Improved Technology for Ribosomal RNA (rRNA) Removal from Formalin-fixed Paraffin-Embedded Samples

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

EPICENTRE Biotechnologies (an Illumina® company), Madison, WI, USA

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 first preparing rRNA-depleted RNA from total RNA samples that are usually of good quality, followed by the synthesis of the di-tagged cDNA sequencing templates. In the past few years, however, it has also become evident from microarray and qPCR studies that formalin-fixed paraffin-embedded (FFPE) cancer tissues hold valuable secrets about disease states. However, RNA-Seq libraries prepared from such highly fragmented FFPE RNA samples yield limited information, since these libraries contain a majority of rRNA reads, with consequent decreases in sequencing depth and coverage. Current commercially available rRNA removal kits are not designed to remove fragments of rRNA, which poses a significant limitation for preparing highly informative RNA-Seq libraries from FFPE RNA samples.

Here, we present RNA-Seq results obtained using Ribo-Zero™ "single-pass" rRNA removal technology, and ScriptSeq™ ligation-free technology for preparing directional, di-tagged cDNA libraries for next-generation sequencing. Using these methods, directional di-tagged DNA fragment libraries can be prepared in under 6 hours from either intact or fragmented (e.g., FFPE) total RNA samples (as little as 100 ng FFPE total RNA sample required). Less than 1% 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 from FFPE RNA samples improves sequence depth and coverage, and increases the percentage of uniquely mapped reads, thereby increasing the information obtained from these disease samples.

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 (Fig. 1). Intact or degraded total RNA samples (100 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 (2 minutes). The rRNA-depleted RNA is recovered either by ethanol precipitation or a column-purification method of choice.

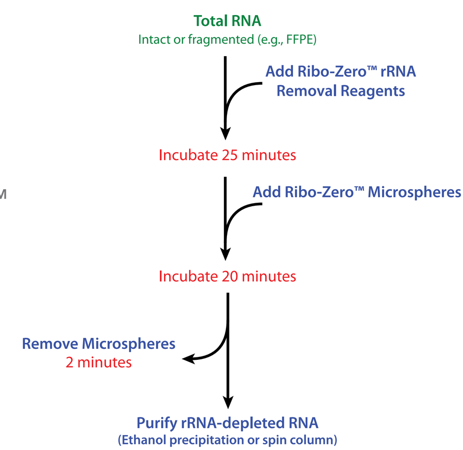

Figure 1. Schematic overview of Ribo-Zero™ rRNA removal technology.

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. 2). 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 are added to the di-tagged cDNA molecules by limited-cycle PCR, and the products are purified (70 minutes). The di-tagged PCR amplicons are now ready for cluster generation in preparation for deep sequencing.

Covered by issued and/or pending patents.

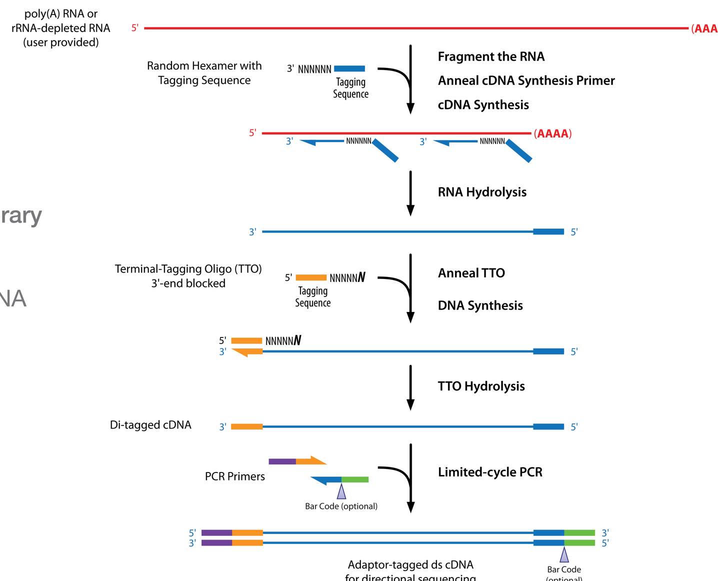

Figure 2. Schematic overview of the ScriptSeq™ directional, di-tagged library preparation method. The process is complete in less than 3 hours, with no intermediate purification steps from RNA to di-tagged cDNA fragments.

RNA-Seq workflow comparison

Table 1 compares the ScriptSeq library preparation workflow to that provided by conventional RNA-Seq methods. The ScriptSeq 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)		
Ligate adaptors	(2:00)	Clean up cDNA	(0:10)
Size-select from gel	(1:30)	—	
Enrich library by PCR	(1:00)	Enrich library by PCR	(1:00)
Total Time (hr:min):	10:00	Total Time: *single-tube reaction*	~3:00

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) as outlined in Fig. 1. Additional aliquots of each RNA was processed similarly, but without adding the rRNA Removal Reagents (mock treatment). One microgram of each RNA sample was loaded on a 1.5% agarose gel and Northern blot analysis performed for GAPDH mRNA (Fig. 3). The remainder of the RNA samples were then converted to first-strand cDNA using random hexamers. The cDNA samples were used as template in qRT-PCR with primers spanning multiple regions of the 28S rRNA and 18S rRNA transcripts, and GAPDH mRNA. There was no measurable change in the GAPDH mRNA content both by Northern blot and qRT-PCR analyses for the complete 5.8S and 5S rRNA transcripts. Greater than 99.9% removal of 28S, 18S, and 5.8S rRNA sequences was measured for both intact and partially fragmented HeLa total RNA (Table 2). For 5S rRNA, the removal was ~97% for either RNA sample.

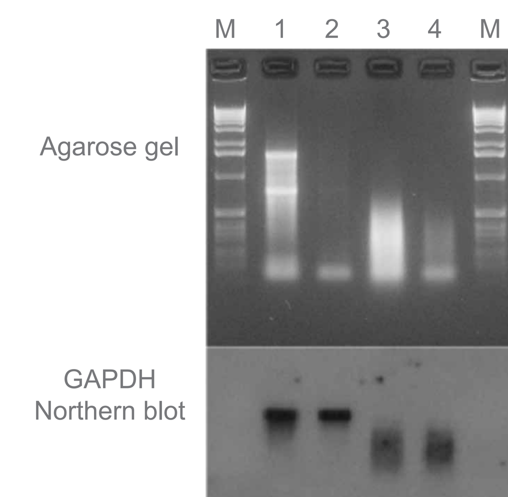

Figure 3. Northern blot analysis for GAPDH mRNA from mock-treated and Ribo-Zero™-treated intact and fragmented HeLa total RNA. Each sample contains 1 µg of RNA. Lane 1, intact HeLa total RNA (mock-treated); lane 2, intact HeLa total RNA (Ribo-Zero™-treated); lane 3, fragmented HeLa total RNA (mock-treated); lane 4, fragmented HeLa total RNA; lanes M, DNA molecular weight marker.

Table 2. Performance of the Ribo-Zero™ rRNA Removal Kit (Human/Mouse/Rat).

RNA 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%

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 (Intact or FFPE RNA) or a commercial oligo(dT)-based mRNA enrichment kit (Intact RNA only). For UHRR and BrRR, ScriptSeq libraries were prepared from 50-ng aliquots of the resulting rRNAdepleted or poly(A)-enriched RNA, as outlined in Fig. 2. 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. 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.

Results

Reduction in rRNA background improves uniquely mappable reads

Sequencing data for the various samples are summarized in Table 3 and Fig. 4. The results show that the method of rRNA reduction greatly influences sequencing results. The Ribo-Zero Kit resulted in maximal rRNA removal with corresponding improvements in the percentage of uniquely mapped reads from the ScriptSeq libraries. A "single-pass" Ribo-Zero rRNA purification was performed compared to two rounds of purification with the poly(A) enrichment kit.

Table 3. The ScriptSeq™ mRNA-Seq kits generate high-quality RNA-Seq libraries from rRNA-depleted intact and FFPE RNA and poly(A)-enriched RNA. The specified RNAs were treated with the Ribo-Zero™ rRNA Removal Kit (Human/Mouse/Rat) or a commercial poly(A) RNA enrichment kit. ScriptSeq (Illumina®-compatible) libraries were prepared and sequenced on an Illumina® GAIIx sequencer.

RNA Sample	rRNA Removal Method	Number of Reads	% rRNA	% Reads Mapped to Genome	% Uniquely Mapped Reads
UHRR	Ribo-Zero™	27,524,073	0.04%	87.0%	79.7%
BrRR	Ribo-Zero™	31,998,097	0.04%	88.0%	81.3%
UHRR	Poly(A) enrichment	30,123,553	5.29%	80.7%	72.5%
BrRR	Poly(A) enrichment	17,763,706	2.57%	79.7%	74.7%
FFPE (breast tumor)	None	21,322,112	48.82%	53.62%	48.5%
FFPE (breast tumor)	Ribo-Zero™ Kit	26,960,546	0.24%	85.6%	79.8%

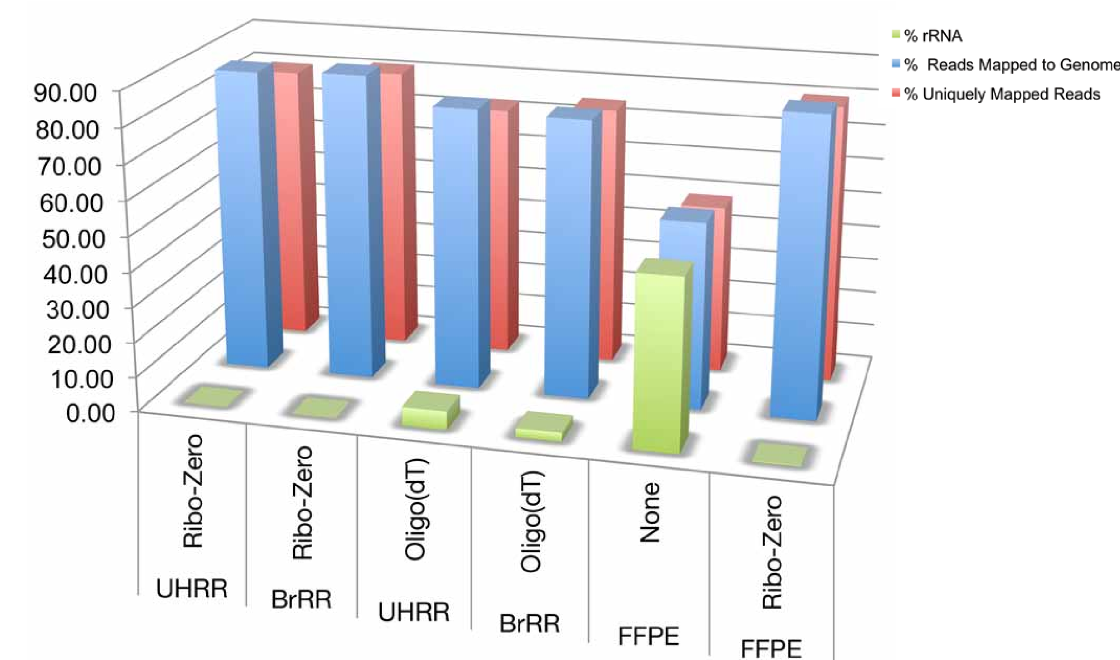

Figure 4. Summary of sequencing data from ScriptSeq™ libraries. Libraries were prepared as described above from UHRR and BrRR, and FFPE breast tumor RNA samples. The indicated method of rRNA removal or mRNA enrichment was used.

ScriptSeq library characteristics

RNA-Seq gene expression data were compared to the corresponding gene expression data obtained from MAQC (Fig. 5) and using the standard Illumina® protocol (Fig. 6). Approximately 92% correlation of differential gene expression (DGE) ratios was observed for Ribo-Zero-treated RNA (n = 706 genes) for UHRR and BrRR, and between the ScriptSeq and Illumina® protocols.

In addition, the ScriptSeq libraries also exhibited excellent directionality, with >97% of reads corresponding to the correct strand (Fig. 5). High expression correlation (R = 0.932) was also observed for libraries prepared from both Ribo-Zero-treated and oligo(dT)-enriched RNA (Fig. 6).

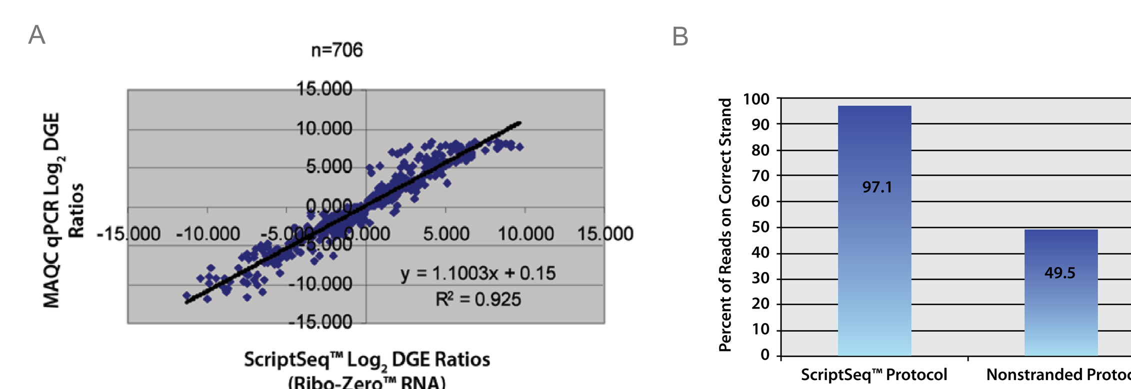

Figure 5. Correlation of gene expression between data obtained from ScriptSeq™ libraries and corresponding MAQC data, and the directionality (strandedness) of ScriptSeq libraries. A) DGE ratios of ScriptSeq libraries prepared from rRNA-depleted UHRR and BrRR. B) Directionality of ScriptSeq libraries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

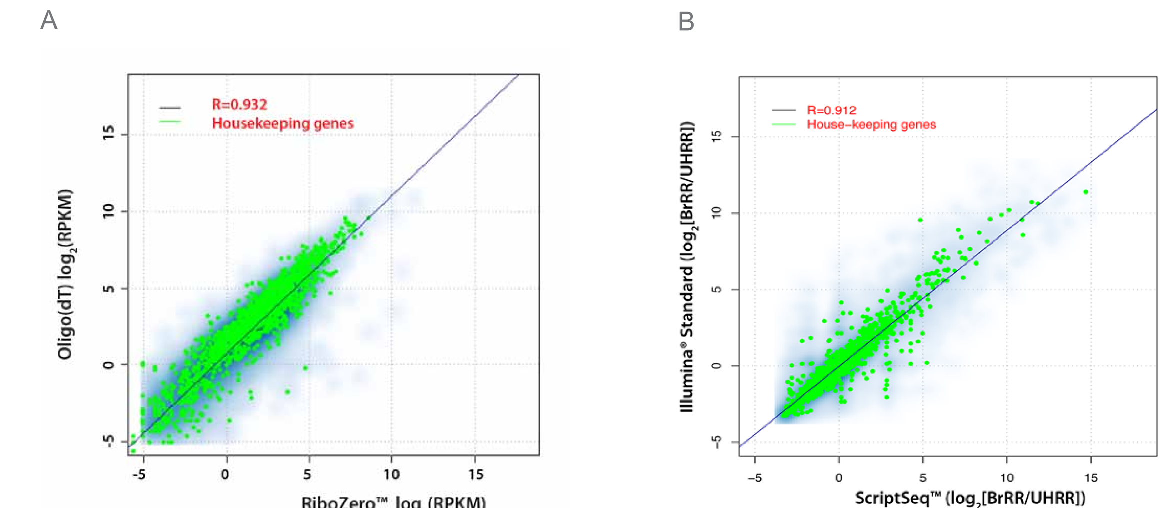

Figure 6. High correlation of relative gene expression levels between ScriptSeq™ and standard Illumina® RNA-Seq libraries. Libraries were prepared from: A) Ribo-Zero and poly(A)-enriched RNA samples; and B) using the standard Illumina® protocol from a similar set of poly(A)-enriched RNA samples. The correlation was examined for housekeeping genes. ScriptSeq libraries were prepared from UHRR and BrRR RNA samples in-house, and the data for standard Illumina® libraries was downloaded from the NCBI Short Read Archive, a total of ~173.7 million 35-bp tags.

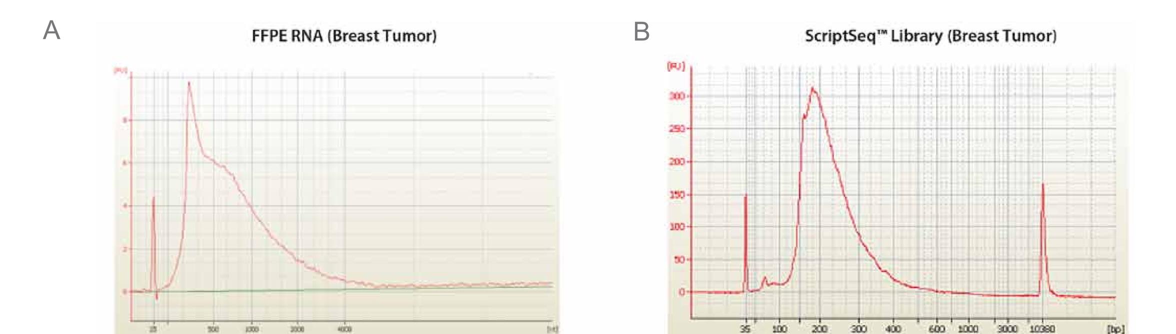

Figure 7. Size range of ScriptSeq™ RNA-Seq libraries from FFPE RNA. Agilent 2100 Bioanalyzer traces are shown for the FFPE breast tumor RNA sample (A) and the ScriptSeq library prepared after Ribo-Zero treatment (B).

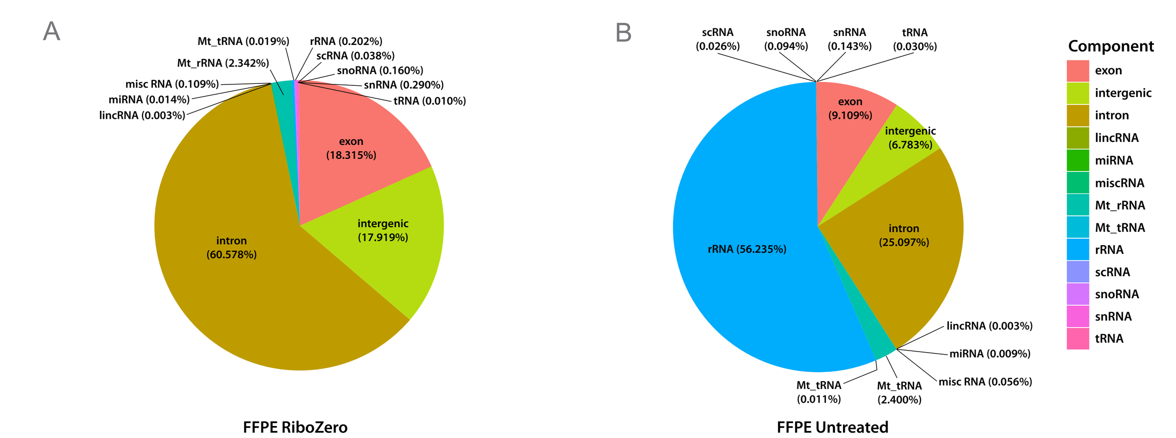

Figure 8. Ribo-Zero™ treatment increases exon, intergenic, and intron mapped reads. FFPE breast tissue RNA was treated with the Ribo-Zero kit; RNA-Seq libraries (A) were compared to the corresponding untreated sample (B). The Ribo-Zero-treated sample contained 76% of all 49,222 possible expressed genes compared to 62% for the untreated samples.

Conclusions

Ribo-Zero rRNA Removal

- Highly efficient removal of rRNA from both intact and fragmented (e.g., FFPE) RNA samples (50 ng to 5 µg total RNA).
- Single-pass rRNA removal process.
- Enables sequencing of degraded (e.g., FFPE) RNA samples by significantly lowering the rRNA background.
- Enables recovery of both poly(A)+ and non-rRNA poly(A)– transcripts.
- Excellent correlation (R = 0.932) between poly(A)+-selected and Ribo-Zero-treated RNA samples
- Kits for human/mouse/rat (mammalian) (high and low inputs), Gram-negative and Gram-positive bacteria, and plant leaves 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® GAII and Roche FLX-Titanium chemistry.
- High-quality libraries from rRNA-depleted total RNA, poly(A) RNA, or fragmented 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 (R2 = 0.925) with MAQC microarray data set and the standard Illumina® RNA-Seq library preparation protocol.
- Barcoding option available for Illumina® GAIIx libraries.

EPICENTRE® Biotechnologies

726 Post Rd, Madison, WI 53713
Tel: (800) 284-8474 • Fax (608) 258-3088

illumina®