Improved Detection of Unique Transcripts from FFPE-Derived RNA Using a Novel, Rapid Whole Transcriptome Sequencing Workflow

Travis Sanders, Lee French, Julie Walker, Jennifer Pavlica, Clara Ross, Thomas Harrison, Ross Wadsworth
Watchmaker Genomics – Boulder, CO USA

Introduction

FFPE samples are an invaluable resource for oncology researchers, providing access to a vast library of diseased tissue samples paired with relevant donor information. Despite their broad utility, FFPE-derived RNA samples often vary wildly in performance with fixation process, block age and storage, and extraction methodology having large impacts on resulting template quality. As a result, robust and reproducible RNA sequencing with FFPE-derived RNA remains a challenge. To address this need, we developed a novel, streamlined whole-transcriptome library preparation workflow specifically tailored for FFPE and low-input RNA performance.

To simplify the protocol while minimizing off-target effects, we built algorithms for optimal probe design and streamlined probe hybridization. We specifically engineered enzymes and reformulated buffers in parallel to improve yields. A novel de-crosslinking step was integrated to improve FFPE performance. Combining reactions, reducing incubation times, and eliminating purifications streamlined the overall workflow.

Experimental Approach

General. We compared our solution to two commercial products (NEBNext® Ultra II Directional RNA Library Prep Kit with Globin & rRNA Depletion and Illumina® Stranded Total RNA Prep with Ribo-Zero™ Plus). RNA was extracted from whole blood and four FFPE blocks, and libraries were prepared in triplicate with inputs ranging from 1 ng to 500 ng (blood) and in duplicate with 100 ng inputs (FFPE). Libraries were sequenced on a NovaSeq™ 6000 S2 flow cell with 2 x 75 bp read lengths. Depleted data sets were randomly subsampled to 24M paired reads, while undepleted total RNA libraries were subsampled to 300M paired reads.

Off-target analysis. Sites with a high potential for off-target depletion were identified by blasting the Watchmaker Genomics probe sequences to hg38 and extracting all hits above an e-value of 5. Each base's depth of coverage in these sites was assessed and normalized to the total number of non-rRNA mapped bases. The median normalized coverage was calculated for each candidate site. Fold_undercovered is the median value of the median normalized coverage of a depleted sample over an undepleted sample. For example, a value of 0.5 implies that the depleted sample has half as much normalized coverage of these sites compared to an undepleted sample. This metric was then normalized to Watchmaker to generate estimates of percent increase or decrease in off-targeting effects.

Benchmarking Workflow Improvements

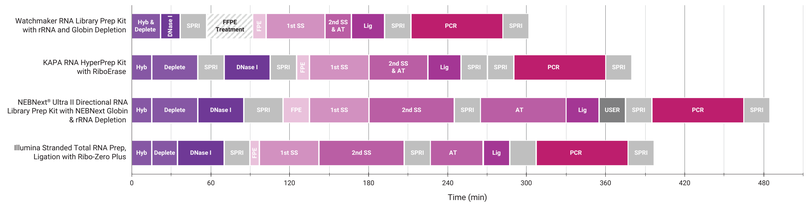

Figure 1. Comparison of total workflow time. The Watchmaker Genomics solution combines and shortens enzymatic steps and has fewer bead purifications in comparison to commercially available kits, resulting in a highly automatable workflow with significantly reduced hands-on time (up to one hour per plate) and consumable requirements (approximately 1,000 tips per 96 libraries).

Low-Bias, High-Efficiency Depletion

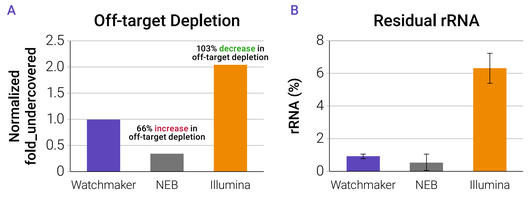

Figure 2. Optimal depletion is a balance of **(A)** minimizing off-target effects and **(B)** removing rRNA reads. NEB has increased off-target depletion with low residual rRNA, and Illumina has very low off-target depletion but with much higher residual rRNA. Watchmaker delivers a balance of minimized off-target depletion with maximized depletion efficiency.

Robust Low-Input Performance: Whole Blood

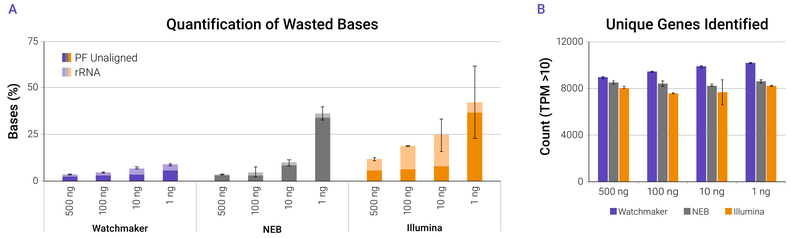

Figure 3. Analysis of **(A)** the percentage of bases wasted due to either failure to align to the reference or aligning to rRNA regions, and **(B)** unique genes identified using featureCounts. Raw counts were converted to TPM and cutoff filters of 10 were applied. Results show the robust performance of the Watchmaker solution with respect to low input amounts, with a minimal increase in wasted bases and no drop in genes detected.

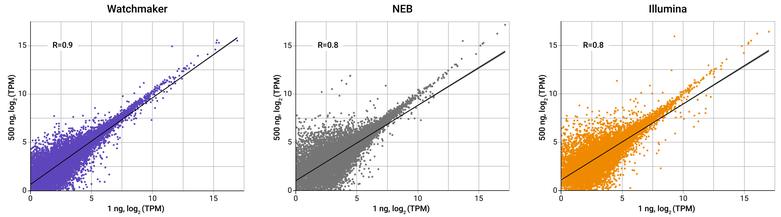

Figure 4. Inter-input reproducibility assessment between 500 ng and 1 ng samples with respect to gene counts using a TPM cutoff of 1. Genes for which there was a zero-value in either sample were omitted from gene count correlation analysis. Trend lines are the best fit lines by linear regression, and their corresponding R-values are shown. Results show increased correlation when using the Watchmaker solution in comparison to other workflows, further demonstrating the robust performance with low input amounts.

Reproducible FFPE Performance

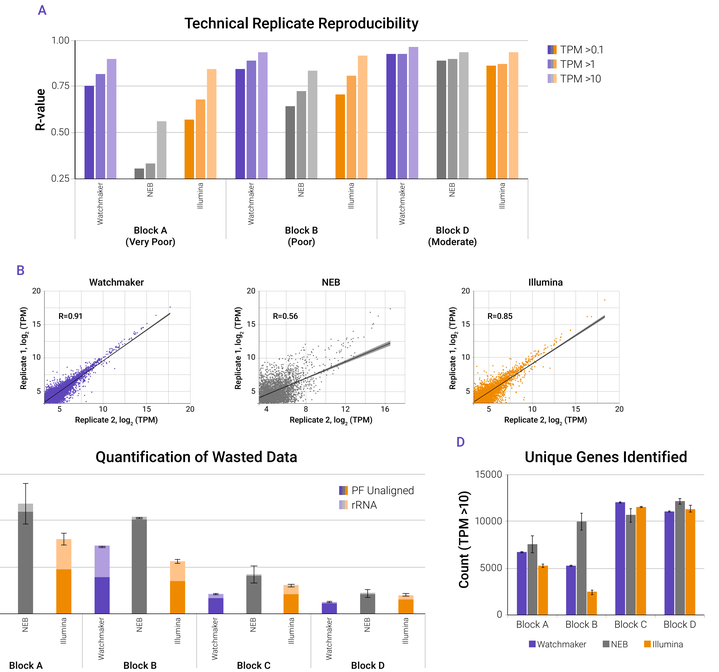

Figure 5. (A) Watchmaker FFPE technical replicate libraries show improved reproducibility, especially with very-poor quality samples, as evidenced by the high and stable correlation coefficient values with TPM cutoffs ranging from 0.1 to 10. This is further illustrated with **(B)** technical replicate gene count correlation plots for Block A (very poor quality) with a TPM cutoff of 10. **(C)** The Watchmaker solution shows reduced percentages of passed filter bases lost due to either failure to align to the reference or aligning to rRNA regions. **(D)** The number of unique genes identified using a TPM cutoff of 10. The absolute number of unique genes identified should be assessed in conjunction with the reproducibility data shown in **(A)** and **(B)**. For example, the NEB solution detects more genes with Block A but has substantially more noise in the data. Though Watchmaker detects slightly fewer genes, the technical replicate data indicate that these are true genes with accurate count measurements.

Conclusions

- Watchmaker Genomics' vertically integrated approach to development—which layers specifically engineered enzymes into co-optimized NGS workflows—has delivered a solution with improved performance with inputs as low as 1 ng and more reproducible performance with FFPE material.
- This novel, simplified workflow enables library construction within five hours, improves automatability, and reduces hands-on time and consumable use by 25% in comparison to commercially available products.