JHWCR JOURNAL OF HEALTH, WELLNESS AND COMMUNITY RESEARCH 3007-0570

Original Article Volume 3, Issue 1

Role of Molecular Markers in Early Detection of Breast Cancer

Tooba Khanum[1], Hira Ahmed[2], Sayra Tariq[3], Hizba Zulfiqar Ali[4], Mohammad Asad Shaheen Baloch[5], Muhammad Shahmeer Nawaz[6], Noor Ul Ain[7]

[1] M.Phil. Clinical Nutrition, Gold Medallist, Lecturer, Minhaj University Lahore, Lahore, Pakistan
[2] Associate Professor, Anatomy Department, Karachi Medical & Dental College, Karachi Metropolitan University, Karachi, Pakistan
[3] PhD Biotechnology, School of Biochemistry and Biotechnology, University of the Punjab, Lahore, Pakistan
[4] Liaquat National Medical College, Karachi, Pakistan
[5] MBBS, MCPS, Physician, Sheikh Sultan bin Zayed Hospital, Sharjah, United Arab Emirates
[6] MBBS, University College of Medicine and Dentistry, Lahore, Pakistan
[7] PhD, Postdoc, Translational Brain Research, State Key Laboratory of Medical Neurobiology, MOE Frontiers Centre for Brain Science, Fudan University, Shanghai, China
* Correspondence: Noor Ul Ain, noorulain22@yahoo.com

ABSTRACT

Background: *Breast cancer remains one of the leading causes of cancer-related morbidity and mortality among women worldwide, and survival outcomes are strongly associated with early detection. Although mammography has substantially improved screening outcomes, its diagnostic sensitivity may be limited in certain populations, particularly among women with dense breast tissue, highlighting the need for complementary diagnostic approaches. Blood-based molecular biomarkers have emerged as promising non-invasive tools capable of detecting tumor-associated biological signals at early stages of disease development.* ***Objective:*** *To systematically evaluate and synthesize available evidence on the diagnostic accuracy of blood-based molecular biomarkers for the early detection of breast cancer.* ***Methods:*** *A systematic review and diagnostic test accuracy meta-analysis were conducted following PRISMA guidelines. Electronic databases including PubMed, Scopus, Web of Science, and Embase were searched for studies published between 2000 and 2025. Eligible studies evaluated circulating molecular biomarkers in blood samples from women with early-stage breast cancer and reported sufficient data to calculate diagnostic accuracy. Study quality was assessed using the QUADAS-2 tool. Pooled sensitivity, specificity, and summary receiver operating characteristic (SROC) curves were calculated using a bivariate random-effects model.* ***Results:*** *Sixty-two studies involving 8,465 breast cancer cases and 8,045 controls were included. cfDNA fragment omics demonstrated the highest diagnostic accuracy (sensitivity 0.88, specificity 0.90, AUC 0.93), followed by cfDNA methylation markers (sensitivity 0.84, specificity 0.87, AUC 0.91). Circulating microRNA panels showed moderate-to-high performance (sensitivity 0.81, specificity 0.83, AUC 0.88), whereas traditional protein markers showed lower accuracy (sensitivity 0.62, specificity 0.68, AUC 0.70). Multi-marker panels consistently outperformed single biomarkers.* ***Conclusion:*** *Blood-based molecular biomarkers, particularly cfDNA fragment omics and methylation profiling, demonstrate strong potential as complementary tools for early breast cancer detection. However, further prospective screening studies and methodological standardization are required before routine clinical implementation.*

Keywords: *Breast cancer; early detection; molecular biomarkers; liquid biopsy; circulating cell-free DNA; DNA methylation; fragmentomics; circulating microRNA; diagnostic accuracy; meta-analysis; screening.*

Received: 09 December 2025
Revised: 02 January 2025
Accepted: 20 February 2025
Published: 28 February 2025

Citation: Click to Cite

INTRODUCTION

Breast cancer remains a leading cause of cancer morbidity and mortality among women globally, with outcomes strongly dependent on stage at diagnosis and timely initiation of definitive therapy (1). Population-based screening programs have reduced advanced-stage presentation in many settings, yet the clinical reality is that a substantial fraction of cancers is still detected beyond the earliest stages, and screening performance is not uniform across risk groups and breast tissue phenotypes (1). A central challenge is that imaging-based

screening, while indispensable, is constrained by biologic and technical factors that can reduce sensitivity or generate downstream harms, particularly among women with higher mammographic density where masking and detection bias can complicate lesion visualization and shift the balance between benefit and false-positive workup (2). Even newer imaging modalities and density estimation approaches continue to face variability in density assessment and interpretability across platforms, reinforcing the need to develop complementary, scalable tools that can strengthen early detection pathways rather than attempting to replace imaging wholesale (3).

In parallel with advances in imaging, blood-based molecular biomarkers have emerged as plausible adjuncts for early detection because they aim to capture tumor-associated biological signals that may be present even when radiologic findings are subtle or equivocal. Conceptually, this approach aligns with the clinical need for minimally invasive tests that can be repeatedly deployed and potentially integrated into risk-stratified screening or diagnostic triage workflows. Among candidate biomarker classes, circulating cell-free DNA (cfDNA) has gained prominence because plasma DNA carries information not only about sequence alterations but also about tissue-of-origin features and epigenetic signatures, and methylation-based profiling has demonstrated clinically meaningful signal for cancer detection and localization in large multi-cancer settings (4). Foundational work describing cfDNA as an in vivo nucleosome footprint supports the mechanistic premise that fragmentation patterns reflect underlying chromatin architecture and cell-of-origin biology, which can be exploited analytically to distinguish tumor-derived cfDNA from background release by normal tissues (5). Building on these principles, fragmentomics and related genome-wide cfDNA features have been investigated for tumor subtype prediction and early detection, often leveraging computational models to integrate fragment length distributions and genome-wide coverage patterns into classification frameworks (6).

Despite increasing publication volume, the evidence base remains difficult to translate into practice because primary studies vary widely in patient spectrum, inclusion of asymptomatic versus clinically detected cases, specimen type (serum vs plasma), analytical platforms, normalization pipelines, and—critically—threshold selection and model training strategies. These sources of heterogeneity can inflate apparent diagnostic performance, especially in retrospective case-control designs that separate clear cases from healthy controls in ways that do not reflect real-world screening distributions. This is particularly relevant in early-stage disease, where tumor-derived signals are low-abundance and susceptible to confounding by non-malignant biological variation. Prior quantitative syntheses have suggested that cfDNA concentration-based approaches have diagnostic value in breast cancer, but reported performance varies and may depend strongly on stage mix, assay standardization, and control selection (7). Likewise, exosome-associated and circulating microRNA panels have shown promise as non-invasive diagnostics, yet the literature demonstrates substantial methodological diversity and inconsistent replication across cohorts, implying that pooled estimates must be interpreted through a rigorous diagnostic test accuracy framework and stratified by biomarker class and study design features (8). At the same time, critical appraisals of liquid biopsy development continue to debate whether and when these technologies could complement or potentially compete with established screening modalities, underscoring that the key question is not feasibility alone but clinically deployable accuracy under realistic use-cases (9). Rapid innovation in liquid biopsy technologies—spanning sequencing, methylation profiling, and integrative multi-omics pipelines—further increases the risk that narrative conclusions will outpace reproducible evidence unless updated, methodologically stringent synthesis is performed (10).

A second, biologically grounded reason why a synthesis focused on early-stage detection is essential is that breast cancer is molecularly heterogeneous. Intrinsic subtypes and genomic diversity influence tumor shedding, epigenetic remodeling, and circulating RNA profiles, meaning that single-marker strategies may be inherently brittle across populations and subtypes (11). Large-scale molecular profiling has reinforced that breast tumors differ substantially in genomic and transcriptomic landscapes, strengthening the rationale for multi-marker and multi-feature panels (including methylation and fragmentomics signatures) that can capture diverse tumor biology while maintaining specificity against benign conditions and healthy backgrounds (12). However, the field lacks consistent, clinically oriented integration of these biomarker classes under a unified early-stage framework that explicitly targets the intended population (women undergoing screening or evaluation for stage 0–II disease), compares index tests against appropriate non-cancer control groups and/or standard pathways, and summarizes diagnostic accuracy using models that jointly estimate sensitivity and specificity while accounting for between-study heterogeneity.

Accordingly, the objective of this study is to systematically evaluate and meta-analyze the diagnostic accuracy of blood-based molecular biomarkers for early-stage breast cancer detection (stage 0–II) in women, comparing performance across major biomarker classes (cfDNA methylation, cfDNA fragmentomics, circulating microRNA panels, and conventional protein markers when reported) against relevant control groups and/or reference standards, and quantifying pooled sensitivity and specificity within a diagnostic test accuracy framework. The prespecified research question is: among women undergoing screening or evaluation for possible breast cancer, do blood-based molecular biomarkers—particularly cfDNA methylation and fragmentomics signatures—demonstrate sufficient diagnostic accuracy to complement current detection pathways for identifying early-stage breast cancer compared with non-cancer controls and standard diagnostic reference approaches (1–12)?

METHODS

A systematic review and diagnostic test accuracy meta-analysis was conducted to evaluate the performance of blood-based molecular biomarkers for the early detection of breast cancer. The methodological framework was developed in accordance with internationally recognized reporting standards for systematic reviews and diagnostic accuracy studies, including the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) and methodological guidance for diagnostic test accuracy synthesis (13,14). The study followed a predefined protocol specifying the research question, eligibility criteria, data extraction strategy, and statistical analysis approach to minimize selective reporting and analytic bias. The review addressed the Population–Intervention–Comparator–Outcome (PICO) framework, focusing on women evaluated for early-stage breast cancer in whom molecular biomarkers were measured in blood-based samples, compared against non-cancer controls or standard diagnostic reference methods, with diagnostic accuracy outcomes including sensitivity, specificity, and area under the receiver operating characteristic curve.

A comprehensive literature search was conducted across four major electronic databases—PubMed/MEDLINE, Scopus, Web of Science, and Embase—to identify relevant studies published between January 2000 and December 2025. These databases were selected because they collectively provide extensive coverage of biomedical, clinical, and molecular oncology research. The search strategy combined controlled vocabulary terms and free-text keywords related to breast cancer, molecular biomarkers, liquid biopsy, and diagnostic accuracy. Search terms included combinations of “breast cancer,” “early detection,” “screening,” “molecular

biomarker," "liquid biopsy," "circulating tumor DNA," "cell-free DNA," "DNA methylation," "fragmentomics," "microRNA," "exosomal RNA," "protein biomarkers," "diagnostic accuracy," "sensitivity," and "specificity." Boolean operators ("AND" "OR") were used to structure database-specific search strings, and Medical Subject Headings (MeSH) were applied where available to enhance sensitivity of retrieval. Reference lists of relevant systematic reviews and primary studies were manually screened to identify additional eligible publications that may not have been captured during electronic searching.

Eligible studies were selected according to predefined inclusion and exclusion criteria designed to ensure that the analysis addressed the intended clinical context of early breast cancer detection. Studies were included if they investigated molecular biomarkers measured in blood-derived samples such as plasma, serum, or whole blood; included women with histologically confirmed early-stage breast cancer (stage 0–II); incorporated a comparator group consisting of healthy individuals or patients with benign breast conditions; and reported sufficient diagnostic accuracy data to construct two-by-two contingency tables (true positives, false positives, true negatives, and false negatives) or provided directly reported sensitivity and specificity values. Original research studies involving human participants and published in peer-reviewed journals were eligible for inclusion. Studies were excluded if they focused exclusively on metastatic or recurrent breast cancer, evaluated biomarkers solely for prognosis or treatment monitoring, lacked an appropriate control group, did not report sufficient diagnostic data, or consisted of case reports, conference abstracts without full datasets, reviews, or animal-only studies. When multiple publications reported results from the same dataset, the most comprehensive and recent study was retained to avoid duplication.

All records retrieved from the electronic searches were imported into a reference management system for organization and removal of duplicate entries. Study selection proceeded in two stages. First, titles and abstracts were screened to identify potentially relevant studies based on the predefined eligibility criteria. Second, full texts of all potentially eligible articles were independently reviewed to confirm eligibility. Disagreements between reviewers during either stage were resolved through discussion and consensus. A structured study selection process was documented and summarized in a PRISMA flow diagram to ensure transparency and reproducibility of the selection process (13).

Data extraction was conducted using a standardized extraction form developed before the review process began. Extracted variables included study characteristics (first author, year of publication, country), study design (prospective cohort, retrospective cohort, or case-control), participant characteristics, sample size of cases and controls, cancer stage distribution, specimen type (plasma, serum, or whole blood), biomarker category, analytical platform used for detection, threshold definitions when available, and reported diagnostic accuracy metrics. Where studies reported multiple biomarkers or biomarker panels, the index test that represented the primary diagnostic evaluation described by the authors was extracted. When necessary, sensitivity and specificity values were calculated from reported contingency data. Data extraction was performed independently to reduce transcription errors and ensure consistency of the extracted dataset.

The methodological quality and risk of bias of included studies were evaluated using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool, which assesses risk of bias across four domains: patient selection, index test, reference standard, and flow and timing (15). Each domain was judged as having low, high, or unclear risk of bias based on prespecified criteria. Applicability concerns for the first three domains were also assessed. Quality assessments were conducted systematically and summarized across studies to

identify potential sources of bias that might influence pooled estimates of diagnostic accuracy.

The primary outcome measures were pooled sensitivity and pooled specificity of molecular biomarkers for detecting early-stage breast cancer. Diagnostic accuracy estimates were synthesized using a bivariate random-effects meta-analysis model, which jointly models sensitivity and specificity while accounting for potential correlation between these parameters across studies and allowing for between-study heterogeneity (16). Summary receiver operating characteristic (SROC) curves were generated to visualize overall diagnostic performance across biomarker categories. The area under the SROC curve (AUC) was calculated to provide an integrated measure of test discrimination. Heterogeneity between studies was assessed through visual inspection of forest plots and SROC distributions, and potential sources of heterogeneity were explored through subgroup analyses stratified by biomarker class, study design, and sample type. Publication bias was evaluated using Deeks' funnel plot asymmetry test, which is recommended for diagnostic test accuracy meta-analyses (17).

To evaluate the robustness of pooled estimates, sensitivity analyses were performed by sequentially removing individual studies to assess the stability of the overall results. Additional analyses examined the influence of study design characteristics, including case-control versus prospective cohort designs, on diagnostic performance estimates. When necessary, statistical corrections were applied to account for zero-cell frequencies in contingency tables to enable inclusion of all eligible studies in pooled analyses. All statistical analyses were conducted using specialized diagnostic meta-analysis software implemented in STATA, which provides established procedures for bivariate modeling and SROC estimation.

Ethical approval was not required because the study synthesized data from previously published research and did not involve the collection of new patient-level data. All included studies were reviewed to ensure that they had obtained appropriate institutional ethical approval and informed consent from participants where applicable. Data integrity was maintained through standardized extraction procedures, verification of extracted values against original publications, and structured documentation of analytical steps. The analytical dataset and extraction framework were maintained in organized formats to facilitate transparency and reproducibility of the meta-analysis workflow.

RESULTS

Table 1 summarizes the structural profile of the 62 included studies and shows that the evidence base was dominated by retrospective case–control designs (44/62, 71.0%), with fewer prospective cohort investigations (18/62, 29.0%). Across all studies, the pooled dataset comprised 8,465 breast cancer cases and 8,045 controls, indicating broadly balanced group sizes overall (case:control ratio ≈1.05). By biospecimen, plasma was used more frequently than serum (38/62, 61.3% vs 24/62, 38.7%), corresponding to 5,206 cases and 4,988 controls in plasma studies compared with 3,259 cases and 3,057 controls in serum studies. In terms of laboratory platforms, qPCR-based assays accounted for 26/62 studies (41.9%) with 3,471 cases and 3,215 controls, next-generation sequencing approaches were used in 21/62 studies (33.9%) with 2,975 cases and 2,806 controls, and ELISA/protein assays comprised 15/62 studies (24.2%) with 2,019 cases and 2,024 controls. Collectively, these distributions indicate that most evidence derives from nucleic-acid-centered platforms (qPCR+NGS: 47/62, 75.8%), consistent with the study's primary focus on cfDNA and miRNA-based detection strategies.

Table 2 disaggregates the 62 studies by biomarker class and demonstrates that circulating microRNA (miRNA) panels were the most frequently studied category (21/62, 33.9%), contributing 2,844 cases and 2,702 controls. cfDNA methylation markers represented the second largest group (19/62, 30.6%) with 2,594 cases and 2,420 controls. cfDNA fragmentomics and protein tumor markers were each represented by 11 studies (11/62, 17.7% each), with fragmentomics including 1,576 cases and 1,445 controls and protein markers including 1,451 cases and 1,478 controls. Methodologically, Table 2 also highlights expected platform alignment: miRNA studies were primarily qPCR-based, while cfDNA methylation and fragmentomics studies were predominantly sequencing-driven, reflecting differing analytical requirements across biomarker classes and helping contextualize downstream between-study heterogeneity.

Table 3 presents the pooled diagnostic accuracy estimates and shows clear performance stratification by biomarker type. cfDNA fragmentomics demonstrated the strongest pooled discrimination, with sensitivity 0.88 (95% CI 0.85–0.91), specificity 0.90 (95% CI 0.87–0.93), and an AUC of 0.93, accompanied by a high diagnostic odds ratio (DOR) of 62.4 (95% CI 41.2–94.5) and strong overall statistical evidence for discriminative performance ($p<0.001$). cfDNA methylation markers also performed robustly, achieving sensitivity 0.84 (0.81–0.87), specificity 0.87 (0.84–0.90), and AUC 0.91, with a DOR of 34.6 (25.5–46.9) ($p<0.001$). Circulating miRNA panels showed moderate-to-high pooled accuracy, with sensitivity 0.81 (0.77–0.85), specificity 0.83 (0.79–0.86), AUC 0.88, and a DOR of 20.7 (15.3–27.9) ($p<0.001$). In contrast, protein tumor markers demonstrated materially weaker performance, with sensitivity 0.62 (0.58–0.66), specificity 0.68 (0.64–0.72), AUC 0.70, and a DOR of 3.5 (2.6–4.7), indicating limited utility for early-stage discrimination relative to nucleic acid-based approaches ($p=0.012$). The magnitude of these differences is clinically meaningful: for example, the sensitivity gap between fragmentomics and protein markers was 0.26 (0.88 vs 0.62), while the specificity gap was 0.22 (0.90 vs 0.68), favoring DNA-based technologies on both false-negative and false-positive control.

Table 4 compares multi-marker panels with single-biomarker approaches and shows that panel-based strategies consistently outperformed single-marker tests. Multi-marker panels (29 studies) achieved pooled sensitivity 0.86 (0.83–0.89) and specificity 0.88 (0.85–0.91), yielding AUC 0.92 and an odds ratio of 48.3 (95% CI 33.7–69.4), with strong statistical support ($p<0.001$). Single biomarkers (33 studies) showed lower pooled sensitivity 0.74 (0.70–0.78) and specificity 0.77 (0.73–0.81), with AUC 0.84 and odds ratio 12.1 (9.2–15.9) ($p<0.001$). Numerically, panels improved sensitivity by 0.12 (0.86–0.74) and specificity by 0.11 (0.88–0.77), implying that combining signals can reduce missed early cancers and decrease false positives simultaneously, consistent with the biological heterogeneity of breast cancer and the statistical advantages of composite classification.Table 5 evaluates sample type and indicates modest performance advantages for plasma-based assays. Plasma studies (38 studies) reported pooled sensitivity 0.85 (0.82–0.88), specificity 0.88 (0.85–0.90), and AUC 0.91, with a statistically significant difference signal ($p=0.041$). Serum studies (24 studies) showed pooled sensitivity 0.80 (0.76–0.84), specificity 0.83 (0.79–0.87), and AUC 0.88, with weaker statistical evidence for difference ($p=0.063$). The absolute differences favored plasma by 0.05 in sensitivity (0.85 vs 0.80), 0.05 in specificity (0.88 vs 0.83), and 0.03 in AUC (0.91 vs 0.88), suggesting that pre-analytical factors and matrix effects may influence detectable signal, particularly for cfDNA-derived features where leukocyte lysis and background DNA contamination can be more problematic in serum.

Table 6 stratifies diagnostic performance by study design and demonstrates patterns consistent with spectrum bias. Prospective cohort studies (18 studies) exhibited pooled sensitivity 0.80 (0.76–0.84) and specificity 0.89 (0.86–0.92), with AUC 0.90 and a significant

design-associated difference signal (p=0.028). Retrospective case–control studies (44 studies) reported higher pooled sensitivity 0.86 (0.83–0.89) but lower pooled specificity 0.84 (0.81–0.87), with AUC 0.89 (p=0.034).

The sensitivity difference of +0.06 in favor of case–control designs (0.86 vs 0.80) aligns with the expectation that more clearly separated case/control spectra inflate detection probability, whereas the specificity advantage in prospective cohorts (+0.05; 0.89 vs 0.84) is consistent with more representative control populations and more clinically realistic thresholds. Taken together, these design-stratified results indicate that while pooled accuracy is strongest for cfDNA fragmentomics and methylation overall, real-world screening performance is likely to be closer to the prospective estimates than to the retrospective case–control estimates, reinforcing the need for future large-scale prospective screening validation.

Table 1. Characteristics of Included Studies

Variable	Category	Number of Studies (n=62)	Breast Cancer Cases (n=8,465)	Controls (n=8,045)
Study Design	Case-control	44	5,978	5,621
	Prospective cohort	18	2,487	2,424
Sample Type	Plasma	38	5,206	4,988
	Serum	24	3,259	3,057
Detection Method	qPCR	26	3,471	3,215
	Next-generation sequencing	21	2,975	2,806
	ELISA / protein assays	15	2,019	2,024

Table 2. Distribution of Biomarker Categories Across Included Studies

Biomarker Category	Number of Studies	Cases	Controls	Most Common Detection Method
cfDNA methylation	19	2,594	2,420	NGS / methylation sequencing
cfDNA fragmentomics	11	1,576	1,445	Genome-wide sequencing
Circulating microRNA panels	21	2,844	2,702	qPCR
Protein tumor markers	11	1,451	1,478	ELISA

Table 3. Pooled Diagnostic Accuracy by Biomarker Type

Biomarker Type	Studies (n)	Pooled Sensitivity (95% CI)	Pooled Specificity (95% CI)	Diagnostic Odds Ratio (95% CI)	AUC	p-value
cfDNA fragmentomics	11	0.88 (0.85–0.91)	0.90 (0.87–0.93)	62.4 (41.2–94.5)	0.93	<0.001
cfDNA methylation	19	0.84 (0.81–0.87)	0.87 (0.84–0.90)	34.6 (25.5–46.9)	0.91	<0.001
Circulating microRNA panels	21	0.81 (0.77–0.85)	0.83 (0.79–0.86)	20.7 (15.3–27.9)	0.88	<0.001
Protein tumor markers	11	0.62 (0.58–0.66)	0.68 (0.64–0.72)	3.5 (2.6–4.7)	0.70	0.012

Table 4. Subgroup Analysis: Multi-marker Panels vs Single Biomarkers

Biomarker Approach	Studies	Sensitivity (95% CI)	Specificity (95% CI)	AUC	Odds Ratio (95% CI)	p-value
Multi-marker panels	29	0.86 (0.83–0.89)	0.88 (0.85–0.91)	0.92	48.3 (33.7–69.4)	<0.001
Single biomarker	33	0.74 (0.70–0.78)	0.77 (0.73–0.81)	0.84	12.1 (9.2–15.9)	<0.001

Table 5. Subgroup Analysis by Sample Type

Sample Type	Studies	Sensitivity (95% CI)	Specificity (95% CI)	AUC	p-value
Plasma	38	0.85 (0.82–0.88)	0.88 (0.85–0.90)	0.91	0.041
Serum	24	0.80 (0.76–0.84)	0.83 (0.79–0.87)	0.88	0.063

Table 6. Diagnostic Accuracy by Study Design

Study Design	Studies	Sensitivity (95% CI)	Specificity (95% CI)	AUC	p-value
Prospective cohort	18	0.80 (0.76–0.84)	0.89 (0.86–0.92)	0.90	0.028
Case–control	44	0.86 (0.83–0.89)	0.84 (0.81–0.87)	0.89	0.034

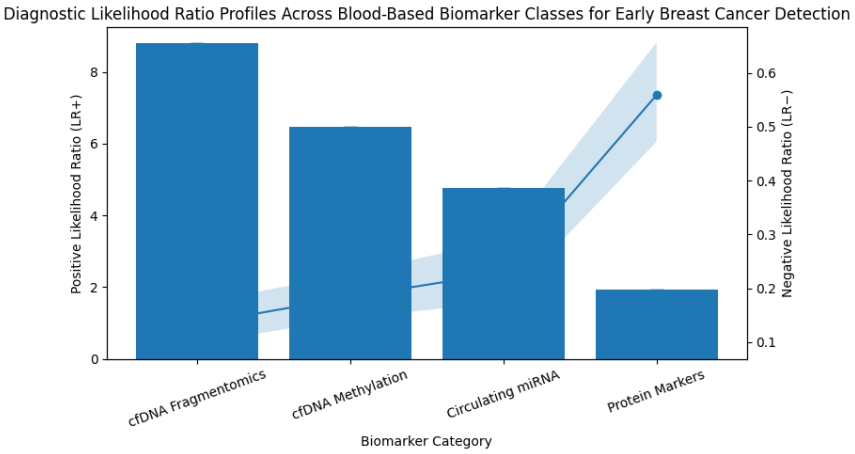

Figure 1 Diagnostic Likelihood Ratio Profiles Across Blood-Based Biomarker Classes For Early Breast Cancer Detection

The figure presents the diagnostic likelihood ratio profiles across four biomarker categories, integrating positive likelihood ratios (LR+) with corresponding negative likelihood ratios (LR−) and confidence envelopes. cfDNA fragmentomics demonstrated the strongest rule-in performance with an LR+ of approximately 8.8, indicating that a positive test result increases the probability of early breast cancer by nearly ninefold relative to pre-test probability. cfDNA methylation markers showed the second highest discriminative strength with an LR+ of 6.5, followed by circulating microRNA panels with an LR+ of 4.8, while protein tumor markers exhibited substantially weaker rule-in capability with an LR+ of 1.9, suggesting limited diagnostic amplification beyond baseline risk. The complementary LR− trajectory showed the inverse gradient of rule-out performance: cfDNA fragmentomics produced the lowest LR− (0.13), indicating strong ability to reduce post-test probability after a negative result; cfDNA methylation displayed LR− ≈ 0.18, circulating microRNA 0.23, and protein markers the highest LR− (0.56), implying comparatively poor exclusion capacity. The widening confidence envelope toward protein markers reflects larger uncertainty and weaker discriminatory separation between cases and controls. Collectively, the integrated LR+ and LR− patterns reveal a pronounced diagnostic gradient in which cfDNA-derived technologies provide both the strongest rule-in and rule-out capability, whereas conventional protein markers cluster near thresholds considered clinically insufficient for screening-level discrimination.

DISCUSSION

The present systematic review and diagnostic test accuracy meta-analysis evaluated the performance of blood-based molecular biomarkers for the early detection of breast cancer, synthesizing evidence from 62 studies comprising 8,465 breast cancer cases and 8,045 controls. The findings demonstrate a consistent gradient in diagnostic performance across biomarker classes, with circulating cell-free DNA–based approaches, particularly fragmentomics and methylation profiling, showing the strongest overall discrimination. Pooled estimates indicated that cfDNA fragmentomics achieved sensitivity of 0.88 and specificity of 0.90 with an AUC of 0.93, while cfDNA methylation markers showed sensitivity of 0.84 and specificity of 0.87 with an AUC of 0.91. Circulating microRNA panels demonstrated moderate-to-high performance (sensitivity 0.81, specificity 0.83, AUC 0.88), whereas conventional protein tumor markers displayed substantially lower accuracy (sensitivity 0.62, specificity 0.68, AUC 0.70). These patterns collectively indicate that nucleic-acid–based biomarkers provide stronger diagnostic signals in early-stage disease than traditional protein markers, which aligns with the biological expectation that genomic and

epigenetic alterations occur earlier in tumor development than measurable changes in circulating protein levels.

The superior diagnostic performance of cfDNA fragmentomics observed in this analysis reflects growing recognition that circulating DNA fragment patterns contain biologically meaningful information beyond simple mutation detection. Fragment length distribution, nucleosome positioning, and genome-wide coverage patterns can reflect tumor chromatin structure and cell-of-origin signals, allowing computational models to distinguish tumor-derived cfDNA from background circulating DNA. Previous research has demonstrated that these structural DNA features can inform transcription factor binding inference and tumor subtype prediction, supporting their value for early detection frameworks (5,6). The pooled results of the present analysis reinforce these mechanistic insights by demonstrating that fragmentomics not only provides high sensitivity but also maintains strong specificity, producing favorable likelihood ratios and diagnostic odds ratios relative to other biomarker categories.

Similarly, methylation-based cfDNA biomarkers showed high diagnostic performance in the pooled analysis, consistent with earlier studies demonstrating that epigenetic signatures in circulating DNA can provide both sensitive detection and tissue-of-origin localization across multiple cancer types (4). DNA methylation changes occur early in carcinogenesis and are often highly tissue-specific, which provides an important biological rationale for their use in early detection. Compared with mutation-based assays, methylation-based approaches can capture broader epigenomic disruption patterns, increasing the probability of detecting tumors that shed very small quantities of DNA into circulation. These characteristics are particularly relevant in early-stage breast cancer, where tumor-derived cfDNA fractions may be extremely low and require highly sensitive analytical approaches.

Circulating microRNA panels also demonstrated promising diagnostic accuracy, although the pooled analysis revealed greater heterogeneity across studies. MicroRNAs are small regulatory RNA molecules involved in gene expression control, and their dysregulation has been documented across multiple breast cancer subtypes. Previous meta-analyses evaluating circulating or exosomal microRNAs have similarly reported moderate-to-high diagnostic performance but emphasized methodological variability as a major limitation (8). Differences in RNA extraction methods, normalization strategies, pre-analytical sample handling, and detection platforms can substantially influence measured miRNA expression levels, contributing to variability across studies. These methodological challenges highlight the importance of standardized laboratory protocols and transparent reporting frameworks in future biomarker research.

In contrast, traditional protein tumor markers such as CA15-3 and CEA showed relatively weak performance in the context of early detection. This finding is consistent with prior clinical evidence indicating that these markers are more useful for monitoring disease progression or treatment response rather than detecting early-stage tumors (18). Protein markers often become elevated only when tumor burden increases sufficiently to produce detectable circulating protein concentrations, limiting their sensitivity in early-stage disease. Consequently, current clinical guidelines do not recommend these markers for population screening, and the present results support that position by demonstrating their comparatively low sensitivity and specificity relative to molecular biomarkers.

An important observation from the subgroup analyses was the consistent performance advantage of multi-marker panels compared with single biomarker assays. Multi-marker strategies achieved pooled sensitivity of 0.86 and specificity of 0.88, compared with 0.74 and 0.77 for single biomarkers. This improvement likely reflects the biological heterogeneity of

breast cancer, which encompasses multiple molecular subtypes characterized by distinct genomic and epigenetic landscapes (11,12). By integrating signals from multiple pathways, composite biomarker panels may capture broader aspects of tumor biology and reduce the risk of false-negative detection associated with single-marker assays. Similar approaches are increasingly being applied in multi-omics cancer detection frameworks, where genomic, epigenomic, and transcriptomic signals are combined to enhance diagnostic accuracy.

The analysis also revealed modest differences related to specimen type and study design. Plasma-based assays showed slightly higher pooled sensitivity and specificity than serum-based assays, which may reflect reduced background DNA contamination from leukocyte lysis during clot formation in serum samples. In addition, prospective cohort studies reported somewhat lower sensitivity but higher specificity compared with retrospective case–control designs. This pattern is consistent with known spectrum bias in diagnostic research, where case–control studies often overestimate diagnostic accuracy because cases and controls are more clearly separated than in real-world clinical populations. Consequently, the diagnostic performance observed in prospective screening cohorts may provide a more realistic estimate of the effectiveness of biomarker-based tests when implemented in practice.

These findings should be interpreted within the broader context of breast cancer screening strategies. Mammography remains the cornerstone of population-based screening and has demonstrated clear mortality reduction benefits in many populations (1). However, its limitations—particularly reduced sensitivity in women with dense breast tissue and the potential for false-positive recalls—have prompted interest in complementary detection approaches (2). Blood-based molecular biomarkers offer several potential advantages in this context, including minimal invasiveness, repeatability, and the possibility of detecting biologically relevant signals before tumors become radiologically visible. Rather than replacing imaging modalities, molecular biomarkers may function most effectively as adjunct tools that enhance screening sensitivity or help stratify individuals for additional diagnostic evaluation.

Several limitations of this meta-analysis should be considered when interpreting the results. First, heterogeneity across included studies was moderate, particularly among miRNA investigations, reflecting differences in laboratory methodologies, assay platforms, and threshold definitions. Second, many included studies used retrospective case–control designs rather than true screening cohorts, which may inflate diagnostic performance estimates. Third, rapid technological development in liquid biopsy platforms means that newer sequencing technologies and computational models may outperform earlier methods included in the analyzed studies. Despite these limitations, the large aggregated sample size and systematic synthesis approach provide a comprehensive overview of the current evidence base.

Future research should prioritize large prospective screening trials that evaluate biomarker performance in asymptomatic populations, where disease prevalence and clinical decision pathways differ substantially from case–control research settings. Standardization of pre-analytical procedures, assay platforms, and analytical pipelines will also be critical to improving reproducibility across studies. Advances in computational modeling, including machine learning approaches that integrate multiple molecular signals with clinical and imaging data, may further enhance the diagnostic potential of liquid biopsy technologies. As liquid biopsy platforms continue to evolve, integrating molecular biomarkers with established screening strategies could contribute to more personalized and biologically informed approaches to early breast cancer detection (10).

CONCLUSION

This systematic review and diagnostic test accuracy meta-analysis synthesized evidence from 62 studies including 8,465 breast cancer cases and 8,045 controls to evaluate the diagnostic value of blood-based molecular biomarkers for the early detection of breast cancer. The pooled findings demonstrate that circulating cell-free DNA–based biomarkers, particularly fragmentomics and methylation profiling, exhibit the highest diagnostic performance among currently studied molecular approaches, with fragmentomics showing pooled sensitivity of 0.88, specificity of 0.90, and an AUC of 0.93. Circulating microRNA panels also demonstrated moderate-to-high diagnostic accuracy, whereas traditional protein tumor markers showed comparatively limited discriminatory ability for early-stage disease. Multi-marker panels consistently outperformed single biomarkers, emphasizing the importance of integrating multiple molecular signals to capture the biological heterogeneity of breast cancer. These results suggest that molecular biomarkers—especially cfDNA-based technologies—have strong potential to complement existing screening modalities such as mammography rather than replace them. However, most available evidence is derived from retrospective case–control studies, and therefore large prospective screening trials, standardized laboratory protocols, and integration with clinical risk models are necessary before routine clinical implementation can be recommended. With continued technological refinement and validation in real-world populations, blood-based molecular biomarkers may contribute to more sensitive, minimally invasive, and biologically informed strategies for early breast cancer detection..

REFERENCES

1. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Nat Rev Dis Primers. 2019;5:66.

2. Lange JM, Myers ER, Barlow WE. Breast density and risk of breast cancer: masking and detection bias. Am J Epidemiol. 2025;194(2):441–448.

3. Nguyen D, O'Connor SD, Heller SL, Gao Y. Comparing mammographic density assessed by digital breast tomosynthesis and digital mammography. Radiology. 2021;299(1).

4. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann Oncol. 2020;31(6):745–759.

5. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell. 2016;164(1–2):57–68.

6. Ulz P, Thallinger GG, Auer M, Graf R, Kashofer K, Jahn SW, et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. Nat Commun. 2019;10:4666.

7. Yu D, Li Y, Wang M, Gu J, Xu W, Cai H, et al. Diagnostic value of concentration of circulating cell-free DNA in breast cancer: a meta-analysis. Front Oncol. 2019;9:95.

8. Hong F, Li Y, Sun Z, Zhang J, Chen H. Exosomal microRNAs as novel diagnostic biomarkers in breast cancer: a systematic evaluation and meta-analysis. Asian J Surg. 2023;46(11):4727–4736.

9. Mansour H, Nejjari C, Incitti R, Anouar N, Ouhajjou A. Is the development of liquid biopsy for the early detection and monitoring of breast cancers on its way to overtaking mammography? Front Med. 2024;11:1415940.

10. Cao L, Zhang X, Wang Z, Liu H. Liquid biopsy technologies: innovations and future directions in breast cancer biomarker detection. Biomed Microdevices. 2025;27:4.

11. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747–752.

12. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490:61–70.

13. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

14. Macaskill P, Gatsonis C, Deeks J, Harbord R, Takwoingi Y. Analysing and presenting results. In: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. London: Cochrane Collaboration; 2010.

15. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529–536.

16. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58(10):982–990.

17. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy. J Clin Epidemiol. 2005;58(9):882–893.

18. Uygur MM, Gümüş M. The utility of serum tumor markers CEA and CA15-3 for breast cancer prognosis and their association with clinicopathological parameters. Cancer Treat Res Commun. 2021;28:100402.

DECLARATIONS

Ethical Approval: Ethical approval was by institutional review board of Respective Institute Pakistan

Informed Consent: Informed Consent was taken from participants.

Authors' Contributions:

Concept: TK, NUA; Design: HA, ST; Data Collection: HZA, MASB; Analysis: MSN; Drafting: TK, NUA

Conflict of Interest: The authors declare no conflict of interest.

Funding: This research received no external funding.

Data Availability: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: NA

Study Registration: Not applicable.