ASHG 2019: 1874F

Integration of Molecular Chemistries Supporting a Full-Length mRNA Sequencing Library Preparation Method on a Microfluidic Circuit

Michael Phelan, Julie Alipaz, Joel Brockman, Sangpen Chamnongpol, Brian Fowler, Jennifer A. Geis, Tom Goralski, Christopher Kubu, Raphael Kung, Benjamin Lacar, Charles Park, David A. King

Introduction

RNA sequencing (RNA-seq) has become the gold standard of expression profiling methods. We have developed an elegant microfluidics-based chemistry and workflow leveraging the Juno™ instrument called the Advanta™ RNA-Seq NGS Library Prep Kit (PN 101-9187). Our RNA-Seq Kit supports simultaneous processing of up to 48 samples with a one-click script on our instrument.

The Juno NGS system automates the RNA-seq workflow along with a new, nanoscale integrated fluidic circuit (IFC) called 48.Atlas™, which is the size of a standard microtiter plate. The Advanta RNA-Seq NGS Library Prep Kit also includes reagents necessary to generate full-length, stranded RNA-seq libraries from the polyadenylated RNA fraction present in as little as 10 ng of total RNA from eukaryotic organisms. Herein, we show comprehensive performance characteristics from several of our internal studies.

Methods and Materials

In order to determine assay robustness, we conducted an internal analytical validation study using 3 different operators on 6 different instruments with 3 different reagent and 3 different IFC lots. Over 900 samples were sequenced comprising ~5 billion total paired-end 75 bp reads. RNA samples were Universal Human Reference RNA (UHRR, Agilent® PN 740000) and human brain RNA (BioChain® PN R1234035). Input amounts were 10 ng and 100 ng of total RNA.

In order to further assess assay robustness, we performed two additional studies. In the first study, we used 20 different tissue types with varying RIN scores (RNA Integrity Number; Agilent). The different tissue samples were correlated with their RINs against number of genes and transcripts detected. In the second study, we conducted a dilution series of UHRR that included input RNA amounts below our minimum of 10 ng. The various input amounts were assessed for percent of reads mapped to the transcriptome (RefSeq). Also, average gene-level Pearson's correlation at all comparison input amounts was calculated. Note: Averages were calculated and projected back after the appropriate Fisher transformation.

The final study was conducted to support differential expression analyses of alternatively spliced isoforms. We used four replicates each of 10 ng of UHRR and human brain RNA. The replicates were sequenced on an Illumina® HiSeq® 2500 system using paired-end 75 bp reads and then combined for analysis by sample type.

For all of the above studies, FASTQ files were first aligned using STAR and kallisto to the human genome reference GRCh38 and to the NCBI RefSeq transcriptome followed by quantification of both genes and transcripts. Sequencing metrics and figures were generated using BEDTools, SAMtools, Picard, Python®/pandas and R/ggplot2/limma/GenomicAlignments/Rsamtools.

Conclusions

Results herein have shown excellent mapping rates of >80% with low-percent rRNA reads of <10%. Technical replicate correlations were observed to be >98% in all conditions. Also, our additional studies demonstrated our kit is robust to input amounts below 10 ng and supports RNA with RIN numbers less than 7. We also provided clear evidence of differential isoform expression.

The Advanta RNA-Seq NGS Library Prep Kit provides true walkaway automation that substantially minimizes manual pipetting steps and hands-on time. This technology enables significantly reduced reagent consumption, which can be helpful for laboratories in managing costs.

Results

Juno system enables an automated, cost-effective approach to RNA sequencing

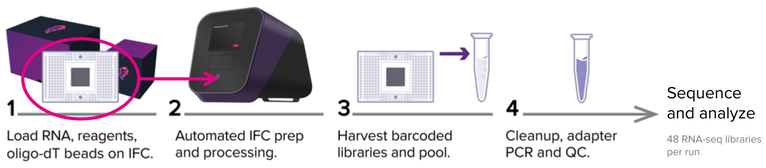

Figure 1. Samples and Advanta RNA-Seq reagents are added to the 48.Atlas IFC, which is subsequently processed on the Juno instrument. The system solution automates many tedious hands-on steps to generate up to 48 RNA-seq libraries. The nanoscale design of the 48.Atlas IFC significantly reduces reagent consumption, which helps minimize overall costs per sample.

New 48.Atlas IFC format for solid-phase poly(A) RNA capture and multi-step reactions enabling walkaway automation

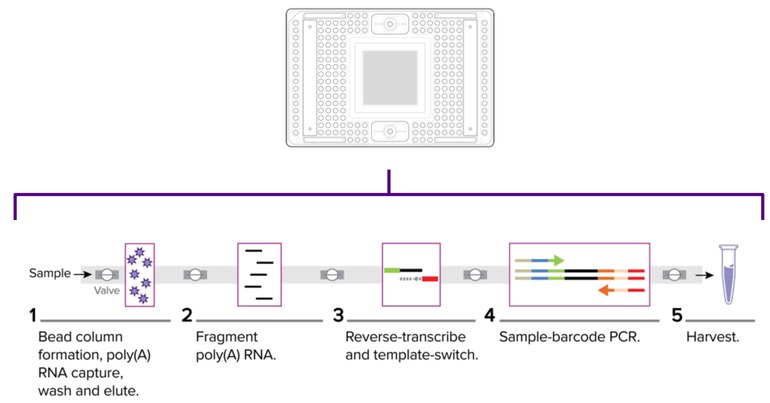

Figure 2. The 48.Atlas IFC architecture automates multiple workflow steps otherwise performed manually, including poly(A) RNA capture, RNA fragmentation, reverse-transcription, sample-barcode PCR and multiple wash steps. The entire process can be completed in ~12.5 hr with only ~2 hr of hands-on time. Following on-IFC barcode PCR, libraries are harvested, followed by purification and quantification prior to sequencing.

Metrics	Human Total RNA Samples, Average (SD)				
	UHRR 10 ng N = 234	UHRR 100 ng N = 231	Brain 10 ng N = 218	Brain 100 ng N = 234	Overall N = 917
Percent reads mapped to genome (not including mtRNA and rRNA)	87.4% (3.2%)	86.8% (2.0%)	82.7% (2.4%)	80.1% (2.2%)	84.3% (3.9%)
Percent reads mapped to transcriptome (RefSeq)	77.2% (3.0%)	74.2% (2.1%)	72.1% (2.2%)	67.0% (2.0%)	72.6% (4.4%)
Percent ribosomal RNA (rRNA) reads	3.0% (0.6%)	4.9% (0.7%)	5.4% (0.5%)	8.4% (0.7%)	5.4% (2.1%)
Percent unmapped reads	4.1% (3.1%)	3.8% (1.8%)	3.7% (2.2%)	4.1% (1.9%)	4.0% (2.3%)
Pearson's correlation of technical replicates within input amounts (10 ng vs. 10 ng; 100 ng vs. 100 ng)	0.984 (0.227)	0.992 (0.374)	0.980 (0.393)	0.992 (0.346)	0.988 (0.394)
Pearson's correlation of technical replicates between input amounts (10 ng vs. 100 ng)	0.983 (0.272)		0.978 (0.357)		0.981 (0.324)
Percent correct strandedness	98.1% (0.2%)	98.6% (0.12%)	98.1% (0.2%)	98.5% (0.1%)	98.3% (0.3%)
Library fragment size	199.7 (19.4)	193.8 (20.2)	195.9 (18.7)	192.5 (17.9)	195.5 (19.2)

Table 1. Performance characteristics of the Advanta RNA-Seq Kit on Juno were assessed in an analytical validation study. The study was conducted using 3 Advanta reagent lots and 3 48.Atlas IFC lots across 6 Juno instruments by 3 different operators. In total, more than 900 samples were sequenced comprising ~5 billion reads.

Gene and transcript detection across tissues with varying sample quality

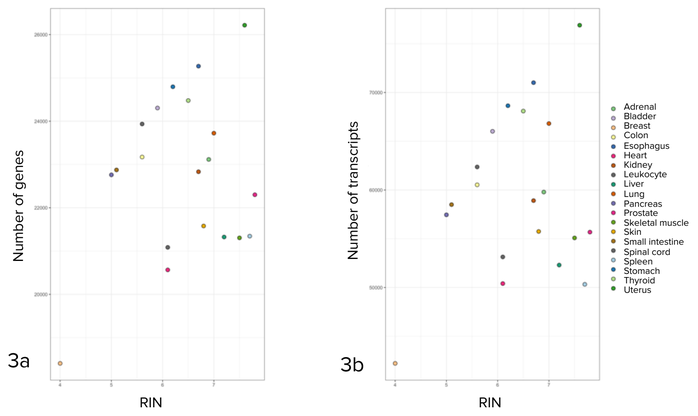

Figure 3. Robust gene (3a) and transcript (3b) detection was observed from RNA extracted from samples of varying tissue origin and quality (RIN) using the Advanta RNA-Seq NGS Library Prep Kit. 100 ng was the input amount and samples were sequenced on an Illumina NextSeq™ instrument to an average read depth of ~20M reads per sample.

Percent of reads mapped to transcriptome (RefSeq) and average gene expression correlation between various input amounts

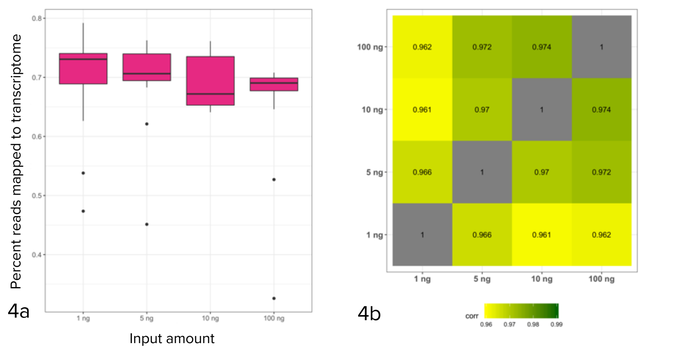

Figure 4. UHRR samples were sequenced with input amounts of 1 ng, 5 ng, 10 ng and 100 ng for a total of 64 samples. 10 ng to 100 ng are the recommended inputs amounts. Percent of reads mapped to the transcriptome (RefSeq) were calculated (4a) for the varying input amounts along with average pairwise correlations (4b) between inputs amounts. This demonstrates no cliff effect below 10 ng, supporting excellent system robustness.

Differential transcript and exon expression between brain and UHRR samples for the gene *GFAP*

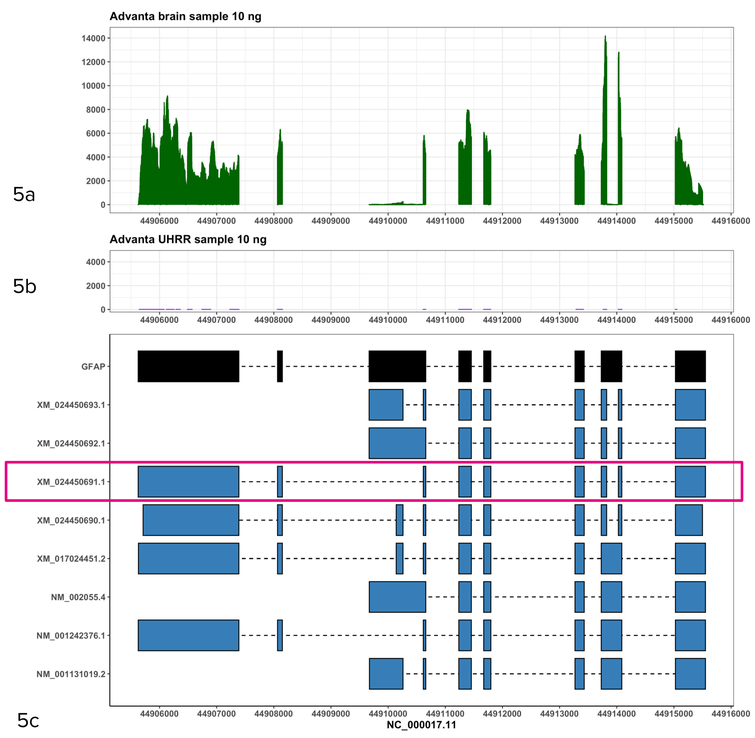

Figure 5. Read pile-ups for brain (5a) 10 ng (n = 4) and UHRR (5b) 10 ng (n = 4) samples processed with the Advanta Kit for the glial fibrillary acidic protein gene (*GFAP*). Corresponding reference transcripts plotted in panel 5c. As expected, little to no expression was observed in UHRR samples (relative to brain samples), consistent with its known expression in the central nervous system. The most likely isoform being expressed is shown in 5c, boxed in red. FASTQ files were first down-sampled to 30M reads per sample to equalize read depth.

Fluidigm Corporation
7000 Shoreline Court, Suite 100
South San Francisco, CA 94080 USA
+1 650 266 6000 • Toll-free in the US and Canada: 866 359 4354
fluidigm.com