Combining gene expression data with SNP association studies to identify genes affecting bovine milk production traits

Vander Jagt, C.J.[1,2], Goddard, M.E.[1,2], Cocks, B.G.[2]

[1]Department of Agriculture & Food Systems, Melbourne School of Land & Environment, The University of Melbourne, Australia. [2]Department of Primary Industries (DPI), Biosciences Platform, Victoria, Australia.

Background

Complex traits, such as milk production, are thought to be controlled by very many genes but there has been little success in identifying those genes. Two different types of experiments that might help identify these genes are microarray experiments and genome wide association studies (GWAS) but there have been few systematic studies to determine if these two approaches identify the same genes. Here we compare genes that are differentially expressed in mammary cells in response to treatments that affect milk synthesis with genes that contain single nucleotide polymorphisms (SNPs) that are associated with milk production.

Objective

To identify genes that have a functional role in the control of bovine milk production.

Methods

- In a collaboration between the Department of Primary Industries (DPI) Victoria and AgResearch (New Zealand) a series of cDNA microarray experiments were conducted examining gene expression changes in the bovine mammary gland during lactation. In this study, we have analysed five of these experiments and they are listed below:
  1) *In vivo* bromocriptine-mediated prolactin loss
  2) *In vitro* bovine mammary epithelial cells in response to prolactin and extracellular matrix
  3) Extended lactation program in cows examining the difference between:
     - Early and late lactation
     - Persistent and non-persistent dairy cows
     - Cows in extended lactation with low, medium or high nutritional status
  4) Genetic merit experiment: comparing cows of high and low genetic merit under normal conditions, when fasted for 36 hrs and when on 50% rations for 7 days
  5) Time since milking: 6hrs, 12hrs, 18hrs, 24hrs and 36hrs post milking in cows
- Microarray images generated were quantified using ImaGene 5.5 (Biodiscovery Inc) and then analysed with the LIMMA (Linear Models for Microarray Data) software (Smyth, 2004; Smyth and Speed, 2003).

.

- Expressed sequence tags (ESTs) displaying minimum 1.5 fold differential expression ($p < 0.1$) were identified and annotated using Ensembl gene identifiers, Gene Ontology terms and Kegg pathways.
- Functional clusters associated with genes were obtained using DAVID (Huang *et. al.*, 2009 (a); Huang *et. al.*, 2009 (b))
- Genes significantly differentially regulated across multiple experiments were selected, making up a candidate list and mapped to the bovine genome (UMD3.0).
- Single nucleotide polymorphisms (SNPs) falling either within the candidate genes or 50kb on either side of each gene were identified.
- These SNPs had previously been tested for an association with bovine milk production traits in genotyping experiments using 800k (Illumina, n = 2000 Holstein cows) SNP chip.

Results & Discussion

- 680 ESTs were found to be significantly differentially expressed in more than one microarray experiment.
- These 680 ESTs mapped to 368 annotated bovine genes (UMD3.0).
- The top 5 functional clusters associated with these 368 genes were:
  - Milk / mammary gland $p < 3.0E-9$
  - mRNA processing $p < 3.5E-5$
  - Lipid metabolism $p < 3.4E-4$
  - Regulation of apoptosis $p < 7.2E-4$
  - Extracellular matrix $p < 2.3E-3$
- The total number of SNPs in the 368 genes ± 50kb was 12,604.
- The number of SNPs significantly associated with milk production traits within the candidate genes ± 50kb are outlined in Table 1.
- Significance testing (Chi-squared) of the results revealed that for all results, more significant SNPs were found within the candidate genes ± 50kb than what you would expect by chance ($P < 0.0001$).
- More stringent and thorough significance testing is required.
- Of the 368 genes, the number of individual genes ± 50kb that contained significant SNPs for milk production traits was:
  - Fat: 37
  - Volume: 65
  - Protein: 71
- One of the genes ± 50kb containing significant SNPs for protein was MAF1 or Repressor of RNA polymerase III transcription MAF1 homolog.
- MAF1 displayed significant differential regulation in both prolactin microarray experiments (experiments 1 and 2).
- 11 SNPs significant ($F < 0.001$) for milk protein were found in MAF1 ± 50kb.

Milk production trait	Expected number of significant SNPs	Observed number of significant SNPs	Is result significant? ($p < 0.0001$)
Fat	113	206	Yes
Volume	181	386	Yes
Protein	195	410	Yes

Table 1. Number of significant SNPs ($F < 0.001$) for milk: fat, volume and protein found in candidate genes ± 50kb. Expected values were obtained by applying the proportion of all significant SNPs on the SNP-chip to the 12,604 SNPs found in all candidate genes ± 50kb. Significance testing was performed using 2 x 2 Chi-squared tests.

Conclusions

Genes that are differentially expressed in the mammary gland in response to treatments that affect milk synthesis contain significantly more SNPs associated with milk production than expected by chance. The 173 genes that were differentially expressed in more than one microarray experiment and contained SNPs significantly associated with milk production are likely to have a functional role in milk synthesis. This approach can identify new genes that are important in the functioning of the mammary gland and this may improve selection of dairy cattle for milk production.

References

Huang DW, Sherman BT, Lempicki RA (2009). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc. 4(1):44-57*

Huang DW, Sherman BT, Lempicki RA (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res 37(1):1-13*

Smyth, G.K. (2004). Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 1, Article 3.

Smyth, G.K., and Speed, T.P. (2003). Normalization of cDNA Microarray Data. *Methods* **31**, pp. 265-273.

Acknowledgments: Dr Jennie Pryce, Liz Ross, Lesley Gray and Charlotte Anderson for their assistance with script writing and the SNP analysis work and Alan McCulloch and Adrian Molenaar from AgResearch NZ for all their help with obtaining the microarray data.

Victoria The Place To Be | The University of Melbourne | Department of Primary Industries | biosciences research