

Transcriptome-wide identification and validation of reference genes of Black rockfish (*Sebastes schlegelii*) for quantitative RT-PCR

Chaofan Jin

Ocean University of China

Weihao Song

Ocean University of China

Mengya Wang

Ocean University of China

Jie Qi

Ocean University of China

Quanqi Zhang

Ocean University of China

Yan He (✉ yanhe@ouc.edu.cn)

<https://orcid.org/0000-0001-7346-6758>

Research article

Keywords: reference genes, black rockfish, transcriptome-wide, CV value, qRT-PCR

Posted Date: March 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-17096/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The quantitative real-time reverse transcription PCR (qRT-PCR) is a widely used technique that relies on the reference gene for gene expression normalization. Selecting a suitable reference gene is a crucial step to obtain an accurate result in qRT-PCR. However, most previous studies of fishes adopted reference genes that were commonly used in mammals without validation.

Results: In this study, we utilized 89 transcriptome datasets covering early developmental stages and adult tissues, and carried out transcriptome-wide identification and validation of reference genes in *Sebastes schlegelii*. Finally, 121 candidate reference genes were identified based on four criteria. Eight candidates (METAP2, BTF3L4, EIF5A1, TCTP, UBC, PAIRB, RAB10, and DLD) and four commonly used reference genes (TUBA, ACTB, GAPDH, RPL17) in mammals were selected for validation via qRT-PCR and four statistical methods (delta-Ct, BestKeeper, geNorm, and NormFinder). The results revealed that the candidate reference genes we recommended are more stable than traditionally used ones.

Conclusions: This is the first study to conduct transcriptome-wide identification and validation of reference genes for quantitative RT-PCR in the black rockfish, and lay an important foundation for gene expression analysis in teleost.

Introduction

The quantitative real-time reverse transcription PCR (qRT-PCR) is a popularized technique for quantification of relative mRNA levels due to the advantage of its low cost, specificity, sensitivity and large dynamic range [1, 2]. In qRT-PCR, the expression levels of target genes of interest are usually normalized to the endogenous controls, also called reference genes. The purpose of reference genes is to remove or reduce the variations in the quantity and quality of RNA among samples, thus making it possible for accurate comparison of transcript abundance of a gene of interest among various samples [2, 3]. Selecting an ideal reference gene that showing constant expression across all samples is an essential step for normalization and quantitative analysis in RT-qPCR. Despite the significance of a suitable reference gene has been widely confirmed, many studies are using commonly used reference genes without validation in qRT-PCR assays. Some commonly used reference genes were confirmed to be unsuitable in different conditions [3]. For example, GAPDH, a widely used reference gene in human, has been reported to be an unsuitable reference gene in some species [4, 5], and so is many other traditionally reference genes, such as 18sRNA and ACTB [6].

However, there are many emerging studies to identify and validate reliable reference genes with various methods with the development of biotechnology. For example, serial analysis of gene expression, microarray analysis, and expressed tag sequences have been widely used for reference genes identification in human and other mammals [3, 7, 8], but there are few similar studies conducted in other species. In recent years, RNA-seq has been becoming a useful experimental tool and the genes with stable expression patterns calculated from the transcriptome datasets could be a reliable source for

reference genes selection. There are many studies carried out utilizing RNA-seq for reference genes selection not only in model organisms, but also in many non-model organisms, like kiwi, seaweed, and scallop [9–11].

The black rockfish (*S.schlegelii*, hereafter denoted 'rockfish') evolving live birth, is a good model of teleost to study the adaption to viviparity [12]. Some studies have been carried out focusing on the genes related to development, growth, immunity, and reproduction in this group of fish [13–15]. 18sRNA, ACTB, and GAPDH, are usually adopted as the endogenous control of qRT-PCR in rockfish studies. A previous study of our group compared the stabilities of several commonly used reference genes (18S RNA, ACTB, GAPDH, TUBA, RPL17, EF1A, HPRT, and B2M) and defined RPL17, EF1A and ACTB as the most reliable three of them [16]. However, some inconsistent results are still observed even we used the “most stable” reference gene in rockfish studies. These can be explained that the “most stable reference gene” is selected by “expression keeps stable in adult tissues” criteria but it is not confirmed if it is still stable during early developmental stages. It is the best way to combine as most samples as possible to select the reference genes. Our group sequenced the genome of rockfish and created a large set of transcriptomes. These provides valuable resources to conduct the transcriptome-wide analysis of suitable reference genes selection.

In this study, we analyzed 89 transcriptome datasets, and 121 candidate reference genes were identified based on four criteria. The expression stabilities of eight candidates and four commonly used genes were validated by qRT-PCR and four statistical methods. The results revealed that candidates we recommended are more stable than traditionally used reference genes. This will benefit further gene quantification analysis of the rockfish and other teleost species.

Results

1. Selection of the reference genes in Rockfish from the transcriptome data

We identified rockfish candidate reference genes among 24094 transcripts from 89 RNA-seq datasets. Firstly, we obtained 8357(34.7%), 439(2%), 242(1%) genes respectively, according to the first three criteria (Fig. 1). Then, we figured out the median $\text{Log}_2(\text{TPM})$ value among genes that met the first three criteria was approximately six. Finally, the fourth criteria were applied to screen genes with an average $\text{Log}_2(\text{TPM}) > 6$, resulting in 121 candidate reference genes for all tissues and developmental stages in rockfish (Fig. 1). As shown in Table 2, the 15 most stable genes have mean $\text{Log}_2(\text{TPM})$ values ranging from 6.33 to 11.00, and the lowest CV values varying from 0.069 to 0.093. Then, we performed functional enrichment analysis of candidate reference genes. As shown in Fig. 2, GO enrichment analysis revealed that candidate reference genes are highly associated with metabolic process, cellular metabolic process and primary metabolic process. Some of these genes are also associated with protein process and protein complex.

2. RNA-seq analysis of candidate and commonly used reference genes expression level

To validate the candidate reference genes we identified, eight genes chosen from the top 15 candidates and four commonly used were selected to compare their expression profiles. The eight candidates include METAP2, EIF5A1, BTF3L4, TPT1, UBC, RAB10, DLD and PAIRB, the four commonly used reference genes include TUBA, ACTB, RPL17 and GAPDH (Table 3).

Concerning Log_2TPM values, candidate reference genes have lower variances than commonly used reference genes, as shown in Fig. 3. For example, GAPDH, a widely used reference genes in mammals, in which Log_2TPM values (ranging from 2 to 14) exhibit high variances in rockfish tissues and developmental stages. Besides, RPL17 and ACTB, also show relatively high variances. In contrast to these commonly used reference genes, candidates identified from the transcriptome datasets have smaller variances in Log_2TPM values among different tissues and developmental stages. As METAP2, which expression level is relatively stable with the Log_2TPM values ranging from 7.1 to 10.2. Our transcriptome datasets suggest the candidate reference genes identified are more stable than commonly used reference genes in rockfish tissues and developmental stages.

3. qRT-PCR validation of candidates and commonly used reference genes

To validate the results from our transcriptome datasets, we conducted the RT-qPCR of twelve genes and four methods (comparative delta-Ct, geNorm, Bestkeeper, and NormFinder) were used for further expression stability evaluation. We used the Pearson coefficient to determine the correlation between Ct values and $\text{Log}_2(\text{TPM})$ values, and there is a negative correlation between two values ($R = -0.636$, $P < 0.01$).

As shown in Fig. 4, the Ct values of each gene in different tissues and developmental stages were determined. Among commonly used genes, GAPDH displays the highest variation of Ct values (from 17 to 32), and RPL17 (Ct 21–29) also expresses variable in different tissues and developmental stages. TUBA and ACTB, two of commonly used reference genes, have relative stable Ct values. In contrast to commonly used genes, most candidates have smaller variances than commonly used reference genes. For example, METAP2 (Ct 22–25), RAB10 (25–27), EIF5A1 (24–27) display very low expression variations. The results that preliminary comparison of Ct values showed most candidates are relatively more stable than commonly used reference genes, in consistent with our transcriptome data.

4. Stability analysis between candidates and commonly used reference genes

After having summarized the RT-qPCR results, it's necessary to compare the expression stability between the candidates and commonly used reference genes by the statistical methods, for determining the optimal reference genes in further normalization. Here, we applied the four statistical methods for stability analysis: delta-Ct, BestKeeper, geNorm and NormFinder.

According to the delta-Ct method, RAB10 was identified as the most stable reference gene with the highest ranking value. And UBC, METAP2, TPT1/PAIRB were most stable genes by the BestKeeper, NormFinder, geNorm methods respectively. In consistent with our observation of CV values and Ct Values, GAPDH and RPL17 are most unstable with the lowest ranking values among twelve genes by four methods. Although the stability rankings are variable of these genes by different methods, the candidates selected from transcriptome data tend to have higher ranking values compared to the commonly used reference genes. The result of stability analysis by statistical methods further confirmed the reliability of candidates selected based on transcriptome data.

Discussion

The inappropriate choice of reference genes for normalizing transcript levels of test genes before the comparative analysis of different biological samples will cause systematic errors in the application of methods for qRT-PCR analysis, which can compromise the interpretation of results. As a crucial step, selecting optimal reference genes is vital important for getting an accurate result in gene quantitative analysis under different experimental conditions. But the same reference genes in different species could have dramatic expression levels, especially between mammals and lower vertebrates like fish. For example, GAPDH has been widely identified and used as a housekeeping genes in daily experiments with high stabilities in mammals [17, 18], but it has a significant expression variation in Atlantic halibut and zebrafish [19, 20]. Even so, most reference genes for qRT-PCR in fish species are selected from commonly used reference genes in mammals, which would lead to unnecessary errors in gene quantitative analysis. With the wide application of omics in molecular biology, transcriptomic data has been becoming a reliable source for the selection of optimal reference genes for gene quantitative analysis [10, 21, 22].

In this study, we used a modified method, developed by the study of Eisenberg and Levanon, to select suitable reference genes in rockfish. The reasonability and advantage of the method we used was detailed described in a previous work [11]. We aimed to screen reference genes with high stability among different tissues and early developmental stages, so we integrated the transcriptome data of different adult tissues and developmental stages. By four criteria, only 121 genes with CV values < 1 were selected from all 24094 genes, suggesting the method we used is very stringent. Besides, we have analyzed the transcriptome data of tissues and early developmental stages respectively by four criteria. Interestingly, there were 375 and 752 genes that were obtained, suggesting a significant difference in gene expression patterns between tissues and developmental stages, which is similar to some other studies [23–25]. After comparing two gene sets obtained from tissues and developmental stages, we obtained a core set of genes (122 genes), most genes are overlapped with the candidate reference genes from 121 gene sets, confirming the stringency and reliability of the methods we applied.

Top 15 candidate reference genes with the lowest CV values are all annotated proteins, and most of them have been widely studied in model organisms. For example, UBC has been identified as a reliable reference gene in some organisms [26–28]. By GO enrichment analysis of all candidate reference genes, as expected, candidates are highly enriched in biological process associated with metabolic process, cellular metabolic process, and protein metabolic process, suggesting they may play an essential part in maintenance of basal cellular functions, as described in other organisms of human, mouse, maize, and scallop [29–31].

The results of RNA-seq analysis were validated by RT-qPCR of eight from the top 15 candidates and four commonly used reference genes, and four statistical methods were used to evaluate their expression stabilities. To be honest, the samples used for qRT-PCR differ from the samples used for RNA-seq, but the stability ranking values of most reference genes obtained from $\text{Log}_2(\text{TPM})$ values and Ct values are still consistent. Further, candidates, selected in this study usually have higher stabilities than these commonly used reference genes. ACTB and GAPDH are the least stable genes among twelve genes, which is consistent with a previous study comparing some commonly used reference genes in rockfish, and they have been widely described to be unsuitable reference genes for gene quantitative analysis in some species [25, 32, 33].

Intriguingly, TUBA gets a medium ranking value by qRT-PCR, showing it is more stable than the other three commonly used reference genes. But TUBA is not in our candidate reference genes list. In the process of screening reference genes based on four criteria, TUBA was filtered out during the second step for screening genes with low dispersion degree over all tissues and developmental stages [$\text{Log}_2(\text{TPM})$] < 1 (The standard deviation value < 1). The CV of TUBA is 0.131. Combining these two results, we don't think TUBA is a perfect reference gene in rockfish, since it is not like the candidates we recommended like RAB10, EIF5A1, PAIRB and BTF3L4, with the high stability ranking values evaluated both from RNA-seq and qRT-PCR. Based on our analysis and validation, the candidate reference genes we screened are the most stable.

In non-model organisms, especially in teleost, the reference genes used for normalization in qRT-PCR experiments usually come from mammals, like GAPDH, ACTB and TUBA. Their unstable expression under different conditions may cause misleading results during normalization. By combining RNA-seq and qRT-PCR analysis, we efficiently identify the most stable reference genes, which could be widely applied in teleost studies.

Conclusions

In this study, 89 transcriptome datasets of rockfish were utilized to identify reference genes, and 121 candidate reference genes were selected from the transcriptomes based on four criteria. Combining RNA-seq and qRT-PCR assays, the expression stabilities of candidates we identified were validated to be higher than commonly used reference genes. Our studies provide valuable resources for further gene expression analysis in teleost.

Material And Methods

1. Sample collection

All adult rockfish used for our study were obtained from the fish hatchery of Qingdao Beibao Aquatic Co., Ltd, Shandong, China. They were transported to our laboratory and maintained in aerated seawater for acclimation. Then, they were sacrificed and the following tissues were dissected: heart, liver, spleen, kidney, brain, gill, intestine, testis and ovary. All samples were frozen in liquid nitrogen and stored at -80°C for further RNA extraction. Each tissue was collected in triplicate or more.

Embryos at six different stages (1-cell, 32-cell, blastula, gastrula, somite and pre-hatching) were collected as described in our previous study [12]. Then, they were stored at -80°C for further use.

2. RNA-seq datasets

In total, 89 RNA-libraries including 63 from different tissues, and 26 from six developmental stages (1-cell, 32-cell, blastula, gastrula, somite and pre-hatching) with three biological replicates or more were sequenced 100 bp from each end using BGI-500 platform. Then, the transcriptome data were processed and the raw counts were converted into TPM values using the RNA-seq by Expectation Maximization software package [34].

3. Identification of reference genes using transcriptome data

Modified methods as described in [11], referring to Eisenberg and Levanon [8] were used to select reference genes among all tissues and different developmental stages of rockfish. Briefly, four criteria utilized to screen reference genes were as follows: expression could be detected in all tissues and developmental stages of embryo, low dispersion degree over all tissues and developmental stages [$\text{Log}_2(\text{TPM}) < 1$], no $\text{Log}_2(\text{TPM})$ differed from the mean $\text{Log}_2(\text{TPM})$ by two or more in each tissue and developmental stage regarding expression level, and $\text{Log}_2(\text{TPM}) > 6$, the median of all $\text{Log}_2(\text{TPM})$ values over all tissues and developmental stages. As described in [11], the stability of reference genes was further evaluated according to the CV value (stdev/mean). Based on the criteria mentioned above, mean $\text{Log}_2(\text{TPM}) > 6$, and dispersion degree [$\text{Log}_2(\text{TPM}) < 1$], so the CV values of reference genes we identified are always less than 0.17.

4. RNA extraction and cDNA synthesis

Total RNA was extracted using Trizol reagents(Invitrogen) according to the manufacturer's protocol. After RNA purification, the integrity and concentration of RNA extracted was assessed via 1.5% agarose gel electrophoresis and determined by a Nano photometer. cDNA was synthesized by Reverse Transcriptase M-MLV kit(Takara) and final volume was 20 ul, the cDNA products were stored at -20°C for further use.

5. Primer design and RT-PCR

In total, we selected ten candidate reference genes from transcriptome data, and four commonly used reference genes for further RT-PCR validation. Primers of these genes were designed by Primer 5.0 software. The criteria of primer design were as follows: amplicon lengths of 80–200 bp, annealing temperature of 60–63°C, primer lengths of 18–25 bp, and GC content of 40–60%. The information of primers is listed in Table S1. All primers amplification efficiencies were determined by the standard curve generated from two-fold dilution series of cDNA concentration (0.625 ng/ul-20 ng/ul).

RT-PCR was conducted in a 384-well plate using a Light Cycler 480 RT-PCR system (Roche Molecular Biochemical, Mannheim, Germany). Each reaction contained 2 ul cDNA(5 ng/ul), 10 ul Light Cycler 480 SYBR Green I Master, 0.5 ul of each primer and 7 ul ddH₂O. Three biological replicates were required for each sample, and three technical replicates were performed for each biological replicate. RT-PCR cycling conditions were 95°C for 5 minutes, followed by 45 cycles at 95°C(15 s), and 61°C(45 s). The specificity of each primer was determined by melting curve and gel picture.

6. RT-PCR data analysis

The stabilities of reference genes determined in our study were evaluated by ReFinder (<https://www.heartcure.com.au/reffinder/>) [23], which integrates four commonly used statistical approaches: geNorm, NormFinder, Bestkeeper, and comparative delt-Ct method [35–37]. Then the geometric mean was calculated to guarantee a comprehensive analysis [38].

Abbreviations

METAP2: Methionine aminopeptidase 2; BTF3L4: Transcription factor BTF3 homolog 4; EIF5A1: Eukaryotic translation initiation factor 5A-1; TCTP: Translationally-controlled tumor protein; UBC: Ubiquitin C; PAIRB: Plasminogen activator inhibitor 1 RNA-binding protein; RAB10: Ras-related protein Rab-10; DLD: Dihydrolipoamide dehydrogenase; TUBA: α -Tubulin; RPL17: Ribosomal protein L17; ACTB: β -actin; GAPDH: Glyceraldehyde-3-phosphate dehydrogenase; qRT-PCR: Quantitative reverse transcription PCR; TPM: Transcripts per million

Declarations

Ethics approval and consent to participate

The animal materials of rockfish used in this study were obtained from Beibao Aquatic Co., Ltd. This study was conducted in accordance with the Animal Care and Use Committee of the Ocean University of China and the China Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research, and Training.

Availability of data and materials

All data generated or analyzed in this article are referred to or included within the article and its additional files.

Consent for publication

Not applicable.

Competing interests

We declare that we have no financial and personal conflicts with other people or organizations.

Author Contributions

CJ contributed to the experimental process, data analysis and manuscript writing; YH conceived and designed the experiments; WS and MW contributed to the experimental process; JQ and QZ contributed to the acquisition of experiment samples; and all authors have read and approved the final manuscript.

Acknowledgements

This study was supported by the National Key R&D Program of China (2018YFD0900101).

Author details

¹MOE Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China; ²Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, China

References

- [1] J. Huggett, K. Dheda, S. Bustin, and A. Zumla, "Real-time RT-PCR normalisation; strategies and considerations," *Genes and immunity*, vol. 6, no. 4, p. 279, 2005.
- [2] K. Dheda, J. F. Huggett, S. A. Bustin, M. A. Johnson, G. Rook, and A. Zumla, "Validation of housekeeping genes for normalizing RNA expression in real-time PCR," *Biotechniques*, vol. 37, no. 1, pp. 112-119, 2004.

- [3] J. Zhu, F. He, S. Song, J. Wang, and J. Yu, "How many human genes can be defined as housekeeping with current expression data?," *BMC genomics*, vol. 9, no. 1, p. 172, 2008.
- [4] X.-Y. Long *et al.*, "Genome-wide identification and evaluation of novel internal control genes for Q-PCR based transcript normalization in wheat," *Plant molecular biology*, vol. 74, no. 3, pp. 307-311, 2010.
- [5] M. Ghani, C. Sato, and E. Rogaeva, "Segmental duplications in genome-wide significant loci and housekeeping genes; warning for GAPDH and ACTB," *Neurobiology of aging*, vol. 34, no. 6, pp. 1710. e1-1710. e4, 2013.
- [6] P. A. Olsvik, K. K. Lie, A.-E. O. Jordal, T. O. Nilsen, and I. Hordvik, "Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon," *BMC molecular biology*, vol. 6, no. 1, p. 21, 2005.
- [7] V. E. Velculescu *et al.*, "Analysis of human transcriptomes," *Nature genetics*, vol. 23, no. 4, p. 387, 1999.
- [8] E. Eisenberg and E. Y. Levanon, "Human housekeeping genes, revisited," *TRENDS in Genetics*, vol. 29, no. 10, pp. 569-574, 2013.
- [9] J. Liu *et al.*, "Genome-wide identification and validation of new reference genes for transcript normalization in developmental and post-harvested fruits of *Actinidia chinensis*," *Gene*, vol. 645, pp. 1-6, 2018.
- [10] D. Gao, F. Kong, P. Sun, G. Bi, and Y. Mao, "Transcriptome-wide identification of optimal reference genes for expression analysis of *Pyropia yezoensis* responses to abiotic stress," *BMC genomics*, vol. 19, no. 1, p. 251, 2018.
- [11] Y. Li *et al.*, "Systematic identification and validation of the reference genes from 60 RNA-Seq libraries in the scallop *Mizuhopecten yessoensis*," *BMC genomics*, vol. 20, no. 1, p. 288, 2019.
- [12] Y. He *et al.*, "A chromosome-level genome of black rockfish, *Sebastes schlegelii*, provides insights into the evolution of live birth," *Molecular ecology resources*, 2019.
- [13] R. Kugapreethan, N. Umasuthan, Q. Wan, W. S. Thulasitha, C. Kim, and J. Lee, "Comparative analysis of two thioredoxin-like genes in black rockfish *Sebastes schlegelii* and their possible involvement in redox homeostasis and innate immune responses," *Developmental & Comparative Immunology*, vol. 67, pp. 43-56, 2017.
- [14] L. Ma *et al.*, "Characterization of the *Dmrt1* gene in the black rockfish *Sebastes schlegelii* revealed a remarkable sex-dimorphic expression," *Fish physiology and biochemistry*, vol. 40, no. 4, pp. 1263-1274, 2014.
- [15] H. S. Kim and S. H. Cho, "Dietary inclusion effect of feed ingredients showing high feeding attractiveness to rockfish (*Sebastes schlegelii* Hilgendorf 1880) on the growth performance, feed

utilization, condition factor and whole body composition of fish (II)," *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, vol. 231, pp. 66-73, 2019.

[16] M. Liman *et al.*, "Selection of reference genes for reverse transcription quantitative real-time PCR normalization in black rockfish (*Sebastes schlegeli*)," *Marine genomics*, vol. 11, pp. 67-73, 2013.

[17] Y. Sun, Y. Li, D. Luo, and D. J. Liao, "Pseudogenes as weaknesses of ACTB (Actb) and GAPDH (Gapdh) used as reference genes in reverse transcription and polymerase chain reactions," *PloS one*, vol. 7, no. 8, p. e41659, 2012.

[18] G. Niu *et al.*, "Identifying suitable reference genes for gene expression analysis in developing skeletal muscle in pigs," *PeerJ*, vol. 4, p. e2428, 2016.

[19] J. M. Fernandes, M. Mommens, Ø. Hagen, I. Babiak, and C. Solberg, "Selection of suitable reference genes for real-time PCR studies of Atlantic halibut development," *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, vol. 150, no. 1, pp. 23-32, 2008.

[20] R. Tang, A. Dodd, D. Lai, W. C. McNabb, and D. R. Love, "Validation of zebrafish (*Danio rerio*) reference genes for quantitative real-time RT-PCR normalization," *Acta biochimica et biophysica Sinica*, vol. 39, no. 5, pp. 384-390, 2007.

[21] Y. Fu, W. He, L. Wang, and Y. Wei, "Selection of appropriate reference genes for quantitative real-time PCR in *Oxytropis ochrocephala* Bunge using transcriptome datasets under abiotic stress treatments," *Frontiers in plant science*, vol. 6, p. 475, 2015.

[22] H. Yang, J. Liu, S. Huang, T. Guo, L. Deng, and W. Hua, "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*, vol. 538, no. 1, pp. 113-122, 2014.

[23] C. Liu *et al.*, "Reference gene selection for quantitative real-time RT-PCR normalization in the half-smooth tongue sole (*Cynoglossus semilaevis*) at different developmental stages, in various tissue types and on exposure to chemicals," *PLoS One*, vol. 9, no. 3, p. e91715, 2014.

[24] W. C. Carlyle, C. A. Toher, J. R. Vandervelde, K. M. McDonald, D. C. Homans, and J. N. Cohn, "Changes in β -actin mRNA expression in remodeling canine myocardium," *Journal of molecular and cellular cardiology*, vol. 28, no. 1, pp. 53-63, 1996.

[25] S. M. Jorgensen, E. J. Kleveland, U. Grimholt, and T. Gjoen, "Validation of reference genes for real-time polymerase chain reaction studies in Atlantic salmon," *Marine biotechnology*, vol. 8, no. 4, pp. 398-408, 2006.

[26] Y.-Y. C. Mosley and H. HogenEsch, "Selection of a suitable reference gene for quantitative gene expression in mouse lymph nodes after vaccination," *BMC research notes*, vol. 10, no. 1, p. 689, 2017.

- [27] X. Zhao *et al.*, "Reference gene selection for quantitative real-time PCR of mycelia from *Lentinula edodes* under high-temperature stress," *BioMed research international*, vol. 2018, 2018.
- [28] M.-L. Wang, Q.-H. Li, H.-H. Xin, X. Chen, X.-J. Zhu, and X.-H. Li, "Reliable reference genes for normalization of gene expression data in tea plants (*Camellia sinensis*) exposed to metal stresses," *PloS one*, vol. 12, no. 4, p. e0175863, 2017.
- [29] J. Zeng *et al.*, "Identification and analysis of house-keeping and tissue-specific genes based on RNA-seq data sets across 15 mouse tissues," *Gene*, vol. 576, no. 1, pp. 560-570, 2016.
- [30] F. Lin, L. Jiang, Y. Liu, Y. Lv, H. Dai, and H. Zhao, "Genome-wide identification of housekeeping genes in maize," *Plant molecular biology*, vol. 86, no. 4-5, pp. 543-554, 2014.
- [31] X. She, C. A. Rohl, J. C. Castle, A. V. Kulkarni, J. M. Johnson, and R. Chen, "Definition, conservation and epigenetics of housekeeping and tissue-enriched genes," *BMC genomics*, vol. 10, no. 1, p. 269, 2009.
- [32] Y. Du, L. Zhang, F. Xu, B. Huang, G. Zhang, and L. Li, "Validation of housekeeping genes as internal controls for studying gene expression during Pacific oyster (*Crassostrea gigas*) development by quantitative real-time PCR," *Fish & shellfish immunology*, vol. 34, no. 3, pp. 939-945, 2013.
- [33] H. J. De Jonge *et al.*, "Evidence based selection of housekeeping genes," *PloS one*, vol. 2, no. 9, p. e898, 2007.
- [34] B. Li and C. N. Dewey, "RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome," *BMC bioinformatics*, vol. 12, no. 1, p. 323, 2011.
- [35] J. Vandesompele *et al.*, "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes," *Genome biology*, vol. 3, no. 7, p. research0034. 1, 2002.
- [36] C. L. Andersen, J. L. Jensen, and T. F. Ørntoft, "Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets," *Cancer research*, vol. 64, no. 15, pp. 5245-5250, 2004.
- [37] N. Silver, S. Best, J. Jiang, and S. L. Thein, "Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR," *BMC molecular biology*, vol. 7, no. 1, p. 33, 2006.
- [38] D. Chen, X. Pan, P. Xiao, M. A. Farwell, and B. Zhang, "Evaluation and identification of reliable reference genes for pharmacogenomics, toxicogenomics, and small RNA expression analysis," *Journal of cellular physiology*, vol. 226, no. 10, pp. 2469-2477, 2011.

Tables

Due to technical limitations, the tables are only available as a download in the supplemental files section.

Table 1: The top 15 candidate reference genes identified from the 89 transcriptome datasets of *S.schlegeli*.

Table 2: Detailed information on the 8 selected candidate reference genes and the 4 commonly used reference genes.

Figures

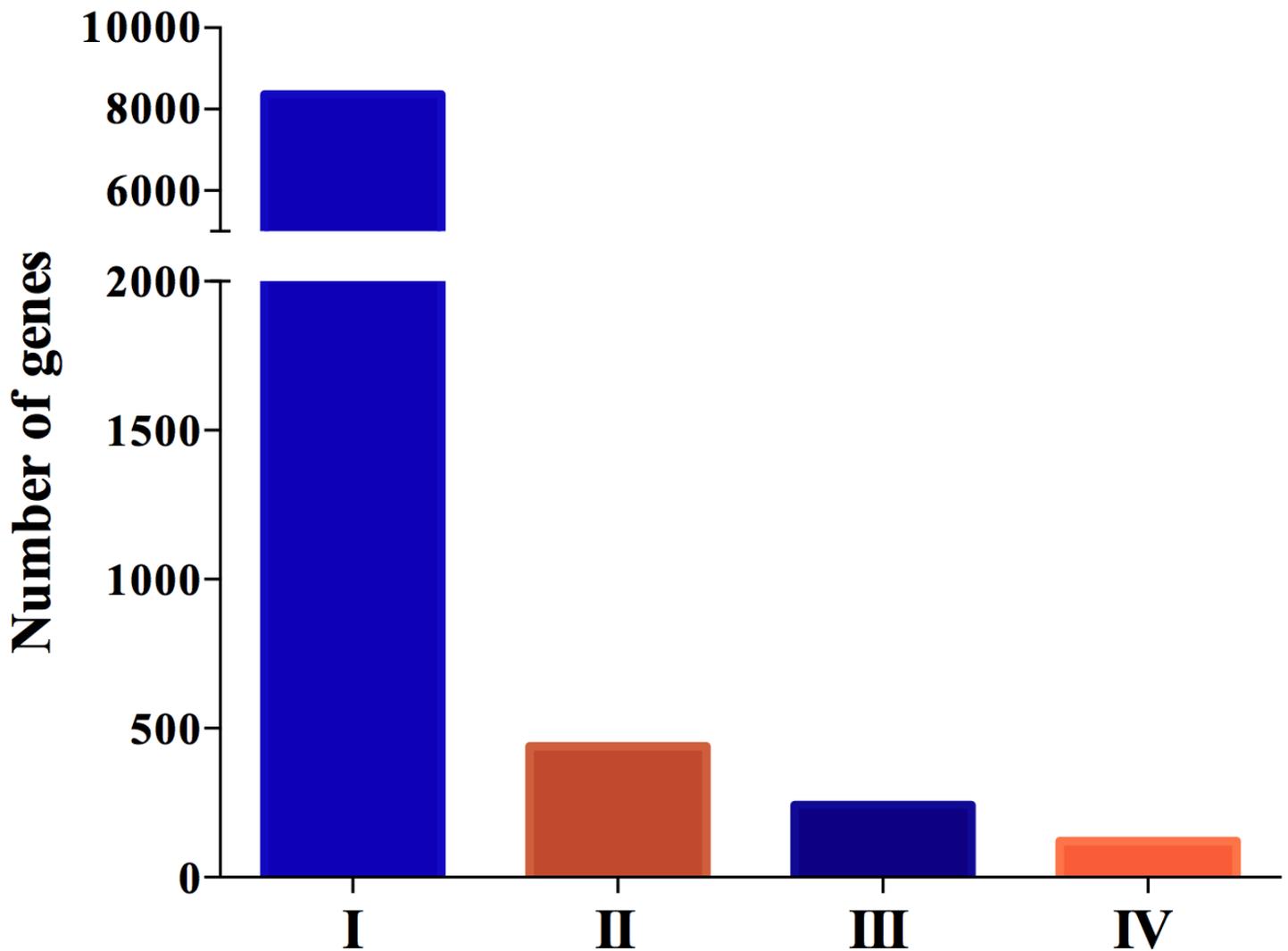


Figure 1

The number of genes that met our fourth criteria are shown. (I) $TPM > 0$; (II) $standard-deviation[\log_2(TPM)] < 1$; (III) no $\log_2(TPM)$ differed from the mean $\log_2(TPM)$ by two or more; (IV) $mean[\log_2(TPM)] > 6$.

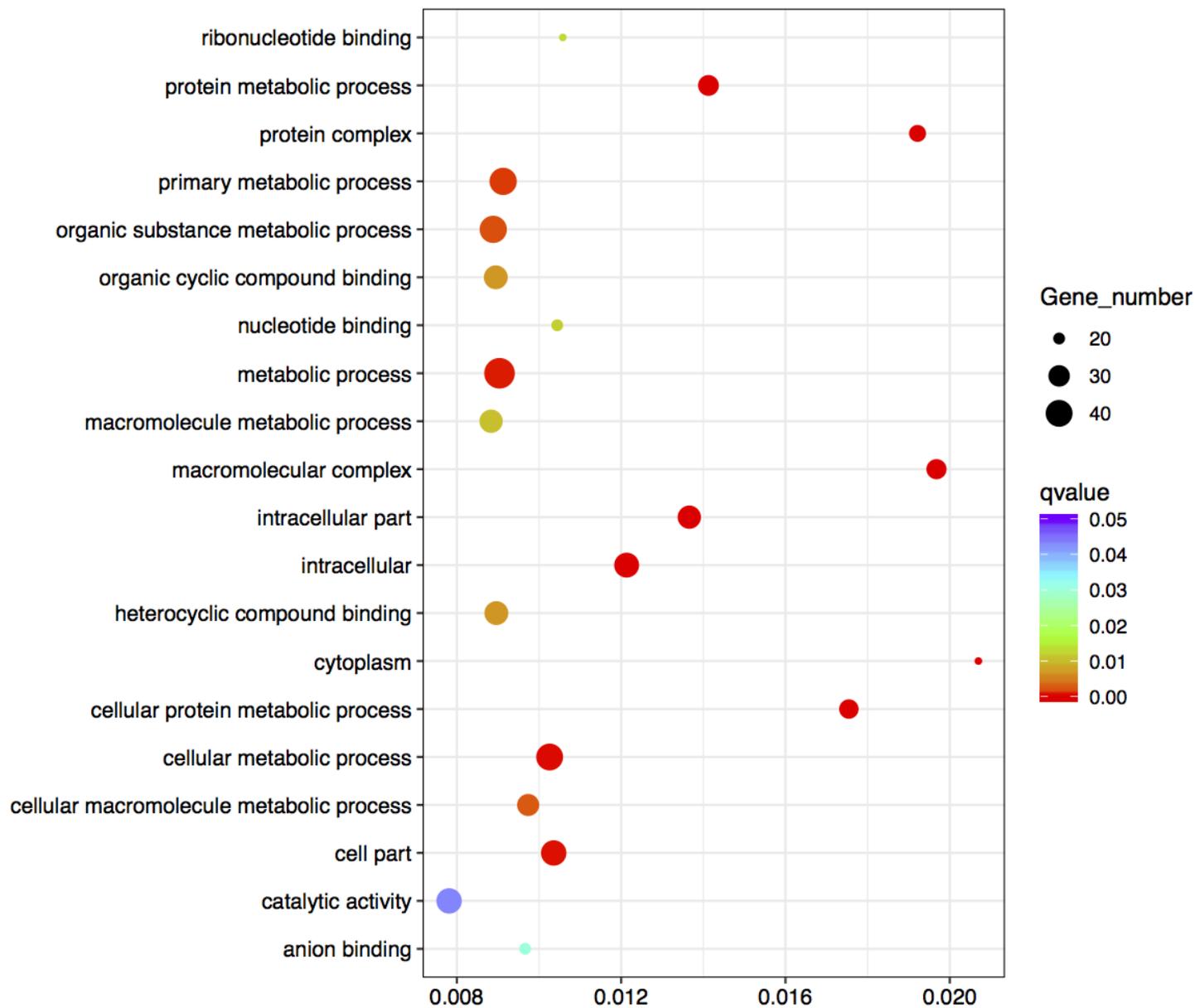


Figure 2

Functional distribution of GO annotation of candidate reference genes. The solid circle represents the number of genes. The x-axis represents the q-value. The y-axis represents the Gene Ontology functional category.

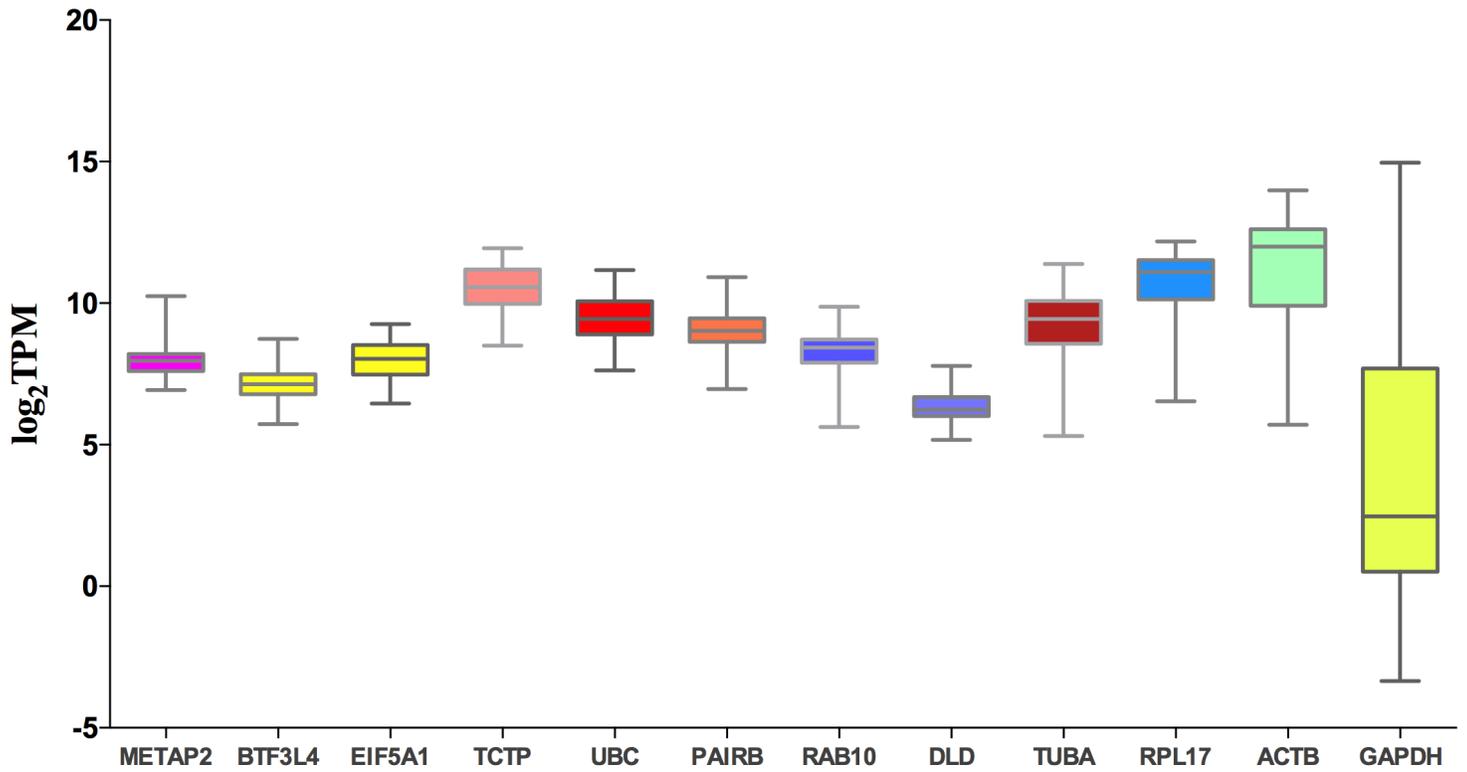


Figure 3

Evaluation of the reference gene candidates and commonly used reference genes based on RNA-seq analysis. A boxplot showing the log₂(TPM) values of the 8 candidate reference genes and 4 commonly used reference genes among tissues and developmental stages.

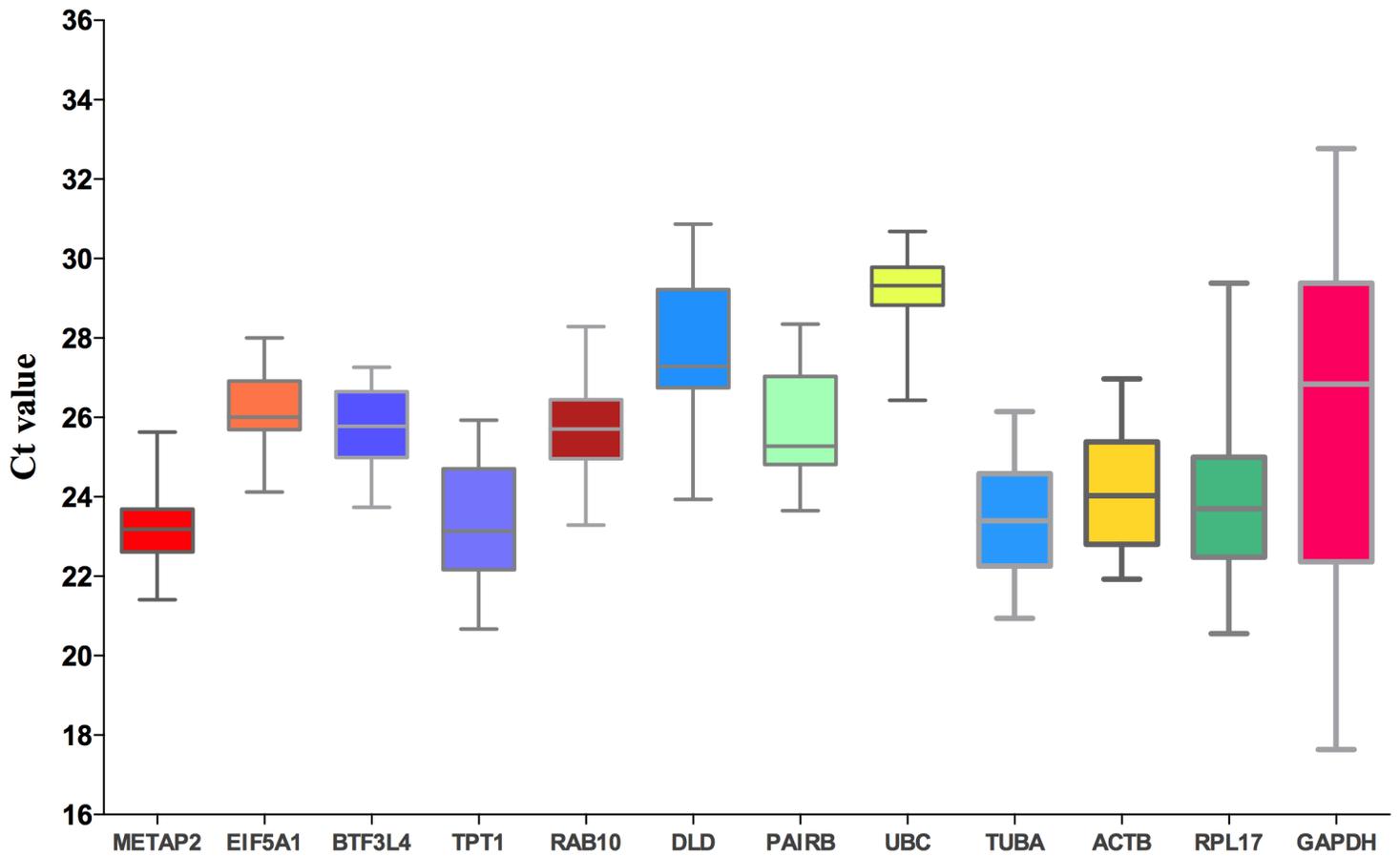


Figure 4

Evaluation of the reference gene candidates and commonly used reference genes based on RT-qPCR analysis. A boxplot showing the Ct values of the 8 candidate reference genes and 4 commonly used reference genes among tissues and developmental stages.

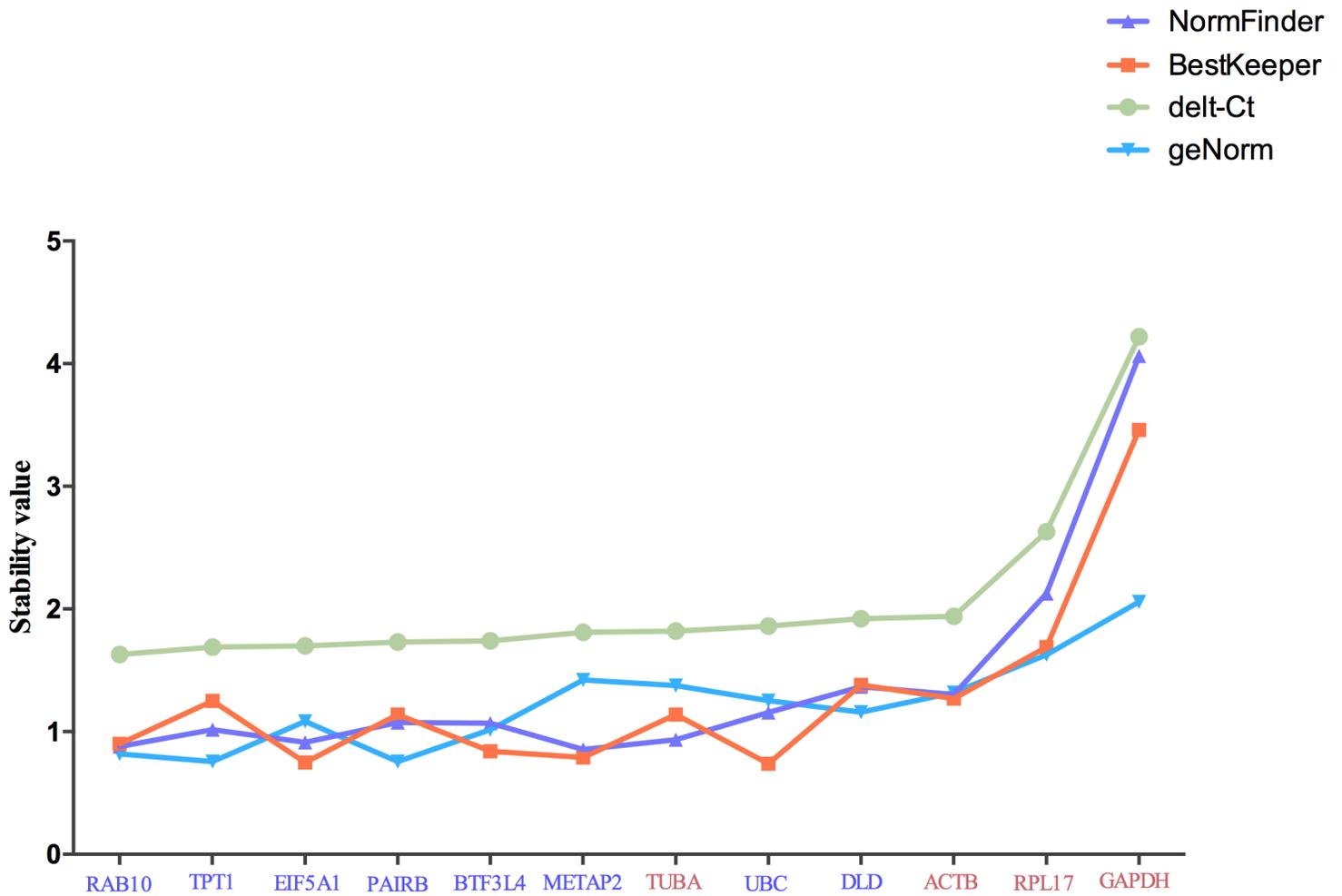


Figure 5

Expression stability of the twelve genes based on RT-qPCR experiments. The stability was evaluated based on geNorm, NormFinder and comparative delta-Ct analyses of the RT-qPCR data.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS3.xls](#)
- [table1.tiff](#)
- [table2.tiff](#)
- [TableS2.xls](#)
- [tableS1.tiff](#)