Maximizing RNA sequencing economy from blood samples using a combined mRNA capture and globin depletion workflow

Ann Meyers, Jen Pavlica, Kailee Reed, Thomas Harris, and Ariele Hanek — *Watchmaker Genomics, Boulder, CO USA*

INTRODUCTION

Blood contains a high abundance of polyadenylated globin mRNAs (cited as up to 80% in some samples),[1] which are naturally enriched during mRNA capture. For researchers focused on gene expression analysis, isoform detection, splice variant and gene fusion identification, single nucleotide variant analysis, or novel transcript discovery, this excess of globin mRNA can be undesirable, as it consumes valuable sequencing resources. Removing globin mRNA from libraries is often utilized to improve sequencing efficiency.

The Watchmaker Polaris™ Depletion Kit – rRNA/Globin (HMR) addresses this need by depleting both rRNA and globin mRNA while preserving non-coding information. However, in some cases, users may not require non-coding RNA and aim to maximize sequencing economy. Combining mRNA capture with depletion provides an optimal solution, achieving efficient mRNA enrichment and globin mRNA removal, reducing the need for deep sequencing and enhancing sequencing efficiency by minimizing non-essential content.

HIGHLIGHTS

- Combining the Watchmaker mRNA Library Prep Kit and the Polaris™ Depletion Kit enables the enrichment of mRNAs and subsequent depletion of globin mRNAs in blood samples to maximize sequencing economy.
- This approach focuses sequencing on exonic transcripts and reduces duplication rates, increases gene detection, and delivers even transcript coverage relative to mRNA capture alone.
- The overall workflow takes approximately 6.25 hours to complete — enabling single day library prep.

OPTIMIZATION AND METHODS

Optimization

To combine the two workflows, purified human whole blood RNA was used as input into the Watchmaker mRNA Library Prep Kit protocol. Following the first and second rounds of poly(A) capture and the final mRNA bead wash, the mRNA was eluted from the capture beads and subsequently underwent rRNA and globin depletion then RNA library preparation (Figure 1).

mRNA capture physically enriches 3' polyadenylated transcripts by hybridizing them with poly(T) primers attached to paramagnetic beads, meaning breaks along the transcripts can result in the loss of 5' information.

Since Polaris Depletion includes a heating step, which can cause RNA degradation and transcript breaks, performing mRNA capture first helps mitigate this issue.

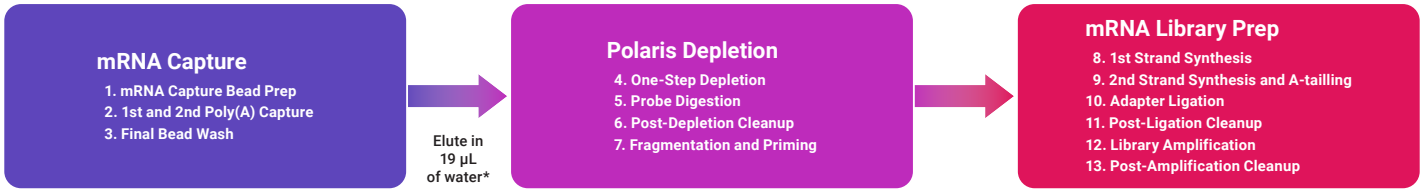
This combined workflow was optimized through systematic testing of adapter concentrations and PCR cycles to balance yield and minimize adapter-dimer formation. Three different adapter concentrations and two PCR cycle numbers were evaluated, and the combination that provided sufficient library yields with minimal adapter-dimer was selected for sequencing.

Experimental Methods

Library Preparation: Libraries were prepared in duplicate from human whole blood RNA across varied input amounts — 10, 50, 100, and 250 ng — using the optimized protocol for combined mRNA capture and globin depletion. Table 1 outlines the adapter concentrations, post-ligation clean-up steps, and PCR cycles used per input amount. A summary of the modified workflow can be found in Table A1 in the Appendix. For control comparisons, mRNA-only libraries were prepared from 100 ng of human whole blood RNA according to standard recommendations (Table 1). IDT xGen Stubby Adapter-UDI Primers (8 nt) were used for all libraries.

Library QC and Sequencing: Library quality and yield were assessed by TapeStation (D1000). Sequencing was conducted on the Illumina NovaSeq 6000 S1 with paired-end 150 nt reads. Libraries were randomly subsampled to 10 million read pairs.

Bioinformatics: Subsampling was performed with Seqtk, followed by quality control using FastQC and adapter trimming with Cutadapt. Reads were aligned with the STAR aligner, and the duplicates were marked using GATK's Picard tool. Gene expression was measured through gene/transcript abundance estimation with featureCounts and a cutoff of 5 unique reads. Various metrics concerning RNA quality, coverage biases, GC content, read distribution, and potential contamination were subsequently assessed using tools like Picard, and RSeQC, among others.



**Denotes a deviation from standard protocol. See additional details in Table A1.*

Figure 1. Combined mRNA capture/globin depletion workflow diagram. Illustration of the workflow combining mRNA capture, rRNA and globin depletion, and mRNA library prep to minimize globin mRNA in libraries generated from blood samples. Total workflow time is estimated to be approximately 6.25 hours, accommodating single day library prep.

Table 1. Details of library preparation by workflow and input mass

RNA enrichment method	Workflow variable	Details			
Combined mRNA capture and globin depletion	Input	10 ng	50 ng	100 ng	250 ng
	Adapter concentration	1 µM	1 µM	1 µM	4 µM
	Post-ligation cleanups	2	2	2	1
	PCR cycles	17	15	13	12
mRNA-only	Input	–	–	100 ng	–
	Adapter concentration	–	–	1 µM	–
	Post-ligation cleanups	–	–	1	–
	PCR cycles	–	–	12	–

RESULTS

Library QC and high-level sequencing metrics

All libraries generated were of sufficient concentration and size for sequencing (Table 2). Fragment analysis (Figure 2) shows little to no adapter-dimer present. The combined mRNA capture and globin depletion protocol yielded libraries of excellent quality and sequencing accuracy across input ranges. The percentage of aligned, passing filter reads were between 88.5% and 96.8%. Correct strand reads were consistently high, reaching 99.2%.

Base composition

As expected and previously described, globin mRNAs are abundant in mRNA capture-only libraries (approx. 14%). Combining the two workflows reduced globin mRNA and rRNA bases to negligible levels (0.16% and 0.014%, respectively), significantly increasing mRNA content in the library to 70%. This ensured a greater proportion of coding sequences, minimizing wasted sequencing on transcripts of disinterest. Intergenic base levels remained consistent across both workflows.

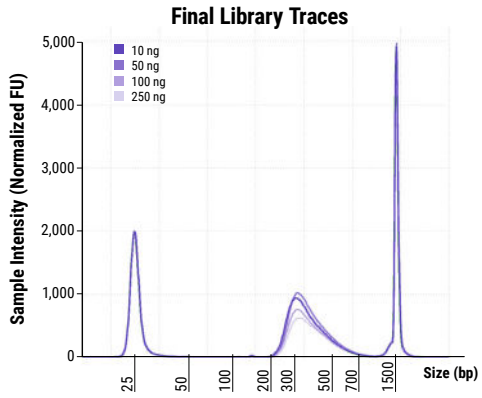


Figure 2. Final library size distributions. TapeStation (D1000) traces of libraries generated using the combined mRNA capture/globin depletion workflow. One library trace shown per condition.

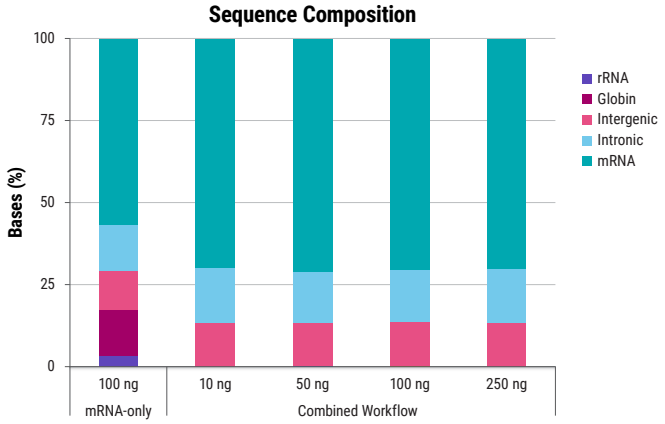


Figure 3. Distribution of sequencing bases for various workflows assessed. The combined workflow delivers the highest percentage of mRNA bases as a result of reduced globin bases.

Table 2: Averaged library and sequencing QC data for the combined workflow

RNA enrichment method	Metric	Details			
Combined mRNA capture and globin depletion	Input	10 ng	50 ng	100 ng	250 ng
	Library yield	34.1 nM	37.9 nM	46.9 nM	36.9 nM
	Mean library size	427 bp	422 bp	374 bp	370 bp
	Mean insert size	201 bp	203 bp	201 bp	218 bp
	Passing filter reads aligned	88.5%	92.4%	92.3%	96.8%
	Correct strand reads	99.2%	99.2%	99.2%	99.2%

Library complexity and gene detection

Library complexity is a critical factor in RNA-seq, representing the variety of unique RNA molecules captured. High complexity supports comprehensive gene expression analysis, minimizes duplication rates, and provides a more accurate snapshot of the transcriptome. Metrics like the number of unique genes detected and the percentage of duplicate reads offer valuable insight into library complexity, especially when sequencing read numbers are consistent across conditions.

Figures 4A and 4B highlight duplication and unique gene detection as a function of workflow and RNA input mass. With the combined workflow, percent duplication naturally decreases with increasing RNA input. This occurs because higher input provides more unique RNA molecules for conversion into cDNA, creating a more diverse and representative library. Conversely, when RNA input is low, there's a higher chance of amplifying the same molecules repeatedly, leading to duplicate sequences and reduced library complexity due to fewer unique fragments. The input-dependent duplicate read improvements translate to modest increases in the number of unique genes detected.

The combined workflow results in fewer duplicate reads in comparison to mRNA capture-only when RNA mass is kept constant at 100 ng. The improvement is likely due to the reduction of globin mRNA transcripts in the combined workflow, as these highly abundant transcripts can inflate duplication rates.

As expected, there is a high degree of overlap between the genes detected in the mRNA-only control and the combined workflow when using 100 ng of RNA (Figure 5).

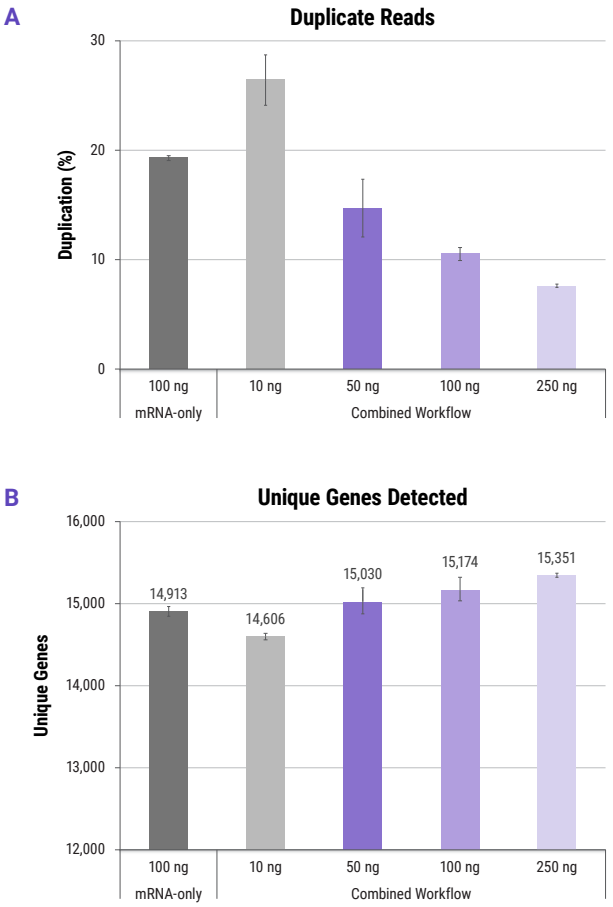


Figure 4. Library complexity as a function of workflow and RNA mass.(A) Average percentage of duplicate reads, and **(B)** average number of unique genes detected with varying RNA mass and RNA enrichment workflows. Increasing RNA input mass leads to a higher number of unique genes detected and fewer duplicate reads, reflecting the anticipated improvement in library complexity. These metrics also improve with the combined workflow, thanks to the reduction of globin mRNAs.

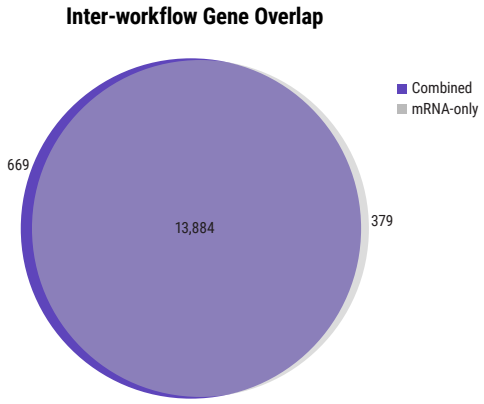


Figure 5. Gene overlap analysis. A majority of the genes identified in the 100 ng combined workflow or mRNA-only libraries were common genes to both workflows. Only genes identified in both technical replicates were included in the overlap analysis.

Transcript coverage uniformity

Even 5' and 3' coverage in RNA sequencing is important because it allows for accurate profiling of gene expression and regulation by capturing information from both ends of a transcript, which can reveal crucial details like alternative splicing events, transcription start sites (TSS),[2] and potential regulatory elements located at the 5' and 3' untranslated regions (UTRs),[3] providing a more complete picture of gene function and expression patterns.

Use of the combined workflow resulted in even transcript coverage across the full range of RNA inputs assessed (Figure 6).

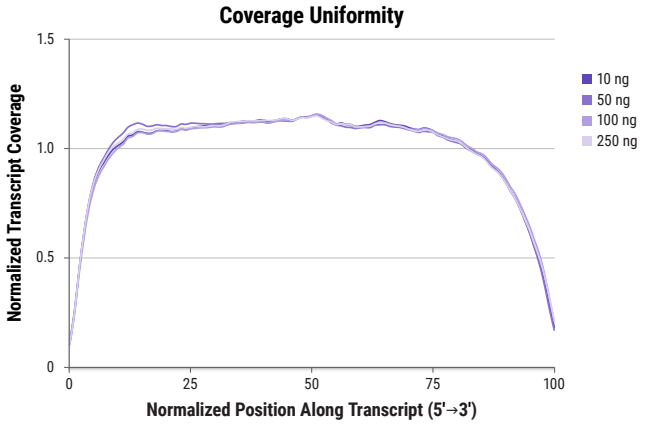


Figure 6. Combined workflow transcript coverage uniformity. Results show exceptional transcript coverage uniformity across the full range of RNA mass assessed when using the combined workflow.

SUMMARY

The combined workflow of Watchmaker Genomics' mRNA Library Prep Kit with Polaris Depletion achieves efficient enrichment of mRNA while effectively removing globin and rRNA, delivering high-quality libraries optimized for gene expression analysis. This approach significantly increases mRNA content, enhances sequencing economy, and minimizes non-essential transcripts, making it ideal for applications requiring focused coverage of coding sequences from blood or blood-derived samples.

Results demonstrate high library complexity, improved gene detection, and uniform transcript coverage across a range of RNA inputs. By reducing duplication rates and maintaining even 5' and 3' coverage, this workflow ensures accurate and comprehensive RNA-seq data, maximizing sequencing resources.

REFERENCES

1. Jang, J.S., Berg, B., Holicky, E. et al. Comparative evaluation for the globin gene depletion methods for mRNA sequencing using the whole blood-derived total RNAs. BMC Genomics 21, 890 (2020). https://doi.org/10.1186/s12864-020-07304-4

2. Zhiyi Qin, Peter Stoilov, Xuegong Zhang, Yi Xing, SEASTAR: systematic evaluation of alternative transcription start sites in RNA, Nucleic Acids Research, Volume 46, Issue 8, 4 May 2018, Page e45, https://doi.org/10.1093/nar/gky053

3. González-Sánchez AM, Castellanos-Silva EA, Díaz-Figueroa G, Cate JHD (2024) JUN mRNA translation regulation is mediated by multiple 5' UTR and start codon features. PLOS ONE 19(3): e0299779. https://doi.org/10.1371/journal.pone.0299779

APPENDIX A: MODIFICATIONS TO STANDARD PROTOCOLS

Refer to Watchmaker mRNA Library Prep Kit User Guide (v2.1.1024 or later) and Watchmaker RNA Library Prep Kit with Polaris User Guide (v4.1.1024 or later) for the library construction protocol, as outlined in Figure A1 and employing the specific modifications detailed in Table A1.

Table A1. Summary of Protocol Modifications

Step Name	Protocol Adjustment	Step Number
Input RNA	2.5 ng – 500 ng	N/A
Follow the Watchmaker mRNA Library Prep Kit User Guide (v2.1 or later)		
RNA and mRNA Capture Bead Preparation	N/A	1.1 – 1.12
1st Poly(A) RNA Capture	N/A	2.1 – 2.11
2nd Poly(A) RNA Capture	Follow the 2nd poly(A) capture protocol up until and including Step 3.6.	3.1 – 3.6
Elution of mRNA from mRNA Capture Beads	Remove the tubes from the magnet and gently **resuspend each bead pellet in 19 µL RNAse-free water** using a pipette. Put the mRNA-bead mixture back on the magnet and wait for at least 2 min, or until all beads have been collected on the tube wall and the solution is clear. Carefully remove 18 µL of supernatant to a fresh tube and store on ice.	Additional new step
Follow the Watchmaker RNA Library Prep Kit with Polaris™ Depletion User Guide: Library Construction Protocol A (v4.1 or later)		
Globin mRNA Depletion	Proceed with rRNA and Globin Depletion Steps A1.1 to A1.9, using the 18 µL of supernatant prepared above as the RNA input sample.	A1.1 – A1.9
Probe Digestion	N/A	A2.1 – A2.5
Post-depletion Cleanup	Complete to the end of Step A3.10 ensuring that the SPRI beads are sufficiently dry.	A3.1 – A3.10
Fragmentation and Priming	Add the Frag & Prime Master Mix to the SPRI beads from Step A3.10 of the RNA Library Prep Kit with Polaris Depletion workflow.	A4.1 – A4.8
Follow the Watchmaker mRNA Library Prep Kit User Guide (v2.1 or later)		
1st Strand Synthesis	N/A	5.1 – 5.7
2nd Strand Synthesis and A-tailing	N/A	6.1 – 6.7
Adapter Ligation	N/A	7.1 – 7.10
Post-ligation Cleanup	Proceed with post-ligation cleanup.	8.1 – 8.14
2nd Post-ligation Cleanup (Optional)	A 2nd post-ligation cleanup is recommended when working with lower than 100 ng of input RNA.	9.1 – 9.14
Library Amplification and Strand Selection	If using truncated adapters, proceed with library amplification using the following PCR cycle number recommendations:	10.1 – 10.6

RNA input into Library Preparation (ng)	No. of PCR cycles to generate 10 – 50 nM library
>250	9 – 11
101 – 250	11 – 12
51 – 100	13 – 14
11 – 50	14 – 16
≤10	16 – 18

	If using full-length adapters, use the number of PCR cycles recommended in the Watchmaker mRNA Library Prep Kit User Guide for the appropriate RNA input.	10.2, Table 3
Post-amplification Cleanup	N/A	11.1 – 11.15

WATCHMAKER GENOMICS

For Technical Support, please contact
support@watchmakergenomics.com.

5744 Central Avenue, Suite 100
Boulder, CO 80301
www.watchmakergenomics.com

WMTN002 v1.0.0325