Efficient generation of cfDNA libraries that are highly enriched in short fragments

Chris Boles[1], Anthony Popkie[2], Sadaf Hoda[1], Egon Ranghini[2] and Kamran Shazand[2]

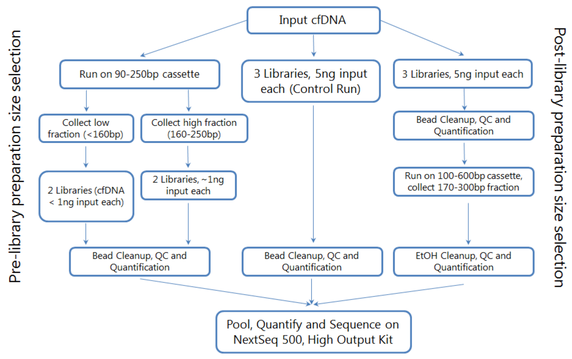
[1]Sage Science, Beverly, MA; [2]Rubicon Genomics, Ann Arbor, MI

Introduction

Cell-free DNA (cfDNA) isolated from blood plasma allows unprecedented non-invasive access to valuable genetic information. The most abundant form of cfDNA shows a characteristic peak of approximately 170bp, the size of the mononucleosomal unit. Increasing evidence suggest that shorter cfDNA fragments (40-170bp) retain genetic information about the fetus, tumor-associated copy number aberrations and the cellular origin of cfDNA (Jiang et al., PNAS, 2015; Snyder et al., Cell, 2016). Next generation sequencing (NGS) is a key sensitive and accurate technology that allows the investigation of the genetic content of cfDNA isolated from liquid biopsies, and already presents numerous applications both in research and precision medicine.

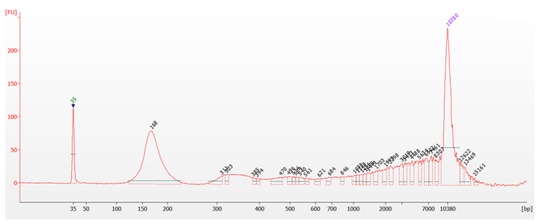
Rubicon Genomics' ThruPLEX® Plasma-Seq kit is specifically designed for sequencing low input cfDNA samples. We used the Pippin Prep system (Sage Science) to perform size selection of cfDNA before ThruPLEX library preparation. Briefly, we fractionated cfDNA from human plasma into <160bp and >160bp fragments on the Pippin Prep, and generated ThruPLEX Plasma-Seq libraries with less than 1ng of size selected input material. As control, ThruPLEX Plasma-Seq libraries were prepared without prior size selection of cfDNA.

Experimental Plan



Size selection conditions on Pippin Prep

Input cfDNA



Bioanalyzer electropherogram of input cfDNA, isolated from 50ml of human plasma (Bioreclamation IVT). The trace shows the typical mononucleosomal peak at approximately 170bp. The electropherogram also reveals the presence of the di-nucleosomal and tri-nucleosomal peaks at approximately 340bp and 510bp, respectively. A small amount of high molecular weight DNA is also detected, however is not incorporated into ThruPLEX libraries. Bioanalyzer markers have a size of 35bp and 10,380bp.

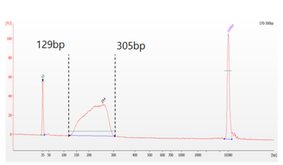
Pippin Prep™

Pre- and Post-Library Size Selection Optimization

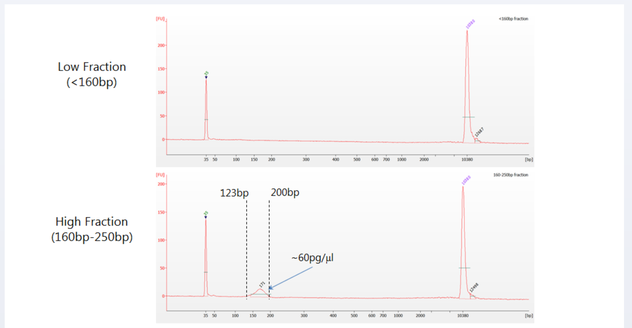
- Protocols optimized using gDNA

3% Agarose Gel, gDNA sheared at 200bp			
Fraction	Start Elution	Pause	End Elution
<160bp	00:45:00	01:15:00	
160bp-250bp		01:15:00	01:45:00

2% Agarose Gel, gDNA sheared at 300bp		
Fraction	Start Elution	End Elution
170bp-300bp	00:53:00	01:05:00



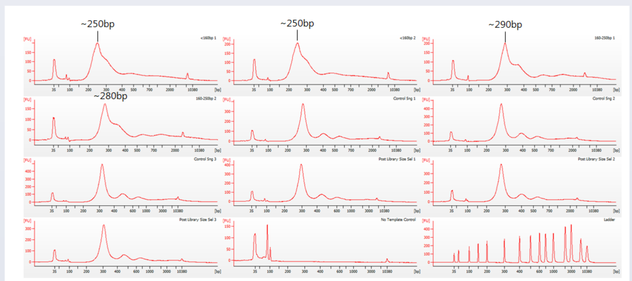
Analyses of Pre-Library Size Selection input cfDNA



Pre-Library Preparation Size Selection

Bioanalyzer electropherograms of small (<160bp) and large (160bp-250bp) fractions. The concentration of the small fraction was too low to be quantified by peak integration.

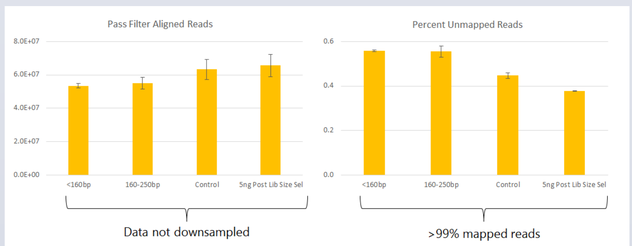
Analyses of Amplified ThruPLEX Libraries



Analyses of Amplified Libraries

Bioanalyzer analysis was performed before cleaning the libraries with Agencourt AMPure XP beads (Beckman Coulter), hence the presence of minor fragments around 100bp.
Libraries prepared with the <160bp fraction show a peak maximum about 30-50bp smaller than libraries prepared with the 160-250bp fraction.

Sequencing Library Quality



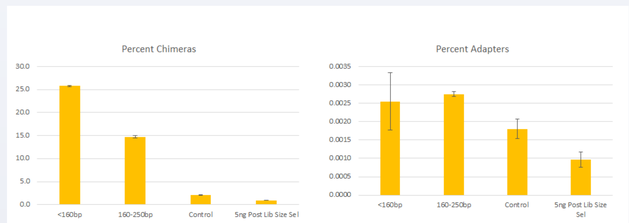
Pass Filter Reads & Percent Unmapped Reads

Libraries were downsampled to 5.25 * 10^7 pass filter reads and aligned for downstream analyses.
More than 99% of reads generated by ThruPLEX Plasma-Seq mapped to the reference genome.

Estimated Library Size & Percent Duplication

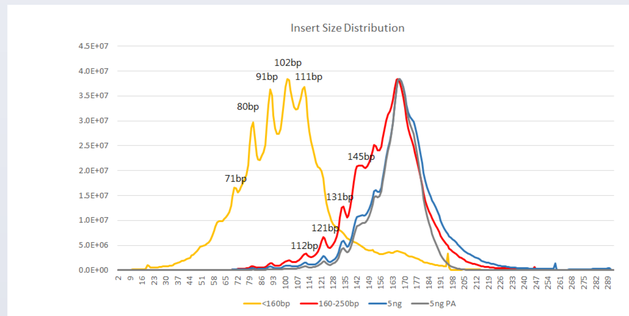
All libraries achieved a complexity between 10^8 and 10^9 unique molecules, which is typical for libraries prepared with cfDNA. Importantly, even when preparing libraries with the smallest fractions, library complexity exceeded 10^8 unique molecules, despite the extremely low input amount of cfDNA. Even with very low input amounts of cfDNA, libraries generated with the shortest fractions showed less than 9% duplicate reads.

Sequencing Library Quality (cont.)



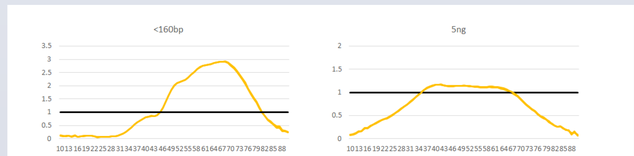
Percent Chimeras & Percent Adapter Reads

The number of adapter reads was extremely low across all conditions tested.



Insert Size Distribution

Insert size distributions normalized to the total number of reads for libraries generated with 5ng of cfDNA (Control Run). Libraries generated using the <160bp fractions were enriched in cfDNA fragments whose length ranged between ~60bp and ~ 120bp.



GC Coverage

Libraries generated with 5ng of cfDNA show uniform GC coverage between 30% and 70% GC. The curves to the left are considerably shifted toward an over representation of GC dinucleotides.

Conclusions

Our results showed that the ThruPLEX Plasma-Seq technology generated highly diverse libraries whose complexity exceeded 10^8 unique molecules. Importantly, using the Pippin Prep system to select cfDNA fragments shorter than 170bp allowed us to generate libraries that were strongly enriched in cfDNA fragments ranging between 60bp and 120bp in length. These fragments showed a characteristic 10bp periodicity in the 70bp – 110bp range. Significant enrichment of small cfDNA fragments was only seen when size selection was performed prior to library construction -- performing size selection post-library-construction did not provide enrichment in small fragments compared to control libraries using no size selection.

Thanks to the combination of Pippin's accurate fractionation capability and ThruPLEX Plasma-Seq's superior ability to make highly diverse libraries with sub-nanogram input amounts, we were able to enrich for the smaller fragments (<160bp) of cfDNA. This method improves access to the shorter fragments of cfDNA that generally are under-represented in a regular protocol due to their lower concentration.

Contact

Sage Science: chris.boles@sagescience.com
Rubicon Genomics: eranghini@rubicongenomics.com