**snRNAseq analysis Homework #3**

This homework is going to build on the previous week’s homework while making three important evolutions.

1. Introduce current standards for dataset quality control (removal of ambient RNA with SoupX and doublet detection).
2. Writing forloops to improve efficiency
3. Saving differentially expressed genes (DEG) lists.

For this homework we will compare how the current pipeline of analysis compares to the old pipeline you used in Week 2 homework when it comes to making clean populations of cells. This question set will require a lot of problem solving but hopefully the reward for the comparison will be satisfying.

Afterwards, we will practice creating a modular score to visualize trends in the dataset as well as perform differential expression gene (DEG) analysis and saving the outputs as Rdata files and excel files.

1. Look at your answers for last week’s homework and make sure to save an .RDS file that labels the cells that you considered doublets (these labels should come from the doublet/mixed clusters).
2. Analyze the “1” dataset (same as week 2 homework) with the current pipeline:
   1. Uploaded as .h5 files, run SoupX and Doublet Detector.
      1. Save an object that has both singlets and doublets
      2. Save an object that is only the singlets
   2. Make a table that compares how many doublets the current pipeline identifies and how that relates to the populations of doublets you identified last week.
      1. What percentage of doublets did the current pipeline identify from the old doublet population?
   3. Use the metadata from the previous week’s homework to graph where the cells that were previously labeled as doublets are localized to.
      1. Compare where the cells go between the object that has “singlets and doublets” versus only “singlets”.
         1. Where did the doublet cells go?
3. Supercluster and identify the cell types by adding 2 levels of metadata.
4. Create a modular score that includes at least 5 genes that define mesangial cells
   1. Is there a cluster that contains mesangial cells?
   2. Is it specific to mesangial cells…or are there more cell types there?
5. Write a forloop that graphs a vlnplot for the top 3 genes that define each cluster
6. Perform DEG analysis on the clusters of the saved object and save an Rdata file and an Excel Spreadsheet.
   1. Bonus points if you write the code to do it based on cell-type names and not cluster numbers.

**Week 3 Learning Objectives:**

1. Merge and integrate two datasets together
2. Remove Ambient RNA with the SoupX package
3. Remove doublets with the doublet detector package
4. Create modular scores to look for patterns of changes with groups of genes
5. Exporting metadata from a Seurat object
6. Merging metadata from a Seurat object
7. Write a forloop
   1. Write a forloop for graphing genes individually with unique labels
   2. Write a forloop to store DEG lists
8. Move DEG lists into new environment
9. Save DEG lists as an Rdata file
10. Save DEG lists as an excel file