

Statistical Design of Studies that Measure Gene Expression using cDNA Arrays

Bob Delongchamp
DBRA/NCTR/FDA

Experimental Design Goals for 'microarray' Center

- Use standard statistical methods
 - Linear models
 - Components of variation
- Deal with high throughput
 - Plug-n-play designs
 - Redundancy
 - Self-checks

Statistical Model

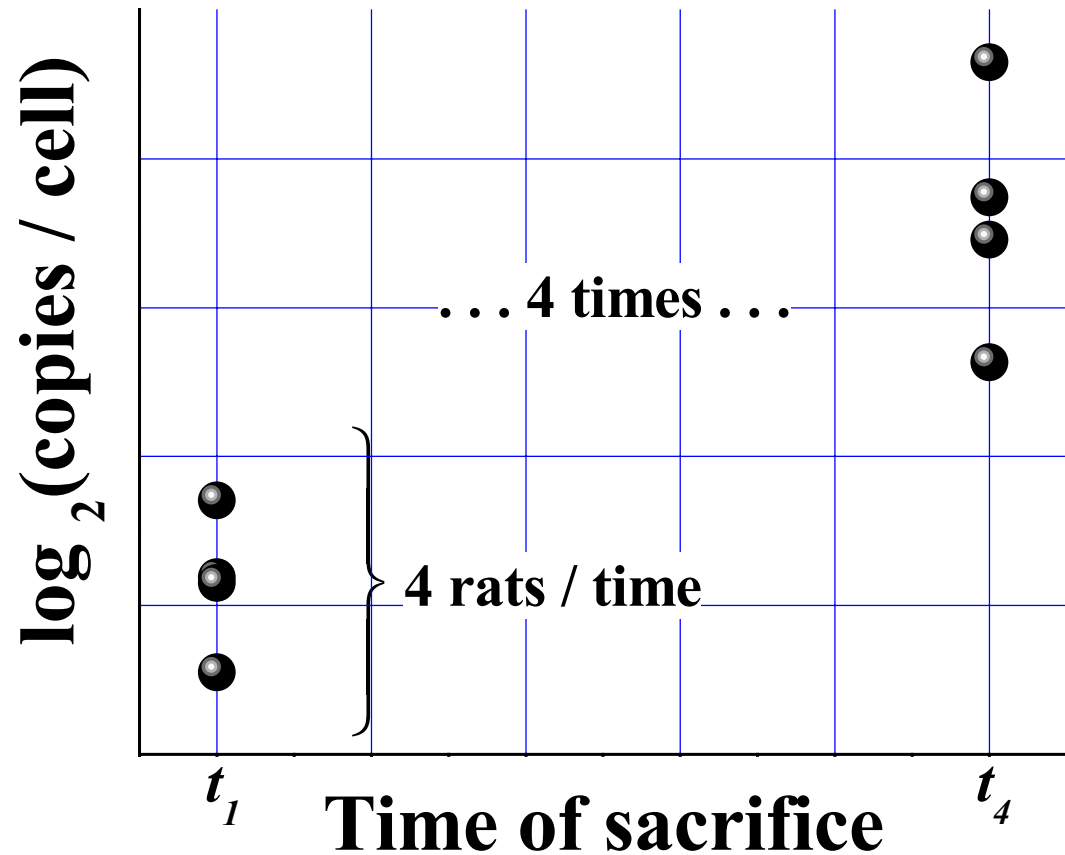
Observation = treatment effect

+ variation among exp. units

+ measurement error(s)

Circadian Study of a mRNA

- $E(y | t_i) = \mu_i$
- $\text{var}(y | t_i) = \sigma^2$
- Time effect:
ANOVA



Measuring Expression

- Copies of a mRNA specie per cell in a sample: 2^ξ ; $\xi = \mu + a$

- Basic premise: observed intensity proportional to expression:

$$I \propto 2^\xi \Rightarrow \log_2 I = k + \xi$$

- Many steps, each proportional

$$\Rightarrow y = \log_2 I = \xi + \sum_j k_j$$

Statistical Model

- $E(\log_2 I) = \mu$;
 $\text{var}(\log_2 I) = \sigma^2 = \sigma_a^2 + \sigma_m^2$
 - variance among rats (exp. units): σ_a^2
 - measurement variance: $\sigma_m^2 = \sum_j \sigma_j^2$
- $$E(k_j) = 0; \text{var}(k_j) = \sigma_j^2$$

Some Sources / Design Technique

- Variation among exp. Units (rats)
- Tissue extraction
- mRNA extraction
- cDNA prep & labeling
- Hybridization / washing / reading
- DNA spot / array
- Replicate & Randomize
- Standard method
- Standard method & blocking with a concurrent reference (BwCR)
- Randomize & BwCR

Randomized Block Design

Layout [§] array, dye and sample <i>w/i</i> block	Slide 1	Slide 2
Cy3	Sample	Reference
Cy5	Reference	Sample

[§] This layout is a dye-flip

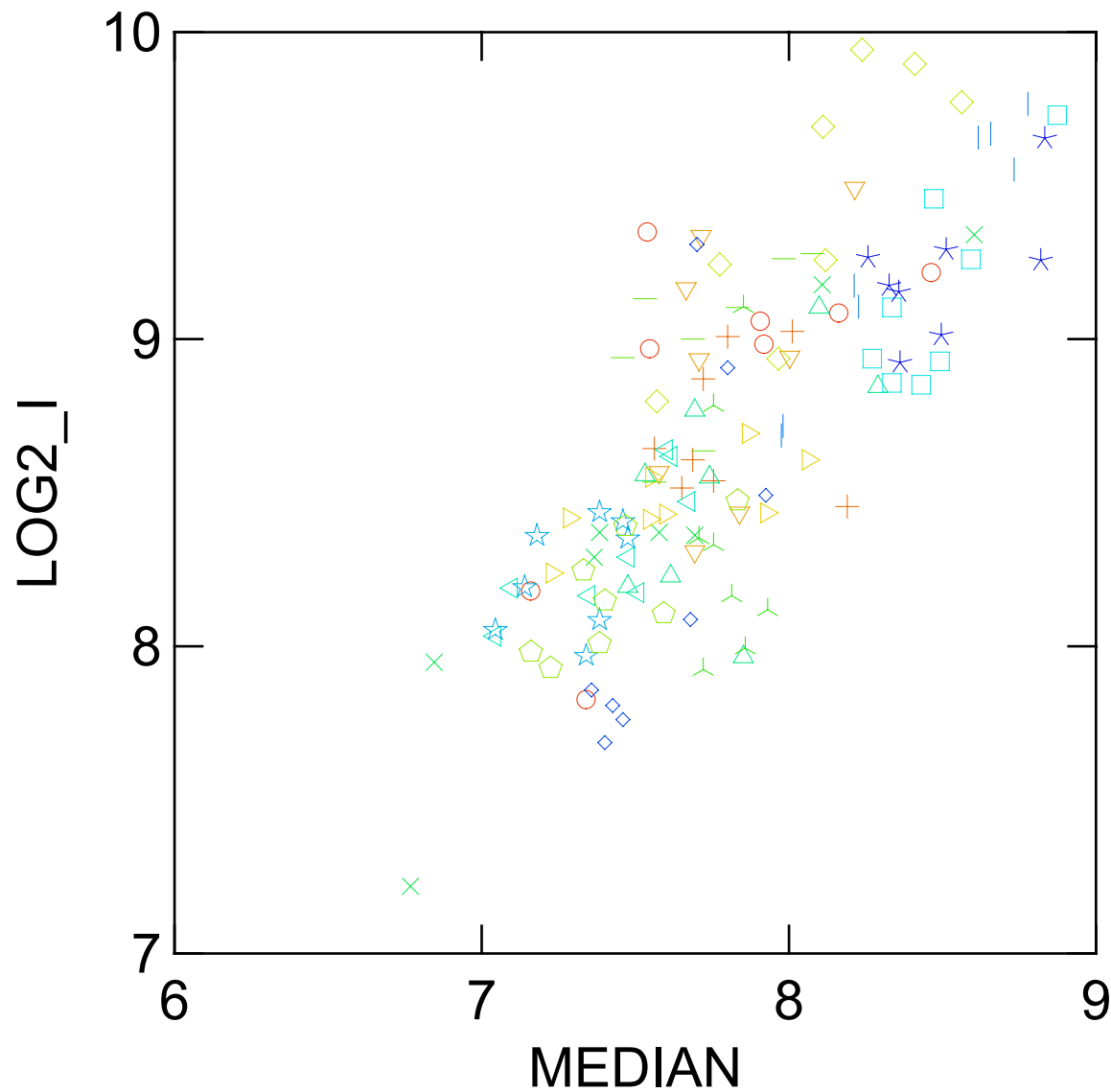
Data for Block '1a'

defender against cell death 1 (DAD-1)

block	Slide	Array	dye	sample	median	$\log_2(I)$
1a	4-13	bottom	g	1	8.16	9.08
1a	4-13	bottom	r	0	7.92	8.98
1a	4-13	top	g	1	7.91	9.06
1a	4-13	top	r	0	8.46	9.22
1a	4-26	bottom	g	0	7.55	8.97
1a	4-26	bottom	r	1	7.54	9.35
1a	4-26	top	g	0	7.16	8.18
1a	4-26	top	r	1	7.34	7.83

defender against cell death 1 (DAD-1)

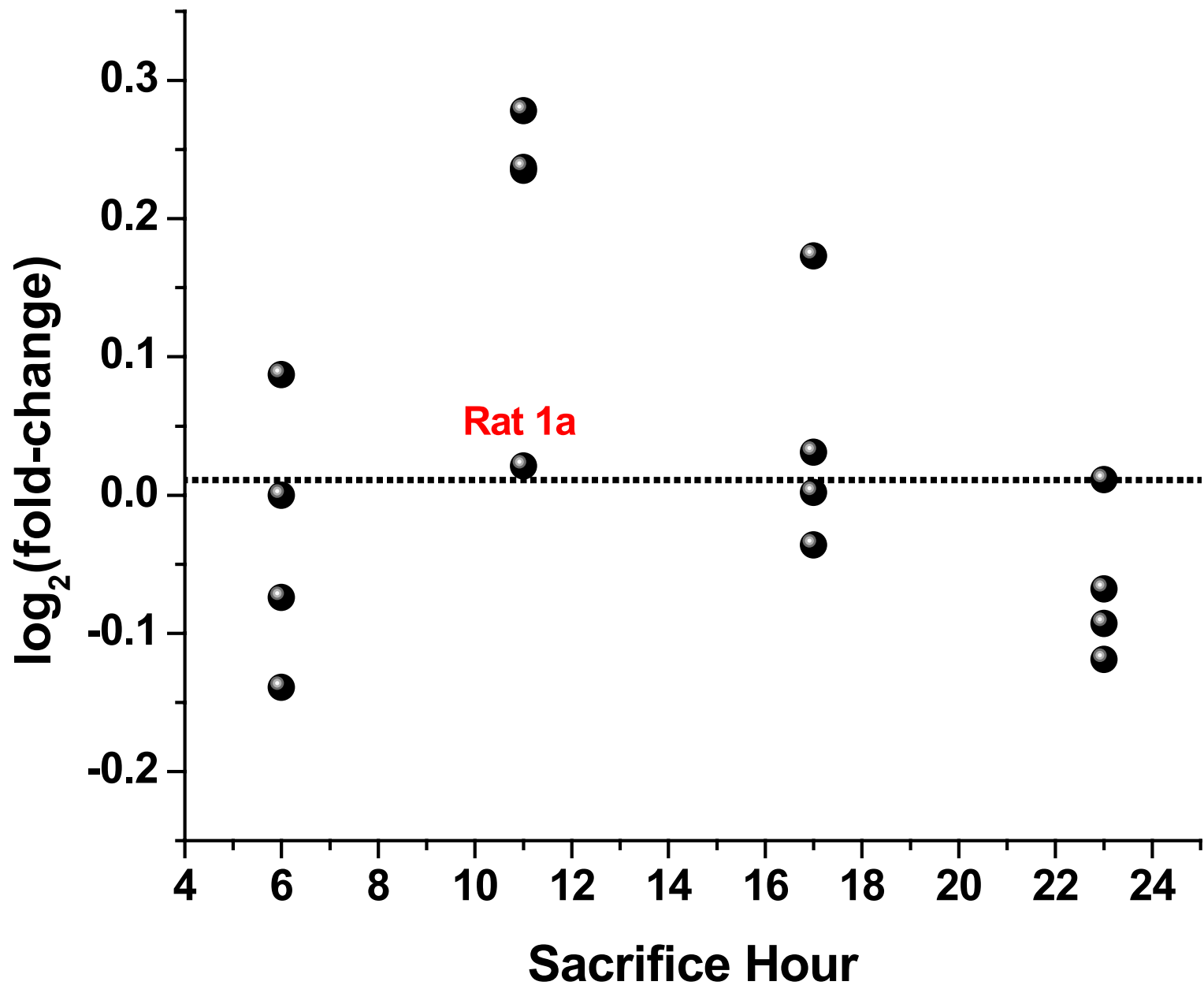
BLOCK



$\log_2(\text{sample}) - \log_2(\text{reference})$

defender against cell death 1 (DAD-1)

block	Estimate	StdErr	t	prob	Sacrifice Hour
1a	0.021	0.196	0.110	0.916	11
1b	0.235	0.197	1.190	0.238	11
1c	0.278	0.196	1.420	0.162	11
1d	0.237	0.199	1.190	0.237	11
2a	0.011	0.198	0.050	0.957	23
2b	-0.119	0.199	-0.590	0.554	23
2c	-0.068	0.196	-0.350	0.729	23
2d	-0.093	0.200	-0.470	0.643	23
3a	0.002	0.200	0.010	0.991	17
3b	0.031	0.196	0.160	0.874	17
3c	0.173	0.198	0.870	0.386	17
3d	-0.036	0.196	-0.180	0.855	17
4a	0.087	0.203	0.430	0.669	6
4b	-0.139	0.197	-0.710	0.483	6
4c	0.000	0.196	0.000	0.999	6
4d	-0.074	0.198	-0.370	0.710	6



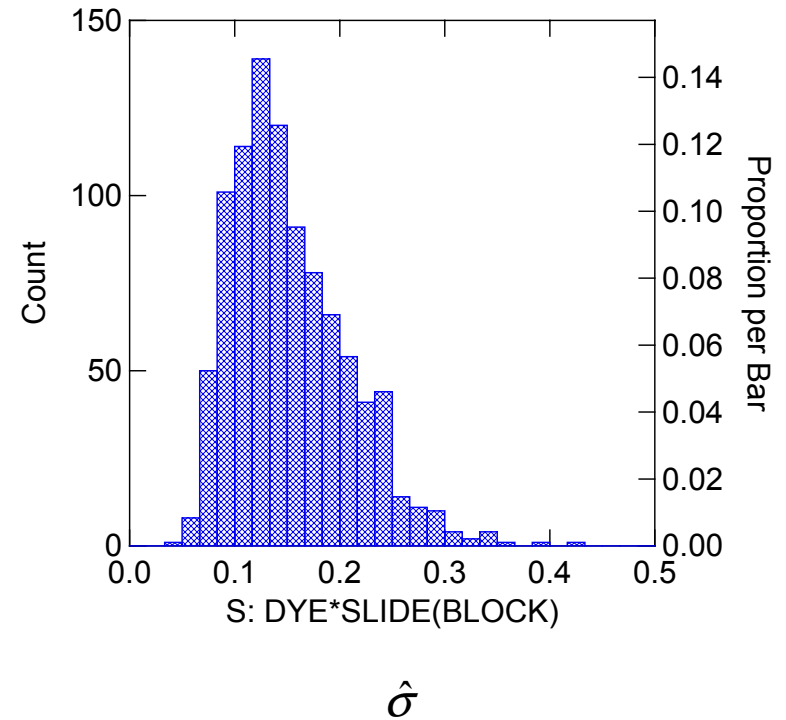
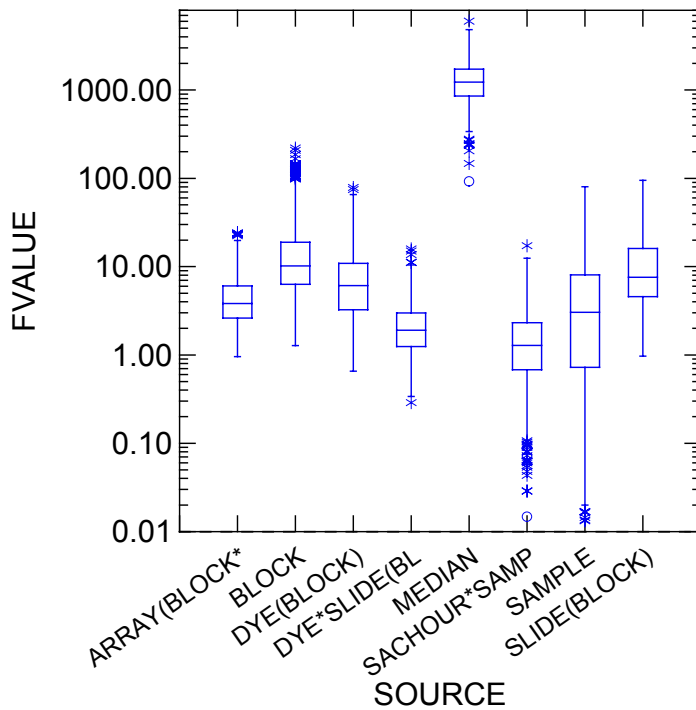
Name = defender against cell death 1 (DAD-1)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
median	1	23.366	23.366	931.01	<.0001
block	15	5.843	0.389	15.52	<.0001
Slide(block)	16	2.940	0.183	7.32	<.0001
dye(block)	16	0.507	0.031	1.26	0.2790
sample	1	0.046	0.046	1.84	0.1849
SacHour*sample	3	0.309	0.103	4.11	0.0144
dye*Slide(block)	12	0.205	0.017	0.68	0.7564
Array(block*Slide)	32	4.052	0.126	5.05	<.0001
Error	31	0.778	0.025		
Corrected Total	127	38.049			

Tests Using the Type I MS for dye*Slide(block) as an Error Term

sample	1	0.046	0.046	2.70	0.1262
SacHour*sample	3	0.309	0.103	6.04	0.0095

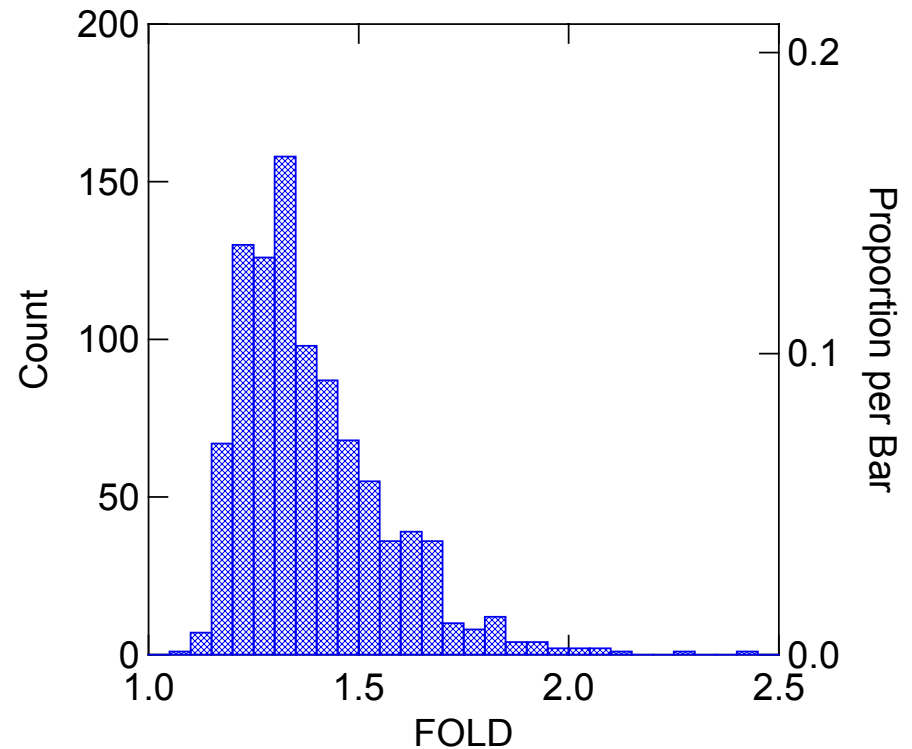
Sources of Variation



955 genes

Statistical Power

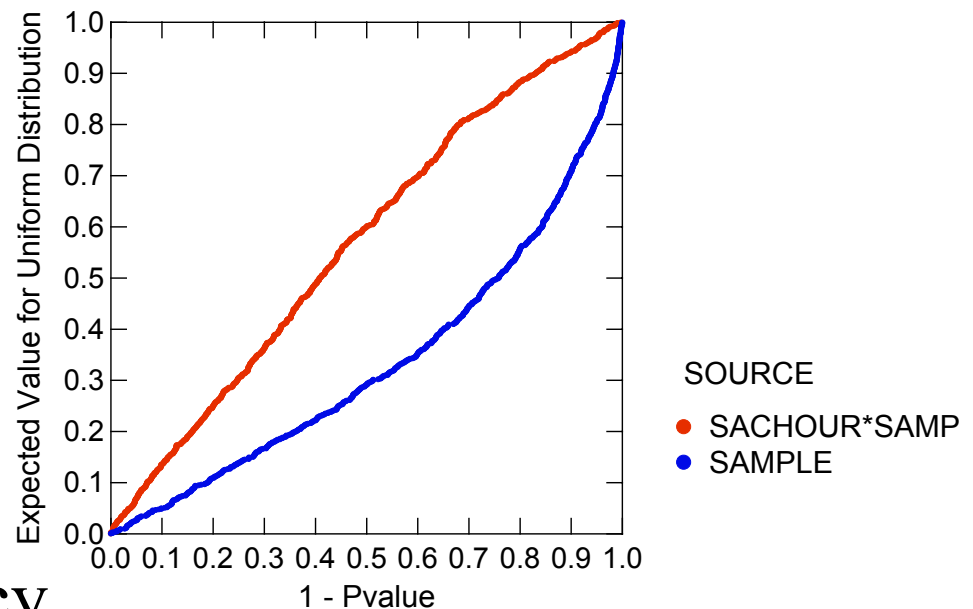
- Pairwise comparison of two sacrifice times
- Fold changes that would be significant at $p = 0.001$ level
- Basis for power calculations in future studies



955 genes

Multiple testing / checks

- P-value plot suggests no excess in significant genes
- Sample checks the hypothesis that average of samples equals average of reference (consistency check?)



Acknowledgements

Division of Biometry and Risk Assessment

- **Cruz Velasco (LSU)**
- **Jim Chen**
- Ralph Kodell
- Visiting ORISE faculty
 - Honshik Ahn, SUNY Stony Brook
 - Mehdi Razzaghi, Bloomsberg U
 - Segun George, U of Memphis
 - Ronnie West, U South Carolina
- **Gary Churchill**, The Jackson Laboratory

Functional Genomics Center

- **Jim Fuscoe**
- **Varsha Desai**

Data Management / Informatics

- Weida Tong
- Steve Harris
- Xiaoxi (Megan) Cao

NCTR Investigators

- Angela Harris
- John Bowyer

