Poster ID: P1-05-18 San Antonio Breast Cancer Symposium® – December 6-10, 2022

Early-stage breast cancer detection using orphan noncoding RNAs

Taylor Cavazos[1], Jeffrey Wang[1], Oluwadamilare I. Afolabi[1], Alice Huang[1], Dung Ngoc Lam[1], Seda Kilinc[1], Jieyang Wang[1], Lisa Fish[1], Xuan Zhao[1], Andy Pohl[1], Helen Li[1], Kimberly H. Chau[1], Patrick Arensdorf[1], Fereydoun Hormozdiari[1], Hani Goodarzi[2], Babak Alipanahi[1]

[1]Exai Bio Inc., Palo Alto, CA, [2]UCSF School of Medicine, University of California, San Francisco, CA

Background

- Small non-coding RNAs (sncRNAs) have established roles as post-transcriptional regulators of cancer pathogenesis.
- We previously reported a novel and 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 and event-free survival among breast cancer patients.[3]
- Early detection of breast cancer is crucial for optimal patient outcomes but cannot always be accomplished based on symptoms or mammography.
- We hypothesize that oncRNAs can be used as biomarkers in a liquid biopsy strategy to detect breast cancer across a range of cancer stages and tumor sizes.

Goals

- Develop and validate a methodology using machine learning to accurately predict breast cancer status based on oncRNA profiles detected in patient sera.

Samples

- The study cohort includes clinically diagnosed female breast cancer patients (*N*=96) and age- and sex-matched individuals from the general population with no known diagnosis of cancer (*N*=95).
- Breast cancer patients were treatment-naive at sample collection and were selected for this study to represent all stages of breast cancer (I–IV) and a broad range of ages, including patients <45 years old.
- Samples were acquired from two commercial biobanks and processed for small RNA sequencing. Dates of blood draw for serum collection ranged from 2010 to 2022.
- Patients had provided informed consent and contributing centers had obtained IRB approval.

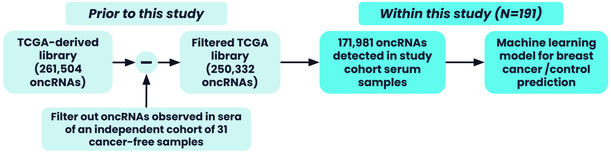
Methods

- RNA was extracted from 191 frozen serum samples of 1.0ml volume and prepared for sequencing. Sample libraries were sequenced to an average depth of 17.7 million 50 bp single-end reads per sample.
- Previously, 261,504 oncRNAs were found to be significantly associated with cancer across multiple tissues, using data from The Cancer Genome Atlas (TCGA) as a discovery cohort through a pan-cancer study.[2] To refine our TCGA library of tissue-derived oncRNAs for applications in serum samples, all oncRNA sequences detected in >1 serum sample of an independent cohort of 31 cancer-free controls were removed, yielding a library of 250,332 oncRNAs.
- This filtered library of 250,332 oncRNAs was used as a reference database to generate oncRNA expression profiles in our study cohort (*N*=191). Of these, 171,981 were detected in at least one individual serum sample.
- oncRNA expression profiles were used to build an ensemble of logistic regression models to make predictions of breast cancer vs. control. The ensemble model was trained and evaluated using a 5-fold cross-validation setup. Within each training fold only oncRNAs observed in >3% of samples and yielding an odds ratio for breast cancer >1 were used to train and validate the model.

Study Cohort

Demographics		Cancer-Free Controls (*N*=95)	Breast Cancer Cases (*N*=96)
Age	Mean (SD)	56.3 (13.0)	59.4 (13.9)
	≥ 65 years, N (%)	29 (30.5%)	33 (34.4%)
Sex	Female, N (%)	95 (100%)	96 (100%)
Smoking	Never-Smoker, N (%)	87 (91.6%)	45 (46.9%)
	Ever-Smoker, N (%)	8 (8.42%)	51 (53.1%)
BMI	BMI < 30, N (%)	67 (70.5%)	74 (77.1%)
	BMI ≥ 30, N (%)	28 (29.5%)	22 (22.9%)

oncRNA Library Creation and Profiling

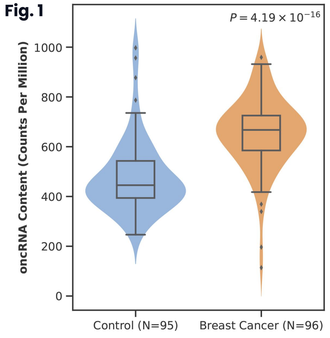


Result 1: oncRNA Content Differentiates Cancer Status

Figure 1. oncRNA Content in Control and Breast Cancer Serum Samples

- Of the 250,332 oncRNAs in the filtered TCGA multicancer library, 171,981 (68.7%) were detected in the study cohort (*N*=191) with 33,043 observed in >3% of all study samples.
- Total oncRNA content in a sample -- the aggregate count of those 33,043 frequently observed oncRNAs, normalized by sequencing depth -- was significantly greater among cancer samples (one-sided Mann-Whitney *U* test, $P=4.19\times10^{-16}$).

Fig. 1



Result 2: Ability of oncRNA-Based Model to Discriminate Between Breast Cancer Patients & Cancer-Free Controls in Serum

Fig. 2

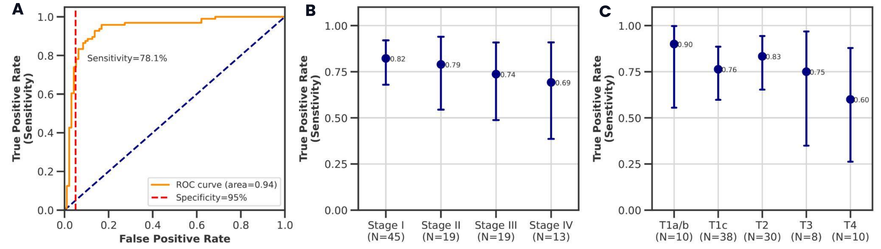


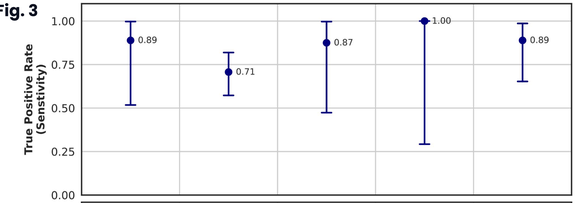
Figure 2. (A) ROC Curve, (B) Sensitivity by Cancer Stage, and (C) Sensitivity by Tumor T Category

- An ensemble of logistic regression models trained on serum oncRNA measurements with 5-fold cross validation (*see Methods*) discriminated effectively between samples from patients with breast cancer vs. cancer-free controls (A-C).
- On average, 10,416 [range: 9,769–11,312] oncRNAs were used as features within each training fold.
- The ROC curve demonstrated an AUC of 0.94 (95% CI: 0.88–0.98), with a sensitivity of 78.1% (68.5%–85.9%) at 95% specificity (A).
- Sensitivities at 95% specificity were highest for early cancer stages (B) and small tumor sizes (C) and tended to be similar across these categories. 95% CI were calculated using the Clopper-Pearson method.

Figure 3. Model Sensitivity by Breast Cancer Molecular Subtypes

- Breast cancer samples (*N*=96) were assigned molecular subtypes based on ER, PR, HER2 expression through immunohistochemistry (Top).
- Each subtype was subdivided into early (I/II) and late (III/IV) stage at diagnosis (Bottom).
- Sensitivities at 95% specificity were uniformly high, at least 87% for each molecular subtype except Luminal A (Top) and were often greater among stage I/II than stage III/IV (Bottom).

Fig. 3



	HER2+ (N=9)	Luminal A (N=58)	Luminal B (N=8)	Triple-negative (N=3)	Triple-positive (N=18)
Stage I/II	0.86 (0.42–1.00) N = 7	0.75 (0.58–0.88) N = 36	0.80 (0.28–0.99) N = 5	1.00 (0.29–1.00) N = 3	0.92 (0.64–1.00) N = 13
Stage III/IV	1.00 (0.16–1.00) N = 2	0.64 (0.41–0.83) N = 22	1.00 (0.29–1.00) N = 3	NA N = 0	0.80 (0.28–0.99) N = 5

Conclusions

- Analyzing oncRNA data with machine learning models accurately predicted breast cancer across all cancer stages (I–IV) and tumor categories (T1–T4).
- Early stage tumors and small tumors were detected with high sensitivity: 82% for stage I and 90% for T1a/b, respectively (at 95% specificity).
- This oncRNA-based liquid biopsy technology is compatible with standard blood sample requirements enabling integration into conventional clinical workflows.
- The results will be validated prospectively in further population studies.

Disclosures: TC, JW, OA, AH, DNL, SK, JW, LF, XZ, AP, 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.

References:
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.