Original Research

Validation and identification of reference genes in Chinese hamster ovary cells for Fc-fusion protein production

Xiaonan Ma, Ling Zhang, Luming Zhang, Chenglong Wang, Xiaorui Guo, Yu Yang, Lin Wang, Xiangru Li and Ningning Ma

Wuya college of Innovation; College of life science and biopharmaceutics, Shenyang Pharmaceutical University, Shenyang 110016, China

Corresponding author: Ningning Ma. Email: nnm_sy@126.com

Impact statement

In order to reveal potential genotype-phenotype relationship, RT-qPCR reactions are frequently applied which require validated and reliable reference genes. With the investigation on long-term passage and fed-batch cultivation of CHO cells producing an Fc-fusion protein, four new reference genes–Akr1a1, Gpx1, Aprt, and Rps16, were identified from 20 candidates with the aid of geNorm, NormFinder, BestKeeper, and ΔCt programs and methods. This article provided more verified options in reference gene selection in related research on CHO cells.

Abstract

Chinese hamster ovary cells are the predominant cell lines used for bio-therapeutic production. Real-time quantitative PCR (RT-qPCR) and transcriptomics are powerful tools to understand and optimize the Chinese hamster ovary cells for higher productivity or better control of product qualities. Reliable reference genes, which were proved to be experiment-specific, are critical yardsticks. In this study, we compared expression stability of 20 candidate reference genes at mRNA level, including commonly used housekeeping genes, previous literature reported genes in Chinese hamster ovary cells producing an intact antibody, and new candidates suggested by our RNA-seq transcriptomic database, in RT-qPCR reactions in Fc-fusion protein-producing Chinese hamster ovary cells with various productivity during long-term cultivation and fed-batch cultures at 26 different conditions. geNorm, NormFinder, BestKeeper, and ΔCt programs and methods were utilized to analyze the gene expression stability and gave an overall ranking. Akr1a1, Gpx1, and Aprt in long-term cultivation and Akr1a1, Rps16 in fed-batch culture, which have not been reported previously, exhibited the highest stability of gene expression, while Pabpn1, Hirip3, and Actb in both sets of experiments together with Atp5f1 in long-term passage process showed the weakest stability. The results were then validated using GLP1-Fc (Glucagon-like peptide-1 Fc fusion protein) gene as the target with determined expression level which were doubly confirmed by both absolute RT-qPCR and confocal microscopy. These new references should be considered for the investigations on Chinese hamster ovary cells in related research.

Keywords: Chinese hamster ovary cell, reference gene, RNA-seq, real-time quantitative PCR

Experimental Biology and Medicine **2020; 245: 690–702. DOI: 10.1177/1535370220914058**

Introduction

Chinese hamster ovary (CHO) cell has served as the most widely used mammalian expression system for the production of majority of recombinant therapeutic proteins in biopharmaceutical field.[1–3] Significant efforts have been taken to achieve higher yield through optimization of cloning vectors, culture media, cultivation conditions, and genetic engineering of cell lines.[4–6] With the development of omics study, outstanding clones with super producing ability are often isolated and analyzed with transcriptomic, proteomic, and metabolomic approaches which can be used as engineering targets to enhance cell line performance.[7] Correlations between productivity and genetic mechanisms were discovered such as gene copy number and integration site of vector into the chromosome.[8] Subtle variations like the ratio of light chain/heavy chain gene copy number or alteration of amino acids for the same mAb (mono-clonal antibody) can lead to dramatic difference in yield.[9] Meanwhile, efforts were made by some laboratories to reveal certain genes correlated with higher expression level of recombinant protein (r-protein).[2] Still, the anfractuous regulatory networks of

Copyright © 2020 by the Society for Experimental Biology and Medicine

the biological system in CHO cells make the identification of genotype–phenotype relationships highly challenging.

In all omics studies, transcriptomic analysis is quite frequently used which can be performed mainly by RNA-seq and microarrays.[3] RNA-seq is widely applied due to its high sensitivity and specificity. With the availability of the first CHO genome data in 2011[10] and the following updated information,[11] RNA-seq offers insightful comprehension in CHO transcriptomics. When prospective genes emerge, real-time quantitative PCR (RT-qPCR) is often performed for validation after the in-silico detection.[8,12]

RT-qPCR is a sensitive technique to measure the gene expression level where relative quantitation is more frequently used as it does not need to prepare standards of the target gene, but it needs internal control genes, also referred to as reference genes.[13] An ideal reference gene should stably express in all physiological and experimental conditions. However, previous studies showed that reference genes can vary considerably. For instance, the gene expression level of some frequently used housekeeping genes (HKGs), involved in basic cellular functions like actin, ubiquitin and Gapdh, was found to fluctuate depending on tissue type, metabolic process, or different physiological conditions even within the same species.[14–16] Therefore, it was speculated that no universal reference genes existed across all experimental conditions,[17,18] which calls for the validation of reference genes for a specific experiment or a specific cell line. However, unverified internal reference genes were still used directly for normalization.[19] Applicable reference genes for specific CHO cells expressing certain intact antibody were identified in static and fed-batch cultures,[18,20] but Leon's work[9] revealed that even the antibodies bound to the same target with different amino acid sequence in particular regions showed various titers and growth conditions in CHO cells. Therefore, the reference genes in CHO cells producing an Fc-fusion protein which have never been investigated are more likely to be different from the previously identified ones. Production instability of r-protein in CHO cells during long-term cultivation is an unneglected serious problem in industrial production.[21–23] Validated reference genes for long-term passages were never investigated either. Therefore, this calls for the verification study because the previously identified genes might not be suitable for the analyses of Fc-fusion protein during RT-qPCR experiments.

In this study, validation scales of candidate reference genes were expanded to 20 genes including seven commonly used HKGs, seven previously identified ones in mAb-manufacturing CHO cells, and six genes from RNA-seq transcriptomic database in our Fc-fusion protein-producing CHO cell lines with different specific productivities. The experimental conditions involved not only the fed-batch culture with diverse media and culture conditions but also the 75-day-continuous passage process which is often performed to investigate the production stability during cell line development.[21–23] The performance of all candidate reference genes was statistically analyzed with four different algorithms of geNorm,[24] NormFinder,[25] BestKeeper,[26] and a comparative ΔCt method.[27] RefFinder[28] program was used to give the comprehensive rank. The exact gene copy number of GLP1-Fc protein for each cell line was measured and used for further validation of the reference genes with various stabilities at the mRNA level.

Materials and methods

Cell culture and experiment conditions

Six CHO cell lines which were well adapted in a serum-free suspension medium CD-CHOK1 (Vbiosci, Shenzhen, China) were investigated in this study including one parental and five GLP1-Fc-fusion protein-producing strains named as CHO-host, CHO-12, CHO-16, CHO-39, CHO-40, and CHO-69, respectively (CHO-K1 lineage).

For the 75-day-continuous cultivation (long-term cultivation), all cell lines were routinely passaged every threedays at the seeding density of 0.3×10^6 cells/mL with the working volume of 30 mL in a 125-mL shake flask at 37°C, 10% CO_2, and 120 r/min. The media were supplemented with 20 μM MSX (Sigma-Aldrich, St. Louis, MO). Cell viability and viable cell count were determined at the sampling time using the Counter Star (Shanghai RuiYu Biotech, China).

For the fed-batch culture, three cell lines (CHO-12, CHO-16 and CHO-69) which were taken as the highest-, the lowest-, and a middle-yielding producers were analyzed. Seeding density was 0.5×10^6 cells/mL with the working volume of 30 mL in a 125-mL shake flask at 37°C, 10% CO_2, and 120 r/min. The process lasted for 14 days until the viability decreased below 60%. Cell count and viability were measured every day. Different culture media with various supplements were conducted (Supplementary table S1). Incubation at 33°C from day 5 until the end of the fed-batch process was also tested. Samples were harvested at days 2, 4, 6, 8 (2, 4, 6 and 8 incubation days after seeding during every fed-batch culture) which represented the lag phase, middle and late log phase, and stationary phase of the incubation process. All experimental conditions were performed in triplicates.

Specific productivity determination

For the long-term cultivation, culture samples of the subculture day (day 0) and two days after subculture (day 2) in a three-day passage process were collected to calculate the cell-specific productivity (Qp). Samples were harvested every five passages. To determine the recombinant protein titer, 0.2 mL of broth was removed from the flask and centrifuged at 1000 r/min for 5 min. All samples were stored at −70°C until use. ELISA assay (E80-104, Bethyl Laboratories, USA) was performed according to the manufacturer's instructions with the appropriately diluted supernatant sample. A purified GLP1-Fc protein sample was utilized as a standard by the calibration of a purchased Dulaglutide (Lilly, Ireland) using a Nanodrop 2000 Spectrophotometer (Thermo Scientific, Waltham, USA).

Daily specific production rate (Qp in pg/cell/day, which means the amount of target protein that a single cell can

produce every 24 h) was calculated with titer (μg/mL) and viable cell count (10^6 cells/mL) as follows[9]

$$Qp = \left(\frac{T_1 - T_0}{\frac{V_1+V_0}{2}}\right) \div t$$

where T_1 = titer of day 2, T_0 = titer of day 0. V_1 = viable cell count of day 2, V_0 = viable cell count of day 0. t = 2.

Pearson correlation coefficient (ρ) was calculated to evaluate the strength of association between Qp and gene copy number of r-protein.

Immunofluorescence observation

Cells were fixed using 4%PFA for 10 min on a glass slide. After the permeabilization of 0.25% Triton X-100 for 10 min, 1% BSA was added onto the cells for 30 min known as blocking. Cells were then incubated with the 500× diluted antibody (goat-anti human IgG Fc (DyLight® 488), Abcam) solution for 1 h at room temperature in the dark and mounted with a drop of mounting medium (with DAPI, Abcam). Fluorescence observation was carried out by the confocal microscope (Nikon, Japan) under 10×40 magnification.

Selection of candidate genes

The RNA-seq analyses of our six CHO cell lines at day 2 (two days after subculture) in long-term cultivation were performed by Beijing Genomics Institute. All extracted RNA samples had RNA Integrity Number (RIN) >9 and were used for RNA-Seq analysis. Sequencing was performed on BGISEQ-500 platform. Clean reads were aligned to the *Cricetulus griseus* reference genome using Bowtie2.[29] The expression level for each transcript was calculated using RSEM.[30]

Six novel genes with highest gene expression stability in different FPKM (fragments per kb of transcript per million reads mapped) value ranges were selected as part of the candidate reference genes. Coefficient of variation (CV%, standard deviation/mean) of the FPKM values of each gene across six cell lines was calculated for the preliminary ranking (Table 1). The lower the CV value was, the better expression stability the gene exhibited.

Seven commonly used reference genes (the house keeping genes)[17,24,31–37] were also included in the following validation. Another seven testing genes were identified in CHO cells with good expression stability at mRNA level from the previous papers.[18,20] The corresponding variability of the above 14 genes in our transcriptomic data is also listed in Table 1. The gene copy number of GLP1-Fc protein in five r-protein-producing cell lines was measured as well for further validation.

Primer design

All primers were designed following MIQE guidelines[38] with the help of Primer-BLAST software[39] to ensure the specificity of the target amplicons. This was confirmed by PCR, agarose gel electrophoresis, and melting curves of RT-qPCR (Supplementary Figure S1).

Table 1. Confirmation of candidate reference genes.

Gene	Mean FPKM	CV%	MFC
Novel genes from RNA-seq analyses			
Eif3k	315.12	2.88	1.08
Akr1a1	753.30	5.51	1.15
Rps16	3140.85	6.02	1.21
Atp5mg	459.44	6.10	1.17
Aprt	315.27	6.81	1.21
Gpx1	810.72	7.34	1.20
Genes from literatures in CHO cells			
Mmadhc[18]	36.36	8.38	1.25
Fkbp1a[18]	565.20	11.33	1.39
Hirip3[20]	48.27	11.80	1.34
Gnb1[18]	704.76	12.16	1.38
Pabpn1[20]	96.84	12.34	1.48
Eif3i[20]	754.66	14.73	1.56
Actr5[20]	43.62	15.51	1.60
HKGs			
Pgk1[17,31]	500.88	7.74	1.28
Actb[17,24,31–37]	3329.08	12.46	1.39
Atp5f1[31]	153.48	18.05	1.67
Hprt[17,24,32,36,37]	61.54	20.67	1.76
Gapdh[17,31,32,34–37]	8088.57	22.11	2.09
Gusb[35,36]	47.02	22.36	1.84
B2m[17,24,35,36]	128.07	28.59	2.05

Note: Three panels of candidate genes were investigated in this study. "Novel genes from RNA-Seq analyses" were chosen from the transcriptomic data of six CHO-K1 cell lines with highest gene expression stabilities. "Genes from literatures in CHO cells" were the previously identified reference genes in CHO cell lines. "HKGs" were the commonly used housekeeping genes (HKGs) in RT-qPCR analyses. In each panel, genes were ranked according to coefficient of variation (CV%) across six cell lines.
FPKM: fragments per kb of transcript per million reads mapped.

RNA extraction and cDNA synthesis

Cells (5×10^6) were harvested by centrifugation at 1000 r/min for 5 min for each sample. Total RNA was extracted using the TRIzol reagent (ThermoFisher) according to the manufacturer's instructions. RNA purity was estimated by measuring A_{260}/A_{280} and A_{260}/A_{230} ratios using a Nanodrop 2000 Spectrophotometer (Thermo Scientific, Waltham, USA). RNA integrity was examined by agarose gel electrophoresis. cDNA was synthesized with 2 μg of total RNA for each sample using the HiFiScript cDNA Synthesis kit (CWBIO, China) in a total volume of 20 μL according to the manufacturer's instruction.

Real-time quantitative PCR

RT-qPCR was performed using a CFX96 Connect Apparatus (Bio-Rad, Japan). All reactions were done in triplicates containing 5 μL of iQ SYBR green PCR supermix (170–8882, Bio-Rad), 1 μL cDNA, 1 μL/primer with the final concentration of 300 nM, and 2 μL nuclease free water to give a final volume of 10 μL. Reactions with no template or RNA instead of template were used as negative controls. The amplification conditions were: 95°C for 3 min, followed by 40 cycles at 95°C for 5 s, and 60°C for 30 s. The melting curve and resulting amplicons were analyzed with CFX

Manager, version 2.2.1 (Bio-Rad). For each reference gene assay, standard curve was generated by a series of four 10-fold dilutions of cDNA (0.1–100 ng). For the gene expression level of GLP1-Fc protein assay, solutions of a plasmid containing the r-protein sequence were used as standards and for the specific gene copy number estimation. Gene copy number was found to relate to the plasmid molecular weight (DNA length) and concentration (C). It could be calculated as (dsDNA)[40]

$$\text{copy}/\mu\text{L} = \frac{6.02 \times 10^{23} \times C \times 10^{-9}}{DNA\,length \times 660}$$

where *C* is the plasmid concentration (ng/μL). *DNA length* is the plasmid size (bp). A series of seven 10-fold dilutions of GLP1-Fc plasmid solution (10^1–10^7 copies/μL) was applied for the standard curve. Therefore, the gene copy number of the r-protein in the 1 μL cDNA sample of each cell line can be obtained from the standard curve which was generated by plotting the Ct values on a logarithmic scale along with corresponding concentrations. It is a linear regression curve through the data points. The amplification efficiency (E) was determined by the slope of the trend line of the standard curve as follows[26]

$$\text{E}(\%) = [10^{-\frac{1}{slope}} - 1] \times 100\%$$

The linearity was obtained from the correlation coefficients (R^2).

Statistical analysis

Four algorithms were performed to evaluate the gene expression stability of candidate reference genes. For geNorm analysis,[24] the rank of stability was obtained according to the M value of each tested gene. The optimal number of reference genes could also be determined. NormFinder[25] calculated the normalization factor to produce the ranking. BestKeeper program[26] gave the rank using the pair-wise correlation analysis of candidate genes with standard deviation of CP values and coefficient of variance. The comparative ΔCt method[27] estimated the most stable reference genes by comparing the relative expression of "pair of genes" within each sample. RefFinder algorithm[28] calculated the geometric mean of each gene with the four previous algorithms to give the comprehensive ranking.

Results

Different r-protein performance in recombinant-CHO cell lines

In order to identify yield-related molecular mechanisms, transcriptomic analyses on one parental and five Fc-fusion-protein-producing CHO cell lines with various specific productivity (Qp) (CHO-K1 lineage) were investigated and the database was used for partial candidate selection in this study. Qp of each cell line was therefore firstly assessed. Five CHO cell lines expressing an Fc-fusion protein were cultured in 125 mL-shake flasks under typical cell culture conditions. Cells were subcultured every three days and samples were harvested every five passages. Five pairs of samples were collected for each cell line at the end of the 75-day-continuous cultivation. The specific productivity (Qp) and Fc-protein gene copy number (GCN) per milliliter DNA sample for each cell line are shown in Figure 1(a) and (b), respectively. CHO-12 was the highest yielding producer at 10.12 pg/cell/day, whereas CHO-16, CHO-39, and CHO-40 were the lowest ones at around 0.7 pg/cell/day. CHO-69 was the middle yielding producer at 5.3 pg/cell/day. Individual gene copy number of the Fc-fusion protein cDNA sample was calculated from the corresponding RT-qPCR standard curve (E = 100.3%, $R^2 = 0.998$). The highest r-protein producer CHO-12 still possessed

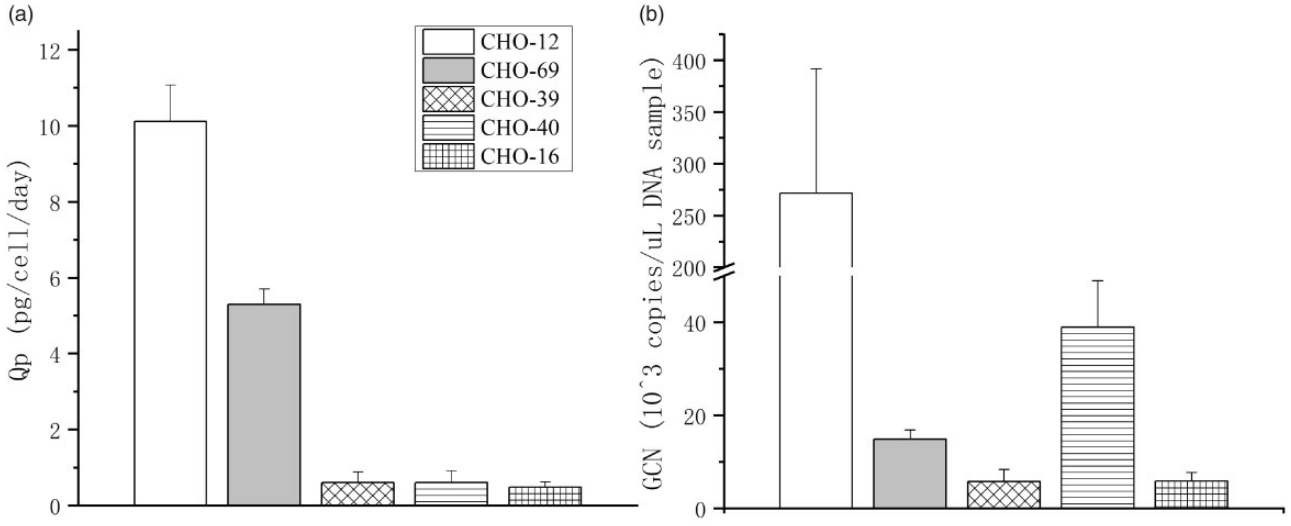

Figure 1. Specific productivity (Qp) and gene copy number (GCN) of GLP1-Fc fusion protein. (a) Qp of each cell line was calculated based on cell numbers and titers at day 0 (sub-culture day) and day 2 (two days incubation after sub-culture). (b) GCN was calculated using the standard curve (E = 100.3%, R2 = 0.998) in absolute RT-qPCR analysis at day 2. Error bars represent the standard deviation of three biological replicates ($P < 0.05$).

the top GCN of 2.71×10^5 copies/μL, whereas CHO-40, a low r-protein producer, had the second highest GCN, and CHO-69, a middle one, had low GCN. Even so, a general correlation of Qp and GCN existed with a correlation coefficient of 0.732 (P value = 0.002).

In order to compare the relationship between Qp and GCN, all data were normalized based on those of CHO-16 with the lowest Qp and GCN values. In Figure 2, the obvious difference in GCN/Qp ratio between the five

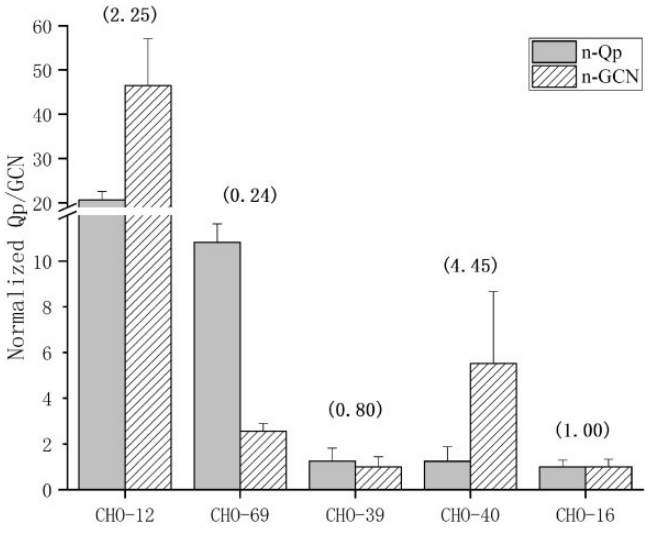

Figure 2. Normalized Qp and GCN. Values of Qp and GCN for each cell line were normalized based on CHO-16 which were measured to be the lowest of both parameters. The GCN/Qp ratio is labeled in brackets above each group of bars. Error bars represent the standard deviation of three biological replicates.

r-protein-producing cell lines revealed that the specific productivity was not always proportional to the GCN. In CHO-12 cell line, the fold changes of Qp were 20.65-fold and that of GCN was 46.4-fold (GCN/Qp = 2.25). One low yielding producer of CHO-40 exhibited similar trends of higher GCN/Qp ratio. However, the middle r-protein producer of CHO-69 seemed to be a more efficient manufacturer with the ratio of 0.24. This proved that GCN was not the only factor responsible for r-protein production.

Confocal microscopy was performed with goat-anti human IgG Fc antibody (Ex: 493 nm, Em: 518 nm) to give a visual display. Figure 3 shows the distribution of extracellular and intracellular GLP1-Fc fusion protein of five cell lines. CHO-12 showed the strongest fluorescence surrounding the cell followed by CHO-69 which were the top two producers in yield. CHO-40 exhibited the strongest fluorescence inside the cell in keeping with the relatively high GCN result. The barely fluorescent peripheral environment was in accordance with its low specific productivity of r-protein. For CHO-39, image of DAPI (Ex: 360 nm, Em: 460 nm) labeled nucleus was shown to indicate the cell position due to weak green luminous intensity. It could be found that the immunofluorescence observation results basically coincided with the Qp and GCN of r-protein measurement.

Absolute GCNs for one CHO-12 fed-batch culture condition at four different incubation time points (days 2, 4, 6 and 8 during the incubation process) were also measured which were 5.45×10^6, 9.08×10^6, 2.93×10^7, and 5.98×10^7 copies/μL DNA sample, respectively. Compared with the long-term cultivation process, gene expression levels of

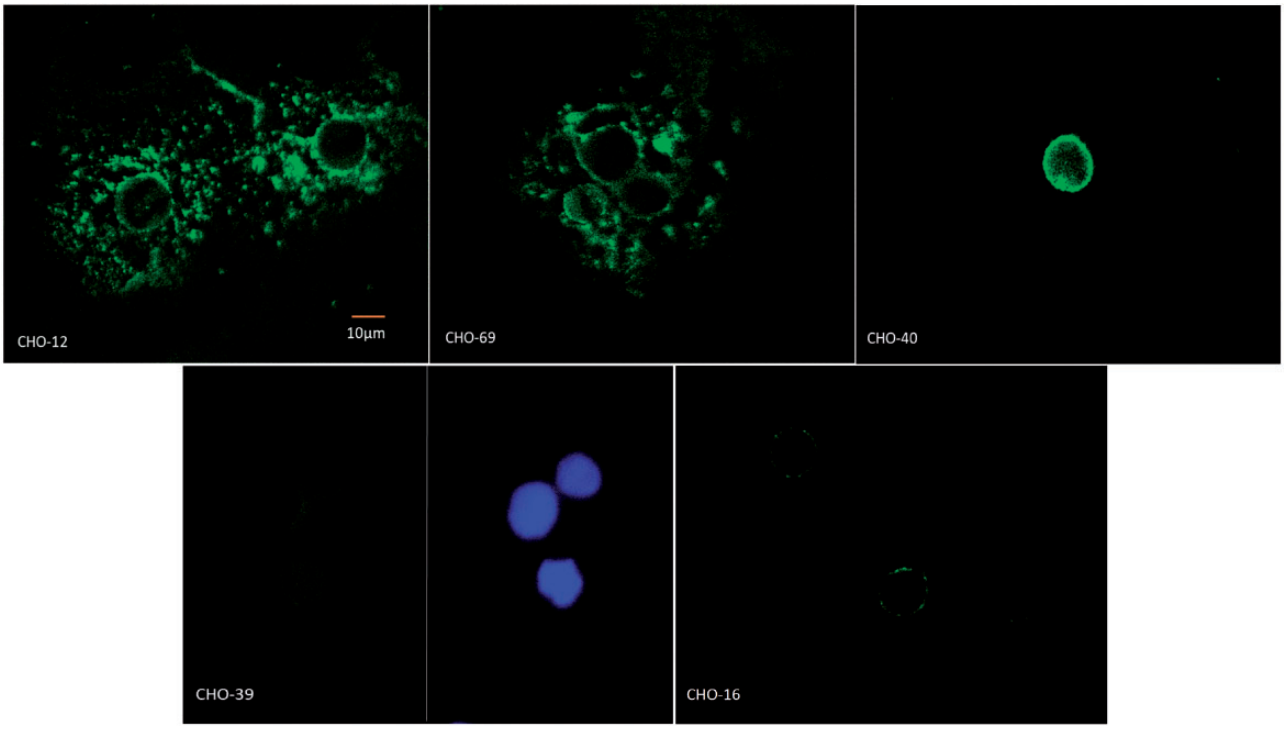

Figure 3. Confocal microscopy (10 × 40 magnification) differentiated the distribution of intracellular and extracellular GLP1-Fc fusion protein of each cell line. The r-protein was marked by goat-anti human IgG Fc antibody (Ex: 493 nm, Em: 518 nm). For CHO-39, DAPI image (nucleus marked) was used to indicate the cell position due to weak green fluorescence.

r-protein were much higher in fed-batch culture and increased with the extension of incubation time. All the GCN data of GLP1-Fc would be used for reference gene validation later.

Selection and gene expression level of candidate reference genes in six CHO cell lines

In order to explore the latent reference genes, three panels of candidates were involved in this study which were six genes from our RNA-seq database (Eif3k, Akr1a1, Rps16, Atp5mg, Aprt, and Gpx1) which exhibited the highest expression stability according to the CV% results of the FPKM values, seven previously identified reference genes in CHO cells (Mmadhc, Fkbp1a, Hirip3, Gnb1, Pabpn1, Eif3i, and Actr5),[18,20] and seven commonly used housekeeping genes (Pgk1, Actb, Atp5f1, Hprt, Gapdh, Gusb, and B2m)[17,24,31–37] in RT-qPCR analyses, respectively. Coefficient of variation (CV%) of the FPKM values of each gene in our RNA-seq database was calculated to rank the expression stability preliminarily. Maximum fold change (MFC) obtained from the division of the highest and smallest FPKM values across six cell lines was also taken into account. Genes with different expression abundance were contained and investigated to provide reference for the research on target genes with various expression levels, since an ideal reference gene should share a similarity in the expression level with the detection gene.

From Table 1, gene expression level of the novel six genes in the first panel displayed higher stability than the other two panels, having CV% of 2.88–7.34% and MFC of 1.08–1.21. Five out of the seven HKGs (Atp5f1, Hprt, Gapdh, Gusb, and B2m) exhibited the most fluctuant situation (CV%: 18.05–28.59%, MFC: 1.67–2.09) at transcriptional level. The other two HKGs (Pgk1 and Actb) expressed more steadily and Pgk1 was surpassed only by the first panel of candidates. The seven previously identified genes possessed CV% of 8.38–15.51% and MFC of 1.25–1.60.

Performance of RT-qPCR assay was evaluated by the amplification efficiency and R^2. As shown in Table 2, efficiencies ranged from 90.5% to 103.5% and $R^2 > 0.98$ which were both in acceptable scopes.

Ct values of all candidate reference genes across long-term cultivation and fed-batch cultivation processes are plotted in Figure 4(a) and (b). In the long-term cultivation process, the average of each gene varied from 18.06 (Rps16) to 32.49 (Hirip3). Variable expression abundance of these genes was found that Rps16, B2m, Atp5mg, Gapdh, and Gpx1 exhibited high gene expression level with median below 20, while Eif3k, Hirip3, and Pabpn1 showed low expression abundance with Ct median above 30. Standard deviation and coefficient of variation of the Ct values for each gene across all cell lines were calculated to give a preliminary assessment of the stability at transcriptional level (data not shown). The top five genes with smallest CV% were Aprt, Eif3k, Gnb1, Atp5mg, and Hprt, respectively, with the value no more than 4%. Atp5f1, Pabpn1, Gapdh, and Actb were revealed to be the most variable ones at gene expression level with CV% ranging from 9.02 to 13.14%.

In the fed-batch process, the gene expression abundance was in the range of Ct values from 13.64 (Gapdh) to 30.23 (Eif3k). B2m and Atp5mg possessed the most stable expression at mRNA level with the CV% values less than 4%. Gapdh, Hirip3, Actb, and Pabpn1 showed the weakest stability with the CV% higher than 9%.

Four widely used algorithms (geNorm, NormFinder, BestKeeper, and ΔCt method) would be applied to analyze the gene expression stability of all candidate genes to rank comprehensively.

Estimation of candidate reference genes stability

geNorm analysis. geNorm analysis ranked the reference genes according to the expression stability (M value) based on logarithmically transformed expression ratios from the most variable to the most stable two genes by step-wise elimination of the least stable one. M, referred as the expression stability measure, was obtained by calculating the average pair-wise variation of a particular gene with all other genes. The genes with M value higher than 1.5 will not be appropriate for ideal reference genes due to high expression fluctuation. Two most stable genes were given ultimately which cannot be further ranked.[24] This applet can also determine the minimum number of reference genes needed for accurate calibration in RT-qPCR reactions. As depicted in Figure 5(a), Akr1a1/Gpx1 > Aprt > Atp5mg > Hprt were identified as the most stably expressed five genes across six CHO cell lines in long-term cultivation process. Whereas, Pabpn1 > Atp5f1 > Hirip3 > Fkbp1a were considered as the most variable genes at mRNA level with high M value above 1.5. In fed-batch cultures (Figure 5(c)), Atp5mg, Gpx1, and Akr1a1 were still in the top five positions but in different orders. Rps16 and B2m also showed good expression stability. Gapdh and Actb which were sometimes taken as "gold standards"[41] were found to be strongly fluctuant at mRNA level during the fed-batch culture. Final ranking result of geNorm algorithm is shown in Tables 3 and 4. Pairwise variation recommended that in our study the most stable genes–Akr1a1, Gpx1, and Aprt in long-term passage and Atp5mg together with Rps16 in fed-batch culture should be used as the minimal reference genes for qPCR normalization with 0.15 as a cut-off value.[24]

NormFinder analysis. NormFinder evaluated the gene expression stability with a model-based approach. It analyzed the candidate separately of both the intra- and the inter-group expression variation. Similar to geNorm, the input data of NormFinder were also logarithmically transformed. Stability value of each candidate gene was calculated to give the ranking. The lower the value was, the better stability the gene expressed.[25] As presented in Table 3, Akr1a1 > Mmadhc > Eif3i > Hprt > Gpx1 were rated as the most stably expressed genes in the long-term cultivation. During the fed-batch culture (Table 4), Akr1a1, Atp5f1, and Rps16 were ranked as the most stable ones. The most variable genes in both sets of experiments were the same as corresponding geNorm results.

Table 2. Primer sequences and performance of candidate reference genes in RT-qPCR analyses.

Gene	Accession No.	Primer sequences (5′-3′)	E(%)	R^2	Size (bp)
Eif3k	XM_007644386.2	AGCCCAAAAACATCGTGGAGA TCATTGAGGAAGGGGCAGAAG	103.4	0.994	134
Akr1a1	XM_003497372.4	GGCTTGGAGGTGACTGCTTAT GAACCTGCCATCTGAGCAAGA	97.6	0.999	154
Rps16	XM_003503853.3	TGAAGGGTGGTGGTCATGTG TCAGCTACAAGCAGGGTTCG	90.5	0.994	154
Atp5mg	XM_007651137.2	ATGGCCAAGTTCGTCCGTAA GGGTTGGGGGAAACAGTTCA	94.7	0.995	127
Aprt	XM_003495090.2	CTCCTTCCGAGCTTCCATCC CTAGGGAGGGGCCAAACAAG	91.8	0.992	116
Gpx1	NM_001256788.1	CGGACATCAGGAGAATGCCA GTAAAGAGCGGGTGAGCCTT	97.7	0.988	138
Mmadhc	XM_003513988.2	TGTCACCTCAATGGGACTGC CAGGTGCATCACTACTCTGAAAC	98.4	0.994	145
Fkbp1a	XM_003499952.2	CTCTCGGGACAGAAACAAGC GACCTACACTCATCTGGGCTAC	98.4	0.997	95
Hirip3	XM_027441928.1	CGTTATATTCGGGCTTGTGG GGTCGACCTGAACTGCTGAT	95.4	0.995	224
Gnb1	NM_001246701.1	CCATATGTTTCTTTCCCAATGGC AAGTCGTCGTACCCAGCAAG	101.9	0.994	184
Pabpn1	XM_027390612.1	GTGGCCATCCTAAAGGGTTT CGGGAGCTGTTGTAATTGGT	98.4	0.996	205
Eif3i	XM_027399389.1	CCACAACTTCCACCAGGATT ATGCGGACGTAACCATCTTC	95.0	0.991	166
Actr5	XM_027421238.1	CCAGAATGGAGAAGGAGCTG GGCACAATGTTCCTTGAGGT	97.4	0.992	197
Pgk1	NM_001246725.1	CAAGGGGAACCAAGTCCCTC GAGCTCTAGACTGGCACCAC	101.2	0.999	161
Actb	NM_001244575.1	CTGTGCTATGTTGCCCTGGA GCCACAGGATTCCATACCCAG	93.2	0.990	174
Atp5f1	XM_027393891.1	GAGCACATGATGGACTGGGT TGCCTGTTTAACTCTCTCTGGC	103.5	0.992	180
Hprt	XM_027432079.1	TACCTCACCGCTTTCTCGTG TCACTAATCACGACGCTGGG	95.9	0.998	124
Gapdh	NM_001244854.2	TGTAAAGCTCATTTCCTGGTATGAC TGTGGGGGTTATTGGACAGG	97.6	0.990	178
Gusb	XM_027413803.1	TTGGTGCCAACTCCTTTCGT TTGTCCCTGCGTACCAGTTC	92.1	0.998	187
B2m	NM_001246674.2	TGGCTCACACGGAGTTTACA CATGTCTCGTTCCCAGGTGA	98.4	0.999	104

Note: Amplification efficiency and R^2 were calculated from the standard curves made of a series of four 10-fold dilutions of cDNA.

ΔCt method. The comparative ΔCt method estimated the expression stability by calculating the average SD of ΔCt between every two samples of each gene.[27] Similar as the other programs, lower value indicated better gene expression stability. Mmadhc > Akr1a1 > Gpx1 > Atp5mg > Aprt were designated as the most stable genes at expression level in long-term passage experiment (Table 3). Hprt was newly determined as one of the least stably expressed genes together with Fkpb1a, Actb, Atp5f1, and Pabpn1. In the fed-batch process, Akr1a1, Atp5mg, Gnb1, and Rps16 showed the highest expression stability. No matter which analysis program and culture process was performed, Actb, Hirip3, Atp5f1, and Pabpn1 were all at the ranking bottom indicating the high expression variation at transcriptional level.

BestKeeper. BestKeeper analyzed the gene expression stability based on the coefficient of correlation to the BestKeeper Index, which is the geometric mean of CP values of candidate reference genes.[26] Variations of gene expression, displayed as the standard deviation (SD) of the CP values, were determined. The smaller the SD value was, the more stably the gene expressed. Candidate with SD > 1 was considered as an inappropriate internal reference gene. Due to the input limitation of BestKeeper, the software tool can only process 10 reference genes at most at a time. We divided the 20 candidate genes into two groups according to the ranking orders of the geNorm, NormFinder, and ΔCt method results. By calculating the geomean of the ranking places in the above three programs, top 10 genes (assigned as group 1) and the rest 10 genes (group 2) were analyzed respectively to give the first ranking. We then used BestKeeper tool to analyze the bottom five genes in group 1 and top five genes in group 2 again to give the second ranking. Final ranking of BestKeeper was made by the combination of the above two ranking results (top five genes in group 1 from the first ranking followed by the 10 genes from the second

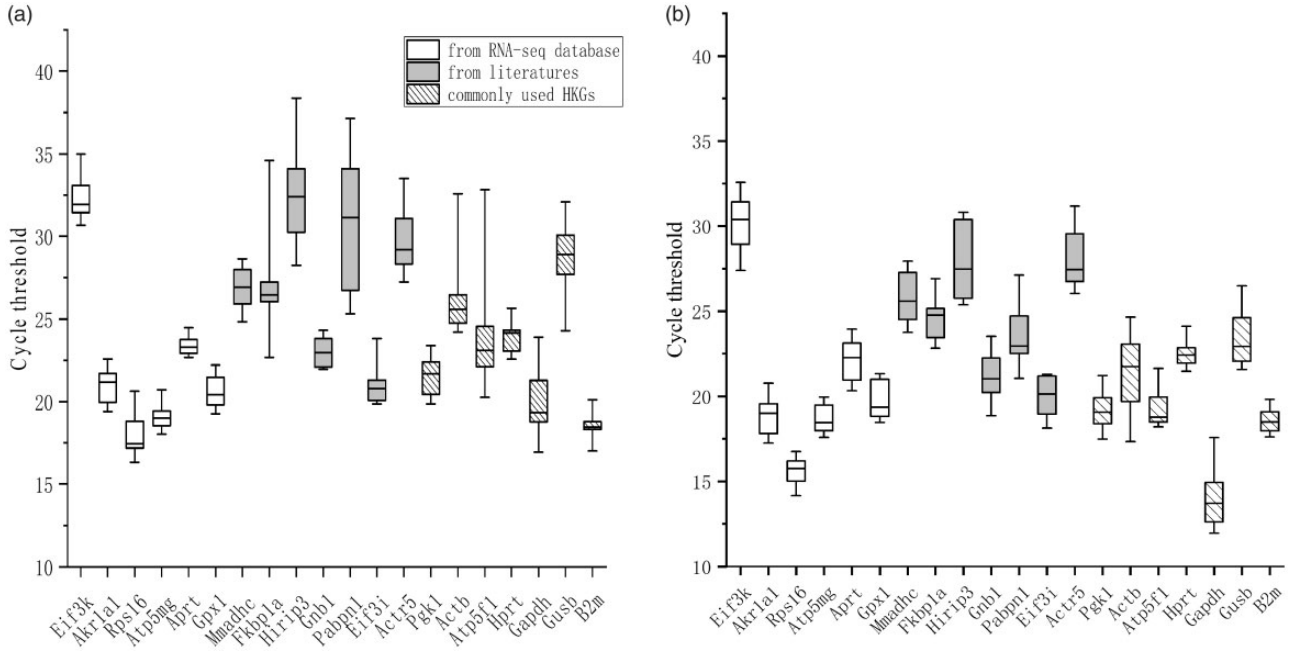

Figure 4. Box-and-whiskers plots of Ct values of candidate reference genes in RT-qPCT reactions. For each plot, lowest value, lower quartile, median, upper quartile, and highest value were indicated by the horizontal lines. (a) Ct values of candidate genes in long-term cultivation. (b) Ct values of candidate genes in fed-batch cultures. All experiments were performed in triplicates.

ranking and the bottom five genes in groups 2 from the first ranking). In Table 3, Aprt > B2m > Atp5mg > Eif3i > Gnb1 were identified as the most stably expressed genes in the long-term cultivation. Fkbp1a was taken the place by Actr5 in the most variable group. In the fed-batch culture (Table 4), B2m, Rps16, Atp5mg, and Akr1a1 showed the best gene expression stability and Hirip3, Actb, Pabpn1, and Gapdh were found to be most fluctuant again like geNorm and NormFinder results.

In order to give a comprehensive ranking of the 20 candidate genes and identify the most stably expressed ones at the mRNA level, RefFinder was applied to calculate the geometric mean from the previous four algorithms of each gene (Tables 3 and 4). Four newly identified reference genes (Akr1a1 > Gpx1 > Aprt in long-term cultivation and Akr1a1> Rps16 in fed-batch culture) exhibited the most stable expression. Following the optimal number of reference genes recommended by geNorm, of a combination of these three and two genes, respectively, should be used for the normalization in RT-qPCR studies. Hirip3, Pabpn1, and Actb were marked as the least stably expressed genes at transcriptional level in the Fc-fusion protein-producing CHO cells.

Impact of different reference genes on target gene expression levels in relative quantitative RT-qPCR reaction

In order to evaluate the performance of different reference genes during the relative quantitative RT-qPCR, previously assessed gene expression level (result of "Different r-protein performance in recombinant-CHO cell lines") of GLP1-Fc protein in five CHO cell lines was used as the target in the long-term cultivation process. Three top ranked genes (Akr1a1, Gpx1, and Aprt) and two bottom ranked genes (Atp5f1 and Pabpn1) were compared. The fold changes of gene expression levels of the r-protein in five CHO cell lines to that of CHO-16 (the lowest GCN value) were calculated by the 2ΔΔCt method. The GCNs of GLP1-Fc protein (obtained by absolute quantification method) across five cell lines were used to obtain the directly calculated fold change indicated as "DC" in Figure 6(a). It could be found that the three most stably expressed genes exhibited similar fold changes to DC across all five cell lines. Atp5f1 and Pabpn1 displayed significant variation results.

Similar validation was performed on the fed-batch culture. GCN data of one CHO-12 fed-batch culture were used as the study object. The fold changes of gene expression levels of the r-protein at four time points to that of day 2 (two days incubation after the seeding day) were calculated by the 2ΔΔCt method. DC was still obtained by computation from the absolute GCNs. The most stably expressed genes—Akr1a1 and Rps16 showed according trends with DC, while the least stable genes—Pabpn1 and Hirip3 gave the opposite numerical relationship. It could be found that normalization to Rps16 yielded better results than using Akr1a1 in Figure 6(b). Here is just one fed-batch culture. Akr1a1 was rated as the most stably expressed reference gene across all the fed-batch culture conditions.

Discussion

CHO cell has been of great interest in biopharmaceutical field for decades due to its high capacity and human-like post-translational modification on recombinant protein. Even significant progress in production has been made,

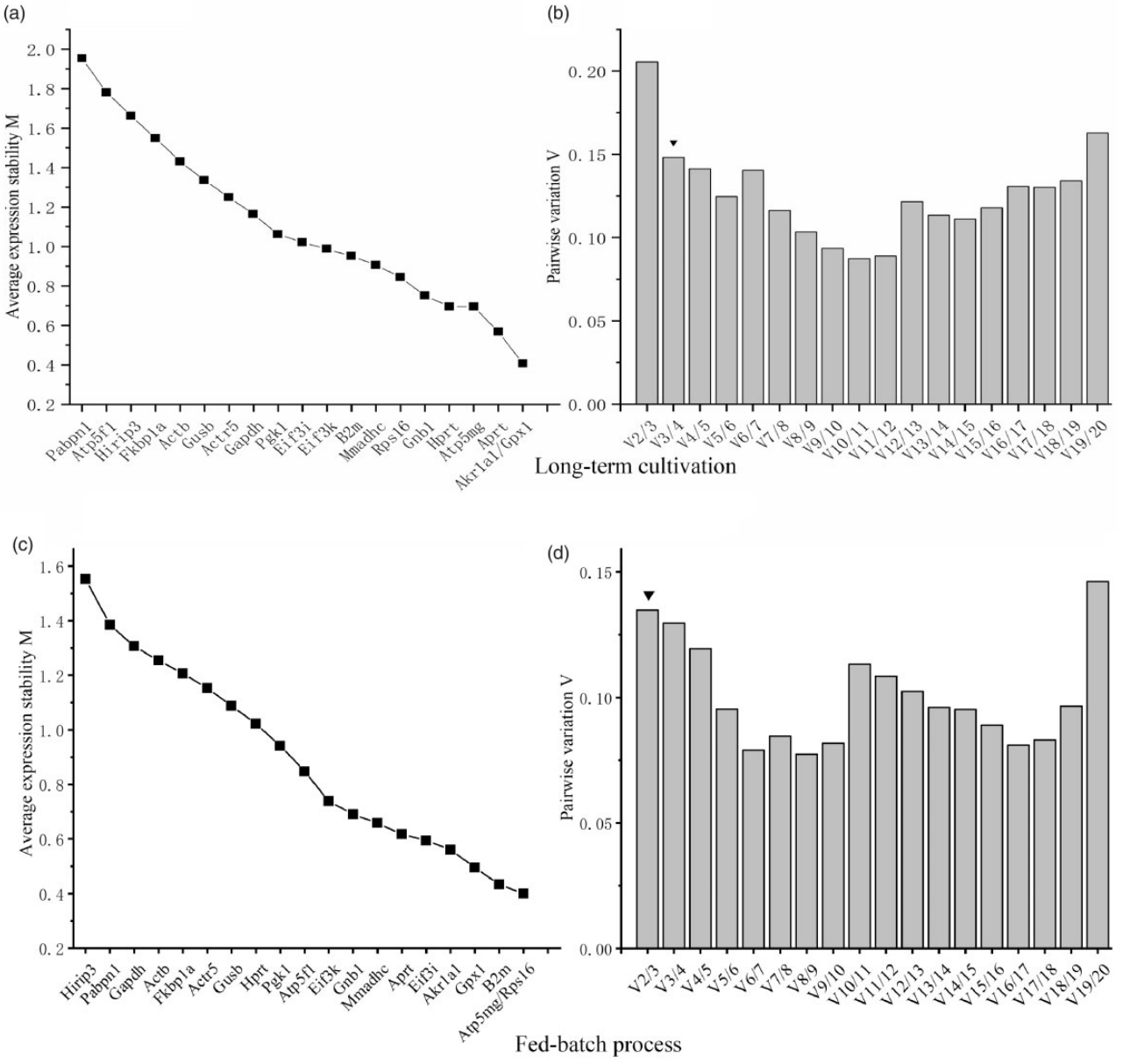

Figure 5. geNorm analysis results. (a) and (c) M values (Average expression stability) calculated by step-wise elimination of the least stable gene indicated the stability of reference genes. (b) and (d) Pairwise variation ($V_{n/n+1}$) calculated from the normalization factor indicates the number of reference genes required for qPCR normalization. Here, $V_{3/4}$ ($V = 0.148$) for long-term cultivation and $V_{2/3}$ ($V = 0.134$) for fed-batch culture revealed that three and two reference genes should be used for calibration, respectively.

the absolute yield is still low compared with prokaryotic *Escherichia coli* and eukaryotic *Saccharomyces cerevisiae* expression systems. Therefore, biological mechanisms related to productivity are highly concerned. Cell lines with various productivities are no doubt to be ideal objects of study.

Recombinant protein production is a complex process which involves multiple biological aspects, such as transcription, translation, metabolism, and secretion.[42,43] That is why in our experiment result, the GCN (transcriptional level) was found not to be proportional to theyield all the time. Relationship between phenotype of high yield and genotype of related genetic mechanisms was investigated a lot. When researching at genetic level, relative quantitative RT-qPCR is one of the most frequently used methods for its high sensitivity and accuracy which needs calibration by the reference genes. It is proposed that reference genes are experiment-specific, and the use of inappropriate reference gene can lead to opposite conclusion. Several reference genes were identified in CHO cells producing an intact antibody in previous research,[18,20] but the ones in Fc-fusion-protein-producing CHO cells have never been reported.

Based on our analyses, out of the 20 candidates, Akr1a1, Gpx1, and Aprt in the long-term cultivation and Akr1a1 and Rps16 in the fed-batch culture surpassed the rest genes in expression stability at the mRNA level. Akr1a1, as a member of the aldo-keto reductase superfamily, was ranked as the most stably expressed gene in all conditions. It is involved in the reduction of aldehydes to

Table 3. Ranking of expression stability analyzed by four algorithms in the long-term cultivation.

	GeNorm		NormFinder		ΔCt		BestKeeper		Overall ranking	
Rank	Gene	M value	Gene	Stability	Gene	Mean SD	Gene	SD	Gene	geoMean
1	Akr1a1[1]/Gpx1[1]	0.41	Akr1a1	0.47	Mmadhc	1.11	Aprt	0.45	Akr1a1	2.00
2		0.41	Mmadhc	0.51	Akr1a1	1.19	B2m	0.55	Gpx1	3.08
3	Aprt[1]	0.57	Eif3i	0.53	Gpx1	1.21	Atp5mg	0.56	Aprt	3.20
4	Atp5mg[1]	0.62	Hprt	0.55	Atp5mg	1.25	Eif3i	0.69	Mmadhc	3.64
5	Hprt[3]	0.70	Gpx1	0.57	Aprt	1.25	Gnb1	0.73	Atp5mg	4.12
6	Gnb1[2]	0.75	Atp5mg	0.61	Pgk1	1.25	Gpx1	0.73	Eif3i	5.51
7	Rps16[1]	0.84	Aprt	0.61	Eif3i	1.27	Hprt	0.78	B2m	6.67
8	Mmadhc[2]	0.91	Pgk1	0.62	Eif3k	1.27	Akr1a1	0.89	Gnb1	7.02
9	B2m[3]	0.95	Gnb1	0.68	Gnb1	1.29	Pgk1	0.94	Hprt	7.18
10	Eif3k[1]	0.99	B2m	0.68	Gapdh	1.34	Eif3k	0.96	Pgk1	8.49
11	Eif3i[2]	1.02	Eif3k	0.72	B2m	1.36	Mmadhc	0.96	Eif3k	9.69
12	Pgk1[3]	1.06	Rps16	0.75	Rps16	1.39	Rps16	1.08	Rps16	10.49
13	Gapdh[3]	1.16	Gapdh	0.89	Gusb	1.51	Fkbp1a	1.28	Gapdh	12.40
14	Actr5[2]	1.25	Actr5	0.93	Actr5	1.51	Gapdh	1.56	Gusb	14.47
15	Gusb[3]	1.34	Gusb	1.08	Hirip3	1.76	Gusb	1.58	Actr5	14.48
16	Actb[3]	1.43	Actb	1.21	Fkpb1a	1.81	Actr5	1.65	Fkbp1a	15.66
17	Fkbp1a[2]	1.55	Fkbp1a	1.56	Actb	1.88	Actb	1.69	Actb	16.49
18	Hirip3[2]	1.66	Hirip3	1.58	Atp5f1	2.40	Hirip3	2.02	Hirip3	17.20
19	Atp5f1[3]	1.78	Atp5f1	1.74	Hprt	2.97	Atp5f1	2.23	Atp5f1	18.74
20	Pabpn1[2]	1.95	Pabpn1	2.25	Pabpn1	3.02	Pabpn1	3.30	Pabpn1	20.00

Note: Superscript of the genes presents different panels of candidate reference genes. 1 = RNA-seq database, 2 = genes from literatures, 3 = HKGs.

Table 4. Ranking of expression stability analyzed by four algorithms in the fed-batch process.

	GeNorm		NormFinder		ΔCt		BestKeeper		Overall ranking	
Rank	Gene	M value	gene	Stability	Gene	Mean SD	Gene	SD	Gene	GeoMean
1	Atp5mg[1]/Rps16[1]	0.40	Akr1a1	0.39	Akr1a1	1.14	B2m	0.47	Akr1a1	2.11
2			Atp5f1	0.42	Atp5mg	1.22	Rps16	0.61	Rps16	2.21
3	B2m[3]	0.43	Rps16	0.51	Gnb1	1.22	Atp5mg	0.62	Atp5mg	2.63
4	Gpx1[1]	0.50	Eif3i	0.52	Rps16	1.22	Akr1a1	0.82	B2m	4.55
5	Akr1a1[1]	0.56	Gnb1	0.55	Eif3i	1.23	Hprt	1.84	Eif3i	5.73
6	Eif3i[2]	0.59	Aprt	0.58	Aprt	1.23	Atp5f1	1.88	Gnb1	6.21
7	Aprt[1]	0.62	Pgk1	0.58	Pgk1	1.26	Pgk1	1.89	Aprt	7.09
8	Mmadhc[2]	0.66	Atp5mg	0.60	Gpx1	1.32	Gpx1	1.92	Atp5f1	7.17
9	Gnb1[2]	0.69	Gusb	0.63	Gusb	1.34	Eif3i	1.99	Gpx1	7.74
10	Eif3k[1]	0.74	Eif3k	0.65	Mmadhc	1.38	Aprt	2.10	Pgk1	8.01
11	Atp5f1[3]	0.85	B2m	0.67	Hprt	1.39	Gnb1	2.31	Hprt	9.62
12	Pgk1[3]	0.94	Hprt	0.68	Actr5	1.40	Mmadhc	2.33	Mmadhc	10.95
13	Hprt[3]	1.02	Actr5	0.70	B2m	1.40	Fkbp1a	2.36	Gusb	11.42
14	Gusb[3]	1.09	Gpx1	0.75	Fkbp1a	1.54	Eif3k	2.45	Eif3k	12.60
15	Actr5[2]	1.15	Mmadhc	0.79	Gapdh	1.61	Gusb	2.56	Actr5	13.91
16	Fkbp1a[2]	1.21	Actb	0.85	Pabpn1	1.77	Actr5	2.86	Fkbp1a	14.92
17	Actb[3]	1.25	Fkbp1a	0.87	Actb	1.87	Gapdh	2.97	Gapdh	16.95
18	Gapdh[3]	1.31	Gapdh	1.02	Eif3k	2.34	Pabpn1	3.27	Actb	17.22
19	Pabpn1[2]	1.38	Pabpn1	1.11	Hirip3	2.56	Actb	3.49	Pabpn1	17.96
20	Hirip3[2]	1.55	Hirip3	1.71	Atp5f1	4.77	Hirip3	6.66	Hirip3	19.75

Note: Superscript of the genes presents different panels of candidate reference genes. 1 = RNA-seq database, 2 = genes from literatures, 3 = HKGs.

corresponding alcohols[44,45] as well as the reduction of D-glucuronic acid in the biosynthesis of ascorbic acid.[46,47] Gpx1, the second most stable gene in long-term cultivation, is coding for glutathione peroxidase 1 and is responsible for the balance of intracellular redox environment.[48] Neither of these two genes had been reported to use as reference genes in RT-qPCR analyses before. It might be because they are not stably expressed in the conditions of most reference gene identification situations, such as in various cell lines or living organisms without any treatments or at different stress conditions. [13,15,25,27,31,49,50] It was proved that Akr1a1 and Gpx1 expressed stably and were suitable in the calibration of target genes in our Fc-fusion protein-producing CHO cells. Aprt, playing an important role in the purine salvage pathway,[51,52] was in the third place of stability analyses in the continuous passages. This result is in agreement with previous work on reference genes in sugarcane.[53,54] Rps16, encodes a ribosomal protein component of the 40S

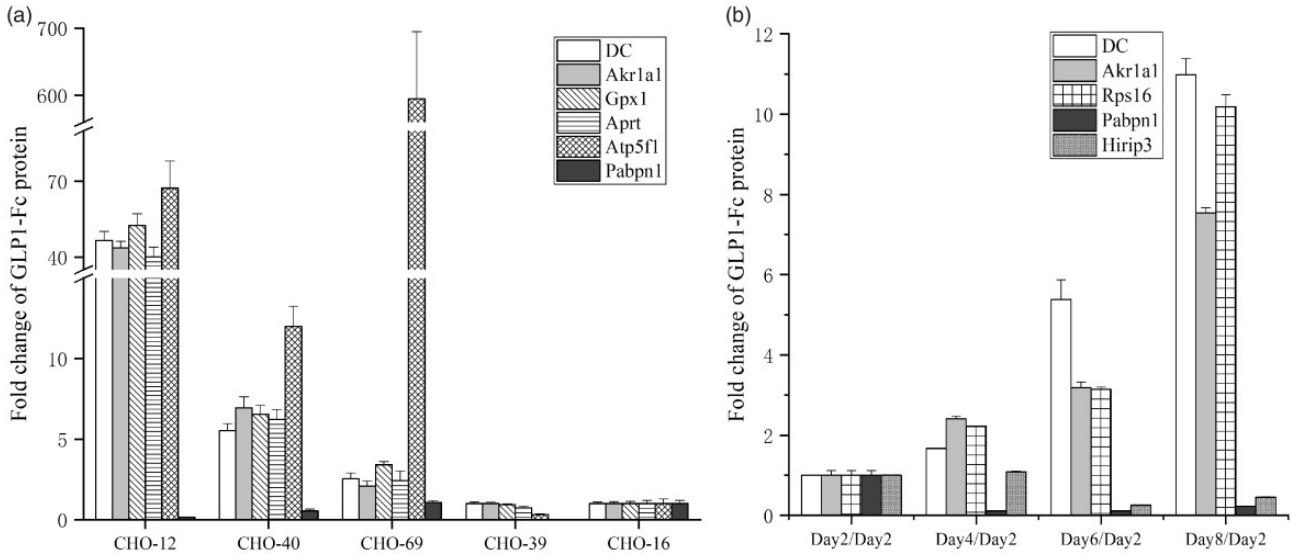

Figure 6. Relative quantification of GLP1-Fc-protein gene expression calibrated by reference genes with different expression stabilities at mRNA level. (a) Long-term cultivation. The fold change of r-protein at transcript level was calculated by the comparison of Ct values of five cell lines to that of CHO-16 with the lowest GCN for each reference gene. (b) Fed-batch culture. The fold change of transcripts of r-protein was calculated by the comparison of Ct values of days 2, 4, 6, 8 to that of day 2 for each reference gene. "DC" represents "directly calculation" from the absolute GCN in the result of "Different r-protein performance in recombinant-CHO cell lines." Error bars represent the standard deviation of three biological replicates.

ribosome subunit, is identified as the second most stable gene in the fed-batch culture. It belongs to the S9P family of ribosomal proteins[55] and showed stable gene expression in certain insects.[56] It is recommended by geNorm program that normalization with the combination of different genes is preferable for RT-qPCR analysis.[50] The involvement in various biological metabolism pathways would more likely to avoid gene co-regulation when used as reference.

In this study, 5 (Akr1a1, Gpx1, Aprt, Atp5mg, and Gnb1 in long-term cultivation process and Akr1a1, Rps16, Atp5mg, Eif3i and Aprt in fed-batch culture) out of the top 10 stably expressed candidate reference genes obtained by the four methods overlapped. Discrepancies in the ranking orders by the four methods might have occurred because of the different computational principles followed. Thus, a comprehensive analysis of the multiple methods was further performed to reduce errors in screening optimal internal reference genes by a single method.

On the contrary, Pabpn1, Atp5f1, and Hirip3 showed great variation at gene expression level across different experimental conditions. Pabpn1 and Hirip3 were identified as the least stable genes which showed opposite result to Bahr's work.[20] Pabpn1 is involved in multi metabolic pathways including translation initiation together with eIF4F, deadenylation, and stabilization of mRNA.[57,58]Hirip3, coding for a nuclear phosphoprotein involved in some aspects of chromatin and histone metabolism.[59,60] The difference in gene expression stability in CHO cells between the production of an intact antibody and an Fc-fusion protein indicated the necessity to verify the reliable reference genes for specific experiments. Atp5f1, a necessary component of ATP synthase, was marked as another most variable gene with the opposite result to some other studies.[13] Inference might be given that these genes are fluctuant during the progress of Fc-fusion protein production.

Impact of using genes with different stabilities at the mRNA level as reference on GLP1-Fc gene analyses was also assessed. It was obvious that inappropriate or even opposite trend in gene expression can lead to unstable reference genes. Accordingly, validation of suitable reference genes before every gene expression analysis experiment was proved indispensable. Due to the fact that no universal reference genes exist, this study provided more applicable reference genes for the research on CHO cell lines.

In conclusion, high stability at transcriptional level of reference genes is the prerequisite in the analyses of relative gene expression level when performing RT-qPCR. This work identified four new reference genes–Akr1a1, Gpx1, Aprt for long-term cultivation, Akr1a1 and Rps16 for fed-batch culture, as the most stable ones at the expression level which were all from the transcriptomic database of our CHO cells during the normal passage and verified in both the long-term and fed-batch cultivation processes. It will facilitate precise gene expression analysis in the CHO cells expressing an Fc-fusion protein.

Authors' contributions: All authors participated the review of the manuscript. NM, and XM designed the experiments. XM, LZ, LZ, and CW performed cell culture and ELISA. XM and LZ performed RT-qPCR. XM, XG, YY, LW and XL analyzed the data. XM wrote the manuscript. NM coordinated its revision. NM, and XM acquired the funds support. All authors have read and approved the final version.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Education Department of Liaoning Province [grant number 201610163L24] and Science and Technology Planning Project of Liaoning Province [grant number 2014226034].

ORCID iD

Ningning Ma https://orcid.org/0000-0001-5986-6632

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

1. Kaneyoshi K, Kuroda K, Uchiyama K, Onitsuka M, Yamano-Adachi N, Koga Y, Omasa T. Secretion analysis of intracellular "difficult-to-express" immunoglobulin G (IgG) in Chinese hamster ovary (CHO) cells. *Cytotechnology* 2019;**71**:305–16
2. Fischer S, Handrick R, Otte K. The art of CHO cell engineering: a comprehensive retrospect and future perspectives. *Biotechnol Adv* 2015;**33**:1878–96
3. Stolfa G, Smonskey MT, Boniface R, Hachmann A-B, Gulde P, Joshi AD, Pierce AP. CHO-Omics review: the impact of current and emerging technologies on Chinese hamster ovary based bioproduction. *Biotechnol J* 2018;**13**:17002–27
4. Datta P, Linhardt RJ, Sharfstein ST. An 'omics approach towards CHO cell engineering. *Biotechnol Bioeng* 2013;**110**:1255–71
5. Kim JY, Kim YG, Lee GM. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl Microbiol Biotechnol* 2012;**93**:917–30
6. Kramer O, Klausing S, Noll T. Methods in mammalian cell line engineering: from random mutagenesis to sequence-specific approaches. *Appl Microbiol Biotechnol* 2010;**88**:425–36
7. Shridhar S, Klanert G, Auer N, Hernandez-Lopez I, Kandula MM, Hackl M, Borth N. Transcriptomic changes in CHO cells after adaptation to suspension growth in protein-free medium analysed by a species-specific microarray. *J Biotechnol* 2017;**257**:13–21
8. Fomina-Yadlin D, Mujacic M, Maggiora K, Quesnell G, Saleem R, McGrew JT. Transcriptome analysis of a CHO cell line expressing a recombinant therapeutic protein treated with inducers of protein expression. *J Biotechnol* 2015;**212**:106–15
9. Pybus LP, Dean G, West NR, Smith A, Daramola O, Field R, James DC. Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells. *Biotechnol Bioeng* 2014;**111**:372–85
10. Wurm FM, Hacker D. First CHO genome. *Nat Biotechnol* 2011;**29**:718–20
11. Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Palsson BO. Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *cricetulus griseus* draft genome. *Nat Biotechnol* 2013;**31**:759–65
12. Vishwanathan N, Le H, Jacob NM, Tsao YS, Ng SW, Loo B, Hu WS. Transcriptome dynamics of transgene amplification in Chinese hamster ovary cells. *Biotechnol Bioeng* 2014;**111**:518–28
13. Panina Y, Germond A, Masui S, Watanabe TM. Validation of common housekeeping genes as reference for qPCR gene expression analysis during iPS reprogramming process. *Sci Rep* 2018;**8**:8716
14. Suzuki T, Higgins PJ, Crawford DR. Control selection for RNA quantitation. *Biotechniques* 2000;**29**:332–7
15. Derks NM, Muller M, Gaszner B, Tilburg-Ouwens DT, Roubos EW, Kozicz LT. Housekeeping genes revisited: different expressions depending on gender, brain area and stressor. *Neuroscience* 2008;**156**:305–9
16. Guenin S, Mauriat M, Pelloux J, Van Wuytswinkel O, Bellini C, Gutierrez L. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *J Exp Bot* 2009;**60**:487–93
17. Lee PD, Sladek R, Greenwood CM, Hudson TJ. Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res* 2002;**12**:292–7
18. Brown AJ, Gibson S, Hatton D, James DC. Transcriptome-Based identification of the optimal reference CHO genes for normalisation of qPCR data. *Biotechnol J* 2018;**13**:1700259–65
19. Dijkstra JR, van Kempen LC, Nagtegaal ID, Bustin SA. Critical appraisal of quantitative PCR results in colorectal cancer research: can we rely on published qPCR results? *Mol Oncol* 2014;**8**:813–8
20. Bahr SM, Borgschulte T, Kayser KJ, Lin N. Using microarray technology to select housekeeping genes in Chinese hamster ovary cells. *Biotechnol Bioeng* 2009;**104**:1041–6
21. Paredes V, Park JS, Jeong Y, Yoon J, Baek K. Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. *Biotechnol Lett* 2013;**35**:987–93
22. Kim M, O'Callaghan PM, Droms KA, James DC. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol Bioeng* 2011;**108**:2434–46
23. Jia YL, Guo X, Lu JT, Wang XY, Qiu LL, Wang TY. CRISPR/Cas9-mediated gene knockout for DNA methyltransferase Dnmt3a in CHO cells displays enhanced transgenic expression and long-term stability. *J Cell Mol Med* 2018;**22**:4106–16
24. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002;**3**:Research0034.1–11
25. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* 2004;**64**:5245–50
26. Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–excel-based tool using pair-wise correlations. *Biotechnol Lett* 2004;**26**:509–15
27. Silver N, Best S, Jiang J, Thein S. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol Biol* 2006;**7**:33–41
28. Xie F, Xiao P, Chen D, Xu L, Zhang B. miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol Biol* 2012;**80**:75–84
29. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods* 2012;**9**:357–9
30. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;**12**:323–38
31. Khimani AH, Mhashilkar AM, Mikulskis A, O'Malley M, Liao J, Golenko EE, Lott ST. Housekeeping genes in cancer: normalization of array data. *Biotechniques* 2005;**38**:739–45
32. Murphy CL, Polak JM. Differentiating embryonic stem cells: GAPDH, but neither HPRT nor beta-tubulin is suitable as an internal standard for measuring RNA levels. *Tissue Eng* 2002;**8**:551–9
33. Serazin-Leroy V, Denis-Henriot D, Morot M, de Mazancourt P, Giudicelli Y. Semi-quantitative RT-PCR for comparison of mRNAs in cells with different amounts of housekeeping gene transcripts. *Mol Cell Probes* 1998;**12**:283–91
34. Savonet V, Maenhaut C, Miot F, Pirson I. Pitfalls in the use of several "housekeeping" genes as standards for quantitation of mRNA: the example of thyroid cells. *Anal Biochem* 1997;**247**:165–7
35. Gabrielsson BG, Olofsson LE, Sjogren A, Jernas M, Elander A, Lonn M, Carlsson LM. Evaluation of reference genes for studies of gene expression in human adipose tissue. *Obes Res* 2005;**13**:649–52
36. Gorzelniak K, Janke J, Engeli S, Sharma AM. Validation of endogenous controls for gene expression studies in human adipocytes and preadipocytes. *Horm Metab Res* 2001;**33**:625–7

37. Hamalainen HK, Tubman JC, Vikman S, Kyrola T, Ylikoski E, Warrington JA, Lahesmaa R. Identification and validation of endogenous reference genes for expression profiling of T helper cell differentiation by quantitative real-time RT-PCR. *Anal Biochem* 2001;**299**:63–70
38. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Wittwer CT. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 2009;**55**:611–22
39. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 2012;**13**:134–44
40. Lee C, Kim J, Shin SG, Hwang S. Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*. *J Biotechnol* 2006;**123**:273–80
41. Chen D, Pan X, Xiao P, Farwell MA, Zhang B. Evaluation and identification of reliable reference genes for pharmacogenomics, toxicogenomics, and small RNA expression analysis. *J Cell Physiol* 2011;**226**:2469–77
42. Chong WP, Reddy SG, Yusufi FN, Lee DY, Wong NS, Heng CK. Metabolomics-driven approach for the improvement of Chinese hamster ovary cell growth: overexpression of malate dehydrogenase II. *J Biotechnol* 2010;**147**:116–21
43. Tabuchi H, Sugiyama T, Tanaka S, Tainaka S. Overexpression of taurine transporter in Chinese hamster ovary cells can enhance cell viability and product yield, while promoting glutamine consumption. *Biotechnol Bioeng* 2010;**107**:998–1003
44. Bohren KM, Bullock B, Wermuth B, Gabbay KH. The aldo-keto reductase superfamily. cDNAs and deduced amino acid sequences of human aldehyde and aldose reductases. *J Biol Chem* 1989;**264**:9547–51
45. Kurahashi T, Kwon M, Homma T, Saito Y, Lee J, Takahashi M. Reductive detoxification of acrolein as a potential role for aldehyde reductase (AKR1A) in mammals. *Biochem Biophys Res Commun* 2014;**452**:136–41
46. Gabbay KH, Bohren KM, Morello R, Bertin T, Liu J, Vogel P. Ascorbate synthesis pathway: dual role of ascorbate in bone homeostasis. *J Biol Chem* 2010;**285**:19510–20
47. Akihara R, Homma T, Lee J, Yamada K, Miyata S, Fujii J. Ablation of aldehyde reductase aggravates carbon tetrachloride-induced acute hepatic injury involving oxidative stress and endoplasmic reticulum stress. *Biochem Biophys Res Commun* 2016;**478**:765–71
48. Zhang Q, Xu H, You Y, Zhang J, Chen R. High Gpx1 expression predicts poor survival in laryngeal squamous cell carcinoma. *Auris Nasus Larynx* 2018;**45**:13–9
49. Long C, Xiao Y, Li S, Tang X, Yuan Z, Bai Y. Identification of optimal endogenous reference RNAs for RT-qPCR normalization in hindgut of rat models with anorectal malformations. *Peer J* 2019;**7**:e6829
50. Zhang Y, Han X, Chen S, Zheng L, He X, Liu M. Selection of suitable reference genes for quantitative real-time PCR gene expression analysis in *salix matsudana* under different abiotic stresses. *Sci Rep* 2017;**7**:40290
51. Furrow E, Pfeifer RJ, Osborne CA, Lulich JP. An APRT mutation is strongly associated with and likely causative for 2,8-dihydroxyadenine urolithiasis in dogs. *Mol Genet Metab* 2014;**111**:399–403
52. Garcia-Huertas P, Mejia-Jaramillo AM, Gonzalez L, Triana-Chavez O. Transcriptome and functional genomics reveal the participation of adenine phosphoribosyltransferase in *Trypanosoma cruzi* resistance to benznidazole. *J Cell Biochem* 2017;**118**:1936–45
53. Ling H, Wu Q, Guo J, Xu L, Que Y. Comprehensive selection of reference genes for gene expression normalization in sugarcane by real time quantitative rt-PCR. *PLoS One* 2014;**9**:e97469
54. Xue B, Guo J, Que Y, Fu Z, Wu L, Xu L. Selection of suitable endogenous reference genes for relative copy number detection in sugarcane. *Int J Mol Sci* 2014;**15**:8846–62
55. Ian'shina Dd Malygin AA, Karpova GG. [Binding of human ribosomal protein S16 with the 18S rRNA fragment 1203-1236/1521]. *Molekul Biol* 2007;**41**:1023–30
56. Livramento KGD, Freitas NC. Gene expression profile analysis is directly affected by the selected reference gene: the case of Leaf-Cutting atta sexdens. *Insects* 2018;**9**:18–33
57. de Melo Neto OP, da Costa Lima TDC, Merlo KC, Romao TP, Rocha PO, Assis LA. Phosphorylation and interactions associated with the control of the leishmania Poly-A binding protein 1 (PABP1) function during translation initiation. *RNA Biol* 2018;**15**:739–55
58. Smith RW, Blee TK, Gray NK. Poly(a)-binding proteins are required for diverse biological processes in metazoans. *Biochem Soc Trans* 2014;**42**:1229–37
59. Assrir N, Filhol O, Galisson F, Lipinski M. HIRIP3 is a nuclear phosphoprotein interacting with and phosphorylated by the serine-threonine kinase CK2. *Biol Chem* 2007;**388**:391–8
60. Lorain S, Quivy JP, Monier-Gavelle F, Scamps C, Lecluse Y, Almouzni G, Lipinski M. Core histones and HIRIP3, a novel histone-binding protein, directly interact with WD repeat protein HIRA. *Mol Cell Biol* 1998;**18**:5546–56

(Received December 2, 2019, Accepted February 25, 2020)