



OPEN Integrating circulating microRNAs with epidemiological factors enhances breast cancer detection across subtypes: the MCC-Spain study

Inés Gómez-Acebo^{1,2,3,20}✉, Sara Valero-Dominguez^{1,2,20}, Javier Llorca¹, Jessica Alonso-Molero^{1,2,3}, Thalía Belmonte⁴, Gemma Castaño-Vinyals^{3,5,6,7}, Víctor Moreno^{3,8,9,10}, Aina Romaguera^{11,12}, Pilar Amiano^{3,13}, Juan Alguacil^{3,14}, Vicente Martín¹⁵, Beatriz Pérez-Gómez^{3,16}, Rosana Burgui^{3,17}, Ana Molina-Barceló¹⁸, Paz Rodríguez-Cundín¹⁹, Manolis Kogevinas^{3,5,6,7}, Marina Pollán^{3,16} & Trinidad Dierssen-Sotos^{1,2,3}

Circulating microRNAs (miRNAs) are promising non-invasive biomarkers for cancer detection; however, their diagnostic performance across breast cancer molecular subtypes and their incremental value beyond demographic and epidemiological variables remain incompletely characterized. We conducted a multicenter case–control study including 317 breast cancer cases and 127 population-based controls. Serum levels of 44 literature-derived miRNAs were quantified by RT-qPCR. Feature selection was performed using LASSO penalization, followed by multivariable logistic regression to estimate odds ratios (ORs) with 95% confidence intervals (CIs). Models were adjusted for demographic and epidemiological covariates. Predictive performance was assessed using repeated fivefold cross-validation and reported as area under the curve (AUC) with bootstrap bias-corrected 95% CIs. Incorporating demographic and epidemiological covariates enhanced discrimination overall (AUC = 0.908 vs. 0.802 unadjusted) and across subtypes. The most notable improvements were observed in Luminal A (0.896 vs. 0.751) and Luminal B (0.894 vs. 0.768), while HER2-positive and Basal-like tumors already showed high performance (AUC = 0.965 and 0.989, respectively). Among the 12 miRNAs selected by LASSO, miR-21-5p and miR-423-3p were consistently elevated in cases, particularly in HER2-positive and Basal-like tumors, whereas miR-101-3p, miR-146a-5p, and miR-29a-3p showed reproducibly lower levels across multiple subtypes, consistent with oncogenic and tumor-suppressive roles, respectively. Circulating miRNA signatures, especially when integrated with demographic and epidemiological information, demonstrate high discriminatory power for breast cancer detection across molecular subtypes. These results support subtype-aware, minimally invasive strategies for screening and risk stratification using miRNA-based models. Prospective validation in independent cohorts is warranted to confirm clinical utility.

Keywords Breast cancer, Molecular subtypes, Circulating miRNA, LASSO, ROC-AUC, Risk stratification

¹Department of Preventive Medicine and Public Health, University of Cantabria, Santander, Spain. ²IDIVAL-Valdecilla Health Research Institute, Santander, Spain. ³Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Institute of Health Carlos III, Madrid, Spain. ⁴Instituto de Investigación Sanitaria del Principado de Asturias, Oviedo, Spain. ⁵Barcelona Institute for Global Health (ISGlobal), Barcelona, Spain. ⁶Department of Medicine and Life Sciences, Universitat Pompeu Fabra (UPF), Barcelona, Spain. ⁷Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain. ⁸Colorectal Cancer Group, ONCOBELL Program, Instituto de Investigación Biomédica de Bellvitge (IDIBELL), Hospitalet de Llobregat, Barcelona, Spain. ⁹Oncology Data Analytics Program, Catalan Institute of Oncology, Hospitalet de Llobregat, Barcelona, Spain. ¹⁰Department of Clinical Sciences, Faculty of Medicine and Health Sciences and Universitat de Barcelona Institute of Complex Systems (UBICS), University of Barcelona, Barcelona, Spain. ¹¹Josep Carreras Leukaemia Research Institute, Badalona, Spain. ¹²Epidemiology Unit and Girona Cancer Registry, Oncology Coordination Plan, Department of Health, Autonomous Government

of Catalonia, Catalan Institute of Oncology (ICO), Girona Biomedical Research Institute (IdIBGi-CERCA), Girona, Spain. ¹³Epidemiology of Chronic and Communicable Diseases Group, Biodonostia Health Research Institute, Basque Government, San Sebastián, Spain. ¹⁴Natural Resources, Health and Environment Research Centre (RENSMA), Universidad de Huelva-Campus El Carmen, Huelva, Spain. ¹⁵Gene-Environment Interactions and Health Research Group (GIIGAS), Institute of Biomedicine (IBIOMED), University of León, Leon, Spain. ¹⁶National Centre for Epidemiology, Carlos III Institute of Health, Madrid, Spain. ¹⁷Institute of Public and Occupational Health of Navarre (ISPLN), Pamplona, Spain. ¹⁸Cancer and Public Health Unit Foundation for the Promotion of Health and Biomedical Research (FISABIO-Salud Pública) in the Valencia Region, Valencia, Spain. ¹⁹Preventive Medicine, Hospital Universitario Marqués de Valdecilla, Santander, Spain. ²⁰Inés Gómez-Acebo and Sara Valero-Dominguez contributed equally to this work. ✉email: ines.gomez@unican.es

In 2022, breast cancer was the most frequently diagnosed malignancy and the leading cause of cancer-related death among women worldwide, accounting for over 2.3 million new cases and approximately 670,000 deaths annually^{1,2}. In Spain, it remains the most common cancer among women, with more than 37,682 new cases in 2024, and it is the third leading cause of cancer mortality among Spanish women³. Despite advances in early detection and treatment, prognosis varies considerably depending on tumor subtype, stage at diagnosis, and other clinical and pathological characteristics⁴.

Breast cancer is a biologically heterogeneous disease⁴. Current molecular classification systems categorize it into four major intrinsic subtypes—Luminal A, Luminal B, HER2-positive, and triple-negative breast cancer (Basal-like)—each with distinct clinical, prognostic, and therapeutic profiles^{4–6}. Differences also exist in patterns of detection: luminal tumors are more often screen-detected, while HER2-positive and Basal-like tumors tend to present symptomatically, at younger ages and with more aggressive features^{4,7}. This biological and clinical diversity underscores the need for complementary biomarkers that can improve risk stratification and support individualized clinical decision-making.

Early detection plays a crucial role in reducing breast cancer mortality and improving long-term outcomes. By identifying cancer at earlier, more treatable stages, the likelihood of successful treatment and survival increases significantly^{8,9}.

Mammography remains the cornerstone of population-based screening programs. While it has contributed significantly to mortality reduction, its sensitivity is limited, especially in women with dense breast tissue^{9,10}. Consequently, it is often complemented by ultrasound, while magnetic resonance imaging (MRI), with higher sensitivity, is reserved for high-risk patients or cases with inconclusive mammographic findings⁹.

Despite these diagnostic tools, contemporary challenges in breast cancer diagnosis and treatment persist. These include incomplete implementation of screening programs, limited methods for real-time monitoring of tumor progression, and the lack of robust, subtype-specific biomarkers to guide therapeutic decisions⁹.

MicroRNAs (miRNAs) are small (~22 nucleotides), non-coding RNAs that regulate gene expression post-transcriptionally and are implicated in several hallmarks of cancer, including proliferation, apoptosis, invasion, metastasis, and immune evasion^{5,11–13}. It is estimated that they regulate up to 60% of protein-coding genes^{14,15}, and depending on the cellular context, they can function as tumor suppressors or oncogenes^{11,13}.

Importantly, miRNAs can be detected in extracellular fluids such as serum and plasma, where they are highly stable due to encapsulation in extracellular vesicles or binding to protein complexes^{11,16}. These properties make circulating miRNAs attractive candidates for non-invasive cancer biomarkers, with potential applications in diagnosis, prognosis, and treatment monitoring^{13,16}.

Several studies have proposed circulating miRNA panels associated with breast cancer risk or progression, but many have limitations such as small sample sizes, inconsistent methodologies, or lack of molecular subtype-specific analyses^{17,18}. In our previous work using the MCC-Spain cohort, we identified serum-based miRNA signatures capable of distinguishing breast cancer cases from healthy controls, as well as across different clinical states¹⁹. However, the potential of circulating miRNAs to differentiate intrinsic molecular subtypes of breast cancer remains underexplored, particularly in multicenter case-control studies with harmonized clinical annotation and robust statistical modelling.

Recent studies have explored the diagnostic and prognostic value of circulating miRNAs in breast cancer, including early efforts to classify molecular subtypes based on serum or plasma profiles^{20,21}. Some subtype-specific patterns have been suggested, such as downregulation of miR-101-3p in Basal-like²² or upregulation of miR-423-3p in hormone receptor-positive tumors²³, but findings remain inconsistent due to differences in patient selection, sample processing, and statistical methodology¹⁹.

To date, few population-based studies have systematically evaluated circulating miRNAs across the major breast cancer subtypes using multivariable penalized regression models with internal validation. Evidence is particularly scarce from case-control designs with matched controls, harmonized pre-analytical procedures, and clinically confirmed molecular subtypes.

In this context, the primary objective of the present study was to evaluate the association between a panel of 46 circulating miRNAs and breast cancer molecular subtype (Luminal A, Luminal B, HER2-positive, and Basal-like), using serum samples from a multicenter case-control study. Supplementary Table 1 provides a detailed overview of the selected miRNAs, including their putative biological functions and previously reported associations with breast cancer biology.

Methods

Study population

The MCC-Spain study is a multicenter, population-based case-control study conducted across ten Spanish provinces (Asturias, Barcelona, Cantabria, Girona, Guipúzcoa, Huelva, León, Madrid, Navarra, and Valencia). For the present analysis, a total of 444 women were included: 317 incident breast cancer cases and 127 population-

based controls. Participants from the pilot Screening Phase (n = 40) were excluded to ensure independence of the validation cohort.

Breast cancer cases included women aged 27–85 years with newly diagnosed, histologically confirmed invasive breast cancer between 2008 and 2013. Molecular subtypes, Luminal A-like, Luminal B-like, HER2-enriched-like, and Basal-like, were defined based on immunohistochemical (IHC) profiling, including estrogen receptor (ER), progesterone receptor (PR), HER2 status, and Ki-67 proliferation index, in accordance with the St. Gallen International Consensus Guidelines²⁴. In this study, the Basal-like subtype corresponds to triple-negative breast cancer (ER-negative, PR-negative, HER2-negative) as defined by IHC.

Controls were women with no previous history of cancer, residing in the same catchment areas as cases. They were frequency-matched to cases by age (in 5-year intervals) and geographic region, and were randomly selected from general practitioner lists at participating primary healthcare centers.

Recruitment protocols and data collection procedures have been described in detail elsewhere^{25,26}. All participants provided written informed consent. The study protocol was approved by the Ethics Committees of all participating institutions and was registered with the Spanish Data Protection Agency (Registration No. 2102672171). Further details are available at <http://www.mccspain.org>. All procedures were performed in accordance with relevant ethical guidelines and regulations.

Data collection

Face-to-face interviews conducted by trained personnel gathered information on demographic and epidemiological covariates, including age, body mass index (BMI), educational level, menopausal status, family socioeconomic status, smoking habits, alcohol consumption, use of hormonal contraceptives, history of hormone replacement therapy (HRT), and leisure-time physical activity. Dietary intake was assessed using a validated food frequency questionnaire and expressed as grams per day (g/day) and total energy intake (kcal/day).

Clinical-pathological information was obtained from medical records, including tumor receptor status (estrogen receptor [ER], progesterone receptor [PR], and HER2), Ki-67 proliferation index, and histological subtype. Breast cancer cases were stratified into molecular subtypes (Luminal A, Luminal B, HER2-positive, and Basal-like) according to the St. Gallen International Consensus Guidelines²⁴.

Biological samples

Blood samples were collected from both cases and controls in the early morning after an overnight fast, following a standardized protocol. Samples were centrifuged at 3000 × g for 20 min at 10 °C, followed by a second centrifugation at 15,000 × g for 10 min at the same temperature to remove residual cellular debris. The resulting serum aliquots were stored at –80 °C until analysis.

Screening phase

The selection of candidate miRNAs followed a two-phase design (screening and validation) consistent with our previous work¹⁹, combining a structured literature review with an exploratory small RNA sequencing phase conducted in 40 serum samples. For the present study, we expanded this framework to specifically evaluate molecular subtypes of breast cancer (Luminal A, Luminal B, HER2-positive, and Basal-like), incorporating a larger and clinically diverse set of cases, including more advanced tumors. An updated review of the literature (2020–2024) focusing on subtype-specific circulating miRNAs was also performed to refine biomarker selection and develop subtype-specific predictive models.

Ten control women and 30 breast cancer cases were selected. As this pilot set was originally recruited for a previous analysis of detection modes, the distribution of molecular subtypes was unbalanced (13 Luminal A, 11 Luminal B, 2 Basal-like, 0 HER2-positive, and 4 unclassified/other; Supplementary Table 5). Consequently, this screening set was not powered to identify subtype-specific markers for HER2 or Basal-like tumors, necessitating the integration of literature-based candidates for these subgroups. These 40 participants were excluded from the validation phase to avoid overlap.

RNA isolation and library preparation

RNA was isolated from serum samples using the miRNeasy Serum/Plasma Kit (QIAGEN) at QIAGEN Genomic Services, following the manufacturer's instructions. Libraries were prepared with the QIAseq miRNA Library Kit (QIAGEN), and quality control was performed with a Bioanalyzer 2100 (Agilent) or TapeStation 4200 (Agilent). A total of 200 µL of RNA were used to generate miRNA NGS libraries. Adapters containing Unique Molecular Indexes (UMIs) were ligated to eliminate amplification bias. The RNA was reverse-transcribed into cDNA, amplified by PCR, and purified. Library pools were quantified by qPCR and sequenced on a NextSeq 500 (Illumina) platform.

Bioinformatics processing

Raw FASTQ files were generated for each sample. Reads were processed using Cutadapt for UMI correction, Bowtie2 for alignment to *Homo sapiens* miRNA sequences from miRBase v22.1, and edgeR (Bioconductor) for differential expression analysis. Counts were normalized using the Trimmed Mean of M-values (TMM) method and log₂-transformed for downstream statistical analyses. Sequence annotation was performed against the GRCh37 human reference genome and miRBase v22.1.

Quality control

Rigorous quality control procedures were applied, including:

- (i) UMI collapsing (unique sequence/UMI combination required),
- (ii) Phred quality score thresholds,
- (iii) Read length > 15 nucleotides,
- (iv) Mapping to GRCh37 reference genome, and
- (v) Background filtering to remove low-abundance reads.

Samples not meeting these criteria were excluded from the dataset.

The most promising miRNAs identified in this exploratory screening were considered for inclusion in the validation panel, along with additional candidates selected from external sources.

Validation of miRNAs based on previous studies

This phase included 444 participants (127 cancer-free controls and 317 breast cancer cases) recruited from the MCC-Spain study across ten Spanish provinces. Blood samples were collected prior to any oncological treatment (surgery, chemotherapy, or radiotherapy). Cases were stratified into molecular subtypes (Luminal A, Luminal B, HER2-positive, and Basal-like) according to the St. Gallen International Consensus Guidelines²⁴.

A total of 46 miRNAs were selected for validation based on two complementary sources: (1) updated literature review on subtype-specific circulating signatures (2020–2024)^{27–119}, and (2) differentially expressed miRNAs identified in our screening phase.

Two endogenous control miRNAs (let-7d-5p and let-7i-5p) were monitored for technical stability and were not included in differential expression or predictive modeling. Expression values were normalized using TMM. Differential expression analysis was conducted for 44 miRNAs, and 12 were retained using LASSO regression models. Candidate miRNAs were chosen according to prior evidence and biological plausibility as circulating biomarkers of breast cancer (see Supplementary Table 1).

RNA extraction and RT-qPCR

The miRNA expression analysis was performed at QIAGEN Genomic Services using standardized protocols. Serum samples were thawed on ice and centrifuged at 3,000 × g for 5 min at 4 °C. From each sample, 200 µL of serum were transferred into FluidX tubes, and 60 µL of Buffer RPL (containing 1 µg of carrier RNA and an RNA spike-in template mix) were added. After vortexing (1 min) and incubation (7 min at room temperature), 20 µL of Buffer RPP was added. Total RNA was extracted with the miRNeasy Serum/Plasma Advanced Kit (QIAGEN) and eluted into 50 µL.

For reverse transcription, 2 µL of extracted RNA were used in 10 µL reactions with the miRCURY LNA RT Kit (QIAGEN). The cDNA was diluted 50 × and analyzed via quantitative real-time PCR (RT-qPCR) using a custom miRCURY LNA PCR panel and SYBR Green master mix (QIAGEN). Amplifications were performed on a LightCycler® 480 Real-Time PCR System (Roche) in 384-well plates. Negative controls without RNA templates were included in each batch.

Data normalization and quality control

Amplification curves were analyzed with Roche LC software using the second derivative method to determine Cq values and assess melting curve profiles. Endogenous miRNAs let-7d-5p and let-7i-5p, consistently detected across all samples, were used for normalization. Data were normalized using the TMM method¹²⁰, and normalized values were subsequently log₂-transformed for statistical analyses.

Additional QC criteria included:

- (i) exclusion of reactions with > 1 melting temperature,
- (ii) melting temperatures deviating from reference values, and
- (iii) low amplification efficiency.

RT-qPCR assays were performed in duplicate (or triplicate when feasible) to ensure reliability and minimize measurement bias.

Statistical analysis

Analyses were performed for the overall model (all breast cancer cases vs. controls) and for models stratified by molecular subtype (Luminal A, Luminal B, HER2-positive, and Basal-like).

Normalized expression values (log₂-transformed) from controls were compared with each case group using Student's t-test without further adjustment. Results are reported as log fold change (logFC), *p* values, and false discovery rate (FDR), calculated using the Benjamini–Hochberg method¹²¹. All reported *p* values are two-sided and were not corrected for multiple testing. Consequently, associations with a *p* value < 0.05 are considered nominal and should be interpreted as exploratory. A positive logFC indicates overexpression in cases relative to controls, whereas a negative logFC reflects underexpression. The greater the absolute logFC value, the larger the difference between groups.

To assess the predictive performance of circulating miRNAs, we applied LASSO (least absolute shrinkage and selection operator) penalized logistic regression¹²². This approach enables the inclusion of multiple predictors while shrinking coefficients of less informative variables toward zero, thus generating parsimonious models. The analysis began with an initial model that included all candidate miRNAs selected during the validation phase. The regularization parameter (λ) was chosen through tenfold cross-validation.

Odds ratios (ORs) with 95% confidence intervals (CIs) were obtained by refitting standard logistic regression models that included only the variables selected by LASSO. To ensure out-of-sample performance, AUC values were estimated using repeated stratified fivefold cross-validation. Model discrimination, reflecting the ability to

distinguish cases from controls, was evaluated by calculating the area under the receiver operating characteristic curve (AUC) with bootstrap bias-corrected 95% CIs.

Each model was re-estimated after adjusting for demographic and epidemiological covariates (age, body mass index [BMI], menopausal status, family history of breast cancer, smoking status, alcohol intake, use of hormonal contraceptives, history of hormone replacement therapy, educational level, and recruitment area). Variables that showed notable differences between cases and controls (Supplementary Table 2) were considered potential confounders and included as covariates, together with additional covariates supported by prior epidemiological evidence. This strategy ensured appropriate control for confounding.

All statistical analyses were performed using Stata 18/SE (StataCorp, College Station, TX, USA).

Results

Characteristics of the study population

Supplementary Table 2 summarizes the baseline characteristics of the 444 women included in the analysis (317 breast cancer cases and 127 controls). Several variables showed different distributions between cases and controls, including menopausal status (premenopausal women: 38.2% vs. 26.8%; $p = 0.023$), family history of breast cancer (29.4% vs. 15.6%; $p = 0.003$), hormonal contraceptive use (50.3% vs. 37.0%; $p = 0.009$), and hormone replacement therapy (less frequent in cases; $p = 0.056$). The distribution of recruitment areas also showed differences between groups ($p < 0.001$).

In contrast, no notable differences were observed in age, body mass index (BMI), educational level, alcohol intake, smoking status, parity, age at first childbirth, total energy intake, or adherence scores to the World Cancer Research Fund (WCRF) lifestyle guidelines.

Table 1 further details the distribution of breast cancer molecular subtypes, showing that among the 317 cases, Luminal A was the most frequent ($n = 142$; 44.8%), followed by Luminal B ($n = 42$; 13.2%), HER2-positive ($n = 47$; 14.8%), and Basal-like ($n = 86$; 27.1%).

Clinical and pathological characteristics of breast cancer cases

Supplementary Table 3 describes the clinical and pathological features of the 317 women diagnosed with breast cancer, including tumor histology, stage, receptor status, and grade. This description illustrates the clinical heterogeneity of the study population, which was considered in the subtype-stratified analyses and is essential to interpreting miRNA-based model performance.

The most common histological subtype was ductal carcinoma (83.9%), followed by lobular carcinoma (4.4%) and other subtypes (10.1%).

Tumor staging data revealed that nearly half of the tumors were classified as T1 (51.7%), with T2 tumors accounting for 29.0%. In situ tumors (Tis) and T0 lesions each represented 2.5% of cases. Regional lymph node involvement was present in 45.1% of patients (N1–N3), while 2.8% had distant metastases at diagnosis (M1). AJCC clinical staging indicated that most patients were diagnosed at early stages (Stage I–II: 70.3%), with 14.8% at stage III and only 2.8% at stage IV. Based on diagnosis timing, 79.3% were classified as early-stage cases.

Regarding receptor status, 57.7% of tumors were estrogen receptor (ER)-positive, 51.1% progesterone receptor (PR)-positive, and 24.0% HER2-positive. Based on surrogate definitions, molecular subtypes were distributed as follows: Luminal A (44.8%), Luminal B (13.2%), HER2-positive (14.8%), and Basal-like (27.1%).

Histological grade was available in 64% of cases. Among those with reported data, 27.4% of tumors were poorly differentiated (grade III), 22.7% moderately differentiated (grade II), and 14.2% well differentiated (grade I).

miRNA signatures in breast cancer

Table 2 presents the multivariate analysis comparing miRNA expression between breast cancer cases and controls. Among the miRNAs associated with an increased risk of breast cancer, miR-21-5p showed the highest odds ratio in the fully adjusted model (OR = 27.11, 95% CI 6.98–105.33), followed by miR-423-3p (OR = 4.65, 95% CI 1.99–10.82), miR-331-3p (OR = 2.19, 95% CI 1.14–4.24), miR-139-5p (OR = 1.81, 95% CI 1.16–2.83), and miR-141-3p (OR = 1.44, 95% CI 1.03–2.01). These miRNAs tended to be overexpressed in cases compared to controls.

Conversely, several miRNAs were associated with a decreased risk of breast cancer. Notably, miR-146a-5p demonstrated a strong inverse association (OR = 0.23, 95% CI 0.10–0.52), as did miR-29a-3p (OR = 0.31, 95% CI 0.17–0.57), miR-101-3p (OR = 0.34, 95% CI 0.16–0.75), miR-142-3p (OR = 0.37, 95% CI 0.17–0.81), and miR-186-5p (OR = 0.36, 95% CI 0.19–0.70). These results are consistent with lower expression levels in cases across different models.

Group of analysis	Number of patients (%)
Control Group	127 (28.6%)
Luminal A	142 (31.98%)
Luminal B	42 (9.46%)
HER2-positive	47 (10.59%)
Basal-like Subtype	86 (19.37%)

Table 1. Distribution of study participants by molecular subtype of breast cancer and control status.

In addition to the model adjusted for all miRNAs included in the final panel, a more comprehensive model was developed that further adjusted for demographic and epidemiological variables (age, menopausal status, BMI, family history of breast cancer, smoking status, alcohol intake, use of hormonal contraceptives, history of HRT, education level, and region of recruitment). These additional adjustments did not substantially alter the direction or magnitude of the associations, reinforcing the robustness of the identified profiles. For instance, miR-21-5p remained the most strongly associated with increased breast cancer risk, while miR-101-3p, miR-146a-5p, and miR-29a-3p consistently showed associations consistent with lower expression levels in cases, potentially reflecting inverse correlations with risk.

These consistent associations across models suggest that a subset of circulating miRNAs may serve as candidate biomarkers warranting further evaluation for breast cancer risk assessment.

Subtype-specific associations

In Luminal A tumors (Table 3), miR-423-3p and miR-21-5p showed marked associations (OR=3.81 and OR=14.16, respectively). In contrast, miR-186-5p (OR=0.19), miR-29a-3p (OR=0.30), and miR-146a-5p (OR=0.32) showed consistent inverse associations. miR-141-3p was also retained in the adjusted model (OR=1.61), suggesting that its association became more evident after accounting for demographic and epidemiological variables.

The Luminal B subgroup was characterized by an inverse association for miR-186-5p (OR=0.27). miR-101-3p showed an OR of 0.39 in crude analysis and 0.45 after full adjustment, indicating a stable direction of association with modest attenuation (Table 4).

Among HER2-positive tumors, miR-21-5p showed an extremely elevated OR (1,300,000), but with an extremely wide confidence interval, reflecting limited stability of the estimate—likely influenced by adjustment for recruitment region and the small sample size. miR-101-3p and miR-146a-5p showed strong inverse associations (both OR ≈ 0.01). Other miRNAs such as miR-29a-3p and miR-331-3p showed extreme ORs with wide intervals, again reflecting instability (Table 5).

In Basal-like tumors, the largest OR was again observed for miR-21-5p (OR=4.6 × 10⁶), although the very wide confidence intervals suggest model instability, likely exacerbated by adjustment for recruitment region. miR-331-3p and miR-423-3p also showed strong positive associations. In contrast, miR-101-3p, miR-142-3p, miR-146a-5p, and miR-29a-3p consistently showed ORs close to zero, reflecting inverse associations that remained stable in direction after adjustment (Table 6).

Subtype-specific circulating miRNA profiles

Table 7 synthesizes the patterns across subtypes. miR-21-5p and miR-423-3p tended to show higher odds across most subtypes. Conversely, miR-146a-5p, miR-29a-3p, and miR-101-3p were consistently retained with ORs below 1, suggesting potential inverse associations. miR-186-5p showed divergent behavior: inverse in Luminal A and B but positive in HER2-positive tumors.

As expected, models for HER2-positive and Basal-like tumors yielded extreme OR values with wide confidence intervals, likely reflecting smaller sample sizes and model instability, potentially exacerbated by adjustment for recruitment area.

miRNA	TMM cases*	TMM controls*	log ₂ (FC)	Fold change*	Odds ratio (95% CI)**	P**	Odds ratio (95% CI)***	P***
miR-101-3p	1.390	1.965	-0.575	0.671	0.31 (0.19–0.52)	<0.001	0.34 (0.16–0.75)	0.007
miR-139-5p	-3.907	-4.494	0.587	1.502	1.55 (1.15–2.11)	0.004	1.81 (1.16–2.83)	0.009
miR-142-3p	3.084	2.906	0.178	1.131	0.55 (0.31–0.96)	0.037	0.37 (0.17–0.81)	0.013
miR-186-5p	-3.680	-3.501	-0.179	0.883	0.72 (0.47–1.11)	0.137	0.36 (0.18–0.70)	0.003
miR-21-3p	-4.690	-4.970	0.280	1.214	1.27 (0.94–1.70)	0.116	1.18 (0.76–1.84)	0.466
miR-29a-3p	-0.036	0.283	-0.319	0.802	0.58 (0.36–0.91)	0.0019	0.31 (0.17–0.57)	<0.001
miR-331-3p	-3.649	-3.835	0.186	1.138	1.84 (1.14–2.95)	0.012	2.19 (1.14–4.24)	0.019
miR-423-3p	-0.979	-1.276	0.297	1.229	3.13 (1.78–5.49)	<0.001	4.65 (1.99–10.82)	<0.001
miR-146a-5p	0.322	0.474	-0.152	0.900	0.23 (0.12–0.43)	<0.001	0.23 (0.10–0.52)	<0.001
miR-141-3p	-5.567	-5.674	0.107	1.077	1.16 (0.92–1.45)	0.218	1.44 (1.03–2.01)	0.033
miR-21-5p	3.437	3.473	-0.036	0.975	4.58 (2.09–10.01)	0.006	27.11 (6.98–105.33)	<0.001

Table 2. Differential expression of candidate circulating miRNAs in breast cancer cases versus controls. TMM (trimmed mean of M-values): Method used for miRNA expression normalization. Fold change (FC): Ratio of expression between cases and controls (FC > 1 indicates overexpression; FC < 1 indicates underexpression). log₂(FC) is the base-2 logarithmic transformation of FC. Crude (unadjusted) vs. Adjusted (** and ***): The unadjusted model does not control for any confounders; the second model is adjusted for all miRNAs included in the final set; the third model also adjusts for age, menopausal status, body mass index (BMI), family history of breast cancer, smoking status, alcohol intake, use of hormonal contraceptives, history of hormone replacement therapy (HRT), education level, and region of recruitment. P values are nominal and not corrected for multiple testing; results should be interpreted as exploratory.

miRNA	TMM cases*	TMM controls*	log ₂ (FC)	Fold change*	Odds ratio (95% CI)**	P**	Odds ratio (95% CI)***	P***
miR-101-3p	1.537	1.965	-0.428	0.743	0.64 (0.37–1.11)	0.110	1.21 (0.49–3.02)	0.680
miR-139-5p	-4.117	-4.494	0.377	1.299	1.35 (0.98–1.85)	0.069	1.67 (0.99–2.79)	0.051
miR-186-5p	-3.863	-3.501	-0.362	0.778	0.51 (0.30–0.84)	0.008	0.19 (0.08–0.46)	<0.001
miR-29a-3p	-0.004	0.283	-0.287	0.820	0.58 (0.36–0.94)	0.028	0.30 (0.14–0.62)	0.001
miR-324-5p	-3.341	-3.397	0.056	1.040	1.40 (0.84–2.31)	0.194	1.05 (0.50–2.22)	0.898
miR-331-3p	-3.595	-3.835	0.240	1.181	1.55 (0.98–2.46)	0.062	1.68 (0.77–3.68)	0.190
miR-423-3p	-1.032	-1.276	0.244	1.184	2.51 (1.37–4.60)	0.003	3.81 (1.53–9.52)	0.004
miR-146a-5p	0.438	0.474	-0.036	0.975	0.35 (0.18–0.69)	0.002	0.32 (0.12–0.84)	0.020
miR-141-3p	-5.627	-5.674	0.047	1.033	1.17 (0.90–1.52)	0.254	1.61 (1.04–2.48)	0.032
miR-21-5p	3.373	3.473	-0.100	0.933	2.50 (1.00–5.68)	0.029	14.15 (3.13–63.97)	0.001

Table 3. Differential expression of candidate circulating miRNAs in luminal A breast cancer cases versus controls. TMM (trimmed mean of M-values): Method used for miRNA expression normalization. Fold change (FC): Ratio of expression between cases and controls (FC > 1 indicates overexpression; FC < 1 indicates underexpression). log₂(FC) is the base-2 logarithmic transformation of FC. Crude (unadjusted) vs. Adjusted (** and ***): The unadjusted model does not control for any confounders; the second model is adjusted for all miRNAs included in the final set; the third model also adjusts for age, menopausal status, body mass index (BMI), family history of breast cancer, smoking status, alcohol intake, use of hormonal contraceptives, history of hormone replacement therapy (HRT), education level, and region of recruitment. *P* values are nominal and not corrected for multiple testing; results should be interpreted as exploratory.

miRNA	TMM cases*	TMM controls*	log ₂ (FC)	Fold change*	Odds ratio (95% CI)**	P**	Odds ratio (95% CI)***	P***
miR-101-3p	1.385	1.965	-0.580	0.669	0.39 (0.19–0.82)	0.012	0.45 (0.12–1.68)	0.233
miR-139-5p	-3.994	-4.494	0.500	1.414	1.37 (0.85–2.20)	0.196	1.62 (0.75–3.49)	0.221
miR-186-5p	-3.822	-3.501	-0.321	0.801	0.67 (0.32–1.38)	0.276	0.30 (0.10–0.88)	0.053
miR-423-3p	-0.944	-1.276	0.332	1.259	2.25 (1.06–4.77)	0.034	2.34 (0.53–3.39)	0.110

Table 4. Differential expression of candidate circulating miRNAs in Luminal B breast cancer cases versus controls. TMM (trimmed mean of M-values): Method used for miRNA expression normalization. Fold change (FC): Ratio of expression between cases and controls (FC > 1 indicates overexpression; FC < 1 indicates underexpression). log₂(FC) is the base-2 logarithmic transformation of FC. Crude (unadjusted) vs. Adjusted (** and ***): The unadjusted model does not control for any confounders; the second model is adjusted for all miRNAs included in the final set; the third model also adjusts for age, menopausal status, body mass index (BMI), family history of breast cancer, smoking status, alcohol intake, use of hormonal contraceptives, history of hormone replacement therapy (HRT), education level, and region of recruitment. *P* values are nominal and not corrected for multiple testing; results should be interpreted as exploratory.

Predictive performance of miRNA-based models

We first assessed the discriminatory performance of miRNA-based panels selected through LASSO regression for the detection of breast cancer and its molecular subtypes. Initial models included only circulating miRNA expression levels (unadjusted models), followed by models incorporating demographic and epidemiological covariates. Performance was evaluated using both LASSO-based AUCs and repeated cross-validated AUCs, with results summarized in Fig. 1 and Supplementary Table 4.

Figure 1 displays ROC curves for models distinguishing (A) all breast cancer cases from controls and stratified by molecular subtypes: (B) Luminal A, (C) Luminal B, (D) HER2-positive, and (E) Basal-like. Dashed lines represent models based solely on circulating miRNAs, while solid lines correspond to models additionally adjusted for age, BMI, menopausal status, family history of breast cancer, smoking, alcohol use, hormonal therapy (contraceptives/HRT), education level, and recruitment region.

In the overall analysis (Fig. 1A), adjustment substantially improved discrimination, increasing the AUC from 0.802 to 0.908. The largest gain was observed in Luminal A (Fig. 1B), with the AUC rising from 0.751 to 0.896. A similar improvement was seen in Luminal B tumors (Fig. 1C; AUC 0.768 vs. 0.894). In HER2-positive tumors (Fig. 1D), adjustment led to a modest improvement in discrimination (AUC 0.914 vs. 0.965). In Basal-like tumors (Fig. 1E), where the unadjusted model already showed excellent performance, covariates still increased the AUC from 0.950 to 0.989.

Supplementary Table 4 further details these comparisons, presenting both LASSO-derived AUCs and mean AUCs from repeated fivefold cross-validation. In the general model, cross-validated AUC increased from 0.778 (95% CI 0.765–0.792) to 0.869 (95% CI 0.849–0.889) after adjustment, supporting the consistency of the observed improvements.

miRNA	TMM cases*	TMM controls*	log ₂ (FC)	Fold change*	Odds ratio (95% CI)**	P**	Odds ratio (95% CI)***	P***
miR-101-3p	1.437	1.965	-0.528	0.694	0.06 (0.02–0.24)	<0.001	0.01 (0.00–0.82)	0.042
miR-139-5p	-3.726	-4.494	0.768	1.703	1.78 (0.85–3.74)	0.125	9.35 (0.88–99.03)	0.063
miR-142-3p	2.951	2.906	0.045	1.032	0.38 (0.12–1.20)	0.099	0.04 (0.00–2.23)	0.116
miR-186-5p	-3.161	-3.501	0.340	1.266	5.71 (1.79–18.24)	0.003	0.60 (0.05–7.70)	0.692
miR-21-3p	-4.462	-4.97	0.508	1.422	1.69 (0.88–3.24)	0.117	4.21 (0.64–27.60)	0.134
miR-29a-3p	0.104	0.283	-0.179	0.883	0.29 (0.09–0.91)	0.034	0.03 (0.00–1.06)	0.054
miR-324-5p	-3.305	-3.397	0.092	1.066	0.45 (0.18–1.11)	0.081	0.27 (0.01–5.15)	0.386
miR-331-3p	-3.658	-3.835	0.177	1.131	5.40 (1.64–17.80)	0.006	19.47 (0.76–498.92)	0.073
miR-423-3p	-0.829	-1.276	0.447	1.363	4.15 (1.10–15.70)	0.036	39.69 (0.42–3788.41)	0.113
miR-146a-5p	0.184	0.474	-0.290	0.818	0.09 (0.02–0.32)	<0.001	0.01 (0.00–0.46)	0.017
miR-21-5p	3.596	3.473	0.123	1.089	50.87 (7.71–335.74)	<0.001	1.3 × 10 ⁶ (20.36–7.8 × 10 ¹⁰)	0.013

Table 5. Circulating miRNA profiles in HER2-positive breast cancer versus controls. TMM (trimmed mean of M-values): Method used for miRNA expression normalization. Fold change (FC): Ratio of expression between cases and controls (FC > 1 indicates overexpression; FC < 1 indicates underexpression). log₂(FC) is the base-2 logarithmic transformation of FC. Crude (unadjusted) vs. Adjusted (** and ***): The unadjusted model does not control for any confounders; the second model is adjusted for all miRNAs included in the final set; the third model also adjusts for age, menopausal status, body mass index (BMI), family history of breast cancer, smoking status, alcohol intake, use of hormonal contraceptives, history of hormone replacement therapy (HRT), education level, and region of recruitment. *P* values are nominal and not corrected for multiple testing; results should be interpreted as exploratory.

miRNA	TMM cases*	TMM controls*	log ₂ (FC)	Fold change*	Odds ratio (95% CI)**	P**	Odds ratio (95% CI)***	P***
miR-101-3p	1.122	1.965	-0.843	0.557	0.03 (0.01–0.09)	<0.001	0.00 (0.00–0.00)	0.002
miR-139-5p	-3.617	-4.494	0.877	1.837	1.90 (0.98–3.69)	0.056	8.08 (0.63–104.36)	0.110
miR-142-3p	2.906	2.906	-0.000	1.000	0.08 (0.02–0.27)	<0.001	0.00 (0.00–0.14)	0.008
miR-21-3p	-4.351	-4.97	0.619	1.536	2.07 (1.10–3.89)	0.024	1.49 (0.21–10.47)	0.686
miR-331-3p	-3.747	-3.835	0.088	1.063	4.30 (1.66–11.14)	0.003	150.39 (4.54–4979.62)	0.005
miR-423-3p	-0.993	-1.276	0.283	1.217	3.58 (1.24–10.38)	0.0019	56.75 (0.88–3660.98)	0.057
miR-146a-5p	0.112	0.474	-0.362	0.778	0.06 (0.02–0.22)	<0.001	0.00 (0.00–0.04)	0.006
miR-21-5p	3.504	3.473	0.031	1.022	55.80 (10.31–302.02)	<0.001	4.6 × 10 (3698.89–5.7 × 10 ¹⁷)	0.003

Table 6. Circulating miRNA profiles in Basal-like breast cancer versus controls. TMM (trimmed mean of M-values): Method used for miRNA expression normalization. Fold change (FC): Ratio of expression between cases and controls (FC > 1 indicates overexpression; FC < 1 indicates underexpression). log₂(FC) is the base-2 logarithmic transformation of FC. Crude (unadjusted) vs. Adjusted (** and ***): The unadjusted model does not control for any confounders; the second model is adjusted for all miRNAs included in the final set; the third model also adjusts for age, menopausal status, body mass index (BMI), family history of breast cancer, smoking status, alcohol intake, use of hormonal contraceptives, history of hormone replacement therapy (HRT), education level, and region of recruitment. *P* values are nominal and not corrected for multiple testing; results should be interpreted as exploratory.

The effect was most pronounced in Luminal A and Luminal B. In Luminal A, the cross-validated AUC rose from 0.719 (95% CI 0.701–0.736) to 0.829 (95% CI 0.797–0.861). In Luminal B, the adjusted AUC reached 0.845 (95% CI 0.820–0.870) compared with 0.749 (95% CI 0.713–0.785) in the unadjusted model. These results reinforce the added predictive value of integrating demographic and lifestyle variables in hormone receptor-positive subtypes, which are typically more heterogeneous.

HER2-positive and Basal-like tumors also benefited from adjustment, although the unadjusted models already showed strong discrimination. The adjusted cross-validated AUC remained within a similar range, showing a slight decrease to 0.911 (95% CI 0.876–0.946) for HER2-positive tumors. Similarly, for the Basal-like tumors, the adjusted cross-validated AUC remained consistent at 0.928 (95% CI 0.911–0.946). Taken together, these findings suggest that the integration of molecular biomarkers with demographic and lifestyle information provides a robust and consistent framework for breast cancer subtype classification and supports high predictive accuracy across different model specifications.

miRNA	Global	Luminal A	Luminal B	HER2	Basal-like
miR-21-5p	↑ 21.18	↑ 9.83	–	↑ 49.66	↑ 64.78
miR-423-3p	↑ 3.81	↑ 4.03	–	↑ 3.92	↑ 5.01
miR-141-3p	↑ 1.41	↑ 1.64	–	–	–
miR-331-3p	↑ 2.19	–	–	↑ 5.06	↑ 5.45
miR-139-5p	↑ 1.66	–	–	–	–
miR-21-3p	–	–	–	–	↑ 2.24
miR-146a-5p	↓ 0.25	↓ 0.34	–	↓ 0.01	↓ 0.00
miR-29a-3p	↓ 0.28	↓ 0.30	–	↓ 0.32	–
miR-101-3p	↓ 0.35	–	↓ 0.47	↓ 0.01	↓ 0.00
miR-142-3p	↓ 0.49	–	–	–	↓ 0.00
miR-186-5p	↓ 0.45	↓ 0.29	↓ 0.30	↑ 5.96	–

Table 7. Subtype-specific adjusted odds ratios (OR) for selected circulating miRNAs in breast cancer. Odds ratios adjusted for age, body mass index (BMI), educational level, socioeconomic status, and smoking. Upwards arrows (↑) indicate overexpression (OR > 1); downwards arrows (↓) indicate underexpression (OR < 1); dashes (–) denote non-significant associations.

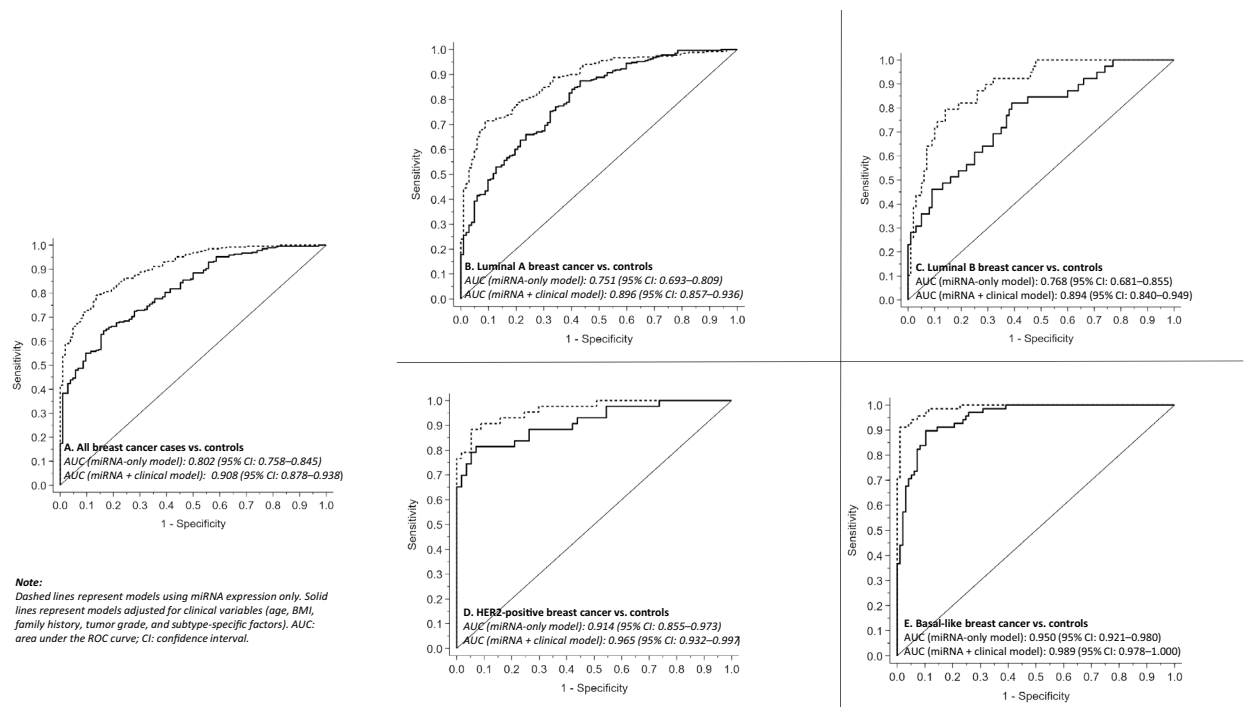


Fig. 1. Receiver operating characteristic (ROC) curves for miRNA-based models with and without clinical adjustment across breast cancer subtypes.

Discussion

In our multicenter study, we found that circulating miRNA panels selected by LASSO achieved strong discriminatory performance for breast cancer detection, particularly in aggressive subtypes. The Basal-like and HER2-positive models reached AUCs of 0.989 and 0.965, respectively, while Luminal A and B subtypes showed more moderate classification performance (AUCs of 0.894–0.896) with improved model performance after adjustment for epidemiological covariates. This highlights the potential contribution of miRNAs as complementary predictive markers in aggressive tumors and as complementary components when combined with patient characteristics in more heterogeneous luminal subtypes.

Our findings are consistent with prior studies reporting comparable predictive performance using circulating miRNA signatures. For instance, Shimomura et al. identified a serum miRNA combination with AUC > 0.85 for early breast cancer detection, and Zhao et al. demonstrated the utility of plasma-based profiles in case-control models. More recently, Gómez-Acebo et al. confirmed in prostate cancer cohorts that LASSO-derived miRNA panels achieve high predictive accuracy (AUC = 0.824), supporting the consistency of this modelling approach across cancer types^{123–125}. Biologically, the higher predictive accuracy observed in Basal-like and

HER2-positive tumors in our data is consistent with the presence of highly active oncogenic programs that drive distinctive circulating miRNA signals; for example, oncomiRs such as miR-21-5p—implicated in epithelial–mesenchymal transition (EMT) and metastatic behavior—tend to be overexpressed in aggressive phenotypes, which may increase the detectability of tumor-derived signals in liquid biopsy^{108,126}. Moreover, evidence that exosomal miR-21-5p attains high diagnostic accuracy¹²⁷ supports the notion that vesicle-packaged miRNAs may concentrate tumor-derived molecules in circulation, contributing to the observed AUCs in aggressive subtypes, while pre-analytical and compartment differences (plasma vs. EVs) likely explain part of the variability reported across studies and subtypes^{37,67,70,128}.

At the individual biomarker level, **miR-21-5p** was the most consistently upregulated in our data across subtypes, with higher expression levels in Basal-like and HER2-positive tumors, while **miR-423-3p** was also consistently upregulated across subtypes. In prior studies, miR-21-5p overexpression has been associated with advanced stage, nodal involvement, and poor outcomes, and with EMT/metastatic behavior in breast cancer stem cells^{108,126}, whereas other works have reported associations with early disease or improved survival^{37,67,128}. Exosomal miR-21-5p has shown high diagnostic performance, but the optimal biological source and pre-analytical conditions remain debated^{70,127}.

For miR-423-3p, the evidence is also mixed: silencing reduced proliferation and connected to PCNA/mutant p53 pathways in cell models¹²⁹; overexpression was reported in ER+/PR+ tumors⁸⁷ and in serum from cases¹⁹, while underexpression was described in HR+/HER2- serum cohorts⁴⁹. Taken together, our findings of overexpression associations for miR-21-5p and miR-423-3p support their involvement in proliferative and EMT-related pathways and with the possibility that extracellular-vesicle packaging enriches tumor-derived miRNAs in circulation, potentially explaining the diagnostic signals noted in aggressive subtypes^{70,108,126,127}.

miR-141-3p in our analyses showed a positive association that was observed only after full adjustment (e.g., an OR of 1.61; 95% CI, 1.04–2.48 in Luminal A), suggesting a potential relationship that was not apparent in crude models. Evidence is mixed: reduced expression and a tumor-suppressive role via HMGB1/HIF-1 α have been reported, with lower levels in aggressive phenotypes and higher levels in HR-positive tumors^{52,130}, whereas other studies found overexpression in high-grade disease and poorer survival, including Basal-like and HER2-amplified contexts^{28,131}. A systematic review emphasizes pleiotropy modulated by subtype and microenvironment⁸⁰. Our adjustment-dependent observation is therefore compatible with a context-sensitive role, potentially hypoxia-responsive, and with the possibility that patient-level factors (age, menopausal status, BMI) confound or influence its circulating association, aligning with the heterogeneity highlighted in these reports.

miR-331-3p showed nominal risk-oriented associations within the HER2-positive and Basal-like subgroups, albeit with very wide confidence intervals that reflect limited precision. Most prior reports indicate downregulation in triple-negative tissues with adverse clinicopathologic features and direct suppression of NRP2, lower levels in serum cohorts, and functional inhibition of migration/invasion in MDA-MB-231 models^{49,91,132}. A potential explanation for this observed discrepancy could involve the combination of compartment and subgroup effects, including tissue versus serum differences, the small size of HER2/Basal-like strata in our data (reflected by wide CIs), and platform variability. These factors can potentially invert the direction observed in liquid biopsy while the intratumoral profile remains suppressive in tissue-based assays, as described by these studies.

For **miR-139-5p**, our data showed a nominal risk association with moderate overexpression in circulation, while multiple tissue-based studies describe downregulation and tumor-suppressive behavior, including in HER2-enriched, luminal B, and Basal-like/triple-negative contexts^{23,53,133–135}. Other works, however, report oncogenic roles depending on regulatory interactions with lncRNAs and signaling pathways^{19,64}. Mechanistically, miR-139-5p interfaces with DNA repair and cell-cycle genes (MAT2A, POLQ, TOP1/2A, XRCC5, NOTCH1, RAD54L) and with lncRNA networks, such as TTN-AS1, and participates in TGF β , Wnt, Rho, and MAPK/PI3K signaling^{30,107,129,136,137}. These pathways may provide a rationale for potentially divergent roles, as depending on pathway activation, subtype composition, and the circulating compartment, miR-139-5p may appear increased in blood while being suppressed in tumor tissue. This interpretation is consistent with observed differences between circulating and tissue-based profiles. In contrast, **miR-101-3p**, **miR-29a-3p**, and **miR-146a-5p** were frequently underexpressed across multiple subtypes, including HER2-positive and Basal-like, yielding nominal inverse associations after adjustment, suggesting a potential role as circulating markers with inverse associations across subtypes. Prior work is largely aligned: miR-101-3p downregulation is linked to poor prognosis, with restoration after therapy and repression of oncogenic programs; mechanistically it targets JAK2, CXCR7, STMN1, AMPK, and EZH2, and can be sequestered by MALAT1 to activate mTOR/PKM2^{19,22,42,55,138–140}. miR-29a-3p is underexpressed in tissue/serum studies with regulation of COL5A1 and AGR2 and associations with better prognosis at higher levels, though EMT-linked invasiveness has been described under specific experimental conditions^{35,77,95,141,142}. For miR-146a-5p, human cohort data are mixed: higher levels have been related to resistance (trastuzumab/NAC) and worse outcomes in HER2+ disease, whereas other studies include it in signatures of lower relapse risk or better survival^{36,46,75,143,144}. Biologically, our observed inverse associations are compatible with the repression of key anti-oncogenic checkpoints: sustained loss of miR-101-3p permits EZH2/NT5E and JAK2/AMPK axis activation; reduced miR-29a-3p de-represses ECM-remodeling nodes (COL5A1/AGR2); and the variability in miR-146a-5p likely reflects differential engagement of treatment-response pathways in HER2-enriched tumors^{22,36,42,46,55,75,95,139,141,142}.

We also observed nominal subtype-dependent patterns, most notably for **miR-186-5p**, which was suppressed in Luminal A/B but elevated in HER2-positive tumors, and for **miR-142-3p**, which showed inverse associations in Basal-like while appearing elevated in some luminal cohorts. The literature mirrors this duality: downregulation of miR-186-5p in luminal tumors and Basal-like tissues/lines is linked to poorer survival and invasion^{113,145,146}, whereas overexpression and FOXM1-related pro-proliferative activity have been described in triple-negative models⁵⁹. For miR-142-3p, underexpression and suppressor roles are described in tissues

and cell lines, targeting the HMG family and EMT axes, though serum studies in luminal disease have shown overexpression versus controls or between luminal subtypes^{19,39,45,100,147,148}. These divergent behaviors may reflect context-specific regulation shaped by subtype biology and compartment (tissue vs. serum vs. EVs), which our stratified framework aims to characterize.

Finally, **miR-142-3p** showed nominal inverse associations in our Basal-like analyses, whereas serum studies in luminal disease have noted higher circulating levels versus controls or between luminal subtypes^{19,148,149}. In contrast, tissue and cell-line work frequently shows underexpression linked to worse prognosis and metastasis, targeting the HMG family members and EMT-related axes, with supporting observations in clinical samples and Luminal A lines^{39,45,100,147,150,151}. Together with our stratified findings, this is compatible with a potential suppressive role for miR-142-3p in Basal-like biology, while emphasizing how subtype and sample type may shape the observed circulating profile.

We observed substantial concordance between nominal differential expression and LASSO-based feature selection: 11 of the 12 miRNAs retained by LASSO showed observable differences in expression between cases and controls. This overlap supports the internal consistency of the pipeline and is compatible with the suggestion that penalized selection may capture potentially relevant associations. While LASSO does not optimize p values by design, we deliberately emphasized changes in odds ratios and their stability after adjustment, which is aligned with the predictive aim of the study. The agreement between LASSO-based performance and repeated cross-validated AUCs further reinforces our results, and is consistent with our prior experience in prostate cancer cohorts showing high discrimination with LASSO-derived serum miRNA panels (AUC = 0.824)¹²³. Together, these findings suggest that the identified panels could reflect the underlying tumor biology while retaining practical predictive utility in a liquid-biopsy setting.

From a methodological standpoint, the close correspondence between LASSO AUCs and the mean AUCs obtained via repeated stratified fivefold cross-validation with bootstrap correction points toward limited optimism in predictive performance estimates. The observation that adjusted models showed improved discrimination, especially in Luminal A and B, suggests that demographic and epidemiological covariates may capture residual confounding or clinically relevant heterogeneity that pure molecular profiles do not fully encode. This “data fusion” approach is aligned with prior serum/plasma studies that reported notable diagnostic performance^{124,125} and with our own experience using penalized models across cancer sites¹²³, compatible with the translatability of the analytic framework.

This study builds upon our previous work on circulating miRNAs and breast cancer detection in the MCC-Spain cohort, where previous analyses focused on overall case-control differences, without stratification by subtype. In the current study, we expanded the analytical scope by incorporating a larger and more diverse set of breast cancer cases, including a higher number of clinically aggressive subtypes such as HER2-positive and Basal-like tumors. Additionally, we conducted an updated literature review specifically focused on miRNAs associated with breast cancer molecular subtypes, providing a biologically driven selection of candidate miRNAs. This approach allowed for the development of exploratory subtype-specific models and the investigation of potential heterogeneity in miRNA expression and diagnostic potential across distinct tumor profiles.

This work has several strengths that enhance the potential utility and clinical relevance of its findings. First, the multicenter recruitment across Spanish hospitals provides geographic and clinical diversity, enhancing generalizability and mitigating single-site biases. Second, the use of penalized regression (LASSO) enables parsimonious, data-driven feature selection in a high-dimensional setting, reducing overfitting relative to stepwise procedures and prioritizing predictive potential rather than nominal p value thresholds. Third, internal validation through repeated fivefold cross-validation with bootstrap resampling (10 repetitions) addresses sampling variability and selection uncertainty, yielding more conservative AUC estimates. Fourth, explicit stratification by molecular subtype accounts for breast cancer heterogeneity and uncovers subtype-specific miRNA patterns with differential predictive performance. Fifth, integrating demographic and epidemiological covariates with circulating miRNAs meaningfully improved discrimination, particularly in Luminal A/B, supporting multimodal risk stratification that reflects both biological and patient-level context. Finally, the convergence of our observed circulating profiles with mechanistic pathways reported in the literature, such as EMT, DNA repair/cell cycle, ECM remodeling, exosomal transport, provides biological plausibility for the predictive panels^{30,107,108,126,127}.

Several limitations should be acknowledged. First, the sample sizes of certain subgroups, particularly Luminal B and HER2-positive tumors, were limited, which reduced the statistical power to detect modest associations and likely contributed to wide confidence intervals (CIs) and occasional extreme odds ratios (ORs) in the HER2/Basal-like analyses. Second, we did not correct for multiple comparisons across the large number of statistical tests performed. Therefore, all reported associations are nominal and exploratory; while this approach maintains sensitivity for discovery, it increases the risk of false positives (Type I errors). Third, some subtype-specific models (notably HER2 and Basal-like) relied on unadjusted or minimally adjusted ORs due to model instability, and adjustment for recruitment site may have further increased variance, as reflected in the wide CIs. Fourth, although internal validation such as cross-validation or bootstrapping was conducted, external validation in independent cohorts is required to assess the generalizability of the predictive panels. Fifth, pre-analytical variability and biological compartment differences, including plasma versus extracellular vesicles, may influence measured miRNA levels, as previously highlighted in translational studies of miR-21-5p⁷⁰, and could explain some discrepancies across studies despite the promising diagnostic performance of exosomal assays¹²⁷. Sixth, the single-country recruitment may limit generalizability to populations with different genetic backgrounds or environmental exposures. It is also worth noting that although penalized regression techniques such as LASSO can be affected by multicollinearity, particularly in datasets with highly correlated biomarkers like miRNAs, the overlap between selected features and those identified as differentially expressed suggests that the model captured biologically relevant signals. While this internal consistency does not replace external validation, it

supports the robustness of the selected predictors. Finally, although NGS-based screening guided the initial selection of candidate miRNAs, the limited sample size and underrepresentation of key molecular subtypes, especially HER2-positive and Basal-like, precluded robust subtype-specific differential expression analysis. Future research should expand NGS profiling across all major breast cancer subtypes to enable subtype-specific biomarker discovery.

Clinically, our findings indicate a pragmatic path toward precision screening and subtype-informed triage using liquid biopsy. The strong discriminatory performance observed in Basal-like and HER2-positive tumors suggests that circulating miRNAs may facilitate early detection or expedite diagnostic workups in aggressive disease, while the improved performance observed in Luminal A/B after integrating demographic and epidemiological covariates supports multimodal models in more heterogeneous subtypes. In practice, standardized pre-analytical procedures, harmonized assay platforms, and direct comparisons between plasma- and EV-based detection methods (as discussed for miR-21-5p) will be essential to ensure reproducibility and scalability^{70,127}. Prospective cohort studies with serial sampling should assess temporal stability and prognostic value (monitoring, minimal residual disease, relapse), and external validation efforts should evaluate generalizability across different healthcare settings.

In conclusion, this study suggests that several circulating miRNAs may be consistently dysregulated across breast cancer subtypes and may serve as non-invasive biomarkers for early detection and molecular classification. In line with our results, miR-21-5p and miR-423-3p emerged as reproducible oncomiRs in circulation, while miR-101-3p, miR-29a-3p, and miR-146a-5p exhibited consistent tumor-suppressive patterns across models. Predictive performance was highest in Basal-like and HER2-positive tumors even with miRNA-only models, and improved further in all settings when demographic and epidemiological covariates were incorporated, particularly in Luminal A and B tumors, underscoring the value of integrated clinical and molecular approaches. These findings support further research toward standardized and multimodal liquid-biopsy strategies that combine circulating miRNA panels with patient-level information to enhance early detection, molecular subtyping, and equitable access to precision diagnostics.

Data availability

The raw small RNA sequencing data generated during the screening phase have been deposited in the Gene Expression Omnibus (GEO) under accession number GSE319901. The RT-qPCR validation dataset is available in GEO under accession number GSE320031.

Received: 19 October 2025; Accepted: 23 February 2026

Published online: 07 March 2026

References

- Filho, A. M. et al. The GLOBOCAN 2022 cancer estimates: Data sources, methods, and a snapshot of the cancer burden worldwide. *Int. J. Cancer* **156**, 1336–1346 (2025).
- Liao, L. Inequality in breast cancer: Global statistics from 2022 to 2050. *Breast Edinb. Scotl.* **79**, 103851 (2025).
- Red Española de Registros de Cáncer (REDECAN). *Estimaciones de La Incidencia Del Cáncer En España, 2025*. <https://redcan.org/storage/documentation/442e1d1a-4040-4674-81cf-5e6a67af6458.pdf> (2025).
- Xiong, X. et al. Breast cancer: Pathogenesis and treatments. *Signal Transduct. Target. Ther.* **10**, 49 (2025).
- Kumar, S. & Ranga, A. Role of miRNAs in breast cancer development and progression: Current research. *BioFactors Oxf. Engl.* **51**, e2146 (2025).
- Waks, A. G. & Winer, E. P. Breast cancer treatment: A review. *JAMA* **321**, 288–300 (2019).
- Howlander, N. et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J. Natl. Cancer Inst.* **106**, dju055 (2014).
- Qi, Y.-J. et al. Radiomics in breast cancer: Current advances and future directions. *Cell Rep. Med.* **5**, 101719 (2024).
- Spear, G., Lee, K., DePersia, A., Lienhoop, T. & Saha, P. Updates in breast cancer screening and diagnosis. *Curr. Treat. Options Oncol.* **25**, 1451–1460 (2024).
- Tsarouchi, M. I., Hoxhaj, A. & Mann, R. M. New approaches and recommendations for risk-adapted breast cancer screening. *J. Magn. Reson. Imaging JMRI* **58**, 987–1010 (2023).
- Azani, A. et al. MicroRNAs as biomarkers for early diagnosis, targeting and prognosis of prostate cancer. *Pathology* **248**, 154618 (2023).
- Hanahan, D. Hallmarks of cancer: New dimensions. *Cancer Discov.* **12**, 31–46 (2022).
- Lee, W. X. et al. MicroRNAs and their role in breast cancer metabolism (Review). *Int. J. Oncol.* **66**, 7 (2025).
- Khan, M. M., Sharma, V. & Serajuddin, M. Emerging role of miRNA in prostate cancer: A future era of diagnostic and therapeutics. *Gene* **888**, 147761 (2023).
- Menon, A., Abd-Aziz, N., Khalid, K., Poh, C. L. & Naidu, R. miRNA: A promising therapeutic target in cancer. *Int. J. Mol. Sci.* **23**, 11502 (2022).
- Siniscalco, D., Galderisi, U., Peluso, G. & Finicelli, M. Circulating microRNAs in cancer: A 5-Year update with a focus on breast and lung cancers. *Int. J. Mol. Sci.* **25**, 3140 (2024).
- Nguyen, T. H. N. et al. Panels of circulating microRNAs as potential diagnostic biomarkers for breast cancer: a systematic review and meta-analysis. *Breast Cancer Res. Treat.* **196**, 1–15 (2022).
- van Schooneveld, E. et al. Dysregulation of microRNAs in breast cancer and their potential role as prognostic and predictive biomarkers in patient management. *Breast Cancer Res. BCR* **17**, 21 (2015).
- Gómez-Acebo, I. et al. Circulating miRNAs signature on breast cancer: the MCC-Spain project. *Eur. J. Med. Res.* **28**, 480 (2023).
- Davey, M. G., Davies, M., Lowery, A. J., Miller, N. & Kerin, M. J. The role of MicroRNA as clinical biomarkers for breast cancer surgery and treatment. *Int. J. Mol. Sci.* **22**, 8290 (2021).
- Garrido-Palacios, A. et al. MicroRNA dysregulation in early breast cancer diagnosis: A systematic review and meta-analysis. *Int. J. Mol. Sci.* **24**, 8270 (2023).
- Liu, P. et al. miR-101-3p is a key regulator of tumor metabolism in triple negative breast cancer targeting AMPK. *Oncotarget* **7**, 35188–35198 (2016).
- Hong, H.-C. et al. A panel of eight microRNAs is a good predictive parameter for triple-negative breast cancer relapse. *Theranostics* **10**, 8771–8789 (2020).

24. Coates, A. S. et al. Tailoring therapies—improving the management of early breast cancer: St Gallen international expert consensus on the primary therapy of early breast cancer 2015. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **26**, 1533–1546 (2015).
25. Alonso-Molero, J. et al. Cohort profile: The MCC-Spain follow-up on colorectal, breast and prostate cancers: Study design and initial results. *BMJ Open* **9**, e031904 (2019).
26. Castaño-Vinyals, G. et al. Population-based multicase-control study in common tumors in Spain (MCC-Spain): Rationale and study design. *Gac. Sanit.* **29**, 308–315 (2015).
27. Ahmadpour, F. et al. Methylation-mediated silencing of miR-125a-5p facilitates breast cancer progression by inducing autophagy. *Mol. Biol. Rep.* **49**, 6325–6339 (2022).
28. Almohaywi, M. et al. Deregulated miRNA expression in triple-negative breast cancer of ancestral genomic-characterized Latina patients. *Int. J. Mol. Sci.* **24**, 13046 (2023).
29. Amirfallah, A. et al. Hsa-miR-21-3p associates with breast cancer patient survival and targets genes in tumor suppressive pathways. *PLoS ONE* **16**, e0260327 (2021).
30. Aranza-Martínez, A. et al. Non-coding RNAs associated with radioresistance in triple-negative breast cancer. *Front. Oncol.* **11**, 752270 (2021).
31. Ashirbekov, Y. et al. Combination of circulating miR-145-5p/miR-191-5p as biomarker for breast cancer detection. *PeerJ* **8**, e10494 (2020).
32. Banerjee, S., Kalyani Yabalooru, S. R. & Karunakaran, D. Identification of mRNA and non-coding RNA hubs using network analysis in organ tropism regulated triple negative breast cancer metastasis. *Comput. Biol. Med.* **127**, 104076 (2020).
33. Bhattarai, S. et al. Dysregulated miRNA expression and androgen receptor loss in racially distinct triple-negative breast cancer. *Int. J. Mol. Sci.* **25**, 13679 (2024).
34. Bouz Mkabaah, L. et al. Assessing the role of MicroRNAs in predicting breast cancer recurrence—a systematic review. *Int. J. Mol. Sci.* **24**, 7115 (2023).
35. Bozkurt, S. B., Ozturk, B., Kocak, N. & Unlu, A. Differences of time-dependent microRNA expressions in breast cancer cells. *Non-coding RNA Res.* **6**, 15–22 (2021).
36. Cabello, P. et al. miR-146a-5p promotes angiogenesis and confers trastuzumab resistance in HER2+ breast cancer. *Cancers* **15**, 2138 (2023).
37. Chang, J.T.-H., Wang, F., Chapin, W. & Huang, R. S. Identification of MicroRNAs as breast cancer prognosis markers through the Cancer Genome Atlas. *PLoS ONE* **11**, e0168284 (2016).
38. Chen, C. et al. Clinical significance of let-7a-5p and miR-21-5p in patients with breast cancer. *Ann. Clin. Lab. Sci.* **49**, 302–308 (2019).
39. Chen, Z.-H. et al. PAX5-miR-142 feedback loop promotes breast cancer proliferation by regulating DNMT1 and ZEB1. *Mol. Med. (Camb. Mass.)* **29**, 89 (2023).
40. Cheng, H., Kuang, S., Tan, L. & Sun, S. Circ_0001955 plays a carcinogenic role in breast cancer via positively regulating GLUT1 via decoying miR-1299. *Thorac. Cancer* **13**, 913–924 (2022).
41. Chiorino, G. et al. Plasma microRNA ratios associated with breast cancer detection in a nested case–control study from a mammography screening cohort. *Sci. Rep.* **13**, 12040 (2023).
42. da Silva, F. C. et al. Is the regulation by miRNAs of NTPDase1 and ecto-5'-nucleotidase genes involved with the different profiles of breast cancer subtypes?. *Purinergic Signal* **18**, 123–133 (2022).
43. Dai, W. et al. miR-148b-3p, miR-190b, and miR-429 regulate cell progression and act as potential biomarkers for breast cancer. *J. Breast Cancer* **22**, 219–236 (2019).
44. Das, S. C. et al. Comprehensive bioinformatics and machine learning analyses for breast cancer staging using TCGA dataset. *Brief. Bioinform.* **26**, bbae628 (2024).
45. Dastmalchi, N. et al. The combined restoration of miR-424-5p and miR-142-3p effectively inhibits MCF-7 breast cancer cell line via modulating apoptosis, proliferation, colony formation, cell cycle and autophagy. *Mol. Biol. Rep.* **49**, 8325–8335 (2022).
46. Di Cosimo, S. et al. End-of-neoadjuvant treatment circulating microRNAs and HER2-positive breast cancer patient prognosis: An exploratory analysis from NeoALTTO. *Front. Oncol.* **12**, 1028825 (2023).
47. Dong, L. et al. Urinary extracellular vesicle-derived miR-126-3p predicts lymph node invasion in patients with high-risk prostate cancer. *Med. Oncol. Northwood Lond. Engl.* **41**, 169 (2024).
48. Dong, M. et al. NCAPG upregulation mediated by four microRNAs combined with activation of the p53 signaling pathway is a predictor of poor prognosis in patients with breast cancer. *Oncol. Lett.* **21**, 323 (2021).
49. Escuin, D. et al. Circulating microRNAs in early breast cancer patients and its association with lymph node metastases. *Front. Oncol.* **11**, 627811 (2021).
50. Estupiñan-Jiménez, J. R. et al. MicroRNA-1307-3p contributes to breast cancer progression through PRM2. *Thorac. Cancer* **15**, 2298–2308 (2024).
51. Fang, Y., Zhang, Q., Chen, Z., Guo, C. & Wu, J. Clinical significance and immune characteristics analysis of miR-221-3p and its key target genes related to epithelial-mesenchymal transition in breast cancer. *Aging* **16**, 322–347 (2024).
52. Fontana, A. et al. Combined analysis of miR-200 family and its significance for breast cancer. *Sci. Rep.* **11**, 2980 (2021).
53. Fu, H., Li, K., Wang, S. & Li, Y. High expression of CCNB1 driven by ncRNAs is associated with a poor prognosis and tumor immune infiltration in breast cancer. *Aging* **14**, 6780–6795 (2022).
54. Fu, L. et al. Serum expression levels of microRNA-382-3p, -598-3p, -1246 and -184 in breast cancer patients. *Oncol. Lett.* **12**, 269–274 (2016).
55. Fuso, P. et al. Let-7a-5p, miR-100-5p, miR-101-3p, and miR-199a-3p hyperexpression as potential predictive biomarkers in early breast cancer patients. *J. Pers. Med.* **11**, 816 (2021).
56. Gallo, M. et al. The EGFR signaling modulates in mesenchymal stem cells the expression of miRNAs involved in the interaction with breast cancer cells. *Cancers* **14**, 1851 (2022).
57. Gao, X.-H. et al. MicroRNA-96-5p represses breast cancer proliferation and invasion through Wnt/ β -catenin signaling via targeting CTNND1. *Sci. Rep.* **10**, 44 (2020).
58. García-Magallanes, N. et al. Underexpression of circulating miR-145-5p and miR-133a-3p are associated with breast cancer and immunohistochemical markers. *J. Cancer Res. Ther.* **16**, 1223–1228 (2020).
59. Hamurcu, Z. et al. MicroRNA profiling identifies Forkhead box transcription factor M1 (FOXO1) regulated miR-186 and miR-200b alterations in triple negative breast cancer. *Cell. Signal.* **83**, 109979 (2021).
60. Hao, H. et al. miRNA-186-5p inhibits migration, invasion and proliferation of breast cancer cells by targeting SBEM. *Aging* **15**, 6993–7007 (2023).
61. Hashimoto, K., Inada, M., Yamamoto, Y. & Ochiya, T. Preliminary evaluation of miR-1307-3p in human serum for detection of 13 types of solid cancer using microRNA chip. *Heliyon* **7**, e07919 (2021).
62. Hosseinpour, Z., Rezaei Tavirani, M. & Akbari, M. E. Stage Analysis of Breast Cancer Metabolomics: A System Biology Approach. *Asian Pac. J. Cancer Prev. APJCP* **24**, 1571–1582 (2023).
63. Hu, Z. et al. Identification and validation of screening models for breast cancer with 3 serum miRNAs in an 11,349 samples mixed cohort. *Breast Cancer* **31**, 1046–1058 (2024).
64. Itani, M. M. et al. A signature of four circulating microRNAs as potential biomarkers for diagnosing early-stage breast cancer. *Int. J. Mol. Sci.* **22**, 6121 (2021).

65. Jia, H., Wu, D., Zhang, Z. & Li, S. Regulatory effect of the MAFG-AS1/miR-150-5p/MYB axis on the proliferation and migration of breast cancer cells. *Int. J. Oncol.* **58**, 33–44 (2021).
66. Jing, Y. et al. Diagnostic value of 5 miRNAs combined detection for breast cancer. *Front. Genet.* **15**, 1482927 (2024).
67. Kim, M. et al. Exploring miRNA-target gene profiles associated with drug resistance in patients with breast cancer receiving neoadjuvant chemotherapy. *Oncol. Lett.* **27**, 158 (2024).
68. Komoto, T. T. et al. Exploring the therapeutic potential of trans-chalcone: Modulation of microRNAs linked to breast cancer progression in MCF-7 cells. *Int. J. Mol. Sci.* **24**, 10785 (2023).
69. Kujala, J. et al. Circulating micro-RNAs predict the risk of recurrence in triple-negative breast cancer. *Cells* **13**, 1884 (2024).
70. Lee, S. H. & Brianna. Association of microRNA-21 expression with breast cancer subtypes and its potential as an early biomarker. *Pathol. - Res. Pract.* **254**, 155073 (2024).
71. Li, H. et al. A serum microRNA signature predicts trastuzumab benefit in HER2-positive metastatic breast cancer patients. *Nat. Commun.* **9**, 1614 (2018).
72. Li, S.-Y. et al. Deciphering the TCF19/miR-199a-5p/SP1/LOXL2 pathway: Implications for breast cancer metastasis and epithelial-mesenchymal transition. *Cancer Lett.* **597**, 216995 (2024).
73. Li, Y., Wang, Y., Fan, H., Zhang, Z. & Li, N. miR-125b-5p inhibits breast cancer cell proliferation, migration and invasion by targeting KIAA1522. *Biochem. Biophys. Res. Commun.* **504**, 277–282 (2018).
74. Li, Y. et al. MiR-183-5p promotes proliferation, metastasis and angiogenesis in breast cancer cells through negatively regulating four and a half LIM protein 1. *J. Breast Cancer* **23**, 355–372 (2020).
75. Liang, W. et al. Inhibition of lymphangiogenesis: A protective role of microRNA 146a-5p in breast cancer. *Breast J.* **2024**, 7813083 (2024).
76. Lisencu, L. A. et al. The role of miR-375-3p, miR-210-3p and Let-7e-5p in the pathological response of breast cancer patients to neoadjuvant therapy. *Medicina* **58**, 1494 (2022).
77. Liu, J., Xiao, S., Chen, J., Lou, W. & Chen, X. A comprehensive analysis for expression, diagnosis, and prognosis of m5C regulator in breast cancer and its ncRNA-mRNA regulatory mechanism. *Front. Genet.* **13**, 822721 (2022).
78. Lou, W., Ding, B., Wang, S. & Fu, P. Overexpression of GPX3, a potential biomarker for diagnosis and prognosis of breast cancer, inhibits progression of breast cancer cells in vitro. *Cancer Cell Int.* **20**, 378 (2020).
79. Lu, Z. et al. Expressions of miR-27a-3p mRNA and YAP1 mRNA in breast cancer and the relationship with clinicopathology and survival prognosis. *Sichuan Da Xue Xue Bao Yi Xue Ban* **56**, 521–527 (2025).
80. Luo, Q., Tian, Y., Qu, G., Kun-Huang & Luo, S. Functional mechanism and clinical implications of miR-141 in human cancers. *Cell. Signal.* **95**, 110354 (2022).
81. Ma, J. et al. CircEGFR reduces the sensitivity of pirarubicin and regulates the malignant progression of triple-negative breast cancer via the miR-1299/EGFR axis. *Int. J. Biol. Macromol.* **244**, 125295 (2023).
82. Ma, S. et al. MiR-26b-5p inhibits cell proliferation and EMT by targeting MYCBP in triple-negative breast cancer. *Cell. Mol. Biol. Lett.* **26**, 52 (2021).
83. Mardi, A., Ghovahi, A., Abbasvandi, F. & Amani, D. Experimental validation of miR-4443, miR-572, and miR-150-5p in serum and tissue of breast cancer patients as a potential diagnostic biomarker: A study based on bioinformatics prediction. *Biochem. Genet.* <https://doi.org/10.1007/s10528-025-11057-8> (2025).
84. Maryam, M., Naemi, M. & Hasani, S. S. A comprehensive review on oncogenic miRNAs in breast cancer. *J. Genet.* **100**, 15 (2021).
85. Mendes, D. C. C. et al. Could be FOXO3a, miR-96-5p and miR-182-5p useful for Brazilian women with luminal A and triple negative breast cancers prognosis and target therapy?. *Clinics* **78**, 100155 (2023).
86. Minini, M. et al. miR-125a-5p impairs the metastatic potential in breast cancer via IP6K1 targeting. *Cancer Lett.* **520**, 48–56 (2021).
87. Murria, R. et al. Immunohistochemical, genetic and epigenetic profiles of hereditary and triple negative breast cancers. Relevance in personalized medicine. *Am. J. Cancer Res.* **5**, 2330–2343 (2015).
88. Naser Al Deen, N. et al. Over-expression of miR-183-5p or miR-492 triggers invasion and proliferation and loss of polarity in non-neoplastic breast epithelium. *Sci. Rep.* **12**, 21974 (2022).
89. Nashtahosseini, Z., Aghamaali, M. R., Sadeghi, F., Heydari, N. & Parsian, H. Circulating status of microRNAs 660-5p and 210-3p in breast cancer patients. *J. Gene Med.* **23**, e3320 (2021).
90. Pei, J. et al. Integrated analysis reveals FLI1 regulates the tumor immune microenvironment via its cell-type-specific expression and transcriptional regulation of distinct target genes of immune cells in breast cancer. *BMC Genomics* **25**, 250 (2024).
91. Perez-Moreno, E. et al. Suppression of breast cancer metastatic behavior by microRNAs targeting EMT transcription factors. A relevant participation of miR-196a-5p and miR-22-3p in ZEB1 expression. *Breast Cancer Res. Treat.* **212**, 277–290 (2025).
92. Piasecka, D., Braun, M., Kordek, R., Sadej, R. & Romanska, H. MicroRNAs in regulation of triple-negative breast cancer progression. *J. Cancer Res. Clin. Oncol.* **144**, 1401–1411 (2018).
93. Pizzamiglio, S. et al. Circulating microRNAs and therapy-associated cardiac events in HER2-positive breast cancer patients: An exploratory analysis from NeoALTTO. *Breast Cancer Res. Treat.* **206**, 285–294 (2024).
94. Rajarajan, D., Kaur, B., Penta, D., Natesh, J. & Meeran, S. M. MiR-145-5p as a predictive biomarker for breast cancer stemness by computational clinical investigation. *Comput. Biol. Med.* **135**, 104601 (2021).
95. Razi, S., Mozdarani, H. & Behzadi Andouhjerdi, R. Evaluation of the potential diagnostic role of the Lnc-MIAT, miR-29a-3p, and FOXO3a ceRNA networks as noninvasive circulatory bioindicator in ductal carcinoma breast cancer. *Breast Cancer: Basic Clin. Res.* **17**, 11782234231184378 (2023).
96. Salum, G. M. et al. Evaluation of tumorigenesis-related miRNAs in breast cancer in Egyptian women: A retrospective, exploratory analysis. *Sci. Rep.* **14**, 29757 (2024).
97. Sánchez-Sánchez, R. et al. MicroRNA-4732-3p is dysregulated in breast cancer patients with cardiotoxicity, and its therapeutic delivery protects the heart from doxorubicin-induced oxidative stress in rats. *Antioxidants* **11**, 1955 (2022).
98. Santana, MdeFM. et al. Increased expression of miR-223-3p and miR-375-3p and anti-inflammatory activity in HDL of newly diagnosed women in advanced stages of breast cancer. *Int. J. Mol. Sci.* **24**, 12762 (2023).
99. Shao, X. et al. Hsa-miR-221-3p promotes proliferation and migration in HER2-positive breast cancer cells by targeting LASS2 and MBD2. *Histol. Histopathol.* **37**, 1099–1112 (2022).
100. Sharma, P. et al. HMGB3 inhibition by miR-142-3p/sh-RNA modulates autophagy and induces apoptosis via ROS accumulation and mitochondrial dysfunction and reduces the tumorigenic potential of human breast cancer cells. *Life Sci.* **304**, 120727 (2022).
101. Shen, B.-J., Yang, Y.-F. & Zhang, X.-X. Hsa_circ_0001925 promotes malignant progression in triple-negative breast cancer via miR-1299/YY1 axis. *Thorac. Cancer* **14**, 746–757 (2023).
102. Song, H. et al. Novel exosomal circEGFR facilitates triple negative breast cancer autophagy via promoting TFEB nuclear trafficking and modulating miR-224-5p/ATG13/ULK1 feedback loop. *Oncogene* **43**, 821–836 (2024).
103. Song, Y. et al. EPAS1 targeting by miR-152-3p in Paclitaxel-resistant breast cancer. *J. Cancer* **11**, 5822–5830 (2020).
104. Spychalski, G. B. et al. miRNA panel from HER2+ and CD24+ plasma extracellular vesicle subpopulations as biomarkers of early-stage breast cancer. *Breast Cancer Res.* **27**, 90 (2025).
105. Sumer, O. E. et al. Selective arm-usage of pre-miR-1307 dysregulates angiogenesis and affects breast cancer aggressiveness. *BMC Biol.* **23**, 25 (2025).
106. Sun, D., Luo, X., Ma, L., Wang, Y. & Zhang, F. Identifying of miR-98-5p/IGF1 axis contributes breast cancer progression using comprehensive bioinformatic analyses methods and experiments validation. *Life Sci.* **261**, 118435 (2020).

107. Sun, H. et al. miR-139-5p was identified as biomarker of different molecular subtypes of breast carcinoma. *Front. Oncol.* **12**, 857714 (2022).
108. Syed, R. U. et al. MicroRNA-21 (miR-21) in breast cancer: From apoptosis dysregulation to therapeutic opportunities. *Pathology* **262**, 155572 (2024).
109. Tiberio, P. et al. Unlocking the potential of circulating miRNAs in the breast cancer neoadjuvant setting: A systematic review and meta-analysis. *Cancers* **15**, 3424 (2023).
110. Uhr, K. et al. MicroRNAs as possible indicators of drug sensitivity in breast cancer cell lines. *PLoS ONE* **14**, e0216400 (2019).
111. Wang, S. et al. Identification of three circulating MicroRNAs in plasma as clinical biomarkers for breast cancer detection. *J. Clin. Med.* **12**, 322 (2022).
112. Wang, T. et al. Bioinformatics analysis reveals the related role of miR-511-5p in the progression of breast cancer. *J. Healthc. Eng.* **2022**, 7146338 (2022).
113. Wang, Z. et al. miR-186-ANXA9 signaling inhibits tumorigenesis in breast cancer. *Front. Oncol.* **13**, 1166666 (2023).
114. Wu, X., Chen, H., Wu, M., Peng, S. & Zhang, L. Downregulation of miR-182-5p inhibits the proliferation and invasion of triple-negative breast cancer cells through regulating TLR4/NF- κ B pathway activity by targeting FBXW7. *Ann. Transl. Med.* **8**, 995 (2020).
115. Wu, X., Ding, M. & Lin, J. Three-microRNA expression signature predicts survival in triple-negative breast cancer. *Oncol. Lett.* **19**, 301–308 (2020).
116. Yang, M. et al. MiR-93-5p regulates tumorigenesis and tumor immunity by targeting PD-L1/CCND1 in breast cancer. *Ann. Transl. Med.* **10**, 203 (2022).
117. Yu, J. et al. Hsa_circ_0007823 overexpression suppresses the progression of triple-negative breast cancer via regulating miR-182-5p-FOXO1 axis. *Breast Cancer Dove Med. Press* **15**, 695–708 (2023).
118. Yu, X. et al. Promotion effect of FOXCUT as a microRNA sponge for miR-24-3p on progression in triple-negative breast cancer through the p38 MAPK signaling pathway. *Chin. Med. J. (Engl.)* **137**, 105–114 (2024).
119. Zayakin, P. et al. Extracellular vesicles—a source of RNA biomarkers for the detection of breast cancer in liquid biopsies. *Cancers* **15**, 4329 (2023).
120. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
121. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).
122. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* **58**, 267–288 (1996).
123. Gómez-Acebo, I. et al. Role of circulating MicroRNAs in prostate cancer diagnosis and risk stratification in the MCC Spain study. *Sci. Rep.* **15**, 17517 (2025).
124. Shimomura, A. et al. Novel combination of serum microRNA for detecting breast cancer in the early stage. *Cancer Sci.* **107**, 326–334 (2016).
125. Zhao, H. et al. A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer. *PLoS ONE* **5**, e13735 (2010).
126. Babadag, S., Altundag-Erdogan, Ö., Akkaya-Ulum, Y. Z. & Çelebi-Saltik, B. The role of telocytes and miR-21-5p in tumorigenicity and metastasis of breast cancer stem cells. *Mol. Biol. Rep.* **51**, 395 (2024).
127. Liu, M. et al. Exosomal hsa-miR-21-5p is a biomarker for breast cancer diagnosis. *PeerJ* **9**, e12147 (2021).
128. Yu, X. et al. Identification and validation of circulating microRNA signatures for breast cancer early detection based on large scale tissue-derived data. *J. Breast Cancer* **21**, 363–370 (2018).
129. Zhao, H. et al. Genetic analysis and preliminary function study of miR-423 in breast cancer. *Tumor Biol* **36**, 4763–4771 (2015).
130. Sun, S., Ma, J., Xie, P., Wu, Z. & Tian, X. Hypoxia-responsive miR-141-3p is involved in the progression of breast cancer via mediating the HMGB1/HIF-1 α signaling pathway. *J. Gene Med.* **22**, e3230 (2020).
131. Taha, M., Mitwally, N., Soliman, A. S. & Yousef, E. Potential diagnostic and prognostic utility of miR-141, miR-181b1, and miR-23b in breast cancer. *Int. J. Mol. Sci.* **21**, 8589 (2020).
132. Zhao, M. et al. miR-331-3p suppresses cell proliferation in TNBC cells by downregulating NRP2. *Technol. Cancer Res. Treat.* **19**, 1533033820905824 (2020).
133. Haakensen, V. D. et al. Subtype-specific micro-RNA expression signatures in breast cancer progression. *Int. J. Cancer* **139**, 1117–1128 (2016).
134. Turkoglu, F., Calisir, A. & Ozturk, B. Clinical importance of serum miRNA levels in breast cancer patients. *Discov. Oncol.* **15**, 19 (2024).
135. Wu, T. et al. Sevoflurane suppresses the migration, invasion, and epithelial-mesenchymal transition of breast cancer cells through the miR-139-5p/ARF6 axis. *J. Surg. Res.* **258**, 314–323 (2021).
136. Fang, J. et al. LncRNA TTN-AS1 facilitates proliferation, invasion, and epithelial-mesenchymal transition of breast cancer cells by regulating miR-139-5p/ZEB1 axis. *J. Cell. Biochem.* **121**, 4772–4784 (2020).
137. Pajic, M. et al. miR-139-5p modulates radiotherapy resistance in breast cancer by repressing multiple gene networks of DNA repair and ROS defense. *Cancer Res.* **78**, 501–515 (2018).
138. Harati, R., Mohammad, M. G., Tlili, A., El-Awady, R. A. & Hamoudi, R. Loss of miR-101-3p promotes transmigration of metastatic breast cancer cells through the brain endothelium by inducing COX-2/MMP1 signaling. *Pharmaceuticals* **13**, 144 (2020).
139. Jiang, H. et al. MiR-101-3p and Syn-Cal14.1a synergy in suppressing EZH2-induced progression of breast cancer. *OncoTargets Ther.* **13**, 9599–9609 (2020).
140. Shao, J., Zhang, Q., Wang, P. & Wang, Z. LncRNA MALAT1 promotes breast cancer progression by sponging miR101-3p to mediate mTOR/PKM2 signal transmission. *Am. J. Transl. Res.* **13**, 10262–10275 (2021).
141. Wang, J., Huang, K., Shi, L., Zhang, Q. & Zhang, S. CircPVT1 promoted the progression of breast cancer by regulating MiR-29a-3p-mediated AGR2-HIF-1 α pathway. *Cancer Manag. Res.* **12**, 11477–11490 (2020).
142. Zhao, B., Song, X. & Guan, H. CircACAP2 promotes breast cancer proliferation and metastasis by targeting miR-29a/b-3p-COL5A1 axis. *Life Sci.* **244**, 117179 (2020).
143. Tashkandi, H., Shah, N., Patel, Y. & Chen, H. Identification of new miRNA biomarkers associated with HER2-positive breast cancers. *Oncoscience* **2**, 924–929 (2015).
144. Wei, G. et al. Expression of miR-146a-5p in breast cancer and its role in proliferation of breast cancer cells. *Oncol. Lett.* <https://doi.org/10.3892/ol.2018.8589> (2018).
145. Jing, L., Lan, L., Mingxin, Z. & Zhaofeng, Z. METTL3/LINC00662/miR-186-5p feedback loop regulates docetaxel resistance in triple negative breast cancer. *Sci. Rep.* **12**, 16715 (2022).
146. Wang, X. et al. Up-regulated circBACH2 contributes to cell proliferation, invasion, and migration of triple-negative breast cancer. *Cell Death Dis.* **12**, 412 (2021).
147. Mansoori, B. et al. miR-142-3p is a tumor suppressor that inhibits estrogen receptor expression in ER-positive breast cancer. *J. Cell. Physiol.* **234**, 16043–16053 (2019).
148. Zhang, K. et al. A two-miRNA signature of upregulated miR-185-5p and miR-362-5p as a blood biomarker for breast cancer. *Pathol. Res. Pract.* **222**, 153458 (2021).

149. Jusoh, A. R. et al. Plasma circulating Mirnas profiling for identification of potential breast cancer early detection biomarkers. *Asian Pac. J. Cancer Prev.* **22**, 1375–1381 (2021).
150. Mansoori, B. et al. MiR-142-3p targets HMGA2 and suppresses breast cancer malignancy. *Life Sci.* **276**, 119431 (2021).
151. Mansoori, B. et al. miR-142-3p as tumor suppressor miRNA in the regulation of tumorigenicity, invasion and migration of human breast cancer by targeting Bach-1 expression. *J. Cell. Physiol.* **234**, 9816–9825 (2019).

Acknowledgements

Some samples included in this study were provided by the Basque Biobank www.biobancovasco.org and the Biobanco La Fe (B.0000723) and they were processed following standard operation procedures with the appropriate approval of the Ethical and Scientific Committees.

Author contributions

IGA and JL were responsible for data acquisition and curation. IGA and JL secured funding for the study. IGA and JL supervised the research. IGA and SVD drafted the first version of the manuscript. All authors contributed to the study design and conception, reviewed previous versions of the manuscript, and approved the final version.

Funding

This study was partially funded by the “Accion Transversal del Cancer,” approved on the Spanish Ministry Council on the 11 October 2007, Instituto de Salud Carlos III-FEDER (PI08/1770, PI08/0533, PI08/1359, PS09/00773, PS09/01286, PS09/01903, PS09/02078, PS09/01662, PI11/01889, PI11/02213, PI12/00488, PI12/01270, PI12/00715, PI14/01219, PI14/0613, PI17/01388 and PI18/00171), Fundación Marqués de Valdecilla (API 10/09), the ICGC International Cancer Genome Consortium CLL [The ICGC CLL-Genome Project was funded by Spanish Ministerio de Economía y Competitividad (MINECO) through the Instituto de Salud Carlos III (ISCIII) and Red Temática de Investigación del Cáncer (RTICC) del ISCIII (RD12/0036/0036)], the Junta de Castilla y León (LE22A10-2), the Consejería de Salud of the Junta de Andalucía (PI-0571–2009, PI-0306–2011, and salud201200057018tra), the Conselleria de Sanitat of the Generalitat Valenciana (AP_061/10), the Recercaixa (2010ACUP 00310), the Regional Government of the Basque Country, the Consejería de Sanidad de la Región de Murcia, by the European Commission grants FOOD-CT-2006–036224-HIWATE, the Spanish Association Against Cancer (AECC) Scientific Foundation, by the Catalan Government—Agency for Management of University and Research Grants (AGAUR) grants 2017SGR723 and 2014SGR850, the Fundación Caja de Ahorros de Asturias, and the University of Oviedo. ISGlobal acknowledges support from the Spanish Ministry of Science and Innovation through the “Centro de Excelencia Severo Ochoa 2019–2023” Program (CEX2018-000806-S) and support from the Generalitat de Catalunya through the CERCA Program. AP-C was supported by the MINECO (Ministry of Economy in Spain) Grant no. PRE2019-089038, fellowship.

Declarations

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by CEIC 2018.280 and 2008/3123/I. The patients/participants provided their written informed consent to participate in this study.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-026-41660-7>.

Correspondence and requests for materials should be addressed to I.G.-A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026