Microbiology Spectrum
American Society for Microbiology Press

Functional Transcriptomics for Bacterial Gene Detectives

BLANCA M. PEREZ-SEPULVEDA[1] and JAY C. D. HINTON[1]
[1]Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom

ABSTRACT Developments in transcriptomic technology and the availability of whole-genome-level expression profiles for many bacterial model organisms have accelerated the assignment of gene function. However, the deluge of transcriptomic data is making the analysis of gene expression a challenging task for biologists. Online resources for global bacterial gene expression analysis are not available for the majority of published data sets, impeding access and hindering data exploration. Here, we show the value of preexisting transcriptomic data sets for hypothesis generation. We describe the use of accessible online resources, such as SalComMac and SalComRegulon, to visualize and analyze expression profiles of coding genes and small RNAs. This approach arms a new generation of "gene detectives" with powerful new tools for understanding the transcriptional networks of *Salmonella*, a bacterium that has become an important model organism for the study of gene regulation. To demonstrate the value of integrating different online platforms, and to show the simplicity of the approach, we used well-characterized small RNAs that respond to envelope stress, oxidative stress, osmotic stress, or iron limitation as examples. We hope to provide impetus for the development of more online resources to allow the scientific community to work intuitively with transcriptomic data.

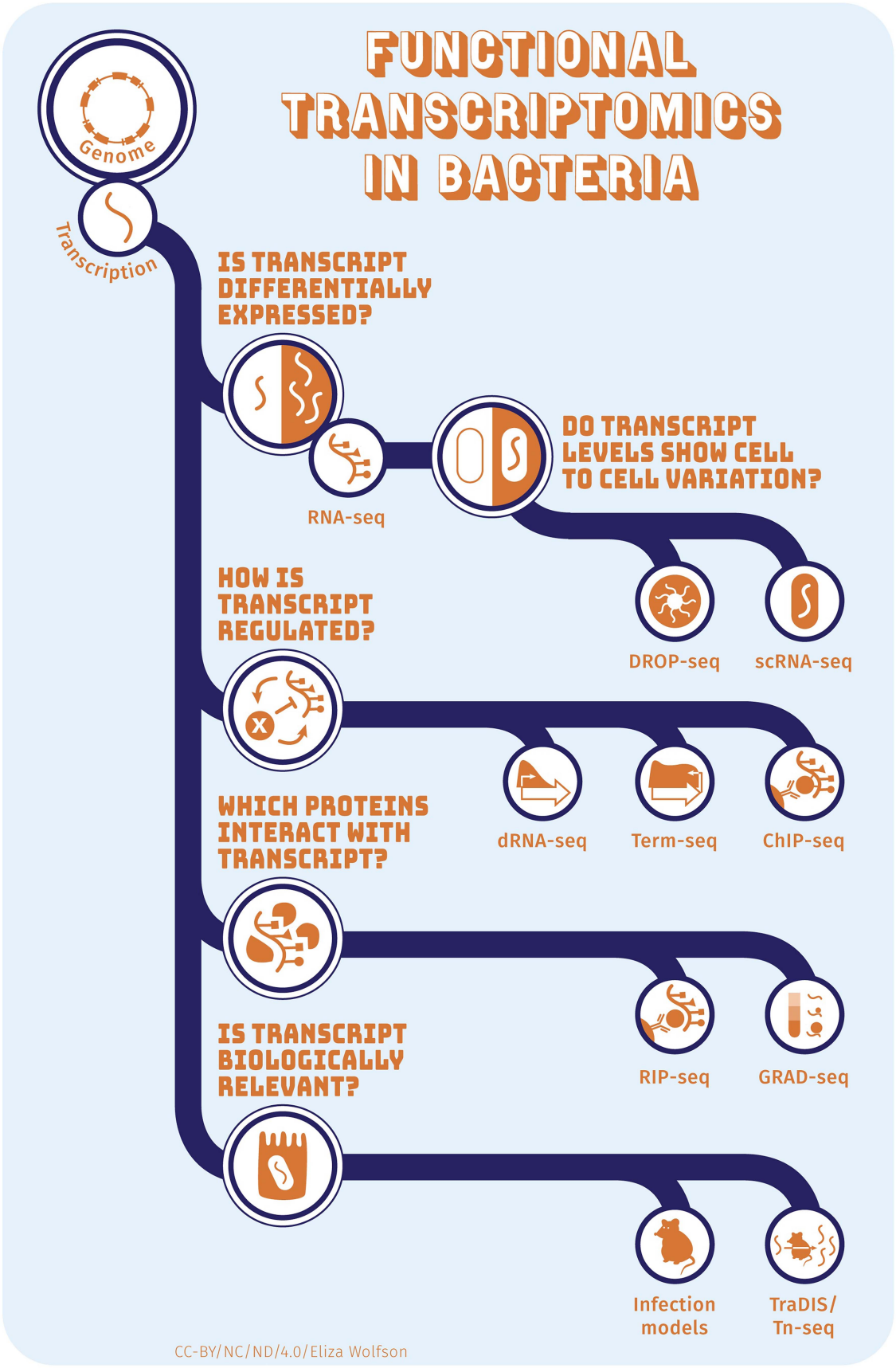
INTRODUCTION

Transcriptional profiling is a valuable part of the functional genomics toolbox. Since the developments in nanotechnology and imaging that led to the invention of next-generation sequencing (1), study of the bacterial transcriptome at the level of the individual nucleotide has proved fruitful. Scientists are now generating increasing amounts of transcriptomic data that need to be managed, analyzed, and stored in an appropriate manner (2). The current need for systematic and accessible approaches for the analysis of gene expression has focused bioinformatic efforts into developing tools for processing transcriptomic data.

Since the amount of transcriptomic data being generated is increasing exponentially, and already exceeds the interpretive capacity of the human brain, we need to improve the ways in which we interact with complex information (3) and choose the best methods for data display and interpretation. "Big data" visualization constitutes a significant challenge, as inappropriate approaches can lead to conclusion bias and other errors (2, 4, 5).

In recent years, the molecular microbiological community has focused on approaches to simplify data visualization and analysis. The creation of online resources by some research labs has provided access to large data sets. A good example is Listeriomics (https://listeriomics.pasteur.fr), a user-friendly online platform that includes curated genomic, transcriptomic, and proteomic data sets generated from *Listeria* species (6). Other valuable online resources include PneumoBrowse (https://veeninglab.com/pneumobrowse) for the analysis of the *Streptococcus pneumoniae* D39V genome (7) and AcinetoCom (http://bioinf.gen.tcd.ie/acinetocom) for investigation of the primary transcriptome of *Acineto-*

Received: 29 June 2018, **Accepted:** 4 September 2018, **Published:** 14 September 2018
Editor: Gisela Storz, Division of Molecular and Cellular Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD; Kai Papenfort, Department of Biology I, Microbiology, LMU Munich, Martinsried, Germany
Citation: Perez-Sepulveda BM, Hinton JCD. 2018. Functional transcriptomics for bacterial gene detectives. *Microbiol Spectrum* 6(5):RWR-0033-2018. doi:10.1128/microbiolspec.RWR-0033-2018.
Correspondence: Jay C. D. Hinton, jay.hinton@liv.ac.uk



bacter baumannii ATCC 17978 (8). Researchers can interrogate and explore the data in these online resources in creative ways. However, other online data repositories are less intuitive, and it can be difficult to access the transcriptomic data needed to make appropriate biological conclusions. The lack of ongoing technical support or suitable funding has led to the demise of some online platforms, leading to valuable resources becoming inaccessible to the wider scientific community.

Here, we will discuss the challenges of visualizing transcriptomic expression data and explain the role that "gene detectives" can play in the interrogation of online resources to develop hypotheses and answer biological questions. We hope to inspire the reader to use online resources to explore transcriptomic data sets.

ENVIRONMENTAL REGULATION OF GENE EXPRESSION

Bacterial pathogens face intense competition in the microbial world. Enterobacterial species that inhabit the gastrointestinal tract have evolved the ability to sense the environment and to use physicochemical information to control gene expression. Environmental modulation of gene expression enables the bacterium to express "the right gene in the right place at the right time," a prerequisite for fitness (9). John Mekalanos commented in 1992 that to understand how bacterial virulence genes are regulated we must first understand when and where the genes are expressed (10).

Transcriptional gene fusions were an invaluable tool in the early days of molecular microbiology, but only allowed the examination of individual genes (11). Soon after the first bacterial genomes were published, it became possible to monitor the expression of all genes in a single experiment. In subsequent years, these transcriptome-based approaches (Fig. 1) have evolved alongside proteomic, metabolomic, and comparative genomic techniques.

TECHNOLOGICAL DEVELOPMENTS THAT HAVE ENABLED FUNCTIONAL TRANSCRIPTOMICS

The transcriptomic revolution began with the advent of microarrays in the late 1990s, which involved the hybridization of target bacterial mRNA transcripts to specific DNA probes (12). DNA microarray technology was developed in the late 1990s by Pat Brown and Joe DeRisi at Stanford University, making it possible to profile the expression of all genes in a single experiment (12, 13). Microarrays became a popular high-throughput tool for studying functional transcriptomics in bacteria. A PubMed (https://www.ncbi.nlm.nih.gov/pubmed) search of the term "bacteria microarray" shows that the number of publications using microarray technology peaked in 2012 with 918 hits, followed by a slow decline in subsequent years.

Microarrays have been largely superseded by high-throughput RNA sequencing (RNA-seq), which was developed in 2008 (14). RNA-seq allows the visualization of every bacterial transcript at the nucleotide level and can provide accurate differential gene expression data from whole-genome transcriptomic analyses. To generate RNA-seq sequence reads, bacteria are grown under defined conditions, and RNA is extracted and then reverse-transcribed to cDNA before high-throughput sequencing (15, 16). Levels of gene expression are assessed by quantifying the RNA-seq reads that map to annotated coding or noncoding genes (tRNA, rRNA, transfer messenger RNA [tmRNA], small RNA [sRNA], etc.). Importantly, RNA-seq allows the identification of elements such as promoters and gene boundaries and the discovery of regulatory RNA species, such as sRNAs and antisense RNAs (17). RNA-seq is now an established technique that can be performed in many sequencing centers worldwide, making it relatively easy to generate transcriptomic data.

RNA-seq has stimulated functional transcriptomic research by enabling a plethora of innovative technologies that have broadened the number of biological

FIGURE 1 Bacterial functional transcriptomics is facilitated by RNA-seq technology. The development of RNA-seq has expanded the range of transcriptome-based techniques that address a variety of biological questions. DROP-seq, RNA-seq of single cells compartmentalized in a droplet; scRNA, single-cell RNA-seq; dRNA-seq, differential RNA-seq; Term-seq, global mapping of 3′ ends of transcripts; ChIP-seq, chromatin immunoprecipitation followed by sequencing; RIP-seq, native RNA immunoprecipitation followed by RNA-seq; GRAD-seq, gradient profiling by RNA-seq; TraDIS, transposon-directed insertion site sequencing; Tn-seq, transposon sequencing. See reference 20 for more details of these techniques. Image by Eliza Wolfson (https://lizawolfson.co.uk) is used under the terms of a creative commons CC-BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode).

questions that can be addressed (18; Fig. 1). These technologies have been the subject of two excellent reviews (15, 19).

THE CHALLENGE OF ACCESSING AND UNDERSTANDING TRANSCRIPTOMIC DATA

Powerful software tools have been developed for the analysis of RNA-seq-based transcriptomic data (20). These approaches are excellent for the generation of lists of differentially expressed genes that meet certain statistical criteria. However, a key barrier to understanding gene expression profiles at a global level has been the huge tab-delimited spreadsheets that can languish untouched in the Supplementary Data section of journals.

The Gene Expression Omnibus (GEO; www.ncbi.nlm.nih.gov/geo) and the EBI ArrayExpress (www.ebi.ac.uk/arrayexpress) are curated public repositories of transcriptomic data from all organisms, including results from microarray and RNA-seq experiments. Although the idea of sharing produced expression data in a public repository has been a positive step forward, GEO is not intuitive to use and data mining can be a challenge. Bioinformatic expertise is too often required to examine and interpret complex data sets or even to successfully download and interpret a transcriptomic experiment from a public repository.

Individual RNA-seq experiments are generally planned with a particular hypothesis in mind, but the entirety of the data collected can subsequently be mined or reinterpreted by the wider community to answer new research questions. It is time to rethink the way in which the worldwide microbiological community manages and accesses our data sets.

IMPROVING THE SHARING AND VISUALIZATION OF BACTERIAL TRANSCRIPTOMIC DATA

To facilitate collaborative analysis, online resources are needed to put global expression data into the hands of biologists. Many RNA-seq-based transcriptomic data sets have now been generated for a range of microbial model systems including yeasts, *Synechocystis*, *Bacillus*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Listeria*, and *Salmonella enterica* (21). Depending on the RNA-seq-based experimental approach, different features can be visualized. These include strand-specific information for the correct annotation of coding sequences, and the identification of transcriptional start sites (TSSs) and 3′ ends (22, 23). To extract accurate differential gene expression values, it is important to normalize RNA-seq data, and several algorithms are available for this purpose (20). Ideally, statistically significant changes at the transcriptional level are generated from biological replicates, allowing mechanistic deductions to be made.

However, the number of online resources or platforms that are available to analyze bacterial expression data does not reflect the amount of published experiments. Of the gene expression data sets that can currently be browsed interactively, many are presented in a way that is not informative, perhaps reflecting the fact that creating, curating, and maintaining online platforms for expression data analysis is a significant effort. Good examples are provided by other research areas that have extremely large data sets, such as the human online resource for transcriptomic analysis, the Genotype-Tissue Expression (GTEx) project (24), which allows intuitive exploration of a variety of features.

The intelligent use of online resources for transcriptomic data analysis can allow gene expression profiles to be used to identify important environmental responses that could otherwise be overlooked. For example, transcriptomic information for 1,292 *E. coli* microarray experiments can be visualized in the GenExpDB (*E. coli* Gene Expression Database; https://genexpdb.okstate.edu), developed by Joe Grissom and Tyrrell Conway.

TOWARD A GLOBAL UNDERSTANDING OF *SALMONELLA* TRANSCRIPTIONAL REGULATION

The analysis and reconstruction of genome-scale transcriptional regulatory networks represents the next frontier in microbial bioinformatics (25), and requires information about the expression of every transcript and the regulatory inputs for every promoter. This challenging aim involves the combination of "-omics" data, including transcriptomic information and a range of chromatin immunoprecipitation (ChIP)-based strategies. Using methods detailed in Fig. 1, binding sites for key regulatory proteins can be identified, and an appreciation of the role of every RNA-binding protein may also be involved. A range of ChIP studies have already been conducted, focused on proteins like H-NS, as well as regulatory factors that control *Salmonella* pathogenicity island 1 and 2 (SPI1 and SPI2) (26–28).

An overarching analytical approach is needed to integrate different types of transcriptional regulatory data together into a single transcriptional network (25). This necessitates an intense bioinformatic effort, which is already being addressed by a few laboratories across the

world (20). Specifically, network inference-based bioinformatic approaches have been used to study transcriptional networks and shed light upon *Salmonella* gene function (29–31), and the SalmoNet network, which integrates metabolic, transcriptional, and protein-protein interaction data, is now available (32). The end result of this long-term, systems-level approach will be a genome-scale regulatory network that will be a significant challenge to visualize effectively (3).

It will be some time before we have a comprehensive transcriptional network that reveals the entire regulatory complexity of *Salmonella*. However, an alternative strategy can already be used to gain insights to the transcriptional world of *Salmonella*. This intuitive approach relies on the ability of the human brain to find patterns within complex data sets (33), coupled with the prior knowledge of molecular microbiologists.

DETECTIVE WORK: INFERRING sRNA FUNCTION FROM GENE EXPRESSION DATA

The development of RNA-seq technology has led to the identification of sRNAs at a global level in many bacterial species (34). The characteristic expression profiles of individual sRNAs (35) allow an understanding of the environmental stresses that modulate sRNA expression to provide functional insights. For *S. enterica* serovar Typhimurium, transcriptional signatures can be visualized in the SalComMac compendium (http://tinyurl.com/SalComMac), which shows the expression of all coding genes and sRNAs in 21 environmental conditions (35, 36). To investigate the impact of infection-relevant environmental stress upon the *Salmonella* transcriptional network, we devised a suite of *in vitro* conditions that reflect particular aspects of the infection process. The conditions include exposing the bacteria to oxidative stress, osmotic shock, acid shock, anaerobic shock, and a low-iron environment (36), plus growth of *S.* Typhimurium within mammalian macrophages (35).

The SalComMac data show that the level of expression of the majority of coding genes and sRNAs is environmentally regulated. Genes that encode key *Salmonella* virulence systems have characteristic transcriptional signatures that we have reported previously (35, 36). For example, coding genes that encode both the structural subunits and the associated effector proteins of the SPI1 type III secretion system show an "SPI1-like" pattern of expression, focused on early stationary phase in rich media. In parallel, we reported a distinct "SPI2-like" transcriptional signature in defined acidic, low-phosphate media (35). More recently, we identified 13 *S.* Typhimurium sRNAs with either an SPI1-like or SPI2-like expression profile (37).

The SalComMac resource reveals changes in the expression of individual sRNAs by particular environmental stressors, and can be used to visualize the global sRNA expression landscape. This approach can generate hypotheses about gene function for experimental investigation. To demonstrate how transcriptomic data can be used to investigate sRNA function, we have explored the expression profiles of 14 well-characterized sRNAs using two data repositories that are currently available for *S.* Typhimurium.

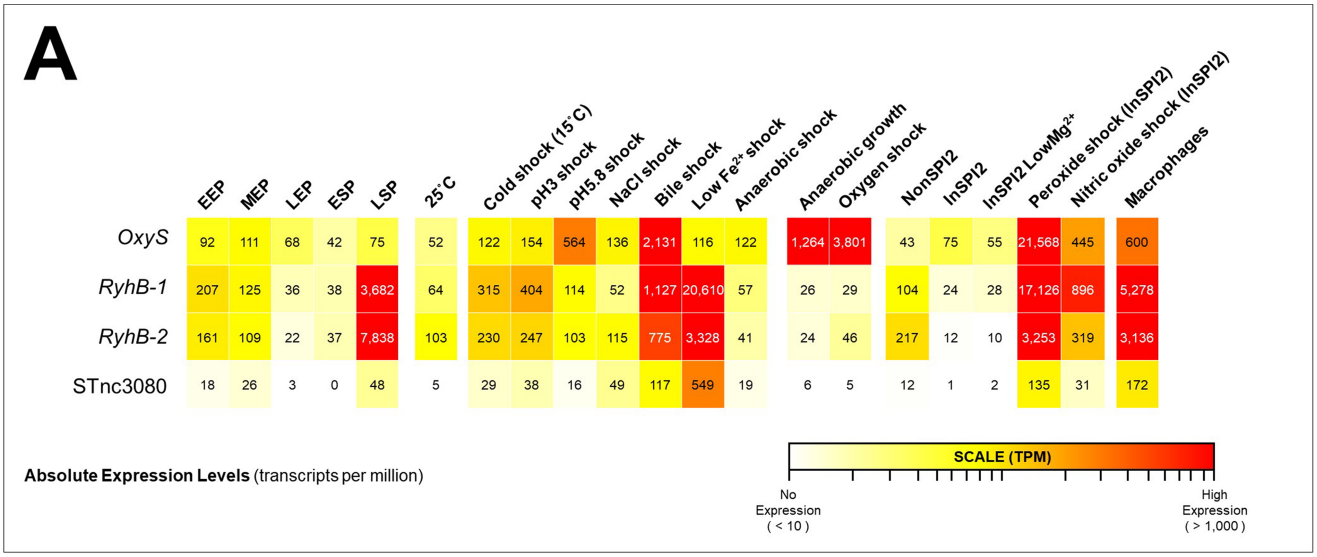
USING ONLINE RESOURCES TO DETERMINE *S.* TYPHIMURIUM sRNA FUNCTION

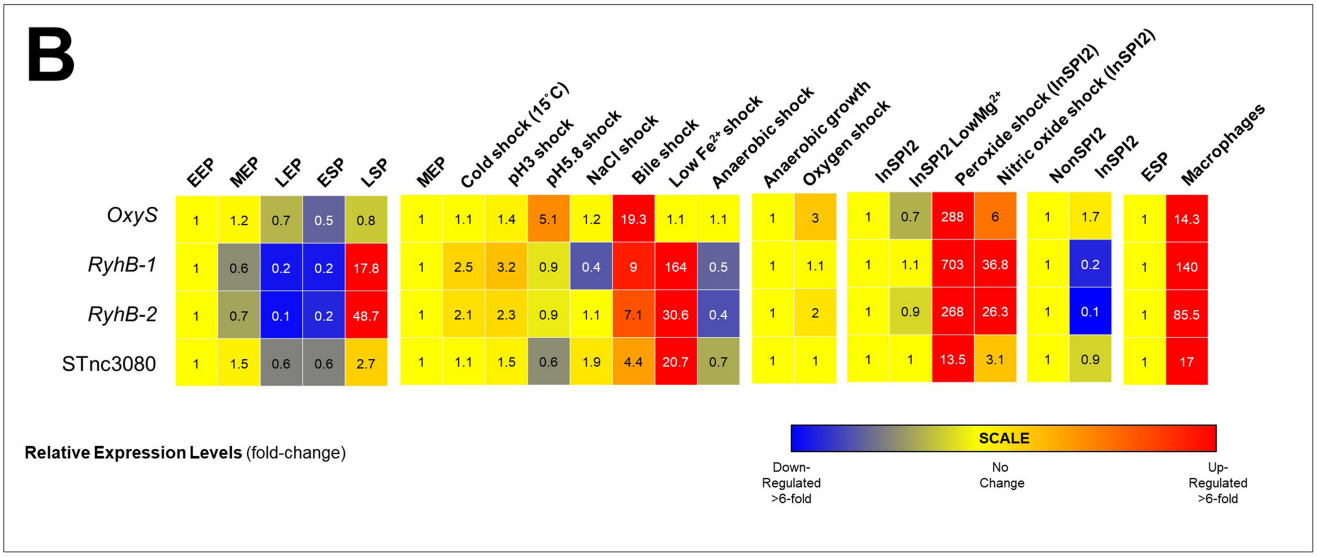
The availability of comprehensive RNA-seq data sets, coupled with the work of many outstanding geneticists and RNA biologists, has made *S. enterica* a model organism for the study of RNA-mediated regulation. The pathogen *S.* Typhimurium is responsible for hundreds of thousands of deaths each year, largely caused by systemic infections (38). RNA-seq-based transcriptomics involving *in vitro* infection-relevant growth conditions have identified characteristic transcriptional signatures (35). To determine whether *in vitro* transcriptional patterns are relevant to intracellular infection, bacterial RNA has been isolated from murine macrophages infected with *S.* Typhimurium and used for RNA-seq and functional transcriptomic analysis (36). These data allow the transcriptional differences observed under *in vitro* conditions to be related to the infection of mammalian cells.

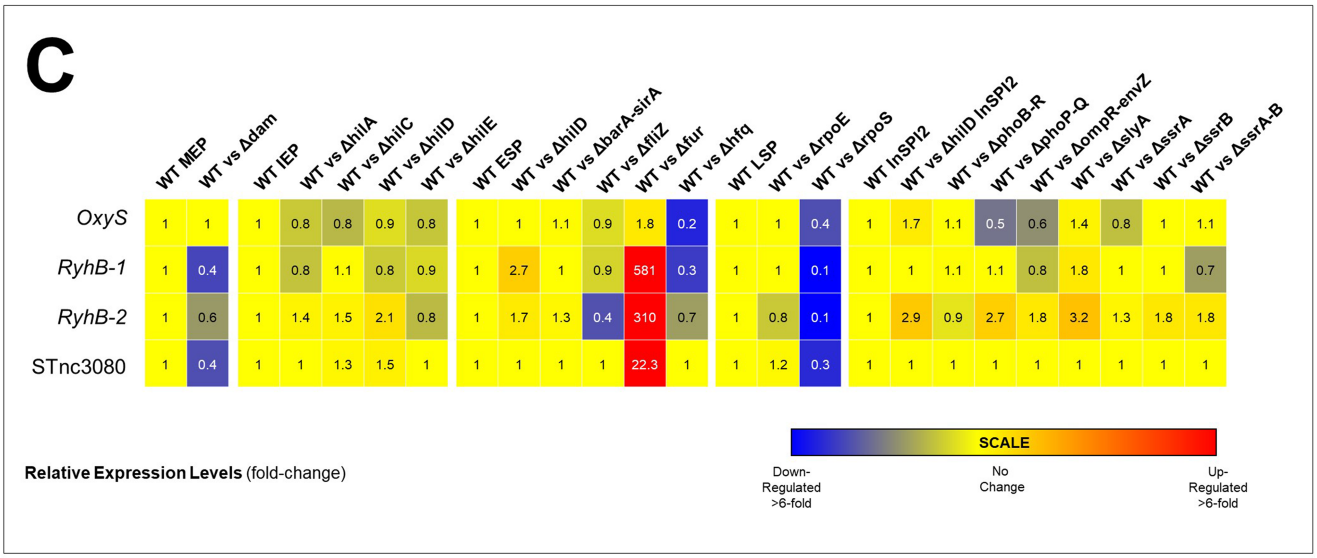
One way to address gene function is to integrate transcriptomic data with global transposon mutagenesis of bacterial pathogens during infection, a concept that has been discussed previously (39). These types of experiments can answer the crucial question: Are these transcripts biologically relevant? (Fig. 1).

As well as investigating the effect of environmental stress upon *Salmonella* gene expression, we also studied the impact of 18 regulatory proteins upon expression of all coding genes and sRNAs. These data are available in a separate online resource, SalComRegulon (http://tinyurl.com/SalComRegulon), that presents transcriptomic data for a selection of mutants that lack transcriptional activators, repressors, sigma factors, and the Hfq RNA chaperone, reflecting both direct and indirect regulatory effects (37).

Following "traditional" data analysis, heat maps of unfiltered gene lists can be a valuable tool for subsequent







transcriptomic data exploration. Figures 2, 3, and 4 show the impact of 21 environmental conditions and key regulatory factors upon the expression of three groups of sRNAs. In each of these three figures, panel A shows the "absolute" level of expression of each gene, in TPM (transcripts per million) units, which have been defined elsewhere (36). A TPM value of <10 indicates that a particular gene is not expressed, and the color scale (yellow through orange to red) indicates the level of expression. Panel B shows the "relative" level of expression between a stress environment and a control culture. The use of relative values allows the effect of environmental stress upon gene expression to be visualized as distinct colored patterns. Panel C also uses a "relative" approach that highlights differences in gene expression between wild-type and mutant bacteria cultivated in particular growth conditions.

Figure 2 shows sRNAs that respond to low iron levels or oxidative stress. OxyS was one of the first sRNAs found to be environmentally regulated (40), being induced by almost 300-fold upon exposure to hydrogen peroxide (peroxide) (Fig. 2B). RyhB is a key iron-regulated sRNA, described in reference 41. *S.* Typhimurium carries two RyhB paralogs called RyhB-1 and RyhB-2 (42). The RyhB-1 sRNA shares extensive sequence similarity to the *E. coli* RyhB ortholog, and has been reported to be highly induced by iron starvation, upregulated by peroxide, and activated by OxyR (43, 44). Consistent with this, the SalComMac data show that RyhB-1 is induced by 165-fold in the low-iron environment and upregulated by 700-fold in response to peroxide. In contrast, the RyhB-2 paralog, which is more distantly related to *E. coli* RyhB, is only upregulated by 30-fold in the low-iron environment and is peroxide-induced at a lower level than RyhB-1. These findings are consistent with the literature (see reference 41). Figure 2B shows that the levels of expression of RyhB-1 and RyhB-2 also increase in response to treatment with nitric oxide and bile and during the infection of macrophages (Fig. 2B). These new findings could be worthy of study in the future.

STnc3080 is an *S.* Typhimurium sRNA that has previously been shown to be upregulated in a low-iron environment (36), and here we observe that STnc3080 is induced by peroxide (Fig. 2B). The impact of various regulatory factors upon the expression of OxyS, RyhB-1, RyhB-2, and STnc3080 is apparent (Fig. 2C). The most significant regulatory input appears to be the action of Fur to repress expression of the three iron-responsive sRNAs. The *E. coli* RyhB has previously been shown to be Fur repressed (45), and this is the first time that a role for Fur in the expression of STnc3080 has been suggested. Expression of both RyhB-1 and RyhB-2 is reduced in the absence of RpoS, suggesting that the σ^{38} sigma factor could activate expression of the two paralogs.

Figure 3 focuses on the expression of four sRNAs involved in the envelope stress response, namely RybB, RyeF, MicA, and RprA. Figure 3A shows that three of the four sRNAs are expressed in all environmental conditions, at relatively high levels. In contrast, RyeF is only expressed in 4 of the 21 conditions at medium to high levels. Figure 3B shows that three of the sRNAs are induced in a low-magnesium environment, a condition known to induce bacterial envelope stress (46). It is likely that the upregulation of the same three sRNAs during macrophage infection also reflects an envelope stress response. Two of the envelope stress-associated sRNAs were osmoinducible, with MicA and RprA showing between 15- and 21-fold upregulation following a 10-min exposure to 0.3 M NaCl. The RyeF (MicL) and MicA sRNAs were significantly upregulated following cold shock at 15°C, which, to our knowledge, has not been reported previously.

FIGURE 2 Environmental and genetic regulation of four sRNAs that are iron responsive and/or induced by oxidative stress. Gene expression data are presented for the sRNAs OxyS, RyhB-1, RyhB-2, and STnc3080 (these data can be visualized online at https://tinyurl.com/ya7s466m and https://tinyurl.com/yb5wz7dt). Data are shown as differential expression profiles involving six discrete heat-map blocks, each block being normalized to the condition on the left-hand side. The heat maps show differential expression, a strategy that lacks accuracy when expression levels are extremely low. Absolute (A) and relative (B) expression levels of *S.* Typhimurium grown under 21 different conditions (SalComMac). (C) Relative expression levels of the wild-type (WT) and mutant *S.* Typhimurium 4/74 grown under different conditions (SalComRegulon). Before experimental validation is considered, it should be ensured that the levels of absolute expression of particular sRNAs are above the expression threshold of 10 TPM units (35–37). EEP, early exponential phase; MEP, mid-exponential phase; LEP, late exponential phase; ESP, early stationary phase; LSP, late stationary phase; InSPI2, SPI2-inducing minimal media.

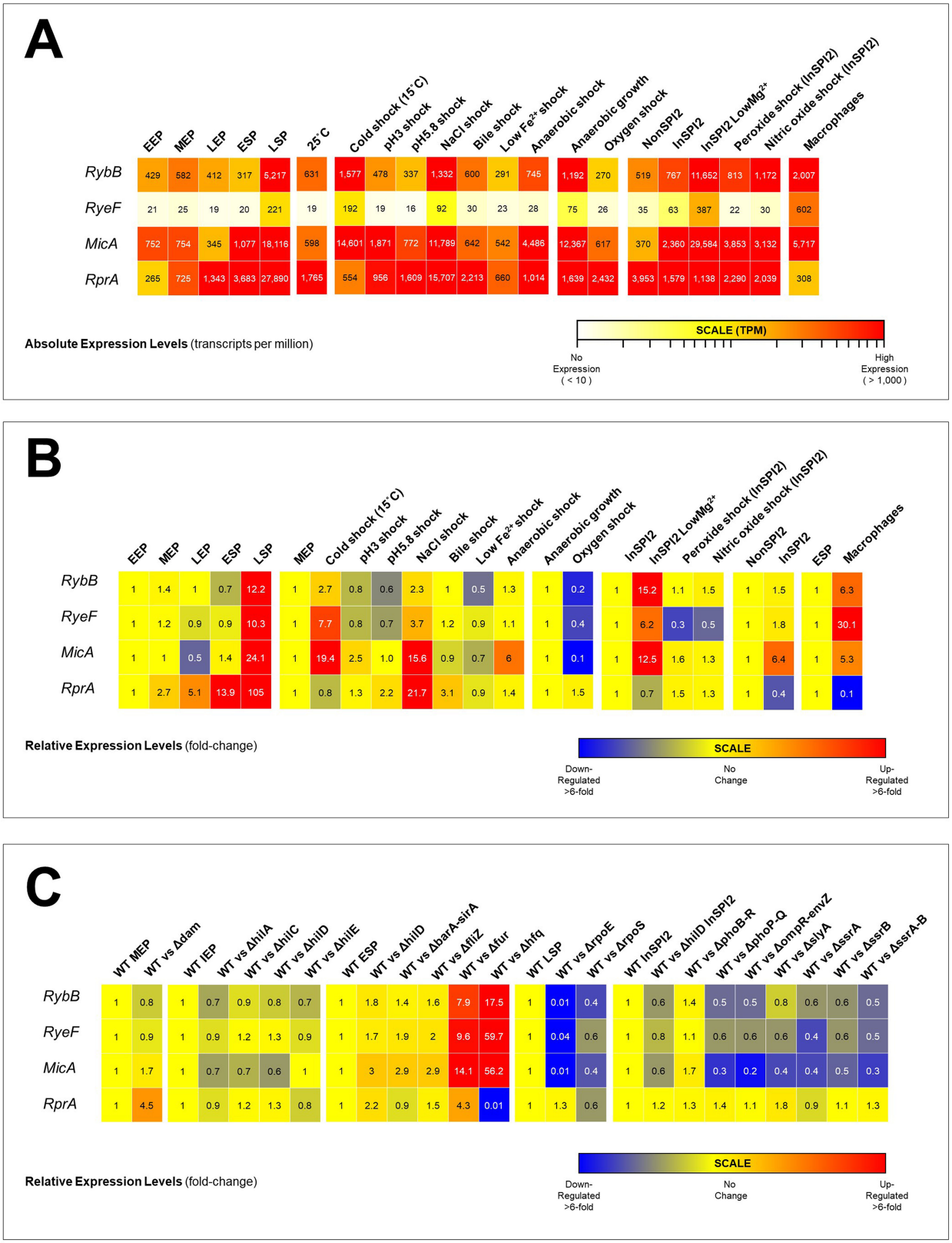


FIGURE 3 Environmental and genetic regulation of four sRNAs involved in the envelope stress response. Gene expression data are shown for the sRNAs RybB, RyeF, MicA, and RprA (these data can be visualized online at https://tinyurl.com/y9mskb6j and https://tinyurl.com/ybnr6jja). Panels A, B, and C are as described in the legend to Fig. 2.

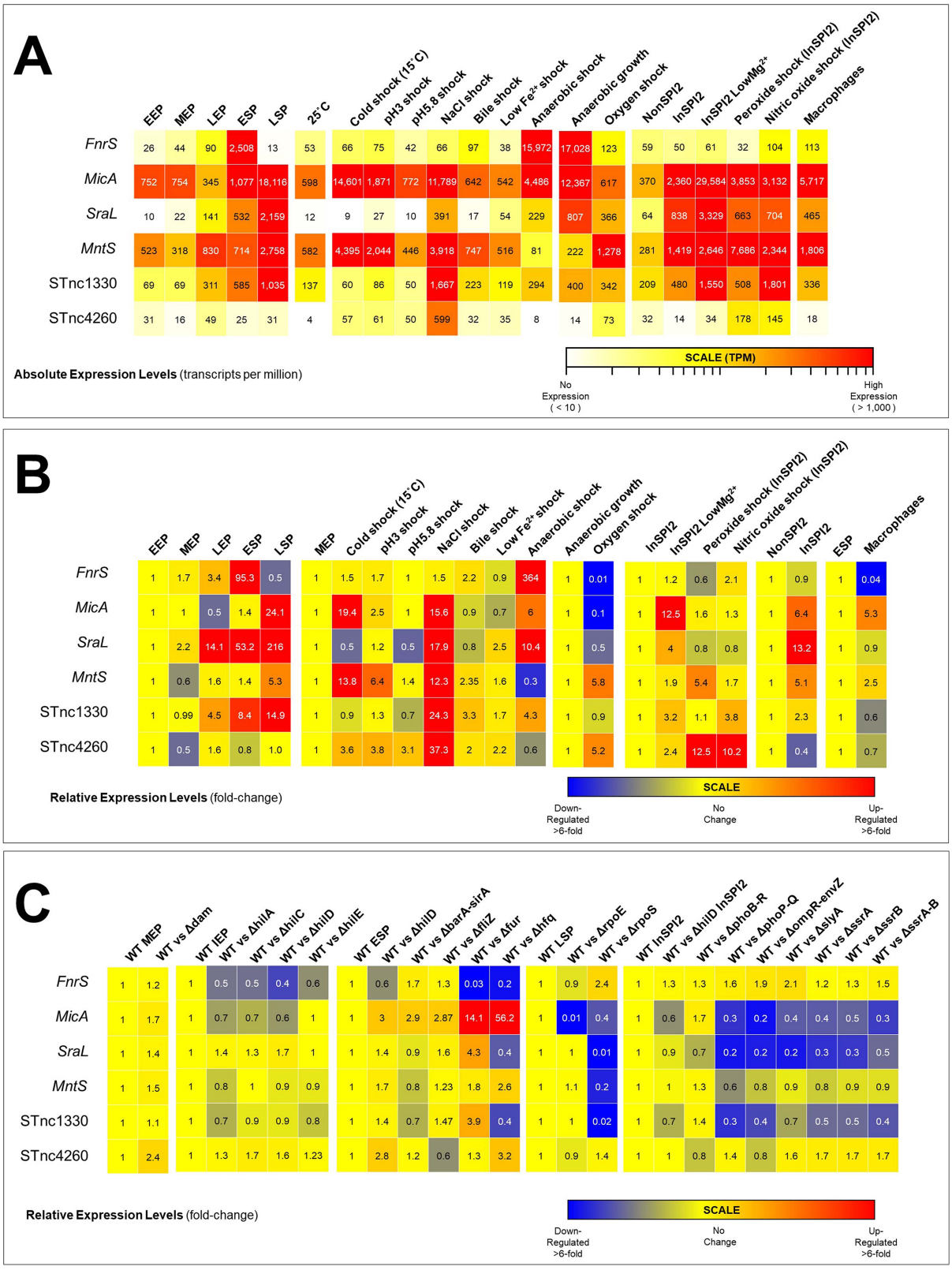


FIGURE 4 Environmental and genetic regulation of six sRNAs that respond to oxygen or osmolarity. Gene expression data are shown for the sRNAs FnrS, MicA, SraL, MntS (RybA), STnc1330, and STnc4260 (these data can be visualized online at https://tinyurl.com/yat8qrql and https://tinyurl.com/y8f533gy). Panels A, B, and C are as described in the legend to Fig. 2.

In terms of transcriptional inputs, the SalCom Regulon data suggest that RybB, RyeF, and MicA are activated by RpoE and are repressed by Fur (Fig. 3C). Activation of MicA and RyeF (MicL) by RpoE is consistent with the literature (46, 47). The data summarized in Fig. 3C raise the possibility of cross talk between PhoPQ/OmpR-EnvZ and MicA, and it is noteworthy that the *phoP* gene is a confirmed target of MicA (48).

Figure 4 presents the expression profiles of six sRNAs that respond to oxygen or osmolarity. MicA is included in both Fig. 3 and 4 for comparative purposes. Two of the sRNAs, MicA and MntS (RybA), are expressed at medium to high levels in all environmental conditions. In contrast, FnrS is only upregulated under conditions of oxygen limitation. MicA and SraL also show a low level of upregulation under anaerobic conditions. Four of the sRNAs are osmoregulated, showing between 12- and 37-fold induction following a 10-min exposure to 0.3 M NaCl (Fig. 4B). As well as being osmoregulated, MntS is induced by cold shock and by pH 3 shock, which could have implications for manganese homeostasis (49).

The SraL sRNA is conserved in many enteric bacteria and is upregulated when *S.* Typhimurium is grown under conditions that induce expression of the SPI2 pathogenicity island (50). As well as visualizing this finding, Fig. 4B shows that SraL is both osmoregulated and induced under anaerobic conditions. Because SraL has not been studied extensively in *Salmonella*, SalComRegulon adds a useful perspective (Fig. 4C). The data suggest that RpoS, PhoPQ, and OmpR-EnvZ activate SraL, ideas that need to be tested experimentally. In contrast, MntS shows a distinct pattern of expression and does not appear to be RpoE dependent. The Hfq chaperone does not bind to the MntS transcript (51), and so is unlikely to play a stabilizing role.

Figure 4 also shows the expression profiles of two uncharacterized *S.* Typhimurium sRNAs that are highly osmoregulated (STnc1330 and STnc4260). The SalComRegulon data suggest that STnc1330 has an RpoS-activated, PhoPQ-activated pattern of expression (Fig. 4C). Taken together, the panels of Fig. 4 show that sRNAs that respond to osmolarity and/or anaerobic growth are controlled by a variety of transcriptional inputs.

As well as facilitating data visualization as heat maps, SalComMac and SalComRegulon can also show RNA-seq sequence reads in a genomic context. This feature can be accessed via the "view in JBrowse" option on the SalComMac and SalComRegulon Web pages, which is located between the Absolute and Relative gene expression panels.

JBrowse is an open-source, JavaScript-based genome browser that is fast and zoomable, ideal for the analysis of multiple sets of mapped transcriptomic sequence reads (52, 53). The JBrowse feature of SalComMac and SalComRegulon allows transcripts of interest to be seen in the context of neighboring genes and TSSs. For example, the transcript of STnc1330, an *S.* Typhimurium 141-nucleotide sRNA (54), can be seen in the context of neighboring TSSs. Figure 5 shows the appearance of the STnc1330 transcript during growth in different environmental conditions; STnc1330 is expressed at low levels during growth at mid-exponential phase, and is highly abundant following osmotic or anaerobic shock (Fig. 5A). STnc1330 is also expressed in SPI2-inducing minimal media and at increased levels under magnesium limitation (Fig. 5B). The impact of regulatory factors is apparent in Fig. 5C and D, which show that PhoPQ and RpoS are required for expression.

THE USE OF EXPRESSION DATA TO INFER REGULATORY INTERACTIONS REQUIRES CAUTION!

Genome-scale expression data always have caveats. It should be noted that monitoring the levels of individual mRNAs and sRNAs by conventional transcriptomics shows the average expression profiles in a large population of bacteria. Furthermore, the expression level of an individual transcript represents a combination of the positive impact of transcriptional initiation and elongation versus the negative effect of RNA decay. Consequently, the stability of individual transcripts can influence global gene expression levels.

Some of the transcriptomic data shown in SalComMac and SalComRegulon are based on RNA harvested between 10 and 30 min after an environmental perturbation (36). As environmental shocks were studied at a single time point, and individual stressors are just used at one concentration, these experimental conditions can differ from those used in other publications. Some coding genes and sRNAs may not show the "expected" patterns of expression. Despite these caveats, the expression of most well-known genes in SalComMac and SalComRegulon is consistent with the literature (36).

CONCLUSIONS AND FUTURE DIRECTIONS

We have come a long way since the first bacterial transcriptomic experiment was reported (12). Nowadays, many laboratories are generating informative RNA-seq data, but key interpretive challenges remain to be solved.

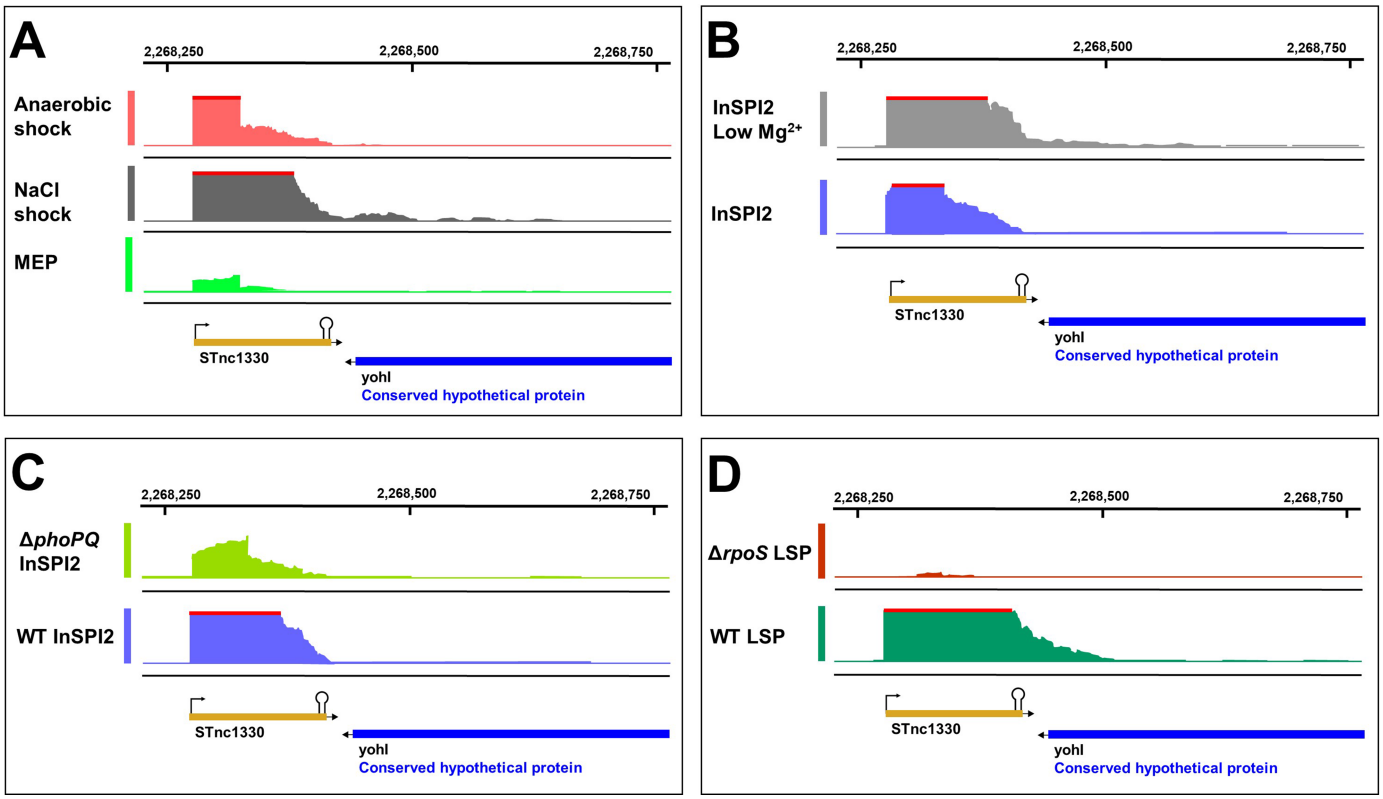


FIGURE 5 Visualization of the STnc1330 sRNA transcript. RNA-seq reads are mapped to the *S.* Typhimurium 4/74 genome (plus strand), showing STnc1330 expression under different conditions (35, 37). (A) MEP, anaerobic shock, and NaCl shock (https://tinyurl.com/STnc1330-NaCl); (B) InSPI2 and InSPI2 low $Mg2^{+}$ (https://tinyurl.com/STnc1330-LowMg); (C) WT InSPI2 versus Δ*phoPQ* (https://tinyurl.com/STnc1330-PhoPQ); (D) WT LSP versus Δ*rpoS* (https://tinyurl.com/STnc1330-RpoS). Height of colored tracks represents the normalized sequencing reads at that locus (scale, 0 to 100). All arrows indicate the direction of transcription; TSSs are indicated by bent arrows and predicted Rho (ρ)-independent terminators are denoted by stem-loop structures.

Making the microbiology community's valuable gene expression data available for intuitive interrogation by current and future generations of microbiologists must be a priority. Currently, a variety of online resources are available for a small selection of model organisms. We need to improve and optimize these resources to ensure that the availability of crucial data sets will be maintained in the future, and not depend exclusively on particular scientists or laboratories.

Transcriptomic experiments can yield surprising biological insights. Recently, we showed that in African *S.* Typhimurium ST313, one nucleotide change was responsible for the upregulation of the *pgtE* transcript and the PgtE protein. This single regulatory difference accounted for the ability of the African *Salmonella* to resist killing by human serum (55), a discovery that was made by the comparison of transcriptomes of two pathovariants of *S.* Typhimurium.

For the future, a new era of comparative transcriptomics will necessitate moving beyond the analysis of individual bacterial strains. To understand key host-specificity phenotypes, the comparative transcriptomic analysis of closely related bacterial pathogens that have different lifestyles will be necessary. The continued development of tools to visualize and interrogate gene expression data will further our understanding of bacterial functional transcriptomics and, hopefully, inspire a new generation of gene detectives.

ACKNOWLEDGMENTS

We are grateful to many people for contributing to the visualization of *Salmonella* transcriptomic data over recent years, particularly members of the Hinton lab (including Aoife Colgan, Rocío Canals, Carsten Kröger, and Nicolas Wenner), Tyrrell Conway, and Karsten Hokamp. This work was supported by funding from a Wellcome Trust Senior Investigator award (to J.C.D.H.) (Grant 106914/Z/15/Z).

REFERENCES

1. MacLean D, Jones JD, Studholme DJ. 2009. Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* **7:**287–296.

2. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE. 2015. Big data: astronomical or genomical? *PLoS Biol* **13:**e1002195. http://dx.doi.org/10.1371/journal.pbio.1002195.

3. Wang R, Perez-Riverol Y, Hermjakob H, Vizcaíno JA. 2015. Open source libraries and frameworks for biological data visualisation: a guide for developers. *Proteomics* **15:**1356–1374. http://dx.doi.org/10.1002/pmic.201400377.

4. Toker L, Feng M, Pavlidis P. 2016. Whose sample is it anyway? Widespread misannotation of samples in transcriptomics studies. *F1000 Res* **5:**2103. http://dx.doi.org/10.12688/f1000research.9471.1.

5. Heiss JA, Just AC. 2018. Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. *Clin Epigenetics* **10:**73. http://dx.doi.org/10.1186/s13148-018-0504-1.

6. Bécavin C, Koutero M, Tchitchek N, Cerutti F, Lechat P, Maillet N, Hoede C, Chiapello H, Gaspin C, Cossart P. 2017. Listeriomics: an interactive web platform for systems biology of *Listeria*. *mSystems* **2:**e00186-e16. http://dx.doi.org/10.1128/mSystems.00186-16.

7. Slager J, Aprianto R, Veening JW. 2018. Deep genome annotation of the opportunistic human pathogen *Streptococcus pneumoniae* D39. *Nucleic Acids Res* doi:10.1093/nar/gky725.

8. Kröger C, MacKenzie KD, Alshabib EY, Kirzinger MW, Suchan DM, Chao TC, Akulova V, Miranda-CasoLuengo AA, Monzon VA, Conway T, Sivasankaran SK, Hinton JC, Hokamp K, Cameron AD. The primary transcriptome, small RNAs, and regulation of antimicrobial resistance in *Acinetobacter baumannii* ATCC 17978. *Nucleic Acids Res* doi:10.1093/nar/gky603.

9. Ilyas B, Tsai CN, Coombes BK. 2017. Evolution of *Salmonella*-host cell interactions through a dynamic bacterial genome. *Front Cell Infect Microbiol* **7:**428. http://dx.doi.org/10.3389/fcimb.2017.00428.

10. Mekalanos JJ. 1992. Environmental signals controlling expression of virulence determinants in bacteria. *J Bacteriol* **174:**1–7. http://dx.doi.org/10.1128/jb.174.1.1-7.1992.

11. Silhavy TJ. 2000. Gene fusions. *J Bacteriol* **182:**5935–5938. http://dx.doi.org/10.1128/JB.182.21.5935-5938.2000.

12. DeRisi JL, Iyer VR, Brown PO. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278:**680–686. http://dx.doi.org/10.1126/science.278.5338.680.

13. Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:**467–470. http://dx.doi.org/10.1126/science.270.5235.467.

14. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320:**1344–1349. http://dx.doi.org/10.1126/science.1158441.

15. Colgan AM, Cameron AD, Kröger C, Colgan AM, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, Canals R, Grissom JE, Conway T, Hokamp K, Hinton JC. 2017. If it transcribes, we can sequence it: mining the complexities of host-pathogen-environment interactions using RNA-seq. *Curr Opin Microbiol* **36:**37–46. http://dx.doi.org/10.1016/j.mib.2017.01.010.

16. Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10:**57–63 http://dx.doi.org/10.1038/nrg2484.

17. Mäder U, Nicolas P, Richard H, Bessières P, Aymerich S. 2011. Comprehensive identification and quantification of microbial transcriptomes by genome-wide unbiased methods. *Curr Opin Biotechnol* **22:**32–41. http://dx.doi.org/10.1016/j.copbio.2010.10.003.

18. Höfer K, Jäschke A. 2018. Epitranscriptomics: RNA modifications in bacteria and archaea. *Microbiol Spectr* **6:**RWR-0015-2017. http://dx.doi.org/10.1128/microbiolspec.RWR-0015-2017.

19. Saliba AE, C Santos S, Vogel J. 2017. New RNA-seq approaches for the study of bacterial pathogens. *Curr Opin Microbiol* **35:**78–87. http://dx.doi.org/10.1016/j.mib.2017.01.001.

20. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szcześniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol* **17:**1–19.

21. Aikawa C, Maruyama F, Nakagawa I. 2010. The dawning era of comprehensive transcriptome analysis in cellular microbiology. *Front Microbiol* **1:**118. http://dx.doi.org/10.3389/fmicb.2010.00118.

22. Creecy JP, Conway T. 2015. Quantitative bacterial transcriptomics with RNA-seq. *Curr Opin Microbiol* **23:**133–140. http://dx.doi.org/10.1016/j.mib.2014.11.011.

23. Levin JZ, Yassour M, Adiconis X, Nusbaum C, Thompson DA, Friedman N, Gnirke A, Regev A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat Methods* **7:**709–715. http://dx.doi.org/10.1038/nmeth.1491.

24. The GTEx Consortium. 2013. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45:**580–585. http://dx.doi.org/10.1038/ng.2653.

25. Faria JP, Overbeek R, Xia F, Rocha M, Rocha I, Henry CS. 2014. Genome-scale bacterial transcriptional regulatory networks: reconstruction and integrated analysis with metabolic models. *Brief Bioinform* **15:**592–611. http://dx.doi.org/10.1093/bib/bbs071.

26. Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, Hinton JC. 2006. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog* **2:**e81. http://dx.doi.org/10.1371/journal.ppat.0020081.

27. Smith C, Stringer AM, Mao C, Palumbo MJ, Wade JT. 2016. Mapping the regulatory network for *Salmonella enterica* serovar Typhimurium invasion. *mBio* **7:**e01024-16. http://dx.doi.org/10.1128/mBio.01024-16.

28. Tomljenovic-Berube AM, Mulder DT, Whiteside MD, Brinkman FS, Coombes BK. 2010. Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB two-component system. *PLoS Genet* **6:**e1000875. http://dx.doi.org/10.1371/journal.pgen.1000875.

29. Cloots L, Marchal K. 2011. Network-based functional modeling of genomics, transcriptomics and metabolism in bacteria. *Curr Opin Microbiol* **14:**599–607. http://dx.doi.org/10.1016/j.mib.2011.09.003.

30. McDermott JE, Yoon H, Nakayasu ES, Metz TO, Hyduke DR, Kidwai AS, Palsson BO, Adkins JN, Heffron F. 2011. Technologies and approaches to elucidate and model the virulence program of *Salmonella*. *Front Microbiol* **2:**121. http://dx.doi.org/10.3389/fmicb.2011.00121.

31. Yoon H, McDermott JE, Porwollik S, McClelland M, Heffron F. 2009. Coordinated regulation of virulence during systemic infection of *Salmonella enterica* serovar Typhimurium. *PLoS Pathog* **5:**e1000306. http://dx.doi.org/10.1371/journal.ppat.1000306.

32. Métris A, Sudhakar P, Fazekas D, Demeter A, Ari E, Olbei M, Branchu P, Kingsley RA, Baranyi J, Korcsmáros T. 2017. SalmoNet, an integrated network of ten *Salmonella enterica* strains reveals common and distinct pathways to host adaptation. *NPJ Syst Biol Appl* **3:**31. http://dx.doi.org/10.1038/s41540-017-0034-z.

33. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, der Maaten L, Vlassis N, Wilmes P. 2015. VizBin—an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* **3:**1–7. http://dx.doi.org/10.1186/s40168-014-0066-1.

34. Barquist L, Vogel J. 2015. Accelerating discovery and functional analysis of small RNAs with new technologies. *Annu Rev Genet* **49:**367–394. http://dx.doi.org/10.1146/annurev-genet-112414-054804.

35. Srikumar S, Kröger C, Hébrard M, Colgan A, Owen SV, Sivasankaran SK, Cameron AD, Hokamp K, Hinton JC. 2015. RNA-seq brings new insights to the intra-macrophage transcriptome of *Salmonella* Typhimurium. *PLoS Pathog* **11:**e1005262. http://dx.doi.org/10.1371/journal.ppat.1005262.

36. Kröger C, Colgan A, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, Canals R, Grissom JE, Conway T, Hokamp K, Hinton JC. 2013. An infection-relevant transcriptomic compendium for *Salmonella enterica* serovar Typhimurium. *Cell Host Microbe* **14:**683–695. http://dx.doi.org/10.1016/j.chom.2013.11.010.

37. Colgan AM, Kröger C, Diard M, Hardt WD, Puente JL, Sivasankaran SK, Hokamp K, Hinton JC. 2016. The impact of 18 ancestral and horizontally-acquired regulatory proteins upon the transcriptome and sRNA landscape of *Salmonella enterica* serovar Typhimurium. *PLoS Genet* **12:**e1006258. http://dx.doi.org/10.1371/journal.pgen.1006258.

38. Havelaar AH, Kirk MD, Torgerson PR, Gibb HJ, Hald T, Lake RJ, Praet N, Bellinger DC, de Silva NR, Gargouri N, Speybroeck N, Cawthorne A, Mathers C, Stein C, Angulo FJ, Devleesschauwer B, World Health Organization Foodborne Disease Burden Epidemiology Reference Group. 2015. World Health Organization global estimates and regional comparisons of the burden of foodborne disease in 2010. *PLoS Med* **12:** e1001923. http://dx.doi.org/10.1371/journal.pmed.1001923.

39. Hammarlöf DL, Canals R, Hinton JC. 2013. The FUN of identifying gene function in bacterial pathogens; insights from *Salmonella* functional genomics. *Curr Opin Microbiol* **16:**643–651. http://dx.doi.org/10.1016/j.mib.2013.07.009.

40. Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G. 1997. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* **90:**43–53. http://dx.doi.org/10.1016/S0092-8674(00)80312-8.

41. Chareyre S, Mandin P. 2018. Bacterial iron homeostasis regulation by sRNAs. *Microbiol Spectr* **6:**RWR-0010-2017. http://dx.doi.org/10.1128/microbiolspec.RWR-0010-2017.

42. Kim JN. 2016. Roles of two RyhB paralogs in the physiology of *Salmonella enterica*. *Microbiol Res* **186-187:**146–152. http://dx.doi.org/10.1016/j.micres.2016.04.004.

43. Calderón IL, Morales EH, Collao B, Calderón PF, Chahuán CA, Acuña LG, Gil F, Saavedra CP. 2014. Role of *Salmonella* Typhimurium small RNAs RyhB-1 and RyhB-2 in the oxidative stress response. *Res Microbiol* **165:**30–40. http://dx.doi.org/10.1016/j.resmic.2013.10.008.

44. Padalon-Brauch G, Hershberg R, Elgrably-Weiss M, Baruch K, Rosenshine I, Margalit H, Altuvia S. 2008. Small RNAs encoded within genetic islands of *Salmonella* Typhimurium show host-induced expression and role in virulence. *Nucleic Acids Res* **36:**1913–1927. http://dx.doi.org/10.1093/nar/gkn050.

45. Massé E, Gottesman S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* **99:**4620–4625. http://dx.doi.org/10.1073/pnas.032066599.

46. Coornaert A, Lu A, Mandin P, Springer M, Gottesman S, Guillier M. 2010. MicA sRNA links the PhoP regulon to cell envelope stress. *Mol Microbiol* **76:**467–479. http://dx.doi.org/10.1111/j.1365-2958.2010.07115.x.

47. Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, Storz G. 2014. MicL, a new σ^E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes Dev* **28:**1620–1634. http://dx.doi.org/10.1101/gad.243485.114.

48. Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA. 2011. Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proc Natl Acad Sci U S A* **108:**12875–12880. http://dx.doi.org/10.1073/pnas.1109379108.

49. Waters LS, Sandoval M, Storz G. 2011. The *Escherichia coli* MntR miniregulon includes genes encoding a small protein and an efflux pump required for manganese homeostasis. *J Bacteriol* **193:**5887–5897. http://dx.doi.org/10.1128/JB.05872-11.

50. Silva IJ, Ortega ÁD, Viegas SC, García-Del Portillo F, Arraiano CM. 2013. An RpoS-dependent sRNA regulates the expression of a chaperone involved in protein folding. *RNA* **19:**1253–1265. http://dx.doi.org/10.1261/rna.039537.113.

51. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. 2012. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* **31:**4005–4019. http://dx.doi.org/10.1038/emboj.2012.229.

52. Skinner ME, Uzilov AV, Stein LD, Mungall CJ, Holmes IH. 2009. JBrowse: a next-generation genome browser. *Genome Res* **19:**1630–1638. http://dx.doi.org/10.1101/gr.094607.109.

53. Westesson O, Skinner M, Holmes I. 2013. Visualizing next-generation sequencing data with JBrowse. *Brief Bioinform* **14:**172–177. http://dx.doi.org/10.1093/bib/bbr078.

54. Kröger C, Dillon SC, Cameron AD, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hébrard M, Händler K, Colgan A, Leekitcharoenphon P, Langridge GC, Lohan AJ, Loftus B, Lucchini S, Ussery DW, Dorman CJ, Thomson NR, Vogel J, Hinton JC. 2012. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci U S A* **109:**E1277–E1286. http://dx.doi.org/10.1073/pnas.1201061109.

55. Hammarlöf DL, Kröger C, Owen SV, Canals R, Lacharme-Lora L, Wenner N, Schager AE, Wells TJ, Henderson IR, Wigley P, Hokamp K, Feasey NA, Gordon MA, Hinton JC. 2018. Role of a single noncoding nucleotide in the evolution of an epidemic African clade of *Salmonella*. *Proc Natl Acad Sci U S A* **115:**E2614–E2623. http://dx.doi.org/10.1073/pnas.1714718115.