

1 **An effort to make sense of antisense transcription in bacteria**

2 Iñigo Lasa<sup>a\*</sup>, Alejandro Toledo-Arana<sup>a</sup>, and Thomas R. Gingeras<sup>b</sup>

3

4 <sup>a</sup>Laboratory of Microbial Biofilms. Idab-CSIC-Universidad Pública de Navarra-Gobierno de Navarra.

5 Pamplona 31006, Spain; <sup>b</sup>Laboratory of Functional Genomics. Cold Spring Harbor Laboratory, Cold

6 Spring Harbor, New York, USA

7

8 Corresponding author: Laboratory of Microbial Biofilms. Idab-CSIC-Universidad Pública de Navarra-

9 Gobierno de Navarra. Pamplona 31006, Spain

10 Tel: 34 948 168007

11 Fax: 34 948 232191

12 Email: [ilasa@unavarra.es](mailto:ilasa@unavarra.es)

13

14 Keywords: overlapping transcription; RNase III; RNA processing; bacteria; transcriptome

15

16

17

1 **Abstract**

2 Analysis of bacterial transcriptomes have shown the existence of a genome-wide process  
3 of overlapping transcription due to the presence of antisense RNAs, as well as messenger  
4 RNAs that overlapped in their entire length or in some portion of the 5'- and 3'-UTR  
5 regions. The biological advantages of such overlapping transcription are unclear but may  
6 play important regulatory roles at the level of transcription, RNA stability and translation.  
7 In a recent report, the human pathogen *Staphylococcus aureus* is observed to generate  
8 genome-wide overlapping transcription in the same bacterial cells leading to a collection  
9 of short RNA fragments generated by the endoribonuclease III, RNaseIII. This processing  
10 appears most prominently in Gram-positive bacteria. The implications of both the use of  
11 pervasive overlapping transcription and the processing of these double stranded  
12 templates into short RNAs are explored and the consequences discussed.

13

## 1 **Pervasive overlapping transcription in bacteria**

2           Implementation of high-throughput RNA analysis techniques to the identification  
3 of the entire collection of RNA molecules (transcriptome) produced by a bacterial  
4 population has directed our view of RNA landscapes away from a protein-centric genome  
5 annotation. As with studies involving eukaryotic cells, the first bacterial transcriptomic  
6 studies also revealed the existence of a genome-wide process of overlapping transcription  
7 <sup>1,2,3,4,5,6,7,8,9,10,11,12,13,14,15</sup>. To be clear overlapping transcription is defined as a process that  
8 generates overlapping sense/antisense RNAs from a genomic region. The resulting RNA  
9 transcripts show perfect complementarity at least in some portion of the length of the  
10 overlapping RNAs. There are at least four different mechanisms to generate overlapping  
11 transcription in bacteria (Fig. 1): (i) *bona fide* antisense RNAs (asRNA), RNA molecules that  
12 do not encoded for proteins and show complementarity with part of a gene, a complete  
13 gene or a group of genes; (ii) 5' overlapping UTRs between mRNAs of contiguous genes  
14 (head-to-head) that are transcribed in divergent directions; (iii) 3' overlapping UTRs  
15 between mRNAs of contiguous genes transcribed in convergent directions (tail-to-tail). In  
16 this case, the overlapping process can be caused by read-through of transcriptional  
17 terminator, the presence of anti-terminator elements or the location of the  
18 transcriptional terminators inside the contiguous gene; and (iv) overlapping operons,  
19 genes that being located in the middle of an operon are transcribed in opposite direction  
20 to the rest of the operon <sup>12,16</sup>. In this definition of overlapping transcription, we exclude  
21 short transcripts that are encoded at genomic locations distant from the RNAs they  
22 regulate and sharing only limited complementarities with their targets, because they are  
23 not produced from complementary strands of the same DNA region.

24           The possibility that well understood technical artefacts had the potential to generate  
25 pseudo-antisense transcription data prompted initial criticism <sup>17</sup>. While these early  
26 uncertainties were partially justified and the signals detected from the antisense strand of

1 highly expressed genes in the earliest studies were in many cases due to the DNA-  
2 dependent DNA polymerase activity of the reverse transcriptase<sup>18</sup>, these technical issues  
3 have been solved in more recent studies. Examples of technical improvements included  
4 (Fig. 2): (i) cDNA synthesis performed in the presence of actinomycin D, which specifically  
5 inhibits the DNA-dependent DNA polymerase activity of the reverse transcriptase enzyme  
6<sup>18</sup>; (ii) enrichment of RNA samples for primary transcripts by use of terminator  
7 exonuclease treatment that degrades 5'P (processed transcript) but not 5'PPP (primary  
8 transcript) allowing for the identification of the transcript start and not the extent of the  
9 asRNA<sup>2</sup>; (iii) direct labeling of 3' and 5' ends of the RNA molecules with adapters before  
10 cDNA synthesis, preserving the strand orientation of each RNA molecule<sup>12,19</sup>; and (iv)  
11 synthesis of the cDNA second strand in the presence of dUTP, which allows the selective  
12 removal of the strand with UNG (Uracil-N-Glycosylase) after ligation of 5' and 3' adaptors  
13<sup>20</sup>. The introduction of these technical modifications in the cDNA synthesis and labeling  
14 methodologies far from contradicting initial observations have confirmed that overlapping  
15 transcription is a very common process. Thus, the percentage of genes that have been  
16 associated with at least one antisense transcript in recent studies ranged from 13% in  
17 *Bacillus subtilis*<sup>4</sup>, 27% in *Synechocystis* PCC6803<sup>10</sup>, 30% in *Anabaena*<sup>11</sup>, 46% in  
18 *Helicobacter pylori*<sup>2</sup>, to 49% in *Staphylococcus aureus*<sup>12</sup>. The first impression of these  
19 data is that strong differences in the extent of overlapping transcription between bacteria  
20 exist. However, it is uncertain whether these differences reflect real biological differences  
21 or a combination of both biological and methodological bias. The answer to this question  
22 would require comparative transcriptomic studies using standardized protocols and  
23 computational tools.

24 What mechanisms regulate the transcription of overlapping transcripts? The origin of the  
25 messenger RNAs participating in overlapping transcription involves promoters recognized  
26 by sigma factors. The few studies that have been done on bona-fide antisense transcripts  
27 reveal that they are transcribed from similar promoters as their sense counterparts<sup>3,4</sup>.

1 Therefore, it is conceivable that the sense and antisense transcripts are regulated via the  
2 same mechanisms. However, it is noteworthy that expression of antisense RNAs is for  
3 most genes lower than those of the corresponding sense transcript, suggesting that the  
4 promoters of antisense transcripts have evolved and modulated their strength to that of  
5 the sense transcript or that transcription process of sense and antisense transcripts is  
6 coordinated by unknown mechanisms.

### 7 **Role of RNase III in antisense regulation**

8 In a recent study devoted to analyzing the transcriptome of the human pathogen  
9 *Staphylococcus aureus* the total RNA sample was fractionated in long and short (<50  
10 nucleotides) RNA fractions. RNA sequencing of both fractions revealed a genome-wide  
11 process of overlapping sense/antisense RNA processing by the activity of double stranded  
12 endoribonuclease, RNase III <sup>12</sup>. The end products of the process are a collection of short  
13 RNA fragments (20 nucleotides on average) that accumulate in every genome region  
14 where overlapping transcription is detected. Given that short RNA fragments originate  
15 from the digestion of overlapping transcripts, the total amount of short RNA fragments is  
16 similar in both strands and is proportional to the amount of double stranded RNA  
17 molecules.

18 This process of overlapping RNA digestion and production of a collection of short  
19 RNA molecules that are symmetrically distributed in both strands of the annotated genes  
20 is not exclusive of *S. aureus* and it also occurs in different Gram-positive bacteria such as  
21 *Bacillus subtilis*, *Listeria monocytogenes* and *Enterococcus faecalis*. In contrast, analysis of  
22 the transcriptome of the Gram-negative bacteria *Salmonella enterica* ser. Enteritidis using  
23 the same approach could not identify the collection of short RNA fragments. The absence  
24 of this process in *Salmonella* supports that a technical artifact does not generate the  
25 collection of short RNAs during the preparation of the RNA libraries. Different reasons can  
26 be envisioned to explain why the complement of short RNA fragments is not detected in

1 *Salmonella*. It is possible that the size of the RNA fragments produced by RNase III enzyme  
2 of *Salmonella* are longer than 50 nucleotides, in which case the RNA fragments would be  
3 excluded from the RNA fraction used to prepare the short RNA libraries. Alternatively,  
4 overlapping transcripts might be processed by a different mechanism or the resulting  
5 short RNA molecules might be unstable in Gram-negative bacteria. Indeed, although  
6 fundamental principles govern RNA degradation in bacteria, significant differences have  
7 been also identified in the degradosome composition of Gram-positive and Gram-  
8 negative bacteria <sup>21,22,23</sup>. Analysis of the short RNA fraction of other Gram-negative  
9 bacteria as well as isogenic mutants in different RNases would aid to clarify whether the  
10 digestion of overlapping transcripts occurs through different mechanisms in both types of  
11 bacteria.

12 Irrespective of the length of the sense/antisense complementarity region, the  
13 formation of the RNA duplexes between overlapping transcripts have been shown to  
14 affect the final amount of the protein encoded by the sense RNA in different ways.  
15 Examples have been described in which sense-antisense duplex formation results in the  
16 sense RNA degradation by RNases such as RNase III and RNase E, an endoribonuclease  
17 that cleaves single strand RNA molecules, thus lowering the amount of translatable sense  
18 RNA <sup>24,25</sup>. Other interactions between overlapping transcripts have been shown to  
19 increase the amount of sense RNA coding protein since the duplex formation process  
20 protects the sense transcript from degradation or increases the likelihood that sense  
21 transcripts will be made at levels exceeding the amount degraded due to the formation of  
22 a double stranded substrate of RNase III <sup>26,27</sup>. Finally, the overlapping of sense and  
23 antisense transcripts can inhibit the binding of the sense transcript to the ribosome and  
24 translation process <sup>28</sup>.

25 **What is the role of the genome-wide overlapping transcription process?**

1           The presence of a stable collection of short RNA fragments derived from the  
2 digestion of overlapping transcripts by RNase III demonstrates that both overlapping  
3 transcripts are present at the same time in the same cell. The observation of co-  
4 expression in the same cell is informative because current transcriptomic studies are  
5 carried out with RNA purified from at least few millions of bacteria and without the RNase  
6 III results it is impossible to determine whether the expression of overlapping transcripts  
7 occurs in the same bacteria or it is mutually exclusive.

8           The question then arises as to what is the role of overlapping transcription and  
9 RNase III mediated digestion for bacterial gene regulation? Two possible alternatives may  
10 be considered to answer this question: short RNA fragments are residual non-functional  
11 products of the overlapping RNA digestion or such fragments are functional molecules  
12 with a specific role in gene regulation. With respect to the first possibility, our results  
13 support the hypothesis that overlapping transcription provides a simple mechanism to  
14 remove all those transcripts that are produced in response to transitory stimuli or escape  
15 the regular transcription repression process. For this purpose, the antisense transcript  
16 would establish the threshold level that the sense RNA have to reach in order to be  
17 translated, removing all the residual RNA molecules whose level are not enough to  
18 produce the minimal amount of protein required to be functional. It has been speculated  
19 that stochastic variations on transcriptional levels might be beneficial to enhance the  
20 phenotypic heterogeneity of the cells within a genetically uniform microbial population <sup>29</sup>.  
21 However, if the transcription initiation process is more leaky than expected and all mRNAs  
22 are indiscriminately translated into protein, then, the cytoplasm will accumulate hundreds  
23 of unintended proteins in insufficient amounts to achieve their function. The presence of  
24 these proteins would have adverse effects in a particular environmental condition.  
25 Alternatively, we cannot exclude the possibility that the RNA transcripts resulting from  
26 the RNaseIII-mediated digestion process could be more stable or translate more  
27 efficiently than the primary transcripts.

1           In the case of 5' and 3' overlapping UTRs, the consequences of the digestion could  
2 be different for the 5' divergent overlapping UTRs or for the 3' convergent overlapping  
3 UTRs. Regardless of the specific consequences, digestion of overlapping UTRs would allow  
4 coordination of the expression of neighboring genes. This finding is in line with the idea  
5 that distribution of bacterial genes within the genomes is not random <sup>30</sup>. Thus, to the  
6 deeply rooted concept that genes encoding proteins of the same metabolic pathway are  
7 clustered together (operons) to facilitate the regulation of their expression, such a second  
8 regulatory level coordinating the expression of adjacent transcription units should be  
9 considered when investigating bacterial gene regulation. Needless to say that overlapping  
10 transcription between adjacent genes also has important consequences when phenotypes  
11 associated with insertion or deletion mutants are investigated. Additional thoughts  
12 related with the function of the overlapping RNA digestion process are the binding  
13 kinetics between overlapping transcripts and the digestion rate of the RNA duplex.  
14 Extensive experimental efforts would be necessary to uncover how these factors affect to  
15 the overlapping RNA digestion process.

16           Concerning the possibility that short RNA fragments may fulfill a function by  
17 themselves, the average size of the RNA fragments generated by overlapping RNA  
18 digestion is 20-22 nucleotides depending on the bacterial species in which they are  
19 generated. The size and double stranded structure of the fragments is similar to that of  
20 the eukaryotic microRNAs (miRNAs). miRNAs are produced by the successive actions of  
21 two RNase III enzymes, Drosha and Dicer, in precursor RNA molecules. Following their  
22 processing, one strand of the miRNAs is loaded into a ribonucleoprotein complexes, which  
23 key component is the Argonaute (AGO) protein. Then, miRNA-AGO complex interact with  
24 their mRNA target by based pairing and direct the inactivation of target RNAs by mRNA  
25 degradation or translational arrest and heterochromatin formation <sup>31,32,33,34</sup>.



1           The existence of a miRNA-based regulatory mechanism in prokaryotic cells was  
2 not considered due to the absence of the required machinery to generate the miRNAs and  
3 more importantly to the absence of argonaute-like proteins. However, very recently a  
4 highly conserved protein (SMc01113/YbeY) sharing structural homology with the MID  
5 domain of the Argonaute protein has been described <sup>35</sup>. YbeY protein is required for  
6 maturation of bacterial 5S, 16S and 23S ribosomal RNAs and it seems to facilitate the  
7 establishment of interactions between small RNA and the mRNA targets, in a similar way  
8 to Hfq protein <sup>35,36</sup>. Furthermore, structural and docking analysis suggests that YbeY could  
9 contribute catalytically, like an RNase, to RNA cleavage after binding to a guide RNA. Thus,  
10 it is tempting to speculate that similarly to what happens in eukaryotes, binding of one  
11 strand of the short RNA to the MID domain of YbeY protein can facilitate the mRNA target  
12 recognition and subsequently affect the mRNA stability or translation efficiency. YbeY and  
13 RNase III are both required for correct maturation of ribosomal RNAs and double knock-  
14 out of both genes encoding these proteins caused a strong defect on bacterial growth <sup>36</sup>.  
15 This overlap in function and sharing of components between RNA processing and  
16 ribosomal RNA maturation constitute an additional difficulty for understanding the  
17 biological relevance of short RNA molecules because it is puzzling to distinguish whether  
18 phenotypes associated to the absence of these enzymes are due to defects on ribosomal  
19 maturation or functions related with post-transcriptional gene regulation. A definitive  
20 strategy to answer this question and demonstrate the functionality of the short RNAs  
21 would require the depletion of the short RNA pool or trans-complementation of the  
22 bacterial cell with a collection of short RNA molecules, two challenging approaches that  
23 warrants methodological developments.

#### 24 **Perspectives and unresolved issues associated with sense and antisense transcription**

25 The mechanisms through which overlapping transcription can affect sense RNA  
26 expression are diverse and are thought to be primarily based on direct interactions

1 between sense/antisense transcripts. It is intriguing to consider what determines the  
2 accessibility of the sense/antisense RNA duplex to RNase III, or why all RNA duplexes are  
3 not degraded by RNase III-like activities.

4           If sense and antisense transcripts are being transcribed at the same locus within  
5 the same cell it is possible that transcriptional interference may play a role in their  
6 regulation <sup>37,38,39</sup>. Several mechanisms have been proposed for transcriptional  
7 interference including the collision between both RNA polymerase complexes and  
8 removal of the transcription initiation complex by the continuous passing of the  
9 elongation RNA polymerase complex in the opposite strand have been proposed. In  
10 addition, the fate of both sense and antisense transcripts depend upon several factors  
11 including: the availability of complementary sequences to interact given that nascent  
12 RNAs are immediately bound and coated with a variety of RNA binding proteins, the  
13 affinity with which both RNA molecules will interact depending on the length of  
14 complementarity of the molecules, and the local and global folding predictions that may  
15 further decide the possible annealing fates of both RNAs.

16           The first insights into the enzymes involved in regulation of overlapping  
17 transcription and the function of this conserved biological process are emerging. Due to  
18 their simplicity and feasibility for genetic manipulation, investigations with bacteria can  
19 provide clues to understanding of the function of overlapping transcription in eukaryotic  
20 cells. However, in order to fulfil this mission at least two methodological difficulties  
21 associated with the particularity of overlapping transcription process needs to be solved.  
22 One difficulty inherent to the double stranded DNA structure is how to genetically  
23 manipulate one of the strands without perturbing the expression of the complementary  
24 overlapping strand. The second difficulty is the necessity of evaluating the results of the  
25 experiments at single cell level, which implies development of specific reporter tools.

1           Finally, with reference to many topics it is often said that “size matters”. An  
2 important lesson that emerges from these studies is that bacterial short RNA fraction  
3 deserves much more attention than has been paid to date and only the combination of  
4 long and short RNA fractions together with complementary sequencing strategies, as it is  
5 shown in figure 2C, can provide the complete and accurate landscape of bacterial  
6 transcriptomes.

7

## 8 **Acknowledgments**

9 A.T.-A. is recipient of “Ramon y Cajal” contracts from the Spanish Ministry of Science and  
10 Innovation. This research was supported by grants ERA-NET Pathogenomics  
11 (PIM2010EPA-00606), BIO2008-05284-C02, BIO2011-30503-C02 and BFU2011-23222 from  
12 Spanish Ministry of Economy and Competitiveness.

13

14

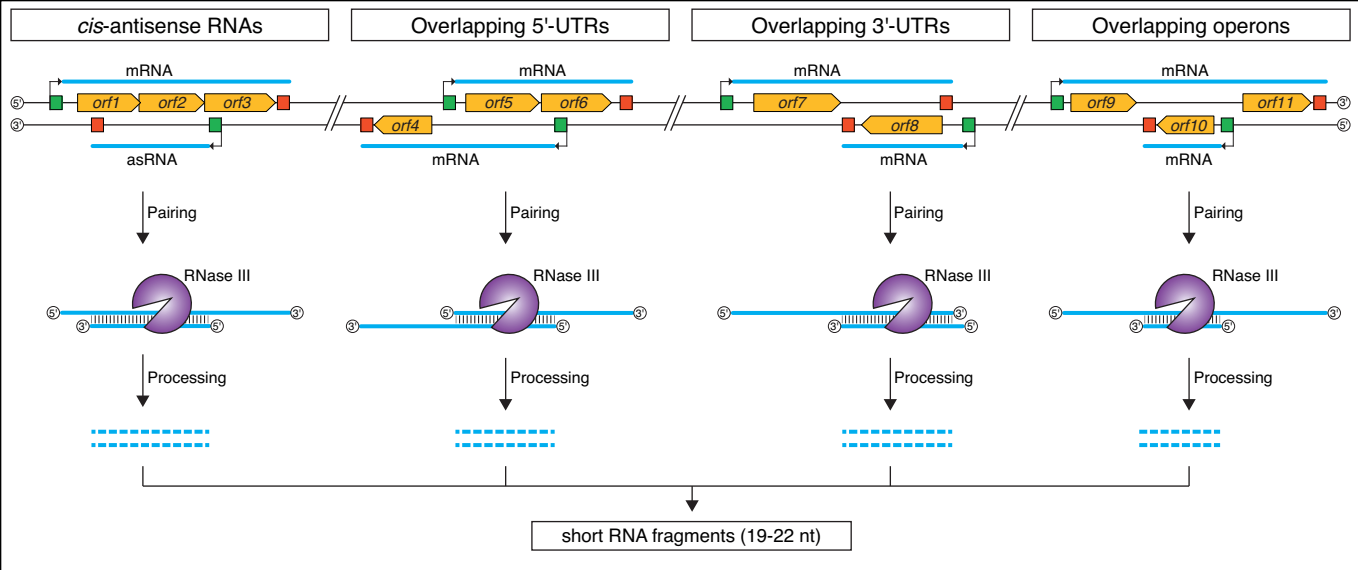
## 15 **References**

- 16 1. Güell M, van Noort V, Yus E, Chen W-H, Leigh-Bell J, Michalodimitrakis K, *et al.*  
17 Transcriptome complexity in a genome-reduced bacterium. *Science* 2009; 326:1268–  
18 1271.
- 19 2. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, *et al.* The primary  
20 transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* 2010; 464:250–  
21 255.
- 22 3. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R. A single-base resolution map  
23 of an archaeal transcriptome. *Genome Res* 2010; 20:133–141.
- 24 4. Nicolas P, Mäder U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, *et al.* Condition-  
25 dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*.  
26 *Science* 2012; 335:1103–1106.
- 27 5. Mendoza-Vargas A, Olvera L, Olvera M, Grande R, Vega-Alvarado L, Taboada B, *et al.*

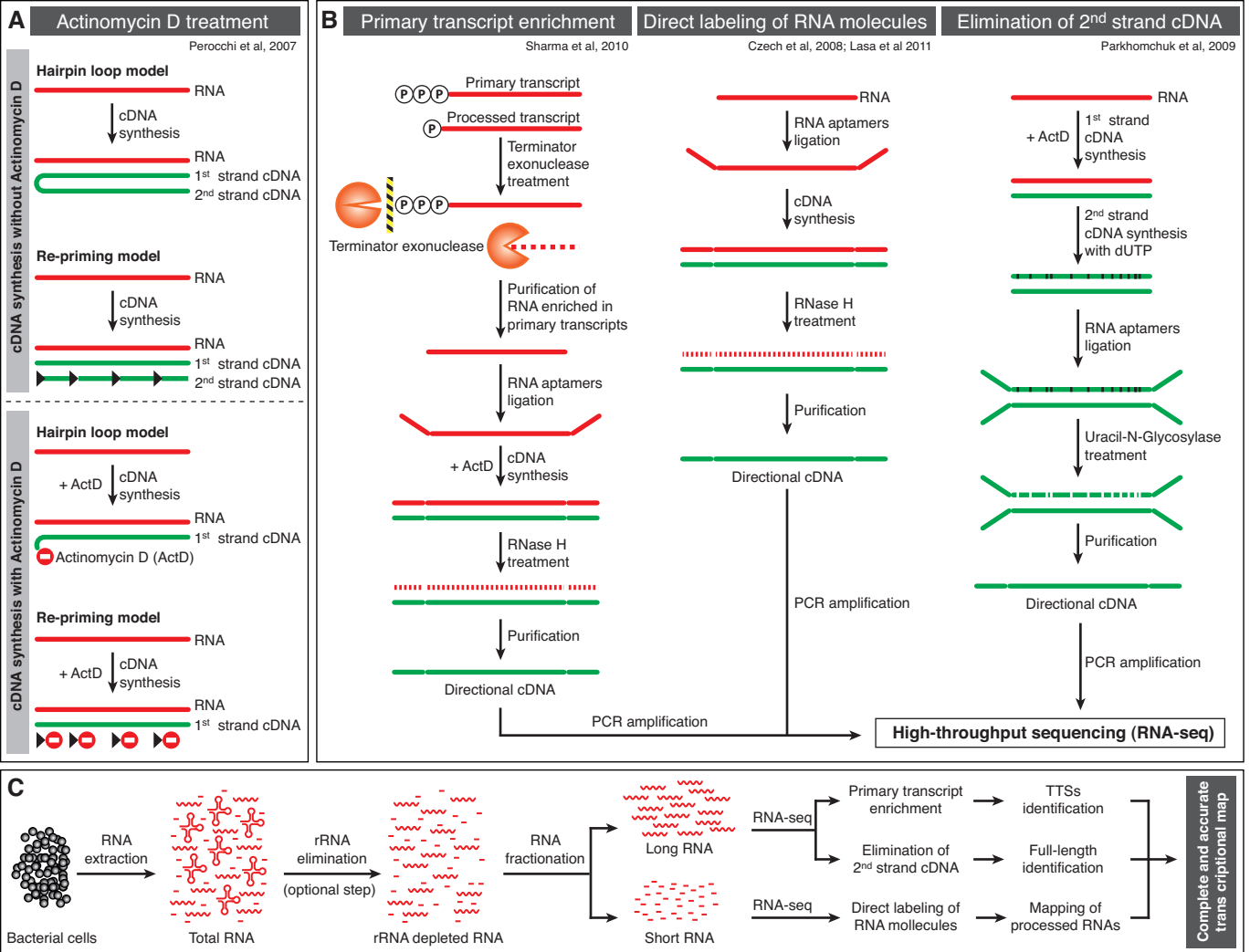
- 1 Genome-wide identification of transcription start sites, promoters and transcription factor  
2 binding sites in *E. coli*. PLoS ONE 2009; 4:e7526.
- 3 6. Dornenburg JE, Devita AM, Palumbo MJ, Wade JT. Widespread antisense transcription  
4 in *Escherichia coli*. mBio 2010; 1:
- 5 7. Filiatrault MJ, Stodghill PV, Bronstein PA, Moll S, Lindeberg M, Grills G, *et al.*  
6 Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs,  
7 and antisense activity. J Bacteriol 2010; 192:2359–2372.
- 8 8. Vijayan V, Jain IH, O'Shea EK. A high resolution map of a cyanobacterial transcriptome.  
9 Genome Biol 2011; 12:R47.
- 10 9. Schmidtke C, Findeiss S, Sharma CM, Kuhfuß J, Hoffmann S, Vogel J, *et al.* Genome-wide  
11 transcriptome analysis of the plant pathogen *Xanthomonas* identifies sRNAs with putative  
12 virulence functions. Nucleic Acids Res 2012; 40:2020–2031.
- 13 10. Mitschke J, Georg J, Scholz I, Sharma CM, Dienst D, Bantscheff J, *et al.* An  
14 experimentally anchored map of transcriptional start sites in the model cyanobacterium  
15 *Synechocystis* sp. PCC6803. Proceedings of the National Academy of Sciences 2011;  
16 108:2124–2129.
- 17 11. Mitschke J, Vioque A, Haas F, Hess WR, Muro-Pastor AM. Dynamics of transcriptional  
18 start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp.  
19 PCC7120. Proceedings of the National Academy of Sciences 2011; 108:20130–20135.
- 20 12. Lasa I, Toledo-Arana A, Dobin A, Villanueva M, de Los Mozos IR, Vergara-Irigaray M, *et*  
21 *al.* Genome-wide antisense transcription drives mRNA processing in bacteria. Proceedings  
22 of the National Academy of Sciences 2011; 108:20172–20177.
- 23 13. Wilms I, Overlöper A, Nowrousian M, Sharma CM, Narberhaus F. Deep sequencing  
24 uncovers numerous small RNAs on all four replicons of the plant pathogen *Agrobacterium*  
25 *tumefaciens*. RNA biology 2012; 9:
- 26 14. Høvik H, Yu W-H, Olsen I, Chen T. Comprehensive transcriptome analysis of the  
27 periodontopathogenic bacterium *Porphyromonas gingivalis* W83. J Bacteriol 2012;  
28 194:100–114.
- 29 15. Georg J, Voss B, Scholz I, Mitschke J, Wilde A, Hess WR. Evidence for a major role of  
30 antisense RNAs in cyanobacterial gene regulation. Molecular Systems Biology 2009; 5:305.
- 31 16. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, *et al.*  
32 The *Listeria* transcriptional landscape from saprophytism to virulence. Nature 2009;  
33 459:950–956.
- 34 17. Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome:  
35 evidence of widespread transcription detected by microarray tiling experiments. Trends  
36 Genet 2005; 21:93–102.
- 37 18. Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM. Antisense artifacts in

- 1 transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res*  
2 2007; 35:e128.
- 3 19. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, *et al.* An endogenous  
4 small interfering RNA pathway in *Drosophila*. *Nature* 2008; 453:798–802.
- 5 20. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, *et al.*  
6 Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic*  
7 *Acids Res* 2009; 37:e123.
- 8 21. Arraiano CM, Andrade JM, Domingues S, Guinote IB, Malecki M, Matos RG, *et al.* The  
9 critical role of RNA processing and degradation in the control of gene expression. *FEMS*  
10 *Microbiology Reviews* 2010; 34:883–923.
- 11 22. Carpousis AJ. The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine  
12 assembled on RNase E. *Annu Rev Microbiol* 2007; 61:71–87.
- 13 23. Condon C. Maturation and degradation of RNA in bacteria. *Current Opinion in*  
14 *Microbiology* 2007; 10:271–278.
- 15 24. Krinke L, Wulff DL. RNase III-dependent hydrolysis of lambda *cII-O* gene mRNA  
16 mediated by lambda OOP antisense RNA. *Genes & Development* 1990; 4:2223–2233.
- 17 25. Lee E-J, Groisman EA. An antisense RNA that governs the expression kinetics of a  
18 multifunctional virulence gene. *Mol Microbiol* 2010; 76:1020–1033.
- 19 26. Storz G, Vogel J, Wassarman KM. Regulation by Small RNAs in Bacteria: Expanding  
20 Frontiers. *Molecular Cell* 2011; 43:880–891.
- 21 27. Stazic D, Lindell D, Steglich C. Antisense RNA protects mRNA from RNase E  
22 degradation by RNA-RNA duplex formation during phage infection. *Nucleic Acids Res*  
23 2011; 39:4890–4899.
- 24 28. Kawano M, Aravind L, Storz G. An antisense RNA controls synthesis of an SOS-induced  
25 toxin evolved from an antitoxin. *Mol Microbiol* 2007; 64:738–754.
- 26 29. Dubnau D, Losick R. Bistability in bacteria. *Mol Microbiol* 2006; 61:564–572.
- 27 30. Lawrence JG. Gene organization: selection, selfishness, and serendipity. *Annu Rev*  
28 *Microbiol* 2003; 57:419–440.
- 29 31. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional  
30 regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 2008; 9:102–114.
- 31 32. Moazed D. Small RNAs in transcriptional gene silencing and genome defence. *Nature*  
32 2009; 457:413–420.
- 33 33. Cernilogar FM, Onorati MC, Kothe GO, Burroughs AM, Parsi KM, Breiling A, *et al.*  
34 Chromatin-associated RNA interference components contribute to transcriptional  
35 regulation in *Drosophila*. *Nature* 2011; 480:391–395.

- 1 34. Nakama M, Kawakami K, Kajitani T, Urano T, Murakami Y. DNA-RNA hybrid formation  
2 mediates RNAi-directed heterochromatin formation. *Genes to Cells* 2012; 17:218–233.
- 3 35. Pandey SP, Minesinger BK, Kumar J, Walker GC. A highly conserved protein of  
4 unknown function in *Sinorhizobium meliloti* affects sRNA regulation similar to Hfq. *Nucleic  
5 Acids Res* 2011; 39:4691–4708.
- 6 36. Davies BW, Köhrer C, Jacob AI, Simmons LA, Zhu J, Aleman LM, *et al.* Role of  
7 *Escherichia coli* YbeY, a highly conserved protein, in rRNA processing. *Mol Microbiol* 2010;  
8 78:506–518.
- 9 37. Shearwin KE, Callen BP, Egan JB. Transcriptional interference--a crash course. *Trends  
10 Genet* 2005; 21:339–345.
- 11 38. Thomason MK, Storz G. Bacterial antisense RNAs: how many are there, and what are  
12 they doing? *Annual review of genetics* 2010; 44:167–188.
- 13 39. Georg J, Hess WR. *cis*-antisense RNA, another level of gene regulation in bacteria.  
14 *Microbiol Mol Biol Rev* 2011; 75:286–300.
- 15



**Fig. 1 Processing of different types of overlapping transcripts by RNase III.** Schematic representation of examples of different type of overlapping transcripts in bacteria. These include bona fide antisense RNAs (asRNAs), overlapping 5' and 3' untranslated regions (UTRs) of messenger RNA (mRNA) and overlapping operons. The sense/antisense RNA duplex are processed by RNase III to short RNA fragments (average of 20 nucleotides) that accumulate in similar amounts in both strands of all genome regions where these types of overlapping transcription are taking place.



**Fig. 2 Flowchart illustrating technical improvements used to preserve the polarity of RNA molecules for transcriptome analysis.** (A) Hypothetical model proposed by Perocchi et al<sup>18</sup> showing how spurious second-strand synthesis can occur during reverse transcription and the mode of inhibition by Actinomycin D. During first-strand cDNA synthesis from RNA molecules by reverse transcription, unintended second-strand cDNA synthesis could occur using the first-strand cDNA as a template. (B) Methods to preserve the polarity of RNA molecules during preparation of libraries for RNA sequencing. Primary transcript enrichment: 5'-monophosphate dependent terminator exonuclease (TEX) specifically degrades RNAs with 5' monophosphates (5' P), while primary transcripts with a 5' triphosphate (5' PPP) or RNA with other termini are protected. Direct labelling of RNA molecules: RNA fraction is ligated to a linker in the 3' end. After removal of non-ligated oligonucleotide, the RNA is ligated to 5' RNA adapter by using T4 RNA ligase. For the first strand synthesis of cDNA, an oligonucleotide complementary to the 3' linker is used<sup>12, 19</sup>. Removal of second strand cDNA: After the first strand cDNA synthesis non-incorporated nucleotides are removed and dTTP is substituted by dUTP during the synthesis of the second strand. After ligation with a Y-shaped adaptor, the dUTP-containing strand is selectively removed with UNG (Uracil-N-Glycosylase), leaving the first cDNA strand intact<sup>20</sup>. (C) Summary flowchart suggesting an experimental design to define a complete and accurate transcriptional map.