

Received July 28, 2020, accepted August 11, 2020, date of publication August 20, 2020, date of current version September 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3018148

# Survival Marker miRNA-Mediated Subpathways of Breast Invasive Carcinoma Derived From Activity Profile

ZIYU NING<sup>1,2</sup>, SHUANG YU<sup>1</sup>, YANQIAO ZHAO<sup>1</sup>, YANG WANG<sup>1</sup>, QI FAN<sup>1</sup>, TIANMING SONG<sup>1</sup>, AND XIAOYANG YU<sup>1</sup>

<sup>1</sup>Higher Educational Key Laboratory for Measuring and Control Technology and Instrumentations of Heilongjiang Province, Harbin University of Science and Technology, Harbin 150080, China

<sup>2</sup>School of Medical Informatics, Harbin Medical University, Daqing Campus, Daqing 163319, China

Corresponding author: Xiaoyang Yu (xiyangyang\_yu@163.com)

This work was supported in part by Heilongjiang Postdoctoral Financial Assistance (LBH-Z19071).

**ABSTRACT** We not only collected 1102 samples from 3 different breast invasive carcinoma (BRCA) datasets, which included a TCGA dataset and 2 GEO datasets, but also collected 23 other cancer datasets, one of which included more than 100 samples. Using these datasets, we topologically inferred miRNA-mediated subpathway activity profiles, which integrated gene expression profiles, prior gene interaction information and target relations between miRNAs and genes, and topological information. Then we constructed the Global Directed Pathway Network (GDPN) with genes as nodes, and from 3 BRCA datasets and other 23 cancer datasets identified a set of miRNA-mediated subpathways that are survival-related risk markers. The results showed large activity values correlated with poor prognosis, such as hsa-miR-107 and hsa-miR-142-3p of BRCA datasets. We assessed the stability and robustness of the miRNA-mediated subpathway survival markers with 2 GEO datasets and 23 external independent datasets. The results showed that the proposed method can significantly reduce noise from sequencing errors and samples heterogeneity by integrating pathway topological information, and can break down the boundary of pathways and provide a new measure for detecting survival-related markers. The top miRNA-mediated subpathways are more reliable in stratifying high risk group and low risk group and selecting therapeutic strategies.

**INDEX TERMS** miRNA-mediated subpathway, survival marker, prognosis, topological information, machine learning.

## I. INTRODUCTION

Breast invasive carcinoma (BRCA) is not only the most commonly diagnosed malignancy in women around the globe but also the world's first cause of female deaths from cancer [1]. BRCA is a heterogeneous disease since from one patient to another it differs in clinical presentation, molecular characteristics, and prognosis [2]. The heterogeneity of BRCA increases the complexity of pathological diagnosis, deciding subtypes and stratification for clinic treatment. For BRCA patients, personalized medicine is becoming a promising way to overcome large variations between individual's treatment responses [3], [4]. In this study, we need to detect robust risk survival markers related to BRCA and evaluate the

performance of our method based on independent datasets. Furthermore, the purpose of our study based on survival data was to find consistent risk miRNA-mediated subpathways (miRNAs), together with facilitation of cancer samples stratification, which could be associated with the prognosis of cancers.

MicroRNAs (miRNAs) are short, endogenous, non-coding RNAs that regulate post-transcription by inhibiting the expression of target genes, thereby affecting the initiation, progression, and prognosis of cancers [5]–[7]. Many studies about high-throughput miRNA-expression profiling have been performed with the aim to identify cancer-related miRNAs for clinical utility in diagnostic and prognostic applications [8]–[10]. MiRNAs are stable and protected from RNase-mediated degradation, thereby enabling its detection in biological fluids and archival tissues for prognosis

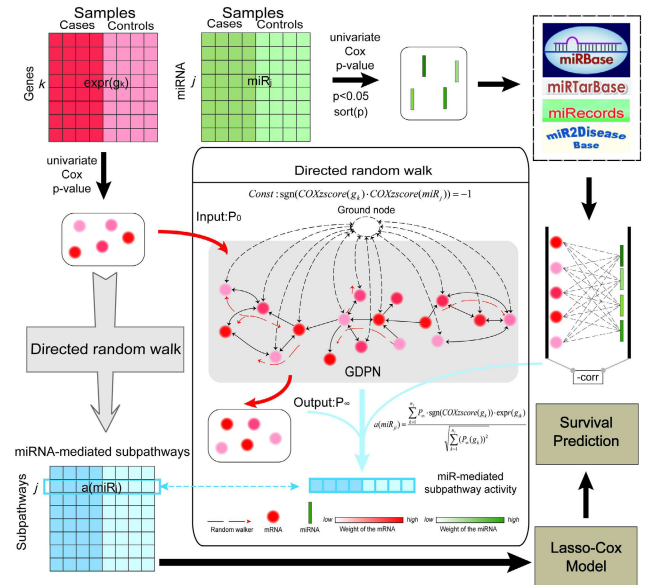
The associate editor coordinating the review of this manuscript and approving it for publication was Quan Zou<sup>1</sup>.

biomarker studies [11]. Thus, several methods were proposed to detect miRNA survival-related biomarkers of the cancers, such as mirwalk analysis platform [12], SVR-luad [13] and miMe [14]. Despite these efforts, at present it is quite difficult to find a clear compendium of miRNA markers for BRCA survival. The stratification performance of miRNA biomarkers often descends drastically in other independent datasets for the same disease phenotype. These problems may be caused by cellular heterogeneity within tissues, the racial differentiation of the patients, the measurement errors in microarray platforms and the sample shortages [15].

Previous studies demonstrated that pathways could be used as a crucial feature in identification of cancer-related survival markers [16]. Therefore, pathway topological analysis can also be used to mine candidate markers for prognosis prediction. Therefore, we incorporated multiple information, such as the differentially expressed level of genes (univariate Cox regression  $p < 0.05$ ), the topological importance of genes in the Global Directed Pathway Network (GDPN), clinical information of samples and target relationships between miRNAs and genes, to identify miRNA-mediated subpathways that could be levered as markers of survival and prognosis for BRCA. The purpose of this study is to topologically infer active miRNA-mediated subpathways toward precise identification of risk survival-related markers. To undertake this work, we proposed a machine learning method, which included data preparation, survival markers selection and risk markers evaluation. First, the collection provided a TCGA (The Cancer Genome Atlas) and 2 GEO (the Gene Expression Omnibus) datasets that included 1102 human BRCA samples, and obtained the target relationships between miRNAs and genes from TarBaseV6.0 [17], miRecords [18], miRTarBase [19], and miR2Disease ([20] databases. We also downloaded other 23 cancer datasets, which included 7086 samples. Second, the GDPN was constructed with genes as nodes. Third, we inferred the expression profiles of miRNA-mediated subpathway activity and extracted the active miRNA-mediated subpathways as risk survival markers by Lasso-Cox model. Finally, we evaluated the performance of our method based on 2 independent datasets and other 23 cancer datasets. After internal and external cross-validation, the top 20 high frequency miRNA-mediated subpathways selected as risk survival-related markers were used to construct risk predictors to stratify patients with respect to their relative risk.

## II. MATERIALS AND METHODS

Figure 1 depicts the method of survival markers identification. The method contains three parts: data preparation, survival markers selection and method evaluation. The data preparation step consists of gene expression matrix and miRNA expression matrix, in which each row represents a gene (or miRNA) and each column represents a sample. Then, the method integrates the gene expression matrix, topology information and target relations between miRNAs and genes into miRNA-mediated subpathway activity profile which is



**FIGURE 1.** The workflow of survival marker identification. The miRNA-mediated subpathways are obtained from the gene profiles by directed random walk. The  $z(g_i)$  is a row vector of gene  $g_i$  expression value across all samples. The  $a(miR_j)$  (i.e. miRNA-mediated sub-pathway activity) is also a row vector which is the row  $j$  of the miRNA (namely  $miR_j$ ) expression value across all samples. The middle panel is the overview illustration of miDRW-based miRNA-mediated subpathway activity inference. The GDPN is constructed on 343 KEGG pathways, which include 7159 gene nodes and 39930 directed edges. The dotted line circle is a virtual node which ensures gene weights to flow through the GDPN.  $P$  is the initial weights of the genes and  $P_\infty$  is the output weight vector. For the  $miR_j$ , we reverse the edge direction when merging the pathways into the GDPN. The miRNA-mediated subpathway activity  $a(miR_j)$  only integrates expression value vector of the univariate Cox regression  $p < 0.05$  genes of  $miR_j$  into  $P_\infty$ .

topologically inferred by the directed random walk. We select risk survival-related markers by ranking frequency based on the outcome of 500 Lasso-Cox models. Finally, we evaluate the performance of our method based on 2 different independent datasets and other 23 cancer datasets.

### A. GENE AND miRNA SAMPLE-MATCHED EXPRESSION DATA

We analyzed and collected 3 independent BRCA datasets (Table S1: Data). One dataset (referred as “BRCA-TCGA”) was downloaded from UCSC Xena Cncter Browser (<https://xena.ucsc.edu/>). We obtained the processed Illumina HiSeq level 3 miRNA and gene expression profiles. Then we removed miRNAs and genes whose expression values equaled to 0, and mapped the gene symbols to Entrez gene ID. Finally, a total of 749 samples with 11574 genes and 153 miRNAs were profiled and used for further analysis. We obtained one dataset [21]–[23] (GSE19783) from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) for independent validation. The platform was Agilent-014850 whole human genome microarray (GPL6480) and Agilent-019118 human miRNA microarray (GPL8227). We mapped the probes to Entrez gene ID and obtained 80 samples with sample-matched 273

miRNAs and 19596 genes. We also obtained the other independent dataset from the GEO database with accession number GSE22220 [24]. The platform was Illumina humanRef-8 v1.0 expression beadchip (GPL6098) and Illumina Human v1 microRNA expression beadchip (GPL8178). A total of 207 samples with 303 miRNAs and 10220 genes profiled were included for further analysis. We also downloaded other 23 cancer datasets from UCSC Xena Cncr Browser. All GEO profiles and 23 TCGA datasets were processed in the similar manner with “BRCA-TCGA” profile.

### B. miRNA-TARGETS ASSOCIATED WITH BRCA

In this study, the target relationships between miRNAs and genes were derived from TarBase V6.0, miRecords, miRTarBase, and miR2Disease databases. After filtering, a total of 755 226 non-repeated human specific interactions among 1137 miRNAs and 20 263 genes were obtained as follows: 598 pairs from miR2Disease, 1749 pairs from miRecords, 26 388 pairs from TarBase, and 750 381 pairs from miRTarBase. We integrated predicted and experimentally verified miRNA-target relationships in our study.

### C. CONSTRUCTING THE PATHWAY GRAPH WITH THE GLOBAL DIRECTION NETWORK (GDPN)

The global pathway graph was constructed based on the interaction data from 343 KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways by using the R package “SubpathwayMiner” (<https://github.com/chunquanlipathway>) [25]. Firstly, each of 343 KEGG pathways was converted into a directed graph based on the interaction information of the KEGG. Then the resulting 343 graphs were merged into a global directed pathway graph (GDPN). We merged the genes that overlapped among the multiple pathways and retained the topological structure of each pathway. The GDPN covers 7159 nodes and 39930 edges. Moreover, a virtual node was added to the GDPN. Each node pointed to the virtual node and virtual node pointed to all other nodes in the GDPN [15], [26]. Each node of the graph represents a gene or a metabolite, and each directed edge represents interaction or regulation relationships between genes in the GDPN. The direction of the edge is derived from the type of interaction between two nodes, which is available from KEGG. For example, if gene A inhibits or activates gene B, the direction of edge points to B. The directed random walk [15], [26] is similar to the PageRank algorithm [27], which is used to search important Web Pages. For PageRank algorithm, if there are more other pages pointing to a Web Page, the Web Page is more important. However, a gene is important if it regulates more other genes [28]. This gene is often called hub genes. The GDPN is followed by power law distributions with an  $R^2 = 0.72, 0.77$  and  $0.71$  for in-degree, out-degree and total degree, respectively. Only a small fraction of genes have higher degrees in the GDPN. It is the most important basic of random walk algorithm [29].

### D. miRNA-MEDIATED SUBPATHWAY ACTIVITY INFERENCE IN THE GDPN

We evaluated the topological importance of genes by the directed random walk in the GDPN. The formula is defined as follows:

$$P_{t+1} = (1 - r)M^T P_t + rP_0 \quad (1)$$

$M$  is the row-normalized adjacency matrix of the GDPN, which was to divide the sum of all elements in the row by each element of the row;  $r \in [0, 1]$  ( $r = 0.7$  in this study) is restart probability;  $P_t$  is a vector in which the  $k$ th element holds the probability of being at gene  $k$  at time step  $t$ ;  $P_0$  is assigned to each gene with its  $-\log(\text{p-value})$ , which is derived from univariate Cox regression analysis, and normalized to a unit vector;  $P_t$  will converge to steady state  $P_\infty$  after several iterations until  $|P_{t+1} - P_t| \leq 10^{-10}$ .  $P_\infty$  are topological weights, which give a measure of global topological importance of genes in the GDPN. So, we suppose that if genes (1) have large degree; (2) have significant p-value and (3) are closely associated with many survival genes that also have significant p-value, these genes will obtain larger weights. The larger weights of genes enhance the inference of a more robust miRNA-mediated subpathway activity. The miRNA-mediated subpathway activity inference formula is defined as follows:

$$a(miR_{ji}) = \frac{\sum_{k=1}^{n_j} P_\infty \cdot \text{sgn}(\text{COXzscore}(g_k)) \cdot \text{expr}(g_{ik})}{\sqrt{\sum_{k=1}^{n_j} (P_\infty(g_k))^2}} \quad (2)$$

$$\text{Constraint :sgn}(\text{COXzscore}(g_k) \cdot \text{COXzscore}(miR_j)) = -1 \quad (3)$$

According to the formula (2), the significantly differential target genes (the univariate Cox regression  $p < 0.05$ , the same as significantly differential target genes below) of each miRNA, could be integrated into a special value, which was called miRNA-mediated subpathway activity. Therefore, the miRNA-mediated subpathway with different sizes are comparable under this activity score function. The  $\text{expr}(g_{ik})$  is the normalized expression value gene  $k$  on the  $i$ th samples;  $\text{COXzscore}(\ )$  is the  $z$  statistics of  $g_k$  and  $miR_j$  from a univariate Cox regression analysis on their expression values and the surviving time;  $\text{sgn}(\ )$  returns  $-1$  for negative numbers and  $+1$  for positive numbers;  $n_j$  represents the number of differentially expressed target genes of each miRNA-mediated subpathway in the GDPN;  $j$  represents the number of miRNA. The formula (3) is a constraint, which is added to formula (2) to detect the negative regulation between miRNAs and genes ( $\text{sgn}(\text{COXzscore}(g_k) \cdot \text{COXzscore}(miR_j))$  equal to  $-1$ ). For example, for upregulated miRNAs, we only integrated downregulated target genes into the gene set and calculated miRNA-mediated subpathway activity, and vice versa. In other words, if  $\text{COXzscore}(g_k) \cdot \text{COXzscore}(miR_j)$  is negative and  $\text{COXzscore}(g_k)$  is positive, the expression values of  $g_k$  in high risk samples will be larger than those in low

risk samples; if  $COXzscore(g_k) - COXzscore(miR_j)$  is negative and  $COXzscore(g_k)$  is negative, the expression values of  $g_k$  in high risk samples will be smaller than those in low risk samples. Thus, the value of  $a(miR_{ji})$  will be large in high risk samples and small in low risk samples for all negatively correlated miRNAs and genes. Finally, we obtain an activity profile of miRNA-mediated subpathway, of which each row represents a miRNA and each column represents a sample, and the values represent the activity of the miRNA in BRCA.

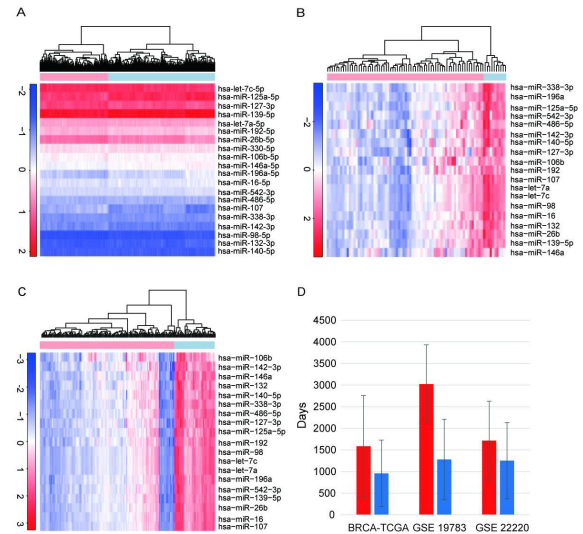
**E. THE SELECTION OF RISK miRNA-MEDIATED SUBPATHWAY AND THE EVALUATION OF METHOD**

For selecting the risk miRNA-mediated subpathway, we sorted the miRNA-mediated subpathways activity by their univariate Cox regression p-value in increasing order and selected miRNA-mediated subpathways with p-value < 0.05 as candidate survival markers. Then we used Lasso-Cox model to select risk survival-related markers, and used C-statistic to estimate the number of survival markers. Firstly, we randomly selected one-fifth of samples for test and the rest for training. For unbiased evaluation, we repeated five-fold cross validation 100 times and calculated the average C-statistics. A total of 500 concordance statistics were generated from five test sets in turn. The R package “glmnet” was used to implement Lasso-Cox model based on activity profile of candidate survival markers. The input of Lasso-Cox model is the data  $(a_{ji}, t_{ji}, \sigma_{ji})$ , where  $a_{ji}$  is the  $j$ th miRNA-mediated subpathway activity of  $i$ th sample;  $t_{ji}$  is the survival time when  $\sigma_{ji} = 1$ , and is the censoring time when  $\sigma_{ji} = 0$ . Thus, we could study the relationship between the miRNA-mediated subpathway activity values and the survival times. A few more important miRNA-mediated subpathways are selected with nonzero coefficients (lambda.min). Then we counted their frequency over 500 Lasso-Cox models. The top ranked miRNA-mediated subpathways in the frequency list are used as candidate risk survival markers. Finally, we use the R package “survAUC” to calculate the C-statistic and estimate the number of survival markers.

**III. RESULTS**

**A. THE METHOD INFERRED SURVIVAL-RELATED miRNA-MEDIATED SUBPATHWAY ACTIVITY**

First, the method is applied to analyze the BRCA-TCGA dataset. We calculated the topological importance of each gene in the GPDN with the formula (1), and inferred miRNA-mediated subpathway (miRNAs) activity profiles with the formula (2) and formula (3). In miRNA-mediated subpathway activity profiles, each row represented a miRNA-mediated subpathway (miRNAs) and each column represented a sample. For BRCA-TCGA dataset, a total of 153 miRNA-mediated subpathway and 749 samples were contained in the inferred activity profiles. To reveal the survival predictive ability of the miRNA-mediated subpathway activity, we sorted the activity profile of miRNA-mediated subpathways by their univariate Cox regression p-value



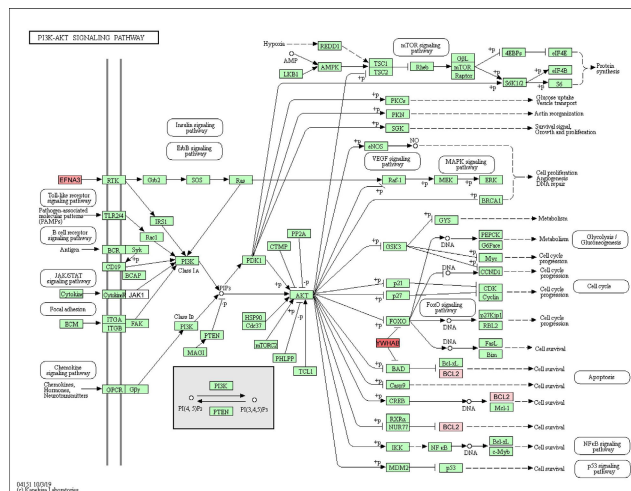
**FIGURE 2.** The activity value of top 20 miRNA-mediated subpathways were used in the hierarchical cluster analysis. (A)-(C) The activity values of top 20 miRNA-mediated subpathways are used in the BRCA-TCGA, GSE19783 and GSE22220, respectively. The row and column represent miRNA-mediated subpathway and samples (The red and blue bar represent good and poor survival), respectively. (D) The height of the bar represents the survival times (days). The clusters are generated with hierarchical cluster analysis.

in increasing order. Then, the candidate miRNA-mediated subpathways ( $p < 0.05$ ) inputted the Lasso-Cox model to perform 100 times five-fold cross validation, and calculated the average C-statistics. The range of average C-statistics was from 0.6102 to 0.8125. It was a stable point of the C-statistics when the N equaled to the 20. Therefore, we called the most frequent miRNA-mediated subpathways (the top 20) survival-related risk markers. These miRNAs have been shown to be related to BRCA survival (Table S1: miRNAs). Hierarchical clustering of the risk markers activity profiles was performed using the R package “gplots” with complete linkage and Euclidean distance (Figure 2). In the heatmap, each row represented a miRNA-mediated subpathway, and each column represented a sample. The survival patterns of BRCA-TCGA dataset were shown in the cluster analysis. The blue cluster samples had poor survival ( $958.34 \pm 769.51$ , Figure 2A). It implied that samples had larger values of the miRNA-mediated subpathway activities in the high risk cluster, and the miRNA-mediated subpathway activities could effectively distinguish different survival pattern of samples. Furthermore, we performed hierarchical cluster analysis based on GSE19783 and GSE22220 using these risk markers except hsa-miR-330 (no hsa-miR-330 in GSE19783 and GSE22220). We obtained the consistent results across these independent datasets (Figure 2B-2C). The results showed that our method could infer robust miRNA-mediated subpathway activities, which can distinguish samples with more significance. Furthermore, these most frequent miRNA-mediated subpathways not only divided samples into two clusters, but also divided them into more detailed subtypes with different survival times (Figure S1).

**B. THE miRNA-MEDIATED SUBPATHWAY INTEGRATED TOPOLOGICAL IMPORTANCE OF THE TARGET GENES**

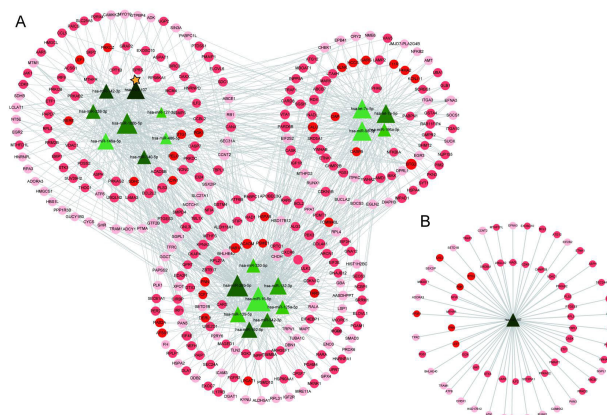
The risk markers that are selected based on activity profile can give us insight into the biological basis about why breast cancer patients have good and poor outcome. For simplicity, if the target genes of miRNA are univariate Cox regression  $p < 0.05$ , we called them differentially expressed target genes in this study. We inferred the activity profile by directed random walk method, which integrated the topological importance of differentially expressed target genes in the GPDN. For evaluating the value of the topology information, we annotated differentially expressed target genes of risk markers into KEGG pathways, respectively. We ranked those pathways with the  $p < 0.01$  in ascending order, and selected the pathways of  $FDR < 0.05$  (Benjamini and Hochberg method). A total of 189 pathways were annotated, and each pathway of the 189 KEGG pathways was selected with the risk markers. Then we counted the frequency occurrences of each pathway and identified the most frequent pathways with over 10 occurrences, of which 82.76% (24/29) were supported by the existing literature (Table S1: Pathways). “Metabolic pathways” (hsa01100) and “HIF-1 signaling pathway” (hsa04066) were ranked first and second on the most frequent pathways list respectively. The “Metabolic pathways” had been the research hotspots in the therapy of breast cancer [30-33]. Metabolic pathways promoted cancer cell survival and growth [34]. The “Metabolic pathways” had been shown to play important roles in early cancer detection such as amino acid metabolism, arachidonic acid (AA) metabolism, fatty acid metabolism, linoleic acid metabolism, and retinol metabolism, showing significant differences between the breast cancer group and the control group in a Korean prospective cohort [30]. Many researches had shown that the effects of hypoxia-inducible factor 1 (HIF-1), a master regulator of the hypoxic response, had been extensively studied during breast cancer progression and metastasis [35]. Many studies showed that HIF-1 signaling pathway played a role in breast cancer progression and metastasis with the PI3K-Akt signaling pathway [36]–[38] (hsa04151, Figure 3).

Furthermore, we calculated the topological weights of differentially expressed target genes in the top 20 miRNA-mediated subpathways, and sorted them in descending order. 38 of the top 100 topologically important target genes were annotated to the most frequent pathways (over 10 occurrences), such as PGK1 and BTG2. PGK1 was a potential survival biomarker and invasion promoter by regulating the HIF-1 $\alpha$ -Mediated epithelial-mesenchymal transition process in breast cancer, and high PGK1 expression predicted poor overall survival (OS) in breast cancer and some other cancers [39], [40]. BTG2 inhibited mTORc1 activity by reducing Raptor-mTOR interaction along with upregulation of tsc1 expression, which led to significant reduction of p70S6K activation as opposed to AKT1S473, but not AKT2, phosphorylation via downregulating PHLPP2 (AKT1-specific phosphatase) in breast cancers [41]. Additionally, high-level



**FIGURE 3.** A snapshot of the PI3K-Akt signaling pathway (hsa04151). The target genes of hsa-miR-107 is annotated to PI3K-Akt signaling pathway. A gene is colored darker red if one of its target genes is more significantly disordered ( univariate Cox regression  $p < 0.05$ ).

BTG2 protein expression correlated with prolonged survival in patients with breast carcinoma. [42]. The result indicated that the topological integration was valuable strategy, which could detected the risk miRNA-mediated subpathways and the key genes of these subpathways (Figure 3, Table S1: Genes). Furthermore, these miRNA-mediated subpathways located smaller pathways region, and broke down the boundary of pathways, which helped us to understand the interaction between molecules and detect survival-related markers from the whole pathway network.

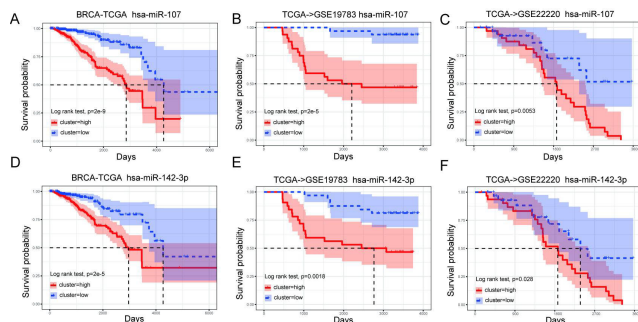


**FIGURE 4.** The full view of miRNA-mediated subpathways. (A) The full view of miRNA-mediated subpathways, which include 1087 edges, 20 miRNAs and 292 genes. (B) One node represents a dysregulated miRNA or target gene, and the color represents the level of differential expression. Some survival-related star genes targeted by hsa-miR-107 (with five-pointed star in Figure 4A), such as PGK1 and BTG2.

**C. CASE STUDY**

There were a total of 1087 edges between 20 miRNAs and their differentially expressed 292 target genes (Figure 4A). Each miRNA-mediated subpathway activity was integrated by their differentially expressed target genes, and considered the topological importance of genes. Then miRNA-mediated

subpathways were more robust survival markers. For evaluating predictive performance of the survival markers, we further investigated whether the miRNA-mediated subpathway activities of the most frequently selected risk markers could stratified the patients into low-risk group and high-risk group. We ranked the breast cancer patients by activity values of each miRNA-mediated subpathway in descending order, and assigned the top 40% patients as the high risk group and the the bottom 40% patients patients as the low risk group. The activity values of hsa-miR-107, hsa-miR-142-3p, hsa-miR-192, hsa-miR-338-3p, hsa-miR-106b, hsa-miR-542-3p, hsa-miR-125a-5p and hsa-miR-127-3p significantly separated breast cancer patients in independent GSE19783 and GSE2220 datasets (there is no hsa-miR-330-5p in GSE19783 and GSE2220). The hsa-miR-107 and hsa-miR-142-3p ranked the first and the second in the frequency list. Figure 4B showed the network based on hsa-miR-107, which targeted many star genes of breast cancer, such as PGK1, BTG2 and CDK8. Furthermore, we used the R package “survminer” to visualize the Kaplan-Meier survival curves and the log rank test of subtypes (Figure 5). Figure 5B and Figure 5E showed the Kaplan-Meier curves for two risk groups in GSE19783 (log-rank,  $p = 2e-5$  and  $p = 2e-3$ ), and Figure 5C and Figure 5F showed the Kaplan-Meier curves for two risk groups in GSE2220 (log-rank,  $p = 5e-3$  and  $p = 3e-2$ ), respectively. It is noted that the follow-up time of GSE2220 was less than 10 years. Thus we removed those samples, of which the follow-up time equaled to 10 years.



**FIGURE 5.** Kaplan-Meier curves stratified by risk miRNA-mediated subpathways (markers) for patients. (A)-(C) Kaplan-Meier curves for patients in BRCA-TCGA, GSE19783 and GSE2220 stratified by the hsa-miR-107, respectively. (D)-(F) Kaplan-Meier curves for patients in BRCA-TCGA, GSE19783 and GSE2220 stratified by the hsa-miR-142-3p, respectively.

Many researches had shown hsa-miR-107 and hsa-miR-142-3p were the survival-related markers. The hsa-miR-107 was associated with tumor recurrence and reduced OS in TNBC patients as it was exclusively up-regulated in relapsing TNBC compared with non-relapsing TNBC, healthy subjects or ER + patients [43], [44]. The expression levels of hsa-miR-107 were significantly elevated in the metastatic group compared with the disease-free group [45]. Moreover, 20 differentially expressed genes of hsa-miR-107 were annotated to 39 pathways, and four differentially expressed genes (JAK1, BCL2, YWHAB and EFNA3) of

hsa-miR-107 were annotated to the PI3K-Akt signaling pathway. JAK1 expression was significantly lower in breast invasive carcinoma compared with adjacent normal tissues, and JAK1 was a prognostic marker and its correlation with immune infiltrated in breast cancer [46]. BCL2 and YWHAB were significantly associated with overall survival in breast cancer [47], [48]. EFNA3 long noncoding RNAs induced by hypoxia promote metastatic dissemination [49] (Figure 3). On the other hand, The hsa-miR-142-3p played an important role in the therapy of breast cancer. The miR-142-3p overexpression resulted in a strong decline in breast cancer stem cell characteristics with a decrease in CD44, CD133, ALDH1, Bod1, BRCA2, and mammosphere formation as well as reduced survival after irradiation [50]. Re-expressing the hypoxia-repressed miR-142-3p, which targeted HIF1A, LOX and ITGA5, and caused further suppression of the HIF-1 $\alpha$ /LOX/ITGA5/FN1 axis [51]. Notably, higher LOX, ITGA5, or FN1, or lower miR-142-3p levels are associated with shorter survival in chemotherapy-treated TNBC patients [51].

The above results indicated that the miRNA-mediated subpathways identified by our method could be served as potential markers for breast cancer prognosis,

#### D. EXTERNAL VALIDATION OF SURVIVAL MARKERS USING ON ANOTHER 23 CANCERS

The analyses done so far provided a ranked collection of miRNA-mediated subpathways found as robust markers of survival in BRCA. The consistency of the results obtained with the independent validation gave strong support to the risk markers found, but we had to consider the activity values used to evaluate our method in other cancers. We downloaded other 31 cancer datasets from UCSC Xena Cncr Browser (<https://xena.ucsc.edu/>). After data preparation, 23 cancer datasets with more than 100 samples were reserved for further analysis, which contained sample-matched miRNA expression, gene expression, clinical information. For unbiasedly evaluating the performance of the method, we calculated the activity profiles of 23 datasets, then we also selected top 20 miRNA-mediated subpathways (the most frequent miRNA-mediated subpathways) as risk survival markers. We ranked the breast cancer patients by activity values of risk markers in descending order, and assigned the top 40% patients as the high risk group and the the bottom 40% patients patients as the low risk group. In 23 cancers datasets, we identified the most frequent common 89 risk markers, of which 17/89 occurred over 10 times and were supported by the existing literature (Table S1: Other independent datasets). These risk markers could significantly separate cancer patients in their datasets with Kaplan-Meier curves. However, the PRAD showed an unsatisfactory stratification effect because the dataset included multi-subtypes, and PRAD was a highly heterogeneous cancer.

The hsa-miR-107 and hsa-miR-125a-5p ranked first and second in the frequency list. Clinically, the signature of a hsa-miR-107 high, DAPK low, and KLF4 low expression profile

correlated with the extent of lymph node and distant metastasis in patients with colorectal cancer and served as a prognostic marker for metastasis recurrence and poor survival [52]. High hsa-miR-107 expression was associated with poor clinicopathological parameters and prognosis in pancreatic ductal adenocarcinoma patients [53]. On the other hand, overexpression of hsa-miR-125a-5p significantly enhanced the ability of cell proliferation, migration and invasion in HNSCC, hsa-miR-125a-5p [54]. The hsa-miR-125a-5p induced apoptosis via a p53-dependent pathway in human lung cancer cells [55]. In other words, hsa-miR-125a-5p, a diagnosis and prognosis marker for multiple cancers, was the star miRNA to regulate important cancer-related pathways and targeted key genes.

Furthermore, we manually retrieved a lot of literatures, whereas there were no similar methods. We implemented five famous pathway-based identification of survival markers methods, which contained the Mean and Median methods [56], the PCA method [57] and the previous DRWPSurv method [16]. We repeated five fold validation experiments 100 times on each dataset and cross datasets, and generated 500 C-statistics and 100 C-statistics, respectively (Figure S2).

The external validation of survival analysis showed that the method could detect the survival risk markers of cancers and the key target genes of miRNA-mediated subpathway. The results indicated that the method was designed to be independent of diseases, and could detect robust survival-related markers (genes and miRNAs) in other cancers. Meanwhile, our method was capable of maintaining the stable identification for different measurement platforms and patient cohorts.

#### IV. DISCUSSION

BRCA is a heterogeneous cancer composed of diverse subtypes, which presents multiple clinical symptoms. This heterogeneity causes the molecular stabilization of BRCA to remain deficient, with a lack of clear risk markers related to the prognosis of the BRCA [58]. General risk factors of the breast cancer in females are age, infertility, age of first full time pregnancy, age of menopause, an inherited mutation in the BRCA1/BRCA2 breast cancer gene [59]. In the recent years, considerable progress has been made to understand the mechanisms responsible for aberrant miRNA expression in BRCA and several miRNAs or miRNA families have been found as key regulators of BRCA hallmarks [60].

The identification of risk miRNA markers with prognostic value in BRCA has been a challenging task [61]. With the development of high-throughout sequencing technology, more and more genomes have been successfully sequenced, many miRNA markers associated with survival have been found, such as hsa-miR-107 [43], [44], hsa-miR-142-3p [51], hsa-miR-192 [62], hsa-miR-338-3p [63], hsa-miR-106b [64], hsa-miR-542-3p [65], hsa-miR-125a-5p [66], [67] and hsa-miR-127-3p [68], [69]. Thus, many researchers have increasingly focused on the detection of robust survival-related miRNA markers. However, at present there is not an explicit list of survival-related miRNA markers for BRCA and it is quite difficult to find consistency in the literature.

The recent advance of multi-omics integration technologies applied to the study of clinical samples does open the way to provide a new measure of risk survival-related markers. A clear limitation comes from the fact that, previous studies, focused on genes or miRNAs how to improve the performance of prognostic predictors, ignored effect of the joint impact from genes and miRNAs. It is well known that miRNAs can disrupt biological pathways and cause diseases by regulating their target genes. Thus, there must be complicated relations among miRNAs, genes and pathways. Another view about risk survival markers is that, disease phenotypes are found to be highly associated with the key local subpathways, rather than entire pathways [25], [70], [71]. We believe that it is a promising way to understand the biological mechanism of cancer survival with deep mining of miRNA-mediated subpathways. If subpathways are to be analyzed effectively, we shall consider the topology structure of pathway network.

The purpose of this study is to detect robust risk markers for precise survival outcome prediction. The method contains three parts: data preparation, survival marker selection and method evaluation. The core of this study is selection of survival markers, which integrate gene expression profiles, prior gene interaction information and target relations between miRNAs and genes, and survival markers are topologically inferred by the directed random walk. Results have been shown that the our method obtained a better predictive performance on breast cancer independent and other cancer datasets.

We integrated weak signals of pathway member genes with mRNA-miRNAs target relations, miRNA-mediated subpathway activities accumulate larger discriminative power and are more robust in predicting survival outcome than individual gene or miRNA markers. The reliable performance of our method could be attributed to the strategy of incorporating multiple information at the step of miRNA-mediated subpathway activity inference. Our method reassigned the topological importance of each gene in a GPDN by directed random walk. The more topologically important the genes were, the larger topology weights they could obtain, and the larger activities they contributed to the miRNA-mediated subpathways. The directed random walk method adjusted gene weights according to the topological importance, the dysregulated signals of key genes were amplified in the GDPN, and vice versa. Finally, the survival-related miRNAs were identified with larger topology weights of target genes. As the hub genes often had subtle expression changes [72], the robustness of the inferred miRNA-mediated subpathway activities might be enhanced, leading to better prognosis prediction.

#### V. CONCLUSION AND FUTURE WORK

In conclusion, we consider that the results presented in this study provide strong evidence for the prognostic value of a list of miRNA-mediated subpathways in BRCA and for their potential to predict other cancers prognosis. In our method, the set of miRNA-mediated subpathways is ranked with their

selection frequency of 500 Lasso-Cox models, being the top ones in the list that provide best prognostic performance. In fact, our results show that the top 20 miRNA-mediated subpathways applied to independent datasets provided very good separation of BRCA datasets and other 23 cancer datasets in two distinct groups of high risk and low risk. The method can significantly reduce noise from sequencing errors and samples heterogeneity by integrating pathway topological information, and can break down the boundary of pathways and provide a new measure to detect survival-related markers. However, our method depends on the data collection and credibility of target relationships. Thus, with the rapid development of human interaction databases and the sequence technology, we believe that the method is a promising way to precisely predict the state of prognosis, and provide a better guide for patient treatment.

#### ACKNOWLEDGMENT

The authors would like to thank all the research staff for their contributions to this project. (Ziyu Ning, Shuang Yu, and Yanqiao Zhao contributed equally to this work.)

#### CONFLICTS OF INTEREST

The authors declare no conflict of interest.

#### REFERENCES

- [1] T. Sarosiek, "Systemic treatment of early breast cancer-current state of knowledge after the conference St Gallen 2017," *Polski Merkurizus Lekarski*, vol. 43, pp. 232–236, Nov. 2017.
- [2] S. K. Yeo and J.-L. Guan, "Breast cancer: Multiple subtypes within a tumor?" *Trends Cancer*, vol. 3, no. 11, pp. 753–760, Nov. 2017.
- [3] E. S. McDonald, A. S. Clark, J. Tchou, P. Zhang, and G. M. Freedman, "Clinical diagnosis and management of breast cancer," *J. Nucl. Med.*, vol. 57, no. 1, pp. 9S–16S, Feb. 2016.
- [4] J. A. de la Mare, L. Contu, M. C. Hunter, B. Moyo, J. N. Sterrenberg, and K. C. Dhanani, "Breast cancer: Current developments in molecular approaches to diagnosis and treatment," *Recent Patents Anti-Cancer Drug Discovery*, vol. 9, pp. 75–153, May 2014.
- [5] S.-S. Zhou, J.-P. Jin, J.-Q. Wang, Z.-G. Zhang, J. H. Freedman, Y. Zheng, and L. Cai, "MiRNAs in cardiovascular diseases: Potential biomarkers, therapeutic targets and challenges," *Acta Pharmacol. Sinica*, vol. 39, no. 7, pp. 1073–1084, Jul. 2018.
- [6] A. M. Mohr and J. L. Mott, "Overview of microRNA biology," *Seminars Liver Disease*, vol. 35, pp. 3–11, Feb. 2015.
- [7] Y. Cai, X. Yu, S. Hu, and J. Yu, "A brief review on the mechanisms of miRNA regulation," *Genomics, Proteomics Bioinf.*, vol. 7, no. 4, pp. 147–154, Dec. 2009.
- [8] E. S. Martens-Uzunova, S. E. Jalava, N. F. Dits, G. J. van Leenders, S. Moller, and J. Trapman, "Diagnostic and prognostic signatures from the small non-coding RNA transcriptome in prostate cancer," *Oncogene*, vol. 31, pp. 978–991, Feb. 2012.
- [9] C. Jay, J. Nemunaitis, P. Chen, P. Fulgham, and A. W. Tong, "MiRNA profiling for diagnosis and prognosis of human cancer," *DNA Cell Biol.*, vol. 26, no. 5, pp. 293–300, May 2007.
- [10] E. Giovannetti, A. van der Velde, N. Funel, E. Vasile, V. Perrone, L. G. Leon, N. De Lio, A. Avani, S. Caponi, L. E. Pollina, V. Gallá, H. Sudo, A. Falcone, D. Campani, U. Boggi, and G. J. Peters, "High-throughput MicroRNA (miRNAs) arrays unravel the prognostic role of MiR-211 in pancreatic cancer," *PLoS ONE*, vol. 7, no. 11, Nov. 2012, Art. no. e49145.
- [11] K. K. To, C. W. Tong, M. Wu, and W. C. Cho, "MicroRNAs in the prognosis and therapy of colorectal cancer: From bench to bedside," *World J. Gastroenterol.*, vol. 24, no. 27, pp. 2949–2973, Jul. 2018.
- [12] H. M. Zhang, B. Yang, H. F. Chen, X. H. Chi, Y. B. Xi, and X. M. Chen, "Prognosis-related miRNA bioinformatics screening of lung adenocarcinoma and its clinical significance," *Zhongguo Ying Yong Sheng Li Xue Za Zhi*, vol. 34, no. 6, pp. 530–535, Jun. 2018.
- [13] S. Yerukala Sathipati and S.-Y. Ho, "Identifying the miRNA signature associated with survival time in patients with lung adenocarcinoma using miRNA expression profiles," *Sci. Rep.*, vol. 7, no. 1, p. 7507, Aug. 2017.
- [14] S. H. Strand, E. Bavafaye-Haghighi, H. Kristensen, A. K. Rasmussen, S. Hoyer, M. Borre, P. Mouritzen, S. Besenbacher, T. F. Orntoft, and K. D. Sorensen, "A novel combined miRNA and methylation marker panel (miMe) for prediction of prostate cancer outcome after radical prostatectomy," *Int. J. Cancer*, vol. 145, no. 12, pp. 3445–3452, Dec. 2019.
- [15] Z. Ning, C. Feng, C. Song, W. Liu, D. Shang, and M. Li, "Topologically inferring active miRNA-mediated subpathways toward precise cancer classification by directed random walk," *Mol. Oncol.*, vol. 13, no. 10, pp. 2211–2226, Oct. 2019.
- [16] W. Liu, W. Wang, G. Tian, W. Xie, L. Lei, J. Liu, W. Huang, L. Xu, and E. Li, "Topologically inferring pathway activity for precise survival outcome prediction: Breast cancer as a case," *Mol. Biosyst.*, vol. 13, no. 3, pp. 537–548, 2017.
- [17] T. Vergoulis, I. S. Vlachos, P. Alexiou, G. Georgakilas, M. Maragkakis, M. Reczko, S. Gerangelos, N. Koziris, T. Dalamagas, and A. G. Hatzigeorgiou, "TarBase 6.0: Capturing the exponential growth of miRNA targets with experimental support," *Nucleic Acids Res.*, vol. 40, no. D1, pp. D222–D229, Jan. 2012.
- [18] F. Xiao, Z. Zuo, G. Cai, S. Kang, X. Gao, and T. Li, "MiRecords: An integrated resource for microRNA-target interactions," *Nucleic Acids Res.*, vol. 37, pp. D105–D110, Jan. 2009.
- [19] C. H. Chou et al., "miRTarBase update 2018: A resource for experimentally validated microRNA-target interactions," *Nucleic Acids Res.*, vol. 46, pp. D296–D302, Jan. 2018.
- [20] Q. Jiang, Y. Wang, Y. Hao, L. Juan, M. Teng, X. Zhang, M. Li, G. Wang, and Y. Liu, "MiR2Disease: A manually curated database for microRNA deregulation in human disease," *Nucleic Acids Res.*, vol. 37, pp. D98–D104, Jan. 2009.
- [21] V. D. Haakensen, I. Steinfeld, R. Saldova, A. A. Shehni, I. Kifer, B. Naume, P. M. Rudd, A.-L. Børresen-Dale, and Z. Yakhini, "Serum N-glycan analysis in breast cancer patients—Relation to tumour biology and clinical outcome," *Mol. Oncol.*, vol. 10, no. 1, pp. 59–72, Jan. 2016.
- [22] E. Enerly, I. Steinfeld, K. Kleivi, S.-K. Leivonen, M. R. Aure, H. G. Russnes, J. A. Rønneberg, H. Johnsen, R. Navon, E. Rødland, R. Mäkelä, B. Naume, M. Perälä, O. Kallioniemi, V. N. Kristensen, Z. Yakhini, and A.-L. Børresen-Dale, "MiRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors," *PLoS ONE*, vol. 6, no. 2, Feb. 2011, Art. no. e16915.
- [23] M. R. Aure, I. Steinfeld, L. O. Baumbusch, K. Liestøl, D. Lipson, S. Nyberg, B. Naume, K. K. Sahlborg, V. N. Kristensen, A.-L. Børresen-Dale, O. C. Lingjærde, and Z. Yakhini, "Identifying in-trans process associated genes in breast cancer by integrated analysis of copy number and expression data," *PLoS ONE*, vol. 8, no. 1, Jan. 2013, Art. no. e53014.
- [24] F. M. Buffa, C. Camps, L. Winchester, C. E. Snell, H. E. Gee, and H. Sheldon, "microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer," *Cancer Res.*, vol. 71, no. 17, pp. 5635–5645, Sep. 2011.
- [25] C. Li, X. Li, Y. Miao, Q. Wang, W. Jiang, C. Xu, J. Li, J. Han, F. Zhang, B. Gong, and L. Xu, "SubpathwayMiner: A software package for flexible identification of pathways," *Nucleic Acids Res.*, vol. 37, no. 19, p. e131, Oct. 2009.
- [26] W. Liu, C. Li, Y. Xu, H. Yang, Q. Yao, and J. Han, "Topologically inferring risk-active pathways toward precise cancer classification by directed random walk," *Bioinformatics*, vol. 29, no. 17, pp. 2169–2177, Sep. 2013.
- [27] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, pp. 107–117, 1998.
- [28] S. Draghici, P. Khatri, A. L. Tarca, K. Amin, A. Done, and C. Voichita, "A systems biology approach for pathway level analysis," *Genome Res.*, vol. 17, no. 10, pp. 1537–1545, Oct. 2007.
- [29] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.
- [30] H. J. Yoo, M. Kim, M. Kim, M. Kang, K. J. Jung, S.-M. Hwang, S. H. Jee, and J. H. Lee, "Analysis of metabolites and metabolic pathways in breast cancer in a Korean prospective cohort: The Korean cancer prevention study-II," *Metabolomics*, vol. 14, no. 6, p. 85, Jun. 2018.
- [31] S. R. Lord, W. C. Cheng, D. Liu, E. Gaude, S. Haider, and T. Metcalf, "Integrated pharmacodynamic analysis identifies two metabolic adaption pathways to metformin in breast cancer," *Cell Metabolism*, vol. 28, no. 5, pp. 679–688, Nov. 2018.



- [32] Y. Gokmen-Polar, Y. Neelamraju, C. P. Goswami, Y. Gu, X. Gu, and G. Nallamothu, "Splicing factor ESRP1 controls ER-positive breast cancer by altering metabolic pathways," *EMBO Rep.*, vol. 20, no. 2, Feb. 2019, Art. no. e46078.
- [33] R. C. Geck and A. Toker, "Nonessential amino acid metabolism in breast cancer," *Adv. Biol. Regulation*, vol. 62, pp. 11–17, Sep. 2016.
- [34] L. K. Boroughs and R. J. DeBerardinis, "Metabolic pathways promoting cancer cell survival and growth," *Nature Cell Biol.*, vol. 17, pp. 351–359, Mar. 2015.
- [35] Z.-J. Liu, G. L. Semenza, and H.-F. Zhang, "Hypoxia-inducible factor 1 and breast cancer metastasis," *J. Zhejiang Univ.-Sci. B*, vol. 16, no. 1, pp. 32–43, Jan. 2015.
- [36] V. R. Sharma, G. K. Gupta, A. K. Sharma, N. Batra, D. K. Sharma, A. Joshi, and A. K. Sharma, "PI3K/Akt/mTOR intracellular pathway and breast cancer: Factors, mechanism and regulation," *Current Pharmaceutical Des.*, vol. 23, no. 11, pp. 1633–1638, May 2017.
- [37] L. Du, X. Li, L. Zhen, W. Chen, L. Mu, Y. Zhang, and A. Song, "Everolimus inhibits breast cancer cell growth through PI3K/AKT/mTOR signaling pathway," *Mol. Med. Rep.*, vol. 17, pp. 7163–7169, Mar. 2018.
- [38] N. Dey, P. De, and B. Leyland-Jones, "PI3K-AKT-mTOR inhibitors in breast cancers: From tumor cell signaling to clinical trials," *Pharmacol. Therapeutics*, vol. 175, pp. 91–106, Jul. 2017.
- [39] D. Xu, J. A. Aka, R. Wang, and S.-X. Lin, "17beta-hydroxysteroid dehydrogenase type 5 is negatively correlated to apoptosis inhibitor GRP78 and tumor-secreted protein PGK1, and modulates breast cancer cell viability and proliferation," *J. Steroid Biochem. Mol. Biol.*, vol. 171, pp. 270–280, Jul. 2017.
- [40] D. Fu, C. He, J. Wei, Z. Zhang, Y. Luo, and H. Tan, "PGK1 is a potential survival biomarker and invasion promoter by regulating the HIF-1 $\alpha$ -mediated epithelial-mesenchymal transition process in breast cancer," *Cellular Physiol. Biochem.*, vol. 51, no. 5, pp. 2434–2444, 2018.
- [41] S. Sundaramoorthy, P. Devanand, M. S. Ryu, K. Y. Song, D. Y. Noh, and I. K. Lim, "TIS21<sup>BTG2</sup> inhibits breast cancer growth and progression by differential regulation of mTORc1 and mTORc2-AKT1-NFAT1-PHLPP2 signaling axis," *J. Cancer Res. Clin. Oncol.*, vol. 144, no. 8, pp. 1445–1462, Aug. 2018.
- [42] E. Möllerström, A. Kovács, K. Lövgren, S. Nemes, U. Delle, A. Danielsson, T. Parris, D. J. Brennan, K. Jirström, P. Karlsson, and K. Helou, "Up-regulation of cell cycle arrest protein BTG2 correlates with increased overall survival in breast cancer, as detected by immunohistochemistry using tissue microarray," *BMC Cancer*, vol. 10, no. 1, p. 296, Jun. 2010.
- [43] Q. Tang, H. Ouyang, D. He, C. Yu, and G. Tang, "MicroRNA-based potential diagnostic, prognostic and therapeutic applications in triple-negative breast cancer," *Artif. Cells, Nanomedicine, Biotechnol.*, vol. 47, no. 1, pp. 2800–2809, Dec. 2019.
- [44] E. D'Ippolito and M. V. Iorio, "MicroRNAs and triple negative breast cancer," *Int. J. Mol. Sci.*, vol. 14, no. 11, pp. 22202–22220, Nov. 2013.
- [45] S. Shen, Q. Sun, Z. Liang, X. Cui, X. Ren, H. Chen, X. Zhang, and Y. Zhou, "A prognostic model of triple-negative breast cancer based on miR-27b-3p and node status," *PLoS ONE*, vol. 9, no. 6, Jun. 2014, Art. no. e100664.
- [46] B. Chen, J. Lai, D. Dai, R. Chen, X. Li, and N. Liao, "JAK1 as a prognostic marker and its correlation with immune infiltrates in breast cancer," *Aging*, vol. 11, no. 23, pp. 11124–11135, Dec. 2019.
- [47] G. Kallergi, V. Tsintari, S. Sfakianakis, E. Bei, E. Lagoudaki, A. Koutsoyopoulos, N. Zacharopoulou, S. Alkahtani, S. Alarifi, C. Stournaras, M. Zervakis, and V. Georgoulas, "The prognostic value of JUNB-positive CTCs in metastatic breast cancer: From bioinformatics to phenotypic characterization," *Breast Cancer Res.*, vol. 21, no. 1, p. 86, Aug. 2019.
- [48] Y. Gu, P. Li, F. Peng, M. Zhang, Y. Zhang, and H. Liang, "Autophagy-related prognostic signature for breast cancer," *Mol. Carcinogenesis*, vol. 55, no. 3, pp. 292–299, Mar. 2016.
- [49] L. Gómez-Maldonado, M. Tiana, O. Roche, A. Prado-Cabrero, L. Jensen, and A. Fernandez-Barral, "EFNA3 long noncoding RNAs induced by hypoxia promote metastatic dissemination," *Oncogene*, vol. 34, pp. 2609–2620, May 2015.
- [50] F. M. Troschel, N. Bohly, K. Borrmann, T. Braun, A. Schwickert, and L. Kiesel, "miR-142-3p attenuates breast cancer stem cell characteristics and decreases radioresistance in vitro," *Tumor Biol.*, vol. 40, no. 8, Aug. 2018, Art. no. 1010428318791887.
- [51] O. Saatci et al., "Targeting lysyl oxidase (LOX) overcomes chemotherapy resistance in triple negative breast cancer," *Nature Commun.*, vol. 11, p. 2416, May 2020.
- [52] H. Y. Chen, Y. M. Lin, H. C. Chung, Y. D. Lang, C. J. Lin, and J. Huang, "miR-103/107 promote metastasis of colorectal cancer by targeting the metastasis suppressors DAPK and KLF4," *Cancer Res*, vol. 72, no. 14, pp. 3631–3641, Jul. 2012.
- [53] J. Xiong, D. Wang, A. Wei, H. Lu, C. Tan, A. Li, J. Tang, Y. Wang, S. He, X. Liu, and W. Hu, "Deregulated expression of miR-107 inhibits metastasis of PDAC through inhibition PI3K/Akt signaling via caveolin-1 and PTEN," *Exp. Cell Res.*, vol. 361, no. 2, pp. 316–323, Dec. 2017.
- [54] S. Jin, M. D. Liu, H. Wu, P. Pang, S. Wang, and Z. N. Li, "Overexpression of hsa-miR-125a-5p enhances proliferation, migration and invasion of head and neck squamous cell carcinoma cell lines by upregulating C-C chemokine receptor type 7," *Oncol. Lett.*, vol. 15, no. 6, pp. 9703–9710, Jun. 2018.
- [55] L. Jiang, Q. Huang, J. Chang, E. Wang, and X. Qiu, "MicroRNA HSA-miR-125a-5p induces apoptosis by activating p53 in lung cancer cells," *Exp. Lung Res.*, vol. 37, no. 7, pp. 387–398, Sep. 2011.
- [56] Z. Guo, "Towards precise classification of cancers based on robust gene functional expression profiles," *BMC Bioinf.*, vol. 6, p. 58, Mar. 2005.
- [57] A. H. Bild, G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M.-B. Joshi, D. Harpole, J. M. Lancaster, A. Berchuck, J. A. Olson, J. R. Marks, H. K. Dressman, M. West, and J. R. Nevins, "Oncogenic pathway signatures in human cancers as a guide to targeted therapies," *Nature*, vol. 439, no. 7074, pp. 353–357, Jan. 2006.
- [58] H. Jia, C. I. Truica, B. Wang, Y. Wang, X. Ren, H. A. Harvey, J. Song, and J.-M. Yang, "Immunotherapy for triple-negative breast cancer: Existing challenges and exciting prospects," *Drug Resistance Updates*, vol. 32, pp. 1–15, May 2017.
- [59] M. S. Donepudi, K. Kondapalli, S. J. Amos, and P. Venkateshan, "Breast cancer statistics and markers," *J. Cancer Res. Therapeutics*, vol. 10, pp. 506–511, Jul./Sep. 2014.
- [60] G. Bertoli, C. Cava, and I. Castiglioni, "MicroRNAs: New biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer," *Theranostics*, vol. 5, no. 10, pp. 1122–1143, 2015.
- [61] S. Suman, T. Basak, P. Gupta, S. Mishra, V. Kumar, and S. Sengupta, "Quantitative proteomics revealed novel proteins associated with molecular subtypes of breast cancer," *J. Proteomics*, vol. 148, pp. 183–193, Oct. 2016.
- [62] Q. Wu, Z. Lu, H. Li, J. Lu, L. Guo, and Q. Ge, "Next-generation sequencing of microRNAs for breast cancer detection," *J. Biomed. Biotechnol.*, vol. 2011, May 2011, Art. no. 597145.
- [63] L. Zhang and F. Ding, "Hsa\_circ\_0008945 promoted breast cancer progression by targeting miR-338-3p," *OncoTargets Therapy*, vol. 12, pp. 6577–6589, Aug. 2019.
- [64] R. Zheng, L. Pan, J. Gao, X. Ye, L. Chen, and X. Zhang, "Prognostic value of miR-106b expression in breast cancer patients," *J. Surgical Res.*, vol. 195, no. 11, pp. 158–165, May 2015.
- [65] T. He, F. Qi, L. Jia, S. Wang, N. Song, L. Guo, Y. Fu, and Y. Luo, "MicroRNA-542-3p inhibits tumour angiogenesis by targeting angiopoietin-2: MiR-542-3p inhibits tumour angiogenesis via Angpt2," *J. Pathol.*, vol. 232, no. 5, pp. 499–508, Apr. 2014.
- [66] H.-D. Zhang, L.-H. Jiang, J.-C. Hou, S.-L. Zhong, S.-Y. Zhou, L.-P. Zhu, J. Li, D.-D. Wang, D.-W. Sun, Z.-L. Ji, and J.-H. Tang, "Circular RNA hsa\_circ\_0052112 promotes cell migration and invasion by acting as sponge for miR-125a-5p in breast cancer," *Biomed. Pharmacotherapy*, vol. 107, pp. 1342–1353, Nov. 2018.
- [67] T.-H. Hsieh, C.-Y. Hsu, C.-F. Tsai, C.-Y. Long, C.-Y. Chai, M.-F. Hou, J.-N. Lee, D.-C. Wu, S.-C. Wang, and E.-M. Tsai, "MiR-125a-5p is a prognostic biomarker that targets HDAC4 to suppress breast tumorigenesis," *Oncotarget*, vol. 6, no. 1, pp. 494–509, Jan. 2015.
- [68] M. Umeh-Garcia, C. Simion, P.-Y. Ho, N. Batra, A. L. Berg, K. L. Carraway, A. Yu, and C. Sweeney, "A novel bioengineered miR-127 prodrug suppresses the growth and metastatic potential of triple-negative breast cancer cells," *Cancer Res.*, vol. 80, no. 3, pp. 418–429, Feb. 2020.
- [69] R. Elango, K. A. Alsaleh, R. Vishnubalaji, M. Manikandan, A. M. Ali, N. Abd El-Aziz, A. Altheyab, A. Al-Rikabi, M. Alfayez, A. Aldahmash, and N. M. Alajez, "MicroRNA expression profiling on paired primary and lymph node metastatic breast cancer revealed distinct microRNA profile associated with LNM," *Frontiers Oncol.*, vol. 10, p. 756, May 2020.
- [70] Y. Xu, Q. Dong, F. Li, Y. Xu, C. Hu, J. Wang, D. Shang, X. Zheng, H. Yang, C. Zhang, M. Shao, M. Meng, Z. Xiong, X. Li, and Y. Zhang, "Identifying subpathway signatures for individualized anticancer drug response by integrating multi-omics data," *J. Transl. Med.*, vol. 17, no. 1, p. 255, Aug. 2019.

- [71] C. Li, J. Han, Q. Yao, C. Zou, Y. Xu, C. Zhang, D. Shang, L. Zhou, C. Zou, Z. Sun, J. Li, Y. Zhang, H. Yang, X. Gao, and X. Li, "Subpathway-GM: Identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways," *Nucleic Acids Res.*, vol. 41, no. 9, p. e101, May 2013.
- [72] X. Lu, V. V. Jain, P. W. Finn, and D. L. Perkins, "Hubs in biological interaction networks exhibit low changes in expression in experimental asthma," *Mol. Syst. Biol.*, vol. 3, no. 1, p. 98, Jan. 2007.



**ZIYU NING** was born in Qiqihar, Heilongjiang, China, in 1982. She received the B.E. degree in electronic information science and technology from Heilongjiang University, in 2005, and the master's degree in signal and information processing from the Harbin University of Science and Technology, in 2009.

Since 2009, she has been working with the School of Medical Informatics, Daqing Campus, Harbin Medical University, Daqing, China. She is the author of three articles and five financial supports. Her research interests include the applications and exploitation of deep learning algorithm, biological feature extraction, the construction of biological networks, and integration and measurement of big data.



**SHUANG YU** was born in Harbin, China, in 1984. He received the B.S. degree in automation and the M.S. degree in detection technology and automatic equipment from the Harbin University of Science and Technology, Harbin, China, in 2007 and 2010, respectively, and the Ph.D. degree in computer science and technology from Harbin Engineering University, Harbin, in 2018.

From 2010 to 2012, he was a Teaching Assistant with the Harbin University of Science and Technology. Since 2012, he has been a Lecturer with the Harbin University of Science and Technology. He is the author of a book, 15 articles, and 11 inventions. His research interests include three-dimensional visual inspection and tissue optical characteristic imaging.

Dr. Yu is a member of the council in Sensor sub-Society of China Instrument and Control Society. He was a recipient of a Heilongjiang Provincial Science and Technology Award awarded by Heilongjiang Provincial Government, China.



**YANQIAO ZHAO** was born in Harbin, Heilongjiang, China, in 1983. He received the B.S. degree in measurement-control technology and instrumentation, and the M.S. and Ph.D. degrees in instrument science and technology from the Harbin Institute of Technology, Harbin, in 2006, 2008, and 2014, respectively.

From 2015 to 2018, he was a Lecturer, and since 2018, he has been an Associate Professor with the Harbin University of Science and Technology. His research interests include prediction of respiratory movement on thoracic and abdominal surface, digital signal processing, medical digital image processing, deep learning, neural networks, and Gaussian process regression.



**YANG WANG** was born in Harbin, Heilongjiang, China, in 1978. He received the B.S. degree in detection technology and instrumentation, and the Ph.D. degree in measuring and testing technologies and instruments theory from the Harbin University of Science and Technology, Harbin, in 2001 and 2014, respectively.

He worked as an Instructor with the Department of Measurement and Control Technology and Instruments, Harbin University of Science and Technology, in 2005, and moved up to an Associate Professor, in 2012, acting as a Master Supervisor later the same year. In 2018, he got a full professorship. He has worked for many years on 3D measurement by structured light and machine vision measurement. In recent years, he has presided over and completed nearly four research projects funded by provinces or Harbin city and been working on another one; published over 20 research articles, over ten of which were included in EI. Besides, he owns more than ten patents of invention and reaps 1 award in science and technology presented by provincial authorities.



**QI FAN** was born in Jixi, Heilongjiang, China, in 1992. She received the B.S. and M.S. degrees from the Harbin University of Science and Technology, in 2014 and 2016, respectively, where she is currently pursuing the Ph.D. degree with the School of Measurement and Communication. Her current research interests include three-dimensional visual inspection and respiratory movement prediction.



**TIANMING SONG** was born in Harbin, Heilongjiang, China, in 1989. He received the B.S. degree from the Heilongjiang Institute of technology, in 2012, and the M.S. degree from the Harbin University of Science and Technology, Harbin, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Measurement and Communication. His current research interests include three-dimensional visual inspection and cerenkov luminisence imaging.



**XIAOYANG YU** was born in Harbin, China, in 1962. He received the B.S. degree in electromagnetic measurement technology and instruments from the Harbin Institute of Electrical Engineering, Harbin, in 1984, and the M.S. and Ph.D. degrees in precision instruments and machinery from the Harbin Institute of Technology, Harbin, in 1989 and 1998, respectively.

From 1984 to 1991, from 1991 to 1995, and from 1995 to 1999, he was a Teaching Assistant, a Lecturer, and an Associate Professor with the Harbin University of Science and Technology, Harbin. Since 1999, he has been a Professor with the Harbin University of Science and Technology. He is the author of three books, more than 200 articles, and more than 40 inventions. His research interests include three-dimensional visual inspection, tissue optical characteristic imaging, and image processing. He is an Editorial Board Member of the journal of Electric Machines and Control, the journal of Harbin University of Science and Technology, and the journal of Electrical measurement and instrument.

Dr. Yu is a Vice Chairman in Heilongjiang Province Instrumentation Society and a member of the council in China Instrument and Control Society. He was a recipient of eight Heilongjiang Provincial Science and Technology Awards awarded by Heilongjiang Provincial Government, China.

...