Chapter

HOW MANY MICE AND HOW MANY ARRAYS? REPLICATION IN MOUSE cDNA MICROARRAY EXPERIMENTS

Subtitle

Xiangqin Cui and Gary A. Churchill

The Jackson Laboratory

Abstract: Biological and technical variances were estimated from the Project Normal data using the mixed model analysis of variance. The technical variance is larger than the biological variance in most genes. In experiments for detecting treatment effects using a reference design, increasing the number of mice per treatment is more effective than pooling mice or increasing the number of arrays per mouse. For a given number of arrays, more mice per treatment with fewer arrays per mouse is more powerful than fewer mice per treatment with more arrays per mouse. A formula is provided for computing the optimum number of arrays per mouse to minimize the total cost of the experiment.

Keywords: cDNA microarray, mixed model, variance components, experimental design, reference design, replication

1. INTRODUCTION

cDNA microarrays are widely used in gene expression profiling. This complex technology involves many steps. Each step can introduce variation (technical variation), which accumulates in the final observations. Some of the systematic variation can be minimized by data transformation and normalization [Cui and Churchill, 2002; Quackenbush, 2001]. However the

intrinsic variation from each step cannot be completely eliminated. Therefore, it is desirable to estimate these variance components from data and to use them to improve the statistical inference.

Biological variation is another source of variation. In order to make general claims about a treatment effect, multiple experimental units (biological replicates) from the population should be assayed. Otherwise, conclusions about the treatment effect will be restricted to the samples tested [Churchill, 2002; Cui and Churchill, 2003].

Variance components from each source can be estimated by modeling microarray data using the mixed model analysis of variance (ANOVA). Unlike fixed effect models, which treat the effects of factors as if they would be repeated exactly if the experiment were to be repeated, mixed models treat some factors, such as the array effect, as random samples from a population. In other words, we assume that if the experiment were to be repeated, the same effects would not be exactly reproduced, but that similar effects would be drawn from a hypothetical population of effects. The variation of these random factors is considered when inferences are made about the treatment effects [Littell et al., 1996; Searle et al., 1992; Witkovsky, 2002]. Therefore, the mixed model results are more general and reproducible.

Knowledge about variance components also provides a basis for making experimental design decisions regarding the allocation of resources [Churchill, 2002; Yang and Speed, 2002]. Replication in a microarray experiment can be present at any levels. For example, multiple samples per treatment (sample level), multiple RNA extractions per sample (RNA level), multiple labelling reactions per RNA source (labelling level), Multiple arrays per label (array level) and multiple spots per gene on each array (spot level). Replication at levels that have large variance components can significantly improve the overall sensitivity of the experiment. Current literature on microarray replication is mainly at array level [Pan et al., 2002; Wolfinger et al., 2001; Zien et al., 2002] and spot level [Lee et al., 2000], with little attention given to biological sample level. Here we use a linear mixed model ANOVA to estimate the variance components at the mouse and the array levels and explore the implications of these variances in the allocation of biological and array replication.

2. MATERIALS AND METHODS

2.1 Data pre-process

The corrected Project Normal microarray data (text files) were downloaded from http://www.camda.duke.edu/camda02/contest.asp. Only

the foreground signals were used. The background signals and flags were ignored. All genes including blanks (5776 spots on each array) are included in the analysis. The data were $\log_2$- and intensity-lowess- transformed and then normalized by subtracting the channel mean from each signal, which is the same as fitting a normalization model [Wolfinger et al., 2001] or fitting some global factors for normalization in global ANOVA models [Kerr et al., 2000].

2.2 Model

The following mixed linear model was fitted to each gene in each organ to estimate variance components,

$$Y_{ij} = \mu + A_i + D_j + M_k + R_h + \varepsilon_{ij}, \quad (1)$$

where μ is the gene mean; A_i (i = 1 ... 24) is array effect; D_j (j = 1, 2) is dye effect. M_k (k = 1 ... 6) is the effect of individual mice, where value of k is determined by the array and dye combination (i,j). R_h (h = 1, 2) is an indicator of reference (h = 1) versus tissue sample (h = 2), which is again determined by the combination of array and dye. ε_{ij} is the residual measurement error. D_j and R_h are fixed effects. A_i, M_k, and ε_{ij} are random effects with assumed normal distributions $N(0,\sigma_A^2)$, $N(0,\sigma_M^2)$, and $N(0,\sigma_\varepsilon^2)$, respectively. σ_A^2, σ_M^2, and σ_ε^2 were estimated using the restricted maximum likelihood (REML) method [Searle et al., 1992; Wikovsky, 2002].

A bigger model that includes the organ effect,

$$Y_{ij} = \mu + A_i + D_j + O_n + M_k + R_h + \varepsilon_{ij}, \quad (2)$$

is fitted to the combined data of all three organs. The organ effect O_n (n = 1, 2, 3) is a fixed effect with three levels representing kidney, liver, and testis. The number of levels in array effect A_i becomes 72 (i = 1 … 72) in this model. The remaining terms are the same as in equation (1).

2.3 Proportion of detectable genes

The power calculation for testing the treatment effect in each gene follows Wolfinger et al. [2001]. The error variance (*EV*) for the treatment effect was calculated using formula $EV = \sigma_M^2 / m + \sigma_\varepsilon^2 /(mn)$, where *m* and *n* are the number of mice per treatment and number of array pairs per mouse, respectively. The effectiveness of the whole experiment is represented by the proportion of detectable genes, where a gene is detectable if we have at least 50% power to detect differential expression at a given significance level and fold change.

3. RESULTS

3.1 Variance components

The variance components from the random effects - mouse (σ_M^2), array (σ_A^2), and measurement (σ_ε^2) - were estimated from the corrected Project Normal data set for each gene in each organ using the mixed model in equation (1). The mean and median of each variance component for each organ across all genes are shown in Table 1. In general, the array variance is the largest component and it is more than 10 times larger than the mouse and the measurement variances. The mouse variance is the smallest for both mean and median. Among the three organs, liver has the smallest mouse variances.

Table 1. Mean and median of variance components. The degrees of freedom for estimating variance components from mouse, array, and measurement in each gene are 5, 18, and 22, respectively.

	variance components	kidney	liver	testis
Mean	mouse	0.0252	0. 0092	0. 0126
	array	0. 3221	0. 2957	0. 3068
	measurement	0. 0250	0. 0308	0. 0244
Median	mouse	0. 0090	0. 0014	0. 0046
	array	0. 2168	0. 1848	0. 2075
	measurement	0. 0168	0. 0178	0.0154

The distributions of variance components across genes in each organ are shown in Figure 1. The histogram of the mouse variance components for

individual tissue shows a characteristic bimodal pattern. The majority of the genes have small mouse variance in all three organs. Only a small portion of genes have larger mouse variances. For example, 18%, 5%, and 6% of the genes have mouse variance over 0.04 in kidney, liver and testis, respectively. This suggests that the majority of the genes are under tight control and a small portion of the genes are less tightly controlled. In addition, the identities of the less tightly controlled genes are different in the three organs. Among all genes that have mouse variance over 0.04, only 18% are common among all three organs. The low overlapping of the less tightly controlled genes could cause the mouse variances to average out when the data from all three organs were combined for variance component analysis. That is exactly what we saw when we fitted a bigger ANOVA model (equation 2) to the combined data with organ effect accounted for. The "shoulder" in the mouse variances of kidney and testis disappeared from the mouse variance of the combined data (Figure 1D).

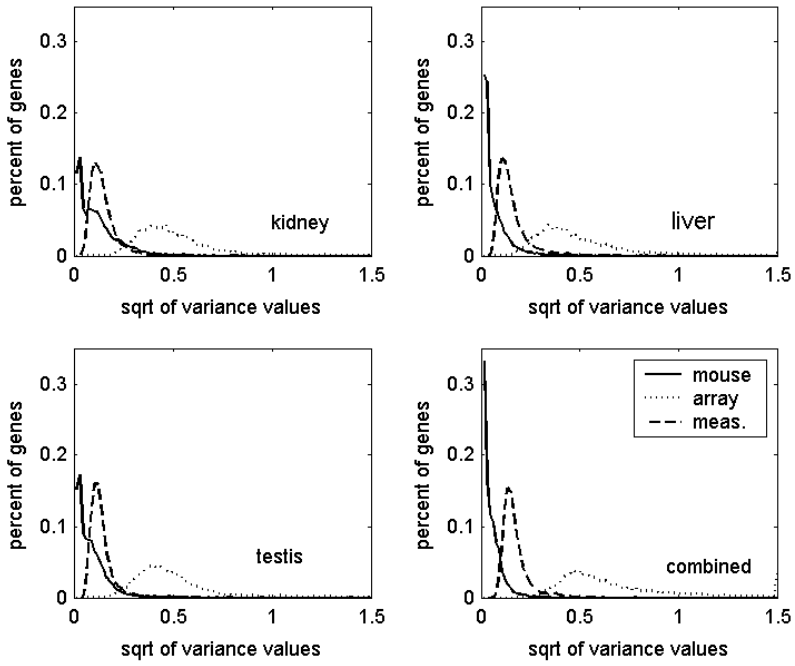

Figure 1. Distributions of the estimated variance components. Smoothed histograms of standard deviation in stead of the variance are plotted for clarity.

For the purpose of making comparisons, variance components from two other experiments were estimated and are shown in Figure 2. The Gallstone experiment [Henning Wittenburg and Beverly Paigen, personal communication] looked at diet and strain effects on RNA expression in mouse liver. Two diets and three strains were assayed in a 2 x 3 factorial design with two mice for each diet and strain combination. The experiment used direct comparisons among samples on a total of 28 arrays. Each array has two adjacent spots for each gene. The Brain Cortex experiment [http://pga.tigr.org/MouseText.shtml] looked at the variation of mRNA expression of brain cortex in mice using arrays with duplicated spots. In this case the duplicated spots were dispersed across the array. Because these data sets have duplicated spots, there are four measurements obtained for each clone, two in the red channel and two in the green channel. Measurements obtained on the same spot (one red and one green) will be correlated because they share common variation in the spot size. Measurement obtained in the same color (both red or both green) will be correlated because they share variation through a common labeling reaction. Therefore, additional random factors for spot and labeling effects were estimated using the mixed model.

The estimated mouse variance components in these two data sets are comparable with those estimated from the Project Normal data set (Figure 1). Some of the technical variance components (array, spot and measurement) are larger in the Brain Cortex experiment than in the Gallstone experiment. One possible explanation for this difference is that the dispersal of replicated spots in the Brain Cortex experiment will pick up spatial variation on the array. In addition, the difference may reflect different levels of control over spot size and morphology or hybridization quality.

Including data from all three organs in the bigger model (2) allows us to estimate the variance components from the combined data and enables us to find genes that are differentially expressed among the three organs. In this case the three organs represent three different "treatments" and finding genes that are differentially expressed could be regarded as one goal of the experiment. At significance level of 0.05 after Bonferroni correction (a stringent adjustment for multiple testing), 2279 genes were identified as differentially expressed among kidney, liver, and testis. It is not surprising to see so many differentially expressed genes given that these are three very different organs. We did not try to characterize these genes any further.

3.2 Power for detecting treatment effects

If the goal of the experiment is to detect treatment effects, statistical inference should be based on the total variance including the biological variance. This will typically be a weighted average of the variance

components and the relative weighting will depend on the design of the experiment. The error variance (*EV*) for treatment effect is a combination of all the variance components that are nested under treatment.

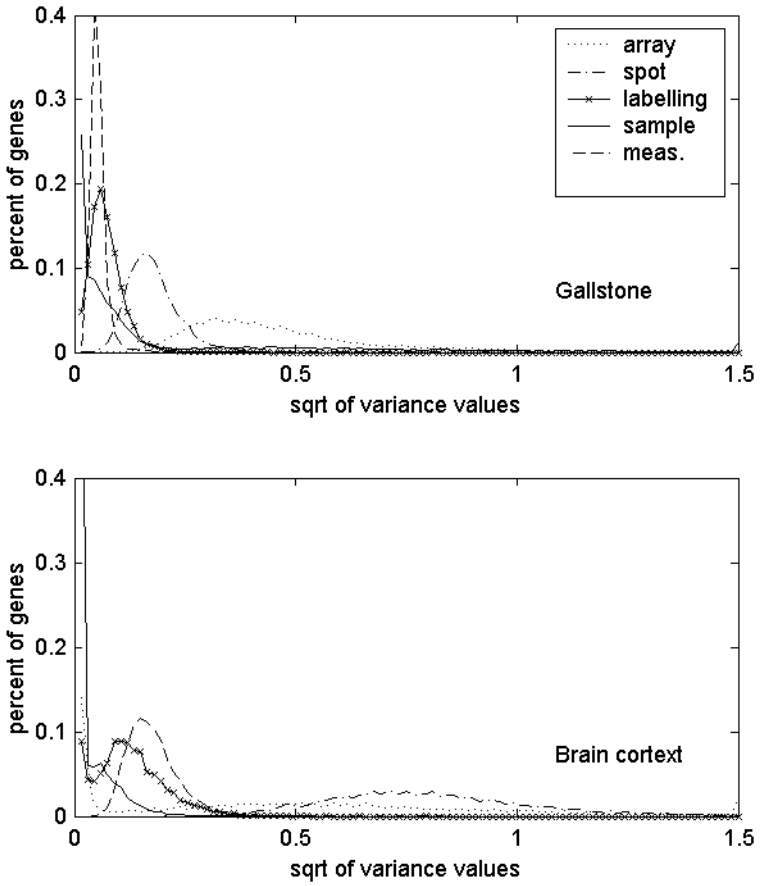

Figure 2. Variance components from mouse Gallstone and Brain Cortex data sets. The smoothed histograms of standard deviations in stead of the variances are plotted for clarity.

In a simple reference design for comparing two treatments, TrtA and TrtB (Figure 3), the *EV* for treatment effect can be computed by

$$EV = \frac{\sigma_M^2}{m} + \frac{\sigma_\varepsilon^2}{mn} \qquad (3)$$

with *m* as number of mice per treatment and *n* as pairs of dye-swap arrays per mouse. This equation is obtained using the within-array information only as in tests based on log ratios. The mixed model can extract between-array information and combine it with the within-array information [Littell et al., 1996]. The *EV* from the mixed model (equation 1) will be

$$EV = \frac{\sigma_M^2}{m} + \frac{\sigma_\varepsilon^2}{mn} - \frac{\sigma_\varepsilon^4}{2mn(\sigma_A^2 + \sigma_\varepsilon^2)} \qquad (4)$$

We have seen that array variance σ_A^2 dominates, thus the correction factor will generally be quite small and we use equation (3) throughout the rest of this paper. Note that the array variance is not included here. This is a consequence of the pairing in two-dye microarray experiments. In one-color systems, array variance will be a major component of *EV*.

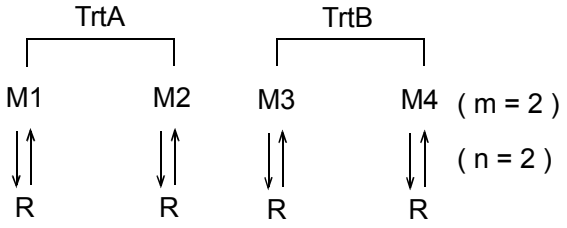

Figure 3. Reference design for testing treatment effect TrtA versus TrtB. Arrows represent arrays with head as Cy3 and tail as Cy5. M, mouse; R, reference; m, number of mice per treatment; n, number of array pairs per mouse.

A small error variance is desirable in order to increase the power of statistical tests (*t* or *F*) for treatment effects. From equation (3) we can see that this can be achieved by increasing the number of mice per treatment, *m*, to reduce both components proportionally. However, increasing *m* may mean substantial increase in cost when mice are expensive. Therefore, it may be desirable to simply increase the number of arrays per mouse, *n*. However, this strategy will only reduce the technical component of the variance, therefore, it is most effective when measurement variance is larger than the mouse variance.

Figure 4 shows the proportion of genes in which a two-fold difference between two treatments can be detected with at least 50% power at a significance level of 0.05 after Bonferroni correction at various combinations of *m* and *n* in the three organs. The *m=2* (2 mice per treatment) case does not show any power, while the 4, 6 and 8 mice per

treatment cases show substantially increased power. The lack of power in the *m=2* case is mainly due to the small degrees of freedom (*df*) for estimating the variance in the *t* test. Since the right side of equation (3) is the sum of two variance components, we use the smaller *df* as a conservative approximation to the *df* of the error variance. Using this method, the *df* of the *m=2* case is 2 while those of the other cases are 6, 10, and 14. At least 5 degrees of freedom are generally recommended for a *t* test.

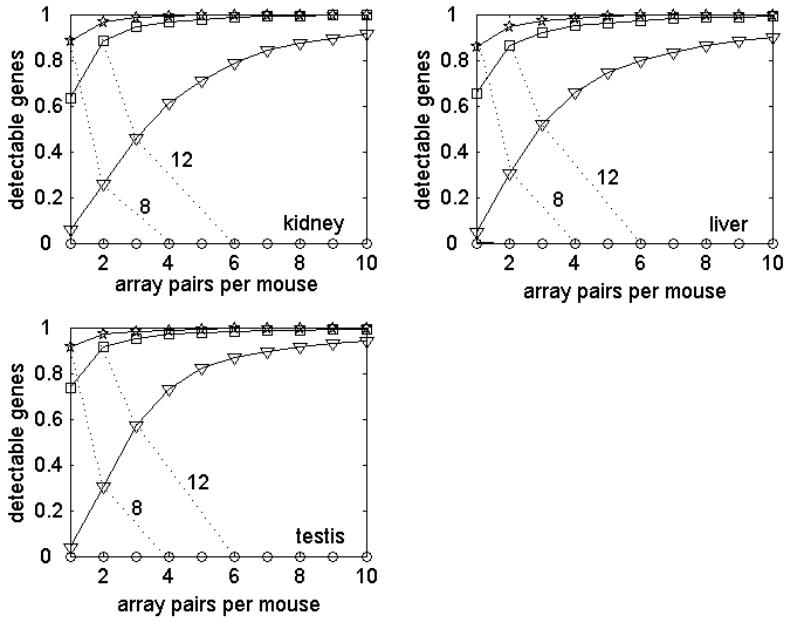

Figure 4. Power for detecting 2-fold change between two treatments at various combinations of number of mice per treatment and number of arrays per mouse. Circle, triangle, square, and star represent 2, 4, 6, and 8 mice per treatment. Dotted lines represent the same number of array pairs (8 or 12) for each treatment. Significance level is 0.05 after Bonferroni correction.

Increasing the number of arrays per mouse generally increases the proportion of detectable genes. The increase is most obvious with small numbers of arrays (from 1 to 4 pairs). Once there are 6 pairs of arrays per mouse, there is hardly any gain from increasing arrays further. If the goal is to achieve detectability for more than 90% of the genes, at least 6 mice per treatment is recommended. For a small number of genes which have significantly larger mouse variances, such as the genes declared significant in Pritchard *et al*. [2001], increasing the number of arrays per mouse can never achieve the same precision as increasing the number of mice per treatment. For the same total number of arrays per treatment, more mice per

treatment with fewer arrays per mouse will have more statistical power as shown by the dotted lines in Figure 4. Therefore, if mice are relatively inexpensive, more mice and fewer arrays per mice is a better choice.

The percent of genes that can be detected at various fold changes is shown in Figure 5 for various combinations of array and mice. Increase of array number per treatment can greatly increase the detection of genes with smaller fold changes. In addition, the experiment with six mice per treatment has substantially higher power than that with four mice per treatment at small fold changes.

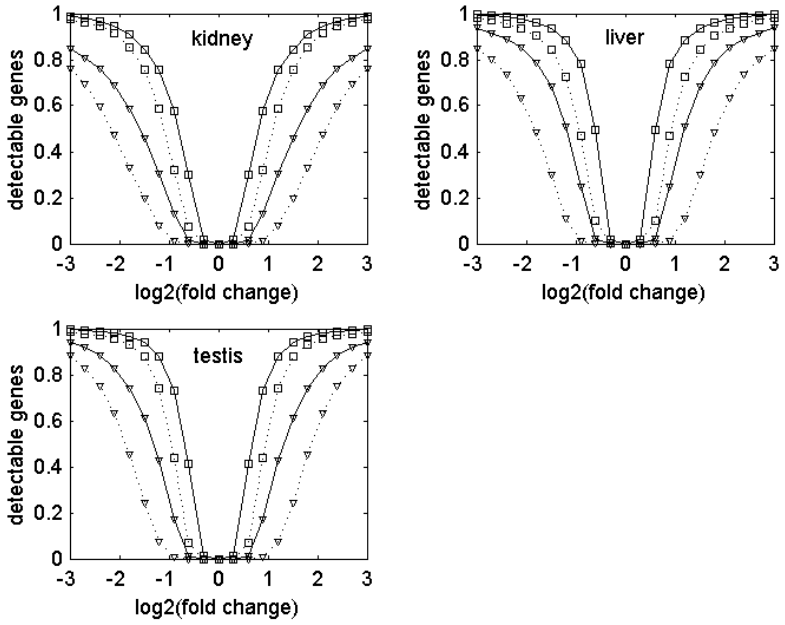

Figure 5. Power for detecting treatment effects at different fold changes. Triangles and squares represent 4 and 6 mice per treatment, respectively. Dotted and solid lines represent 1 and 3 pairs of arrays per mouse, respectively. Significance level is 0.05 after Bonferroni correction.

3.3 Optimum resource allocation

In practice, there is often a limited budget. In order to utilize the resource most effectively, we need to balance mice and arrays to minimize the cost of the whole experiment. Let C_M represent the cost of a mouse and C_A represent the cost of a pair of arrays. Suppose there are m mice per treatment and each mouse will be measured using n pairs of arrays. The total cost will be

$$Cost = m \cdot C_M + m \cdot n \cdot C_A \quad (5)$$

Combining equations (3) and (5), we can derive the optimum number of arrays per mouse [Kuehl, 2000] in order to keep the total cost minimum as

$$n = \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_M^2} \cdot \frac{C_M}{C_A}} \quad (6)$$

For example, suppose that the array price is $300 each and three different mouse strains are $15, $300, and $1500 per mouse. If we use the median values of $\sigma_\varepsilon^2 / \sigma_M^2$ obtained from each organ of the Project Normal data, the optimum pairs of arrays per mouse for differently priced mouse strains can be computed using equation (6). Because all the calculations are based on reference design with dye swap (Figure 3), n is rounded up to the nearest integer (Table 2).

Table 2. Optimum pairs of arrays per mouse (N = number of pairs of arrays per mouse)

Array price (pair)	Mouse price	N (kidney)	N (liver)	N (array)
$600	$15	1	1	1
$600	$300	1	3	2
$600	$1500	2	6	3

If budget is fixed, equation (6) can be plugged into equation (5); the resulting *m* and *n* values will give a minimum *EV* for the fixed cost. If there is a certain *EV* to achieve, equation (6) can be plugged in equation (3) to find *m* and *n* that will result in minimum cost for the desired *EV*.

When there are *r* replicated spots for each clone on each array, additional variance components for labeling (σ_L^2) and spot can be fitted to the data to capture the covariance shared among observations within a spot and a labeling reaction as shown in Figure 2. The *EV* of treatment mean can be approximated as

$$EV = \frac{\sigma_M^2}{m} + \frac{\sigma_L^2}{mn} + \frac{\sigma_\varepsilon^2}{mnr} \quad (7)$$

and the optimum number of dye-swap pairs per mouse will be

$$n = \sqrt{\frac{r\sigma_L^2 + \sigma_\varepsilon^2}{r\sigma_M^2} \cdot \frac{C_M}{C_A}} \tag{8}$$

3.4 Pooling mice

Pooling mice is another possible strategy to increase the precision of treatment tests in this experiment. It reduces *EV* by reducing the mouse variance component. The between-pool variance for a pool size of *k* mice will be approximately

$$\sigma_{pool}^2 = \frac{1}{k^\alpha}\sigma_M^2 \tag{9}$$

for some constant $0 < \alpha < 1$, which is related to the pooling procedure. In the case of $\alpha = 0$ pooling will have no effect. In the case of $\alpha = 1$ the mouse variance is reduced in direct proportion to the pool size *k*. Unfortunately, there is no information about α in the Project Normal data set. Therefore, we cannot estimate how effective pooling will be in this experiment. Suppose that we can reach the maximum effect of pooling, $\alpha = 1$, the effect of pooling 3 mice in each sample is shown in Figure 6. The power for detecting treatment effect increases slightly in kidney and testis, even less in liver. The gain is the biggest when there are four samples per treatment in all three organs.

Due to the presence of α in equation (9), for a fixed total number of mice, more pools with fewer mice in each pool will result in a smaller *EV*, given that a fixed number of arrays for each treatment will be used. For example, 10 pools with 4 mice in each pool is better than 4 pools with 10 mice per pool.

4. DISCUSSION

In this paper we estimated the biological and technical variance components from the Project Normal data and found that the technical variance is the major component in the treatment error variance. Therefore, reducing the technical variance will be effective in increasing the power of the test for treatment effect. One way to reduce the technical variance is through increasing the number of arrays per mouse as discussed in the result section. Another approach, not considered here, is to use more efficient experimental designs, such as loop design [Kerr and Churchill, 2001a; Kerr

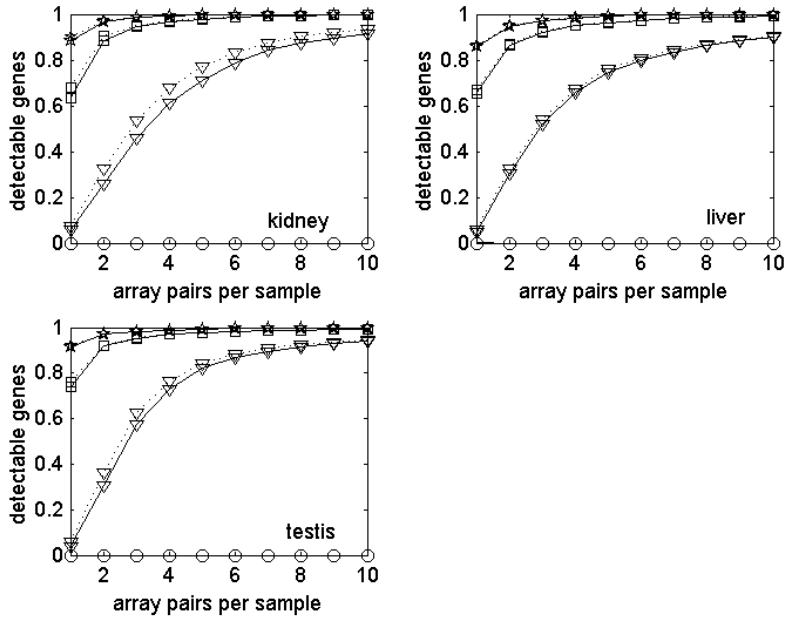

Figure 6. Effect of pooling mice in increasing the power of detecting treatment effect. Circle, triangle, square, and star lines represent 2, 4, 6, and 8 mice (solid) or pools (dotted). Each pool has 3 mice. Significance level is 0.05 after Bonferroni correction.

and Churchill, 2001b] and block design [Simon et al., 2002]. No reference sample is used in these designs and the treatment samples are compared directly. Direct comparisons have higher efficiency than indirect comparisons through the reference sample, because no measurement is wasted on the reference sample [Kerr and Churchill, 2001a; Yang and Speed, 2002]. For comparing two samples, the technical variance from a direct comparison is ¼ of that from an indirect comparison through a reference sample using the same number of arrays. For more than two samples, the reduction of technical variance depends on the positions of comparing samples in the design and it becomes more complicated to calculate when complicated loop or block designs are used [Yang and Speed, 2002; Churchill 2002].

The power calculations in this paper are based on the assumption that each gene has unique variance components, which is a relatively unstable method in experiments with limited number of data points per gene. In those experiments, we could assume that the variance components are the same across all genes; in this case the proportion of detectable genes will be either 0% or 100% because the test statistic (t or F) value will only depend on the

magnitude of the fold change, not the variance components from each gene. However, this type of test is subject to bias if the data have not been properly normalized. If we combine the above two situations and assume that each variance component for each gene is the combination of common variation for all genes and some variation from individual gene [Baldi and Long, 2001; Cui and Churchill, 2003; Lönnstedt and Speed, 2002], the proportion of detectable genes will again be a useful concept. Combining information about variance components across the genes is a potentially powerful approach that we are currently investigating.

Multiple testing adjustments are usually applied to control the false positive errors in microarray experiments when thousands of genes are tested one at a time as in this experiment. In this experiment, we used the Bonferroni correction, which is a stringent family-wise error rate (FWER) correct. The 0.05 significance level after Bonferroni correction means that we expect a probability of 0.05 to have one or more errors in the whole list of the identified significant genes. There are other less stringent multiple testing adjustment methods, such as FDR (false discovery rate) adjustments, which controls the percent of genes in the declared significant gene list that are false [Benjamini and Hochberg, 1995; Storey, 2002]. These methods are appropriate for exploratory experiments in which a list of candidate genes will be confirmed using other technologies. When less stringent multiple test adjustments are used, the power of the test in Figures 4, 5, and 6 will all increase.

All calculations of power and resource allocation in this paper are based on the estimation of the variance components, which relies on replication. In the Project Normal data set, mouse and array are replicated; therefore, the variance introduced by these two factors can be estimated. Similarly, variance components from any other factor, such as RNA extraction, labeling, spot, can be estimated as long as there is replication at an appropriate level of the experimental design. Estimation of all the possible variance components from microarray technology can help to determine where the largest variances come from. These steps are targets for microarray technology improvement. The estimation of variances becomes even more important when new microarray platforms, new techniques, or new facilities are implemented. For large and complicated experiments, pilot studies are recommended for estimating the biological and technical variances in order to customize the power and cost calculation in the design stage.

5. CONCLUSION

We estimated the biological and technical variance components from the Project Normal data set using a mixed model ANOVA. Comparison among the estimated variance components revealed that the technical variance is larger than the biological variance for most of the genes. To detect treatment effect using reference design in experiments with similar variance component values, reducing the biological variance by pooling mice will not be as effective as increasing the replication of arrays to reduce the technical variance. For fixed number of arrays per treatment, designs with more mice per treatment and fewer arrays per mouse are more powerful than designs with fewer mice per treatment and more arrays per mouse.

ACKNOWLEDGEMENTS

We thank Hao Wu for helping the software development and Jason Stockwell for valuable comments. We also thank Henning Wittenburg and Beverly Paigen at the Jackson Lab for the Gallstone data and John Quackenbush at TIGR for the Brain Cortex data. This research is supported by grants CA88327, HL66620, and HL55001 from the National Institute of Health.

REFERENCES:

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

Benjamini, Y, and Hochberg, Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B 57: 289-300.

Churchill, GA (2002). Fundamentals of experimental design for cDNA microarrays. Nat Genet 32 Suppl 2: 490-495.

Cui, X, and Churchill, GA (2002). Data transformation for cDNA microarray data. http://www.jax.org/staff/churchill/labsite/pubs/index.html.

Cui, X, and Churchill, GA (2003). Statistical tests for differential expression in cDNA microarray experiments. Genome Biol 4: 210.

Kerr, MK, and Churchill, GA (2001a). Experimental design for gene expression microarrays. Biostatistics 2: 183-201.

Kerr, MK, and Churchill, GA (2001b). Statistical design and the analysis of gene expression microarray data. Genet Res 77: 123-128.

Kerr, MK, Martin, M, and Churchill, GA (2000). Analysis of variance for gene expression microarray data. J Comput Biol 7: 819-837

Kuehl, R (2000). Designs of experiments: statistical principles of research design and analysis, 2 edn, Duxbury Press).

Lee, ML, Kuo, FC, Whitmore, GA, and Sklar, J (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci U S A 97: 9834-9839.

Littell, RC, Milliken, GA, Stroup, WW, and Wolfinger, RD (1996). SAS system for mixed models (Cary, NC, SAS institute Inc.,).

Lönnstedt, I, and Speed, T (2002). Replicated Microarray Data. Statistica Sinica 12: 31-46.

Pan, W, Lin, J, and Le, CT (2002). How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach. Genome Biol 3: research0022.

Pritchard, CC, Hsu, L, Delrow, J, and Nelson, PS (2001). Project normal: defining normal variance in mouse gene expression. Proc Natl Acad Sci U S A 98: 13266-13271.

Quackenbush, J (2001). Computational analysis of microarray data. Nat Rev Genet 2: 418-427.

Searle, SR, Casella, G, and McCulloch, CE (1992). Variance components, John Wiley and sons, Inc.).

Simon, R, Radmacher, MD, and Dobbin, K (2002). Designs of studies using DNA microarrays. Genet Epidemiol 23: 21-36.

Storey, J (2002). A direct approach to false discovery rates. J R Statist Soc B 64: 479-498.

Witkovsky, V (2002). MATLAB algroithm mixed.m for solving Henderson's mixed model equations. http://www.mathpreprints.com

Wolfinger, RD, Gibson, G, Wolfinger, ED, Bennett, L, Hamadeh, H, Bushel, P, Afshari, C, and Paules, RS (2001). Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 8: 625-637.

Yang, YH, and Speed, T (2002). Design issues for cDNA microarray experiments. Nat Rev Genet 3: 579-588.

Zien, A, Fluck, J, Zimmer, R, and Lengauer, T (2002). Microarrays: how many do you need? Proc. RECOMB'02. 321-330.