Application Note

Whole Exome Enrichment of Cell-Free DNA in Plasma Samples using ThruPLEX® DNA-seq Kit and Agilent SureSelect® Target Enrichment Systems

Kamran Shazand, Edward Jan & John Paul Jerome, Rubicon Genomics, Inc.

Introduction

The presence of circulating DNA has long been observed in human blood and is primarily attributed to apoptosis.[1] Circulating tumor DNA (ctDNA) isolated from the plasma of cancer patients has been the subject of many research studies.[2-3]

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 of preparing NGS libraries are required. Rubicon Genomics' 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, ThruPLEX DNA-seq Kit, with the downstream enrichment tools from Agilent Technologies, namely the SureSelect®XT, SureSelectXT2 and SureSelectQXT Target Enrichment Systems.

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 (SureSelectXT), 100 ng (SureSelectXT2), and 50 ng (SureSelectQXT), 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.

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 1500 × 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 was measured using Qubit® (ThermoFisher) and a Bioanalyzer® (Agilent), respectively.

Library preparation: Libraries were prepared from either cfDNA isolated from plasma samples or Covaris®-sheared (200 bp average size) NA12878 genomic DNA (gDNA) using 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 reference, libraries were also prepared with SureSelect Library Prep Kits using the lowest input amounts recommended by the manufacturer.

Whole Exome Enrichment. Amplified libraries were purified using Agencourt® AMPure® XP (Beckman Coulter, CAT. NO. 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 SureSelectXT2 platform, purified libraries were pooled to obtain 1.5 µg of indexed DNA. For SureSelectXT and QXT platforms, the entire volume of each ThruPLEX DNA-seq library was used for exome enrichment **(Figure 2)**.

Exome enrichment was performed using the SureSelect Reagent Kits and SureSelect Human All Exon V5 probe sets **(Figure 2)**. To integrate ThruPLEX DNA-seq Kit with the SureSelect platforms, reagent use was modified **(Figure 3)**. For all three platforms, IDT xGen® Universal Blocking Oligos (TS HT-i7 and TS HT-i5) were spiked into the blocking mixture containing ThruPLEX DNA-seq libraries prior to hybridization with the probes. The xGen Universal Blocking Oligos were each resuspended to 1 µL per reaction (1 nmole) in nuclease-free water prior to use. For SureSelectXT and QXT platforms, Illumina® P5 and P7 primers were used for post-capture amplification of the 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 – 1.5M 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.

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

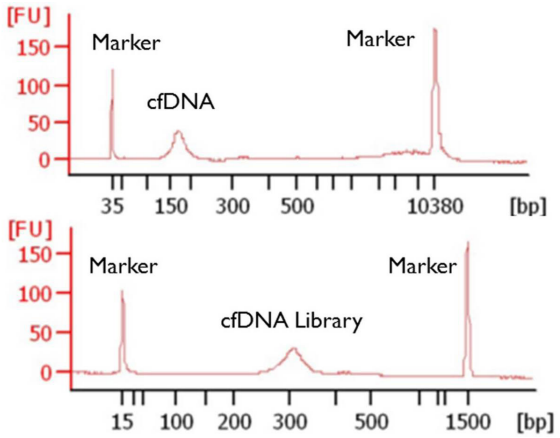

Figure 1. Library preparation from low input amounts. Left: *The amount of cfDNA isolated from plasma samples was determined by Qubit. All samples yielded less than 10 ng of cfDNA per mL of plasma, demonstrating the low level of DNA in plasma samples.* ***Right:*** *Bioanalyzer traces.* ***Top right:*** *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 Reagent Kit and Capture Library	Sample	Library Preparation Kit	Library Prep Input	PCR Cycles	Yield	Capture Input
SureSelectXT Reagent Kit SureSelectXT Human All Exon V5	cfDNA	ThruPLEX DNA-seq Kit	0.5 ng	14	572 ng	572 ng
			2 ng	11	388 ng	388 ng
			10 ng	9	577 ng	577 ng
	gDNA	ThruPLEX DNA-seq Kit	10 ng	7	449 ng	449 ng
		SureSelectXT Library Prep Kit	200 ng	10	2430 ng	750 ng
SureSelectXT2 Reagent Kit SureSelectXT2 Human All Exon V5	cfDNA	ThruPLEX DNA-seq Kit	0.5 ng	14	610 ng	610 ng
			2 ng	11	343 ng	343 ng
			10 ng	9	430 ng	430 ng
	gDNA	ThruPLEX DNA-seq Kit	10 ng	7	268 ng	268 ng
		SureSelectXT2 Library Prep Kit	100 ng	8	1366 ng	375 ng
SureSelectQXT Reagent Kit SureSelectXT Human All Exon V5	cfDNA	ThruPLEX DNA-seq Kit	0.5 ng	14	837 ng	837 ng
			2 ng	11	396 ng	396 ng
			10 ng	9	496 ng	496 ng
	gDNA	ThruPLEX DNA-seq Kit	10 ng	7	365 ng	365 ng

Figure 2. Experimental design. *For each SureSelect Target Enrichment System, libraries were prepared from cfDNA or gDNA using 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.*

	Additional Reagents Required	Omitted Reagents
SureSelectXT 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
SureSelectXT2 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

Figure 3. SureSelect compatibility. *List of reagents used when integrating ThruPLEX DNA-seq Kit with SureSelect Target Enrichment Systems.*

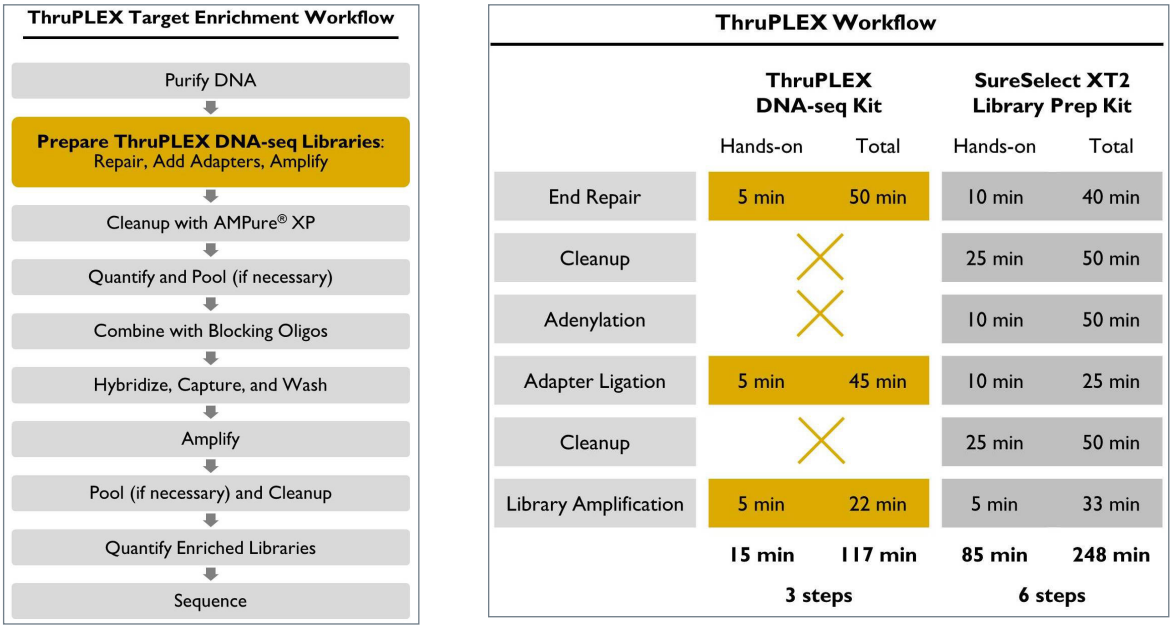

Figure 4. ThruPLEX workflow. Left: *General workflow diagram integrating ThruPLEX DNA-seq Kit with Agilent SureSelect platforms.* **Right:** *ThruPLEX DNA-seq Kit converts double-stranded DNA samples to indexed libraries in 3 simple steps: end repair, adapter ligation, and high-fidelity library amplification. The streamlined workflow prevents sample loss and handling errors and is performed in a single tube or well in less than 2 hours.*

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[4], to generate BAM files for each sample. BAM files were down-sampled to obtain equal numbers of reads, and duplicates were marked using Picard Mark Duplicates.[5] Output files from Picard Mark Duplicates were used to determine quality metrics related to the whole exome capture and sequencing using Picard CalculateHsMetrics.

Results & Discussion

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 ThruPLEX DNA-seq Kit with Agilent SureSelectXT, XT2, and QXT Target Enrichment Systems (**Figure 4**). As a reference, sheared gDNA was also enriched and sequenced following library preparation with the SureSelectXT, XT2, and QXT Library Prep Kits. The key sequencing metrics are summarized in **Figure 5**.

Results from whole exome enrichment of the ThruPLEX DNA-seq cfDNA libraries with each of the SureSelect platforms were comparable to those of the ThruPLEX DNA-seq gDNA libraries in terms of key sequencing metrics **(Figure 5)** and on-target specificity **(Figure 6).** These data demonstrate the exceptional repair capacity of ThruPLEX technology. When comparing the number of enriched (on-bait plus near-bait) bases, ThruPLEX DNA-seq cfDNA libraries were 99% (QXT), 88% (XT2), and 79% (XT) efficient relative to the gDNA libraries **(Figure 6)**. 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
SureSelectXT	ThruPLEX DNA-seq	cfDNA	500 pg	1,396,048	31.6	3.65E+06	4.67%
			2 ng	1,444,625	34.0	1.51E+07	1.32%
			10 ng	1,458,560	31.5	4.57E+07	0.40%
		gDNA	10 ng	1,450,470	37.9	3.50E+07	0.78%
	SureSelectXT	gDNA	200 ng	1,465,623	34.5	2.11E+08	0.12%
SureSelectXT2	ThruPLEX DNA-seq	cfDNA	500 pg	619,266	32.1	7.36E+06	1.47%
			2 ng	621,686	32.9	1.22E+07	1.04%
			10 ng	624,013	32.0	3.77E+07	0.31%
		gDNA	10 ng	622,495	34.0	2.64E+07	0.38%
	SureSelectXT2	gDNA	100 ng	625,674	29.5	8.05E+07	0.55%
SureSelectQXT	ThruPLEX DNA-seq	cfDNA	500 pg	1,052,819	44.9	2.62E+06	8.53%
			2 ng	1,122,128	45.3	9.12E+06	3.30%
			10 ng	1,142,897	45.3	3.36E+07	0.82%
		gDNA	10 ng	1,139,272	43.1	2.38E+07	1.06%

Figure 5. High quality exome enriched libraries. *Summary of sequencing metrics from whole exome sequencing of cfDNA and gDNA libraries prepared using ThruPLEX DNA-seq Kit or SureSelect Library Prep Kits and enriched with SureSelectXT, XT2, and QXT target enrichment systems.*

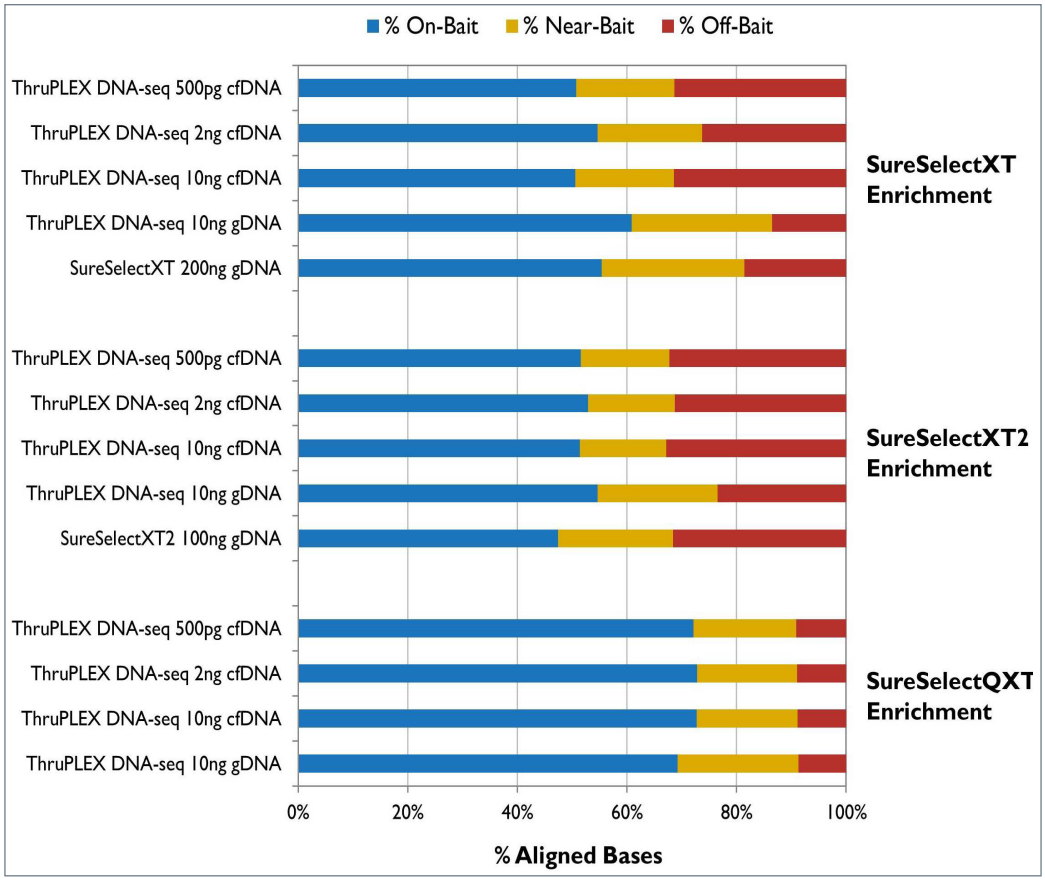

Figure 6. High quality exome data. *ThruPLEX DNA-seq cfDNA libraries demonstrated high performance when enriched using SureSelectXT, XT2, and QXT Target Enrichment Systems. ThruPLEX DNA-seq cfDNA libraries yielded performance similar to SureSelect gDNA libraries at low input amounts that are beyond the reach of the SureSelect Library Prep Kits. ThruPLEX DNA-seq gDNA libraries (10 ng input) generated higher percentage of on-bait bases when compared to gDNA libraries prepared from SureSelectXT (200 ng input) and XT2 (100 ng input) Library Prep Kits.*

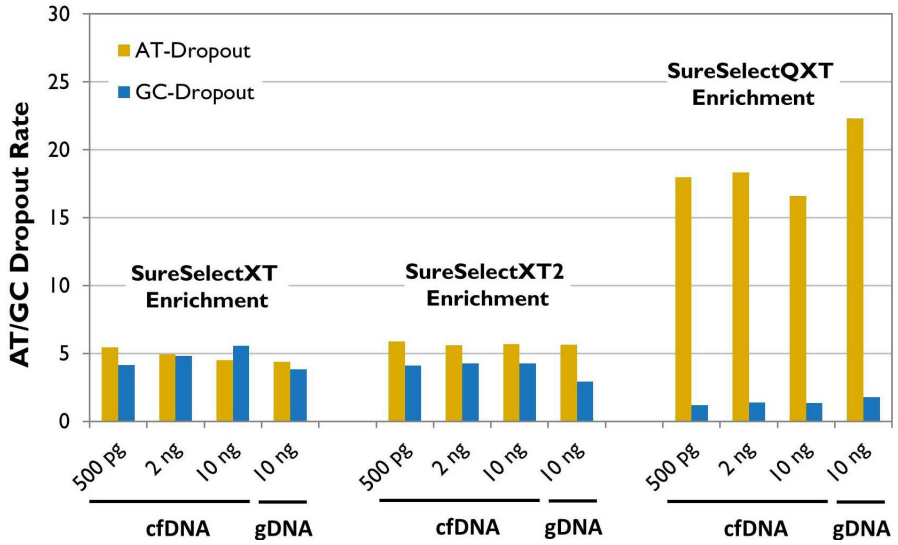

Figure 7. AT/GC dropout. *AT- and GC-dropout rates of ThruPLEX DNA-seq gDNA and cfDNA libraries enriched with SureSelectXT, XT2, and QXT Target Enrichment Systems. SureSelectQXT enriched libraries suffered from high AT-dropout rates. Approximately 15 to 20% of total reads that should have mapped to GC≥ 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.*

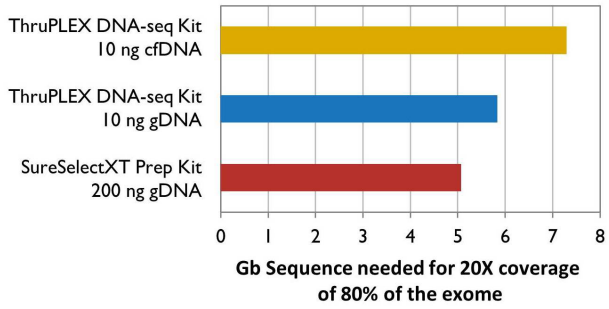

Figure 8. Sequencing requirements. *Estimated amount of sequencing required to achieve 20X coverage of 80% of the exome. Samples were prepared with ThruPLEX DNA-seq Kit or SureSelectXT Library Prep Kit, exome-enriched using the SureSelectXT platform, and sequenced on an Illumina NextSeq 500 using a 2 x 75bp run.*

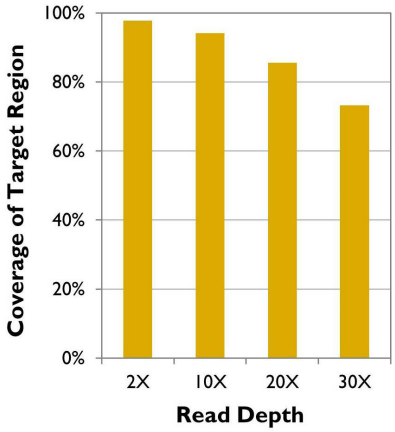

Figure 9. Exome coverage. *ThruPLEX DNA-seq libraries were prepared from 10 ng of cfDNA and exome enriched using the SureSelectXT platform. Up to 8 libraries may be sequenced on a NextSeq 500 high output 2x75 base run, yielding up to 800 M reads to achieve at least 80% coverage of the exome at 20X or higher.*

In general, the SureSelectXT2 platform showed similar performance compared to SureSelectXT. In the SureSelectXT2 workflow, samples are pooled prior to hybridization, which confers ease of use and cost advantages. ThruPLEX DNA-seq Kit can be integrated very conveniently with the SureSelectXT2 platform, requiring only minor adjustments to the protocol and additional universal blocking oligos from IDT **(Figure 3)**. The SureSelectQXT platform provided higher mean target coverage and required the shortest hybridization time. Enrichment with SureSelectQXT also appeared to be more efficient despite variable input amounts. However, SureSelectQXT resulted in much higher AT-dropout rates **(Figure 7)**, which may be the consequence of the temperature cycling during hybridization used in its protocol.

Deep sequencing data was also generated using an Illumina NextSeq 500. ThruPLEX DNA-seq gDNA/ SureSelect XT libraries at 10 ng input required less than 1 Gb of additional sequencing data than the SureSelectXT gDNA library at 200 ng input to yield 20X coverage of at least 80% of the exome **(Figure 8)**. 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 ThruPLEX DNA-seq/SureSelectXT library, 100M total 75-base reads per sample were adequate for SNV calling of at least 85% of the exome **(Figure 9)**. From the NextSeq 500 high output run (2x75), up to 8 cfDNA libraries prepared from 10 ng input could be sequenced to achieve at least 80% coverage of the exome.

Conclusion

By integrating the ThruPLEX DNA-seq Kit with Agilent SureSelect Target Enrichment Systems, we were able to exploit 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 ThruPLEX DNA-seq Kit with SureSelect platforms can be easily attained with minor adjustments to the SureSelect protocols and with addition of universal blocking oligos and sequencing primers. The SureSelectXT2 platform, in which samples are pooled prior to hybridization, is the simplest to integrate. In addition to its higher sensitivity and excellent performance, ThruPLEX DNA-seq Kit offers a faster and simpler workflow with a single-tube, three-step protocol. An integrated enrichment method combining ThruPLEX and SureSelect technologies will be instrumental in translational genomic research where the DNA of interest is present in limiting quantities.

For more information on target enrichment using ThruPLEX libraries visit: http://rubicongenomics.com/applications/request-application-notes/

References

1. Williamson, R. Properties of rapidly labelled deoxyribonucleic acid fragments isolated from the cytoplasm of primary cultures of embryonic mouse liver cells. *J. Mol. Biol.* **1970**, *51*, 157-168.
2. Murtaza, M. et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature*. **2013**, *497*, 108–112.
3. Klevebring, D. et al. Evaluation of exome sequencing to estimate tumor burden in plasma. *PLoS ONE*. **2014**, *9 (8)*, e104417.
4. Li, H. and Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*. **2009**, *25*, 1754-1760.
5. Picard - A set of tools (in Java) for working with next generation sequencing data in the BAM format. http://broadinstitute.github.io/picard.

Trademarks

RUBICON GENOMICS 4743 VENTURE DRIVE, ANN ARBOR, MI 48108 **T +1.734.677.4845** F +1.734.477.9902 RUBICONGENOMICS.