

Authors

Theresa Ten Eyck¹,
Kyoko E. Yuki²,
Stephanie Bannister³,
Lianna Kyriakopoulou^{2,4},
Adam Shlien^{2,4}, and
Michael D. Wilson^{2,4}

¹ Agilent Technologies, Inc.

² The Hospital for Sick
Children,
Toronto, ON, Canada

³ Lexogen GmbH,
Vienna, Austria

⁴ University of Toronto,
ON, Canada

Automation of the Lexogen QuantSeq 3' mRNA Kit on the Agilent NGS Workstation Produces High-Quality Sequencing Libraries

Abstract

Lexogen QuantSeq 3' mRNA kits use oligo-dT primers to capture short fragments at the 3' end of poly-adenylated RNA, allowing for accurate expression profiling from various sample types, including formalin-fixed, paraffin-embedded (FFPE) tissue. The kits have options for low (0.1 to 10 ng total RNA) or standard (10 to 2,000 ng total RNA) input as well as dual or single indexing. QuantSeq was automated on the Agilent NGS Workstation. The various protocol options have been implemented in a single user interface for flexibility and ease of use. The NGS Workstation reduces hands-on time and maintains final library metrics such as concentration, mappability, and input-output correlation. Here, sequencing data generated from the automated protocol are analyzed. Universal Human Reference RNA (UHRR) with External RNA Control Consortium (ERCC) spike-ins was used to prepare eight libraries each of 100 ng and 250 ng input. The results show that libraries prepared on the NGS Workstation produce high-quality libraries with consistent size and concentration, high mappability, low percentage of ribosomal RNA (rRNA), and high input-output correlation, as calculated from ERCC spike-ins.

Introduction

QuantSeq is a straightforward method to prepare highly strand-specific mRNA libraries without the use of an additional poly-adenylated RNA enrichment or ribosomal depletion step. The kit comes in three types:

- FWD, generates reads toward the poly(A) tail that correspond to the mRNA sequence during Read 1 sequencing
- REV, generates reads corresponding to the cDNA sequence during Read 1 sequencing
- Flex, to sequence targeted regions using custom primers

Figure 1 displays and describes the QuantSeq method. This method produces one fragment per transcript, allowing for straightforward analysis because the number of reads mapped to a transcript is proportional to its expression. Samples can be highly multiplexed using the 384 available barcode combinations. The data analysis is simpler and more accurate than with standard RNA-Seq, as it is based on gene counting, and does not require FPKM calculations.

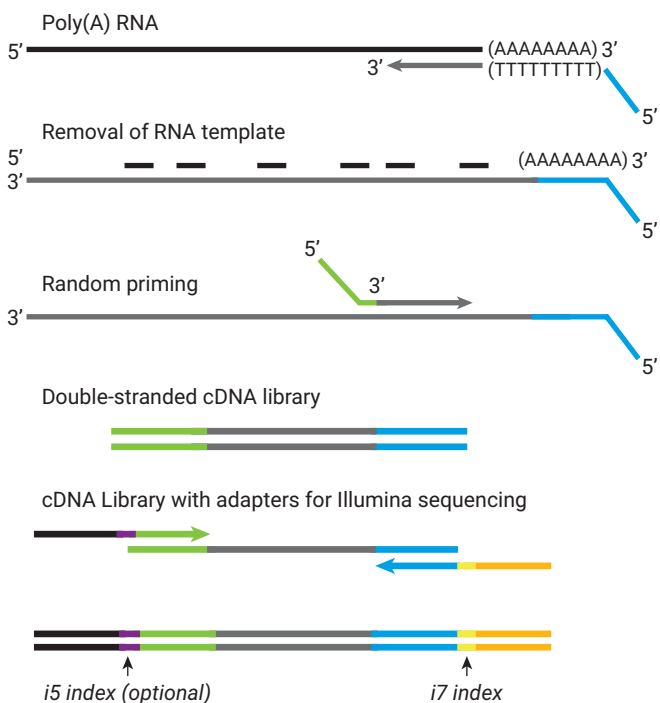


Figure 1. The QuantSeq 3' mRNA protocol produces mRNA libraries in 4.5 hours. Total RNA is reverse transcribed using oligo-dT primers. The RNA template is removed and the second strand is synthesized using random or targeted primers. The cDNA library is purified before being amplified with barcoded Illumina adapters. A final purification removes extra adapters that can interfere with sequencing.

Although the QuantSeq method requires minimal hands-on time, it can be reduced further by introducing automation. Additionally, automation reduces user error and increases throughput and reproducibility. The Agilent NGS Workstation is an automation platform featuring the Agilent Bravo Automated Liquid Handling Platform, Agilent BenchCel 4R Microplate Handler, and the Agilent Labware MiniHub (Figure 2). The Bravo has accessories for heating, cooling, shaking, and magnetic bead separation, allowing for most incubations to be completed on deck.

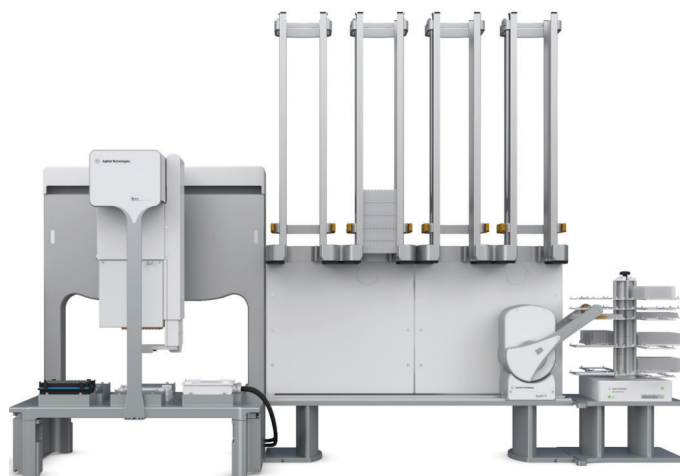


Figure 2. The Agilent NGS Workstation facilitates processing of up to 96 samples with maximum walkaway time. The NGS Workstation features the Agilent Bravo liquid-handler (left), Agilent BenchCel 4R microplate handler with labware stackers (center), and MiniHub reagent carousel (right). The Bravo deck has nine deck positions, including two heating and cooling positions, one chiller position, one orbital shaker, and a removable 96-well magnet. The BenchCel brings tip boxes, labware, and reagents to and from the Bravo deck, facilitating maximum walkaway time.

The automation protocols were written in VWorks Automation Control software. A VWorks form interface was created to capture user input, such as protocol and reagent version, and display the Workstation setup (Figure 3). The form allows discrete, flexible protocols to be run from a familiar, easy-to-use interface. The protocol can accommodate simultaneous processing of 8, 16, 24, 36, 48, or 96 libraries per run.

The protocol supports two versions, V1 and V2, of the QuantSeq kit. V1 and V2 version terms are introduced for automation purposes only. V1 refers to the group of reagents named as QuantSeq 3' mRNA-Seq Library Prep Kit for Illumina supplied before February 7, 2017. V2 refers to the group of reagents named as QuantSeq 3' mRNA-Seq Library Prep Kit for Illumina supplied after February 7, 2017.

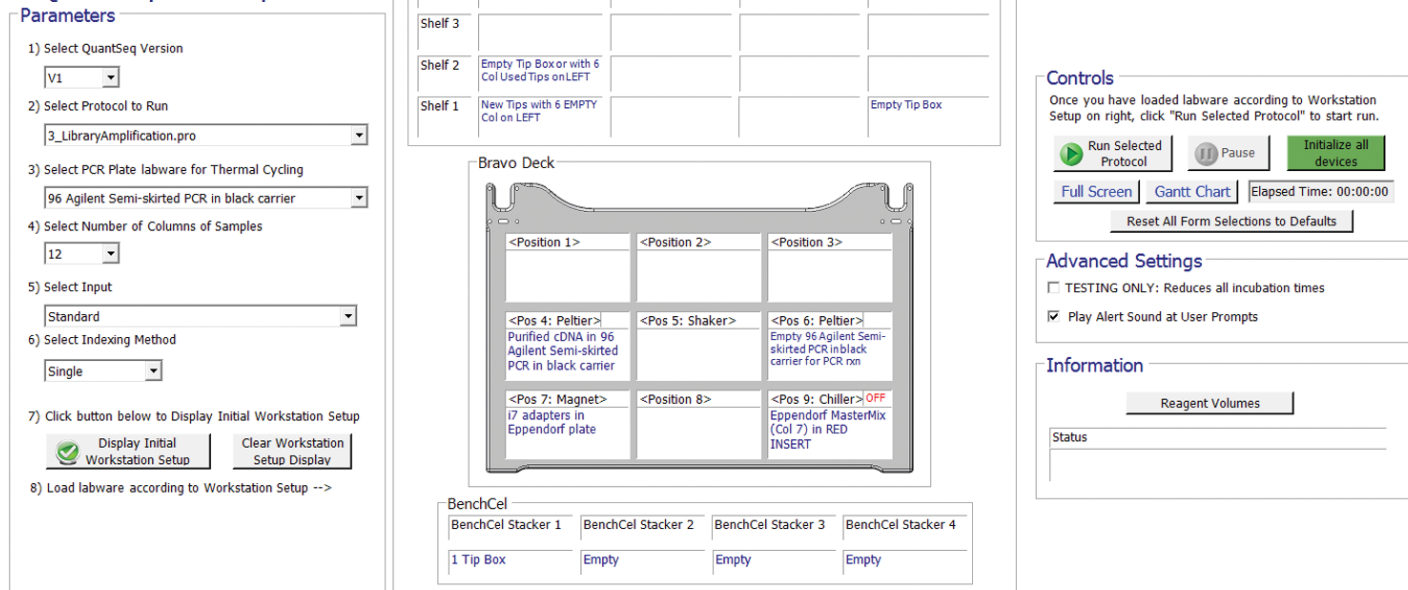


Figure 3. The VWorks interface for Lexogen QuantSeq enables easy and flexible automation of multiple desired workflows. The VWorks form has adjustable parameters that permit the user to run V1, V2, REV, FWD, and Flex kits from the same interface. On the left side, the user selects (1) the reagent version, (2) which protocol they would like to run, (3) what labware they would like to use, (4) the number of samples they would like to process, (5) the input, and (6) the indexing method. Once the parameters have been set, the user can click **Display Initial Workstation Setup** to update the central Workstation Setup display with the labware that needs to be placed on the instrument prior to running the protocol. Once the deck has been set up, the user can click **Run Selected Protocol** on the right side to start the protocol.

Materials and Methods

Before validating the NGS Workstation, the Bravo temperature settings, pipetting, and required dead volumes were optimized using the autoQuantSeq Dummy Automation Module (Lexogen), which provides liquids with the same physical properties as the standard reagent kits. After optimization using the Dummy Module, real reagents were used to verify pipetting accuracy.

To validate the performance of the NGS Workstation, UHRR (Agilent Technologies) libraries were prepared at two different input quantities using the QuantSeq 3' FWD kit. Eight 100 ng libraries and eight 250 ng libraries were prepared simultaneously with the NGS Workstation, using the V2 protocol.

ERCC RNA Spike-in Mix (Thermo Fisher) was added to samples as per the manufacturer's protocol¹. Briefly, ERCCs were diluted 1:100 or 1:1,000, and 0.5 µL or 2 µL were added to each sample for 250 ng or 100 ng total RNA input, respectively.

The libraries were prepared using the NGS Workstation with VWorks software version 11.3.0.1195. The following parameters were used:

QuantSeq version	FWD (V2)
PCR Plate type	ABI MicroAmp Optical 96-well plate
Number of columns of sample	2
Input	Standard
Indexing	Single

The protocols used were as follows:

- 1_1_FirstStrandSynthesis.rst
- 2_2_V2_SecondStrandSynthesis_Purification.rst
- 3_3_LibraryAmplification.pro
- 4_4_FinalPurification.pro

Reagent volumes, including the dead volume necessary for automation, were calculated from the reagent volume spreadsheet, which is linked from the VWorks form (Figure 4).

The 100 ng input samples underwent 20 cycles of amplification, and 250 ng input samples underwent 17 cycles of amplification. Final purified libraries were assessed on the Agilent Bioanalyzer using the High Sensitivity DNA assay, prepared according to manufacturer specifications². The libraries were pooled such that 20 fmol per sample (based on Bioanalyzer concentrations) was present in one pool and sequenced across two lanes of Rapid Run on an Illumina HiSeq 2500.

Sequencing reads were mapped using STAR³. ERCC correlations were calculated from \log_{10} ERCC-counts (number of reads) compared to \log_{10} ERCC-input (number of molecules) for each sample.

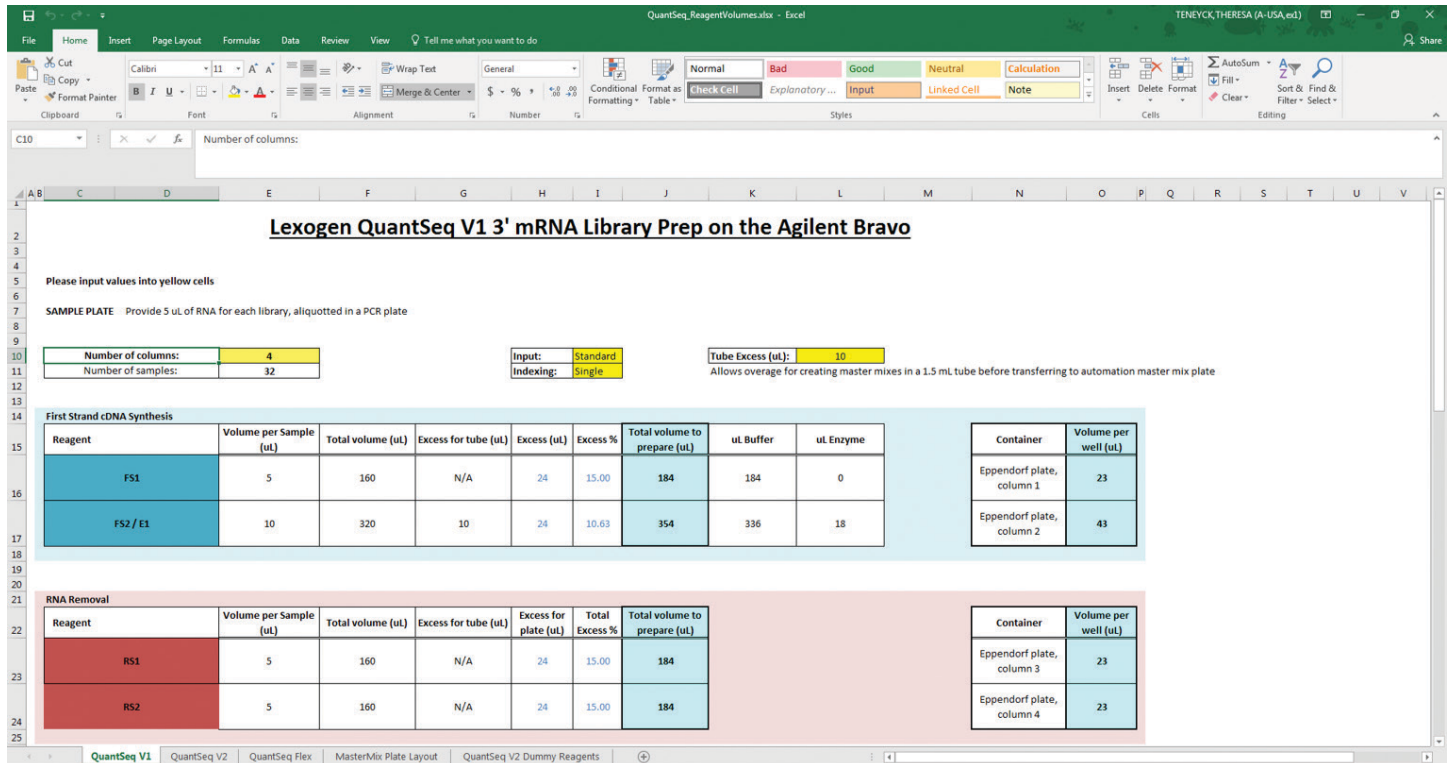


Figure 4. Reagent volumes have been precalculated with the reagent spreadsheet. The form (see Figure 3) links to a reagent volume spreadsheet. The user enters parameters such as the number of samples and input amount, and the spreadsheet outputs how much of each reagent to prepare, including dead volume for both manual and Agilent Bravo pipetting steps.

Results and Discussion

The Bioanalyzer results show fragments between 135 and 500 base pairs (bp) with a peak at approximately 205 to 225 bp (Figure 5). The Bioanalyzer-estimated concentrations

range from 5.8 to 11.3 ng/ μ L for the 100 ng input samples and from 3.1 to 12.1 ng/ μ L for the 250 ng input samples (Table 1). The samples were of consistent fragment sizes, high concentration, and comparable to manually-produced libraries.

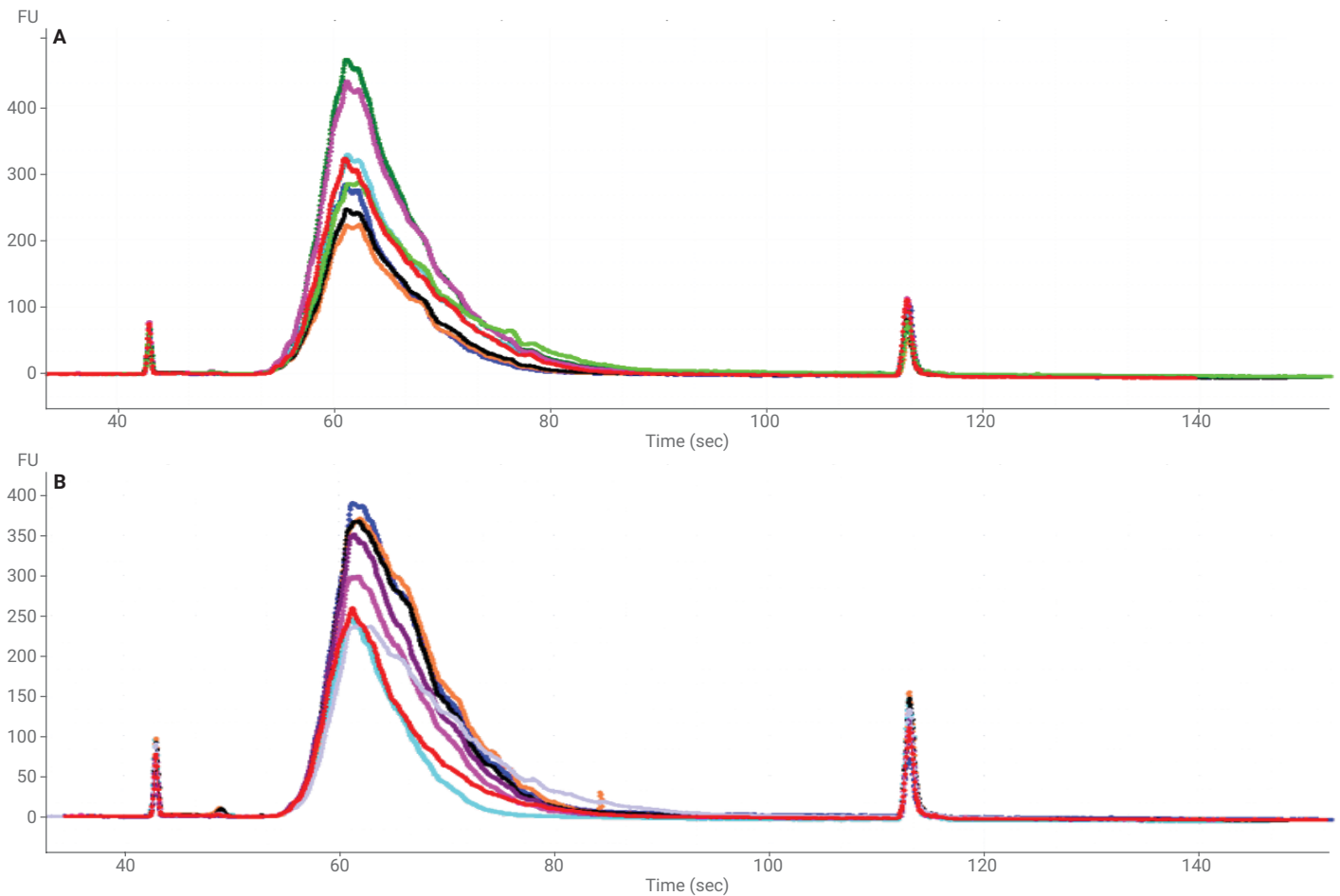


Figure 5. Bioanalyzer results show consistency across automated samples. The final libraries were analyzed using the Agilent High Sensitivity DNA Bioanalyzer assay. The libraries produced from the 100 ng input are displayed in (A), and the libraries produced with 250 ng input are displayed in (B). The electropherograms show reproducible size distribution and concentrations. Note that some samples within a group were run on separate Bioanalyzer chips; therefore, it is impossible to display the x-axis in base pairs. The 60-second tick mark corresponds to approximately 200 bp.

Table 1. Universal human reference RNA (UHRR) was used as input to the QuantSeq method. Sixteen samples were processed on the Agilent NGS Workstation, eight using 250 ng and eight using 100 ng input. The final library yields range from 3.1 to 12.1 ng/μL.

Sample ID	Input (ng)	Number of PCR cycles	Yield (ng/μL)
A1	100	20	7.2
A2			5.8
A3			11.3
A4			8.9
A5			10.1
A6			9.0
A7			7.0
A8			9.9
A9	250	17	5.1
A10			12.1
A11			3.1
A12			6.2
A13			6.3
A14			6.0
A15			7.1
A16			5.4

Table 2 provides the results from the sequencing data analysis. On average, the samples had 9.5 million reads each. The 100 ng samples had an average of 4.9 million reads each, while the 250 ng samples had an average of 14 million reads each. This variance in read counts may be due to the use of the Bioanalyzer for quantification before pooling. The 250 ng samples had a higher percentage of uniquely mapped reads, which is to be expected from the higher input. Additionally, the 250 ng samples had a lower proportion of rRNA (2.69 % ± 0.8) compared to the 100 ng samples (7.33 % ± 2.4).

Table 2. Libraries prepared with the Agilent NGS Workstation produce good quality sequencing data. The number of reads per library ranged from 3,130,794 to 29,917,053. On average, the libraries prepared with 250 ng input had higher uniquely mapped reads, a lower percentage of rRNA, and a higher number of detectable ERCC transcripts. Both inputs produced high-quality data.

Sample ID	Number of reads	Uniquely mapped reads (%)	rRNA (%)	Number of ERCC transcripts detected
A1	3,442,286	84.4	3.13	61
A2	6,546,644	78.7	9.40	60
A3	3,130,794	82.3	5.21	58
A4	5,708,823	81.2	6.66	63
A5	3,754,829	80.7	7.31	58
A6	4,006,940	77.9	10.87	63
A7	8,225,667	80.1	7.65	67
A8	4,683,424	78.8	8.44	62
A9	7,906,501	85.4	1.89	69
A10	6,042,257	84.0	3.20	70
A11	10,750,477	83.6	2.76	76
A12	12,295,976	85.1	2.04	70
A13	13,155,973	83.8	3.47	73
A14	29,917,053	83.4	3.90	77
A15	9,352,055	84.6	2.57	70
A16	22,481,940	85.7	1.65	75
Overall average	9,462,602	82.5	5.01	67
100 ng Average	4,937,426	80.5	7.33	62
250 ng Average	13,987,779	84.4	2.69	73

On a per sample basis, the number of reads mapping to each ERCC transcript was correlated with the number of molecules of that transcript in the sample. Of the 92 unique ERCC transcripts in each sample, an average of 62 were detected in the 100 ng samples, and 73 in the 250 ng samples. For both, transcripts that were not detected were present at the lowest concentrations (data not shown). The Pearson correlations are plotted in Figure 6. The data show that ERCC input and output are highly correlated across all samples. At the 250 ng input level, the average correlation was slightly higher (0.964 ± 0.011) than the 100 ng input level (0.957 ± 0.008).

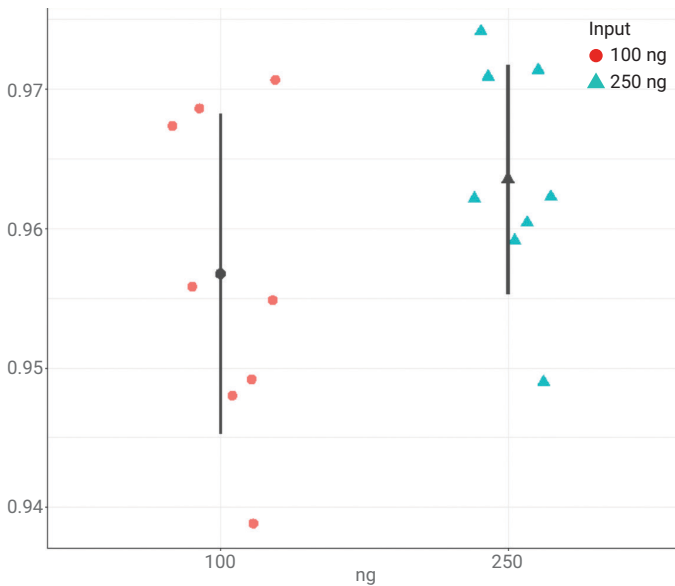


Figure 6. ERCC input and output are highly correlated across samples. The number of reads mapped to ERCCs was correlated with their concentration in each sample. The average Pearson correlation for 100 ng input libraries was 0.957. The average Pearson correlation for 250 ng input libraries was 0.964. In the plot, the red circles represent the 100 ng samples, and the blue triangles represent the 250 ng samples. The gray circle and gray triangle correspond to the means, and the gray bars represent the standard deviation.

Conclusion

The implementation of the Lexogen QuantSeq 3' mRNA protocol on the Agilent NGS Workstation has been shown to produce high-quality sequencing libraries. The resultant libraries, from two different input amounts, are of consistent size, good concentration, high unique mappability, low rRNA percentage, and high correlation between ERCC input and output. The NGS Workstation allows production of up to 96 libraries in 8.5 hours, with only 2 hours of hands-on time (total and hands-on time is lower with lower sample volumes). The automation protocol is easy to implement, and allows flexible automation of the many different types of QuantSeq kits.

Acknowledgements

The authors would like to acknowledge Sergio Pereira and Karen Ho (The Centre for Applied Genomics the Hospital for Sick Children, Toronto, Canada) for technical support and sequencing, Huayun Hou (SickKids and the University of Toronto) for bioinformatics support, and Lexogen bioinformaticians Andreas Tuerk, Gregor Wiktorin, and Michael Ante.

This work was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (OGI-110) to MDW and AS. MDW is supported by a Tier II Canada Research Chair and an Early Researcher Award from the Ontario Ministry of Research, Innovation and Science. KEY is supported by a Garron Family Cancer Centre and SickKids Research Training Fellowship.

References

1. Thermo Fisher, ERCC RNA Control Spike-in Mixes User Guide, [Online]. https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_086340.pdf
2. Agilent High Sensitivity DNA Kit Quick Start Guide, [Online], *Agilent Technologies*, http://www.agilent.com/cs/library/usermanuals/public/G2938-90322_HighSensitivityDNAKit_QSG.pdf
3. Dobin, A.; *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*(1), 15-21.

For More Information

Contact Agilent or Lexogen for a copy of the QuantSeq NGS Workstation protocol.

www.agilent.com/chem

For Research Use Only. Not for use in diagnostic procedures.

Information, descriptions, and specifications referenced herein are based upon the state of knowledge of Lexogen, GmbH at the time of publication and are subject to change without notice. No warranty as to their accurateness, completeness or suitability for a particular purpose is expressed or implied. Lexogen, GmbH shall not be liable for errors contained herein or for any damages in connection with the furnishing of the information, descriptions, and/or specifications referenced herein.