

# Weighted Co-Expression Gene Modules Reveals Candidate Biomarkers for Sheep Intestinal Parasite Resistance

Haja N. Kadarmideen<sup>1</sup> Nick Andronicos<sup>2</sup> Nathan S. Watson-Haigh<sup>1</sup>  
Haja.Kadarmideen@csiro.au Nick.andronicos@csiro.au Nathan.watson-haigh@csiro.au

Peter Hunt<sup>2</sup> Ross Windon<sup>2</sup>  
Peter.hunt@csiro.au Ross.windon@csiro.au

<sup>1</sup>CSIRO Livestock Industries, JM Rendel Laboratory, PO Box 5545, Rockhampton QLD 4702, Australia

<sup>2</sup>CSIRO Livestock Industries, FD McMaster Laboratory, Chiswick, Armidale, NSW 2350, Australia

## Abstract

Microarray gene expression data collected from a sheep gastrointestinal parasite challenge experiment was used in a weighted gene co-expression network analysis (WGCNA) to detect parasite resistance gene modules. Subsequently, we applied integrative systems biology approach in detecting two distinct gene modules and candidate biomarkers within each module for parasite resistance in sheep.

**Keywords:** Sheep, intestinal parasite, gene co-expression networks, gene modules, systems biology

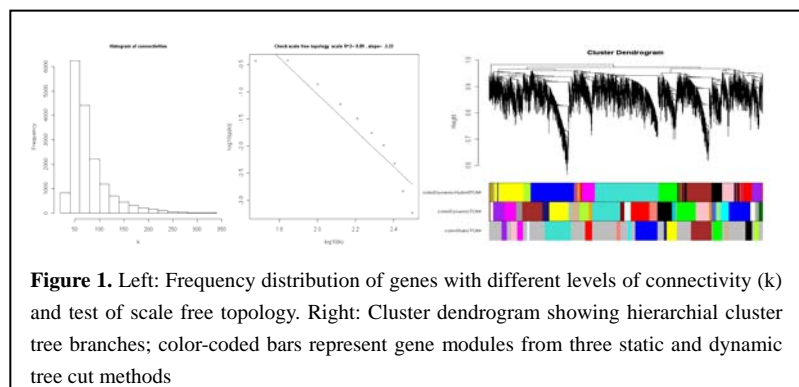
## 1 Introduction

This paper is based on microarray gene expression data collected from a sheep gastrointestinal nematode (GIN) parasite challenge experiment that was designed to examine the gene expression profile of a primary immune response in genetically resistant sheep in order to identify candidate biomarkers that correlate with infection in resistant animals. WGCNA and systems biology methods were used to detect gene modules and candidate genes or biomarkers for parasite resistance in sheep based on various criteria.

## 2 Method and Results

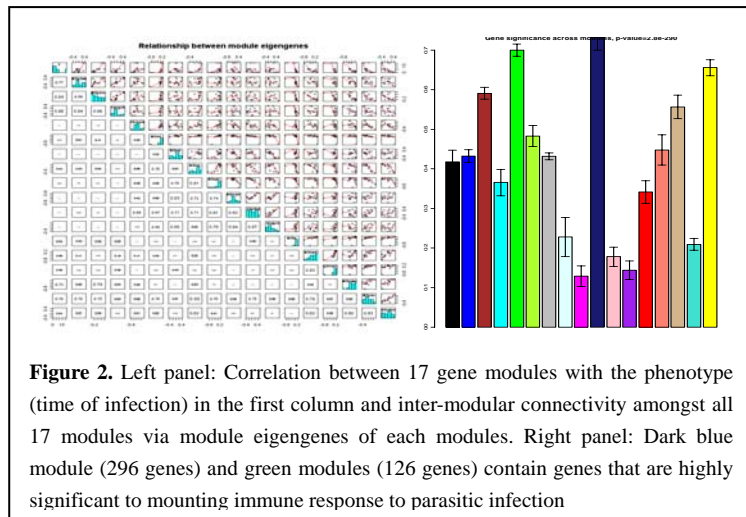
Sheep from the resistant line of the *T. colubriformis* selection flock which are genetically resistant to GIN infections were given a single oral challenge of either 8000 L3 *Haemonchus contortus* (Hc) or 20000 L3 *Trichostrongylus colubriformis* (Tc) nematodes to examine the host gene expression response of resistant lambs to a primary challenge with pathogenic GIN. For the Hc challenge,

abomasal tissue (site of infection response) and blood (systemic response) were taken; and for the Tc challenge, jejunum tissue (site of infection response) and blood (systemic response) were taken at 0, 3, 7 and 21 days post-challenge with 4 biological replicates for each sample at each time point (64 arrays total). The co-expression networks and gene modules were constructed separately for Hc and Tc parasites, and for each one of the 3 tissues. However, we only present result for the 16 Hc challenged abomasum samples (HcA).



The WGCNA distinguishes from other unweighted network construction methods in that it assigns weights to each edge by soft thresholding rather than hard thresholding, where the later could result in some genes not making it into the network due to cut-off threshold values. WGCNA begins with calculation of the absolute value of the Pearson correlation ( $\rho_{ij}$ ) between two genes and the adjacency function,  $a_{ij}$ , among two genes as:  $(\rho_{ij})^\beta$ , which is a kernel of weighted co-expression gene network [1]. To determine the threshold parameter ( $\beta$ ), scale-free topology criterion was used. We applied the above WGCNA method for the 16936 most varying non-redundant transcripts across 16 ( $\times 4$ ) arrays, with a power of 6 used to calculate  $a_{ij}$ .

Fig 1 shows the frequency of genes and their level of connectivity,  $k$ . It shows a large number of genes with  $k = 50-70$  and small number of genes with  $k > 200$ . Those genes with  $k > 200$  and high correlation with time of infection (gene significance) should be considered as major hub acting genes. The straight line relationship shows approximate scale free topology. There were 17 different modules detected for HcA with the blue module (296 genes) being the most significantly related to time post infection (Fig.2). For each module, the module eigengenes were calculated in order to compute inter-modular connectivity as



in [2]. Fig.2 shows that different modules may be highly correlated. We then carried out a functional enrichment analysis of each module by linking with external biological metadata. For instance, for HcA, genes in the blue module that had the most significant relationship with time of infection were filtered out based on their module membership  $> 0.7$  and their gene significance  $> 0.8$ . There were 16 genes that met these criteria; they were subsequently uploaded into EASE (David: <http://david.abcc.ncifcrf.gov/summary.jsp>) to link them to biological metadata and choose candidate biomarkers. Gene ontology results showed that 7 of the genes

in the blue module were involved in the activation of immune response, inflammatory response, cell mediated and humoral immune response, response to wounding, regulation of endocytosis and other functions. Four of the genes were represented in KEGG pathways relating to autophagy, notch signaling and glycolysis. Same systems biology investigations were conducted for other important modules to retrieve biological function and ascertain which candidate genes are druggable targets and hence serve as biomarkers.

### 3 Discussions

In this paper we have constructed co-expression gene modules and related them to time of infection as well as to a biological metadata – a systems biology approach; this identified a few candidate genes in green and dark blue gene modules. These genes were highly enriched in cell mediated and humoral immune functions offering resistance to gastrointestinal nematodes. Hence the expression levels of these genes may function as diagnostic tools indicating a parasite resistant phenotype in sheep and possibly other livestock species.

### 4 Acknowledgments

This project was financially supported by SheepGenomics, an initiative of Australian Wool Innovation Limited and Meat and Livestock Australia.

### References

- [1] Zhang B, Horvath S (2005). A General Framework for Weighted Gene Co-Expression Network Analysis. *Statistical Applications in Genetics and Molecular Biology*, 4 (1), Article 17
- [2] Langfelder P, Horvath S (2007). Eigengene networks for studying the relationships between co-expression modules. *BMC systems biology*, 1, Article: 54