BIOINFORMATICS ORIGINAL PAPER Vol. 22 no. 10 2006, pages 1251–1258
doi:10.1093/bioinformatics/btl068

Gene expression

A calibration method for estimating absolute expression levels from microarray data

Kristof Engelen[1], Bart Naudts[2], Bart De Moor[1] and Kathleen Marchal[1,3,*]

[1]BIOI@SCD, Department of Electrical Engineering, K.U.Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium, [2]ISLab, Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, B-2020 Antwerpen, Belgium and [3]CMPG, Department of Microbial and Molecular Systems, K.U.Leuven, Kasteelpark Arenberg 20, B-3001 Leuven, Belgium

Received on November 10, 2005; revised on February 13, 2006; accepted on February 21, 2006
Advance Access publication March 7, 2006
Associate Editor: Joaquin Dopazo

ABSTRACT

Motivation: We describe an approach to normalize spotted microarray data, based on a physically motivated calibration model. This model consists of two major components, describing the hybridization of target transcripts to their corresponding probes on the one hand, and the measurement of fluorescence from the hybridized, labeled target on the other hand. The model parameters and error distributions are estimated from external control spikes.

Results: Using a publicly available dataset, we show that our procedure is capable of adequately removing the typical non-linearities of the data, without making any assumptions on the distribution of differences in gene expression from one biological sample to the next. Since our model links target concentration to measured intensity, we show how absolute expression values of target transcripts in the hybridization solution can be estimated up to a certain degree.

Contact: kathleen.marchal@biw.kuleuven.be

Supplementary information: Supplementary data are available at *Bioinformatics* online.

INTRODUCTION

Normalization of microarray measurements, the first step in a microarray analysis trajectory, aims at removing consistent and systematic sources of variations to allow mutual comparison of measurements acquired from different slides and experimental settings. Obviously, normalization largely influences the results of all subsequent analyses (such as clustering), and therefore is a crucial phase in the analysis of microarray data. For normalization of spotted microarrays, different methods have been described [for overviews, see for instance Leung and Cavalieri (2003); Quackenbush (2002) and Bilban *et al*. (2002)]. In general, preprocessing of spotted microarrays largely depends on the calculation of the log-ratios of the measured intensities. For complex designs, using ratios complicates the comparison of different experimental conditions, especially when they are not measured with the same reference condition. To cope with this, some approaches inherently work with absolute intensities [e.g. ANOVA (Wolfinger *et al*., 2001; Kerr *et al*., 2000)], or use a universal reference to estimate absolute expression levels from the ratios (Dudley *et al*., 2002). A common ratio-normalization step consists of the linearization of the Cy3 versus Cy5 intensities [e.g. LOESS (Yang *et al*., 2002)], sometimes followed by, or inherently combined with, techniques for variance stabilization (Durbin *et al*., 2002; Huber *et al*., 2002). These methods assume that the distribution of gene expression shows little overall change and is balanced between the biological samples tested (from here on referred to as the 'Global Normalization Assumption'). If this assumption is violated, for instance when comparing two drastically different biological conditions or when working with dedicated arrays, using such a normalization may yield erratic results. Normalization algorithms that do not require this Global Normalization Assumption have been proposed (Wang *et al*., 2005; Zhao *et al*., 2005), but a more reliable strategy to avoid making any assumptions regarding the distribution of the gene expression is to use external control spikes (exogenous RNA species that are added to the hybridization solution in known concentrations, prior to labeling) to estimate normalization parameters. Other types of experimental normalization controls, such as housekeeping genes, spotted clone pools or spotted genomic DNA, have also been proposed [for an overview, see Kroll and Wölfl (2002)], but none of these are able to compensate for unbalanced gene expression changes. By using external control spikes, it has been shown that global mRNA changes, resulting in an uneven distribution of expression changes, occur more frequently than what was previously believed (van Bakel and Holstege, 2004; van de Peppel *et al*., 2003), and that these changes can have a significant impact on the interpretation of data normalized according to the Global Normalization Assumption (Radonjic *et al*., 2005).

External control spikes have previously been employed for quality control and normalization (Radonjic *et al*., 2005; van de Peppel *et al*., 2003; Badiee *et al*., 2003; Wang *et al*., 2003; Benes and Muckenthaler, 2003; Hughes *et al*., 2001; Girke *et al*., 2000; Eickhoff *et al*., 1999), but have seldom (Carter *et al*., 2005) been exploited to their full potential. In fact, spikes are genuine calibration points, in that they relate the measured intensity to the actual RNA concentration in the hybridization solution. In this paper, we propose a normalization procedure that can be used to estimate absolute expression levels, and is based on spike measurements and a calibration model. This procedure is capable of adequately removing the typical non-linearities of the data, without making any assumptions on the distribution of gene expression from one

*To whom correspondence should be addressed.

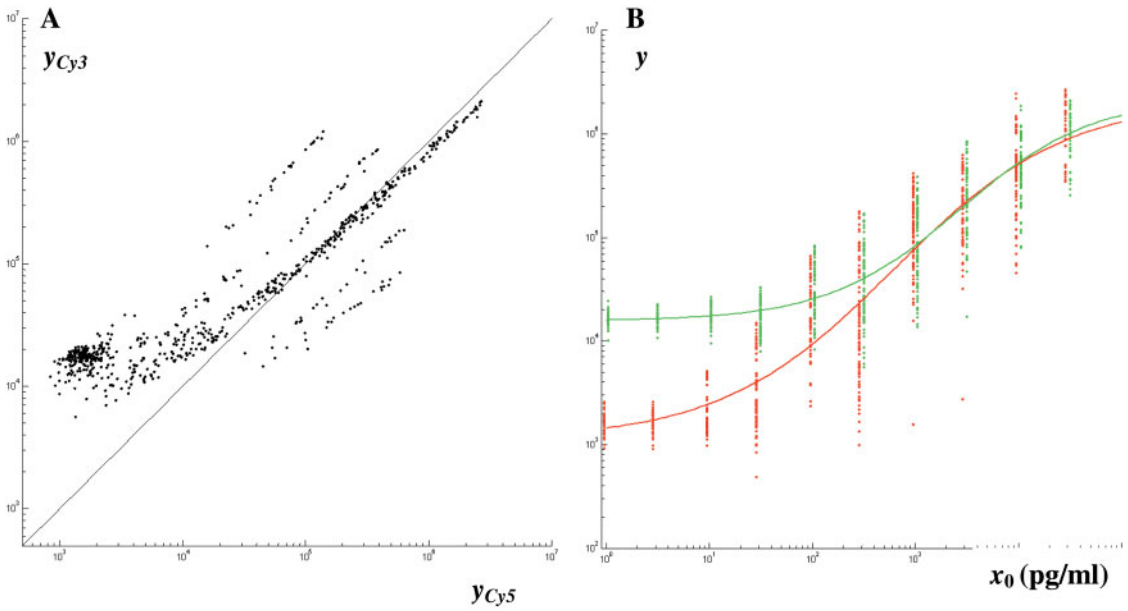

Fig. 1. External control spikes. (**A**) Measured Cy5 intensities (y_{Cy5}) plotted against Cy3 intensities (y_{Cy3}) for all external control spikes (Cy5/Cy3 ratios 1:10, 1:3, 1:1, 3:1 and 10:1). This plot illustrates the relatively small scanner errors, especially compared with the large variation in intensities that is observed in panel B. (**B**) Non-linear relationship between measured intensity y and corresponding concentrations x_0 (pg/ml) of target transcripts in the hybridization solution for all external control spikes with a Cy5/Cy3 ratio of 1:1. A colour version of this figure can be found in the Supplementary data.

biological sample to the next. Moreover, estimates of absolute expression levels instead of expression ratios, can greatly simplify inter platform comparisons and the analysis of large, complex designs comparing multiple biological conditions.

MODELS AND ALGORITHMS

The proposed normalization procedure is straightforward in principle: intensity measurements of external control spikes serve to estimate the parameters of a calibration model. These parameters can then be used to obtain absolute expression levels for every gene in each of the tested biological conditions. The calibration model consists of two components, a hybridization reaction and a dye saturation function. In the following sections a more detailed description of this model is given, along with its corresponding parameters and error distributions.

Hybridization reaction

This component of the model takes spot related errors into account, which have been shown to have a large effect on the final, observed signal (Rocke and Durbin, 2001). How these errors manifest themselves in the measured intensities, becomes clear when comparing the behavior of the data in Figure 1. A plot of the Cy3 versus Cy5 spike intensities (Fig. 1, panel A) illustrates the relatively small scanner errors: ratios of these controls seem highly conserved, especially at upper intensity levels. Figure 1, panel B on the other hand, displays the relation between the measured intensities of these external control spikes to their actual concentration in the hybridization solution. A large variation in intensity for a single spike concentration can be observed. In view of the relatively small scanner errors, the level of variation seen in this plot is remarkable. Heterogeneous 'spot capacities', in terms of the available quantity of probe, offer an explanation: imperfections in the spotting process allow distinct spots to bind different amounts of target from the hybridization solution. Whether the main source of this variation in 'spot capacity' can be attributed to the actual amount of deposited cDNA, or to a measure of spot quality [e.g. probe density (Peterson *et al*., 2001), cDNA probe length (Stillman and Tonkinson, 2001), etc.], the implications are equivalent.

To explain these large variations of absolute intensities observed for a single spike concentration, a hybridization component was included in our model to account for these spot errors. The relation between the amount of hybridized target (x_s) and the concentration of the corresponding transcript in the hybridization solution (x_0) is modeled by the steady state of the following reaction:

$$x_0 + s \overset{K_A}{\Leftrightarrow} x_s. \tag{1}$$

In our model the hybridization constant K_A is assumed to be equal for all spots on a single microarray. Differences in hybridization constants should therefore be interpreted as variations caused by microarray related factors such as temperature, salt concentrations, hybridization time, etc., but do not account for gene specific hybridization efficiencies.

A second assumption underlying our model is that the hybridization is a first order reaction, and that x_0 is in excess (i.e. x_0 is constant). The latter assumption ensures that the amount of hybridized target at the end of the reaction only depends on the initial concentration in the hybridization solution. The amount of probe of a spot (s) available for hybridization will decrease with an increasing amount of hybridized target x_s ($s = s_0 - x_s$, s_0 being the spot size or maximal amount of available probe), so that we can write at thermodynamic equilibrium:

$$\frac{x_s}{x_0(s_0 - x_s)} = K_A. \tag{2}$$

The spot capacity s_0 follows a certain distribution around an average spot capacity μ_s: $s_0 = \mu_s + \varepsilon_s$ (i.e. additive spot error) or $s_0 = \mu_s\, e^{\varepsilon_s}$ (i.e. multiplicative spot error) with $\varepsilon_s \sim N(0,\sigma_s)$. Whichever distribution is more appropriate in any particular case will depend largely on the type of microarray slide and spotting procedure used, and should be evaluated after performing the normalization procedure, e.g. by testing the normality assumptions of the spot error distribution. The distribution parameters μ_s and σ_s can be considered equal for all measurements of a single array, or treated differently on a per pin group basis to compensate for spotting pin related variations. Finally, we assume that the presence of distinct labels (Cy3 and Cy5) does not influence the hybridization efficiency of the differentially labeled target transcripts, i.e.

$$\frac{x_{0,\mathrm{Cy5}}}{x_{0,\mathrm{Cy3}}} = \frac{x_{s,\mathrm{Cy5}}}{x_{s,\mathrm{Cy3}}}$$

where

$$x_0 = x_{0,\mathrm{Cy3}} + x_{0,\mathrm{Cy5}}$$

$$x_s = x_{s,\mathrm{Cy3}} + x_{s,\mathrm{Cy5}}. \quad (3)$$

In the above equations, it would be more accurate to explicitly model the amount of non-labeled target in the solution (i.e. to write $x_0 = x_0^* + x_{0,\mathrm{Cy3}} + x_{0,\mathrm{Cy5}}$, with x_0^* being the amount of non-labeled target), and to include parameters for labeling efficiencies. However, since the external control spikes are added to the hybridization solution before the actual labeling reaction, effects attributed to labeling efficiency are accounted for in the dye saturation function, described below.

Dye saturation function

A second component of our model is the dye saturation function, which describes the relationship between the measured intensity y and the amount of labeled target x_s, hybridized to a single spot on the microarray:

$$y = p_1 x_s e^{\varepsilon_m} + p_2 + \varepsilon_a. \quad (4)$$

This dye saturation function is a simple linear equation incorporating an additive and multiplicative intensity error, respectively represented by $\varepsilon_a \sim N(0,\sigma_a)$ and $\varepsilon_m \sim N(0,\sigma_m)$. This type of function has already been used in other normalization strategies (Durbin *et al*., 2002; Rocke and Durbin, 2001).

In all, there are three different error distributions that are assumed to influence intensity measurements: additive intensity error ε_a, multiplicative intensity error ε_m and spot capacity error ε_s. The parameters of the saturation function and the variances of the intensity error distributions are considered specific for all measurements of a single array and dye combination. The parameters of the hybridization reaction and variance of the spot error on the other hand apply to all measurements of a single array. As such, Cy3 and Cy5 intensities obtained from the same array element are modeled with different saturation parameters and intensity errors, but will share the same hybridization parameters and spot error. Based on Equations (2)–(4), the intensities y_{Cy3} and y_{Cy5}, measured on a single spot s_0 of the array, are related to the amount of corresponding target $x_{0,\mathrm{Cy3}}$ and $x_{0,\mathrm{Cy5}}$ in the hybridization solution as

$$y_{\mathrm{Cy3}} = p_{1,\mathrm{Cy3}}\left(\frac{x_{0,\mathrm{Cy3}} s_0}{K_A + x_{0,\mathrm{Cy3}} + x_{0,\mathrm{Cy5}}}\right) e^{\varepsilon_{m,\mathrm{Cy3}}} + p_{2,\mathrm{Cy3}} + \varepsilon_{a.\mathrm{Cy3}} \quad (5)$$

$$y_{\mathrm{Cy5}} = p_{1,\mathrm{Cy5}}\left(\frac{x_{0,\mathrm{Cy5}} s_0}{K_A + x_{0,\mathrm{Cy5}} + x_{0,\mathrm{Cy3}}}\right) e^{\varepsilon_{m,\mathrm{Cy5}}} + p_{2,\mathrm{Cy5}} + \varepsilon_{a,\mathrm{Cy5}}. \quad (6)$$

The differentially labeled targets $x_{0,\mathrm{Cy3}}$ and $x_{0,\mathrm{Cy5}}$ will compete for the same spotted probe DNA s_0. As shown in the equations above, the intensity measured for the Cy3 channel (y_{Cy3}) is not only dependent on the amount of Cy3 labeled target ($x_{0,\mathrm{Cy3}}$), but also on the amount of target labeled with Cy5 ($x_{0,\mathrm{Cy5}}$), and vice versa.

Parameter estimation

The model parameters are estimated separately for each microarray, based on the measured intensities y of the external control spikes and their known concentration in the hybridization solution x_0. In order to determine these model parameters, it is important to have initial, reliable values for σ_m and σ_a. Estimates for $\sigma_{a,\mathrm{Cy3}}$ and $\sigma_{a,\mathrm{Cy5}}$ can easily be obtained by computing the standard deviation of the intensities for the negative control spikes (not present in the hybridization solution). Finding a reliable measure for $\sigma_{m,\mathrm{Cy3}}$ and $\sigma_{m,\mathrm{Cy5}}$ is less evident. Although the additive intensity error can be neglected, the multiplicative errors are still confounded with the influence of spot errors at high intensity levels. Estimating $\sigma_{m,\mathrm{Cy3}}$ and $\sigma_{m,\mathrm{Cy5}}$ independently for both channels from these higher intensity replicate measurements is not feasible. Obtaining an adequate approximation is nevertheless possible. In the higher intensity range where the calibration controls (ratio 1:1) exhibit a log linear behavior in a y_{Cy3} versus y_{Cy5} plot (Supplementary Figure S1), the main contribution to the observed variation can be assigned to the multiplicative intensity error. Indeed in this range, differences in spot size will obviously nullify themselves and the additive intensity error can be neglected. If we then assume that $\sigma_{m,\mathrm{Cy3}}$ and $\sigma_{m,\mathrm{Cy5}}$ contribute equally to the observed variation ($\sigma_m = \sigma_{m,\mathrm{Cy3}} = \sigma_{m\mathrm{Cy5}}$), a value for σ_m can be obtained (Supplementary Figure S1). Performing an orthogonal regression of Cy5 versus Cy3 intensities on the selected data points will yield an error distribution of which the standard deviation is an estimate of $\sigma_m\sqrt{2}$.

Obtaining a solution for the remaining parameters (dye saturation and hybridization parameters $p_{1,\mathrm{Cy3}}$, $p_{1,\mathrm{Cy5}}$, $p_{2,\mathrm{Cy3}}$, $p_{2,\mathrm{Cy5}}$ and K_A respectively; μ_s is kept constant at an arbitrary value) is done in a least squares sense. The error sum of squares that is minimized is that of spot capacity errors, i.e.

$$\min(\mathrm{SSE}_s = \sum_i \varepsilon_s(i)^2) \quad (7)$$

with respect to $p_{1,\mathrm{Cy3}}$, $p_{2,\mathrm{Cy3}}$, $p_{1,\mathrm{Cy5}}$, $p_{2,\mathrm{Cy5}}$ and K_A; i indicates a single spot.

The minimization of SSE_s is done numerically. The individual spot errors $\varepsilon_s(i)$, necessary to calculate the SSE_s in every iteration (i.e. for any given set of parameter values), are of course unknown. For every spot on the microarray, they are estimated by comparing the expected intensity [a function of target concentration $x_{0,\mathrm{Cy3}}$ and $x_{0,\mathrm{Cy5}}$, and a set of parameter values as indicated by (5) and (6)] to the measured intensity values (y_{Cy3} and y_{Cy5}) for both channels, and

scoring the difference based on the estimators of additive and multiplicative intensity variances. More precisely, for each pair of measurements obtained from a single spot, the following object function is minimized with respect to that spots error $\varepsilon_s(i)$, i.e.

$$\min(Q_{\text{estim}} = Q_{\text{estim}}^{\text{Cy3}} + Q_{\text{estim}}^{\text{Cy5}}) \tag{8}$$

with respect to $\varepsilon_s(i)$, where

$$Q_{\text{estim}}^{D} = \underset{\varepsilon_m, \varepsilon_a}{\text{argmin}} \left(\left(\frac{\varepsilon_m}{\sigma_m \sqrt{2}} \right)^2 + \left(\frac{\varepsilon_a}{\sigma_a \sqrt{2}} \right)^2 \right)_D, \quad D = Cy3, Cy5 \tag{9}$$

subject to Equations (5) and (6), i.e.

$$\begin{aligned} Q_{\text{estim}}^{D} &= \underset{\varepsilon_m}{\text{argmin}} \left(\left(\frac{\varepsilon_m}{\sigma_m \sqrt{2}} \right)^2 + \left(\frac{y - p_1 x_s e^{\varepsilon_m} - p_2}{\sigma_a \sqrt{2}} \right)^2 \right)_D \\ &= \underset{\varepsilon_a}{\text{argmin}} \left(\left(\frac{ln(y - p_2 - \varepsilon_a) - ln(p_2 x_s)}{\sigma_m \sqrt{2}} \right)^2 + \left(\frac{\varepsilon_a}{\sigma_a \sqrt{2}} \right)^2 \right)_D. \end{aligned}$$

This object function is related to the probability of observing the measured Cy3 and Cy5 intensities given the amount of hybridized target [can be calculated according to (5) and (6) as target concentrations of spikes are known] and intensity error distributions. The procedure for an entire microarray is illustrated in Figure 2. The parameters of the intensity error distributions, σ_m and σ_a, determine the spread of measurements around the Cy3 and Cy5 saturation curves. The gray dots in Figure 2 depict the relation between measured intensity and amount of hybridized target under the assumption of equal spot sizes [i.e. all $\varepsilon_s(i)$ are zero]. Most of these are localized in regions of high intensity error and are therefore very unlikely. However, by allowing errors $\varepsilon_s(i)$ on individual spot's capacities, and thus altering the amount of hybridized target per spot for both dyes ($x_{s,\text{Cy3}}$ and $x_{s,\text{Cy5}}$), a good correspondence between intensities and saturation curves can be obtained for both channels, and across the entire measurement range (indicated by the black dots). It is notable how well the Cy3 and Cy5 intensities, and the relationships between them, can be explained by our model. For instance in the example given, at lower intensities, Cy3 intensities are persistently higher than Cy5 for equal amounts of hybridized target, while the opposite is true for higher levels, a trend that is nicely reflected by the fitted model. Notice also that, while the ratios between Cy3 and Cy5 intensities are highly conserved—at least at higher intensity levels—absolute intensities may vary to a large extent for transcripts with the same target concentration x_0 owing to spot inhomogenities.

Normalization: estimation of target expression levels

The obtained parameter values can be used to estimate a single $x_0(t,u)$ (i.e. the absolute expression level of a single gene t in a single biological condition u) based on all measurements that were obtained for this combination of gene and condition. Although each array and dye combination is attributed with its own set of parameters, the normalization can be considered a global one. Namely, for each combination of a gene and a tested biological condition, a single expression level is estimated, irrespective of the number of microarray slides, or the number of replicate spots on a slide, for which this gene condition combination was measured. In this sense, the results format of this normalization is

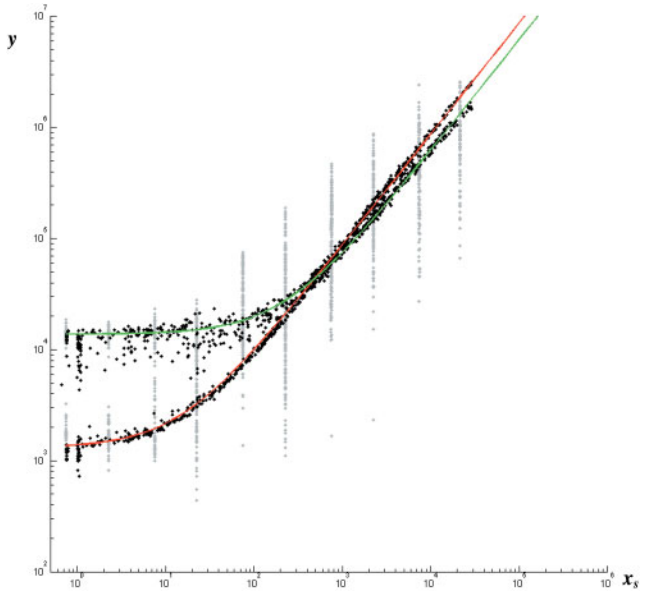

Fig. 2. Parameter estimation. At given parameter values (red and green curve), spot errors are obtained by estimating the amount of hybridized target x_s for the measured intensities y of the external control spikes (black dots). Grey dots depict the amount of hybridized target, assuming equal spot capacities (no spot errors). A colour version of this figure can be found in the Supplementary data.

comparable with the VarietyGene interaction factor effects in the models of Kerr *et al*. (2000), or similar factors in other ANOVA-models.

Although this procedure can be applied to any design, its complexity does depend on the used experimental setup. For a single gene, it requires the estimation of expression values for all the biological conditions at once. These $x_0(t,u)$ can be estimated by minimizing the following object function (an extension of the one used to estimate the model parameters):

$$\min(Q_{\text{norm}} = \sum_{C} \sum_{S_u} Q_{\text{norm}}^{S_u(k)}) \tag{10}$$

with respect to $x_0(t,C)$ and where

$$Q_{\text{norm}}^{S_u(k)} = \left(\underset{\varepsilon_m, \varepsilon_a}{\arg\min} \left(\left(\frac{\varepsilon_m}{\sigma_m \sqrt{2}} \right)^2 + \left(\frac{\varepsilon_a}{\sigma_a \sqrt{2}} \right)^2 \right) + \left(\frac{\varepsilon_s}{\sigma_s \sqrt{2}} \right)^2 \right)_{S_u(k)} \tag{11}$$

subject to Equations (5) and (6)

The subscript C indicates the set of biological conditions under survey; it applies to all conditions that are present in the experimental design. The set of intensities, and the relevant array-dye combinations of parameters, that measure an expression value $x_0(t,u)$, is represented by S_u [a single measured intensity belonging to this set is designated by $S_u(k)$]. So for a single gene t, expression values for all of the biological condition present in the experiment are estimated simultaneously (and together with all the relevant spot errors), and in such a way that the total contribution of the three random errors (i.e. the combined spot errors and additive and multiplicative intensity errors for all intensity data points that are a measure of gene t) is minimized as dictated by the cost function in (10).

RESULTS

A publicly available dataset (Hilson *et al*., 2004), specifically designed for quality control and the assessment of experimental variation (Allemeersch *et al*., 2005; Hilson *et al*., 2004), was chosen to illustrate the workings of our normalization method. This experiment was ideally suited to validate our procedure because first, it contained the necessary spots for measuring external control spikes, which are required for estimating the parameters of our model. A series of external controls (Lucidea Universal Scorecard; Amersham Biosciences) consisted of 10 calibration spikes (added to the hybridization solution in a ratio 1:1 and spanning up to 4.5 orders of magnitude), eight ratio spikes provided at both low and high concentration and two negative controls, was spotted once per pin group, resulting in a total of 24 repeats of each spike probe per array. Second, the experimental design included only a single biological condition (self–self experiments; all hybridizations were conducted with the same RNA sample, extracted from aerial parts of germinating *Arabidopsis thaliana* seedlings), which allows assessing the performance of our normalization method in removing non-linear tendencies present in microarray data. Finally, they were outfitted with an additional set of control spikes that could be used to verify to what extent our method was capable of approximating the absolute target concentrations.

The results presented in this paper were obtained from non-background corrected measurements, since no marked improvements were observed after performing a background subtraction (data not shown). The distribution of spot capacities s_0 was modeled as $s_0 = \mu_s e^{\varepsilon_s}$ with $\varepsilon_s \sim N(0,\sigma_s)$. The distribution parameters μ_s and σ_s were assumed to be equal for all measurements of a single array.

Removal of non-linear artifacts

Figure 3 illustrates the result of applying our method on a selection of two arrays from the 14-array experiment. As this is a self–self design, the same biological sample was measured four times on these two arrays (twice labeled with Cy3 and twice with Cy5). For the purpose of our test, we treated this self–self experiment as a dye swap design with two hypothetically different samples (designated C1 and C2). Estimated expression levels x_0 of the ~19.000 genes are plotted in Figure 3 for C1 versus C2. Because in reality C1 and C2 represent the same biological condition, all estimates being centered along the bisector indicates that our model adequately accounts for the major sources of non-linear variation in the data. The increased variance of the estimates observed at lower target levels is inherent to microarray technology. This range of expression corresponds to the saturation observed in the lower intensity region, i.e. where the additive error has a significant influence, considerably blurring the relationship between measured intensity y and target expression level x_0. Because of these saturation effects, estimates of lower concentration are prone to be less reliable.

As mentioned previously, our method is not bound by experimental design. To illustrate that these results are not only achievable with simple experimental setups, such as a color flip, we normalized a set of four arrays as if it concerned a loop design with four different biological conditions. A comparison of the estimated expression levels is shown in Figure 4.

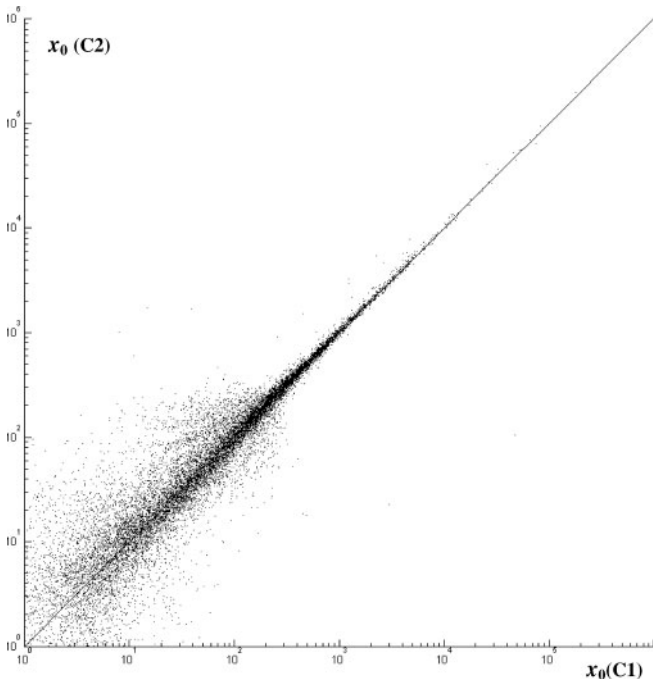

Fig. 3. Removal of non-linear artifacts. Estimated expression levels for C1 are plotted against estimated levels for C2 after normalizing a color flip experiment. C1 and C2 in fact represent the same biological mRNA sample. The centering of data points around the bisector (solid line) indicates that typical microarray non-linearities are adequately accounted for.

Evaluation of target expression level estimates

Although we have shown that our method is capable of estimating absolute expression levels that respect true ratios between the different conditions compared, the previous experiment does not reveal anything about the accuracy of these absolute estimates, i.e. it does not show to what extent these absolute expression levels approximate the actual concentrations of target in the hybridization solution.

To verify the accuracy of estimated target concentrations, they should be compared with their actual concentrations in the hybridization solution. Doing this for the entire population of transcripts is impossible; as for most of the genes this concentration is unknown. However, the dataset contains an additional set of non-commercial spikes for which the absolute concentrations in the hybridization solution are known. The extracted RNA samples were complemented with 14 external controls at amounts of 10^4, 10^3, 10^2, 10, 1, 0.1 or zero copies per cell. In all 14 hybridizations, these controls were compared with a unique reference RNA, capable of binding to all of the 14 spike cDNA probes, always added at a concentration of 100 copies per cell. The experimental design for these control spikes is summarized in Table 1. Results obtained after performing our normalization are shown in Figure 5 [one spike was omitted from analysis because of quality issues (Allemeersch *et al*., 2005)]. Because the estimated target concentrations, expressed in pg/ml, were not directly comparable with the units of copy number per cell, a linear rescaling of these values by a factor that set our estimate of the unique reference RNA to '100' (copies per cell) was performed. Figure 5 shows that, except for the

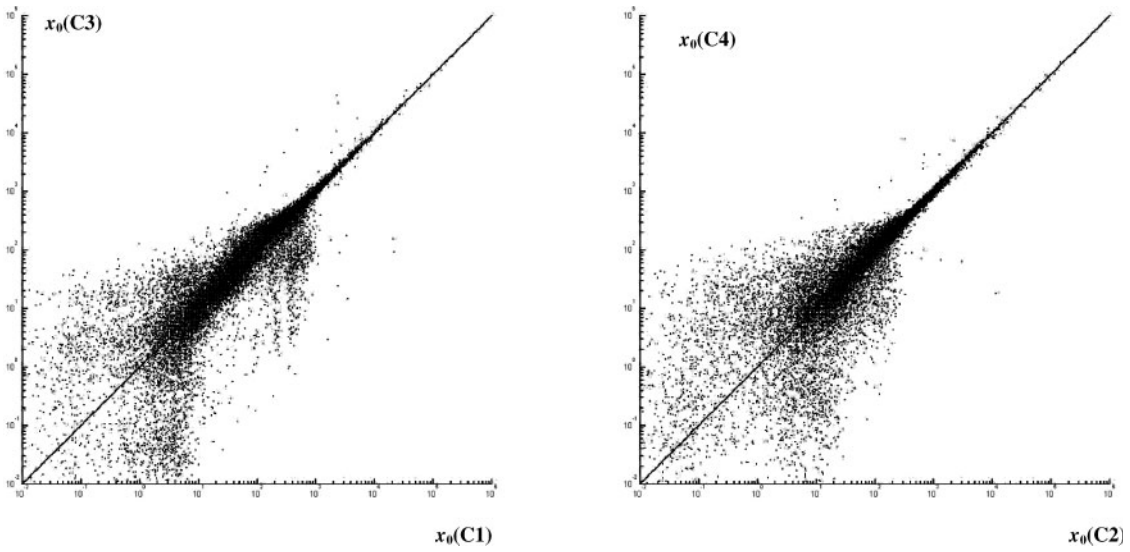

Fig. 4. Removal of non-linear artifacts. Estimated expression levels are plotted after normalizing a loop design experiment with four different hypothetical conditions (designated C1, C2, C3 and C4). Expression levels for conditions that were never measured together on the same microarray slide are directly compared in the plots (i.e. estimated expression levels for C1 are plotted versus those for C3, and estimated expression levels for C2 are plotted versus those for C4). All of these conditions in fact represent the same biological mRNA sample. The centering of data points around the bisector (solid line) indicates that typical microarray non-linearities are adequately accounted for.

Table 1. Mixes of the 14 control spikes

Spike	Spike mix 1	Spike mix 2	Spike mix 3	Spike mix 4	Spike mix 5	Spike mix 6	Spike mix 7	Reference mix
DilA1, DilB1	10 000	0	0.1	1	10	100	1000	100
DilA2, DilB2	1000	10 000	0	0.1	1	100	100	100
DilA3, DilB3	100	1000	10 000	0	0.1	1	10	100
DilA4, DilB4	10	100	1000	10 000	0	0.1	1	100
DilA5, DilB5	1	10	100	1000	10 000	0	0.1	100
DilA6, DilB6	0.1	1	10	100	1000	10 000	0	100
DilA7, DilB7	0	0.1	1	10	100	1000	10 000	100

These spike mixes were added to the hybridization samples, prior to labeling. From the total of 14 arrays, 7 were hybridized with the respective spike mixes labeled in Cy5, each time against the reference mix labeled in Cy3. The remaining seven arrays were hybridized with the respective spike mixes labeled in Cy3, each time against the reference mix labeled in Cy5. Concentrations are given in copy number per cell. *DilB6* was omitted from analysis owing to quality issues (Allemeersch *et al*., 2005).

lowest concentrations, estimated values correspond fairly well to the true target concentrations as present in the hybridization solution. As explained above, also here estimates of the lowest concentrations show a higher error variance.

Comparison of target concentrations between genes

Although Figure 5 shows that concentrations can be accurately estimated, there are several gene-dependent factors that could influence the obtained results, possibly hampering the comparison of estimated concentrations between different genes. Gene specific hybridization efficiencies, for instance, are not taken into account by our model. 'Consistent spot errors' are another factor for which it is theoretically impossible to compensate. Microarrays are usually spotted in batch: experimental errors that influence the DNA probe solutions used for spotting will affect an entire set of microarrays in a similar way. This type of 'consistent spot error' will manifest itself on individual spots across multiple microarray slides, contrary to e.g. variations related to the spotting pins themselves, which would also affect multiple spots on a single array. The particular setup of the 13 external controls, used for assessing the accuracy of estimated target levels, can provide some insight. Because the universal reference RNA can hybridize to all the probes of these spikes, it couples the spot errors of all probes during the estimation of target concentrations. As a consequence of this coupling, consistent spot errors could partially be compensated for, as illustrated in Figure 6. For certain spikes (e.g. *Dil2a*), estimated spot capacities were persistently above or below the average spot capacity μ_s, a feature that was only detectable through the presence of the universal reference RNA. As a result, estimated target concentrations can be subject to gene specific rescaling, hampering the comparison of these concentrations between genes. They can nevertheless be interpreted as absolute values of expression when comparing different concentrations for a single gene.

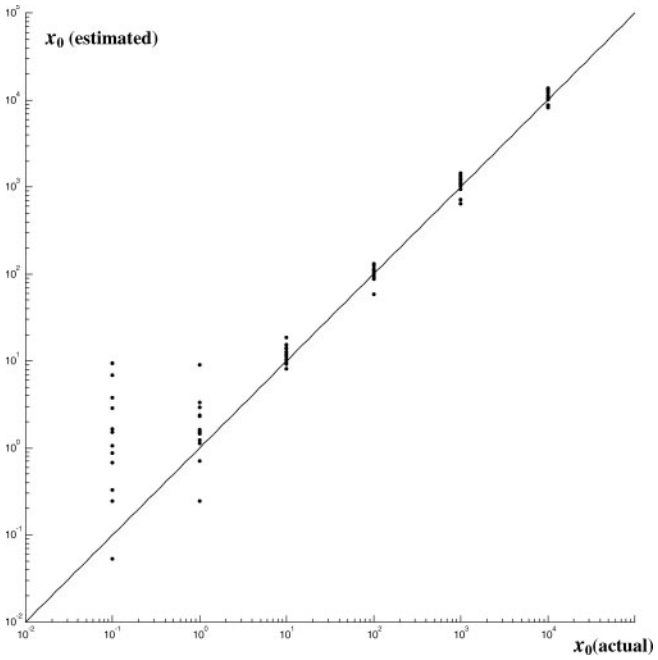

Fig. 5. Evaluation of absolute expression level estimates. Estimated target concentrations (copy number per cell) for all of the 13 controls are plotted against the actual, spiked concentrations. The solid line depicts the bisector.

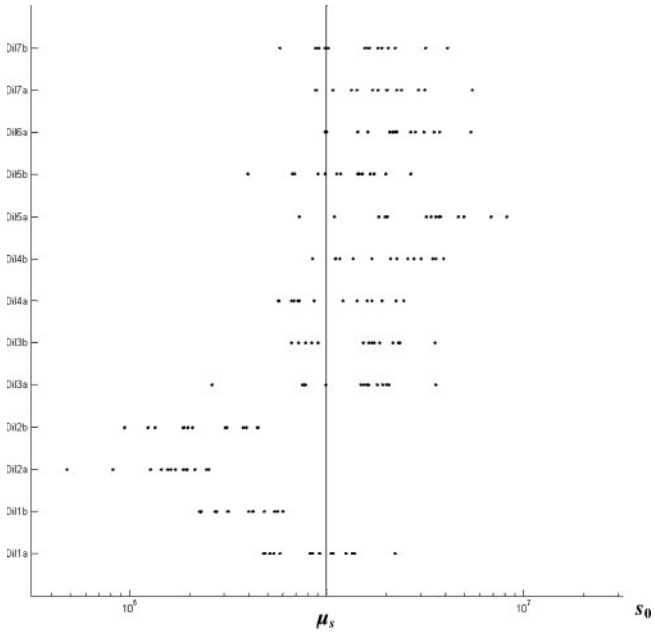

Fig. 6. Consistent spot errors. Estimated spot capacities, corresponding to the 14 microarrays of the experimental design, are plotted for each of the 13 external controls, revealing consistent across-array spot errors. The solid line represents the mean spot capacity.

Influence of background corrections

In our model the combination of the additive intensity error ε_a and intercept of the dye saturation function p_2 can be regarded as an elementary model for the entire slide's background. Having a single background for all spots is different from the spot specific background corrections performed during standard microarray analysis, which estimate a spot specific background from pixels corresponding to the area of the glass slide surrounding the spotted probe. This background model is by no means a restriction concerning the use of background corrected values; our normalization can be applied to both raw and background corrected intensities. Moreover, our method is perfectly capable of working with negative intensity values that may arise when measurements are below background. Whether or not using background corrected measurements is advisable, depends largely on the data quality. This is illustrated in Supplementary Figure S2. Performing a spot specific background correction prior to applying our model would ideally result in the lower saturation limit of our model (p_2) becoming zero. In reality, the estimate for p_2 will indeed be lower, but never reaches a zero level. In general, we have observed a trade-off: background corrected measurements have a larger linear range, but at the expense of increased measurement errors for lower concentrations.

DISCUSSION

In this paper we present an approach for normalizing microarray data using external control spikes to fit a calibration model. This model incorporates parameters and error distributions representing both the hybridization of labeled target to complementary probes and the subsequent measurement of fluorescence intensities. External control spikes serve to estimate the model parameters. The obtained parameter values are then employed to estimate absolute levels of expression for the remaining genes. For each combination of a gene and a tested biological condition, a single absolute target level is estimated, taken the specificities of the design.

The model in itself is fairly basic, in that, with the exception of spot size errors, it is aimed at capturing the global characteristics of an experiment and their overall influence on intensity measurements, generalizing on hard to quantify local sources of variation. The combination of the additive intensity error ε_a and intercept of the dye saturation function p_2, for instance, can be regarded as a global model for the entire slide's background.

The array specific hybridization constant K_A, another global factor, obviously does not account for transcript specific hybridization efficiencies. Therefore, care should be taken when interpreting the estimated expression levels as actual concentrations or when comparing estimated target levels between genes. On the other hand, probe sequences for spotted microarrays are often specifically selected to have properties that obviate large differences in transcript specific hybridization effects. Besides these gene specific hybridization effects, comparison of estimated target levels between genes is also complicated by 'consistent spot errors' across multiple slides. These errors, resulting from experimental inaccuracies in the probe preparation, can arise when microarray slides are spotted in batch. Owing to the characteristics of microarray technology, they cannot be dealt with modelwise.

Although our model is a simplification of physical reality dealing with errors in a global, non-gene specific way, results show that our method is capable of adequately linearizing and normalizing microarray data. An important difference over most existing normalization methods is that our procedure does not rely on any assumptions on the distribution of gene expression levels from one biological sample to the next. Hence, our procedure is particularly well-suited to normalize experiments for which the

Global Normalization Assumption may not be entirely valid, i.e. experiments for which there is no symmetry in the amount of genes that are up-regulated versus down-regulated. Such is typically the case with experiments comparing drastically contrasting biological conditions or with dedicated microarrays, containing only a limited number of probes, representing genes involved in the studied biological process.

In contrast to other normalization methods that use spikes to circumvent the Global Normalization Assumption (van de Peppel *et al*., 2003), our procedure computes absolute expression levels, avoiding the use of ratios. Moreover, for the described experiment, the estimated absolute expression levels approximate the actual concentrations fairly well. Some caution is nevertheless advised when interpreting estimated concentrations as such. This is only problematic as far as comparing expression levels between different genes; the points discussed above have little or no consequence if a comparison is made between estimated target levels across biological conditions for a single gene. Conclusively, our method offers a novel approach to normalize spotted microarrays that combines the advantages of some ANOVA based approaches, which also estimate absolute expression levels, and methods that perform data linearization (e.g. LOESS). The procedure offers independence of assumptions concerning the distribution of gene expression and retains much of the inherent calibration information of external control spike measurements.

ACKNOWLEDGEMENTS

K.E. is a research assistant of the IWT; B.N. was a postdoctoral researcher of the FWO-Vlaanderen for a major part of this work. This work is partially supported by (1) IWT projects: GBOU-SQUAD-20160, GBOU-ANA; (2) Research Council KULeuven: GOA Mefisto-666, GOA-Ambiorics, IDO genetic networks, EF/05/007 SymBioSys; (3) FWO projects: G.0115.01, G.0241.04 and G.0413.03 and (4) IUAP V-22 (2002-2006), 4. FP5 CAGE.

Conflict of Interest: none declared.

REFERENCES

Allemeersch,J. *et al*. (2005) Benchmarking the CATMA microarray. A novel tool for *Arabidopsis* transcriptome analysis. *Plant Physiol*., **137**, 588–601.

Badiee,A. *et al*. (2003) Evaluation of five different cDNA labeling methods for microarrays using spike controls. *BMC Biotechnol*., **3**, 23.

Benes,V. and Muckenthaler,M. (2003) Standardization of protocols in cDNA microarray analysis. *Trends Biochem. Sci*., **28**, 244–249.

Bilban,M. *et al*. (2002) Normalizing DNA microarray data. *Curr. Issues Mol. Biol*., **4**, 57–64.

Carter,M.G. *et al*. (2005) Transcript copy number estimation using a mouse whole-genome oligonucleotide microarray. *Genome Biol*., **6**, R61.

Dudley,A.M. *et al*. (2002) Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc. Natl Acad. Sci. USA*, **99**, 7554–7559.

Durbin,B.P. *et al*. (2002) A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics*, **18** (Suppl. 1), S105–S110.

Eickhoff,B. *et al*. (1999) Normalization of array hybridization experiments in differential gene expression analysis. *Nucleic Acids Res*., **27**, e33.

Girke,T. *et al*. (2000) Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol*., **124**, 1570–1581.

Hilson,P. *et al*. (2004) Versatile gene-specific sequence tags for *Arabidopsis* functional genomics: transcript profiling and reverse genetics applications. *Genome Res*., **14**, 2176–2189.

Huber,W. *et al*. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18** (Suppl. 1), S96–S104.

Hughes,T.R. *et al*. (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol*., **19**, 342–347.

Kerr,M.K. *et al*. (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol*., **7**, 819–837.

Kroll,T.C. and Wolfl,S. (2002) Ranking: a closer look on globalisation methods for normalisation of gene expression arrays. *Nucleic Acids Res*., **30**, e50.

Leung,Y.F. and Cavalieri,D. (2003) Fundamentals of cDNA microarray data analysis. *Trends Genet*., **19**, 649–659.

Peterson,A.W. *et al*. (2001) The effect of surface probe density on DNA hybridization. *Nucleic Acids Res*., **29**, 5163–5168.

Quackenbush,J. (2002) Microarray data normalization and transformation. *Nat. Genet*., **32** (Suppl.), 496–501.

Radonjic,M. *et al*. (2005) Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon *S.cerevisiae* stationary phase exit. *Mol. Cell*, **18**, 171–183.

Rocke,D.M. and Durbin,B. (2001) A model for measurement error for gene expression arrays. *J. Comput. Biol*., **8**, 557–569.

Stillman,B.A. and Tonkinson,J.L. (2001) Expression microarray hybridization kinetics depend on length of the immobilized DNA but are independent of immobilization substrate. *Anal. Biochem*., **295**, 149–157.

van Bakel,H. and Holstege,F.C. (2004) In control: systematic assessment of microarray performance. *EMBO Rep*., **5**, 964–969.

van de Peppel,J. *et al*. (2003) Monitoring global messenger RNA changes in externally controlled microarray experiments. *EMBO Rep*., **4**, 387–393.

Wang,D. *et al*. (2005) A robust two-way semi-linear model for normalization of cDNA microarray data. *BMC Bioinformatics*, **6**, 14.

Wang,H.Y. *et al*. (2003) Assessing unmodified 70mer oligonucleotide probe performance on glass-slide microarrays. *Genome Biol*., **4**, R5.

Wolfinger,R.D. *et al*. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *J. Comput. Biol*., **8**, 625–637.

Yang,Y.H. *et al*. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res*., **30**, e15.

Zhao,Y. *et al*. (2005) An adaptive method for cDNA microarray normalization. *BMC Bioinformatics*, **6**, 28.