

## RESEARCH ARTICLE

## miR-361-5p as a promising qRT-PCR internal control for tumor and normal breast tissues

Sogol Ghanbari<sup>1‡</sup>, Adel Salimi<sup>1‡</sup>, Saeid Rahmani<sup>2</sup>, Nahid Nafissi<sup>3</sup>, Ali Sharifi-Zarchi<sup>2</sup>, Seyed Javad Mowla<sup>1\*</sup>

**1** Molecular Genetics Department, Biological Sciences Faculty, Tarbiat Modares University, Tehran, Iran, **2** Computer Engineering Department, Sharif University of Technology, Tehran, Iran, **3** Surgical Department, School of Medicine, Iran University of Medical Sciences, Tehran, Iran

‡ These authors are co-first authors and contributed equally to this work.

\* [sjmowla@modares.ac.ir](mailto:sjmowla@modares.ac.ir)



## Abstract

## Background

One of the most widely used evaluation methods in miRNA experiments is qRT-PCR. However, selecting suitable internal controls (IC) is crucial for qRT-PCR experiments. Currently, there is no consensus on the ICs for miRNA qRT-PCR experiments in breast cancer. To this end, we tried to identify the most stable (the least expression alteration) and promising miRNAs in normal and tumor breast tissues by employing TCGA miRNA-Seq data and then experimentally validated them on fresh clinical samples.

## Methods

A multi-component scoring system was used which takes into account multiple expression stability criteria as well as correlation with clinical characteristics. Furthermore, we extended the scoring system for more than two biological sub-groups. TCGA BRCA samples were analyzed based on two grouping criteria: Tumor & Normal samples and Tumor subtypes. The top 10 most stable miRNAs were further investigated by differential expression and survival analysis. Then, we examined the expression level of the top scored miRNA (hsa-miR-361-5p) along with two commonly used ICs hsa-miR-16-5p and U48 on 34 pairs of Primary breast tumor and their adjacent normal tissues using qRT-PCR.

## Results

According to our multi-component scoring system, hsa-miR-361-5p had the highest stability score in both grouping criteria and hsa-miR-16-5p showed significantly lower scores. Based on our qRT-PCR assay, while U48 was the most abundant IC, hsa-miR-361-5p had lower standard deviation and also was the only IC capable of detecting a significant up-regulation of hsa-miR-21-5p in breast tumor tissue.

## Conclusions

miRNA-Seq data is a great source to discover stable ICs. Our results demonstrated that hsa-miR-361-5p is a highly stable miRNA in tumor and non-tumor breast tissue and we

## OPEN ACCESS

**Citation:** Ghanbari S, Salimi A, Rahmani S, Nafissi N, Sharifi-Zarchi A, Mowla SJ (2021) miR-361-5p as a promising qRT-PCR internal control for tumor and normal breast tissues. PLoS ONE 16(6): e0253009. <https://doi.org/10.1371/journal.pone.0253009>

**Editor:** Elda Tagliabue, Fondazione IRCCS Istituto Nazionale dei Tumori, ITALY

**Received:** February 14, 2021

**Accepted:** May 27, 2021

**Published:** June 8, 2021

**Copyright:** © 2021 Ghanbari et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the manuscript and its [Supporting information](#) files.

**Funding:** The author(s) received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.

recommend it as a suitable reference gene for miRNA expression studies in breast cancer. Additionally, although hsa-miR-16-5p is a commonly used IC, it's not a suitable one for breast cancer studies.

## Introduction

Identifying dysregulated genes involved in the carcinogenesis and tumor progression is an important component in cancer research [1]. Recently, high-throughput sequencing techniques is being used to conduct the whole transcriptome profiling, however, the main molecular diagnosis tests in clinic still relies on the cheaper quantitative real-time RT-PCR technique [2]. qRT-PCR is one of the most reliable and powerful tools which promises high specificity, sensitivity, and reproducibility to precisely detect the changes in gene expressions in a broad range of clinical samples, collected under different conditions [3, 4]. However, the precision of the results largely depends on choosing correct internal control to normalize the data [5]. Ideally, the expression of a reliable internal control (IC) gene should not be altered in tested tissues or cells under experimental conditions. The ideal internal control gene is universally valid, with a stable expression level across all tissue samples, cells, and experimental treatments. Although such an ideal IC has not yet been found [6, 7].

miRNAs represent an important new class of regulatory biomolecules that play fundamental biological roles including development, differentiation, apoptosis, and metabolism. Over the past decade numerous studies have been done on misregulation of miRNAs expression in various cancers, and the suitability of those miRNAs as novel biomarkers to diagnose, classify, prognose and treat patients with cancer [8, 9]. To use miRNAs as a biomarker in the clinic, it is crucial to standardize miRNA testing and make it reliable and reproducible in routine diagnostic applications. Over the past several years, some technical approaches had been employed to quantify miRNAs in clinical samples, among them, qRT-PCR has become the most popular one. Because of its sensitivity and specificity, it can detect low copy number of precursor and mature miRNA [10].

There is currently no consensus on suitable ICs for quantitative analysis of miRNA in human breast tissue. It is also clear that the traditionally used *GAPDH* and  $\beta$ -*actin* (*ABTB*) house-keeping genes are less validated as suitable ICs for miRNAs quantification [11, 12]. Ribosomal RNAs are another choice, however, they are expressed at much higher levels than target RNAs, making it difficult to normalize rare transcripts and rRNAs at same biological or clinical samples [13]. Additionally, although small-nucleolar RNAs such as RNU6, RNU6B and RNU48 are frequently used as reference genes, there existed evidence that snoRNAs can introduce some bias to miRNA expression in cancer studies [14]. Moreover, it has been argued that miRNAs must be normalized with internal genes from the same RNA class [12, 15]. Up to date, only a few candidate reference miRNAs (miR-16 and let-7) have been reported for breast tissue miRNA quantification [11, 12, 14, 15]. Conversely there has been evidence about the role of aforementioned ICs in cell differentiation and carcinogenesis [16–18] reinforcing the need for a better miRNA IC. To choose a study-specific IC, usually, a panel of 5–20 ICs are tested using qRT-PCR and analyzed by tools like Normfinder, geNorm and BestKeeper [5, 19, 20]. However, this approach is limited to a small number of ICs and the logic behind the assumption of these methods is that the mean expression of ICs is constant between samples which is not satisfied when the number of ICs is too low in a qRT-PCR experiment. Additionally, most researchers simply choose their IC based on literature [21].

Thanks to accession to the total read count, whole transcriptomic RNA-Seq data is a great source to discover stably expressed reference genes. Unlike qRT-PCR data, within samples technical variation can be normalized without control genes. Some studies have taken advantage of this source of data to identify stably expressed reference genes in various diseases, and have reported its adequacy for this purpose [22–24].

This study aimed to find the most suitable references for normalizing qRT-PCR data on miRNA expression in breast tissues, using a multi-component scoring system on miRNA-Seq data. Our scoring method takes into account multiple expression stability criteria as well as finding their correlation with clinical aspects of the samples. The top 10 most stable and promising miRNAs were introduced and evaluated. hsa-miR-361-5p was the best miRNA in overall scores for both tumor & normal samples as well as tumor subtype grouping criteria. We validated our findings using qRT-PCR assay and revealed hsa-miR-361-5ps superiority over hsa-miR-16-5p and U48.

## Materials & methods

### miRNA-Seq dataset

The Cancer Genome Atlas (TCGA) is a landmark cancer genomics program which includes molecular datasets for various type of cancer tissues [25]. The breast cancer (BRCA) miRNA expression data of TCGA was obtained via TCGAblinks package v2.12.3 [26] specifying data type as Isoform Expression Quantification. 1097 tumor tissue and 104 adjacent normal tissue samples were obtained. miRNA names were annotated by miRBase v21 using miRBaseConverter v1.8 [27]. CPM (read counts per million) values were used as a measure of expression level of miRNAs in stability analysis.

### Multi-component reference gene scoring system

A multi-component scoring system introduced by Krasnov et al. [28] was used in the context of miRNA expression. We modified and extended it in order to give it the ability of handling more than two subgroups. This system consists of several scoring components  $S_i$  which examine gene's expression value and its dispersion in subgroups and pooled samples, as well as correlation with clinical and pathological features of patients. Detailed description of each component is presented in Table 1. Overall expression stability score  $S^{exp}$  is calculated as weighted geometric mean of scoring components Eq (1).

$$S^{Exp} = \left( \prod_{i=1}^N (S_i + CA_i)^{W_i} \right)^{1/\sum_{i=1}^N W_i} \quad (1)$$

Here:  $W_i$  specifies each component's importance.

$CA_i$  is a constant to prevent zero values of  $S_i$  from making whole expression zero.

$N$  is the number of components.

Each component is calculated based on a parameterized (1-sigma)-like function as in Eq (2):

$$S_i = \frac{100}{1 + S_q + \left( \frac{\max(x-IV,0)}{IP-IV} \right)^{CS}} \quad (2)$$

Detailed description can be found in the associated paper [28].

**Table 1. Components of scoring function (obtained from associated paper).**

Component	factor	Variable (x = . . .)*	IV	IP	CS	Sq	CA	W	Number of Times applied
$S_{DP}$	T-N expression level difference (pooled samples)	Abs ( $\log_2FC_p$ ) <sub>10-90</sub>	0.05	0.25	2.5	1	0	4	1 (all samples)
$S_{DL}$	T-N expression level difference (paired samples)	Abs (Average ( $\log_2FCL$ ) <sub>10-90</sub> )							1 (paired samples)
$S_{DoO}$	T-N expression level difference: outliers, overexpression	Abs (Average ( $\log_2FCL$ ) <sub>90-100</sub> )	0.1	0.7	2.5	1	10	1	1 (paired samples)
$S_{DoU}$	T-N expression level difference: outliers, underexpression	Abs (Average ( $\log_2FCL$ ) <sub>0-10</sub> )							1 (paired samples)
$S_{DLc}$	Cumulative T-N expression difference among paired samples	Average (Abs ( $\log_2FCL$ ) <sub>10-90</sub> )	0.1	0.5	2.5	1	5	2	1 (paired samples)
$S_{ESTD}$	Expression level stability: standard deviation	StDev (CPM) <sub>10-90</sub> / Average(CPM) <sub>10-90</sub>	0.1	0.3	2	1	5	1.5	2 (all samples: normal and tumor)
$S_{EoH}$	Expression level stability: outliers (high expression)	$\log_2$ (Average(CPM) <sub>90-100</sub> / Average (CPM) <sub>10-90</sub> )	0.1	0.7	2.5	1	5	0.75	2 (all samples: normal and tumor)
$S_{EoL}$	Expression level stability: outliers (low expression)	$\log_2$ (Average(CPM) <sub>10-90</sub> / Average (CPM) <sub>0-10</sub> )							2 (all samples: normal and tumor)
$S_{EA}$	Average expression level	1/ $\log_2$ (CPM) <sub>10-90</sub>	0.30**	0.43**	3	1	0	6	1 (all tumor samples)
$S_{Cp}$	Correlations of expression with clinical parameters (p-values)	$-\log_2$ (p-value)	2	4	3	0.3	5	0.3	15 (3 × 5; 3: CPM <sub>10-90</sub> all tumor samples, CPM <sub>10-90</sub> all normal samples, ( $\log_2FCL$ ) <sub>10-90</sub> ; 5: pathologic TNM classification, pathologic stage, neoplasm cancer status)
$S_{Cr}$	Correlations of expression with clinical parameters ( $r_s$ )	Abs ( $r_s$ )	0.1	0.25	2.5	0.3	5	0.2	15 (the same as above)

IV, ideal value; IP, inflection point; CS, curve slope; Sq, “squeeze”; CA, constant add; W, weight; Abs (. . .), absolute value; Average (. . .), mean value; CPM, counts per million, gene expression level; FCP, ratio of the average CPM in a pool of tumor samples to the average CPM in a pool of normal samples; FCL, ratio of CPM values between tumor and matched normal tissue (per each paired sample); StDev (. . .), standard deviation;  $r_s$ , Spearman’s correlation coefficient.

\* Percentiles, which were taken into calculation, are indicated as a subscript.

\*\* These two parameters are different from the associated paper. as we filtered out miRNAs based on their expression level threshold of 5.9 CPM, here IV and IP were set as  $1/\log(10)$  and  $1/\log(5)$ .

<https://doi.org/10.1371/journal.pone.0253009.t001>

## Extended scoring system for more than two subgroups

In order to extend the scoring system for multiple subgroups, following manipulations were carried out:

1. Components associated with paired samples were removed.
2. For  $S_{DP}$ ,  $\log_2FC$  was calculated between every two subgroups of the data.
3. Beside  $S_{EA}$ , all other scoring components were applied to each subgroup separately and the weights were distributed according to the number of subgroups.

## Sample collection

Primary breast tumor tissues (n = 68) were obtained from patients undergoing surgery, at Khatam-al-Anbia and Rasoul-Akram Hospitals, Tehran, Iran. This research involved collecting human tissues with no experimenting on human subjects or animals. In vitro experiments on commercial cell lines and pathological samples were approved by Ferdowsi University of

Table 2. Clinicopathologic characteristics of 34 breast cancer patients.

Age (mean ± SD)	50 ± 10	TNM classification	
Tumor Subtype	Cases (n)	T = 1	Cases (n)
luminal A	22	T = 2	12
luminal B	2	T = 3	3
Her2 Overexpressed	2	Undefined T	6
TNBC	3	N = 0	17
Undefined Subtype	5	N = 1	5
Grade	Cases (n)	N = 2	Cases (n)
I	4	N = 3	1
II	16	Undefined N	6
III	9	M = 0	34
Undefined Grade	5		
Stage	Cases (n)		
I	9		
IIA	10		
IIB	3		
IIIA	5		
IIIC	1		
Undefined Stage	6		

T, N and M refer to the primary tumor size, nodal status and distant metastases status according to the TNM classification system. TNBC, Triple Negative Breast Cancer.

<https://doi.org/10.1371/journal.pone.0253009.t002>

Mashhad (code number: IR.UM.REC.1399.104). Samples were categorized into 34 pairs of breast tumor and their adjacent apparently normal tissues from October 2018 to June 2019. The studied specimens were examined by pathologists and classified according to the standard histopathological parameters. Tissues were immediately snap-frozen in liquid nitrogen and stored at -80°C until RNA extraction. Clinicopathologic characteristics of patients are summarized in Table 2 as well as S1 Table.

### Total RNA isolation

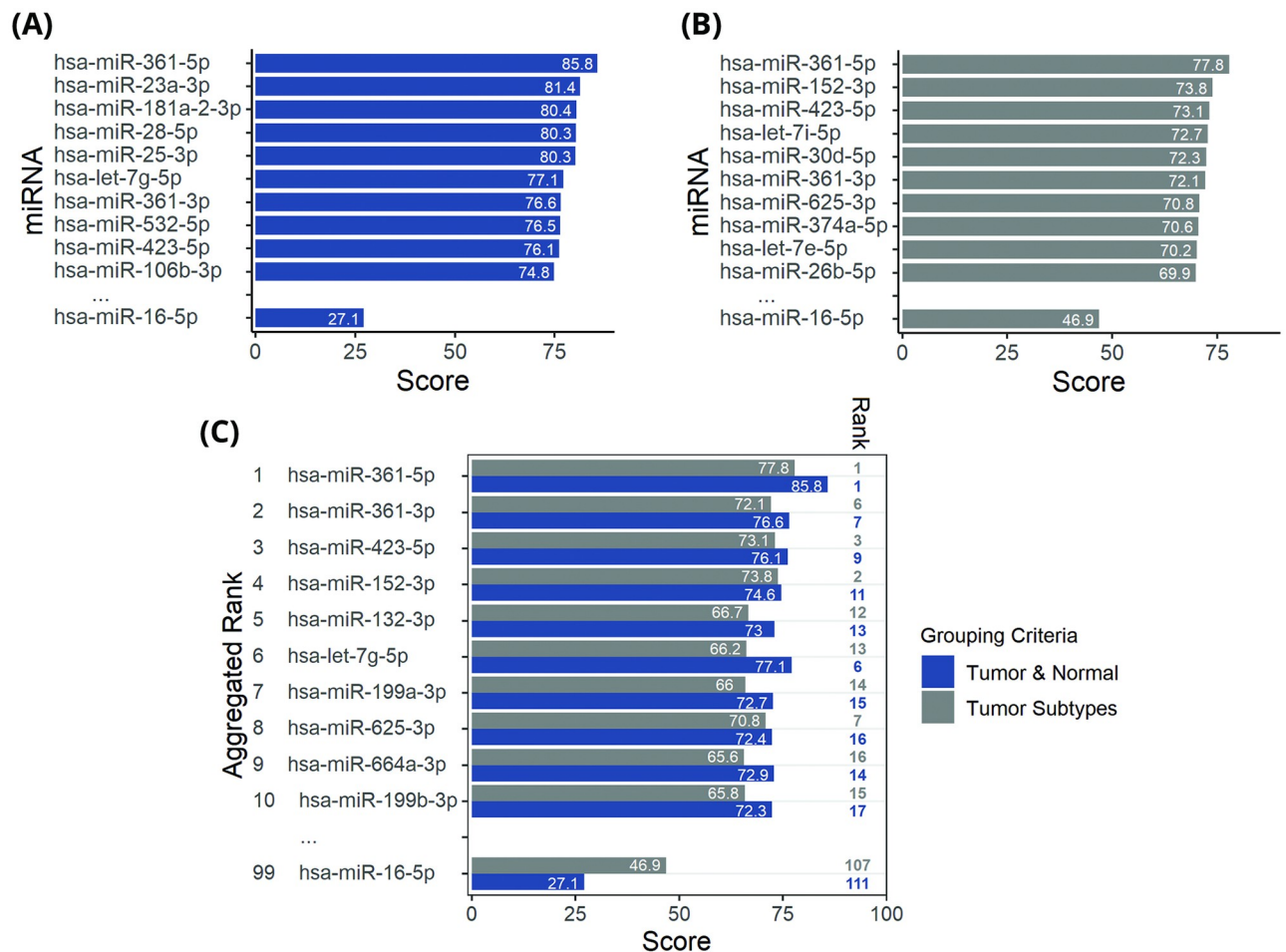
Total RNA was isolated from all samples (approximately 100 mg) using the RiboEx Total RNA reagent (GeneAll Biotechnology, South Korea). The amount of extracted RNA was quantified by measuring the absorbance at 260 nm. The purity of the RNA was determined by calculating the ratio of the absorbance at 260 and 280 nm. The absence of degradation of the RNA was confirmed by electrophoresis of the RNA on a 1% agarose gel containing ethidium bromide.

### Polyadenylation and reverse transcription

For the S-Poly(T) method, extracted total RNA was polyadenylated with Poly(A) Polymerase Tailing Kit (New England Biolabs., UK., Ltd.). Briefly, a 10 µl reaction including 1 µg total RNA, 1 µl of 10 × reaction buffer, 1 µl of 10 mM ATP and 1 unit of Poly(A) polymerase was incubated at 37°C for 30 min, followed by enzyme inactivation at 65°C for 5 min. After polyadenylation, reverse transcription was performed in a 20 µl reaction containing 10 µl of the polyadenylation reaction product, 2 µl of Anchored Oligo(dT), 1 µl of RiboLock RNase Inhibitor (20 U/µL), 4 µl of 5X Reaction Buffer, 2 µl 10 mM dNTP Mix, and 1 µl of RevertAid M-MuLV RT (200 U/µL) (Thermo Fisher Scientific., UK). The reaction was incubated at 42°C for 70 min and then terminated by heating at 70°C for 5 min.







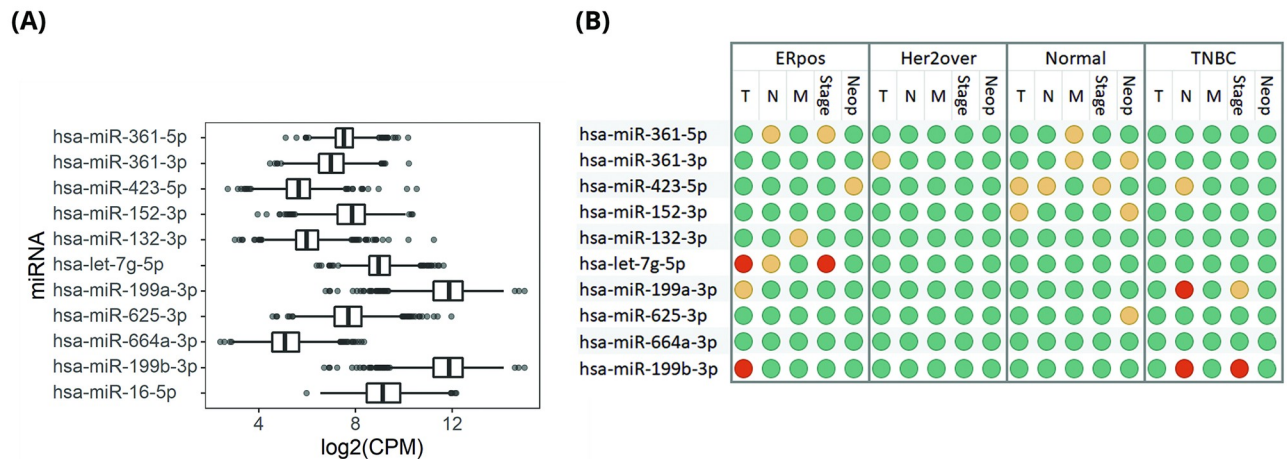
**Fig 1. Multi-component scoring results for different grouping criteria.** (A) for Tumor & Normal grouping criterion and (B) Tumor Subtypes. (C) is the aggregated Rank results using scorings in (A) and (B). The Rank column in (C) represents miRNAs rank in that grouping criterion. All scores are out of 100.

<https://doi.org/10.1371/journal.pone.0253009.g001>

million, equivalent to average read count of 20) in more than 5 percent of samples. Using this filter, 185 miRNAs remained for further analysis. In order to obtain the most promising and stable internal miRNA controls we ran our multi-component scoring system on two different grouping criteria. Firstly, based on Tumor & Normal samples and secondly based on Tumor Subtypes. Tumor Subtypes were defined as: Estrogen receptor status positive ( $n = 795$ ), Her2 receptor over expressed ( $n = 39$ ) and Triple Negative Breast Cancer ( $n = 125$ ; Fig 1). The hsa-miR-361-5p had the highest stability score in both grouping criteria, reaching scores of 85.7 and 76.1 out of 100. By using Robust Ranking Aggregation (RRA) [36] we aggregated rankings of the two grouping criteria. According to the aggregated ranks, hsa-miR-361-3p, hsa-miR-423-5p and hsa-miR-152-3p were the best ICs after hsa-miR-361-5p. Expression level of top 10 miRNAs along with hsa-miR-16-5p are represented in Fig 2A. Among them, hsa-miR-199a-3p and hsa-miR-199b-3p were the most abundant miRNAs.

### Correlation with clinical and pathological characteristics

In each biological subgroup, Spearman's correlation coefficient was calculated between the most stable miRNA expression and the following 5 features: TNM (Tumor, Node, and



**Fig 2. Expression level and correlation with clinical aspects in top 10 most stable miRNAs.** (A) Expression level of top 10 most stable miRNAs + hsa-miR-16-5p for comparison. CPM, Count Per Million. (B) most stable miRNAs correlation with clinical aspects in various biological subgroups. Colors represent statistical significance: red,  $p < 0.01$ ; yellow,  $0.01 < p < 0.1$ ; green,  $p > 0.1$ . ERpos, Estrogen positive; Her2over, Her2 Overexpressed; Normal, adjacent Normal Tissue; TNBC, Triple Negative Breast Cancer; T, N and M are TNM classification indexes; Stage, pathological Stage; Neop, Neoplasm Status.

<https://doi.org/10.1371/journal.pone.0253009.g002>

Metastasis) classification indexes, pathological stage and follow-up person neoplasm cancer status. The most significant correlations were as follows: hsa-let-7g-5p with Tumor ( $p \leq 0.001$ ) and pathological stage ( $p \leq 0.001$ ) and hsa-miR-199a-3p with Node ( $p \leq 0.005$ ; Fig 2B).

### Differential expression analysis

A differential expression analysis for Tumor vs Normal samples was performed on TCGA BRCA miRNA-Seq data considering both paired samples and all samples. As presented in Table 4, although there are some significant differentiations in the most stable miRNAs, their abs (logFC) are smaller than 0.32.

### Survival analysis

Kaplan-Meier survival analysis was carried out on METABRIC breast cancer dataset (n = 1262) for each of 10 most stable miRNAs. As it is shown in Table 5, three of them had

**Table 4. Differential expression analysis for Tumor vs Normal samples on TCGA breast cancer miRNA-Seq data.**

	logFC	adj.P.Val	logFC_paired	adj.P.Val_paired	Mean Exp.
hsa-miR-361-5p	-0/13	0/05	-0/31	<b>1/38E-05</b>	726
hsa-miR-361-3p	-0/15	0/11	-0/11	0/31	535
hsa-miR-423-5p	-0/11	0/20	-0/26	<b>0/01</b>	193
hsa-miR-152-3p	-0/31	<b>0/0005</b>	-0/21	<b>0/02</b>	905
hsa-miR-132-3p	-0/18	<b>0/03</b>	-0/32	<b>0/006</b>	284
hsa-let-7g-5p	0/11	0/07	0/06	0/5	1849
hsa-miR-199a-3p	-0/16	0/11	0/19	<b>0/04</b>	14361
hsa-miR-625-3p	0/04	0/73	0/08	0/57	889
hsa-miR-664a-3p	-0/19	0/06	-0/09	0/44	140
hsa-miR-199b-3p	-0/15	0/11	0/19	0/03	14315

significant p-values are in bold style. Mean Exp., mean read count expression; logFC, log Fold Change.

<https://doi.org/10.1371/journal.pone.0253009.t004>



**Table 5. Association with overall survival (OS) on METABRIC dataset.**

	<b>log-rank p</b>	<b>HR</b>
hsa-miR-361-5p	0/84	0/98
hsa-miR-361-3p	0/86	1/02
hsa-miR-423-5p	0/6	0/95
hsa-miR-152	0/65	0/96
hsa-miR-132	0/2	0/88
hsa-let-7g-5p	<b>0/02</b>	0/79
hsa-miR-199a-3p	<b>0/004</b>	0/75
hsa-miR-625	<b>0/000059</b>	0/67

HR, Hazard Ratio.

<https://doi.org/10.1371/journal.pone.0253009.t005>

significant association with overall survival: hsa-miR-625-3p ( $p \leq 0.00005$ ), hsa-miR-199a-3p ( $p \leq 0.004$ ) and hsa-let-7g-5p ( $p \leq 0.02$ ). Data for hsa-miR-664a-3p and hsa-miR-199b-3p were not available in the METABRIC dataset. Detailed Kaplan-Meier plots are shown in Fig 3.

### miR-361-5p comparison with miR-16-5p on TCGA data

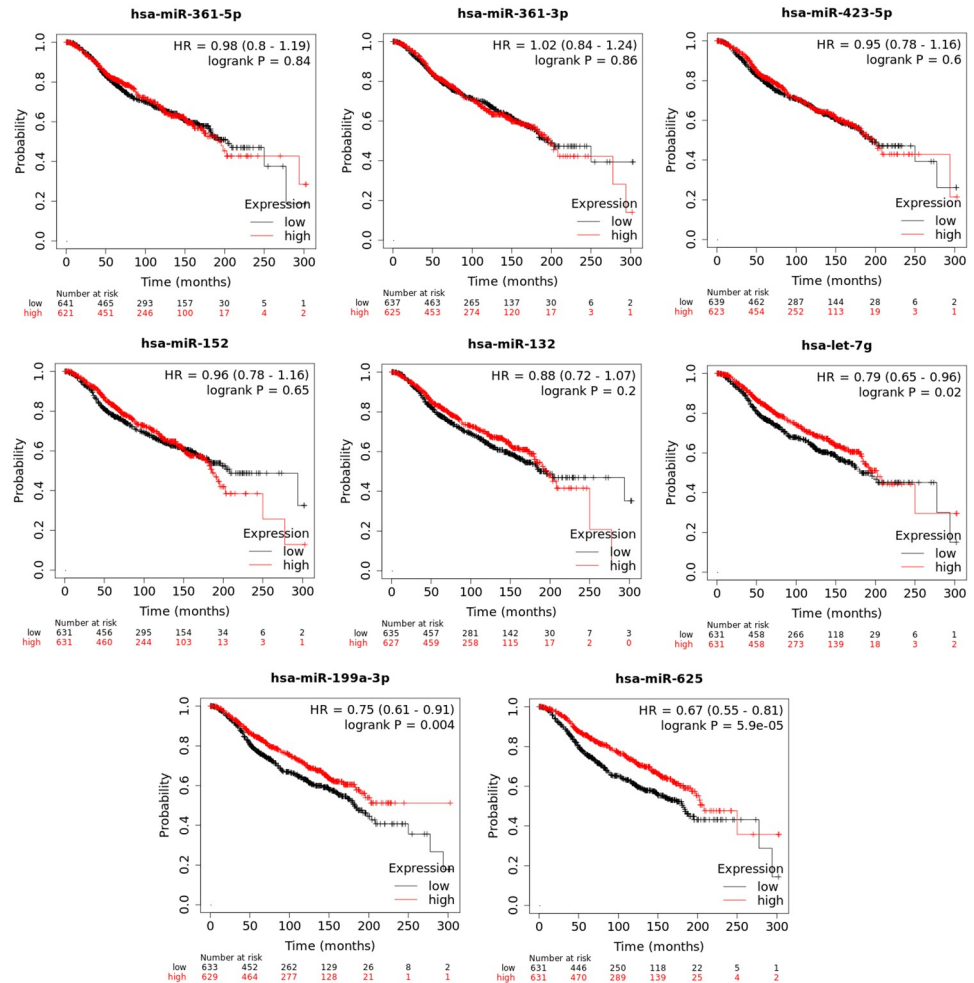
As a promising IC for breast cancer qPCR experiments, hsa-miR-16-5p is a well-known candidate. However, according to our ranking system it was ranked 99th among 185 short-listed miRNAs. Its detailed scores are exhibited in Fig 1. Fig 4A depicts its expression level (in CPM) beside hsa-miR-361-5p, showing hsa-miR-361-5p's lower standard deviation. Fig 4B illustrates that hsa-miR-361-5p had a higher score in most of our scoring components. As a result of differential expression analysis on tumor vs normal samples, hsa-miR-16-5p was significantly differentiated (adj.p.value:  $1.6e-13$  lfc: 0.62) while hsa-miR-361-5p showed minor differentiation (adj.p.value: 0.05 lfc: -0.13). As displayed in Fig 4C, hsa-miR-361 had lower abs (LFC) compared to hsa-miR-16-5p in paired tumor and normal samples. Overall, based on our analysis of TCGA miRNA-Seq data, hsa-miR-361-5p is more stable and reliable IC than hsa-miR-16-5p.

### Experimental validation using qRT-PCR

A profile of 34 paired samples for breast cancer tumor and adjacent normal tissues were assessed by qRT-PCR to validate hsa-miR-361-5p as a promising IC. We examined the expression of hsa-miR-361-5p, hsa-miR-16-5p, as well as U48 control gene. hsa-miR-16-5p was not detected in two samples, so we excluded them from our analysis. U48 and miR-361-5p had a higher expression level in comparison with miR-16-5p (Fig 5A). Among them, U48 was the most abundant control with a median Ct of 26.64. Standard deviation of raw Ct values was used as a stability measure. Fig 5B shows that hsa-miR-361-5p had the lowest standard deviation.

### Effect on relative expression of hsa-miR-21-5p

hsa-miR-21-5p is one of the most up-regulated miRNAs in breast cancer [37]. In order to evaluate our candidate IC, we measured hsa-miR-21-5p's expression level on 12 pairs of tissues randomly, using hsa-miR-361-5p, hsa-miR-16-5p as well as U48. As presented in Fig 5C, hsa-miR-361-5p was the only one which could detect hsa-miR-21-5p's up-regulation, while there was no significant change in hsa-miR-21-5p when normalization was done with U48 or hsa-miR-16-5p.



**Fig 3. Overall survival Kaplan-Meier plots of top 10 most stable miRNAs.** The dataset is METABRIC (n = 1262). Low- and high-risk groups were split based on median expression. Data for hsa-miR-664a-3p and hsa-miR-199b-3p were not available in the METABRIC dataset.

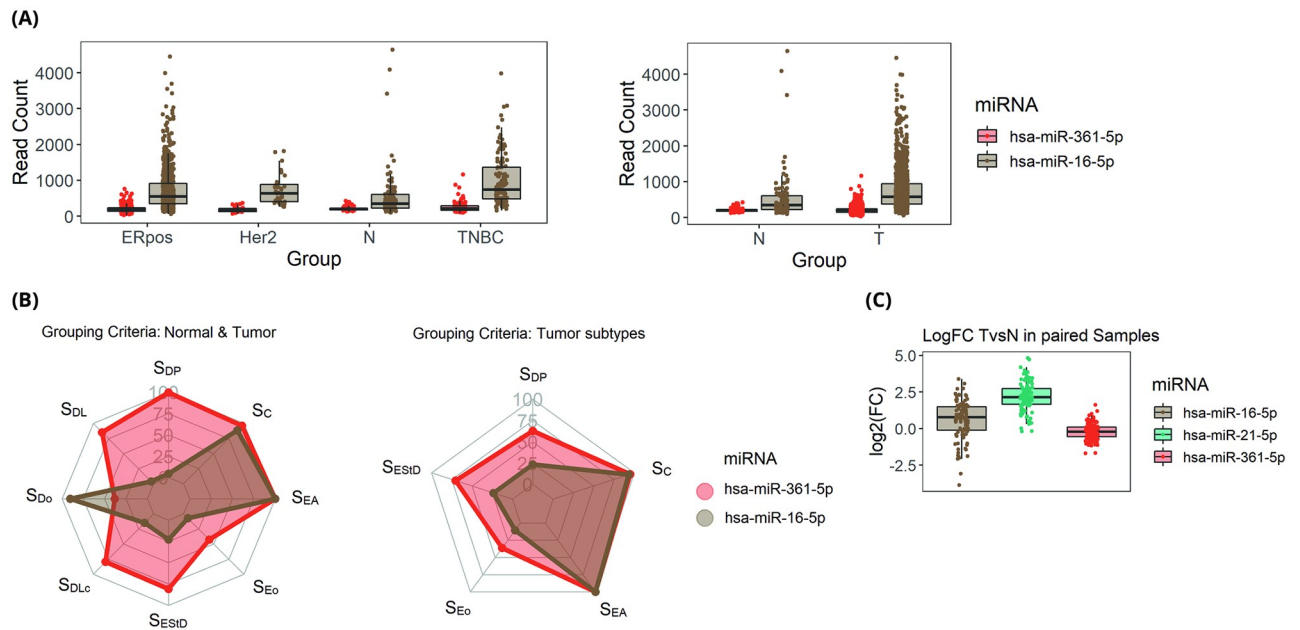
<https://doi.org/10.1371/journal.pone.0253009.g003>

### Discussion

Since their discovery, miRNAs emerged as important molecules in cancer initiation, progression and pathogenesis [8]. Similar to mRNA expression analysis, qPCR for miRNA quantification requires proper normalization strategies to compensate any non-biological variations [38]. Internal reference genes are currently used as the most universal and accurate method of normalization in qRT-PCR studies [5]. Reference genes might ideally have constant and high expression levels under various circumstances in almost all tissue types. However, such a reference gene does not exist [6]. Luckily, this is not a serious issue since most experimental designs are restricted to few tissue or different histological types of the same tissue.

Different RNA species, including rRNAs, tRNAs, snRNAs, and miRNAs have previously been used as ICs in miRNA real-time q-PCR studies of breast cancer. However, researchers are concerned about their use in normalization, mostly due to their very high expression levels as well as some biases in their stability [14, 39, 40].

Several miRNA expression analysis studies on different tissues have used miRNAs like miR-16 to normalize the expression of interested miRNAs. However, there are controversies



**Fig 4. Comparison between hsa-miR-361-5p and hsa-miR-16-5p.** (A) Expression level of hsa-miR-361-5p and hsa-miR-16-5p in two different grouping criteria based on TCGA data. T stands for Tumor; N, Adjacent Normal; ERpos, Estrogen Receptor Positive; Her2, Her2 Receptor overexpressed; TNBC, Triple Negative Breast Cancer. (B) Radar plot of scoring components of the multi-component scoring system. To make the plot more compact, following components are averaged. SEo: SEoL, SEoH; SDo: SDoO SDoU; SC: SCr, SCp. (C) Log2 Fold Change of Tumor Vs Normal paired samples based on TCGA data for candidate reference miRNAs along with hsa-miR-21-5p as a well-known oncomiR.

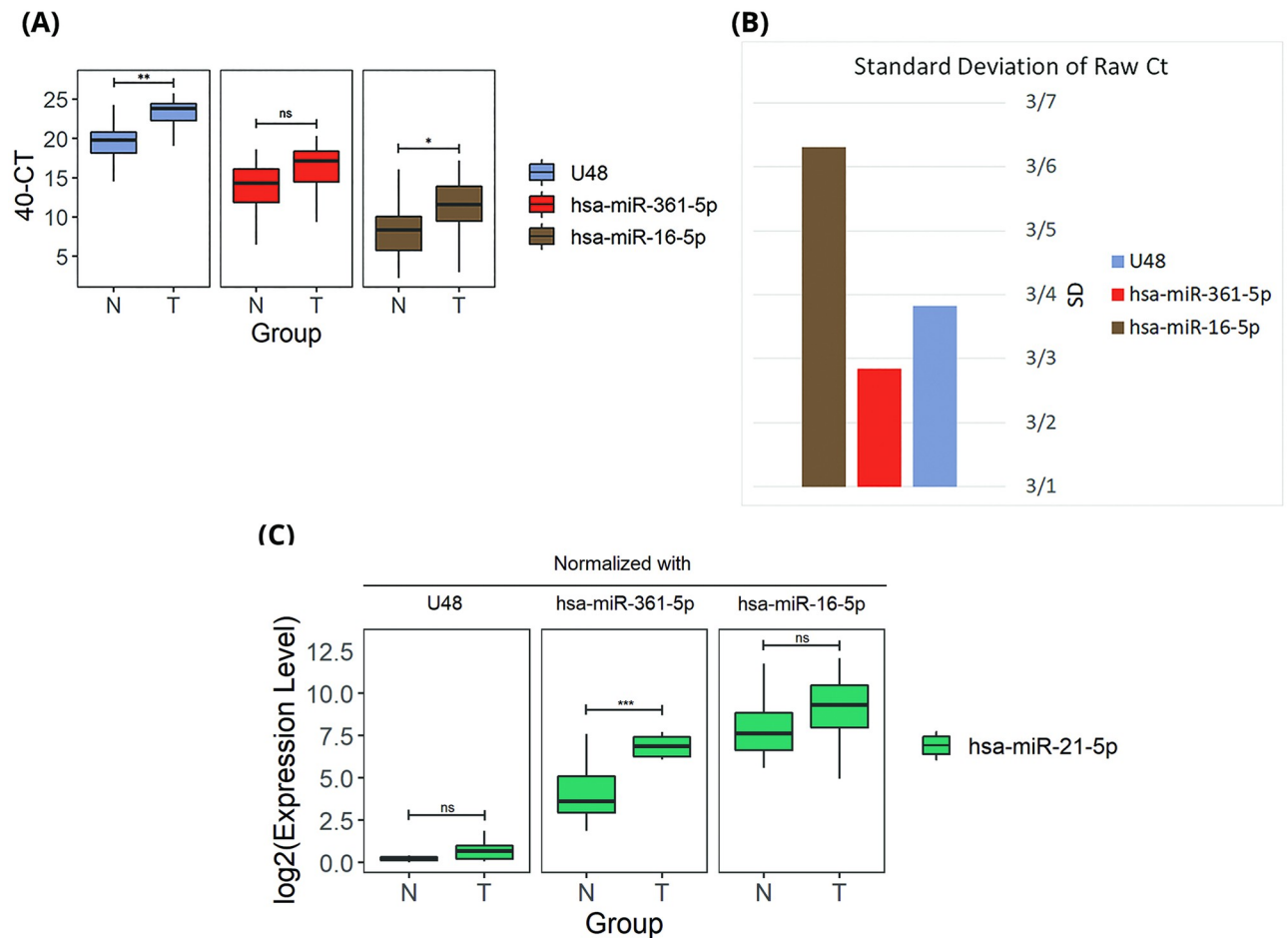
<https://doi.org/10.1371/journal.pone.0253009.g004>

about the suitability of these miRNAs as a good normalizer. Mattie et al. used miR-16 and let-7 to normalize miRNA expression on breast tissue sample. Later, Davoren et al. demonstrated their stable expression in malignant, benign, and normal breast tissues [11, 41, 42]. Also, Early studies on miRNAs expressions in breast cancer utilized miR-16 as a normalizer [43, 44]. Interestingly, there are evidence that miR-15a/16-1 is involved in the regulation of cell proliferation, apoptosis and invasion [45]. In two other studies, it is reported that miR-16 was significantly downregulated in malignant prostate and breast tissues [17, 46].

To our knowledge there has been only one study that had been miRNA-Seq data to detect promising ICs for breast cancer qRT-PCR studies, in which let-7i-5p and miR-361-5p have been suggested as a suitable one. However, they have used only three criteria to select the best reference miRNAs: mean expression, coefficient of variation and expression fold change between normal & tumor samples [47]. The approach doesn't take into account other features like clinical or pathologic aspects and also miRNA differentiation between subtypes of tumor samples which are commonly investigated in many cancer studies.

Here, we have discovered the most stable and reliable miRNA ICs for breast cancer qRT-PCR studies, based on TCGA miRNA-Seq data using a multi-component scoring system. This system not only considers the expression stability of miRNAs in different biological groups like tumor subtypes but also checks out for correlation with clinical features of samples.

We investigated TCGA miRNA-Seq breast cancer data for most stable miRNAs based on two sample grouping criteria. First Tumor & Normal samples, second Tumor Subtypes. This approach not only considered tumor-normal differentiation but also within tumor samples variation due to subtypes of breast cancer. This is important because lots of breast cancer



**Fig 5. qRT-PCR validation.** (A) qRT-PCR raw 40-Ct values of candidate internal controls. The qRT-PCR was ran for 40 cycles so 40-Ct is a measure of expression level. (B) Standard Deviation of candidate internal controls based on raw Ct values of qRT-PCR (C) Expression level of hsa-miR-21-5p normalized with candidate internal controls based on qRT-PCR data. ns: not significant differentiation in Tumor vs Normal samples; \*\*\*: p-value < 0.001.

<https://doi.org/10.1371/journal.pone.0253009.g005>

experiments are focused on tumor subtypes. After applying our scoring system, top 10 most stable miRNAs were selected for further investigation.

hsa-miR-361-5p had the highest stability score among well-expressed miRNAs, followed by hsa-miR-361-3p and hsa-miR-423-5p. hsa-miR-361-5p ranked first in both grouping criteria showing its convenience for both paired samples and tumor subtype studies.

There are few reports on miR-361-5p role in breast cancer. Zhan, et al., used miRNA-Seq data to detect stably expressed miRNAs in 14 human tumor types. hsa-miR-361-5p was reported as a candidate reference miRNA in eight of 14 cancer types including breast cancer [47]. On the other hand, there are some reports on down-regulation of miR-361-5p in breast cancer [48] and also TNBC subtype [49, 50]. However, both researches have used U6 small nuclear RNA as normalizer for their qRT-PCR validation, which is argued to be up-regulated in breast tumor tissue itself and could potentially lead to data-misinterpretation in breast cancer qRT-PCR experiments [15]. It has also not escaped our notice that, as exhibited in Table 4, hsa-miR-361-5p had some down-regulation in TCGA paired samples, however, its fold change was low (1.2 for normal vs tumor). Moreover, while in our analysis hsa-miR-361-5p had no

significant association with overall survival, in one study hsa-miR-361-5p has been reported as a prognostic biomarker for disease free survival, specifically in TNBC subtype [49].

To compare hsa-miR-361-5p usefulness with a commonly used IC, we compared hsa-miR-16-5p expression in the same samples. hsa-miR-361-5p had a better performance in all scoring components except for  $S_{Do}$  which is related to the expression of outliers in paired samples. We set the lower and upper 10% of all samples as outliers.

To validate the suitability of hsa-miR-361-5p as an IC, we performed a qRT-PCR experiment on 34 pairs of tumor and adjacent normal tissue samples. We then quantified hsa-miR-361-5p along with two commonly used ICs, hsa-miR-16-5p and U48. hsa-miR-361-5p had lower standard deviation compared to the others and thus turned out to be a better IC for normal and tumor tissues of breast.

One of the main methods to evaluate the suitability of an IC is to test its performance as a normalizer for quantification the expression of a well-examined gene or miRNA. hsa-miR-21-5p is an oncomiR which is up-regulated in almost all tumor samples [37]. This is further confirmed based on our analysis on TCGA dataset, as shown in Fig 4B. We examined the performance of hsa-miR-361-5p, hsa-miR-16-5p and U48 as an IC and to assess how they can affect the expression level of hsa-miR-21-5p in tumor samples. Surprisingly, hsa-miR-361-5p was the only IC capable of showing the significant up-regulation of hsa-miR-21-5p in our real-time experiment. Moreover, while U48 was a stable reference gene in tumor and normal samples, it could not efficiently detect the significant up-regulation of hsa-miR-21 in tumor samples.

There have been two limitations to our study that should be considered. First, in the qRT-PCR validation phase, most of our tumor samples subtypes were luminal A (64%) and second, the number of samples for evaluating hsa-miR-21-5p expression was low (12 pairs).

In Conclusion, we first introduced the top most stable miRNAs in breast cancer and normal tissues, using a multi-component scoring system on TCGA miRNA-Seq data. This system takes into account the expression stability along with clinical and pathological characteristics of samples. Secondly, we validated that hsa-miR-361-5p is a promising IC for breast cancer qRT-PCR studies and compared it with two commonly used ICs to show its superiority.

## Supporting information

**S1 Table. Clinical and pathological data on malignant tumor samples where available.** T, N and M refer to the primary tumor size, nodal status and distant metastases status according to the TNM breast cancer classification system. ER: = estrogen receptor status; PR: = progesterone receptor status and HER2 = v-erb-b2 erythroblastic leukemia viral oncogene status. Samples marked with star are those checked for miR-21-5p expression.  
(DOCX)

**S1 Fig. General workflow.**  
(TIF)

## Acknowledgments

The authors would like to thank Dr. Fatemeh Yousefi and Dr. Akram Gholipour for their experimental supports. We also thank Mohammadjavad Karimi Taheri for his comments and suggestions that improved the project.

## Author Contributions

**Data curation:** Sogol Ghanbari, Adel Salimi.

**Formal analysis:** Sogol Ghanbari, Adel Salimi.

**Investigation:** Sogol Ghanbari.

**Methodology:** Sogol Ghanbari, Adel Salimi, Saeid Rahmani.

**Project administration:** Ali Sharifi-Zarchi, Seyed Javad Mowla.

**Resources:** Nahid Nafissi.

**Software:** Adel Salimi, Saeid Rahmani.

**Supervision:** Ali Sharifi-Zarchi, Seyed Javad Mowla.

**Validation:** Sogol Ghanbari.

**Visualization:** Sogol Ghanbari, Adel Salimi.

**Writing – original draft:** Sogol Ghanbari, Adel Salimi.

**Writing – review & editing:** Sogol Ghanbari, Adel Salimi, Saeid Rahmani, Nahid Nafissi, Ali Sharifi-Zarchi, Seyed Javad Mowla.

## References

1. Kontomanolis EN, Koutras A, Syllaios A, Schizas D, Mastoraki A, Garmpis N, et al. Role of Oncogenes and Tumor-suppressor Genes in Carcinogenesis: A Review. *Anticancer Research*. 2020; 40(11):6009–15. <https://doi.org/10.21873/anticancer.14622> PMID: 33109539
2. Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular detection and quantification*. 2015; 3:9–16. <https://doi.org/10.1016/j.bdq.2015.01.005> PMID: 27077029
3. Chen Y, Gelfond JA, McManus LM, Shireman PK. Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC genomics*. 2009; 10(1):1–10. <https://doi.org/10.1186/1471-2164-10-407> PMID: 19715577
4. Barra GB, Santa Rita TH, Mesquita PG, Jácomo RH, Nery LFA. Analytical Sensitivity and Specificity of Two RT-qPCR Protocols for SARS-CoV-2 Detection Performed in an Automated Workflow. *Genes*. 2020; 11(10):1183.
5. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*. 2002; 3(7):research0034. <https://doi.org/10.1186/gb-2002-3-7-research0034> PMID: 12184808
6. Zhang Y, Li D, Sun B. Do housekeeping genes exist? *PLoS One*. 2015; 10(5):e0123691. <https://doi.org/10.1371/journal.pone.0123691> PMID: 25970694
7. Taylor SC, Nadeau K, Abbasi M, Lachance C, Nguyen M, Fenrich J. The ultimate qPCR experiment: producing publication quality, reproducible data the first time. *Trends in Biotechnology*. 2019; 37(7):761–74. <https://doi.org/10.1016/j.tibtech.2018.12.002> PMID: 30654913
8. Yang Z, Liu Z. The emerging role of microRNAs in breast cancer. *Journal of Oncology*. 2020; 2020.
9. Forterre A, Komuro H, Aminova S, Harada M. A comprehensive review of cancer MicroRNA therapeutic delivery strategies. *Cancers*. 2020; 12(7):1852. <https://doi.org/10.3390/cancers12071852> PMID: 32660045
10. Benes V, Castoldi M. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods*. 2010; 50(4):244–9. <https://doi.org/10.1016/j.ymeth.2010.01.026> PMID: 20109550
11. Davoren PA, McNeill RE, Lowery AJ, Kerin MJ, Miller N. Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer. *BMC molecular biology*. 2008; 9(1):76. <https://doi.org/10.1186/1471-2199-9-76> PMID: 18718003
12. Chugh P, Dittmer DP. Potential pitfalls in microRNA profiling. *Wiley Interdisciplinary Reviews: RNA*. 2012; 3(5):601–16. <https://doi.org/10.1002/wrna.1120> PMID: 22566380
13. Tricarico C, Pinzani P, Bianchi S, Paglierani M, Distante V, Pazzagli M, et al. Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. *Analytical biochemistry*. 2002; 309(2):293–300. [https://doi.org/10.1016/S0003-2697\(02\)00311-1](https://doi.org/10.1016/S0003-2697(02)00311-1) PMID: 12413463



14. Gee H, Buffa F, Camps C, Ramachandran A, Leek R, Taylor M, et al. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *British journal of cancer*. 2011; 104(7):1168–77. <https://doi.org/10.1038/sj.bjc.6606076> PMID: 21407217
15. Lou G, Ma N, Xu Y, Jiang L, Yang J, Wang C, et al. Differential distribution of U6 (RNU6-1) expression in human carcinoma tissues demonstrates the requirement for caution in the internal control gene selection for microRNA quantification. *International journal of molecular medicine*. 2015; 36(5):1400–8. <https://doi.org/10.3892/ijmm.2015.2338> PMID: 26352225
16. Bianchi N, Zuccato C, Finotti A, Lampronti I, Borgatti M, Gambari R. Involvement of miRNA in erythroid differentiation. *Epigenomics*. 2012; 4(1):51–65. <https://doi.org/10.2217/epi.11.104> PMID: 22332658
17. Cascione L, Gasparini P, Lovat F, Carasi S, Pulvirenti A, Ferro A, et al. Integrated microRNA and mRNA signatures associated with survival in triple negative breast cancer. *PloS one*. 2013; 8(2): e55910. <https://doi.org/10.1371/journal.pone.0055910> PMID: 23405235
18. Cai B, Ma M, Chen B, Li Z, Abdalla BA, Nie Q, et al. MiR-16-5p targets SESN1 to regulate the p53 signaling pathway, affecting myoblast proliferation and apoptosis, and is involved in myoblast differentiation. *Cell death & disease*. 2018; 9(3):1–15.
19. Andersen CL, Jensen JL, Ørntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research*. 2004; 64(15):5245–50. <https://doi.org/10.1158/0008-5472.CAN-04-0496> PMID: 15289330
20. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. *Biotechnology letters*. 2004; 26(6):509–15. <https://doi.org/10.1023/b:bile.000019559.84305.47> PMID: 15127793
21. Chapman JR, Waldenström J. With reference to reference genes: a systematic review of endogenous controls in gene expression studies. *PloS one*. 2015; 10(11):e0141853. <https://doi.org/10.1371/journal.pone.0141853> PMID: 26555275
22. Drobna M, Szarzyńska-Zawadzka B, Dąca-Roszak P, Kosmalka M, Jaksik R, Witt M, et al. Identification of Endogenous Control miRNAs for RT-qPCR in T-Cell Acute Lymphoblastic Leukemia. *International journal of molecular sciences*. 2018; 19(10):2858. <https://doi.org/10.3390/ijms19102858> PMID: 30241379
23. Tilli TM, da Silva Castro C, Tuszynski JA, Carels N. A strategy to identify housekeeping genes suitable for analysis in breast cancer diseases. *BMC genomics*. 2016; 17(1):639. <https://doi.org/10.1186/s12864-016-2946-1> PMID: 27526934
24. Dai Y, Cao Y, Köhler J, Lu A, Xu S, Wang H. Unbiased RNA-Seq-driven identification and validation of reference genes for quantitative RT-PCR analyses of pooled cancer exosomes. *BMC genomics*. 2021; 22(1):1–13. <https://doi.org/10.1186/s12864-020-07350-y> PMID: 33388042
25. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nature genetics*. 2013; 45(10):1113–20. <https://doi.org/10.1038/ng.2764> PMID: 24071849
26. 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 research*. 2016; 44(8):e71–e. <https://doi.org/10.1093/nar/gkv1507> PMID: 26704973
27. Xu T, Su N, Liu L, Zhang J, Wang H, Zhang W, et al. miRBaseConverter: an R/Bioconductor package for converting and retrieving miRNA name, accession, sequence and family information in different versions of miRBase. *BMC bioinformatics*. 2018; 19(19):179–88.
28. Krasnov GS, Kudryavtseva AV, Snezhkina AV, Lakunina VA, Beniaminov AD, Melnikova NV, et al. Pan-cancer analysis of TCGA data revealed promising reference genes for qPCR normalization. *Frontiers in Genetics*. 2019; 10:97. <https://doi.org/10.3389/fgene.2019.00097> PMID: 30881377
29. R Core Team. R: A Language and Environment for Statistical Computing. 2020. <https://www.R-project.org>.
30. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology*. 2014; 15(2):R29. <https://doi.org/10.1186/gb-2014-15-2-r29> PMID: 24485249
31. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015; 43(7):e47–e. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792
32. Lániczky A, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, et al. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast cancer research and treatment*. 2016; 160(3):439–46. <https://doi.org/10.1007/s10549-016-4013-7> PMID: 27744485

33. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012; 486(7403):346–52. <https://doi.org/10.1038/nature10983> PMID: 22522925
34. Gómez-Rubio V. ggplot2-elegant graphics for data analysis. *Journal of Statistical Software*. 2017; 77(1):1–3.
35. Minato Nakazawa. Functions for Medical Statistics Book with some Demographic Data 2019-12-15. <https://cran.r-project.org/web/packages/fmsb/index.html>.
36. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*. 2012; 28(4):573–80. <https://doi.org/10.1093/bioinformatics/btr709> PMID: 22247279
37. Wang H, Tan Z, Hu H, Liu H, Wu T, Zheng C, et al. microRNA-21 promotes breast cancer proliferation and metastasis by targeting LZTFL1. *BMC cancer*. 2019; 19(1):738. <https://doi.org/10.1186/s12885-019-5951-3> PMID: 31351450
38. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Oxford University Press; 2009.
39. Gharbi S, Shamsara M, Khateri S, Soroush MR, Ghorbanmehr N, Tavallaei M, et al. Identification of reliable reference genes for quantification of microRNAs in serum samples of sulfur mustard-exposed veterans. *Cell Journal (Yakhteh)*. 2015; 17(3):494. <https://doi.org/10.22074/cellj.2015.9> PMID: 26464821
40. Lim Q, Zhou L, Ho Y, Wan G, Too H. snoU6 and 5S RNAs are not reliable miRNA reference genes in neuronal differentiation. *Neuroscience*. 2011; 199:32–43. <https://doi.org/10.1016/j.neuroscience.2011.10.024> PMID: 22051575
41. Mattie MD, Benz CC, Bowers J, Sensinger K, Wong L, Scott GK, et al. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. *Molecular cancer*. 2006; 5(1):24. <https://doi.org/10.1186/1476-4598-5-24> PMID: 16784538
42. McDermott AM, Kerin MJ, Miller N. Identification and validation of miRNAs as endogenous controls for RQ-PCR in blood specimens for breast cancer studies. *PloS one*. 2013; 8(12). <https://doi.org/10.1371/journal.pone.0083718> PMID: 24391813
43. Zhao H, Shen J, Medico L, Wang D, Ambrosone CB, Liu S. A pilot study of circulating miRNAs as potential biomarkers of early stage breast cancer. *PloS one*. 2010; 5(10):e13735. <https://doi.org/10.1371/journal.pone.0013735> PMID: 21060830
44. Heneghan HM, Miller N, Lowery AJ, Sweeney KJ, Newell J, Kerin MJ. Circulating microRNAs as novel minimally invasive biomarkers for breast cancer. *Annals of surgery*. 2010; 251(3):499–505. <https://doi.org/10.1097/SLA.0b013e3181cc939f> PMID: 20134314
45. Liu T, Xu Z, Ou D, Liu J, Zhang J. The miR-15a/16 gene cluster in human cancer: A systematic review. *Journal of cellular physiology*. 2019; 234(5):5496–506. <https://doi.org/10.1002/jcp.27342> PMID: 30246332
46. Schaefer A, Jung M, Miller K, Lein M, Kristiansen G, Erbersdobler A, et al. Suitable reference genes for relative quantification of miRNA expression in prostate cancer. *Experimental & molecular medicine*. 2010; 42(11):749–58. <https://doi.org/10.3858/emm.2010.42.11.076> PMID: 20890088
47. Zhan C, Yan L, Wang L, Jiang W, Zhang Y, Xi J, et al. Identification of reference miRNAs in human tumors by TCGA miRNA-seq data. *Biochemical and Biophysical Research Communications*. 2014; 453(3):375–8. <https://doi.org/10.1016/j.bbrc.2014.09.086> PMID: 25264191
48. Liu J, Yang J, Yu L, Rao C, Wang Q, Sun C, et al. miR-361-5p inhibits glioma migration and invasion by targeting SND1. *OncoTargets and therapy*. 2018; 11:5239. <https://doi.org/10.2147/OTT.S171539> PMID: 30214229
49. Cao Z-G, Huang Y-N, Yao L, Liu Y-R, Hu X, Hou Y-F, et al. Positive expression of miR-361-5p indicates better prognosis for breast cancer patients. *Journal of thoracic disease*. 2016; 8(7):1772. <https://doi.org/10.21037/jtd.2016.06.29> PMID: 27499968
50. Han J, Yu J, Yu'na Dai JL, Guo M, Song J, Zhou X. Overexpression of miR-361-5p in triple-negative breast cancer (TNBC) inhibits migration and invasion by targeting RQCD1 and inhibiting the EGFR/PI3K/Akt pathway. *Bosnian journal of basic medical sciences*. 2019; 19(1):52. <https://doi.org/10.17305/bjbm.2018.3399> PMID: 29924958