Increasing sensitivity of transcriptome profiling in Prokaryotic and Eukaryotic samples by depleting abundant RNAs

Amanda Hoppers, Keerthana Krishnan, Deyra N. Rodriguez, Bradley W. Langhorst, Gautam Naishadham, Kaylinnette Pinet, Jian Sun, Brittany S. Sexton, Anetta Nowosielska-Vecchietti, Christine J. Sumner, Lynne Apone, Fiona J. Stewart, Eileen T. Dimalanta and Theodore B. Davis

New England Biolabs, Inc.

INTRODUCTION

RNA-Seq is a widely used technology with a broad range of applications, including differential expression analysis and alternative splice forms identification, in normal and disease contexts as well as in developmental studies. The technology has been pushed to extremes of very low and degraded samples but still battles with the challenge of having a large dynamic range of transcript expression. Highly expressed transcripts with minimal biological interest can dominate readouts, masking detection of more informative low-abundance transcripts. Here, we present an improved method to enrich for RNAs of interest by eliminating abundant, typically unwanted, RNAs.

This method is based on hybridization of probes to the target RNA followed by degradation of the bound RNAs and probes. Optimized probes target rRNAs from human, mouse, rat or bacteria. The probe design has also been expanded to remove hemoglobin transcripts from multiple derivate blood samples. Incorporation of new enzymes and streamlining of the method make it more robust with increased specificity.

Using strand-specific RNA sequencing we measured sequencing metrics before and after depletion across a wide range of samples, including UHRR, RNA from blood, bacterial monocultures and community across various input amounts (10ng – 1ug). Across all samples we achieved high depletion efficiency (up to 99.9 %) with minimal off target effects. We detect a high number of transcripts, with even coverage across the transcript length, while retaining transcript complexity even at the lowest inputs. The method works efficiently with low input and highly degraded total RNA including FFPE samples.

We conclude that the reduction of abundant transcripts for RNA-Seq studies significantly increases the ability to detect true biological variation that could not be detected in non-depleted samples. The method described here is a reliable and simple solution that greatly improves sensitivity in transcriptome studies and is amenable to high-throughput automation.

METHODS

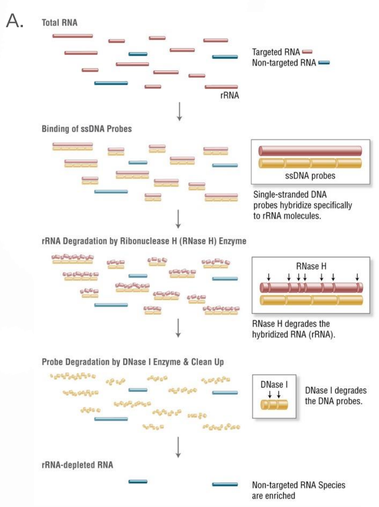


Figure 1. NEBNext Depletion Workflow
A) Total RNA is hybridized with probes targeting unwanted, abundant RNAs (e.g. ribosomal RNA), followed by RNase H digestion to degrade targeted RNA. Finally, DNA probes are digested with DNase I, and the reaction is cleaned using magnetic beads. The entire workflow can be done in <2 hours with only 8 minutes of hands-on time. Depletion of unwanted RNA species, such as ribosomal RNA, can be immediately followed by RNA-seq
B) RNA species targeted for depletion in the Globin mRNA and rRNA Depletion, and Bacterial rRNA Depletion Reagents

Bacterial rRNA depletion

- Total RNA (100ng) from *Escherichia* and *Clostridium* species, as well as the ATCC® MSA-2002™ mock bacterial community were depleted of rRNA using NEBNext Bacterial Depletion reagents.
- Reads were aligned using bowtie 2 [1] end-to-end to a composite reference constructed using Genbank entries from each of the ATCC-20 organisms. Duplicate reads were marked (Picard 1.56) before counting reads overlapping annotated rRNA regions (bedtools 2.26) [2].
- Depletion efficiency was calculated by aligning reads to the reference genomes and counting overlapping rRNA regions.

Globin mRNA and rRNA Depletion

- Human, Mouse and Rat whole blood total RNA (1ug, 100ng, 10ng) was depleted of rRNA and globin mRNA transcripts using the NEBNext Globin & rRNA Depletion Reagents and Globin-Zero Gold rRNA Depletion (*Illumina*®).
- Reads were identified as ribosomal or globin using Mirabait (6 or more shared 25-mers) [3]. rRNA sequences (28S, 18S, 5.8S, 5S 12S, 16S, ITS, ETS) and globin mRNA sequences (HBA1/2, HBB, HBG1/2, HBD, HBE1, HBZ, HBM, HBQ1) were used as baits. Depletion efficiency of rRNA and globin was calculated by dividing matched reads by total number of reads passing instrument quality filtering.

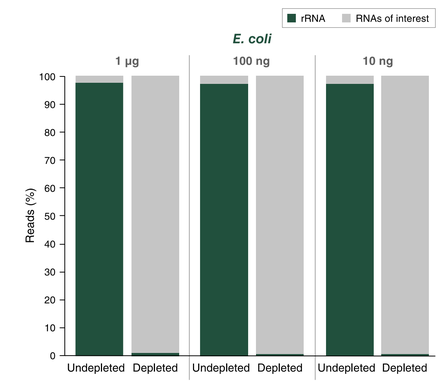
Human, Mouse, Rat V2 rRNA depletion

- Universal Human, Mouse and Rat total RNA (1ug, 100ng, 10ng) was depleted of rRNA using the NEBNext Depletion Kit v2.
- Reads were identified as ribosomal using Mirabait (6 or more, 25-mers) [1]. rRNA sequences (28S, 18S, 5.8S, 5S 12S, 16S, ITS, ETS) were used as baits for human. rRNA sequences (28S, 18S, 5.8S, 5S 12S, 16S) were used as baits for mouse and rat. Depletion efficiency was calculated by dividing matched reads by total number of reads passing instrument quality filtering.

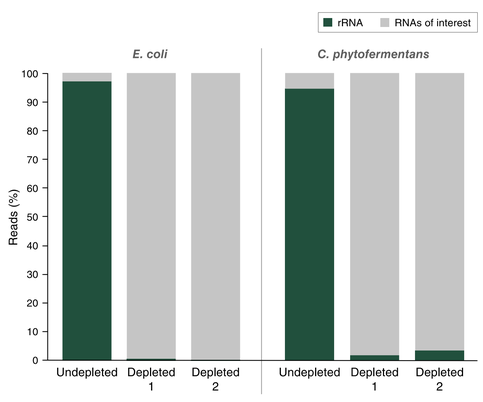
RESULTS: Bacterial rRNA Depletion

Efficient Removal of rRNA in Individual Bacterial Species and Mock Community

A. % rRNA Left in *Escherichia* species



B. % rRNA Left in *Clostridium* species



C. % rRNA Left in the ATCC 20 Mock Community

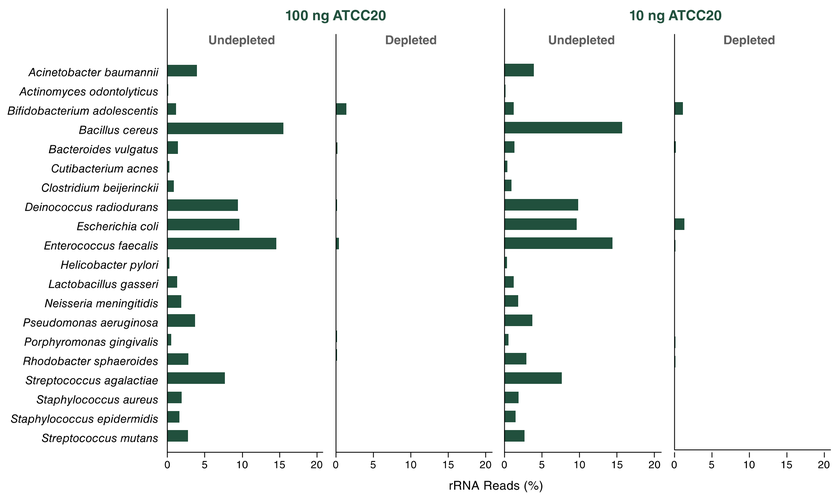


Figure 2. Highly efficient depletion of rRNA in bacteria
NEB bacterial depletion reagents achieved 90% or greater rRNA depletion for different species across individual samples or in a mock community of 20 different species. Similar depletion rates are observed for various species of A) *Escherichia coli*, B) *Clostridium phytofermentans.* C) Mock bacterial community.

Transcript Expression Correlation Is Maintained Across Inputs

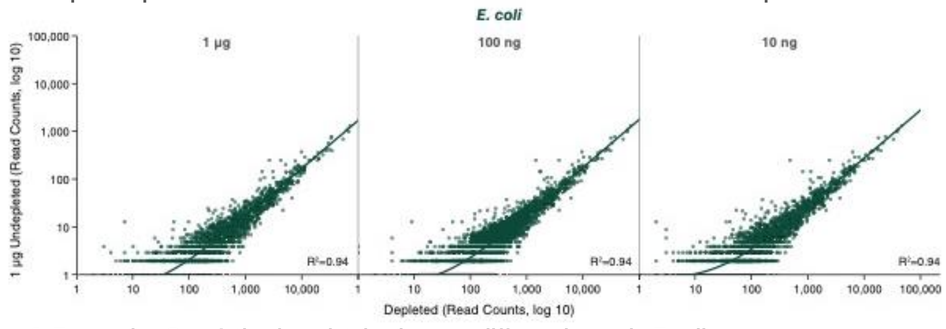


Figure 3. Expression Correlation is maintained across different inputs in *E.coli*
Depletion of rRNA with the NEBNext NEB rRNA depletion kit (Bacteria) does not affect transcript abundance. Consistent transcript expression (R2 > 0.94) was obtained across a wide input range in rRNA depleted samples compared to an undepleted library.

RESULTS: Globin and rRNA Depletion

Depletion of globin and ribosomal RNA enriches for RNAs of interest across species

Figure 4. Highly efficient depletion of globin and rRNA
The NEBNext Globin and rRNA Depletion Kit efficiently removes rRNA and globin RNA from human, mouse and rat total RNA enabling a higher percentage (>95%) of reads to map to RNAs of interest. Shown here is depletion on 1ug of Total RNA.

Consistent Depletion Across Species and Across Inputs

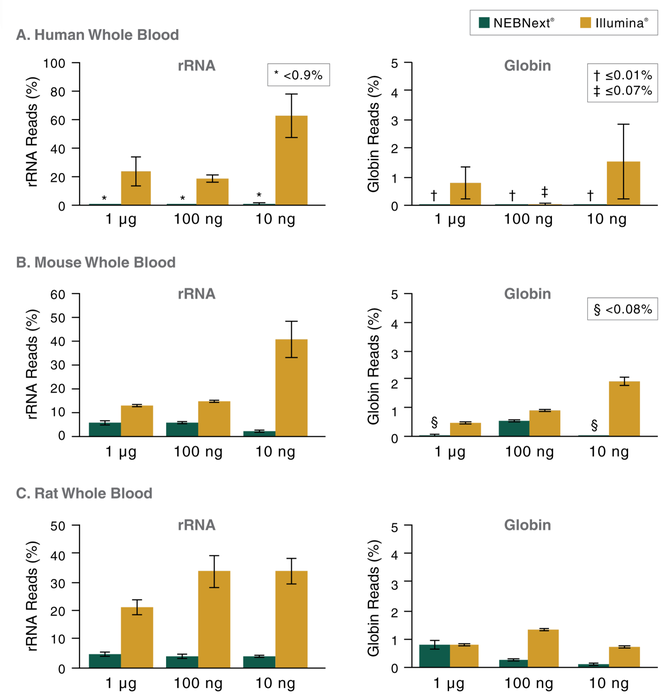


Figure 5: Superior Depletion Across Species and Across Inputs
The NEBNext Globin Depletion Kit is superior at depleting rRNA across species, and at depleting over 99% of globin mRNA. The data displays the % of reads mapping to rRNA and globin after depletion of rRNA and globin mRNA using the NEBNext Globin & rRNA Depletion Kit or Globin-Zero® Gold rRNA Depletion Kit (Illumina®). RNA-seq libraries were prepared from untreated and depleted RNA using the NEBNext Ultra™ II RNA Library Prep Kit for Illumina®. The data represents an average of 3 replicates and error bars indicate standard error. The NEBNext Globin Depletion kit is superior at depleting rRNA across species, and at depleting over 99% of globin mRNA.

Transcript Expression Correlation Is Maintained after Depletion and Across Inputs

A. Transcript correlation is maintained after depletion

B. Transcript correlation is maintained across inputs

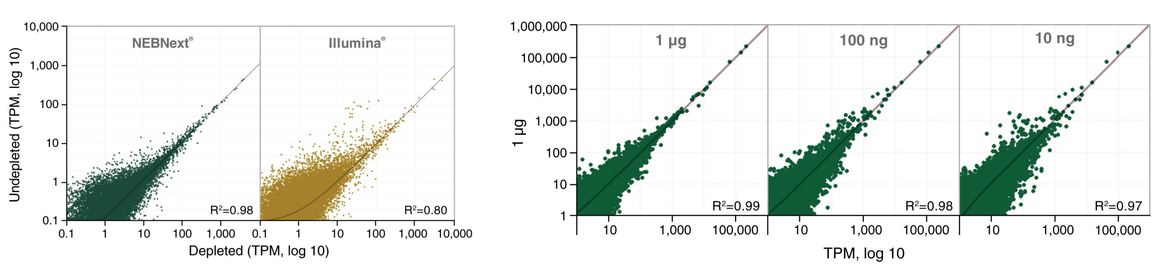


Figure 6. Expression Correlation is maintained after depletion and across different inputs
Depletion with the NEBNext Globin and rRNA Depletion Kit does not affect transcript abundance. A) High transcript expression correlation between depleted and undepleted samples (R^2>0.98) was observed for the NEBNext depletion kit. B) Consistent transcript expression (R^2> 0.97) was observed across a wide input range. For A and B, GENCODE v27 transcript abundances were estimated using Salmon [4]. TPM (Transcript per million mapped reads). Protein Coding transcripts shown.

RESULTS: rRNA Depletion V2

Efficient rRNA depletion across species and inputs amounts

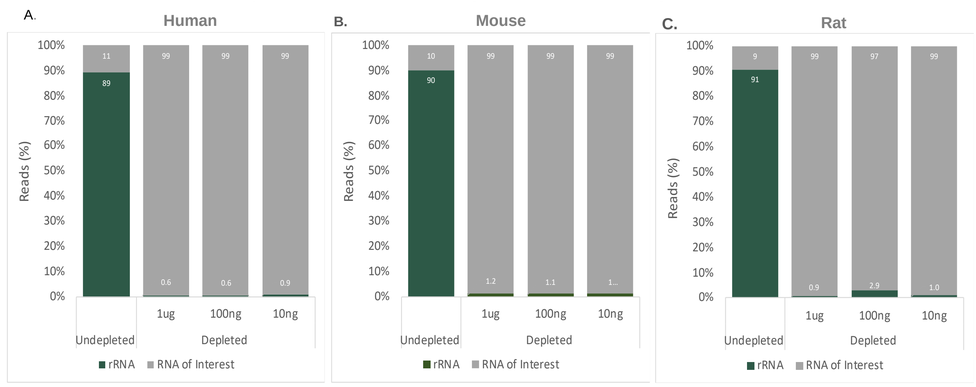


Figure 7. Efficient rRNA depletion enriches for RNAs of interest in human mouse and rat
Universal human (A), mouse (B), rat (C) reference total RNA samples (1 µg, 100 ng and 10 ng) depleted of rRNA using the NEBNext rRNA Depletion Kit V2 consistently achieves >97% rRNA depletion across species and a wide range of input amounts. The data represents an average of 3 replicates. (D) and (E) Human adult normal liver tissue FFPE Total RNA (100 ng and 10 ng) was depleted of rRNA using either the NEBNext rRNA Depletion Kit V2 or the depletion reagents in the TruSeq Stranded Total RNA Gold kit, followed by downstream library prep using Ultra II Directional RNA Library Prep kit. The NEBNext depletion kit is superior at depleting rRNA in FFPE (degraded) samples, and offers the flexibility of lower total RNA input amounts. Correlation analysis of the transcripts indicates consistent transcript expression after treatment.

CONCLUSIONS

- The NEBNext Depletion workflow is a quick and effective method to deplete unwanted RNAs for RNA-Seq
- The workflow is suitable for intact and degraded RNA
- The reagents are compatible with any downstream NGS library prep workflow, and amenable to automation
- Works for a broad input range: 10ng - 1ug total RNA
- Abundance of non-targeted transcripts is maintained after depletion and across a range of inputs
- Highly effective in Human, Mouse and Rat (Globin mRNA and rRNA Depletion Kit) and across diverse bacterial species, single cultures and complex communities (Bacterial rRNA Depletion)

REFERENCES

1. Langmead and Salzberg, *"Fast Gapped-Read Alignment with Bowtie 2."*
2. Quinlan and Hall, *"BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features."*
3. Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A. J., Müller, W. E. G., Wetter, T., & Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Research, 14(6), 1147–1159. http://doi.org/10.1101/gr.1917404
4. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2015). Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. bioRxiv, 21592. http://doi.org/10.1101/021592

ACKNOWLEDGEMENTS

1. NEB Sequencing core (Laurie Mazzola, Danielle Fuchs, Kirsten Augulewicz) for all their sequencing support
2. NEB Scientists for help with biological samples, eg. Bacterial RNA used in this study.