SUPPLEMENTAL INFORMATION

Global transcriptional start site mapping using dRNA-seq reveals novel antisense RNAs in *Escherichia coli*

Maureen K. Thomason,[a, ‡,*], Thorsten Bischler,[b,*] Sara K. Eisenbart,[a,b] Konrad U. Förstner,[b] Aixia Zhang,[a] Alexander Herbig,[c] Kay Nieselt,[c] Cynthia M. Sharma,[a,b,#] Gisela Storz[a,#]

Cell Biology and Metabolism Program, Eunice Kennedy Shriver National Institutes of Health, Bethesda, Maryland, USA[a]; Research Centre for Infectious Diseases (ZINF), University of Würzburg, Würzburg, Germany[b]; Integrative Transcriptomics, ZBIT (Center for Bioinformatics Tübingen), University of Tübingen, Tübingen, Germany[c]

[‡]Present address: Department of Microbiology, University of Washington, Seattle, Washington

[*]Joint First Authors.

[#]Address correspondence to Gisela Storz, storzg@mail.nih.gov or Cynthia M. Sharma, cynthia.sharma@uni-wuerzburg.de.

SUPPLEMENTAL MATERIALS AND METHODS

Deep sequencing sample preparation.

RNA extraction. Frozen cell pellets were thawed on ice and resuspended in 880 µl of lysis buffer (0.5 mg/ml lysozyme dissolved in TE pH 8.0, 1% SDS), mixed by inversion and incubated at 65˚C for 2 min or until the samples cleared. The samples were cooled and 88 µl of 1M sodium acetate, pH 5.2 was added along with 1 ml of acid phenol:chloroform (Ambion). Samples were incubated at 65˚C for 6 min with mixing and spun 10 min at 13,000 rpm, 4˚C. The aqueous layer was extracted a second time with chloroform using Phase Lock Gel 2.0 tubes (5Prime) after which the aqueous layer was ethanol precipitated, washed and resuspended in 100 µl of DEPC-H_2O. RNA concentration was measured by reading the absorbance at OD_{260} and the integrity was checked by running ~2 µg aliquots of each sample on a denaturing 1% agarose 1X TBE gel followed by ethidium bromide staining.

DNase I treatment. Total RNA (40 µg) was denatured at 65˚C for 5 min. The RNA was then combined with 1X DNase I buffer + $MgCl_2$ (Fermentas), 20 U of RNase Inhibitor (Invitrogen), and 10 U of DNase I (Fermentas) in a final volume of 100 µl. The mixture was incubated for 45 min at 37˚C and then extracted with phenol:chloroform:isoamylalcohol (Invitrogen) in 2 ml Phase Lock Gel Heavy tubes (5Prime). Samples were precipitated, washed, and resuspended in 40 µl of DEPC-H_2O. RNA concentration and integrity of ~100 ng aliquots were checked as above, and the absence of genomic DNA contamination was confirmed by PCR using primer MK0095 and MK0096.

Terminator exonuclease (TEX) and tobacco acid pyrophosphatase (TAP) treatment. TEX treatment was performed as described previously (1). Briefly, 7 µg of DNase I-treated RNA was denatured for 2 min at 90˚C, cooled on ice for 5 min and combined with 10 U RNase Inhibitor (Invitrogen), 1X Terminator Exonuclease Buffer A (Epicentre), and 7 U of Terminator Exonuclease (Epicentre) in a final reaction volume of 50 µl. Control reactions lacking terminator exonuclease were run in parallel for each sample. Reactions were incubated at 30˚C for 1 h and stopped by the addition of 0.5 µl of 0.5 M EDTA, 50 µl DEPC-H_2O and extraction with phenol:chloroform:isoamylalcohol with Phase Lock Gel 2.0 Tubes. The supernatant was precipitated, washed and resuspended in 11 µl of DEPC-H_2O. RNA concentration was determined by reading the absorbance at OD_{260}. TAP treatment was performed by incubating TEX-treated and untreated control samples with tobacco acid pyrophosphatase (TAP) for 1 h at

37˚C with 1X TAP Buffer (Invitrogen), 10 U RNase Inhibitor and 5 U Tobacco Acid Pyrophosphatase (Invitrogen) in a final reaction volume of 20 µl. The samples were extracted with phenol:chloroform:isoamylalcohol (Invitrogen), precipitated, washed, and resuspended in 20 µl of DEPC-H_2O. RNA concentration was determined by reading the absorbance at OD_{260}, and RNA integrity was checked on a denaturing 4% acrylamide-7M urea gel in 1X TBE and visualized with Stains-all nucleic acid stain (Sigma).

cDNA library construction. cDNA libraries for Illumina sequencing were constructed by vertis Biotechnology AG, Germany (http://www.vertis-biotech.com/) in a strand specific manner as described previously for eukaryotic microRNA (2) but omitting the RNA size-fractionation step prior to cDNA synthesis. In brief, equal amounts of RNA samples were poly(A)-tailed using poly(A) polymerase. Then, the 5'-PPP structures were removed using tobacco acid pyrophosphatase (TAP). Afterwards, an RNA adapter was ligated to the 5´-phosphate of the TAP-treated, poly(A)-tailed RNA. In the case of the GAIIx-libraries, the 5'linker contained the barcode sequence at its 3'end. For HiSeq 2000 libraries, the barcode was introduced in a later step during PCR-amplification of the cDNA library. First-strand cDNA was synthesized by using an oligo(dT)-adapter primer and the M-MLV reverse transcriptase. In a PCR-based amplification step using a high fidelity DNA polymerase the cDNA concentration was increased to 20-30 ng/µl. For all libraries the Agencourt AMPure XP kit (Beckman Coulter Genomics) was used to purify the DNA, which was subsequently analyzed by capillary electrophoresis.

For the GAIIx-libraries, PCR products for sequencing were generated using the following primers designed for amplicon sequencing according to the instructions of Illumina/Solexa:
5'-end_primer
5'-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGACGATCNNNN-3'
3'-end_primer
5'-CAAGCAGAAGACGGCATACGATTTTTTTTTTTTTTTTTTTTTTT-3'
The samples were run on an Illumina GAIIx instrument with 120 cycles in single-read mode.

For the HiSeq2000 libraries, a library-specific barcode for multiplex sequencing was included as part of a 3'-sequencing adapter. The following adapter sequences flank the cDNA inserts:
TrueSeq_Sense_primer

5’-
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC
T-3’
TrueSeq_Antisense_NNNNNN_primer (NNNNNN = 6n barcode for multiplexing)
5’-CAAGCAGAAGACGGCATACGAGAT-NNNNNN-
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC(dT25)-3’
The samples were run on an Illumina HiSeq 2000 instrument with 100 cycles in single-read mode.

Analysis of deep sequencing data.

Read mapping and coverage plot construction. To assure high sequence quality, the Illumina reads in FASTQ format were trimmed with a cutoff phred score of 20 by the program fastq_quality_trimmer from FASTX toolkit version 0.0.13. After trimming, poly(A)-tail sequences were removed and a size filtering step was applied in which sequences shorter than 12 nt were eliminated. The collections of remaining reads were mapped to the *E. coli* MG1655 genome (NCBI Acc.-No: NC_000913.2; Jun 24, 2004) using our RNA-seq pipeline READemption (3) and *segemehl* (4) with an accuracy cutoff of 95%. Coverage plots representing the numbers of mapped reads per nucleotide were generated. Reads that mapped to multiple locations contributed a fraction to the coverage value. For example, reads mapping to three positions contributed only 1/3 to the coverage values. Each graph was normalized to the number of reads that could be mapped from the respective library. To restore the original data range, each graph was then multiplied by the minimum number of mapped reads calculated over all libraries.

Normalization of expression graphs. Prior to the comparative analysis, the expression graphs with the cDNA coverages that resulted from the read mapping were further normalized. A percentile normalization step was applied to normalize the +TEX graphs. To this end, the 90^{th} percentile of all data values was calculated for each +TEX graph. This value was then used to normalize the +TEX graph as well as the respective −TEX graph. Thus, the relative differences between each +TEX and −TEX graph were not changed in this normalization step. Again, all graphs were multiplied with the overall lowest value to restore the original data range. To account for different enrichment rates, a third normalization step was applied. During this step, prediction of TSS candidates was performed for each replicate of each strain. These candidates

were then used to determine the median enrichment factor for each +/−TEX library pair. Using these medians all −TEX libraries were then normalized against the library with the strongest enrichment. Besides annotation of transcriptional start sites, the resulting graphs were also used for visualization in the Integrated Genome Browser (5).

Correlation analysis. To assess similarity between different libraries, nucleotide and gene-wise Spearman and Pearson correlation coefficients were calculated based on concatenated values of forward and reverse strand position-wise coverage files and visualized in a correlation matrix for both the +TEX and –TEX libraries using the R package *corrplot*. Gene-wise correlation values utilized read overlap counts based on NCBI annotations (Acc.-No: NC_000913.2). Each read with a minimum overlap of 10 nt was counted with a value based on the number of locations where the read was mapped. If the read overlapped more than one annotation, the value was divided by the number of regions and counted separately for each region (e.g. 1/3 for a read mapped to 3 locations).

Transcriptional start site (TSS) annotation. Based on the normalized expression graphs we conducted automated TSS prediction in a similar manner as described in Dugar *et al.* 2013 (6) utilizing TSSpredator (http://it.inf.uni-tuebingen.de/TSSpredator). In brief, for each position (i) in the expression graph corresponding to the TEX treated libraries, the algorithm calculates an expression height, e(i), and compares that expression height to the preceding position by calculating e(i) – e(i-1), which is termed the flank height. Additionally, the algorithm calculates a factor of height change e(i)/e(i-1). To determine if a TSS is a primary TSS and not a processed transcript end an enrichment factor is calculated as $e_{+TEX}(i)/e_{-TEX}(i)$, where $e_{+TEX}(i)$ is the expression height for the terminator exonuclease treated sample and $e_{-TEX}(i)$ is the expression height for the untreated sample. For all positions where these parameters exceed the predefined thresholds a TSS is annotated.

We set the thresholds for the "minimum flank height" and the "minimum factor of height change" which are used to determine if a TSS is "detected" to 0.3 and 2.0, respectively. Here, the value for the "minimum flank height" is a factor to the minimum 90^{th} percentile over all libraries resulting in an absolute value of 1.62. If the TSS candidate reaches these thresholds in at least one replicate of one condition, the thresholds are decreased for the other replicates to 0.1 (0.54 absolute) and 1.5, respectively. Furthermore, we set the "matching replicates" parameter which determines the number of replicates in which a TSS must exceed these thresholds in order to be

marked as “detected” within a condition to 2 for M63 0.4 and LB 0.4 and to 3 for LB 2.0. If a TSS was detected in a certain condition, the lowered thresholds also apply to all remaining libraries of the other conditions. Furthermore, we consider a TSS candidate to be enriched in a condition if the respective enrichment factor for at least one replicate is not less than 2.0. A TSS candidate has to be enriched in at least one condition and is discarded otherwise. If a TSS candidate is not enriched in a condition but still reaches the other thresholds it is only indicated as “detected”. However, a TSS candidate can only be labeled as detected in a condition if its enrichment factor is above 0.66. Otherwise we consider it to be a processing site. In order to take into account slight variations between TSS positions the respective parameters for clustering between replicates and conditions were set to a value of 1. In doing so a consensus TSS position in a three nucleotide window is determined based on the maximum “flank height” among the respective libraries. The same parameters were recently used in our comparative dRNA-seq analysis of multiple *Campylobacter jejuni* strains (6).

Comparison to Database of prOkaryotic OpeRons (DOOR). A table containing all operon annotations (1,526 single gene operons and 851 operons consisting of multiple genes) was downloaded from the DOOR 2.0 website (7) (on Sep 18, 2013). Two genes (b0816 and b1470) not included in the NCBI annotations used for the TSS prediction were discarded. We divided our genes with predicted primary TSS (2,672) into two groups. The first consisted of genes (2,057) for which the TSS was classified only as primary (1,707) or primary and antisense (350). The second consisted of genes (615) with a primary and internal TSS (611) or primary, internal and antisense (4). Furthermore, 231 genes from both groups lacking DOOR locus tags and information regarding operon structure were excluded from the analysis. This resulted in a final set of 2,441 TSS that were considered for the DOOR overlap: group one contained 1,562 primary and 566 primary and secondary and group two contained 309 primary and antisense and 4 primary, internal and antisense. The overlap was calculated by comparing the primary TSS-associated genes of either group to the set of single-standing genes and first genes in an operon from DOOR.

Analysis of iTSS localization. The 5,574 iTSS were split into two groups, iTSS that were also annotated as pTSS or sTSS (968) and the remaining iTSS (4,606) where iTSS in both groups can also be annotated as asTSS. Each gene in which an iTSS was detected was split into 10 equal-sized sections ordered from 5′ to 3′ end (the first section covers the first 10% of the

gene, the second section the second 10%, etc.) and the number of iTSS localizing to each section was counted over all genes. Histogram plots were generated to visualize the distribution of iTSS in each group.

Comparison of expression under different growth conditions. Expression values for TSS (based on a 50 nt window downstream of the TSS position) were calculated as described above for the gene-wise correlation analysis but using only libraries from the M63 0.4 and LB 0.4 conditions. Differential expression between these two conditions was assessed for pTSS only, iTSS only and asTSS only based on all replicates using DESeq2 (8). TSS with an adjusted p-value ≤ 0.05 were defined as differentially expressed. A Fisher exact test based on all TSS in the three classes was performed to determine if there is over- or underrepresentation of any class.

Detection of promoter motifs. Sequences from −50 to +1 as well as from −50 to +5 relative to the TSS were extracted for pTSS only, iTSS only and asTSS only and analyzed using MEME (9). For the −50 to +1 sequences a motif of length 48 nt was predicted to allow a distance of 0 to 3 nt to the TSS position. For the −50 to +5 sequences a motif length of 56 nt was applied.

Identification of overlapping 5' UTRs. Based on all primary and secondary TSS (see Data Set S1) all possible pairs of overlapping 5' UTRs with the first UTR on the forward and second one on the reverse strand were computed with the restriction that the overlapping region must have a minimum length of 10 nucleotides. The data for this is depicted in Data Set S2A.

Comparison of asRNAs detected in our and previous studies. To compare the annotations of previously detected antisense RNAs with our predictions, asRNA annotations were retrieved from the Supplemental Materials and Methods sections from previously published studies (10-12) or were downloaded from RegulonDB (13, 14). For the Raghavan *et al.* data, 90 asTSS were identified antisense to distinct genes which, combined with the gene names, was used to infer strand information for the asTSS (11). For the Shinhara *et al.* data, 229 novel candidate sRNAs were assembled from the mapped sequencing reads (12). From these reads we extracted the TSS positions of 112 candidate sRNAs labeled as "cis-antisense". For the Dornenburg *et al.* data, we used all 1,005 putative TSS located antisense to genes and 385 putative TSS located antisense to predicted untranslated regions (10). The TSS data obtained from RegulonDB consists of one data set with 1,490 TSS from which we extracted 165 TSS described as antisense to a gene (13) (Data set version 2.0), and a second set which includes 5,197 single TSSs and TSS clusters of varying length (14) (Data set version 3.0). We selected 182 single TSS and TSS clusters defined as either

located antisense to a gene or to be "convergent", i.e. the TSS is located sense to one gene and antisense to another. In these cases, we used the specific nucleotide position for a single TSSs and the respective region for a TSS cluster. Additionally, we included set of 89 asRNA predicted by Conway *et al*. (15). We compared the asTSS from each data set, including our 6,379 predicted asTSS, to the asTSS of all other data sets in a pair-wise manner, requiring either a precise match of the annotated positions or allowing a variation of 1, 2, 3 or 10 nt.

Comparison of asTSS to IP-dsRNAs. IP-dsRNAs were extracted from Lybecker *et al.* (16). Afterwards, all dsRNAs assigned to the category "Convergent" were excluded since overlapping 3'UTRs might not be covered in our dRNA-seq analysis as our protocol sequences from the 5' ends of genes. For the comparison we used all TSS annotated as asTSS including the ones also assigned to other classes. An overlap is reported if an asTSS is annotated in a region between 10 nucleotides upstream of the IP-dsRNA 5' end and 10 nucleotides upstream of the dsRNA 3' end on at least one strand.

Northern analysis. The oligonucleotides used to create the riboprobes are listed in Table S2. To synthesize the riboprobe, 3.5 µg of gene-specific PCR product was added to a reaction mix containing 5 µl of α-^{32}P-UTP, 1X T7 RNA polymerase buffer, 2000 U/ml T7 RNA polymerase (NEB), 20 U RNasin, 4 mM DTT (Invitrogen), 0.16 mg/ml BSA (NEB), 0.4 mM GTP, CTP, and ATP, and 0.01 mM UTP (Invitrogen) in a total volume of 25 µl. After 1 h incubation at 37˚C, 2000 U/ml of T7 RNA polymerase was added, and the samples were incubated an additional 1 h at 37˚C. After a final incubation with 1 U/µl of DNase I (Fermentas) at 37˚C for 15 min, probes were purified using G50 columns (GE Healthcare). RNA samples were separated on denaturing 8% acrylamide-7M urea gels, transferred and UV crosslinked to Zeta Probe GT membranes (Bio-Rad) as before (17). The membranes were subsequently incubated for 2 h at 50˚C in 20 ml hybridization buffer (50% formamide, 1.5X SSPE, 1% SDS, 0.5% dry milk), after which the hybridization buffer was exchanged for 20 ml of fresh buffer. Riboprobes were denatured with 100 µl of 10 mg/ml yeast RNA at 95˚C for 4 min, added to the membrane and left to hybridize overnight at 50˚C. The membranes were subsequently rinsed once with 2X SSC + 0.1% SDS, washed once for 20 min at 50˚C with 2X SSC + 0.1% SDS, and twice for 20 min at 50˚C with 0.1X SSC + 0.1% SDS. The membranes were rinsed two more times with 0.1X SSC + 0.1% SDS, allowed to dry, and exposed to Hyperfilm (Amersham) at −80˚C.

SUPPLEMENTAL FIGURES

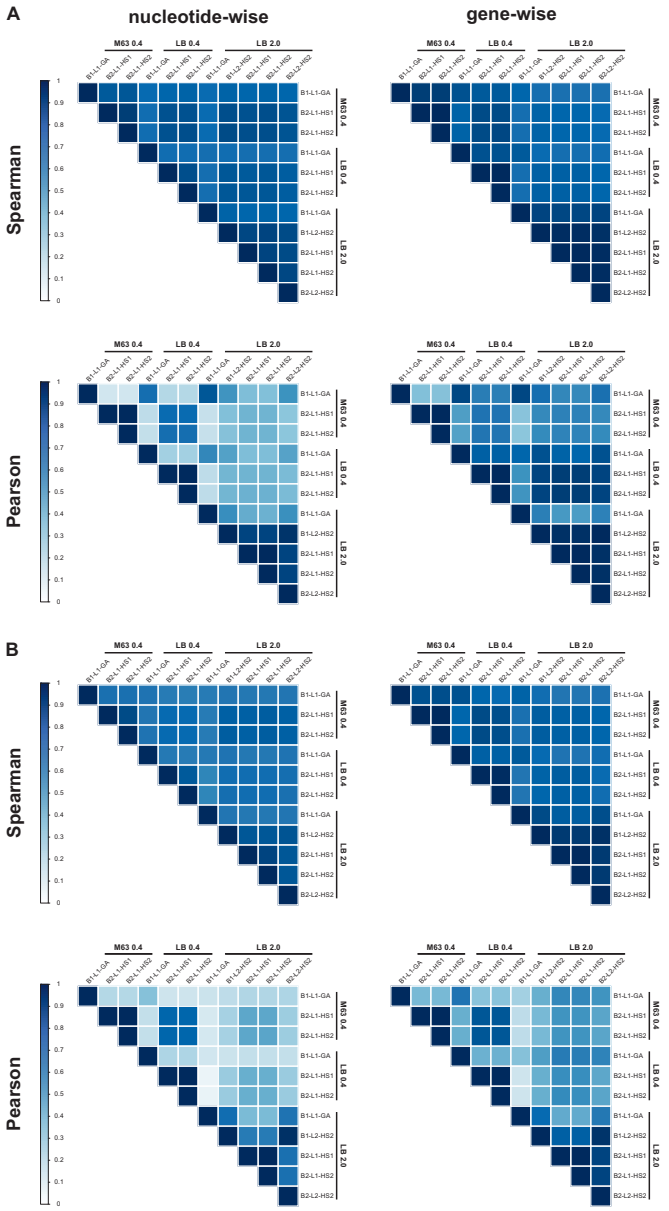

FIG. S1. Reproducibility of cDNA coverage among –TEX (A) and +TEX (B) cDNA libraries. Nucleotide-wise Spearman and Pearson correlation values for all possible combinations of library pairs from all growth conditions were based on expression values from both strands. Gene-wise Spearman and Pearson correlation values for all possible combinations of library pairs were based on expression values for genes as annotated by NCBI. A legend for the extent of correlation is given on the left

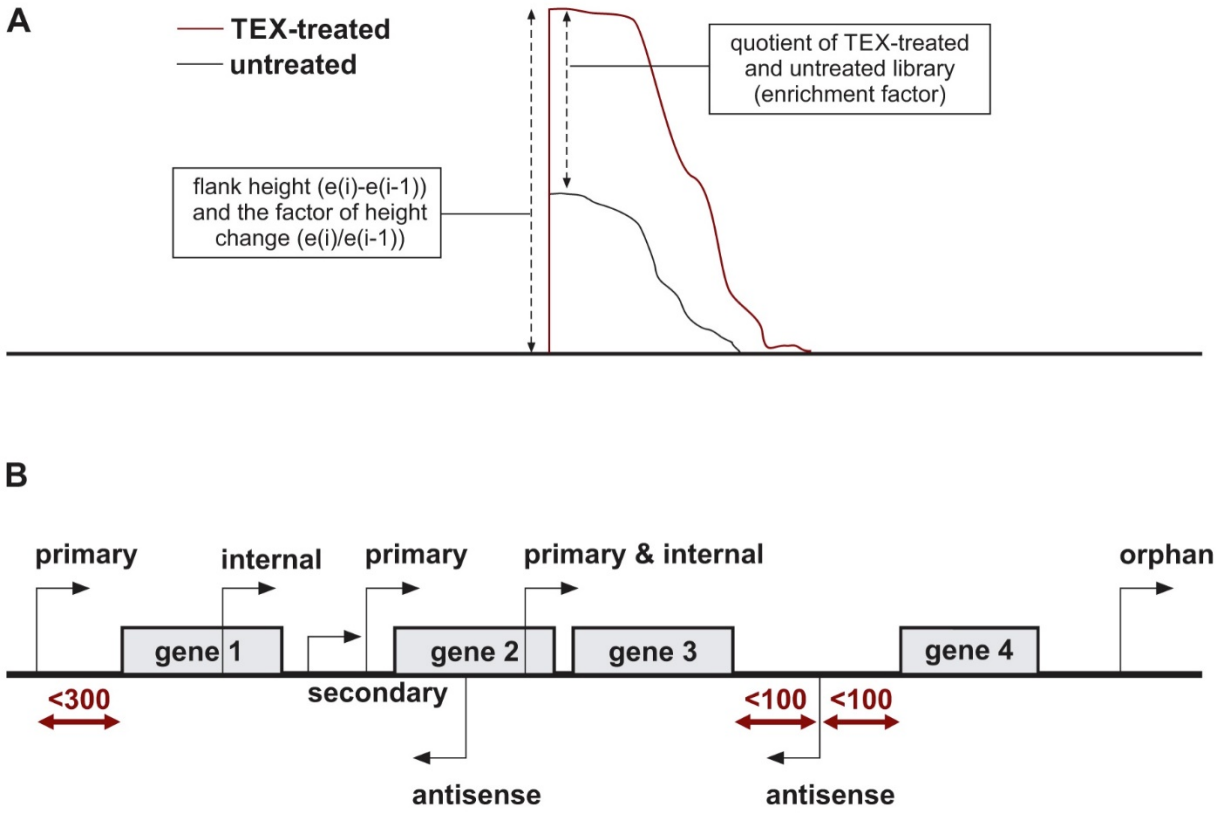

FIG. S2. TSS prediction parameters and classification. (A) Schematic representation of the criteria used for TSS prediction. (B) The different TSS classes (primary, secondary, internal, antisense, and orphan) are depicted according to their location relative to annotated genes. The height of the black arrows depicts differences in expression strength while the distance cutoffs for flanking genes are shown in red. The figures are adapted from (6).

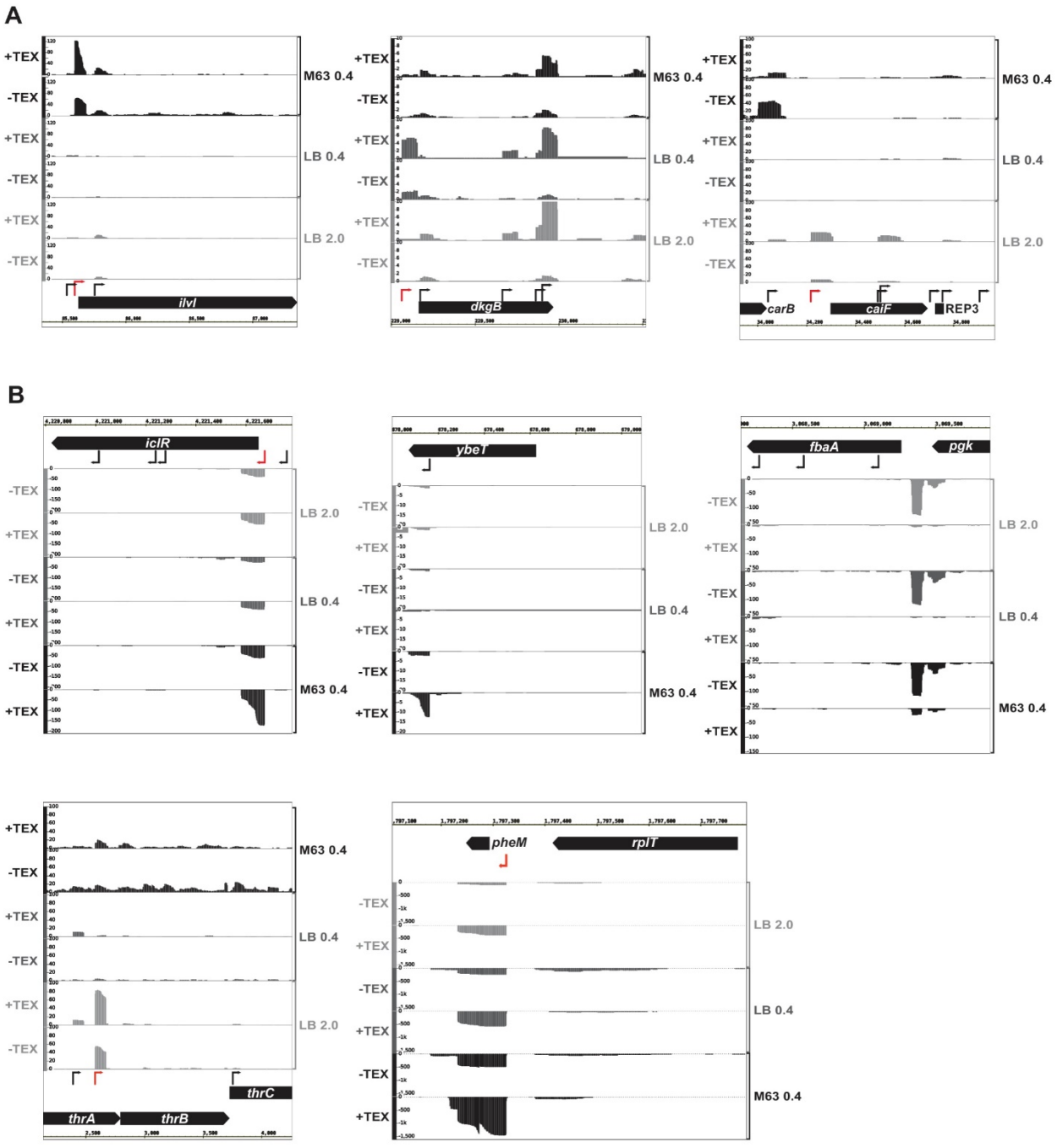

FIG. S3. cDNA coverage plots for examples of TSS detection. (A) Examples of TSS that were detected in only one of three conditions. Each panel shows a primary TSS (indicated by red arrow) that was detected only in one of the three examined growth conditions. (B) Examples of primary TSS (indicated by red arrows) detected in our study that agree or disagree with annotation from the DOOR database (7). The primary TSS detected for the *iclR* gene is in agreement with its annotation as a single gene transcription unit. We did not detect a primary

TSS for *ybeT* and *fbaA*. However, enrichment in the –TEX libraries for *fbaA* indicates the presence of an upstream processing site. The *thrB* gene, located within an operon, was assumed to not have a primary TSS. However, we detect an internal TSS within the upstream *thrA* gene that serves as a primary TSS for *thrB*. A primary TSS upstream of *pheM*, in the *rplT-pheM* operon, indicates *pheM* could be independently transcribed. Screenshots were taken for the B2 L1 HS1 libraries.

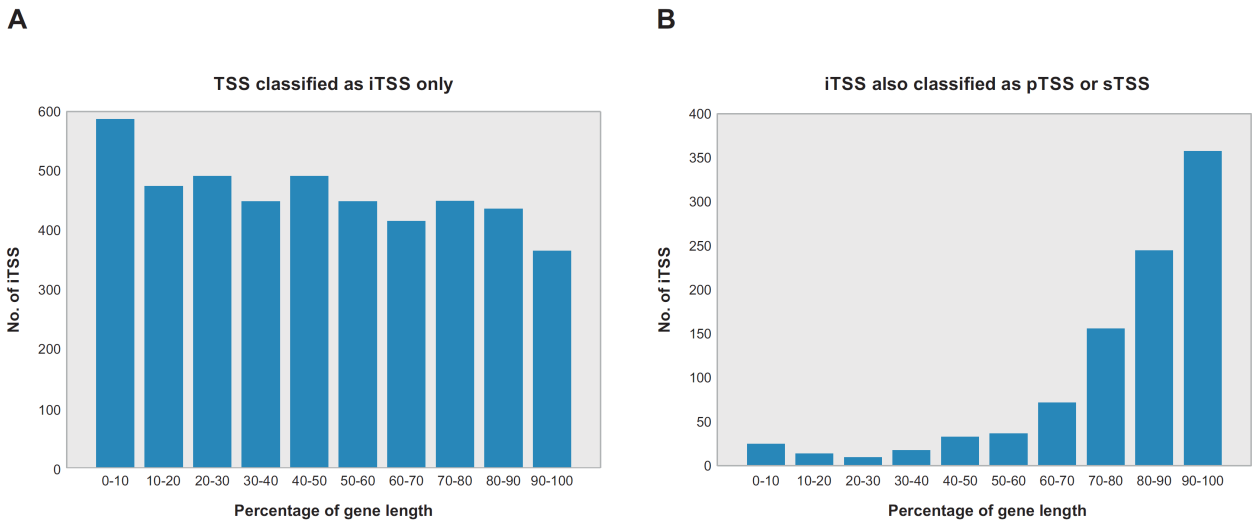

FIG. S4. Location of internal TSS (iTSS) relative to gene annotations. The relative location was determined for all TSS that are only classified as iTSS (A) or for iTSS that are also pTSS or sTSS (B). Each gene in which an iTSS was detected was divided into 10 equal sections and the number of iTSS located in each section was counted over all genes. Eight iTSS in (A) are internal to two distinct but overlapping gene annotations.

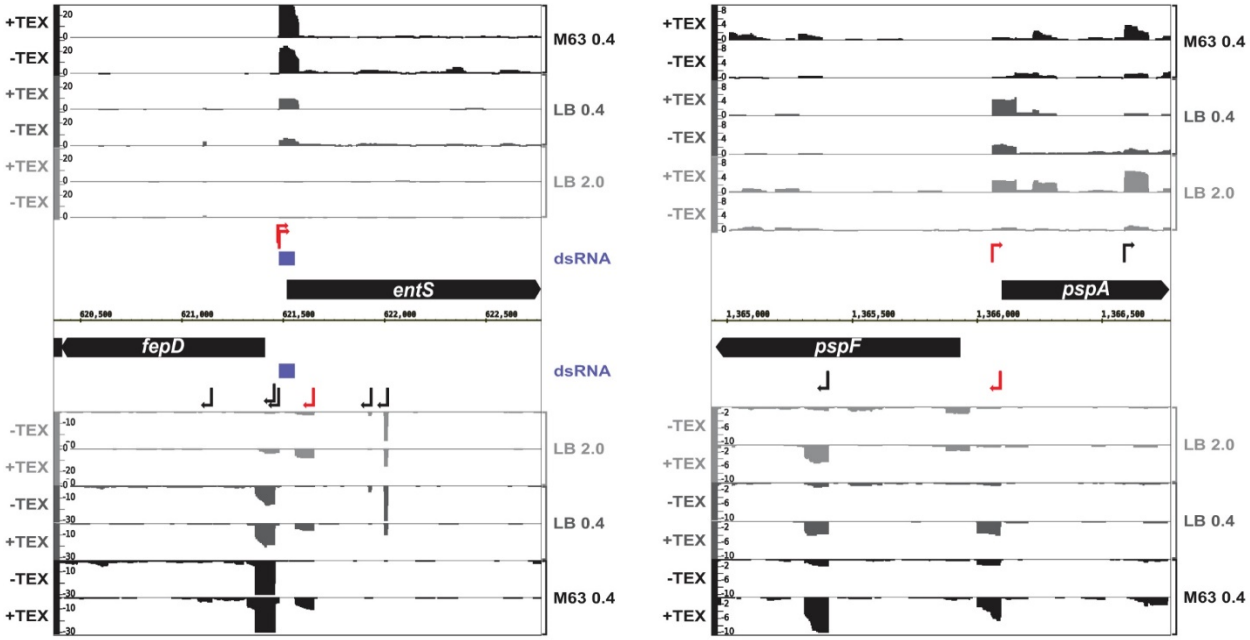

FIG. S5. cDNA coverage plots for examples of overlapping 5' UTRs. The TSS for overlapping 5' UTRs of divergently transcribed gene pairs *entS/fepD* and *pspA/pspF* are indicated in red. The locations of the annotated IP-dsRNAs reported by Lybecker *et al.* (16) are shown in blue where present. Screenshots were taken for the B2 L1 HS1 libraries.

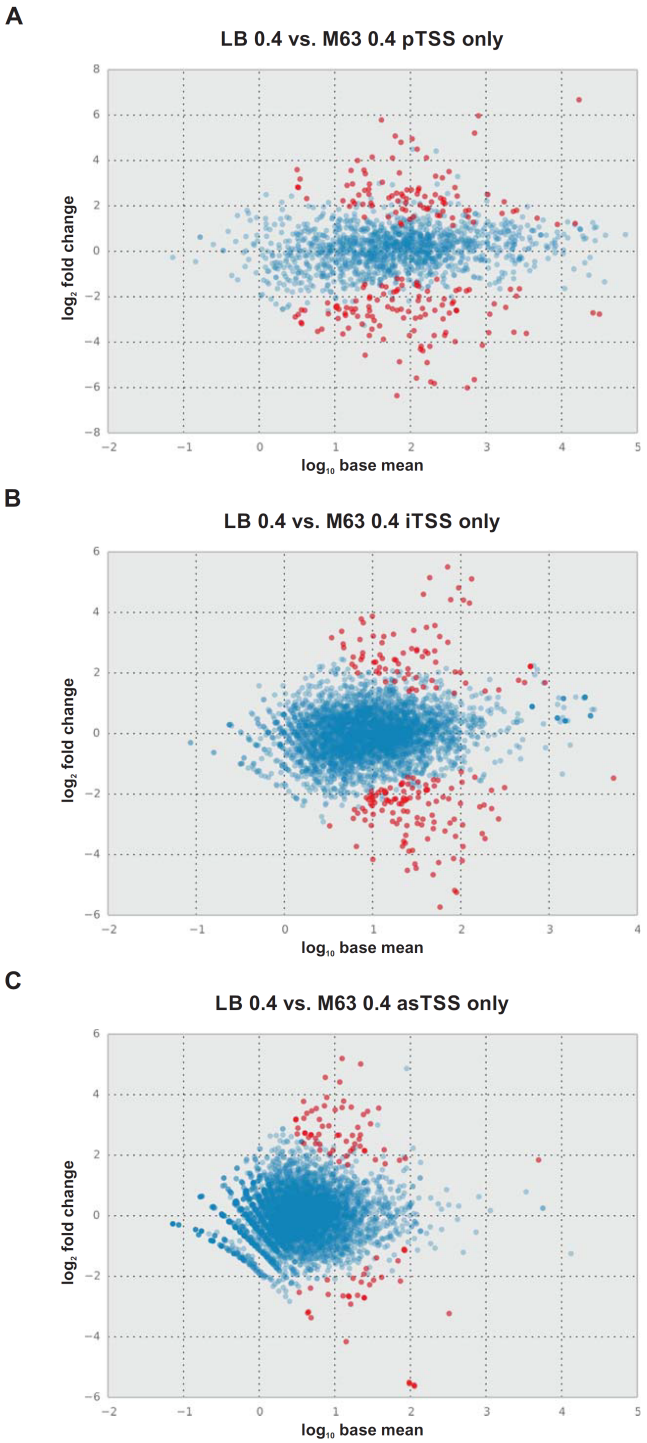

FIG S6. Comparison of mean expression levels and differential expression of TSS classified as pTSS only, iTSS only and asTSS only in LB 0.4 and M63 0.4. There is an overrepresentation of pTSS that show regulation (odd ratio 3.06, p-value: 3.75×10^{-32}) and underrepresentation of asTSS that show regulation (odd ratio 0.41, p-value: 2.98×10^{-18}). The iTSS show neither of these tendencies (odd ratio: 1.02, p-value: 0.79).

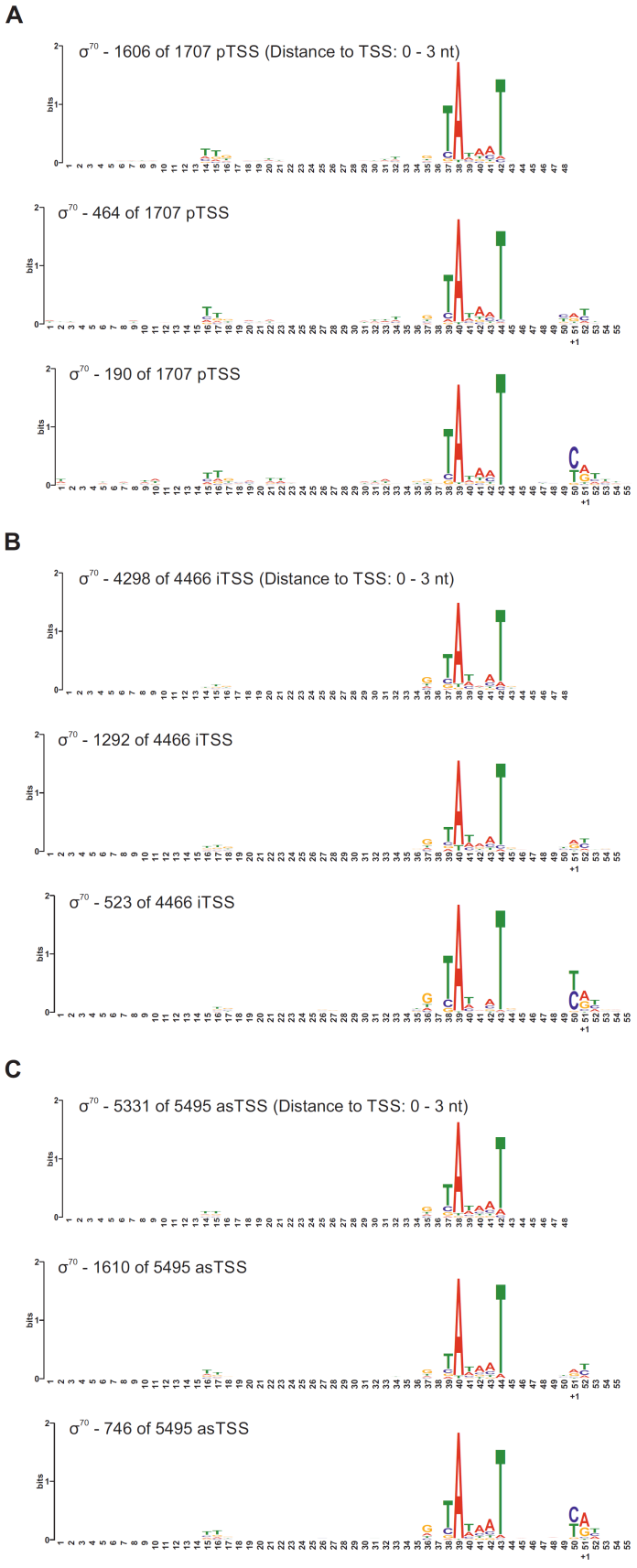
A
σ70 - 1606 of 1707 pTSS (Distance to TSS: 0 - 3 nt)
σ70 - 464 of 1707 pTSS
σ70 - 190 of 1707 pTSS
B
σ70 - 4298 of 4466 iTSS (Distance to TSS: 0 - 3 nt)
σ70 - 1292 of 4466 iTSS
σ70 - 523 of 4466 iTSS
C
σ70 - 5331 of 5495 asTSS (Distance to TSS: 0 - 3 nt)
σ70 - 1610 of 5495 asTSS
σ70 - 746 of 5495 asTSS
bits
+1

FIG S7. Promoter motifs for pTSS, iTSS and asTSS. The motif search was conducted in sequences extracted based on the 1,707 pTSS only (A), 4,466 iTSS only (B) and 5,495 asTSS only (see Supplemental Materials and Methods). The first motif in A, B and C was predicted for sequences ranging from position −50 to +1 relative to the TSS while the second and third represent the two top-scoring motifs based on sequences ranging from position −50 to +5. The first motif in A,B and C shows a canonical σ^{70} -10 sequence found in almost all sequences of the respective TSS class while the second and third motifs reveal a subset of TSS showing a slight enrichment for a pyrimidine at position −1, a purine at position +1 and a pyrimidine at position +2 as described previously (18).

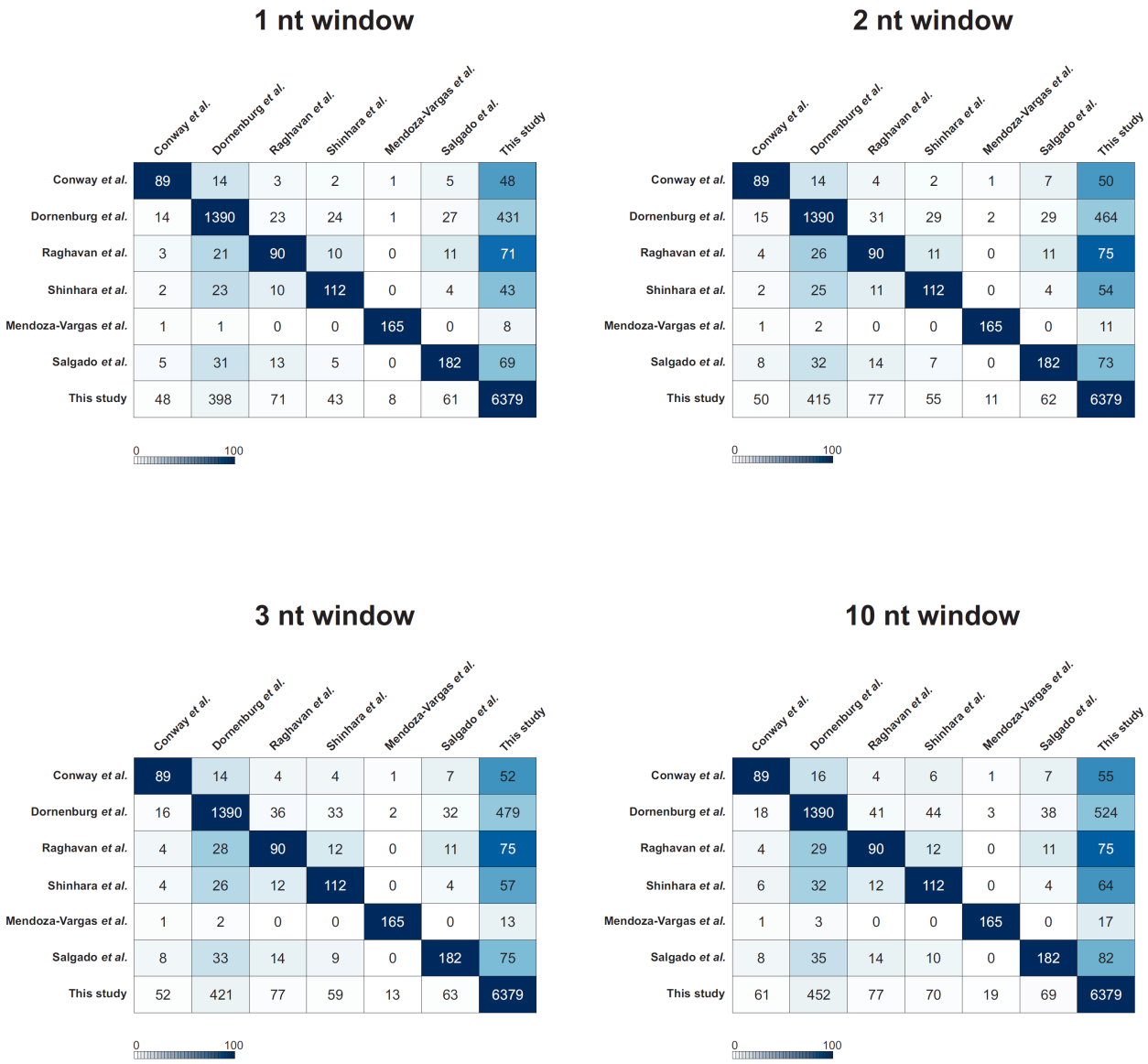

FIG. S8. Pair-wise comparisons of asTSS from this and other studies. The total numbers of annotated asTSS are shown on the main diagonal of the matrix. asTSS positions from the studies in the rows are compared to asTSS positions from the studies in the columns and the number of overlapping asTSS positions within a maximum distance of 1, 2, 3 and 10 nt is listed in the respective matrix entries. Differences in the number of matching TSS for a given pair of studies,

if either one or the other study is used as the basis for comparison, can be explained by cases where unequal numbers of TSS are matched (for example, a single TSS in study I matches several TSS from study II located in close proximity). Background color depicts the percentage of overlapping asTSS relative to the total number of asTSS from the respective study in the row.

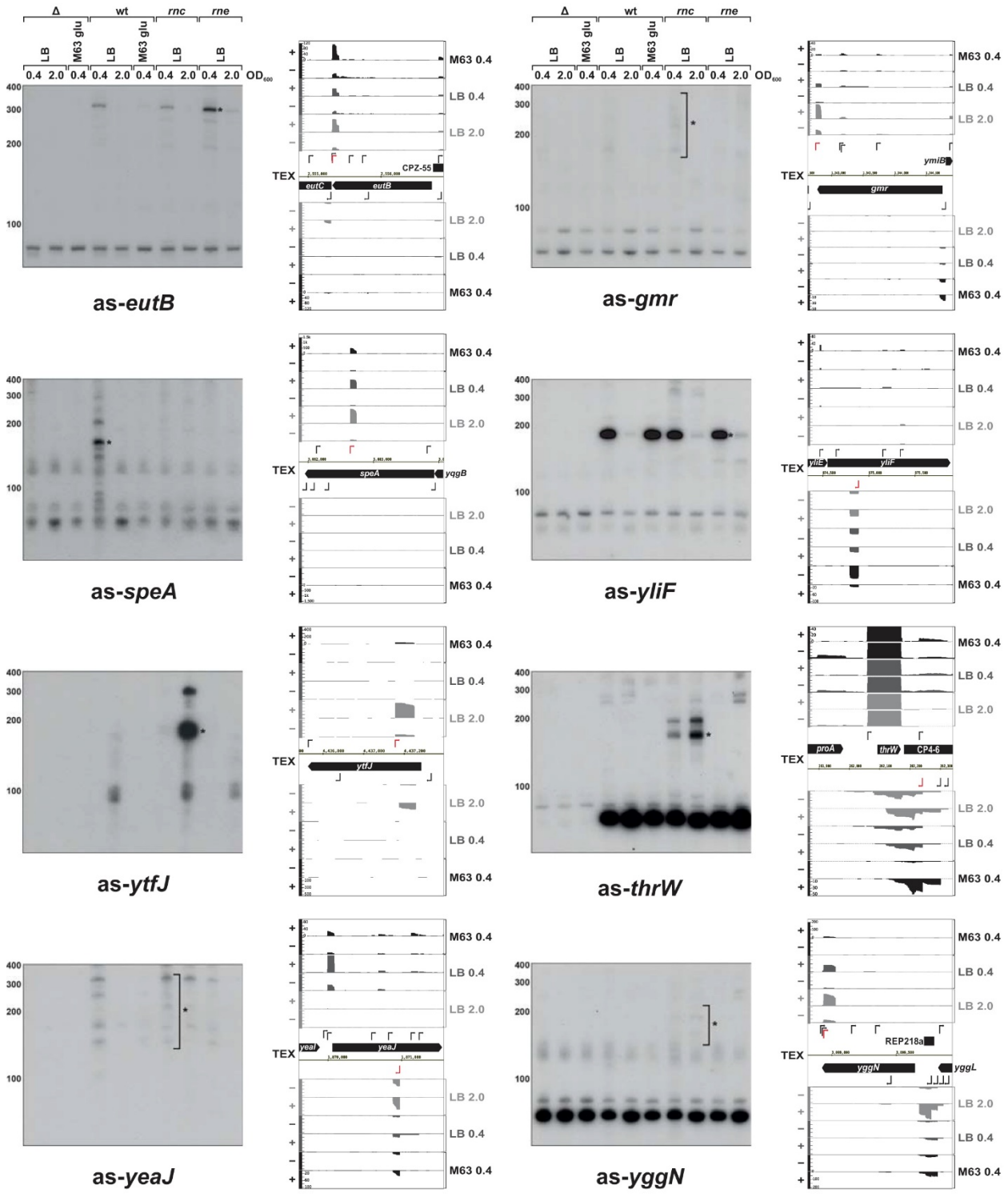

FIG. S9. Northern analysis for asRNA candidates. Northern blots and corresponding cDNA coverage plots of the –/+ TEX libraries for the three growth conditions are shown for the tested

asRNA candidates: as-*eutB*, as-*speA*, as-*ytfJ,* as-*yeaJ*, as-*gmr*, as-*yliF*, as-*thrW,* and as-*yggN*. Black stars indicate the primary bands detected for each asRNA candidate. Red arrows in the coverage plots indicate the positions of the asTSS relative to the corresponding sense gene.

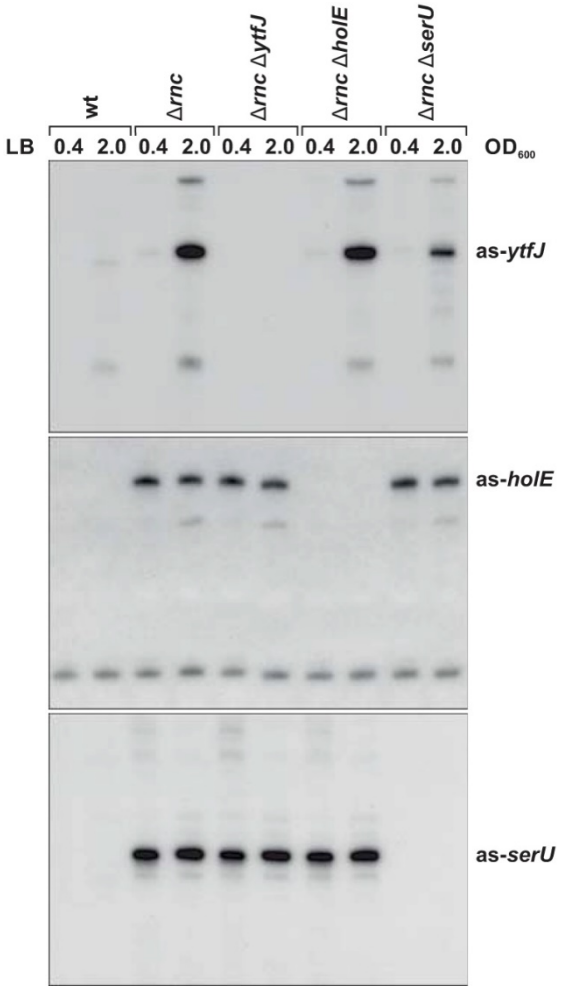

FIG. S10. Northern analysis for control *rnc* mutant strains. Strains deleted for both the *rnc* and the as-*ytfJ,* as-*holE* or as-*serU* loci were grown in LB to the indicated OD_{600} and northern analysis was performed as in Fig. 5 and Supplemental Fig. S9.

SUPPLEMENTAL TABLES

TABLE S1. Strains used in this study.

Name	MPK number	Genotype	Source
MG1655		*E. coli* F- λ- *ilvG- rfb-50 rph-1*	lab stock
NM500		MG1655 *mini-λ:tet*	N. Majdalani
NB478	MPK0331	MG1655 Δ*rnc::cat*	(19)
EM1377	MPK0330	MG1655 Δ*lac X174, rne131 zce-726::Tn10*	(20)
GSO659	MPK0310	MG1655 Δ*as-eutB::kan*	This study
GSO660	MPK0313	MG1655 Δ*as-ymfL::kan*	This study
GSO661	MPK0298	MG1655 Δ*as-qorA::kan*	This study
GSO662		MG1655 Δ*as-argR::kan*	This study
GSO663		MG1655 Δ*as-gmr::kan*	This study
GSO664		MG1655 Δ*as-holE::kan*	This study
GSO665		MG1655 Δ*as-yggN::kan*	This study
GSO666		MG1655 Δ*as-yliF::kan*	This study
GSO667		MG1655 Δ*as-speA::kan*	This study
GSO668		MG1655 Δ*as-gsiB::kan*	This study
GSO669		MG1655 Δ*as-yeaJ::kan*	This study
GSO670		MG1655 Δ*as-serU::kan*	This study
GSO671		MG1655 Δ*as-thrW::kan*	This study
GSO672		MG1655 Δ*as-ytfJ::kan*	This study
GSO718		MG1655 Δ*rnc::cat,* Δ*as-holE::kan*	This study
GSO719		MG1655 Δ*rnc::cat,* Δ*as-ytfJ::kan*	This study
GSO673		MG1655 Δ*rnc::cat,* Δ*as-serU::kan*	This study
GSO674		MG1655 Δ*rnc::cat,* Δ*as-thrW::kan*	This study

TABLE S2. Oligonucleotides used in this study.

Number	Sequence	Use
MK0095	CTTCCATGGCGTCAAGAAACAGC	Forward primer for gDNA contamination PCR check; use with MK0096
MK0096	GTTCTCAGCCTGTATCAGTCT	Reverse primer for gDNA contamination PCR check; use with MK0095
Oligos used for antisense deletion construction		
MK0383	GAACGAGAGTGATCGGCCAGGAAACGCAG CAGCGCCTGCGGTGTAGGCTGGAGCTGCT TC	Forward primer for deletion of antisense *eutB* strain construction; use with MK0399
MK0399	AAGGCCAATACCACCATCGGTATTCCGGGC ACCTTTAGCGATTCCGGGGATCCGTCGACC	Reverse primer for deletion of antisense *eutB* strain construction; use with MK0383
MK0385	CACACTGCGCTTTCAGCGCTTCAACAG	Forward primer for antisense *eutB* PCR check; use with MK0400
MK0400	AGCATCAGCGAACTGCGTGAG	Reverse primer for antisense *eutB* PCR check; use with MK0385
MK0415	CGGCTCCCAACGAATGACTCTGACGGGCA CTCCGTAGTGAGTGTAGGCTGGAGCTGCTT C	Forward primer for deletion of antisense *ymfL* strain construction; use with MK0416
MK0416	TTGCCGAGTGGTTACGCTGAAGCGGCTGA CTGGCTCGATGATTCCGGGGATCCGTCGA CC	Reverse primer for deletion of antisense *ymfL* strain construction; use with MK0415
MK0417	GCCAGCACACACCTTCGTCATTACT	Forward primer for antisense *ymfL* PCR check; use with MK0418
MK0418	GACTTAAGACCGGATATCTATC	Reverse primer for antisense *ymfL* PCR check; use with MK0417
MK0387	GCCGTGGTTACTACTGGCGATTGCGGTGGT GTGGGTTATTGTGTAGGCTGGAGCTGCTTC	Forward primer for deletion of antisense *qor* strain construction; use with MK0388
MK0388	ACCACGCGGGAGGAATTAACCGAGGCCAG TAATGAACTGATTCCGGGGATCCGTCGACC	Reverse primer for deletion of antisense *qor* strain construction; use with MK0387
MK0389	GTCATGCTGATGGTCACCGGCG	Forward primer for antisense *qor* PCR check; use with MK0389
MK0390	CTGCAACGCCGCGGCTTAATGGTCAG	Reverse primer for antisense *qor* PCR check; use with MK0390
AZ1312	GTACGCTTAATGCAGCAACAGTGGGTGTAG GCTGGAGCTGCTTC	Forward primer for deletion of antisense *argR* strain construction; use with AZ1313
AZ1313	GAAATGGTTTACTGCCTGCCAGCTATTCCG GGGATCCGTCGACC	Reverse primer for deletion of antisense *argR* strain construction; use with AZ1312
AZ1314	CGTTGCGCCTTTCTCTTCCGC	Forward primer for antisense *argR* PCR check; use with AZ1315
AZ1315	GCTCGGCTAAGCAAGAAGAACTAG	Reverse primer for antisense *argR* PCR check; use with AZ1314
AZ1282	GCGCCGGGCATCCTCAAATAGGTGTAGGC	Forward primer for deletion of antisense *gmr* strain

	TGGAGCTGCTTC	construction; use with AZ1283
AZ1283	CAACAATTTAGCCAACTAGGTGCGCATTCC GGGGATCCGTCGACC	Reverse primer for deletion of antisense *gmr* strain construction; use with AZ1282
AZ1284	GAAACTCCAGTGGCTTTTGCCAG	Forward primer for antisense *gmr* PCR check; use with AZ1285
AZ1285	CTGCACGTCAGCTCGCCG	Reverse primer for antisense *gmr* PCR check; use with AZ1284
AZ1300	GAGAGATCGGGTGGGGCA GGTGTAGGCTGGAGCTGCTTC	Forward primer for deletion of antisense *holE* strain construction; use with AZ1301
AZ1301	CGTAGCGAAGGGAGCGTGCATTCCGGGGA TCCGTCGACC	Reverse primer for deletion of antisense *holE* strain construction; use with AZ1300
AZ1302	GACATGCACCATGACTCTGATGG	Forward primer for antisense *holE* PCR check; use with AZ1303
AZ1303	CACCACTGAATCCTGTTTCAACACC	Reverse primer for antisense *holE* PCR check; use with AZ1302
AZ1306	CGCCAACTTTACCCACTGTGT AGGTGTAGGCTGGAGCTGCTTC	Forward primer for deletion of antisense *yggN* strain construction; use with AZ1307
AZ1307	GCAGCAAAATGCGCAGCCGTCATTCCGGG GATCCGTCGACC	Reverse primer for deletion of antisense *yggN* strain construction; use with AZ1306
AZ1308	GAAGCCGTCGGTCTTATCGCAG	Forward primer for antisense *yggN* PCR check; use with AZ1309
AZ1309	CGGTAAGCAATATTCCCTGAATGCC	Reverse primer for antisense *yggN* PCR check; use with AZ1308
AZ1288	CGCAAATACATACAATCCGGTCGG CGTGTAGGCTGGAGCTGCTTC	Forward primer for deletion of antisense *yliF* strain construction; use with AZ1289
AZ1289	CATTCTCCGCCTGGGATAAAAGTGGATTCC GGGGATCCGTCGACC	Reverse primer for deletion of antisense *yliF* strain construction; use with AZ1288
AZ1290	GCCCGCCATCCAGTACGCG	Forward primer for antisense *yliF* PCR check; use with AZ1291
AZ1291	CGGTACTGCGGGAAAAATTGTGCG	Reverse primer for antisense *yliF* PCR check; use with AZ1290
AZ1261	GTGGTCGATAGCACCGTCAGAGTGTGTAG GCTGGAGCTGCTTC	Forward primer for deletion of antisense *speA* strain construction; use with AZ1262
AZ1262	CAGTGCTTCGACGTCGGCGGATTCCGGGG ATCCGTCGACC	Reverse primer for deletion of antisense *speA* strain construction; use with AZ1261
AZ1263	CGAACACGTCAACCGCTTCGGTA	Forward primer for antisense *speA* PCR check; use with AZ1264
AZ1264	GTTGAAACCCTGCGTGAAGCCG	Reverse primer for antisense *speA* PCR check; use with AZ1263
AZ1255	CTGCGCGGTGCTGTGGTTATGGTGTAGGCT GGAGCTGCTTC	Forward primer for deletion of antisense *gsiB* strain construction; use with AZ1256
AZ1256	CGGAAGCGATCGATCCGACAACATTCCGG GGATCCGTCGACC	Reverse primer for deletion of antisense *gsiB* strain construction; use with AZ1255

AZ1257	GGTTGAAGCCGACCAGCCAG	Forward primer for antisense *gsiB* PCR check; use with AZ1258
AZ1258	GTGCTGGCGGAGAGTTATACCG	Reverse primer for antisense *gsiB* PCR check; use with AZ1257
AZ1273	CGAATCGACTGTTTAATCGCCTGAGGTGTA GGCTGGAGCTGCTTC	Forward primer for deletion of antisense *yeaJ* strain construction; use with AZ1274
AZ1274	GTCCTTCCTGCAGTCAGGAAGTAATTCCGG GGATCCGTCGACC	Reverse primer for deletion of antisense *yeaJ* strain construction; use with AZ1273
AZ1275	CCGGAAGAGAAGCCGATCTCTTTC	Forward primer for antisense *yeaJ* PCR check; use with AZ1276
AZ1276	GCTTTTGATCAGGCAGTGGAAGGC	Reverse primer for antisense *yeaJ* PCR check; use with AZ1275
AZ1249	GATACAAAGGCTTTCAAAAAAGCTGCGGTG TAGGCTGGAGCTGCTTC	Forward primer for deletion of antisense *serU* strain construction; use with AZ1250
AZ1250	TGTAAAAAACGTTCGGCAAGAGTGACATTC CGGGGATCCGTCGACC	Reverse primer for deletion of antisense *serU* strain construction; use with AZ1249
AZ1251	CAACGCGTCATAAATGTTTACGCAAGTG	Forward primer for antisense *serU* PCR check; use with AZ1252
AZ1252	CCACAAATGGCGCAGGATAAATTAAGAC	Reverse primer for antisense *serU* PCR check; use with AZ1251
AZ1267	CACCACTACAGCGGAACTTTCTTCAGTGTA GGCTGGAGCTGCTTC	Forward primer for deletion of antisense *ytfJ* strain construction; use with AZ1268
AZ1268	CTTAGTGACTATAGACTATCCGGGCATTCC GGGGATCCGTCGACC	Reverse primer for deletion of antisense *ytfJ* strain construction; use with AZ1267
AZ1269	CCCCTATCCCATAGATAACGATAGG	Forward primer for antisense *ytfJ* PCR check; use with AZ1270
AZ1270	GTATAACCGTCCACGGAACAGGATC	Reverse primer for antisense *ytfJ* PCR check; use with AZ1269
AZ1294	CGTAACAACGTAGTACGATGAACATTGCGT GTAGGCTGGAGCTGCTTC	Forward primer for deletion of antisense *thrW* strain construction; use with AZ1295
AZ1295	CGGAAGTGGCGGTAAGCACACATTCCGGG GATCCGTCGACC	Reverse primer for deletion of antisense *thrW* strain construction; use with AZ1294
AZ1296	CAC CCT AGC CGA TGC CGT G	Forward primer for antisense *thrW* PCR check; use with AZ1297
AZ1297	CGATGCCATCGCCCATATTCGTG	Reverse primer for antisense *thrW* PCR check; use with AZ1296
Oligos used for riboprobe construction		
MK0372	CTTGTCCCATTGACGCCATCACGCT	Forward primer for riboprobe construction for antisense *eutB*; use with MK0373
MK0373	CAGAGATGCATAATACGACTCACTATAGGG GGTGATGACATCATGCTCAACTAC	Reverse primer for riboprobe construction for antisense *eutB*; use with MK0372
MK0374	GTTATGACCGCTGGCGTTACTAAGG	Forward primer for riboprobe construction for antisense *qor*; use with MK0375
MK0375	CAGAGATGCATAATACGACTCACTATAGGG	Reverse primer for riboprobe construction for antisense

	CTGTTCTCTTTGATTGCCAGCGGTGTGA	*qor*; use with MK0374
MK0413	CGGTACAGGAAGCGCAATCAGTTGCGAG	Forward primer for riboprobe construction for antisense *ymfL*; use with MK0414
MK0414	CAGAGATGCATAATACGACTCACTATAGGG GAAGACGGTGTAGTGGAACCGCATGA	Reverse primer for riboprobe construction for antisense *ymfL*; use with MK0413
AZ1310	GTATTCATTGTGTGAATGACATGTCGC	Forward primer for riboprobe construction for antisense *argR* ; use with AZ1311
AZ1311	CAGAGATGCATAATACGACTCACTATAGGG ACCACCCCTGCTAACGGTTTC	Reverse primer for riboprobe construction for antisense *argR*; use with AZ1310
AZ1280	GCAAAAGGGGGAAAATGAATAATGC	Forward primer for riboprobe construction for antisense *gmr*; use with AZ1281
AZ1281	CAGAGATGCATAATACGACTCACTATAGGG CCAAGAACGGGATCAATGAGC	Reverse primer for riboprobe construction for antisense *gmr*; use with AZ1280
AZ1298	GCAGGCGTTATGTAAGAAAGTGTAACTC	Forward primer for riboprobe construction for antisense *holE*; use with AZ1299
AZ1299	CAGAGATGCATAATACGACTCACTATAGGG GGCATTTAAAGAACGCTACAATATGCCG	Reverse primer for riboprobe construction for antisense *holE*; use with AZ1298
AZ1304	GGAGATCTACAAAGTTAGAGGCAGG	Forward primer for riboprobe construction for antisense *yggN*; use with AZ1305
AZ1305	CAGAGATGCATAATACGACTCACTATAGGG TGGGGGGGCTGCAATCCTC	Reverse primer for riboprobe construction for antisense *yggN*; use with AZ1304
AZ1286	GGTTTTACCGTCAAAAGAGATAAACCCTG	Forward primer for riboprobe construction for antisense *yliF*; use with AZ1287
AZ1287	CAGAGATGCATAATACGACTCACTATAGGG CAGCTGTTGGTTGGTTTGCGCAC	Reverse primer for riboprobe construction for antisense *yliF*; use with AZ1286
AZ1259	CATGCTCAAATAAAGCTGCTCAGCC	Forward primer for riboprobe construction for antisense *speA*; use with AZ1260
AZ1260	CAGAGATGCATAATACGACTCACTATAGGG CTGCAGAAGATGCGCCGCG	Reverse primer for riboprobe construction for antisense *speA*; use with AZ1259
AZ1253	GCGTCACGTTCATACTGATATAACGC	Forward primer for riboprobe construction for antisense *gsiB*; use with AZ1254
AZ1254	CAGAGATGCATAATACGACTCACTATAGGG GCCCAAACTGGACAGCATAACCTG	Reverse primer for riboprobe construction for antisense *gsiB*; use with AZ1253
AZ1271	GCACCATATTCAGCAAAATTAACGCCG	Forward primer for riboprobe construction for antisense *yeaJ*; use with AZ1272
AZ1272	CAGAGATGCATAATACGACTCACTATAGGG CTAACCGATACCGATTCGGGGC	Reverse primer for riboprobe construction for antisense *yeaJ*; use with AZ1271
AZ1247	GAATCAAGTGCTGAATGTCACAGTATCG	Forward primer for riboprobe construction for antisense *serU*; use with AZ1248
AZ1248	CAGAGATGCATAATACGACTCACTATAGGG GGACCGGTCTCGAAAACCGGAG	Reverse primer for riboprobe construction for antisense *serU*; use with AZ1247
AZ1265	CAGCAATATGTTGCAGTACTCGCAC	Forward primer for riboprobe construction for antisense *ytfJ*; use with AZ1266

AZ1266	CAGAGATGCATAATACGACTCACTATAGGG CTACGCAAGATTCTGGCACTCAC	Reverse primer for riboprobe construction for antisense *ytfJ*; use with AZ1265
AZ1292	GTG AGC GAA GCC CTA TCA GGC	Forward primer for riboprobe construction for antisense *thrW*; use with AZ1293
AZ1293	CAGAGATGCATAATACGACTCACTATAGGG GCGCATTCGTAATGCGAAGGTCG	Reverse primer for riboprobe construction for antisense *thrW*; use with AZ1292

TABLE S3. Summary of *E. coli* K-12 MG1655 dRNA-seq libraries analyzed in this study.

Biological replicate	Library replicate	Sequencing run	Biological condition			Sequencing technique
			M63 0.4	LB 0.4	LB 2.0	
B1	L1	GA	x	x	x	GAIIx
B2	L1	HS1	x	x	x	HiSeq 2000
B2	L1	HS2	x	x	x	HiSeq 2000
B1	L2	HS2			x	HiSeq 2000
B2	L2	HS2			x	HiSeq 2000

TABLE S4. Read mapping statistics for the *E. coli* dRNA-seq libraries. This table contains the total number of reads after quality trimming, the number of mapped and uniquely mapped reads for each growth condition, library replicate and sequencing replicate (see Supplemental Materials and Methods). Percentage values are relative to the total number of reads after quality trimming.

	Library	Total number of reads after quality trimming	Mapped reads	% mapped reads	Uniquely mapped reads	% uniquely mapped reads
L1 GA	M63 0.4 B1 +TEX	2,226,993	2,179,278	97.86	1,353,514	60.78
	M63 0.4 B1 –TEX	3,618,430	3,587,012	99.13	1,641,287	45.36
	LB 0.4 B1 +TEX	3,386,772	3,319,935	98.03	2,450,714	72.36
	LB 0.4 B1 –TEX	2,348,269	2,313,292	98.51	1,448,100	61.67
	LB 2.0 B1 +TEX	1,812,576	1,662,957	91.75	1,255,341	69.26
	LB 2.0 B1 –TEX	2,950,320	2,833,237	96.03	1,369,732	46.43
L1 HS1	M63 0.4 B2 +TEX	8,676,235	8,320,773	95.90	5,616,514	64.73
	M63 0.4 B2 –TEX	8,180,585	7,878,531	96.31	4,840,039	59.16
	LB 0.4 B2 +TEX	6,173,388	6,034,843	97.76	4,521,459	73.24
	LB 0.4 B2 –TEX	7,901,590	7,050,172	89.22	3,178,987	40.23
	LB 2.0 B2 +TEX	7,039,151	6,785,582	96.40	4,724,268	67.11
	LB 2.0 B2 –TEX	5,486,064	4,662,787	84.99	2,771,432	50.52
L1 HS2	M63 0.4 B2 +TEX	8,785,626	8,433,738	95.99	5,688,313	64.75
	M63 0.4 B2 –TEX	9,814,115	9,463,702	96.43	5,799,209	59.09
	LB 0.4 B2 +TEX	5,878,169	5,753,542	97.88	4,301,601	73.18
	LB 0.4 B2 –TEX	8,321,493	7,434,035	89.34	3,356,765	40.34
	LB 2.0 B2 +TEX	7,792,864	7,520,559	96.51	5,221,142	67.00
	LB 2.0 B2 –TEX	6,648,330	5,654,792	85.06	3,352,443	50.43
L2 HS2	LB 2.0 B1 +TEX	6,336,509	5,383,288	84.96	3,970,765	62.66
	LB 2.0 B1 –TEX	7,782,920	5,842,996	75.07	3,399,792	43.68
	LB 2.0 B2 +TEX	6,543,088	5,961,511	91.11	3,984,821	60.90
	LB 2.0 B2 –TEX	5,327,191	3,753,567	70.46	2,217,871	41.63

TABLE S5. Mapping statistics based on strand and RNA class. This table indicates the number of reads mapped to the different RNA classes (mRNA, ncRNA, rRNA, tRNA, and tmRNA) for each strand across all libraries and biological conditions. The numbers for both the mapped and uniquely mapped reads per RNA class are shown with percentage values calculated from the total number of mapped reads regardless of mapped location (taken from Table S4) for the respective biological conditions.

		Biological condition											
		M63 0.4				LB 0.4				LB 2.0			
		Mapped reads		Uniquely mapped reads		Mapped reads		Uniquely mapped reads		Mapped reads		Uniquely mapped reads	
	Total*	39,863,034		24,938,876		31,905,819		19,257,626		50,061,276		32,267,607	
sense	mRNA	7,542,584	(19%)	7,459,916	(30%)	4,754,092	(15%)	4,679,204	(24%)	12,621,458	(25%)	12,449,865	(39%)
	ncRNA	2,144,882	(5%)	2,134,932	(9%)	1,974,888	(6%)	1,972,296	(10%)	6,865,538	(14%)	6,845,341	(21%)
	rRNA	9,337,211	(23%)	1,281,358	(5%)	8,409,540	(26%)	843,630	(4%)	16,392,800	(33%)	1,723,029	(5%)
	tRNA	8,608,679	(22%)	5,226,638	(21%)	5,500,223	(17%)	4,821,845	(25%)	3,695,560	(7%)	1,600,498	(5%)
	tmRNA	283,482	(1%)	283,431	(1%)	182,898	(1%)	182,884	(1%)	510,495	(1%)	510,408	(2%)
antisense	mRNA	976,736	(2%)	945,150	(4%)	566,133	(2%)	535,981	(3%)	1,900,343	(4%)	1,834,877	(6%)
	ncRNA	544,883	(1%)	541,313	(2%)	131,872	(0%)	131,022	(1%)	301,004	(1%)	300,376	(1%)
	rRNA	1,820	(0%)	70	(0%)	2,498	(0%)	30	(0%)	3,385	(0%)	435	(0%)
	tRNA	2,471	(0%)	2,022	(0%)	2,542	(0%)	2,152	(0%)	30,965	(0%)	29,200	(0%)
	tmRNA	8	(0%)	7	(0%)	12	(0%)	12	(0%)	26	(0%)	26	(0%)

*Total reads were calculated by summing all the reads mapped to the *E. coli* genome from all libraries (+TEX and –TEX) from a particular condition including reads that mapped to locations other than the listed RNA classes. The numbers were generated from the data in Supplemental Table S4.

TABLE S6. Known annotated asRNAs. This table contains the previously known annotated asRNAs taken from NCBI annotation.

Gene	Locus tag	Strand	Start	End	Description
arrS	b4704	-	3656009	3656077	asRNA ArrS, GadE-regulated, function unknown
gadY	b4452	+	3662887	3662991	asRNA regulator of transcriptional activator GadX mRNA
ohsC	b4608	+	2698542	2698618	asRNA regulator of *shoB* toxin
rdlA	b4420	+	1268546	1268612	asRNA RdlA affects LdrA translation; proposed addiction module in LDR-A repeat, with toxic peptide LdrA
rdlB	b4422	+	1269081	1269146	asRNA RdlB affects LdrB translation; proposed addiction module in LDR-B repeat, with toxic peptide LdrB
rdlC	b4424	+	1269616	1269683	asRNA RdlC affects LdrC translation; proposed addiction module in LDR-C repeat, with toxic peptide LdrC
rdlD	b4454	+	3698159	3698224	asRNA RdlD affects LdrD translation; proposed addiction module in LDR-D repeat, with toxic peptide LdrD
sibA	b4436	+	2151333	2151475	asRNA regulator of toxic IbsA protein; in SIBa repeat
sibB	b4437	+	2151668	2151803	asRNA regulator of toxic IbsB protein; in SIBb repeat
sibC	b4446	+	3054871	3055010	asRNA regulator of toxic IbsC protein; in SIBc repeat
sibD	b4447	-	3192745	3192887	asRNA regulator of toxic IbsD protein; in SIBd repeat
sibE	b4611	-	3193121	3193262	asRNA regulator of toxic IbsE protein; in SIBe repeat
sokB	b4429	+	1490143	1490198	asRNA blocking *mokB*, and hence *hokB*, translation
sokC	b4413	+	16952	17006	asRNA blocking *mokC*, and hence *hokC*, translation
sokE	b4700	+	606957	607015	asRNA at remnant *mokE/hokE* locus
sokX	b4701	+	2885376	2885431	asRNA, function unknown
symR	b4625	+	4577858	4577934	asRNA destabilizing divergent and overlapping *symE* mRNA

TABLE S7. Published *E. coli* transcriptome studies reporting asRNAs used for comparisons.

Reference	**Strain**	**Growth conditions**	**Special treatment**	**Sequencing**	**Total number of reads**	**Annotated asTSS**
Conway *et al*. (15)	K12 BW38028/ BW39452 Δ*rpoS*	Fermentor MOPS + 0.2% glucose OD_{600} ~0.1-stationary phase	Terminator Exonuclease (TEX) treatment	ABI SOLiD	72,147,745	89
Dornenburg *et al.* (10)	K12 MG1655	LB OD_{600} 0.7	rRNA depletion	Illumina (44 cycles)	8,967,903	1,390
Raghavan *et al.* (11)	K12 MG1655	LB OD_{600} ~0.5	rRNA depletion	Illumina GA II (35 cycles)	30,206,434	90
Shinhara *et al.* (12)	K12 BW25113	M63 glucose OD_{600} 0.76 at 37°C	Extraction of low-molecular weight RNAs	Illumina 1G (35 cycles)	12,473,172	112[#]
Mendoza-Vargas *et al.* (13)	K12 MG1655	LB and M63 glucose at 37°C and 30°C	rRNA depletion	Roche 454 GS20 (reads ~100 nt long)	~350,000	165
Salgado *et al.* *(14)*	K12 MG1655 / MG1655 Δ*rppH*	LB or MOPS media with 0.2% Glucose or 0.2% acetate at 37°C	rRNA depletion, enrichment for 5'-P or 5'-PPP	Illumina GAIIx (36 cycles)	77,628,858[*]	182

*Number of mapped non-rRNA sequences

[#]TSS represent extracted 5' end positions of reported antisense transcripts

TABLE S8. Candidate asRNAs tested by northern analysis.

Strand	Sense gene	Start-end	TSS expression bin	Size	Expression	Detected by others
+	*qorA*	4,261,162-4,261,303	>100,000	70, 200	*rne* LB 0.4	No
-	*gsiB*	869,344-869-393	10,000-100,000	170	LB 0.4; M63	(10)
-	*holE*	1,923,282-1,923,331	10,000-100,000	230	*rnc* LB 0.4 + 2.0	(10, 11, 14)
+	*serU*	2,041,439-2,041,488	10,000-100,000	170	*rnc* LB 0.4 +2.0; *rne*	(10)
+	*eutB*	2,555,334-2,555,448	10,000-100,000	310-320	LB 0.4; *rnc* 0.4; *rne* 0.4	(10, 12, 15)
+	*speA*	3,082,647-3,082,696	10,000-100,000	160	LB 0.4	No
+	*ytfJ*	4,437,157-4,437,206	10,000-100,000	180-190	*rnc* 0.4	(10)
-	*yeaJ*	1,870,929-1,870,978	1,000-10,000	multiple bands	LB 0.4; *rnc* 0.4+ 2.0; *rne* 0.4	(10, 11)
+	*gmr*	1,342,751-1,342,800	1,000-10,000	multiple bands	*rne* 0.4	(10)
-	*yliF*	874,841-874,890	1,000-10,000	185	LB 0.4; M63 0.4; *rnc* 0.4; *rne* 0.4	(10-12)
-	*thrW*	262,193-262,242	1,000-10,000	180	*rnc* LB 0.4 +2.0	No
+	*yggN*	3,098,913-3,098,986	1,000-10,000	multiple bands	*rnc* LB 0.4 +2.0	No
-	*argR*	3,383,214-3,383,263	1,000-10,000	210	LB 0.4; *rnc* 0.4; *rne* 0.4 + 2.0	No
-	*ymfL*	1,202,757-1,203,131	1,000-10,000	>50	LB 0.4 + 2.0; M63; *rnc* 0.4 + 2.0	(12, 15)

SUPPLEMENTAL DATA SETS

Data Set S1. TSS map. This table contains information on positions and assigned classes of all annotated TSS. It lists all TSS that were detected in at least one of the three conditions (column "detected" = 1). In case a TSS was not detected in a certain condition the value of detected is "0". Also, if the TSS is assigned to more than one class, there is one row for each class assignment and each associated gene.

Data Set S2. Overlapping 5' UTRs (Tab S2A) and comparison to IP-dsRNAs (Tab S2B). Tab S2A contains all pairs of overlapping 5' UTRs based on primary and secondary TSS, which can also be classified as internal and/or antisense, with a minimum overlap of 10 nt. Tab S2B contains all IP-dsRNAs described by Lybecker *et al.* (16) for which we found at least one matching asTSS.

Data Set S3. Exclusively asTSS bin expression table. This data set contains information on the expression of exclusively asTSS as well as the overlap to asRNAs from previous studies. The TSS are separated into bins according to the maximum RPKM value over all libraries. Each worksheet contains the TSS assigned to a specific bin.

SUPPLEMENTAL REFERENCES

1. **Sharma, C. M., S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiss, A. Sittka, S. Chabas, K. Reiche, J. Hackermuller, R. Reinhardt, P. F. Stadler, and J. Vogel.** 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori.* Nature **464:**250-255.
2. **Berezikov, E., F. Thuemmler, L. W. van Laake, I. Kondova, R. Bontrop, E. Cuppen, and R. H. Plasterk.** 2006. Diversity of microRNAs in human and chimpanzee brain. Nat Genet **38:**1375-1377.
3. **Förstner, K. U., J. Vogel, and C. M. Sharma.** 2014. READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. Bioinformatics **pii:**btu533.
4. **Hoffmann, S., C. Otto, S. Kurtz, C. M. Sharma, P. Khaitovich, J. Vogel, P. F. Stadler, and J. Hackermuller.** 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput Biol **5:**e1000502.
5. **Nicol, J. W., G. A. Helt, S. G. Blanchard, Jr., A. Raja, and A. E. Loraine.** 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics **25:**2730-2731.
6. **Dugar, G., A. Herbig, K. U. Forstner, N. Heidrich, R. Reinhardt, K. Nieselt, and C. M. Sharma.** 2013. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. PLoS Genet **9:**e1003495.
7. **Mao, X., Q. Ma, C. Zhou, X. Chen, H. Zhang, J. Yang, F. Mao, W. Lai, and Y. Xu.** 2014. DOOR 2.0: presenting operons and their functions through dynamic and integrated views. Nucleic Acids Res. **42:**D654-D659.
8. **Love, M. I., W. Huber, and S. Anders.** 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. bioRxiv **In press**.
9. **Bailey, T. L., and C. Elkan.** 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. **2:**28-36.
10. **Dornenburg, J. E., A. M. Devita, M. J. Palumbo, and J. T. Wade.** 2010. Widespread antisense transcription in *Escherichia coli*. MBio **1:**e00024-00010.
11. **Raghavan, R., D. B. Sloan, and H. Ochman.** 2012. Antisense transcription is pervasive but rarely conserved in enteric bacteria. MBio **3:**e00156-00112.

12. **Shinhara, A., M. Matsui, K. Hiraoka, W. Nomura, R. Hirano, K. Nakahigashi, M. Tomita, H. Mori, and A. Kanai.** 2011. Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. BMC Genomics **12:**428.

13. **Mendoza-Vargas, A., L. Olvera, M. Olvera, R. Grande, L. Vega-Alvarado, B. Taboada, V. Jimenez-Jacinto, H. Salgado, K. Juarez, B. Contreras-Moreira, A. M. Huerta, J. Collado-Vides, and E. Morett.** 2009. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. PLoS One **4:**e7526.

14. **Salgado, H., M. Peralta-Gil, S. Gama-Castro, A. Santos-Zavaleta, L. Muñiz-Rascado, J. S. García-Sotelo, V. Weiss, H. Solano-Lira, I. Martínez-Flores, A. Medina-Rivera, G. Salgado-Osorio, S. Alquicira-Hernández, K. Alquicira-Hernández, A. López-Fuentes, L. Porrón-Sotelo, A. M. Huerta, C. Bonavides-Martínez, Y. I. Balderas-Martínez, L. Pannier, M. Olvera, A. Labastida, V. Jiménez-Jacinto, L. Vega-Alvarado, V. Del Moral-Chávez, A. Hernández-Alvarez, E. Morett, and J. Collado-Vides.** 2013. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res. **41:**D203-D213.

15. **Conway, T., J. P. Creecy, S. M. Maddox, J. E. Grissom, T. L. Conkle, T. M. Shadid, J. Teramoto, P. San Miguel, T. Shimada, A. Ishihama, H. Mori, and B. L. Wanner.** 2014. Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. mBio **5:**e01442-01414.

16. **Lybecker, M., B. Zimmermann, I. Bilusic, N. Tukhtubaeva, and R. Schroeder.** 2014. The double-stranded transcriptome of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **111:**3134-3139.

17. **Thomason, M. K., F. Fontaine, N. De Lay, and G. Storz.** 2012. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. Mol Microbiol **84:**17-35.

18. **Kim, D., J. S. Hong, Y. Qiu, H. Nagarajan, J. H. Seo, B. K. Cho, S. F. Tsai, and B. O. Palsson.** 2012. Comparative analysis of regulatory elements between *Escherichia coli* and *Klebsiella pneumoniae* by genome-wide transcription start site profiling. PLoS Genet **8:**e1002867.

19. **Opdyke, J. A., E. M. Fozo, M. R. Hemm, and G. Storz.** 2011. RNase III participates in GadY-dependent cleavage of the *gadX-gadW* mRNA. J. Mol. Biol. **406:**29-43.

20. **Massé, E., F. E. Escorcia, and S. Gottesman.** 2003. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Genes Dev. **17:**2374-2383.