Bayesian Robust Inference for Differential Gene Expression in cDNA Microarrays with Multiple Samples

Raphael Gottardo*, Adrian E. Raftery*, Ka Yee Yeung** and Roger E. Bumgarner**
*Department of Statistics, University of Washington,
Box 354322 Seattle, WA 98195-4322 (E-mail: *raph@stat.washington.edu*)
**Department of Microbiology, University of Washington,
Box 358070, Seattle, WA 98195.

Technical Report no. 455
Department of Statistics
University of Washington

July 5, 2004

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
05 JUL 2004		**00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Bayesian Robust Inference for Differential Gene Expression in cDNA Microarrays with Multiple Samples	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Washington,Department of Statistics,Box 354322,Seattle,WA,98195-4322**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We consider the problem of identifying differentially expressed genes under different conditions using cDNA microarrays. Standard statistical methods cannot be used because typically there are thousands of genes and few replicates. Because of the many steps involved in the experimental process, from hybridization to image analysis, cDNA microarray data often contain outliers. For example, an outlying data value could occur because of scratches or dust on the surface, imperfections in the glass, or imperfections in the array production. We develop a robust Bayesian hierarchical model for testing for differential expression. Outliers are modeled explicitly using a t-distribution. The model includes an exchangeable prior for the variances which allow different variances for the genes but still shrink extreme empirical variances. Our model can be used for testing for differentially expressed genes among multiple samples, and can distinguish between the different possible patterns of differential expression when there are three or more samples. Parameter estimation is carried out using a novel version of Markov Chain Monte Carlo that is appropriate when the model puts mass on subspaces of the full parameter space. The method is illustrated using two publicly available gene expression data sets. We compare our method to five other commonly used techniques, namely the one-sample t-test, the Bonferroni-adjusted t-test, Significance Analysis of Microarrays (SAM), and EBarrays in both its Lognormal-Normal and Gamma-Gamma forms. In an experiment with HIV data, our method performed better than these alternatives, on the basis of between-replicate agreement and disagreement.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**25**	

Abstract

We consider the problem of identifying differentially expressed genes under different conditions using cDNA microarrays. Standard statistical methods cannot be used because typically there are thousands of genes and few replicates. Because of the many steps involved in the experimental process, from hybridization to image analysis, cDNA microarray data often contain outliers. For example, an outlying data value could occur because of scratches or dust on the surface, imperfections in the glass, or imperfections in the array production. We develop a robust Bayesian hierarchical model for testing for differential expression. Outliers are modeled explicitly using a t-distribution. The model includes an exchangeable prior for the variances which allow different variances for the genes but still shrink extreme empirical variances. Our model can be used for testing for differentially expressed genes among multiple samples, and can distinguish between the different possible patterns of differential expression when there are three or more samples. Parameter estimation is carried out using a novel version of Markov Chain Monte Carlo that is appropriate when the model puts mass on subspaces of the full parameter space. The method is illustrated using two publicly available gene expression data sets. We compare our method to five other commonly used techniques, namely the one-sample t-test, the Bonferroni-adjusted t-test, Significance Analysis of Microarrays (SAM), and EBarrays in both its Lognormal-Normal and Gamma-Gamma forms. In an experiment with HIV data, our method performed better than these alternatives, on the basis of between-replicate agreement and disagreement.

KEY WORDS: Bayesian hierarchical model; Bonferroni adjustment; Empirical Bayes; Gene expression; Heteroscedasticity; Markov chain Monte Carlo; Mixture distribution; Outlier; Singular distribution; t distribution.

Contents

List of Tables

List of Figures

1 Introduction

cDNA microarrays (Schena et al. 1995) consist of thousands of individual DNA sequences printed on a high density array on a glass microscope slide using a robotic arrayer. A microarray works by exploiting the ability of a given labeled cDNA molecule to bind specifically to, or hybridize to, a complemetary sequence on the array. By using an array containing many DNA samples, scientists can measure—in a single experiment—the expression levels of hundreds or thousands of genes within a cell by measuring the amount of labeled cDNA bound to each site on the array. In a typical two-color microarray experiment, two messenger RNA (mRNA) samples, from control and treatment situations, are compared for gene expression. Treatment is taken in a broad sense to mean any condition different from the control. Both mRNA samples are reverse-transcribed into cDNA, labeled using different fluorescent dyes (red and green dyes), and then mixed and hybridized with the arrayed DNA sequences. The hybridized arrays are then imaged to measure the red and green intensities for each spot on the glass slide (Yang et al. 2002). The estimates of the red and green intensities are the starting point of any statistical analysis such as testing for differential expression, discriminant analysis, and clustering (Yeung et al. 2001). Similarly, microarrays can be used to compare the mRNA levels of thousands of genes under several experimental or biological conditions by using several samples. One of the main research areas is to detect genes that are differentially expressed across the different conditions.

In recent years, there has been a considerable amount of work on the detection of differentially expressed genes. Perhaps the first statistical treatment of differential expression with microarrays can be found in Chen, Dougherty, and Bittner (1997). A common approach is to test a hypothesis for each gene and then try to correct for multiple testing. Most of the statistics used are variants of t or F statistics. This was done by Dudoit et al. (2002) using Welsh's t-statistic with p-values estimated by permutations. Tusher, Tibshirani, and Chu (2001) used a modification of the t statistic where the denominator was modified by adding a constant to improve the estimate of the standard deviation. This idea is similar to a regularized t-test (Baldi and Long 2001), where the estimate of the standard deviation in the t-test was regularized by adding a global estimate of the standard deviation computed using an empirical Bayes approach.

In each of these situations, two types of error can occur: a false positive (Type I error) or a false negative (Type II error). When many hypothesis are tested at the same time, the chance of committing a type I error increases. One approach to overcoming this problem is

to try to control the total number of type I errors or false positives. This can be done using multiple testing procedures to control some measure of the overall type I error. The most two common measures in the area of microarrays are the familywise error rate (FWER), which is the probability of making at least one type I error, and the false discovery rate (FDR), which is the proportion of false positives among the total number of discoveries reported. One of the most used FWER-based adjustments is that of Bonferroni. Tusher et al. (2001) used a permutation technique to estimate and control the FDR. Dudoit, Shaffer, and Boldrick (2003) review multiple adjustment procedures in the context of microarrays.

Newton et al. (2001) introduced an empirical Bayes approach to detect changes in gene expression on a single two-channel cDNA slide using a hierarchical gamma-gamma model, and Kendziorski et al. (2003) generalized their work to multiple sample response experiments. Ibrahim, Chen, and Gray (2002) and Tadesse, Ibrahim, and Mutter (2003) have introduced more fully Bayesian approaches where "exact" estimation is carried out by Markov chain Monte Carlo.

In this paper we introduce a Bayesian hierarchical model to test for differentially expressed genes in a robust way. The model is a generalization of a model used by Gottardo et al. (2003) to estimate the true mean intensities from replicates. Robustness is achieved by using a hierarchical-t formulation (Besag and Higdon 1999), which is more robust to outliers than the usual Gaussian model. The model includes an exchangeable prior for the variances, allowing each gene to have a different variance while still achieving some shrinkage. We elaborate this model by introducing a prior that allow us to detect differentially expressed genes in multiple sample experiments, where the number of samples can be greater than two. The prior is written as a mixture of singular Gaussian distributions. We show how one can use Markov Chain Monte Carlo to estimate the parameters, even though the model contains a component that is a mixture of singular distributions. Inference is based on the posterior probabilities of differential expression calculated from our model. We refer to our method as BRIDGE (Bayesian Robust Inference of Differential Gene Expression).

The paper is organized as follows. Section 2 introduces the data structure and the notation. In Section 3 we present the Bayesian hierarchical model, and in Section 4 we show how it is used to test for differential expression. Section 4 also reviews five other commonly used methods to test for differentially expressed genes in the two-sample case. In Section 5, we apply the methods to experimental data and compare the results. In Section 6, we show how one can extend our model to multiple sample experiments and use it to detect differentially expressed genes in a three sample experiment. Finally, in Section 7 we discuss

our results and possible extensions.

2 Data

We used two data sets that are fairly typical of data in this area. In this paper, we use the word sample to describe different experimental or biological conditions.

The HIV data: This data set, described by van't Wout et al. (2003), consists of four experiments using the same RNA preparation on 4 different slides. The expression levels of 7680 cellular RNA transcripts were assessed in CD4-T-cell lines at time $t = 24$ hour after infection with HIV virus type 1. This dataset contains two samples, one that correspond to the HIV infected cells and one for the non infected cells. The dataset contains 12 HIV-1 genes used as positive controls. This dataset is the result of a balanced dye-swap experiment. Two of the four (technical) replicates were hybridized with the green dye (Cy3) for the control and the red dye (Cy5) for the treatment; then the dyes were reversed on the other two replicates. The values in the Cy3 and Cy5 channels were extracted from each image using customized software written at the University of Washington (Spot-On Image, developed by R.E. Bumgarner and Erick Hammersmark). The image analysis provided two numbers for each gene in each replicate: a Green spot intensity and a Red spot intensity (these are background adjusted estimates).

After the image analysis, the data take the form

y_{iscr}, $i = 1, \ldots, I$; $s = 1, 2$; $c = 1, 2$; $r = 1, \ldots, R$,

where y_{iscr} are the estimated intensities of gene i in sample s with color c from replicate r. Note that we use different indices for the color and the sample to allow for dye-swap experiments.

The BRCA data: Hedenfalk et al. (2001) conducted a study to examine breast cancer tissues from patients carrying mutations in the predisposing genes, BRCA1 or BRCA2, or from patients not expected to carry a hereditary mutation. Hedenfalk et al. (2001) examined 22 breast cancer tumor samples: 7 tumors with BRCA1, 8 tumors with BRCA2 and 7 sporadic tumors, i.e. with neither mutation. In this data, "samples" refer to tissue sample types and there is no color swap. A set of 3226 genes was pre-selected by Hedenfalk et al. (2001) by filtering the raw images. The data take the form $\log_2(y_{isr}/\text{ref}_{ir})$, $i = 1, \ldots, I$; $s = 1, 2, 3$; $r = 1, \ldots, R_s$, where y_{isr} is the intensity from gene i of the r-th (biological) replicate in sample s, and ref_{ir} is the intensity from a common reference sample. Note that here,

Hedenfalk et al. (2001) used a reference sample because there are three samples of interest: BRCA1, BRCA2 and sporadic. The log-ratios were normalized so that the mean across genes in each replicate of each sample is zero.

3 Differential Expression with Two Samples

In this section we introduce the Bayesian hierarchical model used to test for differentially expressed genes in the two-sample case. We use a Bayesian linear model (Lindley and Smith 1972) with t-distributed sampling errors to allow for outliers (Besag and Higdon 1999). We also explicitly model the non-constant variances by using an exchangeable prior for the gene precisions (Lewin et al. 2003). Our model includes design effects that deal with normalization issues (Kerr, Martin, and Churchill 2000). The model is an extension of a previous model used by Gottardo et al. (2003) to estimate the intensities from replicates. In this section we extend the model in order to detect differentially expressed genes with two samples.

3.1 The Model

We model $y^*_{iscr} = g_\kappa(y_{iscr}) \equiv \log_2(y_{iscr} + \kappa)$ where κ is a positive additive constant. The parameter κ is estimated beforehand and is treated as fixed in the estimation of the full model, as described in Section 3.3. The model is as follows:

$$\begin{aligned} y^*_{iscr} = g_\kappa(y_{iscr}) &= \mu + \alpha_s + \beta_c + \eta_r + \gamma_{is} + \delta_{sc} + \frac{\epsilon_{iscr}}{\sqrt{w_{icr}}}, \\ (\epsilon_{i1cr}, \epsilon_{i2cr})'|\mathbf{V}_i &\sim \mathrm{N}_2(\mathbf{0}, \mathbf{V}_i), \\ (w_{icr}|\nu_r) &\sim \mathcal{G}a(\nu_r/2, \nu_r/2), \end{aligned} \tag{1}$$

where w_{icr} and $(\epsilon_{i1cr}, \epsilon_{i2cr})'$ are independent. Since the w's are independent of the ϵ's, we have $\frac{\epsilon_{iscr}}{\sqrt{w_{icr}}} \sim \mathcal{T}_{(\nu_r, \mathbf{0}, \mathbf{V}_i)}$, i.e. the (bivariate) errors have a bivariate t distribution with ν_r degrees of freedom and covariance matrix $\mathbf{V}_i$. The advantage of writing the model this way is that, conditionally on the w_{icr}, the sampling errors are again normal, but with different precisions. The interpretation is also easier, as, conditionally on the w_{icr}, estimation becomes a weighted least squares problem.

In (1), μ is the baseline intensity. The sample effect α_s is used to remove the bias between the two samples. The dye effect is represented by β_c, and accounts for the well-known fact that the green dye (Cy3) tends to be brighter than the red dye (Cy5). The interaction of the sample s with the sample c is denoted by δ_{sc}, and is present because the different dyes tend

to have different biases in different samples. The array effect of replicate r, η_r, is intended to normalize the overall intensity of each array across replicates. This parameter is needed because differences between replicates in overall intensity are frequent in microarray data. Finally, γ_{is}, the effect of gene i in sample s, is the quantity of interest. We model it as a random effect with a mixture of two singular Gaussian distributions, the first one of which is singular with respect to the two dimensional Lebesgue measure:

$$(\boldsymbol{\gamma_i}|\boldsymbol{\lambda_\gamma}, p) \sim (1-p)\mathrm{N}(\gamma_{i1}; 0, \lambda_{\gamma_{12}})\mathbf{1}_{[\gamma_{i1}=\gamma_{i2}]} + p\mathrm{N}(\gamma_{i1}; 0, \lambda_{\gamma_1})\mathrm{N}(\gamma_{i2}; 0, \lambda_{\gamma_2})\mathbf{1}_{[\gamma_{i1}\neq\gamma_{i2}]}, \qquad (2)$$

where $\boldsymbol{\gamma_i} = (\gamma_{i1}, \gamma_{i2})'$ and $\boldsymbol{\lambda_\gamma} = (\lambda_{\gamma_1}, \lambda_{\gamma_2}, \lambda_{\gamma_{12}})$. The first component corresponds to the genes that are not differentially expressed ($\gamma_{i1} = \gamma_{i2}$) whereas the second component corresponds to the genes that are differentially expressed ($\gamma_{i1} \neq \gamma_{i2}$). Note that the formulation is not standard as it is not absolutely continuous with respect to the two-dimensional Lebesgue measure. However it defines a proper distribution with respect to a more general dominating measure, namely the sum of a one dimensional Lebesgue measure on the line $\gamma_{i1} = \gamma_{i2}$ and the two-dimensional Lebesgue measure (Gottardo and Raftery 2004).

For a given gene, the correlation matrix, $\mathbf{V}_i$, allows the measurements from the two samples to be correlated, and also allows each gene to have its own variance. The precision matrix (i.e. the inverse of the covariance matrix) is given by

$$\begin{aligned}(\mathbf{V}_i^{-1}|\rho, \lambda_{\epsilon_{i1}}, \lambda_{\epsilon_{i2}}) &= \frac{1}{(1-\rho^2)}\begin{pmatrix} \lambda_{\epsilon_{i1}} & -\sqrt{\lambda_{\epsilon_{i1}}\lambda_{\epsilon_{i2}}}\rho \\ -\sqrt{\lambda_{\epsilon_{i1}}\lambda_{\epsilon_{i2}}}\rho & \lambda_{\epsilon_{i2}} \end{pmatrix}, \\ (\lambda_{\epsilon_{is}}|a_\epsilon, b_\epsilon) &\sim \mathcal{G}a(a_\epsilon^2/b_\epsilon, a_\epsilon/b_\epsilon),\end{aligned}$$

where ρ is the correlation between samples, $\lambda_{\epsilon_{is}}$ is the precision of gene i in sample s, and $\mathcal{G}a(a_\epsilon^2/b_\epsilon, a_\epsilon/b_\epsilon)$ denotes a Gamma distribution with shape parameter a_ϵ^2/b_ϵ and scale parameter a_ϵ/b_ϵ, so that it has mean a_ϵ and variance b_ϵ. We use an exchangeable prior for the precisions, so that information is shared between the genes. This allows shrinkage of very small and very large empirical within-gene between-replicate variances.

3.2 Priors

We use a vague but proper prior for the parameters $\lambda_{\gamma_{12}}, \lambda_{\gamma_1}, \lambda_{\gamma_2}$ of the distributions of the gene effects γ_{is} in (2). This is exponential with mean 200, so that $\lambda_\gamma \sim \mathcal{G}a(1, 0.005)$. The fixed effects μ, α_s, β_c, η_r and δ_{sc} in (1) also have vague but proper priors, normal with a large variance, namely $\mathrm{N}(0, 25)$.

For identifiability, we impose the constraints $\alpha_1 = 0$, $\beta_1 = 0$, $\delta_{11} = \delta_{12} = \delta_{21} = 0$, $\eta_1 = 0$ and $\eta_R = 0$. We also need two constraints on the γ_{is}, such as $\sum_i \gamma_{is} = 0$ for $s = 1, 2$. However, instead of including these constraints as part of the model definition, we let the $\boldsymbol{\gamma}$'s be "free" during the sampling process, and identify the parameters afterwards from the sampled values. These constraints were used in Gottardo et al. (2003). Here, we need to impose an additional constraint to be able to identify the non-differentially expressed genes. If we did not impose any additional constraint, the two components of the $\boldsymbol{\gamma}$ would be unidentifiable, since one could add a constant to all of the $\boldsymbol{\gamma}_2$'s and subtract the same constant to α_2 without changing the overall mean. To avoid this problem, we fixed the sample effect α_2 to the least squares estimate.

We also use vague but proper priors for the error precisions, specified by $a_\epsilon \sim \mathcal{U}_{[0,1000]}$ and $b_\epsilon \sim \mathcal{U}_{[0,1000]}$. The prior for the correlation between the two samples is given by $\rho \sim \mathcal{U}_{[-1,1]}$. The prior for the mixing parameter p is uniform over $[0, 1]$. The prior for the degrees of freedom ν_r is uniform on the set $\{1, 2, \ldots, 10, 20, \ldots, 100\}$.

3.3 Parameter Estimation

Realizations were generated from the posterior distribution via Markov chain Monte Carlo (MCMC) algorithms (Gelfand and Smith 1990). All updates are rather straightforward except for $\boldsymbol{\gamma}$ since the distribution is formed by two singular components. However, there exists a common dominating σ-finite measure and so the Metropolis-Hastings algorithm can be used (Gottardo and Raftery 2004). If the errors were taken to be independent, i.e. $\rho = 0$, the full conditional of $\boldsymbol{\gamma}$ would be given by

$$(\boldsymbol{\gamma}_i | \ldots) \propto (1-p) c_i \mathrm{N}(\gamma_{i1}; \mu_i^*, \lambda_i^{*-1}) \mathbf{1}_{[\gamma_{i1}=\gamma_{i2}]} + p c_{i1} c_{i2} \mathrm{N}(\gamma_{i1}; \mu_{i1}^*, \lambda_{i1}^{*-1}) \mathrm{N}(\gamma_{i2}; \mu_{i2}^*, \lambda_{i2}^{*-1}) \mathbf{1}_{[\gamma_{i1}\neq\gamma_{i2}]},$$

where

$$\lambda_i^* = \sum_{r,s} w_{icr} \lambda_{\epsilon_{is}} + \lambda_{\gamma_{12}}, \quad \mu_i^* = \lambda_i^{*-1} \sum_{r,s} w_{icr} \lambda_{\epsilon_{is}} r_{iscr},$$

and

$$\lambda_{is}^* = \lambda_{\epsilon_{is}} \sum_{r} w_{icr} + \lambda_{\gamma_s}, \quad \mu_{is}^* = \lambda_{is}^{*-1} \sum_{r} w_{icr} \lambda_{\epsilon_{is}} r_{iscr}.$$

The constants c_i, c_{i1} and c_{i2} are given by

$$c_i = \sqrt{\frac{\lambda_{\gamma_{12}}}{\lambda_i^*}} \exp\left\{ -0.5 \sum_{r,s} w_{icr} \lambda_{\epsilon_{is}} r_{iscr}^2 + 0.5 \lambda_i^{*-1} \left(\sum_{r,s} w_{icr} \lambda_{\epsilon_{is}} r_{iscr} \right)^2 \right\},$$

and

$$c_{is} = \sqrt{\frac{\lambda_{\gamma_s}}{\lambda^*_{i1}}} \exp\left\{ -0.5\lambda_{\epsilon_{is}} \sum_r w_{icr} r^2_{iscr} + 0.5\lambda^{*\,-1}_{is} \left(\sum_r w_{icr}\lambda_{\epsilon_{is}} r_{iscr} \right)^2 \right\},$$

where the residuals r_{iscr} are defined by $r_{iscr} = y_{iscr} - \mu + \alpha_s + \beta_c + \eta_r + \delta_{sc}$. To update $\boldsymbol{\gamma}$, one draws new pairs $(\gamma_{i1}, \gamma_{i2})$ from the null component of the full conditional with probability $1 - p^* \equiv (1-p)c/[pc_1c_2 + (1-p)c]$ or from the other component with probability p^*. We refer the reader to Gottardo and Raftery (2004) for more details. Then when $\rho \neq 0$, we used the full conditional given above as proposal and correct the acceptance probability with a Hastings correction factor. Since the number of replicates is usually small, this gives a high acceptance rate, about 90% in the example presented here.

We estimated the shift κ in advance by fitting (1) with $w_{icr} \equiv 1$, $\lambda_{\epsilon_{is}} \equiv \lambda_{\epsilon_s}$ and $p = 1$ via MCMC, and treating κ as a parameter with vague uniform prior $\kappa \sim \mathcal{U}_{[0,10000]}$. We then estimated κ by its posterior mean; see Gottardo et al. (2003) for more details.

4 Methods to be Compared

In this section, we describe six methods for detecting differentially expressed genes with cDNA microarrays. These six methods will be compared in section 5. Our method (BRIDGE) deals with normalization whereas the other methods do not. As recommended in the literature (Chu et al. 2002), SAM and the one-sample t-test will be applied to the mean zero normalized log-ratios $x_{icr} - \bar{x}_{\cdot cr}$, where $x_{icr} = \log_2(y_{i1cr}/y_{i2cr})$. Similarly, EBarrays (Kendziorski et al. 2003) will be applied to the normalized (bivariate) measurements $(y_{i1cr}/\bar{y}_{\cdot 1cr}, y_{i2cr}/\bar{y}_{\cdot 2cr})'$.

Bayesian Robust Inference for Differential Gene Expression (BRIDGE): From model (1) we can compute the marginal posterior probability of differential expression of gene i, namely $\varrho_i = \Pr(\gamma_{i1} \neq \gamma_{i2}|\mathbf{y})$. From the MCMC output, one can estimate the posterior probabilities of differential expression for all the genes. For each gene i, the marginal posterior probability of differential expression, ϱ_i, corresponds to the posterior probability that $\gamma_{i1} \neq \gamma_{i2}$ given the data. For a given posterior sample S of size B we estimate the posterior probabilities by

$$\hat{\varrho}_i = \tfrac{1}{B}\#\{k \in S : \gamma^{(k)}_{i1} \neq \gamma^{(k)}_{i2}\},$$

where $\gamma^{(k)}_{i1}$ and $\gamma^{(k)}_{i2}$ are the values generated at the k-th MCMC iteration. We declare a gene to be differentially expressed if its estimated posterior probability, $\hat{\varrho}_i$, is greater than 0.5.

Raw and Bonferroni-adjusted one-sample t-tests: A classical procedure for testing null hypothesis about the mean of a distribution is the t-test. Here we apply one-sample t-tests to the zero-mean normalized log-ratios. Because of the large number of hypotheses tested we also report the results of adjusting the p-values using the Bonferroni adjustment. The Bonferroni adjustment method controls the familywise error rate (FWER), which is the probability of yielding one or more false positives. If P is the raw p-value, then the Bonferroni-adjusted p value is $\min\{IP, 1\}$. We declare a gene to be differentially expressed if its (raw or adjusted) p-value is less than 0.05.

Significance Analysis of Microarrays (SAM): This is a statistical technique for finding significant genes in a set of microarray experiments, proposed by Tusher, Tibshirani, and Chu (2001). In the one-sample case SAM is similar to a regularized t-test where the estimate of the standard deviation is regularized with a common estimate of the standard deviation. SAM controls an estimate of the FDR, which is the proportion of falsely identified genes among the genes declared to be differentially expressed.

The method for one-sample testing (one class response in the SAM software) is as follows.

1. For each gene i, compute a statistic $d_i = \frac{\bar{x}_i}{s_i + s_0}$ where $\bar{x}_i$ is the sample mean of the log-ratios of gene i, s_i is the sample standard deviation of the log-ratios of gene i and s_0 is a fudge factor common to all the genes.

2. Fix a rejection region based on the statistics d_i and estimate the number of falsely called genes by permutations. Permutations are obtained by randomly permuting the signs of the log-ratios.

3. The FDR is estimated by the median of the number of falsely called genes divided by the number of genes called differentially expressed.

4. Repeat steps 3-4 for a large number of rejection regions

Based on the estimated FDR values, the user chooses a rejection region. See Chu et al. (2002) for further details. Using SAM we select the largest rejection region with estimated FDR less than 0.1. In the literature FDR values between 5% and 10% are commonly used (Tusher, Tibshirani, and Chu 2001; Efron and Tibshirani 2002; Dudoit, Shaffer, and Boldrick 2003). In our experiments, the results were not too sensitive to the choice of the FDR value and we decided to use 10%.

Empirical Bayes (EBarrays) Lognormal-Normal and Gamma-Gamma models: Newton et al. (2001) and Kendziorski et al. (2003) developed methods for testing for differential gene expression using hierarchical lognormal-normal (LNN) and gamma-gamma (GG) models. Newton et al. (2001) developed a method for detecting changes in gene expression on a single two-channel cDNA slide using a hierarchical gamma-gamma model, and Kendziorski et al. (2003) extended this to replicated chips. Here we report results obtained using the hierarchical modeling framework of Kendziorski et al. (2003). For a given gene, Kendziorski et al. (2003) assume that the data arise from a mixture distribution of the form $f(\mathbf{y}_i) = (1-p)f_0(\mathbf{y}_i)+pf_1(\mathbf{y}_i)$, where p is the expected proportion of differentially expressed genes, f_0 is the joint probability density function under the null hypothesis of no expression, f_1 is the joint probability density function under the alternative hypothesis of differential expression, and $\mathbf{y}_i$ is the bivariate vector of measurements for gene i. Given the mean expression in each sample (γ_1 and γ_2), they assume that the measurements arise independently and identically from an observation component $f_{obs}(.|\gamma_1, \gamma_2)$ where γ_1 and γ_2 are the mean expression levels in each sample. If the gene is not differentially expressed, $\gamma \equiv \gamma_1 = \gamma_2$ is the overall mean expression level. They assume that the $\boldsymbol{\gamma}$ arise from some genome wide distribution $\pi(\boldsymbol{\gamma})$, which represents fluctuations in mean expression levels among genes. Then the marginal densities become $f_0(\mathbf{y}_i) = \int \prod_r f_{obs}(\mathbf{y}_{ir}|\gamma)\pi(\gamma)d\gamma$, and $f_1(\mathbf{y}_i) = \iint \prod_r f_{obs}(\mathbf{y}_{ir}|\gamma_1, \gamma_2)\pi(\gamma_1, \gamma_2)d\gamma_1 d\gamma_2$, where the product is over replicates. They summarize the inference about differential expression by the posterior probabilities of differential expression,

$$\Pr(z_i = 1|\mathbf{y}_i) = \frac{pf_1(\mathbf{y}_i)}{(1-p)f_0(\mathbf{y}_i) + pf_1(\mathbf{y}_i)},$$

where z_i is an indicator variable equal to 1 if the gene i is differentially expressed, and to 0 otherwise.

Kendziorski et al. (2003) used two distributional forms: a gamma-gamma (GG) model and a lognormal-normal (LNN) model. For the GG model, the observation component is a gamma distribution with mean value γ and scale α/γ. The coefficient of variation α is taken to be constant across genes. Then α/γ is taken to be inverse gamma with shape parameter α_0 and ν. For the lognormal-normal (LNN) model the observation component is a lognormal distribution with mean value γ and variance σ^2 taken to be the same for all the genes. The conjugate prior for γ is normal with mean γ_0 and variance τ_0^2. For both models, the prior can be integrated out and the EM algorithm can be used to estimate the unknown parameters. A gene is declared to be differentially expressed if its posterior probability is greater than 0.5.

5 Results

We fitted the model described in Section 3.1 to the HIV data. The posterior modes of the degrees of freedom of the t-distribution, ν_r, ranged from 4 to 100, indicating that the sampling errors can be heavier-tailed than the Gaussian distribution and that the proportion of outliers varies from array to array. There is substantial between-sample correlation, estimated as 0.63, even after removing design effects and gene effects. Our model also captures the non-constant variance with posterior means 8.3 and 16.5 for a_ϵ and b_ϵ respectively.

The proportion of differentially expressed genes is estimated to be 0.02. Figure 1 is a plot of the posterior probabilities against the posterior means of the log-ratios computed from our model. A relatively small number of genes seem to be differentially expressed. Most of the log-ratios are shrunk close to zero and have very low posterior probabilities of differential expression. Most of the BRIDGE posterior probabilities are smaller than 0.2 (Figure 1).

In contrast, the unadjusted t-test yields many p-values less than 0.05. Figure 2 shows both the raw and the adjusted p-values. More than 700 genes have raw p-values less than 0.05 whereas only 2 have adjusted p-values less than 0.05. The adjustment is clearly too conservative since we know for sure from external information that at least 12 genes are differentially expressed. Table 1 summarizes the number of genes called differentially expressed by each method. BRIDGE, SAM and EBarrays report roughly the same number of differentially expressed genes, all around 100. All the methods except the t-test with adjusted p-values correctly detect the 12 positive control genes, i.e. those known to be differentially expressed.

In order to evaluate the performance of each method, we divided the four replicates of the HIV24 data into two groups of two replicates. We applied each method to each group of replicates and looked at the agreement and disagreement between the genes declared to be differentially expressed. Table 1 contains the number of genes declared to be differentially expressed when two replicates are used. All the methods except the Bonferroni-adjusted t-test detect the 12 positive controls. Overall, the number of genes declared to be differentially expressed is much smaller when two replicates are used than when four replicates are used, for all the methods except EBarrays. This may be because EBarrays assumes a constant variance, whereas there seems to be evidence that the actual variances do differ between genes. When the true variance of a gene is high, the empirical mean across replicates may also be high by chance alone, leading to the gene to be (incorrectly) declared to be differentially expressed by EBarrays. This is more likely to happen when there are two replicates than

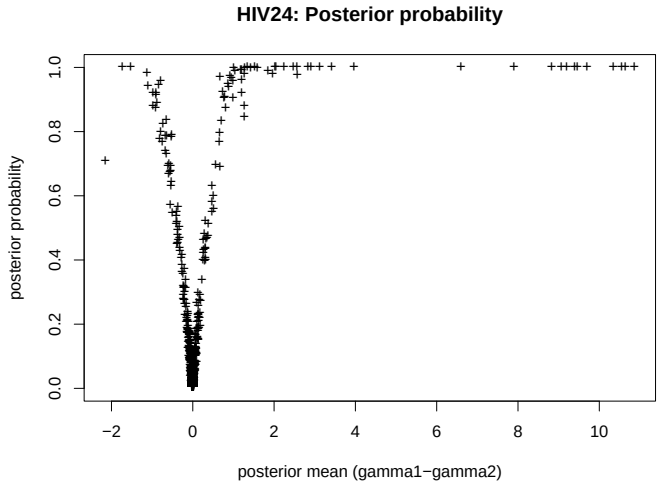

Figure 1: Posterior probabilities from the BRIDGE method plotted against the posterior differences between γ_1 and γ_2 (estimated log-ratios) for the HIV24 data. Most of the log-ratios are shrunk close to zero and have very low posterior probabilities of differential expression.

when there are four, because the empirical means of replicates for non-differentially expressed genes are more variable with fewer replicates.

Table 1 shows the number of agreements and disagreements between the two groups of replicates, for each of the six methods we consider. The raw and adjusted t-tests both perform poorly: the raw t-test records a very high level of disagreement, presumably corresponding to a large number of false positives, and the adjusted t-test is clearly too conservative. Among the other methods, the number of genes on which there was agreement are comparable, but EBarrays recorded much larger numbers of genes on which there was disagreement than BRIDGE or SAM. Comparing BRIDGE with SAM, BRIDGE recorded agreement on more genes than SAM (21 compared to 17), with disagreement on only half as many genes (8 compared to 16). It thus seems that for this particular dataset, using the criterion of agreement between replicates, BRIDGE did better than the other methods at identifying differentially expressed genes.

6 The Multiple Sample Case

Even though the model introduced in Section 3.1 is intended to test for differential expression between two samples, it can be extended to situations where there are more than two samples,

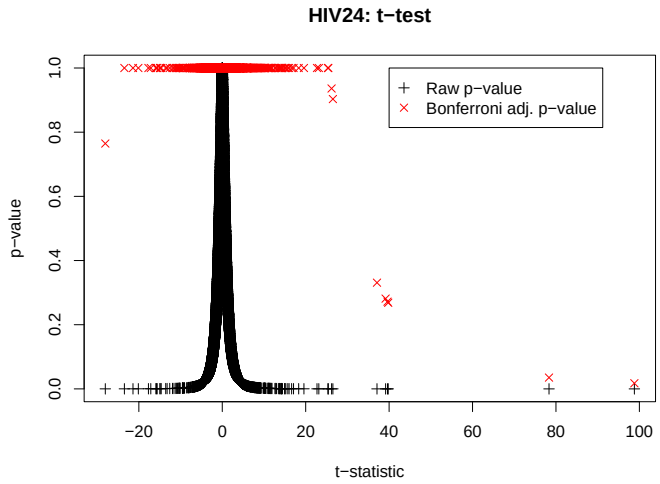

Figure 2: Raw and Bonferroni-adjusted p-values plotted against the t-statistics for the HIV24 data. The adjusted p-values are too conservative, with only two p-values less than 0.05.

Table 1: Number of genes declared to be differentially expressed by each method for the HIV data using 4 and 2 replicates.

			EBarrays		t-test	
	BRIDGE	SAM	GG	LNN	Raw	Adj
Replicates 1–4	102	125	83	75	748	2
Replicates 1&3	28	31	69	114	301	0
Replicates 2&4	22	19	53	79	262	0

Table 2: Agreement and disagreement on the differential expression of genes in the HIV data when the 4 replicates are divided into two sets of two. For each method, "Agreement" denotes the number of genes declared to be differentially expressed based on both sets of two replicates, while "Disagreement" refers to the number of genes that were declared to be differentially expressed based on one set of two replicates, and not to be differentially expressed based on the other set of two replicates.

			EBarrays		t-test	
	BRIDGE	SAM	GG	LNN	Raw	Adj
Agreement	21	17	22	30	12	0
Disagreement	8	16	78	135	539	0

and differences in expression of the same gene between any two samples may be of interest. Here the null hypothesis is not as simply defined as before, because there are many possible patterns of differential gene expression. In order to account for all possible patterns, the prior for the γ's in (2) needs to be modified. Because most microarray technologies allow the direct comparison of at most two samples, the general model (1) needs to be modified for changes in the experimental design. In this section we show how our model can be modified to the case where we have three samples. The generalization to more than three samples should be straightforward. We use the Hedenfalk et al. (2001) experiment as an example.

6.1 The model

This time we model the log-ratios $y^*_{isr} = \log_2(y_{isr}/\text{ref}_{ir})$, where ref_{ir} is a common reference sample as defined in Section 2. We assume that the measurement errors for different samples are independent. The log-ratios are normalized and mean centered, so we do not have any design effects, and the model becomes:

$$\begin{aligned} y^*_{isr} &= \gamma_{is} + \frac{\epsilon_{isr}}{\sqrt{w_{isr}}}, \\ (\epsilon_{isr}|\lambda_{\epsilon_{is}}) &\sim \text{N}(0, \lambda^{-1}_{\epsilon_{is}}), \\ (w_{isr}|\nu_r) &\sim \mathcal{G}a(\nu_r/2, \nu_r/2). \end{aligned} \tag{3}$$

We need to modify the distribution of $\boldsymbol{\gamma}$, the vector of gene effects in each sample, to allow each possible pattern of differential expression to have positive probability. We still model it as a random effect, but this time with a mixture of five singular Gaussian distributions, of which four are singular with respect to the three dimensional Lebesgue measure, as follows:

$$\begin{aligned} (\boldsymbol{\gamma}|\boldsymbol{\lambda_\gamma}, \boldsymbol{p}) \sim\ & p_1 \text{N}(\gamma_1; 0, \lambda_{\gamma_1}) \mathbf{1}_{[\gamma_1=\gamma_2=\gamma_3]} \\ +\ & p_2 \text{N}(\gamma_1; 0, \lambda_{\gamma_1}) \text{N}(\gamma_2; 0, \lambda_{\gamma_{23}}) \mathbf{1}_{[\gamma_1\neq\gamma_2=\gamma_3]} \\ +\ & p_3 \text{N}(\gamma_2; 0, \lambda_{\gamma_2}) \text{N}(\gamma_1; 0, \lambda_{\gamma_{13}}) \mathbf{1}_{[\gamma_1=\gamma_3\neq\gamma_2]} \\ +\ & p_4 \text{N}(\gamma_3; 0, \lambda_{\gamma_3}) \text{N}(\gamma_1; 0, \lambda_{\gamma_{12}}) \mathbf{1}_{[\gamma_1=\gamma_2\neq\gamma_3]} \\ +\ & p_5 \text{N}(\gamma_1; 0, \lambda_{\gamma_1}) \text{N}(\gamma_2; 0, \lambda_{\gamma_2}) \text{N}(\gamma_3; 0, \lambda_{\gamma_3}) \mathbf{1}_{[\gamma_1\neq\gamma_2\neq\gamma_3]}, \end{aligned}$$

where $\boldsymbol{\lambda_\gamma} = (\lambda_{\gamma_1}, \lambda_{\gamma_2}, \lambda_{\gamma_3}, \lambda_{\gamma_{12}}, \lambda_{\gamma_{13}}, \lambda_{\gamma_{23}}, \lambda_{\gamma_{123}})$ is the vector of precisions and $\boldsymbol{p}$ is the vector of probabilities for the five patterns, constrained to sum to one. The five components correspond to all five possible patterns of expression. As before, the formulation is not standard since it is not absolutely continuous with respect to the three-dimensional Lebesgue measure, but

it does define a proper distribution with respect to a more general dominating measure (Gottardo and Raftery 2004).

6.2 Priors

We keep the priors described in Section 3.2. All the λ_γ's have exponential prior distributions with mean 200. The mixing probabilities have a prior distribution that is uniform over the simplex $S = \{\mathbf{p} : \sum_i p_i = 1\}$, i.e. the prior is $\mathcal{D}(1, 1, 1, 1, 1)$, Dirichlet with common hyperparameters 1.

This time the parameter estimation is easier. Since there are no design effects, there are no identifiability issues. Moreover, the log-ratios being independent, all parameters can be updated by Gibbs sampling except a_ϵ and b_ϵ. Again the only difficulty is in the update of $\boldsymbol{\gamma}$ because of the singularity between the different components. This time, due to the independence of the errors, the full conditional is explicitly available and so Gibbs sampling can be used as described in Section 3.3.

6.3 Results

We fitted the model given by (3) to the BRCA data. The posterior modes of the degrees of freedom of the t-distribution, ν_r, ranged as low as 4, indicating that the sampling errors can be more heavy-tailed than the Gaussian distribution. The model captures the non-constant variance: the posterior mean of a_ϵ, the mean of the gene-specific variances, was 9.8, and the posterior mean of b_ϵ, the variance of the gene-specific variances, was 29.0. The posterior mixing probabilities, $p_1, \ldots, p_5$, of the five patterns of differential expression, are summarized in Table 3 by their posterior means. The mixing probabilities clearly favor the null pattern of no differential expression, suggesting that most of the genes are not differentially expressed. They also suggest that there is more difference between the BRCA1 and the BRCA2 tumors than any BRCA with sporadic tumors. This confirms results from Hedenfalk et al. (2001).

In this section, we compare our method only with EBarrays because, like our method, it can handle multiple samples by considering all possible patterns of expression. We do not describe the methodology as it is very similar to the two-sample case described in Section 4; see Kendziorski et al. (2003) for further details. This time we used only the LNN model, because if the intensity measurements arise from a lognormal distribution, then so should the ratio. However this is not the case for a gamma distribution.

For each gene, we wish to compare the five different patterns of differential expression:

$$\begin{aligned}
\mathcal{P}_1 &: \gamma_1 = \gamma_2 = \gamma_3, \\
\mathcal{P}_2 &: \gamma_1 \neq \gamma_2 = \gamma_3, \\
\mathcal{P}_3 &: \gamma_1 = \gamma_3 \neq \gamma_2, \\
\mathcal{P}_4 &: \gamma_1 = \gamma_2 \neq \gamma_3, \\
\mathcal{P}_5 &: \gamma_1 \neq \gamma_2 \neq \gamma_3.
\end{aligned}$$

For each gene and pattern, we can compute the posterior probability that the gene conforms to that pattern. We computed the BRIDGE posterior probabilities from our model and the posterior probabilities using EBarrays. The difference between the BRIDGE posterior probabilities and the posterior probabilities computed using EBarrays is quite large. For comparison purposes, we followed Kendziorski et al. (2003) in classifying a gene as conforming to a pattern if the corresponding posterior probability was greater than 0.5. The number of genes classified into each pattern is given by Table 4.

As in the two-sample problem, EBarrays tends to declare more genes to be differentially expressed. This may be due to two main differences between the two methods. First, EBarrays models within-gene variances as constant, while BRIDGE allows for them to vary between genes. If within-gene variances do indeed vary between genes, then a method such as EBarrays that assumes constant variance may be more likely to (incorrectly) declare genes with high variances to be differentially expressed. Second, in its lognormal-normal version, EBarrays assumes measurement errors to be normally distributed, while BRIDGE allows their distribution to have heavier tails. If the tails are indeed heavier, for example if there are outliers, then a method such as EBarrays that assumes normality may be more likely to (incorrectly) declare genes with outlying measurements in some replicates to be differentially expressed.

To illustrate the latter point, consider Figure 3, which shows the expression levels for the three genes classified into pattern $\mathcal{P}_5$ by EBarrays. The posterior probabilities of being in pattern $\mathcal{P}_5$ reported by EBarrays are $0.99, 0.75$ and 0.96. The same three genes are not classified into the $\mathcal{P}_5$ pattern by our method but into patterns $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_2$, respectively, with BRIDGE posterior probabilities $0.59, 0.92$ and 0.86. The first gene in Figure 3 seems to contain several outliers, e.g. measurements 9 and 10. The posterior mean of the weights, $\mathbf{w}$, from our model shows that the same outliers have been downweighted. As a result, the gene is not declared to be differentially expressed in any way by our method. Similarly, the sporadic tumor samples of the two other genes (samples 16-22) contain several outliers

that are downweighted by our model. The last two tumor types (sporadic and BRCA2) no longer seem differentially expressed and our method classifies the two genes into pattern $\mathcal{P}_2$. Visual inspection suggests that the choices made by EBarrays were quite influenced by a few outliers, which our method downweights.

Table 3: Estimates of the mixing probabilities for the five patterns of expression on the BRCA data from model (3) and from EBarrays.

	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$
BRIDGE	0.88	0.073	0.03	0.016	0.001
EBarrays (LNN)	0.79	0.067	0.11	0.019	0.00025

Table 4: Number of genes declared to conform to each pattern of differential expression. A gene was classified into a pattern if the corresponding posterior probability of its conforming to that pattern was greater than 0.5.

	$\mathcal{P}_1$	$\mathcal{P}_2$	$\mathcal{P}_3$	$\mathcal{P}_4$	$\mathcal{P}_5$
BRIDGE	3048	107	41	5	0
EBarrays (LNN)	2721	147	272	36	3

We compared the genes declared to be differentially expressed by each method with two previous studies (Hedenfalk et al. 2001; Lee et al. 2003). In the original analysis, Hedenfalk et al. (2001) selected a list of 53 genes using F-tests. Lee et al. (2003) used a gene selection approach to select a group of 28 genes that best discriminate between BRCA1, BRCA2 and sporadic tumors. Table 5 summarizes the agreement of EBarrays and BRIDGE with the two previous studies. BRIDGE found far fewer genes to be differentially expressed than did EBarrays — 178 as against 505 — but BRIDGE nevertheless detected almost as many of the genes identified by the previous studies as did EBarrays. We did not expect our 178 genes to include all of the previously selected genes by Hedenfalk et al. (2001) and Lee et al. (2003) as they used quite different methods. For example, Hedenfalk et al. (2001) used F-tests to select the 53 genes. It is known that the F-test can be sensitive to outliers, and so it is possible that some of the differential expression identifications of Hedenfalk et al. (2001) may be due to outliers, as may also be the case with EBarrays.

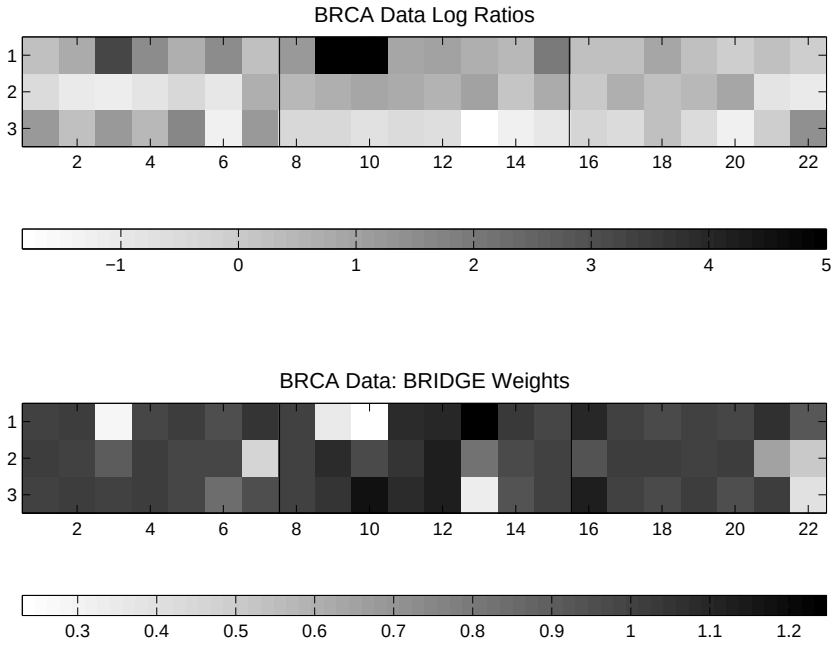

Figure 3: The three genes with the greatest EBarrays posterior probabilities of conforming to pattern $\mathcal{P}_5$ (all three samples different) based the LNN model (top). The corresponding posterior weights from our model (bottom). The first 7 samples (1-7) correspond to BRCA1 tumors, the next 8 samples (8-15) correspond to BRCA2 tumors and the last 7 samples (16-22) correspond to sporadic tumors. Several possibly outlying samples were heavily downweighted by our model. The two dark outliers, measurements 9-10 from gene 1, were truncated to 5 for clarity. The actual data values are 6.64 and 7.13, respectively.

Table 5: Agreements among the genes declared to be differentially by two previous studies and those detected by BRIDGE and EBarrays. For BRIDGE and EBarrays, a gene is called differentially expressed if its posterior probability of conforming to the null pattern is less than 0.5. The total number of genes selected by each method is shown in parentheses.

	Hedenfalk et. al. (53)	Lee et. al. (28)
EBarrays (505)	45	21
BRIDGE (178)	37	20

7 Discussion

We have developed a Bayesian hierarchical model for testing cDNA microarray intensities in a way that is robust to outlying measurements but still powerful even with a small number of replicates. Our Bayesian hierarchical model is based on a model used by Gottardo et al. (2003) to estimate microarray intensities in a robust fashion. We modified the model by using a novel form of prior that allows us to detect differentially expressed genes in multiple sample experiments. When there are three or more samples, the model allows us to detect the differentially expressed genes, and also to classify them into the different patterns of differential expression. In an experiment with HIV two-sample data, we compared our method with five other commonly used methods, and it performed better, at least in terms of agreement and disagreement between groups of replicates.

In order to compare our model with other methods, we identified a gene as differentially expressed if the posterior probability of this was greater than 0.5. In practice, though, we would not use such a cutoff, but instead would report the posterior probabilities themselves. The posterior probabilities from our model are well calibrated and easy to interpret.

In this paper, we have compared our model with five alternatives, but there are many other methods for detecting differentially expressed genes with gene expression data. We chose these five because they are widely used and representative of other methods. For example, there are several other empirial Bayes methods that we could have used. These includes the lognormal-normal models of Lönnstedt and Speed (2002) and Gottardo et al. (2003) and the less parametric approaches of Efron et al. (2001) and Newton et al. (2004). More comparisons between statistical tests can be found in Cui and Churchill (2003). Among explicit adjustments for multiple testing, we considered only the Bonferroni adjustment and the FDR control given by SAM; these are widely used and easy to understand. But other adjustments for multiple comparisons have been proposed recently, mainly controlling the FWER or the FDR (Storey and Tibshirani 2001; Dudoit et al. 2002). We refer the reader to Dudoit, Shaffer, and Boldrick (2003) for more comparisons of multiple testing procedures.

Finally, even though our model was illustrated with two color microarray data, it could also be easily applied to two-color Agilent data or one-color Affymetrix data.

For example, one could replace the log-ratio used in model (3) by the (log) gene expression estimates.

8 Acknowledgements

The authors thank Nema Dean for careful reading of the manuscript and Angelique van't Wout for providing us with some of the data. This research was supported by NIH Grant 8 R01 EB002137-02, and Raftery's research was also partially supported by ONR Grant N00014-01-10745. Yeung's research was funded by NIH-NCI grant 1K25CA106988-01 and Bumgarner's was funded by NIH-NIAID grants 5P01 AI052106-02 and 1U54AI057141-01, NIH-NIEHA grant 1U19ES011387-02, NIH-NHLBI grants 5R01HL072370-02 and 1P50HL073996-01, and NIH-NCRR grant 1S10RR019423-01.

References

Baldi, P. and A. Long (2001). A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes. *Bioinformatics 17*, 509–519.

Besag, J. E. and D. M. Higdon (1999). Bayesian analysis of agricultural field experiments (with discussion). *Journal of the Royal Statistical Society, Series B 61*, 691–746.

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

Chu, G., B. Narasimham, R. Tibshirani, and V. Tusher (2002). *SAM "Significance Analysis of Microarrays", Users guide and technical document*. Stanford University.

Cui, X. and G. Churchill (2003). Statistical tests for differential expression in cDNA microarray experiments. *Genome Biology 4*, 210.

Dudoit, S., J. Shaffer, and J. Boldrick (2003). Multiple hypothesis testing in microarray experiments. *Statistical Science 18*, 71–103.

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

Efron, B. and R. Tibshirani (2002). Empirical Bayes methods and false discovery rates for microarrays. *Genetic Epidemiology 23*, 70–86.

Efron, B., R. Tibshirani, J. D. Storey, and V. Tusher (2001). Empirical Bayes analysis of a microarray experiment. *Journal of the American Statistical Association 96*, 1151–1160.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Gottardo, R., J. A. Pannucci, C. R. Kuske, and T. Brettin (2003). Statistical analysis of microarray data: A Bayesian approach. *Biostatistics 4*, 597–620.

Gottardo, R. and A. Raftery (2004). Markov chain Monte Carlo computations with mixture of singular distributions. Technical report, Statistics Department, University of Washington.

Gottardo, R., A. E. Raftery, K. Y. Yeung, and R. Bumgarner (2003). Robust estimation of cDNA microarray intensities. Technical Report 438, Statistics Department, University of Washington, Seattle.

Hedenfalk, I., D. Duggan, Y. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, O. Kallioniemi, B. Wilfond, A. Borg, and J. Trent (2001). Gene-expression profiles in hereditary breast cancer. *The New England Journal of Medicine 344*, 539–548.

Ibrahim, J., M. H. Chen, and R. Gray (2002). Bayesian models for gene expression with DNA microarray data. *Journal of the American Statistical Association 97*, 88–99.

Kendziorski, C., M. Newton, H. Lan, and M. N. Gould (2003). On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Statistics in Medicine 22*, 3899–3914.

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

Lee, K., N. Sha, E. Dougherty, M. Vannucci, and B. Mallick (2003). Gene selection: a Bayesian variable selection approach. *Bioinformatics 19*, 90–97.

Lewin, A., S. Richardson, C. Marshall, A. Glazier, and T. Aitman (2003). Bayesian modelling of differential gene expression. Technical report, Imperial College, London.

Lindley, D. V. and A. F. M. Smith (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B 34*, 1–41.

Lönnstedt, I. and T. P. Speed (2002). Replicated microarray data. *Statistica Sinica 12*, 31–46.

Newton, M., A. Noueiry, D. Sarkar, and P. Ahlquist (2004). Detecting differential gene

expression with a semiparametric hierarchical mixture method. *Biostatistics 5*, 155–176.

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

Schena, M., D. Shalon, R. W. Davis, and P. Brown (1995). Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. *Science 270*, 467–470.

Storey, J. and R. Tibshirani (2001). Estimating false discovery rates under dependence, with applications to DNA microarrays. Technical Report 2001-28, Department of Statistics, Stanford University.

Tadesse, M., J. Ibrahim, and G. Mutter (2003). Identification of differentially expressed genes in high-density oligonucleotide arrays accounting for the quantification limits of the technology. *Biometrics 59*, 542–554.

Tusher, V., R. Tibshirani, and G. Chu (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences 98*, 5116–5121.

van't Wout, A. B., G. K. Lehrma, S. A. Mikheeva, G. C. O'Keeffe, M. G. Katze, R. E. Bumgarner, G. K. Geiss, and J. I. Mullins (2003). Cellular gene expression upon human immunodeficiency virus type 1 infection of $CD4^+ - T -$ Cell lines. *Journal of Virology 77*, 1392–1402.

Yang, Y. H., M. J. Buckley, S. Dudoit, and T. P. Speed (2002). Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics 11*, 108–136.

Yeung, K. Y., C. Fraley, A. Murua, A. E. Raftery, and W. L. Ruzzo (2001). Model-based clustering and data transformations for gene expression data. *Bioinformatics 17*, 977–987.