P1997

Early-stage colorectal cancer detection using artificial intelligence and whole-genome sequencing of cell-free DNA

Katherine Niehaus, Nathan Wan, David Weinberg, Brandon White, Ajay Kannan, Erik Gafni, Tzu-Yu Liu, Imran S. Haque, Girish Putcha

Freenome, Inc., South San Francisco, CA

BACKGROUND

- Despite population screening programs and availability of several stool-based, non-invasive screening methods, nearly 60% of colorectal cancer (CRC) cases are detected with regional or distant metastases (Siegel et al., 2018)
- Blood-based methods using cell-free DNA (cfDNA) are under development as an alternative to stool-based tests
- Early-stage detection of cancer using only tumor-derived mutations in cfDNA (i.e., circulating tumor DNA, or ctDNA) is challenging for practical, technical, and biological reasons, such as the small proportion of cfDNA derived from tumor tissue (i.e., tumor fraction, or ctDNA/cfDNA) in early-stage disease (Haque et al., 2017)
- Using machine learning (ML) to discover signatures in cfDNA that may reflect both tumor and non-tumor (e.g., immune) contributions represents a promising direction for the early detection of cancer
- Confounders, including variation in preanalytical and analytical processes, can affect the performance of ML models, especially in retrospective studies, and must be controlled to limit bias and improve generalizability

OBJECTIVE

- As part of a program to develop a blood-based screening test for CRC, a machine learning approach for representing and learning associations between cfDNA profiles and cancer status was evaluated in a large cohort of non-cancer controls and early-stage CRC patients (predominantly stages I and II), with a focus on the importance of accounting for known confounding variables

METHODS

- **Sample collection:** De-identified plasma samples were received from academic medical centers and commercial biobanks (**Table 1**)
- **Whole-genome sequencing of cfDNA:** cfDNA was isolated from 250 µL of plasma and converted into Illumina-compatible libraries, which were sequenced to a minimum of 400 million reads
- **Bioinformatics and feature generation:** Reads aligning to annotated protein-coding genes were extracted, and read counts were normalized to account for variability in read depth, sequence-content bias, and technical batch effects (Pertea et al., 2018)
- **Machine learning:** ML models were trained using different cross-validation techniques including k-fold, k-batch, and balanced k-batch (**Figures 2, 3**)

Figure 1. Methods from sample processing to results

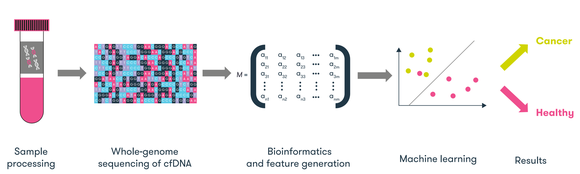
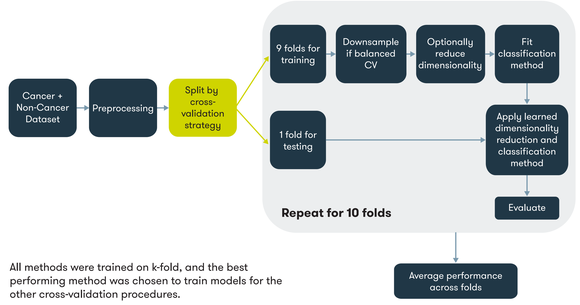


Table 1. Clinical characteristics and demographics of patients with CRC and non-cancer controls

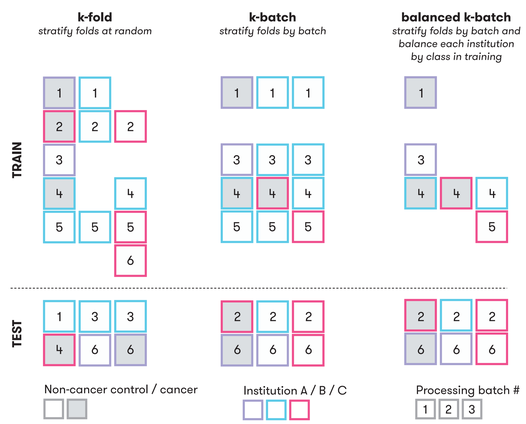
	CRC N=797	Control N=456	Total Samples N=1253
Gender N (%)			
Female	377 (47%)	279 (61%)	656 (52%)
Male	411 (52%)	122 (27%)	533 (43%)
Unknown	9 (1%)	55 (12%)	64 (5%)
Stage N (%)			
I	239 (30%)	N/A	N/A
II	417 (52%)		
III	114 (14%)		
IV	10 (1%)		
Unknown	17 (2%)		
Age (yrs)			
Median (IQR)	69 (61-77)	59 (53-64)	65 (57-74)

Figure 2. Model training and cross-validation (CV) procedures



All methods were trained on k-fold, and the best performing method was chosen to train models for the other cross-validation procedures.

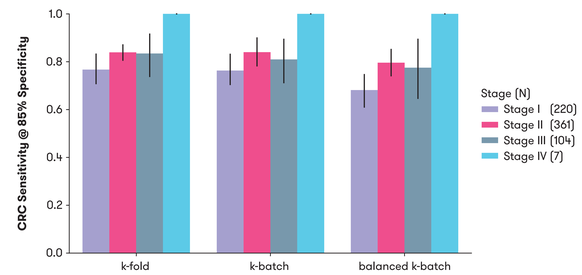
Figure 3. Training schemas for k-fold, k-batch, and balanced k-batch



Each square represents a single sample, with the fill color indicating class label (CRC or non-cancer control), the border color representing the institution of origin, and the number indicating processing batch.

RESULTS

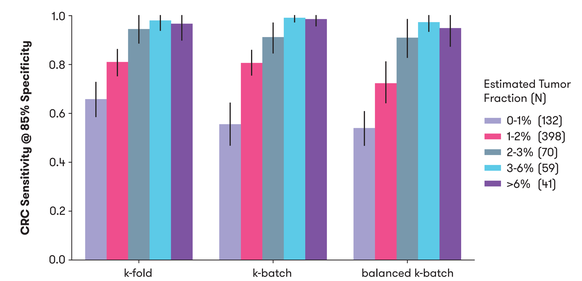
Figure 4. Sensitivity by CRC stage in patients aged 50-84



Classification performance for CRC within the intended-use age range (50-84) across all validation methods. Threshold for sensitivity was defined at 85% specificity in each test fold. N is number of samples for each stage. CI=95% bootstrap confidence interval.

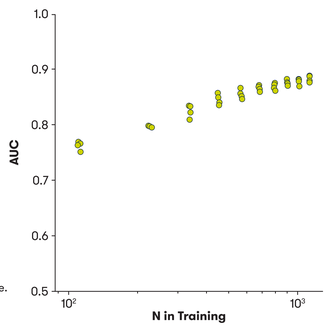
- 82% of samples were from patients with early-stage CRC (stages I and II)
- All validation methods achieved approximately equivalent sensitivity across stages I through III (based on confidence intervals). Stage IV cancer was always classified correctly

Figure 5. Sensitivity by tumor fraction in patients aged 50-84



Classification performance for CRC within the intended-use age range (50-84) across all validation methods. Threshold for sensitivity was defined at 85% specificity in each test fold. N is number of CRC samples. Tumor fraction is the proportion of cfDNA derived from tumor tissue (i.e., ctDNA/cfDNA) and was estimated using IchorCNA (Adalsteinsson et al., 2017). CI=95% bootstrap confidence interval.

Figure 6. Non-linear relationship between total number of samples in training set and AUC in test set



AUC=area under the receiver operating characteristic curve.

- Classifier performance continued to improve with the addition of more training samples

Table 2. CRC performance by cross-validation procedure in patients aged 50-84

Validation	Average Training Set Size (N)	Mean AUC (95% CI)	Mean Sensitivity at 85% Specificity (95% CI)
k-fold	1128	0.89 (0.87–0.91)	82% (78–85%)
k-batch	1128	0.89 (0.87–0.91)	80% (76–85%)
balanced k-batch	592	0.86 (0.83–0.89)	75% (68–81%)

AUC=area under the receiver operating characteristic curve; CI=95% bootstrap confidence interval.

- Batch-to-batch technical variability was evaluated using k-batch validation
- Institution-specific differences in population or sample handling were evaluated using balanced k-batch validation
- Sensitivity increased with increasing tumor fraction across all validation methods (**Figure 5**)
- AUC for IchorCNA-estimated tumor fraction alone was 0.63, which was lower than results from the ML model under any cross-validation scheme (**Table 2**)

CONCLUSIONS

- A prototype blood-based CRC screening test using cfDNA and machine learning achieved high sensitivity and specificity in a predominantly early-stage CRC cohort (stages I and II)
- Classifier performance suggests contributions from both tumor and non-tumor (e.g., immune) derived signals
- Assessing genome-wide cfDNA profiles at moderate depth of coverage enables the use of low-volume plasma samples
- Cross-validation methods highlighted the importance of performing similar confounder analyses for retrospective (and prospective) studies
- Prospective validation of a similar machine learning method using cfDNA is underway (NCT03688906), along with research evaluating the potential of a multi-analyte approach that integrates other cell-free, blood-based analytes (e.g., proteins) to improve performance

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Dr. Andrew Godwin and the Biospecimen Repository Core Facility staff (funded in part by the National Cancer Institute Cancer Center Support Grant (P30 CA168524) and NHS Research Scotland Tayside Biorepository), Geneticist Inc., iSpecimen Inc., and Indivumed for support of this research by providing de-identified plasma samples. We also thank Signe Fransen for her extensive suggestions, feedback, and editorial support.

REFERENCES

Adalsteinsson et al. 2017; *Nature Communications* 8 (1): 1324

Haque et al. 2017; *bioRxiv* https://doi.org/10.1101/237578

Pertea, et al. 2018; *bioRxiv* https://doi.org/10.1101/332825

Siegel et al. 2018; *CA: A Cancer Journal for Clinicians* 68 (1): 7–30

Presented at American College of Gastroenterology (ACG) 2018 Annual Scientific Meeting, October 8–10, 2018, Philadelphia, PA, USA