



Identification of reference miRNAs in human tumors by TCGA miRNA-seq data



Cheng Zhan^a, Li Yan^b, Lin Wang^a, Wei Jiang^a, Yongxing Zhang^a, Junjie Xi^a, Li Chen^a, Yulin Jin^a, Yulei Qiao^{a,*}, Yu Shi^{a,*}, Qun Wang^a

^a Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai 200032, China

^b Department of Radiation Oncology, Eye and ENT Hospital, Fudan University, Shanghai 200031, China

ARTICLE INFO

Article history:

Received 18 September 2014

Available online 27 September 2014

Keywords:

Reference miRNA

miRNA-seq

TCGA

ABSTRACT

Although the accuracy of detecting the expression of miRNAs by quantitative real-time polymerase chain reaction (qRT-PCR) is highly dependent on reliable reference miRNAs, many commonly used reference miRNAs are not stably expressed and as such are not suitable for quantification and normalization of qRT-PCR data. To solve this problem, we analyzed the global expression profiles of thousands of samples in 14 types of common human tumors released by The Cancer Genome Atlas (TCGA), and identified the most stably and highly expressed miRNAs as candidate reference miRNAs in each type of tumor. We found that miR-361-5p and let-7i-5p were the most recommended candidate reference miRNAs in nine and eight types of tumors, respectively, followed by let-7a-5p, miR-28-5p and miR-99b-5p. Our results are of important value to those researchers focused on miRNA; however, these candidate reference miRNAs still need to be validated prior to their use in qRT-PCR studies.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

With nearly two decades of history, quantitative real-time polymerase chain reaction (qRT-PCR) is still the most frequently used method to accurately quantitate the expression of genes and miRNAs. It is fast, economical, easy to use and especially suitable for the measurement of a limited number of genes or miRNAs. Given the highly sensitive and specific nature of qRT-PCR, however, its accuracy is mostly dependent on reliable references for normalization control. Accordingly, reference genes and miRNAs, are frequently used to normalize experimental deviation in qRT-PCR datasets arising from differences in RNA content, quantity and quality, transcriptional activity, and operational deviation, among others. The expression of these references should ideally be stably and constitutively elevated, and should not vary in different samples, nor be affected by experimental treatments. However, since actually many commonly used references do not meet the aforementioned criteria and are not suitable as a normalization control, the proper validation of references is required prior to their use in qRT-PCR studies [1,2].

With the development of microarray and sequencing in recent years, a large number of high-throughput gene expression data

have accumulated. These data confirmed that many conventional reference genes such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β -actin (ACTB) varied considerably in some experimental conditions [2–7]. Meanwhile, many studies have used these data to identify and validate suitable reference genes for special biological conditions: for example, evolution, differentiation, development and treatment of cancer, and other diseases or for comparing different physiological stages of a single organ [2–11].

However, there are few studies on reference miRNAs due to the rarity of high-throughput miRNA expression data. The validity of reference miRNAs has been unproven until now, and many commonly used reference miRNAs have been found to be unstable, such as U6, RNU6B, and RNU48 in cancers [12–16]. Since the recent release of many miRNA-seq data of human tumor samples by The Cancer Genome Atlas (TCGA), we have tried to analyze these data and have attempted to identify the appropriate reference miRNAs in human tumors.

2. Material and methods

2.1. miRNA expression data collection and pre-processing

Level 3 miRNA-seq isoform quantification data released by TCGA before April 15, 2014 were obtained from the TCGA data por-

* Corresponding authors. Fax: +86 21 64041990 2915.

E-mail addresses: qiao.yulei@zs-hospital.sh.cn (Y. Qiao), shi.yu@zs-hospital.sh.cn (Y. Shi).

tal (<https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>). Only those cancer types with more than five normal control samples were involved in this analysis, including bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), prostate adenocarcinoma (PRAD), stomach adenocarcinoma (STAD), thyroid carcinoma (THCA) and uterine corpus endometrial carcinoma (UCEC), 14 cancer types, 589 normal samples, and 5727 tumor samples in all. The number of samples of each tumor type is detailed in Table 1. The expression data of mature miRNAs and star miRNAs (3' arms of pre-miRNA) from each pre-miRNA were sorted and calculated according to their MIMAT serial number based on miRbase V21.0 (<http://www.mirbase.org/>), whereas stem-loop, precursor, or unannotated transcript data were not included in our analysis [17]. Reads per million miRNA mapped (RPM) values were used to represent miRNA expression levels, and data were presented as mean \pm standard deviation (SD).

2.2. Selection of reference miRNA candidates

In this study, the definition of a reference miRNA is a miRNA that is highly and stably expressed across normal and tumor physiological states. Reference miRNA candidates for each cancer type were filtered using the following criteria, based on previous studies [2,4,6]:

- (1) Mean (N) > 100 and Mean (T) > 100.
- (2) CV (N) < 0.5 and CV (T) < 0.5.
- (3) Mean (N)/Mean (T) < 1.3 and Mean (T)/Mean (N) < 1.3.

Here, Mean (N) and Mean (T) denote the mean of the RPM value of the miRNA in normal samples and tumor samples, respectively. CV (N) and CV (T) denote the coefficient of variation (CV); that is, the SD divided by the mean RPM value of the miRNA in normal samples and tumor samples, respectively. The first criterion identified miRNAs that are highly expressed in tissues and can be easily detected by qRT-PCR. For each miRNA, expression RPM values were averaged for both normal and tumor states. A miRNA was retained if the average expression level exceeded the selected threshold value of 100. The second criterion used the CV to verify whether the miRNAs were stably expressed in normal and tumor states. The third criterion used expression fold change to filter out miRNAs that are differentially expressed across normal and tumor states.

Table 1
Sample numbers of each tumor type.

Tumor type	Normal	Tumor	Total
Bladder urothelial carcinoma	19	260	279
Breast invasive carcinoma	103	1055	1158
Colon adenocarcinoma	8	430	438
Esophageal carcinoma	9	72	81
Head and neck squamous cell carcinoma	44	513	557
Kidney chromophobe	25	66	91
Kidney renal clear cell carcinoma	71	531	602
Liver hepatocellular carcinoma	50	200	250
Lung adenocarcinoma	46	499	545
Lung squamous cell carcinoma	45	467	512
Prostate adenocarcinoma	50	330	380
Stomach adenocarcinoma	38	276	314
Thyroid carcinoma	59	507	566
Uterine corpus endometrial carcinoma	22	521	543
Total	589	5727	6316

3. Results

Table 2 shows the number of miRNAs that meet different criteria in each tumor type. For example, in bladder urothelial carcinoma, by applying the first criterion, the number of miRNAs highly expressed in normal and tumor samples was 116 and 144, respectively. By applying the second criterion, 23 and 9 miRNAs were stably expressed in normal and tumor states, respectively. By applying the third criterion, there were 137 miRNAs stably expressed across normal and tumor states. Finally, two miRNAs that met all three criteria were identified as candidate reference miRNAs for bladder urothelial carcinoma.

Table 3 displays a list of candidate reference miRNAs for each tumor type. The number of candidate reference miRNAs varied from 2 to 26 in different cancer types. Fifty-eight miRNAs were identified as candidate reference miRNAs for one or more cancer types. Table 4 shows the corresponding numbers of tumor types for which candidate reference miRNAs were identified. Among them, miR-361-5p and let-7i-5p were highly consistent in nine and eight of 14 human tumor types, respectively, followed by let-7a-5p, miR-28-5p, and miR-99b-5p. These five miRNAs covered all 14 tumor types. Fig. 1 depicts the expression profiles of these five miRNAs in different cancer types. In addition, our results did not reveal any miRNAs that were suitable as reference miRNAs for all cancer types. Additionally, those candidate reference miRNAs, such as miR-99b-5p, were stably expressed in some kinds of tumor; however, their expression varied widely in other types of tumors.

4. Discussion

For a long time, reference genes like GAPDH were used as normalization controls in miRNA detection of qRT-PCR [18]. However, because of the huge difference in length between mRNA and miRNA, reference genes are not ideal controls, and as such, are no longer used. 5s rRNA was another often-used control, but its expression level is far more than some miRNAs with low content, and its expression is affected by apoptosis and tumorigenesis; thus, 5s rRNA is rarely used as well [12]. Currently, the most commonly used controls are U6 and other snRNAs such as RNU6B, RNU48, and RNU64. As a control, U6 is widely used in miRNA research in a variety of organs and tissues. However, many studies have revealed that these snRNAs are not stable, and in certain circumstances, there was a nearly 62-fold difference of the U6 expression in samples [13–16,19,20]. Therefore, researchers should not haphazardly select snRNAs like U6 as a normalization control without rigorous validation.

It has become widely recognized that the normalization control in the detection of miRNAs should be the same type as the sample miRNAs. However, currently there is still no uniform standard about reference miRNAs. Schaefer et al. recommended miR-130b-3p as a reference miRNA in prostate cancer [14]; Lamba et al. recommended miR-23b-3p and miR-152-5p in hepatic tumors [13]; Ratert et al. recommended that a combination of miR-101-3p, miR-125a-5p, miR-148b-3p, and miR-151a-5p in urothelial carcinoma [16]. In addition, there are many other studies on reference miRNAs in solid tissues and serum [15,19–31]. However, most of these studies are mainly based on the measurement of a limited set of miRNAs by qRT-PCR, rather than a global investigation of miRNA expression profiles.

Here, we investigated high throughput data generated by miRNA-seq from TCGA, and we analyzed the expression profiles of thousands of miRNAs from thousands of samples in 14 types of common tumors. Finally, we identified candidate reference miRNAs in each type of cancer, which should prove to be valuable

Table 2

Summary of the number of miRNAs in 14 tumor types that met our inclusion criteria.

Tumor type	Normal samples		Tumor samples		Stable across normal and tumor States	Candidate reference miRNAs
	Highly expressed	Stable	Highly expressed	Stable		
Bladder urothelial carcinoma	116	23	144	9	137	2
Breast invasive carcinoma	133	85	133	7	203	2
Colon adenocarcinoma	119	222	131	25	71	2
Esophageal carcinoma	117	157	133	76	183	13
Head and neck squamous cell carcinoma	131	57	144	18	46	5
Kidney chromophobe	118	238	113	92	231	23
Kidney renal clear cell carcinoma	111	176	122	59	214	10
Liver hepatocellular carcinoma	135	247	140	16	239	6
Lung adenocarcinoma	124	68	138	13	177	4
Lung squamous cell carcinoma	130	211	147	21	152	8
Prostate adenocarcinoma	95	140	110	79	260	15
Stomach adenocarcinoma	122	58	138	16	198	6
Thyroid carcinoma	142	182	137	82	321	26
Uterine corpus endometrial carcinoma	118	65	146	10	119	4

Table 3

Candidate reference miRNAs identified in each tumor type.

Tumor type	Candidate reference miRNAs
Bladder urothelial carcinoma	miR-28-5p, miR-28-3p
Breast invasive carcinoma	let-7i-5p, miR-361-5p
Colon adenocarcinoma	let-7i-5p, let-7d-5p
Esophageal carcinoma	let-7a-5p, let-7b-5p, let-7e-5p, let-7f-5p, let-7g-5p, miR-22-3p, miR-27b-3p, miR-28-5p, miR-29a-3p, miR-186-5p, miR-361-5p, miR-361-3p, miR-378a-3p
Head and neck squamous cell carcinoma	let-7i-5p, miR-99b-5p, miR-361-5p, miR-374a-3p, miR-532-5p
Kidney chromophobe	let-7a-5p, let-7d-5p, let-7e-5p, let-7f-5p, let-7g-5p, let-7i-5p, miR-17-5p, miR-20a-5p, miR-25-3p, miR-26a-5p, miR-28-3p, miR-92a-3p, miR-99a-5p, miR-99b-5p, miR-101-3p, miR-103a-3p, miR-125a-5p, miR-126-5p, miR-140-3p, miR-186-5p, miR-200a-5p, miR-361-5p, miR-374a-3p
Kidney renal clear cell carcinoma	let-7d-5p, miR-22-3p, miR-23b-3p, miR-24-3p, miR-26a-5p, miR-26b-5p, miR-30d-5p, miR-30e-3p, miR-99b-5p, miR-191-5p, miR-361-5p
Liver hepatocellular carcinoma	let-7a-5p, let-7i-5p, miR-24-3p, miR-28-3p, miR-148b-3p, miR-361-5p
Lung adenocarcinoma	miR-23a-3p, miR-24-3p, miR-103a-3p, miR-361-5p
Lung squamous cell carcinoma	let-7g-5p, let-7i-5p, miR-15a-5p, miR-23a-3p, miR-28-5p, miR-30e-3p, miR-361-5p, miR-374a-3p
Prostate adenocarcinoma	let-7b-5p, let-7c-5p, miR-22-3p, miR-26a-5p, miR-27a-3p, miR-28-5p, miR-29a-3p, miR-29c-3p, miR-30e-3p, miR-30e-5p, miR-99b-5p, miR-100-5p, miR-101-3p, miR-134-5p, miR-151a-3p
Stomach adenocarcinoma	let-7a-5p, miR-28-5p, miR-101-3p, miR-140-3p, miR-152-3p, miR-374a-3p
Thyroid carcinoma	let-7d-5p, let-7i-5p, miR-15b-5p, miR-17-3p, miR-22-3p, miR-23a-3p, miR-24-3p, miR-27a-3p, miR-27b-3p, miR-29a-3p, miR-30d-5p, miR-30e-5p, miR-93-5p, miR-99a-5p, miR-99b-5p, miR-101-3p, miR-103a-3p, miR-128-3p, miR-141-5p, miR-151a-5p, miR-151a-3p, miR-192-5p, miR-194-5p, miR-200c-3p, miR-361-5p, miR-423-3p
Uterine corpus endometrial carcinoma	let-7a-5p, let-7i-5p, miR-28-3p, miR-30e-3p

Table 4

Candidate reference miRNAs and the corresponding numbers of tumor types for which they were identified.

No. of tumor types	Candidate reference miRNAs
9	miR-361-5p
8	let-7i-5p
5	let-7a-5p, miR-28-5p, miR-99b-5p
4	let-7d-5p, miR-22-3p, miR-24-3p, miR-28-3p, miR-30e-3p, miR-101-3p, miR-374a-3p
3	let-7g-5p, miR-23a-3p, miR-26a-5p, miR-29a-3p, miR-103a-3p
2	let-7b-5p, let-7e-5p, let-7f-5p, miR-27a-3p, miR-27b-3p, miR-30d-5p, miR-30e-5p, miR-99a-5p, miR-140-3p, miR-151a-3p, miR-186-5p
1	let-7c-5p, miR-100-5p, miR-20a-5p, miR-23b-3p, miR-25-3p, miR-26b-5p, miR-29c-3p, miR-92a-3p, miR-93-5p, miR-125a-5p, miR-126-5p, miR-128-3p, miR-134-5p, miR-141-5p, miR-148b-3p, miR-151a-5p, miR-152-3p, miR-15a-5p, miR-15b-5p, miR-17-3p, miR-17-5p, miR-191-5p, miR-192-5p, miR-194-5p, miR-200a-5p, miR-200c-3p, miR-361-3p, miR-378a-3p, miR-423-3p, miR-532-5p

for researchers studying miRNA. However, these candidate miRNAs also should be properly validated prior to their use in qRT-PCR studies.

The let-7 miRNA family is highly conserved in organs and species [32]. Many members of this family are on our candidate reference miRNA list, and are frequently mentioned in other studies on identification of reference miRNAs [12,22,27,31]. In contrast, the potential of miR-361-5p as a reference miRNA has not yet been mentioned. Additionally, we

observed that there are more miRNAs with stable expression and more candidate reference miRNAs in thyroid carcinoma, kidney chromophobe, and prostate adenocarcinoma, presumably due to the relatively low malignancy of these types of tumors.

As a rising high-throughput technology, miRNA-seq based on deep sequencing brings many advantages, such as high dynamic range in detection, high precision, high reproducibility, and a convenient comparison between different miRNAs and samples. It has

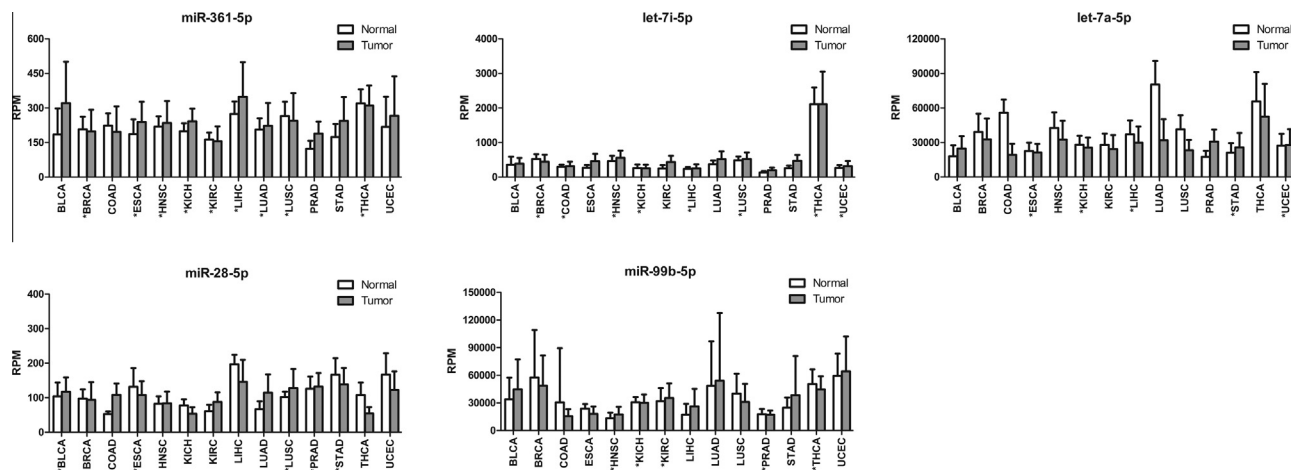


Fig. 1. Expression profiles of five candidate reference miRNAs in 14 tumor types. * Denotes the miRNA identified as a candidate reference miRNA in that tumor.

greatly expanded our knowledge and led to the discovery of many new miRNAs. To note, a new study discovered that miRNA measurement results derived from miRNA-seq were quite different from results derived from microarrays, which are based on nucleotide hybridization [33]. This phenomenon needs to be further investigated.

Acknowledgment

The results published here are based upon data generated by the TCGA Research Network (<http://cancergenome.nih.gov/>) and this analysis is supported by the National Natural Science Foundation of China (Grant Nos. 81401875, 81472225) (<http://www.nsfc.gov.cn/>) and Natural Science Foundation of Shanghai, China (Grant No. 14ZR1406000) (<http://www.stcsn.gov.cn/>).

References

- [1] J. Lin, C. Redies, Histological evidence: housekeeping genes beta-actin and GAPDH are of limited value for normalization of gene expression, *Dev. Genes. Evol.* 222 (2012) 369–376.
- [2] C. Zhan, Y. Zhang, J. Ma, et al., Identification of reference genes for qRT-PCR in human lung squamous-cell carcinoma by RNA-Seq, *Acta Biochim. Biophys. Sin. (Shanghai)* 46 (2014) 330–337.
- [3] R. Wei, E.A. Stewart, W.M. Amoaku, Suitability of endogenous reference genes for gene expression studies with human intraocular endothelial cells, *BMC Res. Notes* 6 (2013) 46.
- [4] W.C. Cheng, C.W. Chang, C.R. Chen, et al., Identification of reference genes across physiological states for qRT-PCR through microarray meta-analysis, *PLoS ONE* 6 (2011) e17347.
- [5] C. Wu, X. Wang, M. Zhong, et al., Evaluation of potential reference genes for qRT-PCR studies in human hepatoma cell lines treated with TNF- α , *Acta Biochim. Biophys. Sin. (Shanghai)* 45 (2013) 780–786.
- [6] D.V. Maltseva, N.A. Khaustova, N.N. Fedotov, et al., High-throughput identification of reference genes for research and clinical RT-qPCR analysis of breast cancer samples, *J. Clin. Bioinf.* 3 (2013) 13.
- [7] F.R. Meyer, H. Grausgruber, C. Binter, et al., Cross-platform microarray meta-analysis for the mouse jejunum selects novel reference genes with highly uniform levels of expression, *PLoS ONE* 8 (2013) e63125.
- [8] Z. Zhai, Y. Yao, Y. Wang, Importance of suitable reference gene selection for quantitative RT-PCR during ATDC5 cells chondrocyte differentiation, *PLoS ONE* 8 (2013) e64786.
- [9] N.C. Noriega, S.G. Kohama, H.F. Urbanski, Microarray analysis of relative gene expression stability for selection of internal reference genes in the rhesus macaque brain, *BMC Mol. Biol.* 11 (2010) 47.
- [10] B. Gur-Dedeoglu, O. Konu, B. Bozkurt, et al., Identification of endogenous reference genes for qRT-PCR analysis in normal matched breast tumor tissues, *Oncol. Res.* 17 (2009) 353–365.
- [11] L. Folkersen, S. Kurtovic, A. Razuvaev, et al., Endogenous control genes in complex vascular tissue samples, *BMC Genomics* 10 (2009) 516.
- [12] H.J. Peltier, G.J. Latham, Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues, *RNA* 14 (2008) 844–852.
- [13] V. Lamba, Y. Ghodke-Puranik, W. Guan, et al., Identification of suitable reference genes for hepatic microRNA quantitation, *BMC Res. Notes* 7 (2014) 129.
- [14] A. Schaefer, M. Jung, K. Miller, et al., Suitable reference genes for relative quantification of miRNA expression in prostate cancer, *Exp. Mol. Med.* 42 (2010) 749–758.
- [15] G. Zheng, H. Wang, X. Zhang, et al., Identification and validation of reference genes for qPCR detection of serum microRNAs in colorectal adenocarcinoma patients, *PLoS ONE* 8 (2013) e83025.
- [16] N. Ratert, H.A. Meyer, M. Jung, et al., Reference miRNAs for miRNAome analysis of urothelial carcinomas, *PLoS ONE* 7 (2012) e39309.
- [17] A. Kozomara, S. Griffiths-Jones, MiRBase: annotating high confidence microRNAs using deep sequencing data, *Nucleic Acids Res.* 42 (2014) D68–73.
- [18] G.J. Hurteau, S.D. Spivack, G.J. Brock, Potential mRNA degradation targets of hsa-miR-200c, identified using informatics and qRT-PCR, *Cell Cycle* 5 (2006) 1951–1956.
- [19] F. Benz, C. Roderburg, C.D. Vargas, et al., U6 is unsuitable for normalization of serum miRNA levels in patients with sepsis or liver fibrosis, *Exp. Mol. Med.* 45 (2013) e42.
- [20] X. Liu, L. Zhang, K. Cheng, et al., Identification of suitable plasma-based reference genes for miRNAome analysis of major depressive disorder, *J. Affect. Disord.* 163 (2014) 133–139.
- [21] A. Torres, K. Torres, P. Wdowiak, et al., Selection and validation of endogenous controls for microRNA expression studies in endometrioid endometrial cancer tissues, *Gynecol. Oncol.* 130 (2013) 588–594.
- [22] O. Timoneda, I. Balcells, S. Cordoba, et al., Determination of reference microRNAs for relative quantification in porcine tissues, *PLoS ONE* 7 (2012) e44413.
- [23] Y. Shen, Y. Li, F. Ye, et al., Identification of miR-23a as a novel microRNA normalizer for relative quantification in human uterine cervical tissues, *Exp. Mol. Med.* 43 (2011) 358–366.
- [24] K.H. Chang, P. Mestdagh, J. Vandesompele, et al., MicroRNA expression profiling to identify and validate reference genes for relative quantification in colorectal cancer, *BMC Cancer* 10 (2010) 173.
- [25] J. Carlsson, G. Helenius, M. Karlsson, et al., Validation of suitable endogenous control genes for expression studies of miRNA in prostate cancer tissues, *Cancer Genet. Cytogenet.* 202 (2010) 71–75.
- [26] P.A. Davoren, R.E. McNeill, A.J. Lowery, et al., Identification of suitable endogenous control genes for microRNA gene expression analysis in human breast cancer, *BMC Mol. Biol.* 9 (2008) 76.
- [27] J. Song, Z. Bai, W. Han, et al., Identification of suitable reference genes for qPCR analysis of serum microRNA in gastric cancer patients, *Dig. Dis. Sci.* 57 (2012) 897–904.
- [28] S. Jurcevic, B. Olsson, K. Klinga-Levan, Validation of suitable endogenous control genes for quantitative PCR analysis of microRNA gene expression in a rat model of endometrial cancer, *Cancer Cell Int* 13 (2013) 45.
- [29] T.C. Roberts, A.M. Coenen-Stass, M.J. Wood, Assessment of RT-qPCR normalization strategies for accurate quantification of extracellular microRNAs in murine serum, *PLoS ONE* 9 (2014) e89237.
- [30] A.M. McDermott, M.J. Kerin, N. Miller, Identification and validation of miRNAs as endogenous controls for RQ-PCR in blood specimens for breast cancer studies, *PLoS ONE* 8 (2013) e83718.
- [31] X. Chen, H. Liang, D. Guan, et al., A combination of Let-7d, Let-7g and Let-7i serves as a stable reference for normalization of serum microRNAs, *PLoS ONE* 8 (2013) e79652.
- [32] S. Roush, F.J. Slack, The let-7 family of microRNAs, *Trends Cell Biol.* 18 (2008) 505–516.
- [33] Y.W. Wan, C.M. Mach, G.I. Allen, et al., On the reproducibility of TCGA ovarian cancer microRNA profiles, *PLoS ONE* 9 (2014) e87782.