Generating Higher Quality Transcriptomic Data from Challenging Sample Types

Nicole Francis[1], Dan Dubinsky[1], Can Kockan[1], Jason Lam[1], Cole Walsh[1], Stacey Gabriel[1] and Niall Lennon[1]

[1]*Broad Clinical Labs, Broad Institute, Cambridge, MA 02141*

Introduction

High-quality transcriptomic data from tissues provides valuable insights into gene expression patterns specific to various tissue types, enhancing our understanding of biological processes, disease mechanisms, and treatment effects. However, current RNA library preparation methods often exclude samples of lower quality or are prohibitively costly and inefficient. Tissues that are challenging to collect or typically yield poor-quality RNA present significant barriers, increasing costs and complicating library generation. Factors such as preservation methods, degradation over time, and the intrinsic characteristics of the tissue can contribute to diminished RNA quality. Collaborators may find it difficult to acquire optimal tissue samples and often have to work with what is available to them.

To support researchers working to advance our understanding of biology, Broad Clinical Labs evaluated products from Watchmaker's RNA portfolio to establish an automated, scalable, robust, and cost-effective bulk RNA library preparation method that accommodates tissues and RNA of varying quality. This evaluation included input titrations, quality titrations as determined by RIN/DV200 metrics, tissues sourced from diverse sites and samples with poor 260/230 ratios.

Incoming Sample Quality

While processing samples for large projects, BCL often notices discrepancies in quality for incoming RNA within a single project. Often these discrepancies are connected to different sites doing the extractions and upfront processing (Figure 1). Generating high quality data for these projects, regardless of incoming quality requires a process that produces consistent data regardless of input.

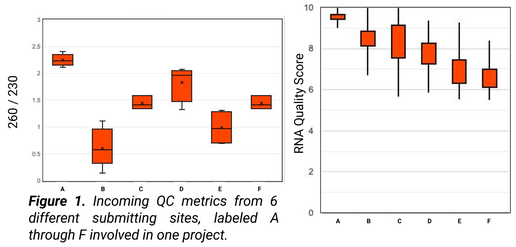


Figure 1. *Incoming QC metrics from 6 different submitting sites, labeled A through F involved in one project.*

Sample performance through a standard poly(A) method is significantly impacted by the incoming quality of RNA (Figure 2).

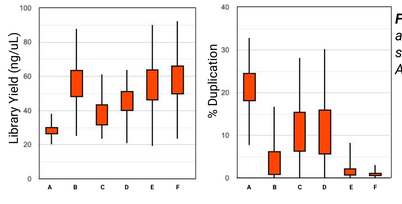
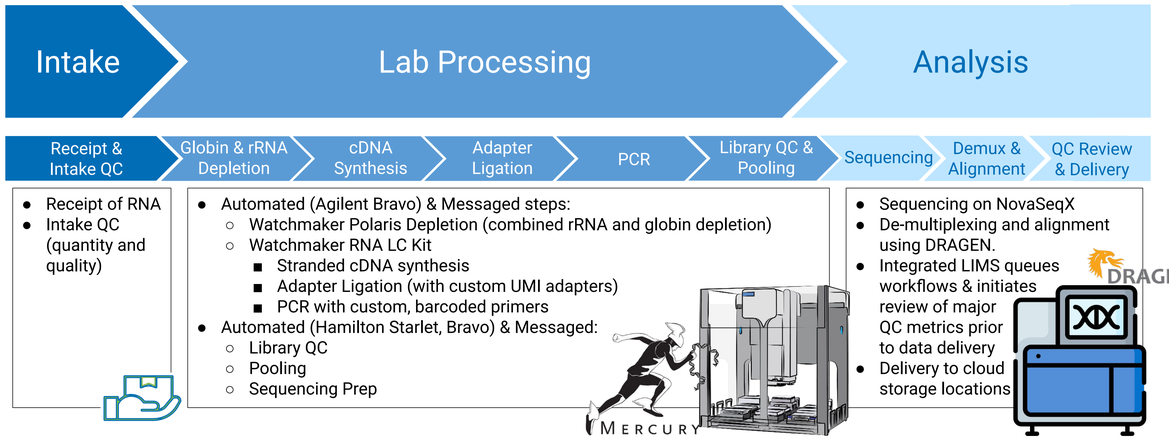


Figure 2. *Library yield and duplication for samples from sites A-H from figure 1.*

Fully Automated Workflow



Degradation & Library Performance

RNA with a range of RIN scores were run through Watchmaker's two total RNA methods alongside a standard poly(A) method. Library and sequencing performance across samples demonstrated that the total RNA method performed more reliably across all RINs. Importantly, RNA with lower RINs (<6) performed more consistently and with improved quality metrics when processed through a total RNA workflow, where the same samples saw a decrease in library quality as RINs decreased.

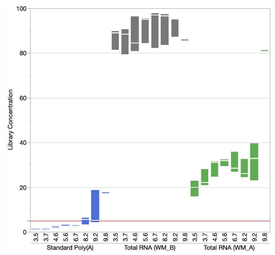


Figure 3. *Library yields of the same samples (RIN 3.5 - 9.8) processed through a standard poly(A) method, and Watchmaker's Total RNA methods A&B.*

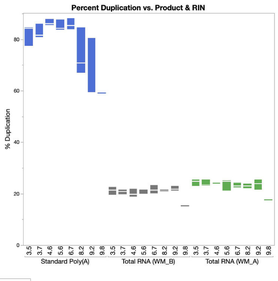


Figure 4. *Duplication rates of samples with RINs from 3.5 - 9.8 processed through a standard poly(A) method, and Watchmaker's Total RNA methods A&B.*

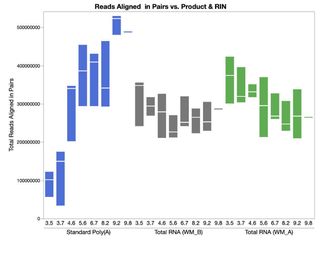


Figure 5. *Distribution of reads from a NovaSeqX run where the samples from each method were pooled together for sequencing.*

Degradation & Gene Coverage

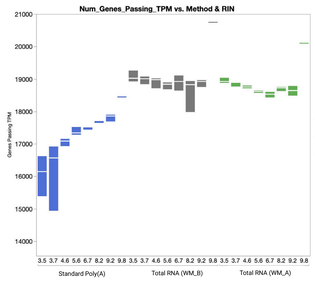
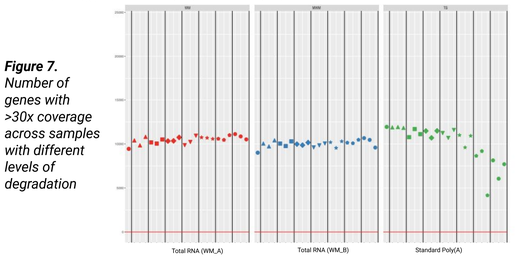


Figure 6. *Comparison of genes passing TPM >0.1 between samples processed through a standard poly(A) method, and Watchmaker's Total RNA methods A&B.*

Figure 7. *Number of genes with >30x coverage across samples with different levels of degradation*



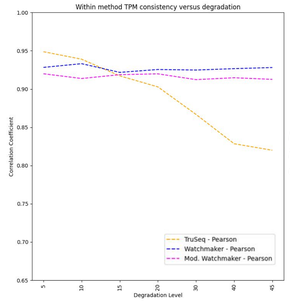


Figure 8. *Correlation of genes identified in the same sample at different levels of degradation (RIN 3.5 - 9.8)*

Evaluation of gene coverage, TPM and correlation of genes identified in the same sample across different levels of degradation show the consistency of sample performance and gene coverage across RIN scores when utilizing either of the Watchmaker total RNA methods.

260 / 230 & Library Performance

High salt contamination inhibits cDNA synthesis and drastically reduces library yields to below passing yields in standard poly(A) methods. Both Watchmaker protocols are able to handle some levels of salt contamination, however Protocol B for lower quality samples produces ~6x yield (Figure 9). which enables low 260/230 samples to meet a minimum number of genes with TPM > 0.1 (Figure 10).

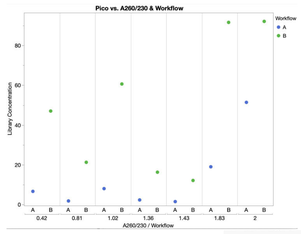
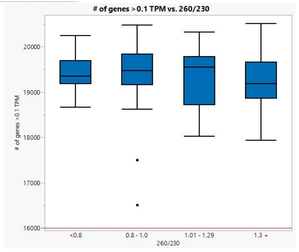


Figure 9. *Comparison of library performance between samples processed through Watchmaker's Total RNA methods A&B. With all ranges of salt contamination performing better with protocol B.*

Figure 10. *Number of genes >0.1 TPM by 260/230 measure of RNA before Watchmaker Total RNA method.*



Next Steps

- While evaluating the impact of incoming quality on library and sequencing performance it was noted that high levels of salt contamination changed the distribution of reads between intronic and exonic regions (Figure 11).
- We are working on methods to reduce levels of salt contamination for RNA samples to have a more even distribution of reads across regions to better identify SNVs consistently across samples within a study.

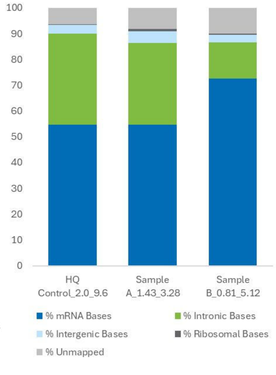


Figure 11. *Breakdown of read distribution for samples with low & high 260/230 values.*

Acknowledgments

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

WATCHMAKER GENOMICS