**Application Note**

# A New RNA-Seq Workflow for the Singular Genomics G4® Sequencing Platform Using the KAPA RNA HyperPrep Kit with RiboErase (HMR)


## Authors

**Sandra Theall**
*Applications Scientist*

**Sarah Trusiak***
*Senior Application Scientist*

**Mariana Fitarelli-Kiehl**
*Principal Applications Scientist*

**Lindsey Cambria**
*Support Team Manager*

**Spencer Debenport***
*Applications Team Manager*

**Rachel Kasinskas**
*Director of Support & Applications*

**Jieqiong Dai**
*Senior Bioinformatics Scientist*

**Nikita Raymond Dsouza**
*Bioinformatics Scientist*

**Alejandro Quiroz Zarate**
*Bioinformatics Manager*

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

**Author no longer an employee of Roche*

**Sabrina Shore**
*Associate Director, Sequencing Applications*

**Ryan Shultzaberger**
*Director, Bioinformatics*

**Chrystal Day**
*Scientist, Sequencing Applications*

**Timothy Looney**
*Senior Director, Scientific Affairs*

**Martin M Fabani**
*Vice President, Sequencing Applications*

**Yu Xiang**
*Senior Scientist, Bioinformatics*

Singular Genomics Systems, Inc.
San Diego, CA



*Transcriptomic analysis using next-generation sequencing (NGS) of RNA can provide a snapshot of the total transcripts present in a sample at a specific time point. RNA sequencing (RNA-seq) facilitates quantifying gene expression, discovery of novel transcripts, and identification of gene fusions; together, these increase our understanding of human biology and disease.[1] With the expanding use of RNA-seq in clinical research, there is a demand for rapid and cost-effective sequencing. The Singular Genomics G4® Sequencing Platform lends itself as a powerful solution to fulfill this burgeoning need. Equipped with the ability to run up to 4 flow cells in parallel, each with 4 independently addressable lanes, G4 facilitates streamlined laboratory functions by combining novel 4-color rapid sequencing-by-synthesis (SBS) chemistry with advanced engineering to deliver results in a single day. This application note presents the integration of the KAPA RNA HyperPrep Kit with RiboErase (HMR) and Singular Genomics' adapters for RNA-seq on the G4® Platform, showcasing the reliability of sequencing with G4 and Roche's KAPA chemistry.*


## Introduction

For decades, gene expression analysis has played a vital role in the research of human diseases.[1] More recently, next-generation sequencing (NGS) has catapulted transcriptomics into what is now known as RNA-sequencing (RNA-seq). RNA-seq is a high-throughput method for quantifying the abundance of all RNA transcripts present in a given sample. Breakthroughs in the simplification of RNA-seq workflows, especially in the preparation of RNA-seq NGS libraries, has made RNA-seq a central component of basic, translational and clinical research. However, as the demand for high-throughput RNA-seq in disease research expands, so does the need for fast turnaround times and flexible workflows.[2]

To address this, Singular Genomics has launched the G4® Sequencing Platform to deliver rapid, flexible, accurate, and cost-effective NGS sequencing. G4 is an innovative benchtop sequencer that leverages a 4-color rapid sequencing by synthesis (SBS) chemistry with advanced optics and fluidics engineering to provide single-day turnaround times across sequencing applications. By combining fast run times and the ability to run up to 4 flow cells with a total of 16 independently addressable lanes, the G4 Platform enables highly efficient laboratory operations. Details regarding G4 specifications, including run time, accuracy, and quality metrics, can be found on the Singular Genomics website. Together, the utility of the trusted KAPA RNA HyperPrep with RiboErase (HMR) kit and the powerful G4 platform offers a new solution for researchers for the sequencing of RNA, with both low and high RNA input amounts.

## Materials and Methods

### Library Preparation

***Input, Ribosomal RNA Depletion, Fragmentation and cDNA synthesis:*** Triplicate libraries were prepared from 25 ng and 500 ng high-quality RNA inputs, using Universal Human Reference (UHR) RNA (Agilent Technologies) and following the standard KAPA RNA HyperPrep with RiboErase workflow, up to the adapter ligation step. Briefly, following ribodepletion, the RNA was fragmented by heating for 8 minutes at 94°C in the presence of magnesium to achieve 100-200 bp fragments, and was then converted to cDNA.

***Adapter Ligation and Cleavage:*** SG Universal Library Prep Adapter (Catalog # 700,111), which is a stem-loop adapter, was diluted to 0.075 µM for 25 ng input samples and 1.5 µM for 500 ng input samples. Diluted adapter (5 uL) was incorporated through a ligation reaction **(Figure 1)** using KAPA Adapter Ligation mastermix and a 15 minute incubation at 20°C. The stem-loop adapters were cleaved by adding 3 µL of Singular Genomics Cleave Enzyme, 25 µL of 4X Cleave Reaction Buffer and 3.5 µL of nuclease-free water to the libraries and incubating at 37°C for 10 minutes followed immediately by incubation at 67°C for 30 minutes. A 0.8X bead cleanup using KAPA Pure Beads was performed to remove residual reagents from the adapter ligation and digestion steps.

***Library Amplification and Index Incorporation:*** Indices were incorporated during PCR amplification by adding 5 µL of the SG UDI Primers (Catalog #700,134) (2 µM final concentration) and 25 µL of the KAPA 2X HiFi HotStart ReadyMix to each library. PCR amplification was carried out using the thermal cycling conditions described in the manufacturer's instructions; the number of PCR cycles used were optimized based on input for this specific workflow. Fifteen cycles were used for 25 ng libraries and 10 cycles were used for 500 ng libraries. A 1X bead cleanup using KAPA Pure beads was performed to remove residual PCR amplification reagents.

***Library QC:*** Sequencing-ready libraries were diluted 1:10 using Qiagen™ Elution Buffer and quantified with Invitrogen™ Qubit 1X HS dsDNA assay. The fragment size distributions and adapter-dimer contamination levels were analyzed for each sample with a 2100 Bioanalyzer instrument and Agilent High Sensitivity DNA Kit (Agilent Technologies). The percentage of adapter-dimer contamination was calculated by setting one region from 120 bp to 170 bp and a second region from 120bp to 1000 bp in the 2100 Expert Software (Agilent Technologies) and dividing the concentration of the first region by the second region.

### Sequencing

Standard protocols were performed using the G4® Sequencing Platform. Libraries were pooled and sequenced over two F2 flow cells using 300 cycle kits (Catalog #700,106) at 2 X 150 bp (12 bp dual indices), generating a total of 565 million paired-end reads. Libraries were sequenced in triplicate.

### Bioinformatics Analysis

Raw reads were trimmed using Fastp (v0.23.2), subsampled to 55 million pairs per sample using Seqtk (v1.3), and then aligned to the GRCh38 human reference genome (GCA_000001405.15_GRCh38) using STAR (v2.7.10a). Transcript-level alignment and quantification was performed using Salmon (v1.6.0). Multiple ways of quality control were performed using FastQC (v0.11.9), Qualimap (v.2.2.2-dev), RNA-SeQC (v2.3.5), and deepTools (v3.5.1).

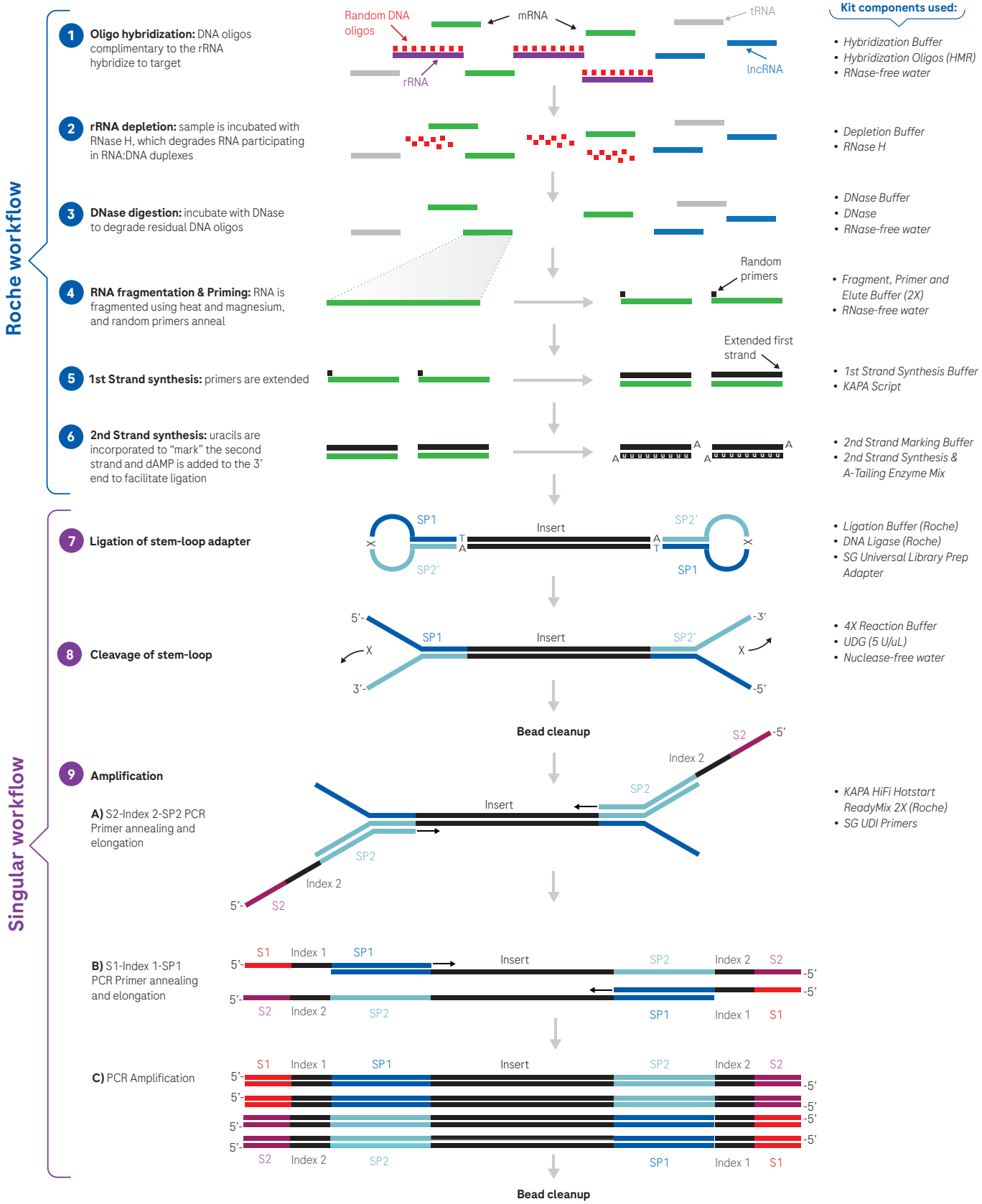


**Figure 1. Combined workflow for de novo library prep using KAPA RNA HyperPrep with RiboErase (HMR) and the Singular Genomics G4 workflow.** Details on the library preparation workflow may be found on the Singular Genomics website at *www.singulargenomics.com*. Singular Genomics uses stem loop-shaped adapters as part of the library preparation process. After end-repair and A-tailing, stem loop adapters are ligated to the DNA ends. These adapters contain a cleavable site which, when cleaved, generates the substrate for PCR amplification. During PCR amplification, the primers add the S1/S2 sequences required for cluster generation on the Singular Genomics flow cell as well as 12 base pair unique dual indices, if indices are used. For more information regarding Singular Genomics adapter design, please refer to the Adapters and Indices for the G4 Sequencing Platform Reference Guide.[3]

# Results and Discussion

## Library QC Metrics

The final, ribo-depleted RNA libraries were assessed as described in the methods. Both low-input (25 ng) and high-input (500 ng) RNA samples exhibited minimal adapter-dimer formation with levels below 5% **(Figure 2A)** and produced libraries of similar sizes **(Figure 2B)**, despite the difference in Singular adapter concentration required for each input. Both low- and high-input RNA samples generated sufficient library material for library QC, sequencing with the low, 0.1 nM minimum concentration requirement for Singular denaturation and dilution protocol, and archiving **(Figure 2C)**. Library yields from low-input samples were greater than from the high-input samples, most likely due to the difference in number of amplification cycles; 15 and 10 cycles, respectively.

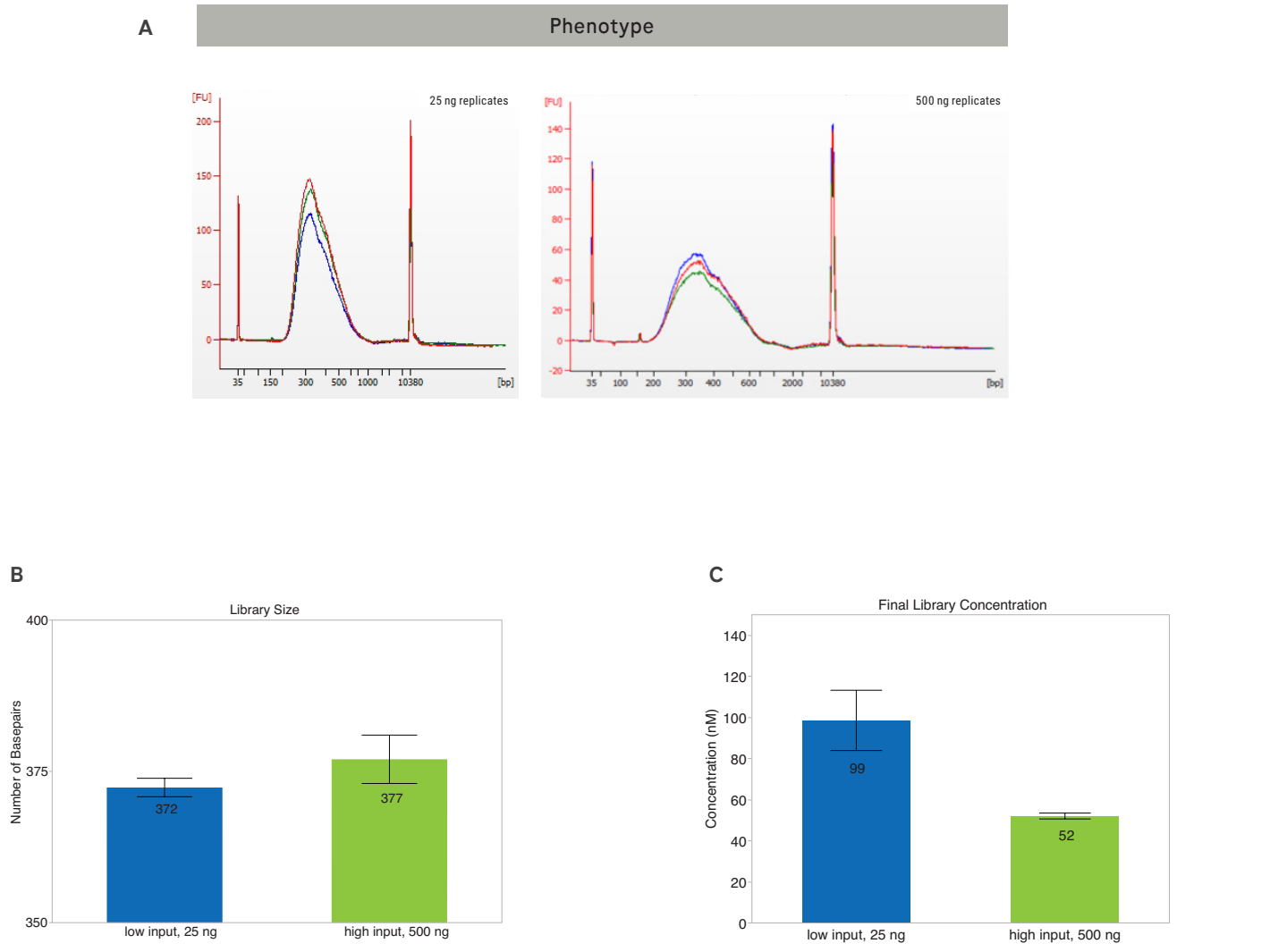


**Figure 2. High-quality libraries were generated across both sample input amounts. (A)** The library phenotype and **(B)** size distribution show that both sample input amounts produced consistently-sized libraries that appear free of contamination by unligated adapters. **(C)** Libraries exceeded minimum concentration of 0.1 nM required to proceed to Singular "denature and dilute library" protocol for sequencing on the G4 platform. Data and error bars reflect the mean and standard deviation respectively. Triplicate libraries sequenced.

## Sequencing Metrics

The first step in the assessment of the overall sequencing data was to identify the prevalence of raw sequencing reads such as duplicate reads, residual rRNA reads, and residual adapter content. The observed duplication rate did not cause concern for overall sequencing quality, however, cycle number can be optimized by the end-user to reduce duplication rate. Elevated duplication rates in samples with lower input mass, as opposed to those with higher input mass, are plausibly attributed to a greater number of PCR amplification cycles undergone by lower input mass libraries **(Figure 3A)**. Both sets of libraries demonstrated effective depletion of rRNA **(Figure 3B)** and approximately 0.1% of adapter content was present across all samples indicated by the overlapped lines that represent the samples in the FastQC: Adapter Content plot. The FastQC: Adapter Content module shows a low abundance (below 5%) of uninformative adapter content present in each library, indicating an efficient library preparation workflow **(Figure 3C)**. Additionally, the gene coverage profiles were also comparable across low-input and high-input RNA libraries showing that sample input mass does not affect coverage across transcripts. **(Figure 4)**.

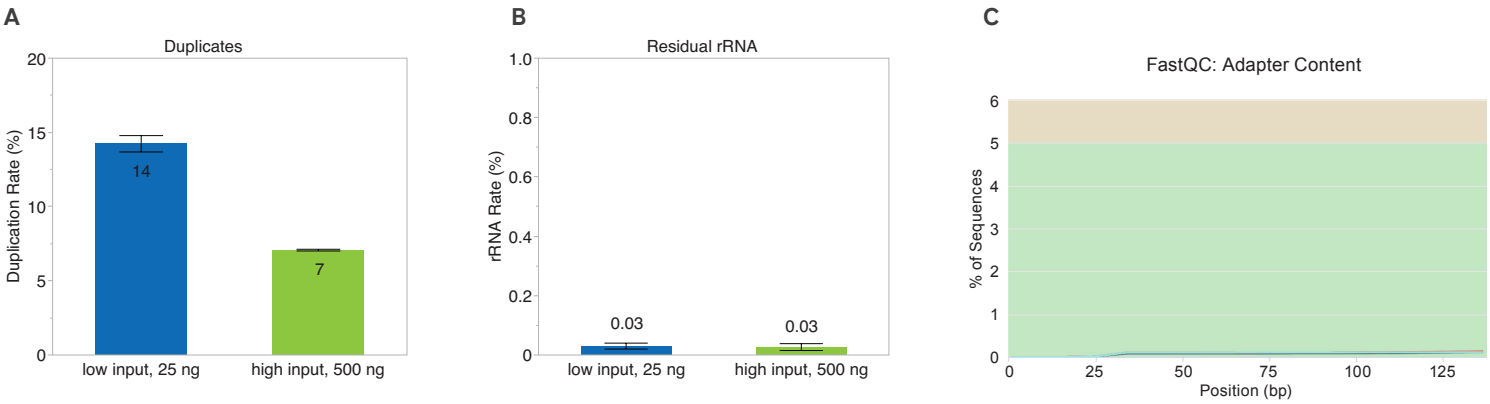


**Figure 3: Low abundance of raw reads containing duplicates, residual rRNA and adapter content across low and high RNA inputs. (A)** The observed duplication rate fell within expected levels across both input mass amounts and higher level shown from 25 ng input mass samples is likely indicative to having a greater amount of PCR amplification cycles compared to 500 ng input mass samples. The data and error bars reflect the mean and standard deviation respectively. **(B)** The consistency and low abundance of residual rRNA reads indicated that the efficacy of ribosomal RNA removal was maintained using both low and high input mass into library preparation. **(C)** The FastQC: Adapter Content module plot shows the cumulative percentage count of the proportion of reads in each library that has adapter sequences detected at any specific position. Once an adapter sequence has been detected in a read, it is counted as being present through the entire read; the module will issue a warning if any adapter sequence is present in more than 5% of all reads, or a failure notification if more than 10% of all reads contain adapter sequence. Triplicate libraries sequenced.

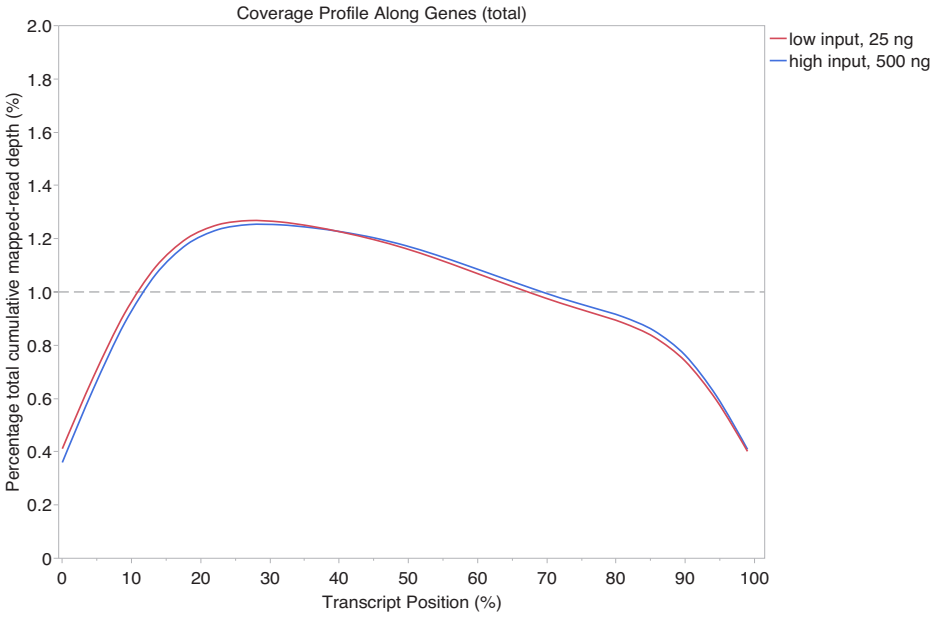


**Figure 4: Sample input mass does not impact coverage profiles along genes.** Gene coverage profile shows the mean distribution of coverage depth across the length of all mapped transcripts, normalized by total read counts, and was comparable between low input (25 ng) and high input (500 ng) libraries. Triplicate libraries were sequenced.

Strand-specificity was also assessed, as strand-specific RNA-seq enables more accurate identification of transcripts by identifying which strand the transcript originated from, thus avoiding ambiguity when genes on the opposite DNA strand may overlap. Strand-specificity was maintained for both low- and high-input RNA libraries sequenced on the G4 system **(Figure 5)**.

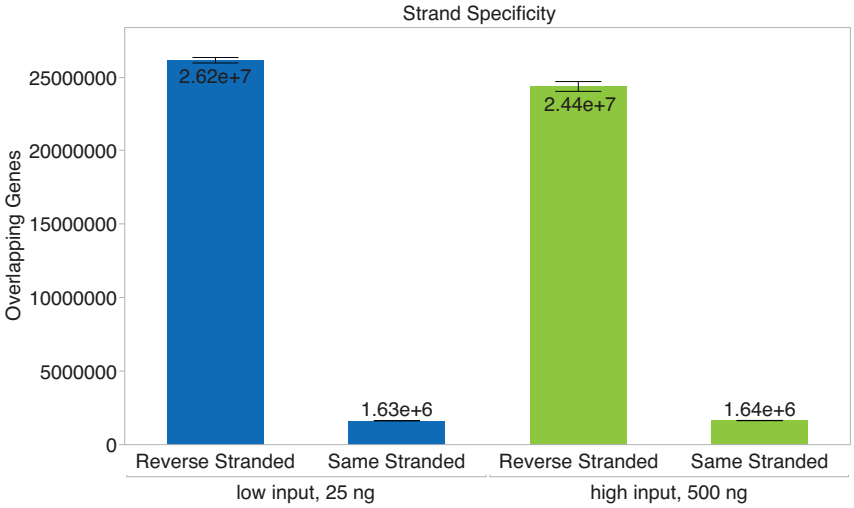


**Figure 5. Strand specificity was maintained, ensuring correct alignment of reads across both low- and high-RNA inputs.** Strand-specificity performance was assessed quantitatively by measuring the total read count of genes that are sorted as "reverse stranded" or "same stranded". The "same stranded" represents the total read counts for the 1st read strand aligned with RNA and the "reverse stranded" represents the total read counts for the 2nd read strand aligned with RNA. If the libraries did not maintain strand specificity, then the total read count would be approximately split equally between "reverse stranded" and "same stranded" for each condition and if the libraries did maintain library specificity, then the total read count would be in the majority of either "reverse stranded" or "same stranded". Data and error bars reflect the mean and standard deviation respectively. Triplicate libraries were sequenced.

Expression efficiency and genomic mapping rates were consistent across both low and high input conditions **(Figures 6 and 7)**. Similar numbers of protein coding and non-protein coding genes and transcripts were detected across both sample input mass conditions as well, with the number of transcripts exceeding the number of genes due to most genes having multiple transcripts associated with them.

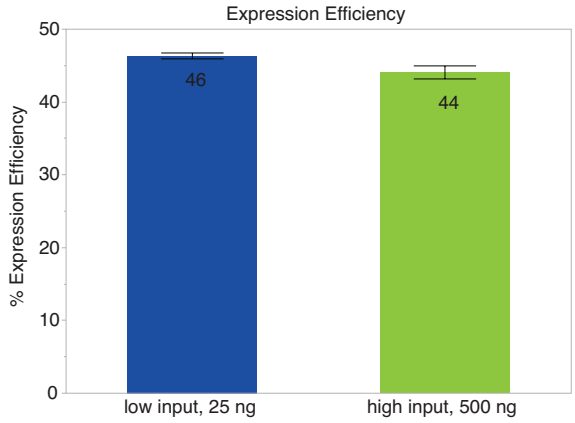


**Figure 6. Sample input mass does not impact expression efficiency** performance. Data bars represent the ratio of exon-mapped reads to the total reads sequenced. The percentage of unique mapping reads aligned to the reference genome were comparable between low input (25 ng) and high input (500 ng) libraries generated. Data and error bars reflect the mean and standard deviation respectively. Triplicate libraries sequenced.

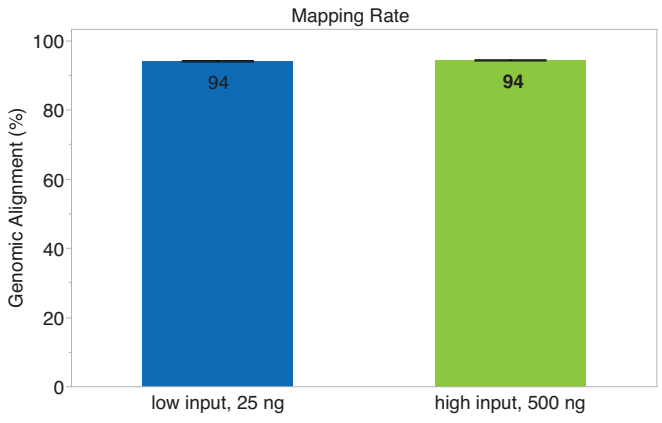


**Figure 7: Similar genomic mapping rate across both low and high RNA input.** The percentage of unique mapping reads aligned to the reference genome was comparable between low-input (25 ng) and high-input (500 ng) libraries. Data and error bars reflect the mean and standard deviation respectively. Triplicate libraries sequenced.

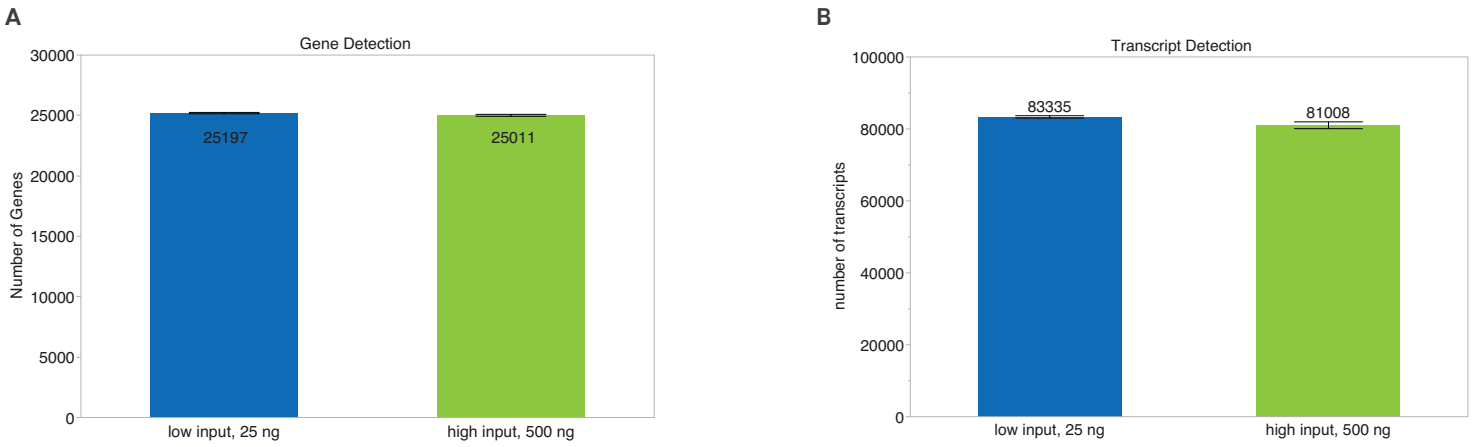


**Figure 8: High abundance of genes and transcripts detected across low and high RNA input.** Low input (25 ng) libraries yielded similar numbers of identified genes **(A)** and transcripts **(B)** compared to high input (500 ng) libraries. Data and error bars reflect the mean and standard deviation respectively. Triplicate libraries sequenced.

## Conclusion

This study demonstrates that Singular Genomics adapters and indices can be seamlessly integrated into Roche's existing KAPA RNA HyperPrep Kit with RiboErase (HMR) library preparation workflow, allowing for versatile and accessible benchtop sequencing on the Singular Genomics G4® Sequencing Platform. RNA-seq libraries constructed using high-quality RNA with the KAPA RNA HyperPrep Kit with Riboerase (HMR) and Singular Genomics adapters and indices generated high quality sequencing data and metrics for both low-input (25 ng) and high-input (500 ng) samples. Pre-sequencing metrics showed sufficient library yield was achieved for sequencing as well as additional sequencing runs and archiving if needed, along with consistent sizes and minimal to no adapter dimer across both conditions. Sequencing metrics showed that quality is maintained across key RNA-seq metrics on the G4 Platform, providing a comprehensive view of the human transcriptome and highlighting the versatility and robustness of the KAPA RNA HyperPrep Kit. Thus, the trusted KAPA RNA Library Preparation Kit with RiboErase (HMR) coupled with the powerful G4 platform offers researchers a fast and flexible workflow for RNA-seq.

## References

1. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. *PLOS Computational Biology*. 2017;13(5):e1005457. doi: https://doi.org/10.1371/journal.pcbi.1005457.
2. Casamassimi A, Federico A, Rienzo M, Esposito S, Ciccodicola A. Transcriptome Profiling in Human Diseases: New Advances and Perspectives. *International Journal of Molecular Sciences*. 2017;18(8):1652. doi: https://doi.org/10.3390/ijms18081652.
3. Adapters and Indices for the G4 Sequencing Platform Reference Guide. Singular Genomics, Inc. 2022. https://singulargenomics.com/wp-content/uploads/2022/10/Adapters-and-Indices-for-G4-600007-087.pdf

For more information about Roche KAPA RNA HyperPrep Kits, please visit: **https://go.roche.com/RNA**

Published by:
**Roche Sequencing and Life Science**
9115 Hague Road
Indianapolis, IN 46256

**sequencing.roche.com**

Project name: KAPA RNA HyperPrep libraries for the Singular G4 system


MC-US-15377 APP105003 A773 08/24