



OPEN

Validation of reference genes for use in untreated bovine fibroblasts

T. Toorani^{1,2}, P. M. Mackie¹ & G. F. Mastromonaco^{1,2}✉

Proper normalization of RT-qPCR data is pivotal to the interpretation of results and accuracy of scientific conclusions. Though different approaches may be taken, normalization against multiple reference genes is now standard practice. Genes traditionally used and deemed constitutively expressed have demonstrated variability in expression under different experimental conditions, necessitating the proper validation of reference genes prior to utilization. Considering the wide use of fibroblasts in research and scientific applications, it is imperative that suitable reference genes for fibroblasts of different animal origins and conditions be elucidated. Previous studies on bovine fibroblasts have tested limited genes and/or samples. Herein, we present an extensive study investigating the expression stability of 16 candidate reference genes across 7 untreated bovine fibroblast cell lines subjected to controlled conditions. Data were analysed using various statistical tools and algorithms, including geNorm, NormFinder, BestKeeper, and RefFinder. A combined use of *GUSB* and *RPL13A* was determined to be the best approach for data normalization in untreated bovine fibroblasts.

The inherent properties of fibroblasts render them ideal cellular models for numerous research and scientific applications. Fibroblasts can be retrieved non-invasively and from a multitude of tissues, and are easily cultured and maintained in vitro¹. Fibroblast culture and cryopreservation techniques are well-established, straightforward, and do not require specialized protocols even among different taxa^{1,2}. Historically, fibroblast cell lines established from human and animal tissues have been used to improve our understanding of disease pathogenesis³, wound healing^{4,5}, and normal fibroblast physiology. More recently, fibroblasts have been used as donor cells in somatic cell nuclear transfer (SCNT)^{6,7}, and induced pluripotent stem cell applications (iPSC)^{8–10}, which present powerful tools for disease modeling in vitro^{11–16}, personalized and regenerative medicine, oncogenic applications¹⁷, and wildlife conservation¹, among others. In many of these studies, describing transcriptional changes of key regulatory genes within the fibroblast has been crucial to the complete understanding of the cellular mechanisms underpinning their function. As a measure of transcriptional dynamics, quantification of mRNA abundance via reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) has become standard practice across many disciplines, owing to its theoretical and logistical simplicity¹⁸. However, while RT-qPCR permits the quantification of small amounts of nucleic acid and the detection of minute variability, including single copy differences¹⁹, it also presents a potential pitfall to this technique, unless proper normalization steps are taken. Accurate data analysis and interpretation following RT-qPCR are contingent upon suitable normalization methods to control for potential errors introduced throughout the multi-step process, including normalization to an internal reference gene (RG)²⁰. Further, the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines recommend the use of multiple internal RGs for data normalization¹⁸, owing to the resolution of the data being defined by the stability of the reference genes under the given experimental conditions²⁰.

The most commonly used “classical” RGs, including β -actin (*ACTB*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), hypoxanthine-guanine phosphoribosyl transferase (*HPRT*) and 18S ribosomal RNA (*18S rRNA*), are carryovers from references used in Northern blotting, RNase protection and conventional RT-PCR assays, which were suitable for these non- and semi-quantitative techniques where qualitative changes were evaluated²⁰. However, the advent of quantitative techniques such as RT-qPCR should have instigated the evaluation of more suitable normalization approaches²⁰. Instead, many investigators continue to use these classical RGs, assuming consistent expression without adequate experimental validation. For example, according to Chapman

¹Reproductive Sciences, Toronto Zoo, Scarborough, ON M1B 5K7, Canada. ²Department of Biomedical Sciences, Ontario Veterinary College, University of Guelph, Guelph, ON N1G 2W1, Canada. ✉email: gmastromonaco@torontozoo.ca

Gene	Name	Function/pathway	NCBI accession no	Qiagen catalog no
<i>ACTB</i>	Actin, beta	Cytoskeletal structural protein	NM_173979.3	PPB00173A-200
<i>B2M</i>	Beta-2-microglobulin	Beta chain of MHC class 1 molecule	NM_173893.3	PPB00031A-200
<i>GAPDH</i>	Glyceraldehyde-3-phosphate dehydrogenase	Glycolysis and gluconeogenesis enzyme	NM_001034034.2	PPB00298A-200
<i>GUSB</i>	Glucuronidase, beta	Catalysis of complex carbohydrates breakdown	NM_001083436.1	PPB06553A-200
<i>HMBS</i>	Hydroxymethylbilan synthase	Heme biosynthetic pathway	NM_001046207.1	PPB06519A-200
<i>HPRT1</i>	Hypoxanthine phosphoribosyltransferase 1	Purine synthesis in salvage pathway	NM_001034035.2	PPB00330A-200
<i>HSP90AB1</i>	Heat shock 90 kDa protein 1, beta	Molecular chaperone	NM_001079637.1	PPB14507A-200
<i>PPIA</i>	Peptidylprolyl isomerase A	Protein folding	NM_178320.2	PPB00426A-200
<i>RAD50</i>	Radiation sensitive 50	DNA double-stranded break repair protein	NM_001206868.1	PPB15504A-200
<i>RPL13A</i>	Ribosomal protein L13a	Ribosome structural constituent; protein synthesis	NM_001076998.2	PPB14550A-200
<i>RPS18</i>	Ribosomal protein S18	Ribosomal protein, component of 40S subunit	NM_001033614.2	PPB01408A-200
<i>SDHA</i>	Succinate dehydrogenase complex, subunit A	Electron transporter in TCA cycle and respiratory chain	NM_174178.2	PPB00460A-200
<i>SF3A1</i>	Splicing factor 3a, subunit 1	pre-mRNA splicing; as a component of pre-catalytic spliceosome "B" complexes	NM_001081510.1	PPB02067A-200
<i>TBP</i>	TATA box binding protein	TATA box binding protein, general RNA polymerase II transcription factor	NM_001075742.1	PPB06797A-200
<i>UBC</i>	Ubiquitin C	Protein modifier; attaches to lysine	NM_001206307.2	PPB01883A-200
<i>YWHAZ</i>	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta	Signal transduction via binding to phosphoserine-containing proteins	NM_174814.2	PPB01343A-200

Table 1. Overview and details of candidate RGs evaluated in this study.

and Waldenström, *ACTB* and *GAPDH* continue to be the most widely used RGs, with 62% of studies using one of these genes as the single normalizing RG, of which only 8% did the due diligence of evaluating stability prior to use²¹. This is despite evidence showing that the expression stability of both *ACTB*^{22–26} and *GAPDH*^{24,27–29} are not impervious to experimental conditions. A further examination of RG stability validation studies showed a significant ($P < 0.001$) and inverse relationship between the number of RGs screened and the probability that *ACTB*, *GAPDH*, or *18S rRNA* were selected for normalization; where the likelihood of these three genes being selected is significantly decreased when more RGs are tested²¹. The importance of proper RG selection has been demonstrated by numerous studies reporting significant differences in results owing to normalization against RGs of varying stability^{23,30,31}. It is unlikely that any universal RGs exist, and it is therefore important for the suitability of RGs to be experimentally validated for a given set of samples and conditions prior to use.

Optimal RGs have been reported for RT-qPCR use in various bovine cells^{25,26,29,32–34}, including fibroblasts^{31,35}. Interestingly, Zhou et al. reported larger variation when only one RG (*ACTB*) was used to calculate target gene expression compared to two RGs³¹. However, limitations in these studies, including small sample size ($n = 1$)³⁵, number of RGs tested (5)³¹, and the use of at most two RG determination algorithms³¹, may have impacted the identification of optimal RGs for bovine fibroblast cells. The current study addresses a gap in knowledge concerning suitable RGs for use in RT-qPCR studies investigating untreated bovine fibroblast cell lines. We evaluated the expression stability of 16 candidate RGs across 7 untreated bovine fibroblast cell lines grown under controlled conditions and standardization by morphology and growth kinetics. Candidate genes were selected based on an extensive review of the literature (Supplementary Table S1) and included RGs previously described in fibroblast and/or bovine studies, as well as the classical RGs (Table 1). Special consideration was given to select genes from various pathways and functional classes to avoid co-regulation. Data were analysed using the most common RG determination methods: geNorm³⁶, NormFinder³⁷, BestKeeper³⁸, and RefFinder³⁹, which integrates the algorithms of the first three methods, as well as delta C⁴⁰. The algorithms for each differ and, as such, their combined application should improve confidence in the selection of ideal candidate RGs when there is congruence. To our knowledge, this is the most extensive validation of RGs for use in untreated bovine fibroblasts undertaken to date.

Results

Range of fluctuation of RGs determined by box-and-whiskers plot. An extensive evaluation of the literature describing the use of RGs, including those specific to fibroblast and/or bovine cells, culminated in the selection of 16 candidate RGs for inclusion in this study. Following RT-qPCR, gene expression variability (Ct) was assessed for each of the candidate genes (Fig. 1). As a measure of stability, the range of fluctuation of each gene was determined by finding the difference between 25 and 75th percentiles, the interquartile range. *YWHAZ*, *PPIA* and *HSP90AB1* displayed the greatest variability, with ranges of fluctuation of 0.5798, 0.4909, and 0.4500 respectively, while the most stable genes were *RPL13A*, *GUSB* and *TBP* with ranges of 0.2368, 0.2442, and 0.2691, respectively.

geNorm analysis of RG stability. The stability of the candidate RGs was analysed using four statistical methods. The algorithm, geNorm, is a comprehensive tool, performing an initial measure of expression stability (M) value assessment for all genes, followed by a stepwise exclusion analysis, wherein the least stable gene is eliminated, and the remaining genes are reassessed for M values. Although M values for all genes fell within range for inclusion ($M < 0.5$), *B2M*, *HPRT1*, and *RAD50* had the lowest stability while *GUSB*, *RPL13A*, and

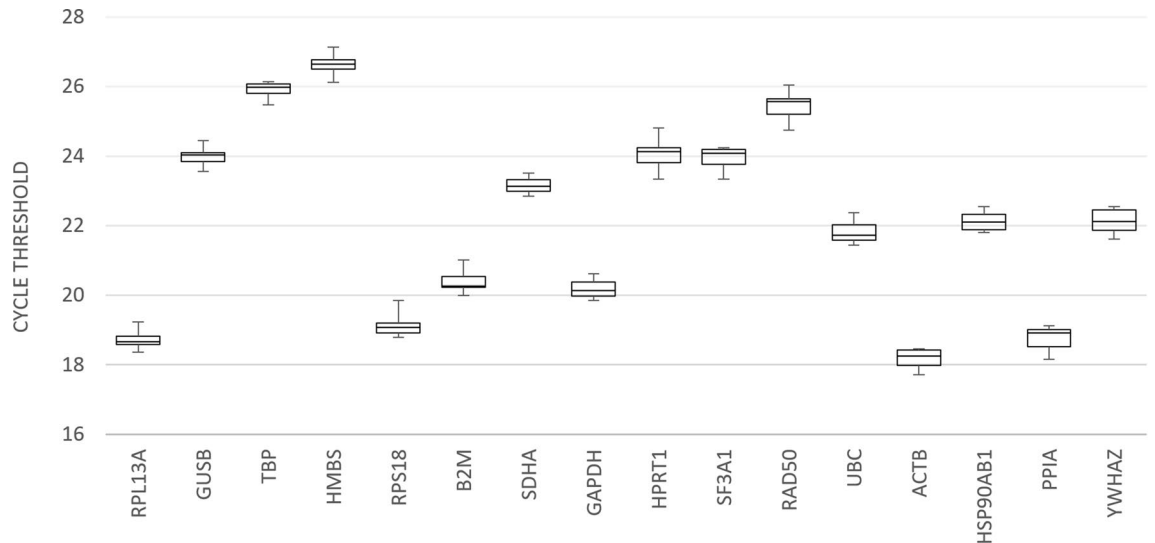


Figure 1. Range of fluctuation of candidate RG Ct values. Box-and-whiskers plot illustrating range of fluctuation of candidate RG Ct values across fibroblast cell lines ($n=7$) in order (from left to right) of increasing interquartile range. Each cell line was reverse transcribed in triplicate and combined, and subsequently assayed in technical triplicates. Mean of technical replicates were used in this analysis. The first quartile, median, and third quartile are indicated by the lower limit of the box, the line within the box, and the upper limit of the box, respectively. The whiskers represent the minimum and maximum values. Data represent one independent experiment.

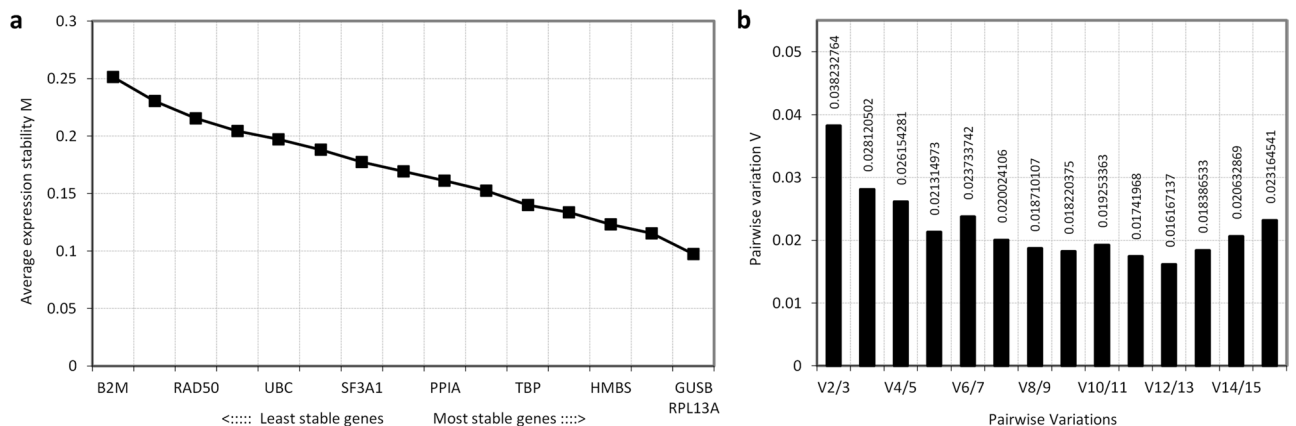


Figure 2. geNorm candidate RG output. Results of (a) the average stepwise exclusion analysis to determine gene values with highest stability expressed as average expression stability (M) values, and (b) pairwise variation (V) values, to determine the optimal number of RGs. Consensus for M and V value cut-off is 0.5 and 0.15, respectively. Sixteen candidate RGs were assessed across $n=7$ cell lines (reverse transcribed in triplicate and combined, and subsequently assayed in technical triplicates). Calculated relative quantity (RQ) of mean Ct values were used as input. Data represent one independent experiment.

SDHA had the highest stability (Fig. 2a). Pairwise variation (V) values were also calculated to determine the suitable number of RGs required for untreated bovine fibroblast gene expression studies using RT-qPCR (Fig. 2b). Low pairwise variation (V) value for $V_{2/3}=0.0382$ indicated that the two most stable RGs (*RPL13A* and *GUSB*) were suitable for data normalization and that adding a third reference gene would not provide additional value⁴¹. Though geNorm shows bias in selecting co-regulated genes, *RPL13A* and *GUSB* are not co-regulated (see Supplementary Tables S2 and S3), further corroborating their suitability as reference genes.

NormFinder analysis of RG stability. Gene stability rankings were compared between the Microsoft Excel plug-in and R version of NormFinder (see Supplementary Table S4). While the former generated stability values, the latter computed GroupSD values for individual genes and combinations of two genes to only two decimal places, precluding the determination of gene rankings with certainty. However, rankings of both versions generally coincided: based on the output, *B2M*, *HPRT1*, and *RAD50* showed the lowest stability, while *HMBS*, *GUSB*, and *TBP* had the highest stability. The most stable combination of genes was *GAPDH* and *YWHAZ* (0.03),

Gene1	Gene2	GroupSD	RANK
<i>GAPDH</i>	<i>YWHAZ</i>	0.03	1
<i>ACTB</i>	<i>SF3A1</i>	0.04	2
<i>GAPDH</i>	<i>SF3A1</i>	0.04	2
<i>GUSB</i>	<i>PPIA</i>	0.04	2
<i>HMBS</i>	<i>RPL13A</i>	0.04	2
<i>ACTB</i>	<i>HMBS</i>	0.05	6
<i>GAPDH</i>	<i>SDHA</i>	0.05	6
<i>PPIA</i>	<i>TBP</i>	0.05	6
<i>ACTB</i>	<i>GUSB</i>	0.06	9
<i>ACTB</i>	<i>TBP</i>	0.06	9
<i>GAPDH</i>	<i>GUSB</i>	0.06	9
<i>GUSB</i>	<i>RPL13A</i>	0.06	9
<i>PPIA</i>	<i>SF3A1</i>	0.06	9
<i>RPL13A</i>	<i>SF3A1</i>	0.29	14
<i>PPIA</i>	<i>SDHA</i>	0.35	15

Table 2. NormFinder candidate RG combination stability output.

	<i>ACTB</i>	<i>GAPDH</i>	<i>GUSB</i>	<i>HMBS</i>	<i>PPIA</i>	<i>RPL13A</i>	<i>SDHA</i>	<i>SF3A1</i>	<i>TBP</i>	<i>YWHAZ</i>
Geometric mean [Ct]	18.19	20.18	23.97	26.64	18.75	18.71	23.17	23.94	25.92	22.13
Arithmetic mean [Ct]	18.20	20.18	23.97	26.64	18.75	18.71	23.17	23.94	25.92	22.13
Min [Ct]	17.73	19.85	23.55	26.13	18.13	18.35	22.88	23.34	25.48	21.61
Max [Ct]	18.49	20.60	24.46	27.11	19.08	19.23	23.48	24.24	26.14	22.56
Standard deviation [\pm Ct]	0.25	0.23	0.21	0.22	0.29	0.20	0.19	0.28	0.18	0.31
Coefficient of variance [% Ct]	1.36	1.14	0.86	0.82	1.54	1.06	0.83	1.16	0.71	1.40
Min [x-fold]	-1.38	-1.26	-1.33	-1.43	-1.53	-1.28	-1.23	-1.52	-1.35	-1.44
Max [x-fold]	1.23	1.33	1.41	1.39	1.26	1.43	1.24	1.24	1.17	1.35
Standard deviation [\pm x-fold]	1.19	1.17	1.15	1.16	1.22	1.15	1.14	1.21	1.14	1.24

Table 3. BestKeeper candidate RG descriptive statistics, including standard deviation [\pm Ct], which is used as a cut-off for subsequent BestKeeper Index calculation.

BestKeeper vs	<i>ACTB</i>	<i>GAPDH</i>	<i>GUSB</i>	<i>HMBS</i>	<i>PPIA</i>	<i>RPL13A</i>	<i>SDHA</i>	<i>SF3A1</i>	<i>TBP</i>	<i>YWHAZ</i>
Coeff. of corr. [r]	0.974	0.869	0.937	0.940	0.954	0.883	0.906	0.872	0.947	0.955
p-value	0.001	0.011	0.002	0.002	0.001	0.008	0.005	0.011	0.001	0.001

Table 4. BestKeeper Index calculation, used in determining coefficients of correlation [r] and p values for candidate RGs.

followed closely by *ACTB* and *SF3A1*, *GAPDH* and *SF3A1*, *GUSB* and *PPIA*, and *HMBS* and *RPL13A* (0.04) (Table 2).

BestKeeper analysis of RG stability. BestKeeper generated a set of descriptive statistics (Table 3), where genes falling short of certain cut-offs (standard deviation [\pm Ct] > 1 and/or standard deviation [\pm x-fold] > 2) were omitted from the subsequent BestKeeper index calculation that determines coefficients of correlation [r] and p values (Table 4). All genes satisfied the initial criterion and were kept in the subsequent analysis. Based on both standard deviation values (\pm Ct/ \pm x-fold), *YWHAZ*, *PPIA* and *SF3A1* displayed the lowest stability, while *TBP*, *SDHA* and *RPL13A* had the highest stability. Using the coefficient of correlation [r] value, *GAPDH*, *SF3A1* and *RPL13A* displayed the lowest stability while *ACTB*, *YWHAZ* and *PPIA* showed the highest stability.

Reffinder analysis of RG stability. Reffinder assesses gene stability based on the algorithms of geNorm, NormFinder, BestKeeper, and delta Ct, and generates a comprehensive ranking based on the geometric mean of ranking values derived from the four other algorithms. As Reffinder does not specify input requirements, results derived from the use of median and mean Ct values were compared. Remarkably, for all four imbedded algorithms and the comprehensive ranking, the top four most stable genes were the same between the median Ct and mean Ct datasets, albeit in slightly different orders for geNorm, NormFinder and the comprehensive rankings.

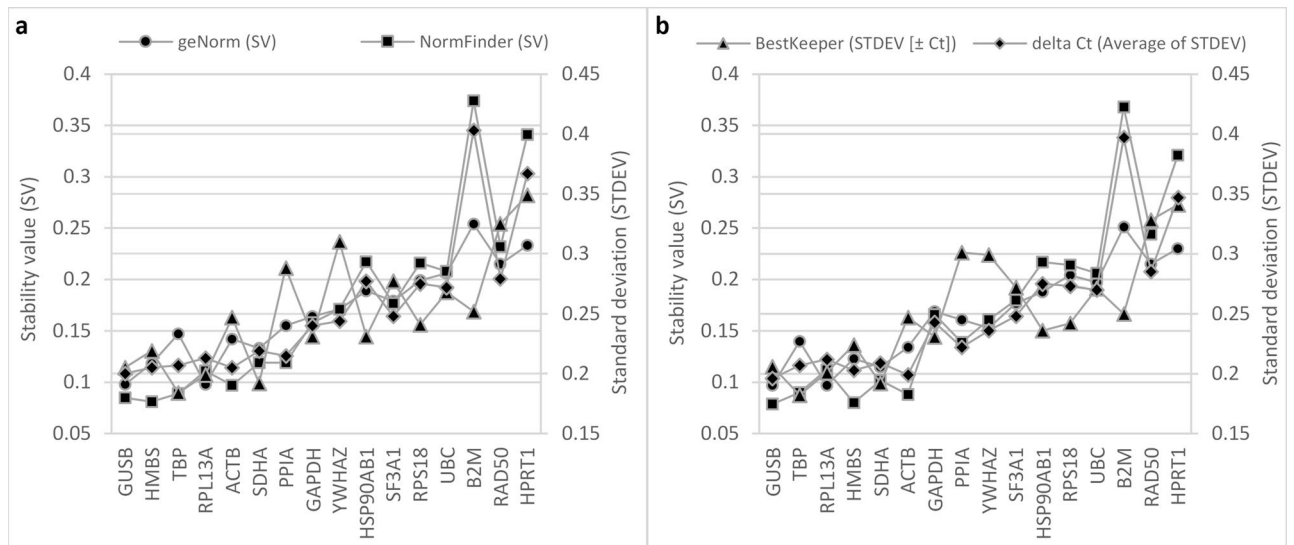


Figure 3. RefFinder candidate RG output. RefFinder results using (a) median, and (b) mean Ct values of technical replicates. Genes showing the highest stability, as determined by RefFinder's comprehensive ranking, are ordered left to right on the x-axis. Individual results from embedded algorithms (geNorm, NormFinder, BestKeeper, and delta Ct) are also illustrated. Lower stability values (SV) and standard deviation (STDEV) are indicative of more stably expressed genes. Sixteen candidate RGs were assessed across $n=7$ cell lines, which were reverse transcribed in triplicate and combined, and consequently assayed in technical triplicates. Data represent one independent experiment.

Each algorithm generated slightly different outcomes (Fig. 3). For the median Ct dataset, results from geNorm, NormFinder, and delta Ct identified genes *B2M*, *HPRT1*, and *RAD50* as the least stable, while BestKeeper determined *HPRT1*, *RAD50*, and *YWHAZ* as least stable. RGs with the highest stability were as follows: geNorm determined the combination of *GUSB* and *RPL13A*, and *HMBS*; NormFinder identified *HMBS*, *GUSB*, and *TBP*; delta Ct identified *GUSB*, *ACTB*, and *HMBS*; and BestKeeper determined *TBP*, *SDHA*, and *RPL13A* as the most suitable. Overall, *HPRT1*, *RAD50*, and *B2M* had the poorest rankings, while *GUSB*, *HMBS*, and *TBP* showed the highest stability. With respect to the mean Ct dataset, the least stable genes were determined by geNorm, NormFinder, and delta Ct to be *B2M*, *HPRT1*, and *RAD50*, while BestKeeper flagged *HPRT1*, *RAD50*, and *PPIA*. The highest stability was shown by the combination of *GUSB/RPL13A*, and *SDHA* according to geNorm; *GUSB*, *HMBS*, and *ACTB* based on NormFinder; *TBP*, *SDHA*, and *RPL13A* according to BestKeeper; and finally, *GUSB*, *ACTB*, and *HMBS* based on delta Ct. The comprehensive ranking determined *HPRT1*, *RAD50*, and *B2M* as the least stable genes and *GUSB*, *TBP*, and *RPL13A* as the most stable genes. For detailed results from both datasets, see Supplementary Table S5.

Overview of rankings from algorithms. The results generated by the box-and-whiskers analysis, geNorm, NormFinder, RefFinder, and BestKeeper have been summarized in Table 5.

Discussion

The selection of suitable RGs for RT-qPCR data normalization is pivotal to the integrity and reliability of results¹⁸. The present study aimed to identify stably expressed RGs for use in RT-qPCR normalization of untreated bovine fibroblasts by assessing a panel of 16 candidate RGs using numerous algorithms and statistical tools.

Several considerations were made in selecting the most optimal RG in this instance. Overall, rankings by different algorithms were generally in agreement, except for BestKeeper, which consistently generated different outcomes. The differences in BestKeeper outcomes, as compared to other RG determination algorithms, may be explained, at least partly, by the use of raw Ct values as input (as compared to the RQ values required by geNorm and NormFinder). Rydbirk and colleagues offer an alternate explanation, referring to the BestKeeper Index, a tool unique to the eponymous software, as the source of this discrepancy in outcome⁴². The BestKeeper Index is calculated from the geometric mean of all RGs satisfying the initial exclusion criterion (standard deviation [\pm Ct] > 1) and is subsequently used to generate Pearson correlation (r) values (coefficient of correlation) for each RG in comparison to this index, thereby determining expression stability. Setting aside BestKeeper results, some genes displayed consistently poor rankings (*B2M*, *HPRT1*, *HSP90AB1*, *RAD50*, *RPS18*, *SF3A1*, and *UBC*), while others consistently showed higher expression stability and the tightest range of Ct values (*GUSB*, *HMBS*, *TBP*, and *RPL13A*). A summary of rankings can be found in Table 5. Regarding similarities between the same algorithms used in different software, both versions of NormFinder were largely in agreement. RefFinder outputs and those of their original counterparts were also highly similar (other than BestKeeper, since RefFinder ranks RGs according to SG, and not the BestKeeper Index). Results from the mean and median were comparable, with small differences observed mostly between genes that do not show much difference in stability ranking.

Candidate RGs	Box-and-whiskers	geNorm			NormFinder			ReffFinder (median Ct/mean Ct)				*BestKeeper (Excel plug-in)	
		Excel plug-in	Excel plug-in	R-based	geNorm	NormFinder	BestKeeper	delta Ct	Comprehensive ranking	Standard deviation of Ct	Coefficient of correlation [r]		
<i>GUSB</i>	2	1	2	1	1/1	2/1	4/4	1/1	1/1	4	6		
<i>HMBS</i>	4	4	1	1	3/4	1/2	5/5	3/3	2/4	5	5		
<i>TBP</i>	3	6	3	3	6/6	3/4	1/1	4/4	3/2	1	4		
<i>RPL13A</i>	1	1	5	5	1/1	5/6	3/3	5/6	4/3	3	8		
<i>ACTB</i>	13	5	4	4	5/5	4/3	9/9	2/2	5/6	7	1		
<i>SDHA</i>	7	3	7	6	4/3	7/5	2/2	7/5	6/5	2	7		
<i>PPIA</i>	15	8	6	6	7/8	6/7	13/14	6/7	7/8	9	3		
<i>GAPDH</i>	8	9	8	8	8/9	8/9	7/6	8/9	8/7	6	10		
<i>YWHAZ</i>	16	7	9	9	9/7	9/8	14/13	9/8	9/9	10	2		
<i>HSP90AB1</i>	14	11	13	12	11/11	13/13	6/7	13/13	10/11	–	–		
<i>RPS18</i>	5	13	12	12	12/13	12/12	8/8	12/12	12/12	–	–		
<i>UBC</i>	12	12	11	11	13/12	11/11	11/11	11/11	13/13	–	–		
<i>SF3A1</i>	10	10	10	10	10/10	10/10	12/12	10/10	11/10	8	9		
<i>RAD50</i>	11	14	14	14	14/14	14/14	15/15	14/14	15/15	–	–		
<i>B2M</i>	6	16	16	16	16/16	16/16	10/10	16/16	14/14	–	–		
<i>HPRT1</i>	9	15	15	15	15/15	15/15	16/16	15/15	16/16	–	–		

Table 5. Compilation of rankings of candidate RGs by all statistical tools and algorithms. *The six genes most frequently ranked with the lowest stability were omitted from analysis using BestKeeper, which is limited to evaluating only 10 genes.

Considering each algorithm uses a different approach and evaluates different parameters when determining RG expression stability, it is expected that variations in the results will be observed, though many studies have reported general consensus between algorithms⁴³. As such, difficulties arise when trying to reconcile different rankings and several considerations must be made in selecting optimal RG(s), including the (i) overlap of rankings by various algorithms, (ii) strengths and weaknesses of each algorithm, as applicable to the experimental design of each study, and (iii) use of different complementary tools for merging results into a comprehensive ranking. GeNorm's pairwise correlation is known to be a robust algorithm for studies with small sample sizes but shows bias towards selecting RGs that are co-regulated. NormFinder's model-based approach confers the advantage of differentiating intragroup variation from intergroup variation, and as such, is a suitable tool for the assessment of RGs in experiments with different sample groups, but requires larger sample sizes (> 8) in comparison to geNorm^{37,43,44}. Different tools have been reported by investigators to merge results of different algorithms into a comprehensive ranking, the most well-known of which is ReffFinder³⁹. Concerns have been raised about the validity of this software mainly based on the fact that ReffFinder does not integrate primer efficiencies in calculations, and that its BestKeeper output is solely based on the initial standard deviation [\pm Ct] values and not the BestKeeper Index. Another limitation of ReffFinder is that results of the four methods are not weighed given the unavailability of their cut-offs and applicable weights⁴⁵. Lastly, cycle threshold coefficient of variation (Ct CV%)⁴⁶ can also be used to determine comprehensive ranking whereby a lower Ct CV% is indicative of higher stability and is one of the parameters found in the initial descriptive statistics generated by BestKeeper. Though this presents a simple approach in assessing RG stability, Ct CV% should be considered carefully, as it is not always inversely correlated to the degree of stability of a given gene⁴⁷. Selection of the most suitable algorithm for a given experimental design must take into account the different strengths and weaknesses as described above. In our case, given our small sample size, geNorm was the most suitable algorithm, but the results from all algorithms were taken into consideration before selecting the optimal RGs.

The panel of genes evaluated showed good stability (e.g. all genes satisfied geNorm's M value criterion of 0.5, NormFinder's standard deviation [\pm Ct] of 1, and geNorm's V value criterion of 0.15), compared to the range of stabilities seen for candidate RGs assessed in other studies^{26,32}. This is to be expected due to the set of standardized untreated fibroblasts used, with average M values between 0.2–0.5 typically being observed when assessing candidate RGs across a homogeneous sample set⁴¹. *GUSB* showed the most consistently stable expression amongst all RGs, ranking first or second in 12 of the 16 rankings. The use of a single RG, however, is discouraged and can create variability in the expression of the target genes independent of experimental treatments³¹, including more than threefold deviation from true values in 25% of cases, and at least sixfold in 10% of cases³⁶. The appropriate number of RGs required must be experimentally validated. GeNorm determines the number and combinations of genes showing stability while NormFinder evaluates all possible pairs of genes. In our study, according to geNorm, the best approach was a combined use of *GUSB* and *RPL13A*, which was corroborated by both versions of NormFinder, where this combination, though tied for 9th in rankings, showed greater stability than the number 1 ranked individual genes *HMBS* and *GUSB* (Group SD of 0.06 vs 0.08, respectively). Further investigation of *GUSB* and *RPL13A* revealed no-coregulation between these two genes, an important consideration when selecting RGs^{37,43,44}. Individually, *GUSB* and *RPL13A* show high and medium/medium–high expression stabilities, respectively. It is their combined expression (geometric mean of expression) that shows high stability

and was therefore determined as the optimal approach for data normalization in untreated bovine fibroblasts. These two genes also display the smallest range of Ct values (Fig. 1). Previous RG validation in bovine fibroblasts concluded the combined use of *ACTB* and *YWHAZ* as the most suitable data normalization approach (using geNorm and NormFinder), and showed that data normalization against one RG (*ACTB*; which showed high expression stability within the genes assessed) can yield results deviating from true values³¹. However, this study only assessed five candidate genes using RNA from treated (serum-starved cells) and co-regulation between *ACTB* and *YWHAZ* did not appear to be considered³¹.

The information provided here addresses a gap in knowledge concerning suitable reference genes for use in RT-qPCR studies of untreated bovine fibroblast cell lines, a commonly studied species. This type of validation, though essential, is costly and time consuming. To this end, the current study was designed as the most extensive RG validation in untreated fibroblast cell lines to date. Our results provide targeted information for future RT-qPCR studies, alleviating the researcher of the burden of repeating the validation. For example, the results presented herein were used to normalize gene expression when comparing inherent expression between various bovine fibroblast cell lines⁴⁸. The limitation of this paper, however, is that any deviations applied in future studies (i.e. different experimental treatments, cell culture conditions), will warrant novel validation of suitable reference genes. However, investigators can use the reference genes found suitable herein as a starting point, or broaden their evaluation to include any of the 16 candidate genes carefully selected from the literature and further tested for co-regulation. Finally, we also highlight the varying strengths and weaknesses of the most published algorithms, and recommend using a multitude of approaches when evaluating gene stability to ensure accurate RT-qPCR normalization under your given experimental conditions.

We conclude that the combined use of *GUSB* and *RPL13A* presents the best approach in normalizing RT-qPCR data in untreated bovine fibroblasts. Though these genes have been less frequently used for normalization, they have previously been shown as suitable RGs in human adipose tissue- and Wharton's Jelly-derived mesenchymal stromal cells (*RPL13A*)²⁴, human nasal epithelium cells (*GUSB*)⁴⁹, human papillary thyroid carcinoma (*GUSB*)⁵⁰, human dermal fibroblasts and mammary epithelial cells (*GUSB*)⁵¹, rat bone mesenchymal stem cell-derived neuronal cells (*RPL13A*)⁵², and treated human ovarian cancer cells (*RPL13A*)⁵³.

Materials and methods

Tissue processing and fibroblast cell culture. Ears from healthy, age-matched (< 30 mos.) Angus bulls were removed post-mortem at the local abattoir, covered with sterile saline and transported to the lab for processing within 4 h of collection. Tissue samples (n=7) were processed, and fibroblast cells cultured as previously described⁵⁴ with minor modifications. Previous experiments have shown comparable outcome between sample sizes of n=11 and n=6⁴³. Given the homogeneity of the samples (i.e. untreated bovine fibroblasts), it was deemed that n=7 was adequate for the determination of suitable reference genes. All materials and reagents were obtained from Thermo Fisher Scientific (Mississauga, ON, Canada) unless otherwise noted. Briefly, tissue samples spanning the entire thickness of the ears were collected using an 8 mm biopsy punch, the skin was discarded, and the remaining cartilage was manually dissected into 1 mm² pieces before being collagenase-digested (Sigma-Aldrich, Oakville, ON, Canada) for 5 h at 38.5 °C and 5% CO₂. Cells and tissue fragments were pelleted by centrifugation at 200 rcf for 5 min, resuspended in Dulbecco's Modified Eagle's Medium (DMEM; Sigma-Aldrich, Oakville, ON, Canada) supplemented with 20% fetal bovine serum (FBS; Sigma-Aldrich, Oakville, ON, Canada) and 1% antibiotic-antimycotic (ABAM), and cultured at 38.5 °C and 5% CO₂ in 25 cm² tissue culture flasks. Media were partially and fully replaced on days 4 and 6, respectively, until confluency, at which point cells were passaged at a dilution of 1:4 into DMEM containing 10% FBS and 1% penicillin-streptomycin (P/S). After 2 passages (P3), confluent cell lines were trypsin-harvested (0.25% trypsin-EDTA), diluted 1:4 in freezing medium containing DMEM supplemented with 20% FBS, 10% dimethyl sulfoxide (DMSO; Sigma-Aldrich, Oakville, ON, Canada), and 1% P/S, and cryopreserved. For consistency, only one ear was processed per day and all steps were conducted by the same operator. Cell lines that did not reach confluency within 9-days from the initiation of culture or contained > 10% cells with abnormal morphology (assessed visually) were discarded. All cell lines were authenticated for cell type homogeneity, chromosome normality (2n=60) and sex, as described by Toorani et al.⁴⁸. Inclusion and exclusion criteria were pre-established.

RNA extraction and cDNA synthesis. Fibroblasts were thawed, cultured to confluency (P4), pelleted, and frozen at - 80 °C for RNA extraction using a Total RNA Purification Kit (Norgen Biotek Corp., Thorold, ON, Canada) following manufacturer's instructions, including an on-column gDNA removal step (RNase-Free DNase I kit, Norgen Biotek Corp., Thorold, ON, Canada). Isolated RNA was quantified using the Qubit RNA HS Assay Kit. RNA integrity was assessed visually (28S and 18S rRNA) following separation on 1% E-GEL, EX-Agarose gels using the Invitrogen E-Gel Power Snap Electrophoresis System.

Reverse transcription (RT) was performed using 1000 ng of RNA, and in accordance with, Norgen Biotek Corp's TruScript First Strand cDNA Synthesis Kit (Thorold, ON, Canada). Samples were stored at - 20 °C before proceeding to reverse transcription quantitative real-time PCR (RT-qPCR).

Selection of candidate genes. A panel of 16 candidate RGs was selected based on an extensive review of the literature (see Supplementary Table S1). Classical RGs were chosen, as well as others used or validated for use in bovine or fibroblast cells. Special consideration was taken to select genes from different functional classes and pathways to avoid co-regulation, which may skew the output of some of the algorithms used in RG analysis^{37,43}. Custom RT² Profiler PCR arrays were ordered from Qiagen (Cat No./ID 330,171, Toronto, ON, Canada) and used as per manufacturer's instructions. Selected RGs were further assessed for co-regulation using the Database

of Gene Co-Regulation (dGCR; www.dGCR.org)⁵⁵ based on data available for humans, mice, and rats from Affymetrix, Illumina, and Agilent (see Supplementary Tables S2 and S3).

Reverse Transcription Quantitative Real-time PCR (RT-qPCR). Custom 96-well (0.2 ml) RT² Profiler PCR arrays pre-coated with the primers for the 16 candidate RGs were purchased from Qiagen and RT-qPCR was performed following manufacturer's instructions on a QuantStudio3 Real-Time PCR System (Thermo Fisher Scientific, Applied Biosystems, Foster City, CA, USA). Briefly, a master-mix containing 4.5 ng of cDNA and 12.5 µl of RT² SYBR Green ROX Mastermix (Qiagen, Toronto, ON, Canada) per reaction was prepared for each sample and loaded into the array plates in triplicate for each of the 16 genes (48 reactions per sample). Plates were briefly centrifuged. Thermocycling profile consisted of an initial cycle of 95 °C for 10 min, followed by 40 cycles at 95 °C for 15 s and 1 min at 60 °C. Dissociation curves (95 °C for 15 s, 60 °C for 1 min, and 60 °C to 95 °C at 0.15 °C s⁻¹ with signal acquisition) were evaluated to confirm primer specificity. To correct for inter-plate variation one sample was repeated on all plates. All Qiagen RT² primers have been previously validated and optimized to work at uniform and near perfect efficiency to allow for simultaneous analysis of multiple genes in arrays. Primers are designed using proprietary algorithms and are subjected to rigorous testing for high performance: the average amplification efficiency of a representative set of assays for 4000 genes used in RT² Arrays was shown to be 99%, with a 95% confidence interval about the mean between 90 and 110%⁵⁶. Therefore, a base of 2 was used for relative quantity calculations^{38,43,57}.

Data analysis. Ct values and baseline corrections to account for inter-plate gene expression variation were calculated using Thermo Fisher Connect (Thermo Fisher Scientific, Applied Biosystems, Foster City, CA, USA). Relative quantification values (RQ) were calculated using $RQ = E^{(min\ Ct - sample\ Ct)}$ where E , the amplification efficiency, is 2 (assume 100% efficiency for commercially designed primers)^{38,43,57}, $min\ Ct$ is the sample with the lowest Ct value (highest expression) and $sample\ Ct$ is the Ct of each sample. This will result in a set of relative quantities between 1 and 0, where the sample with the lowest Ct is 1. The calculation ($min\ Ct - sample\ Ct$) is also referred to as ΔCt .

The calculated stability of candidate RGs and their relative ranking was compared across four statistical methods, geNorm, NormFinder, RefFinder, and BestKeeper. Calculated RQ of mean Ct values were input in geNorm (Version 3.5) to determine expression stability (M) for each gene, as well as the optimal number of RGs required by comparing pairwise variation (V)³⁶. Accepted cut-off values are $M < 0.5$ and $V < 0.15$; when the V_n/V_{n+1} ratio is less than 0.15, it indicates that the optimum number is n ³⁶. Two versions of NormFinder, Microsoft Excel add-in (Version 0.953) and NormFinder for R (Version 5, 2015–01-05), were evaluated using RQ values of median Ct as input for both versions to calculate stability for individual genes and gene combinations (<https://moma.dk/normfinder-software>)³⁷. RefFinder (<https://www.heartcure.com.au/reffinder>)³⁹ evaluates expression stability based on the algorithms of geNorm, NormFinder, BestKeeper, and delta Ct, generating a comprehensive ranking based on the geometric mean of ranking values. Input data format (median or mean Ct values) was not specified and so results from both methods were evaluated. Finally, BestKeeper (Version 1) requires input of raw median Ct values to determine the expression stability by using pairwise correlation analysis of all pairs of candidate genes³⁸. Standard deviations, coefficients of correlation [r] and p-values determined for each gene are used as measures of RG stability, with coefficients of correlation being the superior metric. BestKeeper allows for the assessment of up to 10 candidate RGs, therefore, the 6 genes most often ranking poorly according to the box-and-whiskers assessment and algorithms geNorm, NormFinder and RefFinder (*UBC*, *RAD50*, *HPRT1*, *B2M*, *HSP90A1*, and *RPS18*) were omitted from this analysis.

Data availability

Datasets generated and analysed during the current study are available in the figshare repository, <http://dx.doi.org/10.6084/m9.figshare.1266254064>.

Received: 11 August 2020; Accepted: 27 April 2021

Published online: 13 May 2021

References

- Mastromonaco, G. F., González-Grajales, L. A., Filice, M. & Comizzoli, P. Somatic cells, stem cells, and induced pluripotent stem cells: How do they now contribute to conservation? In *Reproductive Sciences in Animal Conservation Advances in Experimental Medicine and Biology* (eds Holt, W. V. et al.) 385–427 (Springer, New York, 2014).
- Freshney, R. I. *Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications* (Wiley, 2010).
- Villegas, J. & McPhaul, M. Establishment and culture of human skin fibroblasts. *Curr. Protoc. Mol. Biol.* **74**, 28–33 (2005).
- des Jardins-Park, H. E., Foster, D. S. & Longaker, M. T. Fibroblasts and wound healing: An update. *Regen. Med.* **13**, 491–495 (2018).
- Larson, B. J., Longaker, M. T. & Lorenz, H. P. Scarless fetal wound healing: A basic science review. *Plast. Reconstr. Surg.* **126**, 1172–1180 (2010).
- Gouveia, C., Huyser, C., Egli, D. & Pepper, M. S. Lessons learned from somatic cell nuclear transfer. *Int. J. Mol. Sci.* **21**, 2314 (2020).
- Campbell, K. H. et al. Somatic cell nuclear transfer: Past, present and future perspectives. *Theriogenology* **68S**, S214–231 (2007).
- Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
- Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Gu, B. W. et al. Impaired telomere maintenance and decreased canonical WNT signaling but normal ribosome biogenesis in induced pluripotent stem cells from X-linked dyskeratosis congenita patients. *PLoS ONE* **10**, e0127414 (2015).
- Freitas, B. C. G. et al. Stem cells and modeling of autism spectrum disorders. *Exp. Neurol.* **260**, 33–43 (2012).
- Chang, T. et al. Brief report: Phenotypic rescue of induced pluripotent stem cell-derived motoneurons of a spinal muscular atrophy patient. *Stem Cells* **29**, 2090–2093 (2011).

14. Marchetto, M. C. *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* **143**, 527–539 (2010).
15. Dimos, J. T. *et al.* Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. *Science* **321**, 1218–1221 (2008).
16. Park, I. H. *et al.* Disease-specific induced pluripotent stem cells. *Cell* **134**, 877–886 (2008).
17. Navarro, A. M., Susanto, E., Falk, A. & Wilhelm, M. Modeling cancer using patient-derived induced pluripotent stem cells to understand development of childhood malignancies. *Cell Death Discov.* **4**, 7 (2018).
18. Bustin, S. A. *et al.* The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).
19. Palmer, S. *et al.* New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* **41**, 4531–4536 (2003).
20. Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun.* **6**, 279–284 (2005).
21. Chapman, J. R. & Waldenström, J. With reference to reference genes: A systematic review of endogenous controls in gene expression studies. *PLoS ONE* **10**, e0141853 (2015).
22. Banfi, F., Colombini, A., Perucca Orfei, C., Parazzi, V. & Ragni, E. Validation of reference and identity-defining genes in human mesenchymal stem cells cultured under unrelated fetal bovine serum batches for basic science and clinical application. *Stem Cell Rev. Rep.* **14**, 837–846 (2018).
23. Panina, Y., Germond, A., Masui, S. & Watanabe, T. M. Validation of common housekeeping genes as reference for qPCR gene expression analysis during iPS reprogramming process. *Sci. Rep.* **8**, 8716 (2018).
24. Amable, P. R., Teixeira, M. V. T., Carias, R. B. V., Granjeiro, J. M. & Borojevic, R. Identification of appropriate reference genes for human mesenchymal cells during expansion and differentiation. *PLoS ONE* **8**, e73792 (2013).
25. Spalenza, V. *et al.* Identification of internal control genes for quantitative expression analysis by real-time PCR in bovine peripheral lymphocytes. *Vet. J.* **189**, 278–283 (2011).
26. Goossens, K. *et al.* Selection of reference genes for quantitative real-time PCR in bovine preimplantation embryos. *BMC Dev. Biol.* **5**, 27 (2005).
27. Roy, J. G., McElhaney, J. E. & Verschoor, C. P. Reliable reference genes for the quantification of mRNA in human T-cells and PBMCs stimulated with live influenza virus. *BMC Immunol.* **21**, 4 (2020).
28. Sugden, K., Pariante, C. M., McGuffin, P., Aitchison, K. J. & D'Souza, U. M. Housekeeping gene expression is affected by antidepressant treatment in a mouse fibroblast cell line. *J. Psychopharmacol.* **24**, 1253–1259 (2010).
29. Pérez, R., Tupac-Yupanqui, I. & Dunner, S. Evaluation of suitable reference genes for gene expression studies in bovine muscular tissue. *BMC Mol. Biol.* **9**, 79 (2008).
30. Nielsen, S. *et al.* Optimal reference genes for normalization of qPCR gene expression data from proton and photon irradiated dermal fibroblasts. *Sci. Rep.* **8**, 12688 (2018).
31. Zhou, W. *et al.* Transcript levels of several epigenome regulatory genes in bovine somatic donor cells are not correlated with their cloning efficiency. *Cloning Stem Cells* **11**, 397–405 (2009).
32. Anstaett, O. L., Brownlie, J., Collins, M. E. & Thomas, C. J. Validation of endogenous reference genes for RT-qPCR normalisation in bovine lymphoid cells (BL-3) infected with Bovine Viral Diarrhoea Virus (BVDV). *Vet. Immunol. Immunop.* **137**, 201–207 (2010).
33. Lisowski, P., Pierzchała, M., Gościk, J., Pareek, C. S. & Zwierzchowski, L. Evaluation of reference genes for studies of gene expression in the bovine liver, kidney, pituitary, and thyroid. *J. Appl. Genet.* **49**, 367–372 (2008).
34. Emam, M., Thompson-Crispi, K. & Mallard, B. The effect of immunological status, in-vitro treatment and culture time on expression of eleven candidate reference genes in bovine blood mononuclear cells. *BMC Immunol.* **16**, 33 (2005).
35. Zhou, W. *et al.* Global gene expression analysis of bovine blastocysts produced by multiple methods. *Mol. Reprod. Dev.* **75**, 744–758 (2008).
36. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, research0034 (2002).
37. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**, 5245–5250 (2004).
38. Pfaffl, M. W., Tichopad, A., Prgomet, C. & Neuvians, T. P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations. *Biotechnol. Lett.* **26**, 509–515 (2004).
39. 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.* **80**, 75–84 (2012).
40. Silver, N., Best, S., Jiang, J. & Thein, S. L. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Mol. Biol.* **7**, 33 (2006).
41. Hellemans, J. & Vandesompele, J. Selection of reliable reference genes for RT-qPCR analysis. In *Quantitative Real-Time PCR. Methods in Molecular Biology (Methods and Protocols)* (eds Biassoni, R. & Raso, A.) 19–26 (Humana Press, New York, 2014).
42. Rydbirk, R. *et al.* Assessment of brain reference genes for RT-qPCR studies in neurodegenerative diseases. *Sci. Rep.* **6**, 37116 (2016).
43. De Spiegelaere, W. *et al.* Reference gene validation for RT-qPCR, a note on different available software packages. *PLoS ONE* **10**, e0122515 (2015).
44. Mehta, R. *et al.* Validation of endogenous reference genes for qRT-PCR analysis of human visceral adipose samples. *BMC Mol. Biol.* **11**, 39 (2010).
45. Wu, H., Taki, F. A., Zhang, Y., Dobbins, D. L. & Pan, X. Evaluation and identification of reliable reference genes for toxicological study in *Caenorhabditis elegans*. *Mol. Biol. Rep.* **41**, 3445–3455 (2014).
46. Caradec, J. *et al.* 'Desperate house genes': The dramatic example of hypoxia. *Br. J. Cancer* **102**, 1037–1043 (2010).
47. Wang, X. *et al.* Identification and validation of appropriate reference genes for qRT-PCR analysis in *Corynebacterium glutamicum*. *FEMS Microbiol. Lett.* **365**, fy030 (2018).
48. Toorani, T., Mackie, P. M. & Mastro Monaco, G. F. Investigating markers of reprogramming potential in somatic cell lines derived from matched donors. *Cell Repogr.* <https://doi.org/10.1089/cell.2020.0075> (2021).
49. Masvidal, L. *et al.* *GUSB* and *ATP2B4* are suitable reference genes for *CFTR* gene expression data normalization in nasal epithelium cells. *J. Cyst. Fibros.* **11**, 398–404 (2012).
50. Razavi, S. A. *et al.* Validation of reference genes for normalization of relative qRT-PCR studies in papillary thyroid carcinoma. *Sci. Rep.* **9**, 15241 (2019).
51. González-Bermúdez, L., Anglada, T., Genescà, A., Martín, M. & Terradas, M. Identification of reference genes for RT-qPCR data normalisation in aging studies. *Sci. Rep.* **9**, 13970 (2019).
52. He, Y.-X., Zhang, Y., Yang, Q., Wang, C. & Su, G. Selection of suitable reference genes for reverse transcription-quantitative polymerase chain reaction analysis of neuronal cells differentiated from bone mesenchymal stem cells. *Mol. Med. Rep.* **12**, 2291–2300 (2015).
53. Bian, Z. *et al.* RPL13A as a reference gene for normalizing mRNA transcription of ovarian cancer cells with paclitaxel and 10-hydroxycamptothecin treatments. *Mol. Med. Rep.* **11**, 3188–3194 (2014).

54. Mastro Monaco, G. F., Perrault, S. D., Betts, D. H. & King, W. A. Role of chromosome stability and telomere length in the production of viable cell lines for somatic cell nuclear transfer. *BMC Dev. Biol.* **6**, 41 (2006).
55. Williams, G. Database of gene co-regulation (dGCR): A web tool for analysing patterns of gene co-regulation across publicly available expression data. *J. Genom.* **3**, 29–35 (2015).
56. Quellhorst, G. & Rulli, S. A *Systematic Guideline for Developing the Best Real-Time PCR Primers: Lessons Learned from Designing Assays for More Than 14,000 Genes*. <https://www.qiagen.com/us/resources/download.aspx?id=d6191d0e-701b-4eb1-bafa-d7ab7677875f&lang=en> (2012).
57. Kałużna, M., Kuras, A. & Puławska, J. Validation of reference genes for the normalization of the RT-qPCR gene expression of virulence genes of *Erwinia amylovora* in apple shoots. *Sci. Rep.* **7**, 2034 (2017).
58. Bionaz, M. & Looor, J. J. Identification of reference genes for quantitative real-time PCR in the bovine mammary gland during the lactation cycle. *Physiol. Genomics* **29**, 312–319 (2007).
59. Rekawiecki, R., Rutkowska, J. & Kotwica, J. Identification of optimal housekeeping genes for examination of gene expression in bovine corpus luteum. *Reprod. Biol.* **12**, 362–367 (2012).
60. Rekawiecki, R., Kowalik, M. K. & Kotwica, J. Validation of housekeeping genes for studying differential gene expression in the bovine myometrium. *Acta Vet. Hung.* **61**, 505–516 (2013).
61. Robert, C. *et al.* Quantification of housekeeping transcript levels during the development of bovine preimplantation embryos. *Biol. Reprod.* **67**, 1465–1472 (2002).
62. Robinson, T. L., Sutherland, I. A. & Sutherland, J. Validation of candidate bovine reference genes for use with real-time PCR. *Vet. Immunol. Immunopathol.* **115**, 160–165 (2007).
63. Ross, P. J., Wang, K., Kocabas, A. & Cibelli, J. B. Housekeeping gene transcript abundance in bovine fertilized and cloned embryos. *Cell. Reprogram.* **12**, 709–717 (2010).
64. Toorani, T., Mackie, P. M. & Mastro Monaco, G. F. Validation of reference genes for use in untreated bovine fibroblasts RT-qPCR dataset. Figshare <https://doi.org/10.6084/m9.figshare.12662540> (2021).

Acknowledgements

This project was supported by grant# RGPIN-2015-05116 from the Natural Sciences and Engineering Research Council of Canada (G.F.M.). The authors wish to thank Dr. Mehdi Emam for his technical advice.

Author contributions

T.T. conceived and designed the reference gene pilot experiment. All experimental work was performed by T.T, aside from the loading of array plates, which was done by P.M.M. T.T. analyzed the data and wrote the manuscript. G.F.M. supervised the project and provided funding. All authors have reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-89657-8>.

Correspondence and requests for materials should be addressed to G.F.M.

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

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021