BMC
Bioinformatics

METHODOLOGY ARTICLE   **Open Access**

Separate-channel analysis of two-channel microarrays: recovering inter-spot information

Gordon K Smyth[1,2] and Naomi S Altman[3*]

Abstract

Background: Two-channel (or two-color) microarrays are cost-effective platforms for comparative analysis of gene expression. They are traditionally analysed in terms of the log-ratios (M-values) of the two channel intensities at each spot, but this analysis does not use all the information available in the separate channel observations. Mixed models have been proposed to analyse intensities from the two channels as separate observations, but such models can be complex to use and the gain in efficiency over the log-ratio analysis is difficult to quantify. Mixed models yield test statistics for the null distributions can be specified only approximately, and some approaches do not borrow strength between genes.

Results: This article reformulates the mixed model to clarify the relationship with the traditional log-ratio analysis, to facilitate information borrowing between genes, and to obtain an exact distributional theory for the resulting test statistics. The mixed model is transformed to operate on the M-values and A-values (average log-expression for each spot) instead of on the log-expression values. The log-ratio analysis is shown to ignore information contained in the A-values. The relative efficiency of the log-ratio analysis is shown to depend on the size of the intraspot correlation. A new separate channel analysis method is proposed that assumes a constant intra-spot correlation coefficient across all genes. This approach permits the mixed model to be transformed into an ordinary linear model, allowing the data analysis to use a well-understood empirical Bayes analysis pipeline for linear modeling of microarray data. This yields statistically powerful test statistics that have an exact distributional theory. The log-ratio, mixed model and common correlation methods are compared using three case studies. The results show that separate channel analyses that borrow strength between genes are more powerful than log-ratio analyses. The common correlation analysis is the most powerful of all.

Conclusions: The common correlation method proposed in this article for separate-channel analysis of two-channel microarray data is no more difficult to apply in practice than the traditional log-ratio analysis. It provides an intuitive and powerful means to conduct analyses and make comparisons that might otherwise not be possible.

Keywords: Loop design, Unconnected design, Reference design, Intraclass correlation, False discovery rate, Power, Efficiency

Background

Microarrays have been the most popular technology for genome-wide profiling of gene expression for the past 15 years. The earliest microarrays used two channels, with two RNA samples separately labeled and competitively hybridized to the same array, as a means of controlling inter-probe variability [1]. Despite the rise of one channel

microarrays and other expression profiling technologies, two-channel arrays, also known as two-color arrays, continue to be a cost-effective platform for assessing relative gene expression. The use of two channels is more efficient than one channel for many comparative experiments [2,3]. Microarrays constructed from EST libraries may also be most effective when used with two channels. For species with few genomic resources a common strategy for differential expression studies is to use EST libraries or high throughput sequencing methods to obtain partial transcriptome information and then use the resulting

*Correspondence: nsa1@psu.edu
[3]Department of Statistics, The Pennsylvania State University, University Park, PA 16802–2111, USA
Full list of author information is available at the end of the article

transcripts to develop microarray probes. For example, [4] developed a Nimblegen microarray to investigate disease resistance in apple and [5] used a custom microarray to assess differential gene expression in diseases of a marine flatfish. Bay LK, 2009 [6] used a custom spotted cDNA array to assess differential expression between populations of a reef-building coral.

The traditional and most common approach to the analysis of two-channel gene expression microarrays is to summarize the intensity values in terms of the log-ratios of the two channel intensities for each probe on each array [7]. Most papers on the statistical analysis of two-channel microarrays have also taken this general approach [8,9]. Expression levels measured by neighboring spots on the same array have been shown to be highly correlated [10]. Observations of the two channels from the same physical spot are expected to be even more highly correlated. The practice of analyzing log-ratios in effect takes advantage of this correlation, as the variance of the log-ratio should be smaller than the sum of the variances of two individual positively correlated log-intensity values.

It has been argued however that log-ratio analyses are not fully efficient in that they do not use all the information available in the data [11-13]. A number of papers have popularized the idea of analyzing the individual channel intensities as separate observations [14-17]. One approach is to an estimate a random effect for each microarray spot to account for the correlation between the two channels [16,17]. This approach can be implemented by fitting a mixed model, a linear model with both fixed and random effects, to the expression data for each gene [16,17]. Separate channel analysis has been used to analyze experiments for which two channels were available for some arrays and only one channel for others [18]. Separate-channel approaches to the normalization of two-channel microarray data have also been discussed [19].

Apart from other efficiency gains, separate channel analysis gives the possibility of comparing treatment conditions that are not connected in a two-color experimental design [15]. A two-channel microarray design is said to be connected if every pairwise comparison among treatments can be expressed as a difference of log-ratios. For example, reference designs and loop designs are connected [14]. Unconnected designs contain islands of arrays with treatments that are unlinked by hybridization to the same arrays. For example, if an experiment includes 4 treatments B, C, D and E with treatments B and C always hybridized together and treatments D and E always hybridized together, there is no way to compare treatments B and D using the log-ratio approach. Separate channel models are necessary to analyze such unconnected designs.

A number of papers have shown that the two channel intensities from each spot are usefully represented in terms of the log-ratio (M-value) and the average log-intensity over the two channels (A-value) for each spot [20,21]. This article reinterprets the M and A-values in terms of within and between spot contrasts. The usefulness of this partition is shown to arise in good part from the fact that the M and A-values for a given spot are statistically independent even though the individual log-intensities are highly correlated.

This article goes on to reformulate the mixed model approach in terms of the M and A-values. This approach not only presents an efficient algorithm for estimating the mixed model but also elucidates the difference between the traditional log-ratio based approach and the analysis of separate-channels. Use of the M-values alone for the analysis is shown to discard the between-spot information. The separate-channel approach amounts to recovering information from the between-spot error stratum, i.e., from comparisons among the A-values. The efficiency gains of separate-channel analysis are quantified in terms of the intra-spot correlation.

The idea of regularized statistical methods that borrow strength between genomic features is an important recurring theme in genomic data analysis. Statistical research has shown that, across a wide range of multi-parameter problems, improvements in parameter estimation can be made by combining information across the datasets used to estimate the individual parameters [22,23]. This principle is especially important for microarray data analysis with tens of thousands of probe-wise variances to be estimated, and the idea has led to a number of popular regularized microarray methods [24-26]. Regularization of variance components has also been shown to be beneficial for mixed model analyses of microarray data [17]. The article introduces a simple but effective regularization for the spotwise random effects by forcing the intra-spot correlation to be a constant value across all genes in the study. The genewise variances are subsequently regularized using a conjugate empirical Bayes procedure [26]. The treatment of the intra-spot correlation as a global parameter leads to a number of important advantages. The global estimator is very precise, so the intra-spot correlation can be treated as a known parameter at the individual gene level. The permits the mixed model to be transformed into a form suitable for entry into the well-established empirical Bayes analysis pipeline of the `limma` package [26]. Unlike previous mixed model approaches, this approach leads to test statistics with exact parametric distributions under the null hypothesis, even for experiments with small numbers of samples [26].

The method proposed in this article for separate-channel analysis is no more difficult to apply in practice than the traditional log-ratio analysis and provides an intuitive and powerful means to make comparisons that might otherwise not be possible. The separate channel

tests of differential expression are shown to be more statistically powerful than those from the log-ratio approach, leading to reductions in both false discovery and false nondiscovery rates. This article includes several cases studies, which demonstrate the advantages of the separate channel approach and the performance of the proposed regularization approach for the intra-spot correlation and the genewise variances.

Results and discussion

Intra-spot correlation

Suppose that a gene expression experiment has been conducted in which $2n$ RNA samples have been hybridized to n two-channel microarrays, each printed with the same set of P probes. In general, there may be more than one spot on each microarray for the same probe DNA sequence [27], but we will suppose here that all spots are to be treated as if they correspond to independent probes. Note we use "probe" here to refer to the cDNA sequence used in the array design and "spot" to refer to the physical feature printed onto each array. Two-color microarrays yield two intensities values for each spot, one for each channel. Following usual practice, we will call the shorter wave-length channel "green" and the other "red". Most image analysis softwares yield a foreground and a background intensity for each channel for each spot. We will assume that the foreground intensities have been background corrected, normalized and log-transformed to yield $\log_2$ intensities. Write y_{gi1} and y_{gi2} for the green and red channel log-intensities, respectively, for probe g on array i. We will assume that there are no missing values, that is, a finite y_{gic} is available for all gic. In particular, all the background-corrected intensities are assumed to be positive so that the log-intensities are properly defined, as can be ensured by a variety of model-based background correction strategies [28]. The log-intensities are assumed to be normally distributed,

$$y_{gic} \sim N(\mu_{gic}, \sigma_g^2) \tag{1}$$

where μ_{gic} is an unknown mean that depends on the probe and the treatment conditions applied to channel c on array i. The unknown σ_g^2 is probe-specific but common across arrays for each probe.

We can view each physical spot on each array as a block giving rise to two observations, one for each channel. While it is reasonable to treat log-intensities observed on different arrays as independent, two observations from the same physical feature of the same array must almost inevitably be highly correlated. Hence we assume y_{gic} and $y_{g'i'c'}$ are independent if $i \neq i'$ but that

$$\mathrm{corr}(y_{gi1}, y_{gi2}) = \rho_g$$

where ρ_g is the unknown intra-spot correlation. We expect ρ_g to be positive to reflect the fact that the two

channel observations share any characteristics local to that spot on that array [19].

The separate channel linear model

Probewise linear models provide a flexible and powerful approach to microarray data analysis [12,16]. We assume that the true mean log-expression values μ_{gic} can be represented by a linear model

$$\mu_{gic} = x_{ic}^T \beta_g$$

where the design vector x_{ic} indicates which treatment conditions are applied to the RNA sample hybridized to channel c of array i, and β_g is an unknown vector of coefficients representing population mean log-intensities or log-fold-changes between the treatment conditions. Often x_{ic} is a vector of zeros and ones. In matrix notation, the linear model may be written

$$E(y_g) = X\beta_g$$

where y_g is the $2n$-vector of y_{gic} values for probe g and X is the design matrix with rows x_{ic}.

The covariance matrix $\mathrm{Var}(y_g)$ is block diagonal, with diagonal elements equal to σ_g^2 and off-diagonal elements either zero or equal to $\rho_g\sigma_g^2$ for pairs of observations from the same spot. If an estimate of ρ_g is available, then the linear model coefficient vector β_g can be estimated from the data by generalized least squares,

$$\hat{\beta}_g = \left(X^T C_g^{-1} X\right)^{-1} X^T C_g^{-1} y_g$$

where C_g is the correlation matrix obtained by dividing the variances out of $\mathrm{Var}(y_g)$.

Estimating the intra-spot correlation by mixed models

One approach to the separate channel linear model is to treat it as a mixed model with each spot a randomized block of size two [12,16]. The mixed linear model can be represented as

$$y_{gic} = \mu_{gic} + b_{gi} + \epsilon_{gic}$$

where b_{gi} is the spot effect for probe g on array i and ϵ_{gic} is a residual error term. The spot and residual effects are both random, with $b_{gi} \sim N(0, \sigma_{bg}^2)$ and $\epsilon_{gic} \sim N(0, \sigma_{eg}^2)$. Here σ_{bg}^2 is the variance component for spot-level variation or, equivalently, for array to array variation. The marginal variance for y_{gic} is

$$\sigma_g^2 = \sigma_{gb}^2 + \sigma_{ge}^2$$

and the intra-spot correlation is

$$\rho_g = \frac{\sigma_{gb}^2}{\sigma_{gb}^2 + \sigma_{ge}^2}.$$

The mixed linear model obviously assumes that $\rho_g \geq 0$, because σ_{gb}^2 cannot be negative. A typical mixed model

analysis will obtain REML estimators [29] for σ_{bg}^2 and σ_{eg}^2, leading to an estimate for ρ_g as above, and hence to a generalized least squares estimator for β_g. Obviously the mixed model constrains the intra-spot correlation to be positive.

Fitting mixed models of this form has been used as the basis for two-channel microarray data analysis, although the test statistics have complex distributions that cannot be derived exactly [16,17]. In this article, we will reformulate the mixed model in order to obtain an exact distributional theory for empirical Bayes test statistics, and in order to clarify the relationship between the mixed model and the traditional log-ratio analysis of microarray data.

Spot-wise means and differences

Following [20,21], write $M_{gi} = y_{gi2} - y_{gi1}$ for the log-ratio of red to green intensity for spot gi and $A_{gi} = (y_{gi2} + y_{gi1})/2$ for the average log-intensity for the spot. These quantities were originally proposed for their use in normalization and quality assurance graphics, with "M" and "A" as mnemonics for *Minus* and *Add* respectively. Here we make use of these same quantities as part of a formal statistical analysis.

A key observation is that M_{gi} and A_{gi} are uncorrelated. They represent independent contrasts of the correlated observations y_{gi2} and y_{gi1}. In terms of the above mixed model, M_{gi} can be interpreted as a within-spot contrast, whereas any comparison of the A-values represents a between-spot contrast. The variances are given by

$$\text{var}(M_{gi}) = \sigma_{Mg}^2 = 2\sigma_g^2(1 - \rho_g)$$

and

$$\text{var}(A_{gi}) = \sigma_{Ag}^2 = \sigma_g^2(1 + \rho_g)/2.$$

We see that the variance of M decreases with the intra-spot correlation whereas that of A increases. Note in fact that

$$\frac{1}{2} \log \frac{4\sigma_{Ag}^2}{\sigma_{Mg}^2} = \tanh(\rho_g),$$

so that ρ_g can be estimated from the ratio of the two variances.

Estimating the intra-spot correlation from the *M*-value and *A*-values

We now transform the mixed model for y into a model for the M-values and A-values. This has the effect of transforming the mixed model, in which the observations are not independent but all have the same variance, into a model in which all the observations are independent but the variances are unequal.

Write $M_g = (M_{g1}, \ldots, M_{gn})^T$ and $A_g = (A_{g1}, \ldots, A_{gn})^T$ for the vectors of M and A-values respectively for probe g.

Note that these vectors are linearly related to y_g through $M_g = C_M^T y_g$ and $A_g = C_A^T y_g$ with transformation matrices $C_M^T = (-1, 1) \otimes I_n$ and $C_A^T = (1/2, 1/2) \otimes I_n$. Also write $z_g^T = (M_g, A_g)^T$ for the combined $2n$-vector of M and A-values. Then z_g satisfies the linear model

$$E(z_g) = Z\beta_g$$

where Z is the related to the previous design matrix by $Z = (C_M \ C_A)^T X$.

The linear model for z_g is heteroscedastic, because the first n values of z_g have variance σ_{Mg}^2, whereas the remaining n values have variance σ_{Ag}^2. The model can be fitted using an efficient REML algorithm designed for heteroscedastic regression models [30]. This yields REML estimators $\hat{\sigma}_{Mg}^2$ and $\hat{\sigma}_{Ag}^2$ from which an estimate of ρ_g can be obtained. This regression approach is more general than the mixed model approach described above in that the estimated intra-spot correlation can take negative as well as positive values. The REML estimators $\hat{\sigma}_{Mg}^2$ and $\hat{\sigma}_{Ag}^2$ can be shown to follow approximate chisquare distributions with fractional degrees of freedom d_{Mg} and d_{Ag} respectively (see Methods).

Estimating the common intra-spot correlation

The intra-spot correlation ρ_g arises from the technical design of two channel arrays rather than from biological variation or from characteristics of the RNA sources being compared. It is therefore reasonable to assume that the intra-spot correlation will be relatively consistent across the probes. This leads to the argument that the correlation estimators $\hat{\rho}_g$ may be pooled between probes, an approach similar to that used by [27] when treating within-array replicate probes. From this point, we assume therefore that the intra-spot correlations are equal across probes, $\rho_g = \rho$ for all g.

If the data from different probes were statistically independent, then the REML estimator of $\tau = \tanh^{-1}(\rho)$ would be

$$\hat{\tau} = \frac{1}{2} \log \frac{4 \sum_{g=1}^{G} \hat{\sigma}_{Ag}^2}{\sum_{g=1}^{G} \hat{\sigma}_{Mg}^2}. \tag{2}$$

Although theoretically efficient under the independence assumption, this estimator is not robust against outliers.

A more robust estimator can be constructed by obtaining probe-wise estimates of τ. The individual probe version of equation (2) is

$$\hat{\tau}_g = \frac{1}{2} \log \frac{4\hat{\sigma}_{Ag}^2}{\hat{\sigma}_{Mg}^2}.$$

It is shown in Methods that $\hat{\tau}_g$ is a biased estimator of τ, but with a constant bias that can be computed from the degrees of freedom for σ_{Ag}^2 and σ_{Mg}^2. The bias correction term is similar to the analogous term derived in [27].

A trimmed mean of the bias-corrected $\hat{\tau}_g$ across all probes then provides a suitably robust estimator of τ and hence of $\rho = \tanh(\tau)$. Our software implementation trims 15% of the probes from each tail by default.

Common correlation inference reduces to ordinary linear modeling

The pooled estimator $\hat{\rho} = \tanh(\hat{\tau})$ is estimated from all the probes on the microarray, typically tens of thousands of probes, and hence can be considered to be a highly precise estimator on which each individual probe has little influence. A consequence of this is that $\hat{\rho}$ can be treated as known when undertaking inference about each individual probe. In particular, the heteroscedastic regression model for M and A-values described above can be transformed to an ordinary homoscedastic regression model by rescaling all the M-values by $\{2(1 - \hat{\rho})\}^{1/2}$ and all the A-values by $\{(1 + \hat{\rho})/2\}^{1/2}$. This leads to a re-scaled version z_g^* of z_g and a re-scaled version Z^* of the design matrix Z. The estimator of β_g finally is the ordinary least squares estimator

$$\hat{\beta}_g = \left(Z^{*T}Z^*\right)^{-1} Z^{*T}z_g^*.$$

In this way, the common correlation model permits us to undertake a separate channel analysis without incurring the inferential complexities of mixed models or heteroscedastic regression. Once the common intra-spot correlation is estimated, the separate channel analysis can utilize the established framework of linear modeling for microarray data [26]. In particular, empirical Bayes methods can be used to borrow strength between probes, leading to moderated t-statistics with exact distributions, as previously used for log-ratio analysis or for one-channel microarrays [26]. Details are given in Methods.

Efficiency gains for separate channels over log-ratios
Recovering information from the A-values

This section considers the relative efficiency gains of separate-channel versus log-ratio analysis for some commonly used designs. Efficiency is considered on a gene by gene level, so the subscript g is suppressed in this section.

Our reformulation of separate channel analysis in terms of M and A-values clarifies the relationship with log-ratio analysis, because the traditional log-ratio analysis is exactly equivalent to the M-value analysis. In other words the log-ratio analysis is equivalent to ignoring the second set of n observations in the heteroscedastic regression described above. The extra information recovered in the separate-channel approach compared to analysis of the log-ratios corresponds exactly to the information contained in the A-values about the treatment effects of interest. Recovering this information in general improves our power to detect treatment differences.

Paired design

The simplest comparative microarray experiment has only 2 treatments and only biological replication. On each array, one channel is hybridized to a sample from each treatment. In this case there is no information gained from recovering intra-probe information. Both the log-ratio and the separate channel approaches yield the same t-statistic (the paired t-test) for testing for differential expression for a given probe.

Common reference design

The second simplest design also has only 2 treatments (say B and C) and only biological replication. However, one channel on each array is hybridized to a common reference sample R, which is usually a technical replicate from a large RNA pool, while the other channel is used for a sample from one of the 2 treatments of interest. Suppose that there are $n/2$ arrays comparing B with the reference and $n/2$ comparing C with the reference. We will assume that the reference is always labeled green, but the analysis below is readily modified to accommodate dye-swap or dye-balanced designs, in which B and C may be hybridized with either label.

Consider the analysis for a given probe. Let $\bar{M}_B$ be the mean of the M-values for the arrays hybridized with B and $\bar{M}_C$ be the mean of the M-values for the arrays hybridized with C. Define $\bar{A}_B$ and $\bar{A}_C$ similarly. Then the laws of probability for sums and differences of random variables give:

$$\bar{M}_B \sim N(\beta_B - \beta_R, 4\sigma^2(1 - \rho)/n)$$
$$\bar{M}_C \sim N(\beta_C - \beta_R, 4\sigma^2(1 - \rho)/n)$$
$$\bar{A}_B \sim N(\beta_B/2 + \beta_R/2, \sigma^2(1 + \rho)/n)$$
$$\bar{A}_C \sim N(\beta_C/2 + \beta_R/2, \sigma^2(1 + \rho)/n)$$

where β_B, β_C and β_R are the log-expression values for the probe in RNA samples B, C and R respectively. Write γ for the log-ratio of interest, $\beta_B - \beta_C$. The M-values yield the estimator

$$\hat{\gamma}_M = \bar{M}_B - \bar{M}_C$$

with variance

$$\text{var}\hat{\gamma}_M = 8\sigma^2(1 - \rho)/n$$

The A-values yield the estimator

$$\hat{\gamma}_A = 2(\bar{A}_B - \bar{A}_C)$$

with variance

$$\text{var}\hat{\gamma}_A = 8\sigma^2(1 + \rho)/n$$

The statistical (Fisher) information provided by the M-values is $1/\text{var}\hat{\gamma}_M$ while that from the A-values is $1/\text{var}\hat{\gamma}_A$. Taking the ratio of these two variances shows that

the extra information provided by the A-values relative to that provided by the M-values is

$$\frac{1-\rho}{1+\rho}.$$

If ρ is close to one, then the added information is close to zero and the log-ratio analysis is nearly fully efficient. If ρ is small, however, then the A-values can contribute nearly as much information as the M-values, effectively doubling the statistical information that is used.

These formulas also allow us to explore the efficiency of the common reference design itself. Suppose that, instead of hybridizing the reference sample to each array, one channel of each array had been left empty. In that case, the log-fold-change $\beta_B - \beta_C$ would be estimated using a one channel analysis with variance $4\sigma^2/n$. Comparing this with $\text{var}\hat{\gamma}_M$ shows that the use of competitive hybridizations with a common reference improves the precision of the experiment if and only if the intra-spot correlation is greater than 0.5, assuming that the usual log-ratio analysis is performed. When the intra-spot correlation is less than 0.5, subtracting the reference from each M-values increases the variance of the observation rather than decreasing it. On the other hand, when a separate channel analysis is used, the use of the common reference channel always improves the precision of the experiment relative to not observing the reference channel at all. By using all the information available, the separate channel analysis restores the intuition that adding extra observations to a data set should not worsen the results.

Unconnected designs

Separate channel analysis is most useful in the case of unconnected designs for which some comparisons cannot be made through the M-values. Suppose that $n/2$ arrays are hybridized with RNA from sources B and C and $n/2$ arrays are hybridized with sources D and E. Then the comparisons B versus C and D versus E can be made using M-values but B or C vs D or E can not. Let $\bar{M}_{BC}$ be the mean of the M-values for the arrays hybridized with B and C and $\bar{M}_{DE}$ be the mean of the M-values for the arrays hybridized with D and E. Similarly for $\bar{A}_{BC}$ and $\bar{A}_{DE}$. Then

$$\bar{M}_{BC} \sim N(\beta_B - \beta_C, 4\sigma^2(1-\rho)/n)$$
$$\bar{M}_{DE} \sim N(\beta_D - \beta_E, 4\sigma^2(1-\rho)/n)$$
$$\bar{A}_{BC} \sim N(\beta_B/2 + \beta_C/2, \sigma^2(1+\rho)/n)$$
$$\bar{A}_{DE} \sim N(\beta_D/2 + \beta_E/2, \sigma^2(1+\rho)/n)$$

where β_B, β_C, β_D and β_E are the log-expression values for the probe in RNA samples B, C, D and E respectively. Let $\gamma = \beta_B - \beta_C$ be the population mean log-ratio between B and C for a given probe and let $\delta = \beta_B - \beta_D$ be the population mean log-ratio between B and D for the same probe. The A-values provide no information about

γ so the mixed model approach yields the same estimate and t-statistic as the M-value linear model approach. This estimator is simply

$$\hat{\gamma} = \bar{M}_{BC}$$

with variance

$$\text{var}\hat{\gamma} = 4\sigma^2(1-\rho)/n.$$

The M-values provided no information about δ meaning that the M-value approach is unable to estimate δ. The combined M and A-value approach yields the estimator

$$\hat{\gamma} = \bar{M}_{BC}/2 - \bar{M}_{DE}/2 + \bar{A}_{BC} - \bar{A}_{DE}$$

for δ with variance

$$\text{var}\hat{\gamma} = 4\sigma^2/n$$

Notice that the relative efficiency of $\hat{\delta}$ versus $\hat{\gamma}$ is $1 - \rho$. This represents the reduced efficiency of making an indirect contrast between arrays versus a direct contrast between channel hybridized to the same arrays.

As a published example of an unconnected design, consider the experiment of [15]. Jin et al. [15] considered gene expression differences by gender and age (1 and 6 weeks) in two genotypes of *Drosophila melanogaster*. Each array had a single combination of gender and genotype with both ages in the different channels. Thus this is a paired design in the age main effect, and this is the only effect that can be tested using the log-ratio analysis. Instead a mixed model separate channel analysis was used to analyze the design as a $2 \times 2 \times 2$ factorial design.

Size of efficiency gains in practice

As noted above, there are two extreme cases in which the efficiency gain from separate channel analysis is either zero or 100%. For the simplest two-color experiment in which two treatments are competitively hybridized on the same arrays, and probe-specific dye-effects are absent, then the log-ratios are fully efficient and there is no information to be gained from a separate channel analysis. The other extreme is that in which unconnected treatments are to be compared. In that case the log-ratios contain no information, and the information gained from separate channel analysis is 100% (or infinity relative to the log-ratio information).

For all other designs, the efficiency gain from separate channel analysis depends on the size of the intra-spot correlation. The intra-spot correlation measures the proportion of total variability arising from technical variability of arrays. Generally speaking, the more biological variability there is relative to technical inter-array variability in an experiment, the lower the intra-spot correlation and the greater the information gain. The titration experiment (see second case study) provides an example of an experiment with technical replication only. In this case, all the

variation is technical in nature and the intra-spot correlation is very high at 0.95. The ApoAI knockout experiment (see first case study) is conducted on genetically identical mice. This experiment has a moderate degree of biological variability and yields an intermediate intra-spot correlation of 0.85. The California poppy data (third case study) shows a greater degree of biological variability and yields a relatively low intra-spot correlation of 0.65. For a common reference design, the relative efficiency gain from separate channel analysis with these three intra-spot correlation scenarios would be 2%, 8% and 21% respectively. Note that an efficiency gain of 21% achieves the same improvement in precision as increasing the number of biological samples and arrays by the same percentage.

Case studies
Analysis pipeline
This section considers three case studies. The ApoAI knockout data is from academic spotted arrays while the titration example used commercial Agilent arrays. The California poppy study used custom Agilent arrays designed from a shallowly sequenced transcriptome. Unless otherwise noted, all analyses were undertaken using the `limma` software package [10].

The intensity data were background corrected and normalized prior to differential expression analysis as described in Methods. Briefly, M-values were loess normalized [20] using `normalizeWithinArrays` while A-values were quantile normalized [31] using `normalizeBetweenArrays`. This pipeline normalizes both the M-values and the A-values in a way that agrees with the usual M-value normalization for a traditional log-ratio analysis. The log-ratio and separate channel analyses used the same pre-processed data in each case to ensure the analyses are directly comparable.

Linear modeling used the `lmFit` function for log-ratio analyses and the `intraspotCorrelation` and `lmscFit` functions for separate channel analysis. In all cases the `eBayes` function was used to construct empirical Bayes moderated t-statistics and p-values [26].

ApoAI knockout experiment - a 2 sample common reference design
The apolipoprotein AI gene (known as either Apoa1 or ApoAI) is known to play a pivotal role in high density lipoprotein (HDL) metabolism. Mice which have the ApoAI gene knocked out have very low HDL cholesterol levels. Callow MJ, 2000 [32] discusses an experiment to determine how ApoAI deficiency affects the action of other genes in the liver, with the idea that this will help determine the molecular pathways through which ApoAI operates. A common reference design was used with 16 arrays, 8 wild-type and 8 knock-out mouse liver samples labeled with Cy5 and a common reference sample, created

by pooling RNA from the 8 wild-type mice labeled with Cy3. Samples were hybridized to a custom spotted cDNA microarray with 6384 probes.

Both log-ratio and separate channel analyses were conducted to find genes differentially expressed in the ApoAI samples as compared to wild-type. The intra-probe correlation was estimated to be 84.9%, suggesting an efficiency gain of 8% for the separate channel over the log-ratio analysis.

Figure 1 relates the p-values from the two analyses in a scatterplot on a $\log_{10}$ scale. Both analyses assign very low p-values to the same top eight genes, but the p-values from the separate channel analysis are several orders of magnitude lower than those from the log-ratio analysis. The separate channel analysis also detects more differentially expressed genes than the log-ratio approach at any false discovery rate (FDR) greater than 0.05 (Table 1).

The `qvalue` package [33] implements Storey's 2003 method [34] for estimating positive FDRs (q-values) and for estimating the total proportion of probes that are not differentially expressed (π_0). The log-ratio and separate channel analyses yielded similar values for $\hat{\pi}_0$ (86% and 88% respectively). This suggests that there are actually over 700 probes that are truly differentially expressed but were not detected at conventional FDR levels because of small fold changes or high variability. As expected, the separate channel approach shows a gain in power and is able to detect more of these probes. At the same time, the fact that $\hat{\pi}_0$ does not decrease for the separate channel analysis shows that the separate channel analysis is not
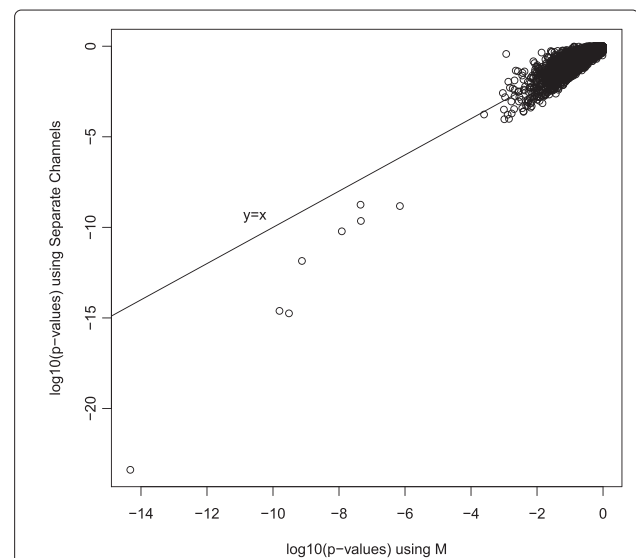

Figure 1 ApoAI data $\log_{10}$(p-values). $\log_{10}$(p-values) from the separate channel plotted against the $\log_{10}$(p-values) from the log ratio analysis for the ApoAI data. The p-values for the two methods are correlated, but the p-values from the separate channel analysis are smaller.

Table 1 Number of significant genes for the ApoAI knockout experiment

FDR	Log-ratio	Separate channel
0.01	8	8
0.05	8	8
0.10	8	15
0.25	9	53

The second and third columns show the number of genes declared differently expressed at various FDR levels by the log-ratio and separate channel analyses respectively. FDRs were estimated using the `qvalue` package. The separate channel analysis detects more significant genes.

on average decreasing the *p*-values of non-differentially expressed genes.

The results in Table 1 are for `qvalue` FDRs that incorporate the estimate of π_0. The nonadaptive Benjamini and Hochberg FDR estimator [35] gives similar results but is slightly more conservative. If Benjamini and Hochberg FDRs are used, the number of probes detected by the separate channel method decrease slightly to 13 and 44 at FDRs 0.1 and 0.25 respectively.

Titration data - a multi-treatment common reference design

Two cell lines with very different gene expression profiles were used to prepare a titration series of mRNA as a test of microarray technologies [2]. The two cell lines were MCF7, a cell line derived from breast epithelial cancer cells, and Jurkat, derived from T cell leukemia cells. RNA samples from the two cell lines were mixed in a titration series with 0%, 50%, 76%, 88%, 94%, and 100% MCF7 mRNA. Each mixture was labeled with both red and green and hybridized to 2 arrays. A separate 0% MCF7 sample was used as a reference on each array, in a dye-swap reference design. The data analyzed here were hybridized to Agilent commercial human arrays. The data are available at [36]. After normalization and filtering of very low expressing genes as described in Methods, 17985 spots were used in the analysis.

The analysis was performed as a comparison of each sample with the control sample with 0% MCF7. We expect that the number of genes detected as differentially expressed should increase with the percentage of MCF7 in the sample. Four analysis methods are compared: log-ratio analysis using `limma` [10] and three separate channel analyses: the ordinary linear mixed model as in [16] computed with `lme` [37], the common correlation separate channel method implemented in `limma` and the variance component shrinkage method implemented in `maanova` [17].

The ordinary mixed model [16] is the classical model for analysis of variance in an incomplete random block design. It includes a block (spot) effect, or equivalently an intra-spot correlation, and an error variance computed for each gene. In the context of microarray data analysis, it has the disadvantage of not borrowing information between probes, so the analysis of each probe relies solely on the data for that probe.

For `maanova` we used the linear mixed model with array as a random effect, and used the "Fs" option with tabulated p-values to estimate the differential expression p-values. This option imposes shrinkage of the gene-wise mean squared error similar to the method in `limma` as well as shrinkage of the gene-wise estimate of the random spot variance. However, there are two important differences between `maanova` and the common correlation separate channel analysis implemented in `limma`. Firstly, `maanova` moderates the gene-wise estimate of the spot effect, while the common correlation model imposes a common value ρ for the intra-spot correlation for all genes and models the spot variance σ_{gb}^2 for gene g as $\sigma_{gb}^2 = \sigma_{ge}^2 \rho/(1 - \rho)$. Because the common correlation model estimates the common intra-spot correlation ρ from the (typically) thousands of spots on the arrays, it is treated as a known constant. Hence the common correlation model adjusts only for the estimation of σ_{ge}^2 yielding larger error degrees of freedom. As well, because the implementation of the common correlation model in `limma` uses an empirical Bayes estimator for σ_{ge}^2, the posterior error degrees of freedom are used for evaluating p-values. `maanova` uses an empirical method of shrinkage of the estimates of both σ_{gb}^2 and σ_{ge}^2 which does not provide a degrees of freedom adjustment. Instead `maanova` has an option to compute the permutation null distribution of test statistic. However, in this analysis we chose to use the tabulated p-values, which use the usual ANOVA degrees of freedom. These match the degrees of freedom for the tests performed in the mixed model and adjust for estimation of both σ_{gb}^2 and σ_{ge}^2.

Figure 2 displays the number of genes significant at estimated FDR less than 0.01 using the Benjamini and Hochberg method [35]. The separate channel analysis discovers the most significant genes at all levels of dilution, followed by the log-ratio analysis. The other two analyses have less power. However, all 4 analyses agree on 14774 genes over all levels of dilution above 0. The conclusion is similar if we consider a false discovery rate less than 0.05 for the significance cut-off. While we might expect all the separate channel analyses to outperform the log-ratio analysis, the large intra-spot correlation of 0.95 assures high relative efficiency of the log-ratio analysis with an efficiency loss of less than 2.6%. Hence, the differences in power among the methods is due to both the additional information from recovery of inter-spot information and the variance regularization. To highlight this, the curve for
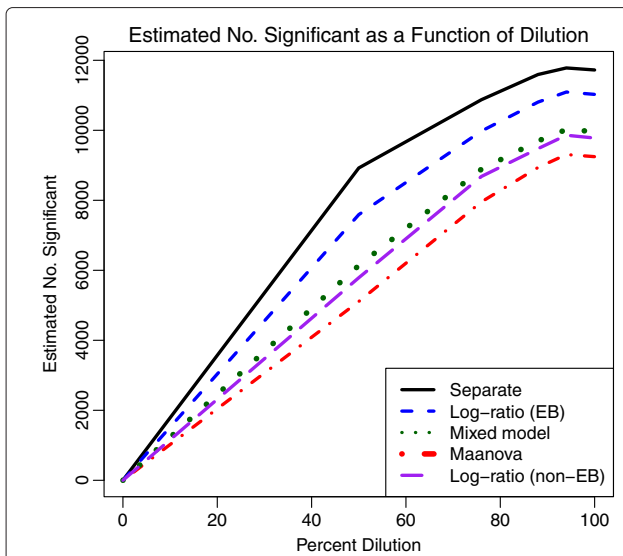
Figure 2 Number of significant genes for titration data. The number of significant genes increases with percent MCF7 for all analysis methods (Benjamini and Hochberg FDR < 0.01). The separate channel analysis makes the most discoveries at each dilution. The log-ratio method with variance shrinkage is the next most powerful. The linear mixed model which is uses a genewise variance estimator is more powerful than the log-ratio method with genewise variance estimator, but not as powerful as either of the methods using variance regularization. For these data, `maanova` is the least powerful.

Figure 3 Cumulative distributions of p-values for the titration data. The empirical cumulative distribution function of the log10(p-values) for the titration data using four analysis methods. The separate channel p-values are shown in black at 100% and 50% MCF7. The other methods are shown at 100% only. This shows that the separate channel analysis at 50% MCF7 is more powerful than the other methods at 100% MCF7. The log-ratio analysis with variance regularization is more powerful than `maanova` and the mixed model, but less powerful than the separate channel analysis with variance regularization, although the estimated efficiency gain for the latter is only 2.5%.

the log-ratio analysis with variance computed genewise without regularization is also plotted. As expected, it lies below the curve for the separate channel analysis using the mixed model.

Of the 8923 genes found significant by the separate channel analysis at 50% dilution and FDR< 0.01, 8354 (94%) are found significant at all higher dilutions and only 134 (1.5%) are found only at this dilution, which corresponds well to the estimated false discovery rate. By contrast, the next most powerful method, the log-ratio method, finds only 7591 significant genes at 50% dilution and FDR< 0.01.

To see the comparative results in more detail, Figure 3 shows the empirical cumulative distribution function of the log10 of the p-values for some of the comparisons. Curves farther to the left indicate smaller p-values and hence more powerful tests. For each method, the power of the tests increases as the percentage of MCF7 in the sample increases. In the interests of clarity, we show only a few of the curves. The two black lines are the common correlation separate channel analyses at 50% MCF7 (right-most curve) and 100% MCF7 (left-most curve). The remaining curves are for the 100% MCF7 comparison. The log-ratio analysis, which also uses variance regularization, is the most powerful among the remaining methods, but is only
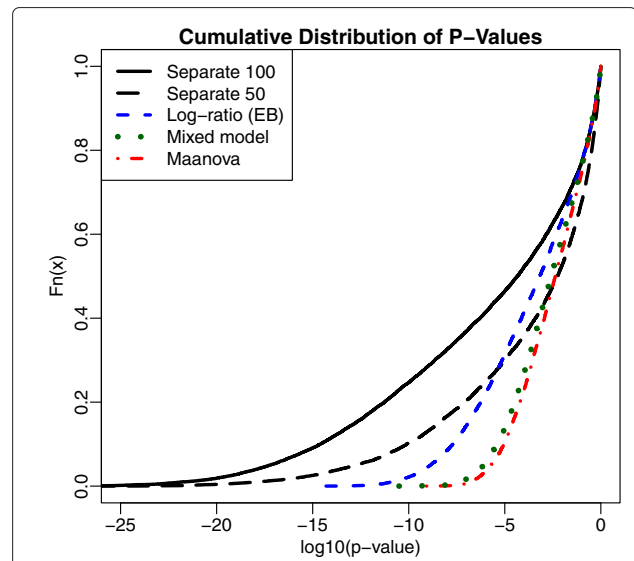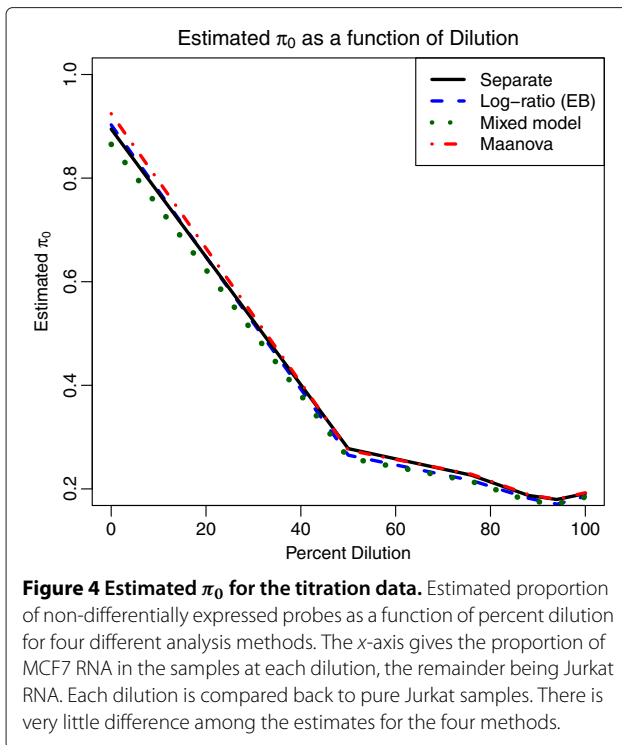
about as powerful as the separate channel analysis at 50% MCF7. The log-ratio analysis without variance regularization yields a curve that is almost indistinguishable from the mixed model. The correlation among the p-values for any pair of methods for the 100% MCF7 treatment is over 93%.

Figure 4 shows the estimated proportion π_0 of nondifferentially expressing genes using the method of Storey [34] as a function of the percentage of MCF7 in the sample for the four analysis methods. Note that the true set of differentially expressing genes does not depend on the percentage dilution (except of course for 0%) so that differences in the estimate of π_0 among the dilution levels is simply due to differences in detection power due to the increasing effect size induced by the dilution. We see that the estimate of π_0 decreases as the percentage of MCF7 in the sample increases, for all analysis methods. The estimates of π_0 at each level of dilution are very similar for all four analysis methods, varying by less than 2% except at the 0% dilution.

We expect $\pi_0 = 1$ at the 0% dilution, because the comparison is pure Jurkat vs Jurkat, but the estimates actually ranged from 0.87 (for the mixed model) to 0.93 (for `maanova`). While this cannot be explained on

Figure 4 Estimated π_0 for the titration data. Estimated proportion of non-differentially expressed probes as a function of percent dilution for four different analysis methods. The x-axis gives the proportion of MCF7 RNA in the samples at each dilution, the remainder being Jurkat RNA. Each dilution is compared back to pure Jurkat samples. There is very little difference among the estimates for the four methods.

biological grounds, it is possible that some unknown technical effects in handling the arrays introduced some subtle batch effects. Figure 5 displays the unadjusted p-values computed by each of the 4 methods for the 0% MCF7 comparison. The excess of small p-values accounts for the low estimates of π_0. For this comparison, the separate channel method gives 2 significant genes at 1% FDR and 29 at 2% FDR. While this is implausible from the biological point of view, all of the analysis methods indicate that $\pi_0 \leq 0.93$ as discussed above, indicating that there may be some unexplained technical effects inducing differential expression among these samples.

California poppy data - a double loop design
For an evolutionary study of gene expression in plant organs, tissues from 8 above ground organs were collected from California poppy in 4 biological replicates [38]. The mRNA was labeled so that 2 biological replicates of each tissue were labeled with each dye, and the samples were hybridized to custom Agilent microarrays with 6446 unique poppy probes in a double loop design. The experiment is described in detail in [38] and is illustrated in Figure 6. The design was selected to optimize the pairwise comparisons between certain tissues, while keeping the power of all pairwise comparisons as close as possible with 16 arrays and maintaining balance among the biological replicates.

The estimated intraclass correlation for this experiment is 65.2%. The relative efficiency of the comparisons depends on their relative locations on the double loop in Figure 6 [11].

Table 2 displays the number of differentially expressed genes detected for the leaf versus sepal and leaf versus carpel comparisons using log-ratios and separate channel analyses. In Figure 6, leaf tissue is in position A, sepal in position B and carpel in position E. The separate channel analysis detects substantially more significant genes for all comparisons and all FDR cutoffs. Almost all genes detected by the log-ratio approach were also detected by the separate channel method. At FDR < 0.01, all but 10 of the 668 genes detected as differentially expressed for leaf versus sepal comparison by the log-ratio analysis were also found by the separate channel analysis. For the leaf vs carpel comparison, all but 1 of the 132 genes found by log-ratio analysis were also found by the separate channel analysis. At the same time, the separate channel analysis does not decrease the estimate of π_0, suggesting that it is not decreasing the *p*-values of non-differentially expressed genes.

Conclusions
Two-channel microarrays continue to provide cost-effective platforms for whole-genome studies. Technological advances in printing and hybridization have greatly reduced the technical variance associated with microarray studies, while greatly increasing the number of features and reducing cost. Improvements in the statistical analysis of two-channel microarray data further improves cost-effectiveness by improving both sensitivity and specificity especially for small sample sizes.

This study has demonstrated the improvement in efficiency of differential expression analysis that can be achieved for most designs through the use of separate channel analysis. The separate channel analysis can be used to perform comparisons between treatments that may not be possible using log-ratio analysis in unconnected designs. By reparametrizing in terms of spotwise means and differences, the extra information in the separate channel analysis was shown to be that recovered from the *A*-values about the treatment conditions. The insight that the *M* and *A*-values are statistically independent throws some light on why they have been useful quantities for normalization and plotting purposes of two-channel data.

A common correlation estimation strategy has been proposed for the separate channel model, achieving considerable simplification as well as further gains in efficiency by stabilizing the intra-spot correlation estimates. The common intra-spot correlation is more readily interpreted than a set of varying probe-wise estimates, and yields a greater theoretical understanding of how much
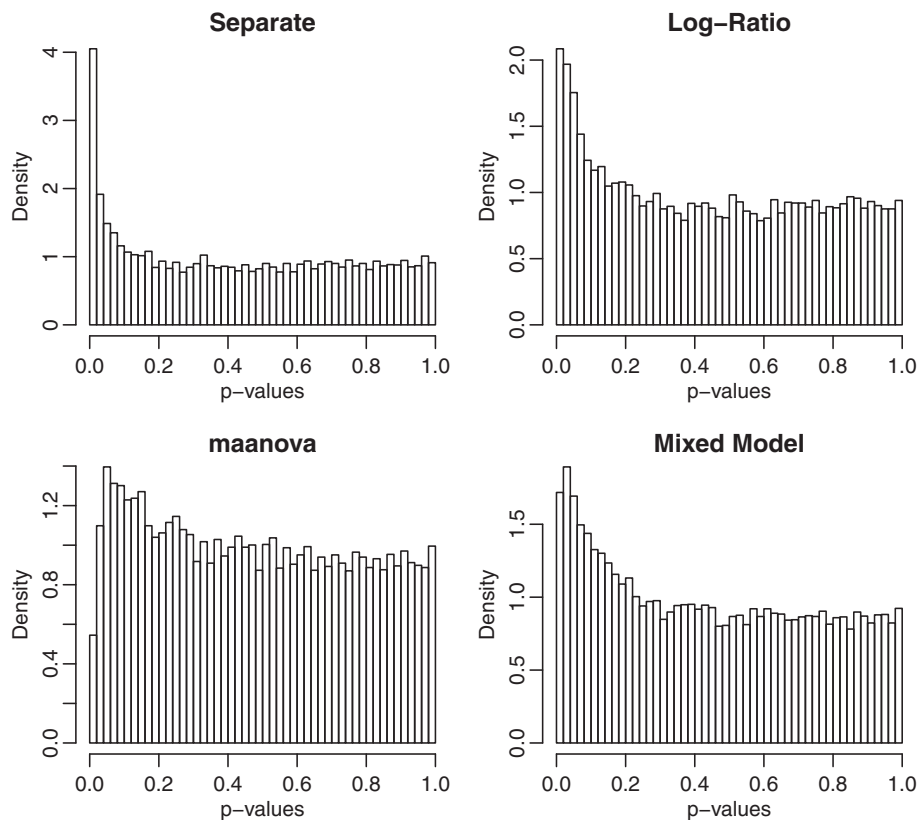
Figure 5 P-values for the control treatment of the titration data. For the control treatment (0% MCF7) the 100% of the genes should express nondifferentially. However, for all four analysis methods, the histogram of p-values shows an excess of small p-values indicative of differential expression.

information is gained from the separate channel analysis. The analysis shows that the efficiency gains are greatest when the intra-spot correlation is small, and this in turn occurs when biological variation is large. In other words, the gain in efficiency is greatest when it is most needed, when biological variation is large.

The common intra-spot correlation strategy is analogous to a common inter-spot correlation strategy proposed previously for combining information across multiple probes for the same gene on the same array [27]. The common correlation strategy can be viewed as an extreme form of shrinkage estimation in which the prior information about an individual probe *g* is reduced to a point mass at the common value. The current usage applies the same idea to observations from the same spot instead of from different spots in the same array.

Experiments with modest numbers of biological replicates produce variance estimators that are very unreliable on an individual probe-wise basis, so it is important to borrow strength between genes to achieve good statistical power and false discovery control on a genome-wide

basis. The common intra-spot correlation approach proposed here facilitates the use of empirical Bayes test statistics with an exact distributional theory in small samples [26]. This approach cannot usually be applied to mixed models with multiple variance components, because the variance component estimators are not independent and do not follow scaled chisquare distributions. General methods to shrink variance components using James-Stein type estimation have been proposed by Cui et al [17], and these offer an alternative route to regularized separate channel analysis. The resulting test statistics do not however have known null distributions in small samples, leading Cui et al [17] to suggest the use of permutation analysis to establish the null distributions. The common correlation approach of this article provides a more convenient strategy for complex multi-factor experiments.

The separate channel analysis reduces exactly to the log-ratio analysis of [26] if array is included as a factor in the linear model and the intra-spot correlation is set to zero. This exhibits the fact that the log-ratio analysis uses only within-array information while the separate channel
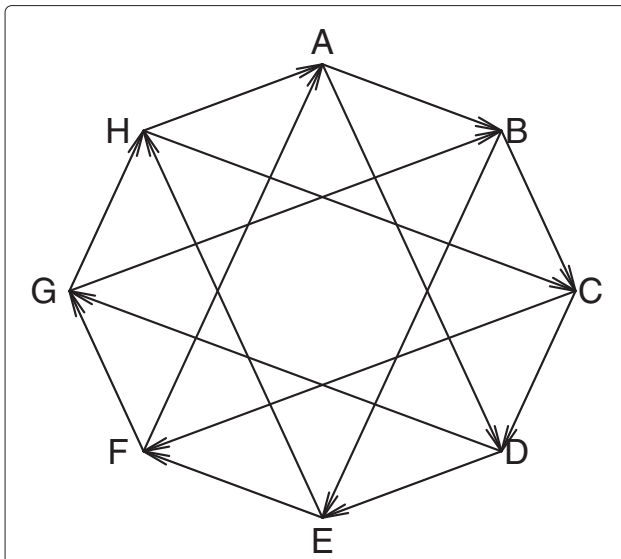
Figure 6 The double loop design for the poppy experiment. The double loop design for the poppy experiment. Each letter represents a tissue. Each edge represents a microarray. The head of the arrow represents the red channel and the tail represents the green. Each tissue is represented by 4 samples on 4 arrays.

analysis recovers information from between-array variation.

Another possible approach that we have explored is to apply empirical Bayes squeezing to the M and A-value variances $\hat{\sigma}^2_{Mg}$ and $\hat{\sigma}^2_{Ag}$. Unlike the usual variance components $\hat{\sigma}^2_{eg}$ and $\hat{\sigma}^2_{bg}$, the M-A variance components are approximately independent and chisquare distributed, and this is exact if the mixed model is balanced, in which case the M-A components are scaled versions of the within and between spot mean squares. Nevertheless the common correlation model, with its aggressive regularization of the intra-spot correlations, appears to be better motivated in terms of technical design of two-channel microarrays.

A number of real data examples with various experimental designs have been used to demonstrate that

the common correlation mixed model provides a powerful method for differential expression analysis that outperforms both the log-ratio method and separate channel analyses using the ordinary linear mixed model.

Methods

Empirical Bayes moderated *t*-statistics

This section briefly reviews the empirical Bayes differential expression approach implemented in the limma software package. Suppose that probe-wise linear models have been fitted to the expression data as described in Results. Suppose we wish to detect genes for which the jth coefficient β_{gj} is non-zero. This coefficient might represent for example a log-fold-change between two treatment conditions. The least squares estimator $\hat{\beta}_g$ is given in Results. The ordinary t-statistic for testing the null hypothesis that $\beta_{gj} = 0$ is

$$t_{gj} = \frac{\hat{\beta}_{gj}}{s_g/\sqrt{v_j}}$$

where s_g is the residual standard deviation for probe g and v_j is the effective sample size for estimating β_{gj} derived from the design matrix. In the notation used in Results, the effective sample size v_j is the jth diagonal element of the inverse of $Z^{*T}Z^*$. Under the null hypothesis, t_{gj} follows a t-distribution on $d = n - p$ degrees of freedom, where p is the number of coefficients in the linear model.

Following [25,26], an improved test can be obtained by computing the posterior variances

$$\tilde{s}^2_g = \frac{d_0 s^2_0 + d\, s^2_g}{d_0 + d} \tag{3}$$

and moderated t-statistics

$$\tilde{t}_{gj} = \frac{\hat{\beta}_{gj}}{\tilde{s}_g/\sqrt{v_j}}$$

Under the null hypothesis, $\tilde{t}_{gj}$ follows a t-distribution on $d_0 + d$ degrees of freedom. The gain in degrees of freedom of the moderated over the ordinary t-statistic reflects the information which is borrowed from other probes

Table 2 Number of significant genes for the poppy data

			Number of differentially expressed genes	
Analysis	Comparison	$\hat{\pi}_0$	FDR < 0.05	FDR < 0.01
Log-ratio	Leaf vs Sepal	0.52	1157	668
Separate channel	Leaf vs Sepal	0.64	1331	954
Log-ratio	Leaf vs Carpel	0.84	254	132
Separate channel	Leaf vs Carpel	0.87	505	314

Rows give results for different tissue comparisons and different analysis methods. The column $\hat{\pi}_0$ gives the estimated proportion of probes that are not differentially expressed between the tissue types. FDRs and π_0 values were estimated using the `qvalue` package. The separate channel method detects more significant genes but does not decrease $\hat{\pi}_0$ overall.

when making inferences about an individual probe. The moderated t-test has been shown to be superior to other tests in comparative studies [39-41].

The hyper-parameters s_0 and d_0 in the prior distribution for σ_g^2 are estimated from the expression data on all G probes as described by [26]. After this step, s_0 and d_0 are treated as known.

Approximate degrees of freedom for the *M* and *A*-value variances

Let $h = (h_1, \ldots, h_{2n})^T$ be the $2n$-vector of leverages that arises from fitting the heteroscedastic regression model described in the section on the common correlation model. The leverages are defined to be the diagonal elements of the hat-matrix

$$H_g = Z^* \left(Z^{*T} Z^* \right)^{-1} Z^{*T}$$

with Z^* defined in Results. The effective degrees of freedom associated with $\hat{\sigma}_{Mg}^2$ and $\hat{\sigma}_{Ag}^2$ are the residual degrees of freedom associated with the M_{gi} and the A_{gi} components respectively of the regression. In terms of the leverages, the effective degrees of freedom are $d_M = n - \sum_{i=1}^{n} h_i$ and $d_A = n - \sum_{i=n+1}^{2n} h_i$ respectively. This means that approximately

$$\hat{\sigma}_{Mg}^2 \sim \sigma_{Mg}^2 \chi_{d_M}^2 / d_M$$
$$\hat{\sigma}_{Ag}^2 \sim \sigma_{Ag}^2 \chi_{d_A}^2 / d_A$$

The fact that $\hat{\sigma}_{Ag}^2$ and $\hat{\sigma}_{Mg}^2$ follow approximate chisquare distributions implies that $\hat{\sigma}_{Ag}^2 / \hat{\sigma}_{Mg}^2$ follows a scaled F-distribution. If we define,

$$\hat{\tau}_g = \frac{1}{2} \log \frac{4 \hat{\sigma}_{Ag}^2}{\hat{\sigma}_{Mg}^2}$$

then it follows that from [27] that

$$E(\hat{\tau}_g) = \tanh^{-1}(\rho) + \text{bias}$$

with

$$\text{bias} = \psi(d_A/2) - \psi(d_M/2) + \log(d_A/2) - \log(d_M/2)$$

where $\psi()$ is the digamma function. Hence $(\hat{\tau}_g - \text{bias})$ is approximately unbiased for $\tau = \tanh^{-1} \rho$.

Software implementation

The common correlation separate channel approach is implemented in the R [42] package `limma` distributed as part of the Bioconductor project [43]. The intra-spot correlation is estimated by the `intraspotCorrelation` function, and the separate channel linear model is fitted by the function `lmscFit`. In other respects, separate channel analyses follow exactly the same framework as other analyses using the `limma` package. The package includes normalization methods, differential expression analysis

and output tabulation for both log-ratio and separate channel analysis.

Case study datasets

The ApoAI knockout experiment was described by [32]. RNA samples were hybridized to academic spotted arrays and images were quantified using SPOT. A log-ratio analysis of the microarray data from the experiment is described in Section 8.4.2 of the limma User's Guide [44]. The data were pre-processed for this study as described in the limma User's Guide except for the addition of quantile normalization of the A-values. The microarray data are available as an R data object from the limma package home page [45].

The titration data have been previously described by [2]. The original study considered four microarray platforms and three image analysis programs. The current study considers only the Agilent arrays because these were found to be the most precise of the two-color platforms [2]. Twenty four RNA samples were hybridized to 12 Agilent Human 1A Oligo microarrays printed with 22,000 oligonucleotide probes. Images were quantified using Agilent Feature Extraction software. Intensities were background corrected by the maximum likelihood normexp method [28,46] using the `backgroundCorrect` function of the limma package. Intensities were global loess normalized using `normalizeWithinArrays` and A-quantile normalized using `normalizeBetweenArrays`. All positive control probes were filtered before background correction and normalization. After normalization, negative control probes and probes with average log-intensity less than the 75% quantile of the negative controls were filtered from subsequent analysis.

The poppy data have been previously described by [38] where the probe selection was given in detail. Due to a complex probe design, probes were filtered prior to normalization to select one representative probe per unigene and exclude numerous control probes as described in [38]. After filtering, the data were preprocessed by background correction using `backgroundCorrect` with `method=half`. Intensities were global loess normalized using `normalizeWithinArrays` and A-quantile normalized using `normalizeBetweenArrays`.

Author details
[1] Walter and Eliza Hall Institute of Medical Research, Melbourne, Vic 3052, Australia. [2] Department of Mathematics and Statistics, University of Melbourne, Vic 3010, Australia. [3] Department of Statistics, The Pennsylvania State University, University Park, PA 16802–2111, USA.

References
1. Shalon D, Smith SJ, Brown PO: **A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization.** *Genome Res* 1996, **6**:639–645.
2. Holloway AJ, Oshlack A, Diyagama DS, David D, Bowtell DDL, Smyth GK: **Statistical analysis of an RNA titration series evaluates microarray precision and sensitivity on a whole-array basis.** *BMC Bioinformatics* 2006, **7**:511.
3. Passos V, Tan F, Winkens B, Berger M: **Optimal designs for one- and two-color microarrays using mixed models: a comparative evaluation of their efficiencies.** *J Comput Biol* 2009, **16**:67–83.
4. Jensen P, Halbrendt N, Fazio G, Makalowska I, Altman N, Praul C, Maximova S, Ngugi HK, Crassweller R, Travis J, McNellis T: **Rootstock-regulated gene expression patterns associated with fire blight resistance in apple.** *BMC Genomics* 2012, **13**:9.
5. Small HJ, Williams TD, Sturve J, Chipman JK, Southam AD, Bean TP, Lyons BP, Stentiford GD: **Gene expression analyses of hepatocellular adenoma and hepatocellular carcinoma from the marine flatfish Limanda limanda.** *Dis Aquat Organ* 2010, **88**:127–141.
6. Bay LK, Ulstrup KE, Nielsen HB, Jarmer H, Goffard N, Willis BL, Miller DJ, Van Oppen MJH: **Microarray analysis reveals transcriptional plasticity in the reef building coral Acropora millepora.** *Mol Ecol* 2009, **18**:3062–3075.
7. Chen Y, Dougherty ER, Bittner ML: **Ratio based decisions and the quantitative analysis of cDNA microarray images.** *J Biomed Opt* 1997, **2**:364–374.
8. Smyth GK, Yang YH, Speed TP: **Statistical issues in microarray data analysis.** *Methods Mol Biol* 2003, **224**:111–136.
9. Yang YH, Speed TP: **Design and analysis of comparative microarray experiments.** In *Statistical Analysis of Gene Expression Microarray Data*. Edited by Speed TP: Chapman & Hall/CRC Press; 2003:35–91.
10. Smyth GK: **Limma: linear models for microarray data.** In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. Edited by Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W. New York: Springer; 2005:397–420.
11. Altman NS, Hua J: **Extending the loop design for 2-channel microarray experiments.** *Genet Res* 2006, **88**:153–163.
12. Kerr MK: **Linear models for microarray data analysis: hidden similarities and differences.** *J Comput Biol* 2003, **10**:891–901.
13. Smyth G: **Individual channel analysis of two-colour microarray data Invited session IPM 11: computational tools for microarray analysis.** In *55th Session of the International Statistics Institute*. Sydney: International Statistics, Institute; 2005.
14. Kerr MK, Martin M, Churchill GA: **Analysis of variance for gene expression microarray data.** *J Comput Biol* 2000, **7**:819–837.
15. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G: **The contributions of sex, genotype and age to transcriptional variance in** *Drosophila melanogaster*. *Nat Genet* 2001, **29**:389–395.
16. Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS: **Assessing gene significance from cDNA microarray expression data via mixed models.** *J Comput Biol* 2001, **8**:625–637.
17. Cui X, Hwang JG, Qiu J, Blades NJ, Churchill GA: **Improved statistical tests for differential gene expression by shrinking variance components estimates.** *Biostatistics* 2005, **6**:59–75.
18. Lynch A, Neal D, Kelly J, Burtt G, Thorne N: **Missing channels in two-colour microarray experiments: combining single-channel and two-channel data.** *BMC Bioinformatics* 2007, **8**:26.
19. Yang YH, Thorne NP: **Normalization for two-color cDNA microarray data.** In *Science and Statistics: A Festschrift for Terry Speed*. Edited by Goldstein DR. Hayward: Institute of Mathematical Statistics; 2003:403–418. [IMS Lecture Notes - Monograph, Series, Volume 40].
20. Yang YH, Dudoit S, Luu P, Speed TP: **Normalization for cDNA microarray data.** In *Microarrays: Optical Technologies and Informatics*. Edited by Bittner ML, Chen Y, Dorsel AN, Dougherty ER. San, Jose: Society for Optical Engineering; 2001:141–152. [Proceedings of SPIE, Vol 4266].
21. Dudoit S, Yang YH, Callow MJ, Speed TP: **Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments.** *Statistica Sinica* 2002, **12**:111–139.
22. Stein C: **Inadmissibility of the usual estimator for the mean of a multivariate distribution.** In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press; 1956:197.
23. Efron B, Morris C: **Data analysis using Stein's estimator and its generalizations.** *J Am Stat Assoc* 1975, **70**(350):311–319.
24. Baldi P, Long AD: **A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes.** *Bioinformatics* 2001, **17**:509–519.
25. Wright GW, Simon RM: **A random variance model for detection of differential gene expression in small microarray experiments.** *Bioinformatics* 2003, **19**(18):2448–2455.
26. Smyth GK: **Linear models and empirical Bayes for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**: Article 3.
27. Smyth GK, Michaud J, Scott H: **The use of within-array replicate spots for assessing differential expression in microarray experiments.** *Bioinformatics* 2005, **21**:2067–2075.
28. Ritchie M, Silver J, Oshlack A, Holmes M, Diyagama D, Holloway A, Smyth G: **A comparison of background correction methods for two-colour microarrays.** *Bioinformatics* 2007, **23**:2700–2707.
29. Searle SR, Casella G, McCulloch CE: *Variance Components*. New York: Wiley; 1992.
30. Smyth GK: **An efficient algorithm for REML in heteroscedastic regression.** *J Comput Graphical Stat* 2002, **11**:836–847.
31. Bolstadt BM, Irizarry RA, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19**:185–193.
32. Callow MJ, Dudoit S, Gong EL, Speed TP, Rubin EM: **Microarray expression profiling identifies genes with altered expression in HDL-deficient mice.** *Genome Res* 2000, **10**(12):2022–2029.
33. Dabney A, Storey JD: **qvalue: Q-value estimation for false discovery rate control.** Package version 1.26.0 [http://www.bioconductor.org]
34. Storey JD: **The positive false discovery rate: A Bayesian interpretation and the q-value.** *Ann Stat* 2003, **31**:2013–2035.
35. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: A practical and powerful approach to multiple testing.** *J R Stat Soc Ser B* 1995, **57**:289–300.
36. Oshlack A, Smyth GK: **Supplementary Materials for Holloway et al "Statistical analysis of an RNA titration series evaluates microarray precision and sensitivity on a whole-array basis".** 2006. [http://bioinf.wehi.edu.au/folders/mixture]
37. Pinheiro JC, Bates DM: *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag; 2000.
38. Zahn LM, Ma X, Altman NS, Zhang Q, Wall PK, Tian D, Gibas CJ, Gharaibeh R, Leebens-Mack JH, dePamphilis CW, Ma H: **Comparative transcriptomics among floral organs of the basal eudicot** *Eschscholzia californica* **as reference for floral evolutionary developmental studies.** *Genome Biol* 2010, **11**:R101.
39. Kooperberg C, Aragaki A, Strand AD, Olson JM: **Significance testing for small microarray experiments.** *Stat Med* 2005, **24**(15):2281–2298.

40. Diboun I, Wernisch L, Orengo CA, Koltzenburg M: **Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma.** *BMC Genomics* 2006, **7:**252.
41. Murie C, Woody O, Lee AY, Nadon R: **Comparison of small n statistical tests of differential expression applied to microarrays.** *BMC Bioinformatics* 2009, **10:**45.
42. R Development Core Team: **R: A language and environment for statistical computing R Foundation for Statistical Computing.** 2011. [http://www.R-project.org/] [Vienna, Austria. ISBN 3-900051-07-0].
43. **Bioconductor: Open Source Software for Bioinformatics.** [http://bioconductor.org]
44. Smyth GK, Ritchie M, Thorne N, Wettenhall J, Shi W: **limma: Linear Models for Microarray Data User's Guide.** Package version 3.14.0 2013, [http://www.bioconductor.org]
45. Smyth GK: **LIMMA: Linear Models for Microarray Data.** [http://bioinf.wehi.edu.au/limma]
46. Silver J, Ritchie M, Smyth G: **Microarray background correction: maximum likelihood estimation for the normal–exponential convolution.** *Biostatistics* 2009, **10**(2):352–363.