

RESEARCH ARTICLE

Open Access

# Inferring the relation between transcriptional and posttranscriptional regulation from expression compendia

Ivan Ishchukov<sup>1†</sup>, Yan Wu<sup>1†</sup>, Sandra Van Puyvelde<sup>1</sup>, Jos Vanderleyden<sup>1</sup> and Kathleen Marchal<sup>1,2,3\*</sup>

## Abstract

**Background:** Publicly available expression compendia that measure both mRNAs and sRNAs provide a promising resource to simultaneously infer the transcriptional and the posttranscriptional network. To maximally exploit the information contained in such compendia, we propose an analysis flow that combines publicly available expression compendia and sequence-based predictions to infer novel sRNA-target interactions and to reconstruct the relation between the sRNA and the transcriptional network.

**Results:** We relied on module inference to construct modules of coexpressed genes (sRNAs). TFs and sRNAs were assigned to these modules using the state-of-the-art inference techniques LeMoNe and Context Likelihood of Relatedness (CLR). Combining these expressions with sequence-based sRNA-target interactions allowed us to predict 30 novel sRNA-target interactions comprising 14 sRNAs. Our results highlight the role of the posttranscriptional network in finetuning the transcriptional regulation, e.g. by intra-operonic regulation.

**Conclusion:** In this work we show how strategies that combine expression information with sequence-based predictions can help unveiling the intricate interaction between the transcriptional and the posttranscriptional network in prokaryotic model systems.

**Keywords:** sRNA, Gene, Module network, Network inference, *Escherichia coli*

## Background

Transcriptional regulation plays a predominant role in prokaryotic organisms. Although the transcription factor (TF) - mediated network of *Escherichia coli* (*E. coli*) is amongst the best documented network, recent advances in next-generation sequencing unveiled the unprecedented role of sRNA-mediated posttranscriptional regulation by *trans*-acting sRNAs [1-3]. In *E. coli* more than 80 sRNA genes have been identified [4]. Most of the currently known sRNAs are *trans*-acting and assumed to be dependent on the chaperone Hfq. Hfq-dependent sRNAs have a short and imperfect binding region with their

mRNA targets (10–25 bases) [5] and they execute their regulatory mechanism by direct base pairing with the mRNA of their target genes. The outcome of this interaction can be negative, for example by blocking ribosome entry (translational repression) or positive, where positive regulation can be mediated by a plethora of possible mechanisms, for example by melting inhibitory secondary structures, stabilizing the transcript or by sequestration of endonucleolytic cleavage [6,7]. Regulation is often coupled to nuclease mediated cleavage of the mRNA [8,9]. Another small number of posttranscriptional regulatory RNAs also act through antisense RNA base pairing with their targets, but in general in an Hfq independent way. This base pairing usually involves a longer stretch of complementary DNA. Targets of those sRNAs are positioned at their exact same location in the genome (cis location), but on the opposite DNA strand.

Although it is known that sRNAs are involved in a complex regulation network, it is still unclear how this sRNA-mediated posttranscriptional network relates to

\* Correspondence: kathleen.marchal@ugent.be

<sup>†</sup>Equal contributors

<sup>1</sup>Center of Microbial and Plant Genetics, Kasteelpark Arenberg 20, B-3001 Leuven, Belgium

<sup>2</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Gent, Belgium

Full list of author information is available at the end of the article

the TF-mediated regulatory network. Publicly available expression compendia that measure both mRNAs and sRNAs provide useful information to infer such interactions. Modi et al., 2011 [10] have already exploited the potential of network inference to compile an updated sRNA-regulatory network of *E. coli* using a well-known network inference approach CLR [11]. Compared to their work we relied on an alternative module inference framework (rather than a single gene based approach) to simultaneously assign TFs and sRNAs to target genes. In addition, we combined interactions inferred from expression compendia with sequence-based predictions to infer novel sRNA-target interactions. This integrative suite allowed us to both infer the sRNA-target network and to reconstruct the relation between the sRNA and the transcriptional network.

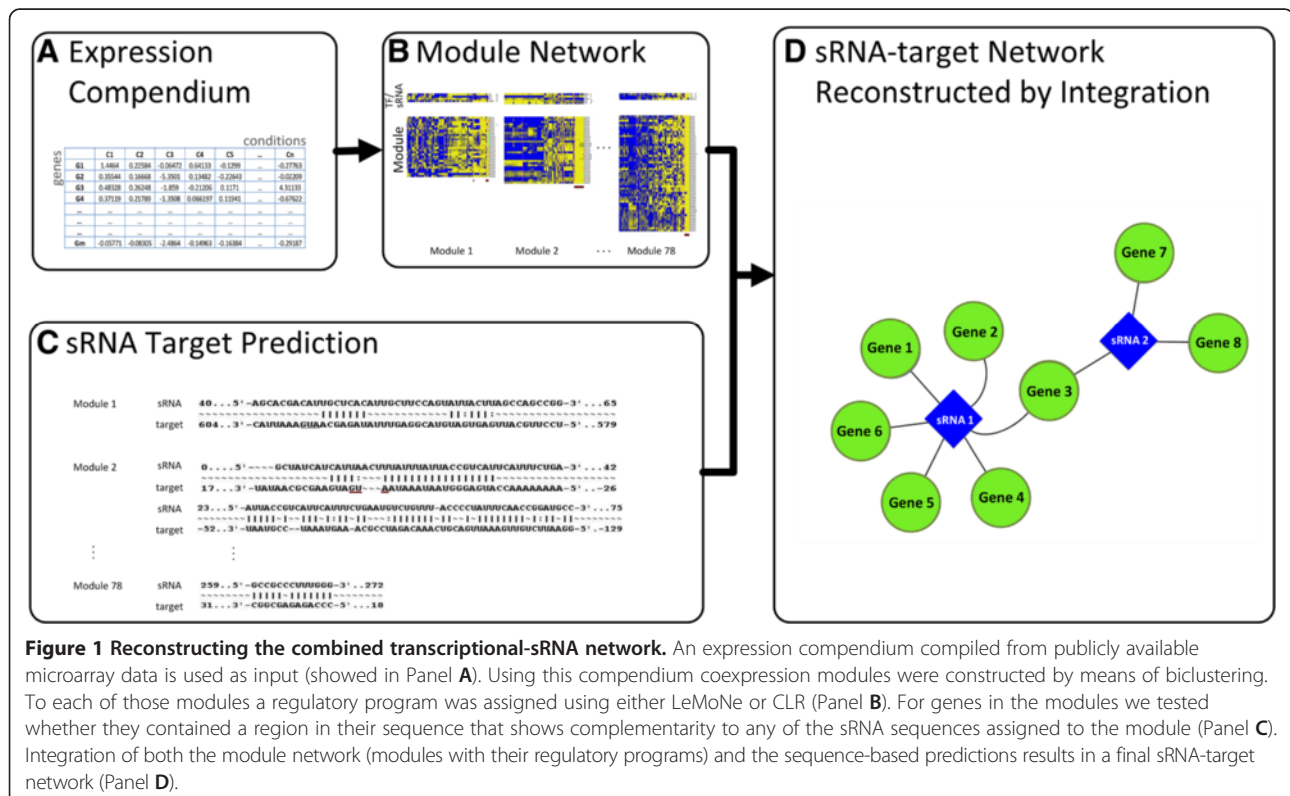
## Results

### Overview of the analysis flow

The analysis flow used to reconstruct the combined transcriptional-sRNA network from the expression compendium is depicted in Figure 1. We first inferred a module network from the expression compendium (Panel A and B). A module consists of a set of genes that is co-expressed, and the conditions under which these genes are co-expressed. Because genes in a module behave similarly, we assume they might be co-regulated either at the

transcriptional or post-transcriptional level. Possible TFs or sRNAs that could explain their co-expression behavior were assigned to each of the obtained modules using expression-based network inference methods that assess whether there exists a similarity in the profile of the assigned TF/sRNA and that of the genes in the module to which the TF/sRNA is assigned. Because it has been shown that network inference approaches differing in their underlying principles often give complementary predictions [12], we used a combination of two different methods (LeMoNe [13] and CLR [11]) to make our final predictions.

Expression-based inference methods cannot distinguish whether the regulators affect the modules to which they are assigned in a direct versus an indirect way, i.e. whether the assigned regulators directly interact with the target genes in the modules to affect their regulation or whether they affect another regulator which on its turn physically interacts with the targets in the module. To infer for the assigned sRNAs direct from indirect modes of regulation, we complemented the expression-based inferences with sequence-based information (Panel C): direct interactions as summarized in the sRNA-target interaction network (Panel D) were inferred by identifying genes in the module that contained a region in their sequence that was complementary to a region present in an sRNA assigned to the module (results obtained from IntaRNA [14] and TargetRNA [15,16]).



**Figure 1 Reconstructing the combined transcriptional-sRNA network.** An expression compendium compiled from publicly available microarray data is used as input (shown in Panel A). Using this compendium coexpression modules were constructed by means of biclustering. To each of those modules a regulatory program was assigned using either LeMoNe or CLR (Panel B). For genes in the modules we tested whether they contained a region in their sequence that shows complementarity to any of the sRNA sequences assigned to the module (Panel C). Integration of both the module network (modules with their regulatory programs) and the sequence-based predictions results in a final sRNA-target network (Panel D).

### Module inference

To infer modules, we relied on a previously developed global biclustering algorithm (ISA [17,18]). With ISA, we identified 78 modules in our dataset of which 57 were functionally enriched. All 78 modules contained at least one predicted sRNA target (based on IntaRNA and TargetRNA predictions (see Methods)) and 21 modules contained an experimentally validated sRNA target. For several modules which showed a clear functional overrepresentation, sRNA targets within the module had a function related to the functional category assigned to the module (see below for a more detailed description of those modules). An overview of the modules is given in Additional file 1: Table S1: Characteristics of module network as reconstructed by CLR and LeMoNe. 37 out of the 108 experimentally verified sRNA targets ended up in a module, while the remaining sRNA targets remained unclustered. In some cases, e.g. for *OmrA*, *OmrB*, *OxyS*, *DsrA*, *GcvB* targets of the same sRNA, were clustered together. For the other cases it seems that targets, despite being regulated by the same sRNA exhibit a profoundly different expression pattern (Additional file 2: Table S2: overview of the sRNAs in different modules). This indicates an intricate interaction between the sRNAs and the TF-mediated transcriptional network.

### Assigning a regulatory program

To map the interaction between the transcriptional and the posttranscriptional network, we reconstructed a module network by assigning to each of the modules a regulatory program that consists of a combination of sRNAs and TFs. Because a module expression profile is more robust than a single gene profile, we assigned regulatory programs to modules rather than to single genes (here referred to as module networks). To this end, we used two previously described inference tools that exploit expression information to reconstruct networks, CLR and LeMoNe [12,19]. Both tools assume that the expression profile of the regulator is a proxy for its activity. They thus assign a regulator to a module if the expression profiles of both entities show a relation. The first method, LeMoNe [13] is inherently module-based: it assigns a regulatory program to pregrouped gene sets (or module). LeMoNe first partitions for the selected gene set in a module, the conditions according to different levels of over (under) expression (multivariate distribution). Then it assigns to each module those regulators for which the expression profiles best fits all or part of the condition partitions in the module. CLR [11], on the other hand assigns a regulatory program based on the degree of mutual information between the expression profile of a regulator and that of each possible target gene. Although initially developed as a direct inference method that assigns a regulatory program on a gene by gene basis, we apply CLR here to assign a regulatory program at module

level (see Methods). As input for both CLR and LeMoNe we used the gene selection of the 78 ISA modules, but instead of only using the conditions selected in the module, we used all conditions in the compendium to perform the regulatory assignments as both CLR and LeMoNe can weigh the conditions 'according' to their relevance for the assignment of the regulatory program. An overview of the assignment of a regulatory program to the different modules is shown in Additional file 1: Table S1.

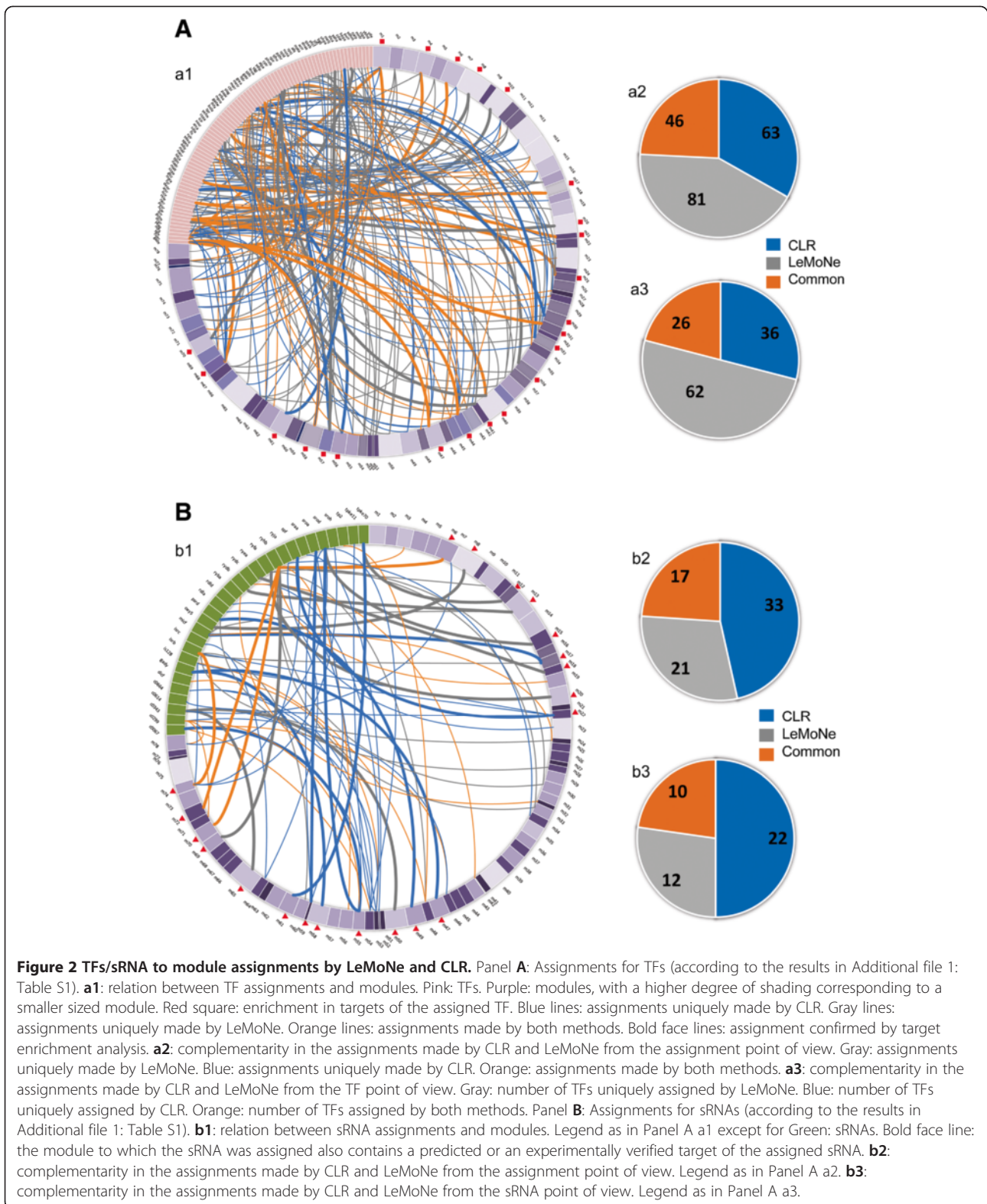
### Complementarity between CLR and LeMoNe

Both CLR and LeMoNe assign a score to the TF/sRNA module assignments. To set the threshold for CLR, we used the FDR-based strategy described in the original paper [11]. For LeMoNe we relied on a previously optimized threshold [19]. These criteria resulted in assigning on average 2–3 regulators (with a regulator being defined as a TF or sRNA) per module for CLR (159 assignments comprising 75 regulators) and 1–2 for LeMoNe, (165 assignments comprising 89 regulators). To four modules no regulators were assigned, given the defined thresholds.

Figure 2 gives an overview of the TF and sRNA assignments by either LeMoNe or CLR to the different modules (Panel A). Figure 2 panel B and C summarizes quantitatively the complementarity between LeMoNe and CLR for respectively the TFs and sRNAs. Because in principle the same regulator (TF/sRNA) can be assigned to different modules, the number of assignments is larger than the number of assigned regulators. The complementarity between both methods is therefore displayed both from the perspective of the assignments and of the assigned number of TFs.

For TFs a total of 190 assignments, covering 84 TFs were made of which 127 assignments covering 71 TFs were made by LeMoNe and 109 assignments covering 49 TFs were made by CLR. Of the total number of assignments, 46 assignments covering 26 different TFs were consistent between CLR and LeMoNe. For about 39 assignments made by either CLR or LeMoNe covering 17 different TFs, the assignments were confirmed by target enrichment analysis: that is the module to which the TF was assigned indeed was overrepresented in known targets (according to RegulonDB) of the assigned TF (Figure 2 Panel A). 66.7% of these assignments that were confirmed by target enrichment analysis can be found in the intersection of targets predicted by both methods.

For the sRNAs, a total of 71 assignments for 30 different sRNAs were made. Of the 38 assignments for 18 sRNAs made by LeMoNe and the 50 assignments for 26 sRNAs made by CLR, 17 assignments for 10 sRNAs were consistently predicted by both algorithms. For 3 cases (one assignment for *MicF* (predicted by CLR), one assignment for *Spf* (predicted by CLR), and one assignment for *RyhB* (predicted by LeMoNe and CLR)), the module to which



the sRNA was assigned also contained at least one experimentally verified target of this assigned sRNA. 21 out of the 43 modules to which an sRNA was assigned, contained

either a predicted or an experimentally verified target for the assigned sRNA, indicating that probably in the other cases the assigned sRNAs are involved in the indirect rather



than the direct regulation of the genes in the module e.g. by regulating the TFs that were assigned to the module rather than by regulating the genes in the module itself.

Regarding the complementarity between LeMoNe and CLR, results show that both methods tend to make more predictions for TFs than for sRNAs which indicates that the signal in the dataset is more pronounced for TFs than for sRNAs. This is to be expected as the probes in the used array platforms were designed to measure protein coding genes (and thus TFs), but not sRNAs. sRNAs are in general represented by very few probes and not necessarily on all platforms (see Methods). Other differences between the results obtained by either method can be explained by the method's specificities. In general, CLR performs well if the profile of the assigned regulator matches the profile of the module for a minimal number of conditions of any type (it tests a global similarity rather than a condition specific profile). LeMoNe on the other hand assigns more importance in finding a fit between the expression profile of the regulator and that of the module for those condition partitions that are homogenous. The latter can contain a large part of the conditions in the dataset (in which case the interaction would also be recovered by CLR) or it can be restricted to only a subset of the conditions (in which case the interaction would only be recovered by LeMoNe). However, LeMoNe's ability to assign regulators that only match part of the conditions in the dataset comes at the expense of also penalizing more small mismatches between the profiles of the regulators and those of the modules for the condition partitions of importance. So even if the global profile seems to match quite well (high CLR score), such mismatches can result in low LeMoNe scores. Both properties can explain why for the same thresholds used on both TFs and sRNAs, LeMoNe tends to find more unique assignments for TFs than for sRNAs compared to CLR (for which the opposite is true): because for TFs the signal is rather robustly measured, LeMoNe can also assign regulators for which the expression profile only matches the module's profile for a subset of the conditions, whereas CLR can not assign those regulators, explaining why LeMoNe assigns relatively more TFs than CLR. However, in the case of sRNAs, the expression signal is much less robustly measured and small mismatches in the expression profile of the regulator and that of the module become a major limitation for LeMoNe. This results in CLR being able to assign more sRNAs than LeMoNe under those conditions.

The most reliable assignments of regulators (TF or sRNA) to modules evidently consist of the predictions that were made by both LeMoNe and CLR (indicated in bold in the Additional file 1: Table S1). A selection of interesting modules is described in more detail below.

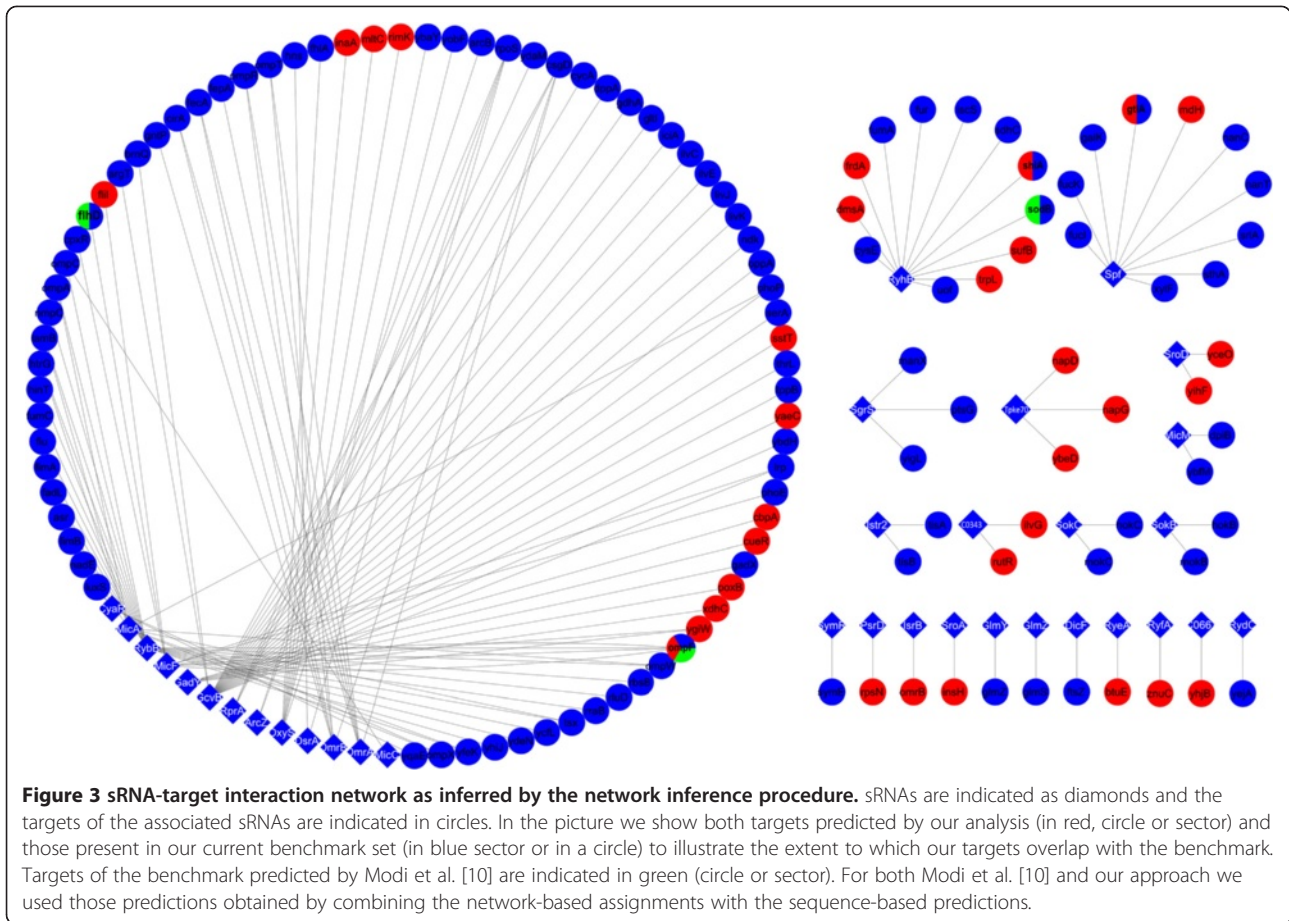
### sRNA-target interaction network

To derive from the sRNA-module assignments the underlying sRNA-target network, we defined as potential targets of a certain sRNA those module genes that also contained in their upstream regions a putative recognition site of the sRNA assigned to the module (based on the union of IntaRNA and TargetRNA). This resulted in 30 different sRNA-target interactions (corresponding to 33 sRNA-target assignments) comprising 14 sRNAs (out of the 72 for which we tried to make an assignment) regulating 30 genes (Figure 3).

Recently, Modi et al. [10] have also applied CLR to predict novel sRNA-target interactions for 24 sRNAs (which are all contained in the set we used in our analyses). Our approach is intrinsically quite different from the one of Modi et al. [10] in its experimental set up: whereas Modi et al. [10] assigns sRNAs to single genes, we assign sRNAs to modules (module-base inference). In addition both approaches, ours and the one of Modi et al. [10] use slightly different parameter settings for running CLR and the sequence-based sRNA-target predictions. We compared to what extent the results of our approach corresponded to those of Modi et al. [10] (Additional file 3: Table S3).

Whereas we integrate sequence-based predictions with the module-based sRNA assignments to infer the direct sRNA-target interaction network (Figure 3 and Additional file 3: Table S3 column f), Modi et al. [10] in their study used expression profile-based sRNA assignments to construct an sRNA-target interaction network which is thus composed of both direct and indirect targets. The original number of predictions made by Modi et al. [10] based on expression data only is shown column c of the table in the Additional file 3: Table S3. To make the results of both networks more comparable, we removed from the predictions of Modi et al. [10] the indirect interactions using the sequence-based prediction approach adopted in their original paper (column d).

Results on a benchmark dataset show that both methods ours and the one of Modi et al. [10] perform poorly in recovering true benchmark interactions and have a very low sensitivity. When considering the targets of Modi et al. [10], predicted based on high scoring CLR assignments, benchmark interactions could be recovered for 5 sRNAs (Additional file 3: Table S3 column e) (MicF (1 target)), GadY (1 target), GcvB (2 target), RyhB (1 targets)). Of those, three benchmark interactions (MicF, RyhB, GadY) were retained if the target was also required to contain a recognition site for the respective sRNA (column e between brackets). We had a similarly low sensitivity (column h between brackets) and could also only recover the known MicF target from the benchmark. Comparing columns d and g of Additional file 3: Table S3, containing respectively the number of direct sRNA targets predicted by



Modi et al. [10] and our study by combining expression and sequence data shows that both methods predict novel targets for a very different set of sRNAs. Besides for MicF and RyhB for which both our method and Modi et al. [10] can predict targets, Modi et al. [10] predicts targets for GcvB whereas we do not. In contrast, with our approach we were able to predict targets for OxyS, GadY, RyeA, PsrD, SroD, Tpk70, SroA, IsrB and C0343 that were not detected by Modi et al. [10] (the last 7 sRNAs were not analyzed by Modi et al. [10]). This indicates the complementarity between single gene and module-based approaches [12].

#### Description of interesting modules

An overview of the modules can be found in Additional file 1: Table S1. The more detailed content of the modules, together with their regulatory program can be found at Additional file 4: Module Overview.

#### Module 6

Module 6 (Figure 4 panel A) is a rather large module being overrepresented for genes involved in iron transport and iron-sulfur cluster assembly. Two regulators, one TF (IscR) and one sRNA (RyhB) were assigned to this module

with a high reliability (as their assignment was confirmed by both LeMoNe and CLR): IscR, a sulfur-cluster containing TF, known to regulate the expression of operons that encode components of a pathway of iron-sulfur cluster assembly, iron-sulfur proteins, anaerobic respiration enzymes and biofilm formation [20,21]. Module 6 contains two known targets being regulated by IscR (the operons *nrdHIEF* and *sufABCDEs*, involved in iron-sulfur cluster assembly [22]). Interestingly, the regulator IscR belongs to a polycistronic mRNA *iscRSUA* known to be regulated by the sRNA RyhB which was also assigned to this module.

This assignment of RyhB was further confirmed by that fact that the module contained one known target of RyhB (*shiA*) [23]. In addition to this known target we also have one predicted RyhB target in the module, *sufB* (Figure 4 panel B), which is a component of the SufBC<sub>2</sub>D Fe-S cluster assembly scaffold complex [24] that is responsible for the synthesis of Fe-S clusters [24-26].

Down-regulation of proteins involved in assembly of Fe-S clusters by RyhB would make sense given the known function of RyhB during Fe homeostasis: RyhB is known to reduce iron consumption under low-iron conditions by downregulating expression of iron-containing proteins [27-29] and the Fur regulon, of which also several genes





(See figure on previous page.)

**Figure 4 Module 6 and its regulatory program.** Panel **A**: the regulatory program assigned to this module. The module to which the regulatory program was assigned; yellow indicates high expression levels and blue refers to low expression levels of genes in the module. The genes correspond to the genes present in the original module discovered by ISA. Conditions present in the original ISA module are indicated by a horizontal bar. As both LeMoNe and CLR use all conditions when assigning their respective regulatory program, we indicated also the additional compendium conditions that were relevant for assigning the respective regulatory programs. Genes in the module correspond to likely targets of the assigned regulators. Targets indicated by a square correspond to known targets of the assigned TF (s). Targets indicated by empty triangle correspond to predicted targets of the assigned sRNA, targets indicated by filled triangle correspond to known targets of the assigned sRNA. Panel **B**: sRNA-target interaction as predicted by the sequence-based analysis for both known and predicted targets of the sRNAs assigned to the module. Indicated sequence positions refer to the location of the recognition sequence relative to the translation start of the gene the sRNA is predicted to interact with (if ATG is indicated in red). If the ATG is indicated in black it refers to the start codon of a neighboring gene. In this case the underlined ATG corresponds to the start codon of *IscS* as the target site of RyhB is located in the intergenic region between the predicted target *IscR* and *IscS*.

are present in the module and which is known to have a central role in iron metabolism. In addition the direct binding of RyhB to *sufB* seems likely as the interacting region identified in RyhB is located in an unstructured region and overlaps with the binding region of previously detected targets (see Additional file 1: Table S1).

Interestingly, both the ISC assembly system (*iscRSUA* operon), which is responsible for Fe-S cluster production under normal conditions and the SUF assembly system (*sufABCDES* operon) responsible for Fe-S cluster production under oxidative stress conditions are encoded by polycistronic operons [30]. The polycistronic *iscRSUA* mRNA to which also *IscR*, the regulator assigned to the module belongs, is known to be processed by RyhB by inducing a cleavage of the operonic transcript between *iscR* and *iscSUA* [31]. Our predicted interaction between RyhB and *sufB*, the second gene of the *sufABCDES* operon suggests that RyhB would also interact with the *sufABCDES* operon through an intraoperonic regulation mechanism [31].

#### Module 17

Module 17 contains 31 genes most of which relate to membrane encoded transport systems. The module was predicted to be regulated by the TFs MalT, of which also the targets were found to be enriched in this module, CueR and YgiV. For the latter TFs no known targets were found in the module. For the sRNA MicF assigned to this module, one of its known targets, *ompF*, belongs to the module [9]. Although the module does not contain any other predicted targets of MicF, we found evidence of a MicF recognition site upstream the coding region of the *ygiV*, a TF also assigned to the module. The recognition site is located in the region from -102 to -52 bp upstream of *ygiV*, a region that comprises the short intergenic region between *ygiW* and *ygiV*, and the C terminal end of the coding region of *ygiW*. The region on MicF, predicted to bind with *ygiV*, although being quite large partially overlaps with the unstructured 5' end of MicF to which also the known MicF target *OmpF* binds (see Additional file 4: Modules overview). YgiV is known

as a repressor of McbR (also known as YncC), a regulator of biofilm formation [32].

Also interesting is the assignment of the sRNA *IsrB* with unknown function to the same module. *IsrB* in *E. coli* has no documented targets yet, but its genomic location overlaps with the coding regions of *azu*-genes, a set of inner membrane encoding genes with unknown function. A link between *IsrB* and membrane encoded functions is plausible viewing the large subset of membrane related functionalities in this module. However, relying on our sequence-based sRNA target prediction, no direct target of *IsrB* was found to be present in this module so *IsrB* could be involved in the indirect regulation of this module (e.g. by regulating other regulators that on their turn regulate the genes in the module).

#### Module 58

Module 58, mainly expressed under stationary growth contains genes involved in acid response, amino acid starvation (purine salvage, amino acid uptake) and induction of microaerobiosis (represented by Ecocyc enrichment analysis). Three regulators were assigned to Module 58, GadE, CueR and the sRNA GadY, all of which are known targets of RpoS [33]. Module 58 was also found to be enriched in direct targets of GadE (see Additional file 1: Table S1), indicating that the assignment of GadE as a regulator to module 58 is true. GadE, the central activator of the acid response system controls genes involved in the maintenance of pH homeostasis through its direct targets involved in the glutamate-dependent acid resistance system (here represented by *gadA* and *gadBC* genes) and is involved in multidrug efflux (*mdtE*, *mdtF*) through controlling the expression of the TFs, GadW and GadX both related to acid resistance (of which only GadW was found in the module) [34].

The sRNA GadY which is highly expressed during entry into stationary phase and regulated by low pH [35] is related to the GadE dependent acid response through an intricate network of interactions with GadW (also in module 58) and GadX (according to Regulon DB).



A last regulator assigned to module 58 was CueR “Cu efflux regulator”, which was also predicted to be a target of GadY using our sequence-based predictions. CueR, regulates genes related to the primary copper homeostasis system in response to the presence of copper, silver, or gold ions [36]. None of the known CueR targets related to its function in Cu<sup>2+</sup> homeostasis were found in module 58. However, CueR being a target of GadY and also being assigned as a regulator to module 58 points towards a connection between Cu<sup>2+</sup> and pH homeostasis, a link that has been suggested before. Yamamoto et al., for instance, showed that pH changes affect the genome-wide transcription pattern of copper-balance genes in the presence of CuSO<sub>4</sub> [37].

Besides CueR, module 58 contained three additional predicted targets of GadY (assuming that genes belonging to a module with an assigned sRNA as regulator that also contain a recognition sequence of that sRNA in their upstream region are direct targets of the sRNA). A first one, CbpA has a functionality related to the one of DnaJ and functions as a co-chaperone with DnaK. A second one, PoxB is pyruvate oxidase and the last one XdhA-XdhB-XdhC is a putative heterotrimeric xanthine dehydrogenase [38]. How their functionalities link to the role of GadY is less clear. GadY is known to be an antisense binding sRNA that acts on its cis encoded target GadX. So far GadX is the only characterized target of GadY. However, it cannot yet be excluded that GadY would have additional targets encoded elsewhere on the genome [39], the more because GadY has been shown to share the Hfq binding property of transacting sRNAs [33]. The fact that the regions to which GadY would bind in its predicted targets CueR, DnaJ and PoxB are located quite far upstream of their respective annotated TSSs could explain why such non-conventional targets have largely been overlooked by computational predictions.

#### **Module 8**

Module 8 contains pathways which relate to oxidative membrane stress (osmotic stress response, efflux pumps, membrane remodeling). Three regulators have been assigned to this module by LeMoNe: MarA, Fur and OxyS. MarA, is a “multiple antibiotic resistance” regulator of which indeed part of its known regulon was found in the module. MarA is an outer membrane porin involved in the efflux of several hydrophobic and amphipathic molecules and is known to be involved in resistance to antibiotics and oxidative stress [40]. The module indeed contains MarA targets such as TolC, an outer membrane porin. Although the Fur regulon members are not well represented in this module, the autoregulated TF Fur has not only been assigned to the module, but also belongs to the module itself, further supporting its assignment. Besides its well documented role in iron homeostasis, Fur is

also known to be involved in oxidative stress responses by downregulating iron uptake systems [41].

Next to these TFs also the sRNA OxyS known to play a regulatory role in the oxidative stress response [42] was assigned to this module. Three targets regulated by OxyS were predicted with our approach and were found in module 8, implying that OxyS regulates together with MarA the genes in module 8: RimK, a ribosomal protein S6 modification protein belonging to the *ybjC-nfsA-rimK-ybjN* operon, an operon which indeed is known to be regulated by (Rob/MarA/SoxS) and OxyR. So, the additional regulation of *rimK* (intraoperonic promotor site) by OxyS is plausible. InaA, a second predicted target of OxyS present in module 8 is pH-inducible protein involved in stress response [43,44]. A third target of OxyS which we could predict is *mltC*, a membrane-bound lytic murein transglycosylase C, known to be induced by oxidative stress via SoxS [45].

#### **Module 20**

Module 20 contains genes belonging to pathways involved in transport, oxidative stress response (Mar and Sox operons) and gluconate, ascorbate utilization. SoxR which was assigned to the module is also part of the module. The regulator IclR also assigned to the module is known to regulate the glyoxylate bypass operon [46,47].

According to our predictions, the sRNA assigned to this module RyfA, which has no assigned function yet would have one predicted target in the module, i.e. ZnuC, the ATP-binding component of an ABC transporter involved in high-affinity zinc uptake (ZnuABC). *znuC* transcripts were shown to disappear or markedly decreased at 5 min after zinc addition [37]. Such quick induction or repression of the zinc-responsive genes upon increasing environmental zinc levels suggests a regulation mechanism mediated by sRNAs. Some of the enzymes being expressed in the module are indeed known to depend upon a Zn<sup>2+</sup> containing active site (e.g. UlaE [48]). In literature we found an indirect relation between SoxS and RyfA through the regulation of the predicted target ZnuC. SoxS is known to increase the expression of the zinc uptake system ZnuACB in *E. coli*, although no direct binding of SoxS to the promoter of *znuACB* has been observed [49].

#### **Module 61**

Module 61 (44 genes) contains genes related to oxidation-reduction, electron transport and energy generation. Three TFs, AdiY and YahA were assigned to this module by LeMoNe and one CdaR by CLR. To our knowledge the role of YahA, a c-di-GMP-specific phosphodiesterase is yet unknown (YahA contains an EAL domain close to an N-terminal putative DNA-binding domain). AdiY was previously shown to be strongly upregulated after a rapid decrease in external pH [50]. Its known target, the arginine

decarboxylase system (*adi*) is known to be induced in rich medium, under anaerobic conditions, and at low pH [50,51] conditions under which genes present in the module are also known to be expressed. CdaR, regulates genes involved in the uptake and metabolism of galactarate and glucarate and is also found to be one of the regulators for which the targets are enriched in the module. Besides these TFs also Tpk70 a sRNA of approximately 40 nt in length with yet unknown function was assigned to this module [52]. Module 61 also contains two predicted targets of Tpk70 that is NapG and NapD (predicted using sequence-based methods) both parts of the periplasmic nitrate reductase system in *E. coli* [53-55].

## Methods

### Microarray compendium

The *E. coli* compendium used in this study consisted of 348 contrasts taken from an initial compendium of 610 contrasts, covering 4311 genes and 78 sRNAs and profiling a diverse set of conditions such as (an)aerobic growth, growth in ethanol, different pH levels, various *E. coli* strains included, etc. In an initial test we used the full compendium to generate our results (610 contrasts). We noted that a subset of 262 contrasts (row 19 in Additional file 5: Table S4), corresponding to the previous compendium of [11] did not exhibit a large variability across conditions and did therefore bias our results. That is why we excluded this dataset for further analysis (resulting in a total of 348 remaining contrasts). The final compendium was composed of 31 experiments performed on 21 Affymetrix platforms AffymetrixGene Chip *E. coli* Antisense Genome Array [Ecoli\_ASv2], 12 AffymetrixGeneChip *E. coli* Genome 2.0 Array [E\_coli\_2] (Additional file 5: Table S4: Description of expression compendium). The platform design files did not yet contain probe annotation for sRNAs, but some of the 'intergenic' probes corresponded to regions containing sRNAs. By blasting the probe sequences against a more recent annotation of the *E. coli* genome, we could link 1331 (1323 probes for sRNAs from Additional file 6: Table S5: Probe Information) probes to 72 sRNAs that were covered by all 3 Affymetrix platforms (a detailed list with probe annotation can be found in the supplementary information Additional file 6: Table S5).

### sRNA target prediction

sRNA target prediction was based on IntaRNA [14] and TargetRNA [15]. For both IntaRNA and TargetRNA, we searched for sRNA interaction sites in a region close to the start codon (ranging from 50 bp upstream to the ATG until -150 downstream from the ATG as this range showed to have the best sensitivity in recovering known sRNA-target interactions). For both algorithms the seed length was set to 8 nucleotides. Other settings had default values

TargetRNA was performed without scores accounting for GU pairs, without thermodynamic information and without orthology information).

Target predictions were considered selected stringently if they were in the top 25 list of both methods (stringent selection criterium). If they were only predicted in the top 25 list of one of the two methods, they were considered less reliable (and we referred to this as the non-stringent selection criterium). Using the stringent selection criteria approximately 12% of the known targets for sRNAs could be retrieved with an average PPV of 19%. With the non-stringent criteria these numbers are respectively 18% with a PPV of 1%. However, as we could recover so few benchmark interactions with the stringent criterium, we relied on the non-stringent criterium, assuming that we would compensate for the lower PPV by integrating the sequence-based predictions with those obtained from the expression-based assignments (Additional file 7: Performance tests to optimize sequence-based sRNA-target predictions). For a full list of predictions we refer to Additional file 8: Table S6: Predictions of sRNA targets based on IntaRNA and TargetRNA.

To identify putative interactions between TFs and sRNAs assigned to the same modules, we screened the sequences of the assigned regulators for recognition sites of the sRNAs assigned to the same module in a region near the start codon. Here we used a more relaxed screening to identify putative interactions that is we extended the screened region from -150 bp upstream to the ATG to 200 downstream from the ATG and screened an additional region near the transcript end (200 bp upstream of the stopcodon to the stopcodon).

To further validate the sequence-based predictions of the most promising sRNA-target interactions (those described in the main text), 1) we tested to what extent the location of the recognition region of sRNA-target interactions in the respective sRNAs overlapped with an unstructured regions [56] and, with the binding region of previously described targets of the same sRNAs, 2) we tested to what extent the location of the recognition region of the respective sRNA-target interactions in the target sequences were positioned relative to the start codon and transcription start positions of those targets. Results of this analysis are displayed Additional file 4: Module overview.

### Module detection

Modules (biclusters) were generated by running ISA [17,18] obtained from ISA website (<http://www2.unil.ch/cbg/index.php?title=ISA>) on the *E. coli* expression compendium mentioned above, using default thresholds on the minimum number of genes (3) and on the minimum number of conditions (2). The ctrh parameter for checking convergence was set to be 50. The default number of 100 random seeds

needed to initiate the global biclustering were generated using `generate.seeds()`. Running ISA on the *E. coli* compendium resulted in 78 modules.

### Inferring module networks

To assign a regulatory program to the 78 modules, we used two previously described inference methods CLR [11] and LeMoNe [12,19]. As input for both CLR and LeMoNe we used the gene selection of the 78 ISA modules, but instead of only using the conditions selected in the module, we used all conditions in the compendium to perform the regulatory assignments as both CLR and LeMoNe. Both tools require as input a regulator list to reconstruct the regulatory program. To this end we compiled a list of 311 potential regulators (transcription factors and sRNA regulators) from RegulonDB release 8.1 (RegulonDB version 7.0) [57], Ecocyc [58], RFAM [59], and literature (see Additional file 9: Table S7: List of Regulators). These regulators consisted of 170 annotated TFs (RegulonDB), 69 predicted transcription factors predicted (Ecocyc), 72 annotated sRNAs (RFAM), covering about 50% of all TFs in *E. coli* [60] and 30% of all estimated sRNAs (based on the analysis of all the non-coding transcripts in *E. coli* [61]). LeMoNe was obtained from source website (<http://bioinformatics.psb.ugent.be/software/details/lemone>). It is an unsupervised, module-based method that assigns regulators (TFs or sRNAs) to an expression module (here our ISA modules) by first fitting a multivariate normal distribution to the expression profiles of all genes in the module such that the module conditions are regrouped in partitions with a coherent expression value (being either over or underexpressed according to a multivariate normal distribution that fits all genes in the module). It subsequently searches for the set of regulators for which the expression profiles best fit to part or all of the condition partitions in the module. LeMoNe was run with the default parameter settings. Regulators assigned by LeMoNe were ranked according to their regulator score (we only withheld regulators which had a score higher than 100). In our previous work we showed that this threshold value was shown to be well suited for assigning TFs as regulators to modules in *E. coli* [19]. CLR was obtained from the developer [11]. CLR was run with default parameter settings except for the statistical mode settings (we selected NORMAL mode which is recommended on larger networks). CLR was initially developed as a direct inference method [12], that assigns a regulator to an individual gene if the expression profiles of the gene and its assigned regulator show a sufficiently high mutual information. CLR was initially not developed to explicitly exploit modularity. To be able to use CLR also at module level, we calculated for each module its mean expression profile (over all conditions of the compendium). We assigned regulators to both individual gene expression profiles and module mean

profiles. We observed that CLR was able to more confidently assign regulators to the mean module profiles than to the individual gene expression profiles (op scoring assignments made by CLR almost always involved mean module profiles when we used as input to CLR a mix of mean profiles and the profiles of the individual genes). This indicates that using modules instead of individual profiles helps reducing the noise and makes the assignments more robust. As a final threshold for CLR we selected a cut-off that corresponds to an FDR < 0.05 using the FDR (false discovery rate) calculation described in the original paper.

### Calculating functional enrichment

Functional enrichment was based on GO (category Biological process) [62] using a hypergeometric test and using false discovery rate (FDR, based correction for multiple testing (Benjamini and Hochberg correction), as implemented in BinGO [63]. Categories overrepresented with a significance level < 0.05 were withheld. Testing whether coexpression modules were enriched for known targets of TFs (according to RegulonDB) was also based the hypergeometric test with FDR correction ( $p < 0.05$  were considered significant). Because of the low number of known sRNA targets, the enrichment value of expression modules in sRNA targets was unreliable and therefore replaced by indicating for each expression module the number of known or predicted targets it contained for the assigned sRNA.

Ecocyc pathways functional enrichment was based on pathways description [58] using a hypergeometric test and using false discovery rate (FDR, i.e. the expected proportion of false positives among the positively identified tests), based correction for multiple testing (Benjamini and Hochberg correction, build-in Genecodis tool [64]). Categories overrepresented with a significance level < 0.001 were withheld.

### Benchmarks

We compiled a benchmark of 118 experimentally verified sRNA-target interactions in *E. coli* covering 25 different sRNAs (both Hfq-dependent and non Hfq-dependent) from curated databases (RegulonDB [55], Ecocyc [58]) and literature. A full table of curated interactions can be found on the supplementary table (Additional file 10: Table S8: Description of our benchmark of known sRNA-target interactions). This benchmark is an updated version of the one described in Modi et al. [10] with more recently validated interactions [65-68].

Compared to Modi et al. [10] our benchmark contained 71 additional interactions for 25 sRNAs (containing known interactions for 5 sRNAs that were not covered by Modi et al. [10], and 65 additional interactions for 20 sRNAs that were covered in the initial benchmark of Modi et al. [10]).



### Comparison with the predicted sRNA network of Modi et al. 2011

The CLR based predictions made by Modi et al. [10] were derived from its supplementary file (st02.doc). Sequence-based predictions on sRNA-target interactions were not available as such from the original paper. To generate these predictions in a way that was described in the original paper we did the following: we used the TargetRNA [69] with the same default settings [15] mentioned also above, but including a default setting for the seed length (9 nucleotides), sRNA sequences were used from the source TargetRNA server [69], or, if not available on this server from RFAM or Ecocyc. We searched for sRNA recognition sites in a region close to the start codon, ranging from -30 bp upstream of the ATG to 20 bp downstream of the ATG. Target predictions were considered selected stringently if they had a p-value less than 0.01. Predictions on sRNA-target interactions prior to and after the sequence-based filtering were compared for both our analysis and the analysis of Modi et al. [10] on our updated benchmark dataset (Additional file 3: Table S3).

### Discussion

This work studied the relation between the transcriptional and the posttranscriptional network by applying two complementary module-based inference frameworks LeMoNe and CLR to simultaneously assign TFs and sRNAs to coexpression modules. In general, we observed that in contrast to targets of the same TF, targets of the same sRNA rarely show a similar coexpression behavior, indicating that the level of posttranscriptional regulation has been evolved towards an additional layer of regulation that is largely involved in finetuning gene-specific expression behavior. This is further confirmed by the observation that for several of our novel detected sRNA-target interactions, the recognition region of the interaction is located within an operon (e.g. *MicF-ygiV*, *RyhB-sufB*, *GadY-xdhC*), pointing towards an extended role of these sRNAs in intraoperonic regulation [70]. We also observed that in contrast to TFs, sRNAs assigned to a module were most frequently only involved indirectly in the regulation of the module genes that is by modulating the expression of the TF responsible for the coexpression behavior of the genes in the module.

Combining sequence-based target predictions with the module-based sRNA assignments allowed us to draw a predicted sRNA-target interaction network. A similar sRNA-target reconstruction was recently performed by Modi et al. [10] using CLR in a non-module based setting. Our approach is intrinsically different from the single gene based approach applied by Modi et al. [10]. In a module-based approach one assumes that different genes in the same module all support the inferred interactions, which should contribute to the reliability of the assignments, but might

come at the expense of missing interactions with genes that will not end up in a module which might be recovered by a single gene based approach [12]. Secondly, rather than presenting the interactions inferred through the expression data as a final sRNA-target interaction network as was done in Modi et al. [10] we combined the expression-based inference with a sequence-based inference. Because of these differences in the number of interactions, it is not trivial to compare the results of our analysis and the one of Modi et al. [10] directly. Applying also a sequence-based filtering to distinguish direct from indirect targets on the results of Modi et al. [10] indicated that for both methods the recovery of known interactions is still very limited. At this stage this does not say much about the quality of the novel predicted interactions, but rather indicates the limitation of the currently available expression compendia that are mainly compiled from expression arrays that were not yet designed to measure sRNAs.

### Conclusion

In this work we show the potential of combining network-based inference with sequence-based prediction techniques for identifying sRNA-target interactions and for studying the intricate relation between the transcriptional and the posttranscriptional network. Such integrative data-analysis techniques can be expected to become increasingly attractive as more specialized compendia measuring simultaneously protein coding genes and sRNAs will become available.

### Additional files

**Additional file 1: Table S1.** Characteristics of module network as reconstructed by CLR and LeMoNe.

**Additional file 2: Table S2.** Overview of the sRNAs in the different modules.

**Additional file 3: Table S3.** Performance of inferring sRNA-target interactions.

**Additional file 4: Module overview.**

**Additional file 5: Table S4.** Description of the expression compendium.

**Additional file 6: Table S5.** Probe information.

**Additional file 7: Performance tests to optimize sequence-based sRNA-target predictions.**

**Additional file 8: Table S6.** Predictions of sRNA targets based on intaRNA and TargetRNA.

**Additional file 9: Table S7.** List of regulators.

**Additional file 10: Table S8.** Description of our benchmark of known sRNA-target interactions.

### Authors' contribution

II performed the analyses and helped designing the experiments and drafting the text. YW performed the module detection, contributed to the benchmark analysis and the draft. SVP contributed extensively to the biological interpretation and helped drafting the text. KM supervised the work, designed the analyses and wrote the manuscript. JV contributed to helpful discussions and critically read the manuscript. All authors read and approved the final manuscript.



## Acknowledgements

We would like to thank Pieter Monsieurs for initiating the study and also Daniel Ryan for his contribution during his master thesis. This work is supported by: 1) KU Leuven: GOA/08/011, PF/10/010 (NATAR), personal grant Ivan Ishchukov. 2) Agentschap voor Innovatie door Wetenschap en Technologie (IWT): NEMOA and personal grant IWT of Sandra Van Puyvelde, 3) Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO) G.0428.13 N; 4) Ghent University [Multidisciplinary Research Partnership "N2N"].

## Author details

<sup>1</sup>Center of Microbial and Plant Genetics, Kasteelpark Arenberg 20, B-3001 Leuven, Belgium. <sup>2</sup>Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Gent, Belgium. <sup>3</sup>Department of Information Technology, Ghent University, IMinds, VIB, 9052 Gent, Belgium.

Received: 24 June 2013 Accepted: 9 January 2014

Published: 27 January 2014

## References

- Sorek R, Cossart P: Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* 2010, **11**(1):9–16.
- Kroger C, Dillon SC, Cameron AD, Pappenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hebrard M, Handler K, et al: The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci USA* 2012, **109**(20):E1277–E1286.
- Sharma CM, Hoffmann S, Darfeuille F, Reigier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, et al: The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 2010, **464**(7286):250–255.
- Raghavan R, Groisman EA, Ochman H: Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res* 2011, **21**(9):1487–1497.
- Pappenfort K, Vogel J: Regulatory RNA in bacterial pathogens. *Cell Host Microbe* 2010, **8**(1):116–127.
- Pappenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J: Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. *Cell* 2013, **153**(2):426–437.
- Frohlich KS, Vogel J: Activation of gene expression by small RNA. *Curr Opin Microbiol* 2009, **12**(6):674–682.
- Pappenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JC, Vogel J: *SigmaE*-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol* 2006, **62**(6):1674–1688.
- Storz G, Vogel J, Wassarman KM: Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* 2011, **43**(6):880–891.
- Modi SR, Camacho DM, Kohanski MA, Walker GC, Collins JJ: Functional characterization of bacterial sRNAs using a network biology approach. *Proc Natl Acad Sci U S A* 2011, **108**(37):15522–15527.
- Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS: Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007, **5**(1):e8.
- De Smet R, Marchal K: Advantages and limitations of current network inference methods. *Nat Rev Microbiol* 2010, **8**(10):717–729.
- Michoel T, Maere S, Bonnet E, Joshi A, Saeys Y, Van den Bulcke T, Van Leemput K, van Remortel P, Kuiper M, Marchal K, et al: Validating module network learning algorithms using simulated data. *BMC Bioinforma* 2007, **8**(2):55.
- Busch A, Richter AS, Backofen R: IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 2008, **24**(24):2849–2856.
- Tjaden B, Goodwin SS, Opdyke JA, Guillier M, Fu DX, Gottesman S, Storz G: Target prediction for small, noncoding RNAs in bacteria. *Nucleic Acids Res* 2006, **34**(9):2791–2802.
- Tjaden B: Prediction of small, noncoding RNAs in bacteria using heterogeneous data. *J Math Biol* 2008, **56**(1–2):183–200.
- Ihmels J, Friedlander G, Bergmann S, Sarig O, Ziv Y, Barkai N: Revealing modular organization in the yeast transcriptional network. *Nat Genet* 2002, **31**(4):370–377.
- Bergmann S, Ihmels J, Barkai N: Iterative signature algorithm for the analysis of large-scale gene expression data. *Physical review E, Statistical, nonlinear, and soft matter physics* 2003, **67**:031902. 3 Pt 1.
- Michoel T, De Smet R, Joshi A, Van de Peer Y, Marchal K: Comparative analysis of module-based versus direct methods for reverse-engineering transcriptional regulatory networks. *BMC Syst Biol* 2009, **3**:49.
- Tokumoto U, Takahashi Y: Genetic analysis of the *isc* operon in *Escherichia coli* involved in the biogenesis of cellular iron-sulfur proteins. *J Biochem* 2001, **130**(1):63–71.
- Wu Y, Outten FW: IscR controls iron-dependent biofilm formation in *Escherichia coli* by regulating type I fimbria expression. *J Bacteriol* 2009, **191**(4):1248–1257.
- Takahashi Y, Tokumoto U: A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids. *J Biol Chem* 2002, **277**(32):28380–28383.
- Whipp MJ, Camakaris H, Pittard AJ: Cloning and analysis of the *shiA* gene, which encodes the shikimate transport system of *Escherichia coli* K-12. *Gene* 1998, **209**(1–2):185–192.
- Layer G, Gaddam SA, Ayala-Castro CN, Ollagnier-de Choudens S, Lascoux D, Fontecave M, Outten FW: SufE transfers sulfur from SufS to SufB for iron-sulfur cluster assembly. *J Biol Chem* 2007, **282**(18):13342–13350.
- Outten FW, Wood MJ, Munoz FM, Storz G: The SufE protein and the SufBCD complex enhance SufS cysteine desulfurase activity as part of a sulfur transfer pathway for Fe-S cluster assembly in *Escherichia coli*. *J Biol Chem* 2003, **278**(46):45713–45719.
- Bolstad HM, Wood MJ: An in vivo method for characterization of protein interactions within sulfur trafficking systems of *E. coli*. *J Proteome Res* 2010, **9**(12):6740–6751.
- Masse E, Gottesman S: A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA* 2002, **99**(7):4620–4625.
- Masse E, Vanderpool CK, Gottesman S: Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol* 2005, **187**(20):6962–6971.
- Semsey S, Andersson AM, Krishna S, Jensen MH, Masse E, Sneppen K: Genetic regulation of fluxes: iron homeostasis of *Escherichia coli*. *Nucleic Acids Res* 2006, **34**(17):4960–4967.
- Shepard EM, Boyd ES, Broderick JB, Peters JW: Biosynthesis of complex iron-sulfur enzymes. *Curr Opin Chem Biol* 2011, **15**(2):319–327.
- Desnoyers G, Morissette A, Prevost K, Masse E: Small RNA-induced differential degradation of the polycistronic mRNA *iscRSUA*. *EMBO J* 2009, **28**(11):1551–1561.
- Zhang XS, Garcia-Contreras R, Wood TK: *Escherichia coli* transcription factor YncC (McbR) regulates colanic acid and biofilm formation by repressing expression of periplasmic protein YbiM (McbA). *ISME J* 2008, **2**(6):615–631.
- Opdyke JA, Kang JG, Storz G: GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *J Bacteriol* 2004, **186**(20):6698–6705.
- Tramonti A, De Canio M, De Biase D: GadX/GadW-dependent regulation of the *Escherichia coli* acid fitness island: transcriptional control at the *gady-gadW* divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Mol Microbiol* 2008, **70**(4):965–982.
- Hayes ET, Wilks JC, Sanfilippo P, Yohannes E, Tate DP, Jones BD, Radmacher MD, BonDurant SS, Slonczewski JL: Oxygen limitation modulates pH regulation of catabolism and hydrogenases, multidrug transporters, and envelope composition in *Escherichia coli* K-12. *BMC Microbiol* 2006, **6**:89.
- Hobman JL, Wilkie J, Brown NL: A design for life: prokaryotic metal-binding MerR family regulators. *Biomaterials* 2005, **18**(4):429–436.
- Yamamoto K, Ishihama A: Transcriptional response of *Escherichia coli* to external copper. *Mol Microbiol* 2005, **56**(1):215–227.
- Xi H, Schneider BL, Reitzer L: Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *J Bacteriol* 2000, **182**(19):5332–5341.
- Opdyke JA, Fozo EM, Hemm MR, Storz G: RNase III participates in GadY-dependent cleavage of the *gadX-gadW* mRNA. *J Mol Biol* 2011, **406**(1):29–43.
- Cohen SP, Hachler H, Levy SB: Genetic and functional analysis of the multiple antibiotic resistance (*mar*) locus in *Escherichia coli*. *J Bacteriol* 1993, **175**(5):1484–1492.
- Imlay JA: Cellular defenses against superoxide and hydrogen peroxide. *Annu Rev Biochem* 2008, **77**:755–776.
- Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G: A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* 1997, **90**(1):43–53.
- Slonczewski JL, Gonzalez TN, Bartholomew FM, Holt NJ: Mu d-directed *lacZ* fusions regulated by low pH in *Escherichia coli*. *J Bacteriol* 1987, **169**(7):3001–3006.

44. White S, Tuttle FE, Blankenhorn D, Dosch DC, Slonczewski JL: pH dependence and gene structure of *inaA* in *Escherichia coli*. *J Bacteriol* 1992, **174**(5):1537–1543.
45. Pomposiello PJ, Koutsolioutsou A, Carrasco D, Demple B: SoxRS-regulated expression and genetic analysis of the *yggX* gene of *Escherichia coli*. *J Bacteriol* 2003, **185**(22):6624–6632.
46. Gui L, Sunnarborg A, Pan B, LaPorte DC: Autoregulation of *iclR*, the gene encoding the repressor of the glyoxylate bypass operon. *J Bacteriol* 1996, **178**(1):321–324.
47. Peskov K, Goryanin I, Prank K, Tobin F, Demin O: Kinetic modeling of ace operon genetic regulation in *Escherichia coli*. *J Bioinform Comput Biol* 2008, **6**(5):933–959.
48. Shi R, Pineda M, Ajamian E, Cui Q, Matte A, Cygler M: Structure of L-xylulose-5-Phosphate 3-epimerase (UlaE) from the anaerobic L-ascorbate utilization pathway of *Escherichia coli*: identification of a novel phosphate binding motif within a TIM barrel fold. *J Bacteriol* 2008, **190**(24):8137–8144.
49. Warner DM, Levy SB: SoxS increases the expression of the zinc uptake system *ZnuACB* in an *Escherichia coli* murine pyelonephritis model. *J Bacteriol* 2012, **194**(5):1177–1185.
50. Stim-Herndon KP, Flores TM, Bennett GN: Molecular characterization of *adiY*, a regulatory gene which affects expression of the biodegradative acid-induced arginine decarboxylase gene (*adiA*) of *Escherichia coli*. *Microbiology* 1996, **142**(Pt 5):1311–1320.
51. Stim-Herndon KP: Effect of the local anesthetics phenethyl alcohol and procaine on *hns* mutants of the acid-induced biodegradative arginine (*adi*) and lysine (*cad*) decarboxylases of *Escherichia coli*. *Curr Microbiol* 1995, **30**(5):281–285.
52. Rivas E, Klein RJ, Jones TA, Eddy SR: Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Current biology: CB* 2001, **11**(17):1369–1373.
53. Cole J: Nitrate reduction to ammonia by enteric bacteria: redundancy, or a strategy for survival during oxygen starvation? *FEMS Microbiol Lett* 1996, **136**(1):1–11.
54. Brondijk THC, Nilavongse A, Filenko N, Richardson DJ, Cole JA: NapGH components of the periplasmic nitrate reductase of *Escherichia coli* K-12: location, topology and physiological roles in quinol oxidation and redox balancing. *Biochem J* 2004, **379**(Pt 1):47–55.
55. Bueno E, Mesa S, Bedmar EJ, Richardson DJ, Delgado MJ: Bacterial adaptation of respiration from oxic to microoxic and anoxic conditions: redox control. *Antioxid Redox Signal* 2012, **16**(8):819–852.
56. Beisel CL, Updegrave TB, Janson BJ, Storz G: Multiple factors dictate target selection by Hfq-binding small RNAs. *EMBO J* 2012, **31**(8):1961–1974.
57. Gama-Castro S, Salgado H, Peralta-Gil M, Santos-Zavaleta A, Muniz-Rascado L, Solano-Lira H, Jimenez-Jacinto V, Weiss V, Garcia-Sotelo JS, Lopez-Fuentes A, et al: RegulonDB version 7.0: transcriptional regulation of *Escherichia coli* K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res* 2011, **39**(Database issue):D98–D105.
58. Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muniz-Rascado L, Bonavides-Martinez C, Paley S, Krummenacker M, Altman T, et al: EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Res* 2011, **39**(Database issue):D583–D590.
59. Gardner PP, Daub J, Tate J, Moore BL, Osuch IH, Griffiths-Jones S, Finn RD, Nawrocki EP, Kolbe DL, Eddy SR, et al: Rfam: Wikipedia, clans and the “decimal” release. *Nucleic Acids Res* 2011, **39**(Database issue):D141–D145.
60. Babu MM: Computational approaches to study transcriptional regulation. *Biochem Soc Trans* 2008, **36**(Pt 4):758–765.
61. Shinohara A, Matsui M, Hiraoka K, Nomura W, Hirano R, Nakahigashi K, Tomita M, Mori H, Kanai A: Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. *BMC Genomics* 2011, **12**:428.
62. Lewis S, Ashburner M, Reese MG: Annotating eukaryote genomes. *Curr Opin Struct Biol* 2000, **10**(3):349–354.
63. Maere S, Heymans K, Kuiper M: BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005, **21**(16):3448–3449.
64. Tabas-Madrid D, Nogales-Cadenas R, Pascual-Montano A: GeneCodis3: a non-redundant and modular enrichment analysis tool for functional genomics. *Nucleic Acids Res* 2012, **40**:W478–W483.
65. Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA: 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* 2011, **108**(31):12875–12880.
66. Beisel CL, Storz G: Discriminating tastes: physiological contributions of the Hfq-binding small RNA Spot 42 to catabolite repression. *RNA biology* 2011, **8**(5):766–770.
67. Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC, Vogel J: Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol* 2011, **81**(5):1144–1165.
68. Coornaert A, Chiaruttini C, Springer M, Guillier M: Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS genetics* 2013, **9**(1):e1003156.
69. Tjaden B: TargetRNA: a tool for predicting targets of small RNA action in bacteria. *Nucleic Acids Res* 2008, **36**:W109–W113.
70. Waters LS, Storz G: Regulatory RNAs in bacteria. *Cell* 2009, **136**(4):615–628.

doi:10.1186/1471-2180-14-14

Cite this article as: Ishchukov et al.: Inferring the relation between transcriptional and posttranscriptional regulation from expression compendia. *BMC Microbiology* 2014 **14**:14.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

