Normalization for Two-color cDNA Microarray Data

*Yee Hwa Yang and Natalie P. Thorne**

Abstract

There are many sources of systematic variation in microarray experiments which affect the measured gene expression levels. Normalization is the term used to describe the process of removing such variation. Two-color cDNA microarray experiments are comparative in nature; therefore, commonly used normalization methods focus on adjusting the value of log-intensity ratios between the red and the green channels. This paper reviews some normalization procedures required to ensure that observed differences across spots both within and between slides are reliably measured. In addition, the paper investigates the possibility of obtaining meaningful single-channel information from two-color microarray experiments after careful single-channel normalization.

Keywords: cDNA microarray; normalization; dye bias; robust smoother; single-channel normalization

1 Introduction

Microarray experiments measure the expression of thousands of genes simultaneously and generate large and complex multivariate datasets. One of the challenges imposed by the enormous growth in this area of biology is the development of computational and statistical tools for processing such datasets. Pre-processing steps such as image analysis and normalization are important aspects of microarray experiments, since they can have a potentially large impact on subsequent data analyses such as clustering or the identification of differentially expressed genes. This paper is concerned with the normalization of two-color cDNA microarray data and examines various procedures applicable to different types of datasets. Normalization is essential to extract reliable measures of the fluorescence intensities and to ensure that the observed differences in intensity indeed reflect differential gene expression and not artefactual bias inherent to the experiment.

We begin in Section 2 with a brief introduction to the biology and technology of cDNA microarrays. This is followed by a discussion in Section 3 on the motivation behind the two main types of normalization procedures: two-channel and single-channel.

*Both authors contributed equally to this work.

Sections 4 and 5 review a number of two-channel and single-channel normalization methods respectively. In particular, Section 5 investigates the possibility of getting useful information from the normalization and analysis of single-channel data from cDNA microarrays. Finally, Section 6 discusses the implications for assessing these different normalization procedures and outlines some of the open questions that remain on this topic.

2 Background on DNA microarrays

DNA microarrays are part of a new class of biotechnologies which allow the monitoring of expression levels for thousands of genes simultaneously. Applications of microarrays range from the study of gene expression in yeast under different environmental stress conditions [6, 10, 13, 22] to the comparison of gene expression profiles for tumors from cancer patients [2, 3, 9, 12, 18, 19]. In addition to the enormous scientific potential of microarrays to help in understanding gene regulation and gene interactions, microarrays are being used increasingly in pharmaceutical and clinical research. Our focus here is on complementary DNA (cDNA) microarrays, where thousands of distinct DNA sequences representing different genes are printed in a high-density array on a glass microscope slide using a robotic arrayer. The relative abundance of each of these genes in two RNA samples may be estimated by fluorescently labeling the two samples, mixing them in equal amounts, and hybridizing the mixture to the sequences on the glass slide. More fully, the two samples of messenger RNA (mRNA) from cells (known as *target*) are reverse-transcribed into cDNA, and labeled using differently fluorescing dyes (usually the red fluorescent dye Cyanine 5 and the green fluorescent dye Cyanine 3). The mixture then reacts with the arrayed cDNA sequences (known as *probes* following the definitions adopted in "*The Chipping Forecast*", a January 1999 supplement to Nature Genetics). This chemical reaction, known as competitive hybridization, results in complementary DNA sequences from the targets and the probes base-pairing with one another. The slides are scanned at wavelengths appropriate for the two dyes, giving fluorescence measurements for each dye for each spot on the array. The underlying assumption in microarray analysis is that these red and green fluorescence intensities for a typical spot represent the amount of mRNA (gene expression) from the corresponding gene in the respective samples. We refer the reader to Schena [21] for a more detailed introduction to the biology and technology of cDNA microarrays.

3 Normalization

Microarray experiments are performed to investigate relationships between different biological samples based on their genes expression. A general approach is to identify genes with relative differential expression between different target samples. The rela-

tive expression from each array is usually measured as the ratio of the red and green fluorescence intensities for each spot. This ratio represents the relative abundance of the corresponding DNA probe in the two mRNA samples. Although these ratios, or fold-changes, provide an intuitive measure of relative expression, they have the disadvantage of treating up- and down- regulated genes differently. Using a log (base 2) scale for intensity is preferred for a number of reasons, including: variation of log-ratios is less dependent on absolute magnitude, and taking the log of the ratio evens out the highly skewed distribution, providing a more realistic sense of variation. For the rest of this review, we base our discussion on log-ratios and log-intensities.

In general, before performing statistical analysis, it is necessary to identify and adjust for artefactual systematic variation in intensities between samples on the same slide and also between slides; that is, variation which cannot be attributed to true biological differences between mRNA samples. This process is known as *normalization.* We define normalization methods based on adjusting the log-ratios as *two-channel* normalization. The need for normalization can be seen most clearly in Figure 1, which shows a plot of a *self–self hybridization.* Here, two identical mRNA samples are labeled with different dyes and hybridized to the same slide. The data are represented by an M versus A plot, or MA plot, where the log-ratios are given by $M = \log_2(R/G)$ and average log-intensity by $A = \log_2\sqrt{RG}$. Because there is no true differential expression in a self-self hybridization, one would expect the red and green intensities to be equal. However, we observe from Figure 1 that the red intensities tend to be lower than the green intensities. This systematic variation may be a consequence of different labeling efficiencies and scanning properties of the Cy3 and Cy5 dyes; different scanning parameters, such as PMT (photo multiplier tube) settings; print-tip, spatial, or PCR plate effects. Furthermore, the imbalance in the red and green intensities is usually not constant across the spots within and between arrays, and can vary according to overall spot intensity A, location on the array, plate origin, and possibly other variables. Section 4 describes procedures for two-channel normalization.

The advantage of relying on the *log-ratio* for measuring relative gene expression within two samples on the same slide rather than considering *log-intensity* values for individual channels is because log-ratios are considered to be more stable than the absolute intensities across slides. Absolute log-intensities are often confounded by spot-spot variation inherent to printed microarrays. This is demonstrated in Figure 2, where we show the spatial plots of an experiment comparing stages E15 and E18 of the olfactory epithelium (OE) in embryonic mice. Panels (a) and (b) show spatial plots of log-intensities from the red channel (Cyanine 5) and green channel (Cyanine 3) respectively. Panel (c) shows the same spatial plot of log-ratios. We observe reproducible spatial effects of the single channels within a slide that are effectively canceled out by the log ratios. This demonstrates the stability of log-ratios in general compared to log-intensities, and provides a clear warning that analysis of single-channel data should proceed with great care.

The main disadvantage of an analysis based solely on log-ratios is that it constrains

researchers to comparative investigations. At times the nature of the research problem requires *single-channel* analysis, for example, when the aim is to identify genes that are expressed in a certain sample, or perhaps at particular time points in a time series experiment. In this case, the quantity of interest is a separate log-intensity measurement for each channel. Compared to log-ratios, separate log-intensities are usually less stable in cDNA microarrays.

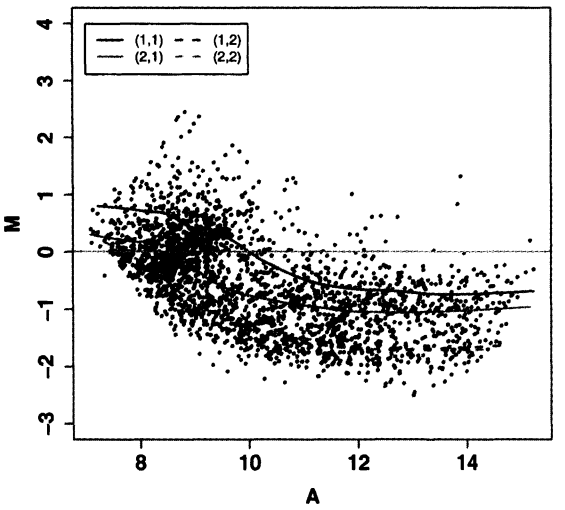


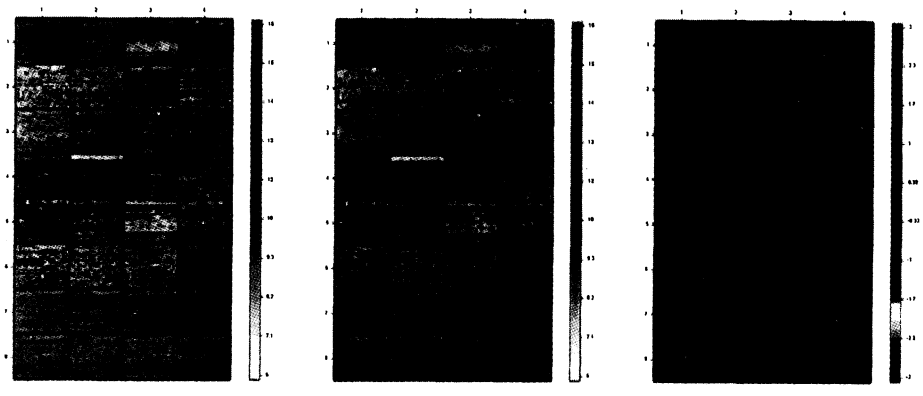
Figure 1: Self-self hybridization illustrating systematic variation. Colored lines indicate the `loess` fit for each of 4 print-tips used to spot the array.

Given the breadth and nature of systematic variation observed in log-ratios, there is an inevitable step-up in complexity of biases for single-channel data. Therefore, the problem of normalization to make the channels from multiple arrays comparable is a more challenging one. Section 5 presents some procedures for single-channel normalization and a discussion on the assessment of single-channel normalization methods.

4 Two-channel normalization

The process of two-channel normalization can be separated into two main components: *location* and *scale*. In general, methods for location and scale normalization adjust the center and spread, respectively, of the distribution of log-ratios. The normalized intensity log-ratios M_{norm} are generally given by

$$M_{norm} = \frac{M - l}{s},$$



(a) (b) (c)

Figure 2: Illustration of the spatial effects that exist in the log-intensities from single-channels that are not observed in the log-ratios. Shown are spatial plots from a single slide in the OE dataset. (a) Spatial plot of red channel. (b) Spatial plot of green channel. (c) Spatial plot of log-ratios.

where l and s denote the location and scale normalization values respectively.

Location normalization

The location value l can be obtained by a wide range of methods. The most commonly used method is *global normalization* with l equal to a constant c and $s = 1$; that is, log-ratios are corrected by subtracting a constant c with $M_{norm} = M - c$. Common choices for this constant c are the median or the mean of the log-intensity ratios (M) for a specified set of genes assumed not to be differentially expressed. There are also many other estimation methods for the constant c. For example, Chen *et al.* [5] propose an iterative method based on ratio statistics for estimating normalization constants. In another approach, Kerr *et al.* [16] and Wolfinger *et al.* [23] propose an ANOVA model for the single channels and perform normalization by including a dye main effect and treatment and array interaction terms in the model. This is followed by adjusting every gene on the array by the same fitted value obtained from model. Figure 3(a) shows an *MA* plot of a mutant *swirl* versus *wild type* comparison of zebrafish prior to normalization. The goal of the *swirl* experiment is to identify genes with altered expression in the mutant compared to *wild type* zebrafish. In this instance, the vast majority of genes on the microarray should show no difference in expression level. This figure depicts a clear dye bias which appears to be dependent on spot intensity. All global methods which subtract the same constant c from every log-ratio on the array do not correct such intensity-dependent biases.

It follows that *location normalization* methods which account for such biases are

often necessary. The intensity-dependent bias is noticeable in an *MA* plot (Figure 3(a)) as a distinct curve in the scatter plot varying with spot intensity. The log-ratios can be normalized by $M_{norm} = M - c(A)$, where $c(A)$ is a function of average spot intensity A. Several intensity-dependent methods have been proposed for location normalization. In Yang *et al.* [24, 25], estimates of $c(A)$ are made using the local scatter plot smoother function `loess` [7, 8] within the software package R. Kepler *et al.* [15] propose a similar approach using a different local regression method. Finkelstein *et al.* [11] present an iterative linear normalization, also known as a robust linear regression, which can be viewed as a constrained version of the robust locally–weighted intensity-dependent normalization.

Figure 3(b) shows boxplots of log-ratios stratified by print-tip groups after intensity-dependent normalization. This figure shows that after intensity-dependent normalization, other systematic biases still remain. We can generalize further to account for other bias by fitting different intensity-dependent curves to different regions of the array: $M_{norm} = M - c_i(A)$, where i indexes different regions of the array. For example, Yang *et al.* [24] to use i to index print-tip groups. Often, systematic differences result from such differences between the print-tips as slight variations in length or in the size of the tip opening, or variable tip deformation after many hours of printing. In addition, because each tip prints DNA spots on different areas of the slide, print-tip groups are proxies for spatial effects on the slide. Figure 3(c) shows an *MA* plot after print-tip group `loess` normalization.

Scale normalization: within and between slides

The effect of location normalization is to center log-ratios around zero by accounting for intensity- and spatially-dependent bias. In addition, it is important to consider *scale normalization*, since large scale differences between multiple slides can lead some slides giving undue weight to an average of log-ratios across slides. One common method of scale normalization is to divide each intensity by the total of the intensities on the slide, so that all slides then have the same total intensity. Yang *et al.* [24] instead propose a robust estimate of scale, such as the *median absolute deviation (MAD)*, for both within-slide and multiple-slide (across slide) scale adjustment. Yang *et al.* [24] also discuss that the need for scale normalization is often determined empirically, as there is a trade-off between the gains achieved by scale normalization and the possible increase in variability introduced by this additional step. In cases where scale differences appear fairly small, it may thus be preferable to perform only a location normalization.

Comparing different methods

We can compare different within-slide normalization methods by examining their effects on the location and scale of the normalized log-ratios M_{norm}. Figure 3(d) shows density plots of the log-ratios for different normalization methods. Without normaliza-

tion (black curve), the log-ratios are centered around -0.5 indicating a bias toward the green (Cy3) dye. A global median normalization (red curve) shifts the center of the log-ratio distribution to zero but does not affect the spread. The dependence of the log-ratio M on the overall intensity A is also still present. Both the intensity-dependent (green curve) and within print-tip group (blue curve) location normalization methods reduce the spread of the log-ratios compared to a global normalization. It is important to note that these approaches implicitly assume that relatively few genes are differentially expressed, or there is no systematic relationship between differential gene expression and intensity or location of the spots on the slide.

Control genes

For most of the methods described, the set of genes to use for the normalization must be decided. In general, the set of genes most appropriate for normalization depends on the nature of the experiment, the amount of observed variation in gene expression, and possibly also on the normalization method applied to the data. Frequently, biological comparisons made on microarrays are of a very specific nature, and differences in gene expression are only detected in a small proportion of genes. In these experiments, it is usual to use most of the genes on the array. Instead of using all genes for normalization, one may use a selected subset of constantly expressed genes. These include the traditional "housekeeping genes", spiked controls, genomic DNA, Microarray Sample Pooled (MSP) titration series [24] and rank-invariant genes. Further details on the effect of different sets of control genes on normalization procedures are provided in [24].

5 Single-channel normalization

Single-channel normalization aims to remove systematic intensity bias, that is, intensity not due to real gene expression, from the red (Cy5) and green (Cy3) channels separately, both within and between arrays. This normalization allows comparisons of absolute intensities between arrays.

Jin *et al.* [14] performed a factorial experiment on age, sex and genotype (two levels for each factor) of *Drosophila melanogaster* flies, where age was the only factor compared within slides. The main effects for the remaining factors were estimable only via single-channel analysis, not by analysis of the log-ratios. Notably, a different experimental design would have enabled all main effects and interactions to be estimated from log-ratios while still maintaining a reasonable level of replication for each comparison type. Here we draw attention to the fact that complex multi-factor designs may not facilitate the estimation of all contrasts of interest from log-ratios alone. In such cases, it may be desirable to recover information from single-channel analysis. Indeed, future complex microarray experiments may be specifically designed to incorporate both log-ratio and log-intensity single-channel analysis methods. In time series experiments, absolute intensity estimates at each time could tell us which genes are

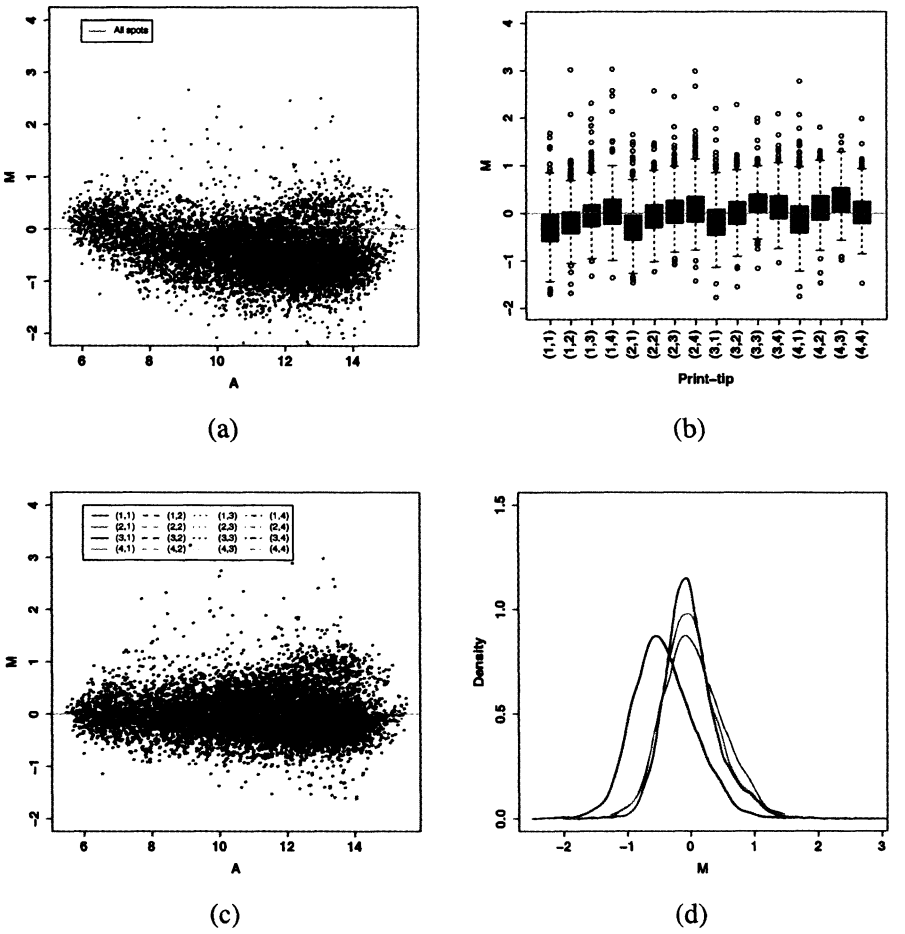


Figure 3: Illustration of two-channel normalization using the *swirl* dataset. (a) *MA* plot before normalization; the green curve corresponds to the `loess` fit for the entire dataset. (b) Boxplots, stratified by print-tip group, of log-ratios after intensity-dependent (`loess`) normalization, but before within print-tip group normalization. (c) *MA* plot after within print-tip group normalization. (d) Density plots of the log-ratios for different normalization procedures. The solid black curve represents the density of the log-ratios without normalization. The red, green, and blue curves represent the densities after global median normalization, intensity-dependent location normalization, and within print-tip group location normalization, respectively.

expressed or not at any given time, or allow estimation of between array single-channel comparisons of time points.

Analysis methods that use ANOVA to model the log-intensities rather than the log-ratios have been investigated by Kerr *et al.* [16] and Wolfinger *et al.* [23]. As mentioned in Section 4, these ANOVA models essentially perform constant global normalization and are therefore inadequate for correcting the nonlinear and spatial systematic variation observed, *e.g.* in Figure 2. Analysis methods that model single-channel intensities have been proposed for "one-color" technologies such as nylon filter arrays and Affymetrix GeneChip. Unlike the cDNA arrays, these technologies generate only a single channel of absolute expression data from each array. Various methods have been proposed [1, 4, 17, 20] to normalize multiple Affymetrix arrays. In this section, we look at extending some of these methods for single-channel normalization of cDNA arrays.

We illustrate the problem of single-channel normalization with a time series dataset examining the olfactory epithelium (OE) of embryonic mice with all possible pairwise dye-swap comparisons of stages E13, E14, till E18. In this paper, we do not explicitly investigate the biological problem of which genes are expressed over time, but rather use the dataset for illustrative purposes only. In addition to the balanced, highly replicated design of this experiment, this dataset is appealing because it contains many controls of different known concentrations. Every print-tip group on every slide includes two different Microarray Sample Pooled (MSP) titration controls of 5 and 6 concentrations respectively [24]. We later outline possible uses for this in assessing single-channel normalization methods.

Single-channel normalization of two-color cDNA microarray experiments can be considered as a two stage process: *within-array* normalization followed by *between-array* (between all channels from multiple arrays) normalization.

In addressing the within-array single-channel normalization problem we see that many parallels can be drawn from the two-channel location normalization approach, such as removing systematic imbalances between the $\log R$ and $\log G$ intensities and correcting for spatial effects within slides. For *dye bias* correction, we can adjust the log-red and log-green intensity by $\log R_p = \log R - \frac{1}{2}c_i(A)$ and $\log G_p = \log G + \frac{1}{2}c_i(A)$ where $c_i(A)$ denotes the normalization adjustment estimated from "print-tip loess" normalization within each slide. We notice that in addition to normalizing spatial effects based on the log-ratios, we must also address spatial effects of the absolute intensity of both channels. This is evident in Figure 2, where we see that even though there is no observable systematic spatial variation in the log-ratios we can still observe reproducible spatial effects of the single-channels. We refer to such arrays as having *systematic spatial variation in intensity within slides*. Efforts are underway to investigate spatial normalization methods which will be robust to extreme local intensity values.

The second stage of single-channel normalization, between-array single-channel normalization, is concerned with comparability of the distributions of log-intensities between arrays. Like the two-channel problem, we wish for the single-channels to have similar scale and location values. At this stage, we do not distinguish which channel is

red and which green, and assume that red-green imbalances were removed by within-array normalization.

For the OE dataset in Figure 4 we see that the distributions of all 60 channels from the 30 arrays are quite varied. The density curves differ in location, variation and shape. Interestingly, the red and green channels within arrays are very close in distribution (data not shown). To adjust for the difference in distribution between channels from multiple arrays, we consider methods developed for *Affymetrix* technology. In particular, we adapt the *quantile normalization* method proposed in Bolstad *et al.* [4]. This method extends the idea of normalizing for equivalent medians or quartiles of the single-channels by requiring *every quantile* across channels be equivalent, and thus forcing each channel to share a common distribution. The distribution is estimated by averaging across channels for each quantile. We refer the reader to Bolstad *et al.* [4] for further details on this method and an algorithm for its implementation. Of particular concern with the use of this method is that replacing quantile values with an average might attenuate log-intensity values, particularly in the tails of the distribution where real expression is potentially affected.

In assessing the performance of these methods, we recommend constructing *MA* plots based on normalized log-intensities to check that dye-biases have been removed. Figure 5 displays *MA* plots for a typical array from the OE dataset showing the effect of different single channel normalization methods. Panel (a) shows the data before any normalization. Between-array quantile normalization (Panel (c)), based on the entire OE dataset, appears to be just as effective at removing intensity dependent dye-bias as the within-array "print-tip loess" single-channel normalization shown in panel (b). We advise using boxplots of the red and green channels to assess red-green imbalances and to check the location and scale of log-intensity distributions after different levels of normalization. It is beneficial to highlight any previously known differentially expressed genes on the *MA* plots to check that they remain distinguishable after normalization.

In the OE dataset, the intensity values of MSP titration controls should remain constant across all 60 channels regardless of what is hybridized. Thus, we can easily determine whether normalization decreases the variability of these control measurements in the single-channels. However, determining bias before and after normalization is more challenging. To measure bias we must be able to compare observed intensities with something known; that is, some truth must be available. The truth regarding absolute intensities for the MSP titration controls is unknown, but there is some knowledge about their relative absolute intensities based on the concentrations of the titration series. We can check (data not shown) that the ratios of intensities between different controls get closer to what we expect. Currently in progress is a variance-bias assessment of the performance of the normalization methods on OE and other similar datasets.

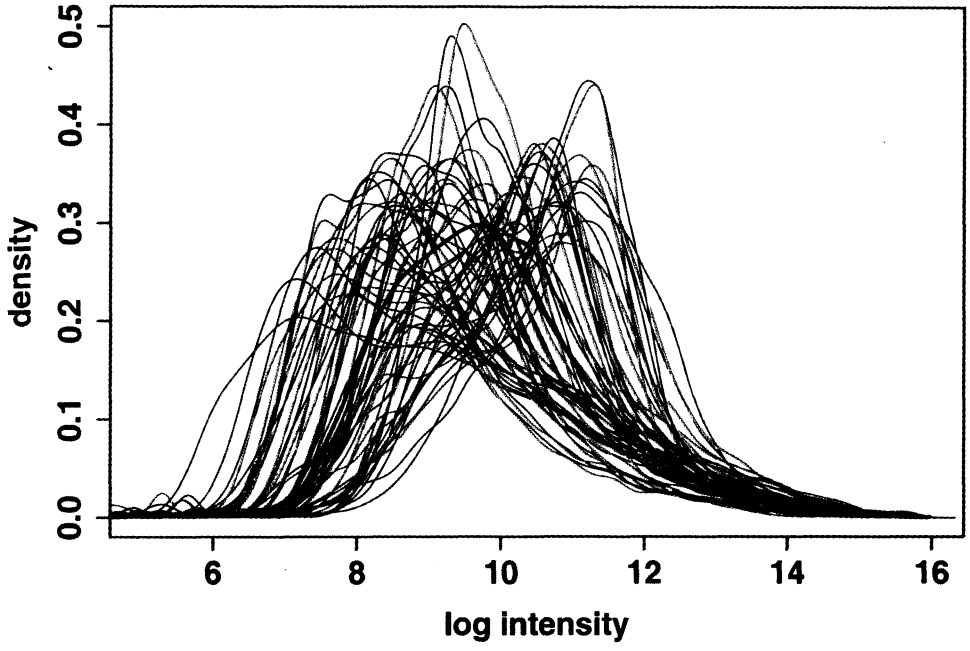


Figure 4: *Single-Channel Quantile Normalization.* Density plots of each of 30 red and green log-transformed single-channels from the OE dataset. The densities of red and green channels within slides are usually very similar. The solid black curve represents the density of all channels after quantile normalization.

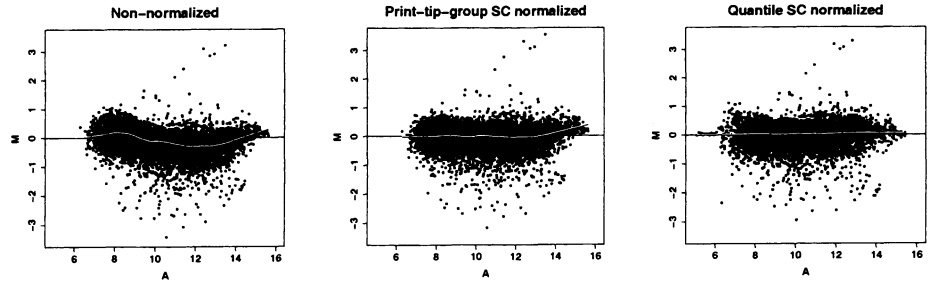


Figure 5: *MA* plots with `loess` curve for a typical array from the OE data (a) before normalization, (b) after single-channel quantile normalization and (c) after "print-tip group" single-channel normalization.

6 Discussion

We have reviewed various normalization approaches applicable to different types of microarray experiments. Most of the normalization work to date is based on two-channel normalization procedures that adjust the log-ratios. However, we have also considered the problem of normalizing single-channels from two-color cDNA microarrays. We have also raised the question of how to assess single-channel normalization is preferable, and what aspects to consider when comparing normalization methods. We neither advocate nor promote the notion of single-channel data analysis in general, but instead suggest that satisfactory normalization of single-channel data is what is lacking for it to be considered a promising option for researchers.

We demonstrate that single-channel analysis is potentially useful in certain circumstances where the nature of the research problem suggests single-channel analysis. Though analyzing microarray data based solely on single-channels is not a new concept [4, 20], limited attention has been given to single-channel normalization of two-color cDNA microarrays. As a place to begin, we have adapted existing procedures from both two-color cDNA and from single-color (*e.g.* Affymetrix) normalizations. The investigation into single-channel normalization raises many other issues of interest, including, in particular, the implications for normalization of log-ratios, for experimental design and analysis and for the replication required for reasonable precision of between array single-channel contrasts.

In any microarray experiment it is important to adjust for the inherent artefactual bias, as well as to understand the assumptions behind any procedure used. In addition, it should be checked that systematic errors are reduced after normalization and that any observed gene expression differences are meaningful (scientific validation). Diagnostic plots such as *MA* plots, spatial plots, density and boxplots can assist in the decision of the level of adjustment needed for both single- and two-channel normalization, and can be used to check that artefacts have been removed by normalization. For example, investigators may decide whether to perform within-slide scale normalization for a dataset by examining boxplots of log-ratios stratified by different print-tip groups.

In general, one should be careful that the gains achieved by further levels of normalization do not introduce a large increase in variability. An important problem that should be addressed is to define formal criteria to assess the effectiveness of various normalization procedures. That is, the issues of bias and variance should be addressed simultaneously. In practice, it is relatively easy to show whether a new normalization method decreases variance. However, it is more challenging to establish that this reduction in variance did not come at the cost of attenuating absolute and relative intensity values (increased bias). To fully address this issue, it is important to obtain a specially constructed dataset with known levels of absolute and differential gene expression, as well as a reasonable number of replications. Examples of such datasets are available for Affymetrix technology `http://qolotus02.genelogic.com/datasets.nsf/` and some initial analyses of these data are available

at `http://www.stat.Berkeley.EDU/users/terry/zarray/Affy/affy_index.html`. In conclusion, until such datasets are available for two-color cDNA microarrays, or until further understanding of the effects of different normalization procedures is gained, it is important to apply normalization algorithms with caution.

Acknowledgments

The authors wish to thank our advisor Terry Speed for all his patience, guidance and encouragement in all aspects of our studies. His wealth of ideas and scientific enthusiasm have made our research both interesting and exciting. We appreciate the discussions, ideas and support given by Gordon Smyth on many aspects of our research, especially on single-channel normalization. Sandrine Dudoit is much appreciated for many discussions on normalization topics, as is Matthew Ritchie. We would like to thank David Kimelman and David Raible at the University of Washington for providing the *swirl* embryos and Katrin Wuennenberg-Stapleton from the Ngai Lab at UC Berkeley for performing the *swirl* microarray experiment. We would like to thank Cynthia Duggan and Carina Howell from the Ngai Lab at UC Berkeley for providing us with the dataset from the OE experiment. Many thanks to our colleagues at the Walter and Eliza Hall Institute who provide a breadth of general and technical support. Finally, we thank the anonymous referees and Darlene Goldstein for comments on an earlier version of this paper.

(Jean) Yee Hwa Yang, Division of Biostatistics, University of California, San Francisco, `jean@biostat.berkeley.edu`

Natalie P. Thorne, Division of Genetics and Bioinformatics, The Walter and Eliza Hall Institute, Australia, `thorne@wehi.edu.au`

References

[1] Affymetrix. *Statistical alogrithms reference guide*, 2001. Technical report.

[2] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson Jr, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt. Different types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, 403:503–511, 2000.

[3] U. Alon, N. Barkai D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine. Broad patterns of gene expression revealed by clustering analysis of tumor

and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences, USA*, 96:6745–6750, 1999.

[4] B. M. Bolstad, R. A. Irizzary, M. Åstrand, and T. P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. In press.

[5] Y. Chen, E. R. Dougherty, and M. L. Bittner. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics*, 2:364–374, 1997.

[6] S. Chu, J. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P. O. Brown, and I. Herskowitz. The transcriptional program of sporulation in budding yeast. *Science*, 282:699–705, 1998.

[7] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74:829–836, 1979.

[8] W. S. Cleveland and S. J. Devlin. Locally-weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83:590–610, 1988.

[9] J. DeRisi, L. Penland, P. O. Brown, M. L. Bittner, P. S. Meltzer, M. Ray, Y. Chen, Y. A. Su, and J. M. Trent. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genetics*, 14:457–460, 1996.

[10] J. L. DeRisi, V. R. Iyer, and P. O. Brown. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278:680–685, 1997.

[11] D. B. Finkelstein, J. Gollub, R. Ewing, F. Sterky, S. Somerville, and J. M. Cherry. Iterative linear regression by sector: renormalization of cDNA microarray data and cluster analysis weighted by cross homology. In *CAMDA 2000*, 2000.

[12] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M.L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286:531–537, 1999.

[13] T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaniants, D. D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, and S. H. Friend. Functional discovery via a compendium of expression profiles. *Cell*, 102:109–126, 2000.

[14] W. Jin, R. M. Riley, R. D. Wolfinger, K. P. White, G. Passador-Gurgel, and G. Gibson. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics*, 29:389–395, 2001.

[15] T. B. Kepler, L. Crosby, and K. T. Morgan. Normalization and analysis of DNA microarray data by self-consistency and local regression. *Genome Biology*, 3(7):research0037.1–12, 2002.

[16] M. K. Kerr, M. Martin, and G. A. Churchill. Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, 7:819–837, 2000.

[17] C. Li and W. H. Wong. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proceedings of the National Academy of Sciences, USA*, 98:31–36, January 2001.

[18] C. M. Perou, S. S. Jeffrey, M. van de Rijn, C. A. Rees, M. B. Eisen, D. T. Ross, A. Pergamenschikov, C. F. Williams, S. X. Zhu, J. C. F. Lee, D. Lashkari, D. Shalon, P. O. Brown, and D. Botstein. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proceedings of the National Academy of Sciences, USA*, 96:9212–9217, 1999.

[19] D. T. Ross, U. Scherf, M. B. Eisen, C. M. Perou, P. Spellman, V. Iyer, S. S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J. C. F. Lee, D. Lashkari, D. Shalon, T. G. Myers, J. N. Weinstein, D. Botstein, and P. O. Brown. Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics*, 24:227–234, 2000.

[20] E. E. Schadt, C. Li, B. Eliss, and W. H. Wong. Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *Journal of Cellular Biochemistry*, 84:120–125, 2002.

[21] M. Schena, editor. *Microarray Biochip Technology*. Eaton, 2000.

[22] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, 9:3273–3297, 1998.

[23] R. D. Wolfinger, G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules. Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, 8(6):625–638, 2001.

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

[25] Y. H. Yang, S. Dudoit, P. Luu, and T. P. Speed. Normalization for cDNA microarray data. In M. L. Bittner, Y. Chen, A. N. Dorsel, and E. R. Dougherty, editors,

Microarrays: Optical Technologies and Informatics, volume 4266 of *Proceedings of SPIE*, May 2001.