Abstract Number: 1274
Presentation Number: PD5-01

Deep Proteomics and AI Classifier for Early Breast Cancer Detection

Alec Horrmann[1], Yash Travadi[1], Jacob Carey[1], Ella Boytim[2], Grant Schaap[1], Kevin Mallery[1], Kaylee Judith Kamalanathan[1], Nathaniel Bristow[1], Catalina Galeano-Garces[1], Song Yi Bae[1], Adam Groth[1], Alexa Hesch[1], Carissa Rungkittikhun[1], Pooja Advani[3], Justin Hwang[2], Badrinath R. Konety[1,4], Justin M. Drake[1,2,5]
[1]Astrin Biosciences, MN, [2]Masonic Cancer Center, University of Minnesota, MN, [3]Divison of Hematology and Oncology, Mayo Clinic, FL [4]Allina Health, MN [5]Department of Pharmacology, University of Minnesota, MN

Introduction

Mammography has enabled breast cancer to be diagnosed at an early stage in a large percentage of patients; however, in the half of women with dense breast tissue, mammography sensitivity is as low as 30%, underscoring the need for more sensitive and accessible screening methods. The FDA's 2024 mammography regulation highlighting this issue has left patients more informed but without solutions, increasing patient anxiety.[1-2]

Liquid-based biopsies for early breast cancer detection are emerging but currently remain out of reach for clinical use. Recent data on nucleotide assessment from plasma in breast cancer have been mixed with 87% sensitivity for late-stage disease (stage 3-4) but only 20% sensitivity for early-stage disease (stage 1-2).[3] Proteomics is an exciting area for early cancer screening, with advances in sample preparation and equipment enabling deep proteome assays that detect proteins 8-9 orders of magnitude lower in abundance than common plasma proteins.

Sample Cohort

Banked plasma samples were purchased from two different vendors and shipped to Astrin Biosciences. Plasma from treatment naïve breast cancer patients were selected alongside demographically matched healthy donors.

Table 1. Summary of patient demographics between cohorts

Characteristics	Training Set		Validation Set	
	Healthy n = 466	Cancer n = 379	Healthy n = 195	Cancer n = 202
Age				
Mean (SD)	55.4 (10.8)	57.2 (12.4)	56.4 (8.8)	59.8 (13.0)
BMI				
Mean (SD)	26.2 (4.4)	27.8 (6.3)	26.2 (4.3)	26.7 (5.4)
Race, n (%)				
White	429 (92.1%)	355 (93.6%)	176 (90.3%)	197 (97.5%)
Black	12 (2.6%)	8 (2.1%)	4 (2.1%)	1 (0.5%)
Asian	21 (4.5%)	15 (4.0%)	10 (5.1%)	4 (2.0%)
Other/ Unknown	4 (0.8%)	1 (0.3%)	5 (2.5%)	0 (0.0%)
Cancer Stage, n (%)				
Stage 0		45 (11.9%)		13 (6.4%)
Stage 1		130 (34.3%)		69 (34.2%)
Stage 2		152 (40.1%)		98 (48.5%)
Stage 3		35 (9.2%)		16 (7.9%)
Stage 4		14 (3.7%)		2 (1.0%)
Unknown		3 (0.8%)		4 (2.0%)
Source, n (%)				
ProteoGenex	316 (67.8%)	240 (63.3%)	105 (53.8%)	129 (63.9%)
BioIVT	150 (32.2%)	139 (36.7%)	90 (46.2%)	73 (36.1%)

Methods

Proteins were isolated using a superparamagnetic bead solution then digested to create peptides. The peptide solution was loaded onto a mass spectrometer in data-independent acquisition (DIA) mode. Peptides were quantified by DIA-NN then normalized. A five-fold cross-validation model was developed on the training set then applied to the validation set.[4]

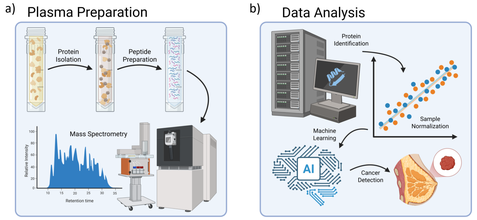


Figure 1. Sample Processing Overview. Created using BioRender.[5]

Cancer Classifier Development

Samples were randomly assigned to training and validation sets. Validation samples were blinded during sample preparation and analysis and batched separately from training samples to mimic real world patient specimens.

A machine learning classifier was developed using samples from 845 women and implemented using an Exponentiated Gradient method, with L2-norm regularized logistic regression and the constraint of demographic parity across sources. Training performance was assessed using 5-fold cross-validation, yielding an AUC of 0.961.

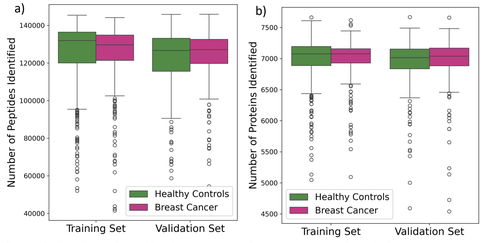


Figure 2. Consistent identification of (a) peptides and (b) proteins within the cohort.

Results

The validation set had a sensitivity of 92.3% at a specificity of 92.6%. The cancer cohort was analyzed further by stage and subtype. In both analyses, sensitivity remained above 84% among all groups. Sensitivity was higher than 90% for triple negative breast cancer, notoriously the most aggressive breast cancer subtype. Both sensitivity and specificity were analyzed across demographic groups and no demographic cofounders were identified.

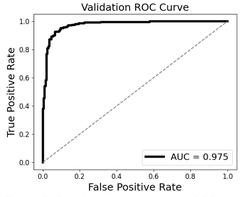


Figure 3. Receiver operator curve (ROC) for the validation data set.

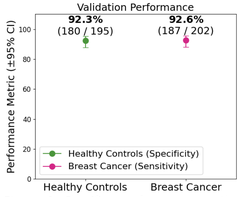


Figure 4. Classifier performance on the validation data set.

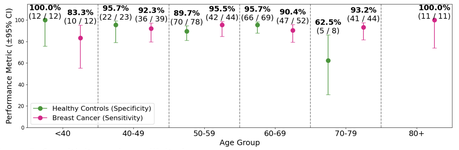


Figure 5. Sensitivity and specificity by age group.

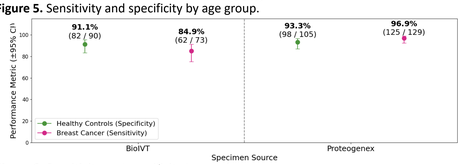


Figure 6. Sensitivity and specificity by specimen source.

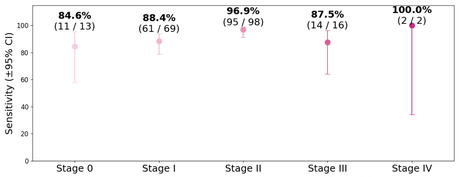


Figure 7. Sensitivity remained high across all cancer stages in the validation data set.

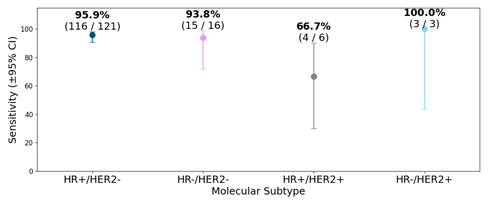


Figure 8. Sensitivity remained high across all molecular subtypes.

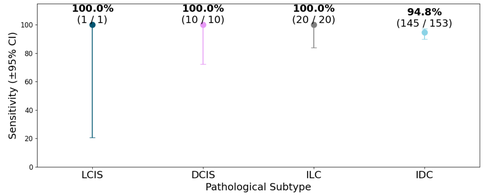


Figure 9. Sensitivity remained high across all pathological subtypes.

Pathway Analysis

Gene set enrichment analysis (GSEA) was performed on the breast cancer samples to ensure we were identifying cancer associated proteins and pathways.

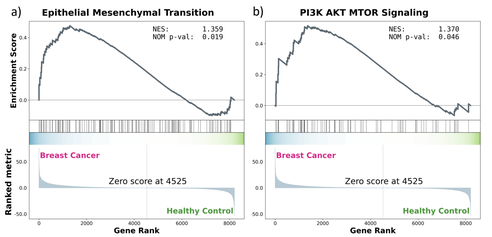


Figure 10. Enriched pathways for (a) epithelial-to-mesenchymal transition (EMT) and (b) PI3K-AKT signaling.

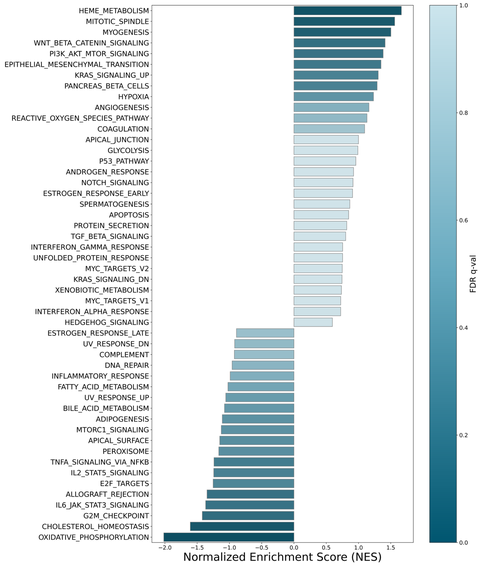


Figure 11. GSEA pathway enrichment analysis.

Conclusions

We have developed a highly sensitive blood-based assay that utilizes deep proteomic profiling to identify distinctive cancer specific signatures in women who are undergoing screening for breast cancer. This work enabled us to develop a protein-based classifier from plasma for early detection of breast cancer with both high specificity and high sensitivity even at early stages.

Clinical Impact

This assay advances clinical diagnostics by identifying proteomic markers for early detection of breast cancer. This innovative blood-based assay enhances screening strategies for women, especially those with dense breasts who are at average or high-risk for breast cancer due to current imaging limitations.

Justin M. Drake, Ph.D.
Astrin Biosciences
Email: justin.drake@astrinbio.com
Website: astrinbio.com

References

1. Mammography Quality Standards Act. (ed. Food and Drug Administration, H.) (2023).
2. Force, U.S.P.S.T., et al. Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement. JAMA 331, 1918-1930 (2024).
3. Klein, E.A., et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol 32, 1167-1177 (2021).
4. Horrmann, A., et al. A Plasma-based Deep Proteomic Platform for Early-Stage Breast Cancer Detection. medRxiv, 2025.2009.2022.25336353 (2025).
5. Created in BioRender. Biosciences, A. (2025) https://BioRender.com/ym22k44.

This presentation is the intellectual property of Astrin Biosciences. Contact Dr. Justin Drake at justin.drake@astrinbio.com for permission to reprint and/or distribute.