Authors

Laura de Jager
RED Scientist

Michael Berry
Bioinformatics Scientist

Luke Esau
RED Senior Scientist

Ross Wadsworth
RED Team Lead

Roche Sequencing Solutions
Cape Town, South Africa

Jennifer Wendt
Manager of Product Development

Dan Burgess
Director of Development

Roche Sequencing Solutions
Madison, WI, USA

Maryke Appel
Sr International Product Manager

Roche Sequencing Solutions
Pleasanton, CA, USA

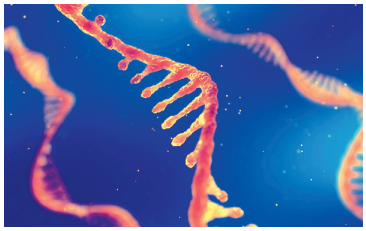
Date of first publication:
February 2018

Publication date of this version:
September 2018

KAPA RiboErase (HMR) Kits offer a flexible technology for selective transcript depletion prior to library construction for whole transcriptome analysis

RNA-Seq has enabled advances in translational and clinical research. However, the abundance of rRNA and highly expressed mRNAs pose a challenge to the sensitivity and economy of this technology. KAPA RiboErase (HMR) Kits support the efficient depletion of rRNA and globin transcripts from human, mouse, and rat samples in an automation-friendly workflow. The protocol is highly flexible and can be modified to deplete other transcripts of choice with user-supplied oligos.

Introduction



High-resolution transcriptome analysis using next-generation sequencing (RNA-Seq) can be used to study the levels and structure of both coding and non-coding RNA, and is now routinely utilized in life sciences, translational, and clinical research. A challenge to the economy and sensitivity of RNA-Seq lies in the fact that ribosomal RNA (rRNA) constitutes up to 80% of the total RNA in a cell.[1] Library preparation protocols for Illumina sequencing systems therefore provide for the removal (depletion) of these "unwanted" transcripts from RNA samples prior to cDNA synthesis, or enrichment of cDNA libraries for sequences of interest using hybridization-based capture.

The KAPA RiboErase (HMR) Kit is designed for the depletion of cytoplasmic and mitochondrial transcripts from human, mouse, and rat samples. Kits contain DNA oligos that are complementary to rRNA transcripts. After hybridization, RNA:DNA hybrids are enzymatically removed using RNase H. The 2.5-hour workflow is automation-friendly and the technology is inherently flexible. Recently, the KAPA RiboErase (HMR) family of kits was expanded to provide options for the depletion of globin transcripts from blood samples, which can account for 50% to 80% of residual transcripts after rRNA depletion.[2]

The depletion of highly abundant mRNA transcripts further improves the sensitivity and economy of RNA-Seq. Fewer sequencing reads are associated with unwanted transcripts, which improves the detection of rare transcripts, or those with subtle changes in expression levels between cell types or biologial conditions. The overall number of reads required to meet experimental objectives may, consequently, also be reduced.

In this study, we demonstrate that the KAPA RNA HyperPrep Kit with RiboErase (HMR) may be used to efficiently deplete transcripts of choice by using user-supplied oligos. User-defined, custom oligos are conveniently added with minimal changes to the workflow, maintaining the single-day turnaround time from RNA sample to sequencing-ready library. As previously shown for commercially available rRNA and globin oligos, negligible off-target depletion was observed.[3] An associated Application Note4 further demonstrates the utility of the technology for the depletion of non-human transcripts.

Experimental design and methods

Depletion targets

This study was designed to demonstrate the utility of KAPA RiboErase (HMR) kits for the depletion of highly abundant human transcripts, as well as non-human content. To this end, a set of custom oligos were designed to target the twelve most highly expressed transcripts in human blood (other than globin). In addition, oligos complementary to eight of the 92 transcripts in the Ambion® ERCC Spike-In Controls Mix 1 (ThermoFisher Scientific[5]) were included. A summary of the targets for this "RiboErase Plus" oligo set is given in Table 1.

Table 1. "RiboErase Plus" target list

Human blood transcripts[a]	ERCC Mix 1 transcripts ID (subgroup; concentration)[b]
RN7SL1-201	00130 (A; 30,000)
RN7SL2-201	00096 (B; 15,000)
RN7SL4P-201	00074 (C; 15,000)
RN7SL5P-201	00002 (D; 15,000)
RN7SK-201	00004 (A; 7,500)
RPPH1	00171 (B; 3,750)
SNORD3A	00113 (C; 3,750)
B2M-207	00046 (D; 3,750)
EEF1A1	
MT-CO1-201	
MT-CO2-201	
MT-CO3-201	

a. Identified from previous sequencing data.

b. ERCC Control Mixes are pre-formulated blends of unlabeled, polyadenylated transcripts, derived from plasmids certified by the National Institute of Standards and Technology (NIST). Mixes are added to RNA expression analysis experiments after sample isolation, to enable assessment of defined performance standards. See **Appendix** for more details.

RNA depletion and library construction workflow

The full workflow, from input RNA to sequencing-ready library, is depicted in Figure 1. RNA depletion was performed with the KAPA RiboErase (HMR) Kit, whereas the KAPA RNA HyperPrep Kit was used for cDNA synthesis and library construction.

The entire process takes approximately 6.5 hours to complete, and all steps and reagents are automation-friendly.

For this study, triplicate libraries were prepared from either 25 ng and 100 ng inputs of total RNA from human blood (BioChain), supplemented with Ambion ERCC RNA Spike-In Control Mix 1 to a final concentration of 0.25% (m/m).

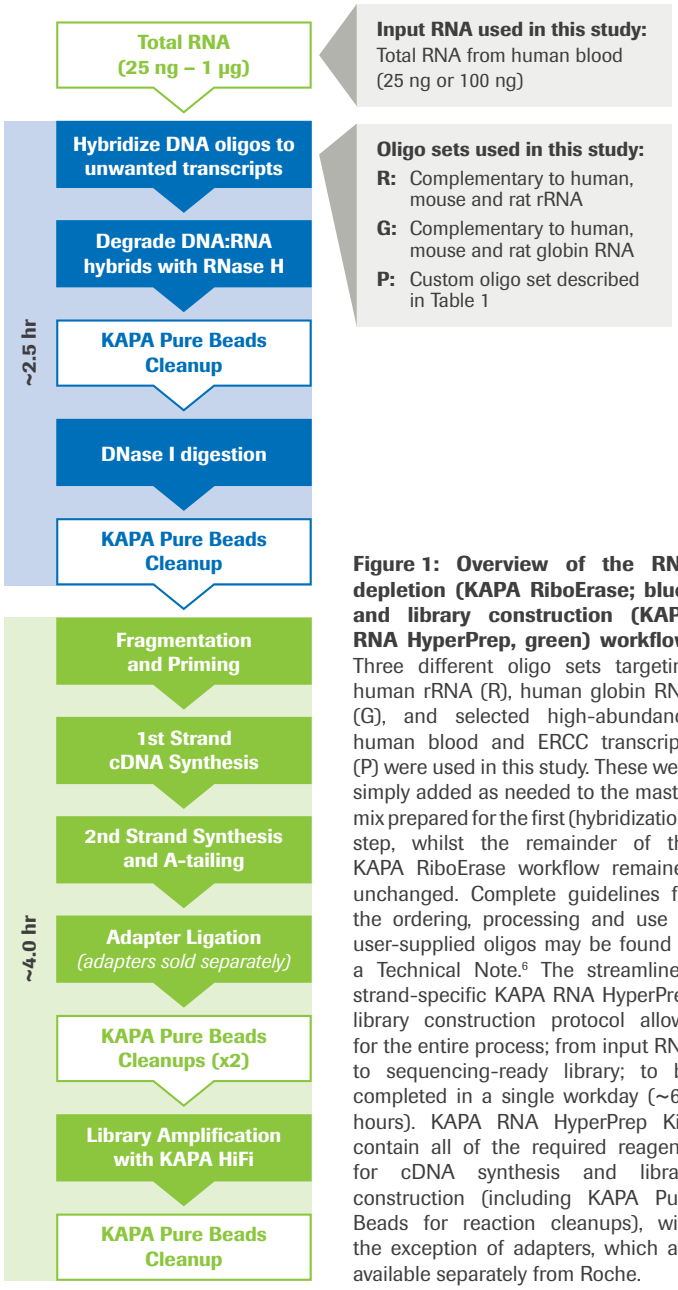


Figure 1: Overview of the RNA depletion (KAPA RiboErase; blue) and library construction (KAPA RNA HyperPrep, green) workflow. Three different oligo sets targeting human rRNA (R), human globin RNA (G), and selected high-abundance human blood and ERCC transcripts (P) were used in this study. These were simply added as needed to the master mix prepared for the first (hybridization) step, whilst the remainder of the KAPA RiboErase workflow remained unchanged. Complete guidelines for the ordering, processing and use of user-supplied oligos may be found in a Technical Note.[6] The streamlined, strand-specific KAPA RNA HyperPrep library construction protocol allows for the entire process; from input RNA to sequencing-ready library; to be completed in a single workday (~6.5 hours). KAPA RNA HyperPrep Kits contain all of the required reagents for cDNA synthesis and library construction (including KAPA Pure Beads for reaction cleanups), with the exception of adapters, which are available separately from Roche.

Three oligo sets were used for different experiments:

- HMR (R); the standard KAPA RiboErase (HMR) oligo set, targeting human, mouse and rat cytoplasmic (5S, 5.8S, 18S, and 28S) and mitochondrial (12S and 16S) rRNA transcripts;
- Globin (G); the KAPA RiboErase (HMR) oligo set targeting human, mouse and rat globin transcripts, and
- Plus (P); the custom oligo set described in Table 1.

One, two, or all oligo sets were included in the hybridation step of the depletion workflow, depending on experimental objectives.

Sequencing and data analysis

Libraries were quantified using the KAPA Library Quantification Kit, normalized and pooled, and sequenced on an Illumina® NextSeq® 500 instrument (≥24 million reads per library). Adapter sequences and low-quality bases were trimmed by cutadapt prior to splice-aware alignment to known rRNA sequences with HISAT2, to identify reads aligning to residual rRNA. Remaining reads were first mapped to Globin or "Plus" targets, and then aligned to the GRCh38 human and ERCC reference sequences.

Average transcript expression levels were calculated from trimmed FASTQ files as Transcripts Per Million (TPM) using Kallisto.

Results

Targeted transcripts are effectively removed

Depletion efficiency was evaluated in three different ways. Firstly, the number of reads aligning to different transcript classes were quantified, when Globin and/or "Plus" oligo sets were added to the standard KAPA RiboErase (HMR) workflow (Figure 2).

Both oligo sets were shown to result in highly efficient depletion of target transcripts, with a concomitant increase in the percentage of reads associated with transcripts of interest.

Secondly, the percentage reduction in transcripts per kilobase million (TPM) for the twelve highly expressed human blood transcripts was calculated by comparing reads assigned to each transcript when only the rRNA oligos vs. the rRNA and "Plus" oligo sets were used (Table 2). The reduction rate varied from 80% to 100%, with an average of 96% across the twelve targets for both inputs. The average depletion efficiency of 88% for the three mitochondrial targets (MT-CO1-201, MT-CO2-201, and MT-CO3-201) was lower than the same for the nine nuclear transcripts (99%). This was attributed to suboptimal oligo design. Design optimization was not attempted, but is envisaged to improve results when specific transcripts are not sufficiently depleted with an initial design.

Finally, visual inspection of reads aligned to regions of the "Plus" oligo targets (performed with the Integrative Genome Viewer, IGV) confirmed that targeted exonic regions of the twelve highly expressed transcripts were very effectively depleted. A representative example is given in Figure 3.

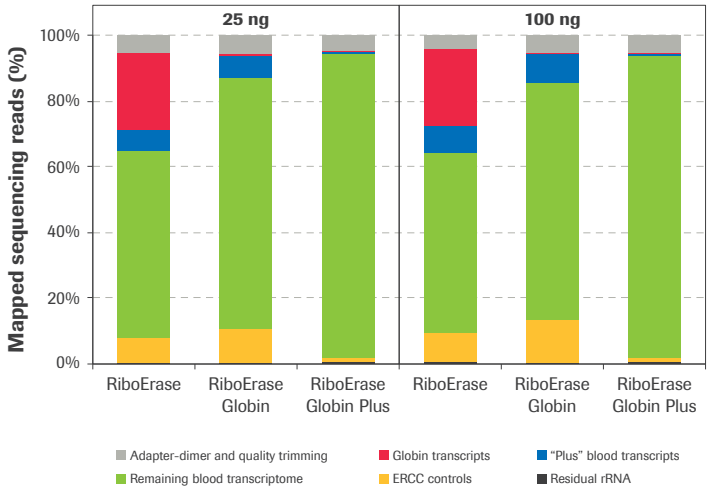


Figure 2: A higher number of reads map to transcripts of interest after selective depletion. The average percentage of sequencing reads that map to transcripts not targeted for depletion significantly increased as the Globin and "Plus" targets were selectively removed from RNA samples prior to library construction. The standard KAPA RiboErase (HMR) Kit removed the majority of cytoplasmic (5S, 5.8S, 18S, and 28S), and mitochondrial (12S and 16S) rRNA, with an average of 0.3% sequencing reads mapping to residual rRNA across all samples. The addition of globin depletion oligos significantly increased the alignment of sequencing reads to the remainder of the transcriptome. Further improvements were achieved when the custom depletion oligos were added. Comparable results were obtained for both inputs into the RNA depletion workflow (25 ng and 100 ng).

Table 2. Reduction in target gene TPM for highly expressed human blood transcripts targeted with the "Plus" oligo set

Target	25 ng input	100 ng input
RN7SL1-201	99.8%	99.8%
RN7SL2-201	99.7%	99.7%
RN7SL4P-201	99.6%	99.4%
RN7SL5P-201	96.3%	94.8%
RN7SK-201	99.9%	100.0%
RPPH1	99.9%	99.9%
SNORD3A	99.3%	99.4%
B2M-207	96.1%	95.8%
EEF1A1	98.9%	98.9%
MT-CO1-201	91.0%	90.7%
MT-CO2-201	91.8%	92.0%
MT-CO3-201	80.6%	80.6%
Average	96.1%	95.9%

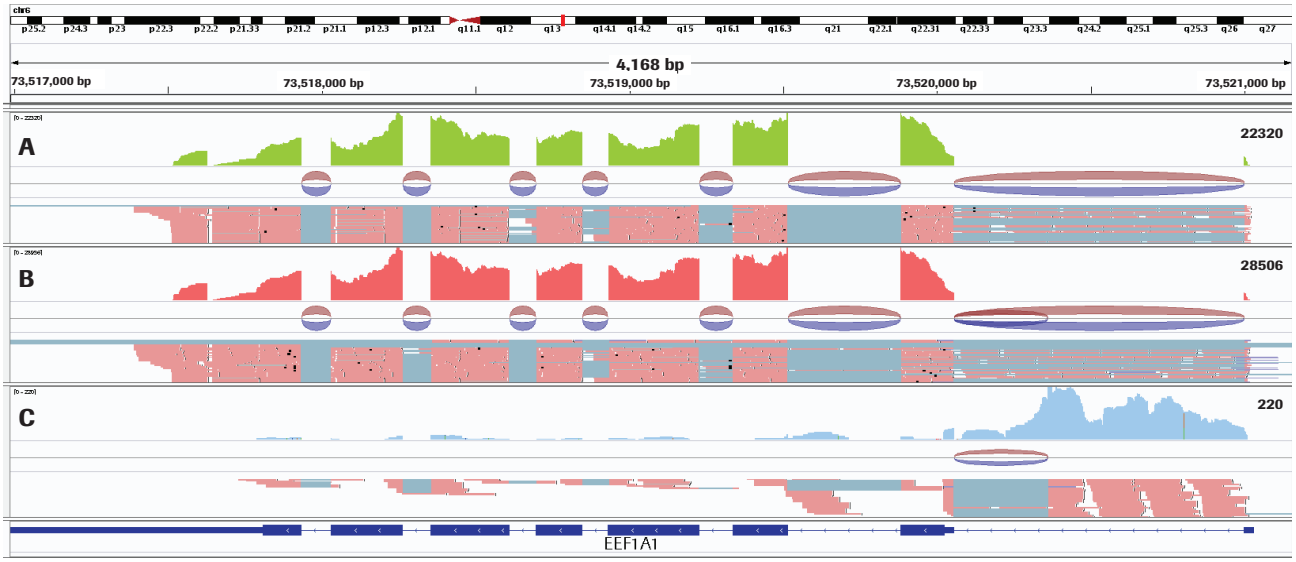


Figure 3. IGV screenshot showing reads from 25 ng KAPA RiboErase (A), RiboErase Globin (B), and RiboErase Globin Plus (C) libraries, aligned to the EEF1A1 gene sequence. The RiboErase Globin Plus data shows very few reads mapping to the targeted exonic regions of EEF1A1, confirming the efficiency of depletion. The majority of remaining reads are located in intronic regions, which were not targeted.

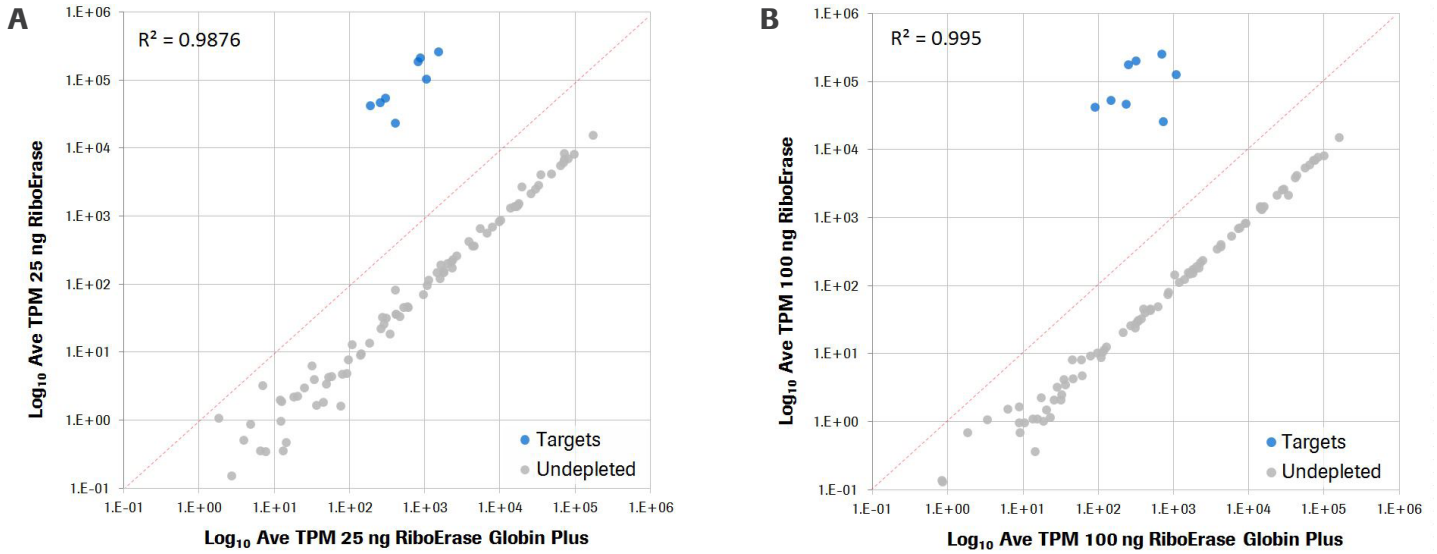


Figure 4. Expression correlation plots for ERCC Mix 1 transcripts, for the standard KAPA RiboErase (HMR) workflow (y-axis) vs. the RiboErase Globin Plus workflow (x-axis). Results for libraries constructed from 25 ng **(A)** and 100 ng **(B)** RNA (input into the depletion reaction) are given. Targeted ERCC Mix 1 transcripts (blue) were selectively depleted from the RiboErase Globin Plus libraries, with no evident off-target depletion.

(C) Selective depletion of the eight targeted ERCC transcripts resulted in a significant increase in the number of ERCC transcripts with a TPM >5. Of the total number of 92 ERCC transcripts, (which span a 10^6-fold concentration range[5]), 60 – 70 (colored bars) were detected with the RiboErase and RiboErase Globin workflows, when spiked into the input RNA at the recommended concentration. The remaining 20 – 30 transcripts (grey bars) were either not detected, or returned a TPM <5 from the given amount of sequencing. The RiboErase Globin Plus workflow allowed for the detection of 15 – 20 additional ERCC transcripts with a TPM >5, for both RNA input amounts.

Data on file.
For Research Use Only. Not for use in diagnostic procedures.

Depletion of unwanted transcripts improves the detection of low-abundance genes

To confirm whether selective depletion resulted in improved detection of low-level transcripts, average transcript expression levels (expressed in transcripts per kilobase million, TPM) were calculated from trimmed fastq files using Kallisto. Three types of expression correlation plots comparing the standard KAPA RiboErase (HMR), RiboErase Globin, and/or RiboErase Globin Plus workflows are given in Figures 4 – 6.

Figure 4 focuses on the ERCC Mix 1 transcripts only. Correlation plots (A and B) for reactions that contained the standard, commercial RiboErase (HMR) oligo set (y-axis) vs. the RiboErase Globin Plus oligo set (which includes oligos complementary to globin transcripts, the additional highly expressed blood transcripts and the eight ERCC transcripts; x-axis), clearly show that the eight targeted ERCC transcripts were efficiently depleted. No impact was observed on the 84 ERCC transcripts that were not targeted. Removal of those eight transcripts (that include some of the most abundant transcripts in ERCC Mix 15) meant that more sequencing reads were available to detect the ERCC transcripts present in Mix 1 at very low concentrations. With the amount of sequencing performed, an additional 15 – 20 lower-abundance ERCC Mix 1 transcripts returned TPM values >5, as opposed to being undetectable, or returning TPM values ≤5 without the custom oligos (Figure 4C).

In contrast, Figure 5 focuses on the blood transcriptome. In this case, correlation plots for the standard RiboErase (HMR) workflow vs. the RiboErase (HMR) Globin workflow (including commercially available rRNA and globin-targeting oligos; A and B), as well as the standard workflow vs. the RiboErase (HMR)

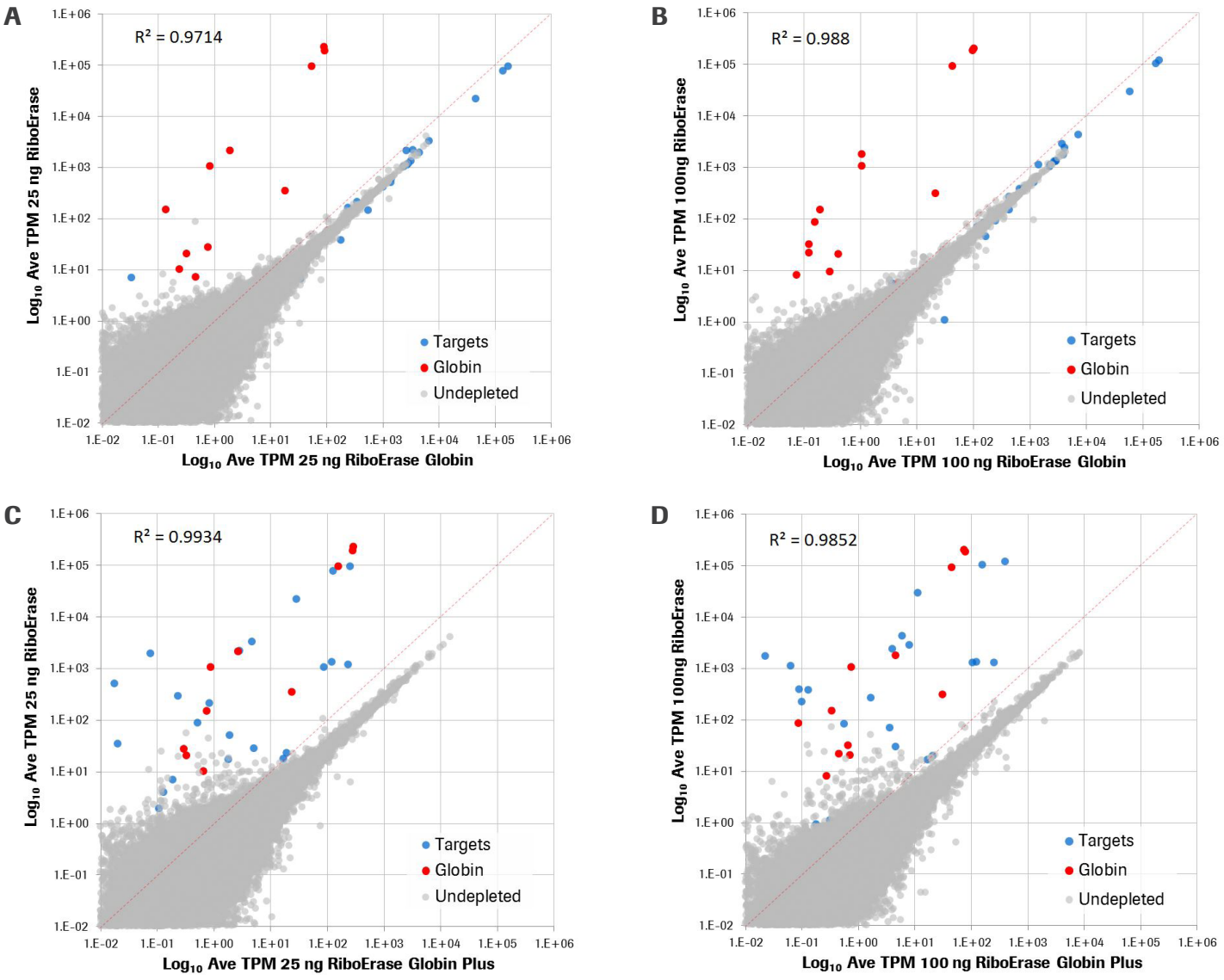


Figure 5. Expression correlation plots for the human blood transcriptome, for the standard KAPA RiboErase (HMR) workflow, vs. the RiboErase Globin workflow (A, 25 ng input and B, 100 ng input), or RiboErase Globin Plus workflow (C, 25 ng input and D, 100 ng input). TPM values for transcripts mapping to the human reference sequence shows that the majority of globin and targeted high-abundance blood transcripts were effectively depleted. The average TPM of the remaining human blood transcripts increased significantly, as seen by a shift to the right of the red trend line, indicating that more reads are mapping to untargeted transcripts.

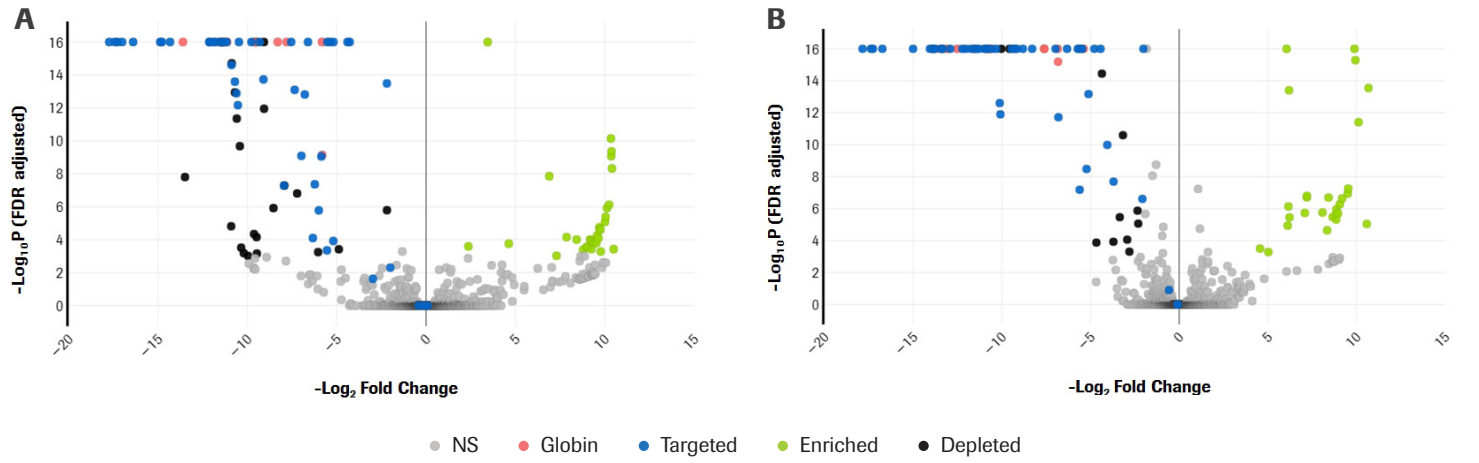


Figure 6. Differential expression analysis for the standard RiboErase vs. RiboErase Globin Plus workflows. Volcano plots for both the **(A)** 25 ng and **(B)** 100 ng inputs are shown. The analysis was done with edgeR, with significance levels set at q >0.001. Each dot represents one transcript that was detected in both workflows, and shows the fold change and the significance thereof. Transcripts that displayed no significant change in expression between the workflows are labeled as not significant (NS, grey). For both inputs, effective depletion of the globin (red) and additional targeted blood transcripts (blue) was achieved with the RiboErase Globin Plus workflow. As shown before, depletion of highly abundant transcripts translated to the enrichment of lower-abundance transcripts (green). Some degree of off-target depletion was detected, but appeared to be stochastic, as impacted transcripts (black) varied between inputs and technical replicates.

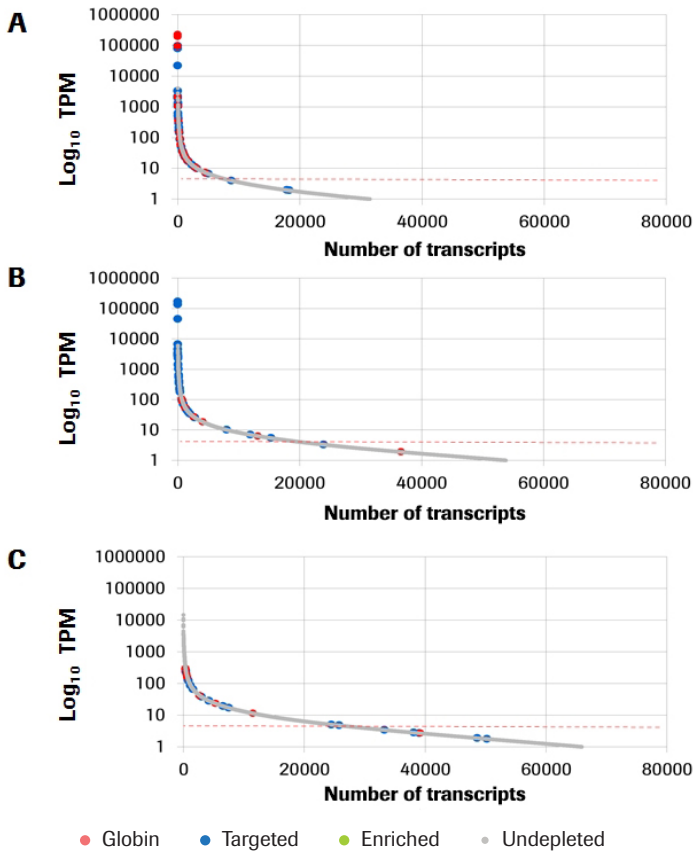


Figure 7. Number of transcripts with a TPM >5 detected in libraries constructed from 25 ng human blood total RNA, using the standard KAPA RiboErase (HMR)workflow (A), vs. the RiboErase Globin (B) and RiboErase Globin Plus (C) workflows. The red dotted line represents a TPM = 5. In the standard workflow, globin (red) and other targeted blood transcripts (blue) dominate. Effective depletion with the Globin and "Plus" oligos not only significantly increases the number of transcripts with a TPM >5, but also the relative abundance of many of these in the final library.

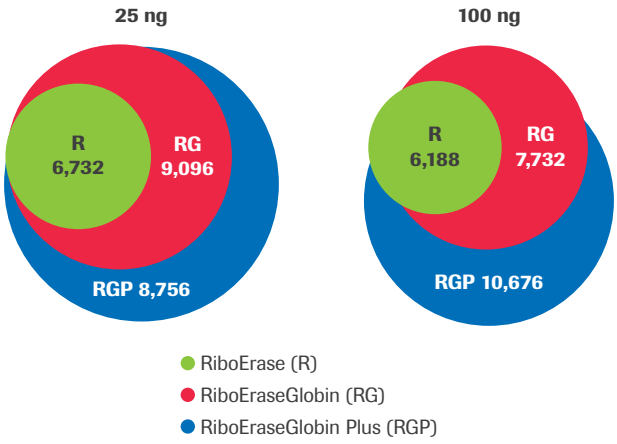


Figure 8. Venn diagram representation of gene expression patterns between workflows. Venn diagrams illustrate the overlap of transcripts detected with the standard KAPA RiboErase (HMR) workflow (R), vs. the RiboErase Globin (RG) and RiboErase Globin Plus (RGP) workflows. Depletion of highly abundant globin and other blood transcripts increased the total number of transcripts detected (TPM >5) by almost 4-fold for both inputs.

Globin Plus workflow (rRNA and globin plus all custom oligos; C and D) are given for both inputs. Plots A and B show that the globin transcripts (red dots) are efficiently depleted when globin oligos were added, with little or no impact on the expression levels of the other targeted blood transcripts (blue dots). Those twelve transcripts (which were confirmed in plots A and B to be highly abundant), were effectively depleted (plots C and D) when the "Plus" oligos were added to the rRNA and globin oligos during setup of the hybridization mix (ref. Figure 1). Because the Globin and Globin Plus workflows target highly abundant transcripts, removal of these transcripts left more sequencing reads for the detection of the rest of the blood transcriptome. In both experiments, the average TPM for untargeted transcripts increased, as shown by a shift to the right of the red trendline.

For Figure 6, the blood transcriptome was analyzed in a different way. The "Volcano plots" shown in this figure depict the fold change in the expression for each transcript detected in both the standard and the Globin Plus workflows. Three groups of transcripts were identified:

- Transcripts displaying **no significant change** in expression level between the two workflows (grey dots).
- Transcripts with **reduced expression levels in the RiboErase Globin Plus workflow** (red, blue, and black dots). A very high fold reduction in expression levels was observed for all of the globin transcripts (red dots), and most of the transcripts targeted with the "Plus" oligos (blue dots). In this group, black dots represent transcripts that were not specifically targeted with the Globin or "Plus" oligos, and indicate off-target depletion. Upon further interrogation, these events appeared to be stochastic, as the identity of the impacted transcripts varied between inputs and technical replicates. Furthermore, some of these were found to be short RNAs (<400 bp), which may have been lost during library construction (as opposed to depletion). For these reasons, off-target depletion events were regarded to be of low significance.
- Transcripts that were enriched in the **RiboErase Globin Plus workflow** (green dots). The identity of these transcripts also differed between the two inputs, indicating that their enrichment was a non-specific consequence of having more reads available to detect lower-abundance transcripts.

To further quantify transcript enrichment when the Globin and "Plus" oligos were included in the hybridization step, the total number of transcripts with a TPM >5 was determined (Figure 7, 25 ng input only). The Venn diagrams in Figure 8 illustrate the overlap of transcripts detected with the three workflows for both input amounts. When only rRNA was depleted (RiboErase workflow; R), an average of approximately 6,500 transcripts with a TPM >5 was observed. Merely removing the globin transcripts (RiboErase Globin workflow; RG) more than doubled this number for both inputs, whereas depletion of the other highly abundant blood transcripts yielded an additional 55% and 75% transcripts with a TPM >5 for the 25 ng and 100 ng inputs, respectively.

Conclusions

KAPA RNA HyperPrep Kits with RiboErase (HMR) offer a flexible, single-day library preparation solution for whole-transcriptome sequencing on Illumina® platforms. Commercially available kits provide oligos for the depletion of human, mouse, and rat rRNA or rRNA and globin transcripts.

In this study, we have shown that user-supplied oligos may easily be added to the standard workflow to selectively deplete any additional transcripts of choice. Using oligos targeting the twelve most abundant transcripts in human blood (other than globin), as well as oligos complementary to artificial, experimental control transcripts, we have demonstrated effective and specific depletion of targeted transcripts with reproducible results across two input amounts (25 ng and 100 ng RNA into the depletion reaction). An accompanying study[4] has confirmed that efficient, selective depletion is not dependent on the presence of the commercially available human, mouse, and rat rRNA oligos. These may be entirely replaced with customized oligo mixes, to allow for the selective depletion of user-defined content from the total RNA of any species.

When highly expressed transcripts are depleted prior to the construction of cDNA libraries, significantly less sequencing is wasted on RNA species that often have a low information value. As a result:

- more sequencing reads are associated with lower-abundance RNA species, which are typically of higher biological interest;
- more data (e.g., higher TPM values) for lower-abundance transcripts improve confidence in results;
- more subtle fold changes in gene expression levels between biological or experimental conditions can be detected; and/or
- the amount of sequencing per sample required to achieve experimental objectives may be reduced.

In addition to potential improvements in sequencing economy, and robust performance in RNA depletion workflows, KAPA RNA HyperPrep Kits with RiboErase (HMR) offer:

- the convenience of a streamlined, automation-friendly workflow, from input RNA to sequencing-ready library, which can be completed in a standard workday.[3]
- the ability to collect QC data at several points in the workflow, to inform key decisions about library construction parameters, and downstream sequencing.[6]
- the flexibility to fine-tune key library construction parameters, to improve outcomes for low-quality samples.[7]
- integrated service and support for a complete workflow solution that includes adapters and qPCR-based library quantification of final libraries.

References

1. Lodish H, Berk A, Zipursky SL, et al. *Molecular Cell Biology*. 4th edition. New York: W. H. Freeman; 2000. Section 11.6, Processing of rRNA and tRNA. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21729/. Accessed August 2018.
2. Krjutškov, K, Koel M, Roost A-M et al. Globin mRNA reduction for whole-blood transcriptome sequencing. *Scientific Reports* 2016;6:31584.doi: 10.1038/srep31584
3. Roche. Roche Sample Prep Solutions for RNA-Seq. 2018. Accessed August 2018.
4. Roche. Application Note: High-efficiency species-specific ribosomal RNA depletion with the KAPA RNA HyperPrep Kit. 2018. Accessed December 2018.
5. Ambion/Life Technologies Corporation. ERCC RNA Spike-In Control Mixes User Guide, rev. D. 2012. Accessed August 2018.
6. Roche. Sequencing Solutions Technical Note: Selective RNA transcript depletion using KAPA RNA HyperPrep Kit with RiboErase and customized DNA oligos. 2018. Accessed September 2018.
7. Roche. Sequencing Solutions Technical Note: How To … Prepare libraries from degraded RNA inputs with the KAPA RNA HyperPrep Kit with RiboErase (HMR) for whole transcriptome sequencing. 2018. Accessed May 2018.

Appendix: Depletion oligos for targeted ERCC Mix 1 transcripts

As indicated in **Experimental design and methods**, ERCC Control Mixes (ThermoFisher Scientific) are easily added to input RNA prior to library construction for RNA-Seq, and enable the assessment of defined performance standards.[5] Similarly, the oligos used in this study (to selectively deplete 12 of the 84 transcripts in ERCC Control Mix 1) may be included as a control in any custom depletion experiment. Comparable results to those documented here should be obtained, provided that similar experimental conditions and data analysis pipelines are employed.

The sequences of the twelve targeted ERCC Control Mix 1 transcripts listed in Table 1 on p. 2 may be obtained from the supplier (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_095047.txt; accessed August 2018). Depletion oligos were designed, ordered, processed and utilized as described in Sections 5 – 8 of the Technical Note entitled Selective RNA transcript depletion using KAPA RNA HyperPrep Kit with RiboErase and customized, user-supplied DNA oligonucleotides.[6] The Technical Note, as well as an accompanying file containing the depletion oligo sequences, is available from your local Roche technical support team or **technical-support.roche.com**.

Published by:

Roche Sequencing and Life Science
9115 Hague Road
Indianapolis, IN 46256

sequencing.roche.com

MC-US-06587 APP111005 A546 4/20