A Bayesian method for analysing spotted microarray data

Colin D. Meiklejohn and Jeffrey P. Townsend

Data received (in revised form): 1st June 2005

Abstract
In the decade since their invention, spotted microarrays have been undergoing technical advances that have increased the utility, scope and precision of their ability to measure gene expression. At the same time, more researchers are taking advantage of the fundamentally quantitative nature of these tools with refined experimental designs and sophisticated statistical analyses. These new approaches utilise the power of microarrays to estimate differences in gene expression levels, rather than just categorising genes as up- or down-regulated, and allow the comparison of expression data across multiple samples. In this review, some of the technical aspects of spotted microarrays that can affect statistical inference are highlighted, and a discussion is provided of how several methods for estimating gene expression level across multiple samples deal with these challenges. The focus is on a Bayesian analysis method, BAGEL, which is easy to implement and produces easily interpreted results.

INTRODUCTION
Both single-channel photolithographic microarrays and two-channel spotted microarrays provide researchers with the ability to measure differences in gene expression for thousands of genes simultaneously. Within the past five years, these tools have proven their utility for studying biological systems in contexts as diverse as identifying regulatory sequence motifs, uncovering the mechanisms of behavioural plasticity, identification of cancer cell types and investigating the molecular underpinnings of animal development. Many applications demonstrate the wide utility of microarray data, but many researchers have also used microarrays to study the fundamental nature of gene expression itself (eg Brem et al., Brem and Kruglyak, Townsend et al., Meiklejohn et al., Ranz et al., Gibson et al. and Wayne et al.). Although the concept that gene expression can adopt binary states (on or off) has been of great heuristic utility to biologists, the number of transcripts of a given messenger RNA in a cell is fundamentally quantitative, and the power of microarrays lies in their ability to measure this parameter for thousands of genes in parallel. Single-channel arrays have been considered quantitative since their inception, whereas historically, two-channel arrays have often been consigned to the identification of lists of qualitatively differentially expressed genes rather than fully utilised for the quantitative estimation of gene expression levels. This history may reflect the fact that spotted arrays usually assess relative gene expression between two or more samples (Figure 1A). However, two-channel microarrays can estimate expression across multiple samples on a single scale, which may be converted to absolute transcript abundance if absolute scaling can be incorporated from another source, such as counts of expressed sequence tags or data from serial analysis of gene expression (SAGE).

Whether on relative or absolute scales, inference of differential expression requires statistical analysis of microarray data. The reliability of a single measurement of the direction of change of gene expression increases with the
Figure 1: (A) Outline of a spotted microarray experiment. (B) Spot saturation can confound estimation of relative expression levels. This conceptual microarray experiment measures the expression level of a single gene across four expression nodes on four microarrays which express the gene over a range of one to ten arbitrary expression units. In all four cases, the ratio of Cy5 fluorescence to Cy3 fluorescence is 2:1, accurately reflecting relative expression between the two nodes being directly compared. However, only in the upper two arrays, where there are sufficient binding sites in the target spot for all of the labelled probe from node D, do the intensities accurately reflect that nodes C and D express this gene at a five-fold higher level than nodes A and B. In the lower two arrays, where the target spot contains at most sufficient binding sites for the labelled probe from nodes A and B, the lack of direct comparison between nodes A/B and C/D obscures this five-fold difference.
magnitudes of differential expression observed, and the choice of a large enough change for inferring these differences (such as a two-fold cut-off) may result in very few false positives (genes that are incorrectly determined to be more abundantly expressed in a particular sample). However, the relationship between statistical power and fold-change varies from gene to gene,\textsuperscript{21–23} and most researchers would probably prefer not to incur the large number of false negatives (genes that are incorrectly determined not to be differentially expressed) that such cut-offs entail. Furthermore, inferred magnitudes of differential expression vary across microarray platforms and other methods of gene expression analysis, making the choice of fold-change cut-off even more arbitrary. Many relevant changes in gene expression may also be more subtle than two-fold.\textsuperscript{13,24–26} Assessment of differential expression should therefore be performed with a statistical methodology that takes the inherent noise associated with microarray measurements into account.

This paper provides a brief overview of several statistical methods for multifactorial studies involving spotted microarrays, ranging from classical to Bayesian. Differences in the conceptual assumptions that go into the methods, and also practical differences in how they deal with artefacts that arise due to the nature of spotted microarray technology are highlighted. The majority of our discussion is on the method with which we have greatest familiarity, BAGEL (Bayesian Analysis of Gene Expression Levels\textsuperscript{19}).

**TWO-CHANNEL SPOTTED MICROARRAY ANALYSIS**

Raw spotted microarray data consist of fluorescence intensity measurements across an array of immobilised probe DNA spots to which target DNA or RNA labelled with two different fluorescent dyes has been competitively hybridised (Figure 1A). Typically, multiple competitive hybridisations are performed between biological samples that are of interest to a researcher, such as different tissues, genotypes, drug treatments or species. Such samples are referred to as ‘expression nodes’ hereafter, as they often appear as nodes in graphic representations of microarray experimental designs.

There are many potential sources of variation that can impinge upon the result of a single microarray experiment. These include uncontrollable differences among experimental organisms, the application of experimental treatments, RNA extraction, cDNA labelling and hybridisation efficiency, to name a few. In order to measure these sources of error, it is necessary to include both biological replicates (for example, repeated application of a drug to independently derived cell populations) and technical replicates (such as repeated cDNA synthesis and fluorescent labelling of a given RNA sample) in a microarray experimental design. For a detailed discussion of technical and biological replication and statistical inference, see Churchill.\textsuperscript{27}

A number of manipulations of raw spotted microarray data are required to implement a statistical analysis to identify differential expression between nodes. The first step is usually to filter poor quality spots from the data, based on criteria such as signal-to-noise ratio, spot size or morphology, or technical errors on the array, as can result from dust or scratches. While stringent spot selection criteria can significantly improve the quality of the data and strengthen the robustness of conclusions, such selection may reduce the number of spots for which there are measurements across all arrays. For statistical methods that require balanced experimental designs, the loss of spots on one or a few arrays may require a separate analysis of these genes, or they may have to be excluded from the analysis altogether.

One significant source of error that must be taken into account arises from the different chemical properties of the most commonly used dyes, Cy5 and Cy3.
These dyes differ in relevant aspects of their chemistry including their half-life, dynamic range, linearity of fluorescence over their range, and even susceptibility to degradation by ozone; these differences can further cause the signal from each dye to vary independently across the physical area of the array. To correct for these dye effects, there are a number of normalisation methods that are commonly used. The simplest of these matches the mean or median signal intensities for the two dyes across the slide, following the premise that under ideal circumstances, the two samples compared started with equal amounts of RNA. More sophisticated methods allow for intensity-dependent and local normalisations, and expression node-by-dye interactions.

Two particular challenges for the analysis of spotted microarray data are (i) correlations between signal intensities coming from the two dyes on a single spot, and (ii) variation from spot to spot in signal intensities within a dye. Both factors are often referred to as ‘spot effects’. Spot effects result from a number of factors that can influence the signal intensity of labelled DNA hybridising to a microarray spot, such as the binding energy or concentration of immobilised probe molecules in the spot, reverse transcription or labelling efficiency of a particular RNA, or degree of sequence similarity between the probe and labelled target. Quantitative inferences about the expression level of a given gene are therefore difficult to disentangle from the quantitative intensity conferred by the particular spot(s) on which expression is measured. The original solution that eliminates factors that are shared by samples hybridising to the same slide is to consider the ratio of the fluorescence intensities from the two dyes, rather than the intensities themselves, thus implicitly eliminating spot effects.

Another issue that can arise in spotted microarray experiments is spot saturation, which occurs when all of the probe DNA molecules available for hybridisation in a spot are bound by labelled target molecules. This is a distinct issue from the more commonly discussed scanner saturation, which occurs when the fluorescence of a spot is outside the dynamic range of the laser scanner used to measure the signal across the array. Given the wide range of absolute transcript abundances across genes expressed in cells (Figure 2), the limited amount of DNA that can be placed in a spot, and the fixed dynamic range of most scanners, it can be difficult to simultaneously measure meagrely expressed genes and avoid saturating spots with labelled DNA for the most abundantly expressed genes.

When considering fluorescence intensities from two samples competitively hybridised to the same slide, spot saturation should not (in theory) confound accurate inference of the relative levels of expression between these two samples. However, if statistical analyses that use ratios as variates are to infer relative expression levels of genes between nodes that are not directly compared in competitive hybridisations, they must include unbroken chains of comparisons between all nodes, or spot saturation can obscure differences in expression (Figure 1B). Such unbroken chains of comparisons are in general desirable for rigorous statistical analyses.

**ANOVA**

Analysis of variance (ANOVA) methods were among the first statistical tools to be adapted for microarrays, and the hierarchical nature of many well-designed microarray experiments lends itself well to such approaches. These approaches fit models in which fluorescence intensity is a linear function of a number of factors, some of which are of interest to the researcher. Common implementations include factors that are assumed to be consistent across arrays, such as variation attributable to specific genes, treatments or dye, and effects that cause fluorescence to vary between hybridisations. These variable (or random) effects include differences between arrays, as well as...
interactions between arrays and genes, dyes and treatments. Using a linear model to account for the effects of arrays, dye, and array-by-dye interactions is one method of normalising data as mentioned above. The gene-by-treatment interaction term measures that which is presumably of interest to the researcher, namely the effect of the treatment (expression node) on the expression of specific genes. These methods are mixed-model approaches, as they assume some normally distributed random effects with mean zero and some fixed effects that are constant across replicates.

In analyses of variance where the variates are normalised fluorescence intensities (rather than ratios) including a gene-by-array interaction term will correct for spot effects when making comparisons between arrays within a single gene. Comparing intensities across genes, however, will still be confounded by spot effects that vary from gene to gene, such as might arise from different concentrations of DNA which are printed onto a slide, unless the correlations between the two channels within a spot are explicitly included in the model. The impact of spot saturation on ANOVA inferences of differential expression depends on whether ratios or normalised single channel intensities are used as variates. Spot saturation presents a challenge for methods that use fluorescence intensities as variates, as the fluorescence intensity for one expression node will depend on the identity of the other expression node against which the first is being competitively hybridised. More specifically, significant amounts of spot saturation should cause such methods to underestimate instances and degrees of differential expression. One solution to this problem would be to include an expression node-by-expression node interaction term; however, such extensions would significantly complicate the analyses.

Appropriate statistical inference using ANOVA methods requires that the data conform to the classical assumptions of analysis of variance, such as that the data be normally distributed and homoscedastic. These requirements may be relaxed if the significance of the gene-by-treatment terms are assessed by testing terms of interest against resampled errors using permutation tests. One further restriction of traditional ANOVAs is the requirement of a balanced experimental design. In some cases it may be possible with further effort to infer statistical significance despite an unbalanced design, but other implementations require that all
spots with missing data be removed from
the analysis entirely.

**LIMMA**
Linear models with Bayesian components

**LIMMA**
Another set of microarray analysis methods combines classical linear models with Bayesian inference for assessing statistical significance. The benefits of such approaches can include fewer distributional assumptions for statistical rigour and more flexibility in dealing with missing data. The Cyber-T program\(^{37}\) uses Bayesian inference to better infer the standard deviation used in a two-sample comparison t-test, but is restricted to comparisons between two expression nodes. The limma program\(^{34}\) extends this concept to an arbitrary number of expression nodes and contrasts between them. Limma fits a linear model that estimates contrasts between nodes hybridised to the same array from ratio data, although it can also model individual channel data as described above for standard ANOVAs.

Like Cyber-T, limma utilises a Bayesian framework in its variance estimation, and derives a prior distribution on the variance from the full set of ratios across all arrays, which avoids the problem of artificially reduced variance estimates for genes that show high reproducibility by chance alone.\(^{38}\) However, this approach may necessarily inflate variance estimates for genes that show an unusually high, but real, level of reproducibility. Both spot effects and spot saturation are appropriately incorporated into limma analyses by using ratios as variates. The Bayesian approach to variance estimation allows one to analyse unbalanced experimental designs with limma, and allows it to deal well with missing data.

**BAGEL**
Bayesian analysis of gene expression levels

**BAGEL**
BAGEL is a statistical framework designed specifically for analysis of spotted microarray data. The statistical model employed by BAGEL assumes that the measured fluorescence intensity for one dye on a given spot is a function of (i) the true quantity of the labelled mRNA hybridising to the spot; (ii) some number of multiplicatively and/or additively confounding factors that are specific to the spot in question and shared by the measured intensity from the other dye (spot effects); and (iii) some number of unbiased, randomly distributed error terms that are independent between the two dyes (such as reverse transcription or DNA labelling efficiency). If the error terms contribute multiplicatively, then this ratio can be approximated by the ratio of two log-normal distributions.\(^{23}\) Note that, as with limma, spot effects and spot saturation are accounted for in BAGEL by using only ratio data. Further, this method has no requirements regarding balanced data, and can accommodate any variability in replication across expression nodes as long as they are connected by an unbroken chain of comparisons, although nodes with greater replication will have estimated gene expression levels with greater confidence (see below).

Given \(n\) expression nodes, these models require the estimation of \(2n - 1\) parameters \((n - 1\) expression levels and \(n\) variances\) for each gene. The number of parameters may be reduced by assuming that, for a given gene, all the nodes have the same error variance. This reduces the number of parameters to \(n\), and consequently reduces the number of replicates required for statistical inference. Alternatively, one may assume that, for a given gene, all samples have a constant relationship with the expression level, ie that they have a common coefficient of variation. This approach also requires estimating \(n\) free parameters. Note that BAGEL, unlike hierarchical but not fully multifactorial Bayesian methods,\(^{39,40}\) makes no explicit distinction between biological replicates and technical replicates. One must therefore ensure that all potential sources of variation (biological and technical) are included for
all expression nodes in order to obtain results that may be generalised.

The likelihood function derived from either of the error models above is explored using a standard Markov chain Monte Carlo (MCMC) approach (for details, see Metropolis et al., Hastings and Gelman et al.). This method starts with a random vector of parameters and then changes one or more of the parameters by a small, random step. At each step the likelihood of the data given the model and the parameter values is calculated. If the new parameters give a better fit to the data, then the new values are accepted. If the new parameters give a worse fit to the data, then the new values are accepted with a probability proportional to the ratio of the likelihood of the data with the new parameter values to the likelihood of the data with the old parameter values. In this way the Markov chain searches the parameter space, finding vectors of relative gene expression levels that produce the greatest likelihood given the model. Samples of parameter values from the chain are used to construct the Bayesian posterior probability of the parameters given the data.

Relative expression levels and statistical significance can be inferred from the parameter values sampled from the Markov chain. The relative expression level of a node for a given gene is the median value across the samples from the chain, normalised by the node with the lowest relative expression level. Ninety-five per cent credible intervals correspond to the values within which 95 per cent of the samples from the chain are bounded, and the P-value for the hypothesis that a given gene’s expression in node A is greater than node B are the proportion of samples from the chain where node A’s expression level was greater than node B’s. These P-values, which may be interpreted as hypothesis tests rather than Bayesian measures of the strength of belief, are not adjusted to compensate in any way for multiple hypothesis tests. The high-throughput nature of microarrays requires that any acceptance or rejection of such hypotheses of differential expression be addressed at some level in the analysis, either by close examination of consistency with other biological data or by a statistical summary such as the false discovery rate (see below).

EXAMPLES USING MICROARRAY DATA

For the purpose of illustration, some analyses using a published set of microarray experiments are discussed below. These experiments examined gene expression in young adult males and females of Drosophila melanogaster and a closely related species, D. simulans, raised under standard laboratory conditions, using cDNA microarrays containing ~6000 expressed sequence tags (ESTs). The purpose of these experiments was to assess the degree of interspecific divergence in gene expression and its relationship to differential expression between the two sexes. This data set has four nodes, and the subset of the experiments used for the analyses here is shown in Figure 3. To facilitate ease of comparison with the results using limma (see below), the raw data were normalised with limma’s print-tip loess method before analysis with either program.

The combinations of additive or multiplicative errors and common or node-specific variances or coefficients of variation produces four error models that BAGEL can use. Figure 4 shows the results of BAGEL analysis of four genes
from the Drosophila data set under these four error models. Each error model was run twice in order to assess variation between independent runs within a model. Changing between additive and multiplicative error models, and constraining variances or CVs make very little difference to the results. For three of the genes, choosing unconstrained variances also has little effect on the results, aside from increasing the width of the 95 per cent credible intervals. However, for the gene BcDNA:LD09936, choosing an unconstrained variance model does have a significant impact on the relative expression levels inferred by BAGEL. This result suggests that only the most highly reproducible and replicated experiments should be analysed with an unconstrained variance/CV model.

In order to directly compare results from a BAGEL analysis to limma, the Drosophila data set was reanalysed using the limma program in the R environment. Figure 5 shows the log-ratio of expression between females of D. melanogaster and D. simulans estimated...
by BAGEL and limma for 6,022 genes. The concordance is nearly perfect except for extremely differentially expressed genes. Table 1 gives the number of genes called differentially expressed at a number of significance thresholds. In all cases, the two methods agree more than they disagree, and genes called differentially expressed by only one method are generally due to the use of a cut-off. This can be seen in that only 29 genes were called differentially expressed at \( P < 0.0001 \) by one method while having a \( P \)-value greater than 0.01 by the other (Table 1). In previously published comparisons between BAGEL and other ANOVA methods, the two approaches have also produced quite similar and compatible results.\(^5\,45\)

The empirical power of an experiment to reveal differences in gene expression level may be summarised by the magnitude of inferred differential expression (or gene expression level, GEL) at which there is a 50 per cent chance of statistical significance (GEL\(_{50}\)), as inferred from a logistic regression of the proportion of significant calls on estimated expression level.\(^23\) As compared with the relatively balanced design of Ranz \emph{et al.},\(^15\) a companion set of hybridisations used to compare gene expression among males of different strains of \textit{D. melanogaster}\(^14\) includes nodes that were assayed on as many as 12 and as few as 3 arrays (Figure 6). Figure 7 shows the logistic regression from comparisons between highly and poorly replicated nodes; more comparisons yield greater power to reveal smaller differences in gene expression.

Because microarray experiments typically assess statistical significance of differential expression for a great many genes at once, an issue arises when one engages in classical hypothesis testing: numerous false positives may arise solely because of the large number of statistical tests performed. Traditional methods for controlling the family-wise false positive rate, such as Bonferroni corrections,\(^35\) tend to produce an extremely high false negative rate, given the large number of tests involved. One response is to report or control the false discovery rate (FDR\(^46\,47\)) which is the proportion of genes with statistically significantly different expression that are expected to be false positives. A number of fairly accurate methods have been developed for controlling the FDR including the Benjamini-Hochberg procedure.\(^48\)

---

**Figure 5:** The inferred log-ratio of expression levels for 6,022 genes between \textit{D. simulans} and \textit{D. melanogaster} females, as estimated by limma and BAGEL. Least-squares regression, \( y = 0.958x + 0.006 \), and \( r^2 = 0.978 \)
simple analytical methods for controlling the FDR have been developed. It can also be estimated through resampling techniques, in which data are shuffled to produce randomly constructed samples, which are then analysed using the chosen statistical framework. The number of significantly differentially expressed genes at a given significance threshold found in the randomised data is an estimate of the false discovery rate for that threshold. Such non-parametric testing has become more practical as computing power has increased, and has the advantage of not relying on assumptions such as normality or log-normality of the data. However, these methods can only identify how many false positives one should expect, not which genes in a list of significantly differentially expressed genes are false positives. This emphasises the importance of confirming any results of interest from a microarray experiment with alternate methods.

There are many considerations that may affect the effectiveness and appropriateness of such randomly permuted samples. For instance, a permuted data matrix should have the same structure (number of expression nodes and comparisons between specific nodes) and proportion of missing data as the true data set, as these parameters will influence the FDR. Different methods of permutation also make different assumptions regarding the error structure in the data. Shuffling appropriately normalised ratios across all comparisons, but within a given gene, produces gene-specific FDR estimates that are insensitive to differences in error variances between genes. Alternatively, the full matrix of ratios across all genes and comparisons can be resampled, which assumes that the error associated with the measurement of gene expression levels is distributed approximately identically and independently across all genes. Finally, ratios taken from self–self hybridisations, where RNA samples from two replicates of the same expression node are labelled with alternate dyes and competitively hybridised to an array, can be resampled to create the permuted data. This last approach has the advantage that deviations from a ratio of one in self–self hybridisation ratios should be due solely to the sources of experimental error that can lead to false positives, whereas many

**Table 1:** Significance calls by method

<table>
<thead>
<tr>
<th>Method</th>
<th>P &lt; 0.05</th>
<th>0.01</th>
<th>0.001</th>
<th>0.0001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both</td>
<td>2,298</td>
<td>1,347</td>
<td>580</td>
<td>235</td>
</tr>
<tr>
<td>limma</td>
<td>222</td>
<td>299</td>
<td>307</td>
<td>220</td>
</tr>
<tr>
<td>BAGEL</td>
<td>342</td>
<td>247</td>
<td>149</td>
<td>80</td>
</tr>
<tr>
<td>Neither</td>
<td>3,180</td>
<td>4,149</td>
<td>3,006</td>
<td>5,507</td>
</tr>
</tbody>
</table>

The number of genes assigned to non-contiguous significance categories by the two methods. In each entry, the first number indicates those genes where the more significant call came from limma and the less significant from BAGEL, and vice versa for the second number. No attempt was made to control the number of false positives in either case.

**Figure 6:** Experimental design for microarray analysis of gene expression in adult males of eight strains of D. melanogaster. For each hybridisation the arrowhead indicates the sample labelled with Cy5.
or most of the deviations from one in heterologous hybridisations reflect true differences in gene expression as well as experimental error. However, if self–self hybridisations are used as the source of ratios for inferring false positive rates, it is important that each sample to be hybridised be exposed to all potential sources of experimental error, ie independent RNA extraction, cDNA synthesis, labelling, etc. In practice, resampling self–self and heterologous hybridisation ratios have resulted in quite similar FDR estimates (Table 2).14,15

CONCLUSIONS
While the relative simplicity of the experimental designs of most spotted microarray experiments should make statistical analysis of the data straightforward, there are a number of technical challenges associated with this technology that can cause difficulties if not properly addressed. Some of those issues have been raised here, such as spot saturation, and the problem of missing data due to stringent spot selection criteria. These considerations, as well as others, should guide researchers in choosing a statistical method with which to analyse their data. Methods that can incorporate ratios as variates will remove the need to correct for spot effects in the analysis, which can be beneficial for all but the most precisely manufactured arrays. However, the comparisons presented here and elsewhere5,45 indicate that, for well-replicated array experiments, the results of different analysis methods are largely in agreement.

At the present time, BAGEL provides node-specific gene expression estimates. Any higher-level comparisons must therefore be made post hoc using individual relative expression levels and credible intervals. However, composite comparisons of gene expression level

Table 2: Inferred false discovery rate

<table>
<thead>
<tr>
<th></th>
<th>CS</th>
<th>OR</th>
<th>HikR</th>
<th>StL</th>
<th>Z53</th>
<th>Z30</th>
<th>Z29</th>
<th>Z2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>25</td>
<td>30</td>
<td>26</td>
<td>15</td>
<td>16</td>
<td>40</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td>15</td>
<td>21</td>
<td>27</td>
<td>24</td>
<td>18</td>
<td>27</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>HikR</td>
<td>31</td>
<td>21</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>19</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>StL</td>
<td>19</td>
<td>21</td>
<td>19</td>
<td>38</td>
<td>30</td>
<td>16</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Z53</td>
<td>20</td>
<td>19</td>
<td>21</td>
<td>34</td>
<td>22</td>
<td>29</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Z30</td>
<td>16</td>
<td>10</td>
<td>25</td>
<td>22</td>
<td>12</td>
<td>21</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>Z29</td>
<td>22</td>
<td>15</td>
<td>26</td>
<td>19</td>
<td>15</td>
<td>19</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Z2</td>
<td>33</td>
<td>25</td>
<td>23</td>
<td>21</td>
<td>22</td>
<td>20</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>

The number of pairwise false positives between eight expression nodes in the experiments of Meiklejohn et al.14 at a significance cut-off (P-value) of 0.01. The numbers above the diagonal were obtained by resampling the full matrix of normalised ratios; those below the diagonal were obtained by resampling normalised ratios from self–self hybridisations. The number of genes significantly differentially expressed at this cut-off ranged from 218 (StL v. HikR) to 928 (Z53 v. StL).
among nodes or sets of nodes could be tested by sampling appropriately from the MCMC chain output. Among freely available microarray analysis packages, the most widely used methods have been written for the R statistical environment (such as maanova and limma). For the microarray practitioner with little statistical experience, implementing these programs can prove challenging. Of the methods discussed here, BAGEL is the easiest to implement and obtain intuitive, immediately interpretable results. Such ease of analysis can enable biologists to move quickly beyond the microarray and on to the next set of experiments suggested by the results.

Acknowledgments
We would like to thank the Miller Institute for Basic Research in Science and the NIH (grant GM068087) for supporting this research.

References


