TECHNICAL REPORT

A Comparison of Two Sample Preparation Methods for Strand-specific RNA-Seq Using Human Whole Blood

Introduction

When performing RNA-Seq on libraries from human whole blood, reads from globin RNAs and rRNAs may overwhelm sequencing reads of greater interest. Several commercial sample preparation kits are intended to address this problem. These kits selectively deplete only globin and rRNA transcripts, effectively enriching for other transcripts. Two of these kits are the Ovation® Human Blood RNA-Seq Multiplex System 1–8 (Part No. 0337-32) from NuGEN Technologies and ScriptSeq™ Complete Gold Kit (Blood) Low Input (Cat# SCL24GBL) from Epicentre — an Illumina company. Both of these kits start with total RNA to produce strand-specific RNA-Seq libraries with reduced globin and rRNA content.

Ovation Human Blood RNA-Seq Multiplex System 1–8 uses NuGEN's Insert Dependent Adaptor Cleavage (InDA-C) technology to enrich for non-rRNA and non-globin sequences during NGS library construction. This process effectively minimizes rRNA and globin sequences from the finished libraries.

ScriptSeq Complete Gold Kit (Blood) Low Input uses Epicentre's Globin-Zero™ technology and Magnetic Core Kit to remove globin and rRNA from the sample. In this workflow, globin and rRNA transcripts are removed from the total RNA pool by hybridization-based capture.

In this technical report, we describe a comparison of the performance these two RNA-Seq library preparation kits.

Materials and Methods

RNA-Seq libraries were prepared from PAXGene™ human whole blood total RNA using both the NuGEN Ovation Human Blood RNA-Seq Multiplex System 1–8 (PN 0337-32) and Epicentre ScriptSeq Complete Gold Kit (Blood) Low Input kits. The Epicentre system required additional PCR primers and PCR Enzyme mix to be purchased separately from the vendor. One pooled sample and one individual sample were used as input for each of the two kits. The protocol for each kit was run with both 100 ng (the recommended minimum for both kits) and 50 ng total blood RNA as input. Libraries were sequenced on the Illumina Genome Analyzer IIx platform using single-end sequencing with 40-bp reads.

Results

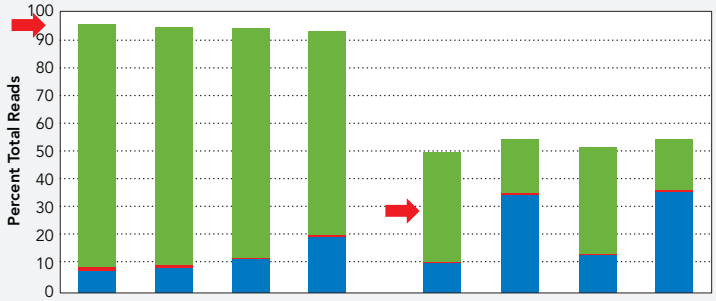
Library Preparation

The Ovation Human Blood RNA-Seq Multiplex System 1–8 produced sufficient yields to proceed to a standard workflow for cluster generation and sequencing on the GAIIx. The yields from the Epicentre kit were insufficient for a standard cluster generation workflow and required three additional rounds of PCR amplification before the yield was great enough for cluster generation.

Sequencing

As shown in **Figure 1**, the Ovation Human Blood RNA-Seq Multiplex System 1–8 produced libraries with

FIGURE 1. **Distribution of sequence reads**



	NG_A 100 ng	NG_A 50 ng	NG_B 100 ng	NG_B 50 ng		EC_A 100 ng	EC_A 50 ng	EC_B 100 ng	EC_B 50 ng
Informative	87.7%	86.4%	83.6%	74.1%		40.3%	20.2%	39.2%	18.9%
Globin	1.1%	1.2%	0.7%	0.7%		0.2%	0.8%	0.3%	0.7%
rRNA	6.9%	7.9%	11.3%	19.2%		9.8%	34.2%	12.7%	35.3%

Distribution of sequence reads for two libraries constructed using the NuGEN Human Blood RNA-Seq Multiplex System 1–8 (NG_A and NG_B) and two libraries constructed using the Epicentre ScriptSeq Complete Gold Kit (Blood) Low Input (EC_A and EC_B) with 100 ng and 50 ng total blood RNA input. Libraries were sequenced on the Illumina Genome Analyzer IIx platform using single-end sequencing with 40-bp reads.

NuGEN
imagine more from less®

a significantly higher percentage of aligned and informative reads. The NuGEN Ovation System had overall alignment rates greater than 90% for all sample types and inputs, whereas the Epicentre kit alignment rates were below 60%. As a result, percentage of 'Informative' reads, defined as reads that are not derived from rRNA or globin genes, were 2–3x greater when using the NuGEN Ovation System.

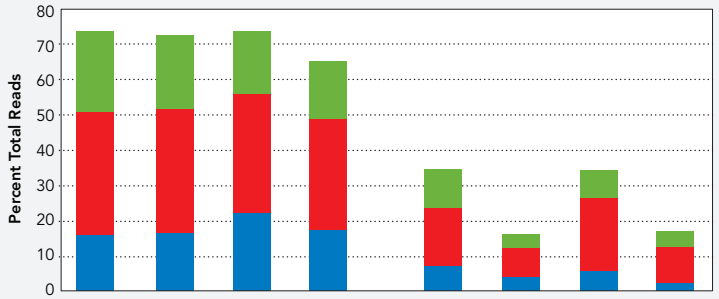
Figure 2 illustrates that the Ovation Human Blood RNA-Seq Multiplex System 1–8 also produced libraries with a greater percentage of reads mapping to RefSeq. More than double the percentage of reads mapped to RefSeq with the NuGEN System for each sample tested. The NuGEN system enables much more efficient use of sequencing space to provide information on non-rRNA and non-globin mRNAs at low inputs. This effect is even more pronounced when the input quantities fall below the recommendations of each manufacturer. The NuGEN System has a greater tolerance for mis-measurement of input or otherwise compromised samples that may be challenging to obtain.

Table 1 shows the number of RefSeq genes detected in each sample based upon the kit used. NU_A and NU_B represent the two libraries generated using the Ovation Human Blood RNA-Seq Multiplex System. EC_A and EC_B represent the two libraries generated using the Epicentre ScriptSeq Complete Gold Kit (Blood) Low Input. Data is shown for both 50 ng and 100 ng input of total blood RNA. The Ovation Human Blood RNA-Seq Multiplex System 1–8 kit detected significantly more genes under all conditions than were detected in the libraries produced by the Epicentre system.

Both systems demonstrated good 5′–3′ transcript coverage, although the inclusion of both random priming and poly dT first strand priming in the NuGEN kit led to better 3′ coverage (data not shown).

Table 2 provides a comparison of the performance of the two kits in several key areas. The results show that both the NuGEN Ovation Human Blood RNA-Seq Multiplex System 1–8 produced a greater percentage of aligned reads with far fewer duplicates

FIGURE 2. **Distribution of reads within RefSeq genes**



	NG_A 100 ng	NG_A 50 ng	NG_B 100 ng	NG_B 50 ng		EC_A 100 ng	EC_A 50 ng	EC_B 100 ng	EC_B 50 ng
Exon	22.7%	21.1%	17.4%	16.3%		10.9%	4.2%	8.2%	4.6%
Intron	35.0%	34.7%	34.1%	31.6%		16.5%	7.9%	20.6%	10.0%
Intergenic	16.0%	17.0%	22.2%	17.4%		7.3%	4.1%	5.9%	2.6%

TABLE 1. **Genes detected**

NG_A 100 ng	NG_A 50 ng	NG_B 100 ng	NG_B 50 ng	EC_A 100 ng	EC_A 50 ng	EC_B 100 ng	EC_B 50 ng	NuGEN / Epicentre	FPKM Threshold
11546	11627	11763	11690	10008	9063	10038	9623		0.3
10486	10482	10312	10427	9320	8561	9261	9068		0.5
9072	8971	8447	8683	8109	7450	7867	7974		1.0

TABLE 2. **Performance at 100 ng total RNA input**

System	NuGEN Ovation Human Blood RNA-Seq Multiplex System 1–8	Epicentre ScriptSeq Complete Gold Kit (Blood) Low Input
Alignment	>95%	50–75%
% duplicates	21–27%	60–75%
Globin reduction*	<2% globin	<2% globin
rRNA reduction*	<12% rRNA	<25% rRNA
Mapped to Exon**	22–34%	8–11%
RefSeq Coverage	More transcripts detected	Fewer transcripts detected

* Percentage of aligned reads
** Percentage of all reads

and two-fold greater rRNA reduction when compared to the Epicentre kit. The NuGEN kit resulted libraries with more RefSeq transcripts detected and nearly twice the percentage of reads mapping to exons.

Conclusion

Both the Ovation Human Blood RNA-Seq Multiplex System 1–8 kit and the Epicentre ScriptSeq Complete Gold Kit (Blood) Low Input displayed excellent ability to reduce the number of globin RNAs incorporated into the RNA-Seq libraries while retaining strand information. However, the NuGEN kit was clearly superior in reducing the percentage of rRNAs and generating informative reads aligning to RefSeq. Ultimately, the NuGEN kit produced a greater percentage of informative reads, creating greater capabilities for discovery.

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 109
9350 AC Leek
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

M01361 v1