

Kansas State University Libraries

New Prairie Press

---

Conference on Applied Statistics in Agriculture

2001 - 13th Annual Conference Proceedings

---

## DESIGNING MICROARRAYS

B. A. Craig

O. Vitek

M. A. Black

M. Tanurdzic

R. W. Doerge

*See next page for additional authors*

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Craig, B. A.; Vitek, O.; Black, M. A.; Tanurdzic, M.; and Doerge, R. W. (2001). "DESIGNING MICROARRAYS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1223>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

---

**Author Information**

B. A. Craig, O. Vitek, M. A. Black, M. Tanurdzic, and R. W. Doerge

## DESIGNING MICROARRAYS

B.A. Craig<sup>1,2</sup>, O. Vitek<sup>1,2</sup>, M.A. Black<sup>1,2</sup>, M. Tanurdzic<sup>3</sup> & R.W. Doerge<sup>1,2,4</sup><sup>1</sup>Department of Statistics, Purdue University, West Lafayette, IN 47907<sup>2</sup>Computational Genomics, Purdue University, West Lafayette, IN 47907<sup>3</sup>Purdue Genetics Program, Purdue University, West Lafayette, IN 47907<sup>4</sup>Department of Agronomy, Purdue University, West Lafayette, IN 47907**Abstract**

Microarrays allow the simultaneous assessment of expression levels for thousands of genes across various treatment conditions and time. It has been shown that in these experiments expression levels can also be affected by factors in the printing of the slide, in the hybridization process, and in the post-hybridization process. Recently, variations of the incomplete block design were proposed as a way to avoid confounding the expression levels of interest with several of these nuisance factors. In this paper, we propose additional design procedures to remove factors that contribute to the spatial variability on a slide. This approach requires the use of replication, and involves designing how the replicates are to be positioned on the slide. We demonstrate these techniques using an experiment involving sixty-four genes, four replicates per slide, and five treatment conditions.

# 1 Introduction

## 1.1 Overview of microarray technology

The role of an organism's genes is to regulate cell behavior by the production of proteins. Determining which genes actively manufacture (i.e., express) proteins under various conditions, or at certain times, is one of the fundamental questions in genetics, and is the focus of what are termed *gene expression* experiments. Such studies seek to determine when a gene is active by measuring the amount of messenger RNA (mRNA) being produced by that gene, with the amount of mRNA present strongly associated with the level of protein expressed. For many years now, it has been possible to estimate the expression level for a single gene by measuring its mRNA production. However, the advent of DNA microarray technology (Schena *et al.*, 1996) has made such studies possible on a much larger scale.

The use of microarrays allows researchers to obtain data from thousands of genes under various treatment conditions, and/or over time, in a single experiment. This wealth of data has led to the proposal of many different analytic approaches, with an early preference for clustering techniques (Eisen *et al.*, 1998; Tamayo *et al.*, 1999; Holter *et al.*, 2000) gradually giving way to methods for determining significant differential gene expression across treatments (Chen *et al.*, 1997; Lee *et al.*, 2000; Kerr, Martin, and Churchill, 2000; Newton *et al.*, 2001; Kerr and Churchill, 2001). Although two popular technologies currently exist (the Affymetrix oligonucleotide array, and the cDNA array developed at Stanford), the focus here is on the cDNA microarray, since it is the most customizable, and thus best suited to the application of experimental design.

A cDNA microarray consists of a collection of *spots* arranged in rows and columns on a glass slide (i.e., microarray). Each spot is comprised of thousands of identical segments of

complementary DNA (cDNA) representing the sequence of a single gene. The basic form of a cDNA microarray experiment involves investigating expression levels for genes under two different treatment conditions (e.g., a treatment and a control). To accomplish this, mRNA is extracted from cells under each condition, and labeled with fluorescent tags. These tags are often a green (*Cy3*) label for the control mRNA, and a red (*Cy5*) label for the treatment mRNA. The labeled mRNA is then mixed together and placed on the microarray to allow *hybridization* to occur (i.e., the mRNA bonding to its complementary cDNA sequence). After hybridization the array undergoes a washing process in order to remove any unbound mRNA from the slide.

By knowing the identity of each spot (i.e., knowing the gene from which the sequence at that spot was derived), the presence of hybridized mRNA quantifies expression by that particular gene. To determine the level of expression recorded by each spot, a multi-frequency laser scanner is used to excite the fluorescent labels. Scanning produces a pair of image files, one for each color of label (often these are referred to as *channels*), which are then processed by a software package to quantify the intensity of the fluorescent signal. By comparing the green and red intensities at each spot, changes in intensities between the treatment conditions can be observed for individual genes. These changes are referred to as *differential expression*.

An obvious statistical hurdle encountered in this situation lies in determining whether an observed fold change in expression is “real” (i.e., significant), or whether it is a result of the inherent variability of the experimental process. The estimation of this inherent variability, therefore, is a key component in any statistical analysis that is proposed. This estimation also includes adjusting for, or separating out, known sources of variability in the process. The first such separation generally involves subtracting a local background intensity from each spot to remove the effect of non-hybridization factors such as glare from the slide. A

second correction is usually made via a data *normalization* procedure, which attempts to standardize the data to allow comparisons between channels on an array, as well as across multiple arrays. This standardization across channels, however, is not straightforward, with a number of different techniques available, each relying on various assumptions (Hegde *et al.*, 2000).

Various statistical procedures have been proposed for the analysis of microarray data from experiments involving multiple treatments (and/or time points) conducted across multiple slides (Chen *et al.*, 1997; Lee *et al.*, 2000; Newton *et al.*, 2001). Most of these approaches assume that the data have been normalized prior to analysis, and that this normalization allows valid comparisons to be made across all these situations. This is a rather strong assumption, since it requires that relatively simple normalization procedures (some of which involve major assumptions) are able to remove all experimental effects from the data, *except* those resulting from differences in mRNA expression. In an effort to overcome the difficulties of data normalization and analysis, a linear models approach to the design and analysis of microarray experiments has been proposed (Kerr, Martin, and Churchill, 2000; Kerr and Churchill, 2001). This approach advocates identifying factors which can introduce variability into the data, and designing microarray experiments which allow the statistical estimation of these factors. The benefit to incorporating these potential sources of variation into an analysis of variance (ANOVA) model is that data analysis can be performed without having to normalize the data *a priori*. The use of such models in conjunction with appropriate experimental designs means that the analysis can maximize the amount of information available from the microarray experiment.

The purpose of this work is to extend the design and analysis of microarray experiments to include sources of variation related to the construction of each slide in an experiment.

The sources are 1) the effect of using different pins (on a multiple pin printing tool) to print the spots on the slide, 2) the effect of varying amounts of genetic material being retrieved by the pins on subsequent visits to the printing tool template, and 3) the effect of washing the excess genetic material from the slide. The incorporation of these terms is possible because each gene is replicated (i.e., each gene is located at more than one spot) and because the printing tool template (i.e., the layout of the slide) is designed to avoid confounding these effects with specific genes. The results of a pilot study are presented to assess the importance of each of these effects and the results of this study are used to demonstrate the improvement in performance of this new model (i.e., power) over the basic model of Kerr, Martin, and Churchill (2000), as well as the importance of using replicate spots in microarray experiments (Lee *et al.*, 2000; Black and Doerge, 2001). The design of the printing tool template is discussed in reference to a study of retrotransposon expression involving five treatments.

## 1.2 The retrotransposon experiment

The experiment under consideration is designed to study the activity of retrotransposons in maize, specifically what stress mechanisms activate retrotransposons and which retrotransposons act together. In short, a retrotransposon is a transposable element whose transposition involves a process of reverse transcription with an RNA intermediate similar to that of a retrovirus. Experimental factors include the age of the plant, the location of the cells within the plant, and the stress mechanism. The levels of each of the factors are presented in Table 1. For our purposes, we simplify the experiment and consider the age of the plant and the location of the cells within the plant fixed, and focus solely on the five stress conditions. A total of sixty-four retrotransposons were selected for the experiment. By construction

Table 1: Experimental layout under consideration for the study of retrotransposons in maize.

Stress	One week old plants			Two week old plants		
	Roots	Leaves	Root Tips	Roots	Leaves	Root Tips
A. Heat	x	x	x	x	x	x
B. Freeze	x	x	x	x	x	x
C. Flood	x	x	x	x	x	x
D. Drought	x	x	x	x	x	x
E. Control	x	x	x	x	x	x

each slide contains 256 spots, therefore four replicates of each retrotransposon sequence are possible. For simplicity the term “gene” will be used in place of “retrotransposon” for the remainder of this paper.

## 2 Design of Microarray Experiments

Microarray experiments typically have mRNA from two treatment conditions on each array. Thus, if an experiment involves more than two treatments, multiple arrays are required. If the arrays are regarded as blocks in an experimental design setting, this suggests an incomplete block approach to the analysis. Kerr and Churchill (2001) proposed a “loop” design, which is a cyclic block design with block size of two (David, 1963), so as to achieve balance between the effect of using different labels for the various treatments (i.e., dye swapping). The simplest “loop” design has  $t$  slides for  $t$  treatments and each treatment appears in two blocks. This results in a balanced incomplete block design when  $t = 3$  and a partially balanced block design when  $t \geq 4$ . With five different stress conditions being considered in this study, the “loop” design would consist of five slides (Table 2).

If both dye and slide are used as blocking factors in a simple “loop” design, there are no



Table 2: Loop design layout for a microarray experiment involving five (A-E) treatments.

Slide	Dye	
	Red	Green
1	A	B
2	B	C
3	C	D
4	D	E
5	E	A

remaining degrees of freedom to test for treatment differences. However, since the goal of a microarray experiment is to compare the simultaneous behavior of numerous genes across treatments, the interest is in the treatment by gene effects present in the subplot portion of the experiment. It is also important to note that by cycling through the treatments, each treatment is labeled with both dyes. This balance means potential dye by gene effects are not confounded with treatment by gene effects.

The linear model is

$$\log(y_{ijklr}) = \mu + S_i + T_j + C_k + G_l + SG_{il} + TG_{jl} + CG_{kl} + \epsilon_{ijklr}$$

where  $y_{ijklr}$  is the background corrected intensity,  $S_i$  is the  $i$ th slide effect,  $T_j$  is the  $j$ th treatment effect,  $C_k$  is the  $k$ th dye effect and  $G_l$  is the  $l$ th gene effect. Fitting all three interaction terms and error is only possible when there are replicate spots of each gene. Kerr, Martin, and Churchill (2000) didn't consider replicates so the  $SG$  term was used as error. Additionally, Kerr, Martin, and Churchill (2000) assume all terms to be fixed, while we consider slide to be random. The use of a mixed model has been suggested by Wolfinger *et al.* (to appear). One of the key benefits of this approach is the recovery of interblock information (Cochran and Cox, 1957), which can be quite substantial when the block size is

two.

## 2.1 Comparisons of Interest

We first present standard errors for two hypotheses of interest under the fixed effects model. While the mixed effects model standard errors will be different due to the slide and slide by gene interaction variances, the general pattern that the standard error depends on the distance between the two treatments still holds. Thus, the fixed effects standard errors can be useful in designing an experiment. Because of the incomplete structure of the “loop” design ( $t \geq 3$ ), the slide effects are partially confounded with the treatment effects and the treatment by gene effects are partially confounded with the slide by gene effects. Because there is general interest in comparisons that involve the treatment by gene effects, the standard error depends on whether the slide by gene interaction is included in the model.

The first comparison we consider assesses whether there is significant differential expression of a particular gene across two treatments,  $H_0 : \mu_{jl} = \mu_{j'l}$ , where  $\mu_{jl}$  is the mean log intensity for treatment  $j$  and gene  $l$ . If the slide by gene interaction term is not included in the model, the standard error of this comparison is

$$\sqrt{(t(g + 1 + 2d) - 2(d + 1)^2) \sigma^2 / tgr}, \quad (1)$$

where  $t$  is the number of treatments,  $g$  is the number of genes,  $r$  is the number of gene replicates per slide and  $d$  is minimum number of slides the two treatments are apart. Referring back to Table 2, Treatment A is one slide away from treatment C (and D) but appears on the same slide as B (and E). Thus, the standard error for a comparison of a particular gene between Treatment A and B (or E) will be smaller than a comparison between Treatment A and C (or D).

When the slide by gene interaction is included in the model, the standard error is

$$\sqrt{2(d+1)(t-d-1)\sigma^2/tr}. \quad (2)$$

Again, the standard error is smaller for those treatments that are closer together in the loop. In the former situation (Eq. 1), the difference in standard errors is reduced as the number of genes increases. Thus, for large experiments involving thousands of genes, the placement of the treatments in the loop has little consequence. However, in the latter situation (Eq. 2), the number of genes does not factor into the standard error. Thus, if an experimenter were designing a “loop” design and the slide by gene interaction were to be included, it would be beneficial to place those treatments that are of more interest to compare closer together in the design. If all the comparisons take an equal weight and  $t$  is fairly large, it would be beneficial to consider other sorts of incomplete block designs.

In some situations, a particular treatment could lower/increase the expression level in all the genes and a pairwise comparison would find differences that may not be of interest. An alternative comparison would be to see if two gene expression levels behave in a similar manner across treatments. In other words, testing for the pairwise interaction,  $H_0 : \mu_{jl} - \mu_{j'l'} = \mu_{j'l} - \mu_{j'l'}$ . If there is no slide by gene interaction, the standard error is constant and simply  $\sqrt{2\sigma^2/r}$ . When the interaction is included, the variance is twice that when comparing just a single gene (Eq. 2) and the standard error is  $\sqrt{4(d+1)(t-d-1)\sigma^2/tr}$ . Again with this hypothesis, if the  $SG$  term is included in the model, it is important to consider the placement of the treatments within the loop.

## 3 The Printing and Washing Process

### 3.1 Array construction

Most experimental design for microarrays has considered the normalization effects but not the effects that occur during the printing and washing of a slide. Several recent papers have considered a spot effect to account for the spot-to-spot (i.e., spatial) variability inherent in microarray slides. Dudoit *et al.* (2000) adjust for this variability by considering a nonlinear smoother for each pin. We take a more systematic approach towards understanding the effects that cause this spot-to-spot variability and use this information to determine how the gene replicates should be placed on the slide to avoid confounding these effects with the treatment by gene effects.

The general cDNA microarray printing process involves a multi-pin printing tool and a printing tool template, which contains a collection of wells, each well containing genetic material which represents the sequence of a particular gene. When a solid pin printing tool is used, it first dips into a collection of template wells, each pin into a separate well, and collects genetic material on each pin tip. The printing tool then moves to the slide and prints this genetic material on the slide. The printing tool is washed off and the process is repeated for a different set of wells. The sets of template wells and order of printing depends on the particular printing equipment.

For this experiment, a  $4 \times 4$  printing tool is used. Since there are 256 spots on each slide, the printing tool will dip and print 16 times per slide (Figure 1). The upper array of this figure represents the printing tool template and the lower array represents a magnification of the slide. The numbers 1 through 64 represent the 64 genes in the study. The shaded region in the upper array represents the first dip of the printing tool. Because the slide

is approximately the same size as the printing tool, the spots that are printed are spread throughout the slide. The next set of 16 template wells to be spotted is to the right of the shaded region and are printed to the right of each shaded spot on the slide. This process is continued until all spots have been printed. This printing process means that there is a one-to-one correspondence between the wells on the template and spots on the slide. While the sets of 16 wells ( $4 \times 4$  subarrays) on the template are associated with one dip of the printing tool, a similar  $4 \times 4$  subarray on the slide is associated with a particular pin of the printing tool.

### 3.2 Pilot study

Since it is not known whether the different pins and dips involved in the printing process have any effect on the size and composition of the spots on the array, both pins and dips are considered as potential spot-to-spot effects. For example, if the printing tool dipped a little deeper in one set of template wells, more solution could potentially collect on the pins and larger spots printed on the slide. Likewise, if one of the pins were larger (or smaller) than the others, it could collect and print a different amount of cDNA at the spots. To assess the amount of systematic variability due to these two factors, a pilot study was carried out. In this study, only one dye (Cy3) and one treatment were considered. On each of three slides, the same “gene” (salmon sperm DNA) was printed at all 256 spots. Thus, if there were no printing/spatial effects, there should be no systematic differences between spots within a slide.

The observed intensities after hybridization and washing are shown in Figure 2. The darker spots represent a lower intensity. For each slide, the intensities were scaled to be

between 0 and 1 so the three slides are not directly comparable. It is clear on each slide, however, that there is a nonrandom distribution of intensities. For slides 1 and 3, there appears to be an increasing trend in intensities from the lower left to upper right while in slide 2, it appears to be in the opposite direction.

This spatial trend suggests another potential source of the spot-to-spot variability. After the slide is printed and the mRNA is hybridized, the remaining genetic material is washed off, and dried using a centrifuge. Depending on the orientation of the slide within the centrifuge, labeled genetic material may be unequally washed off. Because each pin is associated with a particular region of the slide, it would be difficult to separate what is due to the pin and what is due to the washing effect if all the slides were dried in the same direction. With this in mind, slides 1 and 3 were oriented in the opposite direction as slide 2. While Figure 2 shows washing to be an important source of variability, the intensity patterns in slides 1 and 3 are different enough to suggest that other factors also play a role.

To assess the sources of variability in this pilot study, the linear model

$$\log(y_{ijk}) = \mu + S_i + D_{j(i)} + P_k + SP_{ik} + \epsilon_{ijk},$$

where  $S_i$  is the  $i$ th slide effect,  $P_k$  is the  $k$ th pin effect, and  $D_{j(i)}$  is the  $j$  dip effect within the  $i$ th slide, was fit to the data. Both slide and dip are considered random effects.

Table 3 summarizes the analysis for all three slides and when slide 2 is omitted. By omitting slide 2, the variation due to slide orientation is removed. It shows, in both cases, that a large percentage of the total variability in the spots can be explained by slide and pin effects. Considering all the slides, if dip were omitted from the linear model, the residual variance only increases from 0.017 to 0.019. If *both* dip and pin are removed from the model, however, the residual variance increases to 0.100.

Differences in the power of a pairwise comparison (e.g., when  $d = 0$  and  $t = 5$ ) under the fixed effects model (Eq. 2) based on the number of replications and inclusion of printing and washing effects are shown in Figure 3. Ignoring both pin and dip effects is comparable to the original Kerr, Martin, and Churchill (2000) model (with replicates). Including only pin effects in the model is possible with at least two replicates (and proper slide design). The increase in power when pin effects are included is quite substantial. The probability that a fold change of 1.5 would be detected increases from approximately 54% to 92%, even though the number of replicates drops from four to two. When four replicates are used with pin effects, the power increases to approximately 100%. There appears to be little difference in power when adding dip effects to the model.

## 4 Designing the Printing Tool Template

In order to separate the printing and washing effects from the treatment by gene effects, gene replicates (i.e., multiple spots on an array representing the same gene) are needed and they must be placed on the slide in such a way to avoid confounding. One method of replication

Table 3: Summary of the variance components in the pilot study when considering all three slides and when slide 2 is omitted.

Effect	Slides 1,2,3		Slides 1,3	
	Estimate of Variance	Percent of Total Variability	Estimate of Variance	Percent of Total Variability
Slide	0.007	6.2	0.001	0.9
Dip	0.003	2.2	0.004	5.6
Slide*Pin	0.089	77.0	0.043	67.0
Residual	0.017	14.6	0.017	26.5

is to create just one replicate within the template and dip numerous times into the same sets of wells. Each additional dip is known as an offset (i.e., two dips is known as a double offset). The reason this method is not recommended is because the same pin is associated with each gene thereby placing the replicate in the same region of the slide and potentially confounding it with washing and printing effects.

Instead, we propose that the replicates be included in the template and arranged in such a way that the replicates spots are spread throughout the slide. From a design only perspective, this can easily be done on a well by well basis. The problem is that no experimenter would be willing to create a template well by well. Instead, an experimenter would use a multi-tip pipette to fill several template wells at once. With this in mind, the protocol pictured in Figure 4 was developed for this microarray experiment involving four replicates of sixty-four genes and an 8-tip pipette.

First, there is a 96 ( $8 \times 12$ ) well sample plate and the 64 genes are randomly arranged on this plate in an  $8 \times 8$  square (Figure 4). The contents of these wells will be passed to the template using (for the most part) an 8-tip pipette. The 8-tip pipette allows one to pick up material from an entire row or column (i.e., 8 genes) and place the material in the template. The template, however, is a 384 ( $16 \times 24$ ) well plate so the pipette places the material into every other well. In order to separate the pin and dip effects from the gene effect, various genes need to be printed with different pins and with different genes. This means each replicate of 64 genes must be ordered differently. To alter the order of each replicate, but also use the 8-tip pipette, the replicates are created as follows.



**Applied Statistics in Agriculture**

**Rep 1** The first column of the sample plate is placed every other row in the first column of the template. The second column of the sample plate is then placed one well below the first in the template thereby completing the first column of the template. This same procedure is continued for the 3rd thru 8th columns of the sample plate.

**Rep 2** The same procedure as the first replicate is used except that the first row of the sample plate is skipped and then placed at the bottom.

**Rep 3** Rotate the plate 90° and repeat the procedure of Rep 1. This is comparable to now interweaving the rows of the sample plate instead of the columns.

**Rep 4** This replicate is generated the same way as Rep 2 using the rotated sample plate.

This protocol is designed to allow both pin and dip effects in the model. Skipping a row in Reps 2 and 4 is necessary only if dips are included. If dips are not considered, rotating the sample plate each replicate and interweaving the columns would suffice.

Combining this design of the slide within the previous “loop” design introduces additional terms in the subplot portion of the design. The linear model becomes

$$\log(y_{ijklmn}) = \mu + S_i + T_j + C_k + G_l + TG_{jl} + CG_{kl} + SG_{il} + D_{m(i)} + P_n + SP_{in} + \epsilon_{ijklmn}$$

Because each gene cannot be printed with each pin and included in each dip, the treatment by gene effects are partially confounded with these effects. We also consider both slide and dip effects to be random.

Table 4 summarizes the sources and degrees of freedom for the dip/pin model and the spot effect model. Fitting a spot effect is comparable to fitting gene, slide by gene, and

Table 4: ANOVA tables (source and degrees of freedom) for this microarray experiment using a spot effect (left column) and dip/pin effect (right column)

Effect	Spot DF	Pin, Dip DF
Slide	4	4
Dye	1	1
Trt	4	4
Trt*Gene	252	252
Dye*Gene	63	63
Spot	1275	
Gene	63	63
Slide*Gene	252	252
Rep(Slide*Gene)	960	
Pin		15
Dip(Slide)		75
Pin*Slide		60

replicate within slide by gene effects which total 1275 degrees of freedom. When dip and pin effects are fit in place of a spot effect, the model has 810 fewer parameters. Thus, provided that pin and dip factors account for a large portion of the spatial variability, a great reduction in the number of model parameters is possible.

## 5 Discussion

While there has been a lot of recent work in the area of design for microarray experiments, there has been little investigation into the printing and washing effects that contribute to the spatial variability within a slide. This small microarray experiment provides the opportunity to set up a design protocol for the slide itself that greatly improves the power of detecting differential gene expression. While our proposed protocol is experiment specific, the ideas of

replication on a slide and designing the slide so that the replicates aren't confounded with various printing and washing effects should be considered in any microarray experiment. For example, in our pilot study, it was the pin and slide effects that contributed the most to the overall variability of a spot on the array. Including pin effects is comparable to printing the gene replicates with different pins. Since pins are associated with a particular location of the slide, this means that the replicates should be spread out on the slide and not printed near each other, as an offset would do.

The pilot study provided us with an opportunity to investigate the printing and washing effects while keeping everything else constant. Further studies involving the use of two dyes would allow comparison of this model with the spot effect model. The pilot studies results suggest that a lot of the spot-to-spot variability is removed by pin and slide effects. If this were the case in general, then designing the slide appropriately would substantially reduce the number of model parameters making the analysis of these large data sets more manageable with current software packages.

One comparison issue not considered in this paper is the multiplicity issue. For example, an experiment involving arrays of  $g$  genes and  $t$  treatments would contain  $gt(t-1)/2$  pairwise comparisons. Currently, various multiple comparison procedures such as adjusted p-values (Westfall and Young, 1993) and controlling the false discovery rate (Benjamini and Hochberg, 1995) have been proposed, although in at least one case the techniques employed have been criticized for their conservatism (Dudoit *et al.*, 2000).

Biotechnological advances, such as microarray technologies, are obviously providing levels and magnitudes of genomic data that were unimaginable even five years ago. As such, how science is being performed is changing simply because of the data being generated. While the sheer magnitude of these data may answer some very demanding questions, there are more

intricate questions that can only be addressed through proper statistical design and analysis of experiments. Our contribution via this work is to provide the scientific community with a perspective for analysis that relies on design of experiments to address sources of variation inherent in all microarray studies. Once designed properly, the data supplied by these array experiments can be easily analyzed using standard statistical software procedures (SAS, S-Plus, or R). We also predict that much of what we are learning, suggesting, and implementing for the design and analysis of microarray experiments will be directly applicable to the next generation of genomic data, namely protein expression data.

### Acknowledgments

M.A. Black, R.W. Doerge, and O. Vitek are funded by a USDA-IFAFs (00-52100-9615) grant to R.W. Doerge. We thank Dr. Jeff Bennetzen for his continual support of the microarray studies associated with these investigations.

### References

1. Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, **57**:289-300.
2. Black, M.A. & Doerge, R.W. (To appear) Calculation of the minimum number of replicate spots required for detection of significant gene expression fold change in microarray experiments, *Proceedings of the 2001 Kansas State University Conference on Applied Statistics in Agriculture*, G.A. Milliken ed., Manhattan, KS : Department of Statistics.
3. Chen, Y., Dougherty, E.R. & Bittner, M.L. (1997) Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics* **2**, 364–374.

4. Cochran, W.G. and Cox, G.M. (1957). *Experimental Designs*. Wiley, New York.
5. David, H.A. (1963) The structure of cyclic paired-comparison designs. *J. Australian Math Soc* **3**, 117–127.
6. Dudoit, S., Yang, Y.H., Callow, M.J. and Speed, T.P. (2000) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Technical Report #578, Statistics Department, Stanford University.
7. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. (1998) Cluster analysis of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863-14868.
8. Hegde, P. *et al.* (2000) A concise guide to cDNA microarray analysis. *BioTechniques* **29**, 548-562.
9. Holter, N.S., Mitra, M., Maritan, A., Clepak, M., Banavar, J.R. & Fedoroff, N.V. (2000) Fundamental patterns underlying gene expression profiles: simplicity from complexity. *Proc. Natl. Acad. Sci. USA* **97**, 8409–8414.
10. Kerr, M.K., Martin, M. & Churchill, G.A. (2000) Analysis of variance for gene expression microarray data. *J. Comput. Biol.* **7**, 819–837
11. Kerr, M.K. & Churchill, G.A. (2001) Statistical Design and the Analysis of Gene Expression Microarrays. *Genetical Research* **77**, 123–128.
12. Lee, M.T., Kuo, F.C., Whitmore, G.A. & Sklar, J. (2000) Importance of replication in microarray gene expression studies: Statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl. Acad. Sci. USA* **97**, 9834–9839.
13. Newton, M.A., Kendziorski, C.M., Richmond, C.S., Blattner, F.R. & Tsui, K.W. (2001) On differential variability of expression ratios: Improving statistical inference about gene

- expression changes from microarray data *J. Comput. Biol.* **8**, 37–52.
14. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O. & Davis, R.W. (1996) Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* **93**, 10614–10619.
  15. Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. & Golub, T.R. (1999) Interpreting patterns of gene expression with self-organizing maps: Methods and application to hemotopoietic differentiation. *Proc. Natl. Acad. Sci. USA* **96**, 2907–2912.
  16. Westfall, P.H. & Young, S.S. (1993). Resampling-based based multiple testing: examples and methods for p-value adjustment. Wiley, New York.
  17. Wolfinger, R.D., Gibson, G., Wolfinger, E.D., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C., and Paules, R.S. (To appear). Assessing Gene Significance from cDNA Microarray Expression Data via Mixed Models. *J. Comp. Biol.*

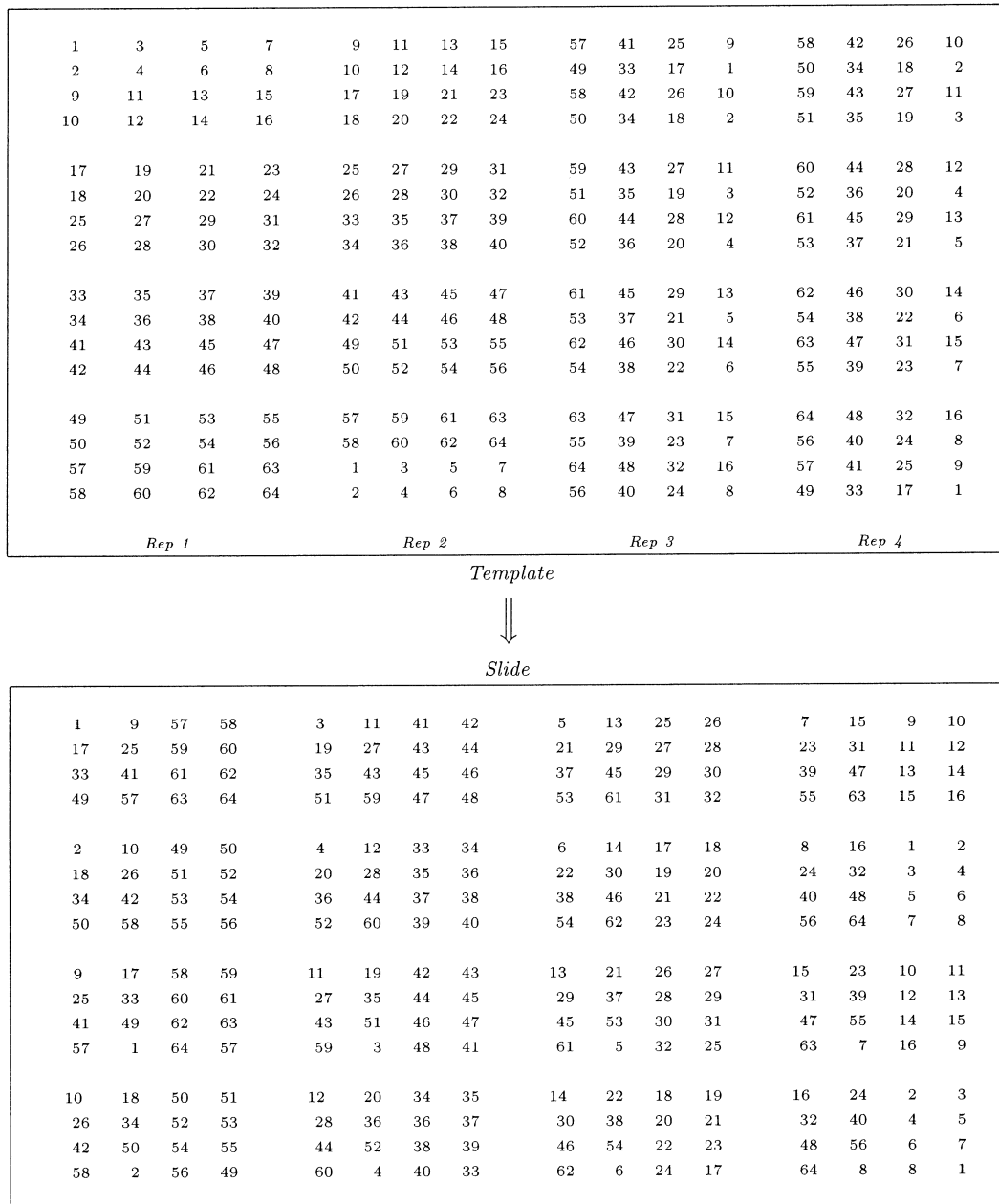


Figure 1: The printing process for one slide in this experiment. The template contains four replicates of the 64 genes. The shaded numbers represent the first dip in the template and where these spots are placed on the slide.

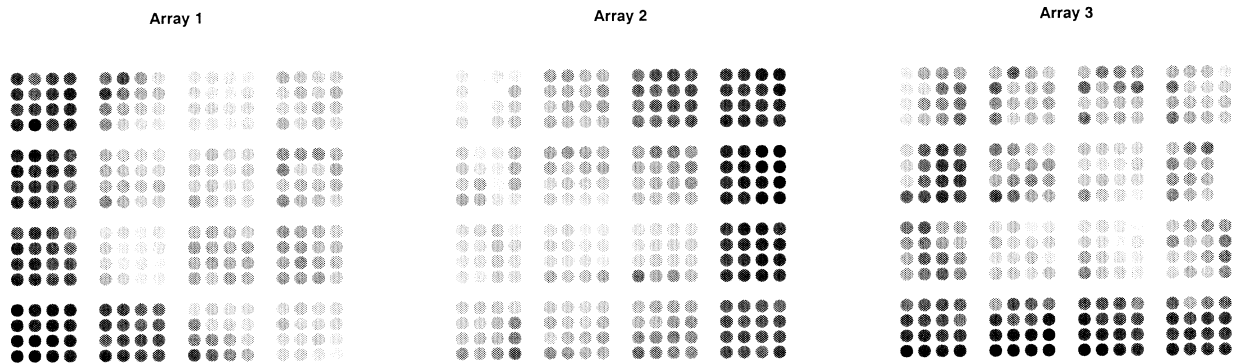


Figure 2: Distribution of intensities for the three slides of the pilot study. Slides 1 and 3 were washed in a direction opposite of that for slide 2. There appears to be a right to left, top to bottom trend in intensities.



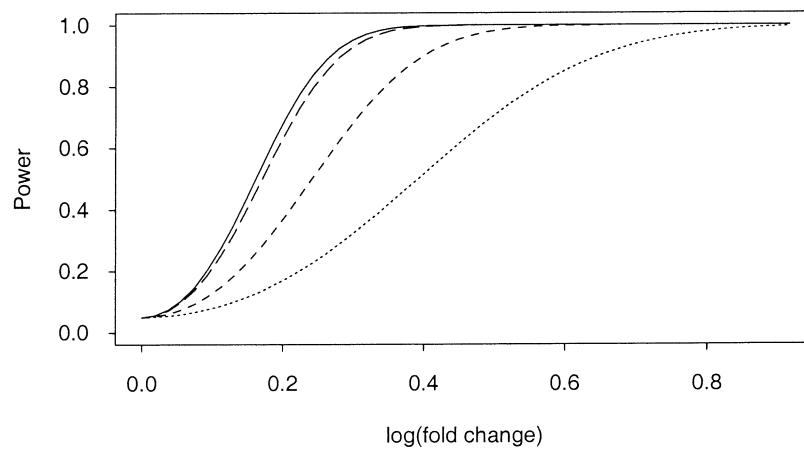


Figure 3: Power curve for a pairwise comparison ( $\alpha = .05$ ) involving treatments on the same slide (Eq. 2). The solid line is for the model that includes both pin and dip effects ( $\sigma^2 = 0.016, r = 4$ ), the dotted curve is for the model the does not include these effects ( $\sigma^2 = 0.110, r = 4$ ), the short dashed curve is for the model that includes pin effects ( $\sigma^2 = 0.019, r = 2$ ), and the long dashed curve is for the model that includes pin effects ( $\sigma^2 = 0.019, r = 4$ ).

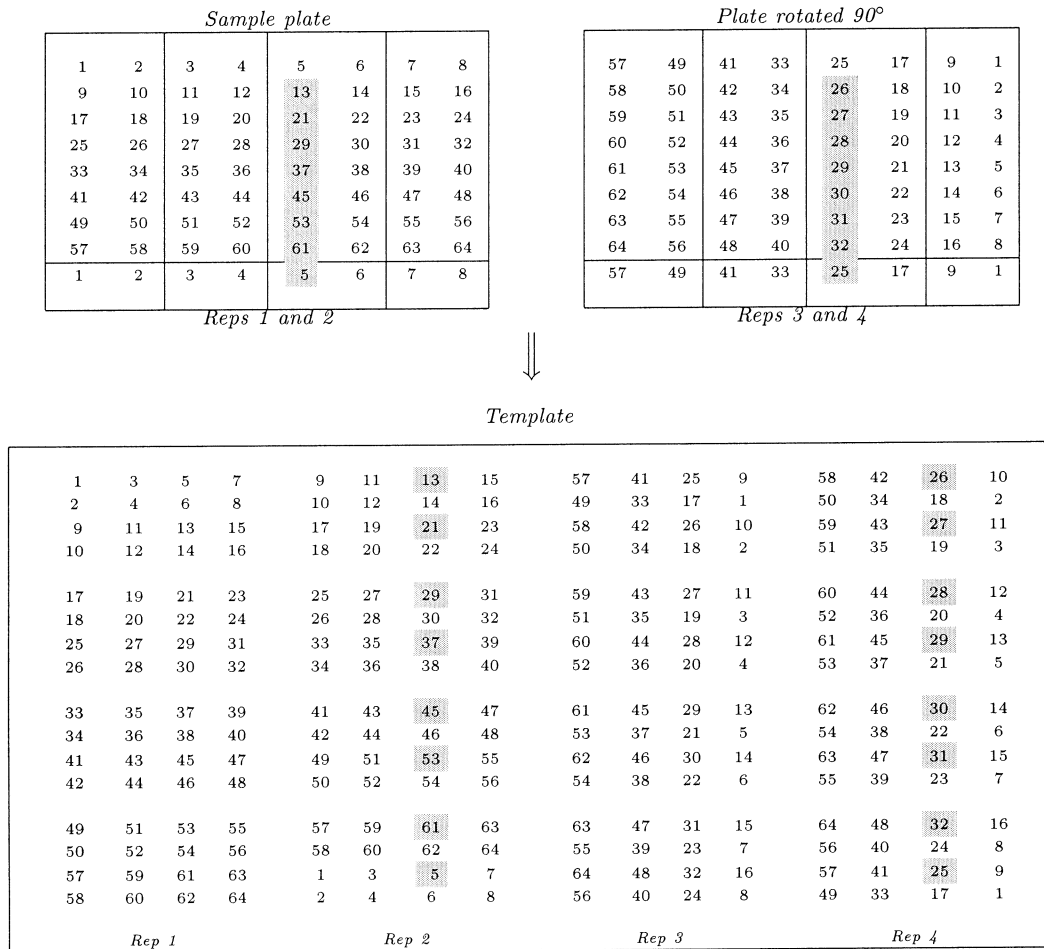


Figure 4: The proposed design of the template for this experiment involving 64 genes and four replicates. Replicates 1 and 2 are generated from the plate shown on the left. Replicates 3 and 4 are generated in a similar manner after the plate has been rotated 90°.