FUNCTIONAL GENOMICS OF ESTIMATED BREEDING VALUE FOR EYE MUSCLE DEPTH IN SHEEP

L.J.A. Kogelman[1,2], N.S. Watson-Haigh[2], H.N. Kadarmideen[3], J.W. Kijas[4], T. Vuocolo[4], K. Byrne[4], H. Oddy[5], G.E. Gardner[6], C. Gondro[5] and R.L. Tellam[4]

[1]Wageningen Univ., Animal Breeding and Genetics, Wageningen, The Netherlands; [2]CSIRO Livestock Industries, Townsville QLD 4814, Australia; [3]Faculty of Life Sciences, Univ. Copenhagen, 1870 Frederiksberg C, Denmark; [4]CSIRO Livestock Industries, QLD Bioscience Precinct, St. Lucia, Brisbane, QLD 4067, Australia; [5]School of Environmental and Rural Science, Univ. New England, Armidale, NSW 2351, Australia; [6]School of Veterinary and Biomedical Sciences, Murdoch Univ., Perth, WA 6150, Australia.

SUMMARY

A weighted gene co-expression network analysis (WGCN) and a differential network analysis were applied to microarray gene expression data for skeletal muscle samples from progeny of 6 Poll Dorset sires characterised as having high or low estimated breeding values (EBVs) for Eye Muscle Depth (EMD). There was strong genetic architecture to the gene expression data. Gene network analyses identified expression modules that were enriched for several biological themes including protein catabolism, ribosome function, mRNA processing, mitochondria and muscle structural proteins. These biological pathways likely contribute to the genetics of enhanced muscling in sheep.

INTRODUCTION

Many genetic loci typically contribute to complex traits, such as enhanced muscling. Gene expression studies may provide valuable insight into the genetic architecture of this trait. By defining gene co-expression modules and correlating them to the physiological trait, a network can be constructed which may lead to the identification of biologically important pathways underpinning the genetics of the trait. The objective of the current research was to identify gene co-expression modules providing insight into the biology contributing to enhanced EMD in the progeny of Poll Dorset sires characterized as having high and low EBVs for the trait. Two different gene network strategies were employed.

MATERIALS AND METHODS

Samples and microarray analysis. Nineteen Poll Dorset sheep from 3 high muscling sires (HM; sire groups 1-3) and 21 Poll Dorset sheep from 3 low muscling sires (LM; sire groups 4-6) were used. HM or LM status was assigned based on sire EBV for EMD. Sire EBVs (range: +2.95 to -1.07 mm) were in the top 1-15% (HM sires) or 60-95% (LM sires) percentiles (all EBV accuracies > 89%). Microarray gene expression analyses (Bovine Affymetrix microarray) of skeletal muscle samples used GC-RMA to generate expression summary values (Byrne *et al*. 2010; Wu *et al.* 2004). Statistical analyses were performed using Bioconductor (Gentleman *et al*. 2004).

Weighted gene co-expression network (WGCN) construction. To efficiently analyse the data-set its size was first reduced by removal of genes with low mean expression levels ($\log_2$ <2.35) or little variation in expression (S.D. <0.01). The latter genes provide no significant information in a co-expression analysis. The number of genes was then further reduced based on connectivity (the sum of the connection strengths between a particular gene and all other genes in the network) to the 3,500 most highly connected transcripts in each of the HM and LM datasets. The union of

these 2 sets resulted in 5,394 unique genes, which was used for WGCN analysis (Langfelder and Horvath, 2008). The absolute value of the Pearson correlation between gene expression and EMD EBVs was raised to a power β to create the adjacency matrix which was then used to calculate the topological overlap measure (TOM), which shows the degree of overlap in shared neighbours between pairs of genes in the network. Gene modules were defined using the Dynamic Tree Cutting algorithm on a dendrogram created from the dissimilarity-TOM matrix. Forty two modules were initially identified.

Differential network analysis. CoXpress was used to identify genes within the 42 modules that were highly correlated in the HM state but not the LM state, and visa versa (Watson 2006). A module of genes was defined as differentially co-expressed when it was significantly different from random in one condition (HM or LM) but not the other.

Functional enrichment analysis. Functional enrichment analysis was employed to assign biological relevance to the gene network modules by using AgriGO (Zhou *et al.* 2010) and DAVID (Huang *et al.* 2009). The entire microarray was used as the statistical background. Conservative default parameters were selected. All p-values were Benjamini corrected.

RESULTS AND DISCUSSION

WGCN. Initial analyses revealed strong sire structure in the gene expression data. This indicated that there was a genetic basis to the gene expression data (data not shown). Forty two network modules were initially defined and then selected on their module correlation (MC), which is the absolute correlation between the module eigengene (a representative gene expression pattern for the module) and the EBVs for EMD. Four modules were identified based on their MC being >0.4. Genes in these modules were retained in the selected modules if: 1) their intra-modular connectivity (the connectivity of a gene in a module with respect to other genes in that module) was >0.6; 2) their intra-modular connectivity with other modules was <0.6, and; 3) the absolute correlation of the gene expression with EMD EBV was >0.5. These 4 modules were characterised as: Module A (MC = 0.54, 39 genes), Module B (MC = -0.52, 88 genes), Module C (MC = -0.52, 33 genes) and Module D (MC = -0.42, 42 genes).

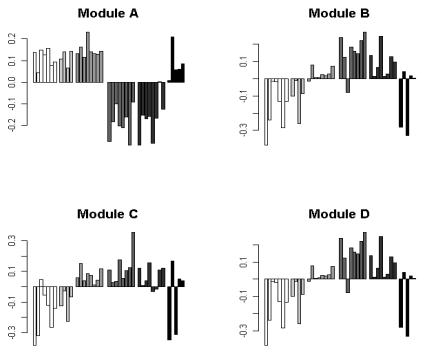

Figure 1. Expression profiles of module eigengenes for the four modules identified by WGCN. The first 3 sire groups are HM sheep and the last 3 are LM sheep. Sire groups are differentially shaded.

Figure 1 shows the expression patterns of the module eigengenes for the 4 identified modules. Conceptually an eigengene is the average expression profile for the module. In general, these patterns highlighted similarities within the HM sire groups and similarities within the LM sire groups. However, sire group 3 (animals 12-19, HM sire) behaved somewhat differently from sire groups 1 and 2 (both HM sires). This variation may be due to: (i) biological variation leading to different mechanisms promoting muscling in the sire groups, and/or; (ii) the offspring of this sire could have been atypical of its EBV status. Likewise, sire group 6 (animals 36-40, LM sire) was somewhat discordant with sire groups 4 and 5 (both LM sires).

Figure 2 shows a representative AgriGO analysis for Module B, which was strongly enriched for aspects of protein catabolism. This is also apparent from analysis of individual GO categories (not shown) as well as other analyses e.g. KEGG Pathway (p=2.16E-09) and INTERPRO Protein Domain (p=1.42E-10). The module eigengene suggests decreased proteasome activity in the HM group, which is consistent with increased muscling in HM animals. Module D was strongly enriched for functional terms representing protein synthesis at the level of Ribosome Protein Function (KEGG Pathway; p=1.22E-29) while module C was enriched for RNA Processing (KEGG pathway Splicosome; p=0.03). Module A did not achieve significance however the striking relationship between this module and sire group EBV status indicated that further analysis was warranted. Consequently, AgriGO functional analysis was performed using less stringent parameters ($p<0.1$ and ≥ 2 genes/term). The Biological Process analysis identified Muscle Sarcomere Organisation (p=0.02) and Muscle Development (p=0.02) and is therefore consistent with up-regulation of this module in progeny from high muscling sires.

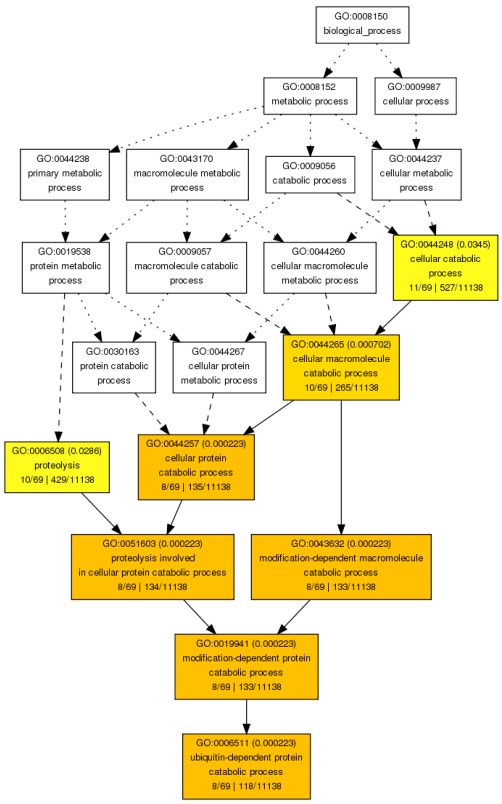

Figure 2. AgriGO gene ontology analysis. The diagram shows a representative analysis for Module B (Biological Process).

Differential network. Eight of the 42 modules were differentially co-expressed (Table 1). Four of these modules were found to be non-random in the HM group and random in the LM group (modules Diff.E, Diff.F, Diff.G and Diff.H) and visa versa for the remaining 4 modules (Diff.A, Diff.B, Diff.C and Diff.D).

Enriched biological terms were associated with three differentially co-expressed modules i.e. Ribosome and Mitochondrial Function (KEGG Pathway; p=1.3E-79 and p=0.1, respectively); RNA Processing (AgriGO analysis (p=0.04), and; Muscle Contractile Fibres (AgriGO; p=1.7E-6).

Table 1. Differentially co-expressed modules defined by coXpress. A module is differentially co-expressed when the pairwise correlations were nonrandom (p<0.05) in one condition (HM or LM) but random in the other condition (p>0.3).

Module	Number of genes	P-value [1]		Mean correlation		Mean difference correlation
		LM [2]	HM [3]	LM	HM	LM – HM
Diff.A	103	0.00	0.66	0.39	0.02	0.37
Diff.B	128	0.00	0.77	0.37	0.01	0.36
Diff.C	141	0.00	0.88	0.30	0.03	0.27
Diff.D	74	0.02	1.00	0.27	0.01	0.26
Diff.E	8	0.56	0.00	0.02	0.47	0.45
Diff.F	12	0.99	0.00	0.02	0.51	0.49
Diff.G	51	0.61	0.00	0.03	0.59	0.56
Diff.H	3	0.34	0.00	0.04	0.62	0.58

[1] P-value; [2] Low muscling; [3] High muscling

CONCLUSIONS

There was strong genetic structure in gene expression data obtained from skeletal muscle samples of progeny from sires with contrasting EBVs for EMD. Functional gene expression networks were identified that are likely to be directly contributing to the muscling EBV status of the sires. There were also indications that multiple mechanisms could be contributing to the high muscling trait. The WGCNA and differential network analyses identified specific functional pathways likely to be directly contributing to the muscling trait. These pathways included protein catabolism, protein biosynthesis at the level of ribosome function, myofibril function, mitochondrial function and mRNA processing. The future challenge is to link these pathways to genetic polymorphisms in specific genes.

ACKNOWLEDGEMENT

This research was supported by the MLA/AWI SheepGenomics Initiative.

REFERENCES

Byrne K., Vuocolo T., Gondro C., White J.D., Cockett N.E., Hadfield T., Bidwell C.A., Waddell J.N. and Tellam R.L. (2010) *BMC Genomics,* **11**: 378.

Gentleman R., Carey V. and Bates D. (2004) *Genome Biology*. **5**: R80.

Huang D.W., Sherman B.T. and Lempicki R.A. (2009) *Nature Protocols,* **4**: 44.

Langfelder P. and Horvath S. (2008) *BMC Bioinformatics.* **9**: 559.

Watson M. (2006) *BMC Bioinformatics*. **7**: 509.

Wu Z., Irizarry R.A., Gentleman R., Martinez Murillo F., Spencer F. (2004) *J. Amer. Stat. Assoc.*, **99**: 909.

Zhou D., Xin Z., Yi L., Zhenhai Z. And Zhen S. (2010) *Nucleic Acids Research,* **38**: W64.