1015

Clontech Laboratories, Inc.
A Takara Bio Company

SMART® Tools for Strand-Specific Transcriptome Analysis from Mammalian Total RNA

Nathalie Bolduc, Magnolia Bostick, Simon Lee, Tommy Duong, Craig Betts[1], Andrew Farmer*

*Corresponding Author: Andrew_Farmer@clontech.com
[1] Current Affiliation: Cellular Research, Inc., Palo Alto, CA

Abstract

Next-generation sequencing is empowering a deeper understanding of biology by enabling RNA expression analysis over the entire transcriptome with high sensitivity and wide dynamic range. One powerful application within this field is stranded RNA-seq, which is necessary to distinguish overlapping genes and to conduct comprehensive annotation and quantification of long noncoding RNAs.

Commonly used methods for generating strand-specific RNA-seq libraries are plagued by protocols that require several rounds of enzymatic treatments and clean-up steps. This makes them time-intensive, insensitive, and challenging when processing several samples simultaneously. Moreover, generation of RNA-seq libraries from total RNA is challenged by the high amounts of ribosomal RNA (rRNA) in the starting material. Here we present two novel, streamlined workflows, based on Clontech's patented SMART technology, which, together, are able to generate strand-specific RNA-seq libraries from 200 pg to 1 µg of total RNA in less than six hours with minimal carryover rRNA.

For higher inputs (between 10 ng and 1 µg of total RNA), combining SMARTer® Stranded RNA-seq technology with our RNase H-based RiboGone™ rRNA removal system enables depletion of over 95% of rRNA in mammalian samples, while directly producing Illumina®-ready stranded libraries. However, for input below 10 ng, no technology currently allows for rRNA depletion prior to first-strand synthesis. We have circumvented this limitation by modifying the SMARTer Stranded RNA-seq workflow to include a post-library amplification removal of molecules originating from rRNA. This extremely sensitive workflow shows excellent reproducibility within and across input levels ranging from 200 pg to 10 ng, with fewer than 20% of reads mapping back to rRNA. We will show application data illustrating the usefulness of these tools to the identification and quantification of long noncoding RNAs (lncRNAs) over a wide range of inputs.

Introduction

SMARTer Stranded kits are an ideal solution for next-generation sequencing experiments focusing on total RNA, and particularly for identifying non-polyadenylated lncRNAs. These kits use random priming for cDNA synthesis and maintain representation of all transcripts >180 nt. They take advantage of template switching technology to generate robust RNA-seq libraries for Illumina sequencing that retain strand of origin information in a time-saving manner. RiboGone technology for rRNA removal from total RNA is directly built into available higher-input (100 ng–1 µg) SMARTer Stranded kits, and is available as a separate kit for lower-input samples (10–100 ng). The new technique presented here for post-cDNA synthesis removal of library fragments originating from rRNA expands the input range of total RNA that can benefit from SMARTer Stranded RNA-seq technology.

Materials and Methods

Libraries were prepared according to specified protocols using Human Brain Total RNA, Mouse Brain Total RNA, or Mouse Liver Total RNA (Clontech). Sequencing was performed on an Illumina MiSeq® instrument (1.7–2 million 1 x 60 nt single-end reads per library). Bioinformatics, including mapping and determination of RPKM values, was performed using CLC Genomics Workbench. Reads were trimmed and mapped to rRNA and the mitochondrial genome. Unmapped reads were then mapped to the human (hg19) or mouse (mm10) genome with RefSeq masking, or to the GENCODE long non-coding RNA transcripts (version 21 for human and M4 for mouse). For determination of the percent reads mapping to the correct strand (per biological annotation), reads were mapped using STAR against hg19 or mm10 with Ensembl annotation, and the correct strand was defined by Picard analysis.

Conclusions

SMARTer Stranded technology is well suited to RNA-seq library preparation for experiments that require exceptional sensitivity and strand of origin information, like research involving lncRNAs. Here we present two different strategies for targeting and significantly reducing rRNA and mitochondrial rRNA reads. Both strategies generate final sequencing libraries that produce high-quality, reproducible data from a variety of RNA sources. These methods make it possible to identify lncRNA transcripts with high sensitivity and reproducibility, even from extremely low input samples.

- **Wide input range**: 200 pg–10 ng, 10–100 ng, or 100 ng–1 µg total RNA; total RNA with RIN 3–10 (for more information about stranded RNA-seq from degraded samples visit: www.clontech.com/RNA-seq-HI)
- **Effective rRNA removal**: Low percentage of rRNA reads regardless of pre- or post-cDNA synthesis removal strategy
- **Reproducible sequencing metrics**: Both the existing SMARTer Stranded RNA-seq kits and the pico method generate sequencing libraries with excellent mappability, and great representation of lncRNAs.
- **Proven quality**: To see data showing how the inherently stranded SMART reaction preserves strand of origin information visit: www.clontech.com/SMARTer-stranded

For Research Use Only. Not for use in diagnostic or therapeutic procedures. Not for resale. Illumina and MiSeq are registered trademarks of Illumina, Inc. Takara and the Takara logo are trademarks of TAKARA HOLDINGS, Kyoto, Japan. Clontech, the Clontech logo, RiboGone, SMART, and SMARTer are trademarks of Clontech Laboratories, Inc. All other marks are the property of their respective owners. Certain trademarks may not be registered in all jurisdictions.

03.15 US (633751)

1

Method for medium- and high-input (10 ng–1 µg) total RNA

10–100 ng: RiboGone - Mammalian kit & SMARTer Stranded RNA-Seq Kit

100 ng–1 µg: SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian

Method for picogram-input (200 pg–10 ng) total RNA

Pico method—in development

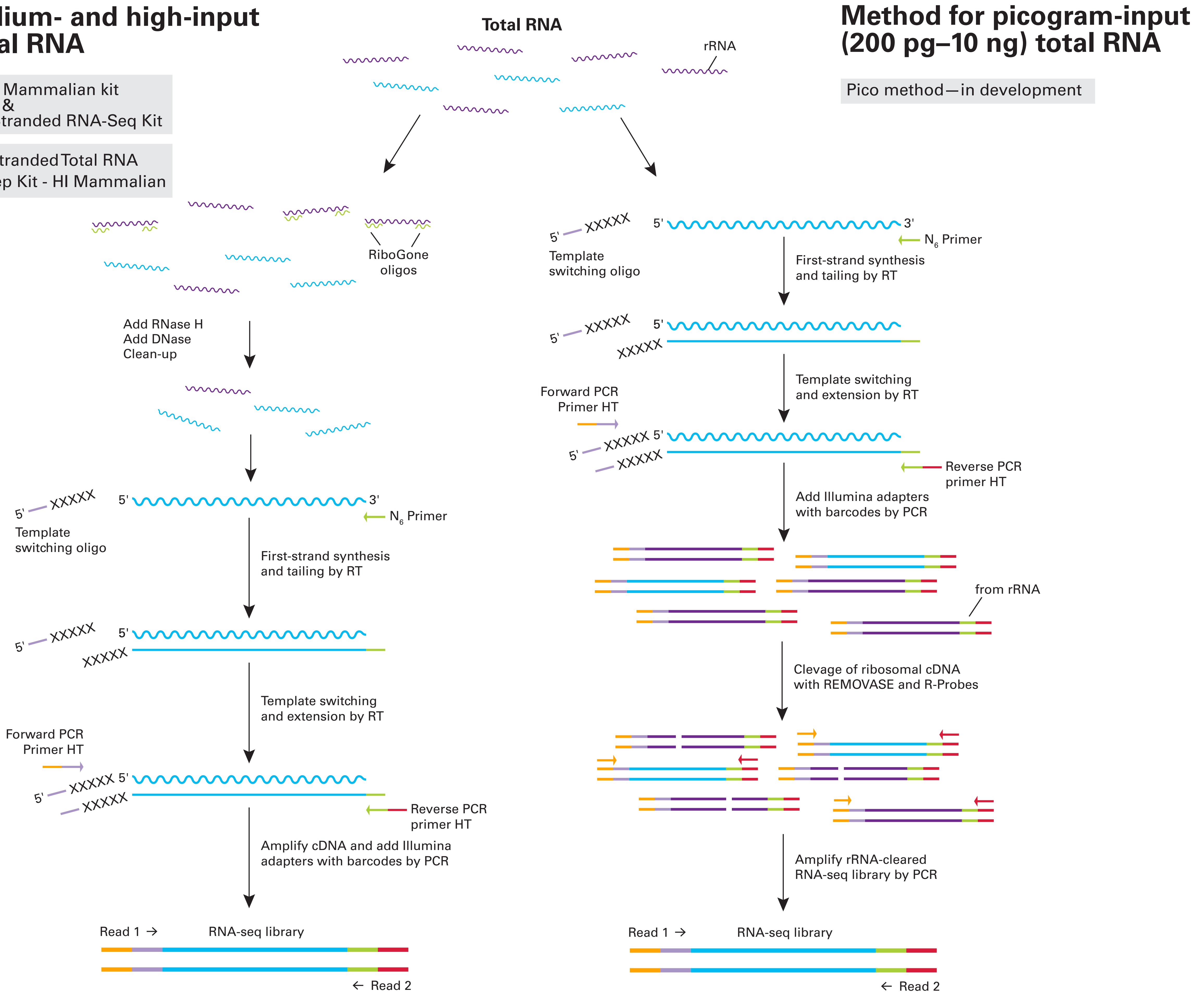
Total RNA
rRNA
RiboGone oligos
Add RNase H
Add DNase
Clean-up
5'
XXXXX
3'
Template switching oligo
N6 Primer
First-strand synthesis and tailing by RT
Template switching and extension by RT
Forward PCR Primer HT
Reverse PCR primer HT
Amplify cDNA and add Illumina adapters with barcodes by PCR
Add Illumina adapters with barcodes by PCR
from rRNA
Clevage of ribosomal cDNA with REMOVASE and R-Probes
Amplify rRNA-cleared RNA-seq library by PCR
Read 1 →
RNA-seq library
← Read 2

Figure 1. Comparison of methods for removal of rRNA reads pre- or post-cDNA synthesis. For total RNA samples >10 ng, rRNA can be removed directly from the input RNA, using specific oligonucleotides and RNase H digestion, prior to cDNA synthesis and sequencing library preparation using SMART (**S**witching **M**echanism **a**t 5´ End of **R**NA **T**emplate) technology. However, very low input (200 pg–10 ng) total RNA samples cannot use this strategy. Our solution for these samples is to prepare RNA-seq libraries from the total RNA sample, then specifically cleave ribosomal cDNA (ds cDNA produced from rRNA), and enrich for intact, non-ribosomal cDNA, fragments by PCR (pico method).

2

Reproducible RNA-seq data generated using the pico method

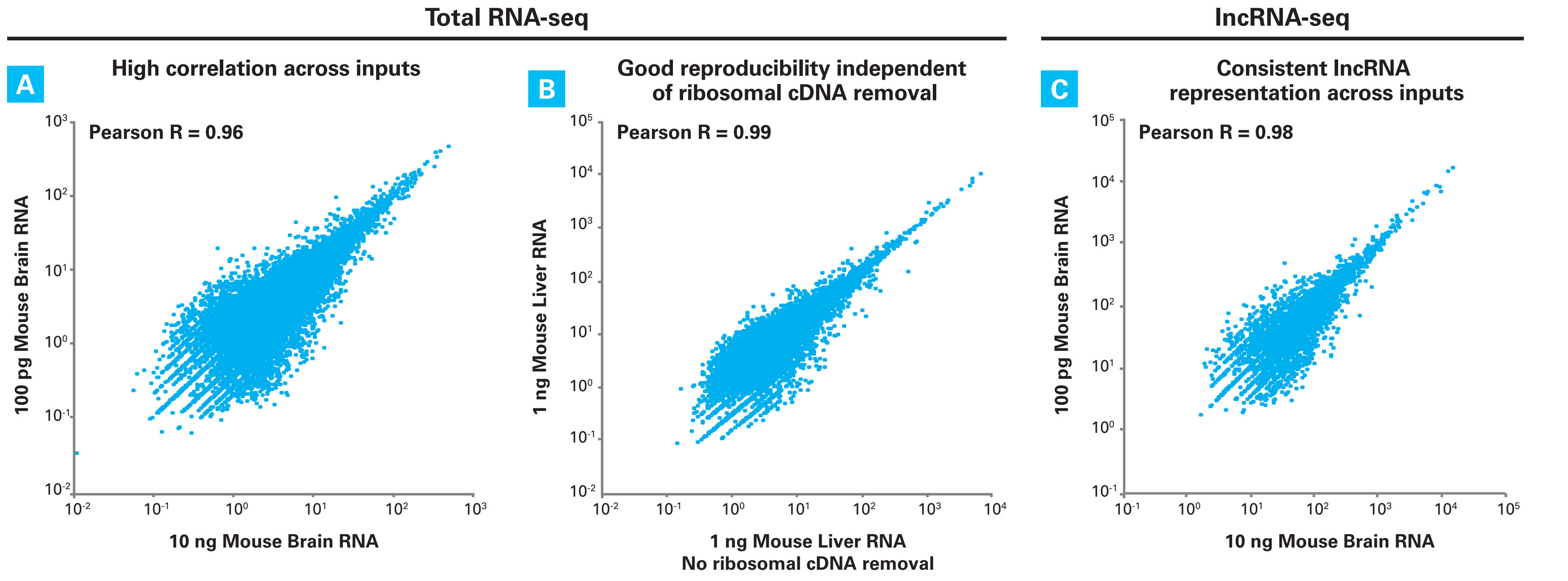

Figure 2. The pico method produces highly reproducible sequencing libraries. Total RNA-seq libraries were generated using the indicated amount of Mouse Brain or Mouse Liver Total RNA using the pico method. RPKMs were compared using scatterplots. **Panel A.** Libraries generated from 100 pg and 10 ng total RNA inputs showed high reproducibility (see more data in Table I). **Panel B.** Comparison between libraries generated with and without ribosomal cDNA removal (sequencing metrics not shown in Table I) showed that no significant biases were introduced by ribosomal cDNA removal. **Panel C.** Reads from libraries shown in Panel A were mapped to lncRNA (annotation from GENCODE M4) and compared across inputs. The high correlation showed that lncRNAs can be reproducibly identified using the pico method.

Table I: Sequencing Metrics from Various Amounts of Total RNA

RNA Source	**Human Brain**				**Mouse Brain**		
Kit or method	Stranded HI	Pico method			Stranded HI	Pico method	
Input amount	1 µg	10 ng	1 ng	1 ng	1 µg	10 ng	100 pg
rRNA/ribosomal cDNA removal?	Yes	Yes	Yes	**No**	Yes	Yes	Yes
Distribution of reads:							
rRNA (%)	0.3	14.5	15.6	**81.6**	0.7	13.7	11.6
Mapped to mitochondrial genome (%)	9.2	11.9	12.0	6.1	9.7	9.6	8.0
Mapped to genome (%)	85.9	65.5	62.4	2.4	86.3	64.7	59.2
Mapped uniquely to genome (%)	79.4	61.3	58.1	2.2	81.6	61.3	56.1
Exonic (%)	32.7	27.2	26.9	0.3	24.5	20.8	19.7
Intronic (%)	43.0	30.1	27.8	0.9	52.4	35.6	31.9
Intergenic (%)	10.3	8.1	7.6	1.3	9.4	8.2	7.6
Total number of reads mapped (%)	95.4	91.8	90.0	90.2	96.7	88.1	78.7
No. of transcripts with RPKM >1	15,001	16,023	16,239	3,252	11,589	12,911	12,601
Mapped to correct strand (%)	97.2	96.8	96.4	96.8	97.6	96.8	96.5
Mapped to lncRNA*:							
Total (%)	6.0	7.0	6.9	0.5	7.0	5.7	5.3
Unique (%)	4.4	5.0	4.9	0.5	5.6	4.5	4.1
No. of lncRNA transcripts with RPKM >1	10,076	9,665	9,192	3,033	5,145	4,513	3,877

* The total number of lncRNA transcripts was 26,414 for the human GENCODE 21 data set, and 9,962 for the mouse GENCODE M4 data set.

Table I. High-quality RNA-seq libraries from high-input and picogram-input workflows. Libraries were generated from Human Brain or Mouse Brain Total RNA using either the SMARTer Stranded Total RNA Sample Prep Kit - HI Mammalian (Stranded HI; for 1 µg input RNA) or the pico method (described above in Figure 1). The pico method was able to generate consistent sequencing metrics even from as little as 100 pg of total RNA input. The number of lncRNA transcripts identified in libraries generated from nanogram and picogram amounts of total RNA was comparable to the number identified from 1 µg of the same RNA sample (Stranded HI), highlighting the robustness of the pico method.

1290 Terra Bella Ave., Mountain View, CA 94043
Orders and Customer/Technical Service: 800.662.2566
Visit us at www.clontech.com

To download a copy of this poster please visit:
www.clontech.com/lncRNA2015