

CONNECTING GENE EXPRESSION AND PHENOTYPE – PRELIMINARY RESULTS FROM RNA SEQUENCING OF 150 LAMBS

S. Bolormaa¹, R. Behrendt¹, M.I. Knight¹, B.A. Mason¹, C.P. Prowse-Wilkins¹, L. Slocombe¹, C.J. Vander Jagt¹, B.J. Hayes^{1,2}, A.J. Chamberlain¹ and H.D. Daetwyler^{1,2}

¹ Department of Economic Development, Jobs, Transport and Resources, Victoria, Australia

² School of Applied Systems Biology, La Trobe University, Victoria, Australia

SUMMARY

Gene expression analysis can aid in prioritising regions or classes of variants for genomic prediction and they increase our understanding of quantitative traits. The number of reads from RNA sequencing that align to a gene can be used to quantify gene expression. We sampled liver and muscle tissues of 150 lambs at slaughter. Their dams had been managed to high, medium, and low body condition scores (BCS) during mid-to-late pregnancy and the lambs were fed three different finishing diets. Differential expression of genes (DEG) was investigated contrasting tissue, BCS, lamb diets, other treatment differences, as well as high and low lamb carcass eye muscle width (CEMW). A large number of DEG were identified between tissues, but only the low versus high BCS comparison resulted in DEG for treatments. DEG were also found when we contrasted high and low CEMW. A strong trend toward down regulation was observed in all tests, except in BCS where all DEG were overexpressed in fatter ewes.

INTRODUCTION

Gene expression refers to the production of RNA transcripts which ultimately result in a gene product. Genes may be expressed only at certain time points in the animal life cycle and expression may be tissue specific. Linking gene expression to measured phenotypes or even to management strategies may help us optimise the production system. Using expression analysis to prioritise candidate genes together with marker or DNA sequence data could ultimately increase genomic prediction accuracies for key traits.

The rate at which genes are expressed can be investigated by measuring the abundance of RNA transcripts in a tissue. Sequencing RNA is an attractive option for this analysis because the number of reads that align to an annotated gene is a reliable measure of abundance (Mortazavi et al. 2008). The analysis of RNA sequence data has a number of uses. At the most basic level, it results in a set of genes expressed in the starting material (i.e. tissue, cell line, etc). Additionally, contrasts can be performed within and across tissues, depending on the variety of tissues sampled or the number of treatments or other phenotypes measured in a study (e.g. liver versus muscle, high versus low body weight). Contrasts across animals may result in sets of differentially expressed genes (DEG), whose degree of difference can be measured via statistical tests and by the Log₂ fold change between classes.

Here we present preliminary results from a differential expression analysis using RNA sequence data on liver and muscle tissues of 150 lambs. Our initial aim was to investigate whether dam body condition score (BCS) during gestation and lamb finishing diets affect gene expression, potentially linking these treatments to genetics.

MATERIALS AND METHODS

The experimental design involved 648 pregnant ewes (Merino x Border Leicester and Maternal/Coopworth Composites) that resulted from artificial insemination to 5 Polled Dorset (PD) and 4 White Suffolk (WS) sires (that had been selected for high, medium, or low lean meat yield genomic breeding values). The core design was based on 3 BCS - ewe live-weight change -

nutritional treatments during gestation that were targeted in ewes reaching BCS2.5, BCS3.0 and BCS3.5 by lambing. The ewe BCS was managed to a flock average of BCS3.0 from prior to artificial insemination and until ultrasound scanning at day 50 to confirm pregnancy. Following scanning, ewes were distributed amongst 18 management cells and nutrition was managed by allocation of feed-on-offer during the last 2 trimesters to achieve the condition score targets. Ewes were maintained within one management group and were given ad libitum access to pasture after lambing to weaning.

Lambs were weaned at 12-13 weeks of age. Male lambs were backgrounded in pasture prior to the finishing with 3 different feedlot diets: 1) high protein, moderate energy, 2) high energy, moderate protein, and 3) high protein and energy. All male progeny (N=436) were slaughtered in three blocks at a commercial abattoir. Full-bone out was performed on 100 lambs, with a partial bone out on the remainder (Pearce et al. 2010). Additionally, a large number of carcass phenotypes were recorded. Here we will present early results on only eye muscle width (CEMW).

Liver and loin muscle tissue samples from slaughter lambs were taken within 10 minutes of death to determine the influence of the experimental treatments on aspects of meat quality, lean meat yield, and gene expression. Samples were flash frozen in liquid nitrogen and stored at -80°C. 150 lambs (100 full and 50 partial bone outs) randomly selected across all dam lamb nutritional treatments, birth types, breeds and sires, were selected for RNA sequencing. Frozen muscle and liver tissues were ground using the Genogrinder2010 (SPEX). Ground tissue was homogenised in Trizol® (Life Technologies™) and RNA extracted using the Trizol® Plus RNA extraction kit (Life Technologies™). Individually barcoded strand specific RNA sequence libraries were produced using the SureSelect Strand Specific RNA Library Prep Kit (Agilent Technologies). The 300 libraries were combined into one of four pools and 120 bp paired-end sequencing performed on a HiSeq2000 genome analyser (Illumina Inc) with the aim of producing 40 million paired reads per library. Fastq files were called using CASAVA v1.8 (Illumina Inc).

Fastq files were quality controlled using quadtrim (<https://bitbucket.org/arobinson/quadtrim>) as follows. Low quality bases were trimmed from read ends (phred score < 20). Reads were removed if they: failed the chastity filter, contained more than three Ns, had a mean base quality score of <20, or were < 50% of original read length. Filtered reads were aligned to the SheepOAR3.1 assembly using the Ensembl v78 SheepOAR3.1 (Jiang et al. 2014) annotation file containing 25,202 genes using Tophat2 (Trapnell et al. 2012). A gene by tissue count matrix containing all animal results was generated using the python package HTSeq (Anders et al. 2015). The R program DESeq (Anders and Huber 2010) was used for DEG analysis of tissues, traits and treatments, evaluating whether read counts per gene were significantly different when testing multiple samples belonging to two classes (e.g. liver and muscle) based on a negative binomial regression test. Counts were normalised for mean read depth per sample. The model fitted contrasted two classes across both tissues and within liver or muscle. The following pairwise contrasts were tested: liver and muscle across all samples, liver and muscle in PD and WS sire groups, sire breed within either tissue, 3 BCS levels of dam at gestation, 3 lamb finishing diets, 3 kill days, 3 sire LMY ASBVs, single versus multiple births, and extremes of CEMW (top 10% versus bottom 10%). Genes were reported as DE if their false discovery rate percentage (FDR%) was below 40% at a p value of 0.001 ($FDR\% = (25202 * 0.001 [\text{number of DEG}]^{-1}) * 100$).

RESULTS AND DISCUSSION

An average of 70 million reads (range 19-333 million) were generated for each of 298 samples after filtering for chastity, which is substantially more than the target 40 million. Quality control reduced the reads to 87% of which a mean of 85% of paired reads (range 78-90%) were aligned to the assembly, which was comparable to other sheep studies (Chen *et al.* 2015).

The normalised counts matrix was then used to test contrasts. The comparison of liver and

muscle across all samples yielded 10,116 DEG, which was the largest number in the study (Table 1). It is expected that different genes would be expressed in discrete tissue types. The animals were approximately balanced between PD and WS sire breeds. Analysis of liver versus muscle contrasts within PD and WS revealed 7,990 and 7,060 DEG in each sire breed group, respectively. The DEG overlap between the sire groups was 6,559, indicating that in smaller and independent samples the same genes are expressed. Single breed analysis resulted in negligible additional DEG when compared with the analysis across all animals (data not shown). Contrasts of sire breed resulted in a non-significant small number of DEG. The FDR% of liver versus muscle comparisons was always < 1%. Filtering DEG by increasing Log₂ fold change stringency gradually reduced the number of significant genes. However, a large number resulted in Log₂ fold changes of => 10 (e.g. all samples liver vs muscle, 621). The majority of these DEGs exhibited no expression in one tissue and strong expression in the other. This may be suspicious with small sample size, but in our case of testing 149 lambs it is unlikely that all libraries would miss true transcripts in these genes. DEG in muscle were much more likely to be down regulated (~80-90%) than up regulated, and the percentage of down regulation decreased slightly with increasing Log₂ fold changes before plateauing at ~80% (Figure 1).

Table 1. Number of DEGs at progressive Log₂ Fold Change thresholds and FDR% at Log₂ Fold threshold 1. BCS2.5vs3.5 of ewes refers to low and high body condition during pregnancy.

<i>LIVER vs MUSCLE</i>	FDR%	Absolute Log ₂ Fold Change									
		1	2	3	4	5	6	7	8	9	10+
All Samples	0.3	10116	6553	3751	2350	1633	1205	972	813	683	621
Polled Dorset	0.3	7990	5565	3336	2156	1506	1151	929	784	683	627
White Suffolk	0.4	7060	5820	3443	2186	1528	1155	933	786	690	627
Overlap PD-WS	-	6559									
<i>TREATMENTS</i>	FDR%	Absolute Log ₂ Fold Change									
TEST WITHIN		1	2	3	4	5	6	7	8	9	10+
LIVER											
BCS2.5vs3.5 of ewes	37.6	67	11	2	1	1	1	1	1	1	1
<i>TRAITS</i>	FDR%	Absolute Log ₂ Fold Change									
BOTH TISSUES		1	2	3	4	5	6	7	8	9	10+
CEMW-top10vsbot10%	4.1	616	481	67	21	8	6	6	6	6	6
LIVER											
CEMW-top10vsbot10%	7.4	340	241	63	19	6	5	5	5	5	5

The contrasts across treatments detected fewer DEG. In fact, no significant DEGs were found for treatments in muscle. The negative control of kill day also revealed no effect on expression levels, which increases the likelihood that systematic problems during slaughter and related processing were avoided. The only treatment that exhibited DEG was BCS, where the contrast of low (2.5) versus high (3.5) BCS of the ewe resulted in 67 significant genes in liver of the lambs (Table 1). Interestingly, all DEG were up regulated in BCS3.5 when compared to BCS2.5, potentially linking an increase in body condition of ewes during gestation to up regulation of genes in their offspring at slaughter. This is different to all other expression directions observed in this study.

The contrast in extreme lamb CEMW phenotypes resulted in between 616 and 340 DEG across tissues and within liver, respectively (Table 1). The number of DEGs met our FDR% threshold of 40% up to and including Log₂ fold change 3, which was a stronger signal than for ewe body

condition disparity. In the CEMW tests, all DEG were down regulated, which indicates that a reduction in trait phenotype could be shown to be associated with down regulation of gene expression (Figure 1).

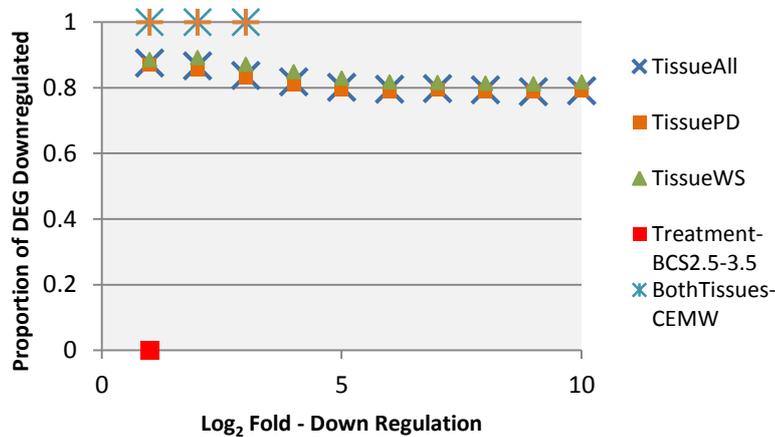


Figure 1. Proportion of down regulated genes at progressing thresholds of Log₂ fold change for all significant contrasts, where All is all samples, BCS is body condition score, and CEMW are eye muscle width, respectively. Only points with FDR<40% shown.

KEGG and GO analysis of DEG identified two major gene pathway groups related to fat/cholesterol (liver) and muscle fibre. Several genes (*APOA-1*, *-2*, *-4*, *-5*, *APOC3*, and *APOF/CAV3*) were involved in cellular cholesterol and phospholipid efflux, homeostasis and transport (Bonferroni $P < 5.5 \times 10^{-5}$), whereas 5 genes (*ACT*, *ACTG1*, *MYL*, *BMP10*, and *CAV3*) were weakly linked to skeletal muscle differentiation (Bonferroni $P < 8.8 \times 10^{-1}$). The DEG identified when contrasting dam BCS contained candidate genes involved in stress response and embryonic development (e.g. *KLHDC10*, *BMP4*, *MAPKAPK3*, *ABL1*)

The preliminary analysis of this large RNA sequencing dataset has revealed widespread DE of genes between tissues. It has also connected ewe body condition during gestation to liver gene expression in their lambs. Additionally, differences in CEMW were shown to be mirrored in gene expression patterns. Further analyses on other phenotypic traits and allele specific expression will be performed. The outcomes of this study will contribute towards more precise annotation of the sheep reference genome and it will aid in prioritising genetic markers for genomic prediction.

ACKNOWLEDGMENTS

The authors would like to thank Meat and Livestock Australia for funding the Lamb Foundations Project, and the Growing Food and Fibre Initiative of the Victorian State Government for funding the expression analysis. We are grateful to all staff involved in animal care and the collection of tissues and phenotypes at and after slaughter.

REFERENCES

- Anders S, Huber W (2010) *Genome Biology* **11**:R106
 Anders S, Pyl PT, Huber W (2015) *Bioinformatics* **31**:166.
 Chen HY, Shen H, Jia B, et al (2015) *PLoS ONE* **10**:e0120170
 Jiang Y, Xie M, Chen W, et al (2014) *Science* **344**:1168.
 Mortazavi A, Williams BA, McCue K et al (2008) *Nat Meth* **5**:621.
 Pearce KL, van de Ven R, Mudford C, et al (2010) *Anim. Prod. Sci.* **50**:1107.
 Trapnell C, Roberts A, Goff L, et al (2012) *Nat Protocols* **7**:562.