Selection and Validation of Reference Genes for Pan-Cancer in Platelets Based on RNA-Sequence Data

Xiaoxia Wen†, Guishu Yang†, Yongcheng Dong†, Liping Luo†, Bangrong Cao†, Birga Anteneh Mengesha, Ruiling Zu, Yulin Liao, Chang Liu, Shi Li, Yao Deng, Kajjiong Zhang, Xin Ma, Jian Huang, Dongsheng Wang, Keyan Zhao†, Ping Leng† and Huaichao Luo†

1Chongqing Key Laboratory of Sichuan-Chongqing Co-construction for Diagnosis and Treatment of Infectious Diseases Integrated Traditional Chinese and Western Medicine, College of Medical Technology, Chengdu University of Traditional Chinese Medicine, Chengdu, China, 2Department of Clinical Laboratory, Guangyuan Central Hospital, Guangyuan, China, 3GenomCan Inc., Chengdu, China, 4Sichuan Cancer Hospital and Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China, 5Center for Informational Biology, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China, 6Department of Clinical Laboratory, Sichuan Cancer Hospital and Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

Many studies in recent years have demonstrated that some messenger RNA (mRNA) in platelets can be used as biomarkers for the diagnosis of pan-cancer. The quantitative real-time polymerase chain reaction (RT-qPCR) molecular technique is most commonly used to determine mRNA expression changes in platelets. Accurate and reliable relative RT-qPCR is highly dependent on reliable reference genes. However, there is no study to validate the reference gene in platelets for pan-cancer. Given that the expression of some commonly used reference genes is altered in certain conditions, selecting and verifying the most suitable reference gene for pan-cancer in platelets is necessary to diagnose early stage cancer. This study performed bioinformatics and functional analysis from the RNA-seq of platelets transcriptome of pan-cancer and 73 commonly known reference genes. These candidate reference genes were verified by another platelets expression data set (GSE89843). Then, we used RT-qPCR to confirm the expression levels of these seven genes in pan-cancer patients and healthy individuals. These RT-qPCR results were analyzed using the internal stability analysis software programs (the comparative Delta CT method, geNorm, NormFinder, and BestKeeper) to rank the candidate genes in the order of decreasing stability. By contrast, the GAPDH gene was stably and constitutively expressed at high levels in all the tested samples. Therefore, GAPDH was recommended as the most suitable reference gene for platelet transcript analysis. In conclusion, our result may play an essential part in establishing a molecular diagnostic platform based on the platelets to diagnose pan-cancer.

Keywords: platelets, reference genes, quantitative real time polymerase chain reaction, normalization, pan-cancer
INTRODUCTION

Platelets are derived from the megakaryocytes of the bone marrow, which are abundant in the peripheral blood (Jain et al., 2013). Platelets have long been considered to only stimulate coagulation after tissue trauma or vascular injury (Holinstat, 2017; Roweth and Battinelli, 2021). However, recent studies have shown that platelets are involved in multiple stages of cancer and are potential cancer diagnostic biomarkers (Zu et al., 2020). In the past, it was believed that the platelet content was static because platelets are cell fragments lacking a nucleus, and therefore no transcription and translation were expected (t Veld and Wurdinger, 2019) until some researchers demonstrated that platelets have the ability for protein synthesis (Warshaw et al., 1966; Burkhart et al., 2012), and the mRNA is involved in the protein synthesis reaction in platelets (Harrison and Goodall, 2008). It has been well appreciated that platelets can obtain a diverse range of mRNAs from megakaryocytes, translating into protein under external stimuli (Raslova et al., 2007). Studies have proved that tumor cells can directly stimulate platelet protein synthesis, while platelets can also sequester tumor-associated biomolecules such as proteins and RNA (t Veld and Wurdinger, 2019; Klement et al., 2009; Nilsson et al., 2016). The combination of specific splicing events in response to external signals and the ability of platelets to directly splice the circulating mRNA provides a highly dynamic transcriptome for platelets potentially suitable for liquid biopsies for cancer diagnosis (t Veld and Wurdinger, 2019; Harrison and Goodall, 2008; Best et al., 2018; Nassa et al., 2018). Given this situation, the concept of tumor-educated platelets (TEPs) has been proposed in recent years, referring to those platelets that can interact with the tumor cells and change the RNA profile (Best et al., 2015). TEP mRNAs have been confirmed to be dynamically affected by tumor conditions and may serve as biomarkers for cancer diagnosis, prognosis, prediction, or monitoring (Xue et al., 2018; Wurdinger et al., 2020).

RT-qPCR has been considered a sensitive, efficient, and reliable molecular technique to determine the mRNA levels (Xiong et al., 2018; Lin et al., 2019). Studies have proved that RT-qPCR can also amplify platelet-derived mRNA even though the concentration of mRNA is low in the whole platelets (Newman et al., 1988). Accurate and reliable relative RT-qPCR is highly dependent on reliable reference genes (Deng et al., 2020). The use of inappropriate reference genes can result in incorrect findings (Zhou et al., 2018). Therefore, the selection of reference genes depends on various species and under different experimental conditions (Coulson et al., 2008; Wang et al., 2021). However, the reference genes in the current studies of differential gene expression between the platelets and different cancers have not been uniform (Table 1). Most reference genes in platelets were found to directly use tissues’ or cells’ reference genes. Different reference genes were also used in the same cancer study. Firstly, we cannot determine whether the reference genes of cells and tissues can be applied to platelets. Furthermore, it is unclear whether the most appropriate selection of reference genes in platelets will differ due to the different cancers. Therefore, selecting and verifying the most suitable reference gene for pan-cancer in platelets is necessary.

This study is aimed to screen out the candidate genes expressed stably through the platelets’ transcript data set analysis and verify their expression stability in the platelets of pan-cancer patients by the RT-qPCR method. Then, the computer program Delta CT method (Silver et al., 2006), BestKeeper (Pfaffl et al., 2004), geNorm (Vandesompele et al., 2002), and NormFinder (Andersen et al., 2004) were used for a comprehensive analysis of the expression stability of the candidate genes. The reference gene, expressed most stably in the platelets, can be used as an internal control for the quantitative gene assay. It will promote establishing a molecular diagnostic platform based on the TEPs, to diagnose and monitor pan-cancer.

MATERIALS AND METHODS

Data Collection and Bioinformatics Analysis

We used data set GSE68086, the RNA-sequencing data of multiple malignant cancers (non–small-cell lung cancer, colorectal cancer, pancreatic cancer, glioblastoma, breast cancer, and hepatobiliary carcinomas). It is available in the public repository of the Gene Expression Omnibus (GEO) database supported by the National Center for Biotechnology Information (NCBI) (Best et al., 2015). For further downstream

| References gene | Cancer type | Sample size | PMID |
|-----------------|-------------|-------------|------|
| GAPDH | Colorectal cancer (CRC) | 286 CRC patients and 41 healthy controls and 22 patients with ulcerative colitis and 23 patients with Crohn’s disease | 31639773 |
| ACTB | Lung cancer | 48 lung cancer patients and 48 healthy donors | 31552488 |
| GAPDH | Colorectal cancer (CRC) and non–small-cell lung cancer (NSCLC) | 19 CRC patients, 16 NSCLC patients, and 4 healthy volunteers | 33955587 |
| ACTB | Non–small-cell lung cancer (NSCLC) | 243 NSCLC patients, 150 healthy controls, and 141 benign pulmonary nodules patients | 31523198 |
| ACTB and GAPDH | Non–small-cell lung cancer (NSCLC) | 10 NSCLC patients and 7 healthy subjects | 33287695 |
| ACTB | Hepatocellular carcinoma (HCC) | 20 HCC patients, 20 liver cirrhosis patients, and 10 healthy subjects | 34469466 |
| ACTB | Lung cancer | 58 healthy donors and 158 lung cancer patients | 30201066 |
analyses, the reads were quality controlled using Trimmomatic (Bolger et al., 2014), mapped to the human reference genome using STAR (Dobin et al., 2013), and intron-spanning reads were summarized using HTseq (Anders et al., 2015). The processed data include 285 samples (columns) and 57,736 ensemble gene ids (rows). Firstly, the samples that yielded less than $0.4 \times 10^6$ intron-spanning reads were excluded. Genes with a count of 0 in more than 70% of the total sample size were also deleted. Besides, genes were further excluded by following the three filtration criteria (Cheng et al., 2011; Maltseva et al., 2013; Zhan et al., 2014) for being highly and stably expressed in platelets across normal and tumor samples.

1) Mean (normal)/mean (tumor) < 1.2 and mean (tumor)/mean (normal) < 1.2. The mean of the log2CPM value of the mRNA in normal and tumor samples. We retained the positive and negative 1.2-fold genes in the normal and tumor samples.

2) Top 10% mean normal and top 10% mean tumor samples were included. We retained the first 10% of genes in the normal and tumor samples.

3) CV (coefficient of variation) (normal) < 10% and CV (tumor) < 10%. We retained genes with CV < 10% in the tumor and normal samples. CV = standard deviation (SD)/mean.

### Participants in the Validation Group

The tumor participants were included as follows: 1) patients with clinically suspected cancer were admitted to the Sichuan Cancer Hospital based on the guidelines; 2) patients without preoperative chemotherapy or radiotherapy; and 3) all the final diagnoses were based on pathology examinations. The tumor patients were excluded as follows: 1) patients with a previous history of antiplatelet medications such as aspirin; 2) pregnant patients; 3) patients with infections; and 4) patients without comprehensive clinical information. All the healthy participants were included with no disease. This study was approved by the medical ethical committee of the Sichuan Cancer Hospital (SCCHEC-02-2020-043).

### Platelet Isolation

The blood samples of all tumor participants were collected preoperatively. 1.5 ml of EDTA anticoagulated blood was added to 2 ml of EP tube. Platelet-rich plasma (PRP) was separated from the nucleated blood cells by a 20-min $120 \times g$ centrifugation step using the centrifuge (Shuke Instrument, Sichuan, China), while the platelets were separated from PRP by centrifuging at $360 \times g$ for 20 min. To minimize the impact of time, the isolation was supposed to be completed within 2 h after blood collection.

### RNA Isolation and cDNA Synthesis

Total RNA was extracted from the platelets using a TRizol reagent (Ambion, United States). The concentration and quality of the total RNA were assessed using Thermo Scientific NanoDrop 2000 Spectrophotometer (Thermo Scientific, United States). Reverse transcription was performed using a PrimeScript RT reagent kit with a gDNA eraser (TaKaRa Bio, Dalian, China) following the manufacturer’s instructions.

### Quantitative Real-Time Polymerase Chain Reaction

The primers used in the study are listed in Table 2. All the primers were designed and synthesized by Tsingke Biological Technology (Beijing, China). Quantitative real-time polymerase chain reaction (RT-qPCR) was carried out using the CFX

### TABLE 2 | Primer sequences of the seven candidate reference genes.

| Gene | Primer sequences (5’–3’) |
|------|-------------------------|
| ACTB | F:GCTATACGACCTGCTGCCTTTCT R:CTGCTTAATGTCAACCGAAGAT CTCTTTATGCGATGCTGATQG R:TTCACTTCAGGATGTCCTTT |
| GAPDH | F:ACCGAGAACAGCTGCTGATQG R:TTCACTTCAGGATGTCCTTT |
| YWHAZ | F:CCTGCAATAGGCTGTGAAGG R:GACCTACGGAAGGCTCCTAAAC |
| B2M | F:GGAGCTTACCAAGGATGCTAAG GTACCTTTCTCCTCAATTAAAC |
| GNAS | F:TCGCTGGGGAACAGTAAAGAC AGCTGCGCTGGGGAGACTG |
| OAZ1 | F:CTGACTGCTGGCTAAGAACCG R:GATCCTCCCTGACCTTTCAAC |
| PTMA | F:TCAGACGGGACCGTGAGACA R:GGATTTCCCTGTTACAGGGG |

### FIGURE 1 | The overall workflow of bioinformatical statistics for screening the candidate reference genes from the platelet RNA sequencing data set.
Then, the reference gene that we finally screened out was analyzed for stability of its expression in various cancers via the Platelet Expression Atlas website (http://bioinfo.life.hust.edu.cn/PEA/#/). This is a comprehensive platelet expression atlas (PEA) resource and platelet transcriptome landscape website which collects platelet expression data sets, including 1260 RNA-seq, 358 RNA microarray, 21 miRNA-seq, and 430 miRNA microarray data sets from 27 disease types and healthy controls from the gene expression omnibus of the National Center For Biotechnology Information (NCBI GEO) and sequence read archive (SRA) databases (Xie et al., 2022).

**Validation of Reference Gene**

The reference gene we selected was further used to verify the differential gene expression between healthy subjects and lung cancer patients through RT-qPCR experiments to better evaluate its clinical application value as a reference gene. Statistical analysis was performed with GraphPad 8.4. Student’s t-test or two-sided $\chi^2$ test was used to compare the differences in other variables among the groups. A $p$ value $< 0.05$ was considered to be statistically significant.

**RESULTS**

**Shortlisting of Reference Genes**

A total of 285 candidate genes were obtained after processing the data set GSE68086. The overall workflow of the present study is
shown in Figure 1, and the details are stated in the Materials and Methods section. We further have 95 candidate genes after optimizing the mean >1 and CV < 1 from the 285 of our pre-evaluation reference genes (Supplementary Table S1). After that, we compared 95 genes with 73 known reference genes (Radonić et al., 2004; Zhang et al., 2005; Tratwal et al., 2014; Ayakannu et al., 2015; Sharan et al., 2015; Walter et al., 2016; Panina et al., 2018; Zhao et al., 2018; Zhang et al., 2022) and finally got seven candidate genes (YWHAZ, GNAS, GAPDH, OAZ1, PTMA, B2M, and ACTB). These seven genes are known as the reference genes and are also stably expressed genes selected from the platelet data set.

**FIGURE 3** The verification of expression stability of seven candidate reference genes of platelet in another data set (GSE89843). (A) Seven candidate reference genes all expressed stably in the platelet sequencing data from 377 healthy individuals. (B) Seven candidate reference genes all expressed stably in the platelet sequencing data from 402 NSCLC patients. Blue: selected (stable expression), red: removed (unstable expression).
The distribution relationship between the candidate genes and the six tumor groups [glioblastoma (GBM), breast cancer (BrCa), pancreatic cancer (PAAD), non–small-cell lung cancer (NSCLC), hepatobiliary cancer (HBC), and colorectal cancer (CRC)] is shown in Figure 2. We then used the same bioinformatics analysis conditions (Materials and Methods section) of data set GSE68086 to analyze data set GSE89843, to verify the stability of the seven candidate genes we selected. The data set GSE89843 consists of 402 platelet samples from NSCLC patients in different stages and 377 from the healthy subjects. The result show that all the seven candidate genes that we selected also expressed stably in another platelet data set (GSE89843) (Figure 3).

### Stability Assessment of Seven Candidate Reference Genes

The baseline characteristics of all the participants are listed in Table 3. A total of 30 subjects were included in the first validation step: non–small-cell lung carcinoma (NSCLC, n = 5), colorectal cancer (CRC, n = 6), hepatobiliary cancer (HBC, n = 6), breast cancer (BrCa, n = 6), and healthy subjects (HC, n = 7). Therefore, we selected 30 subjects to perform the first validation step (Figure 4A). Then, all the measurements were calculated for all the reference genes in a specific tumor and healthy group (Supplementary Table S2). The reference gene B2M was highly stable and more expressed, scoring a mean = 25.40, median = 25.19, and SD = 1.33 (Supplementary Table S3).
We also found that B2M was more stable in hepatobiliary and breast cancers than the other genes. The expressions of GAPDH in colon cancer, PTMA in healthy control, and GNAS in non–small-cell lung cancer were also higher (Figure 4B–H). The overall results indicated that GAPDH, B2M, and ACTB were more stable in each cancer than the other four candidate genes.

Stability Assessment of GAPDH, B2M, and ACTB

Then, we selected these three reference genes GAPDH, B2M, and ACTB with higher expression stability to validate in the second step. A total of 50 subjects were included in this step, that is, 10 NSCLC, 10 CRC, 10 HBC, 10 BrCa, and 10 HC (Table 3). The mean Ct values of the three reference genes in the 50 subjects are shown in Supplementary Table S4. The web-based four algorithms were applied (Xie et al., 2012) to compare stability among the three reference genes. The Delta CT method analyses were performed to rank the genes according to the overall stability across the 50 individuals (Figure 5A). The average expression stability (M) value from the GeNorm analysis was lower than 2.7 for the most stable candidates. According to geNorm, B2M was highly expressed parallel to GAPDH (Figure 5B). The ranking of the genes in the NormFinder analysis was almost similar to the Delta CT ranking (Figure 5C). The BestKeeper algorithm calculated the correlation coefficient r, SD, and CV of the gene pairing, and the results showed that GAPDH is the most stably expressed reference gene (Figure 5D). Furthermore, the candidate reference genes were ranked in the increasing order of their stability values, and the GAPDH was the best reference gene in platelets for pan-cancer (Table 4).

Finally, according to the analysis results from the website of the Platelet Expression Atlas, the expression of GAPDH in platelets was stable in a variety of cancers when compared

![Figure 5](image-url)
with that in healthy subjects, including some uncommon cancers. GAPDH in platelets were differentially expressed only in ST elevation myocardial infarction and HIV, dengue, and H1N1 ($p < 0.05$) (Supplementary Table S5). Therefore, the results further proved that GAPDH was suitable as a reference gene in the platelets for pan-cancer.

**Validation of the Clinical Application Value for GAPDH As a Reference Gene**

The abovementioned analysis results showed that GAPDH was more suitable as a reference gene for pan-cancer platelet transcriptome quantitative analysis. To evaluate its clinical value as a reference gene more comprehensively, a new RT-qPCR experiment was designed. We selected the differential gene FLNA, which was significantly different in lung cancer patients when compared with healthy subjects by analysis via the Platelet Expression Atlas website, to verify the differential expression of the FLNA in the platelets of lung cancer and healthy subjects by using GAPDH as the reference gene.

As shown in Table 5, 42 subjects were enrolled in this step: Lung cancer (LC, n = 21) and healthy subjects (HC, n = 21). There was no statistical difference in gender between the two groups ($p > 0.05$), while the age between the groups was statistically different and the patients with lung cancer were significantly older than the healthy subjects ($p < 0.05$).

The results of RT-qPCR analysis showed that the expression of the FLNA gene in the two groups of patients was statistically significant ($p < 0.05$) (Supplementary Table S6). The expression of FLNA was significantly higher in lung cancer patients than in normal people (Figure 6), indicating that GAPDH can be used as a reference gene for RT-qPCR analysis of tumor platelets, and also had a profound clinical application value for the early diagnosis of cancer.

### DISCUSSION

Liquid biopsy technology based on blood biomarkers has developed rapidly in recent years, and various studies have shown that liquid biopsy is considered an important tool for early cancer detection (Chen and Zhao, 2019; Igatiadis et al., 2021). Platelets are highly concerned as an emerging biological source of liquid biopsy (‘t Veld and Wurdinger, 2019). TEPs mRNA has been confirmed to be dynamically influenced by tumor conditions and may be used as a biomarker for many cancer diagnoses, prognosis, prediction, and monitoring (Best et al., 2015). There have been many studies on cancer detection and monitoring through differential expression of TEPs mRNA. Yang et al. (2019) used RT-qPCR to find significantly higher TEP TIMP1 mRNA in colorectal cancer patients than in healthy individuals and in patients with ulcerative colitis and Crohn’s disease. Yao et al. (2019) found that the expression of TEP TPM3 mRNA is significantly increased in BrCa patients, by using an RT-qPCR assay. Xing et al. (2019) proved that TEP ITGA2B mRNA expression is higher in NSCLC patients than in healthy individuals and in patients with benign lung nodules, by using RT-qPCR.

RT-qPCR is a technique with high sensitivity and specificity, which is widely applied in quantifying gene expression levels (Hellemans et al., 2007). It is important to set the reliable internal controls by the reference gene in the RT-qPCR quantification assay (Bustin et al., 2009). Many studies have shown that there is no single reference gene that could be effectively used in the RT-qPCR in all species or under all experimental conditions (Suzuki

| TABLE 5 | Basic clinical characteristics of the verified subjects. |
|----------|----------|----------|
|          | Lung cancer | Healthy subjects | $p$ value |
| Total    | 21         | 21         | —         |
| Gender   | —          | —          | 0.298     |
| Male     | 14 (66.7%) | 9 (42.9%)  | —         |
| Female   | 7 (33.3%)  | 12 (57.1%) | —         |
| Age      | 61.90 (7.01) | 49.48 (8.41) | 0.001     |

Italic values represent the statistically significant difference of age between the healthy subjects and lung cancer patients.
et al., 2000; Zhou et al., 2018). For example, Brzeszczyńska et al. (2020) proved that the classical reference gene in HepaRG cells such as GAPDH was altered by drug treatment. Vorachek et al. (2013) found that the commonly used reference genes, PGK1, ACTB, and B2M for neutrophils were not reliable reference genes under different conditions. The lack of gene expression stability makes it difficult to quantify and normalize RT-qPCR data. Therefore, reference genes with systematic identification and validation are essential for solving these problems. With more and more studies using TEPs mRNA for cancer detection and monitoring, it is urgent to screen out the stable reference genes in platelets for early cancer detection.

This study identified the stable reference gene in the platelets of pan-cancer patients and normal participants. In the past, there were few studies reported on the normalization of transcript levels for platelets. Two of the seven reference genes, ACTB and GAPDH, have been reported as normalization control in mRNA detection of RT-qPCR (Hurteau et al., 2006), which have also been confirmed to be expressed in neuroendocrine lung cancer (Walter et al., 2016). Our approach is different from the earlier study, in which the mRNAs were extracted and sequenced from the platelets. In addition, an analysis of platelets in patients with myocardial infarction showed that three reference genes, HDGF, GNAS, and ACTB, were reported as the most stable reference genes (Zsóri et al., 2013), while ACTB was one of the three most stable reference genes in our study. Another study on lung cancer cell division and platelets provided a potential platelet miRNA-based treatment strategy for lung cancer. It showed the importance of internal control in the detection of miRNA expression (Liang et al., 2015). But this study did not give a more detailed description of the selection of reference genes. Our research covered a wide range of cancers and selected the most suitable reference gene for platelet transcript research.

Surprisingly, our results revealed that reference genes’ stability and expressions varied from one cancer group to another. The B2M gene was expressed higher in hepatic carcinoma and breast cancer while being more stably expressed in liver cancer, and it has not yet been reported as per the knowledge we have. The GAPDH gene was more stable in colon cancer, and the GNAS gene was highly stable in lung cancer. Despite the possibility of different stable genes in various types of tumors, the overall reference genes validation indicated that GAPDH, B2M, and ACTB were the highly stable genes in the order of first to third consecutively in all the subjects. This study recommends GAPDH as a reference gene for pan-cancer normalization, providing a standard for quantitatively detecting the gene expression levels in platelets by using this reference gene as an internal control. We also suggested that further research has to be done on this reference gene with different systematic techniques on cancer-specific normalization for internal control.

There are some limitations to the present study that can be addressed in future work. On the one hand, the sample size and the type of cancer are not enough, which may introduce errors in this type of study. On the other hand, we only selected and validated the intersection among 95 candidate reference genes in the RNA-seq data set of the TEPs and 73 known reference genes. However, it is also necessary to consider choosing more specific platelet reference genes than the currently known reference genes for validation.

CONCLUSION

In conclusion, we recommend GAPDH as the most suitable reference gene in platelets for pan-cancer normalization, providing a reference standard for quantitatively detecting the gene expression levels in platelets for the diagnosis of pan-cancer by using this reference gene as an internal control.

DATA AVAILABILITY STATEMENT

The data sets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in https://www.ncbi.nlm.nih.gov/, GSE68086, GSE89843.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Sichuan Cancer Hospital and Institute (SCCHEC-02-2020-043). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Conceptualization: XW, GY, KZ, and PL; methodology: HL and WZ; software: XM and YD; validation: XW, LL, and BC; formal analysis: XW; investigation: YL, CL, SL, RZ and KZ; resources: XW, HL, DW, JH, GY, PL, and KZ; data curation: XW; writing—original draft preparation: XW and GY; writing—review and editing: XW, GY, BAM, and PL; visualization: KZ and XW; supervision: GY and PL; funding acquisition: HL and RZ. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was supported by the Sichuan Provincial Cadre Health Research Project. (Chuan Gan Yan 2021-804), the Health and Family Planning Commission of Sichuan Province universal application project (No. 20P1115), the Science and Technology Bureau of Chengdu technological innovation project (No. 2019-YF05-01279-SN), and the Health and Family Planning Commission of Sichuan
Province universal application project (No. 19P)275). YD, XM, and KZ were supported by grants from Chengdu Science and Technology Bureau (2019-YF09-00034-CG) and Science and Technology Department of Sichuan Province (2018Y0311, 2019JDRC0049).

REFERENCES

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python Framework to Work with High-Throughput Sequencing Data. *Bioinformatics* 31, 166–169. doi:10.1093/bioinformatics/btu638

Andersen, C. L., Jensen, J. L., and Orntoft, T. F. (2004). 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.* 64, 5245–5250. doi:10.1158/0008-5472.CAN-04-0496

Ayakannu, T., Taylor, A. H., Willets, J. M., Brown, L. D. G., McDonald, J., et al. (2015). Validation of Endogenous Control Reference Genes for Normalization of Gene Expression Studies in Endometrial Carcinoma. *Mol. Hum. Reprod.* 21, 723–735. doi:10.1093/molehr/gav033

Best, M. G., Sol, N., Kooi, I., Tannous, J., Westerman, B. A., Rustenburg, F., et al. (2015). RNA-seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics. *Cancer Cell* 28, 666–676. doi:10.1016/j.ccell.2015.09.018

Best, M. G., Wesseling, P., and Wurdinger, T. (2018). Tumor-Educated Platelets as Monitoring. *J. Clin. Bioinform.* 8, e170. doi:10.1186/s13188-018-0375-y

Brzeszczyński, J., Brzeszczyńska, F., Samuel, K., Morgan, K., Morley, S. D., Plevris, J. N., et al. (2020). Validation of Reference Genes for Gene Expression Studies by RT-qPCR in HepaRG Cells during Toxicity Testing and Disease Modelling. *Cells* 9, 770. doi:10.3390/cells9030770

Burkhart, J. M., Vaudel, M., Gambaryan, S., Radau, S., Walter, U., Martens, L., et al. (2012). The First Comprehensive and Quantitative Analysis of Human Platelet Protein Composition Allows the Comparative Analysis of Structural and Functional Pathways. *Blood* 120, e73–e82. doi:10.1182/blood-2012-04-415694

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin. Chem.* 55, 611–622. doi:10.1373/clinchem.2008.112797

Chen, M., and Zhao, H. (2019). Next-generation Sequencing in Liquid Biopsy: Cancer Screening and Early Detection. *Hum. Genomics* 13, 34. doi:10.1186/s12920-019-0220-8

Chen, W.-C., Chang, C.-W., Chen, C.-R., Tsai, M.-L., Shu, W.-Y., Li, C.-Y., et al. (2011). Identification of Reference Genes across Physiological States for qRT-PCR through Microarray Meta-Analysis. *PLoS One* 6, e17347. doi:10.1371/journal.pone.0017347

Coulton, D. T., Brockbank, S., Quinn, J. G., Murphy, S., Ravid, R., Irvine, G. B., et al. (2008). Identification of Valid Reference Genes for the Normalization of RT qPCR Gene Expression Data in Human Brain Tissue. *RMC Mol. Biol.* 9, 46. doi:10.1146/1741-2199-9-46

Deng, Y., Li, Y., and Sun, H. (2020). Selection of Reference Genes for RT-qPCR Normalization in Blueberry (Vaccinium Corymbosum x Angustifolium) under Various Abiotic Stresses. *FEBS Open Bio* 10, 1418–1435. doi:10.1002/ febs.26312

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29, 15–21. doi:10.1093/bioinformatics/bts635

Harrisons, P., and Goodall, A. H. (2008). "Message in the Platelet" - More Than Just Vestigial mRNAs. *Platelets* 19, 395–404. doi:10.1080/09576450802000582

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase Relative Quantification Framework and Software for Management and Automated Analysis of Real-Time Quantitative PCR Data. *Genome Biol.* 8, R19. doi:10.1186/gb-2007-8-2-r19

Holinstat, M. (2017). Normal Platelet Function. *Cancer Metastasis Rev.* 36, 195–198. doi:10.1007/s10555-017-9677-x

Hurteau, G., Spivack, D. S., and Brock, G. J. (2006). Potential mRNA Degradation Targets of Hsa-mir-200c. *Cell Cycle* 5, 1951–1956. doi:10.4161/cc.5.17.3133

Ignatiadis, M., Sledge, G. W., and Jeffrey, S. S. (2021). Liquid Biopsy Enters the Clinic - Implementation Issues and Future Challenges. *Nat. Rev. Clin. Oncol.* 18, 297–312. doi:10.1038/s41571-020-00457-x

Jain, S., Kapetanaki, M. G., Raghavachari, N., Woodhouse, K., Yu, G., Barge, S., et al. (2013). Expression of Regulatory Platelet microRNAs in Patients with Sickle Cell Disease. *PLoS One* 8, e60932. doi:10.1371/journal.pone.0060932

Lakka Klement, G., Yip, T.-T., Cassidy, F., Kuchi, L., Cervi, D., Podust, V., et al. (2009). Platelets Actively Sequester Angiogenesis Regulators. *Blood* 113, 2835–2842. doi:10.1182/blood-2008-06-159541

Li, N., Mean, J., Yang, S., and Huang, L. (2019). Reference Gene Selection for Real-Time Quantitative PCR Normalization in Hemorrhia Compressa and Hemorrhia Altissima Leaf Tissue. *Mol. Biol. Rep.* 46, 4763–4769. doi:10.1007/s10555-017-9677-x

Maltesa, D. V., Khuastova, N. A., Fedotov, N. N., Matveeva, E. O., Lebedov, A. E., Shkurnikov, M. U., et al. (2013). High-throughput Identification of Reference Genes for Research and Clinical RT-qPCR Analysis of Breast Cancer Samples. *J. Clin. Bioinform.* 3, 13. doi:10.1186/2043-9113-3-13

Nassa, G., Giurato, G., Cimmino, G., Rizzo, F., Ravo, M., Salvati, A., et al. (2018). Splicing of Platelet Resident Pre-mRNAs upon Activation by Physiological Stimuli Results in Functionally Relevant Proteome Modifications. *Sci. Rep.* 8, 498. doi:10.1038/s41598-017-18985-5

Newman, P. J., Gorski, J., White, G. C., Gidwizis, S., Cretney, C. J., and Auster, R. H. (1988). Enzymatic Amplification of Platelet-specific Messenger RNA Using the Polyemerase Chain Reaction. *J. Clin. Invest.* 82, 739–743. doi:10.1172/JCI13656

Nilsson, R. J. A., Karachaliou, N., Berenguer, J., Gimenez-Capitan, A., Schellen, P., Teixido, C., et al. (2016). Rearranged EML4-ALK Fusion Transcripts Sequestrate Cells Invasions via Targeting Tumor Suppressor EBP41L3. *Mol. Cancer* 14, 58. doi:10.1186/s12943-015-0327-z

Lin, Y., Zhang, A., Yang, S., and Huang, L. (2019). Reference Gene Selection for Real-Time Quantitative PCR Normalization in Hemorrhia Compressa and Hemorrhia Altissima Leaf Tissue. *Mol. Biol. Rep.* 46, 4763–4769. doi:10.1007/s10555-017-9677-x

Roweth, H. G., and Battinelli, E. M. (2021). Lessons to Learn from Tumor-Educated Platelets. *Blood* 137, 3174–3180. doi:10.1182/blood.2019003976

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.913886/full#supplementary-material
Sharan, R. N., Vaiphei, S. T., Nongrum, S., Keppen, J., and Ksoo, M. (2015). Consensus Reference Gene(s) for Gene Expression Studies in Human Cancers: End of the Tunnel Visible? Cell Oncol. 38, 419–431. doi:10.1007/s11103-015-0244-6

Silver, N., Best, S., Jiang, J., and Thein, S. L. (2006). Selection of Housekeeping Genes for Quantitative RT-PCR in Human Reticulocytes Used as a Real-Time Control. BMC Mol. Biol. 7, 33. doi:10.1186/1471-2199-7-33

Suzuki, T., Higgins, P. J., and Crawford, D. R. (2000). Control Selection for RNA Quantitation. Biotechniques 29, 332–337. doi:10.2144/0292r02

Tratwal, J., Follin, B., Eklööb, A., Kastrup, J., and Haack-Sørensen, M. (2014). Identification of a Common Reference Gene Pair for qPCR in Human Mesenchymal Stromal Cells from Different Tissue Sources Treated with VEGF. BMC Mol. Biol. 15, 11. doi:10.1186/1471-2199-15-11

’t Veld, S. G. J., and Wurdinger, T. (2019). Tumor-educated Platelets. Blood 133, 2359–2364. doi:10.1182/blood-2018-12-852830

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., et al. (2002). Accurate Normalization of Real-Time Quantitative RT-PCR Data by Geometric Averaging of Multiple Internal Control Genes. Genome Biol. 3 (1), research0034. doi:10.1186/gb-2002-3-7-research0034

Vorachek, W., HugiJetuBoBe, G., Bobe, G., and Hall, J. (2013). Reference Gene Selection for Quantitative PCR Studies in Sheep Neutrophils. IJMS 14, 11484–11495. doi:10.3390/ijms140611484

Walter, R. F. H., Werner, R., Vollbrecht, C., Hager, T., Flom, E., Christoph, D. C., et al. (2016). ACTB, CDKN1B, GAPDH, GRB2, RHOA and SDCBP Were Identified as Reference Genes in Neuroendocrine Lung Cancer via the nCounter Technology. PLoS One 11, e0165181. doi:10.1371/journal.pone.0165181

Wang, M., Ren, T., Marowa, P., Du, H., and Xu, Z. (2021). Identification and Selection of Reference Genes for Gene Expression Analysis by Quantitative Real-Time PCR in Suaceda Glaucu’s Response to Salinity. Sci. Rep. 11, 8569. doi:10.1038/s41598-021-88151-5

Warshaw, A. L., Laster, L., and Shulman, N. R. (1966). The Stimulation by Thrombin of Glucose Oxidation in Human Platelets. J. Clin. Invest. 45, 1923–1934. doi:10.1122/jc.1965.4072

Wurdinger, T., ’t Veld, S. G. J., and In ’t Veld, M. G. (2020). Platelet RNA as Pan-Tumor Biomarker for Cancer Detection. Cancer Res. 80, 1371–1373. doi:10.1158/0008-5472.CAN-19-3684

Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. (2012). miRDeepFinder: a miRNA Identification Tool for Deep Sequencing of Plant Small RNAs. Plant Mol. Biol. 80, 75–84. doi:10.1007/s11103-012-9885-2

Xie, G. Y., Liu, C. J., Miao, Y. R., Xia, M., Zhang, Q., and Guo, A. Y. (2022). A Comprehensive Platelet Expression Atlas (PEA) Resource and Platelet Transcriptome Landscape. Am. J Hematol. 97. doi:10.1002/ajh.26393

Xing, S., Zeng, T., Xue, N., He, Y., Lai, Y.-Z., Li, H.-L., et al. (2019). Development and Validation of Tumor-Educated Blood Platelets Integrin Alpha 2b (ITGA2B) RNA for Diagnosis and Prognosis of Non-small-cell Lung Cancer through RNA-Seq. Int. J. Biol. Sci. 15, 1977–1992. doi:10.7150/ijbs.36284

Xiong, D.-D., Dang, Y.-W., Lin, P., Wen, D.-Y., He, R.-Q., Luo, D.-Z., et al. (2018). A circRNA-miRNA-mRNA Network Identification for Exploring Underlying Pathogenesis and Therapy Strategy of Hepatocellular Carcinoma. J. Transl. Med. 16, 220. doi:10.1186/s12967-018-1593-5

Xue, L., Xie, L., Song, X., and Song, X. (2018). Identification of Potential Tumor-Educated Platelets RNA Biomarkers in Non-small-cell Lung Cancer by Integrated Bioinformatical Analysis. J. Clin. Lab. Anal. 32, e22450. doi:10.1002/jcla.22450

Yang, L., Jiang, Q., Li, D.-Z., Zhou, X., Yu, D.-S., and Zhong, J. (2019). TIMP1 mRNA in Tumor-Educated Platelets Is Diagnostic Biomarker for Colorectal Cancer. Aging 11, 8998–9012. doi:10.18632/aging.102366

Yao, B., Qu, S., Hu, R., Gao, W., Jin, S., Ju, J., et al. (2019). Delivery of Platelet TPM3 mRNA into Breast Cancer Cells via Microvesicles Enhances Metastasis. FEBS Open Bio 9, 2159–2169. doi:10.1111/2211-5463.12759

Zhan, C., Zhang, Y., Ma, J., Wang, L., Jiang, W., Shi, Y., et al. (2014). Identification of Reference Genes for qRT-PCR in Human Lung Squamous-Cell Carcinoma by RNA-Seq. Acta Biochim. Biophys. Sin. (Shanghai) 46, 330–337. doi:10.1093/abbs/gmt153

Zhang, C., Wang, Y., Jin, G., Wu, S., Cui, J., and Wang, R.-F. (2022). (Corrigendum) Selection of Reference Genes for Gene Expression Studies in Human Bladder Cancer Using SYBR-Green Quantitative Polymerase Chain Reaction. Oncol. Lett. 23, 85. doi:10.3892/ol.2022.13205

Zhang, X., Ding, L., and Sandford, A. J. (2005). Selection of Reference Genes for Gene Expression Studies in Human Neutrophils by Real-Time PCR. BMC Mol. Biol. 6, 4. doi:10.1186/1471-2199-6-4

Zhao, H., Ma, T.-F., Lin, J., Liu, L.-L., Sun, W.-J., Guo, L.-X., et al. (2018). Identification of Valid Reference Genes for mRNA and microRNA Normalisation in Prostate Cancer Cell Lines. Sci. Rep. 8, 1949. doi:10.1038/s41598-018-19458-z

Zhou, L., Chen, F., Ye, J., and Pan, H. (2018). Selection of Reliable Reference Genes for RT-qPCR Analysis of Bursaphelenchus mucronatus Gene Expression from Different Habitats and Developmental Stages. Front. Genet. 9, 269. doi:10.3389/fgene.2018.00269

Zsöri, K., Muszbek, L., Cai, Z., and Shekimari, A. (2013). Validation of Reference Genes for the Determination of Platelet Transcript Level in Healthy Individuals and in Patients with the History of Myocardial Infarction. Ijms 14, 3456–3466. doi:10.3390/ijms14023456

Zu, R., Yu, S., Yang, G., Ge, Y., Wang, D., Zhang, L., et al. (2020). Integration of Platelet Features in Blood and Platelet Rich Plasma for Detection of Lung Cancer. Clin. Chim. Acta 509, 43–51. doi:10.1016/j.cca.2020.05.043

Conflict of Interest: Authors YD, XM, and KZ were employed by the company GenomCan Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.