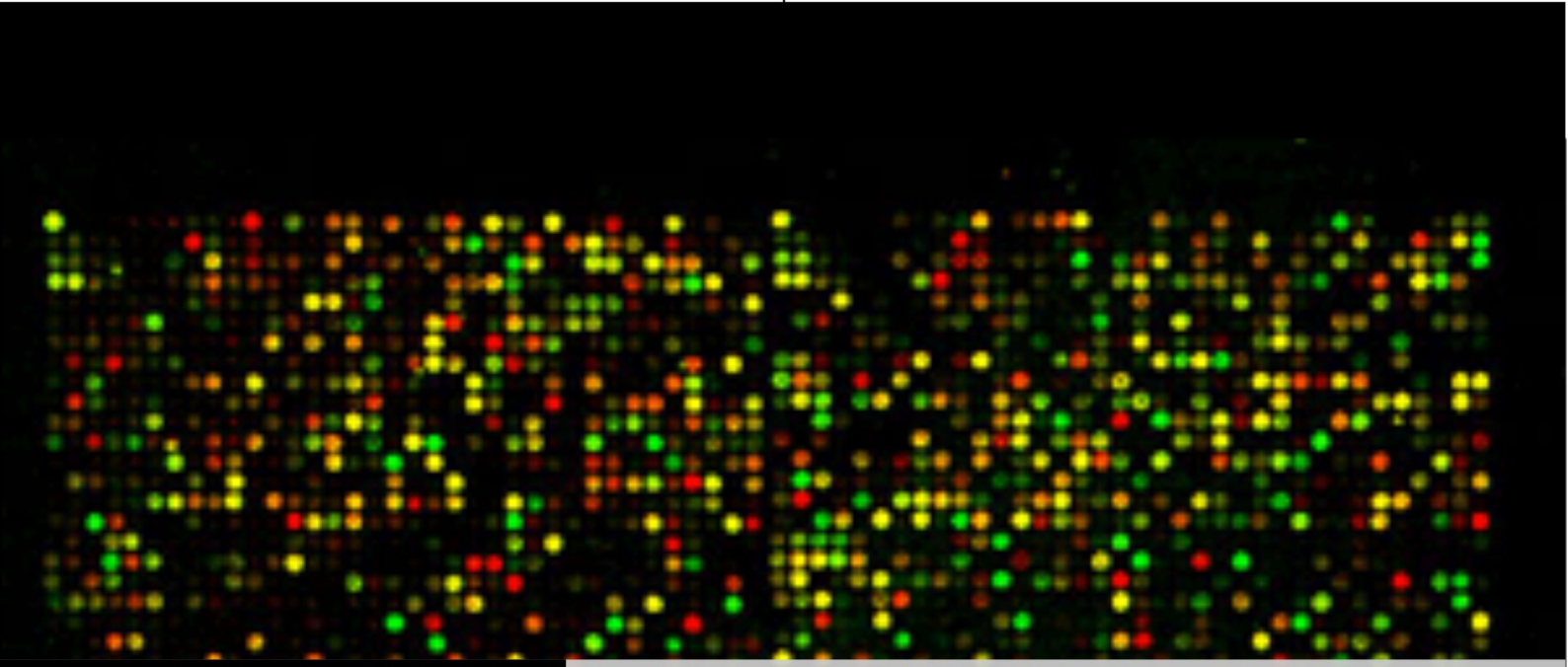
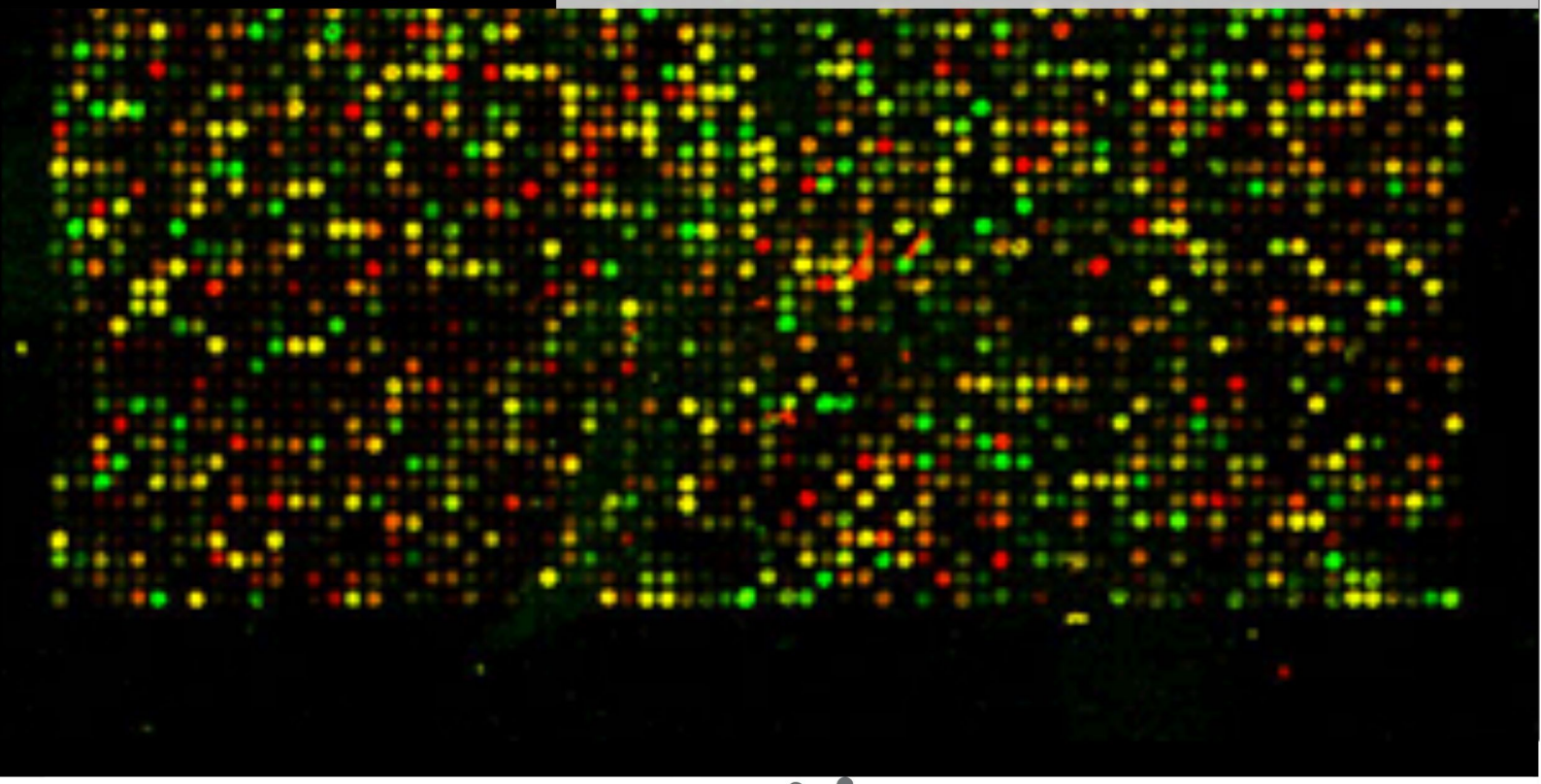


TESIS DOCTORAL
Olatz Ruiz-Larrabeiti
Leioa, 2016

Discovery and study of small RNAs in *Escherichia coli* using custom microarrays and next generation sequencing



eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

ZTF-FCT
Zientzia eta Teknologia Fakultatea
Facultad de Ciencia y Tecnología

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

ZTF-FCT
Zientzia eta Teknologia Fakultatea
Facultad de Ciencia y Tecnología

Departamento de Inmunología, Microbiología y Parasitología
Immunologia, Mikrobiologia y Parasitologia Saila
Department of Immunology, Microbiology, and Parasitology

Grupo de Respuesta Bacteriana al Estrés

Tesis Doctoral

Discovery and study of small RNAs in *Escherichia coli* using custom microarrays and next generation sequencing

Olatz Ruiz Larrabeiti

Supervisor
Dr. Vladimir Kaberdin

This study was funded by Spanish Ministry of Economy and Competitiveness grant BFU2011-25455, IKERBASQUE foundation, and the University of Basque country (EHU/UPV).

During this study I was funded by pre-doctoral grant from the University of the Basque Country (UPV/EHU).

Nada de lo que sucede se olvida jamás,
aunque tú no puedas recordarlo»

Sen to Chihiro no Kamikakushi,
Hayao Miyazaki

A mis padres.

Agradecimientos

Primero, quiero dar las gracias a mi supervisor, el Dr. Vladimir Kaberdin, por su paciencia, su perseverancia y su entrega, y por enseñarme por medio de su ejemplo. También quiero agradecer mi tutora, la Dra. Inés Arana, por mostrarse siempre disponible y ayudarme en cada ocasión que ha tenido.

Gracias a todo el profesorado del departamento, por haber hecho siempre lo posible por ayudarme, así como a Jose Ignacio Ruiz Sanz por facilitarme el uso del equipo del Laboratorio de Fisiología y por su atención, así como a los investigadores del dicho laboratorio, por su comprensión y su ayuda.

I would also like to thank Dr. Kenneth McDowall, Dr. Justin Clarke and Dr. Ayad Hassan for all their help, their support and their generosity during my visit to Leeds. Thanks to all the friends I met in the laboratory at University of Leeds, for welcoming me and making me feel at home.

Je tiens également à exprimer ma profonde gratitude à Dr. Eliane Hajnsdorf, à Céline Gracia, à Dr. Fanette Fontaine et à Joel Caillet pour leur aide, leurs conseils et leur soutien pendant mon séjour au laboratoire de l'IBPC à Paris. Notre collaboration m'est d'une grande valeur. Je garde de très bons souvenirs de tous mes amis à l'Institut.

Gracias a Montse y a Josemi por brindarme todas las oportunidades. Gracias a Aiala y a Eire, que son un motivo de orgullo sin fin. Gracias a Kepa por su apoyo infinito y su fe en mí. Gracias a mi Laia por todo su amor.

Gracias a mis hermanos científicos, Mikel e Itxaso, por serlo. Gracias por vuestro respeto, vuestra paciencia, y vuestro compañerismo. Para mí es un honor haber compartido esta experiencia con vosotros. También quiero agradecer a mis compañeros, David, Claudia y Zaloa, por haberme prestado su ayuda siempre que han podido. La convivencia en el laboratorio no es fácil, por eso gracias a todos por crear un ambiente en el que caben la alegría y la redención.

Quiero agradecer a todas las personas con las que he tenido la suerte de coincidir estos años en nuestro laboratorio y en el departamento. Son unos cuantos años y mucha gente. Me llevo recuerdos preciosos de este tiempo.

Quiero agradecer en particular a mis amigas Ainhoa, Naiara y Aitziber, que me han animado en cada bache y eran las primeras en celebrar mis pequeños triunfos. Gracias por todos los momentos que hemos compartido, y gracias por vuestra amistad. Gracias a mis amigas Irantzu y Esti, por su apoyo constante, y su cariño incondicional.

Me siento muy afortunada de decir que nombrar a cada una de las personas con las que he compartido este viaje y a cuya amistad debo gran parte de la fuerza que me ha traído hasta aquí ocuparía demasiado espacio. Gracias a mis amigos de París, gracias a mis amigos de Leeds. Estando con vosotros me sentía en casa. Gracias a todos mis amigos de la universidad.

Me llevo una tremenda experiencia y un montón de amigos.

Gracias a todos.

Index

General introduction

Bacteria largely owe their pervasiveness and ubiquitous presence in natural habitats to their ability to adapt and survive under hostile and continuously changing conditions. Their efficient and timely responses to a wide range of stimuli usually involve a large variety of complex mechanisms controlling gene expression at the transcriptional and post-transcriptional levels.

Among the cellular factors that control bacterial fitness and adaptation, antisense RNAs are known to function as key post-transcriptional regulators of gene expression (reviewed by Kaberdin and Bläsi, (2006) and Storz *et al.* (2011)). In addition to small antisense RNAs (sRNAs) that act on *cis*-encoded mRNA targets (often transcribed from extrachromosomal genetic elements such as transposons, plasmids and bacteriophages), a large number of antisense RNAs, namely *trans*-encoded sRNAs, basepair with target mRNAs that are transcribed from different genomic locations. Since *trans*-encoded sRNAs and their mRNA targets normally share imperfect complementarity, their interaction is usually assisted by the RNA chaperone Hfq (reviewed by Vogel and Luisi (2011) and Sobrero and Valverde (2012)). The formation of a sRNA-mRNA duplex can promote or impede the RNase dependent degradation of the target mRNA and/or repress or enhance its translation (reviewed by Kaberdin and Bläsi, (2006)), thus exerting rapid control of gene expression.

Very often transcription of *trans*-encoded sRNAs is induced in response to specific environmental cues, thus making sRNAs important regulators during bacterial responses to various stress conditions, such as limitation of iron, cell envelope stress, switching to utilization of alternative carbon sources, sub-optimal temperature and oxidative stress (reviewed by Hoe *et al.* (2013)).

There are several advantageous features concerning the regulation of gene expression by sRNAs. One of them is the capacity of single sRNAs to regulate multiple targets. For example, according to BSRD database (bac-srna.org) (Li *et al.*, 2013), the small RNAs DsrA, RyhB, RybB, and GcvB can regulate 5, 9, 11 and 21 different mRNA targets, respectively. Moreover, there are also examples of several sRNAs acting on the same target mRNA. While MicC, InvR, RybB and MicA sRNAs control the expression of *ompC* and *ompD* mRNAs encoding components of outer membrane porins, DsrA, RprA and ArcZ are some examples of sRNAs that affect the translation of *rpoS* mRNA, encoding the sigma factor σ^S, which functions as a master regulator of the

general stress response in *E. coli* (reviewed by Wagner and Romby, (2015)). Such diverse modes of action allow the simultaneous control of multiple targets and their fine-tuning, which ensures the high efficiency and specificity of bacterial stress responses. Furthermore, some sRNAs regulate transcription factors under conditions when both of them (sRNAs and transcription factors) are integral parts of extensive regulatory networks. Thus, their concerted action can precisely control target mRNA expression at both transcriptional and post-transcriptional levels, resulting in accurate and rapid responses (reviewed by Beisel and Storz, (2010) and Wagner and Romby, (2015)). Another advantage of the sRNA-mediated control is that the short half-lives of sRNAs (Vogel *et al.*, 2003), linked to their co-degradation with their mRNA targets, permits a fast inactivation and elimination of the regulator once the stress is over (reviewed by Kaberdin and Bläsi, (2006) and Saramago *et al.* (2014). Finally, as the energy expenditure required for synthesis of a sRNA is lower than that required for the synthesis of a transcription factor or translational repressor, regulation by sRNAs appears to be more economical.

The first non-coding small RNAs were discovered in early 70s (e.g. 4.5S, RnpB and Spot 42, reviewed by Wassarman *et al.* (1999)), and in the two following decades a few more sRNAs were added, including the first *trans*-encoded antisense sRNAs, such as OxyS, MicF, DicF and DsrA, discovered by co-purification with RNA-binding proteins (e.g. Hfq), induction during exposure to stress or by cloning approaches (reviewed by Wassarman *et al.* (1999)). The gradually increasing number of characterized *trans*-encoded sRNAs allowed defining their common features (i.e., their average length of 50 to 300 nucleotides (Hershberg *et al.*, 2003), transcription from intergenic regions (IGR), the conservation of their sequence among closely related species (Wassarman *et al.*, 2001), the presence of a promoter and ρ-independent terminator) (reviewed by Updegrove *et al.* (2015)). These features were used to develop general criteria employed for sRNAs prediction and analysis. Some examples of studies devoted to search of sRNAs in *E. coli* according to the above criteria include theoretical and experimental studies carried out by Argaman *et al.*, (2001), Rivas *et al.* (2001), Wassarman *et al.* (2001), Chen *et al.* (2002), Tjaden *et al.* (2002), Vogel *et al.* (2003), Zhang *et al.* (2003), Kawano *et al.* (2005) and Yachie *et al.* (2006)), as well as those using next-generation sequencing performed by Raghavan *et al.* (2011), Shinhara *et al.* (2011), Conway *et al.* (2014), Romero *et al.* (2014), Thomason *et al.* (2015) and Weirick *et al.* (2015). These studies and others greatly expanded the landscape of identified small RNAs in *E. coli,* currently including *circa* one hundred of

sRNAs annotated in sRNA databases (e.g. 108 sRNAs are included in the BSRD database according to the last update on October 2015 (Li *et al.*, 2013)).

Although some known sRNAs are well characterized, the biological functions and mRNA targets of many other sRNAs are still unknown. Moreover, given the large number of putative sRNAs predicted *in silico* and detected by high-throughput techniques, one of the current challenges is to develop efficient and reliable tools for distinguishing the signals potentially belonging to putative sRNAs from transcriptional noise (Lindgreen *et al.*, 2014), subsequently selecting the most promising sRNA candidates for further validation and analysis.

As further validation of putative sRNA and their analysis can also be a time-consuming and challenging task, there is a necessity for further improvement and development of new high-throughput approaches that enable sRNA detection and analysis to reveal their biological functions. In this direction, the recent work of Melamed *et al.* (2016) led to development of RIL-seq approach, aiming at identification of sRNAs and their mRNA targets through their interaction with the RNA chaperone Hfq.

The above considerations determined the two major aims of the present study, respectively focused on (i) optimization of high-throughput methods for validation and analysis of the sRNAs and (ii) utilization of new generation sequencing for discovery of new sRNAs in *E. coli*.

In the course of this study, we designed and tested a new custom microarray for sRNA profiling in *E. coli*. The results obtained with custom microarrays (**Chapters 1 and 2**) as well as those obtained by analysis of RNA-seq data (**Chapter 3**) led to the discovery and validation of several novel sRNAs. While providing new information about the expression patterns of known and previously reported putative sRNAs, this work also led to new insights regarding the possible involvement of known and putative sRNAs in the regulation of the central carbon metabolism and associated pathways in *E. coli*. Some of the results presented in **Chapter 1** and **Chapter 3** have been recently published (Romero A. *et al.*, 2014; Ruiz-Larrabeiti *et al.*, 2016b).

Bibliography

Argaman, L.; Hershberg, R.; Vogel, J.; Bejerano, G.; Wagner, E. G.; Margalit, H.; Altuvia, S. Novel Small RNA-Encoding Genes in the Intergenic Regions of Escherichia

Beisel, C. L.; Storz, G. Base Pairing Small RNAs and Their Roles in Global Regulatory Networks. *FEMS Microbiol. Rev.* **2010**, *34* (5), 866–882.

Chen, S.; Lesnik, E. a; Hall, T. a; Sampath, R.; Griffey, R. H.; Ecker, D. J.; Blyn, L. B. A Bioinformatics Based Approach to Discover Small RNA Genes in the *Escherichia coli* Genome. *Biosystems.* **2002**, *65* (2–3), 157–177.

Conway, T.; Creecy, J. P.; Maddox, S. M.; Grissom, J. E.; Conkle, T. L.; Shadid, T. M.; Teramoto, J. Unprecedented High-Resolution View of Bacterial Operon Architecture Revealed by RNA Sequencing. *MBio* **2014**, *5* (4), :e01442-14.

Hershberg, R.; Altuvia, S.; Margalit, H. A Survey of Small RNA-Encoding Genes in *Escherichia coli*. *Nucleic Acids Res.* **2003**, *31* (7), 1813–1820.

Hoe, C.-H.; Raabe, C. a; Rozhdestvensky, T. S.; Tang, T.-H. Bacterial sRNAs: Regulation in Stress. *Int. J. Med. Microbiol.* **2013**, *303* (5), 217–229.

Kaberdin, V. R.; Bläsi, U. Translation Initiation and the Fate of Bacterial mRNAs. *FEMS Microbiol. Rev.* **2006**, *30* (6), 967–979.

Kawano, M.; Reynolds, A. a; Miranda-Rios, J.; Storz, G. Detection of 5'- and 3'-UTR-Derived Small RNAs and Cis-Encoded Antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.* **2005**, *33* (3), 1040–1050.

Li, L.; Huang, D.; Cheung, M. K.; Nong, W.; Huang, Q.; Kwan, H. S. BSRD: A Repository for Bacterial Small Regulatory RNA. *Nucleic Acids Res.* **2013**, *41* (D1), 233–238.

Melamed, S.; Peer, A.; Faigenbaum-Romm, R.; Gatt, Y. E.; Reiss, N.; Bar, A.; Altuvia, Y.; Argaman, L.; Margalit, H. Global Mapping of Small RNA-Target Interactions in Resource Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol. Cell* **2016**, *63* (5), 884–897.

Michaux, C.; Verneuil, N.; Hartke, A.; Giard, J. Physiological Roles of Small RNA Molecules. *Microbiology* **2014**, No. 33, 1–33.

Raghavan, R.; Groisman, E. A.; Ochman, H. Genome-Wide Detection of Novel

Regulatory RNAs in E . Coli. *Genome Res.* **2011**, *21*, 1487–1497.

Richards, G. R.; Vanderpool, C. K. Molecular Call and Response: The Physiology of Bacterial Small RNAs. *Biochim. Biophys. Acta* **2011**, *1809* (10), 525–531.

Rivas, E.; Klein, R. J.; Jones, T. a; Eddy, S. R. Computational Identification of Noncoding RNAs in E. Coli by Comparative Genomics. *Curr. Biol.* **2001**, *11* (17), 1369–1373.

Romero A., D.; Hasan, A. H.; Lin, Y.-F. fei; Kime, L.; Ruiz-Larrabeiti, O.; Urem, M.; Bucca, G.; Mamanova, L.; Laing, E. E.; van Wezel, G. P.; *et al.* A Comparison of Key Aspects of Gene Regulation in Streptomyces Coelicolor and *Escherichia coli* Using Nucleotide-Resolution Transcription Maps Produced in Parallel by Global and Differential RNA Sequencing. *Mol. Microbiol.* **2014**, *94* (5), 963–987.

Ruiz-Larrabeiti, O.; Plágaro, A. H.; Gracia, C.; Gallego, L.; Hajnsdorf, E.; Kaberdin, V. R. A New Custom Microarray for sRNA Profiling in *Escherichia coli*. *FEMS Microbiol. Lett.* **2016**, 1–20.

Saramago, M.; Bárria, C.; Dos Santos, R. F.; Silva, I. J.; Pobre, V.; Domingues, S.; Andrade, J. M.; Viegas, S. C.; Arraiano, C. M. The Role of RNases in the Regulation of Small RNAs. *Curr. Opin. Microbiol.* **2014**, *18*, 105–115.

Shinhara, A.; Matsui, M.; Hiraoka, K.; Nomura, W.; Hirano, R.; Nakahigashi, K.; Tomita, M.; Mori, H.; Kanai, A.; Kana, A. Deep Sequencing Reveals as-yet-Undiscovered Small RNAs in *Escherichia coli*. *BMC Genomics* **2011**, *12* (1), 428.

Sobrero, P.; Valverde, C. The Bacterial Protein Hfq: Much More than a Mere RNA-Binding Factor. *Crit. Rev. Microbiol.* **2012**, *38* (4), 276–299.

Storz, G.; Vogel, J.; Wassarman, K. M. Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell* **2011**, *43* (6), 880–891.

Thomason, M. K.; Bischler, T.; Eisenbart, S. K.; Förstner, K. U.; Zhang, A.; Herbig, A.; Nieselt, K.; Sharma, C. M.; Storz, G. Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in *Escherichia coli*. *J. Bacteriol.* **2015**, *197* (1), 18–28.

Tjaden, B.; Saxena, R. M.; Stolyar, S.; Haynor, D. R.; Kolker, E.; Rosenow, C. Transcriptome Analysis of *Escherichia coli* Using High-Density Oligonucleotide Probe Arrays. *Nucleic Acids Res.* **2002**, *30* (17), 3732–3738.

Updegrove, T. B.; Shabalina, S. A.; Storz, G. How Do Base-Pairing Small RNAs Evolve? *FEMS Microbiol. Rev.* **2015**, *39* (3), 379–391.

Vogel, J.; Luisi, B. F. Hfq and Its Constellation of RNA. *Nat. Rev. Microbiol.* **2011**, *9* (8), 578–589.

Vogel, J.; Bartels, V.; Tang, T. H.; Churakov, G.; Slagter-Jäger, J. G.; Hüttenhofer, A.; Wagner, E. G. H. RNomics in *Escherichia coli* Detects New sRNA Species and Indicates Parallel Transcriptional Output in Bacteria. *Nucleic Acids Res.* **2003**, *31* (22), 6435–6443.

Wagner, E. G. H.; Romby, P. *Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It*; Elsevier Ltd, **2015**; Vol. 90.

Wassarman, K. M.; Zhang, A.; Storz, G. Small RNAs in *Escherichia coli*. *Trends Microbiol.* **1999**, *7* (1), 37–45.

Wassarman, K. M.; Repoila, F.; Rosenow, C.; Storz, G.; Gottesman, S. Identification of Novel Small RNAs Using Comparative Genomics and Microarrays. *GENES Dev.* **2001**, *15*, 1637–1651.

Weirick, T.; Militello, G.; Müller, R.; John, D.; Dimmeler, S.; Uchida, S. The Identification and Characterization of Novel Transcripts from RNA-Seq Data. *Brief. Bioinform.* **2015**, No. May, 1–8.

Yachie, N.; Numata, K.; Saito, R.; Kanai, A.; Tomita, M. Prediction of Non-Coding and Antisense RNA Genes in *Escherichia coli* with Gapped Markov Model. *Gene* **2006**, *372*, 171–181.

Zhang, A.; Wassarman, K. M.; Rosenow, C.; Tjaden, B.; Storz, G.; Gottesman, S. Global Analysis of Small RNA and mRNA Targets of Hfq. *Mol. Microbiol.* **2003**, *50* (4), 1111–1124.

Chapter 1

Design and testing of a custom microarray for sRNA profiling in *Escherichia coli*

Introduction

Small RNAs (sRNAs) play essential regulatory roles in the post-transcriptional control of gene expression in *E. coli* and many other bacteria. Previous work revealed that sRNAs are often involved in regulation of bacterial stress responses and adaptation to changing environmental conditions. Some of the best characterized sRNAs function by regulating iron homeostasis, carbon metabolism, envelope and oxidative stresses (Gottesman *et al.*, 2006). Many *trans*-encoded sRNAs exert their regulatory functions by base-pairing with their target mRNAs to form imperfect sRNA-mRNA duplexes, often in an Hfq-dependent manner (Wagner, 2013), thereby affecting the stability and/or the translation of their target mRNAs (Kaberdin and Bläsi, 2006; Gottesman and Storz, 2011).

Small RNAs are typically 50-300 nt long (Hershberg *et al.*, 2003). *Trans*-encoded small RNAs are transcribed from intergenic regions (IGR) and usually do not contain open reading frames (ORF). Their small size and non-coding nature previously hampered their discovery. However, development of sRNA prediction tools and new experimental approaches expedited sRNA discovery and analysis of their regulation of gene expression in *E. coli* and other bacteria (Vogel and Sharma, 2005). These studies revealed numerous putative sRNA candidates (see **Chapter 3**), of which only a small fraction was annotated in major databases such as EcoCyc (Keseler *et al.*, 2013), EcoGene (Zhou and Rudd, 2013) and BSRD (Li *et al.*, 2013). Moreover, many of the annotated sRNAs have not been characterized so far with respect to their functions and mechanisms of action.

Microarray is a robust technique for which advanced analytical tools have been developed and improved reproducibility has been achieved (Yang *et al.*, 2002), thus enabling an economical high-throughput gene expression analysis that can offer a similar performance as standard RNA-seq approaches (Malone and Oliver, 2011). Recently microarray analysis has been used for sRNA discovery and characterization (Wassarman *et al.*, 2001; Zhang *et al.*, 2003; Pulvermacher *et al.*, 2009). Both low and

high density microarrays are currently available from different suppliers (such as Affymetrix and Agilent Technologies) to carry out routine gene expression profiling in pro- and eukaryotic organisms. These commercial microarrays are generally effective in detection of annotated transcripts and protein-coding genes. Nevertheless, detection efficiency of short transcripts such as sRNAs by these microarrays cannot be guaranteed unless the probes are properly designed (i.e. to ensure that the properties of each probe match the characteristics of the corresponding sequence and hybridization conditions) and tested to verify specificity and hybridization efficiency. For this reason, even high-density microarrays containing probes that cover the sequence of the entire genome, including all intergenic regions, cannot assure detection of *trans*-encoded sRNAs due to the lack of probes specifically designed for their detection.

The aim of this study was to design a custom microarray with a new set of probes specific for many known and a number of putative sRNAs of *E. coli* and to test its capacity to detect the corresponding transcripts. Towards this goal, we used the eArray platform (Agilent Technologies) for designing and manufacturing custom microarrays with the capacity to simultaneously monitor expression of protein-coding genes and non-coding transcripts originated from selected intergenic regions. We tested our microarray by analysing the differences in the transcriptome profiles of *E. coli* MG1655 strain cells grown in rich and minimal media. We also compared the transcriptome profiles of a mutant strain lacking poly(A)polymerase I (PAP I) (an enzyme capable of affecting the stability of all types of RNA in *E. coli*, including sRNAs (Viegas *et al.*, 2007; Urban and Vogel, 2008; Maes *et al.*, 2013)) and its isogenic wild type strain of *E. coli*. The results of the microarray analysis revealed differential expression of a number of known sRNAs and the corresponding regulation of sRNA targets. Moreover, we could confirm the existence of some putative sRNAs previously predicted *in silico,* and validated this result by northern blotting. Thus, our data demonstrated the efficiency of this custom microarray for simultaneous analysis and expression profiling of small RNAs and protein-coding transcripts of *E. coli*.

Materials and Methods

Strains and growth conditions

The experiments presented in this chapter were performed with *Escherichia coli* MG1655 strain grown at 37 °C with shaking (100 rpm) in either Luria–Bertani (LB) or M9 minimal medium supplemented with 0.4% glucose (M9/glucose), and the *Escherichia coli* N3433 (wild-type) strain and its isogenic mutant strain IBPC903 (*pcnB::kanR*) (Joanny *et al.*, 2007) grown in LB as specified above.

Custom microarray design

The nucleotide sequences of intergenic regions encoding known and putative sRNAs (Hershberg *et al.*, 2003) were processed by a web-based application, eArray (Agilent Technologies), to generate a set of custom microarray probes (Supplementary Tables S1 and S2 in Ruiz-Larrabeiti *et al.* (2016), a link to the tables can be found in Supplementary data section, p. 91). These probes were combined with the standard set of microarray probes previously designed by Agilent Technologies to detect *E. coli* protein-coding transcripts to create the custom microarray used in this study. The final microarray design (Agilent-032948 *E. coli* Gene Expression Microarray VK1020097) was deposited in the Array-Express database (https://www.ebi.ac.uk/arrayexpress/).

RNA isolation and microarray analysis

Aliquots of 8 mL of cell cultures grown in triplicate as described above until $O.D._{600}$~0.5 were mixed with 1 mL of ice-cold stop solution (5% phenol in ethanol). Cells were collected by centrifugation (15 min, 4 °C, 4000 g), suspended in 250 µl of 1X TE (pH 8.0) supplemented with lysozyme (0.5 mg/mL, Sigma), mixed with an equal volume of lysis buffer (200 mM NaCl, 20 mM Tris–HCl (pH 7.5), 40 mM EDTA, 0.5% SDS) and incubated at 95 °C for 30 sec. RNA was extracted from the cell lysates by sequential extractions with phenol, phenol/chloroform/isoamyl alcohol (25:24:1) and chloroform/isoamyl alcohol (24:1), precipitated with isopropanol, washed with 80% cold ethanol, air-dried for 5 min and dissolved in RNase-free water (Accugene). To remove genomic DNA, RNA was further treated with the RNase free DNase I (Thermo Scientific) for 30 min at 37 °C, extracted with phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol-precipitated. The resulting RNA pellets were washed with 80%

cold-ethanol, air-dried and suspended in RNase-free water (Accugene). Further analysis of RNA samples was carried out at the General Genomic Service (SGiker) of the University of the Basque Country as described previously (Kaberdin *et al.*, 2015). Briefly, after verifying RNA quality and integrity by Lab-chip technology on an Agilent 2100 Bioanalyzer with Agilent RNA6000 Nano Chips, RNA was retro-transcribed with Superscript III Reverse Transcriptase (Invitrogen) and labelled using the SuperScript Indirect cDNA labelling System (Invitrogen) to incorporate amino-modified nucleotides. After removal of unincorporated nucleotides, the amino-modified cDNA was coupled to fluorescent dyes (Cy5 or Cy3), used for hybridization with custom microarrays designed as described above and subsequently manufactured by Agilent Technologies. Raw data from Feature Extraction Software (FE processed signals) were subsequently processed on GeneSpring MultiOmic Analysis Software 11.5.1 (Agilent Technologies) and were further subjected to LIMMA statistical analysis by using the MultiExperiment Viewer application (version 4.7.1) available at http://www.tm4.org/mev.html (Saeed *et al.*, 2003, 2006). Three sets of data containing genes with threshold p-values equal or smaller than 0.05, 0.02 and 0.01, respectively, were selected to reveal transcripts differently expressed in LB or M9/glucose medium. The data sets obtained for the MG1655 strain were deposited in the Array Express database (https://www.ebi.ac.uk/arrayexpress/) under EMTAB-3329 accession number.

Northern blotting

Aliquots of total RNA (5-10 µg each) were individually mixed with equal volumes of 2X RNA-loading dye (0.03% Bromophenol Blue, 0.03% Xylene Cyanol, 10 mM EDTA in formamide), incubated at 90 °C for 90 s, chilled on ice, and separated on 6% polyacrylamide-urea denaturing gels. Then the fractionated RNA were electro-transferred to Hybond-N+ membranes (GE Healthcare Life Sciences) and UV-crosslinked. The membranes with the cross-linked RNAs were cut into smaller pieces and, after pre-hybridization with ULTRAhyb-Oligo Hybridization Buffer (Ambion) at 42 °C for 30 min, were further hybridized with specific 5'-end radio-labelled oligonucleotide probes (listed in **Table S3**).

For radioactive labelling 5 pmol of oligonucleotide probe (Sigma) was mixed with 2 µl of PNK buffer A (Thermo Scientific) and molecular biology grade water (AccuGENE) to get the final volume of 16 µl. Samples were denatured by incubation at 90 °C for 90 s, chilled on ice, and radiolabelled by incubation with 3 µl of γ-^{32}P-ATP (3000 Ci/mmol 10

mCi/mL, 250 µCi, Perkin Elmer) and 1 µl of T4 polynucleotide kinase (Thermo Scientific) at 37 °C for 30 minutes. Reaction was stopped by addition of 30 µl of EDTA 2 mM, and radio-labelled probes were purified by ethanol precipitation by adding 0.1 volumes of 3M NaCl and 2.5 volumes of ethanol, incubating in dry ice for 1h and centrifugation at 14,500 rpm for 1 h. Supernatant was disposed and pellet was suspended in 20 µl of molecular biology grade water (AccuGENE).

To generate the size marker (ladder), membranes were hybridized with a mixture of four radio-labelled probes complementary to tRNA (95 nt), 5S rRNA (120 nt), 6S RNA (183 nt) and RnpB RNA (also known as M1) (377 nt), respectively. All the probes were denatured before addition to hybridization buffer. After hybridization at 42 °C overnight, the membranes were washed twice with 20 mL of preheated wash buffer (5X SSC with 0.5 % w/v SDS) at 49 °C for 30 min and exposed to Imaging Screen-K (Bio-Rad). After their capturing by Molecular Imager FX (Bio-Rad), the images were further processed using Quantity One (Bio-Rad) and GeneSys (Syngene) software.

Results

Testing custom microarrays via comparison of gene expression in *E. coli* cells grown in rich *versus* minimal media

To test the efficiency of sRNA detection and profiling by our custom microarray, we analysed the gene expression of *E. coli* cells exponentially grown in rich Luria-Bertani media (LB) compared to those grown in M9 minimal media supplemented with glucose (M9/glucose). The complete set of raw and processed microarray data were generated as described in Materials and Methods section of this chapter. Based on the statistical analysis of experimental data obtained in three independent experiments, we found genes that were differentially expressed in cells grown LB and in M9/glucose. To analyse the metabolic significance of their differential regulation we used STRING database software (Franceschini *et al.*, 2013) to cluster the differentially expressed genes according to their biological functions. This procedure revealed the regulation of groups of genes involved in the central carbon metabolism, biosynthesis and degradation of amino acid, lipids, vitamins and related metabolites, as well as genes involved in bacterial stress responses, cell motility and adhesion (**Table 1**).

Apart from the transcripts detected with the standard set of probes provided by Agilent Technologies, which include probes for detection of protein-coding genes, our custom microarray contained a set of probes specifically designed for detection of known (**Supplementary Table S1** in Ruiz-Larrabeiti *et al.* (2016)) and putative sRNAs (**Supplementary Table S2** in Ruiz-Larrabeiti *et al.* (2016)). Nearly all the probes included in the microarray for detection of known sRNAs provided a reasonably strong (i.e. well above the background) and statistically significant signal (according to Agilent Feature Extraction standard protocols). Microarray analysis revealed that the expression level of the majority of the sRNAs analysed in *E. coli* cells during the logarithmic growth remained largely unchanged in cells grown in LB compared to those grown in M9/glucose, and only two known small RNAs (GcvB and RydB) (**Table 1**) showed significantly different expression levels. As found by the gene expression profiling and validated by northern blot (**Fig. 1**), GcvB was more efficiently expressed in cells grown in rich media (LB), while RydB was considerably more abundant in cells grown in M9/glucose.

Table 1. Functional clusters of the differentially expressed genes revealed by comparison of *E. coli* cells grown in LB *versus* M9/glucose media. Genes expressed at considerably higher (↑) or lower (↓) levels are included (p-value≤0.02, expression fold change>2).

Biological pathways				
General category	Specific pathway	Gene name	Fold difference and regulation	
Central carbon metabolism	Glycolysis/Gluconeogenesis	*fbaB*	12	↓
	Pyruvate metabolism	*poxB*	8	↓
	Mixed acid fermentation	*adhP*	7	↓
	Pentose phosphate pathway	*talA*	7	↓
	Trehalose degradation	*treF*	6	↓
	Glyoxylate cycle	*aceB*	13	↓
	Glyoxylate and dicarboxylate metabolism	*dmlA*	4	↓
	Amino sugar and nucleotide sugar metabolism	*nanA, nanE, frlB*	8-27	↑
Environmental and metabolic stress responses	Acid stress resistance	*hdeA, hdeB, hdeD, ybaS, gadA, gadB, gadC, gadE, gadX, slp*	11-55	↓
	Oxidative stress resistance	*patA, potF, potG, ordL,* *puuD, aldH*	11-30	↓
	Osmotic stress resistance	*yciF, otsA*	4-11	↓
	Antimicrobial peptide hydrolysis	*ompT*	9	↓
	Small RNA	*gcvB*	7	↑
	Small RNA	*rydB*	2-3	↓
Processing of genetic information	Aminoacyl-tRNA biosynthesis	*lysU*	9	↑
Amino acid metabolism	Tryptophan metabolism	*tnaA, ydeN*	6-73	↑
	Cysteine and methionine metabolism	*dcyD*	7	↓
	Tyrosine biosynthesis, phenylalanine biosynthesis	*pheA*	3	↓

	Serine biosynthesis	*serA*	41	↓
	Tryptophan biosynthesis	*trpE, trpD, trpC, trpB*	22-34	↓
	Arginine and proline metabolism	*astB, astD, argG*	5-19	↓
	Phenylalanine, valine, alanine, isoleucine, leucine biosynthesis	*ilvE, Ilv, ilvH, ilvI, ilvM, leuA, leuB, leuC, leuD aroG*	6-66	↓
	Methionine biosynthesis	*metE*	15	↓
	Homoserine and lysine biosynthesis	*asd*	5	↓
	Ornithine biosynthesis	*argE*	5	↓
	Amino acid transport	*fliY, livH, livJ, livK, livM, nlpA, yecC, yecS, sstT, aroP, argT, cycA*	5-77	↓
Transport	Sialic acid transport	*nanT*	28	↑
	Drug transporters	*emrD*	43	↑
	Phosphotransferase system	*treB*	35	↑
	Glycerol transporter	*glpT, glpF*	40-48	↑
	Nicotinamide riboside transport	*pnuC*	32	↓
	Sugar transport superfamily protein	*yhjE*	98	↓
Lipid metabolism	Glycerolipid metabolism	*glpK*	45	↑
	Glycerophospholipid metabolism	*glpA, glpC, glpD*	26-32	↑
		ybhO	8	↓
Adhesion and motility	Curli and biofilm formation	*csgD, csgF*	14-20	↓
	Fimbriae biosynthesis	*fimA, fimC, fimF, fimH*	6-9	↓
Vitamin and cofactor metabolism and energy production	Nicotinate and nicotinamide metabolism	*nadA, nadB*	25-27	↓
	Folate biosynthesis	*folE, ybiC*	5-8	↓
	Thiamine metabolism	*thiC, thiD, thiE, thiF, thiG, thiH, thiM*	7-45	↓

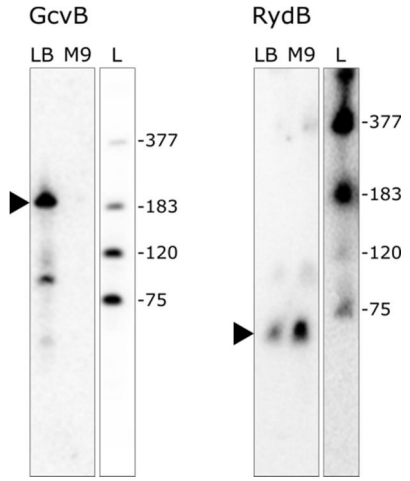


Figure 1. Two known sRNAs differentially expressed in LB *versus* M9/glucose media. Total RNA isolated from the *E. coli* MG1655 strain grown in rich (LB) or minimal (M9) media was used for northern blot detection of known sRNAs (indicated by arrowheads) differentially expressed in these two growth media, according to the microarray analysis (**Table 1**). Ladder (L) was obtained by using radiolabelled probes complementary to RnpB (M1) RNA (377 nt), 6S RNA (183 nt), 5S rRNA (120 nt) and $tRNA^{Asn}$ (75 nt). The sequence of each probe can be found in **Table S3**.

Additional examples of differentially regulated sRNAs revealed by comparison of the *E. coli* wild-type strain and its isogenic *pcnB*⁻ mutant

To provide further proof of the capacity of our custom microarray to detect differential regulation of sRNAs, we used it to compare transcriptome profiles of the *E. coli pcnB*⁻ mutant strain IBPC603 and its isogenic wild-type strain N3433. The *pcnB*⁻ mutation functionally inactivates poly(A)polymerase I (PAP I), known for its global role in RNA stability regulation (Regnier & Hajnsdorf, 2009). Because several studies have reported that a number of sRNAs are polyadenylated or degraded by a poly(A) dependent pathway (Viegas *et al.*, 2007; Urban and Vogel, 2008; Andrade *et al.*, 2012; Maes *et al.*, 2013), we anticipated that this approach would expand the number of known sRNAs which differential expression was verified by our custom microarray. Indeed, analysis of microarray data comparing transcriptome profiles of *E. coli* wild-type strain with that of the *pcnB*⁻ mutant during their exponential growth in LB revealed differential expression of five sRNAs (GlmY, GlmZ, RyjA, RybB and SroH) (**Table 2**). Moreover, these results were further confirmed by northern blotting (**Fig. 2**). The positive effects of the PAP I inactivation on expression of GlmY, GlmZ and RyjA (also known as SraL) were largely confirmatory and were consistent with the results of previous studies (Viegas *et al.*, 2007; Reichenbach *et al.*, 2008; Urban and Vogel, 2008). In contrast, the

data obtained for RybB and SroH were novel and therefore reveal new examples of sRNAs which expression level is controlled by poly(A) polymerase I *in vivo.*

Table 2. Small RNAs differentially expressed in the *pcnB*⁻ and the wild-type *E. coli* strains. Microarray data revealed differential regulation of five sRNAs (p-value ≤ 0.05).

Name, Synonyms	Orientation of flanking (→) and sRNA (→) genes (sRNA coordinates)	Fold change, regulation	References
RybB	*rcdA-ybjL* →←← (887977-888057)	3 ↑	Johansen *et al.*, 2006, Thompson *et al.*, 2007
RyjA, SraL	*soxR-ghxP* →←→ (4277927-4278066)	16 ↑	Argaman *et al.*, 2001, Wassarman *et al.*, 2001 Viegas *et al.* 2007
SroH,	*yjaZ-thiH* →←← (4190327-4190487)	8 - 21 ↑	Vogel *et al.*, 2003, Hobbs *et al.*, 2010
GlmY, SroF	*glrK-purL* ←←← (2691157-2691340)	16 ↑	Urban & Vogel, 2008 Reichenbach *et al.* 2008
GlmZ, SraJ, RyiA	*aslA-hemY* ←→← (3986432-3986603)	6 - 9 ↑	Kalamorz *et al.*, 2007, Urban & Vogel, 2008 Reichenbach *et al.* 2008

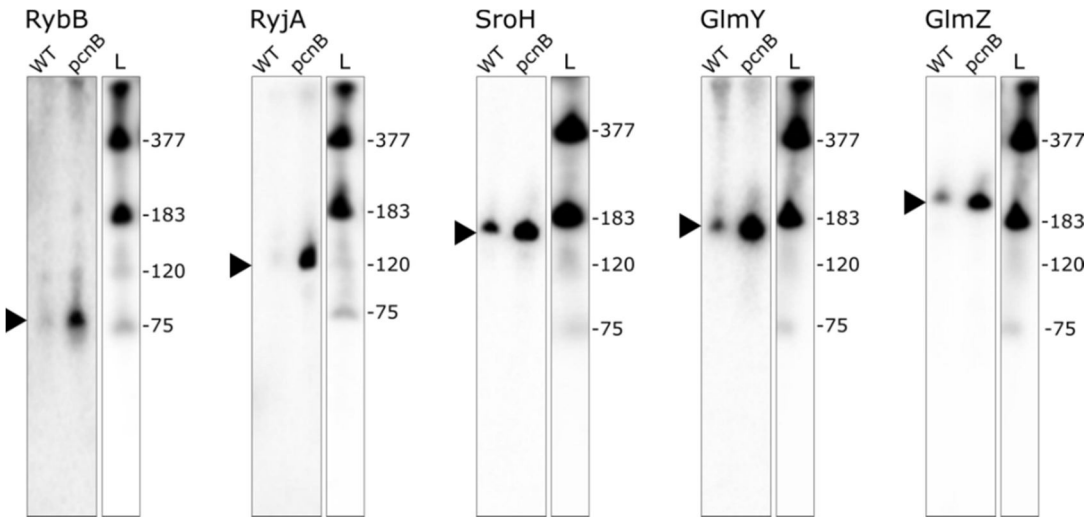


Figure 2. Known sRNAs differentially expressed in the wild-type and *pcnB*⁻ mutant. Total RNA isolated from the *E. coli* N3433 wild-type strain (WT) and its isogenic *pcnB*⁻ mutant strain IBPC603 (pcnB) was used for northern blot detection of known sRNAs (indicated by arrowheads) found to be differentially expressed in these two strains by microarray analysis (**Table 2**). Ladder (L) is detailed in **Fig. 1** and the sequences of probes used are listed in **Table S3**.

Validation of previously predicted sRNAs

By the time this study was started, many high-throughput approaches and *in silico* search methods had been directed to sRNA discovery in *Escherichia coli* (Carter *et al.*, 2001; Rivas *et al.*, 2001; Wassarman *et al.*, 2001; Chen *et al.*, 2002; Tjaden *et al.*, 2002), revealing a large number of putative sRNAs (Hershberg *et al.*, 2003). Some of them were further validated and included in databases such as EcoCyc and BSRD (Keseler *et al.*, 2013; Li *et al.*, 2013). Although their existence *in vivo* and potential involvement in the antisense control of gene expression was tested in some cases, many of these putative sRNAs have never been subjected to experimental validation. This prompted us to include additional probes (Supplementary **Table S2** in Ruiz-Larrabeiti *et al.* (2106)) potentially able to detect some candidate sRNAs. The intergenic regions that did not show evidence of overlap with the 5' or 3' untranslated regions of the flanking transcripts and for which microarray probes provided positive hybridization signals were selected for the final list (**Table 3**). Further northern blot analysis was employed to assess the existence of sRNAs potentially transcribed from six randomly selected regions, which yielded positive signals corresponding to two new sRNAs (**Fig. 3**). The first one (G0-10699) has been envisaged recently based on the results of RNA sequencing (Raghavan *et al.*, 2011). The second putative sRNA, IS001 (Chen *et al.*, 2002), has been predicted along with many others by using bioinformatic

tools (Hershberg *et al.*, 2003). Thus, the above results confirm the expression of G0-10699 and IS001 genes *in vivo*, thereby expanding the number of *E. coli* sRNAs verified experimentally.

Table 3. Putative sRNA candidates selected based on the signals obtained with custom microarrays. The results showed in this table were compiled based on the results of the microarray and northern blot analyses of total RNA isolated from *E. coli* MG1655 strain grown in LB or M9/glucose. NT marks sRNAs "not tested" by northern blotting.

Name, synonyms	Orientation of flanking (→) and sRNA (→) genes (sRNA coordinates)	Microarray	Northern Blot	References
STnc20, IS001	*insA1-rpsT* ← → ← (20520-20760)	✓	✓	(Sittka *et al.*, 2008; Chen *et al.*, 2002)
C0074	*ykfA-perR* ← →← (269148-269239)	✓	✘	(Tjaden *et al.*, 2002)
C0086, HB_33	*frmR-yaiO* ← → ← (379973-380044)	✓	✘	(Carter *et al.*, 2001; Tjaden *et al.*, 2002)
C0087	*ampH→sbmA* ← → → (396403-396577)	✓	NT	(Tjaden *et al.*, 2002)
IS017	*ybcK-ybcL* → ←→ (570682-570824)	✓	NT	(Chen *et al.*, 2002)
IS021	*appY-ompT* → → ← (584477-584677)	✓	✘	(Chen *et al.*, 2002)
k32	*rnk-rna* ← ← ← (643977-644177)	✓	NT	(Rivas *et al.*, 2001)
C0168	*ybfF –seqA* ← ← → (712827-712977)	✓	NT	(Tjaden *et al.*, 2002)
C0267, HB_32	*putP - efeU-1* → ← → (1080847-1081097)	✓	NT	(Carter *et al.*, 2001; Chen *et al.*, 2002)
C0073	*yciC- ompW* ← → → (1313748-1313991)	✓	NT	(Tjaden *et al.*, 2002)
IS079	*uidR-hdhA* ← → ← (1697117-1697277)	✓	NT	(Chen *et al.*, 2002)

C0441, STnc190a	*yobH-kdgR* → → ← (1909176-1909276)	✓	NT	(Tjaden *et al.*, 2002; Sittka *et al.*, 2008)
G0-10699, RNA0_361, k41	*sdiA-yecC* ← → ← (1996944-1997063)	✓	✓	(Rivas *et al.*, 2001; Raghavan *et al.*, 2011)
C0593, IS125, HB_413	*yffP-yffQ* → ← → (2563235-2563463)	✓	NT	(Carter *et al.*, 2001; Tjaden *et al.*, 2002)
IS139	*stpA- alaE* ← ← → (2798762-2798941)	✓	NT	(Chen *et al.*, 2002)
HB_428	*ygaM-nrdH* → ← → (2800512-2800591)	✓	NT	(Carter *et al.*, 2001)
IS143	*ygdH-sdaC* → → → (2927849-2927996)	✓	✘	(Chen *et al.*, 2002)
C0696, HB_447	*ygfF-gcpV* ← ← ← (3045917-3046044)	✓	NT	(Carter *et al.*, 2001; Tjaden *et al.*, 2002)
Tpk16	*speC-yqgA* ← → → (3109178-3109478)	✓	NT	(Rivas *et al.*, 2001)
STnc770	*yhiL-yhiJ* ← ← ← (3630627-3630927)	✓	NT	(Sittka *et al.*, 2008)

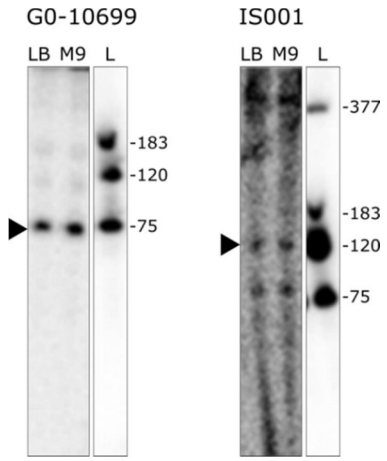


Figure 3. Northern blot analysis of putative sRNAs. Total RNA isolated from the *E. coli* MG1655 strain grown in rich (LB) or minimal (M9) media was used for northern blot detection of small RNA candidates detected by microarray analysis (indicated by arrowheads) (**Table 3**). Ladder (L) is described in **Fig. 1**, and the sequence of the corresponding probes are listed in **Table S3**.

Discussion

Due to their robustness, affordable cost and well-developed statistical tools, microarrays are widely used in gene expression studies to reveal data sets of differentially regulated transcripts that normally overlap well with those identified by using RNA-seq approaches (Malone and Oliver, 2011; Raghavachari *et al.*, 2012). The use of microarrays to discover sRNA functions requires simultaneous profiling of non-coding sRNAs and protein-coding transcripts (*i.e.* mRNAs). Many commercial microarrays have not been optimized for specific detection of small transcripts, including many *trans*-encoded sRNAs that are involved in bacterial stress responses. This motivated us to create a custom microarray and test its capacity to detect *E. coli* sRNAs. Towards this goal, we selected a list of known and putative sRNAs in *E. coli*, for which we designed multiple probes that were incorporated in the custom microarray along with the standard set of probes currently used by Agilent Technologies for detection of protein-coding transcripts.

Following their manufacturing by Agilent Technologies, custom microarrays were further employed for gene expression analysis of *E. coli* grown in rich (LB) and minimal (M9/glucose) media. Analysis of the gene expression data revealed differential expression of a number of genes that control bacterial metabolism and cell adaptation to environmental cues (**Table 1**). We found that, in agreement with the need to use glucose as a primary carbon source during cell growth in M9/glucose medium, numerous genes involved in the central carbon metabolisms (glycolysis, tricarboxylic acids cycle (TCA) as well as amino acid biosynthesis) showed higher expression levels. These data are consistent with the results of previous studies that demonstrated that cell growth on minimal media normally leads to an increase in expression of genes involved in uptake and biosynthesis of amino acids when compared to cells growing on rich media (Tao *et al.*, 1999).

Apart from providing insights into differential contribution of metabolic and stress-related genes to bacterial growth on minimal and rich media, the results of microarray analysis have also demonstrated a high capacity of our custom microarray to detect nearly all known and some putative sRNAs targeted by the probes included in our custom microarray (i.e., they provided signals that were both well above the background and statistically reliable). Detection of these sRNAs points to a good tool for pre-screening and selection of putative sRNAs for further validation or characterization, which is usually time-consuming and often challenging.

Although most of the detected known sRNAs did not exhibit considerable differences in their expression levels, we found that two sRNAs, GcvB and RydB, were differentially expressed in rich and minimal media (**Table 1, Fig. 1**). The first one (GcvB) was more abundant in cells grown in LB when compared to cells grown on M9/glucose. Previous work suggested that this antisense RNA controls expression of several genes involved in acid stress resistance (*hdeA* and *hdeB*; (Stauffer & Stauffer, 2012)) and in uptake of amino acids (*argT*, *cycA, livK, livJ, sstT*) in *Escherichia coli* (Pulvermacher *et al.*, 2009; Stauffer & Stauffer, 2012; Urbanowski *et al.*, 2000) and *Salmonella typhimurium* (Sharma *et al.*, 2007).

The higher expression of GcvB detected in *E. coli* grown in rich media when compared to M9/glucose (**Table 1, Fig. 1**) and a parallel decrease in the abundance of target mRNAs encoded by several genes (namely, *argT*, *cycA, livK, livJ, sstT, hdeA* and *hdeB*, see **Table 1**) were fully consistent with the previous findings. On the contrary, the level of the second differentially expressed known sRNA (RydB) was higher in cells grown in the minimal medium (**Table 1, Fig. 1**). Although overexpression of RydB causes a decrease in the level of the *rpoS* transcript in stationary phase (Komasa *et al.*, 2011), the actual mechanism of RydB action and its specific role in the cellular metabolism remain largely unknown.

Besides testing gene expression of *E. coli* by employing different growth conditions, our custom microarrays were also used to reveal sRNAs differentially expressed in the *pcnB*$^{-}$ mutant of *E. coli* compared to its isogenic wild-type strain. We found several known sRNAs (i.e. GlmY, GlmZ, RyjA, RybB and SroH, see **Table 2** and **Fig. 2**) whose abundance varied significantly between the wild-type strain and the mutant lacking PAP I, known for its role in the control of mRNA and sRNA stability (Regnier & Hajnsdorf, 2009). Stabilization of GlmY, GlmZ and RyjA in the absence of PAP I had been previously reported (Viegas *et al.*, 2007; Reichenbach *et al.*, 2008; Urban and Vogel, 2008), and detection of their differential expression by our microarray (also validated by northern blot analysis) not only provided additional proof of the efficiency of the custom microarray for expression profiling of sRNAs, but also revealed that RybB and SroH were differentially expressed in the wild-type and *pcnB*$^{-}$ mutant strain. These findings and comparison of transcriptome profiles of these strains will be further analysed in future work and reported elsewhere.

Finally, our custom microarray also included specific probes for detection of putative sRNAs. Some of these probes yielded positive signals (see **Table 3**), and six of them were selected for validation by northern blot analysis. As a result, we confirmed

the existence of two putative sRNAs (namely, G0-10699 (Raghavan *et al.*, 2011) and IS001 (Chen *et al.*, 2002), see **Fig. 3**), thus broadening the landscape of experimentally verified putative sRNAs.

Collectively, the results of this study suggest that our custom microarray with improved capacity for detection of sRNAs not only enables detection of differential expression of sRNAs under different growth conditions or in different genetic backgrounds, but also offers an attractive option for pre-screening of putative sRNAs predicted *in silico* (Hershberg *et al.*, 2003) or recently identified by RNA-seq (Rau *et al.*, 2015).

Bibliography

Andrade, J. M.; Pobre, V.; Matos, A. M.; Arraiano, C. M. The Crucial Role of PNPase in the Degradation of Small RNAs That Are Not Associated with Hfq. *RNA* **2012**, *18* (4), 844–855.

Carter, R. J.; Dubchak, I.; Holbrook, S. R. A Computational Approach to Identify Genes for Functional RNAs in Genomic Sequences. *Nucleic Acids Res.* **2001**, *29* (19), 3928–3938.

Chen, S.; Lesnik, E. a; Hall, T. a; Sampath, R.; Griffey, R. H.; Ecker, D. J.; Blyn, L. B. A Bioinformatics Based Approach to Discover Small RNA Genes in the *Escherichia coli* Genome. *Biosystems.* **2002**, *65* (2–3), 157–177.

Franceschini, A.; Szklarczyk, D.; Frankild, S.; Kuhn, M.; Simonovic, M.; Roth, A.; Lin, J.; Minguez, P.; Bork, P.; von Mering, C.; *et al.* STRING v9.1: Protein-Protein Interaction Networks, with Increased Coverage and Integration. *Nucleic Acids Res.* **2013**, *41* (D1), D808–D815.

Gottesman, S.; Storz, G. Bacterial Small RNA Regulators -- Versatile Roles and Rapidly Evolving Variations.pdf. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, 1–16.

Gottesman, S.; Mccullen, C. A.; Guillier, M.; Vanderpool, C. K.; Majdalani, N.; Benhammou, J.; Thompson, K. M.; Fitzgerlad, P. C.; Sowa, N. A.; Fitzgerlad, D. J. Small RNA Regulators and the Bacterial Response to Stress Small RNA Regulators and the Bacterial Response to Stress. *Cold Spring Harb. Symp. Quant. Biol.* **2006**, *LXXI*, 1–11.

Hershberg, R.; Altuvia, S.; Margalit, H. A Survey of Small RNA-Encoding Genes in *Escherichia coli*. *Nucleic Acids Res.* **2003**, *31* (7), 1813–1820.

Joanny, G.; Derout, J. Le; Bréchemier-Baey, D.; Labas, V.; Vinh, J.; Régnier, P.; Hajnsdorf, E. Polyadenylation of a Functional mRNA Controls Gene Expression in *Escherichia coli*. *Nucleic Acids Res.* **2007**, *35* (8), 2494–2502.

Kaberdin, V. R.; Bläsi, U. Translation Initiation and the Fate of Bacterial mRNAs. *FEMS Microbiol. Rev.* **2006**, *30* (6), 967–979.

Kaberdin, V. R.; Montánchez, I.; Parada, C.; Orruño, M.; Arana, I.; Barcina, I. Unveiling the Metabolic Pathways Associated with the Adaptive Reduction of Cell Size During Vibrio Harveyi Persistence in Seawater Microcosms. *Microb. Ecol.* **2015**, *70* (3), 689–

700.

Keseler, I. M.; Mackie, A.; Peralta-Gil, M.; Santos-Zavaleta, A.; Gama-Castro, S.; Bonavides-Martínez, C.; Fulcher, C.; Huerta, A. M.; Kothari, A.; Krummenacker, M.; *et al.* EcoCyc: Fusing Model Organism Databases with Systems Biology. *Nucleic Acids Res.* **2013**, *41* (Database issue), D605-12.

Li, L.; Huang, D.; Cheung, M. K.; Nong, W.; Huang, Q.; Kwan, H. S. BSRD: A Repository for Bacterial Small Regulatory RNA. *Nucleic Acids Res.* **2013**, *41* (D1), 233–238.

Maes, A.; Gracia, C.; Dominique Bréchemier, P. H.; Chatre, E.; Lemelle, L.; Bertin, P. N.; Hajnsdorf, E.; Maes, A.; Gracia, C.; Bréchemier, D.; *et al.* Role of Polyadenylation in Regulation of the Flagella Cascade and Motility in E. Coli. *Biochimie* **2013**, *95* (2), 410–418.

Malone, J. H.; Oliver, B. Microarrays, Deep Sequencing and the True Measure of the Transcriptome. *BMC Biol.* **2011**, *9* (1), 34.

Pulvermacher, S. C.; Stauffer, L. T.; Stauffer, G. V. Role of the sRNA GcvB in Regulation of cycA in *Escherichia coli*. *Microbiology* **2009**, *155* (1), 106–114.

Raghavachari, N.; Barb, J.; Yang, Y.; Liu, P.; Woodhouse, K.; Levy, D.; O'Donnell, C. J.; Munson, P. J.; Kato, G. J. A Systematic Comparison and Evaluation of High Density Exon Arrays and RNA-Seq Technology Used to Unravel the Peripheral Blood Transcriptome of Sickle Cell Disease. *BMC Med. Genomics* **2012**, *5* (1), 28.

Raghavan, R.; Groisman, E. A.; Ochman, H. Genome-Wide Detection of Novel Regulatory RNAs in E . Coli. *Genome Res.* **2011**, *21*, 1487–1497.

Reichenbach, B.; Maes, A.; Kalamorz, F.; E, H.; Görke, B. The Small RNA GlmY Acts Upstream of the sRNA GlmZ in the Activation of glmS Expression and Is Subject to Regulation by Polyadenylation in *Escherichia coli*. *Nucleic Acids Res.* **2008**, *36* (8), 2570–2580.

Rivas, E.; Klein, R. J.; Jones, T. a; Eddy, S. R. Computational Identification of Noncoding RNAs in E. Coli by Comparative Genomics. *Curr. Biol.* **2001**, *11* (17), 1369–1373.

Saeed, A. I.; Sharov, V.; J., W.; Li, J.; Liang, W.; Bhagabati, N.; Braisted, J.; Klapa, M.; Currier, T.; Thia-Garajan, M.; *et al.* TM4: A Free, Open-Source System for Microarray

Data Management and Analysis. *Biotechniques* **2003**, *34* (2), 374–377.

Saeed, A. I.; Bhagabati, N. K.; Braisted, J. C.; Liang, W.; Sharov, V.; Howe, E. A.; Li, J.; Thiagarajan, M.; White, J. A.; Quackenbush, J. [9] TM4 Microarray Software Suite. *Methods Enzymol.* **2006**, *411*, 134–193.

Tjaden, B.; Saxena, R. M.; Stolyar, S.; Haynor, D. R.; Kolker, E.; Rosenow, C. Transcriptome Analysis of *Escherichia coli* Using High-Density Oligonucleotide Probe Arrays. *Nucleic Acids Res.* **2002**, *30* (17), 3732–3738.

Urban, J. H.; Vogel, J. Two Seemingly Homologous Noncoding RNAs Act Hierarchically to Activate glmS mRNA Translation. *PLoS Biol.* **2008**, *6* (3), e64.

Viegas, S. C.; Pfeiffer, V.; Sittka, A.; Silva, I. J.; Vogel, J.; Arraiano, C. M. Characterization of the Role of Ribonucleases in Salmonella Small RNA Decay. *Nucleic Acids Res.* **2007**, *35* (22), 7651–7664.

Vogel, J.; Sharma, C. M. How to Find Small Non-Coding RNAs in Bacteria. *Biol. Chem.* **2005**, *386* (12), 1219–1238.

Wagner, E. G. H. Cycling of RNAs on Hfq. *RNA Biol.* **2013**, *10* (4), 619–626.

Wassarman, K. M.; Repoila, F.; Rosenow, C.; Storz, G.; Gottesman, S. Identification of Novel Small RNAs Using Comparative Genomics and Microarrays. *GENES Dev.* **2001**, *15*, 1637–1651.

Yang, Y. H.; Dudoit, S.; Luu, P.; Lin, D. M.; Peng, V.; Ngai, J.; Speed, T. P. Normalization for cDNA Microarray Data: A Robust Composite Method Addressing Single and Multiple Slide Systematic Variation. *Nucleic Acids Res.* **2002**, *30* (4), e15.

Zhang, A.; Wassarman, K. M.; Rosenow, C.; Tjaden, B.; Storz, G.; Gottesman, S. Global Analysis of Small RNA and mRNA Targets of Hfq. *Mol. Microbiol.* **2003**, *50* (4), 1111–1124.

Zhou, J.; Rudd, K. E. EcoGene 3.0. *Nucleic Acids Res.* **2013**, *41* (Database issue), D613-24.

Chapter 2

Identification of sRNAs differentially expressed in *E. coli* cells grown on glucose or pyruvate as a sole carbon source

Introduction

The pervasiveness of bacteria and their presence in adverse environments are the results of their ability to adapt and survive under hostile and continuously changing conditions. These properties of bacteria are associated with their capacity to efficiently utilize the resources available in their surroundings and to promptly adjust their cellular metabolism. Among the cellular factors that control bacterial fitness and adaptation, *trans*-encoded small RNAs (sRNAs) are known to play an essential role. Previous work revealed their contribution to regulation of biofilm formation (Bardill *et al.*, 2011), virulence (Porcheron *et al.*, 2014) and chemotaxis (Papenfort *et al.*, 2009). Moreover, a large number of sRNAs are involved in responses to environmental cues and stress conditions, such as limitation of iron (Massé *et al.*, 2007) and exposure to reactive oxygen species (Altuvia *et al.*, 1997). A number of sRNAs are also involved in control of central carbon metabolism (CCM) and related pathways. Some of the best characterized mechanisms of antisense regulation of CCM include those mediated by Spot 42 and SgrS, which regulate galactose metabolism and glucose-phosphate stress response (Møller *et al.*, 2002; Vanderpool and Gottesman, 2004), as well as CsrB and CsrC, positively regulating glycogen biosynthesis (Liu *et al.*, 1997; Weilbacher *et al.*, 2003).

The central carbon metabolism of *E. coli* includes pathways that are important for generation of energy and production of intermediate metabolites. The latter serve as precursors for biosynthesis of amino acids and other essential cellular components. Despite a large number of studies that have contributed to the discovery of the key reactions and metabolites that take part in bacterial carbon metabolism, still the regulatory networks that control the expression of genes involved in CCM are not fully defined, and current studies continue to reveal new features (Basan *et al.*, 2015; Cao *et al.*, 2016; Morin *et al.*, 2016). Moreover, because of its complexity and the number of

interconnected pathways, it seems likely that the post-transcriptional control of CCM potentially involves new antisense mechanisms mediated by known and putative sRNAs.

To expand our understanding of the regulatory networks and key factors controlling CCM of *E. coli*, we compared gene expression profiles of *E. coli* cells grown in M9 minimal media supplemented with pyruvate or glucose (M9/pyruvate and M9/glucose, respectively), thereby comparing the transcriptome profiles of cells primarily using glycolytic (growth on M9/glucose) or gluconeogenetic (growth on M9/pyruvate) pathways. Gene expression analysis was performed by using the custom microarray recently developed in our group for simultaneous profiling of protein-coding genes and sRNAs in *E. coli* (Ruiz-Larrabeiti *et al.*, 2016).

We found that, in addition to the expected regulation of protein-coding genes whose products contribute to glycolysis and gluconeogenesis, a number of genes encoding several sRNAs (GcvB, RprA, RyhB, GlmY, GlmZ and CyaR) with known roles in CCM and overlapping pathways (reviewed in Richards and Vanderpool, 2011 and Michaux *et al.*, 2014) were differentially expressed too. Moreover, differential regulation of some putative sRNAs including RyeA, a previously identified putative sRNA with unknown regulatory roles (Wassarman *et al.*, 2001), was demonstrated for the first time, thus expanding the number of sRNAs potentially implicated in the regulation of CCM in *E. coli*.

Materials and Methods

Strains and growth conditions

The experiments described here were performed with the *Escherichia coli* K-12 MG1655 strain routinely grown at 37 °C with shaking (100 rpm) in M9 minimal medium supplemented with 0.4% (w/v) glucose or 0.48% (w/v) pyruvate.

RNA isolation and microarray analysis

Cell cultures were grown in triplicate in M9/glucose or M9/pyruvate medium at 37 °C with shaking (100 rpm) until $O.D._{600}$~0.5, and isolation of total RNA as well as its treatment for microarray analysis was performed following the protocols described in the Materials and Methods section of **Chapter 1**. Three sets of data containing genes with threshold p-values equal or smaller than 0.05, 0.01 and 0.005, respectively, were selected to reveal transcripts differently expressed in cells grown in M9/glucose and M9/pyruvate.

Northern blotting

Details of the protocol for northern blotting are provided in **Chapter 1**. Briefly, total RNA was isolated from *Escherichia coli* K-12 MG1655 strain grown in M9 minimal media supplemented with glucose or pyruvate and was separated by denaturing electrophoresis, electro-transferred to a nylon membrane, and hybridized with radiolabelled probes specifically designed for complementarity to the corresponding sRNAs. The list of the probes is provided in **Table S5**.

Results

Gene expression analysis

To learn more about central carbon metabolism regulation, we performed a comparative gene expression analysis of *E. coli* cells grown in M9 minimal media supplemented with either pyruvate or glucose by using a custom microarray previously developed in our group (Ruiz-Larrabeiti *et al.*, 2016). The transcripts showing at least two fold difference in the expression level in cells grown in M9/pyruvate when compared to those grown in M9/glucose were further analysed with STRING database software (Franceschini *et al.*, 2013) to cluster the differentially expressed genes according to their biological roles. This analysis revealed several groups of up- and downregulated genes involved in CCM (**Table 1**). The data obtained indicated that during their growth on pyruvate, *E. coli* cells expressed numerous genes involved in gluconeogenesis (*pck, fbp*) and pyruvate-dependent pathways (*ackA, acs*) more efficiently than when they were grown on glucose. In contrast, the transcripts encoding the enzymes that function in glycolysis (*pykF, eno*), sugar storage (*otsA, glgB*), pentose phosphate pathway (*talA, tktB*) and mixed acid fermentation (*adhE, tdcE*) were more abundant when glucose was used as the only carbon source. Among the differentially expressed genes, there were also those involved in biosynthesis and degradation of cellular components (e.g., fatty acids, amino acids, nucleotides, vitamins and cofactors) essential for regular cell growth. The data on the differential expression of genes involved in CCM and related pathways are presented in **Fig. 1**.

In addition, microarray analysis also revealed differences in expression level of genes controlling other cellular functions, such as stress response, genetic information processing and electron transport. The full set of genes grouped according to their biological functions is presented in **Table S4**.

Small RNAs differentially expressed in *E. coli* cells grown on M9/pyruvate *versus* M9/glucose

Apart from regulation of protein-coding transcripts, microarray analysis also revealed a group of small RNAs differentially expressed in *E. coli* cells grown on pyruvate *versus* glucose (**Table 2**). We found that some known sRNAs (namely, RyeA, RprA, CyaR, RyhB, GcvB, GlmY and GlmZ) were expressed at higher levels in cells grown on

pyruvate, and their expression patterns were further confirmed by northern blot analysis (**Fig. 2**). In addition, microarray analysis also disclosed significant differences in the expression level of some putative sRNAs (C0593, C0652, I343, tpk16, IS001 and IS021) previously predicted *in silico* (Rivas *et al.*, 2001; Wassarman *et al.*, 2001; Chen *et al.*, 2002; Sætrom *et al.*, 2005; Yachie *et al.*, 2006) and/or revealed by using high-throughput techniques (Tjaden *et al.*, 2002; Vogel *et al.*, 2003; Zhang *et al.*, 2003; Kawano *et al.*, 2005; Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Kang *et al.*, 2013; Ruiz-Larrabeiti *et al.*, 2016).

Table 1. Groups of genes involved in central carbon metabolism and differentially expressed in *E. coli* cells grown in M9/pyruvate compared to those grown on M9/glucose. Functional clustering of transcripts showing higher (↑) or lower (↓) relative expression as detected by microarray analysis (p-value≤0.005, expression fold change>2) revealed groups of genes involved in central carbon metabolism and related pathways.

General category	Specific pathway	Examples of genes differentially expressed	Regulation and fold change	
Central carbon metabolism	Gluconeogenesis	*pck, fbp, ppsA*	↑	2-6
	Glycolysis and gluconeogenesis	*pykF, eno, fbaA, gapA, pgi*	↓	2-5
	Pyruvate metabolism	*ackA, acs, aldA, ldhA, mqo*	↑	2.2-5.9
	Mixed acid fermentation	*adhE, adhP, tdcE, fumC*	↓	2-3
	Starch and sucrose metabolism	*malP, malQ, otsA, glgB*	↓	2.3-2.7
	Pentose phosphate pathway	*talA, tktB, edd, gntK*	↓	2.9-7.6
Amino acids metabolism	Biosynthesis of serine, aspartate, glutamate and aromatic amino acids	*dapB, lysC, phea, argA, aroH, tyrA, carA, metL, hisB, dsdC, trpA*	↓	2-150
	Biosynthesis of lysine, phenylalanine and the amino acids of pyruvate family	*ilvA, ilvD, ilvE, ilvM, leuA* *ilvB, ilvC, ilvN*	↑ ↓	2.8-5.4 2.6-14.9
	Biosynthesis and degradation of amino acids of asparagine, glutamine and serine	*iaaA, glnA, ybaS, gltB, gltD, ggt, cysK*	↓	2-114
	Alanine biosynthesis and degradation	*dadA, dadX*	↑	24
	Degradation of amino acids of glutamate family	*gadA, gadB* *astA, astB, astC, astD, astE*	↓ ↑	6.-9.4 2.1-3.5

	Degradation of amino acids of aspartate, serine and aromatic families	*paaJ, tdh, kbl, tnaA*	↑	2-3
Nucleotides metabolism		*guaA, purB, ndk, pyrB, upp, yjjG, deoD, cmk, codA*	↓	2.1-28
Fatty acids	Biosynthesis	*accB, accC, fabA, rfaL, gpsA*	↓	2.1-7
	Degradation	*fadA, fadB, fadE, fadI, fadJ*	↑	2-5.4
Cofactors and vitamins metabolism	NAD metabolism	*nadB, pntA, pntB, pncC*	↓	2-4
	Folate, biotin and quinone biosynthesis	*pabB, folA, folK, menB, ydiI, bioC*	↓	2-3.8
	Iron-sulfur cluster assembly	*hscA, iscX*	↓	2.2

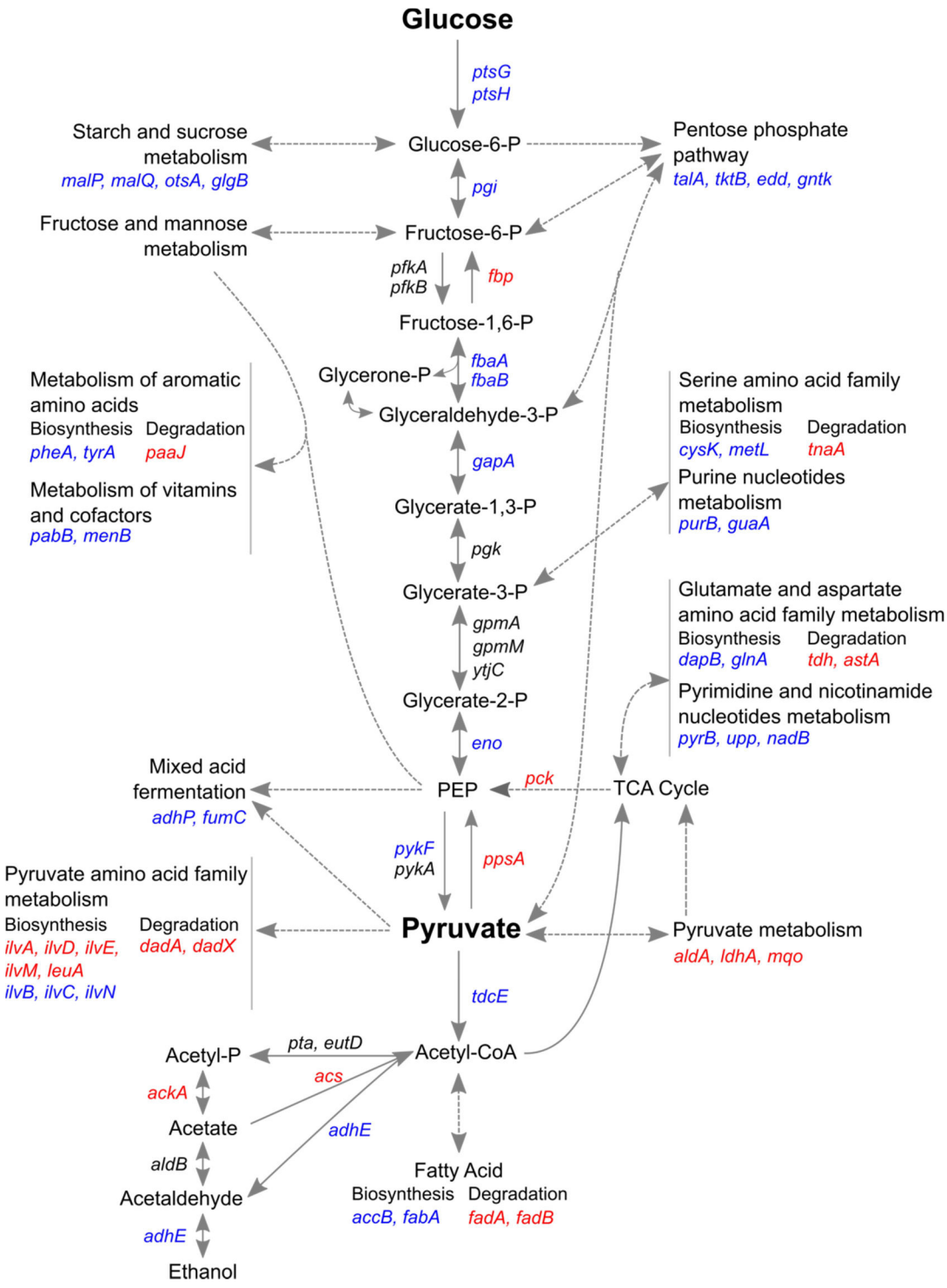


Figure 1. Differentially expressed genes involved in the central carbon metabolism in *E. coli*. Comparison of transcriptome profiles of *E. coli* cells grown on pyruvate *versus* glucose revealed that the level of some CCM-related transcripts was higher (red), lower (blue), or remained nearly the same (black).

Table 2. Small RNAs differentially expressed in *E. coli* grown in M9/pyruvate *versus* M9/glucose. Microarray analysis revealed several known and putative sRNAs showing higher (↑) or lower (↓) relative expression (p-value≤0.005, expression fold change>2) in cells grown in M9/pyruvate compared to those grown on M9/glucose.

Gene name, EcoCyc identifier	Orientation of flanking (→) and sRNA (→) genes (sRNA coordinates)	Expression fold change, regulation		References
RybA G0-8881	*yliL-mntR* → ← → (852725-853064)	2.0	↑	(Gerstle *et al.*, 2012)
RyeA, G0-8865	*pphA-yebY* ← → ← (1923066-1923314)	3.3 - 4.7	↑	(Argaman *et al.*, 2001; Wassarman *et al.*, 2001; Gutierrez *et al.*, 2013)
RprA, G0-8863	*ydiK-ydiL* → → → (1770372-1770476)	3.8	↑	(Majdalani *et al.*, 2002; Jørgensen *et al.*, 2013)
CyaR, G0-8878	*yegQ-ogrK* → → ← (2167114-2167200)	9.8	↑	(Johansen *et al.*, 2008; De Lay and Gottesman, 2009)
RyhB, G0-8872	*yhhX-yhhY* ← ← → (3580927-3581016)	130.8	↑	(Massé *et al.*, 2003; Afonyushkin *et al.*, 2005)
GcvB, G0-8867	*gcvA-ygdI* ← → ← (2942696-2942900)	4.1 - 4.8	↑	(Urbanowski *et al.*, 2000; Sharma *et al.*, 2007)
GlmY, G0-8910	*glrK-purL* ← ← ← (2691157-2691340)	2.1 - 2.3	↑	(Urban and Vogel, 2008; Reichenbach *et al.*, 2009)
GlmZ, G0-8873	*aslA-hemY* ← → ← (3986432-3986603)	2.1	↑	(Kalamorz *et al.*, 2007; Urban and Vogel, 2008)
SroA	*tbpA-sgrR* ← ← ← (75516-75608)	2.4	↓	(Rivas *et al.*, 2001; Vogel *et al.*, 2003)
C0593	*yffP-yffQ* → ← → (2563235-2563463)	2.5	↑	(Carter *et al.*, 2001; Tjaden *et al.*, 2002)
C0652 (HB_428)	*ygaM-nrdH* → ← → (2800512-2800591)	4.8	↑	(Carter *et al.*, 2001)
IS143	*ygdH-sdaC* → → → (2927849-2927996)	2.2 - 4.3	↓	(Chen *et al.*, 2002)
tpk16	*speC-yqgA* ← → → (3109178-3109478)	2.1	↓	(Rivas *et al.*, 2001)
IS001	*insA1-rpsT* ← → ← (20520-20760)	2.4	↑	(Chen *et al.*, 2002; Sittka *et al.*, 2008)
IS021	*appY-ompT* → → ← (584477-584677)	3.2 -14.2	↓	(Chen *et al.*, 2002)

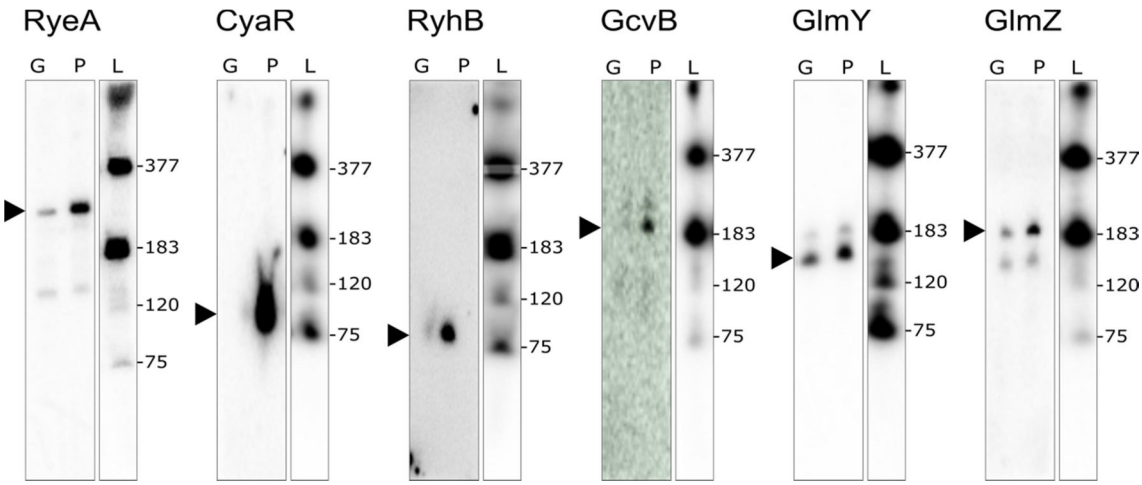


Figure 2. Northern blotting of known sRNAs differentially expressed according to microarray data. Validation of the differential expression on known sRNAs by northern blot analysis performed with total RNA isolated from *E. coli* cells grown in M9/glucose or M9/pyruvate (lanes G and P, respectively) and further hybridized to radiolabelled oligonucleotide probes complementary to these sRNAs. Molecular ladder was obtained by hybridization of total RNA with radiolabelled probes specific for RnpB (M1) RNA (377 nt), 6S RNA (183 nt), 5S rRNA (120 nt) and $tRNA^{Asn}$ (75 nt), the sequences of the probes are listed in **Table S5**.

Discussion

A profound understanding of the mechanisms used by *E. coli* during adaptation to different and often challenging environmental conditions necessarily includes comprehensive insights into post-transcriptional mechanisms controlling CCM and related pathways. Although the central carbon metabolism of *E. coli* has been extensively studied, the map of regulatory networks that control and fine-tune CCM is still under construction. A number of sRNAs have already been found to have regulatory roles in CCM. To learn more about their regulatory functions and to search for new sRNAs involved in regulation of CCM we compared transcriptome profiles of *E. coli* cells grown in M9/pyruvate *versus* M9/glucose.

Regulation of genes with major roles in the *Escherichia coli* central carbon metabolism

Glucose is the preferential carbon source in *Escherichia coli* (Inada *et al.*, 1996; Stülke and Hillen, 1999; Brückner and Titgemeyer, 2002; Aidelberg *et al.*, 2014) and it is normally metabolized through glycolysis. Consistent with the current view of the central carbon metabolism (Holms, 1996), comparison of transcriptome profiles of *E. coli* cells grown in M9/pyruvate with those grown in M9/glucose has shown (i) a decrease in the relative expression of genes involved in glycolysis and sugar metabolism (mixed acid fermentation, starch and sucrose metabolism, pentose phosphate pathway, sugar storage) and (ii) a concomitant increase in the relative expression of genes involved in gluconeogenesis, pyruvate metabolism, acetate assimilation and related pathways (**Table 1, Fig. 1**). In addition, we found that the relative abundance of transcripts encoding proteins for import and utilization of sugars other than glucose (namely *gatCYD* galactitol and sorbitol PTS transport system components and galactitol degradation machinery encoding transcripts) was also higher in cells grown on pyruvate, probably due to partial relieve of the catabolite repression.

Growth on pyruvate reduces expression of genes involved in biosynthesis of a number of major biomolecules and cellular components

E. coli cells grow two-fold slower on pyruvate than on glucose, which apparently leads to the overall reduction in the efficiency of cellular metabolism and energy production. Indeed, our data disclosed that *E. coli* growth on pyruvate (compared to its growth on glucose) leads to less efficient expression of genes involved in biosynthesis of amino acids (i.e., *dapB, lysC*), nucleosides and nucleotides (i.e., *guaA, purB*), vitamins and cofactors (i.e., *nadB, pntA*) and fatty acids (i.e., *accB, fabA*), whereas the corresponding catabolic genes (i.e. genes controlling recycling and degradation of essential cellular components (**Fig. 1**, **Tables 1 and S4**) including fatty acids (i.e., *fadA, fadE*), amino acids (i.e., *paaJ, tdh*), and cell envelope components (i.e., *ampD, ldcA*)) were expressed at higher levels.

Similar to expression of biosynthetic genes, a number of genes involved in the processing of genetic information (DNA replication, folding and repair, transcription, translation, RNA folding, ribosome biogenesis, nucleoside and nucleotide metabolism, etc.) also showed lower relative expression in cells grown on pyruvate compared to those grown on glucose. Consistent with the lower replication rate of DNA in cells grown on pyruvate, the *cspD* gene encoding an inhibitor of DNA replication was highly expressed.

We also found that nearly all the genes related to generation of energy were less abundant in cells grown in M9/pyruvate than in M9/glucose, which is likely due to the possible link between energy needs and growth rate of bacteria. Consistent with the expected reduction in energy needs during slower growth (i.e. growth on M9/pyruvate) we observed a decrease in expression level of genes involved in biofilm formation, motility and molecular transport.

While most of the transcripts encoding *trans*-membrane importers (uptake of cellular components such as nucleotides (i.e., *uraA, codB*), amino acids (i.e., *gadC, mmuP*) and compounds containing metallic ions (i.e., *btuD, fhuB*) or exporters (bacterial secretion systems and drug-proton antiporters (i.e., *emrA, gspC*) were more abundant in cells grown in M9/glucose, a very small fraction of them was found to be regulated in an opposite manner. Namely, the transcripts more abundant in cells grown on pyruvate included the *actP* and *cstA* mRNA encoding acetate and peptide importer, respectively. Similar regulatory patterns were observed for the *phoS*, *pstA* and *pstC*

transcripts encoding the components of the ATP-dependent phosphate ABC transporter, thus suggesting the importance of these transporters during cell growth on pyruvate.

Escherichia coli growth on different carbon sources alters the expression of genes involved in stress responses

Most of the stress related genes were expressed at higher level in *E. coli* cells grown on glucose (**Table S4**). The majority of them encoded enzymes with known roles in resistance to oxidative and acid stress (i.e., *katE, gadE, slp, marA, rnfC*). The regulation of some stress-related genes is likely attributed to a higher rate of *E. coli* growth on glucose (when compared to growth on pyruvate) accompanied by more efficient energy production through oxidative phosphorylation. As the latter can give rise to higher levels of reactive oxygen species (ROS) that are often toxic for the cell, bacteria need to cope with the oxidative stress by upregulating the corresponding genes.

Regulation of known sRNAs

Microarray data revealed differential regulation of several known and putative sRNAs (**Table 2**). We found that CyaR sRNA was more abundant in *E. coli* cells grown on pyruvate compared to glucose (**Table 2, Fig. 2**). This result could be anticipated, as CyaR transcription is activated by CRP regulator upon binding to cAMP, which is more abundant in the absence of glucose.

Like CyaR, RybA, RyhB and GcvB, were more efficiently expressed in *E. coli* grown on pyruvate (**Table 2, Fig. 2**). These sRNAs negatively regulate genes involved in oxidative and acid stress responses and biosynthesis of amino acids and cofactors. Accordingly, we found that some of their target mRNAs encoding enzymes involved in these cellular processes were more abundant in cells grown on glucose (*aroL*, *aroF*, *tyrA*, *sodA*, *sodB*, *ilvC* and *artJ*, **Tables 1 and S4**). Surprisingly, some transcripts encoding importers of amino acids (*brnQ, sstT, cycA*, (Pulvermacher *et al.*, 2009; Sharma *et al.*, 2011)), that are normally downregulated by GcvB, showed higher abundance in cells grown on pyruvate (**Table S4**), even though GcvB was also more abundant under these conditions (**Table 2**). Other examples of genes showing unexpected regulation include *glmS*, which was more abundant in *E. coli* cells grown

on glucose, although GlmY and GlmZ sRNAs, that normally stabilize *glmS* (Kalamorz *et al.*, 2007; Reichenbach *et al.*, 2008), were less abundant in cells grown on glucose than in those grown on pyruvate. Likewise, there was no correlation between the level of *gadE* mRNA, which was also more abundant in cells grown on glucose (**Table S4**), and its positive regulator RprA (Mika *et al.*, 2012), detected at lower concentration in cells grown on this carbon source (**Table 2**). The lack of correlation between expression levels of these small RNAs and some of their target mRNAs might reflect the counterbalancing effects caused by additional regulatory factors acting in an opposite manner on these sRNA targets, thereby illustrating the complexity and interconnection of the regulatory networks that control CCM in bacteria.

Other differentially expressed sRNAs with unknown functions

Microarray analysis also revealed differential expression of RyeA and SroA sRNAs, which are annotated in EcoCyc database (Keseler *et al.*, 2013) but their targets and regulatory roles have not been clearly defined yet. RyeA showed higher relative expression in the cells grown on pyruvate (**Table 2, Fig. 2**), and although it shares perfect complementarity with SdsR, a regulator of multidrug resistance and motility (De Lay and Gottesman, 2012; Gutierrez *et al.*, 2013), no interaction between them has been described. As to SroA, it was less abundant in cells grown in M9/pyruvate (**Table 2**). The regulatory roles of SroA have been studied in *Salmonella*, where it has shown to be upregulated upon cell exposure to tigecycline antibiotic (Yu and Schneiders, 2012). SroA was proposed to have a role in host infection (Santiviago *et al.*, 2009). In *E. coli*, the regulatory functions of SroA are still to be discovered.

In addition to these annotated but poorly characterized sRNAs, we also observed differential expression of several putative sRNAs (IS001, IS021, C0593, C0652, IS343 and tpk16, see **Table 2**), which were predicted *in silico* (Carter *et al.*, 2001; Rivas *et al.*, 2001; Chen *et al.*, 2002) and/or identified by high-throughput approaches such as deep sequencing (Sittka *et al.*, 2008) or microarray analysis (Tjaden *et al.*, 2002), but are not annotated in EcoCyc database. Further studies will be necessary to reveal their regulatory functions. Even though the regulatory functions of these known and putative sRNAs in *E. coli* remain unknown, the present work shows for the first time their possible involvement in the regulation of central carbon metabolism or related pathways.

In summary, our microarray analysis disclosed unexpected regulation of some known sRNAs and their mRNA targets, suggesting the existence of additional regulatory mechanisms controlling carbon metabolism and other related pathways, which need to be characterized in future. We also discovered the differential expression of a number of putative sRNAs for the first time, possibly indicating their functionality and involvement in central carbon metabolism and overlapping pathways.

Bibliography

Afonyushkin, T.; Vecerek, B.; Moll, I.; Bläsi, U.; Kaberdin, V. R. Both RNase E and RNase III Control the Stability of sodB mRNA upon Translational Inhibition by the Small Regulatory RNA RyhB. *Nucleic Acids Res.* **2005**, *33* (5), 1678–1689.

Aidelberg, G.; Towbin, B. D.; Rothschild, D.; Dekel, E.; Bren, A.; Alon, U. Hierarchy of Non-Glucose Sugars in *Escherichia coli*. *BMC Syst. Biol.* **2014**, *8*, 133.

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.

Andersen, J.; Delihas, N.; Kazuhiro, I.; Green, P. J.; Pines, O.; Ilercil, O.; Inouye, M. The Isolation and Characterization of RNA Coded by the micF Gene in *Escherichia coli*. *Nucleic Acids Res.* **1987**, *15* (5), 2089–2101.

Argaman, L.; Hershberg, R.; Vogel, J.; Bejerano, G.; Wagner, E. G.; Margalit, H.; Altuvia, S. Novel Small RNA-Encoding Genes in the Intergenic Regions of *Escherichia coli*. *Curr. Biol.* **2001**, *11* (12), 941–950.

Bardill, J. P.; Zhao, X.; Hammer, B. K. The Vibrio Cholerae Quorum Sensing Response Is Mediated by Hfq-Dependent sRNA/mRNA Base Pairing Interactions. *Mol. Microbiol.* **2011**, *80* (5), 1381–1394.

Basan, M.; Hui, S.; Okano, H.; Zhang, Z.; Shen, Y.; Williamson, J. R.; Hwa, T. Overflow Metabolism in *Escherichia coli* Results from Efficient Proteome Allocation. *Nature* **2015**, *528* (7580), 99–104.

Brückner, R.; Titgemeyer, F. Carbon Catabolite Repression in Bacteria: Choice of the Carbon Source and Autoregulatory Limitation of Sugar Utilization. *FEMS Microbiol.*

Lett. **2002**, *209* (2), 141–148.

Cao, Y.; Förstner, K. U.; Vogel, J.; Smith, C. J. Cis-Encoded sRNAs, a Conserved Mechanism for Repression of Polysaccharide Utilization in the *Bacteroides*. *J. Bacteriol.* **2016**, 198(18):2410-8.

Carter, R. J.; Dubchak, I.; Holbrook, S. R. A Computational Approach to Identify Genes for Functional RNAs in Genomic Sequences. *Nucleic Acids Res.* **2001**, *29* (19), 3928–3938.

Chen, S.; Lesnik, E. a; Hall, T. a; Sampath, R.; Griffey, R. H.; Ecker, D. J.; Blyn, L. B. A Bioinformatics Based Approach to Discover Small RNA Genes in the *Escherichia coli* Genome. *Biosystems.* **2002**, *65* (2–3), 157–177.

Franceschini, A.; Szklarczyk, D.; Frankild, S.; Kuhn, M.; Simonovic, M.; Roth, A.; Lin, J.; Minguez, P.; Bork, P.; von Mering, C.; *et al.* STRING v9.1: Protein-Protein Interaction Networks, with Increased Coverage and Integration. *Nucleic Acids Res.* **2013**, *41* (D1), D808–D815.

Gerstle, K.; Klätschke, K.; Hahn, U.; Piganeau, N. The Small RNA RybA Regulates Key-Genes in the Biosynthesis of Aromatic Amino Acids under Peroxide Stress in E. Coli. *RNA Biol.* **2012**, *9* (4), 458–468.

Gutierrez, a; Laureti, L.; Crussard, S.; Abida, H.; Rodríguez-Rojas, a; Blázquez, J.; Baharoglu, Z.; Mazel, D.; Darfeuille, F.; Vogel, J.; *et al.* β-Lactam Antibiotics Promote Bacterial Mutagenesis via an RpoS-Mediated Reduction in Replication Fidelity. *Nat. Commun.* **2013**, *4*, 1610.

Holms, H. Flux Analysis and Control of the Central Metabolic Pathways in *Escherichia coli*. *FEMS Microbiol. Rev.* **1996**, *19* (2), 85–116.

Inada, T.; Kimata, K.; Aiba, H. Mechanism Responsible for Glucose-Lactose Diauxie in *Escherichia coli*: Challenge to the cAMP Model. *Genes to Cells* **1996**, *1* (3), 293–301.

Johansen, J.; Eriksen, M.; Kallipolitis, B.; Valentin-Hansen, P. Down-Regulation of Outer Membrane Proteins by Noncoding RNAs: Unraveling the cAMP-CRP- and sigmaE-Dependent CyaR-ompX Regulatory Case. *J. Mol. Biol.* **2008**, *383* (1), 1–9.

Jørgensen, M. G.; Thomason, M. K.; Havelund, J.; Valentin-Hansen, P.; Storz, G. Dual Function of the McaS Small RNA in Controlling Biofilm Formation. *Genes Dev.* **2013**,

27 (10), 1132–1145.

Kalamorz, F.; Reichenbach, B.; März, W.; Rak, B.; Görke, B. Feedback Control of Glucosamine-6-Phosphate Synthase GlmS Expression Depends on the Small RNA GlmZ and Involves the Novel Protein YhbJ in *Escherichia coli*. *Mol. Microbiol.* **2007**, *65* (6), 1518–1533.

Kang, S.-M.; Choi, J.-W.; Lee, Y.; Hong, S.-H.; Lee, H.-J. Identification of microRNA-Size, Small RNAs in *Escherichia coli*. *Curr. Microbiol.* **2013**, 67(5):609-13..

Kawano, M.; Reynolds, A. a; Miranda-Rios, J.; Storz, G. Detection of 5'- and 3'-UTR-Derived Small RNAs and Cis-Encoded Antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.* **2005**, *33* (3), 1040–1050.

Keseler, I. M.; Mackie, A.; Peralta-Gil, M.; Santos-Zavaleta, A.; Gama-Castro, S.; Bonavides-Martínez, C.; Fulcher, C.; Huerta, A. M.; Kothari, A.; Krummenacker, M.; *et al.* EcoCyc: Fusing Model Organism Databases with Systems Biology. *Nucleic Acids Res.* **2013**, *41* (Database issue), D605-12.

De Lay, N.; Gottesman, S. The Crp-Activated Small Noncoding Regulatory RNA CyaR (RyeE) Links Nutritional Status to Group Behavior. *J. Bacteriol.* **2009**, *191* (2), 461–476.

De Lay, N.; Gottesman, S. A Complex Network of Small Non-Coding RNAs Regulate Motility in *Escherichia coli*. *Mol. Microbiol.* **2012**, *86* (3), 524–538.

Liu, M. Y.; Gui, G.; Wei, B.; Iii, J. F. P.; Oakford, L.; Giedroc, D. P.; Romeo, T. The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagonizes Its Activity in *Escherichia coli* The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagoniz. *J. Biol. Chem.* **1997**, *272* (28), 17502–17510.

Majdalani, N.; Hernandez, D.; Gottesman, S. Regulation and Mode of Action of the Second Small RNA Activator of RpoS Translation, RprA. *Mol. Microbiol.* **2002**, *46* (3), 813–826.

Massé, E.; Majdalani, N.; Gottesman, S. Regulatory Roles for Small RNAs in Bacteria. *Curr. Opin. Microbiol.* **2003**, *6* (2), 120–124.

Massé, E.; Salvail, H.; Desnoyers, G.; Arguin, M. Small RNAs Controlling Iron Metabolism. *Curr. Opin. Microbiol.* **2007**, *10* (2), 140–145.

Michaux, C.; Verneuil, N.; Hartke, A.; Giard, J. Physiological Roles of Small RNA Molecules. *Microbiology* **2014**, No. 33, 1–33.

Mika, F.; Busse, S.; Possling, A.; Berkholz, J.; Tschowri, N.; Sommerfeldt, N.; Pruteanu, M.; Hengge, R. Targeting of csgD by the Small Regulatory RNA RprA Links Stationary Phase, Biofilm Formation and Cell Envelope Stress in *Escherichia coli*. *Mol. Microbiol.* **2012**, *84* (1), 51–65.

Møller, T.; Franch, T.; Udesen, C.; Gerdes, K.; Valentin-Hansen, P. Spot 42 RNA Mediates Discoordinate Expression of the E. Coli Galactose Operon. *Genes Dev.* **2002**, *16* (13), 1696–1706.

Morin, M.; Ropers, D.; Letisse, F. The Post-Transcriptional Regulatory System CSR Controls the Balance of Metabolic Pools in Upper Glycolysis of *Escherichia coli* . **2016**, *100* (February), 686–700.

Papenfort, K.; Said, N.; Welsink, T.; Lucchini, S.; Hinton, J. C. D.; Vogel, J. Specific and Pleiotropic Patterns of mRNA Regulation by ArcZ, a Conserved, Hfq-Dependent Small RNA. *Mol. Microbiol.* **2009**, *74* (1), 139–158.

Porcheron, G.; Habib, R.; Houle, S.; Caza, M.; Lépine, F.; Daigle, F.; Massé, E.; Dozois, C. M. The Small RNA RyhB Contributes to Siderophore Production and Virulence of Uropathogenic *Escherichia coli*. *Infect. Immun.* **2014**, *82* (12), 5056–5068.

Pulvermacher, S. C.; Stauffer, L. T.; Stauffer, G. V. Role of the sRNA GcvB in Regulation of cycA in *Escherichia coli*. *Microbiology* **2009**, *155* (1), 106–114.

Raghavan, R.; Groisman, E. A.; Ochman, H. Genome-Wide Detection of Novel Regulatory RNAs in E . Coli. *Genome Res.* **2011**, *21*, 1487–1497.

Reichenbach, B.; Maes, A.; Kalamorz, F.; E, H.; Görke, B. The Small RNA GlmY Acts Upstream of the sRNA GlmZ in the Activation of glmS Expression and Is Subject to Regulation by Polyadenylation in *Escherichia coli*. *Nucleic Acids Res.* **2008**, *36* (8), 2570–2580.

Reichenbach, B.; Göpel, Y.; Görke, B. Dual Control by Perfectly Overlapping Sigma 54- and Sigma 70- Promoters Adjusts Small RNA GlmY Expression to Different Environmental Signals. *Mol. Microbiol.* **2009**, *74* (5), 1054–1070.

Richards, G. R.; Vanderpool, C. K. Molecular Call and Response: The Physiology of

Bacterial Small RNAs. *Biochim. Biophys. Acta* **2011**, *1809* (10), 525–531.

Rivas, E.; Klein, R. J.; Jones, T. a; Eddy, S. R. Computational Identification of Noncoding RNAs in E. Coli by Comparative Genomics. *Curr. Biol.* **2001**, *11* (17), 1369–1373.

Ruiz-Larrabeiti, O.; Plágaro, A. H.; Gracia, C.; Gallego, L.; Hajnsdorf, E.; Kaberdin, V. R. A New Custom Microarray for sRNA Profiling in *Escherichia coli*. *FEMS Microbiol. Lett.* **2016**, 1–20.

Sætrom, P.; Sneve, R.; Kristiansen, K. I.; Snøve Jr., O.; Grünfeld, T.; Rognes, T.; Seeberg, E. Predicting Non-Coding RNA Genes in *Escherichia coli* with Boosted Genetic Programming. *Nucleic Acids Res.* **2005**, *33* (10), 3263–3270.

Santiviago, C. A.; Reynolds, M. M.; Porwollik, S.; Choi, S. H.; Long, F.; Andrews-Polymenis, H. L.; McClelland, M. Analysis of Pools of Targeted Salmonella Deletion Mutants Identifies Novel Genes Affecting Fitness during Competitive Infection in Mice. *PLoS Pathog.* **2009**, *5* (7).

Sharma, C. M.; Darfeuille, F.; Plantinga, T. H.; Vogel, J. A Small RNA Regulates Multiple ABC Transporter mRNAs by Targeting C/A-Rich Elements inside and Upstream of Ribosome-Binding Sites. *Genes Dev.* **2007**, *21* (21), 2804–2817.

Sharma, C. M.; Papenfort, K.; Pernitzsch, S. R.; Mollenkopf, H.-J.; Hinton, J. C. D.; 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.

Shinhara, A.; Matsui, M.; Hiraoka, K.; Nomura, W.; Hirano, R.; Nakahigashi, K.; Tomita, M.; Mori, H.; Kanai, A.; Kana, A. Deep Sequencing Reveals as-yet-Undiscovered Small RNAs in *Escherichia coli*. *BMC Genomics* **2011**, *12* (1), 428.

Sittka, A.; Lucchini, S.; Papenfort, K.; Sharma, C. M.; Rolle, K.; Binnewies, T. T.; Hinton, J. C. D.; Vogel, J. Deep Sequencing Analysis of Small Noncoding RNA and mRNA Targets of the Global Post-Transcriptional Regulator, Hfq. *PLoS Genet.* **2008**, *4* (8), e1000163.

Stülke, J.; Hillen, W. Carbon Catabolite Repression. *Curr. Opin. Microbiol.* **1999**, *2*, 195–201.

Tjaden, B.; Saxena, R. M.; Stolyar, S.; Haynor, D. R.; Kolker, E.; Rosenow, C. Transcriptome Analysis of *Escherichia coli* Using High-Density Oligonucleotide Probe Arrays. *Nucleic Acids Res.* **2002**, *30* (17), 3732–3738.

Urban, J. H.; Vogel, J. Two Seemingly Homologous Noncoding RNAs Act Hierarchically to Activate glmS mRNA Translation. *PLoS Biol.* **2008**, *6* (3), e64.

Urbanowski, M. L.; Stauffer, L. T.; Stauffer, G. V. The gcvB Gene Encodes a Small Untranslated RNA Involved in Expression of the Dipeptide and Oligopeptide Transport Systems in *Escherichia coli*. *Mol. Microbiol.* **2000**, *37* (4), 856–868.

Vanderpool, C. K.; Gottesman, S. Involvement of a Novel Transcriptional Activator and Small RNA in Post-Transcriptional Regulation of the Glucose Phosphoenolpyruvate Phosphotransferase System. *Mol. Microbiol.* **2004**, *54* (4), 1076–1089.

Vogel, J.; Bartels, V.; Tang, T. H.; Churakov, G.; Slagter-Jäger, J. G.; Hüttenhofer, A.; Wagner, E. G. H. RNomics in *Escherichia coli* Detects New sRNA Species and Indicates Parallel Transcriptional Output in Bacteria. *Nucleic Acids Res.* **2003**, *31* (22), 6435–6443.

Wassarman, K. M.; Repoila, F.; Rosenow, C.; Storz, G.; Gottesman, S. Identification of Novel Small RNAs Using Comparative Genomics and Microarrays. *GENES Dev.* **2001**, *15*, 1637–1651.

Weilbacher, T.; Suzuki, K.; Dubey, A. K.; Wang, X.; Gudapaty, S.; Morozov, I.; Baker, C. S.; Georgellis, D.; Babitzke, P.; Romero A, D. A Novel sRNA Component of the Carbon Storage Regulatory System of *Escherichia coli*. *Mol. Microbiol.* **2003**, *48* (3), 657–670.

Yachie, N.; Numata, K.; Saito, R.; Kanai, A.; Tomita, M. Prediction of Non-Coding and Antisense RNA Genes in *Escherichia coli* with Gapped Markov Model. *Gene* **2006**, *372*, 171–181.

Yu, J.; Schneiders, T. Tigecycline Challenge Triggers sRNA Production in Salmonella Enterica Serovar Typhimurium. *BMC Microbiol.* **2012**, *12* (1), 195.

Zhang, A.; Wassarman, K. M.; Rosenow, C.; Tjaden, B.; Storz, G.; Gottesman, S. Global Analysis of Small RNA and mRNA Targets of Hfq. *Mol. Microbiol.* **2003**, *50* (4), 1111–1124.

Chapter 3

Discovery of new small RNAs by using RNA sequencing approaches

Introduction

Early work on characterization of sRNAs (reviewed by Wassarman *et al.* (1999)) established their important roles in bacterial stress responses. Since then, numerous studies have focused on identification of sRNAs and their characterization in diverse prokaryotic organisms. As the number of identified and studied sRNAs increased, their common characteristics were deduced and further utilized to develop bioinformatic tools and experimental approaches used to search for novel small RNAs and characterize their biological functions. These studies greatly expanded our knowledge of the regulatory networks that control bacterial metabolism and response to environmental signals at the post-transcriptional level.

Discovery of new sRNAs usually requires a combination of *in silico* and experimental approaches. The common characteristics of annotated *trans*-encoded sRNAs (average length, location in intergenic regions (IGR), the conservation of their sequence among closely related species, the presence of a promoter and a ρ-independent terminator) were used to eventually create complex algorithms for sRNA prediction (Rivas *et al.*, 2001; Wassarman *et al.*, 2001; Chen *et al.*, 2002; Sætrom *et al.*, 2005; Yachie *et al.*, 2006). In addition to *in silico* tools, several experimental approaches were developed to discover novel sRNAs in *E. coli* and other enterobacteria. Some of them identified sRNAs by analysing RNA molecules that co-immuno-precipitated with the RNA chaperone Hfq, whereas others employed shot-gun cloning approaches to characterize the low molecular weight fraction of total RNA (Tjaden *et al.*, 2002; Vogel *et al.*, 2003; Zhang *et al.*, 2003; Kawano *et al.*, 2005).

In the last years RNA sequencing (RNA-seq) has emerged as a new tool enabling the analysis of the entire transcriptomes of pro- and eukaryotic organisms at single nucleotide resolution. Development of RNA-seq was revolutionary because it allowed detection of novel transcripts and measurement of their expression level without prior knowledge of their sequences with great sensitivity, which, in contrast to

detection by high-density microarrays, does not depend on hybridization efficiency. Recently RNA-seq has been employed to validate *in silico* predicted transcripts and discover new putative sRNAs in *E. coli* (Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Kang *et al.*, 2013). Moreover, recent studies have adapted RNA-seq for direct identification of transcription start sites (TSS). This modification of RNA-seq enabled differentiation between metastable decay intermediates derived from larger protein-coding transcripts and small (50 to 300 nucleotides long, (Hershberg *et al.*, 2003)), non-coding primary transcripts which could potentially represent regulatory RNAs (Sharma *et al.*, 2010; Lin *et al.*, 2013).

Similar to other high-throughput approaches, RNA-seq has a number of limitations, and can occasionally lead to errors. Thus, further validation of RNA-seq data by independent methods is often required (Weirick *et al.*, 2015). A number of techniques such as northern blot, RT-qPCR and 3'-RACE have been used to confirm the presence, length, and relative abundance of individual transcripts.

In this work we analysed RNA-seq data from various sources to identify new sRNAs and discuss the differences in sRNA identification obtained by different RNA-seq approaches. The analysis was performed for a set of putative sRNAs envisaged by previous RNA-seq studies (Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Conway *et al.*, 2014) as well as those discovered by the dRNA-seq approach performed in collaboration with the group of Dr. Kenneth McDowall from the University of Leeds (Romero *et al.*, 2014), followed by testing their presence in the total pool of *E. coli* transcripts by northern blotting. While some of the putative sRNAs provided good signals, thereby confirming the existence of these sRNAs *in vivo* and validating their size, analysis of other candidates revealed signals corresponding to longer transcripts, or yielded no signal at all. Additionally, we analysed the differences in the experimental procedures of the four independent RNA-seq approaches to compare their potential to detect known and discover putative small RNAs. Here we present the results of our analysis and discuss advantages and limitations of RNA-seq approaches currently used for detection and discovery of bacterial sRNAs.

Materials and Methods

Part of the work presented in this chapter is the result of our collaboration with the group of Dr. Kenneth McDowall (University of Leeds) (Romero. *et al.*, 2014). The initial efforts undertaken by his group in collaboration with others yielded a set of RNA-seq data, which was further analysed and validated by us. Namely, our work was focused on analysis of the signals derived from the sequences corresponding to *E. coli* intergenic regions with the aim of identifying putative sRNA transcripts and to further validate their existence by northern blot analysis.

Bacterial strains and growth conditions

E. coli K-12 strain BW25113 (Baba *et al.*, 2006) was used by Romero *et al.* (2014) for the differential RNA sequencing experiments. For testing and validation of their results by northern blotting, we used *E. coli* K-12 MG1655 strain. Both *E. coli* strains were cultured at 37 °C with shaking (100 rpm) in Luria-Bertani broth (Sigma or Amresco) or in M9 minimal media (Sigma) supplemented with glucose (0.4% w/v) until mid-logarithmic phase ($O.D._{600}$~0.5).

RNA isolation

An aliquot of 8 mL of cell culture was mixed with 1 mL of stop solution (5% phenol in ethanol) to inhibit cell metabolism and the cells were harvested by centrifugation (15 min, 4 °C, 4000 *g*).

To carry out RNA-seq analysis, total RNA was isolated from *E. coli* BW25113 as described previously for this organism (Kime *et al.*, 2008). To remove contaminating genomic DNA, samples were treated with DNase I using conditions described by the vendor (Ambion) and RNA was extracted with phenol: chloroform (1:1). The concentration and integrity of RNA samples were determined using a NanoDropTM 1000 spectrophotometer (Thermo Fisher Scientific) and agarose gel electrophoresis (Kime *et al.*, 2008) respectively. Samples were then enriched for mRNA using MICROBExpress^{TM}-Bacteria beads, as described by the manufacturer (Ambion).

To perform northern blot analysis, total RNA was isolated as described for the RNA-seq analysis, with the exception that no depletion of rRNA was done. Further

details on RNA isolation for northern blotting are specified in the Materials and Methods in **Chapter 1**.

Differential RNA sequencing analysis

RNA-seq data were generated by Vertis Biotechnologie AG (Germany) as a service to the group of Dr. Kenneth McDowall (University of Leeds). Briefly, the procedure included the construction of cDNA libraries before and after treatment with tobacco acid pyrophosphatase (TAP), RNA sequencing and alignment of RNA sequence reads to the corresponding genome positions (Lin *et al.*, 2013), which were retrieved from the NCBI database (Pruitt *et al.*, 2007). For each genome position, the number of times it was the first nucleotide in sequence reads, i.e. associated with a 5′ end *in vivo*, was counted. This was done separately for each of the two libraries and the counts were compared. It should be noted that as described previously (Lin *et al.*, 2013) the 5′-sequencing adaptor was ligated to transcripts prior to fragmentation, thereby allowing the 5′ ends of both long and short transcripts to be detected. Global RNA-seq was performed at the WellcomeTrust Sanger Centre (Cambridge, UK) using FRT-seq, in which reverse transcription takes place in the flowcell, avoiding PCR biases and duplicates (Mamanova *et al.*, 2010b), and the sequences were processed as described previously (Lin *et al.*, 2013). After aligning the gRNA-seq reads to the genome, the number of times each genome position was present in a read irrespective of its position was counted. The reference genome for *E. coli* K-12 strain BW25113 was U00096.2. The RNA-seq data were deposited into in the GEO archive (Barrett *et al.*, 2013) under the GSM1126845 accession number.

Northern blotting

Details of the protocol for northern blotting are provided in **Chapter 1**. Briefly, total RNA was isolated from *Escherichia coli* K-12 MG1655 grown in Luria-Bertani or M9 minimal media supplemented with glucose and was separated by denaturing electrophoresis, electro-transferred to a membrane, and hybridized to radiolabelled probes specifically designed to be complementary to each potential sRNA (listed in **Table S6**).

Results

Bioinformatic analysis and northern blot validation of potential sRNAs found by differential RNA sequencing

During our collaboration with McDowall's team (University of Leeds, UK) we searched for new putative *trans*-encoded small RNAs by using a single-nucleotide resolution map of *E. coli* transcriptome including transcriptional start sites (TSS) identified by differential RNA sequencing (dRNA-seq). To identify TSS, *E. coli* RNA samples were divided into two aliquots. One aliquot was treated with tobacco acid pyrophosphatase (TAP), which converts the triphosphorylated 5′ ends of primary transcripts into monophosphorylated ones, whereas the second aliquot was used without TAP treatment. Then, 5′-sequencing adaptor was ligated to the 5'-monophosphorylated ends of transcripts of each sample and they were separately analysed by RNA-seq and compared. A significant increase in the number of sequencing reads at a specific position in the TAP-treated aliquot provided an identifier of a TSS (Romero. *et al.*, 2014). Analysis of these dRNA-seq data in the UCSC microbial genome browser (Chan *et al.*, 2012) enabled to compile a list of primary transcripts originated from intergenic regions (i.e., those showing clear evidence for TSS). In order to find the best putative sRNA candidates, we refined this list by eliminating transcripts potentially originated from 5' and 3' untranslated regions (UTRs) of flanking genes that were possibly produced by premature termination or represented potential metastable decay intermediates, and intergenic regions containing repetitive and palindromic REP sequences related to transposition of mobile elements (Higgins *et al.*, 1988; Tobes and Pareja, 2006) listed in EcoCyc database (Keseler *et al.*, 2013). Other parameters taken into account during the preliminary screening were the presence of a potential ρ-independent terminator (predicted *in silico* using the mfold Web Server (Zuker, 2003)), and a favouring orientation of the putative sRNA and its flanking genes allowing a clear distinction between them (i.e., divergent orientation of the putative sRNAs and the flanking genes). This preliminary screening revealed 10 putative sRNAs (**Table 1**) that had not been previously identified in other studies, except one sRNA candidate (i.e. p-sRNA-6 / ecr1743d+), independently found by Conway *et al.* (2014). We further tested these small RNA candidates by northern blotting and found that 6 of them yielded positive signals (**Table 1**, **Fig. 1**). The good rate of sRNA detection demonstrated the high efficiency of sRNA discovery achieved by differential RNA sequencing in combination with bioinformatic curation of the putative sRNA list.

Table 1. New putative sRNAs found by dRNA-seq. The table contains small RNA candidates identified by Romero *et al.* (2014) that were pre-screened (see text) and tested by northern blotting. Arbitrary names have been given to the putative sRNAs (p-sRNA-#) to facilitate their discussion in the text.

sRNA names	Flanking genes, orientation sRNA (→) and flanking genes (→), sRNA left-right ends	Length (nt)	NB	Found by
p-sRNA-1 0174u+	*dxr-ispU* →→→ 194784-194848	64	✓	Romero
p-sRNA-2 TSS-54	*ygcW-yqcE* ←→→ 2898527-2898590	63	✓	Romero
p-sRNA-3 4051d-	*pspG-qor* →→← 4261162-4261273	111	✓	Romero
p-sRNA-4 TSS-81	*metF-katG* →←→ 4131585-4131791	206	✓	Romero
p-sRNA-5 2775u+	*ppiC-yifO* ←←← 3957925-3958004	79	✓	Romero
p-sRNA-6 1743d+	*ves-spy* ←←← 1823080-1823177	97	✓	Romero, Conway
p-sRNA-7 0770d-	*ybhI-ybhJ* →←→ 802544-802660	116	✘	Romero
p-sRNA-8 TSS-40	*psuK-fruA* ←→← 2257512-2257740	228	✘	Romero
p-sRNA-9 TSS-41	*yeiQ-yeiR* →←→ 2265741-2265816	75	✘	Romero
p-sRNA-10 TSS-50	*pepB-iscX* ←→← 2654399-2654550	151	✘	Romero

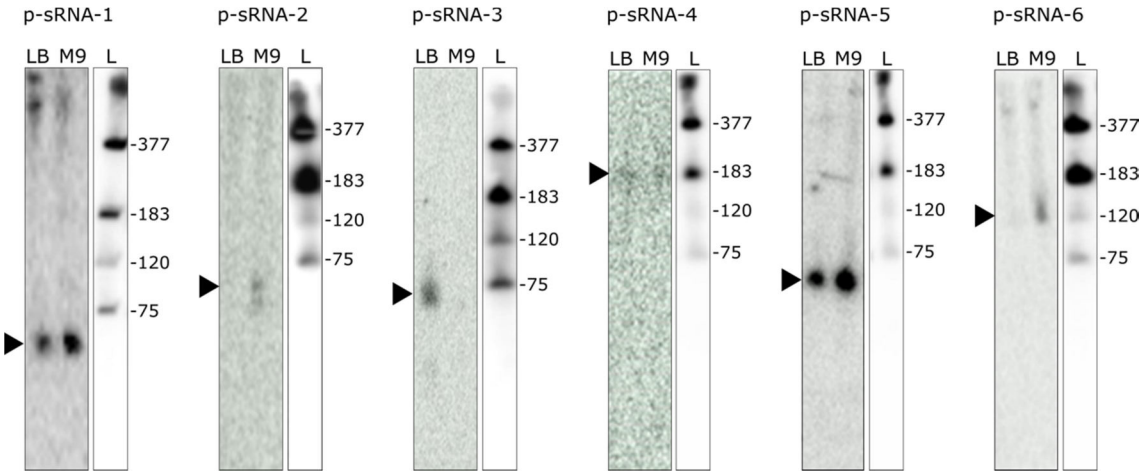


Figure 1. Northern blot validation of new putative sRNAs discovered by dRNA-seq. Total RNA isolated from the *E. coli* MG1655 strain grown in LB or M9/glucose medium (lanes LB and M9, respectively) was used to test the existence of the selected intergenic primary transcripts (**Table 1**) found by dRNA-seq (Romero *et al.*, 2014) by northern blotting (details on the experimental procedure can be found in Materials and Methods in **Chapter 1**, probes are listed in **Table S8**). The positions of small RNAs are indicated by arrowheads. Ladder (L) was generated by hybridization of total RNA with radiolabelled probes complementary to RnpB (M1) RNA (377 nt), 6S RNA (183 nt), 5S rRNA (120 nt) and $tRNA^{Asn}$ (75 nt). The sequences of the probes are listed in **Table S8**.

Comparison of four independent RNA sequencing studies and their capacity to detect small RNAs

By the time our collaboration with the group of Dr. Kenneth McDowall on the analysis of the dRNA-seq data had been completed, several other groups had also used RNA-seq approaches to analyse *E. coli* transcriptome (Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Conway *et al.*, 2014) and disclosed a number of new potential sRNAs. Comparison of their results revealed a considerable variability in the number and nature of the putative sRNAs discovered by each study.

To learn more about the factors and experimental conditions that could be critical for efficient search and detection of small RNAs by using RNA-seq, we compared the procedures employed by each study (Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Conway *et al.*, 2014; Romero. *et al.*, 2014). For this purpose, we analysed the experimental workflows of their approaches (**Fig. 2**). We found some major differences in the RNA isolation protocols. One of them is the use of RNeasy Kit columns (Qiagen) in the study carried out by Conway *et al.* (2014) and avoided by the rest of the studies, which might result in some loss of low-molecular-weight RNAs, thus

possibly causing under-representation of sRNAs in the output data. Another difference is the special treatment of RNA done only by Shinhara *et al.* (2011) to protect labile sRNAs from degradation. In addition to the above differences, two groups used differential RNA-seq procedures for TSS identification (dRNA-seq) (Conway *et al.*, 2014; Romero. *et al.*, 2014), whereas others utilized standard RNA-seq protocols. Moreover, in the study of Romero. *et al.* (2014) the preparation of cDNA libraries was also different from others because the reverse transcription was carried out in the flowcell (FRT-seq) to avoid the bias caused by PCR. Finally, the different studies also applied different parameters for the processing of RNA-seq data and sRNA search (**Fig. 2**). The latter was used to select either only *trans*-encoded candidate sRNA (Raghavan *et al.*, 2011; Romero. *et al.*, 2014) or both *cis*- and *trans*-encoded transcripts (Shinhara *et al.*, 2011; Conway *et al.*, 2014)).

	Conway 2014	Romero 2014	Raghavan 2011	Shinhara 2011
Bacterial strain and culture conditions 	*Escherichia coli* BW38028 MOPS minimal medium supplemented with glucose Logarithmic- and stationary-phase	*Escherichia coli* BW25113 Luria-Bertani broth Mid-Logarithmic-phase	*Escherichia coli* MG1655 N-minimal media supplemented with $MgCl_2$, casamino acids and glucose Mid-Logarithmic-phase	*Escherichia coli* BW25113 M63 glucose minimal medium Late-Logarithmic-phase
Treatment of isolated RNA 	Columns were used for RNA isolation RNA was divided in TEX-treated and not treated aliquots	RNA sample was enriched for mRNA Use of columns was avoided for RNA isolation RNA was divided in TAP-treated and not treated aliquots	RNA sample was enriched for mRNA Use of columns was avoided for RNA isolation	sRNAs with short half-lives were protected Use of columns was avoided for RNA isolation RNA was size-selected (<200 nt) RNA sample was TAP-treated
cDNA preparation, sequencing and sRNA search 	Unnanotated long primary transcripts not containing ORFs were listed as long ncRNAs	5' adapter was ligated prior to RNA fragmentation FRT-seq was employed Primary transcripts in IGRs were selected REP sequences were removed sRNA candidates were tested by NB	cDNA was size-selected (100-400 bp) IGR>150 nt were selected REP sequences were removed TSS and terminators were predicted *in silico* sRNA candidates were tested by 3' -RACE Hfq dependence was tested by RT-qPCR	Non-annotated multiple regions were removed Transcribed regions with computational or experimental evidence of TSS were selected Transcripts >50 nt were selected sRNA candidates were tested by NB

Figure 2. Workflow of four independent RNA-seq studies of the *E. coli* transcriptome. This table shows the differences in the methods applied in the RNA-seq studies performed by Raghavan *et al.* (2011), Shinhara *et al.* (2011), Conway *et al.* (2014) and Romero *et al.* (2014).

To evaluate how the differences in the experimental methods employed in these studies affected the efficiency of sRNA discovery, we compared the number and nature of the known sRNAs detected by each work. Small RNAs were considered as 'known' if they were among the transcripts annotated in EcoCyc database (www.ecocyc.org (Keseler *et al.*, 2013)) as 'small regulatory RNA', 'antisense RNA' or had similar definitions, excluding rRNAs, tmRNAs, tRNAs or transcripts originated from phantom genes and pseudo genes. In total 76 known sRNAs were detected in these studies (**Table S6**). **Fig. 3A** represents the number of known sRNAs detected by each study (in grey). There were clear differences in the capacity of each study to detect known sRNAs. While Conway *et al.* (2014) detected only 14 known sRNAs, the number of sRNAs detected by Romero. *et al.* (2014), Raghavan *et al.* (2011), Shinhara *et al.* (2011) were considerable higher (50, 61 and 65, respectively). The lower number of known sRNAs detected by Conway *et al.* (2014) compared to other studies is probably caused by the use of RNeasy kit (Qiagen) columns during RNA purification steps, which, according to the manufacturer's instructions, is mainly efficient for isolation of RNA molecules longer than 200 nt. The rate of detection of new putative sRNAs was consistent with that observed for detection of known sRNAs (**Fig. 3A**, in blue). Namely, Raghavan *et al.* (2011) and Shinhara *et al.* (2011) could detect more putative sRNAs (177 and 108, respectively) than Conway *et al.* (2014) and Romero. *et al.* (2014) (18 and 41 in each case).

The comparison also revealed that among the known sRNAs that were detected by RNA-seq, the majority was found by at least two independent studies, and only 14 of them were exclusively detected in a single study (**Fig. 3B**).

Testing new putative sRNAs found by several RNA sequencing studies

The results obtained by comparison of the known sRNAs detected in different RNA-seq studies (**Fig. 3B**) led us to the idea that identification of the same putative sRNAs in independent studies could be indicative of their true existence. While testing this idea, we found that only 4 putative sRNAs were detected by three RNA-seq studies, and none of them was detected by all of them (**Fig. 3C**, **Table S7**).

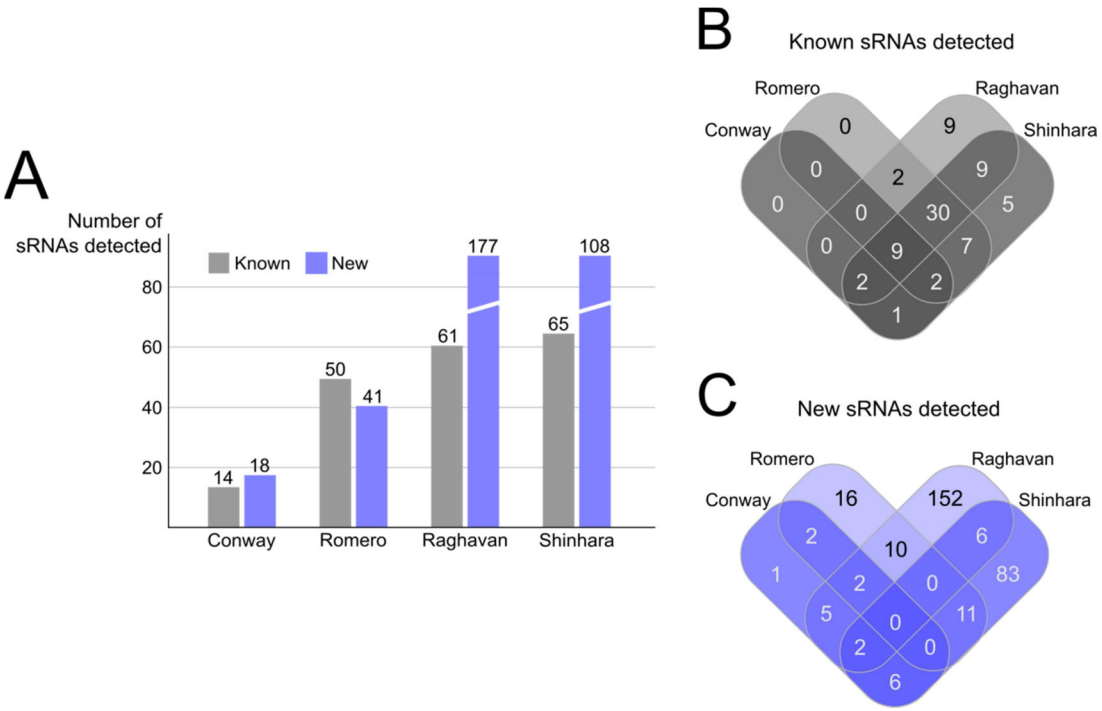


Figure 3. Known and new putative small RNAs detected by four independent RNA-seq studies. A) Total number of sRNAs (Y-axis, grey and blue bars correspond to known and new putative sRNAs, respectively) detected in each of the four independent studies included in the comparison (Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Conway *et al.*, 2014; Romero A. *et al.*, 2014), (X-axis). **B) C)** Venn diagrams representing overlaps between the sets of known sRNAs **(B)** or new putative **(C)** sRNAs detected in the above RNA-seq studies.

We used northern blot analysis to validate three potential sRNAs that were identified independently by three of the RNA-seq studies (IS019, and G0-10698 and G0-10699, **Table 3**), and found that they yielded positive signals (**Fig. 4**). The putative sRNA left undetected here (ECS698) was not recognized as a primary transcript by Romero *et al.* (2014), has not been further analysed to our knowledge, and may not exist *in vivo*.

To further test the idea that identification of putative sRNAs by independent approaches might constitute an important criterion that needs to be taken into account during pre-screening of sRNA candidates for validation and/or characterization, we selected a group of putative sRNAs found by us (Romero A. *et al.*, 2014) and other RNA-seq studies (Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Conway *et al.*, 2014), and subsequently chose the best candidates for further experimental validation based on the criteria mentioned above (i.e. no overlap with a REP sequence, presence of TSS and a potential ρ-independent terminator, optimal orientation of sRNA and flanking

genes). This preliminary analysis led to selection of 9 candidates for northern blot verification (**Table 3**), of which 6 yielded positive signals (**Figure 4**).

Table 3. Putative sRNAs found in several independent RNA-seq studies. These sRNAs were selected for validation by northern blot analysis. References indicate the studies that led to detection of these putative sRNAs by using RNA-seq (in bold), or other approaches.

sRNA name	Orientation of sRNA (→) and flanking genes (→), sRNA left-right ends	Length (nt)	NB	References
IS019	*nmpC-essD* ←→→ 576438-576505	67	✓	**Romero *et al.*, 2014,** **Raghavan *et al.*, 2011,** **Conway *et al.*, 2014,** Chen *et al.*, 2002
G0-10698	*ynfM-asr* →←→ 1668973-1669160	188	✓	**Shinhara *et al.*, 2011,** **Raghavan *et al.*, 2011,** **Conway *et al.*, 2014** Wassarman *et al.*, 2001, Chen *et al.*, 2002, Yachie *et al.*, 2006
G0-10699	*sdiA-yecC* ←→← 1994930-1995121	192	✓	**Romero *et al.*, 2014,** **Raghavan *et al.*, 2011,** **Conway *et al.*, 2014** Chen *et al.*, 2002, Tjaden *et al.*, 2002, Zhang *et al.*, 2003, Yachie *et al.*, 2006
p-sRNA-11	*ydfU-rem* ←←← 1642532-1642655	123	✓	**Romero *et al.*, 2014,** **Raghavan *et al.*, 2011**
ECS054	*intZ-yffL* →←→ 2558320-2558395	75	✓	**Romero *et al.*, 2014,** **Shinhara *et al.*, 2011**
ECS674	*rodZ-rlmN* ←←← 2640965-2641027	62	✘	**Romero *et al.*, 2014,** **Shinhara *et al.*, 2011**
C0614	*sseA-IS628* →←→ 2651370-2651478	108	✓	**Romero *et al.*, 2014,** **Shinhara *et al.*, 2011** Tjaden *et al.*, 2002
p-sRNA-12 3777d+	*zupT-ribB* →←← 3181345-3181648	303	✓	**Romero *et al.*, 2014,** **Shinhara *et al.*, 2011,**
ECS663	*thiC-rsd* ←←← 4194200-4194325	125	✓	**Romero *et al.*, 2014,** **Shinhara *et al.*, 2011**
G0-10700	*yejG-bcr* ←←← 2276280-2276520	241	✘	**Shinhara *et al.*, 2011,** **Raghavan *et al.*, 2011** Wassarman *et al.*, 2001

G0-10701 (och5)	*ygfI-yggE* ←→← 3065209-3065366	158	✓	**Shinhara *et al.*, 2011, Raghavan *et al.*, 2011**, Bak *et al.*, 2015
G0-10705	*glnA-typA* ←→→ 4056194-4056265	72	✘	**Shinhara *et al.*, 2011, Raghavan *et al.*, 2011** Rivas *et al.*, 2001, Sætrom *et al.*, 2005

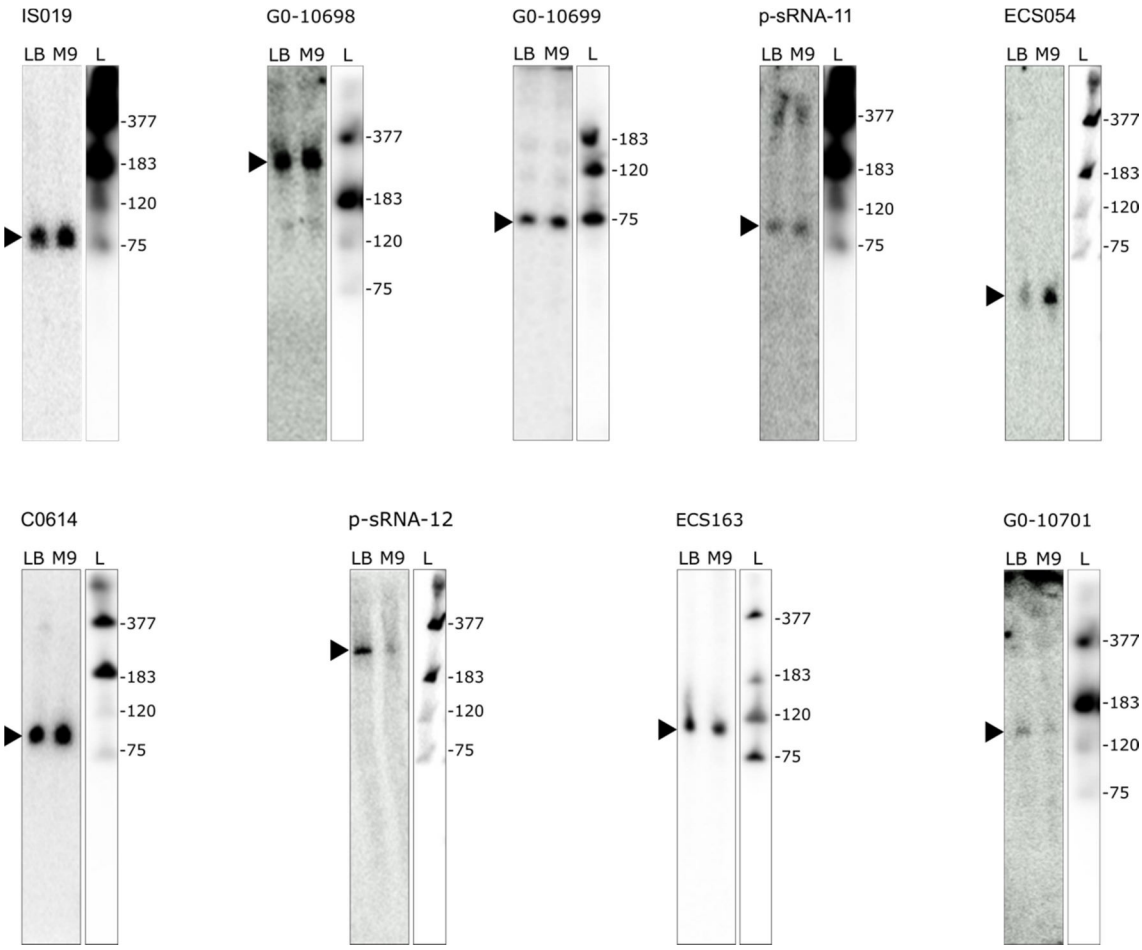


Figure 4. Northern blot detection of selected putative sRNAs previously identified by independent RNA-seq studies. The sRNA candidates of **Table 3** were analysed by northern blot as described in the figure legend of **Figure 1**. The positions of small RNAs are indicated by arrowheads and molecular ladder sizes are indicated in nucleotides. The sequences of the probes are listed in **Table S8**.

Differential expression of putative sRNAs in *E. coli* grown in LB and M9/glucose

Trans-encoded small RNAs often play significant roles in bacterial responses to stress (e.g., suboptimal temperature, media composition and pH (reviewed in Wagner and Romby, 2015)) and are transcriptionally regulated under those conditions. In order to test whether their transcription/stability is affected by the composition of growth medium used, validation of sRNA by northern blot analysis was done by using total RNA isolated from *E. coli* cells grown in rich (Luria-Bertani, LB) and minimal (M9 supplemented with glucose, M9/glucose) media. These experiments revealed considerable variations in the abundance of a number of putative sRNAs. We found that p-sRNA-3 (**Fig. 1**) and p-sRNA-12 (**Fig. 4**) seemed to be more efficiently expressed in *E. coli* cells grown in LB, whereas p-sRNA-1, 2, 5 and 6 (**Fig. 1**) and ECS054 (**Fig. 4**) were apparently more abundant in cells grown in M9/glucose. The differential expression of these sRNAs could be indicative of their possible role in regulation of targets involved in the metabolic pathways (e.g. glycolysis, amino acid metabolism, biosynthesis of membrane transporters, etc.) affected by the composition of the growth media (for examples, see **Chapter 1**).

Discussion

RNA-seq is a powerful tool for high-throughput analysis of transcriptomes and it has provided new means to both investigate and corroborate previous findings on transcript boundaries, RNA processing and degradation events and operon architecture, in pro- and eukaryotic cells. New variants of RNA-seq have been developed for specific purposes, such as sRNA interactome analysis (Melamed *et al.*, 2016), and detection of transcription start sites (Sharma *et al.*, 2010). The latter has extensively contributed to identification of small RNAs and transcriptomic mapping, helping to reveal the high complexity of the RNA landscape of more than 25 bacterial and archaeal species (Sharma and Vogel, 2014). Like every other technique, RNA-seq has limitations and challenging aspects (Wang *et al.*, 2009; Hirsch *et al.*, 2015), and the results obtained require validation by alternative methods to assure their reliability (Weirick *et al.*, 2015). In collaboration with the group of Dr. McDowall (University of Leeds), we carried out *in silico* analysis of the primary intergenic transcripts identified by dRNA-seq (Romero *et al.*, 2014) and originated from intergenic regions of the *E. coli* genome. Based on the results of this analysis, we selected the best 10 sRNA candidates (**Table 1**) for further validation by northern blotting. We found that 6 of them were expressed *in vivo* (**Figure 1**). The good rate of detection of selected sRNA candidates (**Table 1**) proves both the capability of the dRNA-seq approach used by Romero *et al.* (2014) to discover primary intergenic transcripts and the eficacy of the bioinformatic curation of sRNA candidates found by dRNA-seq. The dRNA-seq approach applied by Romero *et al.* (2014) is an adaptation of dRNA-seq procedure developed by Sharma *et al.* (2010). Both the modified dRNA-seq used by McDowall's team and the original method enabled the identification of new putative sRNAs in *Eschercihia coli*, *Streptomyces coelicolor* (Romero A. *et al.*, 2014), *Propionibacterium acnes* (Lin *et al.*, 2013) and *Helicobacter pylori* (Sharma *et al.*, 2010). The dRNA-seq procedure used by McDowall's team included two major modifications, such as use of flowcell reverse transcription (FRT-seq, (Mamanova and Turner, 2011)), to avoid the PCR caused bias during construction of cDNA libraries, and the use of tobacco acid pyrophosphatase (TAP) instead of Terminator^{TM} exonuclease (TEX), which removes 5′-monophosphorylated transcripts. The reason for the replacement of TEX was its poor efficiency on RNA of *Streptomyces coelicolor*, presumably due to the high prevalence of stable secondary structures (Romero A. *et al.*, 2014). The convenience of using TAP or TEX for dRNA-seq analysis is partly controversial given that the sensitivity of identification of TSS by dRNA-seq with TAP may be hindered by the de-pyrophosphorylation caused by RppH *in vivo*.

However, the use of FRT-seq for cDNA library preparation instead of the traditional reverse transcription followed by PCR can arguably be regarded as a valuable improvement (Mamanova *et al.*, 2010a).

Apart from Romero *et al.* (2014), other groups have also used RNA-seq approaches to analyse *E. coli* transcriptome and found a number of new small RNAs in *E. coli* (Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Conway *et al.*, 2014). However, the number and nature of known and putative sRNAs detected in these studies were different (**Fig. 3**). Comparison of their workflows (**Figure 2**) revealed a number of experimental features potentially accountable for the above differences. In particular, it seems likely that the low efficiency of sRNA detection (both known and putative, **Fig. 3A**) by Conway *et al.* (2014) was due to loss of short RNA molecules during the RNA purification procedure. Nevertheless, it is noteworthy that, unlike Shinhara *et al.* (2011), Raghavan *et al.* (2011) and Romero *et al.* (2014), Conway *et al.* (2014) primarily focused their work on mapping promoters and operons and largely neglected sRNAs, which seemed to be less important in the context of their study.

In contrast to Romero *et al.* (2014) and Conway *et al.* (2014), Shinhara *et al.* (2011) and Raghavan *et al.* (2011) reported the largest pools of new putative sRNAs. However, their RNA-seq approach did not allow distinguishing reads corresponding to 5'-monophosphorylated and 5'-triphosphorylated RNAs, i.e., they were not able to distinguish small primary transcripts and relatively stable decay intermediates of long transcripts. Although both studies could partly overcome this obstacle by using previously reported TSS data and *in silico* promoter prediction, it seems that not all promoters have consensus motifs that can be identified by computer algorithms (Tompa *et al.*, 2005). This obstacle, however, could be overcome by employing dRNA-seq, that in combination with flowcell reverse transcription prevents bias caused by conventional PCR based cDNA libraries construction, thereby ensuring a high rate of sRNA discovery (Romero *et al.*, 2014).

Despite a large number of experimental data suggesting the existence of many new sRNAs, *in vivo* expression of these sRNA candidates needs to be validated. It seems likely that detection of a putative sRNA by independent RNA-seq studies provides a good indication of its existence *in vivo*, and therefore these putative sRNAs could be good candidates for further validation by other methods (e.g., northern blot and RT-qPCR). To test this idea, we compiled a list of 296 putative sRNAs detected in four independent RNA-seq studies (**Table S7**), and found that 44 had been detected by

at least two different research teams (**Fig. 3C**). We bioinformatically analysed these putative transcripts, and selected the best 12 candidates for northern blot validation (**Table 3**), of which 9 yielded positive detection signals (**Fig. 4**). The high rates of successful detection demonstrates the efficiency of the tested approach, thus providing a strategy for effective sRNA discovery based on the comparison of results provided in independent RNA-seq experiments, and further *in silico* selection of the best candidates for experimental validation.

We believe that it is unlikely that the putative sRNAs validated in this work (**Fig. 1 and 4**) are metastable decay intermediates or artefacts of the RNA sequencing approach, because they are associated with TSSs and were detected by northern blotting. Moreover, we also examined their abundance in cells grown in rich and minimal media by northern blotting, and found that some of them were differentially expressed. These findings indicate that their expression is affected by the composition of the growth media used, and suggest their possible involvement in regulation of metabolic pathways respectively regulated in cells grown in these media. Validation of the differential expression of these putative sRNAs by using quantitative methods such as RT-qPCR in the future could provide further support of these observations.

In summary, analysis of RNA-seq approaches and their capacities to discover sRNAs demonstrated that combination of *in silico* selection of promising sRNA candidates represents an efficient strategy for putative sRNA search and discovery. The benefits of this approach are particularly clear, given that hundreds of putative sRNAs have been reported so far by several groups, and only a small fraction of them have been further verified and characterized (**Tables S6** and **S7**).

Bibliography

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.

Altuvia, S.; Zhang, a; Argaman, L.; Tiwari, a; Storz, G. The *Escherichia coli* OxyS Regulatory RNA Represses fhlA Translation by Blocking Ribosome Binding. *EMBO J.* **1998**, *17* (20), 6069–6075.

Andersen, J.; Delihas, N.; Kazuhiro, I.; Green, P. J.; Pines, O.; Ilercil, O.; Inouye, M. The Isolation and Characterization of RNA Coded by the micF Gene in *Escherichia coli*. *Nucleic Acids Res.* **1987**, *15* (5), 2089–2101.

Antal, M.; Bordeau, V.; Douchin, V.; Felden, B. A Small Bacterial RNA Regulates a Putative ABC Transporter. *J. Biol. Chem.* **2005**, *280* (9), 7901–7908.

Argaman, L.; Hershberg, R.; Vogel, J.; Bejerano, G.; Wagner, E. G.; Margalit, H.; Altuvia, S. Novel Small RNA-Encoding Genes in the Intergenic Regions of *Escherichia coli*. *Curr. Biol.* **2001**, *11* (12), 941–950.

Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K. A.; Tomita, M.; Wanner, B. L.; Mori, H. Construction of *Escherichia coli* K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection. *Mol. Syst. Biol.* **2006**, *2:2006.0008*.

Bak, G.; Lee, J. Y. J.; Suk, S.; Kim, D.; Lee, J. Y. J.; Kim, K.; Choi, B.-S.; Lee, Y.; Young Lee, J.; Kim, K.; *et al.* Identification of Novel sRNAs Involved in Biofilm Formation, Motility, and Fimbriae Formation in *Escherichia coli*. *Sci. Rep.* **2015**, *5* (October), 15287.

Barrett, T.; Wilhite, S. E.; Ledoux, P.; Evangelista, C.; Kim, I. F.; Tomashevsky, M.; Marshall, K. A.; Phillippy, K. H.; Sherman, P. M.; Holko, M.; *et al.* NCBI GEO: Archive for Functional Genomics Data Sets - Update. *Nucleic Acids Res.* **2013**, *41*, D991–D995.

Chan, P. P.; Holmes, A. D.; Smith, A. M.; Tran, D.; Lowe, T. M. The UCSC Archaeal Genome Browser: 2012 Update. *Nucleic Acids Res.* **2012**, *40*, D646–D652.

Chen, S.; Lesnik, E. a; Hall, T. a; Sampath, R.; Griffey, R. H.; Ecker, D. J.; Blyn, L. B. A

Bioinformatics Based Approach to Discover Small RNA Genes in the *Escherichia coli* Genome. *Biosystems.* **2002**, *65* (2–3), 157–177.

Chen, S.; Zhang, A.; Blyn, L. B.; Storz, G. MicC , a Second Small-RNA Regulator of Omp Protein Expression in *Escherichia coli*. **2004**, *186* (20), 6689–6697.

Conway, T.; Creecy, J. P.; Maddox, S. M.; Grissom, J. E.; Conkle, T. L.; Shadid, T. M.; Teramoto, J. Unprecedented High-Resolution View of Bacterial Operon Architecture Revealed by RNA Sequencing. *MBio* **2014**, *5* (4), :e01442-14.

Fozo, E. M.; Kawano, M.; Fontaine, F.; Kaya, Y.; Mendieta, K. S.; Jones, K. L.; Ocampo, A.; Rudd, K. E.; Storz, G. Repression of Small Toxic Protein Synthesis by the Sib and OhsC Small RNAs. *Mol Microbiol.* **2008a**, *70* (5), 1076–1093.

Fozo, E. M.; Hemm, M. R.; Storz, G. Small Toxic Proteins and the Antisense RNAs That Repress Them. *Microbiol Mol Biol Rev* **2008b**, *72* (4), 579–89, Table of Contents.

Gerstle, K.; Klätschke, K.; Hahn, U.; Piganeau, N. The Small RNA RybA Regulates Key-Genes in the Biosynthesis of Aromatic Amino Acids under Peroxide Stress in E. Coli. *RNA Biol.* **2012**, *9* (4), 458–468.

Hershberg, R.; Altuvia, S.; Margalit, H. A Survey of Small RNA-Encoding Genes in *Escherichia coli*. *Nucleic Acids Res.* **2003**, *31* (7), 1813–1820.

Higgins, C. F.; Mclaren, R. S.; Newbury, S. F. Repetitive Extragenic Palindromic (REP) Sequences Are Highly Conserved Inverted Repeats Present in. *Gene* **1988**, *72*, 3–14.

Hirsch, C. D.; Springer, N. M.; Hirsch, C. N. Genomic Limitations to RNAseq Expression Profiling. *Plant J.* **2015**, *84*, 491–503.

Hoe, C.-H.; Raabe, C. a; Rozhdestvensky, T. S.; Tang, T.-H. Bacterial sRNAs: Regulation in Stress. *Int. J. Med. Microbiol.* **2013**, *303* (5), 217–229.

Ikemura, T.; Dahlberg, J. E.; Ikemura T and Dahlberg JE; Ikemura, T.; Dahlberg, J. E. Small Ribonucleic Acids of *Escherichia coli* characterization by polyacrylamide gel electrophoresis and fingerprint analysis. *JBC* **1973**, 248(14):5024-32..

Joyce, C. M.; Grindley, N. D. Identification of Two Genes Immediately Downstream from the polA Gene of *Escherichia coli*. *J. Bacteriol.* **1982**, *152* (3), 1211–1219.

Kang, S.-M.; Choi, J.-W.; Lee, Y.; Hong, S.-H.; Lee, H.-J. Identification of microRNA-Size, Small RNAs in *Escherichia coli*. *Curr. Microbiol.* **2013**, 67(5):609-13.

Kawano, M.; Oshima, T.; Kasai, H.; Mori, H. Molecular Characterization of Long Direct Repeat (LDR) Sequences Expressing a Stable mRNA Encoding for a 35-Amino-Acid Cell-Killing Peptide and a Cis-Encoded Small Antisense RNA in *Escherichia coli*. *Mol. Microbiol.* **2002**, *45* (2), 333–349.

Kawano, M.; Reynolds, A. a; Miranda-Rios, J.; Storz, G. Detection of 5'- and 3'-UTR-Derived Small RNAs and Cis-Encoded Antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.* **2005**, *33* (3), 1040–1050.

Keseler, I. M.; Mackie, A.; Peralta-Gil, M.; Santos-Zavaleta, A.; Gama-Castro, S.; Bonavides-Martínez, C.; Fulcher, C.; Huerta, A. M.; Kothari, A.; Krummenacker, M.; *et al.* EcoCyc: Fusing Model Organism Databases with Systems Biology. *Nucleic Acids Res.* **2013**, *41* (Database issue), D605-12.

Kime, L.; Jourdan, S. S.; McDowall, K. J. *Chapter 12 Identifying and Characterizing Substrates of the RNase E/G Family of Enzymes*, 1st ed.; Elsevier Inc., **2008**; Vol. 447.

Lin, Y.; A, D. R.; Guan, S.; Mamanova, L.; McDowall, K. J. A Combination of Improved Differential and Global RNA-Seq Reveals Pervasive Transcription Initiation and Events in All Stages of the Life-Cycle of Functional RNAs in Propionibacterium Acnes, a Major Contributor to Wide-Spread Human Disease. *BMC Genomics* **2013**, *14* (1), 620.

Liu, M. Y.; Gui, G.; Wei, B.; Iii, J. F. P.; Oakford, L.; Giedroc, D. P.; Romeo, T. The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagonizes Its Activity in *Escherichia coli* The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagoniz. *J. Biol. Chem.* **1997**, *272* (28), 17502–17510.

Majdalani, N.; Chen, S.; Murrow, J.; St John, K.; Gottesman, S. Regulation of RpoS by a Novel Small RNA: The Characterization of RprA. *Mol. Microbiol.* **2001**, *39* (5), 1382–1394.

Mamanova, L.; Turner, D. J. Low-Bias, Strand-Specific Transcriptome Illumina Sequencing by on-Flowcell Reverse Transcription (FRT-Seq). *Nat. Protoc.* **2011**, *6* (11), 1736–1747.

Mamanova, L.; Andrews, R. M.; James, K. D.; Sheridan, E. M.; Ellis, P. D.; Langford, C. F.; Ost, T. W. B.; Collins, J. E.; Turner, D. J. FRT-Seq: Amplification- Free, Strand-Specific Transcriptome Sequencing. *Nat. Methods* **2010a**, *7* (2), 130–132.

Mamanova, L.; Andrews, R. M.; James, K. D.; Sheridan, E. M.; Peter, D.; Langford, C.

F.; Ost, T. W. B.; Collins, J. E.; Turner, D. J. FRT-Seq: Amplification-Free, Strand-Specific, Transcriptome Sequencing. *Nat Methods* **2010b**, *7* (2), 130–132.

Massé, E.; Gottesman, S. A Small RNA Regulates the Expression of Genes Involved in Iron Metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (7), 4620–4625.

Melamed, S.; Peer, A.; Faigenbaum-Romm, R.; Gatt, Y. E.; Reiss, N.; Bar, A.; Altuvia, Y.; Argaman, L.; Margalit, H. Global Mapping of Small RNA-Target Interactions in Resource Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol. Cell* **2016**, *63* (5), 884–897.

Pedersen, K.; Gerdes, K. Multiple Hok Genes on the Chromosome of *Escherichia coli*. *Mol. Microbiol.* **1999**, *32* (5), 1090–1102.

Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI Reference Sequences (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins. *Nucleic Acids Res.* **2007**, *35* (SUPPL. 1), 61–65.

Raghavan, R.; Groisman, E. A.; Ochman, H. Genome-Wide Detection of Novel Regulatory RNAs in E . Coli. *Genome Res.* **2011**, *21*, 1487–1497.

Rivas, E.; Klein, R. J.; Jones, T. a; Eddy, S. R. Computational Identification of Noncoding RNAs in E. Coli by Comparative Genomics. *Curr. Biol.* **2001**, *11* (17), 1369–1373.

Romero A., D.; Hasan, A. H.; Lin, Y.-F. fei; Kime, L.; Ruiz-Larrabeiti, O.; Urem, M.; Bucca, G.; Mamanova, L.; Laing, E. E.; van Wezel, G. P.; *et al.* A Comparison of Key Aspects of Gene Regulation in Streptomyces Coelicolor and *Escherichia coli* Using Nucleotide-Resolution Transcription Maps Produced in Parallel by Global and Differential RNA Sequencing. *Mol. Microbiol.* **2014**, *94* (5), 963–987.

Rudd, K. E. Linkage Map of *Escherichia coli* K-12, Edition 10: The Physical Map. *Microbiol. Mol. Biol. Rev.* **1998**, *62* (3), 985–1019.

Sætrom, P.; Sneve, R.; Kristiansen, K. I.; Snøve Jr., O.; Grünfeld, T.; Rognes, T.; Seeberg, E. Predicting Non-Coding RNA Genes in *Escherichia coli* with Boosted Genetic Programming. *Nucleic Acids Res.* **2005**, *33* (10), 3263–3270.

Sharma, C. M.; Vogel, J. Differential RNA-Seq: The Approach behind and the Biological Insight Gained. *Curr. Opin. Microbiol.* **2014**, *19* (1), 97–105.

Sharma, C. M.; Hoffmann, S.; Darfeuille, F.; Reignier, J.; Findeiss, S.; Sittka, A.; Chabas, S.; Reiche, K.; Hackermüller, J.; Reinhardt, R.; *et al.* The Primary Transcriptome of the Major Human Pathogen Helicobacter Pylori. *Nature* **2010**, *464* (7286), 250–255.

Shinhara, A.; Matsui, M.; Hiraoka, K.; Nomura, W.; Hirano, R.; Nakahigashi, K.; Tomita, M.; Mori, H.; Kanai, A.; Kana, A. Deep Sequencing Reveals as-yet-Undiscovered Small RNAs in *Escherichia coli*. *BMC Genomics* **2011**, *12* (1), 428.

Sledjeski, D.; Gottesman, S. A Small RNA Acts as an Antisilencer of the H-NS-Silenced rcsA Gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92* (6), 2003–2007.

Tjaden, B.; Saxena, R. M.; Stolyar, S.; Haynor, D. R.; Kolker, E.; Rosenow, C. Transcriptome Analysis of *Escherichia coli* Using High-Density Oligonucleotide Probe Arrays. *Nucleic Acids Res.* **2002**, *30* (17), 3732–3738.

Tobes, R.; Pareja, E. Bacterial Repetitive Extragenic Palindromic Sequences Are DNA Targets for Insertion Sequence Elements. *BMC Genomics* **2006**, *7*, 62.

Tompa, M.; Li, N.; Bailey, T.; Church, G. M.; De Moor, B.; Eskin, E.; Favorov, A.; Frith, M. C.; Fu, Y.; Kent, W.; *et al.* Assessing Computational Tools for the Discovery of Transcription Factor Binding Sites. *Nat. Biotechnol.* **2005**, *23* (1), 137–144.

Urban, J. H.; Vogel, J. Translational Control and Target Recognition by *Escherichia coli* Small RNAs in Vivo. *Nucleic Acids Res.* **2007**, *35* (3), 1018–1037.

Vanderpool, C. K.; Gottesman, S. Involvement of a Novel Transcriptional Activator and Small RNA in Post-Transcriptional Regulation of the Glucose Phosphoenolpyruvate Phosphotransferase System. *Mol. Microbiol.* **2004**, *54* (4), 1076–1089.

Vogel, J.; Bartels, V.; Tang, T. H.; Churakov, G.; Slagter-Jäger, J. G.; Hüttenhofer, A.; Wagner, E. G. H. RNomics in *Escherichia coli* Detects New sRNA Species and Indicates Parallel Transcriptional Output in Bacteria. *Nucleic Acids Res.* **2003**, *31* (22), 6435–6443.

Wagner, E. G. H.; Romby, P. *Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It*; Elsevier Ltd, 2015; Vol. 90.

Wang, Z.; Gerstein, M.; Snyder, M. RNA-Seq: A Revolutionary Tool for Transcriptomics. *Nat. Rev. Genet.* **2009**, *10* (1), 57–63.

Wassarman, K. M.; Zhang, A.; Storz, G. Small RNAs in *Escherichia coli*. *Trends Microbiol.* **1999**, *7* (1), 37–45.

Wassarman, K. M.; Repoila, F.; Rosenow, C.; Storz, G.; Gottesman, S. Identification of Novel Small RNAs Using Comparative Genomics and Microarrays. *GENES Dev.* **2001**, *15*, 1637–1651.

Waters, L. S.; Sandoval, M.; Storz, G. The *Escherichia coli* MntR Miniregulon Includes Genes Encoding a Small Protein and an Efflux Pump Required for Manganese Homeostasis. *J. Bacteriol.* **2011**, *193* (21), 5887–5897.

Weel-Sneve, R.; Kristiansen, K. I.; Odsbu, I.; Dalhus, B.; Booth, J.; Rognes, T.; Skarstad, K.; Bjørås, M. Single Transmembrane Peptide DinQ Modulates Membrane-Dependent Activities. *PLoS Genet.* **2013**, *9* (2), e1003260.

Weilbacher, T.; Suzuki, K.; Dubey, A. K.; Wang, X.; Gudapaty, S.; Morozov, I.; Baker, C. S.; Georgellis, D.; Babitzke, P.; Romero A, D. A Novel sRNA Component of the Carbon Storage Regulatory System of *Escherichia coli*. *Mol. Microbiol.* **2003**, *48* (3), 657–670.

Weirick, T.; Militello, G.; Müller, R.; John, D.; Dimmeler, S.; Uchida, S. The Identification and Characterization of Novel Transcripts from RNA-Seq Data. *Brief. Bioinform.* **2015**, No. May, 1–8.

Yachie, N.; Numata, K.; Saito, R.; Kanai, A.; Tomita, M. Prediction of Non-Coding and Antisense RNA Genes in *Escherichia coli* with Gapped Markov Model. *Gene* **2006**, *372*, 171–181.

Zhang, A.; Wassarman, K. M.; Rosenow, C.; Tjaden, B.; Storz, G.; Gottesman, S. Global Analysis of Small RNA and mRNA Targets of Hfq. *Mol. Microbiol.* **2003**, *50* (4), 1111–1124.

Zuker, M. Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* **2003**, *31* (13), 3406–3415.

Summary of results

Previous studies revealed a class of small antisense RNAs (sRNAs) that play important roles in regulatory mechanisms that enable *Escherichia coli* and other bacteria (including essential genera of pathogenic enterobacteria) to adapt and survive to environmental stresses. Analysis of several stress-related sRNAs demonstrated that they can exert their regulatory functions by base-pairing with their target mRNAs to alter their translation and stability, often in an Hfq-dependent manner. Despite significant progress in discovery of new sRNAs and defining their biological functions, still a large fraction of experimentally identified and *in silico* predicted sRNAs has not been validated and characterized in detail. The above consideration defined the main goals of this study, focused on (i) utilization of high-throughput custom microarrays for validation and analysis of sRNAs and (ii) screening and validation of putative sRNAs revealed by RNA sequencing in *E. coli*.

To address the first objective, we designed and tested a new custom microarray for high-throughput detection of sRNAs in *E. coli*. We selected microarray analysis as the method of choice because of its robustness, affordable cost and statistical reliability, constituting an excellent and extensively used gene expression profiling technique with a similar analytical power as the more recently developed next-generation sequencing approaches (Malone and Oliver, 2011; Raghavachari *et al.*, 2012). The efficiency and reliability of our custom microarray was tested by comparing transcriptome profiles of *E. coli* cells grown under different conditions of nutrient availability (i.e. growth in Luria-Bertani rich media versus M9 minimal media supplemented with glucose), as well as in cells with different genetic backgrounds (i.e. in a mutant strain lacking poly(A)polymerase (PAP I), known for its role in regulation of RNA stability, *versus* its isogenic wild-type strain) (**Chapter 1**).

Once the efficiency of our custom microarray was established, we further used it to identify sRNAs associated with two opposite pathways, glycolysis and gluconeogenesis, which play the central role in *E. coli* carbon metabolism. Gene expression analysis experiments were carried out on *E. coli* cells cultured in M9 minimal media supplemented with either pyruvate or glucose as only carbon sources (M9/pyruvate and M9/glucose, respectively). This analysis revealed the differential expression of a number of sRNAs (namely, GcvB, RprA, RyhB, GlmY, GlmZ and CyaR) with known roles in central carbon metabolism and overlapping pathways (reviewed by Richards and Vanderpool, (2011) and Michaux *et al.* (2014)). Moreover,

the differential regulation of some putative sRNAs including RyeA, a previously identified sRNA with unknown regulatory roles (Wassarman *et al.*, 2001), was demonstrated for the first time, suggesting its potential role in regulation of central carbon metabolism in *E. coli* (**Chapter 2**).

Finally, the last part of our study was focused on discovery of sRNAs by RNA sequencing, and their experimental validation. This part included the results of our joint project with Dr. Kenneth McDowall (University of Leeds). During this collaboration we analysed a list of intergenic transcripts found in *E. coli* by using differential RNA sequencing (dRNA-seq) and selected the best sRNA candidates to be tested by northern blot analysis. Apart from being involved in the above collaboration, we also compared the results of different RNA-seq studies dealing with the analysis of the *E. coli* transcriptome and analysed the factors affecting their capacity to discover new sRNAs. During this analysis, we compared the number and nature of known sRNAs detected in four independent studies carried out in *E. coli* (Raghavan *et al.*, 2011; Shinhara *et al.*, 2011; Conway *et al.*, 2014; Romero A. *et al.*, 2014). We found that most of the known sRNAs (i.e., annotated in EcoCyc database) detected by the four independent RNA-seq studies were detected by at least two of them (**Chapter 3**). We therefore hypothesized that detection of a putative sRNA by independent RNA-seq approaches might be a good indicative of its actual existence. To test this idea, we compiled a list of 44 sRNA candidates detected by more than one study. We analysed them *in silico*, and selected the best candidates for further validation, subsequently obtaining evidence of *in vivo* expression of 9 of them (**Chapter 3**). Moreover, we found that the abundance of several detected sRNAs was differentially affected in cells grown in rich (LB) and minimal media (M9/glucose), thus suggesting their possible involvement in regulation of central carbon metabolism or associated pathways (**Chapter 3**).

Bibliography

Conway, T.; Creecy, J. P.; Maddox, S. M.; Grissom, J. E.; Conkle, T. L.; Shadid, T. M.; Teramoto, J. Unprecedented High-Resolution View of Bacterial Operon Architecture Revealed by RNA Sequencing. *MBio* **2014**, *5* (4), :e01442-14.

Malone, J. H.; Oliver, B. Microarrays, Deep Sequencing and the True Measure of the Transcriptome. *BMC Biol.* **2011**, *9* (1), 34.

Michaux, C.; Verneuil, N.; Hartke, A.; Giard, J. Physiological Roles of Small RNA Molecules. *Microbiology* **2014**, No. 33, 1–33.

Raghavachari, N.; Barb, J.; Yang, Y.; Liu, P.; Woodhouse, K.; Levy, D.; O'Donnell, C. J.; Munson, P. J.; Kato, G. J. A Systematic Comparison and Evaluation of High Density Exon Arrays and RNA-Seq Technology Used to Unravel the Peripheral Blood Transcriptome of Sickle Cell Disease. *BMC Med. Genomics* **2012**, *5* (1), 28.

Raghavan, R.; Groisman, E. A.; Ochman, H. Genome-Wide Detection of Novel Regulatory RNAs in E . Coli. *Genome Res.* **2011**, *21*, 1487–1497.

Richards, G. R.; Vanderpool, C. K. Molecular Call and Response: The Physiology of Bacterial Small RNAs. *Biochim. Biophys. Acta* **2011**, *1809* (10), 525–531.

Romero A., D.; Hasan, A. H.; Lin, Y.-F. fei; Kime, L.; Ruiz-Larrabeiti, O.; Urem, M.; Bucca, G.; Mamanova, L.; Laing, E. E.; van Wezel, G. P.; *et al.* A Comparison of Key Aspects of Gene Regulation in Streptomyces Coelicolor and *Escherichia coli* Using Nucleotide-Resolution Transcription Maps Produced in Parallel by Global and Differential RNA Sequencing. *Mol. Microbiol.* **2014**, *94* (5), 963–987.

Shinhara, A.; Matsui, M.; Hiraoka, K.; Nomura, W.; Hirano, R.; Nakahigashi, K.; Tomita, M.; Mori, H.; Kanai, A.; Kana, A. Deep Sequencing Reveals as-yet-Undiscovered Small RNAs in *Escherichia coli*. *BMC Genomics* **2011**, *12* (1), 428.

Wassarman, K. M.; Repoila, F.; Rosenow, C.; Storz, G.; Gottesman, S. Identification of Novel Small RNAs Using Comparative Genomics and Microarrays. *GENES Dev.* **2001**, *15*, 1637–1651.

Conclusions

1- Our custom microarray is an efficient tool for detection of sRNAs and gene expression profiling of *E. coli* sRNAs and their target mRNAs at the whole transcriptome level.

2- The use of our custom microarrays for gene expression analysis of *E. coli* grown in LB and M9/glucose media demonstrated that only two known sRNAs (GcvB and RydB) showed considerable differences in their expression level.

3- Comparison of the *E. coli* wild-type strain and its isogenic *pcnB*$^{-}$ mutant revealed five sRNAs (GlmY, GlmZ, RyjA, RybB and SroH) differentially expressed in these strains. Moreover, the data obtained for RyhB and SroH were novel and therefore reveal new examples of sRNAs whose level is controlled by poly(A)polymerase I *in vivo*.

4- The use of the custom microarray for pre-screening previously reported putative sRNAs detected the presence of G0-10699 and IS001, and validation by northern blot analysis confirmed their expression *in vivo*, thereby expanding the number of *E. coli* sRNAs verified experimentally.

5- Gene expression analysis carried out with custom microarrays on *E. coli* cells cultured in M9/pyruvate and M9/glucose revealed the differential expression of a number of sRNAs (namely, GcvB, RprA, RyhB, GlmY, GlmZ and CyaR) with known roles in central carbon metabolism and overlapping pathways. Moreover, the differential regulation of some putative sRNAs including RyeA, a previously identified sRNA with unknown regulatory roles, was demonstrated for the first time, thus expanding the number of sRNAs potentially involved in the regulation of central carbon metabolism in *E. coli*.

6- Our collaboration with the group of Dr. Kenneth McDowall demonstrated that combination of differential RNA sequencing and subsequent *in silico* curation of the sRNA candidate list results in a good rate of discovery of *bona fide* putative sRNAs.

7- *In silico* pre-selection of sRNA candidates independently detected in different RNA sequencing studies ensures the high rate of putative sRNA discovery too.

8- Some new putative sRNAs (p-sRNA-1, 2, 3, 4, 5 and 6) as well as previously reported (but not experimentally verified) putative sRNAs (IS019, G0-10698, p-sRNA-11, ECS054, C0614, p-sRNA-12, ECS163 and G0-10701) were found to be expressed *in vivo*.

9- Some new putative sRNAs (p-sRNA-1, 2, 3, 4, 5 and 6) as well as ECS054 putative sRNA were differentially expressed in *E. coli* cells grown in LB *versus* M9/glucose, thus suggesting their possible role in the regulation of central carbon metabolism.

Resumen y Conclusiones

Descubrimiento y estudio de ARNs pequeños (small RNAs) en *Escherichia coli* utilizando microarrays de diseño propio y secuenciación de última generación

sRNAs: Características y funciones

La ubicuidad y extensa presencia de bacterias en amplia diversidad de hábitats se debe en gran medida a su capacidad de adaptarse y sobrevivir bajo condiciones hostiles y continuamente cambiantes. La eficiencia y sincronía de las repuestas bacterianas a multitud de estímulos implican una gran variedad de complejos mecanismos de control de la expresión génica a nivel transcripcional y post-transcripcional.

Entre los factores celulares que controlan el estado fisiológico y la adaptación bacteriana, las moléculas de ARN anti-sentido tienen funciones clave en la regulación post-transcripcional de la expresión génica (revisado por Kaberdin and Bläsi (2006) y Storz *et al.* (2011)). En particular, las moléculas de ARN pequeñas (small RNAs o sRNAs) tienen un papel esencial en la regulación de la respuesta y adaptación a muchas condiciones de estrés en bacterias, y controlan la expresión génica por medio de mecanismos anti-sentido, es decir, hibridando con ARN mensajeros (mRNAs) diana y formando un duplex sRNA-mRNA que puede promover o impedir la degradación dependiente de RNasas (enzimas especificas implicadas en la degradación, procesamiento y maduración del ARN) del mRNA diana y/o reprimir o inducir su traducción (revisado por Kaberdin and Bläsi (2006)). Existen dos grandes sub-tipos de sRNAs, los llamados *cis*-codificados y los *trans*-codificados, siendo una de las principales diferencias entre ambos que los sRNAs *cis*-codificados se transcriben del mismo *locus* genómico que los mRNAs diana, mientras que los sRNAs *trans*-codificados y sus mRNAs dianas se transcriben de *locus* distintos. Por lo tanto, los sRNAs codificados en *cis* respecto a sus dianas comparten con estas una complementariedad perfecta, en tanto que los *trans*-codificados y sus mRNAs diana comparten una complementariedad imperfecta y parcial, y la interacción entre ambos suele estar mediada por la chaperona de ARN Hfq (revisado por Vogel and Luisi (2011) y Sobrero and Valverde (2012)). A menudo, la transcripción de sRNAs *trans*-

codificados es inducida en respuesta a señales ambientales específicas cuya respuesta regulan, haciendo de estos sRNAs importantes reguladores durante las respuestas bacterianas a diversas condiciones de estrés, tales como la escasez de hierro o nutrientes, daño en la envoltura celular, temperatura o pH sub-óptimos, o estrés oxidativo, entre otras (revisado por Hoe *et al.* (2013)).

La regulación de la expresión génica ejercida por sRNAs cuenta con varias características ventajosas, como la capacidad de un solo sRNA *trans*-codificado para regular múltiples mRNA dianas y de varios sRNAs de regular una misma diana (por ejemplo, cada uno de los sRNAs DsrA, RyhB, RybB y GcvB, regula 5, 9 11 y 21 mRNAs diana, respectivamente (Li *et al.*, 2013), mientras que los sRNAs DsrA, RprA Y ArcZ, entre otros, controlan la traducción de mRNA *rpoS*, que codifica σ^S, un regulador clave de la respuesta general al estrés en *E. coli* (revisado por Wagner and Romby (2015)). Esto permite tanto una respuesta amplia y sincronizada, como la matización y el ajuste preciso de la expresión de algunos genes, lo que garantiza la alta eficiencia y especificidad de las respuestas al estrés bacteriano. Además, algunos sRNAs regulan y son regulados por factores de transcripción, siendo ambos partes integrales de amplias redes de regulación. Por lo tanto, la acción concertada de sRNAs y factores de transcripción puede controlar con precisión la expresión de mRNAs diana tanto a nivel transcripcional como post-transcripcional, y la propia expresión de los agentes reguladores, proporcionando una respuesta equilibrada y precisa (revisado por Beisel and Storz (2010) y Wagner and Romby (2015)). Además, la vida media comúnmente corta de los sRNAs (Vogel *et al.*, 2003) relacionada con su codegradación junto con sus mRNAs diana permite una rápida inactivación del regulador una vez que el estrés ha terminado (revisado por Kaberdin and Bläsi (2006) y Saramago *et al.* (2014)). Finalmente, como el gasto energético requerido para la síntesis del ARN es menor que el requerido para la síntesis de un factor de transcripción o un represor de la traducción, la regulación ejercida por sRNAs es rápida, precisa y posiblemente más económica.

Descubrimiento y estado actual del conocimiento acerca de los sRNAs

Los primeros sRNA identificados se descubrieron a principios de los años 70, y durante las dos décadas siguientes apenas una decena de sRNAs fueron identificados, incluyendo los primeros sRNA anti-sentido *trans*-codificados (revisado por Wassarman *et al.* (1999)). El número gradualmente creciente de sRNAs *trans*-

codificados caracterizados permitió definir sus características comunes (tales como su longitud media de 50 a 300 nucleótidos (Hershberg *et al.*, 2003), que se transcriben de regiones intergénicas (IGR), que en algunos casos están bien conservados entre especies estrechamente relacionadas (Wassarman *et al.*, 2001), y que están asociados a un promotor y un terminador ρ-independiente) (revisado por (Updegrove *et al.*, 2015)). Estas características se emplearon en el desarrollo de criterios generales utilizados para la predicción y análisis de sRNAs. Algunos estudios dedicados a la búsqueda de sRNAs en *E. coli* de acuerdo con los criterios descritos incluyen estudios teóricos y experimentales (algunos ejemplos incluyen Argaman *et al.* (2001), Rivas *et al.* (2001), Wassarman *et al.* (2001), Chen *et al.* (2002), Tjaden *et al.* (2002), Vogel *et al.* (2003), Zhang *et al.* (2003), Kawano *et al.* (2005) y Yachie *et al.* (2006)), así como secuenciación de última generación (*next generation sequencing*) (tales como Raghavan *et al.*, (2011), Shinhara *et al.*, (2011), Conway *et al.*, (2014), Romero A. *et al.*, (2014), Thomason *et al.*, (2015) y Weirick *et al.*, (2015)). Estos estudios y otros ampliaron enormemente el panorama de sRNAs identificados en *E. coli*, y actualmente alrededor de un centenar de sRNAs han sido anotados en bases de datos (por ejemplo, en la base de datos BSRD se cuentan 108 sRNAs como anotados en su última actualización en Octubre de 2015 (Li *et al.* (2013)).

Aunque algunos sRNAs anotados están bien caracterizados, las funciones biológicas y mRNA dianas de muchos otros sRNAs aún se desconocen. Por otra parte, dado el gran número de sRNA putativos predichos *in silico* y por medio de técnicas de alto rendimiento, uno de los desafíos actuales es desarrollar herramientas eficientes y robustas para distinguir las señales potencialmente pertenecientes a sRNAs putativos del ruido transcripcional (Lindgreen *et al.*, 2014), con el fin de seleccionar los candidatos de sRNAs de más prometedores para su posterior validación y análisis.

Dado que validar experimentalmente y analizar las posibles funciones de los sRNAs puede ser difícil y suponer mucho tiempo, desarrollar y mejorar nuevos procedimientos de alto rendimiento que permitan la validación y el estudio de sRNAs con el fin de caracterizarlos y revelar sus funciones biológicas es deseable y necesario para profundizar en el conocimiento de las redes reguladoras bacterianas. En la actualidad se están desarrollando procedimientos en esta dirección, como el reciente trabajo de Melamed *et al.* (2016) en el que desarrollaron un método denominado RIL-seq, con el objetivo de identificar sRNAs y sus mRNA diana a través de su interacción con la chaperona Hfq, o el estudio de Lindgreen *et al.* (2014), en el que estudian cómo

establecer una ventana filogenética para el uso efectivo del análisis comparativo con el fin de encontrar nuevos sRNAs verdaderos entre el ruido transcriptómico.

Objetivos de nuestro estudio

El papel regulador de los sRNAs caracterizados y el desconocimiento acerca de las funciones de los muchos sRNAs putativos anotados y predichos determinaron los dos objetivos generales del presente estudio, centrados respectivamente en (i) la optimización de métodos de alto rendimiento para la validación y análisis de sRNAs y (ii) la utilización de secuenciación de última generación para el descubrimiento de nuevos sRNAs en *E. coli*.

Diseño y prueba de un microarray optimizado para la detección de sRNAs en *Escherichia coli*

Con respecto al primer objetivo, diseñamos y probamos un nuevo microarray optimizado para la detección simultánea de sRNAs y mRNAs en *E. coli*. Seleccionamos el microarray como el método de análisis debido a su robustez, coste asequible y fiabilidad estadística, constituyendo una excelente técnica ampliamente utilizada para el perfilamiento de la expresión génica y que tiene una capacidad analítica similar a la de técnicas más recientemente desarrolladas, como la secuenciación de última generación (Malone and Oliver, 2011; Raghavachari *et al*., 2012).

El diseño de nuestro microarray se basó en el conjunto estándar de sondas de Agilent Technologies para el transcriptoma de *E. coli,* más un nuevo conjunto de sondas para la detección de sRNAs (anotados y predichos), con el fin de detectar simultáneamente el perfil de expresión de los sRNAs y los mRNAs. La diferencia principal entre nuestro diseño y otros microarrays comerciales de alta densidad consiste en el diseño específico de las sondas para la hibridación con transcritos originados en regiones intergénicas, optimizadas en relación a las características de sus secuencia. Los microarrays comerciales tienen sondas que cubren todas las regiones intergénicas del genoma de *E. coli*, pero al no estar diseñadas de manera específica no pueden garantizar la correcta hibridación.

Para probar la eficacia y fiabilidad de nuestro microarray de diseño propio, analizamos y comparamos el transcriptoma de células de *E. coli* cultivadas en medio nutritivo rico (Luria-Bertani, LB) y medio mínimo (M9 suplementado con glucosa, M9/ glucosa), lo que reveló la expresión diferencial de una serie de genes implicados en el metabolismo central de carbono, de forma consistente con lo esperado y previamente descrito (Tao *et al.*, 1999), validando por tanto el correcto funcionamiento del microarray para el perfilamiento transcriptómico general de *E. coli*. Respecto a los sRNAs incluidos en el diseño, el nivel de expresión de la mayoría de los sRNAs anotados detectados por el microarray no se vio afectado por el crecimiento en diferentes medios de cultivo, y sólo dos sRNAs anotados (GcvB y RydB) mostraron una regulación diferencial. Estos resultados fueron consistentes con las diferencias detectadas en el nivel de expresión de algunos mRNAs diana de GcvB, y la participación de estos genes en vías metabólicas reguladas de distinta forma durante el crecimiento en los dos medios de cultivo (tales como la resistencia al estrés ácido y la captación de aminoácidos) era también coherente con el perfil transcriptómico general y las necesidades metabólicas de las células cultivadas en medio rico y medio mínimo. Por otra parte, el mecanismo de acción y el rol específico de RydB en el metabolismo bacteriano aún no han sido descritos, y el hallazgo de la regulación de su expresión en células cultivadas en LB comparado con M9/glucosa podría sugerir su implicación en el control de las vías metabólicas que *E. coli* utiliza de forma diferente durante el crecimiento en medio rico y medio mínimo.

Con el fin de corroborar la capacidad de nuestro microarray para detectar la expresión diferencial de sRNAs, comparamos la expresión génica de una cepa mutante de *E. coli* carente de poli(A)polimerasa I (PAP I) (una enzima que afecta a la estabilidad del ARN (Régnier and Hajnsdorf, 2009) con la de su cepa silvestre (*wild-type*). Este análisis reveló la expresión diferencial de varios sRNAs anotados, entre ellos RybB y SroH, para los cuales el efecto de PAP I en su estabilidad no había sido descrito (Keseler *et al.*, 2013).

Además de los hallazgos sobre la regulación de los sRNAs anotados, la existencia de dos de los sRNAs putativos detectados por el microarray (G0-10699 e IS001 (Chen *et al.* (2002) y Raghavan *et al.* (2011)) se confirmaron por medio de análisis de northern blot.

Análisis de la expresión de sRNAs implicados en las vías de la glucolisis y gluconeogénesis utilizando el microarray de diseño propio

Tras comprobar la capacidad del microarray para el perfilamiento transcriptómico y la detección y medida del nivel de expresión de los sRNAs, lo empleamos para identificar sRNAs potencialmente implicados en la regulación de la glucólisis y la gluconeogénesis, dos vías opuestas que desempeñan un papel central en el metabolismo de carbono de *E. coli*. Los experimentos de análisis de la expresión génica se llevaron a cabo en células de *E. coli* cultivadas en medio mínimo M9 suplementado con piruvato o glucosa como fuentes principales de carbono (M9/piruvato y M9/glucosa), que son sustratos para la vía gluconeogénica y glucolítica, respectivamente. Este análisis, además de revelar la expresión diferencial de un número de sRNAs bien caracterizados implicados en el control de vías metabólicas relacionadas con la glucólisis y la gluconeogénesis, demostró la expresión diferencial de algunos sRNAs putativos, incluyendo RyeA, un sRNA anotado cuyas posibles funciones y dianas aún se desconocen, aportando nueva información que potencialmente vincula las funciones reguladoras de estos sRNAs con la regulación del metabolismo central del carbono.

Descubrimiento y validación de sRNAs putativos nuevos y predichos mediante secuenciación de ARN

Por último, abordamos la búsqueda y la validación de nuevos sRNAs identificados por secuenciación de ARN (RNA-seq). Durante nuestra colaboración con el Dr. Kenneth McDowall (Universidad de Leeds) analizamos una lista de transcritos intergénicos encontrados en *E. coli* por medio de la secuenciación diferencial de ARN (dRNA-seq), seleccionamos los candidatos más prometedores bio-informáticamente y analizamos algunos de ellos por northern blot, probando la existencia de 6 nuevos sRNAs putativos en *E. coli*.

Además de participar en esta colaboración, exploramos la capacidad de diferentes estudios de secuenciación de ARN para descubrir sRNAs, para lo cual comparamos el número y la identidad de los sRNAs anotados detectados en cuatro estudios independientes llevados a cabo en *E. coli* (Raghavan *et al.* (2011), Shinhara *et al.* (2011), Conway *et al.* (2014) y Romero A. *et al.* (2014)), y encontramos que la mayoría de los sRNAs conocidos habían sido

detectados por al menos dos de los estudios. Como consecuencia, formulamos la hipótesis de que la detección independiente por varios estudios de secuenciación de ARN podría ser indicativa de la verdadera existencia de un sRNA candidato. Para probar esta idea, comparamos los candidatos de sRNAs encontrados por los cuatro estudios de secuenciación de ARN y compilamos una lista de candidatos detectados por más de un estudio, que analizamos bio-informáticamente tal y como habíamos hecho previamente con los sRNAs anotados, y seleccionamos a los mejores candidatos para su posterior validación por northern blot, obteniendo evidencias experimentales de la existencia de 9 de ellos. Adicionalmente, se encontró que algunos de los sRNAs putativos validados mostraron niveles de expresión diferentes en células crecidas en un medio rico (LB) y en un medio mínimo (M9/glucosa), señalando su posible funcionalidad.

Resultados

Tras demostrar que nuestro microarray de diseño propio es una herramienta útil y eficaz para la detección de sRNAs y el perfilamiento de su expresión y la de los mRNAs lo empleamos para el estudio de los sRNAs en células de *E. coli* cultivadas en distintas condiciones y con distintos *background* genéticos demostró que la expresión de sólo dos sRNAs anotados (GcvB y RydB) se veía significativamente afectada por el crecimiento celular en LB comparado con M9/glucosa. Así mismo, el uso del microarray para comparar la expresión transcriptómica de una cepa mutante de *E. coli* carente de PAP I y su cepa silvestre (*wild-type*) mostró el efecto desestabilizador, hasta ahora desconocido, de PAP I sobre dos sRNAs anotados, RyhB y SroH. Adicionalmente, empleamos el nuevo microarray para el estudio de células de *E. coli* cultivadas en M9/piruvato y M9/glucosa, lo cual reveló la regulación diferencial de algunos sRNAs putativos, incluyendo RyeA, un sRNA previamente identificado con funciones reguladoras desconocidas, ampliando así el número de sRNAs potencialmente implicados en la regulación del metabolismo central de carbono en *E. coli*. Por último, el análisis por northern blot de varios sRNAs predichos en *E. coli* y detectados por el microarray dio lugar a la verificación de la existencia de dos de ellos (G0-10699 y IS001).

Además del análisis de la expresión génica mediante el microarray de diseño propio, participamos en una colaboración con el grupo del Dr. Kenneth McDowall

(Universidad de Leeds), en la que analizamos y comprobamos la existencia de 6 sRNAs putativos nuevos, lo que se traduce en una ampliación del panorama de sRNAs putativos experimentalmente verificados en *E. coli*, así como en la confirmación de la eficacia de la selección de los mejores candidatos basada en el análisis *in silico* de los candidatos de sRNAs encontrados por secuenciación de ARN, obteniendo como resultado una buena tasa de éxito de verificación de la expresión *in vivo* de sRNAs putativos.

Finalmente, demostramos que una estrategia basada en la pre-selección *in silico* de los candidatos de sRNAs detectados de forma independiente en diferentes estudios de secuenciación de ARN también asegura una alta tasa de detección de sRNAs putativos, y mostramos que mediante procedimiento conseguimos elaborar una lista reducida de sRNAs predichos como candidatos prometedores para su análisis por northern blot, del cual obtuvimos verificación experimental de la existencia de 9 de ellos.

Parte de los resultados obtenidos durante este trabajo se incluyeron dos publicaciones (Romero A. *et al.*, 2014; Ruiz-Larrabeiti *et al.*, 2016b).

Conclusiones

1- Nuestro microarray personalizado es una herramienta eficiente para la detección de sRNAs y perfiles de expresión génica de sRNAs y sus mRNAs diana.

2- El análisis por medio del microarray personalizado de la expresión génica de células de *E. coli* cultivadas en medio rico (LB) y medio mínimo (M9/glucosa) demostró que sólo dos sRNAs anotados (GcvB y RydB) mostraron diferencias considerables en su nivel de expresión.

3- La comparación de la cepa silvestre (*wild-type*) de *E. coli* y su mutante isogénico *pcnB*$^-$ reveló por primera vez la expresión diferencial de los sRNAs RyhB y SroH en estas cepas, mostrando nuevos ejemplos de sRNAs cuyo nivel de expresión está controlado por la poli(A)polimerasa I *in vivo*.

4- El uso del microarray de diseño propio para la detección de sRNAs putativos predichos produjo una señal positiva para G0-10699 y IS001, cuya existencia fue confirmada por northern blot, ampliando el número de sRNAs verificados experimentalmente en *E. coli*.

5- El perfilamiento transcriptómico de células de *E. coli* cultivadas en M9/piruvato y M9/glucosa realizado con el nuevo microarray reveló la expresión diferencial de un número de sRNAs putativos, incluyendo RyeA, un sRNA previamente identificado con funciones reguladoras desconocidas, ampliando así el número de sRNAs potencialmente implicados en la regulación del metabolismo central de carbono en *E. coli*.

6- Nuestro trabajo en colaboración con el grupo del Dr. Kenneth McDowall demostró que la combinación de la secuenciación diferencial de ARN y el subsiguiente análisis *in silico* de la lista de sRNAs candidatos da como resultado una buena tasa de detección de sRNAs putativos expresados *in vivo*.

7- La pre-selección *in silico* de los sRNAs candidatos detectados de forma independiente en diferentes estudios de secuenciación de ARN para su verificación por northern blot también proporciona una alta tasa de detección de sRNAs putativos.

8- Demostramos la expresión *in vivo* de algunos nuevos sRNAs putativos (p-sRNA-1, 2, 3, 4, 5 y 6), así como sRNAs putativos predichos (IS019, G0-10698, p-sRNA-11, ECS054, C0614, p-sRNA-12, ECS163 y G0-10701).

9- Algunos nuevos sRNAs putativos (p-sRNA-1, 2, 3, 4, 5 y 6) así como el sRNA putativo previamente predicho ECS054 se expresaron diferencialmente en células de *E. coli* cultivadas en medio rico (LB) comparado con medio mínimo (M9/glucosa), lo cual es indicativo de su posible papel en la regulación del metabolismo central del carbono.

Bibliografía

Argaman, L.; Hershberg, R.; Vogel, J.; Bejerano, G.; Wagner, E. G.; Margalit, H.; Altuvia, S. Novel Small RNA-Encoding Genes in the Intergenic Regions of *Escherichia coli*. *Curr. Biol.* **2001**, *11* (12), 941–950.

Beisel, C. L.; Storz, G. Base Pairing Small RNAs and Their Roles in Global Regulatory Networks. *FEMS Microbiol. Rev.* **2010**, *34* (5), 866–882.

Chen, S.; Lesnik, E. a; Hall, T. a; Sampath, R.; Griffey, R. H.; Ecker, D. J.; Blyn, L. B. A Bioinformatics Based Approach to Discover Small RNA Genes in the *Escherichia coli* Genome. *Biosystems.* **2002**, *65* (2–3), 157–177.

Conway, T.; Creecy, J. P.; Maddox, S. M.; Grissom, J. E.; Conkle, T. L.; Shadid, T. M.; Teramoto, J. Unprecedented High-Resolution View of Bacterial Operon Architecture Revealed by RNA Sequencing. *MBio* **2014**, *5* (4), :e01442-14.

Hershberg, R.; Altuvia, S.; Margalit, H. A Survey of Small RNA-Encoding Genes in *Escherichia coli*. *Nucleic Acids Res.* **2003**, *31* (7), 1813–1820.

Hoe, C.-H.; Raabe, C. a; Rozhdestvensky, T. S.; Tang, T.-H. Bacterial sRNAs: Regulation in Stress. *Int. J. Med. Microbiol.* **2013**, *303* (5), 217–229.

Kaberdin, V. R.; Bläsi, U. Translation Initiation and the Fate of Bacterial mRNAs. *FEMS Microbiol. Rev.* **2006**, *30* (6), 967–979.

Kawano, M.; Reynolds, A. a; Miranda-Rios, J.; Storz, G. Detection of 5'- and 3'-UTR-Derived Small RNAs and Cis-Encoded Antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.* **2005**, *33* (3), 1040–1050.

Keseler, I. M.; Mackie, A.; Peralta-Gil, M.; Santos-Zavaleta, A.; Gama-Castro, S.; Bonavides-Martínez, C.; Fulcher, C.; Huerta, A. M.; Kothari, A.; Krummenacker, M.; *et al.* EcoCyc: Fusing Model Organism Databases with Systems Biology. *Nucleic Acids Res.* **2013**, *41* (Database issue), D605-12.

Li, L.; Huang, D.; Cheung, M. K.; Nong, W.; Huang, Q.; Kwan, H. S. BSRD: A Repository for Bacterial Small Regulatory RNA. *Nucleic Acids Res.* **2013**, *41* (D1), 233–238.

Lindgreen, S.; Umu, S. U.; Lai, A. S. W.; Eldai, H.; Liu, W.; McGimpsey, S.; Wheeler, N. E.; Biggs, P. J.; Thomson, N. R.; Barquist, L.; *et al.* Robust Identification of Noncoding RNA from Transcriptomes Requires Phylogenetically-Informed Sampling. *PLoS Comput. Biol.* **2014**, *10* (10).

Malone, J. H.; Oliver, B. Microarrays, Deep Sequencing and the True Measure of the

Transcriptome. *BMC Biol.* **2011**, *9* (1), 34.

Melamed, S.; Peer, A.; Faigenbaum-Romm, R.; Gatt, Y. E.; Reiss, N.; Bar, A.; Altuvia, Y.; Argaman, L.; Margalit, H. Global Mapping of Small RNA-Target Interactions in Resource Global Mapping of Small RNA-Target Interactions in Bacteria. *Mol. Cell* **2016**, *63* (5), 884–897.

Raghavachari, N.; Barb, J.; Yang, Y.; Liu, P.; Woodhouse, K.; Levy, D.; O'Donnell, C. J.; Munson, P. J.; Kato, G. J. A Systematic Comparison and Evaluation of High Density Exon Arrays and RNA-Seq Technology Used to Unravel the Peripheral Blood Transcriptome of Sickle Cell Disease. *BMC Med. Genomics* **2012**, *5* (1), 28.

Raghavan, R.; Groisman, E. A.; Ochman, H. Genome-Wide Detection of Novel Regulatory RNAs in E . Coli. *Genome Res.* **2011**, *21*, 1487–1497.

Régnier, P.; Hajnsdorf, E. Chapter 4 Poly(A)-Assisted RNA Decay and Modulators of RNA Stability. *Prog. Mol. Biol. Transl. Sci.* **2009**, *85* (C), 137–185.

Rivas, E.; Klein, R. J.; Jones, T. a; Eddy, S. R. Computational Identification of Noncoding RNAs in E. Coli by Comparative Genomics. *Curr. Biol.* **2001**, *11* (17), 1369–1373.

Romero A., D.; Hasan, A. H.; Lin, Y.-F. fei; Kime, L.; Ruiz-Larrabeiti, O.; Urem, M.; Bucca, G.; Mamanova, L.; Laing, E. E.; van Wezel, G. P.; *et al.* A Comparison of Key Aspects of Gene Regulation in Streptomyces Coelicolor and *Escherichia coli* Using Nucleotide-Resolution Transcription Maps Produced in Parallel by Global and Differential RNA Sequencing. *Mol. Microbiol.* **2014**, *94* (5), 963–987.

Ruiz-Larrabeiti, O.; Plágaro, A. H.; Gracia, C.; Gallego, L.; Hajnsdorf, E.; Kaberdin, V. R. A New Custom Microarray for sRNA Profiling in *Escherichia coli*. *FEMS Microbiol. Lett.* **2016**, 1–20.

Saramago, M.; Bárria, C.; Dos Santos, R. F.; Silva, I. J.; Pobre, V.; Domingues, S.; Andrade, J. M.; Viegas, S. C.; Arraiano, C. M. The Role of RNases in the Regulation of Small RNAs. *Curr. Opin. Microbiol.* **2014**, *18*, 105–115.

Shinhara, A.; Matsui, M.; Hiraoka, K.; Nomura, W.; Hirano, R.; Nakahigashi, K.; Tomita, M.; Mori, H.; Kanai, A.; Kana, A. Deep Sequencing Reveals as-yet-Undiscovered Small RNAs in *Escherichia coli*. *BMC Genomics* **2011**, *12* (1), 428.

Sobrero, P.; Valverde, C. The Bacterial Protein Hfq: Much More than a Mere RNA-Binding Factor. *Crit. Rev. Microbiol.* **2012**, *38* (4), 276–299.

Storz, G.; Vogel, J.; Wassarman, K. M. Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Mol. Cell* **2011**, *43* (6), 880–891.

Tao, H.; Bausch, C.; Richmond, C.; Blattner, F. R.; Conway, T. Functional Genomics: Expression Analysis of *Escherichia coli* Growing on Minimal and Rich Media. *J. Bacteriol.* **1999**,

181 (20), 6425–6440.

Thomason, M. K.; Bischler, T.; Eisenbart, S. K.; Förstner, K. U.; Zhang, A.; Herbig, A.; Nieselt, K.; Sharma, C. M.; Storz, G. Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in *Escherichia coli*. *J. Bacteriol.* **2015**, *197* (1), 18–28.

Tjaden, B.; Saxena, R. M.; Stolyar, S.; Haynor, D. R.; Kolker, E.; Rosenow, C. Transcriptome Analysis of *Escherichia coli* Using High-Density Oligonucleotide Probe Arrays. *Nucleic Acids Res.* **2002**, *30* (17), 3732–3738.

Updegrove, T. B.; Shabalina, S. A.; Storz, G. How Do Base-Pairing Small RNAs Evolve? *FEMS Microbiol. Rev.* **2015**, *39* (3), 379–391.

Vogel, J.; Sharma, C. M. How to Find Small Non-Coding RNAs in Bacteria. *Biol. Chem.* **2005**, *386* (12), 1219–1238.

Vogel, J.; Luisi, B. F. Hfq and Its Constellation of RNA. *Nat. Rev. Microbiol.* **2011**, *9* (8), 578–589.

Vogel, J.; Bartels, V.; Tang, T. H.; Churakov, G.; Slagter-Jäger, J. G.; Hüttenhofer, A.; Wagner, E. G. H. RNomics in *Escherichia coli* Detects New sRNA Species and Indicates Parallel Transcriptional Output in Bacteria. *Nucleic Acids Res.* **2003**, *31* (22), 6435–6443.

Wagner, E. G. H.; Romby, P. *Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It*; Elsevier Ltd, 2015; Vol. 90.

Wassarman, K. M.; Zhang, A.; Storz, G. Small RNAs in *Escherichia coli*. *Trends Microbiol.* **1999**, *7* (1), 37–45.

Wassarman, K. M.; Repoila, F.; Rosenow, C.; Storz, G.; Gottesman, S. Identification of Novel Small RNAs Using Comparative Genomics and Microarrays. *GENES Dev.* **2001**, *15*, 1637–1651.

Weirick, T.; Militello, G.; Müller, R.; John, D.; Dimmeler, S.; Uchida, S. The Identification and Characterization of Novel Transcripts from RNA-Seq Data. *Brief. Bioinform.* **2015**, No. May, 1–8.

Yachie, N.; Numata, K.; Saito, R.; Kanai, A.; Tomita, M. Prediction of Non-Coding and Antisense RNA Genes in *Escherichia coli* with Gapped Markov Model. *Gene* **2006**, *372*, 171–181.

Zhang, A.; Wassarman, K. M.; Rosenow, C.; Tjaden, B.; Storz, G.; Gottesman, S. Global Analysis of Small RNA and mRNA Targets of Hfq. *Mol. Microbiol.* **2003**, *50* (4), 1111–1124.

Supplementary data

Chapter 1 supplementary data

Link to Supplementary Tables S1 and S2 in Ruiz-Larrabeiti *et al.* (2016)
http://femsle.oxfordjournals.org/content/suppl/2016/05/17/fnw131.DC1

Table S3. Oligonucleotide probes used for northern blot analysis of small RNAs in *E. coli* cells grown in LB and M9/glucose as well as those used to generate the molecular size marker.

sRNAs	Sequence (5' to 3')
RybB VK5	GGGCTCCACAAAATGGGGACATCAAAGAAAAGCAGTGGC
RyjA, VK18	GGTCTTTGTCTTTCTCTCTATCCCGCTGGTACACAGG
SroH, VK26	CGGGATGTCTGAATTTCTTCAGTCTGCTGCATCCTGG
GlmY, VK31	AAGTTGGACGGCAGGCACCTTGTTGTGCGTCATTCG
GlmZ, VK32	GCGTCATTCCGGAGTTTATGAGGCACTAAGGCGAAC
IS001, VK88	GTGGGGGAGTCTGGAGATTGAGTAGATATTCTTGTTCAGAATGT ATCAGCCGATGGTTCT CTTAAGCCACGAAGAGTTCAGATAGTACAACGGCATGTCTCTTTT GACTATCTGGCAACC
C0074, VK67	CTCTCACTTTTGCTCGCGGCTCCGAAACTTTGCTGGGTTGAGATT TTAAGGTGTATCACG
C0086, VK63	AAATTTATCCGGTGAATGTGGTCGGAAAACAAAGAGGAAAGGGG GGGGGGCTAATCGGCA
IS021, VK124	GCGTTTTTTTTGATGTGATATTCTGGAACCATTAATTTGTAATTGG GTTGCTGTCGCCTA TTTAAGGGTTAATTGTTACATTGAAATGGCTAGTTATTCCCCGGG GCGATTTTCACCTCG TTTTATACATACTATAATTGATGGTTTTCTATGTGATTTAGTTAATA ACCTTCTGGGTTT
G0-10699, VK76	GGGGCTGGACGCCATTTCAAGCCTGATAAAACTGCTTAACAAAT CAGCATAACTCATTAA
IS143, VK85	CCATTAAACGACATCTGACTTCTTCCGGGG

Ladder	
rnpB (M1)	GCTCTCTGTTGCACTGGTCGTGGGTTTCC
6S (*ssrS*)	GGTGAATGTGTCGTCGCAGTTTTAAGGCTTCTCGGA
5S (*rrfH*)	CGCTACGGCGTTTCACTTCTGAGTTCGGC
tRNA (*asnT*)	TACGCATTAACAGTCCGCCGTTCTACCGACTGAACTACAGA

Chapter 2 supplementary data

Table S4. Functional groups of genes showing higher (↑) or lower (↓) relative expression in *E. coli* cells grown in M9/pyruvate *versus* M9/glucose. Functional clustering of differentially expressed genes (p-value≤0.005, expression fold change>2) revealed by microarray analysis

General category	Specific pathway	Gene name	Regulation, expression fold change	
Central carbon metabolism	Gluconeogenesis	*pck*	↑	6
		fbp	↑	2
		ppsA	↑	5
		maeB	↑	3
	Glycolysis	*pykF*	↓	3.6
		ydbK	↓	5
	Glycolysis and gluconeogenesis	*eno*	↓	2.5
		fbaA, fbaB	↓	2-3
		gapA	↓	4
		pgi	↓	2
	Glycolate and glyoxylate degradation	*glcF*	↑	2-4
		glcB, glcD	↑	2-5.5
	Pyruvate metabolism	*ackA*	↑	2.2
		acs	↑	4.3
		aldA	↑	5.9
		ldhA	↑	5.5
		maeB	↑	3
		mqo	↑	5.2
	Acetoacetate degradation	*atoA, atoD*	↑	2.6-11.1
	Galactitol degradation	*gatD, gatY*	↑	4.2-3.3
	Fucose degradation	*fucA, fucO*	↑	4
	Chitobiose degradation	*celF*	↑	2
	Mixed acid fermentation	*adhE, adhP*	↓	2-3
		tdcE	↓	3
		fumB, fumC	↓	2
	Starch and sucrose metabolism	*malP, malQ*	↓	2.5
		otsA	↓	2.3
		glgb	↓	2.7
	Hexitol biosynthesis and degradation	*gutD*	↓	2.6
		mtlD	↓	2
	Pentose phosphate pathway	*talA*	↓	4.2
		tktB	↓	2.9
		edd	↓	3.3
		gntK	↓	3.9-7.6
Amino acid metabolism	Lysine, Phenylalanine, valine, alanine, isoleucine, leucine biosynthesis	*ilvA, ilvD, ilvE, ilvM*	↑	2.8-5.4
		ilvB, ilvC, ilvN	↓	2.6-14.9
	Lysine biosynthesis	*dapB*	↓	2.6
		lysC	↓	5.6
	Tyrosine, phenylalanine biosynthesis	*phea*	↓	2.2
	Phenylalanine degradation	*paaJ*	↑	3

	Valine, leucine and isoleucine biosynthesis	*leuA*	↑	2-3
	Alanine biosynthesis and degradation	*dadA, dadX*	↑	24
	Arginine degradation	*astA, astB, astC, astD, astE*	↑	2.1-3.5
	Aromatic amino acid biosynthesis	*aroH, aroF, aroL*	↓	2.2-4
		tyrA	↓	3.7
	Arginine and aromatic amino acids biosynthesis	*argA, argB, argC, argD, argF, argG, argH*	↓	7-15
		carA, carB	↓	55-150
	Threonine degradation	*tdh*	↑	2.4
		kbl	↑	2.1
	Tryptophan and cysteine degradation (via pyruvate)	*tnaA*	↑	2
	Methionine and cysteine biosynthesis	*metA, metB, metC, metK, metL*	↓	2.3-8.8
	Asparagine metabolism	*iaaA*	↓	3
	Cysteine metabolism	*cysK, cysM*	↓	2.7-4.5
	Glutamine biosynthesis and degradation	*glnA*	↓	3.2
		glsA (ybaS)	↓	5.25
		gltB, gltD	↓	4
	Glutamate degradation	*gadA, gadB*	↓	6.8-9.4
	Histidine biosynthesis	*hisB, hisC, hisD, hisF, hisG, hisH, hisI, hisL*	↓	2.5-3.6
	Tryptophan biosynthesis	*dsdA, dsdC (LysR)*	↓	2
		trpA, trpB, trpC, trpD, trpE	↓	2.2-12
	Glutathione metabolism	*ggt*	↓	2
Processing of genetic information	DNA repair	*recF, recG, recQ*	↓	2-2.8
	DNA replication, repair, recombination	*priB*	↓	3
		holD	↓	2.8
	DNA replication inhibition	*cspD*	↑	2.5
	DNA folding	*dnaJ*	↓	6
	DNA repair	*xseA*	↓	2.2
	Supercoiled DNA relaxation	*topA*	↓	2
	Transcription regulation	*rsd*	↑	2.1
		crp	↑	3.4
		fis	↓	4.5
		trpR	↓	2.2
		metJ, metR	↓	3-4.5
	Transcription	*rpoA*	↓	2.8
	Regulation of DNA replication	*dsdC*	↓	2
		hda	↓	2.6
	RNA folding, sorting and degradation	*rnpA*	↓	3.8
		pcnB	↓	2.9
	tRNA biogenesis, processing or modification	*dusB*	↓	3.6
		spoU	↓	2
	Ribosome recycling	*hslR*	↓	2.6

	Ribosome biogenesis	*ksgA*	↓	2.3
		rplA,rplB, rplC, rplD, rplF, rplI, rplJ, rplK, rplM, rplQ, rplR, rplS, rplV, rplW, rplX	↓	2-4.3
		rpmB, rpmC, rpmD, rpmE, rpmG	↓	2.1-3.2
		rpsB, rpsC, rpsD, rpsE, rpsF, rpsG, rpsI, rpsK, rpsL, rpsN, rpsO, rpsQ, rpsR, rpsS, rpsT, rpsU	↓	2-4.8
		rimI	↓	2.2
	Nucleosides and Nucleotides metabolism	*aphA*	↑	7.5
		guaA, guaB, guaC	↓	2.1-2.8
		ndk	↓	3.6
		purB, purD, purE, purF, purH, purK, purN, purT	↓	2.5-5.5
		pyrB, pyrC, pyrD, pyrE, pyrF, pyrH, pyrI, pyrG,	↓	2.1-28
		upp	↓	7.3
		yjjG	↓	2.4
		cmk	↓	2.5
		codA	↓	7.5
		deoD	↓	2.1
	Translation	*infA*	↓	2.7
		yciH	↓	3.5
		yeiP	↓	2.3
Cellular envelope metabolism	Cell Wall Biosynthesis	*ddlA, ddlB*	↑	3.6
	Outer membrane biosynthesis	*glmS*	↓	2
	Repression of peptidoglycan degradation	*ycjZ*	↓	2
	Cell wall recycling	*ampD*	↑	2.1
		ldcA	↑	2.5
		ycjH	↑	2
	Cell division	*nlpD*	↑	2.2
	Formation of cell shape	*mreC*	↓	3.8
Response to environmental signals	Response to glucose-6-phosphate	*uhpA, uhpB*	↓	2-2.3
	Response to copper homeostasis	*cueO*	↓	4.4
	Response to nitrogen limitation	*glnL*	↓	2.4
Protein biosynthesis turnover and folding	Protein folding and degradation	*clpA, clpS*	↑	3-2.2
		hslU, hslV	↓	7.8
		groL	↓	13
	Protein synthesis/translation	*efp*	↓	2.3
		prfC	↓	2.4
	Protein folding (stress related)	*clpB*	↓	4.1
		dnaK	↓	9.8
		groS	↓	15
		hslO	↓	2.3
		ibpA, ibpB	↓	4.5-2.5

		ridA	↓	3.5
		degP	↓	2.1
	Protein folding	*grpE*	↓	2.3
		tig	↓	2.3
Transport	Transport of acetate, glycolate and tellurite	*actP*	↑	5.3
	Transport of sugar and polysaccharides	*ptsG, ptsH*	↓	2.2-4.8
		srlB, srlE	↓	2.1-3.6
		mtlA	↓	3.5
		manX	↓	2.3
		treB	↓	3
		gntT	↓	11.5
		gatC	↑	4
	Transport of glutathione	*gsiA, gsiB, gsiC, gsiD*	↓	2.1-2.2
	Transport of taurine	*tauA*	↓	2.5
	Transport of polyamine	*potB, potD, potG, potI*	↓	2-2.3
	Transport of nucleosides	*tsx*	↓	2.3
	Transport of cytosine and uracil	*uraA*	↓	15
		codB	↓	25.5
	Transport of amino acids	*gadC*	↓	3.4
		mmuP	↓	3.5
		abc (metN)	↓	5
		metI, metQ	↓	2.9-3.2
		artI, artJ, artM, artP, artQ	↓	2-19.7
		hisJ, hisM, hisP, hisQ	↓	2.2-2.9
		yehW, yehX, yehY	↓	2-2.6
		aroP	↓	2.1
	Transport of peptides and polypeptides	*tatC* (export)	↓	2.1
		cstA (Import, starvation)	↑	5.7
	Transport of phosphate	*phoS*	↑	4.1
		pstA, pstC	↑	3-3.4
	Transport of organometallic compounds	*btuD*	↓	2.3
	Transport of ferric compounds	*fhuB*	↓	2.1
	Export of copper and silver	*cusA, cusB, cusC, cusF, cusR, cusS*	↓	3-24.9
	Transport of sulfate and selenite and other sulfur compounds	*sbp*	↓	11.3
		ssuA	↓	2.2
		cysP, cysW	↓	19-31
	Transport of thiamine	*tbpA*	↓	2.5
		thiP, thiQ	↓	2.5-2.8
	Transport of drugs	*emrA, emrB*	↓	2-3
		mdtC, mdtD (yegF)	↓	2.2
		mdtE, mdtF	↓	9.2-15.8
	Membrane Transport and bacterial secretion system	*gspC, gspD, gspG*	↓	3.6, 4.1, 2.6
		secG	↓	2.7
Fatty acid metabolism	Fatty acid and lipid biosynthesis	*fabA, fabF*	↓	2.1-2.8

	Fatty acid biosynthesis initiation	*accB, accC*	↓	3.2-6
	Lipopolysaccharide biosynthesis and glycan biosynthesis and metabolism	*rfaC, rfaG*	↓	7
	CDP-diacylglycerol biosynthesis	*gpsA*	↓	2.7
	Fatty Acid degradation	*fadA, fadB, fadE, fadI, fadJ*	↑	2-5.4
Generation of precursor metabolites and energy	Sulfur metabolism	*sseA*	↑	2.4
		cysC, cysD, cysI, cysJ, cysN	↓	28-114
	Oxidative phosphorylation	*ccmf, ccmH*	↓	2.2, 2.1
		appB, appC	↓	2.6, 3.7
		nuoC, nuoE, nuoF, nuoG, nuoH, nuoI, nuoL	↓	2.6
	Nitrogen metabolism	*hyaB*	↓	2.7
		hybA, hypB, hypD,	↓	2.3-2.8
		napA, napB, napF, napH,	↓	2.7-4.3-
		nirB	↓	3.1
	Hydrogenase maturation	*hypD*	↓	2.7
Metabolism of cofactors and vitamins	Lipoate biosynthesis and incorporation	*lipA*	↑	3.9
	Nicotinate and nicotinamide metabolism	*nadB, nadD*	↓	2-2.3
		pntA, pntB,	↓	3.2-4
		pncC	↓	2.1
	Folate biosynthesis	*pabB*	↓	2.3
		folA	↓	2
		folK	↓	3.7
		metF	↓	5.2
	Biotin metabolism	*bioC*	↓	2.7
	Quinone biosynthesis	*menB, menC, menD, menE, menH*	↓	3.8
		ydiI	↓	2.4
	Lipoate biosynthesis and incorporation, biotin biosynthesis, methylerythritol phosphate pathway and thiazole biosynthesis	*fdx*	↓	2.2
	Iron-sulfur cluster assembly	*hscA, hscB*	↓	2.2
		iscx	↓	2.1
Stress response	Transcription under stress conditions	*rpoS*	↑	2.4
	Response to hydrogen peroxide and cadmium stress, inhibition of biofilm formation	*ychH*	↑	6.2
	Blue light-responsive regulator of BluR	*ycgF*	↑	2.6
	Superoxide radicals degradation	*katE*	↓	5
		katG	↓	2.8
		sodA, sodB	↓	2.1-9.3
	Bacteriocin production	*cvpA*	↓	4.4
	Acid resistance	*gadE, gadW, gadX*	↓	2-6.4
		hdeA, hdeB, hdeD	↓	4.2-4.4

		yhim	↓	3.6
	Starvation lipoprotein/ Acid stress response	*slp*	↓	3.1
	Proteolysis of bacteriophage protein	*hflD*	↓	2.6
	Component of restriction-modification system	*hsdM, hsdR, hsdS*	↓	2.0-9.6
	Part of the defective prophage	*iss (borD)*	↓	8.3
	Transcriptional regulation and resistance to antibiotics, oxidative stress, organic solvents, and heavy metals	*marA, marB, marR*	↓	5.7-8
	Restriction endonuclease	*mcrB, mcrC*	↓	4.6-6.2
	Antimicrobial peptide	*ompT*	↓	5.6
	Oxidative, antibiotic and other environmental stress response	*rnfC (rsxC)*	↓	2.1
Amines and Polyamines Biosynthesis	Putrescine biosynthesis	*speB*	↓	2.1
	Spermidine biosynthesis	*speD, speE*	↓	3.1-2.4
Adhesion and motility	Biofilm formation induction	*dosC*	↓	3.8
		ydaM	↓	5.1
	Biofilm formation inhibition	*yahA*	↑	3
		gmr	↓	2.5
	Biofilm and motility inhibition	*ymgA, ymgC*	↓	2.3-3.3
	Promotion of adhesion, inhibition of motility	*uspF*	↑	2.7
	Flagellar biosynthesis	*fliO, fliN, fliM*	↓	2.5-2.9
	Chemotaxis	*cheB, cheR, cheY, cheZ*	↓	2-2.4
	Acid stress and biofilm formation	*ariR*	↓	2.8

Table S5. Oligonucleotide probes used for northern blot analysis of small RNAs in *E. coli* cells grown in M9/pyruvate and M9/glucose.

sRNAs	Sequence (5' to 3')
RyeA	GCGTCGTCATCTATTCTTAAAGGGCAAGGCAAC
CyaR	CACAGGCTAAGGAGGTGGTTCCTGGTACAGCTAGCA
RyhB	CTGGAAGCAATGTGAGCAATGTCGTGCTTTCAGGTTC
GcvB	GGACAGACAGGGTAAATGTACAGGAAGTGAAAAAAGGTAGCTTTGCTACCATG GTCTGAA
GlmY	AAGTTGGACGGCAGGCACCTTGTTGTGCGTCATTCG
GlmZ	GCGTCATTCCGGAGTTTATGAGGCACTAAGGCGAAC

Chapter 3 supplementary data

Table S6. A complete list of known sRNAs detected by four independent RNA-seq analyses. This list compiles the known sRNAs detected by Raghavan *et al.* (2011), Shinhara *et al.* (2011), Conway *et al.* (2014), and/or Romero *et al.* (2014) (detection by any of these studies is indicated in **bold**). The references include some of the studies involved in the discovery and characterization of these sRNAs.

sRNA names	Flanking genes sRNA coordinates and orientation	RNA-seq studies				References
SokC Sof	*mokC-nhaA* 16952→17030	Romero	Raghavan	Shinhara	Conway	(Pedersen and Gerdes, 1999)
ChiX MicM SroB, RybC,	*ybaK-ybaP* 506426→506515	Romero	Raghavan	Shinhara	Conway	(Vogel *et al.*, 2003)
RyeA SraC, psrA8, tkpe79, IS091	*pphA-yebY* 1921090→1921365	Romero	Raghavan	Shinhara	Conway	(Argaman *et al.*, 2001)
DsrA	*yodD-yedP* 2023243←2023337	Romero	Raghavan	Shinhara	Conway	(Sledjeski and Gottesman, 1995)
MicF Stc	*ompC-rcsD* 2311106→2311203	Romero	Raghavan	Shinhara	Conway	(Andersen *et al.*, 1987)
RyeG IS118, ECS009	*yfdI-tfaS* 2468483←2468688	Romero	Raghavan	Shinhara	Conway	(Chen *et al.*, 2002)
MicA SraD, psrA10	*luxS-gshA* 2812824→2812900	Romero	Raghavan	Shinhara	Conway	(Argaman *et al.*, 2001)
ArcZ psrA16, SraH, RyhA	*elbB-arcB* 3348599→3348721	Romero	Raghavan	Shinhara	Conway	(Argaman *et al.*, 2001)
GadY IS183	*gadW-gadX* 3662887→3662995	Romero	Raghavan	Shinhara	Conway	(Chen *et al.*, 2002)
RybB p25	*rcdA-ybjL* 887197←887281	Romero	-	Shinhara	Conway	(Vogel *et al.*, 2003)
RydC ECS211	*cybB-ydcA* 1489466←1489551	Romero	-	Shinhara	Conway	(Zhang *et al.*, 2003; Antal *et al.*, 2005)
SroA tpe79	*tbpA-sgrR* 75470←75609	Romero	Raghavan	Shinhara	-	(Vogel *et al.*, 2003)
SgrS RyaA	*sgrR-setA* 77367→77595	Romero	Raghavan	Shinhara	-	(Vanderpool and Gottesman, 2004)
tp2	*pdhR-aceE* 122859←123017	Romero	Raghavan	Shinhara	-	(Rivas *et al.*, 2001)
rybA	*ybiP-mntR* 852035←852271	Romero	Raghavan	Shinhara	-	(Wassarman *et al.*, 2001; Waters *et al.*, 2011; Gerstle *et al.*, 2012)
RdlA	*ldrA-ldrB* 1268546→1268658	Romero	Raghavan	Shinhara	-	(Kawano *et al.*, 2002)

rdlB	*ldrB-ldrC* 1269081→1269146	Romero	Raghavan	Shinhara	-	(Kawano *et al.*, 2002, 2005)
RdlC	*ldrC-chaA* 1269616→1269715	Romero	Raghavan	Shinhara	-	(Kawano *et al.*, 2002, 2005)
McaS IS061, IsrA	*abgR-smrA* 1403678←1403777	Romero	Raghavan	Shinhara	-	(Chen *et al.*, 2002)
SokB	*mokB-trg* 1490143→1490200	Romero	Raghavan	Shinhara	-	(Pedersen and Gerdes, 1999; Kawano *et al.*, 2005)
rydB	*sufA-ydiH* 1762730←1762794	Romero	Raghavan	Shinhara	-	(Rivas *et al.*, 2001; Wassarman *et al.*, 2001; Hershberg *et al.*, 2003)
RprA psrA5, IS083	*ydik-ydiL* 1768396→1768505	Romero	Raghavan	Shinhara	-	(Argaman *et al.*, 2001; Majdalani *et al.*, 2001)
RyeB tkpe79, SdsR	*pphA-yebY* 1921125←1921231	Romero	Raghavan	Shinhara	-	(Rivas *et al.*, 2001; Wassarman *et al.*, 2001; Vogel *et al.*, 2003)
SibA QUAD1a, tp11, ryeC	*yegL-SibB* 2151335→2151555	Romero	Raghavan	Shinhara	-	(Wassarman *et al.*, 2001; Fozo *et al.*, 2008b)
SibB QUAD1b, tpe60, RyeD	*sibA-mdtA* 2151670→2151815	Romero	Raghavan	Shinhara	-	(Wassarman *et al.*, 2001; Fozo *et al.*, 2008b)
RyfA PAIR3, tp1	*IS128-sseB* 2651879→2652180	Romero	Raghavan	Shinhara	-	(Rivas *et al.*, 2001; Wassarman *et al.*, 2001; Hershberg *et al.*, 2003)
GlmY tke1, SroF, SraJ, RyiA, k19	*glrK-purL* 2689170←2689363	Romero	Raghavan	Shinhara	-	(Rivas *et al.*, 2001; Vogel *et al.*, 2003; Urban and Vogel, 2007)
RyfD	*clpB-yfiH* 2732170←2732316	Romero	Raghavan	Shinhara	-	(Kawano *et al.*, 2005)
CsrB	*yqcC-syd* 2922178←2922547	Romero	Raghavan	Shinhara	-	(Liu *et al.*, 1997)
GcvB IS145	*gcvA-ygdI* 2940718→2940925	Romero	Raghavan	Shinhara	-	(Argaman *et al.*, 2001)
OmrA psrA12, RygA, PAIR2a, t59, SraE	*aas-OmrB* 2974123←2974213	Romero	Raghavan	Shinhara	-	(Argaman *et al.*, 2001; Wassarman *et al.*, 2001)
OmrB PAIR2b, t59, RygB	*omrA-galR* 2974325←2974410	Romero	Raghavan	Shinhara	-	(Wassarman *et al.*, 2001; Vogel *et al.*, 2003)

SibD tp8, C0730, IS156, QUAD1d, RygD	*yqiK-sibE* 3192735←3192890	Romero	Raghavan	Shinhara	-	(Rudd, 1998; Rivas *et al.*, 2001; Wassarman *et al.*, 2001; Fozo *et al.*, 2008a)
SibE QUAD1e, rygE	*ibsD-rfaE* 3193105←3193265	Romero	Raghavan	Shinhara	-	(Fozo *et al.*, 2008a)
GlmZ SraJ, RyiA	*aslA-hemY* 3984455→3984680	Romero	Raghavan	Shinhara	-	(Argaman *et al.*, 2001; Wassarman *et al.*, 2001; Vogel *et al.*, 2003)
Spot42 Spf	*polA-yihA* 4047920→4048032	Romero	Raghavan	Shinhara	-	(Ikemura *et al.*, 1973; Joyce and Grindley, 1982)
CsrC SraK, psrA21, RyiB, tpk2, IS198	*yihA-yihI* 4049059→4049300	Romero	Raghavan	Shinhara	-	(Argaman *et al.*, 2001; Wassarman *et al.*, 2001; Chen *et al.*, 2002; Weilbacher *et al.*, 2003)
RdlD	*ldrD-yhjV* 3698159→3698270	Romero	Raghavan	Shinhara	-	(Kawano *et al.*, 2002)
SroH IS206	*yjaZ-thiH* 4188340←4188512	Romero	Raghavan	Shinhara	-	(Vogel *et al.*, 2003)
RyjA psrA24, SraL	*soxR-yjcD* 4275945←4276090	Romero	Raghavan	Shinhara	-	(Argaman *et al.*, 2001; Wassarman *et al.*, 2001)
SymR RyjC	*symE-hsdS* 4577858→4577942	Romero	Raghavan	Shinhara	-	(Kawano *et al.*, 2005)
ryjB	*sgcA-sgcQ* 4526000→4526089	-	Raghavan	Shinhara	Conway	(Zhang *et al.*, 2003; Kawano *et al.*, 2005)
psrD psrA4, pke20	*yceF-yceD* 1145812→1145980	-	Raghavan	Shinhara	Conway	(Argaman *et al.*, 2001)
RyhB psrA18, IS176, SraI	*yhhX-yhhY* 3578940←3579042	Romero	Raghavan	-	-	(Argaman *et al.*, 2001; Wassarman *et al.*, 2001; Massé and Gottesman, 2002)
OxyS	*argH-oxyR* 4156298←4156419	Romero	Raghavan	-	-	(Altuvia *et al.*, 1997, 1998)
CyaR RyeE	*yegQ-ogrK* 2165138→2165320	Romero	-	Shinhara	-	(Wassarman *et al.*, 2001)
OhsC RyfC	*shoB-acpS* 2698542→2698620	Romero	-	Shinhara	-	(Kawano *et al.*, 2005)
C0614	*sseA-IS128* 2651370←2651478	Romero	-	Shinhara	-	(Tjaden *et al.*, 2002)
SibC QUAD1c, t27, RygC	*fau-serA* 3054873→3055015	Romero	-	Shinhara	-	(Rudd, 1998; Wassarman *et al.*, 2001; Fozo *et al.*, 2008a)
agrA ECS022	*dinQ-agrB* 3646085→3646172	Romero	-	Shinhara	-	(Weel-Sneve *et al.*, 2013)

agrB ECS004	*agrA-arsR* 3646316→3646400	Romero	-	Shinhara	-	(Weel-Sneve *et al.*, 2013)
IstR-1	*ivbL-tisA* 3851130←3851216	Romero	-	Shinhara	-	(Vogel *et al.*, 2003)
IS128	*C0614-ryfA* 2651537→2651745	-	-	Shinhara	Conway	(Chen *et al.*, 2002)
tff	map-rpsB 189712→189847	-	Raghavan	Shinhara	-	(Rivas *et al.*, 2001)
sroD	rnd-fadD 457952←458008	-	Raghavan	Shinhara	-	(Vogel *et al.*, 2003)
tpke11	dnaK-dnaJ 1229852→1229930	-	Raghavan	Shinhara	-	(Rivas *et al.*, 2001)
sroE	hisS-ispG 1286289←1286459	-	Raghavan	Shinhara	-	(Vogel *et al.*, 2003)
FnrS, rydA ECS161	ydaN-C0343 1407150→407271	-	Raghavan	Shinhara	-	(Wassarman *et al.*, 2001)
isrB IS092	yecJ-yecR 1886041←1886126	-	Raghavan	Shinhara	-	(Chen *et al.*, 2002; Hershberg *et al.*, 2003)
rttR	ychS-tyrV 1985863←1986022	-	Raghavan	Shinhara	-	(Bösl and Kersten, 1991)
C0299	hlyE-umuD 2638617←2638708	-	Raghavan	Shinhara	-	(Tjaden *et al.*, 2002)
sraA psrA3, t15	clpX- lon 3054005→3054187	-	Raghavan	Shinhara	-	(Argaman *et al.*, 2001; Hershberg *et al.*, 2003)
C0293	lcd-ymfD 1195937→1196009	-	Raghavan	-	-	(Tjaden *et al.*, 2002)
C0343	fnrS-dbpA 1407387→1407461	-	Raghavan	-	-	(Tjaden *et al.*, 2002)
tpke70	lpxP-ypdK 2494216←2494651	-	Raghavan	-	-	(Rivas *et al.*, 2001)
isrC IS102	yeeP-flu 2069339→2069542	-	Raghavan	-	-	(Chen *et al.*, 2002)
micC IS063, tke8	ompN-ybdK 1435145→1435253	-	Raghavan	-	-	(Chen *et al.*, 2004)
psrO psrA15, p3, sraG	Pnp-rpsO 3309247→3309420	-	Raghavan	-	-	(Argaman *et al.*, 2001)
eyeA I001	insO-1-ykfC 271804→271878	-	Raghavan	-	-	(Sætrom *et al.*, 2005)
MgrR	yneM-dgcF 1620841←1620938	-	Raghavan	-	-	(Zhang *et al.*, 2003)
SokX	Cas3-cysH 2885376→2885431	-	Raghavan	-	-	(Kawano *et al.*, 2005)
C0719	glcA-glcB 3119380→3119601	-	-	Shinhara	-	(Tjaden *et al.*, 2002)
sroC HB_314	gltJ-gltI 685904←686066	-	-	Shinhara	-	(Vogel *et al.*, 2003)

rseX IS096	yedR-yedS 2031673→2031763	-	-	Shinhara	-	(Chen *et al.*, 2002)
C0664	norW-hypF 2833077→2833189	-	-	Shinhara	-	(Tjaden *et al.*, 2002)
C0362	fdnI-yddM 1550025→1550410	-	-	Shinhara	-	(Tjaden *et al.*, 2002)

Table S7. A complete list of new putative sRNAs detected by four RNA-seq analyses. This list compiles the sRNA candidates identified by Raghavan *et al.* (2011), Shinhara *et al.* (2011), Conway *et al.* (2014), and/or Romero *et al.* (2014) (discovery by any of these studies is indicated in **bold**). The references include some of the studies involved in the discovery and characterization of these sRNAs.

sRNA names	Flanking genes, sRNA coordinates and orientation	RNA-seq studies				References
G0-10699	*sdiA-yecC* 1994930→1995121	Romero	Raghavan	-	Conway	(Chen *et al.*, 2002; Tjaden *et al.*, 2002; Zhang *et al.*, 2003; Yachie *et al.*, 2006)
IS019	*nmpC-essD* 576438→576505	Romero	Raghavan	-	Conway	(Chen *et al.*, 2002)
G0-10698	*ynfM-asr* 1668973←1669160	-	Raghavan	Shinhara	Conway	(Wassarman *et al.*, 2001; Chen *et al.*, 2002; Yachie *et al.*, 2006)
ECS198	*ybfE-ybfF* 712026→712209	-	Raghavan	Shinhara	Conway	
ECS119	*glpX-glpK-glpF* 4116114-4116537	-	Raghavan	Shinhara	-	
ECS028	*sspB-sspA* 3376716←3376831	-	Raghavan	Shinhara	-	(Wassarman *et al.*, 2001; Vogel *et al.*, 2003)
G0-10701	*ygfI-yggE* 3065209→3065366	-	Raghavan	Shinhara	-	(Bak *et al.*, 2015)
G0-10705	*glnA-typA* 4056194→4056265	-	Raghavan	Shinhara	-	(Rivas *et al.*, 2001; Tjaden *et al.*, 2002; Sætrom *et al.*, 2005)
G0-10700	*yejG-bcr* 2276280←2276520	-	Raghavan	Shinhara	-	(Wassarman *et al.*, 2001)
NC093	*yggW-yggM-ansB* 3098850→3098919	-	Raghavan	Shinhara	-	(Yachie *et al.*, 2006)
	1445308-1445542	-	Raghavan	-	Conway	
	3988790-3989175	-	Raghavan	-	Conway	
	4285395-4285786	-	Raghavan	-	Conway	
	4440200-4440404	-	Raghavan	-	Conway	
IS116	2411266→2411418	-	Raghavan	-	Conway	(Chen *et al.*, 2002)
ECS027	637846→637925	-	-	Shinhara	Conway	(Chen *et al.*, 2002)
ECS031	1432570←1432644	-	-	Shinhara	Conway	
ECS168	4434586→4434673	-	-	Shinhara	Conway	

ECS207	3920625→3920699	-	-	Shinhara	Conway	
ECS210	889150→889273	-	-	Shinhara	Conway	
ECS223	1029185→1029249	-	-	Shinhara	Conway	(Rivas *et al.*, 2001)
k2	*ybjE-aqpZ* 914410←914525	Romero	Raghavan	-	-	(Rivas *et al.*, 2001)
t29	*recE-racC* 1415447→1415590	Romero	Raghavan	-	-	
	cspB-cspF 1639658←1639740	Romero	Raghavan	-	-	
	ydfU-rem 1642532←1642655	Romero	Raghavan	-	-	
	flxA-ydfW 1644765←1644847	Romero	Raghavan	-	-	
IS141	*iap-cas2* 2875905←2876486	Romero	Raghavan	-	-	(Chen *et al.*, 2002)
	yjhX-yjhR 4532010←4532248	Romero	Raghavan	-	-	
	nanC-fimB 4537775←4537952	Romero	Raghavan	-	-	
	nanC-fimB 4538709→4538880	Romero	Raghavan	-	-	
	mdtM-yjiP 4566620←4566778	Romero	Raghavan	-	-	
I006	*proA-ykfI* 261977←262243	Romero	-	Shinhara	-	(Sætrom *et al.*, 2005)
ECS020	*InsB-4-cspH* 1049560←1049936	Romero	-	Shinhara	-	
ECS025	*cspF-ydfT* 1640375←1640447	Romero	-	Shinhara	-	
ECS007	*insH-yoeA* 2066405←2066508	Romero	-	Shinhara	-	
ECS054	*intZ-yffL* 2558320←2558395	Romero	-	Shinhara	-	
ECS174	*rodZ-rlmN* 2640965←2641027	Romero	-	Shinhara	-	(Sætrom *et al.*, 2005)
ECS034	*zupT-ribB* 3181345←3181648	Romero	-	Shinhara	-	
ECS050	*mnmE-tnaC* 3886215←3886316	Romero	-	Shinhara	-	
ECS094	*insG-yjhB* 4501450←4501548	Romero	-	Shinhara	-	
ECS109	*epmB-ecnA* 4373926←4373976	Romero	-	Shinhara	-	
ECS163	*thiC-rsd* 4194200←4194325	Romero	-	Shinhara	-	(Vogel *et al.*, 2003)
ecr1743 (d+)	*ves-spy* 1823080←1823177	Romero	-	-	Conway	
	yjgZ-insG 4499861→4500185	Romero	-	-	Conway	

ecr0174 (u+)	*dxr-ispU* 194784→194848	Romero	-	-	-	
ecr0210 (d+)	*yafE-mltD* 232367→232584	Romero	-	-	-	
ecr0375 (u+)	*yaiU-yaiV* 393639→393720	Romero	-	-	-	
ecr1891 (d+)	*motA-flhC* 1975230←1975327	Romero	-	-	-	
ecr4051 (d-)	*pspG-qor* 4261162→4261273	Romero	-	-	-	
ecr4224 (d+)	*chpS-chpB* 4446731→4446835	Romero	-	-	-	
ecr0235 (u+)	*yafP-ykfJ* 253340→253465	Romero	-	-	-	
ecr0573 (u-)	*cusC-cusF* 596198←596332	Romero	-	-	-	
ecr0770 (d-)	*ybhI-ybhJ* 802544←802660	Romero	-	-	-	
ecr2774 (u+)	*ygcW-yqcE* 2898315←2898388	Romero	-	-	-	
ecr2775 (u+)	*ygcW-yqcE* 2898527→2898590	Romero	-	-	-	
ecr4065 (d-)	*yjcE-yjcF* 4279660←4279744	Romero	-	-	-	
RVS(III)-4131791	*metF-katG* 4131585←4131791	Romero	-	-	-	
tss40	2257512→2257740	Romero	-	-	-	
tss41	2265741←2265816	Romero	-	-	-	
tss50	2654399→2654550	Romero	-	-	-	
G0-10697	*ychE-oppA* 1298697→1298951	-	Raghavan	-	-	(Wassarman *et al.*, 2001; Kawano *et al.*, 2005)
G0-10702	*yhcC-gltB* 3352086→3352191	-	Raghavan	-	-	
G0-10703	*yhcF-yhcG* 3365635←3365792	-	Raghavan	-	-	
G0-10704	*yigE-corA* 3999214→3999357	-	Raghavan	-	-	(Rivas *et al.*, 2001; Chen *et al.*, 2002)
G0-10706	*ytfL-msrA* 4439248←4439353	-	Raghavan	-	-	
	11787-12162	-	Raghavan	-	-	
	57110-57363	-	Raghavan	-	-	
	63265←63428	-	Raghavan	-	-	
	131261-131614	-	Raghavan	-	-	
	146695-146967	-	Raghavan	-	-	
	259325-259611	-	Raghavan	-	-	

344216→344627	-	Raghavan	-	-
395512-395862	-	Raghavan	-	-
398558-398816	-	Raghavan	-	-
440568-440772	-	Raghavan	-	-
611718-612037	-	Raghavan	-	-
623734-624107	-	Raghavan	-	-
637797-638167	-	Raghavan	-	-
702835-703166	-	Raghavan	-	-
728045-728356	-	Raghavan	-	-
849321-849672	-	Raghavan	-	-
882612-882895	-	Raghavan	-	-
889043-889311	-	Raghavan	-	-
915271-915695	-	Raghavan	-	-
931274-931817	-	Raghavan	-	-
1050399-1050683	-	Raghavan	-	-
1062999-1063258	-	Raghavan	-	-
1194175-194345	-	Raghavan	-	-
1276842-1277179	-	Raghavan	-	-
1297345-1297820	-	Raghavan	-	-
1324666→1324875	-	Raghavan	-	-
1416254-1416694	-	Raghavan	-	-
1432282←1432981	-	Raghavan	-	-
1521090-1521330	-	Raghavan	-	-
1524005-1524270	-	Raghavan	-	-
1630310-1631062	-	Raghavan	-	-
1638610←1639362	-	Raghavan	-	-
1639579-1639878	-	Raghavan	-	-
1642329←1642674	-	Raghavan	-	-
1644762-1645197	-	Raghavan	-	-
1668977→1669399	-	Raghavan	-	-
1671526-1671936	-	Raghavan	-	-
1735570-1735867	-	Raghavan	-	-
1739147→1739436	-	Raghavan	-	-
1800595-1801117	-	Raghavan	-	-
1903284→1903711	-	Raghavan	-	-
1984153-1984948	-	Raghavan	-	-
2051353-2051666	-	Raghavan	-	-
2085139←2085352	-	Raghavan	-	-
2226781-2227003	-	Raghavan	-	-
2244792←2245084	-	Raghavan	-	-
2288203-2288521	-	Raghavan	-	-
2310772-2311105	-	Raghavan	-	-
2342192-2342886	-	Raghavan	-	-
2463030-2463322	-	Raghavan	-	-
2481362-2481776	-	Raghavan	-	-
2583548-2583752	-	Raghavan	-	-
2759196-2759372	-	Raghavan	-	-
2763176←2763534	-	Raghavan	-	-

2798498→2798744	-	Raghavan	-	-
2997914-2998366	-	Raghavan	-	-
3065196←3065361	-	Raghavan	-	-
3107178-3107574	-	Raghavan	-	-
3198988-3199228	-	Raghavan	-	-
3208566-3208802	-	Raghavan	-	-
3219270-3219487	-	Raghavan	-	-
3352073-3352653	-	Raghavan	-	-
3365665→3365848	-	Raghavan	-	-
3372512-3372890	-	Raghavan	-	-
3376674←3376891	-	Raghavan	-	-
3399110←3399413	-	Raghavan	-	-
3401355-3401505	-	Raghavan	-	-
3403054→3403457	-	Raghavan	-	-
3426785-3427257	-	Raghavan	-	-
3451293←3451529	-	Raghavan	-	-
3474463←3474628	-	Raghavan	-	-
3483841-3484141	-	Raghavan	-	-
3530462-3530839	-	Raghavan	-	-
3572902-3573093	-	Raghavan	-	-
3595584-3596006	-	Raghavan	-	-
3598807←3599050	-	Raghavan	-	-
3628626←3628990	-	Raghavan	-	-
3637744-3638133	-	Raghavan	-	-
3645857-3646550	-	Raghavan	-	-
3655591→3656388	-	Raghavan	-	-
3667212-3667614	-	Raghavan	-	-
3735201-3735519	-	Raghavan	-	-
3833953-3834244	-	Raghavan	-	-
3851281-3851575	-	Raghavan	-	-
3865446-3865750	-	Raghavan	-	-
3881753-3882358	-	Raghavan	-	-
3939351-3939830	-	Raghavan	-	-
3964114→3964439	-	Raghavan	-	-
3984031-3984454	-	Raghavan	-	-
3999080-3999448	-	Raghavan	-	-
4083846-4084038	-	Raghavan	-	-
4124834-4125035	-	Raghavan	-	-
4152871-4153023	-	Raghavan	-	-
4205556-4206169	-	Raghavan	-	-
4254490-4254659	-	Raghavan	-	-
4258027-4258343	-	Raghavan	-	-
4273065-4273493	-	Raghavan	-	-
4328262→4328524	-	Raghavan	-	-
4366351-4366686	-	Raghavan	-	-
4380342-4380665	-	Raghavan	-	-
4422815-4423140	-	Raghavan	-	-
4437286-4437609	-	Raghavan	-	-

	4498815-4499282	-	Raghavan	-	-	
	4499613-4500125	-	Raghavan	-	-	
	4554344-4555015	-	Raghavan	-	-	
	4558704-4558952	-	Raghavan	-	-	
	4566543-4567020	-	Raghavan	-	-	
	4600882-4601499	-	Raghavan	-	-	
IS001	*nhaR-rpsT* 20666←20798	-	Raghavan	-	-	(Chen *et al.*, 2002)
IS009	*yagU-ykgJ* 302835→303034	-	Raghavan	-	-	(Chen *et al.*, 2002)
p14	*psiF-yaiC* 402834←402927	-	Raghavan	-	-	(Rivas *et al.*, 2001)
IS011	*cyoA-ampG* 450965→451296	-	Raghavan	-	-	(Chen *et al.*, 2002)
IS014	*ybaJ-acrB* 480028←480388	-	Raghavan	-	-	(Chen *et al.*, 2002)
IS018	*folD-sfmA* 573636→573747	-	Raghavan	-	-	(Chen *et al.*, 2002)
IS022	584964→585220	-	Raghavan	-	-	(Chen *et al.*, 2002)
IS023	623829←624038	-	Raghavan	-	-	(Chen *et al.*, 2002)
tp28	709961←710158	-	Raghavan	-	-	(Rivas *et al.*, 2001)
IS029	752270→752396	-	Raghavan	-	-	(Chen *et al.*, 2002; Yachie *et al.*, 2006)
c-8	770261→770500	-	Raghavan	-	-	
NC021	854993→855137	-	Raghavan	-	-	(Yachie *et al.*, 2006)
IS038	915522→915655	-	Raghavan	-	-	(Chen *et al.*, 2002)
tk5	921522→921589	-	Raghavan	-	-	(Rivas *et al.*, 2001)
IS042	1014724←1014910	-	Raghavan	-	-	(Chen *et al.*, 2002)
IS055	1223195→1223288	-	Raghavan	-	-	(Chen *et al.*, 2002)
tp21	1255837→1255942	-	Raghavan	-	-	(Rivas *et al.*, 2001; Yachie *et al.*, 2006)
NC044	1565327→1565515	-	Raghavan	-	-	(Yachie *et al.*, 2006)
NC048	1636282←1636395	-	Raghavan	-	-	(Yachie *et al.*, 2006)
tp29	1903368→1903613	-	Raghavan	-	-	(Rivas *et al.*, 2001)
te10	1976232→1976427	-	Raghavan	-	-	(Rivas *et al.*, 2001)
NC059	2040100→2040182	-	Raghavan	-	-	(Yachie *et al.*, 2006)
IS109	2190861→2190980	-	Raghavan	-	-	(Chen *et al.*, 2002)

k27	2720644→2720746	-	Raghavan	-	-	(Rivas *et al.*, 2001)
IS135	2759200→2759301	-	Raghavan	-	-	(Chen *et al.*, 2002)
NC087	2796832←2796960	-	Raghavan	-	-	(Yachie *et al.*, 2006)
IS143	2925871→2926018	-	Raghavan	-	-	(Chen *et al.*, 2002)
IS149	3048695←3048955	-	Raghavan	-	-	(Chen *et al.*, 2002)
tpke85	3081833→3081949	-	Raghavan	-	-	(Rivas *et al.*, 2001)
tpk16	3107239→3107413	-	Raghavan	-	-	(Rivas *et al.*, 2001)
te21	3326110←3326261	-	Raghavan	-	-	(Rivas *et al.*, 2001)
IS166	3358753←3358963	-	Raghavan	-	-	(Chen *et al.*, 2002)
IS173	3517316←3517406	-	Raghavan	-	-	(Chen *et al.*, 2002)
tp46	3679988→3680184	-	Raghavan	-	-	(Rivas *et al.*, 2001)
p1	4049881→4050062	-	Raghavan	-	-	(Rivas *et al.*, 2001)
tpk3	4056084→4056430	-	Raghavan	-	-	(Rivas *et al.*, 2001)
p2	4404011←4404207	-	Raghavan	-	-	(Rivas *et al.*, 2001)
tpke10	4531814→4532310	-	Raghavan	-	-	(Rivas *et al.*, 2001)
IS219	4501648→4501912	-	Raghavan	-	-	(Chen *et al.*, 2002)
ECS001	3382436←3382524	-	-	Shinhara	-	
ECS002	3645971←3646021	-	-	Shinhara	-	(Tjaden *et al.*, 2002)
ECS005	3655821→3655882	-	-	Shinhara	-	
ECS006	2111172←2111260	-	-	Shinhara	-	
ECS014	302907→303013	-	-	Shinhara	-	
ECS016	4484090←4484162	-	-	Shinhara	-	
ECS019	1702813←1702878	-	-	Shinhara	-	
ECS023	192615→192701	-	-	Shinhara	-	
ECS024	407797→407870	-	-	Shinhara	-	
ECS030	502571→502645	-	-	Shinhara	-	(Rivas *et al.*, 2001)
ECS032	2280462→2280520	-	-	Shinhara	-	
ECS035	1741399→1741453	-	-	Shinhara	-	
ECS037	1676332→1676390	-	-	Shinhara	-	
ECS039	89360→89417	-	-	Shinhara	-	
ECS042	1225703→1225791	-	-	Shinhara	-	
ECS044	2806199←2806250	-	-	Shinhara	-	(Tjaden *et al.*, 2002)

ECS046	3984318→3984372	-	-	Shinhara	-	(Wassarman *et al.*, 2001; Tjaden *et al.*, 2002; Yachie *et al.*, 2006)
ECS047	1332921←1332971	-	-	Shinhara	-	
ECS051	579779→579829	-	-	Shinhara	-	
ECS053	59499→59625	-	-	Shinhara	-	(Sætrom *et al.*, 2005)
ECS056	1752876→1752933	-	-	Shinhara	-	
ECS058	573755←573826	-	-	Shinhara	-	(Sætrom *et al.*, 2005)
ECS065	3826754←3826820	-	-	Shinhara	-	
ECS071	3388537←3388594	-	-	Shinhara	-	
ECS073	4530328←4530385	-	-	Shinhara	-	
ECS076	2006160←2006240	-	-	Shinhara	-	
ECS077	4372535←4372614	-	-	Shinhara	-	
ECS078	3524341→3524399	-	-	Shinhara	-	
ECS080	1432461←1432528	-	-	Shinhara	-	
ECS081	1848740←1848862	-	-	Shinhara	-	
ECS082	4626608→4626731	-	-	Shinhara	-	(Sætrom *et al.*, 2005)
ECS083	1066996←1067055	-	-	Shinhara	-	
ECS088	1433069←1433122	-	-	Shinhara	-	(Chen *et al.*, 2002)
ECS089	29303→29371	-	-	Shinhara	-	
ECS093	2536509→2536598	-	-	Shinhara	-	
ECS096	2915878→2915939	-	-	Shinhara	-	
ECS103	62←132	-	-	Shinhara	-	
ECS104	1494707←1494775	-	-	Shinhara	-	
ECS106	2042726←2042792	-	-	Shinhara	-	
ECS107	2763842←2763912	-	-	Shinhara	-	
ECS110	1475507→1475561	-	-	Shinhara	-	
ECS111	2902502←2902556	-	-	Shinhara	-	
ECS114	19625←19706	-	-	Shinhara	-	
ECS120	1630770←1630837	-	-	Shinhara	-	
ECS121	2967133→2967207	-	-	Shinhara	-	(Wassarman *et al.*, 2001)
ECS123	668396←668446	-	-	Shinhara	-	
ECS127	4116322→4116371	-	-	Shinhara	-	
ECS129	2036861←2036912	-	-	Shinhara	-	
ECS131	3683620←3683670	-	-	Shinhara	-	
ECS132	1554498→1554551	-	-	Shinhara	-	
ECS134	2111303→2111362	-	-	Shinhara	-	(Rivas *et al.*, 2001; Kawano *et al.*, 2005)
ECS137	4324995→4325051	-	-	Shinhara	-	
ECS138	4460897→4460952	-	-	Shinhara	-	
ECS143	2515931←2515989	-	-	Shinhara	-	

ECS152	2454057←2454119	-	-	Shinhara	-	
ECS157	1029155←1029221	-	-	Shinhara	-	
ECS160	2380661←2380730	-	-	Shinhara	-	
ECS161	1407150→1407271	-	-	Shinhara	-	(Wassarman *et al.*, 2001)
ECS164	192792→192850	-	-	Shinhara	-	
ECS165	1906151→1906214	-	-	Shinhara	-	
ECS166	986620→986799	-	-	Shinhara	-	(Wassarman *et al.*, 2001; Tjaden *et al.*, 2002)
ECS167	989733→989783	-	-	Shinhara	-	(Tjaden *et al.*, 2002)
ECS169	4148368→4148435	-	-	Shinhara	-	
ECS175	2763277←2763343	-	-	Shinhara	-	
ECS177	4457372→4457430	-	-	Shinhara	-	
ECS180	2404787←2404849	-	-	Shinhara	-	
ECS182	1481038→1481088	-	-	Shinhara	-	
ECS184	2404734←2404785	-	-	Shinhara	-	
ECS185	164658→164715	-	-	Shinhara	-	
ECS186	2576447←2576555	-	-	Shinhara	-	(Chen *et al.*, 2002; Tjaden *et al.*, 2002)
ECS188	3617059←3617187	-	-	Shinhara	-	
ECS191	1019493→1019559	-	-	Shinhara	-	(Tjaden *et al.*, 2002)
ECS192	1790135←1790188	-	-	Shinhara	-	
ECS194	3382362←3382419	-	-	Shinhara	-	(Sætrom *et al.*, 2005)
ECS195	3305974←3306035	-	-	Shinhara	-	
ECS196	2151914←2151987	-	-	Shinhara	-	
ECS200	2902008→2902058	-	-	Shinhara	-	(Sætrom *et al.*, 2005)
ECS205	2559006←2559079	-	-	Shinhara	-	(Sætrom *et al.*, 2005)
ECS206	2068543→2068617	-	-	Shinhara	-	
ECS212	2870925←2870997	-	-	Shinhara	-	
ECS219	4437418→4437471	-	-	Shinhara	-	
ECS228	1993524←1993590	-	-	Shinhara	-	
ECS229	4570223→4570288	-	-	Shinhara	-	
	2651730←2651798	-	-	-	Conway	

Table S8. Oligonucleotide probes used for northern blot analysis of putative small RNAs found in *E. coli* by RNA-seq.

sRNAs	Sequence (5' to 3')
p-sRNA-1 0174u+ (KMD-TSS-2)	CCACGGCATATCTGACCTTATAAAGCCAAC
p-sRNA-2 TSS-54	CCCTTTCAGGTGGATTACTTTTCTCAGGTC
p-sRNA-3 4051d- (KMD-TSS-86 HA04835541)	CTGTATGTAGGGTACAGCACGATGAATCTG
p-sRNA-4 TSS-81	CAGCCTTGTTTTCTCCCTCATTACTTGAAGG
p-sRNA-5 2775u+ (KMD-TSS-73)	CTCAAGGGGAGAAAACTTAGGGCCTCTATG
p-sRNA-6 1743d+ (KMD-TSS-30)	GCAGTTGCAGGCATTTTACCTTTTGCC
p-sRNA-7 0770d- KMD-TSS-18	GGATTCACTCGCTTGTCGCCTTCCTG
p-sRNA-8 TSS-40	ATTGTTCTGTTCTGTGCATCCCGTCAC TCCTGTCCTGTAACATTTGATGAGCGCAA
p-sRNA-9 TSS-41	CCCCTTCTCTTTTCTGCCTTTTTTAGCCAGG
p-sRNA-10 TSS-50	GTTATCTATGCTCCTGGGGCTTCACTCAC
IS019 (KMD-OC-3)	CAGCCAAGAGCCATGAATAGGATTCGATAG
G0-10698 (VK131)	TAGACGAAGCGCCAGCCAATTCCCGCAGCGCGTCTAGC GTCATCA
G0-10699 (VK76)	GGGGCTGGACGCCATTTCAAGCCTGATAAAACTGCTTAA CAAATCAGCATAACTCATTAA
p-sRNA-11 (KMD-OC-9)	GTTTTCTCCGGTGGCACGGTGTTACTCAG

ECS054 (KMD-TSS-47)	GGCGGTATGAGAAACAAGCTAAAGGAG
ECS174 p-sRNA 27 KM - TSS-48	AGCCCACATTCACTAAGGCTAACTTACTG
C0614 (KMD-TSS-49)	TCCAGCAGGTTGACTTGTGTTACATGAGCA
p-sRNA-12 3777d+	ACCTTTTCGGCTGTCTCTTCTCTCGTACTG CCCACCACAACCAGAGCAATCAATACAATC
ECS163 (VK111)	ACATTCAATGCCCCATTTGCGGGGCTAATTTCTTGTCGG AGTGCCTTAACTGG
G0-10700 (VK135)	GATGCCTCGCCTGCGTGGCCTCTATGGCTCTGATTTAAG GGGCTATATGTACCCGTTTCTGCGATCTAACTCAACCTT
G0-10701 (och5) (VK130)	CCTATCTTTGCTCGCGGTTTCATCAAGGATAGGCTGACT GAAACGG GCTAATGCCTTTACCCGATATTGCCTCC
G0-10705 (VK137)	TTGGTGCAACATTCACATCGTGGTGCAGCCCTT