TECH NOTE

Low-input whole-exome sequencing using SMARTer ThruPLEX technology and Agilent SureSelect Target Enrichment Systems

Introduction

The presence of circulating DNA has long been observed in human blood and is primarily attributed to apoptosis (Williamson 1970). Circulating tumor DNA (ctDNA) isolated from the plasma of cancer patients has been the subject of many research studies (Murtaza et al. 2013; Klevebring et al. 2014).

Many of these experiments have been made possible with the advent of next-generation sequencing (NGS) platforms, which allow for identification of copy number and single nucleotide variants. Since the amount of DNA present in plasma is generally low, highly sensitive methods for preparing NGS libraries are required. SMARTer ThruPLEX technology provides an excellent choice due to its low-input requirements, using as little as 1 ng of cfDNA, as well as its single-tube workflow, which helps prevent sample loss and cross-contamination and ensures positive sample identification.

Instead of sequencing the entire genome, many researchers are choosing to direct their attention to the exome or a targeted subset of the exome to focus on the coding regions and reduce costs. This approach consists of enriching the exonic regions of interest to identify various types of genetic alterations in the targeted regions through sequencing and analysis. The present work describes exome enrichment of cfDNA libraries by combining our highly sensitive library preparation product, SMARTer ThruPLEX DNA-seq kit, with the downstream enrichment tools from Agilent Technologies, namely the SureSelect^{XT}, SureSelect^{XT2} and SureSelect^{QXT} Target Enrichment Systems.

SMARTer ThruPLEX DNA-seq kit is an essential addition to the SureSelect platforms because the default SureSelect library preparation kits are limited to input amounts of 200 ng (SureSelect^{XT}), 100 ng (SureSelect^{XT2}), and 50 ng (SureSelect^{QXT}), while cfDNA in plasma samples is frequently present in amounts between 1 ng and 30 ng per 1 ml of plasma. The data we present here clearly demonstrate a powerful method that will allow investigators to obtain high-quality exome data from limited amounts of starting material with minimal protocol adjustment.

Results

The relatively low level of cfDNA in plasma samples presents a major challenge to the detection of genomic variations using next-generation sequencing. We carried out whole-exome enrichment of cfDNA from plasma samples by integrating SMARTer ThruPLEX DNA-seq kit with Agilent SureSelect^{XT}, ^{XT2}, and ^{QXT} Target Enrichment Systems (Figure 1). As a reference, sheared gDNA was also enriched and sequenced following library preparation with the SureSelect^{XT}, ^{XT2}, and ^{QXT} Library Prep Kits. The key sequencing metrics are summarized in Table I.

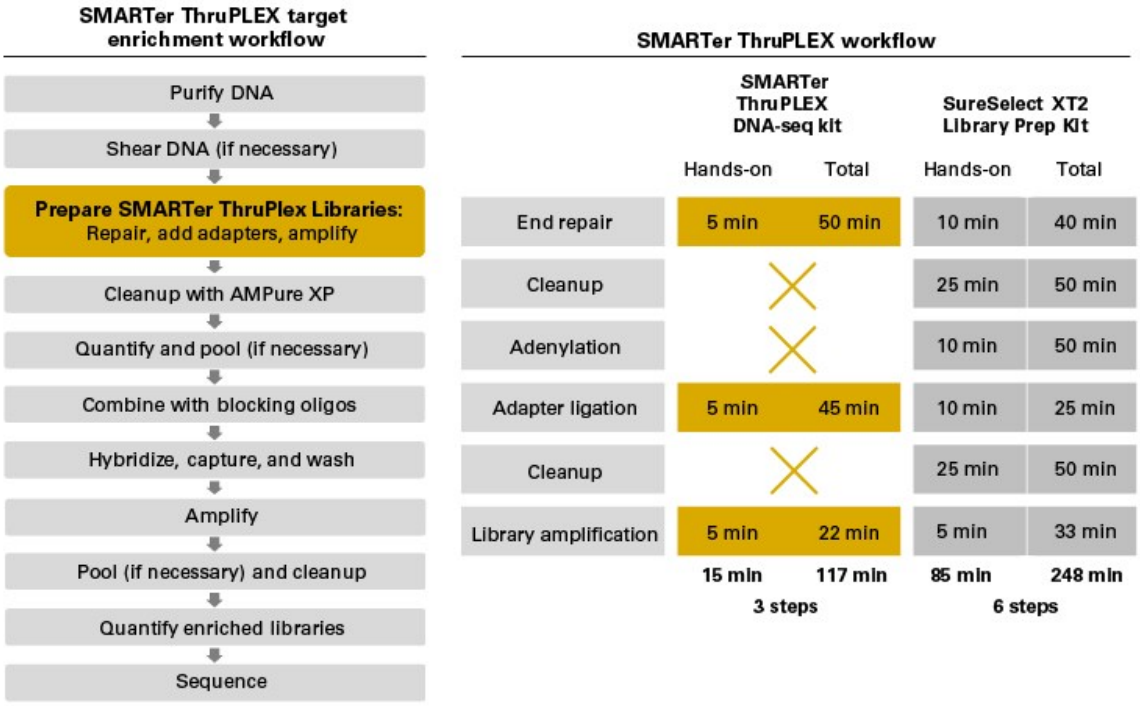

Figure 1. ThruPLEX workflow. General workflow diagram integrating SMARTer ThruPLEX DNA-seq kit with Agilent SureSelect platforms (left panel). SMARTer ThruPLEX DNA-seq kit converts double-stranded DNA samples to indexed libraries in three simple steps: end repair, adapter ligation, and high-fidelity library amplification (right panel). The streamlined workflow prevents sample loss and handling errors and is performed in a single tube or well in less than two hours.

Results from whole-exome enrichment of the SMARTer ThruPLEX DNA-seq cfDNA libraries with each of the SureSelect platforms were comparable to those of the SMARTer ThruPLEX DNA-seq gDNA libraries in terms of key sequencing metrics (Table I) and on-target specificity (Figure 2). These data demonstrate the exceptional repair capacity of SMARTer ThruPLEX technology. When comparing the number of enriched (on-bait plus near-bait) bases, SMARTer ThruPLEX DNA-seq cfDNA libraries were 99% (^{QXT}), 88% (^{XT2}), and 79% (^{XT}) efficient relative to the gDNA libraries (Figure 2). This loss of information can be attributed to factors such as shorter fragment length and lower complexity of cfDNA from plasma samples.

Enrichment platform	Library preparation	Input type	Input amount	Unique reads	Fold enrichment	Library size	% duplication
SureSelect^{XT}	SMARTer ThruPLEX DNA-seq	cfDNA	500 pg	1,396,048	31.6	3.65 x 10^6	4.67
			2 ng	1,444,625	34.0	1.51 x 10^7	1.32
			10 ng	1,458,560	31.5	4.57 x 10^7	0.40
		gDNA	10 ng	1,450,470	37.9	3.50 x 10^7	0.78
	SureSelect^{XT}	gDNA	200 ng	1,456,623	34.5	2.11 x 10^8	0.12
SureSelect^{XT2}	SMARTer ThruPLEX DNA-seq	cfDNA	500 pg	619,266	32.1	7.36 x 10^6	1.47
			2 ng	621,686	32.9	1.22 x 10^7	1.04
			10 ng	624,013	32.0	3.77 x 10^7	0.31
		gDNA	10 ng	622,495	34.0	2.64 x 10^7	0.38
	SureSelect^{XT2}	gDNA	100 ng	625,674	29.5	8.05 x 10^7	0.55
SureSelect^{QXT}	SMARTer ThruPLEX DNA-seq	cfDNA	500 pg	1,052,819	44.9	2.62 x 10^6	8.53
			2 ng	1,122,128	45.3	9.12 x 10^6	3.30
			10 ng	1,142,897	45.3	3.36 x 10^7	0.82
		gDNA	10 ng	1,139,272	43.1	2.83 x 10^7	1.06

Table I. High-quality exome-enriched libraries. Summary of sequencing metrics from whole exome sequencing of cfDNA and gDNA libraries prepared using the SMARTer ThruPLEX DNA-seq kit or SureSelect Library Prep Kits and enriched with SureSelect^{XT}, ^{XT2}, and ^{QXT} target enrichment systems.

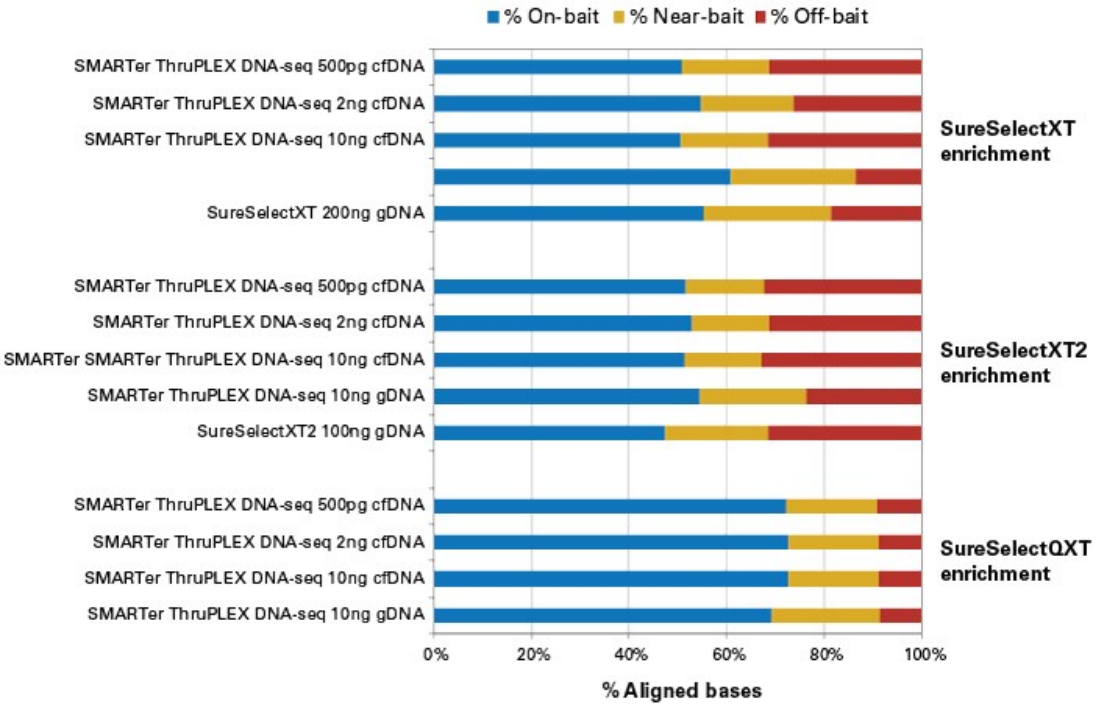

Figure 2. High-quality exome data. SMARTer ThruPLEX DNA-seq cfDNA libraries demonstrated high performance when enriched using SureSelect^{XT}, ^{XT2}, and ^{QXT} Target Enrichment Systems. SMARTer ThruPLEX DNA-seq cfDNA libraries yielded performance similar to SureSelect gDNA libraries at low input amounts that are beyond the reach of the SureSelect Library Prep Kits. SMARTer ThruPLEX DNA-seq gDNA libraries (10 ng input) generated higher percentage of on-bait bases when compared to gDNA libraries prepared from SureSelect^{XT} (200 ng input) and ^{XT2} (100 ng input) Library Prep Kits.

In general, the SureSelect^{XT2} platform showed similar performance compared to SureSelect^{XT}. In the SureSelect^{XT2} workflow, samples are pooled prior to hybridization, which confers ease of use and cost advantages. SMARTer ThruPLEX DNA-seq Kit can be integrated very conveniently with the SureSelect^{XT2} platform, requiring only minor adjustments to the protocol and additional universal blocking oligos from IDT (Table II). The SureSelect^{QXT} platform provided higher mean target coverage and required the shortest hybridization time. Enrichment with SureSelect^{QXT} also appeared to be more efficient despite variable input amounts. However, SureSelect^{QXT} resulted in much higher AT-dropout rates (Figure 3), which may be the consequence of the temperature cycling during hybridization used in its protocol.

	Additional reagents required	Omitted reagents
SureSelect^{XT} Reagent Kit	Illumina P5 Primer Illumina P7 Primer xGen Universal Blocking Oligo i5 xGen Universal Blocking Oligo i7	SureSelect TE Kit Indexing Hyb Module Box #2 • SureSelect ILM Indexing Pre Capture PCR Reverse Primer • SureSelect ILM Indexing Post Capture Forward PCR Primer SureSelect Library Prep Kit
SureSelect^{XT2} Reagent Kit	xGen Universal Blocking Oligo i5 xGen Universal Blocking Oligo i7	^{XT2} Pre-capture Indexes ^{XT2} Library Prep Kit, except • SureSelect Herculase II Master Mix • ^{XT2} Primer Mix
SureSelect^{QXT} Reagent Kit	Illumina P5 Primer Illumina P7 Primer xGen Universal Blocking Oligo i5 xGen Universal Blocking Oligo i7	^{QXT} Library Prep Kit, Box 2, except • Herculase II Fusion DNA Polymerase • 5X herculase II Reaction Buffer • 100 mM dNTP Mix (25 mM each dNTP) ^{QXT} TE Kit, Hyb Module, Box #1 • SureSelect QXT Stop Solution ^{QXT} TE Kit, Hyb Module, Box #2 • ^{QXT} Primer Mix

Table II. SureSelect compatibility. List of reagents used when integrating SMARTer ThruPLEX DNA-seq Kit with SureSelect Target Enrichment Systems.

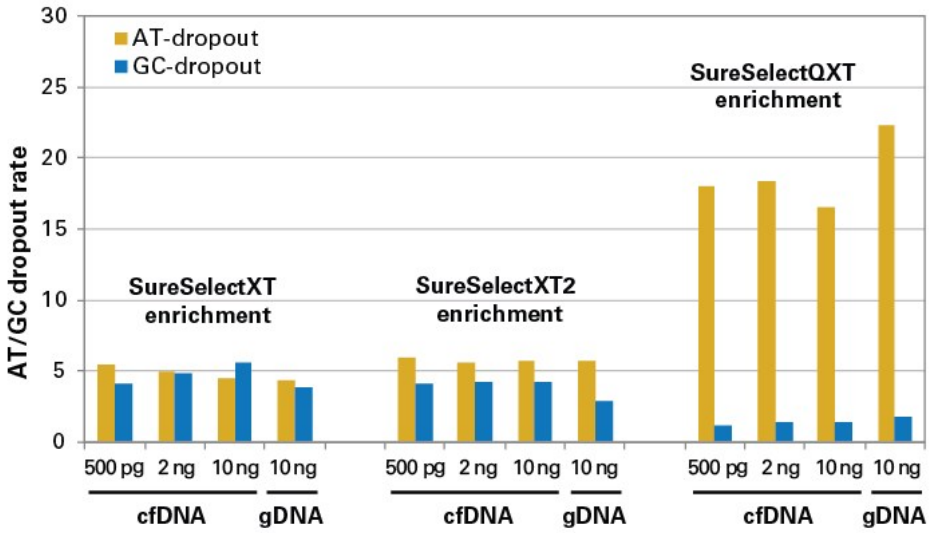

Figure 3. AT/GC dropout. AT- and GC-dropout rates of SMARTer ThruPLEX DNA-seq gDNA and cfDNA libraries enriched with SureSelectXT, XT2, and QXT Target Enrichment Systems. SureSelect^{QXT} enriched libraries suffered from high AT-dropout rates. Approximately 15 to 20% of total reads that should have mapped to GC≥ 50% regions were mapped elsewhere. AT-dropout is a measure of coverage of regions ≤ 50% GC relative to the mean; GC dropout is a measure of coverage of regions ≥ 50% GC relative to the mean.

Deep sequencing data was also generated using an Illumina NextSeq® 500. SMARTer ThruPLEX DNA-seq gDNA/SureSelect^{XT} libraries at 10 ng input required less than 1 gb of additional sequencing data than the SureSelect^{XT} gDNA library at 200 ng input to yield 20X coverage of at least 80% of the exome (Figure 4). As expected, more sequencing data is required for libraries made from 10 ng of cfDNA. This is likely due to decreased diversity due to reduced input amount and lower capture efficiency of plasma cfDNA samples. For the SMARTer ThruPLEX DNA-seq/SureSelect^{XT} library, 100 M total 75-base-reads per sample were adequate for SNV calling of at least 85% of the exome (Figure 5). From the NextSeq 500 high output run (2 x 75 bp), up to 8 cfDNA libraries prepared from 10 ng input could be sequenced to achieve at least 80% coverage of the exome.

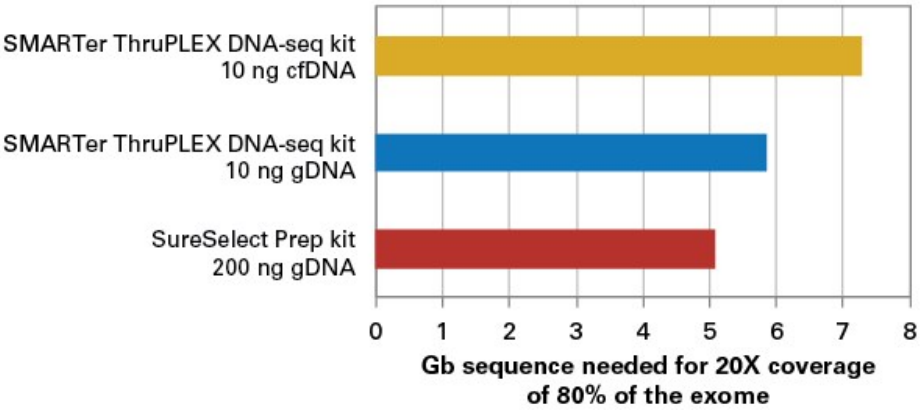

Figure 4. Sequencing requirements. The estimated amount of sequencing required to achieve 20X coverage of 80% of the exome. Samples were prepared with SMARTer ThruPLEX DNA-seq kit or SureSelect^{XT} Library Prep Kit, exome-enriched using the SureSelect^{XT} platform, and sequenced on an Illumina NextSeq 500 using a 2 x 75 bp run.

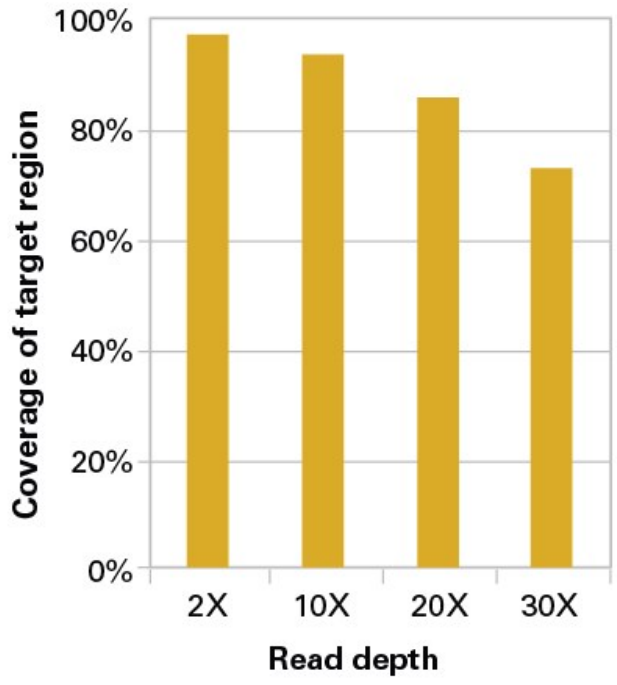

Figure 5. Exome coverage. SMARTer ThruPLEX DNA-seq libraries were prepared from 10 ng of cfDNA and exome enriched using the SureSelect^{XT} platform. Up to eight libraries may be sequenced on a NextSeq 500 high output 2 x 75 bp run, yielding up to 800 M reads to achieve at least 80% coverage of the exome at 20X or higher.

Conclusions

By integrating the SMARTer ThruPLEX DNA-seq kit with Agilent SureSelect Target Enrichment Systems, we were able to exploit SMARTer ThruPLEX technology's high sensitivity to perform library preparation and whole exome enrichment using the low amounts of cfDNA present in plasma samples. The amount of data generated is adequate for SNV calling. Compatibility of SMARTer ThruPLEX DNA-seq Kit with SureSelect platforms can be easily attained with minor adjustments to the SureSelect protocols and with the addition of universal blocking oligos and sequencing primers. The SureSelect^{XT2} platform, in which samples are pooled prior to hybridization, is the simplest to integrate. In addition to its higher sensitivity and excellent performance, SMARTer ThruPLEX DNA-seq Kit offers a faster and simpler workflow with a single-tube, three-step protocol. An integrated enrichment method combining SMARTer ThruPLEX and SureSelect technologies will be instrumental in translational genomic research where the DNA of interest is present in limiting quantities.

Methods

DNA isolation

Plasma samples were acquired from Medical Research Networx, LLC. Blood was collected from healthy donors into BD Vacutainer EDTA tubes and plasma was separated by double centrifugation at 4°C for 12 minutes at 1,500*g*. Processed plasma samples were stored at –80°C until DNA was extracted. Qiagen QIAamp Circulating Nucleic Acid Kit was used to extract DNA from 5 ml of plasma. DNA quantity and size distribution were measured using Qubit (Thermo Fisher Scientific) and a Bioanalyzer (Agilent), respectively.

Library preparation

Libraries were prepared from either cfDNA isolated from plasma samples or Covaris-sheared (average size 200 bp) NA12878 genomic DNA (gDNA) using SMARTer ThruPLEX DNA-seq Kit with dual indexes at different input amounts (Figure 1). The quality of prepared libraries was verified on Qubit and Bioanalyzer (Figure 1). All cfDNA libraries enriched on the same SureSelect platform were prepared from the same plasma sample. As a reference, libraries were also prepared with SureSelect Library Prep Kits using the lowest input amounts recommended by the manufacturer.

Plasma DNA sample	Yield (ng/ml plasma)	Total yield (ng)
1	4.2	21.0
2	2.9	14.5
3	1.9	9.5
4	4.0	20.0
5	8.5	42.5
6	3.4	17.0
7	6.8	27.2
8	4.5	22.5
9	5.3	26.5

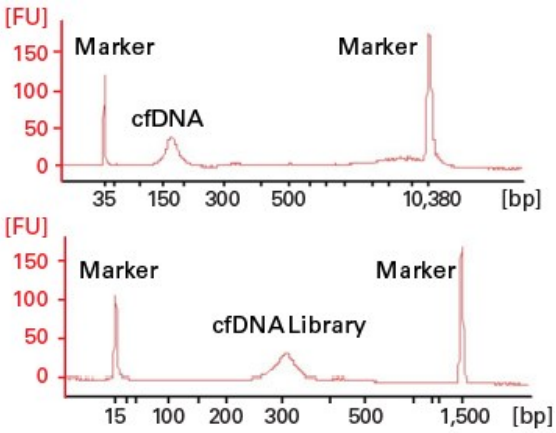

Figure 6. Library preparation from low-input amounts. Left: The amount of cfDNA isolated from plasma samples was determined by Qubit. All samples yielded less than 10 ng of cfDNA per ml of plasma, demonstrating the low level of DNA in plasma samples. Right: Bioanalyzer traces. Top right: Cell-free DNA in plasma is generally found to be between 150 bp and 200 bp in length. Bottom right: ThruPLEX DNA-seq library preparation adds approximately 140 bases to the DNA molecules.

Whole exome enrichment and sequencing

Amplified libraries were purified using Agencourt AMPure XP (Beckman Coulter, Cat. # A63880) and eluted in 20–50 µl of PCR grade water. Prior to enrichment, purified libraries were individually assessed using a Qubit and a Bioanalyzer. For enrichment using the SureSelect^{XT2} platform, purified libraries were pooled to obtain 1.5 µg of indexed DNA. For SureSelect^{XT} and ^{QXT} platforms, the entire volume of each SMARTer ThruPLEX DNA-seq library was used for exome enrichment (Table III).

Whole-exome enrichment reagent kit and capture library	Sample	Library preparation kit	Library prep input (ng)	PCR cycles	Yield (ng)	Capture input (ng)
SureSelect^{XT} Reagent Kit SureSelect^{XT} Human All Exon V5	cfDNA	SMARTer ThruPLEX DNA kit	0.5	14	572	572
			2	11	388	388
			10	9	577	577
	gDNA	SMARTer ThruPLEX DNA kit	10	7	449	449
		SureSelect^{XT} Library Prep Kit	200	10	2,430	750
SureSelect^{XT2} Reagent Kit SureSelect^{XT2} Human All Exon V5	cfDNA	SMARTer ThruPLEX DNA kit	0.5	14	610	610
			2	11	343	343
			10	9	430	430
	gDNA	SMARTer ThruPLEX DNA kit	10	7	268	268
		SureSelect^{XT2} Library Prep Kit	100	8	1,366	375
SureSelect^{QXT} Reagent Kit SureSelect^{XT} Human All Exon V5	cfDNA	SMARTer ThruPLEX DNA kit	0.5	14	837	837
			2	11	396	396
			10	9	496	496
	gDNA	SMARTer ThruPLEX DNA kit	10	7	365	365

Table III. Experimental design. For each SureSelect Target Enrichment System, libraries were prepared from cfDNA or gDNA using SMARTer ThruPLEX DNA-seq kit or the corresponding SureSelect Library Prep Kit. Whole exome enrichment was carried out using the SureSelect Reagent Kits and Human All Exon V5 probe sets.

Exome enrichment was performed using the SureSelect Reagent Kits and SureSelect Human All Exon V5 probe sets (Table III). To integrate SMARTer ThruPLEX DNA-seq Kit with the SureSelect platforms, reagent use was modified (Table II; see Results section). For all three platforms, IDT xGen Universal Blocking Oligos (TS HT-i7 and TS HT-i5) were spiked into the blocking mixture containing SMARTer ThruPLEX DNA-seq libraries prior to hybridization with the probes. The xGen Universal Blocking Oligos were each resuspended to 1 µl per reaction (1 nmol) in nuclease-free water prior to use. For SureSelect^{XT} and ^{QXT} platforms, Illumina P5 and P7 primers were used for post-capture amplification of the SMARTer ThruPLEX DNA-seq libraries. All samples were subjected to 10 cycles of post-capture amplification to produce the final sequencing libraries.

Sequencing

Pooled samples were quantified using KAPA Library Quantification Kit and loaded onto Illumina MiSeq® v3 flow cells. Reactions were carried out as 2 x 75 bp paired-end runs, and approximately 0.6–1.5 M reads per sample were generated. Selected samples were also sequenced on an Illumina NextSeq 500 as a 2 x 75 bp high output paired-end run.

Data analysis

Sequence reads were analyzed using DNANexus. Reads were mapped to the human genome reference, hg19, using the Burrows-Wheeler Alignment algorithm, BWA-MEM, to generate BAM files for each sample. BAM files were downsampled to obtain equal numbers of reads, and duplicates were marked using Picard Mark Duplicates (Li and Durbin 2009; Broad Institute 2017). Output files from Picard Mark Duplicates were used to determine quality metrics related to the whole exome capture and sequencing using Picard CalculateHsMetrics.

References

Broad Institute. Picard Tools - A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. http://broadinstitute.github.io/picard/

Klevebring, D. et al. Evaluation of Exome Sequencing to Estimate Tumor Burden in Plasma. *PLoS One* 9, e104417 (2014).

Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1,754–1,760 (2009).

Murtaza, M. et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 497, 108–112 (2013).

Williamson, R. Properties of rapidly labelled deoxyribonucleic acid fragments isolated from the cytoplasm of primary cultures of embryonic mouse liver cells. *J. Mol. Biol.* 51, 157–168 (1970).

Related Products

Cat. #	Product	Size	License	Quantity	Details
R400406	SMARTer® ThruPLEX® DNA-seq 48D Kit	48 Rxns			
R400427	SMARTer® ThruPLEX® DNA-seq 48S Kit	48 Rxns			
R400407	SMARTer® ThruPLEX® DNA-seq 96D Kit	96 Rxns			
R400428	ThruPLEX® DNA-seq 12S (48) Kit	48 Rxns			

Add to Cart

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999