CALCULATION OF THE MINIMUM NUMBER OF REPLICATE SPOTS REQUIRED FOR DETECTION OF SIGNIFICANT GENE EXPRESSION FOLD CHANGE IN MICROARRA Y EXPERIMENTS

Michael A. Black

R. W. Doerge

Recommended Citation

CALCULATION OF THE MINIMUM NUMBER OF REPLICATE SPOTS REQUIRED FOR DETECTION OF SIGNIFICANT GENE EXPRESSION FOLD CHANGE IN MICROARRAY EXPERIMENTS

Michael A. Black[1,3] & R.W. Doerge[1,2,3]

[1]Department of Statistics, Purdue University, West Lafayette, IN 47907
[2]Department of Agronomy, Purdue University, West Lafayette, IN 47907
[3]Computational Genomics, Purdue University, West Lafayette, IN 47907

Abstract

Calculations for the number of per gene replicate spots in microarray experiments are presented for the purpose of obtaining estimates of the sampling variability present in microarray data, and for determining the minimum number of replicate spots required to achieve a high probability of detecting a significant fold change in gene expression. Our approach is based on data from control microarrays, and employs standard statistical estimation techniques. We have demonstrated the usefulness of our framework by analyzing two experimental data sets containing control array data. The minimum number of replicate spots required on a treatment array were calculated to achieve detection of a 3-fold increase in expression with 90%, 95% or 99% confidence. The inclusion of replicate spots on microarrays not only allows more accurate estimation of the variability present in an experiment, but more importantly increases the probability of detecting genes undergoing significant fold changes in expression, while substantially decreasing the probability of observing fold changes due to chance rather than true differential expression.

1 Introduction

1.1 Gene expression

A major goal of genomic research involves the determination of gene function, the discovery of which ultimately gives investigators fundamental insight into the ways genes act to affect the traits exhibited by an organism. The ability of a gene to influence an organism's characteristics is the result of the manufacture (expression) of proteins by that gene, the identity of which is determined by the gene's genetic sequence (DNA). By observing the expression patterns of genes (i.e., determining when genes are turned on or off) under various treatment conditions, important clues about gene function can be obtained. Studies which investigate such patterns are termed *gene expression experiments*.

A common method for determining the expression of a gene is to measure the amount of mRNA (an intermediate product in protein production) being produced by that gene. Although not exact, there is a relatively high degree of correspondence

between the volume of mRNA present, and the amount of protein produced. Although the ability to measure mRNA levels is not new, microarray technologies[1] have enhanced the performance of such experiments on a grand scale. Specifically, these techniques allow researchers to obtain expression data from thousands of genes under various treatment conditions in a single experiment.

1.2 Overview of microarray technology

A microarray experiment consists of mRNA extracted from cells under different treatment conditions, and specially coated glass slides (the microarrays) to which spots of genetic material are attached. In the case of Affymetrix arrays, this material comprises oligonucleotides (short genetic sequences) which are synthesized on the array itself, while for cDNA arrays the material is complementary DNA (cDNA), which is printed onto the array by a robotic tool. In both cases, a single spot on the array comprises thousands of strands of identical cDNA which represent the sequence of a single gene, with thousands of spots (and thus thousands of genes) able to fit on a single array. Although the two technologies share many similarities, here we focus solely on the cDNA array.

In its most simple form, the aim of a cDNA microarray experiment is to measure the relative difference in mRNA expression between two different treatment conditions for a collection of genes. This is accomplished by extracting mRNA samples from cells under each treatment condition, tagging each sample with fluorescent labels (often green ($Cy3$) for the control condition, and red ($Cy5$) for the treatment condition), mixing the mRNA samples together, and then placing the combined mRNA on the microarray to allow *hybridization* (i.e., the process by which the single-stranded mRNA in the sample bonds with its matching complementary DNA sequence which is attached to the microarray) to occur. Since the sequence for each gene is unique, each gene's mRNA will only bond with the unique complementary sequence *from that gene*, which means that each spot only collects mRNA that was produced by the gene it represents. Using a laser scanner, the labeled mRNA can be fluoressed, with the strength of the observed signal providing an estimate of the amount of mRNA from each treatment condition that has hybridized at each spot on the array. The results of the laser scanning are stored as a digital image (usually a 16-bit tiff file), which is further analyzed by specialized software that locates the position of each spot within the image, and provides a quantification of the signal intensities recorded at each spot. The set of intensity signals from each of the red and green labels are often referred to as *channels* in array experiments.

In addition to determining the red and green signal intensities, the software also calculates a *background intensity* for each label color at each spot. The background intensity aims to remove factors effecting the intensity levels (such as reflective glare from the array surface) directly associated with the presence of hybridized mRNA at the spot. This practice is referred to as *background correction*. Similar corrections must also be made for differing average signal intensities between the two channels, and across multiple slides. Such *normalization* procedures generally consist of standardizing

the intensities from each channel to a median of one (i.e., by dividing through by the median), thus putting all intensities of the same scale[2].

For an array with only a single spot per gene, the ratio of the two normalized background corrected fluorescence intensities at each spot gives an indication of the relative abundance of mRNA being produced by each gene under each treatment condition, thus allowing the identification of genes whose mRNA expression levels differ between the two treatments. Genes whose mRNA production level changes between the two treatment conditions are said to have undergone *differential expression*.

1.3   Statistical issues

When statistically analyzing microarray data there is a common assumption that the experimental process is relatively free of variability, with each spot providing a completely accurate representation of the abundance of each gene product under each treatment condition. In reality this assumption is far from true, with variability introduced by the process of extracting and labeling the mRNA, and also by the random nature of hybridization. These realizations have prompted the use of replication in microarray experiments[3] as well as the proposal of more comprehensive methods of data analysis[4], including methods which incorporate data normalization into the analysis process[5,6].

An obvious statistical hurdle which is encountered in the absence of replication is the question of how to determine whether an observed fold change in expression is real, or whether it is a result of the inherent variability of the experimental process. The estimation of this inherent variability, therefore, is a key component in any statistical analysis. Since the observed fold change at each spot in an array experiment provides an estimate of the level of differential expression for that gene, most statistical procedures aim to determine which observations represent true differential expression by taking into account the variation present in the data. This can be thought of as investigating the *specificity* of the experiment, that is, determining whether the observed fold changes represent real differential expression. The use of replication, in the form of multiple spots for each gene represented on the array, provides additional information about this variability, and thus provides a more reliable measure of whether an observed fold change represents true differential expression.

Whether true differential expression can be detected, given the amount of variability in the observations is an additional aspect to consider in any microarray experiment. That is, if a gene is up-regulated from one condition to another, what is the chance that an intensity fold change will be observed? This question addresses the *sensitivity* of a microarray experiment. Replication increases the sensitivity of an array experiment by improving the accuracy of the estimates of differential expression, thus improving the chance of observing genes which are differentially expressed.

Typically microarray experiments have suffered from a lack of replicate spots, with each gene being represented by a single spot of genetic material on the array. Here we refer to microarrays to which mRNA from treatment and control samples is hybridized as *treatment arrays*. If the treatment array does not contain replicate spots, then it is

Applied Statistics in Agriculture

difficult to gain information about the within-array sampling variation present in the experiment, thus making it hard to determine whether observed differential expression is real, or simply due to chance.

Despite the limitations of single replicate arrays, there is a scenario in which information about the within-array sampling variation can be obtained, without the need for replicate spots. In the situation where mRNA is extracted from a cells under a single (control) condition, split in two, fluorescently labeled (one half with $Cy3$, the other half with $Cy5$), and then hybridized to an array, the array can be referred to as a *control array*. Even though there are no replicate spots on the array, each spot has two intensity readings from the same sample, and as such they can be considered replicates arising from the same sampling distribution for each gene represented on the array, assuming that appropriate normalization[2] has been performed. Using this as our motivation, we outline model-based methods to estimate the sampling variability present in control array data, and use this to quantify the number of gene replicate spots required to detect a given fold change in expression. The overriding benefit of this approach is that this information can then be used to design future experiments which will achieve desired levels of sensitivity and specificity.

2 Methods

2.1 Model fitting

The basic assumption of the models proposed here is that the intensity reading from each distinct genetic segment (gene) on the array has a unique sampling distribution, with some true mean intensity (i.e., each gene's expression is addressed individually). We also assume that the sampling distribution for each spot's normalized observed intensity is strictly positive and unimodal [3,4,7]. In order to meet these conditions, we focus on models which are based on the log-normal and gamma distributions. These distributions were chosen due to the attention they have recently received in the microarray literature[3,4], and also due to their similarity – the behavior of observations arising from these distributions is very similar, often making it difficult to distinguish between the two when dealing with small samples[8].

For each model we propose one common parameter across all spots, and uniquely identify each gene through the second model parameter. The shared parameter is justifiable since there are only two observations per spot (since there is not enough information in the data to accurately estimate two parameters per gene) , and thus by pooling this information, more accurate estimates of overall variability can be obtained. The forms of the three models for the control array data are as follows,

$$\text{Model 1}: G_{i1}, R_{i1} \sim \text{Log-normal}(\mu_i, \sigma^2), \ i = 1, ..., K$$
$$\text{Model 2}: G_{i1}, R_{i1} \sim \text{Gamma}(\alpha, \beta_i), \ i = 1, ..., K$$
$$\text{Model 3}: G_{i1}, R_{i1} \sim \text{Gamma}(\alpha_i, \beta), \ i = 1, ..., K$$

where K is the number of genes represented on the array, G_{i1} and R_{i1} are the green ($Cy3$) and red ($Cy5$) observed intensities for the ith gene. The subscripted 1 on each of G_{i1} and R_{i1} indicates that only one replicate spot is used for each gene. Unlike the linear models approach[5,6] our models do not provide a means to normalize the data, thus we assume that the data to which the models are applied have been suitably normalized[2] in order to remove any undesirable experimental effects (e.g., additional variability due to fluorescent labeling or geographic location on the array). The two gamma models differ in that the first assumes (like the log-normal model) a constant coefficient of variation, while the second assumes non-constancy. In order to determine which model best fits the data from the control array, maximum likelihood methods are used to obtain estimates of the parameters defined for each of the models, with the better fitting model of the two having the higher log-likelihood value.

2.2   Calculation of replicate numbers

The statistical justification for placing replicate spots on an array is to provide an estimate of the within-array uncertainty present in the intensity readings. For a single gene under one of the log-normal or gamma models, it is possible to calculate the probability of detecting an m-fold increase in expression level, given the estimated model parameters (from the control array) for that gene, for a range of replicate numbers. To facilitate these calculations, a confidence based fold-change threshold must be chosen, such that any gene whose observed fold change is above this threshold is considered as having undergone a significant change in expression level. Based on these results the smallest number of replicates achieving a pre-specified detection probability (e.g., 95% confidence) can be determined for each gene.

Under the log-normal model the calculations involve differences of mean logged intensities for each gene $(\overline{\log(R_{i.})} - \overline{\log(G_{i.})})$, allowing the normal distribution, $N(\mu, \sigma^2)$, to be used. In the case of an m-fold increase in expression between control and treatment, this distribution has mean $\log(m)$ and variance $2\sigma^2/N_i$, where N_i is the number of replicate spots for the ith gene. It is important to note that this result is independent of the mean intensity of the control sample, that is, the log-normal model assumes that the magnitude of the intensity is not a factor to consider in calculating the number of replicates required.

For the first gamma model (again, assuming a m-fold increase) the fold-change estimate can be converted to a proportion of total intensity $(\bar{R}_{i.}/(\bar{R}_{i.} + \bar{G}_{i.}))$ allowing the gamma representation of the beta distribution[9] to be used. This formulation then permits a $\text{Beta}(mN_i\alpha, N_i\alpha)$ distribution to be used for replicate number calculations, the benefit being an eventual closed form calculation of replicate number(s). The calculations are identical for the second gamma model, except that the parameters of the Beta distribution are now gene specific, that is, a $\text{Beta}(mN_i\alpha_i, N_i\alpha_i)$ distribution is used for replicate number calculations.

The major benefit of the second gamma model's replicate calculation is that it takes into account the magnitude of the control intensity, which means that differing numbers of replicates are required to detect significant fold changes for genes with different

average control intensities. Specifically, genes with lower control (green) intensities require a higher number of replicates on the treatment array for the detection of a significant fold-change, since the treatment (red) intensity may be small relative to the amount of sampling error. For a gene with a large control intensity, however, fewer replicates are needed to detect a significant change in expression level, since the fold change between control and treatment could potentially be much larger than the sampling variation. The absence of this feature in the log-normal model and the first gamma model translates to an "average" number of proposed replicates, which makes it possible that too few replicates would be recommended for low intensity genes, decreasing the accuracy of the estimate of per gene variability and thus increasing the chance of failing to detect a gene with a significant change in expression level (type II error).

Both the log-normal and gamma models can be used to determine the probability of false positive results (type I error). This situation involves observing an intensity fold-change greater than the threshold value, when there is actually no difference in expression levels for that gene. A good example of this occurs when a large fold-change is observed on a control array, despite the fact that there is no real difference between the two samples. This can also happen on a treatment array, making it impossible to separate real changes in expression level from those due to sampling variation. If the sampling variation is large enough, this situation can easily occur for a single replicate, but as the number of replicates increases, the chance of such an occurrence is reduced, since the observations are averaged.

3   Results

To illustrate our model-based replicate calculation approach, we utilize two publicly available data sets which contain control chip data. The first is the *Escherichia coli* data originally published by Richmond *et al.*[10] and further analyzed by Newton *et al.*,[4] and the second is the yeast sporulation data of Chu *et al.*[11]. The control chip data in both cases consist of an array with one replicate spot for each of the genes represented on the array, to which a control sample of mRNA (half tagged with $Cy3$, half with $Cy5$) was hybridized. The arrays were then scanned to produce red and green intensity values for each spot. Although there was only a single replicate for each gene, since both the red and the green samples were derived from the same source in each case, they can be considered as a pair of replicates from the same sampling distribution. The data were normalized in the manner of Newton *et al.*,[4] by subtracting the background intensity for each gene and then dividing the result by the sum of all the positive background subtracted intensities on the chip. These intensities are then multiplied by a factor of 10,000 to provide a suitable scale for analysis[4]. This procedure reduces the influence of background fluorescence on the observed intensities. No evidence of effects relating to the fluorescent labeling of the samples was noted in either data set.

Standard maximum likelihood methods and results[12] were used to fit the log-normal model to the data from each control chip[10,11], while the EM algorithm[13] was employed

to fit the gamma models. These methods gave maximum likelihood estimates of the μ_i and σ^2 for the log-normal model and (α_i, β) and (α, β_i) parameters for the gamma models, as well as a log-likelihood value for the fit of each model.

Based on the parameter estimates obtained from our model fitting process, each model can be used to determine the minimum number of replicates required to detect an m-fold increase in mean expression intensity with various degrees of certainty for each of the control chips[10,11]. For illustrative purposes, we chose to investigate the sensitivity and specificity of a fixed fold-change threshold of 2.5 for detecting a 3-fold change in gene expression as has been done elsewhere in the literature [14]. Based on this approach, the results presented here reflect the sensitivity and specificity of using a 2.5-fold cutoff to detect a 3-fold change in expression in an experiment involving the *E. coli* or yeast arrays[10,11]. The choice of these values is designed to illustrate the methods presented here, and are not recommended as a general procedure. In order to achieve desired levels of sensitivity and specificity further investigation would be needed over a range of (possibly variable) thresholds.

For the log-normal model, the probability, P^*, of detecting a 3-fold increase using a 2.5-fold cutoff in observed intensity of gene i (averaged over N_i replicates) is given by,

$$P\left(z > log(2.5)\right) = P^*, \text{ where } z \sim \mathrm{N}(log(3.0), \frac{2\hat{\sigma}^2}{N_i}). \tag{1}$$

Inverting this equation demonstrates that N_i depends only on P^* and $\hat{\sigma}^2$, that is, the number of replicates required depends only on the desired certainty of detecting a 3-fold increase and the common variance term, and not the specifics of an individual gene. Similarly, for the first gamma model P^* is given by

$$P\left(w > \frac{2.5}{1 + 2.5}\right) = P^*, \text{ where } w \sim \mathrm{Beta}(3N_i\alpha, N_i\alpha). \tag{2}$$

In contrast, for the second gamma model P^* is given by

$$P\left(w > \frac{2.5}{1 + 2.5}\right) = P^*, \text{ where } w \sim \mathrm{Beta}(3N_i\alpha_i, N_i\alpha_i), \tag{3}$$

so that the number of replicates for gene i, N_i, is dependent on both P^* and α_i (the location parameter for gene i). Simply put, for the gamma model, the number of replicates required to detect a 3-fold change depends on both the desired level of certainty, and the mean control intensity of gene i.

Similar calculations can also be made to investigate the probability of false positive detections. For the log-normal model this probability is given by,

$$P\left(z > log(2.5)\right) = P^*, \text{ where } z \sim \mathrm{N}(0, \frac{2\hat{\sigma}^2}{N_i}). \tag{4}$$

which corresponds to the probability an observing a fold change above the cutoff of 2.5, given that there was no true differential expression. For the gamma models the

probability is

$$P\left(w > \frac{2.5}{1+2.5}\right) = P^*, \text{ where } w \sim \text{Beta}(N_i\alpha, N_i\alpha) \text{ or } \text{Beta}(N_i\alpha_i, N_i\alpha_i). \quad (5)$$

These methods are now applied to the control array data sets[10,11]. For both data sets the distributions have been truncated to reflect the data points in which we are most interested. That is, the lower intensity observations which make up 95%-99% of the data sets.

3.1 *E. coli* data

The Richmond *et al.*[10] control chip data consist of an array of 4290 genes. Table 1 contains the estimated scale parameters and log-likelihood fits for the log-normal and gamma models, applied to the *E. coli* data[10]. The log-likelihood values indicate that the first gamma model tends to fit the data more closely than do the other models. Fig. 1 (a,c,e,g) contains plots of the observed data, and plots of simulated data based on the fitted gamma and log-normal model. The appearance of the plots suggests that the variability predicted by the second gamma model underestimates that of the data, while the log-normal slightly overestimates this variability. The first gamma model appears to produce results which most closely resemble the data.

Using equations (1), (2) and (3), graphs (Fig. 2) relating the required number of replicates for given detection probabilities can be produced specifically for the *E. coli* data[10]. The graphs differ for each model due to the fact that for the second gamma model the mean intensity affects the number of replicates required, whereas this is not true for the log-normal model, nor for the first gamma model. Fig. 2a demonstrates the use of the log-normal model for the *E. coli* data[10], and relates the number of replicates required to detect a 3-fold increase in expression level with 90%, 95% and 99% probability. Since the scale of the horizontal axis is continuous, the next highest integer is suggested in order to maintain each confidence level, that is, three replicates for 90% confidence, four replicates for 95% and six replicates for 99%. Similar data is shown in Fig. 2c for the first gamma model, with two replicates required for 90% confidence, three replicates for 95% and five replicates for 99%.

Fig. 2e provides results for the required number of replicates under the second gamma model (detection probabilities 90%, 95% and 99%), although the presentation is different to allow for the dependence of the replicate numbers on the mean control intensity. For the Richmond *et al.*[10] experiment the numbers of replicates required (based on the mean control intensity observed on the control chip) for the gene of interest is determined by locating the approximate mean intensity (vertical axis) value intersecting the confidence level curve related to the desired detection probability, and reading off the necessary number of replicates (rounding to the next highest integer) required for an individual gene (horizontal axis). An example is illustrated on the graph for any gene with a normalized control intensity of 10.0 (this value is approximately the mode of the intensity distributions for both normalized data sets). In this case the gamma model predicts that two, four or seven replicates of each gene would be needed

to detect a 3-fold increase in expression with respective 90%, 95% or 99% confidence when using an observed 2.5-fold significance threshold.

The situation of false positive detection for the log-normal model is considered in Fig. 2b. The dashed lines on the graph indicate the probability of observing a false positive result in an experiment with only a single replicate spot for each gene. As shown in Fig. 2b, the approximate value 0.016 is in rough agreement with the proportion of false positives (0.009) observed in the control array data. As the number of replicates increases, the chance of observing a false positive result decreases dramatically. The false positive probabilities for the first gamma model are shown in Fig 3d. Fig. 2f contains similar data for the second gamma model, although the probability now depends on the mean control intensity for each gene. As for the other two models, this probability decreases as the number of replicate spots for each gene is increased.

3.2   Yeast sporulation data

The Chu *et al.*[11] control chip data consists of an array of approximately 6200 genes (a single replicate spot per gene), each with a a control sample split in two and labeled with $Cy3$ and $Cy5$ as in the *E. coli* data[10].

The estimated scale parameters and log-likelihood fits for the log-normal and gamma models are shown in Table 1. As for the *E. coli* data[10], the first gamma model appears a better log-likelihood fit than the other two models. Again the spread of the points in Fig. 1 suggests that the variability of the data is underestimated by the second gamma model, and slightly overestimated by the log-normal model, but this time the first gamma model severely underestimates the variability of the data.

Using equations (1), (2) and (3) for detection probability calculation, the minimum number of replicates needed to detect a 3-fold increase in expression were calculated for various confidence levels. Fig. 3a contains replicate number *versus* detection probability for the log-normal model. These results predict that two, three, or four replicate spots for each gene would be needed on a treatment chip to detect a 3-fold increase with 90%, 95% and 99% confidence respectively, when an observed 2.5-fold increase is used to declare significance. Note that these numbers are less than those for the *E. coli* data, and are the result of the sporulation data containing less sampling variation. For the first gamma model (Fig. 3c) only one replicate is required for 90% and 95% confidence, while 2 are required for 99% confidence.

For the second gamma model, the required number of replicates per gene is presented in Fig. 3e. Again, an example is given for a gene with a normalized control intensity of 10.0. For the sporulation data the second gamma model predicts that any such gene would require one, two, or three replicates to detect a 3-fold increase in expression with 90%, 95% and 99% confidence, respectively. Again, the smaller amount of sampling variation in the sporulation data results in a noticeably lower number of replicates being required than for the *E. coli* data. The issue of false detections in the sporulation data is addressed for the log-normal and gamma models in Fig. 3e and 3f. The small amount of sampling variation present in the data, however, makes it extremely unlikely for a large fold change to be seen when no difference between

treatment and control exists, even in the case of genes with very low intensities.

4 Summary

Preliminary analysis utilizing control array data can be employed for the purpose of designing experiments which minimize the number of replicates that are needed to achieve a high probability of detecting significant fold changes on a treatment array. Furthermore, using a gamma model, we have shown that replicate calculations can even be performed on an individual gene basis. That is, the number of replicates needed for a treatment array can be calculated for each gene, with genes having a high control intensity receiving less replicates, and those with a low control intensity receiving more.

Despite the flexibility of the second gamma model, and the better log-likelihood fit of the first gamma model, the approach which appeared to give the best visual fit to the data was the log-normal model. The fact that both gamma models provided a good (log-likelihood) fit is most likely due to the large number of very small observations (i.e., intensities close to zero) in the data. This underlying data structure makes it probable that the gamma models systematically underestimated the variability of the relatively small number of high intensity observations, as can be seen in Fig. 1. Based on these results, we believe that the log-normal distribution is an appropriate choice for the investigation of sampling variability in the microarray experiments examined here.

By analyzing two separate data sets, the effect of differing amounts of sampling variability has been seen. The smaller sampling variation present in the Chu *et al.* data[11] clearly resulted in both models recommending less replicates per gene on a treatment chip than for the Richmond *et al.* data[10]. This underscores the importance of performing a control chip experiment to assess sampling variability before constructing a treatment chip, and also illustrates why a standard number of replicates cannot be proposed to cover all microarray experiments.

Acknowledgments

References

1. Schena M., Shalon D., Heller R., Chai A., Brown P., & Davis R. (1996) Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *PNAS* **93**:10614–10619.

2. Hegde P., Qi R., Abernathy K., Gay C., Dharap S., Gaspard R., Hughes J., Snesrud E., Lee N., & Quackenbush J. (2000) A concise guide to cDNA microarray analysis. *BioTechniques* **29**:548–562.

3. Lee M., Kuo F., Whitmore G., & Sklar J. (2000) Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. *PNAS* **97**:9834–9839.

4. Newton M., Kendziorski C., Richmond C., Blattner F., & Tsui K. (2001) On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *Journal of Computational Biology* **8**:37–52.

5. Kerr M., Martin M., & Churchill G. (2000) Analysis of variance for gene expression microarray data. *Journal of Computational Biology* **7**:819–837.

6. Kerr M. & Churchill G. (2001) Statistical design and the analysis of gene expression microarray data. *Genetical Research* **77**:123–128.

7. Chen Y., Dougherty E., & Bittner M. (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics* **2**:264–374.

8. Kappenman R.F. (1982) On a Method for Selecting a Distributional Model. *Communications in Statistics, Part A – Theory and Methods* **11**:663–672.

9. Johnson N., Kotz S., & Balakrishnan N. (1994) *Continuous Univariate Distributions*. Wiley, New York.

10. Richmond C., Glasner J., Mau R., Jin H., & Blattner F. (1997) Genome-wide expression profiling in Escherichia coli K-12. *Nucleic Acids Research* **27**:3821–3835.

11. Chu S., DeRisi J., Eisen M., Mulholland J., Botstein D., Brown P., & Herskowitz I. (1998) The transcriptional program of sporulation in budding yeast. *Science* **282**:699–705.

12. Casella G. & Berger R.L. (1990) *Statistical Inference*. Wadsworth, Belmont.

13. Dempster A., Laird N., & Rubin D. (1977) Maximum Likelihood From Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society, Series B, Methodological* **39**:1–22.

14. Schenk P., Kazan K., Wilson I., Anderson J., Richmond T., Somerville S., & Manners J. (2000) Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. *PNAS* **97**:11655–11660.

Table 1. Results of Model Fitting

Model	E. coli data [10]		Yeast sporulation data [11]	
	Common parameter	Log-likelihood	Common parameter	Log-likelihood
Log-normal	$\hat{\sigma}^2=0.0458$	-18342.87	$\hat{\sigma}^2=0.0151$	-18127.07
Gamma 1	$\hat{\alpha}=45.7039$	-17356.58	$\hat{\alpha}=132.61$	-16715.70
Gamma 2	$\hat{\beta}=0.319$	-17911.07	$\hat{\beta}=0.124$	-19055.26

Estimated common parameters and log-likelihood values after fitting the gamma and log-normal models to the E. coli control chip data[10] and the yeast sporulation control chip data[11].
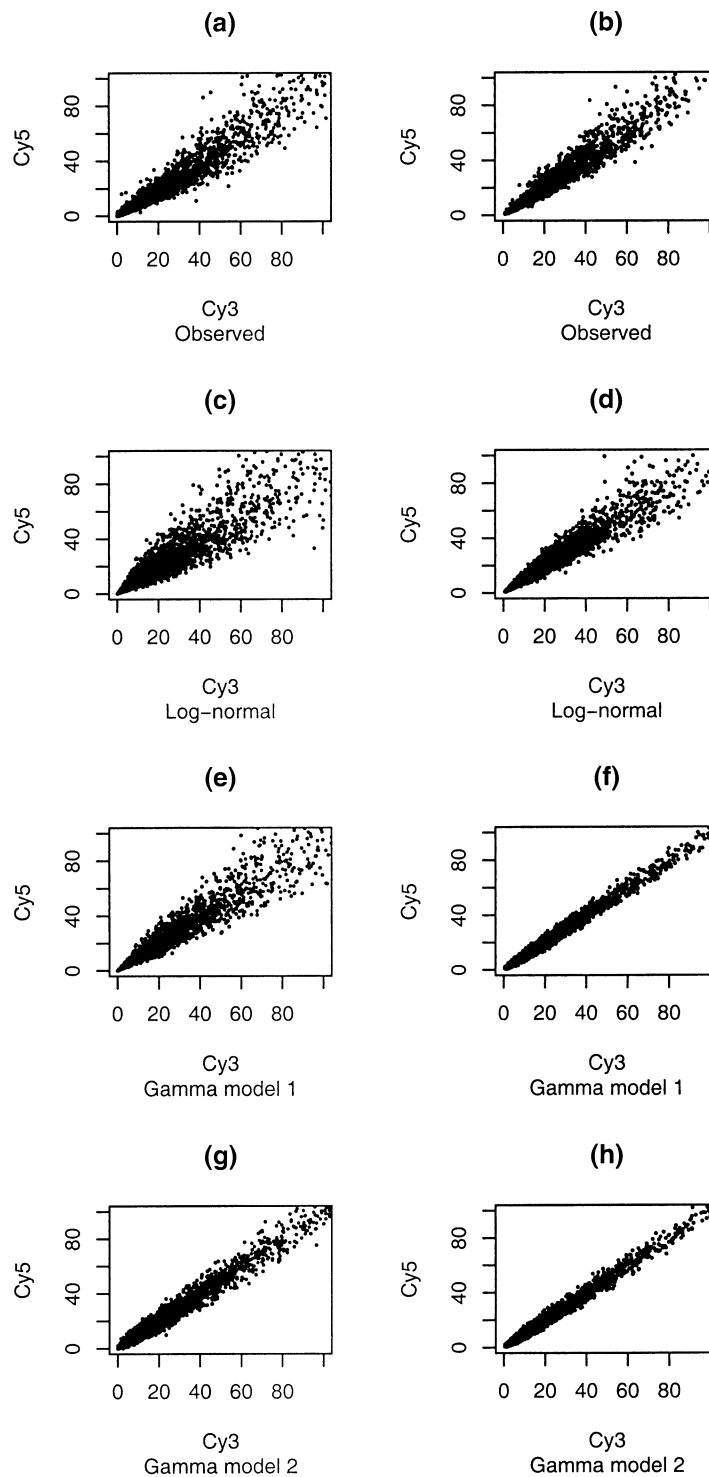
Fig. 1. Plots of observed **(a,b)** and simulated (Log-normal model **(c,d)**, Gamma model 1 **(e,f)**, Gamma model 2 **(g,h)**) $Cy3$ and $Cy5$ data. Plots **(a,c,e,g)** relate to the *E. coli* control array data[10], and **(b,d,f,h)** relate to the yeast sporulation control array data[11].
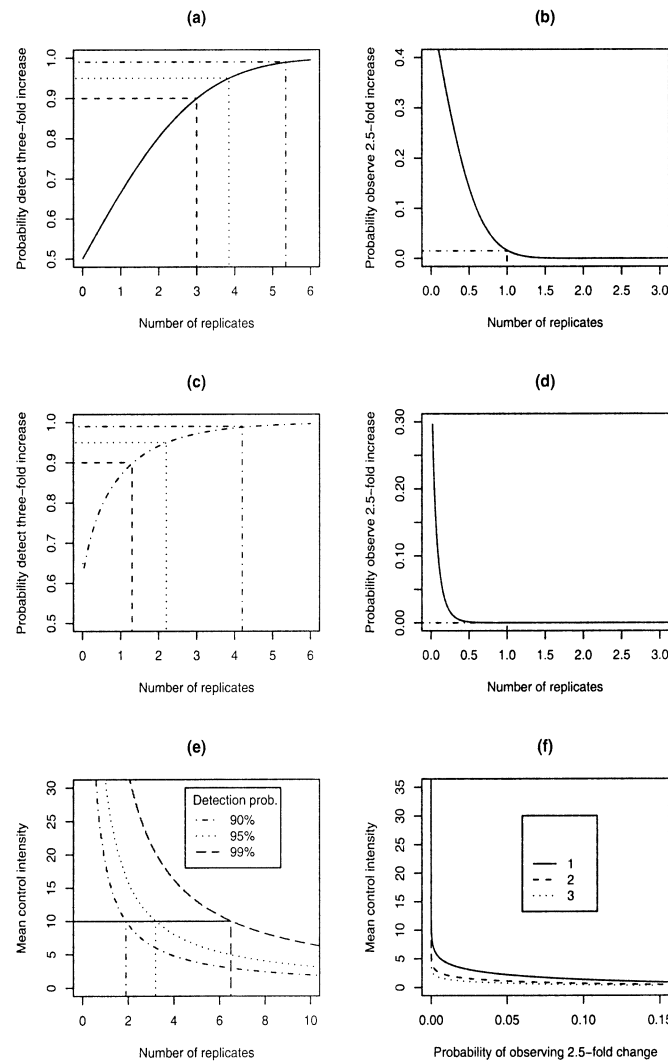
Fig. 2. *E. coli* control array data[10]. **(a)** Log-normal model probability of detecting a 3-fold increase in expression level using a 2.5-fold observed mean cutoff for N replicates of a single gene. Lines indicate 90%, 95% and 99% confidence. **(b)** Probability of false positive (observing a mean 2.5-fold increase in intensity when no true difference in expression level exists) as the number of replicates increases under the log-normal model. Dashed lines on graph indicate the probability of false positive when only one replicate is used. **(c)** Probability of detecting a 3-fold increase in expression level using a 2.5-fold observed mean cutoff for N replicates of a single gene under the first gamma model. Lines indicate 90%, 95% and 99% confidence. **(d)** Probability of false positives under the first gamma model. **(e)** Number of replicates required to detect a 3-fold increase in mean control intensity using a 2.5-fold observed mean cutoff, based on the second gamma model. Curves represent 90%, 95% and 99% confidence levels. Straight lines give example of number of replicates needed for gene with mean control intensity of 10.0. **(f)** Probability of false positive for one, two, or three replicates under the second gamma model.
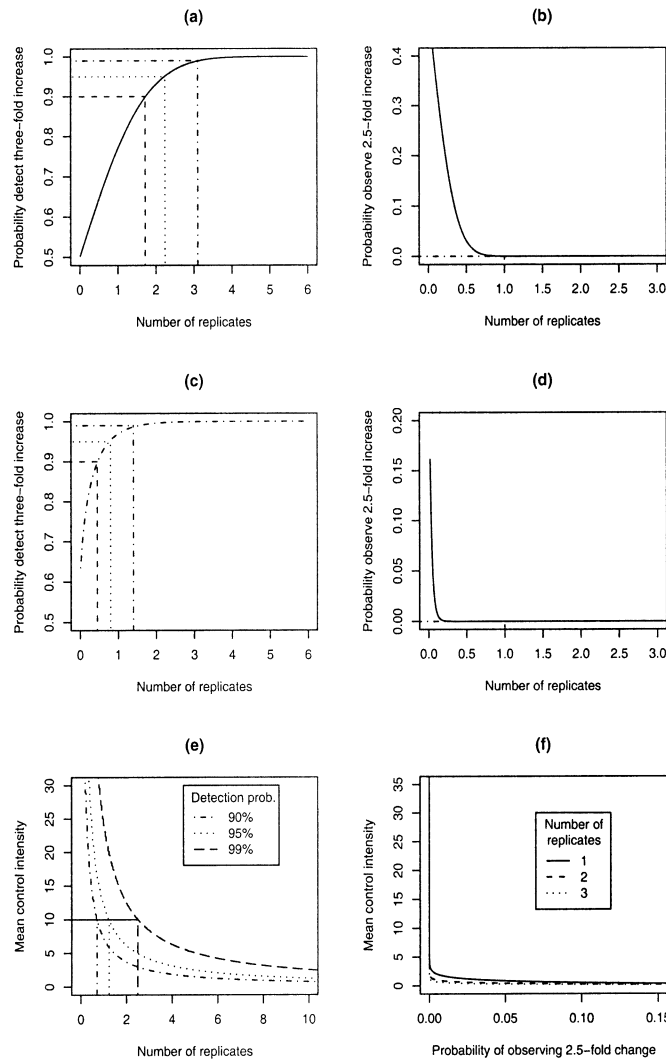
Fig. 3. Yeast sporulation control array data[11]. **(a)** Log-normal model probability of detecting a 3-fold increase in expression level using a 2.5-fold observed mean cutoff for N replicates of a single gene. Lines indicate 90%, 95% and 99% confidence. **(b)** Probability of false positive (observing a mean 2.5-fold increase in intensity when no true difference in expression level exists) as the number of replicates increases under the log-normal model. Dashed lines on graph indicate the probability of false positive when only one replicate is used. **(c)** Probability of detecting a 3-fold increase in expression level using a 2.5-fold observed mean cutoff for N replicates of a single gene under the first gamma model. Lines indicate 90%, 95% and 99% confidence. **(d)** Probability of false positives under the first gamma model. **(e)** Number of replicates required to detect a 3-fold increase in mean control intensity using a 2.5-fold observed mean cutoff, based on the second gamma model. Curves represent 90%, 95% and 99% confidence levels. Straight lines give example of number of replicates needed for gene with mean control intensity of 10.0. **(f)** Probability of false positive for one, two, or three replicates under the second gamma model.