Unravelling the transcriptomic control of growth traits in sheep during prenatal and postnatal development

S.A. Woolley[1*], M. Salavati[1] and E.L. Clark[1]

[1]The Roslin Institute and Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, EH25 9RG, Midlothian, United Kingdom [*]shernae.woolley@roslin.ed.ac.uk

Abstract

Understanding gene expression in sheep can offer numerous opportunities to investigate the genomic control of health and productivity, and can help to inform future breeding programs. This study aims to identify transcriptionally active genes associated with growth through the analysis of RNA-sequencing (RNA-seq) data in Texel x Scottish Blackface sheep. Gene expression levels were estimated across tissues and developmental stages and allele-specific expression analysed. Tissue- and developmental stage-specific differences in gene expression were observed. Between day 100 of gestation and one week of age, differentially expressed genes related to muscle growth and development, were upregulated including *COL2A1* and *CDCA8* in liver and *MYH3*, *GDF5* and *COL9A2* in skeletal bicep muscle. Allele-specific expression analysis revealed genes exhibiting allelic imbalance in gene families related to growth. Integration of these results with GWAS data will assist the identification of expressed variants that are associated with growth traits in sheep.

Introduction

To meet future food production targets, and provide sufficient protein derived from red meat for a growing global population, improving the efficiency of sheep production is paramount. Growth traits in sheep are important for production and are often polygenic (Moghaddar *et al*. 2019), making the identification of trait-specific markers for genomic selection challenging. Genomic studies have accelerated the identification of key genes and variants of interest for growth traits, such as the mutation in the 3'UTR region of the myostatin gene associated with the double-muscling phenotype in Texel sheep (Clop *et al*. 2006). Investigation of the sheep transcriptome can offer insight into how gene expression influences growth traits, by measuring, for example, the up- or down- regulation of genes involved in muscle growth and differentiation through pre and postnatal development. This study aims to identify transcriptionally active genes associated with growth through the analysis of RNA-seq data across different developmental stages and tissue types in Texel x Scottish Blackface sheep.

Materials & Methods

Animals.

A dataset consisting of 48 samples from three prenatal and four postnatal developmental stages from Texel x Scottish Blackface sheep was generated for tissues associated with growth and development. The tissues used in this study were collected for the sheep gene expression atlas project (Clark *et al*. 2017). Prenatal time points: Day 23 whole embryo (n=3), day 23 maternal caruncle (n=2); Day 35 liver (n=3), placentome (n=2), maternal caruncle (n=2) and Day 100 liver, ovary, skeletal bicep muscle (n=4), placentome (n=2). Postnatal time points: Newborn liver (n=3), skeletal bicep muscle (n=3), ovary (n=2); One week of age liver (n=3), skeletal bicep muscle (n=3), ovary (n=2) and eight weeks of age liver (n=3), skeletal bicep muscle (n=3).

RNA extraction and sequencing.
RNA was extracted from <100mg of -80°C frozen tissue samples using Trizol after homogenisation with ceramic beads. Samples with an RNA integrity number > 6 were selected for RNA-seq and were sequenced by Novogene (Cambridge, United Kingdom) using the NEB Next® Ultra™ RNA Library Prep Kit. Libraries were sequenced on the NovaSeq 6000 platform (Illumina®, United States) at an expected depth of 60 million reads per sample with 150bp paired end reads.

Transcriptome analysis.
After trimming and quality control, mapping of the RNA-seq data to the Oar_rambouillet_v1.0 (GCA_002742125.1; v103) assembly was performed by HISAT2 (v2.1.0) using standard parameters (Kim *et al*. 2019). To estimate RNA transcript abundance, transcript per million (TPM) counts were quantified with Kallisto (v0.44.0) (Bray *et al*. 2016). Based on the TPM estimates, gene-gene network analyses were performed in Graphia (v2.1) (Freeman *et al*. 2020). Functional gene enrichment using EnrichR (Chen *et al*. 2013) in RStudio (R v.3.6.1) was performed for the first 20 gene clusters generated from the gene-gene network analysis.

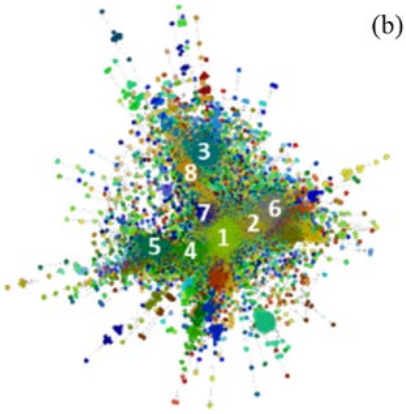
Differential gene expression.
Differential gene expression between prenatal day 100 of gestation and postnatal one week of age was analysed using DESeq2 (Love *et al*. 2014), and the non-averaged transcript abundance counts per tissue type from Kallisto for liver and bicep muscle tissue. A likelihood ratio test (LRT) was used to compare a reduced model across each time point per tissue. The thresholds used for the reduced LRT model analysis included a p-adjusted value (p*adj*) <0.1 and a $\log_2$ change ($\log_2$FC) of 5. DESeq2 results were visualised in R studio using EnhancedVolcano (Blighe *et al*. 2021).

Allele-specific expression.
In brief, allele-specific expression (ASE) analysis was conducted on all 48 RNA-seq samples. Reference mapping bias and ASE analysis was performed as previously described *(*Salavati *et al*. 2019). The Liptak score (transformed z score value) from the GeneiASE output (Edsgärd *et al*. 2016) was used as the proxy for the level of allelic imbalance.

Results

Transcriptome analysis.
Gene-gene network analysis revealed large clusters of genes exhibiting tissue-specific expression in ovary, liver and skeletal bicep muscle tissues (Figure 1a and 1b). For skeletal bicep muscle tissue the genes within cluster four showed high expression levels at day 100 of gestation, and cluster five showed high expression levels in newborn to one-week-old lambs. Functional gene enrichment for cluster four showed associations of genes within this cluster with GO terms for muscle fibre organisation and development, and cluster five showed associations for muscle fibre organisation and contraction. The Myostatin (*GDF8*) gene, in cluster four, for example, which controls muscle development, showed increased expression at day 100 of gestation relative to the other developmental stages.

(a)



(b)

Cluster	Nodes	Tissue
1	6,288	All
2	1,154	Placentome and caruncle
3	767	Liver
4	591	Day 100 skeletal bicep
5	529	Skeletal bicep
6	523	Day 23 caruncle
7	426	Day 23 embryo
8	349	All

Figure 1. (a) Gene-gene network analysis showing large clusters of genes exhibiting tissue- and developmental stage-specific gene expression. Clusters are labelled according to the tissue type with the highest expression level for that cluster, with a Pearson correlation matrix (positive only) of 0.831 and a MCL granularity of 2.2, with the removal of components with a size ≤ 5. (b) The eight largest clusters with the number of nodes (genes) and the associated tissue type.

Differential gene expression.

Rather than presenting all comparisons for differential gene expression analysis, here we focus on liver and skeletal bicep muscle tissue between one prenatal time point (day 100 of gestation) and one postnatal time point (one week of age). For liver tissue, comparing day 100 of gestation and one week of age showed 12% upregulated and 7.9% downregulated genes. In skeletal bicep muscle tissue, 23% of genes were upregulated and 23% were downregulated. Some differentially expressed genes associated with growth had a log2FC>5 (Figure 2). Genes in families associated with growth such as the myosin, TGF-beta and fibrillar collagen families were identified in skeletal bicep muscle tissue. This included the upregulation of the genes *MYH3*, *GDF5* and *COL9A2*. In liver tissue, genes in the fibrillar collagen family were also upregulated, such as the *COL2A1* gene.

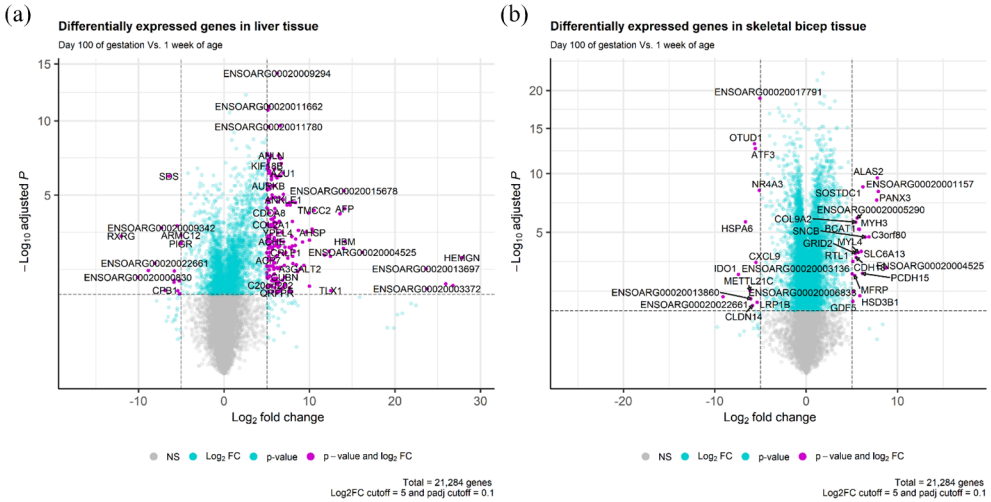


Figure 2. Volcano plots for prenatal time point day 100 of gestation and postnatal one week of age in (a) liver and (b) skeletal bicep muscle tissue, showing differential gene expression using DESeq2.

Allele-specific expression.

Within the top 20 highest liptak scores, for allelic imbalance, the genes *ACTB* and *ANXA2* showed a liptak score of ~946 for a day 100 of gestation sample and ~641 at a one week of age sample, respectively, in liver. For skeletal bicep muscle tissue, the genes *ACTB*, *IGF2* and *MYH2* had liptak scores of ~813 for a day 100 of gestation sample, ~408 for a day 100 of gestation sample and ~334 for a one week of age sample respectively. These genes were found to be members of gene families associated with growth, including either the actin, annexin, insulin family of polypeptide growth factors and myosin family of genes.

Discussion

Improved understanding of the sheep transcriptome provides opportunities to incorporate gene expression information in genomics enabled breeding programmes for sheep (Yuan *et al*. 2021). The initial results from this study identified large tissue-specific clusters of genes associated with growth and development. Differential gene expression analysis revealed upregulation of genes from both the collagen and myosin families, between pre and postnatal time points consistent with previous RNA-seq transcriptome studies (Clark *et al*. 2017). Differential gene expression and initial ASE analysis showed that genes within the myosin and TGF-beta families, including *MYH2*, *MYH3* and *GDF5*, exhibited significant allelic imbalance and/or upregulation in developing skeletal bicep muscle tissue, with *MYH3* associated with growth traits in cattle (Xu *et al*. 2014). Integration of our results with available eQTL datasets for tissues associated with growth (Yuan *et al*. 2021), and with information from GWAS, will identify expressed variants located in genomic regions associated with growth traits, which can be exploited in genomics-enabled breeding programmes for sheep. This work was funded by BBSRC award number BB/S01540X/1.

References

Blighe K., Rana S., Lewis M. (2021) R Packag. version 1.10.0. https://doi.org/doi:10.18129/B9.bioc.EnhancedVolcano.

Bray N. L., Pimentel H., Melsted P. and Pachter L. (2016) Nat. Biotechnol 34:525–527. https://doi.org/10.1038/nbt.3519.

Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z. et al. (2013) BMC Bioinform 14:128. https://doi.org/10.1186/1471-2105-14-128.

Clark E.L., Bush S.J., McCulloch M.E.B., Farquhar I.L., Young R. et al. (2017) PLOS Genet. 13:e1006997. https://doi.org/10.1371/journal.pgen.1006997.

Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X. et al. (2006) Nat. Genet 38:813–818. https://doi.org/10.1038/ng1810.

Edsgärd D., Iglesias M.J., Reilly S.-J., Hamsten A., Tornvall P. et al. (2016) Sci. Rep 6: 21134. https://doi.org/10.1038/srep21134.

Freeman T.C., Horsewell S., Patir A., Harling-Lee J., Regan T. et al. (2020) bioRxiv. https://doi.org/2020.09.02.279349. doi:10.1101/2020.09.02.279349.

Kim D., Paggi J.M., Park C., Bennett C., and Salzberg S. L. (2019) Nat. Biotechnol 37:907–915. https://doi.org/10.1038/s41587-019-0201-4.

Love M.I., Huber W. and Anders, S. (2014) Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8.

Moghaddar N., Khansefid M., Van Der Werf J.H.J., Bolormaa. S., Duijvesteijn N. et al. (2019) Genet. Sel. Evol. 51:72. https://doi.org/10.1186/s12711-019-0514-2.

Salavati M., Bush S.J., Palma-Vera S., McCulloch M.E.B., Hume D.A. et al. (2019) Front. Genet. 10:863. https://doi.org/10.3389/fgene.2019.00863

Xu Y., Shi T., Cai H., Zhou Y., Lan X. et al. (2014) Gene 535:106–111. https://doi.org/10.1016/j.gene.2013.11.057.

Yuan Z., Sunduimijid B., Xiang R., Behrendt R., Knight M. I. et al. (2021) Genet. Sel. Evol. 53:8. https://doi.org/10.1186/s12711-021-00602-9.