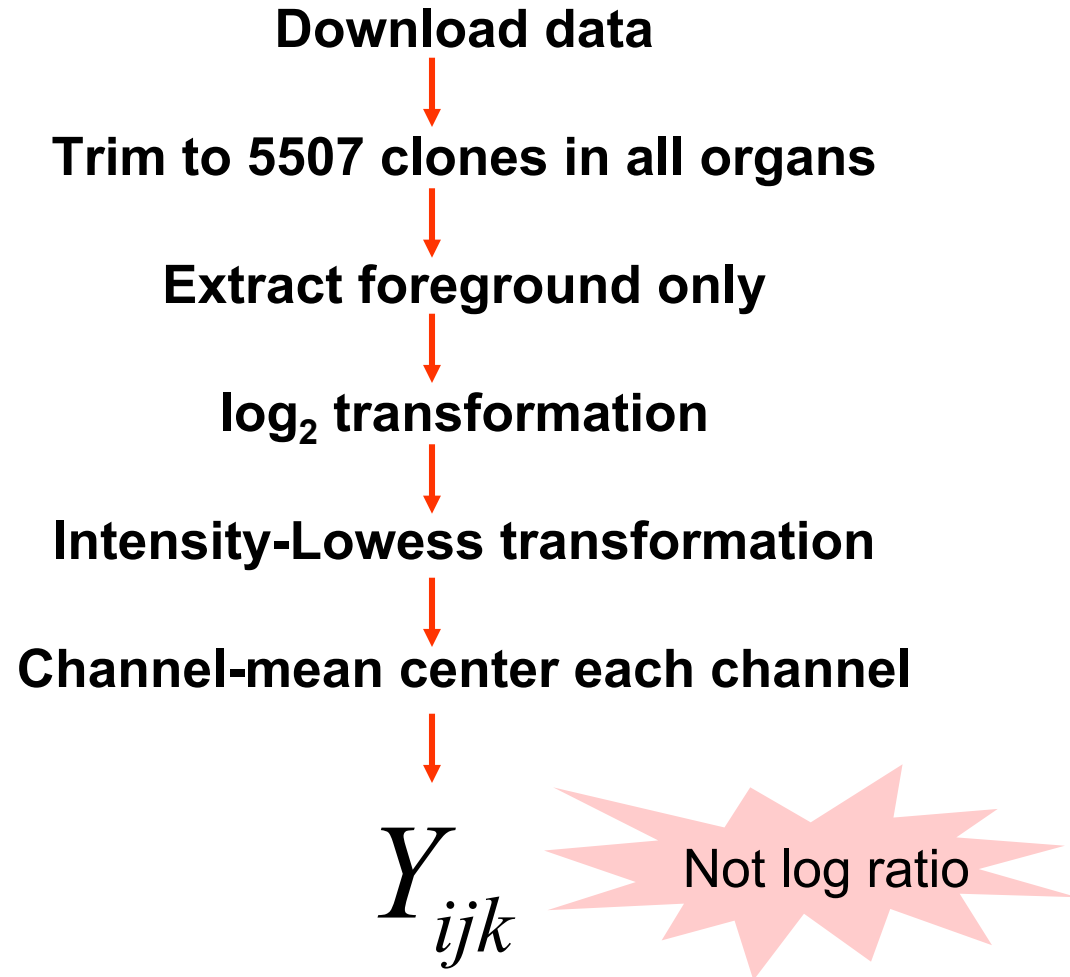


How many mice and how many arrays? Replication in cDNA microarray experiment

Xiangqin Cui

The Jackson Laboratory

Data preprocess:



Mixed linear model for each gene:

Individual organ:

$$y_{ijk} = u + M_i + D_j + A_k + R + \varepsilon_{ijk}$$

y_{ijk} : Log₂(signal)
 u : Gene mean (fixed)
 M_i : Mouse effect (random σ_m^2)
 D_j : Dye effect (fixed)
 A_k : Array effect (random σ_a^2)
 R : Reference mean (fixed)
 ε_{ijk} : Meas. error (random σ_e^2)

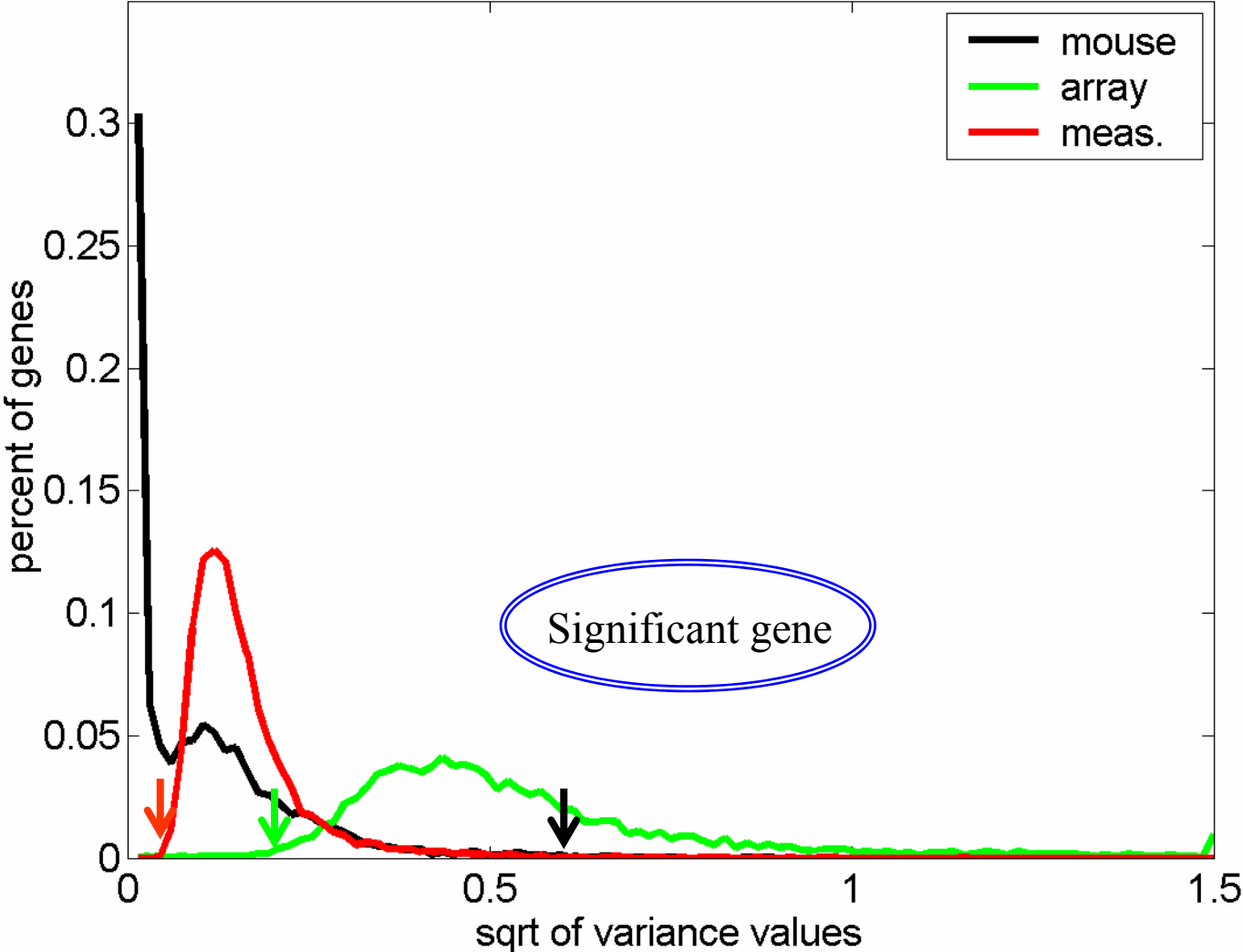
Assumptions:	$M_i \sim N(0, \sigma_m^2)$	$A_k \sim N(0, \sigma_a^2)$	$\varepsilon_{ijk} \sim N(0, \sigma_e^2)$
--------------	-----------------------------	-----------------------------	---

Combined data:

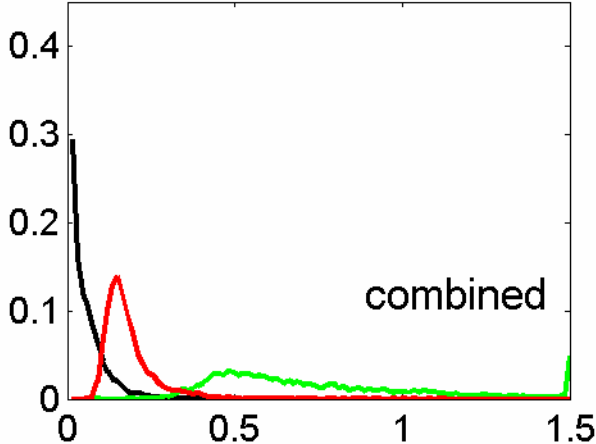
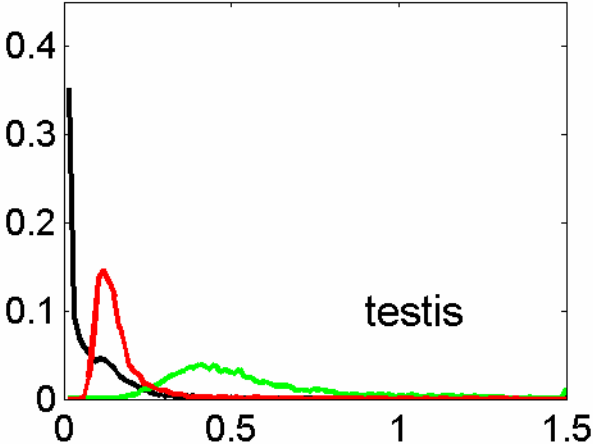
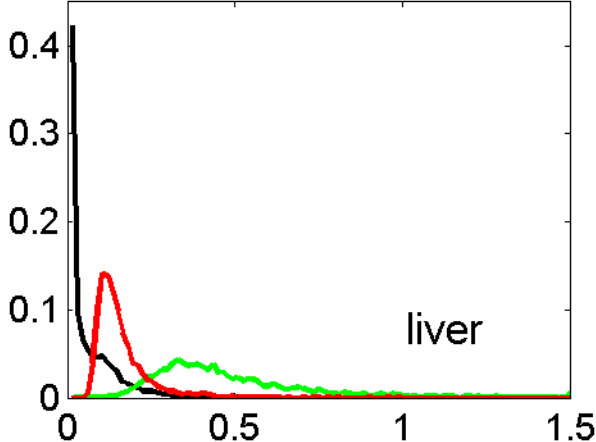
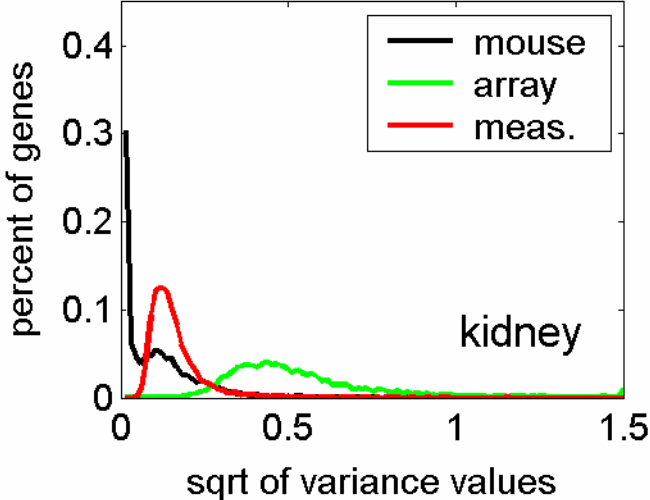
$$y_{lijk} = u + O_l + M_i + D_j + A_k + R + \varepsilon_{lijk}$$

O_l : Organ (fixed effect)

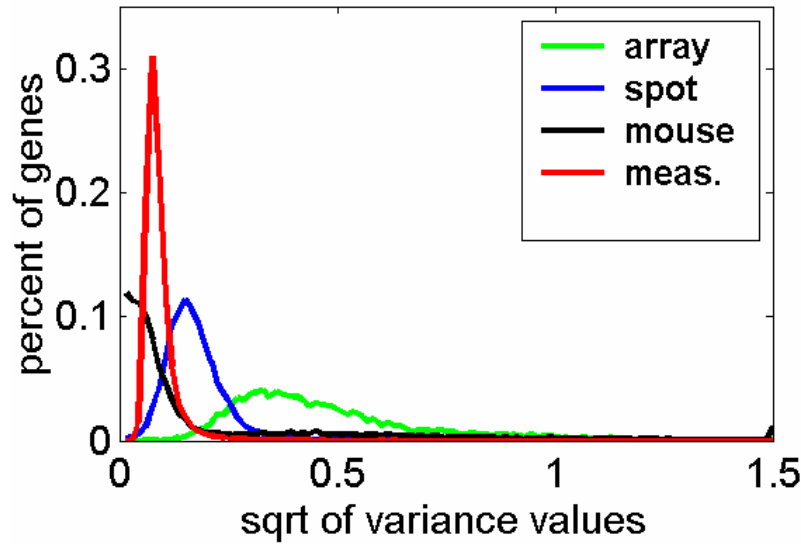
Estimated variance components from kidney



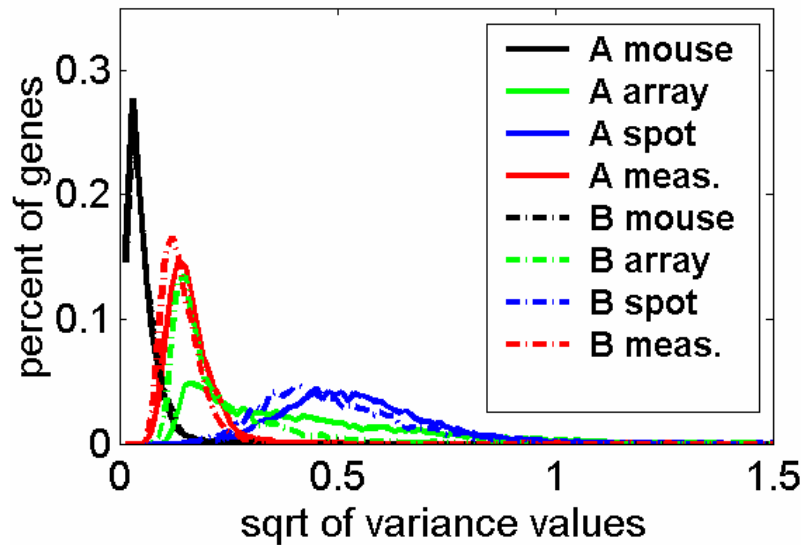
Estimated variance components from all organs



Variance components from two other data sets:

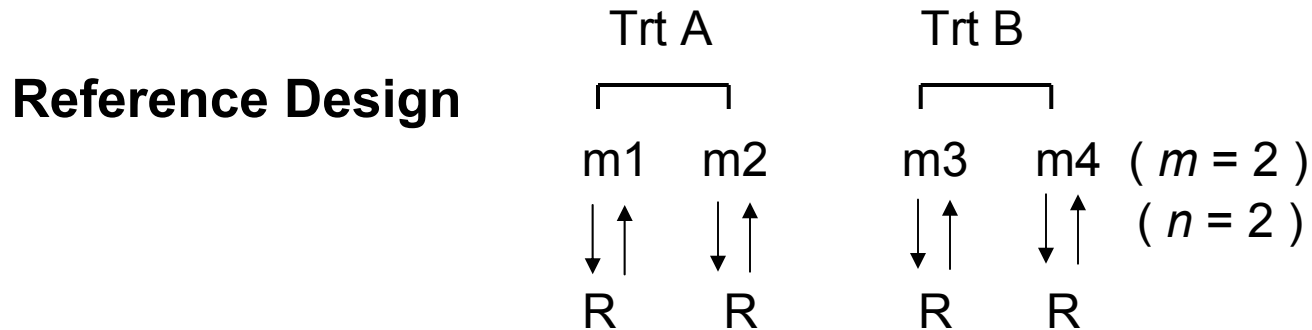


Gallstone



Brain cortex

MSE for treatment effects:



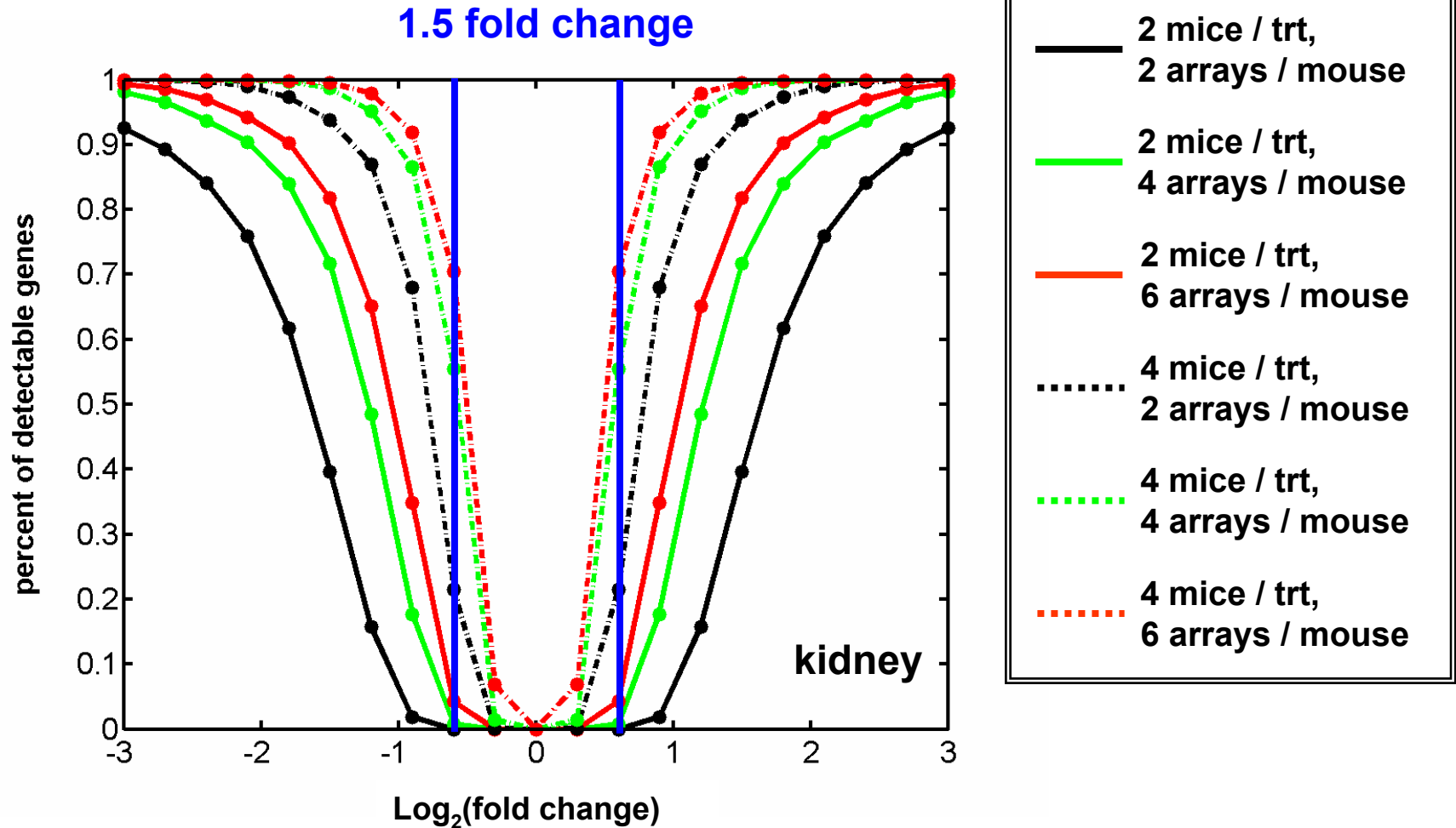
$$MSE = \frac{\sigma_m^2}{m} + \frac{\sigma_a^2}{mn} + \frac{\sigma_s^2}{mnr} + \frac{\sigma_e^2}{mnr}$$

m : number of mice per treatment

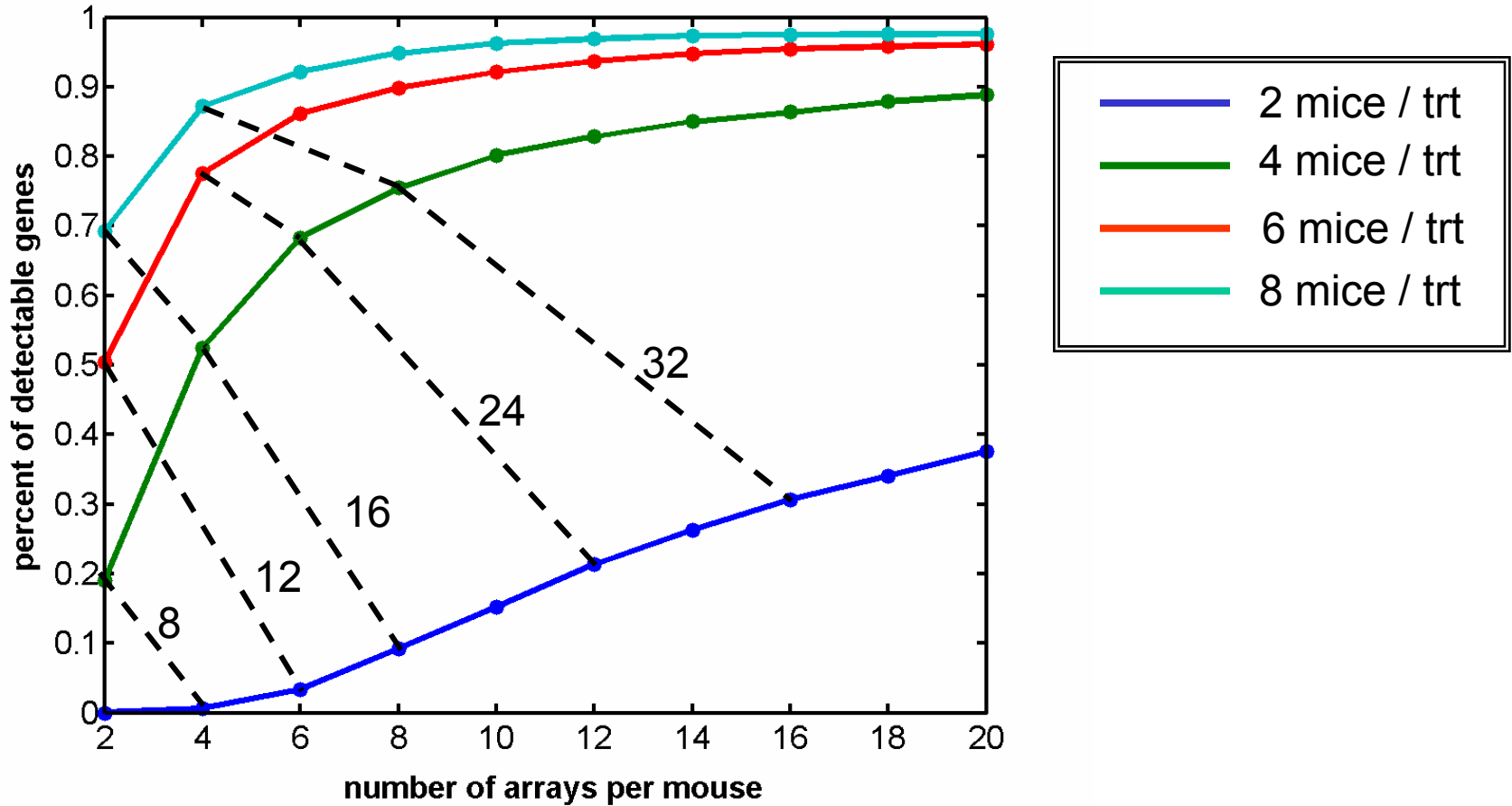
n : number of arrays per mouse

r : number of spots for each clone on array

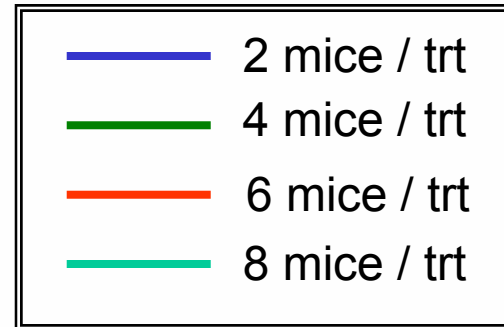
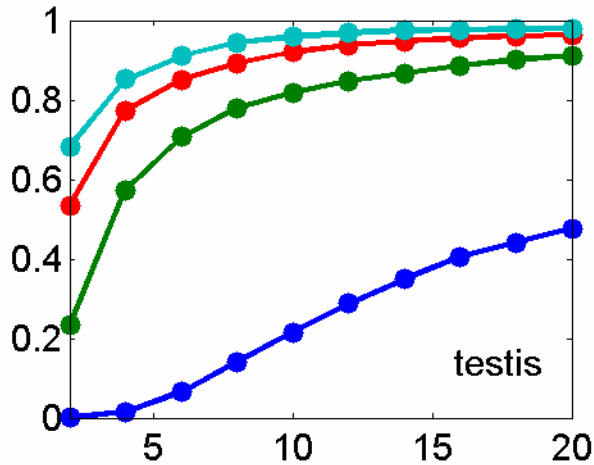
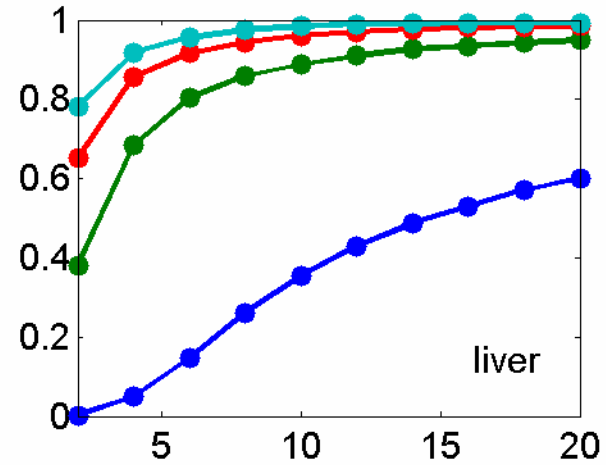
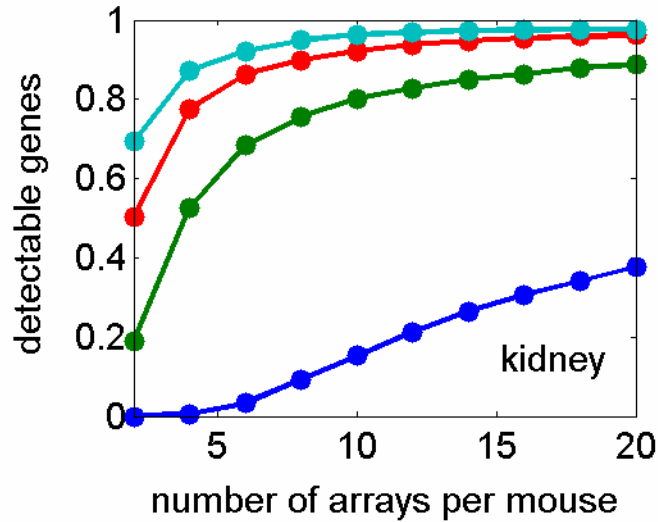
Power at different fold changes



Power to detect 1.5 fold change in kidney



Power to detect 1.5 fold change



Resource allocation:

$$MSE = \frac{\sigma_m^2}{m} + \frac{\sigma_a^2}{mn} + \frac{\sigma_s^2}{mnr} + \frac{\sigma_e^2}{mnr}$$

$$Cost = mC_m + m \cdot nC_a$$

m mice / trt
 n arrays / mouse
 r spot replicates on array

C_m cost / mouse
 C_a cost / array

σ_s^2 spot variation

The optimum number of arrays per mouse:

$$n = \sqrt{\frac{r\sigma_a^2 + \sigma_s^2 + \sigma_e^2}{r\sigma_m^2} \cdot \frac{C_m}{C_a}}$$

Examples for resource allocation

- Based on the variance components estimated from Project Normal data.
- $r = 1$ (no replicated spots on array)
- Reference design

Mouse price	Array price	# of arrays per mouse
\$15	\$300	2
\$300	\$300	7
\$1500	\$300	16

More efficient array level designs, such as direct comparisons and loop designs, can reduce the optimum number of arrays per mouse.

Pooling mice

Pooling can reduce the mouse variance but not the technical variances

$$\sigma_{pool}^2 = \frac{1}{k^\alpha} \sigma_m^2$$

k : pool size

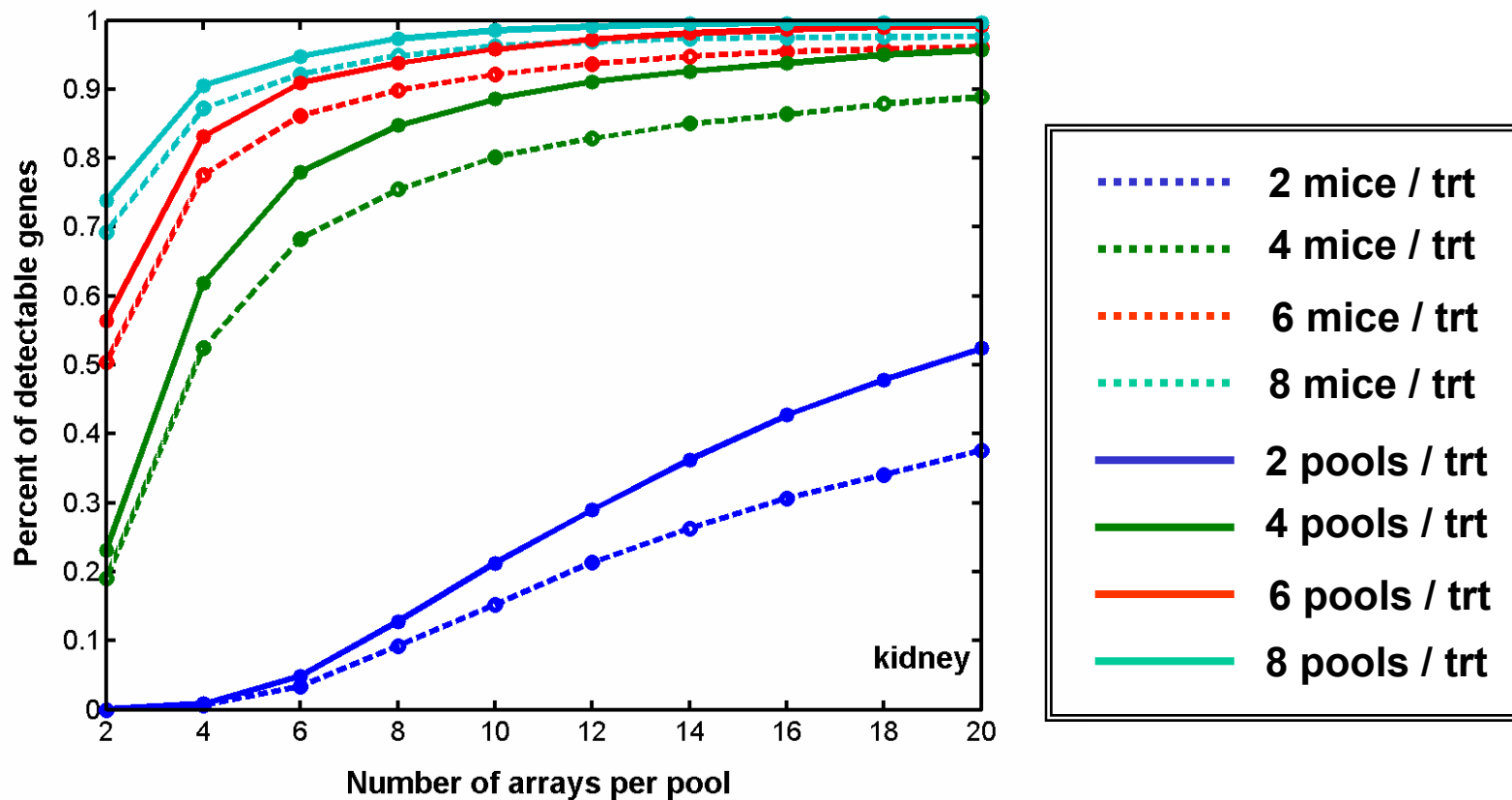
α : constant for the effect of pooling. $0 \leq \alpha \leq 1$

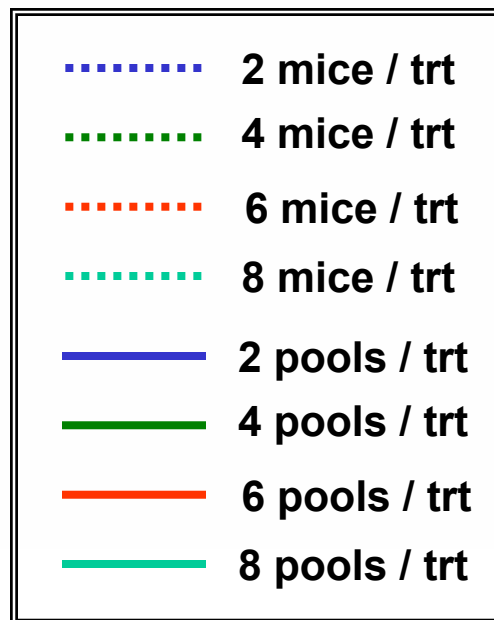
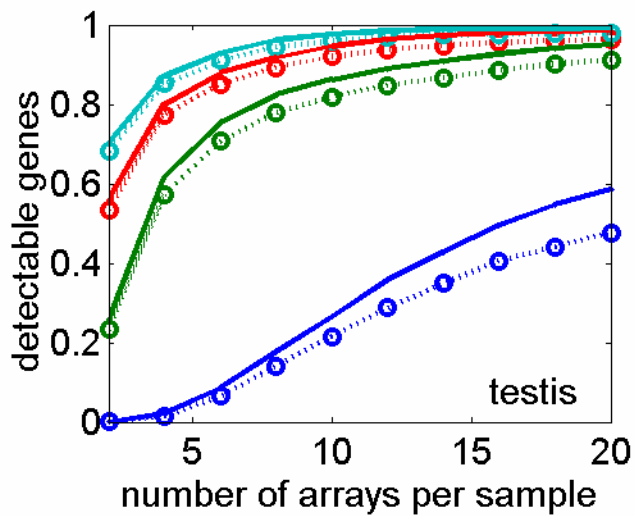
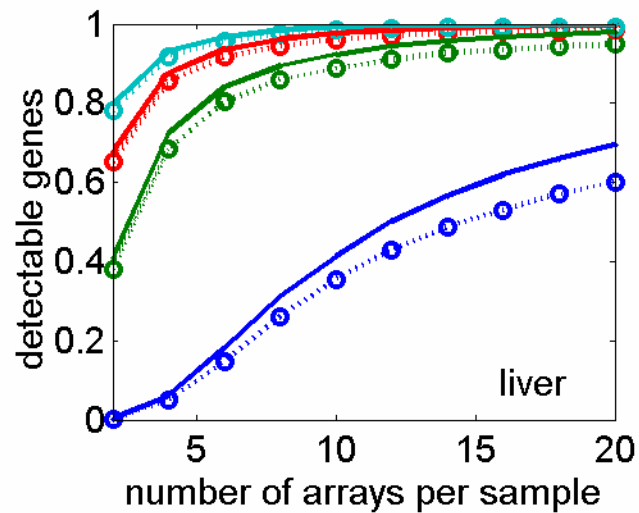
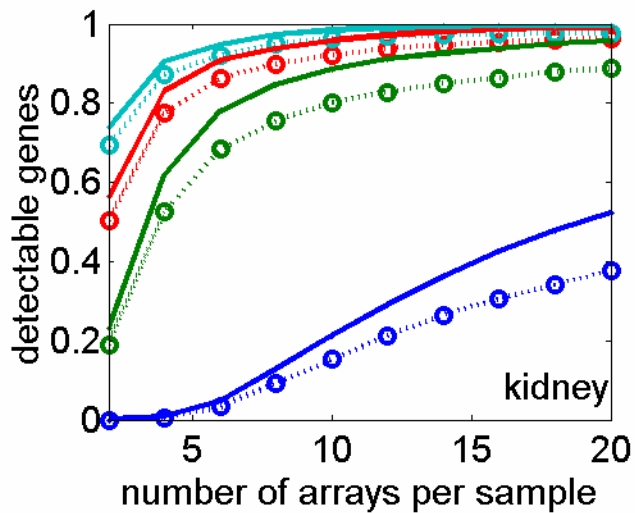
$\alpha = 0$, pooling has no effect.

$\alpha = 1$, pooling has maximum effect.

Power increase to detect 1.5 fold change by pooling

(Pool size $k = 3$, $\alpha = 1$)





Conclusions

- * Technical variation is larger than biological variation for most genes.**
- * Increase of technical replication can improve the power of the experiment effectively.**
- * Biological replication is essential for making broad-sense inferences. Increase of it is more effective in improving the power of the experiment.**
- * Pooling can reduce biological variation. However, the effect can be small due to relative small biological variation.**

Acknowledgements:

The Jackson Laboratory :

Gary A. Churchill

Hao Wu

Qian Li

Microarray facility

Beverly J. Paigen

TIGR:

John Quackenbush