Genetics of Global Gene Expression Patterns and Gene Networks Affecting Muscling in Sheep

H. N. Kadarmideen [*], N. S. Watson-Haigh[*], J. W. Kijas[†], T. Vuocolo[†], K. Byrne[†], C. Gondro[‡], V. H. Oddy[‡], G. E. Gardner[§] and R. L. Tellam[†]

Introduction

The Sheep Genomics research program in Australia has undertaken a range of investigations to identify genes and DNA markers contributing to increase in size and altered distribution of muscling in sheep breeds (Oddy et al. 2007). The main objectives of this study were: (i) to discover key genes and pathways underlying muscling traits by expression profiling of skeletal muscle from sheep born to industry sires with high and low genetic merit (Estimated Breeding Values or EBVs) for eye muscle depth (EMD), and analysing the resulting expression data by advanced bioinformatics / systems genetics approaches, and; (ii) to link high throughput genetic information on these animals (single nucleotide polymorphisms or SNPs) with gene expression profiles of highly differentially expressed genes and highly connected regulatory (hub) genes identified in the first objective. With these approaches, we can generate a comprehensive understanding of the molecular mechanisms underpinning muscle hypertrophy phenotypes in sheep and identify potential candidate genes that can be used in breeding programs to improve meat quantity.

Materials and methods

Sheep longissimus dorsi (LD) skeletal muscle samples were collected from 40 progeny belonging to six Poll Dorset industry sires, with 3 sires having high muscling and 3 sires having low muscling EBV for EMD. Transcription profiling of these samples was undertaken using the Bovine Affymetrix GeneChip® containing 24,128 probe sets (Vuocolo et al., 2007). All animals were also genotyped using ovine Illumina SNP 50k arrays. The gene expression data were subject to stringent hybridisation quality control and were
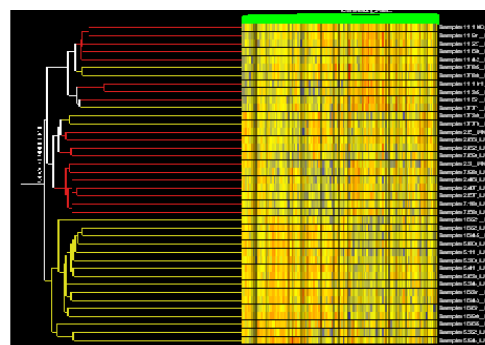

Figure 1: Unsupervised clustering showing distinct gene expression patterns between high EBV (n=19; red) and low EBV (n=21; yellow) sire groups.

[*] CSIRO Livestock Industries, Davies Laboratories, Aitkenvale, Townsville QLD 4814, Australia

[†] CSIRO Livestock Industries, Queensland Bioscience Precinct, St. Lucia, Brisbane, QLD 4067, Australia

[‡] School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia

[§] School of Veterinary and Biomedical Sciences, Murdoch University, Murdoch, WA 6150, Australia

processed using GC-RMA (Wu et al. 2004). The results of preliminary unsupervised clustering of the normalised data are shown in figure 1; it clearly depicted distinct clusters of high and low EBV sire groups. LIMMA in the R statistical framework was used to identify differentially expressed (DE) genes between the two sire groups.

Weighted Gene Co-Expression Network Analysis (WGCNA) of Zhang and Horvath (2005) begins with calculation of the absolute value of the Pearson correlation (ρ_{ij}) between two genes and the adjacency function, a_{ij}, among two genes as: $(\rho_{ij})^{\beta}$. In contrast to other unweighted networks, scale free topology criterion was used here to select the appropriate soft threshold,β which was 7, 6 and 7, for the Entire, High and Low EBV datasets, respectively. The main advantage of WGCNA is that it is robust to the choice of the parameter, β. WGCNA analyses were restricted to the top 3,000 most connected genes for computational reasons. The topological overlap matrix was used as an input to the average hierarchal linkage algorithm to detect gene modules by applying the Dynamic Tree Cut method (Langfelder et al. 2008). The strength of correlation of the ith gene profile with genetic merit (here EBVs for EMD) is called gene significance for the ith gene (GS_i) which indicates the directional effect of the ith gene on EBVs. In order to identify genes of highest relevance to EBV, GS_i was used; in addition, they were used to identify gene modules of highest mean relationship to EBVs. Module membership for the ith gene (MM_i) is a correlation between the ith gene profile and the eigengene of a given module. Genes having high MM_i are also potential hubs. By combing GS_i with MM_i, we set up the criteria for a gene to be candidate biomarker as $MM_i > 0.7$ and $GS_i > 0.8$. These criteria identified major regulators of genes in the network that are also highly significantly related to muscle EBVs.

Identifying causal polymorphisms using genetical genomics. Genome-wide association of genetic (SNP) markers with gene expression traits (e-traits) can identify two types of SNPs affecting e-traits (eSNPs): cis-acting and trans-acting SNP (Kadarmideen et al. 2006; Kadarmideen 2008). Instead of genome-wide eSNP mapping, only e-traits from DE and hub genes were considered. We assigned sheep genes to bovine Affymetrix probe sets indirectly via bovine Ref Seq mapping to the OARV1.0 genome assembly. We were then able to identify cis-SNPs, defined as those SNPs within ±5 Mb of the ovine gene, for the set of DE or hub gene e-traits and perform genome-specific association analyses (GSAS) using the SNPassoc package (Gonzalez et al. 2007).
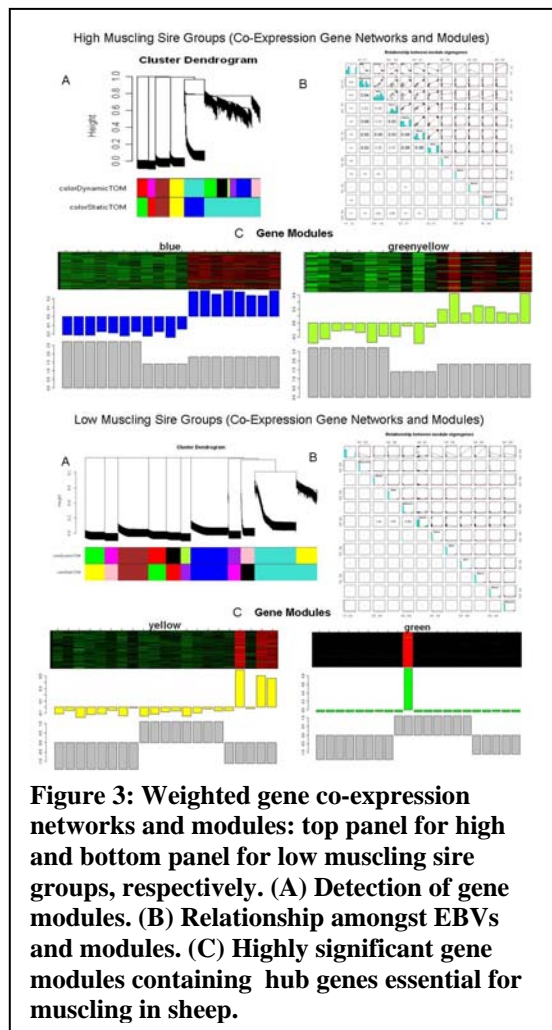
Results and discussion

Differential gene expression. Figure 2 shows the heatplots of all 2,058 statistically significant DE probe-sets (BH adjusted p value of ≤ 0.05 for the High vs. Low EBV contrast). The groups of genes with similar expression profiles in rows

Figure 2: Heatplots of 2,058 statistically significant DE probe-sets for High vs. Low EBV EMD sire groups
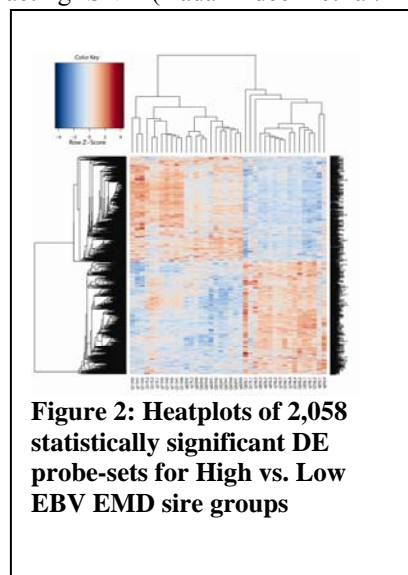
are clustered and plotted against clusters of sire samples with similar expression profiles. As expected with DE genes, the clear distinction between the two EBV groups is seen by visualizing the group of up-regulated genes in the high EBV group which are down-regulated in low EBV group and vice versa. This demonstrates the genetic basis of gene expression in eye muscle depth phenotypes. Gene ontology (GO) enrichment analysis was conducted using GOEAST (Zheng and Wang 2008) with p-value ≤0.05 and log odds ratio ≥3). The analysis showed that these genes were highly statistically enriched for *metabolism*, *anatomical structure formation involved in morphogenesis*, *limb and skeletal muscle development*, *proteolysis*, *protein binding*, and *monooxygenase activity*.

Gene networks and modules in muscle development. Three different gene co-expression networks were derived from gene expression levels i.e. (i) the entire population, (ii) high EBV sire group and, (iii) low EBV sire group. In the whole dataset, there was a large number of genes with gene connectivity or edges with k = 50-70 and a small number of genes with k > 200. Those genes with k > 200 and high correlation with EBVs are considered as major hub acting genes. However, when networks were constructed within the high or low EBV sire groups, there was a large number of genes with k > 200, as expected, due to genes that are involved in the same pathways and cellular functions contributing to the increase or decrease in muscle mass, respectively. Figure 3 shows hierarchical clustering applied to EBV samples and use of the Dynamic Tree Cut method. For the high EBV group, there were 11 modules containing 64 - 491 genes. Of these, 10 genes met our criteria (based on combined score of GS_i with MM_i) of being a "candidate biomarker" for high muscling. Similarly, for the low EBV group, there were also 11 different modules containing 130 - 482 genes. Of these, 25 genes met our criteria for being a "candidate biomarker" gene for low muscling. GOEAST analyses (p-value ≤0.05; log odds

Figure 3: Weighted gene co-expression networks and modules: top panel for high and bottom panel for low muscling sire groups, respectively. (A) Detection of gene modules. (B) Relationship amongst EBVs and modules. (C) Highly significant gene modules containing hub genes essential for muscling in sheep.

ratio ≥3) of the hub genes revealed enrichment for the following terms: *ensheathment of neurons, myelination, myelin sheath, heart, liver and thyroid gland development,* and *regulation of gene-specific transcription, protein-DNA complex assembly, DNA packaging, metabolic process, catabolic process and SMAD protein nuclear translocation.*

Identification of cis-acting SNPs for biomarker profiles: We limited our analyses to cis-SNPs flanking the DE and hub genes of interest within ±5 Mbp of the ovine gene locations. The number of SNP within any 5 Mbp region of interest ranged from 0 to 200. The GSAS of e-traits revealed several genomic regions which have direct effects on transcript abundance (at 1% FDR) and potentially contain causal polymorphisms for increased eye muscle depth in sheep. Further analyses are in progress for detecting cis-eSNPs and trans-eSNPs by varying bin sizes and construction of underlying genetic interaction networks.

Conclusions

We have reported the genetic basis for global gene expression traits affecting eye muscle depth in sheep. This was confirmed by distinct clustering patterns of e-traits across sire groups, differential gene expression patterns between high and low EBV groups and weighted gene expression networks containing hub genes that are predictive of muscling genotypes/phenotypes. The functional annotations and significant GO terms associated with genes identified in these analyses were involved in a variety of pathways, biological, cellular and molecular functions directly relevant to muscling. Further, targeted genome-specific eSNP association mapping revealed causal genetic networks directly underlying muscling phenotypes in sheep. To our knowledge this is the first report of a 'systems genetics' approach for a muscling trait in sheep. The study clearly highlights the strong genetic contribution to gene expression.

Acknowledgments

References

Gonzalez, J. R., Armengol, L., Sole, X., *et al*. (2007). *Bioinformatics,* 23: 644-645

Kadarmideen, H.N., Von Rohr, P. Janss, L.L.G. (2006). *Mamm.Genome.,* 17: 548-564.

Kadarmideen, H.N. (2008). *IET Systems Biol.* 2: 423-441.

Langfelder, P., Zhang, B. and Horvath, S., (2008). *Bioinformatics*, 24: 719-720.

Oddy, V. H., Dalrymple, B.P., McEwan, J.C., *et al*. (2007). *Proc.17th conf. Assoc. Advancement of Anim. Breed. Genet.* 411-417

Vuocolo, T., K. Byrne, et al. (2007). *Physiol. Genomics,* 28: 253-72.

Wu, Z., Irizarry, R.A., Gentleman, R. *et al.* (2004). *J. Amer. Stat.Assoc.* 99: 909-917.

Zhang, B. and Horvath, S. (2005). *Statistical Appl. Genet. and Mol.Biol.,* 4 Art. 17.

Zheng, Q. and Wang, X.J. (2008). *Nucleic Acids Res.*, 36:W358-W363.