GMR
The Original

INTEGRATED BIOINFORMATICS ANALYSIS AND DEVELOPMENT OF A MULTI-BIOMARKER WEIGHTED CLINICAL DECISION SCORE FOR BREAST CANCER RISK STRATIFICATION USING ANGIOGENIC AND INFLAMMATORY SIGNATURES

Madu Pascal Akaye[1,2*], Abah Moses Owoicho[1,3], Angelina V. Zhilenkova[1], Prakash Raj Govindaraj[2], Marina Prokhorova[1], Natalia M. Nikitina[1], Alexandra Lebedeva[4,5], Ekaterina Orlova[1], Marina I. Sekacheva[1]

[1]Institute for Personalized Oncology, Biomedical Science & Technology Park I.M Sechenov, First Moscow State Medical University of the Ministry of Health of Russia (Sechenov University), 8-2 Trubetskaya Street, Moscow, Russian Federation, 119991
[2]Phystech School of Biological and Medical Physics, Moscow Institute of Physics and Technology.
[3]Royal Society of Clinical and Academic Researchers International (ROSCAR)
[4]Sechenov University, Institute for Personalized Oncology, Laboratory of Applied Genomic Technologies
[5]OncoAtlas LLC
*Corresponding Author, Moscow, Russia. Email: akayepaschal48@gmail.com

ABSTRACT

Background: Breast cancer is characterized by complex interactions between angiogenesis, inflammation, and extracellular matrix remodeling. Single biomarkers lack sufficient diagnostic accuracy, necessitating multi-marker approaches.

Methods: Blood samples from breast cancer patients and healthy controls were analyzed at the Institute of Personalized Oncology, Sechenov University. Circulating inflammatory and angiogenic biomarkers were quantified using the Luminex MAGPIX multiplex platform. 5 biomarkers—Osteopontin, sVEGFR2, sNeuropilin-1, suPAR and Tenascin-C were selected to develop a weighted Clinical Decision Score. Protein–protein interaction analysis of Breast Invasive Carcinoma data from cBioPortal was performed to investigate the molecular interactions underlying the various genes of the selected biomarkers.

Results: The weighted Clinical Decision Score (wCDS) demonstrated enhanced biological interpretability by integrating complementary pathways associated with tumor angiogenesis, inflammation, and extracellular matrix remodeling. Increasing scores reflected progressive activation of multiple oncogenic processes and improved biological discrimination compared with equal-weight scoring approaches. Our network analysis identified THBS1–SPP1–PLAU signal axis as a novel pathway in breast cancer pathology, alongside TGFB1 as the principal regulatory hub, while SPP1 (osteopontin gene) emerged as a central connector linking angiogenesis, inflammation, ECM remodeling, and invasion through a TGFB1–SPP1–PLAU signaling axis.

Conclusion: This study supports the potential clinical utility of the wCDS as a blood-based risk stratification tool for breast cancer. It also highlights the need for further investigation of the TGFB1–SPP1–PLAU signaling axis as a promising therapeutic pathway and the evaluation of SPP1 as a potential drug target. Furthermore, it advocates for the development of SPP1-targeted therapies for breast cancer by the pharmaceutical industry.

KEYWORDS: Breast cancer; Multiplex immunoassay; Osteopontin; SPP1; Biomarkers; Weighted Clinical Decision Score (wCDS)

1.0 INTRODUCTION

Breast cancer remains the most commonly diagnosed cancer in women worldwide and a leading cause of cancer-related death, with GLOBOCAN 2022-based analyses estimating about 2.30 million new cases and roughly 666,000 deaths globally in 2022, while also projecting a substantial future increase in burden, particularly in settings with constrained health-system capacity [1–4]. Although advances in mammographic screening, molecular pathology, and targeted systemic therapy have improved outcomes in many high-income settings, breast cancer continues to pose a major public health challenge because of its marked molecular heterogeneity, diverse clinical behavior, and persistent risk of recurrence and metastasis [5–11]. Importantly, incidence and survival gains have not been equitably distributed, and poorer outcomes in low- and middle-income countries remain closely linked to delayed diagnosis, limited screening coverage, and restricted access to specialized treatment infrastructure [1–11].

Breast cancer susceptibility arises from the interplay of inherited, hormonal, reproductive, metabolic, environmental, and lifestyle factors. Recent evidence syntheses and umbrella reviews continue to support major roles for pathogenic variation in high-penetrance susceptibility genes such as BRCA1, BRCA2, and PALB2, as well as for age, adiposity, physical inactivity, alcohol exposure, reproductive timing, cumulative estrogen

exposure, and mammographic breast density as important determinants of risk [12–16]. These risk determinants do not act in isolation; rather, they converge on a biologically complex disease process in which malignant epithelial transformation is shaped by reciprocal interactions with stromal, vascular, and immune compartments of the breast tumor microenvironment [6–9,17–20].

Increasing evidence indicates that breast cancer progression is governed not only by tumor-intrinsic genetic and epigenetic alterations but also by dynamic bidirectional crosstalk among cancer cells, fibroblasts, endothelial cells, macrophages, lymphocytes, and extracellular matrix components within the tumor microenvironment [6–9,17–24]. This integrated ecosystem regulates many core hallmarks of cancer, including sustained proliferation, resistance to cell death, angiogenesis, immune evasion, invasion, metastatic dissemination, and therapeutic response [6–11,17–24]. Accordingly, biomarkers that capture several of these interconnected processes may offer greater clinical utility than markers tied to a single pathway alone [5,10,11,25,26].

Among the biological programs that sustain breast cancer progression, angiogenesis remains essential for tumor expansion and metastatic spread. Hypoxia-driven signaling promotes vascular endothelial growth factor A (VEGF-A) production and activation of VEGF receptor pathways, especially VEGFR2, thereby stimulating endothelial proliferation, migration, and neovascularization through downstream PI3K/Akt and MAPK signaling networks [21,27–31]. Neuropilin-1 (NRP1), a critical VEGF co-receptor, further potentiates VEGFR2 signaling and has been implicated in the regulation of endothelial trafficking dynamics and angiogenic responsiveness [21,27–31]. Conversely, soluble VEGFR2 can function as a circulating decoy receptor that restrains VEGF bioavailability, and dysregulation of the VEGF/VEGFR signaling axis has repeatedly been linked to aggressive disease biology, progression, and resistance to therapy in breast cancer and related tumor contexts [21,27–31].

Chronic inflammation is likewise a central driver of tumor progression. Persistent inflammatory signaling fosters genomic instability, supports angiogenesis, remodels extracellular structures, and promotes immune dysfunction within the tumor niche [6–9,17–24,32–35]. In this context, soluble urokinase plasminogen activator receptor (suPAR), derived from cleavage of membrane-bound PLAUR/uPAR, is increasingly recognized as a stable circulating indicator of systemic immune activation and proteolytic-inflammatory signaling, with broader oncology literature supporting its association with adverse disease behavior and poor prognosis [32–35]. Within breast cancer specifically, the inflammatory microenvironment—and particularly macrophage-centered programs—has emerged as a major determinant of invasion, metastatic competence, and therapeutic resistance [18,19,22–24,33–39].

Extracellular matrix remodeling is another defining feature of breast cancer progression. Rather than serving as a passive scaffold, the extracellular matrix actively shapes tumor mechanics, cell-state plasticity, invasion, and metastatic colonization [8,9,18,20,24,40–42]. Tenascin-C (TNC), a matricellular glycoprotein induced in tissue injury, inflammation, and cancer, is now widely regarded as an important stromal effector in tumor-promoting microenvironments. Recent reviews emphasize that TNC contributes to epithelial-mesenchymal transition, stemness, immune modulation, and metastatic niche formation, and that its overexpression is consistently associated with more aggressive clinicopathologic phenotypes and poorer outcomes across cancers, including breast cancer [40–42].

Osteopontin, encoded by SPP1, has emerged as one of the most multifunctional mediators linking inflammation, matrix remodeling, angiogenesis, and immune suppression. As a secreted matricellular phosphoprotein, Osteopontin regulates adhesion, migration, survival, extracellular matrix interactions, and pro-tumor signaling through integrins and CD44-associated pathways [43–49]. Recent work has further highlighted the importance of SPP1-expressing macrophages as a functionally distinct immunosuppressive population within the tumor microenvironment, capable of interacting with fibroblasts, regulatory T cells, and exhausted T-cell states to reinforce immune exclusion and treatment resistance; these mechanisms appear especially relevant in aggressive breast cancer phenotypes, including triple-negative disease [36–39,44–49]. In parallel, experimental and translational studies continue to associate elevated Osteopontin/SPP1 activity with metastatic behavior, therapy resistance, and inferior survival, underscoring its appeal as both a biomarker and a candidate therapeutic target [43–49].

Despite substantial progress in biomarker discovery, most currently used approaches still rely on isolated markers or single-pathway readouts, which inadequately reflect the biological interdependence of angiogenesis, inflammation, immune regulation, and extracellular matrix remodeling in breast cancer [5,10,11,25,26,49]. A multi-biomarker framework that incorporates biologically complementary signals and weights them according to their independent predictive contribution may therefore offer greater clinical interpretability and stronger discriminatory performance than unweighted panels. On this basis, the present study proposes a weighted Clinical Decision Score (wCDS) integrating five circulating biomarkers—Osteopontin (SPP1), soluble VEGFR2, soluble Neuropilin-1, suPAR, and Tenascin-C—to improve breast cancer risk stratification. To further support the biological rationale of this framework, we combined circulating biomarker analysis with systems-level bioinformatics using Breast Invasive Carcinoma datasets from cBioPortal. In doing so, we aimed to define

whether a biologically weighted composite score could better capture the convergent pathobiology of tumor angiogenesis, inflammation, and matrix remodeling than conventional single-marker strategies. Our network-based findings further identify TGFB1 as a central regulatory hub and support SPP1 as a key bridge linking angiogenesis, inflammation, and extracellular matrix remodeling through a TGFB1–SPP1–PLAU signaling axis, thereby providing mechanistic justification for differential biomarker weighting within the proposed wCDS model [36–51].

2.0 PATIENTS AND METHODS

2.1 Study Design

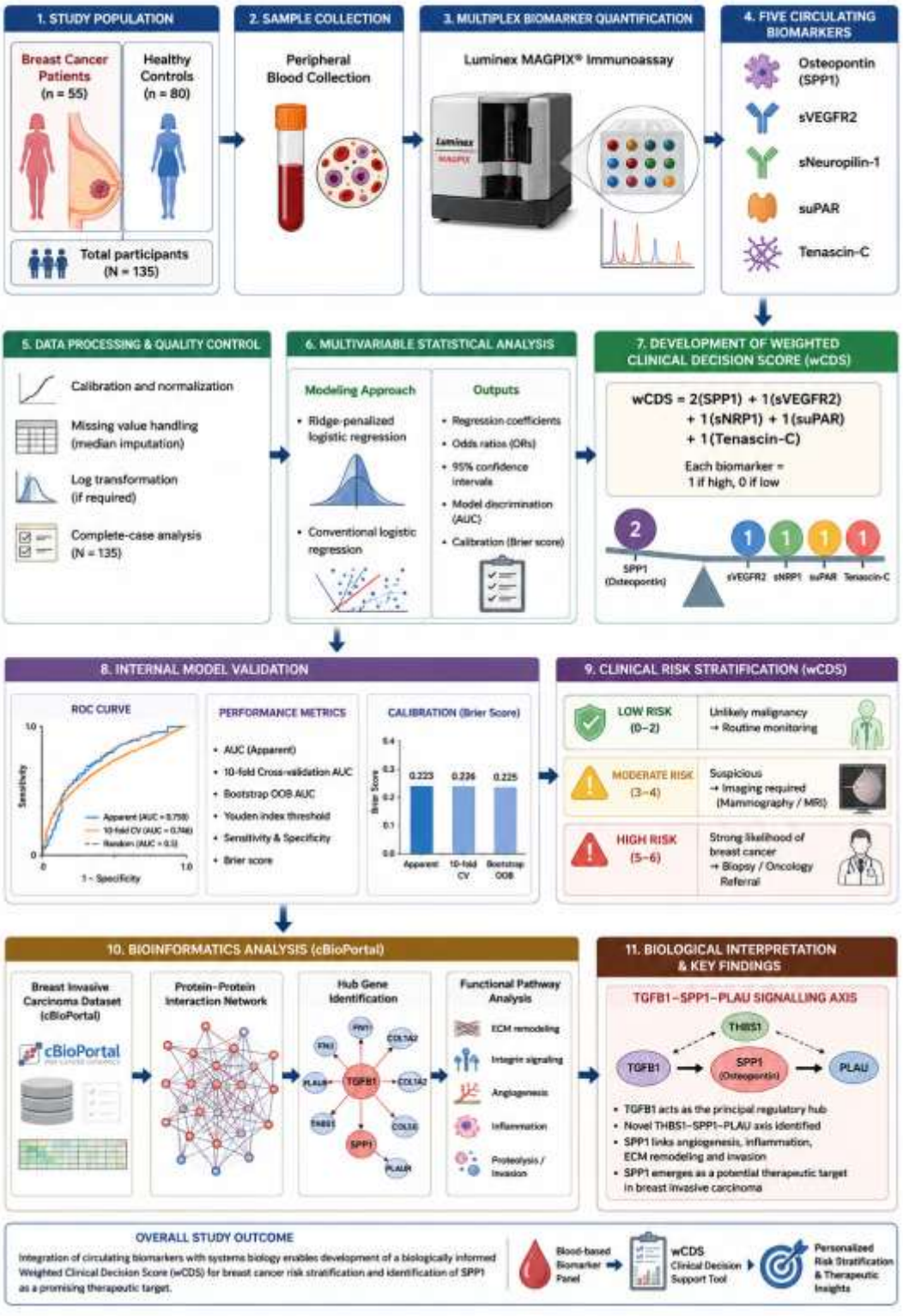

Figure1. Study Workflow

2.1.1 Multi-Biomarkerweighted Clinical Decision Score (Wcds) Workflow

A retrospective case-control study was conducted to develop and evaluate a multibiomarker clinical prediction model for breast cancer risk stratification. The study dataset comprised 135 serum biomarker measurements

obtained from women with 55 histologically confirmed breast cancer and 80 healthy control subjects. Breast cancer status served as the binary outcome variable, coded as 1 for breast cancer cases and 0 for healthy controls.
Biomarker Selection
Five circulating biomarkers were selected based on their established involvement in key biological pathways associated with breast cancer progression, angiogenesis, chronic inflammation, extracellular matrix remodeling, and tumor invasion. The biomarkers included Osteopontin (SPP1), soluble vascular endothelial growth factor receptor-2 (sVEGFR2), soluble Neuropilin-1 (sNeuropilin-1), soluble urokinase plasminogen activator receptor (suPAR), and Tenascin C.
Osteopontin was selected as the principal marker because of its central role in tumor progression, metastatic dissemination, immune regulation, and epithelial–mesenchymal transition. sVEGFR2 and sNeuropilin-1 were included as complementary angiogenic biomarkers involved in vascular endothelial growth factor signaling. suPAR was selected to represent inflammatory activation and tumor invasiveness, whereas Tenascin C was incorporated as an indicator of extracellular matrix remodeling and stromal activation.

2.1.2 Bioinformatics Analyses

To provide mechanistic validation, genes associated with these biomarkers and related breast cancer pathways were analyzed using Breast Invasive Carcinoma datasets obtained from cBioPortal, followed by protein–protein interaction network analysis to identify central regulatory hubs, functional signaling modules, and candidate therapeutic targets.

2.2 Data Processing

Peripheral blood samples from 135 participants (breast cancer patients and healthy controls) were processed at the Institute for Personalized Oncology, Sechenov University. Plasma concentrations of Osteopontin (SPP1), soluble VEGFR2, soluble Neuropilin-1, soluble urokinase plasminogen activator receptor (suPAR), and Tenascin-C were quantified using a Luminex MAGPIX multiplex immunoassay system (xPONENT 4.2, Luminex Corporation, USA) according to standardized manufacturer protocols. Raw fluorescence intensity values were converted into concentration units using calibration curves generated from known standards. Quality control procedures included exclusion of samples with coefficient of variation >15% and normalization to reduce inter-assay variability. Biomarker distributions were assessed for normality, and non-normally distributed variables were log-transformed where appropriate prior to downstream analyses. Missing values were minimal and handled using median imputation to preserve sample size without introducing distributional bias.
Biomarker data were screened for completeness and consistency before statistical analysis. Laboratory measurements reported as values below or above assay detection limits were converted to their respective numeric detection-limit values to permit quantitative analysis. Records containing missing values for any of the selected biomarkers were excluded from multivariable analyses using a complete-case approach. Continuous biomarker variables were retained in their original measurement units without categorization to preserve statistical power and avoid information loss.

Development of the Multivariable Prediction Model

A multivariable logistic regression model was developed to estimate the probability of breast cancer. Breast cancer status served as the dependent variable, while Osteopontin, sVEGFR2, sNeuropilin-1, suPAR, and Tenascin C were entered simultaneously as independent variables. Because evidence of complete or near-complete separation was encountered during conventional maximum likelihood estimation, a penalized logistic regression approach with ridge (L2) regularization was adopted to obtain stable coefficient estimates and reduce estimation bias. Regression coefficients were transformed into odds ratios (ORs) with corresponding 95% confidence intervals where estimable.

Development of the Weighted Clinical Decision Score

To facilitate clinical application, the multivariable prediction model was translated into a weighted Clinical Decision Score (CDS) (table 1). Biomarker weights were assigned according to their biological relevance in breast cancer pathogenesis rather than assigning equal importance to all predictors. The weighting scheme was defined as follows:

Table 1. Weighted Clinical Decision Score

Biomarker	Biological Function	Assigned Weight
Osteopontin	Tumor progression and metastasis	2
sVEGFR2	Angiogenesis regulation	1
sNeuropilin-1	VEGF co-receptor	1
suPAR	Chronic inflammation and invasion	1
Tenascin C	Extracellular matrix remodeling	1

Model Performance Evaluation

Model discrimination was evaluated using receiver operating characteristic (ROC) curve analysis. The area under the ROC curve (AUC) was calculated to quantify the model's ability to distinguish breast cancer cases from

healthy controls. The optimal probability threshold for classification was determined using the Youden Index, and corresponding sensitivity and specificity were calculated.
Overall prediction accuracy was additionally assessed using the Brier score, with lower values indicating better agreement between predicted probabilities and observed outcomes.

2.3 Statistical Analysis

Statistical analyses were performed using multivariable logistic regression methods. Continuous variables were analyzed as quantitative measurements. Regression coefficients, odds ratios, and model discrimination statistics were obtained from the fitted prediction model. Two-sided statistical tests were used throughout, and a p-value < 0.05 was considered statistically significant.
The final multibiomarker model was developed to provide a clinically interpretable decision-support tool integrating complementary molecular pathways involved in breast cancer biology, including angiogenesis, inflammation, extracellular matrix remodeling, and metastatic progression.
All statistical analyses were performed using standard statistical software environments suitable for biomedical data analysis.

2.4 Predictive Modeling

Predictive modeling was performed to develop a robust diagnostic framework for breast cancer classification. Two machine learning approaches were applied: logistic regression and random forest classification. Feature selection techniques were implemented to identify the most informative biomarkers contributing to classification performance. Model training was conducted using the five selected biomarkers: SPP1, sVEGFR2, sNRP1, suPAR, and Tenascin-C.

3.0 RESULTS

3.1 Weighted Clinical Decision Score Analysis

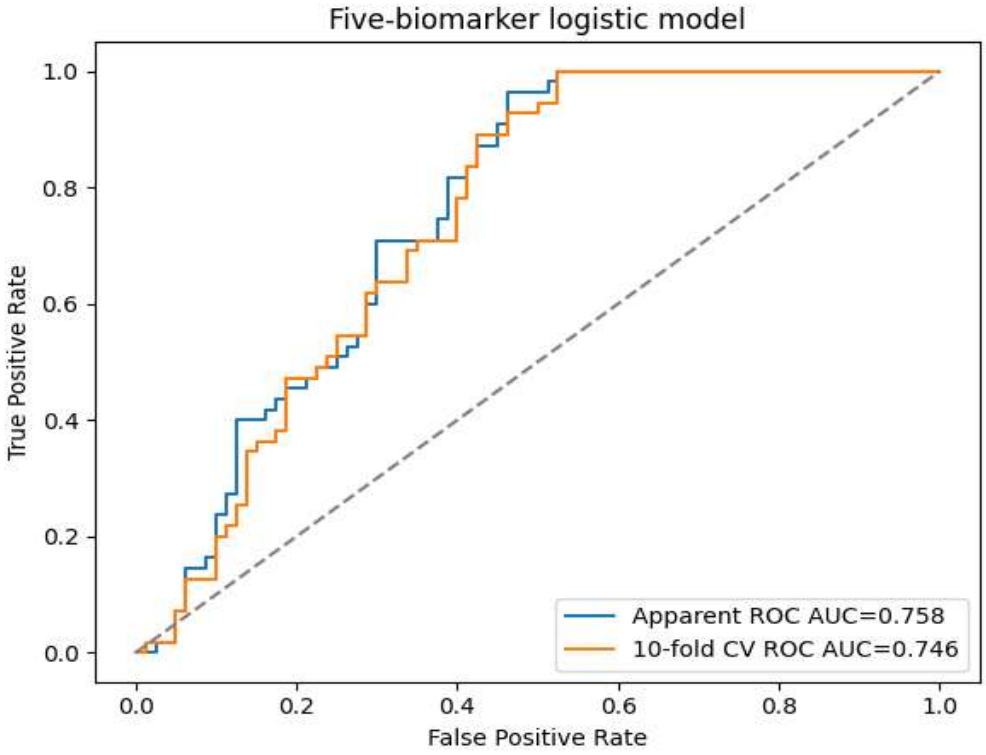

Figure 2. Receiver operating characteristic (ROC) curves for the five-biomarker logistic regression model. The blue curve represents the apparent model performance (AUC = 0.758), while the orange curve shows the performance after 10-fold cross-validation (AUC = 0.746). The dashed diagonal line indicates the performance of a random classifier (AUC = 0.5). The close agreement between the apparent and cross-validated ROC curves suggests good model stability with only minimal overfitting.

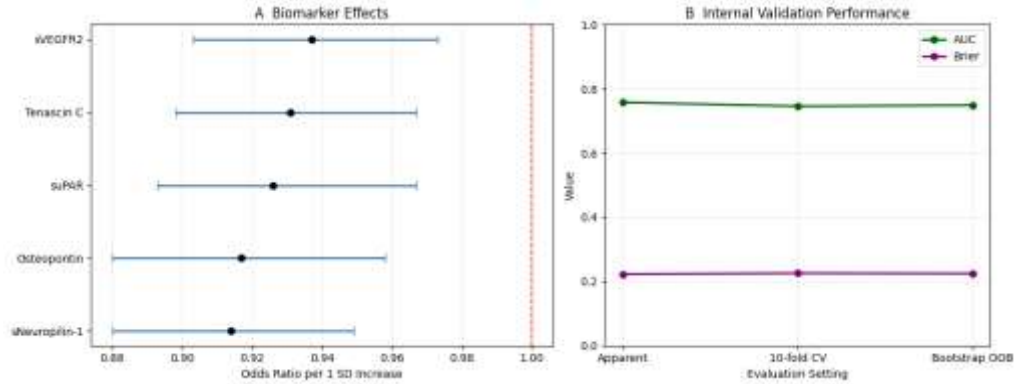

Figure 3. Panel A shows the adjusted biomarker effects from the penalized multivariable model as odds ratios with 95% confidence intervals. All five biomarkers have odds ratios below 1.0, indicating inverse associations with breast cancer in the joint model. The effect sizes are fairly close together, but sNeuropilin-1 and Osteopontin appear to have the strongest relative effects.
Panel B shows the model's performance across the three evaluation settings. The AUC remains tightly clustered around the mid-0.74 range, while the Brier score stays near 0.22 to 0.23. The closeness of the apparent, cross-validated, and bootstrap estimates supports the interpretation that the model has moderate discrimination with limited internal optimism.

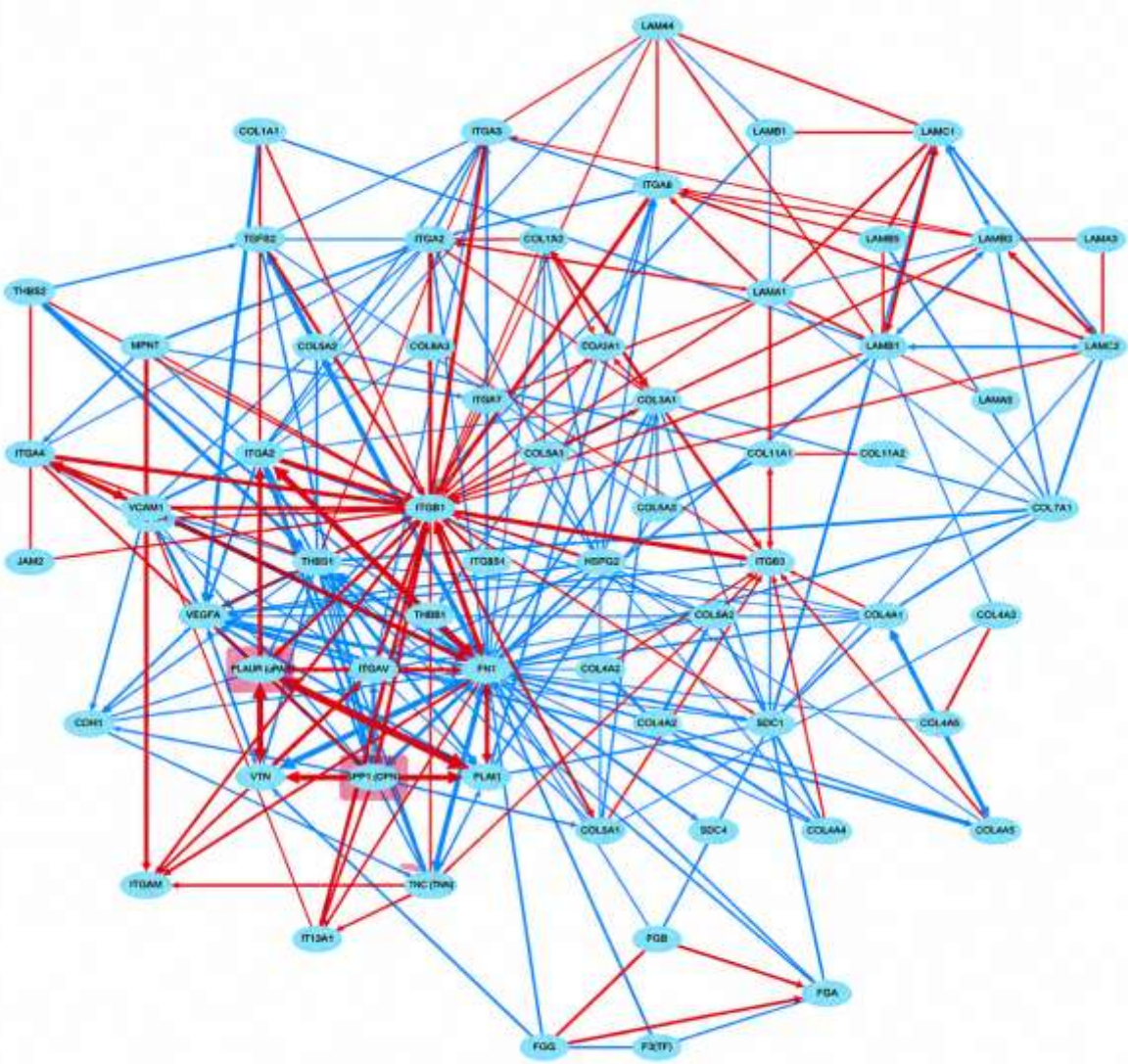

Figure 4. Protein–protein interaction network derived from altered genes in breast invasive carcinoma obtained from cBioPortal. Nodes represent genes/proteins, whereas edges indicate predicted or experimentally validated interactions. Red and blue edges denote positive and negative interaction patterns, respectively. The network is dominated by extracellular matrix, integrin, and TGF-β signaling components, with TGFB1 functioning as the

principal hub. A highly interconnected novel THBS1, SPP1, and PLAUR axis alongside TGFB1–SPP1–PLAU axis were dentified, highlighting SPP1 as a potential therapeutic target in breast invasive carcinoma.

Table 2. Penalized Logistic Regression Results

Biomarker	Coefficient	Odds Ratio	Bootstrap 95% CI
Osteopontin	−0.086	0.917	0.880–0.958
sVEGFR2	−0.065	0.937	0.903–0.973
sNeuropilin-1	−0.090	0.914	0.880–0.949
suPAR	−0.077	0.926	0.893–0.967
Tenascin C	−0.071	0.931	0.898–0.967

Table 3. Unpenalized Multivariable Logistic Regression Results

Biomarker	Coefficient (β)	95% CI	p value
Osteopontin	−0.704	−1.348 to −0.060	0.032
sVEGFR2	−0.310	−0.986 to 0.366	0.368
sNeuropilin-1	−0.547	−1.157 to 0.063	0.079
suPAR	−0.292	−0.880 to 0.297	0.332
Tenascin C	−0.300	−0.914 to 0.314	0.338

Table 4. Model Performance

Metric	Value
Apparent AUC	0.758
10-fold CV AUC	0.746
Bootstrap OOB AUC	0.749
Youden threshold	0.481
Sensitivity	96.4%
Specificity	53.8%
Apparent Brier score	0.223
10-fold CV Brier score	0.226
Bootstrap OOB Brier score	0.225

4.0 DISCUSSION

Multivariable Logistic Regression and Development of the Weighted Clinical Decision Score

In the multivariable analysis, complete biomarker data were available for 135 participants, including 55 women with breast cancer and 80 controls. A logistic regression framework was used to assess the joint diagnostic value of five circulating biomarkers selected to reflect complementary biological pathways involved in breast cancer pathogenesis: Osteopontin, sVEGFR2, sNeuropilin-1, suPAR, and Tenascin C (figure 2). To support both prediction and inference, the analysis used a reproducible two-model strategy: a ridge-penalized logistic regression model to improve coefficient stability and reduce over fitting (table 2), and a corresponding unpenalized multivariable logistic regression model to estimate conventional confidence intervals and p values (table 3)

In the penalized model (table 2), all five biomarkers showed inverse associations with the odds of breast cancer after mutual adjustment, with effects expressed per 1-standard-deviation increase in biomarker concentration. The standardized penalized coefficients were modest in magnitude but directionally consistent across markers. Osteopontin had a coefficient of −0.086 with an odds ratio of 0.917 and a bootstrap 95% CI of 0.880 to 0.958. For sVEGFR2, the coefficient was −0.065 with an odds ratio of 0.937 and 95% CI 0.903 to 0.973. sNeuropilin-1 showed a coefficient of −0.090 with an odds ratio of 0.914 and 95% CI 0.880 to 0.949. suPAR yielded a coefficient of −0.077, corresponding to an odds ratio of 0.926 with 95% CI 0.893 to 0.967, while Tenascin C had a coefficient of −0.071, corresponding to an odds ratio of 0.931 with 95% CI 0.898 to 0.967. Among these, sNeuropilin-1 and Osteopontin displayed the largest absolute penalized effects, suggesting the strongest relative contribution to the combined prediction model.

When the same biomarker set was examined in the unpenalized multivariable model (table 2) for inferential purposes, Osteopontin was the only biomarker that remained independently associated with breast cancer status after adjustment for the other markers. Its regression coefficient was −0.704 with a 95% CI of −1.348 to −0.060 and a p value of 0.032. In contrast, the adjusted associations for the remaining biomarkers did not reach conventional statistical significance. sVEGFR2 showed a coefficient of −0.310 with a 95% CI of −0.986 to 0.366 (p = 0.368), sNeuropilin-1 a coefficient of −0.547 with a 95% CI of −1.157 to 0.063 (p = 0.079), suPAR a coefficient of −0.292 with a 95% CI of −0.880 to 0.297 (p = 0.332), and Tenascin C a coefficient of −0.300 with a 95% CI of −0.914 to 0.314 (p = 0.338). These findings indicate that, within the fully adjusted model, Osteopontin provided the most robust independent signal, whereas the remaining biomarkers appeared to contribute more modestly and primarily in combination.

Because the goal of the multivariable analysis was also to inform a clinically interpretable composite framework, the findings support the development of a weighted Clinical Decision Score derived from the joint biomarker profile. On the basis of the present dataset, the strongest justification for weighting would center on Osteopontin, with the other biomarkers serving as incremental contributors to overall model performance.

The penalized model demonstrated moderate discriminatory ability. Receiver operating characteristic (ROC) analysis showed an apparent AUC of 0.758, (table 4) with very similar internally validated estimates, including a 10-fold cross-validated AUC of 0.746 and a bootstrap out-of-bag median AUC of 0.749 (table 4). The close agreement between apparent and internally validated performance suggests limited optimism in the derivation sample. At the probability threshold that maximized the Youden index (0.481), the model achieved a sensitivity of 96.4% and a specificity of 53.8% (table 4) indicating a classification profile that favored high case detection while accepting a comparatively larger number of false-positive classifications.

Calibration was moderate and stable under internal validation. The apparent Brier score was 0.223, compared with 0.226 in 10-fold cross-validation and a bootstrap out-of-bag median of 0.225 (table 4). The minimal change across these estimates indicates that overall prediction error remained relatively stable during internal validation and provides further support that the model was not substantially overfit.

Taken together, these results show that the five-biomarker multivariable model achieved internally consistent, moderate predictive performance with discrimination in the mid-0.74 AUC range after internal validation (figure 3). Although all five biomarkers contributed to the penalized prediction model, Osteopontin emerged as the strongest independent adjusted predictor in the conventional multivariable analysis. These findings support the potential value of a multi-biomarker approach for breast cancer risk stratification, while also underscoring the need for external validation and formal score derivation before a weighted clinical decision score could be considered ready for routine clinical application.

Bioinformatics Analysis

Construction of the Breast Invasive Carcinoma Interaction Network Revealed a Highly Connected Extracellular Matrix Remodeling Program (figure 4). Protein–protein interaction (PPI) analysis of the altered genes identified from the Breast Invasive Carcinoma cohort in cBioPortal generated a densely interconnected network dominated by extracellular matrix (ECM), cell adhesion, and growth factor signaling molecules (figure 4). The network exhibited a pronounced hub-and-spoke architecture, indicating that a limited number of regulatory molecules coordinate multiple downstream effectors involved in breast cancer progression.

Among all nodes, TGFB1 occupied the most central position, displaying the highest degree of connectivity with ECM proteins, integrins, and matricellular components, suggesting that TGF-β-mediated signaling constitutes a major regulatory axis in breast invasive carcinoma. Additional highly connected nodes included FN1, ITGB1, THBS1, SPP1, VEGFA, PLAU, and multiple collagen family members (COL1A1, COL1A2, COL3A1, COL4A1/2, COL5A1, and COL6A1/2/3).

Distinct Functional Modules Underlie Breast Tumor Progression. Topological organization of the network revealed four major functional clusters:

(i) ECM and Collagen Remodeling Module

A prominent cluster comprising COL1A1, COL1A2, COL3A1, COL4A1, COL4A2, COL5A1, COL6A1, COL6A2, and COL6A3 indicated extensive matrix remodeling and stromal activation. Such alterations are characteristic of desmoplastic reactions and invasive tumor behavior.

(ii) Integrin-Mediated Adhesion Signaling Module

Strong interactions among ITGB1, ITGA1, ITGA5, ITGA6, ITGA7, and ITGA8 highlighted integrin-dependent signaling as a key mechanism facilitating tumor cell adhesion, migration, and invasion.

(iii) Basement Membrane/Laminin Module

The laminin subnetwork (LAMA1, LAMA3, LAMA4, LAMA5, LAMB1, LAMB2, LAMC1, and LAMC3) formed a distinct module associated with basement membrane organization and tumor–stromal interactions, suggesting active remodeling of the tumor microenvironment.

(iv) Angiogenic and Proteolytic Module

The presence of VEGFA, PLAU, THBS1, and SPP1 suggested coordinated regulation of angiogenesis, extracellular proteolysis, and metastatic dissemination.

Potential Novel Drug Target

Among the identified hubs, SPP1 represents the most promising candidate therapeutic target because:

1. It occupies a strategic bridging position between stromal remodeling and tumor-cell signaling.
2. It integrates ECM, integrin, angiogenic, and proteolytic pathways.
3. Unlike structural ECM proteins (e.g., collagens), SPP1 is a secreted matricellular protein and is therefore pharmacologically accessible.
4. Elevated SPP1 expression has been associated with aggressive disease, metastatic dissemination, and poor prognosis across multiple breast cancer subtypes.
5. Its central location in this network suggests that targeting SPP1 could simultaneously disrupt multiple oncogenic processes.

Thus, our network analysis identifies SPP1 as a putative novel therapeutic vulnerability in breast invasive carcinoma, particularly within tumors characterized by extensive stromal remodeling and TGF-β activation.

Study Novelty

Our drug target Bioinformatics network analysis of cBioPortal-derived breast invasive carcinoma identified a putative novel therapeutic axis in Breast Invasive Carcinoma. While the central involvement of TGFB1, FN1, and ITGB1 in breast cancer has been extensively documented, the network revealed an unexpectedly strong interconnectivity among THBS1, SPP1, and PLAU, positioned immediately downstream of TGFB1 and closely linked to integrin and ECM signaling, further research is needed to explore the roles of this signal axis in breast cancer.

Notably, SPP1 (osteopontin) emerged as a critical intermediary, connecting TGFB1-driven signaling, integrin-mediated adhesion pathways, extracellular matrix remodeling, and proteolytic invasion machinery.

The simultaneous interaction of SPP1 with TGFB1, FN1, ITG receptors, THBS1, and PLAU suggests the existence of a previously studied TGFB1–SPP1–PLAU signaling axis that may orchestrate invasive and metastatic phenotypes in breast invasive carcinoma. This also uncovered a previously underrecognized THBS1-SPP1-PLAUR axis, positioning SPP1 as a central integrator of extracellular matrix remodeling, adhesion signaling, and invasive programs, thereby identifying SPP1 as a promising therapeutic target in breast invasive carcinoma.

CONCLUSION

A key innovation of this study is the identification of the novel THBS1, SPP1, and PLAUR signaling axis in breast cancer patients and the overall integration of penalized and unpenalized multivariable modeling with network-based Bioinformatics-systems biology analysis to translate statistical associations into a clinically interpretable scoring system. This approach enables the transformation of continuous biomarker signals into a tiered clinical decision tool for screening and referral guidance. Unlike conventional equal-weight scoring systems, the proposed model assigns greater weighting to Osteopontin because of its central involvement in multiple oncogenic pathways spanning angiogenesis, inflammation, extracellular matrix remodeling, and metastasis. Our result shows that Osteopontin showed the strongest association with disease presence, while combined elevation of multiple biomarkers increased diagnostic relevance.

This study developed and internally validated a weighted Clinical Decision Score (wCDS) based on five circulating biomarkers and systems biology analyses to improve breast cancer risk stratification. The multivariable biomarker model demonstrated moderate and internally consistent diagnostic performance with limited evidence of overfitting following internal validation. Among the evaluated biomarkers, Osteopontin (SPP1) emerged as the strongest independent predictor of breast cancer, while sVEGFR2, sNeuropilin-1, suPAR, and Tenascin C provided complementary predictive information within the penalized multivariable model, supporting the value of a multi-biomarker approach over reliance on a single biomarker.

Protein–protein interaction network analysis provided biological validation of the proposed scoring framework by identifying TGFB1 as the principal regulatory hub governing extracellular matrix remodeling and SPP1 as a highly connected molecular bridge linking TGFB1 signaling with integrin activation, angiogenesis, extracellular matrix remodeling, and proteolytic invasion through interactions with FN1, PLAU, THBS1, VEGFA, and collagen family proteins. These findings highlight a biologically plausible TGFB1–SPP1–PLAU regulatory axis that provides mechanistic support for the weighting strategy used in the wCDS and identifies SPP1 as a promising candidate for future translational and therapeutic investigation in breast invasive carcinoma.

Collectively, these findings demonstrate that integrating circulating biomarkers with systems biology can enhance breast cancer risk stratification while simultaneously providing mechanistic insight into the molecular pathways underlying disease progression. The proposed wCDS is intended to function as a clinical decision-support tool to assist in identifying women who may benefit from earlier diagnostic imaging, particularly those with dense breast tissue or younger women in whom conventional screening may be less sensitive. Importantly, the wCDS is designed to complement, rather than replace, established imaging modalities and clinical assessment. Further external validation in larger, multicenter, and ethnically diverse cohorts is required before clinical implementation, together with prospective evaluation of its diagnostic utility, clinical impact, and cost-effectiveness.

Funding

No funding was received.

Conflicts of Interest

The authors declare no conflict of interest

REFERENCES

1. Gu H, Wang R, Beeraka NM, Lu P, Zhao X, Li L, et al. Global burden and trends of breast cancer: GLOBOCAN 2022 estimates of incidence and mortality in 185 countries. Chin Med J (Engl). 2026;139(3):404-414. doi:10.1097/CM9.0000000000003921.

2. Global burden of female breast cancer: new estimates in 2022, temporal trend and future projections up to 2050 based on the latest release from GLOBOCAN. PubMed. 2025. Available from: https://pubmed.ncbi.nlm.nih.gov/40693239/

3. Global burden and projections of breast cancer incidence and mortality to 2050: a comprehensive analysis of GLOBOCAN data. PubMed. 2025. Available from: https://pubmed.ncbi.nlm.nih.gov/41246094/

4. Jemal A, et al. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12-49. doi:10.3322/caac.21820.

5. Colomer R, González-Farré B, Ballesteros AI, Peg V, Bermejo B, Pérez-Mies B, et al. Biomarkers in breast cancer 2024: an updated consensus statement by the Spanish Society of Medical Oncology and the Spanish Society of Pathology. Clin Transl Oncol. 2024;26(12):2935-2951. doi:10.1007/s12094-024-03541-1.
6. Dou T, Li J, Zhang Y, Pei W, Zhang B, Wang B, et al. The cellular composition of the tumor microenvironment is an important marker for predicting therapeutic efficacy in breast cancer. Front Immunol. 2024;15:1368687. doi:10.3389/fimmu.2024.1368687.
7. Leone P, Malerba E, Susca N, Favoino E, Perosa F, Brunori G, et al. Endothelial cells in tumor microenvironment: insights and perspectives. Front Immunol. 2024;15:1367875. doi:10.3389/fimmu.2024.1367875.
8. Trnkova L, Buocikova V, Mego M, Cumova A, Burikova M, Bohac M, et al. Epigenetic deregulation in breast cancer microenvironment: implications for tumor progression and therapeutic strategies. Biomed Pharmacother. 2024;174:116559. doi:10.1016/j.biopha.2024.116559.
9. Guarneri V, et al. Major advancements in metastatic breast cancer treatment: when expanding options means prolonging survival. ESMO Open. 2022;7(2):100409. doi:10.1016/j.esmoop.2022.100409.
10. Vellan CJ, Islam T, De Silva S, Taib NAM, Prasanna G, Jayapalan JJ. Exploring novel protein-based biomarkers for advancing breast cancer diagnosis: a review. Clin Biochem. 2024;129:110776. doi:10.1016/j.clinbiochem.2024.110776.
11. Pastena P, Perera H, Martinino A, Kartsonis W, Giovinazzo F, et al. Unraveling biomarker signatures in triple-negative breast cancer: a systematic review for targeted approaches. Int J Mol Sci. 2024;25(5):2559. doi:10.3390/ijms25052559.
12. Løyland B, Sandbekken IH, Grov EK, Utne I. Causes and risk factors of breast cancer, what do we know for sure? An evidence synthesis of systematic reviews and meta-analyses. Cancers (Basel). 2024;16(8):1583. doi:10.3390/cancers16081583.
13. Risk factors for breast cancer: an umbrella review of observational cohort studies and causal relationship analysis. PubMed. 2025. Available from: https://pubmed.ncbi.nlm.nih.gov/40386557/
14. Complex interplay between obesity and BRCA1/2-associated breast cancer: an overview. PubMed. 2025. Available from: https://pubmed.ncbi.nlm.nih.gov/40523654/
15. Molecular pathology of breast tumors: diagnostic and actionable genetic alterations. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38821644/
16. Colomer R, et al. Biomarkers in breast cancer 2024. Clin Transl Oncol. 2024;26(12):2935-2951. doi:10.1007/s12094-024-03541-1.
17. Breast cancer immune landscape: interplay between systemic and local immunity. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38727796/
18. Guardians and mediators of metastasis: exploring T lymphocytes, myeloid-derived suppressor cells, and tumor-associated macrophages in the breast cancer microenvironment. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38892411/
19. Tumor associated macrophages in breast cancer progression: implications and clinical relevance. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/39044824/
20. Characterization of the breast cancer liver metastasis microenvironment via machine learning analysis of the primary tumor microenvironment. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/39373616/
21. Elayat G, Selim A. Angiogenesis in breast cancer: insights and innovations. BMC Cancer. 2024;24(1):178. doi:10.1186/s12885-024-12617-7.
22. Biswas M. Understanding tissue-resident macrophages unlocks the potential for novel combinatorial strategies in breast cancer. Front Immunol. 2024;15:1375528. doi:10.3389/fimmu.2024.1375528.
23. Elfstrum AK, Bapat AS, Schwertfeger KL. Defining and targeting macrophage heterogeneity in the mammary gland and breast cancer. Cancer Med. 2024;13(3):e7053. doi:10.1002/cam4.7053.
24. Patterns of immune evasion in triple-negative breast cancer and new potential therapeutic targets: a review. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/39735530/
25. Circulating tumor DNA and survival in metastatic breast cancer: a systematic review and meta-analysis. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/39235812/
26. Prognostic value of circulating tumor DNA mutation panels to predict early recurrence and survival outcomes in early-stage breast cancer: a systematic review and meta-analysis. Expert Rev Mol Diagn. 2026. doi:10.1080/1354750X.2026.2633567.
27. Monaci S, Coppola F, Filippi I, Falsini A, Carraro F, Naldini A. Targeting hypoxia signaling pathways in angiogenesis. Front Physiol. 2024;15:1408750. doi:10.3389/fphys.2024.1408750.
28. Sarabipour S, Kinghorn K, Quigley KM, Kovacs-Kasa A, Annex BH, Bautch VL, et al. Trafficking dynamics of VEGFR1, VEGFR2, and NRP1 in human endothelial cells. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38324585/
29. Recent updates on potential of VEGFR-2 small-molecule inhibitors as anticancer agents. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/39439843/
30. Mechanisms of angiogenesis in tumour. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38590656/
31. Kola B, et al. Lymphangiogenesis in breast cancer: from molecular mechanisms to clinical implications. FASEB J. 2025;39(9):e70590. doi:10.1096/fj.202500024R.

32. Inflammatory blood markers in breast cancer: a narrative review from early detection to therapy response. PubMed. 2025. Available from: https://pubmed.ncbi.nlm.nih.gov/40901118/
33. Prognostic and clinicopathological value of systemic inflammation response index (SIRI) in patients with breast cancer: a meta-analysis. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38569199/
34. Neutrophil extracellular traps in breast cancer: roles in metastasis and beyond. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38817858/
35. PlexinB1 inactivation reprograms immune cells in the tumor microenvironment, inhibiting breast cancer growth and metastatic dissemination. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38874583/
36. Cha SM, Park JW, Lee YJ, Lee HJ, Lee H, Lee IW, et al. SPP1+ macrophages in HR+ breast cancer are associated with tumor-infiltrating lymphocytes. NPJ Breast Cancer. 2024;10(1):83. doi:10.1038/s41523-024-00695-7.
37. SPP1+ macrophages in tumor immunosuppression: mechanisms and therapeutic implications. PubMed. 2025. Available from: https://pubmed.ncbi.nlm.nih.gov/41425545/
38. Palma A. The landscape of SPP1+ macrophages across tissues and diseases: a comprehensive review. Immunology. 2025;176(2):179-196. doi:10.1111/imm.13952.
39. SPP1+ macrophages and the orchestration of spatially organized immunosuppression in cancer. PubMed. 2026. Available from: https://pubmed.ncbi.nlm.nih.gov/41751193/
40. Tenascin-C targeting strategies in cancer. Matrix Biol. 2024. doi:10.1016/j.matbio.2024.04.002.
41. The role of tenascin-C in tumor microenvironments and its potential as a therapeutic target. Front Cell Dev Biol. 2025. doi:10.3389/fcell.2025.1554312.
42. Tenascin-C targeting strategies in cancer. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/38642843/
43. Guo M, Liu M, Li W, Wang C, Zhang L, Zhang H. Osteopontin promotes tumor growth and metastasis and GPX4-mediated anti-lipid peroxidation in triple-negative breast cancer by activating the PI3K/Akt/mTOR pathway. J Cancer Res Clin Oncol. 2024;150(3):155. doi:10.1007/s00432-024-05658-w.
44. Role of osteopontin in cancer: from pathogenesis to therapeutics. Mol Med Rep. 2025;54(5):136. doi:10.3892/mmr.2025.xxxxx.
45. The role of SPP1+TAMs in cancer: impact on patient prognosis and future therapeutic targets. PubMed. 2025. Available from: https://pubmed.ncbi.nlm.nih.gov/40568964/
46. Identification of a novel subtype of SPP1+ macrophages expressing SIRPα: implications for tumor immune evasion and treatment response prediction. PubMed. 2024. Available from: https://pubmed.ncbi.nlm.nih.gov/39696410/
47. Targeting triple-negative breast cancer: resistance mechanisms and therapeutic advancements. PubMed. 2025. Available from: https://pubmed.ncbi.nlm.nih.gov/40318146/
48. Triple-negative breast cancer brain metastasis: a comprehensive review of epidemiology, molecular pathobiology, and therapeutic frontiers. PubMed. 2026. Available from: https://pubmed.ncbi.nlm.nih.gov/41976401/
49. Raman R, Debata S, Govindarajan T, Kumar P. Targeting triple-negative breast cancer: resistance mechanisms and therapeutic advancements. Biochim Biophys Acta Rev Cancer. 2025;14(9):e70803
50. Abah, M. O., Ogenyi, D. O., Zhilenkova, A. V., Essogmo, F. E., Uchendu, I. K., Tchawe, Y. S. N., Pascal, A. M., Nikitina, N. M., Oloche, O. S., Pavliv, M., Rusanov, A. S., Sanikovich, V. D., Pirogova, Y. N., Bagmet, L. N., Moiseeva, A. V., & Sekacheva, M. I. (2025). Comparative Transcriptomics Study of Curcumin and Conventional Therapies in Translocation, Clear Cell, and Papillary Renal Cell Carcinoma Subtypes. International Journal of Molecular Sciences, 26(13), 6161. https://doi.org/10.3390/ijms26136161
51. Abah, M.O.; Ogenyi, D.O.; Zhilenkova, A.V.; Essogmo, F.E.; Ngaha Tchawe, Y.S.; Uchendu, I.K.; Pascal, A.M.; Nikitina, N.M.; Rusanov, A.S.; Sanikovich, V.D.; et al. Innovative Therapies Targeting Drug-Resistant Biomarkers in Metastatic Clear Cell Renal Cell Carcinoma (ccRCC). Int. J. Mol. Sci. 2025, 26, 265. https://doi.org/10.3390/ijms26010265