

Circulating miRNA Expression Profiling in Breast Cancer Molecular Subtypes: Applying Machine Learning Analysis in Bioinformatics

ALEXANDRA TRIANTAFYLLOU¹, NIKOLAOS DOVROLIS²,
ELENI ZOGRAFOS³, CHARALAMPOS THEODOROPOULOS¹,
GEORGE C. ZOGRAFOS¹, NIKOLAOS V. MICHALOPOULOS¹ and MARIA GAZOULI³

¹1st Propaedeutic Surgical Department, Hippocraton General

Hospital, National and Kapodistrian University of Athens, Athens, Greece;

²Department of Medicine, Laboratory of Biology, Democritus University of Thrace, Alexandroupolis, Greece;

³Laboratory of Biology, School of Medicine, National and Kapodistrian University of Athens, Athens, Greece

Abstract. *Background/Aim:* Breast cancer is a leading worldwide cause of female cancer-related morbidity and mortality. Since molecular characteristics increasingly guide disease management, demystifying breast tumor miRNA signature emerges as an essential step toward personalized care. This study aimed to investigate the variations in circulating miRNA expression profiles between breast cancer subtypes and healthy controls and to identify relevant target genes and molecular functions. *Materials and Methods:* MiRNA expression was tested by miScriptTM miRNA PCR Array Human Cancer Pathway Finder kit, and subsequently, a machine learning approach was applied for miRNA profiling of the various breast cancer molecular subtypes. *Results:* Serum samples from patients with primary breast cancer (n=66) and healthy controls (n=16) were analyzed. MiR-21 was the single common molecule among all breast cancer subtypes. Furthermore, several miRNAs were found to be differentially expressed explicitly in the different subtypes;

luminal A (miR-23b, miR-142, miR-29a, miR-181d, miR-16, miR-29b, miR-155, miR-181c), luminal B (miR-148a, let-7d, miR-92a, miR-34c, let-7b, miR-15a), HER2+ (miR-125b, miR-134, miR-98, miR-143, miR-138, miR-135b) and triple negative breast cancer (miR-17, miR-150, miR-210, miR-372, let-7f, miR-191, miR-133b, miR-146b, miR-7). Finally, miRNA-associated target genes and molecular functions were identified. *Conclusion:* Applying a machine learning approach to delineate miRNA signatures of various breast cancer molecular subtypes allows further understanding of molecular disease characteristics that can prove clinically relevant.

Breast cancer (BC) is a leading cause of cancer-related morbidity and mortality, affecting 1 in 8 women and accounting for approximately 15% of all cancer-related deaths among females worldwide (1). Optimal disease outcomes are inextricably linked to early diagnosis, highlighting the need for the wide implementation of screening programs and the development of new sensitive biomarkers without the limited sensitivity of the currently available imaging modalities (2). One of the main challenges is the heterogeneity of the disease, with varying prognosis and treatment responses noted among patients of the same clinical stage and pathological features. Hence, merely describing the clinical macroscopic, and histopathological microscopic or immunohistochemical characteristics proves inadequate to sufficiently decide optimal treatment strategy and foretell patient prognosis (3).

Breast cancer classification into distinct molecular subtypes based on specific immunophenotypic features has become the mainstay approach for drafting the therapeutic strategy in a somewhat personalized manner (4). More than two decades ago, Perou and Sorlie were the first to propose a “Molecular Classification” of the disease, according to which the five

Correspondence to: Maria Gazouli, Ph.D., Prof. of Biology-Nanomedicine, Department of Basic Medical Sciences, School of Medicine, National and Kapodistrian University of Athens, Michalakopoulou 176, 11527 Athens, Greece. Tel: +30 2107462231, +30 2107462244, e-mail: mgazouli@med.uoa.gr

Key Words: miRNAs, breast cancer, machine learning, artificial intelligence.

©2022 International Institute of Anticancer Research
www.iiar-anticancer.org



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) 4.0 international license (<https://creativecommons.org/licenses/by-nc-nd/4.0>).

“molecular subtypes of Perou” are identified as: luminal A, luminal B, basal-like, normal-like, and HER2-positive (5). Each intrinsic molecular subtype has unique risk factors, prognosis, prevalence, survival rate, and responsiveness to targeted therapeutic agents that have been integral to the improvement of clinical outcomes (6). Gene expression profiling is now well integrated into everyday clinical practice, but it once represented a paradigm shift that marked the beginning of a new era of individualized medicine, leaving behind the traditional descriptive “morphological” classification for a more integrative approach, that considers clinical features and immunohistochemical biomarkers.

Since the landmark study of Perou, technological advances increased the resolution for large-scale gene-expression and high throughput transcriptome analyses, that have been employed to fully elucidate the inter- and intra-tumor heterogeneous nature of breast malignancies (7). In this way, breast neoplasms can now be stratified into integrative clusters that reveal patterns of single-nucleotide variants and are associated with distinct clinical outcomes and response to therapy, although the clinical significance of these approaches remains under discussion (8). Recently published mutational signatures, based on whole genome sequencing data, highlighted the variability of a breast tumor’s biological features over time due to phenotypic dynamics and genomic evolution (9). In this context, multiple factors affect the available amount of bioactive RNAs, but one of the key cellular instruments to regulate gene expression levels at the post-transcriptional level are microRNAs (miRNAs).

MiRNAs are a group of evolutionary conserved, non-coding, single-stranded ribonucleic acids of short length ranging from 18 to 22 nucleotides, that regulate gene expression *via* binding to specific mRNA targets and leading to their degradation or translational inhibition. Various forms of circulating miRNAs can be detected in the blood, including free miRNA, lipoprotein-miRNA complexes, and miRNAs embedded in extracellular vesicles (10). Although miRNA synthesis is tightly controlled, aberrant expression of miRNAs is directly associated with numerous malignancies, including breast cancer (4). The differences in miRNA composition and level of expression between normal and neoplastic, but also between different subtypes of a particular malignancy (*e.g.*, molecular subtypes of BC), and at various disease stages (*i.e.*, tumor size, regional node status, predominantly metastatic disease), warrants a wide spectrum of clinical applications. Therefore, circulating miRNAs emerge as critical mediators of cancer development and progression, as well as attractive candidates for the discovery of novel biomarkers and therapeutic targets (2, 4, 6).

The exponential growth and complexity of scientific and clinical data in experimental biology have transformed both the theory and practice of oncology, promising a deeper understanding of cancer and, accordingly, a more personalized

and effective oncological care. However, the overwhelming amount of information on large-scale miRNA profiling poses the additional challenge of developing efficient methods to store and process these data and extract clinically meaningful biological knowledge. “Bioinformatics” has been developed to serve this purpose. Bioinformatics is an interdisciplinary field that combines biology, chemistry, physics, computer science, information engineering, mathematics, and statistics to analyze and interpret extensive and complex biological data (11). Bioinformatics typically exploits artificial intelligence (AI), which is broadly defined as the ability of a machine to accomplish tasks typically associated with intelligent human behavior. A branch of AI, machine learning (ML), refers to computer algorithms that can generate predictive models derived from exposure to training data rather than exhaustive *a priori* design, being also able to iteratively self-adjust to optimize their performance (12). ML comprises several subsets that are preferentially applied depending on the available data and research question. For example, commonly used models in biomedical research include Artificial Neural Networks (ANNs), which process information in a manner inspired by the neurons of the human brain, and Support Vector Machines (SVN), which perform classification or regression by optimizing decision boundaries in multidimensional space. Overall, ML has a growing role in oncology research and practice (11, 12).

Our study investigates the variations in circulating miRNA expression profiles between breast cancer subtypes and healthy controls in serum samples. We also employed a linear Support Vector Machine method, an ML subset, to investigate specific patterns more reliably in differential miRNA expression between the various subtypes and assess the predictive value of our investigated miRNAs for the development of breast cancer. Finally, to deeply comprehend the responsible molecular pathways in breast cancer progression and their differential expression in the various subtypes, we queried gene ontology datasets to identify target genes of differentially expressed miRNAs and to perform functional analysis on the gene sets. This work explores potential and applicable miRNA targets with diagnostic and prognostic values in all breast cancer subtypes. By accumulating knowledge in this field and mapping out the molecular pathways in various disease states, the critical features will stand out, allowing for fully individualized approaches in the true era of personalized medicine.

Materials and Methods

Patients and controls. Patients with primary breast cancer (n=66) were selected for analysis with the miRNA PCR array. All patients received surgery and treatment between 2017 and 2018. Tumor subtypes were defined according to estrogen receptor (ER), progesterone receptor (PR), and HER2 expression. ER and PR were considered positive if more than 1% of nuclei were stained. HER2

expression was determined by immunohistochemistry (IHC) staining. Additional samples from 16 women without cancer history were selected as healthy controls. All serum samples (500 µl) were collected from women with breast cancer before surgery or treatment and stored at -80°C before use. Written informed consent was obtained from all subjects for the collection and research use of breast tumors. Our complete study was approved by the ethics committees of the Hippocraton General Hospital of Athens, Greece and the Medical School of National and Kapodistrian University of Athens, Greece. The clinicopathological characteristics of all breast cancer patients are presented in Table I.

Exosome isolation and RNA extraction. Overall, we have used the manufacturer's instructions during the extraction process. According to these instructions, the total RNAs from the exosome pellet were isolated using the NucleoSpin miRNA kit (Machery-Nagel, Germany). All RNA samples were instantly stored at a temperature of 80°C (and remained there until they were used).

miRNA expression. We have used the miScript II RT Kit (Qiagen, Hilden, Germany), to achieve the reverse transcription of 500 ng of RNA. Subsequently, we completed the expression of a panel, which tested for 84 miRNAs, by using the miScript™ miRNA PCR Array Human Cancer PathwayFinder (MIHS-102Z, Qiagen) and miScript SYBR Green PCR Kit (Qiagen). These are custom panel tests for miRNA that have been correlated with tumor diagnosis, staging, progression, or prognosis. Each array contained six different snoRNA/snRNA as a normalization control for the array data (SNORD61, SNORD68, SNORD72, SNORD95, SNORD96A, RNU6-6P), miRNA reverse transcription control (RTC) and positive PCR control (PPC). We were able to classify our samples into five categories: "Healthy Controls", "HER2-positive", "Luminal A", "Luminal B", and "Triple-negative". We also calculated the miRNA relative expression using the $2^{-\Delta\Delta C_t}$ method (used for each miRNA in each sample), normalizing on the geometric mean of the controls. We also calculated the fold change between our groups (represented by fold regulation) using the $2^{-\Delta\Delta C_t}$ method. Lastly, we calculated *p*-values using a Student's *t*-test of the replicate normalized miRNA expression values for each miRNA in our groups; they were also corrected for FDR using the Benjamini-Hochberg method (*p*. adjust function in R).

Machine learning modeling approach. In order to evaluate our expression results and to refine them by identifying the most critical miRNAs that were present among our groups, we employed the "linear Support Vector Machine method" using the caret package in R on the entire miRNA panel, regardless of previous differential expression results (13, 14). Thus, we were able to have a broader approach and not to exclude several features (miRNAs). We completed data pre-processing with "scale" and "center", and we used the previously calculated relative expression (per miRNA and per sample) to validate the accuracy of our model. We independently tested each tumor sample group *versus* the healthy control group in order to determine its predictive value. This model used a 10-fold cross-validation approach with a 70%-30% partitioning of the samples for training and testing accordingly. Therefore, the top 20 important features for each pairing were selected as an input for our next steps. We used the online tool VENNY v. 2.1.0 that permitted us to further identify differentially expressed genes (DEGs) and differentially expressed miRNAs

Table I. *Clinical characteristics of breast cancer patients.*

Variables	Cases, n=66	Healthy women, n=16
Median age (range) (y)	47 (27-68)	45 (26-64)
Molecular subtype, n		
HER2 positive	11	
Luminal A	24	
Luminal B	20	
Triple Negative	11	
Lymph node involvement, n (%)		
Yes	34 (51.52)	
No	32 (48.48)	
Stage, n (%)		
1	10 (15.15)	
2	31 (46.97)	
3	25 (37.88)	

(DEMs), that were common among all breast cancer subtypes; this tool was also utilized to create "Venn Diagrams" of the important features between our models (15).

miRNA target identification and functional analysis. In our study, we were able to detect the affected genes with the use of the aforementioned 20 important miRNAs for all previous models (HER2+, luminal A, luminal B, triple-negative); for this purpose, we used the "multimir R package" (16). We also used the following databases on validated miRNA-target interactions in order to identify the affected genes: "miRecords" (17), "miRTarbase" (18), and "Tarbase" (19). Finally, we enriched the resulting gene lists that were used as input to the cluster profiler package (20) by using Gene Ontology Molecular Function Terms (GO-MF) (21). We calculated *p*-values for the GO-MF rankings with the use of the one-sided Fisher's exact test and the FDR (adjusted by *q*-value). We finally used VENNY v. 2.1.0 to find intersecting GO-MF terms from our groups.

Results

Differential miRNA expression. As described in the methodology section, we performed a separate differential expression calculation of each molecular subtype group *versus* healthy controls. Only miRNAs with an absolute fold change greater than 2 ($FC > 2$) and an adjusted *p*-value lower than 0.05 (*p*.adjust < 0.05) are reported in Table II. The HER2-positive sample group appeared to be the most perturbed (39 miRNAs), followed by the luminal A (38 miRNAs), triple-negative (36 miRNAs), and luminal B (34 miRNAs) groups. Notably, the HER2-positive group had 16 under-expressed miRNAs (with the highest being miR-134 at approx. -18-fold down-regulation) and 23 over-expressed miRNAs (with the highest being miR-200c with approx. 15-fold up-regulation). In the luminal A group, 14 miRNAs were down-regulated (highest: miR-29a at approx. -20-fold), and 24 miRNAs were up-regulated (highest: miR-21 at approx. 54-fold). In the

Table II. *Differential miRNA expression for each sample group vs. controls.*

Luminal A vs. HC		Luminal B vs. HC		HER2+ vs. HC		T. Negative vs. HC	
miRNA	FR	miRNA	FR	miRNA	FR	miRNA	FR
miR-29a	-19.808	miR-20a	-9.541	miR-134	-17.655	miR-206	-6.524
miR-181d	-19.524	miR-148b	-6.424	miR-34a	-7.067	miR-27b	-6.160
miR-181b	-16.513	miR-205	-5.844	miR-20a	-5.296	miR-100	-5.996
miR-29b	-14.732	miR-100	-4.403	miR-10a	-3.955	miR-130a	-5.402
miR-23b	-9.030	miR-10a	-3.883	miR-27b	-3.478	miR-146a	-4.024
miR-148b	-8.056	miR-15a	-3.628	miR-25	-3.438	miR-19	-3.537
miR-181c	-4.574	miR-155	-3.445	miR-30c	-3.394	miR-181b	-3.273
miR-27b	-3.451	miR-130a	-3.045	miR-128	-3.283	miR-150	-3.150
miR-205	-3.248	let-7f	-2.716	miR-130a	-2.821	let-7f	-3.087
miR-32	-2.688	miR-218	-2.713	miR-135b	-2.651	miR-372	-2.739
let-7f	-2.505	miR-146a	-2.540	miR-155	-2.561	miR-10a	-2.712
miR-148a	-2.458	miR-150	-2.468	miR-301a	-2.417	miR-32	-2.619
miR-146a	-2.402	miR-27b	-2.379	miR-17	-2.372	miR-30c	-2.617
miR-214	-2.155	miR-206	-2.270	miR-181c	-2.119	miR-155	-2.478
miR-146b	+2.021	let-7c	-2.262	miR-215	-2.038	miR-34c	-2.250
miR-98	+2.053	miR-301a	-2.183	miR-191	-2.021	miR-205	-2.197
miR-218	+2.126	miR-32	-2.009	miR-210	+2.122	miR-200c	-2.113
miR-184	+2.229	miR-184	+2.009	let-7b	+2.177	miR-125b	+2.303
miR-34a	+2.255	miR-7	+2.013	miR-193a	+2.337	miR-10b	+2.364
miR-193a	+2.262	miR-210	+2.120	miR-20b	+2.344	miR-98	+2.392
miR-210	+2.265	let-7d	+2.237	miR-140	+2.347	miR-7	+2.578
miR-181a	+2.288	miR-193a	+2.395	miR-15b	+2.412	miR-135b	+2.590
miR-133b	+2.474	miR-181d	+2.500	miR-122	+2.601	miR-133b	+2.657
miR-125b	+2.550	miR-181a	+2.519	miR-133b	+2.803	miR-92a	+2.703
miR-125a	+2.624	miR-125a	+2.617	miR-203a	+2.902	miR-146b	+2.834
miR-144	+2.732	let-7b	+3.411	miR-143	+3.267	let-7e	+2.843
miR-196a	+2.829	miR-92a	+3.423	miR-144	+3.270	let-7b	+2.853
let-7e	+2.891	miR-140	+3.834	miR-125a	+3.407	miR-196a	+2.928
let-7b	+3.483	miR-196a	+3.850	miR-148a	+4.361	miR-181d	+3.356
miR-92a	+3.728	let-7e	+5.960	miR-184	+4.502	let-7a	+4.909
miR-20a	+5.642	miR-203a	+6.025	let-7e	+4.528	miR-20a	+6.247
miR-140	+7.663	miR-148	+8.785	miR-206	+4.640	miR-34a	+7.022
miR-142	+7.707	miR-25	+77.079	let-7a	+4.763	miR-17	+16.184
miR-16	+7.729	miR-21	+117.043	miR-181a	+4.853	miR-210	+38.659
miR-155	+7.865			miR-125b	+5.424	miR-25	+40.403
miR-206	+9.034			miR-196a	+8.197	miR-21	+107.636
miR-203a	+31.052			miR-205	+8.709		
miR-21	+54.142			miR-21	+13.549		
				miR-200c	+14.947		

FR: Fold regulation; HC: healthy controls.

Luminal B group, 17 miRNAs were down-regulated (highest: miR-20a at approx. -10-fold), and 17 miRNAs were up-regulated (highest: miR-21 at approx. 117-fold). Finally, in the triple-negative group, 17 miRNAs were down-regulated (highest: miR-206 at approx. -7-fold), and 19 miRNAs were up-regulated (highest: miR-21 at approx. +108-fold) compared to controls. Overall, there were many commonalities between our subtype groups regarding differential miRNA expression, with the most notable being miR-21 having the highest up-regulation in three out of four groups, and the second highest in the triple-negative subset of patients.

Machine learning models. Applying a ML approach to assess the predictive value of our investigated miRNAs in breast cancer, allowed us to go further than the traditional differential expression results. All four models performed exceptionally well, with the models involving luminal A, luminal B and triple-negative samples achieving 100% accuracy with a Kappa of 1 and the one involving the HER2+ sample group a 95% accuracy and a Kappa of 0.875. Figure 1 shows the top 20 important features for each of our models, whereas Figure 2 depicts a Venn diagram of these features for each group. MiR-21 was the only common

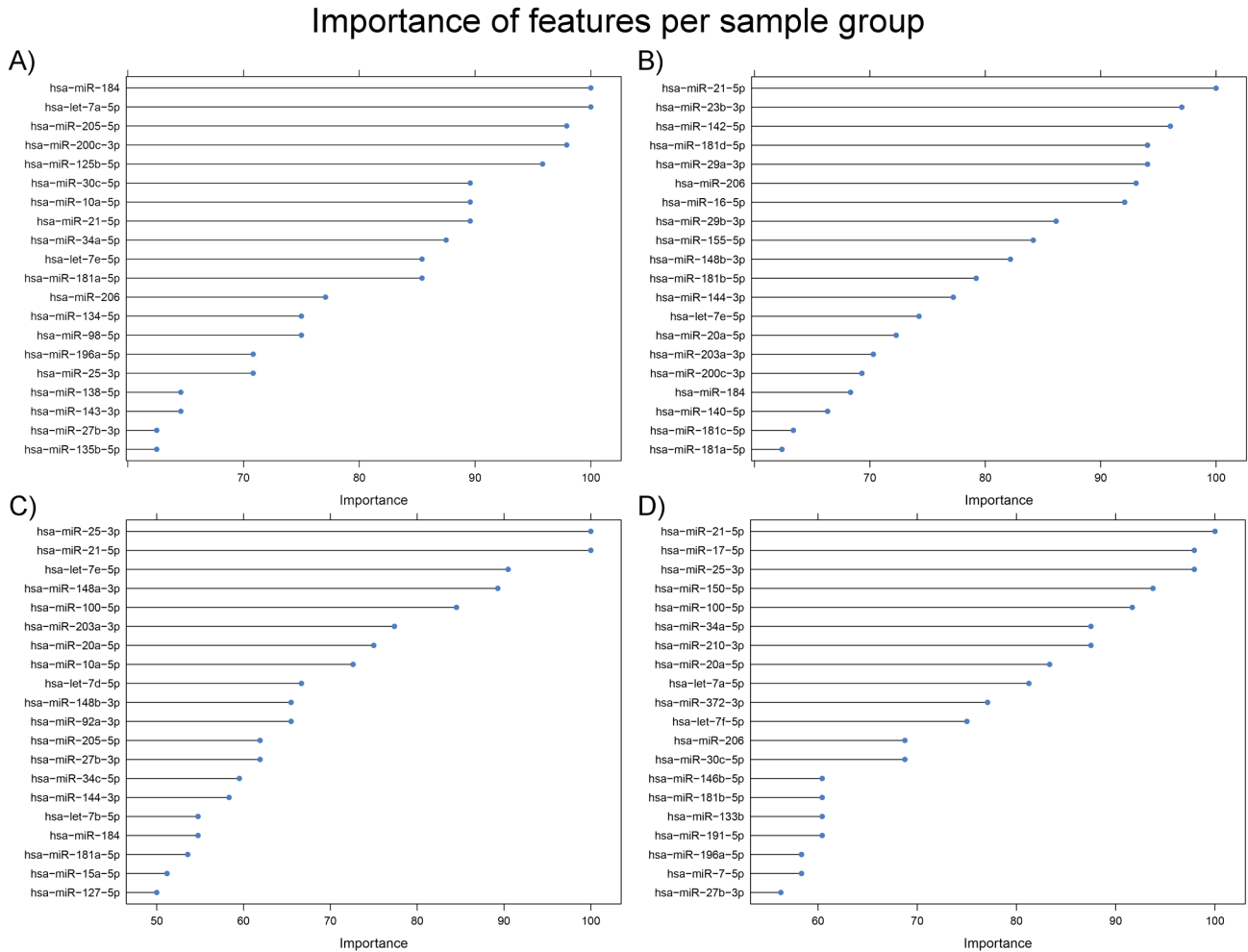


Figure 1. Top 20 important predictive features for our Support Vector Machine models. A) HER2+ group, B) Luminal A group, C) Luminal B group, D) Triple negative group.

molecule between all four models. Other notable findings include: 1) Let-7a, miR-30c, miR-34a, and miR-196a were not present as important features in both luminal groups. 2) MiR-184, let-7e, and miR-181a were not present in the triple-negative group. 3) MiR-148b, miR-144, miR-203a, and miR-140 were only in the top important features of the two luminal groups. 4) Six miRNAs (miR-125b, miR-134, miR-98, miR-143, miR-138, miR-135b) were in the top 20 of important features only for the HER2+ group; 8 miRNAs (miR-23b, miR-142, miR-29a, miR-181d, miR-16, miR-29b, miR-155, miR-181c) were in the top 20 of important features only for the luminal A group; 6 miRNAs (miR-148a, let-7d, miR-92a, miR-34c, let-7b, miR-15a) were in the top 20 of important features only for the luminal B group, and finally, 9 miRNAs (miR-17, miR-150, miR-210, miR-372, let-7f, miR-191, miR-133b, miR-146b, miR-7) were in the top 20 of important features only for the triple-negative group.

miRNA target identification and functional analysis. To further elucidate the biological background of each breast cancer sub-phenotype, we used the previously identified important features to identify their target genes and performed functional analysis on the gene sets. In total, 11,346 genes were identified for the HER2+, 12,181 for the luminal A, 11,502 for the luminal B, and 28,160 for the triple-negative subtype. The top 20 pathways *via* GO-MF for each group are depicted in Figure 3. Since there was an overlap in identified miRNAs, it was only logical that an intersection of the top 20 molecular functions for each sample group existed as well. Thirteen distinct functions appear to be shared among all sub-phenotypes (“cadherin binding”, “transcription coregulator activity”, “DNA-binding transcription factor binding”, “RNA polymerase II-specific DNA-binding transcription factor binding”, “ubiquitin-like protein ligase binding”, “GTPase binding”, “ubiquitin protein ligase binding”, “transcription

coactivator activity”, “small GTPase binding”, “ubiquitin-like protein transferase activity”, “ribonucleoprotein complex binding”, “nuclear hormone receptor binding”, “protein C-terminus binding”); HER2+ and luminal A were the only ones who shared “SMAD binding” and “transcription corepressor activity”; luminal B and triple-negative had “ATPase activity” and “catalytic activity acting on RNA” in common; HER2+ and triple-negative were the only ones who shared functions, namely “protein heterodimerization activity” and “histone deacetylase binding”; luminal A and luminal B had only one unique molecular function common between them (“ubiquitin-like protein binding”), and the HER2+ shared “protein heterodimerization activity” and “histone deacetylase binding” with the triple-negative group. In addition, “ubiquitin-protein transferase activity” was identified in all sub-phenotypes, except luminal B. Finally, only luminal A and luminal B molecular subtypes had unique functions, which characterized them individually. For the former, it was “phosphoric ester hydrolase activity”, “molecular adaptor activity”, “nucleoside binding”, and for the latter, “protein kinase regulator activity” and “histone binding.” These intersections are visualized as a Venn diagram in Figure 4.

Discussion

Efforts to better understand the mechanisms involved in breast cancer initiation, progression, and resistance are more relevant now than ever before, as we enter the era of personalized medical oncology. In this context, miRNAs are highly attractive molecules capable of being clinically utilized as diagnostic, prognostic, disease monitoring, and predictive biomarkers. Due to the high-level of complexity in miRNA research, the differential expression of miRNAs among distinct breast cancer subtypes, and the combined analysis of these profiles, can help elucidate the functional effects of miRNA expression in breast cancer.

In the present study, we performed a comparative analysis of miRNA expression profiles between breast tumors of all four intrinsic molecular subtypes (luminal A, luminal B, HER2-enriched, and basal-like) and healthy controls. Using a quantitative reverse transcription-PCR (qRT-PCR) array approach, we identified 147 differentially expressed miRNAs among BC subtypes; each molecular subtype group was separately analyzed in comparison to healthy controls. As the amount of data was extensive and in order to fully comprehend the emerging expression patterns, we applied a ML approach to assess the predictive value of our investigated miRNAs in breast cancer tumorigenesis. Interestingly, miR-21 was the only common molecule between all four breast cancer models. It was significantly enriched in all breast cancer subtypes compared to healthy counterparts, exhibiting the highest expression in three out of four subtypes (*i.e.*, luminal A, luminal B and TNBC) and the second highest in the

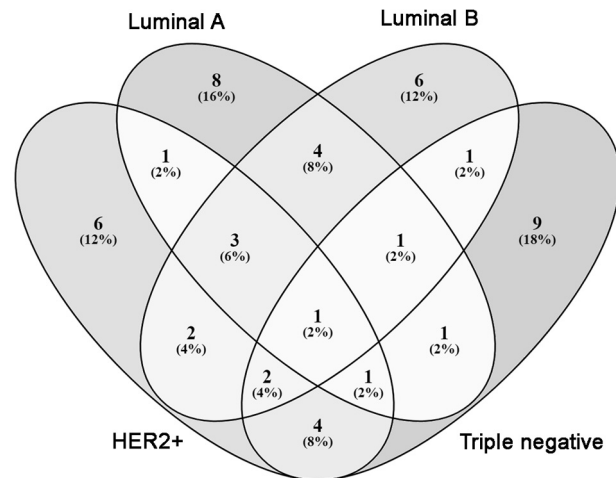


Figure 2. Commonalities and differences in the top 20 important predictive miRNAs between sample groups from our Support Vector Machine models.

remaining one (*i.e.*, HER2+). This finding comes as no surprise, since miR-21 is a well-recognized onco-miRNA located in the FRA17B, a fragile site that is frequently amplified in different types of cancers, including breast cancer (22). High expression levels of miR-21 have been independently correlated with advanced clinical stage, lymph node metastasis, and shortened survival of breast cancer patients (23). MiR-21 enhances tumor growth and invasion by regulating signaling pathways involved in cell cycle, repair of DNA damage, and apoptosis, as well as the NF- κ B signaling pathway (24). Among the genes whose expression levels are altered by miR-21 are the tumor suppressor genes *PDCD4*, *PTEN*, and *TMPI* (22, 25). Recently, *LZTFL1* was identified as a novel target gene of miR-21. This gene may also function as a tumor suppressor, possibly by interacting with E-cadherin and the actin cytoskeleton and thereby, regulating the transition of epithelial cells to mesenchymal cells (EMT). The investigators found that knockdown of *LZTFL1* overcame the suppression of miR-21 inhibitor on cell proliferation, metastasis, and the expression of EMT markers in breast cancer cells (26). Furthermore, plasma miR-21 levels have been proposed as a potential biomarker for the detection of both primary and recurrent breast cancer, as well as a promising strategy in breast cancer therapy since inhibition of miR-21 suppressed cell proliferation and metastasis in breast cancer cells (22, 26). Finally, the suppression of miR-21 has shown to sensitize breast cancer cells to widely available anticancer agents, such as taxol, trastuzumab, topotecan, and doxorubicin (27).

Additional aberrant miRNAs, well-described in breast cancer literature, were identified in our study, confirming their important implication in tumorigenesis. Out of these,

Functional analysis of genes affected by top features

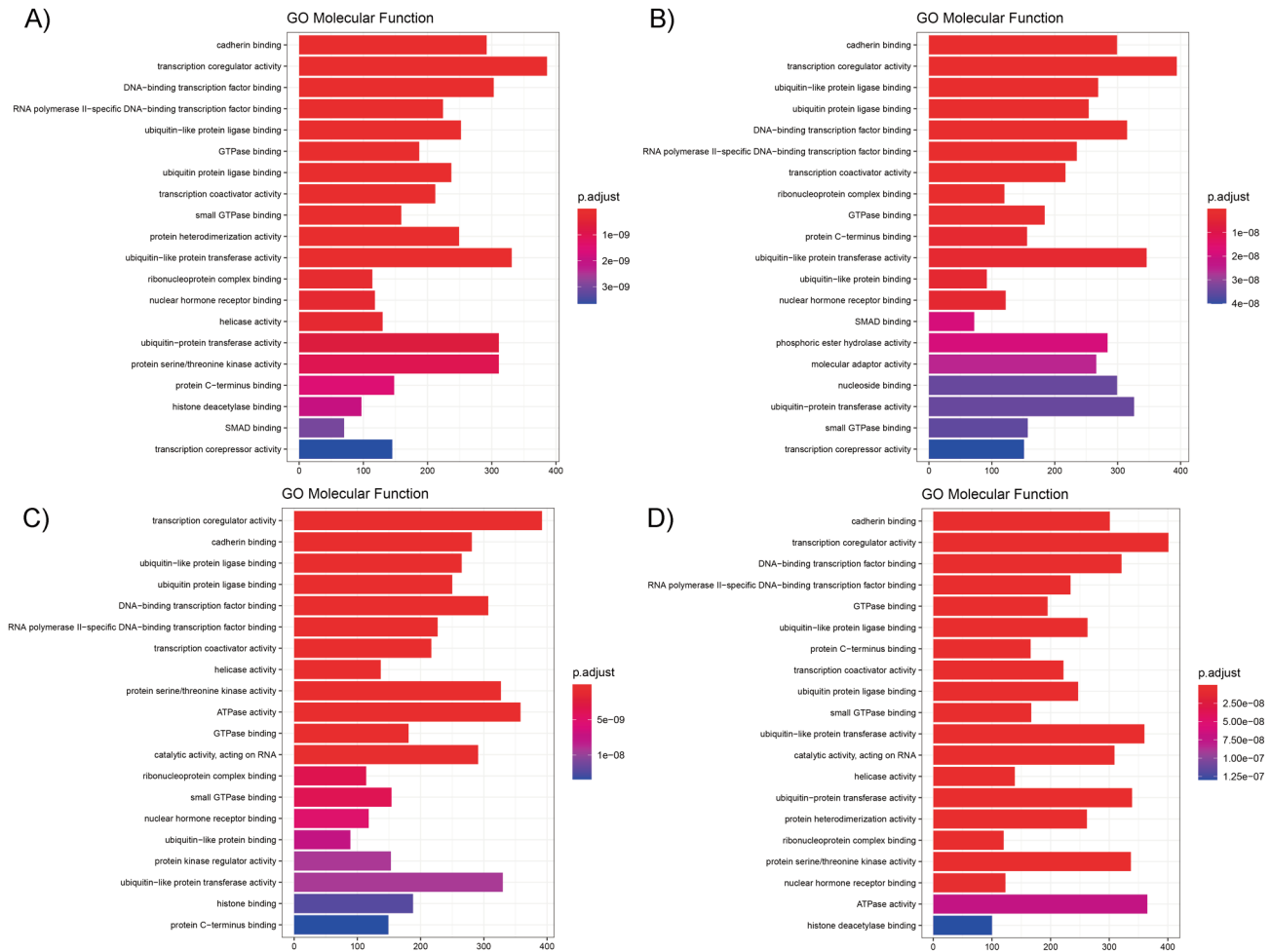


Figure 3. Top 20 molecular functions identified from the target genes of the important predictive miRNAs for each sample group: A) HER2+ group, B) Luminal A group, C) Luminal B group, D) Triple-negative group.

four miRNA molecules, namely miR-148b, miR-144, miR-203a, and miR-140, were identified as top predictive indices exclusively in ER-positive breast cancers (Luminal A and B). Specifically, miR-148b was found to be significantly down-regulated in both luminal subtypes. This is in line with previous studies, showing that miR-148b acts as a tumor-suppressive miRNA in breast cancer progression by targeting a series of cancer-related oncogenes, thus being down-regulated in the serum samples from breast cancer patients (28, 29). However, some contradicting results have shown that inhibition of miR-148b promoted apoptotic cell death *via* the PARP pathway, revealing its complex role in carcinogenesis (30). Another important predictive molecule for both luminal subtypes was miR-144. Extensive

bibliographic evidence suggests that up-regulation of miR-144 acts in a tumor-suppressive manner, with lower miR-144 expression being associated with poor differentiation, higher clinical stage, and lymph node metastasis in patients with breast cancer. These functions are mediated through inhibiting the ZEB1/2 transcription factor, that promotes tumor invasion and metastasis by inducing EMT (31). Interestingly, in our study, miR-144 was significantly up-regulated in luminal A subtype cancers, which is considered the subtype with the best prognosis, exhibiting a significantly lower relapse rate than the others (3). Additionally, miR-203a was found to be significantly up-regulated in both luminal subtypes. Previously published data confirmed the oncogenic role of miR-203a in breast cancer, with gene ontology

analysis demonstrating that miR-203a affects molecular functions associated with “plasma membrane integrity”, “cell surface receptor linked signal and transduction”, and “3′,5′-cyclic nucleotide phosphodiesterase activity”, primarily through altering the insulin-like growth factor receptor (IGF1) gene expression (32). Finally, miR-140 emerged in our study as of high predictive value in hormone receptor-positive disease, being significantly over-expressed in both luminal subtypes. However, throughout the literature, miR-140 is commonly considered a tumor-suppressive miRNA in luminal subtype breast carcinoma. Probably, this discordance is attributed to its role in the maintenance of basal-like features in breast cancer stem cells (BCSCs), characterized by basal epithelium cytokeratin expression and negative ER status (33). Its predictive capacity though remains undisputed since miR-140 can sensitize BCSCs to doxorubicin by down-regulating the Wnt/ β -catenin signaling pathway (34).

Furthermore, ML analysis distinguished the top predictive miRNAs explicitly identified in HER-2 positive breast cancers (miR-125b, miR-134, miR-98, miR-143, miR-138, miR-135b), with the ones ranking highest among them being miR-125b and miR-134, as it is presented in Figure 1A. Characteristically, miR-125b was significantly up-regulated in the HER2-positive group compared to healthy counterparts. Previous literature references concerning miR-125b demonstrate it to be an onco-miRNA whose function is mediated *via* the repression of the pro-apoptotic regulator gene BAK1. Moreover, higher miR-125b expression has been reported in non-responsive patients after admission of 5-fluorouracil, whereas miR-125b deletion on chromosome 11q was correlated with a benefit of anthracycline-based chemotherapy and a low recurrence rate in patients with lymph node-negative breast cancer (35). Herein, miR-134 was significantly down-regulated in the HER2-positive group compared to healthy counterparts. Our findings agree with previously published data suggesting a tumor suppressor role of miR-134. Low expression levels of miR-134 were observed in human breast cancer cell lines and significantly associated with lymph node metastasis, TNM stage, and reduced cell differentiation. It was suggested that miR-134 inhibited the growth, migration, and invasion of breast cancer cells *via* directly down-regulating KRAS (36).

This study identified miRNA molecules that were amongst the top predictors exclusively in the clinically aggressive TNBC subtype (miR-17, miR-150, miR-210, miR-372, let-7f, miR-191, miR-133b, miR-146b, miR-7), among which the top predictive features were exhibited by miR-17, miR-150, and miR-210 (Figure 1D). Specifically, miR-17 was found to be significantly up-regulated in the TNBC group compared to healthy counterparts. Our findings agree with a previously published study regarding miR-17 family over-expression in high-grade triple-negative tumors. The investigators found that miR-17 expression is coregulated with the transcription factor

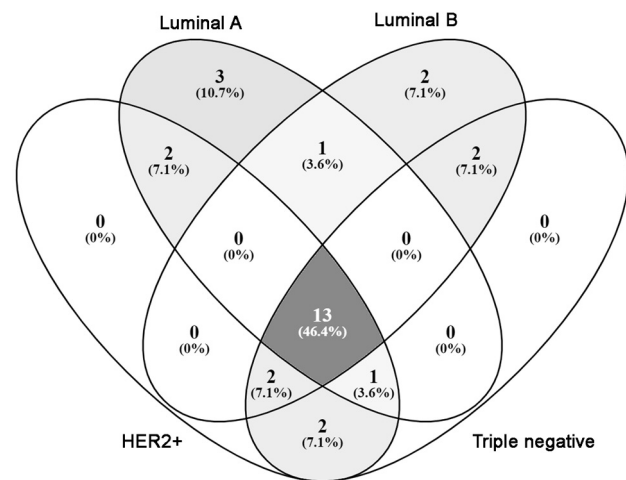


Figure 4. Intersection of the top 20 molecular functions identified by the target genes of the important predictive miRNAs for each sample group.

and proto-oncogene Myc. In addition, the miR-17 promoter has binding sites for HES1, a transcriptional repressor in the Notch signaling pathway, which is also over-expressed in triple-negative breast cancer (37). However, miR-150 was found to be significantly down-regulated in the TNBC group analyzed in our study, in contrast to previously published data. Lu et al. suggested that over-expression of miR 150 enhances breast cancer cell proliferation, invasion, and migration, as well as increases the expression of mesenchymal cell markers (vimentin, N cadherin and β catenin) and decreases the expression of epithelial cell markers (E cadherin and zonula occludens 1), *via* negatively regulating SRCIN1 (38). Although our results differ, it can nevertheless be argued that miR-150-5p expression levels in TNBC are associated with advanced tumor grade and poor patient survival, as it has been recently confirmed (39), factors that were out of the scope of our analysis.

Concurrently, outputs concerning the expression of these miRNAs were associated with specific molecular pathways *via* molecular function gene ontology datasets, allowing us to get a closer look to the preferential involvement of several pathways in breast cancer carcinogenesis as a whole and in specific subtypes. The gene ontology analysis was significantly enriched in transcription coregulation, for all breast cancer subgroups. In fact, many transcriptional coactivators and corepressors have been involved in endocrine response and resistance in breast cancer. Characteristically, ER, the single most important target in breast cancer, regulates gene expression by recruiting transcriptional coregulators and components of the basal transcription machinery, in common with other nuclear receptors (40). Interestingly, gene ontology dataset research did not identify any unique molecular

functions for the triple-negative subset of tumors, inferring what has already been extensively researched, concerning the molecular heterogeneity of this entity (41). Although most triple-negative cases exhibit basal-like features, a discordance has been observed, since some of them turn out to not be basal-like by microarray analysis, whereas others have been found to express hormone receptors or HER2 (42). This observation should be taken into consideration when attempting to interpret the presented interactions and potential regulation regarding miRNAs in triple negative breast cancer.

Our study has several limitations principally associated with the lack of available information about the histological features and clinical stage of the included tumors. As already stated, miRNA expression is altered during carcinogenesis, which may partially explain any inconsistent results regarding miRNA expression between our study and previously published ones. Another issue stems from the relatively small sample size of the patients enrolled. Furthermore, a potential pitfall is related to the ML analysis itself. In essence, we are dealing with an elegant statistical process, which takes advantage of a bundle of training data to draw conclusions for raw data, being susceptible to any heterogeneity between these groups and to statistical errors.

Despite these limitations, our study importantly investigated an actual clinical population instead of merely performing an in vitro analysis of breast cancer cell lines or in silico database research. Furthermore, a parallel analysis of all breast cancer molecular subtypes was achieved in an ethnically homogenous population (Caucasian women of Greek ancestry) under the same research conditions, thus avoiding potential errors arising from interstudy comparisons. Our study greatly enriches the present literature regarding molecular characteristics of breast cancer and the accompanying differential miRNA expression profiles by offering well-founded results stemming from a combination of experimental data, ML analysis, and gene ontology database review.

Conclusion

Currently, breast cancer management increasingly relies on the molecular characteristics of the disease. Accumulating knowledge suggests an ever-deepening level of detail and that miRNA profiling may be an essential key to demystifying subtle biological and clinical features of the various disease subtypes. Besides, miRNAs are closely linked to cancer development and progression, appearing to be promising biomarkers for disease diagnosis, prognosis, and response to treatment. Today, this ever-growing amount of data can be efficiently classified, manipulated, and interpreted through the various subsets of artificial intelligence under the interdisciplinary field of “bioinformatics.” Overall, tumor miRNA profiling appears as a crucial step towards the era of a fully personalized approach to breast cancer.

Conflicts of Interest

All Authors declare no conflicts of interest in relation to this study.

Authors' Contributions

AT and MG conceptualized the project and methodology. ND, with the support of AT and CT analyzed the data and generated the tables and figures. AT, EZ, and CT prepared the original draft under the supervision of GCZ, NVM, and MG. All Authors provided critical feedback, contributed to the manuscript, and approved the final version in accordance with criteria established by the International Committee of Medical Journal Editors (ICMJE).

Acknowledgements

This study was funded by the non-profit organization of the Hellenic Society of Cancer Biomarkers and Targeted Therapy.

References

- 1 Siegel RL, Miller KD, Fuchs HE and Jemal A: Cancer statistics, 2022. *CA Cancer J Clin* 72(1): 7-33, 2022. PMID: 35020204. DOI: 10.3322/caac.21708
- 2 Triantafyllou A, Gazouli M, Theodoropoulos C, Zografos E, Zografos GC and Michalopoulos NV: Exosomes in breast cancer management: Where do we stand? A literature review. *Biol Cell* 114(4): 109-122, 2022. PMID: 35080041. DOI: 10.1111/boc.202100081
- 3 Yersal O and Barutca S: Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol* 5(3): 412-424, 2014. PMID: 25114856. DOI: 10.5306/wjco.v5.i3.412
- 4 Kudela E, Samec M, Koklesova L, Liskova A, Kubatka P, Kozubik E, Rokos T, Pribulova T, Gabonova E, Smolar M and Biringer K: miRNA expression profiles in luminal a breast cancer-implications in biology, prognosis, and prediction of response to hormonal treatment. *Int J Mol Sci* 21(20): 7691, 2020. PMID: 33080858. DOI: 10.3390/ijms21207691
- 5 Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO and Botstein D: Molecular portraits of human breast tumours. *Nature* 406(6797): 747-752, 2000. PMID: 10963602. DOI: 10.1038/35021093
- 6 Mehrgou A, Ebadollahi S, Jameie B and Teimourian S: Analysis of subtype-specific and common Gene/MiRNA expression profiles of four main breast cancer subtypes using bioinformatic approach; Characterization of four genes, and two MicroRNAs with possible diagnostic and prognostic values. *Informatics in Medicine Unlocked* 20: 100425, 2020. DOI: 10.1016/j.imu.2020.100425
- 7 Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Gräf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, METABRIC Group., Langerød A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowitz F, Murphy L, Ellis I, Purushotham A, Børresen-Dale AL, Brenton JD, Tavaré S, Caldas C and Aparicio S: The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486(7403): 346-352, 2012. PMID: 22522925. DOI: 10.1038/nature10983

- 8 Russnes HG, Lingjærde OC, Børresen-Dale AL and Caldas C: Breast cancer molecular stratification: from intrinsic subtypes to integrative clusters. *Am J Pathol* 187(10): 2152-2162, 2017. PMID: 28733194. DOI: 10.1016/j.ajpath.2017.04.022
- 9 Morganello S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, Sieuwerts AM, Brinkman AB, Martin S, Ramakrishna M, Butler A, Kim HY, Borg Å, Sotiriou C, Futreal PA, Campbell PJ, Span PN, Van Laere S, Lakhani SR, Eyfjord JE, Thompson AM, Stunnenberg HG, van de Vijver MJ, Martens JW, Børresen-Dale AL, Richardson AL, Kong G, Thomas G, Sale J, Rada C, Stratton MR, Birney E and Nik-Zainal S: The topography of mutational processes in breast cancer genomes. *Nat Commun* 7: 11383, 2016. PMID: 27136393. DOI: 10.1038/ncomms11383
- 10 Nik Mohamed Kamal NNSB and Shahidan WNS: Non-Exosomal and Exosomal Circulatory MicroRNAs: Which Are More Valid as Biomarkers? *Front Pharmacol* 10: 1500, 2020. PMID: 32038230. DOI: 10.3389/fphar.2019.01500
- 11 Larrañaga P, Calvo B, Santana R, Bielza C, Galdiano J, Inza I, Lozano JA, Armañanzas R, Santafé G, Pérez A and Robles V: Machine learning in bioinformatics. *Brief Bioinform* 7(1): 86-112, 2006. PMID: 16761367. DOI: 10.1093/bib/bbk007
- 12 Nagy M, Radakovich N and Nazha A: Machine learning in oncology: what should clinicians know? *JCO Clin Cancer Inform* 4: 799-810, 2020. PMID: 32926637. DOI: 10.1200/CCI.20.00049
- 13 Kuhn M: Building predictive models in R using the caret package. *Journal of Statistical Software* 28(5): 1-26, 2015. DOI: 10.18637/JSS.V028.I05
- 14 Ihaka R and Gentleman R: R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5(3): 299, 2018. DOI: 10.2307/1390807
- 15 Oliveros J: Venny 2.1.0. An interactive tool for comparing lists with Venn Diagrams, 2007. Available at: <https://bioinfogp.cnb.csic.es/tools/venny/> [Last accessed on June 9, 2022]
- 16 Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, Mahaffey S, Rossi S, Calin GA, Bemis L and Theodorescu D: The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. *Nucleic Acids Res* 42(17): e133, 2014. PMID: 25063298. DOI: 10.1093/nar/gku631
- 17 Xiao F, Zuo Z, Cai G, Kang S, Gao X and Li T: miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* 37(Database issue): D105-D110, 2009. PMID: 18996891. DOI: 10.1093/nar/gkn851
- 18 Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, Tang Y, Chen YG, Jin CN, Yu Y, Xu JT, Li YM, Cai XX, Zhou ZY, Chen XH, Pei YY, Hu L, Su JJ, Cui SD, Wang F, Xie YY, Ding SY, Luo MF, Chou CH, Chang NW, Chen KW, Cheng YH, Wan XH, Hsu WL, Lee TY, Wei FX and Huang HD: miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res* 48(D1): D148-D154, 2020. PMID: 31647101. DOI: 10.1093/nar/gkz896
- 19 Paraskevopoulou MD, Vlachos IS and Hatzigeorgiou AG: DIANA-TarBase and DIANA suite tools: studying experimentally supported microRNA targets. *Curr Protoc Bioinformatics* 55: 12.14.1-12.14.18, 2016. PMID: 27603020. DOI: 10.1002/cpbi.12
- 20 Yu G, Wang LG, Han Y and He QY: clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16(5): 284-287, 2012. PMID: 22455463. DOI: 10.1089/omi.2011.0118
- 21 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM and Sherlock G: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1): 25-29, 2000. PMID: 10802651. DOI: 10.1038/75556
- 22 Motamedi M, Hashemzadeh Chaleshtori M, Ghasemi S and Mokarian F: Plasma level of miR-21 and miR-451 in primary and recurrent breast cancer patients. *Breast Cancer (Dove Med Press)* 11: 293-301, 2019. PMID: 31749630. DOI: 10.2147/BCTT.S224333
- 23 Yan LX, Huang XF, Shao Q, Huang MY, Deng L, Wu QL, Zeng YX and Shao JY: MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. *RNA* 14(11): 2348-2360, 2008. PMID: 18812439. DOI: 10.1261/rna.1034808
- 24 Hong L, Han Y, Zhang Y, Zhang H, Zhao Q, Wu K and Fan D: MicroRNA-21: a therapeutic target for reversing drug resistance in cancer. *Expert Opin Ther Targets* 17(9): 1073-1080, 2013. PMID: 23865553. DOI: 10.1517/14728222.2013.819853
- 25 Wickramasinghe NS, Manavalan TT, Dougherty SM, Riggs KA, Li Y and Klinge CM: Estradiol downregulates miR-21 expression and increases miR-21 target gene expression in MCF-7 breast cancer cells. *Nucleic Acids Res* 37(8): 2584-2595, 2009. PMID: 19264808. DOI: 10.1093/nar/gkp117
- 26 Wang H, Tan Z, Hu H, Liu H, Wu T, Zheng C, Wang X, Luo Z, Wang J, Liu S, Lu Z and Tu J: microRNA-21 promotes breast cancer proliferation and metastasis by targeting LZTFL1. *BMC Cancer* 19(1): 738, 2019. PMID: 31351450. DOI: 10.1186/s12885-019-5951-3
- 27 Najjary S, Mohammadzadeh R, Mokhtarzadeh A, Mohammadi A, Kojabad AB and Baradaran B: Role of miR-21 as an authentic oncogene in mediating drug resistance in breast cancer. *Gene* 738: 144453, 2020. PMID: 32035242. DOI: 10.1016/j.gene.2020.144453
- 28 Cimino D, De Pittà C, Orso F, Zampini M, Casara S, Penna E, Quaglino E, Forni M, Damasco C, Pinatel E, Ponzzone R, Romualdi C, Brisken C, De Bortoli M, Biglia N, Provero P, Lanfranchi G and Taverna D: miR148b is a major coordinator of breast cancer progression in a relapse-associated microRNA signature by targeting ITGA5, ROCK1, PIK3CA, NRAS, and CSF1. *FASEB J* 27(3): 1223-1235, 2013. PMID: 23233531. DOI: 10.1096/fj.12-214692
- 29 Mangolini A, Ferracin M, Zanzi MV, Saccenti E, Ebnaof SO, Poma VV, Sanz JM, Passaro A, Pedriali M, Frassoldati A, Querzoli P, Sabbioni S, Carcoforo P, Hollingsworth A and Negrini M: Diagnostic and prognostic microRNAs in the serum of breast cancer patients measured by droplet digital PCR. *Biomark Res* 3: 12, 2015. PMID: 26120471. DOI: 10.1186/s40364-015-0037-0
- 30 Dai W, He J, Zheng L, Bi M, Hu F, Chen M, Niu H, Yang J, Luo Y, Tang W and Sheng M: miR-148b-3p, miR-190b, and miR-429 regulate cell progression and act as potential biomarkers for breast cancer. *J Breast Cancer* 22(2): 219-236, 2019. PMID: 31281725. DOI: 10.4048/jbc.2019.22.e19
- 31 Pan Y, Zhang J, Fu H and Shen L: miR-144 functions as a tumor suppressor in breast cancer through inhibiting ZEB1/2-mediated epithelial mesenchymal transition process. *Oncotargets Ther* 9: 6247-6255, 2016. PMID: 27785072. DOI: 10.2147/OTT.S103650
- 32 Cai KT, Feng CX, Zhao JC, He RQ, Ma J and Zhong JC: Upregulated miR 203a 3p and its potential molecular mechanism

- in breast cancer: A study based on bioinformatics analyses and a comprehensive meta analysis. *Mol Med Rep* 18(6): 4994-5008, 2018. PMID: 30320391. DOI: 10.3892/mmr.2018.9543
- 33 Niu T, Zhang W and Xiao W: MicroRNA regulation of cancer stem cells in the pathogenesis of breast cancer. *Cancer Cell Int* 21(1): 31, 2021. PMID: 33413418. DOI: 10.1186/s12935-020-01716-8
 - 34 Wu D, Zhang J, Lu Y, Bo S, Li L, Wang L, Zhang Q and Mao J: miR-140-5p inhibits the proliferation and enhances the efficacy of doxorubicin to breast cancer stem cells by targeting Wnt1. *Cancer Gene Ther* 26(3-4): 74-82, 2019. PMID: 30032164. DOI: 10.1038/s41417-018-0035-0
 - 35 van Schooneveld E, Wildiers H, Vergote I, Vermeulen PB, Dirix LY and Van Laere SJ: Dysregulation of microRNAs in breast cancer and their potential role as prognostic and predictive biomarkers in patient management. *Breast Cancer Res* 17: 21, 2015. PMID: 25849621. DOI: 10.1186/s13058-015-0526-y
 - 36 Su X, Zhang L, Li H, Cheng P, Zhu Y, Liu Z, Zhao Y, Xu H, Li D, Gao H and Zhang T: MicroRNA-134 targets KRAS to suppress breast cancer cell proliferation, migration and invasion. *Oncol Lett* 13(3): 1932-1938, 2017. PMID: 28454346. DOI: 10.3892/ol.2017.5644
 - 37 Moi L, Braaten T, Al-Shibli K, Lund E and Busund LR: Differential expression of the miR-17-92 cluster and miR-17 family in breast cancer according to tumor type; results from the Norwegian Women and Cancer (NOWAC) study. *J Transl Med* 17(1): 334, 2019. PMID: 31581940. DOI: 10.1186/s12967-019-2086-x
 - 38 Lu Q, Guo Z and Qian H: Role of microRNA-150-5p/SRCIN1 axis in the progression of breast cancer. *Exp Ther Med* 17(3): 2221-2229, 2019. PMID: 30867707. DOI: 10.3892/etm.2019.7206
 - 39 Sugita BM, Rodriguez Y, Fonseca AS, Nunes Souza E, Kallakury B, Cavalli IJ, Ribeiro EMSF, Aneja R and Cavalli LR: MiR-150-5p overexpression in triple-negative breast cancer contributes to the *in vitro* aggressiveness of this breast cancer subtype. *Cancers (Basel)* 14(9): 2156, 2022. PMID: 35565284. DOI: 10.3390/cancers14092156
 - 40 Ali S: Transcriptional coactivators and corepressors in endocrine response and resistance in breast cancer. *Therapeutic Resistance to Anti-Hormonal Drugs in Breast Cancer*: 27-38, 2021. DOI: 10.1007/978-1-4020-8526-0_2
 - 41 Abramson VG and Mayer IA: Molecular heterogeneity of triple negative breast cancer. *Curr Breast Cancer Rep* 6(3): 154-158, 2014. PMID: 25419441. DOI: 10.1007/s12609-014-0152-1
 - 42 de Ronde JJ, Hannemann J, Halfwerk H, Mulder L, Straver ME, Vrancken Peeters MJ, Wesseling J, van de Vijver M, Wessels LF and Rodenhuis S: Concordance of clinical and molecular breast cancer subtyping in the context of preoperative chemotherapy response. *Breast Cancer Res Treat* 119(1): 119-126, 2010. PMID: 19669409. DOI: 10.1007/s10549-009-0499-6

Received June 14, 2022

Revised July 8, 2022

Accepted July 13, 2022