



INTRODUCTION

Truths, Damn Truths & Statistics in Gene Expression Studies

Disclaimer:

The Biologist and the Statistician are being executed.
They are both granted one last request.

The Statistician asks that he/she be allowed to give one
final lecture on his/her Grand Theory of Statistics.

The Biologist asks that he/she be executed first.

Armidale Animal Breeding Summer Course, UNE, Feb. 2006

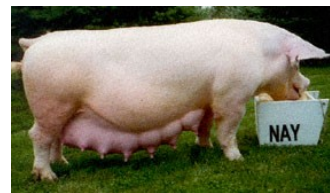


THE RAW MATERIAL

This is a Cow



This is a Pig
(female)



This is a Sheep



This is a Chicken



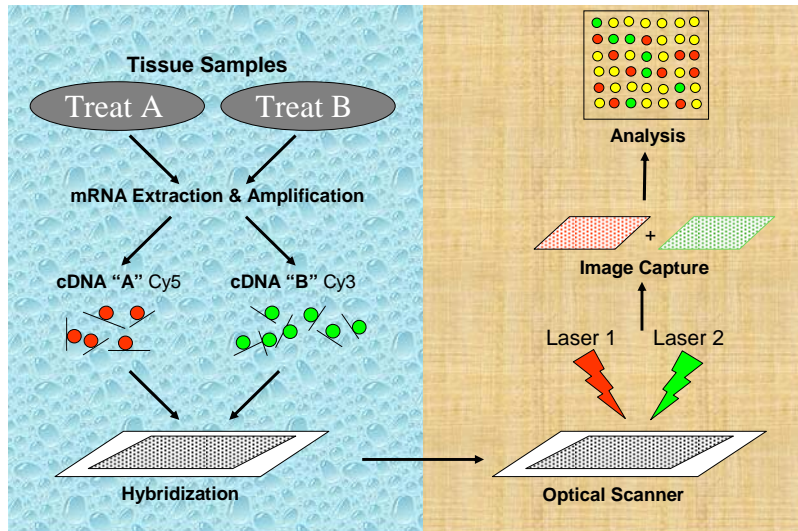
This is a Prawn



Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY PROCESS

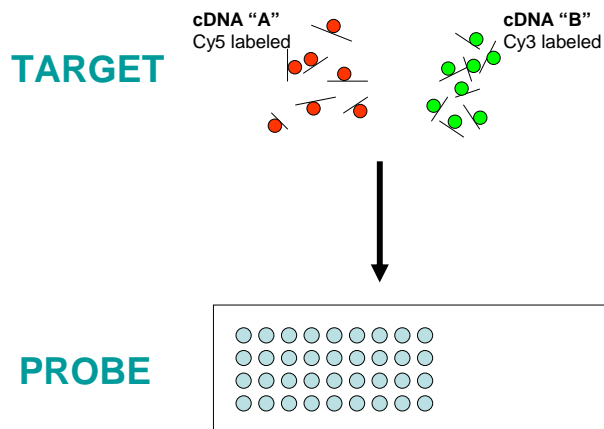


Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY PROCESS

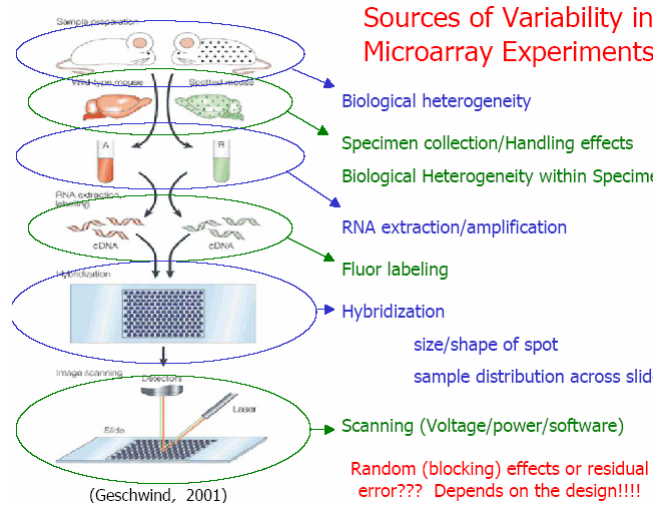
Definition of PROBE and TARGET



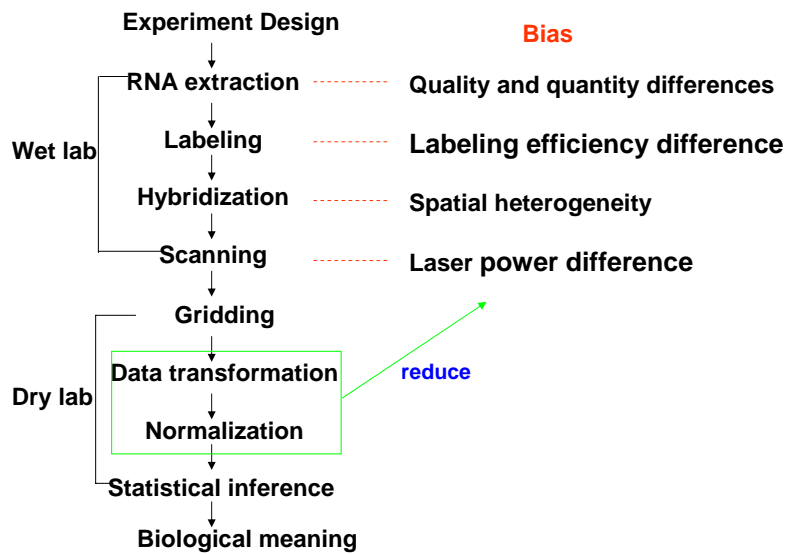
Armidale Animal Breeding Summer Course, UNE, Feb. 2006



MICROARRAY "WET" PROCESS



MICROARRAY "WET & DRY" PROCESSES





MICROARRAY OUTPUT



What you see is what you get ...truly!

- 3 Spatial Features 1. Printing block
- 2. Row
- 3. Column
- 2 Channels: - Red: - Foreground ... 4. Mean
- 5. Median
- 6. Std Dev
- Background ... 7. Mean
- 8. Median
- 9. Std Dev
- Green: - Foreground ... 10. Mean
- 11. Median
- 12. Std Dev
- Background ... 13. Mean
- 14. Median
- 15. Std Dev



MICROARRAY APPLICATIONS

- Determine genes which are differentially expressed (DE).
- Connect DE genes to sequence databases to search for common upstream regions.
- Overlay DE genes on pathway diagrams.
- Relate expression levels to other information on cells, e.g. tumor types.
- Identify temporal and spatial trends in gene expression.
- Seek roles of genes based on patterns of co-regulation.

- **...Applications to Selective Breeding Schemes?**



MICROARRAY PROCESS

CLAIM

“The majority of microarray papers are analysed with substandard methods”

C Tilstone (citing D Allison), Nature 2003, 424:610

REASONS

P Value

- 1. Biologists don't care 10
- 2. Statisticians are bad 20
- 3. Unrealistic expectations 70



CHALLENGES

Time Dependent

Data Dependent

Human Dependent

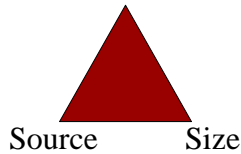
Chronology

Paradigm

Skill Integration

Logical	cDNA
1800s – DATA	√
30-60s – METHODS	×
50-70s – SOFTWARE	×
1980s – COMPUTER	√

Distribution



Quantitative

Computer Sci.
Statisticians
Mathematicians

Non-Q

Biochemists
Physiologists
Pathologists

EGG ↔ BANANA

“banana omelette”

Historical Excitement Balance Interdisciplinary



Challenges

→ Human Dependent

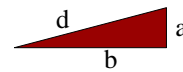
→ ~~Historical~~ *Hysterical*

• **Traditionally:** Statistics grew alongside Agriculture

“Introduction to Statistical Analysis”

- Law of Large Numbers
- Central Limit Theorem
- Pythagoras Theorem

$$SST = SSM + SSE$$



• **Nowadays:** Statistics alongside (Bio)Technology



Challenges

→ Human Dependent

→ Excitement (source of)

Eg. Always log spot intensities and ratios

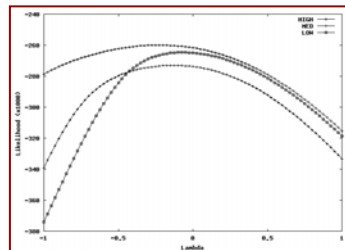
T Speed. “Hints and Prejudices”

• **Biochemist:** My software does it, therefore it’s great!

• **Statistician:** Well, I need further evidence to be convinced

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(x) & \lambda = 0 \end{cases}$$

$$l(\lambda) = -\frac{1}{2} \ln \left[\frac{1}{n} \sum_{j=1}^n (x_j^{(\lambda)} - \overline{x^{(\lambda)}})^2 \right] + (\lambda - 1) \sum_{j=1}^n \ln(x_j)$$





Challenges

→ Human Dependent

→ Balance

•Too many Statisticians:

Evidence: It takes 1 ship, 10 days to cross the ocean

Question: How many days does it take for 10 ships to cross the ocean?

Evidence: It takes 1 builder, 10 days to build a wall

Question: How many days does it take for 10 builders to build a wall?



Challenges

→ Human Dependent

→ Balance

•Too many Statisticians:

PHD SCHOLARSHIP
Statistical Science Program
MATHEMATICAL SCIENCES INSTITUTE
THE AUSTRALIAN NATIONAL UNIVERSITY

Stipend \$22,771 (2002 rate, indexed annually, tax free)

A PhD Scholarship (APAI) is being offered by the Mathematical Sciences Institute at The ANU. An ARC Linkage Grant held by Professors Peter Hall (ANU) and Don Poskitt (Monash University), in conjunction with BAE Systems, Melbourne, will fund the scholarship.

The research problem is in the area of stochastic control applied to **ship motion**, and involves the development and implementation of both parametric and nonparametric methods. The successful applicant will have a strong interest in statistical methodology, computational techniques, theoretical analysis, and the development of statistical research problems.



Challenges

Human Dependent

Balance

•Too many Biochemists:

		Treated?	
		No	Yes
Died?	No	100	150
	Yes	120	120

Survival Rates:
 Treated = $150/270 = 55.55\%$
 Non-Tr = $100/220 = 45.45\%$ **22% Increase!**

Women?

		No	Yes
		60	30
Died?	No	60	30
	Yes	100	60

Survival Rates:
 Treated = $30/90 = 33.33\%$
 Non-Tr = $60/160 = 37.50\%$ **12.5% Decrease!**

Men?

		No	Yes
		40	120
Died?	No	40	120
	Yes	20	60

Survival Rates:
 Treated = $120/180 = 66.66\%$
 Non-Tr = $40/60 = 66.66\%$ **No Difference!**

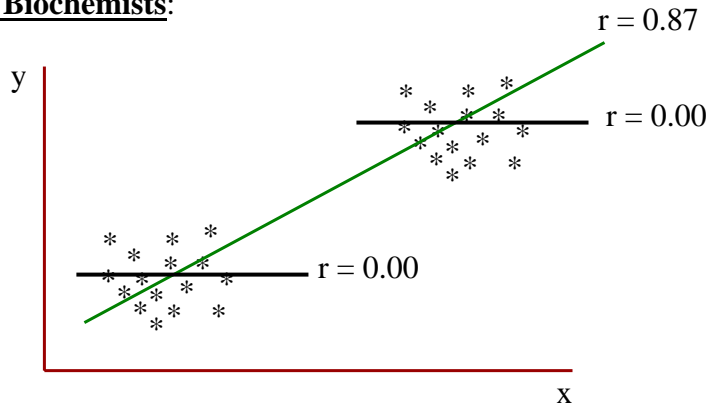


Challenges

Human Dependent

Balance

•Too many Biochemists:





Challenges

Human Dependent

Interdisciplinary Skills



Minimal knowledge of the application discipline is needed

However, simply keeping track of the literature is an enormous task.

- 1. Animal Science
- 2. Journal of Animal Science
- 3. Journal of Dairy Science
- 4. Genetic Selection Evolution
- 5. Livestock Production Science
- 6.a couple or so more.

- 1. JASA
- 2. JABES
- 3. Biostatistics

- 1. NPG: Nature, Nature Genetics, Nature Review Genetics, Nature Biotechnology, Nature Methods, Nature Computational Biology.
- 2. BMC: Bioinformatics, Genomics.
- 3. Science
- 4. Genome Research
- 5. Bioinformatics
- 6. PNAS
- 7. Mammalian Genome
- 8. Nucleic Acids Reseach
- 9.and many many more.



Challenges

Human Dependent

Interdisciplinary Skills



Minimal knowledge of the application discipline is needed

Failing that, the Statisticians will win,
.....but with the wrong weapons.

- 1. Amount of Expression = Amount of Response
- 2. Same cut-off point to judge all genes
- 3. Over-emphasis in normalization (Thus, reject "Boutique Arrays")
- 4. Over-emphasis in variance stabilization



Information: Amount vs Quality



Probability (certainty) of both Female?

- Case 1. No Information1/4
- Case 2. The one on the left is female1/2
- Case 3. One of them is female1/3