SCIENTIFIC REPORTS
natureresearch

OPEN

Efficient and specific oligo-based depletion of rRNA

Amelie J. Kraus[1,2,3], Benedikt G. Brink[1,2] & T. Nicolai Siegel [1,2,3]

In most organisms, ribosomal RNA (rRNA) contributes to >85% of total RNA. Thus, to obtain useful information from RNA-sequencing (RNA-seq) analyses at reasonable sequencing depth, typically, mature polyadenylated transcripts are enriched or rRNA molecules are depleted. Targeted depletion of rRNA is particularly useful when studying transcripts lacking a poly(A) tail, such as some non-coding RNAs (ncRNAs), most bacterial RNAs and partially degraded or immature transcripts. While several commercially available kits allow effective rRNA depletion, their efficiency relies on a high degree of sequence homology between oligonucleotide probes and the target RNA. This restricts the use of such kits to a limited number of organisms with conserved rRNA sequences. In this study we describe the use of biotinylated oligos and streptavidin-coated paramagnetic beads for the efficient and specific depletion of trypanosomal rRNA. Our approach reduces the levels of the most abundant rRNA transcripts to less than 5% with minimal off-target effects. By adjusting the sequence of the oligonucleotide probes, our approach can be used to deplete rRNAs or other abundant transcripts independent of species. Thus, our protocol provides a useful alternative for rRNA removal where enrichment of polyadenylated transcripts is not an option and commercial kits for rRNA are not available.

Massive parallel sequencing has become the gold standard for transcriptome analyses and has been employed in organisms as diverse as humans[1], *Trypanosoma brucei*[2] and *Escherichia coli*[3]. Yet, the large proportion of ribosomal RNA (rRNA), comprising >85% of total RNA in most organisms, complicates detection of low abundant transcripts. To increase the sequence coverage of the transcripts of interest, highly abundant transcripts, such as rRNAs, are typically removed from the pool of total RNA prior to sequencing[4].

Generally, one of two strategies is used to remove rRNA from total RNA, enrichment of mature polyadenylated (poly(A)) mRNA or targeted depletion of rRNA. The former is based on the use of oligo(dT) primers during reverse transcription of RNA into cDNA. This approach typically reduces the levels of rRNA to less than 5%[5], but it cannot be used if the RNA of interest is lacking a poly(A) tail. This is the case for partially degraded samples[6], many short and long ncRNAs[7], newly transcribed, unprocessed transcripts[8] or RNA from bacteria[9]. To study these classes of RNA by RNA-seq, hybridization-based rRNA depletion is typically performed using one of the available kits that follow two approaches. One approach is to capture rRNA with complimentary oligos that are coupled to paramagnetic beads, followed by removal of the bound rRNA. Kits following this approach include, for example, the RiboMinus Transcriptome Isolation Kits (ThermoFisherScientific). An alternative strategy is to hybridize complementary DNA oligos to rRNA followed by degradation of the RNA:DNA hybrids using RNaseH[10,11], e.g. NEBNext rRNA Depletion Kits (New England Biolabs). To ensure efficient rRNA depletion and minimal removal of unrelated transcripts, both approaches require a high degree of sequence homology between rRNA transcripts and DNA probes. This restricts the use of the available commercial kits to organisms with rRNA sequences matching those of the provided probes. As a consequence, no commercially available rRNA depletion kits exist for many organisms. However, even if a suitable kit does exist for the organism studied, that oligo sequences and buffer compositions are typically not disclosed poses problems to reproduce analyses if a manufacturer decides to change a kit's composition or to terminate its production. For example, despite significant differences between human and *T. brucei* rRNA sequences, the RiboMinus Eukaryote Kit (Invitrogen, A1083708) has been successfully used to deplete trypanosomal rRNA[12,13]. However, following a change in the kit's composition, the manufacturer no longer recommends the kit for the depletion of trypanosomal rRNA (personal

[1]Department of Veterinary Sciences, Experimental Parasitology, Ludwig-Maximilians-Universität München, 80802, Munich, Germany. [2]Biomedical Center Munich, Department of Physiological Chemistry, Ludwig-Maximilians-Universität München, 82152, Planegg-Martinsried, Germany. [3]Research Center for Infectious Diseases, University of Würzburg, 97080, Würzburg, Germany. Correspondence and requests for materials should be addressed to T.N.S. (email: n.siegel@lmu.de)
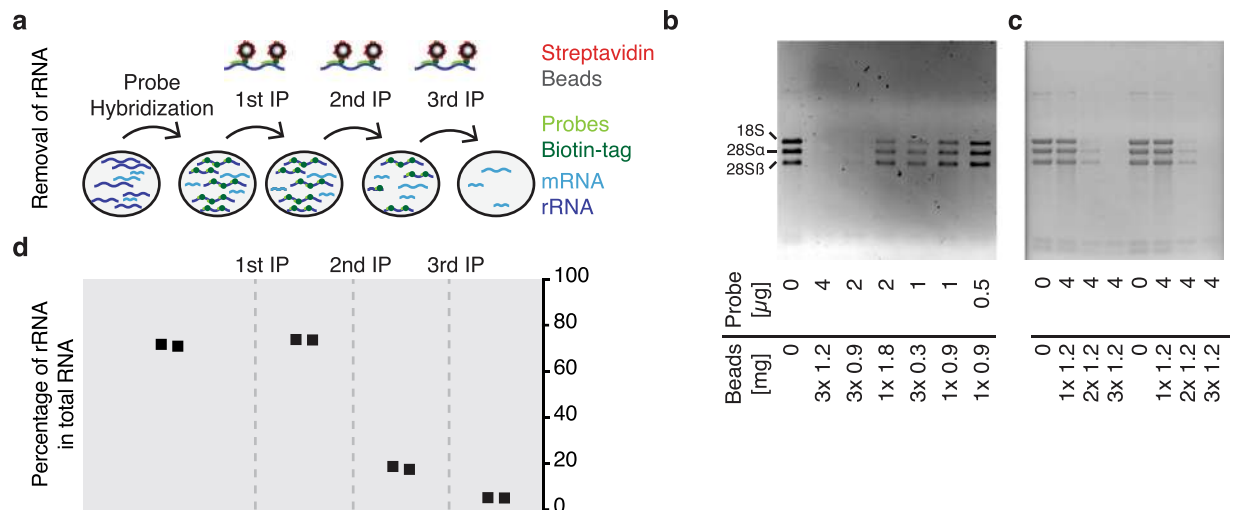
Figure 1. Depletion of rRNA transcripts. (**a**) Outline of depletion strategy. (**b,c**) Agarose gels revealing presence or absence of the three large trypanosomal rRNA transcripts following different depletion conditions. For each condition 2 μg of total RNA was used. (**d**) The samples shown in (**c**) were analyzed by RNA-seq. Shown are the percentages of rRNA (28S, 5.8S, 18S and 5S) relative to total RNA.

communication with ThermoFisherScientific). Similar problems arise in case of production stops. Even though a recent study comparing three kits for the depletion of bacterial rRNA found the Ribo-Zero kit (Illumina) to exhibit the highest efficiency[14], production of this kit has been discontinued as of November 2018.

Thus, our goal was to establish a protocol for the specific depletion of rRNA molecules that would allow the generation of high-quality transcriptome analyses without the need for commercially available kits.

Here, we describe an efficient and highly specific rRNA removal approach that can be easily adapted to deplete rRNA or other transcripts of any species. Using a set of only 12 biotinylated DNA oligos, we were able to reduce the levels of the most abundant rRNA transcripts to less than 5% of the total RNA with minimal off-target effects.

Results

Design of rRNA-specific biotinylated oligonucleotides.
Similar to ribosomes from other eukaryotes, the *T. brucei* ribosome contains more then 50 different rRNAs[15]. Sequencing total RNA, we found that 28S alpha, 28S beta and 18S are the most abundant rRNA transcripts in *T. brucei*, contributing to 75% of the total *T. brucei* RNA (Supplementary Table 1). Thus, to deplete rRNAs of the small and large ribosomal subunits, we established a protocol for the efficient removal of five rRNAs belonging to the major subunits (28S alpha, 28S beta, 18S, 5.8S and 5S) using customized, biotinylated hybridization probes and capturing of the probe:rRNA hybrid by streptavidin-coated beads (outlined in Fig. 1a). To ensure depletion of partially degraded transcripts, rRNA transcripts longer than 1 kb (28S alpha, 28S beta and 18S) were targeted with three probes- one against the 5′-end, one against the middle and the other against the 3′-end of the transcript. The efficiency of rRNA removal strongly relies on sequence-specific hybridization of these probes. Non-specific cross-hybridization can cause biases in transcriptome profiles.

Thus, to obtain highly specific probes, 25 bp probes were designed using Primer3Plus[16] and evaluated for possible cross-hybridizations using BLAST[17]. Next, to increase the annealing temperature, the probe length was extended to 50 bp and re-evaluated for possible cross-hybridization using BLAST. A total of 12 probes (Table 1) were ordered with a 5′-biotin, to allow easy pull-out with streptavidin-coated paramagnetic beads. Alternatively, probes may be designed with specifically developed programs[18].

Using biotinylated oligos, rRNA levels can be decreased to <5% of total RNA.
To establish conditions for efficient rRNA depletion, we compared different template-to-probe ratios and evaluated the effect of multiple rounds of rRNA depletion using different amounts of magnetic beads.

Previously, a template-to-probe mass ratio of 1:2 has been successfully used for the depletion of bacterial rRNA[19]. Thus, we based our tests on similar ratios and compared hybridization of 2 μg of total *T. brucei* RNA with 0.5 μg, 1 μg, 2 μg or 4 μg (240 pmoles) of biotinylated oligos. Based on the binding capacity of the beads (Dynabeads MyOne Streptavidin C1) and the size of our oligos, 1.8 mg and 3.6 mg of beads seemed ideal to capture 2 μg and 4 μg of oligos, respectively.

Analyzing depleted and control RNA by gel electrophoresis, we observed the most efficient rRNA depletion when 4 μg of probes were used (Fig. 1b and Supplementary Fig. 1). In addition, our tests indicated that with 2 μg of total RNA and 4 μg of probes, three rounds of oligo capture, using 1.2 mg of beads per round, were necessary for efficient rRNA removal (Fig. 1c and Supplementary Fig. 2).

Using these conditions, RNA-seq analysis of the rRNA-depleted and control samples indicated that following 0 (control), 1, 2 or 3 rounds of oligo capture, rRNA (28S, 5.8S, 18S and 5S) contributed to 75.6%, 77.6%, 30.4% and 4.3% of total RNA, respectively (Fig. 1d and Supplementary Table 2). Thus, a set of 12 oligos was sufficient to reduce 28S alpha, 28S beta, 18S, 5.8S and 5S rRNA levels to <5%.

50 bp probe	Sequence
1_28Sα_rRNA_S	CTCCGTGGCTAAGGAATGTAATGGGCTGACACCTTCTTTGGCTCAATACT
2_28Sα_rRNA_M	AGGGAAACTTCGACGTGAACCAGCTACTAGGTGGTTCGATGAGTCTTTCG
3_28Sα_rRNA_E	AGGGAGANGTGCGCCGACCACCTTCACTTTCCTAACAAAACGGTGTAACG
4_28Sβ_rRNA_S	ACAACACACCAGAAGGAGGTTAGTAGATACGGCTCGGGAGGGTAGCCTTC
5_28Sβ_rRNA_M	CTTCAAGTCCGTTCCCTTGACTGTCGTTTCGCTAGATAGTAACTCGGGAC
6_28Sβ_rRNA_E	ATTTACGCTTGTGTTCCACGTGAGATTTCTGTTCCCACTGAGCTCGCCTT
7_18S_rRNA_S	ATGTATTAGCTTGGCGTTTCGCCAAGTTATCCAGTAATGCGGAGACCGCA
8_18S_rRNA_M	CACCCTGTCCGGATCTGGTAAAGTTCCCCGTGTTGAGTCAAATTAAACCG
9_18S_rRNA_E	TACGAGACATTCCTCGTTGCGCGACCAATAATTGCAATACTCGGTCCCAA
10_5.8S_rRNA	CTTATCGCACTTTGCTGCGTTCTTCAACGAAATAGGAAGCCAAGTCATCC
11_5S_rRNA	CTAACGAGGCCTGTGGCCGCTTAACTTCACAAATCGGACGGGATATGGTG
12_18S_rRNA	AGTGATTCTGAGGCATGCATGGCTAAGTCCTTGAAACAAGCATATGACTA

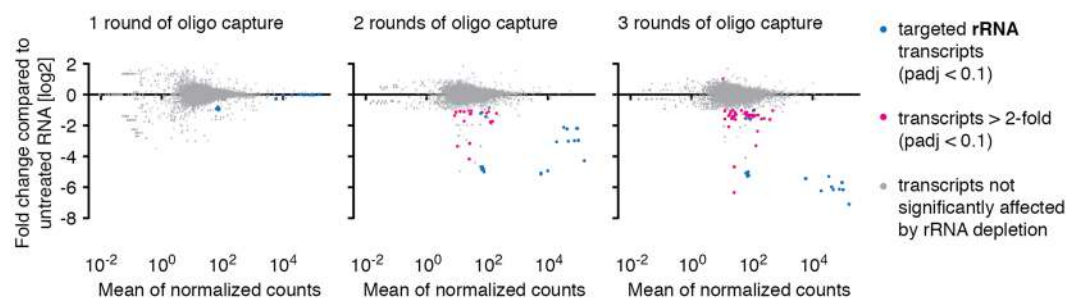**Table 1.** List of oligonucleotide probes.

Figure 2. RNA-seq-based analysis of off-target depletion. Non-rRNA transcript levels were determined in duplicates by RNA-seq following 1, 2 or 3 rounds of oligo capture and compared to non-rRNA transcript levels of an untreated control sample. Reads mapping to rRNA genes were excluded from the mRNA off-target analysis and evaluated separately. Twelve of the 13 transcripts affected by 2 rounds of oligo capture were also among the 50 transcripts affected by 3 rounds of oligo capture.

Depletion of non-rRNA is minimal. The usefulness of hybridization-based rRNA depletion depends strongly on its specificity, since cross-hybridization of rRNA probes to other transcripts will result in incorrect transcript level measurements.

To determine whether the set of 12 biotinylated probes designed for this study cross-hybridized with non-rRNA, we performed differential expression analyses, comparing the effect of different rounds of oligo capture to an untreated control sample.

When RNA was exposed to increasing rounds of oligo capture and analyzed by RNA-seq, we found RNA levels of 0, 13 and 50 non-rRNA transcripts (of 8459 total) to be significantly (padj < 0.1) different by more than 2-fold after 1, 2 and 3 rounds of oligo capture, respectively, compared to the untreated control, (Fig. 2; Supplementary Table 3).

Given the reproducibility of off-target effects, we suspected that the 50 co-depleted transcripts shared some homology with our rRNA-probes. Indeed, using BLAST[17], we observed that some of the co-depleted non-rRNA transcripts did show homology to our probes (Fig. 3 and Supplementary Table 4). However, other transcripts that exhibited a similar degree of homology showed no co-depletion. Overall, our data contained no evidence that the degree of homology between probe and non-rRNA transcript and the degree of co-depletion of non-rRNA transcripts are correlated.

Thus, while we were unable to determine why a set of 50 transcripts was co-depleted with rRNA transcripts, our rRNA-seq data indicated that 99.6% of genes are not affected by the depletion of rRNA transcripts.

rRNA depletion yields a higher percentage of ncRNA than poly(A) enrichment. Most transcriptome analyses benefit from the removal of rRNA. Yet the analysis of transcripts lacking a poly(A) tail, such as many non-protein coding transcripts and immature transcripts, eliminates the possibility of using strategies enriching for polyadenylated RNA. Instead, methods to specifically deplete rRNA are needed. Our RNA-seq data indicated that depletion of rRNA greatly improves the coverage across most genes (Fig. 4).

To better understand the effects of rRNA depletion and mRNA-enrichment-based approaches on transcriptome data, we compared the transcriptome data generated in this study with previously published data generated from poly(A)-enriched RNA and RNA from which the rRNA had been removed using the RiboMinus Eukaryote Kit. We found our dataset to correlate strongly with previously published RNA-seq datasets (Fig. 5 and Supplementary Table 5). Yet, for individual genes we observed marked differences in apparent transcript levels

Target ID	Sequence similarity	Fold change (log2)	e-value
Tb427.10.7650 5.8S rRNA M3	RNA: AACGAAATAAGATGCCATAGTC Probe: AACGAAATAGGAAGCCA–AGTC	-6.35	0.29
Tb427tmp.18.0001 18S rRNA SSU	RNA: AAATTAAACC Probe: AAATTAAACC	-3.33	3.7
Tb427tmp.01.5590 28S rRNA LSU alpha	RNA: CCTTCTTTGGATC Probe: CCTTCTTTGGCTC	-1.58	3.7
Tb427.06.4320 28S rRNA LSU alpha	RNA: AGGAATGGAATGCCCTGACAACTTC Probe: AGGAATGTAATGGGCTGACACCTTC	-0.443	10
Tb427.08.5770 5S rRNA M5	RNA: GCCTGTGGAGCCGCTTAACTTC Probe: GCCTGT–G–GCCGCTTAACTTC	0.30	0.81
Tb427.08.1360 18S rRNA SSU	RNA: CATTCCTCGTTGCGCG Probe: CATTCCTCGTTGCGCG	0.03	0.22

Figure 3. Representative examples of transcripts for which homology to an rRNA probe was observed and that were depleted (top three) or not depleted (bottom three) following rRNA depletion. SSU: small subunit, LSU: large subunit.

(Fig. 6). Since mRNA degradation involves shortening of the poly(A) tail, and transcripts with a short poly(A) are captured inefficiently by oligo-dT-based enrichment methods[20], we investigated whether very long transcripts or transcripts with a short half-life were underrepresented in poly(A)-enriched RNA datasets compared to our rRNA depleted RNA-seq data. While no general correlation could be observed (Supplementary Fig. 3) for individual RNAs, for example the FUTSCH RNA, which is the third-longest RNA in *T. brucei*[21], transcript levels were strongly reduced in the poly(A) data compared to our dataset (fold-change (log2) = -3.298; padj ≤ 0.1).

Overall, we found that transcriptome data from our rRNA-depleted pool of RNA contained a higher percentage of reads aligning outside of ORFs than reads from a poly(A)-enriched RNA pool: 79.8% compared to 54.5%[22]. Such reads could be splicing intermediates or ncRNAs.

Discussion

We describe the establishment of a highly specific and efficient approach to deplete rRNA from total RNA extracts that can be easily adapted to remove any species-specific rRNA or other abundant transcripts that may mask signal from low abundance transcripts.

Performing a systematic RNA-seq analysis following multiple rounds of rRNA depletion, we found that 3 rounds of probe capture were necessary to reduce rRNA levels to <5%. No rRNA depletion was observed following 1 round of oligo capture. We suspect that streptavidin-coated beads have a higher binding affinity for unbound hybridization probes than for probes bound to rRNA. Thus, in the first IP we may have pulled out only unbound probes. In addition, analysis of RNA-seq data revealed a small set of transcripts that was co-depleted with rRNA. Surprisingly, the degree of co-depletion did not correlate with the degree of sequence homology between the co-depleted transcripts and the probes. Thus, we suspect that the off-target effects may be caused by non-specific binding of some transcripts to the streptavidin-coated beads. Given that the off-target depletion was reproducible, it should be possible to normalize for the observed co-depletion if necessary. In addition, the reproducibility means that the co-depletion will not affect comparative RNA-seq analyses. If a specific transcript is co-depleted at similar levels in RNA from wild type and mutant cells, the co-depletion will not affect the measurements of changes in transcript levels. Finally, we suspect that strategies employing enrichment of polyadenylated transcripts introduce biases and that differences in poly(A) tail length may affect the efficiency of poly(T)-primed reverse transcription[20].

To evaluate the robustness of our protocol, we compared the distribution of RNA-seq reads from our rRNA-depleted samples to previously published RNA-seq datasets derived from poly(A)-enriched samples. RNA-seq data from samples in which rRNA was depleted by hybridization to complementary DNA oligos and RNaseH digestion were not included in the comparison since the RNA was isolated either from a genetically manipulated cell line[23] or a different life cycle stage[11]. Comparing reads mapping to ORFs, we observed very high correlations, underscoring the reproducibility of the different methods. However, in our dataset a much higher percentage of reads mapped to regions outside of annotated ORFs. We suspect this increase to be the result of the lack of enrichment for polyadenylated mRNA and to better represent the true distribution of cellular RNA.

In this analysis we focused on the depletion of five rRNA transcripts. However, the *T. brucei* transcriptome contains several other highly abundant transcripts such as VSG-2 transcripts (4% of total RNA after 3 rounds of rRNA depletion) or rRNAs from the gamma, delta and zeta subunits (~10% of total RNA in our control sample). To increase the detection of low abundance mRNA even further, our set of 12 oligos may be extended to target other highly abundant transcripts. While the need for three rounds of depletion adds to the cost of our approach, the cost per rRNA depletion is still lower than it would be with most commercially available kits. To further reduce the costs, we propose regeneration of the streptavidin-coated magnetic beads. It has been shown that the incubation of beads in water at temperatures above 70 °C destroys the biotin-streptavidin interactions without
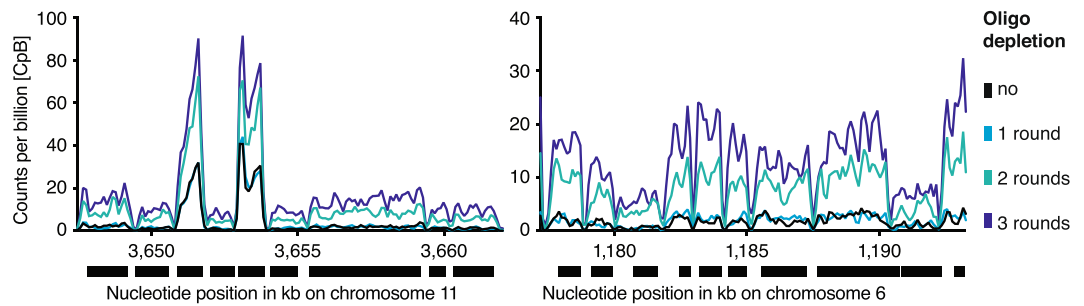
Figure 4. rRNA depletion affects RNA-seq coverage. Shown is the RNA-seq read distribution following multiple rounds of oligo capture across two representative genomic regions.

denaturing the proteins[24]. Alternatively, beads have been regenerated at 90% efficiency by exposure to a solution containing 25% aqueous ammonia and 25% aqueous ammonia in methanol (9:1) at 25 °C[25].

In summary, our study shows that organism-specific design of hybridization probes can be successfully used for the depletion of abundant transcripts. We describe a protocol that allows the efficient depletion of rRNA with little off-target effects. Thus, our protocol provides a useful alternative for the depletion of rRNA where enrichment of polyadenylated transcripts is not an option and commercial kits for rRNA removal are not available.

Materials and Methods

Cell cultivation and RNA isolation. Wild type *T. brucei* derived from the Lister 427 bloodstream-form MITat 1.2 isolate was cultivated at 37 °C and 5% CO_2 in HMI-11 medium (HMI-9 medium[26] without serum plus) up to a density of 0.9×10^6 cells/ml. 45 million cells were harvested at 1,500 g and 4 °C for 10 min. The cell pellet was washed with $1\times$ TDB (5 mM KCl, 80 mM NaCl, 1 mM $MgSO_4$, 20 mM Na_2HPO_4, 2 mM NaH_2PO_4, 20 mM glucose pH 7.4). RNA isolation was performed using the NucleoSpin RNA kit (Macherey-Nagel; cat. no. 740955.10) with minor changes. The cell lysis buffer was modified by adding 3.8 µl 1 M RNAse-free dithiothreitol (Sigma-Aldrich; cat. no. 10197777001) and 1 µl of 1:10 Ambion ERCC RNA Spike-In Mix (ThermoFisherScientific; cat. no. 4456739).

Preparation of biotinylated oligonucleotides. Selected oligonucleotides were ordered with a 5′-biotin tag and HPLC-purified from Sigma-Aldrich. Each oligo was diluted to 100 µM in nuclease-free 10 mM Tris pH 8 (diluted from Ambion 1 M Tris pH 8; Invitrogen; cat. no. AM9856). The rRNA depletion mix was generated by combining equal volumes of each 100 µM oligonucleotide stock.

Removal of ribosomal RNA. The set-up for the hybridization reaction is based on a published protocol[19]. All solutions were kept free from nucleases. For each hybridization reaction, 2 µg of total RNA were mixed with 10 µl of formamide (SigmaAldrich; cat. no. F9037-100ML), 2.5 µl of $20\times$ SSC (3 M NaCl, 0.3 M sodium citrate, the pH was adjusted to 7.0 with HCl), 5 µl of 0.005 M EDTA pH 8 (stock solution 0.5 M; ThermoFisherScientific; cat. no. AM9260G), 2.48 µl of 100 µM rRNA depletion mix (total 4 µg of oligos) and RNAse-free water (ThermoFisherScientific; cat. no. AM9938) to a total volume of 50 µl. Hybridization was performed in a Bio-Rad C1000 Touch Thermal Cycler capable of performing temperature ramps with the following program: 5 min at 80 °C, ramp down to 25 °C at intervals of 5 °C per minute. Subsequently, 2 µl of RNAse-OUT (ThermoFisherScientific; cat. no. 10777019) and 50 µl of 1x SCC containing 20% formamide were added. Dynabead MyOne Streptavidin C1 beads (ThermoFisherScientific; cat. no. 65001) were prepared as recommended by the manufacturer for RNA applications and immobilization of nucleic acids. For each round of oligo capture 120 µl (1.2 mg) of magnetics beads were prepared (unless indicated otherwise). For the first round of depletion, the hybridization reaction was added to the beads, incubated at room temperature (RT) for 15 min using gentle rotation followed by bead separation on a magnetic rack (2 min). For the second and third round of depletion, the supernatant was added to a new batch of beads, incubated at RT for 15 min using gentle rotation followed by bead separation on a magnetic rack (2 min). The resulting supernatant, containing rRNA-depleted RNA, was purified using RNeasy MinElute CleanUp Kit (QIAGEN; cat. no. 74204). Depletion of rRNAs was evaluated on a 1.2% TBE-agarose gel and on an Agilent 2100 Bioanalyzer (Agilent Technologies; cat. no. G2939BA) using the RNA 6000 Nano Kit (Agilent Technologies; cat. no. 5067–1511).

cDNA synthesis, library preparation and sequencing. Synthesis of cDNA was performed using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs; cat. no. E7420) according to the manufacturer's instruction. The concentration of cDNA was measured using Qubit dsDNA HS Assay Kit (Invitrogen, cat. no. Q32854) and a Qubit 2.0 Fluorometer (Invitrogen; cat. no. Q32866). Sequencing libraries were prepared as described previously[27]. To generate strand-specific RNA-seq libraries, uracil excision and removing of the second strand was performed prior to conversion of Y-shaped adapters. Therefore, 3 µl of USER enzyme (New England Biolabs; cat. no. M5505) were mixed with 16 µl of adapter-ligated DNA, 1 µl of TruSeq PCR primer cocktail (50 µM) and 20 µl of KAPA HiFi HotStart ReadyMix (KAPA Biosystems, cat. no. KK2601). USER digestion was performed at 37 °C for 15 min, followed by the published amplification protocol. Library concentrations
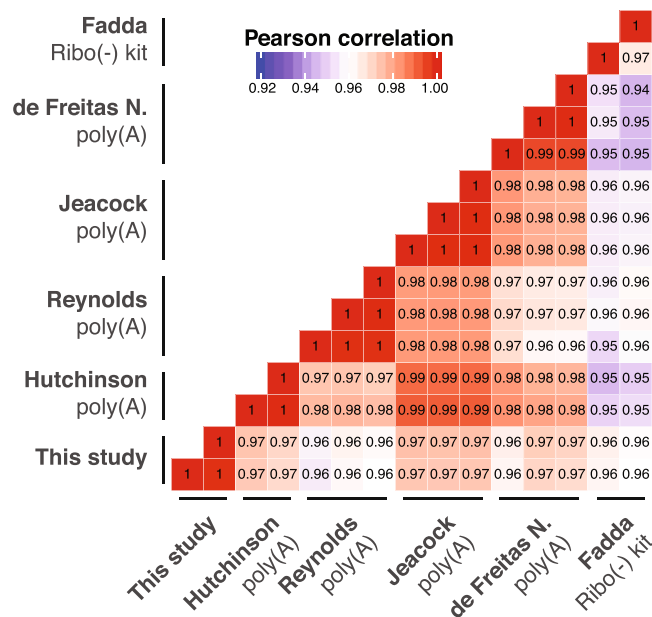
Figure 5. Correlation of different transcriptome analyses[13,22,35–37] and this study (3x oligo capture). Pearson correlation coefficients were calculated based on the relative abundance of transcripts, excluding rRNA transcripts.
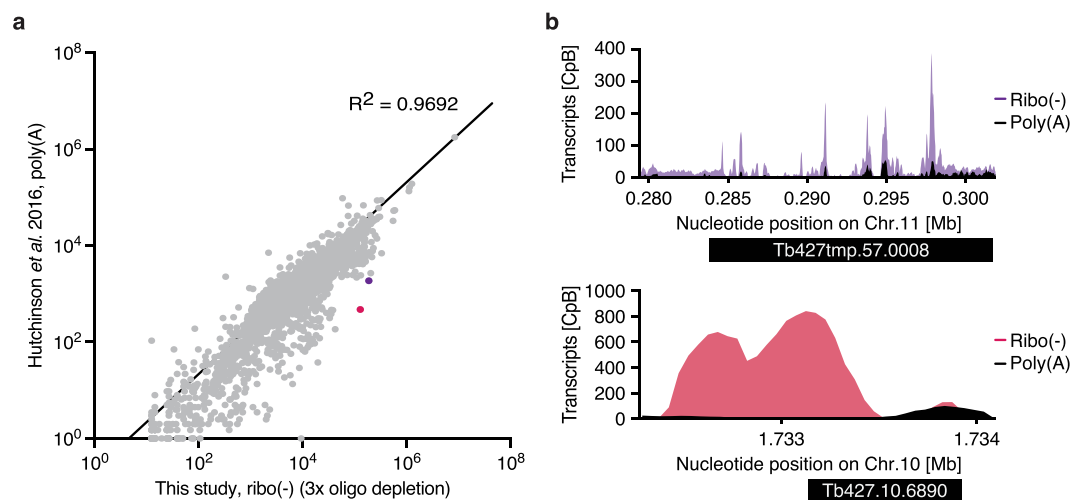

Figure 6. Comparisons of poly(A)-enriched and ribo (−) transcriptome datasets. (**a**) Scatter plot showing the normalized abundance of transcripts for data sets generated from poly(A)-enriched RNA[22] and our ribo (−) datasets generated from RNA following three rounds of oligo capture. Sequence tag abundance was normalized. (**b**) Read coverage plots for two genes, shown as purple and magenta dot in (**a**), for which transcript levels differed between a poly(A)-enriched dataset[22] and our ribo (−) dataset.

were determined in duplicates using Qubit dsDNA HS Assay Kit (Invitrogen, cat. no. Q32854) and a Qubit 2.0 Fluorometer (Invitrogen, cat. no. Q32866) and quantified using the KAPA Library Quantification Kit (KAPA Biosystems, cat. no. KK4824) according to the manufacturer's instruction. Strand-specific RNA-sequencing libraries were sequenced in paired-end mode on an Illumina NextSeq 500 sequencer (2×76 cycles).

Processing of sequencing data. Adapter sequences were removed using Cutadapt[28] and the sequencing datasets were mapped to the *T. brucei* Lister 427 genome assembly (release 36, downloaded from TriTrypDB[29]) using BWA-mem[30]. The alignments were converted from SAM to BAM format, sorted and indexed using SAMtools version 1.8[31]. Additionally, unmapped, PCR or optical duplicate, not primary aligned and supplementary aligned reads were filtered out from the alignment files (SAM flag: 3332).

The same procedure was applied for processing the poly(A) RNA-seq datasets from Hutchinson *et al.*[22]. For on overview of the data analyses performed to generate the different figures see Supplementary Table 6.

Differential expression analysis of RNA-seq data sets.   Using BAM files, reads per gene were counted using the GenomicAlignments package[32] in R[33] and normalized to the total read counts. To determine the fraction of rRNAs in each RNA-seq experiment, rRNA read counts (28S alpha, 28S beta, 18S, 5.8S and 5S) were summed and their proportion of the total reads counts was calculated. Differential gene expression analyses were performed using the DESeq2 package[34] from R/Bioconductor. For analyses of non-rRNA transcripts, rRNA counts were excluded and vice versa.

Comparison of different rRNA depletion approaches.   The ribo(−) kit and different poly(A) RNA-seq datasets derived from previous studies were downloaded from NCBI SRA (accession number: SRP042959[13]) and EBI ENA (accession numbers: PRJEB8747[22]; PRJNA287144[35]; PRJEB22797[36]; PRJEB14403[37]). For correlation analyses, transcripts per million were calculated for each dataset in the non-alignment-based mode with SALMON[38] using the transcriptome of *T. brucei* Lister 427 (release 36, downloaded from TriTrypDB[29]) as reference and normalized to the total number of transcripts excluding the number of rRNA transcripts.

Calculation of reads mapping outside of ORFs.   To calculate the number of reads mapping inside of protein-coding genes, we counted the number of reads mapping to the CDS features of the official annotation using the GenomicAlignments package[32] in R[33]. Given the absence of UTRs in the annotation file of the *T. brucei* Lister 427 genome assembly (release 36, downloaded from TriTrypDB[29]), we extended each CDS with the median UTR length of 89 bp upstream and 400 bp downstream[2]. To calculate reads mapping outside of protein-coding genes, we counted all reads not mapping to CDS + UTRs or rRNA. The total number of reads mapping inside or outside of protein-coding genes was summed up and the percentages were calculated.

Coverage visualization of RNA-seq data-sets.   To visualize read coverage, filtered SAM alignments from replicate sequence experiments were merged using samtools *merge*. The number of reads was normalized per billion mapped reads and coverage files were generated in the wiggle format using COVERnant version 0.3.0 with the subcommand *ratio*, as previously described[39].

Data Availability

All sequencing data generated for this publication have been deposited in the European Nucleotide Archive and can be accessed through the accession number PRJEB31609.

References

1. Sultan, M. *et al*. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* **321**, 956–960 (2008).
2. Siegel, T. N., Hekstra, D. R., Wang, X., Dewell, S. & Cross, G. A. M. Genome-wide analysis of mRNA abundance in two life-cycle stages of *Trypanosoma brucei* and identification of splicing and polyadenylation sites. *Nucleic Acids Res.* **38**, 4946–4957 (2010).
3. Cho, B. K. *et al*. The transcription unit architecture of the Escherichia coli genome. *Nat Biotechnol* **27**, 1043–1049 (2009).
4. Wilhelm, B. T. & Landry, J. R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**, 249–257 (2009).
5. Cui, P. *et al*. A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* **96**, 259–265 (2010).
6. van Hoof, A. & Parker, R. Messenger RNA degradation: beginning at the end. *Curr Biol* **12**, R285–7 (2002).
7. Yang, L., Duff, M. O., Graveley, B. R., Carmichael, G. G. & Chen, L. L. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol* **12**, R16 (2011).
8. Colgan, D. F. & Manley, J. L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11**, 2755–2766 (1997).
9. Sarkar, N. Polyadenylation of mRNA in prokaryotes. *Annu Rev Biochem* **66**, 173–197 (1997).
10. Morlan, J. D., Qu, K. & Sinicropi, D. V. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One* **7**, e42882 (2012).
11. Minia, I., Merce, C., Terrao, M. & Clayton, C. Translation Regulation and RNA Granule Formation after Heat Shock of Procyclic Form Trypanosoma brucei: Many Heat-Induced mRNAs Are also Increased during Differentiation to Mammalian-Infective Forms. *PLoS Negl Trop Dis* **10**, e0004982 (2016).
12. Fadda, A., Farber, V., Droll, D. & Clayton, C. The roles of 3'-exoribonucleases and the exosome in trypanosome mRNA degradation. *RNA* **19**, 937–947 (2013).
13. Fadda, A. *et al*. Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Mol Microbiol* **94**, 307–326 (2014).
14. Petrova, O. E., Garcia-Alcalde, F., Zampaloni, C. & Sauer, K. Comparative evaluation of rRNA depletion procedures for the improved analysis of bacterial biofilm and mixed pathogen culture transcriptomes. *Sci Rep* **7**, 41114 (2017).
15. Berriman, M. *et al*. The genome of the African trypanosome *Trypanosoma brucei*. *Science* **309**, 416–422 (2005).
16. Untergasser, A. *et al*. Primer3–new capabilities and interfaces. *Nucleic Acids Res* **40**, e115 (2012).
17. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
18. Li, S. K. *et al*. Organism-specific rRNA capture system for application in next-generation sequencing. *PLoS One* **8**, e74286 (2013).
19. Stewart, F. J., Ottesen, E. A. & DeLong, E. F. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J* **4**, 896–907 (2010).
20. Guo, Y. *et al*. RNAseq by Total RNA Library Identifies Additional RNAs Compared to Poly(A) RNA Library. *Biomed Res Int* **2015**, 862130 (2015).
21. Kramer, S. Simultaneous detection of mRNA transcription and decay intermediates by dual colour single mRNA FISH on subcellular resolution. *Nucleic Acids Res* (2016).
22. Hutchinson, S., Glover, L. & Horn, D. High-resolution analysis of multi-copy variant surface glycoprotein gene expression sites in African trypanosomes. *BMC Genomics* **17**, 806 (2016).
23. Mugo, E. & Clayton, C. Expression of the RNA-binding protein RBP10 promotes the bloodstream-form differentiation state in Trypanosoma brucei. *PLoS Pathog* **13**, e1006560 (2017).
24. Holmberg, A. *et al*. The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* **26**, 501–510 (2005).
25. Abdul-Hammeda, M., Babalolab, J. O., Breidenc, B. & Sandhoffc, K. Regeneration of Streptavidin-coated Paramagnetic Beads for Multiple uses in Inter-membrane Lipid Transfer Assays. *Annals of Science* **1**, 13–18 (2016).

26. Hirumi, H. & Hirumi, K. Continuous cultivation of *Trypanosoma brucei* blood stream forms in a medium containing a low concentration of serum protein without feeder cell layers. *J. Parasitol.* **75**, 985–989 (1989).
27. Wedel, C. & Siegel, T. N. Genome-wide analysis of chromatin structures in Trypanosoma brucei using high-resolution MNase-ChIP-seq. *Exp Parasitol* **180**, 2–12 (2017).
28. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal* **17**, 10–12 (2011).
29. Aslett, M. *et al*. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res* **38**, D457–62 (2010).
30. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
31. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
32. Lawrence, M. *et al*. Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118 (2013).
33. R Core Team. R: A language and environment for statistical computing. URL http://www.R-project.org/ (2014).
34. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
35. Reynolds, D. *et al*. Histone H3 Variant Regulates RNA Polymerase II Transcription Termination and Dual Strand Transcription of siRNA Loci in Trypanosoma brucei. *PLoS Genet* **12**, e1005758 (2016).
36. Jeacock, L., Faria, J. & Horn, D. Codon usage bias controls mRNA and protein abundance in trypanosomatids. *Elife* **7** (2018).
37. de Freitas Nascimento, J., Kelly, S., Sunter, J. & Carrington, M. Codon choice directs constitutive mRNA levels in trypanosomes. *Elife* **7** (2018).
38. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
39. Wedel, C., Förstner, K. U., Derr, R. & Siegel, T. N. GT-rich promoters can drive RNA pol II transcription and deposition of H2A.Z in African trypanosomes. *EMBO J* **36**, 2581–2594 (2017).

Acknowledgements

Author Contributions

The study was conceptualized by A.J. Kraus and T.N. Siegel. Experimental work was carried out by A.J. Kraus, data analysis was performed by A.J. Kraus and B.G. Brink and the manuscript was written by A.J. Kraus and T.N. Siegel.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-48692-2.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.