An insight into the promoter methylation of *PHF20L1* and the gene association with metastasis in breast cancer

Jose Alfredo Sierra-Ramirez[1,B,F], Emmanuel Seseña-Mendez[2,E,F], Marycarmen Godinez-Victoria[1,B,F], Marta Elena Hernandez-Caballero[2,A,C–E]

[1] Medical School, Graduate Section, National Polytechnic Institute, Mexico City, Mexico
[2] Department of Biomedicine, Faculty of Medicine, Meritorious Autonomous University of Puebla (BUAP), Puebla, Mexico

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation; D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online) *Adv Clin Exp Med*. 2021;30(5):507–515

Address for correspondence
Marta Elena Hernandez-Caballero
E-mail: ehdezc@yahoo.com

Funding sources
This work was supported by the Instituto Politécnico Nacional (SIP20195078) and Benemérita Universidad Autónoma de Puebla PRODEP (BUAP-CA-159).

Conflict of interest
None declared

Acknowledgements
We thank Efrain Atenco for assistance with R software.

Received on September 15, 2020
Reviewed on February 12, 2021
Accepted on February 17, 2021

Published online on April 13, 2021

Cite as
Sierra-Ramirez JA, Seseña-Mendez E, Godinez-Victoria M, Hernandez-Caballero ME. An insight into the promoter methylation of *PHF20L1* and the gene association with metastasis in breast cancer. *Adv Clin Exp Med*. 2021;30(5):507–515. doi:10.17219/acem/133426

DOI
10.17219/acem/133426

Abstract

Background. Plant homeodomain finger protein 20-like 1 (*PHF20L1*) is a protein reader involved in epigenetic regulation that binds monomethyl-lysine. An oncogenic function has been attributed to *PHF20L1* but its role in breast cancer (BC) is not clear.

Objectives. To explore *PHF20L1* promoter methylation and comprehensive bioinformatics analysis to improve understanding of the role of *PHF20L1* in BC.

Materials and methods. Seventy-four BC samples and 16 control samples were converted using sodium bisulfite treatment and analyzed with methylation-specific polymerase chain reaction (PCR). Bioinformatic analysis was performed in the BC dataset using The Cancer Genome Atlas (TCGA) trough data visualized and interpreted in the MEXPRESS website. Methylation, gene expression and survival evaluation were performed with R v. 4.0.2 software. Using multiple bioinformatic tools, we conducted a search for genes co-expressed with *PHF20L1*, analyzed its ontology and predicted associated miRNAs and miRNA-*PHF20L1* networks. The expression and prognostic value of *PHF20L1* and co-expressed genes were analyzed.

Results. We found demethylation in *PHF20L1* promoter in both BC samples and healthy tissues. Data mining with 241 patients demonstrated changes in methylation of promoter regions in basal-like and luminal A subtypes. Expression of the *PHF20L1* gene had a negative correlation with methylation. Twelve genes were co-expressed. *PHF20L1* is a target of miR96-5p, miR9-5p and miR182-5p, which are involved in proliferation and metastasis. *PHF20L1* gene expression was not associated with overall survival (OS), or relapse-free survival (RFS), but was associated with distant metastasis-free survival (DMFS).

Conclusions. Our findings showed differences in methylation of *PHF20L1* promoter region near TSS and upstream in BC subtypes; its overexpression impacted DMFS. We found that *PHF20L1* is targeted by miR96-5p, miR9-5p and miR182-5p, which are involved in proliferation and metastasis, and regulates genes engaged in processes such as alternative splicing.

Key words: metastasis, hypomethylation, *PHF20L1*, miRNA, promoter

Background

DNA mutations are the driving force for cancer initiation, progression and invasion. Nevertheless, accumulating evidence suggests that epigenetic modifications are also involved. In tumors at early stages, it is common to observe hypomethylation of DNA from tumor cells and hypermethylation of CpG islands of specific promoters, which has led to the suggestion that epigenetic dysregulation actually precedes tumor events before classic mutations.[1] Histone acetylation is an essential key for epigenetic regulation. Histone acetyltransferases (HATs) are responsible for transferring an acetyl group from acetyl-Co-A to the ε-amino of histone lysine residues.[2] They do not work in isolation but as part of a complex whose components are responsible for determining the lysine specificity.[3] In vivo, HATs require coactivators that determine which lysine will be acetylated and play a key role in a variety of cellular functions thanks to their various domains.[4,5] *PHF20L1* is part of the non-specific lethal (NLS) complex involved in histone acetylation and post-translational modification.[6] Located in the nucleoplasm and plasma membrane, *PHF20L1* has Tudor, MBT, Lys-rich, and zinc finger plant homeodomain (PHD) type domains (Uniprot KB-A8MW92). It is similar to the PHF20 homolog,[7] with which it maintains 33% homology, especially in the second PHD domain of PHF20, which shares 73% identity. Currently, its role and regulation are being revealed. It participates in avoiding SOX2 proteolysis[8] and regulates the degradation of methylated DNMT1.[9] *PHF20L1* is considered an oncogene[10] and has an important function in breast cancer (BC), which suggests that *PHF20L1* may have a role in cancer treatment.[5] MicroRNAs (miRNAs) are short noncoding RNA that regulate the expression of target genes and are associated with tumorigenesis, invasion and metastasis. An miRNA can regulate multiple genes that participate in the same biological pathway.[11]

The availability of cancer multi-omics databases allows us to decipher the genomic drivers of cancer, and the emergence of user-friendly tools to analyze and visualize a bulk of data is crucial to achieve the full potential of these datasets. In this study, we examined the *PHF20L1* promoter methylation through sodium bisulfite treatment and its participation in BC by analyzing expression in public gene datasets.

Objectives

We investigated PHF20L1 methylation and gene expression with an emphasis on its relationship with co-expressed genes, its contribution to survival (independently and with co-expressed genes), miRNAs that target it, and its involvement in cancer.

Materials and methods

Study subjects

Tissues from 80 confirmed BC cases and 16 healthy adjacent fresh tissue controls were collected in Hospital la Raza (Mexico City, Mexico). The study was approved by institutional ethical committees for research La Raza Hospital, Mexican Social Security Institute (IMSS), Mexico City, Mexico, and informed consent was obtained from all patients. All clinical data were collected from medical records. The state of disease was obtained based on pathological report.

DNA extraction

DNA from tissues was obtained with QIAamp DNA Micro Kit (Qiagen, Valencia USA). DNA concentration was measured using NanoDrop 8000 Spectrophotometer (Thermo Fisher Scientific, Waltham, USA).

Sodium bisulfite treatment and methylation-specific PCR (MSP)

DNA isolated from tissues was bisulfite-modified using EpiTect Bisulfite Kit (Qiagen, Frederick, USA) according to the manufacturer's protocol as previously described.[12] The CpG island from the promoter region was located using Eukaryotic Promoter Database tool (https://epd.epfl.ch/index.php). MSP primer pairs were designed using Methprimer software[13] to detect bisulfite-induced changes affecting unmethylated (U) and methylated (M) alleles. Primer sequences are as follows: *PHF20L1* (MF) 5'-TTAAGAATAATAAATAATGTTTTTCGT-3'; (MR) 5'-GTAACTCACGAAAATTAAACCCG-3'; (UF) 5'-AAGAATAATAAATAATGTTTTTTGT-3'; (UR) 5'-ATAACTCACAAAAATTAAACCCAAA-3'. The size of methylated polymerase chain reaction (PCR) products was 204 bp for methylated and 203 bp for unmethylated amplicon in *PHF20L1*. PCR for bisulfite-converted DNA was performed using EpiTect MSP Kit (Qiagen). Twenty nanograms of DNA, 10 μM of each primer and 2X Master mix MSP (Qiagen, Valencia USA) were combined in a final reaction volume of 10 μL. For methylated *PHF20L1*, cycle conditions were as follows: 95°C for 10 min, 1 cycle; 35 cycles (94°C for 15 s, 52°C for 30 s, 72°C for 30 s); and 72°C for 10 min, 1 cycle. For unmethylated *PHF20L1*, cycle conditions were as follows: 95°C for 10 min, 1 cycle; 35 cycles (94°C for 15 s, 50°C for 30 s, 72°C for 20 s); and 72°C for 10 min, 1 cycle. Each PCR assay included a methylation control, an unmethylated control and genomic DNA (EpiTect PCR Control DNA Set, Qiagen, USA). The PCR products were analyzed using 3.5% agarose gel electrophoresis.

Bioinformatic analysis of data in breast invasive carcinoma

We assessed the gene expression and methylation of *PHF20L1* in breast invasive carcinoma dataset using The Cancer Genome Atlas (TCGA) database (http://tcgaportal.org). Data were visualized and interpreted using MEXPRESS (https://www.mexpress.be/).

Methylation and RNA analysis

Methylation data of 561 BC samples were obtained through MEXPRESS for methylation assay.[14] MEXPRESS is an online user-friendly tool for the visualization and interpretation of TCGA data to assess expression, DNA methylation, and clinical data, as well as the relationships among them.[15] TCGA database was used for analyses of mRNA expression with R v. 4.0.2 (www.r-project.org).[14] Mean and standard deviation of parameters were used as descriptive statistics. Because data did not show normal distribution, a generalized linear model (GLM) of gamma distribution error was used (test analogous to a one-way analysis of variance (ANOVA)) in addition to Kruskal–Wallis analysis. Afterward, Tukey and Dunnet tests were performed with the multcomp[16] and FSA libraries.[17] All graphs were made with ggplot2[18] and ggsignif libraries.[19] An α value of 0.05 was used.

Co-expression analysis

Using ONCOMINE (https://www.oncomine.org), we analyzed pair-wise gene expression correlation (correlation ≥0.60). Database for Annotation, Visualization and Integrated Discovery (DAVID) v. 6.8 was used to perform gene ontology (GO) function analysis of co-expressed genes. In the GO analysis, the categories included were cellular component (CC) and molecular function (MF).

miRNAs analysis

miRNAs associated with *PHF20L1* were predicted using miRDB (http://mirdb.org/index.html)[20] and mirDIP (http://ophid.utoronto.ca/mirDIP/),[21] an online tool that provides 152 million human miRNA-target interactions. Our search was limited to high confidence (integrated score ≥0.90). Furthermore, to obtain a miRNA-*PHF20L1* network with co-expressed genes, we used miRNet (https://www.mirnet.ca.).[22]

Survival

We used R v. 4.0.2 software to determine survival probability with *PHF20L1* gene expression levels between TCGA BC samples. The patients samples were divided in 2 cohorts according to an expression cutoff of 3 (obtained using median value) and analized using the R package named survival.[23] Methylation compared to survival was evaluated using data only from cg with significant results in shorter-survival patients. To evaluate the global prognostic value of *PHF20L1* co-expressed genes, we used Kaplan–Meier plotter (http://kmplot.com), an online database of microarray datasets that assesses the effect of genes on survival in 5143 breast samples among other cancers,[24] and calculates hazard ratio (HR) with 95% confidence intervals (95% CIs) and log-rank p-values. Survival analyses play a central role in identifying potential genes as key genes and biomarkers.

Results

Methylation analysis

The promoter methylation status of the *PHF20L1* gene was examined in 74 sporadic BC tumors (6 patients did not have complete data) and 16 non-tumoral adjacent tissues from some of these same patients. The mean age of the patients was 54.1 ±11 years. The BC stage frequencies were as follows: stage II 39.0%, stage III 45.5% and stage IV 15.6%. The methylation assay using sodium bisulfite revealed no difference in *PHF20L1* promoter methylation status between cancer stages or in comparison to healthy tissues; all samples were demethylated (Fig. 1A). Through data mining of *PHF20L1* gene in the TCGA database with MEXPRESS tool, we analyzed DNA methylation (Infinium HumanMethylation450 microarray) in 241 patients with complete data. Eight promoter probes were analyzed. There was no significant difference regarding tumor stage (stage I 18%, stage II 58.5%, stage III 21.2%, and stage IV 2.3%) or histology type (data from 824 patients were used for this analysis). When we looked for differences between methylation and PAM50 BC molecular classification, we found differences with 2 probes: cg5307234 and cg27342122, of which cg27342122 corresponded to the region we analyzed (Fig. 1B). The luminal A subtype had more methylation and the basal-like subtype had less methylation in promoter region cg5307234 probe, while, the basal-like subtype showed hypomethylation in the promoter region of cg27342122 compared to the other subtypes (Fig. 1C,D).

HF20L1 gene expression

Due to the minimum amount of tissue for our analysis, it was not possible to obtain RNA. The *PHF20L1* gene expression was checked in TCGA BC samples that had subtype information available. We found that the mRNA expression level of *PHF20L1* was higher in the basal-like subtype (Fig. 2A). In addition, a negative correlation between DNA methylation and *PHF20L1* transcription was also observed ($r = -0.19$, $p < 0.001$) (Fig. 2B).

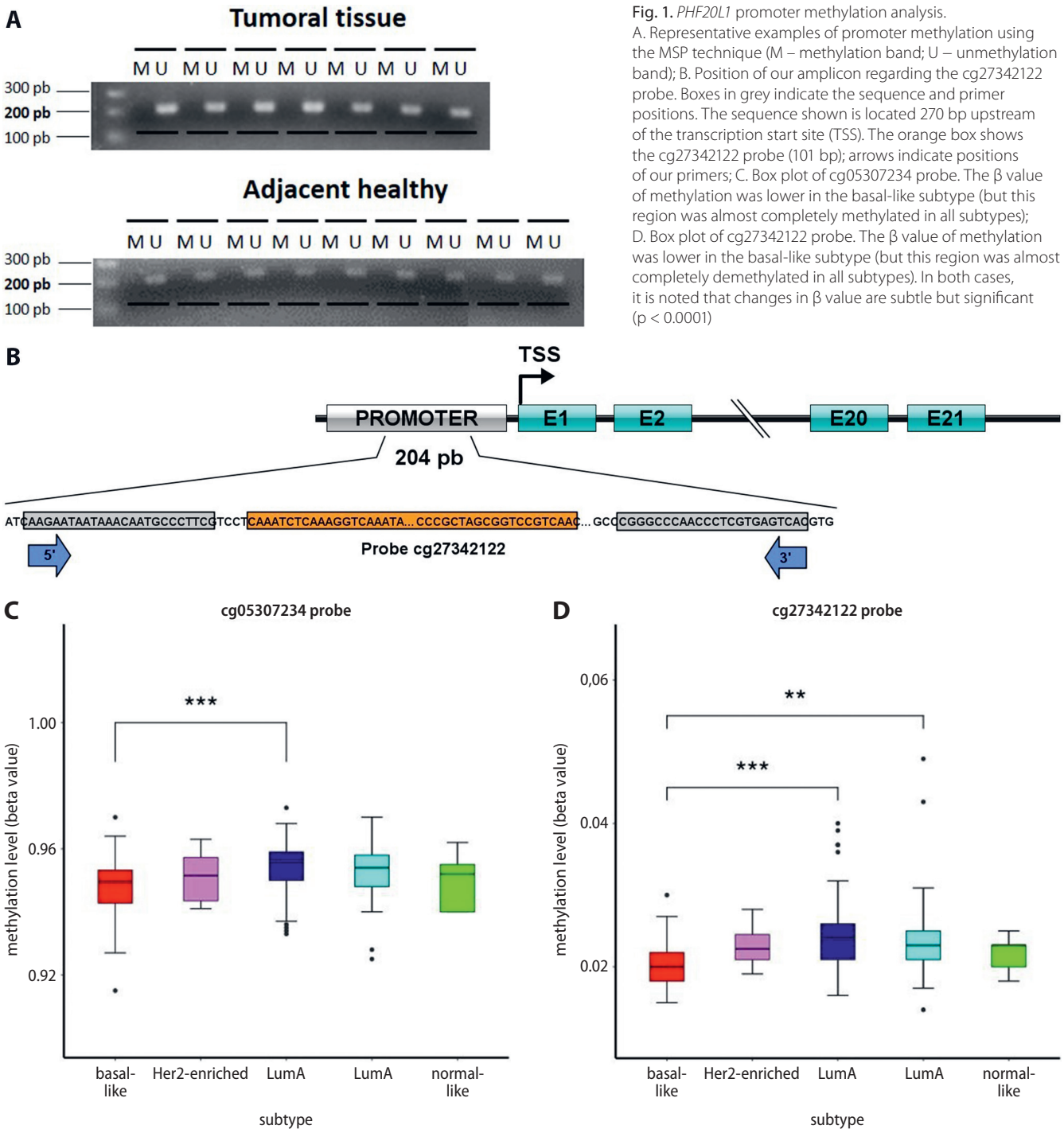

Fig. 1. *PHF20L1* promoter methylation analysis. A. Representative examples of promoter methylation using the MSP technique (M – methylation band; U – unmethylation band); B. Position of our amplicon regarding the cg27342122 probe. Boxes in grey indicate the sequence and primer positions. The sequence shown is located 270 bp upstream of the transcription start site (TSS). The orange box shows the cg27342122 probe (101 bp); arrows indicate positions of our primers; C. Box plot of cg05307234 probe. The β value of methylation was lower in the basal-like subtype (but this region was almost completely methylated in all subtypes); D. Box plot of cg27342122 probe. The β value of methylation was lower in the basal-like subtype (but this region was almost completely demethylated in all subtypes). In both cases, it is noted that changes in β value are subtle but significant ($p < 0.0001$)

Co-expression analysis

Using the ONCOMINE database, we selected co-expression analysis in BC primary sites, using only female mRNA data. We found 12 genes co-expressed with *PHF20L1* across the BC dataset with a correlation value of ≥0.6. Co-expressed genes were clustered through gene ontology analysis using DAVID. The enrichment GO terms considered were CC and MF ontologies (Table 1). In the CC ontology, we obtained 3 GO categories involved with nucleus (7 genes), nucleoplasm (5 genes) and coiled coil (5 genes). The other enrichment category MF comprised items related to alternative splicing (10 genes), splice variant (9 genes) and phosphoprotein (9 genes). Only clusters that had at least 5 genes were included.

miRNAs analysis

miRNAs have a role in post-transcriptional regulation of gene expression which leads to targeting of mRNAs for degradation and/or inhibition of translation. Furthermore, one miRNA can co-regulate several genes. Using mirDIP

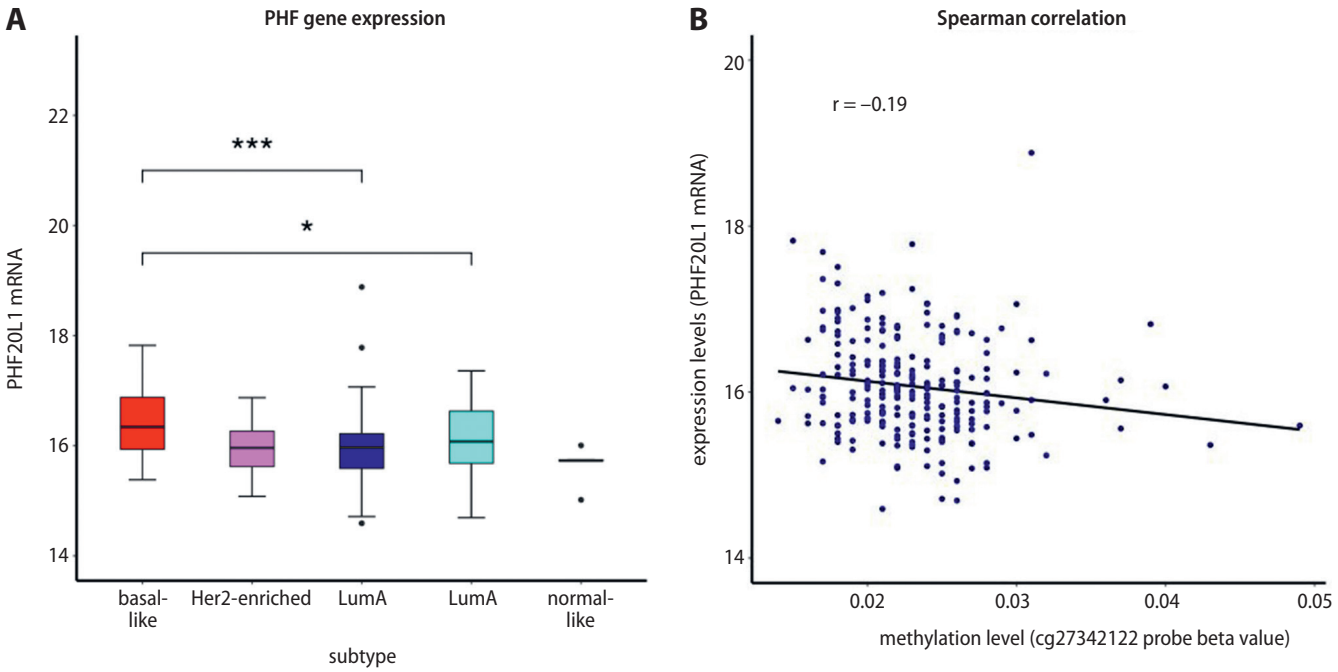

Fig. 2. Relationship between *PHF20L1* expression in BC and methylation. A. Box plot of mRNA expression based on the TCGA database ($p < 0.0001$); B. Correlation between cg27342122 methylation and *PHF20L1* expression. Spearman test showed a negative correlation between DNA methylation and *PHF20L1* transcription ($r = -0.19$, $p < 0.001$)

Table 1. Gene ontology analysis of gene expression. Functional annotation clustering (enrichment score 1.36–0.47)

Gene co-expressed	Category	Term/gene function	Count	%	p-value
N4BP2L2, PHF20L1, PNISR, YLPM1, ATG12, MFF, PON2, PRKRA, PIAS1, ZNF407	GOTERM_MF_DIRECT	alternative splicing	10	83.3	4.2E-2
N4BP2L2, PHF20L1, PNISR, YLPM1, ATG12, MFF, PON2, PRKRA, ZNF407	GOTERM_MF_DIRECT	splice variant	9	75	2.4E-2
PHF20L1, PNISR, STT3B, YLPM1, MFF, PRKRA, PIAS1, ZNF407, SMC5	GOTERM_MF_DIRECT	phosphoprotein	9	75	3.0E-2
N4BP2L2, PHF20L1, YLPM1, PON2, PIAS1, ZNF407, SMC5	GOTERM_CC_DIRECT	nucleus	7	58.3	7.3E-2
PNISR, YLPM1, PRKRA, PIAS1, SMC5	GOTERM_CC_DIRECT	nucleoplasm	5	41.7	7.3E-2
N4BP2L2, PNISR, YLPM1 MFF, SMC5	GOTERM_CC_DIRECT	coiled coil	5	41.7	6.6E-2

CC – cellular component; MF – molecular function.

(corroborated by miRDB), 190 miRNAs were predicted as regulators of *PHF20L1* gene expression, but only miR-96-5p, miR-9-5p and miR-182-5p were obtained using a very high score class for prediction (score ≥0.90). After searching in ONCOMINE, we found 12 genes co-expressed with *PHF20L1*. Next, we made a network with the miRNet tool in BC tissues, setting the cutoff degree on 1. We found 1 node with 442 miRNAs; this network included miRNAs predicted by mirDIP, including 79 miRNAs involved in breast neoplasms and triple-negative breast carcinoma (Fig. 3). Mir-network allowed us to observe that miRNAs predicted as *PHF20L1* regulators participate in cell differentiation, cell cycle and apoptosis ($p = 0.006$), and mir9-5p and mir-182-5p are involved in triple-negative breast carcinoma BC ($p = 0.02$), and mir-96-5p in breast neoplasms ($p = 0.02$).

PHF20L1 prognosis in breast cancer

The prognostic value of *PHF20L1* expression was examined with R software. The expression of *PHF20L1* or the promoter methylation in cg5307234 and cg27342122 probe regions by subtype had no relation with overall survival (OS), but when we analyzed the methylation data from deceased patients alone, i.e., patients with shorter survival, we found that more hypomethylation of *PHF20L1* was observed in the basal-like subtype, with respect to the luminal A subtype, and methylation between the luminal A and luminal B subtypes ($p < 0.01$) was associated with survival (Fig. 4A). The Kaplan–Meier plotter platform was used to analyzed relapse-free survival (RFS) and distant metastasis-free survival (DMFS), which revealed that expression of *PHF20L1* is related with DMFS ($p = 0.02$) (Fig. 4B). Exploring the potential roles of genes co-expressed with *PHF20L1* in OS, RFS and DMFS, we obtained Kaplan–Meier survival curves from the Kaplan–Meier plotter platform. We found that expression of *ZNF407* and *PIAS1* were related with OS. Expression of all genes except *PHF20L1*, *STT3B*, *YLPM1*, and *MFF* was associated with RFS of BC, and *STT3B*, *PRKRA*, *ATG12*, and *PHF20L1* were associated with DMFS (Table 2).

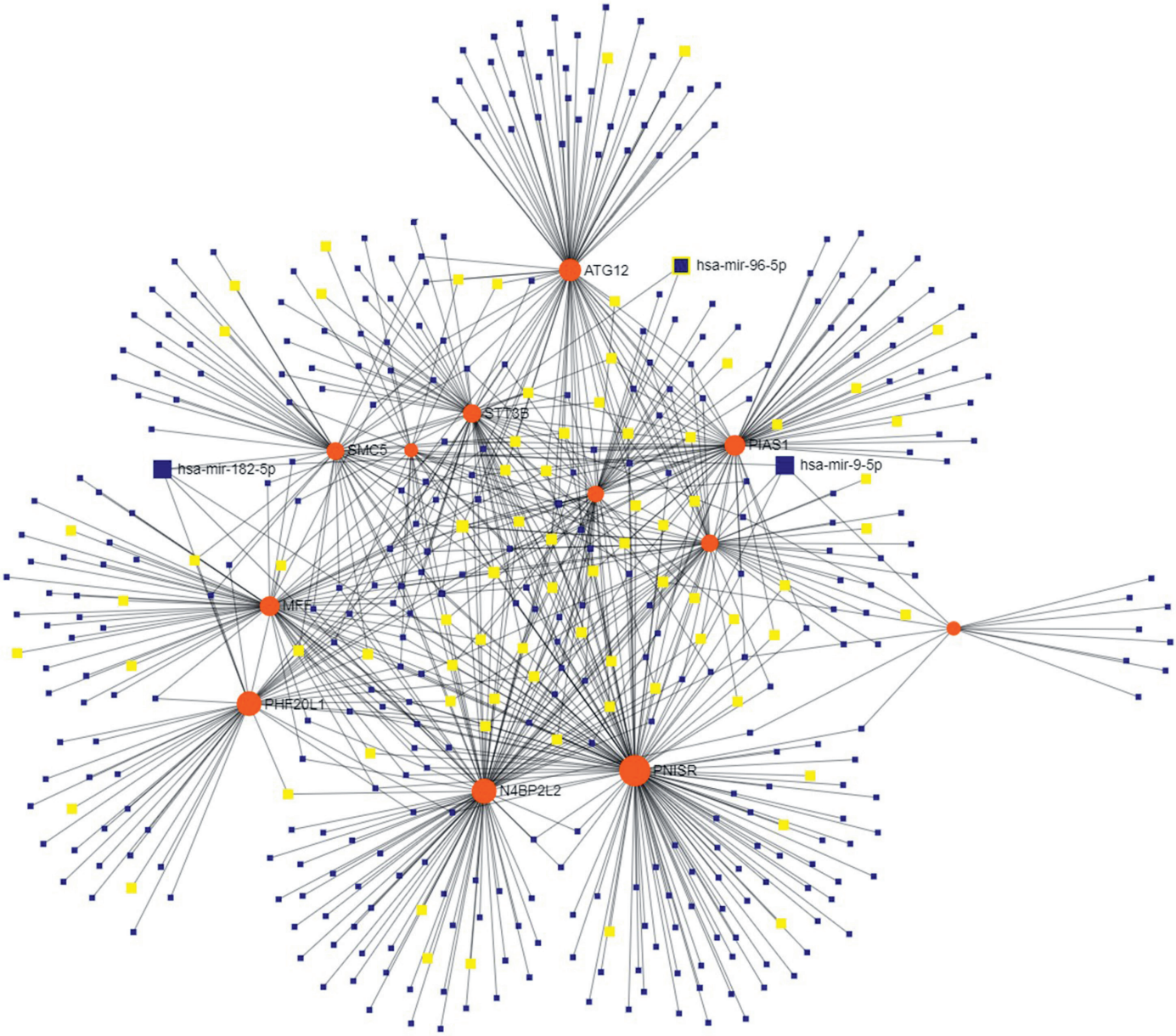

Fig. 3. Interaction miRNet of *PHF20L1* and co-expressed genes. The interactions among all co-expressed genes and associated miRNAs are shown. The red circles represent genes; the blue squares represent miRNAs. miRNAs predicted to regulate *PHF20L1* are larger squares and the yellow ones represent the miRNAs implicated in breast neoplasms (33 genes), triple-negative breast carcinoma (32 genes) or both (14 genes)

Discussion

Epigenetic readers contain diverse methyl–lysine binding motifs including PHD, chromo, Tudor, MBT, PWWP, Ank, BAH, WD40, ADD, and zn-CW domains.[25] *PHF20L1* is a reader that interacts with mono- and dimethylated lysine residues in *H3K4*, *H4K20*, *H3K27*, and *DNMT1*, due to its Tudor and PHD domains, and histone *H4K16* acetylation, due to its MYST (MOF) domain.[5,26] For example, *PHF20L1* is recruited to E2F-responsive promoters through pRb mono-methylated K810, which suggests that *PHF20L1* could participate in cell cycle progression mediating transcriptional repression.[27] *PHF20L1* is overexpressed in the aggressive subtypes basal-like and luminal B, which have been strongly associated with shorter survival in patients with BC.[10] Thus, this gene could be a critical tethering factor regulating molecular mechanisms through methylation signals on both DNA and histones.[10] A growing body of evidence supports the epigenetic reprogramming of cancer cells as a key step in breast carcinogenesis. Teschendorff et al. found that genomic distribution of methylation is not random and is strongly enriched for binding sites of transcription factors specifying chromatin architecture.[28] We found differential methylation in *PHF20L1* promoter in the molecular subtypes basal-like and luminal A in 2 regions: -708 bp to TSS (cg5307234 probe) and -242 bp (cg27342122 probe). Basal-like, luminal A and luminal B subtypes have significant differences

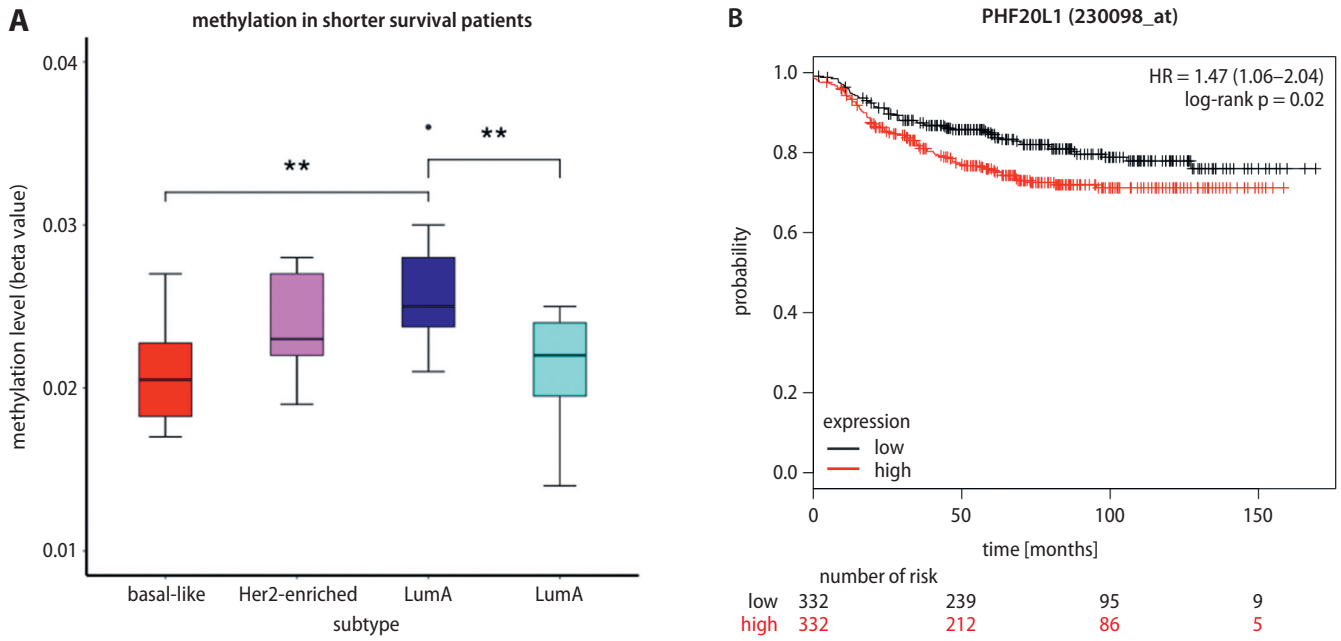

Fig. 4. Prognostic value of *PHF20L1* expression in survival. A. The relationship between mRNA expression of *PHF20L1* and survival by BC subtype (normal-like subtype was not included because it only included 1 patient); B. Kaplan–Meier distant metastasis-free survival analyses for *PHF20L1* expression in BC patients in Kaplan–Meier plotter. Value of $p < 0.05$, p-value was obtained from the log-rank test. See Table 2 for details of overall survival and relapse-free survival

Table 2. Prognostic value of genes co-expressed with PHF20L1 according to ONCOMINE

Gene	Location	Score*	OS p-value	RFS p-value	DMSF p-value
PHF20L1	8q24.22	–	0.24	0.88	0.02
STT3B	3p23	0.740	0.08	0.11	0.014
ZNF407	18q22.3	0.708	0.055	0.00042	0.073
PIAS1	15q23	0.654	0.008	0.002	0.18
YLPM1	14q24.3	0.654	0.09	0.081	0.94
MFF	2q36.3	0.654	0.33	0.59	0.18
PNISR	6q16.2	0.654	0.1	0.00053	0.37
PON2	7q21.3	0.654	0.2	0.021	0.077
PRKRA	2q31.2	0.654	0.95	0.026	0.042
N4BP2L2	13q13.1	0.654	0.22	0.0001	1
SMC5	9q21.12	0.631	0.9	0.0046	0.87
ATG12	5q22.3	0.631	0.62	2.9–E10	0.003

* score correlation ≥0.6; OS – overall survival; RSF – relapse free survival; DMSF – distant metastasis free survival.

in methylation in the promoter region. The methylation pattern was different inasmuch as region -708 to TSS was nearly methylated with a β value mean of 0.95 for luminal A and B subtypes, while in the -242 region both subtypes and basal-like were hypomethylated (β value of 0.023 and 0.020, respectively). Usually, promoters have sites for transcription factor binding. With aid from the TF2DNA database (http://fiserlab.org/tf2dna_db//index.html) and JASPAR CORE 2020 (http://jaspar.genereg.net/), we found that transcription factor EC (TFEC) binds to the cg5307234 region (-708 bp) and participates in regulating multiple cellular processes including survival, growth and differentiation.[29,30] The cg27342122 region (-242 bp) has sites for binding of *GATA3*, *FOXP2* and *FOXP3* transcription factors. Of these, the transcription factor *GATA3* is relevant for its role in determination of cell identity. *GATA3* is expressed in mammary glands in the differentiated luminal epithelial cells.[31] So, differences in methylation pattern may affect the binding of transcription factors deregulating *PHF20L1* expression.

In our analysis, we found that *PHF20L1* overexpression was not related to OS in the analysis by cancer subtype except when the comparison was made only with patients with shorter survival. Similarly, hypomethylation was correlated with survival in these patients in the basal-like and luminal A and B subtypes. The luminal B subtype is distinguished

by a higher proliferative activity than luminal A and worse prognosis.[32] When we analyzed survival without grouping by subtypes and including the co-expressed genes, we found that overexpression of *PHF20L1*, *STT3B*, *PRKRA*, and *ATG12* was related to DMFS. We found that all genes overexpressed except *STT3B*, *YLPM1*, *MFF*, and *PHF20L1* are related to RSF. Interestingly, *PHF20L1* and many co-expressed genes are involved in key processes such as alternative splicing.

miRNAs have an important role in cellular regulation, and we found 3 miRNAs with a high probability to regulate *PHF20L1*: miR96-5p, miR9-5p and miR182-5p. miR96-5p may participate in epithelial–mesenchymal transition[33]; using miRNet, we found that this miRNA is involved in breast neoplasms. miR9-5p could enhance cancer stem cell-like traits of BC, but its role depends on the stage of BC, i.e., it could inhibit cell proliferation (tumor suppressor activity) or play an oncogenic role in metastasis.[34] On the other hand, miR182-5p is a key oncogenic miRNA that promotes cell proliferation and metastasis[35] and could be involved in epigenetic changes.[36] Therefore, these miRNAs might have an important role in methylation and expression changes of the *PHF20L1* gene contributing to its role in BC metastasis.

PHF20L1 is established as an important epigenetic reader whose loss could induce genome hypomethylation. For us, the use of public databases and bioinformatics tools was crucial to obtain a better picture of *PHF20L1* interactions particularly with miRNAs, which in turn are involved in a complex regulatory network affecting transcription.

Limitations

Our study has limitations. First, validation should be carried out both in vitro and in vivo to determine the clinical usefulness in patients with metastatic disease. The second limitation is the modest sample size for some analyses and the difference in our experimental approach to methylation analysis.

Conclusions

Our findings indicate that changes in methylation near TSS of the *PHF20L1* gene may influence its expression in BC subtypes and that *PHF20L1* gene overexpression affects distant metastasis-free survival in BC. Furthermore, the study suggests that miR96-5p, miR9-5p and miR182-5p target and regulate to *PHF20L1*. These results support participation of *PHF20L1* in the metastasis process.

Data availability

The database data supporting this research article is from previously reported studies and datasets, which have been cited. The data used to analyze with R software is available at the MEXPRESS website.

ORCID iDs

Jose Alfredo Sierra-Ramirez https://orcid.org/0000-0001-5060-5882
Emmanuel Seseña-Mendez https://orcid.org/0000-0002-9255-0231
Marycarmen Godinez-Victoria https://orcid.org/0000-0003-3317-7980
Marta Elena Hernandez-Caballero https://orcid.org/0000-0003-4174-2118

References

1. Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. *Epigenomics*. 2015;7(2):247–265. doi:10.2217/epi.14.80
2. Kim Y, Tanner KG, Denu JM. A continuous, nonradioactive assay for histone acetyltransferases. *Anal Biochem*. 2000;280(2):308–314. doi:10.1006/abio.2000.4546
3. Carrozza MJ, Utley RT, Workman JL, Côté J. The diverse functions of histone acetyltransferase complexes. *Trends Genet*. 2003;19(6):321–329. doi:10.1016/S0168-9525(03)00115-X
4. Lee KK, Workman JL. Histone acetyltransferase complexes: One size doesn't fit all. *Nat Rev Mol Cell Biol*. 2007;8(4):284–295. doi:10.1038/nrm2145
5. Hou Y, Liu W, Yi X, et al. PHF20L1 as a H3K27me2 reader coordinates with transcriptional repressors to promote breast tumorigenesis. *Sci Adv*. 2020;6(16):eaaz0356. doi:10.1126/sciadv.aaz0356
6. Mendjan S, Taipale M, Kind J, et al. Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol Cell*. 2006;21(6):811–823. doi:10.1016/j.molcel.2006.02.007
7. Cui G, Park S, Badeaux AI, et al. PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nat Struct Mol Biol*. 2012;19(9):916–924. doi:10.1038/nsmb.2353
8. Wang Q, Yu M, Ma Y, et al. PHF20L1 antagonizes SOX2 proteolysis triggered by the MLL1/WDR5 complexes. *Lab Invest*. 2018;98(12):1627–1641. doi:10.1038/s41374-018-0106-8
9. Estève PO, Terragni J, Deepti K, et al. Methyllysine reader plant homeodomain (PHD) finger protein 20-like 1 (PHF20L1) antagonizes DNA (cytosine-5) methyltransferase 1 (DNMT1) proteasomal degradation. *J Biol Chem*. 2014;289(12):8277–8287. doi:10.1074/jbc.M113.525279
10. Jiang Y, Liu L, Shan W, Yang ZQ. An integrated genomic analysis of Tudor domain-containing proteins identifies PHD finger protein 20-like 1 (PHF20L1) as a candidate oncogene in breast cancer. *Mol Oncol*. 2016;10(2):292–302. doi:10.1016/j.molonc.2015.10.013
11. Lin S, Gregory RI. MicroRNA biogenesis pathways in cancer. *Nat Rev Cancer*. 2015;15(6):321–333. doi:10.1038/nrc3932
12. Borgonio-Cuadra VM, Miranda-Duarte A, Rojas-Toledo X, et al. Association between promoter hypermethylation of the DACT2 gene and tumor stages in breast cancer. *J BUON*. 2018;23(2):361–365. PMID: 29745077
13. Li LC, Dahiya R. MethPrimer: Designing primers for methylation PCRs. *Bioinformatics*. 2002;18(11):1427–1431. doi:10.1093/bioinformatics/18.11.1427
14. R Core Team. R: A language and environment for statistical computing. https://www.r-project.org/. Accessed April 1, 2019.
15. Koch A, De Meyer T, Jeschke J, Van Criekinge W. MEXPRESS: Visualizing expression, DNA methylation and clinical TCGA data. *BMC Genomics*. 2015;16(1):636. doi:10.1186/s12864-015-1847-z
16. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. *Biometrical J*. 2008;50(3):346–363. doi:10.1002/bimj.200810425
17. Ogle DH, Wheeler P, Dinno A. FSA: Fisheries Stock Analysis. R package v. 0.8.25. 2019. https://github.com/droglenc/FSA. Accessed April 1, 2019.
18. Wickham H. *Ggplot2 Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016. doi:10.1007/978-0-387-98141-3
19. Ahlmann-Eltze C. Package 'ggstatsplot'. https://www.rdocumentation.org/packages/ggsignif/versions/0.6.1. Accessed April 1, 2019.
20. Yuhao Chen XW. miRDB: An online database for prediction of functional microRNA targets. *Nucleid Acid Res*. 2020;48(D1):D127–D131. doi:10.1093/nar/gkz757
21. Tokar T, Pastrello C, Rossos AEM, et al. MirDIP 4.1: Integrative database of human microRNA target predictions. *Nucleic Acids Res*. 2018;46(D1):D360–D370. doi:10.1093/nar/gkx1144

22. Chang L, Zhou G, Soufan O, Xia J. miRNet 2.0: Network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Res*. 2020;48(W1):W244–W251. doi:10.1093/nar/gkaa467
23. Therneau T. A Package for Survival Analysis in R. R package v. 3.2-7. https://CRAN.R-project.org/package=survival.
24. Győrffy B, Lanczky A, Eklund AC, et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res Treat*. 2010;123(3):725–731. doi:10.1007/s10549-009-0674-9
25. Musselman CA, Khorasanizadeh S, Kutateladze TG. Towards understanding methyllysine readout. *Biochim Biophys Acta Gene Regul Mech*. 2014;1839(8):686–693. doi:10.1016/j.bbagrm.2014.04.001
26. Avvakumov N, Côté J. The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene*. 2007;26(37):5395–5407. doi:10.1038/sj.onc.1210608
27. Carr SM, Munro S, Sagum CA, Fedorov O, Bedford MT, La Thangue NB. Tudor-domain protein PHF20L1 reads lysine methylated retinoblastoma tumour suppressor protein. *Cell Death Differ*. 2017;24(12):2139–2149. doi:10.1038/cdd.2017.135
28. Teschendorff AE, Gao Y, Jones A, et al. DNA methylation outliers in normal breast tissue identify field defects that are enriched in cancer. *Nat Commun*. 2016;7:10478. doi:10.1038/ncomms10478
29. Zhao GQ, Zhao Q, Zhou X, Mattei MG, de Crombrugghe B. TFEC, a basic helix-loop-helix protein, forms heterodimers with TFE3 and inhibits TFE3-dependent transcription activation. *Mol Cell Biol*. 1993; 13(8):4505–4512. doi:10.1128/mcb.13.8.4505
30. Haq R, Fisher DE. Biology and clinical relevance of the micropthalmia family of transcription factors in human cancer. *J Clin Oncol*. 2011; 29(25):3474–3482. doi:10.1200/JCO.2010.32.6223
31. Asselin-Labat ML, Sutherland KD, Barker H, et al. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol*. 2007;9(2):201–209. doi:10.1038/ncb1530
32. Prat A, Pineda E, Adamo B, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. *Breast*. 2015;24(Suppl 2):S26–S35. doi:10.1016/j.breast.2015.07.008
33. Anderson O, Guttilla Reed IK. Regulation of cell growth and migration by miR-96 and miR-183 in a breast cancer model of epithelial-mesenchymal transition. *PLoS One*. 2020;15(5):e0233187. doi:10.1371/journal.pone.0233187
34. Li X, Zeng Z, Wang J, et al. MicroRNA-9 and breast cancer. *Biomed Pharmacother*. 2020;122:109687. doi:10.1016/j.biopha.2019.109687
35. Shah NR, Chen H. MicroRNAs in pathogenesis of breast cancer: Implications in diagnosis and treatment. *World J Clin Oncol*. 2014;5(2):48–60. doi:10.5306/wjco.v5.i2.48
36. Li Q, Zhu F, Chen P. MiR-7 and miR-218 epigenetically control tumor suppressor genes RASSF1A and claudin-6 by targeting HoxB3 in breast cancer. *Biochem Biophys Res Commun*. 2012;424(1):28–33. doi:10.1016/j.bbrc.2012.06.028