Divergently Transcribed ncRNAs in *Escherichia coli*: Refinement of the Transcription Starts Assumes Functional Diversification

Sergey Kiselev, Natalia Markelova and Irina Masulis *

Department of Functional Genomics and Cellular Stress, Institute of Cell Biophysics Russian Academy of Sciences, Pushchino, Russia

Non-coding regulatory RNAs (ncRNAs) comprise specialized group of essential genetically encoded biological molecules involved in the wide variety of cellular metabolic processes. The progressive increase in the number of newly identified ncRNAs and the defining of their genome location indicate their predominant nesting in intergenic regions and expression under the control of their own regulatory elements. At the same time, the regulation of ncRNA's transcription cannot be considered in isolation from the processes occurring in the immediate genetic environment. A number of experimental data indicate the notable impact of positional regulation of gene expression mediated by dynamic temporal DNA rearrangements accompanying transcription events in the vicinity of neighboring genes. This issue can be perceived as particularly significant for divergently transcribed ncRNAs being actually subjected to double regulatory pressure. Based on available results of RNAseq experiments for *Escherichia coli*, we screened out divergent ncRNAs and the adjacent genes for the exact positions of transcription start sites (TSSs) and relative efficiency of RNA production. This analysis revealed extension or shortening of some previously annotated ncRNAs resulting in modified secondary structure, confirmed stable expression of four ncRNAs annotated earlier as putative, and approved the possibility of expression of divergently transcribed ncRNAs containing repetitive extragenic palindromic (REP) elements. The biogenesis of secreted ncRNAs from divergently transcribed *ffs*, *chiX*, *ral*A, and *ryh*B is discussed taking into account positions of TSSs. Refinement of TSSs for the neighboring genes renders some ncRNAs as true antisense overlapping with 5′UTR of divergently transcribed mRNAs.

Keywords: divergent transcription, non-coding RNAs, genome annotation, REP elements, secreted RNAs

INTRODUCTION

The generally accepted classification of non-coding regulatory RNAs (ncRNAs) is primarily based on their genomic positioning relatively to the target gene subjected to regulation. Thus *cis-* and *trans-* acting subtypes are described (reviewed in Waters and Storz, 2009), implying that the first control the expression of closely located gene (primarily via antisense mechanism and base pairing protruding over 43 to 650 bp) (Brantl, 2007) while *trans*-encoded sRNAs are confirmed to have from one to dozens remote targets detected by means of biochemical and bioinformatics approaches (Vogel and

Wagner, 2007; Hör et al., 2020). Assignment of ncRNAs as powerful and numerous class of metabolic regulators stimulated a flow of works aimed to discover and comprehensively characterize physiologically relevant products. For a long time, the dominated strategies exploited 1) the search of conservative intergenic regions (IGRs), 2) the presence of transcription initiation and termination signals, and 3) the ability of potential ncRNA to adopt stable secondary structure (Altuvia, 2007; Backofen and Hess, 2010). Predicted candidates were experimentally verified using hybridization and reverse transcription approaches. This allowed revealing major ncRNAs as master regulators with defined biological function (Argaman et al., 2001; Dutta and Srivastava, 2018). Selective cloning strategy aimed to hound down small ncRNAs (sRNAs) (Kawano et al., 2005) and high throughput sequence facilities significantly enriching ncRNAs community (Shinhara et al., 2011; Barquist and Vogel, 2015) including independently encoded genes and ncRNAs, originated by cutting-out from mRNAs. In *E. coli*, the observation of non-uniformity in decay rate of entire mRNAs and their internal fragments led to the finding of 58 novel putative ncRNAs partially or completely embedded into the coding regions (Dar and Sorek, 2018). Although recently described ways to produce ncRNAs from 5′ and 3′ untranslated regions of coding sequences contribute to the whole spectrum of riboregulators (Carrier et al., 2018), their transcription from intergenic regions remains one of the primary sources (Adams and Storz, 2020). The mode of ncRNA's biogenesis—from IGRs by means of transcription from specified promoter or from coding genomic region and adjacent loci mediated by RNAse E—implies the involvement of different molecular mechanisms in the regulation of their abundance. ncRNA excised from mRNA precursor follows the pathways implicated in the control of parental RNA synthesis, although additional level of regulation may be imposed by accessibility of distinct structural motifs to RNAses. IGR-derived ncRNA possesses individual control elements including promoter, terminator, and transcription factor binding sites. ncRNAs encoded within IGRs undergo more strict purifying selection if compared to non-coding regions containing promoters or terminators only (Thorpe et al., 2017). Genome-wide analyses encompassing near 900 sRNAs from 13 bacterial species revealed meaningful enrichment of long conserved IGRs by non-coding sRNAs (Tsai et al., 2015). Since these sRNAs may be considered as "self-sufficient" genetic elements, one might expect their conservation between different species and evolutionary distant taxonomic bacterial groups. Nevertheless, comparative analysis highlighted substantial deviation in phylogenetic distances calculated for sRNAs and for protein-coding sequences (Lindgreen et al., 2014). Thus, ncRNAs undergo their own way of emergency and evolution still remaining one of the most obscure matters. It was proposed that promoters for ncRNAs could arise from transcriptional noise in terms of new promoter spontaneously erupting in IGR and conferring beneficial properties (Jose et al., 2019). Such RNAs must go "trial and error" path for the adaptation to the regulatory networks of the cell being probed for the compatibility with different targets among mRNAs and

RNA-binding proteins. ncRNA's implementation in metabolic circuits may occur as coevolution with RNA-binding proteins (ChiX, CyaR, FnrS, MicA, SgrS, DsrA, and MicC) or like prevenient gain of novel ncRNAs searching for a target (ArcZ, MicF, OmrA, OxyS, and RprA) (Peer and Margalit, 2014). The lack of the initial destination for the interaction with a specific partner may partially explain the phenomenon of "multi-target" action of many *trans*-encoded ncRNAs. Other possible mechanisms giving rise to the novel ncRNAs may be horizontal gene transfer and duplication of preexisting ncRNAs followed by the acquisition of new functions (Dutcher and Raghavan 2018). Preferential location of ncRNAs in IGRs assumes that, apart from involvement of their own regulatory elements, transcription of neighboring genes may contribute to the maintenance of their expression at appropriate level. Using *lac* promoter-based reporter system (Bryant et al., 2014) or barcoded construct (Scholz et al., 2019), it was demonstrated that transcriptional output varies in the range of 300- or 20-fold, respectively, depending on the local genetic environment. This positional regulation via mechanistic properties of given DNA segment also may affect the "decision" for certain RNA to be expressed from the certain chromosomal region. The influence of genomic context is expected to be quite different for RNAs having collinear or divergent orientation, respectively, to the adjacent gene accounting concomitant conformational effects induced by transcription (**Figure 1**). The model of twine-supercoiling domains postulates the formation of partially unwound DNA regions behind transcription complexes and accumulation of positive supercoils on the front line (Wu et al., 1988). Since initiation of transcription was shown to be sensitive to DNA superhelical state (Lim et al., 2003; Peter et al., 2004; Zhi et al., 2017; Dorman, 2019), the expression of closely located transcription units in fact may be considered as context-encoded regulatory level implicated in the control of ncRNAs synthesis. Gene expression profiling revealed altered response of divergent genes to the global changes in superhelical density induced by DNA gyrase inhibitor norfloxacin indicating the combined effect of local and global DNA conformational changes on transcription (Masulis et al., 2015). In this respect for the case of divergently transcribed ncRNAs, the relative positions of promoter regions and the rate of their utilization by RNAP appear to be critical for the final transcriptional output. As for any pair of divergent transcriptional units, transcript level for each counterpart is expected to be a result of mutual effects mediated by competition for enzyme binding or by means of non-contact structural DNA deformation or by both mechanisms concurrently. Here, we addressed if codirected or divergent positioning of known ncRNAs has any preference and if there is any bias—antagonistic, synergistic, or neutral—between the expression of ncRNAs and their divergent partners.
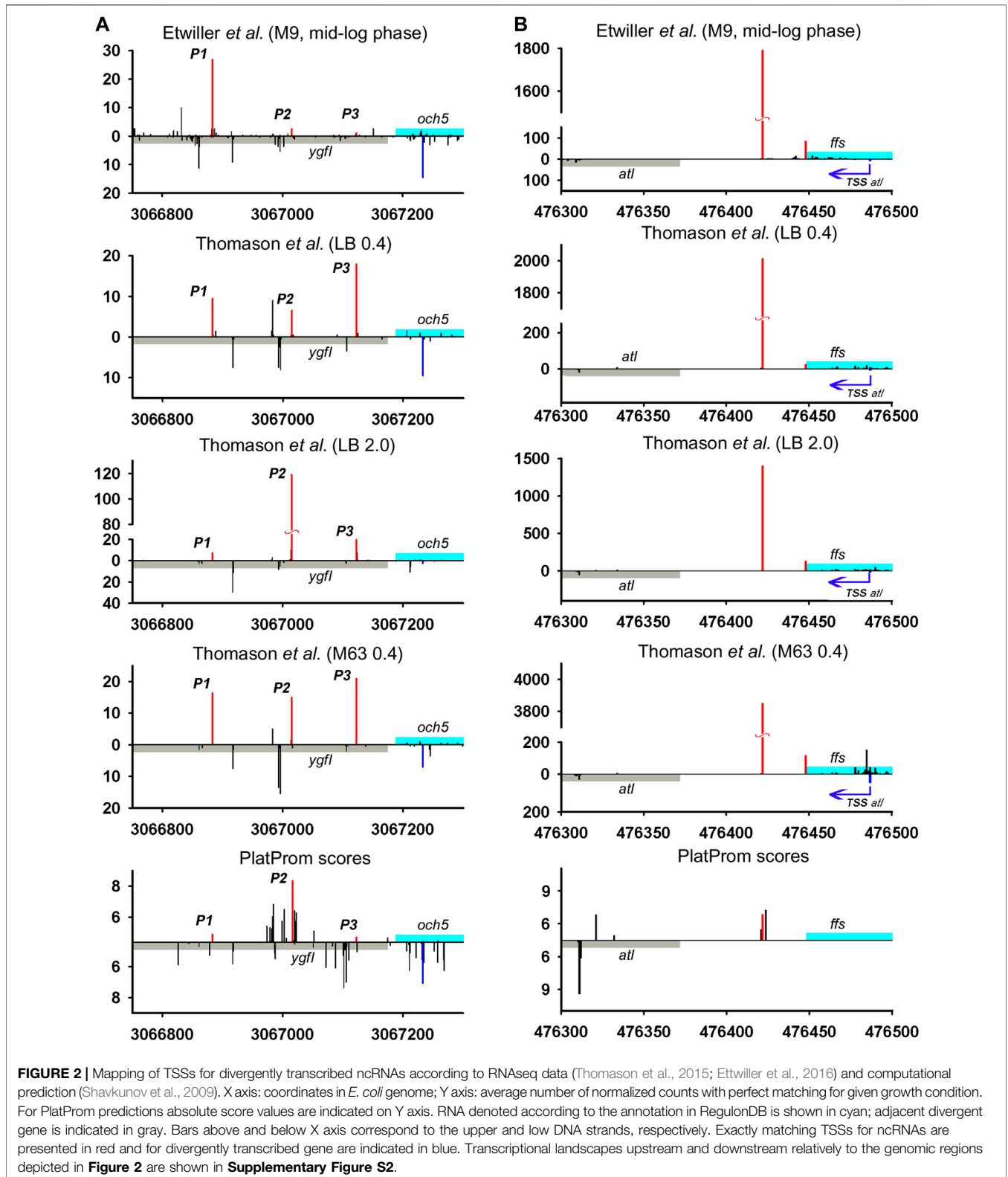
Totally, 125 genes with established or proposed assignment to non-coding small RNAs are annotated in RegulonDB to date (Santos-Zavaleta et al., 2019). 12 of them overlap with 5′ or 3′ terminal parts of protein-coding sequences assuming that transcription of these RNAs *a priori* is superimposed with read-through of parental mRNA. Based on annotation, 61

FIGURE 1 | Schematic presentation of various genomic contexts for ncRNAs annotated as separate transcription units. 5′ end of ncRNA assumed to coincide with transcription start site (TSS). "Stressed DNA" and positions of "transcription bubbles" illustrate topology-mediated effects.
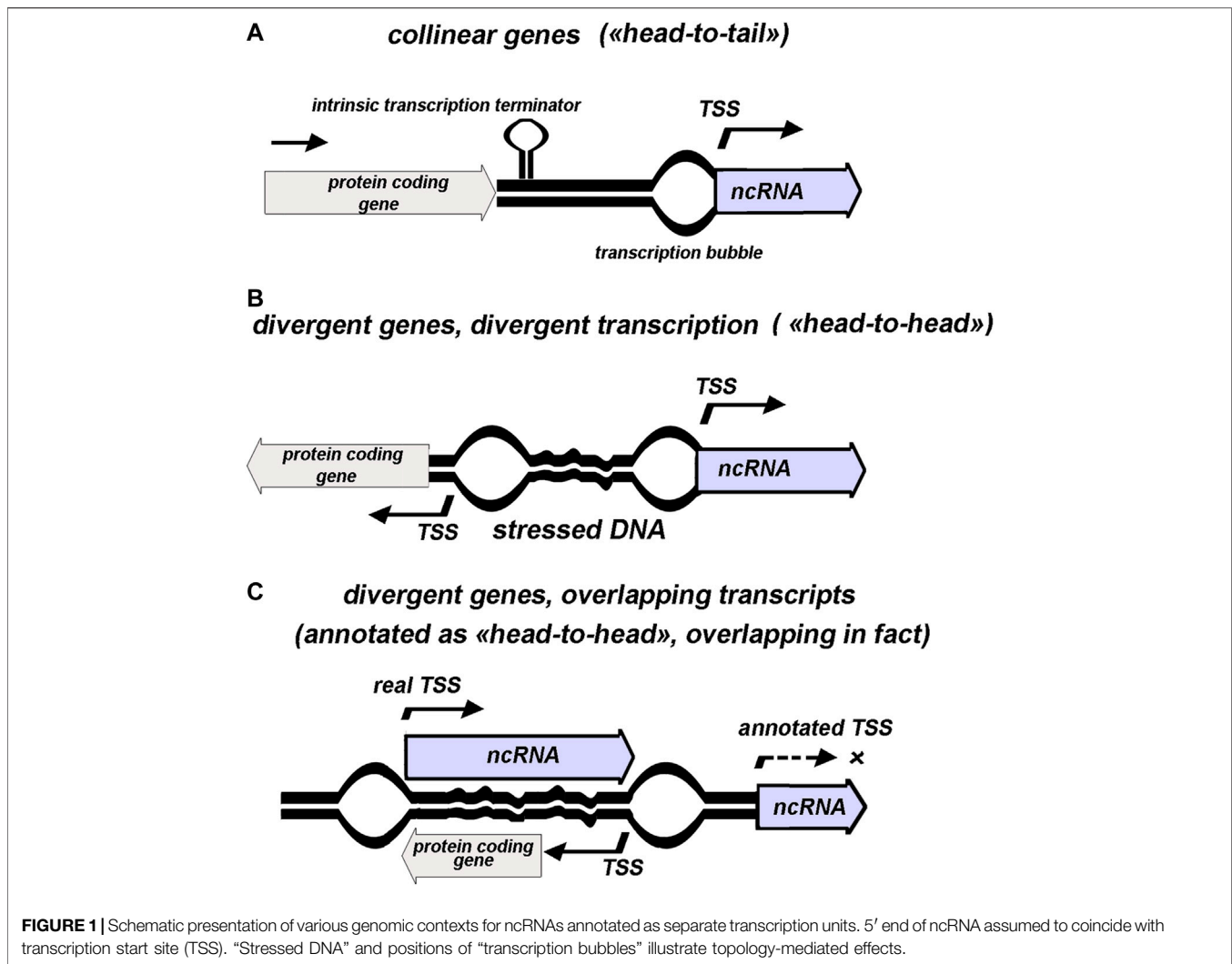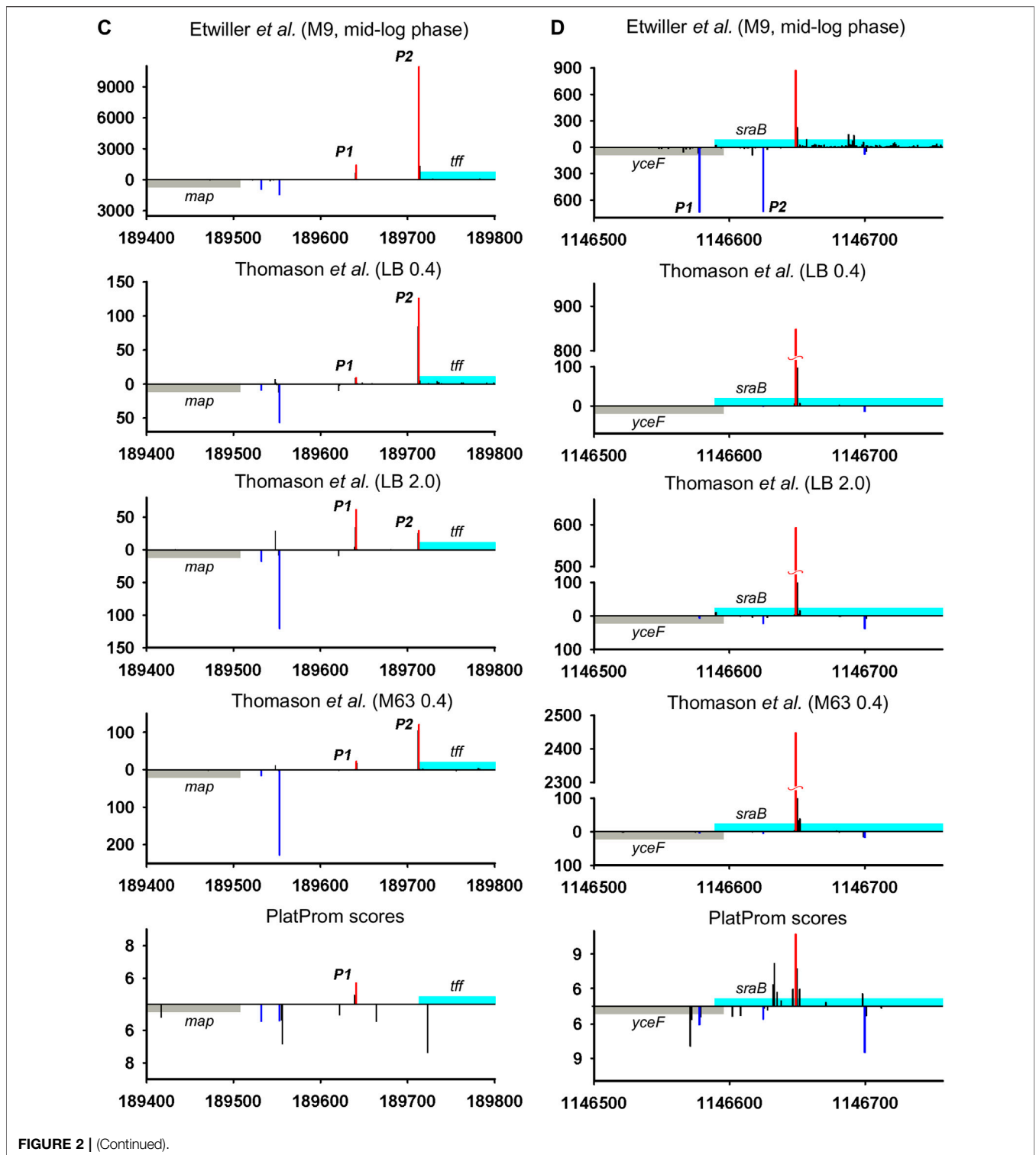
codirected and 64 divergently transcribed ncRNAs are revealed in *E. coli* genome demonstrating the lack of predominant orientation of ncRNAs relatively to the adjacent genes—"head-to-tail", "head-to-head", or overlapping over 5′ termini (**Figure 1**). Unambiguous attribution of ncRNA as an independent transcriptional unit and its assignment to one of the three types of relative arrangement with respect to the neighboring gene are fraught with some deceptions arising from incompleteness of our knowledge regarding the exact positions of promoters and terminators (**Figure 1C**). Thus, meta-analysis of data obtained using high-throughput RNAseq and tiling arrays allowed defining true *bona fide* small RNA as a product of gene not overlapping any other genes from the opposite strand (Liu et al., 2018). According to these requirements, antisense RNAs, UTR-processed products, and transcripts derived from premature termination (riboswitch) were sifted out from almost 500 previously proposed staphylococcal ncRNAs resulting in only 46 *bona fide* representatives with high degree of conservation within *Staphylococcaceae* family. Thus, the displacement of TSS into

the coding region of the neighboring gene compels to reconsider categorization of ncRNA as *cis*-acting or *trans*-acting. Some cases of this type revealed for *E. coli* ncRNAs are discussed below.

We juxtaposed the location of TSSs predicted by promoter-search algorithm PlatProm (Shavkunov et al., 2009) and experimentally found 5′ termini of ncRNAs and divergently transcribed neighboring genes of *E. coli* using data sets available from Thomason et al. (2015) and Ettwiller et al. (2016). PlatProm was developed to search for potential TSSs accounting sequence features of experimentally characterized *E. coli* promoters. It evaluates each position in the genome for the ability to serve as transcription start assuming essential promoter textual elements formalized in weight matrixes. The score value reflecting similarity with "ideal" promoter is considered as the quantitative measure of transcriptional capacity for given position in the genome (Shavkunov et al., 2009). The data of Thomason et al. (2015) also allow estimating the expression level of ncRNAs under various conditions. The positions of primary TSSs in superimposed transcriptional profiles are in general in good agreement with each other

FIGURE 2 | Mapping of TSSs for divergently transcribed ncRNAs according to RNAseq data (Thomason et al., 2015; Ettwiller et al., 2016) and computational prediction (Shavkunov et al., 2009). X axis: coordinates in *E. coli* genome; Y axis: average number of normalized counts with perfect matching for given growth condition. For PlatProm predictions absolute score values are indicated on Y axis. RNA denoted according to the annotation in RegulonDB is shown in cyan; adjacent divergent gene is indicated in gray. Bars above and below X axis correspond to the upper and low DNA strands, respectively. Exactly matching TSSs for ncRNAs are presented in red and for divergently transcribed gene are indicated in blue. Transcriptional landscapes upstream and downstream relatively to the genomic regions depicted in **Figure 2** are shown in **Supplementary Figure S2**.
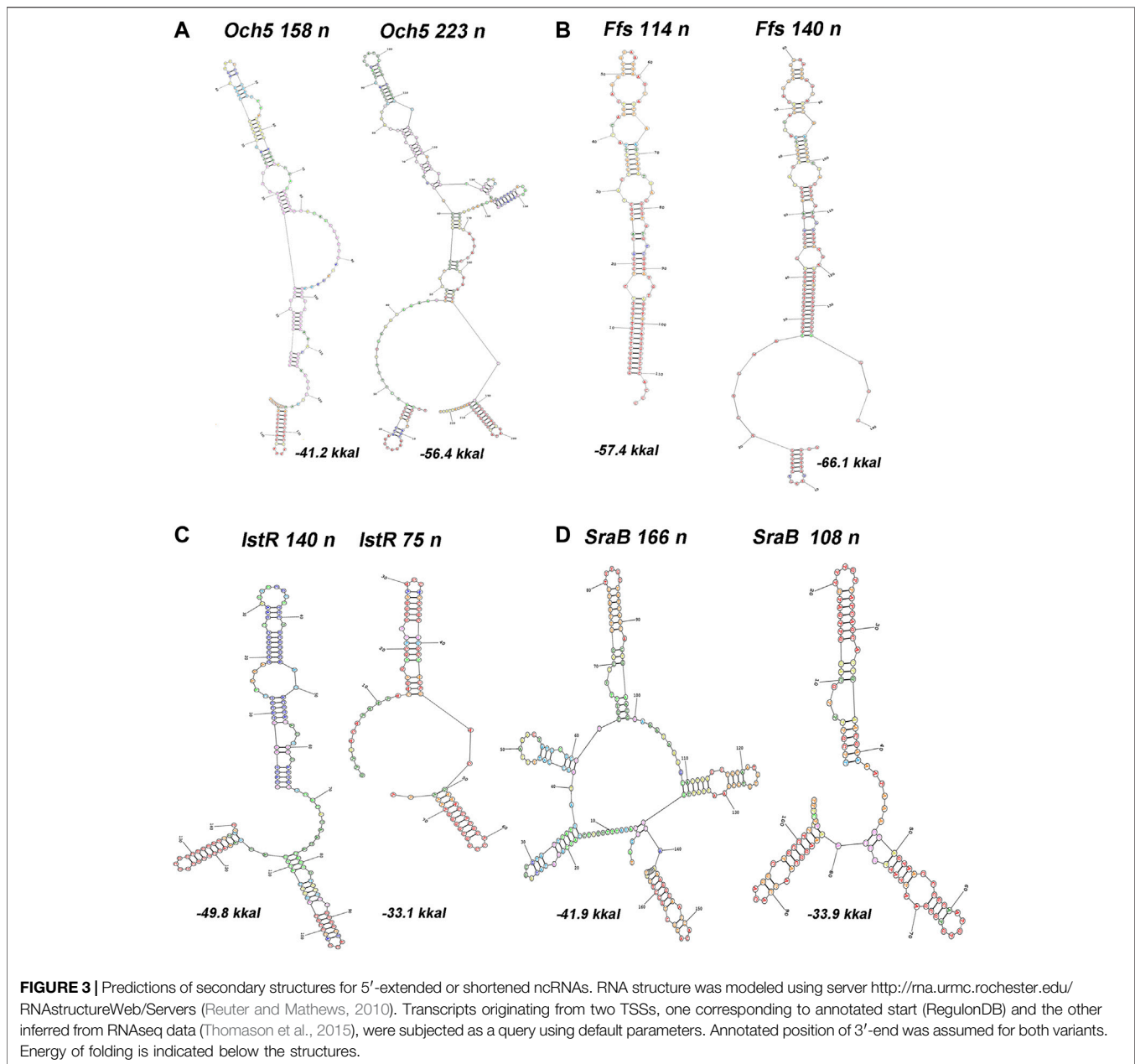
FIGURE 2 | (Continued).

(**Figure 2**) confirming the highest value and objectivity of such data for the reconstruction of the genome-wide transcriptional profile as an integral fingerprint of bacterial physiological status in given growth conditions.

Evidence and Objections
Upshift or Downshift of sRNA 5′ End

RNAseq data allowed verifying the exact positions of transcription starts for divergently transcribed ncRNAs

FIGURE 3 | Predictions of secondary structures for 5′-extended or shortened ncRNAs. RNA structure was modeled using server http://rna.urmc.rochester.edu/ RNAstructureWeb/Servers (Reuter and Mathews, 2010). Transcripts originating from two TSSs, one corresponding to annotated start (RegulonDB) and the other inferred from RNAseq data (Thomason et al., 2015), were subjected as a query using default parameters. Annotated position of 3′-end was assumed for both variants. Energy of folding is indicated below the structures.

(**Supplementary Table S1**). Although the coordinates of primary TSSs for the majority of ncRNAs are in good agreement with their 5′ ends mapped in RegulonDB, the overall transcriptional landscape of near 30 genomic loci needs some corrections. The most notable examples are considered here. Thus, for G0-8898 putative ncRNA annotated as "phantom gene" TSS seems to be located 27 bp upstream from 5′ end indicated in RegulonDB, (**Supplementary Table S1**) for *och*5 this shift is 65 bp, and for *ffs* it is 26 bp (**Figure 2A,B**; **Supplementary Figures S2A,B**). The opposite deviation toward 3′ end was observed for *rse*X, *ytf*L5′ (**Supplementary Table S1**), *sra*B (**Figure 2D**; **Supplementary Figure S2D**) genes—TSSs were moved downstream by 29, 43, 60, and 65 bp, respectively, resulting in shortening of corresponding

ncRNAs. Additional TSS exactly matching computational predictions by PlatProm and prevailing in stationary phase was found for *tff* (**Figure 2C**; **Supplementary Figure S2C**). Thus, *map-tff* genomic loci may produce two divergently transcribed growth-phase dependent ncRNAs differing in length by 71 nt. Nor synergistic or cooperative effect in the expression of *map* and *tff* was observed assuming their autonomous regulation. For *crp*T, double-side deviation was revealed—weak promoter with start point was pushed 164 bp upstream to the annotated start and primary TSS shifted 6 bp downstream (**Supplementary Figure S1B**). Second-order 5′-extended CrpT RNA with considerable expression level only at stationary phase may serve as antisense for yheU mRNA forming

the duplex with 5′UTR (promoter P2). RydC, one of the most abundant and early described sRNAs implicated in the wide spectrum of regulatory networks (Antal et al., 2005) possesses upstream TSS remote from the primary start by 96 bp which enlarges the secondary putative sRNA by a factor 3.

The upshift as well as downshift of ncRNA's 5′ ends if compared to previously accepted positions may account for changing of folding properties and thus selectivity to RNA binding proteins and mRNA targets. Extension or shortening of ncRNA may result in 1) considerable rearrangements of folding pattern (Och5, **Figure 3A**), 2) addition of a new structural motif like when it happens in the case of Ffs and IstR while the "core" ncRNA's structure remains the same (**Figures 3B,C**), or 3) elimination of intramolecular stem-loop structures if TSS appears to be pushed downstream (SraB, **Figure 3D**). Newly attached or removed structural details in ncRNA's architecture may serve as anchors for the interaction with partner molecules, thus determining the spectrum of ncRNA's targets. Noteworthy enlargement or trimming of ncRNAs deduced from refined TSS position does not noticeably change energy of folding and specific impact of nucleotides in RNA stability. SraB and IstR are the exceptions showing increased nucleotide impact in free energy of folding by 26% and 23.8% correspondingly when shortened from 5′ end, respectively, to the annotated length. SraB belongs to the classic type of divergently transcribed overlapping ncRNAs, firstly discovered by Argaman et al., 2001, confirmed by Northern blot and *in vitro* transcription assay, and conserved in *Salmonella typhimurium*, *Klebsiella pneumoniae*, and *Yersinia pestis*. Although in early experiments the expression of *sra*B was proposed to be growth phase dependent and was detected in late stationary phase in LB but not in the mineral medium, RNAseq data pointed out stable production of this RNA in 5′-truncated form in exponential phase in LB as well as in M63 with threefold increase in minimal medium. Surprisingly, in virulent *S. typhimurium* strain, SL1344 *sra*B TSS corresponds to TSS revealed for *E. coli* in Thomason et al. (2015) and Ettwiller et al. (2016), and the length of homologous RNA is expected to be 99 nucleotides, close to that of "shortened" *E. coli* derivative (data from **Supplementary Table S8** in Ramachandran et al., 2014). *S. typhimurium* SraB is shown to be significantly suppressed by ppGpp (Ramachandran et al., 2014) and is proposed to confer resistance to egg albumin (Jiang et al., 2010). The discrepancy in the positioning of SraB 5′-end was probably caused by the presence of deceptively "attractive" canonical sigma70 promoter related to the previously annotated TSS. Downshifted transcription start highlighted by RNAseq on the contrary corresponds to poorly defined promoter although it was indicated as "predicted" in RegulonDB (Santos-Zavaleta et al., 2019). It remains questionable why of the two promoters only one possessing less similarity to consensus is engaged in RNA synthesis. One of the possible explanations may assume collision effects due to transcription initiated at *yce*F P1 promoter (**Figure 2D**, data set of Ettwiller et al. (2016)). At the same time, *sra*B transcription is insensitive to the activity of divergent overlapping promoter *yce*F P2. Another data set used in
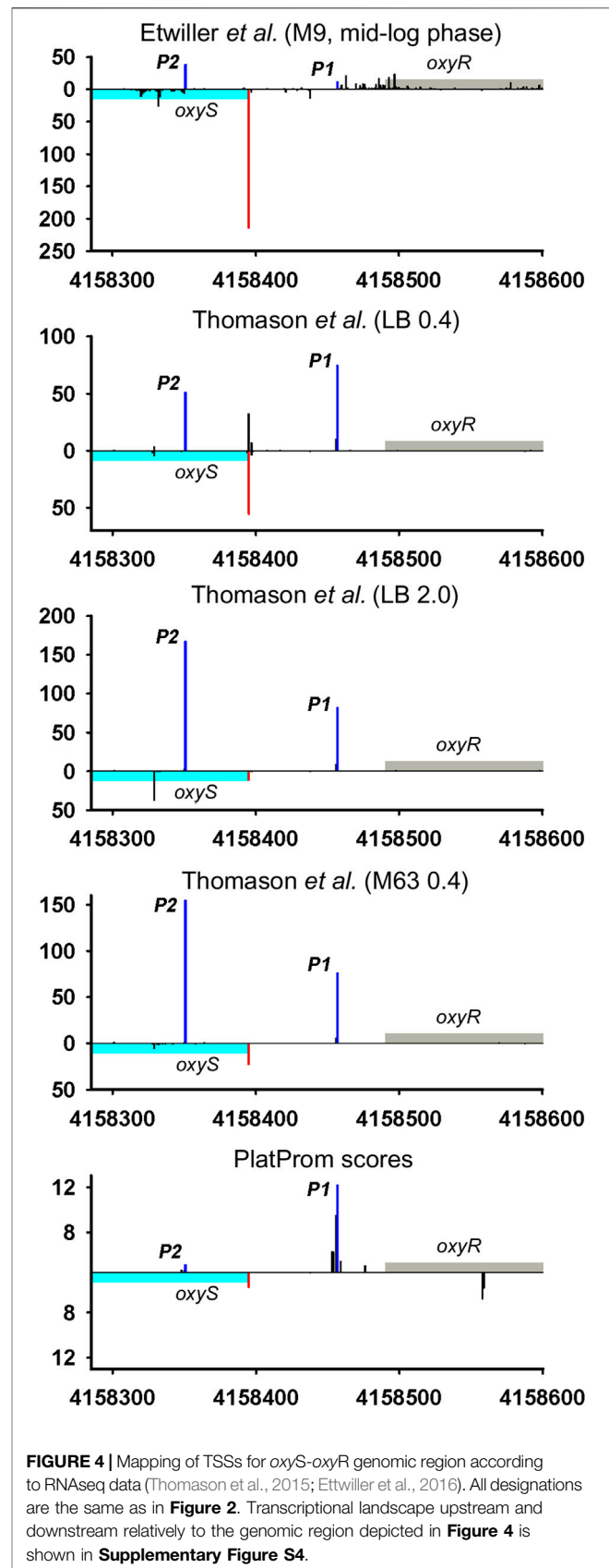
FIGURE 4 | Mapping of TSSs for *oxy*S-*oxy*R genomic region according to RNAseq data (Thomason et al., 2015; Ettwiller et al., 2016). All designations are the same as in **Figure 2**. Transcriptional landscape upstream and downstream relatively to the genomic region depicted in **Figure 4** is shown in **Supplementary Figure S4**.

this study (Thomason et al., 2015) has not confirmed transcriptional activity of *yce*F while the position of SraB TSS perfectly coincides for all RNAseq experiments and computational scanning by PlatProm (**Figure 2D**; **Supplementary Figure S2D**).

Ffs ncRNA overlaps with divergent *atl* transcript by 65 bp although the level of *atl* transcript appeared to be negligible (**Figure 2B**; **Supplementary Figure S2B**). Ffs named also 4.5S is one of the most "deserved" among ncRNAs in the historical aspect due to its function as an essential component of signal recognition particles (SRP). Motif 5′ACCAGGTCAGGTCCG GAAGGAAGCAGC responsible for the formation of apical helix structure was shown to be critical for the interaction with Ffh protein and conserved in bacteria and even high organisms (Peterson and Philips, 2008). Extension by 26 bases from 5′ end does not interfere with the overall folding pattern although additional stem-loop structure appears at the terminus (**Figure 3B**). This appendage as a consequence of TSS upshift is conserved among *Enterobacteriaceae* including *Escherichia*, *Shigella*, *Salmonella*, *Enterobacter*, *Citrobacter*, *Klebsiella*, *Kluyvera*, *Kosakonia*, and *Trabulsiella* as confirmed by Microbial Nucleotide BLAST when extended *ffs* gene was submitted as a query. It is noticeable that in spite of location on plus or minus strand *ffs* retains opposite orientation to the neighboring gene. The relative efficiency of transcription in the pair *ffs*P1 /*atl* P1 for the three experimental conditions used in Thomason et al. (2015) implies antagonistic rather than supporting effect of atlP1 on ffsP1 activity consistent with collision model (Callen et al., 2004).

Duplicated promoters are likely to regulate the expression of *rye*G having two TSSs separated by 26 bp. The amount of corresponding ncRNAs depends on the growth conditions. Both short and long RNAs are consistently activated in stationary phase along with the transcription of the adjacent *tfa*S gene demonstrating cooperative behavior (**Supplementary Figure S1A**, data set of Thomason et al., 2015). Indeed, the distances between TSSs related to promoters P1 and P2 and the start point of TfaS mRNA (57 and 83 bp correspondingly) allow alternate or simultaneous binding of RNA polymerase to transiently unwound DNA and transcription initiation in opposite directions. Two variants of RyeG RNA of 194 and 220 nt in length having identical folding in the core part of the molecule differ in the structure of 3′terminus (**Supplementary Figure S3**) that may result in altered recognition of their targets.

Thus, the same genomic region may be the source of two or more specious ncRNAs differing in length, shape, and stability, thereby assuming to expand the whole spectrum of related functions.

Potential Antisense Transcripts Inferred From TSSs Mapping

Divergently encoded ncRNAs may be added to the family of antisense ncRNAs accounting for newly discovered or corrected positions of TSSs for adjacent divergent genes. Thus, OxyS may overlap 5′UTR of its *vis-à-vis oxy*R by 41bp. Described in 1997

(Altuvia et al., 1997), this RNA for more than 20 years has never ceased to amaze researchers with the variety of its functions. Annotated TSS was concurrently detected in all RNAseq experiments and exactly predicted by PlatProm algorithm (**Figure 4**; **Supplementary Figure S4**). Assuming the activity of sole P1 *oxy*R promoter OxyS may be defined as typical *bona fide* ncRNA although obvious activity of P2 promoter coming to be major in stationary phase and in mineral medium allows to ascribe to OxyS potential role of antisense RNA.

Similarly, MicC over its entire length may serve as antisense for *omp*N mRNA having long 5′UTR with TSS remote from initiation codon by 408 bp (**Supplementary Table S1**). In the pair, *micA/luxS* highly expressed *micA* sRNA is overlapped with 5′UTRs of one of two *lux*S-related mRNAs. MicA is involved in a number of RNA-RNA interactions in *trans* presumably with mRNAs related to cell-envelop stress (PhoP, LamB, OmpA, OmpW, Tsx, OmpX, FimB) (Mihailovic et al., 2018). For example, MicA is capable of forming stable duplex with OmpA mRNA as confirmed *in vitro* using SPR technique (Vincent et al., 2013). Overlapping of ncRNAs with 5′UTRs of divergent mRNAs allowing ascribing to the non-coding transcripts potential antisense functions may be proposed for the pairs G0-10702/YhcC (15 bp duplex), G0-10703/yhcG (128 bp), RdlD/ldrD (136 bp), G0-10705/GlnA (63 bp, significant for stationary phase), SymR/SymE (76 bp, entire length), G0-10703/YhcG, ArrS/GadE (68 bp, entire length) (**Supplementary Table S1**).

Therefore, two types of corrections based on RNAseq data make it possible to consider ncRNA as antisense, identifying a previously unknown TSS for the divergent gene and/or discovering additional TSSs for ncRNA. Transcriptional regulation via multiple promoters is widely exploited in prokaryotes and it seems to be a general principle of promoters evolving in intergenic loci. Thus, from 28 to 40% of experimentally found promoters for mRNA are accompanied by the "second-order" regulatory sites within 250 bp long region upstream to the ORFs (Huerta and Collado-Vides, 2003). The classical example of paired promoters for the transcription of rRNA separated by 120 bp ensures flexible growth rate-dependent regulation of ribosome synthesis (Gourse et al., 1996). *dad*A encoding D-amino acid dehydrogenase is transcribed from three consecutive promoters differing in their activity and selective response to cAMP-CRP (Zhi et al., 1998). In some cases, tandemly arranged promoters may be recognized by RNA polymerase holoenzyme containing alternative sigma-factors and adopted to respond to changing environmental conditions, thus providing reserve facilities for transcriptional regulation. For example, AppY mRNA may be initiated at two promoters separated by 76 bp and recognized by sigma70 or sigma38 RNAP (Rangarajan and Schnetz, 2018). Sigma70 and sigma38 dependent TSSs of *nha*A are separated by 141 bp demonstrating differential sensitivity to HNS and elevated sodium concentrations (Dover and Padan, 2001). Promoter duplication may obviously be exploited in bacteria for the fine turning of ncRNA's expression and the extension of accessibility of divergent mRNAs for regulation via antisense mechanism.
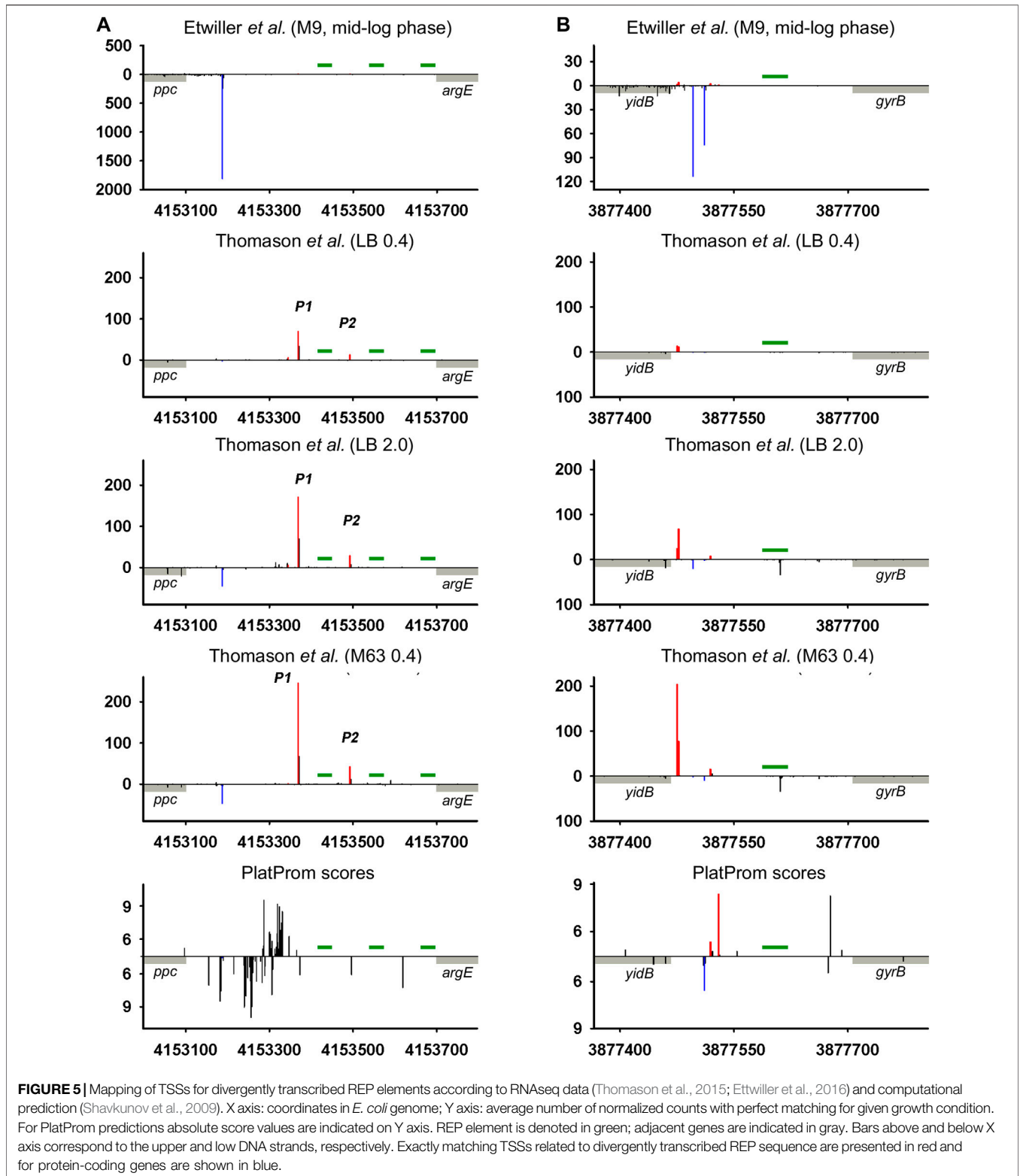
Divergent ncRNAs as Purveyors of Extracellular exRNAs

Besides indubitable universality of ncRNAs in coordination of intracellular regulatory processes, their role as messenger molecules in cell communications is coming to be more and more appreciable. RNA is exported as small fragments (less than 60 nt in length) in a controlled manner exploiting outer membrane vesicles (OMV) or yet unrevealed secretion machinery (Ghosal, 2018). Among the total pool of extracellular RNAs, ncRNA's derivatives comprise a substantial fraction along with rRNA's and tRNA's generated fragments. Divergently transcribed Ffs RNA contributes to this set being preferentially exported in OMV-free manner in *E. coli* (Ghosal et al., 2015). Its function as external signal molecule in bacterial population is likely to be conserved since homologous ncRNA of *Salmonella enterica* serovar Typhimurium LT2 was detected in OMV-enriched samples (Malabirade et al., 2018). Nevertheless, pathways of processing and secretion of Ffs ncRNA differ in two Enterobacteria—in *E. coli* only short fragment of 29 nt was shown to be exported while in *S. Typhimurium* LT2 full-length transcript enclosed in or bound to OMV was revealed. The involvement of Ffs in intercellular communication is confirmed by the observation of considerable increase (up to 13-fold) in its amount in cultivation medium for mixed bacterial population (*E. coli*—*Paenibacillus bisonicum*) if compared to pure *E. coli* culture (Alikina et al., 2018). Another three extracellular exRNAs originated from divergently transcribed precursors (ChiX, RalA, and RyhB) demonstrated opposite reaction on the presence of competing bacterium resulting in 2- to 20-fold decrease in their secretion. It should be noted that 5′ ends of short fragments of ChiX (5′ ACACCGTCGCTTAAAG) and RalA (5′ GAGGACTGAAGT) revealed in cultural fluid (Alikina et al., 2018) exactly correspond to primary TSSs of parental ncRNAs indicating possible endonuclease cutting out or release of aborted transcripts generated upon initiation of RNA synthesis. RyhB-related exRNA (5′CCGGGTGCTGGCTTTTTTT), and vice versa, may be ascribed to 3′ terminus of full-length ncRNA. 5′-end of Ffs-born intracellular fragment (Alikina et al., 2018; **Supplementary Table S1**) is shifted down by 11 nt relatively to TSS refined according to the data of Thomason et al. (2015). It means that additional stem-loop structure expected to form in 5′extended transcript (**Figure 3B**) may exert functional role for nibbling out short RNA. Ffs, ChiX, and RyhB belong to the group of the most abundant ncRNAs with enhanced expression in stationary phase (Thomason et al., 2015) raising the possibility of their involvement in the control of cell density in bacterial population. Otherwise, RalA is accumulated at very low level which makes the presence of its fragment among secreted RNAs rather intriguing, indicating selectivity of export.

Putative Divergent ncRNAs Containing REP Elements

Intergenic regions, along with the fact that they serve as reservoir of ncRNAs, may contain repeating sequences with mirror-symmetry, so called REP (repetitive extragenic palindromic) elements. The transcription of these sequences yielding RNAs with potential regulatory functions is still questionable. REP elements represent conservative universal structural modules of bacterial genomes. Deduced on the base of the alignment of 35 *E. coli* and *S. typhimurium* intergenic regions, REP sequence was proposed to be 35 bp long symmetrical motif GC (g/t) GATGGCG (g/a)GC (g/t)…(g/a)CG (c/t)CTTATC (c/a) GGCCTAC (Stern et al., 1984). Similar stretches homologous to 5′TGCCGGATGCGGCGTAAACGCCTTATCCGGCCTAC (Tobes and Ramos, 2005) and 5′GCCGGATGCGGCGTGAAC GCCTTATCCGGCCTACGA (Markelova et al., 2016) were found in *E. coli* genome in 290 and 224 copies correspondently without any preferential location between convergent or codirectional genes. Self-complementarity assuming the ability to form hairpin structures in RNA or in negatively supercoiled DNA propelled to discuss the involvement of REP elements in regulation of gene expression (Higgins et al., 1988). In spite of efforts to confirm their role in transcription activation, attenuation, or maintenance of mRNA stability (Higgins et al., 1988), the only function related to premature termination of translation can be considered as experimentally approved (Liang et al., 2015). Thus, REP motif located at a distance no longer than 15 bp from translation termination codon in 3′UTR of mRNA affects the abundance of corresponding protein. At the same time, near 80% of REP elements are pushed down from the 3′ terminus of adjacent gene at a longer distance (Markelova et al., 2016) implying another biological function. It was suggested that REP sequences may be read-through as autonomous transcription units contributing to the family of novel non-coding RNAs (Qian and Adhya, 2017). ncRNA CbsR16 originated from REP-containing repeat-reach region of *Coxiella burnetii* chromosome was described recently (Wachter et al., 2018). Here, we analyzed 79 REP-containing genomic loci separating collinear protein-coding genes for the presence of TSSs providing divergently transcribed RNAs encompassing REP elements. Indeed in 35 cases TSSs were found indicating transcription in opposite direction relatively to the juxtaposed gene (**Supplementary Material**, List of REP-containing genomic regions). Two representative examples are shown in **Figure 5**. In the case of *ppc-argE* intergenic region *bona fide* divergent transcription expected to capture 3 and 2 consecutive REP elements was registered (**Figure 5A**; Thomason et al., 2015). For *yidB-gyrB* loci, divergent RNA initiated between two codirected genes seems to be true antisense product complementary to *yidB* 5′UTR (**Figure 5B**). Weak expression of *bona fide* ncRNA trimmed from 5′ end was detected only for stationary phase and mineral medium. Does the presence of REP elements as a structure-forming motifs impact in the functional properties of corresponding RNAs remains unclear. Obviously, REPs may be included in novel divergently transcribed RNAs if there is no intrinsic terminator between TSS and the 5′-end of REP. Only 16 of 35 REP-containing intergenic regions have T_n tracks indicating potential intrinsic terminator between TSS and 5′-end of REP, though not all T-rich blocks are preceded by self-complementary sequence essential for transcription termination. It increases the probability of transcription passing through REP motifs. The distance from TSSs to REP 5′ end varies from 16 to 298 bp allowing also to assume their function in rho-dependent transcription termination proposed in Espeli et al. (2001).

FIGURE 5 | Mapping of TSSs for divergently transcribed REP elements according to RNAseq data (Thomason et al., 2015; Ettwiller et al., 2016) and computational prediction (Shavkunov et al., 2009). X axis: coordinates in *E. coli* genome; Y axis: average number of normalized counts with perfect matching for given growth condition. For PlatProm predictions absolute score values are indicated on Y axis. REP element is denoted in green; adjacent genes are indicated in gray. Bars above and below X axis correspond to the upper and low DNA strands, respectively. Exactly matching TSSs related to divergently transcribed REP sequence are presented in red and for protein-coding genes are shown in blue.

The relative positions of TSSs and the borders of REPs admit the possibility of their inclusion in ncRNAs as well as return us to the early hypothesis regarding the role of REPs as transcription punctuation signals. In the last case, REPs may serve to block transcription of divergent ncRNAs, thus determining their length.
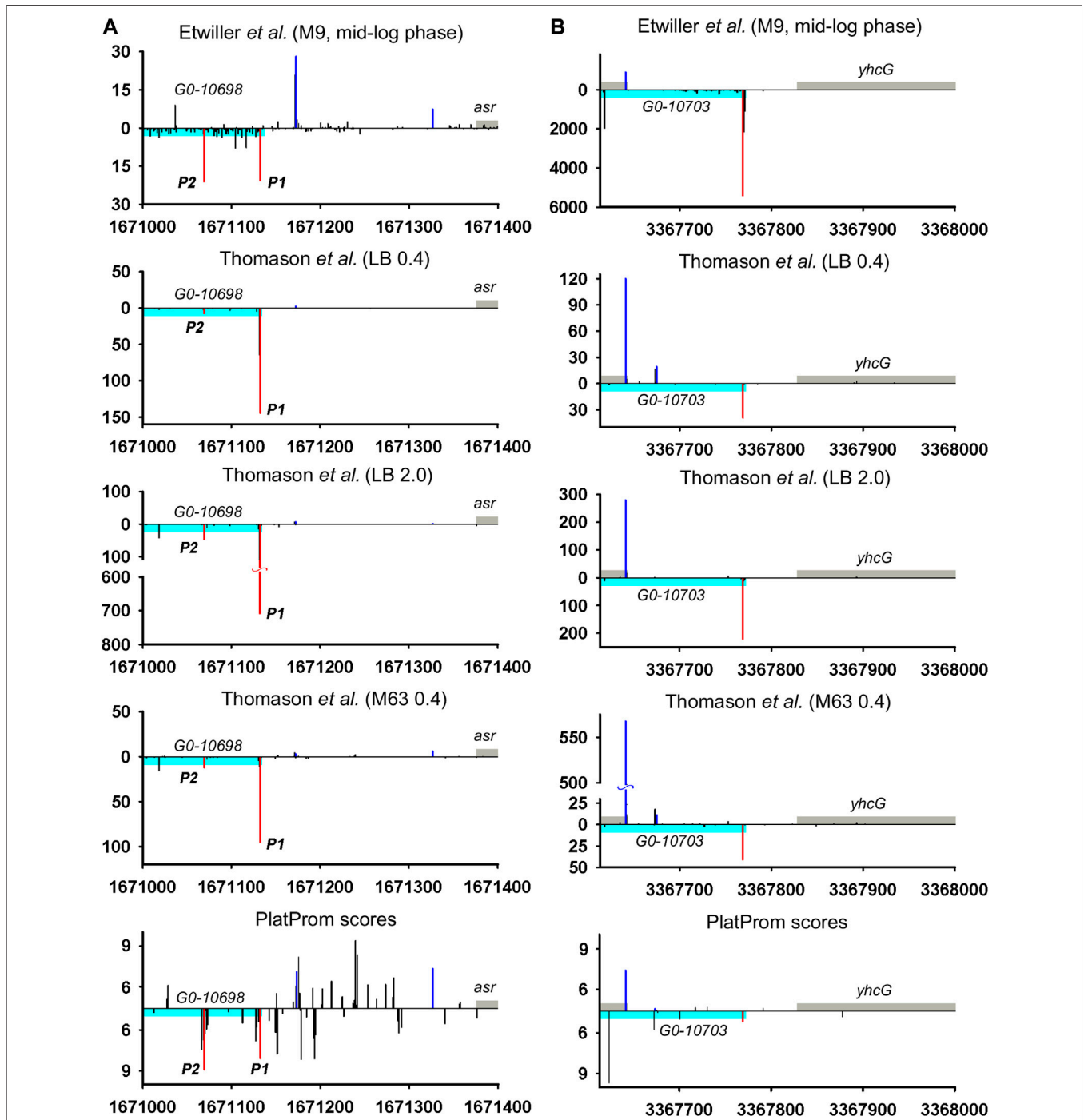
FIGURE 6 | Mapping of TSSs for divergent potential ncRNAs—G0-10698, G0-10703, G0-10702, and G0-10705 according to RNAseq data (Thomason et al., 2015; Ettwiller et al., 2016) and computational prediction (Shavkunov et al., 2009). X axis: coordinates in *E. coli* genome; Y axis: average number of normalized counts with perfect matching for given growth condition. For PlatProm predictions absolute score values are indicated on Y axis. RNA denoted according to the annotation in RegulonDB is shown in cyan; adjacent divergent gene is indicated in gray. Bars above and below X axis correspond to the upper and low DNA strands, respectively. Exactly matching TSSs for ncRNAs are presented in red and for divergently transcribed gene are shown in blue. Transcriptional landscapes upstream and downstream relatively to the genomic regions corresponding to G0-10698-*asr* and G0-10703-*yhc*G are shown in **Supplementary Figure S5**.

FIGURE 6 | (Continued).
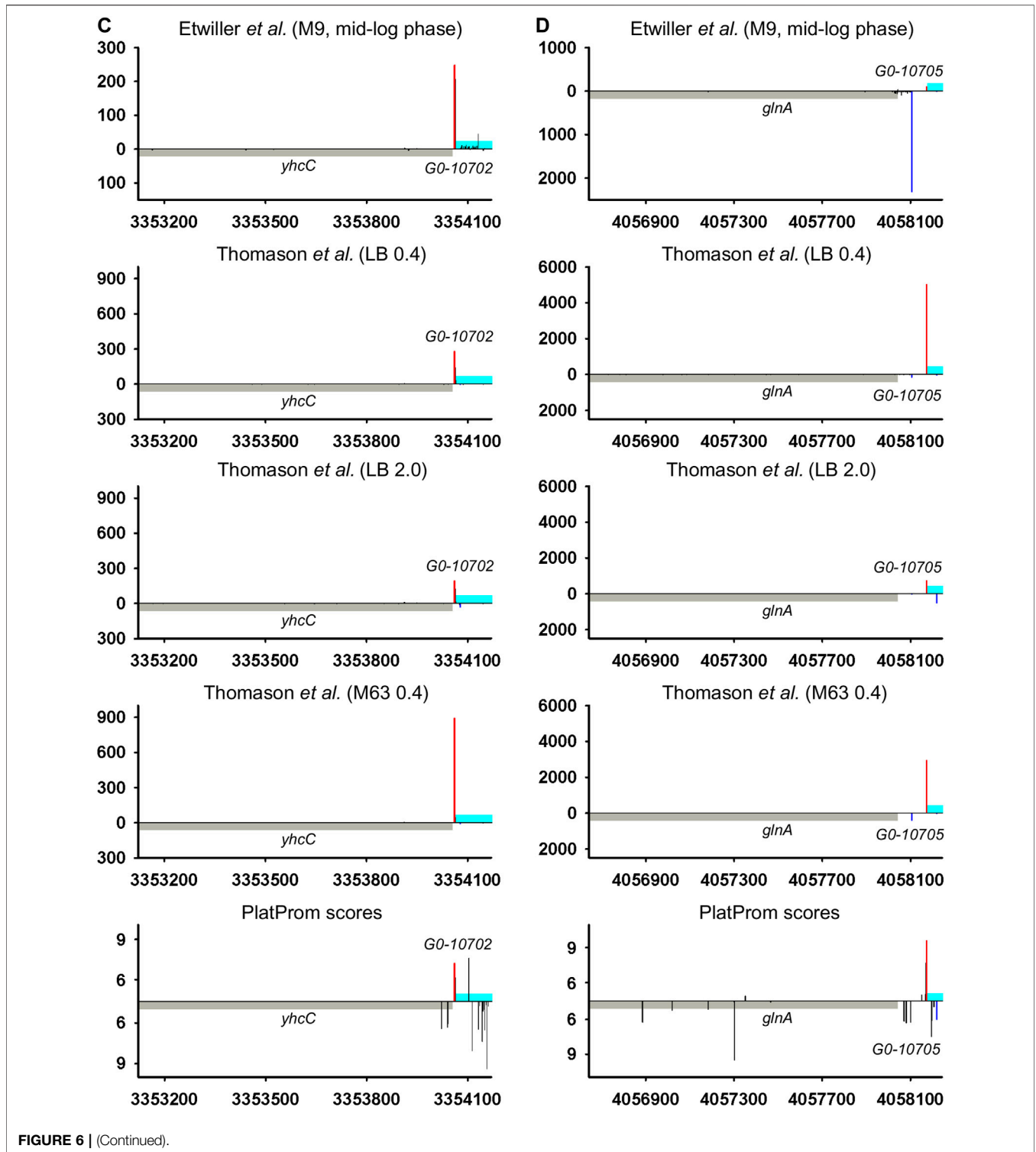
ncRNAs Under the Question—Are They Really Transcribed or Not?

Hypothetical ncRNAs inferred from the data of Raghavan et al. (2011) so far remained as the candidates for the experimental validation and in 2019 were even excluded from EcoCyc database (Keseler et al., 2017) as lacking of sufficient experimental evidences of their expression. 5′-specific RNAseq (Thomason et al., 2015; Ettwiller et al., 2016) unambiguously confirmed the

TABLE 1 | Data used for TSSs mapping.

Accession numbers	Total number of reads after adapter trimming and sifting out of reads less than 16 bp in length	Number of reads perfectly matching the genome	Normalizing coefficient (N_{av}/N_i)
Ettwiller et al., 2016 (M9, mid-log phase)			
ERR930221	14.063,469	11,015,459	0.7322
ERR930222	9,184,197	5,116,134	1.5765
Thomason et al., 2015 (LB 0.4)			
SRR1173969	6,145,538	4886330	0.9927
SRR1173970	5,834,605	4,814,550	1.0075
Thomason et al., 2015 (LB 2.0)			
SRR1173974	5,627,281	3,341,088	1.3391
SRR1173978	6,967,456	5,072,554	0.8821
SRR1173979	7,691,759	5,861,466	0.7633
SRR1173980	6,214,631	3,621,841	1.2353
Thomason et al., 2015 (M63 0.4)			
SRR1173985	8,368,755	5,128,824	1.0292
SRR1173986	8,445,228	5,428,622	0.9724

transcription of four divergent potential ncRNAs—G0-10698, G0-10703, G0-10702, and G0-10705 (**Figure 6**; **Supplementary Figure S5**). Detected TSS in generals correspond to 5′ ends annotated in RegulonDB and are in perfect agreement with PlatProm predictions. Expression of these RNAs demonstrates clear dependence on growth phase, especially for G0-10698 and G0-10703 activated in stationary phase. G0-10703 RNA may be attributed as antisense transcript for *yhc*G gene with yet uncertain function whose expression follows from all sets of experimental data as well as from computational predictions (**Figure 6B**; **Supplementary Figure S5B**). G0-10698 and G0-10702 are conserved among *Enterobacteriaceae* (*E. coli, Shigella*, *Salmonella* sp. HNK130, *Salmonella* sp. S13, and *Enterobacter hormaechei* str. RHBSTW-00218). G0-10703 is found in some strains of *E. coli* and in *Salmonella* sp. HNK130, whereas G0-10705 is more widely spread among *E. coli, Shigella, Salmonella enterica, Salmonella* sp. HNK130, *Salmonella* sp. S13, *Enterobacter hormaechei* str. RHBSTW-00218, *Citrobacter rodentium* ICC168, and *Citrobacter* sp. RHB25-C09. Thus, undeservedly rejected divergently transcribed ncRNAs may regain their place in *E. coli* genome annotation.

DISCUSSION

High-throughput sequencing data provide a truly inexhaustible source for the deduction of new molecular relationships in microbial physiology. Mapping of additional TSSs in upstream regions of annotated ncRNAs not only contributes to our knowledge regarding the diversity of specious RNA produced from the same intergenic region, but also enforces us to reconsider functional attribution of certain ncRNAs. More than a half of divergently transcribed ncRNAs are present as a pair or triplet of collinear products transcribed from different TSSs what is consistent with the phenomenon of promoter duplication or multiplication well known for protein-coding genes (Huerta and Collado-Vides, 2003). Despite the role of RNAP "concentrator" additional promoters may provide

optimal transcription processivity for ncRNAs according to the cooperation model proposed in Kim et al., 2019. The promoters of known ncRNAs often occur as a sequential array with separate TSSs providing the accumulation of ncRNAs of different lengths and potentially different specificities. The ratio of the activity of alternative promoters for ncRNA is not a stable characteristic and depends on growth conditions. In the case of *och*5 in mineral medium, all three promoters (P1, P2, and P3) demonstrate comparable activity, whereas in LB stationary phase P2 becomes the most active (**Figure 2A**; **Supplementary Figure S2A**). For REP-associated potential ncRNA in *ppc-arg*E intergenic region, synergistic effect of individual promoters seems to be rather probable (**Figure 5**; **Supplementary Figure S5**) although for *och*5 and *tff* (**Figure 2A,C**; **Supplementary Figures S2A,C**) cooperation between promoters for ncRNAs may interfere with transcription in opposite direction. It also cannot be excluded that less active promoters represent residual traces or, on the contrary, novel signs on the way of ncRNA selection. It is consistent with the observation that young ncRNAs usually are expressed with low efficiency if compared to evolutionary fixed ones (Kacharia et al., 2017).

Only 30 ncRNAs of 61 chosen as divergently transcribed examples accounting for their genomic context may be attributed as *bona fide* non-overlapped RNAs. The remaining 31 ncRNAs appeared to have TSS within coding part or 5′UTR of the adjacent gene rendering them as potential antisense regulators. Their expression may be subjected to the collision effects (Callen et al., 2004) resulting in substantial decrease in the level of ncRNA caused by convergent transcription. It is illustrated by the case of *oxy*R-*oxy*S pair when engagement of oxyR P2 promoter reduces the level of OxyS (**Figure 4**). We intended to outline correlation between transcriptional rate of true divergent ncRNAs and their counter-partners. However, the level of transcription of the opposite genes was often hardly detectable whereas ncRNA was rather abundant (*atl-ffs, yce*F-*sra*B, **Figure 2B,D**, *yhc*C-G0-10702, **Figure 6C**). The number of examples available from RNA-seq data and used here did not allow us to conclude that the proximity of highly expressed divergent ncRNAs with transcriptionally "sluggish" genes is

more a rule than an exception. Antagonistic rather than supporting effect of divergent transcription on the activity of model promoter mediated by excessive superhelical tension in intervening region was recently reported (Kim et al., 2019). One cannot exclude that this mechanism is implemented in the pairs "ncRNA/adjacent protein-coding gene" when active transcription of ncRNA suppresses the production of mRNA. The extent of mutual promoter interference should obviously depend on the distance between TSSs located on opposite DNA strands, which in the case of true divergent ncRNAs varies in the range from 7 to 290 bp with a median value of 84 bp.

One of the noteworthy observations is that two essential ncRNAs expressed from strong single promoters, Ffs and SraB, appeared to be longer and shorter correspondingly than it was previously assumed. The exact nucleotide sequence of RNA, accounting for the position of TSS, is important for the formation of its secondary structure, which is recognized by target molecules. For Ffs, addition of stem-loop structure at 5′end may serve a hallmark for enzymatic degradation. It is confirmed by the finding that Ffs-derivative protruding upstream from the annotated TSP was detected in intracellular RNA fraction indicating the presence of the extended precursor (Alikina et al., 2018). SraB presumed to start from the position 1,146,589 on plus strand (NC_000913.3) that overlaps with early translated region of *yce*F and is likely to be initiated 60 bp downstream. The refined position of TSS was reproduced for all data sets used in our analysis and was also supported by computational prediction (**Figure 2D**). The exact length of SraB was discussed when it was detected for the first time and the presence of shortened 105 bp product was also mentioned (Argaman et al., 2001). At least seven ncRNAs are likely to wait for rescaling of their size and re-drawing folding pattern. Focusing here on divergently encoded *E. coli* ncRNAs, we did not focus on collinear regulatory RNAs, for which new boundaries can also be established and, accordingly, new functions and ways for the regulation of their expression can be proposed.

METHODS

Raw data available from Thomason et al. (2015) and Ettwiller et al. (2016) (**Table 1**) were used for the mapping of 5′ ends of transcripts on the genome *E. coli* K-12 str. MG1655 (NC_000913.3). Data processing was carried out in the following stages: 1) 3′-terminal adapter sequence (AAAAAAAAAA–in Thomason et al., 2015; AGATCGGAAGA–in Ettwiller et al., 2016) was cut off; 2) sequences less than 16 bp in length were removed from the whole massive of reads, and residual reads were mapped using bowtie2, generating sam-file1 (Langmead and Salzberg, 2012), with parameter L 16; 3) resulting sam-file1 was used for the extraction of

reads perfectly matching the genome (only those containing both labels XM:i:0 and NM:i:0 were selected) generating sam-file2; 4) sam-file2 was converted into bam-format using SAMtools (Li et al., 2009), and the same program was used for the sorting of bam-files; 5) distribution of 5′ ends along both DNA strands was obtained using BEDtools (Quinlan and Hall, 2010). The number of 5′ ends for given position in the genome was normalized using coefficient calculated as ratio of the average number of perfectly matching reads in a series of experiments (N_{av}) to the number of perfectly matching reads in given experiment (N_i). The data shown in **Figure 2** (2 S), **Figure 4** (4 S), **Figure 5**, **Figure 6** (5 S), and **Supplementary Figure S1** represent an average number of normalized counts corresponding to two biological replicates of Ettwiller et al. (2016), two biological replicates for LB OD_{600} = 0.4 (Thomason et al., 2015), 4 biological replicates for LB OD_{600} = 2.0 (Thomason et al., 2015), and two biological replicates for M63 OD_{600} = 0.4 (Thomason et al., 2015). Accession numbers of data used in this study are given in **Table 1**.

DATA AVAILABILITY STATEMENT

Publicly available datasets (from Thomason et al., 2015; Etwiller et al., 2016) were analyzed in this study. This data can be found in European Nucleotide Archive (ERR930221, ERR930222) and NCBI Sequence Read Archive (SRR1173969, SRR1173970, SRR1173974, SRR1173978, SRR1173979, SRR1173980, SRR1173985, SRR1173986).

AUTHOR CONTRIBUTIONS

IM was responsible for writing the manuscript; SK for transcriptome data processing and writing the manuscript; NM for transcriptome data processing.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.610453/full#supplementary-material.

REFERENCES

Adams, P. P., and Storz, G. (2020). Prevalence of small base-pairing RNAs derived from diverse genomic loci. *Biochim. Biophys. Acta Gene Regul. Mech.* 1863, 194524. doi:10.1016/j.bbagrm.2020.194524

Alikina, O. V., Glazunova, O. A., Bykov, A. A., Kiselev, S. S., Tutukina, M. N., Shavkunov, K. S., et al. (2018). A cohabiting bacterium alters the spectrum of short RNAs secreted by *Escherichia coli*. *FEMS Microbiol. Lett.* 365 (24). doi:10.1093/femsle/fny262

Altuvia, S. (2007). Identification of bacterial small non-coding RNAs: experimental approaches. *Curr. Opin. Microbiol.* 10, 257–261. doi:10.1016/j.mib.2007.05.003

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L., and Storz, G. (1997). A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* 90, 43–53. doi:10.1016/s0092-8674(00)80312-8

Antal, M., Bordeau, V., Douchin, V., and Felden, B. (2005). A small bacterial RNA regulates a putative ABC transporter. *J. Biol. Chem.* 280, 7901–7908. doi:10.1074/jbc.M413071200

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E. G. H., Margalit, H., et al. (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr. Biol.* 11, 941–950. doi:10.1016/s0960-9822(01)00270-6

Backofen, R., and Hess, W. R. (2010). Computational prediction of sRNAs and their targets in bacteria. *RNA Biol.* 7, 33–42. doi:10.4161/rna.7.1.10655

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

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr. Opin. Microbiol.* 10, 102–109. doi:10.1016/j.mib.2007.03.012

Bryant, J. A., Sellars, L. E., Busby, S. J., and Lee, D. J. (2014). Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* 42, 11383–11392. doi:10.1093/nar/gku828

Callen, B. P., Shearwin, K. E., and Egan, J. B. (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell* 14, 647–656. doi:10.1016/j.molcel.2004.05.010

Carrier, M. C., Lalaouna, D., and Massé, E. (2018). Broadening the definition of bacterial small RNAs: characteristics and mechanisms of action. *Annu. Rev. Microbiol.* 72, 141–161. doi:10.1146/annurev-micro-090817-062607

Dar, D., and Sorek, R. (2018). Bacterial noncoding RNAs excised from within protein-coding transcripts. *mBio*, 9, e01730–18. doi:10.1128/mBio.01730-18

Dorman, C. J. (2019). DNA supercoiling and transcription in bacteria: a two-way street. *BMC Mol. Cell Biol.* 20, 26. doi:10.1186/s12860-019-0211-6

Dover, N., and Padan, E. (2001). Transcription of *nha*A, the main Na$^+$/H$^+$ antiporter of *Escherichia coli*, is regulated by Na(+) and growth phase. *J. Bacteriol.* 183, 644–653. doi:10.1128/JB.183.2.644-653.2001

Dutcher, H. A., and Raghavan, R. (2018). Origin, evolution, and loss of bacterial small RNAs. *Microbiol. Spectrum* 6 (2), 10. doi:10.1128/microbiolspec.RWR-0004-2017

Dutta, T., and Srivastava, S. (2018). Small RNA-mediated regulation in bacteria: a growing palette of diverse mechanisms. *Gene* 656, 60–72. doi:10.1016/j.gene.2018.02.068

Espéli, O., Moulin, L., and Boccard, F. (2001). Transcription attenuation associated with bacterial repetitive extragenic BIME elements. *J. Mol. Biol.* 314, 375–386. doi:10.1006/jmbi.2001.5150

Ettwiller, L., Buswell, J., Yigit, E., and Schildkraut, I. (2016). A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genom.* 17, 199–210. doi:10.1186/s12864-016-2539-z

Ghosal, A. (2018). Secreted bacterial RNA: an unexplored avenue. *FEMS Microbiol. Lett.* 365. doi:10.1093/femsle/fny036

Ghosal, A., Upadhyaya, B. B., Fritz, J. V., Heintz-Buschart, A., Desai, M. S., Yusuf, D., et al. (2015). The extracellular RNA complement of *Escherichia coli*. *MicrobiologyOpen* 4, 252–266. doi:10.1002/mbo3.235

Gourse, R. L., Gaal, T., Bartlett, M. S., Appleman, J. A., and Ross, W. (1996). rRNA transcription and growth rate-dependent regulation of ribosome synthesis in *Escherichia coli*. *Annu. Rev. Microbiol.* 50, 645–677. doi:10.1146/annurev.micro.50.1.645

Higgins, C. F., McLaren, R. S., and Newbury, S. F. (1988). Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? a review. *Gene* 72, 3–14. doi:10.1016/0378-1119(88)90122-9

Hör, J., Matera, G., Vogel, J., Gottesman, S., and Storz, G. (2020). Trans-acting small RNAs and their effects on gene expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus*. doi:10.1128/ecosalplus.ESP-0030-2019

Huerta, A. M., and Collado-Vides, J. (2003). Sigma70 Promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *J. Mol. Biol.* 333, 261–278. doi:10.1016/j.jmb.2003.07.017

Jiang, H., Cao, M., Cao, X., Gu, H., and Zeng, K. (2010). [sRNA (sraB) regulate the resistant ability of *Salmonella enterica* serovar enteritidis to egg albumen]. *Wei Sheng Wu Xue Bao* 50, 1537–1544. doi:10.13343/j.cnki.wsxb.2010.11.017

Jose, B. R., Gardner, P. P., and Barquist, L. (2019). Transcriptional noise and exaptation as sources for bacterial sRNAs. *Biochem. Soc. Trans.* 47, 527–539. doi:10.1042/BST20180171

Kacharia, F. R., Millar, J. A., and Raghavan, R. (2017). Emergence of new sRNAs in enteric bacteria is associated with low expression and rapid evolution. *J. Mol. Evol.* 84, 204–213. doi:10.1007/s00239-017-9793-9

Kawano, M., Reynolds, A. A., Miranda-Rios, J., and Storz, G. (2005). Detection of 5'- and 3'-UTR-derived small RNAs and cis-encoded antisense RNAs in *Escherichia coli*. *Nucleic Acids Res.* 33, 1040–1050. doi:10.1093/nar/gki256

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., et al. (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 45, D543–D550. doi:10.1093/nar/gkw1003

Kim, S., Beltran, B., Irnov, I., and Jacobs-Wagner, C. (2019). Long-distance cooperative and antagonistic RNA polymerase dynamics via DNA supercoiling. *Cell* 179, 106–119.e16. doi:10.1016/j.cell.2019.08.033

Langmead, B., and Salzberg, S. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi:10.1038/nmeth.1923

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Liang, W., Rudd, K. E., and Deutscher, M. P. (2015). A role for REP sequences in regulating translation. *Mol. Cell* 58, 431–439. doi:10.1016/j.molcel.2015.03.019

Lim, H. M., Lewis, D. E., Lee, H. J., Liu, M., and Adhya, S. (2003). Effect of varying the supercoiling of DNA on transcription and its regulation. *Biochemistry* 42, 10718–10725. doi:10.1021/bi030110t

Lindgreen, S., Umu, S. U., Lai, A. S., Eldai, H., Liu, W., McGimpsey, S., et al. (2014). Robust identification of noncoding RNA from transcriptomes requires phylogenetically-informed sampling. *PLoS Comput. Biol.* 10, e1003907. doi:10.1371/journal.pcbi.1003907

Liu, W., Rochat, T., Toffano-Nioche, C., Le Lam, T. N., Bouloc, P., and Morvan, C. (2018). Assessment of *bona fide* sRNAs in *Staphylococcus aureus*. *Front. Microbiol.* 9, 228. doi:10.3389/fmicb.2018.00228

Malabirade, A., Habier, J., Heintz-Buschart, A., May, P., Godet, J., Halder, R., et al. (2018). The RNA complement of outer membrane vesicles from *Salmonella enterica* serovar Typhimurium under distinct culture conditions. *Front. Microbiol.* 9 (9), 2015. doi:10.3389/fmicb.2018.02015

Markelova, N. Yu., Masulis, I. S., and Ozoline, O. N. (2016). REP-elements of the *Escherichia coli* genome and transcription signals: positional and functional analysis. *Mat. Biolog. Bioinform.* 11 (Issue Suppl), 1–14. doi:10.17537/2016.11.t1

Masulis, I. S., Babaeva, Z. Sh., Chernyshov, S. V., and Ozoline, O. N. (2015). Visualizing the activity of *Escherichia coli* divergent promoters and probing their dependence on superhelical density using dual-colour fluorescent reporter vector. *Sci. Rep.* 5, 11449. doi:10.1038/srep11449

Mihailovic, M. K., Vazquez-Anderson, J., Li, Y., Fry, V., Vimalathas, P., Herrera, D., et al. (2018). High-throughput *in vivo* mapping of RNA accessible interfaces to identify functional sRNA binding sites. *Nat. Commun.* 9, 4084. doi:10.1038/s41467-018-06207-z

Peer, A., and Margalit, H. (2014). Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. *RNA* 20, 994–1003. doi:10.1261/rna.043133.113

Peter, B. J., Arsuaga, J., Breier, A. M., Khodursky, A. B., Brown, P. O., and Cozzarelli, N. R. (2004). Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol.* 5, R87. doi:10.1186/gb-2004-5-11-r87

Peterson, J. M., and Phillips, G. J. (2008). Characterization of conserved bases in 4.5S RNA of *Escherichia coli* by construction of new F' factors. *J. Bacteriol.* 190, 7709–7718. doi:10.1128/JB.00995-08

Qian, Z., and Adhya, S. (2017). DNA repeat sequences: diversity and versatility of functions. *Curr. Genet.* 63, 411–416. doi:10.1007/s00294-016-0654-7

Quinlan, A. R., and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. doi:10.1093/bioinformatics/btq033

Raghavan, R., Groisman, E. A., and Ochman, H. (2011). Genome-wide detection of novel regulatory RNAs in *E. coli*. *Genome Res.* 21, 1487–1497. doi:10.1101/gr.119370.110

Ramachandran, V. K., Shearer, N., and Thompson, A. (2014). The primary transcriptome of *Salmonella enterica* Serovar Typhimurium and its dependence on ppGpp during late stationary phase. *PloS One* 9 (3), e92690. doi:10.1371/journal.pone.0092690/

Rangarajan, A. A., and Schnetz, K. (2018). Interference of transcription across H-NS binding sites and repression by H-NS. *Mol. Microbiol.* 108, 226–239. doi:10.1111/mmi.13926

Reuter, J. S., and Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11, 129. doi:10.1186/1471-2105-11-129

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., et al. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* 47 (D1), D212–D220. doi:10.1093/nar/gky1077

Scholz, S. A., Diao, R., Wolfe, M. B., Fivenson, E. M., Lin, X. N., and Freddolino, P. L. (2019). High-resolution mapping of the *Escherichia coli* chromosome reveals positions of high and low transcription. *Cell Syst.* 8, 212–225.e9. doi:10.1016/j.cels.2019.02.004

Shavkunov, K. S., Masulis, I. S., Tutukina, M. N., Deev, A. A., and Ozoline, O. N. (2009). Gains and unexpected lessons from genome-scale promoter mapping. *Nucleic Acids Res.* 37, 4919–4931. doi:10.1093/nar/gkp490

Shinhara, A., Matsui, M., Hiraoka, K., Nomura, W., Hirano, R., Nakahigashi, K., et al. (2011). Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. *BMC Genom.* 12, 428. doi:10.1186/1471-2164-12-428

Stern, M. J., Ames, G. F., Smith, N. H., Robinson, E. C., and Higgins, C. F. (1984). Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* 37, 1015–1026. doi:10.1016/0092-8674(84)90436-7

Thomason, M. K., Bischler, T., Eisenbart, S. K., Förstner, K. U., Zhang, A., Herbig, A., et al. (2015). Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *J. Bacteriol.* 197, 18–28. doi:10.1128/JB.02096-14

Thorpe, H. A., Bayliss, S. C., Hurst, L. D., and Feil, E. J. (2017). Comparative analyses of selection operating on nontranslated intergenic regions of diverse bacterial species. Genetics. 206, 363–376. doi:10.1534/genetics.116.195784

Tobes, R., and Ramos, J. L. (2005). REP code: defining bacterial identity in extragenic space. *Environ. Microbiol.* 7, 225–228. doi:10.1111/j.1462-2920.2004.00704.x

Tsai, C. H., Liao, R., Chou, B., Palumbo, M., and Contreras, L. M. (2015). Genome-wide analyses in bacteria show small-RNA enrichment for long and conserved intergenic regions. *J. Bacteriol.* 197, 40–50. doi:10.1128/JB.02359-14

Vincent, H. A., Phillips, J. O., Henderson, C. A., Roberts, A. J., Stone, C. M., Mardle, C. E., et al. (2013). An improved method for surface immobilisation of RNA: application to small non-coding RNA-mRNA pairing. *PLoS One* 8 (11), e79142. doi:10.1371/journal.pone.0079142

Vogel, J., and Wagner, E. G. (2007). Target identification of small noncoding RNAs in bacteria. *Curr. Opin. Microbiol.* 10, 262–270. doi:10.1016/j.mib.2007.06.001

Wachter, S., Raghavan, R., Wachter, J., and Minnick, M. F. (2018). Identification of novel MITEs (miniature inverted-repeat transposable elements) in *Coxiella burnetii*: implications for protein and small RNA evolution. *BMC Genomics* 19, 247. doi:10.1186/s12864-018-4608-y

Waters, L. S., and Storz, G. (2009). Regulatory RNAs in bacteria. *Cell* 136, 615–628. doi:10.1016/j.cell.2009.01.043

Wu, H. Y., Shyy, S. H., Wang, J. C., and Liu, L. F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell* 53, 433–440. doi:10.1016/0092-8674(88)90163-8

Zhi, J., Mathew, E., and Freundlich, M. (1998). *In vitro* and *in vivo* characterization of three major *dad*AX promoters in *Escherichia coli* that are regulated by cyclic AMP-CRP and Lrp. *Mol. Gen. Genet.* 258, 442–447. doi:10.1007/s004380050754

Zhi, X., Dages, S., Dages, K., Liu, Y., Hua, Z. C., Makemson, J., et al. (2017). Transient and dynamic DNA supercoiling potently stimulates the leu-500 promoter in *Escherichia coli*. *J. Biol. Chem.* 292, 14566–14575. doi:10.1074/jbc.M117.794628