Serum-based colorectal cancer detection using orphan noncoding RNAs

Exai Bio

Hani Goodarzi[1], Jeffrey Wang[2], Oluwadamilare I. Afolabi[2], Lisa Fish[2], Helen Li[2], Kimberly H. Chau[2], Patrick Arensdorf[2], Fereydoun Hormozdiari[2], Babak Alipanahi[2]

UCSF School of Medicine, University of California, San Francisco, CA, Exai Bio Inc., Palo Alto, CA

UCSF School of Medicine

Background

- Small non-coding RNAs (sncRNAs) have established roles as post-transcriptional regulators of cancer pathogenesis.
- We previously reported a novel and previously unannotated class of sncRNAs that were found in breast cancer tissue but not in normal tissue adjacent to the tumor, which we termed orphan non-coding RNAs (oncRNAs).[1] Since then, we have identified and validated novel oncRNAs in multiple cancer tissues, using data from The Cancer Genome Atlas (TCGA) and other independent cohorts.[2]
- We recently showed that these oncRNAs can also be detected in sera and demonstrated prognostic value for treatment response among breast cancer patients.[3]
- Early detection of colorectal cancer (CRC) can drastically improve survival odds, reduce treatment complexity and side effects, and improve patient quality of life.[4]
- We hypothesize that oncRNAs can be used as biomarkers in a liquid biopsy strategy to detect CRC across a range of cancer stages and tumor sizes.

Goals

- Develop and validate a methodology that uses machine learning (ML) to accurately predict CRC status based on oncRNA profiles detected in patient sera.

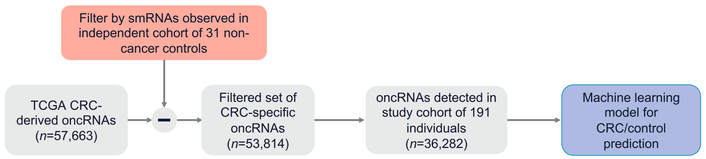
Samples

- Our study cohort consists of 191 frozen serum samples from clinically diagnosed colorectal cancer patients (*n*=96) and age- and sex-matched individuals from the general population with no known diagnosis of cancer (*n*=95). Samples were acquired from three commercial biobanks and processed for small RNA (smRNA) sequencing. Dates of blood draw for serum collection range from 2009 to 2022.
- Subjects were treatment-naive at sample collection and were selected to represent all stages of CRC (I–IV) as well as a broad range of ages of onset, including patients <45 years old.
- Patients had provided informed consent and contributing centers had obtained IRB approval.

Methods

- RNA was extracted from frozen serum samples of ≤1.0ml volume and prepared for sequencing. Sample libraries were sequenced to an average depth of 18.8 million 50 bp single-end reads per sample.
- oncRNAs were previously identified in multiple cancer tissues, using data from TCGA as a discovery cohort. Of this multicancer library of oncRNAs, 57,663 were significantly present in TCGA CRC samples. To refine our TCGA library of CRC-associated oncRNAs for applications in serum, we filtered out smRNA sequences found in sera of an independent non-cancer control cohort (*N*=31). OncRNAs that were detected in more than one control serum sample were filtered out, yielding a final set of 53,814 CRC-significant oncRNAs.
- This filtered library of oncRNAs was used as a reference to generate oncRNA expression profiles by cataloguing and quantifying oncRNAs for each individual serum sample *(N*=191).
- These oncRNA expression profiles were used to build an ensemble of logistic regression models to make predictions of CRC vs. control. The ensemble model was trained and evaluated using a 5-fold cross-validation setup. Within each training fold only oncRNAs observed in >4% of samples and yielding an odds ratio for CRC >1 were used to train and validate the model.

oncRNA Library Creation and Profiling



Study Cohort

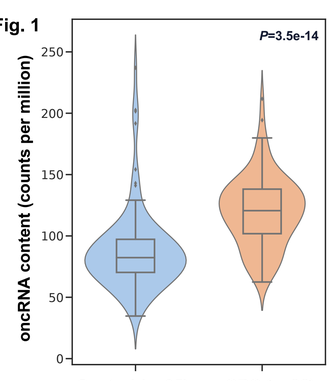
	Non-cancer Controls (*n*=95)	Colorectal Cancer Cases (*n*=96)
Age (mean ± std)	52.8 (±14.9)	53.8 (±12.4)
<45	24	24
45–55	24	24
55–65	24	24
>65	23	24
Sex *(n)*		
Male	48	48
Female	47	48
Cancer Stage *(n)*		
I	NA	24
II	NA	24
III	NA	24
IV	NA	24

Result 1: oncRNA Content Differentiates Cancer Status

Figure 1. oncRNA Content in Cancer and Control Samples

- Of the 53,814 TCGA CRC-specific oncRNA species, 36,282 (67.4%) were observed in the study cohort *(N*=191).
- Total sequencing depth-normalized oncRNA content, the aggregate count of all detected oncRNAs within each sample, was significantly higher in cancer samples (one-sided Mann-Whitney *U* test, *P*=3.5e-14).

Fig. 1

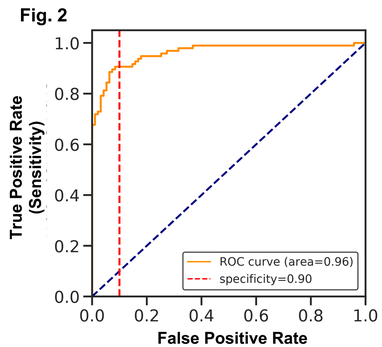


Result 2: Prediction of Colorectal Cancer Status

Figure 2. ROC Curve of an Ensemble Model

- A five-fold cross validation of an ensemble of logistic regression model's CRC prediction performance on our study cohort (*N*=191).
- On average, 3,285 oncRNAs were used as features within each fold.
- Overall area under the ROC curve (AUC) across folds is 0.964 (95% CI: 0.938–0.99).
- The model achieved an overall sensitivity (true positive rate) of 90.6% (95% CI: 82.9%–95.6%) with specificity set at 90% across folds.

Fig. 2



Figures 3 & 4. Sensitivities for CRC Detection by Cancer Stage (I–IV) and Tumor T Category (T1–T4)

- For each subgroup, using the model, sensitivity was calculated with specificity set at 90% (based on recent CMS reimbursement publication). 95% confidence intervals were calculated using the Clopper-Pearson method.
- Sensitivities for CRC detection were similar and high across all stages and tumor T categories.

Fig. 3

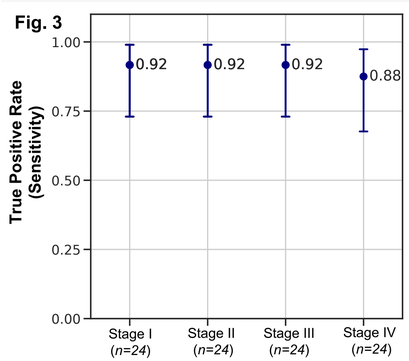
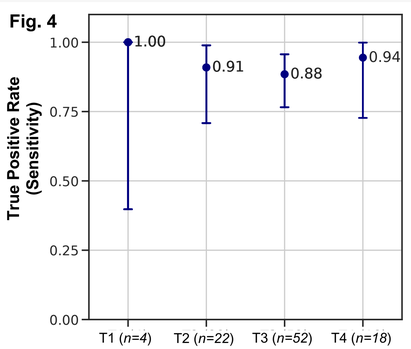


Fig. 4



Conclusion

- Analyzing oncRNA data with machine learning models accurately predicted colorectal cancer (CRC) across all cancer stages (I–IV) and tumor categories (T1–T4).
- This oncRNA-based liquid biopsy technology is compatible with standard sample requirements enabling integration into conventional clinical workflows,
- The results will be validated prospectively in further population studies.

Disclosure: JW, OA, LF, HL, KC, FH are full-time employees of Exai Bio. BA and PA are co-founders, stockholders, and full-time employees of Exai Bio. HG is co-founder, stockholder, and advisor of Exai Bio.

Reference:

1. Fish L., et al. *Nature Med.* 2018;24:1743-51.
2. Wang J, et al. *AACR.* 2022; 3353.
3. Navickas A., et al. *SABCS.* 2021; PD9-04.
4. Xi Y., et al. *Transl Oncol.* 2021;14(10):101174