Selection and Validation of Reference Genes for Gene Expression Studies in *Codonopsis pilosula* Based on Transcriptome Sequence Data

Lijun Liang¹,²,³, Zhigui He⁴,⁵, Haizheng Yu¹,³, Erhuan Wang⁶, Xiaojiang Zhang⁶, Bingxue Zhang⁴,³, Chenlu Zhang⁷ & Zongsuo Liang¹,⁴,⁸*

Relative gene expression analyses by RT-qPCR (reverse transcription-quantitative PCR) are highly dependent on the reference genes in normalizing the expression data of target genes. Therefore, inappropriate endogenous control genes will lead to inaccurate target gene expression profiles, and the selection and validation of suitable internal reference genes becomes essential. In this study, we retrieved the commonly used reference genes in transcriptome datasets of *Codonopsis pilosula* by RNA-Seq (unpublished data), and selected 15 candidate reference genes according to the coefficient of variation (CV) and fold change (FC) value of gene expression. The expression levels of candidate reference genes, which is at different growth stages, undergoing cold stress and drought stress, was determined by RT-qPCR. The expression stability of these genes was evaluated using software packages and algorithms including ΔCt, geNorm, NormFinder and Bestkeeper. Then appropriate reference genes were screened and validated by target gene-UDGPase (UDP glucose pyrophosphorylase). The optimal RGs combinations of *C. pilosula*, including PP2A59γ, CPY20-1, UBCE32, RPL5B and UBC18 for developmental stage, RPL5B, RPL13 and PP2A59γ for cold treatment, RPL13 and PP2A59γ for drought treatment, were found and proposed as reference genes for future work. This paper laid foundations for both the selection of reference genes and exploration in metabolic mechanism of *C. pilosula*.

*Codonopsis pilosula* (Franch.) Nannf is a traditional Chinese medicinal and edible plant¹.² As a history of more than 300 years, *C. pilosula* is mainly distributed in four provinces of China including Gansu, Shanxi, Shanxi and Hubei³.⁴. It is reported that the main ingredients related to its pharmacodynamics are polysaccharide, tangshenosides, lobetynol, atractylonelolide, and so on¹. By means of transcriptome sequencing, Gao J. P. et al. and Li J. et al. revealed that *UDGPase* (UDP glucose pyrophosphorylase) and *UGDH* (UDP-glucose dehydrogenase) were the key enzyme genes in *C. pilosula* controlling the synthesis pathway of polysaccharide⁵–⁸. Consequently, *UDGPase* was cloned and characterized by them adopting GAPDH as reference gene (RG)⁹. *GAPDH* and *SUC4* (sucrose/H⁺ cotransporter 4) were also cloned and characterized by Wang X. L. et al. and Zheng Q. H. et al.⁵,⁶. In the above studies, Real-Time Quantitative PCR (RT-qPCR) was commonly used in expression analysis of those genes. Compared with traditional quantitative analysis techniques like Northern blots or cDNA microarrays, RT-qPCR is much more popular in gene expression because of its time-saving, accuracy, low cost and high throughput characteristics⁵–⁸. RT-qPCR has played an important role in researching the molecular regulation

¹Institute of soil and water conservation, CAS & MWR, Yangling, 712100, P. R. China. ²College of Landscape Architecture, Zhejiang Agriculture & Forestry University, Hangzhou, 311300, P. R. China. ³University of Chinese Academy of Sciences, Beijing, 100049, P. R. China. ⁴College of Life Sciences, Northwest A&F University, Yangling, 712100, P. R. China. ⁵Institute of Food Science and Biological Engineering, Guilin Tourism University, Guilin, 541006, P. R. China. ⁶Buchang Pharmaceuticals Co., Ltd, Xi’an, 712000, P. R. China. ⁷School of Biological Science and Engineering, Shaanxi University of Technology, Hanzhong, 723000, P. R. China. ⁸College of Life Sciences, Key Laboratory of Plant Secondary Metabolism and Regulation of Zhejiang Province, Zhejiang Sci-Tech University, Hangzhou, 310018, P. R. China. *email: liangzs@ms.iswc.ac.cn
of secondary metabolites of medicinal plants. Wu Y. N. et al. and Gao J. P. et al. analyzed several gene expression of C. pilosula, which laid foundations for the gene regulation. Quantitative method is crucial to the accuracy of gene expression in RT-qPCR, which is divided into absolute quantification and relative quantification methods.\(^\text{10-12,14,15,18-20}\). RGs, which expressed in all cells of plants as housekeeping genes, mainly include ACT (actin), GAPDH (glyceraldehyde-3-phosphate dehydrogenase), G6PDH (glucose-6-phosphate dehydrogenase), EF (elongation factor), 18 S rRNA (18 S ribosomal RNA), UBP (ubiquitin), TUB (tubulin), EIF (eukaryotic initiation factor), UBC (ubiquitin conjugating enzyme), H3 (histone), PGK (phosphoglycerate kinase), TEF (translation elongation factor), CYP (Cyclophilin), NADHD (Nicotinamide adenine dinucleotide dehydrogenase), and so on.\(^\text{19,21-27}\). Recent studies have shown that gene expression was highly tissue-specific and often varied with the physiological status of the plant or experimental conditions, and the universal reference genes have not been found at present.\(^\text{19,21,24,28-20}\). Selection of RGs should be conducted amongst the same pool of samples, and carried out according to the specific tissue, developmental stages and experimental conditions, thus the selected RGs can be used for target gene expression.\(^\text{25,31-35}\). In Gao J. Ps study on the polysaccharide of C. pilosula, GAPDH was used as the internal reference gene to normalize the expression of UDGPase. This work started the exploration of synthesis mechanism for C. pilosula metabolites, and GAPDH became the first internal reference gene used in the study of molecular regulation mechanisms. In order to make clear the mechanisms, systematic selection and validation of RGs are required. Reliable internal reference genes were particularly essential for C. pilosula during the growth stages and above all under the drought and cold stress conditions, because C. pilosula is mainly distributed in dry and cold areas with frequently extreme weather such as cold and drought. The impact of extreme climates such as cold and drought on the reliable of RG was extensively reported in many plants other than C. pilosula.\(^\text{21,24,31,32}\)

In this study, we retrieved the commonly used RGs of C. pilosula in transcriptome datasets (unpublished data), and selected 15 candidate RGs according to threshold method of coefficient of variation (CV) and fold change (FC) value in gene expression.\(^\text{36,37}\). RT-qPCR was then adopted to determine the candidate RGs at developmental stages, under cold stress treatments and drought stress treatments. The corresponding stability was evaluated using software packages and algorithms including ΔCt,\(^\text{38}\) geNorm,\(^\text{39}\) NormFinder,\(^\text{40}\) and Bestkeeper.\(^\text{41}\) Based on the above analysis, the appropriate RGs were screened and validated by target gene-UDGPase (UDP glucose pyrophosphorylase). This research for the first time systematically explored the RGs of C. pilosula to lay foundations for the selection of RGs and analysis of gene expression.

**Results**

**Selection of candidate reference genes based on transcriptome data in C. pilosula.** Root of C. pilosula at the initial and late stages of its swelling were collected, and three tissue layers including periderm, phloem and xylem of the roots were separated as samples for transcriptome sequencing. Six replicates for each tissue. The samples were sent to Lianchuan biological co. LTD to perform the transcriptome sequencing by using paired-end sequencing technology on an Illumina Hi-Seq™ 4000 platform. After assembling and annotation, expression profile data was mapped to the transcriptome. TPM (Transcripts per kilobase of exon model per million mapped reads) was calculated by means of EdgeR package (The RSEM package), which was then used to analyze gene expression level.

According to literatures on the conventional RGs, the related genes including ACT, GAPDH, G6PDH, EF, 18 S rRNA, UBP, TUB, EIF, UBC, H3, PGK, TEF, CYP and NADHD were retrieved from the transcriptomic data of C. pilosula. As is reported, low-expression genes would make poor RT-qPCR references due to the difficulties in detecting and quantifying their expression, thus, genes with expression levels lower than five TPM were excluded after the stability analyses. In addition, genes with log fold variance more than two in different tissues were also rejected. After the removals, a total of 207 genes were selected for further study. Calculations for mean value (MV), SD, CV and log2FC were executed in Microsoft Excel according to TPM. Genes with low CV and log2FC also rejected. After the removals, a total of 207 genes were selected for further study. Calculations for mean value detecting and quantifying their expression, thus, genes with expression levels lower than five TPM were excluded.

The expression level of candidate RGs was normally determined by the threshold cycle (Ct) values of RT-qPCR. Small Ct value indicated higher gene expression abundance, and vice versa. Figure 1 displayed the Ct values of RGs and their variability. The genes with lower variability were UPL-RHF2A, UBCE32, UBC18, ABCCR, and RPL5B according to the distribution of raw Ct values, thus could be selected as the RGs. The commonly used ACT, EIF1α, and TUBA3 were the worst candidate RGs because the Ct values spanned multiple units according to the scope of the interquartile. However, analyses on the distribution of the raw Ct values and the range of the interquartile are not sufficient to assess the expression stability of candidate RGs, and other methods are required for more comprehensive and accurate evaluations.

**Expression stability analysis of candidate reference genes of C. pilosula.** The expression stability of fifteen candidate RGs of leaf, stem, and root of C. pilosula under the conditions of different developmental stage,
cold stress, drought stress, as well as the total samples were analyzed by means of ΔCt, geNorm, NormFinder, and BestKeeper methods.

**ΔCt analysis.** The mean SD of ΔCt between one gene and the other 14 genes were calculated, which was used to evaluate the stability of the gene. The lower the mean SD was, the higher the stability of the gene tended to. Table 2 listed the rank of RGs and their expression stability values at different developmental stages and treatments. **PP2A59γ and RPL5B** were the most stable RGs at the different developmental stages. **RPL5B** was the most stable RG in cold stress treatment, and **PP2A59γ** was most stable in drought stress treatment. **UBCE32** was found to be the most stable RG in the total samples, while **ACTβ** was the most unstable RG both at the different developmental stages and under the cold stress or drought stress treatments.

**geNorm analysis.** The M values of fifteen candidate RGs were calculated by using geNorm program and the stability of each candidate RG was ranked by the M value. Genes with the lowest M values are considered to have the most stable expression and denoted as the suitable RG. The samples were divided into four groups: developmental stage (ck, 15d and 30d), cold stress (ck, 4 h, 8 h and 12 h), drought stress (ck, 5d and 10d), and total (all the samples). The average expression stability values (M) and ranking of the candidate RGs was shown in Fig. 2. The optimal RGs were found to vary with the experimental conditions. **CPY20-1** and **UBC18** were the most credible RGs for samples collected at different developmental stages. **RPL5B** and **RPL13** were the best RGs for samples experiencing cold stress treatment. In drought stress treatment, **CPY20-1** and **RPL13** were the most suitable RGs. **CPY20-1** and **RPL5B** were the most stable RGs for all the samples.

To determine the optimal number of RGs for normalization of gene expression level, the pairwise variation ($V_{n+1}$) of two sequential normalization factors (NFs) was computed using geNorm program. If the pairwise variation between genes is larger than or equal to 0.15, one gene should be added for reliable normalization of the pairwise variation. Once the pairwise variation is below 0.15, no additional genes are required for normalization.

Table 1. Primer sequences and amplicon characteristics of 15 candidate RGs and validation gene for RT-qPCR analysis. *gene known to have different levels of expression in C. pilosula was used to validate the candidate RGs.

| Gene symbol | Gene description | Primer sequences (5′→3′) | Amplicon Size(bp) | TM (°C) | Efficiency (%) |
|-------------|------------------|---------------------------|-------------------|---------|----------------|
| CYP20-1     | Cytochrome P450 20-1 | GGACCAGACACCAATGGTTCAC  AAGGGTAGCTCTCCGCTGTC | 188 | 58.0  | 91.5 |
| ABC8        | ABC-transporter C family member 8 | AGGCCCTAACGGGTACAACCAAG | 136 | 58.0  | 96.2 |
| ACTβ        | Beta-Actin       | AGAGAAAGGCGCTGAATGACCA  AGGCTTGGGAGAAACCTTA | 103 | 58.0  | 102.1 |
| UPL-RHF2A   | E3 ubiquitin–protein ligase RHF2A | GAGGGAGGGCTCAAGGGAGCTG  AGCTCGAAATGGGACAGACG | 172 | 58.0  | 92.5 |
| EIF1α       | Elongation factor 1-alpha | GCTCGGTGACAAAGTGGAT  CGGAGGTGTGGCAAATCAAGA | 191 | 58.0  | 107.2 |
| EIF4A-14    | Eukaryotic initiation factor 4A | GAGCCCTAACGGGTACCCAAG  AGCTGAATGGGGAGGGCTTC | 187 | 58.0  | 98.1 |
| RPL5B       | 60S ribosomal protein L5 | CACGTGCTATGCGCAATGAA  CCGCCCGCTCAACAGAGAAA | 180 | 58.0  | 95.0 |
| RPL13       | 60S ribosomal protein L13 | AGGAACTTGTCAACAGCCACA  CCGCAGCCCTTTAAGGCCCTT | 176 | 58.0  | 93.7 |
| G6PDH       | glucose-6-phosphate dehydrogenase | TGTTGTGACGCTAGTGCAGAGTC  AGCATCAACGGTGTCCTCACAGA | 187 | 58.0  | 91.3 |
| GAPDH       | Glyceraldehyde-3-phosphate dehydrogenase | CCAAGAAGGGCAGGCGATGTT  CCTTACGGCGCTGACCTT | 150 | 58.0  | 94.1 |
| TUBα3       | Alpha-tubulin | TGACTTGGTCGCTACTGTT  CCGCCCTTCTTGGCTACATGGA | 178 | 58.0  | 106.4 |
| PP2A579γ    | Protein phosphatase 2A-2A-57 | CTTGCAGCTGGAGAAACCGGG | 183 | 58.0  | 102.5 |
| PP2A59γ     | Protein phosphatase 2A-2A-59 | GCGGACACTAATGGAACAAAG  CAGGGGCTCCAGGTTGATT | 166 | 58.0  | 95.6 |
| UBC18       | Ubiquitin-conjugating enzyme 18 | CCAGCTCTCTACACCCCTCA  TCTCTTGGTCTTGCTTGCC | 152 | 58.0  | 104.7 |
| UBCE32      | Ubiquitin-conjugating enzyme 32 | TGTTGACGCCGGAACACGGATG  CCGTCAAAGCTGGAGGGTGGG | 175 | 58.0  | 91.8 |
| UGPase*     | UDP-glucose pyrophosphorylase | TGATGGCTATGTGACCCGGAA  ACCCTTGAGAACGACGAAGA | 187 | 58.0  | 102.1 |

* denote gene known to have different levels of expression in *C. pilosula* was used to validate the candidate RGs.
Analyzation on the pairwise variation of candidate RGs was displayed in Fig. 3. For both the total samples and samples developed at different stages, V5/6 was below 0.15, accordingly, five RGs was the optimal number. For samples subjected to cold stress treatment, V3/4 was lower than 0.15, thus 3 RGs are required, while for samples subjected to drought stress treatment, 2 RGs are needed. Results of pairwise variation was consistent with that of the stability measurements. RGs of samples subjected to cold stress or drought stress treatment demonstrated lower average pairwise variation and higher stability. In contrast, RGs of both the total samples and samples at various growth stage behaved unstable and showed higher pairwise variation.

**NormFinder analysis.** NormFinder software was also applied to analyze the pairwise variation of candidate RGs. Expression stability values of genes were shown in Table 3. Genes with lower SV was considered to be stable,
and suitable for RGs. Stable genes in samples at different developmental stages were PP2A59γ, UBC32 and RPL5B, RPL13 and RPL5B were the most suitable RGs in cold stress treatment. PP2A59γ was the most stable RG in drought stress treatment. PP2A59γ, RPL13, RPL5B, and UBC32 were the most suitable RGs for total samples. However, ACT/3 was the most unstable gene in all of the samples, which agreed with the results from geNorm analysis.

**BestKeeper analysis.** BestKeeper analysis determined stable RG according to the Ct values of each gene. Genes with high coefficient of determination (r), low SD and CV are considered stable. According to SD, CV, and r of Ct value, the expression stability of candidate RGs was ranked corresponding to the different developmental stages, cold and drought stress treatment, as well as the total samples. Then the geomean value of each RG was calculated and used to rank the fifteen RGs (Table 4). The results indicated that the optimal RG was ABC5C8 for developmental stages, UBC18 for cold treatment, and RPL13 for both drought treatment and the total samples. The most unstable RG was ACT/3 except for developmental stage, in which the most unreliable RG was UPL-RHF2A.

**Comprehensive stability analysis of reference genes.** Based on the ranking results of expression stability by means of the four methods (ΔCt, geNorm, NormFinder, and BestKeeper), the geomean value of each RG was calculated and applied to determine the comprehensive rank of the fifteen RGs in developmental stage, cold treatment, drought treatment, and the total samples. Results in Table 5 showed that ACT/3 was the most unstable RG (single low-ranked RG). The top-ranked single RG is PP2A59γ for developmental stage, RPL5B for cold treatment, and RPL13 for both drought treatment and the total samples. According to the number of RGs suggested by geNorm and the ranking listed in Table 5, the best combination of RGs was PP2A59γ, CPY20-1, UBC32, RPL5B, and UBC18 for developmental stage, RPL5B, RPL13, and PP2A59γ for cold treatment, RPL13 and PP2A59γ for drought treatment, and RPL13, UBC32, RPL5B, CPY20-1 and PP2A59γ for the total samples.

**Validation on the stability of reference genes.** The validity of RG or RGs, denoted as relative quality in Figs. 4 to 6, was verified by normalizing the target genes UDGfase (UDP glucose pyrophosphorylase) with the selected single RG, the combination of RGs and ACT/3. For the developmental stage of *C. pilosula*, when PP2A59γ alone was used as target gene for DGfase expression, relative expression of UDGfase of leaves increased firstly and then decreased, and the gene UDGfase was found to be overexpressed nearly 29 times on the 15th day (Fig. 4A). Relative expression of UDGfase in both the stems and roots decreased and then increased (Fig. 4B,C), reaching the highest on the 30th day for the stems. In comparison, when normalization was performed using the ACT/3 gene, serious deviation produced in all tissues (leaf, stem, and root), and the gene UDGfase was found to be overexpressed nearly 58 times on the 15th day (Fig. 4G–I). Using the combination of RGs including PP2A59γ, CPY20-1, UBC32, RPL5B and UBC18, the expression characteristics of UDGfase in leaves, stems and roots was basically similar to that using PP2A59γ alone. However, the relative expression levels of UDGfase at different time points in different tissues were slightly lower than those using the most stable single RG (Fig. 4D–F).

For the cold treatment, as shown in Fig. 5, the variation characteristics of relative expression of UDGfase are similar in three organs whether using the most stable RPL5B or combination of RPL5B, RPL13 and PP2A59γ. The relative expression levels of UDGfase using single stable RG were slightly higher than those using combination of RGs except for the control. Application of unstable ACT/3 gene for normalizing target gene in root led to larger deviations. In the drought treatment, Both the most stable RPL13 and combination of RPL13 and PP2A59γ were the effective RGs for the relative expression of UDGfase of the three tissues. In contrast, using the most unstable

---

**Table 2.** The rank of RGs for normalization and their expression stability values calculated by the ΔCt method.

| Rank position | Developmental stage | Cold | Drought | Total |
|---------------|---------------------|------|---------|-------|
| Gene          | Average of STDEV    | Gene | Average of STDEV | Gene | Average of STDEV |
| 1             | PP2A59γ             | 1.354 | RPL5B    | 1.015 | PP2A59γ         | 0.812 | UBC32     | 1.185 |
| 2             | RPL5B               | 1.354 | PP2A59γ  | 1.024 | CPY20-1         | 0.857 | RPL5B     | 1.223 |
| 3             | UBC32               | 1.357 | UBC32    | 1.055 | UBC32           | 0.858 | PP2A59γ   | 1.253 |
| 4             | UBC18               | 1.426 | RPL13    | 1.064 | RPL13           | 0.883 | RPL13     | 1.254 |
| 5             | G6PDH               | 1.446 | EFlα     | 1.081 | UBC18           | 0.928 | EFlα      | 1.282 |
| 6             | CPY20-1             | 1.464 | PP2A57γ  | 1.129 | UPL-RHF2A       | 0.931 | CPY20-1   | 1.298 |
| 7             | EFlα                | 1.495 | G6PDH    | 1.172 | RPL5B           | 0.978 | PP2A57γ   | 1.308 |
| 8             | PP2A57γ             | 1.499 | CPY20-1  | 1.177 | PP2A57γ         | 0.990 | G6PDH     | 1.353 |
| 9             | RPL13               | 1.523 | EFlα     | 1.294 | EFlα           | 1.020 | UBC18     | 1.386 |
| 10            | TUBα3               | 1.570 | UPL-RHF2A| 1.299 | TUBα3          | 1.060 | ABC5C8    | 1.464 |
| 11            | GAPDH               | 1.658 | UBC18    | 1.379 | EFlα           | 1.069 | UPL-RHF2A| 1.549 |
| 12            | ABC5C8              | 1.851 | TUBα3    | 1.379 | G6PDH           | 1.206 | GAPDH     | 1.579 |
| 13            | UPL-RHF2A           | 2.077 | GAPDH    | 1.519 | TUBα3          | 1.221 | ABC5C8    | 1.692 |
| 14            | EFlα                | 2.401 | ABC5C8   | 1.815 | G6PDH          | 1.364 | EFlα      | 2.179 |
| 15            | ACT/3               | 2.429 | ACT/3    | 2.977 | ACT/3          | 2.864 | ACT/3     | 2.996 |
ACTB gene for normalization, although the expression of UDGPase in the roots was similar to that of the most stable RG or RGs, the expression bias of UDGPase appeared in the leaves and stems (Fig. 6).

Discussion

Although C. pilosula is a traditional Chinese medicine, insufficient researches have been conducted on the internal reference genes for gene expression normalization in biosynthesis of its main components. As a result, the synthesis mechanism is not clear by now. Fortunately, Gao, J. P. et al. initiated the study of the internal reference gene of C. pilosula. In their work, GAPDH was selected as RG to study the expression characteristics of UDGPase gene in polysaccharide synthesis. In our study, more stable RG and RGs combinations than GAPDH were found. Based on RT-qPCR study of C. pilosula, systematic selection and validation analysis of reliable RGs for gene
algorithms. Taking full consideration of the ranks obtained from the four methods, the geometric means of the stability of candidate RGs. The rankings derived from the four methods were slightly different due to the changed conditions of RGs were selected. Surprisingly, some RGs commonly used in other plants, such as PP2A57, were retrieved from the transcriptome data of *C. pilosula*, and then screened according to the CV, log2FC values of the expression level of each gene in all samples.

RT-qPCR was used to determine the expression of the candidate RGs under the conditions of cold stress, drought stress and developmental stages, and finally the stable RGs were selected. This study confirmed that the single optimal RG is PP2A59, for developmental stage, RPL5B for cold treatment, and RPL13 for drought treatment. Single RG can be used for gene expression analysis, as is used in most studies. However, disturbing factors such as extremely cold and drought climates during the period of growth probably

Table 3. The rank of RGs for normalization and their expression stability values calculated by the NormFinder program.

| Rank position | Developmental stage | Cold | Drought | Total |
|---------------|---------------------|------|---------|-------|
| Gene | Geomean | Gene | Geomean | Gene | Geomean | Gene | Geomean | Gene | Geomean |
| 1 | ABCCC8 | 2.466 | UBC18 | 3.302 | RPL13 | 2.000 | RPL13 | 2.154 |
| 2 | GAPDH | 3.634 | ABCCC8 | 3.557 | CPY20-1 | 2.466 | CPY20-1 | 3.000 |
| 3 | RPL13 | 4.579 | GAPDH | 3.733 | UBC32 | 3.420 | UBC32 | 3.420 |
| 4 | CPY20-1 | 4.718 | RPL5B | 3.826 | PP2A59γ | 3.826 | PP2A59γ | 4.041 |
| 5 | G6PDH | 4.762 | PP2A59γ | 4.642 | UBC18 | 4.160 | RPL5B | 4.932 |
| 6 | EF1α | 4.946 | UBC32 | 5.192 | ABCCC8 | 6.782 | EF1α | 4.946 |
| 7 | UBC18 | 5.848 | EF1α | 6.415 | RPL5B | 6.910 | PP2A59γ | 5.518 |
| 8 | RPL5B | 6.316 | RPL13 | 7.114 | G6PDH | 8.173 | ABCCC8 | 6.952 |
| 9 | UBC32 | 6.542 | UPL-RHF2A | 7.343 | PP2A59γ | 8.243 | PP2A59γ | 7.319 |
| 10 | PP2A59γ | 6.604 | EF1α | 7.764 | UPL-RHF2A | 8.573 | UPL-RHF2A | 9.524 |
| 11 | EF1α | 7.958 | PP2A59γ | 8.143 | EF1α | 10.288 | UPL-RHF2A | 10.260 |
| 12 | PP2A59γ | 9.655 | CPY20-1 | 8.243 | EF1α | 10.298 | GADPH | 10.537 |
| 13 | ACT3 | 12.897 | G6PDH | 10.459 | TUBα3 | 10.483 | G6PDH | 11.334 |
| 14 | TUBα3 | 12.974 | TUBα3 | 12.083 | GADPH | 10.801 | EF1α | 12.209 |
| 15 | UPL-RHF2A | 13.976 | ACT3 | 14.659 | ACT3 | 15.000 | ACT3 | 15.000 |

Table 4. The rank of RGs for normalization calculated by the BestKeeper program.

| Rank position | Developmental stage | Cold | Drought | Total |
|---------------|---------------------|------|---------|-------|
| Gene | Geomean | Gene | Geomean | Gene | Geomean | Gene | Geomean |
| 1 | PP2A59γ | 2.065 | RPL13 | 0.075 | PP2A59γ | 0.084 | PP2A59γ | 0.145 |
| 2 | UBC32 | 0.928 | RPL5B | 0.095 | RPL13 | 0.252 | RPL13 | 0.157 |
| 3 | RPL5B | 0.243 | PP2A59γ | 0.157 | UBC32 | 0.328 | UBC32 | 0.160 |
| 4 | PP2A59γ | 0.291 | EF1α | 0.174 | CPY20-1 | 0.334 | RPL5B | 0.179 |
| 5 | RPL13 | 0.305 | UBC32 | 0.184 | UPL-RHF2A | 0.408 | CPY20-1 | 0.208 |
| 6 | CPY20-1 | 0.313 | CPY20-1 | 0.187 | PP2A59γ | 0.443 | PP2A59γ | 0.222 |
| 7 | G6PDH | 0.313 | PP2A59γ | 0.219 | RPL5B | 0.444 | EF1α | 0.236 |
| 8 | EF1α | 0.352 | G6PDH | 0.268 | UBC18 | 0.460 | G6PDH | 0.311 |
| 9 | TUBα3 | 0.356 | EF1α | 0.281 | EF1α | 0.462 | EF1α | 0.320 |
| 10 | UBC18 | 0.367 | UPL-RHF2A | 0.333 | EF1α | 0.481 | UBC18 | 0.385 |
| 11 | EF1α | 0.494 | TUBα3 | 0.437 | ABCCC8 | 0.573 | TUBα3 | 0.393 |
| 12 | GADPH | 0.528 | UBC18 | 0.496 | TUBα3 | 0.636 | UPL-RHF2A | 0.397 |
| 13 | ABCCC8 | 0.634 | GADPH | 0.521 | G6PDH | 0.689 | GADPH | 0.512 |
| 14 | UPL-RHF2A | 0.730 | ABCCC8 | 0.647 | GADPH | 0.819 | ABCCC8 | 0.542 |
| 15 | ACT3 | 0.903 | ACT3 | 1.276 | ACT3 | 1.669 | ACT3 | 1.167 |

expression normalization were carried out. More importantly, candidate RGs, used in the selection of stable RGs, were retrieved from the transcriptome data of *C. pilosula*, and then screened according to the CV, log2FC values of the expression level of each gene in all samples.

Some candidate RGs that ranked at the top of the RNA-Seq analysis were not always stable in RT-qPCR test. This study also showed that the single optimal RG is PP2A59γ for developmental stage, RPL5B for cold treatment, and RPL13 for drought treatment. Single RG can be used for gene expression analysis, as is used in most studies.
leaded to deviation in gene expression by using single RG. Therefore, two or more RGs are recommended for reliable gene expression\(^4^\). Suitable combinations of RGs were obtained as follows. \(PP2A59\gamma, CPY20-1, UBCE32, RPL5B, \) and \(UBC18\) were the optimal combination RGs for developmental stage, \(RPL5B, RPL13, \) and \(PP2A59\gamma\) for cold treatment, and the combination of \(RPL13\) and \(PP2A59\gamma\) for drought treatment. The combination of RGs could reduce the influencing of experimental factors on the gene expression, accordingly, it was more accurate

| Rank | Developmental stage | Cold | Drought | Total |
|------|---------------------|------|---------|-------|
| 1    | PP2A59\gamma        | RPL5B| RPL13   | RPL13 |
| 2    | CPY20-1             | RPL13| PP2A59\gamma | UBCE32 |
| 3    | UBCE32              | PP2A59\gamma | CPY20-1 | RPL5B |
| 4    | RPL5B               | UBCE32| UBCE32 | CPY20-1 |
| 5    | UBC18               | EIF4A-14| UBC18 | PP2A59\gamma |
| 6    | G6PDH               | UBC18| RPL5B   | EIF4A-14 |
| 7    | RPL13               | PP2A579| UPL-RHF2A | PP2A579 |
| 8    | ABCC8               | CPY20-1| PP2A579 | UBC18 |
| 9    | PP2A579             | ABCC8| ABCC8   | G6PDH |
| 10   | GAPDH               | G6PDH| EIF4A-14| TUB\(\alpha\)3 |
| 11   | EIF4A-14            | GAPDH| EF1\(\alpha\)| UPL-RHF2A |
| 12   | TUB\(\alpha\)3      | EF1\(\alpha\)| G6PDH | ABCC8 |
| 13   | EF1\(\alpha\)       | UPL-RHF2A| TUB\(\alpha\)3 | GAPDH |
| 14   | UPL-RHF2A           | TUB\(\alpha\)3| GAPDH | EF1\(\alpha\) |
| 15   | ACT\(\beta\)        | ACT\(\beta\)| ACT\(\beta\)| ACT\(\beta\) |

Table 5. The comprehensive ranking of RGs for normalization.

Figure 4. Relative quantity of \(UDGPase\) in different developmental stage with different RGs after normalization. Samples A, B, C were normalized with a single top-ranked RG (TR) - \(PP2A59\gamma\), samples D, E, F were normalized with a combination of multiple top-ranked RGs (TRs) - \(PP2A59\gamma, CPY20-1, UBCE32, RPL5B\) and \(UBC18\), while samples G, H, I were normalized with a single low-ranked RG (LR) - \(ACT\(\beta\)).
and reliable than single RG in the normalization of gene expression. Comprehensive analysis on the selected RGs recommended five RGs combination for gene expression of *C. pilosula*, which are *RPL13*, *UBCE32*, *RPL5B*, *CPY20-1*, and *PP2A59*.

**Materials and Methods**

**Plant material.** One-year-old seedlings of *C. pilosula* were collected as test plant from Tanchang, Gansu province, China on April 5th 2018. The seedlings were cultivated in sandy soil at a temperature of 25 ± 2 °C, relative humidity of 55–60%, irradiance of 300 μmol m⁻² s⁻¹ and 14 h photoperiod (6:00–20:00 h). After 30 days cultivation, test plants were selected from the uniformly grown seedlings with the length of vines between 30 and 50 cm. These plants were subjected to three treatments, which are the normal watering and fertilization for developmental stages study, cold-stress treatment and drought-stress treatment. For developmental stages study, samples of leaf, stem and root were collected after 30, 45 and 60 days. Cold-stress treatment were conducted in incubator under 4 ± 2 °C for 0, 4, 8 and 12 h, respectively. The leaf, stem and root were collected correspondingly from the treated and untreated plants. Drought-stress treatment were carried out by controlling the relative water content of soil at 50 ± 5%, with the normally irrigated soil as controls. The leaf, stem and root were collected after 5 and 10 days from the drought-stress and normally irrigated plants. Three plants were sampled for each replicate and three replicates were required for each treatment. All samples were immediately frozen in liquid nitrogen, and kept frozen at −80 °C for analyses.

**RNA isolation and cDNA synthesis.** Total RNA was extracted from each sample using the RNAprepPure Plant Kit DP441 (Tiangen Biotech CO., LTD, Beijing, China) according to the manufacturer’s instructions, followed by DNase I (Tiangen Biotech CO., LTD, Beijing, China) treatment to eliminate DNA contamination. The integrity of RNA was determined by 1.5% agarose gel electrophoresis. The purity and concentration of total RNA was determined using NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA, US). RNA samples with a concentration higher than 60 ng/μL and a ratio of A260/A280 between 1.8 and 2.0 were required for cDNA preparation. Synthesis of cDNA was conducted in the PrimerScript™ RT cDNA Synthesis Kit (TaKaRa Bio Inc., Dalian, China) using 1.0 μg RNA solution. The obtained cDNA was then diluted with 10 times nuclease-free water to prepare RT-qPCR.
Selection of candidate reference genes. Transcriptome sequencing was performed on C. pilosula samples collected from different growth stages adopting paired-end sequencing technology on an Illumina Hi-Seq™ 4000 platform. After assembling and annotating the transcriptome data, the gene expression level (GEL) was analyzed by calculating TPM (Transcripts per kilobase of exon model per million mapped reads) with the help of EdgeR package46. Multiples of gene expression difference was calculated from the log fold change (FC) value. The related data has not been officially published. According to the literature36,37, the conventional RGs include ACT, GAPDH, G6PDH, EF, 18S rRNA, UBQ, TUB, EIF, UBC, H3, PGK, TEF, CYP, and NADHD. The related genes were retrieved from the transcriptomic data of C. pilosula, and the gene expression levels including values of CV(coefficient of variation) and log FC were calculated. In general, the lower the CV value is, the more stable the expression of relevant genes becomes. Under the conditions of CV < 0.5 and |log2FC| < 0.2, the candidate reference RGs were screened from the genes which ranked high in stability.

Primer design and RT-qPCR. Primers were designed using Primer5 software based on the sequences retrieved from the RNA-Seq dataset of C. pilosula. The criteria for primer design were as follows: primer lengths in 20–24 bp, GC content of 45–55%, melting temperature at 58 °C, and amplicon lengths in 100–250 bp. Primer information for the candidate RGs was listed in Table 2. Reactions were conducted using TB Green Premix Ex Taq II kit (Takara) and amplified on QuantStudio™ 6 Flex Real-Time PCR system (ABI Technologies) in a final volume of 10 μL according to the manufacturer’s instructions. Melting curve analyses were done following the amplification cycles in order to examine the specificity of the reactions. Amplification efficiencies were calculated using the QuantStudio™ Real-Time PCR software (ver 1.1). The PCR efficiency of each primer pair was determined by the slope of the standard curve according to the equation $E = \frac{10^{(-1/slope)} - 1}{100\%}$. Efficiency values should be between 90% and 110%.

Data analysis. The four most commonly used methods including $\Delta$Ct38, geNorm39, NormFinder40 and BestKeeper41, were adopt to calculate and identify the expression stability of candidate reference genes. The $\Delta$Ct method displays the pairwise comparisons by calculating the standard deviation (SD) of each pair candidate RGs and the average SD of each gene. The lower the average SD is, the more stable the RG performs. The geNorm algorithm calculates the expression stability value (M value) and pairwise variation (Vn/n + 1) for all candidate genes. Lower M values suggests higher stability of gene expression. The Vn/n + 1 value determines the optimal
number of RGs for accurate normalization. A cut-off value of Vn/n + 1 < 0.15 indicates that an additional RG makes no significant contribution to the normalization. NormFinder program is based on the ANOVA model, which reveals expression variation by calculating the stability value (SV) when using RGs for normalization. Lower SV indicates higher stability. The BestKeeper method is an Excel-based tool using pairwise correlations, which discloses the expression level of all candidate genes by calculating the SD, coefficient of correlation (r) and coefficient of variance (CV). After rejecting the SD values of RGs higher than one, the candidate RGs was ranked according to the three indexes. The most stable gene expression exhibits the lowest CV ± SD value. Based on the results derived from the four methods, the geometric mean of each candidate gene was calculated to obtain its overall final ranking. A lower geometric mean of ranking indicates more stable expression.

Validation of reference genes. In order to validate the selected RGs under different treatment conditions, relative expression analyses of UDPGase with published data were performed using the 2−ΔΔCt method. The effect of the single RG and the combination of RGs in normalizing the relative expression level of the target gene were compared to determine whether the combination of internal RGs can improve the reliability and accuracy of RT-qPCR results. Gene expression was also normalized using the most unstable ACT/β3 gene for comparison.

Conclusions This research provided the optimal single RG and combination of RGs for C. pilosula. The combination of PP2A59γ, CPY20-1, UBCE32, RPL5B, and UBC18 are suitable for developmental stage, RPL5B, RPL13, and PP2A59γ for cold treatment, and in drought treatment, RPL13 and PP2A59γ were the optimal combination of RGs. These combinations of RGs are recommended as reference genes in RT-qPCR analysis of C. pilosula. This study also demonstrated the validity of using RNA-Seq expression data for the determination of candidate reference genes for RT-qPCR analysis.

Received: 18 April 2019; Accepted: 31 December 2019; Published online: 28 January 2020

References
1. Commission, C. P. Pharmacopoeia of the People’s Republic of China. Vol. I (China Medical Science Press, 2015).
2. Wang, J., Deng, C.-Q., Shi, L. & Deng, M.-L. Progress of modern research on Codonopsis tax. Guide of China Medicine 9, 279–281, https://doi.org/10.15912/j.cnki.gocm.2011.31.068 (2011).
3. Liang, L.-J. et al. Study on Hybrid Characteristics of Medicinally Used Cultivated Codonopsis Species Using Ribosomal Internal Transcribed Spacer (ITS) Sequencing. Molecules 23, 1565, https://doi.org/10.3390/molecules23071565 (2018).
4. He, J. Y. et al. The genus Codonopsis (Campanulaceae): a review of phytochemistry, bioactivity and quality control. Journal of Natural Medicines 69, 1–21, https://doi.org/10.1007/s11418-014-0861-9 (2015).
5. Qing-hong, Z., Jing, L., Jiao-jiao, J., Jian-kuan, L. & Jian-ping, G. Cloning and Expression Analysis of the CpSU4 Gene in Codonopsis pilosula. Journal of Shantxi University (Nat. Sci.) 41, 172–179, https://doi.org/10.13451/j.cnki.shanxi.univ(nat. sci.)2018.03.002 (2018).
6. Xiao-lin, W., Jiao-jiao, J. & Jian-ping, G. Clone and expression of CpGAPDH gene in Codonopsis pilosula. Journal of Chinese Materia Medica 43, 712–720, https://doi.org/10.19540/j.cnki.cjcm.20181005.009 (2018).
7. Jing, L., Qiong-qiong, G., Hai-feng, S. & Jian-ping, G. Cloning, sequence analysis, and prokaryotic expression of CpUGPase gene in Codonopsis Radix. Chinese Traditional and Herbal Drugs 47, 3876–3883, https://doi.org/10.7501/jss.0253-2670.2016.21.021 (2016).
8. Gao, J. P., Wang, D., Cao, L. Y. & Sun, H. F. Transcriptome Sequencing of Codonopsis pilosula and Identification of Candidate Genes Involved in Poly saccharide Biosynthesis. PLoS One 10, e0117342, https://doi.org/10.1371/journal.pone.0117342 (2015).
9. Li, T. et al. Selection and Validation of Appropriate Reference Genes for Quantitative Real-Time PCR Analysis in Isatis indigotica Fort. Frontiers in Plant Science 8, 139, https://doi.org/10.3389/fpls.2017.01539 (2017).
10. Wu, Z.-J., Tian, C., Jiang, Q., Li, X.-H. & Zhuang, J. Selection of suitable reference genes for qRT-PCR normalization during leaf development and hormonal stimuli in tea plant (Camellia sinensis). Scientific Reports 6, 19748, https://doi.org/10.1038/srep19748 (2016).
11. Ma, R., Xu, S., Zhao, Y., Xia, B. & Wang, R. Selection and Validation of Appropriate Reference Genes for Quantitative Real-Time PCR Analysis of Gene Expression in Lycoris aurea. Frontiers in Plant Science 7, 536, https://doi.org/10.3389/fpls.2016.00536 (2016).
12. Tian, C. et al. Selection of Suitable Reference Genes for qPCR Normalization under Abiotic Stresses and Hormone Stimuli in Carrot Leaves. PLoS One 10, e0117569, https://doi.org/10.1371/journal.pone.0117569 (2015).
13. Hao, X. et al. Identification and evaluation of reliable reference genes for quantitative real-time PCR analysis in tea plant (Camellia sinensis (L.) O. Kuntze). International Journal of Molecular Sciences 15, 22155–22172, https://doi.org/10.3390/ijms15122155 (2014).
14. Zhu, J. et al. Reference Gene Selection for Quantitative Real-Time PCR Normalization in Caragana intermedia under Different Abiotic Stress Conditions. PLoS One 8, e31396, https://doi.org/10.1371/journal.pone.0031396 (2013).
15. Dekkers, B. J. W. et al. Identification of Reference Genes for RT-qPCR Expression Analysis in Arabidopsis and Tomato Seeds. Plant and Cell Physiology 53, 28–37, https://doi.org/10.1093/pcp/pcn113 (2012).
16. Maroufi, A., Bockstaele, E. V. & Loose, M. D. Validation of reference genes for gene expression analysis in chicory (Cichorium intybus) using quantitative real-time PCR. BMC Molecular Biology 11, 15, https://doi.org/10.1186/1471-2199-11-15 (2010).
17. Liva, K., J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCt Method. Methods 25, 402–408, https://doi.org/10.1016/j.ymeth.2001.12.062 (2001).
18. Liang, W. et al. Selection and evaluation of reference genes for qRT-PCR analysis in Eucapsis konishii Hayata based on transcriptome data. Plant Methods 14, 42, https://doi.org/10.1186/s13007-018-0311-x (2018).
19. Kozera, B. & Rapacz, M. Reference genes in real-time PCR. Journal of Applied Genetics 54, 391–406, https://doi.org/10.1007/s13353-013-0173-x (2013).
20. Chang, E. et al. Selection of Reference Genes for Quantitative Gene Expression Studies in Platycladus orientalis (Cupressaceae) Using Real-Time PCR. PLoS One 7, e33278, https://doi.org/10.1371/journal.pone.0033278 (2012).
21. Dan-dan, L., Bo, H., Qing, W. & Wei, W. The Research on Reference Genes in Medicinal Plant. Molecular Plant Breeding 15, 903–910, https://doi.org/10.13731/j.mpb010.000903 (2017).
22. Xiao-fa, Z., Jin, W., Hong-wei, S. & Hong-wei, X. The Exploration of 18S rRNA for Quantitative RT-PCR as Reference Gene in Plant. Journal of Illus Normal University (Natural Science Edition) 37, 115–119, https://doi.org/10.16862/j.cnki.issn1674-3873.2016.02.025 (2016).
23. Yu-Fang, Z., Li-Juan, Z. & You-Ling, Z. Selection and Application of Reference Genes for Gene Expression Studies. Plant Physiology Journal 50, 1119–1125, https://doi.org/10.13592/cnkj-pp.14.0201 (2014).
24. Joseph, J. T., Poolakkalody, N. J. & Shah, J. M. Plant reference genes for development and stress response studies. Journal of Biosciences 43, 173–187, https://doi.org/10.1007/s13238-017-9728-z (2018).
25. Chapman, J. R. & Waldenstrom, J. With Reference to Reference Genes: A Systematic Review of Endogenous Controls in Gene Expression Studies. PLoS One 10, e0141853, https://doi.org/10.1371/journal.pone.0141853 (2015).
26. Long, X. Y. et al. Genome-wide identification and evaluation of novel internal control genes for Q-PCR based transcript normalization in wheat. Plant Molecular Biology 74, 307–311, https://doi.org/10.1007/s11103-010-9668-8 (2010).
27. Jonge, H. J. M. D. et al. Evidence Based Selection of Housekeeping Genes. PLoS One 2, e989, https://doi.org/10.1371/journal.pone.0000898 (2007).
28. Hong, S.-Y., Seo, P. I., Yang, M.-S., Xiang, F. & Park, C.-M. Exploring valid reference genes for gene expression studies in Brachypodium distachyon by real-time PCR. BMC Plant Biology 8, 112, https://doi.org/10.1186/1471-2228-8-112 (2008).
29. Gutierrez, L. et al. The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnology Journal 6, 609–618, https://doi.org/10.1111/j.1467-7652.2008.00346.x (2010).
30. Gimonen, J., Eattok, N., Deynne, A. V. & Bluemwald, E. Selection and validation of reference genes for gene expression analysis in switchgrass (Panicum virgatum) using quantitative real-time RT-PCR. PLoS One 9, e9174, https://doi.org/10.1371/journal.pone.0091474 (2014).
31. Shivhare, R. & Lata, C. Selection of suitable reference genes for assessing gene expression in pear miller under different abiotic stresses and their combinations. Scientific Reports 6, 23036, https://doi.org/10.1038/srep23036 (2016).
32. Maroufi, A. Selection of reference genes for real-time quantitative PCR analysis of gene expression in Glycyrhriza glabra under drought stress. Biologia Plantarum 60, 1–10, https://doi.org/10.1007/s10535-016-0601-y (2016).
33. Lekshmy, S. & Iha, S. K. Selection of reference genes suitable for qRT-PCR expression profiling of biotic stress, nutrient deficiency and plant hormone responsive genes in bread wheat. Indian Journal of Plant Physiology 22, 101–106, https://doi.org/10.1007/s40502-017-0282-3 (2017).
34. Luo, M. et al. Selection of reference genes for miRNA qRT-PCR under abiotic stress in grapevine. Scientific Reports 8, 4444, https://doi.org/10.1038/s41598-018-22743-6 (2018).
35. Martins, P. K. et al. Selection of reliable reference genes for RT-qPCR analysis during developmental stages and abiotic stress in Setaria viridis. Scientific Reports 6, 28348, https://doi.org/10.1038/srep28348 (2016).
36. Yang, H. et al. Selection and evaluation of novel reference genes for quantitative reverse transcription PCR (qRT-PCR) based on genome and transcriptome data in Brassica napus L. Gene 538, 113–122, https://doi.org/10.1016/j.gene.2013.12.057 (2014).
37. Czechowski, T., Stitt, M., Altman, T., Udvardi, M. K. & Scheible, W. R. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiology 139, 5–17, https://doi.org/10.1104/pp.105.063743 (2005).
38. Silver, N., Best, S., Jiang, J. & Thein, S. L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Molecular Biology 7, 33, https://doi.org/10.1186/1471-2199-7-33 (2006).
39. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology 3, research0034–0031, https://doi.org/10.1186/gb-2002-3-7-research0034 (2002).
40. Andersen, C. L., Jensen, J. L. & Orntoft, T. F. Normalization of realtime quantitative reverse transcription–PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data set. Cancer Research 64, 5245–5250, https://doi.org/10.1158/0008-5472.CAN-04-0496 (2004).
41. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. Biotechnology Letters 26, 509–515, https://doi.org/10.1023/B:BILE.0000019559.84305.47 (2004).
42. Stanton, K. A. et al. A Whole–Transcriptome Approach to Evaluating Reference Genes for Quantitative Gene Expression Studies: A Case Study in Mimulus. G3 (Bethesda), 7, 1085–1095, https://doi.org/10.1534/g3.116.038075 (2017).
43. Zhao, J. et al. Genome-Wide Constitutively Expressed Gene Analysis and New Reference Gene Selection Based on Transcriptome Data: A Case Study from Poplar/Canker Disease Interaction. Frontiers in Plant Science 8, 1876, https://doi.org/10.3389/fpls.2017.01876 (2017).
44. Mallona, I., Lischewski, S., Weiss, J., Hause, B. & Egea-Cortines, M. Validation of reference genes for quantitative real-time PCR during leaf and flower development in Petunia hybrida. BMC Plant Biology 10, 4, https://doi.org/10.1186/1471-2229-10-4 (2010).
45. Sabeh, M., Decepepe, M.-O., St-Arnaud, M. & Mitrice, B. Transcription-wide selection of a reliable set of reference genes for gene expression studies in potato cyst nematodes (Globodera spp.). PLoS One 13, e0193840, https://doi.org/10.1371/journal.pone.0193840 (2018).
46. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140, https://doi.org/10.1093/bioinformatics/btp616 (2010).
47. Radonic, A. et al. Guidance to reference gene selection for quantitative real-time PCR. Biochemical and Biophysical Research Communications 313, 856–862, https://doi.org/10.1016/j.bbrc.2003.11.177 (2004).

Acknowledgements
This research was funded by the 13th five-year plan of Ministry of science and technology of the People's Republic of China, grant number 2015BAC-01B-03.

Author contributions
Z.L. supervised the project and gained financial support; L.L., H.Y. and Z.H. conceived the experiments; L.L., E.W., X.Z., B.Z., C.Z. and Z.L. performed the experiments; L.L. and H.Y. analysed the data; L.L. and Z.H. wrote and edited the manuscript; Z.L. and C.Z. modified the manuscript. All authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-58328-5.
Correspondence
Correspondence and requests for materials should be addressed to Z.L.

Reprints and permissions information
Reprints and permissions information is available at www.nature.com/reprints.
