1. Load dataset and Initialize the Seurat object are slightly different with the Seurat tutorial, since the format of this dataset is different with the one used in the tutorial. You can use the following codes:

# Load the dataset

counts <- Read10X\_h5("data/project4\_data/filtered\_feature\_bc\_matrix.h5")

# Initialize the Seurat object with the raw (non-normalized data)

epi50 <- CreateSeuratObject(

counts = counts$`Gene Expression`

)

2. When calculate percentage of reads that map to the mitochondrial genome, we use the set of all genes starting with mt- (not MT-) as a set of mitochondrial genes.

3. For cell type clustering, let’s use resolution = 0.1. Since the gene expression profiles between EVL and YSL are very different, we don’t need to use a high clustering resolution.

4. For cell type annotation, there are five cell types: Mesoderm cells, Ectoderm cells, EVL cells, YSL cells and Apoptosis-like cells. You can use FeaturePlot to check the expression pattern of the following genes: tbxta, sox3, krt8, slc26a1, tp73, to decide which cluster corresponding to which cell type.

You can check the functional annotations of these genes from https://zfin.org/

5. A list of metabolically related genes saved in file “lipid\_metabolic\_process.txt” in “project4\_data” folder. To find which metabolically related genes are differentially expressed between EVL and YSL, when can first find all differentially expressed genes between EVL and YSL. Then, we can use “subset” function from R to find the intersection between all differentially expressed genes and metabolically related genes. Finally, to show the expression patterns of these differentially expressed metabolically related genes across all cell types, we can use “DoHeatmap” function from Seurat package. We can produce two heatmap figures, one for genes with higher expression in EVL, the other one for genes with higher expression in YSL.