

WEIGHTED CO-EXPRESSION NETWORKS SHED LIGHT ON THE MOLECULAR MECHANISM OF ACTION OF METYRAPONE ON WOOL FOLLICLE DEVELOPMENT

N.S. Watson-Haigh¹, H.N. Kadarmideen², M. McDowall³, G.S. Natrass⁴, H.A. McGrice³ and P.I. Hynd³

¹ CSIRO Livestock Industries, St Lucia ² CSIRO Livestock Industries, Rockhampton; ³ The University of Adelaide and ⁴ South Australian Research & Development Institute (SARDI)

SUMMARY

The density of Merino wool follicles is established early in fetal development. This commercially important trait dictates wool fibre diameter which is the key driver of the price paid for wool. Merino lambs exposed to metyrapone (an inhibitor of cortisol synthesis) *in utero* show a lifetime alteration in wool growth parameters. Microarray data from a metyrapone treatment experiment were analysed within a systems biology framework using Weighted Gene Co-expression Network Analysis (WGCNA). Four networks were created to determine those genes involved in metyrapone mediated improvement of wool growth parameters. Using the WGCNA approach, we were able to detect co-expressed gene modules associated with metyrapone treatment. Gene ontology enrichment analysis of the genes comprising these modules identified a Bone Morphogenetic Protein (BMP4), a ligand known to be involved in hair/wool follicle development and expressed at the time of branching of secondary-derived follicles in Merino sheep.

INTRODUCTION

Merino sheep have a characteristically high follicle density of up to 60/mm² and a total of 10-100 million follicles compared to the estimated 5 million follicles in human skin. Primary follicles are the first to form, followed by secondary follicles and then secondary-derived follicles that branch from the secondary follicles. It is this high concentration of secondary-derived follicles which is a distinctive feature of Merino sheep. Branching of the secondary original follicles is essential in the Merino fleece as it is the major source of fine fibres (Hardy and Lyne 1956; Adelson *et al.* 2004). More detail is provided in a recent review (Rogers 2006). Fleece density, quality and length are important commercial traits, so improvements in these are beneficial.

A previous study showed that lambs exposed to metyrapone, an inhibitor of cortisol synthesis, *in utero* between day 55 and 65 of gestation have been shown to possess lifetime improvements in their fleeces (in prep.). The identification of genes responsible for this fleece improvement would facilitate the development of other pharmaceutical intervention strategies to improve wool growth. A microarray experiment was performed on foetal skin samples in an attempt to determine those genes involved in metyrapone mediated improvement of wool parameters. We follow a general framework for constructing gene co-expression networks (Zhang and Horvath 2005) and used the WGCNA R package (Langfelder and Horvath 2008). For details of network concepts and terminology used see (Langfelder and Horvath 2008; Dong and Horvath 2007).

MATERIALS AND METHODS

Experimental Design. Twenty time-mated (synchronised with progesterone sponges and then artificially inseminated) pregnant Merino ewes were allocated to 4 equally sized treatment groups receiving daily intramuscular injections of a control (5ml of 0.6 M tartaric acid in 5% Tween 20) or metyrapone (5 ml of 1.76g metyrapone in 0.6 M tartaric acid in 5% Tween 20) between day 55

and 65 of gestation. 2cm midside foetal skin samples were collected at two time points (Table 1). Samples were snap frozen in liquid nitrogen prior to RNA extraction.

Microarray Data. The RNA from 16 foetal skin samples was hybridised to Affymetrix GeneChip® Bovine Genome Array. Data were subjected to quality control procedures to ensure that: 1) hybridisations were uniform as assessed by pseudo-array images produced by the Bioconductor (Gentleman *et al.* 2004) affyPLM package and 2) RNA was not degraded as assessed by the Bioconductor affy package (Gautier *et al.* 2004).

Gene expression values were generated using the median-polish summarisation method following GC-RMA background correction of PM probes and quantile normalisation.

Table 1. Experimental design showing treatment groups

Group	Treatment	Treatment period (day of gestation)	Sample collected (day of gestation)	Number of single pregnancies
1	Control	55-59	60	4
2	Metyrapone	55-59	60	5
3	Control	55-65	67	4
4	Metyrapone	55-65	67	3

Weighted Network Construction. Four networks were constructed using sample groups as follows: network A) group 1 and 3; network B) group 2 and 4; network C) group 1 and 2; and network D) group 3 and 4. For each network non-changing genes with low mean levels of expression ($< 10^{0.805}$) and low variance ($< 10^{-6}$) across all the arrays were excluded to reduce computational complexity in later steps. The power adjacency function was applied to the co-expression measurement, the absolute Pearson correlation coefficient, to derive the adjacency matrix: $a_{ij} = |\text{cor}(x_i, x_j)|$. The value of the power function exponent, β , was chosen based on the accuracy of the model fit, R^2 (Zhang and Horvath 2005). We chose β in the interval [0,30] which maximised R^2 while maintaining a high level of mean connectivity. Using the gene connectivity measure, we constructed a network based on the top 8,000 most connected genes.

Modules were defined for the top 8,000 most connected genes using the dynamic hybrid tree cutting algorithm of the dynamicTreeCut R package (Langfelder *et al.* 2008). Interesting modules were defined as those which had a high Pearson correlation (≥ 0.8) with the day of gestation.

Gene Ontology Enrichment Analysis. Gene lists were validated for biological relevance using GOEAST (Zheng and Wang 2008), a gene ontology (GO) enrichment analysis tool. GO enrichment tools are used to identify GO terms which are statistically overrepresented in a list of genes and provide a means to identify biological functions and processes at play.

RESULTS AND DISCUSSION

Microarray Data. Pseudo-array images showed that several of the arrays contained small hybridisation artefacts (data not shown) but nothing so unusual as to warrant their removal (Bolstad *et al.* 2005). The slopes and profiles of our RNA degradation plot indicated no major issues with RNA quality (data not shown).

Weighted Co-Expression Network Analysis. Figure 1 shows a Topological Overlap Measure (TOM) plot for network B with module membership indicated below and to the right of the dendrograms. It shows a high level of topological overlap between genes within a module

(indicated by the dark squares along the diagonal), with only a small amount of inter-module overlap (indicated by mostly pale off-diagonal regions).

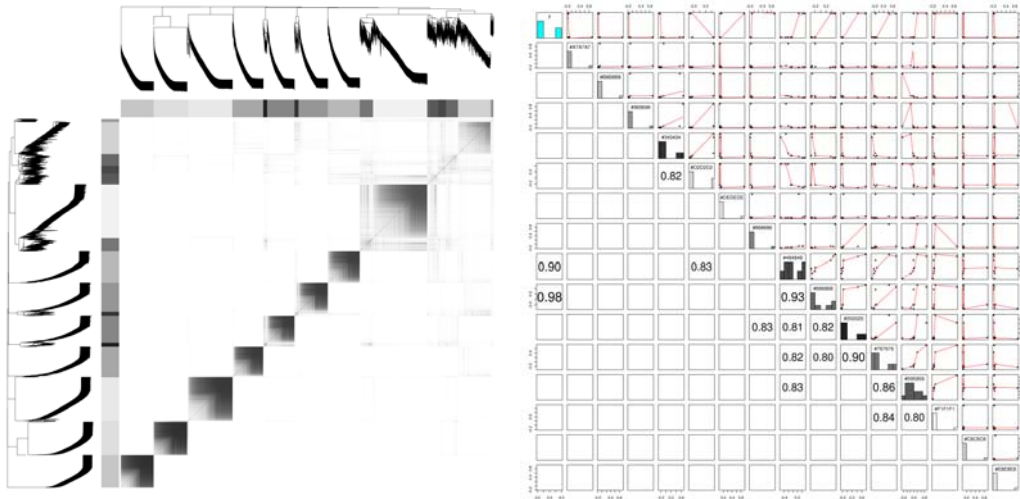


Figure 1. WGCNA results for network B. Left) TOM plot with module membership indicated next to dendrograms. Right) Pairwise correlations plot between module eigengenes and experimental parameter, day (1st column/row). Correlation values ≥ 0.8 are shown.

Effective time-mating resulted in no modules highly correlated with experimental parameter day, as expected, in Network A. On the other hand, network B provided 2 modules, “#484848” and “#686868” with correlations of 0.9 and 0.98 with experimental parameter, day (Figure 1). This indicates differential gene expression between day 60 and 67 in metyrapone samples which is not seen in controls. Module eigengenes, the first principle component of a module, for the two modules show higher levels of gene expression in the metyrapone day 67 samples compared to the metyrapone day 60 samples (Figure 2).

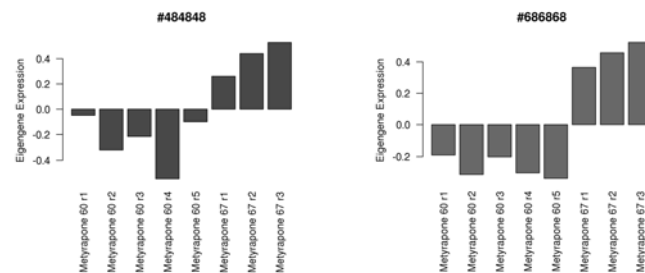


Figure 2. Eigengene expression for modules “#484848” and “#686868”.

Gene Ontology Enrichment Analysis. The “#686868” module (267 genes) showed enrichment for 58 Biological Process GO terms (a selection shown in Table 2), 33 Molecular Function GO terms and 9 Cellular Component GO terms. Among several interesting tube/patterning terms were GO:0048754 (branching morphogenesis of a tube) and GO:0030509 (BMP signalling pathway) which comprise of Bone Morphogenetic Protein 4 (BMP4) and Chemokine (C-X-C motif) Receptor 4 genes. BMPs are members of the transforming growth factor- β (TGF- β) superfamily of ligands, with important roles in a myriad of biological activities (Waite and Eng 2003; Shi and Massagué 2003). They have been shown to be involved in the regulation of hair follicle initiation

and development, including the branching of secondary follicles (Menzies *et al.* 2009) which is important in Merino sheep.

Table 2. A selection of Biological Process GO terms enriched in the “#686868” module

GOID	Definition	No. of genes	P-value
GO:0051056	regulation of small GTPase mediated signal transduction	9	0.004
GO:0007389	pattern specification process	3	0.018
GO:0010646	regulation of cell communication	11	0.018
GO:0001763	morphogenesis of a branching structure	2	0.028
GO:0048754	branching morphogenesis of a tube	2	0.028
GO:0030509	BMP signaling pathway	1	0.051
GO:0001569	patterning of blood vessels	1	0.051
GO:0009880	embryonic pattern specification	1	0.051
GO:0035239	tube morphogenesis	2	0.056
GO:0009799	determination of symmetry	1	0.070

CONCLUSIONS

Metyrapone treatment of pregnant ewes between 55 and 65 days of gestation result in pronounced differences in gene co-expression patterns. Using a WGCNA approach we were able to recover modules, which are highly correlated to experimental parameters of interest, which contain several hundred genes. Using a GO enrichment analysis we showed that these modules contain genes known to be involved in the regulation of hair follicle initiation and development. This information can now be used to refine the timing of drug administration or used to identify compounds with similar or more potent effects on wool growth.

ACKNOWLEDGMENTS

The project was financially supported by Sheep Genomics, an initiative of Australian Wool Innovation Limited and Meat and Livestock Australia. The Systems Biology Project is supported by CSIRO OCE Postdoctoral Fellowship and Transformational Biology Capability Platform.

REFERENCES

- Adelson, D.L., Cam, G.R., DeSilva, U. and Franklin, I.R. (2004) *Genomics*. **83**:95
- Bolstad, B.M., Collin, F., Brettschneider, J., Simpson, K. (2005) In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health)*.
- Dong, J. and Horvath, S. (2007) *BMC Syst. Biol.* **1**:24.
- Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A. (2004) *Bioinformatics*. **20**:307
- Gentleman, R.C., Carey, V.J., Bates, D.M. and Bolstad, B. (2004) *Genome Biology*. **5**:R80.
- Hardy, M. and Lyne, A. (1956) *Aust. J. Biol. Sci.* **6**:423
- Langfelder, P. and Horvath, S. (2008) *BMC Bioinformatics*. **9**:559.
- Langfelder, P., Zhang, B. and Horvath, S. (2008) *Bioinformatics*. **24**:719.
- Menzies, M., Stockwell, S., Brownlee, A. and Cam, G. (2009) *Exp. Dermatol.*:In Press.
- Rogers, G.E. (2006) *Exp. Dermatol.* **15**:931
- Shi, Y. and Massagué, J. (2003) *Cell*. **113**:685
- Waite, K.A. and Eng, C. (2003) *Nat. Rev. Genet.* **4**:763
- Zhang, B. and Horvath, S. (2005) *Stat. Appl. Genet. Mol. Biol.* **4**:Article 17.
- Zheng, Q. and Wang, X. (2008) *Nucl. Acids Res.* **36**(suppl_2):W358.