Normalization of Two-Channel Microarray Experiments: A Semiparametric Approach

Eckel J.E., Gennings C., Therneau T.M., Burgoon L.D.,
Boverhof D.R., Zacharewski T.R.

Technical Report **#69**
July 2004

Normalization of Two-Channel Microarray Experiments: A Semiparametric Approach

Eckel J.E.,[1*] Gennings C.,[2] Therneau T.M.,[1] Burgoon, L.D.,[4,5,6] Boverhof D.R.,[3,5,6] Zacharewski T.R.[3,5,6]

[1]Department of Health Sciences Research, Mayo Clinic, Rochester, MN, 55905, USA; [2]Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, 23298, USA; [3]Department of Biochemistry and Molecular Biology, [4]Department of Pharmacology and Toxicology, [5]National Food Safety and Toxicology Center, and [6]Center for Integrative Toxicology, Michigan State University, East Lansing, MI, 48824, USA

*To whom correspondence should be addressed:

Department of Health Sciences Research

Mayo Clinic

200 First Street SW

Rochester, MN 55905

Tel: (507) 538-6512

Fax: (507) 266-2478

e-mail: eckel@mayo.edu

Keywords: normalization, semiparametric, intensity-dependent bias, dose-response

ABSTRACT

Motivation: An important underlying assumption of any experiment is that the experimental subjects are similar across levels of the treatment variable, so that changes in the response variable can be attributed to exposure to the treatment under study. This assumption is often not valid in the analysis of a microarray experiment due to systematic biases in the measured expression levels related to experimental factors such as spot location (often referred to as a print-tip effect), arrays, dyes, and various interactions of these effects. Thus, normalization is a critical initial step in the analysis of a microarray experiment, where the objective is to balance the individual signal intensity levels across the experimental factors, while maintaining the effect due to the treatment under investigation.

Results: Various normalization strategies have been developed including log-median centering, analysis of variance modeling, and local regression smoothing methods for removing linear and/or intensity-dependent systematic effects in two-channel microarray experiments. We describe a method that incorporates many of these into a single strategy, referred to as two-channel *fastlo*, and is derived from a normalization procedure that was developed for single-channel arrays. The proposed normalization procedure is applied to a two-channel dose-response experiment.

Availability: The **SAS** macro for two-channel *fastlo* is available from the authors and the data used to test the methods is publicly available at http//www.bch.msu.edu/~zacharet/publications/supplementary/ee_dr.

Contact: eckel@mayo.edu

1. INTRODUCTION

Various normalization strategies have been proposed in the literature for two-channel arrays which include, but are not limited to, log-median centering (Jazaeri et al. **2002;** Sotiriou et al. 2003), analysis of variance (ANOVA) models (Kerr et al. **2000;** Wolfinger et al. 2001), and local regression smoothing models (Dudoit et al. **2002;** Yang et al. 2002) for the removal of experimental effects. However, these strategies possess various limitations. For example, simple normaliization procedures such as log-median centering do not take into account overall effects due to arrays, print-tips, and dyes or intensity-dependent biases due to these effects. Thus, although it is a simple normalization procedure to implement, log-median centering will be insufficient for most and if not all situations because it will not eliminate bias due to these overall effects. In contrast, ANOVA models have the capability to normalize with respect to effects due to spot location, arrays, and dyes; however, they assume a linear function and therefore do not account for nonlinear intensity-dependeat biases that are inherently a result of two-channel microarray experiments. The ANOVA model also assumes constant variance, which is often clearly not justified even on the log-transformed scale, as evidenced by an MA-plot on the log-transformed data. As defined by Dudoit et al. (2002), an MA-plot is a representation of the (R,G) data such that $M = \log_2(R/G)$ and $A = \log_2\sqrt{(R \times G)}$ (where R corresponds to the signal intensity produced from the Red channel, Cy5, and G for the Green channel, Cy3). Thus, before implementing an ANOVA approach, a variance-stabilizing transformation is often warranted.

Realizing that nonlinear intensity-dependent biases clearly exist in microarray experiments via MA-plots, Dudoit et al. (2002) and Yang et al. (2002) proposed a within-print tip local regression procedure that is applied to each array separately to normalize the log-ratio intensities from two-channel arrays. Thus, for each print-tip in a 4x4 grid of print-tips on each array, a local regression function f is fit to the log-ratio intensities in the corresponding print-tip. Ideally, the function f would be a horizontal line at $M=0$ for perfectly normalized data. Dudoit et al. (2002) and Yang et al. (2002) provide an example of a relevant print-tip effect and we suspect that such effects are relatively common and deserve attention. Normalizing within a print-tip results in normalization of the data with respect to the lowest fundamental/experimental unit and thus is generally appealing. Although the within-print tip normalization approach is capable of removing intensity-dependent biases in the log ratios, it only normalizes signal intensities from the two channels on a corresponding array and therefore does not normalize with regard to relationships across arrays. After applying the within-print tip normalization the normalized log-ratio intensities for each array should be centered about zero. However, the spread in their log-ratio intensities could vary significantly across arrays. Therefore, Yang et al. (2002) suggest applying a multiple-array-scale adjustment to the within-array-normalized intensities that essentially forces an entire set of arrays to have equivalent spreads in their log-ratio intensities (for implementation of the Dudoit et al. (2002) and Yang et al. (2002) approach see the Limma software in the Bioconductor & R package at http://www.bioconductor.org/).

The Dudoit et al. (2002) and Yang et al. (2002) approach is reasonable for an experiment that utilizes a reference design, and in which it is of interest to express the intensities in terms of a ratio. However, because there are numerous other experimental designs currently being implemented (e.g., the loop design by Kerr et al. 2001), their approach needs to be extended to fit a more general class of experimental designs and for data that need not be expressed as a ratio. In addition, their approach does not account for across-array intensity-dependent effects. Therefore, with respect to normalization, there is still substantial room for improvement and we provide another normalization tool to add to the analyst's toolbox. The proposed normalization procedure for two-channel arrays corrects for intensity-dependent biases both within- and across-arrays and is not specific to any single experimental design. It uses a combination of a parametric model and a nonparametric model and is derived from a procedure that was developed for single-channel oligonucleotide arrays. The true signal intensity for every cDNA is estimated via a parametric model and normalization is applied via a set of local regression curves that corrects for nonlinear intensity-dependent biases. In contrast to the methods of Dudoit et al. (2002), Yang et al. (2002) and the methods available in the Limma software in the Bioconductor & R package, the proposed technique corrects for intensity-dependent biases across channels on a single array as well as across- and within-channels across a set of arrays. Thus, it resembles the cyclic loess procedure of Bolstad et al. (2003) as well as the *fastlo* procedure of Ballman et al. (2003) that were developed for oligonucleotide arrays where each array is normalized against every other array in the experiment. The motivation for the proposed normalization strategy is to balance the effects due to dyes, location (print-tip) and arrays in addition to correcting for intensity-

dependent biases, while maintaining the effect due to the treatment(s) under investigation.

Because the proposed normalization technique is essentially an extension of the *fastlo* procedure developed by Ballman et al. (2003) for single-channel arrays, and so that the reader can gain an appreciation lor the simplicity of the procedure, Section **2** briefly discusses how to apply *fastlo* with regard to single-channel arrays. Section **3** then describes how one-channel *fastlo* is extended to apply to two-channel arrays, which includes a discussion of the additional factors that must be addressed with two-channel arrays in comparison to single-channel arrays. Section **4** provides an application of the proposed normalization procedure to a two-channel dose-response experiment. And lastly, Section 5 is a discussion of the aforementioned work.

2. *FASTLO*: SINGLE-CHANNEL ARRAYS

Ballman et al. (2003) state that *fastlo* can be conceptualized as a loess smooth coupled with a very simple linear model. In general, the data are set up as a matrix where cDNAs comprise the rows and arrays comprise the columns. To implement *fastlo*, first the average signal intensity across the j arrays for the i^{th} cDNA is estimated for each cDNA in the matrix, which corresponds to a vector row means $\hat{y}_{i\cdot}$ that represent estimates of the true signal intensity. Note that the simplest parametric model to estimate the row means is $y_{ij} = \alpha_i + \varepsilon_{ij}$, where y_{ij} is the signal intensity for cDNA i represented on array j and $\hat{y}_{i\cdot} = \hat{\alpha}_i = \bar{y}_{i\cdot}$ is the estimated fit for the row. Second, $\hat{y}_{i\cdot}$ is plotted against $(y_{ij} - \hat{y}_{i\cdot})$, referred to as a modified MA-plot, for each array j separately. Thus, each of the j

modified MA-plots have a point associated with each of the cDNAs in the study. The modified MA-plots depict the bias in using $\hat{y}_{i\cdot}$ to estimate the true signal intensities, Third, a loess curve $f_j(\hat{y}_{i\cdot})$ is fit through the data for each of the j modified MA-plots separately. As stated previously, if the data were perfectly normalized, the function $f_j(\hat{y}_{i\cdot})$ would essentially be a straight line at M=O for each of the modified MA-plots. Fourth, $f_j(\hat{y}_{i\cdot})$ is subtracted from y_{ij}. Lastly, this algorithm is repeated until some convergence criterion has been satisfied. Ballman et al. (2003) suggest that the *fastlo* algorithm has converged when the row means remain unchanged and this typically requires two iterations at most. Ballman et al. (2003) also show that *fastlo* is equivalent to cyclic loess (Bolstad et al. 2003) and computationally appealing since it requires significantly fewer loess smooths in comparison to cyclic less to obtain relatively equivalent results.

A feature of *fastlo* is that while cycling through the algorithm the estimated true signal intensity, the row means of the data matrix, for each cDNA are preserved. An additional desirable feature is that the column means of the data matrix converge to an overall global mean within each experimental unit as defined by the parametric model. For example, suppose the parametric model that contains all cDNA i by treatment k interactions $y_{ijk} = \alpha_i + \beta_k + (\alpha\beta)_{ik} + \varepsilon_{ijk}$ is used to estimate the matrix row means. Note that this is equivalent to implementing *fastlo* separately on each treatment group. Under this parametric model, the average signal intensity across cDNAs for every array within a treatment group (i.e., the matrix column means) will converge to a global mean after implementing *fastlo*. This will be important in two-channel experiments for situations

when it is expected that many of the cDNA clones spotted on the arrays will be differentially expressed. In this scenario, maintaining the average signal intensity within a treatment group is essential.

3. TWO-CHANNEL *FASTLO*

Essentially, the *fastlo* procedure in Section **2** can be conceptualized as the semiparametric model

$$y = \mu + f(\mu) + \varepsilon, \tag{3.1}$$

where μ is an unknown parametric vector that estimates the true signal intensity for each cDNA, $f(\mu)$ is an unknown nonlinear bias function, or set of bias functions, that is assumed to be reasonably smooth, and ε is a vector of random errors. In two-channel arrays, the formation of the parametric component in **(3.1)** is relatively similar to that in one-channel arrays. However, the nonlinear bias function becomes increasingly more complicated because of the additional experimental effects that exist in two-channel experiments. With one-channel arrays, experimental bias due to arrays is the primary concern. On the other hand, in two-channel arrays experimental bias due to additional effects, such as dyes and print-tips for example, are also of concern.

3.I Paranietric Component

As is the case with one-channel arrays, the simplest parametric model to estimate the row means in a two-channel experiment is $y_{ijd} = \alpha_i + \varepsilon_{ijd}$ where y_{ijd} denotes the signal intensity for cDNA i associated with arrayj and dye d. With respect to the semiparametric model in (3.1), this implies that $\mu = \alpha_i$. However, with regard to a k-

sample experiment ($k \geq 2$) where the signal intensity distribution is not expected to be equivalent across the treatment groups due to a majority of the cDNAs being differentially expressed, it may be attractive to normalize within each of the k treatment groups. For example, $\mu = \alpha_i + \beta_k + (\alpha\beta)_{ik}$ contains all cDNA i by treatment k interactions and thus will produce estimates of row means for each treatment group separately. The two parametric models just discussed are simple examples of how to construct the most appropriate set of row means for a two-channel experiment. However, the analyst must realize that the best model for their particular experiment depends on the experimental design and therefore, we have shown that the preceding models can be straightforwardly modified to include any experimental effect of interest.

3.2 *Nonparametric Component* (Bias *Function)*

After the true signal intensities have been estimated via a parametric model, the second step in performing two-channel *fastlo* is to estimate the bias functions via a nonparametric model. With respect to one-channel arrays, a separate bias function is estimated for each of the j modified *MA*-plots, where j denotes the arrays. With two-channel arrays this process becomes ever more complicated with the increasing number of experimental effects as well as the manner in which they should be considered. For example, in a two-channel experiment it is typically of interest to take into consideration intensity-dependent biases due to arrays, dyes, and print-tips (the print-tip effect could reflect an actual print-tip effect and/or a location effect). Therefore, the analyst needs to consider whether the bias functions for these effects are additive or whether interactions exist among the bias functions. And lastly, it is important to discover which effects

actually generate intensity-dependent biases and which, if any can be simply included in the parametric component.

Using a bottom-up modeling approach, the nonparametric component that consumes the most degrees of freedom is a multiplicative model that generates a separate bias function for each array j by dye d by print-tip p combination. For this scenario the nonparametric component in model (3.1) is defined to be $f(\mu) = f_{jdp}(\mu)$. This assumes that the intensity-dependent bias depends on the exact level of each of array, dye, and print-tip effects simultaneously. Conversely, if it is reasonable to assume that the bias function for each of these effects is independent of the other two effects (i.e., the bias functions are additive), then a more parsimonious nonparametric component should be considered. For example, if the array, dye, and print-tip bias functions are additive, then the nonparametric component in (3.1) should be $f(\mu) = f_j(\mu) + g_d(\mu) + h_p(\mu)$. Likewise, there are numerous intermediate models that can be considered as well (Table **2).** Thus, for two-channel *fastlo* the nonparametric component is defined such that it corrects for intensity-dependent biases due to arrays, dyes, and print-tips using the most parsimonious model possible. **A** model fit criterion that takes into account the number of estimated parameters (equivalently, the total degrees of freedom associated with **all** of the bias functions), such as mean squared error (MSE), is used to determine the most appropriate nonparametric component.

4. APPLICATION: DOSE-RESPONSE EXPERIMENT

The objective of this study was to examine dose-dependent changes in hepatic gene expression in liver tissue from immature ovariectomized C57BL/6 mice gavaged with ethynyl estradiol (EE), an orally active estrogen. cDNA microarrays, representing 6,528 cDNA clones, were used *to* assess hepatic changes in gene expression. The five dose-concentrations studied were 0.1, 1, 10, 100, and 250 µg/kg, in addition to an untreated sample (U) and a vehicle control sample (V – referred to as 0 µg/kg). All cDNAs were spotted in duplicate on each microarray, with the bottom half of the array being an exact replicate of the top half. Each half of a microarray was made up of the same 4x4 grid of print-tips, and hence, the entire array consists of **32** blocks. The experimental design, referred to as a 'spokes' design, was replicated in quadruplicate (Figure **1).** Therefore, each of the cDNAs under investigation had 192 total possible observations **(= 48** arrays x 2 dyes x 2 spots). Of the 6,528 cDNA clones spotted on the arrays, only 6,282 clones had at least 90% complete data (≤ 20 abnormal data points; spots were declared abnormal during image analysis) and were used in the proceeding analyses. Each set of dye-swaps (referred to as a spoke) within a design replicate consists of a single independent liver tissue, and each design replicate is an independent biological replicate.

By examining modified MA-plots, intensity-dependent bias is apparent in these data (Figure **2)** and thus neither a simple log-median centering nor a parametric normalization procedure would be appropriate. Note that in Figure 2 the horizontal axis **is** the estimated true signal intensity for every cDNA ($A = \hat{\mu}$) and the vertical axis represents the bias associated with using $\hat{\mu}$ to estimate the true signal intensity ($M = y - \hat{\mu}$). The ultimate goal of these data is to assess dose-response relationships such that (1) each dose

concentration will be standardized relative to the vehicle samples to obtain the effect due to only the dose-concentration and then **(2)** dose-response relationships will be examined. For this reason, expressing the signal intensities as log ratios is not convenient and as a result the Dudoit et al. (2001) and Yang et al. (2001) normalization approach is not applicable. This was the motivation behind developing two-channel *fastlo*.

To apply two-channel justlo in these data the parametric model $\mu = \alpha_i + \beta_k + (\alpha\beta)_{ik}$ was used to estimate the true signal intensity for each cDNA with respect to the spokes design in Figure I, where i (i=1,...,6282) denotes the cDNA and k (k=1,...,7) denotes the treatment. Accordingly, the above parametric model estimates 43,974 average signal intensities (Table 1). A wide assortment of bias functions were considered with these data and **MSE** was used to determine the most parsimonious set of bias functions (Table 2). With these data it is evident from Table **2** that controlling for only a nonlinear array effect eliminates more systematic bias (MSE = 0.2917) than controlling for only a nonlinear dye effect (MSE = 0.3939) or only a nonlinear print-tip effect (MSE = 0.3939). Thus, it appears that most of the experimental bias in these data is linked to the arrays. Modeling the bias functions for arrayj, dye d, and print-tipp in an additive manner is of essentially no benefit since it ultimately results in a MSE (MSE = 0.2897) that is almost equivalent to only controlling for **a** nonlinear array bias.

Moving beyond an additive model and looking **at** interactions, a top-down modeling approach was utilized such that all interactions that included an array effect were considered since the array effect is clearly responsible for most of the systematic bias

with these data. Although the nonparametric model $f(\mu) = f_{jdp}(\mu)$ produced the smallest MSE for these data (**MSE** = 0.2375), it appeared that the model was actually over-fitting the data. This model estimates a bias curve for each array j by dye d by print-tip p combination, such that each curve is estimated from approximately **384** data points. However, there are only a few data points at each end of the curve that inappropriately become influential to the fit of the bias functions (Figure 2a). Therefore, it appears more appropriate to fit a bias curve to each array j by print-tipp combination only (Figure 2b), thus averaging across the two dyes, and then apply a separate additive-shift constant to this curve for each dye at each array j by print-tipp combination. Thus, we are using the nonparametric component $f(\mu) = f_{jp}(\mu) + shift_{jpd}$, which actually includes a parametric component to estimate a mean shift associated with each dye for each array j by print-tip p combination. This model results in a MSE that is only slightly larger than that of the model $f(\mu) = f_{jdp}(\mu)$; however, using toughly 4,608 fewer degrees of freedom (Table 2).

As is the case with the parametric component, the determination of the most appropriate nonparametric component depends on the data at hand. If there had been more data points at each end of the curve for this example, over-fitting with the nonparametric component $f(\mu) = f_{jdp}(\mu)$ would not had been a concern. Thus, it is recommended that in addition to using **MSE** to determine the most parsimonious nonparametric component, visualization tools such as modified *MA*-plots should also be used to determine if the **bias** functions are either under- or over-fitting the data at hand. Modified MA-plots are also effective in determining the functional form of the bias functions. In Figure 2c it is clear

that the amount of dye bias remaining in the current example after adjusting for intensity-dependent biases due to arrays and print-tips is minimal. **As** a result, instead of using a nonlinear function to eliminate dye bias, a simple additive mean shift was applied for each dye at each array-by-print tip combination.

Figure **3** displays the before- and after-effect of the two-channel *fastlo* normalization procedure for the third dose-concentration (10 μg/kg). Before normalization the average signal intensities across systematic effects clearly fluctuates (Figure 3a). However, the fluctuation in average signal intensities across systematic effects was removed after applying two-channel *fastlo* (Figure 3b). Because a within-treatment parametric component was implemented to estimate the true signal intensities for each cDNA, the individual treatment means were maintained and thus the dose-response relationship was preserved.

The Yang et al. (2003) approach and the Limma software in the Bioconductor & R package would have arrived at a similar version of Figure 3 after two sequential steps. First intensity-dependent biases are corrected within each array and second a multiple-array-scale adjustment is applied to the within-array intensity-dependent normalized intensities that forces the entire set of arrays to have equivalent spreads in their normalized intensity values. The proposed two-channel *fastlo* approach accomplishes all this in a single step while also adjusting for intensity-dependent biases across arrays.

5. DISCUSSION

We have described a normalization procedure for two-channel microarray experiments that incorporates various normalization strategies that have previously been described in the literature for both two-channel and single-channel arrays. Because the proposed semiparametric normalization procedure utilizes a linear model, it can be implemented on any experimental design while being capable of handling intensity-dependent biases. Lastly, we have also confirmed that normalizing down to the print-tip effect with the EE dose-response example is optimal with respect to eliminating the most experimental bias.

Kerr et al. (2001) discussed concerns with normalization procedures that use local regression smoothing curves because of the large number **of** parameters that are needed. They suggest that it is unclear how to choose a smoothing parameter because if the smoother is too small there will be over-fitting and if the smoother is too large then the procedure is ineffective. In the analyses of Section **4** we used the default setting (smooth=0.5) in the LOESS PROCEDURE of SAS® (v8.2). With respect to the nonparametric component $f(\mu) = f_{jp}(\mu) + shift_{jpd}$ in the EE dose-response example, a loess fit is estimated from approximately 768 data points (Figure 2b) using approximately five degrees **of** freedom (min = **4.8** and max = **5.4).** Thus, we would argue that this approach resulted in neither over- nor under-fitting of the bias functions.

With respect to normalization, the following two assumptions are often necessary: (1) only a relatively small proportion of cDNAs are differentially expressed or **(2)** there is symmetry in the number of up/down-regulated cDNAs. In two-channel *fastlo*, the construction of the parametric component in model **(3.1)** determines whether the

aforesaid assumptions are necessary. For example, the preceding assumptions are necessary if the parametric component $\mu = \alpha_i$ is used to estimate the true signal intensities, where i denotes the cDNA. After applying two-channel *fastlo*, the average signal intensities across experimental conditions for groups defined in the parametric model should be equivalent (Figure 3). Thus, for the parametric component $\mu = \alpha_i$ the average signal intensities across experimental units as well as across treatment groups will be equivalent. In this scenario it is assumed that the small proportion of differentially expressed cDNAs will be represented as outliers in the modified *MA*-plots. On the other hand, if the nonparametric component $\mu = \alpha_i + \beta_k + (\alpha\beta)_{ik}$ is used to estimate the true signal intensities, where k denotes the treatment, then the aforementioned assumptions are not necessary. In this setting the true signal intensities are estimated within each treatment group and the modified MA-plots are forcing the signal intensities within a treatment group to be equivalent across all other experimental effects, as shown in Figure 3. Thus, the number of differentially expressed cDNAs is not detrimental to the normalization procedure in this scenario.

Given the above discussion, there are problems associated with using the parametric model $\mu = \alpha_i + \beta_k + (\alpha\beta)_{ik}$ to estimate the true signal intensities for each cDNA if there are not an appropriate number of arrays associated with each treatment group. That is, using this model, it is assumed that the estimated true signal intensity represents a treatment effect and not simply an experimental effect such as an array effect. This assumption is inappropriate, for example, if there are only two arrays per treatment group. In a scenario where there are only a small number of arrays per treatment group,

it is difficult to defend whether the effect at hand is actually due to the treatment or in fact is due to a systematic effect, the sample size is simply too small. Therefore, if it is assumed that a large proportion of the cDNA clones spotted on the arrays are going to be differentially expressed, we suggest that the experimental design be chosen carefully with suitable biological replication so that treatment effects can be appropriately estimated.

ACKNOWLEDGEMENTS

We would like to acknowledge Kathy Dawson for her valuable comments with regard to this manuscript as well as Karla Ballman and Ann Oberg for critically reviewing it. Funding was provided to JEE by NCI grant R25 CA92049 and to TRZ by NIEHS grant R01 ES011271.

REFERENCES

Ballman, K.V., Grill, **D.E.,** Oberg, **A.L.** hnd Therneau, T.M (2003) Faster cyclic loess: normalizing **DNA** arrays via linear models. Technical Report #68, Department of Health Sciences Research, Mayo Clinic, Rochester, MN.

Bolstad, B.M., Arizarr, **R.A.,** Astrand, M. and Speed, T.P. **(2003) A** comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**(2), 185-193.

Dudoit, S., Yang, Y.H., Callow, M.J. and Speed, T.P. (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica*, **12,** 111-139.

Jazaeri, A.A., Yee, C.J., Sotiriou, C., Brantley, K.R., Boyd, J. and Liu, E.T. (2002) Gene expression profiles of BRCA 1-linked, BRCA2-linked, and sporadic ovarian cancers. *Journal of the Nationul Cancer Institute,* **94(**13), 990-1000.

Kerr, M.K., Afshari, C.A., Bennett, L., Bushel, P., Martinez, J., Walker, N. and Churchill, G.A. (2002) Statistical analysis of a gene expression microarray experiment with replication. *Statistica Sinicn,* 12,203-217.

Kerr, M.K. and Churchill, G.A. (2001) Experimental design for gene expression microarrays. *Biostatistics,* **2,** 183-20l.

Kerr, M.K., Martin, M. and Churchill, G.A. (2000) Analysis of variance for gene expression microarray data. *Journal of Computational Biology,* **7,** 819-837.

Sotiriou, C., Neo, S.Y., McShane, L.M., Korn, E.L., Long, P.M., Jazaeri, A., Martiat, P., **Fox, S.B.,** Harris, A.L. and Liu, E.T. (2003) Breast cancer classification and prognosis based gene expression profiles from **a** population-based study. *PNAS*, **100**(18), 10393-10398.

Wolfinger, R., Gibson, G., Wolfinger, E., Bennet, L., Hamadeh, H., Bushel, P., Afshari, *C.* and Paules, R. **(2001)** Assessing gene significance from cDNA mciroarray expression data via mixed models. *Journal* of ***Computational Biology*, 8,625-637.**

Yang, Y.H., Dudoit, **S.,** Luu, P., Lin, D.M., Peng, V., Ngai, J. and Speed, T.P. **(2002)** Normalization for cDNA microarray data: a robust composite method for addressing single and multiple slide systematic variation. *Nucleic Acids Research,* **30**(4), e15.

Table 1: ANOVA table for the parametric component in model (3.1), where i denotes the cDNA and k denotes the treatment.

Effect	DF
α_i	6,282
β_k	6
$(\alpha\beta)_{ik}$	37,686
Total	43,974

Table 2: ANOVA table for the semiparametric model in (3.1) such that roughly 5 degrees of freedom were used for each of the loess smooths (smooth=0.5 in the LOESS PROCEDURE of SAS® v8.2), where j denotes the array, d denotes the dye, and p denotes the print-tip, and 43,974 degrees of freedom were used to estimate the parametric component. There are 1,206,144 total possible observations, 6,032 have missing values (marked as abnormal during image analysis), leaving 1,200,112 total observations in all analyses. Under the semiparametric model that does not include any bias functions, $y = \mu = \alpha_i + \beta_k + (\alpha\beta)_{ik}$, the resulting SS_{error} = 456,530.90, DF_{error} = 1,156,138 and MS_{error} = 0.3949.

Bias Functions (BF)	DF_{BF}	SS_{error}	DF_{error}	MS_{error}
$f_j(\mu)$	240	337,124.54	1,155,898	0.2917
$f_d(\mu)$	10	455,341.16	1,156,128	0.3939
$f_p(\mu)$	160	455,383.15	1,155,978	0.3939
$f_j(\mu) + g_d(\mu) + h_p(\mu)$	410	334,761.23	1,155,728	0.2897
$f_{jd}(\mu)$	480	318,814.71	1,155,658	0.2759
$f_{jp}(\mu)$	7,680	294,796.34	1,148,458	0.2567
$f_{jdp}(\mu)$	15,360	270,951.04	1,140,778	0.2375
$f_{jp}(\mu) + Shift_d$	7,682	294,755.10	1,148,456	0.2567
$f_{jp}(\mu) + Shift_{jpd}$	10,752	279,150.16	1,145,386	0.2437

Figure 1: Experimental design. Each arrow represents an array such that the head of the arrow corresponds to the Cy5 dye and the tail of the arrow corresponds to the Cy3 dye. Each node represents a tissue and U, V, D_1, D_2, D_3, D_4 and D_5 correspond to an untreated sample, a vehicle control sample (0 µg/kg), 0.1, 1, 10, 100 and 250 µg/kg respectively.

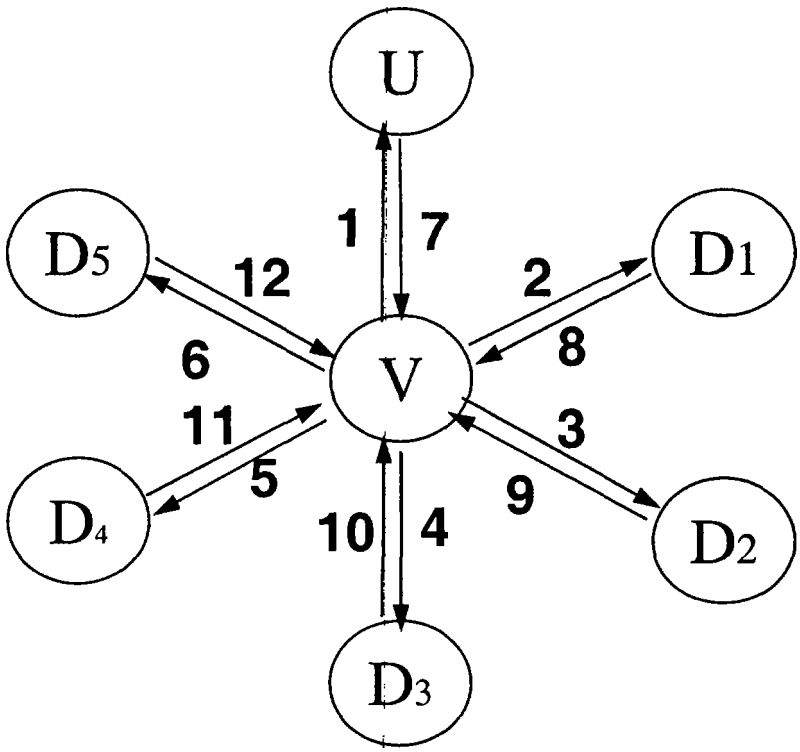
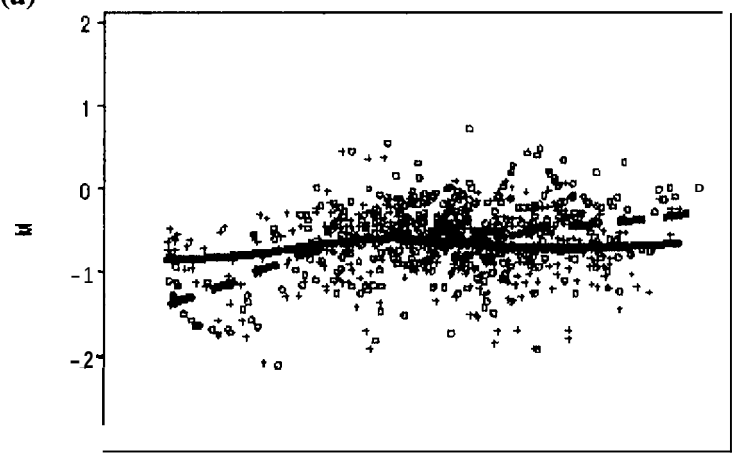
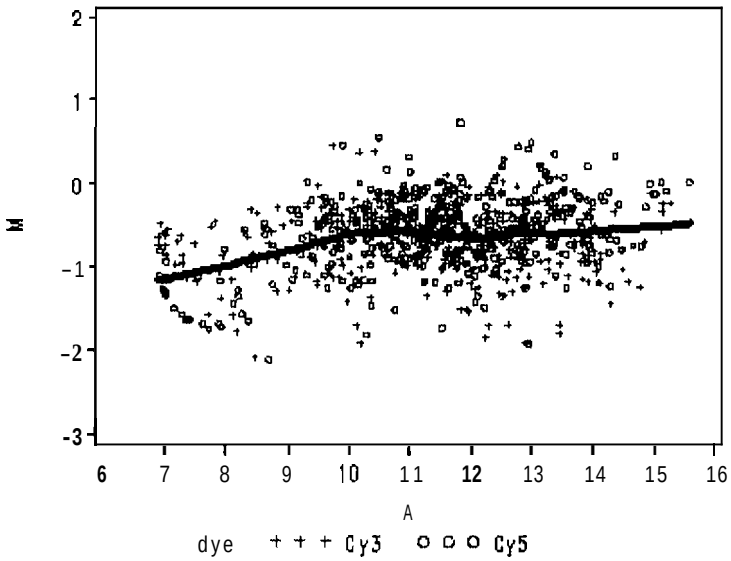


Figure 2: Modified MA-plot for the upper-left-comer print-tip on array 10 for the first design replicate: (a) the estimated nonlinear fit for $f(\mu) = f_{jdp}(\mu)$ (b) the estimated nonlinear fit for $f(\mu) = f_{jp}(\mu)$ (c) residuals after fitting the $f(\mu) = f_{jp}(\mu)$ bias function. The horizontal axis is the estimated true signal intensity for every cDNA ($A = \hat{\mu}$) and the vertical axis represents the bias associated with using $\hat{\mu}$ to estimate the true signal intensity ($M = y - \hat{\mu}$).

(a)





(c)

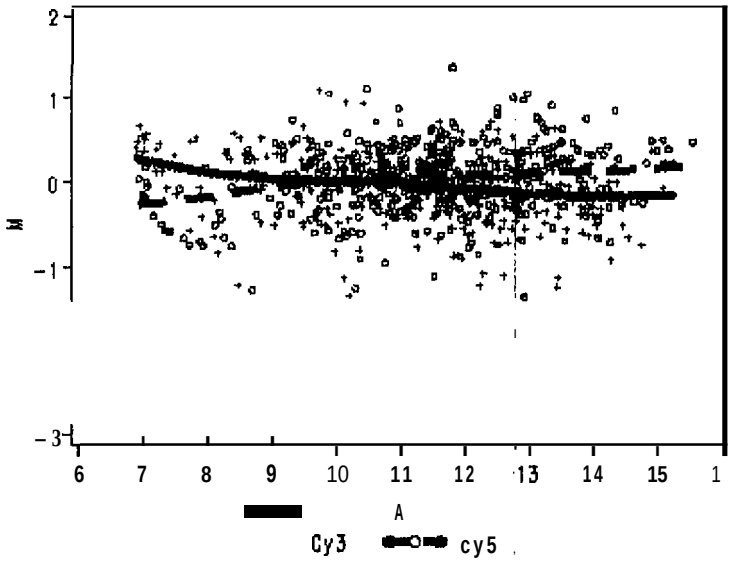

2
1
0
−1
−3
M
6
7
8
9
10
11
12
13
14
15
1
A
Cy3
cy5

Figure 3: Average log*-transformed signal intensities across both technical and biological replicates for the third dose concentration and the lower-right-corner print tip of both array halfs (spot=1 refers to the top half and spot=2 refers to the bottom half of the array), (a) before two-channel *fastlo* and (b) after two-channel *fastlo*. Note that *r1d3s1* denotes the average $\log_2$-transformed signal intensity for design replicate=1, dye=Cy3, and spot=1.