



# Co-Expression

Validation of alternative methods of data normalization in gene co-expression studies

Reverter et al. 2005  
Bioinformatics 21:1112



## Co-Expression

Motivation

**Statistics are like a bikini,  
what they reveal is suggestive,  
but what they conceal is vital.**

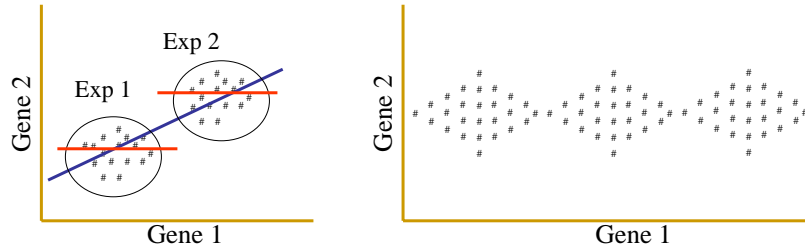
Aaron Levenstein



Co-Expression

Motivation

Correlations Are Dangerous!



$$SE(r) = \frac{\sqrt{1-r^2}}{\sqrt{n-2}}$$

$$P(r) = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} (1-r^2)^{(\nu-2)/2}$$



Co-Expression

Introduction

Validation of alternative methods of data normalization in gene co-expression studies. *Bioinformatics* 2005, 21:1112

A. Reverter, W. Barris, S.M. McWilliam, K.A. Byrne, Y.H. Wang, S.H. Tan, N. Hudson, and B.P. Dalrymple

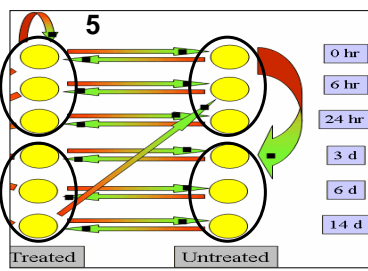
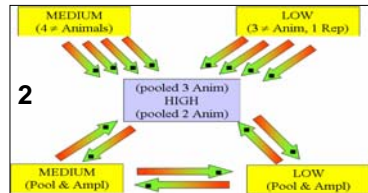
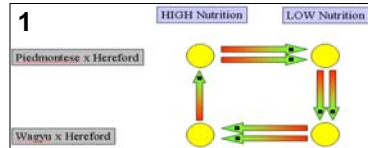
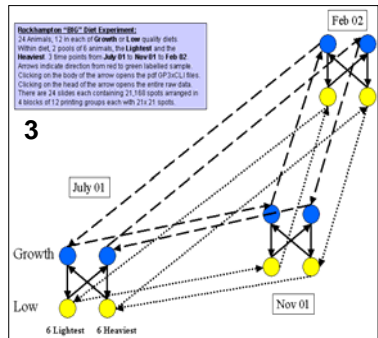
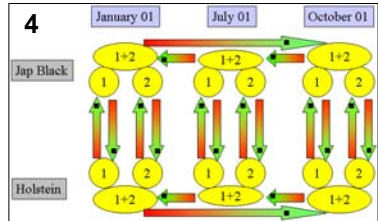
Experiment <sup>a</sup>	Hybs.	Cond.	Signals <sup>b</sup>		
			N	Mean	STD
1. Two breeds by two diets	7	4	193,175	7.61	2.99
2. Three diets	14	3	361,320	8.52	2.84
3. Two diets at three ages	24	6	801,807	11.60	1.89
4. Two breeds at three ages	18	6	459,978	8.34	3.28
5. Two fat treatments at two ages	15	4	418,817	7.04	3.31

Expression of each clone (gene) across 23 conditions



Co-Expression

Introduction



Armidaile Animal Breeding Summer Course, UNE, Feb. 2006



Co-Expression

Methods

Expression of each clone (gene) across 23 conditions

Nine Methods of Data Normalization

	1. RMNA	2. RMCE	3. RMEC
Clone $i$			
Cond			
Experiment 1	$\begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_3 \\ \bar{X}_4 \\ \bar{X}_5 \end{pmatrix}$	$\begin{pmatrix} (\bar{X}_1 - \mu_{E1}) / \sigma_{E1} \\ (\bar{X}_2 - \mu_{E1}) / \sigma_{E1} \\ (\bar{X}_3 - \mu_{E1}) / \sigma_{E1} \\ (\bar{X}_4 - \mu_{E2}) / \sigma_{E2} \\ (\bar{X}_5 - \mu_{E2}) / \sigma_{E2} \end{pmatrix}$	$\begin{pmatrix} (\bar{X}_1 - \mu_{Ci}) / \sigma_{Ci} \\ (\bar{X}_2 - \mu_{Ci}) / \sigma_{Ci} \\ (\bar{X}_3 - \mu_{Ci}) / \sigma_{Ci} \\ (\bar{X}_4 - \mu_{Ci}) / \sigma_{Ci} \\ (\bar{X}_5 - \mu_{Ci}) / \sigma_{Ci} \end{pmatrix}$
Experiment 2			
Experiment 3			
Experiment 4			
Experiment 5			
	$(\mu_{Ci}, \sigma_{Ci})$		

Armidaile Animal Breeding Summer Course, UNE, Feb. 2006



Co-Expression

Methods

Expression of each clone (gene) across 23 conditions

Nine Methods of Data Normalization

				RMNA		RMCE		RMEC			
				Gene 1	Gene 2	Gene 1	Gene 2	Gene 1	Gene 2		
<b>Gene x Exp Means</b>											
			1	11	13	8.60	10.00	-0.42	0.37	-0.33	0.28
	Mean	STD	1	9	10						
			1	8	11						
Gene 1	9.42	1.93	1	8	7						
Gene 2	9.17	2.25	1	7	9						
			1	8	9	10.00	9.33	0.30	0.07	0.31	-0.06
			1	8	9						
			1	14	10						
			1	10	4	10.00	8.00	0.30	-0.52	0.31	-0.75
			1	11	11						
			1	9	9						
			1	10	8						
			2	11	9	9.00	10.00	-0.08	0.13	-0.15	0.28
			2	11	9						
Gene 1	9.20	2.53	2	8	11						
Gene 2	9.80	1.55	2	5	10						
			2	10	11						
			2	7	10	9.40	9.60	0.08	-0.13	0.04	0.07
			2	6	6						
			2	10	10						
			2	12	11						
			2	12	11						
			2	12	11						
			Mean	9.32	9.45						
			STD	2.17	1.95	-0.7864		-0.7644		-0.7864	

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Co-Expression

Methods

Expression of each clone (gene) across 23 conditions

Nine Methods of Data Normalization

				RMNA		RMCE		RMEC			
				Gene 1	Gene 2	Gene 1	Gene 2	Gene 1	Gene 2		
<b>Gene x Exp Means</b>											
			1	11	13	8.60	10.00	-0.42	0.37	-0.33	0.28
	Mean	STD	1	9	10						
			1	8	11						
Gene 1	9.42	1.93	1	8	7						
Gene 2	9.17	2.25	1	7	9						
			1	8	9	10.00	9.33	0.30	0.07	0.31	-0.06
			1	8	9						
			1	14	10						
			1	10	4	10.00	8.00	0.30	-0.52	0.31	-0.75
			1	11	11						
			1	9	9						
			1	10	8						
			2	11	9	9.00	10.00	-0.08	0.13	-0.15	0.28
			2	11	9						
Gene 1	9.20	2.53	2	8	11						
Gene 2	9.80	1.55	2	5	10						
			2	10	11						
			2	7	10	9.40	9.60	0.08	-0.13	0.04	0.07
			2	6	6						
			2	10	10						
			2	12	11						
			2	12	11						
			2	12	11						
			Mean	9.32	9.45						
			STD	2.17	1.95	-0.7864		-0.7644		-0.7864	

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Co-Expression

Methods

Expression of each clone (gene) across 23 conditions

Nine Methods of Data Normalization

				RMNA		RMCE		RMEC				
				Gene 1	Gene 2	Gene 1	Gene 2	Gene 1	Gene 2			
Gene x Exp Means				1	11	8.60	10.00	-0.42	0.37	-0.33	0.28	
Mean	STD	1	1	9	10							
		1	1	8	11							
Gene 1	9.42	1.93	1	1	8	7						
Gene 2	9.17	2.25	1	1	7	9						
			1	2	8	9						
			1	2	8	9						
			1	2	14	10						
			1	3	10	4	10.00	8.00	0.30	-0.52	0.31	-0.75
			1	3	11	11						
			1	3	9	9						
			1	3	10	8						
			2	4	11	9	9.00	10.00	-0.08	0.13	-0.15	0.28
			2	4	11	9						
Gene 1	9.20	2.53	2	4	8	11						
Gene 2	9.80	1.55	2	4	5	10						
			2	4	10	11						
			2	5	7	10						
			2	5	6	6						
			2	5	10	10						
			2	5	12	11						
			2	5	12	11						
				Mean	9.32	9.45						
				STD	2.17	1.95						
							-0.7864		-0.7644			-0.7864

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Co-Expression

Methods

Expression of each clone (gene) across 23 conditions

Nine Methods of Data Normalization

				RMNA		RMCE		RMEC				
				Gene 1	Gene 2	Gene 1	Gene 2	Gene 1	Gene 2			
Gene x Exp Means				1	11	8.60	10.00	-0.42	0.37	-0.33	0.28	
Mean	STD	1	1	9	10							
		1	1	8	11							
Gene 1	9.42	1.93	1	1	8	7						
Gene 2	9.17	2.25	1	1	7	9						
			1	2	8	9	10.00			0.31	-0.06	
			1	2	8	9						
			1	2	14	10						
			1	3	10	4	10.00	8.00	0.30	-0.52	0.31	-0.75
			1	3	11	11						
			1	3	9	9						
			1	3	10	8						
			2	4	11	9	9.00	10.00	-0.08	0.13	-0.15	0.28
			2	4	11	9						
Gene 1	9.20	2.53	2	4	8	11						
Gene 2	9.80	1.55	2	4	5	10						
			2	4	10	11						
			2	5	7	10						
			2	5	6	6	9.40	9.60				0.07
			2	5	10	10						
			2	5	12	11						
			2	5	12	11						
				Mean	9.32	9.45						
				STD	2.17	1.95						
							-0.7864		-0.7644			-0.7864

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Co-Expression

Methods

Expression of each clone (gene) across 23 conditions

Nine Methods of Data Normalization

				RMNA		RMCE		RMEC			
		Exper	Cond	Gene 1	Gene 2	Gene 1	Gene 2	Gene 1	Gene 2		
<b>Gene x Exp Means</b>											
	Mean	STD	1	11	13	8.60	10.00	-0.42	0.37	-0.33	0.28
			1	9	10						
			1	8	11						
Gene 1	9.42	1.93	1	8	7						
Gene 2	9.17	2.25	1	7	9						
			1	8	9	10.00	9.33	0.30	0.07	0.31	-0.06
			1	8	9						
			1	14	10						
			1	10	4	10.00	8.00	0.30	-0.52	0.31	-0.75
			1	11	11						
			1	9	9						
			1	10	8						
			2	11	9	9.00	10.00	-0.08	0.13	-0.15	0.28
			2	11	9						
Gene 1	9.20	2.53	2	8	11						
Gene 2	9.80	1.55	2	5	10						
			2	10	11						
			2	7	10	9.40	9.60	0.08	-0.13	0.04	0.07
			2	6	6						
			2	10	10						
			2	12	11						
			2	12	11						
			2	12	11						
			Mean	9.32	9.45						
			STD	2.17	1.95						

Expected to be the equal!

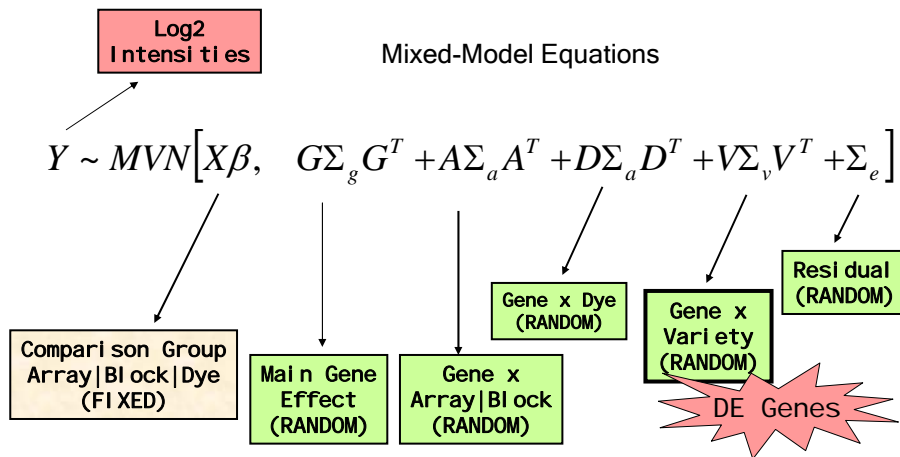
-0.7864      -0.7644      -0.7864



Co-Expression

Methods

Nine Methods of Data Normalization





Co-Expression

Methods

Nine Methods of Data Normalization

- 4. MM1NA
- 5. MM1CE
- 6. MM1EC

Based on Uni-variate Mixed-Models

$$y = X\beta + Z_1c + Z_2a + Z_3d + Z_4t + e$$

- 7. MM5NA
- 8. MM5CE
- 9. MM5EC

Based on Penta-variate Mixed-Models

$$y_E = X_E\beta_E + Z_{E1}c_E + Z_{E2}a_E + Z_{E3}d_E + Z_{E4}t_E + e_E$$



Co-Expression

Methods

Nine Methods of Data Normalization

- 4. MM1NA
- 5. MM1CE
- 6. MM1EC

Based on Uni-variate Mixed-Models

$$\hat{t}_c$$

$$\frac{\hat{t}_c - \mu_{c/E}}{\sigma_E}$$

$$\frac{\hat{t}_c - \mu_{E/c}}{\sigma_c}$$

Vector of BLUP Clone x Condition Interaction

- 7. MM5NA

Based on Penta-variate Mixed-Models

$$\hat{t}_{Ec}$$

$$\frac{\hat{t}_{Ec} - \mu_{c/E}}{\sigma_E}$$

- 9. MM5EC

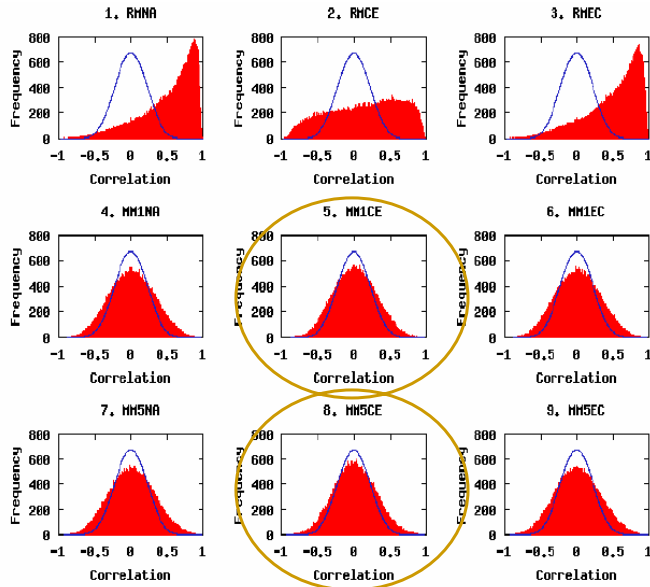
$$\frac{\hat{t}_{Ec} - \mu_{E/c}}{\sigma_c}$$



Co-Expression

Results

Observed (RED) vs Expected (BLUE)



Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Co-Expression

Results

Average across all pair-wise gene correlations (bold, italics and on the diagonal); Correlation of correlations (above diagonal); and maximum absolute discrepancy between methods (below diagonal)

	1	2	3	4	5	6	7	8	9
1. RMNA	<b>0.53</b>	0.20	0.98	0.11	0.12	0.11	0.06	0.07	0.06
2. RMCE	1.84	<b>0.15</b>	0.21	0.14	0.14	0.14	0.09	0.10	0.09
3. RMEC	0.76	1.79	<b>0.51</b>	0.12	0.12	0.12	0.07	0.07	0.07
4. MM1NA	1.78	1.72	1.74	<b>0.02</b>	0.94	0.99	0.57	0.60	0.56
5. MM1CE	1.75	1.75	1.74	0.51	<b>0.01</b>	0.94	0.61	0.69	0.61
6. MM1EC	1.79	1.71	1.74	0.50	0.55	<b>0.02</b>	0.57	0.61	0.58
7. MM5NA	1.80	1.75	1.79	1.25	1.14	1.21	<b>0.01</b>	0.91	0.99
8. MM5CE	1.78	1.74	1.77	1.19	1.09	1.19	0.59	<b>0.01</b>	0.90
9. MM5EC	1.80	1.75	1.79	1.25	1.13	1.20	0.55	0.66	<b>0.01</b>

Armidale Animal Breeding Summer Course, UNE, Feb. 2006





Co-Expression

Results

Redundant genes (ie. those with many probes on the array) allow for Validation

Table 4. Average correlation among all pair-wise clones for genes with more than x clones and by each normalization method

x	Genes	Normalization method								
		1	2	3	4	5	6	7	8	9
1	195	0.50	0.28	0.50	0.29	0.30	0.29	0.41	0.39	0.41
2	106	0.48	0.31	0.48	0.36	0.38	0.36	0.50	0.48	0.50
3	71	0.53	0.32	0.53	0.41	0.43	0.41	0.55	0.52	0.55
4	48	0.55	0.30	0.55	0.45	0.47	0.45	0.58	0.55	0.58
5	40	0.56	0.34	0.56	0.48	0.51	0.48	0.61	0.58	0.61
10	21	0.61	0.36	0.61	0.56	0.58	0.56	0.69	0.66	0.69
15	15	0.57	0.35	0.57	0.54	0.56	0.54	0.66	0.63	0.66
20	11	0.55	0.34	0.55	0.55	0.57	0.55	0.66	0.64	0.66
25	6	0.48	0.33	0.48	0.55	0.57	0.55	0.66	0.64	0.66

Armidale Animal Breeding Summer Course, UNE, Feb. 2006



Co-Expression

Results

Redundant genes allow for Validation

Table 5. Percentage of FDR and SENS for correlation between clone and gene greater than x and by each normalization method

x >	Normalization method									
	1	2	3	4	5	6	7	8	9	
FDR <sup>a</sup>										
0.50	35.63	20.81	34.65	7.09	6.85	8.75	7.25	8.72	8.72	
0.60	31.64	16.47	30.46	3.83	3.92	3.72	5.44	4.44	5.36	
0.70	26.31	11.73	24.91	1.63	1.68	1.58	2.87	2.20	2.81	
0.80	18.59	6.65	17.17	0.41	0.57	0.40	1.05	0.77	1.02	
0.85	13.52	4.18	12.18	0.14	0.21	0.13	0.45	0.31	0.44	
0.90	7.56	1.89	6.52	0.03	0.04	0.03	0.11	0.07	0.10	
0.95	1.73	0.37	1.41	0.00	0.00	0.00	0.01	0.01	0.01	
SENS <sup>b</sup>										
0.50	74.44	54.70	74.01	65.20	65.24	65.50	73.05	72.55	73.74	
0.60	68.23	46.75	68.12	56.59	57.49	56.94	66.30	64.50	67.06	
0.70	60.46	37.58	59.85	46.01	48.31	46.17	58.41	54.57	58.72	
0.80	48.52	25.71	47.35	30.85	33.62	31.26	47.45	43.48	47.65	
0.85	39.64	18.73	37.67	21.74	24.34	21.87	39.87	34.86	39.77	
0.90	27.94	10.96	25.30	12.32	14.09	12.17	30.03	25.34	29.46	
0.95	13.59	4.30	11.52	3.97	4.75	3.68	13.93	10.34	13.34	

$$^a \text{FDR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

Proportion of clones wrongly assigned to a functional annotation.

$$^b \text{SENS} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Proportion of clones correctly assigned to a functional annotation.

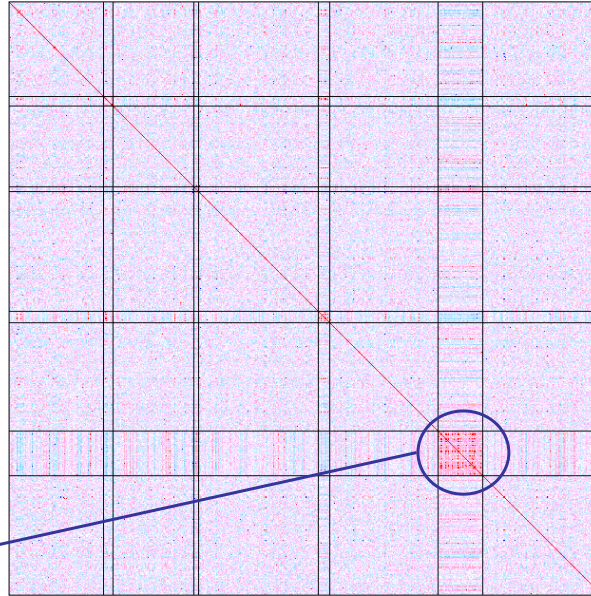
Armidale Animal Breeding Summer Course, UNE, Feb. 2006



### Co-Expression

#### Results

624 Gene Co-Expression  
(Alphabetical order)



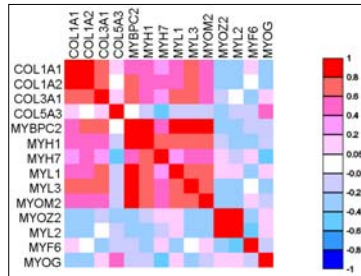
Roisosomal Proteins



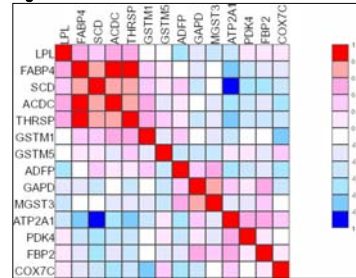
### Co-Expression

#### Applications

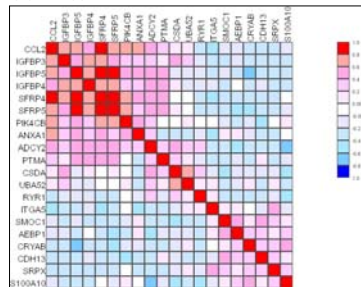
Brian



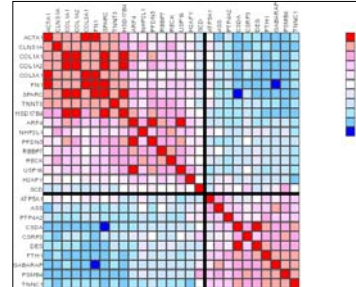
Sigrid



Siok Kwee



Keren

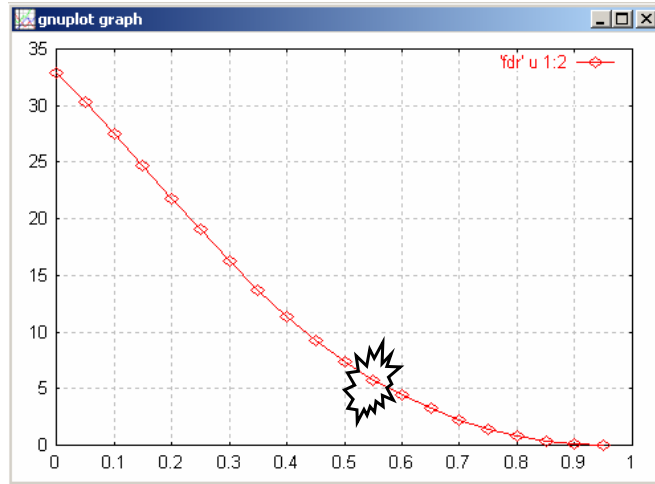




### Co-Expression

#### Applications

False Discovery Rate as a Function of Correlation



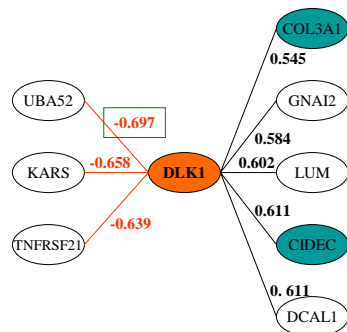
Armidale Animal Breeding Summer Course, UNE, Feb. 2006



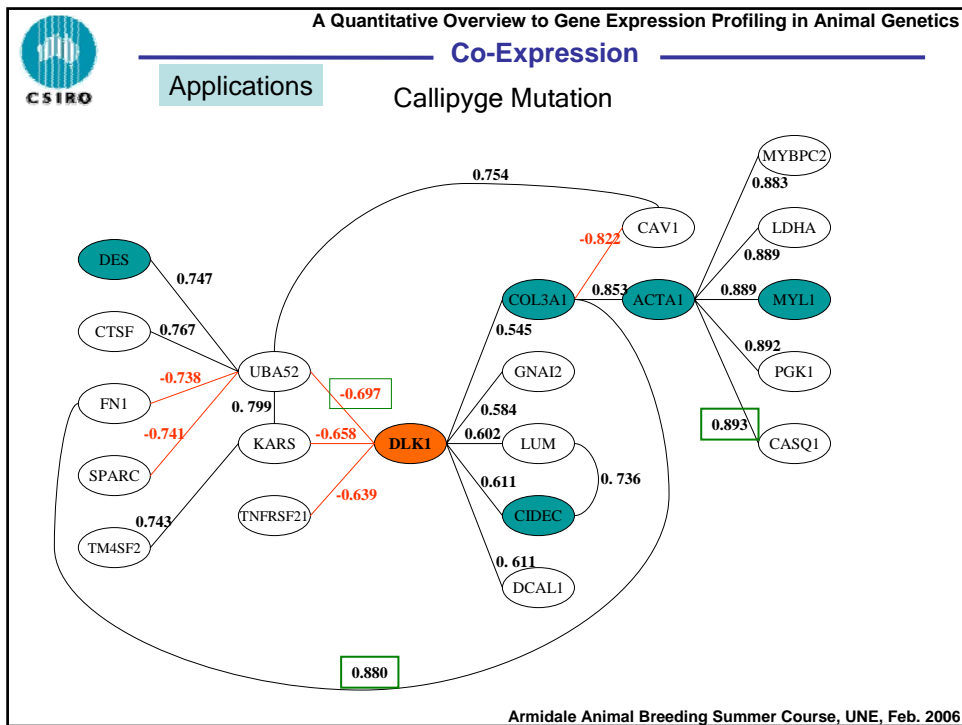
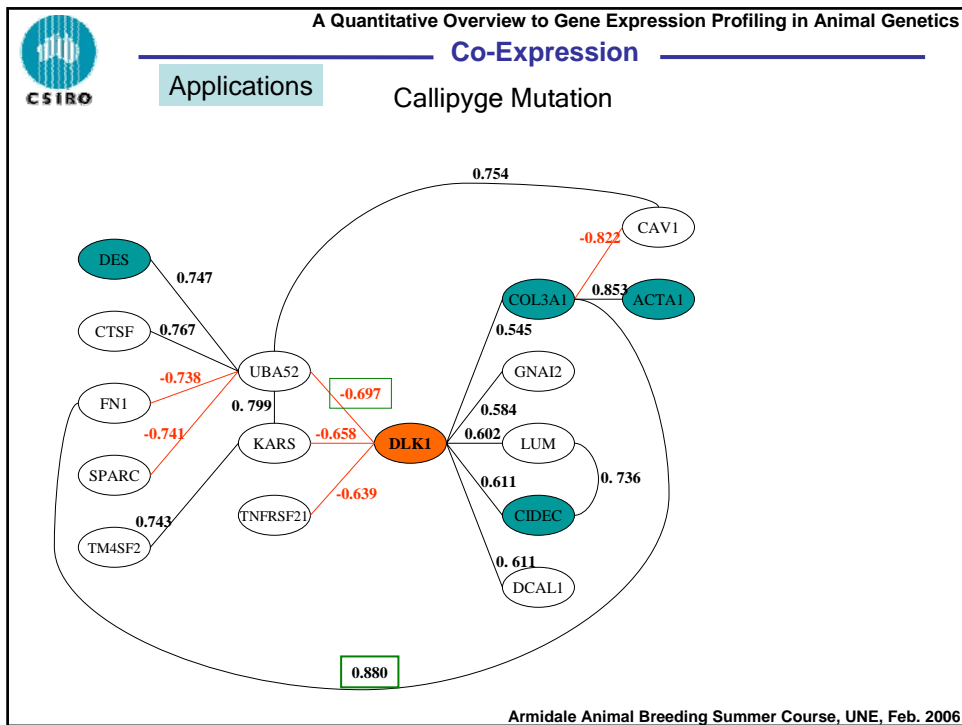
### Co-Expression

#### Applications

#### Callipyge Mutation



Armidale Animal Breeding Summer Course, UNE, Feb. 2006





### Co-Expression

Applications

Callipyge Mutation

