

A Quantitative Overview to Gene Expression Profiling in Animal Genetics

————— **MICROARRAY EXPERIMENTS** —————

————— **Technical Concerns** —————

1. Biochemist Level :

- a. Preparation (**Printing**) of the Chip
- b. RNA Extraction, Amplification and Hybridisation
- c. Optical Scanner (Reading)

2. Quantitative Level :

- a. Design —————→
- b. Image (data) Quality —————→
- c. Data Analysis —————→
- d. Data Storage

Repl i cati on:

1. Ani mal
2. Samp le
3. Array
- 4. Spot**

Note: Randomisation intentionally neglected.

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



TECHNICAL CONCERNS

BASIC PIECES FOR SIGNAL DETECTION

- Foreground **RED** and **GREEN** \Rightarrow R_f G_f
- Background **RED** and **GREEN** \Rightarrow R_b G_b
- Background-corrected **RED** \Rightarrow $R = R_f - R_b$ } *True Signals!*
GREEN \Rightarrow $G = G_f - G_b$
- Log-transformed $\text{Log}_2(\mathbf{R})$
 $\text{Log}_2(\mathbf{G})$
- Difference: "Minus" $\mathbf{M} = \text{Log}_2(\mathbf{R}) - \text{Log}_2(\mathbf{G}) = \text{Log}_2(\mathbf{R}/\mathbf{G})$
- Mean: "Average" $\mathbf{A} = 0.5 * (\text{Log}_2(\mathbf{R}) + \text{Log}_2(\mathbf{G})) = 0.5 * \text{Log}_2(\mathbf{R} * \mathbf{G})$
- **MA-Plots** ...to come



TECHNICAL CONCERNS

2.d – Data Storage:

RELATIONAL DATABASES FOR MICROARRAY

BASE: BioArray Software Environment: A Platform for Comprehensive Management and Analysis of Microarray Data

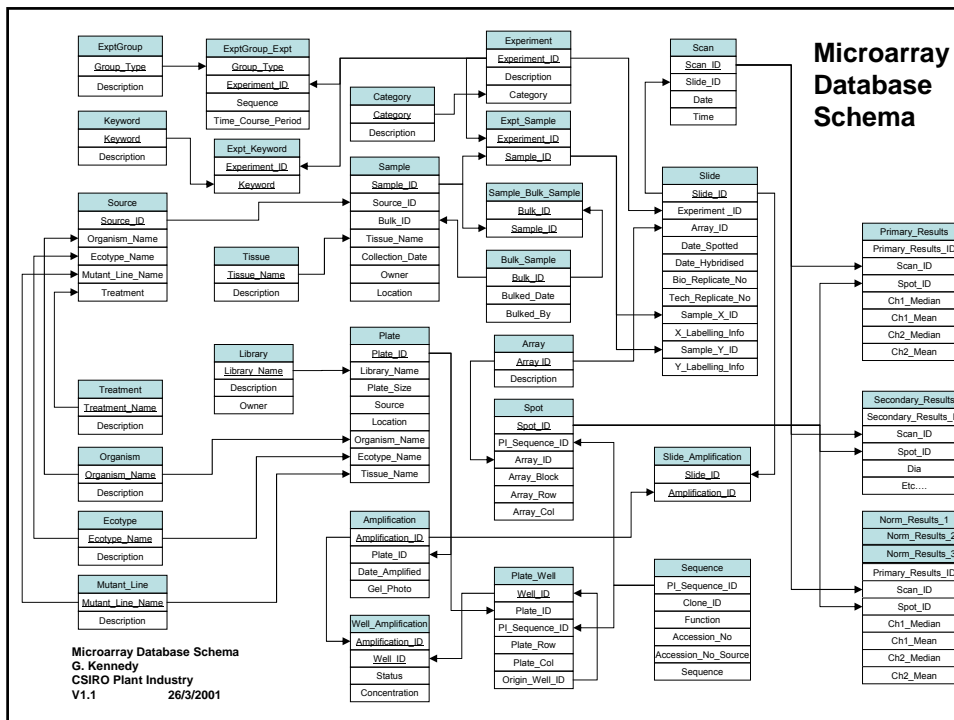
Lao H. Saal, Carl Troein, Johan Vallon-Christersson, Sofia Gruberger, Åke Borg and Carsten Peterson
Genome Biology 2002 3(8): software0003.1-0003.6
<http://base.thep.lu.se/index.phtml>

GENA: Genomics Array Database

CSIRO Plant Industries & CMIS
<http://www.pi.csiro.au/gena/>

GEXEX: Gene Expression Experiments

CSIRO Livestock Industries
<https://www.biolives.li.csiro.au/gexex/>



A Quantitative Overview to Gene Expression Profiling in Animal Genetics

CSIRO **TECHNICAL CONCERNS**

2.a – Data Storage:

OPINION

The level of sophistication becomes so high that it is unrealistic to expect an automatic adoption of this system by the end user.

SOLUTION

A simple intuitive graphical interface warehousing system to simultaneously access (i) details of the design configuration, and (ii) the entire raw data.

Armistead Animal Breeding Summer Course, UNE, Feb. 2006



TECHNICAL CONCERNS

2.a – Data Storage: GEXEX

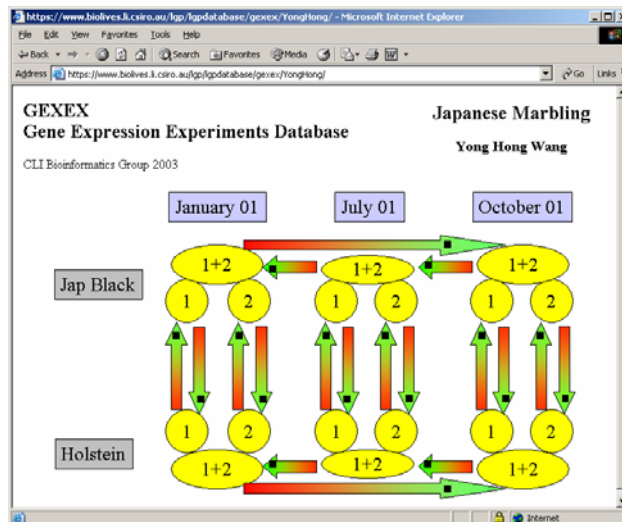


Armidale Animal Breeding Summer Course, UNE, Feb. 2006



TECHNICAL CONCERNS

2.a – Data Storage: GEXEX



Armidale Animal Breeding Summer Course, UNE, Feb. 2006



TECHNICAL CONCERNS

2.a – Data Storage: GEXEX

1	10	6	"LBO40C3"	"CCL014570"	3500	6380	120	71	84	67	57
1	11	6	"LBO40B6"	"CCL014564"	3720	6380	70	1066	1372	1275	49
1	12	6	"LBO40B6"	"CCL014564"	3940	6370	70	1221	2269	3310	58
1	13	6	"LBO39G5"	"CCL014546"	4170	6380	90	138	339	607	52
1	14	6	"LBO39G5"	"CCL014546"	4400	6380	70	228	422	570	63
1	15	6	"LBO39F5"	"CCL014540"	4600	6380	120	68	80	53	55
1	16	6	"LBO39F5"	"CCL014540"	4820	6380	120	66	93	146	53
1	17	6	"LBO39C9"	"CCL014522"	5040	6360	90	109	139	129	59
1	18	6	"LBO39C9"	"CCL014522"	5260	6360	120	96	123	141	62
1	19	6	"LBO39B5"	"CCL014516"	5480	6380	120	56	60	49	56
1	20	6	"LBO39B5"	"CCL014516"	5700	6380	120	39	87	88	38
1	1	7	"LBO56H3"	"CCL015314"	1520	6590	120	71	76	47	60
1	2	7	"LBO56H3"	"CCL015314"	1740	6590	120	59	70	49	59
1	3	7	"LBO56G1"	"CCL015308"	1960	6590	120	71	90	66	54
1	4	7	"LBO56G1"	"CCL015308"	2180	6590	120	70	94	70	55
1	5	7	"LBO56G11"	"CCL015290"	2420	6600	60	211	207	99	58
1	6	7	"LBO56G11"	"CCL015290"	2620	6590	60	140	170	102	50
1	7	7	"LBO56B8"	"CCL015284"	2840	6590	80	547	1043	1094	54
1	8	7	"LBO56B8"	"CCL015284"	3070	6600	100	233	633	821	56
1	9	7	"LBO56G10"	"CCL015266"	3280	6600	120	49	62	44	56
1	10	7	"LBO56G10"	"CCL015266"	3500	6600	120	56	63	42	54
1	11	7	"LBO55F9"	"CCL015260"	3740	6610	70	176	620	1123	50
1	12	7	"LBO55F9"	"CCL015260"	3940	6600	120	05	266	626	53
1	13	7	"LBO48D6"	"CCL014954"	4160	6600	120	46	62	46	33
1	14	7	"LBO48D6"	"CCL014954"	4380	6600	120	37	83	161	54
1	15	7	"LBO48C5"	"CCL014948"	4620	6590	110	202	268	195	53
1	16	7	"LBO48C5"	"CCL014948"	4830	6600	110	193	235	186	54
1	17	7	"LBO47G8"	"CCL014930"	5040	6600	120	58	70	54	57
1	18	7	"LBO47G8"	"CCL014930"	5260	6600	120	52	65	57	55
1	19	7	"LBO47F5"	"CCL014924"	5510	6590	110	202	263	205	52
1	20	7	"LBO47F5"	"CCL014924"	5730	6600	100	186	281	328	36
1	1	8	"LBO66A12"	"CCL015722"	1540	6820	120	202	277	350	56

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

“Biologists interested in gene expression profiling should feel free to match experimental design to their particular situation; **there is no universal microarray design.**”

A careful grounding in the principles of experimental design will help to ensure that we will accumulate knowledge and **not just enormous amount of data.**”

*Churchill & Oliver, 2001.
Sex, flies, and microarrays.
Nature Genetics, 29:355.*

- Accommodate your software to your design, not the other way around.
- Beef CRC Database
- Type I Error (False Positives)
- **Type III Error** (Correctly detecting an effect, but Incorrectly attributing the cause).

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

Key Issues:

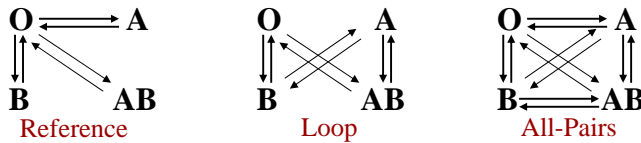
- a. Identify/Prioritise Questions
- b. N of Available Samples
- c. N of Available Arrays
- d. Consider Dye Bias

Put more arrays on key questions

Pooling?

- Dye-Swap
- Dye-Balancing
- Self-Self

Evaluation of Designs:



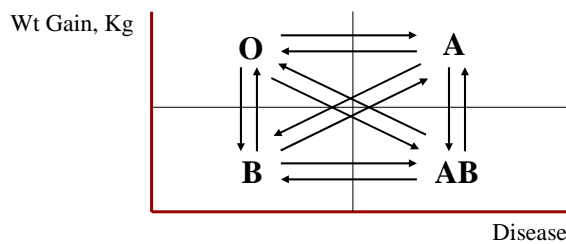
Variance of Estimated Effects (Relative to the All-Pairs)

	Reference	Loop	All-Pairs
Main effect of A	1	4/3	1
Main effect of B	1	1	1
Interaction AB	3	8/3	2
Contrast A-B	2	1	1

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN



Model:

$$\begin{aligned} \mu_O &= \mu \\ \mu_A &= \mu + \alpha \\ \mu_B &= \mu + \beta \\ \mu_{AB} &= \mu + \alpha + \beta + \gamma \end{aligned}$$

The ratio:

$$M_{A,AB} = \text{Log} \frac{R_A}{G_{AB}} = \text{Log}(R_A) - \text{Log}(G_{AB})$$

estimates

$$\mu_A - \mu_{AB} = \beta + \gamma$$

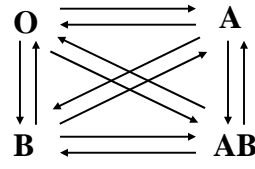
Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

Model: $\mu_O = \mu$
 $\mu_A = \mu + \alpha$
 $\mu_B = \mu + \beta$
 $\mu_{AB} = \mu + \alpha + \beta + \gamma$

$$\begin{bmatrix} M_{O.A} \\ M_{A.O} \\ M_{O.B} \\ M_{B.O} \\ M_{O.AB} \\ M_{AB.O} \\ M_{A.B} \\ M_{B.A} \\ M_{A.AB} \\ M_{AB.A} \\ M_{B.AB} \\ M_{AB.B} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + Error$$



$$M = X\theta + E$$

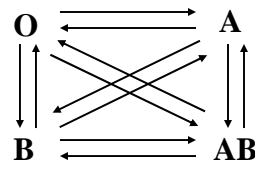
$$\hat{\theta} = (X^T X)^{-1} X^T M$$



All Pairs

Model: $\mu_O = \mu$
 $\mu_A = \mu + \alpha$
 $\mu_B = \mu + \beta$
 $\mu_{AB} = \mu + \alpha + \beta + \gamma$

$$\begin{bmatrix} M_{O.A} \\ M_{A.O} \\ M_{O.B} \\ M_{B.O} \\ M_{O.AB} \\ M_{AB.O} \\ M_{A.B} \\ M_{B.A} \\ M_{A.AB} \\ M_{AB.A} \\ M_{B.AB} \\ M_{AB.B} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + Error$$



$$M = X\theta + E$$

$$(X^T X) = \begin{pmatrix} 8 & 0 & 4 \\ 0 & 8 & 4 \\ 4 & 4 & 6 \end{pmatrix}$$

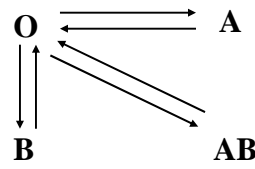


A Quantitative Overview to Gene Expression Profiling in Animal Genetics

Reference

Model: $\mu_O = \mu$
 $\mu_A = \mu + \alpha$
 $\mu_B = \mu + \beta$
 $\mu_{AB} = \mu + \alpha + \beta + \gamma$

$$\begin{bmatrix} M_{O.A} \\ M_{A.O} \\ M_{O.B} \\ M_{B.O} \\ M_{O.AB} \\ M_{AB.O} \\ M_{A.B} \\ M_{B.A} \\ M_{A.AB} \\ M_{AB.A} \\ M_{B.AB} \\ M_{AB.B} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + Error$$



$$M = X\theta + E$$

$$(X^T X) = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

Armidale Animal Breeding Summer Course, UNE, Feb. 2006

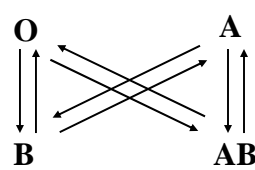


A Quantitative Overview to Gene Expression Profiling in Animal Genetics

Loop

Model: $\mu_O = \mu$
 $\mu_A = \mu + \alpha$
 $\mu_B = \mu + \beta$
 $\mu_{AB} = \mu + \alpha + \beta + \gamma$

$$\begin{bmatrix} M_{O.A} \\ M_{A.O} \\ M_{O.B} \\ M_{B.O} \\ M_{O.AB} \\ M_{AB.O} \\ M_{A.B} \\ M_{B.A} \\ M_{A.AB} \\ M_{AB.A} \\ M_{B.AB} \\ M_{AB.B} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \\ 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} + Error$$



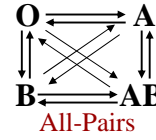
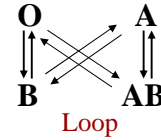
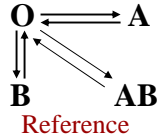
$$M = X\theta + E$$

$$(X^T X) = \begin{pmatrix} 4 & 0 & 2 \\ 0 & 8 & 4 \\ 2 & 4 & 4 \end{pmatrix}$$

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Evaluation of Designs:



$$(X^T X)^{-1} = \begin{pmatrix} 0.5 & 0 & -0.5 \\ 0 & 0.5 & -0.5 \\ -0.5 & -0.5 & 1.5 \end{pmatrix} \quad \begin{pmatrix} 0.5 & 0.25 & -0.5 \\ 0.25 & 0.375 & -0.5 \\ -0.5 & -0.5 & 1.0 \end{pmatrix} \quad \begin{pmatrix} 0.25 & 0.125 & -0.25 \\ 0.125 & 0.25 & -0.25 \\ -0.25 & -0.25 & 0.5 \end{pmatrix}$$

Variance of Estimated Effects (Relative to the All-Pairs)

	Reference	Loop	All-Pairs
Main effect of A	1	4/3	1
Main effect of B	1	1	1
Interaction AB	3	8/3	2
Contrast A-B	2	1	1



MICROARRAY EXPERIMENTAL DESIGN (Time-course)

Yang & Speed, 2002

http://www.nature.com - Hi-res table window - Microsoft Internet Explorer

Nature Reviews Genetics 3, 579-588 (2002); doi:10.1038/nrg863 < previous next >

DESIGN ISSUES FOR CDNA MICROARRAY EXPERIMENTS

Table 2 | Time-course experiments

Design choices	t versus t + 1			Comparisons t versus t + 2		t versus t + 3	Average variance
	t ₁ /t ₂	t ₂ /t ₃	t ₃ /t ₄	t ₁ /t ₃	t ₂ /t ₄	t ₁ /t ₄	
Design I – T1 as common reference T1 → T2 → T3 → T4	1.00	2.00	2.00	1.00	2.00	1.00	1.5
Design II – direct: sequential T1 → T2 → T3 → T4	1.00	1.00	1.00	2.00	2.00	3.00	1.67
Design III – common reference T1 → T2 → T3 → T4	2.00	2.00	2.00	2.00	2.00	2.00	2.00
Design IV – T1 as common reference T1 → T2 → T3 → T4	0.67	0.67	1.67	0.67	1.67	1.00	1.06
Design V – direct: loop T1 → T2 → T3 → T4	0.75	0.75	0.75	1.00	1.00	0.75	0.83
Design VI – direct: mixed T1 → T2 → T3 → T4	1.00	0.75	1.00	0.75	0.75	0.75	0.83

Variance of estimated effects for six different designs of time-course experiments. Designs I and II involve only three slides and the remaining designs involve four. σ² was set to 1 throughout.

3 slides

4 slides



MICROARRAY EXPERIMENTAL DESIGN

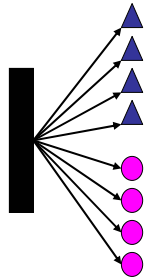
e^2 = Error Variance

n = No. of Replicates

p^2 = Population Variance

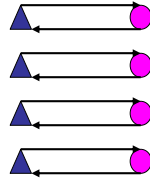
$2n$ = Total No. of Chips

Reference



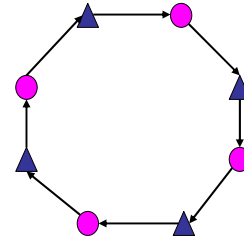
$$\frac{1}{n}(4e^2 + 2p^2)$$

Multiple Dye-Swap



$$\frac{1}{n}(e^2 + 2p^2)$$

Loop



$$\frac{1}{n}(e^2 + 2p^2)$$

Conclusion: Relative size of e^2 to p^2 will dictate the optimal design

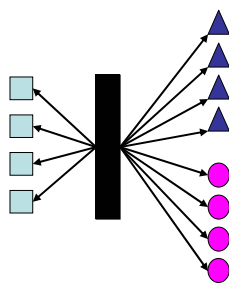
Kerr 2003. Biometrics 59:822-828

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



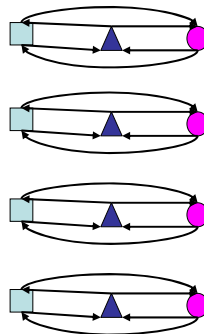
MICROARRAY EXPERIMENTAL DESIGN

Reference



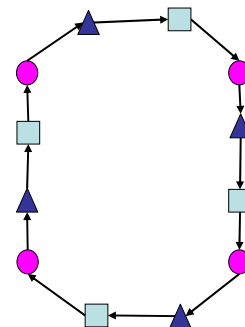
12 Chips

Multiple Dye-Swap



24 Chips

Loop



12 Chips

Conclusion: Loops require as many chips as samples

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

Glonek & Solomon

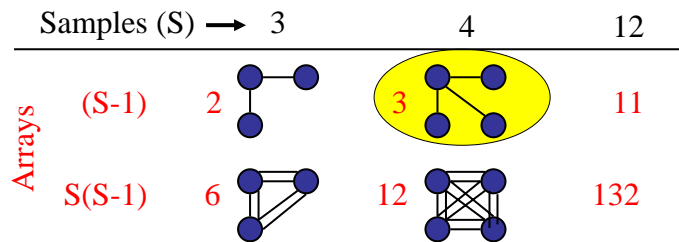
Factorial and Time Course Designs for cDNA Microarray Experiments

• **Definition**

A design with a total of n slides and design matrix X is said to be **admissible** if there exists no other design with n slides and design matrix X^* such that $ci^* \leq ci$

For all i with strict inequality for at least one i . Where ci^* and ci are respectively the diagonal elements of $(X^*X^*)^{-1}$ and $(X'X)^{-1}$.

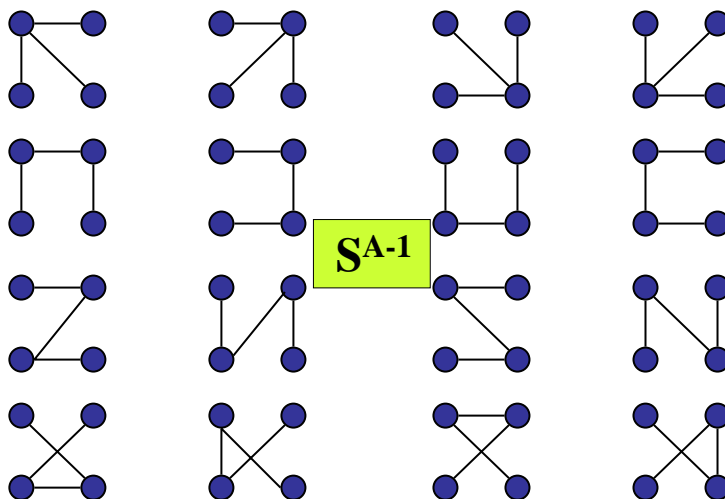
• **Samples vs Slides vs Configurations**



N of Configurations?



N of Configurations?





N of Configurations?

Wool Pigmentation

Pie-Bald black

Non-Pie-Bald black



Normal

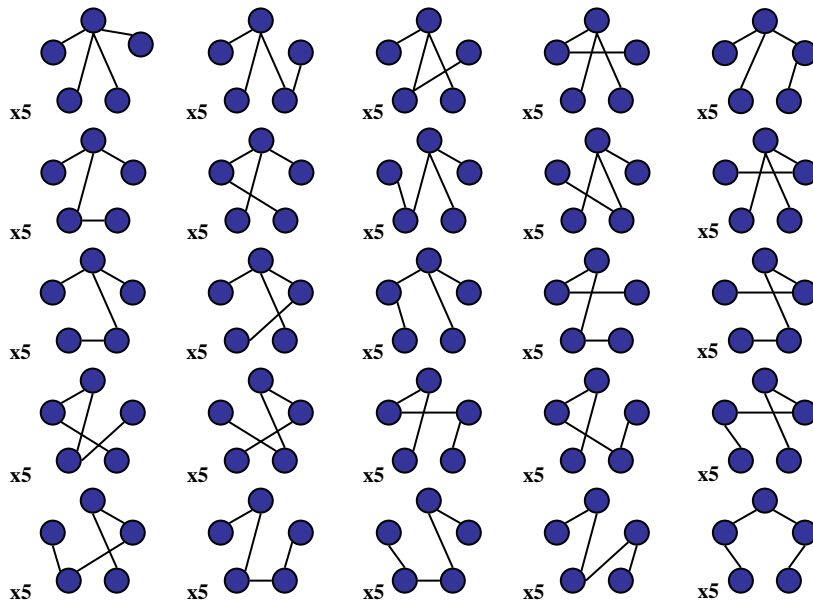


White



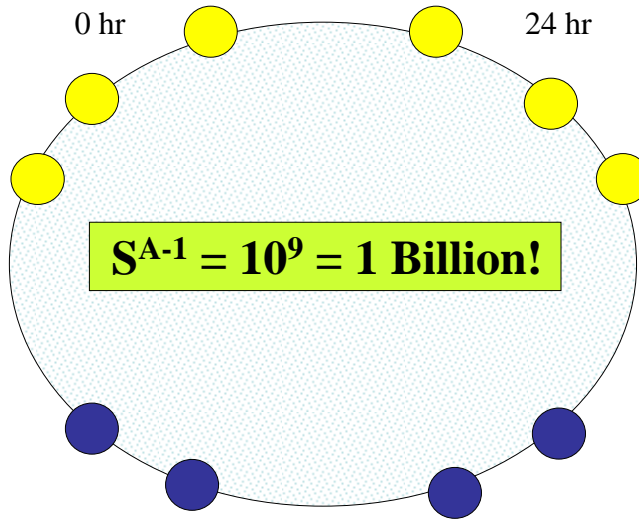
Recessive

$S^{A-1} = 5^3 = 125$



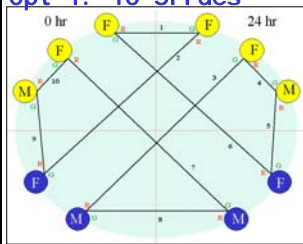


N of Configurations?

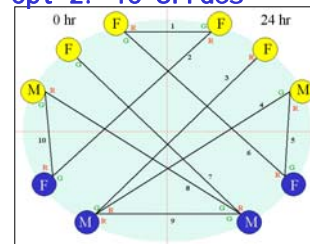


Transitivity (Townsend, 2003) & Extendability (Kerr, 2003)

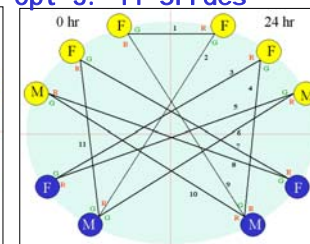
Opt 1: 10 Slides



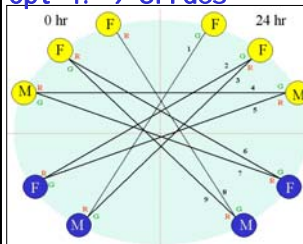
Opt 2: 10 Slides



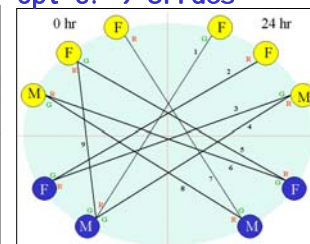
Opt 3: 11 Slides



Opt 4: 9 Slides



Opt 5: 9 Slides



Design Comparison Relative to Option 3					
Contrast	Opt 1 (10)	Opt 2 (10)	Opt 3 (11)	Opt 4 (9)	Opt 5 (9)
Immuno	60	80	100	80	90
Time	60	80	90	90	80
Sex	60	40	60	60	50
Immuno*Sex	0	0	20	0	10



MICROARRAY EXPERIMENTAL DESIGN

Take home message I:
“Identify the effects of interest *a priori*”

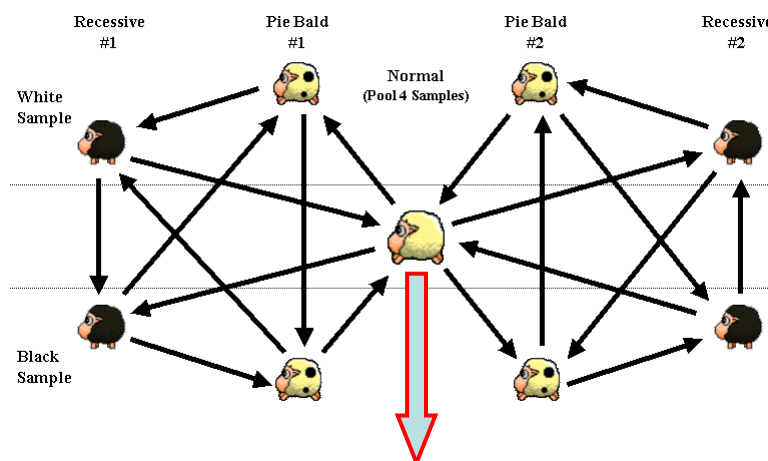
In addition to experimental constraints, design decisions should be guided by the knowledge of which effects are of greater interest to the investigator.

E.g. which main effects, which interactions.

The experimenter should thus decide on the comparisons for which he wants the most precision and these should be made **within slides** to the extent possible.



Wool Pigmentation



Is this pooled sample better than 4 individual "normal" samples?



Handling Constraints (Samples & Arrays):

Pooling & Replication

Pavlidis et al. (2003) The effect of replication on gene Expression microarray experiments. Bioinformatics 19: 1620

>= 5 Replicates
10-15 Replicates

Peng et al. (2003) Statistical implications of pooling RNA Samples for microarray experiments. BMC Bioinformatics 4: 26

Power: n9c9 ≈ 95%, n3c3 ≈ 50%, n9c3 ≈ 90%
n25c5 ≈ n20c20

Kendziorski et al. (2005) On the utility of biological samples in microarray experiments. PNAS 102: 4252.



Handling Constraints (Samples & Arrays):

Pooling & Replication

Peng et al. (2003) Statistical implications of pooling RNA Samples for microarray experiments. BMC Bioinformatics 4: 26

Power: n9c9 ≈ 95%, n3c3 ≈ 50%, n9c3 ≈ 90%
n25c5 ≈ n20c20

Table 1: Agreement of significant genes between "virtual" pooling and no pooling with data from one real experiment. Note: total number of "genes" on the chip = 8799, alpha = 0.05. Pool size=number of subjects per chip (# subjects per group/ # chips per group)



Take home message II:

“In the cases where we do not have enough material from one biological sample to perform one array (chip) hybridizations, Pooling or Amplification are necessary”

Pooling vs Individual Samples

Pooling is seen as “Biological Averaging”.
Trade off between: Cost of performing a hybridization
Cost of the mRNA samples.

IF Cost of mRNA samples << cost per hybridization
THEN Pooling can assist reducing the number of hybridization.

Pooling vs Amplified Samples

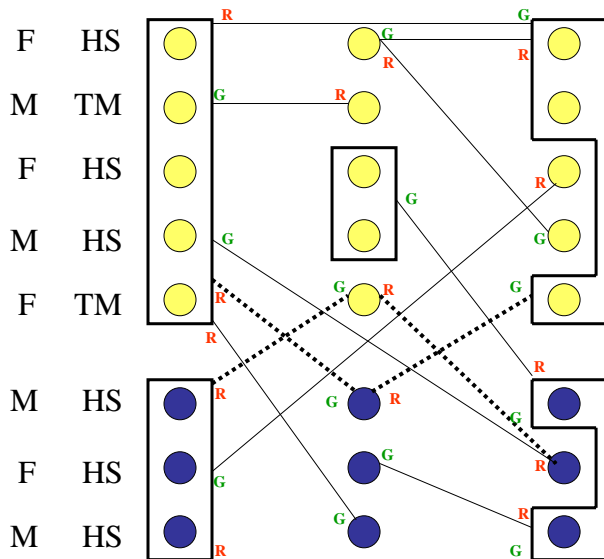
Amplification → Introduces more noise.
Non-linear amplification (??), ≠ genes amplified at ≠ rate.
Able to perform more hybridizations.

Pooling → Less replicates hybridizations.



Pooling & Replication

N of Arrays?
~~24~~: 23 To 552
↓ pooling
~~14~~: 13 To 182





Pooling & Replication

	RES	SUS	0	3	24	M	F	HS	TM
RES	8	-8	1	0	-1	-1.766	1.766	-3.866	3.866
SUS		8	-1	0	1	1.766	-1.766	3.866	-3.866
0			8	-4	-4	-1.335	1.335	0.666	-0.666
3				10	-6	-1.033	1.033	-0.468	0.468
24					10	2.368	-2.368	-0.198	0.198
M						6.247	-6.247	0.493	-0.493
F							6.247	-0.493	0.493
HS								3.798	-3.798
TM									3.798
Sum(ABS)	29.3	29.3	22.0	23.0	27.1	21.7	21.7	17.6	17.6

Reference Design

Sum(ABS)	26.8	26.8	39.1	23.1	17.3	7.1	7.1	14.3	14.3
----------	------	------	------	------	------	-----	-----	------	------



Another (NEW?) Constraint:

Amount of RNA

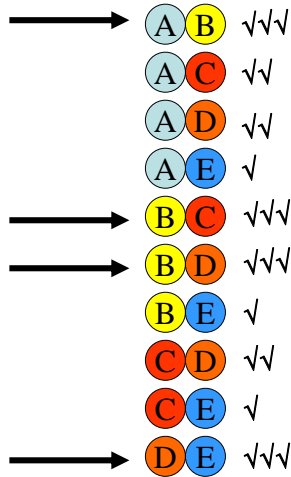
- A** M avium slope 18 days 3 3-3-3
- B** M avium broth 18 days 10 1-2-2-1-2-1-2-1-2-1
- C** M para broth 10 weeks 5 1-2-2-1-1
- D** M para broth 12 weeks 6 1-1-4-5-2-1
- E** M para in-vivo 3 1-1-1

Not interested in Amplifying



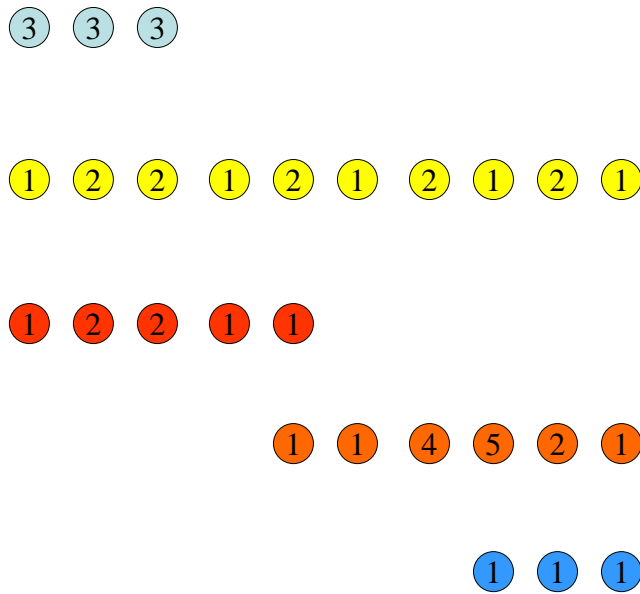
Another (NEW?) Constraint:

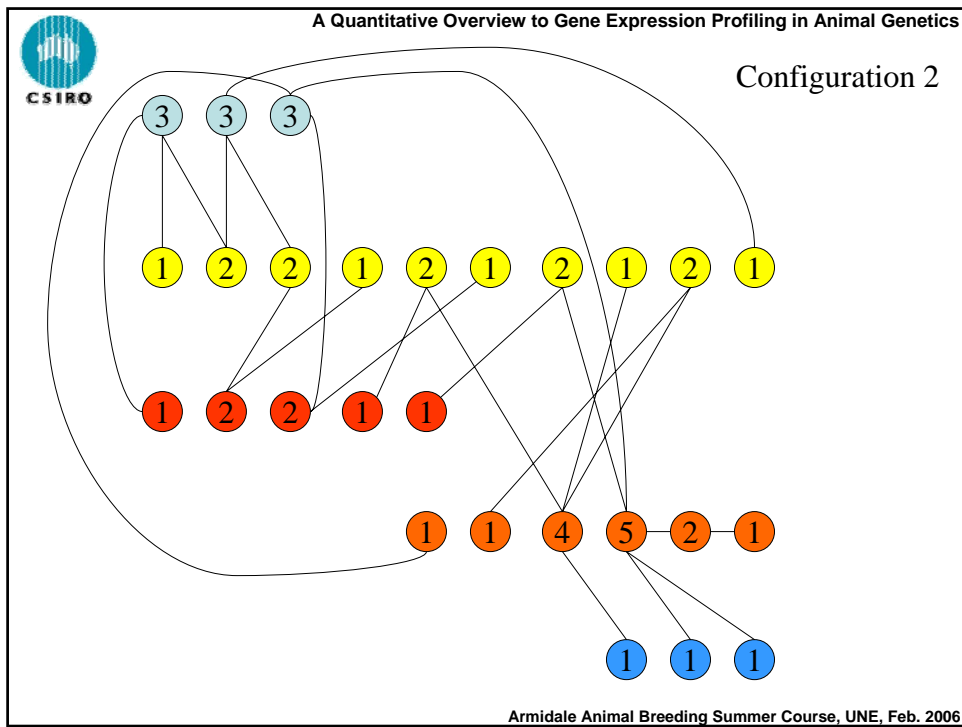
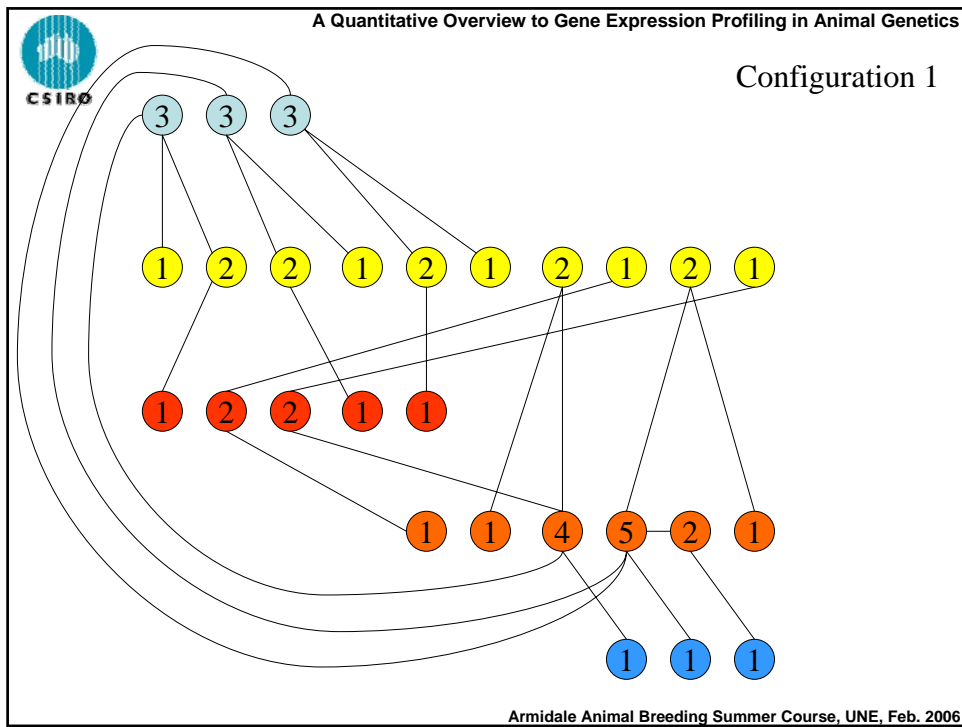
Amount of RNA

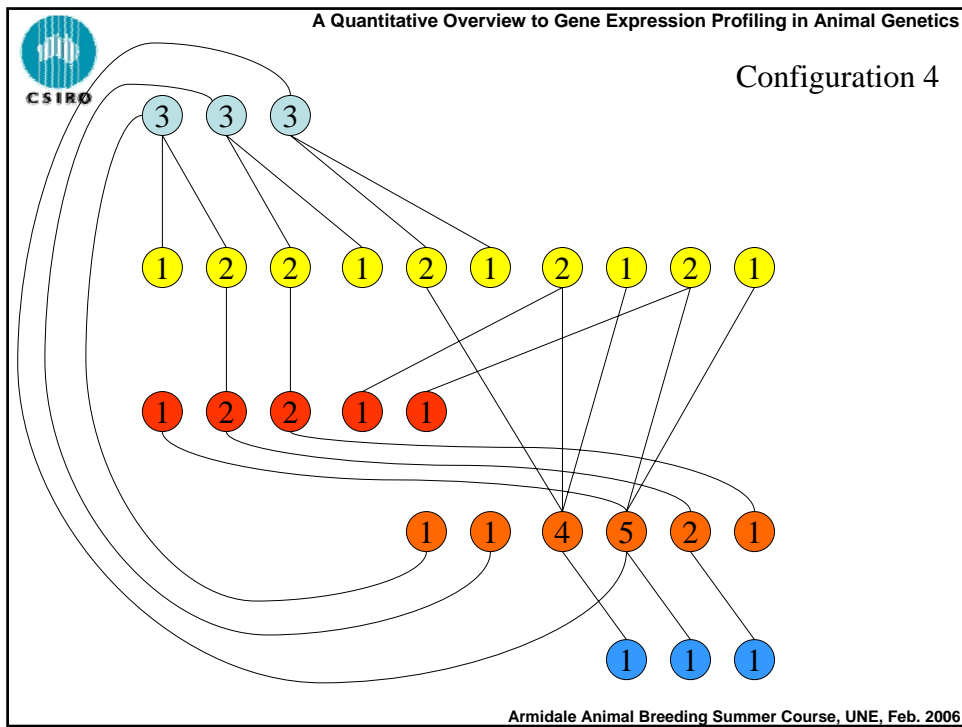
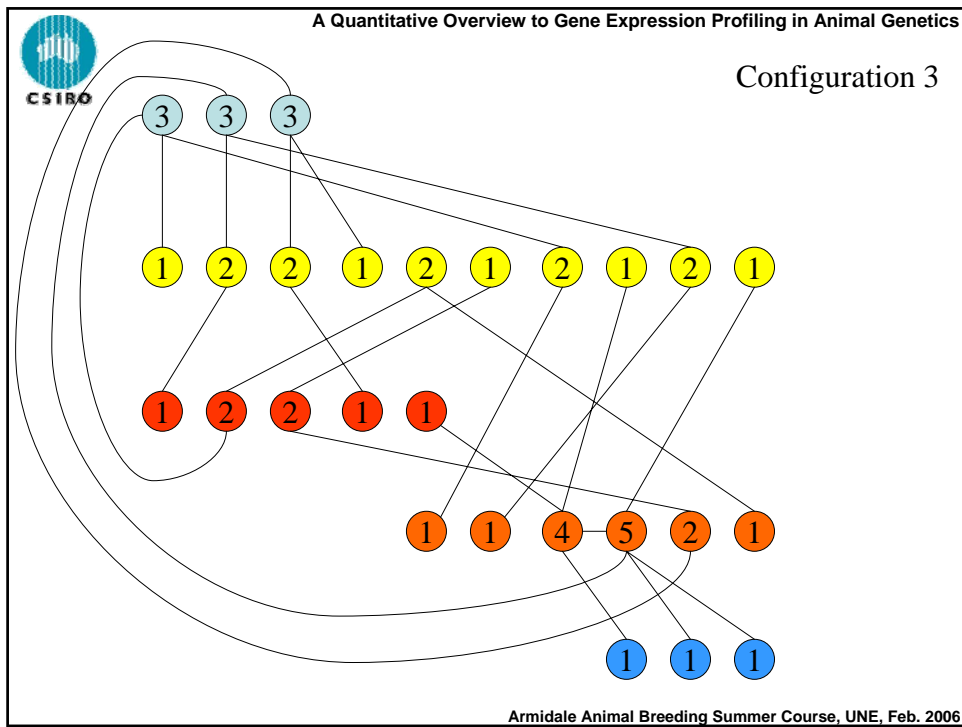


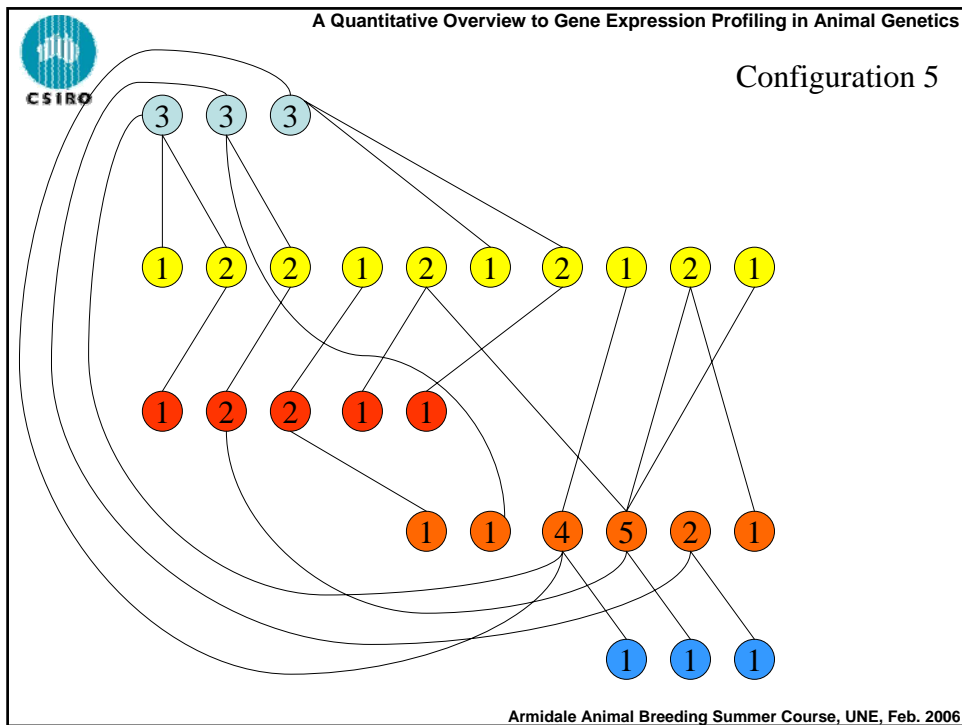
Importance due to Transitivity of AB with BC and BD

Procedure:
Five configurations will be proposed and the statistical optimality of each evaluated.









A Quantitative Overview to Gene Expression Profiling in Animal Genetics

	Imp	Weight					Squared Error					
		1	2	3	4	5	1	2	3	4	5	
(A)(B)	4	6	5	6	6	5	4	1	4	4	1	
(A)(C)	2	0	2	1	0	0	4	0	1	4	4	
(A)(D)	2	3	2	2	3	4	1	0	0	1	4	
(A)(E)	1	0	0	0	0	0	1	1	1	1	1	
(B)(C)	3	5	5	4	4	5	4	4	1	1	4	
(B)(D)	4	4	5	5	5	5	0	1	1	1	1	
(B)(E)	1	0	0	0	0	0	1	1	1	1	1	
(C)(D)	2	2	0	2	3	2	0	4	0	1	0	
(C)(E)	1	0	0	0	0	0	1	1	1	1	1	
(D)(E)	4	3	3	3	3	3	1	1	1	1	1	
Noise							SSE	17	14	11	16	18
(D)(D)	0	1	2	1	0	0	MSE	.74	.64	.48	.66	.75

Conclusion: Configuration 3

Armidale Animal Breeding Summer Course, UNE, Feb. 2006

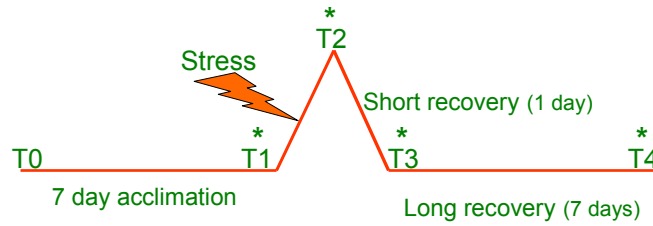


ONE LAST EXAMPLE

(E de la Vega, K Wilson, AIMS, Townsville)



1. Osmotic stress (35 to 10 ppt. stress for 8 hours)
2. Hypoxic stress (1ppm. DO / 8 hours)
3. Thermal stress (35.5 C / 24 hours)
4. Controls (kept at 35ppt, 28 C, >6 ppm. DO)



* Sampled 9 shrimp/treatment for gene expression analysis

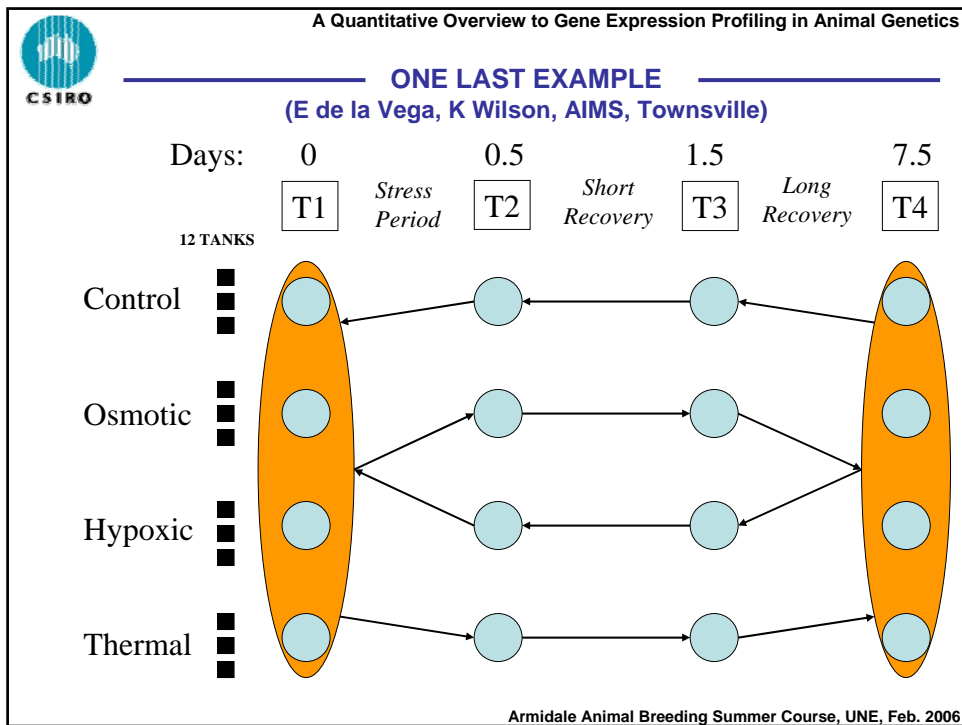
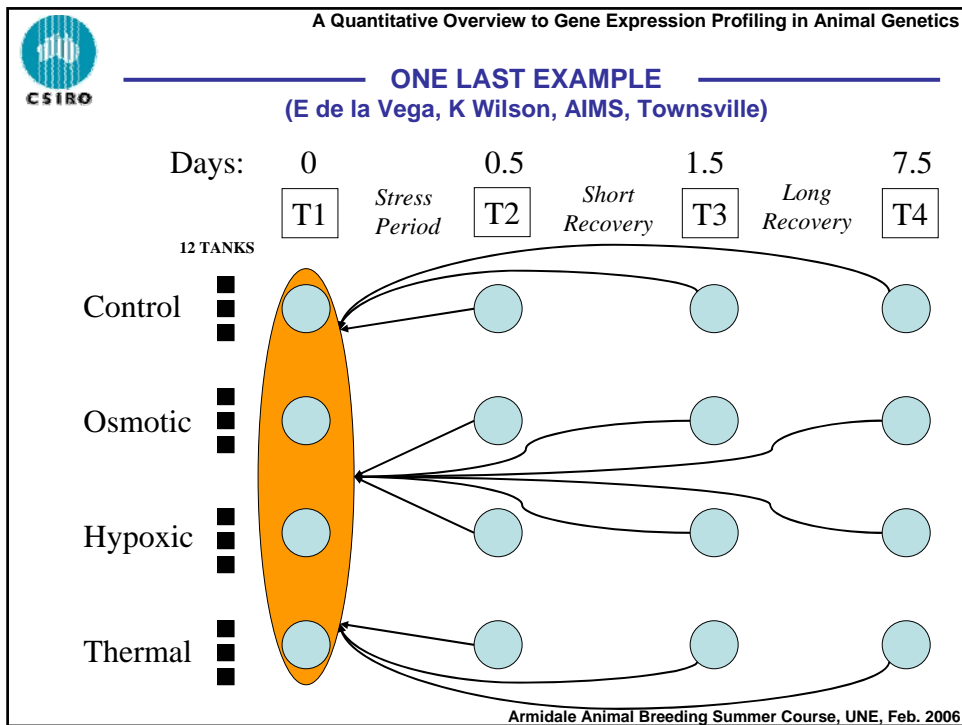
Max. 24 Hybridisations!



ONE LAST EXAMPLE

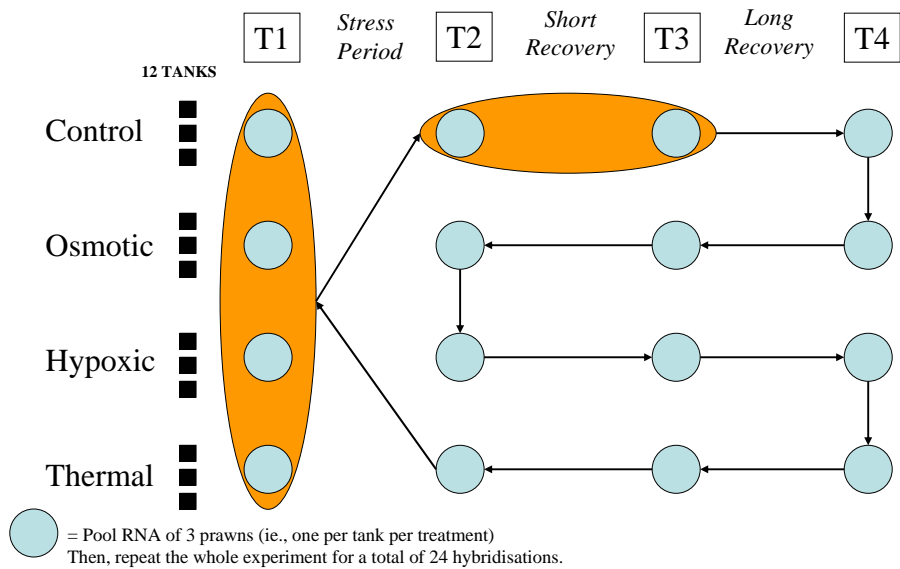
(E de la Vega, K Wilson, AIMS, Townsville)

Days:	0	0.5	1.5	7.5
	T1	T2	T3	T4
	Stress Period		Short Recovery	Long Recovery
12 TANKS				
Control	●	●	●	●
Osmotic	●	●	●	●
Hypoxic	●	●	●	●
Thermal	●	●	●	●





ONE LAST EXAMPLE (E de la Vega, K Wilson, AIMS, Townsville)



Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

Take home message III:
"Graphical representation tells the history"

- 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 (or samples).
 - The precision of the estimated contrast 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.
- Pooling issues can be immediately spotted
 - **Equal amounts** of RNA samples in a pool are essential
 - Samples intervene in a pool **once only** → Avoid messy analysis

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Break here



MICROARRAY EXPERIMENTAL DESIGN

The \$64M Question: How many animals?

As many as possible → The more replicates, the better your estimate of expression (that's an asymptotic process, so if you add at least a few replicates, the effect will be really strong).

Five → Experience shows that for most experiments you get a reasonable number of differentially expressed genes with 5 replicates.

Three → One to convince yourself, one to convince your boss, and one just in case (T. Speed?).

- It Depends On:**
1. the **Quality** of the sample
 2. the **Magnitude** of the expected effect
 3. the experimental **Design**
 4. the **Method** of analysis.

Reference: Flexible
Dye-Swap: Efficient
Loop: Elegant

Parametric more sensitive
than Non-Parametric.

Cell cultures less
noisy than biopsies

Never compare
dogs and donuts
(cf. Knock outs)

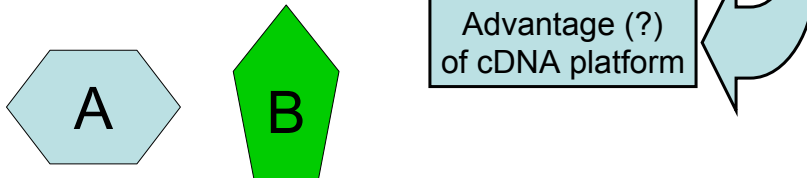


MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

1. Construction of subtracted libraries
2. Microarray hybridisations
3. Validation

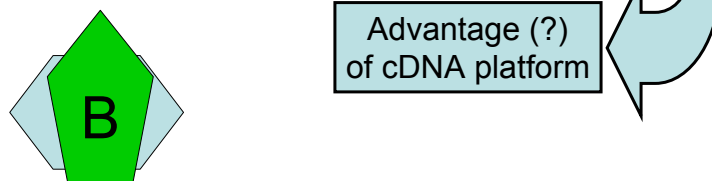


MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

1. Construction of subtracted libraries
2. Microarray hybridisations
3. Validation



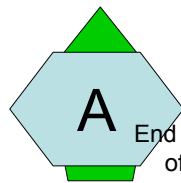


MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

1. Construction of subtracted libraries
2. Microarray hybridisations
3. Validation



End up with a "library" of ESTs (genes) enriched for a condition of interest. These will be printed on your microarray slide.

Advantage (?)
of cDNA platform



MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

✚ Pavlidis et al. (2003) The effect of replication on gene Expression microarray experiments. *Bioinformatics* 19: 1620

>= 5 Replicates
10-15 Replicates

1. Some experiments are still performed with little or none biological replication
2. Nevertheless, they still generate useful results → Big differences are likely to be real
3. They should be treated as **PILOT STUDIES**



MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

Advantages of PILOT STUDIES

- Estimate experimental variability
- Refine laboratory methods/techniques
- Refine experimental design
- Allows for rapid screening
- Provides preliminary data for project funding



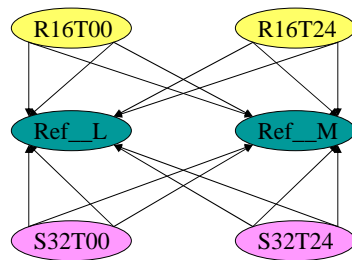
MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

Pilot Studies & Subtracted Libraries

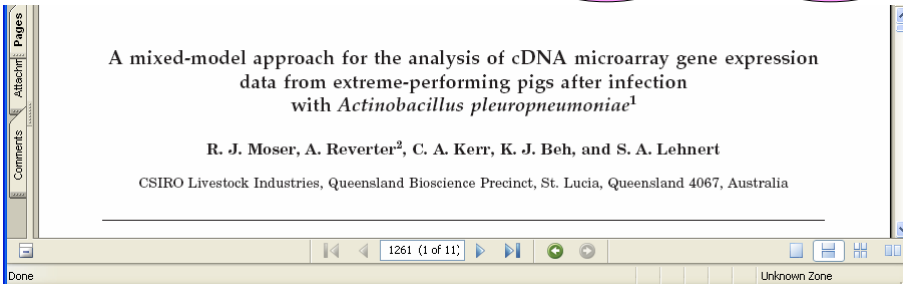
(J Anim Sci, 2004, 82:1261-1271)



A mixed-model approach for the analysis of cDNA microarray gene expression data from extreme-performing pigs after infection with *Actinobacillus pleuropneumoniae*¹

R. J. Moser, A. Reverter², C. A. Kerr, K. J. Beh, and S. A. Lehnert

CSIRO Livestock Industries, Queensland Bioscience Precinct, St. Lucia, Queensland 4067, Australia





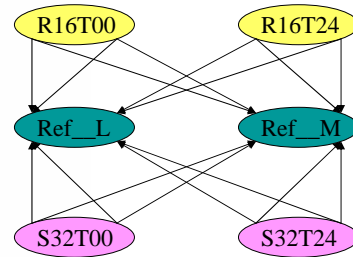
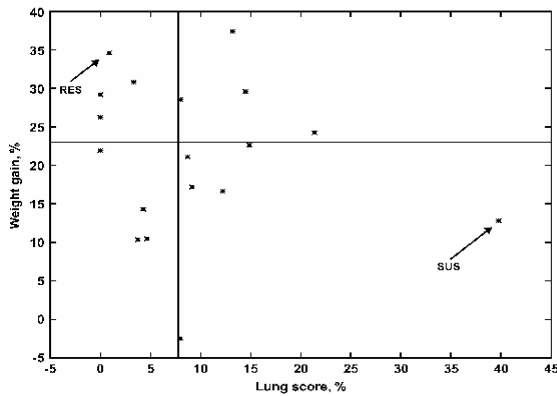
MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

• Pilot Studies & Subtracted Libraries

(J Anim Sci, 2004, 82:1261-1271)



Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

• Pilot Studies & Subtracted Libraries

(J Anim Sci, 2004, 82:1261-1271)

Table 5. Origin of 307 differentially expressed (DEXP) elements and proportion (%) compared with size of respective subtracted cDNA library of porcine descent

cDNA library	cDNA subtraction direction ^a	Library size	DEXP elements (profile) ^b		%
			Rup-Sdown	Rdown-Sup	
SPL	<u>Spleen vs. liver tissue</u>	576	6	—	1.04
LIVa	<u>Liver vs. spleen tissue</u>	480	2	—	0.4
LN	<u>Lymph node vs. liver tissue</u>	672	16	109	18.6
PLC-Fc	<u>Activated^c vs. non-activated lymphocytes</u>	768	5	—	0.7
PLC-R	<u>Nonactivated vs. activated lymphocytes</u>	768	5	—	0.7
PLMU-F	<u>Activated lymphocytes vs. muscle tissue</u>	576	27	4	5.4
PLMU-R	<u>Muscle tissue vs. activated lymphocytes</u>	576	23	2	4.3
WBC-F	<u>Blood leukocytes after^e vs. blood leukocytes before challenge</u>	714	91	13	14.6
WBC-R	<u>Blood leukocytes before vs. blood leukocytes after challenge</u>	714	4	—	0.6

^acDNA pool being selected for in the subtraction process is underlined.
^bRup-Sdown = up-regulated in the resistant and down-regulated in the susceptible animal; Rdown-Sup = down-regulated in the resistant and up-regulated in the susceptible animal.
^cF = forward; R = reverse.
^dTotal RNA and subsequently mRNA was extracted from proliferated lymphocytes after 24-h activation with the lectin concanavalin A (conA).
^eTotal RNA (mRNA) was extracted from white blood cells (buffy coat) from 10 pigs before and 24 h after being challenged with *Actinobacillus pleuropneumoniae*.

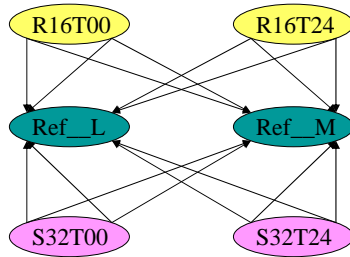
Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

- From Pilot to Final



Pigs Pleuropneumonia Pilot:

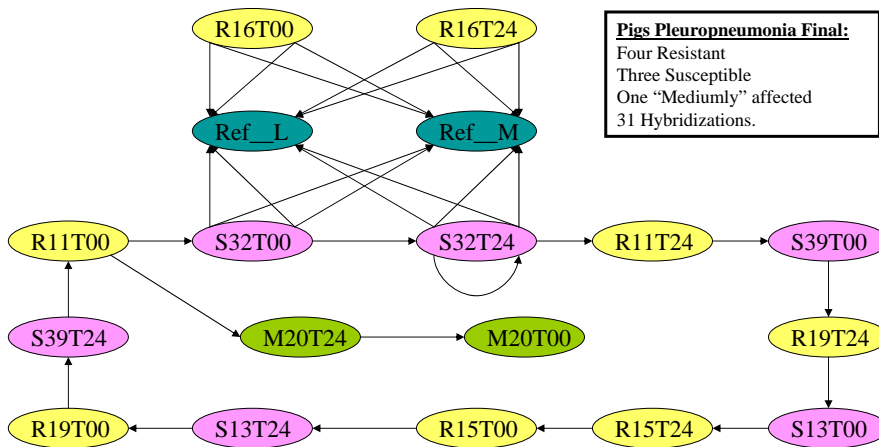
One Resistant
One Susceptible
16 Hybridizations.



MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

- From Pilot to Final



Pigs Pleuropneumonia Final:

Four Resistant
Three Susceptible
One "Mediumly" affected
31 Hybridizations.



MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

Fleece Rot Resistance

1. Construction of subtracted libraries
2. Microarray hybridisations
3. Validation (eg. RT-PCR)



Different animals across these three stages to avoid bias due to sampling



- Two existing lines: Resistant (RES) and Susceptible (SUS)
- Animals to be put through a “wetting trial” in order to obtain a visual assessment of their susceptibility



MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals?

(Fleece Rot Resistance)

Conditions

1. Subtracted libraries

The most extreme animals within each line, RES and SUS (min. = 2) to ensure enrichment in the subtraction. NB: Using one RES and one SUS assumes monotonicity within condition.

2. Microarrays

Could use animals from the entire spectrum, but preferably extremes (also within line, RES and SUS). NB: Need biological replicates. Lee et al 2002 (PNAS, 97:9834-9839) recommends a minimum of 3.

3. Validation

Animals at random from the entire population of RES or SUS. ie. An average sort of RES or an average sort of SUS.

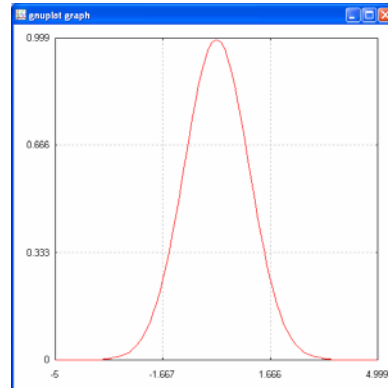
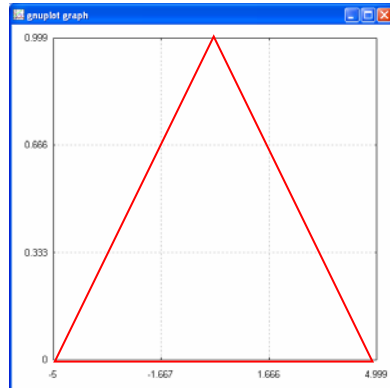


MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals? (Fleece Rot Resistance)

Assuming the distribution of "resistance" is symmetric (not necessarily normal), uni-modal and more leptokurtic than a triangle, then the middle third contains ≥ 3 times as many observations as either extreme third.



Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

...Designing from scratch

How many animals? (Fleece Rot Resistance)

Conclusion: we require 20 animals within immunological categories (RES and SUS) and with the following allocation:

	N of Animals According to Use/Destination		
	To develop the focussed microarray	To perform the microarray experiment	To run the qRT-PCR on candidate genes
Most Extreme	1	3	0
Average	0	9	3
Least Extreme	1	3	0

Armidale Animal Breeding Summer Course, UNE, Feb. 2006

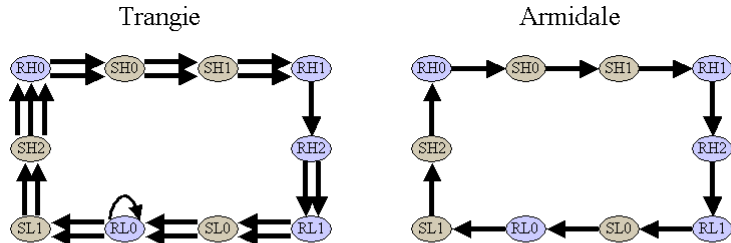


MICROARRAY EXPERIMENTAL DESIGN

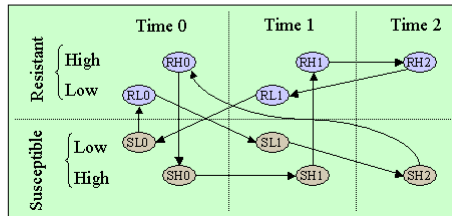
...Designing from scratch

How many animals?

(Fleece Rot Resistance)



LEGEND: Arrows indicate hybridisations and go from the sample labelled with red to the sample labelled with green dye. Clicking of the body of the arrow pups up the 2-page pdf file with summary info for this hybridisation. Clicking of the head pups up the entire raw data. The ppt with full design details is [HERE](#).



Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY EXPERIMENTAL DESIGN

The \$64M Question: How many animals?

$$n = \frac{4(z_{1-\alpha/2} + z_{1-\beta})^2}{(\delta / \sigma)^2}$$

Simon et al., 2002.
Genetic Epidemiology 23: 21-36

Where $z_{\alpha/2}$ and z_{β} are normal percentile values at

false positive rate $\alpha \rightarrow$ Type I error rate

false negative rate $\beta \rightarrow$ Type II error rate,

$1 - \beta \rightarrow$ power to detect differences (Prob. of detecting TP)

$\delta =$ minimum detectable \log_2 ratio;

and $\sigma =$ SD of log ratio values.

NB: Reference Designs Only

Example:

For $\alpha = 0.001$ and $\beta = 0.05$, get $z_{\alpha/2} = -3.29$ and $z_{\beta} = -1.65$.

Assume $\delta = 1.0$ (2-fold change) and $\sigma = 0.25$,

$\rightarrow n = 12$ samples (6 query and 6 control) \leftarrow

Armidale Animal Breeding Summer Course, UNE, Feb. 2006