Small RNAs (sRNAs) have been discovered in every bacterium examined and have been shown to play important roles in the regulation of a diverse range of behaviors, from metabolism to infection. However, despite a wide range of available techniques for discovering and validating sRNA regulatory interactions, only a minority of these molecules have been well characterized. In part, this is due to the nature of posttranscriptional regulation: the activity of an sRNA depends on the state of the transcriptome as a whole, so characterization is best carried out under the conditions in which it is naturally active. In this issue of mSystems, Arrieta-Ortiz and colleagues (M. L. Arrieta-Ortiz, C. Hafemeister, B. Shuster, N. S. Baliga, et al., mSystems 5:e00057-20, 2020, https://doi.org/10.1128/mSystems.00057-20) present a network inference approach based on estimating sRNA activity across transcriptomic compendia. This shows promise not only for identifying new sRNA regulatory interactions but also for pinpointing the conditions in which these interactions occur, providing a new avenue toward functional characterization of sRNAs.

KEYWORDS
network inference, ncRNA, sRNA

Two decades after the first genome-wide screens showed small RNAs (sRNAs) to be a widespread feature of bacterial genomes, these short noncoding transcripts remain enigmatic. High-throughput RNA sequencing (RNA-seq)-based approaches now routinely find hundreds of potential sRNAs in any bacterium examined (1). Despite this, only on the order of dozens of sRNAs have been characterized in any detail, and the number with clearly defined physiological roles is even smaller—in fact, it remains unclear what fraction may even possess a function at all (2). While some characterized sRNAs have specialized functions, such as titrating regulatory RNA-binding proteins (3), the majority appear to effect their function through RNA-RNA interactions with target transcripts. Depending on the molecular details and protein factors recruited, these interactions can have a wide range of consequences, including activating translation (4) or affecting transcript termination (5), but the most common outcome appears to be repression of translation often with a concomitant decay of the target transcript. In the best-characterized bacterial systems, *Escherichia coli* and *Salmonella enterica*, many of these sRNAs appear to form regulatory networks centered on RNA-binding proteins, most famously Hfq (6) and the more recently discovered ProQ (7).

The bottleneck in sRNA characterization is not a lack of techniques. A basic molecular toolkit for identifying and validating sRNA-target interactions was developed in the mid-2000s by several different groups (8), and includes techniques like sRNA pulse induction followed by transcriptomics to identify putative targets, reporter systems to validate in vivo interactions, and structural probing to map the details of base pairs formed in the interaction. Computational tools for predicting interactions based on thermodynamic models of RNA folding have also been long available and are useful for understanding individual interactions (9) but generally have too high a false-positive rate to be useful for genome-wide predictions (10), though methods incorporating...
interaction conservation (11) or experimental evidence (12) are beginning to improve this. More recently, approaches based on coupling pulldowns to high-throughput sequencing have been developed. These approaches include GRIL-seq (global small noncoding RNA target identification by ligation and sequencing) (13) and MAPS (MS2 affinity purification coupled with RNA sequencing) (14), which involve pulling down an sRNA of interest with or without ligation, and RIL-seq (RNA interaction by ligation and sequencing) (15), which relies on pulling down an RNA-binding protein and ligating the RNA species interacting on it. These methods can provide direct evidence for sRNA-target interactions, though they also often report tens to hundreds of potential interactions, and it remains unclear how many of these impact cell physiology.

Beyond any inherent limitations of the techniques available, there is a more fundamental reason that sRNAs are difficult to characterize, namely, the inherently epistatic nature of posttranscriptional regulation. It was recognized early on that sRNA-based regulation is dependent on the overall cellular state in a way that transcriptional regulators are not (16). Most obviously, an sRNA cannot regulate a transcript that is not expressed under the tested conditions. More subtly, artificially induced competition for central RNA-binding proteins can have dramatic global effects on sRNA-based regulation (17), and it is becoming clear that similar phenomena affect regulation under more ordinary conditions (18). Other factors, such as the condition-specific expression of sRNA sponges (19), may reduce the regulatory activity of sRNAs. There is even evidence of cross talk between the networks mediated by RNA-binding proteins (20), adding another layer of complexity. We are then left in a situation where we cannot easily determine sRNA function because we do not know the conditions it is relevant to, and we do not know the conditions it is relevant to because we do not know the function.

Arrieta-Ortiz and colleagues offer one way out of this conundrum (21), using network inference (NI). NI is a collection of techniques that aim to reconstruct an underlying regulatory network from large collections of transcriptomic data. These techniques range from methods that apply clustering or thresholding approaches to simple statistics like Pearson correlations or mutual information to systems of ordinary differential equations that attempt to provide a working dynamic model of regulatory interactions (22). One of the reasons these methods have not been more widely adopted is the lack of large transcriptomic atlases from which to derive networks, outside certain model organisms like E. coli. In fact, the mutual information-based context likelihood of relatedness (CLR) algorithm was applied to sRNA network prediction in E. coli almost a decade ago (23), with promising preliminary results. Now with the wide availability of RNA-seq making such compendia easier to produce (24–26), NI approaches are due for a revival.

The key innovation of Arrieta-Ortiz and colleagues’ method (21) is to account for the epistatic nature of sRNA regulation in applying NI. Building on their previous work with transcription factors (27), they first apply a dimensionality reduction technique called network component analysis (28) that uses an incomplete map of the regulatory architecture of the cell to derive sRNA activity scores across the input transcriptomic compendium accounting for known regulatory interactions. They show that these scores anticorrelate more strongly with the expression level of sRNA targets than the expression level of sRNAs themselves, as would be expected in the presence of potential confounding factors. They then show that using these activity scores in place of sRNA expression improves the recovery of known sRNA-target interactions not included in the prior network with both their own NI approach, Inferelator (29), and a hybrid approach based on CLR. With applications in E. coli, Pseudomonas aeruginosa, Bacillus subtilis, and Staphylococcus aureus using a variety of prior network architectures derived from noisy computational or high-throughput experimental approaches, they show that their method consistently reports sRNA-target interactions with independent experimental support. In a final proof of concept, the authors investigate the distribution of activity scores across conditions of the P. aeruginosa sRNA PrrF, recovering known activity in iron-limited, biofilm, and virulence conditions, and the S. aureus sRNA RsaE, suggesting a role in the response to antibiotic stress.
The work in this study clearly demonstrates the utility of NI approaches for generating hypotheses regarding sRNA function and identifying new potential target mRNAs, and importantly, helps to situate the sRNA in something approaching its natural regulatory context. With the continuing accumulation of transcriptomic data in public repositories, it is easy to imagine that in the not too distant future, NI will join molecular techniques, computational target prediction, and high-throughput sequencing-based approaches as a standard part of the sRNA biologist’s toolkit.

ACKNOWLEDGEMENT

I thank Amy Cain for comments on the draft manuscript.

REFERENCES