Conference on Applied Statistics in Agriculture 2005 - 17th Annual Conference Proceedings

STATISTICAL ANALYSIS OF GENE EXPRESSION MICROARRAYS

Tanzy Love

Alicia Carriquiry

STATISTICAL ANALYSIS OF GENE EXPRESSION MICROARRAYS

Tanzy Love and Alicia Carriquiry
Department of Statistics, Iowa State University
Ames, IA 50011-1210
Phone:(515) 294-3440
Fax:(515) 294-4040
tanzy@iastate.edu

Abstract

This manuscript is composed of two major sections. In the first section of the manuscript we introduce some of the biological principles that form the bases of cDNA microarrays and explain how the different analytical steps introduce variability and potential biases in gene expression measurements that can sometimes be difficult to properly address. We address statistical issues associated to the measurement of gene expression (e.g., image segmentation, spot identification), to the correction for background fluorescence and to the normalization and re-scaling of data to remove effects of dye, print-tip and others on expression. In this section of the manuscript we also describe the standard statistical approaches for estimating treatment effect on gene expression, and briefly address the multiple comparisons problem, often referred to as the big p small n paradox. In the second major section of the manuscript, we discuss the use of multiple scans as a means to reduce the variability of gene expression estimates. While the use of multiple scans under the same laser and sensor settings has already been proposed (Romualdi et al. 2003), we describe a general hierarchical modeling approach proposed by Love and Carriquiry (2005) that enables use of all the readings obtained under varied laser and sensor settings for each slide in the analyses, even if the number of readings per slide vary across slides. This technique also uses the varied settings to correct for some amount of the censoring discussed in the first section. It is to be expected that when combining scans and correcting for censoring, the estimate of gene expression will have smaller variance than it would have if based on a single spot measurement. In turn, expression estimates with smaller variance are expected to increase the power of statistical tests performed on them.

1 Introduction

The rapid increase in the use of high-throughput technologies in molecular biology, such as gene expression microarrays, has produced large amounts of data. This makes a collaboration between quantitative analysts and subject-matter specialists particularly important and potentially very fruitful.

The biological hypotheses that may be tested with gene expression microarrays are numerous and require varied statistical methodologies. The statistical issues associated with the analyses and interpretation of these data include the design, measurement, and analysis of the experiment. Here we examine some of the statistical methods relevant to the analyses of microarray data and use a particular experiment with maize to demonstrate some of these ideas.

Gene expression is the term used to denote the extent to which a gene is used within a cell at a particular time or under a certain stimulus. Cells respond to different stimuli by using genes with different abilities and functions. This is how cells within an organism differentiate and respond to a changing environment. Genes are used by being transcribed into mRNA and then mRNA is translated into proteins within the cell. The complex uses and coordination of these proteins are largely unknown.

Ideally, we would like to assess gene function by measuring the amounts of each protein the genes produce. Instead, we determine gene expression by counting the number of mRNA molecules in the transcription phase. mRNA molecules have complementary molecules which they will hybridize (attach) to and this makes them relatively easy to count. However, mRNA abundances and protein abundances are not perfectly correlated; not all mRNA sequences are translated into proteins and certain proteins can build up in a cell by not degrading. Still, mRNA abundance serves as a proxy for protein abundance.

Typically, we wish to compare gene activity in different tissues within an organism or in the same tissue but under different conditions. The idea is to determine which genes are turned on or off (up or down regulated) in different tissues or in different environments. This information is then used in the pursuit of the holy grail of genomics: determining the function of each gene.

2 Construction and Imaging

Gene expression values are generally measured as relative intensities; there are many sources of variation in expression values (labs, technicians, slides, ...). Therefore, a typical experiment will include two or more biological samples. The gene expression in each of these samples is measured for thousands of genes simultaneously. First, the mRNA is extracted from each of the biological samples. Care is taken to ensure that equal amounts of mRNA are obtained from each sample to prevent a bias in the relative expression values. This follows from the common assumption that the amount of mRNA is constant. Subsequent normalizing procedures (such as used in Newton et al. (2001)) assume that the total mRNA for the two samples is the same. The mRNA is then reverse-transcribed to create cDNA sequences complementary to the original sequences.

There are several gene expression microarray technologies. In our examples, we will

discuss cDNA microarrays. This microarray technology allows two samples of cDNA to be measured on the same glass slide and therefore adjust for the slide specific effects on gene expression. The created cDNA is tagged with fluorescent dye. There are several available dyes and the two samples to be placed on the same array are tagged with different dyes. At least in terms in usage, the two most popular dyes are called Cy3 and Cy5, which dye the samples green and red, respectively.

The arrays consist of known probe sequences of nucleic acids affixed at known positions of the medium. Each probe contains many copies of a known sequence of a gene's coding sequence to allow hybridization by the tagged cDNA sequences. Two differently dyed samples of cDNA prepared from two biological samples are mixed and washed over the array. Dyed sequences will hybridize (attach) to their complementary probe sequences. The amount of dyed cDNA that attached to each of the fixed probes is measured by determining the amount of fluorescence at each spot.

The cDNA slides used in gene expression microarrays are produced in various private and academic laboratories. The number and content of the probes on a cDNA slide are determined by the scientists producing the slide who choose the sequences to be used as probes and the layout of the spots including replicating spots at multiple locations on the slide. Slides are printed by a computer that drives a robot with a set of print-tips at the end of an arm. Print-tips are dipped into vials of solution containing the cDNA probes. The robot then deposits a small drop of solution on the slide creating a spot. Most arrayers spot several probes at once on a slide using multiple print-tips. A standard print-head contains 32 tips arranged in a four by eight block, thus 32 spots are printed at a time. Each tip prints a block of spots on the slide and this may induce spatial correlation among the spots in a print-tip group (sector). Replicated spots may give insights into the spatial variation across a slide (since they measure the same expression), however this is not straightforward and they are rarely used in practice.

Dyed and mixed cDNA from two biological samples is placed on the slide and hybridizes with the complementary probes. The fluorescent dyes are excited with a laser and a sensor records the intensity at each spot. The two dyes fluoresce at different laser wavelengths so we obtain separate images for each sample. The fluorescence intensities measured by the sensor are recorded as numbers between 0 and $2^{16} - 1$=65,535. The value corresponds to a color depth of sixteen bits per pixel. All pixel readings over the upper threshold will be recorded as 65,535 and all readings below the lower threshold will be recorded as 0.

3 Preprocessing

Preprocessing of cDNA microarray data consists of transforming the pixel intensity values from the image scanned into corrected and normalized estimates of gene expression intensity for each gene represented on the array. These preprocessing steps attempt to make use of

knowledge about the many structural biases in the microarray measurement process (such as the fluorescence bias between the different dyes). For further discussion of the issues of preprocessing microarray data see Smyth, Yang, and Speed (2002), Yang et al. (2002)and Lönnstedt and Speed (2001).

The first level of preprocessing consists of image processing, converting the image into numerical summaries for each spot. This must be done before further analyses can be carried out and usually consists of the following steps:

- Segmentation - classify each pixel as spot, background, or other.
- Intensity summary - calculate a summary value for the intensity of signal and background at each spot.
- Background correction - correct signal values for the background intensity.

Most of these operations are carried out using specialized software and sometimes these programs are black boxes.

Image segmentation is the process of determining whether a pixel is signal, background, or other. The intensity of a signal pixel is the sum of the fluorescence due to dyed target molecules hybridized with the probe at each spot and the background fluorescence. Background fluorescence arises due to dust, stray dye molecules, etc. that we do not wish to measure in our gene expression estimates. Background fluorescence varies across the slide, thus we measure and correct for local (around each spot) background.

Once pixels have been assigned to signal or background for each spot, summaries over the pixels must be calculated. Three common summaries are the mean, median, or mode of the pixels though more complicated summaries could also be used. Other characteristics of the spots, such as the area (number of pixels) or the total intensity (sum of pixel intensities), are sometimes used to identify poor quality spots. Censoring of a spot may also indicate a spot of poor quality because the true value of the fluorescence is not measured. Therefore, these spots may be excluded from further analysis or analyzed differently from other spots as will be discussed later in the paper.

Note that using a summary of the pixels masks the censoring (above or below) of the individual pixels. However, this only means that the censoring thresholds for the spots are different from the thresholds of the pixels. The proportion of saturated spots and spots censored below depend on the laser and sensor settings. Figure 1 compares low and high scans of a slide to a medium scan. In (a), some of the spots are censored above in the high scan and not in the medium scan. Similarly in (b) some of the spots are censored below in the low scan, but not in the medium scan. These values are background corrected spot averages, so the thresholds are not the same as those of the censored pixels.

Generally, additive background noise is assumed for microarray pixels. A global background can be estimated as the median of all pixels on the slide that have been classified as background. If we assume this form of background noise, then the the same global background value is subtracted from all spot signal values. When background variation on the slide is large, this is not a good approach. An alternative approach is to estimate and correct for local background before any analysis takes place. In either case, background-corrected spot intensities may be negative if the signal summary value is lower than the background estimate.

Normalization consists in minimizing the non-biological variation in measured gene expression. The lower the 'noise' the more reliably we can estimate biological differences in gene expression across different dyes and slides. There are many sources of non-biological variability:

- Dye differences in non-specific binding (Cy3 often binds with the slide).
- Dye differences in binding with sample cDNA sequences (dyes alter hybridization for certain sequences).
- Different amounts of sample cDNA hybridized to slides (this is kept constant between the two samples on a slide).
- Variability between characteristics of the slides, operators, and labs.

There are several types of normalization methods currently in practice. The simplest of these is constant label normalization to remove (minimize) the difference between the binding proficiency and the fluorescence of the two dyes. Let R_i denote the observed expression level of the ith red-dyed spot and similarly, let G_i denote the observed expression level of the ith gene when it is dyed green. Figure 2(a) shows that a scatter plot of the two dye channels, (R_i, G_i) for $i = 1, \ldots, n$, does not fall along the diagonal line. The spots have a higher median intensity estimate on the green channel than on the red channel. However, we assume that only a small proportion of the spots should be differentially expressed (have different true expression in the two samples on a slide). A simple correction for the label bias is subtracting the difference between the medians of expression for the two channels, $G_i^* = G_i - (\text{median}(G_i) - \text{median}(R_i))$. This will cause the medians to be the same for both dye channels on the slide.

A more complex form of label normalization assumes that the dye bias is intensity dependent instead of constant. To examine this technique, we introduce a different notation for channel intensities. Consider the following two functions of intensities:

$$M_i = \log(R_i/G_i) \tag{1}$$

and

$$A_i = \log \sqrt{R_i * G_i} \tag{2}$$

(e.g., (Speed 2003)). Plotting these values corresponds to a 45^o rotation of the (R_i, G_i) plot. If there are no dye effects on intensities, a scatterplot of M against A should suggest a zero correlation between the two quantities and the M_i should be centered around zero. In general, this does not happen. Figure 2(b) shows the data from the top figure transformed to (A, M) values. Again, we can see that there is a bias from the dyes because the M_i values are not centered around zero. Further, there is a nonlinear relationship between M and A.

The most popular adjustment for intensity-dependent dye bias is to correct the M_i values by fitting a loess curve of M_i on A_i and then subtracting the loess estimates (i.e. obtaining loess residuals). Locally weighted polynomial regression, called loess, is a smoothing method that estimates a line of trend through a dataset by combining polynomial regressions from small subsets of the data (Cleveland 1979). Figure 3 shows the loess fit through the data and some loess residuals. The assumptions of this loess adjustment are the following:

1. Most genes are not differentially expressed between the two samples on a slide, so $R_i = G_i$ for most i.

2. In an experiment with a large number of genes, some genes will be up-regulated and some down-regulated, so positive and negative deviations should cancel each other out.

Unfortunately, this method produces normalized values of M_i (which are used in the further analysis of Smyth et al. (2002)), but not of R_i and G_i. To adjust the individual sample intensities, one method is to add half of the loess adjustment to one dye channel and subtract the same amount from the other channel. The corrected (loess residual) values for M_i are plotted against A_i in Figure 4(a) along with the corrected values for R_i and G_i.

The final common type of normalization, scale normalization, removes intensity-dependent dye bias and equalizes the overall variance between the dye channels. Using loess to do intensity-dependent normalization removes dye bias, but the overall intensities and their variability may still differ across slides.

Finally, when artifacts are apparent within the slide, such as blocks on the slide created by the same print tip, further methods may be necessary to remove the non-biological bias from the expression data prior to analyses. Note that the fourth column in each of the rows of box plots of M in Figure 5 has a higher median than the other columns. This suggests a bias such as an uneven washing of the fourth row of the slide. Print tip groups are also surrogates for spatial variability on the slide and, therefore, we also correct for some spatial effects when normalizing in this way. To minimize the effect of local biases due to print tips, we implement a normalization within each print tip group (sector) on the slide. We again assume that most of the genes are not differentially expressed so that $R_i = G_i$ for most i

within the sector and carry out normalization in the same way as previously described on the data from each sector separately.

In our work, we perform intensity-dependent dye bias normalization on each of the print tip groups. Therefore, the normalized R's and G's are corrected for dye, slide, print tip group (or spatial), and intensity effects.

4 Differential Expression

Scientists are interested in identifying differentially expressed genes to gain clues as to their function. For example, a gene that is up-regulated in plant tissue subjected to light but not in tissue subjected to darkness may be related to photosynthesis and genes whose expressions change in plants subjected to drought may be important in the development of drought-resistant varieties of crops. Of interest is $\mu_{g1} - \mu_{g2}$, the difference in the expected mean expression of gene g over treatments 1 and 2. The effects of more than two treatments may also be of interest.

Consider a simple experiment with two treatments. Let the effect of treatment i be denoted by T_i for $i = 1, 2$. In a basic experimental design, we can account for treatment, dye and slide effects. For gene g, let Y_{gijk} denote the observed log signal intensity in the ith treatment, jth dye and kth slide. Within a classical framework, treatment and dye effects are often modeled as fixed and slide effects as random. Consider the following mixed-effects model:

$$Y_{gijk} = \mu_g + T_{gi} + d_j + s_k + e_{gijk}, \tag{3}$$

with $s_j + e_{gijk} \sim N(0, \sigma_s^2 + \sigma_e^2)$.

In this example, the gene is differentially expressed if $T_{g1} \neq T_{g2}$. An estimate of the difference in treatment effects for gene g is $\bar{Y}_{g1\bullet\bullet} - \bar{Y}_{g2\bullet\bullet}$ with confidence interval

$$\bar{Y}_{g1\bullet\bullet} - \bar{Y}_{g2\bullet\bullet} \pm t_{\{n_{g1}+n_{g2}-2;1-\frac{\alpha}{2}\}} \hat{\sigma}_{\bar{Y}_{g1\bullet\bullet} - \bar{Y}_{g2\bullet\bullet}}, \tag{4}$$

where n_{g1} and n_{g2} are the number of terms in the averages $\bar{Y}_{g1\bullet\bullet}$ and $\bar{Y}_{g2\bullet\bullet}$, respectively. The estimated fold change is $\exp(\bar{Y}_{1\bullet\bullet} - \bar{Y}_{2\bullet\bullet})$, with approximate confidence interval estimated from the log-scale interval. For more than two treatments, ANOVA methods are used to estimate the treatment effects and confidence intervals.

In addition to detecting differences among two or more treatments, scientists are interested in discovering patterns of expression across multiple time points. For example, if we obtain samples during the developmental stages of an animal's organs, can we detect genes with non-null patterns of expression over time? Further, can we group the genes according to the expression pattern over time? Similar patterns of expression suggest a common

regulatory pathway and scientists can infer possible gene functions or relationships by understanding these regulatory pathways. Both heuristic and statistical clustering methods can group the genes according to their expression patterns.

While the analysis of differential expression can be performed in a straightforward way, there are several challenging issues which are not explicitly considered by these methods. First is the fact that signal intensities Y have undergone significant pre-processing. What we optimistically call an observed signal intensity is a noisy estimate of true signal. Secondly, there is a large p, small n problem: for each of a large number p of genes we perform tests using only a small number n of sample individuals. For example, we may test the following for p different genes:

$$H_0 : \text{no fold change versus } H_a : \text{fold change.} \tag{5}$$

The naive approach is to designate a threshold α and declare all genes with a p-value below the threshold to be differentially expressed. We can expect a large number of false positives; the true Type 1 error rate will be much higher than α. Thus the recent thrust to develop methods to control the false positive rate (e.g., (Benjamini and Hochberg 1995), (Dudoit et al. 2002), (Storey 2002), (Benjamini and Yiekutieli 2001)).

An alternative analysis of expression values that considers measurement error and multiple comparisons explicitly is hierarchical modeling. Using the log expression values as before, we can formulate a model hierarchically:

$$\begin{aligned} Y_{gijk} &\sim N(\mu_{gij}, \sigma^2_{ij}) \\ \mu_{gij} &\sim N(\mu_{gi}, \sigma^2_i), \end{aligned} \tag{6}$$

where the μ_{gij}, the mean expression of gene g in treatment i and dye j, is assumed to be exchangeable given treatment and arising from a distribution with mean μ_{gi}. Inferences about the expression change between treatments for gene g are based on the posterior distribution of $\mu_{gi} - \mu_{gi'}$.

Another hierarchical model, Gamma-Gamma, for gene expression for proposed by Newton et al. (2001). This model applies to the intensities $E_{gijk} = \exp(Y_{gijk})$, ignores slide and dye effect by assuming normalization removed them, and is restricted to the two treatment case with no replication (though extensions have since been produced in Kendziorski et al. (2003)). Let $R_g = E_{g1}$ and $G_g = E_{g2}$, then we can create the following hierarchical model:

$$\begin{aligned} R_g &\sim \text{Gamma}(a, \theta_{Rg}) \\ G_g &\sim \text{Gamma}(a, \theta_{Gg}) \\ \theta_{Rg} &\sim \text{Gamma}(a_0, \nu) \\ \theta_{Gg} &\sim \text{Gamma}(a_0, \nu), \end{aligned} \tag{7}$$

where (a, a_0, ν) are constant hyperparameters for all genes. This model assumes that R and G have different mean and variance, but equal coefficient of variation (CV). In this model,

$$\rho = \frac{E(R)}{E(G)} = \frac{\theta_G}{\theta_R} \tag{8}$$

is the quantity of interest. If the gene is not differentially expressed, then $\rho = 1$. The shrinkage estimator for ρ

$$\hat{\rho} = \frac{R + \nu}{G + \nu} \tag{9}$$

is a compromise between the mean and mode of $p(\rho|R, G, a, a_0, \nu)$. This estimate of ρ can be used to assess evidence of differential expression.

The exchangeability in the second stage of the model implies that all genes are differentially expressed. Replacing the independent distributions of θ_R and θ_G with a mixture model where $\theta_R = \theta_G$ with probability p is more appropriate. To incorporate this a third layer is added to the model with indicators z_g were $z_g = 0$ if $\theta_R = \theta_G$ and 1 otherwise where

$$z_g \sim \text{Bernoulli}(p). \tag{10}$$

The Gamma-Gamma-Bernoulli model allows estimation of the posterior probability that $z = 1$ (a gene is differentially expressed). The posterior distribution of ρ will have a spike at 1 if $\Pr(z = 0) > 0$).

5 Maize Embryogenesis

We now introduce an experiment that uses cDNA microarrays to demonstrate some of the techniques discussed above and some recent research results. Somatic embryogenesis in *Zea mays* is an experimental important tool for genetic engineering. Natural plant development from a fertilized egg cell follows zygotic embryogenesis development into a seed and eventually a mature plant. Somatic embryos begin as callus (undifferentiated cells) and are induced to develop into embryos by immersion in an embryogenic medium. Callus can be generated from existing plants by transfer to a callus-generating medium. Mature plants can be grown from existing plant material through another experimental process called organogenesis. Somatic embryogenesis creates embryos that are similar to those arising from sexual reproduction and which have the same genotype as the explant from which they were created.

The first somatic embryos in maize tissue culture were produced by Green and Phillips (1975). Armstrong and Green (1985) found that cell lines derived from sources such as immature embryos are heterogeneous in their ability to produce somatic embryos (have cells with differing embryogenic competence) and that certain types of callus tend to be more embryogenic. Unfortunately, embryogenic competence is genotype-specific in many plant species including *Zea Mays*, and often the most desirable or economically important lines

do not easily produce somatic embryos; they are recalcitrant to regeneration. Since genetic transformation of the plant in the embryonic stage has enormous potential for development of high yielding varieties, it is important to recognize embryogenic cells or tissues and to identify genetic markers for them. In this way we hope to gain tools for improving embryogenesis in recalcitrant maize lines.

In order to identify the genetic traits responsible for highly embryogenic lines, we examined gene expression changes during maize somatic embryo development. Somatic embryos were generated identically in six callus lines (labeled A, B, C, D, E, and F) developed from immature Hi II embryo explants. These lines are assumed to be random samples from the population of Hi II lines. Hi II is a regeneration-proficient hybrid of *Zea mays* which also produces high crop yields and thus is of economic importance. After callus populations were generated from the six lines in a callus-generating medium (N6E medium +2, 4-D, 3% sucrose), embryogenic calli (identifiable by shape) were selected from the total callus for each of the lines. These selected calli were matured into somatic embryos by transferring them to a sucrose-enhanced medium (Regen Medium I -2, 4-D, 6% sucrose). After 21 days, the embryos were exposed to light and transfered to a new medium (Regen Medium II -2, 4-D, 3% sucrose) to encourage germination. Material was sampled at five time points during the development and maturation of the embryos into seedlings, see Figure 6. To reduce the interplant variability without actually losing the opportunity to estimate the line to line variance in measurements, samples from pairs of lines were pooled at each time point, so the material used later in the microarray analysis comes from the pools rather than from individual plant lines. Pools were labeled AB, CD, and EF. The dataset used in this analysis can be obtained by contacting the authors.

Gene expression patterns in the AB, CD and EF lines were profiled using 12,060 element maize cDNA arrays. The experimental design was a loop design with dye-swapping for line pools AB, CD, and EF (Kerr and Churchill 2001), which resulted in a total of 30 slides. The loop design is an efficient design in that each time point is compared to an earlier time period rather than to a reference or time zero. Thus, the design results in additional replicates of measurements at each time point without increasing the number of arrays. Under our design, gene expression is measured at each time point four times within each line pool, so that across line pools, time point samples are repeated 12 times. However, the four measurements of a line pool at one time point are taken on identical material (they are technical replicates. Therefore, the true biological replication is six plants pooled into three pools. Thus, the power of our conclusions is somewhat lessened by the use of technical replications using the same biological material. Still, this design allows for an analysis of the measurement variance across time and across line pools, see Figure 7.

We are interested in identifying the genes or groups of genes which actively participate in somatic embryogenesis. These genes will exhibit significant changes in their expression over the course of tissue development and maturation. While most of the genes participating

in embryonic development are expected to be up-regulated as embryos mature, it is also possible that some genes active in embryogenesis will down-regulate over time or will exhibit some other expression profiles. Therefore, we seek to classify the 12,060 elements of the microarrays into those with constant expression over all time points and those with any other pattern of expression. Some initial results obtained using a subset of these data are reported in (Che et al. 2005).

6 Combining Multiple Readings

As is often the case in microarray experiments, each slide was scanned multiple times using different laser and sensor settings. By varying the settings of the instruments, the operator can strike a balance between over-exposing the highly expressing genes while still picking up a signal from the lowly fluorescing spots. As we have seen, there is truncation of pixels both above (assigning over-exposed pixels the value of 65,535) and below (assigning under-exposed pixels values near or below the background level). In a canonical analysis of gene expression data, only a single scan of each slide is included in the analysis and the rest are discarded, however it has been empirically demonstrated that scans can be repeated up to around 10 times without degrading the slide (depending on the technology used) (Romualdi et al. 2003). We use an approach proposed in Love and Carriquiry (2005) that permits estimating gene expression profiles using all available measurements for each spot. Here we show that by making use of the additional information we obtain estimates of gene expression intensities that are less variable than using the 'best' scan alone. Importantly, the set of genes identified as embryogenic in our experiment changes if all scans, rather than just the best, are used in statistical analyses of the data.

The spot level intensity summaries described earlier (most often mean spot pixel minus median background pixel) display truncation in a different manner than pixel intensities. The averaging of pixels within the spot masks both types of truncation. Further, background correction shifts the observed truncation level away from the limits of the scanner. This is why the visible truncation in Figure 1 appears near 10 and 50,000 and not at the scanner limits of 0 and 65,535.

Multiple readings of the microarray slides can be taken for both fluorescence channels. Since all of the readings at different settings attempt to capture true expression levels for the genes on the slide, it is reasonable to assume that all readings contribute useful information about true expression levels and to think of combining the multiple readings into one estimate of gene expression for each spot. If the readings at different settings contain information about the true expression of the gene, then the variance in estimated gene expression that is due to the measurement process should be reduced in the estimate that is based on all available readings.

Generally, settings for different slides are very different because of the large experimental variation between slides. That is, one slide may result in a good reading at low laser and sensor settings while another may require higher settings to reduce the number of expression levels below threshold while keeping the number of overexposed spots to a minimum. Because of this practice, we are typically unable to assume that the settings act as blocks in a traditional experimental design. Since the settings to read the two channels are almost always chosen separately across slides, we can model each slide/dye combination separately. In what follows, we consider an arbitrary slide and dye channel in the experiment and demonstrate a hierarchical model for estimating gene expression levels that permits incorporating multiple measurements for each gene into a single analysis.

6.1 Bayesian Hierarchical Gamma Model

In order to estimate gene expression, we use a Bayesian hierarchical model proposed in Love and Carriquiry (2005). This model incorporates all slide scans into one estimate of expression per spot. To formulate the model, we rely on the natural ordering of slide readings. In practice, we order the slides from smallest to largest based on median reading.

Suppose that there are $m+1$ readings taken at each of n spots on a particular slide and a dye. In the maize embryogenesis experiment that we discuss here, $m+1=3$ and $n=12,060$ for all 60 slide/dye combinations, but the number of readings need not be constant over slides. For a given gene i, we use $S_{i1},\ldots,S_{i(m+1)}$ to denote the $m+1$ ordered signal measurements after background correction. We assume that all readings measure the same quantity – actual gene expression – with error. Therefore, under suitable scaling the readings would be identically distributed. We assume that the scaled readings (which are strictly positive) can be represented by a Gamma distribution. The Gamma has support on the positive real line and, depending on parameter values, exhibits noticeable skewness. We assume a constant shape parameter, a, for all genes on the slide/dye. The scale parameter for each observation S_{ij} will have two components, θ_i for the intensity due to the gene i expression and δ_j for the intensity due to the scan j settings.

We do not observe intensity of spots in readings where they are censored; however, we do know that they are censored and we also know that the measurement is larger (smaller) than a known value. We define an indicator variable, C_{ij}, where $C_{ij}=0$ if observation S_{ij} is not censored, $C_{ij}=1$ if observation S_{ij} is censored below, and $C_{ij}=2$ if observation S_{ij} is censored above. This variable and the subset of S, $S^{(o)}$, which includes non-censored measurements make up our observed data. The measurements that would have been observed in the absence of censoring are therefore taken to be missing. The set of missing data is denoted by $S^{(m)}$ and $S{=}S^{(o)}\cup S^{(m)}$. In a Bayesian framework, we can estimate missing values along with parameters.

In practice a spot can be designated as censored below if any of its pixels are less than

the background median. A spot can be designated as censored above if any of its pixels are saturated. Alternatively, exploratory data analysis can be used to decide appropriate cut-off values for a particular slide/dye combination, such as 20 and 50,000. To ensure that no gene has all of its values missing, a spot censored below in the highest scan or above in the lowest scan is not recorded as censored for that scan. We will denote the lower and upper truncation points by L and U, respectively.

Following Love and Carriquiry (2005) we now examine the conditional likelihood of S_{ij}, given the censoring indicator, C_{ij}. Let $f(\cdot|\lambda)$ be the density function of the Gamma(a, λ) distribution and $F(\cdot|\lambda)$ be its cumulative distribution function. Then censoring implies that the likelihood for $S_{ij} \in S^{(o)}$, an uncensored point, should have the following form:

$$p(S_{ij}|C_{ij} = 0) = f(S_{ij}|\theta_i\delta_j) \left(F(U|\theta_i\delta_j) - F(L|\theta_i\delta_j)\right)^{-1} I_{(L,U)}(S_{ij}), \tag{11}$$

where $I_A(\cdot)$ is the identity function on the set A. For a gene expression measurement $S_{ij} \in S^{(m)}$, which is censored below, the likelihood has the following form:

$$p(S_{ij}|C_{ij} = 1) = f(S_{ij}|\theta_i\delta_j)F(L|\theta_i\delta_j)^{-1}I_{[0,L]}(S_{ij}). \tag{12}$$

The likelihood of $S_{ij} \in S^{(m)}$ which is censored above is

$$p(S_{ij}|C_{ij} = 2) = f(S_{ij}|\theta_i\delta_j) \left(1 - F(U|\theta_i\delta_j)\right)^{-1} I_{[U,\infty)}(S_{ij}). \tag{13}$$

The restriction on the support of the likelihood will remain in the posterior distributions of the $S_{ij} \in S^{(m)}$.

This leads to the following observed data likelihood (for details, please refer to Love and Carriquiry (2005)):

$$p(S^{(o)}, C) = \prod_{(i,j)\in I_N} \Big(f(S_{ij}|\theta_i\delta_j)I_{(L,U)}(S_{ij})\Big) \times \prod_{(i,j)\in I_L} F(L|\theta_i\delta_j) \times \prod_{(i,j)\in I_U} (1 - F(U|\theta_i\delta_j)) \tag{14}$$

We complete the specification of the model by assigning prior distributions to each parameter. The conjugate prior option, while convenient from a mathematical viewpoint, is unsuitable from a biological viewpoint. We consider instead independent Gamma prior distributions for each of the $n + m$ parameters. Gamma distributions can be justified from a biological point of view because typically genes spotted on a slide exhibit low expression levels and only some of them exhibit high levels of expression. The Gamma distribution would appear to be an appropriate model for the population distribution because the expression values of the genes, estimated by a/θ_i, will be skewed. Thus

$$\theta_i \sim \Gamma(a_0, \nu) \tag{15}$$

for $i = 1, \ldots, n$. The Gamma model may also be reasonable for the strictly positive scaling parameters, so that

$$\delta_j \sim \Gamma(\alpha_1, \alpha_2) \tag{16}$$

for $j = 1, \ldots, m$. The joint Gamma prior has the form

$$p(\theta, \delta) \propto \prod_{i=1}^{n} \theta_i^{a_0} \prod_{i=1}^{m} \delta_j^{\alpha_1} e^{-\nu \sum_{i=1}^{n} \theta_i - \alpha_2 \sum_{j=1}^{m+1} \delta_j}. \tag{17}$$

The conditional posterior distributions of $\theta|\delta$ and $\delta|\theta$ are Gamma distributions under this prior, but the joint posterior of (θ, δ) is not. Therefore, the prior in (17) is a semi-conjugate prior distribution.

The hyperparameters in the model are $\eta = (a, a_0, \nu, \alpha_1, \alpha_2)$. We must either specify prior distributions for these hyperparameters or fix the parameters at some appropriate value. We have investigated several methods for choosing η in Love and Carriquiry (2005). The hyperparameters α_1 and α_2 are both chosen to be 10 to create a relatively noninformative prior on the δ's. Specifying a value for the other hyperparameters a, a_0 and ν, however, requires some thought since these parameters can have a significant effect on the estimates of expression levels. We use the Empirical Bayes estimates of the remaining hyperparameters, given the data from one scan of the slide (Carlin and Louis 1997). This has been found to be the most computationally efficient among methods that have low MSE of expression estimation in simulation studies (Love and Carriquiry 2005).

The joint posterior distribution of (δ, θ) is given by:

$$p(\delta, \theta | S^{(o)}, C, \eta) \propto \left(\prod_{j=1}^{m} \delta_j \right)^{na+\alpha_1-1} e^{-\sum_{j=1}^{m+1} \delta_j (\sum_{i=1}^{n} \theta_i S_{ij} - \alpha_2)} \left(\prod_{i=1}^{n} \theta_i \right)^{a(m+1)+a_0-1} e^{-\nu \sum_{i=1}^{n} \theta_i}.$$

The posterior distribution of $S^{(m)}$ is

$$p(S^{(m)} | S^{(o)}, C, \eta) = \int \int p(S^{(m)} | \theta, \delta, S^{(o)}, C, \eta) p(\theta, \delta | S^{(o)}, C, \eta) d\theta d\delta. \tag{18}$$

We use Markov chain Monte Carlo (MCMC) methods to approximate the joint posterior distribution of the parameters and missing values in the model (Gelman et al. 1995). To do so, we first derive the full conditional distributions for each of them:

$$\delta_j | \eta, \delta_{-j}, \theta, S, C \sim \Gamma(na + \alpha_1, \sum_{i=1}^{n} \theta_i S_{ij} + \alpha_2) \tag{19}$$

for $j = 1, \ldots, m$,

$$\theta_i | \eta, \delta, \theta_{-i}, S, C \sim \Gamma((m+1)a + a_0, S_{i(m+1)} + \sum_{j=1}^{m} \delta_j S_{ij} + \nu) \tag{20}$$

for $i = 1, \ldots, n$,

$$S_{ij} | \eta, \delta, \theta, S^{(o)}, S^{(m)}_{-ij}, C \sim \Gamma(a, \theta_i \delta_j) I_{[0,L]}(S_{ij}) \tag{21}$$

for $(i,j) \in I_L$, and

$$S_{ij}|\eta, \delta, \theta, S^{(o)}, S^{(m)}_{-ij}, C \sim \Gamma(a, \theta_i\delta_j) I_{[U,\infty)}(S_{ij}) \tag{22}$$

for $(i,j) \in I_U$. Here, sampling from the last distribution is equivalent to drawing from $\Gamma(a, \theta_i\delta_j)$ and rejecting the draw if it is less than U.

Notice that all full conditional distributions have standard form, and thus the Gibbs sampler can be used to sequentially draw parameter values from the conditionals. If we have chosen a fully Bayesian estimation of (a, a_0, ν), then the full conditionals of (a, a_0, ν), are included in the Gibbs sampler. A point estimate for the expression of the ith gene is the posterior mean of a/θ_i. These estimates may be subsequently used as the expression values for further analysis after appropriate normalization has been done.

7 Comparing Results

We now revisit the maize embryogenesis experiment that was introduced earlier. The main objective of the experiment was to determine the subset of the 12,060 genes that vary significantly in the process of somatic embryogenesis in maize. Thirty cDNA microarray slides were spotted in the course of the experiment, which resulted in 60 slide/dye combinations on which to implement one of the standard approaches as well as the new method proposed here. We first analyzed the gene expression data using background-corrected expression as the gene expression estimate. In order to make results comparable, however, we also carried out some of the analyses applying the hierarchical modeling approach proposed by Newton et al. (2001). Each of the 60 slide/dye combinations were scanned three times at different laser and sensor settings. Unequal numbers of scans per slide would not have limited the application of the procedure. Some of the results discussed below are also discussed in Love and Carriquiry (2005) in greater detail.

7.1 Single gene expression profiles

We first present a direct comparison of the results that would be obtained from a single slide if one or all available readings are used to estimate expression. Two hierarchical models were fit to one dye channel of one slide: the hierarchical model proposed here, that incorporates all readings and the hierarchical model proposed by Newton et al. (2001) that incorporates information from only one scan. In this example, the Newton model was applied to the highest available reading for the gene. We present the results obtained for two arbitrarily chosen genes. Examination of those results suggests that by combining all readings for a spot we realize several improvements.

For a particular slide, the point estimates (posterior means) for the scaling parameters δ_1, δ_2 were $\hat{\delta}_1 = 0.510$ and $\hat{\delta}_2 = 0.978$ so that scaling of the three gene expression measurements results in $1.96S_{i1} \approx 1.02S_{i2} \approx S_{i3}$. The first gene on which we focus is gene labeled #1 for which we obtained a posterior expression estimate of 213.1 based on its three

measurements of (51.3, 211.0, 227.3). When using only its highest measurement (227.3), the estimate obtained was 593.5 with a 95% credible set of (170, 1735). Notice that the gene expression estimate based on the three readings is within the 95% posterior probability interval Consider now gene #1735, for which the highest reading was censored at 0 due to very large within-spot measurement error. The posterior point estimate of expression for this gene was 108.1 based on three measurements of (59.9, 77.2, 0).When using only the measurement from the highest scan (0), the estimate of gene expression was 368.7 with a 95% credible set of (112, 1080). In this case, the estimate obtained by combining the three readings for the gene was also contained in the 95% posterior probability interval.

We argue that by combining multiple measurements into the estimate of gene expression for a single gene the resulting estimator has lower standard error. In fact, the posterior standard deviations of expression of genes #1 and #1735 in our example were 80 and 42, respectively when using the three measurements but increased to 919 and 307 when only one scan was used for estimation. Therefore, the posterior distributions of gene expression are much less concentrated around the mean when only one reading is used in estimation.

Now, we focus on inferences about gene expression profiles for individual genes and compare results that are obtained by using a standard and the proposed approaches. Here, the standard approach consists of using the observed, background-corrected expression value in the 'best' scan as the estimate of gene expression. We grouped genes according to their expression profiles over time and identified representative genes of interest from six groups (Che et al. 2005). Figure 8 shows errorbar plots for example genes from these six genes as presented in Che et al. (2005). Using all readings in the hierarchical model and the posterior mean as the expression estimate for each gene, we get the errorbar plots in Figure 9. Though the patterns remain almost unchanged, the incorporation of multiple scans changes the scale of the microarray measurement. These data have been normalized in the same way, however, the 'best' scan which was used in the initial analysis was not necessarily the highest scan of the three incorporated. Also, very high and low expressing genes have their estimated expression altered by the treatment of censoring (see the water channel protein in Figure 9).

7.2 Inferences about time and line effects

In order to identify line and time effects we fitted two-way analysis of variance models (ANOVA) to the estimated expression for each of the 12,060 genes. Recall that the power of our conclusions is somewhat lessened by the use of technical replication (See Section 5), however, this affects one and three scan analysis of the data equally. We are interested in identifying those genes for which time effects, line effects or both are statistically significant after controlling for multiple comparisons using the approach proposed by Benjamini and Hochberg (1995).

When gene expression is estimated using the observed background-corrected reading

for each gene, 1,026 genes were identified as exhibiting significantly differential expression over time. When the three slide readings per slide/dye combination were used to estimate gene expression, we found that 2,229 genes appear to have significantly different expression levels at different time points, thus indicating that these genes are regulated during somatic embryogenesis. Note that the number of genes identified as differentially expressed during somatic embryogenesis approximately tripled when all available measurements on each spot were used for analysis. This result was to be expected given the reduction in bias and RMSE in gene expression estimates that was achieved by implementing the procedure we propose. Of the 1,026 genes that were identified as embryogenic using only one scan, 933 were also included in the longer list of differentially expressed genes identified when all readings were utilized.

In Che et al. (2005) we counted the number of genes with estimated two and three-fold changes in expression relative to the day seven measurement (Che et al. 2005). By so doing we obtain an overall assessment of gene activity during somatic embryogenesis. Figure 10 shows the number of genes with different expression values at each time point relative to day seven and was obtained using gene expression estimates based on only one scan per gene. We say that a gene is up-regulated if its expression increased and that it is down-regulated if its expression value decreased. From the figure we see that in the absence of light, more of the active genes are down-regulated than up-regulated. The light was turned on on day 23 of the experiment. At days 23 and 28, many more genes become active and exhibit two- and three-fold expression changes relative to day seven. This was an expected result; light triggers photomorphogenesis, a complex biological process that is known to involve many genes. Note too that about 3% to 5% of the genes are down-regulated at days 14 and 21 (relative to day seven). The number of down-regulated genes then falls off dramatically at days 23 and 28. Upon further investigation, these genes were found to be largely histone and ribosomal protein genes, which may be downregulated as a result of a slowing down in cell proliferation and growth during embryo maturation. We carried out the same analysis using the three readings available for each slide and the hierarchical modeling approach proposed in this manuscript. We did not find noticeable differences in the conclusions that would be drawn from the one-scan or the three-scan analyses. However, there are many more genes identified as significantly down-regulated at day 23. Using gene expression estimates generated from all available readings, we obtain a larger number of significantly differentially expressed genes that follow a similar pattern of genes up- or down-regulation during the course of the experiment (see Figure 11).

8 Discussion

This manuscript is composed of two major sections. In the first section, we discuss in some detail the steps that are typically taken in the collection and pre-processing of gene expression microarray data. We also describe simple analyses of microarray gene expression data that can be implemented to identify differentially expressed genes across treatments.

In the second section of the paper, we present some newer research results that summarize what has been presented in Love and Carriquiry (2005).

Data collected in the course of microarray experimentation is subject to multiple sources of measurement error. Some of the measurement error may actually introduce biases and analysts typically attempt to reduce those biases by re-scaling and normalizing the data prior to analysis. One source of potentially significant measurement error is the settings of the instruments (laser and sensor) that are used to obtain the data. Because the 'optimal' settings may vary from slide to slide, operators often obtain multiple readings of each slide and then choose the 'best', meaning the reading that includes the fewest saturated spots and the fewest under-exposed spots.

In the first half of the manuscript we introduce some of the biological principles that form the bases of cDNA microarrays and explain how the different analytical steps introduce variability and potentially also biases in gene expression measurements that can be sometimes difficult to properly address. We address statistical issues associated to the measurement of gene expression (e.g., image segmentation, spot identification), to the correction for background fluorescence and to the normalization and re-scaling of data to remove effects of dye, print-tip and others on expression. Because normalization has received some attention in the literature, we discuss in some detail the more commonly used approaches to partially account for systematic yet extraneous effects on gene expression data. In this section of the manuscript we also describe the standard statistical approaches for estimating treatment effect on gene expression, and briefly address the multiple comparisons problem, often referred to as the big p small n paradox.

The use of multiple scans obtained under the same laser and sensor settings have been proposed as a means to reduce the variability of gene expression estimates (Romualdi et al. 2003). Yet improving homogeneity of spots and accounting for the purely random measurement error should be possible using effective segmentation and background cleaning methods. It has been only recently that some attention has been focused on analytical methods that might permit incorporating multiple slide scans obtained under different measurement conditions into statistical analyses. Several approaches have been proposed in the literature for doing so (Dudley et al., 2002; Lyng et al., 2004; Garcia de la Nava et al., 2004). In the second half of this manuscript, we describe a general hierarchical modeling approach proposed by Love and Carriquiry (2005) that enables use of all the readings available for each slide for analyses, even if the number of readings per slide vary across slides. The basic premise is that each reading of a spot contains some information about the true expression of the gene and that if an appropriate scaling factor for each spot can be estimated, then all readings for a spot estimate the same quantity and can be combined. If so, then it is to be expected that the estimate of gene expression will have smaller variance than it would have if based on a single spot measurement. To determine whether the modeling approach we propose results in estimators of gene expression with good statistical properties, Love and

Carriquiry (2005) ran a simulation study and assessed bias and root mean squared error of the estimators over repeated sampling. They found that the hierarchical modeling approach they propose had smaller bias and smaller RMSE than all other estimators, suggesting that basing estimation on as many readings for each spot as might be available is probably a reasonable idea.

To illustrate the proposed approach, we applied it to microarray data collected in a maize embryogenesis experiment carried out by scientists in the Plant Sciences Institute at Iowa State University. The complete analyses of these data is presented elsewhere (Che et al., 2005). We show only a subset of the results here, to highlight some of the benefits that appear to accrue when using the three scans available for each slide. When comparing our results to those obtained from fitting the Newton (2001) hierarchical model using only one reading per slide, we note that the variance of expression estimates is lower when based on three readings, as would be expected. We also notice that expression levels are not as shrunken toward the mean expression (2594). Because of the smaller bias and RMSE in gene expression estimates, inferences about the set of genes involved in somatic embryogenesis in maize change drastically when statistical analyses are based on one or on three readings of each slide. As might be expected, the power of tests increases as the RMSE in gene expression estimates decreases which in turn results in more precise time and biological line pool effects. As Skibbe, Nettleton, and Schnable (2004) pointed out, conclusions drawn about differential expression can be dependent on the slide scan used. Here we see that stronger conclusions are possible using all available scans than using only one.

9 Acknowledgments

We would like to thank Kansas State University for inviting us to present our work at the Conference on Statistics Applied to Agriculture. Some of the figures used in the oral presentation were provided by Dan Nettleton and were part of the course notes for his excellent new course on the design and analysis of microarray experiments at Iowa State University. We also thank the following colleagues at Iowa State University, who provided the experimental maize data: Steve Howell and Ping Che of the Plant Sciences Institute and Kan Wang and Bronwyn Frame of the Center for Plant Transformation and the Department of Agronomy. While working on this research, Tanzy Love was partially funded through the NSF grant NSF-DMS#0091953.

References

Armstrong, C.L. and Green, C.E. (1985). Establishment and maintenance of friable embryogenic maize callus and the involvement of L-proline. *Planta 164*, 207–214

Benjamini, V. and Hochberg, V. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B: Methodological 57*, 289–300

Benjamini, V. and Yiekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics 29*, 1165–1188

Carlin, B.P. and Louis, T. (1997). *Bayes and Empirical Bayes*. Chapman & Hall, London, United Kingdom

Che, P., Love, T.M., Frame, B.R., Wang, K., Carriquiry, A.L., and Howell, S.H. (2005). Gene expression program during maturation and germination of somatic embryos in maize cultures. *Plant Molecular Biology*. In press.

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

Dudley, A.M., Aach, J., Steffen, M.A., and Church, G.M. (2002). Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *PNAS 99*(11), 7554–7559

Garcia de la Nava, J., van Hijum, S., and Trelles, O. (2004). Saturation and Quantization Reduction in Microarray Experiments using Two Scans at Different Sensitivities. *Statistical Applications in Genetics and Molecular Biology 3*(1), article 11.

Dudoit, S., Shaffer, J.P., and Boldrick, J.C. (2002) Multiple hypothesis testing in microarray experiments. Division of Biostatistics, University of California at Berkeley. Technical Report 110.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (1995). *Bayesian Data Analysis* Chapman & Hall, London

Green, C.E., and Phillips, R.L. (1975). Plant regeneration from tissue cultures of maize. *Crop Science 15*, 417–420

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

Kerr, M.K., and Churchill, G.A. (2001). Experimental Design for Gene Expression Microarrays. *Biostatistics 2*, 183–201

Lönnstedt, I. and Speed, T.P. (2001). Replicated microarray data. *Statistica Sinica 12* 31–46

Love, T.M. and Carriquiry, A. (2005). Incorporating Multiple cDNA Microarray Slide Scans -Application to Somatic Embryogenesis in Maize. submitted to *The Journal of the American Statistical Association*.

Lyng, H., Badiee, A., Svendsrud, D.H., Hovig, E., Myklebost, O., and Stokke, T. (2004). Profound influence of microarray scanner characteristics on gene expression ratios: analysis and procedure for correction. *BMC Genomics 5*:10.

Newton, M.A., Kendziorski, C.M., Richmond, C.S., Blattner, F.R., and Tsui, K.W. (2001). On Differential Variability of Expression Ratios: Improving Statistical Inference about Gene Expression Changes from Microarray Data. *Journal of Computational Biology*, *8*(1), 37–52.

Romualdi, C., Trevisan, S., Celegato, B., Costa, G., and Lanfranchi, G. (2003). Improved detection of differentially expressed genes in microarray experiments through multiple scanning and image integration. *Nucleic Acids Research 31*(23) e149

Skibbe, D.S., Nettleton, D., and Schnable, P.S. (2004). Scanning microarrays at multiple intensities increases the number of statistically significant differences detected. *Unpublished Manuscript*

Smyth, G.K., Yang, Y.H., and Speed, T. (2002). Statistical Issues in cDNA Microarray Data Analysis. *Functional Genomics:Methods and Protocols* Humana Press, Totowa, NJ.

Speed, T. (ed.) (2003). *Statistical Analysis of Gene Expression Microarray Data*. Chapman & Hall/CRC Press, Boca Raton, Florida.

Storey, J.D. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B (64)* 474–498

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

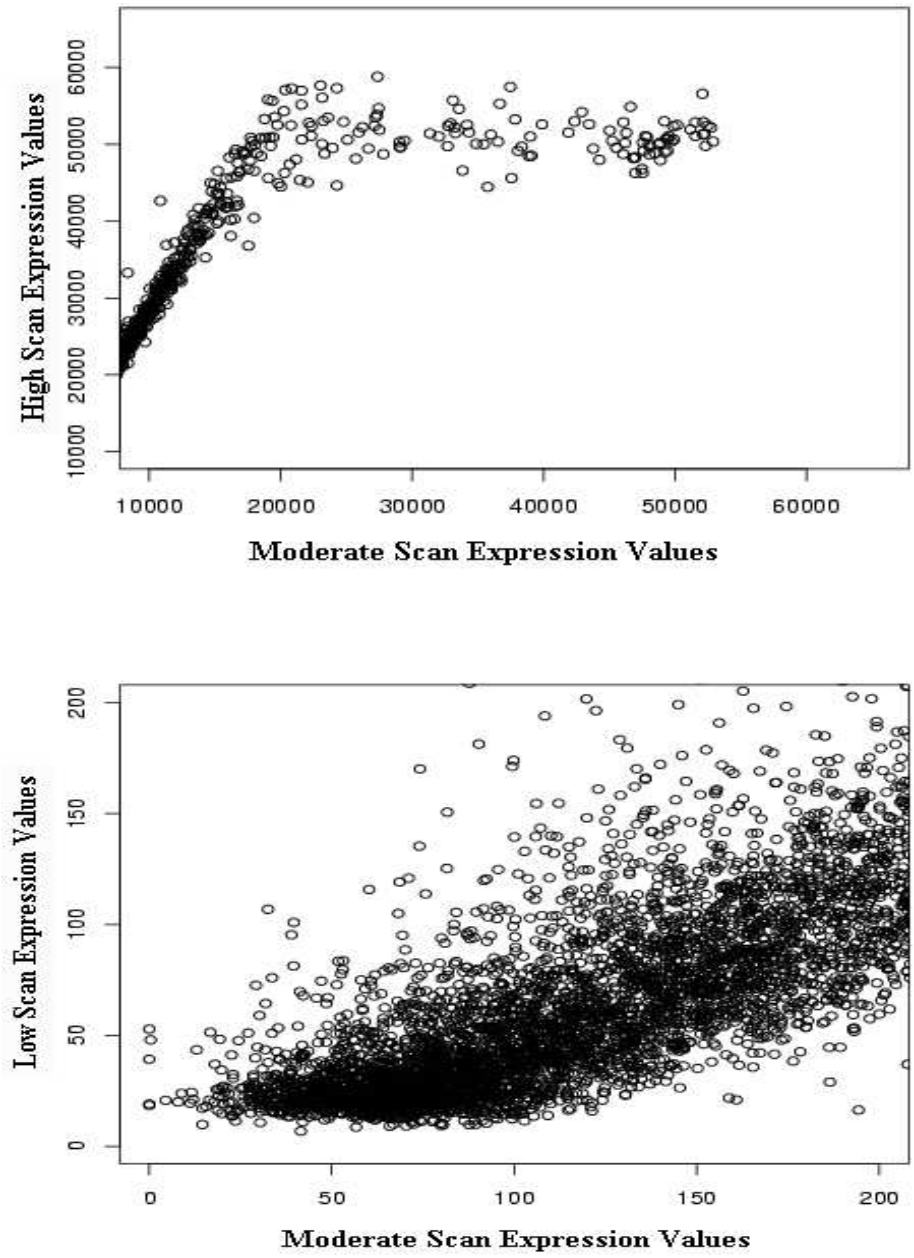

Figure 1: Examples of background corrected spot averages (a) saturated in a high scan and (b) censored below in a low scan. The few negative background correct average values have been set to zero.

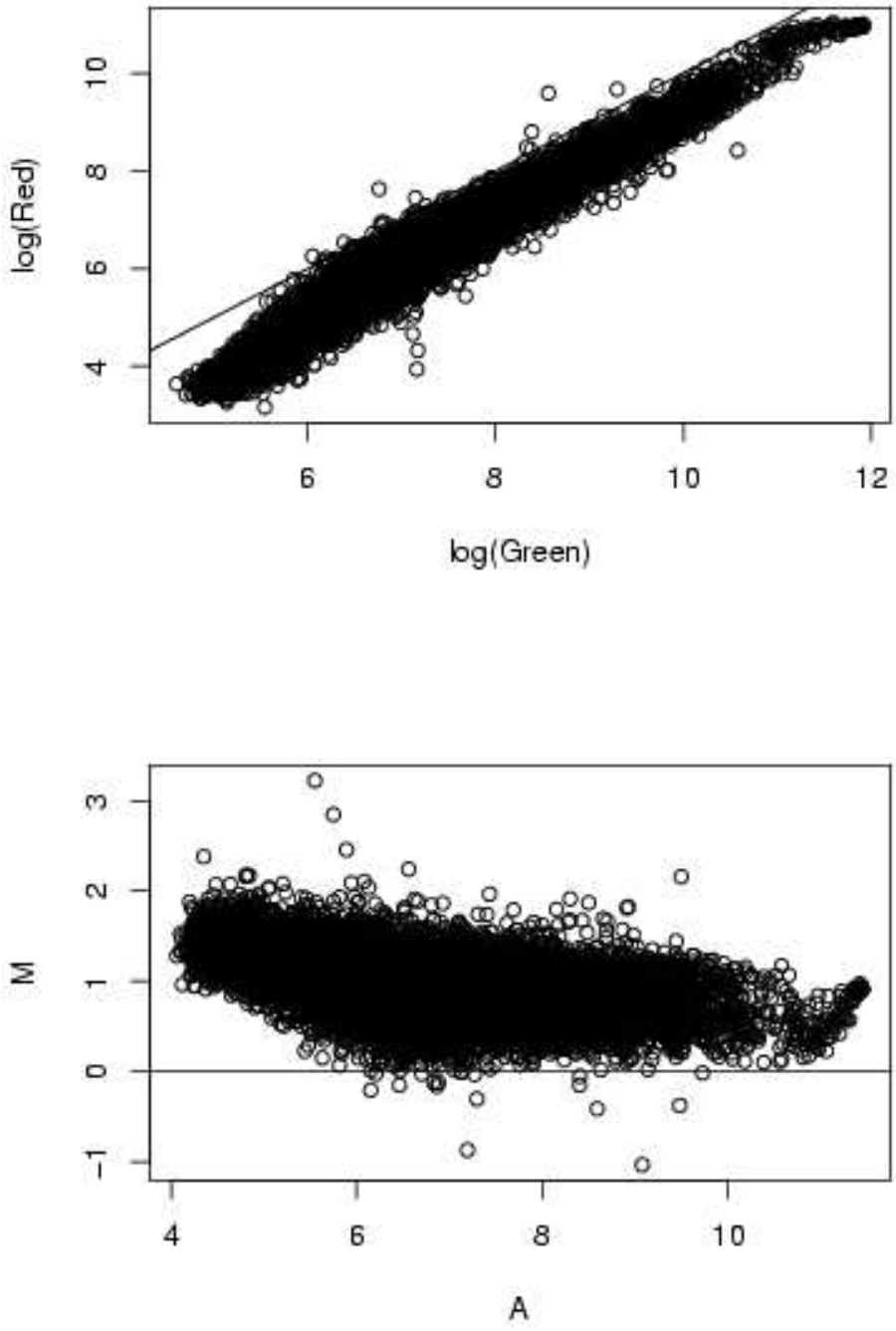

Figure 2: The log intensities for the red and green channels of spots and the M versus A plot for spot intensities on one slide.

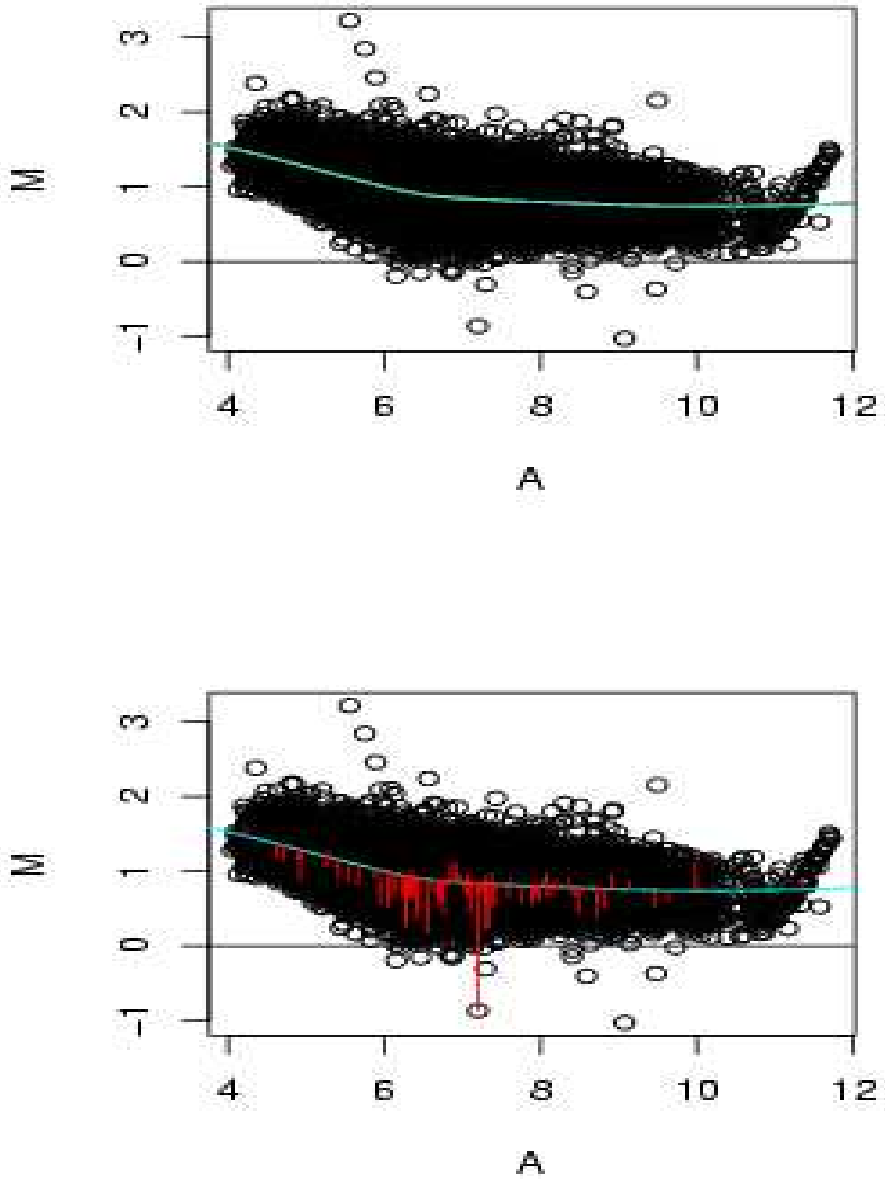

Figure 3: The loess fit of M as a function of A. The normalized values of M are the residuals from this line. One hundred residuals are shown. Note the outlier gene near 7 on the A axis.

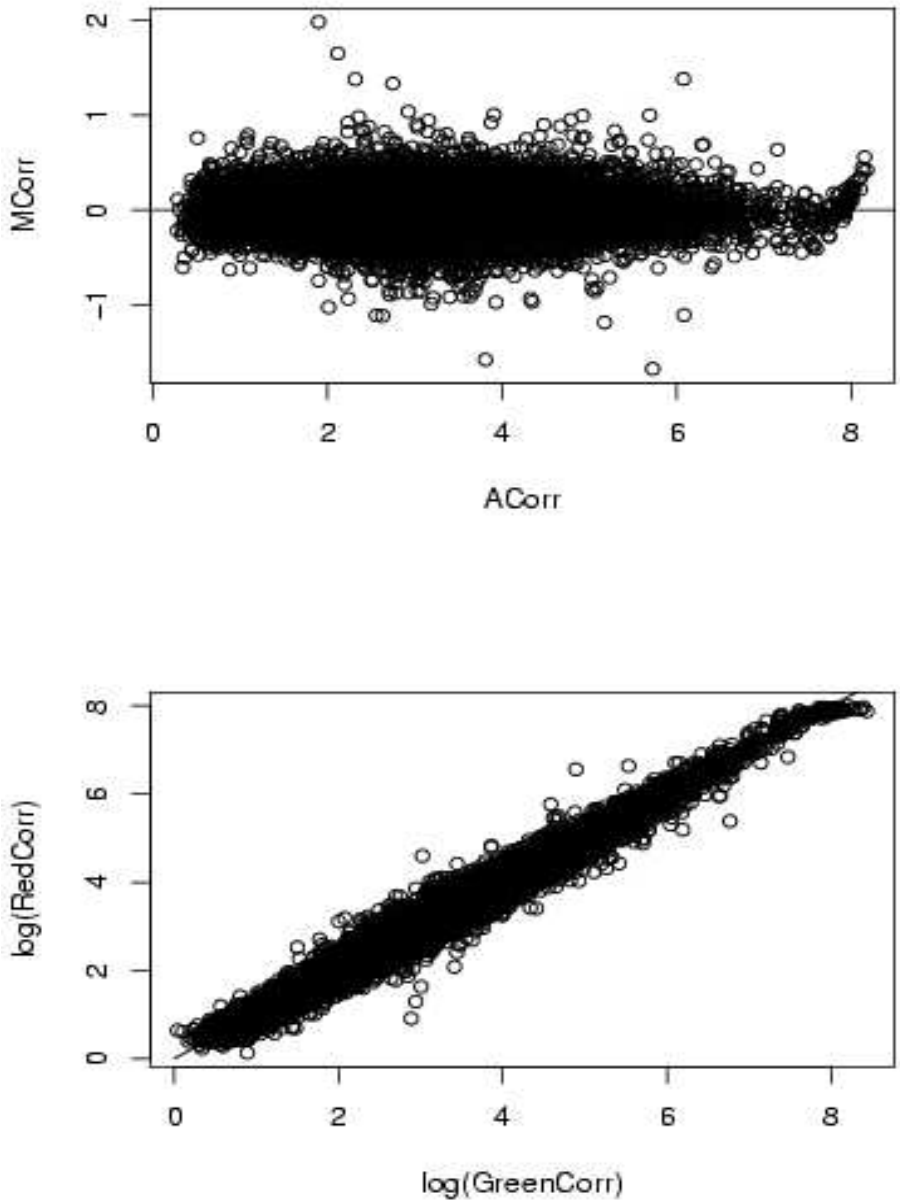

Figure 4: The corrected values for spot intensities on one slide. Note that the dye bias and intensity dependence have been removed in the corrected values.

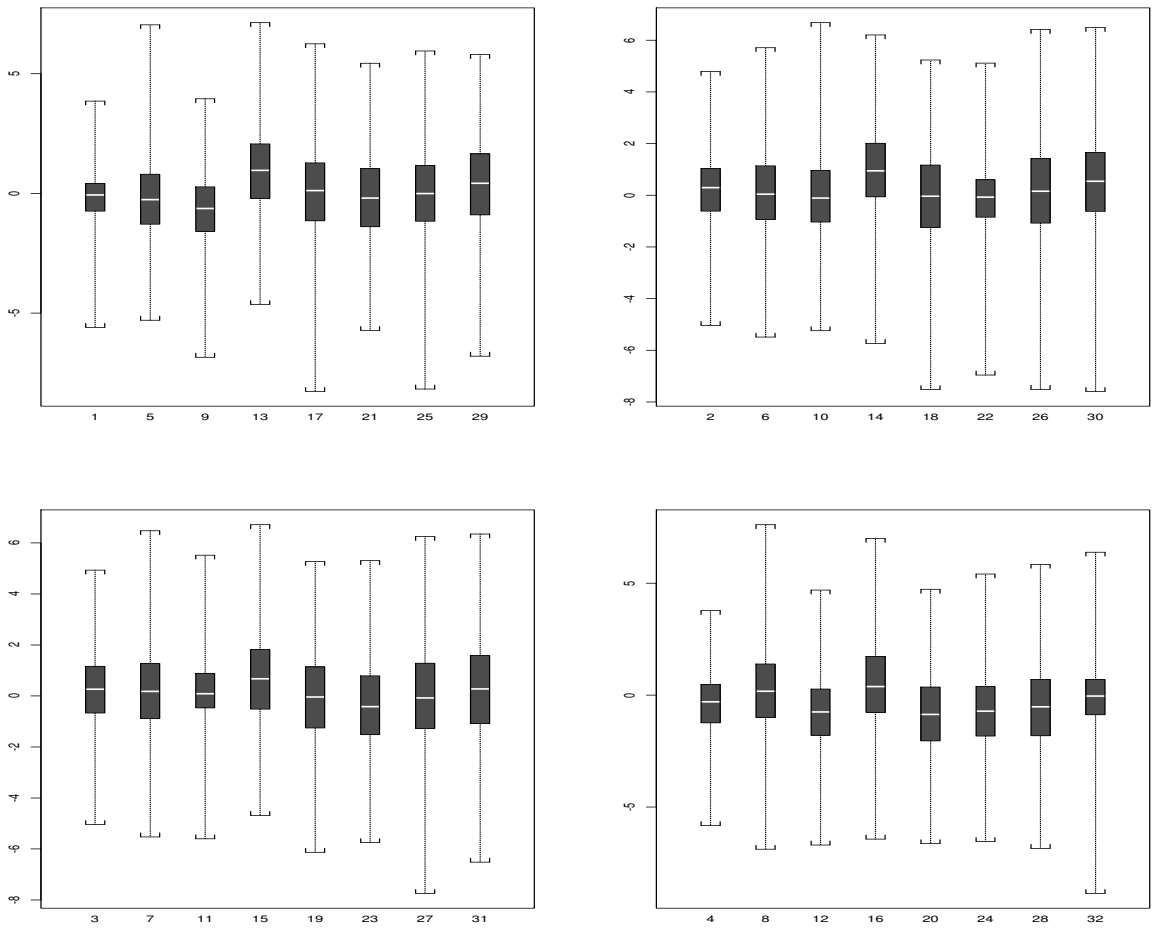

Figure 5: Boxplots for M_i values for each of 32 print tip groups on a slide, grouped by the four metarows on the slide.

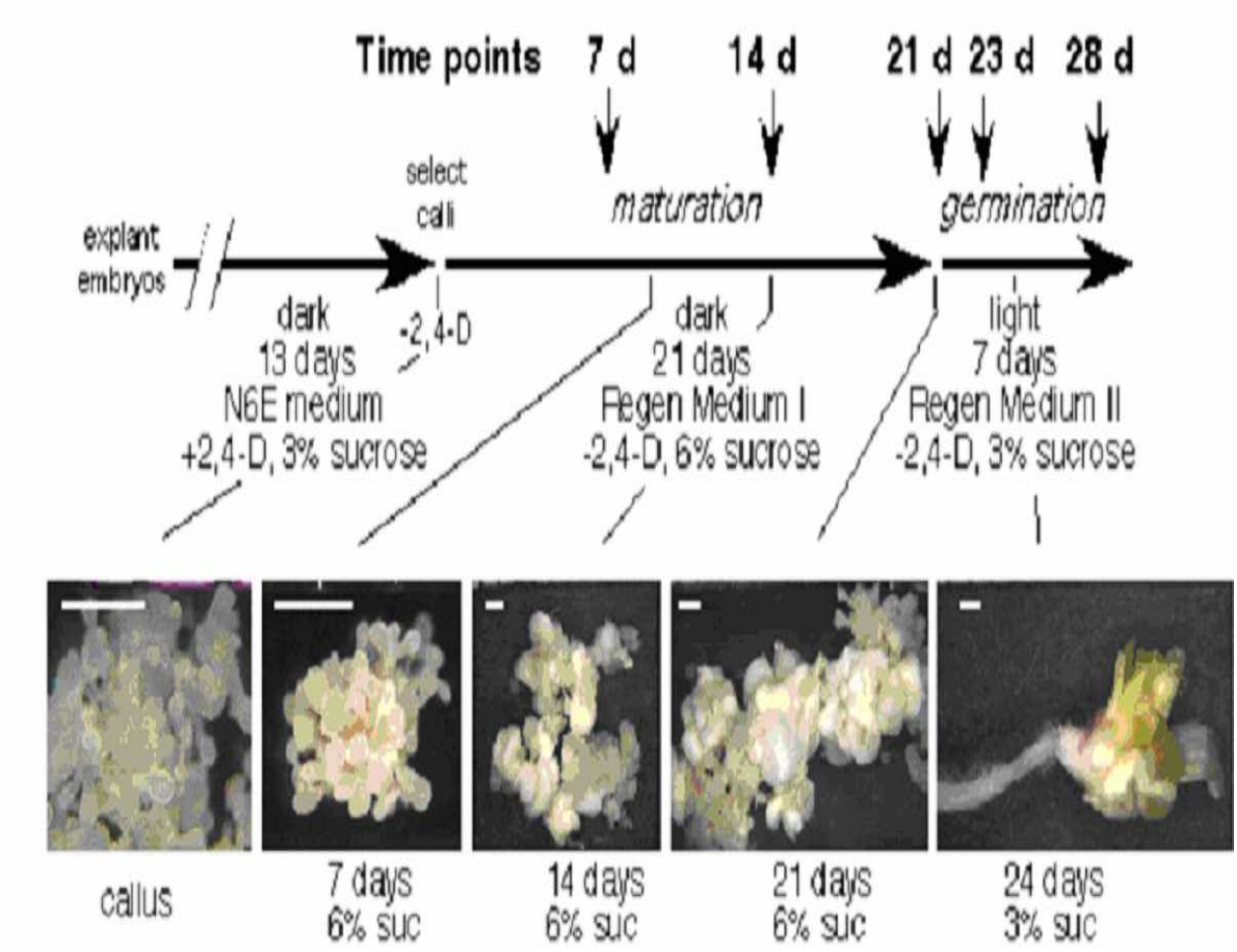

Figure 6: The time course experiment for somatic embryogenesis in maize.

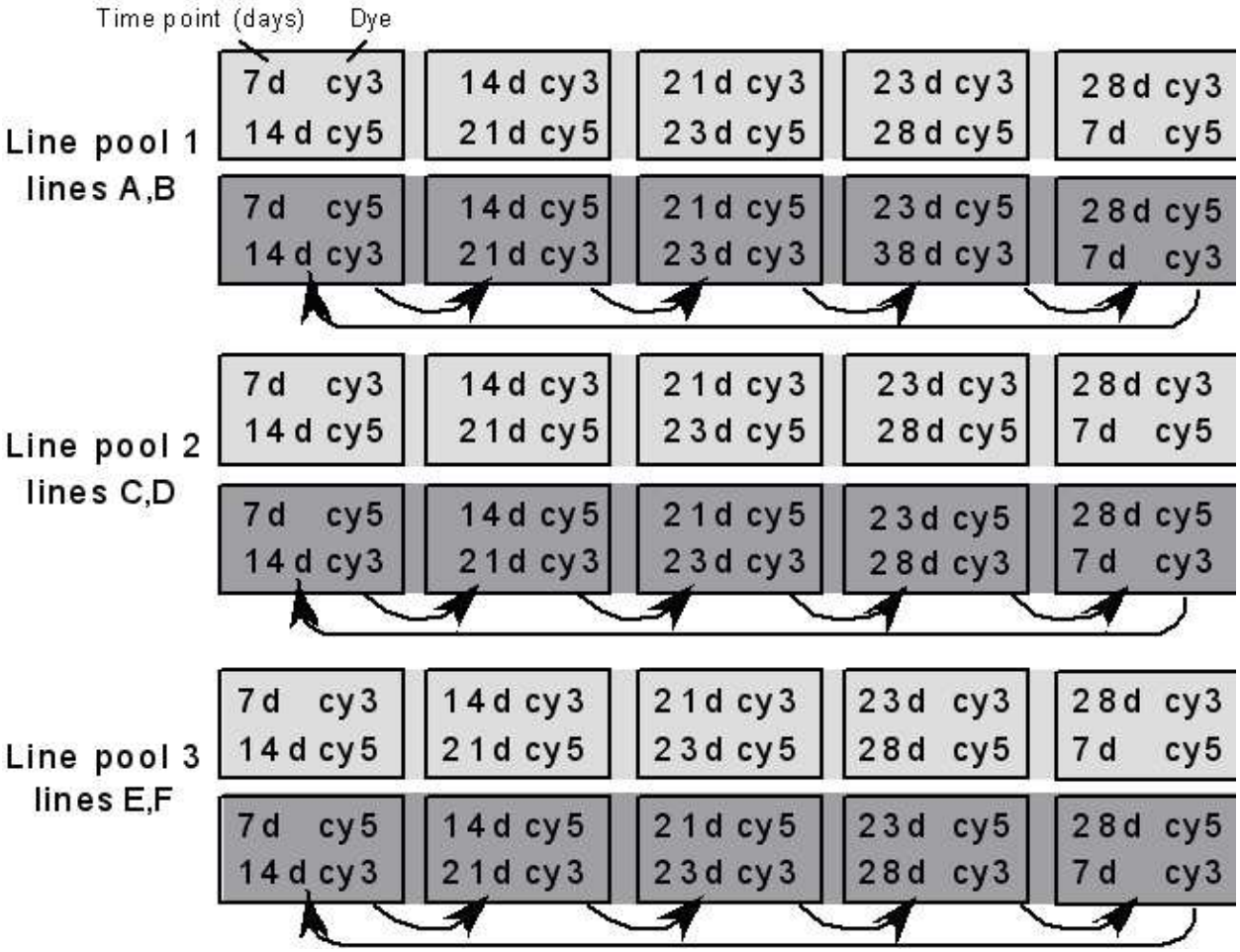

Figure 7: The microarray double loop design with dye swap for the maize embryogenesis experiment. Each box represents one of the 30 slides created. The arrows show the direction of the loop as each time point is compared to its neighbors.

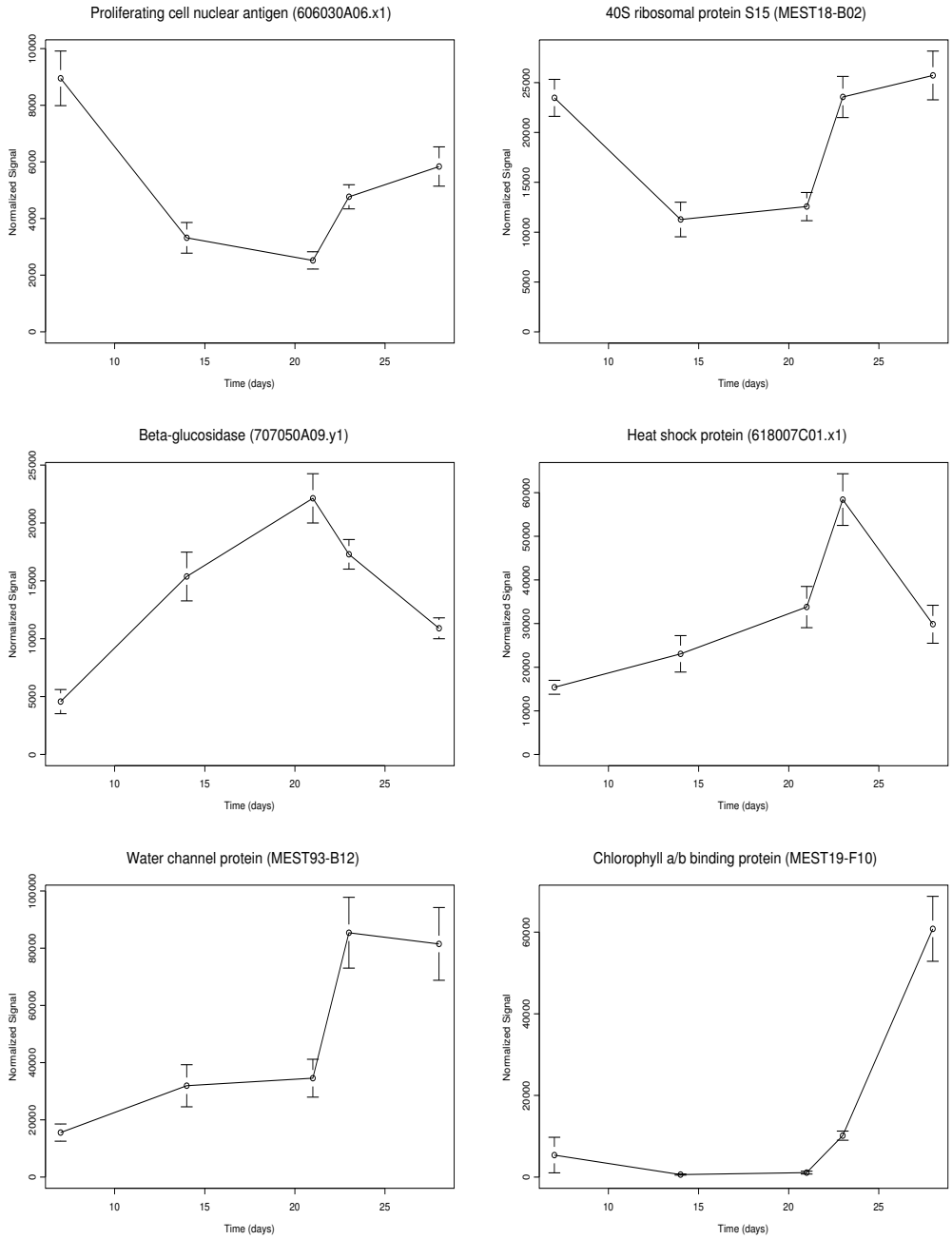

Figure 8: Errorbar plots for some example genes from Che et al. (2005) (Fig 4).

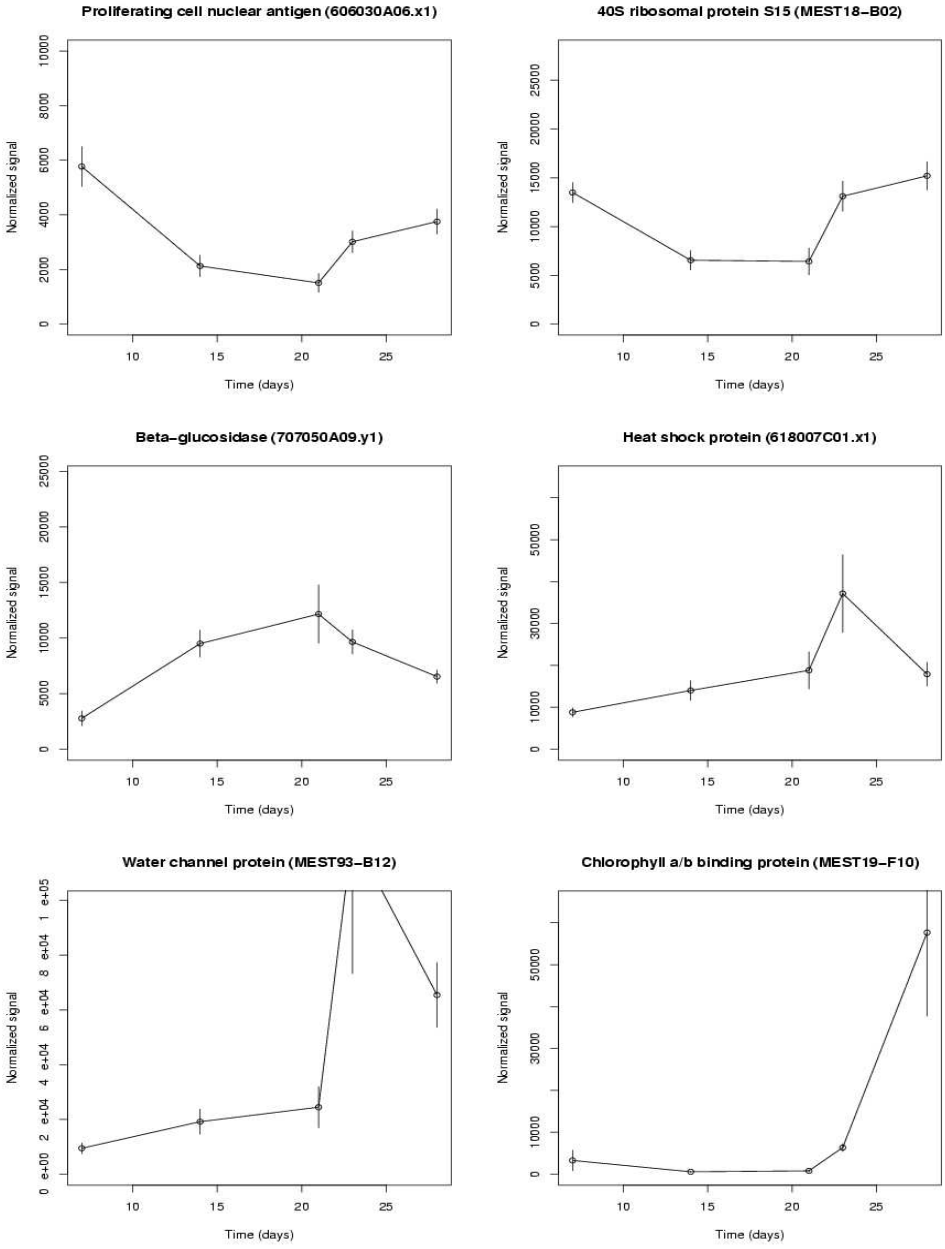

Figure 9: Errorbar plots for the example genes from Che et al. (2005) (Fig 4) after using all readings.

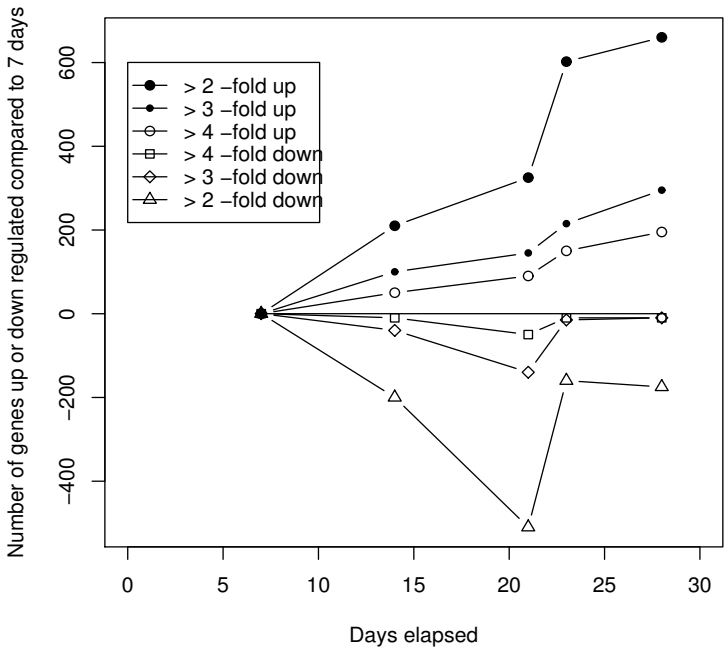

Figure 10: Numbers of genes with fold changes in expression from Che et al. (2005)(Fig 3).

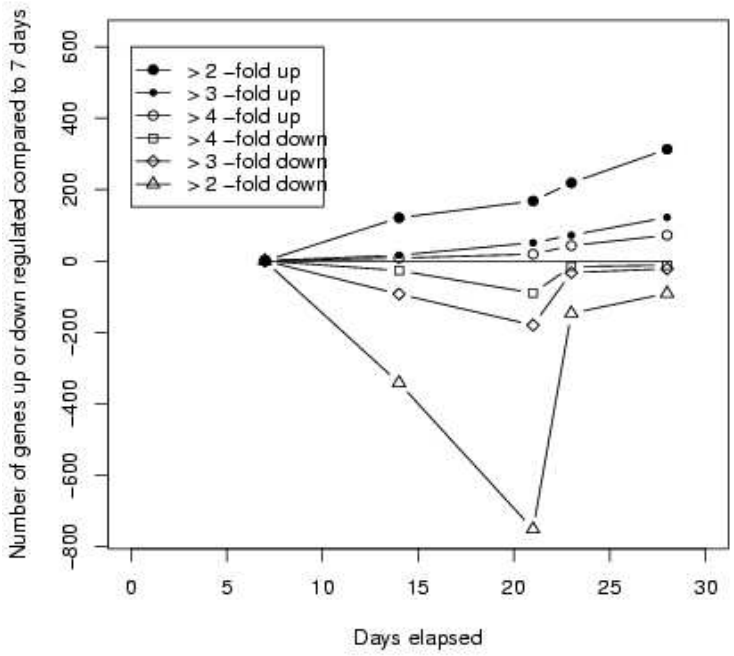

Figure 11: Numbers of genes with fold changes in expression from Che et al. (2005)(Fig 3) after using all readings.