

## INTENSITIES VERSUS INTENSITY RATIOS IN THE ANALYSIS OF cDNA MICROARRAY DATA

A. Reverter, Y.H. Wang, K.A. Byrne, S.A. Lehnert and B.P. Dalrymple  
Cooperative Research Centre for Cattle and Beef Quality  
CSIRO Livestock Industries, Queensland Bioscience Precinct  
306 Carmody Road, St Lucia, QLD 4067

### SUMMARY

Intensity (INT) and intensity ratios (RAT) records from microarray data were compared with respect to their ability to identify differentially expressed genes. Data from two cDNA microarray slides were selected from each of two separate experiments (EXP1 and EXP2). EXP1 compared muscle RNA samples from Brahman steers fed high and low quality diets and yielded 39,654 INT records on 4,785 genes. EXP2 compared muscle RNA samples from Japanese Black and Holstein cattle and produced 42,130 INT records on 4,991 genes. Half as many RAT records were available. INT and RAT were analysed with an equivalent model that included the random effect of gene by treatment interaction. A correlation of 0.98 was observed between BLUPs from the two models indicating an agreement between INT and RAT in ranking genes. Among the 50 most extreme genes, there were three and one discrepancies in EXP1 and EXP2, respectively.

**Keywords:** Gene expression, microarray, beef cattle

### INTRODUCTION

Gene expression technology is becoming increasingly accessible to animal scientists (Moody, 2001). Novel statistical challenges emerge to elucidate the information from the simultaneous expression of thousands of genes in the form of red and green fluorescent intensities. Options include the computation of the ratio (RAT) of red to green intensities for each element in the array or, alternatively, to use the individual intensities (INT) as dependent variables. Although most of the techniques developed for analysis of microarray data use RAT, many of them can be adapted for use with INT. For data quality control, INT have been found to be more useful than RAT (Tran *et al.* 2002). However, a formal comparison of INT versus RAT in the analysis of cDNA microarray data has not previously been reported.

The Cooperative Research Centre for Cattle and Beef Quality undertook genetic profiling of bovine muscle to explore the gene regulatory pathways that control gene expression response at the muscle and fat tissue levels. Within this project, the objective of this paper is to use data from two microarray slides from each of two separate experiments and to compare INT and RAT with respect to their ability to identify differentially expressed genes.

### MATERIALS AND METHODS

**Animals, genes and microarray data.** Data from two cDNA microarray slides were selected from each of two separate experiments, EXP1 and EXP2. The two microarray slides represented what is known as a “dye-swap”. In a dye-swap, a replicate experiment is performed, in which the cDNA target that was previously labelled with Cy5 (red) is now labelled with Cy3 (green) and

vice versa. EXP1 and EXP2 were conducted by two independent laboratory operators. Both experiments were conducted using the same microarray of 9,600 duplicate elements comprising of 2,200 ESTs and 7,400 anonymous cDNAs from cattle muscle and subcutaneous fat libraries. Each array comprised 19,200 spots organised in 48 blocks of 20 columns by 20 rows each. The dye-swap from EXP1 compared the RNA samples from Brahman steers fed high and low quality diets and yielded 39,654 background-corrected and base-2 log-transformed intensities on 4,785 genes. The dye-swap from EXP2 compared the RNA samples from Japanese Black and Holstein cattle at approximately 18 months of age and produced 42,130 INT records on 4,991 genes. Half as many RAT as INT observations were available from each experiment.

In this notation, each experiment contained two arrays (ARR1 and ARR2), and two treatments (TRT1 and TRT2). In EXP1, TRT1 and TRT2 corresponded to low and high quality diets, respectively. In EXP2, TRT1 and TRT2 corresponded to Japanese Black and Holstein breeds, respectively. In both experiments, TRT1 was labelled with red dye in ARR1 and with green dye in ARR2. In contrast, TRT2 was labelled with green dye in ARR1 and with red dye in ARR2.

**Data analysis and comparison.** Preliminary analyses were performed to identify sources of systematic variation using the procedure GLM of SAS (1991). Later and within experiment, the same mixed-model was fitted to INT records with the joint combination of array, block (48 levels nested within array), dye and treatment as fixed effect with 192 levels in total and gene by treatment (GxT) interaction as random effect with 9,570 and 9,982 levels in EXP1 and EXP2, respectively. The mixed-model for the analysis of RAT records included array, block and treatment contrast (TRT1 minus TRT2 and TRT2 minus TRT1) as fixed effect with 96 levels in both experiments and gene by treatment contrast (GxTc) interaction as random effect with equal levels as per the analysis of INT. REML estimates of variance components and BLUP of random effects were obtained with VCE software (Groeneveld and Garcia-Cortes 1998).

Within each experiment, results from the analysis of RAT were compared to those from the analysis of INT by exploring the BLUP for each GxTc in the RAT and comparing them to the BLUP for the equivalent GxT in the INT. Finally, the 50 (~1%) most extreme genes identified from the analysis of RAT were compared to the 50 most extreme genes identified from INT.

## RESULTS AND DISCUSSION

Table 1 shows descriptive statistics for INT and RAT records by experiments and by levels of the main design effects. The fixed effects model for RAT containing array and block nested within array explained 8.0% and 4.3% of the total variation in EXP1 and EXP2, respectively. For EXP2, no differences ( $P > 0.05$ ) were found in RAT due to array. The fixed effects model for INT containing array, block nested within array, dye channel and the array by dye channel interaction explained 8.4% and 7.1% of the total variation in EXP1 and EXP2, respectively. All four effects were significant ( $P < 0.01$ ) sources of variation in INT except dye channel in EXP1 and the array by dye channel interaction in EXP2. The lack of differences in RAT due to array in EXP2 is not entirely unexpected and indicates the potential for users to fine tune the optical scanner in such a way to produce a homogeneous amount of fluorescent red and green intensities across the spots within the array. In the absence of such fine-tuning and assuming no treatment

by dye interaction, the RAT records out of a given array are expected to be of equal magnitude and opposite sign to those RAT resulting from a dye-swap. This was the case in EXP1 (Table 1).

**Table 1. Summary statistics for intensities (INT) and red to green intensity ratios (RAT) for each experiment (EXP1 and EXP2) and by level of main effect**

Trait	Effect	Level <sup>a</sup>	N	Mean	SD	Min.	Max.
EXP1							
INT	Total		39,654	10.45	2.01	0.00	15.99
	Array	ARR1	19,938	10.94	1.64	2.00	15.99
		ARR2	19,716	9.96	2.21	0.00	15.99
	Dye	Red	19,827	10.45	2.12	0.00	15.99
		Green	19,827	10.46	1.89	0.00	15.99
	Treatment	TRT1	19,827	10.55	2.01	0.00	15.99
		TRT2	19,827	10.36	2.00	0.00	15.99
RAT	Total		19,827	-0.02	0.89	-7.38	8.01
	Array	ARR1	9,969	0.17	0.87	-7.38	4.79
		ARR2	9,969	-0.20	0.87	-7.35	8.01
EXP2							
INT	Total		42,130	9.53	2.03	0.00	15.99
	Array	ARR1	21,158	9.43	2.09	0.00	15.99
		ARR2	20,972	9.64	1.95	0.00	15.99
	Dye	Red	21,065	9.49	2.06	0.00	15.99
		Green	21,065	9.58	2.00	0.00	15.99
	Treatment	TRT1	21,065	9.54	1.96	2.32	15.99
		TRT2	21,065	9.53	2.09	0.00	15.99
RAT	Total		21,065	-0.09	0.66	-6.21	5.13
	Array	ARR1	10,579	-0.08	0.67	-5.58	5.13
		ARR2	10,486	-0.09	0.65	-6.21	5.04

<sup>a</sup>Each experiment (EXP1 and EXP2) contained two arrays, ARR1 and ARR2, and two treatments, TRT1 and TRT2.

For EXP1, REML estimates of total variance (GxT plus residual) were 3.729 and 0.755 for INT and RAT, respectively. The percentage of total variance due to GxT was 75.9% and 91.7% for INT and RAT, respectively. For EXP2, total variance was estimated at 3.958 and 0.373 for INT and RAT, respectively, and the percentage of total variance due to GxT was 75.8% and 771% for INT and RAT, respectively. The higher proportion of total variance explained by the interaction would favour RAT against INT, particularly in EXP1. However, this higher proportion comes at the expense of RAT having a variation that is 5 to 10 times smaller than that of INT while both have similar ranges. Furthermore, the exploration of the BLUP for GxT and GxTc resulting from analysing INT and RAT revealed a rank correlation of 0.98 for both

experiments. A strong agreement is expected in the potential of both variables INT and RAT to identify the same set of genes as differentially expressed across treatments.

When the most extreme 50 genes (approximately 1%) according to INT and RAT were further scrutinised, there were 3 genes in EXP1 (Table 2) and only 1 gene in EXP2 that were ranked in the top 50 by one method and not by the other method. The 3 genes with discrepancies in EXP1 that ranked among the top 50 when analysing RAT (last 3 columns of Table 2) were still ranked in the top 70 when analysing INT and the biggest change in rank was by 19 places for gene CCL009304. Similarly, the 3 genes that ranked among the top 50 when analysing INT (first 3 columns of Table 2) were also ranked in the top 72 when analysing RAT although the biggest change in rank was by 46 places for gene CCL008103. A detailed examination of gene CCL008103 revealed the existence of a treatment by dye channel interaction the effect of which was accounted for when using INT as dependent variable. Observations on INT for this gene in TRT1 at channels red and green averaged  $12.84 \pm 0.05$  and  $9.69 \pm 0.31$ , respectively, while in TRT2 at channels red and green were  $11.73 \pm 0.25$  and  $14.70 \pm 0.04$ , respectively. Unaccounted treatment by dye channel interaction was also observed for the other genes when using RAT.

A similar trend was observed for the one gene showing rank discrepancies in EXP2. Gene CCL013373 was ranked 48 and 55 when using RAT and INT, respectively, while gene CCL014972 was ranked 120 and 48 when using RAT and INT, respectively. Observations on INT for this gene in TRT1 at channels red and green averaged  $11.35 \pm 0.06$  and  $11.86 \pm 0.03$ , respectively; while in TRT2 were  $11.02 \pm 0.07$  and  $10.15 \pm 0.09$ , respectively for each channel.

**Table 2. Rank of the three genes showing discrepancies among the top 50 genes in the first experiment when analysing intensities (INT) as opposed to analysing intensity ratios (RAT)**

Gene in the top 50 with INT	Rank when analysing		Gene in the top 50 with RAT	Rank when analysing	
	INT	RAT		INT	RAT
CCL008103	10	56	CCL012284	57	41
CCL011618	49	53	CCL009178	67	49
CCL008010	50	72	CCL009304	69	50

Although RAT can help to reveal some patterns in the data, RAT records contain no information about gene expression levels and ignore various parameters that are dependent on the measured INT, including the confidence limits that are placed on any given microarray process.

## REFERENCES

- Groeneveld, E. and Garcia-Cortes, L.A. (1998) *Proc. 6th World Cong. Genet. Appl. Livest. Prod.* **27**:455.
- Moody, D.E. (2001) *J. Anim. Sci.* **79**(E. Suppl.):E128.
- SAS/STAT (1991) "User's Guide" Release 6.03 Edition, SAS Institute Inc.
- Tran, P.H., Peiffer, D.A., Shin, Y., Meek, L.M., Brody, J.P. and Cho, K.W.Y. (2002) *Nucleic Acids Res.* **30**(12):e54.