Application Note

ThruPLEX® Plasma-seq Kit: Optimized Library Preparation for Whole Genome and Targeted Sequencing of Cell-Free DNA from Plasma

Kamran Shazand, Komal Kunder & John Paul Jerome, Rubicon Genomics, Inc.

Introduction

Cell-free DNA (cfDNA), circulating in blood and found in the plasma component, was first discovered in the 1940s[1] and has been the subject of renewed attention in the research community due to easy access to genetic information. The main source of cfDNA is the apoptotic turnover of hematopoietic cells. DNA fragments are generated by the apoptotic endonuclease, Caspase-Activated DNase (CAD) digesting the chromosomal DNA at regular distances in their nucleosomal arrangement around histones, leading to fragments of various sizes. The cfDNA of primary interest exists as fragments of about 170 bp in length.

This genetic information is being is used by translational scientists to better understand the progression of cancer. Circulating tumor DNA (ctDNA) derived from malignant tumors is found as a component of the cfDNA; libraries prepared from these samples contain genetic information of the tumor.[2-3] For example, Murtaza and colleagues at CRUK performed a research study in which cfDNA libraries were prepared following therapy and the genetic evolution of several metastatic cancers was followed.

One of the major limitations of utilizing next-generation sequencing (NGS) with cfDNA is the difficulty of making sensitive libraries from the relatively low abundance of cfDNA obtained from plasma. Concentrations of cfDNA are quite variable, ranging from 1 ng to 20 ng per mL of plasma and the component of interest is fractionally represented. Rubicon Genomics' ThruPLEX® technology which has a history of use in low input library preparation[4-5], has been reformulated and optimized specifically for cfDNA for input levels starting at less than 1 ng and ranging to over 30 ng. This new member of the ThruPLEX product line, ThruPLEX Plasma-seq Kit, is capable of transforming cfDNA into high complexity libraries for Illumina® NGS platforms. The three-step, single-tube workflow yields indexed libraries from purified cfDNA within 2 hours.

In the present study, we demonstrate the performance and reproducibility of ThruPLEX Plasma-seq Kit in comparison to KAPA® Hyper Prep Kit and NEBNext® Ultra™ DNA Library Prep Kit. Further we show enrichment data which provides a richer view of the genetic variation within the sample.

Methods

Plasma Sample Preparation: Plasma collection was performed by Medical Research Networx, LLC. Blood was collected into BD Vacutainer® EDTA tubes and inverted 10 times to mix. Vacutainer tubes were centrifuged (4°C; 12 minutes; 1500xg) with the centrifuge brake off. The plasma layer was then removed, taking care not to disturb the buffy coat and placed into a 15 mL conical tube. The samples were then centrifuged again (4°C; 12 minutes; 1500xg) before transferring the plasma to a new tube, leaving approximately 0.5 mL to minimize leucocyte carry over. Processed plasma samples were stored at -80°C until DNA was extracted.

Cell-Free DNA Isolation: Qiagen QIAamp® Circulating Nucleic Acid Kit was used according to the manufacturer's protocol without the use of carrier RNA to isolate cfDNA from 5 mL aliquots of plasma samples.

DNA Quality Control and Quantification: Extracted cfDNA eluates from the same individual (15 mL plasma) were pooled, and the quality of these samples was evaluated on an Agilent BioAnalyzer®. The concentration of these samples was measured using Qubit® (ThermoFisher).

Library Preparation: Libraries were prepared from the cfDNA samples following manufacturer instructions using the ThruPLEX Plasma-seq Kit with dual indexes, NEBNext Ultra DNA Library Prep Kit (New England Biolabs) with dual indexes, and KAPA Hyper Prep Kit (KAPA Biosystems) with Roche Nimblegen® SeqCap® EZ adapters diluted to concentrations as recommended in the KAPA protocol for different input amounts. Amplified libraries were pooled and then purified using AMPure® XP beads (Beckman Coulter) and eluted in 30 µL low TE buffer for whole genome sequencing (WGS) or 50 µL ultra-pure water for enrichment. Purified libraries were assessed on the BioAnalyzer and quantified by qPCR using KAPA Library Quantification Kit for Bio-Rad (KAPA Biosystems).
Two WGS experiments and a kinome enrichment were performed (see Table 1). For the first, libraries were prepared from 3 individual plasma samples at input amounts of 0.1 ng, 1 ng, and 30 ng. The amount of mononucleosomal DNA in each sample, as measured by the Bioanalyzer, corresponded to 0.09 ng, 0.62 ng, and 15.44 ng. In the second WGS experiment, 9 individual plasma samples were tested. Cell-free DNA from the equivalent of 1 mL aliquot of plasma sample was used to prepare each library; input amounts ranged from 5 ng to 40 ng. For the kinome sequencing experiment, 2 individual plasma samples were used at input amounts of 6.5 ng and 10 ng in triplicate.

Enrichment: Hybridization and capture of the indexed libraries were carried out using the SureSelect®XT2 ClearSeq® Human DNA Kinome Panel. Briefly, six indexed ThruPLEX Plasma-seq libraries, hybridization buffer mix, blocking mix, RNase block, and the ClearSeq Kinome Panel were combined according to the SureSelectXT2 protocol. In addition, 1 µL (1 nmole) each of i5 and i7 xGen® Universal Blocking Oligo - TS HT (Integrated DNA Technologies) were added into the hybridization reaction which was carried out for 48 hours. Target capture, washes, and final amplification of the enriched libraries were performed according to the SureSelectXT2 protocol to obtain captured libraries ready for Illumina sequencing.

Illumina Sequencing: Pooled libraries were quantified using KAPA Library Quantification Kit and loaded onto an Illumina MiSeq® or NextSeq® 500 flow cell for sequencing. Approximately 17M to 25M reads per library were collected for whole genome sequencing and 5M reads per library for kinome sequencing.

Data Analysis: Sequences were analyzed on the DNANexus platform. Reads were aligned to the human genome, hg19, using Burrows-Wheeler Algorithm, BWA-MEM[6], to generate BAM files. For WGS data, reads were first down-sampled to equal numbers across all samples. Down-sampled BAM files were assessed using Picard Mark Duplicates[7] to count duplicate reads and estimate diversity (Estimated Library Size), and Picard Collect GC Metrics was used to determine biases based on sequence GC content. For kinome sequencing data, after mapping with BWA-MEM, Picard CalculateHsMetrics was used to determine capture quality metrics. For SNV analysis, Agilent SureCall was used to identify variants within the targeted exons of the kinome, and Illumina Variant Caller was used to annotate variants.

	Whole Genome Sequencing			Whole Genome Sequencing	Kinome Sequencing	
Samples	Sample 1	Sample 2	Sample 3	Samples 4 to 12	Sample A	Sample B
Input	0.1 ng	1 ng	30 ng	cfDNA from 1 mL of plasma (5 ng to 40 ng)	6.5 ng	10 ng

Table 1. *Plasma samples and input DNA amount: In the first whole genome sequencing (WGS) experiment, three individual plasma samples were used to construct ThruPLEX Plasma-seq libraries at the indicated input amounts. A second WGS experiment used 9 individual plasma samples in triplicate. Two separate plasma samples were used for kinome sequencing.*

Results & Discussion

Preparation of cfDNA Libraries

There are several library preparation kits for Illumina NGS platforms available, but none have been designed specifically for cell-free DNA. ThruPLEX Plasma-seq Kit is able to create highly reproducible libraries over a wide input range of cfDNA, from less than 1 ng to 30 ng. Preparation of cfDNA for NGS has usually been done by home-brew kits or kits initially designed to work with mechanically sheared gDNA of 200 – 600 bp in size. Many kits, including the Illumina TruSeq® Nano kit require a minimum starting amount of 100 ng of DNA. Kits that employ enzymatic fragmentation such as Nextera® DNA Library Prep Kit or KAPA Hyper Plus are not compatible with this type of sample due to the small initial size of cfDNA. In fact, shearing of the cfDNA is unnecessary. The two kits that were selected for the current test are able to create libraries from as little as 1 ng (KAPA Hyper Prep Kit) or 5 ng (NEBNext Ultra DNA Library Prep Kit). The ThruPLEX Plasma-seq Kit is the only kit designed and optimized to efficiently and reproducibly repair, ligate, and amplify NGS libraries from cfDNA. Key to this efficiency and reproducibility with DNA fragmented as a result of apoptosis is the use of stem-loop adaptors to make libraries thus eliminating clean-up steps and background problems caused by y-adaptors. ThruPLEX Plasma-seq also offers advantages in the workflow over all of the alternative kits (Figure 1). Starting with the isolated cfDNA, ThruPLEX creates indexed libraries in three steps in a single tube in about two hours. No sample transfers or intermediate clean-ups are necessary. All components including adapters and indexing reagents are provided with the kit and no optimization is required. Both KAPA Hyper and the NEBNext Ultra have intermediate clean-up steps; both require the purchase of adapters and/or indexing oligonucleotides that often require optimization of concentration to control the amount of adaptor dimers and other artifacts. Additionally, for low input amounts of DNA (< 25 ng), KAPA recommends optimizing the adapter concentration.

	ThruPLEX Plasma-seq	NEBNext Ultra	KAPA Hyper
Indexing reagents included?	Yes	No	No
Indexes available	Up to 48 single and 96 dual indexes	Up to 24 single and 96 dual indexes	None
Recommended input range	1 ng – 30 ng	5 ng – 1000 ng	1 ng – 1000 ng
Total steps	3	4	4
Workflow	1. End Repair	1. End Repair	1. End Repair & A-Tailing
	2. Adapter Ligation	2. Adapter Ligation	2. Adapter Ligation
	↓	3. Cleanup	3. Cleanup
	3. Library Amplification	4. Library Amplification	4. Library Amplification
Total hands-on time	15 min	50 min	50 min
Total kit time	~2 hours	~3 hours	~2.7 hours
Sample transfer steps	0	1	1

Figure 1. ThruPLEX Plasma-seq workflow and advantages. *ThruPLEX Plasma-seq Kit, which includes optimized adapters and indexing reagents, converts cfDNA from plasma samples to indexed NGS libraries in 3 simple steps in a single tube or well in about 2 hours; no sample transfer or clean-up steps are required.*

Comparative Analysis

Libraries created with each of these products were compared for a number of metrics including library diversity, duplicate reads, and unmapped reads. GC bias was also compared across the kits.

ThruPLEX Plasma-seq yielded significantly higher library diversity while, conversely, very low duplicate rates were detected in low pass sequencing analysis (Figure 2). The duplication rate for ThruPLEX was significantly lower than the other kits, indicating that with deeper sequencing runs ThruPLEX will provide more usable data. ThruPLEX also had the fewest unmapped reads. These metrics all indicate that ThruPLEX Plasma-seq will provide more usable data.

In GC bias analysis (Figure 3), ThruPLEX Plasma-seq showed well balanced coverage of the genome between 20 and 70% of GC content. Identical samples were used to prepare libraries with KAPA Hyper, and there was a lack of coverage in the AT rich region. A separate set of four samples was used to generate libraries for NEBNext Ultra and those, too, lacked the AT coverage. Since the human genome has an average GC content around 42%, libraries prepared from ThruPLEX Plasma-seq best represent the original genetic content of the sample.

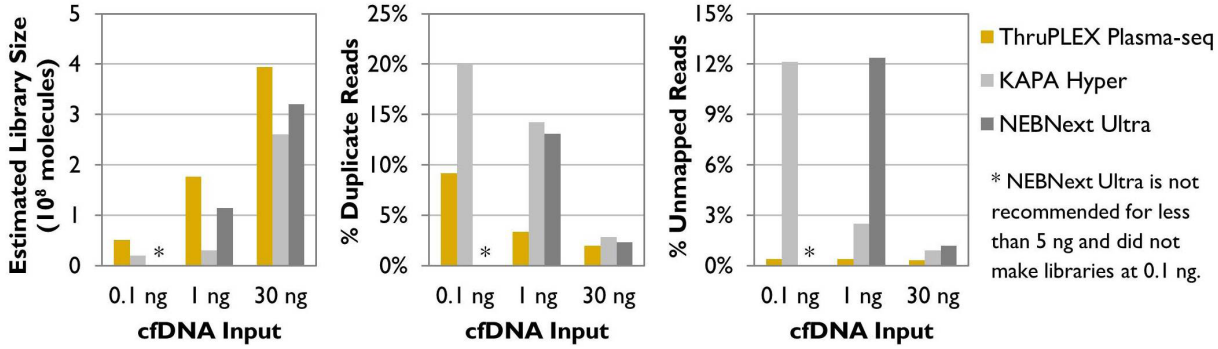

Figure 2. Diverse, reproducible NGS libraries. *Libraries created with ThruPLEX Plasma-seq yielded more unique molecules (left), fewer duplicate reads (middle), and negligible unmapped reads (right). Libraries were sequenced on an Illumina NextSeq 500 as a paired-end run with 17M to 25M reads per library. Duplication rates were calculated after down-sampling the data to 17M reads per library. Representative data from each sample is shown.*

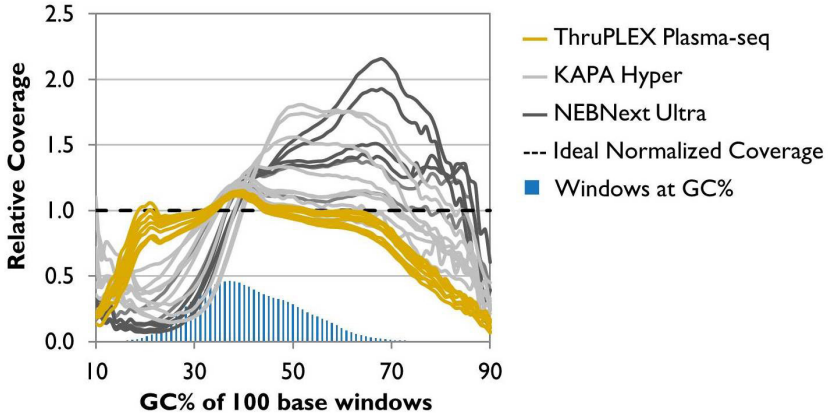

Figure 3. Reproducible, Unbiased GC Coverage. *ThruPLEX Plasma-seq provided the most reproducible and unbiased GC coverage across the human genome, showing minimal variability across the 9 plasma samples tested. Libraries were prepared from cell-free DNA isolated from an equivalent of 1 mL of plasma sample and sequenced on an Illumina NextSeq 500. Four separate plasma samples were used to construct the NEBNext Ultra libraries.*

Enrichment Performance

To better evaluate the performance of ThruPLEX Plasma-seq, libraries were enriched using the Agilent SureSelect XT2 ClearSeq Human DNA Kinome probe set (CAT. NO. 5190-4676) according to the Rubicon Genomics protocol (RDM-153-002) in presence of the Universal xGen Blocking Oligos (IDT). Based on approximately 5M total reads for each sample (Table 2), a 600-fold enrichment of the human kinome (panel size 3.2 Mbp) was obtained. At 30x coverage, an average 77% of bases were covered for the cfDNA samples used in this experiment (Figure 4). Using this data, a highly concordant SNV rate between the replicates for any given sample was found, supporting the ability of ThruPLEX Plasma-seq create libraries that can be used to identify novel allele variants with high efficiency. The identity of the variant calls was confirmed

by identifying 98 - 99% of SNPs in the dbSNP database (Table 2). The other 1 - 2% were novel calls that were generally common to all three replicates of each plasma DNA sample, supporting their biological validity.

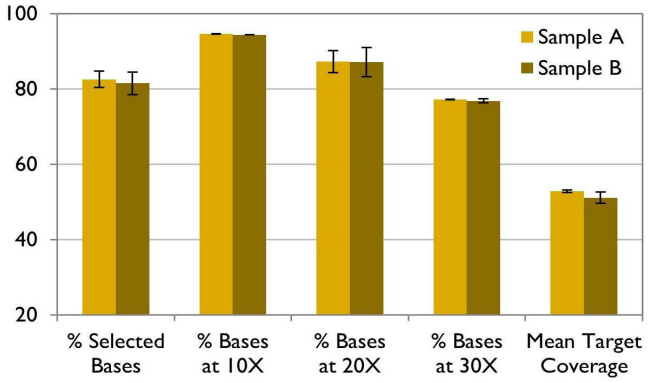

Figure 4. Targeted sequencing metrics of ThruPLEX Plasma-seq libraries enriched using the SureSelectXT2 ClearSeq Human DNA Kinome Panel. *Libraries were sequenced on an Illumina MiSeq, generating ~5M reads per library. Selected bases are identified as successfully captured bases that were in or within 250 bp of the baits.*

	Sample A		
	Replicate 1	Replicate 2	Replicate 3
Total Reads	4,729,478	4,991,598	4,859,650
Total High Quality Uniquely Mapped Reads	3,309,675	3,645,999	3,392,824
Fold Enrichment	645	606	642
Total Number of Variants Identified	1750	1792	1793
Percent of SNPs in dbSNP Database	98.9%	98.8%	98.8%

Table 2. *One ThruPLEX Plasma-seq library that was used in kinome capture (Figure 4) was further analyzed for SNP coverage. Results above indicate the number of variants captured and percent of SNPs identified in the dbSNP database are sufficient to allow mutation detection. Libraries were prepared in triplicate from plasma sample A, enriched using the SureSelectXT2 ClearSeq Human DNA Kinome Panel, and sequenced on an Illumina MiSeq.*

Conclusion

ThruPLEX Plasma-seq was specifically developed to produce high quality libraries from cell-free DNA. Both the repair and ligation reactions have been reformulated to provide superior results with cfDNA. The optimized repair reaction ensures that the ends of each fragment are blunt and polished to provide high ligation efficiency. Likewise, the ligation reaction has been enhanced for cfDNA molecules to provide maximum ligation of the stem-loop adaptor. The elimination of an intermediate clean-up step as well as the lack of transfer steps minimizes loss of molecules, augmenting the formulation changes to provide this cfDNA-specific product. Our data clearly indicate that ThruPLEX Plasma-seq yields better libraries in comparison to its competitors, in terms of diversity, GC bias, and duplicate rates. These libraries are suitable for targeted enrichment and will provide a sensitive tool to allow scientists to easily access and analyze the genetic content of samples from a variety of experimental conditions.

References

1. Mandel P, Metais P. Les acides nucleiques du plasma sanguin chez l'homme. C. R. Seances Soc. Biol. Fil. 1948;142: 241–243.
2. Shaw JA, Stebbing J. Circulating free DNA in the management of breast cancer. Annals of Translational Medicine. 2014;2(1):3.
3. Patel KM, Tsui DW. The translational potential of circulating tumour DNA in oncology. Clin Biochem. 2015 Apr 15.
4. Murtaza et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA.Nature. 2013 May 2;497(7447):108-12.
5. Kitzman et al. Noninvasive whole-genome sequencing of a human fetus. Sci Transl Med. 2012 Jun 6;4(137):137ra76.
6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics. 2009 Jul 15;25(14):1754-60.
7. Picard - A set of tools (in Java) for working with next generation sequencing data in the BAM format. http://broadinstitute.github.io/picard

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

Trademarks

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