

Mohammad Javad Bazaryari and Seyed Hamid Aghaee-Bakhtiari*

MiRNA target enrichment analysis of co-expression network modules reveals important miRNAs and their roles in breast cancer progression

<https://doi.org/10.1515/jib-2022-0036>

Received June 26, 2022; accepted March 21, 2023; published online December 25, 2024

Abstract: Breast cancer has the highest incidence and is the fifth cause of death in cancers. Progression is one of the important features of breast cancer which makes it a life-threatening cancer. MicroRNAs are small RNA molecules that have pivotal roles in the regulation of gene expression and they control different properties in breast cancer such as progression. Recently, systems biology offers novel approaches to study complicated biological systems like miRNAs to find their regulatory roles. One of these approaches is analysis of weighted co-expression network in which genes with similar expression patterns are considered as a single module. Because the genes in one module have similar expression, it is rational to think the same regulatory elements such as miRNAs control their expression. Herein, we use WGCNA to find important modules related to breast cancer progression and use hypergeometric test to perform miRNA target enrichment analysis and find important miRNAs. Also, we use negative correlation between miRNA expression and modules as the second filter to ensure choosing the right candidate miRNAs regarding to important modules. We found hsa-mir-23b, hsa-let-7b and hsa-mir-30a are important miRNAs in breast cancer and also investigated their roles in breast cancer progression.

Keywords: systems biology; gene expression regulation; functional enrichment analysis; protein-protein interaction network; molecular complex detection

Abbreviations

BP	biological process
CC	cellular compartment
FDR	false discovery rate
GO	gene ontology
GSVA	gene set variation analysis
KEGG	Kyoto encyclopedia of genes and genomes
MCODE	molecular complex detection
MF	molecular function
PPI	protein-protein interaction
TCGA	the cancer genome atlas
TOM	topology overlap measure
WGCNA	weighted co-expression network analysis

*Corresponding author: Seyed Hamid Aghaee-Bakhtiari, Bioinformatics Research Center, Basic Sciences Research Institute, Mashhad University of Medical Sciences, Mashhad, Iran; and Department of Medical Biotechnology and Nanotechnology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran, E-mail: aghaeibh@mums.ac.ir. <https://orcid.org/0000-0002-6093-7837>

Mohammad Javad Bazaryari, Department of Medical Biotechnology and Nanotechnology, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran

 Open Access. © 2024 the author(s), published by De Gruyter.  This work is licensed under the Creative Commons Attribution 4.0 International License.

1 Introduction

Breast cancer is the most common cancer in the world and the fifth cause of death due to cancer. Among the women population, it accounts for 25 % of cancer incidence and 16.6 % of cancer mortality [1]. There are several risk factors associated with the development and progression of breast cancer. Factors such as familial history of breast cancer, history of chest radiation therapy and hormone replacement therapy can increase the chance of cancer development. Besides, genetic risk factors also play an important role in breast cancer. Inherited mutations in several genes such as CHEK2, CDH1, ATM, and TP53 are known to be breast cancer risk factors. Mutations in BRCA1 and BRCA2 are the most common cause of hereditary breast cancer and the most notable genetic risk factors [2]. Because of its heterogeneous nature, breast cancer cells can change instantly and adapt to their microenvironments. This leads to the tumor progression and metastasis which are associated to poor survival of patients [3]. There has been an improvement in the survival of patients because of screening and treatment strategies in breast cancer; but still median survival after metastasis or distant recurrence is 38 months [4].

miRNAs are small non coding RNA that play an important role in the regulation of gene expression. Mostly they have negative regulatory effects on their targets by interacting with the 3' untranslated regions (3' UTR) and inducing degradation [5]. Different phenomena in breast cancer such as progression, recurrence, metastasis and resistance to chemotherapy are influenced by these master regulators of gene expression. According to their effects on cancer-related pathways, there are two types of miRNAs: oncogenic miRNAs which are upregulated in cancer tissues and tumor suppressor miRNAs which are downregulated in cancer tissues [6]. For instance miR-200c and miR-141 are both oncogenic miRNAs in breast cancer and are associated with invasion and metastasis. On the other hand, miR-195 is a tumor suppressor miRNA and has anti-metastatic and anti-invasion effect on breast cancer cells [7].

Single miRNA can target multiple mRNAs and also the same gene can be targeted by several miRNAs [8], thus holistic approaches are needed to understand their regulatory effects. Systems biology which has been introduced two decades ago, offers various approaches to understand biological systems in more dynamic and holistic ways. High-throughput technologies like DNA sequencing and RNA sequencing make system-level understanding of biological phenomena more feasible and these system-level insights can be used for drug discovery [9]. Network representation and analysis is one of the popular approaches in systems biology. There are several types of biological networks such as metabolic network, protein-protein interaction (PPI) network, and gene regulatory network. These networks can be constructed based on existing knowledge or can be driven directly from the data and most of the time they are represented as ball-stick diagrams in which components are shown as nodes and their interactions as edges. Analysis of these biomolecular networks can help us to understand the design principles of biological systems and give us a clue about the rational behind them. For instance, scale-free topology, in which connectivity distribution fits a power law is the most common topology of different biological networks. Scale-free topology makes biological networks more robust to random mutation and failures, and also makes them small-world networks, in which there are highly connected clusters connected to each other [10].

Co-expression network is one of the biological networks which has been widely used in recent years. Co-expression networks are based on similarity scores between pairs of genes and their analysis can help to identify and cluster genes with similar patterns of expression. They can be used to understand the biological function(s) of correlated genes and also find a new set of genes that may have important roles in specific biological processes [11]. Co-expression networks are data-driven and they can reveal a different set of insights in comparison to other types of biological network such PPI networks which are based on existing knowledge. Although there can be an overlap between the co-expression network and gene-regulatory network of the same biological system [12]. One of the most successful tools for co-expression network construction and analysis is R package Weighted co-expression network analysis (WGCNA). It recruits different algorithms to find highly correlated genes and cluster them into a module [13].

Several studies have investigated important molecules and pathways in breast cancer by using WGCNA. For instance, WGCNA has been used in several studies to find new important biomarkers and modules which are related to the survival outcome of patients with breast cancer [14, 15]. Also WGCNA has been used to find

important miRNAs and reveal their interaction with their targets by analyzing miRNA co-expression networks, mRNA co-expression networks, or both of them [16, 17]. One way to analyze and find important genes in co-expression networks that may gain little attention in breast cancer study, is enrichment analysis of co-expression network's modules to understand their possible common regulators. Genes in one module have the same expression pattern and this similarity can be due to the regulatory effect of one or more regulators. In this study we aim to examine miRNAs by using systems biology approach to understand their system-level effects in breast cancer progression. To do so, first we use co-expression network of mRNAs to investigate breast cancer and find progression-related important gene modules. Then we examine possible regulatory effects of miRNAs on those modules by miRNA target enrichment analysis.

2 Materials and methods

2.1 Data source and pre-processing

RNA-seq, miRNA-seq, and clinical data of TCGA-BRCA project were downloaded by R package TCGAbiolinks [18], and imported to R programming environment. To eliminate possible effects of treatments on molecular mechanisms, those patients who have received any pharmaceutical or radiation therapy were removed from dataset. From those treatment-naïve patients, tumor samples which have both RNA-seq and miRNA-seq data were chosen for further analysis. In addition, samples without stage information were removed. To filter out lowly expressed features, mRNAs and miRNAs which have less than 50 and 10 total counts number respectively among all samples were removed. Then sequencing data were normalized and log2 transformed. To find targets of miRNAs, Tarbase [19] and mirtarbase [20] which are both curated databases of mRNA-miRNA interactions were used. Both databases were merged into a single matrix and mRNAs that exist in the merged matrix were chosen and used to construct co-expression network. Before network construction, the quality of mRNA data was assessed by hierarchical clustering analysis.

2.2 Co-expression network construction

WGCNA [13], which is a powerful R package was used to construct signed gene co-expression network. To do so, first Pearson's correlation coefficients between genes were calculated and used to construct adjacency matrix as follow:

$$a_{ij} = \left(\left| (1 + \text{cor}(x_i, x_j)) \right| / 2 \right)^\beta$$

where a is adjacency matrix, $\text{cor}(x_i, x_j)$ is Pearson's correlation coefficient between gene i and j , and β is soft threshold which eliminate weak correlation. To find appropriate soft threshold, expression matrix was transformed to similarity matrix and different values of β were tested by pickSoftThreshold function. Finally the smallest value which ensured constructed network has scale-free topology by $R^2 > 0.9$ was chosen. Next, topology overlap measure (TOM) was calculated according to adjacency matrix and different gene modules have been distinguished by using dynamicTreeCut method for dendrogram of dissimilarity (1-TOM). Minimum required size of detected modules were set to 30 ($\text{minModuleSize} = 30$) and modules with less than 0.25 height in dendrogram were merged together ($\text{MergeCutHeight} = 0.25$). Other parameters of blockwiseModules function of WGCNA package, were set as follow: genes were split to blocks with maximum number of 5,000 ($\text{maxBlockSize} = 5,000$); Pearson was used as correlation type ($\text{corType} = \text{"pearson"}$); $\text{deepSplit} = 2$ was used for moderate sensitivity over module splitting; minimum connectivity of 0.5 between module's eigengenes and 10 members was set as threshold to maintain detected modules ($\text{minCoreKME} = 0.5$, $\text{minCoreKMESize} = 10$); also module's member with connectivity less than 0.3 to corresponding module's eigengene were omitted from the module ($\text{minKMEtoStay} = 0.3$).

For each module, a representative value (eigengene) was calculated. To find important modules in cancer progression, correlations between module eigengenes and stage-related traits were calculated. Those modules

that have |correlation| bigger than 0.2 and p -value less than 0.05 with at least two traits related to progression (clinical stage and TNM) were considered as important modules.

2.3 MiRNA target enrichment and correlation analysis

To test whether targets of specific miRNA are enriched in specific network modules, hypergeometric test was performed as follows:

$$P_{HT} = \sum_{k=z}^n \frac{\binom{x}{k} \binom{N-x}{n-k}}{\binom{N}{n}}$$

P_{HT} : hypergeometric test p -value

x : the number of genes in a module

n : the number of targets of a specific miRNA

z : the number of module's members and miRNA's targets

N : total number of genes in the co-expression network

The total number of genes were 17,463 ($N = 17,463$) and the test was performed for 770 miRNAs against significant modules. Because multiple tests were performed, p -value was corrected by Benjamini–Hochberg method and False Discovery Rate (FDR) < 0.05 was chosen as the first threshold. As the second criterion, we used correlation between miRNAs and modules' eigengenes. Here we observed Pearson's correlation coefficients ranging from -0.56 to 0.89 and because miRNAs mostly have negative regulatory effect on gene expression, correlation less than -0.3 was set as the second threshold. Thus miRNA was flagged as significant if it had both significant number of targets in the specific module according to hypergeometric test result (FDR < 0.5) and negative correlation (coefficient < -0.3) with that module.

2.4 MiRNAs' targeted genes GSVA score

Gene set variation analysis (GSVA) [21] is a popular and open-source R package and estimates variation of gene set in each sample of the population. For each significant miRNA, their target genes in targeted module were considered as a gene set and their GSVA score was calculated and compared among different stages of breast cancer.

2.5 Pathway and gene ontology enrichment analysis

DAVID database [22] and Cytoscape plugin, ClueGO [23] were used to performed functional enrichment analysis. When using DAVID, we chose three types of GO terms (biological process [BP], cellular compartment [CC], and molecular function [MF]) and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways as resources. While biological process (BP) gene ontology was used to construct network by ClueGO.

2.6 PPI network construction and cluster analysis

In order to find molecular interaction and possible functional module in targeted gene list of each miRNA, protein-protein interaction network was constructed by CluePedia [24], a powerful plugin for Cytoscape [25]. STRING [26] was used as an interaction resource and the confidence cut-off was set to 0.7 (high confidence). Molecular complex detection (MCODE) algorithm [27] was used to find clusters in PPI networks. Node density cut-off = 0 and node score cut-off = 0.2 were set as MCODE parameters.

3 Results

3.1 Data pre-processing

After selecting treatment-naïve patients and removing samples with missing mRNA, miRNA, or stage data, 106 out of 1098 TCGA-BRCA samples were included. Minimum, average, and maximum age of included patients was 30, 64, and 90 years respectively. Chosen data comprised 21 stage I, 58 stage II, 23 stage III, and 4 stage IV samples. Initial mRNA count matrix had 56602 Ensembl gene ID and after removing lowly expressed genes, it has 17463 ID in common with the merged tarbase and mirtarbase interaction data which had information of 20480 Ensembl gene ID. There was no missing data in count matrix and no outlier was seen in the data. Hierarchical clustering analysis and stage and TNM heatmap can be seen in Figure 1A.

3.2 Co-expression network construction and module detection

To construct co-expression network, correlation coefficients were raised to $\beta = 16$. As shown in Figure 1.B, it fitted scale-free topology by $R^2 = 0.915$ and has mean connectivity = 8.11. Forty distinct modules were found in constructed network. Grey module which consists of unassigned genes was removed in further analysis. The smallest module (mediumpurple3) had 30 genes and the largest one (turquoise) had 1,539 genes. Figure 1c shows

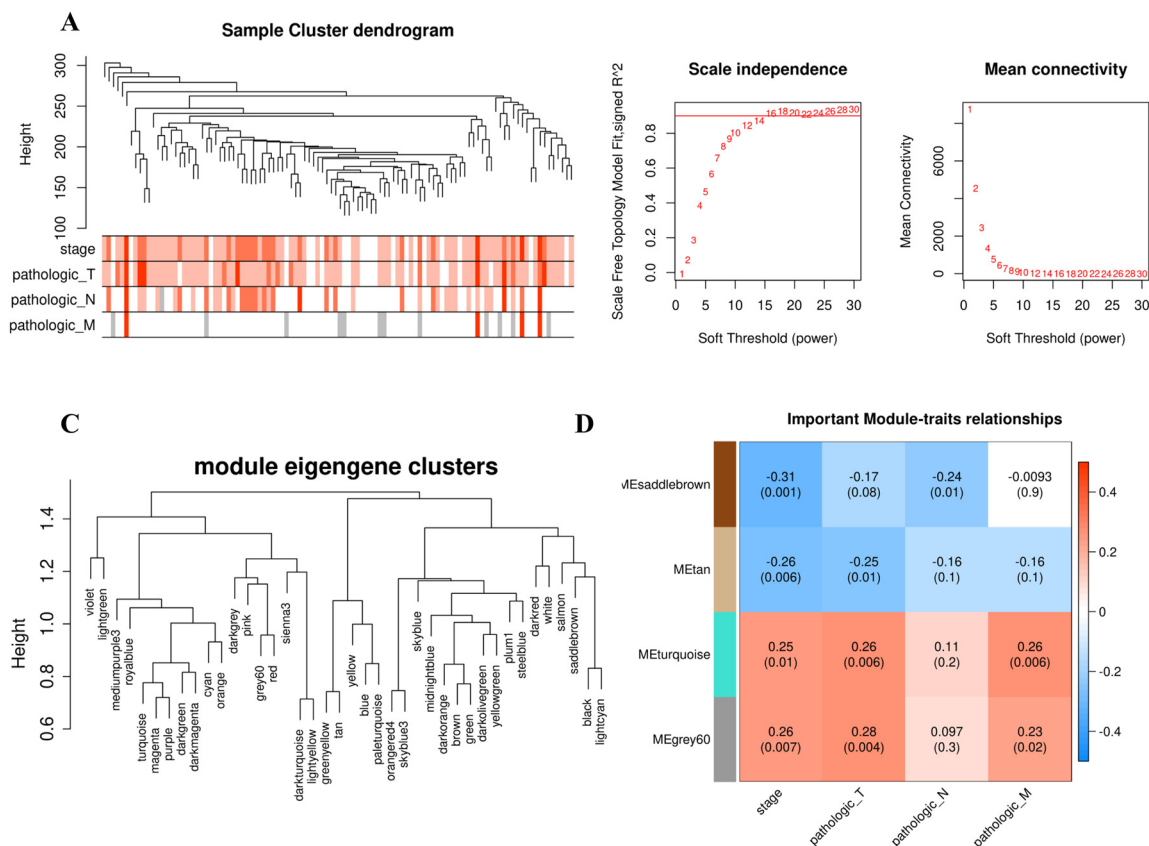


Figure 1: Dendrogram and heatmap represent distance between mRNA data samples and their corresponding stage and TNM data respectively (A) plots show scale free topology fitting R^2 and mean connectivity of different powers. Power = 16 fulfills $R^2 > 0.9$ and was chosen for network construction. (B) Dendrogram shows modules' eigengenes and their distances. (C) Heatmap demonstrates important modules and correlations between their eigengenes and stage and TNM. P -values are also mentioned in parentheses (D).

Table 1: Pathways, and GO analysis of important modules. For each module, the first five functional terms, their resources, and corresponding FDR values are reported.

Modules	Term: resource (FDR)				
saddlebrown	BP: cilium movement (3.485150E-11)	BP: flagellated sperm motility (1.117428E-05)	BP: cilium movement involved in cell motility (3.521998E-05)	BP: cilium assembly (3.887220E-05)	CC: motile cilium (9.376820E-05)
tan	CC: extracellular region (4.142144E-05)	CC: plasma membrane (8.357566E-05)	CC: lipid particle (9.767895E-05)	BP: retinol metabolic process (0.000116)	KEGG: PPAR signaling pathway (0.000241)
turquoise	BP: mitochondrial translational elongation (5.559936E-019)	BP: mitochondrial translational termination (5.559936E-019)	BP: mRNA splicing, via spliceosome (2.402600E-11)	BP: NIK/NF-kappaB signaling (5.491643E-09)	BP: mitochondrial respiratory chain complex I assembly (1.003084E-08)
grey60	CC: cytosol (2.783334E-06)	CC: nucleoplasm (0.004703)	CC: shelterin complex (0.0236595)	CC: nuclear telomere cap complex (0.0236595)	MF: protein binding (0.0302450)

hierarchical clustering of modules' eigengenes. Correlations between eigengenes and stage and TNM were calculated and among all modules, four modules had $|\text{correlation}| > 0.2$ with at least two of stage-related traits. Figure 1D shows stage_modules correlation heatmap and their corresponding p -values and top GO term and pathways for important modules can be seen in Table 1. Saddlebrown and tan modules have negative correlation with cancer progression and top results of their biological process gene ontology enrichment analysis were cilium movements and retinol metabolic process respectively. There were also two other important modules which show positive correlation with cancer progression. Turquoise module was one of them and mitochondrial processes such as mitochondrial translation and oxidative phosphorylation were among its top GO terms and pathways. Another one was grey60 module and it consisted of small group of genes that exist in cytoplasm and nucleoplasm.

3.3 miRNA targets enrichment analysis

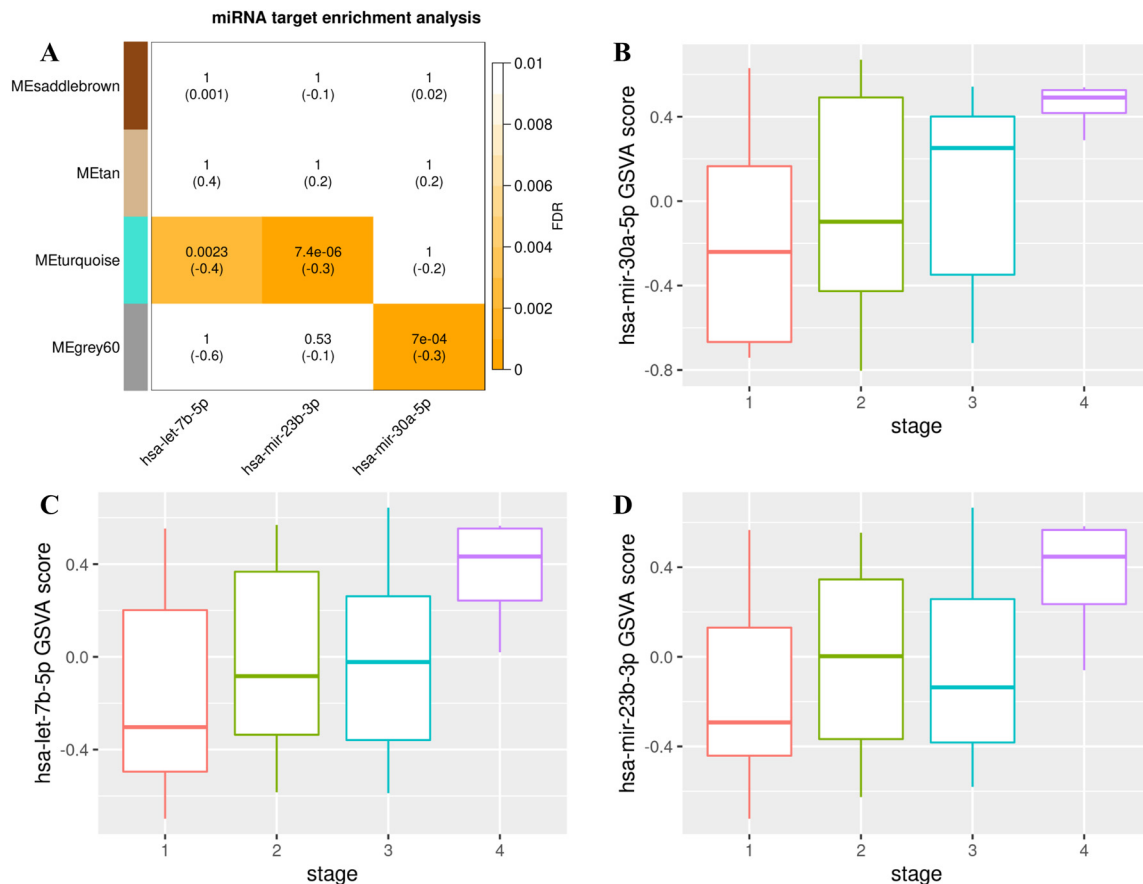
After performing hypergeometric test and calculating false discovery rate (FDR), 29 miRNA passed $\text{FDR} < 0.05$ threshold. Moreover, correlations between normalized miRNA expression and modules' eigengenes were calculated and correlation < -0.3 was used as the second level of threshold and three miRNA passed both threshold. Hsa-let-7b-5p and hsa-mir-23b-3p had 471 and 372 targets in the turquoise module respectively and hsa-mir-30a-5p had 28 targets in grey60 module. The turquoise module had 1,539 and the grey60 module had 92 genes and both modules were positively correlated with stage-related traits which shows suppression effect of mentioned miRNAs on the breast cancer progression. Important miRNAs' target enrichment results and their correlation with important modules are depicted in Figure 2A.

3.4 Important miRNA targets GSVA score

For each important miRNA, their targets were considered as a gene set and gene set variation analysis (GSVA) score was calculated and compared among different stages. As shown in Figure 2B–D, targeted gene set of each important miRNA (hsa-mir-23a-3p, hsa-let-7b-5p, and hsa-mir30a-5p) had a higher GSVA score in higher stages.

3.5 Pathway and gene ontology enrichment analysis of miRNAs' target gene sets

Because hsa-mir-23b-3p and hsa-let-7b-5p targets were significantly enriched in the same module (turquoise), their functional roles were investigated together and their targets were treated as a single gene list. As ClueGO result (Figure 3A) demonstrate, mir-23b and let-7b target related cellular activities as expected. Major cellular



activities such as mRNA metabolism and processing, mitochondrial gene expression, and oxidative phosphorylation enhance in higher breast cancer stages and it can be due to reducing inhibitory effects of mir-23b and let-7b. These miRNAs seem to have complementary roles in breast cancer. For instance, mRNA metabolism is mostly controlled by mir-23b and mRNA processing is affected by let-7b; or mitochondrial gene expression is regulated by let-7b and mir-23b regulates oxidative phosphorylation.

On the other hand, because hsa-mir-30a-5p targeted separated module (grey60), its functional roles were assessed individually. As represented in DAVID analysis (Table 2), targeted genes of mir-30a were mostly in cytosol and autophagy was a frequent term that had been seen in its functional analysis.

3.6 PPI network was constructed, and important modules were found

To investigate targets of mir-23b and let-7b holistically, PPI network was constructed by Cytoscape. Because targeted genes were originally co-expressed, it is possible to see different functional protein modules which work together in the same activity or even different genes which are subunits of one complex protein. To find these modules, MCODE was performed and two important modules with the highest scores were chosen. The first module which its MCODE score was 21.818 (Figure 3B), consists of 21 mitochondrial ribosomal proteins, mitochondrial translation initiation factor (MTIF1) and mitochondrial inner membrane protein OXA1L that interact with mitochondrial ribosomes and require for insertion of proteins to mitochondrial inner membrane. Second

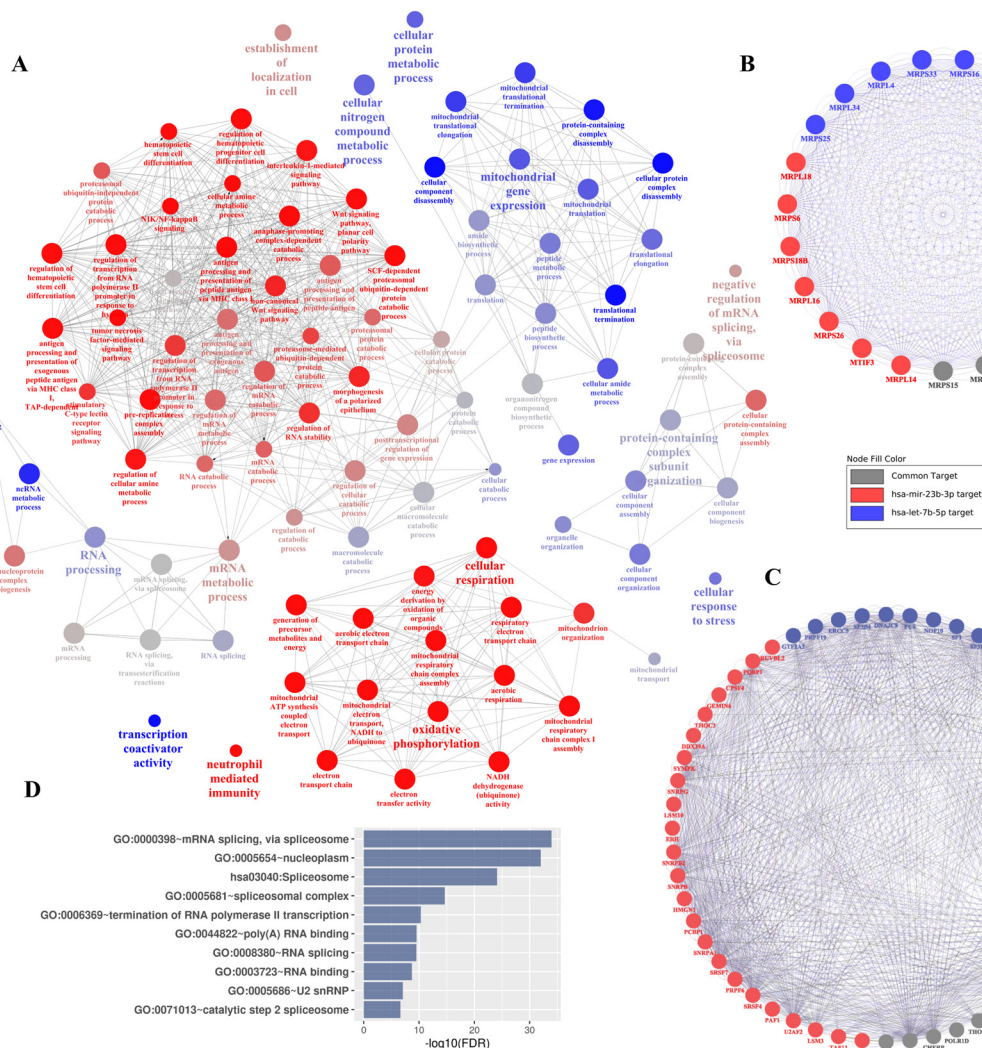
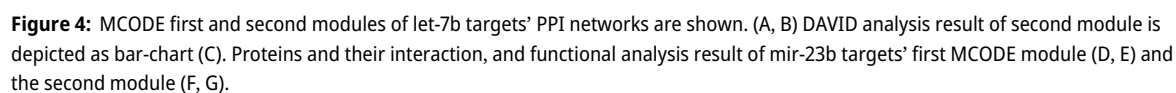


Figure 3: ClueGO result shows functional enrichment analysis for miR-23b and let-7b targets. Nodes show biological process GO terms with p -value < 0.05 and their colors indicate percentage of each miRNA's targets. Blue nodes show terms consist of more than 60 % let-7b targeted genes and red nodes show terms with more than 60 % genes targeted by miR-23b. (A) MCODE first two modules are shown; blue and red nodes indicate let-7b and miR-23b targeted genes respectively and gray nodes are genes which are targeted by both. (B, C) DAVID pathway enrichment analysis result for second MCODE modules are sorted by log10(FDR) and plotted as bar-chart (D).

Table 2: DAVID analysis result of miR-30a targets. Terms and pathways which have $FDR < 0.05$ are shown. Their corresponding number and symbol of genes are also represented.

Term	FDR	Count	Genes
GO:0000421~autophagosome membrane	0.001761	4	GABARAPL2, MAP1LC3B, CHMP1A, MAP1LC3B2
GO:0005829~cytosol	0.001761	18	GABARAPL2, USP10, AKTIP, NIP7, SIAH1, CSNK2A2, MAP1LC3B2, MEAK7, ZNRF1, SLC7A5, MAP1LC3B, PSMD7, OGFOD1, POLR2C, CHMP1A, RANBP10, VPS35, NAE1
GO:0097352~autophagosome maturation	0.006519	4	GABARAPL2, MAP1LC3B, CHMP1A, MAP1LC3B2
hsa04137:Mitophagy – animal	0.012069	4	GABARAPL2, MAP1LC3B, CSNK2A2, MAP1LC3B2
GO:0006995~cellular response to nitrogen starvation	0.018318	3	GABARAPL2, MAP1LC3B, MAP1LC3B2
hsa04139:Mitophagy – yeast	0.021799	3	GABARAPL2, USP10, CSNK2A2
GO:0016236~macroautophagy	0.026575	4	GABARAPL2, MAP1LC3B, CSNK2A2, MAP1LC3B2
GO:0000209~protein polyubiquitination	0.034069	5	ZNRF1, PSMD7, CBFB, AKTIP, SIAH1



module with score = 18.182 is represented in Figure 3C. As DAVID analysis results indicated (Figure 3D), its genes are mostly involved in spliceosome and there are several genes related to transcription in this module such as several RNA polymerase subunits.

To examine mir-23b and let-7b targets in more detail, we also constructed separated PPI networks for each of them and run MCODE algorithm separately to find their important network's clusters. For let-7b, the first module comprised mitochondrial ribosomal proteins and the second module consisted of spliceosomal proteins as same as modules in union PPI network of mir-23b and let-7b targets. Modules of let-7b targets' PPI networks can be seen in Figure 4A and B and DAVID analysis result of second module is shown in Figure 4C. As shown in Figure 4D and E, first module of mir-23b targets mostly consists of proteosomal genes and its DAVID analysis result shows its genes mostly participate in ubiquitin activity in cell cycle and ubiquitin-dependent catabolism processes. Second module of mir-23b targets and their enrichment analysis are depicted in Figure 4F and G, and most of its nodes are spliceosomal protein and it indicates mRNA splicing as important target of both mir-23b and let-7b.

4 Discussion

Breast cancer is the most common cancer around the world and because of its complexity and heterogeneity, its molecular mechanism is not fully understood. Breast cancer cells can grow constantly, invade near tissues, and metastasis to other locations of the body. Better understanding of these complications at the molecular level can enhance our ability in breast cancer diagnosis and treatment.

In recent years, systems biology has revolutionized the way we were looking at biological systems and expanded our knowledge in molecular biology. System-level insights can help us to understand diseases and their molecular mechanisms better and also it can lead us to discovering new drugs and find new prognostic and diagnostic biomarkers. MiRNAs are among the most important components which play pivotal roles in health and diseases and because each one of them has several targets, systems biology approach is suitable to study their regulatory effects.

In this study we investigated TCGA-BRCA mRNA and miRNA data of treatment-naïve patients to find important miRNAs in breast cancer progression. Due to regulatory effects of miRNAs on different targets, it's expected to see common expression patterns in their targets. To unravel these types of similarities in expression pattern, first we used WGCNA to construct co-expression network and find modules of correlated genes. Second we chose four important modules whose representative values (eigengenes) have statistically important correlation with stage; two of them had negative correlation and the other two modules had positive correlation with stage. Then we examine possible regulatory effects of miRNAs on those important modules by target enrichment analysis. We used hypergeometric test to do so and adjusted p-value and negative correlations between miRNAs and modules' eigengenes were used to find important miRNAs with possible regulatory effects on modules. We found three miRNAs by applying these criteria which have both significantly enriched targets and correlation with one of the four important modules. All of those three miRNAs targeted modules with positive correlations with stage.

4.1 Hsa-mir-30a-5p

One of the important miRNAs we found was hsa-mir-30a-5p and pathway enrichment analysis of its target showed autophagy as an important term. Several studies have shown mir-30a is a tumor suppressor miRNA in breast cancer and revealed its suppression effect on progression and metastasis [28–31]. Also its higher level of expression has been shown to be associated with better overall survival of patients [32]. Moreover there are also several studies that show mir-30a has an inhibitory effect on autophagy in different types of cancers including breast cancer [33, 34]. Most of these studies focused on beclin-1 as autophagy-related target of mir-30a, but here our results showed that the inhibitory effect of mir-30a goes behind beclin-1 and there are several autophagy-related genes which are targeted by mir-30a in breast cancer. Autophagy is a crucial process in

different types of cancer and it plays an important role in breast cancer tumorigenicity [35]. There are also several studies indicated that it can have both anti- and pro-metastatic effects on breast cancer cells. One of the possible explanations is autophagy can help unattached cells to survive before dormancy in new environments [36] and it can explain why the lower expression of mir-30a is associated with growth in unattached breast cancer cells [37].

4.2 Hsa-let-7b-5p

Hsa-let-7b-5p was another miRNA that we found to be important in breast cancer progression. Let-7 microRNA family have been shown to be dysregulated in several malignancy in both cancerous tissues and blood sample of patients [38]. This family have been indicated to regulate JAK-STAT3 pathway via STAT3 targeting, and MYC pathway by regulating expression of c-MYC. Also there are evidences which shows the regulatory role of let-7 family on breast cancer metastasis and stemness mostly by targeting oncogenes such as RAS, HMGA2 [39]. Regarding to let-7b, a member of let-7 family, there is also an *in vivo* evidence of their downregulation in breast cancer [40]. Dysregulation of let-7b have been also shown *in vitro*, and it's role in the apoptosis process has been indicated by breast cancer cell line transfection [41]. Moreover, a study indicated the effect of let-7b on cell growth and metastasis through reprogramming aerobic glycolysis via modulation of Hexokinase 2 (HK2) [42]. This had been followed by another evidence which indicated the role of metabolic reprogramming via let-7 family to be associated with tumor progression and autophagy [43].

Here we have found the targets of let-7b had been enriched in one of the important module in the breast cancer co-expression network (turquoise module) which had positive correlation with progression traits. Pathway enrichment analysis showed mitochondrial gene expression is in this module and targeted by let-7b. Moreover PPI network analysis of let-7b targets in the turquoise module, indicated this process maybe targeted mostly by targeting mitochondrial ribosomal gene expression. Localization of precursor and mature form of let-7 family microRNA have been indicated previously by *in situ* hybridization [44]. our results suggested metabolic reprogramming by let-7b could be due to upstream modulation of mitochondrial genes' expression indeed by controlling regulation of mitochondrial ribosomal proteins. Also mitochondrial ribosomal genes belonged to breast cancer progression related module (turquoise). This suggest the importance of mitochondrial ribosome in breast cancer cell growth and metastasis. There are studies that showed the importance of mitochondrial ribosomes in breast cancer and their subunits are differentially expressed in this disease [45, 46]. Here we found this dysregulation maybe caused by let-7b.

4.3 Hsa-mir-23b-3p

Also we found hsa-mir-23b-3p as another important miRNA and together with let-7b, they have significant number of targets in the same module. Mir-23b have been shown to be dysregulated in several malignancies. Mir-23b has been shown to have dual role in different cancer types and evidences have indicated both anti-tumor effects and tumor promoting effects for this microRNA [47]. Also in breast cancer, there are several studies show mir-23b dysregulation but there is also duality of its roles in breast cancer formation and progression. Some studies emphasized on oncogenic effects of mir-23b and their positive effect on breast cancer progression. For instance, knockout of mir-23b and mir-27b have been shown to reduce cell growth in breast cancer cell line (MCF7) but had no effect on migratory behavior of breast cancer cell line [48]. Another study indicated elevation in mir-23b expression in malignant breast cancer tissue comparing to normal tissue; but decreasing in higher stages comparing to lower stages [49]. Other studies addressed their tumor-suppressing roles especially on progression and metastasis. Using mir-23b transfection in breast cancer cell line, one study have shown regulatory role of mir-23b on several cytoskeletal genes' expression which cause cytoskeleton remodeling and increase motility of cancer cells [50]. Also another evidence indicated mir-23b stabilized epithelial phenotype and reverse Epithelial Mesenchymal Transition (EMT) which is an important behavior in metastasis [51].

Here we found mir-23b may have tumor-suppressing role in breast cancer and inhibits progression and metastasis. We have found mir-23b targets have been enriched in turquoise module and they were mostly

involved in oxidative phosphorylation process and NF-KB signaling pathway. Previous study showed suppression of mir-23b expression by c-myc can change metabolism program and enhances glutamine metabolism. Considering converting glucose to lactate in anaerobic metabolism in cancer cells (Warburg effect), glutamine metabolism is pivotal to maintaining mitochondrial metabolism in cancer cells [52]. This phenomenon (glutamine metabolism) have been shown to be an important behavior in breast cancer cells like other malignancy [53]. Here we have shown oxidative phosphorylation in mitochondria is correlated to breast cancer progression and mir-23b maybe the key regulatory element in this process. Also there are a line of evidences that have shown NF-KB pathway can change the expression of mir-23b by increasing its expression [54, 55]. Interestingly our result indicated the regulatory effect of mir-23b on NF-KB pathway. This have been also demonstrated in other studies [56, 57]. This would be suggesting feedback loop between the expression of mir-23b and NF-KB pathway. Further study is necessary to investigate relationship between mir-23b and NF-KB to unravel their regulatory effects on each other in breast cancer.

Also our results show that let-7b and mir-23b affect mRNA metabolism and processing such as mRNA splicing in breast cancer. Spliceosomes play important roles in breast cancer and their components have been shown to be related to breast cancer patients' survival [58]. To our knowledge, no study have demonstrated complementary role of let-7b and mir-23b in breast cancer. Also we showed several new processes which are targeted by these two miRNAs in breast cancer.

In summary, we identified mir-30a, mir-23b, and let-7b as important miRNAs related to breast cancer progression by enrichment analysis of WGCNA modules. They regulate expression of not just few specific genes but two progression-related modules in breast cancer which may contain even several hundreds of genes. We found mir-23b and let-7b targeted module contains metabolic-related genes and mir-30a targets several autophagy-related genes. Our finding revealed new mechanisms of these miRNAs in breast cancer and their possible roles in breast cancer progression.

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: M.J.B. and S.H.A.B. conceptualized the study. S.H.A.B. performed project administration and supervised the study. M.J.B. wrote the original draft. M.J.B. and S.H.A.B. wrote, reviewed, and edited the manuscript.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interest: All authors have no conflicts of interest to declare.

Research funding: This study was supported by Mashhad University of Medical Sciences grants no. 941302 and 992301.

Data availability: Not applicable.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clin* 2021;71:209–49.
2. Feng Y, Spezia M, Huang S, Yuan C, Zeng Z, Zhang L, et al. Breast cancer development and progression: risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes Dis* 2018;5:77–106.
3. Lüönd F, Tiede S, Christofori G. Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. *Br J Cancer* 2021;125:164–75.
4. Caswell-Jin JL, Plevritis SK, Tian L, Cadham CJ, Xu C, Stout NK, et al. Change in survival in metastatic breast cancer with treatment advances: meta-analysis and systematic review. *JNCI Cancer Spectr* 2018;2:pkv062.
5. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA biogenesis, mechanisms of actions, and circulation. *Front Endocrinol* 2018;9. <https://doi.org/10.3389/fendo.2018.00402>.
6. Wang W, Luo YP. MicroRNAs in breast cancer: oncogene and tumor suppressors with clinical potential. *J Zhejiang Univ - Sci B* 2015;16:18–31.
7. Loh HY, Norman BP, Lai KS, Rahman NM, Alitheen NBM, Osman MA. The regulatory role of MicroRNAs in breast cancer. *Int J Mol Sci* 2019;20:4940.
8. Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009;136:215–33.

9. Kitano H. Systems biology: a brief overview. *Science* 2002;295:1662–4.
10. Ma'ayan A. Introduction to network analysis in systems biology. *Sci Signal* 2011;4:tr5–tr.
11. Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. Learning from co-expression networks: possibilities and challenges. *Front Plant Sci* 2016;7. <https://doi.org/10.3389/fpls.2016.00444>.
12. Xulvi-Brunet R, Li H. Co-expression networks: graph properties and topological comparisons. *Bioinformatics* 2010;26:205–14.
13. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* 2008;9:559.
14. Jia R, Zhao H, Jia M. Identification of co-expression modules and potential biomarkers of breast cancer by WGCNA. *Gene* 2020;750:144757.
15. Tian Z, He W, Tang J, Liao X, Yang Q, Wu Y, et al. Identification of important modules and biomarkers in breast cancer based on WGCNA. *OncoTargets Ther* 2020;13:6805–17.
16. Bao C, Lu Y, Chen J, Chen D, Lou W, Ding B, et al. Exploring specific prognostic biomarkers in triple-negative breast cancer. *Cell Death Dis* 2019;10:807.
17. Wani N, Barh D, Raza K. Modular network inference between miRNA–mRNA expression profiles using weighted co-expression network analysis. *J Integr Bioinform* 2021;18:20210029.
18. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* 2016;44:e71.
19. Sethupathy P, Corda B, Hatzigeorgiou AG. TarBase: a comprehensive database of experimentally supported animal microRNA targets. *RNA* 2006;12:192–7.
20. Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Res* 2020;48:D148–d54.
21. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinf* 2013;14:7.
22. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003;4:R60.
23. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009;25:1091–3.
24. Bindea G, Galon J, Mlecnik B. CluePedia cytoscape plugin: pathway insights using integrated experimental and in silico data. *Bioinformatics* 2013;29:661–3.
25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.
26. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47:D607–13.
27. Bader GD, Hogue CWV. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf* 2003;4:2.
28. Fu J, Xu X, Kang L, Zhou L, Wang S, Lu J, et al. miR-30a suppresses breast cancer cell proliferation and migration by targeting Eya2. *Biochem Biophys Res Commun* 2014;445:314–9.
29. Wang X, Qiu H, Tang R, Song H, Pan H, Feng Z, et al. miR-30a inhibits epithelial-mesenchymal transition and metastasis in triple-negative breast cancer by targeting ROR1. *Oncol Rep* 2018;39:2635–43.
30. Xiao B, Shi X, Bai J. miR-30a regulates the proliferation and invasion of breast cancer cells by targeting Snail. *Oncol Lett* 2019;17:406–13.
31. Xiong J, Wei B, Ye Q, Liu W. MiR-30a-5p/UBE3C axis regulates breast cancer cell proliferation and migration. *Biochem Biophys Res Commun* 2019;516:1013–8.
32. Kawaguchi T, Yan L, Qi Q, Peng X, Gabriel EM, Young J, et al. Overexpression of suppressive microRNAs, miR-30a and miR-200c are associated with improved survival of breast cancer patients. *Sci Rep* 2017;7:15945.
33. Yu Y, Cao L, Yang L, Kang R, Lotze M, Tang D. microRNA 30A promotes autophagy in response to cancer therapy. *Autophagy* 2012;8:853–5.
34. Zhu H, Wu H, Liu X, Li B, Chen Y, Ren X, et al. Regulation of autophagy by a beclin 1-targeted microRNA, miR-30a, in cancer cells. *Autophagy* 2009;5:816–23.
35. Gong C, Bauvy C, Tonelli G, Yue W, Delomélie C, Nicolas V, et al. Beclin 1 and autophagy are required for the tumorigenicity of breast cancer stem-like/progenitor cells. *Oncogene* 2013;32:2261–72.
36. Zhou Y, Rucker EB 3rd, Zhou BP. Autophagy regulation in the development and treatment of breast cancer. *Acta Biochim Biophys Sin* 2016;48:60–74.
37. Ouzounova M, Vuong T, Ancey P-B, Ferrand M, Durand G, Le-Calvez KF, et al. MicroRNA miR-30 family regulates non-attachment growth of breast cancer cells. *BMC Genom* 2013;14:139.
38. Chirshv E, Oberg KC, Ioffe YJ, Unternaehrer JJ. Let-7 as biomarker, prognostic indicator, and therapy for precision medicine in cancer. *Clin Transl Med* 2019;8:24.
39. Thammaiah CK, Jayaram S. Role of let-7 family microRNA in breast cancer. *Noncoding RNA Res* 2016;1:77–82.
40. Volinia S, Galasso M, Sana ME, Wise TF, Palatini J, Huebner K, et al. Breast cancer signatures for invasiveness and prognosis defined by deep sequencing of microRNA. *Proc Natl Acad Sci USA* 2012;109:3024–9.

41. Bozgeyik E. Bioinformatic analysis and in vitro validation of let-7b and let-7c in breast cancer. *Comput Biol Chem* 2020;84:107191.
42. Ye Q, Li L, Zhang X, Lin Y, Ren X, Xie T, et al. Let-7b-5p inhibits breast cancer cell growth and metastasis via repression of hexokinase 2-mediated aerobic glycolysis. *Cell Death Discov* 2023;9:114.
43. Li CH, Liao CC. The metabolism reprogramming of microRNA let-7-mediated glycolysis contributes to autophagy and tumor progression. *Int J Mol Sci* 2021;23. <https://doi.org/10.3390/ijms23010113>.
44. Barrey E, Saint-Auret G, Bonnamy B, Damas D, Boyer O, Gidrol X. Pre-microRNA and mature microRNA in human mitochondria. *PLoS One* 2011;6:e20220.
45. Oviya RP, Gopal G, Shirley SS, Sridevi V, Jayavelu S, Rajkumar T. Mitochondrial ribosomal small subunit proteins (MRPS) MRPS6 and MRPS23 show dysregulation in breast cancer affecting tumorigenic cellular processes. *Gene* 2021;790:145697.
46. Zhang Q, Liang Z, Gao Y, Teng M, Niu L. Differentially expressed mitochondrial genes in breast cancer cells: potential new targets for anti-cancer therapies. *Gene* 2017;596:45–52.
47. Grossi I, Salvi A, Baiocchi G, Portolani N, De Petro G. Functional role of microRNA-23b-3p in cancer biology. *MicroRNA* 2018;7:156–66.
48. Hannafon BN, Cai A, Calloway CL, Xu YF, Zhang R, Fung KM, et al. miR-23b and miR-27b are oncogenic microRNAs in breast cancer: evidence from a CRISPR/Cas9 deletion study. *BMC Cancer* 2019;19:642.
49. Taha M, Mitwally N, Soliman AS, Yousef E. Potential diagnostic and prognostic utility of miR-141, miR-181b1, and miR-23b in breast cancer. *Int J Mol Sci* 2020;21. <https://doi.org/10.3390/ijms21228589>.
50. Pellegrino L, Stebbing J, Braga VM, Frampton AE, Jacob J, Buluwela L, et al. miR-23b regulates cytoskeletal remodeling, motility and metastasis by directly targeting multiple transcripts. *Nucleic Acids Res* 2013;41:5400–12.
51. Pellegrino L, Krell J, Roca-Alonso L, Stebbing J, Castellano L. MicroRNA-23b regulates cellular architecture and impairs motogenic and invasive phenotypes during cancer progression. *BioArchitecture* 2013;3:119–24.
52. Gao P, Tchernyshyov I, Chang TC, Lee YS, Kita K, Ochi T, et al. c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. *Nature* 2009;458:762–5.
53. Penkert J, Ripperger T, Schieck M, Schlegelberger B, Steinemann D, Illig T. On metabolic reprogramming and tumor biology: a comprehensive survey of metabolism in breast cancer. *Oncotarget* 2016;7:67626–49.
54. Donadelli M, Dando I, Fiorini C, Palmieri M. Regulation of miR-23b expression and its dual role on ROS production and tumour development. *Cancer Lett* 2014;349:107–13.
55. Naidu S, Shi L, Magee P, Middleton JD, Laganá A, Sahoo S, et al. PDGFR-modulated miR-23b cluster and miR-125a-5p suppress lung tumorigenesis by targeting multiple components of KRAS and NF-κB pathways. *Sci Rep* 2017;7:15441.
56. Hu R, O'Connell RM. MiR-23b is a safeguard against autoimmunity. *Nat Med* 2012;18:1009–10.
57. Sun X, Zheng X, Zhang X, Zhang Y, Luo G. Exosomal microRNA-23b-3p from bone marrow mesenchymal stem cells maintains T helper/Treg balance by downregulating the PI3k/Akt/NF-κB signaling pathway in intracranial aneurysm. *Brain Res Bull* 2020;165:305–15.
58. An J, Luo Z, An W, Cao D, Ma J, Liu Z. Identification of spliceosome components pivotal to breast cancer survival. *RNA Biol* 2021;18:833–42.