Case study 5: analysis of an unknown sample

The analysis is based on an unknown sample obtained from either mice or human tissues. After tissue dissociation, cells were sorted by FACS and scRNA-seq libraries were prepared with the Smart-Seq2 protocol/platform.

The starting point is a digital count matrix with mouse or human genes as features.

Analyze the dataset with the methods that you prefer and try to cover the following points.

1) Quality control and filtering. Quality control and filtering (mitochondrial genes might be missing from the count matrix, but you can use spike-in RNAs, all starting with "ERCC", as an alternative quality control). ERCC: External RNA Controls Consortium.

2) Normalization, identification of variable features, scaling (normal procedure or sctransform)
3) Dimensionality reduction
4) Clustering
5) Identification of marker genes
6) Cell cycle analysis
7) Annotation

Based on your analysis, on the clusters of cells and marker genes you identified, what is your hypothesis on the sample you are analyzing?

8) *Extra: try to integrate and analyze 2 different samples.*