Cost-effective, high-resolution whole transcriptome sequencing for gene fusion detection applications

Jen Pavlica[1], Philip Benson[1], Gila Yanai[2], Sarah Pollock[2], Josh Haimes[1], Nika Iremadze[2], Arwa Kathiria[2], Tyson Clark[2], Ariel Jaimovich[2], Brian Kudlow[1]

1 – Watchmaker Genomics, Boulder, CO | 2 – Ultima Genomics, Newark, CA

Introduction

As per-base sequencing costs continue to decline, there is less of a need to restrict reads to pre-identified regions of interest through targeted sequencing. This is particularly relevant for RNA, where cost-effective sequencing allows for a hypothesis-free whole transcriptome sequencing (WTS) approach — useful for identifying new features such as gene fusions. We optimized the Watchmaker RNA Library Prep Kit with Polaris Depletion for sequencing on the Ultima Genomics platform, an ultra-high capacity, single-end read sequencing system, and evaluated the workflow for gene fusion detection via WTS.

Experimental Approach

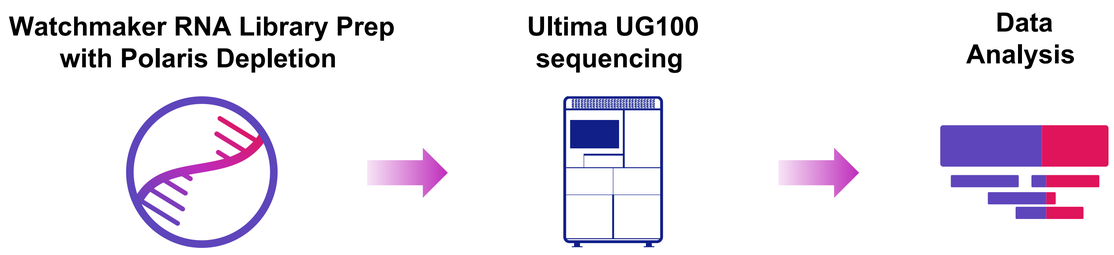

Figure 1. High-level NGS sequencing workflow. Total RNA libraries were prepared with the Watchmaker RNA Library Prep Kit with Polaris Depletion prior to sequencing on Ultima Genomics' UG 100 sequencer and subsequent data analysis for high-level sequencing metrics and fusion calling.

RNA samples: RNA was extracted from Seraseq® FFPE Fusion RNA v4 Reference Material, which contains 16 clinically relevant gene fusions (and two additional isoforms that were not included in our analysis). Universal Human Reference (UHR) RNA was used as a high-quality control. For high-quality RNA, the RIN was 9.8 and DV200 was 94.6%. For FFPE RNA, the RIN was 1.8 and DV200 was 91%.

Library prep: Libraries were prepared from either 50 ng or 250 ng of the Seraseq FFPE material or 100 ng of UHR RNA using the Watchmaker RNA Library Prep Kit with Polaris Depletion and truncated adapters. To maintain strand-specificity, a UDG treatment was integrated into the ligation reaction to make the second strand unamplifiable. Libraries were amplified using UG tsR1 conversion primers and tsR2 universal primer. A variety of post-amplification (PA) cleanup conditions were assessed, as indicated. Control libraries were prepared using a library prep chemistry from Vendor X.

Sequencing: Samples were sequenced using an Ultima Genomics UG100 sequencer, producing single-end reads with a mean average length of 236 bp. The samples were sequenced to a mean depth of ~30M reads for UHR RNA and ~100M reads for Seraseq.

Data analysis: Following adapter trimming, reads were aligned with STAR, using the parameters recommended by STAR-Fusion in order to retain junction information. Duplicate marking was performed using an in-house tool designed for UG single-ended reads. The STAR junction file, along with the deduplicated reads, were input into STAR-Fusion, which was run with "—FusionInspector validate." The minimal FFPM was set to 0.05.

RNAseq metrics were collected using the Picard tool CollectRnaSeqMetrics. FeatureCounts (from the Subread package) was used to assess the number of expressed genes and the duplication fraction.

Workflow Optimization for Ultima Sequencing

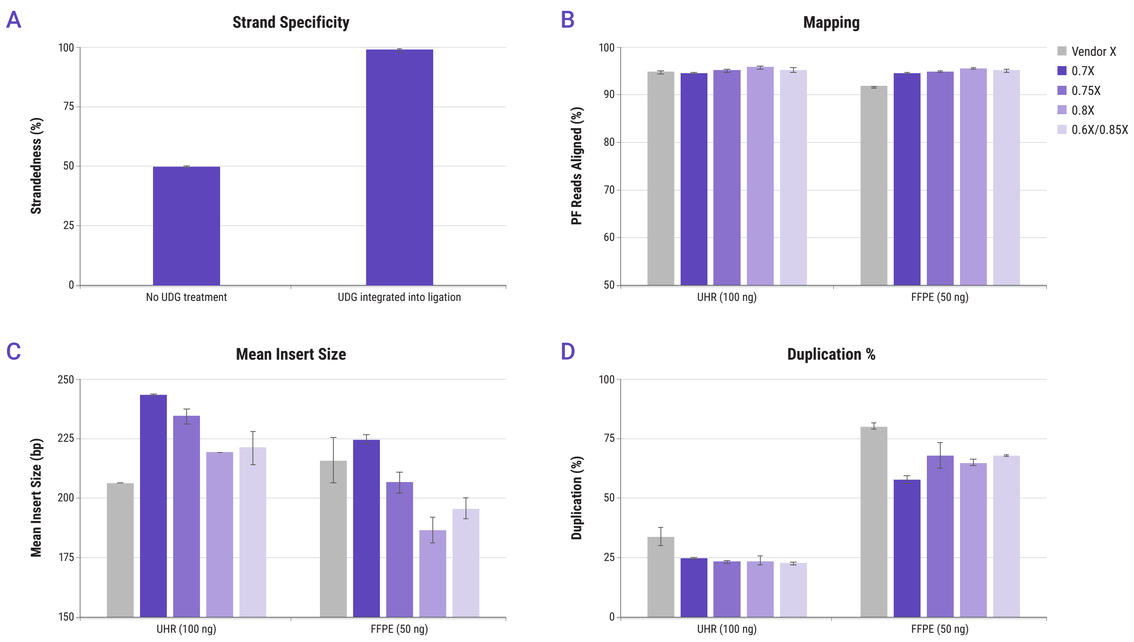

Figure 2. Impact of UDG treatment and post-amplification (PA) cleanup ratio on high-level sequencing metrics. **(A)** To streamline the workflow, the UDG treatment step was incorporated into the ligation reaction. Results show excellent strand specificity relative to a no UDG treatment control. **(B)** PA cleanup ratio had no impact of PF reads aligning to the reference. **(C)** A 0.7X cleanup resulted in the longest libraries and **(D)** lowest duplication rate with FFPE material. Data were randomly subsampled to 15M reads for UHR libraries and 61.5M reads for FFPE libraries.

Workflow Impact on Fusion Calling and Gene Expression

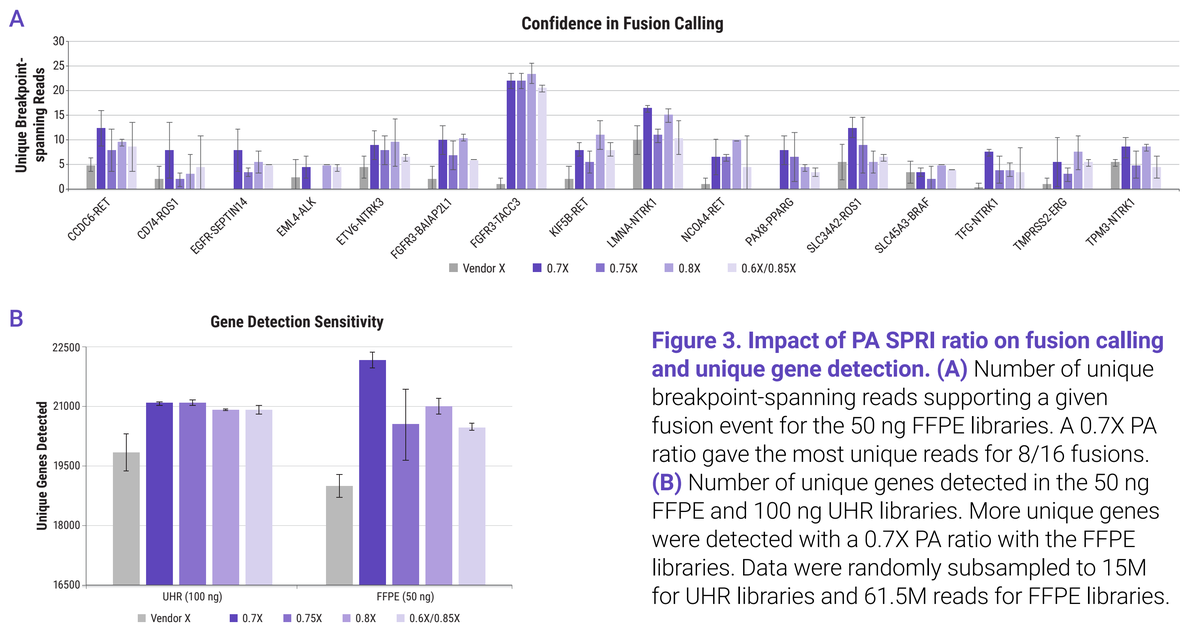

Figure 3. Impact of PA SPRI ratio on fusion calling and unique gene detection. (A) Number of unique breakpoint-spanning reads supporting a given fusion event for the 50 ng FFPE libraries. A 0.7X PA ratio gave the most unique reads for 8/16 fusions. **(B)** Number of unique genes detected in the 50 ng FFPE and 100 ng UHR libraries. More unique genes were detected with a 0.7X PA ratio with the FFPE libraries. Data were randomly subsampled to 15M for UHR libraries and 61.5M reads for FFPE libraries.

Mass and Sequencing Depth Impact on Fusion Calling

	50 ng				250 ng			
	Watchmaker		Vendor X		Watchmaker		Vendor X	
	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2	Rep 1	Rep 2
CCDC6-RET	10	15	6	4	19	6	16	6
CD74-ROS1	4	12	4	0	8	7	4	3
EGFR-SEPTIN14	5	11	0	0	11	8	3	0
EML4-ALK	6	3	0	5	9	9	9	7
ETV6-NTRK3	11	7	3	6	16	14	11	15
FGFR3-BAIAP2L1	8	12	4	0	15	11	10	5
FGFR3-TACC3	21	23	0	2	28	27	12	17
KIF5B-RET	9	7	0	4	23	23	10	15
LMNA-NTRK1	16	17	8	12	16	18	19	24
NCOA4-RET	4	9	0	2	9	12	3	0
PAX8-PPARG	6	10	0	0	7	10	6	6
SLC34A2-ROS1	14	11	3	8	10	14	19	7
SLC45A3-BRAF	3	4	2	5	8	4	21	6
TFG-NTRK1	7	8	0	1	13	10	7	3
TMPRSS2-ERG	9	2	2	0	7	5	22	27
TPM3-NTRK1	10	7	6	5	14	14	21	29
Fusions called	100%	93%	43%	50%	100%	100%	100%	87%

Table 1. Fusions detected with a minimum of 3 breakpoint-spanning reads. For both the Watchmaker and Vendor X libraries, fusion calling was improved when the RNA input mass was increased from 50 ng to 250 ng.

With 50 ng inputs, Watchmaker only missed one fusion in one technical replicate, compared to 17 with Vendor X. No fusions were missed with Watchmaker at 250 ng, while 2 were missed with Vendor X.

Watchmaker libraries were prepared using a 0.7X PA cleanup ratio. Data were randomly subsampled to 61.5M reads.

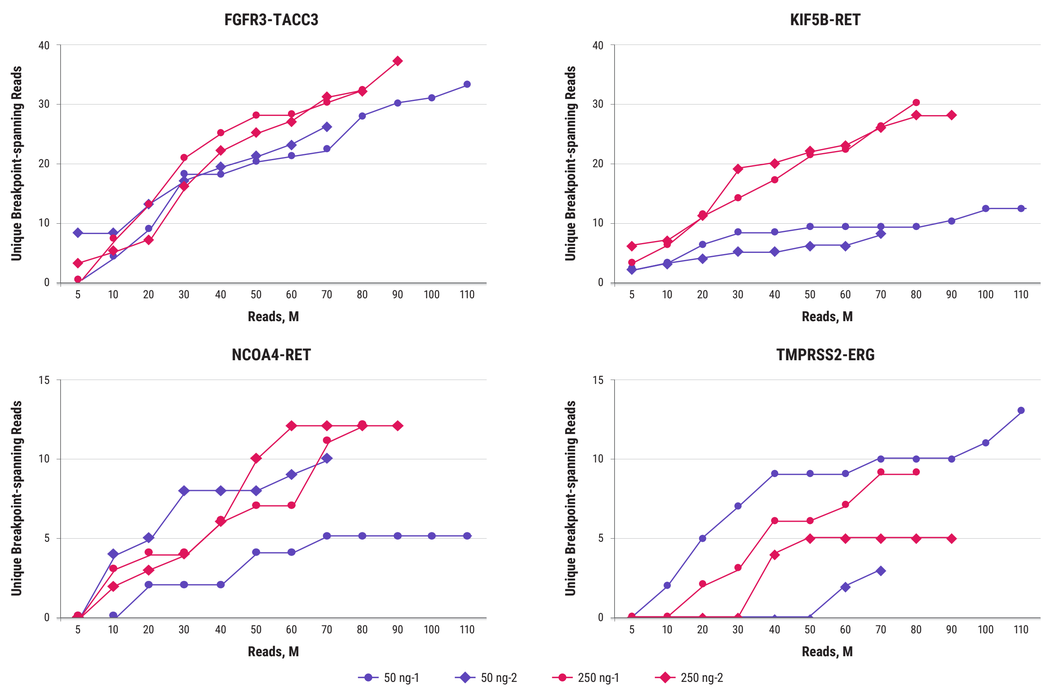

Figure 4. Impact of sequencing depth on fusion calling. A random selection of 4 of the 16 fusions present in the FFPE sample. For all libraries and fusions, the number of unique breakpoint-spanning reads increased with increased sequencing depth.

For the 250 ng libraries, a minimum of 40M reads was required to detect all fusions with a minimum of 3 unique breakpoint-spanning reads. For 50 ng libraries, a minimum of 70M reads was required. This equates to $9.40 and $16.52 per sample in sequencing costs, respectively.

Libraries were prepared using a 0.7X PA cleanup ratio.

Conclusions

- The Watchmaker RNA Library Prep Kit with Polaris Depletion integrates well with the UG 100 sequencer with some minor workflow modifications, including a UDG treatment step integrated into the ligation reaction and a reduced post-amplification cleanup ratio.
- It is feasible to detect fusions with a whole transcriptome sequencing (WTS) approach, and a variety of variables impact fusion detection, including library prep chemistry, RNA mass into library prep, and sequencing depth.