# Analysis of young and aged muscle sc-RNA Sequencing

## Analysis of Young and Aged muscle scRNA-seq

Upon importing the 10X, the total number of cells that exist is in the young and aged matrices are 2061 and 1551 cells respectively, and the total number of genes are 27998 for both cases.

Upon creating the Seurat object, we filter out genes and cells with the following criteria:

* Gene must be expressed in at least 1 cell
* Cells that are expressing less than 200 or more than 5000 genes.
* Cells with more than 20000 Unique Molecular Identifiers (UMIs)
* Cells with more than 25% of their reads mapping to mitochondrial genes.

After filtering, the number of cells in the young and aged matrices are 2039 and 1532, respectively. The number of genes in the young and aged matrices are 16539 and 16892, respectively.

We next Normalize the data, find the highly variable genes in each condition, and scale the datasets. We next find the top variable genes that are highly variable in both datasets.

Next, we use the highly variable genes to perform principal component analysis (PCA) on the datasets. After that we determine the statistically significant PCs which we will then use to cluster the cells. For both young and aged muscle, all 20 of the PCs are significant based on the Jackstraw method of analysis implemented in Seurat.

We next use these PCs to cluster our cells. In young and aged muscle, clustering reveals 9 and 10 populations, respectively. We then use these clusters to run Non-linear dimensional reduction (tSNE) to visualize the clusters in 2D.

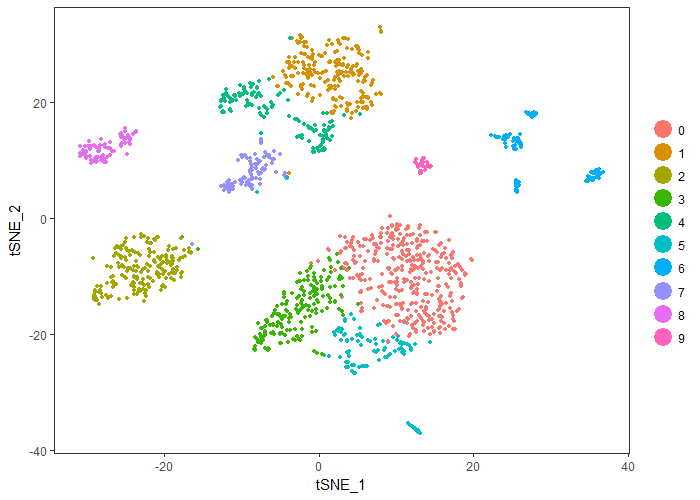
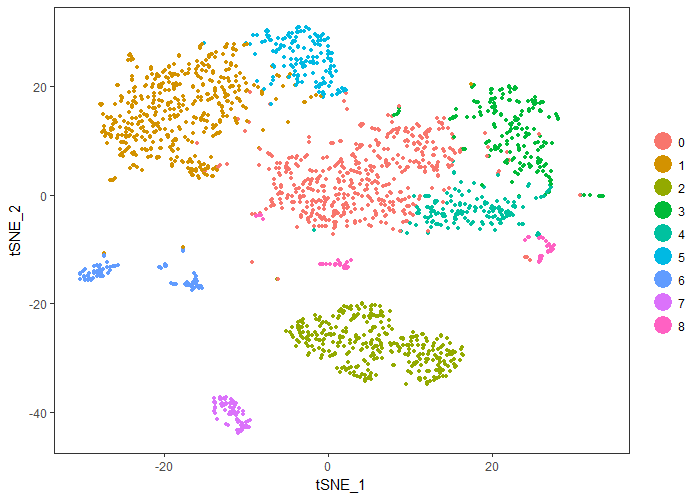


Figure – tSNE plot of young (left) and aged muscle (right). (700x500 pixels)

## Integrated analysis of Young and Aged muscle scRNA-seq;

For integrated analysis of young and aged, we use the Seurat object created for each of the young and aged muscle datasets and combine them together.

First, we take the top 1000 variable genes of each condition, combine it into a list, and select the ones that are expressed in both the aged and young muscle, and use them to perform a canonical correlation analysis (CCA) and combine the 2 datasets together.

Next, we determine the significant CCAs. Ours is ~20, and perform CCA alignment, followed by tSNE reduction and clustering. We use 0.6 as our resolution. And the result is as displayed below:

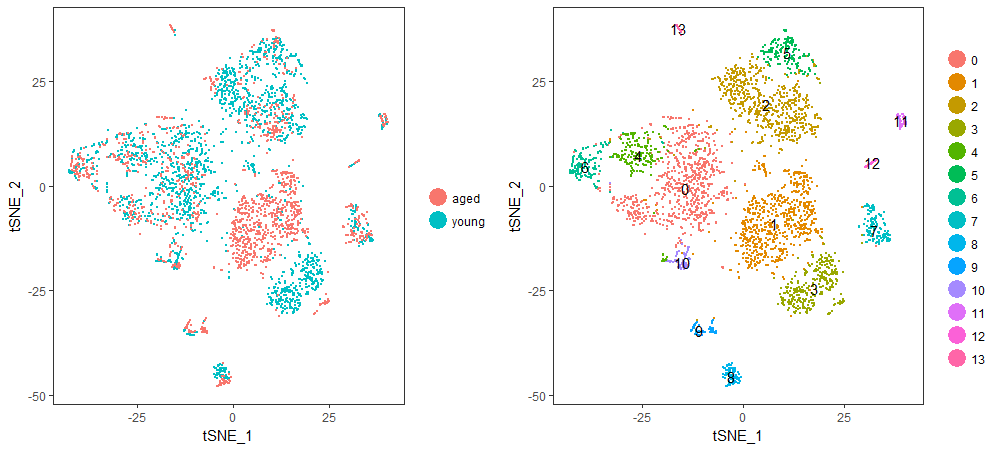


Figure 2 - tSNE plot of aged and young skeletal muscle analysis. (Image dimensions when exporting: 1000x450 (w x h)

Most of the created clusters have both aged and young cells. The major exceptions are clusters 1 and 3 in which the majority are aged and young respectively. Both clusters appear to correspond to myogenic cells.