ThruPLEX® as a High Sensitivity Library Prep Tool for Whole Exome and Target Panel Sequencing

J.P. Jerome, E. Jan, J. Wibbenmeyer, J. Langmore, K. Shazand

Rubicon Genomics, Inc. Ann Arbor, MI USA • www.rubicongenomics.com • +1 734.677.4845

Abstract

Whole exome sequencing (WES) and target panel sequencing (TPS) are currently the most popular applications in the NGS community. Although ThruPLEX® has already been validated for WES applications in the laboratory of current customers, we wanted to answer the following questions: Can ThruPLEX libraries be easily used for enrichment with major capture platforms from Agilent and Roche (protocol establishment)? How do ThruPLEX enrichment metrics compare with the standard Roche (NimbleGen) and Agilent library preps? What are the advantages of using ThruPLEX libraries for capture?

First, libraries were generated from sheared human gDNA with an average size of 200 bp using the ThruPLEX-FD library prep kit for three low input amounts (500 pg, 10 ng, and 50 ng). Whole exomes (SeqCap® EZ Human Exome v3.0 or SureSelect® Human All Exon V5) or a target panel (IDT xGen® Lockdown®) were enriched from ThruPLEX libraries according to vendor protocols (Roche NimbleGen® SeqCap EZ or Agilent SureSelect XT2) with slight modifications made to integrate and optimize the capture of the ThruPLEX libraries. For comparison, NimbleGen and Agilent XT2 library preps were made and captured according to their respective protocols from the lowest recommended inputs (10 ng and 100 ng, respectively). This established a protocol to integrate ThruPLEX libraries with SureSelect XT2 and NimbleGen SeqCap EZ. Subsequently, protocols were established for using ThruPLEX libraries with other major Agilent capture systems (SureSelect XT and SureSelect QXT). For SureSelect platforms, circulating cell-free DNA from human plasma samples was also used for capture. Captured libraries were sequenced on Illumina MiSeq® with V3 reagents to assess the quality of WES. Reads were downsampled before mapping to the hg19 reference genome and to the appropriate exome/panel for analysis using Picard CalculateHsMetrics. Percentage of bases on, off and near the exome, percentage of PCR duplicates, reads on target, target coverage, and an estimate of the amount of sequence data needed to obtain specific coverage levels of the exome was calculated.

Conclusions: 1) ThruPLEX libraries made from low input amounts and plasma-derived DNA may be enriched using major capture platforms with minor protocol modifications; 2) ThruPLEX was as highly enriched as control libraries from NimbleGen, and better than Agilent libraries; 3) the high sensitivity (lower input requirements) and fast, easy workflow make ThruPLEX an ideal library prep for exome or panel sequencing applications from limited, degraded, clinical DNA samples.

Since ThruPLEX DNA-seq has better performance than ThruPLEX-FD (data not shown), it is the preferred library preparation kit for clinical DNA samples such as cell-free DNA from plasma or other biofluids.

High Efficiency Capture of Low Input Amounts on Agilent SureSelect Platforms

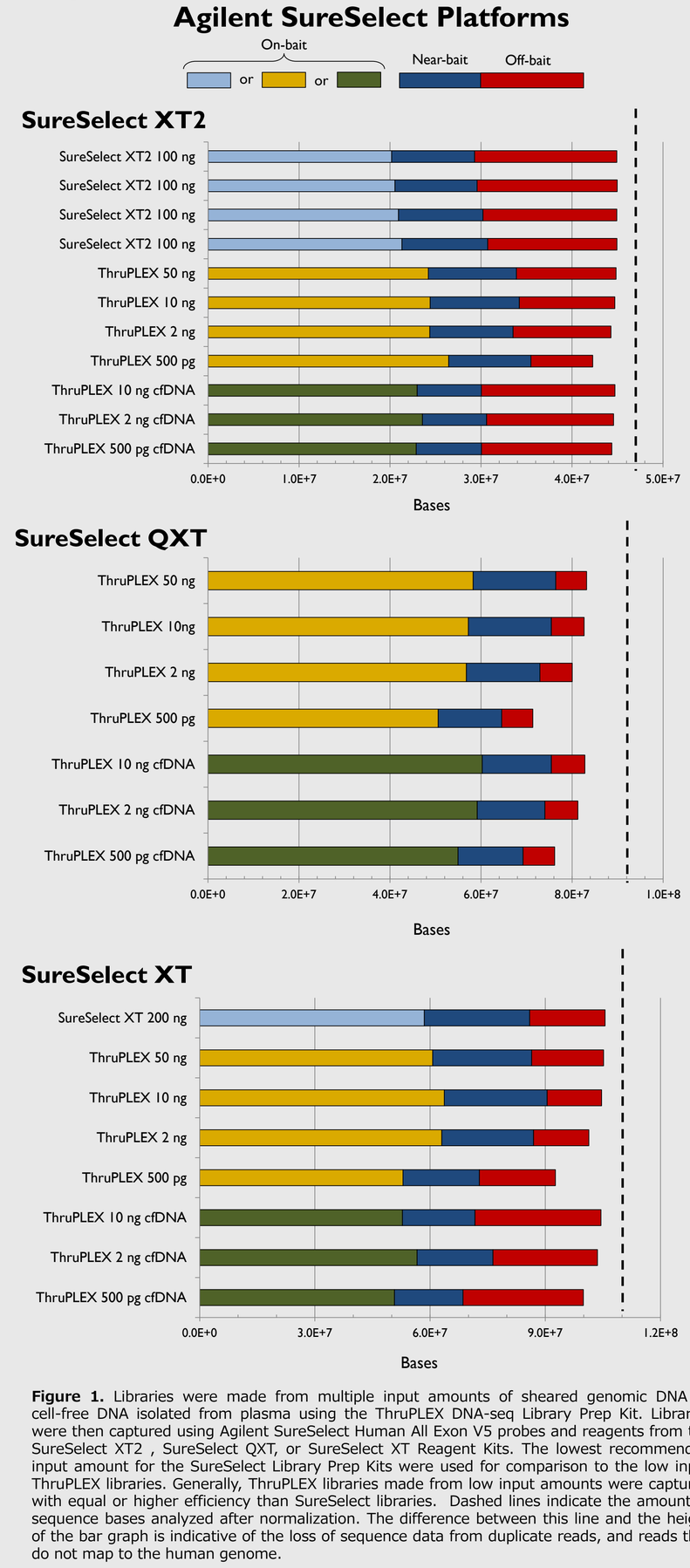


Figure 1. Libraries were made from multiple input amounts of sheared genomic DNA or cell-free DNA isolated from plasma using the ThruPLEX DNA-seq Library Prep Kit. Libraries were then captured using Agilent SureSelect Human All Exon V5 probes and reagents from the SureSelect XT2 , SureSelect QXT, or SureSelect XT Reagent Kits. The lowest recommended input amount for the SureSelect Library Prep Kits were used for comparison to the low input ThruPLEX libraries. Generally, ThruPLEX libraries made from low input amounts were captured with equal or higher efficiency than SureSelect libraries. Dashed lines indicate the amount of sequence bases analyzed after normalization. The difference between this line and the height of the bar graph is indicative of the loss of sequence data from duplicate reads, and reads that do not map to the human genome.

High Quality ThruPLEX Libraries are Captured from Low Amounts of Input DNA

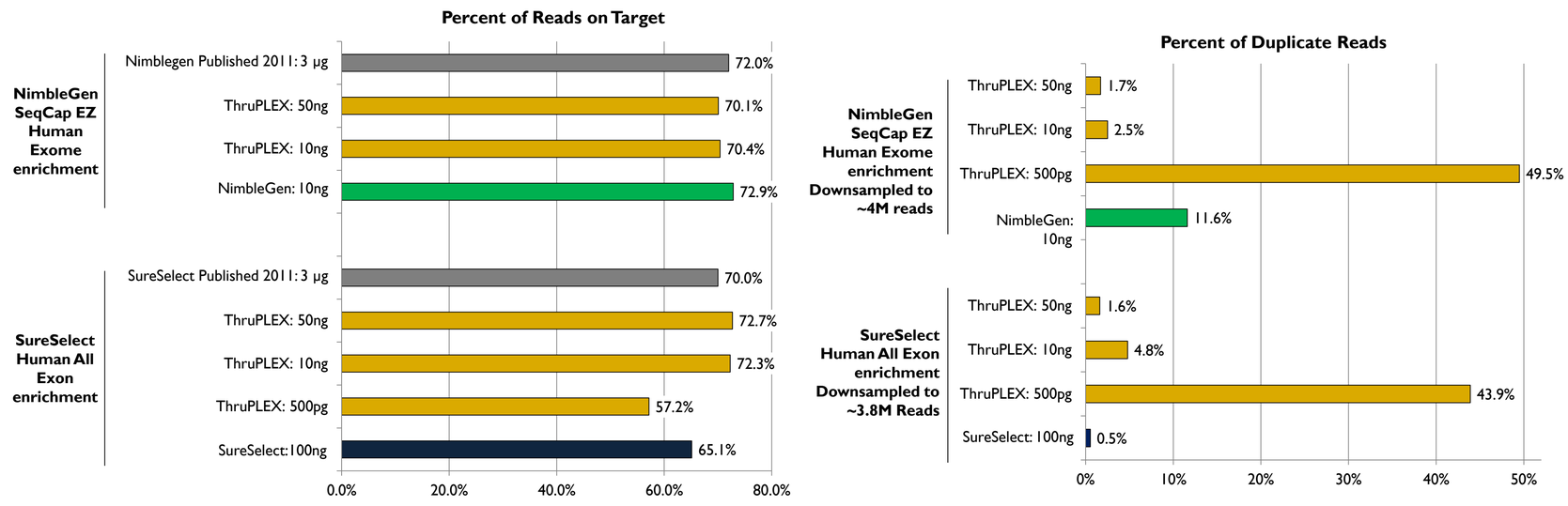


Figure 2. Left: Using two major capture platforms (Roche NimbleGen SeqCap EZ and Agilent SureSelect XT2) we were able to show that ThruPLEX-FD libraries are captured with approximately equal efficiency as NimbleGen libraries and better efficiency than SureSelect XT2 libraries. Efficiency of capture is shown on the left as the percent of reads on target. Published data was adapted from *Clark et al., (2011) Nature Biotechnology 29: 908-914.* **Right:** The number of reads for each sample was normalized, and the percentage of PCR duplicates was measured with Picard Mark Duplicates. This shows that from equal input (10 ng) ThruPLEX-FD WES has less duplicate reads and thus higher sequence diversity than the NimbleGen library. As expected, the Agilent SureSelect library had lower numbers of duplicate reads than the ThruPLEX-FD libraries: This is due to the higher input amount into the SureSelect XT2 library prep.

Less Sequence Data Needed for ThruPLEX WES

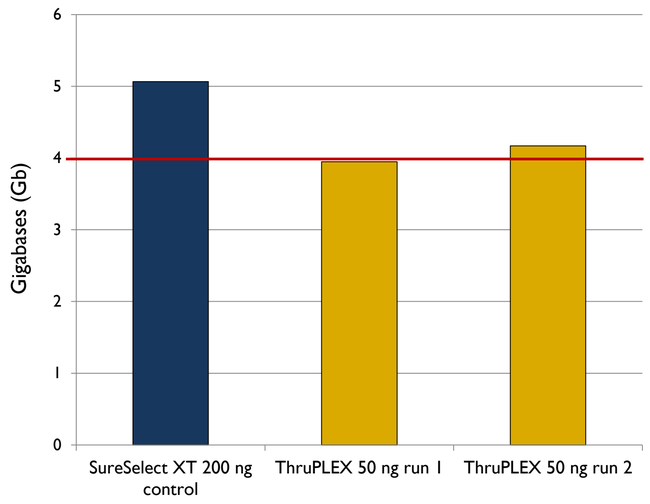


Figure 3. The Agilent SureSelect protocol calls for 4 Gb of sequence data to obtain 20X coverage of 80% of the V5 All Exon exome (denoted by the red line in the chart above). Picard CalculateHsMetrics calculates the amount of sequencing necessary to obtain this level of coverage ("HS_PENALTY_20X"). Using SureSelect XT capture of ThruPLEX DNA-seq libraries we estimated that it would take approximately 4 Gb of sequence to achieve 20X coverage of 80% of the SureSelect V5 exome with 50 ng of sheared DNA as input. This value was consistent from independent WES runs. Using the standard SureSelect XT library prep kit with the lowest recommended input (200 ng) it would take 5 Gb of sequence.

Efficient Target Panel Sequencing from Low Inputs

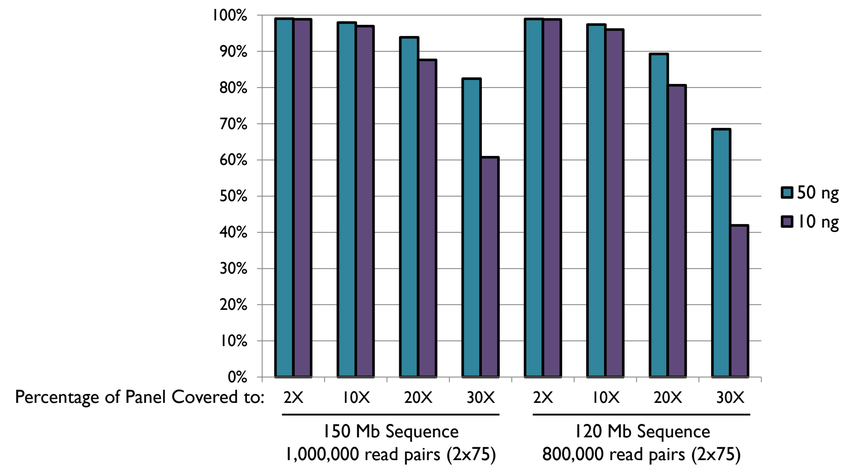


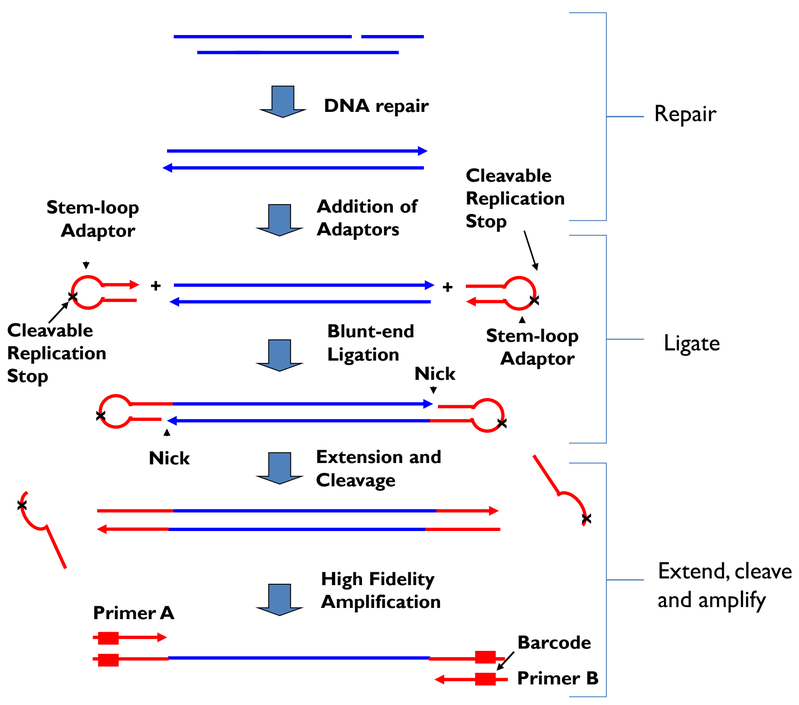
Figure 4. ThruPLEX DNA-seq libraries were used with NimbleGen SeqCap EZ reagents and the xGen Acute Myeloid Leukemia Cancer Panel to perform target panel sequencing (TPS). For ThruPLEX libraries made from 10 ng and 50 ng input, 800K 2x75 reads yielded coverage of at least 20X for over 80% of targets. For 50 ng, 80% of targets were covered to at least 30X with 1M 2x75 read pairs.

ThruPLEX Workflow Advantage

		End Repair	Cleanup	Adenylation	Cleanup	Adapter Ligation	Cleanup	Size Selection	Library Amplification		
ThruPLEX-FD	Hands-on Total	5 min 50 min	X	X	X	5 min 45 min	X	X	5 min 22 min	**15 min 117 min**	**3 steps**
NimbleGen Library Preparation*	Hands-on Total	10 min 40 min	30 min 90 min	10 min 40 min	30 min 90 min	10 min 25 min	30 min 90 min	25 min 60 min	5 min 21 min	**150 min 456 min**	**8 steps**
SureSelectXT2 Library Prep	Hands-on Total	10 min 40 min	25 min 50 min	10 min 50 min	X	10 min 25 min	25 min 50 min	X	5 min 33 min	**85 min 248 min**	**6 steps**

*Kapa Library Preparation Kit is the designated kit for Roche NimbleGen SeqCap EZ System

ThruPLEX Chemistry



ThruPLEX NGS Target Enrichment Workflow

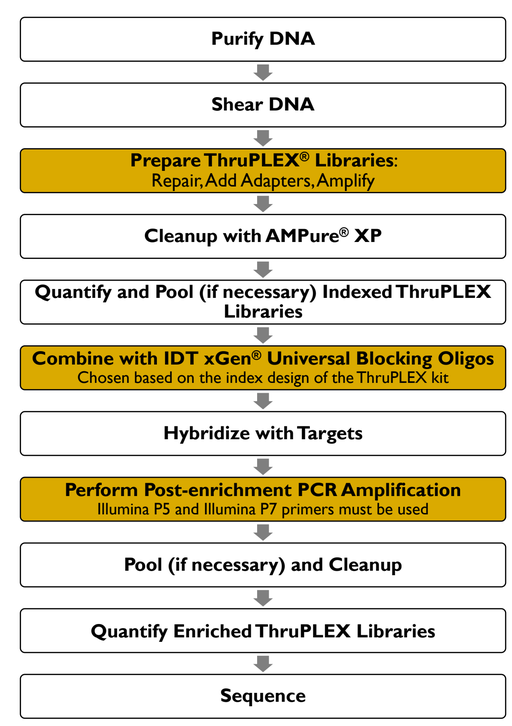


Figure 5. Top: The ThruPLEX workflow is highly streamlined, the reactions occur in a single tube, and there is no need for intermediate clean-up steps. These features greatly lower the risk of sample loss and contamination during library prep, while reducing the amount of time spent making libraries. **Left**: ThruPLEX libraries have an improved DNA repair reaction and a highly efficient ligation of proprietary stem-loop adaptors, followed by amplification where indexes are incorporated to label libraries for subsequent pooling. **Right:** ThruPLEX libraries are easily integrated into the workflows of major capture platforms. Modifications are highlighted in yellow. IDT xGen Universal Blocking Oligos are used during hybridization of ThruPLEX libraries with commercially available probe sets to decrease non-specific pulldown of DNA. Because ThruPLEX libraries are indexed, Illumina P5 and P7 primers must be used for post-capture amplification when using SureSelect XT or QXT. Detailed protocols are available at rubicongenomics.com/applications/enrichment.