

Experimental Design and Other Issues in Microarray Studies

**National Research Council
Committee on Emerging Issues and Data on Environmental Contaminants
September 16, 2003**

**Kathleen Kerr, Ph.D.
Department of Biostatistics
University of Washington**

Topics

- Experimental design
 - Design layout
 - Biological replicates
 - A “forgotten” topic
- A design misconception (sort of)
- The purpose of databases

Experimental Design for Microarrays

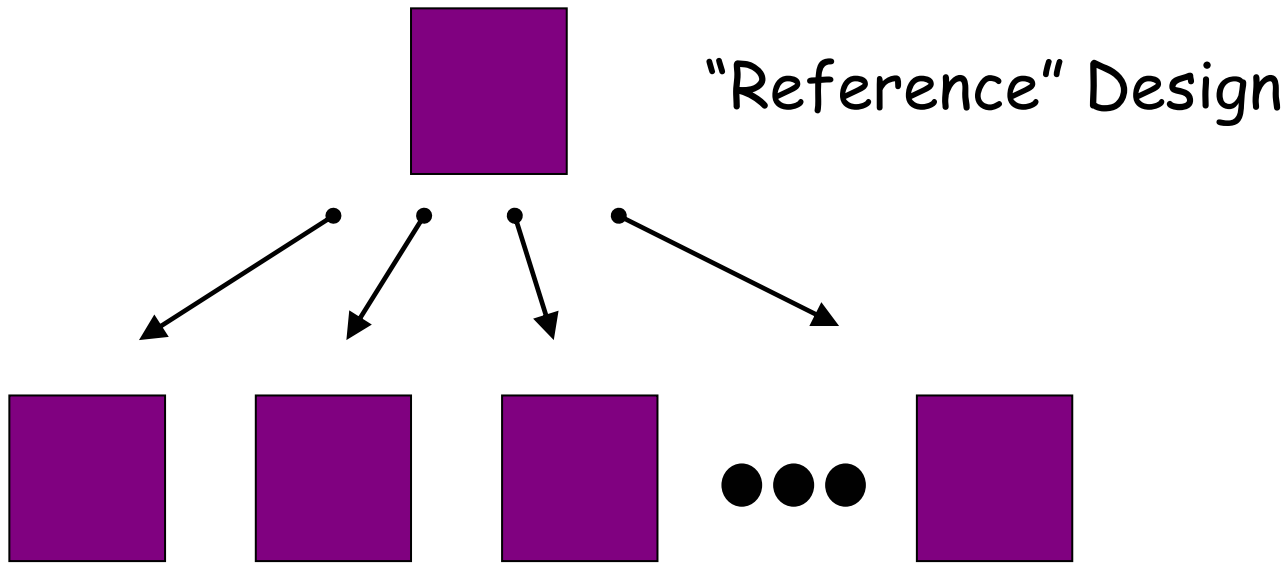
A. A handful of papers. . .

Kerr & Churchill, *Biostatistics*, 2001

Yang & Speed, *Nature Reviews Genetics*, 2002

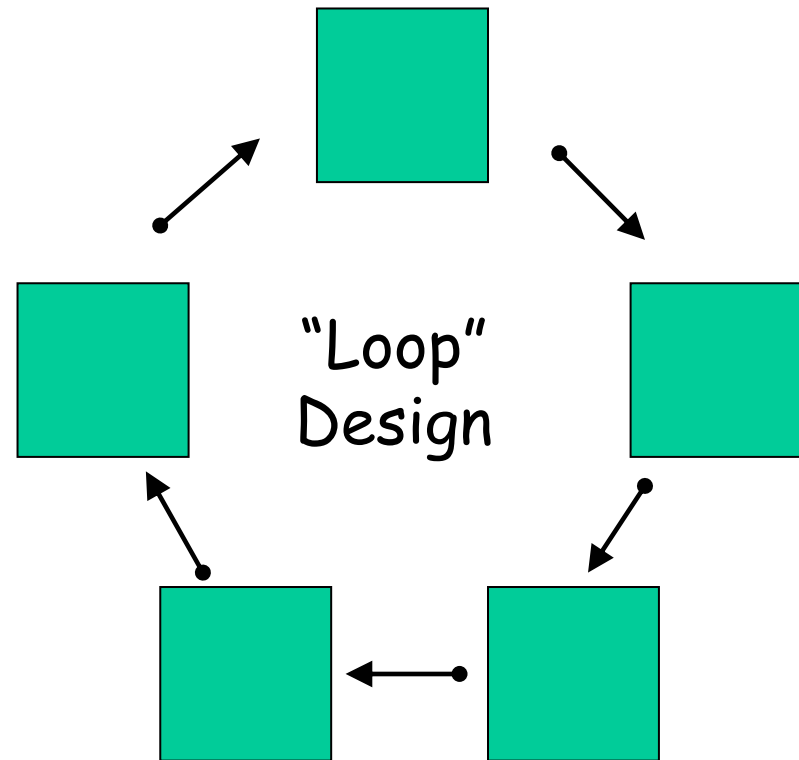
Glonek & Solomon, *Tech Report*, 2003

. . . examine the question of the experimental layout (for two-color assays) using models that reflect the technical error in the microarray assay.

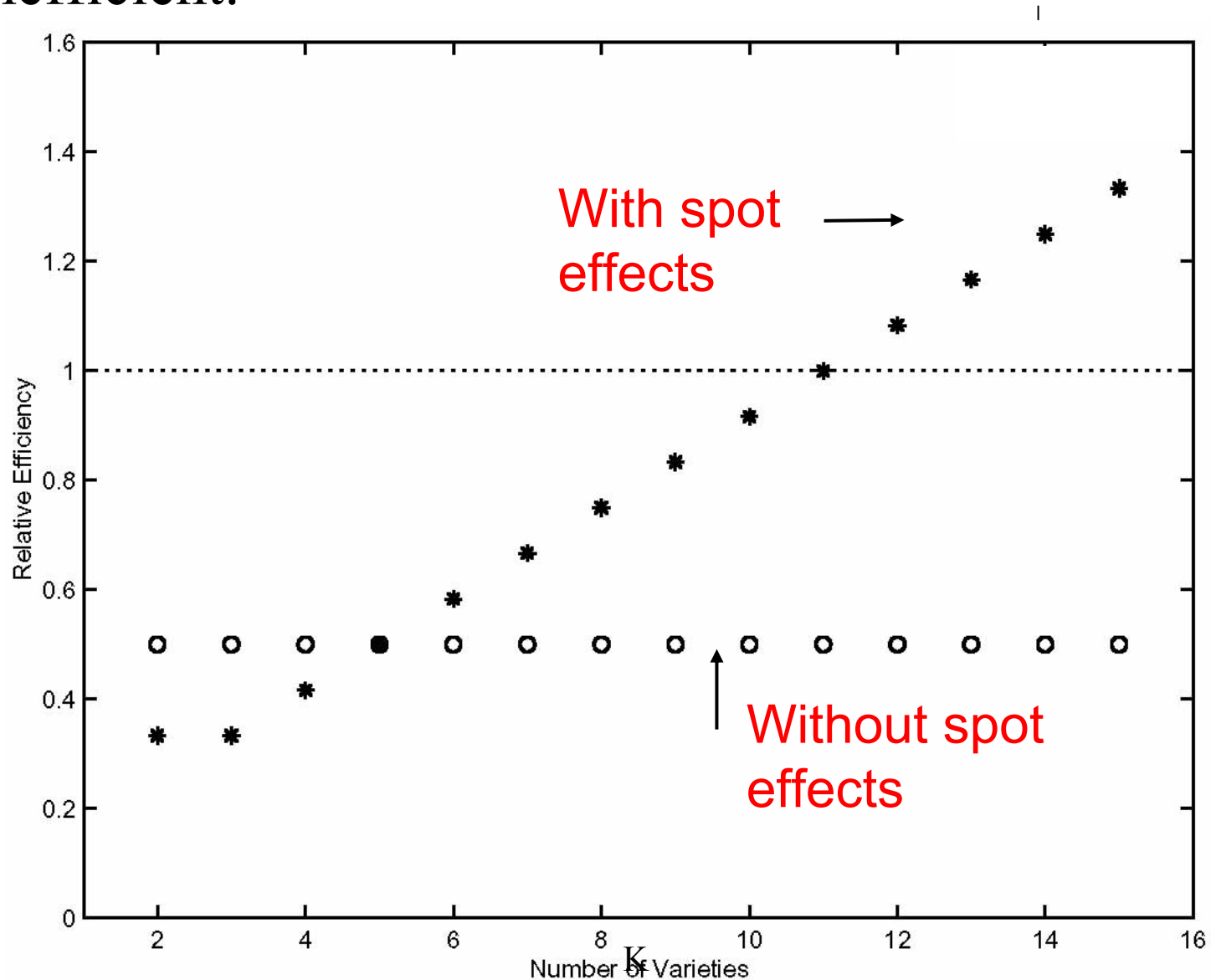


For example, to illustrate the effect of design layout on experimental results, Kerr & Churchill examined the relative efficiencies of these two design strategies.

The next slide shows the average variance of all pairwise comparisons of (non-reference) RNAs.



Relative Efficiency – Small loops are more efficient than reference designs under this criterion, large loops are inefficient.

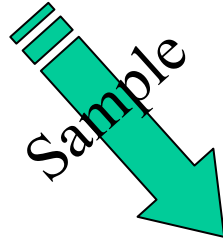
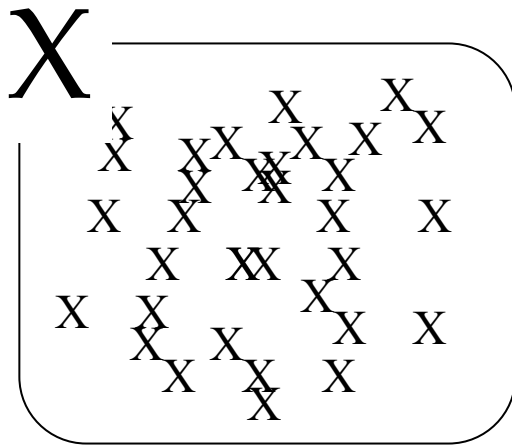


Experimental Design for Microarrays

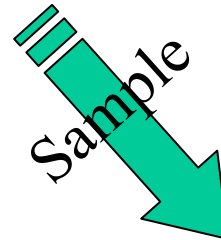
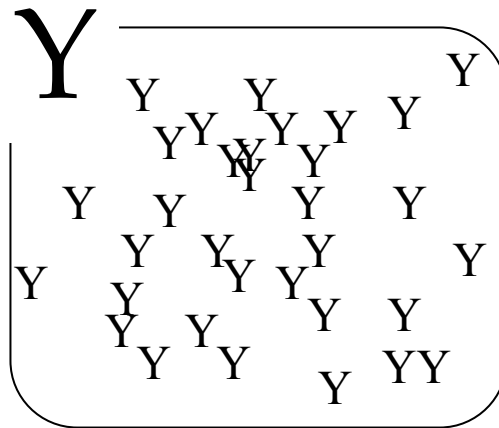
B. Many more papers have stressed the importance of including biological replicates in microarray experiments. Scientists have come to understand that the large amount of data produced by microarrays cannot replace biological replicates.

Technical Replication vs. Biological Replication: Experimental Design Choices

Statistics 101: Two Populations to Compare (unobserved)



Sample



Sample

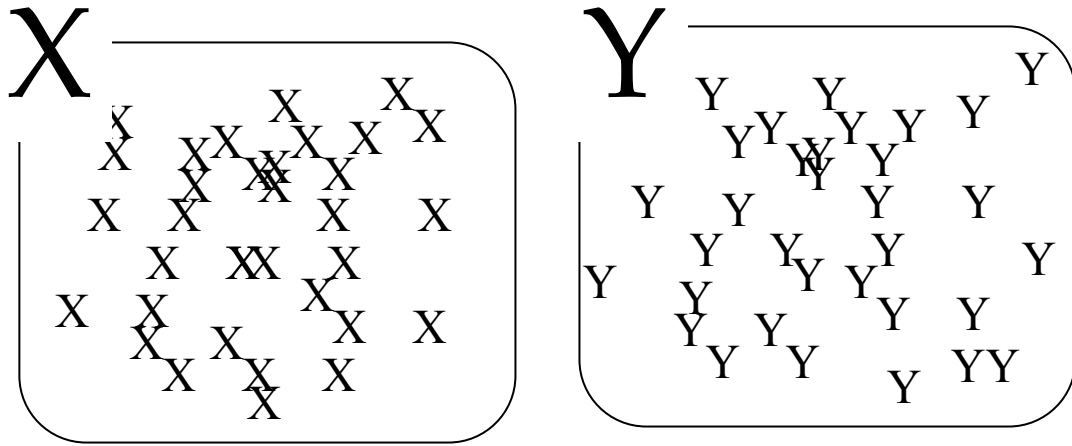
observed:

$X_1, X_2, X_3, \dots, X_m.$

$Y_1, Y_2, Y_3, \dots, Y_n.$

Technical Replication vs. Biological Replication: Experimental Design Choices

Two Populations to Compare (unobserved)



Inference

statistical analysis:

$X_1, X_2, X_3, \dots, X_m.$

$Y_1, Y_2, Y_3, \dots, Y_n.$

The importance of biological replicates

- A rigorous statistical inference cannot be made with a sample size of 1 ! Conversely, the more replicates there are, the better the inference.

Experimental Design for Microarrays

These two messages

A. The effect of design layout

B. The importance of biological replicates

are not yet connected in the literature.

Technical Replication vs. Biological Replication: Experimental Design Choices

Suppose now that observations are made with sizable error: $X_1 + \varepsilon_1, X_2 + \varepsilon_2, \dots, X_m + \varepsilon_m$ and $Y_1 + \varepsilon'_1, Y_2 + \varepsilon'_2, \dots, Y_n + \varepsilon'_n$.

Normally, we use the sampling variability of the X 's and Y 's to make inference about the true distributions. But now we have another source of variability, the measurement error.

Let τ^2 be the variance of the X's and Y's and let σ^2 be the variance of the measurement error.

Suppose our primary cost is measurement (not sampling); we can afford a fixed number N of measurements.

Is it better to take repeated measurements on sampled individuals (because of measurement error) or sample more individuals?

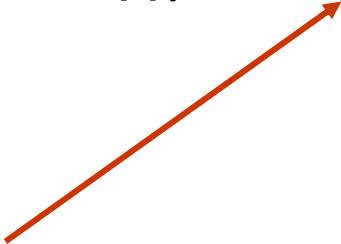
Technical Replication vs. Biological Replication: Experimental Design Choices

Say we can afford N measurements.

If we sample n individuals from each population, and measure each r times, then $N=n \cdot r$.

For given N , τ^2 , and σ^2 , how should we choose n and r ?

For estimating the difference in means,

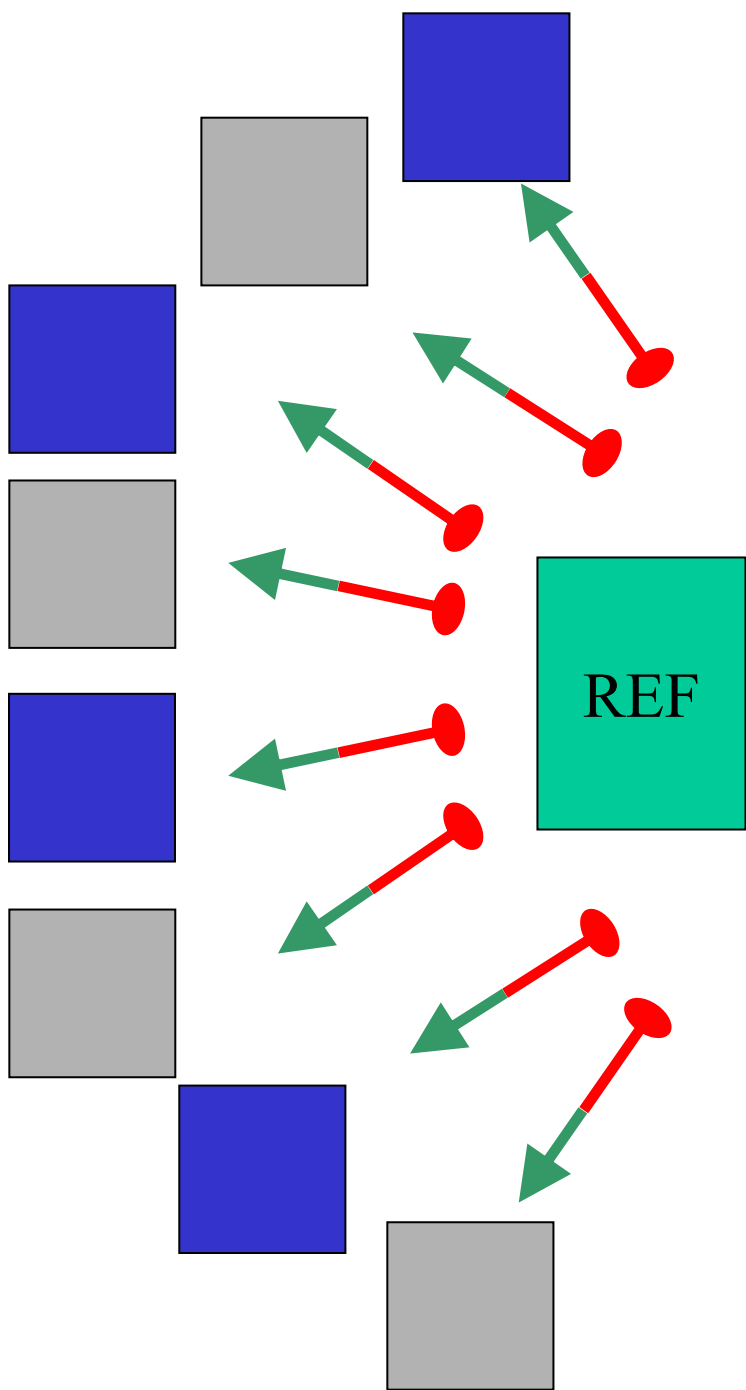
$$\text{var}(\bar{X} - \bar{Y}) = \frac{2}{n} \tau^2 + \frac{2}{rn} \sigma^2 = \frac{2}{n} \tau^2 + \frac{4}{N} \sigma^2$$


Clearly, we should choose n as large as possible, i.e., $r = 1$ repeated measurement.

In other words, biological replicates always beat technical replicates.

Technical and Biological Replicates in Microarrays

- For one-color microarray platforms (e.g. Affymetrix), this answers the question about biological vs. technical replicates.
- For two-color microarrays, the situation is somewhat more complicated because of the paired natures of the observations.
- Consider two groups of size n in a two-color microarray experiment.

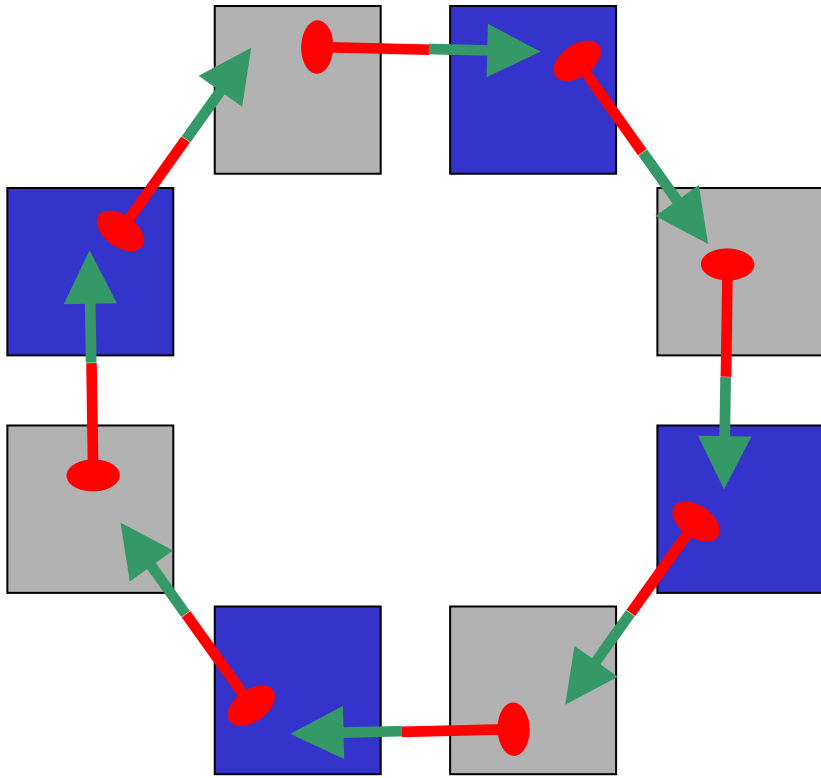


n individuals in each group

$2n$ arrays

Compare the difference in means
between the two groups:

$$\text{variance} = (4\sigma^2 + 2\tau^2)/n$$



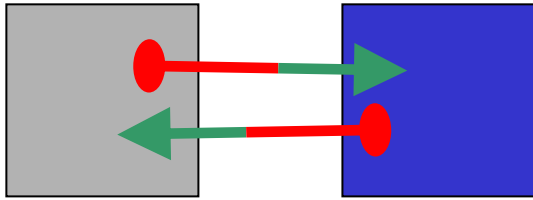
“Alternating Loop” Design

n individuals in each group
 $2n$ arrays

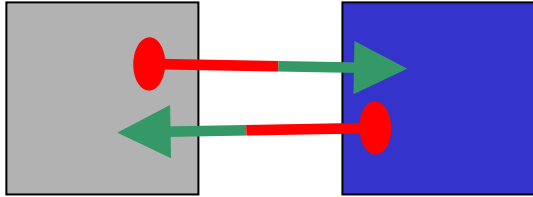
Compare the difference in means
between the two groups:

$$\text{variance} = (\sigma^2 + 2\tau^2)/n$$

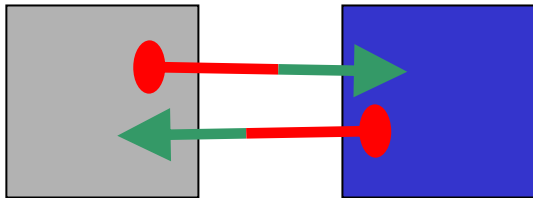
Always smaller than for the
reference design.



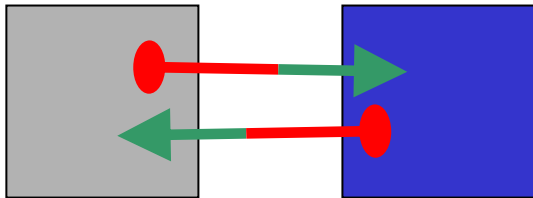
Multiple dye-swap



n individuals in each group
2n arrays



Compare the difference in means
between the two groups:



$$\text{variance} = (\sigma^2 + 2\tau^2)/n$$

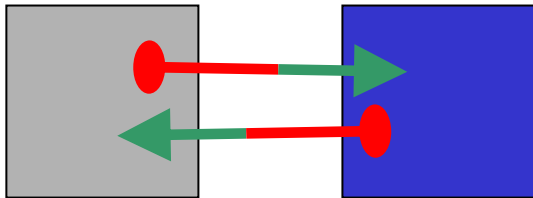
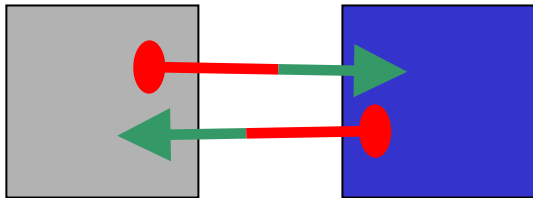
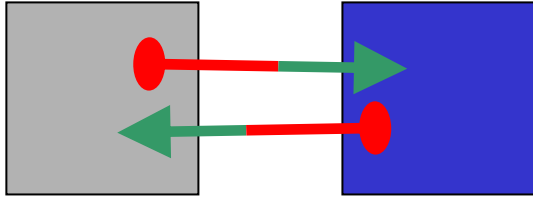
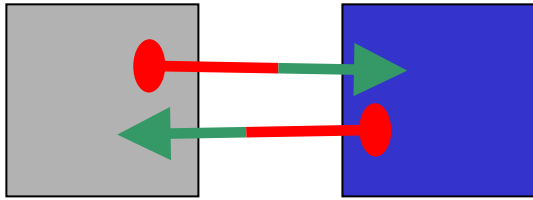
Same as the alternating loop.

Note that the reference design is always less efficient (even as $n \rightarrow \infty$): $(4\sigma^2 + 2\tau^2)/n > (\sigma^2 + 2\tau^2)/n$

However, if $\tau^2 \gg \sigma^2$, then this disadvantage is diminished.

Therefore, for studies with high biological variability, the reference design can be a good choice.

[Reference design has many advantages: robust, easy to plan and execute, easier to analyze.]



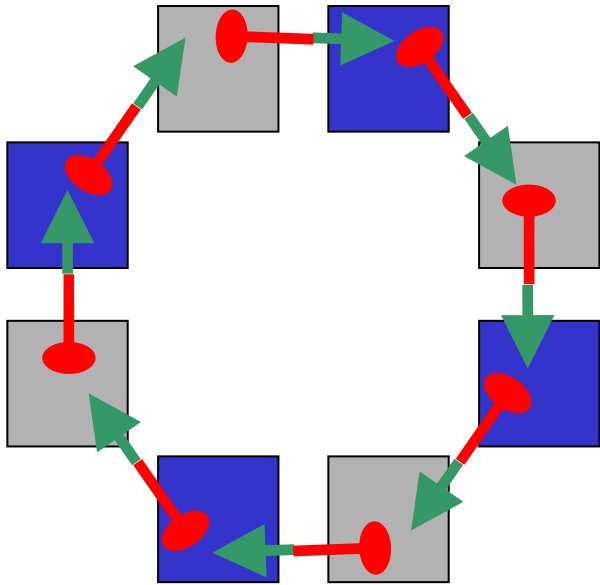
On the other hand, this design is as efficient at the alternating loop design.

It is also simple and robust, like the reference design.

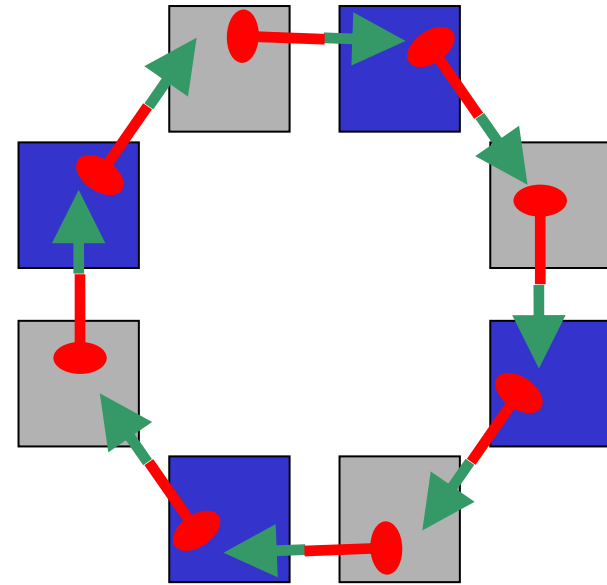
Is it important to include a reference RNA?

Overheard: “Experiments that do not employ a common reference sample cannot be compared.”

This statement is true in one sense, but false in another, important sense.

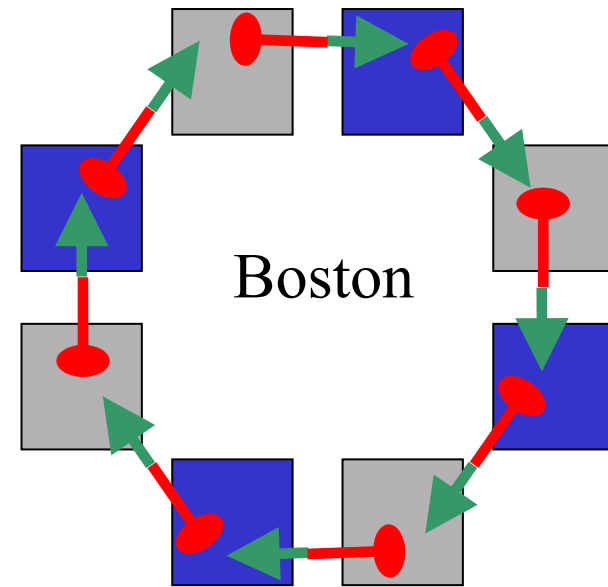
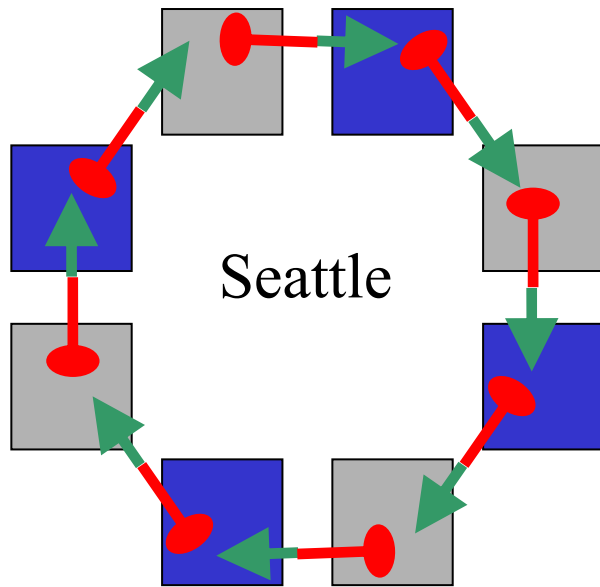


UW Researcher compared
tox exposed B6 mice to
unexposed in a loop
design.



Researcher in Boston
does the same
(maybe even used a
different layout).

For any gene, each experimenter can estimate the change in gene expression between exposed and unexposed mice. And they can compare those estimates – no reference RNA needed!



On the other hand, to compare the expression between RNAs in the different studies, they need to be compared to a common RNA (using conventional analytical methods).

In other words, I cannot compare the expression in any of the Seattle RNAs to any of the Boston RNAs because there is no “connection” between them.

Note, however, that this was not the goal of the experiment.

Experimental Design

- There are generally considered to be three principles of experimental design
 - blocking
 - replication
 - randomization
- For microarrays I have discussed aspects of two issues for experimental design:
 - Design layout – blocking
 - Biological replicates – replication

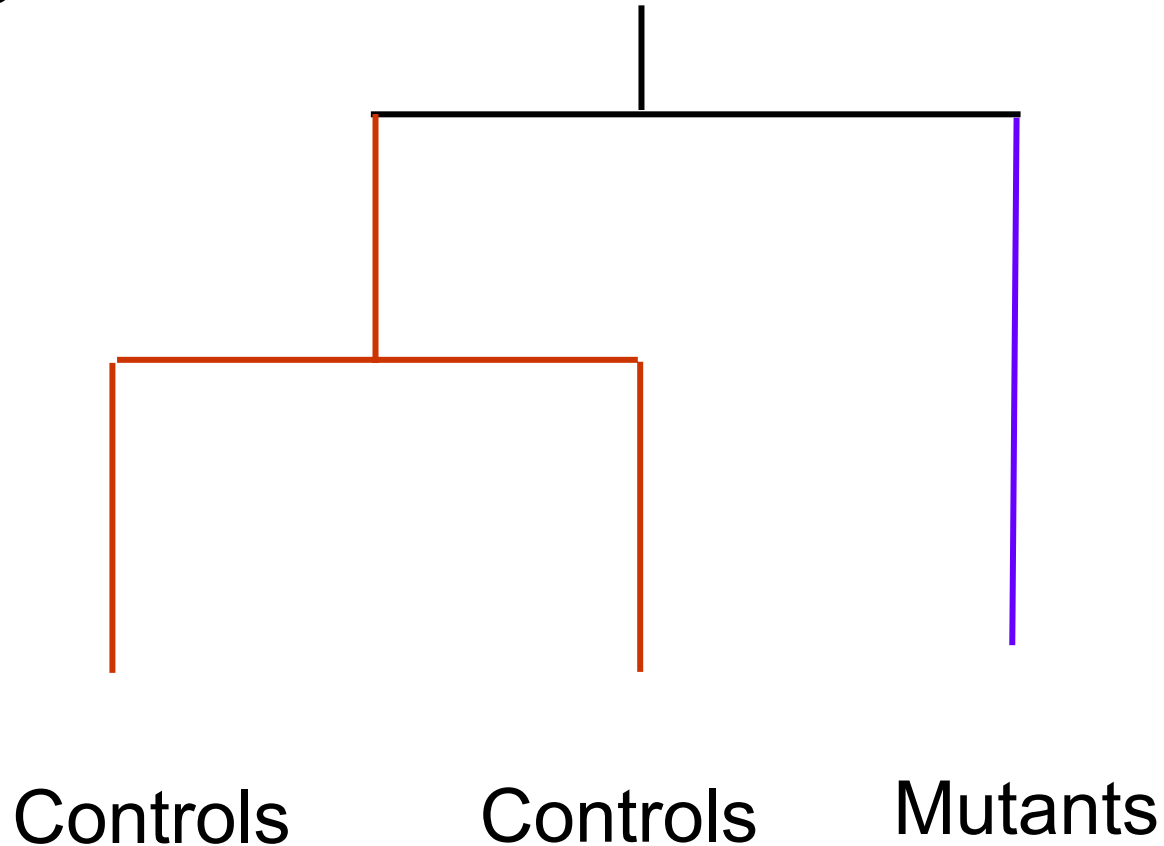
An cautionary tale

Experiment:

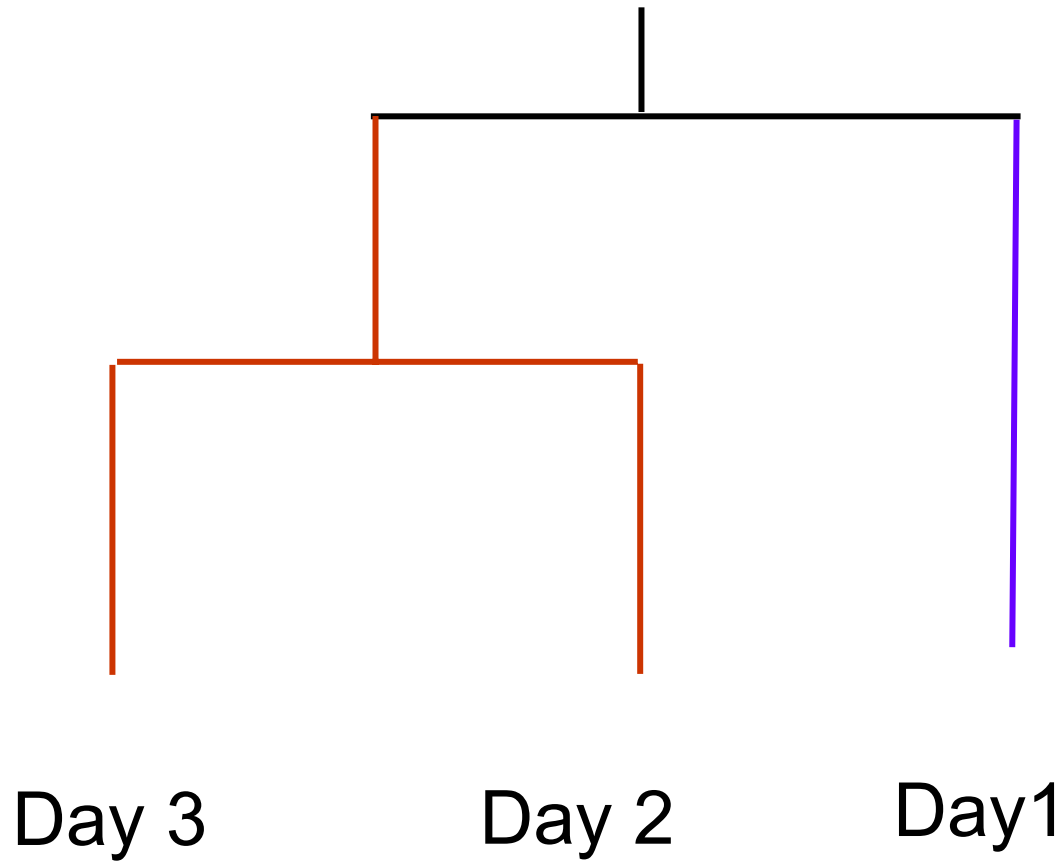
30 RNAs

- 10 carriers of a mutation
- 20 controls
- expression analysis with a single channel platform

Cluster analysis suggested three groups of subjects



Conclusion? Cluster analysis distinguishes mutants and controls and has discovered subclasses within the controls?



Conclusion: Large day-to-day effects are confounded with mutant/control differences; cluster analysis has merely discovered an artifact.

experimental plan:

Day 1: 10 mutants

Day 2: 10 controls

Day 3: 10 remaining controls

Randomization was the important missing component in this study.

Taking a broader view, what about larger studies?

...gene expression data are . . . meaningful only in the context of a detailed description of the conditions under which they were generated, including the particular state of the living system under study and the perturbations to which it has been subjected. In contrast to an organism's genome, there are as many transcriptomes as there are cell types multiplied by environmental conditions.

Commentary by Brazma et al, Nature Genetics, December 2001

Reality: Interpreting gene expression data requires an understanding of the context in which the data were collected.

Meaningful exchange of microarray data is currently difficult because it is rare that published data provide sufficient information depth or are even in the same format from one publication to another. Only when data can be easily exchanged will the entire biological community be able to derive the full benefit from such microarray studies.

Spellman et al, Genome Biology, 2002 v.3

Problem: Data analysis, data storage is ad hoc. Even with the same dataset, results might not be able to be reproduced, much less after repeating the experiment.

Question: What is this illusive “full benefit”?

"Harried editors can rejoice that, at last, the community is taming the unruly beast that is microarray information," Nature editors wrote of the MIAME standard in a recent issue of the journal (419:232).

At the very least, some organization is called for.

At first glance, the extent of the information requested in the MIAME specification may seem daunting ...

Commentary by Brazma et al, Nature Genetics, December 2001

We should not forget that these requirements do carry a price.

“The biggest impediment to large-scale analysis of microarray data is the lack of a common data format”,

-David Haussler, Director
Center for Biomolecular Science and Engineering, UCSC,
as quoted in “The Scientist” October 10, 2002



An invited speaker at RECOMB 2001
(I believe it was Roger Brent) made the following
observation and comment (paraphrase):

Some scientists talk about microarray data today the way some scientists used to talk about computers in the 1960's. They seemed to think that, once computers became powerful enough, they would – by some unspecified mechanism – suddenly become intelligent.

That is, we talk about the importance of getting microarray data into databases, with all the details recorded. The idea seems to be that once enough bits of information are stored in contiguous memory, we will – by some unspecified mechanism – know something about biology we did not know before.

In my opinion

- The purpose of doing a microarray study is not to collect data for a database
- A microarray study should be designed with a primary objective in mind
 - Comparing gene expression in different groups
 - Seek biomarkers or a classification tool for diagnosis, prognosis, etc.
 - Etc.
- The likelihood is low that a large database comprised of data from multiple, un-coordinated experiments will be successfully “mined” and yield new biological understanding.

- This says nothing against collaboration! To the contrary, scientists need to collaborate and design experiments that compliment each other.
- Storing all the details of different experiments cannot substitute for designing experiments to answer questions of interest.
- Large microarray experiments need to be actively planned collaborations
 - study objective should be identified
 - analytical plan from the beginning