



Analysis of (cDNA) Microarray Data: Part VII. Mixed-Model Equations 2

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Mixed-Model Equations

Joint analysis of multiple cDNA microarray studies via multivariate mixed models applied to genetic improvement of beef cattle¹

A. Reverter², Y. H. Wang, K. A. Byrne, S. H. Tan, G. S. Harper, and S. A. Lehnert

The Cooperative Research Centre for Cattle and Beef Quality, CSIRO Livestock Industries,
Queensland Bioscience Precinct, St. Lucia, Queensland 4067, Australia

ABSTRACT: In functional genomic laboratories, it is common to use the same microarray slide across studies, each investigating a unique biological question, and each analyzed separately due to computational limitations and/or because there is no hybridization of samples from different studies on one slide. However, the question of analyzing data from multiple studies is a major current issue in microarray data analysis because there are gains to be made in the accuracy of estimated effects by exploiting a covariance structure between gene expression data across studies. We propose an approach for combining multiple studies using multivariate mixed models, with the assumption of a nonzero correlation among genes across experiments, while imposing a null residual covariance. We applied

this method to jointly analyze three experiments in genetics of cattle with a total of 54 arrays, each with 19,200 spots and 7,638 elements. The resulting seven-variate model contains 752,476 equations and 56 covariances. To identify differentially expressed genes, we applied model-based clustering to a linear combination of the random gene \times variety interaction effect. We enhanced the biological interpretation of the results by applying an iterative algorithm to identify the gene ontology classes that significantly changed in each experiment. We found 118 elements with coordinate expression that clustered into distinct biological functions such as adipogenesis and protein turnover. These results contribute to our understanding of the mechanistic processes involved in adipogenesis and nutrient partitioning.

Key Words: Beef, Complementary DNA, Gene Expression, Mixed Models

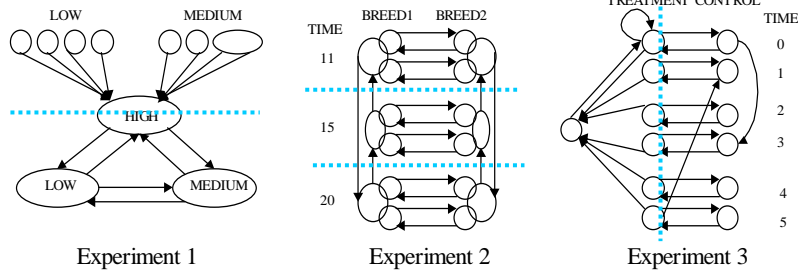
©2004 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2004. 82:3430-3439

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Mixed-Model Equations



Observations

Comparison Groups Levels Observations

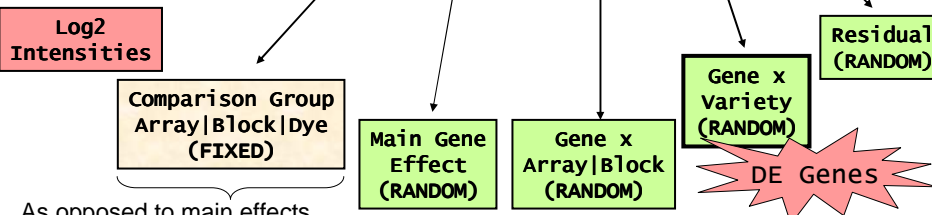
	N	Mean	SD	Min	Max	Levels	Mean	Min	Max
Y ₁₁	197,802	9.33	1.99	5.17	15.99	768	257.5	139	343
Y ₁₂	74,030	10.82	1.91	4.95	15.99	576	128.5	22	243
Y ₂₁	110,308	9.99	2.07	4.25	15.99	576	191.5	27	319
Y ₂₂	116,409	9.89	2.09	5.17	15.99	576	202.1	19	318
Y ₂₃	117,687	10.38	2.04	4.91	15.99	576	204.3	36	320
Y ₃₁	106,591	10.11	1.77	6.60	15.99	672	158.6	37	278
Y ₃₂	236,671	9.44	2.11	5.36	15.99	1,440	164.3	57	269

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Mixed-Model Equations

$$Y \sim MVN[X\beta, G\Sigma_g G^T + A\Sigma_a A^T + V\Sigma_v V^T + \Sigma_e]$$



As opposed to main effects and interactions separately

- 54 Array slides
- 959,498 valid Intensity Records (S2N>1, M2M>0.85)
- 7,638 Elements (genes)
- 752,476 Equations
- 56 (Co)variance Components (REML)
- Mixtures of distributions

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Mixed-Model Equations

56 (co)variance components

$$\Sigma_g = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_{1,2}} & \sigma_{g_{1,3}} & \sigma_{g_{1,4}} & \sigma_{g_{1,5}} & \sigma_{g_{1,6}} & \sigma_{g_{1,7}} \\ \sigma_{g_{1,2}} & \sigma_{g_2}^2 & \sigma_{g_{2,3}} & \sigma_{g_{2,4}} & \sigma_{g_{2,5}} & \sigma_{g_{2,6}} & \sigma_{g_{2,7}} \\ \sigma_{g_{1,3}} & \sigma_{g_{2,3}} & \sigma_{g_3}^2 & \sigma_{g_{3,4}} & \sigma_{g_{3,5}} & \sigma_{g_{3,6}} & \sigma_{g_{3,7}} \\ \sigma_{g_{1,4}} & \sigma_{g_{2,4}} & \sigma_{g_{3,4}} & \sigma_{g_4}^2 & \sigma_{g_{4,5}} & \sigma_{g_{4,6}} & \sigma_{g_{4,7}} \\ \sigma_{g_{1,5}} & \sigma_{g_{2,5}} & \sigma_{g_{3,5}} & \sigma_{g_{4,5}} & \sigma_{g_5}^2 & \sigma_{g_{5,6}} & \sigma_{g_{5,7}} \\ \sigma_{g_{1,6}} & \sigma_{g_{2,6}} & \sigma_{g_{3,6}} & \sigma_{g_{4,6}} & \sigma_{g_{5,6}} & \sigma_{g_6}^2 & \sigma_{g_{6,7}} \\ \sigma_{g_{1,7}} & \sigma_{g_{2,7}} & \sigma_{g_{3,7}} & \sigma_{g_{4,7}} & \sigma_{g_{5,7}} & \sigma_{g_{6,7}} & \sigma_{g_7}^2 \end{bmatrix}$$

$$\Sigma_a = \begin{bmatrix} \sigma_{a_1}^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{a_2}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{a_3}^2 & \sigma_{a_{3,4}} & \sigma_{a_{3,5}} & 0 & 0 \\ 0 & 0 & \sigma_{a_{3,4}} & \sigma_{a_4}^2 & \sigma_{a_{4,5}} & 0 & 0 \\ 0 & 0 & \sigma_{a_{3,5}} & \sigma_{a_{4,5}} & \sigma_{a_5}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{a_6}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{a_7}^2 \end{bmatrix}$$

$$\Sigma_v = \begin{bmatrix} \sigma_{v_1}^2 & \sigma_{v_{1,2}} & 0 & 0 & 0 & 0 & 0 \\ \sigma_{v_{1,2}} & \sigma_{v_2}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{v_3}^2 & \sigma_{v_{3,4}} & \sigma_{v_{3,5}} & 0 & 0 \\ 0 & 0 & \sigma_{v_{3,4}} & \sigma_{v_4}^2 & \sigma_{v_{4,5}} & 0 & 0 \\ 0 & 0 & \sigma_{v_{3,5}} & \sigma_{v_{4,5}} & \sigma_{v_5}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{v_6}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{v_7}^2 \end{bmatrix}$$

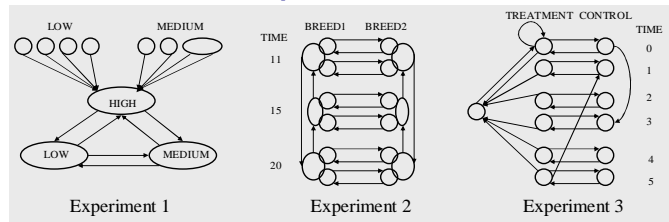
$$\Sigma_e = \text{diag}\{\sigma_{e_1}^2 \ \sigma_{e_2}^2 \ \sigma_{e_3}^2 \ \sigma_{e_4}^2 \ \sigma_{e_5}^2 \ \sigma_{e_6}^2 \ \sigma_{e_7}^2\}$$

A. Reverter – Sept. 2006, UAB, Barcelona, Spain



Mixed-Model Equations

% Total Variance Due to:



• Error	3.0 – 3.6	5.1 – 6.7	3.0 – 3.7
• Gene	83.6 – 90.4	78.3 – 81.9	47.5 – 83.9
• Gene x Array	3.5 – 9.8	10.4 – 12.6	10.6 – 43.5
• Gene x Variety	2.4 – 3.7	2.1 – 2.6	2.5 – 5.4

• Genetic Correlations Moderate (EXP3) to Strong

• Gene x Variety Corr Strong (EXP1) to Moderate (EXP2)

A. Reverter – Sept. 2006, UAB, Barcelona, Spain

**Mixed-Model Equations****Measures of (Possible) Differential Expression**

$$d1_i = \frac{1}{2} \sum_{j=1}^2 [\hat{v}_j(HIGH)_i - \hat{v}_j(LOW)_i]$$

$$d2_i = \frac{1}{3} \sum_{j=3}^5 [\hat{v}_j(BREED1)_i - \hat{v}_j(BREED2)_i]$$

$$d3_i = \frac{1}{12} \sum_{j=6}^7 \sum_{t=0}^5 [\hat{v}_j(TREATMENT)_i - \hat{v}_j(CONTROL)_i]$$

$i = 1, \dots, 7,638$ genes
 $j = 1, \dots, 7$ variables
 $t = 0, \dots, 5$ time points (EXP3 only)

❁ other measure definitions (contrasts) could also be valid

A. Reverter - Sept. 2006, UAB, Barcelona, Spain

**Mixed-Model Equations****Mixtures of Tri-Variate Normal Distributions**

$$f \begin{pmatrix} d1 \\ d2 \\ d3 \end{pmatrix} = 0.67 \times N \left[\begin{pmatrix} -0.21 \\ -0.03 \\ -0.01 \end{pmatrix}, \begin{pmatrix} 0.10 & -0.01 & -0.01 \\ -0.01 & 0.04 & 0.01 \\ -0.01 & 0.01 & 0.02 \end{pmatrix} \right] +$$

$$0.23 \times N \left[\begin{pmatrix} 0.40 \\ 0.19 \\ 0.01 \end{pmatrix}, \begin{pmatrix} 0.67 & 0.09 & -0.04 \\ 0.09 & 0.11 & -0.01 \\ -0.04 & -0.01 & 0.08 \end{pmatrix} \right] +$$

$$0.10 \times N \left[\begin{pmatrix} 0.08 \\ -0.08 \\ 0.23 \end{pmatrix}, \begin{pmatrix} 0.08 & -0.02 & 0.05 \\ -0.02 & 0.16 & 0.06 \\ 0.05 & 0.06 & 0.38 \end{pmatrix} \right].$$

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Mixed-Model Equations

Mixtures of Tri-Variate Normal Distributions

$$f \begin{pmatrix} d1 \\ d2 \\ d3 \end{pmatrix} = 0.67 \times N \left[\begin{pmatrix} -0.21 \\ -0.03 \\ -0.01 \end{pmatrix}, \begin{pmatrix} 0.10 & -0.01 & -0.01 \\ -0.01 & 0.04 & 0.01 \\ -0.01 & 0.01 & 0.02 \end{pmatrix} \right] + \text{Non-DE}$$

$$0.23 \times N \left[\begin{pmatrix} 0.40 \\ 0.19 \\ 0.01 \end{pmatrix}, \begin{pmatrix} 0.67 & 0.09 & -0.04 \\ 0.09 & 0.11 & -0.01 \\ -0.04 & -0.01 & 0.08 \end{pmatrix} \right] + \text{DE in } d_1 \text{ Up-DE in } d_2$$

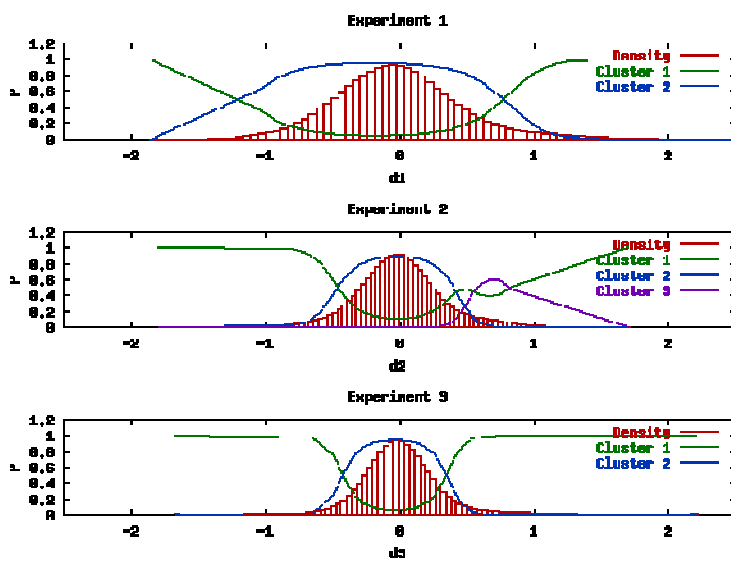
$$0.10 \times N \left[\begin{pmatrix} 0.08 \\ -0.08 \\ 0.23 \end{pmatrix}, \begin{pmatrix} 0.08 & -0.02 & 0.05 \\ -0.02 & 0.16 & 0.06 \\ 0.05 & 0.06 & 0.38 \end{pmatrix} \right] + \text{DE in } d_3 \text{ Down-DE in } d_2$$

A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Mixed-Model Equations

Densities and Posterior Probabilities



A. Reverter - Sept. 2006, UAB, Barcelona, Spain



Mixed-Model Equations

Differentially Expressed Genes

		Exp1		Exp2		Exp3	
		Up	Down	Up	Down	Up	Down
High-Low	Up	409	0	26	13	36	11
	Down		41	3	0	5	0
HOL-JBL	Up			68	0	0	8
	Down				319	10	6
TSS-UTS	Up					252	0
	Down						109

10 DE Elements across the 3 Exp
(2 UP/DOWN/UP; 8 UP/UP/DOWN)