International Journal of Education and Science Research Review
E-ISSN 2348-6457
July –August - 2021, Volume-8, Issue-4 P-ISSN 2349-1817
www.ijesrr.org Email- editor@ijesrr.org

# IDENTIFICATION OF NOVEL BIOMARKERS ASSOCIATED WITH POOR PATIENT'S OUTCOME TO DISTINGUISH LOBULAR AND DUCTAL BREAST CANCER

**Tamanna Walia**

**Research scholar, Kalinga University Naya Raipur Chhattisgarh 492101 India**

**Dr.Sushma Dubey,**

**Prof.&HOD, Department of Biotechnology, Kalinga University Naya Raipur Chhattisgarh 4902101, India**

## ABSTRACT

Breast cancer is one of the leading causes of mortality in women among cancer patients. The identification of individual molecules that are implicated in breast cancer and the understanding of their features have been the focus of a great number of researches. The identification of several biomarkers that are not only simple to measure but also trustworthy, affordable, and possess high levels of both sensitivity and specificity has been accomplished. Both the fast-accelerating development of technology and the availability of epigenetic knowledge play important roles in the prevention and treatment of cancer. A wide variety of databases have been used for the purpose of collecting, storing, and analyzing the gathered data. In order to extract meaningful and accessible data from databases, it is essential to recognize the existence of such data. In today's world, several researchers make use of databases such as The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), Surveillance, Epidemiology and End Results (SEER), and Embase in order to get valuable information on biomarkers for breast cancer. This article provides a summary of the databases that are now available and have been used for the purpose of identifying biomarkers for breast cancer. The information that is offered by this review would be helpful in the search for effective techniques for the diagnosis and treatment of breast cancer.

***Keywords:*** *Biomarkers, Breast cancer, Poor Patient's*

## INTRODUCTION

It is well acknowledged that breast cancer is among the most significant health issues that women face all over the globe. As a result of a lot of investigations, the main causes of breast cancer as well as the individual molecules that are involved in breast cancer have been identified. A number of technological advancements have been made in order to enhance the early identification of breast cancer. Patients who are afflicted with breast cancer now have access to therapeutic chemicals, whether they are synthetic or natural that is capable of efficiently inhibiting or controlling possible molecular targets. This has the potential to raise the survival rate of these patients. Despite this, death rates continue to be somewhat high, and as a result, experts are still looking for novel treatments to combat breast cancer.

In the process of creating new diagnostic and therapeutic procedures, one of the potential ways would be the identification of novel biomarkers. In order to detect illnesses and get a deeper understanding of them, several biomarkers are used. A high level of sensitivity and specificity is possessed by these, in addition to their ease of measurement, dependability, and low cost. The fact that they fluctuate with various stages of the illness and have diagnostic and predictive value makes them useful not just for screening but also for detecting potential recurrences of the disease. Many different methods have been developed in order to uncover new biomarkers that are present in a variety of illnesses. Biological information indexing and database providing is one of these methods. It assists in the discovery of biomarkers and in gaining a deeper comprehension of biological responses such as invasion, metastasis, and proliferation. There have been several technological advancements and methods used in order to discover biomarkers and to make a contribution to the characterization of illnesses. The subject of cancer biology has lately seen an increase in the amount of bioinformation. The current study provides a summary of the meta-data that is currently available, which includes The Cancer Genome Atlas (TCGA), Gene Expression Omnibus (GEO), Embase, and Surveillance and other databases.

**The Cancer Genome Atlas**

Together with the National Cancer Institute (NCI) and the National Human Genome Research Institute, the Translational Cancer Genome Atlas (TCGA) was initiated as a pilot project in the year 2006. Through the collection of high-dimensional and overall genetic alterations in 33 different forms of cancer, including breast, brain, and ovarian cancer, the Translational Cancer Genome Atlas (TCGA) aims to speed up the process of disease knowledge. Through the use of technology for genome analysis and characterization, the Translational Cancer Genome Atlas (TCGA) helps to advance cancer prevention, diagnosis, and therapy. In order to characterize the DNA and RNA of tumors, a variety of different techniques were used at the epigenetic level. The tumor and normal tissue samples from more than 11,000 individuals are stored in the TCGA database. Because it does not impose any restrictions on access, the TCGA data may be used by researchers who are successful in discovering breast cancer biomarkers.

A large number of gene targets have been looked for by researchers using the TCGA data, which has been used to discover a great deal of biomarkers in breast cancer. Among the biomarkers that have been discovered in human breast cancer, for instance, long non-coding RNAs (lncRNAs) have been found. According to the findings of a research that used one thousand instances of TCGA data, LINC00657 plays an essential part in the development and proliferation of tumor cells. In addition, the new long noncoding RNA known as FGF 14 antisense RNA2 has the potential to function as a tumor suppressor gene and decrease the progression of breast cancer. The Piwi-interacting RNAs, which are crucial in the preservation of the germline, have been identified, and PIWI proteins have the potential to serve as biomarkers for cancer. Research discovered the PIWIL3 and PIWIL4 genes by using the TCGA dataset and analyzing their relevant prognostic significance. The molecules of miRNA include a wide variety of possible biomarkers. In the field of breast cancer, miR-660-5p and miR-574-3p are examples of candidates that are associated with overall survival and recurrence-free survival rate. MiR-10b, miR-26a, miR-146a, and miR-153 are the microRNAs that are responsible for regulating the expression of BRCA1 in triple negative breast cancer.

There are several different DNA molecules that have been recognized as potential biomarkers for breast cancer. High levels of promoter methylation were shown to have a substantial association with the hormone-receptor-positive status of breast tumors, according to data obtained from the Technical Cancer Group of America (TCGA) and patients treated at the Breast Cancer Care facility in Chhattisgarh. In addition to the

fact that methylation is associated with matrix metalloproteinase (MMP)-7 productions, the Transcriptional Cancer Genome Analysis (TCGA) has shown that hypomethylation of the MMP-7 promoter is a prognostic signal. There is a possibility that a correlation exists between the methylation status of the DSC2, KCNK4, GSTM1, AXL, DNAJC15, HBII-52, TUSC3, and TES genes with a worse survival rate for breast cancer in African Americans.

Markers may be recognized by the presence of RNA-protein complexes. The Musashi RNA-binding protein 2 is identified as being linked with clinical outcomes and is located upstream of ER1. In addition to its role as a tumor suppressor, the RNA-binding protein known as tristetraprolin (TTP, ZFP36) is associated with the control of cAMP response element-binding protein activity. In cases of breast cancer, a decreased TTP was indicative of a bad prognosis.

Proteins have been investigated by researchers and analyzed in terms of their resistance to anti-cancer drugs and ligands. A decrease in T-cell proliferation and an increase in apoptosis were seen as a result of the overexpression of programmed cell death ligand 1 on the cell surface, which was caused by the loss of phosphatase and tensin homologs. Additionally, drugs that target the PI3K pathway may be able to enhance the antitumor adaptive immune response. Additionally, there is a correlation between the overexpression of pSTAT3 and the resistance to trastuzumab in HER2-positive primary breast cancer. It was discovered that the expression of KLK10 was a contributor to breast cancer that was resistant to trastuzumab. Considering that KLK10 and pSTAT3 are associated with molecules that are resistant to trastuzumab, they have the potential to serve as diagnostic markers for breast cancer. The subtype of breast cancer known as triple-negative breast cancer has a low survival rate and very strong resistance to treatment.

**Gene Expression Omnibus**

The Genome Editing Organisation (GEO) is a global data repository that facilitates the dissemination of high-throughput functional genomics data, including microarray, next-generation sequencing, and other types of data. Gene expression research on a wide variety of biological topics, such as ecology, illness, metabolism, toxicity, development, evolution, and immunology, make up about 90 percent of the data that is included in the Gene Expression Online database. Other types of functional genomic and epigenomic investigations are represented by the non-expression data in GEO. These studies include those that investigate chromatin structure, changes in genome copy number, interactions between the genome and proteins, and genome methylation. Through the use of user-friendly web-based tools, the massive amount of data can be successfully studied, queried, and visualized. At the moment, GEO maintains around one billion unique gene expression measurements, which are collected from more than one hundred different plants and animals.

GEO has been used in a number of investigations to discover previously unknown markers. It has been shown that a number of lncRNAs are effective markers. An example of this would be the discovery of four lncRNA genes via the use of the random survival forest method on expression signature. This finding suggests that lncRNAs have a role in the pathogenesis of breast cancer. A number of lncRNA signatures have the potential to serve as useful biomarkers for determining the risk of metastatic disease in breast cancer patients, therefore enhancing our comprehension of the molecular pathways involved in the invasion of breast cancer. There is a correlation between the overexpression of cancer-secreted miR-105, which is present in metastasis cancer, and the advancement of metastatic breast cancer in early-stage breast cancer. It has been shown that a correlation exists between down-regulated miR-126/miR and poor metastasis-free survival in breast cancer

patients. A marker of breast cancer and breast cancer molecular subtype survival, the DNA methylation pattern is also a sign of breast cancer as well.

For the purpose of identifying EMT and metastatic molecules, the GEO database has been used. In a number of different types of solid cancer, the expression of polyomavirus enhancer activator 3 protein (Pea3), which is a member of the Ets-transcription factor family, is increased with the development of metastatic disease. Transactivation of Snail, which is an EMT activator, is the mechanism by which overexpression of Pea3 leads to an increase in EMT in human breast epithelial cells. Comparing the levels of WNT5A and B expression in MDA-MB-231 triple-negative breast cancer cells to those in MCF-7 ER-positive breast cancer cells, it was discovered that the former were much higher. The enhancement of MCF-7 invasion was facilitated by the degradation of WNT5B by WNT and Jun-N-terminal kinase antagonists. In the same work, alternative WNT receptors ROR1 and 2 were discovered by the use of GEO. There is a strong correlation between the WNT signaling pathway and the spread of breast cancer to the brain in MDA-MB-231 cells.

**OBJECTIVES**

1. The Study Identification of Novel Biomarkers Associated.
2. The Study Patient's Outcome to Distinguish Lobular and Ductal Breast Cancer.

**RESEARCH METHODOLOGY**

In the research, there were 79 female patients who were diagnosed with infiltrating ductal breast cancer. The mean age of these patients was and their ages varied from 24 to 94 years. There was a range of 57 to 117 months in the follow-up period after the operation, with the mean being months. None of the patients had been diagnosed with any kind of cancer or metastasis at the time of diagnosis, and this was one of the criteria for inclusion in the study. Patients had treatment consisting of segmental resection or mastectomy, which included dissection of the lymph nodes in the axillary region. This was then followed by radiation and adjuvant systemic therapy, if it was deemed necessary. In the time leading up to the operation, none of them had undergone chemotherapy or radiation. Following the completion of chemotherapy, 64 patients (81%) were given radiation, and 43 patients (54%) were given tamoxifen (20 mg/day) for a period of sixty months before the conclusion of their treatment. Among the seventeen patients who acquired metastases, all of them had operable breast cancers of stage II and stage III, with bigger tumors frequently being poorly differentiated. Additionally, all of them, with the exception of one, had positive lymph nodes in the axillary region. The bulk of these individuals, thirteen in all, were given chemotherapy after undergoing surgical procedures. A total of eleven patients out of fourteen patients who had a positive ER status were also treated with hormone treatment.

Fresh tissue samples were then utilized for RNA separation and subsequent gene expression analysis. These samples were obtained by performing a macro-dissection of tumor cells. In order to conduct a global gene expression study utilizing oligoarrays, thirty-eight instances were used, and fifty-five cases were utilized for quantitative real-time PCR analysis of candidate genes. In addition, immunohistochemistry was used to analyze tumor tissue samples from 1,276 ductal breast carcinomas in order to determine the expression of proteins. Both the Civil Hospital in NayaRaipur (Chhattisgarh) and the State Cancer Centre (Chhattisgarh) provided the samples that were collected. In addition to obtaining written informed permission from each and

every patient throughout the collection period, the research was also evaluated and approved by the Ethics Committees of both institutions (CEP FHAC 340/04 and CEP ACCC 1155/08).

**DATA ANALYSIS**

Our strategy entails using a deep learning system (DLS) that is comprised of two stages for each biomarker. Individual picture patches that have been cropped to represent tiny sections of tissue are used in the first step to make predictions about the local biomarker status. It is possible that the result of this prediction will fall into one of three categories: a biomarker positive tumor, a biomarker negative tumor, or a non-tumor. In the second step of the DLS, the predictions made in the first stage are applied to each and every patch on the slide in order to make a prediction about the biomarker status at the slide level.

**Patch-level model status prediction**

Using 1.21 billion patches from 576 slides over 200 cases, the first stage of the DLS was constructed (trained and tweaked), and it was assessed on a test set consisting of all patches from 181 slides across 64 cases. Following that, we will give the classification performance of one versus all over all patches, which may be classified as biomarker positive tumor, biomarker negative tumor, or non-tumor. The patch-level area under the receiver operating characteristic curves (AUCs) for ER, PR, and HER2 were 0.939 (95% confidence interval from 0.936 to 0.941), 0.938 (95% confidence interval from 0.936 to 0.940), and 0.808 (95% confidence interval from 0.802 to 0.813), respectively. Examples of the predictions made at the patch level are shown in and Supplementary, together with the accompanying pictures obtained from immunohistochemistry. We made the observation that heterogeneous staining was, in fact, related with heterogeneous predictions. This was in contrast to the homogeneous examples, which were seen to have patch-level predictions that were consistently positive.

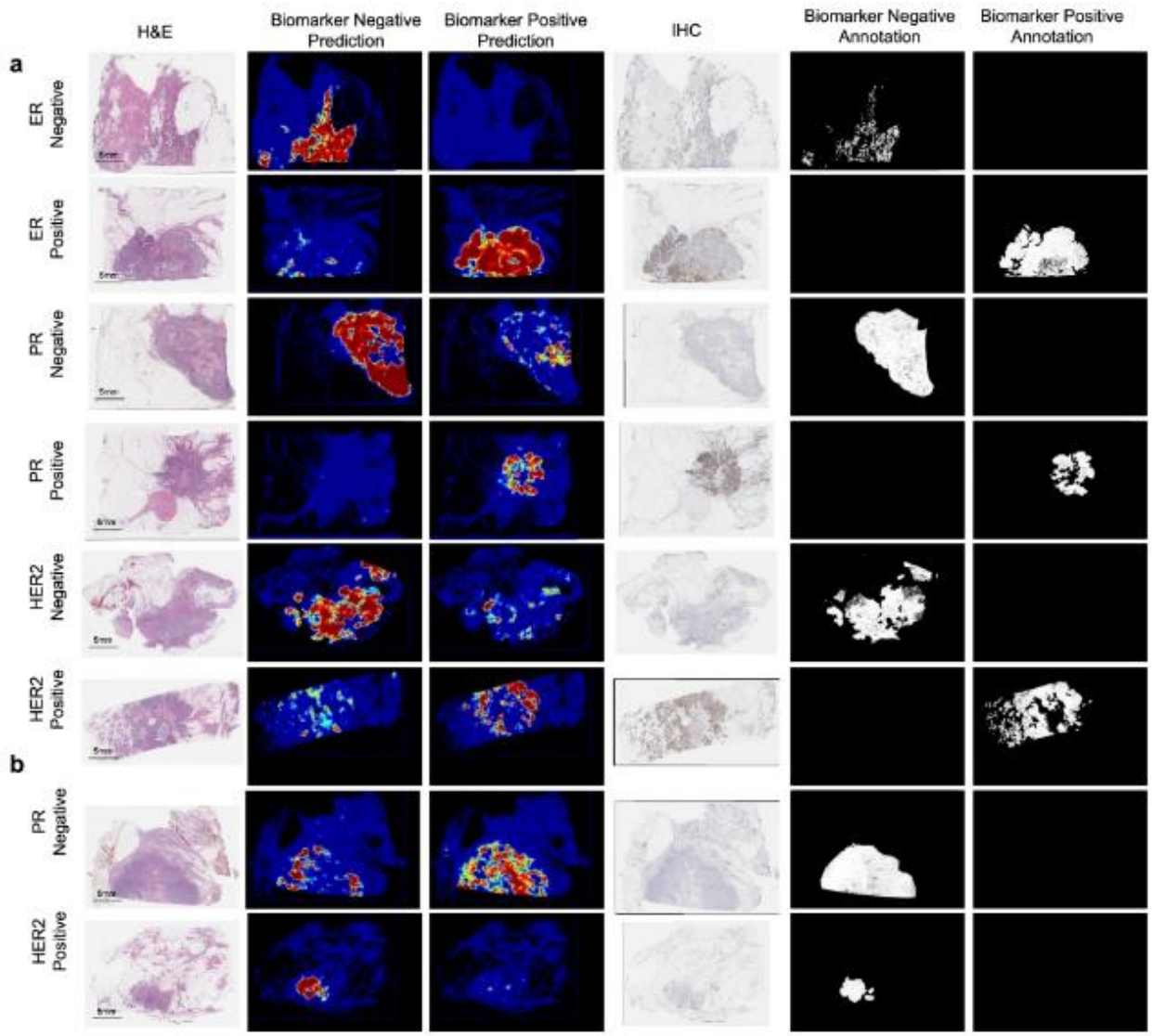


**Figure. 1: Visualization of predictions and annotations.**

a Sample instances for which DLS predictions are concordant with region-level pathologist annotations. b A selection of examples for where the regional-level pathologist annotations and the DLS predictions are in disagreement with one another. Heatmaps provide a visual representation of predictions, with colors ranging from blue (showing a low anticipated likelihood) to red (representing a high projected probability), and black indicating that the prediction does not include tissue. Annotations are represented as white regions in the annotation masks (which are black and white), and the labels that relate to those regions are provided in the column header on white regions. Hematoxylin and eosin (H&E), immunohistochemistry (IHC), deep learning system (DLS), estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER2) are all examples of biological techniques.

**Testing association of DLS patch-level predictions with specific histologic concepts**

We chose six particular histologic traits for which we produced ideas for TCAV analysis based on discussions with breast his to specialists. These aspects are as follows: high-grade carcinoma, low-grade carcinoma, invasive lobular carcinoma, ductal carcinoma in situ (DCIS), tumor-adjacent desmoplastic stromal alterations,

and tumor infiltrating lymphocytes are all types of cancer that may be seen in the bloodstream. For the purpose of this investigation, a high TCAV score for a certain concept (for example, high-grade carcinoma) suggests that the particular DLS biomarker prediction is connected with that concept (for more information, refer to the Methods section). The TCAV score for each idea is shown in the Supplementary section, which includes both positive and negative status forecasts for each relevant biomarker. At the same time as ER-negative expectations were revealed to be connected with TILs, ER-positive predictions were shown to be associated with the low-grade notion. When it came to PR, positive predictions were more significantly related with concepts such as low grade, lobular, DCIS, and desmoplasia, while PR-negative predictions were more firmly connected with the high-grade notion. Negative predictions were more strongly connected with low-grade cancer and lobular carcinoma notions when it came to their association with.

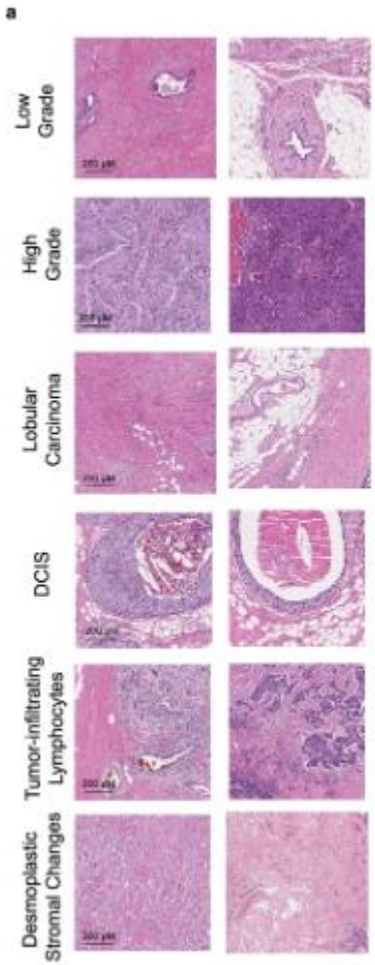


**Figure. 2: Testing with Concept Activation Vector (TCAV) analysis.**

each of the six ideas that were used for the TCAV analysis was represented by a representative concept patch. b The TCAV scores for each of the three biomarkers. These values indicate whether the biomarker status is positive (blue) or negative (red). Using 500 class-of-interest patches, 500 random patches, and 100 concept patches for each trial, the error bars represent confidence intervals with a 95% level of certainty for a total of 20 trials. A greater link between the notion and the model's depiction of that biomarker state is indicated by higher scores. Further information on TCAV scores may be found in Supplementary and Supplementary TIL

tumor infiltrating lymphocytes, ER Estrogen Receptor, PR Progesterone Receptor, human epidermal growth factor receptor 2, and DCIS Ductal Carcinoma in Situ.

## CONCLUSION

Through the use of databases, this study offers information on biomarkers in breast cancer disease. When it comes to the molecular biology of cancer, the majority of databases provide fresh and complete information. For the purpose of identifying and differentiating chemicals and genomic architecture, expression levels, and responses to medications in malignancies, the use of modern technology and bioinformatics tools is a significant contributor. The majority of the databases offered scholars access to a substantial amount of material without encountering any restrictions. For the purpose of developing prospective breast cancer biomarkers, pharmacological and therapeutic targets, analyzing overall survival and recurrence-free survival, and gaining an understanding of cancer genetic and epigenetic profiles, the researchers have used the datasets. In the process of conducting clinical trials, the targets that were discovered via the analysis of the databases are then used for the purposes of treatment, prognosis, and breast cancer prevention, which ultimately leads to advances in personal therapy. These databases are used for curing breast cancer patients in a variety of clinical contexts, which would be advantageous to the diagnosis and treatment of breast cancer. In this study, we summarized many databases that are utilized for curing breast cancer patients.

## REFERTENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011; 61:69–90. doi: 10.3322/caac.20107.

2. Hong M, Park N, Chun YJ. Role of annexin a5 on mitochondria-dependent apoptosis induced by tetramethoxystilbene in human breast cancer cells. Biomol Ther (Seoul) 2014; 22:519–24. doi: 10.4062/biomolther.2014.112.

3. Jung J. Human tumor xenograft models for preclinical assessment of anticancer drug development. Toxicol Res. 2014; 30:1–5. doi: 10.5487/TR.2014.30.1.001.

4. Ko EY, Moon A. Natural products for chemoprevention of breast cancer. J Cancer Prev. 2015; 20:223–31. doi: 10.15430/JCP.2015.20.4.223.

5. Kim DE, Kim Y, Cho DH, Jeong SY, Kim SB, Suh N, et al. Raloxifene induces autophagy-dependent cell death in breast cancer cells via the activation of AMP-activated protein kinase. Mol Cells. 2015; 38:138–44.

6. Malati T. Tumour markers: an overview. Indian J Clin Biochem. 2007; 22:17–31. doi: 10.1007/BF02913308.

7. Liu H, Li J, Koirala P, Ding X, Chen B, Wang Y, et al. Long non-coding RNAs as prognostic markers in human breast cancer. Oncotarget. 2016; 7:20584–96.

8. Yang F, Liu YH, Dong SY, Ma RM, Bhandari A, Zhang XH, et al. A novel long non-coding RNA FGF14-AS2 is correlated with progression and prognosis in breast cancer. Biochem Biophys Res Commun. 2016; 470:479–83.

9. Krishnan P, Ghosh S, Graham K, Mackey JR, Kovalchuk O, Damaraju S. Piwi-interacting RNAs and PIWI genes as novel prognostic markers for breast cancer. Oncotarget. 2016; 7:37944–56.

10. Krishnan P, Ghosh S, Wang B, Li D, Narasimhan A, Berendt R, et al. Next generation sequencing profiling identifies miR-574-3p and miR-660-5p as potential novel prognostic markers for breast cancer. BMC Genomics. 2015; 16:735.

11. Fkih M'hamed I, Privat M, Ponelle F, Penault-Llorca F, Kenani A, Bignon YJ. Identification of miR-10b, miR-26a, miR-146a and miR-153 as potential triple-negative breast cancer biomarkers. Cell Oncol (Dordr) 2015; 38:433–42.

12. Benevolenskaya EV, Islam AB, Ahsan H, Kibriya MG, Jasmine F, Wolff B, et al. DNA methylation and hormone receptor status in breast cancer. Clin Epigenetics. 2016; 8:17. doi: 10.1186/s13148-016-0184-7.

13. Sizemore ST, Sizemore GM, Booth CN, Thompson CL, Silverman P, Bebek G, et al. Hypomethylation of the MMP7 promoter and increased expression of MMP7 distinguishes the basal-like breast cancer subtype from other triple-negative tumors.

14. Conway K, Edmiston SN, Tse CK, Bryant C, Kuan PF, Hair BY, et al. Racial variation in breast tumor promoter methylation in the Carolina Breast Cancer Study. Cancer Epidemiol Biomarkers Prev. 2015; 24:921–30.

15. Kang MH, Jeong KJ, Kim WY, Lee HJ, Gong G, Suh N, et al. Musashi RNA-binding protein 2 regulates estrogen receptor 1 function in breast cancer [published online ahead of print September 5, 2016] Oncogene. doi: 10.1038/onc.2016.327.

16. Fallahi M, Amelio AL, Cleveland JL, Rounbehler RJ. CREB targets define the gene expression signature of malignancies having reduced levels of the tumor suppressor tristetraprolin. PLoS One. 2014;9: e115517.

17. Mittendorf EA, Philips AV, Meric-Bernstam F, Qiao N, Wu Y, Harrington S, et al. PD-L1 expression in triple-negative breast cancer. Cancer Immunol Res. 2014; 2:361–70. doi: 10.1158/2326-6066.CIR-13-0127.

18. Sonnenblick A, Brohée S, Fumagalli D, Vincent D, Venet D, Ignatiadis M, et al. Constitutive phosphorylated STAT3-associated gene signature is predictive for trastuzumab resistance in primary HER2-positive breast cancer. BMC Med.

19. Wang Z, Ruan B, Jin Y, Zhang Y, Li J, Zhu L, et al. Identification of KLK10 as a therapeutic target to reverse trastuzumab resistance in breast cance.

20. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn) 2015;19: A68–77.