# Circulating microRNAs in Early Breast Cancer Patients and Its Association With Lymph Node Metastases

**Daniel Escuin** ( descuin@santpau.cat )
Hospital de la Santa Creu i Sant Pau Institut de Recerca https://orcid.org/0000-0003-3538-0579
**Laura López-Vilaró**
Hospital de la Santa Creu i Sant Pau
**Olga Bell**
Hospital de la Santa Creu i Sant Pau Institut de Recerca
**Josefina Mora**
Hospital de la Santa Creu i Sant Pau
**Antonio Moral**
Hospital de la Santa Creu i Sant Pau
**José Ignacio Pérez-García**
Hospital de la Santa Creu i Sant Pau
**Teresa Ramón y Cajal**
Hospital de la Santa Creu i Sant Pau
**Cristina Arqueros**
Hospital de la Santa Creu i Sant Pau
**Enrique Lerma**
Hospital de la Santa Creu i Sant Pau
**Agustí Barnadas**
Hospital de la Santa Creu i Sant Pau

**Research article**

**Keywords:** Circulating microRNAs, breast cancer, lymph nodes, metastasis, surrogate biomarkers

**Posted Date:** August 21st, 2020

**DOI:** https://doi.org/10.21203/rs.3.rs-62423/v1

**Version of Record:** A version of this preprint was published at Frontiers in Oncology on August 26th, 2021. See the published version at https://doi.org/10.3389/fonc.2021.627811.

# Abstract

**Background**: In recent years, miRNAs have emerged as important regulators of many cellular processes, including the various steps of the metastatic process. In addition, circulating miRNAs appear to be surprisingly stable in peripheral blood making them ideal noninvasive biomarkers for disease diagnosis. Here, we investigated the expression profile of circulating miRNAs and their association with the metastatic lymph node status in early breast cancer patients.

**Methods**: We designed a proof-of-principle study using 16 plasma samples from patients with known sentinel lymph node status (n=12 positive and n=4 negative). We performed RNA-sequencing and validated the results by qPCR. Gene Ontology term enrichment and KEGG pathway analyses were carried out using DAVID tools.

**Results**: We found16 differentially expressed miRNAs after adjusting for false discovery correction ($q < 0.01$) in patients with positive samples. Thirteen miRNAs were down-regulated (miR-339-5p, miR-133a-3p, miR-326, miR-331-3p, miR-369-3p, miR-328-3p, miR-26a-3p, miR-139-3p, miR-493-3p, miR-664a-5p, miR-323b-3pmiR-1307-3p and miR-423-3p) and 3 were up-regulated (miR-101-3p, miR-146a-5p and miR-144-3p). Hierarchical clustering using differentially expressed miRNAs clearly distinguished patients according to their lymph node status. We did not find any difference in the miRNA expression profile between plasma samples associated with macrometastasis or micrometastasis. The expression of 9 miRNAs was validated by qPCR. Moreover, gene ontology analysis showed a significant enrichment of biological processes associated with the regulation of the epithelial mesenchymal transition, cell proliferation and transcriptional regulation.

**Conclusions**: Our results indicated the potential role of several circulating miRNAs as surrogate markers of lymph node metastases in early breast cancer patients. Further validation in a larger cohort of patients will be necessary to confirm our results.


# Introduction

Breast cancer remains a common disease worldwide and the second cause of cancer death in the US (1). Early diagnosis, improvements in treatment and early onset of therapy are important factors determining the prognosis and management of patients with breast cancer. Various factors such as early age at menarche, late age at first birth, and late age at menopause are related to breast cancer risk. However, lymph node (LN) affection remains the most important prognosis factor in breast cancer (2). There are a number of factors associated with metastases to the LN, including tumor size, presence of lymphovascular invasion, poor histological grade and age (3, 4). Nevertheless, about 13% of patients with all this factors being favorable develop positive LN (4). Moreover, about 20–30% of breast cancer patients with negative LN will have distant metastases and, interestingly, a similar percentage of patients with positive LN will not metastasize (5). It is unclear thus, whether distant metastases arise in a sequential manner from LN metastases or in parallel through the blood stream. In addition, it is also

unclear whether other factors such interactions between the tumor and the stroma favor locoregional metastases (6).

Most women diagnosed with breast cancer are initially treated with surgery to remove a breast tumor and to determine the presence of metastases in the sentinel LNs (SLNs). This is currently the recommended procedure for axillary staging of early breast cancer. The accurate evaluation of patients with involved SLN determines further axillary lymph node dissection (ALND), the golden standard procedure for invasive breast cancer. However, ALND has been questioned in recent years because of inherent morbidity following the procedure without directly contributing to survival in primary breast cancer patients (7–9) and the recognition that not all patients with nodal disease may require extensive axillary surgery (10).

Elucidation of breast cancer's molecular biological features have had a dramatic effect on how patients with breast cancer are diagnosed and treated. However, effective management of breast cancer is still difficult because of the lack of sensitive and specific biomarkers for early detection and for diseases monitoring. Accumulating evidence in the last years has highlighted the potential use of peripheral blood circulating nucleic acids in breast cancer diagnosis, prognosis and for monitoring response to anticancer therapy. Among these, circulating microRNAs (miRNAs) are increasingly recognized as a promising non-invasive biomarker, given the ease with which miRNAs can be isolated and their structural stability under different conditions of sample processing and isolation (11–13).

MicroRNAs (miRNAs) are a small (19–25 nt) non-coding RNA, expressed in a wide variety of organisms and highly conserved across species (14). MiRNAs regulate the expression of target genes by binding to complementary regions of messenger transcripts to repress their translation or regulate their degradation. MiRNAs are now recognized as novel post-transcriptional regulators targeting over 30% of the human genome (15). The overall emerging picture is that of a complex regulation level of gene expression, in which a single miRNA may control hundreds of targets (16). Many cellular pathways are affected by the regulatory function of miRNAs and several human pathologies, including cancers, have been associated with misregulation of the miRNAs (17) and their metastases (18). Numerous studies have identified widespread alterations in the expression of miRNAs related to human neoplasias. In breast cancer, analysis of miRNA expression classified the different breast cancer molecular subtypes and correlated these with various clinicopathological factors (19) and numerous miRNAs have been shown to play a pivotal role in various steps of the metastatic process (20). In addition, few studies have investigated the use of circulating miRNAs to determine the LN status, the occurrence of distant metastases and breast cancer recurrence using qPCR approaches and showed promising results for the use of various miRNAs as a prognostic biomarkers of metastatic disease (21–24).

Herein, we sought to examine the miRNA content in plasma samples from early breast cancer patients with known SLN and axillary LN metastatic status. We designed a proof-of-principle study to profile the expression of miRNAs by RNA-sequencing using preoperative peripheral blood from patients with early breast cancer who were not previously treated. Our results support the hypothesis of the existence of differential miRNA expression profile in the peripheral blood from breast cancer patients associated with

the LN status of their tumors. Our data highlights the potential use of circulating miRNAs as surrogate markers of locoregional metastatases in breast cancer.

# Material And Methods

### Patients

We studied 16 patients with early breast cancer treated with surgery and diagnosed for positive SLNs. All patients had confirmed diagnosis based on histopathology of tumor biopsy. All tumors were invasive ductal carcinomas with or without *in situ* component. In 2 cases, tumors were mixed and show presence of invasive lobular carcinoma component. Intraoperative SLN were evaluated by the OSNA assay (25). None of the patients had prior treatment with surgery, chemotherapy or radiation. All patients were hormone receptor (HR) positive, HER2 negative. We collected clinical parameters: age, menopausal status, personal and familiar disease precedents and clinical follow-up; pathological parameters: tumor stage was be determined according to the AJCC/UICC system (26), histological grade was determined using the Elston-Ellis grading system (27), histology (ductal, lobular, special types), associated ductal or lobular carcinoma *in situ*, presence of vascular and lymphatic invasion, tumor infiltrating lymphocytes, type of invasion (expansive/infiltrating), tumor multifocality, tumor necrosis; proliferation of non-tumoral tissue (ductal hyperplasia, atypical ductal/lobular hyperplasia).

**Blood Processing and Isolation of Plasma**. Human plasma samples were collected prospectively from early breast cancer patients who have not received any previous treatment. Peripheral blood was withdrawn before surgery. Approximately, 10-15ml of peripheral blood was collected for plasma processing in EDTA tubes. Plasma tubes were processed within 2 hours of collection and spun at 1200xg for 10 minutes. Plasma was aliquot in 1.5 ml tubes and stored at -80C until further processing. All plasma samples used in this study were inspected for absence of hemolysis as previously described (28). Briefly, the hemolysis score (HS) was determined by ultraviolet-visible (UV-Vis) absorbance measurements using a NanoDrop® 2000 Spectrophotometer (Thermo Scientific, Barrington, IL, USA). Measurements were performed by applying 2 µl of plasma on the micro-volume pedestal after centrifugation at 1000 × g for 5 min at 4°C and using saline (PBS) as a blank. In addition, monitoring of hemolysis was conducted by qPCR for all samples by comparing the level of a highly expressed miRNA in red blood cells (has-miR-451a) with a miRNA unaffected by hemolysis (has-miR-23a-3p) as previously described (29).

### RNA isolation NGS Library preparation and Next Generation Sequencing

RNA was isolated from 300µl of plasma samples with the miRNeasy serum/plasma advanced kit (Qiagen, Cat No/ID: 217204) according to the manufacturer's instructions. A range of spike-ins was added to the plasma samples prior to RNA isolation. A quality check was performed by qPCR previous to sequencing the samples. Sixteen samples were selected to perform NGS, including 12 positive SLNs (n = 6 macrometastasis and n= 6 micrometastasis) and 4 negative SLNs. A total of 5µl total RNA was used to construct the NGS libraries using the QIAseqmiRNA Library Kit (Qiagen, Cat. No: 331505). Briefly, after ligation of 3' and 5' adapters and Unique Molecular Identifier (UMIs) to miRNAs, complementary DNA

library ready for sequencing was constructed by reverse transcription followed by 22 cycles of PCR amplification and cDNA cleaned up using QMN beads. Next, we perform a library preparation quality check using either Bioanalyzer 2100 (Agilent) or TapeStation 4200 (Agilent). Based on quality of the inserts and the concentration measurements, libraries were pooled in equimolar ratios and quantified using the qPCR ExiSEQ LNA™ Quant kit (Exiqon). The library pools were then sequenced with a NextSeq500 platform (Illumina) using sequence runs of 75nt single-end reads with an average number of 10 million reads/sample. Raw data was demultiplexed and FASTQ files for each sample were generated using the bcl2fastq 2.18.0.12 software (Illumina). FASTQ data were checked using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

**Genome annotation and quantification of miRNAs**.

To perform genome annotation we used the Exiqon/Xplore RNA pipeline. Following sequencing, Cutadapt (1.9.1) (30) was used to trimmed adaptor sequences. A quality check (QC) was performed to ensure Q-scores >30 (>99.9% correct) of our data (31). Reads with correct length were analyzed for the presence of UMIs using Cutadapt (1.9.1) and then collapsed by UMIs into FASTQ files. This approach eliminates library amplification bias and allowed for true identification of the miRNAs. Bowtie2 software (2.2.6) was used for mapping the reads. The mapping criterion for aligning reads to spike-ins, abundant sequences and miRBase_20 was for reads to have perfect match to the reference sequences. To map the genome, one mismatch was allowed in the first 32 bases of the read. Small insertions and deletions (INDELs) were not allowed. The resulting sequences were annotated using the human assembly GRCh37 and the miRBase_20 database. IsomiR analysis was performed individually for each sample based on the occurrence of count variants for each detected miRNA. Read variants were merged onto a single count file with a consistent nomenclature across samples. Only isomiRs present at a level of 5% of total reads for a specific miRNA were retained. Tag per million (TPM) was used as a normalization procedure to correct for differences in sequencing depth and to quantified each RNA species.

**Differential expression analysis**

Differential expression (DE) analyses were performed using the trimmed mean of M-values normalization method (TMM) (32), based on the log-fold and absolute gene-wise changes in expression between samples. The TMM normalization compensates for sample specific effects caused by the variation in library size/sequencing depth between samples and also compensates for under- or over-sampling effects by trimming and scaling factors that minimize log fold changes between samples across the majority of the miRNAs. Differential expression analysis was performed using the EdgeR statistical software package (Bioconductor, http://www.bioconductor.org/). The isomiR analysis was done using Exiqon in-house scripts (exq_ngs_mircount). Predicted miRNAs analysis was performed based on the read count distribution using the exiqon_ngs_mirpred in house script and the secondary structure prediction according to the miRPara classification score (33). Volcano plots were constructed using R programming (34) by plotting the p value (-log10) on the y-axis and the expression fold change between the two experimental groups on the x-axis

### Principal Component Analysis (PCA) and Heat map and unsupervised clustering

Principal component analysis (PCA) was performed using R programming and TMM-normalized quantifications from defined collections of samples as input. The same input was used to generate a heatmap and unsupervised hierarchical clustering by samples and gene expression profile with R scripts (34). We selected the top 50 miRNAs with the largest coefficient of variation (% CV) across all samples to obtain a cluster of samples. The data was normalized to TMM and converted to log2 scale.

### Gene Ontology (GO) Enrichment Analysis

Gene Ontology (GO) analyses (35, 36) were done with R TopGO package with experimentally verified targets of significantly differentially expressed miRNAs as input. Two different statistical tests were used and compared. First, a standard Fisher's test was used to investigate enrichment of terms between groups. Second, we used the Emil method (37) to incorporate the topology of the GO network, to compensate for local dependencies between GO that could mask significant GO terms. We compared the predictions from these two methods to highlight truly relevant GO terms.

### Quantitative real-time RT–PCR analysis

Quantitative real-time RT–PCR analysis was done with an ABI Prism 7500 Sequence Detection System using the miRCURY LNA™ Universal RT cDNA Synthesis Kit (Exiqon). The cDNA was diluted 50X and assayed in 10 µl PCR reactions according to the protocol for the miRCURY LNA™ Universal RT microRNA PCR System (Exiqon A/S); each microRNA was assayed twice by qPCR on the Serum/plasma Focus microRNA PCR panel. A no-template control (NTC) of water was purified with the samples and profiled like the samples. Analysis of the data was performed using the relative miRNA expression according to the comparative Ct (ΔΔCt) method using negative metastatic samples as reference. We used the geNorm (38) or the Normfinder algorithm (39) to select the best combination of two reference genes based on our qPCR data. Data from multiples plates were normalized using UniSp3 spike-in as interplate calibrators.

### Statistics

Differentially expressed miRNAs from RNA-sequencing data were detected by an exact test based on conditional maximum likelihood (CML) included in the R Bioconductor package edgeR (40). *P* values from RNA-sequencing and qPCR were corrected (q-values) for multiple testing using the Benjamini-Hochberg procedure (41). A false discovery rate (FDR) $q < 0.05$ was considered significant. In all group comparisons missing expression values were treated as zero. Differences in total numbers of miRNAs between groups were analyzed by two-sided parametric t tests. For analysis of clinicopathological parameters, quantitative variables between groups were compared using the Student's T-test and qualitative variables were compared using the X2 or Fisher exact tests. A two-sided p-value $\leq 0.05$ was considered significant.

## Results

A total of 25 patients were included in this study. However, only samples from 16 patients passed the pre-RNA-sequencing quality check (QC). The main clinicopathological characteristics of the patients are described in Table 1. A total of 12 (75%) patients had SLN-positive tumors (76%), of which 6 were diagnosed as micrometastasis (38%) and 6 as macrometastasis (38%).

Table 1

**Basic patient and tumor characteristics.** Data indicates the total number of patients included in the RNA-sequencing (NGS) and in the data validation by qPCR.

| **Database** | | **NGS** | **qPCR** |
|---|---|---|---|
| **Variable** | | **N (%)** | **N (%)** |
| Patients | Plasma | 16 (100) | 20 (100) |
| Age, years | Mean + SD | 63 ± 13 | 63 ± 14 |
| | Median (range) | 63 (46–89) | 59 (46–89) |
| Tumor status | T1 | 9 (56) | 11 (55) |
| | T2 | 4 (25) | 6 (30) |
| | T3 | 3 (19) | 3 (15) |
| Node status | Negative | 4 (25) | 6 (30) |
| | Micrometastasis | 6 (38) | 6 (30) |
| | Macrometastasis | 6 (38) | 8 (40) |
| Axillary Lymph Node Status | Negative | 4 (67) | 5 (63) |
| | Positive | 2 (33) | 3 (38) |
| Tumor grade | I | 6 (38) | 8 (40) |
| | II | 9 (56) | 11 (55) |
| | III | 1 (6) | 1 (5) |
| Estrogen receptor | Negative | 0 (0) | 0 (0) |
| | Positive | 16 (100) | 20 (100) |
| Progesterone receptor | Negative | 1 (6) | 1 (5) |
| | Positive | 15 (94) | 19 (95) |
| HER2 status | Negative | 16 (100) | 20 (100) |
| | Positive | 0 (0) | 0 (0) |
| Ki67 | < 20% | 15 (94) | 18 (90) |
| | > 20% | 1 (6) | 2 (10) |
| Surgery | Mastectomy | 3 (19) | 4 (20) |
| | Lumpectomy | 13 (81) | 16 (80) |

| Database | | NGS | qPCR |
|---|---|---|---|
| Lymphovascular invasion | Negative | 14 (88) | 15 (75) |
| | Positive | 2 (13) | 5 (25) |

All samples passed the post-sequencing QC, which confirmed that the average read quality and base quality had a Q-score > 30 (99.9% correct) (31) and the expected read length distribution for miRNAs (supplementary Fig. 1). All samples were sequenced in one excellent run with a median 27.2 million read number. Following sequencing and trimming, all reads containing identical insert sequence and UMI sequence (insert-UMI pair) were collapsed into a single read, which were passed into the analysis pipeline. This allowed for true quantification of the miRNAs by eliminating library amplification bias and a better representation of the RNA molecules in the sample. We obtained an average of 1.8 million collapsed reads for each sample. After mapping, we still found good mapping reads to miRNAs with a very dominant miRNA peak in most of the samples, indicating a good sample/data quality (supplementary table 1 and supplementary Fig. 2). Overall, we obtained an average genome mapping rate of 46.2% (Fig. 1A), which are values well within the range for plasma samples. After mapping and counting to relevant entries in mirbase_20, the number of known miRNAs was calculated using TPM to measure expression. We found comparable numbers of identified miRNAs using either TPM > 1 (182 miRNAs) or TPM > 10 (125 miRNAs) Fig. 1B). We did not identify any sequences identical to those of known miRNAs in miRBase_20 for other organisms. However, we were able to predict 80 miRNAs based on the structural properties of the genome in the indicated locations resembling those of known miRNAs (Supplementary Table 2).

Next we sought to investigate whether the patients were assigned into biological groups based on their miRNA expression. We performed an unsupervised two-way hierarchical clustering of miRNAs and samples using the 50 miRNAs with the largest coefficient of variation based on TMM counts (Fig. 2A). Our results show that samples did not cluster according to the SLN outcome of the patients, suggesting that that other factors might be responsible to inflict more variation on the samples. We obtained similar results using a principal component analysis. Interestingly, we observed that the 2 samples showing the greater variability (p18 and p62) corresponded to those patients whose tumors had a mixed pathological component (IDC and ILC) (Fig. 2B).

Despite the unsupervised analysis did not group our samples according to the metastasis status of the patients, we were able to identify differentially expressed miRNAs between groups based on the SLN outcome of the patients. First, we analyzed our cohort of patients according to their positive (n = 12) or negative (n = 4) SLN metastasis status. We found 73 miRNAs with a significant differential expression ($p < 0.05$). However, only 16 miRNAs remained significant after correcting for multiple testing ($q > 0.05$) (Table 2). Thirteen miRNAs were down-regulated (miR-339-5p, miR-133a-3p, miR-326, miR-331-3p, miR-369-3p, miR-328-3p, miR-26a-3p, miR-139-3p, miR-493-3p, miR-664a-5p, miR-323b-3pmiR-1307-3p and miR-423-3p) and 3 miRNAs were up-regulated (miR-101-3p, miR-146a-5p and miR-144-3p) (Fig. 3A). Next,

we analyzed the data based on SLN metastasis status subgroups. When we compared patients with macrometastasis vs. negative SLNs, we found 42 miRNAs differentially expressed, but only miR-339-5p ($p < 0.0001$, $q = 0.0413$) remained significant after adjustment of the p value (Fig. 3B, Table 2). Similar results were obtained when comparing micrometastasis and negative SLNs, we obtained 66 miRNAs differentially expressed, but only miR-376c-3p ($p = 0.0001$, $q = 0.046$), miR-326 ($p = 0.0003$, $q = 0.049$) and miR-323b-3p ($p = 0.0004$, $q = 0.049$) passed the FDR (Fig. 3C, Table 2). Interestingly, we did not find any circulating miRNA differentially expressed between patients with macrometastasis or micrometastasis SLNs (Fig. 3D, Table 2). In addition, we used the 16 significant differentially expressed miRNAs in patients with positive SLNs to build a heatmap and hierarchical clustering. Our results show that these miRNAs clearly separated patients with negative and positive SLNs (Fig. 3E). Further validation was performed by specific qPCR assays on the differentially expressed miRNAs (supplementary Fig. 3).

Table 2

**Differentially expressed miRNAs**. Data shows the 25 most significantly differentially expressed miRNAs between patients with positive and negative SLNs and other comparisons made between groups. The list includes the average TMM values per group compared, logarithmic fold change (logFC), raw p-values and Benjamini-Hochberg FDR corrected q-values. Abbreviations: Macro, macrometastasis; Micro, micrometastasis, TMM, trimmed mean of M-values

| | | TMM | TMM | | | |
|---|---|---|---|---|---|---|
| **Names** | **Sequence (5' – 3')** | **Positive** | **Negative** | **logFC** | **p value** | **q value** |
| hsa-miR-339-5p | TCCCTGTCCTCCAGGAGCTCACG | 37.4 | 127.6 | -1.8 | < 0.0001 | 0.007 |
| hsa-miR-133a-3p | TTTGGTCCCCTTCAACCAGCTG | 6.1 | 28.4 | -2.0 | < 0.0001 | 0.008 |
| hsa-miR-326 | CCTCTGGGCCCTTCCTCCAG | 10.6 | 50.0 | -2.2 | < 0.0001 | 0.008 |
| hsa-miR-331-3p | GCCCCTGGGCCTATCCTAGAA | 0.9 | 10.3 | -2.8 | 0.0001 | 0.011 |
| hsa-miR-369-3p | AATAATACATGGTTGATCTTT | 7.9 | 26.0 | -1.7 | 0.0005 | 0.031 |
| hsa-miR-328-3p | CTGGCCCTCTCTGCCCTTCCGT | 134.1 | 350.9 | -1.4 | 0.0005 | 0.031 |
| hsa-miR-26a-1-3p | CCTATTCTTGGTTACTTGCACG | 2.1 | 10.9 | -2.5 | 0.0007 | 0.034 |
| hsa-miR-139-3p | TGGAGACGCGGCCCTGTTGGAGT | 30.2 | 79.8 | -1.4 | 0.0008 | 0.034 |
| hsa-miR-493-3p | TGAAGGTCTACTGTGTGCCAGG | 1.9 | 11.5 | -2.1 | 0.0010 | 0.034 |
| hsa-miR-664a-5p | ACTGGCTAGGGAAAATGATTGGAT | 49.7 | 108.1 | -1.1 | 0.0010 | 0.034 |
| hsa-miR-101-3p | TACAGTACTGTGATAACTGAA | 6070.1 | 3215.4 | 0.9 | 0.0011 | 0.034 |
| hsa-miR-146a-5p | TGAGAACTGAATTCCATGGGTT | 2960.5 | 6266.3 | -1.1 | 0.0012 | 0.034 |
| hsa-miR-144-3p | TACAGTATAGATGATGTACT | 511.3 | 295.1 | 0.8 | 0.0013 | 0.035 |
| hsa-miR-323b-3p | CCCAATACACGGTCGACCTCTT | 10.5 | 29.9 | -1.4 | 0.0016 | 0.040 |
| hsa-miR-1307-3p | ACTCGGCGTGGCGTCGGTCGTG | 150.3 | 337.4 | -1.2 | 0.0017 | 0.040 |

| | | TMM | TMM | | | |
|---|---|---|---|---|---|---|
| hsa-miR-423-3p | AGCTCGGTCTGAGGCCCCTCAGT | 353.7 | 649.1 | -0.9 | 0.0023 | 0.050 |
| hsa-miR-376c-3p | AACATAGAGGAAATTCCACGT | 3.8 | 14.0 | -1.8 | 0.0028 | 0.056 |
| hsa-miR-1 | TGGAATGTAAAGAAGTATGTAT | 69.5 | 168.7 | -1.3 | 0.0037 | 0.071 |
| hsa-miR-1908-5p | CGGCGGGGACGGCGATTGGTC | 23.1 | 59.1 | -1.3 | 0.0042 | 0.073 |
| hsa-miR-744-5p | TGCGGGGCTAGGGCTAACAGCA | 138.5 | 298.1 | -1.1 | 0.0042 | 0.073 |
| hsa-miR-584-5p | TTATGGTTTGCCTGGGACTGAG | 419.9 | 835.4 | -1.0 | 0.0048 | 0.078 |
| hsa-miR-6721-5p | TGGGCAGGGGCTTATTGTAGGAG | 2.4 | 9.3 | -1.9 | 0.0055 | 0.083 |
| hsa-miR-432-5p | TCTTGGAGTAGGTCATTGGGTGG | 130.1 | 432.4 | -1.7 | 0.0055 | 0.083 |
| hsa-miR-28-3p | CACTAGATTGTGAGCTCCTGGA | 72.1 | 150.4 | -1.0 | 0.0058 | 0.084 |
| | | **Macro** | **Negative** | | | |
| hsa-miR-339-5p | TCCCTGTCCTCCAGGAGCTCACG | 38.9 | 138.0 | -1.8 | 0.0001 | 0.041 |
| | | **Micro** | **Negative** | | | |
| hsa-miR-376c-3p | AACATAGAGGAAATTCCACGT | 1.5 | 13.7 | -3.1 | 0.0001 | 0.046 |
| hsa-miR-326 | CCTCTGGGCCCTTCCTCCAG | 9.8 | 49.2 | -2.3 | 0.0003 | 0.049 |
| hsa-miR-323b-3p | CCCAATACACGGTCGACCTCTT | 7.1 | 29.5 | -2.0 | 0.0004 | 0.049 |
| | | **Macro** | **Micro** | | | |
| hsa-miR-122-5p | TGGAGTGTGACAATGGTGTTTG | 8948.2 | 76103.3 | -3.1 | 0.0002 | 0.062 |
| hsa-miR-125b-2-3p | TCACAAGTCAGGCTCTTGGGAC | 0.3 | 8.9 | -3.6 | 0.0006 | 0.090 |

Next, we sought to understand how our data is related to biological functions by performing a gene ontology (GO) analysis. Selecting *Homo sapiens* as the background of listed target genes, we obtained the GO term annotations and KEGG pathway analysis through the functional annotation summaries. The

results are summarized in Table 3. The top 50 biological process GO terms ($p < 0.05$) associated with differentially expressed circulating miRNAs in patients with positive SLNs compared to the reference background (negative SLNs samples). are shown in Fig. 4A. Our data shows that differentially expressed miRNAs associated with biological processes (BP) markedly focused on epigenetic gene expression reuglation, the epithelial-mesenchymal transition (EMT), transcription, cell motility and proliferation processes ($p < 0.01$) (Fig. 4B, Table 3, supplementary data). For instance, we found that positive regulation of mesenchymal cell proliferation term (GO: 0002053) was significantly enriched ($p < 0.0028$) as well as positive regulation of the histone H3-H4 methylation term (GO:0051571) ($p < 0.0017$). These two GO terms remained significant even when patients with positive SLNs were sub-classified as having macro- or micrometastasis in their SLNs. As for the cellular component (CC), the target miRNAs were significantly located in vesicle and membrane fractions ($p < 0.01$). Moreover, differentially expressed miRNAs were enriched in molecular function (MF) terms associated with transcription factors, G protein-related coupled peptide receptor activity, receptor regulator activity and microtubule motor activity ($p < 0.01$) (Table 3).

Table 3

**Gene Ontology (GO).** Top ten GO functional annotation for the significant pathways associated with differentially expressed circulating miRNAs in patients with positive SLNs

| GO ID | GO Term | Counts | p value |
|---|---|---|---|
| **BIOLOGICAL PROCESS** | | | |
| GO:0051571 | positive regulation of histone H3-K4 methylation | 9/15 | 0.0017 |
| GO:0042462 | eye photoreceptor cell development | 12/26 | 0.0017 |
| GO:0070555 | response to interleukin-1 | 17/40 | 0.0021 |
| GO:0002053 | positive regulation of mesenchymal cell proliferation | 25/62 | 0.0028 |
| GO:0071320 | cellular response to cAMP | 12/23 | 0.0034 |
| GO:0002407 | dendritic cell chemotaxis | 4/5 | 0.0035 |
| GO:0009629 | response to gravity | 6/8 | 0.0042 |
| GO:0007097 | nuclear migration | 6/9 | 0.006 |
| GO:0051573 | negative regulation of histone H3-K9 methylation | 6/8 | 0.0064 |
| GO:0034446 | substrate adhesion-dependent cell spreading | 8/16 | 0.0067 |
| **CELLULAR COMPONENT** | | | |
| GO:0031091 | platelet alpha granule | 18/44 | 0.0065 |
| GO:0031983 | vesicle lumen | 17/46 | 0.0137 |
| GO:0060205 | cytoplasmic membrane-bounded vesicle lumen | 17/46 | 0.0137 |
| GO:0031093 | platelet alpha granule lumen | 16/41 | 0.0171 |
| GO:0034774 | secretory granule lumen | 16/41 | 0.0171 |
| GO:0044306 | neuron projection terminus | 9/20 | 0.0216 |
| GO:1902495 | transmembrane transporter complex | 14/31 | 0.0218 |
| GO:1990351 | transporter complex | 14/31 | 0.0218 |
| GO:0015030 | Cajal body | 8/16 | 0.0239 |
| GO:0034704 | calcium channel complex | 7/12 | 0.0245 |
| **MOLECULAR FUNCTION** | | | |
| GO:0008528 | G-protein coupled peptide receptor activity | 9/15 | 0.0013 |
| GO:0001618 | virus receptor activity | 6/13 | 0.0018 |
| GO:0030955 | potassium ion binding | 5/6 | 0.0064 |

| GO ID | GO Term | Counts | p value |
|---|---|---|---|
| GO:0005161 | platelet-derived growth factor receptor binding | 17/39 | 0.0078 |
| GO:0008798 | beta-aspartyl-peptidase activity | 4/7 | 0.0097 |
| GO:0030545 | receptor regulator activity | 15/36 | 0.0097 |
| GO:0003777 | microtubule motor activity | 11/20 | 0.0115 |
| GO:0046625 | sphingolipid binding | 5/6 | 0.0128 |
| GO:0008307 | structural constituent of muscle | 6/8 | 0.0131 |
| GO:0017022 | myosin binding | 6/11 | 0.0144 |

# Discussion

In recent years, miRNAs have emerged as important regulators of the various steps of the metastatic process (20). Currently, lymph node affection remains the most important prognosis factor in breast cancer (2) and the presence of metastasis in the SLNs is still currently the recommended procedure for axillary staging of early breast cancer. The accurate evaluation of patients with involved SLN determines further axillary lymph node dissection (ALND), the golden standard procedure for invasive breast cancer. However, ALND has been questioned in recent years because of inherent morbidity following the procedure without directly contributing to survival.

In this study, we sought to gain a better understanding of the role of miRNAs in the metastatic process and whether specific expression patterns of miRNAs could predict SLN metastatic status in patients with early breast cancer. Our ultimate goal was to evaluate the use of specific circulating miRNAs in the peripheral blood of patients with breast cancer as tumor biomarkers. In this study, plasma samples were collected prior to any treatment, thus the results obtained from deep sequencing are expected to reflect the basal miRNA expression prior to any therapeutic intervention in these patients. We have performed a proof-of-principle study in plasma samples from 16 breast cancer patients with known SLN metastasis status. Our results show good a good quality sequencing data with mapping rates to miRNAs in most of the samples and comparable miRNA discovery across samples. Thus we are confident in the accuracy of the reported results.

Our data shows that 16 miRNAs that were significantly expressed in plasma samples from patients with positive SLNs. Overall, we found a general down-regulation of miRNAs, with the exception of miR-101-3p and miR-144-3p that showed a 1.9- and 1.7-fold change up-regulation, respectively. MiR-101-3p has been shown to be dysregulated in several malignancies (42), including breast cancer (42, 43). However, it has been noted discrepancies on the direction of the dysregulation. While some reports indicate up-regulation of miR-101, others indicated the opposite. This due to the fact that mature miR-101-3p originates from two different precursors located at different chromosomes. One precursor may be processed to 1 or 2 miRNAs and thus, the mature and precursor miRNA levels might not correlate, and this therefore will

influence the clinical interpretation (43). The same study looked at putative miR-101-3p target genes were analyzed and the most predominant functions were transcription, metabolism, biosynthesis, proliferation, and transcription factor binding. This result indicated that candidate genes have a definitive impact on the pathogenesis of BC (43).

Of those miRNAs down-regulated, miR-339-5p showed a 3.5-fold inhibition in patients with positive SLN metastasis. The expression of miR-339-5p remained significant when we performed the analysis in the subgroup of macrometastatic SLN samples and we observed a non-significant trend towards significance for the subgroup of micrometastatic SLNs (q = 0.071). Our results agree with previous reports showing that reduced miR-339-5p expression in breast cancer is associated with increased metastasis to lymph nodes (44, 45), high clinical stages and worse clinical outcome (44). A similar association with positive LN has been reported in NSCLC patients (46). In addition, miR-339-5p expression appears to be down-regulated in several human cancers including NSCLC (46), ovarian carcinoma (47), hepatocellular carcinoma (48), gliomas (49), colorectal cancer(50), osteosarcoma (51) and breast cancer (45). Mir-339-5p acts as a tumor suppressor gene and its expression is required to inhibit cell migration and invasion in breast cancer cells (44) in a mechanism that involves at least the B-cell lymphoma 6 (BCL6) protein. The authors showed that forced expression of BCL6 results in increased proliferation, anchorage-independent growth, migration, invasion and survival of breast cancer cell lines, whereas knockdown of BCL6 expression reduced these oncogenic properties of breast cancer cells (52). Interestingly, miR-339-5p has been shown to inhibit migration and invasion by targeting BCL6 in breast cancer (53), ovarian cancer cell lines (47) and in NSCLC (54). In addition, miR-339-5p down-regulation in NSCLC inhibits metastasis of NSCLC by regulating the epithelial-to-mesenchymal(EMT) transitionvia BCL6 (54). A recent report has shown that miR-339-5p regulates EMT through regulation of TGF-β (55) in osteosarcoma (51).

The EMT and the TGF-β pathways are one of the most important mechanisms underlying the metastatic ability of cancer cells (56, 57). We have previously shown the importance of the EMT in breast cancer (58) and here, we show that GO term analysis based on the DE miRNAs showed a significant association with the biological process “positive regulation of mesenchymal cell proliferation” (GO:0002053). Another pathway enriched was the positive regulation of H3 K4 methylation (GO:0051571), a mark that on a genome-wide scale is broadly associated with transcriptional regulation, and "negative regulation of H3K9 methylation" (GO:0051573). H3K9 methylation has been associated with the EMT through interactions of KDM1A (a H3K9 demethylase) with the members of the SNAI1 family of zinc finger transcription factors, including SNAI1 (SNAIL) and SNAI2 (SLUG). The expression of SNAI1 and E-cadherin is a hallmark of carcinoma development and metastasis (59). Our data suggest that that MiR-101 could be involved in the regulation of these pathways, as it has been shown to directly target the histone methyltransferase enhancer of zeste homologue 2 (EZH2), which could promote tumor proliferation and invasion (60)

Circulating miR-133a expression was elevated in plasma samples from early-stage BC patients compared to healthy donors (61, 62) and similar results were reported for circulating miR-1307-3p (63) whereas

downregulation of miR-376c-3p (64) miR-376c-3p have been linked to breast cancer recurrence (24). In addition, one study showed that miR-133a expression was reported to be down-regulated in paired breast cancer tumor and serum samples (65), indicating the tumor origin of miR-133a.We found down-regulation of miR-326 in our plasma samples. MiR-326 has been reported to target B7-H3 in breast cancer, an immunoregulatory protein, is overexpressed in several cancers and is often associated with metastasis and poor prognosis (66). Furthermore, its expression has been shown to inhibit tumorogenesis through direct targeting of Nin one binding protein (NOB1) and the MAPK pathway in glioma cells (49).

# Conclusions

Taken together, our results show a significant number of differentially expressed circulating miRNAs from breast cancer patients with SLN metastasis. More importantly, we show an overall down-regulation of miRNAs that have been reported to be direct targets of proteins that promote metastasis. Our study thus shows the feasibility to analyze circulating miRNAs in peripheral blood samples as potential surrogate biomarkers of LN metastasis in breast cancer patients. A further clinical study in a larger patient cohort is warranted to validate these results. Further functional analysis will be required to unveil the molecular mechanisms related to metastasis steps associated with the dysregulation of the miRNAs described herein.

# Abbreviations

aLN, axillary lymph node; ALND, axillary lymph node dissection; CK19, cytokeratin 19; ER, estrogen receptor; FDR, False discovery rate; HR, Hormone receptor; LN, lymph node; miRNA, microRNAs; NGS, next generation sequencing; OSNA, one-step nucleic amplification ; PR, progesterone receptor; QC, quality check; SLN, sentinel lymph node ; TMM, trimmed mean of M-values ; TPM, transcripts per million

# Declarations

### Ethics approval and consent to participate

This study was conducted according to the Declaration of Helsinki principles, with approval from the Clinical Research Ethics Committee at Institut d'Investigacions Biomèdiques Sant Pau. Written informed consent was obtained from all patients under institutional review board-approved protocols.

### Consent for publication

Not applicable

### Availability of data and materials

All data generated or analysed during this study are included in this published article [and its supplementary information files. The datasets used during the current study are available from the corresponding author on reasonable request.


### Funding

This study was supported by a grant from the Instituto de Salud Carlos III (ISCIII, PI13/00110) and co-financed with funds from the European Regional Development Fund (ERDF) to Daniel Escuin, Centro de Investigación Biomédica en Red Cáncer (CB16/12/00471) to Agustí Barnadas and a grant from the Red Temática de Investigación Cooperativa de Cáncer (RTICC; RD12/0036/0076) to Agustí Barnadas. The study sponsors had no involvement in the experimental work, data analysis or the manuscript preparation


### Author's contributions

DE, LLV and AB were involved in the conceptualization of the project. DE, LLV, OB, JM, JIP, AM, CA, TRyC and EL were involved in resources, investigation and methodology. DE, LLV, EL and AB were involved in analysis and interpretation of data. All authors have contributed to the writing of the paper and have critically reviewed it.


### Acknowledgments

The authors would like to thank the patients and their caregivers for their participation and Michael Gombert (Qiagen genomic services) for his advice and excellent technical assistance.


### Conflict of interests

The authors declare that they have no conflict of interest.

# References


1. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. CA: a cancer journal for clinicians. 2019;69(6):438-51.
2. Shek LL, Godolphin W. Model for breast cancer survival: relative prognostic roles of axillary nodal status, TNM stage, estrogen receptor concentration, and tumor necrosis. Cancer Res. 1988;48(19):5565-9.
3. Wildiers H, Van Calster B, van de Poll-Franse LV, Hendrickx W, Roislien J, Smeets A, et al. Relationship between age and axillary lymph node involvement in women with breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(18):2931-7.
4. Rivadeneira DE, Simmons RM, Christos PJ, Hanna K, Daly JM, Osborne MP. Predictive factors associated with axillary lymph node metastases in T1a and T1b breast carcinomas: analysis in more than 900 patients. Journal of the American College of Surgeons. 2000;191(1):1-6; discussion -8.

5. Weigelt B, Peterse JL, van 't Veer LJ. Breast cancer metastasis: markers and models. Nat Rev Cancer. 2005;5(8):591-602.
6. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. The New England journal of medicine. 2002;347(16):1233-41.
7. Giuliano AE, McCall L, Beitsch P, Whitworth PW, Blumencranz P, Leitch AM, et al. Locoregional recurrence after sentinel lymph node dissection with or without axillary dissection in patients with sentinel lymph node metastases: the American College of Surgeons Oncology Group Z0011 randomized trial. Annals of surgery. 2010;252(3):426-32; discussion 32-3.
8. Jagsi R, Chadha M, Moni J, Ballman K, Laurie F, Buchholz TA, et al. Radiation field design in the ACOSOG Z0011 (Alliance) Trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2014;32(32):3600-6.
9. Giuliano AE, Ballman KV, McCall L, Beitsch PD, Brennan MB, Kelemen PR, et al. Effect of Axillary Dissection vs No Axillary Dissection on 10-Year Overall Survival Among Women With Invasive Breast Cancer and Sentinel Node Metastasis: The ACOSOG Z0011 (Alliance) Randomized Clinical Trial. JAMA : the journal of the American Medical Association. 2017;318(10):918-26.
10. Park KU, Caudle A. Management of the Axilla in the Patient with Breast Cancer. The Surgical clinics of North America. 2018;98(4):747-60.
11. Hamam R, Hamam D, Alsaleh KA, Kassem M, Zaher W, Alfayez M, et al. Circulating microRNAs in breast cancer: novel diagnostic and prognostic biomarkers. Cell death & disease. 2017;8(9):e3045.
12. Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nature reviews Genetics. 2018.
13. Condrat CE, Thompson DC, Barbu MG, Bugnar OL, Boboc A, Cretoiu D, et al. miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. Cells. 2020;9(2).
14. Bartel DP, Chen CZ. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nature reviews Genetics. 2004;5(5):396-400.
15. Croce CM. Causes and consequences of microRNA dysregulation in cancer. Nature reviews Genetics. 2009;10(10):704-14.
16. Shen J, Stass SA, Jiang F. MicroRNAs as potential biomarkers in human solid tumors. Cancer letters. 2013;329(2):125-36.
17. Dai R, Ahmed SA. MicroRNA, a new paradigm for understanding immunoregulation, inflammation, and autoimmune diseases. Transl Res. 2011;157(4):163-79.
18. Nicoloso MS, Spizzo R, Shimizu M, Rossi S, Calin GA. MicroRNAs--the micro steering wheel of tumour metastases. Nat Rev Cancer. 2009;9(4):293-302.
19. Blenkiron C, Goldstein LD, Thorne NP, Spiteri I, Chin SF, Dunning MJ, et al. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biol. 2007;8(10):R214.

20. Valastyan S. Roles of microRNAs and other non-coding RNAs in breast cancer metastasis. Journal of mammary gland biology and neoplasia. 2012;17(1):23-32.
21. Chen W, Cai F, Zhang B, Barekati Z, Zhong XY. The level of circulating miRNA-10b and miRNA-373 in detecting lymph node metastasis of breast cancer: potential biomarkers. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2013;34(1):455-62.
22. Shaker O, Maher M, Nassar Y, Morcos G, Gad Z. Role of microRNAs -29b-2, -155, -197 and -205 as diagnostic biomarkers in serum of breast cancer females. Gene. 2015;560(1):77-82.
23. Si H, Sun X, Chen Y, Cao Y, Chen S, Wang H, et al. Circulating microRNA-92a and microRNA-21 as novel minimally invasive biomarkers for primary breast cancer. Journal of cancer research and clinical oncology. 2013;139(2):223-9.
24. Huo D, Clayton WM, Yoshimatsu TF, Chen J, Olopade OI. Identification of a circulating microRNA signature to distinguish recurrence in breast cancer patients. Oncotarget. 2016;7(34):55231-48.
25. Tsujimoto M, Nakabayashi K, Yoshidome K, Kaneko T, Iwase T, Akiyama F, et al. One-step nucleic acid amplification for intraoperative detection of lymph node metastasis in breast cancer patients. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13(16):4807-16.
26. Webber C, Gospodarowicz M, Sobin LH, Wittekind C, Greene FL, Mason MD, et al. Improving the TNM classification: findings from a 10-year continuous literature review. International journal of cancer Journal international du cancer. 2014;135(2):371-8.
27. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology. 1991;19(5):403-10.
28. Appierto V, Callari M, Cavadini E, Morelli D, Daidone MG, Tiberio P. A lipemia-independent NanoDrop((R))-based score to identify hemolysis in plasma and serum samples. Bioanalysis. 2014;6(9):1215-26.
29. Blondal T, Jensby Nielsen S, Baker A, Andreasen D, Mouritzen P, Wrang Teilum M, et al. Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods. 2013;59(1):S1-6.
30. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011;17(1):3.
31. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010;38(6):1767-71.
32. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.
33. Wu Y, Wei B, Liu H, Li T, Rayner S. MiRPara: a SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. BMC bioinformatics. 2011;12:107.
34. R_Development_Core_Team. R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at http://www.R-project.org/. 2011.

35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9.
36. Gene Ontology C. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015;43(Database issue):D1049-56.
37. Alexa A, Rahnenfuhrer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22(13):1600-7.
38. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004;64(15):5245-50.
39. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3(7):RESEARCH0034.
40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.
41. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.
42. Liu P, Ye F, Xie X, Li X, Tang H, Li S, et al. mir-101-3p is a key regulator of tumor metabolism in triple negative breast cancer targeting AMPK. Oncotarget. 2016;7(23):35188-98.
43. Li CY, Xiong DD, Huang CQ, He RQ, Liang HW, Pan DH, et al. Clinical Value of miR-101-3p and Biological Analysis of its Prospective Targets in Breast Cancer: A Study Based on The Cancer Genome Atlas (TCGA) and Bioinformatics. Medical science monitor : international medical journal of experimental and clinical research. 2017;23:1857-71.
44. Wu ZS, Wu Q, Wang CQ, Wang XN, Wang Y, Zhao JJ, et al. MiR-339-5p inhibits breast cancer cell migration and invasion in vitro and may be a potential biomarker for breast cancer prognosis. BMC Cancer. 2010;10:542.
45. Wang B, Li J, Sun M, Sun L, Zhang X. miRNA expression in breast cancer varies with lymph node metastasis and other clinicopathologic features. IUBMB life. 2014;66(5):371-7.
46. Li Y, Zhao W, Bao P, Li C, Ma XQ, Li Y, et al. miR-339-5p inhibits cell migration and invasion in vitro and may be associated with the tumor-node-metastasis staging and lymph node metastasis of non-small cell lung cancer. Oncology letters. 2014;8(2):719-25.
47. Shan W, Li J, Bai Y, Lu X. miR-339-5p inhibits migration and invasion in ovarian cancer cell lines by targeting NACC1 and BCL6. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(4):5203-11.
48. Wang YL, Chen CM, Wang XM, Wang L. Effects of miR-339-5p on invasion and prognosis of hepatocellular carcinoma. Clinics and research in hepatology and gastroenterology. 2016;40(1):51-6.
49. Zhou J, Xu T, Yan Y, Qin R, Wang H, Zhang X, et al. MicroRNA-326 functions as a tumor suppressor in glioma by targeting the Nin one binding protein (NOB1). PloS one. 2013;8(7):e68469.

50. Zhou C, Liu G, Wang L, Lu Y, Yuan L, Zheng L, et al. MiR-339-5p regulates the growth, colony formation and metastasis of colorectal cancer cells by targeting PRL-1. PloS one. 2013;8(5):e63142.
51. Zhang L, Lu XQ, Zhou XQ, Liu QB, Chen L, Cai F. NEAT1 induces osteosarcoma development by modulating the miR-339-5p/TGF-beta1 pathway. Journal of cellular physiology. 2018.
52. Wu Q, Liu X, Yan H, He YH, Ye S, Cheng XW, et al. B-cell lymphoma 6 protein stimulates oncogenicity of human breast cancer cells. BMC Cancer. 2014;14:418.
53. Yan H, Zhao M, Huang S, Chen P, Wu WY, Huang J, et al. Prolactin Inhibits BCL6 Expression in Breast Cancer Cells through a MicroRNA-339-5p-Dependent Pathway. Journal of breast cancer. 2016;19(1):26-33.
54. Li Y, Zhang X, Yang Z, Li Y, Han B, Chen LA. miR-339-5p inhibits metastasis of non-small cell lung cancer by regulating the epithelial-to-mesenchymal transition. Oncology letters. 2018;15(2):2508-14.
55. Meulmeester E, Ten Dijke P. The dynamic roles of TGF-beta in cancer. The Journal of pathology. 2011;223(2):205-18.
56. Chanda A, Sarkar A, Bonni S. The SUMO System and TGFbeta Signaling Interplay in Regulation of Epithelial-Mesenchymal Transition: Implications for Cancer Progression. Cancers. 2018;10(8).
57. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. Nat Rev Cancer. 2018;18(2):128-34.
58. Montserrat N, Gallardo A, Escuin D, Catasus L, Prat J, Gutierrez-Avigno FJ, et al. Repression of E-cadherin by SNAIL, ZEB1, and TWIST in invasive ductal carcinomas of the breast: a cooperative effort? Human pathology. 2011;42(1):103-10.
59. Ismail T, Lee HK, Kim C, Kwon T, Park TJ, Lee HS. KDM1A microenvironment, its oncogenic potential, and therapeutic significance. Epigenetics & chromatin. 2018;11(1):33.
60. Varambally S, Cao Q, Mani RS, Shankar S, Wang X, Ateeq B, et al. Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science. 2008;322(5908):1695-9.
61. Kodahl AR, Lyng MB, Binder H, Cold S, Gravgaard K, Knoop AS, et al. Novel circulating microRNA signature as a potential non-invasive multi-marker test in ER-positive early-stage breast cancer: a case control study. Molecular oncology. 2014;8(5):874-83.
62. Shen J, Hu Q, Schrauder M, Yan L, Wang D, Medico L, et al. Circulating miR-148b and miR-133a as biomarkers for breast cancer detection. Oncotarget. 2014;5(14):5284-94.
63. Shimomura A, Shiino S, Kawauchi J, Takizawa S, Sakamoto H, Matsuzaki J, et al. Novel combination of serum microRNA for detecting breast cancer in the early stage. Cancer science. 2016;107(3):326-34.
64. Cuk K, Zucknick M, Madhavan D, Schott S, Golatta M, Heil J, et al. Plasma microRNA panel for minimally invasive detection of breast cancer. PloS one. 2013;8(10):e76729.
65. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. Clinical cancer research : an official journal of the American Association for Cancer Research. 2013;19(16):4477-87.

66. Nygren MK, Tekle C, Ingebrigtsen VA, Makela R, Krohn M, Aure MR, et al. Identifying microRNAs regulating B7-H3 in breast cancer: the clinical impact of microRNA-29c. British journal of cancer. 2014;110(8):2072-80.

# Figures

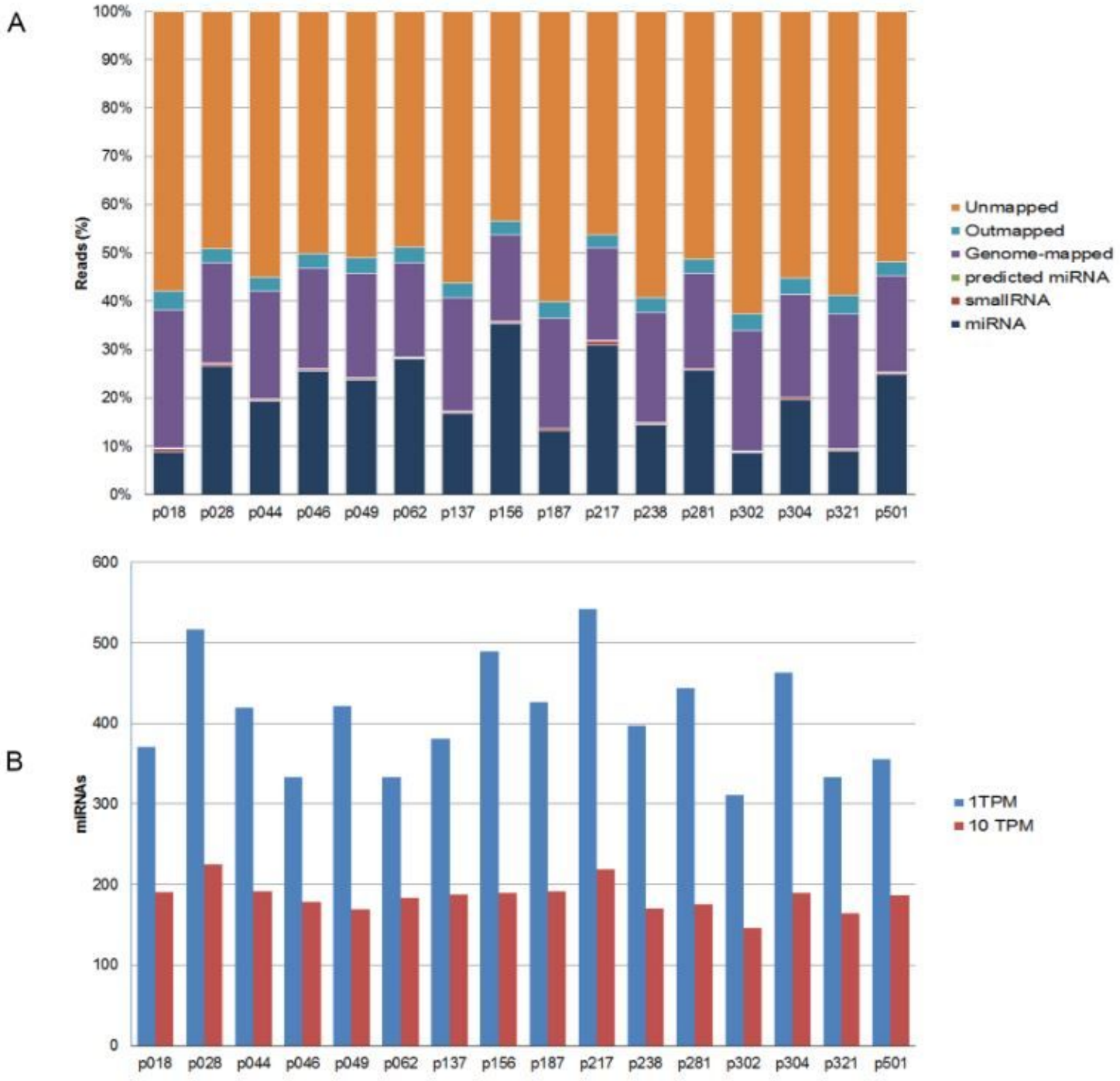
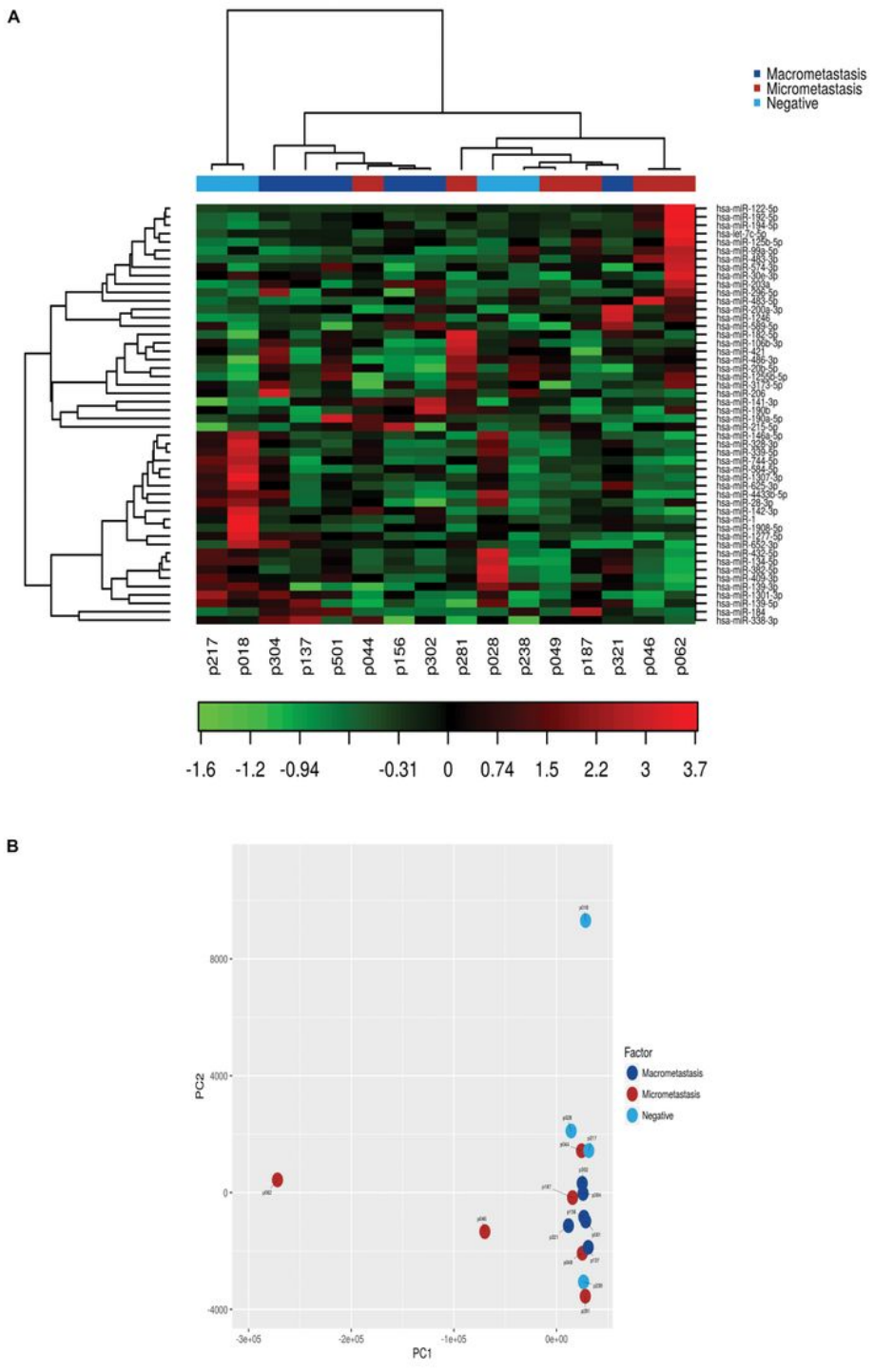


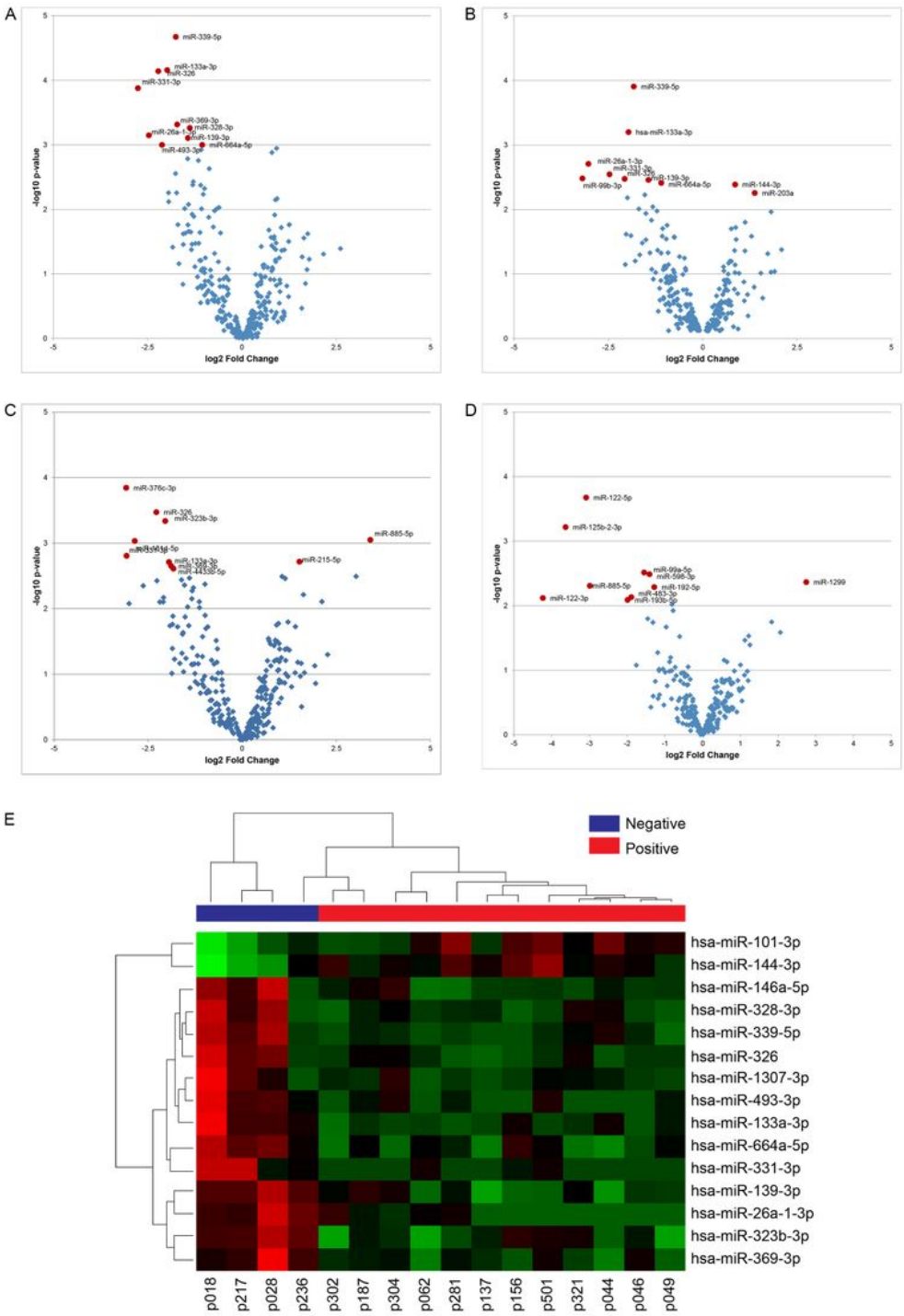

**Figure 1**

Summary of the mapping results for the samples. (A) Percentage of the different reads found for each sample sequenced. Reads are classified as miRNAs, smallRNAs, genome-mapped, outmapped, high abundance (e.g. rRNA, polyA,mtRNA) and unmapped reads. (B) Number of identified miRNAs with TPM normalized numbers of counts >1 (blue bars) or >10 (red bars).



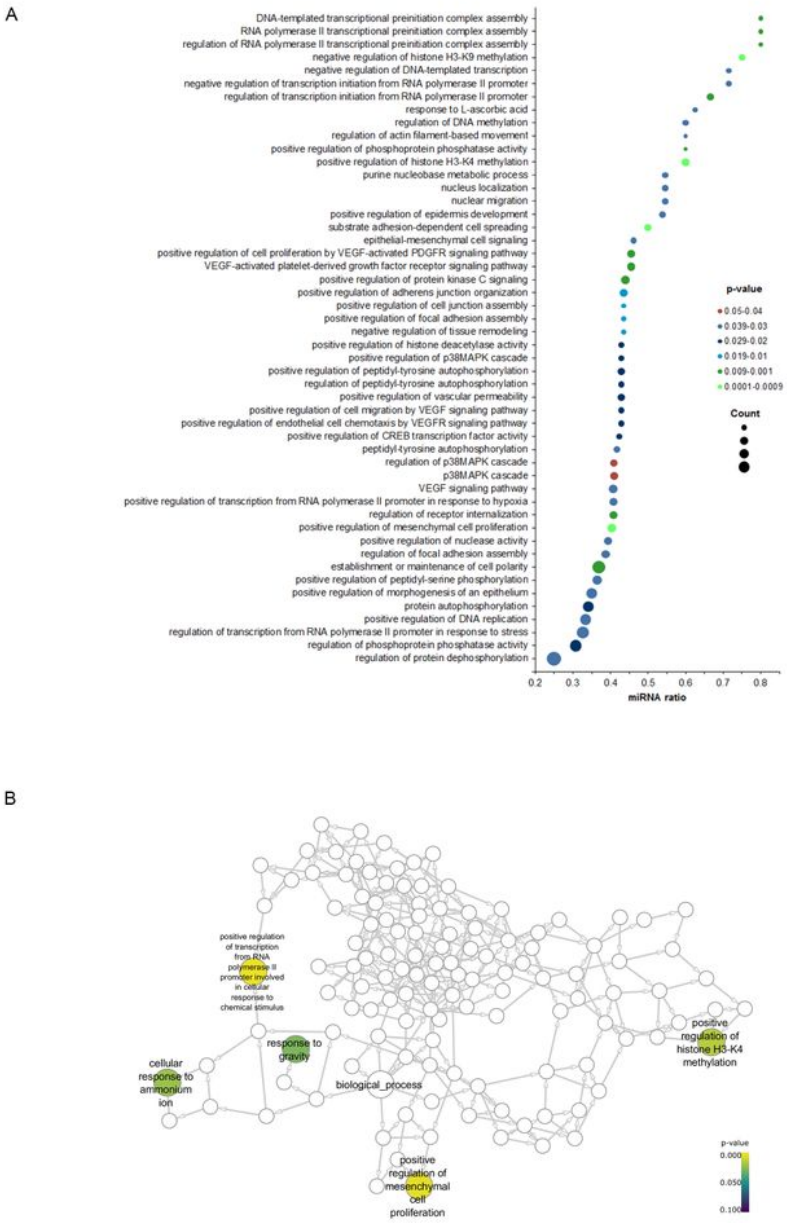

## Figure 2

Class discovery associated with the SLN metastatic status. The analysis was performed on all samples using the 50 miRNAs with the largest coefficient of variation based on TMM counts. (A) Heat map and unsupervised hierarchical clustering analyzed by samples and miRNAs. Each row represents one miRNA and each column represents one sample. The color represents the relative expression level of a miRNA across all samples. The color scale is shown below: red represents an expression level above the mean; green represents an expression level below the mean. (B) Principal Component Analysis (PCA) of samples according to the SLN metastatic status of patients.



## Figure 3

Differential Expressed miRNAs. The volcano plots show the relationship between the p-values and the fold change between the experimental groups. The 10 miRNAs with the lowest p-values are named in the plot. The different experimental groups represent the metastatic SLN status of the patients (A) Positive vs. Negative (B) Macrometastasis vs. Micrometastasis, (C) Macrometastasis vs Negative and (D) Micrometastasis vs Negative, (E) Heat map and hierarchical clustering analyzed by samples and miRNAs. The analysis was performed using the 15 miRNAs differentially expressed between patients with positive and negative SLNs. Each row represents one miRNA and each column represents one sample. The color represents the relative expression level of a miRNA across all samples. The color scale is shown below: red represents an expression level above the mean; green represents an expression level below the mean.




## Figure 4

Gene ontology (GO) enrichment analysis for the significant biological processes associated with positive SLNs. (A) Dot plot graph shows the 50 most significant biological process GO terms (y-axis) and the ratio between the number of expressed miRNAs associated to the GO term and the number of significantly differentially expressed genes associated to the GO term (x-axis). The color of the nodes indicates the p-

value and the size of the nodes the number of miRNAs associated with a specific GO term. (B) Neural network shows the GO terms for the biological processes associated with patients with positive SLNs.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytable3.xlsx
- Supplementarytable2.xlsx
- Supplementarytable1.xlsx
- SuppFig3.tif
- SuppFig2.tif
- SuppFig1.tif
- Additionalinformation.docx