Hfq CLASH uncovers sRNA-target interaction networks enhancing 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

A reason microorganisms are so successful is their ability to rapidly adapt to constantly changing environments. Bacterial small RNAs (sRNAs) play an important role in adaptive responses by shaping gene expression profiles and integrating multiple regulatory pathways. This enables microorganisms to more efficiently counteract environmental insults than could be achieved by simply switching transcription factors on and off. Although a plethora of sRNAs have been identified, the majority have not been functionally characterized and their relative contribution in regulating adaptive responses remains unclear. 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 interaction networks. Using this proximity-dependent RNA-RNA ligation method, we uncovered thousands of sRNA-target duplexes associated with the RNA chaperone Hfq at specific growth stages, many of which have not been described before. Our work revealed that 3'UTR-derived sRNAs and sRNA-sRNA interactions are more prevalent than previously anticipated. We uncovered sRNA-target interaction networks that play a role in adaptation to changes in nutrient availability by enhancing the uptake of nutrients from the environment. We describe detailed functional analyses of a novel sRNA (MdoR), the first example of a bacterial 3'UTR-derived sRNA that functions as part of a mixed coherent feed forward loop. MdoR enhances the effectiveness of maltose uptake by (a) inactivating repressive pathways that blocks the accumulation of specific maltose transporters and (b) by reducing the flux of general porins to the outer membrane. Our work suggests that many mRNAs encode regulatory sRNAs embedded within their 3'UTRs, allowing direct regulatory interactions between functionally related mRNAs. Small RNA sponging interactions appear frequently and reveal that major nutritional stress responses are coordinated post-transcriptionally. This provides striking examples of how cells utilize sRNA regulatory networks to integrate multiple signals and 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. Complex responses are mediated through integrated transcriptional and post-transcriptional networks. Control at the transcriptional level dictate which genes are expressed[1,2] and is well-characterised in *Escherichia coli*. Post-transcriptional regulation has not been extensively studied, but is key for regulating 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[3]), *cis*-acting RNAs[4], and RNA binding proteins (RBPs), such as ribonucleases[5], RNA chaperones and RNA helicases. The sRNAs are the largest class of bacterial regulators, which often work in tandem with RBPs to regulate their RNA targets[3,6]. 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[7–9]. Small RNAs can repress or stimulate translation and transcription, as well as control mRNA stability[3,7,10,11]. Although sRNAs are important components of regulatory networks[12], relatively little is known about how sRNAs help cells to efficiently respond to environmental challenges.

To understand sRNA-mediated adaptive responses, detailed knowledge of the underlying circuits is required. In *E. coli*, hundreds of sRNAs have been discovered, but only a limited number have been thoroughly characterised. A key step to unravel the roles of sRNAs in regulating gene expression is to identify their targets. To tackle this globally, high-throughput methods have been developed that have uncovered a plethora of sRNA-target interactions, many more than previously anticipated[13–18].

During growth in rich media, *E. coli* are exposed to continuously changing conditions, such as carbon-source shifts, and 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 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[19,20]. Although the exponential and stationary phases have been characterised in detail[21,22], little is known about the transition between these two phases, which is a very dynamic stage of the cell population cycle, involving many concurrent changes that span multiple metabolic pathways. There are deep physiological differences between exponentially-growing cells and cells that have reached stationary phase, yet a complete view of how the timings and dynamics are achieved,

and how gene expression is modulated during these changes remains elusive. During these stages the cell population switches between alternative carbon sources as their preferred sources are exhausted from the medium. To adapt, they need to rapidly rewire their transcriptome to be able to scavenge these alternative carbon sources[23–25]. For instance, as substrate availability varies, the expression of specific nutrient transporters and metabolic enzymes is regulated accordingly. The role of sRNAs in regulating this transition is even less clear. Most sRNAs are expressed and associate with Hfq in a growth-dependent manner, with the vast majority being primarily expressed during stationary phase, and only a few being bound by Hfq at the transition stage[26]. However, the function of these sRNAs in regulating the transition phase is still unclear.

Here we applied UV cross-linking, ligation and sequencing of hybrids (CLASH)[27,28] on *E. coli* Hfq to uncover sRNA-target RNA interaction dynamics that take place during the entry from exponential to stationary phase. We demonstrate that the highly stringent purification steps make Hfq CLASH a robust method for direct mapping of Hfq-mediated sRNA-target interactions in *E. coli*. We highlight the dynamics of these interactions during the transition from exponential to stationary phases of growth in Lysogeny broth (LB). Our data revealed Hfq binding to 3'UTRs of hundreds of mRNA transcripts, suggesting that 3'UTR-derived sRNAs[26,29–31] might be more prevalent than previously anticipated. Moreover, our analyses predicted a surprisingly large number of sRNA-sRNA interactions, several of which we experimentally validated. Our data suggests that CyaR, an sRNA that is primarily expressed during the transition from late exponential to stationary phase, plays an important role in enhancing nutrient uptake from the environment by antagonizing the sRNA GcvB, which targets mRNAs involved in amino acid and peptide uptake. This emphasizes the notion that sRNA-sRNA interactions play an important role in environmental adaptation. Our results expand on sRNA-target interactions uncovered by RNase E CLASH[13] and RIL-seq[18], and we show that Hfq CLASH can generate very robust results. We also characterized a novel 3'UTR-derived sRNA, which we refer to as MdoR (mal-dependent OMP repressor). Unlike the majority of bacterial sRNAs, MdoR is very transiently expressed during the progression from exponential to stationary phase. We demonstrate that MdoR is a degradation intermediate of the *malG* 3'UTR, the last transcript of the *malEFG* polycistron that encodes components of the maltose transport system. We show that MdoR directly downregulates a select number of mRNAs, including several encoding major porins. Analysis of the chimeric reads predicted that MdoR forms unusually long base-pairing interactions with many of its targets. We propose that MdoR is part of a regulatory network that, during the transition from exponential to stationary phase, stimulates accumulation of high affinity maltose transporters in the outer membrane by inactivating opposing pathways.

Results

Hfq CLASH in *E. coli*

To unravel the post-transcriptional networks that underlie the transition between exponential and stationary growth phases in *E. coli*, we performed CLASH[27,28] (Fig. 1a) on Hfq to uncover growth-stage dependent sRNA regulatory networks. *E. coli* cells expressing an HTF (His6-TEV-3xFLAG)-tagged Hfq were grown in LB and an equal number of cells were harvested at different optical densities (OD_{600}). Cell samples from seven different optical densities were subsequently subjected to Hfq CLASH (Fig. 1b). To generate high quality Hfq CLASH data we had to make a number of improvements to the original protocol used for RNAse E CLASH[13]. The biggest problem that needed to be resolved was the seemingly rapid degradation of RNA during the cell lysis and first immunoprecipitation steps, which resulted in the recovery of relatively short cross-linked RNA fragments and poor recovery of chimeric reads. To minimize this degradation, we shortened the incubation time with anti-FLAG beads to 1-2 hours. Crucially, actively growing cells were UV irradiated for only seconds in the Vari-X-linker[32] and harvested by rapid filtration to increase the recovery of short-lived protein-RNA and RNA-RNA interactions and to minimize changes in the transcriptome induced by DNA damage. Finally, to improve the detection of chimeras, all libraries were paired-end sequenced. As a negative control, CLASH was performed on an untagged strain. The control samples had ~10 times lower number of reads, demonstrating the specificity of the approach. To complement the CLASH data and for normalization purposes, RNAseq data was generated from cells harvested at the same optical densities.

Based on the growth curve shown in Fig. 1b, we categorized OD_{600} 0.4 and 0.8 as exponential growth phase: 1.2, 1.8, 2.4 as the transition phase from exponential to stationary: 3.0 and 4.0 as early stationary phase. Strikingly, we observed that in LB medium Hfq levels in stationary phase were up to 18-fold higher compared to early log phase (Fig. 1c). To determine the cross-linking efficiency, Hfq-RNA complexes immobilized on nickel beads were radiolabelled, resolved on NuPAGE gels and detected by autoradiography. The data show that the recovery of Hfq and radioactive signal was comparable at each OD_{600} (Fig 1c), suggesting that the beads were saturated with Hfq during the purification.

Hfq binds to the transcriptome in a growth-stage dependent manner

Meta-analyses of the Hfq CLASH sequencing data revealed that the distribution of Hfq binding across mRNAs was very similar at each growth stage. We observed the expected Hfq enrichment at the 5'UTRs and at the 3'UTRs at each growth stage (see Supplementary Fig.1 for examples). After identifying significantly enriched Hfq binding peaks (FDR <= 0.05) we used the genomic coordinates of these peaks to search for Hfq binding motifs in mRNAs. The

most enriched k-mer included poly-U stretches (Supplementary Fig.1b), that resemble the poly-U tracts characteristic to Rho-independent terminators found at the end of many bacterial transcripts[33], and confirms the motif uncovered in CLIP-seq studies in *Salmonella*[34].

Given the established role of Hfq in sRNA stabilization and mediating sRNA-target interactions, it was logical to assume that changes in Hfq binding and availability directly affect sRNA steady-states. This would imply that the Hfq binding data should show a strong correlation with the RNA-seq data. To test this, we compared the Hfq cross-linking data to the RNA-seq data. K-means clustering of the TPM-normalized read count data revealed 7-9 different patterns of changes in read counts in the Hfq cross-linking and total RNAseq data (Fig. 2a). Reminiscent of recent work performed in *Salmonella*[26], most sRNAs in *E. coli* appear to be preferentially expressed when the cells reach stationary phase (Fig. 2a). However, much to our surprise, the Hfq cross-linking profile did not always follow the same trend (Fig. 2a). The scatterplot analyses highlight that, globally, changes in sRNA expression are indeed not always recapitulated in their respective Hfq-binding profile (and *vice versa*) (Fig 2b). The correlation between changes in sRNA expression levels versus changes in Hfq cross-linking was poor at lower cell density (OD_{600} 0.8: r = 0.27) but gradually improved as the cells approach stationary phase (OD_{600} 3.0 r =0.6; Fig. 2b). A very similar result was obtained when doing the same comparison for all Hfq-bound RNAs, including mRNAs (Supplementary Fig.2). Notable examples are ChiX and GcvB. In the case of ChiX, Hfq binding only shows modest changes during the growth phases (Fig 2A; right heat map, cluster 8), whereas sRNA levels steadily increased (Fig. 2a; left heat map; cluster 3). In the case of GcvB, Hfq cross-linking decreased slightly (Fig. 2a; right heat map; cluster 8); however, the sRNA expression levels increased slightly at higher optical densities (Fig. 2a; left heat map; cluster 6). Some sRNAs, such as ArcZ and GlmZ, are close to the trend line in all growth conditions, suggesting Hfq plays an important role in their stabilization (Table S1). However, particularly at low optical densities, many sRNAs deviated from the trend. A striking example is ArrS, an RNA that until recently[18] was thought to act only as a *cis*-encoded sRNA. ArrS is very highly expressed throughout growth (especially during stationary phase) relative to OD_{600} 0.4, but the increase in Hfq cross-linking is comparatively low.

A plausible explanation for the observed results is that Hfq levels were significantly lower during exponential phase (Fig. 1c) and that during this growth phase there may not be sufficient Hfq to bind all sRNAs. We conclude that the dynamics of sRNA expression and binding to Hfq are not always well correlated and that this may be linked to the availability of Hfq in the cell (also see discussion).

Hfq CLASH robustly detects RNA-RNA interactions

To get a complete overview of the RNA-RNA interactions captured by Hfq CLASH, we

merged the data from two biological replicates of CLASH growth phase experiments (Table S2). Overlapping paired-end reads were merged and analyzed using the hyb pipeline[35]. To select RNA-RNA interactions for further analysis, we took into consideration whether the chimera was found in independent experiments, the number of *unique* chimera counts (after collapsing the data), the minimum folding energy, and occurrence of the chimeras in both orientations (gene_A - gene_B and gene_B - gene_A). The distribution of combinations of transcript classes found in all discovered unique chimeric reads indicates sRNA-mRNA interactions as the most frequent Hfq-mediated interaction type (~35%) (Fig. 3a). We suspect that this number might even be higher, as more than a quarter of the chimeras contained sRNA and fragments from intergenic regions. Manual inspection of several of these indicated that a number of the intergenic sequences were located near genes for which the UTRs were unannotated or short.

Analysis of the unique interactions uncovered at different growth stages reveals that the majority are growth stage-specific. A relatively small number of interactions were common to all growth conditions, and the clear majority are occurring at the transition between exponential and stationary phase and at the entry into stationary phase (Fig. 3b). Meta-analysis of the mRNA fragments of the chimeras revealed that the majority of mRNA fragments in the chimeras were strongly enriched in 5'UTRs peaking near the translational start codon (Fig. 3c), consistent with the canonical mode of translational inhibition by sRNAs[36] and demonstrating the robustness of the data. Some enrichment was also found in 3'UTRs of mRNAs (Fig. 3c). Motif analyses revealed a distinct sequence preference in 5'UTR and 3'UTR binding sites (Fig. 3d, Table S3). The 3'UTR-containing chimera consensus motif corresponds to poly-U transcription termination sites. The A-rich 5'UTR motif is more consistent with Hfq binding to Shine Dalgarno-like $(ARN)_n$ sequences[37]. Two examples of experimentally validated sRNA-target interactions are detailed in Fig. 4. ChiX is an abundant sRNA that down-regulates *chiP* expression by occluding the Shine-Dalgarno (SD) sequence[38,39]. In the absence of chitosugars, ChiX also represses *chbC* by base-pairing to an intercistronic region between *chbB* and *chbC*[39–41]. We investigated in detail the chimeras in our data that support the known ChiX-*chiP* and ChiX-*chbC* interactions (Fig. 4a,b). This example also illustrates that even without any statistical filtering of the reads, the Hfq CLASH data has remarkably low background, as shown by the sharp peaks at the known target sites and low background signals elsewhere on the coding sequences. Furthermore, *in silico* RNA secondary structure analysis of the chimeric sequences was sufficient to precisely map the base-pairing interaction between the sRNA and its target (as shown for ChiX and *chiP* in Fig. 4b), and confirm the experimentally-validated interacting regions/sequences. The sRNAs ArcZ and RprA showed the highest number of different interactions and the vast majority overlapped the known seed sequence[42,43], including a large number with relatively low read counts (2-5; Fig. 4c). This

suggests that even low abundance chimeras could represent *bona fide* interactions. Consistent with this idea, comparing our data to RIL-seq (S-chimeras)[18] and validated interactions found in the sRNATarbase 3 database[44] showed that ~50% of the interactions found in all three datasets had low read counts in the Hfq CLASH data (Fig. 4d). Therefore, unique chimeras that were found at least twice in the data were also considered for further analyses. A complete overview of chimeras with a minimum of two unique cDNA counts are provided in Table S2.

Hfq CLASH predicts sRNA-sRNA sponging as a widespread layer of post-transcriptional regulation

We also uncovered a large pool (~160) of sRNA-sRNA interactions, 95 of which were supported by at least two chimeric reads (Table S4). The sRNA-sRNA network is dominated by several abundant sRNAs that appear to act as hubs – sRNAs with many different interacting partners: ChiX, Spot42 (spf), ArcZ and GcvB (Table S4). The vast majority of the interactions were growth stage specific and the sRNA-sRNA networks show extensive rewiring across the exponential, transition and stationary phases (Fig. 5a). In many cases the experimentally-validated sRNA seed sequences were found in the chimeric reads, for both established and novel interactions. For example, the vast majority of ArcZ sRNA-sRNA chimeras contained the known and well conserved seed sequence (Fig. 5b, Supplementary Fig.3). The exception was the ArcZ-ssrS (6S RNA) interaction. Although we found many unique chimeras supporting this interaction, it involves a region upstream of the ArcZ seed sequence, which is not very well conserved (Supplementary Fig.3). Therefore, we did not study this interaction in more detail.

The sRNA-sRNA chimeras containing CyaR fragments were of particular interest, as the sRNA is primarily expressed during the transition from late exponential to stationary phase (Fig. 2a;[45]). In the case of CyaR, the known seed sequence as well as a conserved ~25 nt fragment in the 5' region was found in chimeras (Fig. 5b; Supplementary Fig.3). Similar seed sequences were identified in CLASH experiments using RNase E as a bait[13], suggesting that this region represents a *bona fide* interaction site. Notably, we identified many ArcZ-CyaR chimeras containing the seed sequence from both sRNAs, suggesting that these sRNAs could inhibit each other's activity. To validate these findings, we co-expressed the two sRNAs in *E. coli* and monitored their expression as well as expression of their target mRNAs[37]. The data were subsequently normalized to the results obtained with a scrambled sRNA to calculate fold changes in expression levels. Since it is difficult to predict directly from the CLASH data which sRNA in each pair acts as the decoy/sponge, we tested both directions. ArcZ overexpression not only significantly decreased the expression of its own mRNA targets (*tpx*, *sdaC*) but also CyaR (Fig. 5d, panel I). Concomitantly, we observed a substantial increase in CyaR targets

nadE and *yqaE* (Fig. 5d, panel I). Over-expression of CyaR reduced the level of a direct mRNA target (*nadE*) but it did not significantly alter the level of ArcZ or ArcZ mRNA targets (*tpx* and *sdaC*; Fig. 5d, panel II). This suggests that the regulation is unidirectional. It is, however, important to note that although in this experiment the expression of CyaR increased ~45-fold, it still did not reach the levels of ArcZ (Supplementary Fig.4). Therefore, it is possible that under the tested conditions the CyaR over-expression was not sufficient to see a sponging effect on ArcZ. Nevertheless, we conclude that ArcZ acts as a CyaR anti-sRNA and can trigger its degradation. We next analysed the CyaR-GcvB sRNA-sRNA interaction that was detected in our Hfq CLASH data. Overexpression of CyaR not only significantly reduced expression levels of its targets (*nadE* and *yqaE*) but also that of GcvB and a concomitant increase in the levels of GcvB target *dppA* (Fig. 5d, panel III). Conversely, although GcvB overexpression destabilised its target *dppA*[46], CyaR levels were not significantly affected. However, levels of CyaR targets *nadE* and *yqaE* significantly increased (albeit modestly), suggesting that in this case the regulation is bi-directional and that GcvB can act as a CyaR sponge (Fig. 5d, panel IV). All of the CyaR-GcvB chimeric reads contained fragments of the 5' region of CyaR and a part of the CyaR seed sequence and the data indicate that these base pair with the previously described R1 and R2 seed sequences (Supplementary Fig.4;[46]). In addition, we found chimeras containing the 5' end of GcvB and a portion of the CyaR seed sequence. Interestingly, except for one nucleotide, the first 9 nucleotides in GcvB are identical to the R2 seed sequence (Supplementary Fig.3), suggesting that CyaR interaction with GcvB has alternative binding modes.

In conclusion, this work uncovered two new and functional sRNA-sRNA interactions in *E. coli*. The biological significance of these interactions is detailed in the discussion.

Hfq CLASH identifies novel sRNAs in untranslated regions

Two lines of evidence from our data indicate that many more mRNAs may be harbouring sRNAs in their UTRs or be involved in base-pairing among themselves. Firstly, almost 13% of the intermolecular chimeras mapped to mRNA-mRNA interactions (Fig. 3a, Table S5). Second, we observed extensive binding of Hfq in 3'UTRs near transcriptional terminators, indicating *E. coli* 3'UTRs may contain many functional sRNAs, as observed in *Salmonella*[31]. We uncovered 3'UTRs of mRNAs base-pairing with other mRNA in 5'UTRs, coding sequences and 3'UTRs (Fig. 6a). We identified 148 3'UTR-containing mRNA fragments that were involved in 375 unique interactions (minimum two unique reads), 59 of which were also present in the RIL-seq S-chimeras data[18]. Thirteen of the 3'UTRs were found as part of chimeras in the RIL-seq data, while 12 appeared stabilised upon transient inactivation of RNase E performed in Salmonella (TIER-seq data[31]); Fig. 3a, Table S5). Out of the 375 3'UTR-mRNA chimeras, 70 were 3'UTRs fused to 5'UTRs of mRNAs, suggesting

that these may represent 3'UTR-derived sRNAs that base-pair with 5'UTRs of mRNAs, a region frequently targeted by sRNAs. Strikingly, many of 3'UTR fragments in these chimeras mapped to *cpxP*, which in *Salmonella* is known to harbour the CpxQ sRNA (Table S5). Our analyses greatly increased the number of potential CpxQ mRNA targets. We also identified six mRNA 3'UTRs that were uncovered in all three datasets highlighted in Fig. 6a, suggesting these were most likely to contain sRNAs. Northern blot analyses confirmed the presence of sRNAs in *malG* and *gadE* 3'UTRs (Fig 6b, Supplementary Fig.5a). The latter was also recently experimentally confirmed in the RIL-seq data and was annotated as *gadF*[18]. Furthermore, significant Hfq cross-linking could be detected in the 3'UTRs of these transcripts (Fig. 6c). Our Hfq CLASH data revealed many more 3'UTRs that could potentially harbor sRNAs (Table S5). For example, we could show that the 3'UTR of *ygaM*, which was found in chimeric reads in our data, also likely contains an sRNA as a short fragment originating from the *ygaM* 3'UTR could be detected that cross-linked robustly to Hfq (Fig. 6b,c and Supplementary Fig.5b). We conclude that 3'UTRs may be a larger reservoir of sRNAs than previously anticipated.

The majority of sRNAs are more abundant at higher cell densities (including the *gadE*-, *ygaM*-derived sRNAs and *rybB*; see Fig. 6b, 2a), most likely because Hfq is highly expressed at these stages (Fig. 1c). In sharp contrast, the sRNA fragment identified in the *malG* 3'UTR was expressed very transiently and peaked at an OD_{600} of 1.8 (Fig. 6b), reminiscent of what was observed for CyaR (Fig. 1a). We envisioned that the particularly transient expression of this sRNA may be correlated with a role in the adaptive responses triggered during transition from exponential to stationary phases of growth. We named it MdoR (mal-dependent OMP repressor) and characterized it in detail.

MdoR is a 104 nt sRNA that contains part of the *malG* coding sequence, including the stop codon and it contains a Rho-independent terminator (Supplementary Fig.6a,b). The full-length *malG* and *malEFG* transcripts recapitulate the same expression profile: both peak at OD_{600} of 1.8 and drop to almost undetectable levels at OD_{600} 2.4, whereas a fraction of MdoR molecules remain stable at OD_{600} 2.4, perhaps due to stabilization by Hfq (Fig. 7a,b). Our RNAseq data revealed that all genes involved in maltose transport and metabolism have the same cell-density dependent dynamics in expression (Fig. 7a). Clearly, under our growth conditions, the *mal* regulon is specifically induced as cells reach the late exponential/transition into stationary phase. The similarity in the shape of MdoR expression profile shows that this sRNA is not made under different conditions than the transcription unit it overlaps, corroborating to the hypothesis that it is not independently produced. Additionally, we identified shorter *malG* 3'UTR-containing fragments of intermediate length between *malG* and MdoR that could be degradation intermediates (Fig. 7b). The detection of these intermediate species suggests that the *malEFG* primary transcript is undergoing serial ribonucleolytic cleavage steps. The fact that a *malG* fragment was also detected in the *Salmonella* TIER-seq

data[31] suggests that this sRNAs is released through RNase E processing or degradation. In addition, no clear promoter sequence could be found in the *malG* coding sequence, supporting the notion that MdoR is not generated as a primary transcript. Collectively, this suggests that MdoR is a product of ribonucleolytic processing and/or degradation of the *malEFG* or *malG* transcripts. To test this possibility, we used a temperature-sensitive RNase E mutant strain (rne^{ts} [47]). The expectation was that MdoR levels would decrease at the non-permissive temperature. Unfortunately, MdoR was undetectable after shifting the parental strain to 43ºC, suggesting the sRNA is unstable at elevated temperatures. In the RNase E mutant, however, MdoR became stabilized at the non-permissive temperature (Supplementary Fig.6c). We conclude that RNase E can degrade MdoR, however, the data do not allow us to determine whether the sRNA is also released from the *malG* 3'UTR by the same nuclease. As an alternative approach to test whether these intermediate species and MdoR are degradation products, we asked whether these products contained mono-phosphates at their 5' termini, a hallmark of many ribonuclease cleaved RNAs. Total RNA extracted from cells at OD_{600} 1.2 and 1.8, conditions in which MdoR is most abundant, was treated with Terminator 5'-Phosphate Dependent Exonuclease (5'-P Exo) (Fig. 7c). This exonuclease degrades RNA species with 5' mono-phosphate but not 5'-triphosphate ends. MdoR could not be detected in RNA samples treated with 5'-P Exo (Fig. 7c, lanes 2 and 4), suggesting it bears a 5' mono-phosphate. In contrast, RybB, an sRNA with a 5' tri-phosphate transcribed from an independent promoter[48,49] was a poor substrate for the exonuclease (Fig. 7c). As a positive control we included RaiZ, an sRNA with a 5'-mono-phosphate generated by RNase E processing[8]. Like MdoR, RaiZ was degraded in the presence of 5'-P Exo. 5'-P Exo treatment of the total RNA also reduced the levels of the longer intermediate species as well as the full-length *malG*. These data support a mechanism by which the full-length polycistronic RNA undergoes decay that is initiated at a site in the upstream *malEFG* region. It is known that constitutive processing of *malEFG* mRNA produced from a single promoter allows uncoupling of gene expression within the operon. The distal gene *malE* is clipped off and selectively stabilized, allowing it to be expressed at higher levels than other members of the operon[50]. This maturation process involves the degradosome[51] and turnover generates 5'-mono-phosphorylated mRNA fragments that promote further rounds of cleavage that ultimately degrade *malG*. The *malG* 3'UTR, however, would be less susceptible to degradation as it is stabilized by Hfq binding.

MdoR directly regulates the expression of major outer membrane porins and represses the envelope stress response pathway

The MdoR chimeras frequently contained 5'UTR fragments of two mRNAs encoding major porins, *ompC* and *ompA* (Fig. 8a-c). We also found MdoR chimeras containing

fragments of a *cis*-encoded sRNA, OhsC (Fig. 8d), suggesting that either MdoR controls its expression by sponging, or *vice versa)*. The most abundant and favourable interactions of MdoR with mRNAs (*ompC* and *ompA*) appear to be utilizing roughly the same region of the *malG* 3'UTR for base-pairing (Fig. 8b,c), indicative that the corresponding site on the predicted sRNA may be a main, functional seed. The predicted interaction between MdoR and *ompC* is unusually long and consists of two stems interrupted by a bulge, suggesting that these two RNAs form a stable complex. Conservation analyses and *in silico* target predictions (CopraRNA[52,53]) indicate that the seed sequence predicted by CLASH is relatively well-conserved (Supplementary Fig.7A), and could be utilized for the regulation of multiple targets (Supplementary Fig.7b).

To verify the MdoR CLASH data we pulse-overexpressed the sRNA from a plasmid-borne arabinose inducible promoter followed by total RNA sequencing (Fig. 8e). To minimize secondary changes in gene expression, cells were harvested after 15 minutes of MdoR induction, by which time the sRNA was readily detectable. Note that induction was performed at OD_{660} = 0.4, when endogenous levels of MdoR are very low. As a control we used cells harbouring an empty vector. Differential gene expression analysis (DESeq2;[54]), identified ~20 transcripts that were significantly overrepresented in the control data compared to the MdoR overexpression data (Fig. 8f; Table S6). Thus, these are likely downregulated by MdoR *in vivo*. This included the sigma factor *rpoE* (σ^E), which plays an important role in controlling gene expression during stress responses, including envelope stress[55–58]. The observed reduction in levels of the anti-σ^E protein RseA can be explained by the fact that σ^E and RseA are part of the same operon. Intriguingly, MdoR overexpression also reduced the levels of sRNAs RyeA and MicA, the latter of which depends on σ^E for its expression[59]. In *Salmonella* MicA down-regulates another outer membrane protein, LamB, a high affinity maltose/maltodextrin transporter[60]. Fragments of three mRNAs (*ompC*, *ompA* and ptsH) that were found in MdoR chimeric reads were also differentially expressed in the MdoR overexpression RNA-seq data, providing strong evidence that these are direct MdoR targets. PtsH is a component of the phosphoenolpyruvate-dependent sugar phosphotransferase system. Of note is that five unique MdoR-*ptsH* chimeric reads were identified in the CLASH data, reinforcing the notion that even chimeras with low unique read counts in CLASH can represent genuine interactions.

All the available data suggest that MdoR is part of a mixed feed forward loop that enhances the uptake of maltose/maltodextrin by maltose transporters (also see discussion). By base-pairing with the 5'UTRs of the mRNAs it firstly reduces the flux of the more general porins such as OmpC and OmpA to the outer membrane surface. Secondly, we propose that MdoR stimulates the accumulation of the high-affinity maltose transporter LamB in the OMP by suppressing the inhibitory σ^E pathway and MicA. To further test this model, we performed

additional validation experiments. Firstly, we confirmed the DESeq2 results for a number of the regulated genes (*ompC, rpoE*, *micA* and *ryeA*) by RT-qPCR (Fig. 9a). We included an MdoR mutant in which seed sequence in stem 1 was changed into its complementary sequence (Supplementary Fig.6a). As a control we also included the RybB sRNA, which is known to regulate both *rpoE* and *ompC* expression[48,61,62]. In all cases, the MdoR SM mutations abolished the negative regulatory effect on target expression (Fig. 9a).

To demonstrate direct target regulation *in vivo*, we employed a well-established reporter system where an sRNA is co-expressed with a construct containing the mRNA target region fused to the coding sequence of superfolder green fluorescent protein (sfGFP)[63,64] (Supplementary Fig.8b). Fusions for *ompC*, *ompA* and σ^E were constructed, but only the OmpC and OmpA-sfGFP reporters produced stable fusions that could be analysed. We also included an MdoR sRNA seed sequence mutant (MdoR SM) and an *ompC* mutant containing compensatory mutations (OmpC SM; Supplementary Fig.8a). GFP expression and cell density were quantified in plate readers. As a positive control we used the MicC and RybB sRNAs they regulate *E. coli ompC* expression[62,65]. Fluorescence measurements confirmed that levels of OmpC-sfGFP and OmpA-sfGFP fusions were significantly lower in cells expressing MdoR (Supplementary Fig.8c). Importantly, MdoR over-expression did not change fluorescence in cells only expressing the GFP coding sequence (Supplementary Fig.8c). Mutating the MdoR seed region largely restored OmpA- and OmpC-sfGFP reporter levels to the levels of the no sRNA negative control. The MdoR SM mutant was still able to partially suppress expression of the OmpC SM-sfGFP mutant, suggesting that base-pairing interactions between these two mutants is less stable compared to the wild-type (Supplementary Fig.8c). The wild-type MdoR was also able to partially suppress expression of the *ompC* SM mutant. We suggest that the predicted base-pairing interactions between MdoR and *ompC* in the second stem (Supplementary Fig.8a) might be sufficient to partially suppress *ompC* expression. Regardless, the data strongly imply that MdoR directly regulates *ompC* expression.

To distinguish between RNA degradation and regulation at the level of translation, we performed two types of experiments. We first measured expression of the GFP reporters by RT-qPCR using GFP-specific primers. The results were essentially identical to the GFP fluorescence measurements (Fig. 9b); reduced *ompC*-sfGFP mRNA levels were observed in cells expressing MdoR, but not in cells expressing MdoR SM. MdoR was still able to partially suppress the *ompC* SM mutant, but MdoR SM was able to downregulate *ompC* SM significantly better. Finally, we also performed polysome profiling experiments to assess the level of *ompC* translation upon overexpression of MdoR. Although this did not noticeably affect 70S and polysome levels (Supplementary Fig.9a), we observed a significant (~37%) reduction of *ompC* mRNA in the polysomal fractions (Supplementary Fig.9b). This suggests that MdoR

can also repress translation of *ompC.* However, because in this experiment over-expression of MdoR reduced *ompC* mRNA levels ~5-fold, we conclude that MdoR predominantly regulates OmpC protein expression by targeting the *ompC* mRNA for degradation.

MdoR enhances maltoporin expression during maltose fermentation

To further substantiate these results, we next switched to a more controlled system to investigate the effect of endogenous MdoR on its targets. In order to determine whether MdoR has a role in adaptation to maltose-metabolising conditions, single overnight cultures grown in glucose were split and (re)inoculated in fresh medium containing either glucose or maltose as the sole carbon sources, and expression of several *mal* regulon genes and MdoR targets were quantified. We show that MdoR and its parental transcript *malG* are almost undetectable during growth in glucose, and highly expressed during growth in maltose (~35-fold increase, Supplementary Fig.10a,b). This is consistent with catabolite repression of the *mal* regulon by glucose, and its induction my maltose[66]. Intriguingly, we observed that *ompC* mRNA levels are overall significantly lower during growth in maltose, compared to glucose (Supplementary Fig.10a). This suggests that porin expression is also regulated by nutrient source in *E. coli*. Similarly, MicA, a repressor of LamB synthesis, has reduced levels in maltose compared to glucose (Supplementary Fig.10a). We next mutated the entire seed sequence of the chromosomal copy of MdoR (both stem 1 and 2) in the chromosomal context, to completely disrupt base-pairing with *ompC* mRNA. Notably, the mutant MdoR sRNA is less abundant than the wild-type and longer fragments that contain upstream *malG* regions could be more readily detected (Supplementary Fig.10b). We speculate that the mutation in the upstream MdoR sequence might have affected RNase E recruitment or cleavage, impairing MdoR processing. The MdoR mutant strain also accumulated significantly higher levels of MicA (Fig. 9c) and less *lamB*. *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 strongly 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[67–69], they can regulate the kinetics of gene expression as well

as suppress noisy signals. 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 pieces 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 performing 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[28]) 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[13,18] but we also identified a large number of novel interactions, in particular sRNA-sRNA interactions. Our data indicate that sRNA sponging may be more prevalent than previously anticipated. We validated a number of our CLASH findings and uncovered a novel 3'UTR derived sRNA that plays a role in enhancing uptake of an alternative carbon source during the transition to stationary phase by inhibiting competing pathways.

Improving Hfq CLASH

Our *S. cerevisiae* Cross-linking and cDNA analysis data (CRAC;[70]) showed that a percentage of the cDNAs were formed by intermolecular ligations of two RNA fragments (chimeras) that were known to base pair *in vivo*[28]. These findings prompted the development of protocols that enable enrichment of sRNA-target chimeric reads using Hfq as an obvious bait. The initial Hfq UV cross-linking data (CRAC;[37]) did not yield sufficiently high numbers of chimeric reads to extract new biological insights and it was proposed that duplexes formed by Hfq are rapidly transferred to the RNA degradosome in line with observations from other groups[71–73], greatly reducing the likelihood of capturing sRNA-target interactions with Hfq using CLASH[13]. However, a recent study demonstrated that Hfq can be used effectively as a bait to enrich for sRNA-target duplexes under lower-stringency purification conditions suggesting that sRNA-mRNA duplexes are sufficiently stable on Hfq to allow purification[18]. This encouraged us to further optimize the CLASH method. We made a number of relatively small changes to the protocol that, when combined, enabled us to recover a large number of sRNA-target chimeric reads (detailed in Materials and Methods). We shortened various incubation steps to minimize RNA degradation and performed very long and stringent washes after bead incubation steps to remove background binding of non-specific proteins and RNAs. We very carefully controlled the RNase digestion step that is used to trim the cross-linked

RNAs prior to making cDNA libraries; this ensured that we would recovered longer chimeric RNA fragments. The resulting cDNA libraries were paired-end sequenced to increase the recovery of chimeric reads with high mapping scores from the raw sequencing data. These improvements led to a substantial improvement in the recovery of chimeric reads (9.5% compared to 0.001%. 0.71% were intermolecular chimeras). Another important change we made is that we substantially reduced the UV cross-linking times required for Hfq cross-linking. We recently developed the Vari-X-linker[32], which enables us to cross-link protein-RNA interactions in actively growing cells in seconds and we used filtration devices to rapidly harvest the bacterial cells. We previously showed that filtration combined with short UV cross-linking times dramatically reduces noise introduced by the activation of the DNA damage response and significantly increased the recovery of short-lived RNA species[32]. Encouragingly, there is considerable overlap between published Hfq RIL-seq data and our CLASH data (Table S1), however, we detected a large number of interactions that were not previously described. We speculate that many of these interactions represent short-lived RNA duplexes that are preferentially captured with our UV cross-linking and rapid cell filtration setup.

A major strength of CLASH is that the purification steps are performed under highly stringent and denaturing conditions. During the first FLAG affinity purification steps the beads are extensively washed with high salt buffers and the second Nickel affinity purification step is done under completely denaturing conditions (6M guanidium hydrochloride) to ensure that only RNAs that are covalently bound to the protein of interest are purified. These two purification steps, combined with the more extensive washes we have introduced, can generate high quality Hfq CLASH data. Consistent with its canonical role in RNA metabolism, meta analyses of the raw data showed a strong enrichment of Hfq in 5'-and 3'UTRs (Supplementary Fig.1a, c) and motif analyses of read clusters uncovered the known Hfq RNA-binding motifs (Supplementary Fig.1b and Fig. 3d). Moreover, when we mapped the mRNA fragments from the chimeric reads to the genome (without prior statistical filtering), discrete peaks were detected that predominantly mapped to UTRs and the background signal was low. This demonstrates the overall robustness and the specificity of our improved CLASH protocol. Comparison of our Hfq CLASH data with the Hfq RIL-seq data and previously validated interactions (sRNATarbase3) revealed that ~50% of the shared interactions were relatively low abundant in our data (after collapsing the data; Fig. 4d). This was unexpected as intuitively one would predict that chimeras with low read counts would have a higher likelihood of representing spurious interactions. Analysis of the ArcZ and ChiX chimeric fragments revealed that the vast majority overlapped with the known seed sequences (Fig. 4c), even chimeras with counts as low as two. Altogether, this suggest that many of the low abundance chimeras

in CLASH data represent *bona fide* interactions. Therefore, in this manuscript we reported all chimeras that have at least two read counts (see Supplementary Tables S1-S3).

Correlation between Hfq binding and steady state levels of RNA substrates

Hfq binding to sRNAs is important for stabilizing the sRNA as well as promoting sRNA-target interactions. We therefore assumed therefore that cDNA counts obtained from the Hfq CLASH data would show a strong positive correlation with the steady state RNA levels. However, a strong correlation could only be observed at the end of exponential phase. The prevailing wisdom is that many sRNAs would be degraded if not associated with Hfq, therefore this was quite a surprising finding. We propose that this may be linked to the availability of Hfq, as the protein is ~15 times less abundant at exponential phase as compared to stationary (Fig. 2c). Similar changes in Hfq expression at different growth phases have also been observed in pathogenic bacteria[74,75]. It is conceivable that at different growth stages sRNAs are packaged in different RNPs and that the composition of these complexes is dynamic. The *E. coli* sRNA IsrA/McaS has been shown to associate with a large number of different proteins, including Hfq, ProQ and CsrA; RNA chaperones that are known to bind and stabilize sRNAs and regulate sRNA-target interactions[34,76–78]. It is tempting to speculate that the composition of this RNP may be growth phase dependent. Therefore, Hfq may not be essential to stabilize all sRNAs at low cell densities and their stability may vary at different growth stages. A plausible model is that some sRNAs are sequestered and sufficiently stabilized by other RBPs (such as ProQ and CsrA[77]) during early growth stages and that Hfq can only stably associate with these RNPs once expression levels are sufficiently high.

The 3'UTR derived sRNA MdoR functions in a mixed feed forward loop by suppressing opposing pathways

We found a large number of chimeras that represent over a hundred distinct intermolecular interactions between 3'UTRs and other mRNA features, which implicate direct mRNA-mRNA communication. One mechanism by which mRNA crosstalk is achieved is by sRNA generation through degradation of the parental mRNA. These fragments were primarily described as decoys or sponges for other sRNAs[79], but could act as *trans*-acting sRNAs as well[29]. Among the interactions in our data we found 70 interactions between 3'UTRs and 5'UTRs (Table S5). We speculate many of these represent novel 3'UTR-derived sRNAs that target 5'UTRs of mRNAs. These could either have been processed from 3'UTRs in an RNase E (or other RNase)-dependent manner or transcribed from an internal promoter. Some of these predictions were validated by our follow up work (discussed below) and, while this work was in progress, also by others[18]. One of the 3'UTR-derived sRNAs we uncovered (MdoR) was of particular interest as it is only detected during the transition from late exponential to

early stationary phase. A model for how we believe MdoR functions in maltose uptake is shown in Figure 10a. We showed that MdoR directly regulates the expression of outer membrane porins (OmpA and OmpC) through direct base-pairing interactions and inhibits the membrane stress response activation by suppressing σ^E and the MicA sRNA. Our data do not allow us to determine whether MdoR downregulates σ^E and MicA directly, however, it is worth mentioning that a low number of chimeric reads between the 3'UTR of *malG* and MicA were found in the unfiltered RIL-Seq data[18], suggesting it could well be direct. Unfortunately, we were unable to test direct interactions between MdoR and σ^E *in vivo* as all the σ^E-GFP fusions we generated were not stably expressed. Although several 3'UTR derived sRNAs have been described in diverse bacterial species[26,29,31,79–81], MdoR is unique in a sense that it not only a 3'UTR derived sRNA that targets multiple pathways, but it is also part of a mixed feed forward loop that enhances the uptake of the alternative carbon sources maltose and maltodextrin by suppressing opposing pathways (Fig. 10a). The genetic structure and transcriptional regulation of the *mal* regulon are well understood. However, its post-transcriptional regulation has remained largely unexplored. Our work uncovered new links between the maltose uptake (*mal*) regulon, envelope stress-responses and membrane composition/assembly pathways. While initially cells are initially metabolising more favourable carbon sources such as glucose, these carbon sources are generally rapidly depleted and bacteria need to quickly switch to alternative C-sources, such as maltose and maltodextrins. During maltose utilisation, the *malEFG* operon is transcribed by the MalT transcription factor and the transcript is processed by RNase E as well as other degradosome components[51]. Here we show that the MdoR sRNA is a product of *malEFG* processing, which is protected from further degradation by Hfq. Efficient uptake of maltose/maltodextrin not only requires the inner membrane transporters encoded by the *malEFG* operon, but also the high-affinity transporter LamB (Fig. 10a), which cooperates with the inner membrane proteins to import these carbon sources. LamB is significantly upregulated when cells start to utilize alternative carbon sources. To promote maltose uptake via LamB, MdoR downregulates MicA, an sRNA repressor of LamB. This suppression enhances the translation of *lamB*. Expression of OMPs, however, needs to be carefully coordinated as any changes in the protein composition of the OM, such as changes in the levels of LamB, can lead to induction of the σ^E envelope stress response[82]. Therefore, to ensure efficient LamB production, we propose that MdoR (most likely indirectly) suppresses the σ^E pathway that negatively regulates *lamB* expression through MicA (Fig. 10a). This regulatory cascade triggers envelope remodelling and requires chaperones for correct insertion of LamB into the membrane. We propose that the MdoR-mediated downregulation of *rpoE* (~30% reduction upon MdoR overexpression) moderates σ^E induction caused by the reorganisation of the membrane when maltodextrins are utilised. When ectopically expressed

at high levels, MdoR reduces mRNA levels of *ompC*, *ompA* and other membrane proteins. Reducing the synthesis of OmpA and OmpC proteins may free up these resources enabling more efficient production of LamB. Due to the very high abundance and intrinsic stability of many of these OMP mRNAs (minute-long half-lives), we rationalise that even a mild reduction in their steady state can profoundly relieve the pressure on the OMP synthesis and assembly pathways[83]. The net outcome of *mal* regulon transcription, MdoR biogenesis and regulatory activity, is increased expression of high-affinity components of maltose-specific transport (MalE and LamB). This results in higher diffusion rates of maltodextrins inside the cell.

CLASH uncovers a large network of sRNA-sRNA interactions

Our analyses uncovered an unexpectedly large number of sRNA-sRNA interactions. The majority of these interactions involved known sRNA seed sequences, suggesting that these could represent *bona fide* sponging interactions that prevent sRNAs from base-pairing with their targets. About 40 sRNA-sRNA interactions involving sRNAs from the core genome were detected in the RNase E CLASH dataset, about a quarter of which were also detected in our Hfq data. The relatively low overlap between the datasets is not surprising, given the differences in the growth conditions (virulence inducing conditions for EHEC *E. coli* vs growth transitions for *E. coli* MG1655). Moreover, many of the sRNA-sRNA interactions recovered in association with RNase E are likely duplexes that are in the process of being degraded. Many of the sRNA-sRNA interactions unique to our Hfq data may represent sponging interactions between anti-sRNAs and seed sequences of sRNAs that interfere with target regulation and may not necessarily involve recruitment/activity of RNase E.

The chimeras containing CyaR fragments were of particular interest as CyaR is preferentially expressed during the transition from late exponential to stationary phase (Fig. 2a[45,84]) and may therefore play an important role in adaptation to nutrient availability. We could show that over-expression of ArcZ reduced CyaR levels and that CyaR can down-regulate GcvB levels (Fig. 10b). Interestingly, over-expression of ArcZ in *Salmonella* showed a dramatic reduction in CyaR bound to Hfq and upregulation of CyaR targets, such as *nadE*[42], suggesting that this activity is conserved between these two Gram-negative bacteria. The fate of these sRNA-sRNA duplexes may depend on where the interactions take place. Folding of the chimeric reads suggests that ArcZ preferentially base-pairs with CyaR at the 5' end (Supplementary Fig.3a), which may alter secondary structures that normally help to stabilize the sRNA. The regulation appeared to be unidirectional (Fig. 10b), however, under the tested conditions ArcZ levels were always higher than CyaR, even after over-expressing the latter. Therefore, it is certainly possible that under conditions were ArcZ levels are lower, CyaR may be able to exert a regulatory effect.

Unlike previously reported sRNA decoys or sponges[39,40,79], ArcZ and CyaR are transcribed from independent promoters, and target mRNAs associated with many different processes. Thus, these interactions would be expected to connect many different pathways. For example, ArcZ regulation of CyaR may connect adaptation to stationary phase/biofilm development[45,85] to quorum sensing and cellular adherence[45]. CyaR expression is controlled by the global regulator Crp. Most of the genes controlled by Crp are involved in transport and/or catabolism of amino acids or sugar. It was therefore surprising that, as noted by the authors[45], the mRNAs targeted by CyaR had little to do with these processes. Our data revealed a missing link and suggest that CyaR indeed plays a role in adaptation to nutrient availability by supressing the activity of GcvB, which negatively regulates mRNAs involved in amino acid and peptide uptake (Fig. 10b). We propose that CyaR is part of a mixed coherent feed forward loop that stimulates amino acid uptake during the transition from late exponential to the stationary phase (Fig. 10b). Interestingly, ArcZ downregulates the *sdaCB* dicistron which encodes for proteins involved in serine uptake and metabolism[42]. This operon has been shown to be regulated by Crp as well, suggesting that ArcZ can counteract the activity of Crp.

Materials and Methods

Bacterial strains and culture conditions

An overview of the bacterial strains used in this study is provided in Table S7. The *E. coli* MG1655, TOP10 or TOP10F' strains served as parental strains. The *E. coli* K12 strain used for CLASH experiments, MG1655 *hfq*::HTF was previously reported[37]. Cells were grown in Lysogeny Broth (LB) or minimal medium with supplements (1xM9 salts, 2 mM $MgSO_4$,0.1 mM $CaCl_2$, 0.03 mM thiamine, 0.2% carbon-source) at 37°C under aerobic conditions with shaking at 200 rpm. The media were supplemented with antibiotics where required at the following concentrations: ampicillin - 100 µg/ml, chloramphenicol - 25 µg/ml, kanamycin - 50 µg/ml. Where indicated, 0.2% glucose or maltose were used. For induction of sRNA expression from plasmids, 1 mM IPTG, 200 nM anhydrotetracycline hydrochloride or 0.2% L-arabinose were used.

Construction of sRNA expression plasmids

For the pulse-overexpression constructs, the sRNA gene of interest was cloned at the transcriptional +1 site under P*ara* control by amplifying the pBAD+1 plasmid (Table S7) by inverse PCR using Q5 DNA Polymerase (NEB). The pBAD+1 template is derived from pBAD*myc*His A[37]. The sRNA genes and seed mutants (SM) were synthesized as ultramers (IDT; Table S7), which served as the forward primers, as described(Tree et al., 2014). The reverse primer (oligo pBAD+1_5P_rev) bears a monophosphorylated 5'-end to allow blunt-

end self-ligation. The PCR reaction was digested with 10U DpnI (NEB) for 1h at 37°C and purified by ethanol precipitation. The sRNA-pBAD linear PCR product was circularized by self-ligation, performed as above. Ligations were transformed in DH5α competent cells. Positive transformants were screened by sequencing. The control plasmid pBAD+1 was constructed similarly by self-ligation of the PCR product generated from oligonucleotides pBAD+1_XbaI_fwd and pBAD+1_5P_rev. Small RNA overexpression constructs derived from the pZA21MCS and pZE12*luc* (Expressys) were generated identically, using the indicated ultramers in Table S7 as forward primers, and oligos pZA21MCS_5P_rev and pZE12_5P_rev as reverse primers, respectively, and transformed in *E. coli* TOP10F'.

Construction of mRNA-superfolder GFP fusions

Table S7 lists all the plasmids, gene fragments and primers used for cloning procedures in this work. To construct constitutively expressed, in-frame mRNA-*sfGFP* fusions for the fluorescence reporter studies, the 5'UTR, start codon and first ~5 codons of target genes were cloned under the control of PLtetO-1 promoter in a pXG10-SF backbone as previously described[63,64]. Derivatives of the target–GFP fusion plasmids harboring seed mutations (SM) were generated using synthetic mutated gene-fragments (IDT, Table S7). To prepare the inserts, the target region of mRNA of interest was either amplified by PCR from *E. coli* genomic DNA or synthesized as g-blocks (IDT) and cloned using NheI and NsiI restriction sites. Transformants were screened by restriction digest analysis and verified by Sanger sequencing (Edinburgh Genomics).

Hfq UV Cross-linking, Ligation and AnalysiS of Hybrids (Hfq-CLASH)

CLASH was performed essentially as described[13], with a number of modifications including changes in incubation steps, cDNA library preparation, reaction volumes and UV cross-linking. *E. coli* expressing the chromosomal Hfq-HTF were grown overnight in LB at 37°C with shaking (200 rpm), diluted to starter OD_{600} 0.05 in fresh LB, and re-grown with shaking at 37°C in 750 ml LB. A volume of culture equivalent to 80 OD_{600} per 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 immediately subjected to UV (254 nm) irradiation for 22 seconds (~500 mJ/cm2) in the Vari-X-linker[32]. Cells were harvested using a rapid filtration device [32] onto 0.45 µM nitrocellulose filters (Merck Millipore) and flash-frozen on the membrane in liquid nitrogen. The following day, the membranes were washed with ~15 ml ice-cold phosphate-buffered saline (PBS), and cells were harvested by centrifugation. Cell pellets were lysed by bead-beating in 1 volume per weight TN150 buffer (50mM Tris pH 8.0, 150 mM NaCl, 0.1% NP-40, 5 mM β-mercaptoethanol) in the presence of protease inhibitors (Roche), and 3 volumes 0.1 mm Zirconia beads (Thistle Scientific), by performing 5 cycles of 1 minute vortexing followed by 1-

minute incubation on ice. One additional volume of TN150 buffer was added. To reduce the viscosity of the lysate and remove contaminating DNA the lysate was incubated with RQ1 DNase I (10U/ml Promega) for 30 minutes on ice. Two-additional volumes of TN150 were added and mixed with the lysates by vortexing. The lysates were centrifuged for 20 minutes at 4000 rpm at 4°C and subsequently clarified by a second centrifugation step at 13.4 krpm, for 20 min at 4°C. Purification of the UV cross-linked Hfq-HTF-RNA complexes and cDNA library preparation was performed as described [70]. Cell lysates were incubated with 50 µl of pre-equilibrated M2 anti-FLAG beads (Sigma) for 1-2 hours at 4°C. The anti-FLAG beads were washed three times 10 minutes with 2 ml TN1000 (50 mM Tris pH 7.5, 0.1% NP-40, 1M NaCl) and three times 10 minutes with TN150 without protease inhibitors (50 mM Tris pH 7.5, 0.1% NP-40, 150mM NaCl). For TEV cleavage, the beads were resuspended in 250 µl of TN150 buffer (without protease inhibitors) and incubated with home-made GST-TEV protease at room temperature for 1.5 hours. The TEV eluates were then incubated with a fresh 1:100 dilution preparation of RNaceIt (RNase A and T1 mixture; Agilent) for exactly 5 minutes at 37ºC, after which they were mixed with 0.4g GuHCl (6M, Sigma), NaCl (300mM), and Imidazole (10mM). Note this needs to be carefully optimized to obtain high-quality cDNA libraries. The samples were then transferred to 50 µl Nickel-NTA agarose beads (Qiagen), equilibrated with wash buffer 1 (6 M GuHcl, 0.1% NP-40, 300 mM NaCl, 50 mM Tris pH 7.8, 10 mM Imidazole, 5 mM beta-mercaptoethanol). Binding was performed at 4ºC overnight with rotation. The following day, the beads were transferred to Pierce SnapCap spin columns (Thermo Fisher), washed 3 times with wash buffer 1 and 3 times with 1xPNK buffer (10 mM $MgCl_2$, 50mM Tris pH 7.8, 0.1% NP-40, 5 mM beta-mercaptoethanol). The washes were followed by on-column TSAP incubation (Thermosensitive alkaline phosphatase, Promega) treatment for 1h at 37°C with 8 U of phosphatase in 60 µl of 1xPNK, in the presence of 80U RNasin (Promega). The beads were washed once with 500 µl wash buffer 1 and three times with 500 µl 1xPNK buffer. To add 3'-linkers (App-PE - Table S7), the Nickel-NTA beads were incubated in 80 µl 3'-linker ligation mix with (1 X PNK buffer, 1 µM 3'-adapter, 10% PEG8000, 30U Truncated T4 RNA ligase 2 K227Q (NEB), 60U RNasin). The samples were incubated for 4 hours at 25°C. The 5'-ends of bound RNAs were radiolabeled with 30U T4 PNK (NEB) and 3µl ^{32}P-γATP (1.1µCi; Perkin Elmer) in 1xPNK buffer for 40 min at 37°C, after which 100 mM cold ATP (Roche) was added to a final concentration of 1mM, and the incubation prolonged for another 20 min to complete 5'-end phosphorylation. The resin was washed three times with 500 µl wash buffer 1 and three times with equal volume of 1xPNK buffer. For on-bead 5'-linker ligation, the beads were incubated 16h at 16°C in 1xPNK buffer with 40U T4 RNA ligase I (NEB), and 1 µl 100 µM L5 adapter (Table S7), in the presence of 1mM ATP and 60U RNasin (Promega). The Nickel-NTA beads were washed three times with wash buffer 1

and three times with buffer 2 (50 mM Tris–HCl pH 7.8, 50 mM NaCl, 10 mM imidazole, 0.1% NP-40, 5 mM β-mercaptoethanol). The protein-RNA complexes were eluted in two steps in new tubes with 200 µl of elution buffer (wash buffer 2 with 250 mM imidazole). The protein-RNA complexes were precipitated on ice by adding TCA to a final concentration of 20%, followed by a 20-minute centrifugation at 4°C at 13.4 krpm. Pellets were washed with 800 µl acetone, and air dried for a few minutes in the hood. The protein pellet was resuspended and incubated at 65°C in 20 µl 1x NuPage loading buffer (Novex), resolved on 4–12% NuPAGE gels, and visualised by autoradiography. The cross-linked proteins-RNA were cut directly from the gel and incubated with 160 µg of Proteinase K (Roche) in 600 µl wash buffer 2 supplemented with 1% SDS and 5 mM EDTA at 55°C for 2-3 hours with mixing. The RNA was subsequently extracted by phenol-chloroform extraction and ethanol precipitated. The RNA pellet was directly resuspended in RT buffer and was transcribed in a single reaction with the SuperScript IV system (Invitrogen) according to manufacturer's instructions using the PE_reverse oligo as primer. The cDNA was purified with the DNA Clean and Concentrator 5 kit (Zymo Research) and eluted in 11 µl DEPC water. Half of the cDNA (5 µl) was amplified by PCR using Pfu Polymerase (Promega) with the cycling conditions (95°C for 2 min; 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). The PCR primers are listed in Table S7. PCR products were treated with 40U Exonuclease 1 (NEB) for 1 h at 37°C to remove free oligonucleotide and purified by ethanol precipitation/ or the DNA Clean and Concentrator 5 kit (Zymo Research). Libraries were resolved on a 2% MetaPhor agarose (Lonza) gel and 175-300bp fragments were gel-extracted with the MinElute kit (Qiagen) according to manufacturer's instructions. All libraries were quantified on a 2100 Bionalyzer using the High-Sensitivity DNA assay and a Qubit 4 (Invitrogen). Individual libraries were pooled based on concentration and barcode sequence identity. Paired-end sequencing (75 bp) was performed by Edinburgh Genomics on an Illumina HiSeq 4000 platform.

RNAseq

E. coli MG1655 was cultured and harvested as described for the CLASH procedure. Total RNA was extracted using the Guanidium thiocyanate phenol method. RNA integrity was assessed with the Prokaryote Total RNA Nano assay on a 2100 Bioanalyzer (Agilent). Genomic DNA was removed by incubating 10 µg of total RNA with 2U Turbo DNase (Ambion) in a 50 µl final volume for 30 minutes at 37°C in the presence of 10 U SuperaseIn RNase Inhibitor (Ambion). RNA was subsequently phenol-chloroform extracted and purified by ethanol-precipitation. Ribosomal RNA was removed with the Ribo-Zero rRNA Removal Kit (Gram-Negative Bacteria) (Illumina) according to the manufacturer's instructions. Successful rRNA depletion was verified on the Agilent 2100 Bioanalyzer. The RNA was fragmented for 5

min at 95°C in the presence of Superscript III buffer (Invitrogen) followed by a five minute incubation on ice. Reverse-transcription (RT) was performed with Superscript III (Invitrogen) in 20 µl reactions according to manufacturer's procedures using 250 ng of ribosomal RNA depleted RNA and 2.5 µM random hexamers (PE_solexa_hexamer, oligo 73, Table S7). The RNA and free primers were degraded using 20U of Exonuclease I (NEB) and 50U RNaseIf (NEB) and the cDNA was purified with the DNA Clean & Concentrator 5 kit (Zymo Research). Ligation of the 5' adapter (P5_phospho_adapter, oligo 39) to the cDNA was performed using CircLigase II (EpiCentre) for 6 hours at 60°C, followed by a 10 minute inactivation at 80°C. The cDNA was purified with the DNA Clean & Concentrator 5 kit (Zymo Research). Half of the cDNA library was PCR amplified using Pfu polymerase (Promega) using the P5 forward PCR oligonucleotide and barcoded BC reverse oligonucleotides (200 nM; Table S7; 95°C for 2 min, 95°C for 20s, 52°C for 30s and 72°C for 1 min, and a final extension of 72°C for 5 min. 20 cycles of amplification). The PCR products were treated with Exonuclease 1 (NEB) for 1 h at 37°C and purified by ethanol precipitation. Libraries were resolved on a 2% MetaPhor agarose (Lonza) gel 200-500 bp fragments were gel-extracted using the MinElute kit (Qiagen). All libraries were quantified on a 2100 Bionalyzer using the High-Sensitivity DNA assay. Individual libraries were pooled in equimolar amounts. Paired-end sequencing (75 bp) was performed by Edinburgh Genomics on a Illumina HiSeq 4000 platform.

Small RNA over-expression studies

Individual TOP10F' clones carrying pZA21 and pZE12-derived sRNA constructs and control plasmids combinations (Table S7) were cultured to OD_{600} 0.1 and expression of sRNAs was induced with IPTG and anhydrotetracycline hydrochloride for one hour. Cells were collected by centrifugation for 30 seconds at 14000 rpm, flash-frozen in liquid nitrogen and total RNA was isolated as above. Genomic DNA was digested with Turbo DNase (Ambion), then the RNA was purified with RNAClean XP beads (Beckman Coulter). Gene expression was quantified by RT-qPCR (see below) using 10 ng total RNA as template, and expressed as fold change relative to the reference sample containing pJV300[86] or pZA21.

For pulse-overexpression studies overnight MG1655 cultures containing pBAD::sRNA and empty pBAD+1 control plasmids were inoculated in fresh LB-ampicillin medium at a starting OD_{600} of 0.05, and grown aerobically at 37°C to OD_{600} 0.4. Pre-induction (0 min) and post-induction samples were harvested. For induction, cultures were supplemented with L-arabinose and rapidly collected by filtration and flash-frozen in liquid nitrogen at the indicated time-points. RNA was extracted from three biological replicate time-series, followed by RNASeq library preparation, next generation sequencing and DESeq2 analysis of differentially expressed genes.

GFP reporter system to quantify sRNA effect on target expression

A two-plasmid system was used to express each sRNA, and mRNA-*sfGFP* fusions[63,64] with modifications. The sRNA and sfGFP-fusion plasmids were co-transformed in *E. coli* TOP10 cells by electroporation and cells were maintained on dual selection with ampicillin and chloramphenicol. In TOP10 cells, the mRNA-*sfGFP* constructs are constitutively expressed, whereas sRNA expression requires L-arabinose induction. The expression of sfGFP-fused targets in the presence or absence of sRNAs was quantified at the protein level, by plate reader experiments and at the RNA level, by RT-qPCR.

For the plate reader experiments, a single colony of bacterial strain harbouring a sRNA-target-sfGFP combination was inoculated in a 96-well Flat Bottom Transparent Polystyrene plate (Fisher Scientific) with lid (Thermo Scientific) and cultured overnight at 37°C in 100 µl LB supplemented with antibiotics and L-arabinose to induce expression of sRNAs. Next day, each overnight inoculum was diluted 1:100 by serial dilution, in triplicate, in LB with freshly prepared L-arabinose to a final volume of 100 µl. Cultures were grown in a 96-well plate in an Infinite 200 Pro plate reader (Tecan) controlled by i-control software (Tecan) for 192 cycles at 37°C with 1 min orbital shaking (4 mm amplitude) every 5th minute. To monitor optical density over time, the following parameters were used: wavelength 600 nm, bandwidth 9 nm. Fluorescence was monitored with excitation wavelength 480 nm, bandwidth 9 nm and emission wavelength 520 nm, bandwidth 20 nm. Measurements were recorded at 5 minute intervals, by top reading. Raw data was processed following guidance from previous reports[63]. First, the range of linearity of increase of fluorescence with OD_{600} was identified for all individual triplicates. Only the linearity range common to all triplicates was considered for further analysis. For each set of triplicates, the mean fluorescence was calculated at each OD_{600}. To correct for background and cell autofluorescence, the fluorescence mean of a strain with plasmid pXG-0 was subtracted from all strains with GFP plasmids at the equivalent OD_{600}. Ultimately, a curve was generated for each sample, plotting the background-corrected fluorescence (GFP) versus OD_{600}. The experiments were performed for three biological replicates, and mean values and standard error of the means calculated for each strain.

RT-qPCR

Total RNA (12.5µg) was treated with 2U of Turbo DNase (Ambion) for 1 hour at 37°C in a 10 µl reaction in the presence of 2U SuperaseIn RNase inhibitor (Ambion). The DNase was inactivated by 10 minutes incubation at 75°C. Reverse transcription (RT) was performed in a single reaction for all target genes of interest using a mix of gene-specific RT primers at 3.5 µM concentration each. After addition of 2.5 µl RT primer mix, the RNA and primers were

denatured at 70°C for 3 min, then snap chilled and incubated on ice for 5 min. RT was performed for 1 hour at 55°C with SuperScript III (Invitrogen) using 5 µl of RNA-RT primers mix in 10 µl final volume (100 U Superscript III, 2.5 mM DTT, 1xFS Buffer, 0.75 mM dNTPs) in the presence of 1U RNasin (Promega). RT was followed by treatment with 5U RNase H for 30 min at 37°C to remove the RNA from the RNA-cDNA duplexes. The cDNA was diluted 10-fold with DEPC water. Quantitative PCR was performed on 50ng of DNAseI-treated total RNA using the Brilliant III UltraFast SYBR Green QPCR Master Mix (Agilent) and the Luna Universal One-Step RT-qPCR Kit (NEB) according to manufacturer's instructions. The qPCRs were run on a LightCycler 480 (Roche), and the specificity of the product was assessed by generating melting curves, as follows: 65°C-60s, 95°C (0.11 ramp rate with 5 acquisitions per °C, continuous). The data analyses were performed with the IDEAS2.0 software, at default settings: Absolute Quantification/Fit Points for Cp determination and Melt Curve Genotyping. The qPCR efficiency of primer pairs was assessed by performing standard curves by serial dilution of template RNA or genomic DNA. Negative controls such as -RT or no template control were used throughout, and the qPCR for all samples was performed in technical triplicate. Outliers from the samples with technical triplicate standard deviations of Cp > 0.3 were discarded from the analyses. To calculate the fold-change relative to the control, the 2^{-ddCp} method was employed, using *recA* or 5S rRNA (*rrfD*) as the reference genes where indicated. Experiments were performed for minimum two biological replicates, and the mean fold-change and standard error of the mean were computed. Unless otherwise stated, significance of the fold-change difference compared to the reference sample control (for which fold-change =1) was tested with a one-sample t-test.

Northern Blot analysis

Total RNA was extracted from cell lysates by GTC-Phenol extraction. For large RNA fragments, 10 µg of total RNA was resolved on a 1.25% BPTE- gel (pH 7) and transferred to a nylon membrane (HyBond N+, GEHealthcare) by capillarity. For short RNA fragments, 10 µg total RNA was separated on a 8% polyacrylamide TBE-Urea gel and transferred to a nylon membrane by electroblotting for four hours at 50 V. All membranes were UV-crosslinked in a Stratalinker at intensity 1200. Membranes were pre-hybridised in 10 ml of UltraHyb Oligo Hyb (Ambion) for one hour and probed with ^{32}P-labeled DNA oligo for 12-18 hours in a hybridization oven at 42°C. The sequences of the probes used for Northern blot detection are detailed in Table S7. Membranes were washed two times in 2xSSC with 0.5% SDS solution for 10 minutes, and visualized with a Phosphor imaging screen and FujiFilm FLA-5100 Scanner (IP-S mode). For detection of highly abundant species (5S rRNA) autoradiography was used for exposure.

Western blot analyses

E. coli MG1655 Hfq::*htf* strain was cultured in LB in the same conditions as the CLASH and RNA-seq experiments. Cells were collected at OD_{600} 0.4, 0.8, 1.2, 1.8, 2.4, 3 and 4. A volume of cell lysate containing 40 ug protein was run on PAGE gels for each OD_{600} and transferred to a nitrocellulose membrane. The membranes were blocked for one hour in blocking solution (5% non- fat milk in PBST (1X phosphate saline buffer, 0.1% Tween-20). To detect Hfq-HTF protein, the membrane was probed overnight at 4°C with the Rabbit anti-TAP polyclonal primary antibody (Thermo Fisher, 1:5000 dilution in blocking solution), which recognizes an epitope at the region between the TEV-cleavage site and His6. For the loading control we used a rabbit polyclonal to GroEL primary antibody (Abcam, 1:300000 dilution), for 2 hours at room temperature. After 3x10 min PBST washes, the membranes were blotted for two hours with a Goat anti-rabbit secondary antibody (Thermofisher, 1:5000 in blocking solution) at room temperature. Finally, after three 10-minute PBST washes, the proteins were visualised using Pearce enhanced chemiluminescence (ECL, Thermo Fisher) according to manufacturer's instructions

Primer extension analysis

One microgram total RNA was reverse-transcribed using SuperScript III reverse transcriptase (Invitrogen) using ^{32}P-radiolabelled oligonucleotides as primers (Table S7). Primers were added to the RNA and annealing was performed by heating the samples at 85°C for three minutes and then snap chilling them on ice. The RT was performed for one hour at 45°C, followed by Exonuclease I and RNaseIf (NEB) (0.5 µl each) treatment for 30 minutes at 37°C. Reactions were stopped by mixing with an equal volume of 2XRNA loading dye (NEB), 2 minutes incubation at 95°C and snap chilled. The sequencing ladders were prepared with Sequenase v2.0 (USB/Affymetrix) according to specified instructions. Samples were resolved on 6% PAA/8M TBE-urea gels and visualized using the FLA5100 phosphoimager system.

Construction of the MdoR seed-mutant strain

To mutate the chromosomal copy of MdoR, we used the λRed system [87]. We amplified the integration cassette from plasmid pKD4 with ultramers 895 and 896, containing homology regions to the coding sequence of *malG*, the desired MdoR sequence and to the region immediately downstream of the Rho-independent terminator, respectively (Fig S8). With this design, the scar after removal of the *Kan*^{r} cassette was expected at a site outside the MdoR/*malG* sequence. The PCR product was electroporated in *E. coli* MG1566 strains carrying the pKD46 plasmid from which λRed recombinase was induced with 10 mM L-

arabinose. Correct replacement of the MdoR seed sequence was screened by colony PCR using primer pairs: 725 & 909 and 726 & 910. The antibiotic resistance cassette was removed from substitution mutants by FLP-recombinase expressed constitutively from pE-FLP[88]. Successful allele replacement was confirmed by Sanger sequencing.

Polysome profiling analyses

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

Polysomal profiling was performed according to previously described protocols[89,90] with minor changes in the lysis buffer (10 mM NaCl, 10 mM $MgCl_2$, 10 mM Tris-HCl (pH 7.5), 100 ug/ml cycloheximide, 1% (w/v) Na-Deoxycholate, 1U DNAse, 0.6U/mL RiboLock in DEPC water). Lysates were kept on ice for 30 min, centrifuged 3X at 15000 g for 10 min. The supernatants were loaded on a linear 10%–30% [w/v] sucrose gradient and centrifuged for 4 hours using a SW41 rotor at 40000 rpm in a Beckman Optima XPN-100 Ultracentrifuge. Fractions of 1 mL in volume were collected monitoring the absorbance at 254 nm with the UA-6 UV/VIS detector (Teledyne Isco). Fractions from the entire gradient (total RNA) and from the fractions corresponding to ribosomes (70S) and polysomes (polysomal RNA) were pooled and RNA was purified by acid phenol–chloroform extraction according to[91].

Terminator™ 5′-PhosphateDependent Exonuclease treatment

Ten micrograms of total RNA extracted from cell-samples at OD_{600} 1.2 and 1.8 were treated with 5'-Terminator Dependent Exonuclease (Epicentre) as per manufacturer instructions using Buffer A. The reaction was terminated by phenol extraction and ethanol precipitation, and the RNA was loaded on 8% polyacrylamide-urea gels and transferred to nylon membranes that were probed for MdoR, RaiZ, RybB and 5S rRNA (Table S7).

Inactivation of RNase E

The *E. coli* rne-3071 (Table S7) and wild-type strains were grown in LB medium at 33°C to an OD_{600} of 1.5, then shifted to 43°C for 30 min to inactivate RNase E. Cells were harvested by rapid filtration (0.45 µM filters, Merck Millipore) and flash-frozen in liquid nitrogen. Detection of RNAs was performed by Northern blot.

Seed mutant studies

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

Computational analysis

Pre-processing of the raw sequencing data.

Raw sequencing reads in fastq files were processed using a pipeline developed by Sander Granneman, which uses tools from the pyCRAC package[92]. The entire pipeline is available at https://bitbucket.org/sgrann/). The CRAC_pipeline_PE.py pipeline first demultiplexes the data using pyBarcodeFilter.py and the in-read barcode sequences found in the L5 5' adapters. Flexbar then trims the reads to remove 3'-adapter sequences and poor quality nucleotides (Phred score <23). Using the random nucleotide information present in the L5 5'-adaptor sequences, the reads are then collapsed to remove potential PCR duplicates. The reads were then mapped to the *E. coli* MG1655 genome using Novoalign (www.novocraft.com). To determine to which genes the reads mapped to, we generated an annotation file in the Gene Transfer Format (GTF). This file contains the start and end positions of each gene on the chromosome as well as what genomic features (i.e. sRNA, protein- coding, tRNA) it belongs to. To generate this file, we used the Rockhopper software[93] on *E. coli* rRNA-depleted total RNA-seq data (generated by Christel Sirocchi), a minimal GTF file obtained from ENSEMBL (without UTR information). The resulting GTF file contained information not only on the coding sequences, but also complete 5' and 3' UTR coordinates. We then used pyReadCounters.py with novoalign output files as input and the GTF annotation file to count the total number of unique cDNAs that mapped to each gene.

Normalization steps.

To normalise the read count data generated with pyReadCounters.py and to correct for differences in library depth between time-points, we calculated Transcripts Per Million reads (TPM) for each gene. Briefly, for each time-point the raw counts for each gene was first divided by the gene length and then divided by the sum of all the values for the genes in that time-point to normalize for differences in library depth. Subsequently, only genes with a minimum of 5 TPM in all datasets compared were used. The TPM values for each OD_{600} studied were divided by the TPM values of the first sample (OD_{600} 0.4). Thus, the fold-change starts at 1 for all samples in the OD_{600} series. These ratios were then log_2-normalized. The log_2-normalized fold-changes were used to compare RNAseq and Hfq-cross-linking profiles among samples, and to perform k-means clustering with the python sklearn. cluster.KMeans

class.

Hfq-binding coverage plots.

For the analysis of the Hfq binding sites the pyCRAC package[92] was used (versions. 1.3.2 and 1.3.3). The pyBinCollector tool was used to generate Hfq cross-linking distribution plots over genomic features. First, PyCalculateFDRs.py was used to identify the significantly enriched Hfq-binding peaks (minimum 10 reads, minimum 20 nucleotide intervals). Next, pyBinCollector was used to normalize gene lengths by dividing their sequences into 100 bins, and calculate nucleotide densities for each bin. To generate the distribution profile for all genes individually, we normalized the total number of read clusters (assemblies of overlapping cDNA sequences) covering each nucleotide position by the total number of clusters that cover the gene. Motif searches were performed with pyMotif.py using the significantly enriched Hfq-binding peaks (FDR intervals). The 4-8 nucleotide k-mers with Z-scores above the indicated threshold were used for making the motif logo with the k-mer probability logo tool [94] with the -ranked option (http://kplogo.wi.mit.edu/).

Analysis of chimeric reads.

Chimeric reads were identified using the hyb package using default settings[35] and further analysed using the pyCRAC package[92]. To apply this single-end specific pipeline to our paired-end sequencing data, we joined forward and reverse reads using FLASH [95], which merges overlapping paired reads into a single read. These were then analysed using hyb. The -anti option for the hyb pipeline was used to be able to use a genomic *E. coli* hyb database, rather than a transcript database. Uniquely annotated hybrids (.ua.hyb) were used in subsequent analyses. To visualise the hybrids in the genome browser, the .ua.hyb output files were converted to the GTF format. To generate distribution plots for the genes to which the chimeric reads mapped, the parts of the chimeras were clustered with pyClusterReads.py and BEDtools[96] (intersectBed) was used to remove clusters that map to multiple regions. To produce the coverage plots with pyBinCollector, each cluster was counted only once, and the number of reads belonging to each cluster was ignored.

sRNA density plots.

To visualize the nucleotide read density of sRNA-target pairs for a given sRNA, we first merged the hyb datasets for all OD_{600} and biological replicates, and filtered the interactions represented by at least two chimeric reads in the unified dataset. For each sRNA-target pair in the filtered dataset, the hit counts at each nucleotide position for all chimeras were summed, regardless of the orientation of the chimeras. The count data was log_2-normalised (actually Log_2(Chimera count +1) to avoid NaN for nucleotide positions with 0 hits when log-

transforming the data).

sRNA-sRNA network visualization.

Only the interactions represented by at least two chimeric reads in the merged CLASH dataset were considered. For each such interaction, chimera counts corresponding to sRNA-sRNA in either orientation were summed, $\log_2$-transformed and visualized with the igraph Python package.

Differential expression analyses.

For the differential expression analyses DESeq2 was used[54]. Three MdoR pulse-overexpression datasets were compared to three pBAD Control overexpression datasets. Only differentially expressed genes that had an adjusted p-value of 0.05 or lower were considered significant.

Multiple sequence alignments and conservation analyses.

The homologous sequences of MdoR in other enterobacteria were retrieved by BLAST. JalView was used for the multiple sequence alignments, using the MAFFT algorithm[97].

Data and Code availability.

The next generation sequencing data have been deposited on the NCBI Gene Expression Omnibus (GEO) with accession number GSE123050. The python pyCRAC and kinetic-CRAC software packages used for analysing the data are available from https://bitbucket.org/sgrann. 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.

Author contributions

I.A.I, R.v.N and S.G conceived the experiments. All the authors contributed to designing and performing the experiments. I.A.I performed the majority of the experiments and carried out most of the bioinformatics data analyses. M.M. and G.V's contributions were instrumental for generating the polysome analysis data. I.A.I and S.G drafted the manuscript and all the authors reviewed the manuscript and approved the final version.

Acknowledgement

We are grateful to Lionello Bossi for his valuable feedback on the project and for very

fruitful discussions. We would like to thank the members of the Granneman lab for critically reading the manuscript. 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.) and the Australian National Health and Medical Research Council Project grants (APP1067241 and APP1139315 to J.J.T). G.V is supported by IMMAGINA BioTechnology s.r.l. and the Provincia Autonoma di Trento, Italy. Next generation sequencing was carried out by Edinburgh Genomics (HiSeq4000), The University of Edinburgh. Edinburgh Genomics 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)

Figure Legends

Fig. 1: Experimental set-up of the Hfq CLASH experiment at different growth conditions in *E. coli*.

(**a**) Overview of the critical experimental steps for obtaining the Hfq CLASH data and RNA-seq data. *E. coli* cells were grown, UV crosslinked and harvested at different cell-densities (OD_{600}), and the tagged Hfq-RNA complexes were purified under denaturing conditions. Hfq binds to sRNA-target RNA duplexes, and the RNA ends found in proximity can be ligated together and sequenced after removal of the protein. The single reads can be used to map Hfq-RNA interactions, whereas the chimeric reads can be traced to sRNA-target interactions. (**b**) A representative growth curve of the cultures used for the Hfq CLASH experiments, with

OD_{600} at which cells were crosslinked indicated by circles, and each growth stage is indicated above the plot. The same OD_{600} samples were analysed by RNA-seq.

(**c**) Western blot analyses of Hfq levels during the various growth stages. Hfq-HTF was detected using an anti-TAP antibody. GroEL was used as a loading control. Shown below is the quantification of the result. Levels of Hfq were normalized to that of GroEL.

(**d**) Hfq cross-linking 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.

Fig. 2: Growth-dependent sRNA expression and Hfq binding in *E. coli*.

(**a**) Heatmaps show the clustered, normalized Hfq cross-linking (Right; using the CLASH data) and RNA steady-states (Left: using the RNAseq data) of sRNAs relative to OD_{600} 0.4. The cluster assignment is indicated at the left of each plot. The clusters were generated by k-means clustering using the Log_2-fold changes of all genes with minimum 5 TPM (transcripts per million) at each OD_{600} relative to OD_{600} 0.4. A blue shade indicates a low fold-change, whereas a dark orange shade indicates a high fold-change relative to OD_{600} 0.4. The growth stage is indicated as OD_{600} units below each heatmap; data are representative of one biological replicate for each OD_{600}. Note that not all Hfq-bound sRNAs are mRNA-riboregulators: genes ssrS and ffs encode the cytoplasmic RNAs 6S, a regulator of RNA polymerase, and 4.5S, the signal recognition particle RNA.

(**b**) Comparison of changes in Hfq binding to changes in sRNA levels. Scatter plots comparing changes in Hfq binding (x-axis) to sRNA levels (y-axis) for the indicated OD_{600}. For each $OD_{600,}$ from 0.8 to 4.0, the TPM values at each density-point divided by the OD_{600} 0.4 data, and the resulting ratios were log_2-normalized. Data are representative of one biological replicate for each OD_{600}. The Pearson coefficient (r) calculated for sRNAs only, is indicated at the top left of each plot.

Fig. 3: Hfq CLASH detects RNA-RNA interactions in *E. coli*.

(**a**) Intermolecular transcript combinations found in interactions captured by Hfq CLASH. Combination count of all uniquely annotated hybrids on genomic features. *tRNA-tRNA and rRNA-rRNA chimeras originating from different coding regions were removed.

(**b**) Venn diagram showing the intersection between interactions from unfiltered CLASH data from two experiments, 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).

(**c**) (Left) Distribution of chimeras that uniquely map to mRNA genes. Overlapping fragments of the two individual parts of the chimeras were collapsed into single clusters followed by generation of distribution plots. To make these plots, each gene was divided in 100 bins the number of clusters that map to each bin (hit density; y-axis) was calculated; (Right) For the

distribution plot around the AUG, the gene length was normalized in 601 bins (x-axis) 5'-end overlap (-300) before the start of the coding sequence, and 300 bins downstream AUG (+300); the bins corresponding to the start codon are indicated with a pink line. (**D**) Enriched motifs in chimeras that uniquely overlap the 5'UTRs and 3'UTR; the logo was made as in Supplementary Fig.1 using K-mers with a Z-score > 2.

Fig. 4: Hfq CLASH uncovers known interactions and accurately identifies seed regions.
(**a**) Example of re-discovery of known sRNA-mRNA interactions by Hfq CLASH: Genome-browser plot showing the location and peaks of hybrid reads for ChiX (the counts on the y-axis correspond to all ChiX-containing hybrids recovered by Hfq CLASH); below plots for two of its known targets, *chiP* and *chbC* are shown (the hybrid read count are specific to ChiX-*chiP* and ChiX-*chbC* hybrids, respectively. The boxed arrows indicate the name of the feature and whether the gene is encoded on the Watson or Crick strand.
(**b**) Prediction of the base- pairing between ChiX and *chiP*, using the RNA sequences recovered by Hfq CLASH and RNACofold[98]. The SD sequence within the *chiP* mRNA is highlighted.
(**c**) Heatmaps that show the position of base-pairing and read density as $\log_2$(density = count per nucleotide position) for each interaction with ArcZ and RprA, with minimum two chimeras in the merged CLASH dataset. The boxes above indicate the known, experimentally validated seed sequences for the respective sRNAs.
(**d**) Distribution of the 306 known RNA-RNA interactions retrieved from sRNATarbase3 and RILseq log and stationary phase S-chimera datasets in the Hfq CLASH dataset. The interactions were pooled into bins (x-axis) that represent chimera count intervals per interaction. The y-axis shows the number of interactions in each bin.

Fig. 5: sRNA-RNA interactions identified by CLASH are growth-stage specific.
(**a**) sRNA-sRNA network generated from the unfiltered CLASH data from two experiments, 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). Only sRNAs that are independently transcribed were included in the analysis.
(**b**) Heatmaps that show the position of base-pairing and read density as $\log_2$(unique chimera count) for each interaction within ArcZ and CyaR, with minimum two counts in the merged CLASH dataset. (Top) ArcZ heatmap that focuses on sRNA-sRNA interactions. The box above indicates the known, experimentally validated seed sequence. (Bottom) CyaR heatmaps that show all CyaR chimeras and CyaR-sRNA chimeras, respectively. The box

above the heatmaps indicates the known, experimentally validated seed sequence, as well as a new seed identified by CLASH.

(**c**) Hfq CLASH greatly enriches sRNA-sRNA networks: comparison between sRNA-sRNA interactions supported by minimum two chimeras in the Hfq CLASH data, RIL-seq S-chimeras[18] (log and stationary) and RNase E CLASH[13]. Only sRNAs with independent transcription units and annotated in the core genomes of all datasets were considered. The networks below the Venn diagram represent sRNA-sRNA interactions found in multiple datasets.

(**d**) Small RNA sponging interactions coordinate nutritional stress responses. ArcZ, CyaR were overexpressed and the levels of their targets were monitored by RT-qPCR. The *tpx* and *sdaC* mRNAs are ArcZ mRNA targets. The *nadE* and *yqaE* mRNAs are CyaR targets. The *dppA* mRNA is a GcvB target. Over expression of ArcZ downregulates CyaR (panel a) but not the other way around (panel b). CyaR downregulates GcvB and GcvB can act as a sponge for CyaR (panels c-d). Experiments were performed in technical triplicates; the standard error of the mean (SEM) of three biological replicates are reported as error bars. The dashed horizontal line indicates the fold-change corresponding to the scrambled RNA overexpression control (pJV300).

Fig. 6: 3'UTR containing chimeras may uncover novel 3'UTR-derived sRNAs.

(**a**) The names of protein-coding genes that are part of 3'UTR-mRNA chimeras, and uniquely map to 3'UTRs, were selected from the CLASH data and RIL-seq data. The RIL-seq RNA-RNA interaction set [18] S-chimeras for Log and Stationary phases of growth was filtered for the 3'UTR/ EST3UTR annotations on either orientation of the mRNA-mRNA pairs. Both were intersected with the set of mRNAs that were predicted by TIER-seq studies[31] to harbour sRNAs that get released from 3'UTRs by RNase E processing.

(**b** Northern blot validation of several 3'UTR-derived sRNAs candidates, including *malG*, *gadE* and *ygaM* 3'UTRs. RybB was probed as a sRNA positive control and 5S rRNA as the loading control.

(**c**) Genome-browser snapshots of several regions containing candidate sRNAs for optical densities at which the RNA steady-state was maximal for each candidate; the candidate names and OD_{600} are indicated at the left side of the y-axes; the y-axis shows the normalized reads (RPM: reads per million); red: RNA steady states from the RNA-seq experiment, blue: Hfq cross-linking from the CLASH experiment; black: unique chimeric reads found in this region.

Fig. 7: MdoR is a degradation product that emerges at the transition between exponential and stationary phases of growth.

(**a**) The *mal* regulon gene expression peaks at the transition between exponential and stationary phases of growth: the plot shows RNA steady-states (RNA-seq experiment) normalized as log_2TPM (transcripts per million, y-axis); the x-axis indicates the growth-stage as optical density units (OD_{600}).

(**b**) Northern blot using total RNA from *E. coli* harvested at different cell densities (OD_{600}) probed with an oligo antisense to *malG* 3'UTR; 23 S rRNA was used as the loading control; the identity of the bands is indicated at the left of the panel; horizontal bars indicate *malEFG* degradation intermediates.

(**c**) MdoR is a degradation product: Northern blot using total RNA from cells at indicated OD_{600} with (+; lanes 4 and 4) or without (-; lanes 1 and 3) 5'-Phosphate-Dependent Exonuclease treatment; the absence of a band indicates 5'-P Exo-mediated degradation; the top panel shows the Northern probed for MdoR; RybB was probed as a negative control; RaiZ as a positive control and 5S rRNA was probed as a loading control. The text on the right of the blot indicates the phosphorylation state of the 5'-termini for each sRNA.

Fig. 8: MdoR directly destabilizes *ompC* and *ompA* mRNAs and downregulates key members the σ^E regulon.

(**a**) Genome browser plots showing the position of MdoR-containing chimeras, indicated by the CLASH data for MdoR-*ompC,* MdoR-*ompA* and MdoR-OhsC interactions. The counts are cumulative of all hybrid reads that MdoR is part of.

(**b**, **c**, **d**) MdoR forms stable duplexes with the 5'UTR of porin-encoding mRNAs and a non-coding RNA, OhsC. *In silico* prediction (RNACofold[98]) of hybrid structures derived from the most abundant MdoR chimeric reads with the indicated transcripts. The min. MFE is the minimum of the MFEs assigned to chimeras representing an interaction using our pyCalculateMFE.py python script.

(**e**) Pulse-overexpression of MdoR in wild-type *E. coli*: MdoR expression was induced with L-arabinose from pBAD::MdoR, whereas the pBAD Control served as an empty-plasmid control not expressing a sRNA; the 15-minute induction time was optimal and used for pulse-overexpression followed by RNA-seq.

(**f**) Differential expression analysis of the Control and MdoR overexpression data (DESeq2 with data from three replicates): the red points indicate the transcripts differentially expressed in the data: transcripts with a log_2fold change > 0 were enriched in the Control data; those with a log_2-fold change < 0 were enriched in the MdoR overexpression data. The annotated, enlarged red data points indicate the differentially expressed transcripts discussed in the text.

Fig. 9: Cumulative evidence for MdoR role in modulating expression of key players of the pathways involved in optimisation/finetuning of maltose intake.

(**a**). An MdoR seed-mutant (MdoR SM) is unable to downregulate targets identified by CLASH and differential expression analysis. RT-qPCR analysis of several differentially expressed transcripts (gene names shown at the top of the plots), in the presence of sRNAs expressed for 15 minutes from plasmid-borne arabinose-inducible promoters; sRNA names are indicated below each bar; the 'no sRNA' sample contains the empty plasmid as reference for fold-change calculations; *recA* was used as the internal reference gene; experiments were performed in technical triplicates; the standard error of the mean (SEM) of two biological replicates are reported as error bars.
(**b**) MdoR directly downregulates *ompC* mRNA through base-pairing interactions: RT-qPCR analyses of the *ompC-sfGFP* and *ompC SM-sfGFP* fusions expression in the presence of MdoR. Mutating either the sRNA or target seed to the complementary sequence partially disrupts the regulation, while expressing both seed mutants simultaneously partially restores regulation. The bars indicate the mean fold-change in expression relative to the no sRNA Control (horizontal dashed line); error bars indicate the standard error of the mean, n=2; the significance of the differences between the WT and the MdoR SM was assessed with a two-tailed Student's t-test.
(**c**) Expression of MicA, *lamB* and *ompC* in a MdoR seed mutant strain. MdoR stems I and II were mutated to the complementary sequence in the chromosomal context. Mutant and WT cells were grown overnight in glucose, and next day cells were transferred to media containing maltose or glucose only, respectively. Cells were harvested at OD_{600} 0.5 and RT-qPCR performed on total RNA, and analysed as in A and B using the *rrfD* gene as the internal control. The bars indicate the mean fold-change in expression relative to the WT; error bars indicate the standard error of the mean, n=2; the significance of the differences between the MdoR mutant and WT (set as the reference sample, indicated on the plot with a horizontal dashed line) was assessed with a one-sample t-test.

Fig. 10: An sRNA-target network that regulates adaptation to changes in nutrient availability. (**a**) Hypothetical model of MdoR biogenesis and regulatory role. MdoR increases efficiency of maltose/maltodextrin uptake, thus helping *E. coli* adapt to maltose/maltodextrin utilising conditions. During maltodextrin (grey circles) utilization, the *malEFG* operon, encoding IM components for maltose uptake is transcribed by the housekeeping sigma factor (σ^{70}) and MalT-inducer (and Crp-cAMP). A fraction of the *mallEFG* transcript pool is constitutively degraded by RNase E, generating MdoR, a sRNA derived from the 3'UTR of *malG*. MdoR is stabilised by the RNA chaperone Hfq that protects it from complete ribonucleolytic cleavage. MdoR base-pairs with *ompC* and *ompA* mRNAs, leading to their decay. Additionally, MdoR strongly downregulates expression of MicA, a sRNA repressor of LamB synthesis. Therefore, as part of a feed-forward loop, MdoR promotes accumulation of maltose specific porins,

thereby increasing the efficiency of maltodextrin uptake. The changes in envelope composition caused by increased production of LamB exert additional pressure on the OM assembly machinery, that induce extracytoplasmic stress. This type of stress is tackled by σ^E activation, that controls the expression of stress-responsive genes and most of the assembly and insertion factors. But σ^E activation is not desirable in these conditions for two reasons: it's unbalanced activation has precarious effects on the cell, and it's activation would produce high amounts of MicA, that would ultimately repress LamB production. For these reasons, MdoR regulation includes a regulatory arm that mitigates excessive σ^E activation through: MdoR repression of *rpoE* expression, and inhibition of σ^E activation. Decay of *omp* mRNAs likely relieves envelope stress.

(**b**) sRNA-target networks regulating peptide/amino acids and maltose/maltodextrin uptake. Italicized names are mRNA targets.

References

1. Balleza, E. *et al.* Regulation by transcription factors in bacteria: Beyond description. *FEMS Microbiol. Rev.* **33**, 133–151 (2009).
2. Martínez-Antonio, A., Janga, S. C. & Thieffry, D. Functional organisation of Escherichia coli transcriptional regulatory network. *J. Mol. Biol.* **381**, 238–247 (2008).
3. Waters, L. S. & Storz, G. Regulatory RNAs in Bacteria. *Cell* **136**, 615–628 (2009).
4. Kortmann, J. & Narberhaus, F. Bacterial RNA thermometers: molecular zippers and switches. *Nat. Rev. Microbiol.* **10**, 255–265 (2012).
5. Arraiano, C. M. *et al.* The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol. Rev.* **34**, 883–923 (2010).
6. Storz, G., Vogel, J. & Wassarman, K. M. Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell* **43**, 880–891 (2011).
7. Updegrove, T. B., Zhang, A. & Storz, G. Hfq: The flexible RNA matchmaker. *Curr. Opin. Microbiol.* **30**, 133–138 (2016).
8. Smirnov, A., Wang, C., Drewry, L. L. & Vogel, J. Molecular mechanism of mRNA repression in *trans* by a ProQ-dependent small RNA. *EMBO J.* **36**, 1029–1045 (2017).
9. Smirnov, A. *et al.* Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl. Acad. Sci.* **113**, 11591–11596 (2016).
10. Sedlyarova, N. *et al.* sRNA-Mediated Control of Transcription Termination in E.~coli. *Cell* **167**, 111--121.e13 (2016).
11. Vogel, J. & Luisi, B. F. Hfq and its constellation of RNA. *Nat. Rev. Microbiol.* **9**, 578–589 (2011).
12. Mandin, P. & Guillier, M. Expanding control in bacteria: Interplay between small RNAs and transcriptional regulators to control gene expression. *Curr. Opin. Microbiol.* **16**, 125–132 (2013).
13. Waters, S. A. *et al.* Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *EMBO J.* **36**, 374–387 (2017).
14. Hör, J. & Vogel, J. Global snapshots of bacterial RNA networks. *EMBO J.* **36**, 245–247 (2017).
15. Hör, J., Gorski, S. A. & Vogel, J. Bacterial RNA Biology on a Genome Scale. *Mol. Cell* 785–799 (2018). doi:10.1016/j.molcel.2017.12.023
16. Han, K., Tjaden, B. & Lory, S. GRIL-seq provides a method for identifying direct targets of bacterial small regulatory RNA by in vivo proximity ligation. *Nat. Microbiol.* **2**, 16239 (2016).

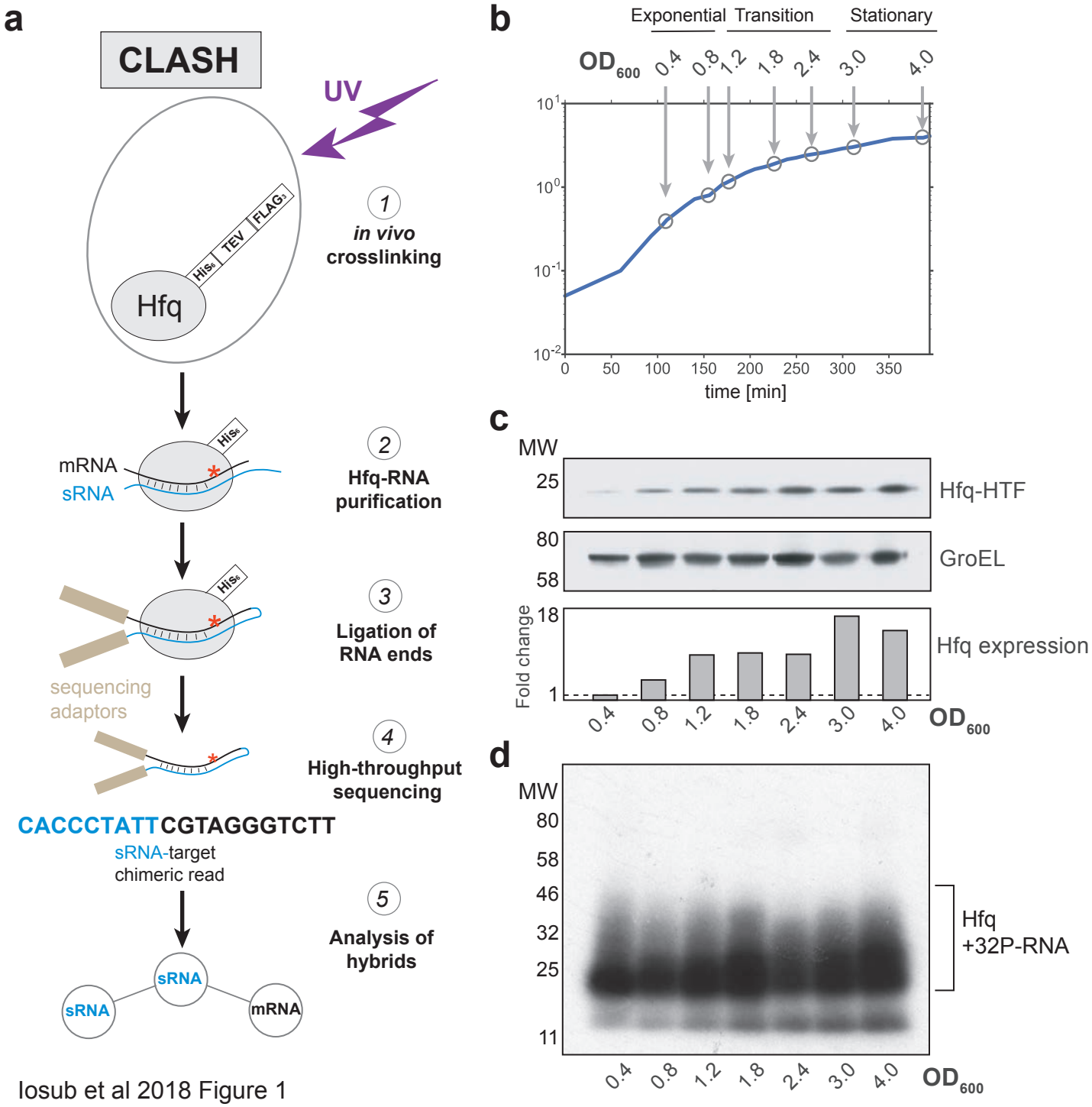
17. Lalaouna, D. *et al.* A 3{\textasciiacute} External Transcribed Spacer in a tRNA Transcript Acts as a Sponge for Small RNAs to Prevent Transcriptional Noise. *Mol. Cell* **58**, 393–405 (2015).
18. Melamed, S. *et al.* Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol. Cell* **63**, 884–897 (2016).
19. Mouali, Y. El *et al.* CRP-cAMP mediates silencing of Salmonella virulence at the post-transcriptional level. 1–26 (2018). doi:10.1371/journal.pgen.1007401
20. Mäder, U. *et al.* Staphylococcus aureus Transcriptome Architecture: From Laboratory to Infection-Mimicking Conditions. *PLoS Genet.* **12**, e1005962 (2016).
21. Pletnev, P., Osterman, I., Sergiev, P., Bogdanov, A. & Dontsova, O. Survival guide: Escherichia coli in the stationary phase. *Acta Naturae* **7**, 22–33 (2015).
22. Navarro Llorens, J. M., Tormo, A. & Martínez-García, E. Stationary phase in gram-negative bacteria. *FEMS Microbiol. Rev.* **34**, 476–495 (2010).
23. Baev, M. V., Baev, D., Jansco Radek, A. & Campbell, J. W. Growth of Escherichia coli MG1655 on LB medium: Monitoring utilization of amino acids, peptides, and nucleotides with transcriptional microarrays. *Appl. Microbiol. Biotechnol.* **71**, 317–322 (2006).
24. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. Escherichia coli physiology in Luria-Bertani broth. *J. Bacteriol.* **189**, 8746–8749 (2007).
25. Baev, M. V., Baev, D., Radek, A. J. & Campbell, J. W. Growth of Escherichia coli MG1655 on LB medium: Determining metabolic strategy with transcriptional microarrays. *Appl. Microbiol. Biotechnol.* **71**, 323–328 (2006).
26. Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C. M. & Vogel, J. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J.* **31**, 4005–4019 (2012).
27. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* (2013). doi:10.1016/j.cell.2013.03.043
28. Kudla, G., Granneman, S., Hahn, D., Beggs, J. D. & Tollervey, D. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10010–10015 (2011).
29. Chao, Y. & Vogel, J. A 3' UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress Response. *Mol. Cell* **61**, 352–363 (2016).
30. Miyakoshi, M., Chao, Y. & Vogel, J. Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Current Opinion in Microbiology* (2015). doi:10.1016/j.mib.2015.01.013
31. Chao, Y. *et al. in vivo* cleavage map illuminates the central cole of RNase E in coding and non-coding RNA pathways. *Mol. Cell* **65**, 39–51 (2017).
32. Van Nues, R. *et al.* Kinetic CRAC uncovers a role for Nab3 in determining gene expression profiles during stress. *Nat. Commun.* (2017). doi:10.1038/s41467-017-00025-5
33. Wilson, K. S. & Hippel, P. H. Von. Transcription termination at intrinsic terminators: The role of the RNA hairpin (Escherichia coli/RNA polymerase/rho-independent termination). *Biochemistry* **92**, 8793–8797 (1995).
34. Holmqvist, E. *et al.* Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in~vivo. *EMBO J.* **35**, e201593360--1011 (2016).
35. Travis, A. J., Moody, J., Helwak, A., Tollervey, D. & Kudla, G. Hyb: A bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and sequencing of hybrids) data. *Methods* **65**, 263–273 (2014).
36. Bouvier, M., Sharma, C. M., Mika, F., Nierhaus, K. H. & Vogel, J. Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol. Cell* **32**, 827–837 (2008).
37. Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D. & Gally, D. L. Identification

of bacteriophage-encoded anti-sRNAs in pathogenic Escherichia coli. *Mol. Cell* **55**, 199–213 (2014).
38. Rasmussen, A. A. *et al.* A conserved small RNA promotes silencing of the outer membrane protein YbfM. *Mol. Microbiol.* **72**, 566–577 (2009).
39. Figueroa-Bossi, N., Valentini, M., Malleret, L. & Bossi, L. Caught at its own game: Regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes Dev.* (2009). doi:10.1101/gad.541609
40. Overgaard, M., Johansen, J., Møller-Jensen, J. & Valentin-Hansen, P. Switching off small RNA regulation with trap-mRNA. *Mol. Microbiol.* (2009). doi:10.1111/j.1365-2958.2009.06807.x
41. Plumbridge, J., Bossi, L., Oberto, J., Wade, J. T. & Figueroa-Bossi, N. Interplay of transcriptional and small RNA-dependent control mechanisms regulates chitosugar uptake in Escherichia coli and Salmonella. *Mol. Microbiol.* **92**, 648–658 (2014).
42. Papenfort, K. *et al.* Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol. Microbiol.* **74**, 139–158 (2009).
43. Mika, F. *et al.* Targeting of csgD by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in Escherichia coli. *Mol. Microbiol.* **84**, 51–65 (2012).
44. Wang, J. *et al.* sRNATarBase 3.0: An updated database for sRNA-target interactions in bacteria. *Nucleic Acids Res.* **44**, D248–D253 (2016).
45. De Lay, N. & Gottesman, S. The crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *J. Bacteriol.* (2009). doi:10.1128/JB.01157-08
46. Sharma, C. M., Darfeuille, F., Plantinga, T. H. & Vogel, J. A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes Dev.* (2007). doi:10.1101/gad.447207
47. Apirion, D. Isolation, genetic mapping and some characterization of a mutation in Escherichia coli that affects the processing of ribonucleic acid. *Genetics* **90**, 659–671 (1978).
48. Papenfort, K. *et al.* σE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay. *Mol. Microbiol.* **62**, 1674–1688 (2006).
49. Johansen, J., Rasmussen, A. A., Overgaard, M. & Valentin-Hansen, P. Conserved small non-coding RNAs that belong to the sigmaE regulon: role in down-regulation of outer membrane proteins. *J. Mol. Biol.* **364**, 1–8 (2006).
50. Newbury, S. F., Smith, N. H. & Higgins, C. F. Differential mRNA stability controls relative gene expression within a polycistronic operon. *Cell* **51**, 1131–1143 (1987).
51. Khemici, V. & Carpousis, A. J. The RNA degradosome and poly(A) polymerase of Escherichia coli are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol. Microbiol.* (2004). doi:10.1046/j.1365-2958.2003.03862.x
52. Wright, P. R. *et al.* CopraRNA and IntaRNA: Predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gku359
53. Wright, P. R. *et al.* Comparative genomics boosts target prediction for bacterial small RNAs. *Proc. Natl. Acad. Sci.* **110**, E3487–E3496 (2013).
54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
55. Bossi, L., Maloriol, D. & Figueroa-Bossi, N. Porin biogenesis activates the σE response in Salmonella hfq mutants. *Biochimie* **90**, 1539–1544 (2008).
56. De Las Peñas, A., Connolly, L. & Gross, C. A. The sigmaE-mediated response to extracytoplasmic stress in Escherichia coli is transduced by RseA and RseB, two negative regulators of sigmaE. *Mol. Microbiol.* **24**, 373–385 (1997).
57. Alba, B. M. & Gross, C. A. Regulation of the Escherichia coli ??E-dependent envelope stress response. *Mol. Microbiol.* **52**, 613–619 (2004).
58. Rhodius, V. A., Suh, W. C., Nonaka, G., West, J. & Gross, C. A. Conserved and

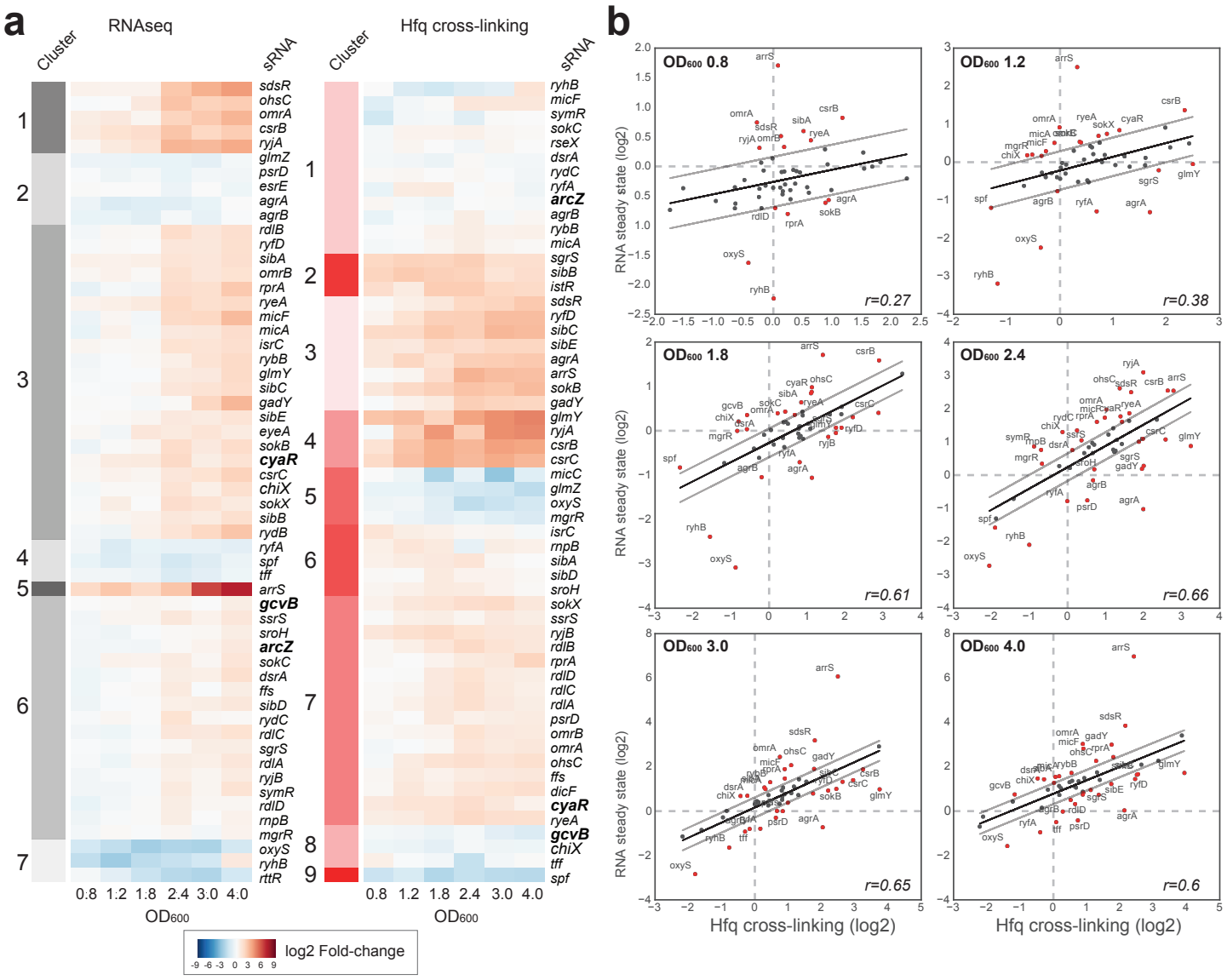
variable functions of the σE stress response in related genomes. *PLoS Biol.* **4**, 0043–0059 (2006).
59. Udekwu, K. I. & Wagner, E. G. H. Sigma E controls biogenesis of the antisense RNA MicA. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkl1154
60. Bossi, L. & Figueroa-Bossi, N. A small RNA downregulates LamB maltoporin in Salmonella. *Mol. Microbiol.* **65**, 799–810 (2007).
61. Thompson, K. M., Rhodius, V. A. & Gottesman, S. σE Regulates and Is Regulated by a Small RNA in Escherichia coli.pdf. *J Bacteriol* **189**, 4243–4256 (2007).
62. Gogol, E. B., Rhodius, V. A., Papenfort, K., Vogel, J. & Gross, C. A. Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 12875–12880 (2011).
63. Urban, J. H. & Vogel, J. Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res.* **35**, 1018–1037 (2007).
64. Corcoran, C. P. *et al.* Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Mol. Microbiol.* **84**, 428–445 (2012).
65. Chen, S., Zhang, A., Blyn, L. B. & Storz, G. MicC, a second small-RNA regulator of omp protein expression in Escherichia coli. *J. Bacteriol.* (2004). doi:10.1128/JB.186.20.6689-6697.2004
66. Boos, W. & Shuman, H. Maltose / Maltodextrin System of Escherichia coli : Transport , Metabolism , and Regulation Maltose / Maltodextrin System of Escherichia coli : Transport , Metabolism , and Regulation. *Microbiol. Mol. Biol. Rev.* **62**, 204–229 (1998).
67. Shimoni, Y. *et al.* Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol. Syst. Biol.* **3**, 138 (2007).
68. Holmqvist, E. & Wagner, E. G. H. Impact of bacterial sRNAs in stress responses. *Biochem. Soc. Trans.* (2017). doi:10.1042/BST20160363
69. Nitzan, M., Rehani, R. & Margalit, H. Integration of Bacterial Small RNAs in Regulatory Networks. *Annu. Rev. Biophys.* **46**, 131–148 (2017).
70. Granneman, S., Kudla, G., Petfalski, E. & Tollervey, D. Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high-throughput analysis of cDNAs. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9613–9618 (2009).
71. Bruce, H. A. *et al.* Analysis of the natively unstructured RNA/protein-recognition core in the Escherichia coli RNA degradosome and its interactions with regulatory RNA/Hfq complexes. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gkx1083
72. Bandyra, K. J. *et al.* The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. *Mol. Cell* **47**, 943–953 (2012).
73. Morita, T., Maki, K. & Aiba, H. RNase E-based ribonucleoprotein complexes: Mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.* (2005). doi:10.1101/gad.1330405
74. Oliva, G., Sahr, T., Rolando, M., Knoth, M. & Buchrieser, C. A unique cis-encoded small noncoding RNA is regulating Legionella pneumophila Hfq expression in a life cycle-dependent manner. *MBio* (2017). doi:10.1128/mBio.02182-16
75. Sonnleitner, E., Schuster, M., Sorger-Domenigg, T., Greenberg, E. P. & Bläsi, U. Hfq-dependent alterations of the transcriptome profile and effects on quorum sensing in Pseudomonas aeruginosa. *Mol. Microbiol.* (2006). doi:10.1111/j.1365-2958.2006.05032.x
76. van Nues, R. W., Castro-Roa, D., Yuzenkova, Y. & Zenkin, N. 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 (2016).
77. Holmqvist, E., Li, L., Bischler, T., Barquist, L. & Vogel, J. Global Maps of ProQ Binding In Vivo Reveal Target Recognition via RNA Structure and Stability Control at mRNA 3′ Ends. *Mol. Cell* (2018). doi:10.1016/j.molcel.2018.04.017

78. Potts, A. H. *et al.* Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics. *Nat. Commun.* (2017). doi:10.1038/s41467-017-01613-1
79. Miyakoshi, M., Chao, Y. & Vogel, J. Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J.* **34**, e201490546--1492 (2015).
80. Eisenhardt, K. M. H., Reuscher, C. M. & Klug, G. PcrX, an sRNA derived from the 3′-UTR of the Rhodobacter sphaeroides puf operon modulates expression of puf genes encoding proteins of the bacterial photosynthetic apparatus. *Molecular Microbiology* (2018). doi:10.1111/mmi.14076
81. Peng, T. *et al.* Regulation of a polyamine transporter by the conserved 3′ UTR-derived sRNA SorX confers resistance to singlet oxygen and organic hydroperoxides in Rhodobacter sphaeroides. *RNA Biol.* (2016). doi:10.1080/15476286.2016.1212152
82. Kenyon, W. J., Thomas, S. M., Johnson, E., Pallen, M. J. & Spector, M. P. Shifts from glucose to certain secondary carbon-sources result in activation of the extracytoplasmic function sigma factor $σ$E in Salmonella enterica serovar Typhimurium. *Microbiology* **151**, 2373–2383 (2005).
83. Guo, M. S. *et al.* MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev.* **28**, 1620–1634 (2014).
84. Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G. & Gottesman, S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* (2001). doi:10.1101/gad.901001
85. Monteiro, C. *et al.* Hfq and Hfq-dependent small RNAs are major contributors to multicellular development in Salmonella enterica serovar typhimurium. *RNA Biol.* (2012). doi:10.4161/rna.19682
86. Sittka, A., Pfeiffer, V., Tedin, K. & Vogel, J. The RNA chaperone Hfq is essential for the virulence of Salmonella typhimurium. *Mol. Microbiol.* (2007). doi:10.1111/j.1365-2958.2006.05489.x
87. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–5 (2000).
88. St-Pierre, F. *et al.* One-step cloning and chromosomal integration of DNA. *ACS Synth. Biol.* (2013). doi:10.1021/sb400021j
89. Bernabò, P. *et al.* In Vivo Translatome Profiling in Spinal Muscular Atrophy Reveals a Role for SMN Protein in Ribosome Biology. *Cell Rep.* (2017). doi:10.1016/j.celrep.2017.10.010
90. Lunelli, L. *et al.* Peering at Brain Polysomes with Atomic Force Microscopy. *J. Vis. Exp.* (2016). doi:10.3791/53851
91. Tebaldi, T. *et al.* Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. *BMC Genomics* (2012). doi:10.1186/1471-2164-13-220
92. Webb, S., Hector, R. D., Kudla, G. & Granneman, S. PAR-CLIP data indicate that Nrd1-Nab3-dependent transcription termination regulates expression of hundreds of protein coding genes in yeast. *Genome Biol.* **15**, R8 (2014).
93. Tjaden, B. De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol.* **16**, 1–10 (2015).
94. Wu, X. & Bartel, D. P. KpLogo: Positional k -mer analysis reveals hidden specificity in biological sequences. *Nucleic Acids Res.* (2017). doi:10.1093/nar/gkx323
95. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* (2011). doi:10.1093/bioinformatics/btr507
96. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
97. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench.

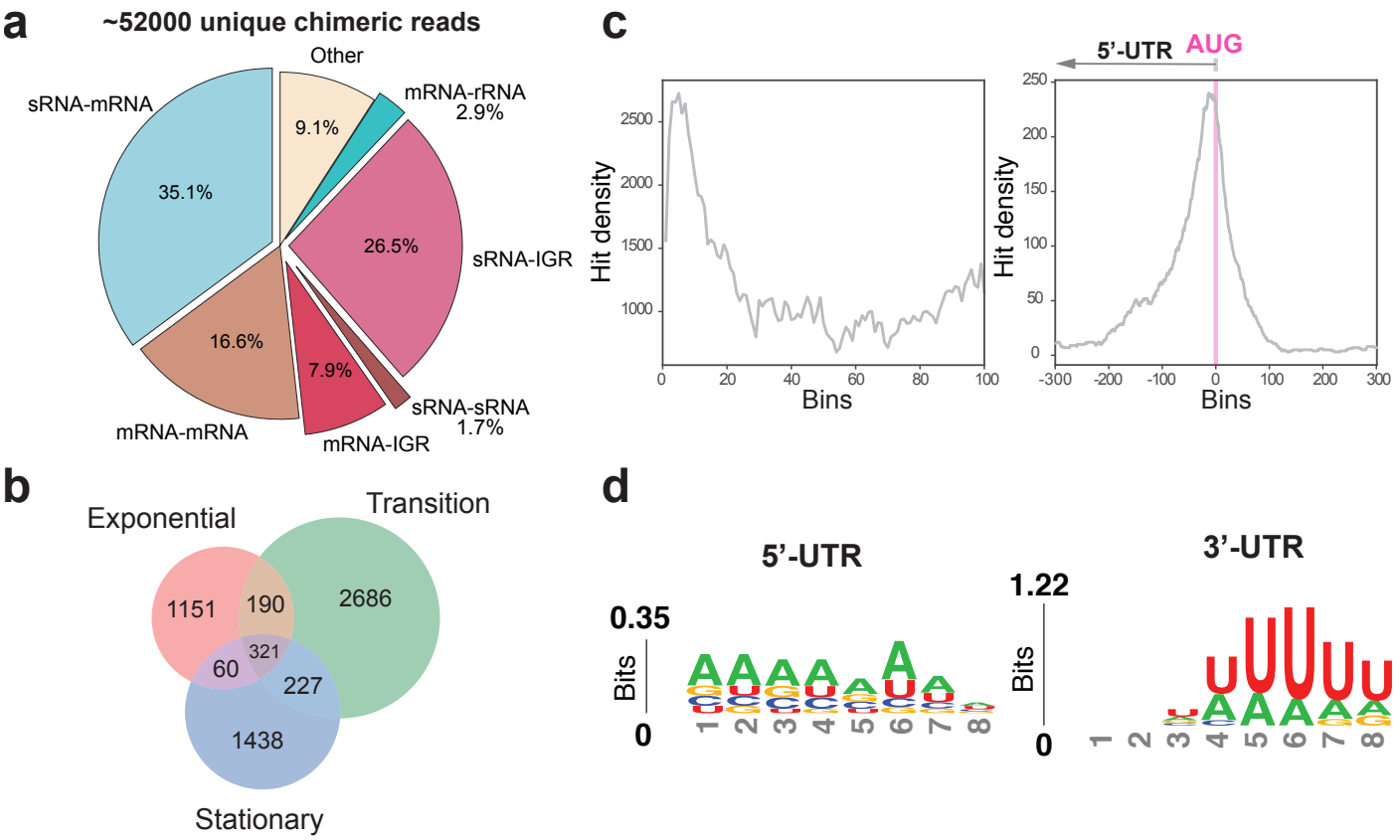
Bioinformatics **25**, 1189–1191 (2009).
98. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).



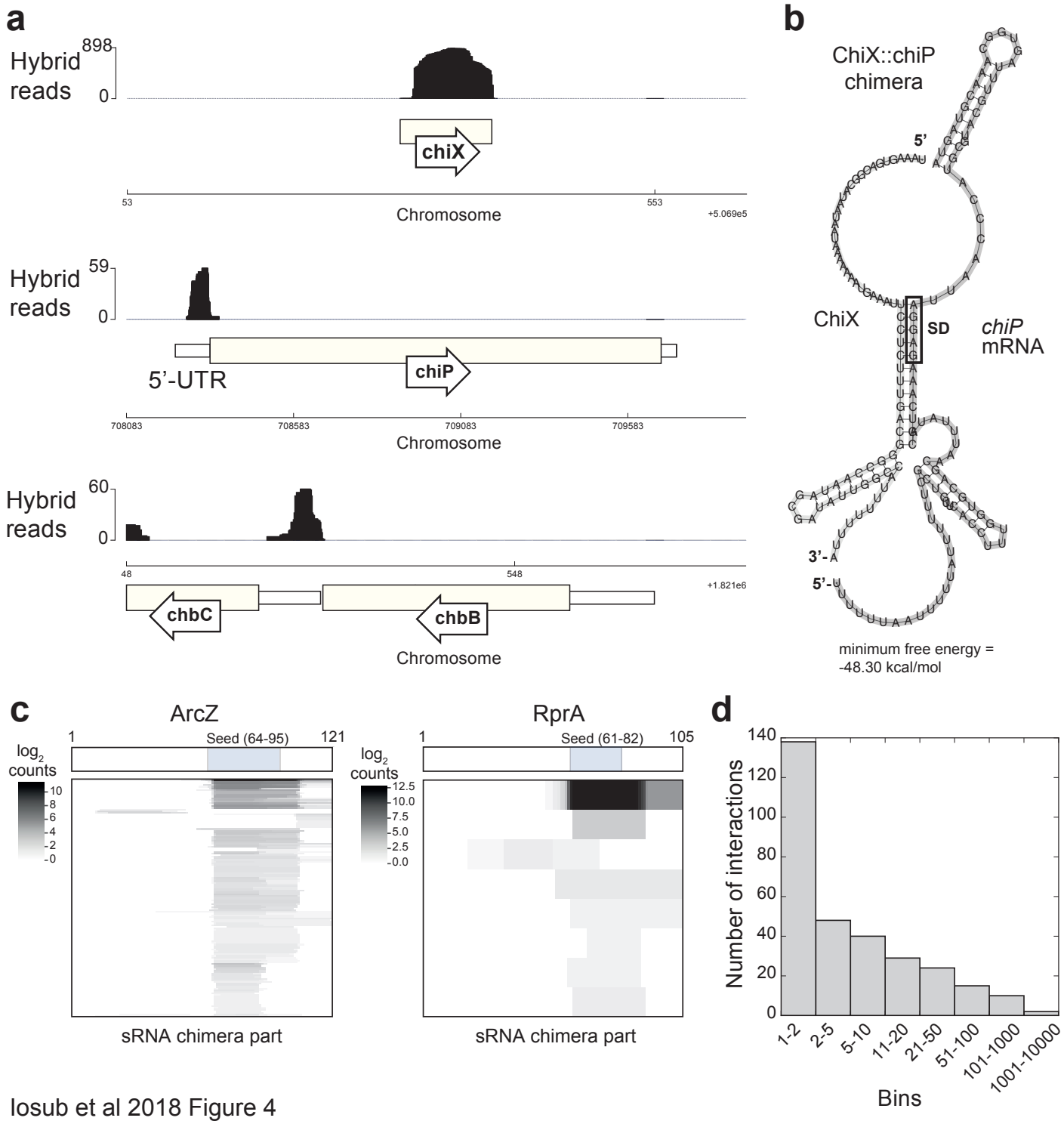
Iosub et al 2018 Figure 1



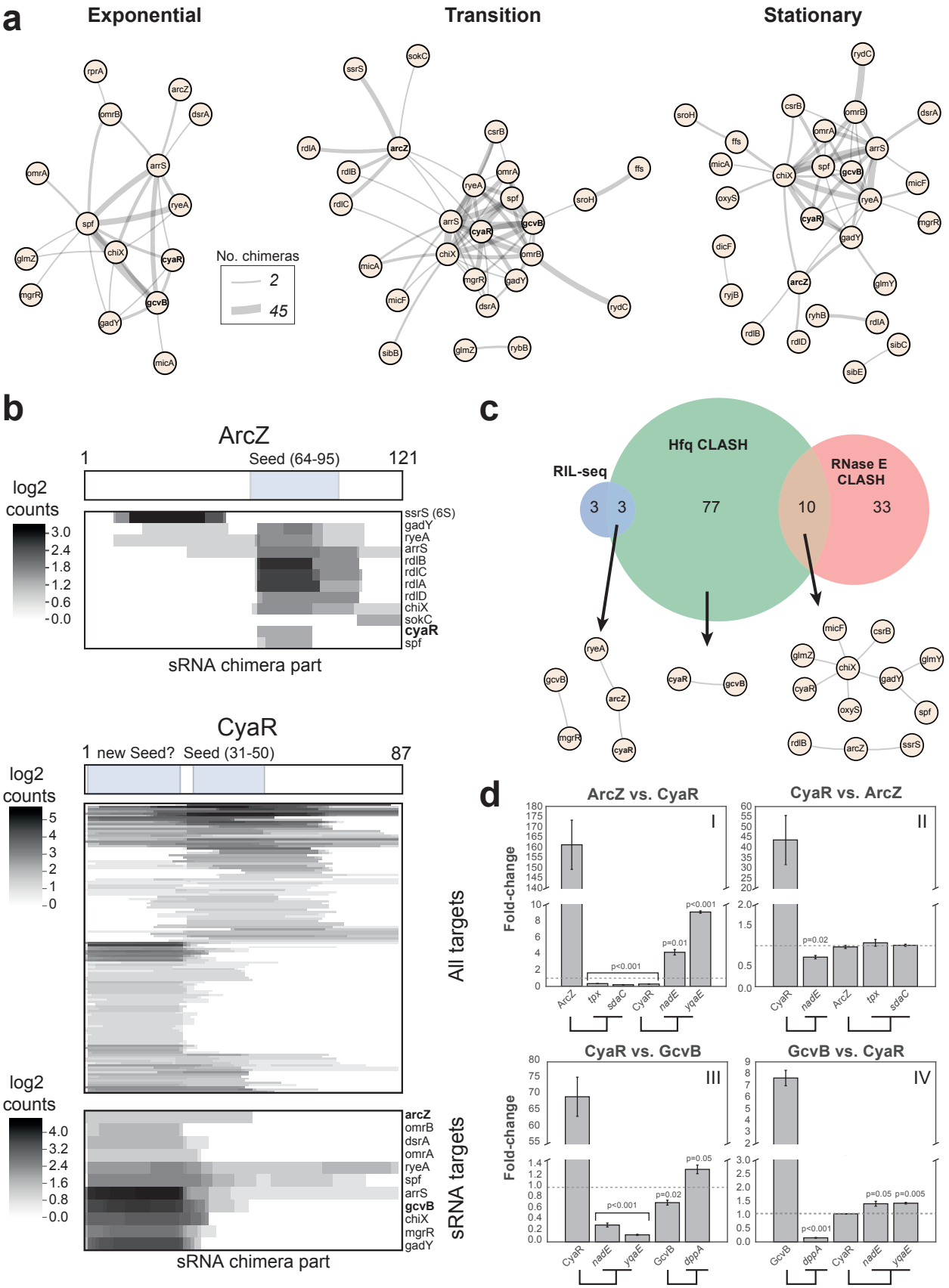
Iosub et al 2018 Figure 2



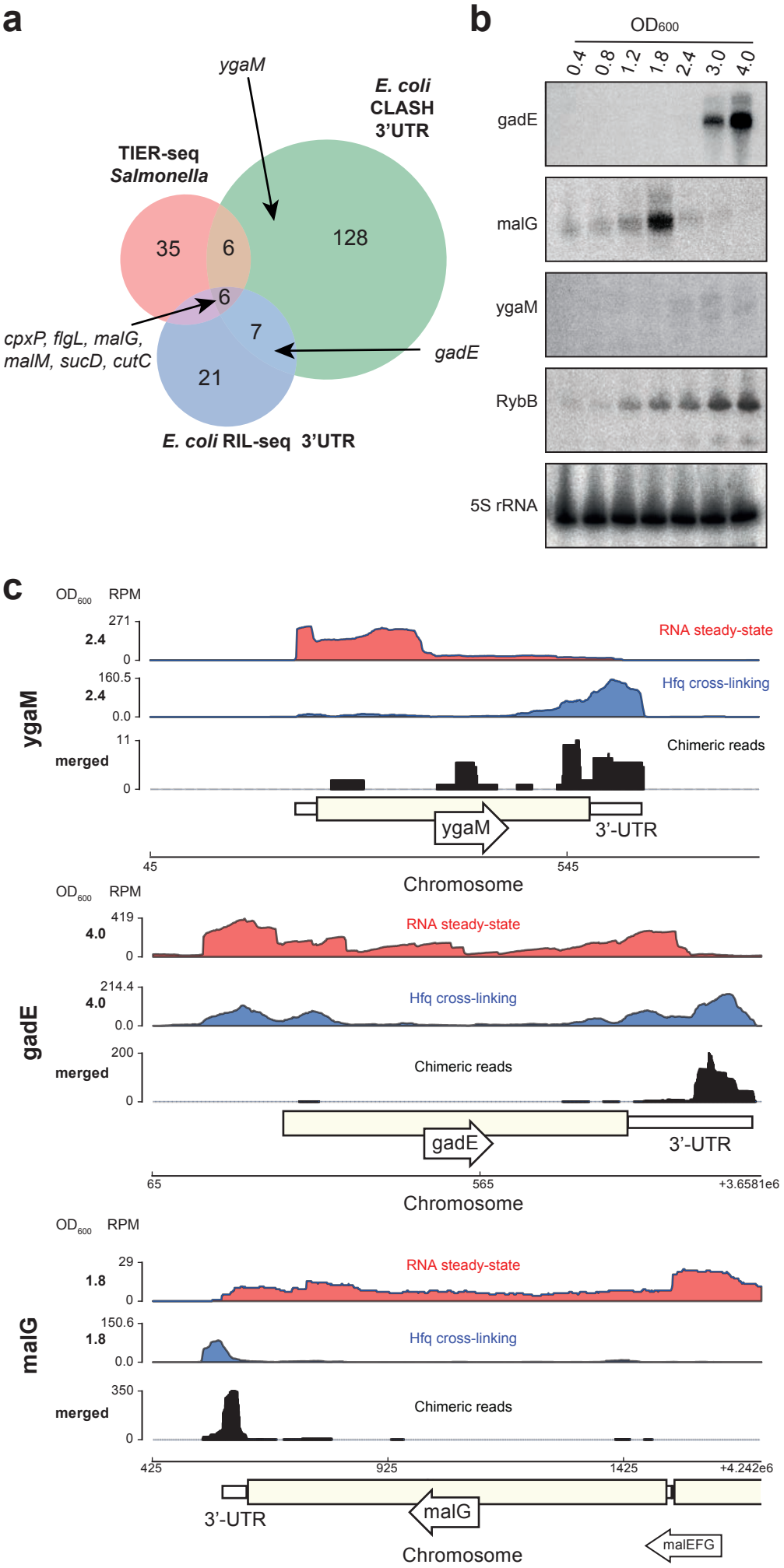
Iosub et al 2018 Figure 3



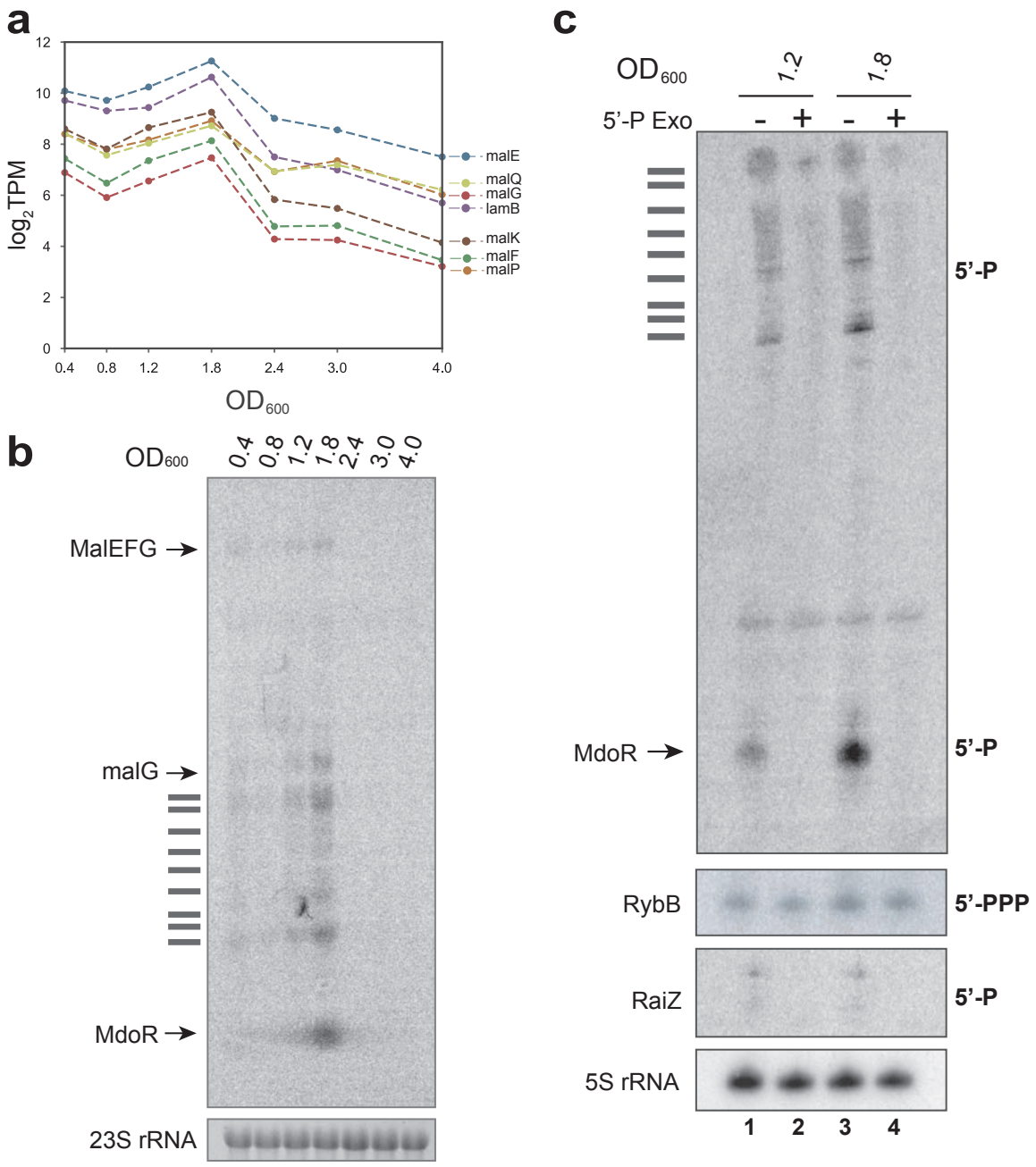
Iosub et al 2018 Figure 4



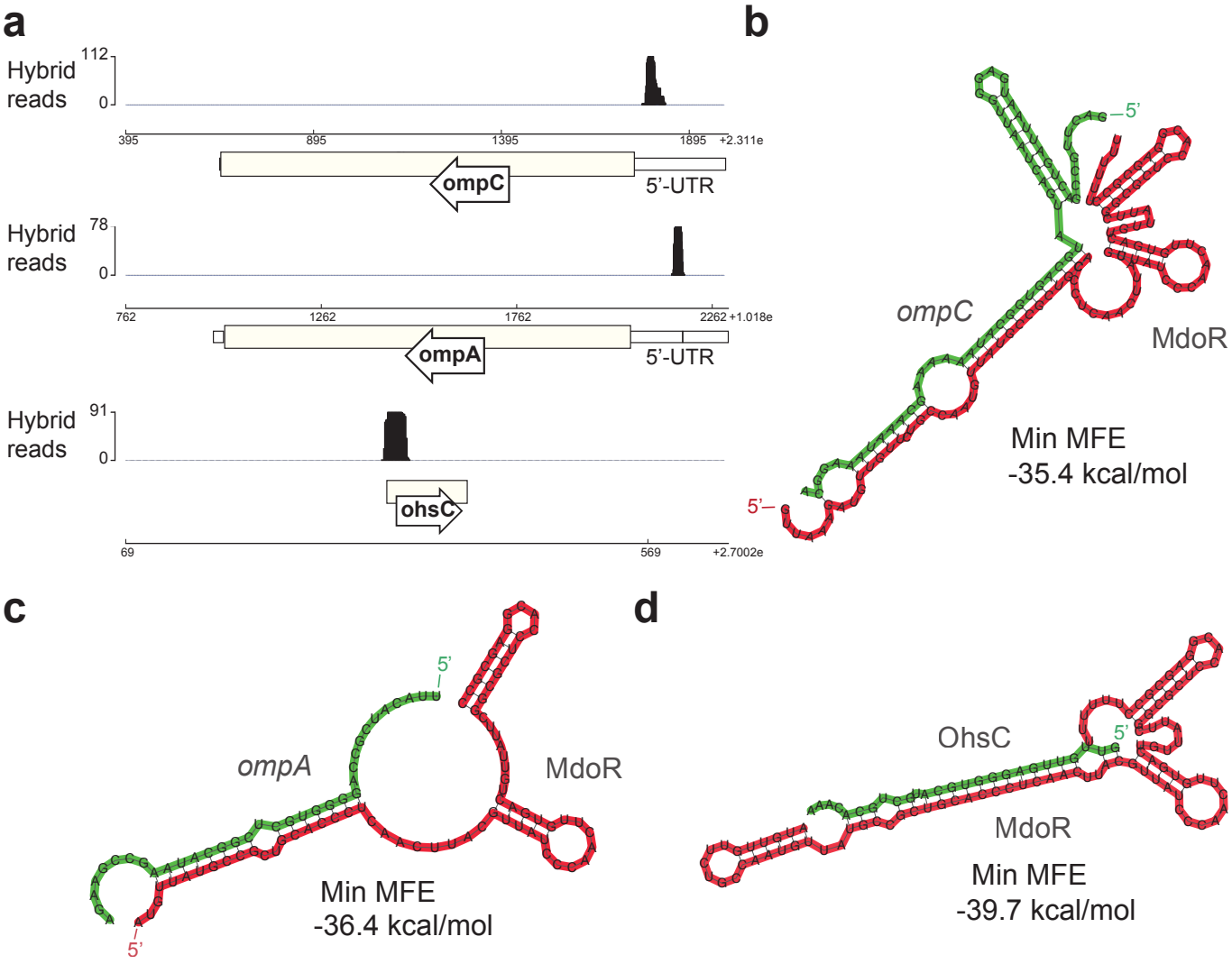
Iosub et al 2018 Figure 5

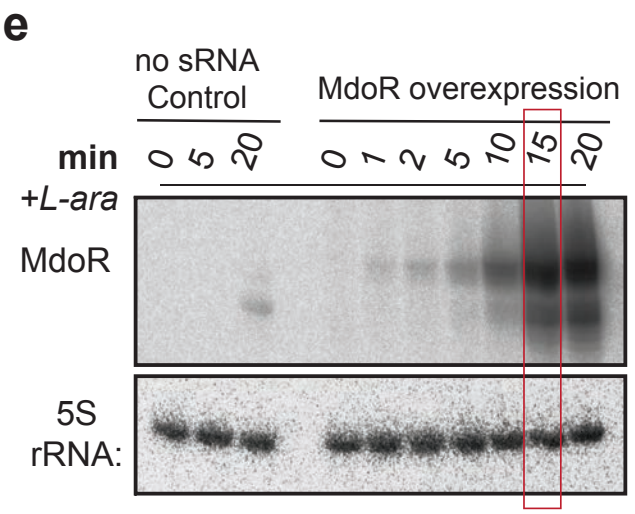


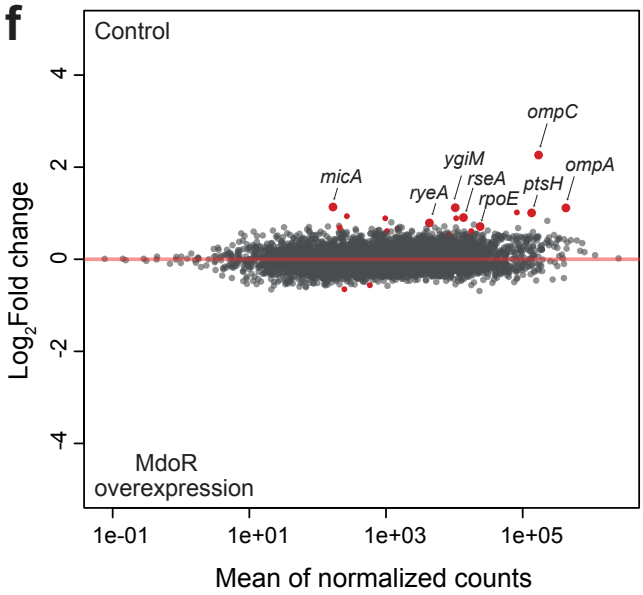
Iosub et al 2018 Figure 6



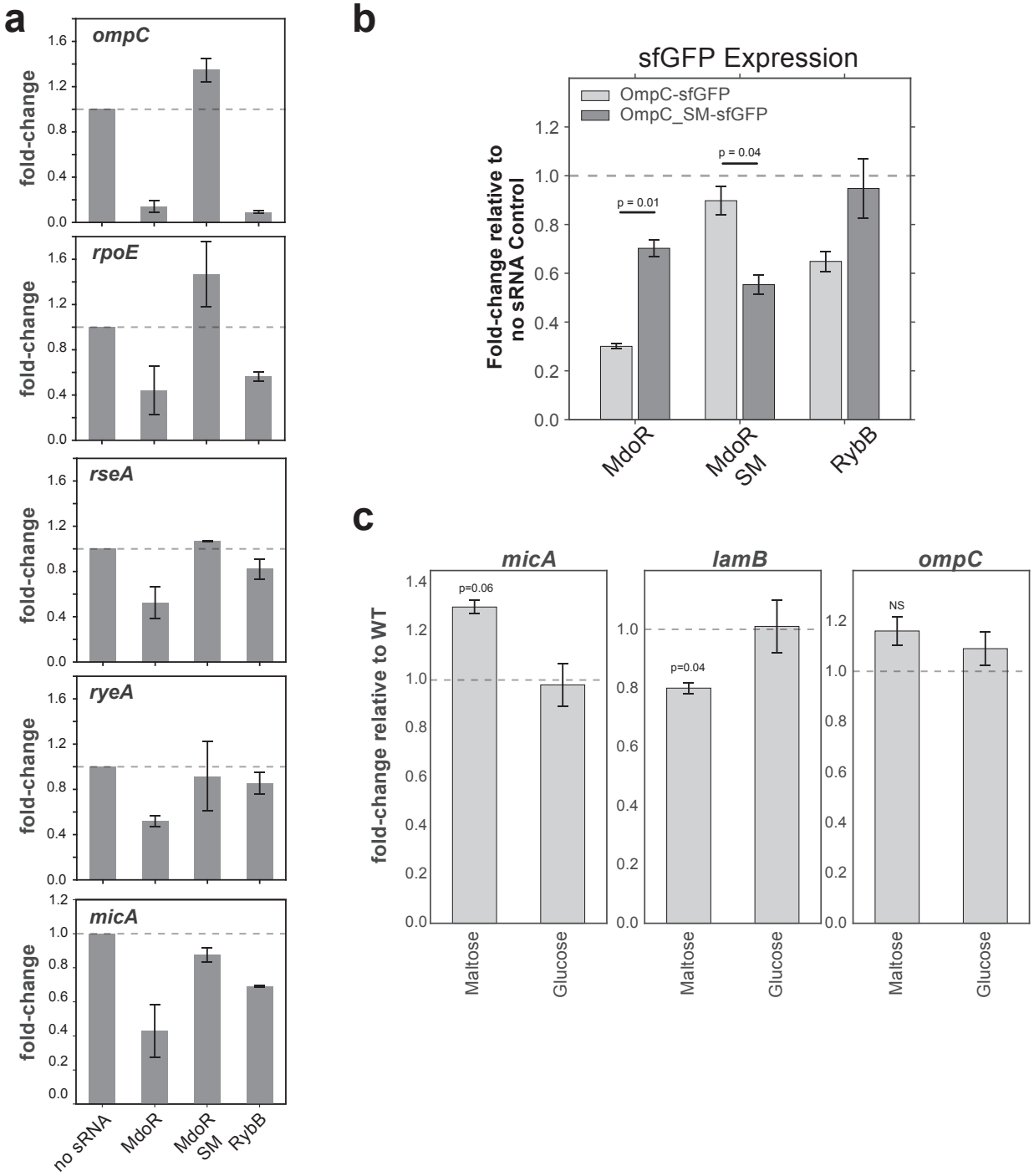
Iosub et al 2018 Figure 7



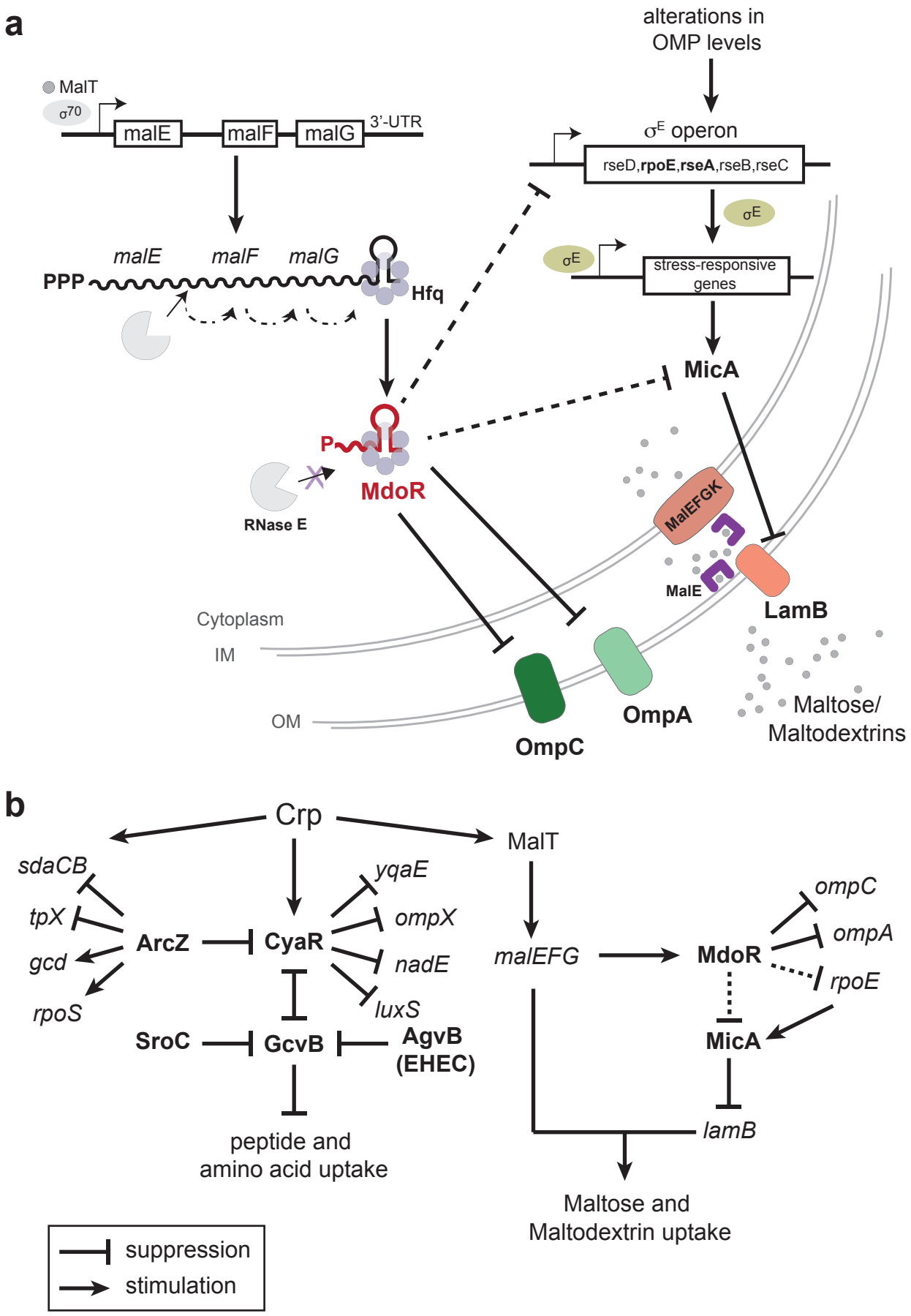




Iosub et al 2018 Figure 8



Iosub et al 2018 Figure 9



Iosub et al 2018 Figure 10