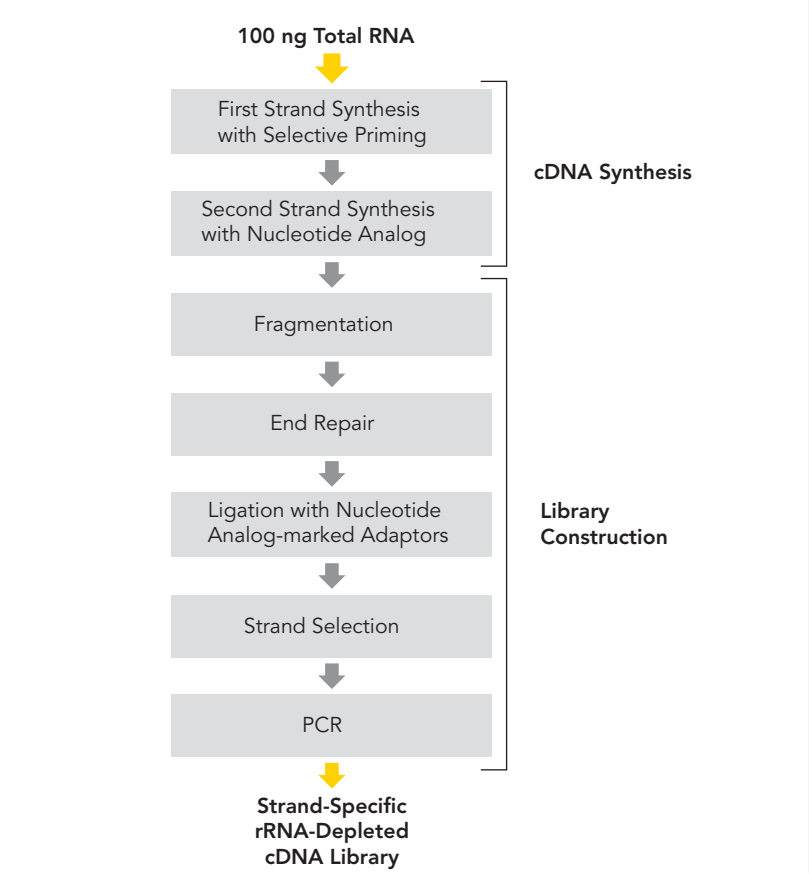
Encore® Complete RNA-Seq Library Systems

Complete solution for strand-specific RNA-Seq library construction using as little as 100 ng of total RNA

Highlights of the Encore Complete RNA-Seq Library Systems

- **A complete solution for strand-specific RNA-Seq —** All required components for low input preparation of strand-specific RNA-Seq libraries from total RNA without the need for rRNA reduction or poly(A) selection
- **Simple, fast, automatable workflow —** Library construction starting with total RNA in as little as seven hours, with only three purification steps and no gel purification
- **Affordable and scalable —** Optional barcoding capability for multiplex sequencing to improve sample throughput and reduce sequencing costs

FIGURE 1 **Construction of Strand-specific RNA-Seq Libraries**



Introduction

The Encore Complete RNA-Seq Library Systems provide an end-to-end solution for strand-specific RNA-Seq library construction using as little as 100 ng of total RNA. The core technology used in this product enriches for non-rRNA in NGS libraries during cDNA synthesis, and can be applied to transcriptomes extracted from a broad range of higher eukaryotes (such as human, mouse, rat and chicken). The cDNA synthesis is carried out using proprietary primers to create cDNA which retains strand information. No dedicated steps are required to reduce sequencing reads derived from rRNA and the final libraries contain sequences from both polyadenylated and non-polyadenylated transcripts. The resulting cDNA is converted to NGS libraries using reagents and adaptors provided in the same kit (see **Figure 1** for workflow). The Encore Complete RNA-Seq Multiplex Systems provide optional barcoding to further optimize efficiencies and cost savings in transcriptome sequencing.

The Encore Complete RNA-Seq Library System (Part No. 0311) contains reagents for production of non-barcoded libraries starting with total RNA, while the Encore Complete RNA-Seq IL Multiplex System 1–8 (Part No. 0312) and IL Multiplex System 9–16 (Part No. 0313) each provide eight unique barcoded adaptors for multiplex sequencing. In combination these latter two kits enable up to 16-plex sequencing. The Encore Complete RNA-Seq DR Multiplex System 1–8 (Part No. 0333) and DR Multiplex System 9–16 (Part No. 0334) offer the same level of multiplexing but use a dedicated read (DR) barcode design.

NuGEN
imagine more from less®

A Complete Solution for Strand-specific Gene Expression Studies

A limitation of some RNA-Seq methods is the inability to determine which genomic strand yields a given transcript or set of transcripts. This directionality or polarity of the transcript is important for correct annotation of novel genes because it provides essential clues about gene function. Many genomic regions can code for transcripts from both strands, and antisense transcription is thought to play an important regulatory role. RNA-Seq data, which includes strand-specific sequencing, provides information about potential transcript overlap as well as transcript orientation for a more complete understanding of expression levels in the presence of antisense transcripts.

As Illustrated in **Table 1**, we examined the utility of the Encore Complete RNA-Seq Library Systems for strand-specific sequencing with input total RNA from different species. For each of the RNA sources tested the results exhibit >95% exon read mapping in the sense orientation for RefSeq sequences. In addition, the overall alignment metrics show a high percentage of total aligned reads

TABLE 1 **Sequencing Alignment Metrics from Different Sources of Total RNA**

	Source of Total RNA				
	Human (MAQC B)	Human (MAQC A)	Chicken	Mouse	Rat
% of Total Reads					
Not aligned	4%	5%	17%	14%	16%
Aligned	96%	95%	83%	86%	84%
% of Mapped Reads by Category					
All non-rRNA	69.1%	76.8%	62.7%	63.9%	80.1%
Non-rRNA single site	50%	51%	54%	37%	55%
Non-rRNA multiple site	18%	25%	7%	23%	21%
All rRNA	30.9%	23.2%	37.3%	36.1%	19.9%
Mitochondrial rRNA (12S and 16S)	10%	5%	10%	15%	4%
Cytoplasmic rRNA (5.8S, 18S, and 28S)	20%	18%	27%	20%	15%
Distribution of RefSeq Reads					
Exons	44%	38%	33%	49%	42%
Introns	36%	36%	7.0%	20%	16%
Intergenic	16%	24%	60%	32%	42%
RefSeq Strand Retention (% of Reads in Sense Orientation)					
Exons	98.2%	95.7%	98.8%	98.1%	97.8%
5′ UTR	96.3%	96.1%	99.2%	98.6%	98.7%
3′ UTR	96.9%	93.6%	98.2%	96.7%	97.4%

The data were generated using 100 ng of total RNA from the indicated sources input to the Encore Complete RNA-Seq IL Multiplex System 1–8. The resulting libraries were sequenced on the Illumina Genome Analyzer IIx using 4 libraries per lane (4-plex) with 40 bp single-read sequencing. Total RNA from chicken, mouse, and rat are whole body male and female reference samples and, therefore, do not represent RNA from any specific tissue or cell types. Non-rRNA single site = % for reads mapping to non-rRNA sequences in the reference genome in a single location. Non-rRNA multiple site = % for reads mapping to non-rRNA sequences in the reference genome in two or more locations. The relative reduction of % aligned reads with chicken total RNA is due to the poor annotation of the available reference genome used in mapping.

FIGURE 2A **Strand-specific Expression of PTMS and MLF2**

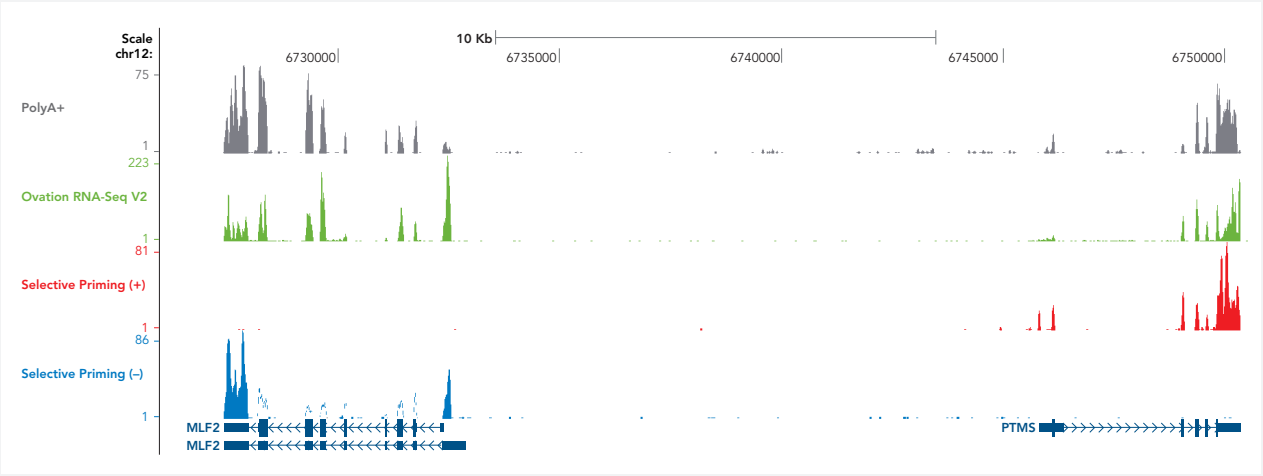
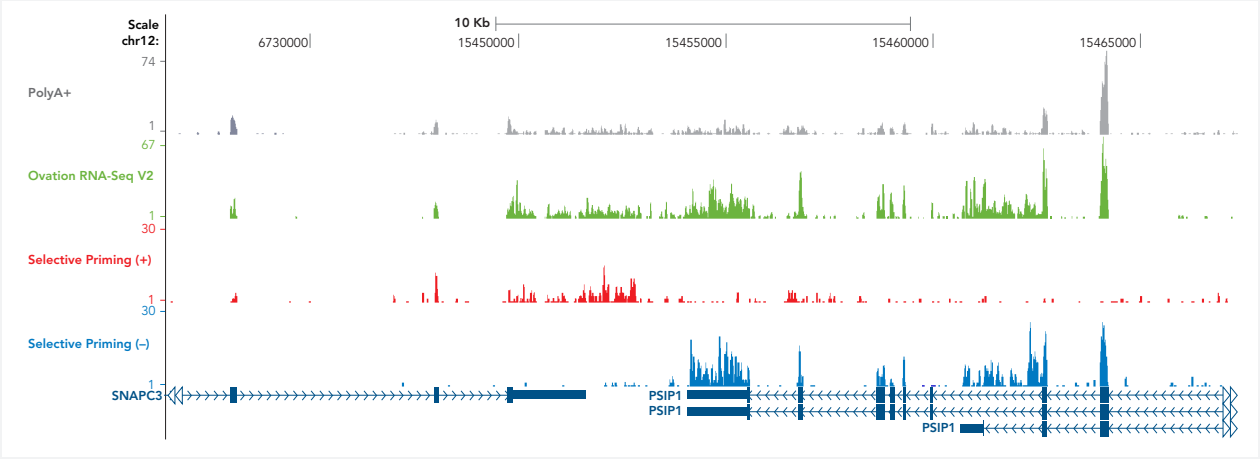


FIGURE 2B **Strand-specific Expression of SNAPC3 and PSIP1**

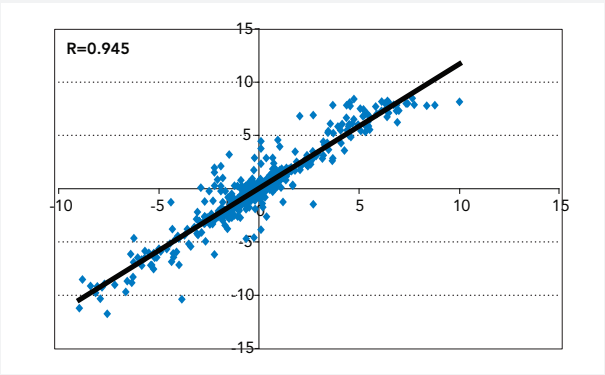


Total RNA from MAQC B (Human brain) was amplified using either the Ovation RNA-Seq System V2 (Part No. 7102) or poly(A)+ RNA selected or processed using the Encore Complete RNA-Seq IL Multiplex System 1–8 (Part No. 0312; tracks labeled Selective Priming). In the case of Ovation RNA-Seq System V2 and poly(A)+ samples, the resulting double-stranded cDNA was input to the Encore NGS Multiplex System I (Part No. 301) to construct NGS libraries. Single-read sequencing results were obtained using the Illumina Genome Analyzer IIx platform with 40-bp reads.

with a significant reduction in both mitochondrial and cytoplasmic rRNA reads. In contrast, protocols using random primers for cDNA synthesis and total RNA as the starting material are expected to yield ~70–75% rRNA reads. The reduction in sequencing reads derived from rRNA enables a higher representation of reads mapping to mRNA sequences in the final library, thereby facilitating more efficient and cost effective interrogation of coding sequences.

As a further example of the high fidelity for strand retention achieved with the Encore Complete RNA-Seq Library Systems, two examples of the sequencing read distribution between the plus and minus strands are shown in **Figure 2A** and **2B**. In Figure 2A sequencing read coverage plots are shown for PTMS and MLF2 transcripts. The results show read coverage across all exons, and illustrate how the sequencing reads segregate to the plus (+) strand for PTMS and the minus (-) strand for MLF2 when using the strand retention capability of the Encore Complete RNA-Seq Library System. In the case of data generated with either poly(A)+ selected RNA or total RNA processed with the Ovation RNA-Seq System V2 (Part No. 7102) the sequencing reads are expected to be distributed evenly over the plus and minus strands, and hence show no preferential strand retention. Similar results are shown in **Figure 2B** for reads mapping

FIGURE 3 **Comparison of Encore Complete RNA-Seq System Data with qPCR**



Differences between Log_2 transformed expression value for MAQC A and MAQC B samples, RNA-Seq data are plotted on X-axis, qPCR data plotted on Y-axis. Six hundred fifty-nine TaqMan probes that uniquely map to the RefSeq annotations used in the RPKM calculations are represented. RPKM stands for Reads Per Kilobase of exon model per Million mapped reads. The RPKM measure of read density reflects the molar concentration of a transcript in the starting sample by normalizing for RNA length and for the total read number in the measurement.

ORDERING INFORMATION

Part No.	Product Name
0311	Encore® Complete RNA-Seq Library System
0312	Encore® Complete RNA-Seq IL Multiplex System 1–8
0313	Encore® Complete RNA-Seq IL Multiplex System 9–16
0333	Encore® Complete RNA-Seq DR Multiplex System 1–8
0334	Encore® Complete RNA-Seq DR Multiplex System 9–16
Related Products	
7102	Ovation® RNA-Seq System V2
7150	Ovation® RNA-Seq FFPE System
Technical Documents	
Encore Complete RNA-Seq Library Systems User Guide	

to SNAPC3 and PSIP1 which are overlapping transcripts. Reads for SNAPC3 segregate to the plus (+) strand, and reads from PSIP1 map preferentially to the minus (–) strand.

Differential Gene Expression

One of the primary uses of RNA-Seq is to determine differential gene expression between different tissues, cells or experimental conditions. As illustrated in **Figure 3**, total RNA derived from MAQC A and B samples was analyzed by RNA-Seq to compare with differential expression data generated by quantitative PCR (qPCR) using the TaqMan assay. Differential expression calls are concordant with this reference expression assay (R = 0.945) without significant data compression. These results demonstrate that differentially expressed genes can be accurately quantified with RNA-Seq libraries prepared with the Encore Complete RNA-Seq Library Systems from NuGEN.

Conclusion

The Encore Complete RNA-Seq Library Systems offer a number of advantages for researchers engaged in RNA-Seq studies:

- An end-to-end solution for strand-specific RNA-Seq — All required components for preparation of strand-specific RNA-Seq libraries from moderate amounts of total RNA without the need for rRNA reduction or poly(A) selection
- Superior strand retention — Using a range of sources for total RNA researchers can expect to achieve >95% read mapping in the sense orientation for accurate and complete understanding of transcript profiling
- Affordable and scalable kit configurations — The barcoding capability provided with Encore Complete RNA-Seq IL Multiplex Systems 1–8 and 9–16, and Encore Complete RNA-Seq DR Multiplex Systems 1–8 and 9–16 improve sample throughput and reduce sequencing costs as your projects grow

NuGEN
imagine more from less®

NuGEN Technologies, Inc.

Headquarters USA
201 Industrial Road, Suite 310
San Carlos, CA 94070 USA
Toll Free Tel: 888.654.6544
Toll Free Fax: 888.296.6544
custserv@nugeninc.com
techserv@nugeninc.com

Europe
P.O. Box 149
6680 AC Bemmel
The Netherlands
Tel: +31-13-5780215
Fax: +31-13-5780216
europe@nugeninc.com

For our international distributors contact information, visit our website

www.nugeninc.com

M01245 v2