Authors

Mariana Fitarelli Kiehl
Senior Applications Scientist

Nancy Nabilsi*
Senior Applications Scientist

Spencer Debenport
Applications Manager

Jennifer Pavlica*
Applications Manager

Rachel Kasinskas
Director of Scientific Support and Applications

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

Jacquelyn DuVall**
Senior Scientist

Jessica Hartman
Director of Product Applications

Cell Microsystems, Inc.
Durham, NC, USA

**Authors no longer employees of Roche*
***Author no longer employee of Cell Microsystems*

Date of first publication:
June 28, 2021

Publication date of this version:
June 28, 2021

KAPA HyperPlus Kit enables robust and flexible library preparation for ultra-low-input and single-cell mRNA sequencing

Understanding the cellular heterogeneity of tissues and tumors is critical to understanding biological processes, and as a result, interest in the molecular profiling of specific cell sub-populations—sometimes down to single-cell resolution—is expanding rapidly. Methods for ultra-low-input and single-cell RNA sequencing are being widely adopted for transcriptome profiling, but the inherently low input quantities of RNA obtained from restricted cell populations pose a challenge to successful next-generation sequencing (NGS) library preparation and reliable transcriptome analysis. This Application Note presents a workflow for generating high-quality libraries from as little as 10 pg of input RNA using a template switching method for cDNA synthesis followed by library preparation with the KAPA HyperPlus Kit, providing a flexible and robust sample process for single-cell RNA-seq applications.

Introduction

Teasing apart cell-to-cell variability in gene expression can provide valuable insights into the biology of cancer, immunology, and developmental biology. Conventional bulk RNA-seq technologies, in which RNA is extracted from thousands to millions of cells in one sample, have been widely used to evaluate differential gene expression between heterogeneous tissues and cell populations. However, the resulting data reports on the average transcriptional responses from that entire cell population, underestimating individual cellular heterogeneity. To address the need for higher-resolution gene expression data, there has been a rapid expansion in the development of more focused techniques such as single-cell RNA sequencing (scRNA-seq) and other ultra-low-input methods, facilitating the understanding of dynamic gene expression at the single-cell level.

The primary challenge for successful NGS library preparation in single-cell transcriptomics is the limited amount of RNA available in a single cell. A typical small eukaryotic cell contains approximately 10 pg of total RNA, of which only an estimated 0.1 pg is mRNA. Moreover, most mammalian genes express multiple alternatively spliced transcripts with variable exon composition and length, requiring methods that generate read coverage for mRNA isoforms as well as across full-length transcripts.

One of the most commonly used strategies for the synthesis and amplification of cDNA from picogram amounts of RNA is the Switch Mechanism at the 5' End of RNA Template method, known as SMART-Seq technology.[1] This method generates full-length cDNAs from even low amounts of input RNA, preserving the complexity of single-cell RNA. SMART-Seq relies on reverse transcription of poly-adenylated mRNA using oligo(dT) primers followed by a template switching reaction at the 5' end of the transcripts. The cDNA is then amplified and used as input into NGS

library preparation protocols; commonly used SMART-Seq workflows use a tagmentation-based library preparation method that combines fragmentation and adapter ligation into a single step. However, while tagmentation methods are highly efficient, they are only compatible with a narrow range of input DNA or cDNA amounts, and are also known to introduce coverage bias.[2]

The streamlined KAPA HyperPlus Kit overcomes the limitations of tagmentation-based workflows, yielding high-quality libraries with uniform coverage across a variety of inputs and sample types while still providing the efficiency of tagmentation-based methods due to its enzymatic fragmentation step. This Application Note describes new scRNA and ultra-low-input RNA-seq methods that combine the KAPA HyperPlus Kit with SMART-Seq cDNA synthesis. It also provides a comparison of these methods to tagmentation-based library preparation for ultra-low-input library preparation. Overall, the results demonstrate that the KAPA HyperPlus Kit provides a more flexible solution for transcriptomic analysis of both ultra-low-input and scRNA-seq applications compared to tagmentation-based methods while also improving data quality.

Materials and methods

Overall experimental design

For a comparison of methods for ultra-low-input RNA-seq **(Figure 1A)**, cDNA was first synthesized using the SMART-Seq method using 10 pg of a commercially available reference sample of human total RNA. Aliquots of the same cDNA sample were then used as input into either the KAPA HyperPlus Library Preparation Kit (150 pg, 1 ng, or 5 ng of cDNA input) or the Nextera XT DNA Library Preparation Kit (150 pg or 1 ng of cDNA input).

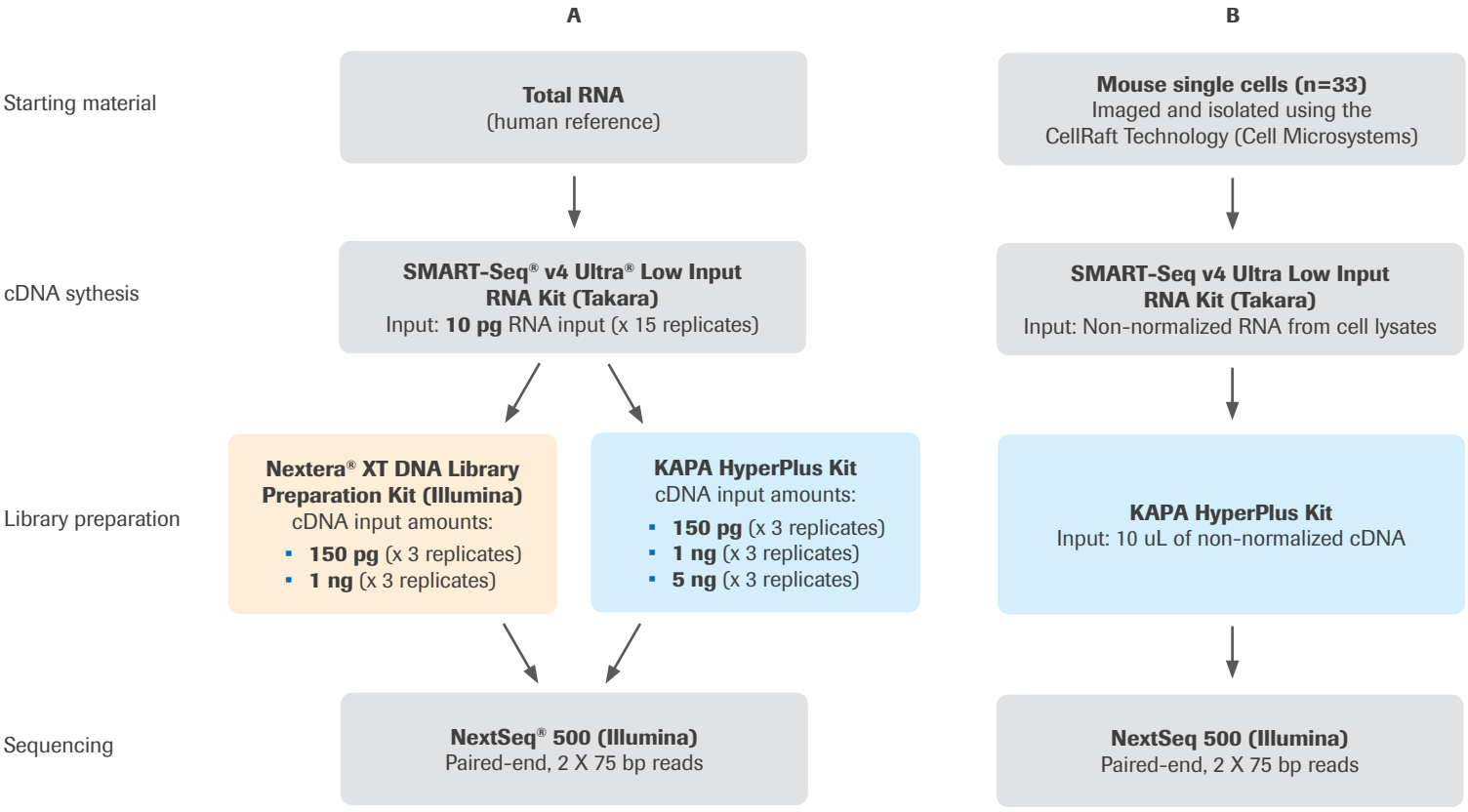

Figure 1. Experimental design. (A) Ultra-low-input RNA-seq using SMART-Seq cDNA synthesis followed by the Nextera XT DNA Library Preparation Kit or the KAPA HyperPlus Kit. **(B)** scRNA-seq with mouse single cells imaged and isolated using CellRaft Technology, followed by SMART-seq cDNA synthesis and the KAPA HyperPlus workflow.

Data on file.

For single-cell RNA-seq libraries, a similar overall strategy was applied to single cells derived from mouse (NIH 3T3). However, in these experiments, single-cell lysates were used directly as input into cDNA synthesis without normalization of RNA, and the resulting cDNA was then used as input for the KAPA HyperPlus Kit, also without normalization **(Figure 1B)**.

Comparison of library preparation workflows and cDNA input quantities using reference human RNA as input

To approximate the amount of RNA that may be present in a typical human cell, 10 pg of Universal Human Reference (UHR) RNA (Agilent Technologies) supplemented with ERCC RNA Spike-In Mix (Ambion) was used as starting material for cDNA synthesis. RNA was converted to cDNA using the SMART-Seq v4 Ultra Low Input RNA Kit (Takara) according to manufacturer's recommendations,[1] which includes a PCR amplification step. Fifteen replicates of amplified cDNA were pooled and purified with 1X AMPure XP beads (Beckman Coulter). The pooled cDNA was quantified with the Qubit dsDNA HS Assay Kit (Invitrogen) and the quality and size distribution were evaluated using a Bioanalyzer 2100 High Sensitivity DNA Kit (Agilent Technologies).

Libraries were prepared with the KAPA HyperPlus Kit using three cDNA input quantities: 150 pg, 1 ng and 5 ng. For each input amount, libraries were prepared in triplicate. Input cDNA was diluted in PCR-grade water and then enzymatically fragmented for 20 minutes at 37°C according to the KAPA HyperPlus Kit instructions, and then ligated to the full-length KAPA Dual-Indexed Adapters (Roche Catalog Number 08278555702). Adapters were used at stock concentrations of 150 nM, 500 nM or 1.5 µM, respectively, for the 150 pg, 1 ng and 5 ng input amounts. Libraries were then amplified by PCR (12 cycles for 150 pg cDNA input samples, and 8 cycles for the 1 ng or 5 ng cDNA input samples).

In parallel, libraries were prepared using the Nextera® XT DNA Library Preparation Kit (Illumina)—hereafter, referred to as "Nextera XT"—using 150 pg or 1 ng of cDNA following manufacturer's instructions. Libraries were generated in triplicate from 150 pg and 1 ng of cDNA input, and were PCR-amplified for 12 or 8 cycles, respectively, using the Nextera XT Index Kit v2 adapters (Illumina).

Fragment size distribution for all libraries was assessed with a Bioanalyzer 2100 High Sensitivity DNA kit and libraries were quantified using the qPCR-based KAPA Library Quantification Kit prior to pooling and sequencing.

Preparation of mouse single-cell RNA-seq libraries using the KAPA HyperPlus Kit

The application of SMART-Seq in combination with the KAPA HyperPlus Kit was evaluated with mouse-derived single cells in a collaboration with Cell Microsystems. To generate input for the SMART-Seq method, mouse-derived cells (NIH 3T3) were seeded onto a CytoSort Array (Cell Microsystems). After the cells adhered to the rafts, the cells were stained with MitoTracker Red, and the array was imaged using the CellRaft AIR System (Cell Microsystems)[3] **(Figure 2)**. Single cells were identified using the Off the AIR software, and thirty-three single cells were isolated from the CytoSort array directly into lysis buffer, followed by a 5 minute incubation at room temperature. Lysates were converted to cDNA with the SMART-Seq v4 Ultra Low Input RNA Kit (Takara) as per manual instructions. The cDNA was then quantified using the Qubit dsDNA HS Assay Kit prior to library preparation.

A 10 µL aliquot of amplified, purified cDNA from each sample was used as input into the KAPA HyperPlus Kit, and libraries were generated using an automated method on a Biomek FX workstation (Beckman Coulter). The automated protocol used parameters recommended for 1 ng DNA input into KAPA HyperPlus Kit (see KAPA HyperPlus Technical Data Sheet), including enzymatic fragmentation for 20 minutes at 37°C, 300 nM adapter stock concentration, and library amplification with 8 cycles of PCR. Fragment size distribution for all libraries was assessed with a Bioanalyzer 2100 High Sensitivity DNA kit and libraries were quantified using the qPCR-based KAPA Library Quantification Kit prior to pooling and sequencing.

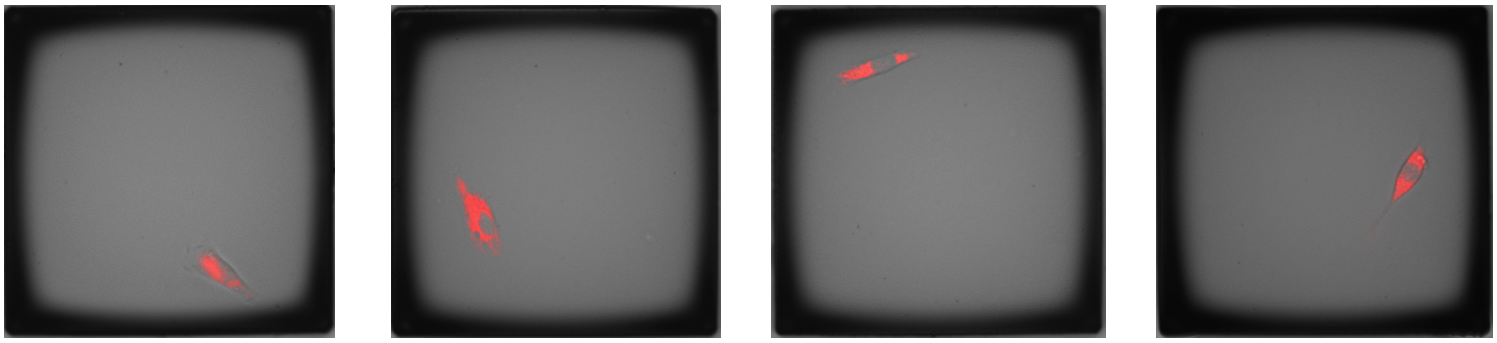

Figure 2. Representative images of NIH-3T3 cells on CytoSort array rafts. NIH-3T3 cells were seeded onto a CytoSort array and allowed to attach. Adherent cells were stained with MitoTracker Red and single cells were identified and isolated for downstream analysis (10X).

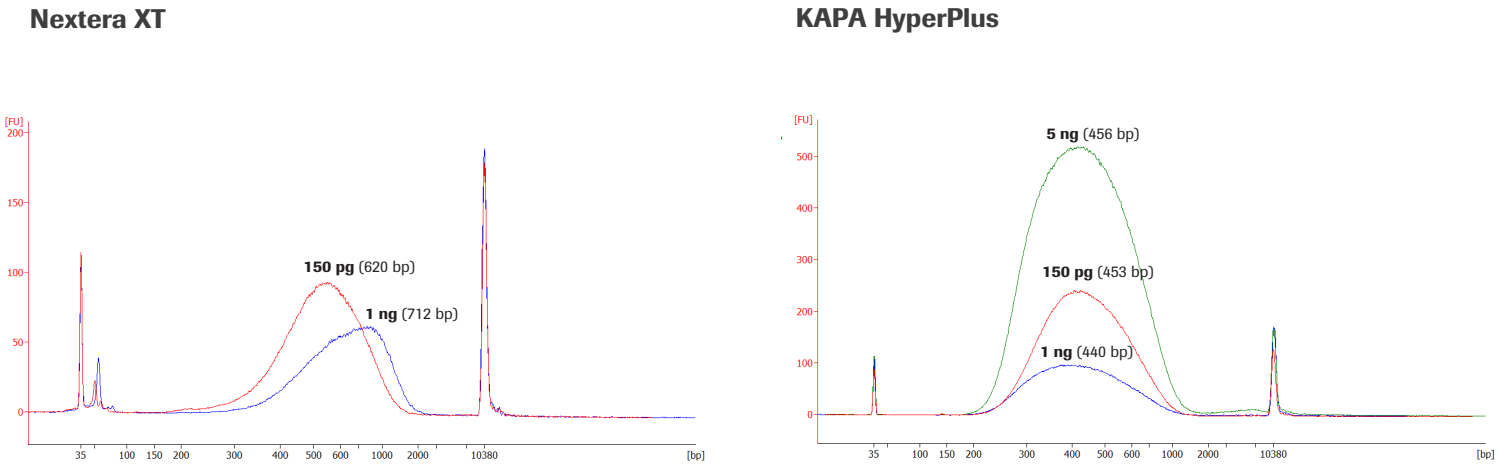

Figure 3. KAPA HyperPlus Kit yielded consistently sized libraries across all cDNA input amounts tested (150 pg, 1 ng, and 5 ng). Representative traces of final amplified libraries for each cDNA input amount, generated using the Nextera XT DNA Library Preparation Kit or the KAPA HyperPlus Kit. Libraries were assessed on a 2100 Bioanalyzer (Agilent) using the High Sensitivity DNA Kit.

Sequencing and data analysis

Normalized libraries were combined into two pools; one containing the libraries generated from human reference RNA and another with libraries from the single mouse cells. Pooled libraries were diluted, denatured, and loaded onto an Illumina® NextSeq® 500 instrument. Paired-end sequencing (2 x 75 bp) was then performed using a NextSeq High Output Kit (Illumina).

Raw reads were downsampled to 2M total reads and processed using an in-house RNA-seq analysis pipeline. Alignments were performed using HISAT2 and transcript expression levels were calculated from trimmed FASTQ files as Transcripts Per Million (TPM) using Kallisto. Principal component analysis (PCA) was performed with ClustVis[4] on TPM values, using unit variance scaling and singular value decomposition. Hierarchical clustering was performed using EdgeR. Pearson's correlation was calculated with GraphPad Prism, using averaged normalized TPM counts of replicates.

Results and discussion

The KAPA HyperPlus Kit produces libraries with uniform size distributions across a range of cDNA input amounts

For library preparation using the Nextera XT Kit, two cDNA input quantities were used: 150 pg (the recommended input based on the SMART-Seq V4 kit instructions), and 1 ng (the maximum input for successful tagmentation, according to the kit manufacturer). For library preparation with the KAPA HyperPlus Kit, both of these cDNA input quantities were used (150 pg and 1 ng) as well as an additional input amount of 5 ng cDNA input; this condition better represents the typical amount of cDNA synthetized from single cells.

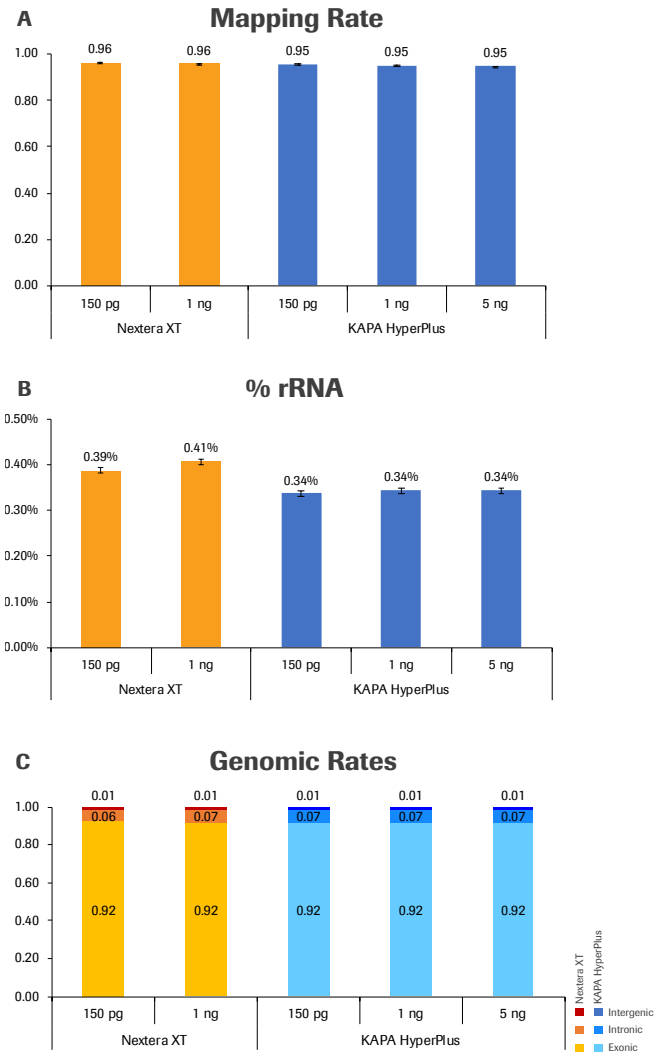

Figure 4. Consistent mapping and alignment metrics were achieved across cDNA inputs for both workflows. (A) Mapping rate. **(B)** Percent of residual rRNA reads. **(C)** Distribution of exonic, intronic, and intergenic mapped reads. Error bars represent the standard deviation for three replicates.

Data on file.

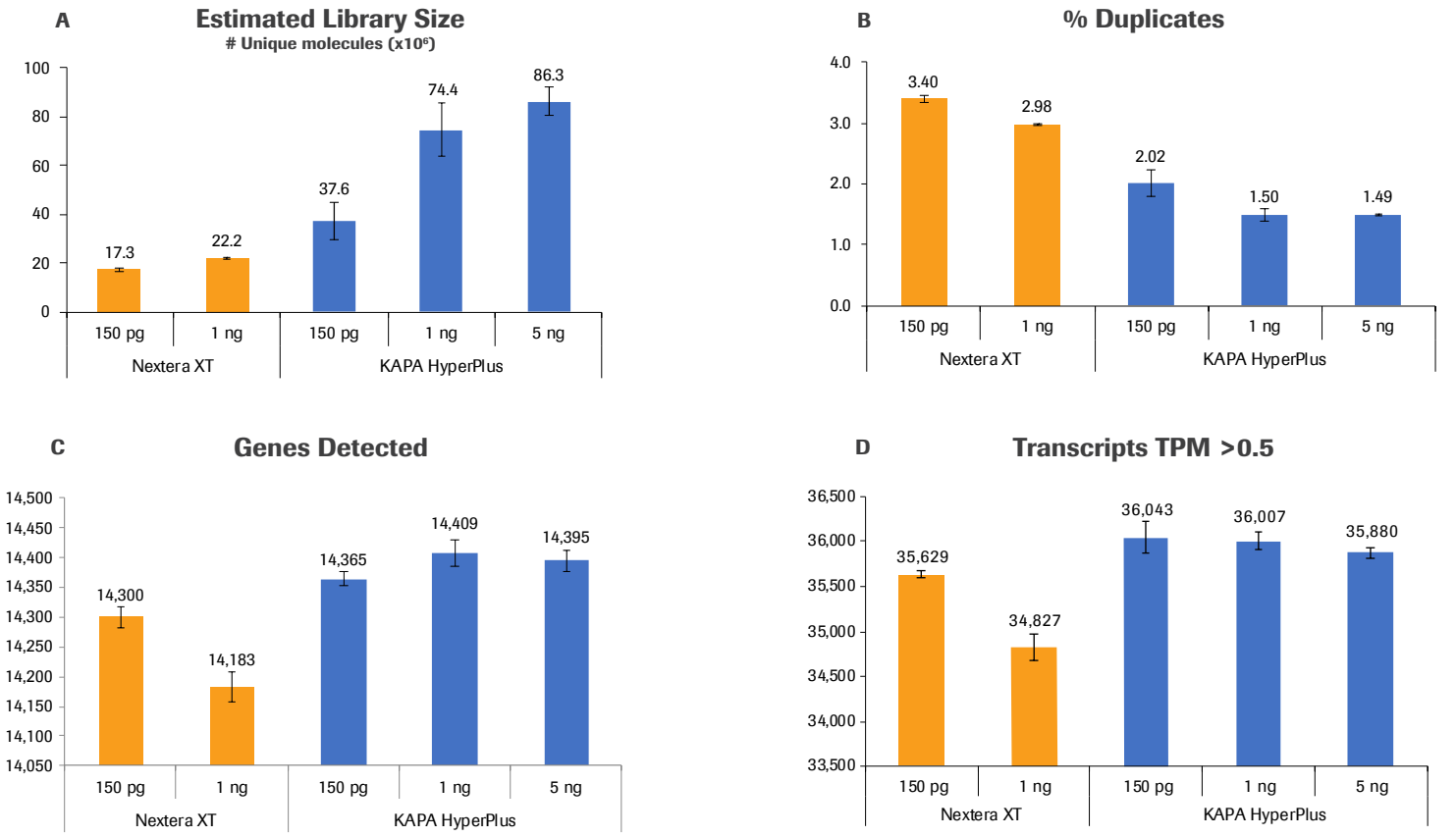

Figure 5. KAPA HyperPlus Kit yields more complex libraries than Nextera XT DNA Library Preparation Kit. (A) Estimated library size, represented as the number of unique molecules in the library. **(B)** Percent of duplicate reads. **(C)** Number of genes detected. **(D)** Number of transcripts expressed above the default minimum expression level of 0.5 TPM (transcripts per million). Error bars represent the standard deviation for three replicates.

Libraries generated with the KAPA HyperPlus Kit displayed consistent library size distribution regardless of cDNA input amounts, whereas the size distribution of libraries produced using the Nextera XT Kit was more variable **(Figure 3)**. The consistent size distribution achieved with the KAPA HyperPlus Kit across a 30-fold range of input amounts suggests that this method could help to alleviate the bottleneck of cDNA normalization prior to library construction, usually required for workflows using SMART-Seq cDNA synthesis.

Libraries prepared with KAPA HyperPlus Kit yielded high-quality sequencing data from 150 pg to 5 ng cDNA inputs

To evaluate how cDNA input quantities impact the performance of the KAPA HyperPlus Kit and how the resulting sequencing data compare against data from libraries prepared with the Nextera XT Kit, sequencing results were analyzed with respect to the following metrics: mapping rate, percent of ribosomal RNA (rRNA) reads, and genomic rates **(Figure 4)**; and library size, percent of duplicate reads, and number of genes and transcripts identified **(Figure 5)**. For all inputs the mapping efficiency exceeded 95%, the percentage of reads from residual rRNA was negligible—as expected for workflows that utilize oligo(dT) priming during cDNA synthesis—and the genomic rates remained consistent across both workflows with KAPA HyperPlus and Nextera XT Kits **(Figure 4)**.

When the number of unique molecules identified (an indicator of library size) was compared across cDNA inputs and between workflows **(Figure 5A)**, the results showed that (1) greater input amounts yielded more unique molecules, and (2) at each cDNA input amount, libraries constructed using the KAPA HyperPlus Kit were more diverse than those constructed using the Nextera XT Kit. For example, 150 pg of cDNA generated, on average, twice as many unique molecules when used as input into the KAPA HyperPlus Kit compared to the Nextera XT Kit, and 1 ng of cDNA yielded three times as many unique molecules with the KAPA HyperPlus Kit. The KAPA HyperPlus Kit also outperformed Nextera XT Kit with respect to percent duplicate reads, the number of genes identified, and the number of transcripts detected **(Figure 5B–D)**.

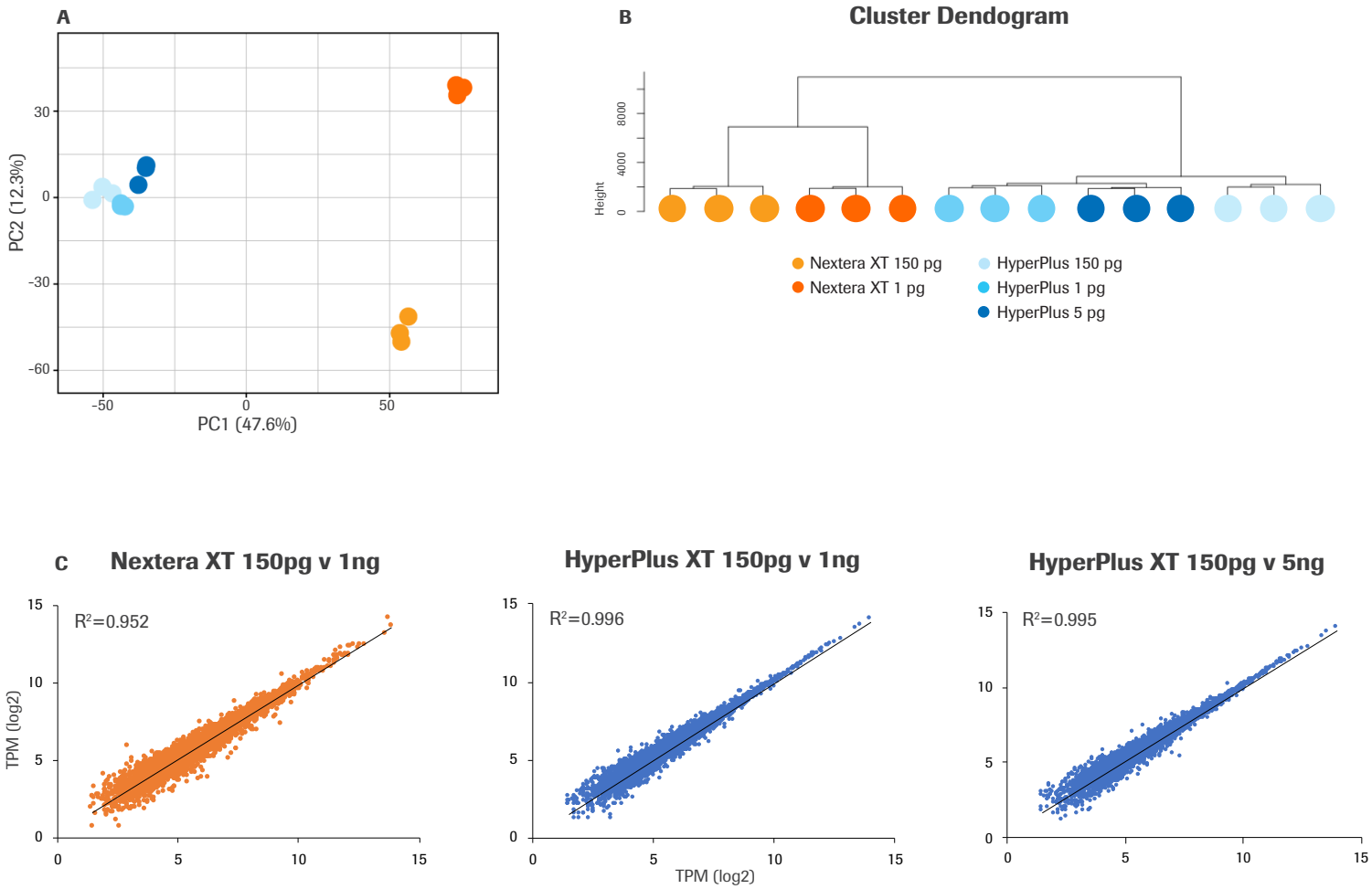

Figure 6. Libraries constructed from the KAPA HyperPlus Kit yield more uniform gene expression data across cDNA input quantities compared to Nextera XT. (A) Principal component analysis of gene expression data, calculated by single value decomposition. **(B)** Hierarchical clustering performed using EdgeR. **(C)** Pearson's correlation of normalized TPM counts.

In summary, libraries prepared with the KAPA HyperPlus Kit showed consistent sequencing metrics across cDNA inputs, with equivalent or better performance than the tagmentation-based Nextera XT workflow. Additionally, the KAPA HyperPlus Kit is compatible with a wider range of DNA input quantities, providing greater flexibility than the Nextera XT Kit, which is limited to a maximum DNA input of 1 ng.

The KAPA HyperPlus Kit yielded reproducible gene expression across a 30-fold range of cDNA input amounts

Global transcript expression values were compared between workflows and across cDNA inputs by principal component analysis (PCA) and hierarchical clustering. Libraries created with the Nextera XT Kit from different cDNA inputs (150 pg, 1 ng) did not cluster together, while libraries created with the KAPA HyperPlus Kit yielded very similar transcriptome profiles across all input quantities (150 pg, 1 ng, 5 ng) **(Figure 6A)**. Hierarchical clustering of the samples recapitulated the patterns of the PCA analysis, with the KAPA HyperPlus Kit demonstrating the highest degree of similarity across samples **(Figure 6B)**. In addition, gene expression levels from libraries prepared with KAPA HyperPlus Kit using cDNA inputs from 150 pg to 5 ng showed a very high correlation (R^2>0.995), while the correlation for libraries generated with the Nextera XT Kit using different cDNA inputs was slightly lower (R^2=0.952) **(Figure 6C)**. These results indicate that the KAPA HyperPlus Kit produced libraries that were functionally equivalent across cDNA inputs and more uniform than libraries created using the Nextera XT Kit.

The KAPA HyperPlus Kit generated high-quality libraries from mouse single-cell lysates without cDNA input normalization

The high-quality results obtained from ultra-low-input RNA-sequencing across a 30-fold range of input amounts suggested that the KAPA HyperPlus Kit in conjunction with the SMART-Seq strategy could be applied to single cells without the need for cDNA input normalization. To assess the performance of the KAPA HyperPlus Kit in a single-cell RNA-seq application without input normalization, single cells derived from mouse were isolated, processed into sequencing libraries, and sequenced **(see Methods and Figure 1B)**.

Table 1 summarizes the characteristics of final libraries generated from 33 isolated single cells. The amount of amplified cDNA derived from these samples varied widely, from <100 pg to ~22 ng. Instead of normalizing the cDNA input prior to library prep, an equivalent volume (10 uL) of each sample was used as input into the KAPA HyperPlus Kit; libraries were then prepared as described in the Methods. All samples yielded libraries with the expected size distribution and sufficient yields for sequencing (>0.5 nM, the minimum library concentration recommended for the Illumina NextSeq 500 platform) **(Table 1 and Figure 7)**.

Libraries generated from single-cell lysates yielded high-quality sequencing metrics, with more than 88% of clusters passing quality filters and 85% of bases showing quality score ≥Q30. Following initial analysis, sequencing data were split into two groups based on cDNA input quantity, as shown in the **Figure 8**. The lower-input group (n=14), represented in green, contained libraries generated with inputs ≤ 5 ng of cDNA (within the range of cDNA inputs tested previously), and the higher-input group (n=19), represented in blue, contained libraries generated with >5 ng of cDNA, which is higher than the recommended maximum input for the Nextera XT method (1ng), but within the acceptable range for the KAPA HyperPlus Kit (1ng to 1,000 ng).

Both input groups yielded consistent sequencing metrics. The average number of total reads was above 1.6M reads, and the mapping rate was ~0.80; this averages >73% of unique mapped reads in both groups **(Figure 8A)**. The genomic rates **(Figure 8B)**, the number of genes detected, and the number of transcripts with expression level above 0.5 TPM **(Figure 8C)** were also similar across cDNA inputs. Although four samples from the higher-input group yielded lower-than-average mapping rates, 29 of the 33 scRNA-seq libraries generated with the KAPA HyperPlus Kit yielded high-complexity and consistent libraries across all cDNA input quantities tested. The mapping results

Table 1. cDNA input quantities and final library characteristics of mouse single-cell samples (n=33).

	cDNA input (ng)	Final library size (bp)	Final library yield (nM)
Mean (±SD)	6.8 (±3.6)	363 (±13.1)	3.6 (±1.2)
Min	<0.1 *	335	1.5
Max	21.8	392	6.6

*below the level of detection of the Qubit DNA High Sensitivity assay; the cDNA input was assumed to be below 0.1 ng.

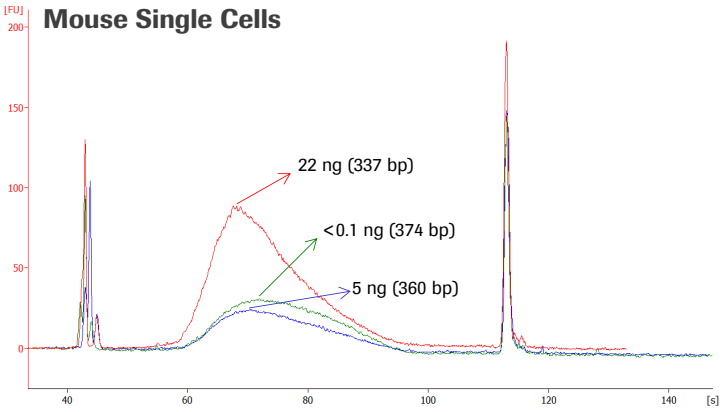

Figure 7. Size distribution of final libraries from mouse single cells was uniform across variable cDNA input amounts. Final amplified library traces for representative samples, assessed on a 2100 Bioanalyzer (Agilent).

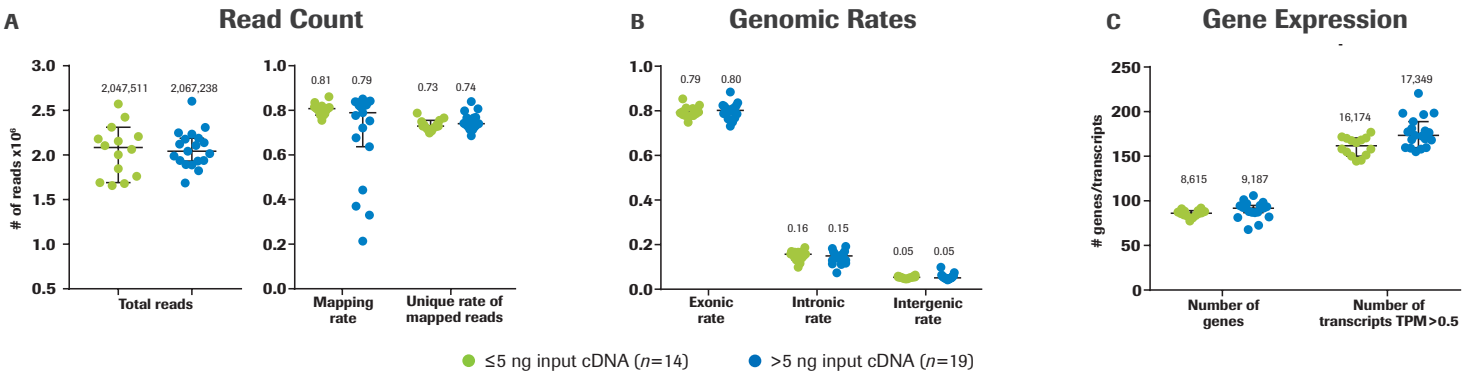

Figure 8. Sequencing metrics from single cells are consistent across cDNA input quantities. Results are grouped according to the amount of cDNA input into the KAPA HyperPlus Kit. The following sequencing metrics are shown: **(A)** Read count metrics; **(B)** Genomic rates; and **(C)** Gene expression analysis based on the number of genes detected and number of transcripts expressed above the default minimum expression level of 0.5 TPM (transcripts per million). Data labels show the median value of the metric in each input-group.

from these four samples suggest that a substantial portion of the measured DNA in these samples was composed of DNA oligo concatamers, rather than true cDNA; the formation of such concatamers during template switching can lead to artificially high cDNA readings, especially in samples with very low input RNA. **(Figure 8C)**.

PCA analysis of transcriptional profiling data showed that data generated by both input groups clustered similarly **(Figure 9A)**. The four blue dots in the far left of the PCA plot are the samples that showed low mapping rates **(Figure 8A)**. In the hierarchical clustering diagram and the heatmap of gene expression levels **(Figure 9B)**, samples from both groups were distributed across the entire dendogram rather than clustered by input quantity, demonstrating that the transcriptional profile of the single-cell samples was not affected by the wide range of cDNA inputs used. Additionally, correlation analysis showed an average R^2 of 0.84 **(Figure 9C)**, which is typical of scRNA-seq experiments.[5] These results show consistency in single-cell transcriptome data generated with the KAPA HyperPlus Kit across a variety of cDNA inputs, complementing the results obtained from ultra-low-input RNA sequencing using the same workflow with known amounts of input cDNA.

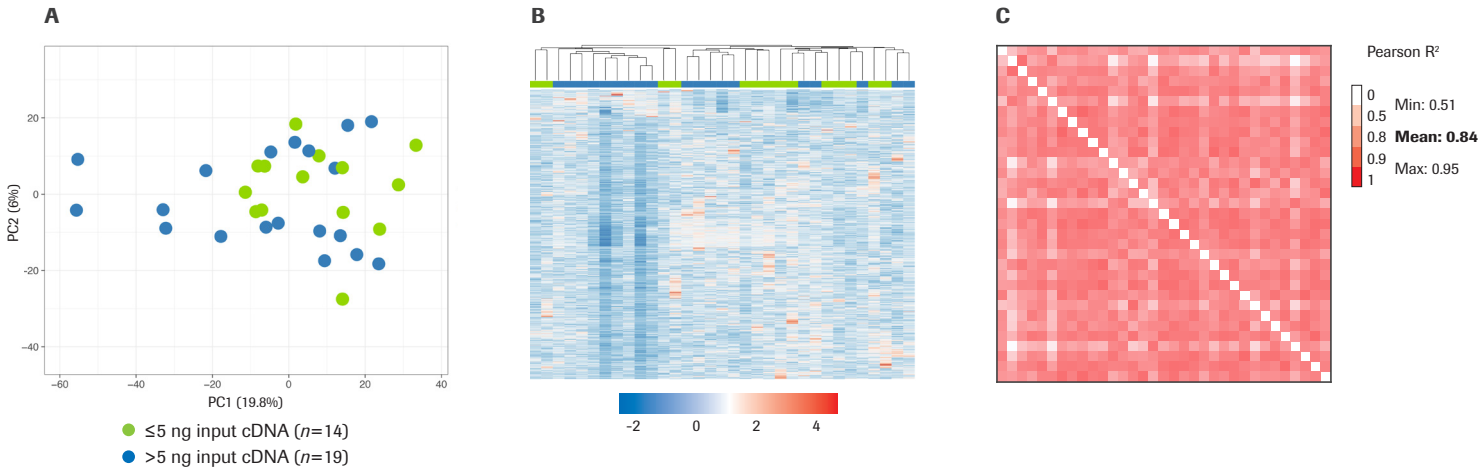

Figure 9. KAPA HyperPlus Kits combined with SMART-Seq technology yield highly reproducible single-cell RNA libraries across a wide range of cDNA input quantities. (A) Principal components analysis (PCA) of transcription levels for all single-cell libraries, grouped according to the cDNA input quantities prior to library preparation. **(B)** Heatmap and hierarchical clustering of expressed genes. Rows and columns are clustered using correlation distance and average linkage. Green and blue bars on top represent the input group of each sample. **(C)** Pairwise Pearson's correlation analysis for all samples. The scale bar represents the range of the R^2 displayed.

Conclusions

The KAPA HyperPlus Kit was successfully combined with SMART-Seq technology using RNA inputs as low as 10 pg. This workflow outperformed the Nextera XT tagmentation-based strategy, generating libraries with higher complexity and greater gene expression concordance across a 30-fold cDNA input range into library preparation. In addition, the KAPA HyperPlus Kit yielded high-quality libraries and sequencing data from mouse single cells, further demonstrating that the KAPA HyperPlus Kit is a robust and flexible workflow that can eliminate the bottleneck of cDNA normalization in scRNA-seq, enabling the development of reproducible higher-throughput workflows.

References

1. Takara Bio USA. SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing User Manual. 2017.
2. Lan J, Yin Y, Reed E et al. Impact of three Illumina library construction methods on GC bias and HLA genotype calling. *Human Immunology* 2015; 76:166–175.
3. Cell Microsystems. RaftNote. DuVall J, Trotta N, McClellan R, La Dine MA, Gebhart S. Imaging and Sorting Living Cells Stained with Vital Dyes on Cell Microsystems' CytoSort Array and Automated CellRaft AIR System. 2017.
4. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research* 2015; 43(W1):W566–70.
5. Ding J, Adiconis X, Simmons SK et al. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nature Biotechnology* 2020; 38:737–746.

For more information about Roche RNA-seq products and solutions, please visit: **sequencing.roche.com/RNA-seq**

Published by:
Roche Sequencing and Life Science
9115 Hague Road
Indianapolis, IN 46256
sequencing.roche.com

For more information on the CellRaft AIR System and the range of suported applications please visit: **www.cellmicrosystems.com** or email **info@cellmicrosystems.com**

MC-US-09371 APP105001 A552 5/21