Tumor Methylation Patterns Accurately Measure Tumor Fraction in Cell-free DNA

2020 ASCO (Virtual)
Annual Meeting

Collin Melton, Pranav Singh, Oliver Venn, Earl Hubbell, Samuel Gross, Yasushi Saito, Josh Newman, Lily Zhang, Christopher Chang, Gordon Cann, Matthew H. Larson, Sarah Stuart, Vasiliki Demas, John Beausang, Alexander M. Aravanis, Arash Jamshidi
GRAIL, Inc.

INTRODUCTION

- Site-specific DNA methylation patterns are characteristic indicators of cellular identity, including neoplastic states.
- The presence of methylation patterns associated with the neoplastic state in cell-free DNA (cfDNA) and rarely found in non-cancer individuals allows estimation of the fraction of circulating tumor DNA and variant containing DNA allele fraction.
  - Tumor fraction (the proportion of molecules in cfDNA that are tumor derived) and allele fraction (the proportion of molecules that are tumor derived and contain a variant) are direct measures of signal for cfDNA-based cancer applications.
- In the Circulating Cell-free Genome Atlas study (CCGA; NCT02889978), we demonstrated multi-cancer detection across >50 cancers and tissue of origin prediction from plasma cfDNA samples by measuring tumor-derived methylation signals.[1]
- Here we extend the idea of SNV allele fraction to methylation patterns in order to estimate allele fractions from methylation sequencing data and subsequently assess the limit of detection of the classifier that was previously used for methylation-based multi-cancer detection.[1]

METHODS

Sequencing and Variant Calling

- In a subset of training samples from the pre-specified second CCGA sub-study, the following data were collected:
  - Tumor biopsy whole-genome bisulfite sequence (WGBS) from DNA extracted from formalin-fixed paraffin-embedded tissue (N = 822).
  - cfDNA targeted methylation (TM) sequencing, cfDNA whole-genome sequencing (WGS), and tissue biopsy WGBS (N = 231).
  - Non-cancer controls (N = 898).
- Somatic single-nucleotide variants (SNVs) were identified using tumor biopsy WGBS.
- Methylation variants were identified by comparing tumor biopsy WGBS to cfDNA WGBS in 898 non-cancer controls.

Allele Fraction and Tumor Fraction Estimation

- Tumor fraction was modeled as follows

$$Prob(tf \mid data) \sim \prod_{i=1}^{n} \text{Pois}\,(x_i;\, \lambda_i) * Prob\,(tf)$$

x_i= observed abnormal counts of site i in cfDNA
tf= tumor fraction
λ_i= lambda for site i = $[tf * vaf_i + (1 - tf) * noise_i] * depth_i$
vaf_i= variant allele fraction for site i in the biopsy
$noise_i$= site specific noise rate in cfDNA
$depth_i$= depth of site i in the cfDNA

- cfDNA allele fraction was modeled as (tumor fraction) * (mean biopsy allele fraction).
- Direct measurement of cfDNA allele fraction was made in subset of samples using WGS cfDNA counts of SNVs called in biopsy WGBS.

Allele Fraction Limit of Detection

- For each tumor sample we estimated allele fraction with the matched targeted methylation cfDNA sample. In addition, we estimated an artificial allele fraction in 100 non-cancer control cfDNA samples to assess the limit of the method in these negative controls.
- Specificity was assessed using the rank of the allele fraction assessed in the matched cfDNA relative to the non-cancer controls.

Classifier Limit of Detection

- Classifier limit of detection (LOD) was assessed in terms of both tumor fraction and allele fraction using Bayesian logistic regression (Bayesian probit regression was also assessed and gave comparable results).
- LOD was reported at 50 and 95% probability of detection.
- To improve the regression fit for 95% probability of detection, samples were filtered to those with concordant SNV and methylation variant (MV) allele fraction estimates.

RESULTS

Methylation Patterns are an Alternative to SNVs for Cancer Detection

- Methylation patterns can be used to determine the tissue of origin.[1]
- Figure 1 depicts how we define a methylation variant as a sequence of 5 adjacent CpGs whose methylation states are indicative of cancer (present in tissue biopsy) and low noise (<1/10,000 frequency in non-cancer cfDNA).
- Methylation variants are plentiful and unlike SNVs are largely recurrent across cancer samples.

Figure 1. Diagram of Tumor Shedding DNA into Bloodstream and Blood-based Methods for Detecting Cancer Signals

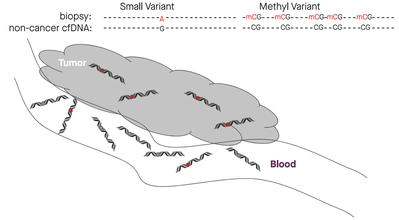
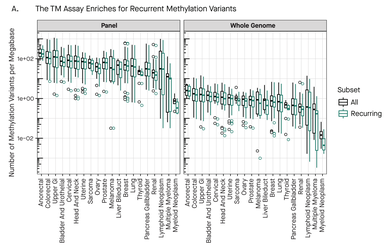


Figure 1: SNV-based cancer detection utilizes single nucleotide changes between cancer and non-cancer cfDNA; methylation variant-based cancer detection utilizes changes in a sequence of cytosine methylation states between biopsy and non-cancer cfDNA.

Methylation Variants Occur Frequently, Recur Across Patients, and can be Easily Assessed with Targeted Sequencing

- There is a high density of methylation variants per megabase on our TM pull-down panel and the majority are recurrent (Figure 2A). This is in contrast to SNVs which are typically not recurrent and thus cannot easily be enriched with targeted sequencing.
- In WGBS of tumor biopsy, we observed a median of 2,635 MVs distributed across the genome.
- A median of 86.8% are recurrent (they are shared with at least one other participant, N = 822 participants) (Figure 2B).
- A median of 69.3% are targeted by our TM assay.

Figure 2. Identification of Broadly Distributed and Recurrent Methylation Variants



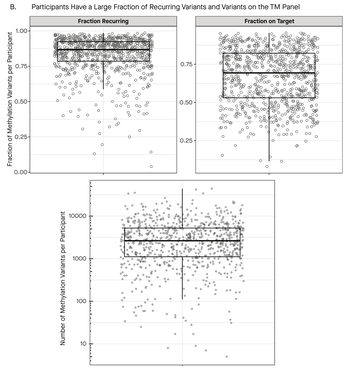


Figure 2: (A) A boxplot of the number of methylation variants per megabase is shown per cancer type for all variants (black) and recurrent variants (green) either on a targeted enrichment panel (left panel) or in the whole genome (right panel). These data demonstrate that a high density of methylation variants can be captured with targeted sequencing across a diverse set of cancer types. (B) The distribution of the fraction of variants on target per sample, the fraction of recurring variants, and the total number of variants per sample are shown, indicating that most samples have a large fraction of variants that recur and that are on this targeted panel.

Methylation Variants can be used to Estimate cfDNA Allele Fraction

- We applied the methylation variant and SNV tumor and allele fraction estimation approaches to 231 participants with 19 cancer types.
- Tumor and allele fraction estimates were concordant between MV and SNV estimates (Spearman's Rho 0.820 and 0.822 respectively) (Figure 3).

Figure 3. Concordant Allele Fraction Estimates

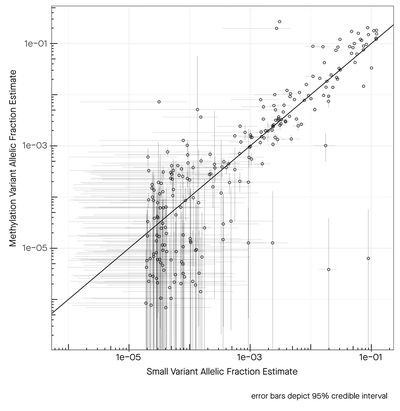
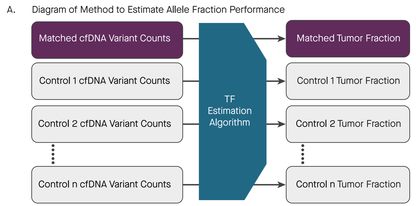


Figure 3: Allele fraction estimates are concordant between methylation and SNV approaches. A scatter plot depicts the concordance of allele fraction estimation using SNVs and methylation variants independently.

- We assess the lower limit of allele fraction estimation by comparing allele fraction estimates using biopsy matched cfDNA to estimates made using control (non-cancer) cfDNA (Figure 4).
- The allele fraction estimation method performs well to an allele fraction of ~10^{-4}.
- Performance is notably worse in samples with less than 100 called variants.

Figure 4. Assessment of the Lower Limit of Allele Fraction Estimation



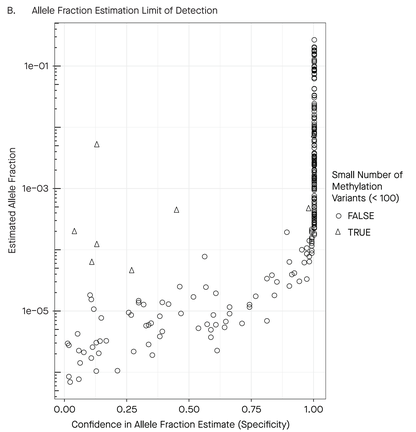


Figure 4: (A) Diagram for determining a per sample allele fraction specificity. For variants called in a single biopsy sample, allele fraction is estimated in matched cfDNA as well as in the cfDNA of 100 non-cancer control subjects. The specificity is determined using the rank of the matched biopsy cfDNA allele fraction relative to the allele fraction estimates in the controls. (B) A scatter plot depicting the estimated allele fraction versus the per sample specificity demonstrates good performance down to an allele fraction of ~10^{-4}. Confidence in the allele fraction estimate increases with the number of called variants per sample as highlighted by those samples with few called variants (triangles) having higher estimated allele fractions despite having low specificity.

Allele fraction estimates can be used to assess classifier performance

- Targeted methylation tumor fraction and allele fraction estimates explained classifier performance (Spearman's Rho 0.856 and 0.866 respectively).
- The classifier tumor fraction LOD50 at 99.4% specificity was 0.00082 (0.00055 - 0.00119 95% CI) and allelic fraction LOD50 is 0.00029 (0.0002 - 0.0004 95% CI).
- The classifier allele fraction LOD95 at 99.4% specificity was 0.00104 (0.00061 - 0.0021).
- The classifier allele fraction LOD95 at a slightly less conservative 99.0% specificity was 0.00081 (0.00048 - 0.00158) (Figure 5); at this threshold, samples were detected down to 0.00006 allelic fraction.

Figure 5. Classifier Limit of Detection

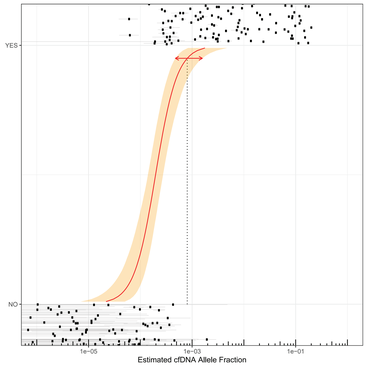


Figure 5: Classifier limit of detection was assessed using Bayesian logistic regression by predicting the probability of cancer detection (at 99.0% specificity) from the allele fraction. The posterior distribution on the regression fit is depicted as the median (red line) with 95% credible interval (shaded orange). The allele fraction giving 95% probability of detection is shown using the dotted line and its credible interval is depicted by the red double sided arrow. Individual participant detection status and allele fraction estimates are depicted by the squares.

CONCLUSIONS

- Tumor-derived fragments with highly specific methylation patterns are present in the cfDNA of participants with cancer.
- The majority of methylation variants recur across patients and can be effectively enriched using a targeted panel.
- Methylation variant abundance in cfDNA is a measure of tumor cfDNA allele fraction, which is a major factor influencing classification performance.
- The low classifier LOD for this multi-cancer detection test provides further evidence of the utility of DNA methylation in early cancer detection.

Reference

1. Liu MC, Oxnard GR, Klein GA, CCGA Consortium, Swanton C, Seiden, MV. Sensitive and specific pan-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol.*, in press.

Disclosures

Study funded by GRAIL, Inc. All authors are current or former employees of GRAIL, Inc. with equity in the company. **CM** is a shareholder of Novartis, Clovis, Cara, Gilead and Bluebird, **MHL** is a shareholder of Illumina, Johnson & Johnson, Procter & Gamble, **SS** is a shareholder of Illumina, **VD** is a shareholder of Alphabet and T2 Biosystems. **AJ** is a shareholder of Illumina. **AMA** is an advisor to and equity holder in Foresite Labs.

Contact

Collin Melton (cmelton@grailbio.com)