

Identification of RNU44 as an Endogenous Reference Gene for Normalizing Cell-Free RNA in Tuberculosis

Wei Gu,¹ Xilin Tu,² Weinan Lu,¹ Yian Yin,¹ Qingtai Meng,¹ Xinyang Wang,¹ Fengmin Zhang,^{1,6} and Yingmei Fu¹

¹Department of Microbiology, Harbin Medical University, Harbin, China, and ²Department of Respiratory Medicine, First Affiliated Hospital, Harbin Medical University, Harbin, China

Background. Normalization of cell-free RNA (cf-RNA) concentration can be affected by variable experimental conditions and thus impact the performance of their diagnostic potential. Our study aimed to identify appropriate endogenous reference genes for cf-RNA biomarker evaluation in the diagnosis of tuberculosis (TB).

Methods. Subjects consisting of patients with TB with and without malignancy, patients with pneumonia, and healthy controls were recruited. Candidate reference genes were screened and identified by literature reviewing and RNA-Seq analysis. Expression levels of the candidate genes were determined by reverse-transcription real-time quantitative polymerase chain reaction in plasma from patients with TB and healthy controls. The stability of gene expression was assessed by geNorm, NormFinder, BestKeeper, the Comparative Delta Ct method, and RefFinder. Differential expression of 2 small RNAs (sRNAs) encoding by genome of *Mycobacterium tuberculosis* in plasma of patients with TB were determined by both absolute quantification and relative quantification with candidate reference genes.

Results. According to the stability ranking analyzed with the 5 computational programs, the top 4 candidates—miR-93, RNU44, miR-16, and glyceraldehyde 3-phosphate dehydrogenase—were used to normalize the transcript levels of 2 mycobacterial sRNAs, MTS2823 and MTS1338, which were observed to have higher copy numbers in the plasma of patients with TB. Normalization with RNU44 displayed significantly higher levels of the 2 *M tuberculosis* sRNAs in the patients' plasma than those of healthy controls.

Conclusions. RNU44 was demonstrated as a proper endogenous gene for cf-RNA normalization in TB diagnosis.

Keywords. biomarker; cf-RNA; plasma; reference gene; tuberculosis.

Tuberculosis (TB) remains one of the top 10 causes of death globally, with about a quarter of the world's population infected with *Mycobacterium tuberculosis* (MTB) [1]. In 2020, only 59% of TB cases globally were bacteriologically confirmed [1]. Cell-free RNA (cf-RNA) molecules can be detected in the extracellular fluid, either leaked from dead cells or secreted actively by vesicles, or transported by RNA-binding proteins [2]. Cf-RNA in biofluid has recently emerged as a promising non-invasive biomarker for TB [3].

To quantify cf-RNA in human biofluid by real-time polymerase chain reaction (PCR) assay, nonhuman spike-in RNAs have commonly been used for normalization. Nevertheless, the evaluation of cf-RNA concentrations can be skewed by preanalytical processing conditions, such as specimen type, storage conditions,

processing protocols, and quantification strategies. These factors, together with variable patient-specific baselines and batch effects, may lead to a lack of reproducibility and poor consistency for cf-RNA biomarkers among different studies [4]. In these cases, endogenous reference genes can outcompete the exogenous reference RNAs in generating better compatibility by removing the sample-to-sample variations.

Nevertheless, selection of suitable endogenous references can be challenging when measuring cf-RNA transcripts. On one hand, some well-established intracellular references, such as the transcripts of β -actin, are liable to be degraded once released into plasma [5]. On the other hand, sequence length, concentration, and profile of extracellular RNAs differ largely from their intracellular counterpart in different pathophysiological conditions [6, 7]. As such, one can note that the applications of endogenous reference genes for cf-RNA measurement lack consistency when comparing the quantitative PCR (qPCR) results from different studies [8]. For TB, there were even contradictory reports on the expression of specific cf-RNA in circulation. It was proposed that miR-93 and miR16 [9, 10] be used as endogenous reference genes for normalizing cf-RNA in TB diagnosis. Conversely, the 2 microRNAs (miRNAs) were found to be differentially expressed in active TB and associated with patient response to therapy [11, 12]. Therefore, identification of appropriate reference genes for normalization is critical to optimize the clinical reliability and

Received 14 July 2022; editorial decision 06 October 2022; accepted 11 October 2022

Correspondence: Yingmei Fu, MD, PhD, Department of Microbiology, Harbin Medical University, 194, Xuefu Road, Nangang District, Harbin, 150081, China (fuyingmei@hrbmu.edu.cn).

Open Forum Infectious Diseases®

© The Author(s) 2022. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
<https://doi.org/10.1093/ofid/ofac540>

reproducibility of cf-RNA markers. This study aimed to identify the reference genes suitable for normalization of cf-RNA by evaluating the stability of selected candidate reference genes and verifying their efficacy in determining diagnostic biomarkers for TB.

METHODS

Patients and Healthy Controls

Blood samples of 50 patients with active pulmonary TB and 30 patients with community-acquired pneumonia (CAP) were collected at the Infectious Disease Hospital of Heilongjiang Province and the First Affiliated Hospital of Harbin Medical University. Patients with TB were diagnosed by clinical manifestations, chest radiography, microscopic detection, and/or MTB culture. Plasma of 12 TB patients with malignancies (8 with lung cancers, 1 with gastric cancer, 1 with nasopharyngeal cancer, 1 with esophageal cancer, and 1 with breast cancer) were collected from Harbin Chest Hospital. The CAP patients were diagnosed by negative acid-fast bacilli smear staining and/or sputum culture but shared similar symptoms to active pulmonary TB. Fifty healthy controls, with no clinical symptoms, normal routine blood tests, and chest radiographs, were collected from the physical examination center of the Second Affiliated Hospital of Harbin Medical University. Written informed consent was obtained from all participants. Ethical permission was approved by the Ethics Committee of Harbin Medical University (certificate number HMUIRB20190014PR1). Venous blood samples collected with ethylenediaminetetraacetic acid tubes were placed in a cooler with ice packs until they were transported to the laboratory within 5 hours. Samples were spun down by centrifugation for 10 minutes at 2500 rpm at 4°C to obtain plasma. Aliquots of 500 µL of plasma in 2 mL microcentrifuge tubes were stored at -80°C until use.

Literature Review of Endogenous Reference Genes

The keywords “reference,” “calculating,” “RNA,” “serum,” and “plasma” were used to search in the PubMed literature database (<https://pubmed.ncbi.nlm.nih.gov/>). We reviewed the literature associated with reference genes in circulating body fluids, mainly plasma and serum, over the last decade. The biological type of reference cf-RNA, sample, disease, and country of the population were retrieved.

Analysis of RNA-Seq Data

The reference genome of *Homo sapiens* GRCh38.p13 and the Genome Annotation file were downloaded from the National Center for Biotechnology Information (NCBI) genome database (<https://www.ncbi.nlm.nih.gov/genome/>). RNA sequencing (RNA-Seq) datasets of serum exosomes of patients with TB (n = 3) and healthy individuals (n = 4) were downloaded from the NCBI Sequence Read Archive Database (SRA; SRP270842) (Supplementary Table 3). The quality of

RNA-Seq reads was assessed, trimmed, and filtered using Fastqc and Trim Galore [13], respectively. Hisat2 [14] was used to build the index of the reference genome and aligned RNA-Seq reads to the reference genome. Samtools [15] was used to sort and index BAM files. Finally, the depth coverage of candidate endogenous was read by Bedtools [16].

Candidate reference genes from different classes of RNA biotypes (messenger RNA [mRNA], ribosomal RNA [rRNA], small nucleolar RNA [snoRNA], small nuclear RNA [snRNA], and miRNA) were screened based on their log₂ fold change (log₂FC), *P* value ($|\log_2\text{FC}| < 2$, $P > .05$), and abundance (Fragments Per Kilobase Million [FPKM]).

RNA Extraction

Total RNA was extracted from 500 µL of human plasma with TRIzol LS Reagent (Thermo Fisher Scientific, Wilmington, North Carolina) according to the manufacturer’s instructions. Concentration of total RNA was measured using a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific). A 260/A 280 was obtained to assess the RNA purity.

Reverse-Transcription and Real-Time PCR Quantification

Complementary DNA was synthesized from 1 µg of total RNA using 5× All-In-One RT MasterMix kit (Applied Biological Materials [abm], Richmond, Canada) in 20 µL of reaction system. For real-time PCR, BlasTaq 2× qPCR MasterMix (abm) was used. Real-time PCR was performed in the LightCycler 96 (Roche, Shanghai, China) with the following conditions: 95°C for 180 seconds, followed by 40 cycles at 95°C for 10 seconds and 60°C for 60 seconds. Primers were designed with primer 3.0 online software (<https://primer3.ut.ee/>) (Supplementary Table 4). Primers and RNA standards were synthesized by Comate Bioscience Company (Changchun, China) and GenScript Company (Nanjing, China), respectively. The specificity of all primers was checked by sequencing qPCR products and melting curves analysis. All experiments were performed in triplicate. The mean cycle threshold (Ct) value of each sample was determined from the triplicate reactions.

Statistical Analysis

IBM SPSS Statistics 26 software (IBM SPSS, Chicago, Illinois) was used for statistical analysis, and GraphPad Prism 9.0 (GraphPad Software, San Diego, California) software was used to generate graphs. One-way analysis of variance was used to compare candidate endogenous reference gene Ct values among patients with TB, patients with CAP, and healthy controls. The Mann-Whitney *U* test was used to compare MTB small RNA (sRNA) copy numbers and relative expression differences between TB patients and healthy controls, and Ct values of RNU44 between TB patients with and without malignancies. *P* < .05 was considered statistically significant. The

relative expression level of each sRNA in plasma was calculated based on $2^{-\Delta Ct}$ ($\Delta Ct = \text{target gene Ct} - \text{reference gene Ct}$).

Four independent computational programs—geNorm [17], NormFinder [18], BestKeeper [19], and Comparative Delta Ct method [20]—were used to evaluate the expression stability of reference genes. RefFinder [21] was used to generate an overall comprehensive ranking.

RESULTS

Preliminary Screening of Endogenous Reference Genes

As shown in [Supplementary Table 1](#), by reviewing the literatures in the last decade, 31 genes were found to be used to normalize cf-RNA concentration as endogenous reference genes. The RNA biotypes of these reference genes belong to miRNA, snRNA, snoRNA, rRNA, and mRNA fragments. They were detected to be expressed primarily from plasma or serum in patients with cancer, pulmonary hypertension, and hepatitis B.

Differential Expression of 31 Reference Genes Between Patients With TB and Healthy Controls

We searched and downloaded a set of RNA-Seq data of serum exosomes from the NCBI-SRA database and performed transcriptome analysis. The FPKM values of 31 reference genes were calculated by the formula $\text{FPKM} = (\text{ExonMappedReads} \times 10^9) / (\text{TotalMappedReads} \times \text{ExonLength})$. Fold changes of reference genes were used to compare the differences between patients with TB and healthy controls. As shown in [Figure 1](#), levels of 20 reference genes were similar between the 2 groups ($|\log_2\text{FC}| < 2, P > .05$).

Based on their abundance and biotype classification, 8 RNAs were chosen as candidate reference genes for further evaluation. These include 5 candidates corresponding to their respective biotypes: RNU6 (snRNA), RNU44 (snoRNA), 18SrRNA and 28SrRNA (rRNA), and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) (mRNA). Among the larger number of miRNAs relative to other 4 classes of RNAs that meet the requirements, 3 candidate miRNAs (let-7a, miR-16, and miR-93) were randomly chosen from the top 50% miRNAs ranked by their FPKM values ([Figure 1](#)). Due to their lower expression abundance (FPKM), miR-101 and miR-222 were excluded, despite their lower fold change between the TB and the healthy control groups.

Detection of Candidate Reference Genes in Plasma Samples

The transcriptional levels of the 8 candidate endogenous reference genes were detected in 90 plasma samples from patients with TB ($n = 30$), patients with CAP ($n = 30$), and healthy controls ($n = 30$). As shown in [Figure 2](#), there was no significant difference among all the levels of candidate endogenous reference genes in the TB patient group, the CAP patient group, and

the healthy control group ($P > .05$). The 8 candidate reference genes could be used for subsequent analysis.

Stability Analysis

Ct values were converted into nonnormalized relative quantities $2^{-\Delta Ct}$, and the ΔCt corresponds to the difference between Ct value of the candidate genes and the lowest Ct value of the respective gene among the 90 study participants [22]. These data were imported into 2 Microsoft Excel-based statistics tools, geNorm and NormFinder, for stability analysis. As shown in [Figure 3A and 3B](#), the 2 programs identified a same panel of top 4 genes (RNU44, miR-16, GAPDH, and miR-93). Results by geNorm showed an acceptable stability range (stability value $[M] \leq 1.5$) of the 4 RNAs—that is, RNU44 ($M = 1.301$), miR-16 ($M = 1.301$), GAPDH ($M = 1.388$), and miR-93 (1.491) ([Figure 3A](#)). NormFinder stability analysis showed a different ranking of the genes in which miR-93 had the best stability ($M = 0.913$), followed by GAPDH ($M = 1.046$), RNU44 ($M = 1.319$), and miR-16 ($M = 1.352$). BestKeeper evaluates and ranks the stability of candidates by calculating the SD and the coefficient of variation (CV) of Ct values, and the correlation factors (r) [19]. It demonstrated that RNU44, GAPDH, and miR-16 were stable ($SD < 1$) ([Figure 3C](#)). Except for miR-16 ($r = -0.024, P > .05$), the expression levels of the other 7 genes had strong positive correlation with each other ($r > 0.5, P < .05$) ([Supplementary Table 2](#)), indicating that the 7 genes are expressed cooperatively and consistently within a same individual. The Comparative Delta Ct method analyzes the expression of a reference gene relative to every other gene in all samples (ΔCt) and calculates the average value of the ΔCt SD [20]. By analysis of qPCR data with Comparative Delta Ct method, miR-93 was shown to be the most stable endogenous reference gene, with an average SD of 1.880, followed by GAPDH (1.908), RNU44 (1.991), and let-7a (2.088) ([Figure 3D](#)).

To obtain the comprehensive ranking of candidate reference genes, RefFinder was used to analyze the 4 datasets from geNorm, NormFinder, BestKeeper, and Comparative Delta Ct method, generating a final comprehensive ranking of 8 candidates. As shown in [Figure 3E](#), the top 4 reference genes are identified, that is, miR-93, RNU44, miR-16, and GAPDH.

Differential Expression of Target cf-RNAs in Patients With TB by Absolute Quantitative Real-Time PCR

The MTB-derived sRNAs MTS2823 and MTS1338 have been found to play a vital role in the stages of MTB infection [23]. These sRNAs were also reported to be highly expressed in the plasma of patients with TB [3]. Hereby the 2 cf-RNAs of bacterial origin were used as the target genes for evaluating the efficacy of the candidate reference genes. Absolute quantitative reverse-transcription PCR (RT-qPCR) analysis results showed that the copy numbers of plasma MTS2823 and MTS1338 in

Differential Expression of Reference Genes (TB vs HC)

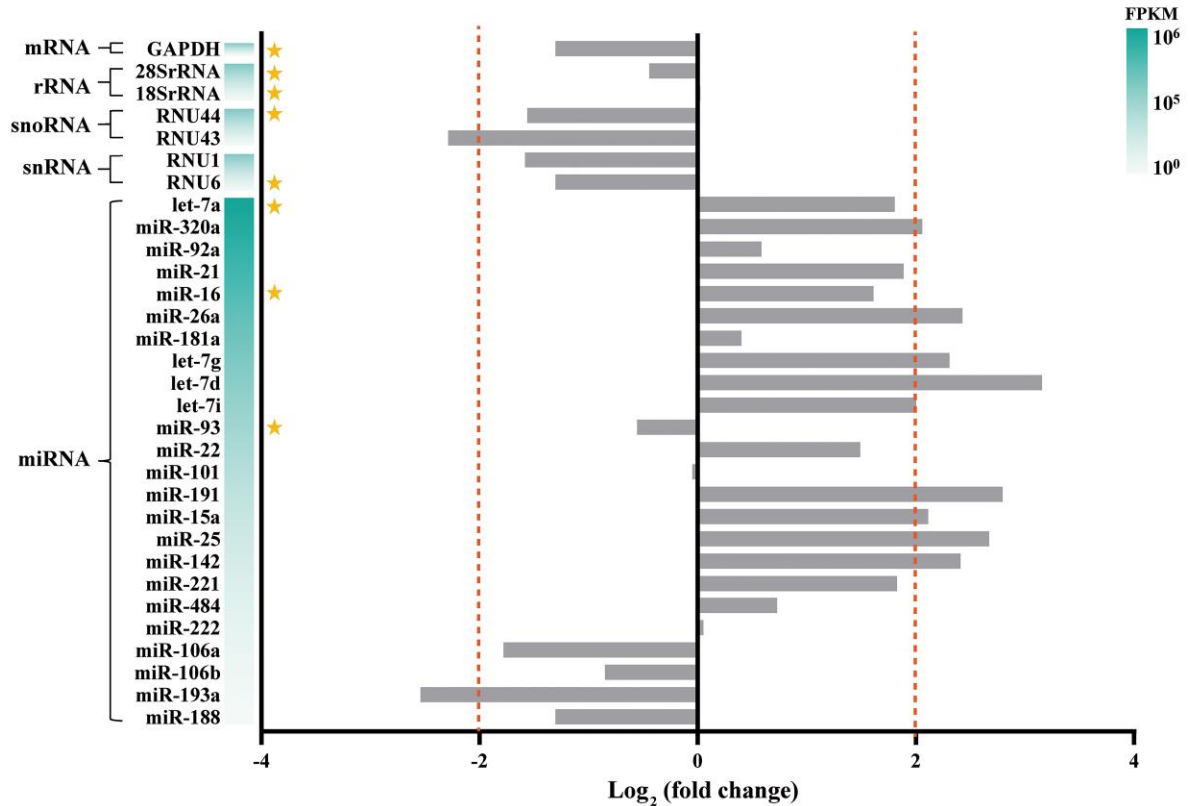


Figure 1. Comparison of preliminary reference gene expression in serum exosomes between patients with tuberculosis (TB) and healthy controls (HC) in terms of fold change as measured by RNA-Seq. Candidate genes are marked with an asterisk. FPKM, Fragments Per Kilobase Million.

TB patients were significantly higher than those of healthy people (Figure 4).

Differential Expression of Target cf-RNAs in TB Patients After Normalization by Reference Genes

According to the comprehensive ranking from RefFinder software, the top 4 RNAs, including miR-93, RNU44, miR-16, and GAPDH, were selected to verify their efficacy as reference genes in reflecting the difference of the bacterial genome-derived sRNA in plasma between the patients and healthy controls. The results showed that when RNU44 was used as the reference gene, the expression levels of MTS2823 and MTS1338 in the plasma of the patients with TB were significantly higher than those of the healthy controls (Figure 5A), which was consistent with the results of absolute quantification. Comparatively, miR-16 can only show the differential level of MTS1338 between the 2 groups (Figure 5B). The other 2 genes, miR93 and GAPDH, failed to exhibit the bacterial sRNA transcript difference.

To compare the fold increase in relative quantification and absolute quantification, we calculated the fold change of plasma

sRNA level by the absolute quantification and the RNU44 normalization method. By absolute quantification, 1.61-fold and 1.88-fold increases were obtained for MTS2823 and MTS1338, respectively. RNU44 normalization demonstrated 2.28-fold and 1.46-fold increases for the 2 sRNAs. As shown in Supplementary Figure 1, a >1-fold increase of plasma sRNA level was determined in TB patients by both methods.

Determination of the Variation in Ct Values of RNU44 in TB Patients With Malignancies

RNU44 was reported to be associated with certain malignancies [24]. Level of RNU44 was determined by RT-qPCR in 12 TB patients with cancer. The concomitant malignancies included lung cancer, gastric cancer, nasopharyngeal cancer, esophageal cancer, and breast cancer. As shown in Supplementary Figure 2, there was no significant difference in the Ct values of RNU44 between TB patients with malignancies and those without malignancies. However, the expression of RNU44 was more variable in the TB with malignancies group (interquartile range [IQR], 2.759) than in the TB without malignancies group (IQR, 1.198). The appropriate reference interval

Ct Values of Candidate Endogenous Genes

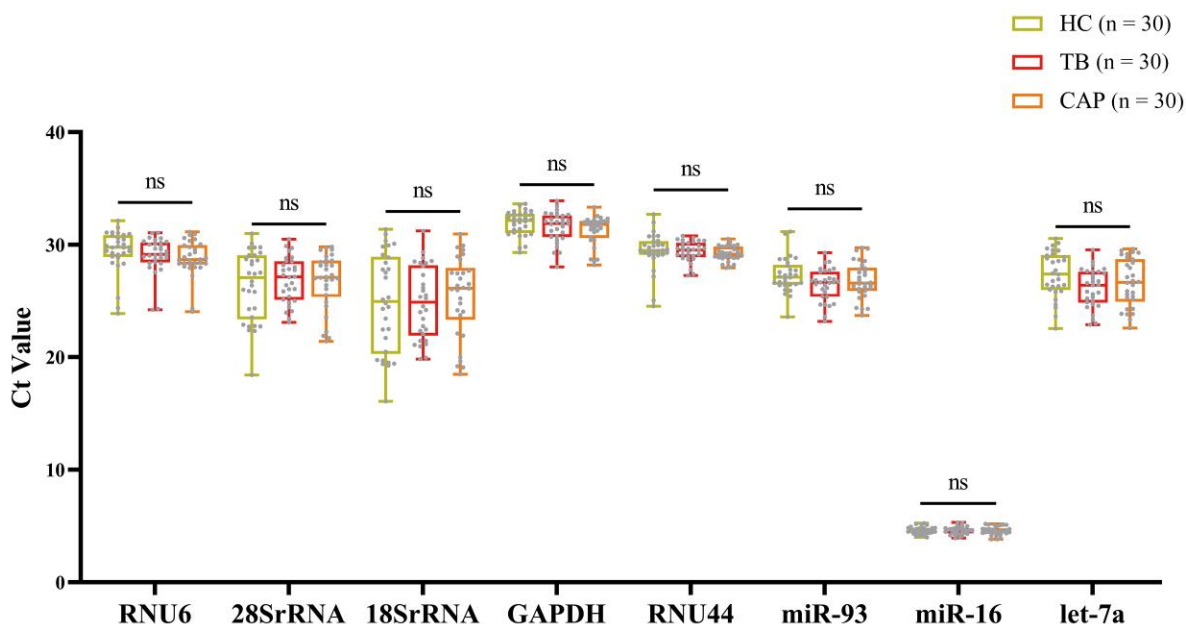


Figure 2. Comparison of cycle threshold (Ct) values of candidate endogenous reference genes in plasma of patients with tuberculosis (TB), patients with community-acquired pneumonia (CAP), and healthy controls (HC). The center line indicates the median; the lower and upper lines show the interquartile range corresponding to the 25th and 75th percentiles, respectively. ns, not statistically significant.

(2.5th–97.5th percentile, 27.487–31.851) was obtained, according to the result of RNU44 Ct values of TB patients without malignancies. Data should be discarded when the Ct values of RNU44 of TB patients with malignancies exceed the reference interval.

DISCUSSION

Emerging studies have demonstrated the potential diagnostic value of cf-RNAs for TB, including host- and pathogen-derived sRNAs [3, 25, 26]. However, many technical and sample-related factors may impact the results of cf-RNA quantification, such as specimen type, storage condition, RNA isolation method, and the donor's physical state [4]. When evaluating circulating cf-RNA levels in diagnostic assays, normalizing the qPCR results with a suitable endogenous reference gene is usually employed to correct intersample variations induced by both preanalytical and qPCR workflows.

Through reviewing the literature involved in cf-RNA over the last 10 years, 31 reference genes, which had been used in various diseases, such as cancer [27], pulmonary hypertension [28], and hepatitis B [29], were retrieved. By analyzing the RNA sequencing data of serum exosomes, 20 of them were found to have similar levels in TB patients when compared with healthy controls. Based on their abundance and biotype classification, 8 RNAs were chosen as candidate reference genes for further

evaluation. These include 5 candidates corresponding to their respective biotypes—RNU6 (snRNA), RNU44 (snoRNA), 18SrRNA and 28SrRNA (rRNA), and GAPDH (mRNA)—and 3 miRNAs (let-7a, miR-16, and miR-93). The 3 selected miRNAs were also frequently reported as reference genes in cf-RNA quantification [9, 30–32].

Four widely used programs (geNorm, NormFinder, BestKeeper, and Comparative Delta Ct) were employed to analyze the stability of candidate reference genes based on RT-qPCR results. geNorm [17] calculates the gene-stability measure M for all reference genes in a given set of samples. NormFinder [18] is based on a mathematical model of gene expression, which can not only estimate the overall variation of candidate reference genes, but also estimates the variation between sample subgroups of a sample set. BestKeeper [19] is an Excel-based spreadsheet software application. In addition to analyzing the reference gene data SD and CV, it also estimates the relationship between all possible reference genes and performs a large number of paired correlation analyses. The Comparative Delta Ct method [20] is used to compare gene pairs, which bypasses the need to accurately quantify input RNA and instead uses Δ Ct comparisons between genes. RefFinder [21] is online software that combines the algorithms of geNorm, BestKeeper, NormFinder, and other software packages. Relying on RefFinder, we obtained the final comprehensive ranking of reference genes stability (Figure 3E). The top

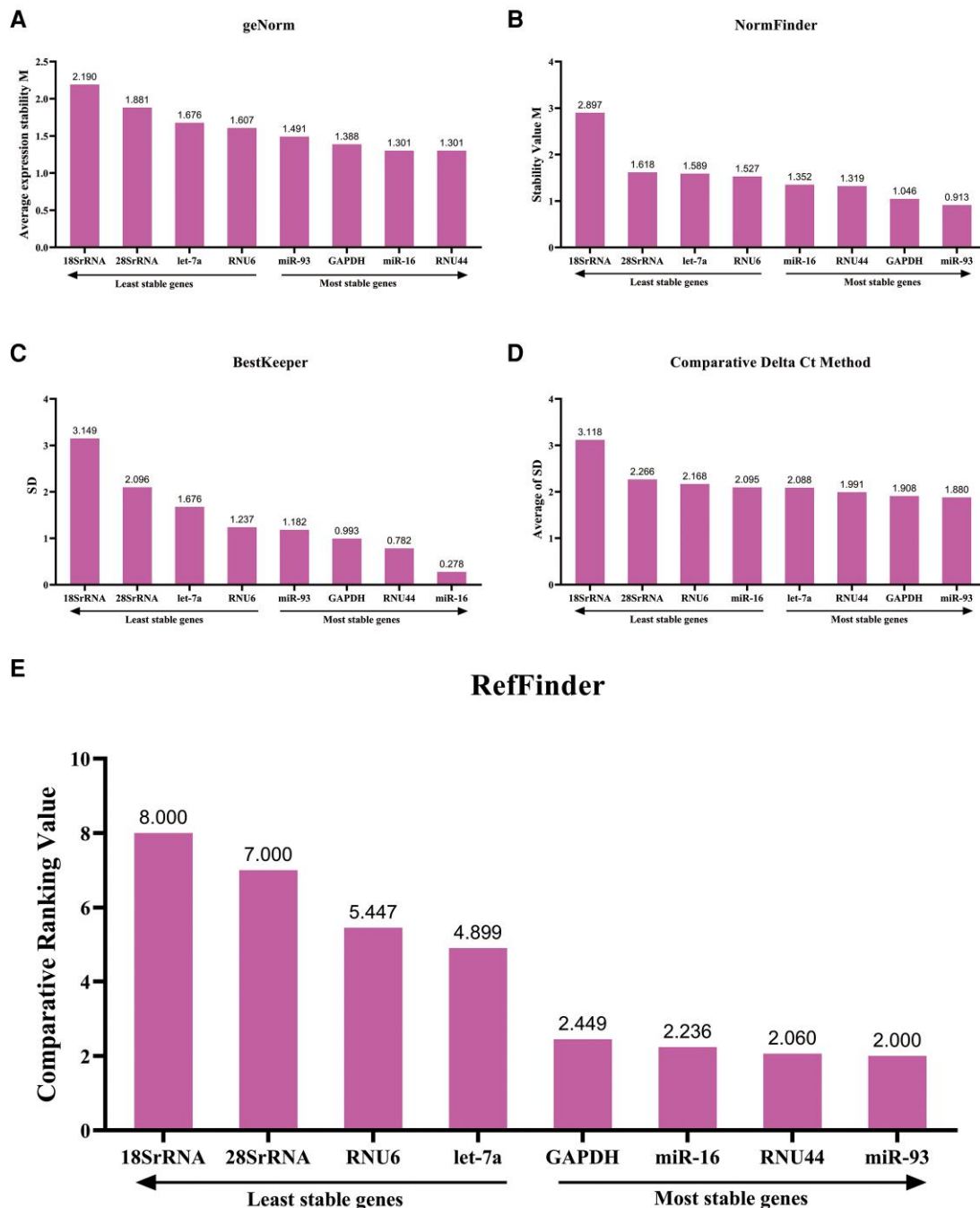


Figure 3. Stability analysis of candidate endogenous reference genes, showing results from geNorm (A), NormFinder (B), BestKeeper (C), comparative Δ Ct method (D), and RefFinder (E). The lower the value, the better the stability of the candidate reference genes. Abbreviations: Ct, cycle threshold; SD, standard deviation.

4 candidate reference genes (miR-93, RNU44, miR-16, and GAPDH) were selected to verify the differences in the expression of plasma target genes of patients with TB and healthy controls.

Numerous studies have reported TB-specific cf-RNAs of human genome origin; however, consistency between studies is very limited [25]. Comparatively, levels of MTS2823 and

MTS1338, sRNAs derived from MTB, have recently been demonstrated to be elevated in infection sites and TB patient plasma [3, 23, 33]. In the validation cohort, absolute quantification of 2 sRNAs in human plasma revealed higher copy numbers of MTS2823 and MTS1338 in people with TB than those of healthy people (Figure 4). MiR-16 could only reflect the difference of MTS1338 (Figure 5B). Although ranking the first in our

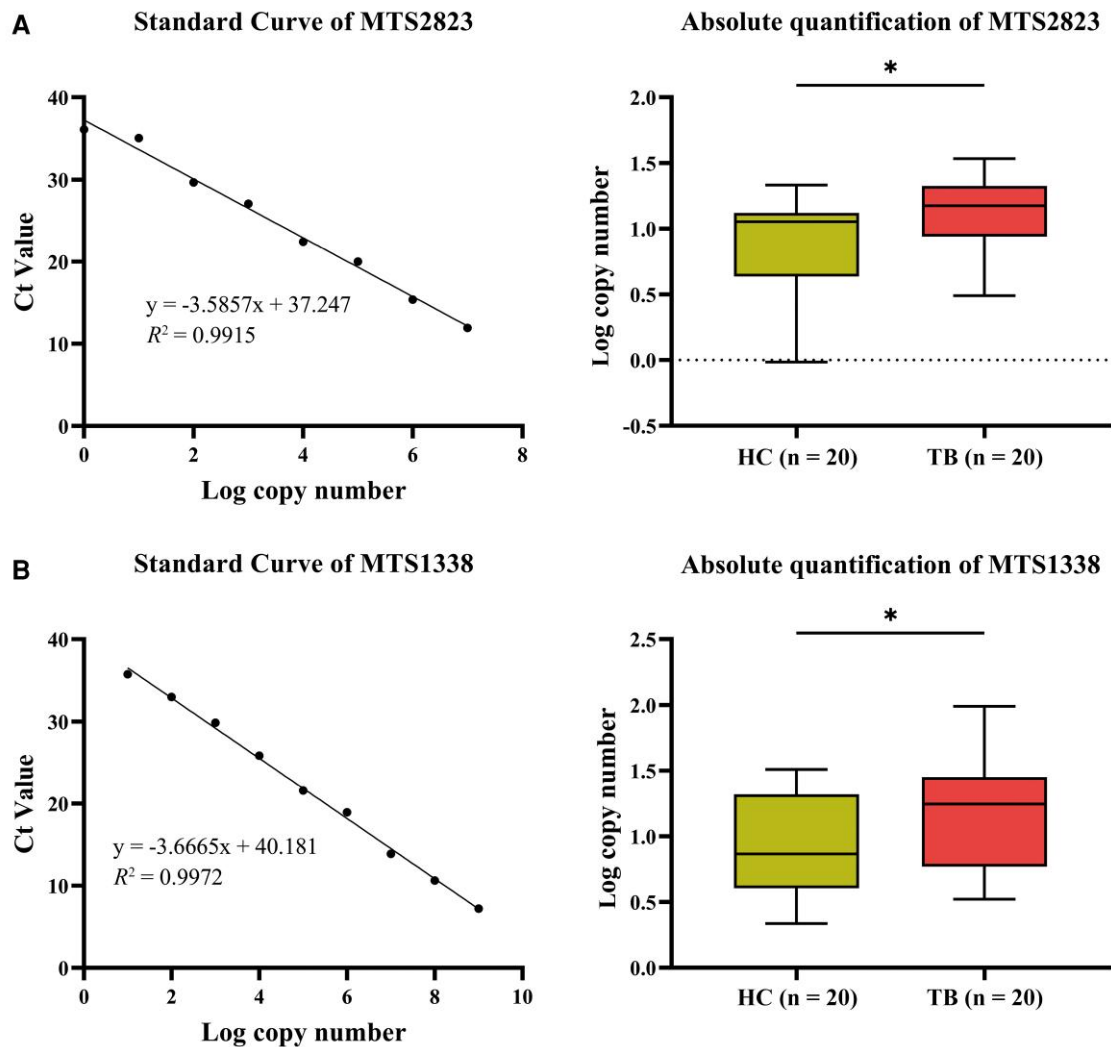


Figure 4. Comparisons of the transcript level of mycobacterial small RNAs by absolute quantitative reverse-transcription polymerase chain reaction in plasma from patients with tuberculosis and healthy controls. A, MTS2823. B, MTS1338. $R^2 > 0.99$. $*P < .05$. Abbreviations: Ct, cycle threshold; HC, healthy controls; TB, tuberculosis.

software analyzing and reported as proper reference gene in TB previously by others [9, 34], miR-93 failed to normalize the level difference of the 2 bacterial sRNAs. The mRNA transcripts, GAPDH, turned inappropriate for the normalization. When we used the candidate endogenous reference genes to standardize the pathogen-derived cf-RNAs, RNU44 displayed differential expression of MTS2823 and MTS1338 in patients' plasma (Figure 5A). RNU44 turned out to be the most suitable one for normalizing the cf-RNA level to distinguish TB patients and healthy controls.

RNU44 is encoded by the intron of growth arrest specific 5 (GAS5) gene. As a snoRNA, its performance as reference gene for quantification of miRNA had been demonstrated in a series of human physiological and pathological conditions, including fetoplacental growth [35, 36], cartilage development

[37], osteoblast differentiation [38], immunoglobulin A nephropathy [39], and tumors (pituitary tumors [40], prostate tumors [41], kidney cancer [42], and breast cancer [43]). Meanwhile, RNU44 has also shown to be associated with certain tumors [24]. Therefore, it is worth to explore how many biases would introduce when determining cf-RNA expression in TB concomitant with malignancies by RNU44 as the reference gene. We found that there was no significant change in the Ct values of RNU44 in the plasma of TB patients with malignancies compared to TB without malignancies. However, the IQR of the 2 groups indicated that the RNU44 expression was more variable in the TB patients with malignancies. These results indicate that RNU44 is applicable to TB combined with tumors when Ct values of RNU44 fall within the reference interval (27.487–31.851).

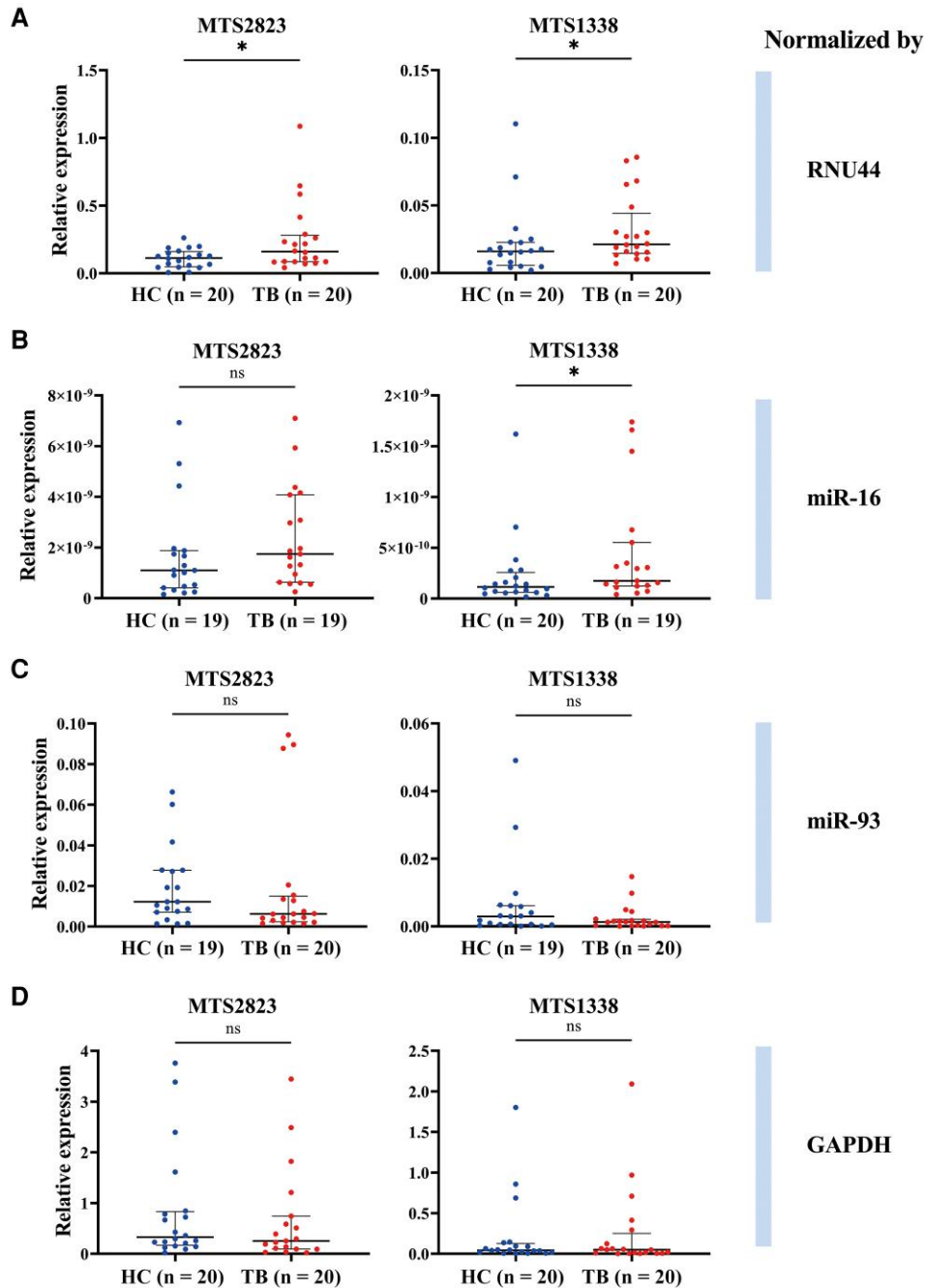


Figure 5. Comparisons of the relative expression of small RNAs MTS2823 and MTS1338 in plasma of patients with tuberculosis (TB) and healthy controls (HC) normalized by reference genes. Target RNAs are normalized by RNU44 (A), miR-16 (B), miR-93 (C), and GAPDH (D). The center line indicates the median, and the lower and upper lines show interquartile range corresponding to the 25th and 75th percentiles, respectively. ns, not statistically significant. * $P < .05$.

It is of interest that a few individuals with higher levels of MTS1338 in healthy controls presented in each of the 4 normalization procedures. These higher outliers were observed in different individuals in the 4 datasheets. In contrast to MTS1338, high expression levels of MTS2823 in healthy controls were not demonstrated when normalizing with RNU44. This gene-specific effect may be attributed to the differences of performance of

reference genes and/or the potential role of MTS1338 in latent tuberculosis infection (LTBI).

MTS1338 was reported to be induced in stationary phase of growth [44], dormancy state [45], macrophage-engulfed mycobacteria [46], and the lungs of mice during chronic TB infection [44]. Transcription of MTS1338 is controlled by the dormant transcriptional regulator DosR, which is activated under

hypoxia and nitric oxide-induced stresses [47]. These results suggest that MTS1338 play important roles for dormancy and LTBI. According to the survey reported in 2015, the incidence of LTBI in healthy persons in China is about 19%–20% [48]. There is possibility that LTBI might be present in healthy controls in our study. Considering that the interferon- γ release assay was not available and the history of MTB exposure was unknown in the healthy controls recruited in this study, further investigation by a separate cohort will clarify the expression level of MTB sRNAs in LTBI and the dynamics of expression of sRNAs during the infection of MTB.

In conclusion, a snoRNA, RNU44, was finally determined as a suitable reference gene for TB cf-RNA quantification. The present study presents valuable data for future clinical evaluation of cf-RNA diagnostic biomarkers in areas with a high burden of TB. With genome-wide screening from the high-throughput RNA sequencing datasets and a larger cohort, more endogenous reference genes might be discovered as proper normalizers for improving the clinical reproducibility and the performance of diagnostic cf-RNA biomarkers for TB.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Author contributions. W. G. contributed to conceptualization, investigation, methodology, formal analysis, data curation, writing—original draft, validation, and software. X. T. contributed to investigation, methodology, formal analysis, data curation, and validation. W. L. and Y. Y. contributed to investigation, validation, and software. Q. M. and X. W. contributed to original draft and validation. F. Z. and Y. F. contributed to conceptualization, validation, supervision, project administration, and writing—review and editing. F. Z. and Y. F. contributed equally to this work.

Acknowledgments. The authors thank Yuqin Liu, Zhiguo Yue, Na Li, and Long Jin from the Infectious Disease Hospital of Heilongjiang Province, and Fanhui Kong from the Harbin Chest Hospital, for their assistance with sample collection. The authors also thank Yanhong Liu (Second Affiliated Hospital of Harbin Medical University) for significant support and effort in promoting the study and helping it run smoothly.

Patient consent. Informed consent was obtained from the patients and healthy volunteers to participate in the study. The design of the work was approved by the Ethics Committee of Harbin Medical University, China (certificate number HMUIRB20190014PR1).

Financial support. This study was supported by the Ministry of Science and Technology of the People's Republic of China (grant number 2017ZX10201301–003–005) and the Natural Science Foundation of China (grant numbers 31370164 and 81701613).

Potential conflicts of interest. The authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

- World Health Organization (WHO). Global tuberculosis report 2020. Geneva, Switzerland: WHO; 2021.
- Zhao C, Sun X, Li L. Biogenesis and function of extracellular miRNAs. *ExRNA* 2019; 1:38.
- Han X, Li T, Fan Y, et al. Screening of 20 *Mycobacterium tuberculosis* sRNAs in plasma for detection of active pulmonary tuberculosis. *Tuberculosis (Edinb)* 2021; 129:102086.
- Lee I, Baxter D, Lee MY, Scherler K, Wang K. The importance of standardization on analyzing circulating RNA. *Mol Diagn Ther* 2017; 21:259–68.
- Holford NC, Sandhu HS, Thakkar H, Butt AN, Swaminathan R. Stability of β -actin mRNA in plasma. *Ann N Y Acad Sci* 2008; 1137:108–11.
- MacLellan SA, MacAulay C, Lam S, Garnis C. Pre-profiling factors influencing serum microRNA levels. *BMC Clin Pathol* 2014; 14:27.
- Zhou Z, Wu Q, Yan Z, et al. Extracellular RNA in a single droplet of human serum reflects physiologic and disease states. *Proc Natl Acad Sci U S A* 2019; 116:19200–8.
- Madadi S, Schwarzenbach H, Lorenzen J, Soleimani M. MicroRNA expression studies: challenge of selecting reliable reference controls for data normalization. *Cell Mol Life Sci* 2019; 76:3497–514.
- Barry SE, Chan B, Ellis M, et al. Identification of miR-93 as a suitable miR for normalizing miRNA in plasma of tuberculosis patients. *J Cell Mol Med* 2015; 19:1606–13.
- Qi Y, Cui L, Ge Y, et al. Altered serum microRNAs as biomarkers for the early diagnosis of pulmonary tuberculosis infection. *BMC Infect Dis* 2012; 12:384.
- Fu Y, Yi Z, Wu X, Li J, Xu F. Circulating microRNAs in patients with active pulmonary tuberculosis. *J Clin Microbiol* 2011; 49:4246–51.
- Wagh V, Urhekar A, Modi D. Levels of microRNA miR-16 and miR-155 are altered in serum of patients with tuberculosis and associate with responses to therapy. *Tuberculosis (Edinb)* 2017; 102:24–30.
- Bush SJ. Read trimming has minimal effect on bacterial SNP-calling accuracy. *Microb Genom* 2020; 6:mgen000434.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* 2019; 37:907–15.
- Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25:2078–9.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010; 26:841–2.
- Vandesompele J, De Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002; 3:Research0034.
- 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 Res* 2004; 64:5245–50.
- 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. *Biotechnol Lett* 2004; 26:509–15.
- Silver N, Best S, Jiang J, Thein SL. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol Biol* 2006; 7:33.
- Xie F, Xiao P, Chen D, Xu L, Zhang B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol Biol* [manuscript published online ahead of print 31 January 2012]. doi:10.1007/s11103-012-9885-2
- Hellems J, Mortier G, De Paepe A, Speleman F, Vandesompele J. Qbase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 2007; 8:R19.
- Cornejo-Granados F, López-Leal G, Mata-Espinosa DA, et al. Targeted RNA-Seq reveals the *M. tuberculosis* transcriptome from an in vivo infection model. *Biology (Basel)* 2021; 10:848.
- Gee HE, Buffa FM, Camps C, et al. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br J Cancer* 2011; 104:1168–77.
- Sinigaglia A, Peta E, Riccetti S, Venkateswaran S, Manganelli R, Barzon L. Tuberculosis-associated microRNAs: from pathogenesis to disease biomarkers. *Cells* 2020; 9:2160.
- Lu G, Jiang X, Wu A, et al. Two small extracellular vesicle sRNAs derived from *Mycobacterium tuberculosis* serve as diagnostic biomarkers for active pulmonary tuberculosis. *Front Microbiol* 2021; 12:642559.
- Hu Z, Dong J, Wang LE, et al. Serum microRNA profiling and breast cancer risk: the use of miR-484/191 as endogenous controls. *Carcinogenesis* 2012; 33:828–34.
- Schlosser K, McIntyre LA, White RJ, Stewart DJ. Customized internal reference controls for improved assessment of circulating microRNAs in disease. *PLoS One* 2015; 10:e0127443.

29. Zhu HT, Dong QZ, Wang G, et al. Identification of suitable reference genes for qRT-PCR analysis of circulating microRNAs in hepatitis B virus-infected patients. *Mol Biotechnol* **2012**; 50:49–56.
30. Wang Z, Lu Y, Zhang X, et al. Serum microRNA is a promising biomarker for osteogenesis imperfecta. *Intractable Rare Dis Res* **2012**; 1:81–5.
31. Wang X, Zhang X, Yuan J, et al. Evaluation of the performance of serum miRNAs as normalizers in microRNA studies focused on cardiovascular disease. *J Thorac Dis* **2018**; 10:2599–607.
32. Donati S, Ciuffi S, Brandi ML. Human circulating miRNAs real-time qRT-PCR-based analysis: an overview of endogenous reference genes used for data normalization. *Int J Mol Sci* **2019**; 20:4353.
33. Ignatov DV, Timoshina O, Logunova NN, Skvortsov TA, Azhikina TL. Expression of *Mycobacterium tuberculosis* small RNAs in mice models of tuberculosis [in Russian]. *Bioorg Khim* **2014**; 40:253–6.
34. Korma W, Mihret A, Tarekegn A, et al. Identification of circulating miR-22-3p and miR-93-5p as stable endogenous control in tuberculosis study. *Diagnostics (Basel)* **2020**; 10:868.
35. Kochhar P, Dwarkanath P, Ravikumar G, et al. Placental expression of RNU44, RNU48 and miR-16-5p: stability and relations with fetoplacental growth. *Eur J Clin Nutr* **2022**; 76:722–9.
36. Lasabová Z, Vazan M, Zibolenova J, Svecova I. Overexpression of miR-21 and miR-122 in preeclamptic placentas. *Neuro Endocrinol Lett* **2015**; 36:695–9.
37. Ragni E, De Luca P, Marmotti A, de Girolamo L. miR-26a-5p is a stable reference gene for miRNA studies in chondrocytes from developing human cartilage. *Cells* **2019**; 8:631.
38. Xu JF, Yang GH, Pan XH, et al. Altered microRNA expression profile in exosomes during osteogenic differentiation of human bone marrow-derived mesenchymal stem cells. *PLoS One* **2014**; 9:e114627.
39. Bockmeyer CL, Säuberlich K, Wittig J, et al. Comparison of different normalization strategies for the analysis of glomerular microRNAs in IgA nephropathy. *Sci Rep* **2016**; 6:31992.
40. Amaral FC, Torres N, Saggioro F, et al. MicroRNAs differentially expressed in ACTH-secreting pituitary tumors. *J Clin Endocrinol Metab* **2009**; 94:320–3.
41. Gordanpour A, Nam RK, Sugar L, Bacopulos S, Seth A. MicroRNA detection in prostate tumors by quantitative real-time PCR (qPCR). *J Vis Exp* **2012**:e3874.
42. Youssef YM, White NMA, Grigull J, et al. Accurate molecular classification of kidney cancer subtypes using microRNA signature. *Eur Urol* **2011**; 59:721–30.
43. Appaiah HN, Goswami CP, Mina LA, et al. Persistent upregulation of U6: SNORD44 small RNA ratio in the serum of breast cancer patients. *Breast Cancer Res* **2011**; 13:R86.
44. Arnvig K, Young D. Non-coding RNA and its potential role in *Mycobacterium tuberculosis* pathogenesis. *RNA Biol* **2012**; 9:427–36.
45. Ignatov DV, Salina EG, Fursov MV, Skvortsov TA, Azhikina TL, Kaprelyants AS. Dormant non-culturable *Mycobacterium tuberculosis* retains stable low-abundant mRNA. *BMC Genomics* **2015**; 16:954.
46. Salina EG, Grigorov A, Skvortsova Y, et al. MTS1338, a small *Mycobacterium tuberculosis* RNA, regulates transcriptional shifts consistent with bacterial adaptation for entering into dormancy and survival within host macrophages. *Front Cell Infect Microbiol*; **2019**; 9:405.
47. Moores A, Riesco AB, Schwenk S, Arnvig KB. Expression, maturation and turnover of DrrS, an unusually stable, DosR regulated small RNA in *Mycobacterium tuberculosis*. *PLoS One* **2017**; 12:e0174079.
48. Chen C, Zhu T, Wang Z, et al. High latent TB infection rate and associated risk factors in the Eastern China of low TB incidence. *PLoS One* **2015**; 10:e0141511.