THE UNIVERSITY
of EDINBURGH

A gene expression atlas of the domestic goat and comparative analysis of immune signatures with sheep

THE UNIVERSITY *of* EDINBURGH

Charity Muthoni Muriuki

Thesis submitted for the degree of Doctor of Philosophy,

College of Medicine and Veterinary Medicine

University of Edinburgh

November 2019

Dedication

To my parents who have generously offered their unwavering support.

To Dan and our future family.

Declaration

I declare that this thesis presented for the degree of Doctor of Philosophy at the College of Medicine and Veterinary Medicine, University of Edinburgh, has been composed by myself and help from others is clearly acknowledged. I confirm that the work has not been submitted for any other academic degree or professional qualification at this University or any other institution.

Charity Muriuki

November 2019

Abstract

Goats are an economically important livestock species providing a resource of meat and milk across the world. They are of particular importance in developing countries contributing to sustainable agriculture, alleviation of poverty and utilisation of marginal grazing. Recently, a highly contiguous refence genome (ARS1) of the domestic goat was released. However, gene expression information on the domestic goat is particularly limited when compared to other ruminants. Despite great genetic similarity, ruminants vary in their susceptibility to similar pathogens, but the underlying molecular mechanisms remain largely unknown. To elucidate the molecular basis of variation in disease response in small ruminants, a gene expression atlas of the domestic goat was generated from a subset of 22 tissue and cell types and compared to the previously developed sheep gene expression atlas.

Fifty-four mRNA-Seq (poly-A selected) 75bp paired-end libraries spanning all major organ systems in the domestic goat were produced, generating a total of 8.7×10^8 paired end sequence reads. The tissues and cell-types sampled were all transcriptionally complex, with each expressing at least 50% of the total protein coding genes at detectable levels. 18,528 protein coding genes (out of a possible 21,343) had detectable expression in at least one tissue sampled, enabling the capture of 90% of the reference transcriptome. Additionally, of the 21,343 protein coding genes in the ARS1 reference transcriptome 7,036 (33%) had no informative gene name. Using the HISAT2 annotation pipeline, informative gene names were assigned to 1,114 (15%) of the previously un-annotated protein coding genes in ARS1, greatly expanding the previously available genetic and genomic resources available for goat. Using network cluster analysis, genes were assigned to specific biological pathways or cell populations based on expression profiles. Clusters of genes in the liver, gastro-intestinal tract and those involved in innate immunity are analysed and discussed in detail.

Additionally, a protocol to isolate goat bone marrow derived macrophages (BMDM) and culture them in the presence of macrophage colony stimulating factor (CSF1) was developed and optimized. The goat BMDM were characterised using light microscopy to confirm morphology as well as flow cytometry to investigate the cell surface markers. Flow cytometry results revealed that goat BMDM express CD14, CD16 and CD172a on the surface similar to sheep macrophages. When exposed to bacterial lipopolysaccharide (LPS), goat BMDM responded by inducing inflammatory cytokines such as *TNF*, interferon-associated genes including *IFI6, IFIT3* and *IFNG* and interleukins such as *IL10RA, IL12B, IL16* and *IL1RAP* similar to sheep BMDM. However, unlike sheep, goat BMDM produced detectable levels of nitric oxide (NO) post-LPS stimulation. The goat BMDM post-LPS stimulation were also analysed with RNA-Seq to reveal hundreds of upregulated genes further expanding the transcriptional data available for goat.

Finally, the data generated from the network cluster analysis of the goat was used to run a comparative analysis with the larger gene expression atlas of the domestic sheep, revealing transcriptional differences between the two species which may underlie the mechanisms controlling disease variation.

Lay Summary

Goats are an economically important livestock species providing meat and milk across the world. There are excellent genetic and genomic resources available for goats, including a highly contiguous reference genome, but the gene expression information available is limited compared to other livestock species.

To support functional annotation of the goat genome and comparative transcriptomics with other livestock species, this project aimed to create an atlas of gene expression for the domestic goat. Using this resource in combination with other similar available resources for sheep gene expression in the two different species was compared. The outcomes of this project will greatly support effective breeding of the domestic goat and enable investigation on how goats differ from sheep and other livestock species especially in their immune response.

Acknowledgements

I would like to thank my supervisors Professor David Hume and Dr Emily Clark for their immense support throughout my studies and for reading and editing this thesis again and again. I'm particularly grateful to Professor David Hume for giving me the incredible opportunity to work with him on my PhD, in such a prestigious institution as The Roslin Institute. I would not have picked a better supervisor and will forever be indebted for the privilege to have been mentored by him. David is extremely knowledgeable on all matter's macrophages, which he generously shared with me and I'm extremely humbled by his patience and enthusiasm which has left a lasting impression on me.

Dr. Emily Clark, thank you for taking me under your wing right from the first day, and patiently teaching me all the lab work. Thank you for your calm demeanour and excellent organisation skills from which I learnt a lot. Thank you for taking on my supervision fully when David left and very ably getting me to the finish line. I wouldn't have done it without you. Thank you for patiently reading and editing this thesis over and over again and for accommodating me even in my lowest moments. I owe this entire thesis to you and your incredible support.

I would also like to appreciate my other supervisor Dr. Mike Coffey for his advice and guidance and facilitation in obtaining the animals used for this project. Thank you too to my academic mentor, distant supervisor and friend Dr. Morris Agaba, without whom I wouldn't have aspired to pursue a PhD. Thank you, Morris, for taking me on as an intern ten years ago, and very patiently and willingly teach me all you knew on all matter's molecular biology and instilling in me a passion and love for research. Thank you for all the pep-talks we had in the ILRI data room, with you drawing out my career-path and continuing to encourage me that I was capable of achieving a PhD. Thank you for your continued support during my PhD studies and being ready and available to reply to my emails. Without you, I wouldn't have been encouraged to get on this journey and I owe it all to you. Thank you!

I would also like to give my sincere thanks to Dr. Mazdak Salavati, the other third of the ‘sheep and goat genomics’ lab and who helped me immensely especially in Chapter 4 of this thesis and his continued encouragement during my write-up period. Thank you too to Dr. Stephen Bush who helped with all the programming and shared his codes for the sheep project that were used to analyse this work. Finally, I would like to appreciate the other members of Hume lab who have helped me in the lab and during ‘team goat’ days; Dr. Clare Pridans, Dr. Rachel Young, Lucas Lefevre, Lindsay Waddell, Dr. Anna Raper, Dr. Zofia Lisowski, Dr. Mary McCulloch, Dr. Lucy Freem, Dr. Tim Regan, Rahki Harne and Prasun Dutta.

On a personal note, I would like to thank my parents James Muriuki and Winfred Muriuki who have always been a source of support and encouragement every step of my life and especially when I embarked on the PhD journey. Thank you for your endless calls from home, your prayers and love. I would also like to sincerely thank my siblings Martin Kihungi and Mercy Njoki for their support and encouragement. To my fiancé Daniel Muraguri, thank you for bearing the brunt of a long-distance relationship together with me, the endless calls and texts and the love that literally crossed oceans. Without your love and support, I wouldn’t have made it to the end. I would also like to thank my dear friends Margaret Mathenge and Linda Koi for their love, support and regular calls from home that made the PhD journey easier to bear.

Thank you to my fellow PhD friends, Dr. Dishon Muloi and Dr. Bridgit Muasa, with whom we started the PhD journey together and continued to encourage each other every step of the way. It was a joy to explore Scotland with you and bear the ‘beautiful’ Scottish weather the best way our Kenyan roots would allow. Thanks Dishon, for the endless cups of tea, pizza and card-game Fridays, that helped us feel less homesick. Thanks too to Dr. Hussein Abkallo who, together with Dishon and Bridgit, we created the tradition of movie and/or dinner Fridays, where we could converse in ‘sheng’ and feel very much at home, despite being miles away. Thank you too to Dr. James Nyagwange,

who being a few steps ahead in the PhD journey offered immense support and encouragement.

To Ms Joan Nadeau, thank you for offering me your friendship. I'm glad I signed up for the Go Connect program through which we met and formed a lasting friendship. Thank you for opening up your home to me, and very kindly offering scrumptious home-made meals that made me feel like I was back in my own mother's kitchen. Thank you for the numerous drives we took for sightseeing in and around Edinburgh, from Linlithgow Castle, North Berwick, Falkirk wheel, St. Andrews, Oban and Isle of Mull, The Roslin Chapel, Stirling to The Kelpies and the Scottish Canals, among many others. Thank you for introducing me to the beauty that was the fringe festival fireworks concert. Thank you for the many conversations we had about everything, It was a joy to have you in my life whilst I was in Edinburgh. Thank you for extending your friendship to my parents when they visited and taking us on drives for sightseeing. I will cherish our friendship forever for it made my experience in Edinburgh very fulfilling.

Finally, I would like to give my sincere appreciation to Professor Appolinaire Djikeng and Professor Steve Kemp for allowing me access to the CTLGH space in ILRI Nairobi, where I spent endless hours during the final write-up phase of my PhD.

List of Publications

Muriuki C, S.J Bush, Salavati M, McCulloch MEB, Z.M Lisowski, M. Agaba, A. Djikeng, M. Coffey, D.A Hume and E.L Clark (2019) A mini-atlas of gene expression for the domestic goat (*Capra hircus*). *Frontiers in Genetics*. 10:1080

Young R, Bush SJ, Lefevre L, McCulloch MEB, Lisowski ZM, **Muriuki C**, Waddell LA, Sauter KA, Pridans C, Clark EL and Hume DA (2018) Species-specific transcriptional regulation of genes involved in nitric oxide production and arginine metabolism in macrophages. *ImmunoHorizons.* 2(1):27-37

Bush SJ, **Muriuki C**, McCulloch MEB, Farquhar IL, Clark EL and Hume DA (2018) Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *GSE* 50:20.

Waddell L., Lefevre L., Bush S.J., Raper A., Young R., Lisowski Z.M., McCulloch M.E., **Muriuki C.**, Sauter K.A., Clark E.L., Irvine K.M., Pridans C., Hope J & Hume D.A (2018) ADGRE1 (EMR1, F4/80) is a rapidly-evolving gene expressed in monocyte-macrophages in all mammalian species. *Frontiers in Immunology*. 9: 2246.

Bush S.J., McCulloch M.E., **Muriuki C**., Salavati M., Davis G.M., Farquhar I.L., Lisowski Z.M., Archibald A.L., Hume D.A. & Clark E.L (2019) Comprehensive transcriptional profiling of the gastrointestinal tract of ruminants from birth to adulthood reveals strong developmental stage specific gene expression. *G3.* 9(*2): 359-373;*

List of Abbreviations

AM	Alveolar macrophage
ARS1	current goat genome assembly
BMDM	Bone marrow derived macrophage
CSF	colony stimulating factor
DNA	deoxyribonucleic acid
ENCODE	Encyclopaedia of DNA Elements
FBS	Fetal bovine serum
FDR	false discovery rate
FACS	fluorescence-activated cell sorting
FANTOM	Functional annotation of mammalian genome
FAANG	Functional Annotation of Animal Genomes
GTEx	Genotype-Tissue Expression
GO	Gene Ontology
GI	Gastro-intestinal
GM-CSF	Granulocyte-macrophage colony-stimulating factor
HGNC	HUGO Gene Nomenclature Committee
IGGC	International Goat Genome Consortium
MYA	Million years ago
MDM	Monocyte derived macrophages
MPS	Mononuclear phagocyte system
NCBI	National Centre for Biotechnology Information
NK	Natural Killer
OEF	ovine embryonic fibroblast
PCR	Polymerase chain reaction

PRR	pathogen recognition receptors
PAMP	pathogen associated molecular patterns
PBMC	Peripheral blood mononuclear cells
RNA	ribonucleic acid
RIN	RNA integrity number
RT-PCR	Reverse-transcription PCR
RBC	Red blood cell
TLR	Toll-like receptor
TPM	Transcripts per million

Contents

List of tables

List of figures

Chapter 1 Introduction

1.1 Improving the genomic resources for goats

Section 1.1 of this chapter is adapted from the authors published work (Muriuki, Bush et al. 2019).

Goats are an important livestock species across the globe and are of particular importance in tropical agri-systems contributing to sustainable agriculture, alleviation of poverty, social cohesion and utilisation of marginal grazing. The International Goat Genomics Consortium (IGGC) (http://www.goatgenome.org) has provided extensive genetic tools and resources for goats including a 52K SNP chip (Tosser-Klopp, Bardou et al. 2014), a functional SNP panel for parentage assessment and breed assignment (Talenti, Palhière et al. 2018) and large-scale genotyping datasets characterising global genetic diversity (Stella, Nicolazzi et al. 2018). In 2017 a highly contiguous reference genome for goat (ARS1) was released (Bickhart, Rosen et al. 2017, Worley 2017), a year later in 2018 the ARS1 assembly was released on the Ensembl genome portal (Zerbino, Achuthan et al. 2018) (https://www.ensembl.org/Capra_hircus/Info/Index) greatly facilitating the utility of the new assembly and providing a robust set of gene models for goat.

RNA-Sequencing (RNA-Seq) has transformed the analysis of gene expression from the single-gene to the whole genome allowing visualisation of the entire transcriptome and defining how we view the transcriptional control of complex traits in livestock (reviewed in (Wickramasinghe, Cánovas et al. 2014)). Using RNA-Seq my lab group generated a large-scale high-resolution atlas of gene expression for sheep (Clark, Bush et al. 2017). This dataset included RNA-Seq libraries from all organ systems and multiple developmental stages and provided a model transcriptome for ruminants. Analysis of the sheep gene expression atlas dataset indicated we could capture approximately 85% of the transcriptome by sampling twenty ‘core’ tissues and cell types (Clark, Bush et al. 2017). One of the main aims of my thesis was to recreate this resource on

a smaller scale for goat. This would provide a smaller, cost-effective, atlas of gene expression for the domestic goat based on transcriptionally rich tissues from all the major organ systems.

In the goat genome there are still many predicted protein-coding and non-coding genes for which the gene model is either incorrect or incomplete, or where there is no informative functional annotation. For example, in the current goat reference genome, ARS1 (Ensembl release 97), 33% of the protein-coding genes are identified only with an Ensembl placeholder ID. Many of these unannotated genes are likely to have important functions. RNA-Seq data can be used to annotate them and assign function (Krupp, Marquardt et al. 2012). With datasets of a sufficient size, genes form co-expression clusters, which can either be ubiquitous, associated with a cellular process or be cell-/tissue specific. This information can then be used to associate a function with genes co-expressed in the same cluster, a method of functional annotation known as the 'guilt by association principle' (Oliver 2000). Using this principle for the sheep gene expression atlas dataset it was possible to annotate thousands of previously unannotated transcripts in the sheep genome (Clark, Bush et al. 2017). By applying this rationale to the goat mini-atlas dataset it would be possible to do the same for the goat genome.

A high-quality functional annotation of existing reference genomes can help considerably in understanding the transcriptional control of complex traits such as immunity and contribute to further improvements in productivity. A substantial component of my thesis focuses on the transcriptional signatures of macrophages in goats and comparative analysis with sheep. Species-specific differences in response to disease, or other traits, are likely to be reflected in gene expression profiles of immune cells. Sheep and goats are both small ruminant mammals and are similar in their physiology. They also share susceptibility to a wide range of viral, bacterial, parasitic and prion pathogens, including multiple potential zoonoses (Sherman 2011), but there have been few comparisons of relative susceptibility or pathology between the species to the same pathogen, nor the nature of innate immunity.

1.2 An overview of Innate Immunity

The immune system is a complex mechanism within an organism that protects against disease broadly classified into innate and adaptive immunity. It functions by distinguishing foreign material from self, identifying and killing a wide range of pathogens including viruses, bacteria and parasitic worms. Innate or native immunity is the evolutionarily primitive, and germ-line specific form of protection that acts as the first line of defence against pathogens. It is universal, antigen-independent and relatively non-specific, making it a more rapid response to infection. Native immunity is considered the most important form of protection, it does not depend on prior exposure to a pathogen and many organisms survive through it alone. Adaptive or acquired immune system is the second layer of protection against disease. It is characterised by its antigen-specific nature, dependence on memory of prior exposure to a pathogen and a time-lapse between exposure and full protection. In many instances, innate and adaptive mechanisms overlap. The focus of my thesis is the innate immune response.

Innate immunity is constituted of physical, cellular and chemical components. The physical component is made up by the mucous membrane that covers all internal organs and poses a barrier against entry by pathogens and environmental contaminants. The chemical component is comprised of cell-associated or soluble molecules known as pathogen recognition receptors (PRR) capable of recognizing pathogens via unique signatures collectively referred to as pathogen associated molecular patterns (PAMPs), as well as enzymes and peptide molecules which possess the ability to hydrolyse microbes. It also includes cytokines and chemokines e.g. IL-1, IL-6 and IL-8 that are responsible for triggering and coordinating the immune response. The cellular component arm of innate immunity is made up of natural killer (NK) cells, T-cells, epithelial cells, mast cells, classical dendritic cells and phagocytic macrophages. Host-pathogen interactions are under continuous selective pressure from rapidly evolving pathogens, and as a result, the innate immune system has evolved to exhibit a high level of diversity across different species (Basset, Holton et al. 2003, Beutler 2004, Tosi 2005).

Pathogens can cross the anatomical barrier posed by the epidermis and penetrate into the tissue. In such situations, the interaction of the pathogen recognition receptors (PRRs) and pathogen through the PAMPs leads to activation of an immune response. As a consequence, an acute inflammatory reaction occurs characterised by symptoms such as high fever. PAMPs are important elements of the microbial cell structure and have remained preserved across evolution. Lipopolysaccharide, also known as endotoxin, is a component of the cell-wall of gram-negative bacteria and is one of the well-studied PAMPs, due to its stability, availability and the wide spectrum of biological responses it can elicit. In small doses, LPS renders a protective property, making an organism resistant to other pathogens, a condition known as endotoxin tolerance (Beutler 2004). While in large amounts it overstimulates the immune response resulting in wide-spread sepsis-shock, a fatal reaction that can result in organ-failure (Dobrovolskaia and Vogel 2002). LPS has been used extensively in *in vitro* assays to mimic immune responses during infection. Studying gene expression profiles following exposure to LPS gives a good indication of how the organism responds to microbial infections (Dobrovolskaia and Vogel 2002, Fujihara, Muroi et al. 2003). The molecular mechanisms that regulate this phenomenon have been the basis of extensive experimental studies on LPS-mediated responses in mouse models, pigs (Kapetanovic, Fairbairn et al. 2012) as well as in humans revealing the large extent of alternative control systems. Little knowledge is available on the differential responses in ruminants (Dobrovolskaia and Vogel 2002, Nilsson, Bajic et al. 2006). During my PhD I aimed to fill this gap and advance our understanding of the ruminant immune response using functional genomics.

The response of pathogen recognition receptors (PRRs) is crucial in determining the outcome of infection and is one of the main components of the innate immune system affected by stimulation with LPS. PRRs are either found on the surface of cells, in intracellular compartments or as soluble components in the blood and tissue fluids. They are associated with most cells of the innate immune system including macrophages, monocytes, granulocytes, mast cells and epithelial cells. They play a role in phagocytosis, opsonisation, induction

of apoptosis and the activation of inflammatory signalling pathways. There are different classes of PRRs from glycan and mannose receptors like macrophage mannose receptor (MRR), scavenger-type PRRs e.g. macrophage scavenger receptor (MSR), complement receptors like CR3, soluble receptors like CD14 and the well-studied toll-like receptors (TLRs) (Basset, Holton et al. 2003).

Toll-like receptors are a type of transmembrane proteins embedded in the cell membrane and contain specific motifs in their structure crucial for the identification of PAMPs. They have been conserved through-out the evolution process, largely remaining unaltered across different species. Depending on their location in the cell and what they recognize, TLRs are broadly classified into groups. Group one are those expressed on the cell surface that mainly recognize microbes e.g. TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10. The second group is made up of intercellular receptors e.g. TLR3 TLR7 and TLR9 which recognize nucleic acid components of pathogens. Activation of TLRs triggers two different signalling pathways: myeloid differentiation primary response gene 88 (MyD 88) dependent pathway and the interferon regulatory factor (IRF) dependent pathway. One of the most studied TLRs is TLR4 in response to LPS where it activates production of pro-inflammatory cytokines and chemokines e.g. IL-1, TNF, IL-6 and IL-8 which in turn recruit other components of the innate immune system and result in the development of an acute inflammatory reaction (Zarember and Godowski 2002, Beutler, Jiang et al. 2006, Kannaki, Shanmugam et al. 2011, Schroder, Masterman et al. 2012).

1.3 Macrophage biology

1.3.1 Mononuclear Phagocyte System

The mononuclear phagocyte system (MPS) is a family of cells of the innate immune system made up of tissue macrophages, bone marrow precursors, circulating blood monocytes and dendritic cells (Hume, Ross et al. 2002, Hume 2006, Hume 2008, Fairbairn, Kapetanovic et al. 2011). Tissue macrophages are ubiquitous in the body, making approximately 15% of all cells. They are known as Kupffer cells in the liver, langerhans in the mucosa, osteoclasts in

the bone, alveolar macrophages in the lungs, microglia in the brain among many other cells. Each of these resident tissue specific macrophages are functionally and biologically different (Davies et al. 2013). MPS cells participate in innate immune mechanisms as well as growth, development and homeostasis and share a lot of similar features especially their phagocytic activity. Differentiation and maturation of cells of the MPS is controlled by special growth factors e.g. CSF-1 and IL-34 both of which activate the macrophage colony stimulating factor receptor (CSF-1R). Regulation of blood monocytes and tissue macrophages through CSF-1R has been widely studied and shows great similarity across species suggesting that interaction between the ligands (CSF1, IL-34) and receptor is crucial for function of macrophages in immune responses (Hume, Ross et al. 2002, Hume 2006). For instance, CSF-1 knock-out mouse and rat models have huge inadequacies of tissue macrophages (Freeman 2014). In addition, blocking of CSF-1R with antibodies leads to reduction of tissue macrophage populations but not of monocytes (Rojo, Raper et al. 2019).

Traditionally, tissue macrophages have been identified by use of specific macrophage surface markers e.g. F4/80 (Hume, Ross et al. 2002). Their development is considered to follow a linear pattern from pluripotent progenitors to committed myeloid progenitors, promonocytes, blood monocytes and eventually tissue macrophages. This view has however since been challenged by more recent studies which have since divided the MPS into distinct lineages, originating from either the yolk sac, foetal liver or bone marrow progenitor. Recent views have also questioned how much of the tissue macrophages derive from circulating monocytes, how frequent they are replenished and the life-span of the resident tissue macrophages in their respective tissues. Debate on the principle definition of MPS is still ongoing and experimental studies to investigate different features continue to be developed (Jenkins and Hume 2014).

Each individual cell-type in the complex biology of a mammal, has its own unique transcriptional profile (Hume, Summers et al. 2010). Therefore, to fully

appreciate the role of macrophages in immune regulation and of the MPS in general, a transcriptomic approach is required. Despite the differing views on the concept of the MPS, its transcriptomic control has been widely studied in mouse and human including macrophage responses during disease (Freeman 2014). Similarly, others have studied transcriptional responses in livestock species including pig (Fairbairn, Kapetanovic et al. 2011, Freeman, Ivens et al. 2012), sheep (Clark, Bush et al. 2017), water buffalo (Young, Lefevre et al. 2019) and chicken (Bush, Freem et al. 2018). Investigation of transcriptomic control of the MPS in goat will help us to gain a better understanding of the complex genetic control of the immune system and how this varies between livestock species.

1.3.2 Transcriptomics of the MPS

Genes with similar functions, are more often than not expressed together in specific cell or tissue types, and are under similar transcriptional control. As such, it is possible to infer the function of a gene by looking at the function of genes it is co-expressed with and those with which it shares transcriptional regulation. This is referred to as the principle of 'guilt-by-association' (Hume, Summers et al. 2010, Freeman 2014). Using this method however depends on the availability of a large amount of transcriptional data to have enough statistical power. Next-generation sequencing technologies such as microarrays, RNA-Seq and the development of advanced bioinformatics tools have enabled the acquisition and analysis of large-scale transcriptomic data. The mouse gene expression atlas created by Su et.al (Su, Cooke et al. 2002, Su, Wiltshire et al. 2004) revolutionized how these analyses were carried out, by measuring gene expression across multiple tissues and cell-types from humans and mice. Extensive transcriptomic analyses have also been undertaken by the Functional Annotation of the Mammalian Genome (FANTOM) consortium that aims to assign functional annotation to full-length transcripts (Consortium 2001, Exploration 2005). The analysis undertaken in these pioneering studies not only revealed clustering of genes with similar function but also those involved in similar processes e.g. genes encoding the cell-cycle, RNA and protein synthesis. Similar analysis has also been

performed in the domestic pig (Freeman, Ivens et al. 2012), sheep (Clark, Bush et al. 2017), water buffalo (Young, Lefevre et al. 2019) and chicken (Bush, Freem et al. 2018) and, as mentioned above, one of the main aims of my PhD was to generate a gene expression atlas of the goat albeit to a smaller extent (compared to the sheep gene expression atlas), with a specific focus on the innate immune system and compare this across the transcriptome of sheep and other ruminants. This comparative analysis will help us to understand the species-specific transcriptional differences that underlie variation in susceptibility to pathogens.

1.3.3 Toll-like receptor polymorphism, genetics of disease resistance and their role in animal production

Domestic ruminants e.g. cattle, sheep and goats all evolved from a common ancestor, the *Caprinae* lineage (sheep and goats) and diverged from the larger *Bovinae* up to 25 million years ago (MYA) while sheep and goats became distinct species about 4 MYA (Jiang, Xie et al. 2014). This close evolutionary relationship is also evidenced by the great genetic similarity and large number of orthologous genes shared. In addition, the first goat reference genome (CHIR_2.0) was assembled into scaffolds by anchoring its chromosomes onto the better studied cattle genome based on the great level of collinearity between goat and cattle chromosomes (Dong, Xie et al. 2013). Despite their great genetic similarity, ruminants vary in their susceptibility and resistance to pathogens. *Mycobacterium bovis* for example hardly causes disease in sheep despite prolonged co-grazing with cattle (Munoz Mendoza, Juan et al. 2012), while Johne's disease caused by *Mycobacterium paratuberculosis* exhibits strain specific variation in pathogenicity between cattle, sheep and goats (Clarke 1997). The molecular mechanisms responsible for these phenotypic differences remain largely unknown. However, considering the great genotypic similarity, the phenotypic variations observed in disease susceptibility can be explained by non-orthologous divergence resulting in genes of different function, or a change in regulation of transcription of orthologous genes. Evidence of divergence of transcriptional control has been widely studied in the human and mouse models (Su, Cooke et al. 2002) and from this,

hypotheses can be formulated on the evolution of ruminant transcriptional control and TLR polymorphism. By performing comparative transcriptomic analysis of macrophages from goats with other livestock species, we will begin to elucidate these mechanisms and understand the evolution of immunity across different ruminant species.

In addition to their role in immunity, TLRs have been implicated in other biological processes of animal production including ovulation, fertilization, gestation, parturition and spermatogenesis in males. Evidence also shows that they are involved in disease processes like mastitis and as such, genetic polymorphisms within TLRs can be targeted as markers for breeding programs (Tirumurugaan, Dhanasekaran et al. 2010, Kannaki, Shanmugam et al. 2011). Toll-like receptors although evolutionarily conserved, have been shown to be alternatively spliced leading to a variation in the signalling cascade of inflammatory response, one consequence of which is the different disease susceptibilities observed across species (Wells, Chalk et al. 2006). This great repertoire of varying innate immune responses is vital to the survival of populations under disease pressure.

By investigating the innate immune responses in different livestock species, we will begin to understand the complex transcriptional patterns of toll-like receptor genes and the functional consequences this might have for disease susceptibility. By identifying key genes involved in the immune response in one species this is likely to highlight genes of interest in the other. These genes could be targets for gene editing or novel therapeutics to improve disease resilience in livestock.

1.4 **Aims of this study**

The overall aims of this study were as follows:

i) To improve the transcriptomic resources for goat by generating a mini-atlas of gene expression. This would provide a valuable resource for the livestock genomics community to complement the available genomic tools.

ii) Provide the first comprehensive characterisation of bone marrow derived macrophages (BMDM) from goat providing a methodology for culturing and characterisation and a detailed analysis of transcription post stimulation with LPS.

iii) Examine the transcriptional basis for differences in the immune response between goats and sheep using comparative analysis of RNA-Seq data from macrophages from these two species.

iv) Provide a foundation for further studies that will investigate transcriptional control of the immune response in both tropical and temperate goat breeds.

Chapter 2 A Gene Expression Atlas of the Domestic Goat

This chapter is adapted from the authors published work (Muriuki, Bush et al. 2019).

2.1 Introduction

The aim of this chapter was to generate a gene expression atlas of the domestic goat from a subset of tissue and cell-types. The Hume lab in which this study was performed, has had a long-standing interest in elucidating the transcriptional landscape of multiple species by generating gene expression atlases such as for pig (Freeman, Ivens et al. 2012, Kapetanovic, Fairbairn et al. 2012), sheep (Clark, Bush et al. 2017), chicken (Bush, Freem et al. 2018) and water buffalo (Young, Lefevre et al. 2019). These studies have demonstrated the utility of a gene expression atlas to better understand the biology and genome-wide transcriptional regulation within cells and tissues and have provided a useful resource of data for comparative transcriptomic analysis studies. This history provides the premise under which the goat gene expression project was initiated, adapting methods and analysis pipelines developed for other species, particularly the sheep gene expression atlas project (Clark, Bush et al. 2017) to profile the transcriptional landscape of the domestic goat.

Whole transcriptome sequencing using RNA-Sequencing has proven to be a high-throughput means of generating genome-wide gene expression data at increasing reliability and reducing cost (Wickramasinghe, Cánovas et al. 2014). Pioneering work in large-scale gene expression projects (Su, Cooke et al. 2002, Su, Wiltshire et al. 2004) revealed the complexity of the genome and transcriptional control of gene expression using micro-arrays. More recent transcriptomics projects using Next Generation Sequencing (NGS) technologies such as FANTOM 5 (Andersson, Gebhard et al. 2014), the ENCODE project (Birney, Stamatoyannopoulos et al. 2007) and the GTEx

consortium (Mele, Ferreira et al. 2015) have provided great insight into the molecular mechanisms underlying human biology and disease. Similarly, the Functional Annotation of Farm Animal Genomes consortium (FAANG) (Andersson, Archibald et al. 2015) is aimed at providing functional genomics information, including transcriptomic data, to improve the annotation of the genomes of livestock and companion animals. This information will help the livestock genomics community investigate how variation in gene expression drives phenotypic diversity and the control of complex traits such as disease resilience and productivity. This study therefore contributes to the objectives of FAANG by providing comprehensive transcriptomic data on the domestic goat. We have generated a gene expression atlas to investigate transcriptional patterns in tissues and cell-types of the goat and to improve the annotation of the current reference genome ARS1 (Derek M Bickhart, Benjamin D Rosen et al. 2017).

Transcriptomic analysis have previously been used to reveal the cell and tissue-type specificity of gene expression patterns (Krupp, Marquardt et al. 2012) and genes encoding protein products required for a cell-specific function or pathway are more likely to be co-expressed. Based on the principle of 'guilt-by-association', co-expression networks can be exploited to assign function to unknown genes (Oliver 2000, Hume, Summers et al. 2010, Mabbott, Baillie et al. 2010). An extension of this principle is that variants associated with a complex trait or phenotype tend to be located within sets of genes that are co-expressed (Baillie, Bretherick et al. 2018). This observation is critical in linking observable phenotype to gene expression by revealing the expression profile of candidate genes associated with key traits. For instance, the sheep expression atlas (Clark, Bush et al. 2017) was utilized to assign functional annotation to >1000 previously unannotated protein-coding genes in the OAR v3.1 sheep reference genome (Jiang, Xie et al. 2014). Similarly, the sheep mastitis study (Banos, Bramis et al. 2017) utilized the sheep gene atlas to assess expression of genes within candidate regions for mastitis resistance, further demonstrating the utility of a gene expression atlas as a resource to

identify genes involved in disease susceptibility and other physiological processes.

Despite the recent release of a greatly improved and contiguous goat reference genome (Derek M Bickhart, Benjamin D Rosen et al. 2017, Worley 2017), functional annotation of the domestic goat remains relatively limited compared to other ruminant species such as sheep and cattle, and thousands of genes have no informative gene name. As such the aim of this chapter was to generate an atlas of gene expression for goat that could improve the existing annotation and provide a resource that is useful to both the livestock genomics community and ruminant researchers more broadly.

2.2 Materials and Methods

2.2.1 Animals

This study was reviewed and approved by The Roslin Institute, University of Edinburgh's Animal Work and Ethics Review Board (AWERB). All animal work was carried out under the regulations of the Animals (Scientific Procedures) Act 1986. All animals used in this study were sourced from the same farm following culling as part of normal husbandry routine. Six healthy males and one female crossbred goat, approximately five to six days old were used for this study.

2.2.2 Tissue collection

A total of seventeen different tissues and two cell types (with two conditions for one of the types), spanning all major organ systems, were harvested from six males and one female crossbred goat. Based on prior knowledge from the sheep expression atlas (Clark, Bush et al. 2017), we sought to collect the most transcriptionally rich tissues to maximise the diversity and complexity of the transcriptome we could capture from a limited sample set. The majority of the tissue and cell samples were collected from four male goats. Tissue from the nervous system and alveolar macrophages were collected from a different set of two males, and four tissue samples of the reproductive tract were collected from the one female. Tissues were collected ensuring, where possible at least two biological replicates per tissue sample. The tissue samples were excised post-mortem in sterile conditions within one hour of death, cut into slices about 0.5cm thick and infused in RNA*later* (Thermo Fisher Scientific, Waltham, USA) to stabilise and protect cellular RNA before transporting back to the lab. Spleen tissue chunks and lipid-rich brain tissue were snap-frozen in dry ice, transported back to the lab on ice, and kept in a -80^{o}C freezer for long-term storage. Tissue samples in RNA*later* were stored in the cold-room at 4^{o}C for short-term storage. Within one week, the infused tissue samples were removed from the RNA*later* put in 1.5ml screw cap cryovials then stored at -80^{o}C for long-term storage and until RNA isolation.

2.2.3 Cell isolation

Alveolar macrophages were isolated from two male goats by pouring 200mL of sterile PBS (Mg^{2+} and Ca^{2+} free) into excised lungs. Cells were collected from the lung walls by gentle massaging and the lung contents collected in a sterile beaker using a funnel to pour. The lavage was repeated twice until the wash ran clear. The alveolar wash was aliquoted into 50mL tubes and placed on ice for transport back to the lab. Once in the tissue culture lab, the alveolar wash was filtered through 100uM cell strainers to remove debris and centrifuged at 400 × g for 10 minutes. The cell pellet was recovered and suspended in 50mL of pre-warmed goat media (RPMI 1640 supplemented with 20% heat-inactivated goat serum, 5mL GlutamaX and 1.25mL Penicillin/Streptomycin), then counted on a haemocytometer. Alveolar macrophages were seeded in 6-well tissue culture treated plates (Nunc, ThermoFisher) in 2mL of goat media at a density of 2.0×10^6cells/mL. Cells were grown in the presence of recombinant human colony stimulating factor 1 (rhCSF1, a gift from Chiron, Emeryville, CA, USA) at a final concentration of 10^4 U/ml (100ng/mL) and plates incubated at 37^0C in 5% CO_2 for 24hrs. Fully differentiated, cells were recovered by washing with 1mL Trizol Reagent (Invitrogen, Darmstadt, Germany) then collected in pre-labelled cryovials and stored at -80 ^{o}C until RNA isolation. The rest of the cells not used for culture were prepared for cryopreservation as described for pig (Kapetanovic, Fairbairn et al. 2012) by suspending in freezing media (90%FBS, 10%DMSO), slowly to avoid shocking the cells with DMSO and aliquoted at a density of 5×10^7cells/mL per cryovial. The cryovials were stored in an isopropanol freezing unit (Mr Frosty) at -80^{o}C for 24hrs to allow a slow reduction in temperature. The next day, cells were transferred to a -155^{o}C freezer for long-term storage.

Bone marrow cells were also collected and differentiated into bone-marrow derived macrophages (BMDM). The isolation, culture and characterisation of BMDM is described in Chapter 3.

2.2.4 Total RNA Isolation

Before isolating RNA from the tissues, all surfaces and equipment were cleaned with RNase-Zap (ThermoFisher Scientific, UK) to remove all traces of RNases and nucleic acid contaminants. Total RNA was isolated from frozen tissue and cells using an organic method. Difficult to lyse tissue samples such those from the gastro-intestinal tract were homogenised using CKMIX (431-0170; VWR, Radnor, USA) lysing matrix ceramic beads while soft tissues were lysed using CK14 (432-3751; VWR, Radnor, USA) beads. 1 ml of chilled Trizol Reagent (Invitrogen, Darmstadt, Germany) was added to 50-100mg of tissue and lysis performed on the Precellys 24 (Bertin Instruments) by pulsing at 5000rpm for 20sec. Alveolar macrophage cells samples that had been collected in 1mL Trizol were lysed by mixing with a pipette. Following lysis, 200ul of BCP (1-bromo-3-chloropropane) (Sigma) was added to the lysis tube and samples shaken vigorously for 15 secs, and then incubated for 3 min at room temperature. Tubes were centrifuged for 15 min at 12000xg at 4 ^{o}C to phase separate the homogenate into clear aqueous layer containing RNA, an interphase containing DNA and a red lower layer containing protein. The aqueous layer containing RNA was recovered carefully and cleaned using spin columns with the RNeasy Mini Kit (Qiagen, UK). The RNeasy Mini Kit Protocol: Purification of Total RNA from Animal Tissues and RNeasy Mini Kit Protocol: Purification of Total RNA from Animal Cells was used from step 5 onwards. Samples were eluted in a final volume of 50uL of RNase-free water. 5uL of sample was set aside for quality control checks while the rest of the sample was stored at - 80^{o}C for long-term storage. This method is widely used and well reported to produce high yields and good quality RNA, without the need for precipitation (Grabmuller, Madea et al. 2015).

2.2.5 RNA quantification and quality check

RNA quality was initially estimated by measuring 1uL from each sample on the NanoDrop spectrophotometer (NanoDrop Products, Wilmington USA), and assessing the absorbance readings at a wavelength of 260nm, 280nm and

230nm. To confirm NanoDrop readings, samples were also analysed on the Qubit RNA BR fluorometric quantitation (ThermoFisher Scientific UK) kit using the RNA Broad Range Assay on the Qubit 2.0 fluorimeter. A working solution of 200uL for each sample and standard was prepared by diluting the Qubit RNA BR reagent with Qubit RNA BR buffer at a ratio of 1:200. 190uL of working solution was added to 10uL of standard (two standards provided as part of kit) while 199uL of working solution was added to 1uL of each RNA sample. The tubes were gently vortexed, allowed to incubate at room temperature for 5mins and measured on the Qubit 2.0 fluorimeter after calibrating with freshly prepared standards.

The RNA integrity number (RIN) was measured on the Agilent 2200 TapeStation (Agilent Technologies, Santa Clara CA USA) using an RNA Screentape (5067-5579, Agilent Technologies), and following the manufacturer's instructions. After diluting RNA samples with RNase-free water to attain a maximum concentration of 200ng, 1 µL of High Sensitivity RNA Sample Buffer (supplied with kit) was added to 2uL of diluted RNA sample, tubes vortexed at 200rpm for 1min then spun down to position the sample at the bottom of the tube. RNA was denatured by heating at 72 ^{o}C for 3min and then cooled on ice for 2min. Samples were loaded onto the Agilent 2200 TapeStation instrument with the corresponding RNA Screentape and measured.

2.2.6 Library preparation and RNA sequencing

Library preparation and RNA sequencing was performed by Edinburgh Genomics (Edinburgh Genomics, Edinburgh, UK). Libraries were run on the Illumina HiSeq 4000 platform (Illumina, San Diego, USA) and sequenced at 30 million; 75bp paired-end reads per sample. All libraries were prepared using the Illumina TruSeq mRNA library preparation protocol (poly-A selected)(Illumina 2017). Briefly, oligo (dT) coated magnetic beads were used to purify polyA containing mRNA molecules which were then fragmented and

primed for cDNA synthesis. The first strand cDNA was synthesised by reverse transcribing cleaved RNA fragments with random hexamers. Actinomycin D was added to the synthesis mix to prevent pseudo DNA-dependent synthesis, only allowing RNA-dependent synthesis to improve strand specificity (Ruprecht, Goodman et al. 1973). The second strand cDNA was synthesised using DNA polymerase 1 and RNase H by replacing deoxythymidine triphosphate (dTTP) with deoxyuridine triphosphate (dUTP) to generate a double stranded cDNA molecule (ds cDNA). One adenine (A) nucleotide was added to the 3' ends of the ds cDNA molecules and a corresponding thymine (T) nucleotide added to 3' end of the adapter to provide a complimentary overhang to ligate the adapter to the fragment. The adapter-fragment chimera was hybridized onto a flow cell and using PCR, and only those fragments having adapter molecules on both ends were amplified.

2.2.7 Data processing

Sequencing data was processed and analysed using methods and analysis pipelines developed for the sheep gene expression atlas (Clark, Bush et al. 2017) and summarised in Appendix E. Kallisto (Bray, Pimentel et al. 2016) was used to calculate gene expression estimates as transcripts per million (TPM). Kallisto employs an algorithm to create an index of short sequences, known as k-mers, about 31 bases in length from a known set of transcripts comprising the reference transcriptome of the species being investigated. It then estimates expression levels from the reads directly without the need to align each read to each transcript, thereby greatly reducing the computing time and power required for analysis. For instance, approximately 30 million reads can be processed in less than ten minutes on a laptop (Bray, Pimentel et al. 2016). The algorithm relies on the availability of a robust reference transcriptome and cannot reconstruct novel transcripts or splice junctions. To enable detection of novel transcripts using Kallisto, a 'two-pass' approach was applied to the Kallisto algorithm. In the first 'pass', all raw reads from the sequencing were pseudo-aligned against the ARS1 goat reference transcriptome (Bickhart, Rosen et al. 2017) available from NCBI

(ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/704/415/GCF_001704415.1_ARS1/GCF_001704415.1_ARS1_rna.fna.gz). Output from the first pass was then parsed to create a revised version of the reference transcriptome for the second pass. The second pass included any transcripts missing in the reference transcriptome in cases where the annotation was incomplete. To do this, any reads that Kallisto could not pseudo-align were assembled *de novo* into putative transcripts and retained only if they could be robustly annotated (by, for instance, encoding a protein similar to one of known function) and if they showed coding potential. Additionally, transcripts with no detectable expression levels in any of the sequenced libraries were removed from the second pass as these were likely to represent transcripts that were highly restricted to a tissue that was not sampled or those in which the expression level was below the detection limit of the sampling depth used.

To complement the Kallisto pipeline, the conventional alignment-based pipeline using HISAT2 (Kim, Langmead et al. 2015) and StringTie (Pertea et al. 2015; Pertea, Kim et al. 2016) was run in parallel. The HISAT2-StringTie pipeline was employed as a confirmatory validation of the estimates derived by Kallisto and to detect novel transcripts.

2.2.8 Network cluster analysis

Miru (Kajeka Ltd, Edinburgh UK), a network analysis program previously used for the analysis of large transcriptomic datasets including sheep (Clark, Bush et al. 2017) and pig (Freeman, Ivens et al. 2012), was used to analyse the gene expression data resulting from this project. The Miru algorithm is based on an earlier program, BioLayout *Express*^{3D} (Theocharidis, van Dongen et al. 2009). The algorithm determines similarities between individual gene expression profiles by calculating a Pearson correlation matrix and rendering the output as a network graph in an interactive three-dimensional graphical interface. Integrated within Miru is the Markov Cluster Algorithm (MCL) (Dongen and Abreu-Goodger 2012), that divides the network graph into co-

expression clusters in which all of the members are correlated to all others at a pre-determined correlation coefficient. As discussed above, genes coding for proteins in the same biological pathway or cell type tend to be located within co-expression clusters. As such, it is possible to infer function of an unknown gene based on the known functions of genes with which it co-expresses (Oliver 2000, Mabbott, Baillie et al. 2010, Freeman, Ivens et al. 2012).

For data validation, and to ensure all samples were correctly labelled, firstly the output from Kallisto (Appendix A) was saved as a '.expression' file and loaded into Miru where it was transposed and then clustered as a sample-to-sample network graph using a Pearson correlation co-efficient of r=0.75 and a Markov Cluster Algorithm (MCL) inflation value of 2.2. No spurious samples were identified. Post validation, the '.expression' file was manually curated using an iterative approach to ensure that a biologically relevant number of nodes and clusters was achieved. A suitable Pearson correlation co-efficient of r=0.83 and MCL inflation value of 2.2, pre-inflation of 6.0, and minimum 20 nodes per cluster, was applied and the gene-to-gene network graph generated. Lowly expressed transcripts (≤ 1 TPM) were excluded from the gene-to-gene graph.

2.2.9 Cluster annotation

Co-expression clusters were annotated visually based on the observed tissue or cell type specific gene expression patterns in Miru. Gene members in each cluster were searched for manually against the GeneCards database (Stelzer, Rosen et al. 2016) and Protein Analysis Through Evolutionary Relationships (PANTHER) (Mi, Dong et al. 2010). Each cluster was assigned a functional class and sub-class. This manual annotation method was supplemented by running Gene Ontology (GO) term enrichment analysis (Ashburner, Ball et al. 2000) for the top 30 largest clusters using the R package topGO (Alexa and Rahnenfuhrer 2016). Finally, cluster annotations were confirmed by broadly comparing gene content with equivalent tissue and cell-specific clusters in other large-scale gene expression atlases for sheep (Clark, Bush et al. 2017) and pig (Freeman, Ivens et al. 2012).

2.3 Results

2.3.1 Establishing the quality of RNA

The ratio of absorbance at 260nm and 280nm (A260/280) was used to assess the purity of RNA, with a value of approximately 2.0 indicating 'pure' RNA and a lower ratio indicating presence of protein contaminants. As a secondary measure of purity, the ratio of absorbance at 260nm and 230nm (A260/230) was assessed. Pure RNA has absorbance values ranging between 2.0 and 2.2, with lower values indicating contamination with carry-over phenol. As illustrated in Table 2-1, the A260/280 and A260/230 ratios for all the RNA samples isolated in this study were within the acceptable range and the RNA purity met the recommended requirement. Further, the integrity of the RNA assessed using the Agilent 2200 TapeStation machine (Appendix C) indicated that all samples had RIN values ranging between 7 and 10 reflecting a high level of purity.

2.3.2 Scope of the goat atlas dataset, sequencing depth and coverage

The goat gene expression atlas dataset includes 54 mRNA-Seq (poly-A selected) 75bp paired-end libraries spanning all major organ systems (

Table **2-2**). Approximately 8.7×10^8 paired end sequence reads were generated in total. The tissues and cell-types sampled were all transcriptionally complex, with each expressing at least 50% of the total protein coding genes at detectable levels (Figure 2-1), ranging from alveolar macrophages which had the lowest transcriptional complexity (56%) to the testes which had the highest (75%) (Appendix G). 18,528 protein coding genes (out of a possible 21,343) had detectable expression (TPM>1) in at least one tissue, representing 90% of the reference transcriptome, ARS1 (Bickhart, Rosen et al. 2017). This goat gene expression atlas expands on previously available goat RNA-Seq datasets (Dong, Xie et al. 2013, Bickhart, Rosen et al. 2017) by adding a new set of seventeen tissues and three immune cell types.

Table 2-1: Quantity and Quality measurements of isolated RNA from all tissue and cell-types

Sample No.	Sample ID	ng/ul	RIN	260/280	260/230	Date of Extraction
Liver						
1	York Goat 1	178.68	9.50	2.12	1.99	15/02/16
2	York Goat 2	2908.65	8.60	2.08	2.11	15/02/16
3	York Goat 3	752.31	8.60	2.07	2.25	15/02/16
4	York Goat 4	319.87	9.20	2.10	2.09	15/02/16
Spleen						
5	York Goat 1	4073.05	7.70	2.04	2.11	23/02/16
6	York Goat 2	2538.46	7.30	2.08	2.28	23/02/16
7	York Goat 4	2120.42	7.60	2.09	2.34	23/02/16
Testes						
8	York Goat 1	4455.89	9.10	1.85	2.05	17/02/16
9	York Goat 2	3429.12	9.40	2.07	2.21	17/02/16
10	York Goat 3	2046.73	9.30	2.09	2.31	17/02/16
11	York Goat 4	2448.05	9.40	2.10	2.28	17/02/16
Muscle						
12	York Goat 1	1852.97	8.90	2.10	2.31	01/03/16
13	York Goat 2	1337.97	8.80	2.09	2.31	01/03/16
14	York Goat 4	1803.1	9.10	2.09	2.24	01/03/16
Adrenal Gland						
15	York Goat 1	1160.78	9.50	2.10	2.30	01/03/16
16	York Goat 2	1005.04	9.10	2.09	2.27	01/03/16

17	York Goat 3	1578.95	9.40	2.10	2.25	01/03/16
18	York Goat 4	2077.97	7.30	2.07	2.32	01/03/16
Rumen						
19	York Goat 3	1568.95	7.70	2.09	2.16	15/02/16
20	York Goat 4	1630.01	8.40	2.09	2.15	15/02/16
Ileum+Peyers Patches						
21	York Goat 1	2100.25	7.50	2.10	2.25	17/02/16
22	York Goat 2	4157.1	7.50	1.95	2.09	17/02/16
Thymus						
23	York Goat 1	4452.07	9.20	1.87	1.95	10/03/16
24	York Goat 2	3733.49	9.80	1.97	2.18	10/03/16
25	York Goat 3	3270.69	9.70	2.04	2.18	10/03/16
26	York Goat 4	3707.73	8.60	2.02	2.19	10/03/16
Kidney Cortex						
27	York Goat 1	3327.04	8.30	2.05	2.20	10/03/16
28	York Goat 2	1805.23	8.90	2.10	2.32	10/03/16
29	York Goat 3	2048.73	7.30	2.10	2.30	10/03/16
30	York Goat 4	2124.14	7.70	2.09	2.31	10/03/16
Large Colon						
31	York Goat 1	1277.45	7.50	2.11	2.28	17/03/16
32	York Goat 2	2024.15	8.90	2.11	2.28	17/03/16
33	York Goat 3	1716.41	9.00	2.10	2.30	17/03/16
34	York Goat 4	1913.33	8.30	2.11	2.26	17/03/16

Skin						
35	York Goat 1	262.94	7.00	2.09	2.26	17/03/16
36	York Goat 2	188	8.10	2.08	2.19	17/03/16
37	York Goat 3	312.48	8.10	2.08	1.84	17/03/16
38	York Goat 4	361.8	7.60	2.07	1.97	17/03/16
BMDM						
39	York Goat 2 0hr	392	10.00	2.07	2.07	30/11/16
40	York Goat 2 7hr	322	10.00	2.06	1.87	30/11/16
41	York Goat 3 0hr	286	10.00	2.08	2.05	30/11/16
42	York Goat 3 7hr	264	10.00	2.03	1.85	30/11/16
43	York Goat 4 0hr	520	9.60	2.10	2.13	30/11/16
44	York Goat 4 7hr	548	9.90	2.08	1.93	30/11/16
AM						
45	York Goat 6 AM	169	9.80	1.98	1.01	20/04/17
46	York Goat 7 AM	159	9.70	2.02	1.93	20/04/17
Cerebellum						
47	York Goat 6	1790	8.30	2.11	2.37	24/04/17
48	York Goat 7	1260	8.50	2.13	2.39	24/04/17

Cortex						
49	York Goat 6	1350	8.50	2.10	2.34	24/04/17
50	York Goat 7	664	7.00	2.13	2.29	24/04/17
Reproductive organs						
51	York Goat 8 Ovaries	1320	8.70	2.12	2.27	03/05/17
52	York Goat 8 Uterus	3420	8.40	2.11	2.26	03/05/17
53	York Goat 8 Fallopian tube	3140	9.30	2.11	2.29	03/05/17
54	York Goat 8 Uterine horn	1670	8.60	2.11	2.24	03/05/17

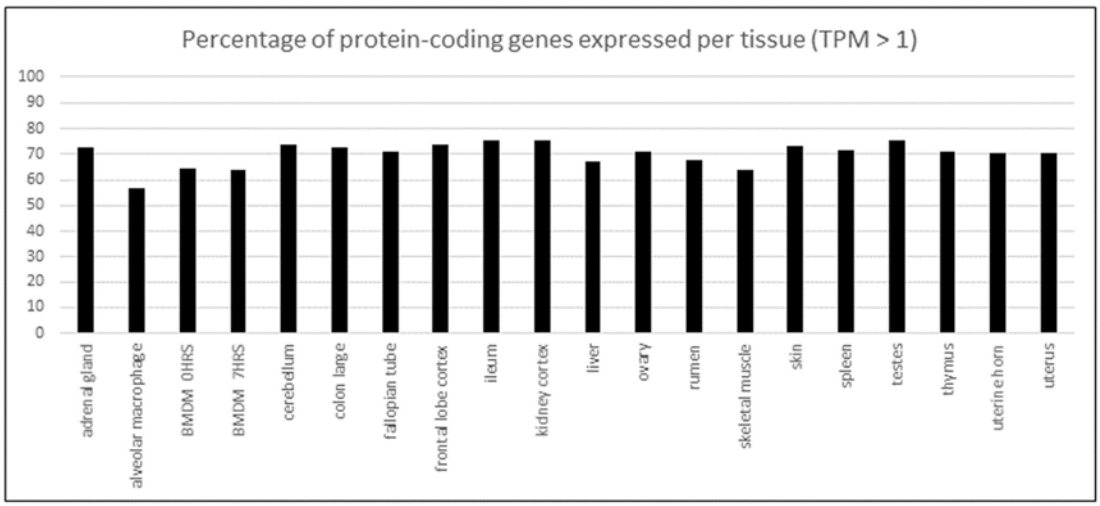


Figure 2-1: Percentage of protein coding genes per tissue. All tissue and cell-types sampled for this project were highly transcriptionally active and expressed >50% of the total protein coding genes

Table 2-2:Details of tissues and cell-types sampled

Tissue/Cell type	Organ System	Protein-coding genes expressed (TPM > 1)	% of protein-coding genes per tissue	No. of replicates
Adrenal gland	Endocrine system	13655	72.39	4
Alveolar macrophage	Immune system	10659	56.51	2
BMDM 0 HRS	Immune system	12193	64.64	3
BMDM 7HRS	Immune system	12038	63.82	3
Cerebellum	Nervous system	13899	73.69	2
Colon large	Digestive system	13679	72.52	4
Fallopian tube	Reproductive system(female)	13432	71.21	1
Frontal lobe cortex	Nervous system	13880	73.59	2
Ileum + Peyers patches	Digestive + Lymphatic system	14173	75.14	2
Kidney cortex	Urinary system	14161	75.08	4
Liver	Endocrine system	12630	66.96	4
Ovary	Reproductive system(female)	13351	70.78	1
Rumen	Digestive system	12815	67.94	2
Skeletal muscle	Muscular system	12011	63.68	3
Skin	Integumentary system	13785	73.08	4
Spleen	Immune system	13530	71.73	3

Testes	Reproductive system(male)	14244	75.52	4
Thymus	Lymphatic system	13403	71.06	4
Uterine horn	Reproductive system(female)	13292	70.47	1
Uterus	Reproductive system(female)	13286	70.44	1

2.3.3 Gene annotation and novel transcript discovery

The proportion of transcripts (lncRNA, protein coding, pseudogenes etc), with detectable expression (TPM >1) in the goat atlas relative to the ARS1 reference transcriptome is summarised at the gene level in Appendix H and at the transcript level in Appendix I. Of the 21,343 protein coding genes in the ARS1 reference transcriptome 7036 (33%) had no informative gene name. Using the HISAT2 annotation pipeline used for the sheep atlas project and described in (Clark, Bush et al. 2017) informative gene names were assigned to 1114 (15%) of the previously un-annotated protein coding genes in ARS1. These genes were annotated by reference to the NCBI non-redundant (nr) peptide database v77 (Pruitt, Tatusova et al. 2007). A short-list containing a conservative set of gene annotations, to HGNC (HUGO Gene Nomenclature Committee) gene symbols, is included in Appendix J. Appendix K contains the full list of genes annotated using the goat atlas dataset and annotation pipeline developed for the sheep atlas project (Clark, Bush et al. 2017). Many unannotated genes can be associated with a gene description, but not necessarily an HGNC symbol; these are also listed in Appendix L. Assigned gene names were manually validated on the full list using network cluster analysis and the ‘guilt by association’ principle.

2.3.4 Network cluster analysis

The gene expression estimates generated by Kallisto (Appendix A), were used to run the network cluster analysis in Miru. First a sample-to-sample network graph (Figure 2-2) was generated, which comprised of 54 nodes (each node represents a library generated for this project) connected by 213 edges (correlations above the set Pearson correlation coefficient R=0.75). All biological replicates clustered closely together, as expected, resulting in seven tissue-specific clusters and two cell-specific clusters. One mixed cluster containing male and female reproductive organs, kidney and lymphatic tissue was observed.

The gene-to-gene network graph generated by Miru, as described in section 2.2.8, was large and highly structured (Fig 2.3). It comprised of 16,172 nodes (transcripts) connected by 1,574,259 edges (correlations above the set threshold, Pearson correlation coefficient of r=0.83). Based on the MCL algorithm, (MCL inflation value of 2.2), the gene network graph separated into 75 distinct co-expression clusters, with the largest cluster (Cluster 1) comprising of 1592 transcripts (Fig 2.4). Only about 759 transcripts (4%) of the detectable transcripts (16,931) failed to cluster at this correlation (r=0.83) and as such were excluded from the resulting gene-to-gene network graph.

The profile and gene content of the top thirty clusters, based on size, was examined in detail and clusters assigned a functional class based on the tissue or cell-types in which genes within each cluster had the highest expression (Table 2-3). GO term enrichment with TopGO was used to validate the assignment of functional class. A list of GO terms associated with each cluster are contained in Appendix B.

Table 2-3:Tissue/cell/pathway association of the 30 largest network clusters in the goat gene expression dataset

Cluster ID	Number of Transcripts	Profile Description	Class	Sub-class
1	1592	Brain (Cortex>cerebellum)	CNS	cortex, neuronal function, synaptic transmission
2	1017	Thymus>Spleen>Ileum	Pathway	Cell-cycle
3	972	General	House Keeping	House Keeping 1, gene regulation
4	467	Liver	GI tract	Liver, cholesterol homeostasis
5	454	Testes	Reproduction	gamete generation, spermatogenesis
6	405	Brain (Cerebellum>cortex)	CNS	cerebellum, signal transduction
7	398	Skin	Skin	keratinocyte differentiation, skin development
8	363	Fallopian Tube	Reproduction	motile cilia, flagella, ciliogenesis
9	353	Skeletal muscle	Musculature	skeletal muscle
10	281	Spleen>Ileum	Immune	B-cell activation, cytokine binding
11	244	Rumen>Skin	GI tract	ruminal epithelium, epidermis morphogenesis
12	243	Colon	GI tract	glycosylation, microvillus
13	206	Adrenal Gland	Endocrine	steroid hormone biosynthesis
14	191	Fibroblasts	Fibroblasts	angiogenesis, fibroblast proliferation
15	160	Macrophages>Spleen	Immune	phagocytosis, pathogen recognition
16	158	Kidney	Early Development	kidney development

17	155	General (highest in macrophages)	House Keeping	House Keeping 2, molecular chaperones
18	129	Ovary	Reproduction	embryogenesis, sperm-egg recognition
19	125	General	House keeping	House Keeping 3
20	122	Kidney	Renal	uptake of intestinal bile acids, hepatic lipid metabolism
21	105	General	Pathway	Krebs cycle, ROS metabolism
22	102	General	House Keeping	House Keeping 4, polyubiquitination
23	100	General	House Keeping	House Keeping 5, protein synthesis
24	97	Uterus	Reproduction	late embryogenesis, fertilization
25	94	Ileum	immune	B-cell signalling, lymphatic system
26	91	General	Pathway	Ribosome biogenesis
27	86	LPS-inducible	Immune	LPS-binding, inflammatory response, cytokine signalling
28	81	General	Pathway	Mitochondrial
29	67	General but not even	Early Development	Development
30	65	General but not even	House Keeping	House Keeping 6

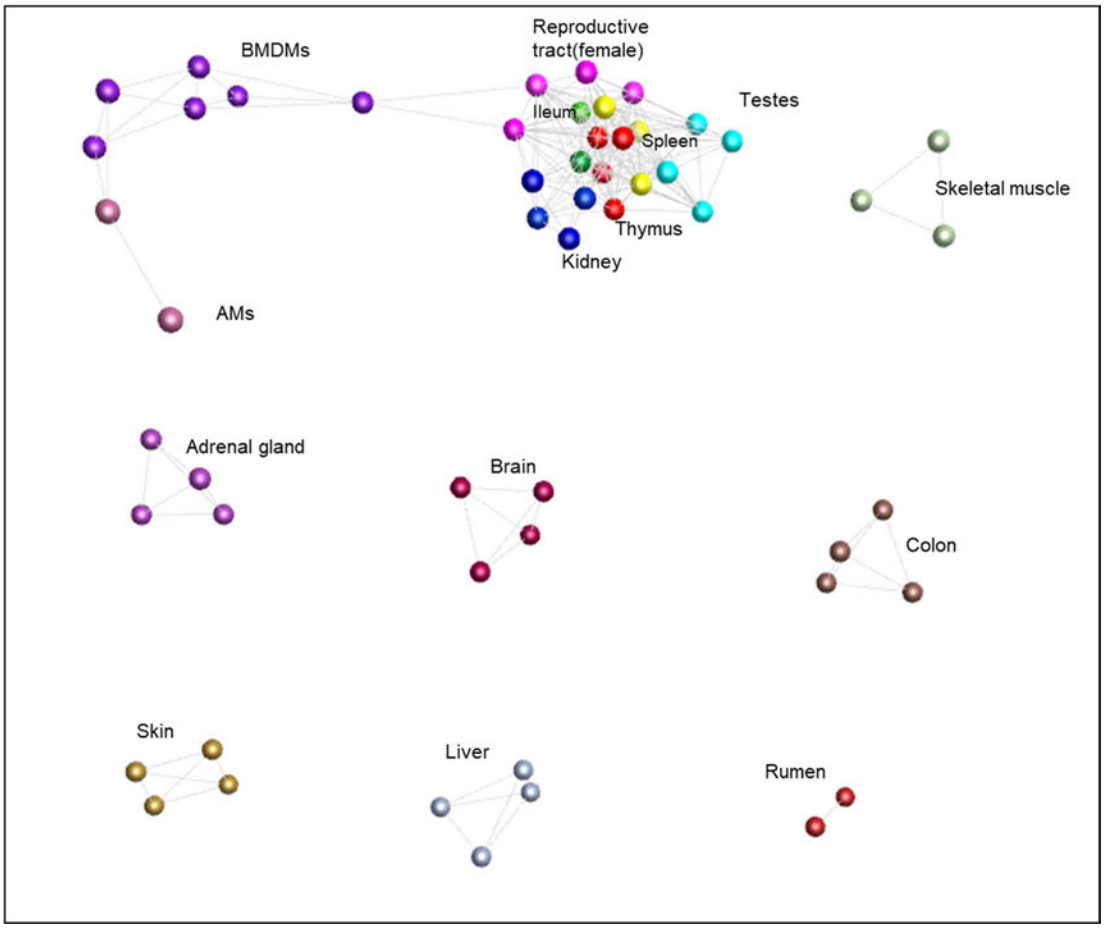


Figure 2-2:Sample to Sample network graph. The network graph shown here is the representation of the sequence libraries generated for the goat gene expression atlas project. Each node represents an individual tissue sample, and edges represent the connections between the samples, resulting in biological replicates clustering together to form a tissue cluster. The graph is comprised of 54 nodes and 213 edges

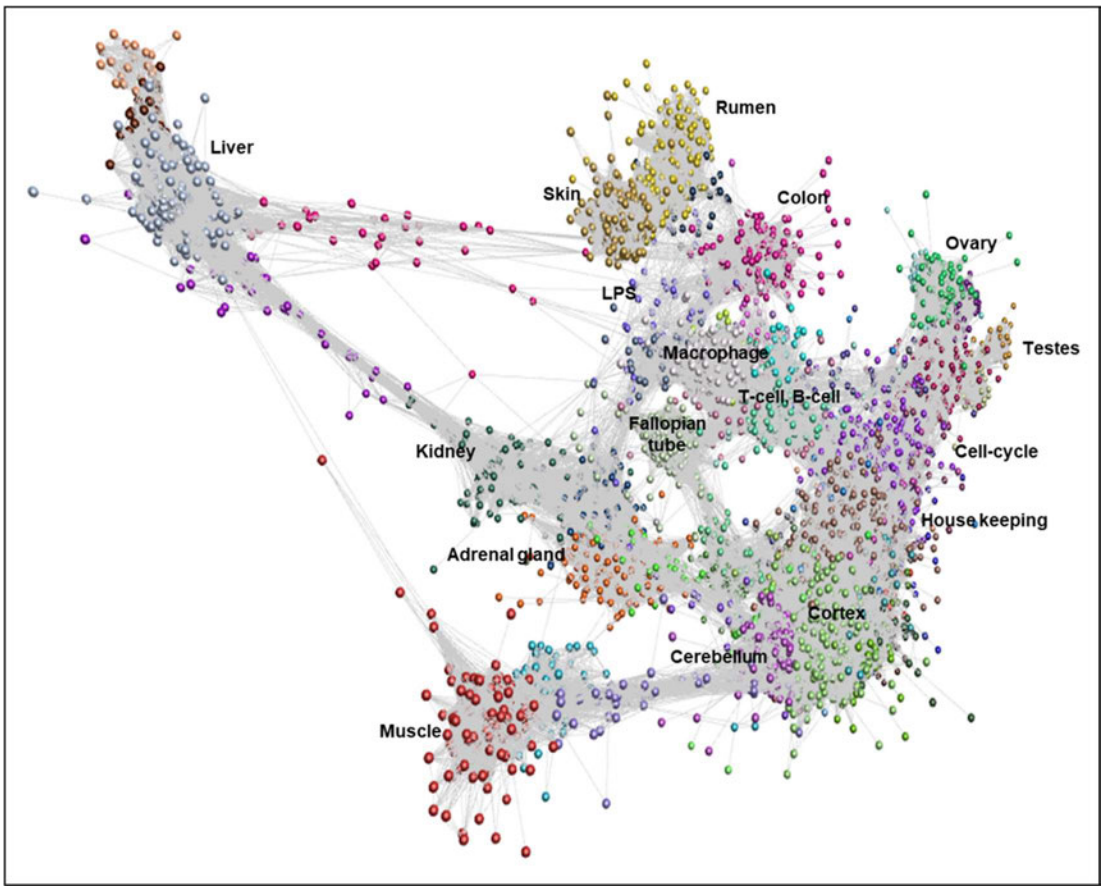


Figure 2-3:Gene to Gene network graph. A three-dimensional visualization of the Pearson correlation gene-to-gene graph of the expression data derived from analysis of goat tissues and cells. Each node represents a gene and the edges (lines) correspond to correlations between individual measurements above the set threshold. The graph is comprised of 16,172 nodes (genes) and 1,574,259 edges (correlations ≥ 0.83). Using the MCL algorithm, groups of similarly expressed genes were connected forming tissue-specific co-expression clusters.

2.3.5 Tissue-specific co-expression clusters

The majority of co-expression clusters comprising the goat gene atlas contained sets of genes showing uniquely high expression signals in a specific tissue or cell-type (Fig 2.4A) e.g. liver (Fig 2.4B (i)), colon (Fig 2.4B (ii)) and muscle (Fig 2.4B (iii)). In each case, the likely function of unannotated protein-coding genes within these clusters could be inferred by association with genes of known function that share the same cell or tissue specific expression pattern, through the principle of 'guilt-by-association'. As observed in previous gene atlas studies such as sheep (Clark, Bush et al. 2017) and pig (Freeman, Ivens et al. 2012), some clusters, show very little evidence of tissue specificity and are commonly labelled as housekeeping clusters. These clusters contain a set of transcripts required for 'generic' cellular functions that are expressed at similar levels in all mammalian cells.

The largest of the clusters (Cluster 1) contained 1,592 genes that were almost exclusively expressed in the central nervous system (cortex, cerebellum) reflecting the high transcriptional activity and complexity in the brain. Significant GO terms for cluster 1 included synaptic transmission ($p<1.0\times10^{-30}$), axon development ($p=2.2\times10^{-22}$), locomotory behaviour ($p=3.10\times10^{-19}$), GABA receptor activity ($p=1.2\times10^{-7}$) and neurotransmitter receptor activity ($p=2.1\times10^{-5}$). Some other clusters had highly tissue specific expression patterns, including Cluster 9 (Fig 2.4B (iii)), which included genes associated with skeletal muscle function and development. One gene found within cluster 9 was *MSTN* which encodes a protein that negatively regulates skeletal muscle cell proliferation and differentiation (Wang, Yu et al. 2012). Several myosin light and heavy chain genes (e.g. *MYH1* and *MYL1*) and transcription factors that are specific to muscle including *MYOG* and *MYOD1* were also found in cluster 9. GO terms for muscle were enriched in cluster 9 e.g. myofibril assembly ($p=4.7\times10^{-18}$) skeletal muscle tissue development ($p=5.5\times10^{-18}$), muscle system process ($p=1.4\times10^{-16}$), structural constituent of muscle ($p=2.9x10^{-8}$) and actin cytoskeleton ($p=5.2\times10^{-12}$). Genes expressed in muscle are of particular biological and commercial interest for livestock production and represent potential targets for gene editing (Yu, Lu et al. 2016). Cluster 8 also

had a highly tissue specific expression pattern and included genes expressed in the fallopian tube with enriched GO terms for cilium morphogenesis ($p=1.2\times10^{-14}$), microtubule-based movement ($p=9.7\times10^{-12}$), motile cilium ($p=1.5\times10^{-9}$) and microtubule organizing centre part ($p=2.5\times10^{-5}$). A motile cilia cluster was identified in the fallopian tube in the sheep gene expression atlas (Clark, Bush et al. 2017) and a similar cluster was enriched in chicken in the trachea (Bush, Freem et al. 2018).

Other tissue-specific clusters; e.g. 4 (liver) Fig 2.4Bi and 12 (colon) Fig 2.4Bii, 25 (ileum), 11 (rumen) and 15 (macrophages) were similarly enriched for genes associated with known tissue-specific functions and these clusters are described in more detail below.

2.3.5.1 Liver cluster

Cluster four (Fig 2.4B (i)) contained 467 transcripts showing highly unique expression in the liver, of which 336 had an informative annotation and 131 had placeholder LOCUS IDs. The identity of the cluster was validated by expression of known liver-specific genes including serum albumin (*ALB*). The over-represented gene sets tested using Gene Ontology (GO) term enrichment analysis (biological process) (Appendix B) were consistent with known metabolic functions of the liver. A significant proportion of genes in this cluster were involved in such processes as oxidation-reduction process ($p=5.7\times10^{10}$), aromatic amino acid family ($p=3.4\times10^{-10}$), cholesterol homeostasis ($p=4.4\times10^{-9}$) and plasminogen activation ($p=1.0\times10^{-8}$). Similarly, the GO terms (molecular function) associated with this cluster included peptidase inhibitor activity ($p=2.8\times10^{-8}$), serine-type endopeptidase activity ($p=9.10\times10^{-8}$), lipoprotein particle receptor binding ($p=2.0\times10^{-6}$) and fatty acid ligase activity ($p=6.7\times10^{-6}$), while GO terms (cellular component) included extracellular vesicular exosome ($p=4.3\times10^{-20}$), high-density and very-low-density lipoprotein particle at ($p=6.7\times10^{-12}$) and ($p=6.2\times10^{-10}$) respectively. This cluster contained genes encoding numerous serine protease inhibitors (SERPINS) such as *SERPINA10, SERPINA11, SERPINA7, SERPINC1, SERPIND1, SERPINF1* and *SERPINF2,* a considerable number of genes in

the solute carrier (*SLC*) gene family including *SLC10A1*, *SLC27A2, SLC27A5, SLC2A2, SLC30A10, SLC39A5 and SLCO2B1* and notable rate-limiting gluconeogenic genes such as *PCK, FBP1* and *G6PC.* The genes encoding growth hormone receptor (*GHR*) and its targets including insulin like growth factor binding acid labile subunit (*IGFALS*) and insulin like growth factor binding protein (*IGFBP4*) were also enriched in cluster 4. Growth promoting hormones play an essential role in hepatic liver metabolism (Fan, Menon et al. 2009) and are therefore likely to be enriched in the liver of neonatal goats.

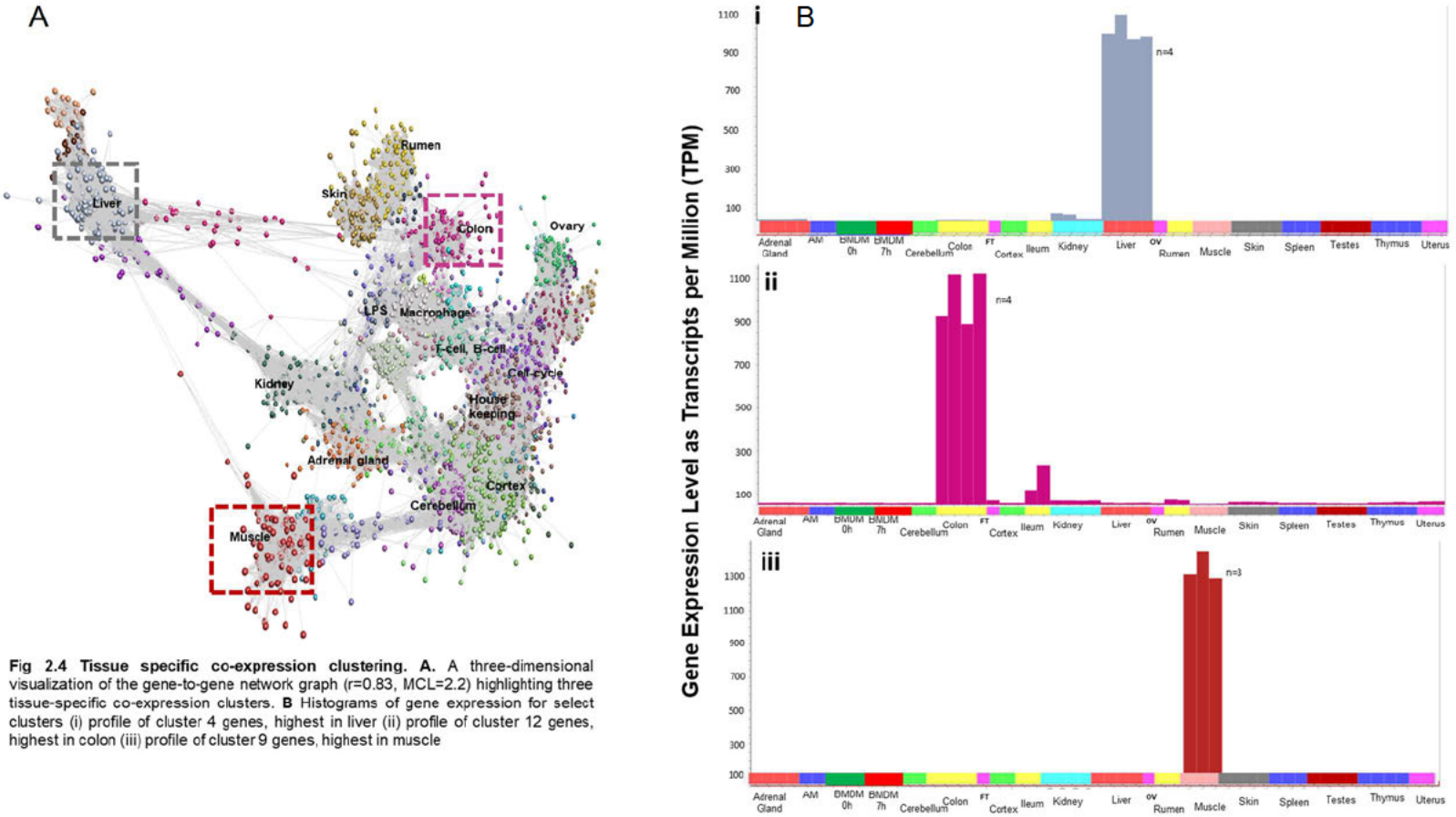


Fig 2.4 Tissue specific co-expression clustering. A. A three-dimensional visualization of the gene-to-gene network graph (r=0.83, MCL=2.2) highlighting three tissue-specific co-expression clusters. **B** Histograms of gene expression for select clusters (i) profile of cluster 4 genes, highest in liver (ii) profile of cluster 12 genes, highest in colon (iii) profile of cluster 9 genes, highest in muscle

Figure 2-4:Tissue specific co-expression clustering. A. A three-dimensional visualization of the gene-to-gene network graph (r=0.83, MCL=2.2) highlighting three tissue-specific co-expression clusters. B Histograms of gene expression for select clusters (i) profile

2.3.5.2 **The gastrointestinal (GI) tract**

The gastrointestinal tract tissues included in the goat expression atlas separated into three main clusters based on the three regions of the digestive system sampled: colon, ileum and Peyer's patches and rumen. This separation of clusters was similar to that observed in a detailed analysis of the transcriptional signatures in the gastrointestinal tract of sheep (Bush, McCulloch et al. 2019). Each cluster was inspected individually to highlight genes and associated biological processes related to GI tract function as described below.

2.3.5.2.1 Ileum and Peyer's patches cluster

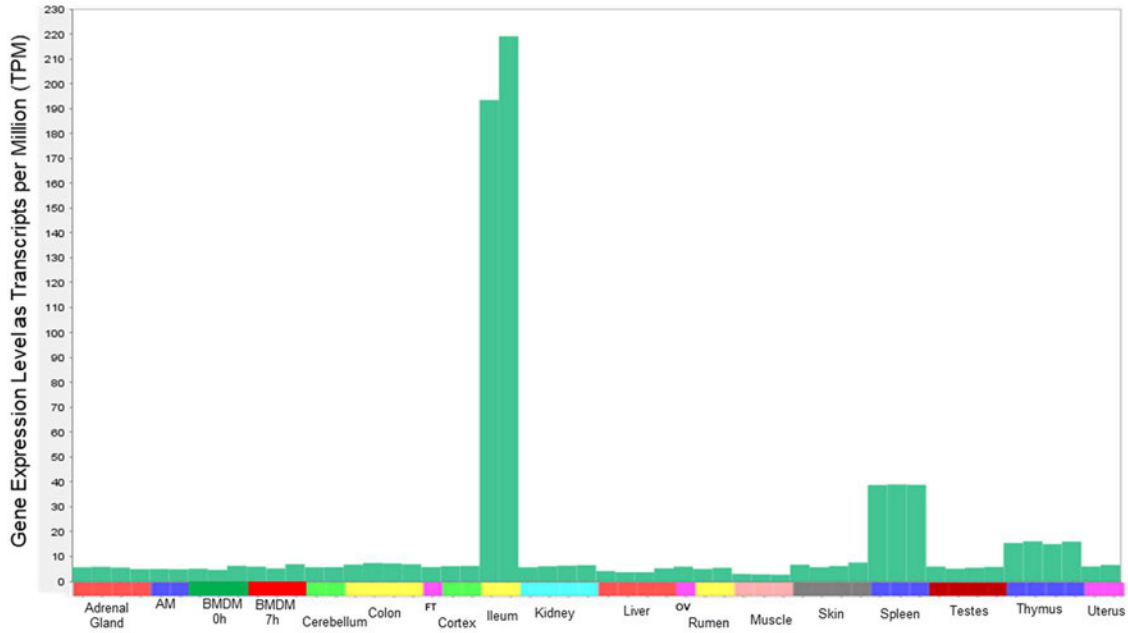


Figure 2-5:Ileum cluster. Expression profile of the Ileum-specific cluster

2.3.5.2.2 Colon cluster

The gene expression pattern uniquely associated with the colon was observed in cluster 12 (Fig 2.4 (ii)). Gene set enrichment analysis revealed GO terms (biological processes) such as microvillus organization ($p= 1.3\times10^{-6}$), positive regulation of peptide hormone secretion ($p= 9.6\times10^{-5}$) and inorganic anion transmembrane transport ($p= 1.7\times10^{-4}$). The molecular functions of GO

associated with the colon cluster included serine-type endopeptidase inhibitor activity ($p= 2.2\times10^{-4}$), ion gated channel activity ($p= 7.43\times10^{-3}$) and acetylglucosaminyltransferase activity ($p= 9.5\times10^{-3}$) while cellular components comprised of brush border ($p= 7.6\times10^{-4}$), golgi membrane ($p= 6.4\times10^{-4}$), microvillus ($p= 6.0\times10^{-4}$), tight junction ($p= 9.6\times10^{-3}$) and myosin complex ($p= 2.2\times10^{-2}$). Notable genes contained in the colon cluster included phospholipid-transporting ATPases *ATP10B* and *ATP8B1*, mucin family genes such as *MUC13* and *MUC2*, as well as protein disulphide isomerases (PDI) *AGR2* and *AGR3.* Additionally, members of the myosin superfamily of genes including *MYO15B*, *MYO1A*, *MYO7B*, keratins such as *KRT19, KRT20, KRT8 and KRTCAP3,* a considerable number of *SLC* genes such as *SLC13A2, SLC6A14, SLC12A2, SLC22A16, SLC26A3, SLC35A1* as well as membrane bound glycoprotein *B3GALT5* and beta galactosyltransferases *B3GNT3, B3GNT6* and *B3GNT7* were enriched in the colon.

2.3.5.2.3 Rumen cluster

Cluster 11 included 244 genes showing very high expression in the rumen and minimal expression in the skin (Fig 2.6). The rumen epithelium and skin share similar transcriptional signatures as has been documented in other studies (Xiang, McNally et al. 2016, Bush, McCulloch et al. 2019). Some of the GO terms (biological processes) associated with this cluster included keratinocyte differentiation ($p= 9.5\times10^{-5}$) and establishment of skin barrier ($p= 3.3\times10^{-5}$), which are characteristic of a rumen epithelial signature. The molecular functions associated with the rumen cluster included such terms as serine-type peptidase activity ($p= 1.3\times10^{-4}$), peroxidase activity ($p= 2.68\times10^{-3}$) and phosphatidic acid binding ($p= 7.72\times10^{-3}$) while cellular components included cornified envelope ($p= 1.5\times10^{-4}$) and basal plasma membrane ($p= 1.68\times10^{-2}$). In concordance with the biological processes and molecular functions associated with the rumen cluster, notable genes enriched genes included members of the keratin family such as *KRT4*, *KRT6A*, *KRT78* and *KRTDAP*, immune genes including *IL17C*, *IL20RA* and *IL36B* and the *SLC* gene *SLC9A2 (*also known as *NHE-2).*

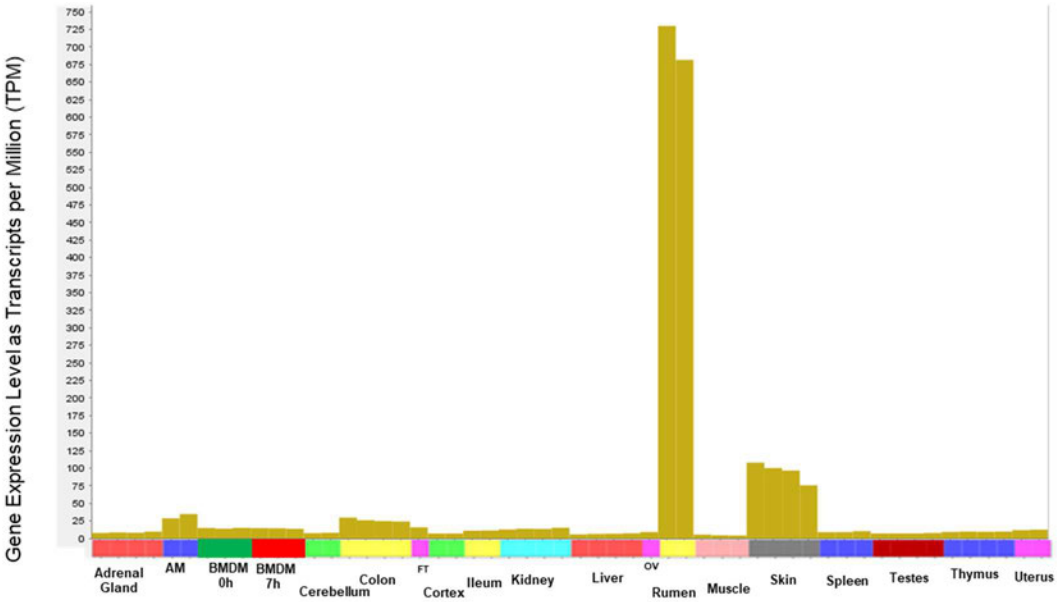


Figure 2-6:Expression profile of cluster 11 highest in the Rumen, with detectable signal in the skin

2.3.6 Cellular processes

In addition to the tissue specific clusters, some of which are described in section 2.3.5 above, a few clusters showed ubiquitous expression across multiple tissues, albeit not at the same level in all tissues, indicating they are involved in a diverse range of cellular processes. For example, the second largest cluster (cluster 2) (Fig 2.7) contained 1017 transcripts with detectable expression across all tissues and cells but showed strong enrichment in the thymus, spleen and ileum. As revealed by gene set over-representation analysis (Appendix B), the GO terms (biological process) significantly enriched in cluster 2 included mitotic sister chromatid segregation ($p= 3.5\times10^{-19}$), DNA repair ($p= 1.2\times10^{-13}$) and centrosome cycle ($p= 8.6\times10^{-10}$). Some of the GO terms (molecular function) associated with this cluster included histone binding ($p= 3.7\times10^{-10}$) and DNA-directed DNA polymerase activity ($p= 3.9\times10^{-5}$) while GO terms (cellular component) included nuclear body ($p= 8.2\times10^{-19}$), chromosome ($p= 4.0\times10^{-16}$) and spindle ($p= 1.2\times10^{-17}$). Notable genes contained in cluster 2 included several RHO GTPases e.g. *ARHGAP15, ARHGAP19, ARHGAP30, ARHGAP4 and ARHGAP45* involved in cell

migration, proliferation, and differentiation, actin remodelling, and G1 cell cycle progression, high mobility group (HMG) genes *HMGN2*, *HMGN5, HMGB1* and *HMGB2* and cyclins *CCND3* and *CCNJL.* The cluster appears enriched for transcripts associated with cell proliferation which are enriched to the greatest extent in immune tissues (e.g. thymus, spleen and ileum). The frequency of cells undergoing mitosis in these tissues, particularly the Peyer's patches, ileum and thymus of neonatal goats might explain the association of cell cycle genes with an immune signature (David, Norrman et al. 2003). As expected with large co-expression clusters, numerous genes contained in cluster 2 lacked meaningful names and were annotated with LocusLink (LOC) gene identifiers.

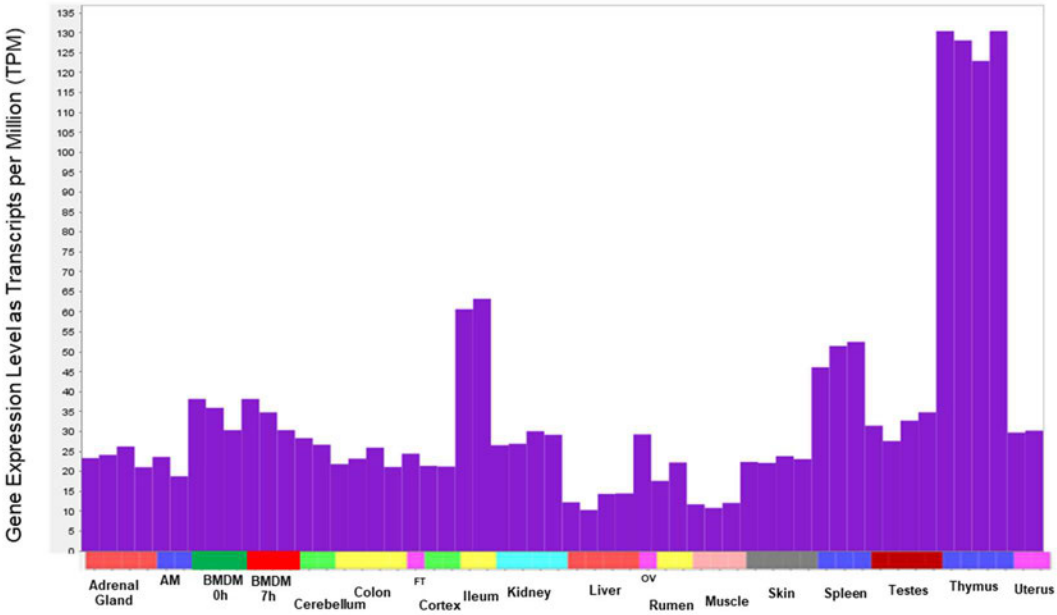


Figure 2-7:Expression profile of the cell-cycle cluster 2 with detectable expression across all tissue but strongly enriched in the thymus, spleen and ileum

2.3.7 **Macrophage-associated clusters**

Some clusters showed strong association with the alveolar and bone marrow derived macrophage cells. Cluster 15 (Fig 2.8) showed highly specific and unique expression pattern in macrophages, with a stronger signature in the alveolar macrophages relative to bone-derived macrophages. The GO terms (biological process) associated with it included integrin-mediated signalling pathway ($p= 3.1\times10^{-5}$), osteoclast differentiation ($p= 2.4\times10^{-4}$) and phagocytosis ($p= 7.0\times10^{-4}$), GO terms (molecular function) included cytokine receptor activity ($p= 1.9\times10^{-3}$) and hydrogen-exporting ATPase activity ($p= 5.5\times10^{-3}$), while the GO terms (cellular component) included immunological synapse ($p= 9.7\times10^{-5}$) and lysosome ($p= 1.8\times10^{-5}$). Colony stimulating factor *CSF1* and *CSF2RA* encoding the receptor for *CSF2* which is required specifically for AM differentiation (Shibata 2001, Schneider, Nobs et al. 2014) and *CD44* which promotes the survival of fetal bone-marrow monocyte derived alveolar macrophages (Dong, Poon et al. 2018) were both found within cluster 15. Other notable genes contained in cluster 15 include, myeloid differentiating protein *LY96, CXCR4*, *IL7R*, mitogen-activated protein kinase *MAPKAPK3,* scavenger receptor *CD84,* transcription factor *MITF* and toll-like receptor *TLR8.* Genes that encode major C-type lectins, a family of genes that play a role in inflammation and immunity such as *CLEC4D*, *CLEC5A* and *CLEC6A* were also found within cluster 15. Many of the genes that were up-regulated in AM in cluster 15, including C-type lectins *CLEC5A* and *CLEC4A*, have been shown to be down regulated in sheep (Clark, Bush et al. 2017, Bush, McCulloch et al. 2019), pigs (Freeman, Ivens et al. 2012) and humans (Baillie, Arner et al. 2017) in the wall of the gut. This highlights functional transcriptional differences in macrophage populations. AM respond to microbial challenge as the first line of defence against inhaled pathogens, while macrophages in the wall of the gut, where this response would be undesirable, down-regulate their response to microorganisms.

Although genes contained in cluster 17 (Fig 2.9) showed detectable expression across all tissues and cells, they exhibited a stronger signal in the alveolar and bone marrow derived macrophages and appeared to be generally

constitutively expressed. The gene set enrichment analysis associated with cluster 17 revealed that the genes were involved in more generalized cellular processes, such as those related to molecular chaperones. Evidently, the GO terms (biological process) enriched in this cluster included ribosome biogenesis ($p= 2.5\times10^{-8}$), regulation of telomere maintenance via telomerase ($p= 1.4\times10^{-6}$) and internal ribosome entry segment (IRES)-dependent translational initiation of linear mRNA ($p= 1.4\times10^{-4}$). GO terms (molecular function) included RNA binding ($p= 2.4\times10^{-24}$), GTP binding ($p= 7.96\text{x}10^{-3}$), and RNA helicase activity ($p= 8.48\times10^{-3}$) while GO terms (cellular component) included nucleolus ($p= 1.0\times10^{-23}$), cajal body ($p= 7.10\times10^{-6}$), and myelin sheath ($p= 1.81\times10^{-3}$).

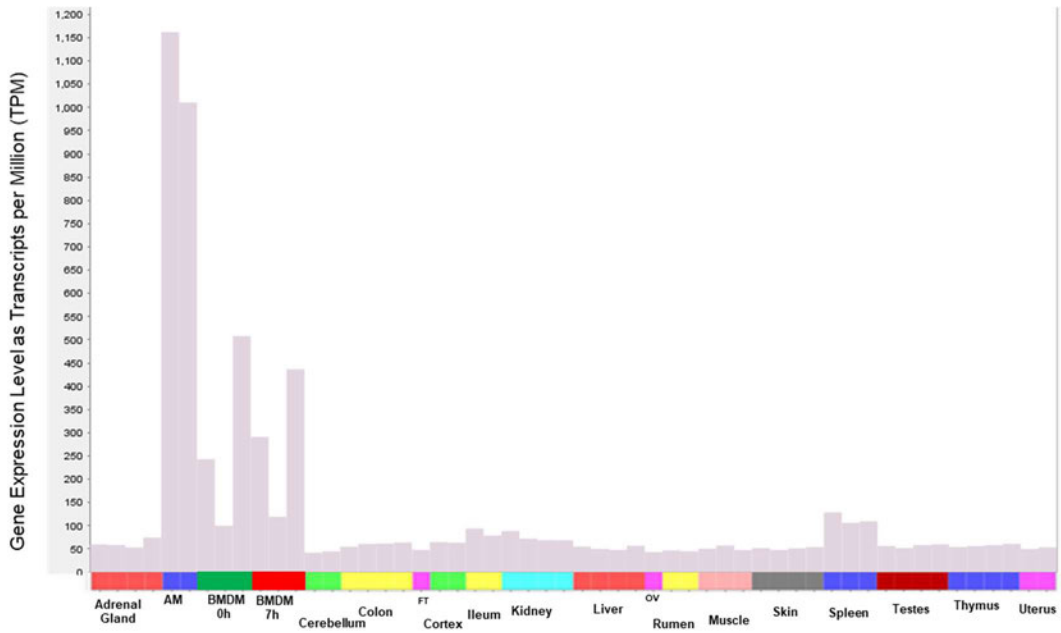


Figure 2-8:Expression profile of the macrophage phagocytic cluster 15. The subset of immune related genes belonging to this cluster showed higher expression in the alveolar macrophages than in the bone-derived macrophages

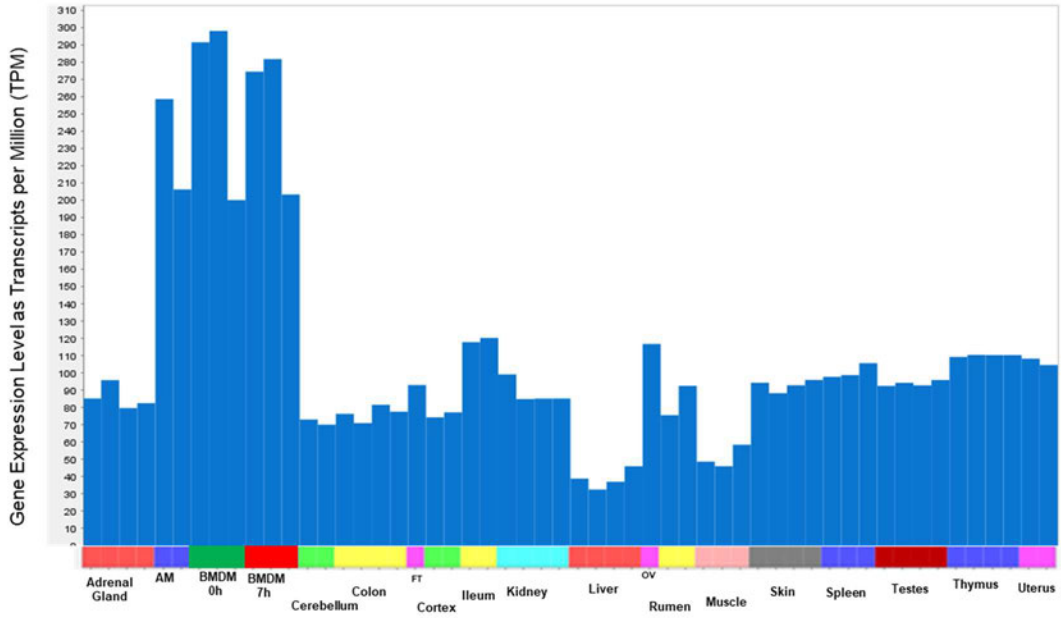


Figure 2-9:Expression profile of cluster 17 showing detectable expression across all tissues and cell types but enriched for in the alveolar and bone-marrow derived macrophages. Majority of the genes contained in this cluster were molecular chaperones

Cluster 27 (Fig 2.10) contains numerous well-studied pro-inflammatory cytokines and chemokines which show induction following challenge of bone marrow derived macrophages with lipopolysaccharide (LPS). The GO terms associated with this cluster included inflammatory response ($p= 1.4\times10^{-10}$), cellular response to lipopolysaccharide ($p= 9.10\times10^{-9}$) and cytokine activity ($p= 4.40\times10^{-10}$). Identification and analysis of LPS-inducible transcripts is described in detail in Chapter 3.

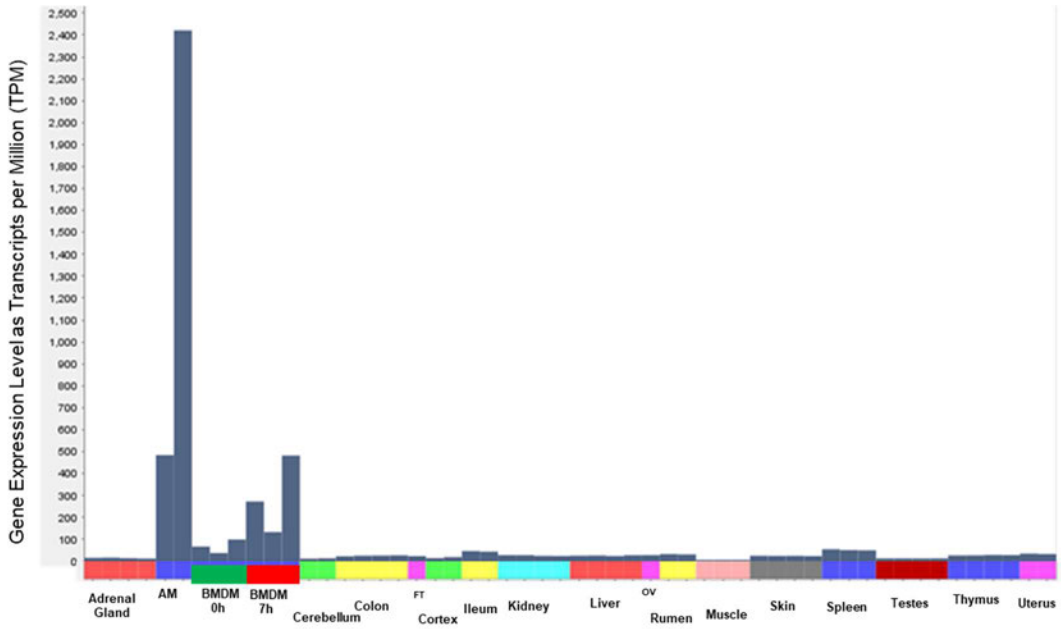


Figure 2-10:Expression profile of the LPS-inducible cluster 27

2.4 Discussion

2.4.1 General Overview

The RNA-Seq data presented in this chapter, generated from six male and one female crossbred dairy goats captured approximately 90% of the goat reference transcriptome by quantifying gene expression across a diverse set of twenty tissue and cell-types, covering all the major organ systems. This dataset includes ten tissues unavailable in previous goat RNA-Seq projects (Dong, Xie et al. 2013, Bickhart, Rosen et al. 2017) contributing a considerable new resource for functional genomics studies in the goat. Most importantly, this dataset provides a resource of tissue-specific gene expression patterns for goat, which can be compared with other livestock.

The overall structure of the goat expression atlas was similar to that observed in other large expression datasets of sheep (Clark, Bush et al. 2017), pig (Freeman, Ivens et al. 2012), human (Ravasi, Suzuki et al. 2010, Andersson, Gebhard et al. 2014) and mouse (Su, Cooke et al. 2002) (Hume, Summers et al. 2010), where transcriptionally active tissues such as liver and brain comprised the largest co-expression clusters. However, the goat testes appeared to be less transcriptionally active than in other species (sheep and pig), presumably because the animals used in this study were neonatal and at this immature stage, the testes have yet to develop the specialized functions, and transcriptional complexity, unique to the adult testes. A detailed comparative analysis of gene co-expression clusters in goat and sheep has been carried out and discussed further in Chapter 4 of this thesis.

The goat reference transcriptome (Bickhart, Rosen et al. 2017) comprised 21,343 transcripts. The goat gene mini-atlas generated by this project captures 18,528 transcripts, approximately 90% of the transcriptome. This provides ‘proof of concept’ that a mini-atlas approach can be utilized for global transcriptomics analysis, by selecting a small sub-set of transcriptionally rich tissues. The remaining 10% (about 2,815 transcripts) are likely to be highly tissue specific genes for a tissue not sampled e.g. from the eye, or transcripts with expression levels was too low to be detected at the sequencing depth

used (>30 million reads per sample). Similarly, the animals used in this project were neonatal from one time point during development and therefore transcripts with highly developmental stage-specific expression would not be captured by this analysis. Additionally, the gene level clustering carried out in Miru, excluded about 4% (750 transcripts) of the detectable transcripts based on their highly unique expression pattern. For example, the *ARG2* gene despite showing substantial expression (estimated at 140 TPM in the alveolar macrophages (Appendix F)) failed to cluster with other immune related genes and was assigned as 'no-class'. Some of these cases arise because individual genes have multiple separate tissue-specific promoters driving expression in different tissues (Anderrson et al. 2014). A notable example in humans is *SERPINA1* encoding alpha-1-antitrypsin, which has separate liver-specific and inducible myeloid-macrophage promoters (Baillie, Arner et al. 2017)). Of note, *SERPINA1* expression is entirely liver-specific in sheep, and goat.

Of the 75 distinct co-expression clusters obtained from the gene level network analysis in Miru, six housekeeping clusters were identified, a common occurrence in large graphs where a relatively low correlation threshold has been employed. The majority of the genes contained in them were unannotated with uninformative LOCIDs. This is similar to observations made in the sheep and pig atlases (Freeman, Ivens et al. 2012, Clark, Bush et al. 2017) and reflects the fact that most functional annotation has focused on tissue-specific genes with important functions leaving the majority of the housekeeping genes uncharacterised. To highlight the utility of the goat gene expression atlas, a sub-set of clusters are discussed in more detail below highlighting their role in ruminant physiology and immunity.

2.4.2 Role of the liver in gluconeogenesis and fatty acid oxidation

The liver is the most important metabolic organ in mammalian animals and plays a crucial role in glucose and fatty acid homeostasis. It is estimated that up to 90% and 75% of glucose requirement in adult and neonatal ruminants respectively is provided by the liver through gluconeogenesis (Nafikov and

Beitz 2007). Hepatic gluconeogenesis is especially important in ruminants due to the uniquely low dietary glucose availability when compared to non-ruminants. From the goat gene expression dataset, high expression of *PCK1, FBP1* and *G6PC* transcripts encoding the rate-limiting gluconeogenic enzymes was detected. As expected, the liver-specific cluster contains *GHR*, encoding the growth hormone receptor, which controls somatic growth in part by regulating liver expression of insulin-like growth factor 1 (*IGF1*) (reviewed in (Chia 2014). *GHR* and its targets including *IGFALS* and *IGFBP4* were also enriched in the liver, which was to be expected as growth promoting hormones play an essential role in hepatic liver metabolism and growth in neonatal animals (Fan, Menon et al. 2009). In dairy cattle the *PDK4* gene, which encodes part of the pyruvate dehydrogenase complex was highly upregulated in the liver, and associated with increased gluconeogenesis from lactate, alanine and pyruvate (Laporta, Rosa et al. 2014). In contrast to the observation in cattle the goat data replicated the observations in the sheep atlas (Clark, Bush et al. 2017) showing higher expression of the *PDK4* gene in the muscle than in the liver, consistent with a role in mitochondrial oxidation of glucose as well as gluconeogenesis.

Peroxisome proliferator-activated receptors *PPAR* are known to be involved in multiple cellular processes including lipid catabolism, glucose homeostasis, inflammation and adipogenesis (reviewed in (Mandard, Muller et al. 2004). Natural *PPAR* ligands include fatty acids and eicosanoids. *PPAR* function as regulators of lipid and lipoprotein metabolism and glucose homeostasis and influence cellular proliferation, differentiation and apoptosis, and is highly expressed in tissues such as liver, muscle, kidney and heart, which are all tissues characterized by a relatively high rate of fatty acid catabolism, where it stimulates the beta-oxidative degradation of fatty acids (Braissant 1996).

PPAR alpha is a master regulator of lipid metabolism in multiple species and has been shown to be highly upregulated in dairy cattle liver especially during winter gestation (Laporta, Rosa et al. 2014). The goat gene expression dataset demonstrates that, as observed in sheep (Clark, Bush et al. 2017), although

expression of *PPAR* alpha gene can be detected in the goat liver, its expression was highest in the kidney. The expression of *PPAR* alpha in both the kidney and liver in goat indicates the kidney is an additional important site for lipid metabolism.

Several solute carriers showed significantly higher expression in the liver. Although the membrane-bound transporters encoded by the solute carrier superfamily have been well demonstrated in humans (Lei 2009), they remain poorly characterised in ruminants. In sheep, as in humans, the high capacity glucose transporter (*SLC2A2*) is most highly expressed in liver, kidney and intestine(Clark, Bush et al. 2017). In the goat atlas, expression was higher in liver than in kidney or intestine, but this may reflect the juvenile status of the animals. Expression of fatty acid (*SLC27A2, SLC27A5),* amino acid (*SLCA14*) and bile acid *(SLC10A1)* transporters, consistent with roles of the liver in fatty acid metabolism, gluconeogenesis and cholesterol homeostasis respectively was detected. Additionally, multiple *SERPINS* such as *SERPINA10, SERPINA11, SERPINA7, SERPINC1, SERPIND1, SERPINF1 and SERPINF2* were detected. *SERPINS* are a family of ubiquitous genes across a wide range of orgasms, involved in diverse functions including inhibition of proteases and controlling proteolytic cascades. Many of these genes encode major plasma protease inhibitors, including inhibitors of enzymes of the coagulation cascade (e.g. *SERPINA10, SERPINC1, SERPIND1*) and have previously been shown to be enriched in human, pig and sheep liver (Law, Zhang et al. 2006). The *SERPINA1* transcript in the goat gene atlas was not initially annotated in the NCBI goat reference transcriptome used in this project (Bickhart, Rosen et al. 2017), instead identified by its locus placement ID *LOC102185401* which as in sheep showed highest expression in the liver.

There are fundamental differences in metabolism across species reflected by the diverse regulatory strategies employed to meet their physiological needs for maintenance and production, which are reflected in transcriptional differences between the species.

2.4.3 Transcriptional profiling of the Goat GI tract

Gene expression analysis of the gastrointestinal tract in the domestic goat is particularly interesting because goats are ruminants (like sheep and cattle) and have a complex digestive system. The sheep (Clark, Bush et al. 2017) and pig (Freeman, Ivens et al. 2012) atlas projects carried out extensive analysis of the gastrointestinal tract, collecting tissue from multiple regions. Despite the limited extent of the GI tract represented within the goat gene atlas (in comparison to sheep and pig), specific marker genes, correlated with the function of the three sampled regions were detected and detailed below.

The GI tract is the largest surface area in the mammalian body exposed to external environment and the lumen of the GI tract is laden with pathogens, antigens and commensal bacteria. As such, it is important that a strong immune response is maintained at the intestinal epithelial barrier. It was therefore, not surprising that many immune related genes were expressed in the gastrointestinal tract. Expression of the majority of immunogenic genes was two- to threefold higher in the ileum and Peyer's patches cluster than in other GI regions. Similar observations were also made in the pig and sheep datasets (Freeman, Ivens et al. 2012, Clark, Bush et al. 2017). For instance, genes encoding the protein components of the B-cell receptor complex such as *CD19, CD22, CD79B, CD180* and *CR2* as well as interleukin genes *IL26* and *IL21R* and *NFKB1D* were expressed at high levels in the ileum and Peyer's patches cluster. The gut associated lymphoid tissue (GALT) is one of the largest lymphoid organs comprising nearly 70% of the body's lymphocytes. The aggregated lymphoid follicles which make up the Peyer's patches are found throughout the ileum and form an integral part of the immune system by surveilling populations of intestinal bacterial and deterring pathogenic bacteria (Jung, Hugot et al. 2010). Other notable immune genes highly expressed within the ileum and Peyer's patches cluster included *SIGLEC10* an immunoglobulin mainly expressed on the cell surface and C-type lectin *CLEC17A*. However, unlike in the sheep and pig gene expression datasets (Freeman, Ivens et al. 2012, Clark, Bush et al. 2017), core components of cell-cycle processes such as cyclins, DNA polymerases and kinesins were not

detectable within the ileum and Peyer's patches cluster of the goat, and were instead identified within the wider atlas as cluster 2 and are detailed below.

Together with the small intestine, the large colon is a major site for nutrient absorption in ruminants. Additionally, the colon plays a role in water and vitamin absorption from the digested material passing through, before excreting the remaining material as faeces from the rectum. A simple columnar epithelium lines the mucosa of the colon with a thin brush border and numerous goblet cells (Pelaseyed, Bergstrom et al. 2014). It was therefore unsurprising that gene members of the cadherin superfamily such as *CDH17*, *CDHR2* and *CDHR5* that function in mucus production (Suli et al. 2019) were highly expressed within the colon cluster. Additionally, the presence of myosin genes *MYO15B*, *MYO1A*, *MYO7B* within the colon cluster could be related to the presence of the thin brush border in the colon (Yu, Planelles-Herrero et al. 2017). Epithelial mucins such as *MUC13* and *MUC2* genes detected within the colon cluster belong to a family of secreted and cell-surface glycoproteins expressed by ductal and glandular epithelial tissue characteristic of that found in the colon. Similarly, as observed in the pig gene atlas(Freeman, Ivens et al. 2012), *KRT8* and *KRT19*, marker genes for columnar epithelium were enriched within the colon cluster. Associated with the role of the colon in nutrient and water absorption numerous solute carrier proteins such as *SLC13A2, SLC6A14, SLC12A2, SLC22A16, SLC26A3* and *SLC35A1* were detectable within the colon cluster.

Lastly, the rumen is the primary site for microbial fermentation of plant material in ruminants. The rumen wall is a strong stratified epithelium related to skin, with multiple layers and surrounded by a muscle layer to enable it contract to move and mix the rumen contents (Xiang, McNally et al. 2016). This was reflected in Fig 2.6 which shows transcriptional similarity in rumen and skin as well as the GO terms associated with the colon cluster such as keratinocyte differentiation ($p= 9.5\times10^{-5}$) and establishment of skin barrier ($p= 3.3\times10^{-5}$) as well as enrichment with numerous keratin family genes such as *KRT4*, *KRT6A*, *KRT78* and *KRTDAP*. The lamina propria layer of the rumen hosts active

immune cells as reflected by the presence of several immune genes, especially those involved in innate immunity of the epithelium, including intestinal bacterial pathogens such as *IL17C* and *IL17RE*, *IL20RA* a subunit of the receptor for *IL20* which is involved in epidermal function and *IL36B* which is part of the *IL36* signalling system that is present in epithelial barriers such as in the rumen where it takes part in local inflammatory response. In relation to regulatory mechanisms in the different processes and layers of the rumen wall, several transcription factors showed significant expression within the rumen cluster. These included key epithelial transcription factors such as *GRHL1, GRHL3 and OVOL1* involved in epithelial differentiation and proliferation and transcription factors involved in the Lipid/oxo-acid metabolism process such as *TP63* and *ZNF750.* These observations correspond to other studies describing gene network analysis of the sheep rumen (Xiang, McNally et al. 2016). As observed in the sheep atlas dataset (Clark, Bush et al. 2017), the goat rumen cluster did not contain a lot of solute carriers, and only *SLC9A2 (*also known as *NHE-2)* showed high expression within the rumen cluster. This is a key Na-H antiporter and has been shown to correlate to rumen-sodium transport.

2.4.4 Cell-cycle processes in the domestic goat

The majority of genes within the cell-cycle cluster 2 were involved in mitosis. Mitotic cell division takes place in all eukaryotic organisms. It is achieved through a highly organised process in four phases, namely G1 (Gap phase), S (DNA replication), G2 and M (mitosis). The molecular mechanisms controlling the cell cycle and transcriptional changes associated with this process have been extensively studied in a wide range of species revealing high conservation (Giotti, Chen et al. 2018). Genes encoding for cell cycle regulators *CDK1, CCNA2, CDC25A, CDC25B* and *CDC25C,* known to play a role in the G2/M boundary were detected in within this cluster as were genes encoding for kinetochore proteins *CENPC, CENPF, CENPH* and *CENPI.* Additionally, numerous motor proteins such as *KIF11, KIF14, KIF15, KIF18A, KIF18B, KIF201B, KIF200, KIF23, KIF2C* and *KIF4A* were expressed with the cell-cycle cluster. Similarly, mitotic cyclins including *CCNB2, CCNB3, CCND3*,

CCNJL and *BUB1* that are involved in the metaphase checkpoint were detected within this cluster. Several high mobility group (HMG) genes, a family of nucleosome binding proteins that bind to chromatin and modulate the structure and activity of chromatin, during the cell cycle such as *HMGN2*, *HMGN5, HMGB1* and *HMGB2* were detected within this cluster. The survival of cells across the body depends on proper progression through cell cycle, ensuring the correct transmission of genetic information to the newly formed cells. As expected, genes within this cluster were detectable across all tissues and cells reflecting that cell-cycle processes in all cells, but the enrichment in thymus, spleen and ileum may indicate that in the neonatal goat, these organs contain the highest proportion of mitotically active cells.

2.4.5 Innate and adaptive immunity

As is the case in the sheep and pig gene atlases, multiple clusters in the goat expression atlas contained genes involved in innate immune responses. For example, cluster 15 contained numerous genes involved in phagocytosis such as *LY96* which associates with toll-like receptor 4 (*TLR4*) and the macrophage-specific surface marker *CXCR4,* that mediates the innate immune response against bacterial lipopolysaccharide (LPS) by inducing expression of the archetypal pro-inflammatory gene *TNF.* A detailed analysis of innate and adaptive immune responses in the domestic goat has been discussed in chapter 3 of this thesis while comparative analysis of goat and sheep is detailed in chapter 4.

2.5 Conclusion

This chapter describes the first detailed analysis of the transcriptional landscape of the domestic goat. As previously demonstrated in sheep (Clark, Bush et al. 2017), cattle (Harhay, Smith et al. 2010) and pig (Freeman, Ivens et al. 2012), gene expression atlases of livestock species are useful resources for functional annotation. Transcriptional profiles for goat varied across tissue and cell types, and we describe the expression pattern of genes involved in metabolism, rumination and immunity. The goat atlas dataset represents a model transcriptome for the domestic goat. The data generated by this project was used to improve the annotation of the goat assembly and incorporated in the most recent Ensembl genebuild for goat (https://www.ensembl.org/Capra_hircus/Info/Index), providing a resource for the livestock genomics community. We have generated a significant new dataset that provides the foundation for further work to investigate candidate genes for numerous complex traits in goats, in the same way the sheep gene atlas was utilized to identify candidate genes for resistance to mastitis (Banos, Bramis et al. 2017).

Chapter 3 Goat bone-marrow derived macrophages and their response to bacterial lipopolysaccharide

3.1 Introduction

This chapter focusses on characterising goat bone-marrow derived macrophages and their response to challenge with bacterial lipopolysaccharide (LPS). As detailed in Chapter 1 of this thesis, the mononuclear phagocyte system (MPS) constitutes a major component of the innate immune system and acts as a first line of defence against invading pathogens (Hume, Ross et al. 2002, Hume 2006, Hume 2008). The maturation and differentiation of cells of the MPS including bone marrow progenitors, monocytes and tissue macrophages is controlled by hemopoietic growth factors particularly colony stimulating factor-1 (CSF-1) and interleukin-34 (IL-34*).*

Lipopolysaccharide (LPS) is a major component of the cell wall of gram-negative bacteria, and a potent activator of macrophages. Its recognition is facilitated through toll-like receptor 4 (TLR4) and CD14 found on the surface of mammalian cells that subsequently trigger the ensuing immune signalling pathways (Dobrovolskaia and Vogel 2002). LPS has been widely used *in vitro* to simulate infection and inflammation and enable investigation of innate immune responses. Studying the cascade of gene expression profiles in macrophages following stimulation with LPS has been the basis for understanding LPS-mediated responses in different species. For example, a deep analysis of the human MDM response to LPS revealed a sequential cascade of transient transcriptional activation over 48 hours (Baillie, Arner et al. 2017). This study and others highlighted the importance of inducible feedback regulation in limiting the duration and magnitude of the response to LPS. Other studies have compared the LPS response in human and mouse macrophages (Schroder, Irvine et al. 2012) and characterised transcriptional

networks in underpinning activation of mouse macrophages (Raza, Barnett et al. 2014). In livestock several studies have characterised the transcriptional response of macrophages to LPS in different species such as pig (Kapetanovic, Fairbairn et al. 2012), horse (Karagianni, Kapetanovic et al. 2017), sheep (Clark, Bush et al. 2017), water-buffalo and cattle (Young, Bush et al. 2018). These studies in livestock have involved the isolation and culture of primary cells, such as bone marrow derived macrophages (BMDM), exposing them to LPS and subsequently characterising the immune responses triggered by analysing gene expression profiles.

However, there is little knowledge available on goat bone marrow derived macrophages, and the study of innate immunity in the domestic goat has progressed relatively slowly compared to other livestock species. A previous study compared the regulation of inducible nitric oxide synthase (iNOS) in cattle and goat by isolating cattle and goat MDM before stimulating them with LPS (Adler et al. 1996). Another study isolated goat peripheral blood mononuclear cells (PBMC) to study the anti-viral immune response in the joint after infection with caprine arthritis encephalitis virus (Lechner and Peterhans 1999). All previous studies have however used RT-qPCR to measure the expression of a small subset of genes. Using RNA-Seq, as for the livestock species mentioned above, we could provide a transcriptome wide profile of the gene expression in response to LPS in goat BMDM. As such the aim of the work presented in this chapter was to optimise culture methods used for other livestock species (Clark, Bush et al. 2017, Young, Bush et al. 2018) to establish a protocol for the isolation, culture and differentiation of goat BMDM and characterise their responses to LPS using RNA-Seq. The dataset we generated would enhance our understanding of the innate immune response in the domestic goat and provide a comparable dataset for cross-species immune transcriptomic analysis.

3.2 Materials and Methods

The focus of this chapter is the characterisation of bone marrow derived macrophages (BMDM), from three male crossbred dairy goats, approximately six days old and their response to challenge with lipopolysaccharide (LPS) (*Salmonella enterica* serotype Minnesota Re 595). It follows similar protocols as described for the domestic pig (Kapetanovic, Fairbairn et al. 2012), sheep (Clark, Bush et al. 2017) and water-buffalo (Young, Bush et al. 2018) over a 7-hour time-course treatment. This dataset comprises a significant component of the macrophage specific and LPS-inducible clusters within the wider goat gene expression atlas discussed in Chapter 2.

3.2.1 Animals

This study was reviewed and approved by The Roslin Institute, University of Edinburgh's Animal Work and Ethics Review Board (AWERB). All animal work was carried out under the regulations of the Animals (Scientific Procedures) Act 1986. Three healthy male crossbred goats, approximately five to six days old were used for this study (a subset of animals described in Chapter 2).

3.2.2 Cell isolation and cryopreservation

The rib cage from each animal was removed, within two hours post euthanasia, and transported to the lab in zip-lock bags containing 2L of sterile PBS while kept on ice. Upon arrival, the ice was replenished, and ribs placed in the cold-room overnight at 4^0C. All cell processing was performed under a tissue culture hood. Ten posterior ribs (five from each side) were stripped clean by scraping of the attached muscle and fat with PM40 knives, washed in 70% ethanol and cut into sections approximately 1 inch long. Bone marrow was flushed from the ends with a 10G needle using RPMI 1640 (Sigma UK) containing 5mM EDTA to prevent clotting and collected in a 50mL tube. The cell suspension was filtered through a 100-µm cell strainer to remove bone debris and centrifuged at 400xg for 5 mins. The supernatant was discarded, and cell pellet resuspended in 5mL of red blood cell (RBC) lysis buffer (10mM $KHCO_3$, 150mM NH_4Cl, 0.1mM EDTA pH 8) and incubated in the dark for 5 mins. Sterile PBS was added to the RBC lysed cells to dilute the red cell lysis buffer and

cells centrifuged again at 400xg for 5 mins and supernatant was discarded. The cell pellet resuspended in 50mL goat media made of 400mL RPMI 1640 (Sigma UK), 100mL goat serum (Sigma UK: G6767), 5mL GlutamaX (Invitrogen UK) and 1.25mL Penicillin/Streptomycin (25U/25µg/mL, Gibco) and cells counted with a haemocytometer.

Bone marrow cells were prepared for cryopreservation as described for pig (Kapetanovic, Fairbairn et al. 2012) and horse (Karagianni, Kapetanovic et al. 2017) by suspending in freezing media (90%FBS, 10%DMSO), slowly to avoid shocking the cells with the DMSO and aliquoted at a density of $5x10^7$cells/mL per cryovial. The cryovials were stored in an isopropanol freezing unit (Mr Frosty) at -80^{o}C for 24hrs to allow a slow reduction in temperature. The next day, cells were transferred to a -155^{o}C freezer for long-term storage. This procedure of isolating cells from all individuals followed by cryopreservation enables downstream in vitro study of immune responses from multiple animals at the same time and in comparable culture conditions.

3.2.3 Cell culture

The steps taken to establish and optimise the protocol for culture and differentiation of goat bone marrow derived macrophages (BMDM) are detailed in section 3.3.1. Briefly, bone marrow cells were recovered from cryopreservation as described for pig (Kapetanovic, Fairbairn et al. 2012) and horse (Karagianni, Kapetanovic et al. 2017). Each vial was rapidly thawed in a water bath (37^0C) for 2mins and the contents transferred to a 50mL falcon tube. 9mL of pre-warmed goat media (RPMI 1640 supplemented with 20% heat-inactivated goat serum, 5mL GlutamaX and 1.25mL Penicillin/Streptomycin) was added to the thawed cells, slowly in a drop-wise manner, to avoid shocking the cells. The tube was gently inverted a few times to mix the cells fully and then centrifuged at 400xg for 5 mins. Supernatant was discarded, and cells counted using a haemocytometer. After assessing the viability of the cells with Trypan Blue 0.4% (ThermoFisher), cells were plated on T75 polystyrene tissue culture treated plates (Nunc, ThermoFisher) at a density of $2.0x10^6$cells/mL. Cells were differentiated in the presence of

recombinant human colony stimulating factor 1 (rhCSF1, a gift from Chiron, Emeryville, CA, USA) at a final concentration of 10^4 U/ml (100ng/mL) and plates incubated at 37^0C in 5% CO^2 for 10 days.

3.2.4 LPS stimulation

On day 11, adherent cells were gently detached from the plate using a cell scraper and washing with 10mL of warm media, taking about 4-5 rounds to recover all cells. Cells were centrifuged at 400xg for 5 mins and the supernatant discarded. Cells were counted and seeded in duplicate in 6-well tissue culture plastic plates (Nunc, ThermoFisher) ensuring each well had 2×10^6cells in a total of 2mL goat media + 20uL rhCSF1. Plates were incubated overnight at 37^0C in 5% CO^2 to allow cells to adhere onto the tissue culture plastic. The following day, cells were stimulated with LPS from *Salmonella enterica* serotype Minnesota Re 595) at a final concentration of 100ng/mL for 7 hours. The 7-hour time point was chosen as it represented the peak activation of marker immune genes including TNF (Freeman, Ivens et al. 2012), which was the only available dataset for a livestock species when these experiments where undertaken. Supernatants from each well were collected in cryovials and stored at -20 ^{o}C to enable measurement of nitrite using the Griess assay (see section 3.2.6). The adherent cells were recovered by washing with 1mL Trizol Reagent (Invitrogen, Darmstadt, Germany) then collected in pre-labelled cryovials and stored at -80 ^{o}C until RNA isolation.

3.2.5 Cell imaging and Flow Cytometry

3.2.5.1 Microscopy

To assess the morphology of the goat BMDM, cells were grown for 10 days until fully differentiated then seeded at $5.0x10^4$ cells on cover slips in 12-well plates and incubated overnight. The next day, cells were washed once with sterile PBS, and then fixed with freshly prepared 4% paraformaldehyde (PFA) for 30min at room temperature. After fixing, cells were permeabilised by adding PBST (0.1% Triton x100) for 1min at room temperature, any unspecific staining was minimised by blocking cells in 3% bovine serum albumin (BSA) for 1 hour. The blocking solution was replaced with a solution of Texas Red-X phalloidin

(ThermoFisher) at a concentration of 1:500 to mark cell boundaries and DAPI (4', 6-diamidino-2-phenylindole) at a concentration of 1:1000 was added to mark the cell nucleus. After incubating the plate in the dark for one hour, cells were washed three times in sterile PBS and cover slips mounted onto glass slides using ProLong Gold mountant (ThermoFisher). Unstained cells were imaged using standard light microscopy and ZEN software (Zeiss, Cambridge, U.K) while the stained cells were imaged using a Leica DMLB-2 upright fluorescent microscope and analysed using the ImageJ software (NIH, Maryland, USA).

3.2.5.2 Phagocytosis assay

The phagocytic activity of goat BMDM was assessed using pHrodo Red *E.coli* BioParticles (ThermoFisher), based on a modified version of the method described in (Heike Bicker, Conny Hoflich et al. 2008) and (Pridans, Davis et al. 2016) . Cells were cultured until they were fully differentiated as described in section 3.2.3. For each sample, 10^5cells in a total volume of 100uL were added to a 1.5mL microcentrifuge tube. An equal volume of pHrodo particles (stock concentration, 1mg/ml, and working dilution 1:20) was added to the cells and the tubes wrapped in foil to keep in the dark then incubated at 37^{o}C for one hour, with respective control tubes kept on ice at 4^{o}C as phagocytosis is inhibited at this temperature. After one hour, the assay tubes were placed on ice for 10min to stop phagocytosis and cells prepared for double staining with different antibodies.

3.2.5.3 CD14, CD16 and CD172a cell surface expression

The pHrodo stained cells were pelleted by centrifuging at 1400 rpm for 4 mins, and the supernatant was discarded. To reduce background and unspecific staining, cells were incubated in 100uL of blocking buffer made up of 2% horse serum in PBS and kept on ice for 15min. Cells were maintained in the dark by wrapping the tubes in foil. The antibodies used in this assay were adopted from a previous study on sheep monocytes (Pridans, Davis et al. 2016) with the assumption that they would cross-react with goat macrophages. Cells were pelleted by centrifuging and resuspended in blocking buffer containing either

of the following primary antibodies: anti-human CD14:AF647 (clone TUK4, 1:20) mouse anti-human CD16: FITC (clone KD1, 1:200), and mouse anti-bovine CD172a (clone CC149, 1:200) all from AbD Serotec. Isotype controls were used at the same concentration as the primary antibodies and included: mouse IgG2a, IgG2b (AbD Serotec) and rat IgG2a (BioLegend. IgG2b-RPE (1:400) (BioLegend) was used as secondary antibody. Sytox Blue (Life Technologies) was used to exclude dead cells and cells analysed on a BD LSR Fortessa X20 (Oxford, UK). Data was acquired after approximately 5000 events.

3.2.6 Griess assay

In rodents, but not in humans or pigs, LPS stimulation of macrophages leads to metabolism of arginine by inducible nitric oxide synthetase (NOS2) and the production of Nitric Oxide (NO)(Kapetanovic, Fairbairn et al. 2012, Kate Schroder, Katharine M. Irvine et al. 2012). Previous studies indicated that goat macrophages do not produce large amounts of NO (Adler et al. 1996, Jungi, Adler et al. 1996). Nitrite, the stable metabolite of NO was measured against sodium nitrite standards using the Griess reagent: 1%sulfanilamide, 0.1%N-(1-naphtyl) ethylenediamine diHCl, 2.5%phosphoric acid. An equal volume of cell culture supernatants collected after LPS stimulation as mentioned in section 3.2.3 was added to 100uL of freshly prepared Griess reagent and incubated at 37°C for 30 min. Absorbance was measured using on a Multiskan v2.6 (ThermoScientific, USA) at 570 nm. Chicken cell culture supernatants from BMDM stimulated with LPS in a similar manner were used as positive control as they have been shown to produce NO in response to LPS (Wu, Hu et al. 2016).

3.2.7 Total RNA isolation

Extraction of good quality RNA is an important pre-requisite to produce high quality expression data. Prior to performing any RNA work, all surfaces and equipment were thoroughly cleaned with RNase-Zap (ThermoFisher) to remove all traces of nucleic acid contaminants. Only RNase free reagents and consumables were used, having been set aside specifically for RNA work. Cell

lysate samples previously stored at -80 °C were thawed on ice and brought to room temperature for 5 min. 200µl BCP (bromochloropropane) (Sigma Aldrich) was added, shaking each tube vigorously for 15sec, then incubating for 3 mins at room temperature. Tubes were centrifuged for 15 minutes at 12,000 x g, at 4°C and the homogenate separated into a clear upper aqueous layer (containing RNA), and interphase and red lower organic layers (containing the DNA and proteins). After centrifugation, the upper aqueous phase was removed and RNeasy Mini Kits (Qiagen UK) were used to extract total RNA from each sample. The protocol to isolate RNA was followed according to the manufacturer's instructions and included an on-column DNase treatment step using the RNase-Free DNase treatment kit (Qiagen UK) to digest any contaminating DNA. The quantity of RNA isolated from each sample was estimated using a NanoDrop spectrophotometer (NanoDrop Products, Wilmington USA) and confirmed using the Qubit fluorometric quantitation (ThermoFisher Scientific UK) kit. RNA quality was analysed by measuring the RNA integrity on the Agilent 2200 TapeStation (Agilent Technologies, Santa Clara CA USA). All samples had RNA integrity number equivalent (RIN) of 9.0 (Appendix M). (RIe≥7.0 is recommended for RNA sequencing).

3.2.8 cDNA synthesis

One microgram of total RNA was used as template for the reverse transcription to complementary DNA (cDNA). The SuperScriptIII First-Strand Synthesis System for RT-PCR (ThermoFisher) was used according to manufacturer's instructions. A 13uL reaction mix was made containing 1ug of total RNA, 1uL of Oligo$(dT)_{20}$ primer, 1uL of dNTPs and RNase/DNase free water. This solution was heated at 65 °C for 5min and then incubated on ice for 1min. The cDNA synthesis mix was prepared by adding 2uL of 10X RT buffer, 4uL of 25mM MgCl2, 0.1M DTT, 1uL RNaseOUT(40U/uL) and 1uL of SuperScriptIII RT enzyme(200U/uL). 10uL of the synthesis mix was added to each RNA/primer mixture, mixed gently and collected by centrifugation. Negative controls to check for genomic DNA contamination, were included by omitting the SuperScriptIII enzyme. The samples were incubated at 50 °C for 50min, and the reaction terminated by incubating at 85 °C for 5min, then chilled on ice

at 4 ºC. The cDNA was stored at -20 ºC until further use in the reverse transcription polymerase chain reaction (RT-qPCR).

3.2.9 RT-qPCR

RT-qPCR analysis against representative pro-inflammatory genes and cytokines was performed to validate the LPS response. All RT-qPCR reactions were carried out using the SYBR Green quantitative PCR system (ThermoFisher) in 20 µL reactions on MicroAmp Optical 96-well Reaction plates (ThermoFisher). Each reaction contained 10uL of 2X *Power* SYBR Green master mix, 10uM forward and reverse primers, 3.8uL of water and 5uL of diluted cDNA. Primers used in this study were selected based on already published and publicly available primer sequences (Zhang, Zhang et al. 2013, Brenaut, Lefevre et al. 2014, Yadav, Dangi et al. 2016, Chopra-Dewasthaly, Korb et al. 2017), and are listed in Table 3-1. The RT-qPCR assay was designed to adhere to the MIQE guidelines (Bustin, Benes et al. 2009) by testing each sample in triplicate and including two negative controls; one with no reverse transcriptase (NRT) and another with no cDNA (or no template control) (NTC). Samples were run on the ABI 7500 RT-qPCR machine (ThermoFisher) using the cycling conditions shown in Table 3-2. Relative quantitation of transcript abundance against housekeeping genes *HPRT* or *GAPDH* was determined using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

Table 3-1:Primers used in the RT-qPCR assay.

Primer name	Sequence (5'-3')	Tm(°C)	Source
cap_TNFa_fwd	CAGAGGGAAGAGCAGTCCCC	60.86	(Brenaut, Lefevre et al. 2014)
cap_TNFa_rev	TGGGCTACCGGCTTGTTATTT	61.53	
cap_IL8_fwd	TGAGAGTGGGCCACACTGC	59.53	(Brenaut, Lefevre et al. 2014)
cap_IL8_rev	CACAACCTTCTGCACCCACTT	59.88	
cap_IL1B_fwd	GACAACAAGATTCCTGTGGCC	59.29	(Brenaut, Lefevre et al. 2014)
cap_IL1B_rev	TCTACTTCCTCCAGATGAAGTGT	55.1	
cap_CCL4_fwd	CAGCCGTGGTATTCCAGAC	55.81	(Brenaut, Lefevre et al. 2014)
cap_CCL4_rev	CTCGGAGCAGCTCAGTTCAGT	59.54	
cap_GAPDH_fwd	GGGTGATGCTGGTGCTGAGT	61.1	(Zhang, Zhang et al. 2013)
cap_GAPDH_rev	TCCCTCCACGATGCCAAA	61.3	
cap_HPRT1_fwd	ATTATGGACAGGACCGAACGG	61.26	(Zhang, Zhang et al. 2013)
cap_HPRT1_rev	CCAACAGGTCGGCAAAGAACT	61.64	
cap_IDO1_fwd	GATACATCACCATGGCGTATGTGT	61.34	(Chopra-Dewasthaly, Korb et al. 2017)
cap_IDO1_rev	AATCCGCATAAAGCAGAATAGGA	60.4	
cap_NOS2_fwd	CCAGCCCAAGGTCTATGTTC	56.98	(Yadav, Dangi et al. 2016)
cap_NOS2_rev	TAGTCCTCCACCTGCTCCTC	56.4	

Table 3-2: RT-qPCR cycling conditions.

Step	Temperature(°C)	Time	Number of cycles
Holding Stage	50	20sec	1
	95	10min	
Cycling Stage	95	15sec	40
	60	1min	
Melt Curve Stage	95	15sec	1
	60	1min	
	95	30sec	
	60	15sec	

3.2.10 RNA sequencing and Data processing

Library preparation and RNA sequencing was performed by Edinburgh Genomics (Edinburgh Genomics, Edinburgh, UK). Libraries were run on the Illumina HiSeq 4000 platform (Illumina, San Diego, USA), sequenced at a depth of 50 million, 75bp paired-end reads per sample. All libraries were prepared using the Illumina TruSeq mRNA library preparation protocol (poly-A selected)(Illumina 2017). Briefly, oligo(dT) coated magnetic beads were used to purify polyA containing mRNA molecules which were then fragmented and primed for cDNA synthesis. The first strand cDNA was synthesised by reverse transcribing cleaved RNA fragments with random hexamers. Actinomycin D was added to the synthesis mix to prevent pseudo DNA-dependent synthesis, only allowing RNA-dependent synthesis which improved strand specificity (Ruprecht, Goodman et al. 1973). The second strand cDNA was synthesised using DNA polymerase 1 and RNase H by replacing dTTP with dUTP to generate a double stranded cDNA molecule (ds cDNA). One adenine (A) nucleotide was added to the 3' ends of the ds cDNA molecules and a corresponding thymine (T) nucleotide added to 3' end of the adapter to provide a complementary overhang to ligate the adapter to the fragment. The adapter-

fragment chimera was hybridized onto a flow cell and using PCR, only those fragments having adapter molecules on both ends were amplified. RNA-Seq data was processed using methods and pipelines developed for the sheep gene expression atlas (Clark, Bush et al. 2017), detailed in Chapter 2 (section 2.2.7) and summarised in Appendix E.

3.2.11 Network Cluster Analysis

Miru (Kajeka Ltd, Edinburgh UK), a network clustering program widely used for the analysis of large transcriptomic datasets including sheep and pig (Theocharidis, van Dongen et al. 2009, Freeman, Ivens et al. 2012, Clark, Bush et al. 2017) was used to analyse the goat BMDM data. Details about the Miru algorithm are included in Chapter 2 (section 2.2.8). The BMDM dataset (Appendix N) was used for the network cluster analysis. Using a Pearson correlation co-efficient cut-off of r=0.99, MCL (Markov Cluster Algorithm) inflation value of 2.2 (Dongen and Abreu-Goodger 2012), pre-inflation of 6.0, and minimum 20 nodes per cluster, a gene-to-gene network graph was generated. A high correlation co-efficient was necessary due to the small number of samples being analysed (two conditions (+/- LPS) and four biological replicates). The resultant graph was made up of nodes representing genes, connected by edges representing the correlations above the set threshold of r=0.99. Lowly expressed transcripts (≤ 1 TPM) were excluded.

3.3 Results

3.3.1 Establishing a protocol for culture and differentiation of goat BMDM

3.3.1.1 Goat cells need autologous serum

Serum is an integral component of the culture media, providing a source of basic nutrients, vitamins, growth and adhesion factors and minerals. Heterologous fetal bovine serum (FBS) is used routinely in culture media compositions to generate macrophages from multiple species including human (Baillie, Arner et al. 2017), mice (Marim, Silveira et al. 2010) and pig (Kapetanovic, Fairbairn et al. 2012). However, in parallel studies in the group working on sheep (M.McCulloch, personal comm) and horse (Z.Lisowski, A. Karagianni; personal comm) autologous serum was optimal. Two media compositions were therefore assessed for the culture of goat BMDM, RPMI 1640 supplemented with either 20% goat serum or 20% FBS.

As shown in Fig 3.1(A and B), freshly isolated cells cultured in media containing 20% goat serum attained the expected stellate morphology of macrophages, differentiated and adhered to the bottom of the plate after one week in culture. BM cells grown in 20% FBS, Fig 3.1 (C and D) appeared viable but failed to develop an adherent population. Accordingly, goat serum was used in all subsequent experiments.

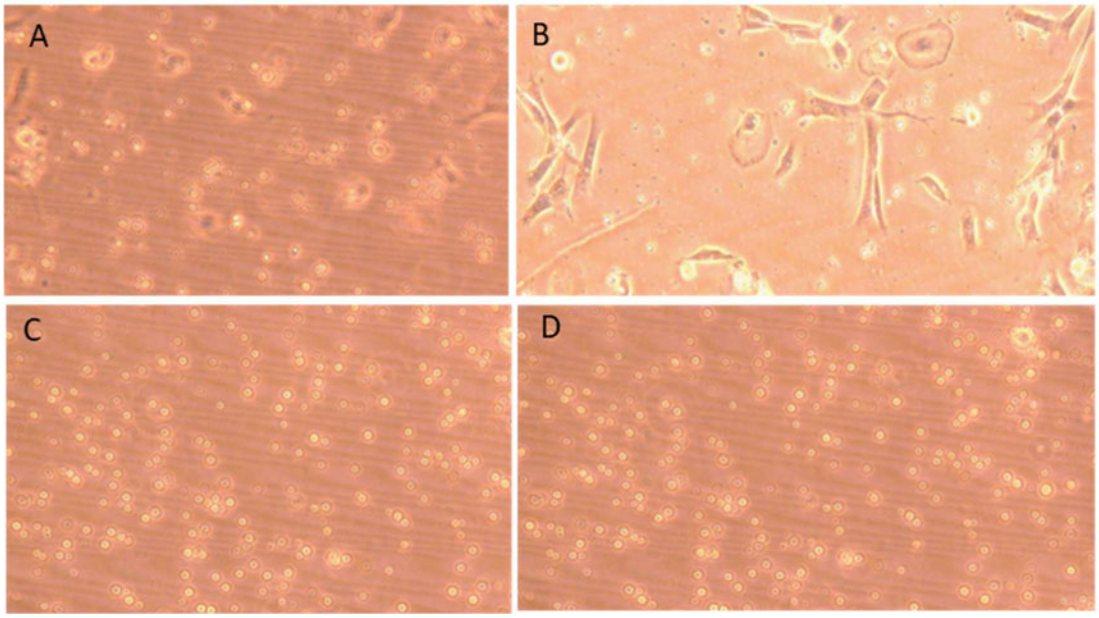


Figure 3-1:Goat 1 bone marrow cells differentiating into BMDM over the week. Top panel: cells growing in media containing 20% goat serum on day 6 (A) and day 8(B). Bottom panel: cells growing in media containing 20% FBS on day 6(C) and day 8(D). Photographs are taken at x20 magnification and are representative in appearance of all bone marrow cells for the other three individuals

3.3.1.2 Goat cells are recoverable from cryopreserved stock and differentiate when plated on tissue culture treated plastic.

Bone marrow isolation yielded a total of 3.52×10^9 cells across all individuals (1.37×10^9 cells, from goat 1, 2.3×10^8 cells from goat 2, 8.2×10^8 cells from goat 3 and 1.1×10^9 cells from goat 4). According to routine practice, cells were cryopreserved using the method described in section 3.2.2. It was therefore important to test the viability of goat bone marrow cells following cryopreservation. In the first instance, recovered cells were plated on $100mm^2$ bacteriological plastic as used for other species pig (Kapetanovic, Fairbairn et al. 2012), mouse (Marim, Silveira et al. 2010), and sheep (Pridans, Davis et al. 2016) and grown in complete goat media in the presence of rhCSF1. To test the optimal cell density required, cells were plated at different cell concentrations. However, as illustrated in Fig 3.2, the thawed cells failed to differentiate, and further optimisation of the protocol was required.

Figure 3-2:Goat 1 bone marrow cells differentiating into BMDM over one week. From right to left, cells cultured at a final density of 1×10^6 cells/mL (A), 2.0×10^6 cells/mL (B) and 2.5×10^6 cells/mL (C). Photographs are all taken at x20 magnification and are representative in appearance of all individuals

Cells from two different individuals were cultured in goat media containing bovine granulocyte-macrophage colony stimulating factor (GM-CSF) (donated by Prof. Gary Entrican, Moredun Research Institute) or media from an ovine embryonic fibroblast (OEF) culture. GM-CSF promotes macrophage maturation, (Francisco-Cruz, Aguilar-Santelises et al. 2014), while OEFs synthesise endogenous CSF and activate macrophage maturation and were added to the goat culture media to facilitate macrophage differentiation. To conserve stocks by minimising the number of cells used for this assessment, cells were seeded in 24-well tissue culture plastic plates at 4×10^4 cells/well and plates incubated at 37^0C in 5% CO_2 for 10 days with monitoring every 2-3 days. Cells grown in control goat media without GM-CSF and OEF were included. As summarised in Fig 3.3, each of these different media compositions tested were successful as cells showed the expected macrophage morphology and adherence after 10 days in culture. Previously, the thawed cells had been cultured on bacteriological plastic. This observation indicated that the key variable was the substratum, rather than the media and seemingly, goat macrophage differentiation requires adhesion to tissue culture (TC) plastic.

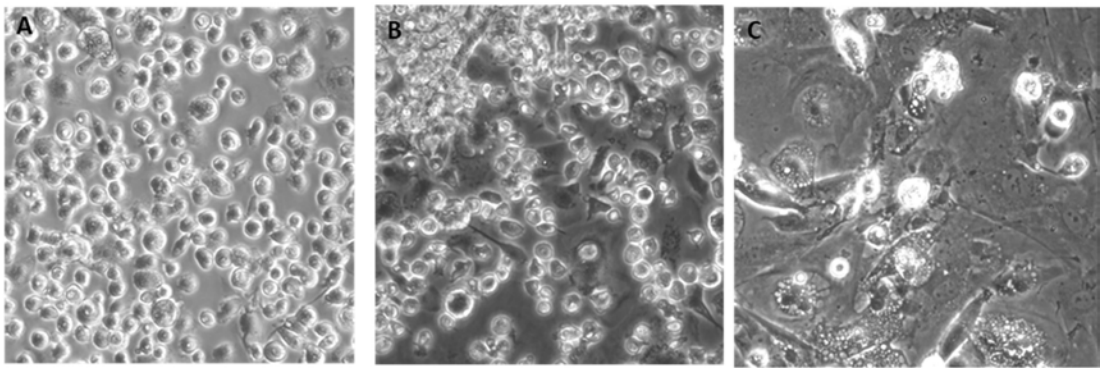


Figure 3-3: Goat 4 differentiated bone marrow derived macrophages after 10 days in culture maintained in different media compositions. Cells were maintained in control goat media (A), goat media containing OEF (B) or goat media containing GM-CSF (C). Cells were grown in the presence of rhCSF1 and photographs are all taken at x20 magnification

Having established that BMDM could differentiate in the conditions described above, it was then important to decide which of the conditions were ideal to use for this study. Since one of the main objectives for this study was to carry out a comparative analysis between goat and other species, it was imperative that culture conditions were as consistent as possible across the two species. As such, conditions as close as possible to those described for sheep (Pridans, Davis et al. 2016, Clark, Bush et al. 2017) and pig (Kapetanovic, Fairbairn et al. 2012) were used, the only difference being that the goats cells required plating out on TC plastic, rather than bacteriological plastic. Cells from the remaining individuals were recovered and cultured in goat media comprised of RPMI 1640 supplemented with 20% heat-inactivated goat serum, 5mL GlutamaX and 1.25mL Penicillin/Streptomycin. Cells were plated on T75 tissue culture treated plastic plates and grown in the presence of rhCSF1. Plates were incubated at 37^{o}C in 5% CO^{2} for 10 days with monitoring every 2-3 days. As demonstrated in Fig 3.4, all cells showed increased size, granularity and adherence to the tissue-cultured treated plastic, typical of observations from other species and studies where bone marrow progenitor cells have been cultured in similar conditions. This protocol was used for all future work.

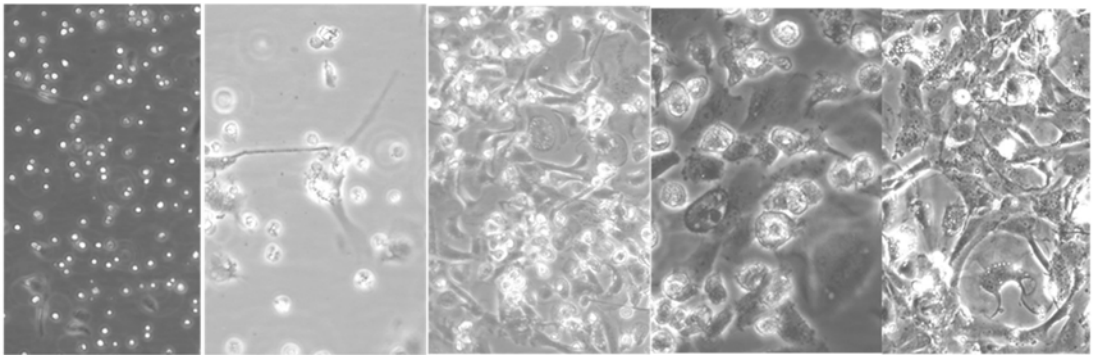


Figure 3-4: Goat bone marrow derived macrophages differentiation over 10 days. From left to right, Goat BMDMs at Day2, 3, 6, 7 and 10. All cells were grown on tissue-culture plastic in the presence of rhCSF1. Photographs are all taken at x20 magnification, and images are representative of all the three individuals used in this study

Generally, as shown in Table 3-3, numbers of monocytes from goat were lower compared to cattle and pig and more comparable to sheep. The progenitor cells were also very fragile and needed to be treated very gently whilst in culture.

Table 3-3:Comparison on monocyte counts across livestock species

Species	Monocyte count %
Pig	2-10
Cattle	0-8
Sheep	0-6
Goat	0-4

3.3.2 Goat BMDM are CD14, CD16 and CD172a positive

To characterise the goat bone marrow derived macrophages, expression of relevant cell surface markers was assessed using flow cytometry. Forward scatter (size) and granularity (side scatter) were used to distinguish the different population of cells and exclude granulocytes from the analysis. Fig 3.5 (C III), demonstrates that up to 95.4% of the cells were positive for CD14, the co-receptor for LPS. The level of expression of CD16 (15.9%) Fig 3.5 (D III) and CD172a (30.4%) Fig 3.5 (E II) was low similar to observations made for sheep monocytes (Pridans, Davis et al. 2016).

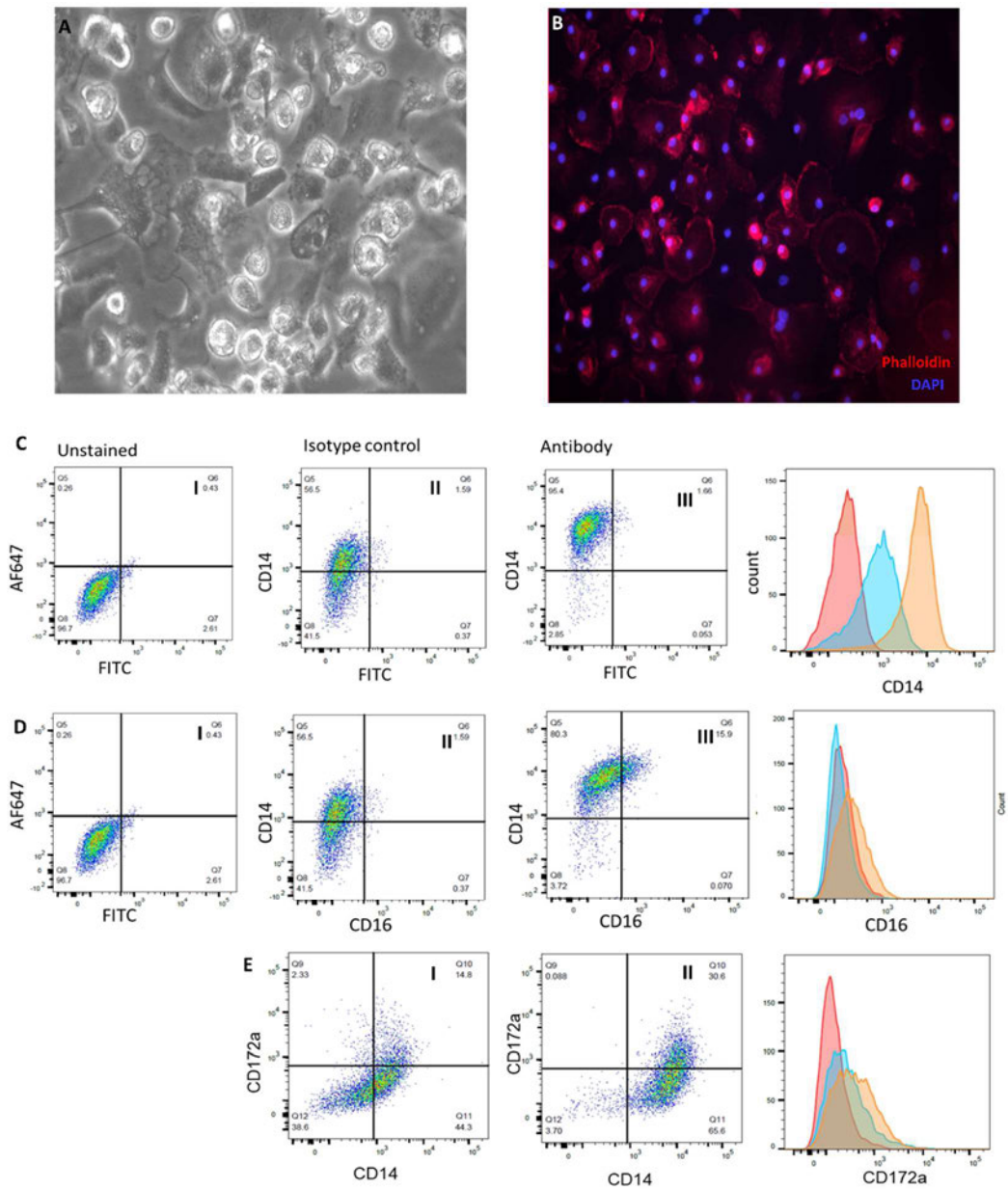


Figure 3-5: Functional characterisation of goat BMDM: Morphology of goat BMDM analysed by light microscopy (A) and fluorescent microscopy (B) shows expected stellate and adherent nature of macrophages. Expression of CD14, CD16 and CD172a was determined in freshly isolated goat BMDM. Threshold levels to determine CD14 and CD16 positivity were set with no antibody control (unstained) for CD14 (C I) and CD16(D I). Isotype controls (C II), (D II) and (E I) were used to measure background staining for CD14, CD16 and CD172a respectively. Antibody expression determined by double staining, demonstrated that

up to 95.4% of the cells were positive for CD14 (C III). The level of expression of CD16 at 15.9% (D III) and CD172a at 30.4% (E II) was low similar to observations made for sheep monocytes. Figures are representative of two goats

3.3.3 LPS induced *TNF*, *IL1 β*, *CCL4* and *IL8* mRNA expression in goat BMDM

Before submitting samples for RNA sequencing, the response of goat BMDM to LPS was validated using RT-qPCR. Messenger RNA (mRNA) expression for a selected set of known LPS-inducible genes including the cytokines *TNF*, *IL1β* and chemokines *IL8* and *CCL4* was measured after stimulating cells with 100ng/mL of LPS for 7hrs. Transcription levels were calculated as relative fold change against the housekeeping gene *HPRT1*. All genes tested showed induced expression in stimulated cells relative to unstimulated cells. The highest induction was observed for *IL1 β* followed by *CCL4* while *TNF* and *IL8* had comparable induction levels. The level of induction varied greatly between the individuals. RT-qPCR results (averaged across individuals for each gene) are presented in Fig 3.6.

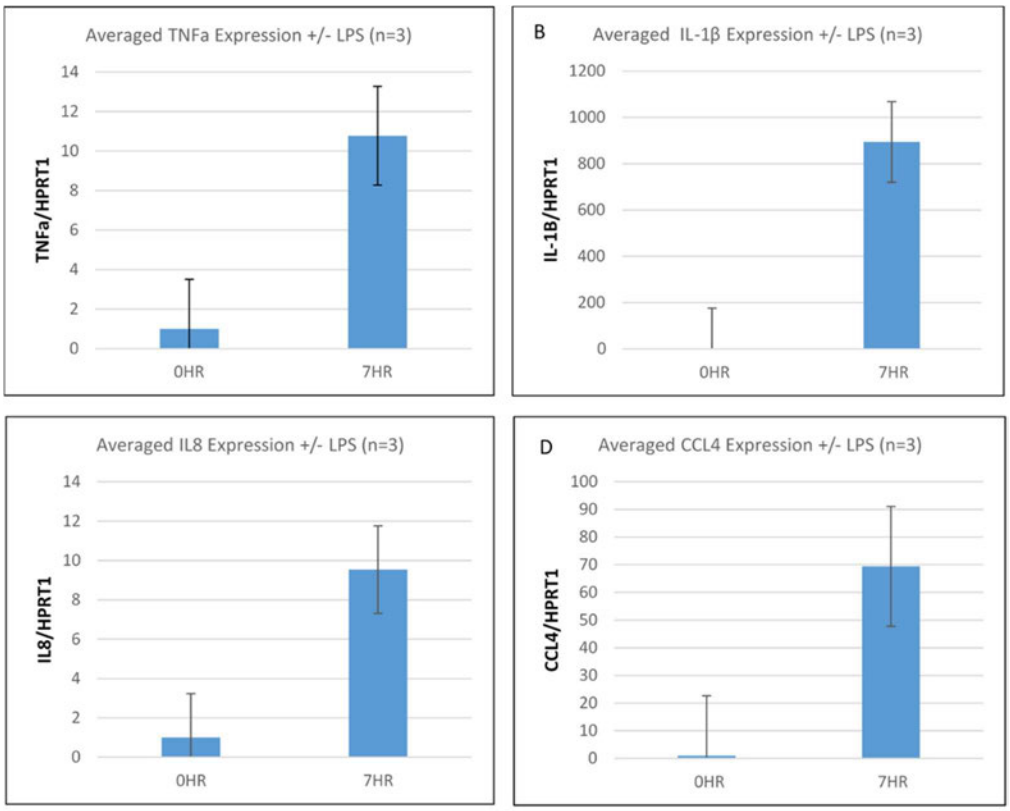
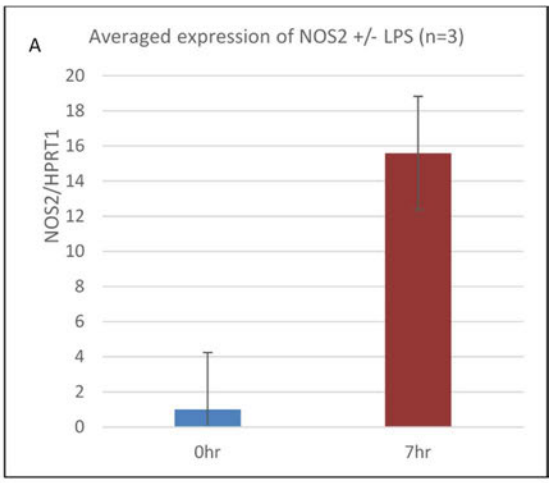
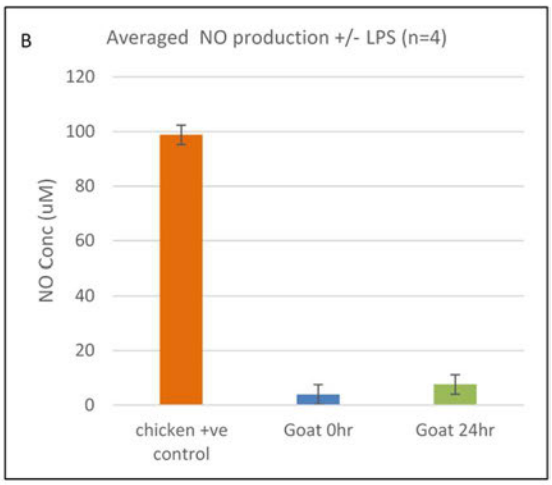


Figure 3-6: Expression of pro-inflammatory cytokines and chemokines in Goat BMDM after stimulation with 100ng/mL LPS for 7hrs. Averaged expression levels of TNFa (A), IL-1B (B), IL8 (C) and CCL4 (D) are presented as relative fold changes against the housekeeping gene HPRT1

3.3.4 Goat BMDM induce NOS2 and produce detectable levels of NO in response to LPS

Nitric Oxide (NO) is widely recognized as an immunological agent and has anti-microbial and anti-tumoricidal activity. Mouse macrophages produce NO in response to LPS while pig and human do not (Kapetanovic, Fairbairn et al. 2012). Transcript levels of inducible NO synthase (*NOS2*) in goat BMDM measured by RT-qPCR relative to the housekeeping gene *HPRT1* showed induced expression after 7hrs stimulation with LPS (Fig 3.7, A). To determine whether there was nitrite induction, nitrite levels were measured in the supernatants for each sample using the Griess assay after 24hr of LPS stimulation. As a positive control, supernatant obtained from chicken BMDM, which produce large amounts of NO in response to LPS (Wu, Hu et al. 2016), was used. By comparison, nitrite produced by goat BMDM was barely detectable (Fig 3.7 ,B). This finding is consistent with the earlier study by (Jungi, Adler et al. 1996). In that study, cattle macrophages produced low but detectable NO, which has been confirmed in our group (Young, Bush et al. 2018). Fig 3.7, C compares goat BMDM to cattle BMDM. In this comparison, there was low but detectable production of NO after 24hr LPS stimulation by goat, indicating goats produce less NO than cattle and chicken although the mRNA transcript is induced after 7 hours stimulation with LPS.





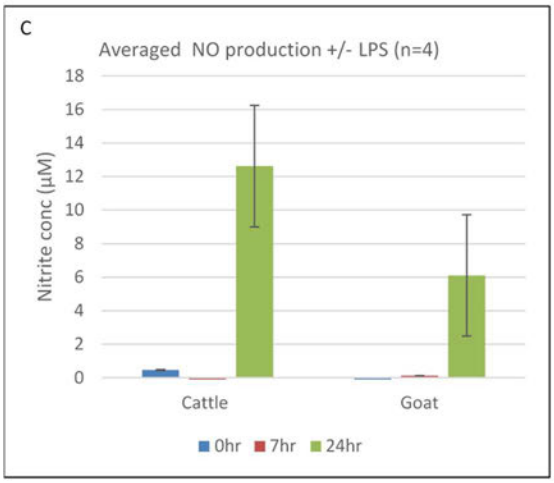


Figure 3-7: Nitric Oxide (NO) production by Goat BMDM post LPS stimulation. Expression of *NOS2* was assayed by RT-qPCR (A) while NO production was assayed by Griess assay (B) and (C).

3.3.5 RNA-Seq analysis of goat BMDM post-LPS stimulation

Further analysis on the gene expression profile of goat BMDM following stimulation with LPS for 7hrs, was performed using RNA sequencing as described in section 3.2.10 and 3.2.11. As expected from studies on human (Martinez, Gordon et al. 2006), mice (Raza, Barnett et al. 2014) and pig (Kapetanovic, Fairbairn et al. 2012), LPS stimulation of goat BMDM led to the induction of a large number of genes. The expression estimates for each gene from the Kallisto output (Appendix N), as transcripts per million (TPM), were analysed using network cluster analysis in Miru. Genes were clustered using a correlation of r=0.99 and MCL of 2.2 resulting in a gene-to-gene network graph which comprised of 14,747 nodes connected by 190,406 edges, representing correlations above the set threshold (r=0.99). As illustrated in Fig 3.8, the BMDM gene-to-gene network graph separated into three main clusters. The largest cluster contained genes, which had a basal expression pattern and remained largely unchanged post-LPS stimulation. Of the other two clusters one contained genes whose expression was down-regulated post-LPS stimulation and the other contained genes that were up-regulated by LPS stimulation. The genes in each cluster were inspected manually to classify them further. This was then confirmed with an automated method using the gene ontology database PANTHER (Mi, Dong et al. 2010).

The cluster of basally expressed genes contained 2,587 expressed genes (Appendix O). The gene enrichment analysis with PANTHER (Appendix P) revealed these genes were mainly involved in metabolic processes and pathways. GO terms associated with the cluster of basally expressed genes included sensory perception of chemical stimulus ($p= 1.10\times10^{-30}$), RNA metabolic process (7.61×10^{-29}), sensory perception of smell ($7.87\text{x}10^{-29}$), metabolic process (1.64×10^{-25}) and primary metabolic process (7.14×10^{-25}). The expression profile of these genes remained largely unaltered under the influence of LPS.

The cluster of down regulated genes 7hrs post-LPS stimulation contained 2,151 genes (Appendix R). The gene enrichment analysis using PANTHER (Appendix Q) indicated that the processes associated with this cluster mainly

involved cell-cycle related processes. The top GO terms associated with this cluster included sensory perception of chemical stimulus (1.28×10^{-16}), RNA metabolic process (7.17×10^{-6}), glycolysis (2.68×10^{-5}), cell adhesion (3.86×10^{-05}) and biological adhesion (3.86×10^{-05}). The down-regulation of the cell-cycle genes, which is also observed in mice (Nilsson, Bajic et al. 2006), indicates there might be an anti-proliferative effect of LPS on macrophages.

The cluster of upregulated genes 7hrs post-LPS stimulation contained 774 genes (Appendix T). Genes in this cluster were mainly involved in innate immune signalling and cellular defence processes as illustrated by the GO term analysis in Appendix S. Significant GO terms associated with this cluster included immune system process (3.93×10^{-10}), cellular defence response (2.19×10^{-7}) and immune response (1.55×10^{-6}). As has been observed in the pig (Kapetanovic, Fairbairn et al. 2012) and mouse studies (Nilsson, Bajic et al. 2006), the upregulated genes cluster included cytokines involved in interferon induced signalling such as: *IFI35, IFI44, IFI6, IFIT3* and *IFNG.* Additionally, expression of interleukin genes such as *IL10RA, IL12B, IL16, IL1RAP, IL21* and *IL4R* was greatly induced by LPS. Also included in this cluster were genes involved in the mitogen activator protein kinase (MAPK) signalling cascade and in particular, the *MAP3K8* gene that is also upregulated in mice and associated with activation of *TNF*. Similarly, *TNF* related genes such as *TNFAIP2, TNFRSF8, TNFSF13B, and TNFSF8* were also upregulated by LPS.

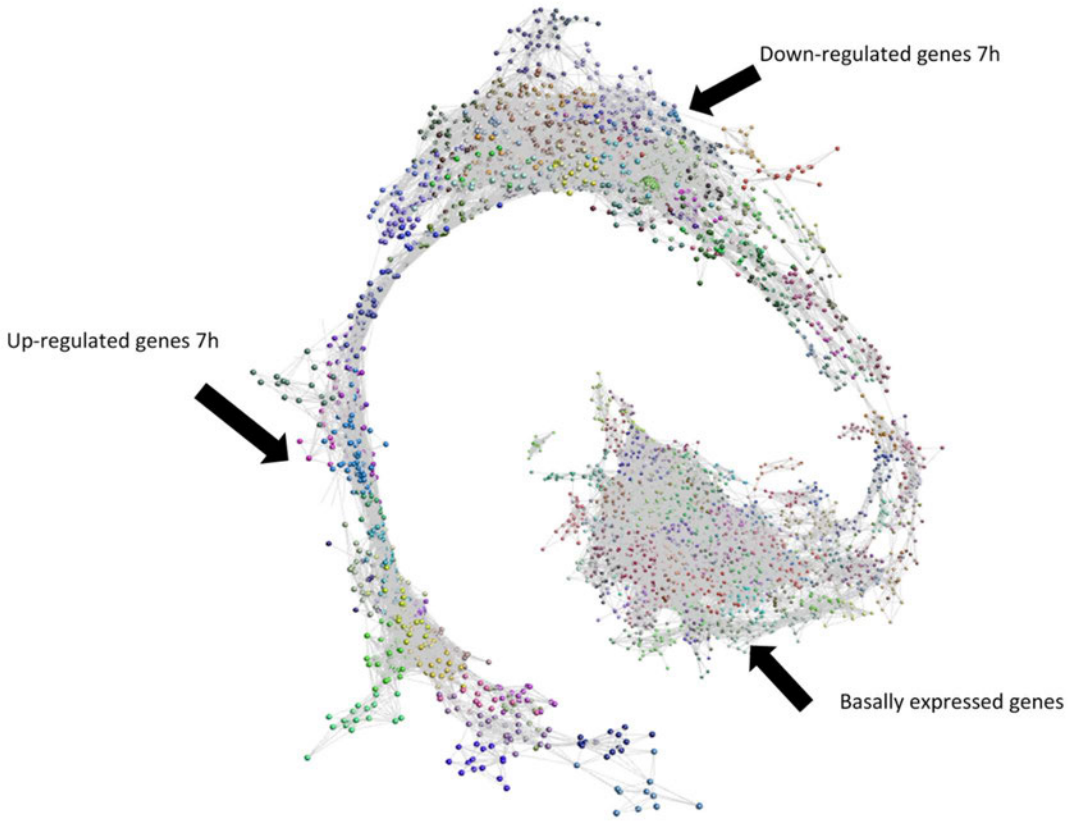


Figure 3-8: Gene-to-Gene network graph of Goat BMDM post LPS stimulation. Each node represents a gene and the edges (lines) correspond to correlations between individual measurements above the set threshold of r=0.99, MCL inflation=2.2, pre-inflation=6.0, and minimum nodes=20

3.4 Discussion

As already demonstrated in previous studies, transcriptomic analysis of macrophages from livestock species like pig (Freeman, Ivens et al. 2012, Kapetanovic, Fairbairn et al. 2012), sheep (Clark, Bush et al. 2017) and cattle (Harhay, Smith et al. 2010, Casey, Meade et al. 2015) provides useful information for understanding the innate immune system. The domestic goat is an economically important livestock species especially in low- and middle-income countries and is susceptible to a wide range of pathogens. However, there are few published studies focused on the transcriptional control of the innate immune response in goats. Few specific reagents are available for immunological studies in goat, with most studies relying on cross-reactivity with sheep and cattle antibodies (Entrican 2002, Hope, Sopp et al. 2012), Most available data on goat macrophages is based upon monocyte derived cells (Taka, Liandris et al. 2013, Walia, Kumar et al. 2015), but the blood monocyte count in goats and sheep is very low (C Pridans Pers. Comm.). As such I chose to focus on bone marrow derived macrophages, to generate a methodology for successful cryopreservation and recovery that would provide an alternative source of progenitors that could be utilised to study the goat innate immune response.

In this chapter, I developed and optimised the culture conditions for goat BMDM, differentiating them successfully and characterising them using FACS and pHrodo staining. Using these cultures of BMDM I was able to perform stimulation with LPS and measure the transcriptional response of a small subset of immune genes using RT-qPCR and the whole transcriptome with RNA-Seq. The dataset I have generated provides a resource of transcriptional information that can be used to improve our understanding of the innate immune response in goat.

The RNA-Seq analysis of goat BMDM carried out in section 3.3.5 revealed more than 700 genes that were up-regulated by LPS stimulation. The up-regulation of important cytokines and chemokines such as *TNF*, interferon-associated genes including *IFI6, IFIT3* and *IFNG* and interleukins such as

IL10RA, IL12B, IL16 and *IL1RAP* was similar to patterns of gene expression post LPS stimulation observed in sheep (http://biogps.org/sheepatlas/#goto=welcome), indicating that the innate immune system in these species has evolved in a similar manner.

One example of shared evolutionary patterns in the innate immune response is Nitric Oxide (NO) production which is widely recognized as an essential component of anti-microbial defence, especially in rodents having anti-microbial and anti-tumoricidal activity (Ito, Koide et al. 2005, Takacs, Swierzy et al. 2012). Neither pig nor human produce NO in response to LPS (Kapetanovic, Fairbairn et al. 2012). A previous study comparing goat and cattle macrophages production of NO in response to LPS (Heiko Adler, Barbara Adler et al. 1996) illustrated that the production of NO by activated goat macrophages was low and nearly undetectable reaching levels lower than that of resting cattle macrophages. Recently, colleagues have also reviewed the species-specific differences in NO production following stimulation with LPS (Young, Bush et al. 2018) and extended the number of species that do not induce *NOS2*/NO production to include sheep and horse. Conversely, water buffalo shared similar expression of *NOS2* with cattle. The differences between ruminant species were attributed to insertion of mobile genetic elements in the *NOS2* promoter region that confer LPS-responsiveness in bovids (Young, Bush et al. 2018). Bovids and small ruminants differ in their susceptibility to a diversity of pathogens and variation in NO production could provide one explanation.

Species-specific variation in response to pathogen challenge are likely to be driven by differences in gene expression post infection. The dataset generated in this chapter from BMDM provides a useful resource for comparative analysis with other species. Sheep and goats are both small ruminant mammals and are similar in their physiology. They also share susceptibility to a wide range of viral, bacterial, parasitic and prion pathogens, including multiple potential zoonoses (Sherman 2011), but there have been few comparisons of relative susceptibility or pathology between the species to the same pathogen, nor the

nature of innate immunity. In chapter 4 a comparative analysis of species-specific gene expression is performed to reveal transcriptional similarities and differences between sheep and goats.

Chapter 4 Comparative transcriptomic analysis of immune related genes in the domestic goat and sheep.

The RNA-Seq and differential expression analysis described in this chapter is adapted from the authors published work (Muriuki, Bush et al. 2019).

4.1 Introduction

Diversity in disease susceptibility in mammals is partly attributed to rapid pathogen evolution, the sustained pressure exerted on the host's immune system to co-evolve, resulting in changes in gene expression and regulation. This chapter aims to investigate the similarities and differences in the LPS-inducible macrophage transcriptional responses of domestic goat and sheep. As mentioned in Chapter 2 of this thesis, the goat gene expression project was undertaken in the same laboratory as the sheep gene atlas study (Clark, Bush et al. 2017) adopting similar experimental strategies to enable a cross-species comparison. In both cases, crossbred animals were used, and macrophages were isolated and cultured using similar protocols. This chapter contains two comparative analyses; the first is a global comparison of transcription between the goat mini-atlas dataset detailed in Chapter 2 and an equivalent subset of data from the sheep gene atlas. The second takes data from a parallel study in our laboratory investigating transcriptomics of sheep macrophages and their response to LPS (M. McCulloch, 2018) and integrates them with results from Chapter 3 of this thesis to enable a comparative analysis of the immune response in goat and sheep BMDM in response to LPS.

Although the basic biology of macrophage signalling is highly conserved in mammals, there are fundamental differences in LPS-regulated gene expression between mammalian species (Young, Bush et al. 2018). Previous studies from our laboratory have highlighted similarities and differences in the macrophage response to LPS between laboratory mice, humans and pigs. In these comparisons, pigs were more similar to humans, providing a more informative model of human disease than mice (Kapetanovic, Fairbairn et al.

2012). The differences between mouse and human macrophage responses to LPS were associated with divergence of regulatory elements in the inducible promoters (Fairbairn, Kapetanovic et al. 2011, Kapetanovic, Fairbairn et al. 2012, Kate Schroder, Katharine M. Irvine et al. 2012).

One of the widely appreciated species-specific transcriptional differences is the LPS induction of nitric oxide synthase (*NOS2*) and subsequent production of the antimicrobial effector nitric oxide (NO) by macrophages (Jungi TW, Adler H et al. 1996). A previous study demonstrated that *NOS2* induction by LPS was observed in cattle macrophages but absent in goat (Heiko Adler, Barbara Adler et al. 1996). This finding was re-examined in our lab (Young, Bush et al. 2018) and has been discussed in detail in Chapter 3 of this thesis. In (Young, Bush et al. 2018) data from the goat BMDM experiment (Chapter 3) was combined with RNA-Seq data from pig, cattle, water buffalo, sheep, horse, rat and human macrophages. These data were used to investigate the mechanisms controlling variation in the expression of genes involved in arginine metabolism, looking specifically at *NOS*2 as well as other LPS-inducible transcripts (Young, Bush et al. 2018). Rats were similar to mice, expressing large amounts of *NOS2,* arginine transporter *SLC7A2,* arginase 1 *(ARG1)*, GTP cyclohydrolase (*Gch1*) and argininosuccinate synthase (*ASS1*). These responses to LPS were not conserved across the other species and the response of goats and sheep are described within this chapter. The results from this multi-species transcriptional comparison confirmed that divergence in expression of *NOS2* and other genes is mainly driven by evolution of *cis*-acting regulatory elements. The insertion of a BOV-A2 retrotransposon in the promoter region of cattle and water buffalo was associated with increased expression of *NOS2* and NO production in the bovid lineage (Young, Bush et al. 2018).

Evolution of transcriptional regulation is linked to species-specific variations in disease susceptibility. Variation in susceptibility to disease has been widely documented across all major farm animals, reviewed in (Bishop and Woolliams 2014). For instance, ruminants differ in their response to a wide range of

economically important infectious bacterial diseases. *Mycobacterium bovis* is an important pathogen of cattle and water buffalo but causes limited disease in sheep despite prolonged co-grazing with cattle (Munoz Mendoza, Juan et al. 2012) while Johne's disease caused by *Mycobacterium avium* subspecies *paratuberculosis* (MAP) exhibits strain specific variation in pathogenicity between cattle, sheep and goats (Stevenson 2015). In some cases, one ruminant species acts as a reservoir for a pathogen that adversely affects another species. For example, Malignant Catarrhal Fever (MCF) caused by alcelaphine herpesvirus 1 (AlHV-1) elicits asymptomatic infection in sheep and African wildebeest but is fatal in cattle and water buffalo (Russell, Stewart et al. 2009, Wambua, Wambua et al. 2016). Although the factors determining disease resistance are polygenic and involve an interplay of multiple genes, quantitative trait loci (QTL) studies have identified several candidate genes associated with variable disease resistance within ruminants. For instance, variations in *CD14, CD18, CXCR2* and MHC genes are associated with mastitis resistance in certain breeds of cattle (Ibeagha-Awemu, Kgwatalala et al. 2008) and *SLC11A1* variation controls the sheep and goat responses to Johne's disease (LA Reddacliff, K Beh et al. 2005, Cecchi, Russo et al. 2017). Similarly, resistance to *Haemonchus contortus* infections in sheep and goats is associated with a stronger Th2-type immune response in the resistant animals (Alba-Hurtado and Munoz-Guzman 2013).

As discussed in the main introduction, goats and sheep are closely related, they diverged four million years ago and commonly occupy similar environmental niches (Jiang, Xie et al. 2014). Immune-associated genes are known to undergo rapid evolution between closely-related species (as evidenced by high rates of non-synonymous amino acid substitution) and are polymorphic within species at the protein level (Ellegren 2008). As discussed previously macrophages are an important immune cell involved in the inflammatory and immune response. Amongst inbred mouse strains, DNA sequence variation in regulatory elements has been associated with differences in both basal and inducible gene expression in macrophages (Link, Duttke et al. 2018). As such we would expect that species-specific differences

in the immune response might be reflected in the transcriptional patterns observed in macrophages under pathogen challenge. This chapter will highlight some of the transcriptional similarities and differences between goat and sheep alveolar macrophages and bone marrow derived macrophages post challenge with LPS. This will help us to understand how the innate immune response is regulated in two closely related ruminant species.

4.2 Materials and Methods

4.2.1 Tissue samples included in the analysis

Approval to carry out this project was issued by the Animal Welfare and Ethics Review Boards of The Roslin Institute and the University of Edinburgh. Details of the goat data included in the analysis for this chapter (see Chapter 2 for a detailed description of the samples) is summarised in Table 4-1. For comparative analysis with sheep, a subset of data from the large RNA-Seq sheep dataset (Clark, Bush et al. 2017) was used. Details of the data from sheep included in the analysis for this chapter are also included in Table 4.1. A total of 54 goat and 52 sheep RNA-Seq libraries were utilised for the analysis described in this chapter.

4.2.2 RNA-sequence data and processing

The experimental protocols for tissue collection, cell isolation, RNA extraction and library preparation used to generate the goat RNA-Seq dataset are described in Chapter 2 and 3 of this thesis. The sheep RNA-Seq dataset was derived from the published sheep gene atlas project (Clark, Bush et al. 2017) carried out in the same laboratory using similar experimental protocols and data processing methods. RNA-Seq data was processed in Kallisto to produce gene level estimates as transcripts per million (TPM) using the methodology described in Chapter 2 of this thesis.

Table 4-1:Details of the data from goat (from Chapter 2) and from sheep (Clark, Bush et al. 2017) used for comparative analysis

Species	Sample type	Age	Tissue type	Number of Individuals
Goat	Core tissues	5-6 days old	adrenal gland,colon,ileum, kidney,liver,rumen, muscle,skin,spleen ,testes,thymus	4, males
	Brain	5-6 days old	cerebellum, cortex	2, males
	Female reproduction	5-6 days old	ovary, fallopian tube, uterus, uterine horn	1, female
	BMDM	5-6 days old	0hr(-LPS), 7hr(+LPS)	3, males
	AM	5-6 days old	24hr(-LPS)	2, males
Sheep	Core tissues	adult	adrenal gland, cerebellum, kidney, liver, muscle, skin, spleen, testes, thymus	3, males
	GI tract	lamb/1week	colon, ileum, rumen,	3, males
	Female reproduction	adult	intercaruncular tissue, fallopian tube, ovary	3, females
	BMDM	adult	0hr(-LPS), 7hr(+LPS)	3, males
	AM	adult	24hr(-LPS)	3, males

4.2.3 Network Cluster Analysis

The network cluster analysis of the goat dataset using the network visualisation tool Miru (Kajeka Ltd, Edinburgh, UK) is described in detail in Chapter 2 of this thesis. To enable comparative analysis, network cluster analysis was performed on the subset of data from sheep (summarised in Table 4.1 and included in Appendix U) using the same parameters on Miru as the goat dataset (MCL 2.2, r=0.83) to generate a gene-to-gene network graph. An all-against-all comparison of the top 30 largest clusters was carried out for every goat with every sheep cluster (top 30 largest clusters). This method allowed identification of the number of annotated genes held in common between the goat and sheep clusters.

4.2.4 Differential Expression Analysis

4.2.4.1 Comparing BMDM response between goat and sheep 0 and 7hr post-LPS stimulation

The differential expression analysis of goat BMDM was performed using the tximport pipeline and edgeR (Robinson, McCarthy et al. 2009) packages. The analysis of differentially expressed genes (DEG) analysis in sheep BMDM (+/--) LPS was performed as part of a parallel study (M. McCullouch, 2018). All comparative analysis of differentially expressed genes in goat and sheep BMDM at 0hr and 7hr post-LPS stimulation was performed using R (version 3.4.0).

The lists of DEGs in goat and sheep BMDM (Appendix V and Appendix W respectively) were first filtered at FDR<10% and then merged based on GENE_ID using the *inner_join* function to only return the observations that overlapped between goat and sheep (i.e. genes with similar annotation in goat and sheep). This was performed using the following R script:

merged <- inner_join(x = goat[goat$FDR < 0.1 ,], y = sheep[sheep$FDR < 0.1,], by = "GENE_ID").

Next, a dissimilarity index (Dis_Index) was determined by taking the absolute difference (ABS) of |log2FC| between sheep and goat using the formula*:*

ABS(|log2FC| *Sheep*- |log2FC| *Goat).*

A high Dis_Index indicated that a gene was differently regulated in goat and sheep, likely due to unique species-specific responses.

The direct comparison of DEGs in goat and sheep BMDM was then merged with the results of the network cluster analysis performed in Miru (section 2.2.3) to identify genes in goat and sheep that were involved in the same biological pathway. To do this, an intersection was calculated using the formula:

Intersection = (N/T) x 100

Where: N= number of genes shared between the sheep and goat cluster and

T = number of genes in the smaller cluster.

A high intersect value indicated that the clustering pattern of the gene was highly similar in goat and sheep, and therefore probably had a similar biological function in the two species.

4.2.4.2 Comparing AM response in goat and sheep

Similarly, differential expression analysis was used to compare transcriptional patterns in alveolar macrophages in goat and sheep. The gene level expression estimates from alveolar macrophages from two male goats and three male sheep were analysed using the same method with tximport and edgeR packages utilised for BMDM. Only genes with the same annotation, expressed at a raw read count of more than 10, FDR<10% and |log2FC| $\geq$ 2 in both goat and sheep were included in the analysis.

4.2.5 Validation of gene expression using qPCR

To validate gene expression estimates for target immune genes, RNA was extracted from BMDMs (+/- LPS) from age-matched one-week old sheep, as described in Chapter 2 and amplified using quantitative PCR. Two genes (*TNF* and *NOS2)* were selected for validation based on results observed from goat BMDM in Chapter 3 (primer sequences and citations are included in Chapter 3). cDNA was synthesised using the same protocol described in Chapter 3

(section 3.2.8). All RT-qPCR reactions were carried out using the SYBR Green quantitative PCR system (ThermoFisher) in 20 µL reactions on MicroAmp Optical 96-well Reaction plates (ThermoFisher). Each reaction contained 10uL of 2X *Power* SYBR Green master mix, 10uM forward and reverse primers, 3.8uL of water and 5uL of diluted cDNA. Positive and negative controls were included for each set of reactions. Relative quantitation of transcript abundance against housekeeping gene *GAPDH* was determined using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

4.3 Results

4.3.1 Overlap of co-expression gene clusters between goat and sheep

An all-against-all comparison of the top 30 largest clusters for each species was performed to identify the number of annotated genes that were shared between the two species (Table 4.2). The percentage of annotated genes shared between the two species for each cluster (or 'functional class') is summarised in Figure 4.1. There was substantial overlap of genes shared between some of the larger clusters, particularly those associated with cellular processes such as cell cycle (85%) and tissue types such as fallopian tube (64%) and cerebellum (60%) (Figure 4.1). There was also significant overlap (>50%) between both species for GI tract clusters including liver, ileum, rumen, and colon. This reflects similarities in the function of these tissues in the two species. Clusters exhibiting less similarity between the two species included testes (30%) and ovary (17%). These are more likely to be developmental stage-specific differences rather than species-specific differences in transcription. Gene expression has been shown to change with developmental stage in the ovary in sheep (Clark, Bush et al. 2017) and in the testes of sheep (Zhang, Zhang et al. 2019) and goats (Faucette, Maher et al. 2014). One of the caveats of this analysis is that data from goats was from neonatal animals and the sheep data is from adult animals. Differences in the age of individuals used in this analysis, adult sheep and neonatal goats, may have had a significant effect on gene expression. The strong overlap of clusters of genes associated with general biological functions such as cell-cycle and skeletal muscle in goat and sheep Table 4-2 does however indicate that the datasets were broadly comparable. Similarly, the close pairing of the colon, ileum and rumen clusters also suggest that transcriptional patterns in the ruminant GI tract are conserved across the two species, irrespective of age-specific differences.

Table 4-2 :Goat and Sheep co-expression cluster comparison

Cluster ID (goat)	Profile Description (Goat)	Cluster ID (sheep)	Profile Description(Sheep)	% of annotated genes in common	No. of genes in goat cluster (annotated/ total)	No. of genes in sheep cluster (annotated/ total)	No. of annotated genes common between goat and sheep
Cluster0001	Cortex	Cluster0002	CNS	56.18	1461/1592	1189/1455	668
Cluster0002	Cell-cycle	Cluster0014	House keeping	84.91	849/1017	159/169	135
Cluster0003	HKG_gene regulation	Cluster0005	House keeping	41.16	881/972	396/456	163
Cluster0004	Liver	Cluster0006	Liver	57.83	336/467	249/370	144
Cluster0005	Testes	Cluster0001	Testes	30.55	275/454	1142/1711	84
Cluster0006	Cerebellum	Cluster0002	CNS	60.51	352/405	1189/1455	213
Cluster0007	Skin	Cluster0012	Skin	47.01	225/398	117/204	55
Cluster0008	Fallopian Tube	Cluster0003	Fallopian Tube	64.16	293/363	471/611	188
Cluster0009	Skeletal muscle	Cluster0007	Skeletal muscle	66.77	326/353	316/355	211

Table 4.2 continued

Cluster ID (goat)	Profile Description (Goat)	Cluster ID (sheep)	Profile Description(Sheep)	% of annotated genes in common	No. of genes in goat cluster (annotated/ total)	No. of genes in sheep cluster (annotated/ total)	No. of annotated genes common between goat and sheep
Cluster0010	Spleen	Cluster0019	Spleen	48.61	175/281	72/135	35
Cluster0011	Rumen	Cluster0015	Rumen	55.56	182/244	135/164	75
Cluster0012	Colon	Cluster0017	Colon	55.45	198/243	110/139	61
Cluster0013	Adrenal Gland	Cluster0016	Adrenal Gland	38.4	175/206	125/159	48
Cluster0014	Fibroblasts	Cluster0008	Immune phagocytosis	5.49	182/191	255/298	10
Cluster0015	Immune phagocytosis	Cluster0008	Immune phagocytosis	33.58	134/160	255/298	45
Cluster0016	Kidney development	Cluster0011	Kidney	10.64	141/158	176/210	15

Table 4.2 continued

Cluster ID (goat)	Profile Description (Goat)	Cluster ID (sheep)	Profile Description(Sheep)	% of annotated genes in common	No. of genes in goat cluster (annotated/ total)	No. of genes in sheep cluster (annotated/ total)	No. of annotated genes common between goat and sheep
Cluster0017	HKG_molecular chaperones	Cluster0008	Immune phagocytosis	7.28	151/155	255/298	11
Cluster0018	Ovary	Cluster0001	Testes	17.91	67/129	1142/1711	12
Cluster0019	House keeping	Cluster0005	House keeping	13.45	119/125	396/456	16
Cluster0020	Kidney	Cluster0011	Kidney	45.74	94/122	176/210	43
Cluster0021	Pathway_metabolism	Cluster0007	Skeletal muscle	9.52	84/105	316/355	8
Cluster0022	House keeping	Cluster0004	House keeping	13.64	88/102	362/545	12
Cluster0023	House Keeping	Cluster0005	House keeping	21.18	85/100	396/456	18
Cluster0024	Uterus	Cluster0001	Testes	6.58	76/97	1142/1711	5

Table 4.2 continued

Cluster ID (goat)	Profile Description (Goat)	Cluster ID (sheep)	Profile Description(Sheep)	% of annotated genes in common	No. of genes in goat cluster (annotated/ total)	No. of genes in sheep cluster (annotated/ total)	No. of annotated genes common between goat and sheep
Cluster0025	Ileum	Cluster0013	Ileum	57.5	80/94	149/183	46
Cluster0026	Pathway_ribosomal	Cluster0027	House keeping	46.43	85/91	56/100	26
Cluster0027	LPS-inducible	Cluster0020	LPS-inducible	40.32	62/86	110/134	25
Cluster0028	Mitochondrial	Cluster0004	House Keeping	7.69	78/81	362/545	6
Cluster0029	Early Development	Cluster0001	Testes	14.29	7/67	1142/1711	1
Cluster0030	House Keeping	Cluster0017	Colon	9.38	64/65	110/139	6

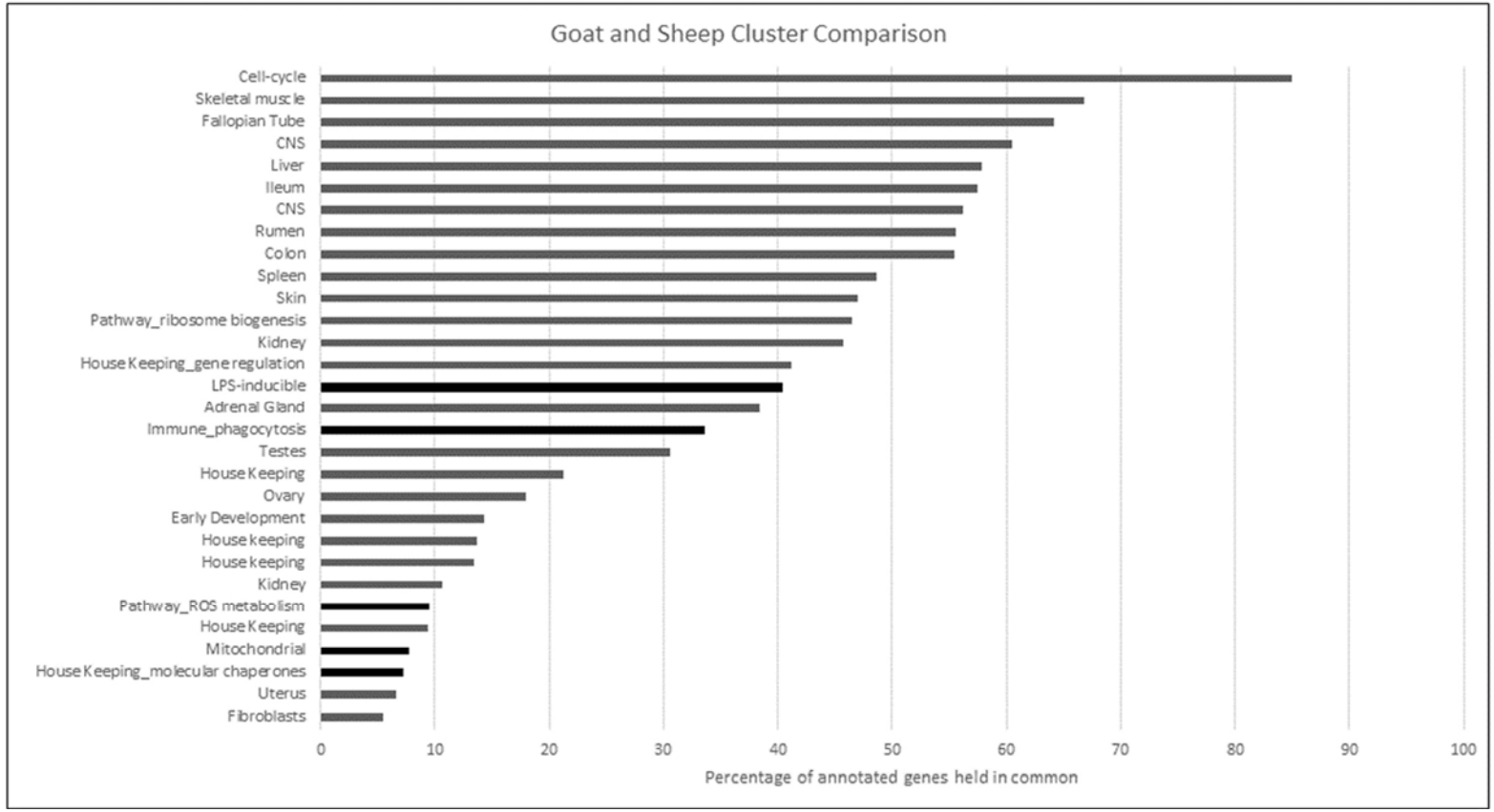


Figure 4-1:Goat-Sheep cluster comparison. Comparison of co-expression clusters in goat and sheep gene expression atlases shows high similarity of cell-cycle functions (84.9%), skeletal muscle (66.7%) and tissues from the GI tract (Ileum, Colon, Rumen). LPS-Inducible and phagocytosis clusters show 40.3% and 33.58% similarity respectively while the house-keeping genes containing molecular chaperones in the goat has a 7.28% similarity with the sheep phagocytosis cluster. The general cluster comparison suggests species-specific differences related to immune response

4.3.2 Comparative expression of key genes required for response to LPS.

Figure 4.2 shows the TLR4 receptor complex, adaptors and key regulators involved in both the MyD88-dependent and MyD88-independent pathways of response to LPS. In principle, global differences between the two species might arise if particular signalling molecules are differentially-expressed. A comparison of the expression of molecules required for signal transduction following activation by TLR4 in goat and sheep macrophages is presented in Table 4.3. All of the signalling components shown in Figure 4.2 were expressed at detectable levels in goat and sheep BMDM, but there were quantitative differences between species. Goat macrophages expressed much lower levels of *TLR4, CD14* and *LY96* but expressed much higher levels of *TRAM1, TAB2* and the transcription factor *IRF3* in comparison to sheep (Table 4.3). In both species, BMDM expressed mRNA encoding CSF1 and its receptor, CSF1R. *CSF1* was highly expressed at both 0 and 7 hours in goat but was LPS inducible in sheep macrophages (Table 4.3).

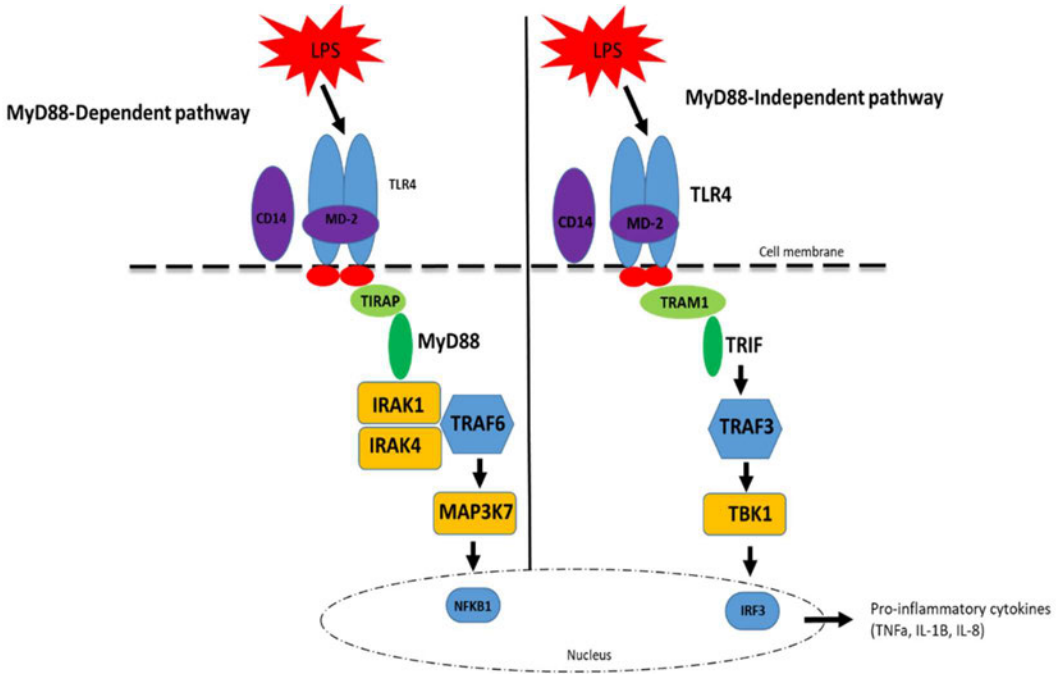


Figure 4-2:The TLR4 receptor complex highlighting MyD88-dependent and MyD88 independent mediators. Figure adapted from (Akira, Uematsu et al. 2006)

Table 4-3:Expression of transcripts required for LPS response in BMDM

Transcript	Gene Name	Goat			Sheep		
		|log2FC|	TPM_0 hr	TPM_7 hr	|log2FC|	TPM_0 hr	TPM_7 hr
CD14	CD14 Molecule	1.18	76.26	159.38	-0.22	668.63	434.36
TLR4	Toll Like Receptor 4	0.74	25.81	82.76	0.93	72.21	110.04
LY96	Lymphocyte Antigen 96 (Myeloid Differentiation Protein-2)	-0.14	81.44	68.57	0.62	262.84	330.67
TIRAP (MAL)	TIR Domain Containing Adaptor Protein	0.38	17.19	21.21	0.50	23.41	26.38
MyD88	Myeloid Differentiation Primary Response 88	-0.22	21.75	17.60	-0.29	70.32	46.95
TRAM1	Translocation Associated Membrane Protein 1	0.37	351.85	425.50	0.35	74.32	77.21
IRAK1	Interleukin 1 Receptor Associated Kinase 1	-0.39	24.81	17.98	-0.09	31.34	24.16
IRAK4	Interleukin 1 Receptor Associated Kinase 4	0.26	18.48	20.86	0.32	43.77	44.46
TICAM1 (TRIF)	Toll Like Receptor Adaptor Molecule 1	0.55	8.12	11.23	0.92	1.21	1.81
TAB1	TGF Beta-Activated Kinase-Binding Protein 1	-0.36	17.23	12.79	-0.27	10.85	7.38
TAB2	TGF Beta-Activated Kinase-Binding Protein 2	0.69	52.92	79.42	1.08	18.35	31.28
MAP3K7(TAK1)	Mitogen-activated protein kinase kinase kinase 7	0.11	67.22	68.41	-0.06	34.85	27.18
TRAF6	TNF Receptor Associated Factor 6	0.14	14.44	15.10	0.22	6.54	6.21
TRAF3	TNF Receptor Associated Factor 3	0.58	11.80	16.61	1.56	6.89	16.21

Table 4.3 continued

Transcript	Gene Name	Goat			Sheep		
		|log2FC|	**TPM_0 hr**	**TPM_7 hr**	|log2FC|	**TPM_0 hr**	**TPM_7 hr**
TBK1	TANK Binding Kinase 1	0.74	22.62	35.28	0.26	23.02	22.41
NFKB1	Nuclear Factor Kappa B Subunit 1	2.55	29.29	158.48	2.34	27.97	111.39
IRF3	Interferon Regulatory Factor 3	0.40	45.43	55.85	0.29	9.69	9.44
CSF1	colony-stimulating factor-1	0.61	115.66	161.47	2.67	35.22	190.01
CSF1R	colony-stimulating factor-1 receptor	-0.49	51.28	34.16	-1.09	152.21	60.97
IL34	Interleukin 34	0.97	2.61	4.73	-1.00	0.64	0.26

4.3.3 Validation of LPS response in sheep BMDM using RT-qPCR

The sheep used to generate BMDM for the sheep gene atlas were approximately two years of age while the goats used in this study were neonatal. As such, it was necessary to validate the LPS response is a small subset of immune genes in age-matched animals. mRNA expression of *TNF* and *NOS2* in age and sex-matched sheep and goats were measured using RT-qPCR and transcription levels calculated as relative fold changes against housekeeping gene GAPDH. As illustrated in Fig 4.3, sheep BMDM showed induced expression of both *TNF* and *NOS2* after 7hr stimulation with LPS, similar to observations made in goat BMDM (Chapter 3).

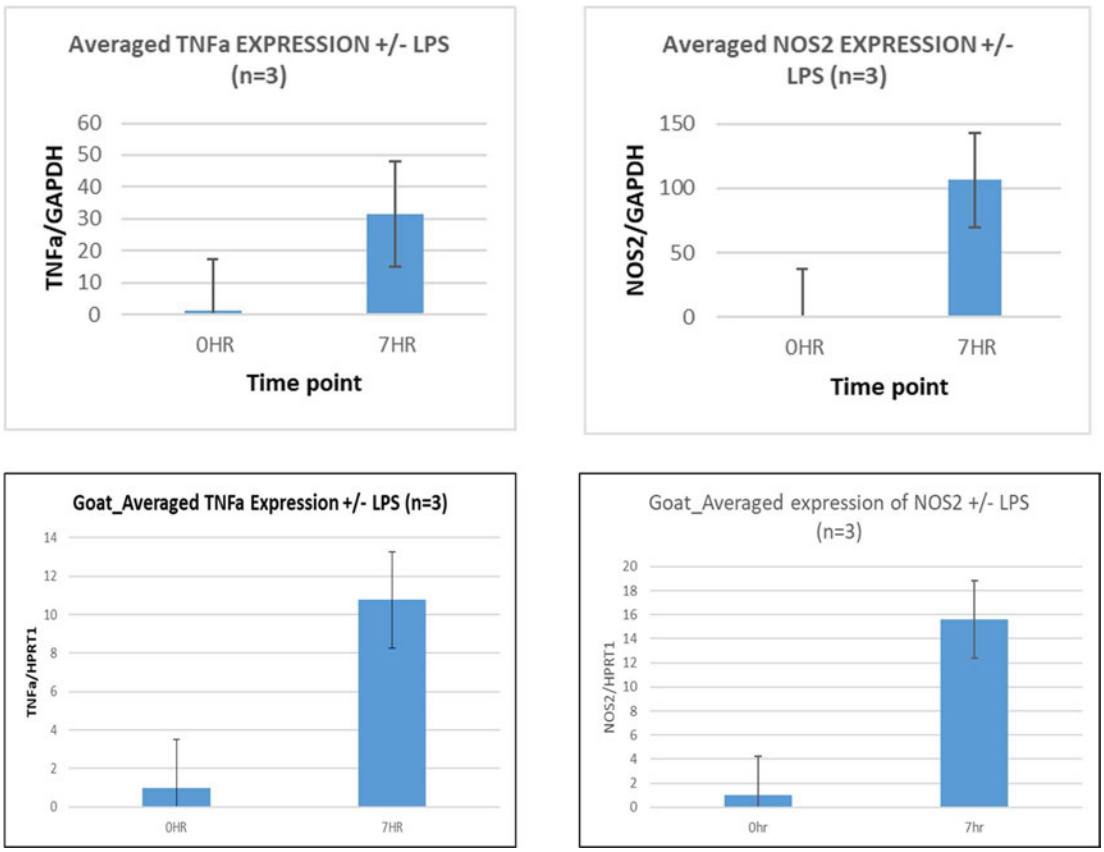


Figure 4-3: Expression of TNF and NOS2 mRNA in sheep BMDM after stimulation with 100ng/mL LPS for 7hours (Top panel). Averaged expression levels are presented as relative fold changes against the housekeeping gene GAPDH. Bottom panel: Expression of TNF and NOS2

mRNA in goat BMDM after stimulation with 100ng/mL LPS for 7hours (adapted from Chapter 3).

4.3.4 Differences in BMDM response to LPS stimulation revealed by comparative analysis

The DEG in goat and sheep BMDM in response to LPS stimulation are included in Appendix V and W respectively. Comparative analysis of the DEG in goat and sheep resulted in 188 genes (Appendix X) that showed significant differences in gene expression response to LPS (FDR<10%, $|\log_2 FC| \geq 2$) that were shared between goats and sheep. The most dissimilar genes between goat and sheep (Dis_Index≥2) are illustrated in Fig 4.4. The top right quadrant of the graph contains genes that were upregulated in both goat and sheep BMDM in response to LPS stimulation. This included several immune genes, *IL33, IL36B, PTX3, CCL20, CSF3* and *CSF2* that showed a higher level of induction by LPS in sheep macrophages compared to goat, and *ICAM1, IL23A, IFIT2, TNFSF10,* and *TNFRSF9* that showed a higher level of induction by LPS in goat macrophages compared to sheep. The top left quadrant contained genes that were upregulated in sheep but downregulated in goat macrophages and included *CARD10, KIT, STC1* and *KCNJ8.* The bottom right quadrant contained genes that were upregulated in goat but downregulated in sheep macrophages including *FOXO1, WWC1, TGFBR3, HEXDC* and *IGFBP4.* Similarly, the proto-oncogene *MERTK* and guanine nucleotide exchange factor *RAB3IL1* showed a high level of intersect between goat and sheep, with similar downregulation response in both goat and sheep macrophages while *KIT* was upregulated in sheep macrophages but downregulated in goat.

An overlay of DEG genes with the clusters generated in Miru to determine whether the differentially expressed genes were involved in the same biological pathway in goat and sheep is illustrated in Fig 4.5. The degree of intersect is represented by the size of the node, and genes with a high intersect (intersect %>1) are shown. *IL33, IFIT2, IL23A* and *TNFSF10* were upregulated in both goat and sheep macrophages (although with different levels of

induction) and showed a high level of intersect between goat and sheep, indicating that they were involved in the same biological pathway in the two species. As the genes were differentially inducible in goat and sheep, it is possible they are differentially regulated and may play a role in species-specific immune responses.

In mammals one of the key innate immune defences against infection is the interferon response (Schoggins, Wilson et al. 2011, Schoggins 2019). A comparison of the absolute level of induction of a select set of interferon inducible-inducible genes is presented in Table 4.4. The majority of these genes showed a similar level of expression in goat and sheep BMDM (+/-) LPS. *IL33* was the only exception, which showed twice as much induction after LPS treatment in sheep relative to goat. Future studies could investigate the species-specific interferon response in sheep and goats by challenging the cells with type 1 interferon as in (Shaw, Hughes et al. 2017).

Table 4-4:Expression of select set of interferon-inducible genes

Transcript	Gene Name	Goat			Sheep		
		|log2FC|	TPM_0 hr	TPM_7 hr	|log2FC|	TPM_0 hr	TPM_7 hr
IL33	Interleukin 33	4.75	1.51	37.41	10.28	0.09	92.55
IFIT2	Interferon Induced Protein With Tetratricopeptide Repeats 2	4.59	1.55	33.38	1.45	16.99	36.96
IL23A	Interleukin 23 Subunit Alpha	4.71	1.26	29.10	1.35	26.28	54.44
TNFSF10	TNF Superfamily Member 10	6.11	0.06	4.26	2.33	1.40	5.54
ISG15	Interferon-Stimulated Protein, 15	5.86	6.23	325.32	5.88	8.54	401.47
MX1	MX Dynamin Like GTPase 1(Interferon-Inducible Protein P78)	4.00	26.01	367.21	4.86	14.89	341.57
IFIH1	Interferon Induced With Helicase C Domain 1	3.46	6.52	64.49	2.74	8.01	42.09
IL1B	Interleukin 1 Beta	7.95	6.93	1552.06	8.72	6.94	2179.39

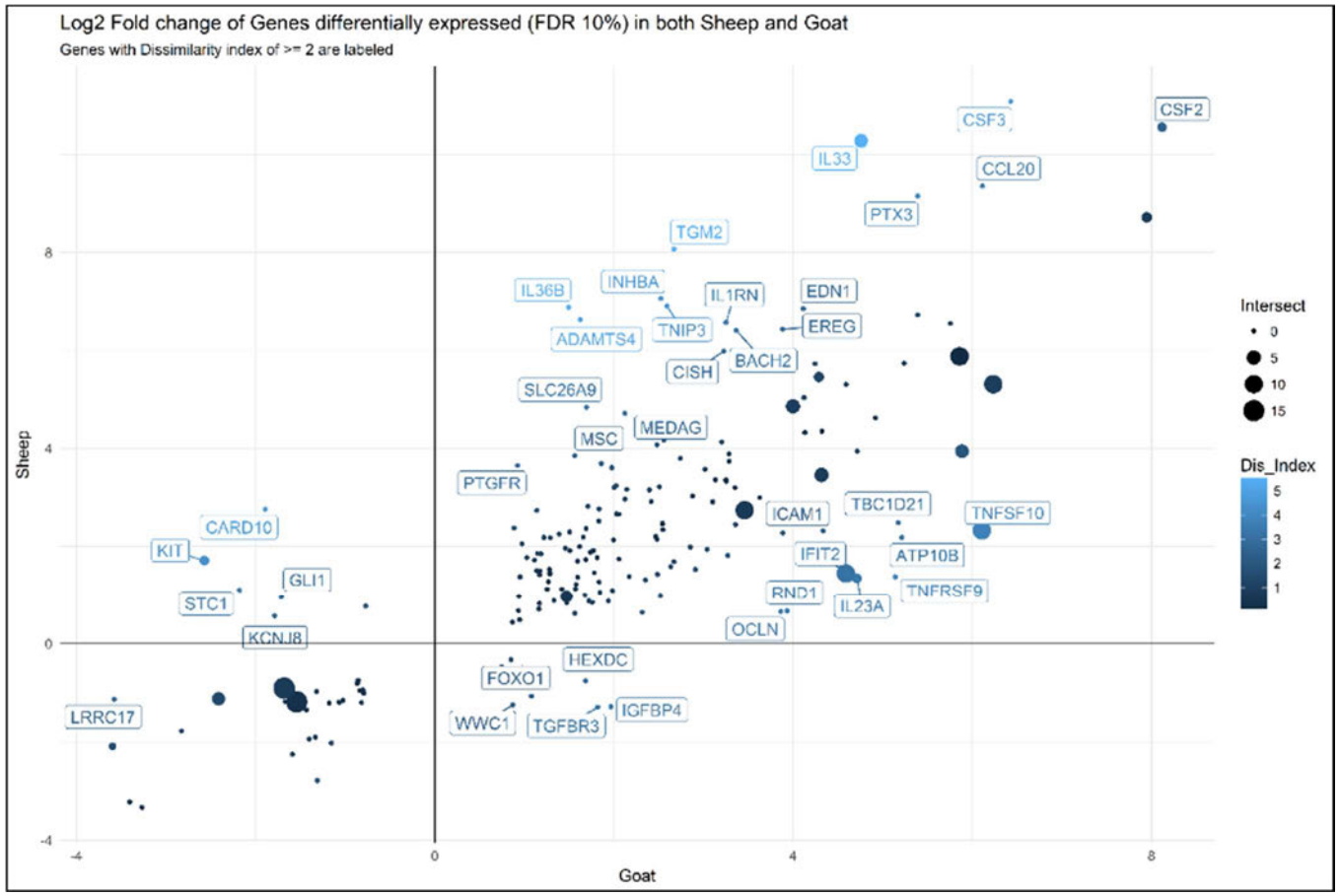


Figure 4-4:Comparative analysis of differentially expressed genes in goat and sheep BMDM. The genes showing the highest level of dissimilarity in response to LPS between goats and sheep (Dis_Index≥2) are shown. (A) Top left quadrant: genes that up-regulated in sheep but down-regulated in goat. Bottom right quadrant: genes up-regulated in goat, but down-regulated in sheep.

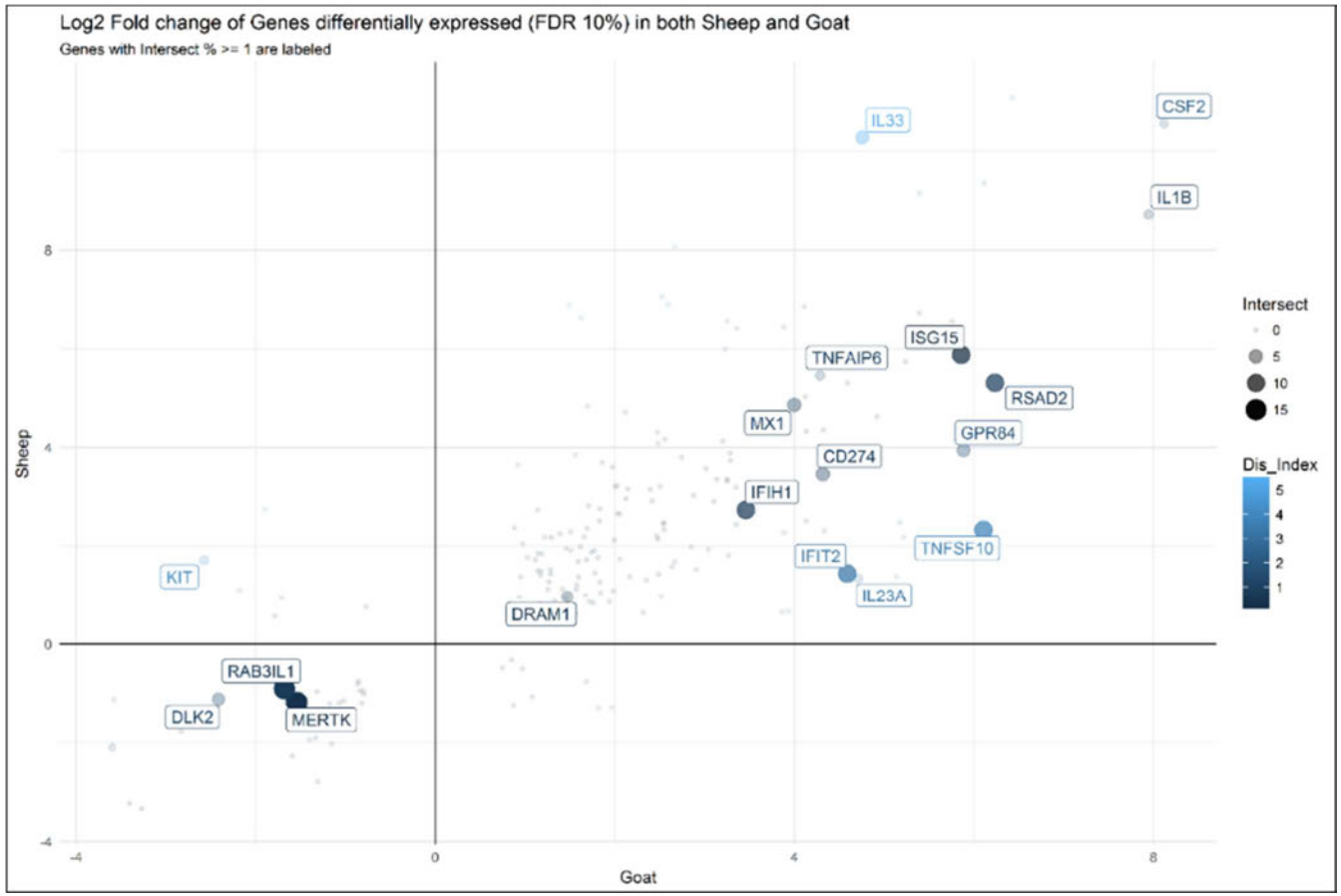


Figure 4-5:Analysis of differentially expressed genes (FDR<10%, $|\log_2 FC| \geq 2$) in goat and sheep BMDM showing high intersect. The genes which showed the highest level of intersect (≥1%), between goats and sheep) are shown. Level of intersect is represented by the size of the node. *IL33, TNFSF10 and IL23A* show the highest level of intersect between the goat and sheep.

4.3.5 Species-specific immune differences in alveolar macrophage response

The full list of all differentially expressed genes between goat and sheep alveolar macrophages is included in Appendix Y. The differentially expressed genes with the top 25 high and 25 low |log2FC| between goat and sheep are shown in Fig 4.4. Goat alveolar macrophages showed elevated expression relative to sheep of several immune genes including *IL33, IL1B, IL34* and *IL1RN,* c-type lectin *CLEC5A,* and colony stimulating factor *CSF1.* Conversely, the set of genes down regulated in sheep relative to goat were not obviously immune-related and included *CMYA5*, collagen *COL8A1* and solute carrier *SLC22A3*.

4.3.6 Comparative Gene Annotation

Some of the unannotated genes in goat were already annotated in sheep (and to a lesser extent, vice versa). To improve the annotation of genes within clusters that shared transcriptional similarity across species, the longest peptide for each unannotated gene in each cluster was used to calculate percentage identity with the EMBOSS Needle (Rice, Longden et al. 2000) which runs global end-to-end alignments for each gene. On this basis, provisional annotations to previously unannotated genes in both goat and sheep were added and are summarised in Table 4.5. The transcript encoding for the *SERPINB2* gene is duplicated in sheep macrophages (personal communication, M. McCulloch), shown in Table 4-5 as transcripts ENSOARG00000006889 and ENSOARG00000005159, which are both strongly induced by LPS. The two sheep *SERPINB2* transcripts show a high similarity with a single goat transcript LOC102181552 at 98.3% and 88.7% respectively indicating that the duplicated *SERPINB2* gene is specific to sheep. The goat *SERPINB2* gene was induced by LPS but was expressed at a much lower level than either of the transcripts that are expressed in sheep macrophages. The sheep genome also contains two transcripts (ENSOARG0000014496 and ENSOARG0000016940) on separate chromosomes that are each annotated on ENSEMBL as *CXCL8*, encoding the major neutrophil chemokine interleukin 8. Both were LPS-inducible in sheep

macrophages. The goat genome contains only one *CXCL8* gene, which is syntenic and identical at the protein sequence level to ENSOARG0000014496.

Table 4-5:Comparative gene annotation

Cluster	Sheep	Goat	% Identity
Muscle	ENSOARG00000001361	*JSRP1*	80.8
	ENSOARG00000006444	*MYH13*	98
	ENSOARG00000017195	*TTN*	98.2
	ENSOARG00000010028	LOC102177638	98.8
	ENSOARG00000012656(1932aa)	LOC102181426(1939aa)	95
	ENSOARG00000012656(1932aa)	*MYH8*(1937aa)	92.3
Fallopian Tube	ENSOARG00000013919	*NEK10* (*NIMA* related kinase 10)	93.4
	ENSOARG00000020208	*ROPN1* (rhophilin associated tail protein 1)	99.5
	ENSOARG00000016978	*FAM47E* (family with sequence similarity 47 member E)	83.3
	ENSOARG00000020032	LOC102178119	100
Cell-cycle	ENSOARG00000005764	*STMN1*	100
	STMN1=ENSOARG00000012293		
	ENSOARG00000019455	*GTSE1* (G2 and S-phase expressed 1)	92.3

LPS	ENSOARG00000014496 (101 aa)	*CXCL8* (101 aa)	100
	ENSOARG00000004253(109 aa)	LOC102180880(92 aa)	83.5
	ENSOARG00000004367 (93 aa)	LOC102181154 (93 aa)	97.8
	ENSOARG00000005159(SERPINB2-like) (416 aa)	LOC102181552 (416 aa)	88.7
	ENSOARG00000006889(SERPINB2) (416 aa)	LOC102181552 (416 aa)	98.3
	ENSOARG00000016940 (102 aa)	*CXCL8* (101 aa)	81.4

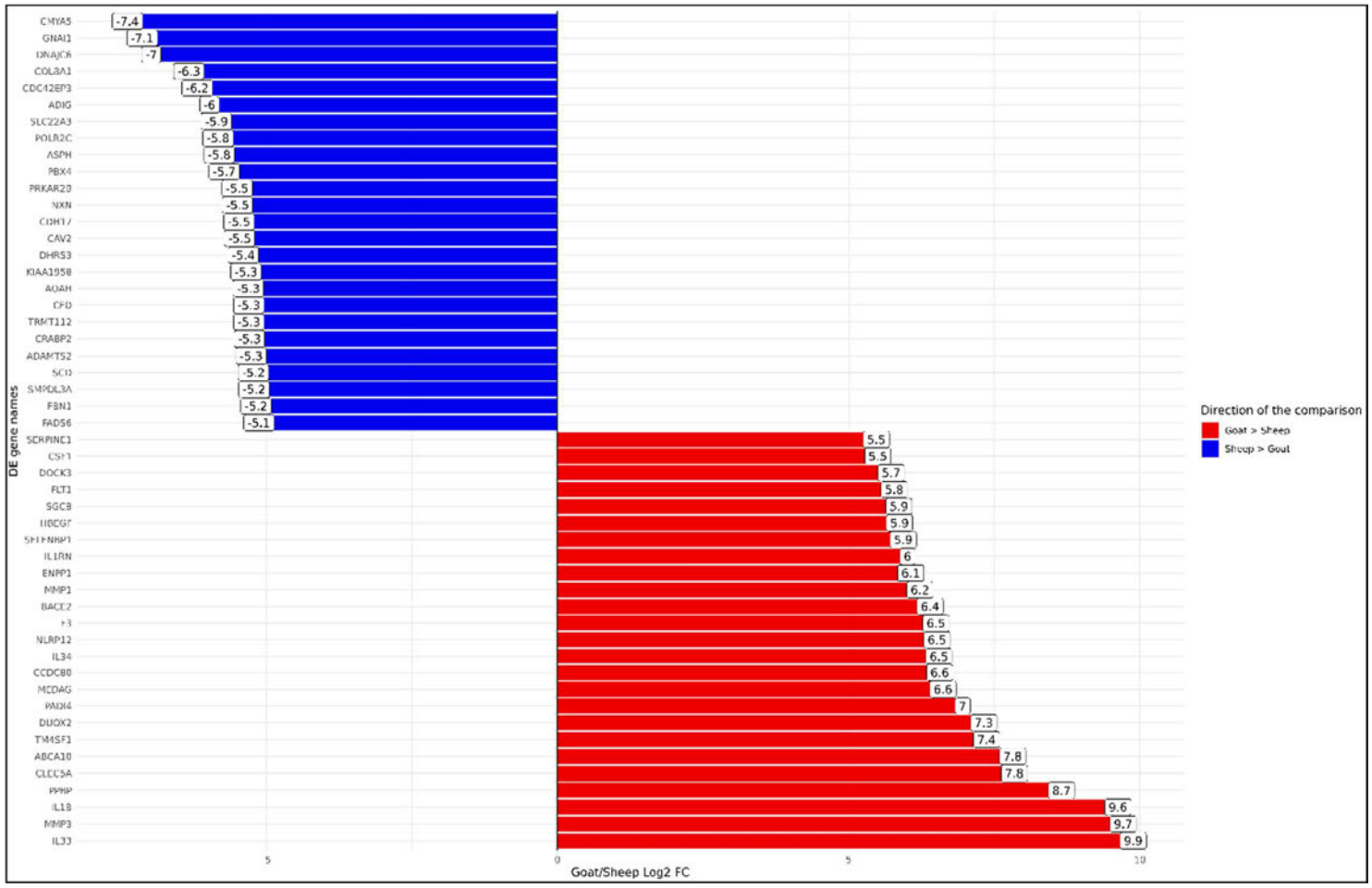


Figure 4-6:Differentially expressed genes (FDR<10%) between goat and sheep alveolar macrophages. The top 25 up-regulated in goat relative to sheep (red) and the top 25 down-regulated in goat relative to sheep (blue) are shown

4.4 Discussion

4.4.1 General overview

The generation of equivalent gene expression atlases of the domestic goat and sheep has enabled a comparative transcriptomic analysis between two closely related species to investigate species-specific differences in gene expression. By selecting sex-matched individuals, the effect of gender on transcriptional control was minimised. However, differences in the age of individuals used in this analysis, adult sheep and neonatal goats, may have had a significant effect on the observed gene expression patterns. To remedy this, we validated the expression of a sub-set of immune genes using RT-qPCR and found that expression patterns of a small subset of immune genes were similar in age-matched sheep and goats. Another caveat is that at the time of the global comparison of transcriptional profiles across tissues in sheep and goat, the annotation of the goat reference assembly (ARS1) was not available on Ensembl and the quality of the annotation was limited compared to the sheep assembly and as such, some of the differences observed may relate to differences in the quality of the goat annotation.

4.4.2 Divergent signalling pathways in response to LPS stimulation

As previously discussed, activation of macrophages by LPS requires the sequential engagement of *CD14* and *TLR4/MD-2* signalling. The active receptor recruits two sets of adaptor proteins to initiate two distinct pathways. The *MyD88*-dependent and *MyD88*-independent (TRIF/TRAM-dependent) pathways target distinct sets of inducible genes (Akira, Uematsu et al. 2006). As shown in Table 4-3 goat macrophages expressed much lower levels of *CD14,* the *TLR4* co-receptor that regulates LPS sensitivity, and somewhat lower levels of both *TLR4* and *MD2*. *CD14*-deficient mice do not respond to LPS (Haziot, Ferrero et al. 1996) while soluble *CD14* increases the sensitivity of mammary epithelial cells to mastitis infections (Yan Wang, Dante S. Zarlenga et al. 2002). While sheep BMDM seemed to downregulate the expression of *CD14* slightly after 7hr stimulation with LPS (|log2FC| of -0.22), goat BMDM upregulated *CD14* expression after 7hr LPS stimulation |log2FC|

of 1.18). Previously, human monocytes were shown to upregulate expression of *CD14* after stimulation with LPS which enabled them to maintain cell-responsiveness (Landmann, Knopf et al. 1996). A more recent study in cattle reported that differentiation of bovine *CD14*-positive monocytes in the presence of the chemokine *CCL5* generated macrophages with reduced *CD14* expression rendering them unresponsive to LPS (Hussen and Schuberth 2017). Both *CD14* and *TLR4* were inducible by LPS in goat macrophages. Assuming *CD14* and *TLR4* mRNA levels reflect expression of the protein, goat BMDM appear to be less sensitive to initial contact with LPS than sheep BMDM but their response mechanism compensates for this by maintaining activation through inducible receptor expression.

The expression of many of the components of the *MyD88*-dependent signalling pathway including *MyD88, IRAK1* and *IRAK4* was lower in goat macrophages compared to sheep. Conversely, *TRAM1*, *IRF3* and *TBK1* specifically involved in the *MyD88*-independent signalling pathway (Yamamoto, Sato et al. 2003) were expressed in much higher levels in goat macrophages than sheep. Taken together, these results suggest that goat BMDM may have a bias towards the *MyD88*-independent pathway of macrophage activation. The *MyD88*-independent pathway through the activation of interferon regulatory factor 3 (*IRF3)* initiates the expression of *IFNβ1* and an autocrine induction of target genes of type 1 interferon. A detailed temporal analysis of this cascade in human macrophages was presented by (Baillie, Arner et al. 2017).

4.4.3 Species-specific differences in immune responses

4.4.3.1 BMDM immune response to LPS stimulation

As previously stated, *IL33* was the most dissimilar gene in expression pattern between goat and sheep BMDM with a |log2FC| of 4.7 in goat compared to 10.28 in sheep. Previous studies have reported that *IL33* plays a protective role in inflammatory bowel disease (IBD) by inducing a Th2 immune response (Lopetuso, Chowdhry et al. 2013). As discussed in section 4.1 of this chapter, an enhanced Th2 response has been associated with *Haemonchus contortus* resistance in sheep (Alba-Hurtado and Munoz-Guzman 2013) by accelerating

parasite expulsion. *IL36B* also had dissimilar gene in expression pattern between goat and sheep BMDM. Little is known about the role of *IL36* in goat biology, but its presence in the IL-1 family of genes (together with *IL33)* is suggestive that it is involved in gut health and homeostasis.

IFIT2 is an interferon induced antiviral protein and has been associated with differential response to bluetongue virus (BTV) (Singh, Prasad et al. 2017), and the enhanced IFN response in goats is in line with our previous observation that goat macrophages may have a bias towards the *MyD88*-independent signalling pathway. Similarly, high *IL23A* levels have been associated with inflammatory infiltration during infection with MAP, which inhibits bacterial proliferation in sheep. Conversely, in nematode infections, higher expression of *IL23A* is associated with lambs susceptible to *T. circumcincta (*Gossner, Venturina et al. 2012*)*. There is not much comparable literature for goats but given the significant similarities in the gene expression profiles observed between sheep and goats, it is reasonable to accept that these genes are regulated in a similar manner to sheep and involved in similar biological pathways. Collectively, these results suggest species-specific differences in the expression of immune genes could underlie variation in susceptibility to nematode and MAP infections between goats and sheep as hypothesised in previous studies (Bishop and Stear 2003, Bishop and Morris 2007).

4.4.3.2 Primed alveolar macrophage response evident in neonatal goats

Alveolar macrophages are the first line of defence against air-borne pathogens. There was a stronger immune response in the neonatal goat alveolar macrophages in comparison to adult sheep alveolar macrophages indicated by elevated expression of such genes as *IL33, IL1B, IL34* and *IL1RN* as well as c-type lectin *CLEC5A,* and colony stimulating factor *CSF1.* This may reflect that the alveolar macrophage immune response is age-dependent, showing a primed response in younger individuals whose adaptive immunity is still primitive, but showing a much more subdued response in older individuals whose adaptive immunity has reached full development. Further

work could examine expression patterns of these genes in alveolar macrophages in age-matched sheep and goats, using RT-qPCR, to determine whether this is indeed the case.

4.5 Conclusion

This chapter describes comparative analysis of transcription profiles in domestic goat and sheep. There was a strong overlap of clusters of genes associated with general biological functions such as cell cycle and skeletal muscle in goat and sheep indicating that broadly the datasets were highly comparable and transcriptional patterns were broadly similar between the two species. Species-specific differences were most obvious in immune tissues and cells. For example, assuming *CD14* and *TLR4* mRNA levels reflect expression of the protein, goat BMDM are seemingly less sensitive to initial contact with LPS than sheep BMDM, but their response mechanism compensates for this by maintaining activation through inducible receptor expression. As demonstrated, only a small subset of *IFN*-induced genes was inducible, but this analysis is based only on the 7hr time-point after LPS stimulation and many of the *IFN*-inducible genes are known to be induced relatively late in response to LPS. Given the observed similarities in gene expression profiles between sheep and goats after LPS stimulation, it is reasonable to accept that the majority of immune genes are regulated in a similar manner in both species and are involved in similar biological pathways. In this chapter comparative analysis, using network cluster analysis was useful for functional annotation of unannotated genes in one species relative to another, improving the genomic resources available for both.

Chapter 5 Summary and Future Directions

5.1 General Discussion

The aim of chapter 2 of my thesis was to generate a mini-atlas of gene expression for goat. From seventeen tissues and 2 cell types I was able to detect 90% of the transcriptome, providing proof of concept that this mini-atlas approach is useful for studying gene expression and for functional annotation. Using the mini-atlas dataset we annotated 15% (representing more than 1000 genes) of the previously unannotated genes in ARS1 (Bickhart, Rosen et al. 2017). The dataset I generated was used by the Ensembl team (Hubbard 2002, Zerbino, Achuthan et al. 2018, Yates, Achuthan et al. 2020) to create a new gene build for the goat ARS1 reference genome (https://www.ensembl.org/Capra_hircus/Info/Index) (Bickhart, Rosen et al. 2017, Worley 2017). This will provide a valuable resource for the livestock genomics community to complement the available genomic tools (Tosser-Klopp, Bardou et al. 2014, Stella, Nicolazzi et al. 2018, Talenti, Palhière et al. 2018). Future work could focus on improving the goat genome further by generating datasets to annotate the regulatory regions of the genome using ATAC-Seq (Corces, Trevino et al. 2017). The annotation of ARS1 could also be improved by adding Iso-Seq (full-length isoform sequencing) data to accurately construct complete transcript models (Gonzalez-Garay 2016). Even undertaking these assays on a small number of tissues would provide a considerable amount of functional annotation information to improve the current goat genome.

One potential area of analysis of the transcriptomic data I generated for goat, that I was not able to explore within the scope of my PhD, was allele-specific expression (ASE), however, this analysis was eventually carried out and included in the main publication of this thesis (Muriuki, Bush et al. 2019) Similar studies have examined ASE across tissues in cattle (Chamberlain, Vander Jagt et al. 2015), sheep (Salavati, Bush et al. 2019) and goat (Cao, Xu et al. 2019). Using the goat mini-atlas dataset it would be possible to analyse

ASE at the gene level across tissues and cell types. It would also be possible to measure ASE at the variant (SNV) level. This would identify variants exhibiting ASE and determine whether they were located within or near important genes for production traits. These variants could then be weighted in genomic prediction models. The sequencing depth used for the goat mini-atlas is, however, insufficient for statistically robust analysis at the SNV level. It does however provide a foundation for further analysis of ASE using a suitable dataset, from more individuals (e.g. for aseQTL analysis (Wang, Hancock et al. 2018)) and at a greater depth.

5.2 The innate immune response in goats and sheep

In chapter 3 of my thesis I provided the first comprehensive characterisation of BMDM from goat providing a methodology for culturing and characterisation and a detailed analysis of transcription post stimulation with LPS. Prior to this study little was known about the transcription in goat macrophages. While more information is available on goat monocyte derived macrophages (Adeyemo, Gao Rj Fau - Lan et al. 1997, Taka, Liandris et al. 2013, Walia, Kumar et al. 2015), there was previously relatively little knowledge available on the characteristics of goat BMDM. In addition, few reagents are available for immunological studies in goat, with most studies relying on cross-reactivity with sheep and cattle antibodies (Entrican 2002, Hope, Sopp et al. 2012). Recently a characterization of goat antibody loci has been published using the new reference genome ARS1 (Schwartz, Philp et al. 2018), demonstrating the usefulness of a highly contiguous reference genome with high quality functional annotation for the development of new resources for livestock species.

The analysis for chapter 4 built on the results from chapter 3 to examine the transcriptional basis for differences in the immune response between the two species. Overall the results of my analysis indicated that BMDM from goat and

sheep do not vary hugely in their transcriptional response to LPS. Several immune genes were upregulated in both goat and sheep BMDM in response to LPS stimulation but differed in their level of induction between the two species. For example, immune genes including *IL33, IL36B, PTX3, CCL20, CSF3* and *CSF2* exhibited higher levels of induction in sheep BMDM relative to goat, and vice versa for *ICAM1, IL23A, IFIT2, TNFSF10,* and *TNFRSF9*.

There were two main caveats to my analysis. The first was that the sheep and goats were not age-matched (the sheep were a year old while the goats were one week old) and though they had been raised in similar environments rearing conditions were not identical. Though the logistics of collecting samples from large animals can be difficult this caveat could be remedied in future studies by collecting bone marrow from age-matched sheep and goats which had been reared in identical environments and comparing transcriptional patterns. The second caveat was that there was a high level of individual variation between goats. This pattern was also observed in other species e.g. (Clark, Bush et al. 2017, Young, Bush et al. 2018). A similar larger study in pig monocytes showed similar variation between individuals but found no evidence for breed-specific differences (Fairbairn, Kapetanovic et al. 2013). A larger number of biological replicates could help to control for individual-specific effects, but as mentioned previously sourcing and collecting sufficient samples from livestock can be difficult. Expression QTL analysis in humans revealed >80% of LPS inducible genes in monocytes exhibited heritable variation in the level of expression (Fairfax, Humburg et al. 2014). An eQTL analysis in 60-100 goats would help to determine the heritable variation in the level of expression and help to narrow down the search space for the variants underlying variation in the immune response. These variants could then be prioritised in breeding programmes for goats. This kind of an experiment would probably need to be based on samples from blood or milk, for the logistical reasons mentioned above. Similar eQTL studies have been performed in 300 cattle from whole blood, milk and somatic cell counts (SSC) (Wang, Hancock et al. 2018).

In general, expanding the analysis I present here to include additional cell types would be interesting. The results of the comparative analysis between sheep and goats provide a foundation for further comparative analysis in both BMDM and additional immune cell types (e.g. monocytes or sorted populations of cells from blood). Similarly, here I measure the response of BMDM to LPS, and further work could measure other components of the immune response, for example, using stimulation with interferon gamma e.g. (Shaw, Hughes et al. 2017). Now I have established a culture system for goat BMDM it would also be possible to investigate transcription after viral challenge with capripox or caprine arthritis encephalitis e.g. (Adeyemo, Gao Rj Fau - Lan et al. 1997, Larruskain and Jugo 2013). The pathogen burden will be very different across the globe, in tropical environments for example, there will be populations of goats that are locally adapted to pathogen or climatic burdens. As such further work could assess the transcriptional response of immune cells from locally adapted goats to *Haemonchus contortus* e.g. (Gill, Altmann et al. 2000) and/or other pathogens that are important in the tropics.

5.3 A gene expression atlas for African goats

Goats are an important livestock species across the globe, but they are of particular importance in tropical agri-systems contributing to sustainable agriculture, alleviation of poverty, social cohesion and utilisation of marginal grazing. In tropical agri-systems they are subject to a broader range of pathogens and environmental stressors than in temperate regions. The goat mini-atlas was generated using samples from crossbred dairy goats from the UK. A logical next step would be to generate a mini-atlas of gene expression for a tropical breed of goat. Transcriptional patterns in tissues that are relevant to metabolism and immunity are likely to be different in tropical breeds. This approach would provide a valuable resource for understanding adaptation to a tropical environment and identifying key loci underlying disease and resilience traits. There is considerable potential for genome engineering in goats using CRISPR/Cas9 and other technologies (Yu, Lu et al. 2016, Kalds,

Zhou et al. 2019), which could be applied to exploit loci underlying resilience to tropical environments.

One example of a breed that would be suitable for generating an atlas of gene expression for a tropical goat would be the Galla goat, which is a multi-purpose breed of goat indigenous to Northern Kenya (Kenya 2019). Galla goats are popular with both small and large-scale farmers, and have been crossbred successfully with European breeds such as the Toggenburg to improve milk yield (Kenya 2019). It would be possible to buy Galla goats and house them at the International Livestock Research Institute (ILRI), Nairobi. As for this project I would collect the same set of tissue samples from neonatal animals to minimise the exposure to natural infection in the field. I would also isolate BMDM and other immune cells including AM and sort populations of cells from blood to characterise immune mediated transcription in a tropical goat breed. This would improve our understanding gene expression in healthy animals and would provide a foundation for investigating transcription during disease infection or under heat or drought stress. Other studies have successfully studied gene expression profiles in immune cells after infection with diseases that are relevant in tropical agri-systems, including Peste-Des-Pestis-Ruminants virus (PPRV) (Wani, Sahu et al. 2019), in live animals. Establishing a culture system for BMDM and other immune cells allows measurement of transcriptional responses in vitro as an alternative or precursor to challenge experiments in vivo.

5.4 Investigating goats with a different genetic background

Goat populations in Africa are genetically different to the managed herds of dairy goats in the UK and Europe. This was reflected in the results of the ADAPT Map project, which genotyped more than 2000 goats from across the globe, including several populations from North, West, East and South East Africa (Colli, Milanesi et al. 2018, Stella, Nicolazzi et al. 2018). This project further characterized the genetic diversity of African goat breeds and identified

signatures of selection related to adaptation to a tropical environment. My research group has sequenced the genomes of 269 goats from twenty different countries for the 1000 Goat Genomes Project (VarGoats 2019). This information will be used to inform the development of new genetic tools for goats, including an update on the current 50K chip. This update will include additional content from African animals and improve the utility of the chip in African agri-systems. Another future direction for this project would be to overlay the expression data from the mini-atlas of expression with the whole genome sequencing dataset from African goats. The intersection of genomic DNA with functional annotation information could provide insights into the molecular basis of population-specific differences and adaptation to a tropical environment. Functional variation particularly in immune genes, for example, between populations, from different ecosystems with varying pathogen or climatic burdens, are likely to underpin differences in susceptibility to changing climatic conditions or disease. This variation can then be utilised for breeding the hardiest individuals while concomitantly improving production traits in community-based breeding programmes.

5.5 Conclusion

The work described in my thesis has made a significant contribution to the understanding of gene expression in the domestic goat and in comparative innate immune responses in small ruminants. The data generated is in the public domain and was used to annotate the current reference genome (ARS1) providing a valuable resource that can be used by the livestock genomics community. It provides a foundation for further studies that will elucidate the genetic basis of the immune response in tropical and temperate goat breeds. The goat is a hugely important livestock species, which contributes significantly to the stability and sustainability of agri-systems across the globe.

References

Adeyemo, O., H. C. Gao Rj Fau - Lan and H. C. Lan (1997). "Cytokine production in vitro by macrophages of goats with caprine arthritis-encephalitis." Cell Mol Biol(0145-5680 (Print)).

Akira, S., S. Uematsu and O. Takeuchi (2006). "Pathogen recognition and innate immunity." Cell **124**(4): 783-801.

Alba-Hurtado, F. and M. A. Munoz-Guzman (2013). "Immune responses associated with resistance to haemonchosis in sheep." Biomed Res Int **2013**: 162158.

Alexa, A. and J. Rahnenfuhrer (2016). topGO: Enrichment Analysis for Gene Ontology. R package version 2.30.1. http://bioconductor.uib.no/2.7/bioc/html/topGO.html.

Andersson, L., A. L. Archibald, C. D. Bottema, R. Brauning, S. C. Burgess, D. W. Burt, E. Casas, H. H. Cheng, L. Clarke, C. Couldrey, B. P. Dalrymple, C. G. Elsik, S. Foissac, E. Giuffra, M. A. Groenen, B. J. Hayes, L. S. Huang, H. Khatib, J. W. Kijas, H. Kim, J. K. Lunney, F. M. McCarthy, J. C. McEwan, S. Moore, B. Nanduri, C. Notredame, Y. Palti, G. S. Plastow, J. M. Reecy, G. A. Rohrer, E. Sarropoulou, C. J. Schmidt, J. Silverstein, R. L. Tellam, M. Tixier-Boichard, G. Tosser-Klopp, C. K. Tuggle, J. Vilkki, S. N. White, S. Zhao and H. Zhou (2015). "Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project." Genome Biol **16**: 57.

Andersson, R., C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jorgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhata, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Muller, A. R. Forrest, P. Carninci, M. Rehli and A. Sandelin

(2014). "An atlas of active enhancers across human cell types and tissues." Nature **507**(7493): 455-461.

Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene Ontology: tool for the unification of biology.pdf." Nat Genet **25**.

Baillie, J. K., A. Bretherick, C. S. Haley, S. Clohisey, A. Gray, L. P. A. Neyton, J. Barrett, E. A. Stahl, A. Tenesa, R. Andersson, J. B. Brown, G. J. Faulkner, M. Lizio, U. Schaefer, C. Daub, M. Itoh, N. Kondo, T. Lassmann, J. Kawai, I. Consortium, D. Mole, V. B. Bajic, P. Heutink, M. Rehli, H. Kawaji, A. Sandelin, H. Suzuki, J. Satsangi, C. A. Wells, N. Hacohen, T. C. Freeman, Y. Hayashizaki, P. Carninci, A. R. R. Forrest and D. A. Hume (2018). "Shared activity patterns arising at genetic susceptibility loci reveal underlying genomic and cellular architecture of human disease." PLoS Comput Biol **14**(3): e1005934.

Baillie, J. K., E. Arner, C. Daub, M. De Hoon, M. Itoh, H. Kawaji, T. Lassmann, P. Carninci, A. R. Forrest, Y. Hayashizaki, F. Consortium, G. J. Faulkner, C. A. Wells, M. Rehli, P. Pavli, K. M. Summers and D. A. Hume (2017). "Analysis of the human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides new insights into genetic aetiology of inflammatory bowel disease." PLoS Genet **13**(3): e1006641

Banos, G., G. Bramis, S. J. Bush, E. L. Clark, M. E. B. McCulloch, J. Smith, G. Schulze, G. Arsenos, D. A. Hume and A. Psifidi (2017). "The genomic architecture of mastitis resistance in dairy sheep." BMC Genomics **18**(1): 624.

Basset, C., J. Holton, R. O'Mahony and I. Roitt (2003). "Innate immunity and pathogen–host interaction." Vaccine **21**: S12-S23.

Beutler, B., Z. Jiang, P. Georgel, K. Crozat, B. Croker, S. Rutschmann, X. Du and K. Hoebe (2006). "Genetic analysis of host resistance: Toll-like receptor signaling and immunity at large." Annu Rev Immunol **24**: 353-389.

Beutler, B. (2004). "Innate immunity: an overview." Molecular Immunology **40**(12): 845-859.

Derek M Bickhart, Benjamin D Rosen, Sergey Koren, Brian L Sayre, Alex R Hastie, Saki Chan, Joyce Lee, Ernest T Lam, Ivan Liachko, Shawn T Sullivan, Joshua N Burton, Heather J Huson, John C Nystrom, Christy M Kelley, Jana L Hutchison, Yang Zhou, Jiajie Sun, Alessandra Crisà, F Abel Ponce de León, John C Schwartz, John A Hammond, Geoffrey C Waldbieser, Steven G Schroeder, George E Liu, Maitreya J Dunham, Jay Shendure, Tad S Sonstegard, Adam M Phillippy, Curtis P Van Tassell and T. P. L. Smith (2017). "Single-molecule sequencing and conformational capture enable de novo mammalian reference genomes." Nature Genetics.

Bickhart, D. M., B. D. Rosen, S. Koren, B. L. Sayre, A. R. Hastie, S. Chan, J. Lee, E. T. Lam, I. Liachko, S. T. Sullivan, J. N. Burton, H. J. Huson, J. C. Nystrom, C. M. Kelley, J. L. Hutchison, Y. Zhou, J. Sun, A. Crisa, F. A. Ponce de Leon, J. C. Schwartz, J. A. Hammond, G. C. Waldbieser, S. G. Schroeder, G. E. Liu, M. J. Dunham, J. Shendure, T. S. Sonstegard, A. M. Phillippy, C. P. Van Tassell and T. P. L. Smith (2017). "Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome." Nat Genet **49**(4): 643-650.

Birney, E., J. A. Stamatoyannopoulos, A. Dutta, R. Guigo, T. R. Gingeras, E. H. Margulies, Z. Weng, M. Snyder, E. T. Dermitzakis, R. E. Thurman, M. S. Kuehn, C. M. Taylor, S. Neph, C. M. Koch, S. Asthana, A. Malhotra, I. Adzhubei, J. A. Greenbaum, R. M. Andrews, P. Flicek, P. J. Boyle, H. Cao, N. P. Carter, G. K. Clelland, S. Davis, N. Day, P. Dhami, S. C. Dillon, M. O. Dorschner, H. Fiegler, P. G. Giresi, J. Goldy, M. Hawrylycz, A. Haydock, R. Humbert, K. D. James, B. E. Johnson, E. M. Johnson, T. T. Frum, E. R. Rosenzweig, N. Karnani, K. Lee, G. C. Lefebvre, P. A. Navas, F. Neri, S. C. Parker, P. J. Sabo, R. Sandstrom, A. Shafer, D. Vetrie, M. Weaver, S. Wilcox, M. Yu, F. S. Collins, J. Dekker, J. D. Lieb, T. D. Tullius, G. E. Crawford, S. Sunyaev, W. S. Noble, I. Dunham, F. Denoeud, A. Reymond, P. Kapranov, J. Rozowsky, D. Zheng, R. Castelo, A. Frankish, J. Harrow, S. Ghosh, A.

Sandelin, I. L. Hofacker, R. Baertsch, D. Keefe, S. Dike, J. Cheng, H. A. Hirsch, E. A. Sekinger, J. Lagarde, J. F. Abril, A. Shahab, C. Flamm, C. Fried, J. Hackermuller, J. Hertel, M. Lindemeyer, K. Missal, A. Tanzer, S. Washietl, J. Korbel, O. Emanuelsson, J. S. Pedersen, N. Holroyd, R. Taylor, D. Swarbreck, N. Matthews, M. C. Dickson, D. J. Thomas, M. T. Weirauch, J. Gilbert, J. Drenkow, I. Bell, X. Zhao, K. G. Srinivasan, W. K. Sung, H. S. Ooi, K. P. Chiu, S. Foissac, T. Alioto, M. Brent, L. Pachter, M. L. Tress, A. Valencia, S. W. Choo, C. Y. Choo, C. Ucla, C. Manzano, C. Wyss, E. Cheung, T. G. Clark, J. B. Brown, M. Ganesh, S. Patel, H. Tammana, J. Chrast, C. N. Henrichsen, C. Kai, J. Kawai, U. Nagalakshmi, J. Wu, Z. Lian, J. Lian, P. Newburger, X. Zhang, P. Bickel, J. S. Mattick, P. Carninci, Y. Hayashizaki, S. Weissman, T. Hubbard, R. M. Myers, J. Rogers, P. F. Stadler, T. M. Lowe, C. L. Wei, Y. Ruan, K. Struhl, M. Gerstein, S. E. Antonarakis, Y. Fu, E. D. Green, U. Karaoz, A. Siepel, J. Taylor, L. A. Liefer, K. A. Wetterstrand, P. J. Good, E. A. Feingold, M. S. Guyer, G. M. Cooper, G. Asimenos, C. N. Dewey, M. Hou, S. Nikolaev, J. I. Montoya-Burgos, A. Loytynoja, S. Whelan, F. Pardi, T. Massingham, H. Huang, N. R. Zhang, I. Holmes, J. C. Mullikin, A. Ureta-Vidal, B. Paten, M. Seringhaus, D. Church, K. Rosenbloom, W. J. Kent, E. A. Stone, N. C. S. Program, C. Baylor College of Medicine Human Genome Sequencing, C. Washington University Genome Sequencing, I. Broad, I. Children's Hospital Oakland Research, S. Batzoglou, N. Goldman, R. C. Hardison, D. Haussler, W. Miller, A. Sidow, N. D. Trinklein, Z. D. Zhang, L. Barrera, R. Stuart, D. C. King, A. Ameur, S. Enroth, M. C. Bieda, J. Kim, A. A. Bhinge, N. Jiang, J. Liu, F. Yao, V. B. Vega, C. W. Lee, P. Ng, A. Shahab, A. Yang, Z. Moqtaderi, Z. Zhu, X. Xu, S. Squazzo, M. J. Oberley, D. Inman, M. A. Singer, T. A. Richmond, K. J. Munn, A. Rada-Iglesias, O. Wallerman, J. Komorowski, J. C. Fowler, P. Couttet, A. W. Bruce, O. M. Dovey, P. D. Ellis, C. F. Langford, D. A. Nix, G. Euskirchen, S. Hartman, A. E. Urban, P. Kraus, S. Van Calcar, N. Heintzman, T. H. Kim, K. Wang, C. Qu, G. Hon, R. Luna, C. K. Glass, M. G. Rosenfeld, S. F. Aldred, S. J. Cooper, A. Halees, J. M. Lin, H. P. Shulha, X. Zhang, M. Xu, J. N. Haidar, Y. Yu, Y. Ruan, V. R. Iyer, R. D. Green, C. Wadelius, P. J. Farnham, B. Ren, R. A. Harte,

A. S. Hinrichs, H. Trumbower, H. Clawson, J. Hillman-Jackson, A. S. Zweig, K. Smith, A. Thakkapallayil, G. Barber, R. M. Kuhn, D. Karolchik, L. Armengol, C. P. Bird, P. I. de Bakker, A. D. Kern, N. Lopez-Bigas, J. D. Martin, B. E. Stranger, A. Woodroffe, E. Davydov, A. Dimas, E. Eyras, I. B. Hallgrimsdottir, J. Huppert, M. C. Zody, G. R. Abecasis, X. Estivill, G. G. Bouffard, X. Guan, N. F. Hansen, J. R. Idol, V. V. Maduro, B. Maskeri, J. C. McDowell, M. Park, P. J. Thomas, A. C. Young, R. W. Blakesley, D. M. Muzny, E. Sodergren, D. A. Wheeler, K. C. Worley, H. Jiang, G. M. Weinstock, R. A. Gibbs, T. Graves, R. Fulton, E. R. Mardis, R. K. Wilson, M. Clamp, J. Cuff, S. Gnerre, D. B. Jaffe, J. L. Chang, K. Lindblad-Toh, E. S. Lander, M. Koriabine, M. Nefedov, K. Osoegawa, Y. Yoshinaga, B. Zhu and P. J. de Jong (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.

Bishop, S. C. and C. A. Morris (2007). "Genetics of disease resistance in sheep and goats." Small Ruminant Research **70**(1): 48-59.

Bishop, S. C. and J. A. Woolliams (2014). "Genomics and disease resistance studies in livestock." Livest Sci **166**: 190-198.

Bishop, S. C. and M. J. Stear (2003). "Modeling of host genetics and resistance to infectious diseases: understanding and controlling nematode infections." Veterinary Parasitology **115**(2): 147-166.

Braissant O, F. F., Scotto C, Dauça M, Wahli W. (1996). "Differential expression of peroxisome proliferator-activated receptors (PPARs)-tissue distribution of PPAR-alpha, -beta, and -gamma in the adult rat.pdf." Endocrinology. **137**(1): 354-366.

Bray, N. L., H. Pimentel, P. Melsted and L. Pachter (2016). "Near-optimal probabilistic RNA-seq quantification." Nat Biotechnol **34**(5): 525-527.

Brenaut, P., L. Lefevre, A. Rau, D. Laloe, G. Pisoni, P. Moroni, C. Bevilacqua and P. Martin (2014). "Contribution of mammary epithelial cells to the immune response during early stages of a bacterial infection to Staphylococcus aureus.pdf." VETERINARY RESEARCH **45**(16).

Bush, S. J., L. Freem, A. J. MacCallum, J. O'Dell, C. Wu, C. Afrasiabi, A. Psifidi, M. P. Stevens, J. Smith, K. M. Summers and D. A. Hume (2018). "Combination of novel and public RNA-seq datasets to generate an mRNA expression atlas for the domestic chicken." BMC genomics **19**(1): 594-594.

Bush, S. J., L. Freem, A. J. MacCallum, J. O'Dell, C. Wu, C. Afrasiabi, A. Psifidi, M. P. Stevens, J. Smith, K. M. Summers and D. A. Hume (2018). "Combination of novel and public RNA-seq datasets to generate an mRNA expression atlas for the domestic chicken." BMC Genomics **19**(1): 594.

Bush, S. J., M. E. B. McCulloch, C. Muriuki, M. Salavati, G. M. Davis, I. L. Farquhar, Z. M. Lisowski, A. L. Archibald, D. A. Hume and E. L. Clark (2019). "Comprehensive Transcriptional Profiling of the Gastrointestinal Tract of Ruminants from Birth to Adulthood Reveals Strong Developmental Stage Specific Gene Expression." G3: Genes|Genomes|Genetics **9**(2): 359.

Bustin, S. A., V. Benes, J. A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M. W. Pfaffl, G. L. Shipley, J. Vandesompele and C. T. Wittwer (2009). "The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments." Clin Chem **55**(4): 611-622.

Cao, Y., H. Xu, R. Li, S. Gao, N. Chen, J. Luo and Y. Jiang (2019). "Genetic Basis of Phenotypic Differences Between Chinese Yunling Black Goats and Nubian Goats Revealed by Allele-Specific Expression in Their F1 Hybrids." Frontiers in Genetics **10**: 145.

Casey, M. E., K. G. Meade, N. C. Nalpas, M. Taraktsoglou, J. A. Browne, K. E. Killick, S. D. Park, E. Gormley, K. Hokamp, D. A. Magee and D. E. MacHugh (2015). "Analysis of the Bovine Monocyte-Derived Macrophage Response to Mycobacterium avium Subspecies Paratuberculosis Infection Using RNA-seq." Front Immunol **6**: 23.

Cecchi, F., C. Russo, D. Iamartino, A. Galiero, B. Turchi, F. Fratini, S. Degl'Innocenti, R. Mazza, S. Biffani, G. Preziuso and C. Cantile (2017). "Identification of candidate genes for paratuberculosis resistance in the native Italian Garfagnina goat breed." Trop Anim Health Prod **49**(6): 1135-1142.

Chamberlain, A. J., C. J. Vander Jagt, B. J. Hayes, M. Khansefid, L. C. Marett, C. A. Millen, T. T. T. Nguyen and M. E. Goddard (2015). "Extensive variation between tissues in allele specific expression in an outbred mammal." BMC Genomics **16**: 993.

Dennis J Chia (2014). "Minireview: mechanisms of growth hormone-mediated gene regulation." Mol Endocrinol **28**(7): 1012-1025.

Chopra-Dewasthaly, R., M. Korb, R. Brunthaler and R. Ertl (2017). "Comprehensive RNA-Seq Profiling to Evaluate the Sheep Mammary Gland Transcriptome in Response to Experimental Mycoplasma agalactiae Infection." PLoS One **12**(1): e0170015.

Clark, E. L., S. J. Bush, M. E. B. McCulloch, I. L. Farquhar, R. Young, L. Lefevre, C. Pridans, H. Tsang, C. Wu, C. Afrasiabi, M. Watson, C. B. Whitelaw, T. C. Freeman, K. M. Summers, A. L. Archibald and D. A. Hume (2017). "A high resolution atlas of gene expression in the domestic sheep (Ovis aries)." PLOS Genetics **13**(9): e1006997.

Clarke, C. J. (1997). "The Pathology and Pathogenesis of Paratuberculosis in ruminants and other species." Journal of Comparative Pathology **116**(3): 217–261.

Colli, L., M. Milanesi, A. Talenti, F. Bertolini, M. Chen, A. Crisà, K. G. Daly, M. Del Corvo, B. Guldbrandtsen, J. A. Lenstra, B. D. Rosen, E. Vajana, G. Catillo, S. Joost, E. L. Nicolazzi, E. Rochat, M. F. Rothschild, B. Servin, T. S. Sonstegard, R. Steri, C. P. Van Tassell, P. Ajmone-Marsan, P. Crepaldi, A. Stella and C. the AdaptMap (2018). "Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes." Genetics Selection Evolution **50**(1): 58.

Consortium, RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium (2001). "Functional annotation of a full-length mouse cDNA collection." NATURE **409**.

Corces, M. R., A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria, S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff,

V. I. Risca, A. Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf and H. Y. Chang (2017). "An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues." Nature methods **14**(10): 959-962.

David C. W., Norrman J., Hammon H. M., Davis W. C. and B. J. W. (2003). "Cell Proliferation, Apoptosis, and B- and T-Lymphocytes in Peyer's Patches of the Ileum, in Thymus and in Lymph nodes of Preterm Calves, and in Full-Term Calves at Birth and on Day 5 of Life." J. Dairy Sci. **86**: 3321–3329.

Davies L.C., et al. (2013) Tissue-resident macrophages. Nat. Immunol. 14, 986-95. 10.1038/ni.2705

Dobrovolskaia, M. A. and S. N. Vogel (2002). "Toll receptors, CD14, and macrophage activation and deactivation by LPS.pdf." Microbes and Infection **4**: 903-914.

Dong, Y., G. F. T. Poon, A. A. Arif, S. S. M. Lee-Sayer, M. Dosanjh and P. Johnson (2018). "The survival of fetal and bone marrow monocyte-derived alveolar macrophages is promoted by CD44 and its interaction with hyaluronan." Mucosal Immunol **11**(3): 601-614.

Dong, Y., M. Xie, Y. Jiang, N. Xiao, X. Du, W. Zhang, G. Tosser-Klopp, J. Wang, S. Yang, J. Liang, W. Chen, J. Chen, P. Zeng, Y. Hou, C. Bian, S. Pan, Y. Li, X. Liu, W. Wang, B. Servin, B. Sayre, B. Zhu, D. Sweeney, R. Moore, W. Nie, Y. Shen, R. Zhao, G. Zhang, J. Li, T. Faraut, J. Womack, Y. Zhang, J. Kijas, N. Cockett, X. Xu and S. Zhao (2013). "Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus)." Nat Biotechnol **31**(2): 135-141.

Dongen, S. v. and C. Abreu-Goodger (2012). "Using MCL to Extract Clusters from Networks.pdf." Methods in Molecular Biology **804**: 281-295.

Ellegren, H. (2008). "Comparative genomics and the study of evolution by natural selection." Mol Ecol **17**(21): 4586-4596.

Entrican, G. (2002). "New technologies for studying immune regulation in ruminants.pdf." Veterinary Immunology and Immunopathology(87): 485-490.

FANTOM Consortium (2005). The Transcriptional Landscape of the mammalian genome. Science **309**.

Fairbairn, L., R. Kapetanovic, D. P. Sester and D. A. Hume (2011). "The mononuclear phagocyte system of the pig as a model for understanding human innate immunity and disease." J Leukoc Biol **89**(6): 855-871.

Fairbairn, L., R. Kapetanovic, D. Beraldi, D. P. Sester, C. K. Tuggle, A. L. Archibald and D. A. Hume (2013). "Comparative Analysis of Monocyte Subsets in the Pig." The Journal of Immunology **190**(12): 6389-6396.

Fairfax, B. P., P. Humburg, S. Makino, V. Naranbhai, D. Wong, E. Lau, L. Jostins, K. Plant, R. Andrews, C. McGee and J. C. Knight (2014). "Innate Immune Activity Conditions the Effect of Regulatory Variants upon Monocyte Gene Expression." Science **343**(6175): 1246949.

Fan, Y., R. K. Menon, P. Cohen, D. Hwang, T. Clemens, D. J. DiGirolamo, J. J. Kopchick, D. Le Roith, M. Trucco and M. A. Sperling (2009). "Liver-specific deletion of the growth hormone receptor reveals essential role of growth hormone signaling in hepatic lipid metabolism." J Biol Chem **284**(30): 19937-19944.

Faucette, A. N., V. A. Maher, M. A. Gutierrez, J. M. Jucker, D. C. Yates, T. H. Welsh, Jr., M. Amstalden, G. R. Newton, L. C. Nuti, D. W. Forrest and N. H. Ing (2014). "Temporal changes in histomorphology and gene expression in goat testes during postnatal development1,2." Journal of Animal Science **92**(10): 4440-4448.

Francisco-Cruz, A., M. Aguilar-Santelises, O. Ramos-Espinosa, D. Mata-Espinosa, B. Marquina-Castillo, J. Barrios-Payan and R. Hernandez-Pando (2014). "Granulocyte-macrophage colony-stimulating factor: not just another haematopoietic growth factor." Med Oncol **31**(1): 774.

Hume DA and Freeman TC (2014). "Transcriptomic analysis of mononuclear phagocyte differentiation and activation." Immunological Reviews **262**: 74-84.

Freeman, T. C., A. Ivens, J. K. Baillie, D. Beraldi, M. W. Barnett, D. Dorward, A. Downing, L. Fairbairn, R. Kapetanovic, S. Raza, A. Tomoiu, R. Alberio, C. Wu, A. I. Su, K. M. Summers, C. K. Tuggle, A. L. Archibald and D. A. Hume (2012). "A gene expression atlas of the domestic pig." BMC Biol **10**: 90.

Fujihara, M., M. Muroi, K.-i. Tanamoto, T. Suzuki, H. Azuma and H. Ikeda (2003). "Molecular mechanisms of macrophage activation and deactivation by lipopolysaccharide: roles of the receptor complex." Pharmacology & Therapeutics **100**(2): 171-194.

Gill, H. S., K. Altmann, M. L. Cross and A. J. Husband (2000). "Induction of T helper 1- and T helper 2-type immune responses during Haemonchus contortus infection in sheep." Immunology **99**(3): 458-463.

Giotti, B., S. H. Chen, M. W. Barnett, T. Regan, T. Ly, S. Wiemann, D. A. Hume and T. C. Freeman (2018). "Assembly of a Parts List of the Human Mitotic Cell Cycle Machinery." J Mol Cell Biol.

Gonzalez-Garay, M. L. (2016). Introduction to Isoform Sequencing Using Pacific Biosciences Technology (Iso-Seq). Transcriptomics and Gene Regulation. J. Wu. Dordrecht, Springer Netherlands**:** 141-160.

Gossner, A. G., V. M. Venturina, A. Peers, C. A. Watkins and J. Hopkins (2012). "Expression of sheep interleukin 23 (IL23A, alpha subunit p19) in two distinct gastrointestinal diseases." Vet Immunol Immunopathol **150**(1-2): 118-122.

Grabmuller, M., B. Madea and C. Courts (2015). "Comparative evaluation of different extraction and quantification methods for forensic RNA analysis." Forensic Sci Int Genet **16**: 195-202.

Harhay, G. P., T. P. Smith, L. J. Alexander, C. D. Haudenschild, J. W. Keele, L. K. Matukumalli, S. G. Schroeder, C. P. Van Tassell, C. R. Gresham, S. M. Bridges, S. C. Burgess and T. S. Sonstegard (2010). "An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation." Genome Biol **11**(10): R102.

Haziot, A., E. Ferrero, F. Kontgen, N. Hijiya, S. Yamamoto, J. Silver and C. L. Stewart (1996). "Resistance to Endotoxin Shock and Reduced Disseminationof Gram-Negative Bacteria in CD14-Deficient Mice." Cell Press **4**: 407-414.

Heike Bicker, Conny Hoflich, Kerstin Wolk, KatrinVogt, Hans-Dieter Volk and R. Sabat (2008). "A simple assay to measure phagocytosis of live bacteria.pdf." Clinical Chemistry **54**(5): 911-915.

Heiko Adler, Barbara Adler, Paola Peveri, Ernst R. Werner, Helmut Wachter, Ernst Peterhans and T. W. Jungi (1996). "Differential Regulation of Inducible Nitric Oxide Synthase Production in Bovine and Caprine Macrophages.pdf." The Journal of Infectious Diseases **173**: 971-978.

Hope, J. C., P. Sopp, S. Wattegedera and G. Entrican (2012). "Tools and reagents for caprine immunology." Small Ruminant Research **103**(1): 23-27.

Hume, D. A., K. M. Summers, S. Raza, J. K. Baillie and T. C. Freeman (2010). "Functional clustering and lineage markers: insights into cellular differentiation and gene function from large-scale microarray studies of purified primary cell populations." Genomics **95**(6): 328-338.

D. A. Hume and C. Pridans (2019). "Deletion of a Csf1r enhancer selectively impacts CSF1R expression and development of tissue macrophage populations." Nat Commun **10**(1): 3215.

David A Hume, Ian L. Ross, S. Roy Himes, R. Tedjo Sasmono, Christine A. Wells and Timothy Ravasi (2002). "The mononuclear phagocyte system revisited." Journal of Leukocyte Biology **72**.

David A Hume (2006). "The mononuclear phagocyte system." Curr Opin Immunol **18**(1): 49-53.

David A Hume (2008). "Differentiation and heterogeneity in the mononuclear phagocyte system." Mucosal Immunol **1**(6): 432-441.

Hume DA and Freeman TC (2014). "Transcriptomic analysis of mononuclear phagocyte differentiation and activation." Immunological Reviews **262**: 74-84.

Hussen, J. and H. J. Schuberth (2017). "Heterogeneity of Bovine Peripheral Blood Monocytes." Front Immunol **8**: 1875.

Ibeagha-Awemu, E. M., P. Kgwatalala, A. E. Ibeagha and X. Zhao (2008). "A critical analysis of disease-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig." Mamm Genome **19**(4): 226-245.

Illumina (2017). TruSeq Stranded mRNA Reference Guide.

Ito, H., N. Koide, A. Morikawa, F. Hassan, S. Islam, G. Tumurkhuu, I. Mori, T. Yoshida, S. Kakumu, H. Moriwaki and T. Yokochi (2005). "Augmentation of lipopolysaccharide-induced nitric oxide production by alpha-galactosylceramide in mouse peritoneal cells." J Endotoxin Res **11**(4): 213-219.

Jenkins, S. J. and D. A. Hume (2014). "Homeostasis in the mononuclear phagocyte system." Trends Immunol **35**(8): 358-367.

Jiang, Y., M. Xie, W. Chen, R. Talbot, J. F. Maddox, T. Faraut, C. Wu, D. M. Muzny, Y. Li, W. Zhang, J. A. Stanton, R. Brauning, W. C. Barris, T. Hourlier, B. L. Aken, S. M. Searle, D. L. Adelson, C. Bian, G. R. Cam, Y. Chen, S. Cheng, U. DeSilva, K. Dixen, Y. Dong, G. Fan, I. R. Franklin, S. Fu, P. Fuentes-Utrilla, R. Guan, M. A. Highland, M. E. Holder, G. Huang, A. B. Ingham, S. N. Jhangiani, D. Kalra, C. L. Kovar, S. L. Lee, W. Liu, X. Liu, C. Lu, T. Lv, T. Mathew, S. McWilliam, M. Menzies, S. Pan, D. Robelin, B. Servin, D. Townley, W. Wang, B. Wei, S. N. White, X. Yang, C. Ye, Y. Yue, P. Zeng, Q. Zhou, J. B. Hansen, K. Kristiansen, R. A. Gibbs, P. Flicek, C. C. Warkup, H. E. Jones, V. H. Oddy, F. W. Nicholas, J. C. McEwan, J. W. Kijas, J. Wang, K. C. Worley, A. L. Archibald, N. Cockett, X. Xu, W. Wang and B. P. Dalrymple (2014). "The sheep genome illuminates biology of the rumen and lipid metabolism." Science **344**(6188): 1168-1173.

Julia A, P., M. Suli, M. Bryan A, K. Evan S, F. James J and T. Matthew J (2019). "Brush border protocadherin CDHR2 promotes the elongation and maximized packing of microvilli in vivo." Molecular Biology of the Cell **30**.

Jung, C., J. P. Hugot and F. Barreau (2010). "Peyer's Patches: The Immune Sensors of the Intestine." Int J Inflam **2010**: 823710.

Jungi TW, Adler H, Adler B, Thöny M, Krampe M and P. E. (1996). "Inducible nitric oxide synthase of macrophages. Present knowledge and evidence for species-specific regulation." Vet Immunol Immunopathol **Nov,54**(1-4): 323-330.

Kalds, P., S. Zhou, B. Cai, J. Liu, Y. Wang, B. Petersen, T. Sonstegard, X. Wang and Y. Chen (2019). "Sheep and Goat Genome Engineering: From Random Transgenesis to the CRISPR Era." Front. Genet. .

Kannaki, T. R., M. Shanmugam and P. C. Verma (2011). "Toll-like receptors and their role in animal reproduction." Anim Reprod Sci **125**(1-4): 1-12.

Kapetanovic, R., L. Fairbairn, D. Beraldi, D. P. Sester, A. L. Archibald, C. K. Tuggle and D. A. Hume (2012). "Pig bone marrow-derived macrophages resemble human macrophages in their response to bacterial lipopolysaccharide." J Immunol **188**(7): 3382-3394.

Karagianni, A. E., R. Kapetanovic, K. M. Summers, B. C. McGorum, D. A. Hume and R. S. Pirie (2017). "Comparative transcriptome analysis of equine alveolar macrophages." Equine Vet J **49**(3): 375-382.

Kenya, L. (2019). "Meat goat breeds in Kenya." Retrieved August 1st 2019, from https://www.livestockkenya.com/index.php/blog/sheep-and-goats/141-meat-goat-breeds-in-kenya.

Kim, D., B. Langmead and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low memory requirements." Nat Methods **12**(4): 357-360.

Krupp, M., J. U. Marquardt, U. Sahin, P. R. Galle, J. Castle and A. Teufel (2012). "RNA-Seq Atlas - A reference database for gene expression profiling in normal tissue by next generation sequencing." Bioinformatics **28**.

Krupp, M., J. U. Marquardt, U. Sahin, P. R. Galle, J. Castle and A. Teufel (2012). "RNA-Seq Atlas-a reference database for gene expression profiling in normal tissue by next-generation sequencing." Bioinformatics **28**(8): 1184-1185.

Landmann R, Knopf HP, Link S, Sansano S, Schumann R and Z. W. (1996). "Human monocyte CD14 is upregulated by lipopolysaccharide." INFECTION AND IMMUNITY **65**(5): 1762-1769.

Laporta, J., G. J. Rosa, H. Naya and M. Carriquiry (2014). "Liver functional genomics in beef cows on grazing systems: novel genes and pathways revealed." Physiol Genomics **46**(4): 138-147.

Larruskain, A. and B. M. Jugo (2013). "Retroviral infections in sheep and goats: small ruminant lentiviruses and host interaction." Viruses **5**(8): 2043-2061.

Law, R. H., Q. Zhang, S. McGowan, A. M. Buckle, G. A. Silverman, W. Wong, C. J. Rosado, C. G. Langendorf, R. N. Pike, P. I. Bird and J. C. Whisstock (2006). "An overview of the serpin superfamily." Genome Biology **7**(216).

F. Lechner, A. S., U. Von Bodungen, G. Beroni, H. Pfister, T. W. Jungi, and E. Peterhans (1999). "Inducible nitric oxide synthase is expressed in joints of goats in the late stage of infection with caprine arthritis encephalitis virus." Clin Exp Immunol **117**: 70-75.

Lei, H. K., Vasiliou Daniel W, Nebert (2009). "Analysis and update of the human solute carrier (SLC) gene superfamily." Human Genomics **3**: 195-206.

Link, V. M., S. H. Duttke, H. B. Chun, I. R. Holtman, E. Westin, M. A. Hoeksema, Y. Abe, D. Skola, C. E. Romanoski, J. Tao, G. J. Fonseca, T. D. Troutman, N. J. Spann, T. Strid, M. Sakai, M. Yu, R. Hu, R. Fang, D. Metzler, B. Ren and C. K. Glass (2018). "Analysis of Genetically Diverse Macrophages Reveals Local and Domain-wide Mechanisms that Control Transcription Factor Binding and Function." Cell **173**(7): 1796-1809 e1717.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.

Lopetuso, L. R., S. Chowdhry and T. T. Pizarro (2013). "Opposing Functions of Classic and Novel IL-1 Family Members in Gut Health and Disease." Front Immunol **4**: 181.

Mabbott, N. A., J. K. Baillie, D. A. Hume and T. C. Freeman (2010). "Meta-analysis of lineage-specific gene expression signatures in mouse leukocyte populations." Immunobiology **215**(9-10): 724-736.

Mandard S, Muller M and Kersten S (2004). "Peroxisome proliferator-activated receptor alpha target genes." Cell Mol Life Sci **61**(4): 393-416.

Marim, F. M., T. N. Silveira, D. S. Lima, Jr. and D. S. Zamboni (2010). "A method for generation of bone marrow-derived macrophages from cryopreserved mouse bone marrow cells." PLoS One **5**(12): e15263.

Martinez, F. O., S. Gordon, M. Locati and A. Mantovani (2006). "Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression." J Immunol **177**(10): 7303-7311.

Mele, M., P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segre, S. Djebali, A. Niarchou, G. T. Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie and R. Guigo (2015). "Human genomics. The human transcriptome across tissues and individuals." Science **348**(6235): 660-665.

Mi, H., Q. Dong, A. Muruganujan, P. Gaudet, S. Lewis and P. D. Thomas (2010). "PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium." Nucleic Acids Res **38**(Database issue): D204-210.

Munoz Mendoza, M., L. Juan, S. Menendez, A. Ocampo, J. Mourelo, J. L. Saez, L. Dominguez, C. Gortazar, J. F. Garcia Marin and A. Balseiro (2012). "Tuberculosis due to Mycobacterium bovis and Mycobacterium caprae in sheep." Vet J **191**(2): 267-269.

Muriuki, C., S. J. Bush, M. Salavati, M. E. B. McCulloch, Z. M. Lisowski, M. Agaba, A. Djikeng, D. A. Hume and E. L. Clark (2019). "A Mini-Atlas of Gene Expression for the Domestic Goat (Capra hircus)." Frontiers in Genetics **10**.

Nilsson, R., V. B. Bajic, H. Suzuki, D. di Bernardo, J. Bjorkegren, S. Katayama, J. F. Reid, M. J. Sweet, M. Gariboldi, P. Carninci, Y. Hayashizaki, D. A. Hume, J. Tegner and T. Ravasi (2006). "Transcriptional network dynamics in macrophage activation." Genomics **88**(2): 133-142.

Oliver, S. (2000). "Proteomics: Guilt-by-association goes global." Nature **403**(6770): 601-603.

Pelaseyed, T., J. H. Bergstrom, J. K. Gustafsson, A. Ermund, G. M. Birchenough, A. Schutte, S. van der Post, F. Svensson, A. M. Rodriguez-Pineiro, E. E. Nystrom, C. Wising, M. E. Johansson and G. C. Hansson (2014). "The mucus and mucins of the goblet cells and enterocytes provide the first

defense line of the gastrointestinal tract and interact with the immune system." Immunol Rev **260**(1): 8-20.

Pertea, M., G. M. Pertea, C. M. Antonescu, T. C. Chang, J. T. Mendell and S. L. Salzberg (2015). "StringTie enables improved reconstruction of a transcriptome from RNA-seq reads." Nat Biotechnol **33**(3): 290-295.

Pertea, M., D. Kim, G. M. Pertea, J. T. Leek and S. L. Salzberg (2016). "Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown." Nat Protoc **11**(9): 1650-1667.

Pridans, C., G. M. Davis, K. A. Sauter, Z. M. Lisowski, Y. Corripio-Miyar, A. Raper, L. Lefevre, R. Young, M. E. McCulloch, S. Lillico, E. Milne, B. Whitelaw and D. A. Hume (2016). "A Csf1r-EGFP Transgene Provides a Novel Marker for Monocyte Subsets in Sheep." J Immunol **197**(6): 2297-2305.

Pruitt, K. D., T. Tatusova and D. R. Maglott (2007). "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins." Nucleic acids research **35**(Database issue): D61-D65.

Rafael A. Nafikov and Donald C. Beitz (2007). "Carbohydrate and lipid metabolism in farm animals." The Journal of Nutrition **137**: 702-705.

Ravasi, T., H. Suzuki, C. V. Cannistraci, S. Katayama, V. B. Bajic, K. Tan, A. Akalin, S. Schmeier, M. Kanamori-Katayama, N. Bertin, P. Carninci, C. O. Daub, A. R. Forrest, J. Gough, S. Grimmond, J. H. Han, T. Hashimoto, W. Hide, O. Hofmann, A. Kamburov, M. Kaur, H. Kawaji, A. Kubosaki, T. Lassmann, E. van Nimwegen, C. R. MacPherson, C. Ogawa, A. Radovanovic, A. Schwartz, R. D. Teasdale, J. Tegner, B. Lenhard, S. A. Teichmann, T. Arakawa, N. Ninomiya, K. Murakami, M. Tagami, S. Fukuda, K. Imamura, C. Kai, R. Ishihara, Y. Kitazume, J. Kawai, D. A. Hume, T. Ideker and Y. Hayashizaki (2010). "An atlas of combinatorial transcriptional regulation in mouse and man." Cell **140**(5): 744-752.

Raza, S., M. W. Barnett, Z. Barnett-Itzhaki, I. Amit, D. A. Hume and T. C. Freeman (2014). "Analysis of the transcriptional networks underpinning the

activation of murine macrophages by inflammatory mediators." J Leukoc Biol **96**(2): 167-183.

LA Reddacliff, H. M. K Beh and R. Whittington (2005). "A preliminary study of possible genetic influences on the susceptibility of sheep to Johne's disease." Australian Veterinary Journal **83**: 435-441.

Rice, P., I. Longden and A. Bleasby (2000). "EMBOSS: The European Molecular Biology Open Software Suite." Trends in Genetics **16**(6): 276-277.

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2009). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.

Rojo, R., A. Raper, D. D. Ozdemir, L. Lefevre, K. Grabert, E. Wollscheid-Lengeling, B. Bradford, M. Caruso, I. Gazova, A. Sanchez, Z. M. Lisowski, J. Alves, I. Molina-Gonzalez, H. Davtyan, R. J. Lodge, J. D. Glover, R. Wallace, D. A. D. Munro, E. David, I. Amit, V. E. Miron, J. Priller, S. J. Jenkins, G. E. Hardingham, M. Blurton-Jones, N. A. Mabbott, K. M. Summers, P. Hohenstein, Russell, G. C., J. P. Stewart and D. M. Haig (2009). "Malignant catarrhal fever: a review." Vet J **179**(3): 324-335.

Ruth M. Ruprecht, N. C. Goodman and S. Spiegelman (1973). "Conditions for the selective synthesis of DNA complementary to template RNA.pdf>." Biochimica et Biophysica Acta **294**: 192-203.

Salavati, M., S. J. Bush, S. Palma-Vera, M. E. B. McCulloch, D. A. Hume and E. L. Clark (2019). "Elimination of reference mapping bias reveals robust immune related allele-specific expression in cross-bred sheep." bioRxiv: 619122.

Schneider, C., S. P. Nobs, M. Kurrer, H. Rehrauer, C. Thiele and M. Kopf (2014). "Induction of the nuclear receptor PPAR-gamma by the cytokine GM-CSF is critical for the differentiation of fetal monocytes into alveolar macrophages." Nat Immunol **15**(11): 1026-1037.

Schoggins, J. W., S. J. Wilson, M. Panis, M. Y. Murphy, C. T. Jones, P. Bieniasz and C. M. Rice (2011). "A diverse range of gene products are

effectors of the type I interferon antiviral response." Nature **472**(7344): 481-485.

Schoggins, J. W. (2019). "Interferon-Stimulated Genes: What Do They All Do?" Annu Rev Virol **6**(1): 567-584.

Kate Schroder, Katharine M. Irvine, Martin S. Taylor, Nilesh J. Bokil, Kim-Anh Le Cao, Kelly-Anne Masterman, Larisa I. Labzin, Colin A. Semple, Ronan Kapetanovic, Lynsey Fairbairn, Altuna Akalin, Geoffrey J. Faulkner, John Kenneth Baillie, Milena Gongora, Carsten O. Daub, Hideya Kawaji, Geoffrey J. McLachlan, Nick Goldman, Sean M. Grimmond, Piero Carninci, Harukazu Suzuki, Yoshihide Hayashizaki, Boris Lenhard, David A. Hume and M. J. Sweet (2012). "Conservation and divergence in Toll-like receptor 4-regulated gene expression in primary human versus mouse macrophages." PNAS **109**(16).

Schwartz, J. C., R. L. Philp, D. M. Bickhart, T. P. L. Smith and J. A. Hammond (2018). "The antibody loci of the domestic goat (Capra hircus)." Immunogenetics **70**(5): 317-326.

Shaw, A. E., J. Hughes, Q. Gu, A. Behdenna, J. B. Singer, T. Dennis, R. J. Orton, M. Varela, R. J. Gifford, S. J. Wilson and M. Palmarini (2017). "Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses." PLoS Biol **15**(12): e2004086.

Sherman, D. M. (2011). "The spread of pathogens through trade in small ruminants and their products." Rev Sci Tech(0253-1933 (Print)): 207-217.

Shibata, Y., Berclaz, P.Y., Chroneos, Z.C., Yoshida, M., Whitsett, J.A. & Trapnell, B.C (2001). "GM-CSF regulates alveolar macrophage differentiation and innate immunity in the lung through PU.1.pdf." Immunity **15**: 557-567.

Singh, A., M. Prasad, B. Mishra, S. Manjunath, A. R. Sahu, G. Bhuvana Priya, S. A. Wani, A. P. Sahoo, A. Kumar, S. Balodi, A. Deora, S. Saxena and R. K. Gandham (2017). "Transcriptome analysis reveals common differential and global gene expression profiles in bluetongue virus serotype 16 (BTV-16)

infected peripheral blood mononuclear cells (PBMCs) in sheep and goats." Genom Data **11**: 62-72.

Stella, A., E. L. Nicolazzi, C. P. Van Tassell, M. F. Rothschild, L. Colli, B. D. Rosen, T. S. Sonstegard, P. Crepaldi, G. Tosser-Klopp, S. Joost, M. Amills, P. Ajmone-Marsan, F. Bertolini, P. Boettcher, R. Boyle Onzima, D. Bradley, D. Buja, M. E. Cano Pereira, A. Carta, G. Catillo, L. Colli, P. Crepaldi, A. Crisà, M. Del Corvo, K. Daly, C. Droegemueller, S. Duruz, A. Elbeltagi, A. Esmailizadeh, O. Faco, T. Figueiredo Cardoso, C. Flury, J. F. Garcia, B. Guldbrandtsen, A. Haile, J. Hallsteinn Hallsson, M. Heaton, V. Hunnicke Nielsen, H. Huson, S. Joost, J. Kijas, J. A. Lenstra, G. Marras, M. Milanesi, C. Minhui, M. Moaeen-ud-Din, R. Morry O'Donnell, O. Moses Danlami, J. Mwacharo, E. L. Nicolazzi, I. Palhière, F. Pilla, M. Poli, J. Reecy, B. A. Rischkowsky, E. Rochat, B. Rosen, M. Rothschild, R. Rupp, B. Sayre, B. Servin, K. Silva, T. Sonstegard, G. Spangler, A. Stella, R. Steri, A. Talenti, F. Tortereau, G. Tosser-Klopp, E. Vajana, C. P. Van Tassell, W. Zhang and C. the AdaptMap (2018). "AdaptMap: exploring goat diversity and adaptation." Genetics Selection Evolution **50**(1): 61.

Stelzer, G., N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T. I. Stein, R. Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N. Rappaport, M. Safran and D. Lancet (2016). "The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses." Current Protocols in Bioinformatics **54**(1): 1.30.31-31.30.33.

Stevenson, K. (2015). "Genetic diversity of Mycobacterium avium subspecies paratuberculosis and the influence of strain type on infection and pathogenesis: a review." Vet Res **46**: 64.

Su, A. I., M. P. Cooke, K. A. Ching, Y. Hakak, J. R. Walker, T. Wiltshire, A. P. Orth, R. G. Vega, L. M. Sapinoso, A. Moqrich, A. Patapoutian, G. M. Hampton, P. G. Schultz and J. B. Hogenesch (2002). "Large-scale analysis of the human and mouse transcriptomes." Proc Natl Acad Sci U S A **99**(7): 4465-4470.

Su, A. I., T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker and J. B.

Hogenesch (2004). "A gene atlas of the mouse and human protein-encoding transcriptomes." Proc Natl Acad Sci U S A **101**(16): 6062-6067.

T. Hubbard, D. B., E. Birney, G. Cameron, Y. Chen, L. Clark, T. Cox, J. Cuff, V. Curwen, T. Down, R. Durbin, E. Eyras, J. Gilbert, M. Hammond, L. Huminiecki, A. Kasprzyk, H. Lehvaslaiho, P. Lijnzaad, C. Melsopp, E. Mongin, R. Pettett, M. Pocock, S. Potter, A. Rust, E. Schmidt, S. Searle, G. Slater, J. Smith, W. Spooner, A. Stabenau, J. Stalker, E. Stupka, A. Ureta-Vidal, I. Vastrik, M. Clamp (2002). "The Ensembl genome database project." Nucleic Acids Res **30**(1).

Taka, S., E. Liandris, M. Gazouli, K. Sotirakoglou, G. Theodoropoulos, M. Bountouri, M. Andreadou and J. Ikonomopoulos (2013). "In vitro expression of the SLC11A1 gene in goat monocyte-derived macrophages challenged with Mycobacterium avium subsp paratuberculosis." Infect Genet Evol **17**: 8-15.

Takacs, A. C., I. J. Swierzy and C. G. Luder (2012). "Interferon-gamma restricts Toxoplasma gondii development in murine skeletal muscle cells via nitric oxide production and immunity-related GTPases." PLoS One **7**(9): e45440.

Talenti, A., I. Palhière, F. Tortereau, G. Pagnacco, A. Stella, E. L. Nicolazzi, P. Crepaldi, G. Tosser-Klopp and C. AdaptMap (2018). "Functional SNP panel for parentage assessment and assignment in worldwide goat breeds." Genetics Selection Evolution **50**(1): 55.

Theocharidis, A., S. van Dongen, A. J. Enright and T. C. Freeman (2009). "Network visualization and analysis of gene expression data using BioLayout Express(3D)." Nat Protoc **4**(10): 1535-1550.

Tirumurugaan, K. G., S. Dhanasekaran, G. D. Raj, A. Raja, K. Kumanan and V. Ramaswamy (2010). "Differential expression of toll-like receptor mRNA in selected tissues of goat (Capra hircus)." Vet Immunol Immunopathol **133**(2-4): 296-301.

Tosi, M. F. (2005). "Innate immune responses to infection." J Allergy Clin Immunol **116**(2): 241-249; quiz 250.

Tosser-Klopp, G., P. Bardou, O. Bouchez, C. Cabau, R. Crooijmans, Y. Dong, C. Donnadieu-Tonon, A. Eggen, H. C. M. Heuven, S. Jamli, A. J. Jiken, C. Klopp, C. T. Lawley, J. McEwan, P. Martin, C. R. Moreno, P. Mulsant, I. Nabihoudine, E. Pailhoux, I. Palhière, R. Rupp, J. Sarry, B. L. Sayre, A. Tircazes, W. Jun, W. Wang, W. Zhang and C. and the International Goat Genome (2014). "Design and Characterization of a 52K SNP Chip for Goats." PLOS ONE **9**(1): e86227.

VarGoats. (2019). "VarGoats: Identification of Variations in Goat genomes related to domestication and adaptation." Retrieved 1st August, 2019, from http://www.goatgenome.org/vargoats.html.

Walia, V., R. Kumar and A. Mitra (2015). "Lipopolysaccharide and Concanavalin A Differentially Induce the Expression of Immune Response Genes in Caprine Monocyte Derived Macrophages." Anim Biotechnol **26**(4): 298-303.

Wambua, L., P. N. Wambua, A. M. Ramogo, D. Mijele and M. Y. Otiende (2016). "Wildebeest-associated malignant catarrhal fever: perspectives for integrated control of a lymphoproliferative disease of cattle in sub-Saharan Africa." Arch Virol **161**(1): 1-10.

Wang, M., T. P. Hancock, A. J. Chamberlain, C. J. Vander Jagt, J. E. Pryce, B. G. Cocks, M. E. Goddard and B. J. Hayes (2018). "Putative bovine topological association domains and CTCF binding motifs can reduce the search space for causative regulatory variants of complex traits." bioRxiv.

Wang, M., H. Yu, Y. S. Kim, C. A. Bidwell and S. Kuang (2012). "Myostatin facilitates slow and inhibits fast myosin heavy chain expression during myogenic differentiation." Biochemical and Biophysical Research Communications **426**(1): 83-88.

Wani, S. A., A. R. Sahu, R. I. N. Khan, A. Pandey, S. Saxena, N. Hosamani, W. A. Malla, D. Chaudhary, S. Kanchan, V. Sah, K. K. Rajak, D. Muthuchelvan, B. Mishra, A. K. Tiwari, A. P. Sahoo, B. Sajjanar, Y. P. Singh, R. K. Gandham, B. P. Mishra and R. K. Singh (2019). "Contrasting Gene Expression Profiles of

Monocytes and Lymphocytes From Peste-Des-Petits-Ruminants Virus Infected Goats." Frontiers in Immunology **10**: 1463.

Wells, C. A., A. M. Chalk, A. Forrest, D. Taylor, N. Waddell, K. Schroder, S. R. Himes, G. Faulkner, S. Lo, T. Kasukawa, H. Kawaji, C. Kai, J. Kawai, S. Katayama, P. Carninci, Y. Hayashizaki, D. A. Hume and S. M. Grimmond (2006). "Alternate transcription of the Toll-like receptor signaling cascade." Genome Biol **7**(2): R10.

Wickramasinghe, S., A. Cánovas, G. Rincón and J. F. Medrano (2014). "RNA-Sequencing: A tool to explore new frontiers in animal genetics." Livestock Science **166**: 206-216.

Worley, K. C. (2017). "A golden goat genome." Nat Genet **49**(4): 485-486.

Wu, Z., T. Hu, L. Rothwell, L. Vervelde, P. Kaiser, K. Boulton, M. J. Nolan, F. M. Tomley, D. P. Blake and D. A. Hume (2016). "Analysis of the function of IL-10 in chickens using specific neutralising antibodies and a sensitive capture ELISA." Dev Comp Immunol **63**: 206-212.

Xiang, R., J. McNally, S. Rowe, A. Jonker, C. S. Pinares-Patino, V. H. Oddy, P. E. Vercoe, J. C. McEwan and B. P. Dalrymple (2016). "Gene network analysis identifies rumen epithelial cell proliferation, differentiation and metabolic pathways perturbed by diet and correlated with methane production." Sci Rep **6**: 39022.

Yadav, V. P., S. S. Dangi, V. S. Chouhan, M. Gupta, S. K. Dangi, G. Singh, V. P. Maurya, P. Kumar and M. Sarkar (2016). "Expression analysis of NOS family and HSP genes during thermal stress in goat (Capra hircus)." Int J Biometeorol **60**(3): 381-389.

Yamamoto, M., S. Sato, H. Hemmi, S. Uematsu, K. Hoshino, T. Kaisho, O. Takeuchi, K. Takeda and S. Akira (2003). "TRAM is specifically involved in the Toll-like receptor 4-mediated MyD88-independent signaling pathway." Nat Immunol **4**(11): 1144-1150.

Yan Wang, Dante S. Zarlenga, Max J. Paape and G. E. Dahl (2002). "Recombinant bovine soluble CD14 sensitizes the mammary gland to

lipopolysaccharide." Veterinary Immunology and Immunopathology **86**: 115-124.

Yates, A. D., P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, A. G. Azov, R. Bennett, J. Bhai, K. Billis, S. Boddu, J. C. Marugan, C. Cummins, C. Davidson, K. Dodiya, R. Fatima, A. Gall, C. G. Giron, L. Gil, T. Grego, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, M. Kay, I. Lavidas, T. Le, D. Lemos, J. G. Martinez, T. Maurel, M. McDowall, A. McMahon, S. Mohanan, B. Moore, M. Nuhn, D. N. Oheh, A. Parker, A. Parton, M. Patricio, M. P. Sakthivel, A. I. Abdul Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, M. Sycheva, M. Szuba, K. Taylor, A. Thormann, G. Threadgold, A. Vullo, B. Walts, A. Winterbottom, A. Zadissa, M. Chakiachvili, B. Flint, A. Frankish, S. E. Hunt, I. I. G, M. Kostadima, N. Langridge, J. E. Loveland, F. J. Martin, J. Morales, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, S. J. Trevanion, F. Cunningham, K. L. Howe, D. R. Zerbino and P. Flicek (2020). "Ensembl 2020." Nucleic Acids Res **48**(D1): D682-D688.

Young, R., L. Lefevre, S. J. Bush, A. Joshi, H. S. Singh, Jadhav S., D. Iamartino, K. M. Summers, A. L. Archibald, J. L. Williams, S. Gokhale, S. Kumar and D. A. Hume (2019). "A gene expression atlas of the domestic water buffalo (Bubalus bubalis). ." Frontiers Genetics

Young, R., S. J. Bush, L. Lefevre, M. E. B. McCulloch, Z. M. Lisowski, C. Muriuki, L. A. Waddell, K. A. Sauter, C. Pridans, E. L. Clark and D. A. Hume (2018). "Species-Specific Transcriptional Regulation of Genes Involved in Nitric Oxide Production and Arginine Metabolism in Macrophages." ImmunoHorizons **2**(1): 27-37.

Yu, B., R. Lu, Y. Yuan, T. Zhang, S. Song, Z. Qi, B. Shao, M. Zhu, F. Mi and Y. Cheng (2016). "Efficient TALEN-mediated myostatin gene editing in goats." BMC Developmental Biology **16**(1): 26.

Yu, I. M., V. J. Planelles-Herrero, Y. Sourigues, D. Moussaoui, H. Sirkia, C. Kikuti, D. Stroebel, M. A. Titus and A. Houdusse (2017). "Myosin 7 and its adaptors link cadherins to actin." Nat Commun **8**: 15864.

Zarember, K. A. and P. J. Godowski (2002). "Tissue Expression of Human Toll-Like Receptors and Differential Regulation of Toll-Like Receptor mRNAs in Leukocytes in Response to Microbes, Their Products, and Cytokines." The Journal of Immunology **168**(2): 554-561.

Zerbino, D. R., P. Achuthan, W. Akanni, M R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates and P. Flicek (2018). "Ensembl 2018." Nucleic Acids Research **46**(D1): D754-D761.

Zhang, G.-M., T.-T. Zhang, S.-Y. An, M. A. El-Samahy, H. Yang, Y.-J. Wan, F.-X. Meng, S.-H. Xiao, F. Wang and Z.-H. Lei (2019). "Expression of Hippo signaling pathway components in Hu sheep male reproductive tract and spermatozoa." Theriogenology **126**: 239-248.

Zhang, Y., X. D. Zhang, X. Liu, Y. S. Li, J. P. Ding, X. R. Zhang and Y. H. Zhang (2013). "Reference gene screening for analyzing gene expression across goat tissue." Asian-Australas J Anim Sci **26**(12): 1665-1671.

5.6 Appendices

Appendices are in files contained in the CD on the inside back cover of this thesis.

Appendix A Unaveraged TPM estimates_kallisto output
Appendix B GO term enrichment analysis cluster 1-30
Appendix C TapeStation Results
Appendix D Annotations of novel transcript models
Appendix E Roslin_SOP_Alignment-based RNA-Seq-Processing
Appendix F Expression profile of ARG2 gene
Appendix G Gene expression representation per tissue
Appendix H Gene level comparison to reference annotation
Appendix I Transcript level comparison to reference annotation
Appendix J Annotations of novel transcript models in reference transcriptome
Appendix K Functional evidence for unknown genes in reference transcriptome
Appendix L Possible names for all unknown protein coding genes in reference transcriptome
Appendix M Goat BMDM sample submission form to Edinburgh Genomics
Appendix N Goat BMDM Unaveraged TPM estimates_kallisto output
Appendix O List of housekeeping genes expressed by goat BMDM post-LPS stimulation
Appendix P GO terms of housekeeping genes expressed by goat BMDM post-LPS
Appendix Q GO terms of downregulated genes in goat BMDM post-LPS
Appendix R List of downregulated genes in goat BMDM post-LPS stimulation
Appendix S GO terms of upregulated genes in goat BMDM post-LPS
Appendix T List of upregulated genes in goat BMDM post-LPS stimulation
Appendix U Sheep RNA-Seq subset data