# Microarray data analysis of gene expression levels in lactating cows treated with bovine somatotropin

Corrado Dimauro[1], Ana H. Dias Francesconi[1], Aldo Cappio-Borlino[1], Mark A. McGuire[2]

[1]Dipartimento di Scienze Zootecniche, Università di Sassari, Italy

[2]Department of Animal and Veterinary Science, University of Idaho, Moscow, USA

*Corresponding author*: Corrado Dimauro. Dipartimento di Scienze Zootecniche, Università di Sassari. Via E. De Nicola 9, 07100 Sassari, Italy - Tel. +39 079 229304 - Fax: +39 079 229302 - Email: dimauro@uniss.it

**ABSTRACT -** Administration of bovine somatotropin (bST) to lactating cows results in an increase in milk production from 10 to 15%. While physiological mechanisms involved in bST administration are well known, there is limited knowledge about the mechanisms that regulate the bST action at genetic level. For this reason, a microarray experiment was conducted to identify differentially expressed genes when bST is given to milking cows. Sixteen high-density microarrays for cattle, each containing 18,263 gene spots, were used. RNA was extracted from the mammary tissue of four lactating Holstein cows, five and two days before, and one and six days after bST administration. A total of 1,251 and 1,167 differentially expressed genes were detected for mean and median expression intensities, respectively. Only the 115 genes which were identified by both mean and median intensities were taken into account. These genes were grouped into 8 clusters according to changes in expression through time points.

*Key words:* Microarray, Somatotropin, Cow.

**Introduction** - Bovine somatotropin (bST), secreted by the cow's pituary gland, is a hormone that directs nutrients towards lactation in milking cows. Administration of additional bST complements the animal's natural ability to partition nutrients toward milk production. In fact, in adequately fed milking cows, bST administration results in an increase in milk production ranging from 10 to 15% during the treatment period. While physiological mechanisms involved in bST administration are well known, there is limited knowledge about the mechanisms that regulate the bST action at genetic level. In order to gain some understanding of the genes involved in the mechanism of bST action on the mammary gland, a microarray experiment was conducted (University of Idaho, Moscow, USA). The aim of this work was to detect genes that express differentially before and after bST administration in the mammary gland of lactating cows.

**Material and methods** - The experiment consisted of 16 high-density microarrays for cattle (Suchyta *et al.*, 2003) from the National Bovine Functional Genomics Consortium (NBFGC) with 18,263 gene spots in each microarray. A web-accessible resource has been established (http://www.cafg.msu.edu) to provide information on all EST clones contained in the microarray. Samples of RNA extracted from mammary tissue of 4 lactating Holstein cows were collected five (-5) and two (-2) days before bST administration, and one (1) and six (6) days after. RNA samples were hybridized with two different dyes (Cy3 and Cy5) in accordance with an incomplete loop design. Scan Array analysis was performed by a two-laser scanner (Model GenePix 400, Axon Instruments, Union City, CA, USA) with the software GenePix Pro 3.0. Time points (-5) *vs.* (-2) were used to ascertain the potential for false-positive expression changes before the bST administration, whereas time points (-5) *vs.* (6) were used to investigate the somatotropin effect on gene expression. Spot quality control, based on a simple automated method (Tran *et al.*, 2002), led to the deletion of spots with mean and median correlation lower than 0.80.

After log-base 2 transformations, the systematic bias of fluorescence intensity was controlled using the LOWESS regression (Yang *et al.*, 2002) between the intensity log-ratio $M = \log_2 R/G$ and the mean intensity $A = \log_2 \sqrt{RG}$, where R and G are the red and green channel, respectively. A linear mixed model (Wolfinger *et al.*, 2001) was then used to remove the global dye effect from the log-base 2 lowess-corrected intensities $y_{gjk}$ from gene g (g=1, ..., 18,263), labeled with dye j (j=1 for R and j=2 for G) in array k (k = 1, …, 16). The model was:

$$y_{gjk} = \mu + D_j + A_k + (AD)_{jk} + \varepsilon_{gjk}$$

where μ represents the global mean value, D is the main effect of dye, A is the random effect of array, AD is the random interaction effect of array per dye and ε is stochastic error. This model corrects effects that are not gene specific. Residuals $r_{ijk}$ from this model, computed by subtracting the fitted values from the $y_{gjk}$, were used as input data for the gene-specific mixed-model that was used to find differentially expressed genes (Wolfinger *et al.*, 2001):

$$r_{gijk} = \mu_g + T_{gi} + D_{gj} + A_{gk} + \gamma_{gijk}$$

where μ is the global mean, T is the main effect of time point, D is the main effect of dye, A is the random effect of array, and γ is the stochastic error. The dye term controls for the fact that Cy3 and Cy5 fluorescent labels could not incorporate into target RNA with the same efficiency for each gene, whereas the array term controls for variation between arrays. Since more than 18,000 genes were studied, the significance of differences in gene expression was corrected for multiple testing error rate using the permutation method (Pounds, 2006). In accordance with this approach, data were assigned randomly to groups (R and G dyes, -5 and 6 time points) and exact tests were computed for the assignments of data to the groups. The p-value was calculated by comparing the test statistic computed from all the assignments with the test statistic calculated for the original dataset. Using the SAS mixed procedure implemented in a macro, F-value of each permuted sample was evaluated, for a total of 1,000 permutations, and final p-value for each gene was:

$$\text{p-value} = \frac{\text{number of F-value} > \text{F-value for original dataset}}{1000}$$

One gene was declared differentially expressed if the p-value was lower than 0.05. All statistical analyses were performed separately for mean and median intensities. Differentially expressed genes that matched for both mean and median intensities were the final detected genes. The trajectory clustering method (Phang *et al.*, 2003) was then used to group differentially expressed genes. This technique is particularly indicated in time course experiments, grouping genes on the basis of the direction of change of gene expression between adjacent time points. The directions of change were: increasing (I), decreasing (D), or flat (F). In general, for a time series containing T points, there are T-1 changes and $3^{T-1}$ possible trajectories. Only time points (-5), (1), and (6) were used and, being T=3, $3^{3-1}=3^2=9$ trajectories were possible. The differentially expressed genes were assigned to the I or D trajectory if the absolute value of fold change between adjacent time points was greater than 0.05. All genes that did not fit this criterion were assigned to the F trajectory.

**Results and conclusions** - A total of 1,251 and 1,167 differentially expressed genes, for mean and median intensity respectively, were detected. Only 115 genes matched for both mean and median intensities (Table 1). These genes were the top differentially expressed genes, whereas for the residual differentially expressed genes (1,136 for mean and 1,052 for median intensities) further studies should follow. Results of trajectory clustering gave 8 clusters that grouped the top differentially expressed genes, as listed in Table 1. The trajectories within each cluster varied as listed in Table 2.

Analysis of directions gave the expression patterns of genes in each cluster. As an example, genes in cluster 1, which were assigned to FF directions, had a flat expression through time points. Differently, FD directions in cluster 2 indicate that gene expression did not decrease immediately after bST administration (time point 1), but later (time point 6), whereas FI directions in cluster 3, indicate that gene expression increased at time point 6.

Results of this study, combined with gene ontology, could increase the knowledge of genes that affect the physiological mechanisms of milk secretion in lactating cows after bST administration.

Table 1. Expressed genes (letter and number code) and related clusters (from 1 to 8).

| Gene | Cluster | Gene | Cluster | Gene | Cluster | Gene | Cluster | Gene | Cluster |
|---|---|---|---|---|---|---|---|---|---|
| AW298846 | 1 | BF606843 | 4 | AW658468 | 6 | BG690227 | 6 | BE683376 | 7 |
| AW653381 | 1 | BF705806 | 4 | AW660666 | 6 | BF602200 | 7 | BE723367 | 7 |
| BE685349 | 1 | AW325448 | 5 | BE236712 | 6 | AW312680 | 7 | BE723560 | 7 |
| AW315465 | 2 | AW353417 | 5 | BE479888 | 6 | AW313994 | 7 | BE749908 | 7 |
| AW484883 | 2 | AW425419 | 5 | BE480182 | 6 | AW315831 | 7 | BE750226 | 7 |
| AW652575 | 2 | AW478998 | 5 | BE590166 | 6 | AW354374 | 7 | BE752589 | 7 |
| AW653506 | 2 | BE477211 | 5 | BE667286 | 6 | AW430157 | 7 | BE756733 | 7 |
| BE751584 | 2 | BE665405 | 5 | BE667343 | 6 | AW484018 | 7 | BF076556 | 7 |
| BE754485 | 2 | BE666109 | 5 | BE683329 | 6 | AW484876 | 7 | BF603152 | 7 |
| AW312024 | 3 | BF074374 | 5 | BE722785 | 6 | AW632107 | 7 | BF603314 | 7 |
| AW357737 | 3 | BF605502 | 5 | BE722842 | 6 | AW656160 | 7 | BF604486 | 7 |
| AW426120 | 3 | BF605866 | 5 | BE723306 | 6 | AW657790 | 7 | BF605433 | 7 |
| AW430389 | 3 | BF651890 | 5 | BE723964 | 6 | AW660289 | 7 | BF774779 | 7 |
| BE485779 | 3 | AW308326 | 6 | BE750795 | 6 | BE236710 | 7 | BF775725 | 8 |
| BE486079 | 3 | AW312212 | 6 | BF073101 | 6 | BE480174 | 7 | AW307905 | 8 |
| BE808112 | 3 | AW313719 | 6 | BF077182 | 6 | BE480217 | 7 | AW311840 | 8 |
| BE845929 | 3 | AW315313 | 6 | BF602365 | 6 | BE486294 | 7 | AW325468 | 8 |
| AW313889 | 4 | AW336299 | 6 | BF651311 | 6 | BE589167 | 7 | AW652590 | 8 |
| AW336277 | 4 | AW427925 | 6 | BF653254 | 6 | BE663576 | 7 | BE588723 | 8 |
| AW428415 | 4 | AW428375 | 6 | BF707071 | 6 | BE663977 | 7 | BE685133 | 8 |
| AW484933 | 4 | AW653219 | 6 | BF707397 | 6 | BE664034 | 7 | BE751252 | 8 |
| BE723922 | 4 | AW656797 | 6 | BF775402 | 6 | BE667059 | 7 | BE809770 | 8 |
| BF074667 | 4 | AW657058 | 6 | BG689092 | 6 | BE682130 | 7 | BF074503 | 8 |

Table 2. Direction changes (I=increasing, D=decreasing, F=flat) in adjacent time points within clusters.

| Clusters | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Directions | FF | FD | FI | DF | DD | DI | ID | II |

*The research was supported by the Ministero dell'Istruzione, dell'Università e della Ricerca of Italy (F.A.R. 2008) and USDA-IFAFS #2001-52100-11211.*

**REFERENCES** - **Phang**, T.L., Neville, M.C., Rudolph, M., Hunter, L., 2003. Trajectory clustering: a non-parametric method for grouping gene expression time courses with applications to mammary development. Pacific Symposium on Biocomputing. 8:351-362. **Pounds**, S.B., 2006. Estimation and control of multiple testing error rates for microarray studies. Brief. Bioinform. 7:25-36. **Suchyta**, S.P., Sipkovsky, S., Kruska, R., Jeffers, A., McNulty, A., Coussens, M.J., Tempelman, R.J., Halgren, R.G., Saama, P.M., *et al.*, 2003. Development and testing of a high-density cDNA microarray resource for cattle. Physiol. Genomics. 15:158-164. **Tran**, P.H., Peiffer, D.A., Shin, Y., Meek, L.M., Brody, J.P., Cho, K.W.Y., 2002. Microarray optimizations: increasing spot accuracy and automated identification of true microarray signals. Nucleic Acids Res. 30:12 e54. **Wolfinger**, R.D., Gibson, G., Wolfinger, E., Bennett, L., Hamadeh, H., Bushel, P., Asfhari, C., Paules, R.S., 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J. Compu. Biol. 8:625-637. **Yang**, Y.H., Buckley, M.J., Dudoit, S., Speed, T.P., 2002. Comparison of methods for image analysis on cDNA microarray data. J. Comput. Graph. Stat. 11:108-136.