Single-stranded library preparation enables unbiased sequencing of damaged, FFPE-derived DNA for both targeted gene capture and WGS applications

Anna Vilborg[1], Francisco M. De La Vega,[1,2] Hanlee P Ji,[2,3].

[1]TOMA Biosciences, Foster City, CA 94404, USA, [2]Stanford University School of Medicine, Stanford, CA, USA, and [3]Stanford Genome Technology Center Stanford, CA, USA.

Introduction

Next Generation Sequencing is increasingly used in translational cancer research and as a diagnostic test. Most tumor specimens available for testing are formalin-fixed, paraffin-embedded (FFPE) blocks. FFPE-derived DNA is typically damaged, causing difficulties for standard library preparation methods. Frayed ends prevent blunt-ended double-stranded adapter ligation, used for whole genome sequencing (WGS) and targeted bait hybridization library preparation. Fragmented DNA and modified bases interfere with PCR, underlying amplification based targeted sequencing. We present a single-stranded library preparation method especially forgiving for low quality FFPE-derived DNA.

Methods

By excising damaged bases and using single-stranded adapter ligation, we omit repair aimed at generating blunt-ended double-stranded DNA, and eliminate the need for whole genome PCR. After ligation of the first adapter, libraries can either undergo second adapter ligation to generate WGS libraries, or be used as input to oligo-selective sequencing (OS-Seq) target capture (Figure 1, Figure 2).

Figure 1: Single stranded library preparation

DNA is sheared to 550 bp, and damaged bases are removed by excision only, without implementing a corrective repair step (A). The DNA is then denatured followed by adapter ligation to single stranded DNA (B). The single stranded approach allows for adapter ligation to DNA regardless of starting material quality, making the adapter ligation step highly efficient for damaged material. Because of the efficient adapter ligation, no whole genome pre-amplification is required. The library then either undergoes capture using primer-probes in solution in ~1h time, followed by second strand extension and low cycle PCR expansion to create a targeted library (C-D), or is subjected to second adapter ligation to create a WGS library (E).

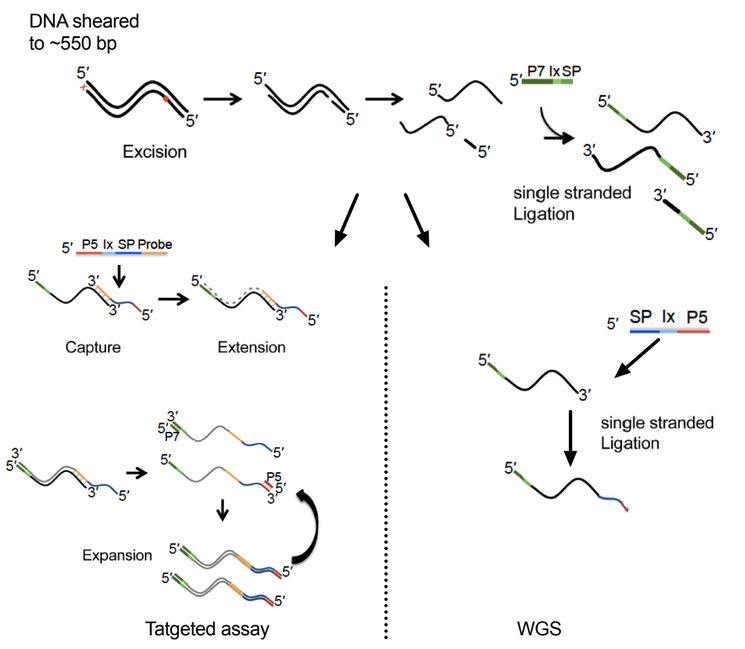
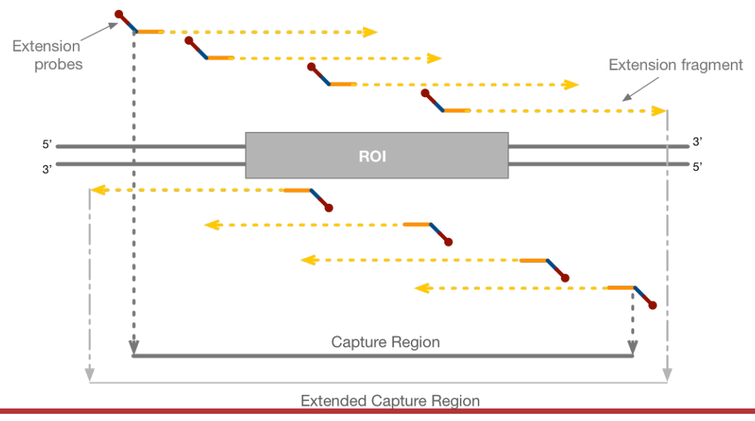


Figure 2: Primer tiling strategy for TOMA OS-Seq

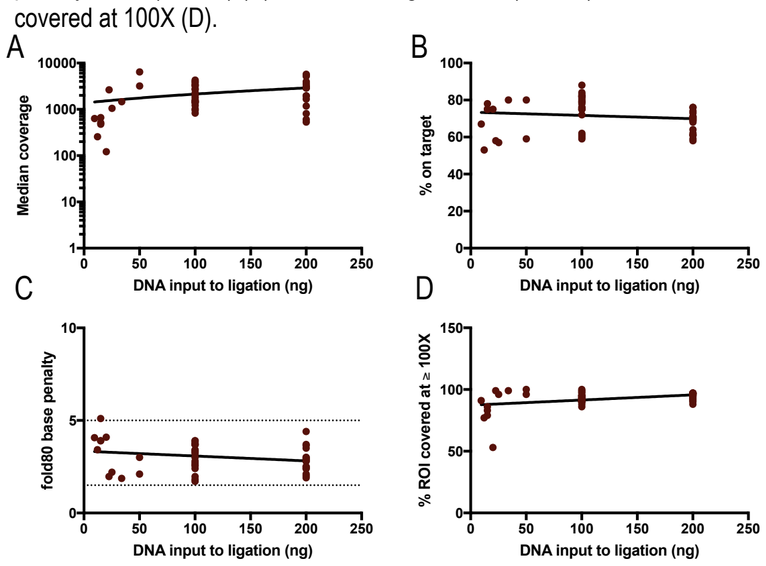
Our primer sequences are short and isothermic, tile the regions of interest from both strands at an average spacing of ~70bp, and are selected to minimize cross-hybridization or false priming sites.



Results

Figure 3: Basic Sequencing metrics

Sequencing metrics from 49 FFPE samples sequenced using our oncology panel of 130 genes (OS-130) are shown as a function of DNA input. For DNA inputs of 30 ng and above, we obtain an average median coverage of 2562 X (SD 1469X) (A), and average on target rate of 72% (SD 8%) (B), fold 80 base penalty of 2.9 (SD 0.7) (C) and an average of 93% (SD 7%) of ROI bases covered at 100X (D).



Dotted lines intersect the y axis at 1.5 and 5, respectively.

Figure 4: OS-Seq is forgiving for low DNA quality

We used a ddPCR assay comparing PCR amplification of a long vs. a short amplicon to determine the DNA QC ratio, or quality, of 49 FFPE derived DNA samples used for library preparation (A). Sequencing metrics for these samples are shown below as a function of quality (QC ratio). We observe no correlation between DNA quality and % on target (B) or uniformity (C), and coverage remains above 500X for all samples (D), demonstrating that the single stranded library preparation method is forgiving for low DNA quality. One example is highlighted in orange: This is a 22 year-old sample of NPC from a Hong Kong biobank, sequenced at 200 ng input.

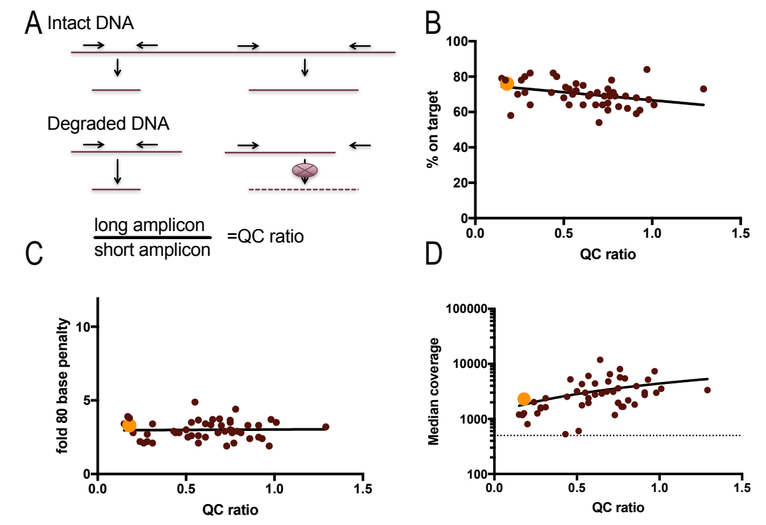
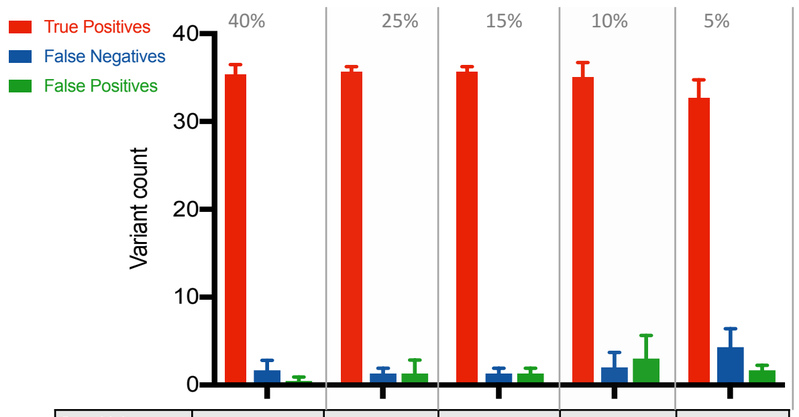


Figure 5: SNV/indel detection

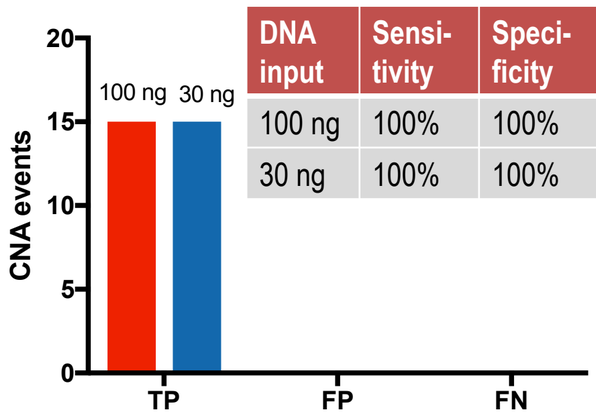
Using SeraCare reference standards, which contain cancer-associated variants at specified VAFs, we detect variants down to to 5.0% expected VAF. DNA input: 100ng. Data was analyzed with TOMA Signome™ analysis pipeline. N=3, error bars show standard deviation.



	40%	25%	15%	10%	5%
Sensitivity	95% (3%)	96% (2%)	96% (2%)	95% (5%)	88% (6%)
Specificity	100% (0%)	100% (0%	100% (0%	100% (0%	100% (0%
PPA	99% (2%)	97% (4%)	96% (2%)	93% (5%)	95% (1%)

Figure 6 and Table 1: CNA detection from reference material

Using TOMA Signome™ CNA calling on SeraCare CNA reference materials with amplifications in ERBB2, FGFR3, EGFR, MYC and MET at either 3, 6, or 12 excess copies (15 CNAs in total), we can obtain highly accurate CNA calling at both 100 ng (red bar) and 30 ng (blue bar) DNA input to ligation. Table 1 shows sensitivity, specificity, and PPV



DNA input	Sensi-tivity	Speci-ficity	PPV
100 ng	100%	100%	100%
30 ng	100%	100%	100%

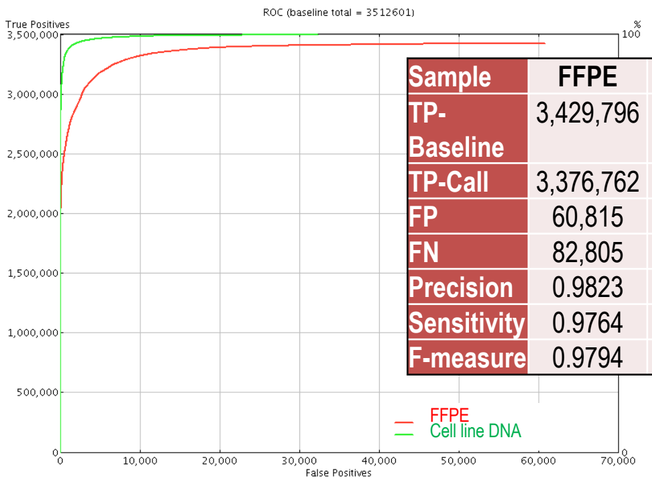
Table 2: DNA based fusion detection

The targeted OS-Seq assay includes detection of 5 common re-arrangements, leading to oncogenic fusions, through targeting introns commonly involved in rearrangements of ROS1, RET, ALK, NTRK1, and FGFR2. Below we use SeraCare reference material to evaluate our ability to detect rearrangements involving ALK and RET at varying %VAFs.

VAF	DNA input	TPR-ALK, supporting reads	NCOA4-RET, supporting reads
2.5%	100 ng	18.5	
5%	100 ng	44	
10%	100 ng	108	12
15%	100 ng	75	22
25%	100 ng	190.5	16
NA24385	100 ng		

Figure 7, Table 3: Whole Genome Sequencing

To test our WGS assay, we used FFPE-embedded GM24385 reference DNA to prepare a PCR-free WGS library, using 200 ng DNA into ligation. The library was sequenced on the HiSeq 2500, 2x250 bp sequencing, resulting in a median depth of 37X and a fold80 base penalty of 1.36. We compared ability to call germline variants from this library with a publicly available Illumina PCR free library of the same DNA source purified from cultured cells, subsampled to the same depth. Figure 8 and Table 2 shows a ROC curve and statistics, respectively, of this comparison



Sample	FFPE	Cell line
TP-Baseline	3,429,796	3,499,775
TP-Call	3,376,762	3,442,467
FP	60,815	32,371
FN	82,805	12,826
Precision	0.9823	0.9907
Sensitivity	0.9764	0.9963
F-measure	0.9794	0.9935

Conclusions

- TOMA OS-Seq generates data with highly uniform coverage and is particularly forgiving to low input quality
- TOMA OS-Seq detects SNVs, indels, select fusions and CNAs, and due to its low reliance on PCR is highly sensitive in the detection of CNAs
- By using the TOMA single stranded approach, FFPE DNA was sequenced to generate similar performance as cell line DNA in a WGS assay

TOMA BIOSCIENCES

// TREATMENT OPTIMIZATION BY MOLECULAR ANALYSIS