Efficient depletion of ribosomal RNA for RNA sequencing in planarians

Iana V. Kim[1,*], Eric J. Ross[2,3], Sascha Dietrich[4], Kristina Döring[4], Alejandro Sánchez Alvarado[2,3] and Claus-D. Kuhn[1,*]

[1] Gene regulation by Non-coding RNA, Elite Network of Bavaria and University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany

[2] Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA

[3] Howard Hughes Medical Institute, Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City MO, 64110, USA

[4] Core Unit Systems Medicine, Institute for Molecular Infection Biology, University of Würzburg, Josef-Schneider-Str. 2, 97080, Würzburg, Germany

* Corresponding authors: iana.kim@uni-bayreuth.de, claus.kuhn@uni-bayreuth.de

Abstract (350 words)

Background

The astounding regenerative abilities of planarian flatworms prompt a steadily growing interest in examining their molecular foundation. Planarian regeneration was found to require hundreds of genes and is hence a complex process. Thus, RNA interference followed by transcriptome-wide gene expression analysis by RNA-seq is a popular technique to study the impact of any particular planarian gene on regeneration. Typically, the removal of ribosomal RNA (rRNA) is the first step of all RNA-Seq library preparation protocols. To date, rRNA removal in planarians was

primarily achieved by the enrichment of polyadenylated (poly(A)) transcripts. However, to better reflect transcriptome dynamics and to cover also non-poly(A) transcripts, a procedure for the targeted removal of rRNA in planarians is needed.

Results

In this study, we describe a workflow for the efficient depletion of rRNA in the planarian model species *S. mediterranea*. Our protocol is based on subtractive hybridization using organism-specific probes. Importantly, the designed probes also deplete rRNA of other freshwater triclad families, a fact that considerably broadens the applicability of our protocol. We tested our approach on total RNA isolated stem cells (termed neoblasts) of *S. mediterranea* and compared ribodepleted libraries with publicly available poly(A)-enriched ones. Overall, mRNA levels after ribodepletion were consisted with poly(A) libraries. However, ribodepleted libraries revealed higher transcript levels for transposable elements and histone mRNAs that remained underrepresented in poly(A) libraries. As neoblasts experience high transposon activity this suggests that ribodepleted libraries better reflect the transcriptional dynamics of planarian stem cells. Furthermore, the presented ribodepletion procedure was successfully expanded to the removal of ribosomal RNA from the gram-negative bacterium *Salmonella typhimurium*.

Conclusions

The ribodepletion protocol presented here ensures the efficient rRNA removal from low input total planarian RNA, which can be further processed for RNA-Seq applications. Resulting libraries contain less than 2% rRNA. Moreover, for a cost-effective and efficient removal of rRNA prior to

sequencing applications our procedure might be adapted to any prokaryotic or eukaryotic species of choice.

Keywords

Planarians, Schmidtea mediterranea, ribosomal RNA removal, rRNA depletion, RNA sequencing

Background

Freshwater planarians of the species *Schmidtea mediterranea* are well known for their extraordinary ability to regenerate. This ability is supported by the presence of a large population of adult pluripotent stem cells, termed neoblasts [1]. Neoblasts are capable of producing all planarian cell types [2]. Moreover, they preserve their potency over the whole lifespan of the animal, which seems to be infinite [3]. Therefore, planarians embody an excellent model to study regeneration, aging and stem cell-based diseases. The phylum Platyhelminthes, to which *S. mediterranea* belongs to, includes multiple other members showing varying degrees of regenerative abilities. While some freshwater species (e.g. *Dugesia japonica* and *Polycelis nigra*) are capable to restore their body from any tiny piece [4, 5], others (*Procotyla fluviatilis*) have limited anterior regeneration abilities [6]. Altogether, the ability to regenerate is not solely based on the presence of pluripotent stem cells, but represents a complex interplay between different signaling pathways. The underlying changes in gene expression therefore need to be studied using transcriptome-wide techniques like RNA sequencing.

For any informative RNA-seq library preparation, ribosomal RNA, comprising >80% of total RNA, has to be removed. To achieve this goal two strategies can be pursued: either polyadenylated (poly(A)) RNA transcripts are enriched or rRNA is removed. Both approaches have advantages and limitations. On the one hand, the enrichment of poly(A) transcripts ensures better coverage of coding genes compared to ribodepleted samples, when sequenced to similar depth [7]. However, this advantage is outweighed by the loss of transcripts lacking poly(A) tails, which include preprocessed RNAs, a large share of all non-coding RNAs, such as enhancer RNAs and other long non-coding RNAs. In addition, long terminal repeat (LTR) retrotransposons and various intermediates of endonucleotic RNA degradation are lost during poly(A) selection [8–13]. Furthermore, most prokaryotic RNAs lack poly(A) tails, making rRNA depletion crucial for the study of bacterial transcriptome [14].

Here, we describe a probe-based subtractive hybridization workflow for rRNA depletion that efficiently removes planarian rRNA from total RNA. The protocol can be applied to input as low as 100 ng total RNA, which corresponds to 100,000 FACS-sorted planarian stem cells (X1 population) [15, 16]. Moreover, the DNA probes developed for *S. mediterranea* were successfully used for the removal of ribosomal RNA in related planarian species of the order Tricladida. The rRNA removal workflow presented here is also easily adapted to other organisms, as demonstrated by the removal of rRNA from *Salmonella typhimurium* total RNA using organism-specific probes.

Results

Development of an efficient rRNA depletion protocol for planarians

To deplete ribosomal RNA from planarian total RNA, we chose to develop a protocol based on the hybridization of rRNA-specific biotinylated DNA probes to ribosomal RNA and the capture of the resulting biotinylated rRNA-DNA hybrids by use of streptavidin-coated magnetic beads (Figure 1A). To that end, we synthesized a pool of 88 3'-biotinylated 40-nt long DNA oligonucleotide probes (siTOOLs Biotech, Martinsried, Germany). We chose probes with a length of 40 nucleotides since their melting temperature in RNA-DNA hybrids was shown to be 80±6.4 °C in the presence of 500 mM sodium ions [17]. This would allow probe annealing at 68°C in agreement with general probe hybridization temperatures [18]. The probes were devised in antisense orientation to the following planarian rRNA species: 28S, 18S type I and type II, 16S, 12S, 5S, 5.8S, internal transcribed spacer (ITS) 1 and ITS 2 (Supplementary Table1).

To assess RNA quality and the efficiency of rRNA removal we used capillary electrophoresis (Fragment Analyzer, Agilent). The separation profile of total planarian RNA only shows a single rRNA peak at about 1500 nucleotides (nts) (Figure 1B). This single rRNA peak is the result of the 28S rRNA being processed into two fragments that co-migrate with the peak for 18S rRNA [19]. Planarian 28S rRNA processing usually entails the removal of a short sequence located in the D7a expansion segment of 28S rRNA. The length of the removed fragment thereby varies from 4 nts to 350 nts between species (e.g. for *Dugesia japonica* 42 nts are removed) [19]. Intriguingly, a similar rRNA maturation process was observed in particular protostomes, in insects such as *D. melanogaster* and in other platyhelminthes [19–21]. In addition to this 28S rRNA maturation phenomenon, *S. mediterranea*

possesses two 18S rDNA copies that differ in about 8% or their sequence. However, only 18S rRNA type I was reported to be functional and predominantly transcribed [22, 23].

As a first step during rRNA removal all 88 DNA probes were annealed to total planarian RNA. Since RNA molecules are negatively charged, the presence of cations facilitates the annealing of probes to RNA by reducing the repulsion of phosphate groups [24, 25]. Although Mg^{2+} ions are most effective in stabilizing the tertiary structure of RNA and in promoting the formation of DNA-RNA hybrids, they are also cofactors for multiple RNases [26] and hence should not be included during ribodepletion. Therefore, we tested several hybridization buffers with varying concentrations of sodium ions (Figure 1C). In the absence of sodium ions we only accomplished the incomplete removal of rRNA. However, sodium concentrations >250 mM in the hybridization buffer led to the complete depletion of rRNA from planarian total RNA (Figure 1C, 1D). Thus, optimal rRNA removal requires the presence of >250 mM NaCl in the hybridization buffer. As we obtained the most consistent results in the presence of 500 mM NaCl, we decided to utilize this salt concentration in our procedure (Figure 1D).

Detailed rRNA depletion workflow

Required buffers:

Hybridization buffer (20 mM Tris-HCl (pH 8.0), 1 M NaCl, 2 mM EDTA)

Solution A (100 mM NaOH, 50 mM NaCl, DEPC-treated)

Solution B (100 mM NaCl, DEPC-treated)

2xB&W (Binding&Washing) buffer (10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2 M NaCl)

Dilution buffer (10 mM Tris-HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA)

Protocol:

1. RNA input

The following protocol efficiently depletes ribosomal RNA from 100 up to 1.5 µg of total RNA (Figure 1E). The procedure can be scaled up for higher RNA input.

2. Hybridization of biotinylated DNA oligonucleotides (40-mers) to ribosomal RNA.

a) For oligonucleotide annealing the following reaction is set up:

10 µl hybridization buffer

10 µl RNA input (1 µg)

1 µl 100 µM biotinylated DNA 40-mers probes

b) Gently mix the solution by pipetting and incubate at 68 °C for 10 min.

c) Immediately transfer the tubes to 37 °C for 30 min.

3. Prepare Dynabeads MyOne Streptavidin C1 (Invitrogen) according to the manufacturer's instruction as follows:

a) For each sample use 120 µl (10 µg/µl) of beads slurry.

b) Wash the beads twice with an equal volume (or at least 1 ml) of Solution A. Add Solution A and incubate the mixture for 2 min. Then, place the tube on a magnet for 1 min and discard the supernatant.

c) Wash the beads once in Solution B. Split the washed beads into two separate tubes for two rounds of subtractive rRNA depletion (Round1 and Round2). Place the beads on a magnet for 1 min and discard Solution B.

d) Resuspend the Round1 beads in 2xB&W buffer to a final concentration of 5 µg/µl (twice the original volume). The Round1 beads will be used during the first round of rRNA depletion. For the second round of depletion, resuspend the Round2 beads to a final concentration of 5 µg/µl in

1xB&W buffer. The Round2 beads will be used in a second depletion step. Keep it at 37 °C until required.

4. Capture of DNA-RNA hybrids using magnetic beads (step 2).

a) Briefly spin down the tubes containing total RNA and probes. Add the following:

100 µl dilution buffer

120 µl washed magnetic beads (5 µg/µl) in 2xB&W (Round1).

Resuspend by pipetting up and down ten times. The final concentration of NaCl during this step is 1 M. Incubate the solution at 37 °C for 15 min. Gently mix the sample occasionally by tapping.

b) Place on magnet for 2 min. Carefully remove the supernatant and add the additional 120 µl of washed magnetic beads in 1xB&W (Round2). Incubate the mixture at 37°C for 15 min with occasional gentle tapping.

c) Place on magnet for 2 min. Carefully transfer the supernatant into a new tube and place it on magnet for another 1 min to remove all traces of magnetic beads from the sample.

d) Transfer the supernatant into a fresh tube.

5. Use the RNA Clean & Concentrator-5 kit (Zymo Research) to concentrate the ribodepleted samples, to carry out size selection and to digest any remaining DNA using DNase I treatment as described [27].

Ribosomal RNA depletion in planarian species related to *S. mediterranea*

Ribosomal DNA genes are among the most conserved sequences in the kingdom of life. They are present in all organisms and are widely used for the construction of phylogenetic trees [28]. The latter is possible because of the low rate of nucleotide substitutions in rRNA sequences (about 1 - 2% substitutions occur per 50 million years based on bacterial 16S rRNA) [29]. The divergence of 18S rRNA sequence between different families of freshwater planarians lays in the range of 6 – 8%, while

interspecies diversity does not exceed 4% [23]. Therefore, low rRNA divergence between taxa can be exploited for the design of universal probes for rRNA depletion in different organisms. To assess the specificity and universal applicability of our DNA probes, we depleted rRNA in flatworm species of the order Tricladida, all related to S. mediterranea (Figure 2A). Total RNA separation profiles were analyzed before and after rRNA depletion of six planarian species from three different families. Two of these, *Dugesia japonica* and *Cura pinguis* belong to the same family as *S. mediterranea*, the Dugesiidae family. In addition, we examined three species from the family Planariidae (*Planaria torva*, *Polycelis nigra* and *Polycelis tenuis*) and one species from the genus *Camerata* of Uteriporidae (subfamily Uteriporinae). For all tested species our DNA probes proved efficient for the complete removal of rRNA, which migrated close to 2000 nts on all electropherograms (Figure 2B). Thus, the probes presented here can be utilized for the removal of ribosomal RNA in a multitude of planarian species and may even be generally applicable to all studied planarian species.

Comparison of RNA-Seq libraries prepared by ribodepletion or poly(A) selection

To assess the efficiency of rRNA removal and the specificity of our DNA probes, we prepared and analyzed RNA-Seq libraries from ribodepleted planarian total RNA from *S. mediterranea*. Total RNA was extracted from 100,000 FACS-sorted planarian neoblasts, resulting in 70 – 100 ng of input RNA. RNA-Seq libraries were prepared and sequenced as described [27] following 15 cycles of PCR amplification. The subsequent analysis of sequenced libraries confirmed the efficient removal of rRNAs. Less than 2% of total sequenced reads constituted ribosomal RNA (Figure 3A). Next, we compared our rRNA-depleted libraries with three publicly available planarian poly(A) enriched RNA-Seq datasets (poly(A) libraries) [30–32]. In case publicly available libraries were sequenced in paired-end mode, we analyzed only the first read of every pair to minimize the technical variation between

libraries [33]. As shown in Figure 3A, the ribodepleted libraries contained significantly less rRNA compared to all poly(A) enriched ones. Interestingly, the major rRNA species that remained after poly(A) selection was mitochondrial 16S rRNA (Figure 3B). Although the planarian genome has a high A-T content (> 70%) [34], we could not attribute the overrepresentation of 16S rRNA in poly(A) libraries to a high frequency or longer stretches of A nucleotides as compared to other rRNA species (Figure 3C). Moreover, using publicly available planarian poly(A)-position profiling by sequencing (3P-Seq) libraries [35] which allow the identification of 3'-ends of polyadenylated RNAs, no polyadenylation sites were detected on 16S rRNA. Therefore, we speculate that upon folding of the 16S rRNA stretches of A nucleotides become exposed and facilitate the interaction with oligo-dT beads during transcript poly(A) selection.

We next assigned the analyzed datasets to the planarian genome. In ribodepleted libraries more than 13% of all mapped reads were assigned to intergenic regions, compared to 7% – 10.5% for poly(A)-enriched ones (Figure 3D). In addition, the percentage of unmapped reads was higher in ribodepleted libraries and constituted about 17.6%, which is on average 2.4% more than in poly(A) datasets. We speculate that for ribodepleted libraries the proportion of reads mapping to intergenic regions will increase in future once complete assemblies of the planarian genome are available. Currently, the planarian genome assembly consists of 481 scaffolds [34]. To detect gene expression variabilities between the analyzed libraries, we performed principal component analysis for the clustering of gene expression data. Although all poly(A) selected libraries were grouped closer together along the PC1 scale, all four analyzed datasets appeared as separated clusters. This indicates the presence of considerable variation even amongst different batches of poly(A) libraries (Figures 3E). One possible source of such variation might be the sequencing depth of the analyzed libraries, which varied considerably from 13 to 64 million of mapped reads (Figure 3F).

Next, to estimate the correlation between ribodepleted and poly(A) libraries, we calculated their Pearson correlation coefficients (Figure 3G). We found the highest Pearson correlation between ribodepleted libraries and polyA B2 samples (R=0.94, $p < 2.2e-16$) (Figure 3F). This could be due to their similar sequencing depth compared to the other polyA libraries. The transcripts whose abundance was most significantly affected by poly(A) selection were found to be histone mRNAs that are known to lack polyA tails (Figure 3G, 3H) [36]. Their expression level appeared to be 8 – 10 fold higher in our ribodepleted libraries. Moreover, in the ribodepleted libraries we detected significantly higher expression levels for transposable elements (Figure 3G, 3I). Out of 316 planarian transposable element families [37] 254 were on average upregulated 5.2, 3.5 and 4.0 fold as compared to polyA B1, polyA B2 and polyA B3 libraries, respectively (Figure 3I). Our ribodepleted libraries revealed that Burro elements, giant retroelements found in planarian genome [34], gypsy retrotransposons, hAT and Mariner/Tc1 DNA transposons are the most active transposable elements in planarian stem cells. Although some of transposable elements are polyadenylated, long-terminal repeat elements (LTRs) lack poly(A)-tails [38]. This renders their detection in poly(A)-enriched sample non-quantitative. Last, our rRNA depletion workflow also allows for the detection of transposable element degradation products that were generated by PIWI proteins guided by specific piRNAs, an abundant class of small RNAs in planarians [39, 40].

Non-specific depletion of coding transcripts in ribodepleted libraries

In using custom ribodepletion probes, our major concern was that the utilized probes would lead to unspecific co-depletion of planarian coding transcripts. To exclude this possibility, we first mapped our pool of 88 DNA probes in antisense orientation to the planarian transcriptome allowing up to 8 mismatches and gaps of up to 3 nts. This mapping strategy requires at least 75% of a DNA

probe to anneal to its RNA target. It resulted in only 11 planarian genes to be potentially recognized by 20 DNA probes from our oligonucleotide pool. Next, we carried out a differential expression analysis of these 11 potentially targeted transcripts between the ribodepleted libraries and poly(A)-selected ones. The analysis revealed that 9 out of 11 potential targets were downregulated at least 1-fold in at least two poly(A) experiments (Figure 4A). As the abundance of three transcripts (SMESG000014330.1 (rhodopsin-like orphan gpcr [41]), SMESG000068163.1 and SMESG000069530.1 (both without annotation)) was very low in all polyA libraries (<0.6 transcripts per million (TPM)) we did not consider these any further. However, we found the remaining six transcripts to be significantly downregulated in ribodepleted libraries. For three of these targeted genes (SMESG000067473.1, SMESG000021061.1 and SMESG000044545.1) the probes map in regions that display significant RNA-seq coverage (Figure 4B, Supplemental Figs. S1A, S1B). Therefore, their lower expression values in ribodepleted libraries is likely attributed to probe targeting. Intriguingly, for the remaining three targets (SMESG000066644.1, SMESG000043656.1 and SMESG000022863.1 annotated as RPL26 (ribosomal protein L26), COX11 (cytochrome c oxidase copper chaperone) and an unknown transcript, respectively) the probes were predicted to map to loci that do not exhibit RNA-seq coverage (Figure 4C, Supplemental Figs. S1C, S1D). The likely reason for this is inaccurate gene annotation. Alternatively, target regions might represent repetitive, multimapping sequences, which we excluded during read mapping. Taken together, our off-target analysis revealed that a maximum of 11 genes might be affected by our rRNA removal procedure - a very low number that underscores the specificity and efficiency of our depletion protocol.

Applicability of the described ribodepletion method to other organisms

296 To demonstrate the applicability of the developed rRNA workflow to other organisms, we
297 employed our protocol to the depletion of ribosomal RNA from *Salmonella typhimurium* using a pool
298 of organism-specific DNA probes (riboPOOL) developed by siTOOLs Biotech (Martinsried, Germany)
299 (Figure 5A). We compared the libraries resulting from the application of our newly developed
300 procedure to the established rRNA depletion workflow that utilizes the Ribo-Zero rRNA Removal Kit
301 (Bacteria) from Illumina. Removal of rRNA from a *S. typhimurium* sample using riboPOOL probes was
302 comparably successful as a depletion reaction using Ribo-Zero, leaving as low as 3.4% rRNA in the final
303 library (Figure 5A). Moreover, an overall comparison of gene expression levels showed a high
304 correlation (Pearson correlation > 0.98) between riboPOOL depleted libraries and libraries prepared
305 with the Ribo-Zero kit (Figure 5B). Taken together, the rRNA depletion workflow described in this
306 manuscript is robust and easily applicable to any bacterial and eukaryotic species of choice utilizing
307 organism-specific probes.
308
309 **Discussion**
310 For samples from typical model organisms, such as human, mouse and rat, there are
311 numerous commercial kits available for the removal of rRNA, e.g. NEBNext from New England Biolabs,
312 RiboGone from Takara and RiboCop from Lexogen. This also applies to typical gram-positive and
313 gram-negative bacteria (MICROBExpress from Thermofisher and Ribominus from Invitrogen).
314 Moreover, these kits can be utilized with a certain degree of compatibility for the depletion of rRNA in
315 organisms of distant phylogenetic groups (e.g. RiboMinus Eukaryote Kit for RNA-Seq, Invitrogen).
316 However, as the breadth of molecularly tractable organisms increases in the past decade, the
317 necessity to develop organism-specific rRNA depletion techniques rises as well [42–44]. To date, such
318 custom protocols were either based on rRNA removal by biotinylated antisense probes along with

streptavidin-coated magnetic beads or they were based on the digestion of DNA-RNA hybrids with RNase H [14, 45–47].

In this study, we describe a novel rRNA depletion workflow for the planarian flatworm *S. mediterranea*. Our protocol is based on the hybridization of biotinylated DNA probes to planarian rRNA followed by the subsequent removal of the resulting rRNA-DNA hybrids with streptavidin-labeled magnetic beads. A comparative analysis between ribodepleted and poly(A)-selected libraries in planarians revealed that our protocol retains all information present in poly(A) selected libraries. Over and above, we found ribodepleted libraries to contain additional information on histone mRNAs, transposable elements (of which many do not have polyA tails) and transcriptional intermediates. This highlights that ribodepletion better retains the dynamic information of transcriptomes. As more and more co-translational decay pathways are being discovered [48], understanding the dynamic of RNA degradation will become more important in the near future. Moreover, by successfully depleting rRNA from other freshwater triclad species, we could demonstrate the versatility of the DNA probes designed for *S. mediterranea*. Last, we validated the efficiency of the developed workflow by removal of rRNA in the gram-negative bacterium *S. typhimurium*. Therefore, the proposed workflow serves as an efficient and cost-effective method for rRNA depletion in any organism of interest.

Conclusions

This study describes an rRNA depletion workflow for the planarian model system *S. mediterranea* and related freshwater triclads. It is based on the hybridization of 40-mer biotinylated DNA oligos to ribosomal RNA followed by the subtraction of formed DNA-RNA hybrids. The protocol is very robust and ensures the efficient removal of rRNA even from low input total RNA. Moreover, we

341 suggest the general applicability of the presented workflow to any prokaryotic or eukaryotic
342 organisms by using organism-specific pools of probes.
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363

Materials and Methods

Ribosomal RNA depletion

Ribosomal RNA depletion was conducted as described in the result section. To evaluate Fragment analyzer separation profiles, planarian total RNA (1000 ng each sample) was subjected to rRNA depletion using varying concentrations of NaCl (0 mM, 50 mM, 250 mM, 500 mM) in the hybridization buffer.

Phylogenetic tree

The phylogenetic tree was constructed using NCBI taxonomic names at phyloT (https://phylot.biobyte.de). The tree was visualized using the Interactive Tree of Life (iToL) tool [49].

Planarian rRNA-depleted RNA-Seq dataset

Raw sequencing reads for planarian rRNA-depleted dataset were downloaded from the project GSE122199 (GSM3460490, GSM3460491, GSM3460492). The libraries were prepared as described [40]. Briefly, planarian rRNA depleted RNA-Seq libraries were prepared from 100,000 FACS-sorted planarian X1 cells as described [27] and sequenced on an Illumina Next-Seq 500 platform (single-end, 75 bp).

Publicly available RNA-Seq datasets

Raw sequencing reads for all datasets were downloaded from the Sequence read archive (SRA). Planarian polyA B1 rep1, polyA B1 rep2, polyA B1 rep3 correspond to SRR2407875, SRR2407876, and SRR2407877, respectively, from the Bioproject PRJNA296017 (GEO: GSE73027) [30].

Planarian polyA B2 rep1, polyA B2 rep2 samples correspond to SRR4068859, SRR4068860 from the Bioproject PRJNA338115 [32]. Planarian polyA B3 rep1, polyA B3 rep2, polyA B3 rep3 correspond to SRR7070906, SRR7070907, SRR7070908, respectively, (PRJNA397855) [31]. Only first read of the pair was analyzed for polyA B2 and polyA B3 from Bioprojects PRJNA338115 and PRJNA397855.

Salmonella typhimurium SL1344 RNA Seq datasets

In total, four samples were sequenced for *Salmonella typhimurium* SL1344 by the IMGM Laboratories GmbH (Martinsried, Germany) on an Illumina NextSeq 500 platform (single-end, 75bp). Of these sequenced samples, one sample was of untreated total RNA, two samples were of RiboZero and one of RiboPools treated total RNA. The sequencing data are available at NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under the accession number GSE132630.

Processing of RNA-Seq libraries

Planarian RNA-Seq data were processed as follows: Reads after removal of 3'-adapters and quality filtering with Trimmomatic (0.36) [50] were trimmed to a length of 50 nt. For libraries sequenced in pair-end mode, only the first read of a pair was considered for the analysis. Next, sequences mapped to planarian rRNAs were removed with SortMeRNA [51]. Reads were assigned to the reference genome version SMESG.1 [34] or consensus transposable element sequences [37] in strand-specific mode and quantified with Kallisto [52] with "--single -l 350 -s 30 -b 30". Differential gene expression analysis was performed with DeSeq2 [53]. To annotate RNA-Seq reads to coding regions (CDS), reads were mapped to the planarian genome using STAR [54] with the following setting: --quantMode TranscriptomeSam --outFilterMultimapNmax 1.

RNA sequencing data from *Salmonella typhimurium* SL1344 were processed with READemption 0.4.3 using default parameters [55]. Sequenced reads were mapped to the RefSeq genome version NC_016810.1 and plasmids NC_017718.1, NC_017719.1, NC_017720.1.

Analysis of DNA probe specificity

DNA probe sequences were mapped to the planarian transcriptome SMEST.1 [56] using the BURST aligner (v0.99.7LL; DB15) [57] with the following settings "-fr -i .80 -m FORAGE". Only sequences that mapped to genes in antisense orientation with no more than 8 mismatches were considered as potential probe targets.

Acknowledgements

We thank Jochen Rink, Mario Ivankovic and Miquel Vila Farré from the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden for generously providing various flatworm species. We thank Jens Hör from the Institute for Molecular Infection Biology, University of Würzburg, for providing the *Salmonella* RNA and IMGM Laboratories for library preparation and sequencing of *Salmonella* samples. We also acknowledge the KeyLab Genomics & Bioinformatics at the University of Bayreuth for Fragment Analyzer (Agilent) measurements. This work was supported by the Elite Network of Bavaria, the University of Bayreuth, the Paul Ehrlich and Ludwig Darmstaedter Prize for Young Researchers (to C.D.K) and the IZKF at the University of Würzburg (project Z-6). A.S.A. is an investigator of the Howard Hughes Medical Institute and the Stowers Institute for Medical Research.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

I.V.K. and C.D.K. conceived and designed the study; I.V.K. and K.D. acquired data; E.J.R. assembled planarian rRNA sequences; I.V.K., E.J.R. and S.D. performed computation analyses; A.S.A. and C.D.K. supervised the study; I.V.K. and C.D.K. drafted the manuscript with input from all authors.

438 **References**

439 1. Elliott SA, Sánchez Alvarado A. The history and enduring contributions of planarians to the study of
440 animal regeneration. Wiley Interdiscip Rev Dev Biol. 2013;2:301–26. doi:10.1002/wdev.82.

441 2. Zeng A, Li H, Guo L, Gao X, McKinney S, Wang Y, et al. Prospectively isolated tetraspanin+ neoblasts
442 are adult pluripotent stem cells underlying planaria regeneration. Cell. 2018;173:1593–608.

443 3. Sahu S, Dattani A, Aboobaker AA. Secrets from immortal worms: What can we learn about
444 biological ageing from the planarian model system? Semin Cell Dev Biol. 2017;70:108–21.
445 doi:10.1016/j.semcdb.2017.08.028.

446 4. Collins JJ. Platyhelminthes. Curr Biol. 2017;27:R252–6. doi:10.1016/J.CUB.2017.02.016.

447 5. Egger B, Gschwentner R, Rieger R. Free-living flatworms under the knife: past and present. Dev
448 Genes Evol. 2007;217:89–104. doi:10.1007/s00427-006-0120-5.

449 6. Sikes JM, Newmark PA. Restoration of anterior regeneration in a planarian with limited
450 regenerative ability. Nature. 2013;500:77–80. doi:10.1038/nature12403.

451 7. Zhao W, He X, Hoadley KA, Parker JS, Hayes D, Perou CM. Comparison of RNA-Seq by poly (A)
452 capture, ribosomal RNA depletion, and DNA microarray for expression profiling. BMC Genomics.
453 2014;15:419. doi:10.1186/1471-2164-15-419.

454 8. Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense
455 transcription in the mammalian transcriptome. Science. 2005;309:1564–6.
456 doi:10.1126/science.1112009.

457 9. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, et al. Transcriptional Maps of 10

458 Human Chromosomes at 5-Nucleotide Resolution. Science (80-). 2005;308:1149–54.
459 doi:10.1126/SCIENCE.1108625.

460 10. Kim T-K, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at
461 neuronal activity-regulated enhancers. Nature. 2010;465:182. doi:10.1038/NATURE09033.

462 11. Finnegan DJ. Retrotransposons. Curr Biol. 2012;22:R432–7. doi:10.1016/J.CUB.2012.04.025.

463 12. Reuter M, Berninger P, Chuma S, Shah H, Hosokawa M, Funaya C, et al. Miwi catalysis is required
464 for piRNA amplification-independent LINE1 transposon silencing. Nature. 2011;480:264–7.
465 doi:10.1038/nature10672.

466 13. Wang W, Yoshikawa M, Han BW, Izumi N, Tomari Y, Weng Z, et al. The initial uridine of primary
467 piRNAs does not create the tenth adenine that Is the hallmark of secondary piRNAs. Mol Cell.
468 2014;56:708–16. doi:10.1016/j.molcel.2014.10.016.

469 14. Kukutla P, Steritz M, Xu J. Depletion of Ribosomal RNA for Mosquito Gut Metagenomic RNA-seq. J
470 Vis Exp. 2013. doi:10.3791/50093.

471 15. Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Sánchez Alvarado A. Developmental biology:
472 SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. Science (80-). 2005;310:1327–30.

473 16. Hayashi T, Asami M, Higuchi S, Shibata N, Agata K. Isolation of planarian X-ray-sensitive stem cells
474 by fluorescence-activated cell sorting. Dev Growth Differ. 2006;48:371–80.

475 17. Wetmur JG. DNA Probes: Applications of the Principles of Nucleic Acid Hybridization. Crit Rev
476 Biochem Mol Biol. 1991;26:227–59. doi:10.3109/10409239109114069.

477 18. Green MR, Sambrook J. Molecular Cloning, 3-Volume Set : A Laboratory Manual. 2012.

19. Sun S, Xie H, Sun Y, Song J, Li Z. Molecular characterization of gap region in 28S rRNA molecules in brine shrimp Artemia parthenogenetica and planarian Dugesia japonica. Biochem. 2012;77:411–7.

20. van Keulen H, Mertz PM, LoVerde PT, Shi H, Rekosh DM. Characterization of a 54-nucleotide gap region in the 28S rRNA gene of Schistosoma mansoni. Mol Biochem Parasitol. 1991;45:205–14. doi:10.1016/0166-6851(91)90087-M.

21. Ware VC, Renkawitz R, Gerbi SA. rRNA processing: removal of only nineteen bases at the gap between 28S alpha and 28S beta rRNAs in Sciara coprophila. Nucleic Acids Res. 1985;13:3581–97. http://www.ncbi.nlm.nih.gov/pubmed/2989775. Accessed 16 Apr 2019.

22. Carranza S, Baguñà J, Riutort M. Origin and Evolution of Paralogous rRNA Gene Clusters Within the Flatworm Family Dugesiidae (Platyhelminthes, Tricladida). J Mol Evol. 1999;49:250–9. doi:10.1007/PL00006547.

23. Carranza S, Giribet G, Ribera C, Riutort M. Evidence that two types of 18S rDNA coexist in the genome of Dugesia (Schmidtea) mediterranea (Platyhelminthes, Turbellaria, Tricladida). Mol Biol Evol. 1996;13:824–32. doi:10.1093/oxfordjournals.molbev.a025643.

24. Draper DE. A guide to ions and RNA structure. RNA. 2004;10:335–43. doi:10.1261/RNA.5205404.

25. Lambert D, Leipply D, Shiman R, Draper DE. The influence of monovalent cation size on the stability of RNA tertiary structures. J Mol Biol. 2009;390:791–804. doi:10.1016/j.jmb.2009.04.083.

26. Garrey SM, Blech M, Riffell JL, Hankins JS, Stickney LM, Diver M, et al. Substrate binding and active site residues in RNases E and G: role of the 5'-sensor. J Biol Chem. 2009;284:31843–50. doi:10.1074/jbc.M109.063263.

498 27. Zhang Z, Theurkauf WE, Weng Z, Zamore PD. Strand-specific libraries for high throughput RNA
499 sequencing (RNA-Seq) prepared without poly(A) selection. Silence. 2012;3:9. doi:10.1186/1758-907X-
500 3-9.

501 28. Woese CR. Interpreting the universal phylogenetic tree. Proc Natl Acad Sci U S A. 2000;97:8392–6.

502 29. Ochman H, Elwyn S, Moran NA. Calibrating bacterial evolution. Proc Natl Acad Sci.
503 1999;96:12638–43.

504 30. Duncan EM, Chitsazan AD, Seidel CW, Alvarado AS. Erratum: Set1 and MLL1/2 Target Distinct Sets
505 of Functionally Different Genomic Loci In Vivo (Cell Reports (2015) 13 (12) (p2741–
506 2755)(S2211124715014084)(10.1016/j.celrep.2015.11.059)). Cell Rep. 2016;17:930.

507 31. Schmidt D, Reuter H, Hüttner K, Ruhe L, Rabert F, Seebeck F, et al. The Integrator complex
508 regulates differential snRNA processing and fate of adult stem cells in the highly regenerative
509 planarian Schmidtea mediterranea. PLoS Genet. 2018;14:e1007828.
510 doi:10.1371/journal.pgen.1007828.

511 32. Mihaylova Y, Abnave P, Kao D, Hughes S, Lai A, Jaber-Hijazi F, et al. Conservation of epigenetic
512 regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells. Nat Commun.
513 2018;9:3633. doi:10.1038/s41467-018-06092-6.

514 33. Williams AG, Thomas S, Wyman SK, Holloway AK. RNA-seq Data: Challenges in and
515 Recommendations for Experimental Design and Analysis. Curr Protoc Hum Genet. 2014;83:11.13.1-20.
516 doi:10.1002/0471142905.hg1113s83.

517 34. Grohme MA, Schloissnig S, Rozanski A, Pippel M, Young GR, Winkler S, et al. The genome of

518 Schmidtea mediterranea and the evolution of core cellular mechanisms. Nature. 2018;554:56–61.
519 doi:10.1038/nature25473.

520 35. Lakshmanan V, Bansal D, Kulkarni J, Poduval D, Krishna S, Sasidharan V, et al. Genome-Wide
521 Analysis of Polyadenylation Events in Schmidtea mediterranea . G3:
522 Genes|Genomes|Genetics. 2016.

523 36. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: Life
524 without a poly(A) tail. Nature Reviews Genetics. 2008;:843–54.

525 37. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic
526 genomes. Mob DNA. 2015;6:11. doi:10.1186/s13100-015-0041-9.

527 38. Kazazian HH. Mobile Elements : Drivers of Genome Evolution. Science (80-). 2004;303:1626–32.
528 doi:10.1126/science.1089670.

529 39. Shibata N, Kashima M, Ishiko T, Nishimura O, Rouhana L, Misaki K, et al. Inheritance of a Nuclear
530 PIWI from Pluripotent Stem Cells by Somatic Descendants Ensures Differentiation by Silencing
531 Transposons in Planarian. Dev Cell. 2016;37:226–37. doi:10.1016/j.devcel.2016.04.009.

532 40. Kim I V., Duncan EM, Ross EJ, Gorbovytska V, Nowotarski S, Elliott SA, et al. Planarians recruit
533 piRNAs for mRNA turnover in adult stem cells. In revision.

534 41. Saberi A, Jamal A, Beets I, Schoofs L, Newmark PA. GPCRs Direct Germline Development and
535 Somatic Gonad Function in Planarians. PLoS Biol. 2016;14:e1002457.
536 doi:10.1371/journal.pbio.1002457.

537 42. Dietrich MR, Ankeny RA, Chen PM. Publication trends in model organism research. Genetics.

538 2014;198:787–94. doi:10.1534/genetics.114.169714.

539 43. Goldstein B, King N. The Future of Cell Biology: Emerging Model Organisms. Trends Cell Biol.
540 2016;26:818–24. doi:10.1016/j.tcb.2016.08.005.

541 44. Russell JJ, Theriot JA, Sood P, Marshall WF, Landweber LF, Fritz-Laylin L, et al. Non-model model
542 organisms. BMC Biol. 2017;15:55. doi:10.1186/s12915-017-0391-5.

543 45. O'Neil D, Glowatz H, Schlumpberger M. Ribosomal RNA Depletion for Efficient Use of RNA-Seq
544 Capacity. In: Current Protocols in Molecular Biology. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2013.
545 p. 4.19.1-4.19.8. doi:10.1002/0471142727.mb0419s103.

546 46. Li S-K, Zhou J-W, Yim AK-Y, Leung AK-Y, Tsui SK-W, Chan T-F, et al. Organism-specific rRNA capture
547 system for application in next-generation sequencing. PLoS One. 2013;8:e74286.
548 doi:10.1371/journal.pone.0074286.

549 47. Morlan JD, Qu K, Sinicropi D V. Selective depletion of rRNA enables whole transcriptome profiling
550 of archival fixed tissue. PLoS One. 2012;7:e42882. doi:10.1371/journal.pone.0042882.

551 48. Ibrahim F, Maragkakis M, Alexiou P, Mourelatos Z. Ribothrypsis, a novel process of canonical
552 mRNA decay, mediates ribosome-phased mRNA endonucleolysis. Nat Struct Mol Biol. 2018;25:302–
553 310.

554 49. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic
555 Acids Res. 2019. doi:10.1093/nar/gkz239.

556 50. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data.
557 Bioinformatics. 2014;30:2114–20. doi:10.1093/bioinformatics/btu170.

558 51. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in
559 metatranscriptomic data. Bioinformatics. 2012;28:3211–7. doi:10.1093/bioinformatics/bts611.

560 52. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat
561 Biotechnol. 2016;34:525–7. doi:10.1038/nbt.3519.

562 53. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq
563 data with DESeq2. Genome Biol. 2014;15:550. doi:10.1186/s13059-014-0550-8.

564 54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-
565 seq aligner. Bioinformatics. 2013;29:15–21. doi:10.1093/bioinformatics/bts635.

566 55. Forstner KU, Vogel J, Sharma CM. READemption--a tool for the computational analysis of deep-
567 sequencing-based transcriptome data. Bioinformatics. 2014;30:3421–3.
568 doi:10.1093/bioinformatics/btu533.

569 56. Rozanski A, Moon H, Brandl H, Martín-Durán JM, Grohme MA, Hüttner K, et al. PlanMine 3.0-
570 improvements to a mineable resource of flatworm biology and biodiversity. Nucleic Acids Res.
571 2019;47:D812–20. doi:10.1093/nar/gky1070.

572 57. Al-Ghalith, Knights G, Knights D. BURST enables optimal exhaustive DNA alignment for big data.
573 2017. doi.org/10.5281/zenodo.806850.

574

Figure 1

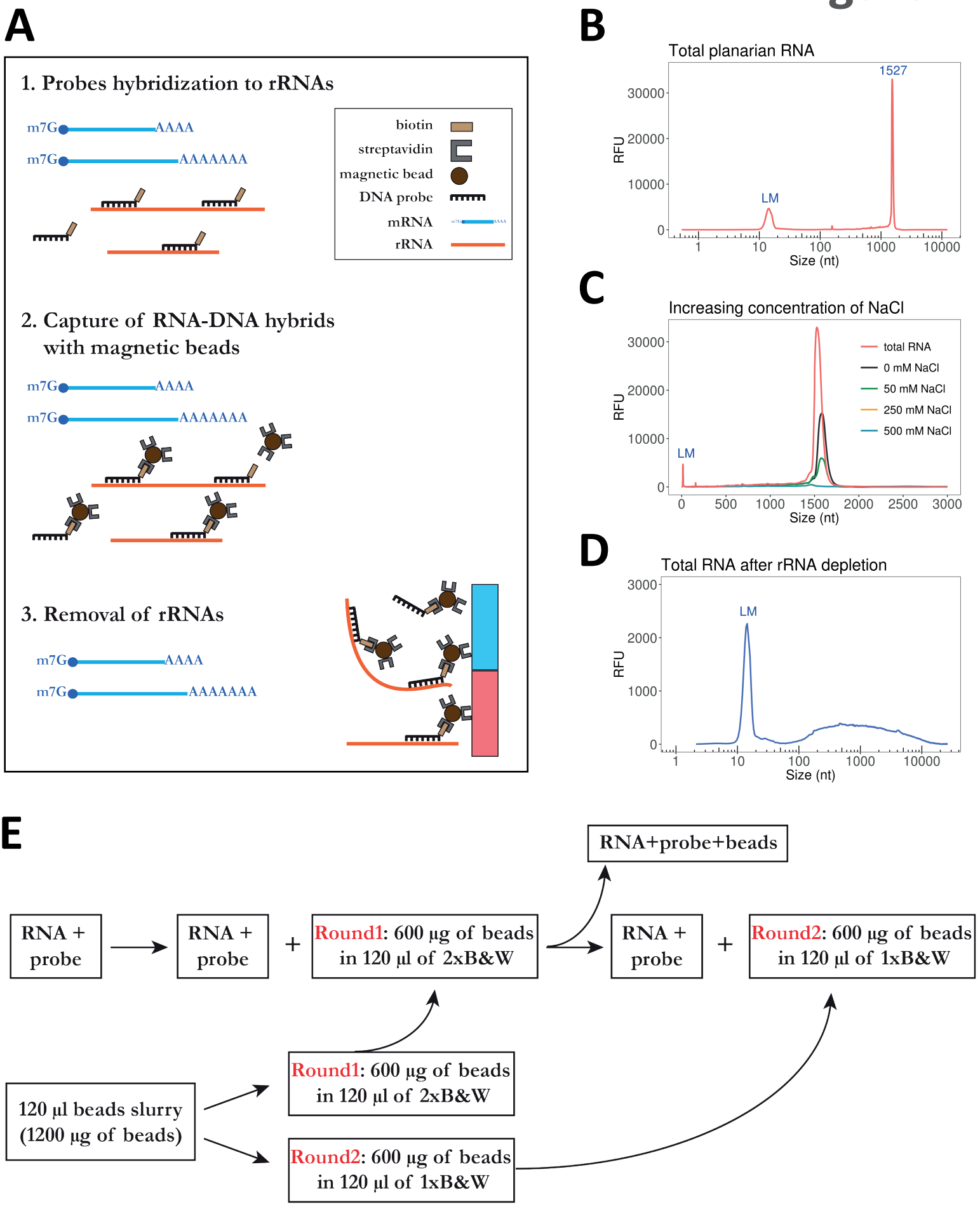


Figure 1. Efficiency of rRNA removal from total planarian RNA.

(A) Schematic representation of rRNA depletion workflow. Biotinylated DNA probes are hybridized to rRNA, followed by subtraction of rRNA-DNA hybrids using streptavidin-labeled magnetic beads. **(B)** Separation profile of planarian total RNA. The large peak at 1527 nts corresponds to the co-migrating 18S rRNAs and the two fragments of processed 28S rRNA. LM stands for lower marker with the length of 15 nt **(C)** Increasing concentration of NaCl improves the efficiency of rRNA removal. **(D)** Total planarian RNA after rRNA depletion. **(E)** Removal of rRNA-DNA hybrids was performed in two consecutive steps using streptavidin-coated magnetic beads resuspended in 2x of 1x B&W buffer.

Figure 2

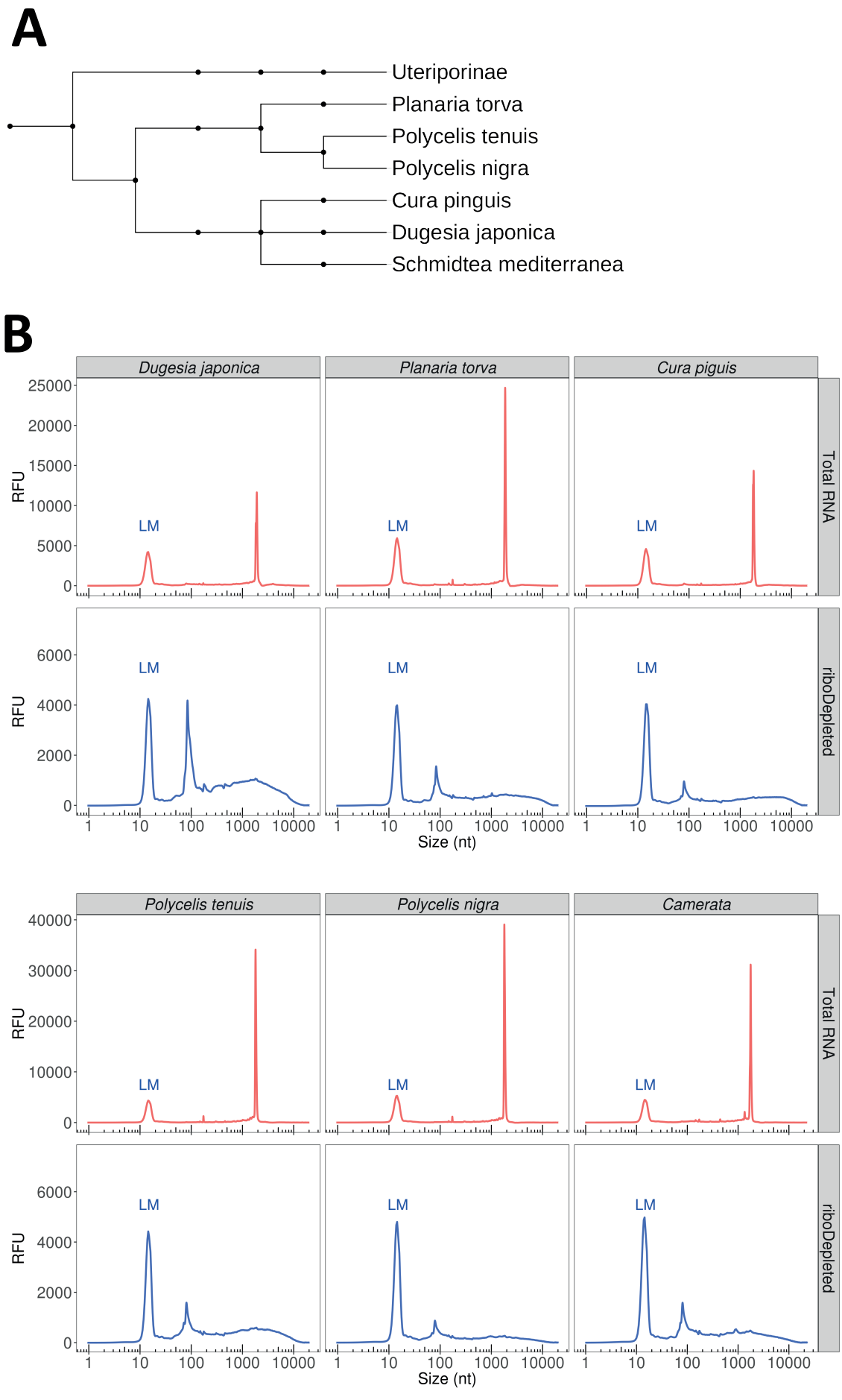
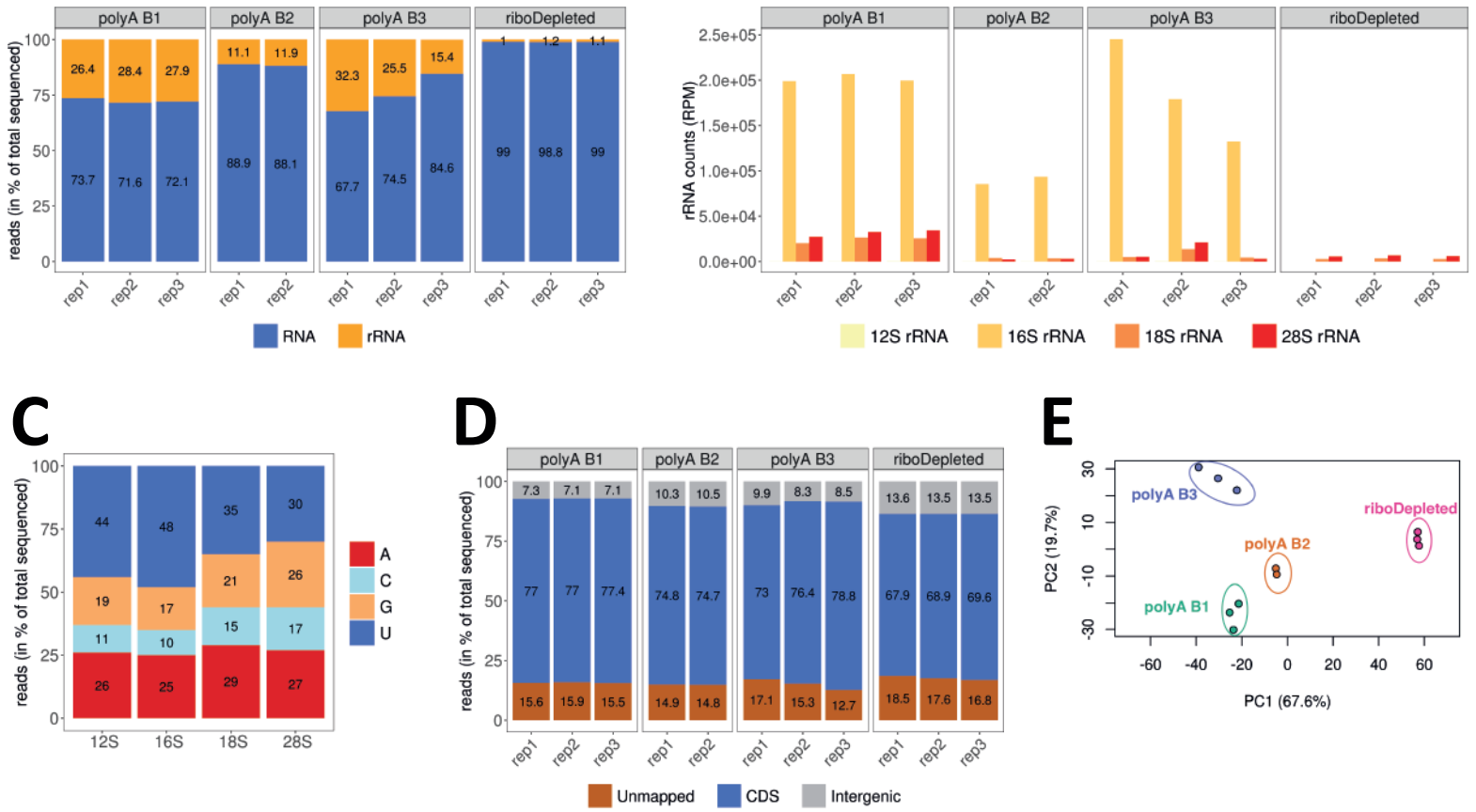


Figure 2. Probes developed for *S. mediterranea* efficiently remove rRNA of other freshwater triclads

(A) Phylogenetic tree showing the taxonomic position of the analyzed planarian species. **(B)** Separation profile of total RNA before and after rRNA depletion. In all species analyzed the 28S rRNA undergoes the "gap deletion" maturation that result in two co-migrating fragments. Both 28S fragments co-migrate with 18S rRNA, resulting in a single rRNA peak. The peak at about 100 nts represents a variety of small RNAs (5S and 5.8S rRNA, tRNAs, and other small RNA fragments) that evaded size selection by the RNA Clean & Concentrator-5 kit (Zymo Research).

Figure 3



F

Sample	Total reads	rRNA reads	Mapped to genome (uniquely)
polyA B1 rep1	22 645 970	5 788 254	13 460 305
polyA B1 rep2	22 363 415	6 145 366	12 846 514
polyA B1 rep3	21 695 759	5 861 887	14 950 109
polyA B2 rep1	51 348 940	5 745 525	38 808 994
polyA B2 rep2	49 044 260	5 854 374	36 796 790
polyA B3 rep1	115 308 395	37 526 234	59 333 489
polyA B3 rep2	93 834 415	24 170 937	58 405 933
polyA B3 rep3	88 752 560	13 853 626	64 843 635
riboDepleted rep1	33 323 527	321 963	26 135 666
riboDepleted rep2	36 314 416	426 284	28 747 336
riboDepleted rep3	32 442 586	331 311	25 862 366

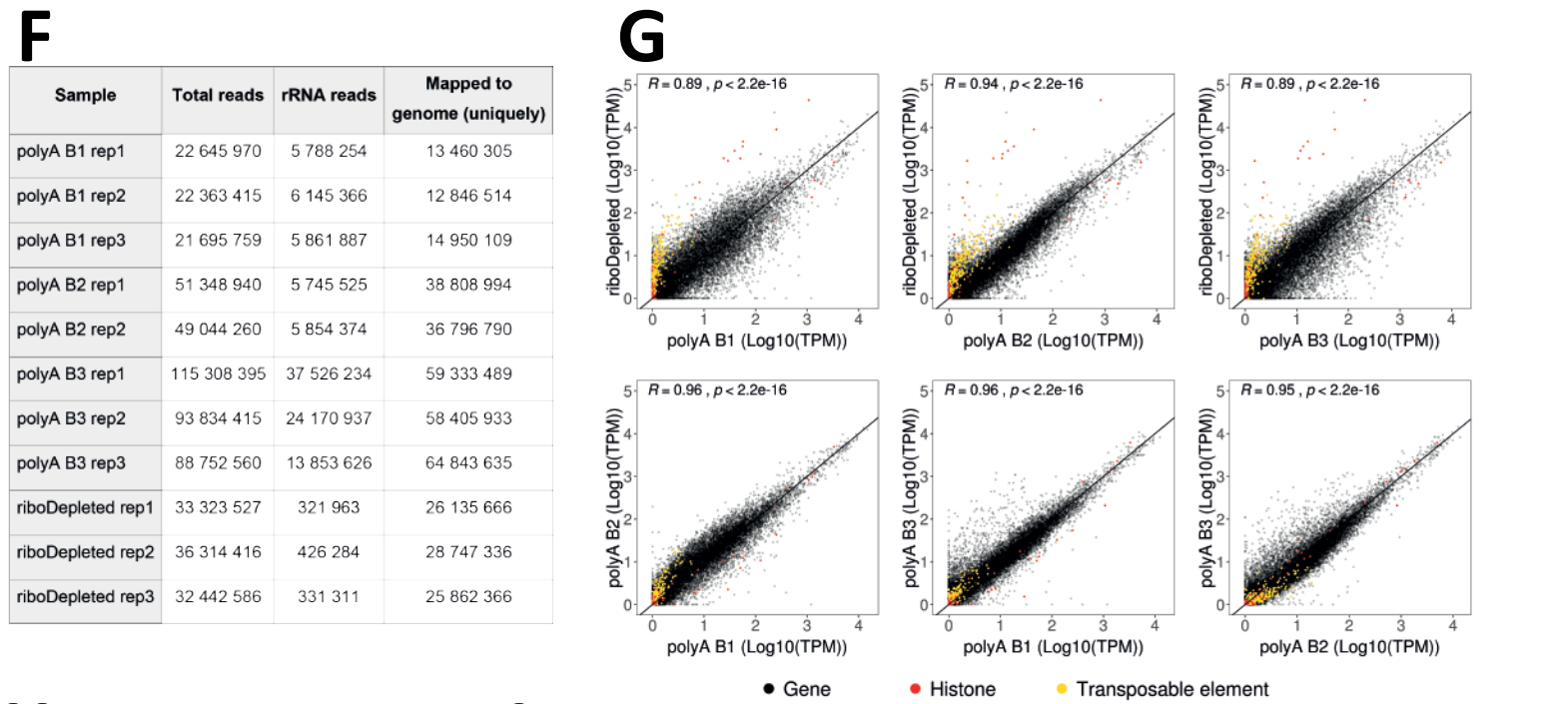


Figure 3. Comparison of rRNA-depleted and poly(A)-enriched planarian RNA-Seq libraries.

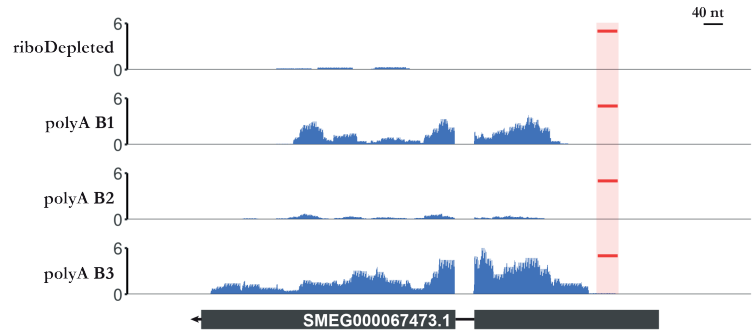
(A) Percentage of rRNAs reads in the sequenced libraries prepared from rRNA-depleted or poly(A)-enriched RNA. **(B)** Species of rRNAs remaining in the final sequenced libraries. **(C)** Nucleotide content of planarian rRNA. **(D)** Percentage of sequenced reads mapped to coding (CDS) and intergenic regions in the planarian genome. **(E)** Principal component analysis (PCA) biplot of log2 expression data for coding genes reveals distinct clustering of all analyzed RNA-Seq experiments. **(F)** Sequencing depth and number of reads mapped to the planarian genome in analyzed rRNA-depleted and poly(A)-enriched samples. **(G)** Comparison of gene expression in transcripts per million (TPM) between planarian ribodepleted and poly(A)-enriched (polyA) RNA-Seq data. The Pearson's correlation coefficient is indicated. **(H)** Increased representation of histone mRNAs in ribodepleted libraries. **(I)** Boxplot of log2 fold changes of the expression of transposable elements between ribodepleted and poly(A)-enriched libraries.

Figure 4

A

Gene_id	TPM in polyA B1	TPM in polyA B2	TPM in polyA B3	TPM in ridoDepleted	LFC polyA+ B1	LFC polyA+ B2	LFC polyA+ B3	Annotation
SMESG000014330.1	0.27	0.24	0.38	0.005	-5,82	-5,30	-6,41	rhodopsin-like orphan gpcr
SMESG000068163.1	0.38	0.15	0.08	0.006	-6,13	-4,39	-3,89	NA
SMESG000069530.1	0.26	0.102	0.027	0.006	-5.99	-4.34	-2.51	NA
SMESG000022863.1	10.05	2.78	10.47	0.60	-4.23	FALSE	-4.34	NA
SMESG000021061.1	10.94	10.32	15.38	5.38	-1.27	-0.82	-1.77	MS3_11207 (MYDGF)
SMESG000067473.1	17.62	6.69	20.57	1.13	-4.21	-2.44	-4.45	MS3_08805 (GID8)
SMESG000044545.1	30.91	32.22	34.31	16.72	-1.11	-0.82	-1.25	REEP2
SMESG000043656.1	48.2	24.8	36.5	12.3	-2,20	-0,88	-1,81	COX11
SMESG000066644.1	1020.6	605.8	798.9	51.9	-4,53	-3,41	-4,19	RPL26

B



C

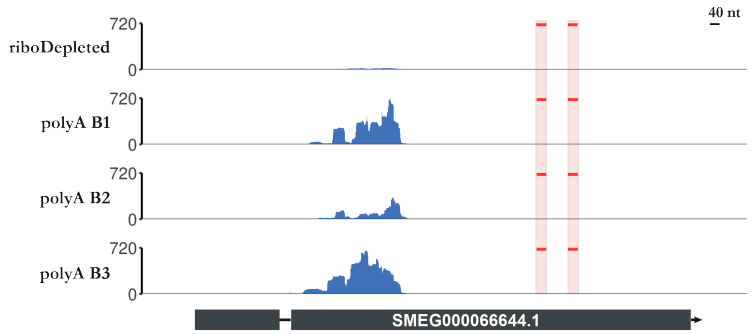
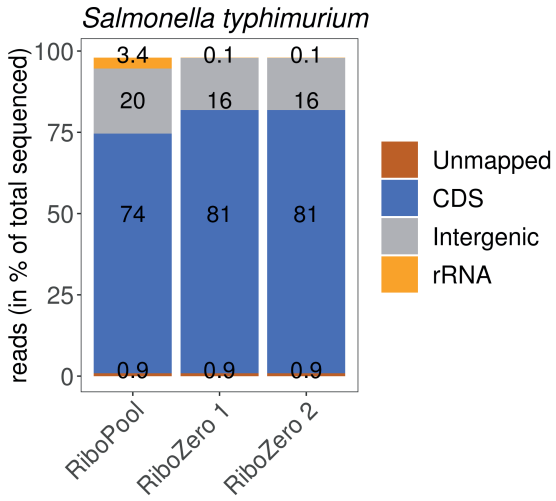


Figure 4. Off-target analysis of DNA probes used for rRNA depletion

(A) Expression level of nine transcripts targeted by designed probes. **(B)** RNA-Seq coverage profile for SMESG000067473.1 in rRNA depleted (ribodepleted) and poly(A) enriched (polyA B1, polyA B2, polyA B3) libraries. The position of antisense probes mapping to the transcripts is marked in red. **(C)** The same as (B) for SMESG000066644.1.

Figure 5

A



B

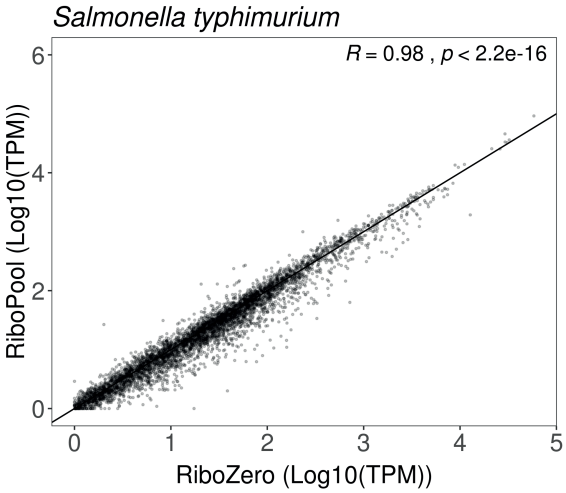
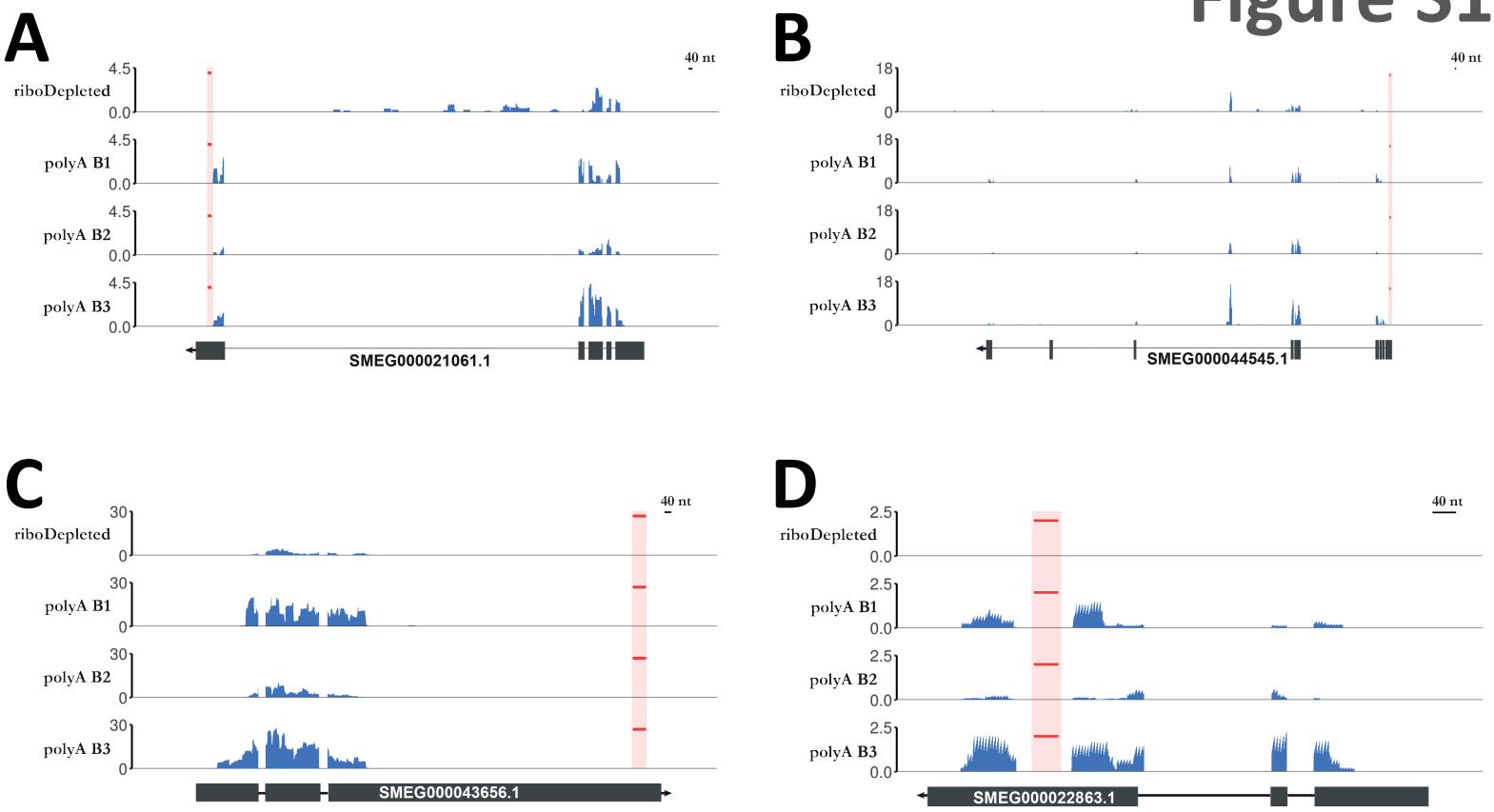


Figure 5. Application of the developed rRNA workflow to other species using organism-specific probes.

(A) Percentage of rRNAs in *Salmonella typhimurium* sequenced libraries prepared using our developed rRNA depletion workflow with organism-specific RiboPool probes or the commercial Ribo-Zero kit. **(B)** Scatter plot showing the transcript abundance (transcript per million (TPM)) between rRNA depleted libraries using our developed workflow or the Ribo-Zero commercial kit. The Pearson's correlation coefficient is indicated.

Figure S1



Supplemental Figure S1.

(A)-(D) RNA-Seq coverage profile for genes potentially targeted by designed probes in rRNA depleted (ribodepleted) and poly(A) enriched (polyA B1, polyA B2, polyA B3) libraries. The position of antisense probes mapping to the transcripts is marked in red.