The Circulating Cell-free Genome Atlas (CCGA) Study: Size Selection of Cell-free DNA (cfDNA) Fragments

ASCO 2019
May 31–June 4, 2019
Chicago, IL, USA

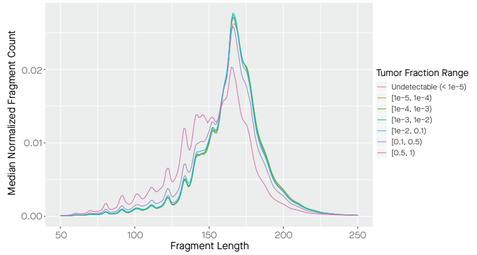
Darya Filippova, PhD; Matthew H. Larson, PhD; M. Cyrus Maher, PhD, MS, MPH; Robert Calef; Monica Pimentel, MS; Yiqi Zhou, MS; Joshua Newman; Sam Gross, PhD; Virgil Nicula; Ting-Chun Liu, MS; Christopher Yakym, MS; Jennifer Berman, PhD; Alexander M. Aravanis, MD, PhD; Arash Jamshidi, PhD

GRAIL, Inc., Menlo Park, CA

BACKGROUND

- Detection of somatic copy number aberrations in individuals with cancer via whole-genome sequencing (WGS) of cfDNA is challenging at low tumor fractions.
- Given that tumor-derived cfDNA fragments are, on average, shorter than those from healthy tissues (Figure 1),[1,2] we performed an exploratory analysis to evaluate the effect of size selection on cancer detection sensitivity.

Figure 1. Fragment Length Distribution by Estimated Tumor Fraction



Fragment length distributions of samples aggregated by estimated tumor fraction. Samples within a given tumor fraction range were grouped together, and the median normalized count at each fragment length from 50–250 bp was calculated across all samples within the group. A notable shift towards shorter fragment lengths was observed in the highest tumor fraction groups.

METHODS

- CCGA1 cfDNA samples were used for an *in silico* experiment; a subset of those that had corresponding WBC were also examined. The in vitro experiment used a distinctive set of CCGA1 samples also with matched tissue.

In Silico Experiment

- We used clinically evaluable CCGA1 training set cfDNA samples (n=1420: 560 non-cancer [NC], 860 cancer [C] stages 0–IV/non-informative) and a hold-out set of individuals under the age of 30 without a cancer diagnosis (referred to as the "calibration set", n=169) (**Table 1**), all of which passed assay QC. The set included only the solid cancer samples and cancers with single or multiple primaries. Two samples (1 C, 1 NC) were removed due to a sample swap. WGS libraries were produced for all samples as previously reported, achieving 36X depth.[3]
  - Briefly, informative features were extracted for the assay (eg, variation in read depth); machine learning classifiers were independently trained on each feature type to estimate probability of cancer and a classification model using 10-fold cross-validation was developed to avoid overfitting to the training data and to discriminate cancer from controls.[3]
- For every sample, including the calibration set, fragments between 90–150 bp were used for analysis, reducing sequencing depth for size-selected samples to 6.91±2.64X. For every *in silico* size-selected sample, we generated a depth-matched control that included all fragment lengths. The calibration set was then used to re-normalize the training set samples to remove assay effects for both size-selected and downsampled data.
- We estimated tumor fraction for both full-depth and size-selected samples using variants called in matched tissue.

In Vitro Experiment

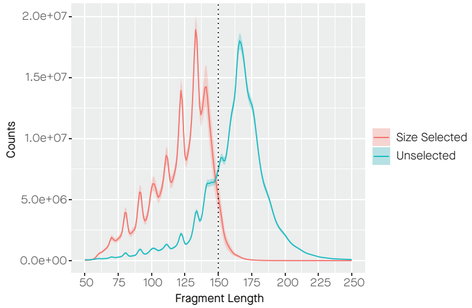
- We performed an *in vitro* experiment to analyze tumor fraction changes through physical size-selection using Pippin Prep (Sage Science).
- Size selection was performed after library preparation and amplification, providing several advantages over pre-amplification size-selection workflows:
  - More material is available for size selection after amplification, which reduces loss of rare tumor-derived fragments.
  - Adapter ligation increases fragment size by 150 bp, which is more separable by conventional size-selection methods, including Pippin Prep.
  - Size selection can be performed on pooled libraries after the incorporation of sample indices, making the process highly scalable.
- For each sample, 1 µg of WGS library was run on a 3% agarose gel cassette. The eluted library was diluted, pooled, and sequenced without further cleanup.
- The mapped fragment length of size-selected libraries showed a decrease in fragment size compared to libraries sequenced prior to size selection (**Figure 2**).

Table 1. Sample Counts for *In Silico* and *In Vitro* Experiments by Stage and Cancer Type

	In Silico		*In Vitro*
	Sample Count	Sample Count with Tissue[a]	Count
By Stage			
I	283	195	13
II	235	178	15
III	156	99	21
IV	157	75	12
Non-informative[b]	29	39	4
By Cancer Type			
Breast	345	283	18
Colorectal	49	33	13
Gastric	12	6	2
Lung	121	47	6
Pancreatic	26	8	3
Other[c]	307	209	23

[a]*Subset of* in silico *sample count.* [b]*Unstaged and stage 0 cancer samples.* [c]*Anorectal; bladder; renal; cervical; esophageal; head/neck; hepatobiliary; lymphoma; melanoma; ovarian; prostate; thyroid; uterine; unknown primary.*

Figure 2. Size Selection with Pippin Prep



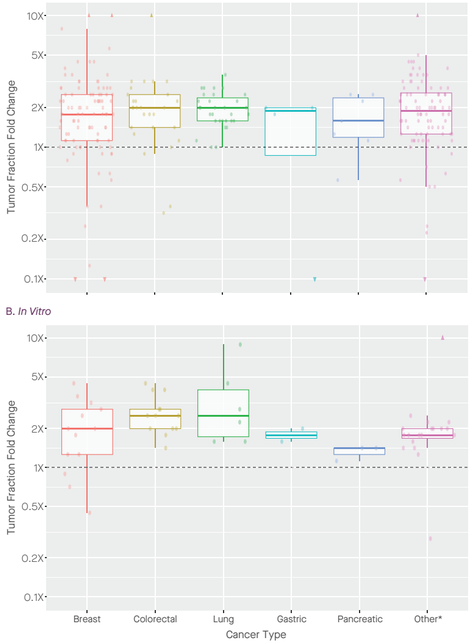
Fragment length distributions of samples before and after size selection (aggregated within condition). Red depicts size-selected libraries; blue depicts unselected libraries. A notable shift towards shorter fragment lengths was observed in the size-selected libraries.

- *In vitro* analyses used a subset of CCGA1 test set samples (n=93: 28 non-cancer individuals, 65 cancer individuals stages I–IV), including C cases sampled across a range of tumor fractions (**Table 1**). NC samples were age matched to C samples.
- When sequencing, the number of fragments was matched to that of the original CCGA1 samples, which resulted in lower depth of 23±4.45X due to an enrichment of shorter fragments.
- To test whether size-selecting to shorter fragment lengths resulted in further increases in tumor fraction, tumor fraction was measured on *in vitro* size-selected samples that were further *in silico* size-selected to shorten fragment lengths starting at 140 bp incrementing down to 50 bp in intervals of 10 bp.

RESULTS

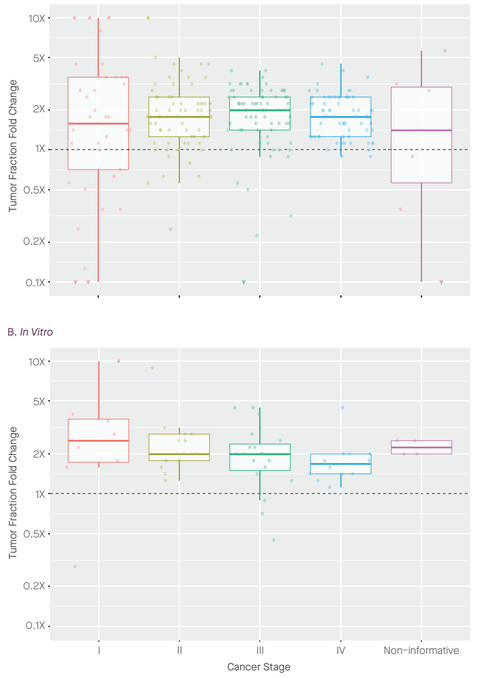
- *In silico* and *in vitro* analyses resulted in median 2.00±0.58-fold and 2.00±0.52-fold increases in overall tumor fraction, respectively (compared to original CCGA1 non-size-selected samples).
- Tumor fraction fold change was consistent across tumor types (*in silico*: 1.78±0.73 breast, 2.00±0.58 CRC, 2.00±0.41 lung [**Figure 3A**]; *in vitro*: 2.00±0.82 breast, 2.51±0.52 CRC, 2.53±0.94 lung [**Figure 3B**]), in contrast to recent work that defined a group of "low ctDNA" cancer types that may not benefit from size selection.[4]
- This was consistent across stages (*in silico*: 2.00±0.74 I-III, 1.78±0.52 IV [**Figure 4A**]; *in vitro*: 2.00±0.55 I-III, 1.68±0.29 IV [**Figure 4B**]).

Figure 3. Tumor Fraction Fold Change After Size Selection by Tumor Types



*Cancer type-specific fold change (log scale) in estimated tumor fraction after in silico (A) and in vitro (B) size selection for all fragments less than 140 bp. Dashed line depicts unchanged tumor fraction. *Anorectal; bladder; renal; cervical; esophageal; head/neck; hepatobiliary; lymphoma; melanoma; ovarian; prostate; thyroid; uterine; unknown primary.*

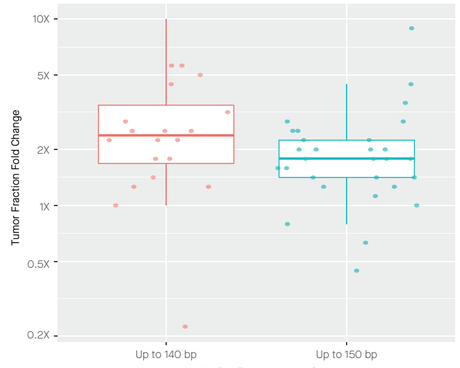
Figure 4. Tumor Fraction Fold Change After Size Selection by Stage



*Stage-specific fold change (log scale) in estimated tumor fraction after in silico size-selection that kept fragments in 90–150 bp range (A) and in vitro size-selection that kept fragments shorter than 140 bp (B). Dashed line depicts unchanged tumor fraction. *Unstaged (in silico and in vitro) and stage 0 cancer (in silico only) samples.*

- When comparing tumor fraction between *in vitro* samples that were size-selected to 140 or 150bp, we observed that, on average, samples size-selected to 140bp had higher tumor fraction (**Figure 5**). When further filtering *in vitro* samples to retain only shorter fragments, tumor fraction did not significantly change relative to the original estimate.

Figure 5. Tumor Fraction Fold Change on the *In Vitro* Samples by Maximum Target Fragment Length



Tumor fraction fold change (log scale) between in vitro samples that kept fragments of up to 140bp length versus those that kept fragments up to 150bp. Samples with undetectable tumor fraction before or after size-selection were removed. Extreme outliers are shown as triangles.

- WGS classifier (as previously described[3]) retrained on *in silico* size-selected data had increased sensitivity of 37.8% at 98% specificity compared to sensitivity of 32.9% on the full-depth data, and 32.1% on the downsampled set ($p<1e-5$).

CONCLUSIONS

- *In silico* and *in vitro* size selection consistently increased tumor fraction across all tested cancer types and stages, and this increase was maximized by tuning size selection.
- Classification performance improved significantly relative to both full-depth data and data downsampled to match depth after size-selection.
- Given that tumor fraction appeared maximal when size-selecting to lengths up to 140 bp, these data suggest that size selection of even modest amounts targeting cfDNA under 140 bp may enhance cfDNA-based cancer detection in WGS assays.

References: 1. Jiang P, et al. *Proc Natl Acad Sci*. Mar 2015, 112 (11): E1317-E1325. 2. Underhill HR, et al. *PLoS Genet*. 2016 Jul 18, 12 (7): e1006162. 3. Klein. E, et al. Development of a comprehensive cell-free DNA (cfDNA) assay for early detection of multiple tumor types: The Circulating Cell-free Genome Atlas (CCGA) study. ASCO. 2018. Chicago, USA. 4. Mouliere, F, et al. *Science Translational Medicine*. 2018 Nov 07, 10 (466): eaat4921

Corresponding author: ajamshidi@grail.com

Funding and Author Disclosures: Study funded by GRAIL, Inc. All authors are employees of GRAIL, Inc. with equity in the company. MHL and AJ are shareholders of Illumina, Inc. MHL is a shareholder of Johnson & Johnson, Procter & Gamble Company. JN is a shareholder of Ultragenyx Pharmaceutical Inc.