

Improved Technologies for Ribosomal RNA Removal and Directional RNA-Seq Library Preparation

Roy Sooknanan, Agnes Radek, John Hitchen, and Anupama Khanna

EPICENTRE Biotechnologies, Madison, WI

Introduction

Deep, massively parallel sequencing of cDNA generated from RNA (RNA-Seq) is rapidly gaining momentum for transcript profiling, discovery of novel transcripts, and identification of alternative splicing events. Current methods for making sequencer-specific di-tagged DNA fragment libraries for RNA-Seq typically comprise depleting ribosomal RNA (rRNA), and either: (i) RNA fragmentation, 5' and 3' adaptor-ligation, size selection, cDNA synthesis, and multiple clean-up steps; or (ii) cDNA synthesis followed by cDNA fragmentation, end-polishing, 5' and 3' adaptor-ligation, size selection, and multiple clean-up steps. These methods are generally time-consuming and require significant hands-on time. Further, even after removal of rRNA from the total RNA samples using commercially available kits, 50% or more of the sequence reads can still be from rRNA, decreasing sequencing depth and coverage.

Here, we present RNA-Seq results obtained using a novel rRNA removal method (Ribo-Zero™ technology) and a novel, ligation-free process for preparing directional di-tagged DNA fragment libraries (ScriptSeq™ technology) for RNA-Seq. Using these methods, directional di-tagged DNA fragment libraries can be prepared in about 6 hours from either intact or fragmented (e.g., FFPE) total RNA samples. Less than 2% of the sequence reads from libraries generated from total RNA from either intact or FFPE samples map to rRNA sequences (28S, 18S, 5.8S, and 5S). This reduction in rRNA sequence reads improves sequence depth and coverage, and increases the percentage of uniquely mapped reads. Further, there is a high correlation ($R^2 = 0.9235$) between differentially expressed transcripts found in the ScriptSeq™ RNA-Seq libraries and the MAQC qPCR panel of genes.

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

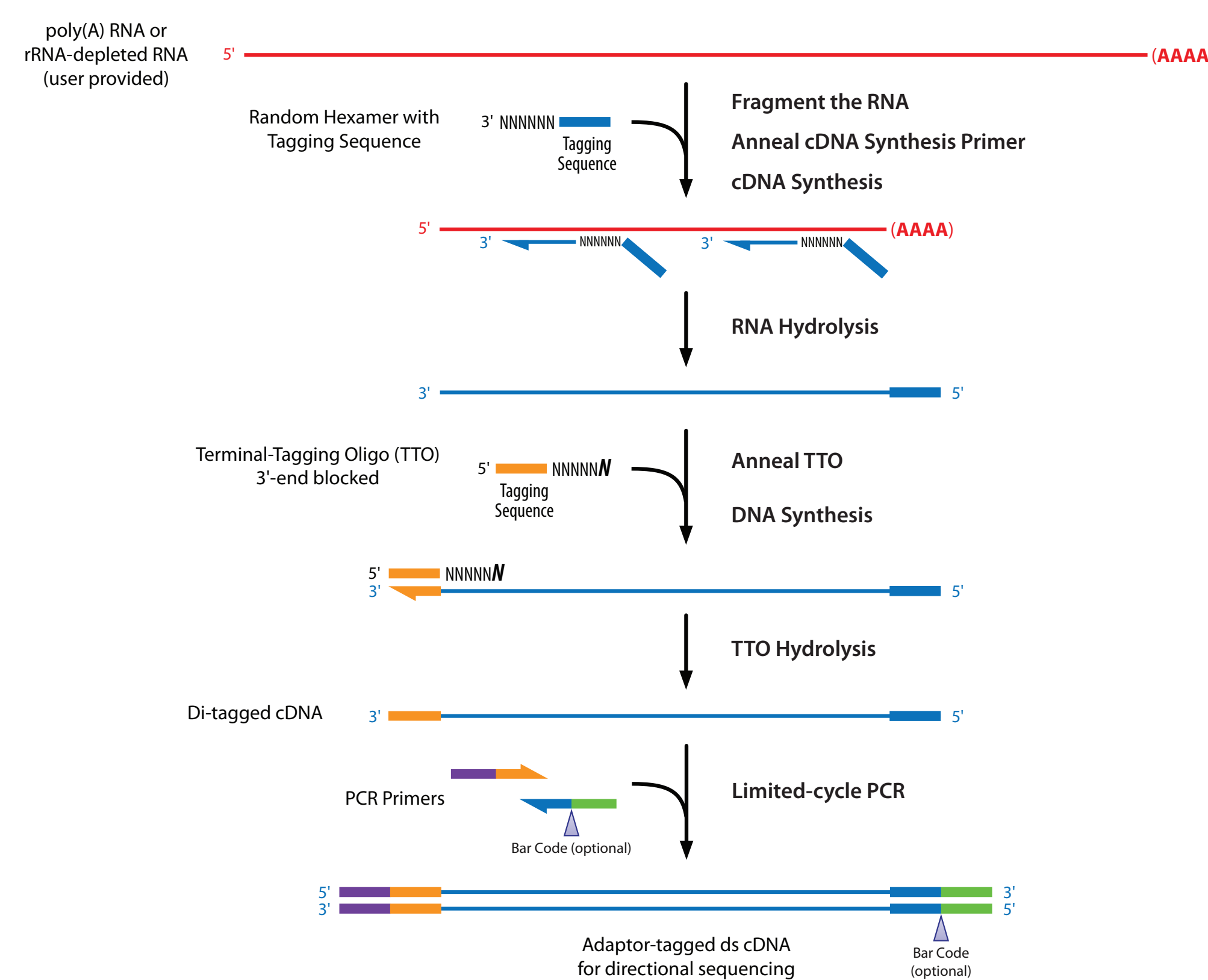


Figure 1. Schematic overview of the ScriptSeq™ directional, di-tagged library preparation method. The process is complete in less than 3 hours, with no intermediate purification steps from RNA to di-tagged cDNA fragments.

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 oligonucleotide (TTO) and a DNA synthesis reagent is added. The TTO contains a known 5'-sequence tag, 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), and forming cDNA molecules with known sequence tags at their 5' and 3' 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 or emulsion PCR in preparation for deep sequencing.

RNA-Seq workflow comparison

Table 1 compares the EPICENTRE ScriptSeq library preparation workflow to that provided by conventional 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. The 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

Results and Discussion

A. Ribo-Zero rRNA Removal

Efficient removal of rRNA from plant leaf, root, and seed total RNA

Aliquots (1 µg and 5 µg) of leaf, root, and seed total RNA were treated with either the Ribo-Zero rRNA Removal Kits for Plant (Leaves) or Plant (Roots/Seeds), accordingly. *Arabidopsis thaliana* total RNA (RIN# 5.1), which appeared partially degraded, was purchased from Amsbio (UK). Soybean root and seed total RNA samples were kind gifts from Dr. Lila O. Vodkin, University of Illinois at Urbana-Champaign. Corresponding mock-treated total RNA samples were also prepared as controls. An aliquot from each sample was removed for Bioanalyzer analysis and the remainder was converted to first-strand cDNA using random hexamers. The cDNA samples were diluted and used as template in qRT-PCR with primers spanning multiple regions of the long rRNA transcripts (25S, 18S, 23S and 16S). For quantifying short (5.8S, 5S, and 4.5S) rRNAs, single primer pairs were used. As controls for mRNA preservation following Ribo-Zero treatment, high copy-number genes (EF-1 α , AT5G60390, and UBQ10, AT4G05320), and a medium/low copy-number gene (F-box protein, AT5G15710) were assayed by qRT-PCR using the same cDNA samples. The results for the rRNA reduction and mRNA preservation are shown in Figs. 2 and 3, and Table 2. It is evident from the electropherogram traces that the major rRNA peaks are depleted (Fig. 2) following Ribo-Zero treatment. For the partially degraded *Arabidopsis thaliana* total RNA sample, the qRT-PCR curves for 18S (Fig. 2A) and 23S (Fig. 2B,C) rRNA reduction are shown as examples. Fig. 3C shows that, for the two mRNA sequences (EF-1 α and F-box protein), the respective amounts remained essentially unchanged before and after Ribo-Zero treatment. The relative levels of depletion (>99.9%) based on qRT-PCR for the different nuclear and chloroplast rRNA for *Arabidopsis thaliana* leaf, and soybean root and seed total RNA samples are shown in Table 2.

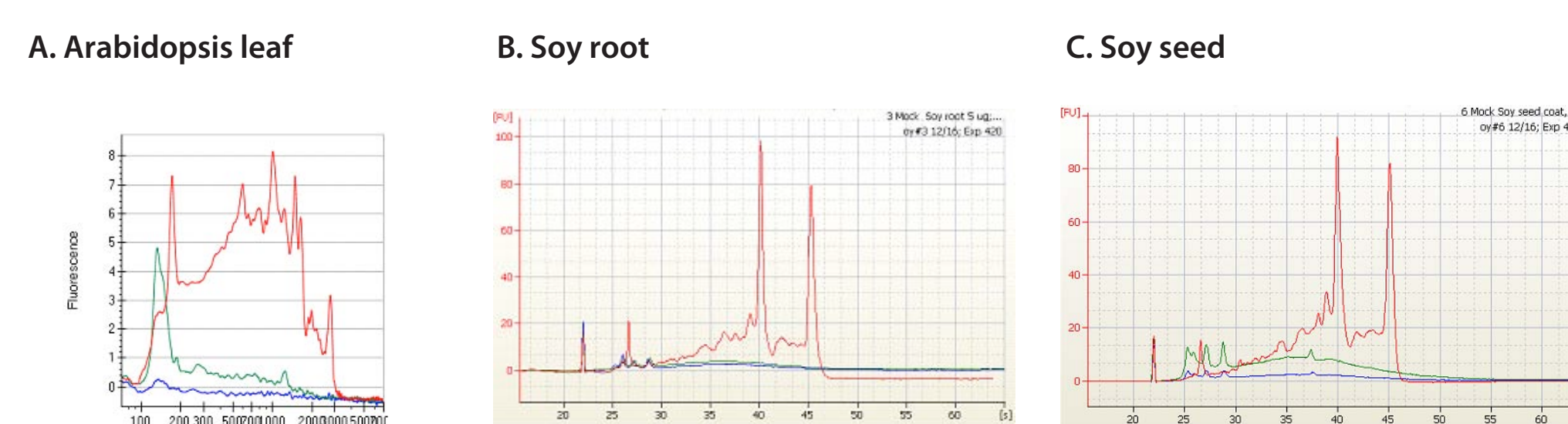


Figure 2. Electropherograms of Ribo-Zero™ and mock-treated total RNA. Aliquots containing 1 µg (blue) and 5 µg (green) aliquots of each total RNA—*Arabidopsis* leaf (A), soy root (B), and soy seed (C)—were treated with the appropriate Ribo-Zero rRNA Removal Kit and the resulting rRNA-depleted RNA samples were analyzed. The corresponding mock-treated samples (red) were similarly analyzed. For the *Arabidopsis* leaf, the Caliper GX instrument was used; whereas, for the soy samples, the Agilent Bioanalyzer was used.

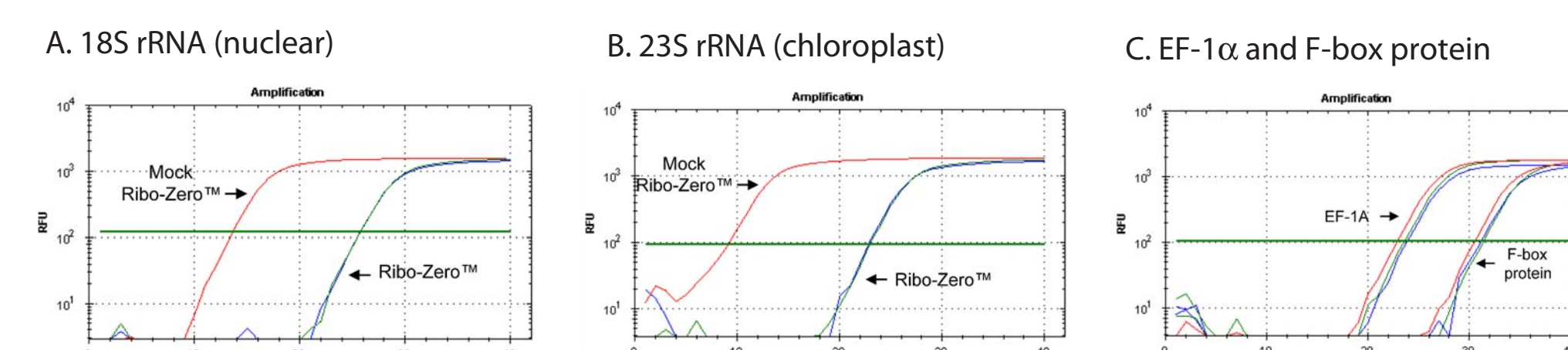


Figure 3. Examples of qRT-PCR curves showing the relative changes in the levels of rRNA (18S and 23S) and two mRNA sequences (EF-1 α and F-box protein) before (mock) and after Ribo-Zero™ treatment. The mock-treated control is shown in red and the treated samples are shown in blue (1 µg) and green (5 µg), respectively.

Table 2. qRT-PCR analysis of plant rRNA removal after Ribo-Zero™ treatment. The percent rRNA removal after Ribo-Zero treatment of *Arabidopsis* leaf, and soybean root and seed total RNA samples was determined using random-primed qRT-PCR with primers to multiple regions of each rRNA where appropriate. Primers were designed with the PrimerSelect module of the LaserGene package (DNASTAR). Primer efficiencies were determined from the slope of C_t vs. dilution plots, and for quantification of rRNA removal, primers with efficiencies between 95% and 100% were selected. ΔC_t values (ΔC_t , difference between mock-treated control and Ribo-Zero-treated sample) for rRNAs were normalized to ΔC_t for EF-1 α . The percent rRNA removal is an average of all primer pairs for each rRNA where applicable.

Input RNA Sample	Input Amounts	% rRNA Removal (Nuclear)			% rRNA Removal (Chloroplast)			
		25S	18S	5.8S	23S	16S	5S	4.5S
<i>Arabidopsis thaliana</i> leaf	1-5 µg	>99.99%	>99.96%	>99.94%	>99.99%	>99.98%	>99.97%	>99.94%
Soybean root	1-5 µg	>99.99%	>99.95%	>99.99%	>99.98%	>99.97%	>99.75%	>99.95%
Soybean seed	1-5 µg	>99.99%	>99.99%	>99.97%	>99.98%	>99.98%	>99.85%	>99.95%

B. ScriptSeq Library Preparation

Universal Human Reference RNA (UHRR), Brain Reference RNA (BrRR), and total RNA isolated from FFPE breast cancer tissue were used as starting material. The specified samples 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 alignment was performed by Beijing Genome Institute (BGI) with both the well-annotated human genome (ENSEMBL56) and an in-house built junction sequences using BWA software allowing for four mismatches.

Reduction in rRNA background improves uniquely mappable reads

Sequencing data for various samples are summarized in Fig. 4. The results show that the method of rRNA reduction greatly influences sequencing results. The Ribo-Zero Kit resulted in maximal rRNA removal compared to other methods, with corresponding improvements in the percentage of uniquely mapped reads from the ScriptSeq libraries.

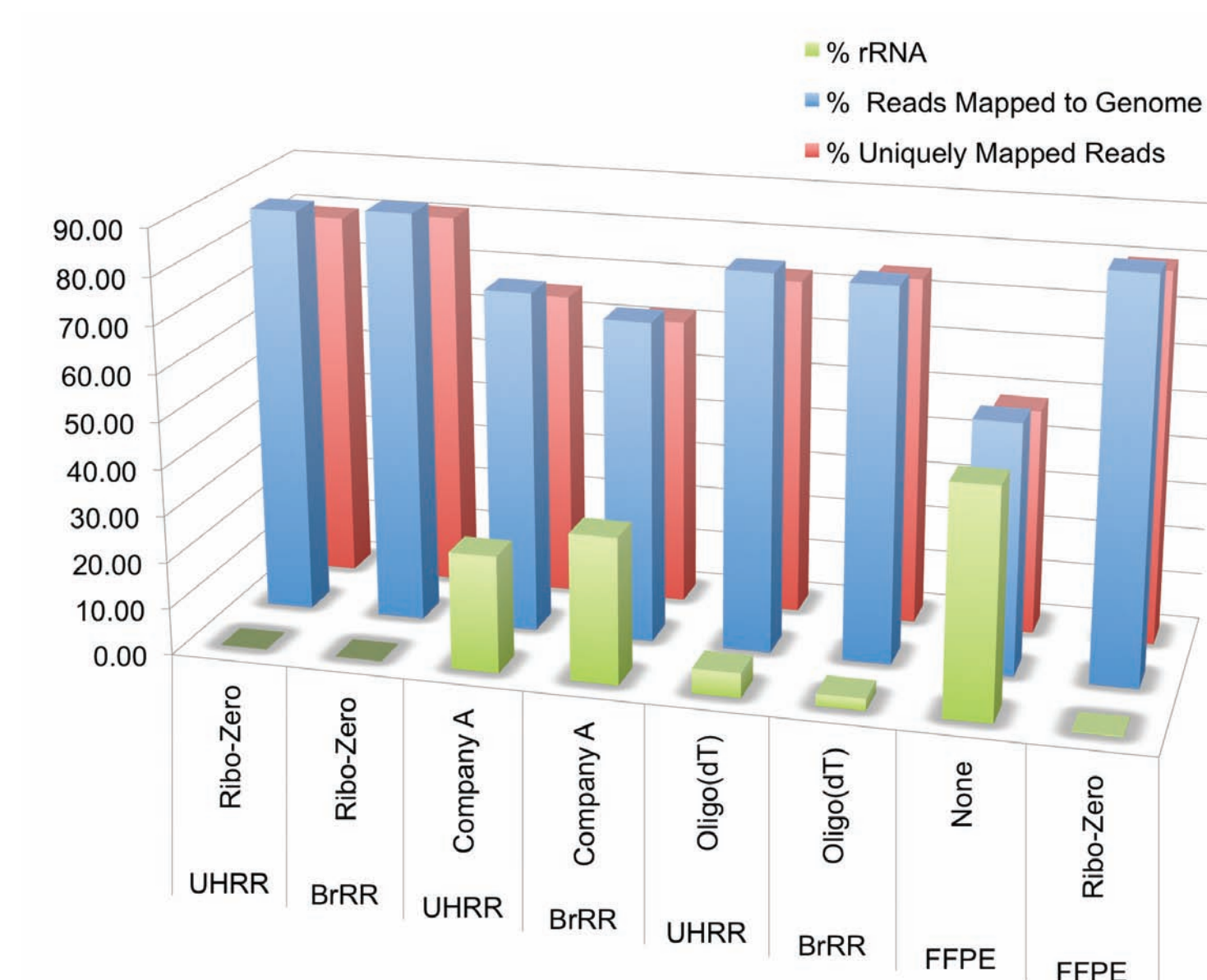


Figure 4. Summary of sequencing data from ScriptSeq™ libraries. Libraries were prepared as described above from Universal Human Reference RNA (UHRR), Brain Reference RNA (BrRR), or RNA extracted from FFPE breast cancer tissue (FFPE). The indicated method of rRNA removal or mRNA enrichment was used.

Excellent correlation of differential gene expression and directionality (strandedness) of ScriptSeq libraries

ScriptSeq libraries were prepared from either Ribo-Zero rRNA-depleted RNA or poly(A)-enriched mRNA from UHRR and BrRR total RNA and sequenced, as described above. RNA-Seq gene expression data were compared to the corresponding gene expression data obtained from the MAQC (Fig. 5A). Approximately 91% correlation of differential gene expression (DGE) ratios was observed for Ribo-Zero treated RNA (n = 706 genes) for UHRR and BrRR (Fig. 5A). A similar correlation to MAQC data was seen for poly(A)-enriched mRNA.

The directionality (strandedness) of a typical ScriptSeq library is shown in Fig. 5B. The percent of correct strand reads following mapping as described above is shown in comparison to a typical nonstranded protocol.

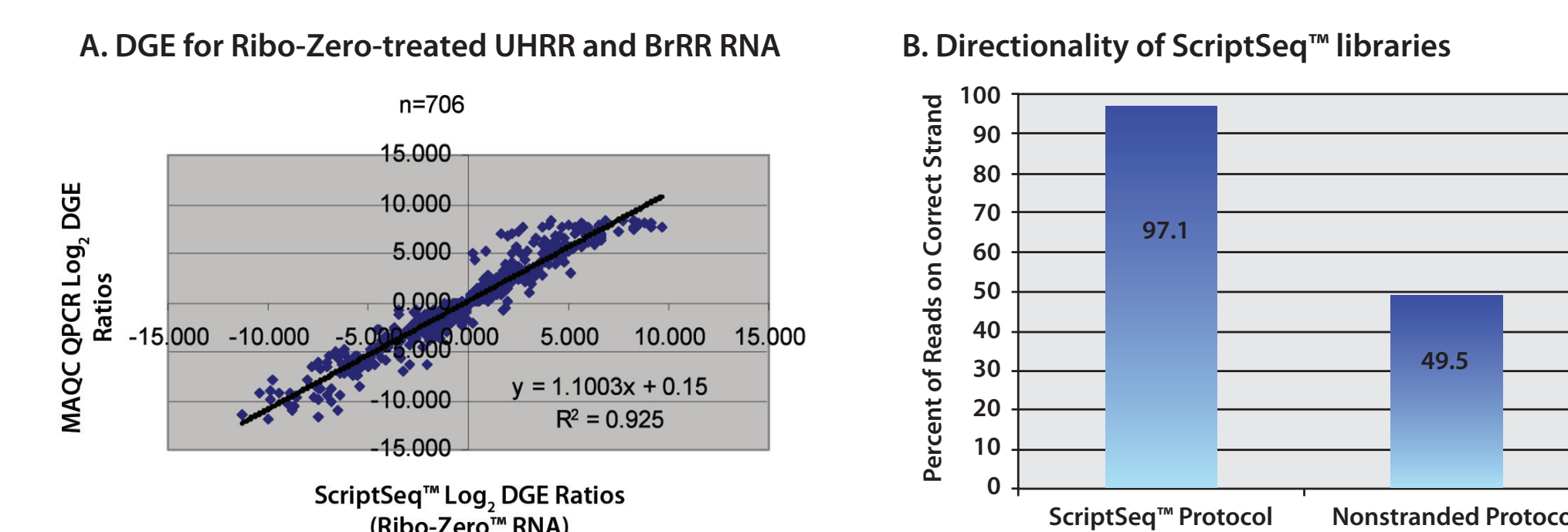


Figure 5. Correlation of gene expression between data obtained from EPICENTRE ScriptSeq™ libraries and corresponding MAQC data, and the directionality (strandedness) of ScriptSeq libraries. A) ScriptSeq libraries prepared from rRNA-depleted, UHRR and BrRR. B) Directionality of ScriptSeq libraries prepared from Ribo-Zero™-treated UHRR.

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) (high and low inputs), and Gram-negative and Gram-positive bacteria currently available.
- Beta kits for plant leaves and roots/seeds are currently available.

ScriptSeq Library Preparation

- Simple, ligation-free, and directional RNA-Seq library preparation workflow with no need for gel purification; compatible with Illumina GAI and Roche FLX-Titanium chemistry.
- High-quality libraries from rRNA-depleted total RNA, poly(A) RNA, or FFPE RNA.
- Cluster generation-ready amplicons in about 3 hours from rRNA-depleted RNA.
- Excellent strand preservation and transcript coverage.
- 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 GAIx libraries.