

Improved Methods for rRNA Removal and RNA-Seq Library Preparation

Roy Sooknanan, John Hitchen, Anupama Khanna, Agnes Radek, Jim Pease, and Nick Caruccio
EPICENTRE Biotechnologies, Madison, WI

Introduction

Deep sequencing of mRNA (mRNA-Seq) is rapidly gaining momentum for transcript profiling, discovery of novel transcripts, and identification of alternative splicing events. Current methods for making RNA-Seq libraries generally consist of first preparing: i) ribosomal RNA (rRNA)-depleted mRNA; and ii) platform-specific, di-tagged cDNA libraries for sequencing. However, these methods are time-consuming with significant hands-on time for cDNA fragmentation, end-polishing, adaptor-ligation, size selection, multiple clean-up steps and require the use of high quality, intact total RNA samples. Just as importantly, rRNA contamination poses a significant barrier to performing transcriptome analysis.

We describe the development of a novel rRNA removal procedure (Ribo-Zero™ technology) and a novel ligation-free cDNA library synthesis procedure (ScriptSeq™ technology) for preparing directional next-generation RNA-Seq libraries in about 6 hours from both intact and degraded RNA samples, and with low rRNA background.

Methods Overview

Ribo-Zero™ rRNA Removal

The Ribo-Zero rRNA removal process uses a proprietary method* that is optimized for removal of all sizes of rRNA. Intact or degraded total RNA samples (50 ng to 5 µg) are mixed with the rRNA Removal Reagents in solution (25 minutes). The mixture is then added to Ribo-Zero Microspheres and incubated for 20 minutes followed by the removal of the Microspheres with a spin-filter column (2 minutes). The rRNA-depleted RNA is recovered either by ethanol precipitation or a column-purification method of choice.

ScriptSeq™ Library Preparation

The ScriptSeq library preparation method* employs random-primed, first-strand cDNA synthesis from rRNA-depleted mRNA (10 ng) that incorporates a platform-specific 3'-sequencing tag (30 minutes; Fig. 1). The RNA and excess oligonucleotides are then enzymatically hydrolyzed (13 minutes) and a mixture comprising a terminal-tagging oligonucleotides (TTO) and a DNA synthesis reagent is added. The TTO contains a known 5'-sequence tag (A), a 3'-random sequence and a terminally blocked 3' end to prevent priming of DNA synthesis. The 3' ends of the cDNA molecules are extended, incorporating a complement to the sequence tag (18 minutes) forming cDNA molecules with known sequence tags at their 5' (A) and 3' (B) ends for directionality. Excess TTO is enzymatically degraded (13 minutes) and the di-tagged cDNA molecules are purified (10 minutes). The complete di-tagged cDNA synthesis process is performed in a single reaction tube. Next, platform-specific capture sequences, which can include a barcode, are added to the di-tagged cDNA molecules by limited-cycle PCR, and the products are purified (70 minutes). The adaptor-tagged library is now ready for cluster generation in preparation for deep sequencing.

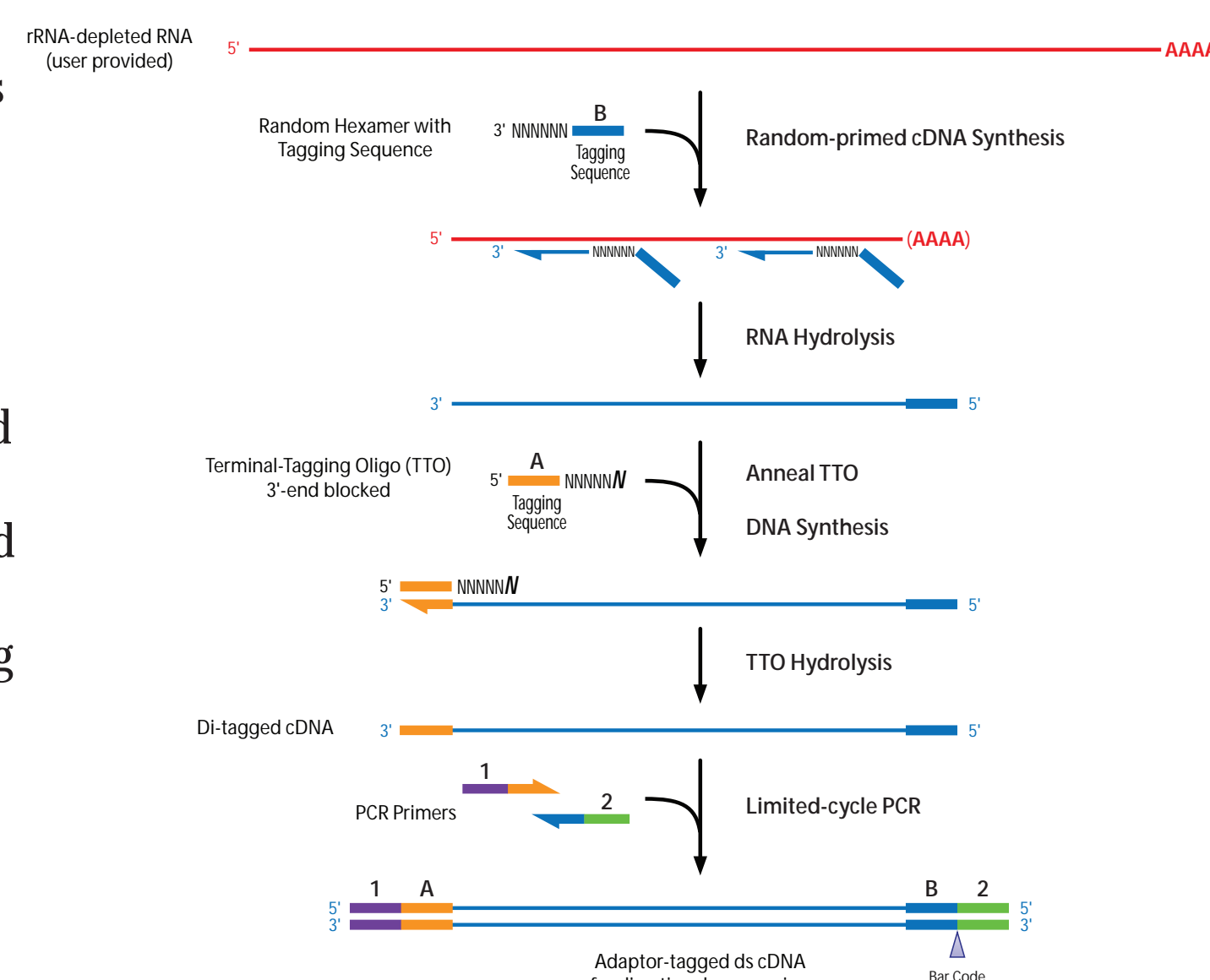


Figure 1. Schematic overview of the ScriptSeq™ directional, di-tagged library preparation method. The process is complete in less than 3 hours, with no intermediate purification steps from RNA to di-tagged cDNA fragments.

RNA-Seq workflow comparison

Table 1 compares the EPICENTRE ScriptSeq library preparation workflow to that provided by current Illumina RNA-Seq methods. The EPICENTRE workflow offers significant savings in the overall reaction and hands-on times, and number of steps required. No intermediate clean-up steps are required from preparing rRNA-depleted RNA to synthesis of di-tagged cDNA fragments.

Table 1. ScriptSeq™ mRNA-Seq Library Preparation Kit (Illumina-compatible) generates libraries for directional sequencing in less than 3 hours. Times for each step are shown in hours:minutes.

Conventional mRNA-Seq Method	ScriptSeq™ Method
Fragment RNA (1:00)	Fragment RNA and synthesize di-tagged cDNA (1:40)*
Synthesize cDNA (4:30)	Clean up cDNA (0:10)
Ligate adaptors (2:00)	—
Size-select from gel (1:30)	—
Enrich library by PCR (1:00)	Enrich library by PCR (1:00)
Total Time: 10:00	Total Time: ~3:00 *single-tube reaction

*Covered by issued and/or pending patents.

Results and Discussion

Efficient removal of rRNA from both intact and fragmented total RNA

A 2.5-µg aliquot of either intact or partially fragmented HeLa total RNA was treated with the Ribo-Zero rRNA Removal Kit (Human/Mouse/Rat). The RNA samples were then converted to first-strand cDNA using random hexamers and the cDNA. The cDNA samples were diluted and used as template in qRT-PCR with primers spanning multiple regions of the 28S rRNA and 18S rRNA transcripts, and the complete 5.8S and 5S rRNA transcripts. Results of several experiments were averaged (Table 2).

Table 2. Analysis of Ribo-Zero™ rRNA Removal Kit (Human) treated RNA preparations.

cDNA Sample Source	% Removal 28S rRNA	% Removal 18S rRNA	% Removal 5.8S rRNA	% Removal 5S rRNA
Intact HeLa total RNA	>99.9%	>99.9%	>99.9%	>96.8%
Partially fragmented HeLa total RNA	>99.9%	>99.9%	>99.9%	>97.8%

Significant reduction of rRNA background and improvement in uniquely mappable reads

Intact and partially fragmented Universal Human Reference RNA (UHRR) (2 x 2.5 µg each) were treated with either the Ribo-Zero Kit or a competitive rRNA removal kit. The respective rRNA-depleted samples were pooled and, for each, RNA-Seq libraries were prepared in triplicate following the ScriptSeq procedure outlined in Fig. 1, using rRNA-depleted RNA from the equivalent of 1 µg total RNA. Replicates of the respective RNA-Seq libraries were pooled and sequencing was performed using Illumina®GAIIx sequencer with 36-nt reads. The data were analyzed using Illumina's Pipeline Eland_rna Module and CASAVA Software as well as the TopHat Software for mapping splice junctions (<http://tophat.ccb.umd.edu/index.html>). The mapping results showed that rRNA background in each library was significantly reduced by the Ribo-Zero Kit (Table 3). Further, for fragmented samples, the Ribo-Zero Kit considerably outperformed the competitive kit, both in terms of reducing rRNA background and obtaining uniquely mappable sequences.

Table 3. Summary of RNA-Seq results from libraries prepared using Ribo-Zero™ rRNA removal technology and a competitive rRNA removal kit.

Total RNA Sample	rRNA Removal Method	% rRNA Background	% Uniquely Mappable Sequences
Intact UHRR	Ribo-Zero™	1.4%	58.1%
Intact UHRR	Competitor	18.4%	51.4%
Fragmented UHRR	Ribo-Zero™	2.1%	59.6%
Fragmented UHRR	Competitor	63.3%	24.6%

High-quality ScriptSeq libraries prepared from rRNA-depleted or poly(A)-enriched RNA samples

We used UHRR, BrRR, and total RNA isolated from FFPE breast cancer tissue as starting material. The specified samples (Figure 2) were treated with either the Ribo-Zero Kit, a competitive rRNA-removal kit (Company A), or a commercial oligo(dT)-based mRNA enrichment kit. For UHRR and BrRR, ScriptSeq libraries were prepared from 50-ng aliquots of the resulting rRNA-depleted or poly(A)-enriched RNA, as outlined in Fig. 1. For FFPE samples, the entire amount of rRNA-depleted RNA recovered from 500 ng total RNA input was used to prepare the libraries. The di-tagged cDNA reactions were amplified by PCR for either 10 cycles (UHRR and BrRR) or 12 cycles (FFPE) followed by Exo I digestion. Each RNA-Seq library was purified using MinElute (Qiagen) and recovered in 15 µl of Elution Buffer. Replicate reactions were pooled and examined using a Bioanalyzer (Agilent). Single-lane, 54-nt unidirectional sequencing reads were obtained for each library using an Illumina GAII sequencer, and sequence analysis was performed using Bowtie.

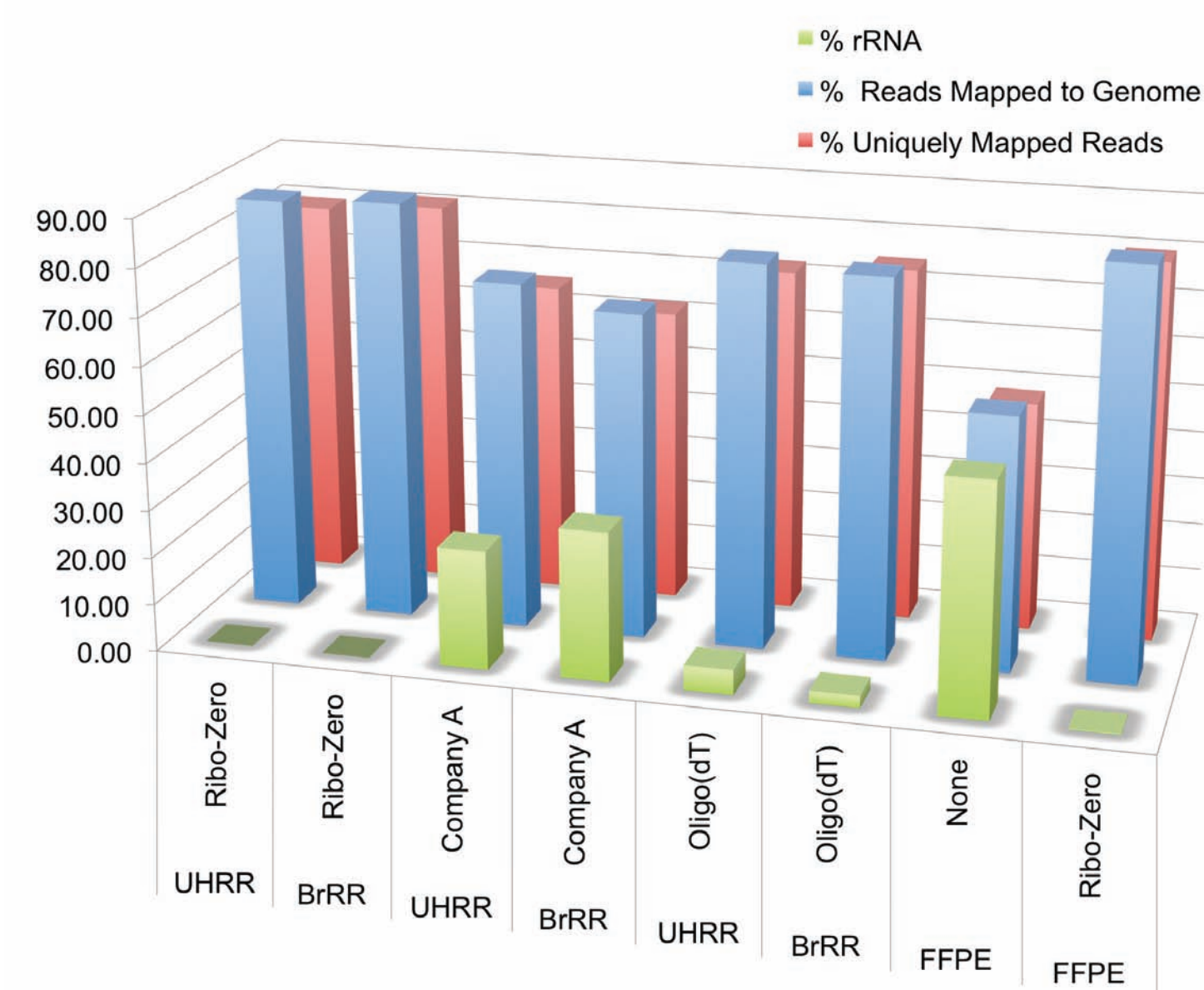


Figure 2. Summary of sequencing data from ScriptSeq™ libraries. Libraries were prepared as described above from Universal Human Reference RNA (UHRR), Brain Reference RNA (BrRR), or RNA extracted from FFPE breast cancer tissue (FFPE). The indicated method of rRNA removal or mRNA enrichment was used.

Good correlation of differential gene expression between ScriptSeq libraries and MAQC microarray data

ScriptSeq libraries were prepared from intact and partially fragmented UHRR and BrRR total RNA and sequenced, as described in the methods for Table 3. RNA-Seq gene expression data were compared to the corresponding gene expression data obtained from the MAQC (Fig. 3). Approximately 91% correlation of differential gene expression ratios was observed for both intact (n = 617 genes) and fragmented (n = 633 genes) UHRR and BrRR.

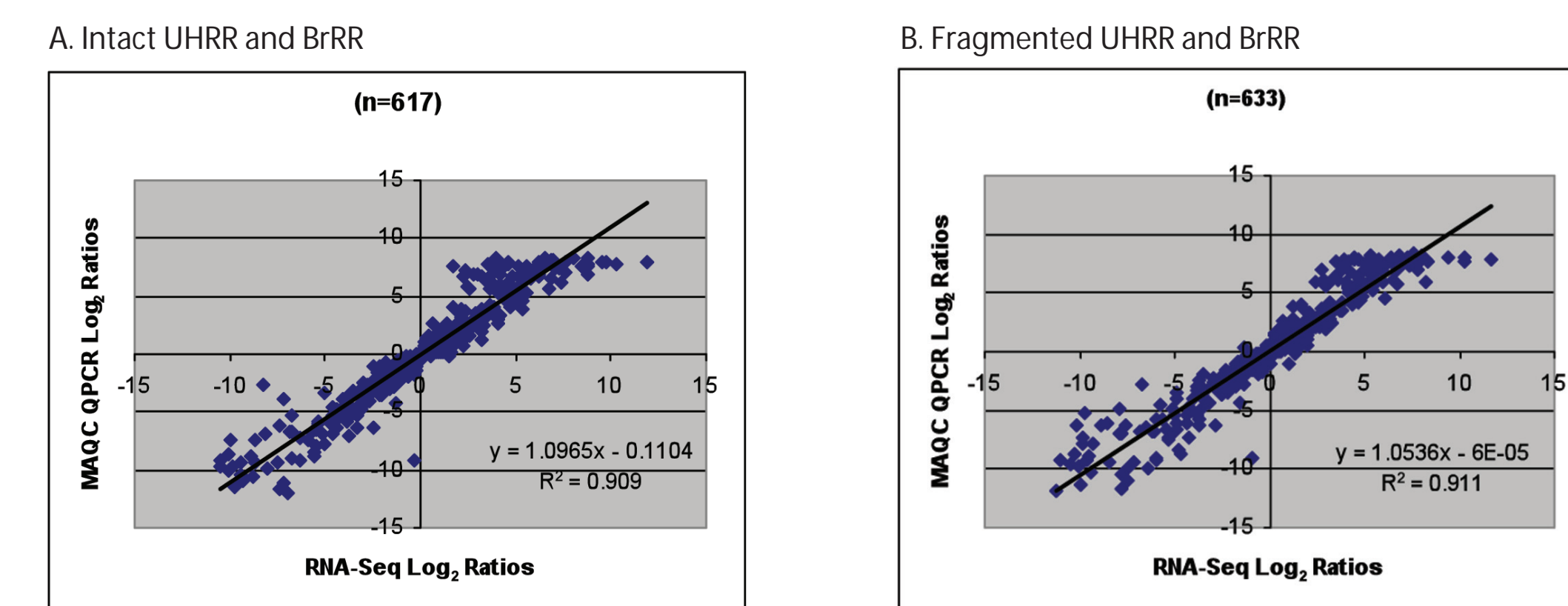


Figure 3. Correlation of gene expression between data obtained from EPICENTRE ScriptSeq™ libraries and corresponding MAQC data. A) Libraries prepared from rRNA-depleted, intact UHRR and BrRR. B) Libraries prepared from partially fragmented, rRNA-depleted UHRR and BrRR.

Directionality and 5'/3' coverage of ScriptSeq libraries

ScriptSeq libraries were prepared from intact and partially fragmented UHRR total RNA and sequenced, as described in the methods for Table 3. Wiggle tracks for the reads mapped using Eland were created with the two strands separated into different tracks. Reads were then mapped to the human genome assembly (GRCh37/hg19) in the UCSC browser and the forward and reverse reads displayed as different colors (Fig. 4). These tracks also describe the mapping data on which the gene expression analysis was performed.

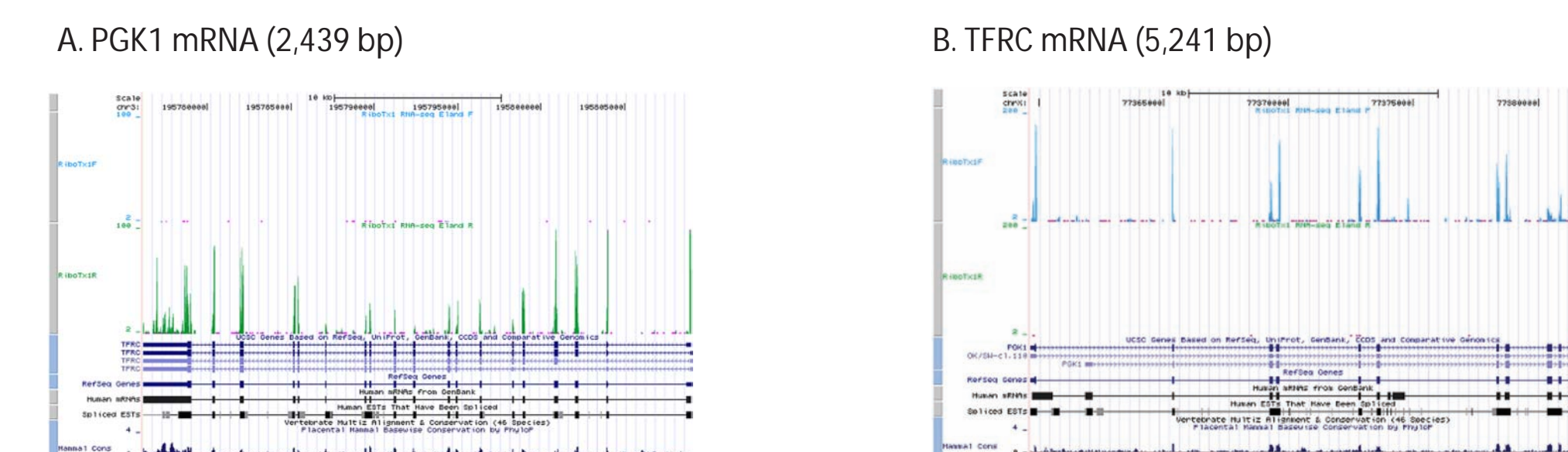


Figure 4. Directionality and 5'/3' coverage of transcripts.

Conclusions

Ribo-Zero rRNA Removal

- Highly efficient removal of rRNA from both intact and fragmented RNA samples (50 ng to 5 µg total RNA).
- Single-pass rRNA removal process.
- Enables sequencing of degraded RNA samples by significantly lowering the rRNA background.
- Enables recovery of both poly(A)⁺ and non-rRNA poly(A)⁻ transcripts.
- Kits for human/mouse/rat (mammalian) and Gram-negative bacteria currently available.
- Kits for Gram-positive bacteria and plants in development.

ScriptSeq Library Preparation

- Simple, ligation-free, and directional RNA-Seq library preparation workflow with no need for gel purification; compatible with Illumina GAII and Roche FLX-Titanium chemistry.
- High-quality libraries from rRNA-depleted total RNA, poly(A) RNA, or FFPE RNA.
- Cluster generation-ready amplicons in under 4 hours from rRNA-depleted RNA (10 ng).
- Equal 5'/3' representation of transcripts.
- Detects both poly(A)⁺ and poly(A)⁻ transcripts with use of random-primed cDNA synthesis.
- High correlation (~91%) with MAQC microarray data set.
- Barcoding option available for Illumina GAIIx libraries.

Contact Information:

Jim Pease: jim.pease@epibio.com
(800) 284-8474 Ext. 6110



www.EpiBio.com • 1-800-284-8474