Authors

Ranjit Kumar
Senior Bioinformatics Scientist

Nicholas Hapshe
Support Scientist

Brian Sogoloff
Support Scientist

Lindsey Cambria
Sample Prep Support Manager

Rachel Kasinskas
Director, Support and Applications

Sequencing and Life Sciences,
Roche Diagnostics Corporation,
Wilmington, MA, USA

Eric Almeida
Graduate Research Assistant

Thais Harder de Palma
Graduate Research Assistant

Dasith Perera
Graduate Research Assistant

Matthew Ramsey
Assistant Professor

Kathryn Ramsey
Assistant Professor

Department of Cell and
Molecular Biology,
University of Rhode Island,
RI, USA

KAPA RiboDesigner—A custom probe design solution for rRNA depletion from single- and multi-species bacterial samples

Ribosomal RNA (rRNA) accounts for the vast majority of total cellular RNA. Efficient depletion of rRNA is thus a critical step in RNA sequencing for maximizing transcript coverage and cost effectiveness. While the KAPA RNA HyperPrep Kit with RiboErase (HMR) is a highly efficient method for the enzymatic depletion of rRNA in human, mouse, and rat samples, this kit can also be expanded for use with other species by designing and including new probes complementary to the target rRNA sequences. However, the manual probe design process is complex and requires extensive bioinformatics expertise. KAPA RiboDesigner is a new custom probe design solution that eliminates the need for manual probe design and expands the ribodepletion capabilities of the KAPA RNA HyperPrep Kit with RiboErase (HMR). This solution automatically designs rRNA depletion probes for any bacterial species when provided with its annotated genome sequence. In addition, KAPA RiboDesigner can simultaneously design non-redundant probes targeting all rRNAs in multi-species mixtures, thus reducing the total number of probes needed. The effectiveness of depletion probes designed by KAPA RiboDesigner was tested using total RNA from several individual species as well as RNA from co-cultured samples and a manually created mixture of RNA from eight bacterial species. In every case, incorporation of these probes into the library preparation workflow in place of the supplied HMR oligos yielded an rRNA depletion efficiency of >98%. Thus, KAPA RiboDesigner greatly expands rRNA depletion capabilities to a vast number of sequenced microbial species.

Introduction

The high amount of rRNA present in total cellular RNA presents several challenges to efficient, accurate transcriptome analysis—including high sequencing costs, decreased coverage of coding RNAs, and difficulty detecting low-abundance transcripts. In human, mouse, and rat species, the rRNA abundance ranges from 70 – 90% of the total RNA; in many bacterial species, rRNA can account for over 90% of the total RNA (Figure 1). To maximize the amount of relevant data obtained from each sample and to increase sequencing efficiency, enrichment of target RNAs prior to library preparation is frequently performed via either rRNA depletion or mRNA enrichment.

The KAPA RNA HyperPrep Kit with RiboErase (HMR) is a robust, reliable RNA-seq library preparation kit that includes highly efficient rRNA depletion of human, mouse, and rat rRNA. This kit can easily be adapted for depletion of bacterial rRNA by replacing the provided HMR oligos with custom-designed oligos that target the appropriate bacterial rRNA (Hapshe, et al., 2018).

Although the substitution of these probes for the HMR probes into the workflow is very easy, the actual probe design process is complex. Manual probe design requires extensive bioinformatics expertise, especially as numerous probes are required for each rRNA gene, and many bacterial species contain multiple copies of each rRNA gene (5S, 16S, and 23S). Furthermore, additional expertise is required to minimize off-target depletion.

KAPA RiboDesigner is a new custom probe design solution that eliminates the need for time-consuming manual probe design for bacterial rRNA depletion. This highly reliable and scalable bioinformatics solution automatically designs high-quality probes using the target organism's annotated genome sequence (provided by the user), targeting all rRNA sequences in that organism. The algorithm is designed to minimize probe redundancy by considering the similarity between rRNA genes, reducing the total number of probes needed for effective depletion of all rRNA copies. In addition, KAPA RiboDesigner can be used to simultaneously design probes for combinations of bacterial genomes, reducing the total number of probes needed for multi-species mixtures.

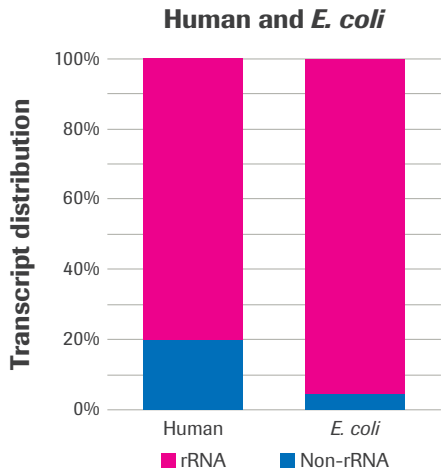

Figure 1. Typical rRNA content in human and *E. coli* total RNA. Very high levels of rRNA are present in the total RNA from humans (~80%) and *E. coli* (~95%).

Probes designed using KAPA RiboDesigner and incorporated into the KAPA RNA HyperPrep Kit with RiboErase (HMR) were tested on three types of bacterial RNA samples: RNA from individual bacterial species; RNA isolated from co-cultures of two different species; and a manually-created mixture of RNA derived from eight individual bacterial species (referred to as the "8-species mix"). The custom probes yielded highly efficient rRNA depletion from all three sample types.

Experimental design and methods

KAPA RiboDesigner Probe design workflow

KAPA RiboDesigner is a bioinformatics solution that designs probes for rRNA depletion for a given microbial species using its annotated genome files (available through NCBI).

The probe design—or probe selection—process consists of two major steps described below. The steps are applied, one at a time, to each rRNA sequence (Figure 2).

The first step, "unique sequence markup," identifies the regions within each rRNA that are not covered by any existing probe in the probe database. The probe database is empty in the first iteration, so the full sequence of the first rRNA is marked as unique for the probe design process. In subsequent iterations, the process uses BLAST algorithm (with similarity threshold ~90%) to map all probes from the probe database against the current rRNA sequence to mark the unmapped—or unique—regions. These marked regions become the new target for probe design. This step is crucial to avoid redundancy in probes as each bacterium has several copies of rRNA sequences with high sequence similarity. This process also enables the user to scale up probe design for multi-species mixes, in which there could be very high redundancy among rRNA sequences.

The second step, "optimal probe selection," uses a sliding window approach to scan each unique marked region defined above, one at a time. The sliding window moves from the beginning to the end of the marked region and selects an optimal probe from a set of possible probes based on features such as GC content, off-target effects, and low complexity in that given window. After an optimal probe is selected, the window moves along the rRNA sequence to select the next possible probe. This approach provides a dynamic way to select optimal probes, improving overall probe pool quality. At the end of this step, all selected probes are added to the probe database. The process then repeats for each rRNA sequence, and the final collection of sequences is used to order custom probe pools (Figure 2).

The KAPA RiboDesigner algorithm also provides the flexibility to modify several design parameters including probe length, gap length, complementarity threshold, etc. via a configuration file. Therefore, a set of user-defined parameters and filters can be applied while designing ribodepletion probes.

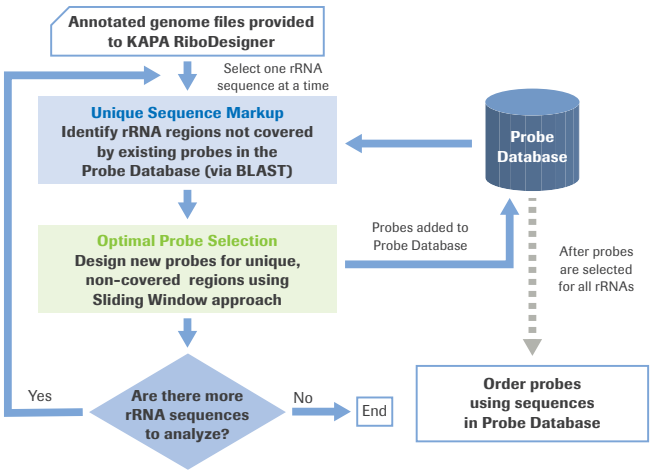

Figure 2: KAPA RiboDesigner algorithm workflow. This workflow describes the steps carried out by the KAPA RiboDesigner algorithm as it selects optimal ribodepletion probes for each rRNA gene.

RNA-seq library preparation workflow

High-quality *E. coli* total RNA was purchased from Thermo Fisher Scientific (AM7940). Total RNA from eight additional bacterial species and two co-cultured samples was provided by collaborators at the University of Rhode Island (Table 1).

KAPA RiboDesigner was used to design rRNA depletion probes for three types of total RNA samples: RNA from individual species; RNA isolated from co-cultures of two different species; and a manually created 8-species mix. The probes for the co-cultured samples were designed and produced for each species separately and then combined prior to ribodepletion. The probes for the 8-species mix were designed simultaneously by KAPA RiboDesigner using a concatenated set of annotated genomes from all eight species.

Each set of custom probes was ordered through IDT and received as a lyophilized pre-pooled "oPool Oligo Pool." Probe pools were resuspended at a concentration of 1 µg/µL (confirmed by NanoDrop); this is the same concentration as the HMR oligos included in the KAPA RNA HyperPrep Kit with RiboErase (HMR), which will be replaced by the custom oligos in the ribodepletion workflow (Figure 3).

Triplicate ribodepleted RNA-seq libraries were prepared for each RNA sample, using 50 ng of high-quality total RNA as input. For custom ribodepletion, the KAPA RiboErase workflow was followed according to the **KAPA RNA HyperPrep Kit with RiboErase (HMR) Technical Data Sheet** except that the custom probes were used in place of the provided HMR oligos; an overview of the RiboErase workflow is depicted in Figure 3. Following depletion, RNA was fragmented at 94°C for 6 minutes. KAPA Unique Dual-Indexed (UDI) Adapters were used at 1.5 µM in the adapter ligation step, and libraries were amplified by PCR (14 cycles) using the included KAPA HiFi HotStart ReadyMix.

Table 1: Bacterial species used to test effectiveness of bacterial rRNA custom depletion with probes selected by KAPA RiboDesigner

Bacterial species	Accession numbers	# of probes in custom probe pool
Escherichia coli	GCF_001723505.1	98
Corynebacterium matruchotii	GCF_000175375.1	90
Haemophilus parainfluenzae	GCA_000191405.1	85
Streptococcus cristatus	GCF_900475445.1	82
Francisella tularensis LVS	NC_007880	83
Francisella noatunensis	NC_023029	88
Streptococcus mitis	GCF_000148585.2	84
Streptococcus sanguinis	GCF_000014205.1	90
Streptococcus gordonii	GCF_000017005.1	82
8-species mix (containing all of the above species except *E. coli*)		404
Co-culture A (*S. sanguinis* and *H. parainfluenzae*)		175
Co-culture B (*S. sanguinis* and *C. matruchotii*)		180

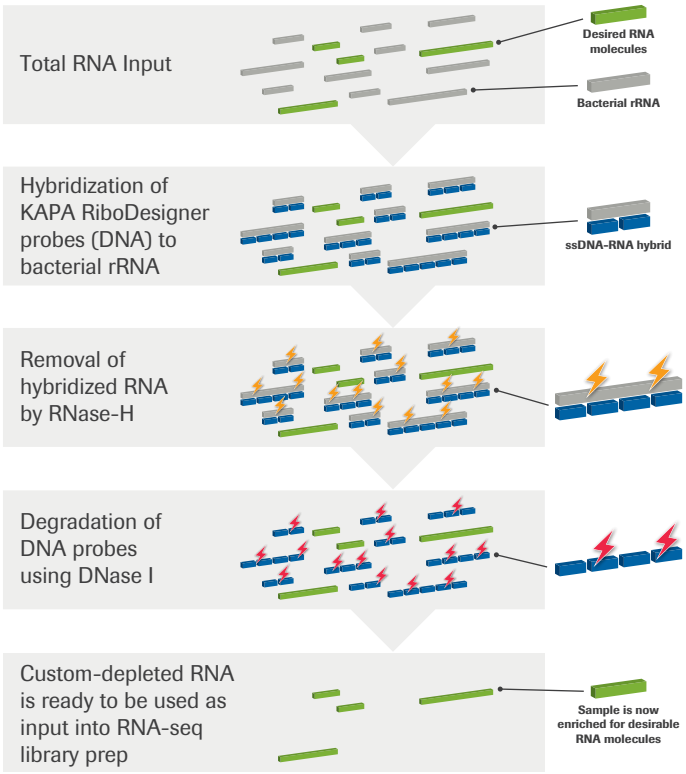

Figure 3: Overview of rRNA depletion workflow using custom-designed probes for bacterial rRNA. This rRNA depletion protocol uses custom probes designed by KAPA RiboDesigner in the RNase H-based depletion workflow employed by the KAPA RNA HyperPrep Kit with RiboErase (HMR).

To measure the potential impact of custom ribodepletion on the evaluation of gene expression, several libraries were also created from non-depleted RNA for comparison. Non-depleted RNA-seq libraries were prepared according to the KAPA RNA HyperPrep Kit workflow, but without the upfront KAPA RiboErase treatment.

Because the rRNA was not removed from these samples prior to adapter ligation and thus the mass of RNA at this step was much greater compared to the above ribodepleted samples, KAPA UDI Adapters were used at 15 µM (compared to 1.5 µM), and the resulting libraries were amplified for only 5 cycles (rather than 14 cycles).

All final libraries were analyzed for library success and adapter-dimer presence using an Agilent Bioanalyzer Instrument with a High Sensitivity DNA Kit. Libraries were then quantified with the qPCR-based KAPA Library Quantification Kit on a Roche LightCycler® 480 II prior to dilution and pooling.

Sequencing and bioinformatics analysis

Libraries were diluted, pooled and sequenced across two high-throughput 2 x 76 bp runs on an Illumina® NextSeq 500 sequencer. Reads were mapped to the appropriate reference genomes and rRNA sequences using HISAT2 (Kim, et al, 2019). Gene expression analysis was performed using KALLISTO (Bray, et al. 2016) and EDGE-R (Robinson, et al. 2010). Plots were generated using R and EXCEL (R Core Team, 2013).

Results and discussion

First, the efficiency of rRNA depletion in a single microbial species—*Escherichia coli*—was evaluated. The *E. coli* genome is ~5 Mb in size and carries seven or eight non-identical but highly similar copies of each rRNA gene (5S, 16S, and 23S). Using the annotated *E. coli* genome sequence available via NCBI, KAPA RiboDesigner selected 98 non-redundant probes to target all of the *E. coli* rRNA genes. The newly designed probes achieved greater than 98% rRNA depletion efficiency when used with KAPA RNA HyperPrep Kit with RiboErase (HMR) (Figures 4 and 5).

To ensure that rRNA depletion did not affect overall transcriptome profiles, correlation analysis was performed between non-depleted and depleted samples. The resulting very high correlation (~94%) demonstrated that the overall transcriptome profiles were very similar and largely unaffected by the rRNA depletion process (Figure 6). To measure the robustness and reproducibility of the assay, the correlation of the transcriptome among technical replicates of post-depletion samples was calculated and found to be very high (>98%) (Figure 7); this is even higher than for non-depleted samples. Together, these findings demonstrate that this method provides a reliable assessment of gene expression in *E. coli*.

The efficiency of rRNA depletion efficiency in three additional species: *Corynebacterium matruchotii*, *Haemophilus parainfluenzae*, and *Streptococcus sanguinis* (Table 1), as well as co-cultures of these bacteria (Co-culture A and Co-culture B; see Table 1). Probes for rRNA depletion were designed individually for each species; for the co-cultures, appropriate probe sets were combined for use in the ribodepletion workflow. A high level of rRNA depletion (>98%) was achieved for both the individual species and the co-cultures, showing the scalability and robustness of the protocol in microbial species other than *E. coli* (Figure 8). Thus, this customized ribodepletion workflow can potentially be used for many, if not all, microbial species for which the annotated genome is available.

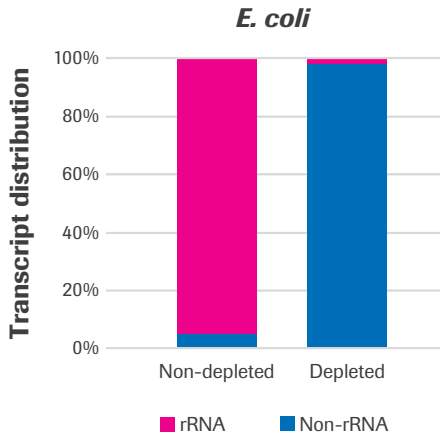

Figure 4: The transcriptome distribution in *E. coli* samples before and after rRNA depletion using custom probes. *E. coli* total RNA is composed mostly of rRNA; when non-depleted samples are sequenced, rRNA reads make up as much as ~95% of the total reads. Following the rRNA depletion protocol using the probes designed with KAPA RiboDesigner, residual rRNA reads were reduced to <2%. To normalize samples for similar non-rRNA content in this comparison, the non-depleted samples were sub-sampled to 90M paired-end reads and depleted samples were subsampled to 3M paired-end reads prior to calculation of the proportion of reads mapping to rRNA genes.

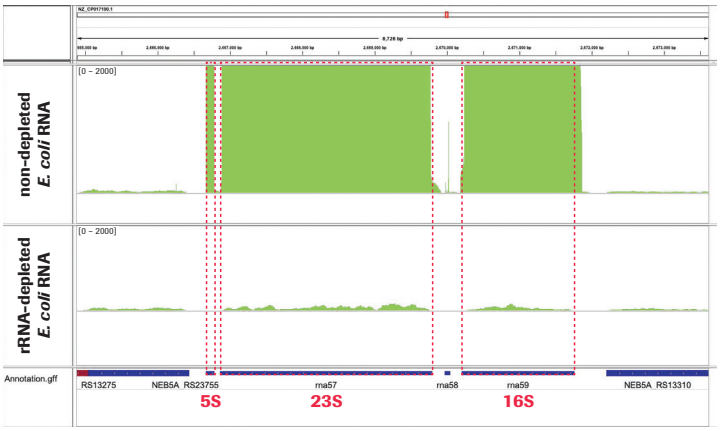

Figure 5: Genome browser visualization of rRNA genes represented in the transcriptome of *E. coli* for non-depleted and depleted samples. Results of non-depleted and depleted samples are presented as two individual tracks. The regions outlined in red represent the rRNA genes (5S, 23S, and 16S). The y-axis is capped to a coverage depth of 2,000; however, in non-depleted samples the rRNA peak reaches ~200,000. After depletion, the coverage depth of rRNA is similar to other nearby genes.

The next experiment tested the ability of KAPA RiboDesigner to select non-redundant rRNA depletion probes for multiple bacterial species simultaneously. The concatenated, annotated genomes of eight bacterial species (Table 1) were provided as input to KAPA RiboDesigner; the algorithm then designed a total of 404 non-redundant probes to deplete all of the rRNA

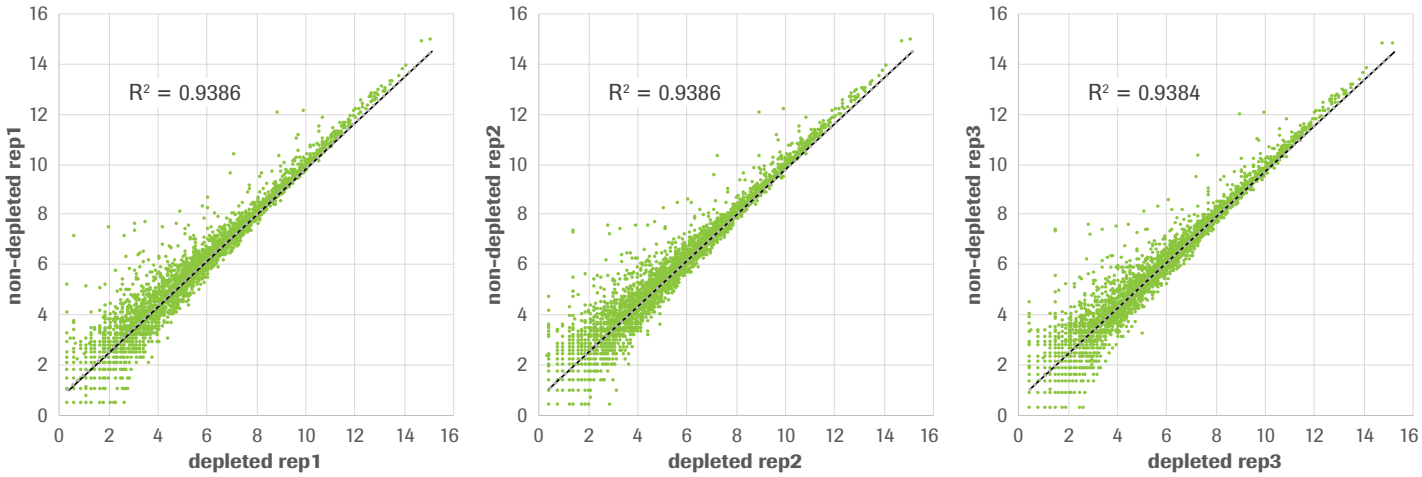

Figure 6: Highly correlated gene expression between rRNA-depleted and non-depleted *E. coli* samples. Gene expression patterns were highly reproducible (94% correlation) for all three replicates, indicating that rRNA depletion using custom probes has minimal impact on the assessment of gene expression. Both the X and Y axes represent gene expression in log2 format.

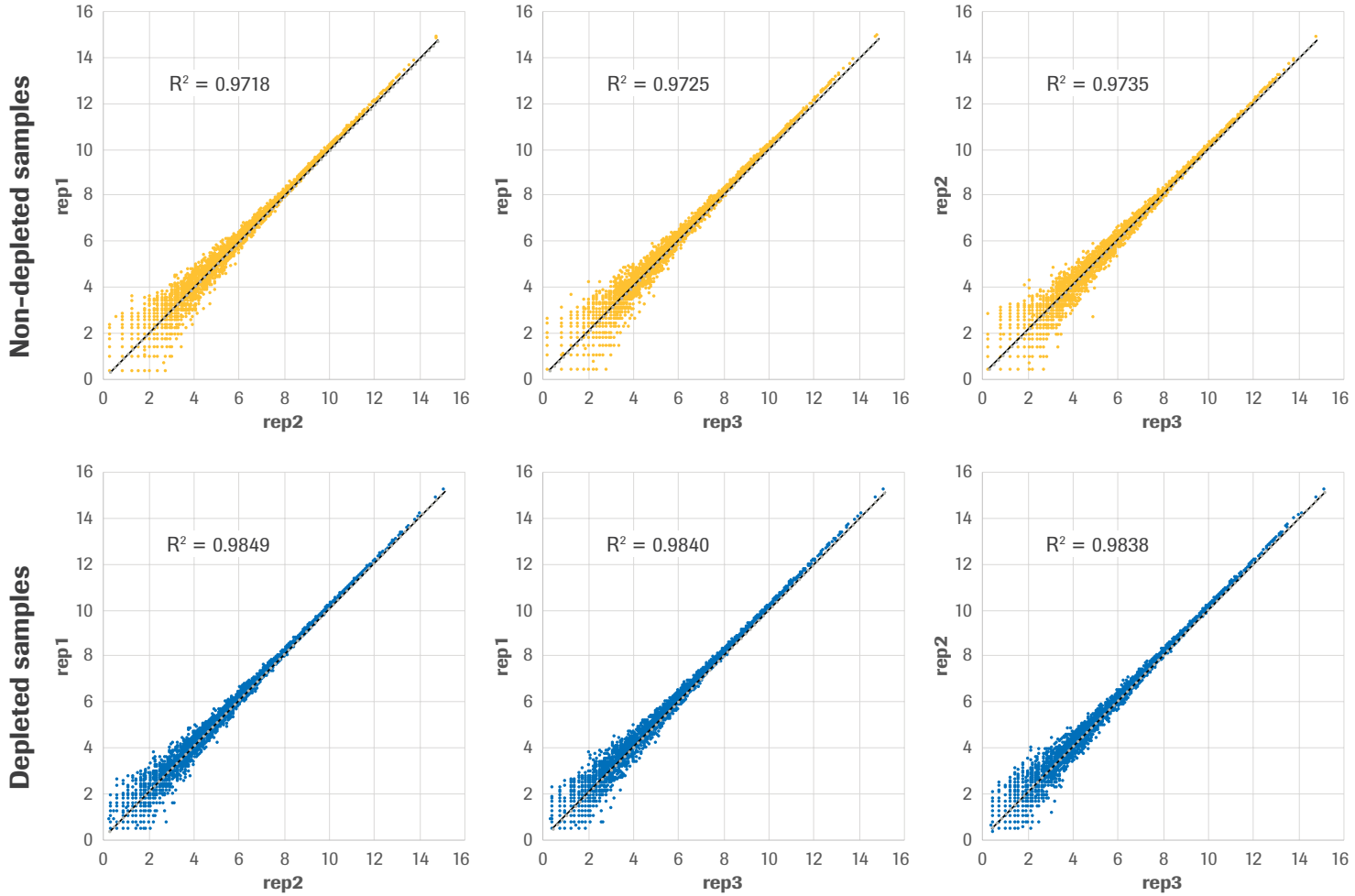

Figure 7: Highly correlated gene expression between technical replicates for both depleted and non-depleted *E. coli* samples. The high level of correlation (>97%) between replicates shows high reproducibility across non-depleted *E. coli* samples prepared with the KAPA RNA Hyper Prep Kit. Following rRNA depletion with custom probes, the correlation between replicates increased to 98%. This indicates that removal of rRNA not only enriches the transcriptome, but also increases its reproducibility and robustness. Both the X and Y axes represent gene expression in log2 format.

molecules from this 8-species mix. For comparison, if each genome had been processed separately and combined, the total number of probes designed for all eight species would have been ~684 (Table 1). Thus, by minimizing sequence redundancy, KAPA RiboDesigner was able to reduce the total number of probes required by ~40%, representing a significant cost savings to the researcher.

The 404-probe pool selected by KAPA RiboDesigner was ordered and their depletion efficiency was tested on a manually created 8-species mixture of total RNA. The KAPA RNA HyperPrep Kit with RiboErase (HMR) Workflow was followed, substituting the custom probe pool for the provided HMR probes.. The custom probes achieved >98% rRNA depletion efficiency (Figure 9).

This highly efficient depletion of rRNA from a multi-species mixture of total RNA demonstrates the utility of this custom probe design solution for complex mixtures of microbes, and opens the possibility for rRNA depletion from a vast array of microbial samples, including microbiome samples. In addition to its utility in creating highly efficient ribodepletion probes for single- and multi-species RNA samples, the flexibility of KAPA RiboDesigner may further expand the potential application of this novel solution for custom probe design.

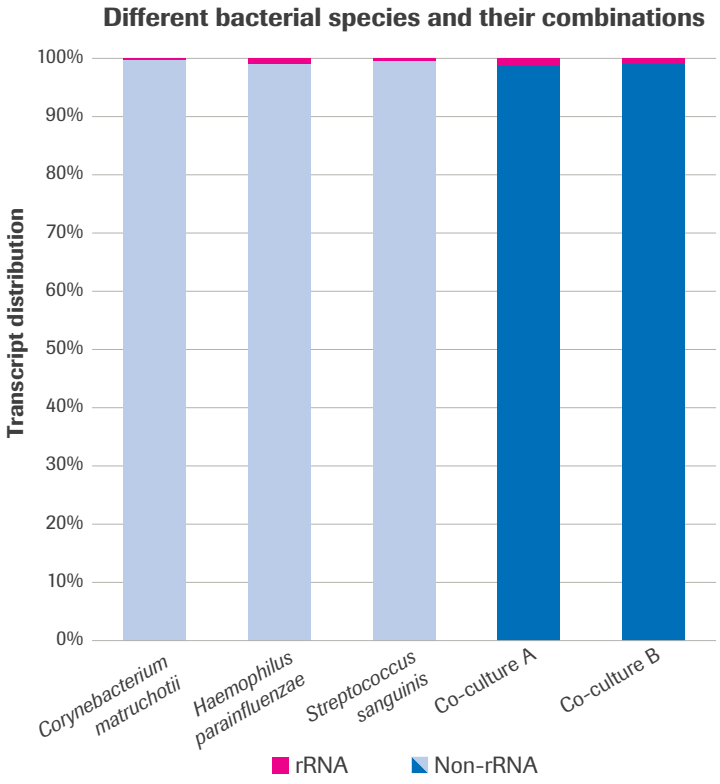

Figure 8: Efficient rRNA depletion from total RNA extracted from three bacterial species and two co-cultured samples. This chart depicts the relative abundance of rRNA post depletion. In all cases, the rRNA depletion was >98%.

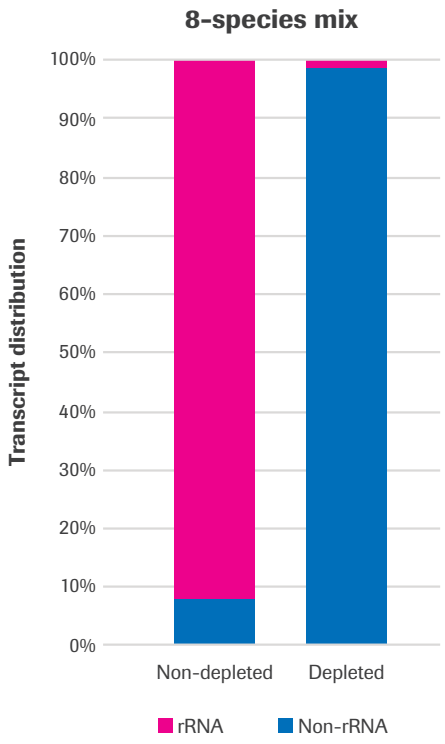

Figure 9. Efficient, simultaneous rRNA depletion from a mixture of RNA from eight bacterial samples. Using 404 custom probes designed with KAPA RiboDesigner, the rRNA content in the multi-species mix was reduced from ~92% in the non-depletion sample to <2% in the rRNA-depleted sample.

Conclusions

The KAPA RiboDesigner probe design solution presented here automatically designs non-redundant probes for rRNA depletion for any bacterial species with an available annotated genome. Custom probes designed by KAPA RiboDesigner were used in place of the HMR oligos included in the KAPA RNA HyperPrep Kit with RiboErase (HMR) in the library preparation workflow. These probes achieved high-quality RNA-seq results with highly efficient (>98%) bacterial rRNA depletion from total RNA isolated from three sample types: RNA from individual bacterial species; RNA isolated from co-cultures of two different species; and a manually created mixture of RNA derived from eight individual bacterial species. In conclusion, KAPA RiboDesigner greatly expands the capabilities of the KAPA RNA HyperPrep Kit with RiboErase (HMR) to include a vast number of sequenced microbial species and mixtures of microbial samples while minimizing the number of probes required.

To learn more about how KAPA RiboDesigner can help your projects, please visit **go.roche.com/RiboDesigner**.

References

1. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37(8):907–915. doi:10.1038/s41587-019-0201-4
2. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification [published correction appears in Nat Biotechnol. 2016 Aug 9;34(8):888]. *Nat Biotechnol*. 2016;34(5):525-527. doi:10.1038/nbt.3519
3. Robinson MD, McCarthy DJ, Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140. doi: 10.1093/bioinformatics/btp616.
4. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
5. Hapshe N, Geldart A, Kasinskas R, Siba A, Becker D, Thompson D. (2018). High-efficiency species-specific ribosomal RNA depletion with the KAPA RNA HyperPrep Kit. Roche Sequencing and Life Science.

Acknowledgement

All bacterial culturing was performed (with the exception of *E. coli*) in Dr. Matthew Ramsey's and Dr. Kathryn Ramsey's lab at the University of Rhode Island, Department of Cell and Molecular Biology.

Published by:

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

sequencing.roche.com

MC-US-07812 APP105002 A553 10/20