Notes & Tips

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

Erik Willems, Luc Leyns, Jo Vandesompele

Abstract

Gene expression analysis by quantitative reverse transcription PCR (qRT–PCR) allows accurate quantification 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.
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.

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].

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 (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 $\mu_i$ (Eq. (1)), the standard deviation of the expression across all conditions $j$ in experiment $i$ as $\sigma_i$ (Eq. (2)), and the mean standard deviation of all experiments as $\overline{\sigma}$ (Eq. (3)).

$$\mu_i = \frac{\sum_{j=1}^{m} a_{ij}}{m}$$  \hspace{1cm} (1)

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

$$\overline{\sigma} = \frac{\sum_{i=1}^{n} \sigma_i}{n}$$  \hspace{1cm} (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 $\sigma_i$ 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):

$$a_{ij} = \left(\frac{a_{ij} - \mu_i}{\sigma_i}\right) \overline{\sigma}$$  \hspace{1cm} (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


References


