

TECH NOTE

SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian: Improved ease of use and sequencing performance for whole transcriptome analysis of high-quality or degraded samples

A versatile solution for low-input total RNA-seq

Kit accommodates 250 pg–10 ng inputs of total RNA of any quality (intact or degraded).

Superior sequencing performance

Reconfigured sequencing libraries perform well on all Illumina platforms (including NextSeq® and HiSeq® 3000/4000), achieving a high %PF without the addition of PhiX.

Improved ease of use

New PCR buffer formulation makes bead purification easier.

High-quality data from challenging samples

Kit yields improved sequencing metrics from FFPE RNA.

Introduction

Obtaining an accurate portrait of expression levels for coding and non-coding RNAs from small sample inputs carries potential for both the fulfillment of basic research objectives and the development of novel therapeutics and clinical diagnostic solutions. While next-generation sequencing (NGS) technology has contributed greatly to our understanding of cellular mRNA composition and dynamics, it has also revealed the existence of a vast assortment of non-coding RNAs that play diverse roles in processes such as gene expression regulation (Mattick and Makunin 2006; Komienko et al. 2013), and are implicated in the development of various human diseases (Hindorf et al. 2009; Wapinski and Chang 2011). Whereas oligo(dT) priming is typically used to capture polyadenylated mRNA for NGS, random priming allows for capture of both coding and non-coding RNA and is often the only feasible option available for processing degraded RNA inputs, such as those obtained from formalin-fixed, paraffin-embedded (FFPE) samples or liquid biopsies. However, a significant challenge associated with random priming is that it also captures ribosomal and mitochondrial RNA molecules, which are typically present in great abundance but not of interest to researchers.

To enable NGS-based analysis of coding and non-coding RNA (i.e., total RNA-seq) from picogram inputs, we previously developed the [SMARTer Stranded Total RNA-Seq Kit - Pico Input Mammalian](#) (referred to below as “Pico v1”), which incorporates a novel technology that enables removal of ribosomal cDNA following cDNA synthesis (as opposed to direct removal of corresponding rRNA molecules prior to reverse transcription).

In keeping with our tradition of continuously refining and improving the performance of our products, we have subsequently developed the [SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian](#) (referred to as “Pico v2”; see workflow in Figure 1). Features that distinguish the Pico v2 kit from its predecessor include superior sequencing performance—particularly for NextSeq and MiniSeq™ instruments that use two-channel SBS technology and for HiSeq 3000/4000—and a new PCR buffer formulation enabling a more user-friendly library-purification process.

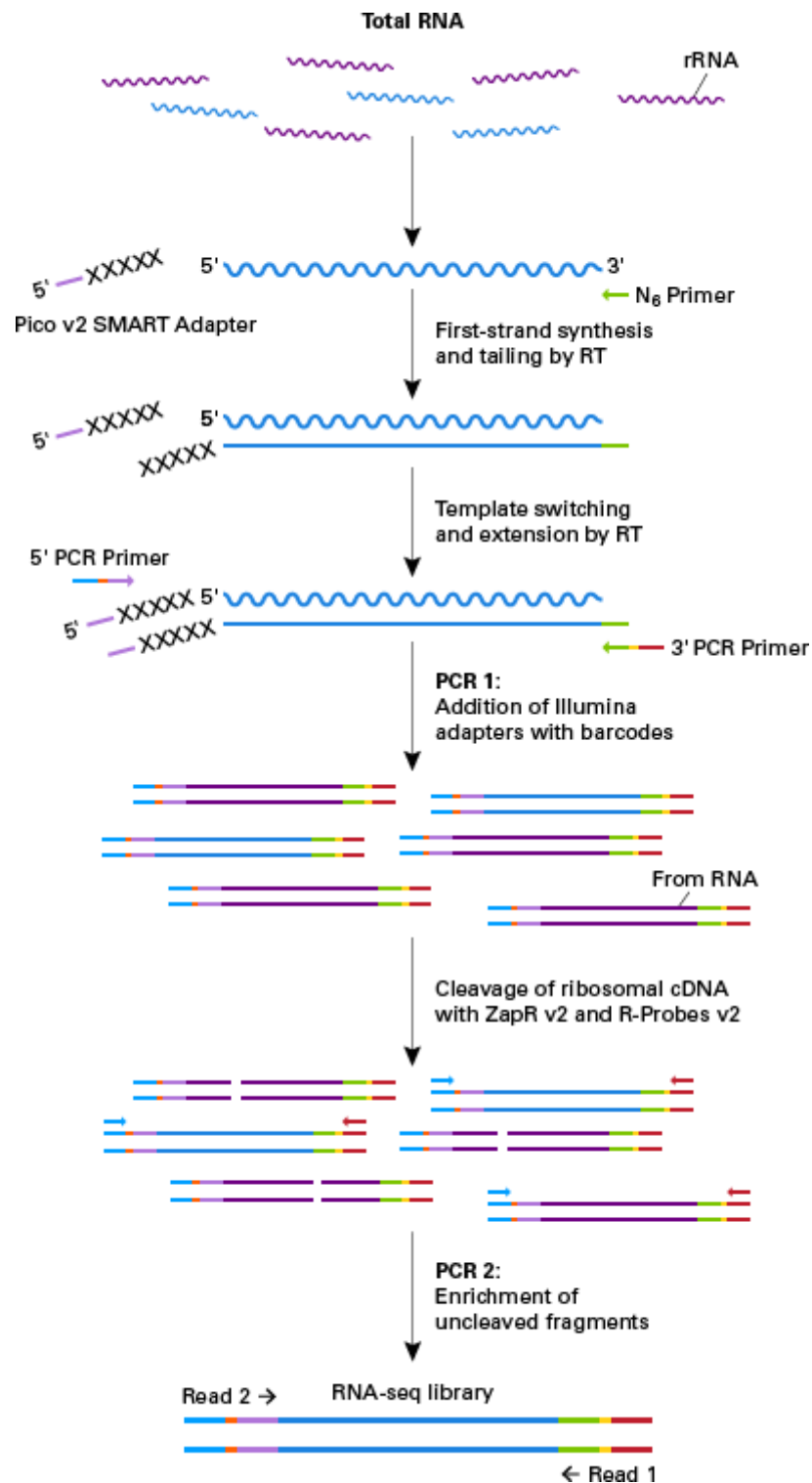


Figure 1. Schematic of technology in the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian. SMART technology is used in this ligation-free protocol to preserve strand-of-origin information. Random priming (represented as the green N₆ Primer) allows the generation of cDNA from all RNA fragments in the sample, including rRNA. When the SMARTScribe Reverse Transcriptase (RT) reaches the 5' end of the RNA fragment, the enzyme's terminal transferase activity adds a few non-templated nucleotides to the 3' end of the cDNA (shown as Xs). The carefully designed Pico v2 SMART Adapter

(included in the SMART TSO Mix v2) base-pairs with the non-templated nucleotide stretch, creating an extended template to enable the RT to continue replicating to the end of the oligonucleotide. The resulting cDNA contains sequences derived from the random primer and the Pico v2 SMART Adapter used in the reverse transcription reaction. In the next step, a first round of PCR amplification (PCR1) adds full-length Illumina adapters, including barcodes. The 5' PCR Primer binds to the Pico v2 SMART Adapter sequence (light purple), while the 3' PCR Primer binds to sequence associated with the random primer (green). The ribosomal cDNA (originating from rRNA) is then cleaved by ZapR v2 in the presence of the mammalian-specific R-Probes v2. This process leaves the library fragments originating from non-rRNA molecules untouched, with priming sites available on both 5' and 3' ends for further PCR amplification. These fragments are enriched via a second round of PCR amplification (PCR2) using primers universal to all libraries. The final library contains sequences allowing clustering on any Illumina flow cell (see details in Figure 2).

The improved sequencing performance provided by the Pico v2 kit is due to reconfiguration of the resulting sequencing libraries (Figure 2). Libraries produced with the Pico v2 kit are generated such that bases corresponding to the random-priming site (located at 3' end of each RNA molecule) are read at the beginning of Read 1, while bases corresponding to nontemplated nucleotides added during the template-switching process are read at the beginning of Read 2. This is essentially a reverse orientation relative to libraries generated with the original version of the kit (in which bases associated with template-switching are read at the beginning of Read 1). The reconfigured libraries produced by the Pico v2 kit provide greater nucleotide diversity at the beginning of Read 1. This in turn eliminates the necessity of adding significant amounts of PhiX control library to the sequencing reaction to achieve a higher percentage of clusters passing filter (%PF), yielding more meaningful data per sequencing run and reducing sequencing costs.



Figure 2. Structural features of final libraries generated with the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian. The adapters added using 5' PCR Primer HT and 3' PCR Primer HT contain sequences allowing clustering on any Illumina flow cell (P7 shown in light blue, P5 shown in red), Illumina TruSeq® HT indexes (Index 1 [i5] sequence shown in yellow, and Index 2 [i7] sequence shown in orange), as well as the regions recognized by sequencing primers Read Primer 2 (Read 2, purple) and Read Primer 1 (Read 1, green). Read 1 generates sequences antisense to the original RNA, while Read 2 yields sequences sense to the original RNA (orientation of original RNA denoted by 5' and 3' in dark blue). The first three nucleotides of the second sequencing read (Read 2) are derived from the Pico v2 SMART Adapter (shown as Xs). These three nucleotides must be trimmed prior to mapping if performing paired-end sequencing.

Results

Improved sequencing performance with the Pico v2 kit

Even with an industry-leading product such as the SMARTer Stranded Total RNA-Seq Kit - Pico Input Mammalian, there is always room for improvement. As described above, a limitation of the Pico v1 kit is that it generates sequencing libraries with relatively low nucleotide diversity at the beginning of Read 1. This low nucleotide diversity results from the nontemplated nucleotides that facilitate adapter binding and incorporation via the template-switching mechanism (see Figure 1, above). Having low nucleotide diversity at the beginning of Read 1 poses challenges for sequencing because the first 25 sequencing cycles are used to determine which clusters pass filtering, and is particularly problematic on platforms using two-channel SBS technology (e.g., NextSeq and MiniSeq). Challenges associated with low library diversity can be mitigated by spiking in a suitable amount of PhiX control library—we recommend adding PhiX at concentrations as high as 30% depending on the platform—however this reduces the amount of relevant sequencing reads generated per sequencing run, consuming time and increasing sequencing costs.

To demonstrate the improved sequencing performance of the Pico v2 kit vs. the original kit, sequencing libraries were generated from various inputs of total RNA using each kit according to the corresponding user manuals and sequenced on both NextSeq and MiniSeq platforms (Figure 3). Whereas libraries generated with the Pico v1 kit yielded %PF values of 81.3% and 77.1% and quantities of reads passing filter that met or approached established benchmarks for NextSeq and MiniSeq instruments, respectively, libraries generated with the Pico v2 kit achieved %PF values of 88.3% and 90.5%, with quantities of reads passing filter that exceeded performance specifications for each platform by a considerable margin. These results demonstrate that the Pico v2 kit provides superior sequencing performance relative to Pico v1.

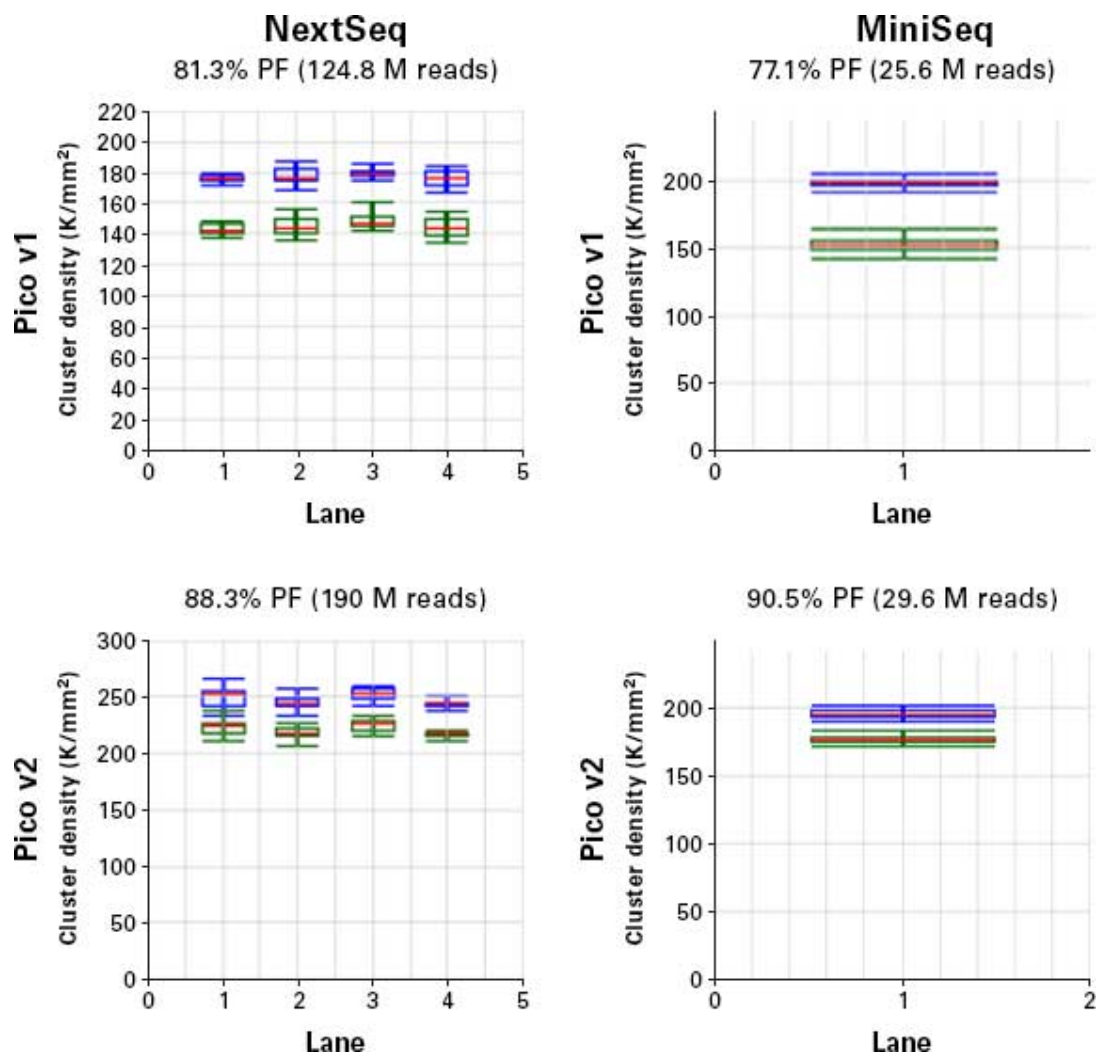


Figure 3. Improved pass-filter rates (%PF) with the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian. Libraries generated with the Pico v1 or Pico v2 kits were pooled and run on NextSeq 500 or MiniSeq instruments, as indicated. For each graph, blue boxplots indicate the distribution of cluster densities for unfiltered (i.e., raw) reads, while the green boxplots indicate the distribution of cluster densities for reads that passed filtering. Quantities of reads passing filter (in millions) and %PF values for each sequencing run are included above each graph. The expected number of reads passing filter according to Illumina specifications was 130 million reads for runs on the NextSeq and 25 million reads for runs on the MiniSeq. Proportions of reads that aligned to PhiX sequences ranged from 0.5% to 1.15% for all sequencing runs. As indicated in the graphs, libraries generated with the Pico v2 kit achieved higher %PF values for both Illumina platforms relative to libraries generated with the Pico v1 kit, and yielded quantities of reads passing filter that greatly exceeded the Illumina specifications.

Improved ease of use during library purification with the Pico v2 kit

As with many NGS library prep kits, Pico v1 and Pico v2 both employ magnetic AMPure beads for multiple library purification steps. Customer feedback regarding the Pico v1 kit indicated that formation, drying, and resuspension of bead pellets during library purification was a common pain point in the kit workflow. To address this, we optimized the PCR buffer for greater compatibility with AMPure bead purification while maintaining its performance for PCR. The new buffer formulation, SeqAmp CB PCR Buffer (CB = “compatible with beads”), allows for the beads to separate more quickly, yielding a tighter bead pellet that dries more uniformly and is easier to resuspend (Figure 4).

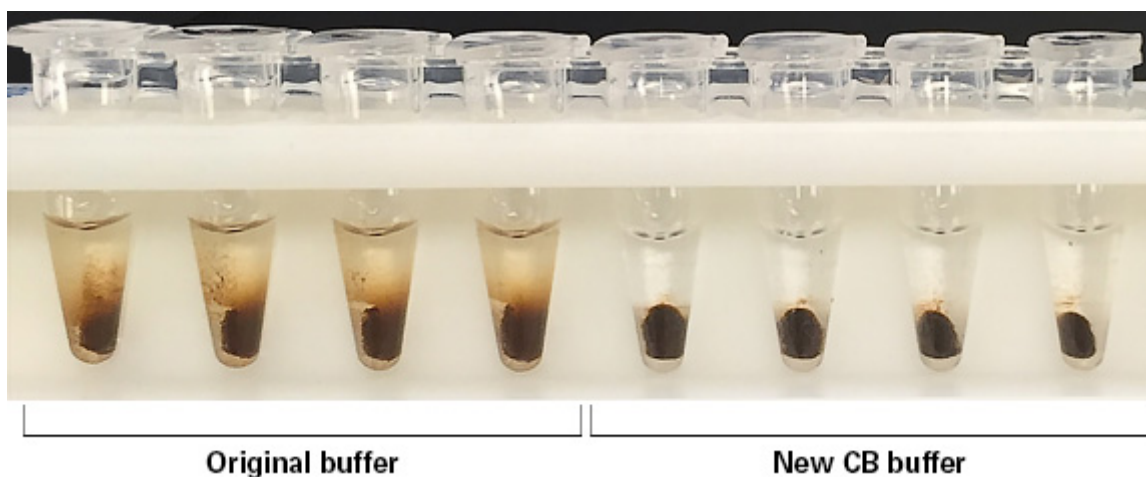


Figure 4. Improved bead-pellet formation with new SeqAmp CB PCR Buffer. The PCR buffer included in the Pico v2 kit was re-formulated to allow for faster, tighter bead-pellet formation. Following magnetic separation for a fixed period, bead pellets formed in the new SeqAmp CB PCR Buffer (right) are tighter than those formed in the original PCR buffer (left). Tighter bead pellets tend to dry more evenly and are easier to resuspend than pellets that are broader and more diffuse.

Comparison of sequencing metrics for FFPE samples processed with Pico v1 and Pico v2

To further demonstrate the enhanced capabilities of the Pico v2 kit relative to its predecessor, particularly for analysis of challenging samples, sequencing libraries were generated from 1-ng and 10-ng inputs of human lung total RNA (DV200 = 68%) obtained from FFPE tissue and sequenced on a NextSeq 500 instrument. In comparison with Pico v1, library yields from the Pico v2 kit were considerably greater for both input amounts (Figure 5A). For the 1-ng input amount, sequencing data for the Pico v2 library identified thousands more transcripts than data for the Pico v1 library, whereas numbers of transcripts identified were comparable at the 10-ng input level. In contrast with the data generated using Pico v1, numbers of transcripts identified for 1-ng and 10-ng inputs using Pico v2 were very similar, suggesting that Pico v2 offers superior sensitivity for detection of low-abundance transcripts in low-input samples.

Proportions of reads mapping to various RNA species were comparable across kits and input amounts, however libraries generated with the Pico v2 kit yielded a lower proportion of reads mapping to rRNA and mtRNA relative to the Pico v1 libraries. For both input amounts, the duplicate rate was lower for Pico v2 libraries, and for the 10-ng input in particular the duplicate rate was ~50% lower. Comparison of transcript expression levels across input amounts for each version of the kit indicated that the correlation was much stronger for the Pico v2 libraries vs. the Pico v1 libraries (Pearson = 0.96 and Spearman = 0.83 vs. Pearson = 0.91 and Spearman = 0.67, Figure 5B). These results suggest that Pico v2 outperforms Pico v1 by providing higher library yields, improved sensitivity, reduced representation of rRNA and mtRNA sequences, and a stronger correlation in gene expression measurements across input amounts.



A

Sequencing Alignment Metrics for 1-ng and 10-ng Inputs of Total RNA				
Kit	Pico v1	Pico v2	Pico v1	Pico v2
RNA source	Human lung FFPE total RNA			
Input amount (ng)	1		10	
Library yield (ng/μl)	0.4	3.2	4.4	21.7
Number of reads (millions)	8.25 (paired-end reads)			
Number of transcripts >1 FPKM	8,481	9,916	10,096	9,878
Number of transcripts >0.1 FPKM	14,347	19,594	20,724	21,325
Proportion of reads (%)				
<i>Exonic</i>	15.9	15.0	16.4	14.9
<i>Intronic</i>	50.5	53.9	54.9	57.9
<i>Intergenic</i>	12.1	12.1	12.8	12.9
<i>rRNA</i>	15.0	13.3	10.3	9.2
<i>Mitochondrial</i>	1.3	0.9	1.5	0.7
Duplicate rate (%)	79.9	67.2	60.1	34.3

B

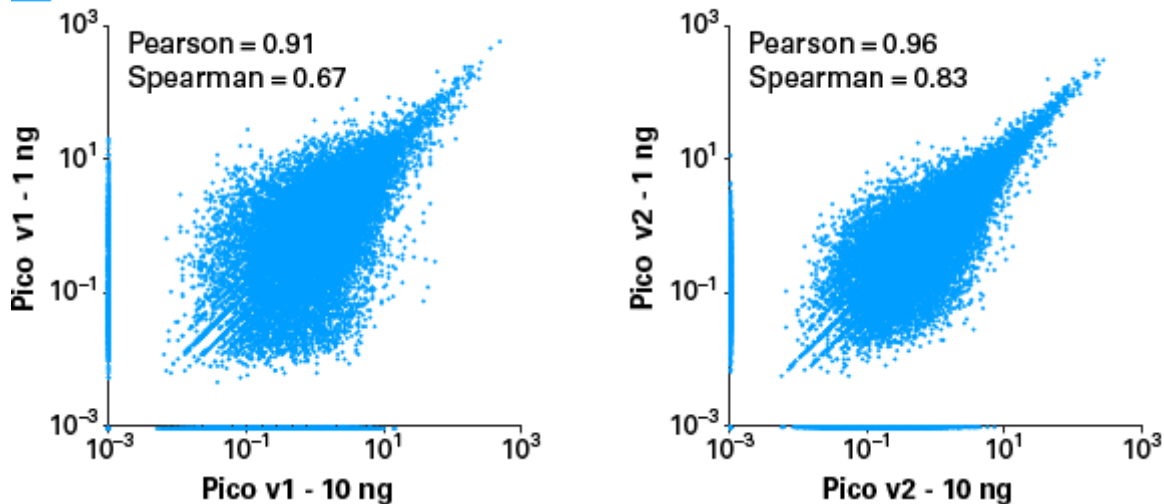


Figure 5. Improved sensitivity and reproducibility with the SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian. Sequencing libraries were generated from 1-ng and 10-ng inputs of total RNA extracted from human lung FFPE tissue using both the Pico v1 and Pico v2 kits, and sequenced on a NextSeq 500 instrument. **Panel A. Sequencing metrics for**

libraries generated from 1-ng or 10-ng inputs using each kit. For both input amounts, the Pico v2 kit resulted in greater library yields, a lower proportion of reads mapping to rRNA and mtRNA, and a lower duplicate rate. For the 1-ng input, sequencing data from the Pico v2 library also identified thousands more transcripts than sequencing data from the Pico v1 library, indicating a higher sensitivity for Pico v2. **Panel B. Comparison of transcript expression levels across input amounts.** Higher reproducibility was observed between 1-ng and 10-ng inputs for data generated with the Pico v2 kit vs. data generated using the Pico v1 kit. FPKM values are shown on a Log_{10} scale. Transcripts represented in only one library can be seen along the X- and Y-axes of the scatter plots.

Summary

To better serve the scientific community, we have incorporated several design improvements into the [SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian](#) that provide superior sequencing performance and a more user-friendly workflow relative to its predecessor. Sequencing libraries generated with the Pico v2 kit demonstrate a higher %PF rate relative to libraries produced with the original kit while requiring little or no addition of PhiX. This improvement will allow researchers to extract more meaningful data from each sequencing run, saving time and conserving resources. The Pico v2 kit also outperforms the Pico v1 kit by providing higher library yields, improved sensitivity, and greater consistency across input amounts, even for challenging samples obtained from FFPE tissue. Optimization of the PCR buffer included with the kit has streamlined the various bead-purification steps, which should also help reduce operational costs for labs performing RNA-seq at high throughput.

Methods

Comparison of pass-filter rates for Pico v1 and Pico v2 libraries

To compare the %PF rates for libraries generated with the Pico v1 and Pico v2 kits, sequencing libraries were generated from varying input types and amounts of total RNA and pooled together. Pools of sequencing libraries were run on the NextSeq 500 using the NextSeq 500/550 Mid Output Kit v2 (150 cycles; Cat. # FC-404-2001) with 2 x 75-bp paired-end reads, and on the MiniSeq using the MiniSeq High Output Kit (75 cycles; Cat. # FC-420-1001) with 2 x 38-bp paired-end reads.

Comparison of sequencing metrics for FFPE samples

To evaluate the performance of the Pico v1 and Pico v2 kits with FFPE samples, total RNA was extracted from a 5- μm curl of FFPE human lung tissue (Cureline) using a NucleoSpin totalRNA FFPE kit (Takara Bio, Cat. # 740982.10). Prior to library preparation, RNA integrity was evaluated on an Agilent Bioanalyzer using an Agilent RNA 6000 Pico Kit (Cat. # 5067-1513), yielding a DV200 value of 68%. Libraries were generated from the extracted RNA using both the Pico v1 and Pico v2 kits without additional RNA fragmentation (protocol option 2). Libraries were sequenced on a NextSeq 500 using the NextSeq 500/550 Mid Output Kit v2 and resulting sequencing datasets were downsampled to 8.25 million paired-end reads.

Sequence analysis

Reads from all libraries were trimmed and mapped to mammalian rRNA and the human mitochondrial genomes using CLC Genomics Workbench. The remaining reads were subsequently mapped using CLC to the human (hg19) genomes with RefSeq annotation. All percentages shown, including the number of reads that map to introns, exons, or intergenic regions, are percentages of the total reads in the library. The number of transcripts identified in each library was determined by the number of transcripts with an FPKM greater than or equal to 1 or 0.1, as shown in Figure 5A. Scatter plots were generated using FPKM values from CLC mapping to the transcriptome. To identify transcripts found in only one replicate (dropouts), 0.001 was added to each value prior to graphing.



References

Hindorf, L. a *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9362–7 (2009).

Kornienko, A. E., Guenzl, P. M., Barlow, D. P. & Pauler, F. M. Gene regulation by the act of long non-coding RNA transcription. *BMC Biol.* **11**, 59 (2013).

Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum Mol Genet* **15 Spec No**, R17–29 (2006).

Wapinski, O. & Chang, H. Y. Long noncoding RNAs and human disease. *Trends Cell Biol.* **21**, 354–361 (2011).

[View web page >>](#)

http://www.clontech.com/US/Products/cDNA_Synthesis_and_Library_Construction/NGS_Learning_Resources/Technical_Notes/v2_Pico_Input_Total_RNA_Seq_Illumina

This is a reprint from a page on our web site. All license, copyright, and trademark information pertaining to this content applies as stated in the original web content. This information can be found at www.clontech.com.