, M0293L) for 1 h at 37°C to remove free oligonucleotide and purified by
961 ethanol precipitation/ or the DNA Clean and Concentrator 5 kit (Zymo Research, D4003T).
962 Libraries were resolved on a 2% MetaPhor agarose (Lonza, LZ50181) gel and 175-300bp
963 fragments were gel-extracted with the MinElute kit (Qiagen, 28004) according to
964 manufacturer's instructions. All libraries were quantified on a 2100 Bionalyzer using the High-
965 Sensitivity DNA assay and a Qubit 4 (Thermo Scientific, Q33226). Individual libraries were
966 pooled based on concentration and barcode sequence identity. Paired-end sequencing (75
967 bp) was performed by Edinburgh Genomics on an Illumina HiSeq 4000 platform.
968
969 **RNA-seq**
970 *E. coli* MG1655 was cultured, UV-irradiated and harvested as described for the CLASH
971 procedure. Total RNA was extracted using the Guanidium thiocyanate phenol method. RNA
972 integrity was assessed with the Prokaryote Total RNA Nano assay on a 2100 Bioanalyzer
973 (Agilent, G2939BA). Sequencing libraries from two biological replicates were prepared by

974 NovoGene using the TruSeq library preparation protocol and 150bp paired-end sequencing
975 was performed on an Illumina NovaSeq 6000 system. This yielded ~7-8 million paired-end
976 reads per sample. For the overexpression analysis of MdoR, we generated RNA-seq libraries
977 using an in-house protocol. Genomic DNA was removed by incubating 10 µg of total RNA with
978 2U Turbo DNase (Thermo Scientific, AM2238) in a 50 µl final volume for 30 minutes at 37°C
979 in the presence of 10 U SuperaseIn RNase Inhibitor (Thermo Scientific, AM2694). RNA was
980 subsequently phenol-chloroform extracted and purified by ethanol-precipitation. Ribosomal
981 RNA was removed with the Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria; Illumina,
982 MRZGN126) according to the manufacturer's instructions. Successful rRNA depletion was
983 verified on the Agilent 2100 Bioanalyzer. The RNA was fragmented for 5 min at 95°C in the
984 presence of Superscript III buffer (Invitrogen) followed by a five-minute incubation on ice.
985 Reverse-transcription (RT) was performed with Superscript III (Thermo Scientific, 18080044)
986 in 20 µl reactions according to manufacturer's procedures using 250 ng of ribosomal RNA
987 depleted RNA and 2.5 µM random hexamers (PE_solexa_hexamer, oligo 73, Supplementary
988 Table 5). The RNA and free primers were degraded using 20U of Exonuclease I (NEB,
989 M0293L) and 50U RNaseIf (NEB, M0243S) and the cDNA was purified with the DNA Clean &
990 Concentrator 5 kit (Zymo Research). Ligation of the 5' adapter (P5_phospho_adapter, oligo
991 39) to the cDNA was performed using CircLigase II (Lucigen, CL9021K) for 6 hours at 60°C,
992 followed by a 10-minute inactivation at 80°C. The cDNA was purified with the DNA Clean &
993 Concentrator 5 kit. Half of the cDNA library was PCR amplified using Pfu polymerase
994 (Promega, M7745) using the P5 forward PCR oligonucleotide and barcoded BC reverse
995 oligonucleotides (200 nM; Supplementary Table 5; 95°C for 2 min, 95°C for 20s, 52°C for 30s
996 and 72°C for 1 min, and a final extension of 72°C for 5 min. 20 cycles of amplification). The
997 PCR products were treated with Exonuclease 1 (NEB, M0293L) for 1 h at 37°C and purified
998 by ethanol precipitation. Libraries were resolved on a 2% MetaPhor agarose gel 200-500 bp
999 fragments were gel-extracted using the MinElute kit. All libraries were quantified on a 2100
1000 Bionalyzer using the High-Sensitivity DNA assay (Agilent, 5067-4627). Individual libraries
1001 were pooled in equimolar amounts. Paired-end sequencing (75 bp) was performed by
1002 Edinburgh Genomics on the Illumina HiSeq 4000 platform.
1003
1004 **Small RNA overexpression studies**
1005 Individual TOP10F' clones carrying pZA21 and pZE12-derived sRNA constructs and
1006 control plasmids combinations (Supplementary Table 5) were cultured to OD_{600} 0.1 and
1007 expression of sRNAs was induced with IPTG and anhydrotetracycline hydrochloride (Sigma,
1008 I6758-1G and 1035708-25MG) for one hour. Cells were collected by centrifugation for 30
1009 seconds at 14000 rpm, flash-frozen in liquid nitrogen and total RNA was isolated as above.

1010 Genomic DNA was digested with Turbo DNase (Thermo Scientific, AM2238) , then the RNA
1011 was purified with RNAClean XP beads (Beckman Coulter, A63987). Gene expression was
1012 quantified by RT-qPCR (see below) using 10 ng total RNA as template, and expressed as fold
1013 change relative to the reference sample containing pJV300 (Sittka et al., 2007) or pZA21.
1014 For pulse-overexpression studies overnight MG1655 cultures containing pBAD::sRNA
1015 and empty pBAD+1 control plasmids were inoculated in fresh LB-ampicillin medium at a
1016 starting OD_{600} of 0.05 and grown aerobically at 37°C to OD_{600} 0.4. Pre-induction (0 min) and
1017 post-induction samples were harvested. For induction, cultures were supplemented with L-
1018 arabinose (Sigma, A3256) and rapidly collected by filtration and flash-frozen in liquid nitrogen
1019 at the indicated time-points. RNA was extracted from three biological replicate time-series,
1020 followed by RNA-seq library preparation, next generation sequencing and DESeq2 analysis
1021 of differentially expressed genes.
1022
1023 **GFP reporter system to quantify sRNA effect on target expression**
1024 A two-plasmid system was used to express each sRNA, and mRNA-*sfGFP* fusions
1025 (Corcoran et al., 2012; Urban and Vogel, 2007) with modifications. The sRNA and sfGFP-
1026 fusion plasmids were co-transformed in *E. coli* TOP10 cells by electroporation and cells were
1027 maintained on dual selection with ampicillin and chloramphenicol. In TOP10 cells, the mRNA-
1028 *sfGFP* constructs are constitutively expressed, whereas sRNA expression requires L-
1029 arabinose induction. The expression of sfGFP-fused targets in the presence or absence of
1030 sRNAs was quantified at the protein level, by plate reader experiments and at the RNA level,
1031 by RT-qPCR.
1032 For the plate reader experiments, a single colony of bacterial strain harbouring a
1033 sRNA-target-sfGFP combination was inoculated in a 96-well Flat Bottom Transparent
1034 Polystyrene plate with lid (Thermo Scientific, 152038) and cultured overnight at 37°C in 100
1035 µl LB supplemented with antibiotics and L-arabinose (Sigma, A3256) to induce expression of
1036 sRNAs. Next day, each overnight inoculum was diluted 1:100 by serial dilution, in triplicate, in
1037 LB with freshly prepared L-arabinose to a final volume of 100 µl. Cultures were grown in a 96-
1038 well plate in an Infinite 200 Pro plate reader (Tecan) controlled by i-control software (Tecan)
1039 for 192 cycles at 37°C with 1 min orbital shaking (4 mm amplitude) every 5th minute. To
1040 monitor optical density over time, the following parameters were used: wavelength 600 nm,
1041 bandwidth 9 nm. Fluorescence was monitored with excitation wavelength 480 nm, bandwidth
1042 9 nm and emission wavelength 520 nm, bandwidth 20 nm. Measurements were recorded at
1043 5-minute intervals, by top reading. Raw data was processed following guidance from previous
1044 reports (Urban and Vogel, 2007). First, the range of linearity of increase of fluorescence with
1045 OD_{600} was identified for all individual triplicates. Only the linearity range common to all

1046 triplicates was considered for further analysis. For each set of triplicates, the mean
1047 fluorescence was calculated at each OD_{600}. To correct for background and cell
1048 autofluorescence, the mean fluorescence of a strain with plasmid pXG-0 was subtracted from
1049 all strains with GFP plasmids at the equivalent OD_{600}. Ultimately, a curve was generated for
1050 each sample, plotting the background-corrected fluorescence (GFP) versus OD_{600}. The
1051 experiments were performed for three biological replicates, and mean values and standard
1052 error of the means calculated for each strain.
1053
1054 **RT-qPCR**
1055 Total RNA (12.5µg) was treated with 2U of Turbo DNase (Thermo Scientific, AM2238)
1056 for 1 hour at 37°C in a 10 µl reaction in the presence of 2U SuperaseIn RNase inhibitor
1057 (Thermo Scientific, AM2694). The DNase was inactivated by 10 minutes incubation at 75°C.
1058 Reverse transcription (RT) was performed in a single reaction for all target genes of interest
1059 using a mix of gene-specific RT primers at 3.5 µM concentration each. After addition of 2.5 µl
1060 RT primer mix, the RNA and primers were denatured at 70°C for 3 min, then snap chilled and
1061 incubated on ice for 5 min. RT was performed for 1 hour at 55°C with SuperScript III (Thermo
1062 Scientific, 18080051) using 5 µl of RNA-RT primers mix in 10 µl final volume (100 U
1063 Superscript III, 2.5 mM DTT, 1xFS Buffer, 0.75 mM dNTPs) in the presence of 1U RNasin
1064 (Promega, N2115). RT was followed by treatment with 5U RNase H for 30 min at 37°C to
1065 remove the RNA from the RNA-cDNA duplexes. The cDNA was diluted 10-fold with DEPC
1066 water. Quantitative PCR was performed on 50ng of DNAse I-treated total RNA using the
1067 Brilliant III UltraFast SYBR Green QPCR Master Mix (Agilent, #600883) and the Luna
1068 Universal One-Step RT-qPCR Kit (NEB, E3005E) according to manufacturer's instructions.
1069 The qPCRs were run on a LightCycler 480 (Roche), and the specificity of the product was
1070 assessed by generating melting curves, as follows: 65°C-60s, 95°C (0.11 ramp rate with 5
1071 acquisitions per °C, continuous). The data analyses were performed with the IDEAS2.0
1072 software, at default settings: Absolute Quantification/Fit Points for Cp determination and Melt
1073 Curve Genotyping.
1074 The qPCR efficiency of primer pairs was assessed by performing standard curves by serial
1075 dilution of template RNA or genomic DNA. Negative controls such as -RT or no template
1076 control were used throughout, and the qPCR for all samples was performed in technical
1077 triplicate. Outliers from the samples with technical triplicate standard deviations of Cp > 0.3
1078 were discarded from the analyses. To calculate the fold-change relative to the control, the 2^{-}
1079 ^{ddCp} method was employed, using *recA* or 5S rRNA (*rrfD*) as the reference genes where
1080 indicated. Experiments were performed for minimum two biological replicates, and the mean
1081 fold-change and standard error of the mean were computed. Unless otherwise stated,

1082 significance of the fold-change difference compared to the reference sample control (for which
1083 fold-change =1) was tested with a one-sample t-test.
1084
1085 **Northern Blot analysis**
1086 Total RNA was extracted from cell lysates by GTC-Phenol extraction. For large RNA
1087 fragments, 10 µg of total RNA was resolved on a 1.25% BPTE- gel (pH 7) and transferred to
1088 a nylon membrane (HyBond N+, GEHealthcare, RPN1210B) by capillarity. For short RNA
1089 fragments, 10 µg total RNA was separated on an 8% polyacrylamide TBE-Urea gel and
1090 transferred to a nitrocellulose membrane by electroblotting for four hours at 50 V. Membranes
1091 were pre-hybridised in 10 ml of UltraHyb Oligo Hyb (Thermo Scientific, AM8663) for one hour
1092 and probed with ^{32}P-labeled DNA oligo at 42°C for 12-18 hours in a hybridization oven. The
1093 sequences of the probes used for Northern blot detection are detailed in Supplementary Table
1094 5. Membranes were washed twice with 2xSSC + 0.5% SDS solution for 10 minutes and
1095 visualized using a Phosphor imaging screen and FujiFilm FLA-5100 Scanner (IP-S mode).
1096 For detection of highly abundant species (5S rRNA) autoradiography was used for exposure.
1097
1098 **Western blot analyses**
1099 A fraction of the *E. coli* MG1655 Hfq::*htf* lysates used for the RNA-seq experiments
1100 using strains cultured, cross-linked, harvested and lysed in identical conditions as the CLASH
1101 experiments containing 40 µg protein was run on PAGE gels and transferred to a nitrocellulose
1102 membrane. The membranes were blocked for one hour in blocking solution (5% non- fat milk
1103 in PBST (1X phosphate saline buffer, 0.1% Tween-20). To detect Hfq-HTF protein, the
1104 membrane was probed overnight at 4°C with the Rabbit anti-TAP polyclonal primary antibody
1105 (Thermo Fisher, 1:5000 dilution in blocking solution), which recognizes an epitope at the
1106 region between the TEV-cleavage site and His6. For the loading control we used a rabbit
1107 polyclonal to GroEL primary antibody (Abcam, 1:150000 dilution, ab82592), for 2 hours at
1108 room temperature. After 3x10 min PBST washes, the membranes were blotted for one hour
1109 with a Goat anti-rabbit IgG H&L (IRDye 800) secondary antibody (Abcam, ab216773, 1:10000
1110 in blocking solution) at room temperature. Finally, after three 10-minute PBST washes, the
1111 blot was rinsed in PBS, and the proteins were visualised with a LI-COR (Odyssey CLx) using
1112 the 800 nm channel and scan intensity 4. Image acquisition and quantifications were
1113 performed with the Image Studio Software.
1114
1115 **Primer extension analysis**
1116 One microgram total RNA was reverse-transcribed using SuperScript III reverse
1117 transcriptase (Thermo Scientific, 18080051) using ^{32}P-radiolabelled oligonucleotides as

1118 primers (Supplementary Table 5). Primers were added to the RNA and annealing was
1119 performed by heating the samples at 85°C for three minutes and then snap chilling them on
1120 ice. The RT was performed for one hour at 45°C, followed by Exonuclease I and RNaseIf
1121 (NEB M0293L and M0243S) (0.5 µl each) treatment for 30 minutes at 37°C. Reactions were
1122 stopped by mixing with an equal volume of 2XRNA loading dye (NEB, B0363S), 2 minutes
1123 incubation at 95°C and snap chilled. The sequencing ladders were prepared with Sequenase
1124 v2.0 (Thermo Scientific, 70775Y200UN) according to specified instructions. Samples were
1125 resolved on 6% PAA/8M TBE-urea gels and visualized using the FujiFilm FLA5100 scanner.
1126

1127 **Construction of the MdoR seed-mutant strain**

1128 To mutate the chromosomal copy of MdoR, we used the λRed system (Datsenko and
1129 Wanner, 2000). We amplified the integration cassette from plasmid pKD4 with ultramers 895
1130 and 896, containing homology regions to the coding sequence of *malG*, the desired MdoR
1131 sequence and to the region immediately downstream of the Rho-independent terminator,
1132 respectively. With this design, the scar after removal of the *Kan*^{r} cassette was expected at a
1133 site outside the MdoR/*malG* sequence. The PCR product was electroporated in *E. coli*
1134 MG1566 strains carrying the pKD46 plasmid from which λRed recombinase was induced with
1135 10 mM L-arabinose. Correct replacement of the MdoR seed sequence was screened by
1136 colony PCR using primer pairs: 725 & 909 and 726 & 910 (Supplementary Table 5). The
1137 antibiotic resistance cassette was removed from substitution mutants by FLP-recombinase
1138 expressed constitutively from pE-FLP (St-Pierre et al., 2013). Successful allele replacement
1139 was confirmed by Sanger sequencing.
1140

1141 **Polysome profiling analyses**

1142 Wild-type *E. coli* MG1655 containing empty pBAD plasmid and an isogenic strain
1143 containing pBAD::MdoR were grown in LB until OD_{600} 0.4, then treated for 15 minutes with L-
1144 arabinose (Sigma, A3256) to induce overexpression of MdoR, and cycloheximide (Sigma,
1145 C4859-1ML) at a final concentration of 100 µg/ml for 3 minutes. 200 ml of cells were harvested
1146 by rapid filtration and flash frozen. The cells were washed in ice-cold PBS supplemented with
1147 100µg/ml cycloheximide.

1148 Polysomal profiling was performed according to previously described protocols
1149 (Bernabò et al., 2017; Lunelli et al., 2016) with minor changes in the lysis buffer (10 mM NaCl,
1150 10 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.5), 100 µg/ml cycloheximide, 1% (w/v) Na-Deoxycholate,
1151 1U RQ DNAse I (Promega, M6101), 0.6U/mL RiboLock (Thermo Scientific, EO0381) in DEPC
1152 water). Lysates were kept on ice for 30 min, centrifuged 3X at 15000 g for 10 min. The
1153 supernatants were loaded on a linear 10%–30% [w/v] sucrose gradient and centrifuged for 4

1154 hours using a SW41 rotor at 40000 rpm in a Beckman Optima XPN-100 Ultracentrifuge.
1155 Fractions of 1 mL in volume were collected monitoring the absorbance at 254 nm with the UA-
1156 6 UV/VIS detector (Teledyne Isco). Fractions from the entire gradient (total RNA) and from the
1157 fractions corresponding to ribosomes (70S) and polysomes (polysomal RNA) were pooled
1158 and RNA was purified by acid phenol–chloroform extraction according to (Tebaldi et al., 2012).
1159

1160 **Terminator™ 5′-PhosphateDependent Exonuclease treatment**

1161 Ten micrograms of total RNA extracted from cell-samples at OD_{600} 1.2 and 1.8 were
1162 treated with 5'-Terminator Dependent Exonuclease (Lucigen, TER51020) as per manufacturer
1163 instructions using Buffer A. The reaction was terminated by phenol extraction and ethanol
1164 precipitation, and the RNA was loaded on 8% polyacrylamide-urea gels and transferred to
1165 nylon membranes that were probed for MdoR, CpxQ, RybB and 5S rRNA (Supplementary
1166 Table 5).
1167

1168 **Seed mutant studies**

1169 Wild-type MG1655 and seed mutant strains were grown overnight in minimal medium
1170 with glucose. Next day, each starter culture was split and inoculated at OD_{600} 0.05 in fresh M9
1171 medium with glucose or maltose as the sole carbon source. Growth was monitored and cells
1172 were harvested at OD_{600} 0.5. Total RNA was extracted, and gene expression was quantified
1173 by RT-qPCR or Northern Blot.
1174

1175 **Computational analysis**

1176 *Pre-processing of the raw sequencing data.*

1177 Raw sequencing reads in fastq files were processed using a pipeline developed by
1178 Sander Granneman, which uses tools from the pyCRAC package (Webb et al., 2014). The
1179 entire pipeline is available at https://bitbucket.org/sgrann/). The CRAC_pipeline_PE.py
1180 pipeline first demultiplexes the data using pyBarcodeFilter.py and the in-read barcode
1181 sequences found in the L5 5' adapters. Flexbar then trims the reads to remove 3'-adapter
1182 sequences and poor-quality nucleotides (Phred score <23). Using the random nucleotide
1183 information present in the L5 5'-adaptor sequences, the reads are then collapsed to remove
1184 potential PCR duplicates. The reads were then mapped to the *E. coli* MG1655 genome using
1185 Novoalign (www.novocraft.com). To determine to which genes the reads mapped to, we
1186 generated an annotation file in the Gene Transfer Format (GTF). This file contains the start
1187 and end positions of each gene on the chromosome as well as what genomic features (i.e.
1188 sRNA, protein- coding, tRNA) it belongs to. To generate this file, we used the Rockhopper
1189 software (Tjaden, 2015) on *E. coli* rRNA-depleted total RNA-seq data (generated by Christel
1190 Sirocchi), a minimal GTF file obtained from ENSEMBL (without UTR information). The

1191 resulting GTF file contained information not only on the coding sequences, but also complete
1192 5' and 3' UTR coordinates. We then used pyReadCounters.py with Novoalign output files as
1193 input and the GTF annotation file to count the total number of unique cDNAs that mapped to
1194 each gene.
1195
1196 *Normalization steps.*
1197 To normalise the read count data generated with pyReadCounters.py and to correct
1198 for differences in library depth between time-points, we calculated Transcripts Per Million
1199 reads (TPM) for each gene. Briefly, for each time-point the raw counts for each gene was first
1200 divided by the gene length and then divided by the sum of all the values for the genes in that
1201 time-point to normalize for differences in library depth. The TPM values for each OD_{600} studied
1202 were divided by the TPM values of the first sample (OD_{600} 0.4) and were then log_2-normalized.
1203 The log_2-normalized fold-changes were used to compare RNA-seq and Hfq-cross-linking
1204 profiles among samples, and to perform k-means clustering with the python sklearn.
1205 cluster.KMeans class.
1206
1207 *Hfq-binding coverage plots.*
1208 For the analysis of the Hfq binding sites the pyCRAC package (Webb et al., 2014) was
1209 used (versions. 1.3.2-1.4.4). The pyBinCollector tool was used to generate Hfq cross-linking
1210 distribution plots over genomic features. First, PyCalculateFDRs.py was used to identify the
1211 significantly enriched Hfq-binding peaks (minimum 10 reads, minimum 20 nucleotide
1212 intervals). Next, pyBinCollector was used to normalize gene lengths by dividing their
1213 sequences into 100 bins and calculate nucleotide densities for each bin. To generate the
1214 distribution profile for all genes individually, we normalized the total number of read clusters
1215 (assemblies of overlapping cDNA sequences) covering each nucleotide position by the total
1216 number of clusters that cover the gene. Motif searches were performed with pyMotif.py using
1217 the significantly enriched Hfq-binding peaks (FDR intervals). The 4-8 nucleotide k-mers with
1218 Z-scores above the indicated threshold were used for making the motif logo with the k-mer
1219 probability logo tool (Wu and Bartel, 2017) with the -ranked option (http://kplogo.wi.mit.edu/).
1220
1221 *Analysis of chimeric reads.*
1222 Chimeric reads were identified using the hyb package using default settings (Travis et
1223 al., 2013) and further analysed using the pyCRAC package (Webb et al., 2014). To apply this
1224 single-end specific pipeline to our paired-end sequencing data, we joined forward and reverse
1225 reads using FLASH (Magoč and Salzberg, 2011), which merges overlapping paired reads into
1226 a single read. These, as well as any remaining single reads, were then analysed using hyb.
1227 The -anti option for the hyb pipeline was used to be able to use a genomic *E. coli* hyb database,

1228 rather than a transcript database. Uniquely annotated hybrids (.ua.hyb) were used in
1229 subsequent analyses. To visualise the hybrids in the genome browser, the .ua.hyb output files
1230 were converted to the GTF format. To generate distribution plots for the genes to which the
1231 chimeric reads mapped, the parts of the chimeras were clustered with pyClusterReads.py and
1232 BEDtools (Quinlan and Hall, 2010) (intersectBed) was used to remove clusters that map to
1233 multiple regions. To produce the coverage plots with pyBinCollector, each cluster was counted
1234 only once, and the number of reads belonging to each cluster was ignored.
1235
1236 *sRNA density plots.*
1237 To visualize the nucleotide read density of sRNA-target pairs for a given sRNA, we
1238 first merged the hyb datasets for all OD_{600} and biological replicates and filtered the interactions
1239 that were found statistically significant in the unified dataset. For each sRNA-target pair in the
1240 filtered dataset, the hit counts at each nucleotide position for all chimeras were summed. The
1241 count data was log_2-normalised (actually log_2(Chimera count +1) to avoid NaN for nucleotide
1242 positions with 0 hits when log-transforming the data).
1243
1244 *sRNA-sRNA network visualization.*
1245 Only the sRNA-sRNA chimeric reads representing statistically significant interactions
1246 in the merged CLASH dataset were considered. For each such interaction, chimera counts
1247 corresponding in either orientation were summed, log_2-transformed and visualized with the
1248 igraph Python package.
1249
1250 *Differential expression analyses*.
1251 For the differential expression analyses DESeq2 was used (Love et al., 2014). Three
1252 MdoR pulse-overexpression datasets were compared to three pBAD Control overexpression
1253 datasets. Only differentially expressed genes that had an adjusted p-value of 0.05 or lower
1254 were considered significant.
1255
1256 *Multiple sequence alignments and conservation analyses*.
1257 The homologous sequences of MdoR in other enterobacteria were retrieved by
1258 BLAST. JalView was used for the multiple sequence alignments, using the MAFFT algorithm
1259 (Waterhouse et al., 2009).
1260
1261 **Data and Code availability.**
1262 The next generation sequencing data have been deposited on the NCBI Gene
1263 Expression Omnibus (GEO) with accession number GSE123050. The python pyCRAC(Webb
1264 et al., 2014), kinetic-CRAC and GenomeBrowser software packages used for analysing the

data are available from https://bitbucket.org/sgrann (pyCRAC up to version 1.3.3), https://git.ecdf.ed.ac.uk/sgrannem/ and pypi (https://pypi.org/user/g_ronimo/). The hyb pipeline for identifying chimeric reads is available from https://github.com/gkudla/hyb. The FLASH algorithm for merging paired reads is available from https://github.com/dstreett/FLASH2. Bedgraph and Gene Transfer Format (GTF) generated from the analysis of the Hfq CLASH, RNA-seq and TEX RNA-seq data(Thomason et al., 2015) are available from the Granneman lab DataShare repository (https://datashare.is.ed.ac.uk/handle/10283/2915).

Acknowledgement

We are grateful to Lionello Bossi and Meriem El Karoui for their valuable feedback on the project and fruitful discussions. We thank Christel Sirocchi for help with preparing *E. coli* RNA-seq libraries, Erica de Leau for expert technical assistance and the members of the Granneman lab for critically reading the manuscript. We thank Jean-Marie Pages and Julia Vergalli for help the analyses of LamB levels and for fruitful discussions. This work was supported by grants from the Wellcome Trust (091549 to S.G and 102334 to I.A.I.), the Wellcome Trust Centre for Cell Biology core grant (092076), a Medical Research Council non Clinical Senior Research Fellowship (MR/R008205/1 to S.G.), the Australian National Health and Medical Research Council Project grants (GNT1067241 and GNT1139313 to J.J.T) and the Autonomous Province of Trento (Axonomix to G.V and M.M.). Sequencing of Hfq CLASH data was carried out by Edinburgh Genomics that is partly supported through core grants from NERC (R8/H10/56), MRC (MR/K001744/1) and BBSRC (BB/J004243/1).

Competing interest.

The authors declare no competing financial interest.

Materials & Correspondence

All requests for code, materials and reagents should be sent to Sander Granneman (sgrannem@ed.ac.uk)

References

Alba BM, Gross CA. 2004. Regulation of the Escherichia coli ??E-dependent envelope stress response. *Mol Microbiol* **52**:613–619. doi:10.1111/j.1365-2958.2003.03982.x

Baev M V., Baev D, Jansco Radek A, Campbell JW. 2006a. Growth of Escherichia coli

1299 MG1655 on LB medium: Monitoring utilization of amino acids, peptides, and
1300 nucleotides with transcriptional microarrays. *Appl Microbiol Biotechnol* **71**:317–322.
1301 doi:10.1007/s00253-005-0310-5
1302 Baev M V., Baev D, Radek AJ, Campbell JW. 2006b. Growth of Escherichia coli MG1655 on
1303 LB medium: Determining metabolic strategy with transcriptional microarrays. *Appl*
1304 *Microbiol Biotechnol* **71**:323–328. doi:10.1007/s00253-006-0392-8
1305 Balleza E, López-Bojorquez LN, Martínez-Antonio A, Resendis-Antonio O, Lozada-Chávez I,
1306 Balderas-Martínez YI, Encarnación S, Collado-Vides J. 2009. Regulation by
1307 transcription factors in bacteria: Beyond description. *FEMS Microbiol Rev* **33**:133–151.
1308 doi:10.1111/j.1574-6976.2008.00145.x
1309 Bandyra KJ, Said N, Pfeiffer V, Górna MW, Vogel J, Luisi BF. 2012. The Seed Region of a
1310 Small RNA Drives the Controlled Destruction of the Target mRNA by the
1311 Endoribonuclease RNase E. *Mol Cell* **47**:943–953. doi:10.1016/j.molcel.2012.07.015
1312 Beisel CL, Storz G. 2011. The Base-Pairing RNA Spot 42 Participates in a Multioutput
1313 Feedforward Loop to Help Enact Catabolite Repression in Escherichia coli. *Mol Cell*
1314 **41**:286–297. doi:10.1016/j.molcel.2010.12.027
1315 Bernabò P, Tebaldi T, Groen EJN, Lane FM, Perenthaler E, Mattedi F, Newbery HJ, Zhou H,
1316 Zuccotti P, Potrich V, Shorrock HK, Muntoni F, Quattrone A, Gillingwater TH, Viero G.
1317 2017. In Vivo Translatome Profiling in Spinal Muscular Atrophy Reveals a Role for SMN
1318 Protein in Ribosome Biology. *Cell Rep*. doi:10.1016/j.celrep.2017.10.010
1319 Boos W, Shuman H. 1998. Maltose / Maltodextrin System of Escherichia coli : Transport ,
1320 Metabolism , and Regulation Maltose / Maltodextrin System of Escherichia coli :
1321 Transport , Metabolism , and Regulation. *Microbiol Mol Biol Rev* **62**:204–229.
1322 Bossi L, Figueroa-Bossi N. 2007. A small RNA downregulates LamB maltoporin in
1323 Salmonella. *Mol Microbiol* **65**:799–810. doi:10.1111/j.1365-2958.2007.05829.x
1324 Bossi L, Maloriol D, Figueroa-Bossi N. 2008. Porin biogenesis activates the σE response in
1325 Salmonella hfq mutants. *Biochimie* **90**:1539–1544. doi:10.1016/j.biochi.2008.06.001
1326 Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J. 2008. Small RNA binding to 5'
1327 mRNA coding region inhibits translational initiation. *Mol Cell* **32**:827–837.
1328 Bruce HA, Du D, Matak-Vinkovic D, Bandyra KJ, Broadhurst RW, Martin E, Sobott F,
1329 Shkumatov A V., Luisi BF. 2018. Analysis of the natively unstructured RNA/protein-
1330 recognition core in the Escherichia coli RNA degradosome and its interactions with
1331 regulatory RNA/Hfq complexes. *Nucleic Acids Res*. doi:10.1093/nar/gkx1083
1332 Chao Y, Li L, Girodat D, Forstner KU, Said N, Corcoran C, Smiga M, Papenfort K, Reinhardt
1333 R, Wieden HJ, Luisi BF, Vogel J. 2017. *in vivo* cleavage map illuminates the central
1334 cole of RNase E in coding and non-coding RNA pathways. *Mol Cell* **65**:39–51.
1335 doi:10.1016/j.molcel.2016.11.002
1336 Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. 2012. An atlas of Hfq-bound
1337 transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J*
1338 **31**:4005–4019.
1339 Chao Y, Vogel J. 2016. A 3' UTR-Derived Small RNA Provides the Regulatory Noncoding
1340 Arm of the Inner Membrane Stress Response. *Mol Cell* **61**:352–363.
1341 doi:10.1016/j.molcel.2015.12.023
1342 Chen S, Zhang A, Blyn LB, Storz G. 2004. MicC, a second small-RNA regulator of omp
1343 protein expression in Escherichia coli. *J Bacteriol*. doi:10.1128/JB.186.20.6689-
1344 6697.2004
1345 Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JCD, Vogel J. 2012.
1346 Superfolder GFP reporters validate diverse new mRNA targets of the classic porin
1347 regulator, MicF RNA. *Mol Microbiol* **84**:428–445.
1348 Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in
1349 Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**:6640–5.
1350 doi:10.1073/pnas.120163297
1351 Davis BM, Waldor MK. 2007. RNase E-dependent processing stabilizes MicX, a Vibrio
1352 cholerae sRNA. *Mol Microbiol*. doi:10.1111/j.1365-2958.2007.05796.x
1353 De Las Peñas A, Connolly L, Gross CA. 1997. The sigmaE-mediated response to

1354 extracytoplasmic stress in Escherichia coli is transduced by RseA and RseB, two
1355 negative regulators of sigmaE. *Mol Microbiol* **24**:373–385. doi:9159523
1356 De Lay N, Gottesman S. 2009. The crp-activated small noncoding regulatory RNA CyaR
1357 (RyeE) links nutritional status to group behavior. *J Bacteriol*. doi:10.1128/JB.01157-08
1358 Feng L, Rutherford ST, Papenfort K, Bagert JD, Van Kessel JC, Tirrell DA, Wingreen NS,
1359 Bassler BL. 2015. A Qrr noncoding RNA deploys four different regulatory mechanisms
1360 to optimize quorum-sensing dynamics. *Cell*. doi:10.1016/j.cell.2014.11.051
1361 Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA. 2011. Small RNAs endow a
1362 transcriptional activator with essential repressor functions for single-tier control of a
1363 global stress regulon. *Proc Natl Acad Sci U S A* **108**:12875–12880.
1364 Granneman S, Kudla G, Petfalski E, Tollervey D. 2009. Identification of protein binding sites
1365 on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of
1366 cDNAs. *Proc Natl Acad Sci U S A* **106**:9613–9618. doi:10.1073/pnas.0901997106
1367 Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G. 2014. MicL, a new
1368 $σ$E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the
1369 major outer membrane lipoprotein. *Genes Dev* **28**:1620–1634.
1370 doi:10.1101/gad.243485.114
1371 Han K, Tjaden B, Lory S. 2016. GRIL-seq provides a method for identifying direct targets of
1372 bacterial small regulatory RNA by in vivo proximity ligation. *Nat Microbiol* **2**:16239.
1373 Helwak A, Kudla G, Dudnakova T, Tollervey D. 2013. Mapping the human miRNA
1374 interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**:654–665.
1375 doi:10.1016/j.cell.2013.03.043
1376 Holmqvist E, Vogel J. 2018. RNA-binding proteins in bacteria. *Nat Rev Microbiol*.
1377 doi:10.1038/s41579-018-0049-5
1378 Holmqvist E, Wagner EGH. 2017. Impact of bacterial sRNAs in stress responses. *Biochem*
1379 *Soc Trans*. doi:10.1042/BST20160363
1380 Holmqvist E, Wright PR, Li L, Bischler T, Barquist L, Reinhardt R, Backofen R, Vogel J.
1381 2016. Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA
1382 revealed by UV crosslinking in~vivo. *EMBO J* **35**:e201593360--1011.
1383 Hör J, Gorski SA, Vogel J. 2018. Bacterial RNA Biology on a Genome Scale. *Mol Cell* 785–
1384 799. doi:10.1016/j.molcel.2017.12.023
1385 Hör J, Vogel J. 2017. Global snapshots of bacterial RNA networks. *EMBO J* **36**:245–247.
1386 doi:10.15252/embj.201696072
1387 Johansen J, Rasmussen AA, Overgaard M, Valentin-Hansen P. 2006. Conserved small non-
1388 coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer
1389 membrane proteins. *J Mol Biol* **364**:1–8.
1390 Kenyon WJ, Thomas SM, Johnson E, Pallen MJ, Spector MP. 2005. Shifts from glucose to
1391 certain secondary carbon-sources result in activation of the extracytoplasmic function
1392 sigma factor $σ$E in Salmonella enterica serovar Typhimurium. *Microbiology*
1393 **151**:2373–2383. doi:10.1099/mic.0.27649-0
1394 Khemici V, Carpousis AJ. 2004. The RNA degradosome and poly(A) polymerase of
1395 Escherichia coli are required in vivo for the degradation of small mRNA decay
1396 intermediates containing REP-stabilizers. *Mol Microbiol*. doi:10.1046/j.1365-
1397 2958.2003.03862.x
1398 Kortmann J, Narberhaus F. 2012. Bacterial RNA thermometers: molecular zippers and
1399 switches. *Nat Rev Microbiol* **10**:255–265.
1400 Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D. 2011. Cross-linking, ligation, and
1401 sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci U S*
1402 *A* **108**:10010–10015. doi:10.1073/pnas.1017386108
1403 Lalaouna D, Carrier M-C, Semsey S, Brouard J-S, Wang J, Wade JT, Massé E. 2015. A 3'
1404 External Transcribed Spacer in a tRNA Transcript Acts as a Sponge for Small RNAs to
1405 Prevent Transcriptional Noise. *Mol Cell* **58**:393–405.
1406 Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker
1407 IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* **6**:26. doi:10.1186/1748-7188-6-
1408 26

1409 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
1410 RNA-seq data with DESeq2. *Genome Biol* **15**:550.
1411 Lunelli L, Bernabò P, Bolner A, Vaghi V, Marchioretto M, Viero G. 2016. Peering at Brain
1412 Polysomes with Atomic Force Microscopy. *J Vis Exp*. doi:10.3791/53851
1413 Mäder U, Nicolas P, Depke M, Pané-Farré J, Debarbouille M, van der Kooi-Pol MM, Guérin
1414 C, Dérozier S, Hiron A, Jarmer H, Leduc A, Michalik S, Reilman E, Schaffer M, Schmidt
1415 F, Bessières P, Noirot P, Hecker M, Msadek T, Völker U, van Dijl JM. 2016.
1416 Staphylococcus aureus Transcriptome Architecture: From Laboratory to Infection-
1417 Mimicking Conditions. *PLoS Genet* **12**:e1005962.
1418 Magoč T, Salzberg SL. 2011. FLASH: Fast length adjustment of short reads to improve
1419 genome assemblies. *Bioinformatics*. doi:10.1093/bioinformatics/btr507
1420 Martínez-Antonio A, Janga SC, Thieffry D. 2008. Functional organisation of Escherichia coli
1421 transcriptional regulatory network. *J Mol Biol* **381**:238–247.
1422 doi:10.1016/j.jmb.2008.05.054
1423 Melamed S, Faigenbaum-Romm R, Peer A, Reiss N, Shechter O, Bar A, Altuvia Y, Argaman
1424 L, Margalit H. 2018. Mapping the small RNA interactome in bacteria using RIL-seq. *Nat*
1425 *Protoc* **13**:1–33.
1426 Melamed S, Peer A, Faigenbaum-Romm R, Gatt YEE, Reiss N, Bar A, Altuvia YYY,
1427 Argaman L, Margalit H. 2016. Global Mapping of Small RNA-Target Interactions in
1428 Bacteria. *Mol Cell* **63**:884–897. doi:10.1016/j.molcel.2016.07.026
1429 Miyakoshi M, Chao Y, Vogel J. 2015a. Regulatory small RNAs from the 3' regions of
1430 bacterial mRNAs. *Curr Opin Microbiol*. doi:10.1016/j.mib.2015.01.013
1431 Miyakoshi M, Chao Y, Vogel J. 2015b. Cross talk between ABC transporter mRNAs via a
1432 target mRNA-derived sponge of the GcvB small RNA. *EMBO J* **34**:e201490546--1492.
1433 doi:10.15252/embj.201490546
1434 Monteiro C, Papenfort K, Hentrich K, Ahmad I, Le Guyon S, Reimann R, Grantcharova N,
1435 Römling U. 2012. Hfq and Hfq-dependent small RNAs are major contributors to
1436 multicellular development in Salmonella enterica serovar typhimurium. *RNA Biol*.
1437 doi:10.4161/rna.19682
1438 Moon K, Gottesman S. 2011. Competition among Hfq-binding small RNAs in Escherichia
1439 coli. *Mol Microbiol* **82**:1545–1562. doi:10.1111/j.1365-2958.2011.07907.x
1440 Morita T, Maki K, Aiba H. 2005. RNase E-based ribonucleoprotein complexes: Mechanical
1441 basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev*.
1442 doi:10.1101/gad.1330405
1443 Mouali Y El, Gaviria-Cantin T, Sá Nchez-Romero MA, Gibert M, Westermann AJ, Rg Vogel
1444 J, Balsalobre C. 2018. CRP-cAMP mediates silencing of Salmonella virulence at the
1445 post-transcriptional level 1–26. doi:10.1371/journal.pgen.1007401
1446 Navarro Llorens JM, Tormo A, Martínez-García E. 2010. Stationary phase in gram-negative
1447 bacteria. *FEMS Microbiol Rev* **34**:476–495. doi:10.1111/j.1574-6976.2010.00213.x
1448 Newbury SF, Smith NH, Higgins CF. 1987. Differential mRNA stability controls relative gene
1449 expression within a polycistronic operon. *Cell* **51**:1131–1143. doi:10.1016/0092-
1450 8674(87)90599-X
1451 Nitzan M, Rehani R, Margalit H. 2017. Integration of Bacterial Small RNAs in Regulatory
1452 Networks. *Annu Rev Biophys* **46**:131–148. doi:10.1146/annurev-biophys-070816-
1453 034058
1454 Nues R van, Schweikert G, Leau E de, Selega A, Langford A, Franklin R, Iosub I,
1455 Wadsworth P, Sanguinetti G, Granneman S, Van Nues R, Schweikert G, De Leau E,
1456 Selega A, Langford A, Franklin R, Iosub I, Wadsworth P, Sanguinetti G, Granneman S,
1457 Nues R van, Schweikert G, Leau E de, Selega A, Langford A, Franklin R, Iosub I,
1458 Wadsworth P, Sanguinetti G, Granneman S, Van Nues R, Schweikert G, De Leau E,
1459 Selega A, Langford A, Franklin R, Iosub I, Wadsworth P, Sanguinetti G, Granneman S.
1460 2017. Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles
1461 during stress. *Nat Commun* **8**:12. doi:10.1038/s41467-017-00025-5
1462 Papenfort K, Pfeiffer V, Lucchini S, Sonawane A, Hinton JCD, Vogel J. 2008. Systematic
1463 deletion of Salmonella small RNA genes identifies CyaR, a conserved CRP-dependent

1464 riboregulator of OmpX synthesis. *Mol Microbiol*. doi:10.1111/j.1365-2958.2008.06189.x
1465 Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JCD, Vogel J. 2006. σE-dependent small
1466 RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA
1467 decay. *Mol Microbiol* **62**:1674–1688.
1468 Papenfort K, Said N, Welsink T, Lucchini S, Hinton JCDD, Vogel J. 2009. Specific and
1469 pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small
1470 RNA. *Mol Microbiol* **74**:139–158. doi:10.1111/j.1365-2958.2009.06857.x
1471 Pletnev P, Osterman I, Sergiev P, Bogdanov A, Dontsova O. 2015. Survival guide:
1472 Escherichia coli in the stationary phase. *Acta Naturae* **7**:22–33.
1473 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
1474 features. *Bioinformatics* **26**:841–842.
1475 Rhodius VA, Suh WC, Nonaka G, West J, Gross CA. 2006. Conserved and variable
1476 functions of the σE stress response in related genomes. *PLoS Biol* **4**:0043–0059.
1477 doi:10.1371/journal.pbio.0040002
1478 Sætrom P, Sneve R, Kristiansen KI, Snøve O, Grünfeld T, Rognes T, Seeberg E. 2005.
1479 Predicting non-coding RNA genes in Escherichia coli with boosted genetic
1480 programming. *Nucleic Acids Res*. doi:10.1093/nar/gki644
1481 Sedlyarova N, Shamovsky I, Bharati BK, Epshtein V, Chen J, Gottesman S, Schroeder R,
1482 Nudler E. 2016. sRNA-Mediated Control of Transcription Termination in E.~coli. *Cell*
1483 **167**:111--121.e13.
1484 Sezonov G, Joseleau-Petit D, D'Ari R. 2007. Escherichia coli physiology in Luria-Bertani
1485 broth. *J Bacteriol* **189**:8746–8749. doi:10.1128/JB.01368-07
1486 Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. 2016. Global Mapping of Human
1487 RNA-RNA Interactions. *Mol Cell* **62**:618–626. doi:10.1016/j.molcel.2016.04.030
1488 Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, Margalit H. 2007.
1489 Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst*
1490 *Biol* **3**:138.
1491 Sittka A, Pfeiffer V, Tedin K, Vogel J. 2007. The RNA chaperone Hfq is essential for the
1492 virulence of Salmonella typhimurium. *Mol Microbiol*. doi:10.1111/j.1365-
1493 2958.2006.05489.x
1494 Smirnov A, Förstner KU, Holmqvist E, Otto A, Günster R, Becher D, Reinhardt R, Vogel J.
1495 2016. Grad-seq guides the discovery of ProQ as a major small RNA-binding protein.
1496 *Proc Natl Acad Sci* **113**:11591–11596. doi:10.1073/pnas.1609981113
1497 Smirnov A, Wang C, Drewry LL, Vogel J. 2017. Molecular mechanism of mRNA repression
1498 in *trans* by a ProQ-dependent small RNA. *EMBO J* **36**:1029–1045.
1499 doi:10.15252/embj.201696127
1500 St-Pierre F, Cui L, Priest DG, Endy D, Dodd IB, Shearwin KE. 2013. One-step cloning and
1501 chromosomal integration of DNA. *ACS Synth Biol*. doi:10.1021/sb400021j
1502 Storz G, Vogel J, Wassarman KM. 2011. Regulation by Small RNAs in Bacteria: Expanding
1503 Frontiers. *Mol Cell* **43**:880–891. doi:10.1016/j.molcel.2011.08.022
1504 Tebaldi T, Re A, Viero G, Pegoretti I, Passerini A, Blanzieri E, Quattrone A. 2012.
1505 Widespread uncoupling between transcriptome and translatome variations after a
1506 stimulus in mammalian cells. *BMC Genomics*. doi:10.1186/1471-2164-13-220
1507 Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, Nieselt K,
1508 Sharma CM, Storza G. 2015. Global transcriptional start site mapping using differential
1509 RNA sequencing reveals novel antisense RNAs in Escherichia coli. *J Bacteriol*.
1510 doi:10.1128/JB.02096-14
1511 Thompson KM, Rhodius VA, Gottesman S. 2007. SigmaE regulates and is regulated by a
1512 small RNA in Escherichia coli. *J Bacteriol* **189**:4243–4256.
1513 Tjaden B. 2015. De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome*
1514 *Biol* **16**:1–10. doi:10.1186/s13059-014-0572-2
1515 Travis AJ, Moody J, Helwak A, Tollervey D, Kudla G. 2013. Hyb: A bioinformatics pipeline
1516 for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data.
1517 *Methods*.
1518 Tree Jai J., Granneman S, McAteer SP, Tollervey D, Gally DL. 2014. Identification of

1519 bacteriophage-encoded anti-sRNAs in pathogenic Escherichia coli. *Mol Cell* **55**:199–
1520 213. doi:10.1016/j.molcel.2014.05.006
1521 Tree J J, Granneman S, McAteer SP, Tollervey D, Gally DL. 2014. Identification of
1522 bacteriophage-encoded anti-sRNAs in pathogenic Escherichia coli.
1523 doi:10.1016/j.molcel.2014.05.006
1524 Udekwu KI, Wagner EGH. 2007. Sigma E controls biogenesis of the antisense RNA MicA.
1525 *Nucleic Acids Res*. doi:10.1093/nar/gkl1154
1526 Updegrove TB, Zhang A, Storz G. 2016. Hfq: The flexible RNA matchmaker. *Curr Opin*
1527 *Microbiol* **30**:133–138. doi:10.1016/j.mib.2016.02.003
1528 Urban JH, Vogel J. 2007. Translational control and target recognition by Escherichia coli
1529 small RNAs in vivo. *Nucleic Acids Res* **35**:1018–1037. doi:10.1093/nar/gkl1040
1530 Vogel J, Luisi BF. 2011. Hfq and its constellation of RNA. *Nat Rev Microbiol* **9**:578–589.
1531 Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. 2001. Identification of novel
1532 small RNAs using comparative genomics and microarrays. *Genes Dev*.
1533 doi:10.1101/gad.901001
1534 Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2-A
1535 multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189–
1536 1191. doi:10.1093/bioinformatics/btp033
1537 Waters LS, Storz G. 2009. Regulatory RNAs in Bacteria. *Cell* **136**:615–628.
1538 doi:10.1016/j.cell.2009.01.043
1539 Waters SA, McAteer SP, Kudla G, Pang I, Deshpande NP, Amos TG, Leong KW, Wilkins
1540 MR, Strugnell R, Gally DL, Tollervey D, Tree JJ. 2016. Small RNA interactome of
1541 pathogenic E.~coli revealed through crosslinking of RNase E. *EMBO J* **36**:e201694639.
1542 doi:10.15252/embj.201694639
1543 Webb S, Hector RD, Kudla G, Granneman S. 2014. PAR-CLIP data indicate that Nrd1-
1544 Nab3-dependent transcription termination regulates expression of hundreds of protein
1545 coding genes in yeast. *Genome Biol* **15**:R8. doi:10.1186/gb-2014-15-1-r8
1546 Wilson KS, Hippel PH Von. 1995. Transcription termination at intrinsic terminators: The role
1547 of the RNA hairpin (Escherichia coli/RNA polymerase/rho-independent termination).
1548 *Biochemistry* **92**:8793–8797. doi:10.1073/pnas.92.19.8793
1549 Wright PR, Georg J, Mann M, Sorescu DA, Richter AS, Lott S, Kleinkauf R, Hess WR,
1550 Backofen R. 2014. CopraRNA and IntaRNA: Predicting small RNA targets, networks
1551 and interaction domains. *Nucleic Acids Res* **42**:119–123. doi:10.1093/nar/gku359
1552 Wright PR, Richter AS, Papenfort K, Mann M, Vogel J, Hess WR, Backofen R, Georg J.
1553 2013. Comparative genomics boosts target prediction for bacterial small RNAs. *Proc*
1554 *Natl Acad Sci* **110**:E3487–E3496. doi:10.1073/pnas.1303248110
1555 Wu X, Bartel DP. 2017. KpLogo: Positional k -mer analysis reveals hidden specificity in
1556 biological sequences. *Nucleic Acids Res*. doi:10.1093/nar/gkx323
1557
1558
1559

Supplementary Figures and legends

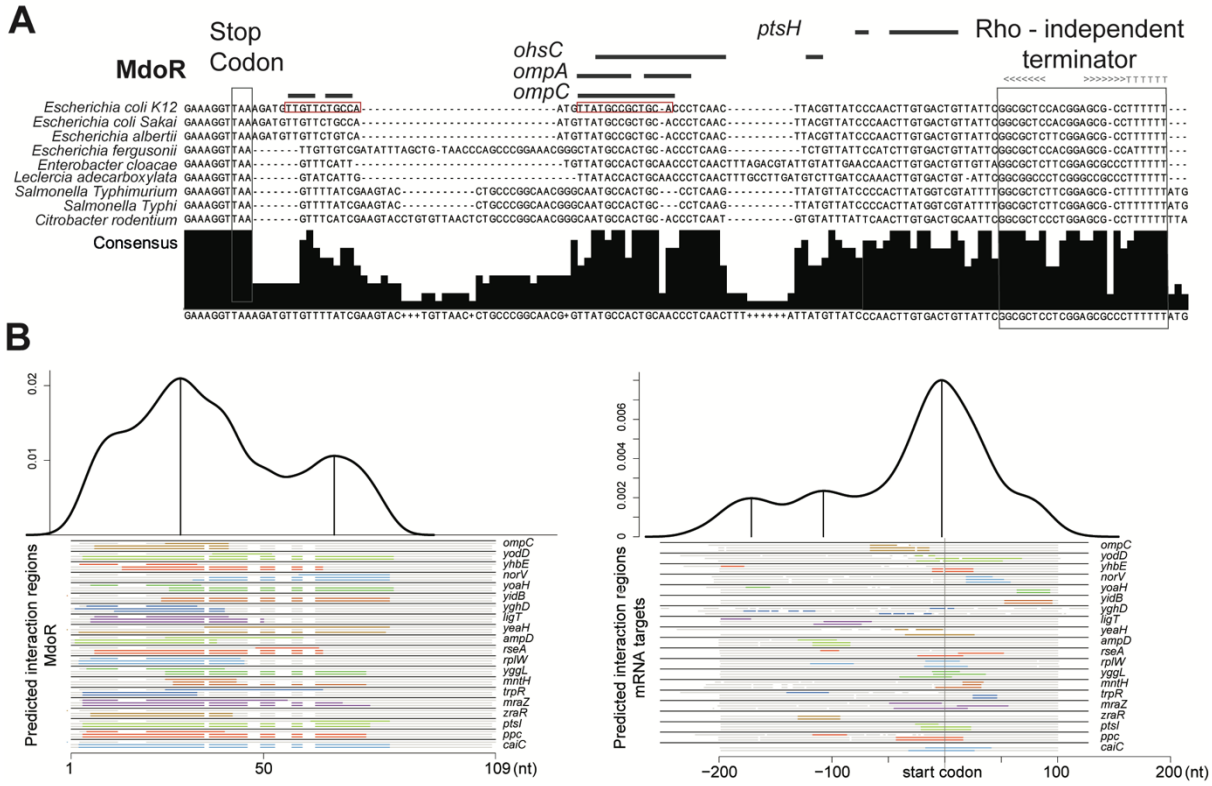


Figure 1–figure supplement 1. Hfq expression and Hfq binding to RNAs at different cell densities in UV-irradiated *E. coli*.
(**A**) Western blot analyses of Hfq levels during various growth stages. Hfq-HTF was detected using an anti-TAP primary antibody, and a fluorescent secondary antibody. GroEL was used as a loading control.
(**B**) Quantification of Hfq levels from the Western blot result. The fluorescent signal for Hfq-HTF and GroEL was measured with the LI-COR from biological replicate experiments. The levels of Hfq were normalised to GroEL and expressed as fold-change relative to OD_{600} 0.4.
(**C**) Hfq crosslinking to RNA is similar at each optical density. Autoradiogram showing the purified radioactively labelled Hfq-RNA complexes for each OD_{600} after elution from the nickel beads. Source data for (**A**-**C**) are provided as a Source Data file.

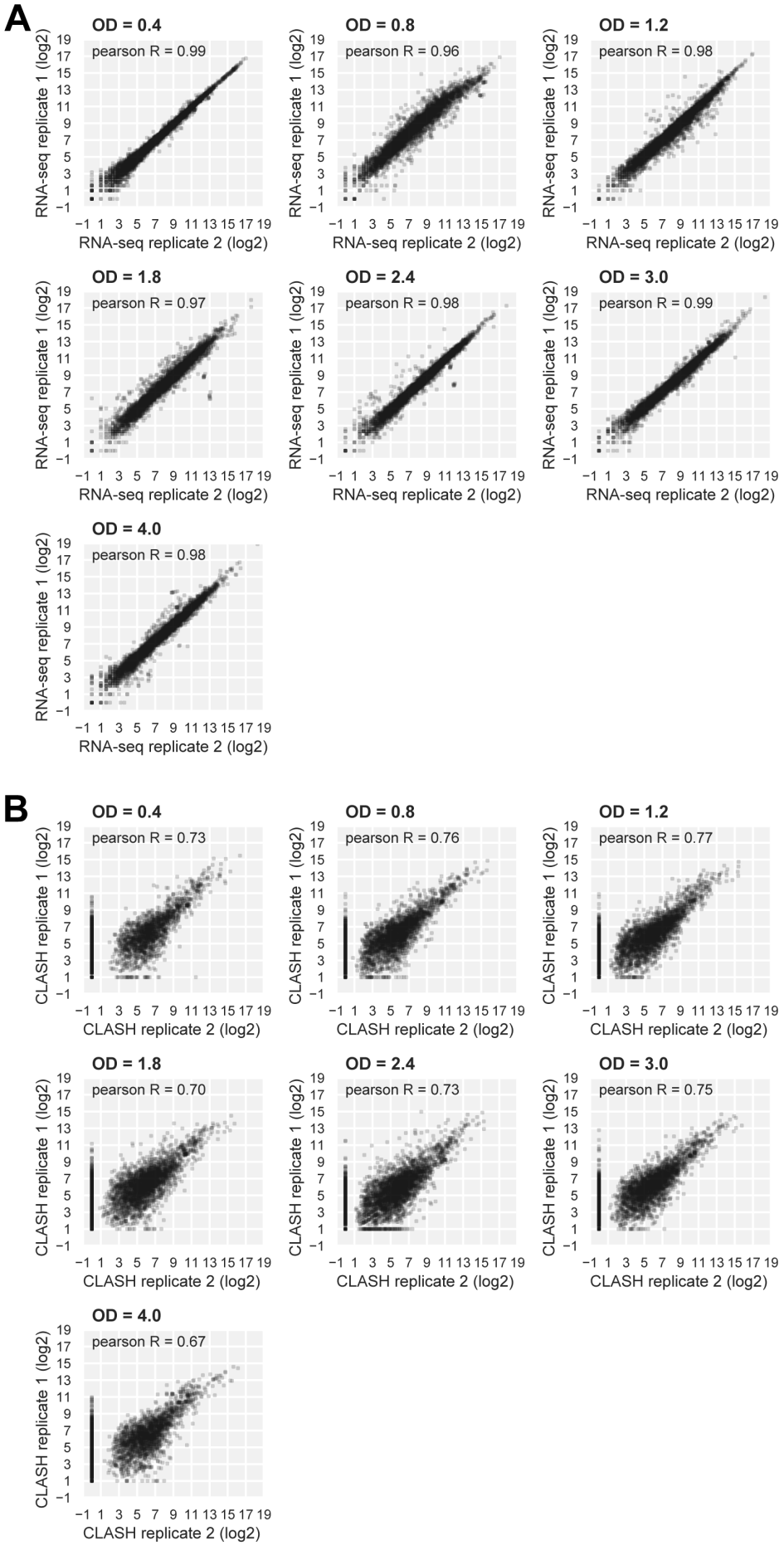

A
OD = 0.4
pearson R = 0.99
OD = 0.8
pearson R = 0.96
OD = 1.2
pearson R = 0.98
OD = 1.8
pearson R = 0.97
OD = 2.4
pearson R = 0.98
OD = 3.0
pearson R = 0.99
OD = 4.0
pearson R = 0.98
RNA-seq replicate 1 (log2)
RNA-seq replicate 2 (log2)
B
OD = 0.4
pearson R = 0.73
OD = 0.8
pearson R = 0.76
OD = 1.2
pearson R = 0.77
OD = 1.8
pearson R = 0.70
OD = 2.4
pearson R = 0.73
OD = 3.0
pearson R = 0.75
OD = 4.0
pearson R = 0.67
CLASH replicate 1 (log2)
CLASH replicate 2 (log2)

Figure 1–figure supplement 2. RNAseq and Hfq CLASH replicate datasets are highly correlated.
(**A, B**) Scatter plots showing the distribution of log_2 Transcripts Per Million (TPM) normalised read counts for Hfq CLASH (**A**) and RNA-seq (**B**) biological replicates. Pearson R coefficients describing the correlation between the two independent experiments at each OD_{600} are included.

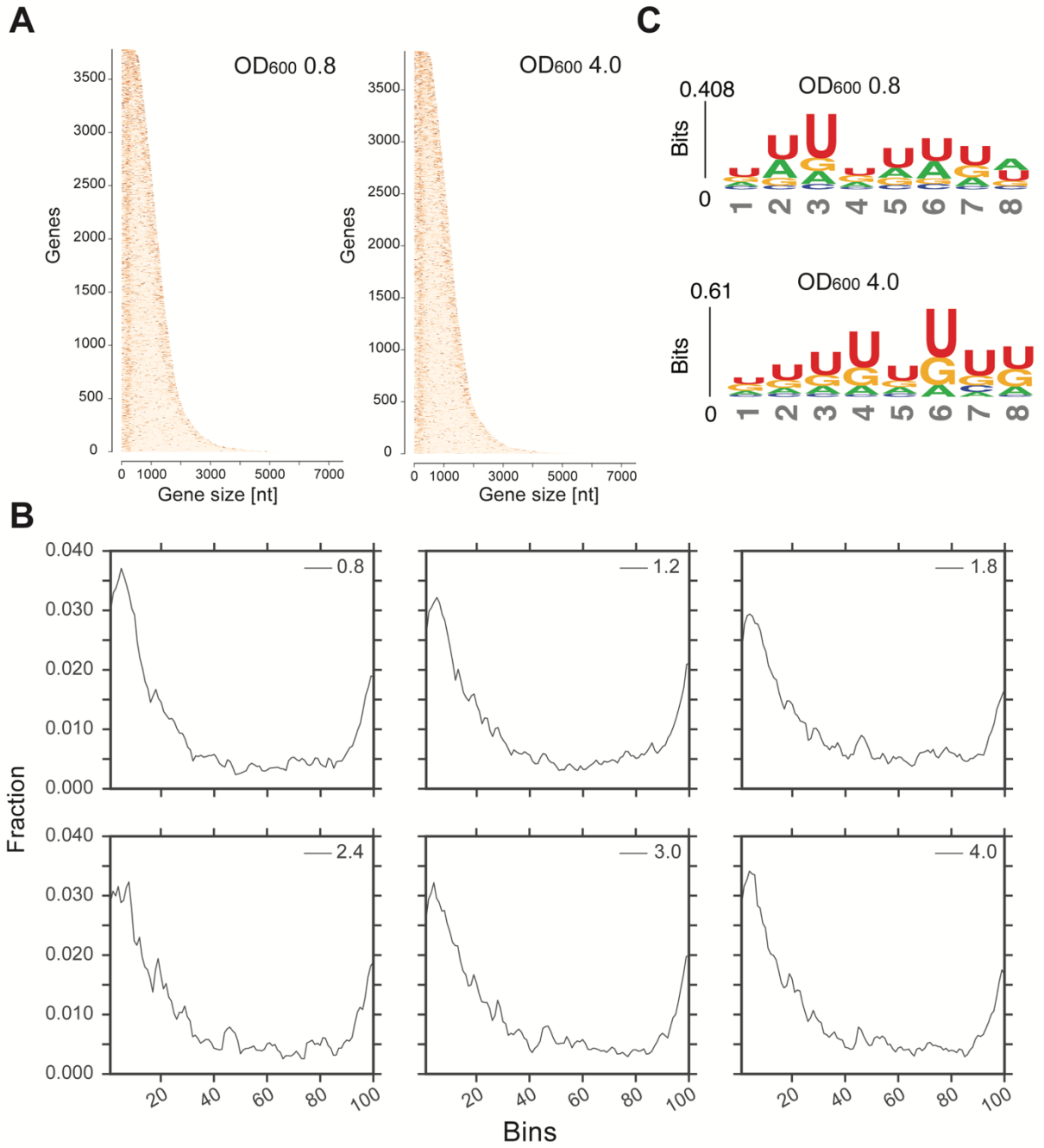


Figure 1–figure supplement 3. Transcriptome-wide maps of Hfq binding to mRNA genes.
(**A**) Heatmaps showing the distribution of Hfq binding sites across all mRNA genes at OD_{600} 0.8 and 4.0. The genes are sorted by their sequence length (x-axis); the darker a nucleotide is, the more Hfq is crosslinked to it. To generate the heatmap, Hfq binding clusters were generated. A 5'-and 3'UTR length of 200bp was used.

(**B**) Hfq binds to poly-U tracks. Significant k-mers (4-8 nt in length) were identified using the pyMotif tool of the pyCRAC package(Webb et al., 2014) and the motif logo was generated using all k-mers with a Z-score > 3, with kpLogo(Wu and Bartel, 2017). (**C**) A more stringent selection of the genes used to generate the distribution of Hfq binding to the transcriptome: all genes with overlapping 5' or 3'UTRs were removed from the analysis to avoid 'duplicate' counting. For all remaining cDNAs, FDR intervals of minimum 20 nt were considered for distribution plotting. The interval length (with UTR flanks as in the GTF annotation file) for each gene was normalized over 100 bins (x-axis), and the fraction of hits in each bin was calculated (y-axis).

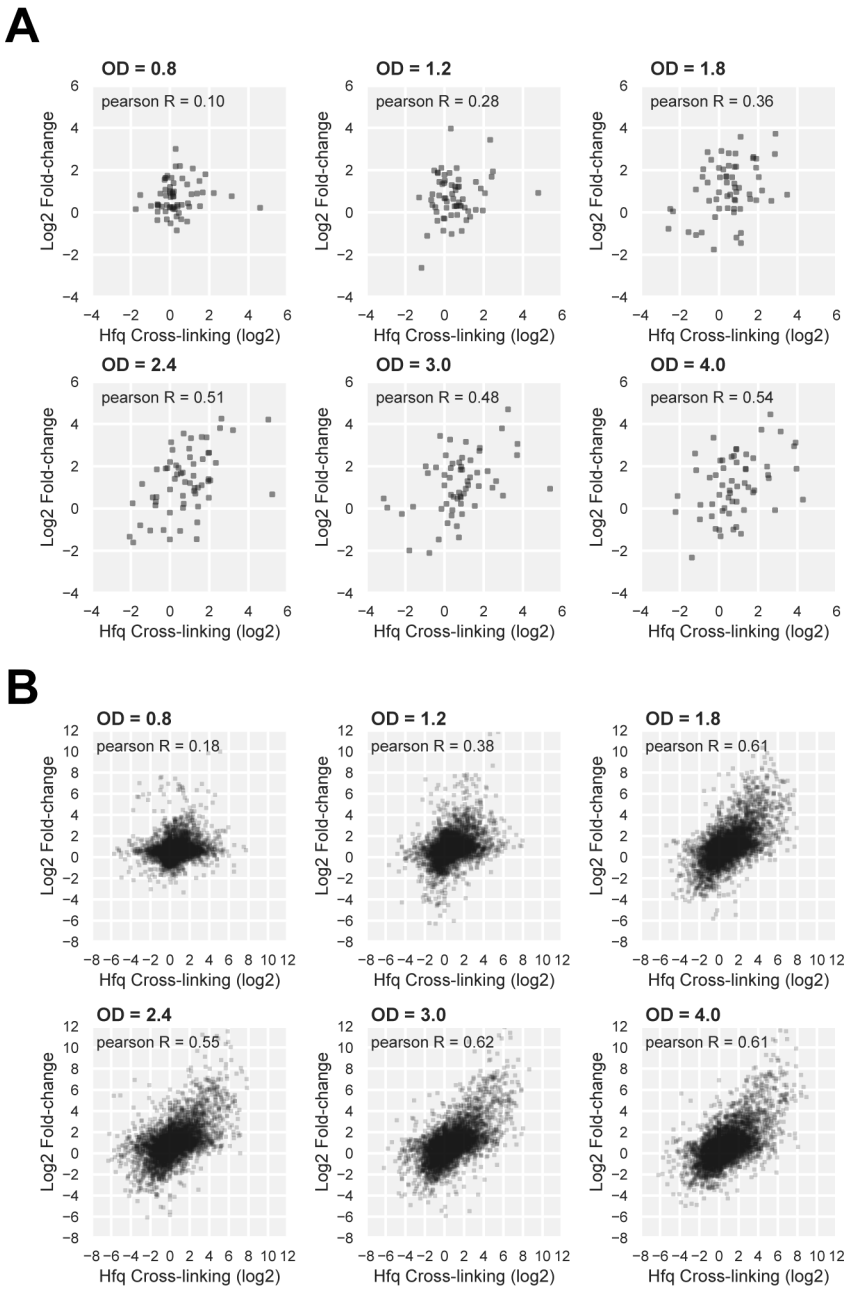


Figure 2–figure supplement 1. The sRNA and mRNA levels are not always strongly correlated with Hfq binding during the various stages of growth.
Scatter plots comparing changes in Hfq binding (x-axis) to RNA levels (y-axis) for the indicated OD_{600} for sRNAs (**A**) and all transcript classes (**B**) show that the correlation improves at higher OD_{600}. For each OD_{600}, the Transcripts Per Million (TPM) normalised read counts at each

density-point were divided by the OD_{600} 0.4 data, and the resulting ratios were log_2-normalized. Pearson R coefficients describing the correlation between the two independent experiments at each OD_{600} is also included for each plot.

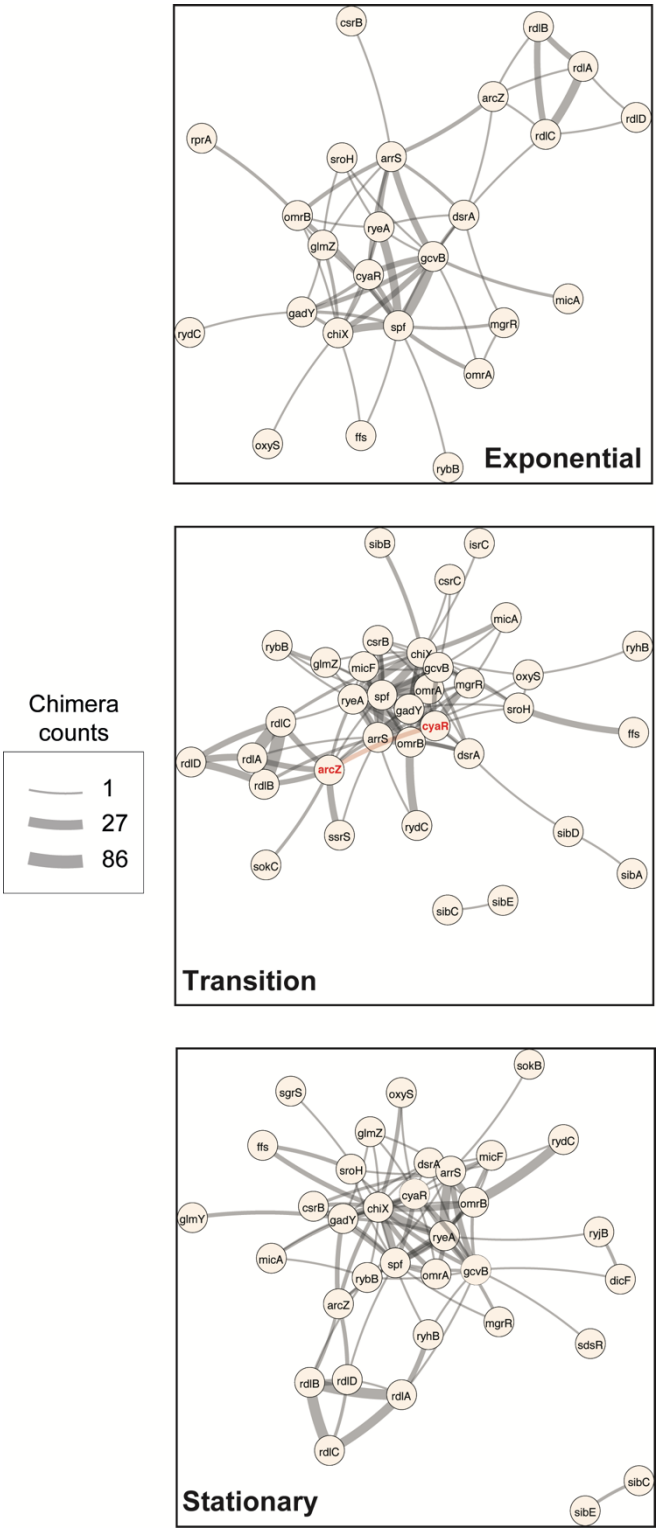


Figure 4–figure supplement 1. sRNA-RNA interactions identified by CLASH are growth-stage specific. sRNA-sRNA network generated from the statistically significant CLASH

interactions from two biological replicates, recovered at three main growth stages: exponential (OD_{600} 0.4 and 0.8), transition (OD_{600} 1.2, 1.8, 2.4) and early stationary (OD_{600} 3.0 and 4.0). The thickness of the edges is proportional to the $\log_2$(unique chimera count for each interaction). Only sRNAs transcribed from independent promoters were included in the analysis.

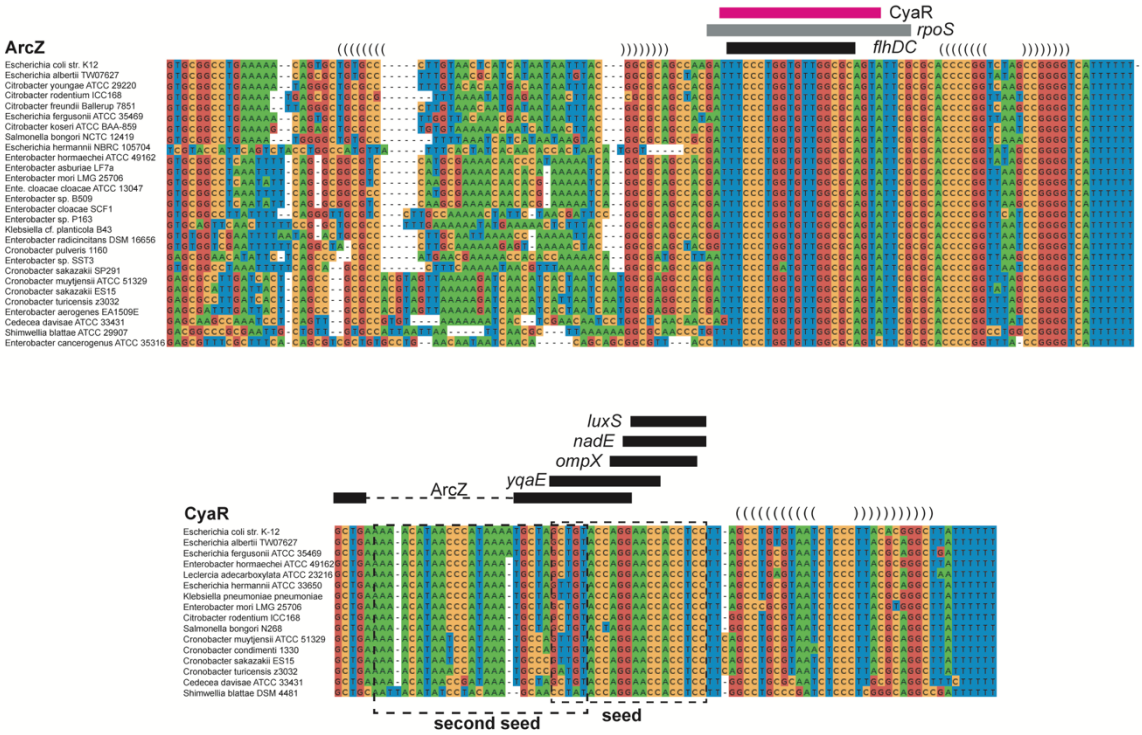


Figure 4–figure supplement 2. Interactions between ArcZ, CyaR and GcvB are conserved. Alignments of ArcZ, CyaR and GcvB were compiled as previously described(van Nues et al., 2016). Names of the enteric bacteria from which the sequence was retrieved are given on the left. Indicated are possible stem-loops (brackets), seed regions (boxed in dashed lines) and their interactions with various sections of ArcZ, CyaR or GcvB (blue and purple bars) or with other sRNAs and mRNAs (black bars). The CyaR sequence indicated with a blue bar is predicted to interact with two regions in GcvB (see blue bars in GcvB alignment), including the second seed sequence. A second interaction (pink bars) involves the seed sequence regions of CyaR and GcvB.

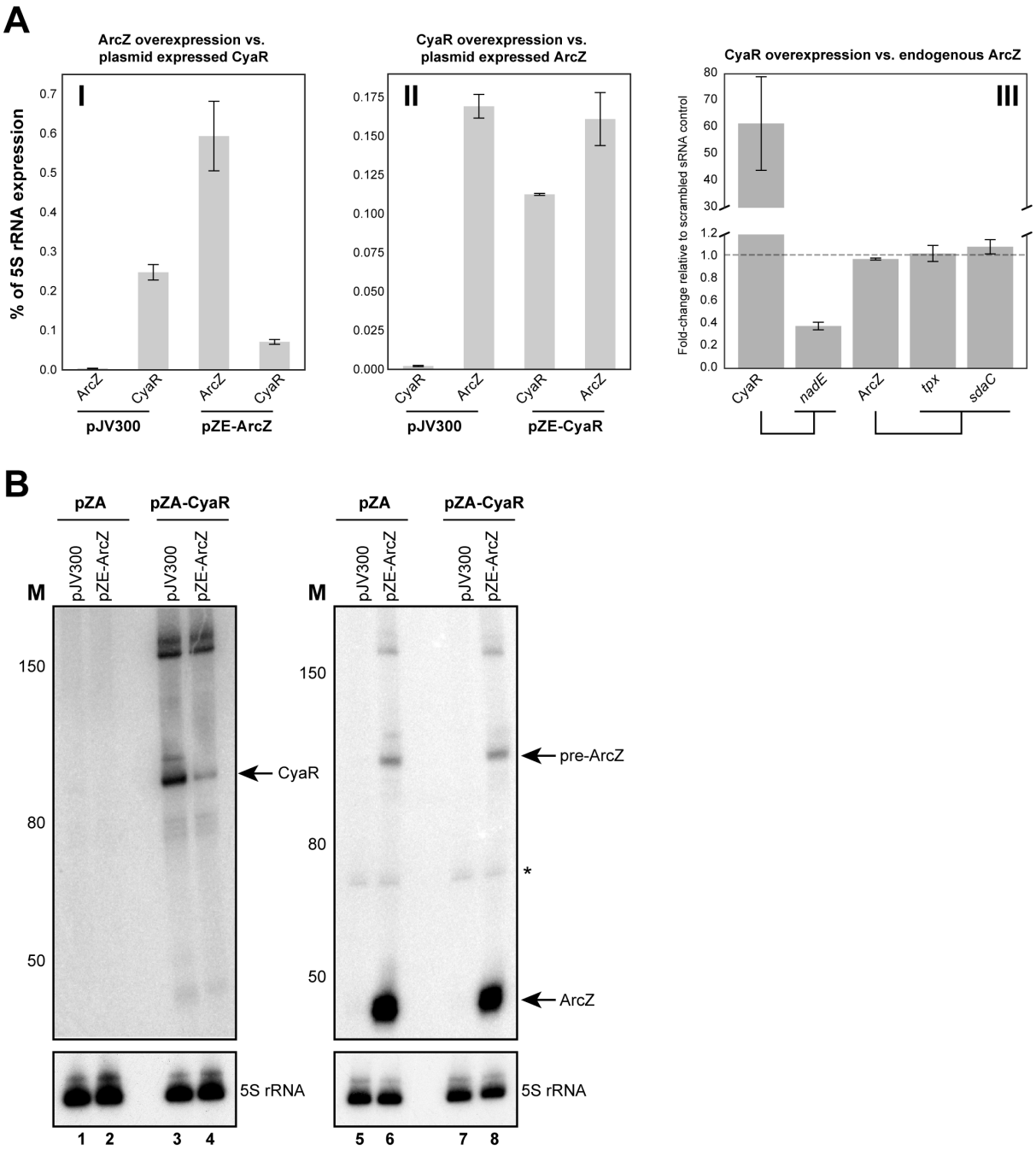


Figure 4–figure supplement 3. ArcZ downregulates CyaR expression.
(**A**) ArcZ and CyaR were overexpressed from a plasmid-borne IPTG inducible promoter (pZE-ArcZ and pZE-CyaR) and the data were compared to data from cells carrying a scrambled RNA plasmid (pJV300). The co-expressed candidate target sRNAs (expressed from pZA-derived backbone) were induced with anhydrotetracycline hydrochloride (panels I and II). The bars indicate the mean fold-change in expression relative to the level of 5S rRNA (*rrfD*) in cells with the indicated vector. In panel III endogenous ArcZ levels were measured upon over-expression of CyaR. Error bars indicate the standard error of the mean from three biological replicates and three technical replicates per experiment.
(**B**) Northern blot analysis of ArcZ and CyaR. The cells containing both the empty pZA and pJV300 plasmids (lanes 1, 5, 9) do not express ArcZ and CyaR at detectable levels.

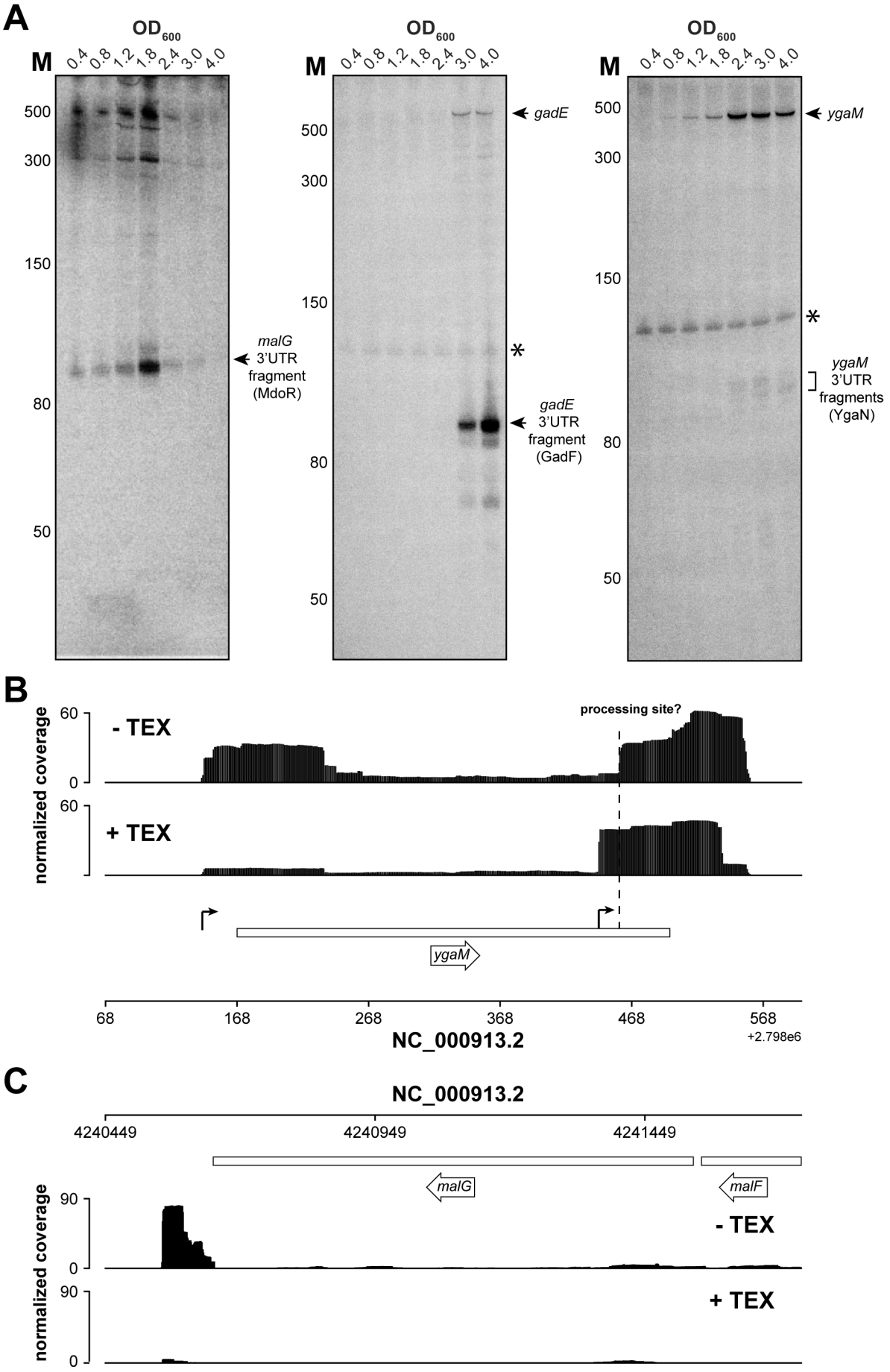

A
OD600
M
0.4
0.8
1.2
1.8
2.4
3.0
4.0
500
300
150
80
50
malG
3'UTR
fragment
(MdoR)
gadE
*
gadE
3'UTR
fragment
(GadF)
ygaM
ygaM
3'UTR
fragments
(YgaN)
B
normalized coverage
60
0
- TEX
+ TEX
processing site?
ygaM
68
168
268
368
468
568
+2.798e6
NC_000913.2
C
NC_000913.2
4240449
4240949
4241449
malG
malF
90
- TEX
+ TEX

Figure 5–figure supplement 1. *YgaM*, *gadE* and *malG* contain sRNAs in their 3'UTRs.
(**A**) Validation of *malG* 3'UTR (MdoR), *ygaM* 3'UTR (YgaN) and *gadE* 3'UTR (GadF) sRNAs by Northern blot. Total RNA extracted from cells at the indicated optical densities (OD_{600}) was resolved on 8% PAA-UREA gels and subjected to Northern blotting using oligos that hybridize with the 3'UTR of the respective transcripts. The asterisk indicates cross-reactivity of the probe with the 5S rRNA. The locations of the 3'UTR-derived fragments are indicated. MdoR and YgaN are ~110nt, whereas GadF fragment is ~ 90nt.
(**B, C**) Analysis of (Terminator 5'-Phosphate Dependent Exonuclease (TEX) RNA-seq datasets(Thomason et al., 2015) indicates that YgaN has an independent promoter, while MdoR is a degradation product of the *malEFG* operon. Genome browser tracks showing the location and normalised reads of *ygaM* and *malG* fragments in the absence of TEX (–TEX) and in the presence of TEX (+TEX). The *ygaM* and putative YgaN promoters are indicated. Independently transcribed YgaN could be further processed by RNases, at the site marked with a dashed vertical line.

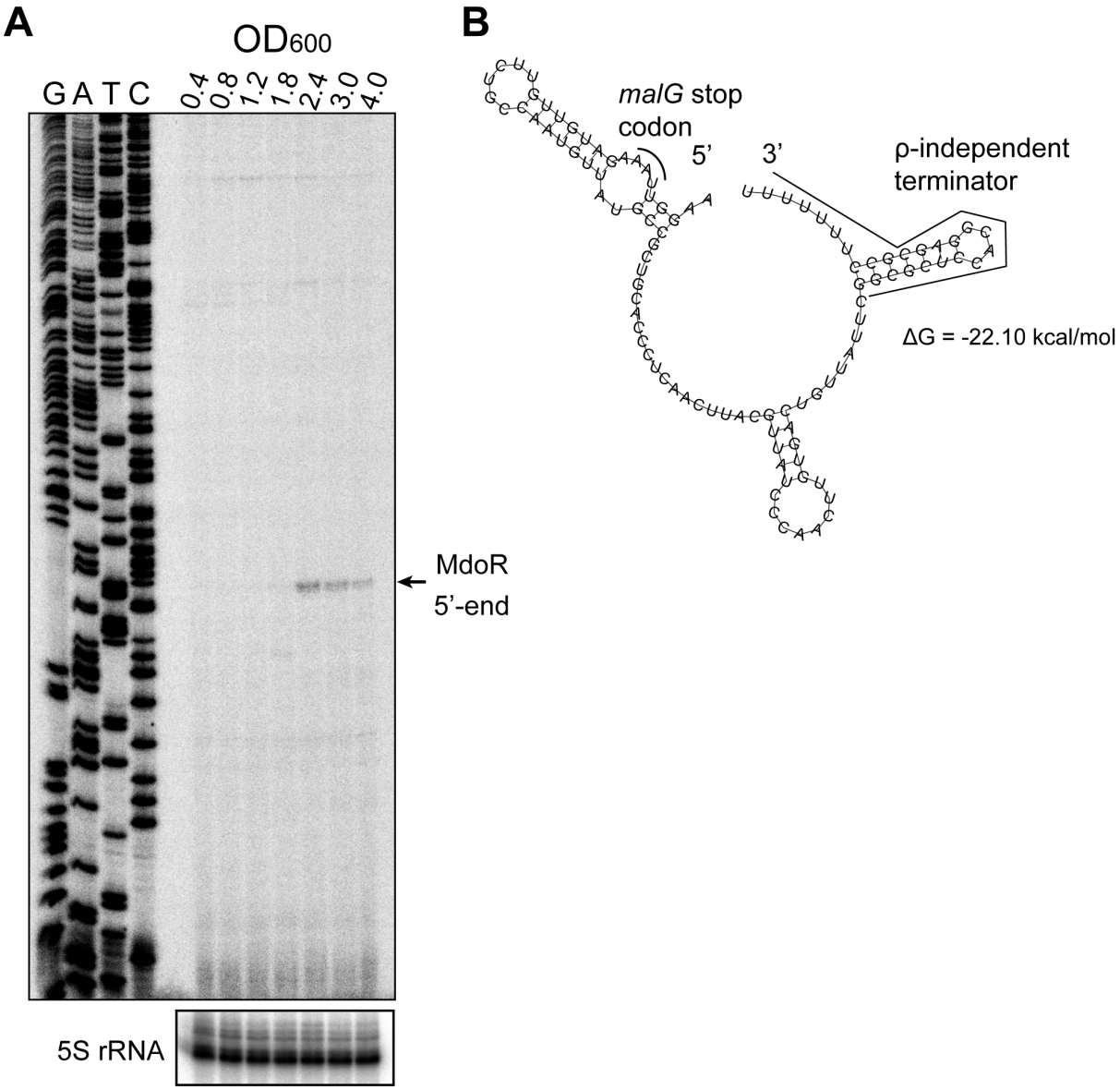


Figure 6–figure supplement 1. Conservation and target prediction analyses of MdoR.
(**A**) MdoR contains conserved and variable regions. Sequence conservation analysis of MdoR in several Gram-negative bacteria species (Mafft algorithm with defaults); the arrow indicates the 5'-end of MdoR; the *malG* stop codon and Rho-independent terminator sites are highlighted; the horizontal black lines indicate the base-pairing regions of MdoR with *ptsH* (top), OhsC, *ompA* and *ompC* (bottom) in *E. coli* as predicted by CLASH combined with *in silico* folding (RNACofold).
(**B**) MdoR is predicted to interact with its targets using two seed regions. Interaction regions within MdoR and top target mRNAs predicted by CopraRNA(Wright et al., 2014, 2013). Density plots showing the relative frequency of a specific MdoR (Left) or mRNA (Right) nucleotide position in all predicted sRNA-mRNA interactions with a p-value < 0.01 in all considered homologs. The vertical lines indicate local maxima; the aligned regions of the homologs are shown in grey, whereas the interacting regions are shown in arbitrary colors; only the top 20 representative clusters members are shown in the aligned regions, with the gene names indicated on the right.

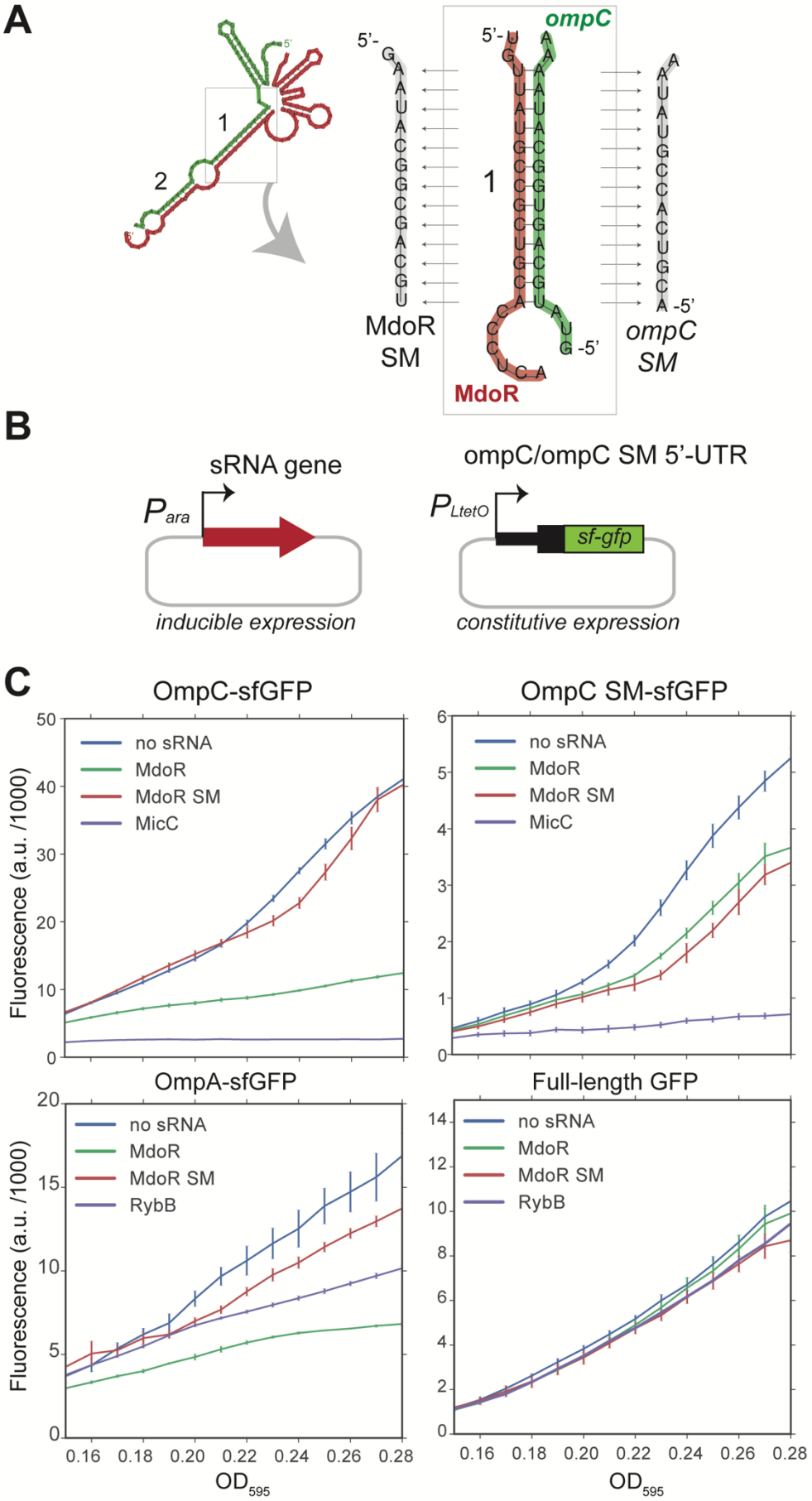

A
ompC
5'-
1
2
MdoR
SM
ompC
SM
MdoR
B
sRNA gene
ompC/ompC SM 5'-UTR
P ara
P LtetO
sf-gfp
inducible expression
constitutive expression
C
OmpC-sfGFP
OmpC SM-sfGFP
OmpA-sfGFP
Full-length GFP
no sRNA
MdoR
MdoR SM
MicC
RybB
Fluorescence (a.u. /1000)
OD595

Figure 8–figure supplement 1. Validation of MdoR-*ompC* interaction using GFP reporters.
(**A**) Design of the wild-type and mutant *ompC* constructs. The panel indicates the base-pairing region within the MdoR-*ompC* duplex that was mutated. We created an MdoR seed mutant (SM) and an *ompC* mutant in which base pairing with MdoR SM was restored.
(**B**) Plasmid system used for the reporter assay: in *E. coli* TOP10 cells, low-copy plasmids constitutively overexpress target 5'UTRs fused to sfGFP and medium-copy plasmids overexpress the full-length sRNAs upon induction with L-arabinose.
(**C**) MdoR downregulates expression of OmpC and OmpA sfGFP fusions. *In vivo* fluorescence measurements of OmpC, OmpC SM and OmpA sfGFP fusion proteins was measured using a Tecan plate reader system. As a negative control we included sfGFP alone in the presence or absence of sRNAs. The 'no sRNA' expressing strains contain the empty pBAD plasmid. The y-axis indicates fluorescence units (F.U.) reported by the plate reader. Experiments were performed in technical and biological triplicates; the fluorescence means and SEM of three biological replicates are reported. Source data for are provided as a Source Data file.

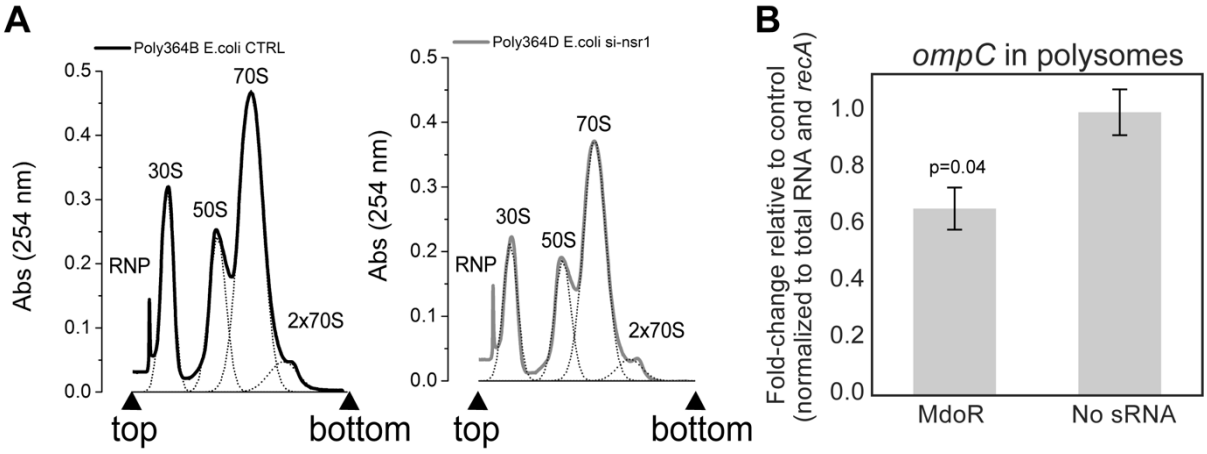


Figure 8–figure supplement 2. MdoR regulates *ompC* mRNA translation in *E. coli*.
(**A**) Cultures at OD_{600} 0.4 overexpressing MdoR or no sRNA for 15 minutes were subjected to polysome profiling. Profiles of the polysomal (2x70S) and subpolysomal fractions obtained for the empty plasmid control and MdoR overexpression samples.
(**B**) RT-qPCR analysis of the polysomal fractions: A 'total' fraction was obtained by mixing equal amounts/volumes of the polysomal and subpolysomal fractions and is representative of the cytosol/cell lysate content. Total RNA was extracted from all fractions (polysomal, subpolysomal and total). Expression of *ompC* in the polysomal fractions was quantified relative to the amount in 'total' fraction, normalized to *recA,* and calculated as fold-change relative to the control sample (y-axis). The experiments were performed in technical triplicates; the standard error of the mean (SEM) of three biological replicates fold changes are reported as error bars. Significance of the difference in *ompC* mRNA level in polysomes was assessed with a two-tailed Student's t-test. Source data for are provided as a Source Data file.

Supplementary References

Thomason MK, Bischler T, Eisenbart SK, Förstner KU, Zhang A, Herbig A, Nieselt K, Sharma CM, Storza G. 2015. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. *J Bacteriol*. doi:10.1128/JB.02096-14

van Nues RW, Castro-Roa D, Yuzenkova Y, Zenkin N. 2016. Ribonucleoprotein particles of bacterial small non-coding RNA IsrA (IS61 or McaS) and its interaction with RNA polymerase core may link transcription to mRNA fate. *Nucleic Acids Res* **44**:2577–2592.

Webb S, Hector RD, Kudla G, Granneman S. 2014. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol* **15**:R8. doi:10.1186/gb-2014-15-1-r8

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**:119–123. doi:10.1093/nar/gku359

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* **110**:E3487–E3496. doi:10.1073/pnas.1303248110

Wu X, Bartel DP. 2017. KpLogo: Positional k -mer analysis reveals hidden specificity in biological sequences. *Nucleic Acids Res*. doi:10.1093/nar/gkx323