SABCS 2025, December 9th–12th, San Antonio, TX

Early-stage breast cancer detection with a plasma cell-free RNA and AI-based liquid biopsy platform

Lee Schwartzberg[1], Mehran Karimzadeh[1], Amir Momen-Roknabadi[1], Taylor B. Cavazos[1], Nae-Chyun Chen[1], Jieyang Wang[1], Michael Multhaup[1], Jeremy Ku[1], Akshaya Krishnan[1], Martha Hernandez[1], Rose Hanna[1], Lisa Fish[1], Magdalena Gebala[1], Anna Hartwig[1], Hani Goodarzi[2], Helen Li[1], Fereydoun Hormozdiari[3], Babak Behsaz[1], Babak Alipanahi[1]

[1]Exai Bio Inc., Palo Alto, CA; [2]Arc Institute, Palo Alto, CA; [3]University of California, Davis, CA, US

Background

- Mammography reduces mortality but has lower sensitivity in women with dense breasts; a substantial fraction of eligible women miss regular screening. Blood tests could complement imaging and improve adherence.
- Early detection of breast cancer with cell-free DNA has been sub-optimal. We quantify orphan non-coding RNAs (oncRNAs), tumor-emergent small RNAs largely absent from healthy tissue, present in tumor tissue and plasma[1], and pair them with a generative AI classifier.
- We use a generative, probabilistic model of cf-oncRNA profiles that learns the joint distribution of markers, technical covariates, and biological heterogeneity[2]. By modeling how true signals and noise are generated, the platform denoises counts, mitigates site/batch effects, and yields a calibrated cancer probability per sample.

Study Design & Cohorts

- Retrospective multi-site study with training and independent validation cohorts from distinct suppliers (>75 collection sites)of plasma from untreated breast cancer patients and controls without breast cancer to determine sensitivity of cancer detection.
- Training: 745 breast cancer, 258 controls. Validation: 92 breast cancer, 98 controls. 1 mL plasma per sample. All passed QC for analyses.
- Case mix: majority stage I/II; broad age and demographic representation aligned with screening-age populations.

Methods

- Biomarker discovery: TCGA small-RNA profiles used to build a pan-cancer oncRNA library enriched in tumors vs adjacent normals.
- Processing: automated cell-free RNA workflow; 100-bp single-end sequencing; ~53M reads per sample on average.
- Modeling: locked generative AI model trained with 5-fold cross-validation; evaluated on an independently sourced test set; sensitivity reported at 90% specificity.

Table 1: Study Demographics

Demographics		Training set		Held-out test set	
		Cancer	Control	Cancer	Control
Sample size	Count, n	745	258	92	98
Age	Mean (SD)	59.56 (13.47)	58.36 (13.02)	63.24 (11.56)	57.09 (9.08)
Smoking status	Never-Smoked, n (%)	528 (70.87%)	207 (80.23%)	79 (85.87%)	41 (41.84%)
BMI	BMI obese (≥ 30), n (%)	238 (31.95%)	82 (31.78%)	30 (32.61%)	22 (22.45%)
Race	White, n (%)	408 (54.77%)	240 (93.02%)	65 (70.65%)	66 (67.35%)
	Black/African American, n (%)	46 (6.17%)	13 (5.04%)	0 (0.00%)	1 (1.02%)
	Asian, n (%)	155 (20.81%)	2 (0.78%)	0 (0.00%)	1 (1.02%)
	Other/Unknown, n (%)	136 (18.26%)	3 (1.16%)	27 (29.35%)	30 (30.61%)
Ethnicity	Hispanic, n (%)	47 (6.31%)	23 (8.91%)	0 (0.00%)	2 (2.04%)
	Non-hispanic, n (%)	191 (25.64%)	235 (91.09%)	92 (100.00%)	96 (97.96%)
	Other/Unknown, n (%)	507 (68.05%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Supplier	Indivumed, n (%)	516 (69.26%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	MT Group, n (%)	163 (21.88%)	258 (100.00%)	0 (0.00%)	0 (0.00%)
	DLS, n (%)	0 (0.00%)	0 (0.00%)	92 (100.00%)	98 (100.00%)
	Others, n (%)	66 (8.86%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Collection tube type	DNA Streck, n (%)	352 (47.25%)	258 (100.00%)	46 (50.00%)	48 (48.98%)
	PPT Tube, n (%)	393 (52.75%)	0 (0.00%)	46 (50.00%)	50 (51.02%)
Menopause status	Post-menopausal, n (%)	133 (17.9%)	1 (0.4%)	0 (0.0%)	60 (61.2%)
	Pre-menopausal, n (%)	44 (5.9%)	2 (0.8%)	0 (0.0%)	6 (6.1%)
	Unknown, n (%)	568 (76.2%)	255 (98.8%)	92 (100.0%)	32 (32.7%)
Subtype	ER+/HER2+, n (%)	141 (18.9%)		20 (21.7%)	
	ER+/HER2-, n (%)	403 (54.1%)		39 (42.4%)	
	ER+/HER2-, n (%)	61 (8.2%)		4 (4.3%)	
	ER-/HER2-, n (%)	97 (13.0%)		18 (19.6%)	
	Unknown, n (%)	43 (5.8%)		11 (12.0%)	
Clinical stage	I, n (%)	277 (37.2%)		30 (32.6%)	
	II, n (%)	316 (42.4%)		30 (32.6%)	
	III, n (%)	115 (15.4%)		23 (25.0%)	
	IV, n (%)	32 (4.3%)		9 (9.8%)	
	Unknown, n (%)	5 (0.7%)		0 (0.0%)	

Overview of Study Design

Participant Inclusion Criteria

- No history of prior cancer diagnosis or therapy
- No systemic immune-modulating therapy within 60 days of collection
- No recent surgery (within 1 month)
- No infusion of blood products within 30 days
- No active COVID-19 infection
- No organ transplantation
- Not pregnant within 12 months of collection
- Minimum age: 18

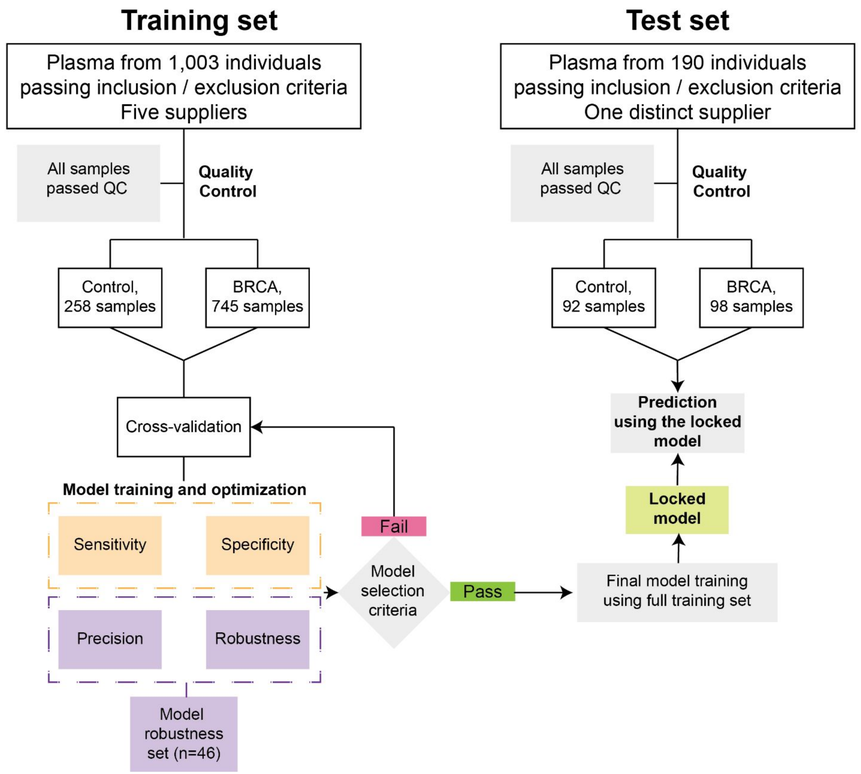

Figure 1. Study design. We collected plasma samples from a total of 1,203 individuals, including 837 with breast cancer and 356 without history of cancer. We set a dataset of 190 samples from a distinct supplier as the held-out test set and used the remaining samples for training a generative AI model through cross validation. Model selection involved evaluating performance as well as robustness. The locked model from the training dataset was evaluated on the held-out test set.

Ability of oncRNA-Based Model to detect breast cancer at different stages

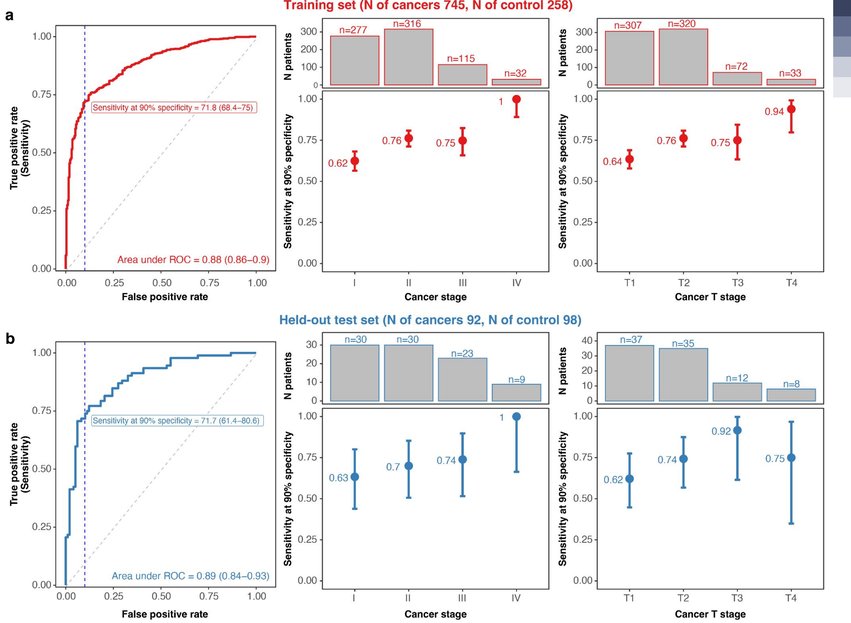

Figure 2. Training set cross-validated and held-out test set performance. (a) Left panel: ROC curve showing cross-validated performance on the training set, averaged across folds. Middle panel: Stage-wise sensitivity at 90% specificity. Right panel: Sensitivity at 90% specificity across different T-stages. **(b)** Similar to (a), but for held-out test set.

Overall Results

- Discrimination: AUROC 0.88 (95% CI 0.86–0.90) in training (CV); AUROC 0.89 (0.84–0.93) in the validation set.
- At a training-calibrated 90% specificity: training sensitivity 71.8% (95% CI 68.4%–75%). Results were comparable between the training and held-out test set.
- Early stage: stage I sensitivity 62% (training) and 63% (validation) consistent with rising sensitivity in higher stages.
- Robust detection across small tumors (T1 and T2); sensitivity ~64–76% in training and ~62–74% in test set,, increasing with tumor burden.

Conclusions

- A 1 mL plasma cf-oncRNA assay coupled with a locked generative AI classifier detects breast cancer—including stage I—with strong accuracy and generalizes to an externally sourced cohort. Signal plausibility (active transcription → higher RNA copy number) supports sensitivity in smaller tumors.
- Cell-free oncRNA + AI offers an orthogonal, minimally invasive complement to screening imaging, ideally suited for the 40–50% of screening-age women with dense breasts who are recommended for supplemental testing. Prospective, site-balanced studies are planned to confirm performance in intended-use populations with well defined demographic and histologic subgroup representation evaluating integration with risk models and imaging pathways.

References:
1) Fish L., et al., *Nature Medicine*, 2018
2) Karimzadeh M. , *et al., Nature Communications*. 2024

Disclosures: LS, MK, AR, TBC, NC, JW, MM, JK, AK, MH, RH, LF, MG, AH, HL, and BB, are or were employees of Exai Bio. BA is a co-founder, stockholder, and full-time employee of Exai Bio. HG is a co-founder, stockholder, and advisor to Exai Bio. FH is an advisor to Exai Bio.