Toward Enhanced MIQE Compliance: Reference Residual Normalization of qPCR Gene Expression Data

Richard C. Edmunds,[1] Jenifer K. McIntyre,[2] J. Adam Luckenbach,[3,4] David H. Baldwin,[1] and John P. Incardona[1]

[1]Ecotoxicology Program, Environmental Conservation Division, and [3]Environmental Physiology Program, Resource Enhancement and Utilization Technologies Division, Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, Seattle, Washington 98112, USA; [2]Washington State University Puyallup Research and Extension Center, Puyallup, Washington 98371, USA; and [4]Center for Reproductive Biology, Washington State University, Pullman, Washington 99164, USA

Normalization of fluorescence-based quantitative real-time PCR (qPCR) data varies across quantitative gene expression studies, despite its integral role in accurate data quantification and interpretation. Identification of suitable reference genes plays an essential role in accurate qPCR normalization, as it ensures that uncorrected gene expression data reflect normalized data. The reference residual normalization (RRN) method presented here is a modified approach to conventional $2^{-\Delta\Delta Ct}$qPCR normalization that increases mathematical transparency and incorporates statistical assessment of reference gene stability. RRN improves mathematical transparency through the use of sample-specific reference residuals (RR_i) that are generated from the mean C_t of one or more reference gene(s) that are unaffected by treatment. To determine stability of putative reference genes, RRN uses ANOVA to assess the effect of treatment on expression and subsequent equivalence-threshold testing to establish the minimum permitted resolution. Step-by-step instructions and comprehensive examples that demonstrate the influence of reference gene stability on target gene normalization and interpretation are provided. Through mathematical transparency and statistical rigor, RRN promotes compliance with Minimum Information for Quantitative Experiments and, in so doing, provides increased confidence in qPCR data analysis and interpretation.

KEY WORDS: ANOVA, normalization, reference genes, threshold cycle, treatment effect

INTRODUCTION

Normalization method plays an essential role in the accurate quantification and interpretation of gene expression data. However, fluorescence-based quantitative real-time PCR (qPCR) normalization methodology is not standardized across gene expression studies, despite attempts to do so (e.g., Minimum Information for Publication of Quantitative Real-Time PCR Experiments).[1,2] To address this issue, we developed a mathematically explicit and statistically rigorous approach to qPCR normalization that enhances $2^{-\Delta\Delta Ct}$ methodology[3,4] and Minimum Information for Quantitative Experiments (MIQE) compliance. The reference residual normalization (RRN) method presented here improves mathematical transparency through the use of sample-specific reference residuals (RR_i), generated from the mean C_t of one or more reference gene(s), to normalize corresponding C_t values (C_{ti}) of one or more target gene(s). RRN also incorporates statistical support (i.e., *P* value) for putative reference genes by determining the effect of treatment (e.g., ANOVA) and minimum permitted resolution (e.g., equivalence threshold test). Through mathematical transparency and statistical rigor, RRN promotes compliance with MIQE to provide increased confidence in qPCR normalization and interpretation.

Data used in examples (see Tables 1–4) were recently generated as part of a larger, yet unpublished, gene expression study on developing zebrafish by R. Edmunds and J. Incardona. Zebrafish embryos were obtained from adults maintained at the Northwest Fisheries Science Center (NOAA; Seattle, WA, USA) using conventional zebrafish animal care protocols.[5] cDNA was synthesized from a normalized quantity (1 μg) of total RNA. All C_t values were determined using a fixed threshold fluorescence of 0.1

ADDRESS CORRESPONDENCE TO: Richard C. Edmunds, Ecotoxicology Program, Environmental Conservation Division, Northwest Fisheries Science Center, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. E., Seattle, WA 98112, USA (Phone: +1 206 860 3310; Fax: +1 206 3335; E-mail: richard.c.edmunds@gmail.com).
doi: 10.7171/jbt.14-2502-003

TABLE 1

Statistical Tests of Treatment on the Expression of Two Technical Reference Genes Using Log_2-Transformed Fold-Change Values

	One-way ANOVA			Dunnett's post-hoc (*P* value)		Expression		Equivalence		
Gene	F ratio	df	*P* value	Control versus T1	Control versus T2	Fold change	Log_2	Threshold	Resolution	Generate RR_i
Ref 1	0.6109	2,11	0.5603	0.5214	0.5401	1.14	0.20	0.56	1.5-fold	Yes
Ref 2	6.8219	2,11	0.0118	0.0113	0.0173	1.77	0.83	1.33	2.5-fold	No

Example of statistically testing two putative technical reference genes (Ref) for an effect of treatment after exposure to low and high treatment concentrations [Treatments 1 and 2 (T1 and T2), respectively], using one-way ANOVA ($\alpha=0.10$) and Dunnett's post-hoc comparisons. Equivalence threshold is the minimum value required for significant equivalence ($P<0.05$) across treatments. This threshold establishes the level of resolution for target-gene responsiveness to treatment that is permitted by reference gene(s) used for normalization. Reference gene expression is considered unaffected when $P > 0.10$ and statistically equivalent if the difference is less than or equal to the equivalence threshold. Statistical analyses conducted using JMP version 10. df, Degrees of freedom.

across all qPCR runs. Gene-specific efficiencies (E) were established off the slopes of standard curves generated from serial dilution of cDNA and pooled equally across all samples ($n=14$).[6] Expression data were collected using Power SYBR Green chemistry (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) on a 7700 Prism (Applied Biosystems, Life Technologies), running standard cycling conditions.[6]

Reference genes (technical and biological) can be used to correct gene expression data for technical variance (e.g., reverse transcription and/or loading differences) as well as biological variance (e.g., cell number and/or RNA population changes), respectively. Normalization to a biological reference gene can be done as a preliminary step before conventional $\Delta\Delta C_t$ normalization (i.e., $\Delta\Delta\Delta C_t$). This type of normalization is most appropriate when unaffected technical reference genes cannot be identified, suggesting substantial sample-to-sample variation in the starting mRNA population (e.g., samples collected across developmental time or tissue type).[7] Given the variability in selection and use of biological reference genes across laboratories and experimental paradigms, explicit examples of $\Delta\Delta\Delta C_t$ normalization are not provided herein.

Regardless of intended application, expression of biological and/or technical reference genes across treatments should be accompanied by statistical support. The expectation is that biological reference gene expression is affected by treatment, developmental stage, or tissue type, whereas technical reference gene expression is unaffected. Statistical tests of biological and technical reference gene expression are of critical importance, as normalization to affected[6] and unaffected[8] reference genes results in the most biologically accurate conclusions, respectively. More specifically, fold-change values are calculated using $2^{\Delta Ct}$ or $E^{\Delta Ct}$ (100% or reaction-specific E, respectively) and log_2-transformed to restore normal distribution before statistical testing of treatment effect (Table 1 and Fig. 1, Step 1). Conventional methods of means comparison (e.g., Student's *t*-test for control vs. one treatment or one-way ANOVA with Dunnett's post-hoc for control vs. two or more treatments) can be used to determine whether reference genes are considered affected ($P<0.05$) or unaffected ($P>0.10$) by treatment (Fig. 1, Step 1.5.1).

Given the importance of identifying technical reference genes that are unaffected by treatment for accurate normalization, increasing α (e.g., $\alpha=0.10$) and consequently decreasing β are recommended for avoidance of Type II error. Moreover, technical reference genes identified as unaffected by treatment should be subjected to an equivalence test to determine statistically the smallest difference in log_2-transformed fold change (e.g., 0.56) required to obtain significant equivalence ($P<0.05$; Table 1). This threshold establishes the minimum change in target gene expression (e.g., 1.5-fold; Table 1) necessary to be considered as a biologically significant response to treatment (i.e., regardless of statistical significance). Note that if preliminary biological reference gene normalization (i.e., $\Delta\Delta\Delta C_t$) is appropriate, then biological RR_i should be applied to C_{ti} values of technical reference gene(s) before expression calculations and statistical tests are conducted (Fig. 1).

Table 1 presents a treatment-effect test for two putative technical reference genes using one-way ANOVA and Dunnett's post-hoc comparisons of log_2-transformed C_t values. In this example, Reference gene 1 expression can be considered unaffected, as there is no significant treatment effect for Treatment 1 ($F=0.6109$, $P=0.5603$) or Treatment 2 ($P=0.5214$ or 0.5401). Subsequent equivalence testing demonstrates that a difference up to 0.56 (1.5-fold) is significantly equivalent ($P<0.05$) between both treatments. However, Reference gene 2 expression cannot be considered unaffected by treatment for Treatment 1 ($F=6.8219$, $P=0.0118$) or Treatment 2 ($P=0.0113$ or 0.0173). For the purpose of this example, subsequent

Step 1: Determine suitability of reference gene(s)

1.1. Calculate mean C_t for replicates of control treatment ($\overline{x}C_{tc}$)
1.2. Calculate difference in C_t from $\overline{x}C_{tc}$ for each control and treatment replicate (ΔC_{ti})
1.3. If efficiency (E) is known, calculate fold-change for each replicate of each treatment ($E^{\Delta C_{ti}}$)
1.3.1. If E is unknown, use $2^{\Delta C_{ti}}$ in Step 1.3
1.4. Log_2-transform fold-change values generated in Step 1.3
1.5. Compare log_2-transformed C_{ti} values among treatments
1.5.1. Use ANOVA for control vs. ≥ 2 treatments or t-test for control vs. 1 treatment
1.6. Consider technical reference gene(s) unaffected by treatment if $P > 0.10$
1.6.1. Consider biological reference gene(s) affected by treatment if $P < 0.05$
1.7. Determine minimum equivalence theshold for log_2-transformed fold-changes ($P \leq 0.05$)
1.7.1. Minimum equivalence theshold is the minimum resolution of target gene quantification

Step 2: Generate reference residuals (RR_i):

2.1. Calculate mean C_t for all control and treatment replicates ($\overline{x}C_t$)
2.1.1. For multiple reference genes, calculate mean C_{ti} values before continuing to Step 2.2
2.1.2. Geometric mean of ≥ 2 reference genes can be obtained using BestKeeper software[9]
2.2. Subtract $\overline{x}C_t$ from C_{ti} for control and treatment replicates to generate RR_i values

Step 3: Normalize target gene C_t values using RR_i:

3.1. Add RR_i values to C_{ti} values for each target gene
3.2. Repeat Steps 1.1-1.5 using RR_i normalized C_{ti} values for each target gene
3.3. Consider target genes as affected by treatment if $P < 0.05$

FIGURE 1

Step-by-step instructions for applying RRN to qPCR data. RRN can be applied to target gene(s) using biological and technical reference genes together (i.e., $\Delta\Delta\Delta C_t$) or technical reference gene(s) alone (i.e., $\Delta\Delta C_t$).

equivalence testing demonstrates that a difference up to 1.325 (2.5-fold) is significantly equivalent ($P<0.05$) between both treatments.

Consideration must also be given to the difference in standard error of the mean (SE) associated with differences in endogenous transcript abundance of technical reference genes (i.e., low- vs. high-abundance transcripts). The C_t values for low-abundance (e.g., tissue-specific loci within a mixed-tissue RNA population) technical reference genes generally exhibit larger standard error (±0.2 to ±1 SE), which translates into statistical stability ($P>0.10$) and results in larger technical RR_i values than do high-abundance (e.g., ubiquitous) technical reference genes. High-abundance technical reference genes can appear affected by treatment ($P<0.10$), as a result of an inherently small error (±0.03 to ±0.10 SE); however, subsequent equivalence testing may establish that a practical equivalence threshold is acceptably low (e.g., ≤1.5-fold). Conversely, if biological and/or technical reference genes exhibit large variation (≥1 SE) and thus, appear unaffected by treatment (e.g., $P>0.10$), subsequent equivalence testing may establish that the practical equivalence threshold is unacceptably high (e.g., greater than two-fold). Therefore, gene expression studies designed to quantify multiple target genes that exhibit a tissue-specific-to-ubiquitous range of endogenous abundance are encouraged to normalize using RR_i generated from the geometric mean of two or more technical reference genes that exhibit a similar tissue-specific-to-ubiquitous range of endogenous abundance. If preliminary normalization to a biological reference gene is appropriate, the use of a high-abundance transcript is recommended, as $\Delta\Delta\Delta C_t$ influences C_{ti} values of technical reference and target genes.

Following the identification of suitable technical reference gene(s), RR_i can be generated from the overall mean of control and treatment C_t values (Fig. 1, Step 2). More specifically, RR_i captures the inherent variance of technical reference genes by subtracting the overall mean C_t from the C_t of each control and treatment replicate, respectively (see Table 2 for example of RR_i generation). Generated RR_i

TABLE 2

Generation of RR_i from Uncorrected C_{ti} Values for One Unaffected (Ref 1) and One Affected (Ref 2) Reference Gene

Control	C_{ti}	RR_i generation	RR_i (Ctrl)	Corrected	Treatment 1	C_{ti}	RR_i generation	RR_i (T1)	Corrected	Treatment 2	C_{ti}	RR_i generation	RR_i (T2)	Corrected
Ref 1 (E=1.95)														
Biological Rep 1	20.24	Overall AVG—Rep 1	0.16	20.39	Biological Rep 1	20.77	Overall AVG—Rep 1	−0.38	20.39	Biological Rep 1	20.11	Overall AVG—Rep 1	0.28	20.39
Biological Rep 2	20.77	Overall AVG—Rep 2	−0.38	20.39	Biological Rep 2	20.13	Overall AVG—Rep 2	0.26	20.39	Biological Rep 2	20.06	Overall AVG—Rep 2	0.33	20.39
Biological Rep 3	20.30	Overall AVG—Rep 3	0.09	20.39	Biological Rep 3	19.93	Overall AVG—Rep 3	0.46	20.39	Biological Rep 3	20.24	Overall AVG—Rep 3	0.15	20.39
Biological Rep 4	20.83	Overall AVG—Rep 4	−0.44	20.39	Biological Rep 4	20.28	Overall AVG—Rep 4	0.11	20.39	Biological Rep 4	20.70	Overall AVG—Rep 4	−0.31	20.39
					Biological Rep 5	20.53	Overall AVG—Rep 5	−0.14	20.39	Biological Rep 5	20.56	Overall AVG—Rep 5	−0.17	20.39
Overall AVG	20.39			20.39	Average	20.33			20.39	Average	20.30			20.39
SD	0.30			0.00	SD	0.33			0.00	SD	0.30			0.00
SE	0.08			0.00	SE	0.15			0.00	SE	0.10			0.00
Ref 2 (E=1.97)														
Biological Rep 1	21.10	Overall AVG—Rep 1	−0.52	20.58	Biological Rep 1	21.02	Overall AVG—Rep 1	−0.44	20.58	Biological Rep 1	20.25	Overall AVG—Rep 1	0.33	20.58
Biological Rep 2	21.08	Overall AVG—Rep 2	−0.50	20.58	Biological Rep 2	20.42	Overall AVG—Rep 2	0.16	20.58	Biological Rep 2	19.94	Overall AVG—Rep 2	0.64	20.58
Biological Rep 3	21.00	Overall AVG—Rep 3	−0.42	20.58	Biological Rep 3	19.62	Overall AVG—Rep 3	0.97	20.58	Biological Rep 3	20.37	Overall AVG—Rep 3	0.21	20.58
Biological Rep 4	21.58	Overall AVG—Rep 4	−1.00	20.58	Biological Rep 4	20.14	Overall AVG—Rep 4	0.44	20.58	Biological Rep 4	20.53	Overall AVG—Rep 4	0.05	20.58
					Biological Rep 5	20.35	Overall AVG—Rep 5	0.23	20.58	Biological Rep 5	20.79	Overall AVG—Rep 5	−0.21	20.58
Overall AVG	20.58			20.58	Average	20.31			20.58	Average	20.40			20.58
SD	0.53			0.00	SD	0.51			0.00	SD	0.30			0.00
SE	0.14			0.00	SE	0.23			0.00	SE	0.10			0.00

Example of RR generation from uncorrected C_t values for one unaffected (Ref 1) and one affected (Ref 2) technical reference gene. RRN uses sample-specific reference residuals (RR_i) to normalize sample-specific C_t values (C_{ti}) of target genes for the variation inherent in Ref 1 and Ref 2 C_{ti} values. Ctrl, control; Rep, replicate; AVG, average.

TABLE 3

Application of Sample-specific Reference Residuals (RR_i) and Fold-Change Calculation ($E^{\Delta\Delta Ct}$) for Target Genes with and without Normalization

	Uncorrected C_{ti} values				Ref 1-corrected C_{ti} values					Ref 2-corrected C_{ti} values				
Control	C_{ti}	Δ	Fold change	Log_2	RR_i	Corrected C_{ti}	Δ	Fold change	Log_2	RR_i	Corrected C_{ti}	Δ	Fold change	Log_2
Target 1 (E=1.96)														
Biological Rep 1	27.20	−1.25	0.43	−1.21	0.16	27.36	−1.55	0.35	−1.50	−0.52	26.69	−1.34	0.41	−1.30
Biological Rep 2	24.33	1.62	2.98	1.58	−0.38	23.96	1.86	3.49	1.80	−0.50	23.83	1.51	2.77	1.47
Biological Rep 3	25.96	−0.01	0.99	−0.01	0.09	26.05	−0.24	0.85	−0.23	−0.42	25.55	−0.20	0.87	−0.20
Biological Rep 4	26.32	−0.36	0.78	−0.35	−0.44	25.88	−0.07	0.96	−0.07	−1.00	25.32	0.03	1.02	0.03
Average	26.00		1.30	0.00		25.80		1.40	0.00		25.30		1.30	0.00
SD	1.20		1.10	1.20		1.40		1.40	1.40		1.20		1.00	1.10
SE	0.60		0.60	0.60		0.70		0.70	0.70		0.60		0.50	0.60
Treatment 1														
Biological Rep 1	20.64	5.31	35.70	5.16	−0.38	20.26	5.55	41.90	5.39	−0.44	20.20	5.15	31.91	5.00
Biological Rep 2	19.97	5.98	55.96	5.81	0.26	20.23	5.58	42.67	5.42	0.16	20.13	5.22	33.47	5.06
Biological Rep 3	19.29	6.66	88.60	6.47	0.46	19.75	6.06	58.94	5.88	0.97	20.26	5.09	30.75	4.94
Biological Rep 4	18.26	7.69	177.20	7.47	0.11	18.37	7.44	149.69	7.23	0.44	18.70	6.65	87.67	6.45
Biological Rep 5	18.16	7.79	189.41	7.57	−0.14	18.03	7.79	188.68	7.56	0.23	18.39	6.95	107.71	6.75
Average	19.27		109.37	6.49		19.33		96.38	6.29		19.53		58.30	5.64
SD	1.08		70.21	1.04		1.06		68.22	1.03		0.91		36.66	0.88
SE	0.48		31.40	0.47		0.47		30.51	0.46		0.41		16.40	0.40
Treatment 2														
Biological Rep 1	20.20	5.76	48.13	5.59	0.28	20.47	5.34	36.31	5.18	0.33	20.53	4.82	25.55	4.68
Biological Rep 2	19.50	6.45	76.87	6.26	0.33	19.83	5.98	55.85	5.80	0.64	20.14	5.21	33.27	5.06
Biological Rep 3	20.42	5.53	41.42	5.37	0.15	20.57	5.24	33.94	5.09	0.21	20.63	4.71	23.87	4.58
Biological Rep 4	19.69	6.27	67.88	6.08	−0.31	19.38	6.43	75.96	6.25	0.05	19.74	5.61	43.48	5.44
Biological Rep 5	21.01	4.94	27.81	4.80	−0.17	20.85	4.97	28.27	4.82	−0.21	20.80	4.55	21.32	4.41
Average	20.16		52.42	5.62		20.22		46.07	5.43		20.37		29.50	4.83
SD	0.60		19.89	0.59		0.60		19.68	0.58		0.43		9.00	0.41
SE	0.27		8.89	0.26		0.27		8.80	0.26		0.19		4.02	0.19
Target 2 (E=1.97)														
Biological Rep 1	29.34	0.00	1.00	0.00	0.16	29.49	−0.77	0.60	−0.75	−0.52	28.82	0.37	1.29	0.36
Biological Rep 2	28.11	1.23	2.30	1.20	−0.38	27.73	1.00	1.97	0.98	−0.50	27.61	1.59	2.93	1.55
Biological Rep 3	29.93	−0.59	0.67	−0.58	0.09	30.02	−1.29	0.42	−1.26	−0.42	29.51	−0.32	0.81	−0.31
Biological Rep 4	29.97	−0.64	0.65	−0.62	−0.44	29.53	−0.81	0.58	−0.79	−1.00	28.97	0.22	1.16	0.22
Average	29.30		1.20	0.00		29.20		0.90	−0.50		28.70		1.50	0.50
SD	0.90		0.80	0.90		1.00		0.70	1.00		0.80		0.90	0.80
SE	0.40		0.40	0.40		0.50		0.40	0.50		0.40		0.50	0.40
Treatment 1														
Biological Rep 1	28.16	1.17	2.21	1.15	−0.38	27.78	1.41	2.60	1.38	−0.44	27.72	1.01	1.98	0.98
Biological Rep 2	28.58	0.75	1.67	0.74	0.26	28.84	0.35	1.27	0.34	0.16	28.74	−0.01	0.99	−0.01
Biological Rep 3	26.41	2.92	7.25	2.86	0.46	26.88	2.32	4.81	2.27	0.97	27.38	1.35	2.50	1.32
Biological Rep 4	26.36	2.98	7.53	2.91	0.11	26.47	2.73	6.35	2.67	0.44	26.80	1.93	3.71	1.89
Biological Rep 5	26.40	2.94	7.32	2.87	−0.14	26.26	2.93	7.29	2.87	0.23	26.63	2.10	4.14	2.05
Average	27.20		5.20	2.10		27.20		4.50	1.90		27.50		2.70	1.20
SD	1.10		3.00	1.10		1.10		2.50	1.00		0.80		1.30	0.80
SE	0.50		1.30	0.50		0.50		1.10	0.50		0.40		0.60	0.40
Treatment 2														
Biological Rep 1	27.58	1.76	3.29	1.72	0.28	27.86	1.34	2.48	1.31	0.33	27.91	0.81	1.74	0.80
Biological Rep 2	26.55	2.79	6.62	2.73	0.33	26.88	2.31	4.80	2.26	0.64	27.18	1.54	2.85	1.51
Biological Rep 3	27.77	1.57	2.90	1.54	0.15	27.92	1.27	2.37	1.25	0.21	27.98	0.75	1.66	0.73
Biological Rep 4	26.70	2.64	5.98	2.58	−0.31	26.39	2.80	6.69	2.74	0.05	26.75	1.97	3.82	1.93
Biological Rep 5	28.59	0.75	1.66	0.73	−0.17	28.42	0.77	1.69	0.76	−0.21	28.37	0.35	1.27	0.35
Average	27.40		4.10	1.90		27.50		3.60	1.70		27.60		2.30	1.10
SD	0.80		2.10	0.80		0.80		2.10	0.80		0.70		1.00	0.60
SE	0.40		0.90	0.40		0.40		0.90	0.40		0.30		0.50	0.30

Example of RR_i application before fold-change ($E^{\Delta\Delta Ct}$) and log_2-transformation calculations of Targets 1 and 2 (low- and high-abundance transcripts, respectively), using uncorrected, Ref 1-corrected, and Ref 2-corrected C_t values.

TABLE 4

Statistical Tests of Uncorrected Versus Ref 1- and Ref 2-Corrected Target Gene Expression

	One-way ANOVA			Dunnett's post-hoc (*P* value)	
	F ratio	df	*P* value	Control versus T1	Control versus T2
Gene-uncorrected					
Target 1	59.7855	2,11	<0.0001	<0.0001	<0.0001
Target 2	6.6404	2,11	0.0128	0.0109	0.0219
Ref 1-Corrected					
Target 1	48.3895	2,11	<0.0001	<0.0001	<0.0001
Target 2	5.1365	2,11	0.0266	0.0219	0.0425
Ref 2-Corrected					
Target 1	57.0851	2,11	<0.0001	<0.0001	<0.0001
Target 2	3.4418	2,11	0.0691	0.0547	0.1012

Example of an appropriate statistical test on detected gene expression patterns of Targets 1 and 2 (low and high abundance, respectively) after treatment using one-way ANOVA and Dunnett's post-hoc comparisons. Treatments 1 and 2 represent low and high concentration treatments, respectively. Expression is considered significant when $P < 0.05$. Statistical analyses conducted using JMP version 10. df, Degrees of freedom.

values are positive or negative, depending on how the C_t value of a given replicate differs from the overall mean C_t value. If normalization to a biological reference gene is appropriate, then generation and application of biological RR_i to corresponding technical reference and target C_{ti} values should be completed as a preliminary step (Fig. 1, Step 2.1).

Technical RR_i can also be generated from multiple reference genes by calculating the mean C_t for each individual sample (e.g., geometric mean)[9,10] before calculating the overall mean of control and treatment C_t values (Fig. 1, Steps 2.1.1 and 2.1.2). Note that RR_i can be generated from the median instead of the mean if nonparametric statistics are considered needed and used to establish suitability of technical reference gene(s).

Once RR_i have been generated, qPCR expression data can be normalized using the RRN method (Fig. 1, Step 3, and Table 3). The application of RR_i to target C_{ti} values corrects for the inherent variance in technical reference gene expression. Table 3 provides a comprehensive example of the mathematically explicit method used by RRN for target gene normalization to affected and unaffected technical reference genes. Normalization of Target 1 and Target 2 to the unaffected technical reference gene (Reference 1) results in ~88% and ~87% retention of uncorrected expression, whereas normalization to the affected technical reference gene (Reference 2) results in only ~55% and ~53% retention of uncorrected expression, respectively (Table 3). Note that the influence of biological reference gene normalization can be larger than technical normalization alone, as it normalizes for biological variation in both technical reference and target gene expression patterns (e.g., differing number of cells in tissue samples with identical weights).[7] The downstream influence of this alteration on data interpretation depends on the magnitude of target gene expression, with assays targeting subtle changes in expression (e.g., less than five-fold; Table 3, Target 2) being more sensitive to a reduction in expression after normalization than assays targeting robust changes (e.g., greater than five-fold; Table 3, Target 1). Accordingly, consideration must also be given to the direction of expression for both technical reference and target genes as their interactions are additive.[1] More specifically, concurrent normalization (i.e., up-regulated target gene normalized to up-regulated technical reference) results in a reductive influence on target gene expression, whereas opposing normalization (i.e., up-regulated target gene normalized to down-regulated technical reference) results in an additive influence.

To avoid incorrect downstream conclusions, as a result of such influence of technical reference gene normalization on target gene expression pattern(s), statistical testing of uncorrected and corrected $\log_2$-transformed fold-change values is recommended (Fig. 1, Steps 1.5.1 and 3.3, and Table 4). Table 4 presents an example of ANOVA analyses on uncorrected and corrected $\log_2$-transformed fold-change values that demonstrate the impact that technical reference gene normalization can have on data interpretation. More specifically, both Targets 1 and 2 exhibit significant changes in expression after treatment when uncorrected ($F_{2,11}=59.7855$, $P<0.0001$; $F_{2,11}=6.6404$, $P=0.0128$) and Reference 1-corrected ($F_{2,11}=48.3895$, $P<0.0001$; $F_{2,11}=5.1365$, $P=0.0266$) values are considered; however, the significant change in Target 2 expression is lost after Reference 2 correction ($F_{2,11}=3.4418$, $P<0.0691$), respectively. This loss of significance is a result of the effect of treatment on Reference 2 expression, which carries through the entire normalization procedure (i.e., significant ANOVA, $P=0.0118$; reduction in retention of

uncorrected expression, ~50%; and loss of significance in Target 2 expression after Treatments 1 and 2, $P=0.0547$ and 0.1012 respectively). Table 4 presents parallel statistical analyses on uncorrected and corrected values to demonstrate the importance of normalizing to technical reference gene(s) that are unaffected by treatment, especially for quantitative molecular studies targeting genes that exhibit subtle (less than five-fold) changes in response to treatment (e.g., tissue-specific mRNAs, non-coding RNAs, or micro RNAs).[11–13]

In summary, RRN improves transparency and confidence regarding qPCR analysis and interpretation. This method uses explicit mathematics and rigorous statistics to increase confidence in identification of suitable reference gene(s), which is essential for accurate normalization of target genes. Through enhanced MIQE compliance and transparency, the use of RRN improves the biological accuracy of quantitative gene expression studies.

ACKNOWLEDGMENTS

The authors thank T. Linbo, A. Aquilina-Beck, and J. T. Dickey for technical assistance; N. Scholz and P. Swanson for guidance; and all members of the NOAA Ecotoxicology Program for unwavering support.

DISCLOSURE

The authors declare no conflict of interest concerning financial support or associations.

REFERENCES

1. Bustin SA, Benes V, Garson JA, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* 2009;55:611–622.
2. Bustin SA. Why the need for qPCR publication guidelines? —*The case for MIQE. Methods* 2010;50:217–226.
3. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 2001;25:402–408.
4. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative Ct method. *Nat Protoc* 2008;3:1101–1108.
5. Linbo TL, Baldwin DH, McIntyre JK, Scholz NL. Effects of water hardness, alkalinity, and dissolved organic carbon on the toxicity of copper to the lateral line of developing fish. *Environ Toxicol* 2009;28:1455–1461.
6. Metzger DC, Luckenbach JA, Shimizu M, Beckman BR. Normalizing for biology: accounting for technical and biological variation in levels of reference gene and insulin-like growth factor 1 (*igf1*) transcripts in fish livers. *Comp Biochem Physiol A* 2010;163:7–14.
7. Tang R, Dodd A, Lai D, McNabb WC, Love DR. Validation of zebrafish (*Danio rerio*) reference genes for quantitative real-time RT-PCR normalization. *Acta Biochim Biophys Sin (Shanghai)* 2007;39:384–390.
8. McCurley AT, Callard GV. Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. *BMC Mol Biol* 2008;9:102.
9. 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;26:509–515.
10. Vandesompele J, De Preter, K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 2002; 3:RESEARCH0034.
11. Etheridge A, Gomes CP, Pereira RW, Galas D, Wang K.The complexity, function, and applications of RNA in circulation. *Front Genet* 2013;4:115.
12. Mattick JS. The genetic signatures of noncoding RNAs. *PLoS Genet* 2009;5:e1000459.
13. Mattick JS, Makunin IV. Small regulatory RNAs in mammals. *Hum Mol Genet* 2005;14(Spec No 1):R121–R132.