Analytical Biochemistry 379 (2008) 127–129

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Notes & Tips

Standardization of real-time PCR gene expression data from independent biological replicates

Erik Willems [a,b,*], Luc Leyns [b], Jo Vandesompele [c,*]

[a] Del E. Webb Neuroscience, Aging, and Stem Cell Research Center, Burnham Institute for Medical Research, La Jolla, CA 92037, USA
[b] Laboratory for Cell Genetics, Vrije Universiteit Brussel, B-1050 Brussels, Belgium
[c] Center for Medical Genetics, Ghent University Hospital, B-9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 21 February 2008
Available online 26 April 2008

ABSTRACT

Gene expression analysis by quantitative reverse transcription PCR (qRT–PCR) allows accurate quantifications of messenger RNA (mRNA) levels over different samples. Corrective methods for different steps in the qRT–PCR reaction have been reported; however, statistical analysis and presentation of substantially variable biological repeats present problems and are often not meaningful, for example, in a biological system such as mouse embryonic stem cell differentiation. Based on a series of sequential corrections, including log transformation, mean centering, and autoscaling, we describe a robust and powerful standardization method that can be used on highly variable data sets to draw statistically reliable conclusions.

Published by Elsevier Inc.

Gene expression analysis by quantitative reverse transcription PCR (qRT–PCR)[1] allows accurate and sensitive measurement of gene expression levels. qRT–PCR has been used to demonstrate alterations in gene expression in a wide range of applications such as in genetic modification of crops, in cancers, during induced immune responses, and during differentiation of embryonic stem (ES) cells.

Because this method enables high-resolution and sensitive quantification, corrections for variations in the qRT–PCR workflow are required to obtain reliable results. Efforts have been made to standardize the technical variability of the PCR reaction by inclusion of a template normalization step, preferably using multiple stably expressed reference genes, and by corrections for PCR efficiency and interrun variations, resulting in more reliable experimental data [1–4].

However, standardization of variable experimental replicates caused by inherent biological variability has not been well studied, nor has an adequate workaround been proposed. Nevertheless, assessment of statistical significance requires a standardized data set because variation in data from multiple replicates (e.g., independent biological replicates of an experiment) might not result in statistically significant differences, even though the biological effect is clearly discernable in each of the individual replicates. Here, we describe a data standardization procedure that can be applied to data sets that display high variation between biological replicates, enabling proper statistical analysis and drawing relevant conclusions. The proposed data processing procedure is not new; it has been used successfully for microarray data transformation, for example, in exploratory analyses such as clustering (in which the magnitude of differences between the genes should be disregarded to attribute equal weight to all genes). However, this procedure has not been recognized as a robust and powerful method to standardize independent biological replicate experiments so as to draw statistically sound conclusions from a data set that otherwise is of limited value due to high interexperimental variation.

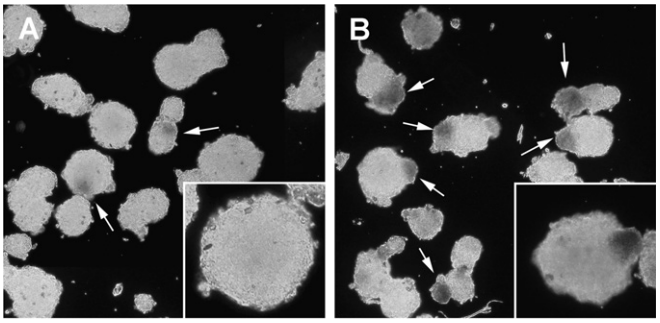
ES cell differentiation is generally accepted as a biological system that is often subject to significant experimental variation between independent replicates because differentiation of ES cells is not fully controllable even by the addition of specific growth factors. Differences between biological repeats can occur in these cultures because of the use of different passages of ES cells or by spontaneous differentiation, leading to substantial variation between biological replicates while showing a similar trend toward a particular differentiation status.

By using *Brachyury* induction by Activin A in embryoid bodies (EBs) as an example of mouse ES (mES) cell differentiation toward mesoderm, we illustrate the experimental variability of biological replicates in this system (Fig. 1). *Brachyury* induction by 3 ng/ml Activin A is demonstrated in all experiments using whole-mount in situ hybridization (WISH) of EBs, whereas higher concentrations (30 ng/ml) reduce *Brachyury* expression in each repeat, corroborating previous studies [5–7] (Figs. 1A and 1B). When performing qRT–PCR on three independent biological repeats run in the same

* Corresponding authors.
E-mail addresses: ewillems@burnham.org (E. Willems), joke.vandesompele@ugent.be (J. Vandesompele).

[1] *Abbreviations used:* qRT–PCR, quantitative reverse transcription PCR; ES, embryonic stem; EB, embryoid body; mES, mouse ES; WISH, whole-mount in situ hybridization; mRNA, messenger RNA; CI, confidence interval.

0003-2697/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.ab.2008.04.036



C

	Ct Brach	Ct Actb	dCt	ddCt	Fold
Control(1)	24.2	17.7	6.5	0.0	1.0
Activin A 3ng/ml(1)	21.3	18.1	3.2	-3.3	9.8
Activin A 30ng/ml(1)	26.0	17.4	8.6	2.1	0.2
Control(2)	24.7	18.9	5.8	0.0	1.0
Activin A 3ng/ml(2)	22.2	17.8	4.4	-1.4	2.6
Activin A 30ng/ml(2)	23.8	17.5	6.3	0.5	0.7
Control(3)	27.0	17.5	9.5	0.0	1.0
Activin A 3ng/ml(3)	24.5	18.1	6.4	-3.1	8.6
Activin A 30ng/ml(3)	29.6	18.7	10.9	1.4	0.4

Fig. 1. Induced *Brachyury* expression in EBs in the presence of Activin A. Whole-mount in situ hybridization shows that in EBs cultured in serum-free conditions (A), *Brachyury* can be induced significantly by Activin A treatment (3 ng/ml) (B). Staining is indicated by white arrows. qRT–PCR reveals that, after correction for *Actb* (reference gene) expression, *Brachyury* (Brach) is induced in three independent biological repeats by low Activin A and is reduced by high Activin A (C). Ct, threshold cycle value; dCt, delta Ct; ddCt, delta-delta Ct. Biological replicates are indicated with (1)–(3). Experimental details are discussed in Ref. [7].

plate to exclude interrun variation, *Brachyury* induction is observed in all repeats (Fig. 1C). However, because of high variation between biological repeats, as is frequently observed in mES cell differentiation experiments, *Brachyury* induction measured by qRT–PCR is not statistically significant on calculation of the mean messenger RNA (mRNA) expression over the three independent experiments (Fig. 2). The two problematic issues in this experiment are the variation in control expression levels between the various replicates and the variation in fold induction between biological replicates as a response to an identical stimulus.

When we compare Experiment 1 with Experiment 3, it is obvious that control levels are largely different (~ eightfold), whereas *Brachyury* fold change inductions by 3 ng/ml Activin A are similar in Experiment 1 and Experiment 3 (Fig. 1C). A second problem is illustrated between Experiment 1 and Experiment 2, where control levels are now similar, whereas *Brachyury* induction is approximately four times higher in Experiment 1 than in Experiment 2 (Fig. 1C).

When the statistical significance of observed differences between the conditions and the controls is determined, these two problems result in high variation of mES cell differentiation data and, therefore, lead to statistically insignificant results even though in the three independent experiments *Brachyury* is clearly induced (Fig. 1C). Because of this, stem cell biologists often opt to show one representative example instead of a more reliable average value with proper error bars or confidence intervals. The observed lack of statistical significance might erroneously raise the question of whether *Brachyury* is truly induced, but our experiments are perfectly in line with reports studying the Activin A pathway [5–7].

Because our findings suggest that the lack of statistical significance might be caused by the experimental variability of the mES cell system, we sought to properly standardize the obtained gene expression data, eliminating or reducing interexperimental variation. Previously, we reported a first standardization step by identifying suitable reference genes in the mES cell differentiation system [8]. Here, we present and apply a standardization procedure for data sets from multiple biological replicates by performing sequential data transformations, enabling correct assessment of statistical significance. In Fig. 2, we illustrate the effect of each of these steps on the mean value and the 95% confidence interval

A

Sequential data transformation steps	Untreated control			Activin A 3ng/ml				Activin A 30ng/ml			
	Average	-95% CI	+95% CI	Average	-95% CI	+95% CI	P<0.05	Average	-95% CI	+95% CI	P<0.05
None	1.0000	-1.2197	2.8635	5.5312	-7.1910	16.8815	No	0.5198	-0.9447	2.2413	No
Log transformation	1.0000	0.0339	29.5004	6.0629	0.3748	98.0747	No	0.3969	0.0076	20.8241	No
+ Mean centering	1.0000	0.7930	1.2611	6.0629	1.1786	31.1888	No	0.3969	0.0888	1.7733	No
+ Autoscaling	1.0000	0.7596	1.3164	6.3100	5.7631	6.9089	Yes	0.4175	0.3472	0.5021	Yes

B

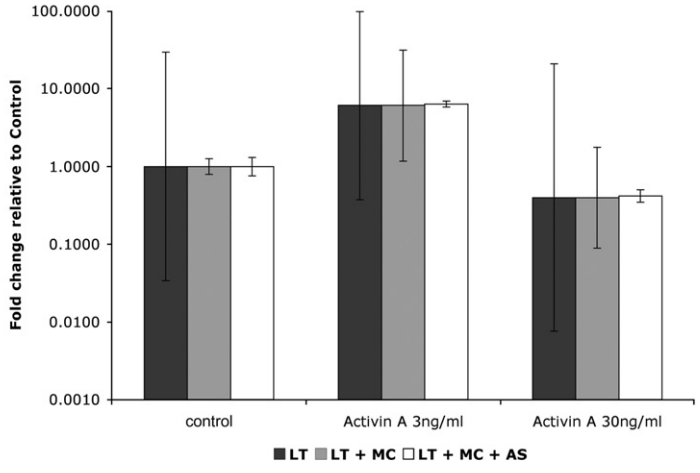


Fig. 2. Effect of sequential standardization steps on the statistical significance of *Brachyury* expression by Activin A in biological repeats. The effect of the sequential standardization steps on the average and the 95% CI (−95% CI, lower interval, and +95% CI, upper interval) is demonstrated in the table (A) as well as in the figure (B), representing the average (histograms) and 95% CI (error bars) after the sequential standardization steps (B). $P < 0.05$ indicates the significance of *Brachyury* induction or reduction by the Activin A treatment compared with untreated controls. LT, log transformation; MC, mean centering; AS, autoscaling.

(CI) calculated at each sequential standardization step performed on the data of the three biological repeats shown in Fig. 1. The standardization calculations for the example data set were performed in MS Excel and are provided as supplementary material.

First, we performed a log transformation of the normalized relative gene expression levels; this makes the data distribution more symmetric, attributing equal weight to conditions with overexpression or underexpression (Fig. 2, black). As such, the influence of outlier values will also be largely eliminated.

For all log-transformed normalized relative quantities a_{ij} from n experiments i and m conditions j, we calculated the mean expression level of all conditions j in experiment i as μ_i (Eq. (1)), the standard deviation of the expression across all conditions j in experiment i as σ_i (Eq. (2)), and the mean standard deviation of all experiments as $\overline{\sigma}$ (Eq. (3)).

$\forall i \in \{1,\ldots,n\}, \forall j \in \{1,\ldots,m\}$:

$$\mu_i = \frac{\sum_{j=1}^{m} a_{ij}}{m} \tag{1}$$

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{m} (a_{ij} - \mu_i)^2}{m-1}} \tag{2}$$

$$\overline{\sigma} = \frac{\sum_{i=1}^{n} \sigma_i}{n} \tag{3}$$

Log transformation, however, does not correct any of the experimental differences observed between biological repeats in ES cell differentiation; therefore, the means of each replicate experiment were mean centered by subtracting the mean normalized *Brachyury* relative expression level across all conditions in a given replicate experiment from that same experiment (Fig. 2, gray; Eq. (4)).

This step does not affect the mean fold induction, but it provides a correction for the difference in background or control level between biological repeats. Even though this second step further reduces variation between replicates caused by different control levels, the difference in fold change between the experimental conditions still results in statistically insignificant conclusions.

By subsequent autoscaling or equalization of the standard deviation across all conditions in each biological replicate, via division of the mean-centered values by the experimental standard deviation σ_i for the same replicate, the influence of varying folds of induction between experiments is greatly reduced (Eq. (4)).

Autoscaling does, however, require a final correction of the fold change by multiplying the autoscaled fold changes with the mean standard deviation of the replicate experiments before autoscaling so as to make the fold changes reflecting the initial observations, resulting in standardized and fold change preserved log-transformed relative quantities a_{ij} (Fig. 2, white):

$$a_{ij} = \left(\frac{a_{ij} - \mu_i}{\sigma_i}\right) \cdot \overline{\sigma} \tag{4}$$

At this point, statistical significance can be determined by calculation of the 95% CI for a limited number of experimental replicates or by another statistical test (e.g., the nonparametric Mann–Whitney test for comparing two groups).

To further validate our method, we investigated whether the sequential standardization steps would lead to false positive outcomes for noninduced genes by using a data set for the *Flk1* gene of which the expression is not affected by Activin A treatment in the described mES cell setup [7]. In the supplementary material, we present the standardization steps performed on *Flk1* expression, illustrating that no statistically significant differences are found in the case of a nonaffected gene such as *Flk1*. Thus, we demonstrate that our method allows distinction between positive samples (induction/reduction in gene expression) and negative samples (unaffected gene expression) without increasing the rate of false positives.

The performance of the proposed standardization procedure was further evaluated on large data sets consisting of at least three biological replicates that had similar variability to the example data sets and was found to be extremely adequate in canceling out interexperimental variation [7].

In summary, we have described a simple method for standardizing gene expression data of biological replicates that shows substantial variation between these replicates, although there is a clear similar trend. By performing a standardization procedure based on log transformation, mean centering, and autoscaling, high interexperimental variation can be canceled out, after which statistical analysis can be used to assess the significance of observed differences.

Acknowledgments

This research was funded by the Horizontal Action of the Vrije Universiteit Brussel (grant HOA1), Fund for Scientific Research Flanders (grant G.0485.06), and the European Union (Sixth Framework, grant NEST012930). Jo Vandesompele is supported by the Fund for Scientific Research Flanders (FWO).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ab.2008.04.036.

References

[1] J. Hellemans, G. Mortier, A. De Paepe, F. Speleman, J. Vandesompele, qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data, Genome Biol. 8 (2007) R19.
[2] M.W. Pfaffl, A new mathematical model for relative quantification in real-time RT–PCR, Nucleic Acids Res. 29 (2001) e45.
[3] M.W. Pfaffl, A. Tichopad, C. Prgomet, T.P. Neuvians, Determination of stable housekeeping genes, differentially regulated target genes, and sample integrity: BestKeeper—Excel-based tool using pair-wise correlations, Biotechnol. Lett. 26 (2004) 509–515.
[4] J. Vandesompele, K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, F. Speleman, Accurate normalization of real-time quantitative RT–PCR data by geometric averaging of multiple internal control genes, Genome Biol. 3 (2002). research0034.1–research0034.11.
[5] P. Gadue, T.L. Huber, P.J. Paddison, G.M. Keller, Wnt and TGF-β signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells, Proc. Natl. Acad. Sci. USA 103 (2006) 16806–16811.
[6] B.M. Johansson, M.V. Wiles, Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development, Mol. Cell. Biol. 15 (1995) 141–151.
[7] E. Willems, L. Leyns, Patterning of mouse embryonic stem cell-derived pan-mesoderm by activin A/Nodal and Bmp4 signaling requires fibroblast growth factor activity, Differentiation (2008), 2 January, E-pub ahead of print.
[8] E. Willems, I. Mateizel, C. Kemp, G. Cauffman, K. Sermon, L. Leyns, Selection of reference genes in mouse embryos and in differentiating human and mouse ES cells, Intl. J. Dev. Biol. 50 (2006) 627–635.