Supplementary information

The Rocks and Shallows of Deep RNA Sequencing: Examples in the *Vibrio cholerae* RNome

Carsten A. Raabe[1], Chee Hock Hoe[2], Gerrit Randau[1], Juergen Brosius[1], Thean Hock Tang[2] and Timofey S. Rozhdestvensky[1]

[1]Institute of Experimental Pathology, University of Muenster, Von-Esmarch-Str. 56, 48149 Muenster, Germany. [2]Advanced Medical and Dental Institute (AMDI), Universiti Sains Malaysia, 13200 Penang, Malaysia

Supplementary introduction

The bacteriology and epidemiology of *V. cholerae* have been described over a century ago; nevertheless cholera continues to be an important public health problem especially among poorer communities in Africa, Asia and South America (WHO, 2009, http://www.who.int/cholera/countries/en/). The etiological agent, *Vibrio cholerae O1*, is commonly subdivided into two biotypes, the *classical* and *E1 Tor*. Nowadays *E1 Tor* constitutes the primary source of human cholera.

Vibrio cholerae belongs to the genus of gram-negative rod-shaped bacteria (Kierek & Watnick, 2003; Sen & Ghosh, 2005). *V. cholerae* species are facultative anaerobes, oxidase positive and non-spores forming. *E1 Tor* often causes epidemics, especially in places where sanitation is poor (WHO, 2010 http://www.who.int/mediacentre/factsheets/fs107/en/index.html). The complete genome of *Vibrio cholerae El Tor N16961* is distributed on two circular chromosomes of 2.961.146 (chromosome 1) and 1.072.314 (chromosome 2) base pairs respectively (Heidelberg et al., 2000). The average G+C content is about 47% (Heidelberg et al., 2000). Genes encoding all proteins required to maintain all important biochemical pathways are localized on the larger chromosome; it also harbours main determinants of *Vibrio cholerae* pathogenicity. The smaller chromosome, which is derived from a megaplasmid within early stages of *Vibrio cholerae* evolution, harbours about 50% hypothetical genes of unknown function (Heidelberg et al., 2000).

Alerted by the abundance of npcRNAs and the broad spectrum of mechanisms they might involve (Repoila et al., 2003; Brantl, 2007; Toledo-Arana et al., 2007; Song & Wai, 2009; Holmqvist et al., 2010; Tu et al., 2010), multiple bioinformatic and experimental attempts were conducted to identify and analyse the npcRNA transcriptome in various bacterial genomes (Vogel et al., 2003; Abu-Qatouseh et al., 2010; Beaume et al., 2010; Bohn et al., 2010; Chinni et al., 2010; Irnov et al., 2010; Sharma et al., 2010). Till recently, most bacterial npcRNAs were identified within the intergenic regions (IGR) of the genome (Sridhar et al., 2010). This bias of small npcRNAs mapping does not reflect a preferential localization, but rather

is a consequence of the computational limitations of npcRNA prediction. Moreover, search strategies or algorithms that rely on the detection of sequence or secondary structure conservation are inappropriate to identify species-specific or unstructured RNAs**.** Therefore, the approach of experimental RNomics has been proven beneficial, as most bio-computational pitfalls were believed negligible in fully randomized experimental screens. However, as it turns out, even experimental surveys suffer from their inherent methodological limitations.

Supplementary Results and Discussion.

Growth stage determination, cDNA library construction and analysis

Apart from constitutively expressed bacterial small npcRNA genes it is now widely accepted that many npcRNAs are subject of growth dependent regulation (Perez et al., 2009). To correlate bacterial growth with defined npcRNA expression profiles growth curves of *V. cholerae clinical isolate VC3321* and N16961 were determined (Figure 1B, C, Supplementary Figure 1B – Figure 6).

To generate a specialized *V. cholerae* library that would account for the global npcRNA transcriptome equal amounts of total RNA of three growth stages (log, lag, stationary) were combined and size-fractionated on preparative denaturing polyacrylamide gels (Supplementary Figure 1B). Two size fractions of 10 - 60 and 60 - 500 nts were collected. Each fraction was separately C-tailed and the DNA SalI 5' adapter was ligated to allow for directional and full-length cDNA cloning. After reverse transcription, 12 cycles of single stranded PCR

amplification primed by the SalI oligonucleotide was performed. 7500 cDNA clones were randomly selected and sequenced (see main text).

. Finally, 6801 cDNA sequences were assembled into contigs using the SeqMan sequence assembler (Lasergene v8 software). The contigs were mapped to the genome of *V. cholerae* by local blast procedures; each functional annotation was challenged manually with Blat (http://archaea.ucsc.edu/) or Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) sequence searches. RNA contigs of sizes less than 15 nts were considered too short and excluded form further analysis.

The vast majority of cDNAs were derived from known npcRNAs of *V. cholerae* (Supplementary Figure 1A). About 61% of all cDNAs represents tRNA (39%) or 5S rRNA (19%) derived reads. Further 11% correspond to fragments of 16S and 23S rRNAs. In addition, 9% reflect reads matching to Rfam npcRNA annotations. The second largest subgroup represented by 660 contigs (14% of all obtained reads) account for RNA sequences that partially or completely overlap with known or hypothetical mRNAs in the same orientation.

We also identified 91 *cis*-antisense encoded npcRNA candidates (2% of all reads) that are expressed complementary to annotated open reading frames (ORFs) (Supplementary Table 1). We subdivided them into two categories: partially or completely overlapped npcRNA candidates. Finally, 132 intergenic npcRNAs (4% of all reads) were detected (Supplementary Table 2).

cis-antisense npcRNA candidates in _V. cholerae_

In general, two different types of bacterial small npcRNAs that supposedly act via base-pairing with corresponding mRNA targets are differentiated: *cis*- and *trans*-encoded npcRNAs. Bacterial antisense transcripts encoded complementary to their respective RNA (*cis*-antisense) targets are best characterized in plasmids where they function in mechanisms as diverse as translational inhibition or inhibition of primer maturation (Brantl, 2002; Baker et al., 2007).

Genomic tiling arrays suggest low levels of *cis*-antisense transcription throughout bacterial growth (Selinger et al., 2000). Selinger *et al.* reported *cis*-antisense transcription for 3000 *E. coli* genes indicating that *cis*-antisense transcription is a general phenomenon and common to many bacterial genes (Selinger et al., 2000). In *V. cholerae* Liu et al. (2009) described 127 novel *cis*-antisense npcRNAs (Liu et al., 2009). We report 91 novel candidates transcribed complementary to annotated ORFs (Supplementary Table 1). Three categories of *cis*-encoded antisense transcripts are distinguished based on CDS region that is complementary to npcRNA candidate (Supplementary Table 1).

When comparing our *cis*-antisense encoded npcRNA candidates with those of Liu et al. (2009), only 7 of them were common to both surveys (Supplementary Table 3). Northern blot analyses indicate that a number of npcRNA candidates from this category are expressed in growth stage specific manner (Figure 1B, C, Supplementary Figures 2, 4 and 6).

Intergenic npcRNAs in *V. cholerae*

Intergenic npcRNAs are defined by their genomic localization between two annotated genes. We identified 132 intergenic npcRNA candidates in our experimental survey, among them four that match to multiple genomic localizations (Supplementary Table 2). Interestingly only 32 intergenic npcRNA candidates are common between our and Liu et al. (2009) surveys (Supplementary Table 3).

About one half of the tested intergenic npcRNAs exhibit growth stage regulated expression profiles (Figure 1B, C, Supplementary Figures 3, 5 and 6). To differentiate candidates that might constitute independent transcriptional units, we scanned our dataset for *rho*-independent termination signals. Only 12 npcRNA candidates contained discernible *rho*-independent terminators (Supplementary Table 2). Therefore and in agreement with the frequently more complex expression profiles detected in northern blot hybridization experiments, many candidates of this category might represent stable RNA processing products (Figure 1B, C, Supplementary Figures 3, 5 and 6). Since most bacterial mRNAs lack UTR annotations it cannot be ruled out that members of this category resemble mRNA processing/degradation products.

Supplementary Material and Methods

Computational analysis

Analysis of the 7500 cDNA dataset

UNIX scripts removed 5' adapter and C-tail from 7500 reads. cDNAs longer than 14 bp were processed further. The SeqMan sequence assembly program of the Lasergene software package (http://www.gatcbiotech.com/de) was used to assemble cDNAs into contigs. The assembly parameters were set to a 10 bp minimum overlap with a minimum sequence similarity of 90% (calculated for the overlap). The resulting contigs were manually inspected and validated. The genomic coordinates and orientation of cDNA sequence contigs were computed by local Wu-blast (http://blast.wustl.edu/) procedures. For all npcRNAs the best match as judged by the lowest e-value is reported. Repeated matches with identical and lowest e-value are denoted. All annotations are further validated by web-based Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Blat (http://archaea.ucsc.edu/cgi-bin/hgBlat) searches. The genomic positions and orientations of all annotated Refseq RNAs, tRNAs or known npcRNAs in *V. cholerae* are downloaded as BED formatted text file via the UCSC table browser (http://archaea.ucsc.edu/cgi-bin/hgTables). AWK scripts were utilized to annotate putative npcRNA candidates and to identify overlaps to known RNAs (www.gnu.org/manual/gawk/gawk.pdf). Similarly, AWK scripts were used to calculate the intersection of putative npcRNAs between both datasets.

Supplementary Figures legends

Supplementary Figure 1

(A) Pie chart representing the read frequencies of npcRNA candidates detected by our cDNA construction and conventional sequencing methods **(B)** growth comparison of *V. cholerae* clinical isolates N16961 and VC3321.

Supplementary Figure 2

Northern blot hybridisation of *cis*-antisense encoded npcRNAs detected in *V. cholerae* VC3321 by our cDNA preparation method. All northern blots are carried out on *V. cholerae* VC3321 total RNA collected from different growth stages, see **Supplementary Figure 1B**. npcRNA names are indicated above each blot and molecular sizes are given to the right.

Supplementary Figure 3

Northern blot hybridisation of intergenic npcRNAs detected in *V. cholerae* VC3321 by our cDNA preparation method; all northern blots are carried out on *V. cholerae* VC3321 total RNA collected from different growth stages, see **Supplementary Figure 1B**. npcRNA names are indicated above each blot and molecular sizes are given to the right.

Supplementary Figure 4

Northern blot hybridisation of *cis*-antisense encoded npcRNAs detected in *V. cholerae* N16961 by RNA-seq. All northern blots are carried out on *V. cholerae* VC3321 total RNA collected from different growth stages, see **Supplementary Figure 1B**. npcRNA names are indicated above each blot and molecular sizes are given to the right.

Supplementary Figure 5

Northern blot hybridisation of intergenic npcRNAs detected in *V. cholerae* N16961 by RNA-seq; all northern blots are carried out on *V. cholerae* VC3321 total RNA collected from different growth stages, see **Supplementary Figure 1B**. npcRNA names are indicated above each blot and molecular sizes are given to the right.

Supplementary Figure 6

Northern blot hybridisation analyses of npcRNA candidates that were detected in both surveys. All northern blots are carried out on *V. cholerae* VC3321 total RNA collected from different growth stages, see **Supplementary Figure 1B**. npcRNA names are indicated above each blot and molecular sizes are given to the right.

Supplementary Tables legends

Supplementary Table 1

Functional annotation of *cis*-antisense encoded npcRNAs detected in *V. cholerae* VC3321 isolate. **(a) Vc_npcR_t_**name: Designation of *V. cholerae* npcRNA candidates. **(b, c) Chr.(#) and Orient.1**: *V. cholerae* chromosomal location and orientation of npcRNA candidate transcription. **(d) npcR_length:** Length of the npcRNA (nt). **(e) cDNAs(#):** cDNA frequency. **(f) rho_independent:** Detection of rho independent terminators for putative npcRNA gene**. (g, h) Cords.1 and Cords.2:** Coordinates of the npcRNA candidates are based on the *Vibrio cholerae O1 biovar El Tor str. N16961* chromosome sequence and annotation (chrI: NC_002505; chrII: NC_002506). **(i) Sense_gene:** ID of sense encoded protein coding gene based on *V. cholerae N16961* annotation. **(j) Sense_gene_des:** Description of the sense gene **(k) Orient.2:** Genomic orientation of the sense gene **(l, m) Cords.3 and Cords.4:** Coordinates of sense encoded protein coding gene based on *V. cholerae N16961* annotation. **(n) Cat:** Region of CDS that is complementary to npcRNA candidate: **comp:** within CDS **start:** Start codon overlap, **stop:** Stop codon overlap

Supplementary Table 2

Functional annotation of intergenic npcRNAs detected in *V. cholerae* VC3321 isolate. **(a) Vc_npcR_t_name:** Designation of *V. cholerae* npcRNA candidates. **(b, c) Chr.(#) and Orient.1**: *V. cholerae* chromosomal location and orientation of npcRNA candidate transcription. **(d) npcR_length:** Length of the npcRNA (nt). **(e) cDNAs(#):** cDNA frequency. **(f) rho_independent:** Detection of rho independent terminator for putative npcRNA gene**. (g, h) Cords.1 and Cords.2:**

Coordinates of npcRNA candidates are based on the *Vibrio cholerae O1 biovar El Tor str. N16961* chromosome sequence and annotation (chrI: NC_002505; chrII: NC_002506). **(l** and **n) Flanking_gene1 and Flanking_gene2:** ID of flanking gene based on *V. cholerae N16961* annotation. (**j** and **o**) **Flanking_gene1_des and Flanking_gene 2_des:** Description of the flanking genes. (**k-m** and **p-r**) **Cords.3, Cords.4, Orient.2** and **Cords.5, Cords.6, Orient.3.** Coordinates and orientations of flanking genes based on *V. cholerae N16961* annotation.

Supplementary Table 3

List of npcRNA candidates detected in both surveys.

(A, E): Designation of *V. cholerae* npcRNA candidates in our and the survey of Liu et al., respectively. **(B, C, D and F, G, H)** *V. cholerae* chromosomal location, orientation and genomic coordinates of npcRNA candidates in our (**B, C, D**) and the dataset by Liu et al. (**F, G, H**) , respectively.

Supplementary Table 4

Oligonucleotides used in this study

References

Abu-Qatouseh LF, Chinni SV, Seggewiss J, Proctor RA, Brosius J, Rozhdestvensky TS, Peters G, von Eiff C, Becker K. 2010. Identification of differentially expressed small non-protein-coding RNAs in Staphylococcus aureus displaying both the normal and the small-colony variant phenotype. *J Mol Med* **88**:565-575.

Baker CS, Eory LA, Yakhnin H, Mercante J, Romeo T, Babitzke P. 2007. CsrA inhibits translation initiation of Escherichia coli hfq by binding to a single site overlapping the Shine-Dalgarno sequence. *J Bacteriol* **189**:5472-5481.

Beaume M, Hernandez D, Farinelli L, Deluen C, Linder P, Gaspin C, Romby P, Schrenzel J, Francois P. 2010. Cartography of methicillin-resistant S. aureus transcripts: detection, orientation and temporal expression during growth phase and stress conditions. *PLoS One* **5**:e10725.

Bohn C, Rigoulay C, Chabelskaya S, Sharma CM, Marchais A, Skorski P, Borezee-Durant E, Barbet R, Jacquet E, Jacq A, Gautheret D, Felden B, Vogel J, Bouloc P. 2010. Experimental discovery of small RNAs in Staphylococcus aureus reveals a riboregulator of central metabolism. *Nucleic Acids Res* **38**:6620-6636.

Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J. 2008. Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol Cell* **32**:827-837.

Brantl S. 2002. Antisense-RNA regulation and RNA interference. *Biochim Biophys Acta* **1575**:15-25.

Brantl S. 2007. Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* **10**:102-109.

Chinni SV, Raabe CA, Zakaria R, Randau G, Hoe CH, Zemann A, Brosius J, Tang TH, Rozhdestvensky TS. 2010. Experimental identification and characterization of 97 novel npcRNA candidates in Salmonella enterica serovar Typhi. *Nucleic Acids Res* **38**:5893-5908.

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleishmann RD, Nierman WC, White O, Salzberg SL, Smith HO, Colwell RR, Mekalanos JJ, Venter JC, Fraser CM. 2000. DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. *Nature* **406**:477-483.

Holmqvist E, Reimegard J, Sterk M, Grantcharova N, Romling U, Wagner EG. 2010. Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *Embo J* **29**:1840-1850.

Irnov I, Sharma CM, Vogel J, Winkler WC. 2010. Identification of regulatory RNAs in Bacillus subtilis. *Nucleic Acids Res* **38**:6637-6651.

Kierek K, Watnick PI. 2003. The Vibrio cholerae O139 O-antigen polysaccharide is essential for Ca2+-dependent biofilm development in sea water. *Proc Natl Acad Sci U S A* **100**:14357-14362.

Liu JM, Livny J, Lawrence MS, Kimball MD, Waldor MK, Camilli A. 2009. Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res* **37**:e46.

Perez N, Trevino J, Liu Z, Ho SC, Babitzke P, Sumby P. 2009. A genome-wide analysis of small regulatory RNAs in the human pathogen group A Streptococcus. *PLoS One* **4**:e7668.

Repoila F, Majdalani N, Gottesman S. 2003. Small non-coding RNAs, co-ordinators of adaptation processes in Escherichia coli: the RpoS paradigm. *Mol Microbiol* **48**:855-861.

Selinger DW, Cheung KJ, Mei R, Johansson EM, Richmond CS, Blattner FR, Lockhart DJ, Church GM. 2000. RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nat Biotechnol* **18**:1262-1268.

Sen A, Ghosh AN. 2005. New Vibrio cholerae O1 biotype ElTor bacteriophages. *Virol J* **2**:28.

Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J. 2010. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**:250-255.

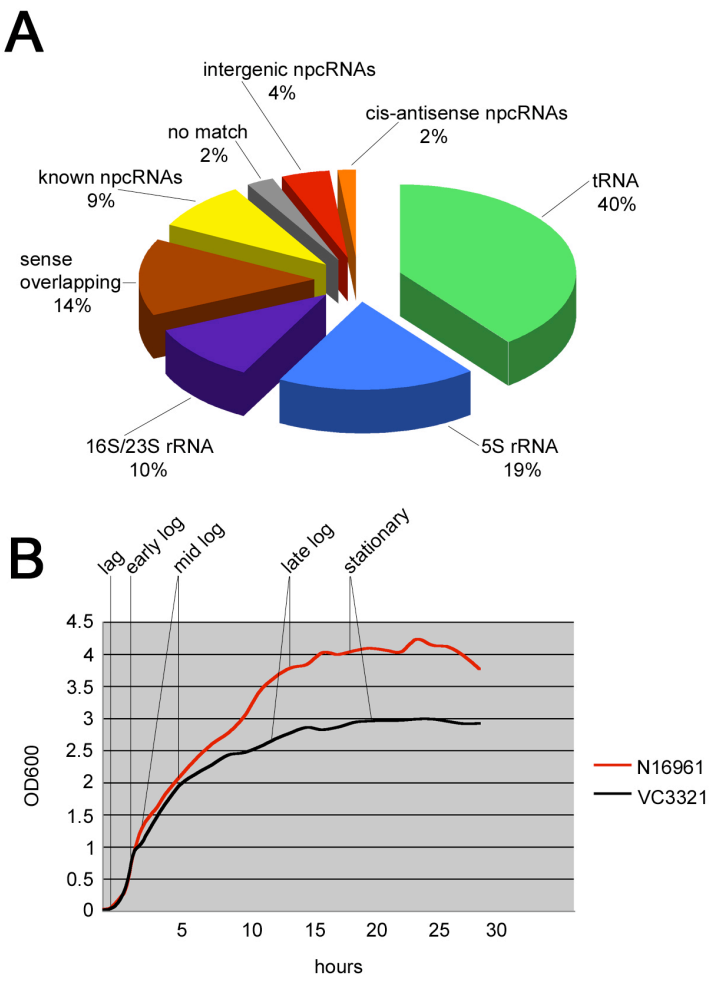
Song T, Wai SN. 2009. A novel sRNA that modulates virulence and environmental fitness of Vibrio cholerae. *RNA Biol* **6**:254-258.

Sridhar J, Narmada SR, Sabarinathan R, Ou HY, Deng Z, Sekar K, Rafi ZA, Rajakumar K. 2010. sRNAscanner: a computational tool for intergenic small RNA detection in bacterial genomes. *PLoS One* **5**:e11970.

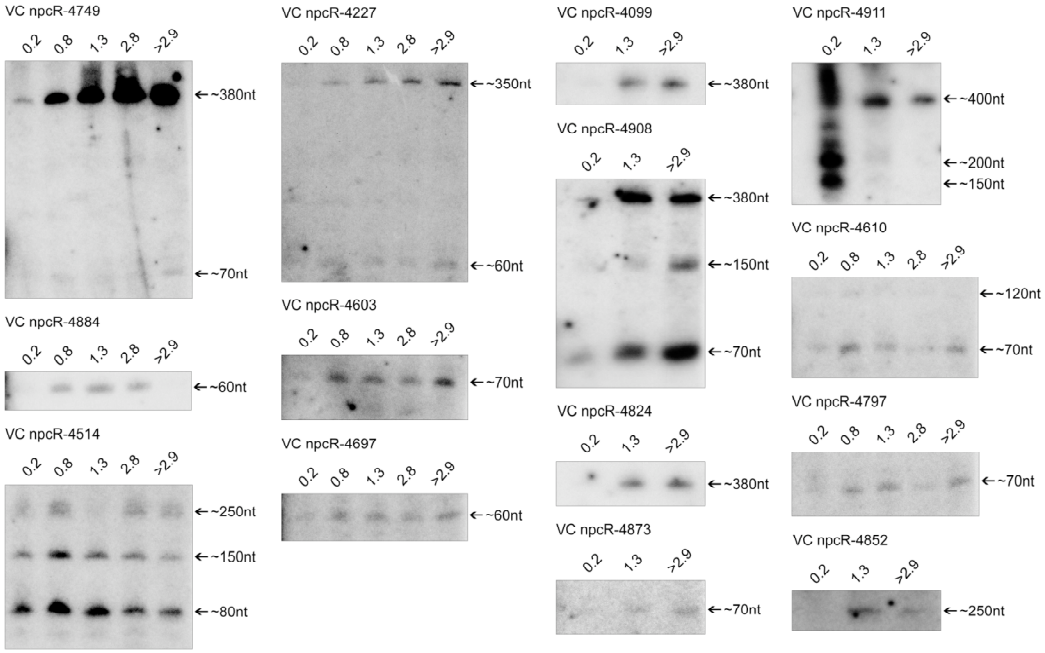
Toledo-Arana A, Repoila F, Cossart P. 2007. Small noncoding RNAs controlling pathogenesis. *Curr Opin Microbiol* **10**:182-188.

Tu KC, Long T, Svenningsen SL, Wingreen NS, Bassler BL. 2010. Negative feedback loops involving small regulatory RNAs precisely control the Vibrio harveyi quorum-sensing response. *Mol Cell* **37**:567-579.

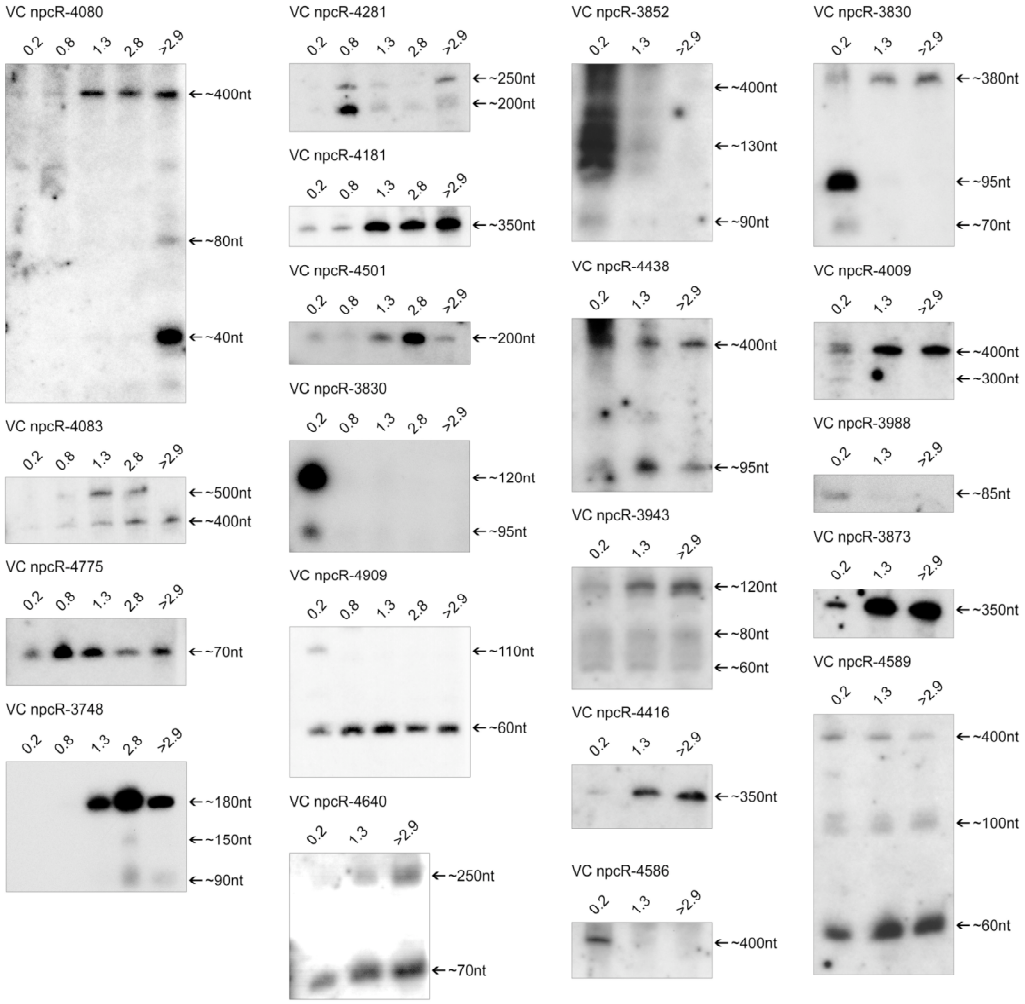
Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, Wagner EG. 2003. RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Res* **31**:6435-6443.



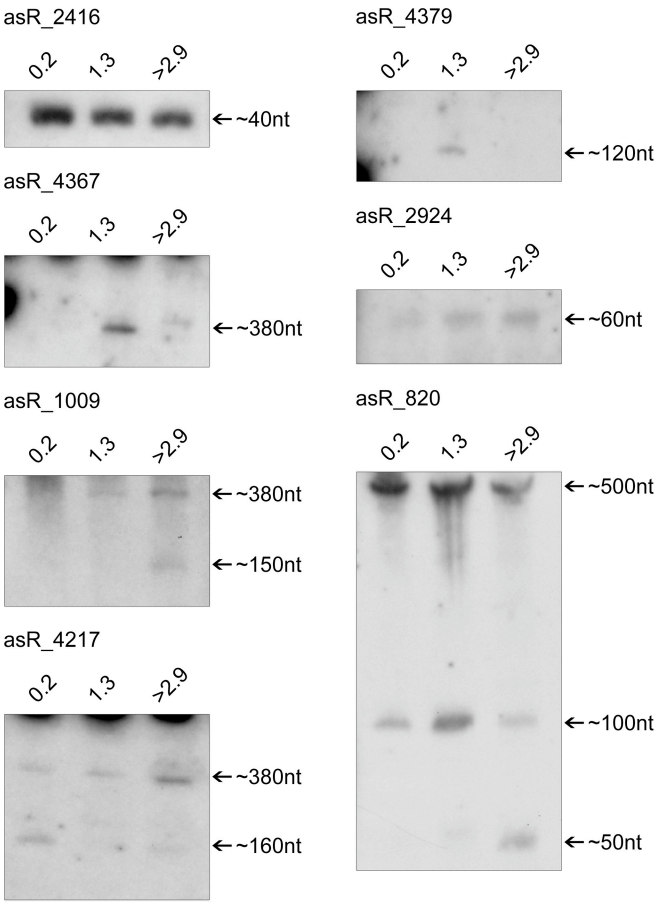
Supplementary Figure 1.



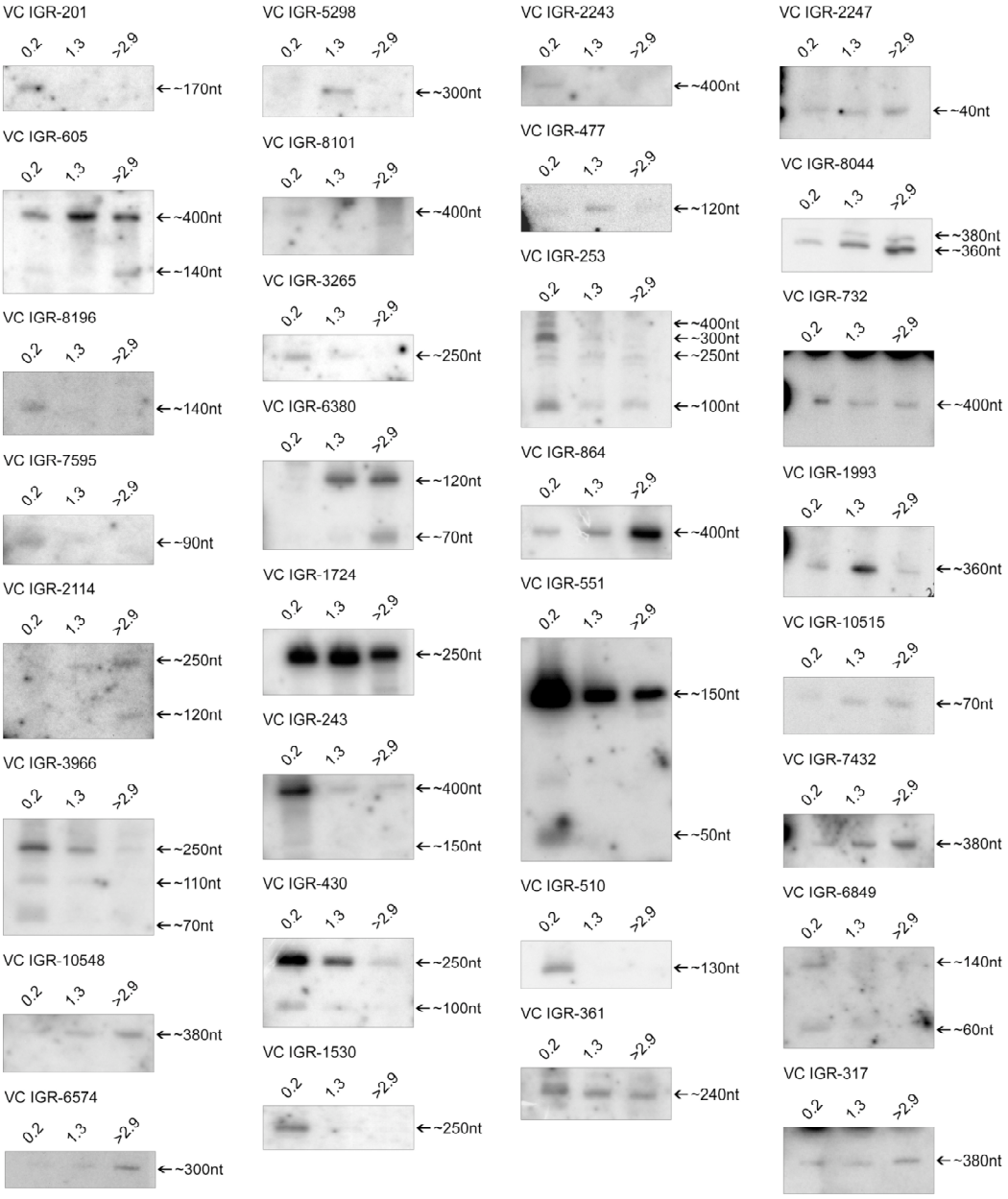
Supplementary Figure 2.



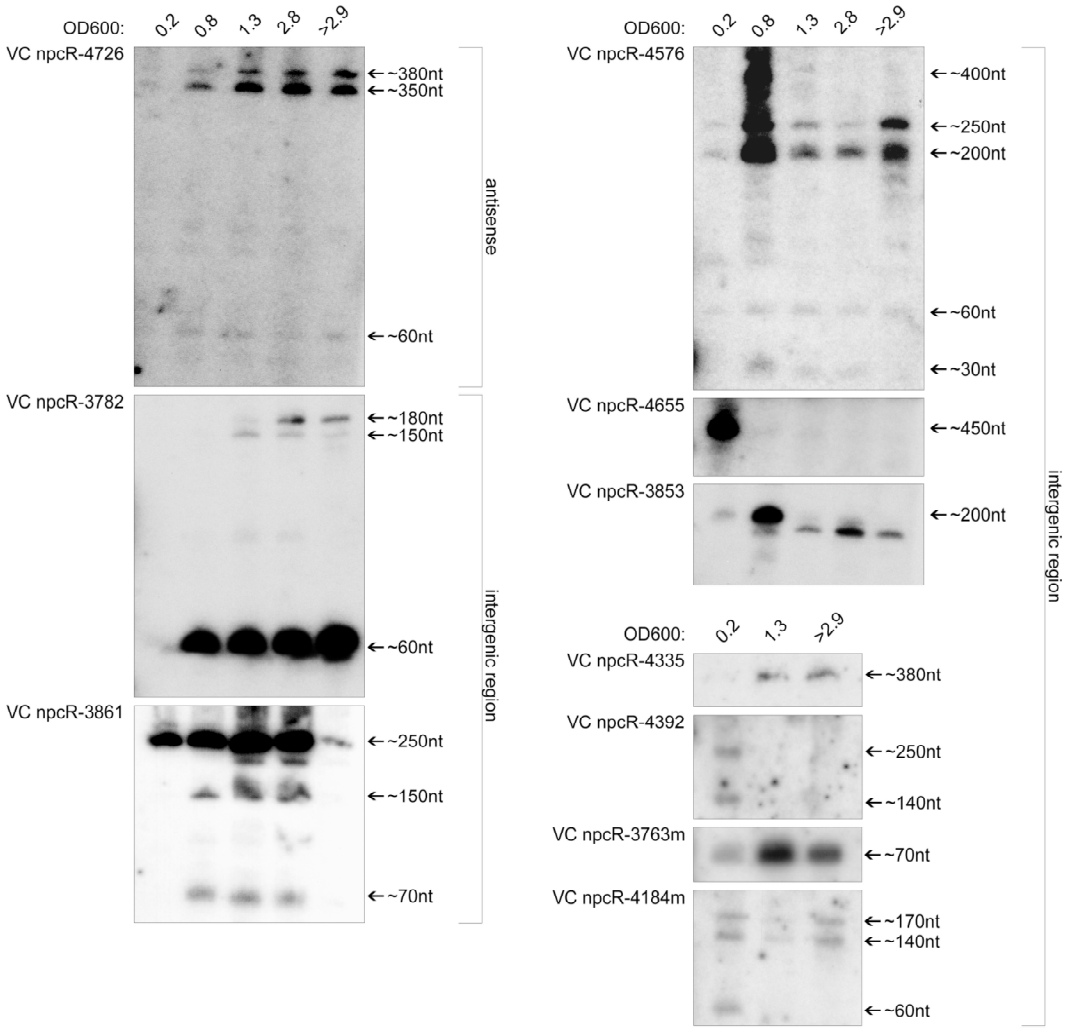
Supplementary Figure 3.



Supplementary Figure 4.



Supplementary Figure 5.



Supplementary Figure 6.