Title: **Insights into the transcriptome of sheep circulating immune cells using single-nuclei RNA-sequencing**

**Application**: Improve the speed of analysing future single-cell experiments in sheep and provide a reference for digital cell sorting from bulk RNA-sequencing data; reducing costs and increasing statistical power.

**Intro**: Exploring the transcriptome at cellular resolution can allow us to detect variations in gene expression across cell and tissue types with unprecedented granularity. This is particularly powerful in heterogenous tissue types where signatures from rarer cell types may be undetectable or considered background noise when using bulk RNA-sequencing, due to their relatively low level of expression.

In sheep, only a few single-cell experiments have been reported in the literature, and none so far have investigated immune tissues. This experiment will investigate the heterogeneity of sheep circulating immune cells, providing new insights into their transcriptional activity, a resource for faster and improved annotation of single-cell datasets in this tissue type, and a basis for deconvolution of bulk RNA-sequencing data to cellular resolution.

**Materials & Methods**: We carried out snRNA-seq of peripheral blood mononuclear cells (PBMCs) sorted from whole blood obtained from adult sheep. In choosing a single-nuclei methodology over whole cells we aimed to improve recovery of all cell types after freezing; reducing the loss of delicate and rarer cells, and allowing the overall cell type proportions to be maintained. After euthanasia with pentobarbital, blood from female commercial cross-breed sheep (n=3) was collected via cardiac puncture. PBMCs were isolated using density gradient centrifugation and stored at -155°C. Nuclei were extracted using a modified version of Ruiz-Daniels et. al. (2023) then fixed and permeabilised using the Parse Biosciences Nuclei fixation kit. Subsequently snRNA-seq libraries were produced using the Parse Biosciences Evercode WT Mini Kit targeting the capture of the transcriptomes of 13,000 nuclei with equal representation of the 3 samples used. Libraries were sequenced on an Illumina Nextseq 2000 P2 flowcell.

Barcoded reads were demultiplexed using the Parse Split-pipe computational pipeline v1.1.0, the latest version of the sheep genome Rambouillet V2 (Davenport et al., 2022), and STAR as the underlying genome mapping tool. The count matrix was imported into R for analysis with Seurat v4. Barcodes with <200 or >25000 transcripts were filtered out, to remove captured ambient RNA and multiplets respectively. After filtering, 9235 nuclei successfully passed quality control criteria.

We normalised the gene expression values before performing PCA and dimensional reduction (UMAP). Clustering was carried out using a Louvain algorithm to co-locate cells based on the similarity of the gene expression. Differential expression testing between clusters allowed us to identify the genes most significantly associated with each cell cluster.

We used known marker genes to interrogate the cell clusters and produce annotations. In addition, we carried out Gene Ontology analysis to investigate the cellular pathways active within the clusters. Two large groups of clusters, thought to represent B-cell and T-cell lineages, were subset from the main dataset and re-clustered to allow finer interrogation of the cell subtypes.

**Results**:

14 clusters of nuclei were identified representing immune cell types and subsets. This shows that sheep PBMCs are highly heterogeneous, with groups of smaller clusters representing cell sub-types. After cluster annotation, superclusters of major immune cell types were identified as well as clusters of rarer cell types (fig. 1).



Figure 1 - UMAP plot of single-nuclei clusters

Canonical cell marker genes were sporadic within the dataset, and some expected markers were absent entirely (CD8, TLR4). Inspecting active pathways, manually reviewing the top significant genes for each cluster (*Bonferroni adjusted p-value <0.05*), and searching for classical cell markers allowed us to identify clusters associated with B cells, T cells, monocyte-derived macrophages, NK cells, Gamma-Delta T cells, and dendritic cells.

**Conclusions**: PBMCs in sheep are a highly heterogenous and complex mixture of immune cell types and sub-types. Classical marker genes are not always helpful when annotating single-nuclei/-cell data, as they may not be captured by the snRNA-seq protocol. However, this analysis has allowed us to capture rare cell types such as dendritic cells, which might otherwise be lost in the averaged expression levels of bulk RNA-seq datasets.

The gene sets identified through this analysis can be used as a basis for cell type annotation in future snRNA-seq experiments in sheep, reducing the time required to manually classify cell types. In addition, this resource will provide a basis for deconvolution of single-cell level gene expression signatures from bulk RNA-seq data; contributing to the economical expansion of experiments from small to larger numbers of animals.

**References**:

Davenport, K.M., Bickhart, D.M., Worley, K., Murali, S.C., Salavati, M., Clark, E.L., Cockett, N.E., Heaton, M.P., Smith, T.P.L., Murdoch, B.M., Rosen, B.D., 2022. An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. Gigascience 11, giab096. https://doi.org/10.1093/gigascience/giab096

Daniels, R.R., Taylor, R.S., Dobie, R., Salisbury, S., Furniss, J.J., Clark, E., Macqueen, D.J., Robledo, D., 2023. A versatile nuclei extraction protocol for single nucleus sequencing in non-model species–Optimization in various Atlantic salmon tissues. PLOS ONE 18, e0285020. https://doi.org/10.1371/journal.pone.0285020