Increasing Complexity and Quality of Total RNA Sequencing at Scale

Nicole Francis[1], Jon Bezney[2], Can Kockan[1], Dan Goodman[1], Jason Lam[1], Stephen Montgomery[2], Stacey Gabriel[1] and Niall Lennon[1]

[1]Broad Clinical Labs, Broad Institute, Cambridge, MA 02141, [2]Stanford University, Stanford, CA 94305

Introduction

Bulk RNA sequencing has been largely centered on generating and analyzing mRNA libraries, as it is an efficient and cost effective approach to studying expression in coding regions. However, total RNA methods can provide a more holistic view of the transcriptome. This can include insight into non-coding regions with roles in regulatory mechanisms, and the detection of alternative splicing events and post-transcriptional modifications that can provide us with a deeper understanding of gene regulation and isoform diversity.

We performed a comparison of sequencing results between libraries generated with the Watchmaker Genomics Polaris Depletion & RNA LC Kit and libraries generated with a globin depletion & poly(A) selection method.

As this was a new library construction method our objectives were to:

1. Evaluate the quantity and quality of library generated as compared to industry-standard poly(A) mRNA methods.
2. Compare standard RNA sequencing metrics from this new workflow to the current RNA workflows used by the Broad Institute's Genomics Platform.
3. Assess transcript counts and protein coding genes identified within the total RNA workflow.
4. Determine if the assay was suitable for automation and was therefore scaleable.

Methods

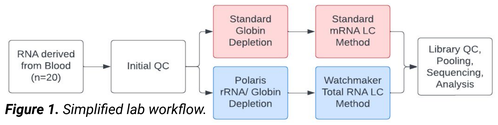


Figure 1. *Simplified lab workflow.*

Twenty total RNA samples derived from whole blood were run through an initial RiboGreen and Caliper QC. The RNA was split into two aliquots. One aliquot proceeded through a standard globin only depletion. Samples proceeded to a standard mRNA workflow, which included poly(A) selection, stranded cDNA synthesis, adapter ligation and PCR.

The second aliquot was run through the Watchmaker Polaris combined rRNA and globin RNA depletion workflow. Samples immediately proceeded into the Watchmaker RNA LC workflow which included stranded cDNA synthesis, adapter ligation and PCR.

Both sets of libraries (20 standard mRNA libraries and 20 total RNA Watchmaker Libraries) were QC'd using PicoGreen, pooled for sequencing, and run on a NovaSeq S2 flowcell. Data was downsampled to 10M reads/sample and aligned to hg38 using STAR aligner. Core RNA metrics including duplication, transcript counts and gene counts were evaluated and compared between the two methods.

Comparison of a standard mRNA workflow to the Watchmaker Total RNA Workflow

Evaluation of these workflows identified that both methods effectively removed globin RNA and both methods had similar rates of unmapped reads. There is a difference in the composition of the reads between the two methods (figure 2). Overall, the total RNA assay yielded more complex libraries at 10x higher quantities. The Watchmaker libraries also identified 15% more transcripts and 3% more protein coding genes than the standard poly(A) method.

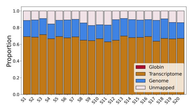


Figure 2a. *Read composition for 20 pilot samples run through the standard poly(A) method with depletion.*

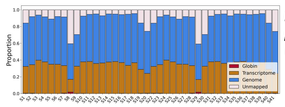


Figure 2b. *Read composition for 40 pilot samples run through Watchmaker Total RNA method.*

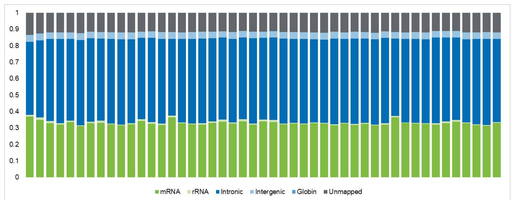


Figure 2c. *Read composition of 48 pilot RNA samples run through the automated total RNA Watchmaker workflow.*

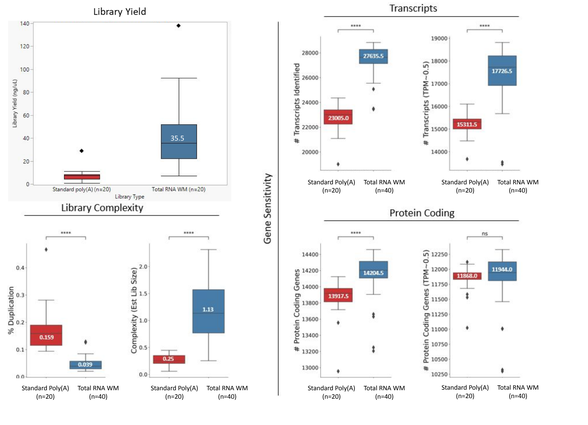
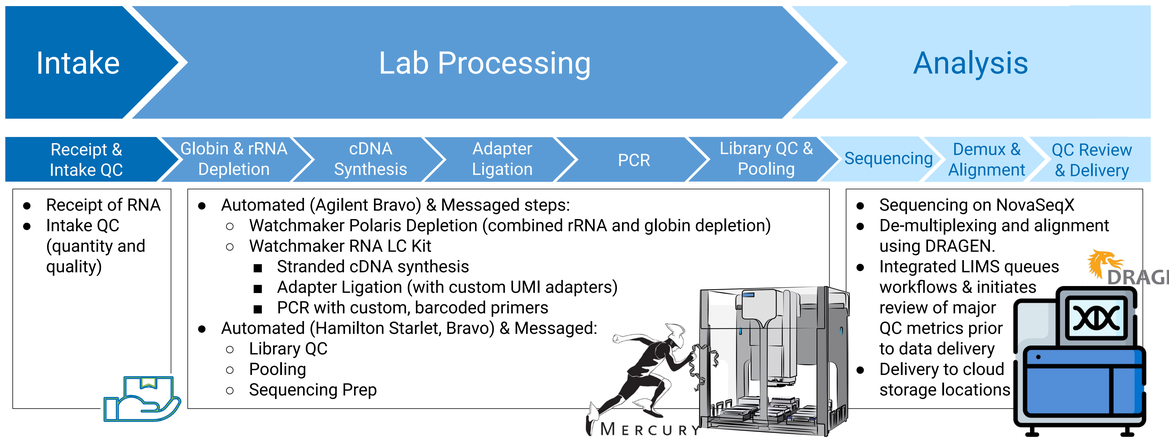


Figure 3*. Comparison of library yields, complexity, duplication and gene sensitivity between a standard poly(A) workflow with globin depletion and the Watchmaker Total RNA workflow.*

Fully Automated Method



Performance of Automated Workflow

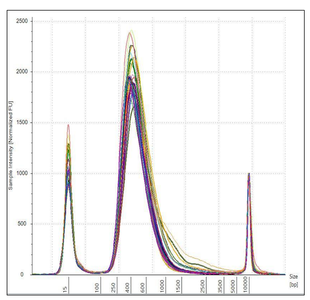


Figure 4. *An overlay of the library traces for 48 libraries that demonstrates the consistent quality and size coming out of automated WM LC.*

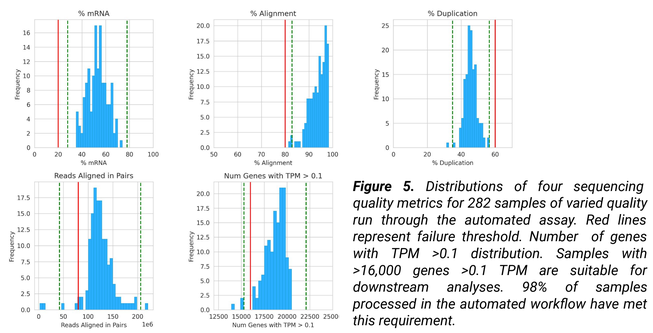


Figure 5. *Distributions of four sequencing quality metrics for 282 samples of varied quality run through the automated assay. Red lines represent failure threshold. Number of genes with TPM >0.1 distribution. Samples with >16,000 genes >0.1 TPM are suitable for downstream analyses. 98% of samples processed in the automated workflow have met this requirement.*

Discussion

- A fully automated Watchmaker Polaris Depletion and RNA LC process drastically reduces the amount of hands on lab processing time a standard poly(A) workflow requires (figure 6).
- The lab workflow is modularly scalable. With a single lab user being able to generate ~960 libraries a week (2 machines / 1 user).

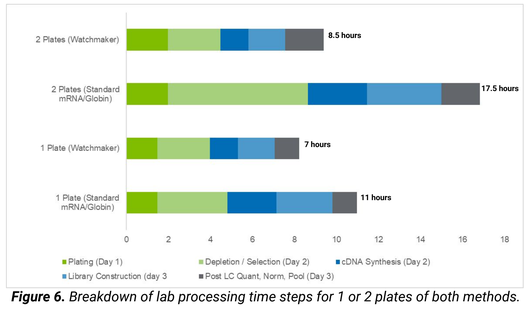


Figure 6. *Breakdown of lab processing time steps for 1 or 2 plates of both methods.*

- Efficient depletion and library construction chemistry generates quality libraries from varied input quantity and quality. This reduces the need for iterative upfront QC required by processes that need precise inputs.

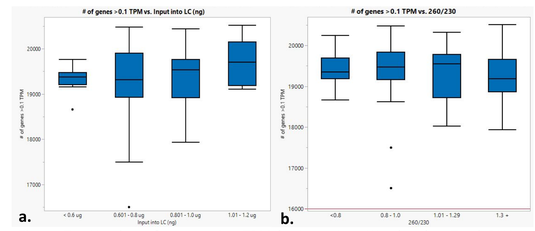


Figure 7a*. Number of genes >0.1 TPM by input into Watchmaker Total RNA method.* ***Figure 7b****. Number of genes >0.1 TPM by 260/230 measure of RNA before Watchmaker Total RNA method.*

- The high complexity and low duplication rates of these libraries allows for higher depth sequencing (80-100M RAP), with 30-40% of those reads mapping to mRNA (figure 2c). This method provides sequencing data that can be used to study the roles of RNA excluded in a poly(A) method without sacrificing the data's utility for standard expression analysis methods.
- An automated Watchmaker total RNA workflow is a scalable, time efficient and robust RNA method suited to generating data for downstream analyses including, but not limited to, expression analysis.

Acknowledgements

Data used in this poster was generated at BCL. For more information please visit: https://broadclinicallabs.org/

WATCHMAKER GENOMICS