Published online 13 January 2017 *Nucleic Acids Research, 2017, Vol. 45, No. 5* ***2329–2340***
doi: 10.1093/nar/gkw1316

Identification of endoribonuclease specific cleavage positions reveals novel targets of RNase III in *Streptococcus pyogenes*

Anaïs Le Rhun[1,2,3,†], Anne-Laure Lécrivain[1,2,†], Johan Reimegård[4], Estelle Proux-Wéra[5], Laura Broglia[2,3], Cristina Della Beffa[3] and Emmanuelle Charpentier[1,2,3,6,*]

[1]The Laboratory for Molecular Infection Sweden (MIMS), Umeå Centre for Microbial Research (UCMR), Department of Molecular Biology, Umeå University, S-90187 Umeå, Sweden, [2]Max Planck Institute for Infection Biology, Department of Regulation in Infection Biology, D-10117 Berlin, Germany, [3]Helmholtz Centre for Infection Research, Department of Regulation in Infection Biology, D-38124 Braunschweig, Germany, [4]Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, S-75123 Uppsala, Sweden, [5]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Box 1031, SE-17121 Solna, Sweden and [6]Humboldt University, D-10115 Berlin, Germany

Received November 10, 2016; Revised December 15, 2016; Editorial Decision December 18, 2016; Accepted January 02, 2017

ABSTRACT

A better understanding of transcriptional and post-transcriptional regulation of gene expression in bacteria relies on studying their transcriptome. RNA sequencing methods are used not only to assess RNA abundance but also the exact boundaries of primary and processed transcripts. Here, we developed a method, called identification of specific cleavage position (ISCP), which enables the identification of direct endoribonuclease targets *in vivo* by comparing the 5′ and 3′ ends of processed transcripts between wild type and RNase deficient strains. To demonstrate the ISCP method, we used as a model the double-stranded specific RNase III in the human pathogen *Streptococcus pyogenes*. We mapped 92 specific cleavage positions (SCPs) among which, 48 were previously described and 44 are new, with the characteristic 2 nucleotides 3′ overhang of RNase III. Most SCPs were located in untranslated regions of RNAs. We screened for RNase III targets using transcriptomic differential expression analysis (DEA) and compared those with the RNase III targets identified using the ISCP method. Our study shows that in *S. pyogenes*, under standard growth conditions, RNase III has a limited impact both on antisense transcripts and on global gene expression with the expression of most of the affected genes being down-regulated in an RNase III deletion mutant.

INTRODUCTION

High-throughput RNA sequencing is a method of choice to provide insights into transcriptional and post-transcriptional regulation of gene expression, allowing the study of transcript expression levels and boundaries. Notably, during the past few years, RNA sequencing methodologies in bacteria have been improved with the introduction of mapping of 5′ transcriptional start sites and 3′ transcriptional termination sites (1–7).

The fate of transcripts is mostly controlled by post-transcriptional regulators like ribonucleases (RNases) that perform RNA processing and/or have decay activities. While processing leads either to maturation (active RNA) or targeted destabilization, RNA decay entails bulk RNA degradation (turnover) (8). Transcriptomic analyses comparing the steady-state levels and/or RNA stability in strains lacking specific RNases with their isogenic wild types (WT) have been useful in providing insights into the biological functions of RNases in bacteria (9–14). However, these techniques do not allow the discrimination between direct and indirect effects thereby making it difficult to assess the exact roles of studied RNases.

With the development of RNA sequencing, the identification of transcripts targeted by RNases became possible; e.g. by combining RNA immunoprecipitation or co-immunoprecipitation with an RNase or a cleavage-defective mutant (15,16). In particular, RNA sequencing offers the possibility to differentiate between the 5′ ends of primary and processed transcripts by specifically depleting the processed transcripts during cDNA library preparation using the Terminator exonuclease (Terminator 5′-Phosphate-

*To whom correspondence should be addressed. Tel: +49 30 28460 410; Fax: +49 30 28460 412; Email: charpentier@mpiib-berlin.mpg.de
† These authors contributed equally to this work as the first authors.

Dependent Exonuclease) enzyme (3,4). Knowing the exact 5′ ends of processed transcripts has improved the understanding of RNase target specificity (17–21). However, until now, except in the case of RNases producing 2′, 3′-cyclic phosphates (22), the exact 3′ ends of RNase processed transcripts have not yet been investigated.

In this study, we present a method, the identification of RNase specific cleavage positions (ISCP), to screen for processing sites (specific cleavage positions, SCPs) obtained by comparing both the 5′ and the 3′ ends of processed transcripts in WT and RNase deletion strains. As a proof of concept, we applied this method to study endoribonuclease III. RNase III cleaves within secondary structures or RNA duplexes (double-stranded (ds) RNA processing) and is widely conserved among bacteria and eukaryotes. Processing by this RNase generates fragments with a distinct 3′ overhang of two nucleotides (nt) (23). In addition to its roles in ribosomal RNA (rRNA) maturation (24), bacterial RNase III regulates genes involved in virulence (25), stress response (26) and other cellular processes. In *Streptococcus pyogenes*, a human pathogen responsible for a wide range of mild to life-threatening diseases, RNase III has been shown to process tracrRNA and the precursor forms of CRISPR RNAs of the CRISPR-Cas9 system (27) and to regulate some of the recently identified sRNAs (28). However, the role of RNase III in this bacterium has not been studied in detail. Using the ISCP method, we retrieved 92 SCPs of RNase III in *S. pyogenes*. To understand the global impact of these *in vivo* cleavages, we combined the obtained SCPs with transcript abundance profiling. We provide here a comprehensive analysis of the effect of RNase III cleavage on transcript expression, with new insights into the biological functions of RNase III in *S. pyogenes*.

MATERIALS AND METHODS

Culture conditions

S. pyogenes SF370 (M1GAS) and isogenic derivatives (Supplementary Table S1) were cultured at 37°C, 5% CO_2, without shaking, in THY medium (Todd Hewitt Broth (THB, Bacto, Becton Dickinson) complemented with 0.2% yeast extract (Servabacter)) or TSA (trypticase soy agar, BD Difco) supplemented with 3% sheep blood. *Escherichia coli* was grown at 37°C with shaking in Luria–Bertani medium and agar. When needed, antibiotics were used at the following concentrations: 100 μg/ml ampicillin and 25 μg/ml kanamycin for *E. coli*; 3 μg/ml erythromycin and 300 μg/ml kanamycin for *S. pyogenes*. Growth curves were performed in triplicates according to the following procedure. Overnight (O/N) cultures were diluted to an optical density at 620 nm ($OD_{620\,nm}$) of 0.2 in 5 ml of medium and pelleted at 3200 × g for 5 min. The pellets were resuspended in 500 μl of THY and used to inoculate 50 ml of medium for a starting $OD_{620\,nm}$ of 0.02. The cell growth was monitored at $OD_{620\,nm}$ using a microplate reader (Biotek PowerWave XS2).

Strain integrity

Plasmid and strain integrity were checked by PCR and DNA sequencing (LGC, Germany). All strains, plasmids and oligos used in this study are listed in Supplementary Table S1. The deletion of *rnc* (encoding RNase III) and *rny* (encoding RNase Y) genes did not affect the expression of surrounding genes.

Bacterial transformation

S. pyogenes competent cells for in-frame deletion of *rnc* and *rny* genes were generated according to the procedure described by (29). Briefly, the cells were grown in THY complemented with 250 mM sucrose (Calbiochem) and 40 mM L-threonine (Sigma-Aldrich) until $OD_{620\,nm}$ reached 0.25. The cells were washed with 0.5 M sucrose and resuspended in 0.5 M sucrose and 20% glycerol. The $OD_{620\,nm}$ of the competent cells was adjusted to 1 before electroporation in a 0.1 cm electrode gap cuvette with a 1.8 kV, 400 Ω and 25 μF pulse. Electro-competent cells for transformation of *S. pyogenes* in-frame deletion mutant with plasmids followed the procedure described by (30). The $OD_{620\,nm}$ of the competent cells was adjusted to 1 before mixing 100 μl of cells with 500 ng of circular plasmid. The cells were electroporated in a 0.1 cm electrode gap cuvette with a pulse of 1.5 kV, 400 Ω and 25 μF. Top10 *E. coli* strain was used for clonings and transformations were done according to the standard heat shock procedure (31).

Construction of gene deletion mutants

Chromosomal deletions of *rnc* and *rny* genes in *S. pyogenes* were generated using the Cre-Lox system (32). The upstream and downstream regions of *rny* were amplified by PCR from the WT genomic DNA, using the pairs of primers OLEC3338/OLEC1999 and OLEC2000/OLEC3339, respectively. The upstream and downstream regions of *rnc* were amplified with OLEC1654/OLEC3342 and OLEC3343/OLEC1655 (the first 50 nt and the last 50 nt of *rnc,* containing the RBS of the downstream gene, were not removed). The fragments were then ligated with the lox71-*ermAM/B*-lox66 cassette (*ermAM/B* amplified from pMSP3535 with OLEC1943/OLEC1932) by PCR-mediated ligation with OLEC3338/OLEC3339 for *rny* and OLEC1654/OLEC1655 for *rnc*. The resulting fragments were cloned into pUC19 (suicide vector for *S. pyogenes*) using BamHI for *rny* and EcoRI for *rnc*. The produced plasmids were linearized with PvuI (cleaving in the ampicillin resistance gene) before transformation in *S. pyogenes*. The integration of the cassette at the correct location was checked by PCR and DNA sequencing in selected erythromycin resistant clones. To remove the *ermAM/B* cassette, electro-competent cells of *S. pyogenes* strains with the integrated cassette were transformed with pEC455 encoding the Cre recombinase (excision of DNA between the two lox sites and replacement of the cassette with a lox72 site). Kanamycin-resistant and erythromycin-sensitive clones were cultured without antibiotic to lose the plasmid. To generate the Δ*rny*_Δ*rnc* double mutant, the *rnc* gene was deleted from the Δ*rny* strain using the procedure described above.

Construction of complemented Δ*rnc* strain

For complementation purposes, *rnc* gene was amplified by PCR from the WT genomic DNA with its constitutive

promoter using the primer pair OLEC1667/OLEC1668. The resulting P*rnc-rnc* fragment was digested with BamHI/EcoRI before being cloned in the shuttle vector pEC85 and transformed in the deletion mutant strain. The control plasmid was transformed in the WT and Δ*rnc* strains as a control.

RNA extraction

Biological triplicates of *S. pyogenes* WT, Δ*rnc*, Δ*rny* and Δ*rny*_Δ*rnc* were collected from independent cultures at the mid-logarithmic growth phase ($OD_{620\,nm}$ of 0.25). Bacterial cultures were mixed with an equal volume of acetone/ethanol (1:1) solution. Total RNAs were extracted using TRIzol (Invitrogen)/chloroform and precipitated with isopropanol.

Primer extension assay

Primer extension was performed on total RNA (10 μg) using the SuperScript III Reverse Transcriptase (Invitrogen), according to the manufacturer's instructions with the following modifications. RNA samples were annealed with 3 μl of radiolabeled primers at 65°C during 30 min and kept on ice for 1 min. The reverse transcription was done at 55°C during 1 h (50°C for OLEC3926). The cDNA samples were then precipitated with ethanol, resuspended in 5 μl of 1X loading dye and resolved on 10% polyacrylamide/8 M urea/TBE gels. The primers were end-radiolabeled as previously described (28) and the AFLP 30–300 bp ladder (Invitrogen) was labeled according the manufacturer's instructions using the T4 Polynucleotide kinase (PNK) (Thermo Fischer Scientific).

RNA sequencing

Following Turbo DNase (Ambion) treatment, the RNA integrity was checked using a bioanalyzer (RIN > 8). Total RNAs were sent for cDNA library preparation and sequencing at the genome analytics platform of the Helmholtz Centre for Infection Research (Braunschweig, Germany). Different procedures for generating cDNA libraries were applied following the description of (33) with some modifications (Supplementary Figure S1). Briefly, most rRNAs were removed using MICROBExpress Bacterial mRNA Enrichment kit (Ambion). Libraries containing both primary (5′ PPP) and processed (5′ P/5′ OH) transcripts (*Pp* libraries) were obtained using RNAs treated with Tobacco acid phosphatase (TAP, Epicentre), which transforms 5′ PPP of primary transcripts to 5′ P. In the libraries enriched in 5′ processed transcripts (*p* libraries), the RNAs were not treated with TAP, hence 5′ adapters could not be ligated to 5′ PPP transcripts. To produce libraries enriched in primary transcripts (*P* libraries), the RNAs were treated with Terminator exonuclease, which degrades 5′ processed transcripts, followed by a TAP treatment (the *P* libraries were only used to define operon boundaries). The RNAs were fragmented by sonication (Covaris) to obtained fragments of 200 nt. The T4 PNK (Fermentas) was used to phosphorylate 5′ OH ends and convert 3′ P to 3′ OH. 3′ and 5′ adapters were ligated to the RNA fragments using T4 RNA ligase (ThermoFisher Scientific). The reverse transcription was done using SuperScript II (Invitrogen) followed by 15 cycles of PCR with Phusion (ThermoFisher Scientific) and agarose gel purification. The sequencing was performed using either 1 × 50 bp single end read (HiSeq2500) for *Pp* libraries (quantification of differential expression) or 2 × 100 bp paired end reads (HiSeq2500) for *Pp*, *P* and *p* libraries (ISCP). The reads were trimmed for quality (-q 10) and adapter sequences were removed using Cutadapt v.1.3 (34). STAR v2.3 (35) was used to map the reads on the *S. pyogenes* reference genome NC_002737.1 allowing for no introns (–alignIntronMin 100 –alignIntronMax 101). The number of mapped reads is available in Supplementary Table S2. Note that this technique does not allow the retrieval of small transcripts. The coverage was visualized using the integrative genomics viewer (36).

Differential expression analysis

A read count table was built from the BAM files using HTseq-count v. 0.6.1 (37) and the *S. pyogenes* reference genome NC_002737.1 combined with novel identified sRNAs (28). This count table was used as input file for the Bioconductor package DESeq2 v1.8.1 (38) to perform the differential gene expression analysis (DEA). The false rate discovery threshold was set at 0.01 and the log2 fold change (log2 FC) at 1.

Transcriptome expression

To calculate the level of expression in the WT strains (Supplementary Table S2A), we used the bioconductor package 'edgeR' (39). The reads per kilobase of transcript per million (RPKM) mapped reads values were then averaged and gene expression was defined using the threshold RPKM ≥ 10. Transcript starting sites of interest were determined by eye screening using the WT *P* library and the rho-independent terminators (predicted by TransTermHP v2.07) (40) were downloaded from http://transterm.cbcb.umd.edu/tt/Streptococcus_pyogenes_M1_GAS.tt.

Identification of specific cleavage positions of endoribonucleases

A python in-house algorithm was used to count the total coverage, the 5′ and 3′ end coverage of all mapped reads at each nucleotide position, for each DNA strand. To reduce possible background noise and potential bias linked to low expression, only positions with a minimum of 10 reads mapped in the WT and the mutant were considered. RNase dependent 5′ or 3′ end positions were identified using four criteria (Figure 1) with a conservative cut-off to avoid false positives: (i) The end coverage from both Δ*rnc* and the reference (WT or Δ*rny*) was ≥ 10 in *p* libraries (≥ 8 in *Pp* libraries). (ii) The percentage of read ends at the tested position compared to all read ends in a window of 20 nt, i.e all reads that end from 9 nt upstream to 10 nt downstream of the position of interest, was equal to or above 16% in *p* libraries and 10% for both in *Pp* libraries in the reference sample. (iii) The percentage of read ends compared to all reads at the specific position was equal to or above to 5% in

p library and 2% in *Pp* library in the reference sample. (iv) The relative (WT/Δ*rnc*) percentage of read ends calculated in (iii) was equal to or above to 3 in *p* library and 2 in *Pp* library. For a more stringent screening, the analysis was also done using Δ*rny*_Δ*rnc* compared to WT and Δ*rny*.

Equations used in the algorithm:

$$TotalCount: x_i^{tot,sample}$$

$$EndCount: x_i^{end,sample}$$

$$end = \{5'end, 3'end\}$$

$$sample = \{mutant, reference\}$$

$$(\boldsymbol{i}) \Rightarrow x_i^{end,reference} + x_i^{end,mutant} \geq EndCountCutoff$$

$$(\boldsymbol{ii}) \Rightarrow \frac{x_i^{end,reference}}{\sum_{j \in \boldsymbol{w}} x_j^{end,reference}} \geq IntraWindowCutoff$$

$$\boldsymbol{w} = \{i - us, i - us + 1, \ldots, i + ds - 1, i + ds\}$$

$$us = floor\left(\frac{windowSize - 1}{2}\right)$$

$$ds = roof\left(\frac{windowSize - 1}{2}\right)$$

$$(\boldsymbol{iii}) \Rightarrow \frac{x_i^{end,reference}}{x_i^{total,reference}} \geq IntraTotalCutoff$$

$$(\boldsymbol{iv}) \Rightarrow \left(\frac{x_i^{end,reference}}{x_i^{tot,reference}}\right) \Big/ \left(\frac{x_i^{end,mutant}}{x_i^{tot,mutant}}\right) \geq InterTotalCutoff$$

$$EndCountCutoff = \begin{cases} 10 & (\text{p } Library) \\ 8 & (\text{Pp } Library) \end{cases}$$

$$windowSize = 20$$

$$IntraTotalCutoff = \begin{cases} 0.05 & (\text{p } Library) \\ 0.02 & (\text{Pp } Library) \end{cases}$$

$$IntraWindowCutoff = \begin{cases} 0.16 & (\text{p } Library) \\ 0.10 & (\text{Pp } Library) \end{cases}$$

$$InterTotalCutoff = \begin{cases} 3 & (\text{p } Library) \\ 2 & (\text{Pp } Library) \end{cases}$$

SF370 genome re-annotation

The genome NC_002737.1 was re-annotated by BILS (Bioinformatics Services to Swedish Life Science, Sweden) using the Prokka package (41). A set of curated protein sequences (taxonomic group '*Streptococcus*') was retrieved from the Uniprot database with proteomic or transcriptomic information. Gene name and functional descriptions were selected for each gene based on the Uniprot and Kegg databases. The new track generated by Prokka was manually checked and curated for the genes that were not assigned their published names after Prokka re-annotation.

Antisense RNAs

To compare the proportion of antisense RNA regions between the strains, the expressed transcriptomes of WT and Δ*rnc* were determined using the number of nucleotides with at least 10 reads mapped on one DNA strand divided by the genome length (1 852 441 nt) in the TAP treated samples (Supplementary Table S2B). We defined regions with antisense transcription as regions longer than 40 nt with at least 10 reads mapped on each DNA strand. When less than 20 nt were present between two asRNA regions, these two regions were combined. Finally, the total length of asRNA regions was normalized against the transcriptome.

RNA folding

RNA folding of RNAs cleaved by RNase III was done using RNA fold (RNAfold -p -d2 –noLP) (42). The sequences selected corresponded to either 50 nt each side around a single cleavage position or 20 nt upstream of the first cleavage position and 20 nt downstream of the last cleavage position (or where the transcription stopped). rRNA base pairing was computed using 50 nt sequence centered on each SCP framing 16S and 23S rRNAs with IntaRNA (43). The 'Max. basepair distance' and 'Folding window size target' parameters were set at 1 and 50, respectively.

Plots

All the statistical analyses and figures were performed in the R statistical environment (Team 2014), R version 3.1.1, using the package 'graphics'.

RESULTS

In this study, we investigated the direct and indirect effects of RNase III from *S. pyogenes*. We compared the transcriptomes of the WT and RNase III deletion strain (Δ*rnc*) performing DEA and ISCP. The DEA allows the identification of both direct and indirect targets of RNase III, while the ISCP pinpoints, to the nucleotide, the cleavage positions in RNA targets.

Development of ISCP to identify direct targets of endoribonucleases

RNA transcripts are divided in primary transcripts (5′ triphosphate: 5′ PPP), starting at transcriptional start sites and processed transcripts, resulting from endoribonuclease or exoribonuclease activities (5′ monophosphate or hydroxyl: 5′ P or 5′ OH). To identify direct targets of RNase III, we generated cDNA libraries containing both primary and processed transcripts (*Pp* libraries) or enriched in 5′ processed transcripts (*p* libraries) for WT, Δ*rnc* and additional control strains (RNase Y deletion (Δ*rny*) and RNase Y and III double deletion (Δ*rny*_Δ*rnc*)) (Supplementary Figure S1; Supplementary Table S2; Materials and Methods). The coverage and the number of reads ending at each nucleotide position (5′ start or 3′ end) were calculated.

The number of read ends resulting from an endoribonucleolytic cleavage by RNase III in the WT (or another reference strain, i.e. the strain deleted from the gene encoding

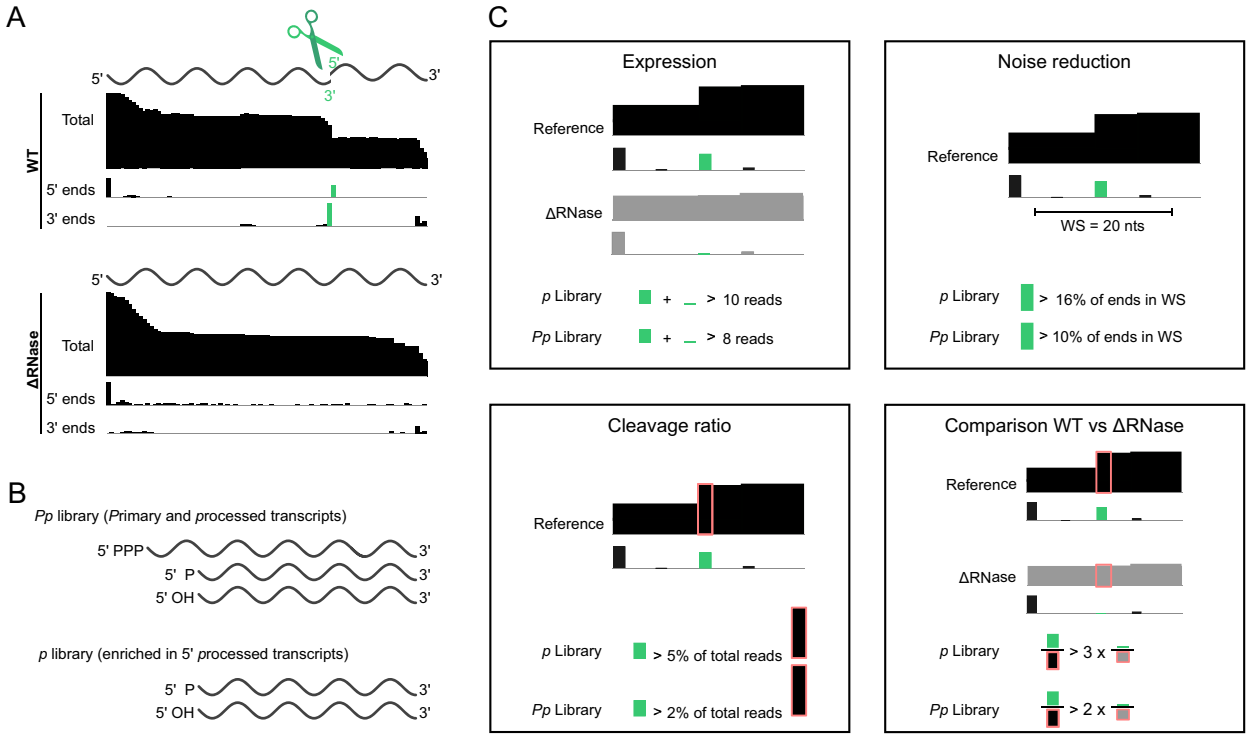

Figure 1. Identification of specific cleavage positions (ISCP). (**A**) Example of an RNA molecule (curved black line) which is processed by an endoribonuclease (green) generating two RNA molecules. The new 3′ and 5′ ends produced (labeled in green) are detected by RNA sequencing in the WT (green bars) but are absent when the RNase responsible for the cleavage is deleted (ΔRNase). 'Total': total read coverage; '5′ ends' and '3′ ends': number of reads starting or ending respectively at each nucleotide (black bars). (**B**) ISCP was done by comparing processed transcript ends in WT and ΔRNase strains. *Pp* libraries containing primary (5′ PPP) and processed (5′ P/OH) transcripts and *p* libraries enriched in 5′ processed transcripts were used to identify specific cleavage positions (SCPs). (**C**) SCPs were defined by several criteria. Lower cut-offs were chosen for the *Pp* libraries compared to the *p* libraries since *Pp* libraries contain a reduced relative amount of processed transcripts. We only considered nucleotides with more than 10 reads mapped as expressed. The first parameter ('Expression') selects positions with a sufficient total number of read ends in both ΔRNase and the reference (WT or a different ΔRNase) strains (green bars). The second parameter ('Noise reduction') evaluates if the 5′ or 3′ end of interest (green bar) is distinct from the noise, defined as the total of 5′ or 3′ ends (black bars) in a window size (WS) of 20 nt. To detect only relevant positions, the third parameter ('Cleavage ratio') computes the amount of cleaved transcripts by comparing the count of the 5′ or 3′ ends of interest (green bar) with the total number of reads covering the position (pink rectangle). The fourth parameter ('Comparison WT versus ΔRNase') selected positions where the cleavage ratio (end count normalized to the total coverage, i.e. ratio green bar/pink rectangle) of the reference was superior to the one in the ΔRNase strain. This parameter also prevents any artefact due to a possible change in RNA level between the two analyzed strains.

another RNase, in the present case Δ*rny*) should substantially decrease in Δ*rnc*. In order to find these SCPs, several parameters described below were chosen to compare Δ*rnc* with WT and Δ*rnc* with Δ*rny*, respectively (Figure 1). Only positions with more than 10 mapped reads were included in the analysis. (i) First, to reduce false positive detection, putative cleavage positions were taken into account only when the sum of the read ends of the reference and of the RNase deletion strains was greater than or equal to 10 or 8 in the *p* and *Pp* libraries, respectively. (ii) Then, we minimized the background noise in the reference strains expressing RNase III by specifying that the number of read ends at the tested positions should account for a sufficient percentage of the total amount of read ends in a 20 nt window (16% and 10% in the *p* and *Pp* libraries, respectively). We used lower cut-offs for the *Pp* libraries than for the *p* libraries that contain more processed transcripts. (iii) To make sure that the cleavages affected a sufficient fraction of the transcripts, we stipulated that the cleavage ratio (i.e. number of read ends/total number of reads at the position) should be at least 5% in *p* libraries (2% in *Pp* libraries). (iv) Finally, the read ends at the cleavage position should be present in both reference strains at least 3 times more than in Δ*rnc* in *p* libraries (2 times in *Pp* libraries). The same analysis was repeated comparing Δ*rny_rnc* (negative control for SCPs of RNase III) with WT and Δ*rny*. Only positions absent both in Δ*rnc* and Δ*rny_rnc* and present both in WT and Δ*rny* were further analyzed.

An SCP corresponds to the nucleotide detected at an end of a transcript, thus the position of the cleavage is directly upstream or downstream of this nucleotide, depending on the identified end of the transcript (5′ or 3′) (Figure 1). In a few cases, for the same cleavage event, we were able to retrieve both the 3′ end and the 5′ end of the two split transcripts. Therefore, this approach allows the identification of cleavage events regardless of the possible degradation of the remaining 5′ or 3′ processed transcripts. RNase III that cleaves ds RNA can generate four SCPs per cleavage event instead of two for single-stranded (ss) specific RNases. Thus, RNase III processing events are a suitable model to test our method.

Using the above-mentioned method with Δ*rnc* strain, we detected 92 putative SCPs located in rRNAs (42), in

mRNAs (27) and in sRNAs (23) (Supplementary Tables S3 and S4). We validated four targets by primer extension analysis (SPy_1227-*bmpA*, SPy_0129, *nrdR* and 23S methyl regulatory element), with both cleavage positions of RNase III visible (Supplementary Figure S2). We also retrieved RNase III SCPs already described in the literature, e.g. RNase III SCPs in the six rRNA operons (Supplementary Figure S3; Supplementary Table S4), and in the duplex formed by two sRNAs, tracrRNA and pre-CRISPR RNA, from the type II CRISPR-Cas9 bacterial immune system (27). This validates both our RNA sequencing method and our analysis algorithm.

RNase III activity on ds RNAs leads to two single cleavage positions for a same cleavage event. We could detect 24 RNase III single cleavages from transcript 5′ and 3′ ends, 17 were identified only from transcript 5′ ends and 24 from transcript 3′ ends. Thus, the study of 3′ processed ends in addition to 5′ processed ends allows the identification of additional SPCs.

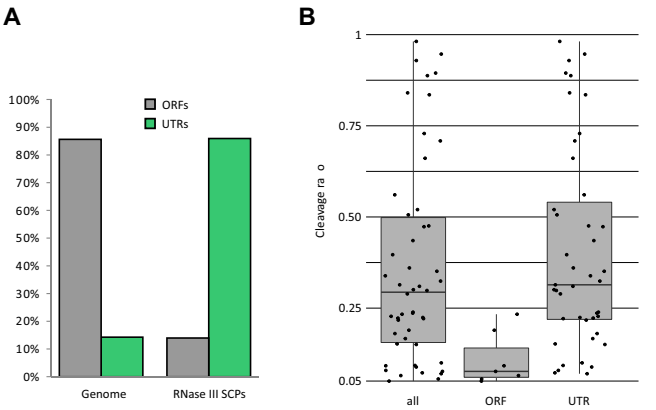

Figure 2. RNase III cleavage ratios. (**A**) Proportion of untranslated regions (UTRs) and open reading frames (ORFs) in the genome and among RNase III specific cleavage positions (SCPs). (**B**) The cleavage ratio of each SCP was plotted for each SCP location (ORF or UTR). Scatter box plot of RNase III cleavage ratios show the first and third quartiles of the data range (top and the bottom of the grey boxes) and the median of the data (line inside the boxes).

pre-rRNA maturation by RNase III

As mentioned previously, a substantial number of RNase III SCPs (42) were identified in each of the six ribosomal RNA operons from *S. pyogenes* (Supplementary Table S4). These ribosomal operons are transcribed as long 30S pre-RNA transcripts containing the 16S rRNA, alanine tRNA, 23S rRNA and 5S rRNA (Supplementary Figure S3A). The identified SCPs were located upstream and downstream of 16S and 23S rRNAs (Supplementary Figure S3A and B) and the base pairing of these regions revealed the characteristic 2 nt 3′ overhang generated by the cleavages (Supplementary Figure S3C). RNase III processing is described as the first step of rRNA and tRNA processing (44,45), which are further matured by other RNases (24).

We also visualized substitute processing events in the 16S and 23S pre-rRNAs occurring in Δ*rnc* but not in the WT (Supplementary Figure S3B). To find substitute SCPs (Sub-SCPs), we applied our ISCP algorithm but this time we compared Δ*rnc* and Δ*rny*_Δ*rnc* strains (carrying Sub-SCPs) with the WT and Δ*rny* strains (no Sub-SCPs). We obtained 53 Sub-SCPs including 44 located in pre-rRNAs (Supplementary Table S5). We detected three Sub-SCPs downstream of the 16S rRNA (including one in the mature alanine tRNA) and four Sub-SCPs (corresponding to two cleavage events) surrounding the 23S RNA for all pre-rRNA transcripts (Supplementary Figure S3A and B).

The Sub-SCP located in the alanine tRNA (orange arrow) would generate 13 nt-truncated alanine tRNAs (Supplementary Figure S3A and B) with unknown effects on tRNA levels or activity. The two others Sub-SCPs located downstream of the 16S rRNA (grey arrows) are probably incorrect Sub-SCPs that are cleavages independent of RNase III presence or absence (Supplementary Figure S3A and B). Each Sub-SCPs, close to RNase III SCP number 2, generated a small fragment in the WT (not sequenced here because of the sequencing size limitation), and a longer fragment in Δ*rnc*. The first 5′ end position was located 2 nt downstream of 16S rRNA mature 3′ end (Supplementary Figure S3B). Di Chiara *et al.* described a similar cleavage event in *B. subtilis* WT and Δ*rnc* strains (21). We predicted that this Sub-SCP corresponds to the expected natural 16S rRNA 3′ end maturation event performed by a still unknown RNase, both in WT and Δ*rnc*. In the WT, a short fragment of 34 nt is created between this processed end and RNase III SCP (number 2). The second processed end detected as Sub-SCP corresponded to the 3′ end (upstream fragment) of the cleavage generating the mature alanine tRNA (Supplementary Figure S3A). The 5′ end of the downstream fragment, originating from the cleavage was visible in both strains. In the WT, the upstream fragment (between RNase III cleavage and this event) was 12 nt and was not sequenced.

The four Sub-SCPs surrounding the 23S were detected at the 3′ end of the upstream fragments and corresponded to two cleavage events (Supplementary Figure S3B; Supplementary Table S5). This indicates that the RNase(s) responsible could be unspecific endoRNase(s) or 3′ exoribonuclease(s). Indeed, the Sub-SCPs downstream of the 23S rRNA were located in the first internal loop (Supplementary Figure S3B), suggesting that the stem (not cleaved by RNase III in Δ*rnc*) could block 3′ exoribonucleic activity.

RNase III SCPs are preferentially located in UTRs

Fifty RNase III SCPs were identified in mRNA (27; 20 in untranslated regions (UTRs) and 7 in open reading frames (ORFs)) and sRNA (23) regions (Supplementary Table S3), including 43 SCPs located in non-coding regions (sRNAs, UTRs). This is six times more than expected if the SCPs were randomly distributed over the genome (ORFs representing 85.7% of the genome) (Figure 2A). Although the level of ds structures in UTRs and ORFs was shown to be comparable in *E. coli* (46), RNase III cleaving preferentially in UTRs could be explained by the fact that RNA structure formation is prevented in ORFs undergoing translation (23). The cleavage ratio by RNase III was extremely diverse, ranking from 5% (cut-off value of the algorithm) to

98% of the reads (36% average, 29% median) (Figure 2B). In the case of SCPs located in ORFs (only seven SCPs in six ORFs), the maximum cleavage ratio observed (25%) was lower than the median, supporting the hypothesis that the translation machinery would limit the access of RNase III to the cleavage sites.

RNase III target classification

In total, 25 transcripts cleaved by RNase III were identified, excluding rRNA transcripts (Supplementary Table S3). The regions surrounding the SCPs were folded and the targets were further divided in two categories, depending on whether RNase III cleavage was in a stem loop (SL) or in an antisense duplex (AS) (Figure 3A; Supplementary Table S3; Supplementary Figure S4). The targets harboring single SCPs, i.e. for which the second position of the cleavage event was not identified, were gathered in a third category and referred to as stand-alone SCPs (SA) (Figure 3A; Supplementary Figure S4).

The 13 targets classified in SL were found in CDSs, UTRs or in sRNAs. For example, a previously described SCP detected in a stem loop located in the 5′ UTR of *pnpA*, generating the 52 nt sRNA (Spy_sRNA1620465) (28), was associated with an upregulation of 2.4 times of *pnpA* mRNA (Figure 3B; Supplementary Table S3; Supplementary Figure S4). A global quantitative proteomic analysis performed in our lab revealed that PNPase, the 3′ to 5′ exoribonuclease PNPase encoded by *pnpA*, was four times more abundant in Δ*rnc* than the WT (data not shown). Thus, RNase III negatively regulates *pnpA* mRNA and PNPase expression in *S. pyogenes*, as described for other bacteria (47,48).

Although transcription antisense to 10 out of the 25 transcripts with SCPs was observed, we retrieved putative cleavage sites with 2 nt 3′ overhangs in the antisense transcripts in only two cases (Supplementary Table S3; Supplementary Figure S4). These regions were located in the UTRs upstream of two integrase genes, *int1* and *int3*, from *S. pyogenes* prophages. In *E. coli*, RNase III negatively regulates the bacteriophage λ integrase expression by cleaving in the 3′ UTR of the transcript and leading to its destabilization (49,50). We also identified a SCP in the UTR of another integrase gene that belongs to the prophage SF370.3, *int2*, but we did not detect the second SCP in the antisense transcript (Supplementary Table S3; Supplementary Figure S4), thus this SCP was classified as SA. We could not detect any expression for the *int1* and *int2* genes in the WT or in Δ*rnc*. The same levels of *int3* transcript were detected in the WT and in Δ*rnc*, thus we could not conclude on the role of RNase III in prophage biology.

The majority of SCPs categorized as SA were positioned in single-stranded regions or 1 nt downstream of a loop (Supplementary Figure S4). We hypothesized that either (i) the part of the RNA containing the retrieved SCP hybridized with another RNA molecule (mRNA, sRNA) that was not identified here; or (ii) the second processed transcript was not visible because it was rapidly degraded; or (iii) one strand was preferentially cleaved by RNase III; (iv) the computed secondary structure was not accurate and we could not predict the second cleavage position in the stem loop or (v) these SCPs were the result of a subsequent exonucleolytic action.

RNase III uncouples *secY and adk* transcript abundance

We identified four SCPs located in the UTR between *secY* (Sec translocase) and *adk* (adenylate kinase) (Figure 3C), two genes that are transcribed from a long operon mainly encoding ribosomal proteins. Two of these SCPs (numbered 3 and 4) formed a 2 nt 3′ overhang in a predicted stem loop (Figure 3C). RNase III deletion only affected the expression of the downstream mRNA, *adk*, which was 3.3 times more abundant in Δ*rnc* (Supplementary Table S6). This indicates that RNase III processing probably uncouples the regulation of *secY-adk* abundance. In *S. aureus, secY* mRNA has already been reported as a target of RNase III, however, the cleavage was shown to occur within the ORF and not in the 3′ UTR (15). It was also described in this bacterial pathogen that *secY* mRNA was more abundant in a Δ*rnc* mutant (15), indicating differences in *secY* regulation between these bacterial species.

Example of overlapping 3′ UTRs independently cleaved by RNase III

We identified eight SCPs in the 3′ UTRs of SPy_0593 (uncharacterized protein) and *lysS* (lysyl tRNA synthetase), both transcribed from opposite strands and partly overlapping (Figure 3D; Supplementary Figure S4). *lysS* (reverse strand) possesses a long 3′ UTR (>600 nt) that is antisense to the entire SPy_0593 transcript (forward strand). Two sRNAs were previously identified in the antisense region; Spy_sRNA477714 in the 3′ UTR of SPy_0593 and Spy_sRNA477741 in the 3′ UTR of *lysS*. Spy_sRNA477741 abundance was shown to be regulated by both RNases III and Y (28). Here, we show that the processing of both UTRs by RNase III leads to the formation of these sRNAs. We first hypothesized that these antisense transcripts could base pair and be cleaved by RNase III at two positions interspaced by ∼30 nt, forming 2 nt 3′ overhang ends (SCPs numbered 3/4 and 5/7) (Supplementary Figure S4). Thus, the four additional SCPs retrieved in this region (SCPs 1, 2, 6, 8) could result from further 5′ to 3′ or 3′ to 5′ degradation of the cleaved transcripts. However, these RNAs are located in a nearly perfect palindromic region of 52 nt and the independent folding of each strand also formed two stem loops with four possible pairs of 2 nt 3′ overhang that could account for all eight detected SCPs (Figure 3D). In this case, one pair would correspond to the cleavage of SPy_0593 mRNA 3′ UTR (SCPs 4/7) and three pairs, separated by 1 or 2 nt, would be located in the 3′ UTR of *lysS* mRNA (SCPs 1/8, 2/6 and 3/5) (Figure 3D). Despite the highly similar sequence of both stem loops, an additional bulge close to the cleavage site that is only present in the 3′ UTR of *lysS* mRNA could explain the different cleavage pattern observed. Thus, we believe that RNase III preferentially processes the two stem loops rather than the duplex of asRNAs. This targeting by RNase III leads to a decrease in abundance of both sRNAs (FC of 2.3 for Spy_sRNA477714 and 6.4 for Spy_sRNA477741) (Supplementary Table S6),

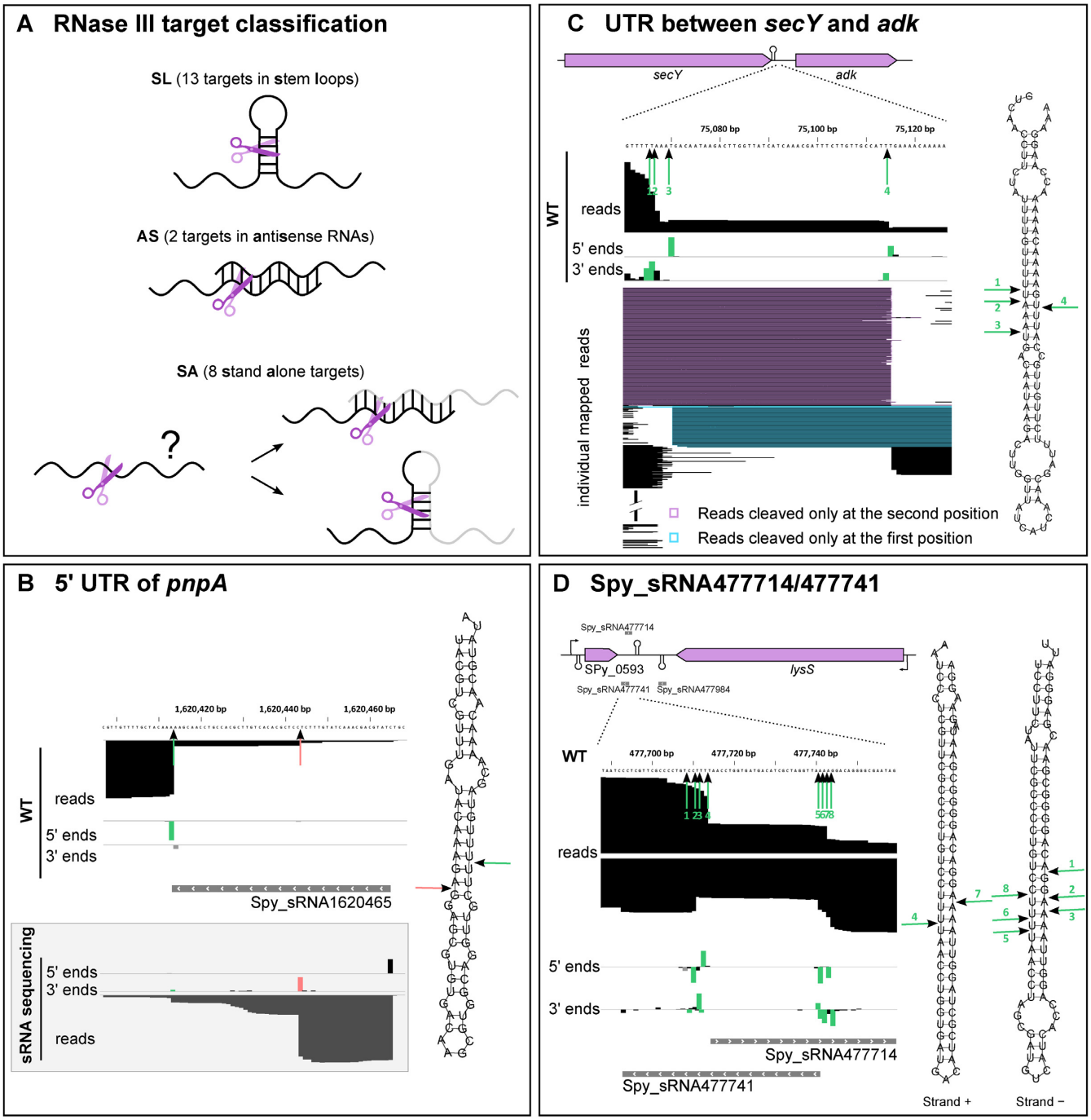

Figure 3. RNase III SCPs. (**A**) RNase III SCPs were classified in three groups. SL: cleavages on both sides of a stem loop forming a 2 nt 3′ overhang; AS: cleavages on both duplex strands forming a 2 nt 3′ overhang; SA: stand-alone positions, where the opposite cleavage position could not be retrieved with the ISCP method or previous sRNA sequencing data. We hypothesized that an RNA with SA SCP could base pair with other mRNAs or sRNAs and be co-processed by RNase III but it is also possible that the second cleavage position in the same transcript could not be retrieved here. (**B**–**D**) RNA coverage visualized using the Integrative Genome Viewer (IGV) (see also Supplementary Figure S4). The genomic coordinates are indicated. 'reads': total read coverage; '5′ ends' and '3′ ends': number of reads starting or ending respectively at each nucleotide (black bars). The positions detected by the ISCP method are shown with green bars and green arrows both on the genomic sequence and RNA folding. The SCPs are numbered to facilitate the visualization. Foldings were generated using RNAfold. Grey rectangles: sRNA genes. UTR: untranslated region. (**B**) Coverage of the 5′ UTR of *pnpA*. The position visualized in the sRNA sequencing data is shown with a pink bar and a pink arrow both on the genomic sequence and RNA folding. (**C**) Coverage of the UTR located between *secY* and *adk*. The individual mapped reads are shown. (**D**) Coverage of Spy_sRNA477714/Spy_sRNA477741 region. Foldings of the RNA from the positive and negative strands are shown.

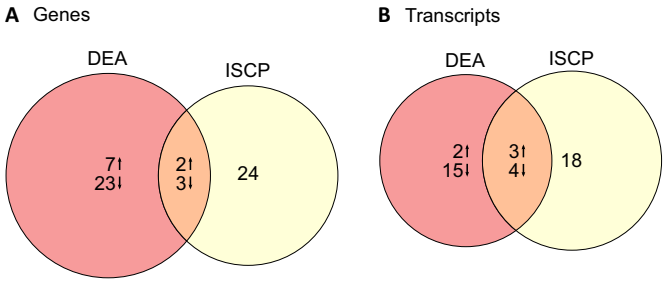

Figure 4. Venn diagram of differential gene expression in the RNase III deletion strain compared to the WT. Comparison of RNase III-regulated (**A**) genes and (**B**) transcripts identified by two approaches: differential expression analysis (DEA) and identification of specific cleavage positions (ISCP). The numbers of affected transcripts are indicated with an up arrow for upregulated genes and a down arrow for downregulated genes.

however, the significance of this process remains unclear since the full-length transcript abundances were not affected.

RNase III has a limited impact on gene expression regulation

The differential expression between Δ*rnc* and WT was computed using DEseq2 (see Material and Methods) with a number of uniquely mapped reads ranging from 4.2 to 7.3 million per sample (Supplementary Table S2A; Supplementary Figure S5). The ISCP and DEA results were compared at the gene level (Figure 4A) or at the transcript level, (i.e. genes expressed from the same operon) (Figure 4B). Only 24 transcripts were differentially expressed in the *rnc* deletion strain, with 19 being downregulated (Figure 4B; Supplementary Table S6). RNase III processing events were associated with a change of RNA abundance in Δ*rnc* compared to the WT for seven transcripts and, for four of them, RNA abundance was downregulated (Figure 4B; Supplementary Table S3).

Given the small effect of RNase III on gene expression, expected indirect effects on other genes would be relatively minor. We listed the RNA-binding proteins that were affected by the absence of RNase III to explain possible indirect effects (Supplementary Table S7). For example, upregulation of the unspecific exoribonuclease PNPase mRNA and protein in Δ*rnc* did not seem to have a detectable effect on global RNA abundance, as very few differentially expressed transcripts were identified in Δ*rnc*. The upregulation of *rpoA*, encoding the RNA polymerase alpha subunit, could potentially lead to some downstream effects, as previously shown in *Bordetella pertussis*, where the alpha subunit diminishes the level of expression of two virulence factors by titrating out a common transcriptional activator (51,52). Few sRNAs were downregulated or cleaved by RNase III, however, most *S. pyogenes* sRNAs have not been characterized and we cannot conclude whether they would regulate downstream targets. RNase III SCPs were also retrieved in *birA* and *nrdR* transcripts (Supplementary Figure S2), encoding two transcriptional repressors of biotin and deoxyribonucleotide biosynthesis, respectively. This did not affect the RNA levels of the repressors themselves and their known downstream regulated genes (53).

DISCUSSION

Identification of specific cleavage positions

In contrast to recently published approaches that are based solely on 5′ end transcript analyses, the ISCP method developed in this study enables not only the identification of *in vivo* transcript starts (5′) but also of the transcript ends (3′) (17–21). In comparison to only corroborating SCPs identified using 5′ ends, the determination of SCPs by the analysis of both 5′ and 3′ ends of the processed transcripts provides additional cleavage positions. With this strategy, an SCP can be identified if at least one of the two produced fragments is stable while in previous approaches, the cleavage sites were detectable only when the downstream fragment was stable.

To increase the specificity of our method and allowing a better clearance of false positive detection, we compared two RNase III deleted strains (Δ*rnc* and Δ*rny_rnc*) with two reference strains (WT and Δ*rny*). Furthermore, the use of two different libraries for both the reference and RNase deleted strains, *Pp* and *P*, allowed a better discrimination between 5′ primary and processed transcripts and ensured the reproducibility of our data. Although SCPs should be direct RNase targets, the deletion of an RNase could in some case affect the expression of other RNases, thus the detected SCPs could possibly be indirect targets. Here, RNase III deletion only leads to the upregulation of PNPase, thus we expect the SCPs detected being direct RNase III targets.

This technique still presents some limitations. To detect a cleavage, at least one of the two fragments should be stable. Furthermore, our approach is possible only when the transcript is expressed in all analyzed strains. As we use a second RNase deletion strain as a negative control (RNase Y for the study of RNase III), we cannot identify RNase III SCPs that are dependent of this second RNase. A way to circumvent this while keeping a second reference strain would be to use another mutant strain.

In vivo RNase III nicking activity

RNase III can generate a single nick or a ds break, depending on the ds RNA local structure (54–56). These two cleavages are possible because the two independent catalytic sites of the RNase III homodimeric enzyme are positioned in each protomer (57). Here, we observed *in vivo* that an important fraction of RNA molecules were nicked at one or the other position. We show as an example that the reads from *secY* and *adk* mRNAs were mostly originating from RNA molecules only cleaved at one position or the other by RNase III (Figure 3C). The very small RNAs (such as the ones generated by RNase III cleavage in a stem loop) were not sequenced here, therefore, we could not assess the percentage of ds cleaved transcripts compared to transcripts cleaved only at one position. Primer extension analyses performed on SPy_1227-*bmpA*, SPy_0129, *nrdR* and the 23S methyl regulatory element also shows presence of the two 5′ ends generated by RNase III, indicating that the first position can be cleaved independently of the second position (Supplementary Figure S2). Overall, the ISCP method highlights that RNase III nicking in duplex regions might be

a general phenomenon and this effect on RNA regulation should be further investigated. For example, it was shown that the single cleavage in a ds region of bacteriophage T7 early transcripts by RNase III generates a structure that stabilizes the transcripts (58), but that RNase III double cleavage in bacteriophage λ *int* transcript accelerates its degradation (50).

RNase III and antisense RNAs

Genomic regions with transcription on both strands (asRNAs) accounted for only 6.08% of the transcriptome in the WT. In Δ*rnc*, antisense transcription accounted for 8.11% of the transcriptome. These data suggest that RNase III has little impact on genome-wide asRNA regulation in *S. pyogenes*, as it was described previously in *B. subtilis* (59), but in contrast to what was shown in *S. aureus* (9). Indeed, Lasa *et al.* (9) showed that 49.2% of the WT transcriptome and 74.2% of the Δ*rnc* transcriptome of *S. aureus* were mapped on both strands and described a global processing by RNase III in these antisense regions.

sRNAs affected by RNase III

RNase III can degrade sRNA:mRNA targets, but can also lead to transcript stabilization (60,61). This enzyme is also important for sRNA maturation, as demonstrated in *S. pyogenes* where it is involved in pre-CRISPR RNA:tracrRNA biogenesis (27). In *S. pyogenes*, a recent study from our laboratory described two putative regulatory RNA elements (Lacto-*rpoB* and 23S methyl), two sRNAs and two asRNAs affected by RNase III (28). In the present study, we notice that even though RNase III preferentially cleaves in UTRs, the influence of this RNase on sRNAs is limited (Supplementary Tables S3 and S6). However the method used here does not allow the sequencing of small transcripts, thus resulting in an underestimation of the impact of RNase III on sRNAs.

Role of RNase III in *S. pyogenes*

Although we showed that RNase III is involved in rRNA processing of *S. pyogenes* using the ISCP method, we concluded that this RNase is not required for rRNA maturation. Indeed, 16S, 23S and 5S rRNA mature forms were also detected in absence of RNase III and Δ*rnc* did not present any growth defects compared to the WT (Supplementary Figure S6). It was previously described in *E. coli* that the 16S rRNA is matured in a strain lacking RNase III (62). The absence of Sub-SCPs around the *S. pyogenes* 16S rRNA in Δ*rnc* indicates that maturation of the 16S RNA can be achieved independently of RNase III processing. The mature 23S rRNA was also present in absence of RNase III, as described for the 5′ maturation of 23S rRNA in *B. subtilis* (63) but opposite to what has been described for *E. coli* (62). We observed Sub-SCPs in absence of RNase III but we do not know whether they are alternative processing steps required for 23S rRNA maturation by the ds specific endoribonuclease Mini-III (63) or whether they occur after the maturation step, in order to degrade precursor fragments generated by Mini-III.

RNase III deletion affected the abundance of 1.6% of *S. pyogenes* transcriptome (considering only gene expression superior or equal to 10 RPKM) under standard growth conditions. The limited number of differentially expressed targets highlights, in the conditions tested, a restricted role of RNase III in *S. pyogenes* compared to what was previously described for *E. coli* and *B. subtilis*, where 12 and 11% of the genome was differentially expressed, respectively (59,64). There could be conditions where RNase III deletion could affect the abundance of more genes, as demonstrated in *E. coli* where it is required for growth in anaerobic/poor carbon conditions (65).

Here, the gene downregulation accounted for 80% of the differentially expressed transcripts in Δ*rnc* compared to the WT (Figure 4). Downregulation of gene expression in RNase deleted strains is often regarded as indirect effects of the deletion, as the removal of an RNA degradative enzyme is expected to stabilize the target transcripts. However, we could not assess whether other RNA-binding proteins are responsible for some possible indirect effects. Additionally, processing by RNase III does not always lead to RNA degradation but can also stabilize/mature transcripts (23). In this study, 5 out of the 8 cleaved and differentially expressed transcripts were downregulated, indicating that RNase III could have a stabilizing effect on transcripts.

Among RNase III primary targets, only one third was differentially expressed (Figure 4). This suggests that RNase III does not have a strong effect on RNA abundance. However, this enzyme can control gene expression without modifying RNA levels. For instance, RNase III can directly affect translation without affecting RNA levels or without cleaving the target (23). Future studies on RNA stability and proteomic complementary to this study may reveal additional direct targets of RNase III.

In conclusion, we successfully developed and validated our ISCP method and identified genome-wide direct targets of RNase III in the human pathogen *S. pyogenes*. The 3′ end detection allowed the identification of targets that would not have been detected with the 5′ end. This method was also suitable to observe RNase III single cleavage *in vivo*. The effects of these cleavages on *S. pyogenes* transcriptome were further investigated using DEA. We showed that surprisingly in *S. pyogenes*, RNase III had little influence on RNA steady state levels.

ACCESSION NUMBERS

RNA sequencing data have been deposited at NCBI under bioproject accession numbers SRP089790 and SRP089789.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

The authors thank Robert Geffers, Sabin Bhuju and Mikael Jarek from the Next Generation Sequencing facility at the Helmholtz Centre for Infection Research (Braunschweig, Germany) for conducting RNA sequencing and demultiplexing of the RNA sequencing data. The authors grate-

fully acknowledge support by Marc Hoeppner and Jeannette Tångrot from the Bioinformatics Infrastructure for Life Sciences (BILS, Uppsala, Sweden) and Mikael Huss from the Bioinformatics Long-term Support, SciLifeLab (WABI, Stockholm and Uppsala Sweden) with RNA sequencing data analysis. The computations were performed using resources provided by the Swedish National Infrastructure for Computing (SNIC) through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project b20133265. The authors acknowledge Thibaud Renault for assistance with bioinformatics. The authors are grateful to Tracy Nissan, Susanne Huch and the members of Charpentier Department for discussions and critical reading of the manuscript.

FUNDING

Alexander von Humboldt Foundation [AvH Professorhip]; Helmholtz Association; Max Planck Society (Max Planck Foundation); Göran Gustafsson Foundation [Göran Gustafsson Prize]; Science for Life Laboratory Wallenberg Advanced Bioinformatics Support [WABI, b2013265]; Carl Kempe Foundation [K2010-57X-21436-01-3]; Umeå University [Dnr: 223-2386-10; Dnr: 223-2989-10-I]; Swedish Research Council [K2013-57X-21436-04-3]. Funding for open access charge: Max Planck Society.
Conflict of interest statement. None declared.

REFERENCES

1. Cho,B.K., Zengler,K., Qiu,Y., Park,Y.S., Knight,E.M., Barrett,C.L., Gao,Y. and Palsson,B.O. (2009) The transcription unit architecture of the *Escherichia coli* genome. *Nat. Biotechnol.*, **27**, 1043–1049.
2. Mendoza-Vargas,A., Olvera,L., Olvera,M., Grande,R., Vega-Alvarado,L., Taboada,B., Jimenez-Jacinto,V., Salgado,H., Juarez,K., Contreras-Moreira,B. *et al.* (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PloS One*, **4**, e7526.
3. Sharma,C.M., Hoffmann,S., Darfeuille,F., Reignier,J., Findeiss,S., Sittka,A., Chabas,S., Reiche,K., Hackermuller,J., Reinhardt,R. *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, **464**, 250–255.
4. Singh,N. and Wade,J.T. (2014) Identification of regulatory RNA in bacterial genomes by genome-scale mapping of transcription start sites. *Methods Mol. Biol.*, **1103**, 1–10.
5. Conway,T., Creecy,J.P., Maddox,S.M., Grissom,J.E., Conkle,T.L., Shadid,T.M., Teramoto,J., San Miguel,P., Shimada,T., Ishihama,A. *et al.* (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio*, **5**, e01442–01414.
6. Creecy,J.P. and Conway,T. (2015) Quantitative bacterial transcriptomics with RNA-seq. *Curr. Opin. Microbiol.*, **23**, 133–140.
7. Dar,D., Shamir,M., Mellin,J.R., Koutero,M., Stern-Ginossar,N., Cossart,P. and Sorek,R. (2016) Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science*, **352**, aad9822.
8. Kennell,D. (2002) Processing endoribonucleases and mRNA degradation in bacteria. *J. Bacteriol.*, **184**, 4645–4657; discussion 4665.
9. Lasa,I., Toledo-Arana,A., Dobin,A., Villanueva,M., de los Mozos,I.R., Vergara-Irigaray,M., Segura,V., Fagegaltier,D., Penades,J.R., Valle,J. *et al.* (2011) Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 20172–20177.
10. Liu,B., Deikus,G., Bree,A., Durand,S., Kearns,D.B. and Bechhofer,D.H. (2014) Global analysis of mRNA decay intermediates in *Bacillus subtilis* wild-type and polynucleotide phosphorylase-deletion strains. *Mol. Microbiol.*, **94**, 41–55.
11. Chen,R., Weng,Y., Zhu,F., Jin,Y., Liu,C., Pan,X., Xia,B., Cheng,Z., Jin,S. and Wu,W. (2016) Polynucleotide phosphorylase regulates multiple virulence factors and the stabilities of small RNAs RsmY/Z in *Pseudomonas aeruginosa*. *Front. Microbiol.*, **7**, 247, 10.3389/fmicb.2016.00247.
12. Redko,Y., Galtier,E., Arnion,H., Darfeuille,F., Sismeiro,O., Coppee,J.Y., Medigue,C., Weiman,M., Cruveiller,S. and De Reuse,H. (2016) RNase J depletion leads to massive changes in mRNA abundance in *Helicobacter pylori*. *RNA Biol.*, **13**, 243–253.
13. Pobre,V. and Arraiano,C.M. (2015) Next generation sequencing analysis reveals that the ribonucleases RNase II, RNase R and PNPase affect bacterial motility and biofilm formation in *E. coli*. *BMC Genomics*, **16**, 72, 10.1186/s12864-015-1237-6.
14. Leskinen,K., Varjosalo,M. and Skurnik,M. (2014) The absence of YbeY ribonuclease compromises the growth and enhances the virulence plasmid gene expression of *Yersinia enterocolitica* O:3. *Microbiology*, **161**, 285-299.
15. Lioliou,E., Sharma,C.M., Caldelari,I., Helfer,A.C., Fechter,P., Vandenesch,F., Vogel,J. and Romby,P. (2012) Global regulatory functions of the *Staphylococcus aureus* endoribonuclease III in gene expression. *PLoS Genet.*, **8**, e1002782.
16. Lybecker,M., Zimmermann,B., Bilusic,I., Tukhtubaeva,N. and Schroeder,R. (2014) The double-stranded transcriptome of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 3134–3139.
17. Clarke,J.E., Kime,L., Romero,A.D. and McDowall,K.J. (2014) Direct entry by RNase E is a major pathway for the degradation and processing of RNA in *Escherichia coli*. *Nucleic Acids Res.*, **42**, 11733–11751.
18. Linder,P., Lemeille,S. and Redder,P. (2014) Transcriptome-wide analyses of 5'-ends in RNase J mutants of a gram-positive pathogen reveal a role in RNA maturation, regulation and degradation. *PLoS Genet.*, **10**, e1004207.
19. Schifano,J.M., Vvedenskaya,I.O., Knoblauch,J.G., Ouyang,M., Nickels,B.E. and Woychik,N.A. (2014) An RNA-seq method for defining endoribonuclease cleavage specificity identifies dual rRNA substrates for toxin MazF-*mt3*. *Nat. Commun.*, **5**, 3538, 10.1038/ncomms4538.
20. Khemici,V., Prados,J., Linder,P. and Redder,P. (2015) Decay-initiating endoribonucleolytic cleavage by RNase Y is kept under tight control via sequence preference and sub-cellular localisation. *PLoS Genet.*, **11**, e1005577.
21. DiChiara,J.M., Liu,B., Figaro,S., Condon,C. and Bechhofer,D.H. (2016) Mapping of internal monophosphate 5' ends of *Bacillus subtilis* messenger RNAs and ribosomal RNAs in wild-type and ribonuclease-mutant strains. *Nucleic Acids Res.*, **44**, 3373–3389.
22. Cooper,D.A., Jha,B.K., Silverman,R.H., Hesselberth,J.R. and Barton,D.J. (2014) Ribonuclease L and metal-ion-independent endoribonuclease cleavage sites in host and viral RNAs. *Nucleic Acids Res.*, **42**, 5202–5216.
23. Court,D.L., Gan,J., Liang,Y.H., Shaw,G.X., Tropea,J.E., Costantino,N., Waugh,D.S. and Ji,X. (2013) RNase III: Genetics and function; structure and mechanism. *Annu. Rev. Genet.*, **47**, 405–431.
24. Deutscher,M.P. (2009) Maturation and degradation of ribosomal RNA in bacteria. *Prog. Mol. Biol. Transl. Sci.*, **85**, 369–391.
25. de los Mozos,I.R., Vergara-Irigaray,M., Segura,V., Villanueva,M., Bitarte,N., Saramago,M., Domingues,S., Arraiano,C.M., Fechter,P., Romby,P. *et al.* (2013) Base pairing interaction between 5′- and 3′-UTRs controls *icaR* mRNA translation in *Staphylococcus aureus*. *PLoS Genet.*, **9**, e1004001 .
26. Lim,B., Sim,M., Lee,H., Hyun,S., Lee,Y., Hahn,Y., Shin,E. and Lee,K. (2015) Regulation of *Escherichia coli* RNase III activity. *J. Microbiol.*, **53**, 487–494.
27. Deltcheva,E., Chylinski,K., Sharma,C.M., Gonzales,K., Chao,Y., Pirzada,Z.A., Eckert,M.R., Vogel,J. and Charpentier,E. (2011) CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature*, **471**, 602–607.
28. Le Rhun,A., Beer,Y.Y., Reimegard,J., Chylinski,K. and Charpentier,E. (2015) RNA sequencing uncovers antisense RNAs and novel small RNAs in *Streptococcus pyogenes*. *RNA Biol.*, **13**, 177–195.
29. Sitkiewicz,I. and Musser,J.M. (2006) Expression microarray and mouse virulence analysis of four conserved two-component gene regulatory systems in group a streptococcus. *Infect. Immun.*, **74**, 1339–1351.

30. Caparon,M.G. and Scott,J.R. (1991) Genetic manipulation of pathogenic streptococci. *Methods Enzymol.*, **204**, 556–586.
31. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular cloning : a laboratory manual*, New York. Cold Spring Harbor Laboratory Press.
32. Lambert,J.M., Bongers,R.S. and Kleerebezem,M. (2007) Cre-lox-based system for multiple gene deletions and selectable-marker removal in *Lactobacillus plantarum*. *Appl. Environ. Microbiol.*, **73**, 1126–1135.
33. Dotsch,A., Eckweiler,D., Schniederjans,M., Zimmermann,A., Jensen,V., Scharfe,M., Geffers,R. and Haussler,S. (2012) The *Pseudomonas aeruginosa* transcriptome in planktonic cultures and static biofilms using RNA sequencing. *PloS One*, **7**, e31092.
34. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. **17**, 10–12.
35. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
36. Robinson,J.T., Thorvaldsdottir,H., Winckler,W., Guttman,M., Lander,E.S., Getz,G. and Mesirov,J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.*, **29**, 24–26.
37. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
38. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550–570.
39. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
40. Kingsford,C.L., Ayanbule,K. and Salzberg,S.L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.*, **8**, R22.
41. Seemann,T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**, 2068–2069.
42. Lorenz,R., Bernhart,S.H., Honer Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26–39.
43. Busch,A., Richter,A.S. and Backofen,R. (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, **24**, 2849–2856.
44. Young,R.A. and Steitz,J.A. (1978) Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **75**, 3593–3597.
45. Bram,R.J., Young,R.A. and Steitz,J.A. (1980) The ribonuclease III site flanking 23S sequences in the 30S ribosomal precursor RNA of *E. coli*. *Cell*, **19**, 393–401.
46. Del Campo,C., Bartholomaus,A., Fedyunin,I. and Ignatova,Z. (2015) Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genet.*, **11**, e1005613.
47. Robert-Lemeur,M. and Portier,C. (1992) *Escherichia Coli* polynucleotide phosphorylase expression is autoregulated through an Rnase-III-dependent mechanism. *EMBO J.*, **11**, 2633–2641.
48. Jarrige,A.C., Mathy,N. and Portier,C. (2001) PNPase autocontrols its expression by degrading a double-stranded structure in the *pnp* mRNA leader. *EMBO J.*, **20**, 6845–6855.
49. Plunkett,G. 3rd and Echols,H. (1989) Retroregulation of the bacteriophage lambda *int* gene: limited secondary degradation of the RNase III-processed transcript. *J. Bacteriol.*, **171**, 588–592.
50. Schmeissner,U., McKenney,K., Rosenberg,M. and Court,D. (1984) Removal of a terminator structure by RNA processing regulates *int* gene expression. *J. Mol. Biol.*, **176**, 39–53.
51. Carbonetti,N.H., Fuchs,T.M., Patamawenu,A.A., Irish,T.J., Deppisch,H. and Gross,R. (1994) Effect of mutations causing overexpression of RNA polymerase alpha subunit on regulation of virulence factors in *Bordetella pertussis*. *J. Bacteriol.*, **176**, 7267–7273.
52. Carbonetti,N.H., Romashko,A. and Irish,T.J. (2000) Overexpression of the RNA polymerase alpha subunit reduces transcription of Bvg-activated virulence genes in *Bordetella pertussis*. *J. Bacteriol.*, **182**, 529–531.
53. Novichkov,P.S., Kazakov,A.E., Ravcheev,D.A., Leyn,S.A., Kovaleva,G.Y., Sutormin,R.A., Kazanov,M.D., Riehl,W., Arkin,A.P., Dubchak,I. *et al.* (2013) RegPrecise 3.0–a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics*, **14**, 745–756.
54. Calin-Jageman,I. and Nicholson,A.W. (2003) RNA structure-dependent uncoupling of substrate recognition and cleavage by *Escherichia coli* ribonuclease III. *Nucleic Acids Res.*, **31**, 2381–2392.
55. Pertzev,A.V. and Nicholson,A.W. (2006) Characterization of RNA sequence determinants and antideterminants of processing reactivity for a minimal substrate of *Escherichia coli* ribonuclease III. *Nucleic Acids Res.*, **34**, 3708–3721.
56. Gan,J., Shaw,G., Tropea,J.E., Waugh,D.S., Court,D.L. and Ji,X. (2008) A stepwise model for double-stranded RNA processing by ribonuclease III. *Mol. Microbiol.*, **67**, 143–154.
57. Nicholson,A.W. (2014) Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdiscip. Rev. RNA*, **5**, 31–48.
58. Gegenheimer,P. and Apirion,D. (1981) Processing of procaryotic ribonucleic acid. *Microbiol. Rev.*, **45**, 502–541.
59. Durand,S., Gilet,L., Bessieres,P., Nicolas,P. and Condon,C. (2012) Three essential ribonucleases-RNase Y, J1, and III-control the abundance of a majority of *Bacillus subtilis* mRNAs. *PLoS Genet.*, **8**, e1002520.
60. Opdyke,J.A., Fozo,E.M., Hemm,M.R. and Storz,G. (2011) RNase III participates in GadY-dependent cleavage of the *gadX-gadW* mRNA. *J. Mol. Biol.*, **406**, 29–43.
61. Lalaouna,D., Simoneau-Roy,M., Lafontaine,D. and Masse,E. (2013) Regulatory RNAs and target mRNA decay in prokaryotes. *Biochimica Biophys. Acta*, **1829**, 742–747.
62. King,T.C. and Schlessinger,D. (1983) S1 nuclease mapping analysis of ribosomal RNA processing in wild type and processing deficient *Escherichia coli*. *J. Biol. Chem.*, **258**, 12034–12042.
63. Redko,Y., Bechhofer,D.H. and Condon,C. (2008) Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in *B. subtilis*. *Mol. Microbiol.*, **68**, 1096–1106.
64. Stead,M.B., Marshburn,S., Mohanty,B.K., Mitra,J., Pena Castillo,L., Ray,D., van Bakel,H., Hughes,T.R. and Kushner,S.R. (2011) Analysis of *Escherichia coli* RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic Acids Res.*, **39**, 3188–3203.
65. Aristarkhov,A., Mikulskis,A., Belasco,J.G. and Lin,E.C. (1996) Translation of the *adhE* transcript to produce ethanol dehydrogenase requires RNase III cleavage in *Escherichia coli*. *J. Bacteriol.*, **178**, 4327–4332.
66. Regnier,P. and Portier,C. (1986) Initiation, attenuation and RNase III processing of transcripts from the *Escherichia coli* operon encoding ribosomal protein S15 and polynucleotide phosphorylase. *J. Mol. Biol.*, **187**, 23–32.