Poster
#6335

Accurate epigenomic estimates of circulating tumor fraction in large-scale clinical data

William W. Young Greenwald, Yupeng He, Sai Chen, Tingting Jiang, Anton Valouev, Jun Min, Catalin Barbacioru,
Daniel P. Gaile, Dustin Ma, Yvonne Kim, Giao Tran, Indira Wu, Ariel Jaimovich, Victoria M. Raymond, Rebecca J. Nagy, and Han-Yu Chuang
Guardant Health, Palo Alto, CA

Abstract

Liquid biopsy offers a rapid and non-invasive alternative to tissue biopsy for identifying biomarkers. More recently, its application has broadened to include assessment of early response to therapy (i.e. molecular response) and in the early-stage settings, detection of minimal residual disease (MRD) and early disease recurrence[1]. Circulating tumor fraction (cTF), the fraction of circulating DNA from tumor cells, is usually estimated by somatic mutations that are well associated with the tumor progression and prognosis. However, interference can occur from clonal hematopoiesis of indeterminate potential (CHIP), and for cell-free DNA (cfDNA) samples that lack detectable somatic mutations, somatic tumor fraction cannot be estimated using this method. In this analysis, we demonstrate that epigenomic signatures accurately measure cTF using orthogonal analytes to somatic mutations and enable cTF estimation even in cases without detectable tumor sequence variants.

Methods

To capture tumor-associated methylated cfDNA, we designed a custom assay on a broad genomic panel (~15.2 Mb) that targets unmethylated regions in plasma cfDNA from healthy individuals. DNA molecules that support methylation were enriched by our assay and this information was post-processed into our machine learning models.

With this panel, we profiled plasma samples from a training set of ~2,000 cancer patients with solid tumor from all stages and ~2,000 cancer-free donors (Table 1). For the prediction of cancer/cancer-free status, we trained a logistic regression model. For cTF prediction, we trained a linear model using the frequency of somatic mutations as the approximation of true cTF to build prediction models. To minimize artifacts, we filtered these mutation calls with our predefined list of common somatic mutations.

On the test dataset, we profiled 559 cancer patients and 131 cancer-free donors. We applied all cancer-specific prediction models onto the test dataset to estimate 1) the cancer/cancer-free classification performance of single models and the aggregated model; 2) the cTF prediction performance.

To further benchmark the accuracy of our methylation models, we built an *in-vitro* and an *in-silico* titration datasets. The *in-vitro* titration dataset was generated by mixing cfDNA from patients with colorectal cancer (CRC) into the plasma from cancer-free donors via experimental titration. The *in-silico* titration dataset was generated by computationally mixing sequencing reads from CRC patients with those from cancer-free donors.

Type	#Samples	
	Training	Test
Cancer-free	2,014	131
CRC	1,596	32
Lung	276	203
Breast	243	146
Other cancer	283	178

Table 1: An overview of training and test datasets in this study

Conclusions

We demonstrate that our methylation approach is capable of accurately quantifying cTFs in somatic-mutation positive and negative cases:

- Our assay can reliably enrich DNA methylation signals in cancer-related regions in the genome.
- Our cancer-specific models achieve >90% detection rate for late-stage cancer patients while maintaining 95% specificity.
- In CRC, our methylation cTF prediction has a Pearson correlation of 0.85 with the orthogonal measure of cTF from somatic mutations. Although this correlation is high, in disparate cases due to potential interference in the genomic-only approach, the methylation cTF approach may be closer to biological truth.
- As we estimate somatic-mutation negative cases to be 30-50% of patients with stage I-III cancer and 15-20% of patients with stage IV cancer, our methylation approach may hold promise for providing better evaluation for patient care and management.

Results

Prediction of cancer status

We built our machine-learning models for cancer/cancer-free classification and cTF prediction on the training dataset of 2,398 cancer patients and 2,014 cancer-free donors. We first evaluated our methods on the training dataset via five-fold cross validation – this validation process was repeated 10 times to estimate the variation of our approach. At 95% specificity, our prediction model for cancer/cancer-free status has an average of 93% detection rate for samples across all stages (Figure 1, blue line). The tumor-fraction prediction model has a similar performance as the status prediction model (Figure 1, orange line).

After the training process, we applied the trained models and their 95% specificity cut-offs to the independent test dataset of 559 cancer patients and 131 cancer-free donors. On the test dataset, we observed a 97% specificity with a total of 4 false positives (FPs) observed across all three models (Table 2). The FP in CRC model is included in the 4 FPs in the lung model.

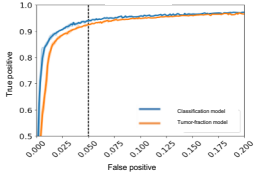

Figure 1: Five-fold cross validation of model performance for the prediction of CRC/cancer-free status in the training set. Shadows indicate variations in the 10 iterations.

Model	Specificity (#True Negative)	Sensitivity (#True Positive)
Lung	96.9% (127)	90% (182)
Breast	100% (131)	95.3% (139)
CRC	99.2% (130)	84% (26)
Aggregated	96.9% (127)	91.3% (409)

Table 2: The performance of cancer prediction models on the independent test dataset

Manual examinations revealed that the FPs are slightly above the tumor-normal cutoff, as well as that some strong signals come from regions with background noise in cancer-free donors. We further refined the model by removing these regions with strong background signals. In the refined model, one out of the four FPs was correctly predicted as TN.

Variance and limit of detection (LoD)

We applied our cTF prediction model to an *in-silico* titrated dataset of 1,000 samples, generated by computationally mixing sequencing reads from 1,000 CRC patients with 1,000 cancer-free donors at different levels. Our method quantified a cTF over 0.1% in >99% of these samples. In contrast, when applied to the dataset of 2,014 cancer-free donors, <5% of the samples resulted in estimated cTFs >0.1%.

We applied our cTF prediction model along with our genomic caller for common CRC somatic mutations on an *in-vitro* dataset comprised of 270 samples that were generated by experimentally titrating plasma from CRC patients into plasma of cancer-free donors at different levels.

At low cTF, the coefficient of variation (CV) of methylation was more robust than the CV from cTF estimated by somatic mutations (Table 3).

TF	#Samples	CV	
		Genomic	Methylation
<0.3%	154	0.72	0.17
0.3%~1%	62	0.24	0.17
>1%	54	0.03	0.05

Table 3: CV of cTF predictions from genomic calls and methylation in the *in-vitro* dataset

Prediction of circulating tumor fractions

With a predefined set of driver genes for CRC from previous studies, we used the frequency of driver mutations called from genomic data as the approximation for underlying true cTFs. We first compared the predicted cTF against this approximated true cTF (Figure 2, left). Methylation-predicted cTFs are consistent with the cTFs inferred based on genomic variants, showing a Pearson correlation of 0.85.

We observed a few outliers that have different methylation vs genomic cTFs; however, it is worth noting that genomic cTFs, even with well developed caller and filters (see methods), may still include artifacts such as CHIP. The straight vertical line at 0% genomic cTF contains cancer-free donors and CRC patients of no detectable somatic mutations. As the cancer-free donors are self-reported, there is a low probability that our negative set includes false negatives, and thus their cTF signals can still be captured by our genomic and/or methylation approaches.

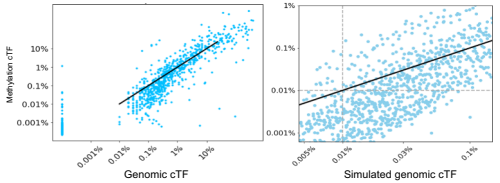

Figure 2: Performance of cTF models (black lines for diagonals). Left: test dataset of CRC and cancer-free samples; Right: *in-silico* dataset for lower truth cTFs.

It is possible that methylation can resolve samples with cTF <0.1%; however, this possibility is difficult to evaluate as cTF<0.1% are below the estimated limit of detection (LoD) for most of current companion diagnostic products[2]. To test this, we *in-silico* titrated data from 100 CRC patients into the data from 100 cancer-free donors at different levels between 0.005% and 0.1% and tested our trained models on this dataset (Figure 2, right).

While cTFs of simulated samples in the 0.03%~0.01% range were consistent with the methylation-based cTF, estimates broke down below 0.01%. We hypothesize that below this level, there aren't enough true methylation signals existing above the noise to allow for a robust cTF prediction.

Patients without detectable somatic mutations

Previous studies have shown that 30-50% of patients with stage I-III cancer, and 15-20% of patients with stage IV cancer, may lack detectable somatic mutations. Current methods relying on somatic mutations thus cannot quantify the cTF for these patients, leaving a large population with unmet need.

We show that methylation-based cTFs enables the prediction of cTF for this 15-50% of patients who lack of detectable somatic mutations (Figure 3, bottom). As expected, the median predicted cTF of these patients is lower than those with detectable somatic mutations (0.3% vs 0.002%).

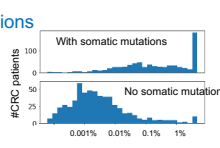

Figure 3: Predicted cTFs for CRC patients in the training set with/without somatic mutations

References

1. Cescon, David W., et al. "Circulating tumor DNA and liquid biopsy in oncology." *Nature Cancer* 1.3 (2020): 276-290.
2. Deveson, Ira W., et al. "Evaluating the analytical validity of circulating tumor DNA sequencing assays for precision oncology." *Nature biotechnology* 39.9 (2021): 1115-1128.