

SYSTEMS GENETICS ANALYSIS REVEALS GENE MODULES AND HERITABLE BIOMARKERS FOR SHEEP INTESTINAL PARASITE RESISTANCE

H. N. Kadarmideen¹, N. Andronicos² and N. S. Watson-Haigh³

CSIRO Livestock Industries, ¹JM Rendel Laboratories, Rockhampton QLD 4702, ²FD McMaster Laboratories, Armidale NSW 2350, ³Queensland Bioscience Precinct, St.Lucia QLD 4067

SUMMARY

Systems genetics methods were applied to microarray gene expression profiling data from a sheep gastrointestinal nematode (GIN) challenge experiment that was designed to detect genes associated with resistance to GIN. Analyses went beyond just detecting and annotating differentially expressed (DE) genes. This included detection of co-expressed (CE) gene modules associated with the duration of infection, essential hub genes, functional enrichment and pathway analyses. Results revealed that DE genes were highly enriched in functions such as cell-mediated and humoral immunological response to GIN. Further, heritabilities were estimated for expression phenotypes of such candidate biomarkers (range 0.05 to 0.9 with high s.e.) indicating their potential for expression-assisted selection. Hence, the systems genetics method is a key step in identifying biologically relevant and heritable genes/biomarkers amongst several sets of DE genes. This approach would provide specific targets for breeding and therapeutic interventions.

INTRODUCTION

This study is based on a sheep GIN challenge experiment that was designed to examine the microarray gene expression profiles obtained from sheep genetically resistant to parasitism to identify candidate genes involved in the resistance response. Identification of differentially expressed (DE) genes using standard methods do not reveal the complex interactions between individual genes (whether they are DE or not) in a given set of biological perturbations. An intuitive way would be to identify which genes are co-expressed (CE) instead of which genes are DE. From microarray data, a strong correlation of gene expression for a pair of genes implies that both genes act within a common functional group and are under similar transcriptional control. Genes with high level of connectivity with other genes in the CE network act as major hubs that are essential for important biological functions (e.g., Weston *et al.* 2008). With this as a basis, we applied a more general weighted gene co-expression network analysis (WGCNA) method for network construction following Zhang and Horvath (2005). The heritability (h^2) of gene expression is a powerful indicator to determine if expression profiles of regulatory hub or candidate genes in individuals can be used in *expression assisted selection* (Kadarmideen *et al.* 2006). In this study, we estimate heritabilities for candidate biomarkers or hub genes revealed by integrated analyses.

MATERIALS AND METHODS

Helminth naïve sheep (n=32) from the resistant line of the *T. colubriformis* (Tc) selection flock (TSF), which are genetically resistant to GIN infections were used in the experiment. They were given a single oral challenge of either 8000 *Haemonchus contortus* (Hc) or 20000 *Trichostrongylus colubriformis* (Tc) nematodes to examine the host gene expression response of resistant lambs to a primary challenge with pathogenic GIN. All animal experiments were approved of the CLI Animal Ethics Committee. For the Hc challenge, abomasal tissue (site of infection response; HcA) and blood (systemic response; HcB) were taken; and for the Tc challenge, jejunum tissue (site of infection response; TcG) and blood (systemic response; TcB)

Disease resistance

were taken at 0, 3, 7 and 21 days post-challenge (length of infection, L) with 4 biological replicates for each sample (4 parasite-tissue groups x 4 time points x 4 replicates = 64 arrays total). RNA extracts were hybridised to Affymetrix GeneChip® Bovine Genome arrays.

Differential Gene Expression Analyses. Microarray data were processed using Affymetrix® Microarray Suite (MAS 5.0) software during normalisation and making detection calls. After editing based on ‘Absent or Marginal’ detection calls, there were 16936 transcripts for further analyses. All statistical and bioinformatic analyses were performed using the Bioconductor package, limma (v.2.10.5). DE genes at 1% and 5% adjusted False Discovery Rate (pFDR) for a range of biological contrasts were detected. Total number of genes that were DE for each contrast was reported, along with their mode of regulation and functional annotation.

Weighted gene co-expression network analyses (WGCNA). The WGCNA distinguishes from other unweighted network construction methods in that it assigns weights to each edge by soft rather than hard thresholding. The latter could result in some borderline wherein important genes drop-off the network due to cut-off threshold values. WGCNA begins with calculation of a co-expression measure, the absolute value of the Pearson correlation (ρ_{ij}) between every pair of genes. A network is represented by an adjacency matrix, \mathbf{A} with elements a_{ij} . The \mathbf{A} matrix is created by applying the adjacency function, β , to co-expression measure, ρ_{ij} , as: $(\rho_{ij})^\beta$, which is a kernel of weighted co-expression gene networks (Zhang and Horvath, 2005). β must be specified in such a way that it corresponds to biological motivation. The adjacency matrix \mathbf{A} is then used to calculate the topological overlap measure (TOM), which reflects the relative interconnectedness between every pair of genes. The TOM matrix was used as input to the average hierarchical linkage algorithm (unsupervised) to create a dendrogram. Modules were then defined using the dynamic hybrid tree cutting algorithm of the *dynamicTreeCut* R package (Langfelder *et al.* 2008). We applied the above WGCNA method for the 16936 transcripts across 16 ($\times 4$) arrays. The co-expression networks were constructed separately for Hc and Tc parasites, and for each one of the 3 tissues (ie., one each for HcA, HcB, TcG and TcB samples using 16 microarray samples each). As there are several modules in each network, significant correlation of gene modules with the trait, L (called “module significance”) was calculated within each network to determine biological relevance with pathogenicity of Hc and Tc worms.

Systems biology of gene modules. Gene modules with high correlation with $L > 0.8$, were subjected to a systems biology approach to provide a biological context. Within such significant modules, genes whose module membership was < 0.7 (ie., those that are not hub genes) were removed. Functional enrichment analyses were carried out by linking the filtered genes with external biological metadata using EASE (<http://david.abcc.ncifcrf.gov/summary.jsp>) and GOEAST (<http://omicslab.genetics.ac.cn/GOEAST/>) to choose candidate genes / biomarkers.

Heritability of DE and module hub genes. Expression profiles of the top 10 DE genes from contrasts and selected hub genes from 4 networks were then prepared for quantitative genetic analysis. A linear model fitted (10) sire effects as random and treatment effects such as time, tissue, flock as fixed, to log-transformed gene expression values (Kadarmideen 2008). This yielded sire variances (σ^2_s) which were then used to calculate h^2 as $4 \cdot \sigma^2_s / (\sigma^2_s + \sigma^2_e)$.

RESULTS AND DISCUSSIONS

Differential Gene Expression Analyses. Comparison between Hc and Tc were made using only the blood sample arrays as the other tissues are affected by Hc or Tc, but not both. Only results

from fitting linear contrasts for L, irrespective of Hc or Tc and contrasts for interaction between parasites and L are discussed here. Figure 1 (left) shows a Venn diagram containing sets of DE genes (Benjamini and Hochberg (BH) adjusted p-value ≤ 0.05 and fold change ≥ 2) for L3 (20 genes), L7 (132 genes) and L21 (23 genes) compared to uninfected samples (L = 0). Of these, there were 2, 113 and 14 DE genes specific to L3, L7 and L21, respectively. The interaction between parasite and L3, L7 and L21 provide 0, 213 and 54 DE genes respectively (Figure 1 right). Of these interaction genes, 200 were parasite:L7 specific and 41 were parasite:L21 specific. This shows that there are a large number of genes at L7 and L21 that respond differently between Hc and Tc infected sheep. Heat plots show hierarchical clustering of expression levels of the Parasite:L7 (Figure 2). The parasite: L7 heat plot shows 2 distinct gene clusters: 1) those more up regulated in Tc L7 and 2) those more up regulated in Hc L7. This indicates the presence of unique co-expressed gene modules for different parasites across different time points; which can be identified via WGCNA.

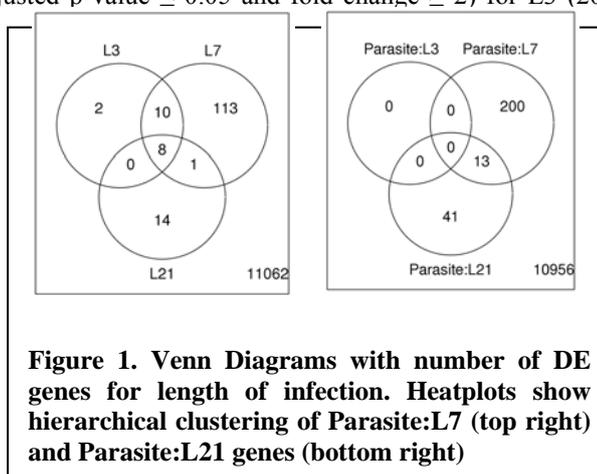


Figure 1. Venn Diagrams with number of DE genes for length of infection. Heatplots show hierarchical clustering of Parasite:L7 (top right) and Parasite:L21 genes (bottom right)

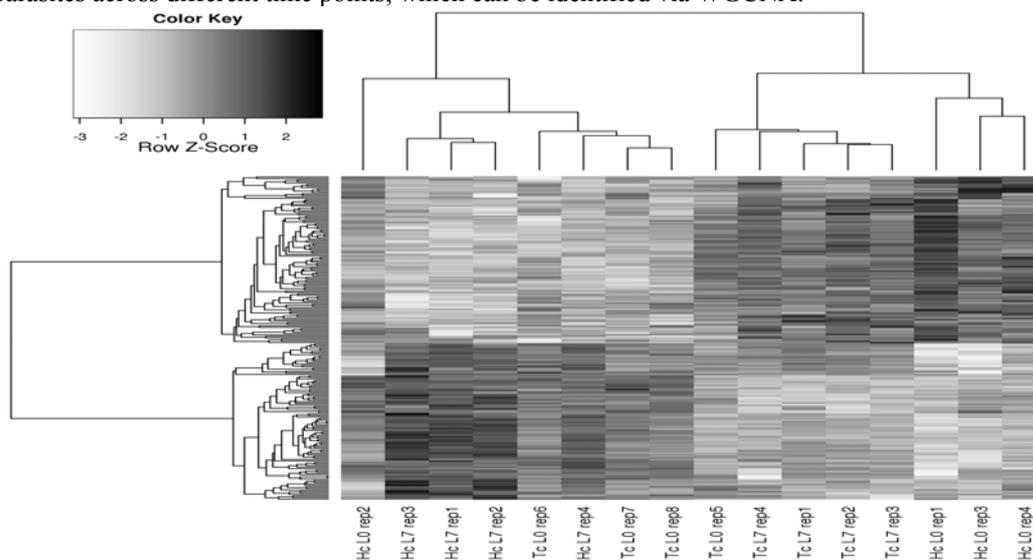


Figure 2. Heatplots show hierarchical clustering of Parasite:L7 interactions

WGCNA and Systems Biology. There were 11 different modules detected for HcA with two major modules (containing 41 and 94 genes, respectively) being the most significantly related to L. Similar results were available for other 3 networks (HcB, TcG and TcB). Systems biology results for different networks showed that they were involved in the activation of immune response, response to antigenic stimulus, cell mediated and humoral immune response, response to wounding, regulation of endocytosis, metabolic and catabolic processes, and signal transduction

Disease resistance

among several dozens of other functions. Several of these genes were represented in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways relating to regulation of autophagy, Notch signalling pathways and glycolysis. Using this systems biology approach we were able to target functionally relevant genes which are candidate drug targets, biomarkers as well as candidates for selection.

Heritability of DE and module hub genes. Heritabilities across a range of DE contrasts and hub genes from WGCNA indicated that about 50% of genes have no significant additive genetic variation in expression while the rest have the estimated h^2 ranging from 0.05 to 0.9 (with high s.e. range 0.09- 0.8). Figure 3 shows h^2 for the top 9 significant DE genes for 21 days of infection with Hc or Tc worms compared to uninfected samples (L0 v L21). Candidate genes with high h^2 of expression may be used in expression assisted selection (Kadarmideen et al., 2006).

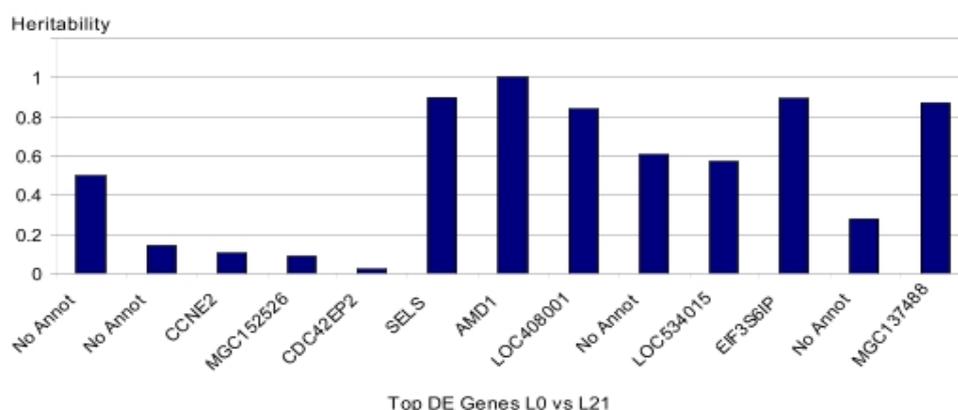


Figure 3. Heritability of Gene Expression of Top 10 DE genes after 21 days of worm infection

CONCLUSIONS

From these studies we conclude that the systems genetics method can be used to identify candidates for biologically relevant and heritable genes/biomarkers from a set of DE genes. This approach has provided a shortlist of a small number of genes for validation, with the eventual aim of developing targets to assist breeding decisions and to design bio-medical interventions.

ACKNOWLEDGMENTS

This project was financially supported by SheepGenomics, an initiative of Australian Wool Innovation Limited and Meat and Livestock Australia. The authors express appreciation to Ross Windon and Peter Hunt for their contribution to FG3 project design.

REFERENCES

- Kadarmideen, H.N., von Rohr, P. and Janss, L.L.G. (2006) *Mamm. Genome* **17**:548.
- Kadarmideen, H.N. (2008). *IET Systems Biology* **2**:423.
- Langfelder, P., Zhang, B. and Horvath, S. (2008) *Bioinformatics*. **24**:719.
- Weston, D.J., Gunter, L.E., Rogers, A. and Wullschleger, S.D. (2008). *BMC Systems biology*, **2**:16.
- Zhang, B., and Horvath, S., (2005). *Stat. Appl. in Genetics and Mol.Biol.* **4**:17.