1977

Size selection of cell-free DNA increases the proportion of tumor-specific variants in cancer patients

Kimberly A. Holden[1], Kerry D. Fitzgerald[1], Graham McLennan[1], Nathan Faulkner[1], Xiaojun Guan[1], Marcia Eisenberg[3], Taylor J. Jensen[1,2]

[1]Sequenom, Inc., a wholly owned subsidiary of Laboratory Corporation of America® Holdings, San Diego, CA; [2]Laboratory Corporation of America® Holdings, Durham, North Carolina; [3]Laboratory Corporation of America® Holdings, Burlington, North Carolina

Integrated ONCOLOGY
LabCorp Specialty Testing Group

I. Introduction

Cell-free DNA (cfDNA) derived from tumor cells is present in the plasma of patients with cancer; however, the proportion of this circulating tumor DNA (ctDNA) is often less than 2%. When the amount of ctDNA is present at such low levels, it presents detection challenges. One biological feature that may be able to be leveraged to improve this signal-to-noise challenge is the fragment length of cfDNA. The median length of cfDNA in circulation from healthy tissue is typically about 167bp, while ctDNA has been demonstrated to be, on average, shorter. We sought to evaluate the feasibility of enriching cfDNA for shorter fragments and whether this could enhance signal for the detection of tumor-specific variants.

II. Methods

SAMPLE COLLECTION AND PROCESSING: Whole blood (~10 mL) was collected from patients with or without known cancer in Streck BCT tubes (Streck, Omaha, NE) and processed to plasma as previously described.[1]

cfDNA EXTRACTION: cfDNA from the plasma of each sample was extracted using a bead-based method as previously described.[1]

SEQUENCING LIBRARY PREPARATION OF cfDNA SAMPLES: Libraries for genome-wide sequencing were created from cfDNA as previously described.[2]

SIZE SELECTION OF PREPARED LIBRARIES: Adapter-ligated libraries were size selected using the Coastal Genomics NIMBUS Select (Coastal Genomics, Inc, Burnaby, British Columbia), an automated platform for gel-based electrophoresis and size selection, targeting cfDNA fragment sizes up to 142bp (+/-15bp).

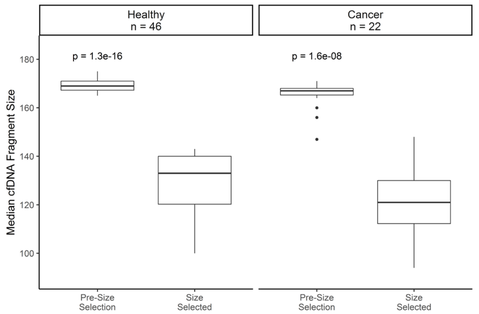
GENOME-WIDE NEXT GENERATION SEQUENCING OF cfDNA SAMPLES: Normalized sequencing libraries before and after size selection were pooled and sequenced using HiSeq2500 (Illumina) instruments as previously described.[3] A mean of 40.1 million sequencing reads were aligned to the human reference genome for each sample.

SEQUENCING DATA ANALYSIS: Sequencing data were processed as previously described.[3] Briefly, sequencing reads were mapped to the human reference genome (hg19) and partitioned into 50 kbp non-overlapping segments. Regions were selected and data were normalized as previously performed for noninvasive detection of fetal copy number alterations (CNAs)[4] and the resultant normalized values were used to calculate a genome instability number (GIN) as previously described.[5] The GIN is a metric intended to capture genome-wide autosomal deviation from empirically derived euploid dosage of the genome in circulation. The area under the curve (AUC) was calculated for each detected CNA to determine the deviation from the empirically derived euploid cfDNA genome. An AUC ratio was subsequently calculated for each detected CNA by dividing the AUC of a CNA after size selection by the AUC of the same CNA before size selection. Fragment length was obtained from paired-end sequencing data and is inferred as the number of nucleotides between the mapped positions of the first nucleotide position of the first and second read, respectively.

STATISTICAL ANALYSIS: Statistical analyses were carried out using custom scripts in an R programming language.[6]

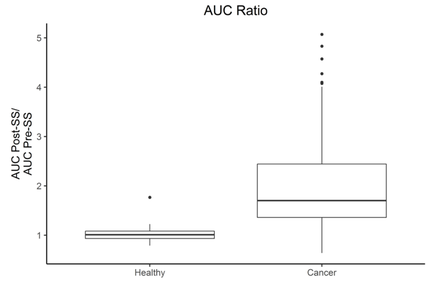
III. Results

Figure 1. Median cfDNA fragment size before and after size selection.



Libraries prior to size selection had an average median cfDNA fragment size of 169bp in healthy patients, and 165bp in cancer patients. Each box extends from the first quartile to the third quartile (25th and 75th percentiles). The upper whisker extends to the largest value no further than 1.5 times the inter-quartile range (IQR), and the lower whisker extends to the smallest value no further than 1.5 times the IQR. Individual points represent outliers to that distribution. After size selection, the average median cfDNA fragment sizes were significantly lower at 129bp and 120bp, respectively ($p=1.3e^{-16}$, $p=1.6e^{-8}$; Wilcoxon Rank Sum)

Figure 2. AUC ratio of CNAs before and after size selection.



An AUC ratio value=1 indicates no difference in CNA detection after size selection. An AUC value >1 indicates an increase in signal for the CNA after size selection. An AUC value <1 indicates a decrease in signal for the CNA after size selection. When evaluating cfDNA from healthy patients, the average AUC ratio was 1.04, consistent with a lack of signal enrichment in the absence of disease. Conversely, detectable CNAs in cancer patients had an average AUC ratio of 1.97, consistent with an enrichment of signal.

Figure 3. Enrichment of shorter cfDNA fragments (cutoff = 152bp) increases the ctDNA signal.

A size selection cutoff of 152bp is shown in this example.

A) The genome-wide profile of the sample before size selection. Each data point represents a 50kB segment of the human genome. Colors alternate by chromosome with the number of each chromosome noted at the top of the figure. Deviations for the empirically derived euploid genome (relative representation=1) above the line indicate an over-representation of cfDNA fragments from that region; values below indicate a loss.

B) The genome-wide profile of the sample after size selection where CNAs increased a small amount in magnitude and the GIN increased.

C) The cfDNA fragment size profile of the sample before and after size selection; the median fragment length after size selection was 148bp whereas the median before size selection was 167bp. Since the magnitude of the difference in fragment sizes is relatively small, there is still considerable overlap in the fragment size distribution between conditions for this sample.

D) Paired analysis of the absolute value of the AUC for a total of 30 CNAs detected before and after size selection. There was significant enrichment of CNAs after size selection, indicated by an average AUC ratio of 1.7 ($p=3.7e^{-9}$; Wilcoxon Rank Sum).

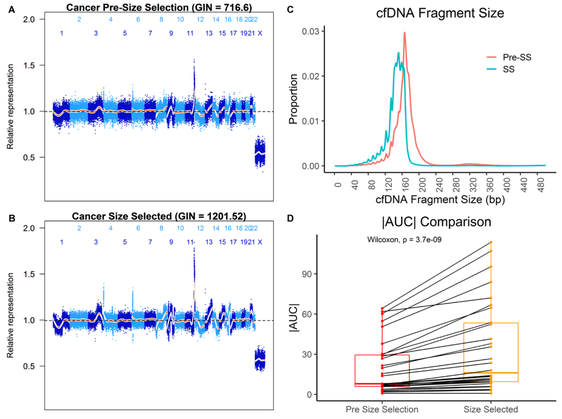


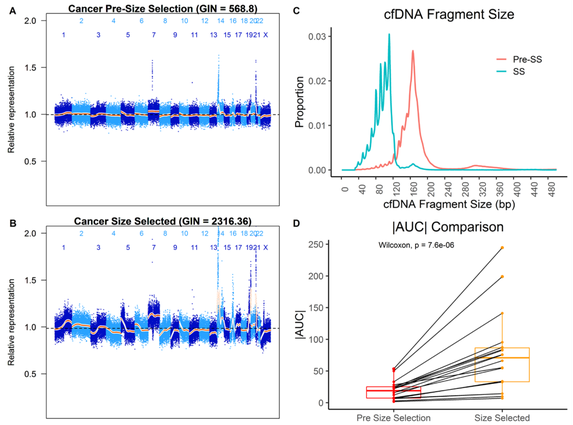
Figure 4. Enrichment of shorter cfDNA fragments (cutoff = 116bp) further increases the ctDNA signal.

A low size selection cutoff of 116bp is shown in this example.

A) The genome-wide profile of the sample before size selection. Each data point represents a 50kB segment of the human genome. Colors alternate by chromosome with the number of each chromosome noted at the top of the figure. Deviations for the empirically derived euploid genome (relative representation=1) above the line indicate an over-representation of cfDNA fragments from that region; values below indicate a loss.

B) The genome-wide profile of the sample after size selection where CNAs increased significantly in magnitude and the GIN increased significantly. Examples of CNAs can be seen on chromosome 7, where the entire chromosome is amplified after size selection. Additionally, three amplifications on chromosomes 7, 14, and 21 are so large after size selection that they have gone above the limits of this figure.

C) The cfDNA fragment size profile of the sample before and after size selection; the median fragment length after size selection was 99bp whereas the median before size selection was 166bp. Since the magnitude of the difference in fragment sizes is large, there is very little overlap in the fragment size distribution between conditions for this sample.



D) Paired analysis of the absolute value of the AUC for a total of 18 CNAs detected before and after size selection. There was significant enrichment of CNAs after size selection, indicated by an average AUC ratio of 3.7 ($p=7.6e^{-6}$; Wilcoxon Rank Sum)

Figure 5. Enrichment of shorter cfDNA fragments (cutoff = 152bp) does not affect the signal from probable germline CNAs.

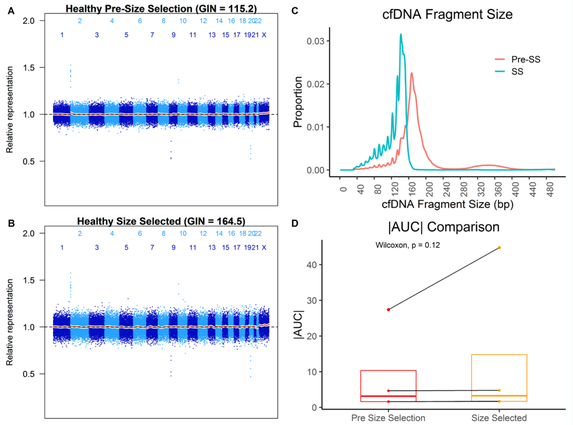
A healthy patient with probable germline CNAs is shown in this example.

A) The genome-wide profile of the sample before size selection. Each data point represents a 50kB segment of the human genome. Colors alternate by chromosome with the number of each chromosome noted at the top of the figure. Deviations for the empirically derived euploid genome (relative representation=1) above the line indicate an over-representation of cfDNA fragments from that region; values below indicate a loss.

B) The genome-wide profile of the sample after size selection.

C) The cfDNA fragment size profile of the sample before and after size selection; the median fragment length after size selection was 138bp whereas the median before size selection was 170bp.

D) Paired analysis of the absolute value of the AUC for a total of 4 CNAs detected before and after size selection. The AUC does not change significantly after size selection, indicated by an average AUC ratio of 1.19 ($p=0.12$; Wilcoxon Rank Sum).



IV. Conclusions

These data demonstrate a proof-of-concept for using size selection to enhance signal for the detection of tumor-specific variants in cancer patients.

V. References

1. Ellison, CK et al., Using Targeted Sequencing of Paralogous Sequences for Noninvasive Detection of Selected Fetal Aneuploidies. *Clin Chem*, 2016. 62(12): p. 1621-1629.
2. Tynan, JA et al., Application of risk score analysis to low-coverage whole genome sequencing data for the noninvasive detection of trisomy 21, trisomy 18, and trisomy 13. *Prenat Diagn*, 2016. 36(1): p. 56-62.
3. Lefkowitz, RB et al., Clinical validation of a noninvasive prenatal test for genomewide detection of fetal copy number variants. *Am J Obstet Gynecol*, 2016. 215(2): p. 227 e1-227 e16.
4. Zhao, C et al., Detection of fetal subchromosomal abnormalities by sequencing circulating cell-free DNA from maternal plasma. *Clin Chem*, 2015. 61(4): p. 608-16.
5. Jensen, T et al., Genome-wide sequencing of cell-free DNA identifies copy number alterations that can be used for monitoring response to immunotherapy in cancer patients. *Mol Cancer Ther* 2019. 18(2): p. 448-458.
6. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL: https://www.R-project.org/.

LabCorp

Poster presented at the *AACR Virtual Annual Meeting* © 2020 June 22-24.