Optimized RNA Seq library preparation for fusion calling from FFPE samples

Giulia Corbet, Josh Haimes, Travis Sanders, Martin Ranik, Thomas Harrison, Eduard Casas, Kailee Reed, Doug Wendel, Kristina Giorda, Ariele Hanek, Jen Pavlica, Brian Kudlow – Watchmaker Genomics

Kurtis Davies – University of Colorado Anschutz Medical Campus

Abstract #321

Introduction

Gene fusions resulting from chromosomal rearrangement, duplication, or deletion can be important drivers of cancer. As such, identification of gene fusions can play a valuable role in selecting targeted therapies. One of the most widely used methods for detecting fusion transcripts is targeted RNA sequencing. This approach utilizes a combination of target enrichment which increases sequencing efficiency by focusing reads on regions of interest coupled with analysis for known fusions within those regions. While effective at detecting known fusions, this methodology limits discovery of novel fusions.

Herein, we set out to determine if whole transcriptome analysis (WTA) — sequencing of the non-ribosomal transcriptome — can be utilized for fusion detection. First, we established detection of 16 fusions using both target enrichment and WTA with Seraseq FFPE control samples. Building off this initial success, we optimized the whole transcriptome library prep to allow for longer insert sizes without reducing library complexity to increase the likelihood of de novo fusion detection. Next, we utilize the optimized whole transcriptome library preparation method with true FFPE samples previously demonstrated to be fusion positive. Finally, we extend the analysis with a de novo fusion calling analysis pipeline. Results indicate that WTA can be used to detect fusions even with challenging samples.

Detect Known Fusions with Whole Transcriptome Analysis or Target Enrichment

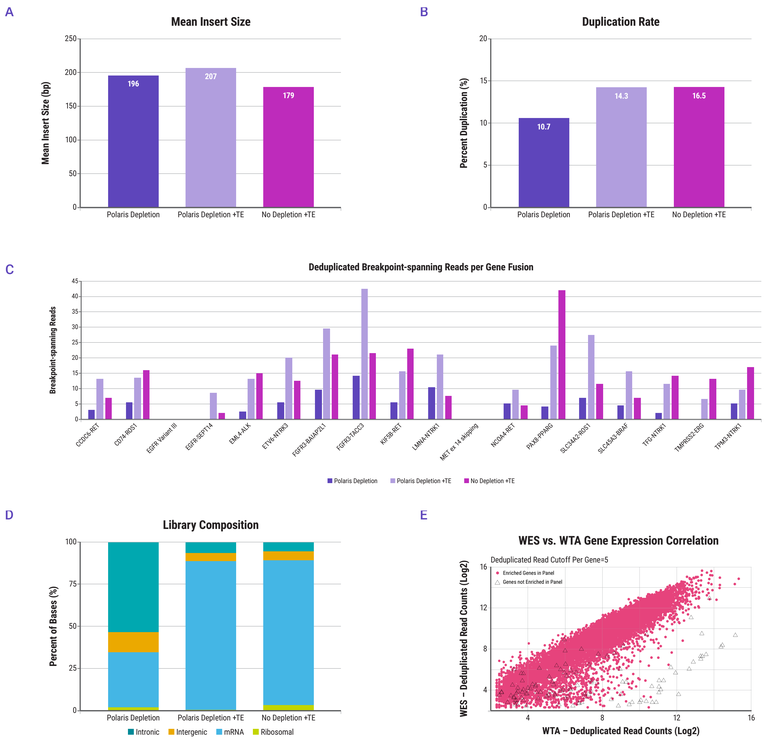

Figure 1. Target enrichment (TE) increases support for fusion calls. Libraries were prepared in duplicate from 100 ng Seraseq FFPE RNA using the Watchmaker RNA Library Prep Kit with (1) Polaris Depletion upstream, (2) Polaris Depletion upstream and TE downstream, or (3) just TE downstream. Data are labelled as "Polaris Depletion", "Polaris Depletion + TE", or "No Depletion + TE" respectively. Data are shown for **(A)** mean insert size, **(B)** duplication rate, **(C)** deduplicated breakpoint-spanning reads per gene fusion, and **(D)** library composition for all 3 workflows. Inclusion of Target Enrichment focuses sequencing on mRNAs and increases the number of breakpoint spanning reads. Incorporation of Polaris Depletion upstream of target enrichment decreases ribosomal RNA reads. **(E)** Expression correlation analysis between ±TE libraries shows that target enrichment alters relative gene expression.

Longer Inserts Improve Fusion Calling

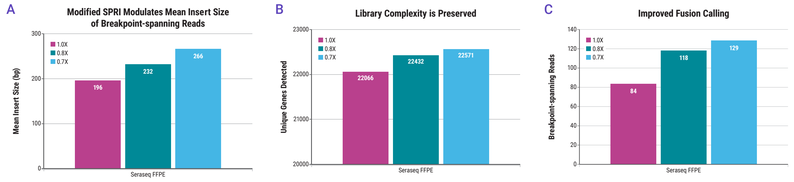

Figure 2. Modified SPRI achieves longer inserts and improves fusion calling. (A) Mean insert size of breakpoint-spanning reads increases with decreasing post amplification cleanup ratio. Magnitude of increase agrees with insert size across the whole library. **(B)** Library Complexity, measured by unique genes detected, is preserved with decreasing SPRI ratios. **(C)** Number of deduplicated breakpoint-spanning reads increases with decreasing post amplification SPRI cleanup ratio. Duplicate libraries were generated with 100 ng of Seraseq FFPE Fusion RNA.

Quality Whole Transcriptome Libraries from FFPE

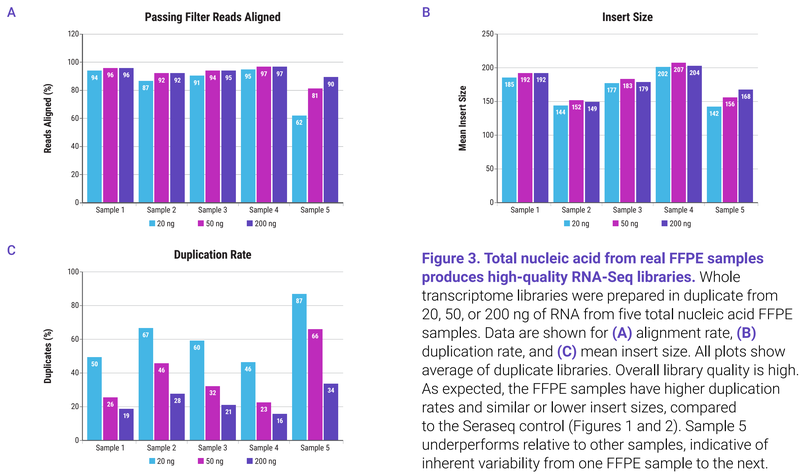

Figure 3. Total nucleic acid from real FFPE samples produces high-quality RNA-Seq libraries. Whole transcriptome libraries were prepared in duplicate from 20, 50, or 200 ng of RNA from five total nucleic acid FFPE samples. Data are shown for **(A)** alignment rate, **(B)** duplication rate, and **(C)** mean insert size. All plots show average of duplicate libraries. Overall library quality is high. As expected, the FFPE samples have higher duplication rates and similar or lower insert sizes, compared to the Seraseq control (Figures 1 and 2). Sample 5 underperforms relative to other samples, indicative of inherent variability from one FFPE sample to the next.

Robust Gene Detection

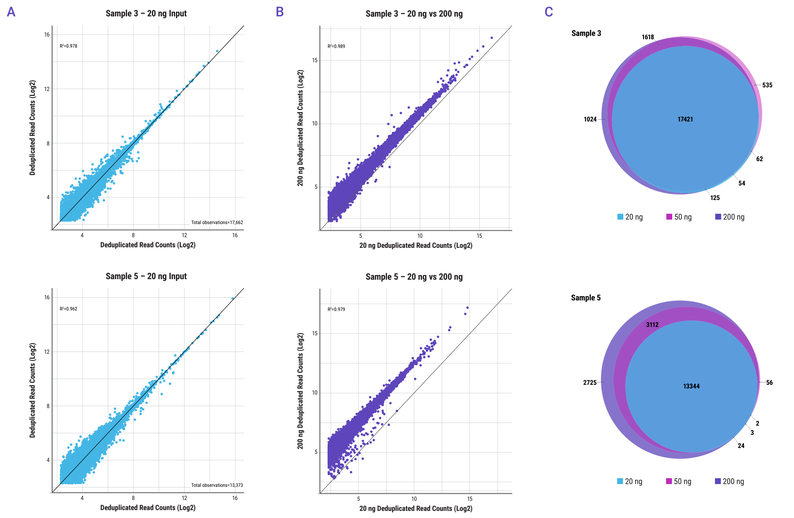

Figure 4. Robust gene detection with FFPE samples. Samples 3 and 5 were chosen as representative of the normal range of FFPE sample quality. Gene expression is highly correlated between **(A)** replicate libraries and **(B)** across inputs. **(C)** Gene detection overlap is high across all inputs.

De Novo Fusion Calling with FFPE and WTA

	Fusion	Input Mass	Detected	Breakpoint Spanning Reads
Sample 1	EML4-ALK	20 ng	yes	3
		50 ng		2
		200 ng		5
Sample 2	MET exon 14 Skipping	20 ng	yes*	14
		50 ng		17
		200 ng		18
Sample 3	EML4-ALK	20 ng	yes	3
		50 ng		3
		200 ng		2
Sample 4	IGFBP5-ALK	20 ng	yes	30
		50 ng		28
		200 ng		34
Sample 5	KIF5B-RET	20 ng	yes	1
		50 ng		3
		200 ng		4

Table 1. Fusion calls for de novo fusion calling analyses with five FFPE samples each at 3 input masses. Counts are averages of duplicate libraries. *MET exon skipping variant was detected and supported by ≥10 deduplicated breakpoint spanning reads per library, when using a targeted fusion analysis method, but was not detected in the de novo fusion analysis, which was not designed to detect intra-gene events.

Materials and Methods

RNA samples: RNA was extracted using the Promega ReliaPrep™ Total RNA Miniprep System from Seraseq FFPE Fusion RNA v4 Reference Materials, which contains 16 clinically relevant gene fusions.

FFPE: Total nucleic acid from fusion-positive FFPE samples was obtained from University of Colorado Anschutz Medical Campus.

Library Prep: Libraries were prepared using the Watchmaker RNA Library Prep Kit with or without Polaris Depletion according to manufacturer's instructions. Post amplification SPRI purifications were modified, where noted, for control samples. Libraries prepared from the FFPE samples utilized a 0.8X post amplification SPRI cleanup. The IDT xGEN stubby adapter system was used.

Target Enrichment: Libraries were enriched using the Twist Exome 2.0 panel and Twist Hybridization Capture Reagents. 200 ng per library was used for capture.

Sequencing: Libraries were sequenced 2 x 150 bp on a NovaSeq 6000 (Seraseq controls) or NextSeq 2000 (FFPE). Data were randomly subsampled to 28.5M read pairs.

Data Analysis: 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 marked using GATK's Picard tool. Gene expression was measured through gene/transcript abundance estimation with featureCounts. Various metrics concerning RNA quality, coverage biases, GC content, read distribution, and potential contamination were subsequently assessed using tools such as Picard, RSeQC, and Kraken2, among others.

De novo fusion calling: Nf-core's rnafusion was used with the following updates: Seqtk sample was added as an upfront process for downsampling reads for rarification and fusion-report was modified to work without the dependency of COSMIC, thus incorporating Mitelman and FusionGDB 2.0.

Conclusions

- **Watchmaker Genomics RNA Library Prep Kit with Polaris Depletion** produces high-quality whole transcriptome libraries from limited inputs and challenging samples enabling confident gene detection and expression analysis.
- Optimization of post amplification SPRI cleanup allows for longer insert sizes without sacrificing library complexity. Longer insert sizes improve fusion detection, especially in de novo analysis pipelines.
- Whole transcriptome sequencing is compatible with de novo fusion calling analysis.
- De novo fusion calling and whole transcriptome approaches enable gene expression analysis and fusion identification from a single dataset.