Optimized Early Cancer Detection from Whole-Genome Sequencing of Cell-Free DNA

2019 ASHG Annual Meeting
October 15–19, 2019
Houston, TX

Collin Melton, PhD; Peter Freese, PhD; Sid Bagaria, MS; Earl Hubbell, PhD; Darya Filippova, PhD; Robert Calef; M. Cyrus Maher, PhD, MS, MPH; Virgil Nicula; Sam Gross, PhD; Oliver Venn, DPhil; Anton Valouev, PhD; Alex Aravanis, MD, PhD; Arash Jamshidi, PhD

GRAIL, Inc., Menlo Park, CA.

INTRODUCTION

- Cancer can be detected through whole-genome sequencing (WGS) of cell-free DNA (cfDNA).
- We previously demonstrated that classification scores from fragment methylation-based features have greater sensitivity at 98% specificity than somatic copy number alteration (SCNA)-derived scores from WGS data[1-2].

OBJECTIVE

- Evaluate the impact of three WGS-derived alternative features (fragment lengths, allelic imbalance, fragment endpoints) on cancer/non-cancer classification performance.

METHODS

- Data from the first substudy of Circulating Cell-free Genome Atlas (CCGA; NCT02889978) were split into training and testing sets.
  - Training set is 561 non-cancer controls and 863 cancer participants from 20 solid tumor types.
  - Testing set is 362 non-cancer controls and 464 cancer participants from 20 solid tumor types.
- Models were optimized using 10-fold cross-validation on the training set.
- Results shown are from the held out testing set.
- A binary classification model was trained for each of the following three alternative WGS features:
  - Fragment lengths: Principal component regression on region-level (100 kb bin) fragment length features. Each region feature was computed as the geometric mean of the per-fragment cancer to non-cancer fragment length likelihood ratio. Cancer fragment length distributions were estimated within the training set from samples with tumor fraction estimates[5] greater than 5% using the following relationship:

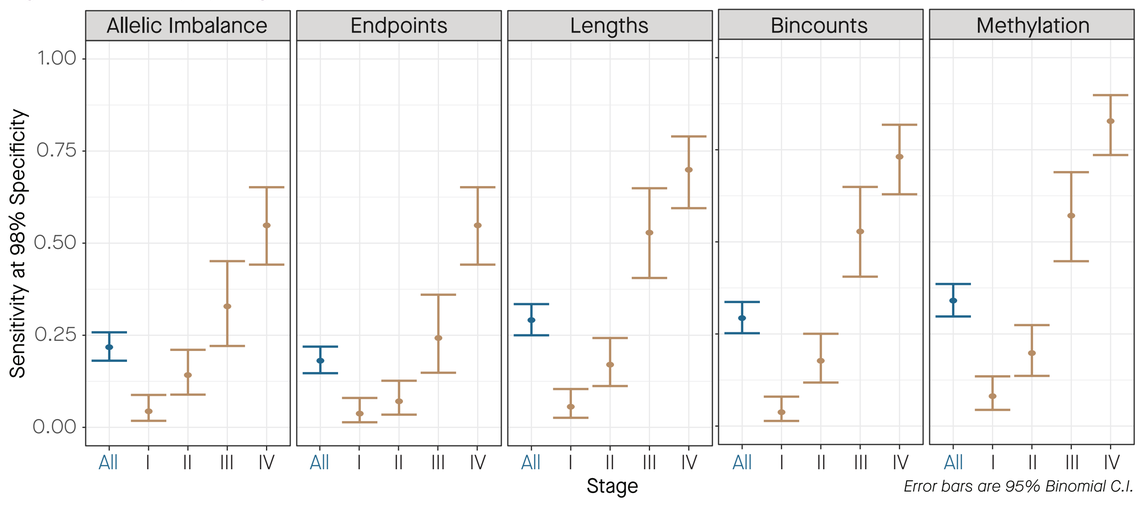
$$P(length \mid sample) = tumor\ fraction * P(length \mid cancer) + (1 - tumor\ fraction) * P(length \mid noncancer)$$

    Per sample estimates of the cancer length distribution were aggregated using the median density per length and renormalized to sum to 1. Per region features were normalized to an independent set of non-cancer reference samples and corrected for GC bias.
  - Allelic imbalance: Logistic regression on chromosome arm-level features that reflect the degree of allelic imbalance. Features were computed as the median of the absolute value of the per-bin (100 kb bin) deviation of the observed from the expected probability of observing a phased haplotype. Bins were filtered to those with at least 50 phased SNPs. Phasing was computed using Eagle2[3] with the 1000 Genomes[4] reference set.
  - Fragment endpoints: Logistic regression on the normalized counts of short (50–140 bp) fragments terminating at cancer-enriched endpoint positions. Cancer-enriched endpoint positions were identified by hierarchical Bayesian modeling of non-cancer endpoint counts as a Gamma/Poisson mixture with a position-specific rate distribution derived from a Gamma prior shared by all positions.
- Performance of the three new alternative WGS classifiers was compared to our previously reported methylation and SCNA bincount classifiers[1-2] using McNemar's test at 98% specificity.
  - Methylation classifier from WGBS (whole-genome bisulfite sequencing) data: Logistic regression on ranked fragment methylation scores. Each score is an estimate of the cancer to non-cancer likelihood ratio.
  - Bincounts classifier from WGS SCNA data: Convolutional neural net using normalized fragment counts per 100kb genomic bin.
- Model complementarity was assessed by comparing ensemble classifiers using logistic regression on scores from two or more classifiers.
  - Each alternative WGS classifier was ensembled with Methylation or Bincounts and compared to the Methylation or Bincounts classifier, respectively.
  - All WGS classifiers (three alternative + Bincounts) were ensembled and compared to the Methylation and Bincounts classifiers.

RESULTS

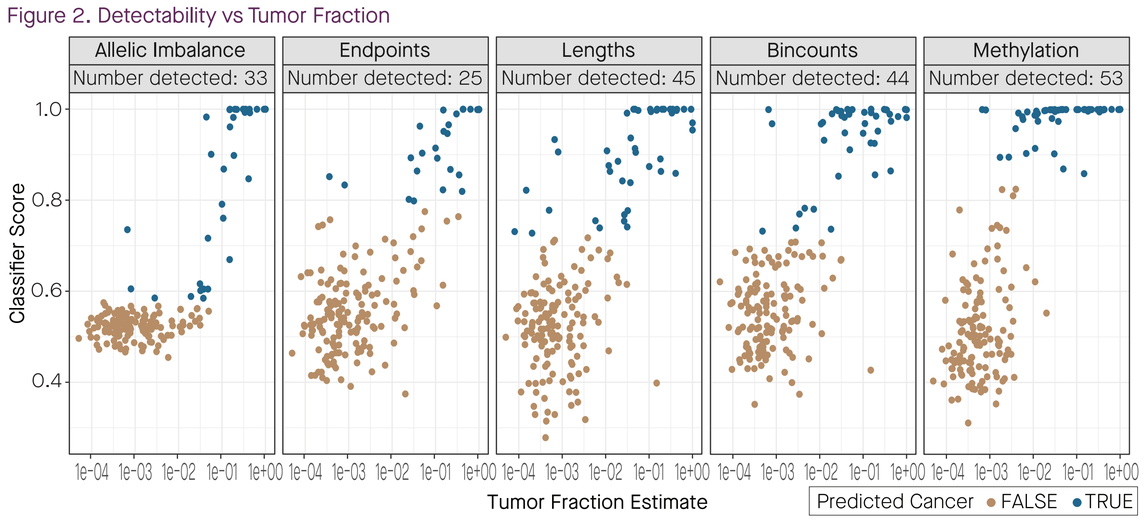
- Overall sensitivity at 98% specificity of the new alternative WGS classifiers was: 18.1% (95% CI, 14.7–21.9) for fragment endpoints, 29.1% (25.0–33.5) for fragment lengths, and 21.8% (18.1–25.8) for allelic imbalance. In comparison, sensitivity was 34.1% (29.7–38.6) for the previous Methylation classifier and 29.3% (25.2–33.7) for the previous Bincount classifier. All classifiers showed an increased sensitivity with stage reflecting availability of biological signal (Figure 1) and sample scores correlated with tumor fraction[5] (Figure 2).

Figure 1. Sensitivity by Stage



Sensitivity at 98% specificity was computed for all stages (blue) as well as by stage (brown). Cancer detectability increased with increasing stage for all classifiers.

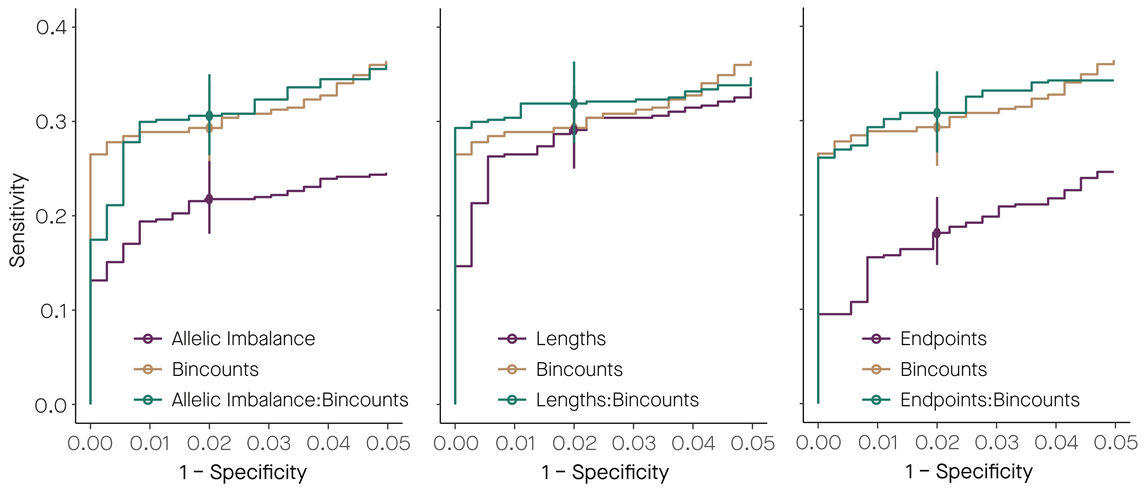
Figure 2. Detectability vs Tumor Fraction



Classification score (y-axis) was plotted against small variant based tumor fraction estimates (x-axis) for the subset of the testing set with available tumor fraction estimates. Blue points indicate samples predicted as cancer given a 98% specificity threshold.

- Classification on fragment lengths performed comparably to and complemented WGS Bincounts sensitivity at 98% specificity. The endpoint and allelic imbalance classifiers show substantially worse performance (Figure 3).

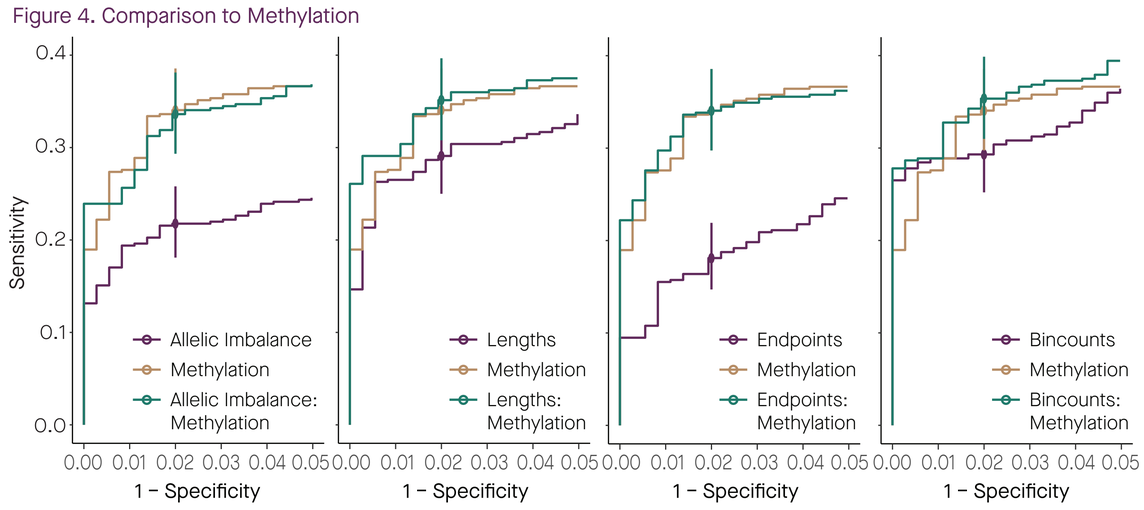
Figure 3. Comparison to Bincounts



ROC curves were generated for individual alternative WGS feature classifiers both alone and ensembled with the Bincounts classifier. Binomial confidence intervals are plotted for sensitivity at 98% specificity.

- The WGS-derived classifiers have lower sensitivity than the Methylation classifier at 98% specificity and do not increase sensitivity when individually ensembled with the Methylation classifier (Figure 4).

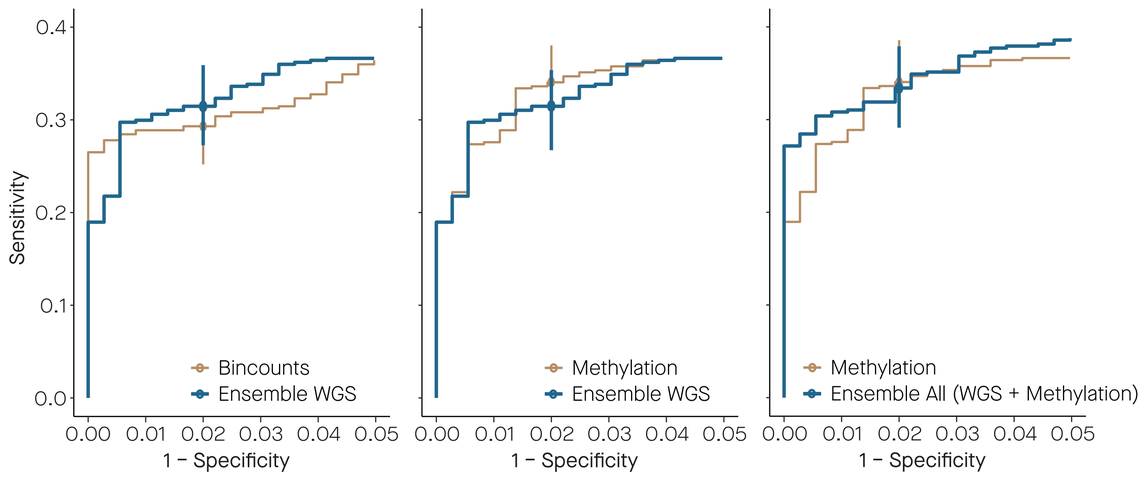
Figure 4. Comparison to Methylation



ROC curves were generated for individual WGS feature classifiers both alone and ensembled with the Methylation classifier. Binomial confidence intervals are plotted for sensitivity at 98% specificity.

- A combined WGS classifier outperforms the bincounts classifier @98% specificity but remains worse than and does not complement the methylation classifier (Figure 5).

Figure 5. Ensemble of All WGS and Ensemble of All WGS + Methylation Compared to Methylation



ROC curves were generated for an ensemble of all the WGS classifiers and all the WGS classifiers plus the Methylation classifier and compared to the Bincounts and Methylation classifiers. Binomial confidence intervals are plotted for sensitivity at 98% specificity.

CONCLUSIONS

- All classifiers showed strong cancer stage and tumor fraction dependence reflecting expectations based on signal availability.
- An ensemble WGS classifier combining fragment endpoints, allelic imbalance, fragment lengths, and bincounts outperformed bincounts alone (+2.2% sensitivity, p=0.044) but had lower sensitivity than the methylation-based classifier (-2.6% sensitivity, p=0.045).
- WGS and methylation features combined did not improve upon the methylation-based classifier alone (-0.6% sensitivity, p=0.546).
- Overall, combining WGS features improved sensitivity at high specificity in the absence of methylation features but was inferior to methylation-based classification; these results, in part, motivated development of an improved methylation assay assessed in the second substudy of CCGA[6].

References

1. Klein EA, 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. *J Clin Oncol.* 2018;36(no. 15_suppl):12021.
2. Liu MC et al. Plasma cell-free DNA (cfDNA) assays for early multi-cancer detection: the Circulating Cell-Free Genome Atlas (CCGA) study. *Annals of Oncology* (2018) 29 (suppl_8): viii14-viii57. 10.1093/annonc/mdy269.
3. Loh et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics,* 2016 Nov;48(11):1443-1448.
4. 1000 Genomes Project Consortium., *Nature.* 2015 Oct 1; 526(7571):68-74.
5. Venn O, Hubbell E, Sakarya O, et al. Poster presented at: The Biology of Genomes Meeting; May 7-11, 2019; Cold Spring Harbor, NY.
6. Liu MC et al. Genome-wide cell-free DNA (cfDNA) methylation signatures and effect on tissue of origin (TOO) performance. *J Clin Oncol.* 2019;37(15_suppl):3049.

Disclosures

Study funded by GRAIL, Inc. All authors are employees of GRAIL, Inc. with equity in the company.