

SPARE: A Point-of-Care Blood Proteomic Rule-Out Test to Reduce Unnecessary Breast Biopsies After a Suspicious Imaging Finding

Reference set requested: EDRN Breast Reference Set (Breast Reference Set and Imaging). **Collaborative group:** Breast & Gynecological. **Organ site:** Breast.

Institution: Novelna Inc.

Specimens requested: 476 EDTA plasma aliquots (190 invasive carcinoma, 55 carcinoma in situ, 231 benign without atypia), released in a single blinded, plate-randomized shipment, minimum 500 μ L per specimen, with the covariate file specified in Section 4.

1. Clinical Relationship

Clinical question. The question addressed is diagnostic, at a single decision point: whether a woman referred for image-guided breast biopsy can safely defer that procedure. Under ACR BI-RADS, a finding sufficiently suspicious to warrant tissue sampling but not diagnostic of malignancy is assessed category 4, defined by a malignancy likelihood above 2% and below 95% and subdivided into 4A (>2–10%), 4B (>10–50%) and 4C (>50–95%).¹ SPARE is scoped to 4A and 4B, which account for 56% and 32% of subdivided category 4 assessments in the National Mammography Database and carry observed malignancy rates of 7.6% and 22% respectively.² Category 4C is outside the intended use: at a 69% malignancy rate, a negative result would leave a residual risk of approximately 14%, which does not support deferral.

Approximately 1.6 million breast biopsies are performed annually in the United States, of which 75–80% return benign pathology, corresponding to 1.19–1.27 million procedures and an estimated \$2.3–2.5 billion in annual expenditure.^{2,3} Documented harms include procedural pain, bleeding and infection risk, scarring that alters subsequent imaging of the affected breast, and psychosocial sequelae measurable 12–14 years after a false-positive result.^{4,5} No blood-based test is currently in use between the suspicious image and the biopsy.

Contribution of the reference set. The performance parameter of interest is discrimination of malignancy from biopsy-referred benign breast disease, estimated in specimens not selected by the developer. This requires cases and benign comparators accrued through a single clinical pathway under a common protocol, with venipuncture preceding tissue diagnosis and outcome ascertained by the index biopsy. The requested reference set satisfies these conditions and was assembled expressly so that circulating markers could be evaluated on blinded material.⁶

2. Background and Significance

Prior approaches. Three lines of work have addressed resolution of a suspicious breast lesion without tissue sampling. Supplemental and advanced imaging improves lesion characterization: meta-analysis indicates that a negative contrast-enhanced MRI could avoid a pooled 80.6% of stereotactic biopsies for BI-RADS 4 microcalcifications at a cost of 3.7% missed cancers,⁷ and modelling found this strategy cost-effective though not cost-saving.⁸ Delivery constrains it: a further appointment, intravenous contrast, capital equipment and a subspecialty reader are required, and the result follows the interval in which the biopsy decision is taken. Nucleic-acid liquid biopsy is constrained by tumor fraction at the lesion sizes presenting at this decision point; circulating tumor DNA reaches 50–90% sensitivity overall but declines substantially in early-stage disease at \$1,000–5,000 per test,⁹ and targeted methylation assays detected 30.5% of breast cancers overall and 2.6% at stage I in their largest validation set.¹⁰ Protein-based blood tests represent the closest prior art. CA 15-3 and CEA track tumor bulk and are not validated for pre-biopsy triage.¹¹ A combinatorial 11-protein and 33-autoantibody panel developed for this population fell from 92.3% sensitivity in training to 66.7% on validation,¹² and to 22% when applied prospectively to BI-RADS 4 calcifications.¹³

In work reported in 2015 during construction and characterization of the resource requested here, ninety candidate circulating proteins were assayed in 505 plasma specimens; none was found useful for distinguishing cancer from benign controls, although elevated CA-125 was identified as a candidate marker in estrogen receptor-negative cancers.⁶ The candidates evaluated were drawn from the abundant plasma fraction and assessed individually. The classifier proposed here differs in two respects that are testable in the same material: analytes were selected multivariately, so discrimination is a property of the classifier rather than of any constituent protein and no member reaches panel-level accuracy alone; and the retained analytes lie predominantly below the concentration range that single-analyte immunoassay reliably reaches. Whether that distinction is sufficient is the hypothesis this study tests, and the design permits its refutation.

Origin of the biomarker and platform. Novella Inc. has developed a proteomic discovery pipeline combining antibody-based proximity extension measurement with penalized multivariate selection, applied across biofluids and disease areas. In a study published in *BMJ Oncology* in 2024, plasma from 440 individuals spanning 18 early-stage solid tumor types — breast cancer among them — was profiled across more than 3,000 proteins; two sex-specific 10-protein detection panels separated cancer from non-cancer at AUC 0.980 in men and 0.983 in women, identifying 93% and 84% of stage I cancers respectively at 99% specificity.¹⁴ Sex-specific modelling was adopted because approximately 80% of the proteins differing significantly in men showed no significant difference in women, and the converse; the retained analytes lay almost entirely in the low-concentration fraction of the plasma proteome. In a study published in *PLOS ONE* in 2026, the same procedure applied to urine yielded a nine-protein bladder cancer panel at AUC 0.970 and extended to neurologic and metabolic indications.¹⁵ Together these studies indicate that the approach generalizes along two axes. It is not specific to a single disease: two sex-specific panels separated cancer from non-cancer across 18 tumor types, and the same procedure has been applied outside oncology. It is not specific to a single biofluid: selection performed comparably in urine and in plasma. The analyte pool, selection procedure and final panel size used in the present work are those employed in both studies. Both are single-cohort studies without external validation, and neither addressed discrimination of malignant from benign disease at a biopsy decision point; they establish the method, not the claim under test here, which is why the present request is for blinded evaluation in independently held material.

Development proceeds in three conditional stages. Aim 1 transfers the panel from the discovery platform to the Siomx CMOS biochip analyzer, a time-gated fluorescence platform running a closed-tube multiplexed immuno-PCR assay and reporting concentration in pg/mL in under one hour at the point of care, at a cost below \$100 per test. This aim replaces relative with absolute quantification, establishes analytical performance against CLSI guidelines, and fixes analytes, coefficients and threshold in a signed, dated lock document. The locked operating point is sensitivity $\geq 97\%$ with specificity $\geq 40\%$. Specificity is expressed as a floor rather than a target: at this prevalence sensitivity governs safety and specificity governs the magnitude of benefit. Across the range of malignancy rates reported for BI-RADS 4A/4B, that operating point leaves a residual malignancy risk of 1.8% to 3.1% after a negative result while deferring approximately 30% of biopsies. At the lower end of that range the residual risk falls below the 2% threshold at which BI-RADS recommends short-interval surveillance in preference to tissue sampling. Tissue diagnosis remains indicated above that level.^{1, 16}

Aim 2, for which the present request is made, applies that locked assay unmodified and blinded to the reference set, with no aliquot opened until the lock is signed and the analysis plan filed. Aim 3 completes characterization of the manufactured product, including between-lot and between-instrument reproducibility and a reference interval, and evaluates the test prospectively at breast imaging practices under the same intended-use definition. Specimens used to build or threshold the assay are procured separately in Aim 1; the reference set enters at one point only and contributes to no other stage.

3. Preliminary Data and Methods

Cohort. The preliminary data below come from a separate and independent study from the multi-cancer work published in *BMJ Oncology*;¹⁴ the cohorts do not overlap, and no specimen contributes to both. EDTA plasma from 480 women: 226 with invasive breast cancer (stage I, 51; stage II, 115; stage III, 60), 87 with benign breast disease (fibroadenoma 44, cyst 21, inflammation 22) and 167 healthy blood donors. Tumor size among cases was 1–10 mm in 31, 10–20 mm in 39, 20–50 mm in 96 and greater than 50 mm in 60. All cancer cases were treatment-naïve, with plasma collected before resection or any therapy. Specimens were processed under a common standard operating procedure with quality control at collection, storage and shipment. A separate cohort of 115 women, used in neither selection nor training, served as replication.

Measurement and quality control. Olink Explore 3072: 3,072 assays covering approximately 2,900 proteins across eight 384-plex panels, each carrying IL6, CXCL8 and TNF controls. All discovery specimens were run in a single batch. Internal incubation, extension and amplification controls monitored intra-assay variability; a buffer negative control and pooled-plasma plate controls supported limit-of-detection assignment and inter-plate normalization; two known specimens per plate served as sample controls. Values were reported as normalized protein expression (NPX) on a $\log_2$ scale.

Selection procedure. L1-penalised logistic regression was fitted across 100 bootstrap resamples; proteins with the highest count of non-zero coefficients were retained, the penalty enforcing sparsity and limiting co-selection of correlated analytes. Performance was assessed by five-fold cross-validation, implemented in scikit-learn.¹⁷

Table 1. Discovery and replication performance.

Comparison	n	AUC	Dispersion and notes
Breast cancer vs. biopsy-referred benign breast disease (20-analyte panel)	226 vs. 87	0.905	0.843–0.960 across 50 cross-validation iterations; sensitivity 0.97 at specificity ≈0.43
Independent replication cohort (benign disease and healthy controls)	115	0.930	Used in neither selection nor training
Reduced 10-analyte panel vs. pooled non-cancer comparator	—	0.932	Basis for the pre-locking panel-reduction step; not evaluated against benign disease alone

Stated limitations of the preliminary data. The discovery set derives from a single geographic source with an ancestry distribution unrepresentative of the US screening population; cases and benign comparators were obtained from a biobank while healthy controls were blood donors, so provenance is partly confounded with outcome; BI-RADS assessment was not recorded, so the intended-use population cannot be identified within the cohort; quantification was relative, on a discovery platform rather than the product; the work was retrospective and single-batch; and benign comparators were predominantly fibroadenoma and cyst. The requested study addresses the first three simultaneously, since all strata derive from a single clinical pathway at four US centers, with confounding handled by adjustment and pre-specified effect-modifier testing. Before any reference-set aliquot is thawed, the panel is transferred to the deployable analyzer and the assay locked — analytes, coefficients and threshold, signed and dated — with absolute quantification in pg/mL replacing the relative discovery scale.

4. Data Analysis Plan

Training set requirement. None. Analytes, coefficients and decision threshold are fixed in a signed lock document before any reference-set specimen is opened, and a blinded test set only is requested. The decision threshold, and every fitting step preceding the lock, are established on prospectively procured CHTN specimens and residual discovery aliquots held by Noveln; no reference-set specimen contributes to any of them. Specimens are requested in one plate-randomized shipment labelled by identifier alone, accompanied by a covariate file carrying age at draw, collection site, calendar year of collection, storage duration, freeze–thaw count and, where recorded, BI-RADS assessment and subcategory, breast density, benign lesion histology, and race and ethnicity. Noveln personnel remain blinded through measurement and analysis; the analysis is executed against a plan filed with the DMCC before shipment, results are filed in writing before outcome disclosure, and no interim analysis is performed.

Cohort composition. Stratum counts are those published for the reference set: incident invasive cancer n=190, carcinoma in situ n=55, benign pathology with atypia n=63, benign disease without atypia n=231, and women with no evidence of breast disease at screening mammography (BI-RADS 1 or 2) n=276.⁶ The primary analysis comprises referral-venue women with venipuncture preceding tissue sampling: invasive carcinoma and carcinoma in situ as cases (n=245), and benign disease without atypia as comparators (n=231), 476 in total. Two groups are excluded from the primary analysis and retained as secondary strata: benign with atypia, since atypical hyperplasia proceeds to excision irrespective of a blood test result and deferral is therefore not the clinical question; and the screening-venue women, who do not constitute a biopsy-referral population.

Variables requested with the specimens. The following are requested per specimen in the masked covariate file, listed in order of importance to the analysis. (i) BI-RADS assessment category and, critically, category 4 subcategory (4A, 4B or 4C), since the intended use is confined to 4A and 4B and the primary analysis is defined on that population; this variable determines whether the primary claim can be stated at subcategory level or must be stated at pooled BI-RADS 4. (ii) Index-biopsy pathology, as reference standard, with histological type of benign lesions. (iii) Age at draw. (iv) Collection site and calendar year of collection. (v) Storage duration and freeze–thaw count. (vi) Breast density. (vii) For cancer cases, invasive versus in situ, stage and receptor status. (viii) Race and ethnicity. (ix) Menopausal status and, where recorded, pre-analytic intervals (draw-to-spin, time-to-freeze). Items (i) through (v) are required for the pre-specified primary and adjusted analyses. Items (vi) through (ix) support pre-specified descriptive strata and are analyzed as available.

Contingency if subcategory is unavailable. If BI-RADS was recorded only as category 4 without subdivision, the primary claim is stated at pooled BI-RADS 4, the departure from the intended-use definition is declared in the analysis plan before shipment, and subcategory-specific performance is deferred to the prospective study

described in Section 5, where subcategory is captured by protocol. This contingency is stated in advance so that the analysis is not altered after the composition of the received data is known. We ask the DMCC to confirm at the time of approval which of items (i) through (ix) are available, so that the plan can be finalized before release.

Endpoints, sample size and power. Co-primary endpoints are sensitivity and specificity of the locked binary result at the locked threshold, with the pathology of the index biopsy as reference standard. Analysis follows one-sided exact binomial tests at $\alpha = 0.025$ with Clopper–Pearson intervals throughout,¹⁸ against pre-specified nulls in a fixed hierarchy, with alpha spent only on the whole-cohort comparison. Against a sensitivity null of 0.90, 245 cases provide 99.7% power at a true sensitivity of 0.97; against a specificity null of 0.30, 231 comparators provide 88.4% power at a true specificity of 0.40. Precision at those values is ± 0.023 for sensitivity (95% CI 0.942–0.988) and ± 0.065 for specificity (0.335–0.465). The requested $n = 476$ is the entire referral-eligible portion of the set and is sufficient for the co-primary comparison; all subgroup and sensitivity analyses are consequently declared descriptive. The design follows the standards for pivotal biomarker evaluation set out by Pepe and colleagues¹⁹ in the respects this resource permits, and departs from them in one, noted below.

Predictive values and re-weighting. The set is biobank-assembled from the intended-use population and therefore constitutes a nested case-control sample: sensitivity and specificity transfer directly, subject to case mix, whereas prevalence does not. The primary clinical endpoint, residual malignancy risk following a negative result, is reported both as observed and re-weighted to published BI-RADS 4A/4B malignancy rates.² Secondary endpoints are AUC with DeLong intervals predictive values, calibration slope and intercept with flexible calibration plots, and decision-curve net benefit against biopsy of all referred women.²⁰

Covariates, fixed in advance in three tiers. Adjusted: age, collection site, storage duration, batch and run, entered in a logistic model of the continuous risk score, from which adjusted AUC and adjusted operating-point performance are reported alongside the unadjusted primary result; site and storage duration carry the greatest weight, provenance confounding being the specific weakness this study exists to remove. Stratified: invasive versus in situ, stage, receptor status, age above and below 50, benign lesion type, and, where recorded, BI-RADS subcategory and breast density, each with exact intervals and interaction tests. Sensitivity analyses: excluding the top decile of storage duration; excluding the largest single collection site; counting every invalid result as a failure rather than a missing value; and restricted to invasive carcinoma alone. Missing covariates are multiply imputed; analyte values are never imputed. Race and ethnicity are reported in full as a pre-specified descriptive stratum.

Invalid results. Three-level controls accompany every plate, and any plate failing control criteria is repeated in full. A specimen returning a non-reportable analyte returns an invalid result, counted against the invalid rate and repeated once from a second aliquot where volume permits, under a repeat policy fixed in the analysis plan before shipment. The requested volume accommodates one repeat.

Statistical resources. Analysis is performed in-house by the Novena data science group, which conducted the discovery analyses described above; code is version-controlled and timestamped against the filed plan and will be supplied with the results. EDRN statistical support is not required beyond DMCC blinding, plate randomization and custody of the outcome key. Deliverables to the DMCC within six months of receipt comprise sensitivity, specificity, negative predictive value, residual malignancy risk both observed and re-weighted, calibration, and decision-curve net benefit, with every pre-specified stratum and sensitivity analysis reported irrespective of direction.

5. Future Plans

Approach to EDRN for Phase II funding and collaboration. An SBIR-funded program covering CLSI analytical validation, prospective in-practice evaluation and an FDA pre-submission is already resourced, so EDRN funding is not sought for that work. What it will not deliver is multi-center prospective validation under independent specimen custody, and for that we would seek collaboration with EDRN in a subsequent Phase II study.

Working within the EDRN collaborative framework. We are amenable to doing so, including protocol development with EDRN biostatisticians, specimen custody and blinding by the DMCC, and joint publication. The design submitted here adopts EDRN conventions — an assay locked before evaluation, prospective specimen collection with retrospective blinded assessment, and pre-specified analysis — independently of this request.

Inclusion in a larger multi-marker panel. We are amenable. The test is itself a multivariate panel, so combination is a methodological extension rather than a departure, and the appropriate comparison is whether a combined model improves net benefit over either component alone. The continuous risk score would be supplied for that purpose. Analyte identities are proprietary and would be shared under an appropriate

agreement; guidance from the DMCC on the form such an agreement should take would be welcome, and the matter is raised now rather than at the point of results posting.

Further development if refinements improve performance. We concur. Two refinements are already pre-specified: recalibration-in-the-large if calibration alone fails, with panel and coefficients unchanged; and a panel-size reduction settled before the lock. Neither is permitted to affect the reference-set result, and no refitting occurs on these specimens under any outcome. EDRN resources would be advantageous where our own data are weakest: ancestry and practice-setting diversity, benign case mix beyond fibroadenoma and cyst, and BI-RADS-subcategory-specific performance.

Reporting in the event of failure. A sensitivity failure is reported as a negative validation and published as such. Given the prior result obtained in this set, a negative outcome for a multivariate low-abundance panel constitutes an informative result and will be reported in full.

References

1. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA, et al. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. 5th ed. Reston, VA: American College of Radiology; 2013.
2. Elezaby M, Li G, Bhargavan-Chatfield M, Burnside ES, DeMartini WB. ACR BI-RADS assessment category 4 subdivisions in diagnostic mammography: utilization and outcomes in the National Mammography Database. *Radiology*. 2018;287(2):416-422.
3. Vlahiotis A, Griffin B, Stavros AT, Margolis J. Analysis of utilization patterns and associated costs of the breast imaging and diagnostic procedures after screening mammography. *Clinicoecon Outcomes Res*. 2018;10:157-167.
4. Lin LLY, Gao Y, Lewin AA, Toth HK, Heller SL, Moy L. Overstated harms of breast cancer screening? A large outcomes analysis of complications associated with 9-gauge stereotactic vacuum-assisted breast biopsy. *AJR Am J Roentgenol*. 2019;212(4):925-932.
5. Gram EG, Siersma V, Brodersen JB. Long-term psychosocial consequences of false-positive screening mammography: a cohort study with follow-up of 12-14 years in Denmark. *BMJ Open*. 2023;13(4):e072188.
6. Marks JR, Anderson KS, Engstrom P, et al. Construction and analysis of the NCI-EDRN breast cancer reference set for circulating markers of disease. *Cancer Epidemiol Biomarkers Prev*. 2015;24(2):435-441.
7. Fueger BJ, Clauser P, Kapetas P, Pötsch N, Helbich TH, Baltzer PAT. Can supplementary contrast-enhanced MRI of the breast avoid needle biopsies in suspicious microcalcifications seen on mammography? A systematic review and meta-analysis. *Breast*. 2021;56:53-60.
8. Fueger BJ, Tollens F, Kaiser CGN, Helbich TH, Pötsch N, Clauser P, Baltzer PAT. Cost-effectiveness of contrast-enhanced breast MRI in suspicious mammographic microcalcifications. *Insights Imaging*. 2025;16:122.
9. van der Poort EKJ, van Ravesteyn NT, van den Broek JJ, de Koning HJ. The early detection of breast cancer using liquid biopsies: model estimates of the benefits, harms, and costs. *Cancers (Basel)*. 2022;14(12):2951.
10. Klein EA, Richards D, Cohn A, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann Oncol*. 2021;32(9):1167-1177.
11. Harris L, Fritsche H, Mennel R, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol*. 2007;25(33):5287-5312.
12. Lourenco AP, Benson KL, Henderson MC, et al. A noninvasive blood-based combinatorial proteomic biomarker assay to detect breast cancer in women under the age of 50 years. *Clin Breast Cancer*. 2017;17(7):516-525.e6.
13. Chen C, Poole L, Sizer L, Carter WB, Frazier TG. Diagnostic accuracy of the Videssa protein-based liquid biopsy for breast cancer in suspicious mammography [abstract]. *J Clin Oncol*. 2019;37(15 suppl):e13034.
14. Budnik B, Amirkhani H, Forouzanfar MH, Afshin A. Novel proteomics-based plasma test for early detection of multiple cancers in the general population. *BMJ Oncol*. 2024;3(1):e000073.
15. Budnik B, Amirkhani H, Weinberger K, Nedjat S, Forouzanfar MH, Afshin A. The promise of deep urine proteomics for diagnosis of cancer, neurologic, and metabolic diseases. *PLoS One*. 2026;21(7):e0354808.
16. Berg WA, Berg JM, Sickles EA, et al. Cancer yield and patterns of follow-up for BI-RADS category 3 after screening mammography recall in the National Mammography Database. *Radiology*. 2020;296(1):32-41.
17. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
18. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*. 1934;26(4):404-413.
19. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *J Natl Cancer Inst*. 2008;100(20):1432-1438.
20. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574.