Exploring the Effect of Pathogen Associated QTL on the Hepatic Acute Phase Response to Experimental *E.coli* Mastitis using Hybrid Markov Network Modelling

D. Edwards[*], C.M. Røntved[†], K.L. Ingvartsen[†] and P. Sørensen[*]

Introduction

Mastitis is a prevalent disease affecting dairy cattle, and knowledge of the genetic basis of susceptibility is of considerable economic importance. Mastitis due to Gram negative bacteria is primarily caused by infection with *E.coli*. Markov networks have been widely used to model gene co-regulation networks, but methods to handle high-dimensional networks with both continuous (Gaussian) and discrete variables (the so-called *hybrid* networks) have only been developed recently: see Edwards et al. (2010). This development allows network modelling to be applied to data from transcription profiling experiments, in which there are both gene expression variables (continuous) and design variables (discrete). Here we apply the approach to data from a study of the genetic basis of the response to *E.coli* infection in Danish Holstein-Friesian dairy cows.

Material and methods

Description of Experiment

In a previous study, a QTL associated with mastitis was found, and haplotypes associated with high or low resistance to clinical mastitis were identified in the QTL region Sørensen et al. (2008). The estimated allele substitution effect of the high resistance haplotype relative to the low resistance haplotype was 0.387 based on the clinical mastitis breeding values (standardized to zero mean and unit variance). Each had a population frequency of 17%.

In the present study, sixteen primiparous Danish Holstein-Friesian dairy cows were selected in early lactation so that eight had the high-resistance haplotype and eight the low-resistance. Fourteen animals carried either one high-resistance or one low-resistance haplotype, and two were homozygous, one for the high-resistance haplotype and the other for the low-resistance haplotype. The animals were challenged intra-mammarily with *E.coli* using a low dose (20-40 CFU/ml) in one quarter. Liver biopsies were collected at times -144, +12, +24, and +192 hours relative to infection using a mild invasive method as described in Vels et al. (2009). The liver biopsies were frozen immediately in liquid nitrogen and transported to the laboratory

[*]Dept of Genetics and Biotechnology, Faculty of Agricultural Sciences, Aarhus University, Denmark
[†]Dept of Animal Health and Bioscience, Faculty of Agricultural Sciences, Aarhus University, Denmark

where the tissues were stored at $-80°$C until RNA extraction. The expression profiles were measured using the Affymetrix Bovine Genome Array as in Jiang et al. (2008).

Statistical analyses.

The arrays contained 24128 probe sets. In a non-specific filtering step, probes whose marginal variance was small or those with duplicate EntrezGene identifiers were omitted. Then RMA normalization was applied. The analysis dataset comprised 5515 log expression values, and two design variables: time and QTL haplotype.

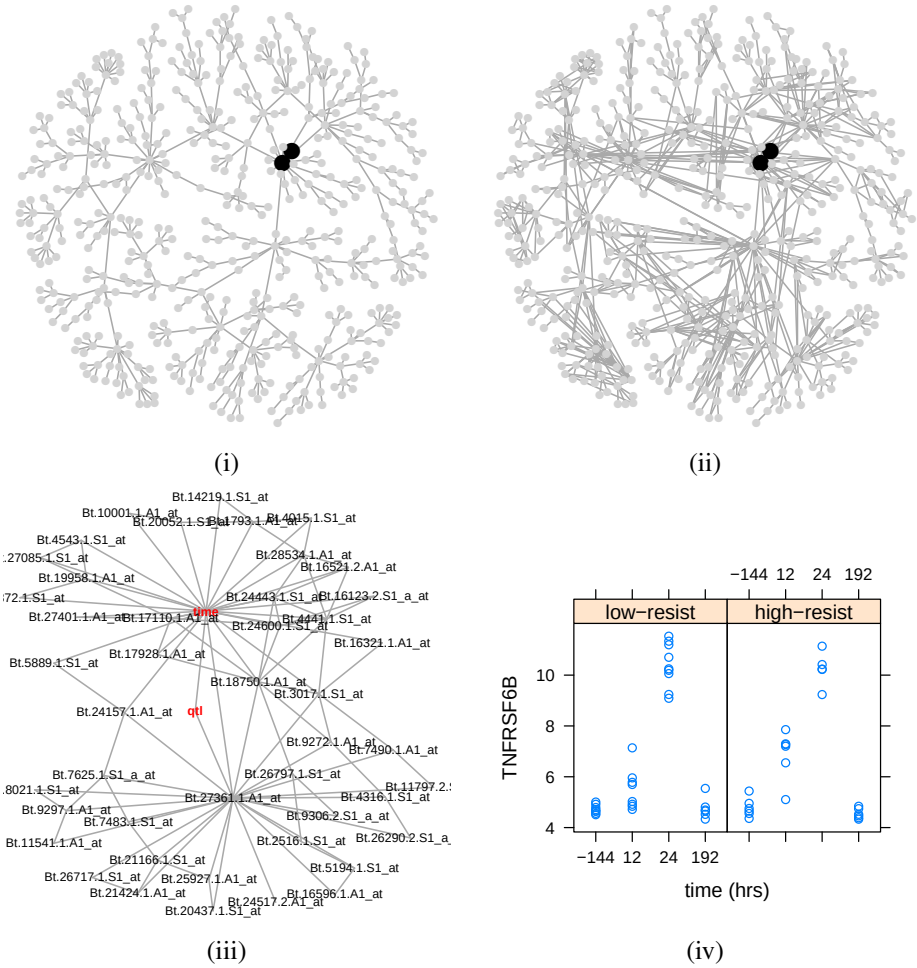

Figure 1: (i) The radius eight neighbourhood of the design variables in the minimal BIC forest. The design variables are shown as black nodes. (ii) The minimal BIC decomposable model for the variables in (i). (iii) The local neighbourhood of the design variables in (ii). The design variables are shown in red. (iv) Expression profiles of TNFRSF6B for the two QTL groups.

Markov network models are statistical models for networks, in which nodes represent random variables and missing edges correspond to conditional independence relations between variables: if the edge between two nodes is missing then the two variables are conditionally independent given the remaining variables.

From this dataset a hybrid Markov network was derived in two steps. First, a simple hybrid network model (the minimum BIC forest) was found using the algorithm described in Edwards et al. (2010). To condition on the design variables, all models considered contain the edge between time and QTL haplotype: see Kirshner et al. (2004). From this network a subset of variables, namely those with a path length to the design variables of eight or less, was identified. Then a greedy algorithm described in de Abreu et al. (2010) was used to find a richer hybrid network model (the minimum BIC decomposable model) for this reduced subset.

Results and discussion

Figure 1(i) shows the radius eight neighbourhood of the design variables in the minimal BIC forest: this contains 546 probes. Figure 1(ii) shows the minimal BIC decomposable model for this reduced subset, and Figure 1(iii) shows a subgraph of (ii) containing the design variables. Whereas many probes are adjacent to the time variable, the QTL haplotype is only adjacent to one: `Bt.27361.1.A1_at`. This indicates that whereas the pathogen challenge had a substantial effect on a large number of genes in the liver, the differential effect of the QTL haplotype was relatively modest. The model suggests that the effect of QTL haplotype on the gene expression time-profiles is mediated by the probe `Bt.27361.1.A1_at`. This is a probe for the TNFRSF6B gene, a member of the tumor necrosis receptor family. The TNF super family is also top-ranked in a limma analysis of QTL effect (not shown). The differing expression profiles of TNFRSF6B for the two QTL groups are shown in Figure 1(iv). Elevated expression of TNFRSF6B at 12 hours after infection is observed in the high-resistance haplotype group. TNFRSF6B is a known immune response gene, but is not located in the QTL region. There are two possible explanations for this. Either the QTL region contains a regulatory element for this gene, or the differential gene expression of TNFRSF6B is an indirect effect of the QTL. More investigations are required to clarify this.

Further work is also required to assess the stability of the network selection algorithms. A possible approach would be to use resampling techniques, that is, apply the algorithms to resampled datasets in order to evaluate the distribution of the nodes adjacent to the QTL variable.

Software

The analyses were performed using the R library gRapHD: see de Abreu et al. (2010). This library is available on the CRAN repository.

Conclusion

The analysis using hybrid Markov networks supplements a conventional limma analysis of gene expression by providing a tentative network model for the differential effect of the QTL

haplotype. This illustrates that recent advances in network modelling enable more informative analyses of complex high-dimensional datasets to be performed.

References

de Abreu, G. C., Labouriau, R., and Edwards, D. (2010). *J. Stat. Software (submitted)*.

Edwards, D., de Abreu, G. C., and Labouriau., R. (2010). *BMC Bioinformatics*, 11(1):18.

Jiang, L., Sørensen, P., Røntved, C., Vels, L., and Ingvartsen, K. L. (2008). *BMC Genomics*, 9:443.

Kirshner, S., Smyth, P., and Robertson, A. W. (2004). In *UAI '04: Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 317–324, Arlington, Virginia, United States. AUAI Press.

Sørensen, L. P., Guldbrandtsen, B., Thomasen, J. R., and Lund, M. S. (2008). *J Dairy Sci*, 91(6):2493–2500.

Vels, L., Røntved, C. M., Bjerring, M., and Ingvartsen, K. L. (2009). *J Dairy Sci*, 92(3):922–934.