Experimental Design

Credit for some of today's materials:
Jean Yang, Terry Speed, and Christina Kendziorski

Experimental design

- Choice of platform
- Array design
  - Creation of probes
  - Location on the array
  - Controls
- Target samples

Outline

- General recommendations
- Types of replicates
- Layouts for two color platforms
- Pooling
- How many replicates

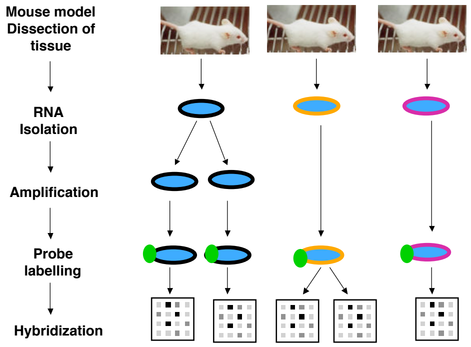
Experimental design

Proper experimental design is needed to ensure that questions of interest ***can*** be answered and that this can be done **accurately and precisely**, given experimental constraints, such as cost of reagents and availability of mRNA.

Avoidance of bias

- **Conditions of an experiment; mRNA extraction and processing, the reagents, the operators, the scanners and so on can leave a "global signature" in the resulting expression data.**
- **Balance**
- **Randomization**

Preparing mRNA samples:



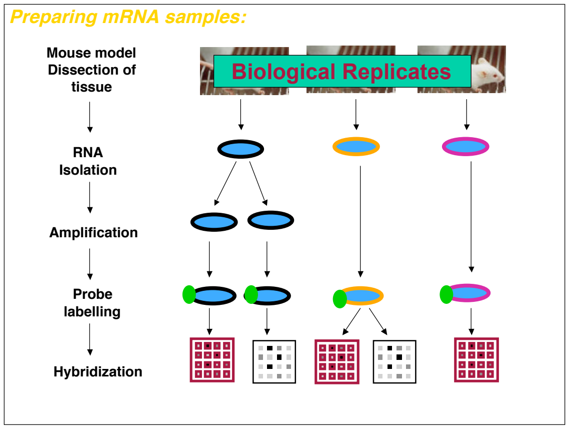
Mouse model Dissection of tissue

RNA Isolation

Amplification

Probe labelling

Hybridization



Preparing mRNA samples:

Mouse model Dissection of tissue

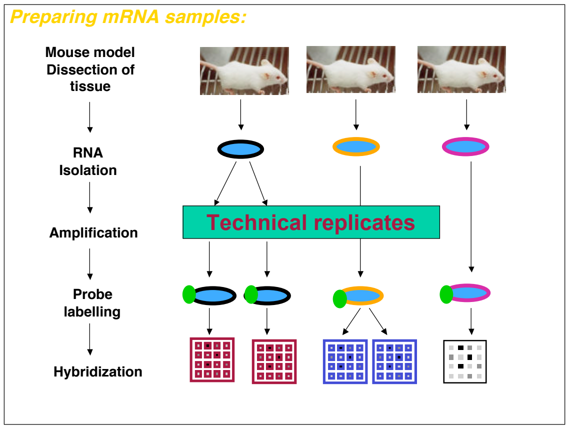
RNA Isolation

Amplification

Probe labelling

Hybridization

Biological Replicates



Preparing mRNA samples:

Mouse model Dissection of tissue

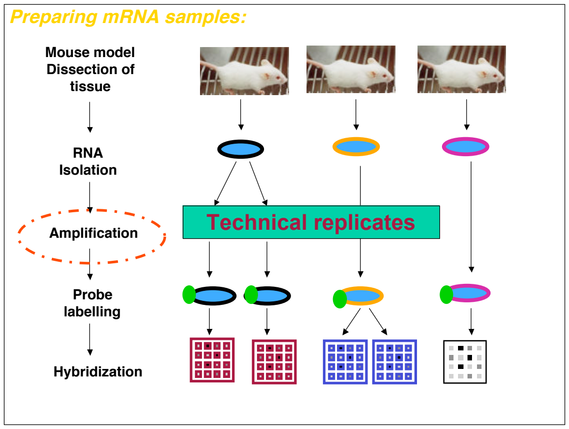
RNA Isolation

Amplification

Probe labelling

Hybridization

Technical replicates



Preparing mRNA samples:

Mouse model Dissection of tissue

RNA Isolation

Amplification

Probe labelling

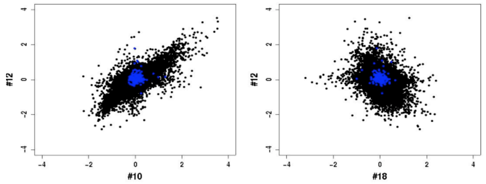
Hybridization

Technical replicates

Technical replication - amplification

Olfactory bulb experiment:

- 3 sets of two different samples performed on different days
- #10 and #12 were from the same RNA isolation and amplification
- #12 and #18 were from different dissections and amplifications
- All 3 data sets were labeled separately before hybridization



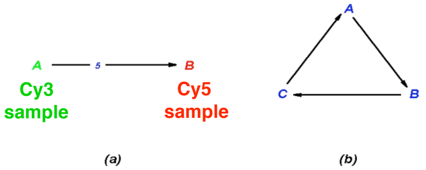
Data provided by
Dave Lin (Cornell)

Layouts for two color platforms

Graphical representation

For two color platforms it is assumed that the size of the spot/probe effect is too big to trust the absolute intensites. Thus we always use relative measurements

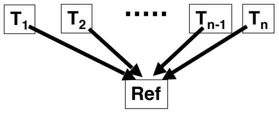
Vertices: mRNA samples; ***Edges:*** hybridization; ***Direction:*** dye assignment.



Graphical representation

- The structure of the graph determines which effects can be estimated and the precision of the estimates.
  - Two mRNA samples can be compared only if there is a path joining the corresponding two vertices.
  - The precision of the estimated contrast then depends on the number of paths joining the two vertices and is inversely related to the length of the paths.
- Direct comparisons within slides yield more precise estimates than indirect ones between slides.
- Experiments studying more than one effect can get complicated if we optimize variance

Common reference design



- Experiment for which the common reference design is appropriate

  Meaningful biological control (C) Identify genes that responded differently / similarly across two or more treatments relative to control.

  Large scale comparison. To discover tumor subtypes when you have many different tumor samples.

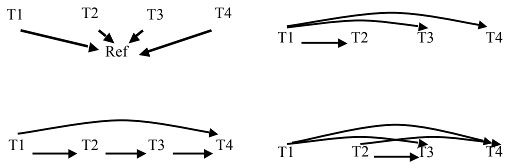
- Advantages:

  Ease of interpretation.

  Extensibility - extend current study or to compare the results from current study to other array projects.

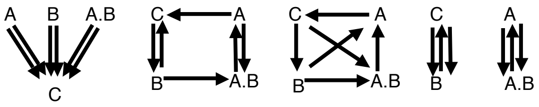
Experiment for which a number of designs are suitable for use

Time Series



Experiment for which a number of designs are suitable for use

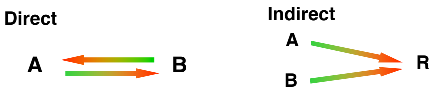
4 samples



Comparing 2 classes of estimates direct *vs* indirect estimates

The simplest design question: Direct versus indirect comparisons

Two samples (A *vs* B)
e.g. KO *vs.* WT or mutant *vs.* WT



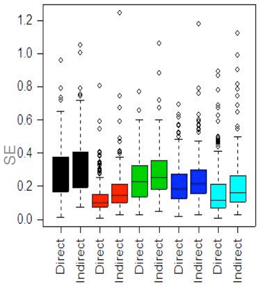
average (log (A/B)) log (A / R) – log (B / R)

$\sigma^2 /2$ $2\sigma^2$

These calculations assume independence of replicates: the reality is not so simple.

Experimental results

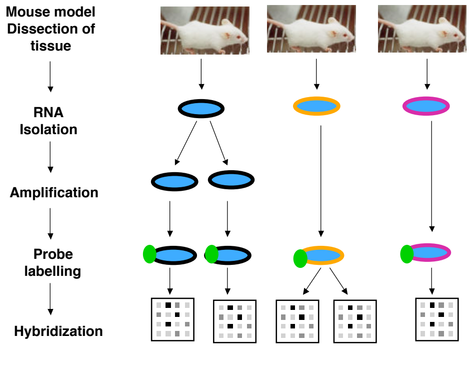
- 5 sets of experiments with similar structure.
- We compare two methods, A and B, and compare SE obtained from replicates.
- Theoretical ratio of (A / B) is 1.6 (assuming independence)
- Experimental observation is 1.1 to 1.4.



Caveat

- The advantage of direct over indirect comparisons was first pointed out by Churchill & Kerr, and in general, we agree with the conclusion. However, you can see in the last MA-plot that the difference is not a factor of 2, as theory predicts.
- Why? Possibly because mRNA from the same extractions - and pools of controls or reference material are the norm - give correlated expression levels. In other words, the assumption of independence between log(T/Ref) and log(C/Ref) is not valid.

Preparing mRNA samples:



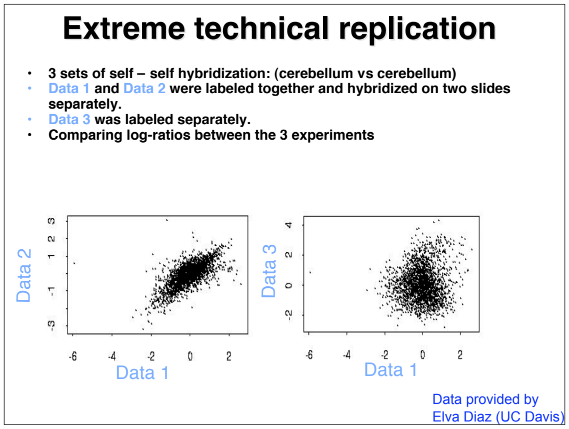
Mouse model Dissection of tissue

RNA Isolation

Amplification

Probe labelling

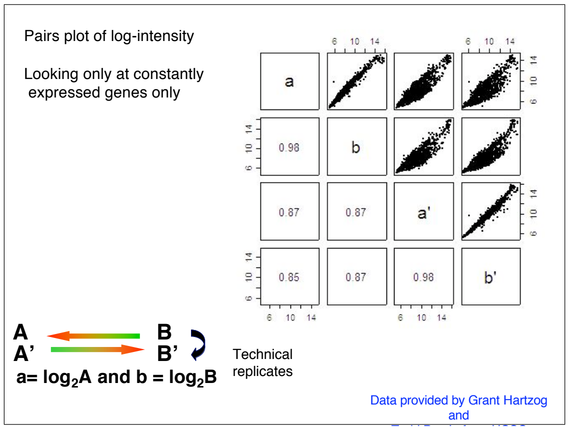
Hybridization



Extreme technical replication

- 3 sets of self – self hybridization: (cerebellum vs cerebellum)
- Data 1 and Data 2 were labeled together and hybridized on two slides separately.
- Data 3 was labeled separately.
- Comparing log-ratios between the 3 experiments

Data provided by
Elva Diaz (UC Davis)



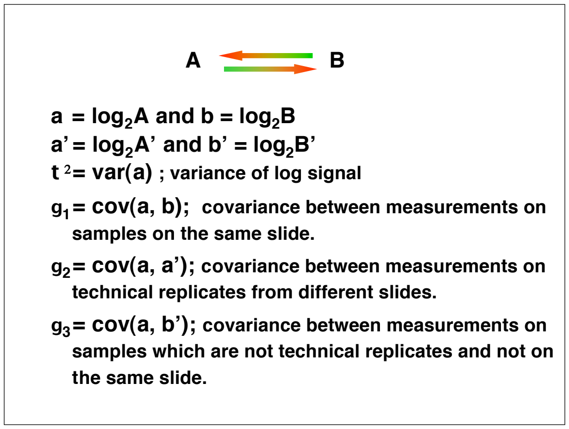
Pairs plot of log-intensity

Looking only at constantly expressed genes only

Technical replicates

a= $\log_2$A and b = $\log_2$B

Data provided by Grant Hartzog
and



a = $\log_2$A and b = $\log_2$B

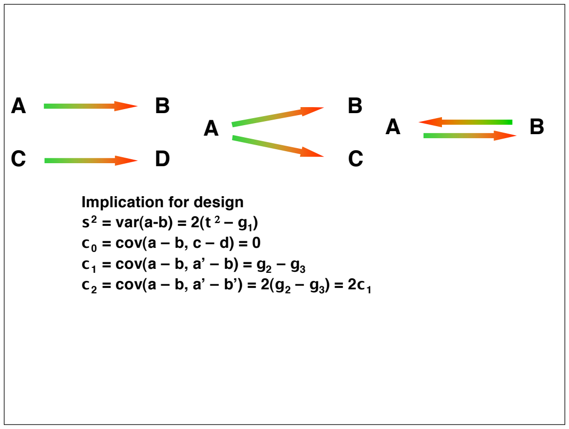
a' = $\log_2$A' and b' = $\log_2$B'

t^2= var(a) ; variance of log signal

g_1= cov(a, b); covariance between measurements on samples on the same slide.

g_2= cov(a, a'); covariance between measurements on technical replicates from different slides.

g_3= cov(a, b'); covariance between measurements on samples which are not technical replicates and not on the same slide.



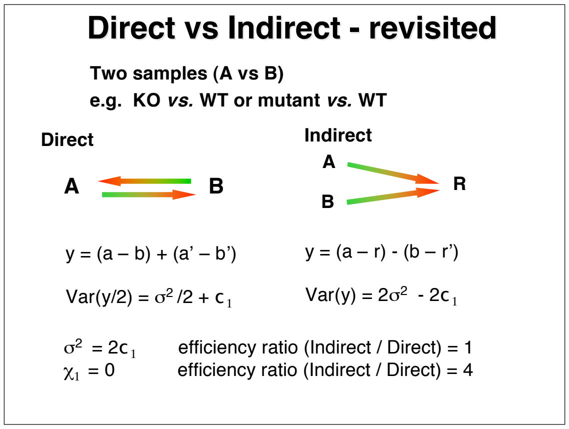
Implication for design

$s^2 = \mathrm{var}(a-b) = 2(t^2 - g_1)$

$c_0 = \mathrm{cov}(a - b, c - d) = 0$

$c_1 = \mathrm{cov}(a - b, a' - b) = g_2 - g_3$

$c_2 = \mathrm{cov}(a - b, a' - b') = 2(g_2 - g_3) = 2c_1$



Direct vs Indirect - revisited

Two samples (A vs B)
e.g. KO *vs.* WT or mutant *vs.* WT

$y = (a - b) + (a' - b')$

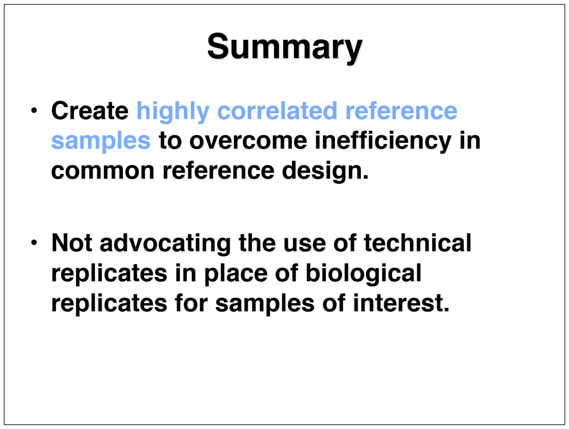
$\mathrm{Var}(y/2) = \sigma^2/2 + c_1$

$y = (a - r) - (b - r')$

$\mathrm{Var}(y) = 2\sigma^2 - 2c_1$

$\sigma^2 = 2c_1$ efficiency ratio (Indirect / Direct) = 1

$\chi_1 = 0$ efficiency ratio (Indirect / Direct) = 4



Summary

- Create highly correlated reference samples to overcome inefficiency in common reference design.
- Not advocating the use of technical replicates in place of biological replicates for samples of interest.

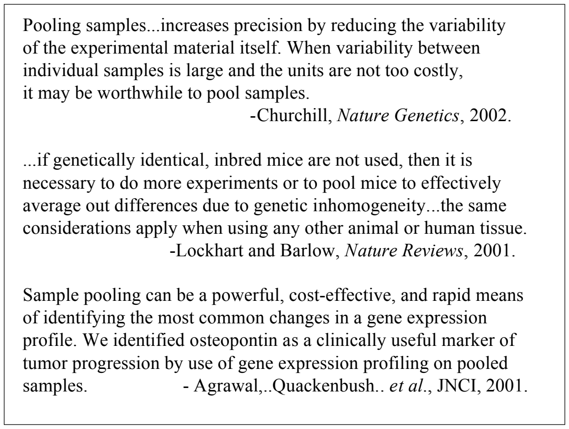
Gene Specific Variance: Pooling and Power Calculations

Most common applications

- Class prediction: In general, do not pool
- Class comparison?
  - Pool everything is generally a bad idea
  - But, other strategies exists

Common question in experimental design

- Should I pool mRNA samples across subjects in an effort to reduce the effect of biological variability?



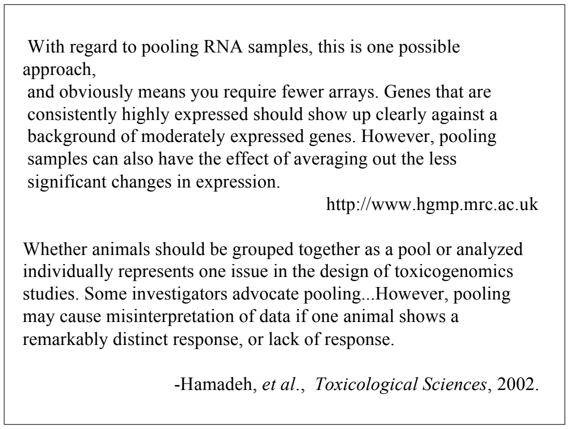
Pooling samples...increases precision by reducing the variability of the experimental material itself. When variability between individual samples is large and the units are not too costly, it may be worthwhile to pool samples.

-Churchill, *Nature Genetics*, 2002.

...if genetically identical, inbred mice are not used, then it is necessary to do more experiments or to pool mice to effectively average out differences due to genetic inhomogeneity...the same considerations apply when using any other animal or human tissue.

-Lockhart and Barlow, *Nature Reviews*, 2001.

Sample pooling can be a powerful, cost-effective, and rapid means of identifying the most common changes in a gene expression profile. We identified osteopontin as a clinically useful marker of tumor progression by use of gene expression profiling on pooled samples. - Agrawal,..Quackenbush.. *et al*., JNCI, 2001.

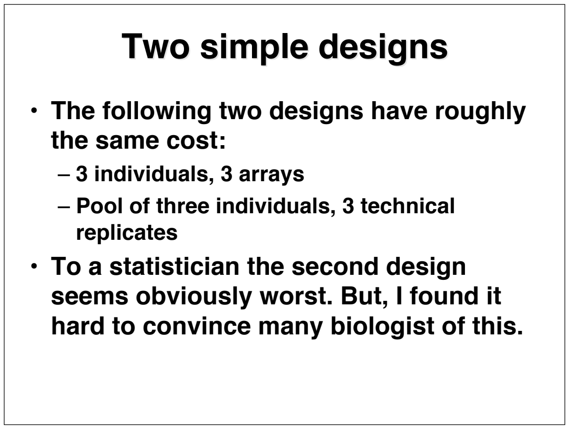


With regard to pooling RNA samples, this is one possible approach,
and obviously means you require fewer arrays. Genes that are consistently highly expressed should show up clearly against a background of moderately expressed genes. However, pooling samples can also have the effect of averaging out the less significant changes in expression.

http://www.hgmp.mrc.ac.uk

Whether animals should be grouped together as a pool or analyzed individually represents one issue in the design of toxicogenomics studies. Some investigators advocate pooling...However, pooling may cause misinterpretation of data if one animal shows a remarkably distinct response, or lack of response.

-Hamadeh, *et al*., *Toxicological Sciences*, 2002.



Two simple designs

- The following two designs have roughly the same cost:
  - 3 individuals, 3 arrays
  - Pool of three individuals, 3 technical replicates
- To a statistician the second design seems obviously worst. But, I found it hard to convince many biologist of this.

Cons of Pooling Everything

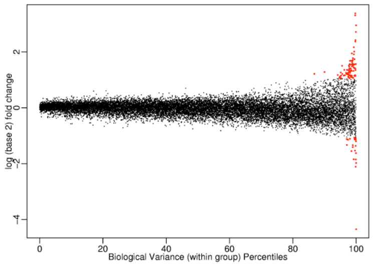
- You can not measure within class variation
- Therefore, no population inference possible
- Mathematical averaging is an alternative way of reducing variance. The standard error of the mean of three numbers is 58% of the variance of each individual measurement
- Pooling may have *non-linear* effects
- You can not take the log before you *average*
- You can not detect outliers

*If the measurements are independent and identically distributed

Cons specific to microarrays

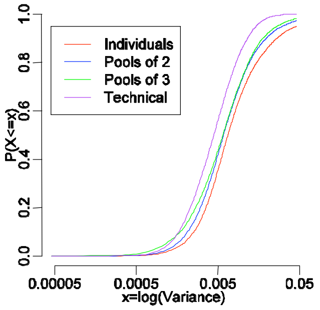
- For now, forget about inference. Let us concentrate on ranking correctly
- Different genes may have different within class *biological* variances
- Not measuring this variance will result in genes with larger biological variance having a better chance of being considered more important

Higher variance: larger fold change

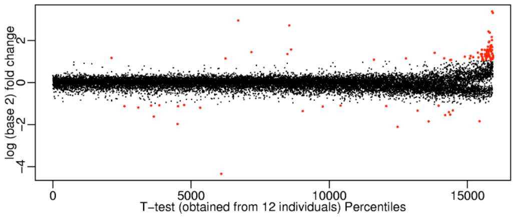


In a three versus three comparison we compute fold change for each gene
From 12 individuals we estimate gene specific variance
If we pool we never see this variance

CDFs of Sample Variances

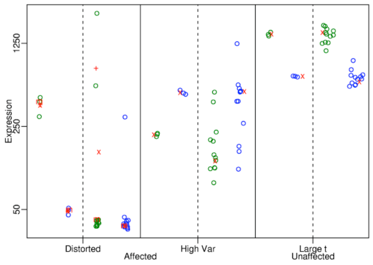


Fold Change False Positives



T-test from 12 versus 12 gives different answer than fold change from 3 versus 3

Problem with pooling everything

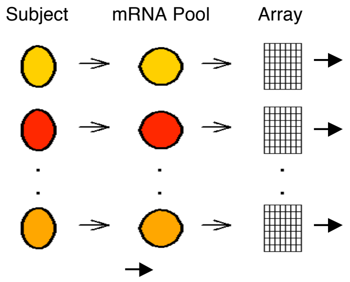


1) You can not measure variability
2) You can not take log before "averaging"

Alternative pooling strategy

- Instead of pooling everything, how about pooling groups?
- For example, will I obtain the same results with 12 individuals on 12 chips as with 12 individuals on 4 chips ?

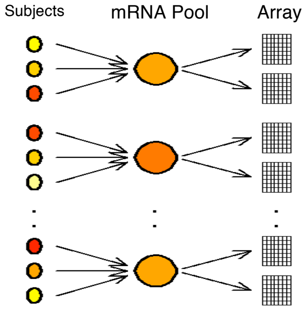
Design I


Subject

mRNA Pool

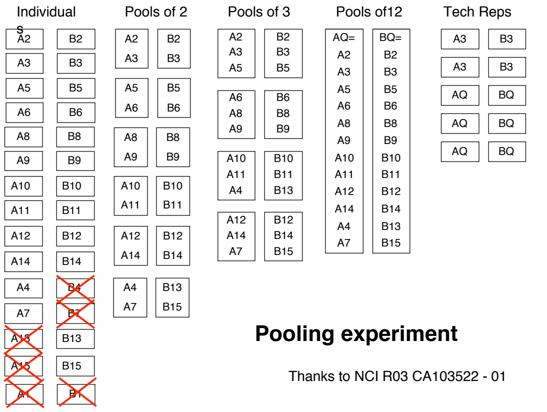
Array

Design II


Subjects

mRNA Pool

Array



Individuals

A2 B2
A3 B3
A5 B5
A6 B6
A8 B8
A9 B9
A10 B10
A11 B11
A12 B12
A14 B14
A4 ~~B4~~
A7 ~~B7~~
~~A13~~ B13
~~A15~~ B15
~~A1~~ ~~B1~~

Pools of 2

A2 A3 | B2 B3
A5 A6 | B5 B6
A8 A9 | B8 B9
A10 A11 | B10 B11
A12 A14 | B12 B14
A4 A7 | B13 B15

Pools of 3

A2 A3 A5 | B2 B3 B5
A6 A8 A9 | B6 B8 B9
A10 A11 A4 | B10 B11 B13
A12 A14 A7 | B12 B14 B15

Pools of12

AQ= A2 A3 A5 A6 A8 A9 A10 A11 A12 A14 A4 A7 | BQ= B2 B3 B5 B6 B8 B9 B10 B11 B12 B14 B13 B15

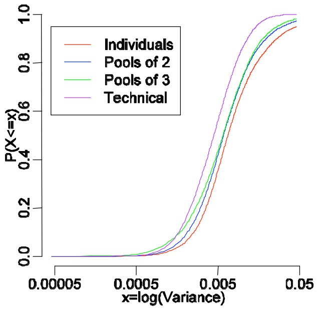
Tech Reps

A3 B3
A3 B3
AQ BQ
AQ BQ
AQ BQ

Pooling experiment

Thanks to NCI R03 CA103522 - 01

CDFs of Sample Variances



More on pooling

- In Kendziorski (2003) some technical details are worked out to determine the best pooling strategy
- These are based on assumptions that can only be checked empirically
- For example, are mathematical and biological averages the same?

Notation

q : nominal level of expression.

r_s : number of subjects that go into one pool.

r_a : number of arrays that probe one pool.

n_p : number of pools.

For a given gene, one experiment results in n_p x r_a observed expression levels, denoted by X_{ij} (i = 1,2,...,n_p), j = 1,2,...,r_a.

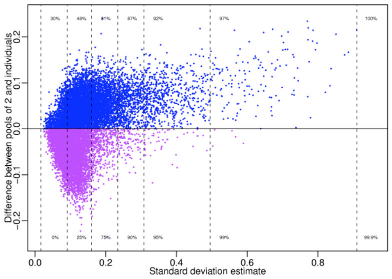
$\overline{X}$ estimates q

Some Issues

- Are the expectations in the previous slide really the same? I.e. is mathematical averaging the same as biological averaging?
- One problem is that the additive error and normality assumptions may only hold if you take the log. But if you take the log then the above assumption certainly doesn't hold because :

$E[\log(X+Y)] \neq E[\log(X)] + E[\log(Y)]$

Empirical evidence of this inequality problem



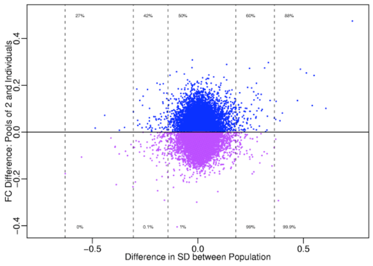
Some Issues

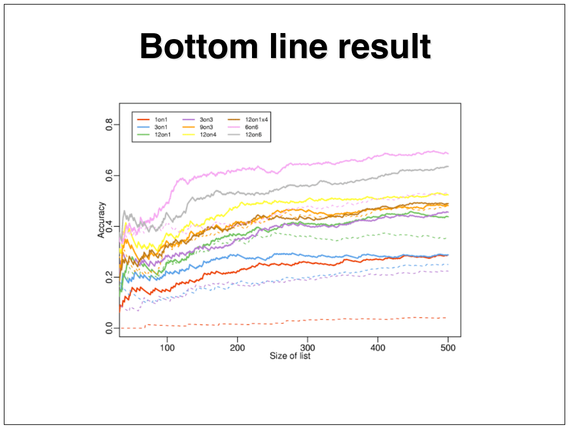
- Some published definition of equivalency are based on gene-specific power calculations. But:
- We are interested in false positives and false negative rates of lists. Various papers describe better approaches, but
- How do we put cost into the equation? Biological samples are usually much cheaper than arrays.

Bottom line

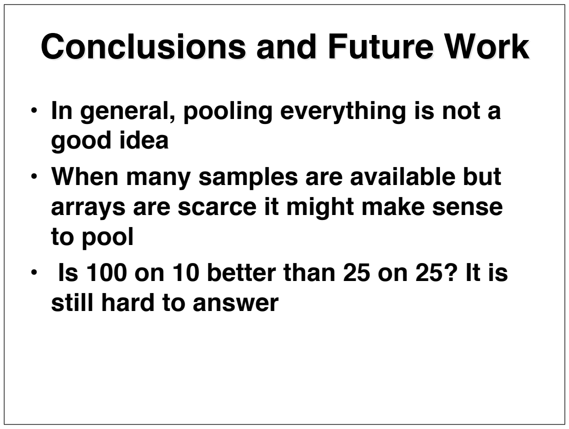
- To certain extent we do not care if the assumption hold perfectly
- More important is that we obtain similar lists of interesting genes
- In this regarding some pooling strategies work pretty well (but not pooling everything)

Not much of a worry when looking at differential expression



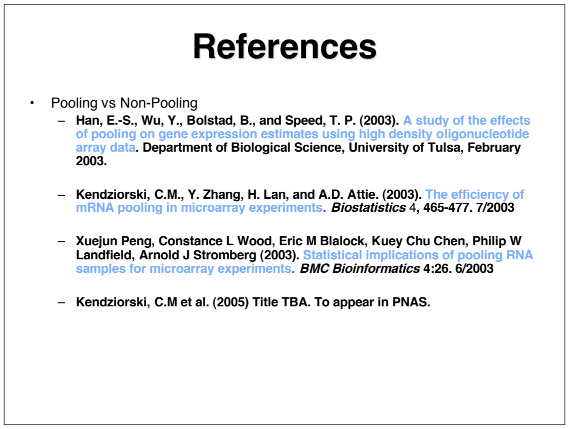


Bottom line result



Conclusions and Future Work

- In general, pooling everything is not a good idea
- When many samples are available but arrays are scarce it might make sense to pool
- Is 100 on 10 better than 25 on 25? It is still hard to answer



References

- Pooling vs Non-Pooling
  - Han, E.-S., Wu, Y., Bolstad, B., and Speed, T. P. (2003). A study of the effects of pooling on gene expression estimates using high density oligonucleotide array data. Department of Biological Science, University of Tulsa, February 2003.
  - Kendziorski, C.M., Y. Zhang, H. Lan, and A.D. Attie. (2003). The efficiency of mRNA pooling in microarray experiments. *Biostatistics* 4, 465-477. 7/2003
  - Xuejun Peng, Constance L Wood, Eric M Blalock, Kuey Chu Chen, Philip W Landfield, Arnold J Stromberg (2003). Statistical implications of pooling RNA samples for microarray experiments. *BMC Bioinformatics* 4:26. 6/2003
  - Kendziorski, C.M et al. (2005) Title TBA. To appear in PNAS.

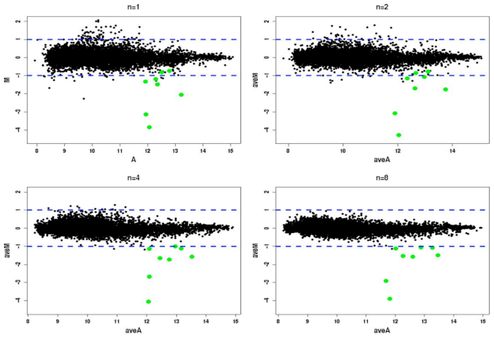
Power Calculations are Hard

- What do we mean by power?
- Are we really doing inference?
- Different tissues will have different variance distributions
- Some papers:
  - Mueller, Parmigiani et al. JASA (2004)
  - Rich Simon's group Biostatistics (2005)

Conclusions

- Spend your money on Biological replicates not technical replicates
- Perform direct comparisons when you can but don't underestimate the logistical advantages of reference designs
- Do not pool everything!
- Don't trust rules of thumb regarding number of replicates: different problems will need different sample sizes

Sample size



Data provided by Matt Callow

- Design 1: Each subject's mRNA is probed individually.

$$X_{i,1} = \theta + \varepsilon_i + \xi_{i,1}$$

e$_i$ represents subject-to-subject variability and x_i denotes array-to-array variability $\lfloor \varepsilon_i \sim N(0,\sigma_\varepsilon^2)$ and $\xi_{i,1} \sim N(0,\sigma_\xi^2)\rfloor$.

- Design II, mRNA from r_s subjects is pooled and probed by r_a arrays.

$$X_{i,j} = \theta + \varepsilon_i' + \xi_{i,j}$$

ε_i' represents pool-to-pool variability $\lfloor \varepsilon_i' \sim N(0,\sigma_\varepsilon^2/r_s)\_$.

For both designs, $E\lfloor \overline{X}_{..} \rfloor = \theta$;

$$\sigma^2_{\overline{X},(1)} = \frac{1}{n_{p1}}\left(\sigma_\varepsilon^2 + \sigma_\xi^2\right) \qquad \sigma^2_{\overline{X},(2)} = \frac{1}{n_{p2}}\left(\frac{\sigma_\varepsilon^2}{r_{s2}} + \frac{\sigma_\xi^2}{r_{a2}}\right)$$

Equivalent Designs according to Kendziorski et al. 2003

When the variance components are not known,

$$R = \frac{E\left(l_1^2\right)}{E\left(l_2^2\right)}$$

For fixed n_{s1} and n_{a1} (total number of subjects and arrays), R=1 when

$$n_{s2} = n_{s1}\left(\frac{\lambda}{K(\lambda+1) - {}^{n_{a1}}\!/_{n_{a2}}}\right)$$

where $\lambda = {}^{\sigma_\varepsilon^2}\!/_{\sigma_\xi^2}$ and K is the ratio of critical values associated with designs I and II (here, $K = t_1^2/t_2^2$).