Improved detection of unique transcripts using a novel, rapid whole transcriptome sequencing workflow

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

Introduction

mRNA-seq is a powerful tool for transcriptome profiling but is not applicable to many clinically relevant sample types. Template damage in FFPE samples generates 3'-bias, while overabundant globin mRNA in blood-derived samples are often uninformative. Whole transcriptome sequencing, where overabundant transcripts are depleted, supports these sample types and offers a comprehensive view of the transcriptome—including biologically relevant non-coding transcripts. However, these workflows are typically long, labor-intensive, and difficult to automate. To address this need, developed the Watchmaker RNA Library Prep Kit with Polaris Depletion—a highly streamlined and automatable solution tailored for challenging samples.

Experimental Approach

RNA-seq library prep comparison. We compared the Watchmaker RNA Library Prep Kit with Polaris Depletion 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). KAPA RNA HyperPrep Kit with RiboErase (HMR) Globin was excluded from experimental analysis due to lack of kit availability. RNA was extracted from whole blood and five 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. Data sets were randomly subsampled to 24M paired reads.

Table 1. FFPE RNA DV200 values

Block ID	DV200
1	47%
2	32%
3	36%
4	55%
5	50%

Gene identification analysis (Figure 2). Unique genes were identified using featureCounts and deduplicated raw reads. Figures 2B and 2D count those unique genes supported by 20 or more deduplicated raw reads.

Overlap analyses (Figures 3 and 4). Overlap analyses only included genes supported by 20 or more deduplicated raw reads and those genes identified in all technical replicates.

Hierarchical clustering (Figures 3 and 4). Hierarchical clustering was performed after normalizing gene counts to the total number of unique, aligned reads for each sample. Agglomerative clustering was performed according to unweighted average linkage method.

Inter-workflow per-gene comparison (Figure 5). Gene counts from technical replicates were averaged and log2 transformed. Analysis only included genes with greater than 5 supporting deduplicated reads, genes that were identified in both technical replicates, and genes that were identified by both workflows.

Benchmarking Workflow Improvements

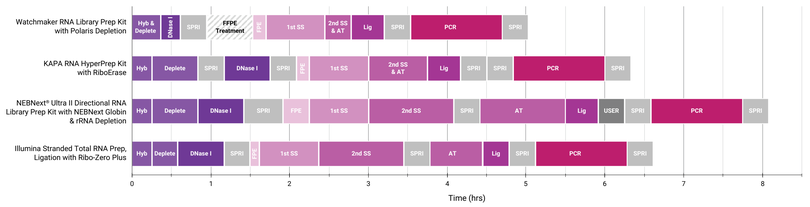

Figure 1. Reduced total turnaround time. The Watchmaker Genomics solution combines and shortens enzymatic steps and has fewer bead purifications in comparison to commercially-available kits, resulting in a highly automatable workflow with significantly reduced hands-on time (up to one hour per plate)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

Robust Gene Detection

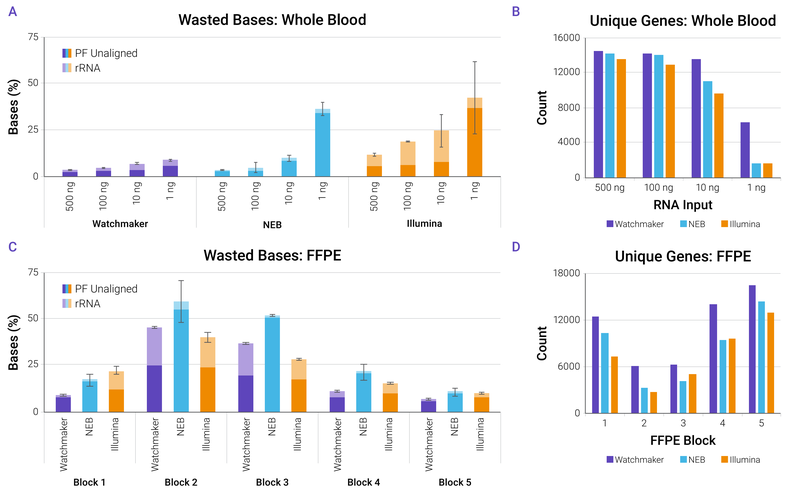

Figure 2. Improved sequencing economy and gene detection. Analysis of **(A)** the percentage of bases wasted due to either failure to align to the reference or aligning to rRNA regions, and **(B)** unique genes identified from whole blood samples. Similar data for FFPE-derived samples are shown in **(C)** and **(D)**.

Increased Sensitivity with Low Inputs

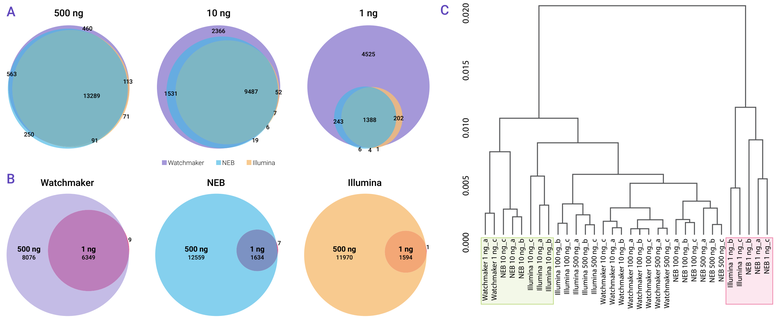

Figure 3. Increased gene detection sensitivity with low input amounts. (A) Inter-workflow overlap analysis of genes identified, stratified by whole blood RNA input amount. **(B)** Inter-input overlap analysis of genes identified, stratified by workflow. **(C)** Hierarchical clustering analysis shows that 1 ng Watchmaker samples cluster with higher input NEB samples, indicating a significant improvement in sensitivity. Conversely, NEB and Illumina 1 ng samples do not cluster with higher inputs. Overall, results indicate that the Watchmaker solution detects a majority of genes identified by NEB and Illumina, with a significant number of additional genes identified as the input amount decreases. Inter-input overlap analysis and clustering data provide confidence that these additional genes are true genes.

Increased Sensitivity with FFPE Samples

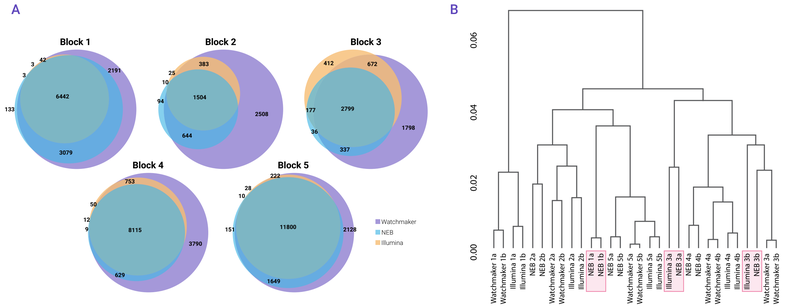

Figure 4. Increased gene detection sensitivity with FFPE. (A) Inter-workflow overlap analysis of genes identified, stratified by FFPE block. **(B)** Hierarchical clustering analysis of all FFPE libraries sequenced. Results indicate that for most FFPE blocks, the Watchmaker solution detects a majority of genes identified by NEB and Illumina, with a significant number of additional genes identified. Clustering data provide confidence that these additional genes are true genes. Watchmaker libraries minimally cluster by FFPE block with Illumina libraries, whereas NEB libraries are less similar. Watchmaker technical replicates always cluster together, whereas Illumina and NEB technical replicates do not.

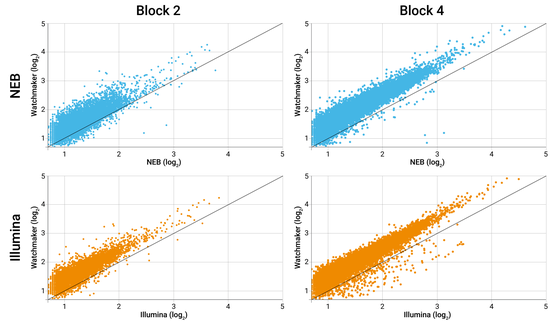

Figure 5. Increased confidence in gene identification. Per-gene comparison between Watchmaker and NEB (top) and Illumina (bottom) for FFPE blocks 2 and 4. An x=y line is included for visual reference. Results show that the Watchmaker solution generally has more supporting unique reads per gene than both NEB and Illumina.

Conclusions

Watchmaker Genomics' vertically integrated approach to development, which layers specifically engineered enzymes into co-optimized NGS workflows, has delivered a RNA-seq solution with:

- Increased gene detection sensitivity with FFPE-derived RNA
- Robust performance with input amounts as low as 1 ng
- A novel, simplified workflow that enables library construction in under five hours and reduces hands-on and consumable use by 25%