



**HAL**  
open science

## Normalization for triple-target microarray experiments

Marie-Laure Magniette, Julie Aubert, Avner Bar-Hen, Samira Elftieh,  
Frédéric Magniette, Jean-Pierre Renou, Jean-Jacques Daudin

► **To cite this version:**

Marie-Laure Magniette, Julie Aubert, Avner Bar-Hen, Samira Elftieh, Frédéric Magniette, et al..  
Normalization for triple-target microarray experiments. BMC Bioinformatics, 2008, 9, pp.216.  
10.1186/1471-2105-9-216 . hal-01197566

**HAL Id: hal-01197566**

**<https://hal.science/hal-01197566v1>**

Submitted on 30 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Research article

Open Access

## Normalization for triple-target microarray experiments

Marie-Laure Martin-Magniette\*<sup>1,2</sup>, Julie Aubert<sup>1</sup>, Avner Bar-Hen<sup>4</sup>, Samira Elftieh<sup>2</sup>, Frederic Magniette<sup>3</sup>, Jean-Pierre Renou<sup>2</sup> and Jean-Jacques Daudin<sup>1</sup>

Address: <sup>1</sup>UMR AgroParisTech-INRA MIA 518, 75231 Paris Cedex05, France, <sup>2</sup>UMR INRA 1165-CNRS 8114-UEVE URGV, 91057 Evry Cedex, France, <sup>3</sup>Unité MOY300, Délégation CNRS ile de France Est, 94532 Thiais Cedex, France and <sup>4</sup>Universite Paris Descartes, MAP 5, Paris cedex 06, France

Email: Marie-Laure Martin-Magniette\* - marie\_laure.martin@agroparistech.fr; Julie Aubert - julie.aubert@agroparistech.fr; Avner Bar-Hen - avner@math-info.univ-paris5.fr; Samira Elftieh - elftieh@evry.inra.fr; Frederic Magniette - frederic.magniette@iledefrance-est.cnrs.fr; Jean-Pierre Renou - renou@evry.inra.fr; Jean-Jacques Daudin - daudin@agroparistech.fr

\* Corresponding author

Published: 28 April 2008

Received: 15 November 2007

BMC Bioinformatics 2008, 9:216 doi:10.1186/1471-2105-9-216

Accepted: 28 April 2008

This article is available from: <http://www.biomedcentral.com/1471-2105/9/216>

© 2008 Martin-Magniette et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Most microarray studies are made using labelling with one or two dyes which allows the hybridization of one or two samples on the same slide. In such experiments, the most frequently used dyes are Cy3 and Cy5. Recent improvements in the technology (dye-labelling, scanner and, image analysis) allow hybridization up to four samples simultaneously. The two additional dyes are Alexa488 and Alexa494. The triple-target or four-target technology is very promising, since it allows more flexibility in the design of experiments, an increase in the statistical power when comparing gene expressions induced by different conditions and a scaled down number of slides. However, there have been few methods proposed for statistical analysis of such data. Moreover the lowess correction of the global dye effect is available for only two-color experiments, and even if its application can be derived, it does not allow simultaneous correction of the raw data.

**Results:** We propose a two-step normalization procedure for triple-target experiments. First the dye bleeding is evaluated and corrected if necessary. Then the signal in each channel is normalized using a generalized lowess procedure to correct a global dye bias. The normalization procedure is validated using triple-self experiments and by comparing the results of triple-target and two-color experiments. Although the focus is on triple-target microarrays, the proposed method can be used to normalize  $p$  differently labelled targets co-hybridized on a same array, for any value of  $p$  greater than 2.

**Conclusion:** The proposed normalization procedure is effective: the technical biases are reduced, the number of false positives is under control in the analysis of differentially expressed genes, and the triple-target experiments are more powerful than the corresponding two-color experiments. There is room for improving the microarray experiments by simultaneously hybridizing more than two samples.

## Background

DNA microarray technology is a high throughput technique by which the expression of the whole genome is studied in a single experiment. In dual label experiments the fluorescent dyes *Cy3* and *Cy5* are used to label the two RNA samples co-hybridized on a same array. Recently two more dyes have been proposed (*Alexa 488* and *Alexa 594*) allowing the simultaneous hybridization of three or four samples. Forster et al. [2] have evaluated triple-target microarray by comparing results of single-target, dual-target and triple-target microarrays. They have concluded that the use of triple-target microarray is valid from an experimental point of view. One year later, Staal et al. [7] have investigated the four-target microarray experiments. Their approach differs from that of [2], but their conclusions are in fair agreement. Their study has shown that *Alexa 594* is best suited as a third dye and that *Alexa 488* can be applied as a fourth dye on some microarray types. These extensions of the microarray technology are promising because they increase throughput, minimize costs and allow more powerful design of experiments. Despite these advantages, triple-target microarrays are only sparsely used [6]. The lack of guidelines for designing these experiments and for normalizing more than two-color microarray data may be an explanation. Recently Woo et al. [8] have proposed experimental designs for three and four-color gene expression microarrays. According to the previous work of [2], the lowess procedure [9] used to normalize data from two-color microarray is still applicable but it normalizes data sequentially because the MA-plot or the lowess correction is defined only for two dyes. Consequently, application of such a normalization method does not globally correct the dye bias due to the three dyes. Moreover the introduction of a third dye induces signal "bleeding". Forster et al. [2] have concluded that "it was considered as negligible between *Cy3* and *Cy5* signals, but seems to be important between *Alexa594* and *Cy3* signals," therefore signal cross-talk cannot be neglected.

We propose in this paper a normalization method for triple-target microarray experiments. First we quantify and correct the signal bleeding. Then we correct the global dye bias using a generalized lowess procedure. Triple-target experiments with *Arabidopsis thaliana* microarrays are used to check if the proposed normalization is effective for correcting the dye bias. Moreover the comparison of the statistical power of the triple-target experiment versus the usual two-color experiment is performed. All programs and data produced for this project are available under request.

## Results and Discussion

### Bleeding

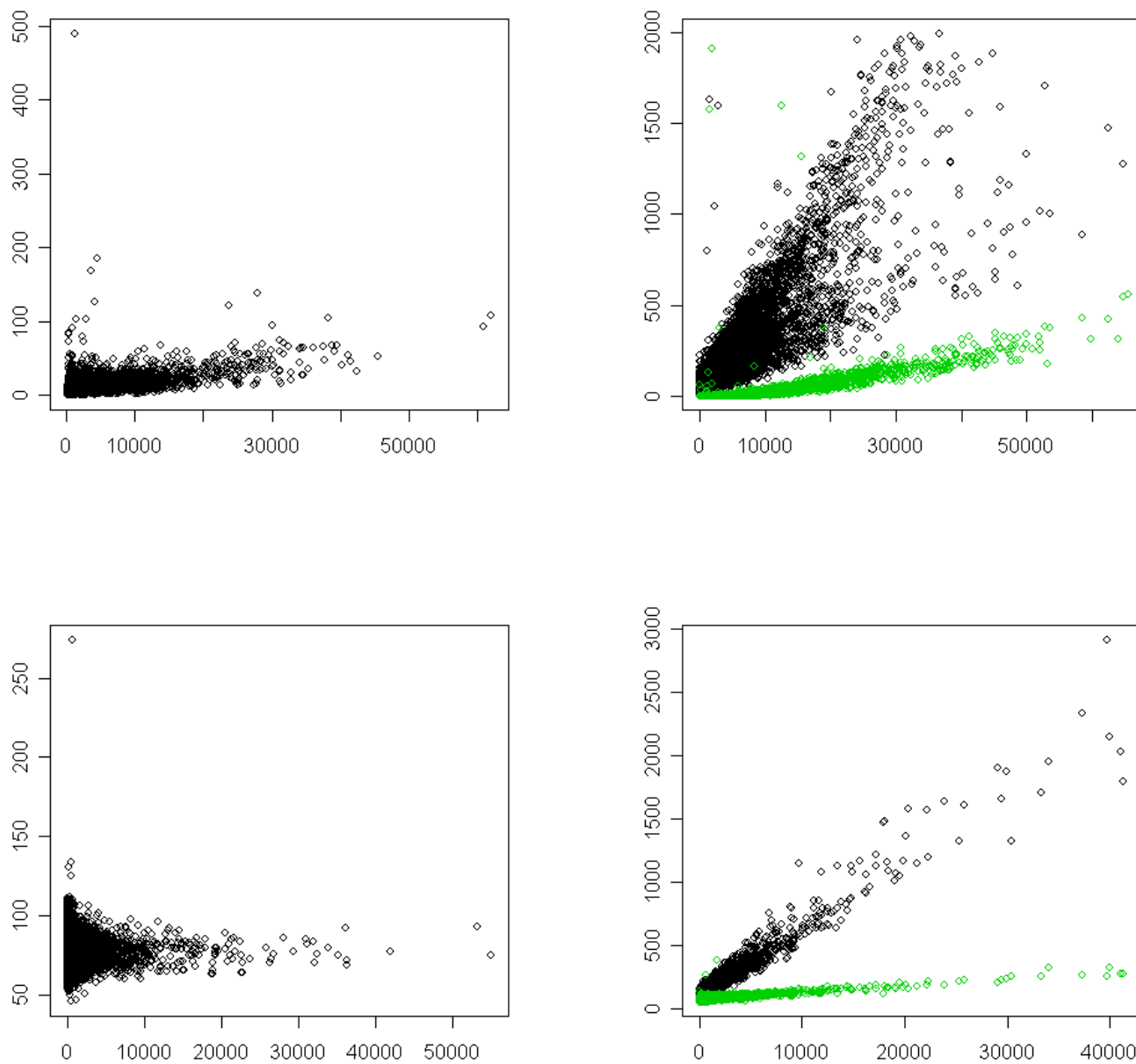
Using the vocabulary of [2], we call a channel, a *blank* channel when no material is hybridized for the associated dye. In theory, this blank channel should produce no signal values, and deviations from this show a bleeding phenomena from one dye-label to another. Signal bleeding from one dye-labelled sample to another is a potential source of bias. Indeed, bleeding artificially increases the signal in other channels of the same spot when the signal is high in one channel. Assume that a gene is highly expressed in condition A and weakly expressed in condition B. The difference between the two conditions is decreased by the bleeding. Therefore bleeding may induce a lowering in the statistical power for detecting differentially expressed genes. Another possibility is that the bleeding effect induces a difference between two channels for the same gene: assume that a gene is highly expressed in condition A and equally expressed in conditions B and C; if the bleeding between the channel corresponding to condition A and the channel corresponding to condition B is higher than the bleeding between A and C, then a difference between signals B and C will appear, which is a technical artifact.

In order to investigate bleeding, we have made a "single target hybridization microarray experiment" where only one dye-labelled sample is hybridized (see the dataset URGV1 in the Methods Section). We also analyzed the single target hybridization data set from Forster [2].

### Experimental indications of the existence of bleeding

Figure 1 gives some plots between the hybridized and blank channels for the Forster and URGV1 datasets. These plots illustrate the bleeding. This bias depends on the channel: the bleeding bias *Cy3* → *Cy5* is negligible but the bleeding bias *Alexa594* → (*Cy5*, *Cy3*) exists. The plots from the Forster single target hybridization experiment and URGV1 datasets present the same patterns with a greater variance for the first set. As the bleeding effect seems to apply on a linear scale, we consider only raw (and not log-transformed) data in this section. Table 1 which contains the Spearman correlation coefficients between the hybridized and the blank channels for the two datasets shows that the bleeding effect exists. The correlations between *Cy5* and *Cy3* are low, but the dye *Alexa594* emits and receives significant cross-talk from the other two dyes.

Since cross-talks exist, we quantify then by using linear regression models. For example, when the sample is hybridized with *Alexa594* and *Cy5* and *Cy3* are the blank channels, the following models are used:  $G_i = \alpha_1 + \beta_1 Y_i + \varepsilon_i$  and  $R_i = \alpha_2 + \beta_2 Y_i + \varepsilon'_i$ , where  $G$ ,  $Y$  and  $R$  stand respec-



**Figure 1**  
**Bleeding.** First row: Forster data, last row: URGVI data. In the first column the hybridized dye is Cy3 and the empty dye is Cy5, in the second column the hybridized dye is Alexa594 and the empty dyes are Cy3 (black) and Cy5(green). x-axis: signal along the hybridized channel, y-axis: signal along the empty channel(s).

tively for green, yellow and red signals and  $i$  denotes the spot index. Similar models are used with swapped dyes. Estimation is performed using a robust method (R-function *rlm*, [3]) to decrease the effect of outliers. Table 2 contains the estimated parameters, which are low. This shows that the impact of bleeding on the signal is low. The greater coefficient is between Cy3 and Alexa594 (0.07). The weakness of the quantitative influence of bleeding is

confirmed by the values of the standard error of the signal in the different channels: the values for the empty channels are between 6 and 200 times lower than the corresponding values for hybridized targets (Table 3). These conclusions are made for only three dyes and two experimental platforms. It is possible that other dyes or other laser technologies induce a greater bleeding bias.

**Table 1: Bleeding: correlations between hybridized and empty channels. Mean (standard error of the mean) Spearman correlations between hybridized and empty channels.**

Data	Cy5 → Cy3	Cy5 → Alexa	Cy3 → Cy5	Cy3 → Alexa	Alexa → Cy5	Alexa → Cy3
Forster	0.29 (0.06)	0.75 (0.002)	0.39 (0.06)	0.84 (0.11)	0.82 (0.02)	0.83 (0.02)
URGV1	0.13 (0.06)	0.58 (0.04)	0.02 (0.03)	0.47 (0.05)	0.26 (0.03)	0.61 (0.04)

**Correction of bleeding**

When there is a high level of bleeding it is necessary to correct it. A procedure is described in the Methods section in order to fulfill this objective. It necessitates a preliminary experiment with three single-target slides. We have used the bleeding correction for the URGV dataset in the following studies. However the results obtained are very similar with and without bleeding correction, because the importance of bleeding is not sizeable, so the data have not been corrected for bleeding in the following studies.

Note that the bleeding bias is cut down by a complete or partially dye-balanced experimental design, because the measure of the expression difference between two conditions is the mean of the individual measures of this difference taken on each slide. For example, if only one difference is distorted by the bleeding bias, its influence on the mean difference of expression is divided by the number of terms in the mean, which is equal to the number of slides containing the two conditions.

**Normalization of the dye bias**

Dye bias is a well characterized technical bias occurring in two-color microarray. It is mainly due to an incorporation difference between the two dyes. We refer to [4,9] for details on this bias and also to [5] for the gene-specific dye bias. This bias is the most important technical bias and must be corrected before any transcriptome data analysis. The most used method is the lowess correction proposed by [9]. In triple-target microarray, this bias also exists and must be corrected. Unfortunately the lowess correction is defined only for two dyes. Thus for the triple-target microarrays, [2] used the lowess correction for three dye-label combinations per array: Cy5/Cy3, Cy5/Alexa594 and Cy3/Alexa594. However, this procedure does not allow a global correction of the dye bias. In this paper we propose a new method generalizing the lowess correction to correct the dye bias in one step.

Let  $i = 1, \dots, n$  be the gene index ( $i$  is actually the spot index, but in the following we call it loosely the gene index),  $j = 1, \dots, p$  the channel index and,  $\gamma_{ij}$  the  $\log_2$  transformed intensity measure of gene  $i$  along the channel  $j$ . Let  $\bar{Y}_i = \frac{1}{p} \sum_j Y_{ij}$ , be the *mean channel* raw data for gene  $i$  on the log scale, and  $D_{ij} = Y_{ij} - \bar{Y}_i$ , the difference between channel  $j$  and the mean channel for gene  $i$ . We generalize the lowess method by modelling  $D_{ij}$  as follows

$$D_{ij} = f_j(\bar{Y}_i) + E_{ij}$$

and by estimating  $f_j$  via a lowess. The value of the channel  $j$  after normalization of intensity dye-bias is defined by:

$$\tilde{Y}_{ij} = Y_{ij} - f_j(\bar{Y}_i) = \bar{Y}_i + E_{ij}. \tag{1}$$

We point out that if this normalization procedure is applied on a two-color microarray, it leads back to the usual lowess method. Figures 2, 3 and 4 illustrate the result of the normalization procedure on an array issued from the Forster triple-self dataset. Figure 2 contains the plots showing the normalization function for each channel. In the context of two-color microarray, the MA-plot is the main graphical representation for visualizing the effect of the global dye-bias normalization. Figure 3 contains the modified MA-plots for three dyes. In such plots, the  $x$ -axis coordinate is the mean intensity of the three channels  $\bar{Y}_i$  and the  $y$ -axis coordinate is the difference between intensity of channel  $j$  and the mean intensity,  $D_{ij} = Y_{ij} - \bar{Y}_i$ . Figure 3 contains the similar modified-MA-plots for the normalized data. The three usual MA-plot of the normalized data for each couple of dyes are represented in Figure 4.

**Table 2: Bleeding: regression coefficient between hybridized and empty channels. Mean (se) of the regression coefficient (x1 000) between hybridized and empty channels (robust regression method).**

Data	Cy5 → Cy3	Cy5 → Alexa	Cy3 → Cy5	Cy3 → Alexa	Alexa → Cy5	Alexa → Cy3
Forster	1 (1)	6 (3)	0.5 (0.5)	26 (14)	2.5 (0.3)	27 (5)
URGV1	1.0 (0.1)	52 (5)	0.0 (0)	26 (2)	5 (0.4)	70(15)

**Table 3: Bleeding: Standard deviation of the signal in each channel. The hybridized target signal values are in bold.**

Forster experiment			
Slide	Alexa	Cy3	Cy5
6s	<b>8043</b>	277	193
11s	<b>6845</b>	251	191
16s	<b>6704</b>	368	245
10s	1132	1210	<b>6802</b>
17s	585	819	<b>6936</b>
18s	264	<b>7240</b>	219
23s	1033	<b>4188</b>	939

URGV experiment			
Slide	Alexa	Cy3	Cy5
3	<b>1249</b>	73	10
6	<b>1124</b>	96	10
2	65	16	<b>1323</b>
5	78	41	<b>1346</b>
1	48	<b>1313</b>	7
4	45	<b>1739</b>	7

**Validation of the normalization**

The normalization procedure has to be validated on two points: first it must suppress or at least cut the technical bias and second it must not reduce the difference of expression between genes. We have used different experiments to check both points. We first use an analysis of variance (Anova) approach, and then a count of the number of differentially expressed genes.

*Analysis of variance of raw and normalized data*

Kerr et al. [4] proposed to validate a given normalization method by analyzing the raw and the normalized data through the same Anova model. A good normalization method should cut the sum of squares due to technical factors or interactions and should not decrease the sum of

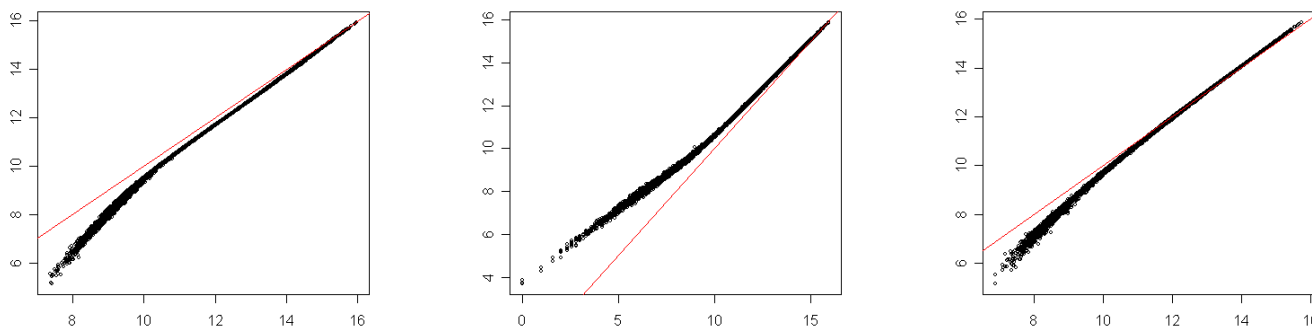
squares due to the interesting biological term under consideration, the gene-condition interaction. As expected, the normalization reduces all the technical biases and the gene-condition interaction is only slightly decreased. This proves that the normalization is effective (see Table 4).

*Number of genes declared differentially expressed*

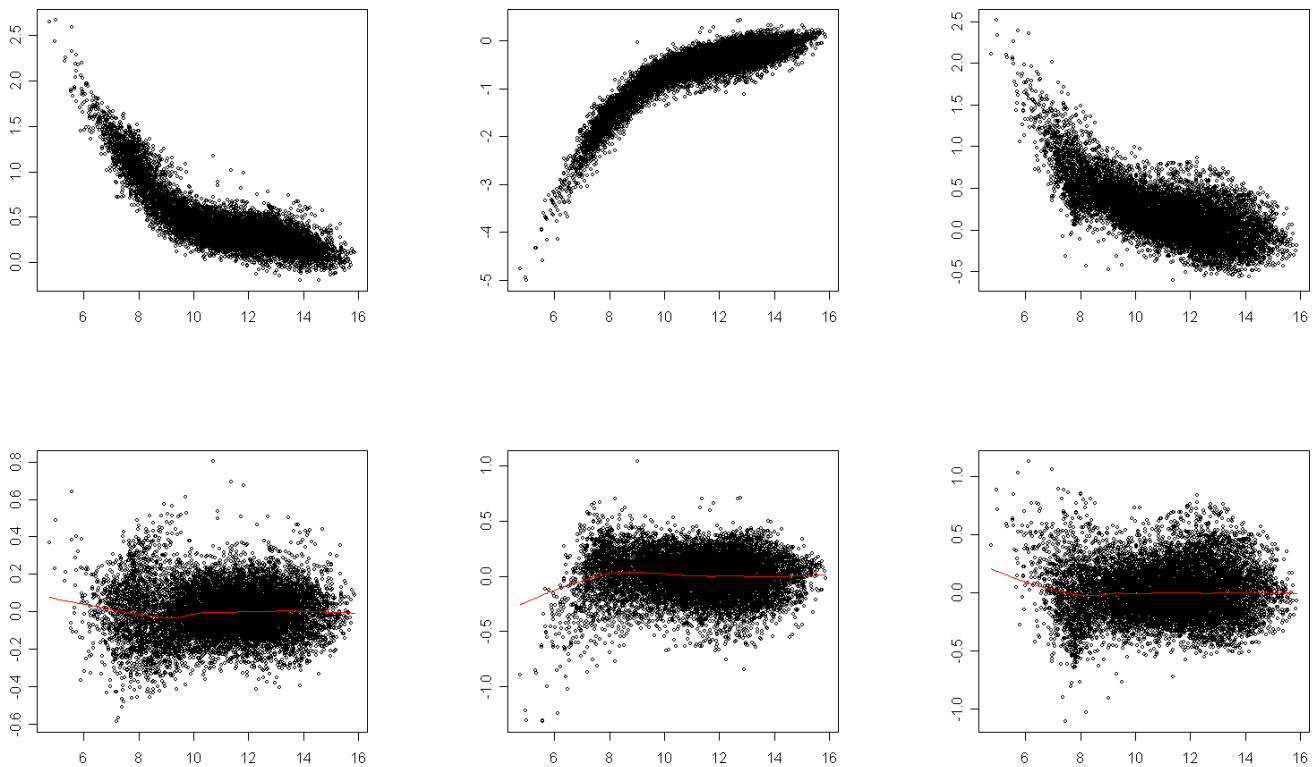
One way for checking the efficiency of a normalization method is to analyze self-experiments, where only one sample is labeled with all the dyes and then hybridized on the same array. In such experiments, no differentially expressed gene is expected. Differential analysis with *varmixt* ([1]) of the triple-self arrays of Forster's experiment and of the URGV2 dataset gives no genes differentially expressed after normalization. A good normalization procedure should not decrease the true difference of expression between genes. We have compared the number of differentially expressed genes for two microarray experiments, studying three conditions:

1. 3 triple-target microarrays (see URGV3 in the Methods Section)
2. 6 two-color microarrays (see URGV4 in the Methods Section), a dye-swap for each comparison between two of the three conditions.

Table 5 states the number of differentially expressed genes for each comparison and for each experiment. The two-color microarrays have been normalized using the usual lowess method and the triple-target microarrays have been normalized by equation (1). All other steps of normalization and the statistical method for differential analysis are the same for the two experiments. The experiment with three triple-target microarrays gives more differentially expressed genes than the experiment with six two-color microarrays, which proves that the proposed normalization for triple-target microarrays does not reduce



**Figure 2**  
**Normalization function.** x-axis: raw data for one channel, y-axis: normalized data from the same channel. First column: Cy5, second column: Cy3, third column: Alexa594.



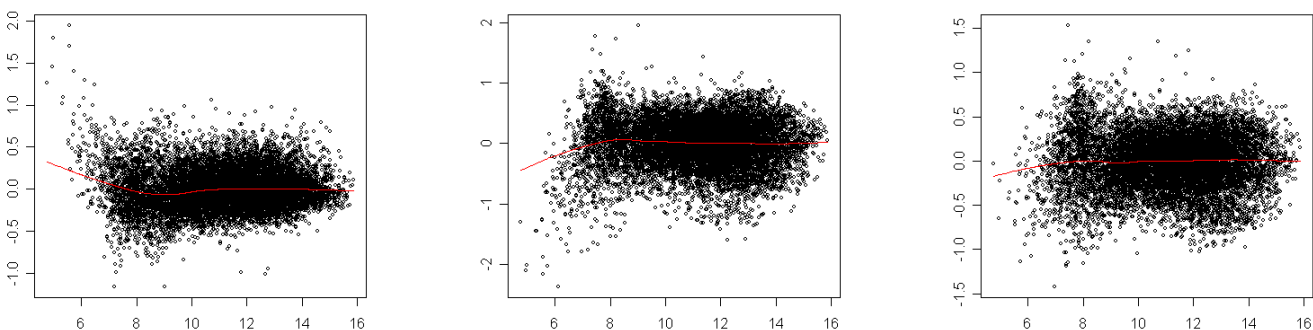
**Figure 3**  
**Modified-MA-plots.** x-axis: mean intensity, y-axis: difference between channel and mean intensities. First row: raw data, last row: normalized data. First column: Cy5, second column: Cy3, third column: Alexa594.

the true difference between gene expression more than the usual lowess method for two dyes does.

**Conclusion**

The proposed normalization procedure is effective: the number of false positives is under control, and the triple-target microarray experiments are more powerful than the corresponding two-color experiments.

There is thus room for improving the routine two-color microarray experiments. The normalization procedure proposed could be used for any number of channels  $p > 2$ , so that it could be tested for four-target microarrays or used to evaluate the bleeding of Alexa 488.



**Figure 4**  
**Usual MA-plots.** x-axis: mean intensity between two channels, y-axis: difference between two channels. First column: Cy5 – Cy3, second column: Cy3 – Alexa594, third column: Cy5 – Alexa594.

**Table 4: Anova Sum of Squares before and after normalization (URGV3 data set)**

Source	Before normalization	After normalization
Array	1191	1184
Dye	13269	11
Array*Dye	425	43
Gene	310836	309177
Array*Gene	6362	6378
Dye*Gene	10595	2739
Condition*Gene	2387	2105
Residual	24890	23929

**Methods**

**Correction of bleeding**

As the bleeding seems to work on a linear scale, a natural idea is to estimate  $p(p - 1)$  bleeding coefficients and correct the raw data using the following expression:

$$\tilde{X}_{ij} = X_{ij} - \sum_{l \neq j} \beta_{lj} X_{il} \tag{2}$$

where  $X_{ij}$  is the raw measure of expression of gene  $i$  on channel  $j$ ,  $\tilde{X}_{ij}$  is the corresponding value corrected for bleeding, and  $\beta_{lj}$  is the coefficient of bleeding from channel  $l$  to channel  $j$ . This bleeding correction works under two assumptions:

1. the bleeding coefficients do not depend on the intensity of the bleeding channel,
2. the effects of the bleeding from several channels are additive on a linear scale.

The first assumption is confirmed by the preceding analysis (see Results Section) and the second one seems realistic. Two ways for estimating the coefficients  $\beta_{lj}$  are possible:

1. use preliminary experiment with  $p$  slides single-target hybridization,
2. use the current data set, with all the  $p$ -target hybridization slides.

**Table 5: Number of genes declared differentially expressed for triple-target and two-color experiments Number of differentially expressed genes (FDR = 5%).**

Comparison	Triple-target experiments	Two-color experiments	Common
C1 versus C2	3353	2188	1925
C1 versus C3	3986	3384	2737
C2 versus C3	4519	3465	2928

The model framework for estimating the bleeding coefficients in  $p$ -target experiments is the following:

$$X_{ija} = \mu + \alpha_i + \gamma_j + \eta_{ij} + \zeta_a + \tau_{ja} + \delta_{c(j,a)} + \theta_{ic(j,a)} + \sum_{l \neq j} \beta_{lja} X_{ila} + E_{ija} \tag{3}$$

where  $a$  is the array index,  $X_{ija}$  is the measure of expression for gene  $i$ , channel  $j$  and array  $a$ ,  $c(j, a)$  is the condition associated with channel  $j$  and slide  $a$ ,  $\alpha_i$  is the gene effect,  $\gamma_j$  the dye effect,  $\eta_{ij}$  the interaction between gene  $i$  and dye  $j$ ,  $\zeta_a$  the effect of array  $a$ ,  $\tau_{ja}$  the interaction between dye  $j$  and array  $a$ ,  $\delta_{c(j, a)}$  the condition  $c(j, a)$  effect,  $\theta_{ic(j, a)}$  the interaction gene-condition and  $\beta_{lja}$  is the bleeding coefficient from channel  $l$  to channel  $j$  for array  $a$ . Note that the global condition effect  $\delta_{c(j, a)}$  is included in the interaction  $\tau_{ja}$ . This is a standard linear model. However the size of the design matrix is huge (more than  $2np$ ) so the computation is not routinely feasible. Even if the computation were feasible, simulations show that there are many confounding effects in this statistical model and consequently the estimates of the  $\beta_{lja}$  are not reliable (data not shown).

Therefore the only possible procedure is to estimate the bleeding coefficients on preliminary one-target slides. This procedure assumes that the coefficients do not depend on the microarray and that the bleeding coefficients of the preliminary single-target experiments are the same as in real  $p$ -target experiments. The bleeding effect may depend on the platform and the technology (laser, PMT tuning, image analysis). This implies that the machine-tuning parameters are not modified during the experiment. For the bleeding correction of the URGV data sets we have used Equation (2) with the coefficients of Table 2. In practice we have only corrected the bleeding from Cy5 to Alexa594, from Cy3 to Alexa594 and from Alexa594 to Cy3.

**Labelling and hybridization protocols for microarray experiments**

Microarray analysis was carried out at the Unité de Recherche en Génomique Végétale (Evry, France), using the CATMA array (Crowe et al., 2003; Hilson et al., 2004), containing 24 576 gene-specific tags from *Arabidopsis thaliana*. Total RNA was extracted from each sample using TRIzol extraction (Invitrogen, Carlsbad, CA) followed by



two ethanol precipitations, then checked for RNA integrity with the Bioanalyzer from Agilent (Waldbronn, Germany). cRNAs were produced from 1  $\mu$ g of total RNA from each sample with the "Message Amp aRNA" kit (Ambion, Austin, TX). Then 5  $\mu$ g of cRNAs were reverse transcribed in the presence of 200 u of SuperScript II (Invitrogen, Carlsbad, CA), in presence of Amino-allyl-dUTP (Sigma-Aldrich, St. Louis, MO). The samples are then labelled by coupling with Cy3 or Cy5 monoreactive dyes (G.E. Healthcare, UK) or Alexa Fluor 594 (Invitrogen, Carlsbad, CA). Labelled samples were purified and concentrated with Qiaquick columns (Qiagen, Hilden, Germany). Slides were pre-hybridized for 1 h and hybridized overnight at 42°C in 25% formamide. Slides were washed in 2  $\times$  SSC+ 0.1% SDS 4', 1  $\times$  SSC 4', 0.2  $\times$  SSC 4', 0.05  $\times$  SSC 1' and dried by centrifugation. The slides were scanned on a Genepix Professionnal 4200A scanner (Molecular Devices Corporation, St. Grégoire, France) and images were analysed by Genepix Pro 6.0 (Molecular Devices, St. Grégoire, France).

#### URGV Dataset description

##### URGV1 single target hybridization microarray experiment

Total RNA sample from *Arabidopsis thaliana* flowers was reverse-transcribed and labelled in a one-dye fashion either with cy3, cy5 or Alexa Fluor 594 and hybridized separately on two slides each (i.e. six hybridizations).

##### URGV2 triple-self hybridization microarray experiment

One pool of total RNA from *Arabidopsis thaliana* roots, leaves and flowers was separated in three aliquots and reverse-transcribed and labelled with the three fluorochromes, then melted and hybridized on the same slides in three technical replicates (i.e. three hybridizations).

##### URGV3 Triple target experiment

Total RNA from *Arabidopsis thaliana* roots, leaves and flowers were labelled independently with the three fluorochromes in a one-dye fashion either with cy3, cy5 or Alexa Fluor 594. Then the three samples were hybridized on the same slide, each being labelled with a different fluorochrome, in three technical replicates with fluorochrome switch (i.e. three hybridizations).

##### URGV4 dual target experiment

Total RNA from *Arabidopsis thaliana* roots; leaves and flowers were labelled independently with the three fluorochromes in a one-dye fashion either with cy3, cy5 or Alexa 594. Then two samples were hybridized on the same slide, each being labelled with a different fluorochrome. Each comparison was performed with a technical replicate with fluorochrome switch: regular dye-swap (i.e. six hybridizations).

#### Authors' contributions

MLMM, JA, ABH and JJD designed the method. MLMM, JA and JJD wrote the manuscript. JA implemented part of the software and performed the statistical analysis. SE made the URGV experiments under the direction of JPR. FM implemented part of the software. All authors contributed to the discussion and have approved the final manuscript.

#### Acknowledgements

This research has been sustained by the Genoplante ANR program ANR05GLA030034 *AgriArray*.

#### References

1. Delmar P, Robin S, Daudin JJ: **Varmixt: efficient variance modeling for the differential analysis of replicated gene expression data.** *Bioinformatics* 2005, **21**(4):502-508.
2. Forster T, Costa Y, Roy D, Cooke HJ, Maratou K: **Triple-target microarray experiments: a novel experimental strategy.** *BMC Genomics* 2004, **5**(1):13.
3. Huber PJ: *Robust Statistics* John Wiley and Sons, New York; 1981.
4. Kerr MK, Afshari CA, Bennett L, Bushel P, Martinez J, Walker NJ, Churchill GA: **Statistical analysis of a gene expression microarray experiment with replication.** *Statistica Sinica* 2002, **12**:203-217.
5. Martin-Magniette M-L, Aubert J, Cabannes E, Daudin J-J: **Evaluation of the gene-specific dye bias in cDNA microarray experiments.** *Bioinformatics* 2000, **21**(9):1995-2005.
6. Reynolds N, Collier B, Maratou K, Bingham V, Speed RM, Taggart M, Semple CA, Gray NK, Cooke Howard J: **Dazl binds in vivo to specific transcripts and can regulate the pre-meiotic translation of mvh in germ cells.** *Human Molecular Genetics* 2005, **14**(24):3899-3909.
7. Staal YCM, van Herwijnen MHM, van Schooten FJ, van Delft JHM: **Application of four dyes in gene expression analyses by microarrays.** *BMC Genomics* 2005, **6**:101.
8. Woo Y, Krueger W, Kaur A, Churchill G: **Experimental design for three-color and four-color gene expression microarrays.** *Bioinformatics* 2005, **21**:459-467.
9. Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP: **Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation.** *Nucleic Acids Research* 2002, **30**(4):e15.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
http://www.biomedcentral.com/info/publishing\_adv.asp

