Experimental design for three-color and four-color gene expression microarrays

Yong Woo¹,³, Winfried Krueger², Anupinder Kaur² and Gary Churchill¹,

¹The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA, ²Laboratory for Microarray Technology Department of Genetics and Developmental Biology, L40760/4085, University of Connecticut Medical School, 263 Farmington Avenue, Farmington, CT 06030, USA and ³Functional Genomics PhD Program, The University of Maine, Orono, ME 04469, USA

Received on January 15, 2005; accepted on March 27, 2005

ABSTRACT

Motivation: Three-color microarrays, compared with two-color microarrays, can increase design efficiency and power to detect differential expression without additional samples and arrays. Furthermore, three-color microarray technology is currently available at a reasonable cost. Despite the potential advantages, clear guidelines for designing and analyzing three-color experiments do not exist.

Results: We propose a three- and a four-color cyclic design (loop) and a complementary graphical representation to help design experiments that are balanced, efficient and robust to hybridization failures. In theory, three-color loop designs are more efficient than two-color loop designs. Experiments using both two- and three-color platforms were performed in parallel and their outputs were analyzed using linear mixed model analysis in R/MAANOVA. These results demonstrate that three-color experiments using the same number of samples (and fewer arrays) will perform as efficiently as two-color experiments. The improved efficiency of the design is somewhat offset by a reduced dynamic range and increased variability in the three-color experimental system. This result suggests that, with minor technological improvements, three-color microarrays using loop designs could detect differential expression more efficiently than two-color loop designs.

Availability: http://www.jax.org/staff/churchill/labsite/software

Contact: garyc@jax.org

Supplementary information: Multicolor cyclic design construction methods and examples along with additional results of the experiment are provided at http://www.jax.org/staff/churchill/labsite/pubs/yong

© The Author 2005. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oupjournals.org
et al., 2004). In this paper, we propose experimental design guidelines to facilitate the use of three-color microarrays.

### 1.2 Approaches in experimental design

Computer software, such as SAS OPTEX (SAS Institute, Inc., Cary, NC), can generate efficient experimental designs using numerical optimizations. Although the resulting designs may be efficient, generally, they are not always intuitive and are limited to specific cases as suggested by Tobias (http://support.sas.com/rnd/app/papers/optex.pdf). Furthermore, numerically optimal designs may lack other important features, such as robustness and balance. In many cases, classical, combinatorial approaches to design may be more appropriate because they offer general solutions to large classes of design problems and can incorporate features that lead to balanced, robust and readily interpreted experiments. For two-color microarrays, designs such as dye-swap, loop and reference have been proposed by various investigators (Kerr and Churchill, 2001; DeRisi et al., 1996). These basic elements can also be combined. For example, reference designs coupled with direct dye-swaps between samples of interest can result in powerful, robust and readily extendible sets of comparisons (Kerr and Churchill, 2001). Similarly, for three- and four-color microarrays, we propose a set of basic design units and guidelines that investigators can combine and modify according to specific situations.

Graphic representations can work synergistically with combinatorial approaches to design. For two-color systems, directed graphs, in which edges represent arrays and nodes represent samples, can represent any experimental design and provide an intuitive summary of the sample pairings. Unfortunately, the directed graph representation does not extend readily to three-color designs. In this paper, we propose a scheme for representing three-color designs graphically. In this scheme, nodes represent samples and ‘triangles’, which consist of three linked edges connecting three distinct vertices, represent arrays. The resulting images provide intuitive and accurate summaries of any three-color experimental design (Figs 1 and 2). This representation can be extended to four or more colors.

### 1.3 Cyclic designs

Cyclic designs are obtained by first defining an initial array. Such an initial array consists of an ordered list of samples to be hybridized and generates subsequent arrays by ‘cycling’ the samples in a regular pattern until the desired number of comparisons is generated (Kuehl, 2000). In the field of microarrays, loop designs belong to the class of cyclic designs. Cyclic designs have been studied extensively and are known to have many desirable properties (John, 1966, 1981; Kuehl, 2000). A complete characterization of cyclic designs is beyond the scope of this paper. Our aim is to provide a simple

---

**Fig. 1.** (A) Two-color simple loop design, \( v = 12, \ [(1, 2) \times 12, \ (2, 1) \times 12] \). (B) Two-color modified loop design, \( v = 12, \ [(1, 2) \times 12, \ (3, 1) \times 12] \). (C) Three-color simple loop design, \( v = 12, \ [(1, 2, 3) \times 12] \). (D) Three-color modified loop design, \( v = 12, \ [(1, 2, 4) \times 12] \). (E) Two-color reference design, \( v = 12 \). (F) Three-color reference design, \( v = 12 \).
set of rules that investigators can use to generate balanced, efficient and robust cyclic designs. The construction of cyclic designs is intuitive and easy, as illustrated in the examples provided here. The construction rules and additional examples appear in the Supplementary information.

2 METHODS

2.1 Calculation of efficiency of designs

In microarray experiments, the experimental performance of the system and the efficiency of the experimental design together determine the overall precision of detecting differential expression. The experimental performance is determined by technical factors that impact the precision of measurements. The efficiency of the experimental design is determined by the arrangement of the samples across arrays. The test statistics, $k_{i,j}$, for comparing two treatment groups $i$ and $j$ are usually in the following form

$$k_{i,j} \propto \frac{\hat{\tau}_i - \hat{\tau}_j}{V(\hat{\tau}_i - \hat{\tau}_j)} = \frac{\hat{\tau}_i - \hat{\tau}_j}{\hat{\sigma}^2 \times e_{i,j}},$$

where $\hat{\tau}_i - \hat{\tau}_j$ is the log scale expression difference, $\hat{\sigma}$ is residual (measurement) error and $e_{i,j}$ is a factor that captures the efficiency of the design (John, 1980). The experimental performance of the system is proportional to $(\hat{\tau}_i - \hat{\tau}_j)/\hat{\sigma}$, whereas the efficiency of the design is inversely proportional to $e_{i,j}$.

Average variance of a design can be defined as the mean of all pairwise variance

$$\frac{\sum V(\hat{\tau}_i - \hat{\tau}_j)}{v(v-1)/2} = \frac{2\hat{\sigma}^2}{v(v-1)} \sum e_{i,j}.$$ 

Efficiency, $E$, can be defined as the reciprocal of the ratio of average variance of the design being tested to that of complete randomized block design (CRBD) (John, 1980)

$$E = \frac{2/(v(v-1)) \sum V(\hat{\tau}_i - \hat{\tau}_j)_1}{\sum V(\hat{\tau}_i - \hat{\tau}_j)_2} = \frac{\hat{\sigma}_{1}^2 \sum (e_{i,j})_1}{\hat{\sigma}_{2}^2 \sum (e_{i,j})_2},$$

where $e_{i,j}$ can be calculated from the $C$-matrix or the information matrix of the design. Assuming $\hat{\sigma}_{1}^2 = \hat{\sigma}_{2}^2$, $E$ is

$$E = \frac{\sum (e_{i,j})_1}{\sum (e_{i,j})_2} = \frac{\sum_{k=2}^v 1/(\mu_{C1})_1}{\sum_{k=2}^v 1/(\mu_{C2})_2},$$

where $(\mu_{C1})_1$ and $(\mu_{C2})_2$ correspond to the non-zero eigenvalues of the information matrices from the CRBD and the design being tested, respectively. Intuitively, designs with high efficiency allocate treatments in each array so that the mean of variances of all treatment pairs is low. For CRBDs, the number of dyes in an array is equal to the number of treatments, $v$, and their efficiency is 1. For ‘incomplete block designs’ the former is greater than the latter, and $E$ is <1. Unless otherwise noted, efficiency is measured relative to CRBD. Details of calculating design efficiencies are found in John (1980), Kerr and Churchill (2001), and Das (2002).

2.2 Construction of cyclic designs for microarrays

The following example illustrates the construction of a cyclic design with $v = 4$ treatment groups using three-color microarrays. Cyclic designs with any number of colors, including two-color loop designs, can be generated in this way: the only requirements are to know the number of arrays and the initial array assignment. More details are provided in the Supplementary information.

(1) Set the number of arrays, $a$, to be used for the experiment to be a multiple of $v$. In this example, we use $a = 8$.

(2) Assign treatment groups to the initial array. We will assign treatment groups 1, 2 and 3 to the initial array in (dye 1, dye 2, dye 3). We use notation $A_1 = (1, 2, 3)$.

(3) Generate subsequent arrays by cyclic development, $A_1 = (1, 2, 3), A_2 = (2, 3, 4), A_3 = (3, 4, 1), A_4 = (4, 1, 2), A_5 = (1, 2, 3), A_6 = (2, 3, 4), A_7 = (3, 4, 1)$ and $A_8 = (4, 1, 2)$. For subsequent arrays, increase
treatment group numbers by increments of one while ‘cycling’ those greater than \( v \) back to 1.

The resulting design can be summarized using the notation \([A_1 \times a]: [(1,2,3) \times 8]\).

2.3 Comparative analysis of two-color and three-color experiments

2.3.1 Experimental design Three liver RNA samples were obtained from three treatment groups (one sample from each group) corresponding to three inbred mouse strains AJ(A), C57BL/6J(B) and the hybrid (C57BL/6J × AJ(J)(F1)) (Beck et al., 2000). Six three-color arrays were hybridized with 18 samples \((6 \times 3 = 18)\), and nine two-color arrays were hybridized with 18 samples \((9 \times 2 = 18)\). A total of 15 hybridizations (six from three-color array and nine from two-color array) were performed in two batches, one with eight and one with seven hybridizations. The designs are summarized in Figures 2A and B.

2.3.2 cDNA arrays A custom cDNA cloneset 6628 ESTs [from the freely available sequence verified clone sets, NIA (15K) and the BMAP (11K) containing non-overlapping GO or Locuslink annotated ESTs] was printed in four non-adherent replicates onto poly-lysine coated substrates using 50% dimethyl sulfoxide as a spotting solution for a spot diameter of 120 \(\mu\)m and a pitch of 170 \(\mu\)m. Printed arrays were then processed using a non-aqueous processing protocol (Diehl et al., 2001).

2.3.3 Labeling and hybridization Total RNA was converted into fluorescently labeled targets using T7 RNA polymerase in a single round linear amplification protocol (Gelder et al., 1990; Pabon et al., 2001). For each sample, three separate reactions were performed with each 5 \(\mu\)g of total RNA and the respective cRNA pooled. Aliquots of 5 \(\mu\)g of the cRNA per channel were then reverse transcribed in the presence of a random (8mer) primer and amino allyl-modified dUTP (Wang et al., 2003). After purification through Qiaquick column chromatography, aminoallyl groups were reacted with the respective Alexa dye esters (555, 594 or 647) for 1 h at room temperature (RT) in an argon environment. The fluorescently labeled cDNA was again purified via Qiaquick column chromatography, ethanol precipitated, dissolved in water and the respective samples combined. After heat denaturation and adjustment to 1 \(\times\) hybridization buffer, samples were hybridized in a microfluidics based chamber (BioMicro Inc.) for 16 h. After hybridization, the slides were washed with decreasing concentrations of SSC \((2 \times \rightarrow 0.2 \times)\) at RT and spin dried.

2.3.4 Data acquisition and importation Raw images were acquired in an environmentally controlled room immediately after post-hybridization processing of the slides using a ScanArray Express scanner (Perkin Elmer) and raw data were extracted from the images using the Imagene software package (Biodiscovery Inc.). Median intensity values (non-background subtracted) for each spot were imported into \( R \) environment (http://www.r-project.org).

2.3.5 Data normalization Data from two- and three-color experiments were normalized using R/MAANOVA package (Woo et al., 2004; Wu et al., 2003). Spatial variations and intensity-dependent biases were corrected using the joint (spatial and intensity) lowess transformation (Yang et al., 2002; Cui et al., 2003). Clones flagged as ‘bad’ at any of the replicated spots in either the two-color experiment or the three-color experiment were removed from the analysis. The four-spot replicates (within array) for each channel were collapsed to their median value. Data for each channel were mean centered prior to gene-level analysis.

2.3.6 Mixed model analysis Using R/MAANOVA, a linear mixed model was used to estimate variance components and expression level changes and to assess statistical significance of differential expression among the samples (Cui and Churchill, 2003; Wolfinger et al., 2001). Total variation was decomposed into batch \((B)\), array \((A)\) and measurement \((\epsilon)\) variance components. We express the gene specific model as

\[
y_{ij} = \mu + T_k + D_j + B_l + A_i + \epsilon_{ij}
\]

where \( y_{ij} \) is the normalized logarithmic intensity on array \( i \) in dye channel \( j \). The data for each gene are decomposed into the overall mean, \( \mu \); an effect of treatment groups, \( T_k \); an effect, \( D_j \); an array effect, \( A_i \); a batch effect, \( B_l \); and measurement errors, \( \epsilon_{ij} \). Array and dye are indexed by \( i \) and \( j \). The treatment index \( k \) is determined by the values of \( i \) and \( j \), and the batch index \( l \) is determined by the values of \( i \). \( B_l, A_i \) and \( \epsilon_{ij} \) were treated as random effects and variance components were estimated separately for each gene using methods described in Littell et al. (1996). The distribution of the variance components was analyzed graphically using smoothed histogram (Fig. 3). The \( F \)-statistic employing shrinkage variance components estimates was used to construct statistical tests (Cui et al., 2005). False discovery rates (FDRs) were calculated from the unadjusted \( P \)-values using \( q \)-value software (Storey and Tibshirani, 2003; http://faculty.washington.edu/storey/qvalue). For illustrative purposes, we focus only on the contrast \( A \) versus \( B \).

3 RESULTS

3.1 Properties of three-color and four-color loop designs For multicolor microarrays, cyclic designs provide a practical approach to experimental designs. We show here that efficiency of multicolor designs can be significantly greater than two-color designs. Cyclic designs balance treatment with respect to dye when the number of arrays is a multiple of the number of treatments, and the partial confounding of treatment and array is less than with two-color systems. These are desirable features of a design that eliminate or minimize
potential confounding of experimental artifacts with treatment effects. Cyclic designs for multicolor experiments are robust to accidental hybridization failures. The loss of efficiency is less than it would be for cyclic two-color arrays. The designs are easy to generate for any number of treatments.

3.1.1 Efficiency We compared the two-, three- and four-color loop designs in terms of their efficiency for different numbers of treatments. We found that three- and four-color loop designs have higher efficiency than their corresponding two-color loop designs when $v > 2$. Two-, three- and four-color simple loop designs: $[(1, 2) \times v], [(1, 2, 3) \times v]$ and $[(1, 2, 3, 4) \times v]$ are compared in Figure 4A. As $v$ increases, efficiency for all loop designs decreases. However, for three- and four-color designs, the rate of decline is less than for two-color designs. In general, three- and four-color loop designs have higher efficiency than two-color loop designs. Two-color $A$-optimal designs from Kerr and Churchill (2001) are less efficient than three-color loop designs. Fewer arrays are required for the same number of RNA samples when using multicolor designs. For example, when $v = 3$, a two-color loop design with six arrays and four replicates per treatment (12 samples) has the same efficiency as a three-color loop designs with three arrays with three replicates per treatment (nine samples). The difference in efficiency increases as $v$ increases.

3.1.2 Balance In some designs, such as the reference design, treatment and dye effects are confounded such that the two types of effects cannot be distinguished. Cyclic designs with the number of arrays equal to a multiple of the number of treatments are always balanced with respect to dye and thus avoid confounding completely. Also, treatment is always partially balanced with respect to array owing to the symmetry of the design, thus minimizing potential for confounding (Kuehl, 2000).

3.1.3 Robustness In microarray experiments, it is not uncommon for hybridizations to fail and thus it is important that a design should be robust enough to provide information after the removal of hybridization failures. Although two-color loop designs are very efficient compared with other two-color designs, they suffer a loss in overall efficiency from a single hybridization failure. However, the additional connectivity of multi-color loops provides robustness and hybridization failures lead to relatively small losses in efficiency. Figure 5 demonstrates that, while the two-color loop designs lose $>75\%$ of the power after loss of one array, three- and four-color loop designs retain $>50\%$. As $v$ increases, the robustness increases for three- and four-color designs, whereas it decreases for two-color designs.

3.2 Loop designs and reference designs

Simple loop designs are generated from initial arrays containing $n$ lowest treatment group numbers, i.e. $(1, 2)$ and $(1, 2, 3)$ for two- and three-color designs. Although it is straightforward to do so, it does not always yield efficient designs; different initial arrays can increase efficiency without any additional samples or arrays. In simple loop designs, we are comparing treatments ‘close’ to each other more efficiently than the ‘distant’ pairs. To achieve optimal efficiency, every pairwise comparison should occur equally often or as close
y = 0.25 and y = 1 indicate that the design is as efficient in treatment pairing as two-color reference design and CRBD. They are adjusted for the number of RNA samples used (not the number of arrays). (A) Plot of efficiency comparing two-, three- and four-color simple loop designs. The designs are [(1,2) x v], [(1,2,3) x v] and [(1,2,3,4) x v], respectively. The number of treatment groups ranges from 3 to 10, 3 ≤ v ≤ 10. (B) Plot of efficiency comparing two-color simple and ‘interwoven’ loop designs. The designs are [(1,2) x 2v] and [(1,2) x v, (1,3) x v] (Fig. 1A and B for v = 12). The number of treatment groups ranges from 4 to 50, 4 ≤ v ≤ 50. Two-color reference designs are generated by ‘dye-flipping’ to every samples to the reference sample. (Fig. 1E for v = 12). (C) Plot of efficiency comparing three-color simple and modified loop designs. The designs are [(1,2,3) x v] and [(1,2,4) x v] (Fig. 1C and D for v = 12). The number of treatment groups ranges from 4 to 50, 4 ≤ v ≤ 50. Three-color reference designs are generated by ‘looping’ every two test samples to the reference sample. (Fig. 1F for v = 12).

Fig. 4. Plots of efficiency. y represents efficiency relative to CRBD. y = 0.25 and y = 1 indicate that the design is as efficient as possible. In two-color designs, ‘interweaving’ the loops by adding arrays comparing ‘distant’ treatments increases the efficiency per RNA samples used (Oleksiak et al., 2002; Kerr and Churchill, 2001; Vinciotti et al., 2005). In Figure 4B, the efficiency of interwoven loops decreases less rapidly than for simple loops as v increases. These loop designs are more efficient than two-color reference designs for v ≤ 10 and v ≤ 23. The efficiency of multicolor designs can be increased without adding more arrays by carefully choosing the initial array to provide connectivity similar to the interwoven loops. The selection of optimal initial array can be guided by graphical diagrams and it is easy to achieve reasonably efficient designs. In Figure 4C, modified three-color loop designs are seen to be more efficient than simple three-color loop designs. The design with initial array (1,2,4) spreads out the comparisons more evenly than does the design with initial array (1,2,3), resulting in a higher overall efficiency for the design with (1,2,4). Loop designs with initial arrays (1,2,3) and (1,2,4) are more efficient than three-color reference designs for v ≤ 18 and v ≤ 38, respectively.

3.3 Comparative analysis of two-color and three-color experiments

3.3.1 Diagnostics Most spatial- and intensity-based biases were removed by joint (spatial and intensity) lowess (data not shown). The visual inspection of the xy-scatterplots of the log ratios indicates that both two-color and three-color experiments exhibit a good reproducibility between technical replicates (see Supplementary information).

3.3.2 Concordance of gene expression changes and statistical significance In lieu of a ‘gold standard’, the agreement between two platforms is used as a surrogate marker for accuracy. Our analysis shows an excellent agreement between the two-color and three-color experiments in terms of log expression changes and the statistical significance. In Figure 3A–C, estimated log ratios between all three treatment pairs, A versus B, A versus F1 and B versus F1, are highly correlated with correlation coefficient ranging from 0.948 to 0.959. The concordance of statistical significance and the direction of gene expression changes are very high (Table 1).

3.3.3 Experimental performance: dynamic range and precision The experimental performance of the two-color experiment, indicated by the dynamic range and measurement precision, is slightly higher than that of the three-color experiment. Standard deviations (SD) of estimated gene-expression level indicate that the two-color microarrays have a greater dynamic range. The median SD of the two-color experiment is 8.1% higher than that of the three-color experiment (0.078 and
Table 1. A table of comparing statistical significance and agreement of gene expression changes

<table>
<thead>
<tr>
<th>α</th>
<th>FDR three-color (%)</th>
<th>FDR two-color (%)</th>
<th>Significant three-color</th>
<th>Significant two-color</th>
<th>None</th>
<th>Three-color only</th>
<th>Two-color only</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>6.97</td>
<td>7.17</td>
<td>1176</td>
<td>1184</td>
<td>807 (483)</td>
<td>232 (187)</td>
<td>240 (203)</td>
<td>944 (941)</td>
</tr>
<tr>
<td>0.05</td>
<td>4.14</td>
<td>4.12</td>
<td>999</td>
<td>1024</td>
<td>1023 (663)</td>
<td>176 (154)</td>
<td>201 (174)</td>
<td>823 (823)</td>
</tr>
<tr>
<td>0.01</td>
<td>1.18</td>
<td>1.20</td>
<td>700</td>
<td>706</td>
<td>1392 (988)</td>
<td>125 (123)</td>
<td>131 (128)</td>
<td>575 (575)</td>
</tr>
<tr>
<td>0.001</td>
<td>0.20</td>
<td>0.23</td>
<td>412</td>
<td>369</td>
<td>1747 (1338)</td>
<td>107 (107)</td>
<td>64 (64)</td>
<td>305 (305)</td>
</tr>
<tr>
<td>0.0001</td>
<td>0.04</td>
<td>0.05</td>
<td>207</td>
<td>160</td>
<td>1997 (1588)</td>
<td>66 (66)</td>
<td>19 (19)</td>
<td>141 (141)</td>
</tr>
</tbody>
</table>

Estimates of FDRs in percentage, number of significant genes, overlapping genes in four combinations (none, two-color only, three-color only and both two-color and three-color) and their concordance in direction of gene expression change (in paranthesis) are shown for five critical $P$-values ($\alpha$). The two experiments have similar numbers of statistically significant genes. The agreements of direction of change for genes significant in both experiments are close to 100% for all critical $P$-values.

0.072, respectively). In Figure 3D, the distribution of residual variance components indicates that the two-color microarrays are performing slightly better the three-color microarrays in terms of the precision of measurements. The median $\hat{\sigma}$ of the two-color experiment is 15% less than that of three-color experiment (0.062 and 0.073, respectively).

3.3.4 Degrees of freedom The degrees of freedom to estimate error variance for the three-color and the two-color experiment are 8 and 6, respectively.

3.3.5 Overall performance to power to detect differential expression The overall power of the system to detect differential expression is similar for both two-color and three-color experiments. Figure 6 shows that the distribution of $P$-values and their FDRs are similar to each other, indicating that the two experiments have comparable power to detect differential expression.

4 DISCUSSION

4.1 Design efficiency and overall performance

Although we expected more power to detect differential expression from the three-color experiment than the two-color experiment based solely on design efficiency, we found them to be similar to each other. The improved design efficiency is offset by a slight decrease in experimental performance of the three-color system. As $v$ increases, the design efficiency of three-color loop designs increases relative to that of two-color loop designs whereas the relative experimental performance remains constant. Therefore, in terms of the overall power to detect differential expression, three-color designs would outperform two-color designs for larger $v$.

4.2 Independence assumption in efficiency calculation

In our efficiency calculation, all effects are considered to be fixed and all observations are considered to be independent out of practical considerations. Although a mixed model can decompose total variances into biological variances and experimental variances from different stages of the experiment, such as extraction, cDNA synthesis and labelling step, several practical difficulties limit its usefulness. First, the calculation of the design efficiency under mixed model assumption can be challenging. In fact, most of the literature in the field of design efficiency is based on a fixed model assumption (Kerr and Churchill, 2001). Second, some literature based on the mixed model assumption suggest that the array or block effects do not make drastic differences (Mukhopadhyay, 1984; Bhattacharya and Shah, 1984; Tempelman, 2005). For example, in Tempelman (2005), there was little difference between treating array
effects as random or fixed effects. Third, the efficiency calculations based on a mixed model required a priori knowledge of the size of each variance component. In the same work of Tempelman (2005), the ratio of the biological variance to the measurement error variance made a significant difference. Many experimental factors and biological factors together determine the relative size of each component and a carefully planned, large-scale experiment will be necessary.

4.3 Biological and technical variation

Our experiment, consisting of technical replicates only, was more suitable for meeting our objectives than those with biological replicates. To focus on the difference between two- and three-color systems while eliminating experimental variations common to both, samples have been pooled appropriately. For testing biological hypotheses, pooling is not generally recommended, and experiment with multiple biological replicates, rather than technical replicates, is essential for drawing biological conclusions.

4.4 Accuracy and agreement

Although it would be ideal to compare the result from three-color microarray experiments with known ‘truth’, such a priori knowledge for most biological systems does not exist. Since two-color experimental system is a more established platform, the concordance between two- and three-color microarray experiments has been used as a surrogate marker for the accuracy of three-color microarray platforms (Woo et al., 2004). Given the high similarity between the two systems, there is no reason to believe that one is more accurate than the other.

4.5 The optimal number of dyes per array

If adding additional dyes results in higher efficiency, what is the optimal number of dyes to be used? Although it is theoretically true that each additional color always results in higher efficiency, from a practical point of view, if too many dyes are hybridized per array, the experiment would be operationally difficult and may be impractical. Furthermore, the marginal gain in efficiency from adding additional dyes always decreases. In other words, the gain in efficiency from using three-color over two-color loop designs is greater than that from using four-color over three-color loop designs. With the use of more colors, the advantage of reducing variability by blocking decreases (Patterson and Williams, 1976). Considering operational difficulties already associated with two-color microarrays, three dyes for an array may be the magic number in terms of practicality and power.

5 CONCLUSION

Three-color microarray experiments that are balanced, efficient and robust can be easily designed with the loop design scheme, we have proposed. But for practical reasons, three-color loop designs with initial array (1, 2, 4) should be used over reference designs and two-color loop designs.

ACKNOWLEDGEMENTS

The authors are thankful to Sarah Williamson for graphics assistance. The research was supported by NIH Grant CA88327 and Cancer Signature Program of the University of Connecticut Medical School. Y.H.W. is supported by a National Science Foundation IGERT training grant (0221625) to The University of Maine (Knowles, PI).

REFERENCES


Three-color and four-color microarray experimental design


