Direct versus indirect designs for cDNA microarray experiments

Terence P. Speed[1,2] and Yee Hwa Yang[1]

April 18, 2002

Technical report # 616, April 2002

1. Department of Statistics, University of California, Berkeley.
2. Division of Genetics and Bioinformatics, The Walter and Eliza Hall Institute, Australia.

Address for correspondence:
Terence P. Speed
Department of Statistics
University of California, Berkeley
367 Evans Hall
Berkeley, CA 94720-3860
Tel: (510) 642-2781
Fax: (510) 642-7892
E-mail: terry@stat.berkeley.edu

Abstract

We calculate the variances of two classes of estimates of differential gene expression based on log ratios of fluorescence intensities from cDNA microarray experiments: direct estimates, using measurements from the same slide, and indirect estimates, using measurements from different slides. These variances are compared and numerical estimates are obtained from a small experiment involving 4 slides. Some qualitative and quantitative conclusions are drawn which have potential relevance to the design of cDNA microarray experiments.

AMS Subject Classification, Primary: 92B15; Secondary: 62H99, 62J10

Keywords: cDNA microarray analysis; differential gene expression ; experimental design; common reference design; technical replicates.

1 Introduction

Microarray experiments measuring the expression of thousands of genes generate large and complex multivariate datasets. Much effort has been devoted to the pre-processing and higher-level analysis of such data, but attention to statistical design is also important if we wish to improve the efficiency and reliability of these experiments.

This paper concerns one very basic design issue: the relative precision of two classes of estimates of differential gene expression in cDNA microarray experiments. With this technology all measurements are paired comparisons, that is, measurements of relative gene expression, with microscope slide playing the role of the block of two units. Below we give a brief explanation of the technology. As with the mixed model for block experiments, the two measurements of a gene's expression on the same slide are correlated, usually quite highly so, and reliance is principally on differences (we use a log scale), and much less on the individual measurements. However, an important difference between block experiments and cDNA experiments is that with the latter, we can have correlations between measurements on different slides. Such correlations arise when the cDNA samples in question are what is known as *technical replicates.*

Suppose that we wish to compare the expression of a gene in two samples A and B of cells. We could compare them on the same slide, i.e. in the *same* microarray experiment, in which case a measure of the gene's differential expression could be $log_2 A/B$, where $log_2 A$ and $log_2 B$ are measures of the gene's expression in samples A and B. We will refer to this as a *direct* estimate of differential expression, direct because the measurements come from the same slide. Alternatively, $log_2 A$ and $log_2 B$ may be estimated in two *different* experiments, A being measured together with a third sample R and B together with a similar sample R', on two different slides. The log ratio $log_2 A/B$ will in this case be replaced by the difference $log_2 A/R - log_2 B/R'$, and we call this an *indirect* estimate of the gene's differential expression, as it is calculated with values $log_2 A$ and $log_2 B$ from different experiments. We note that samples R and R' of the third (reference) cDNA will, in general be technical replicates. Similarly, if the direct comparison of A with B is replicated, then the replicate samples A' and B' are likely to be technical replicates.

Our aim in this paper is to compare the precision of direct and indirect estimates of differential gene expression, both in theory, and on the basis of some experimental data. We begin in Section 2 with a brief introduction to the biology and technology of cDNA microarrays. After outlining recent research applying optimal design principles to microarray experiments, we present the main result of this paper, which compares the variances of direct and indirect estimates of differential expression based on the same number of slides. The novelty of our calculation shown in Section 3 lies in the more realistic covariance matrix we use, incorporating covariances between measurements on the same slide, between measurements on technical replicates across different slides, and between measurements on unrelated samples on different slides. In Section 4 we use a data set comparing two samples to estimate average values of the variances and covariances in our calculation, and hence to illustrate our results. Finally, Section 5 summarizes our findings and discussed the implications for designing more complex cDNA microarray experiments.

2 Background on DNA microarrays

DNA microarrays are part of a new class of biotechnologies which allow the monitoring of expression levels for thousands of genes simultaneously. In addition to the enormous scientific potential of microarrays to help in understanding gene regulation and gene interactions, microarrays are being used increasingly in pharmaceutical and clinical research. Our focus here is on complementary DNA (cDNA) microarrays, where thousands of distinct DNA sequences representing different genes are printed in a high-density array on a glass microscope slide using a robotic arrayer. The relative abundance of each of these genes in two RNA samples may be estimated by fluorescently labeling the two samples, mixing them in equal amounts, and hybridizing the mixture to the sequences on the glass slide. More fully, the two samples of messenger RNA from cells (known as *target*) are reverse-transcribed into cDNA, and labeled using differently fluorescing dyes (usually the red fluorescent dye cyanine 5 and the green fluorescent dye cyanine 3). The mixture then reacts with the arrayed cDNA sequences (known as *probes* following the definitions adopted in NGS (1999)). This chemical reaction, known as competitive hybridization, results in complementary DNA sequences from the targets and the probes base-pairing with one another. The slides are scanned at wavelengths appropriate for the two dyes, giving fluorescence measurements for each dye for each spot on the array. The two absolute fluorescence intensities for any spot should be proportional to the amount of mRNA from the corresponding gene in the respective samples, and the fact that these can be obtained simultaneously for 10-40 thousand genes gives this assay its great power. In practice the absolute intensities are usually noticeably less reproducible across slides than their ratios, and for the most part, the assay read out is the set of (log) ratios. We refer the reader to NGS (1999) for a more detailed introduction to the biology and technology of cDNA microarrays.

3 Experimental design: direct vs indirect comparisons

Proper statistical design is desirable to get the most out of microarray experiments and to ensure that the effects of interest to biologists are accurately and precisely measured. Careful attention to experimental design will ensure that the best use is made of available resources, obvious biases will be avoided, and that the primary questions of interest to the experimenter will be answerable. To date the main work on design with microarray experiments is due to Kerr and Churchill (2001) and

Glonek and Solomon (2002), who have applied ideas from optimal experimental designs to suggest efficient designs for some of the common cDNA microarray experiments. In these papers, and in other similar calculations in the literature, log ratios from different experiments are regarded as statistically independent.

Most biologists conducting simple gene expression comparisons such as between treated and control cells, will carry out replicate experiments on different slides. These will usually involve what are known as *technical replicates*, where we use this term to describe the case where the target mRNA in each hybridization is from the same RNA extraction, but is labeled independently for each hybridization. A more extreme form of technical replication would be when samples from the same extraction and labeling are split, but we do not know of many labs now doing this, though some did initially. We and others have noticed that estimates of differential gene expression based on technical replicates tend to be correlated, whereas the same estimates based on replicates involving different RNA extractions and labellings tend to be uncorrelated. When the more extreme form of technical replication is used, the correlation can be very strong. These observations have led us to re-examine the correlation structure underlying experimental design calculations. When variances are calculated for linear combinations log ratios across replicate slides, it seems desirable to use the most realistic covariance model for the measurements, and this is what we now try to do.

3.1 A simple calculation

Let us consider comparisons between two target samples A and B. For gene i on a typical slide, we denote the intensity value for the two target samples by A_i and B_i. The log base 2 transformation of these values will be written $a_i = \log_2 A_i$ and $b_i = \log_2 B_i$, respectively, and when reference samples R and R' are used, we will write $r_i = log_2 R_i$ and $r'_i = log_2 R'_i$. In addition, we denote the means of the log-signals across slides for gene i by $\alpha_i = E(a_i)$ and $\beta_i = E(b_i)$, respectively. The variances and covariances of the log signals for gene i across slides will be assumed to be the same for all samples, that is, we suppose that differential gene expression is exhibited only through mean expression levels, and we always view this on the log scale. Our dispersion parameters are a common variance σ_i^2, a covariance c_{1i} between measurements on samples from the same hybridization, a covariance c_{2i} between measurements on technical replicate samples from different hybridizations, and a covariance c_{3i} between measurements on samples which are neither technical replicates nor in the same hybridization.

** Place Figure 1 approximately here **

Figure 1 illustrates two different designs for microarray experiments whose purpose is to identify genes differentially expressed between the two target samples A and B. Design I involves two direct comparisons, where the samples A and B are hybridized together on the same slides. Design II illustrates an indirect comparison, where A and B are each hybridized with a common reference sample R. We denote technical replicates of A, B and R by A', B' and R' respectively. Note that both designs involve two hybridizations, and we emphasize that in both cases, our aim is to estimate the expression difference $\alpha_i - \beta_i$ on the log scale. We now calculate the variances of the obvious estimates of this quantity from each experiment. For Design I this is one half of $y_i = a_i - b_i + a'_i - b'_i$.

We have

$$v_{1i} \;=\; var(y_i/2) \;=\; \sigma_i^2 - c_{1i} + c_{2i} - c_{3_i},$$

while for Design II we have

$$v_{2i} \;=\; var(a_i - r_i - b_i + r_i') \;=\; 4(\sigma_i^2 - c_{1i}) - 2(c_{2i} - c_{3i}).$$

We next show that v_{1i} is never greater than v_{2i}. To see this, consider the covariance matrix for the four log intensities from Design I:

$$Cov\begin{pmatrix} a_i \\ b_i \\ a_i' \\ b_i' \end{pmatrix} = \begin{bmatrix} \sigma_i^2 & c_{1i} & c_{2i} & c_{3i} \\ c_{1i} & \sigma_i^2 & c_{3i} & c_{2i} \\ c_{2i} & c_{3i} & \sigma_i^2 & c_{1i} \\ c_{3i} & c_{2i} & c_{1i} & \sigma_i^2 \end{bmatrix}.$$

It is easy to check that the eigenvalues of this matrix are $\lambda_{1i} = \sigma_i^2 + c_{1i} + c_{2i} + c_{3i}$, $\lambda_{2i} = \sigma_i^2 + c_{1i} - c_{2i} - c_{3i}$, $\lambda_{3i} = \sigma_i^2 - c_{1i} + c_{2i} - c_{3i}$, and $\lambda_{4i} = \sigma_i^2 - c_{1i} - c_{2i} + c_{3i}$. In terms of these eigenvalues, we see that $v_{1i} = \lambda_{3i}$ and that $v_{2i} = \lambda_{3i} + 3\lambda_{4i}$. Thus the relative efficiency of the indirect versus the direct design for estimating $\alpha_i - \beta_i$ is

$$\frac{v_{2i}}{v_{1i}} \;=\; 1 + \frac{3\lambda_{4i}}{\lambda_{3i}}. \tag{1}$$

The direct design is evidently never less precise than the indirect one, and the extent of its advantage depends on the values of σ_i^2, c_{1i}, c_{2i} and c_{3i}. Notice that when $\lambda_4 = 0$ ($\sigma_i^2 + c_{3i} = c_{1i} + c_{2i}$), we see that $v_{1i} = v_{2i}$. This shows that under our more general model, the reference design *can* be as efficient as the direct design. At the other extreme, when $c_{2i} = c_{3i}$, that is when the covariance between measurements on technical replicates coincides with that between any two unrelated samples, we have $v_{2i} = 4v_{1i}$. This is the conclusion which is obtained when log ratios from different experiments are supposed independent, and while this is roughly the case in general, we will see that the story is somewhat more complicated. Note that there is no reason in principle why we can't have $c_{2i} < c_{3i}$, equivalently, $\lambda_{3i} < \lambda_{4i}$, in which case $v_{2i} > 4v_{1i}$.

The generic calculation we have just presented is gene-specific. However, it is quite difficult to obtain data permitting estimates these parameters for all 20,000 genes, and so we now turn to finding typical values for them.

3.2 Estimation of v_1 and v_2

In the preceding calculation, our analysis focused on a single gene. In practice we would not design to achieve highest efficiency for a single gene, indeed we would not usually have data of the requisite kind to enable us to estimate the variance and covariances at the single gene level. What we do have in abundance are data which embody the variability and covariability underlying these gene specific variances and covariances *en masse*. It seems reasonable to hope that we can use these data to give us typical or average values of λ_{3i} and λ_{4i} across a gene set. These values could then be used as general guides in a design context.

How can we calculate average values $\bar{\lambda}_3$ and $\bar{\lambda}_4$? We show how this can be done with data from experiments carried out according to Design I, subject to an assumption which, while plausible, is uncheckable at this point. Continuing the notation introduced the previous section, we write $x_i = a_i + b_i - a'_i - b'_i$, $y_i = a_i - b_i + a'_i - b'_i$, and $z_i = a_i - b_i - a'_i + b'_i$.

The following identity shows us how we can connect the averages of gene-specific eigenvalues to quantities we are able to estimate. Note that $E(x_i) = E(y_i) = E(z_i) = 0$, while $var(x_i), var(y_i)$ and $var(z_i)$ are $4\lambda_{2i}, 4\lambda_{3i}$, and $4\lambda_{41}$, respectively. If we suppose that their covariances across i and j are, on average negligible in relation to their average variance, then we get what we want. More fully, if $cov(x_i, x_j) = \gamma_{ij}$, then

$$E\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\} \quad = \quad \frac{4}{n}\sum_{i=1}^{n}\lambda_{2i} - \frac{1}{n(n-1)}\sum_{i \neq j}\gamma_{ij} \quad \approx \quad 4\bar{\lambda}_2. \tag{2}$$

This formula, and its analogue involving the y_i and z_i, will be the basis of our estimates of $\bar{\lambda}_2, \bar{\lambda}_3$ and $\bar{\lambda}_4$. We are unable to present any justification for ignoring the average covariance terms, apart from the fact that the numerical results given below make sense.

4 Illustration

We will illustrate our method to find an approximate for values v_1 and v_2 from the *swirl* experiment provided by Katrin Wuennenberg-Stapleton from the Ngai Lab at UC Berkeley. (The *swirl* embryos for this experiment were provided by David Kimelman and David Raible at the University of Washington.) This experiment was carried out using zebrafish as a model organism to study early development in vertebrates. *Swirl* is a point mutant in the BMP2 gene that affects the dorsal/ventral body axis. Ventral fates such as blood are reduced, whereas dorsal structures such as somites and notochord are expanded. A goal of this *swirl* experiment was to identify genes with altered expression in the *swirl* mutant compared to wild-type (*wt*) zebrafish. The datasets consists of 4 replicate slides (2 sets of dye-swap experiments). For each of these slides, target cDNA from the *swirl* mutant is labeled using one of the cyanine 3 or cyanine 5 dyes and the target cDNA wild-type mutant is labeled using the other dye. The two sets of dye-swaps were performed on two different days. Within each set, the same *swirl* target samples on the two different slides are technical replicate of each other, and similarly for the wild-type (*wt*) samples. Figure 2 shows a graphical representation of this *swirl* experiment.

** Place Figure 2 approximately here **

As described in section 2, the raw data from a cDNA microarray experiments consist of pairs of image files, one for each of the dyes. Careful analysis is required to extract reliable measures of the fluorescence intensities from each image. In addition, a preprocessing step known as normalization is required to allowed comparison across spots between slides. For a typical slide let us denote gene i's log intensity values for the two target samples by $r_i = \log_2 R_i$ and $g_i = \log_2 G_i$. Our method of normalization adjust the r_i and g_i intensity to $r_i^* = r_i - \frac{1}{2}k_i$ and $g_i^* = g_i + \frac{1}{2}k_i$ respectively, where k_i is the normalization adjustment estimated from "print-tip loess" normalization within each slides (Yang et al. (2002)). We are currently developing more sophisticated multiple experiment normalization procedures, but the details of this work are beyond the scope of this paper.

As shown above, we are able to provide noisy estimates of eigenvalues of the covariance matrix for every gene i. While the quantities x_i, y_i and z_i should have mean zero, outliers are a concern. Thus we took the union of the 1,000 most extreme genes from each sample across the four different experiments and removed them from the data before doing our estimation. The removal of $2,817$ spots from the samples of $8,448$ spots provides us with a conservative dataset where spots are reasonably constantly expressed across all experiments. Table 1 presents the estimates of $\bar{\lambda}_2$, $\bar{\lambda}_3$ and $\bar{\lambda}_4$. Using them we can estimate the relative efficiency v_2/v_1 of the indirect and direct designs for estimating $\log_2(swirl) - \log_2(wt)$. We find that it is 4 for the experiments of set 1 and 2.5 for set 2.

Table 1: Estimates of $\bar{\lambda}_2$, $\bar{\lambda}_3$, $\bar{\lambda}_4$ and relative efficiency v_2/v_1 for the *swirl* experiment.

	Set 1		Set 2	
Parameters	Median	Mean	Median	Mean
$\bar{\lambda}_2$	0.06	0.42	0.09	0.33
$\bar{\lambda}_3$	0.01	0.02	0.02	0.04
$\bar{\lambda}_4$	0.01	0.02	0.01	0.02
$1+3\bar{\lambda}_4/\bar{\lambda}_3$	4	4	2.5	2.5

5 Discussion

When we average log ratios, as we do in Design I, we want the terms to be as independent as possible. When we take differences, as we do in Design II, we want the technical replicate terms to be as dependent as possible. This can be achieved by using the same extraction and the same labeling (extreme technical replication). The above results give us an indication of the extent to which this is possible, and would seem to have some implications for experimental design with cDNA microarrays. We turn now to some special issues.

Log-intensity correlations

Our analysis and results throw some light on the correlations we see in the scatter plots of log intensity values within and between slides. Three types of plot are possible, see Figure 3: the familiar one a versus b of log intensities within a slide, and two less commonly presented ones: between log intensities from technical replicates across slides, a versus a', and between log intensities from unrelated samples across slides from the same print batch, a versus b'. The correlation in the first of these is very high indeed, while that in the other two is lower, but still high, and not visibly different across the pair.

Of course the strength of the correlation in these plots is largely driven by the generally similar pairs α_i and β_i of mean log intensities. In order to make this more precise, and to see the role of our covariances c_{1i}, c_{2i} and c_{3i} we use different versions of our earlier identity. The first is

$$E\{\frac{1}{n-1}\sum_{i=1}^{n}(a_i-\bar{a})(b_i-\bar{b})\} \approx \frac{1}{n-1}\sum_{i=1}^{n}(\alpha_i-\bar{\alpha})(\beta-\bar{\beta})+\bar{c}_1 ,$$

where $\bar{c}_1$ is the average of the c_{1i}, and again we have ignored certain average covariance terms. Similarly, if we replace b_i in the above by a'_i, we get

$$E\{\frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a})(a'_i - \bar{a'})\} \approx \frac{1}{n-1}\sum_{i=1}^{n}(\alpha_i - \bar{\alpha})^2 + \bar{c}_2,$$

while replacing a'_i by b'_i gives

$$E\{\frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a})(b'_i - \bar{b'})\} \approx \frac{1}{n-1}\sum_{i=1}^{n}(\alpha_i - \bar{\alpha})(\beta_i - \bar{\beta}) + \bar{c}_3.$$

The first term in all three of these should be approximately the same, and if this is the case, we can conclude that $\bar{c}_1$ is noticeably larger than $\bar{c}_2$ and $\bar{c}_3$, while the latter are approximately equal.

*** Place Figure 3 approximately here. ***

Log-ratio correlations

We turn now to the issue which prompted this analysis: the extent to which log ratios, that is, measures of differential gene expression, are correlated across experiments. It is easy to check that the covariance between estimates of differential expression $a_i - b_i$ and $a'_i - b'_i$, based on technical replicates is $2(c_{2i} - c_{3i})$. Similarly, the covariance between $a_i - a'_i$ and $b_i - b'_i$ is $2(c_{1i} - c_{3i})$. The scatter plots of these quantities are shown as panels (b) and (c) of Figure 4. Panel (a) of this figure is the scatter plot of the pairs $a_i - b'_i$ versus $a'_i - b_i$, which individually should have covariance $2(c_{2i} - c_{1i})$. Once more making use of identities similar to the three just above, we conclude that $\bar{c}_1$ is noticeably larger than $\bar{c}_2$, which in turn is slightly larger than $\bar{c}_3$. This is all consistent with the conclusions of the previous subsection. Note that we are unable to estimate $\bar{c}_2$, $\bar{c}_2$ and $\bar{c}_3$ individually, just certain differences: $\bar{c}_1 - \bar{c}_3 \approx 0.19$ and $\bar{c}_2 - \bar{c}_3 \approx 0.04$.

*** Place Figure 4 approximately here. ***

Next we consider the covariance structure of log ratios, which all have variance $\sigma^2_{wi} = var(a_i - b_i) = 2(\sigma^2_i - c_{1i})$. Formulae for their covariances depend on whether they have 0, 1 or 2 technical replicates in common. Denoting these three covariances by d_{0i}, d_{1i} and d_{2i}, we have the following: $d_{0i} = 0$, $d_{1i} = c_{2i} - c_{3i}$ and $d_{2i} = 2(c_{2i} - c_{3i})$. The preceding analysis, coupled with an examination of Figures 3 and 4, should make it quite clear why we use log ratios of intensities and not the log intensities on their own, in the analysis of cDNA experiments.

Implications for design

We close with a few remarks on the implications of the foregoing for the design of more complex cDNA microarray experiments such as factorials, Glonek and Solomon (2002). A more complete analysis would include the covariances just discussed. Let us extend the design depicted in Figure 1(a) to include samples A^* and B^* which are treated forms of samples A and B, e.g. with some drug. Our interest might then lie in genes whose differential expression between A and B is *different* from that between A^* and B^*, that is, in genes whose expression exhibits a sample by treatment *interaction*.

** Place Figure 5 approximately here **

Figure 5 is a representation of this 2×2 factorial experiment, the number next to the arrows in the diagram being the slide number. The covariance matrix of the parameter estimates can then be calculated using the formula $(X'\Sigma^{-1}X)^{-1}$. In this example, the design matrix and the gene-specific covariance matrix for log ratios are as follows:

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \text{ and } \Sigma_i = Cov \begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \\ y_{5i} \\ y_{6i} \end{pmatrix} = \begin{pmatrix} \sigma^2_{wi} & d_{1i} & d_{1i} & 0 & d_{1i} & d_{1i} \\ d_{1i} & \sigma^2_{wi} & 0 & d_{1i} & d_{2i} & d_{1i} \\ d_{1i} & 0 & \sigma^2_{wi} & d_{2i} & d_{1i} & d_{2i} \\ 0 & d_{1i} & d_{1i} & \sigma^2_{wi} & d_{1i} & d_{2i} \\ d_{1i} & d_{1i} & d_{1i} & d_{1i} & \sigma^2_{wi} & d_{2i} \\ d_{1i} & d_{1i} & d_{2i} & d_{2i} & 0 & \sigma^2_{wi} \end{pmatrix}.$$

The relative efficiencies of different designs for estimating an interaction parameter can now be calculated, but we leave the details for another time.

Acknowledgments

First and foremost, TPS wishes to acknowledge the friendship and advice received from Dev Basu over the years. His ability to make statistical theory and thought experimentation both instructive and enjoyable was unparalleled. He is greatly missed. We would also like to thank David Kimelman and David Raible at the University of Washington for providing the *swirl* embryos and Katrin Wuennenberg-Stapleton from the Ngai Lab at UC Berkeley for performing the *swirl* microarray experiment. We are also very grateful to members of the Ngai Lab at UC Berkeley and Natalie Thorne from University of Melbourne for discussion on design questions arising in microarray experiment. This work was supported in part by the NIH through grants 8R1GM59506A(TPS) and 5R01MH61665-02(YHY).

References

(1999). *The Chipping Forecast*, volume 21. Supplement to *Nature Genetics*.

Glonek, G. F. V. and Solomon, P. J. (2002). Factorial designs for microarray experiments.

Kerr, M. K. and Churchill, G. A. (2001). Experimental design for gene expression microarrays. *Biostatistics*, 2:183–201.

Yang, Y. H., Dudoit, S., Luu, P., Lin, D. M., Peng, V., Ngai, J., and Speed, T. P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Research*, 30(4):e15.

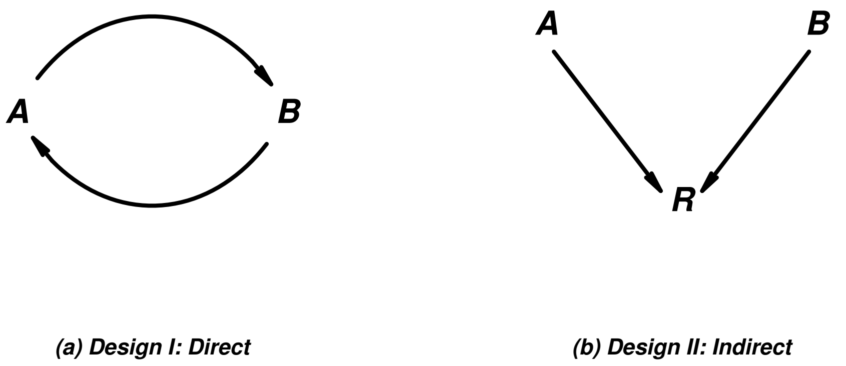

Figure 1: Two possible designs comparing the gene expression in two samples A and B of cells. In this representation, *vertices* correspond to target mRNA samples and *edges* to hybridizations between two samples. By convention, we place the green-labeled sample at the tail and the red-labeled sample at the head of the arrow. (a) *Direct comparison:* this design measures the gene's differential expression in samples A and B directly on the same slide (experiment). (b) *Indirect comparison:* this design measures the expression levels of samples A and B separately on two different slides and estimates the log ratio $\log_2 A/B$ by the difference $\log_2 A/R - \log_2 B/R'$, where the samples R and R' are technical replicate reference samples

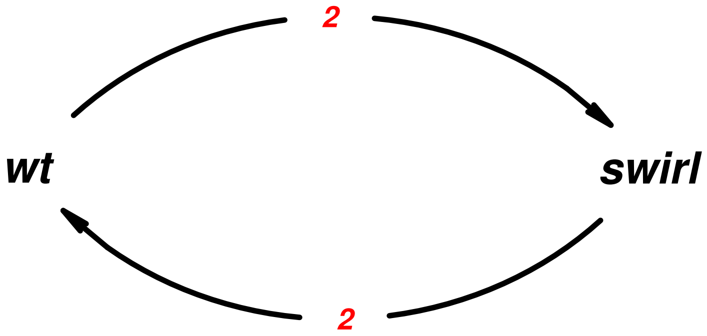

Figure 2: The *swirl* experiments provided by Katrin Wuennenberg-Stapleton from the Ngai Lab at UC Berkeley. This experiment consists of two sets of dye swap experiments comparing gene expression between the mutant *swirl* and wildtype (*wt*) zebrafish. The number on the arrow represents the number of replicated experiments.

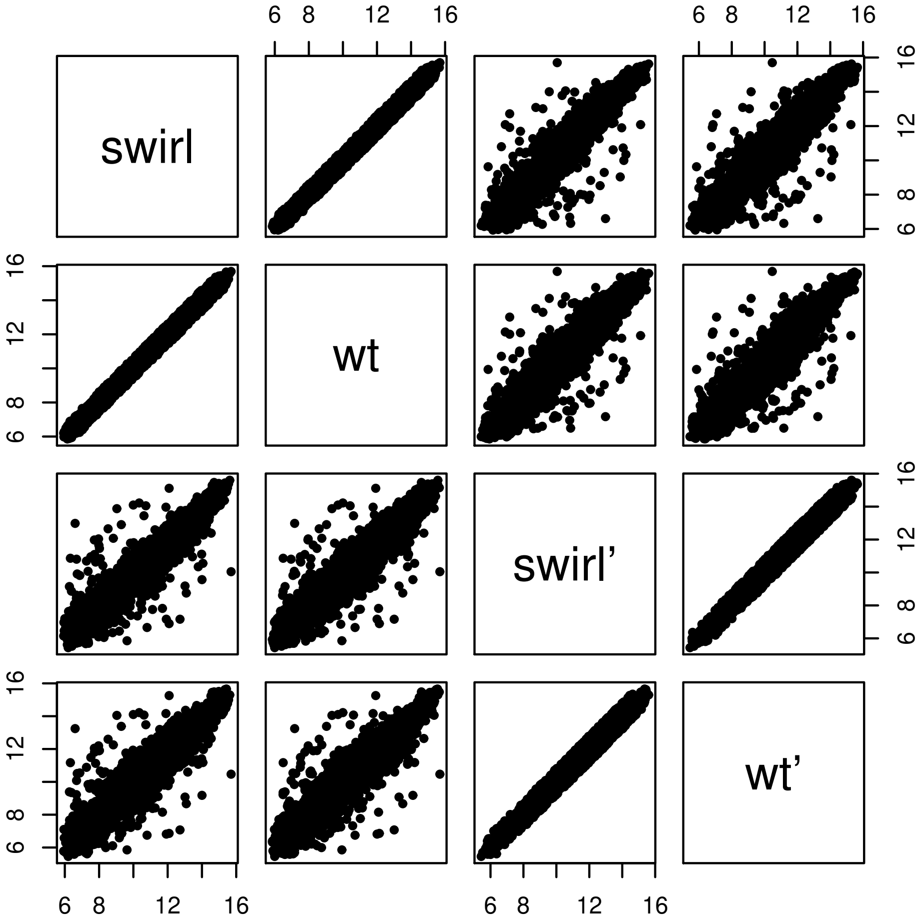

Figure 3: Pairwise scatter plots for set 2 of the *swirl* experiment. This is a set of dye-swap experiment where samples *swirl* and *wt* are compared on the same microarray experiment (i.e. same slide) and another set of samples *swirl'* and *wt'* are compared on another microarray experiment. The samples *swirl* and *swirl'* are *technical replicates*, where the both samples mRNA is from the same RNA extraction, but is labeled independently for each hybridization.

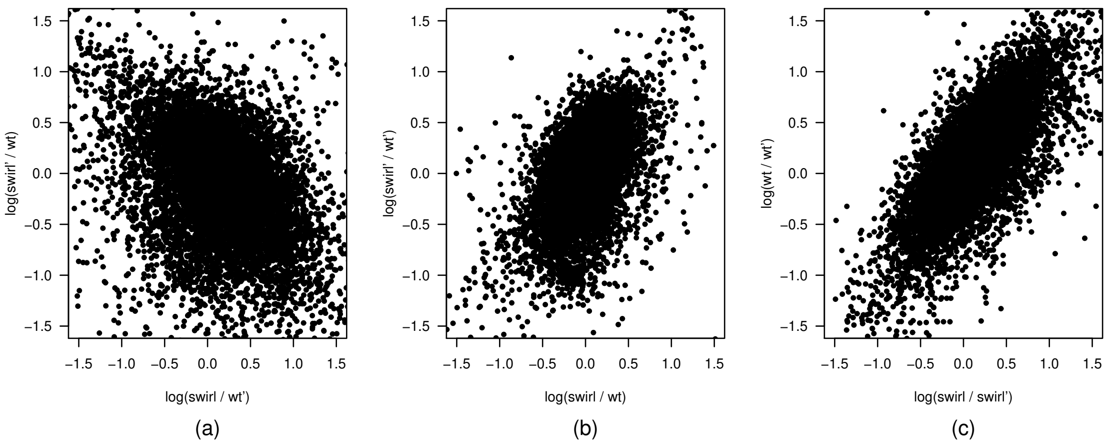

Figure 4: Three types of scatter plots. (a) Scatter plots of $\log_2(swirl/wt')$ versus $\log_2(swirl'/wt)$, where individual spot should have covariance of $2(c_{2i} - c_{1i})$. Similarly, (b) shows the scatter plots of $\log_2(swirl/wt)$ versus $\log_2(swirl'/wt')$, with covariance of $2(c_{2i} - c_{3i})$ for each individual spot, and (c) scatter plots of $\log_2(swirl/swirl')$ versus $\log_2(wt/wt')$, where in theory, each spot has covariance of $2(c_{1i} - c_{3i})$. Making use of identities described in Section 5, we conclude that $\bar{c}_1$ is noticeably larger than $\bar{c}_2$, which in turn is slightly larger than $\bar{c}_3$.

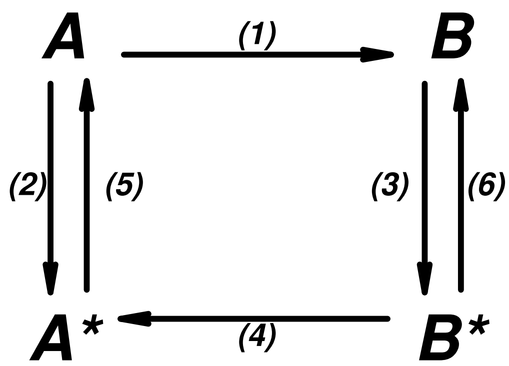

Figure 5: A hypothetical 2 by 2 factorial microarray experiment. The vertices A and B represent two samples A and B of cells, while A^* and B^* represent treated forms of samples A and B, e.g. with some drug. The number next to the arrows in the diagram denotes the slide number corresponding to the experiment.