Single Cell RNA Sequencing Worksheet1: Cell Ranger Count

Author: Chris Ozeroff, July 2023

FastQ Files —**Cell Ranger**→ Filtered Count Matrix

In this tutorial, you will be taking a single cell RNA-sequencing dataset and running it through the Cell Ranger pipeline. Cell Ranger performs alignment, filtering, and unique molecular identifier and barcode counting. It then outputs several files, including a count matrix which we can then analyze in R using software called Seurat.

*Because this can take a while to run, you will get Cell Ranger running and then be given a finished count matrix to analyze in Seurat.

1. On the AWS, mkdir a directory called day8. cd to day8. Make a directory called e_and_o inside of the day8 directory.

2. You will edit a sbatch script called cellrangerCount_sbatch. Git pull the script from github from the day8 scripts.

3. Open cellrangerCount_sbatch in vim. You will need to edit the error and output file path, the path to the transcriptome directory, and the path to the fastq directory. You will also need to set the ntasks AND local cores equal to 8.

4. Path to transcriptome directory:

```
--transcriptome=/scratch/Shares/public/sread2023/data_files/day8/refdata-gex-GRCh38-2020-A \
```

5. Path to the fastq's directory:

```
--fastqs=/scratch/Shares/public/sread2023/data_files/day8/10k_PBMC_3p_nextgem_Chromium_X_fastqs \
```

   Note: Many single cell sequencing data sets are publicly available from 10X genomics. Information about this particular data set can be found at: https://www.10xgenomics.com/resources/datasets/10k-human-pbmcs-3-v3-1-chromium-x-with-intronic-reads-3-1-high . These fastq's were downloaded ahead of time using the following curl command from by the 10x website (you do not need to do this step):

```
-bash-4.2$ curl -O https://s3-us-west-2.amazonaws.com/10x.files/samples/cell-exp/6.1.2/10k_PBMC_3p_nextgem_Chromium_X_intron/10k_PBMC_3p_nextgem_Chromium_X_intron_fastqs.tar
  % Total    % Received % Xferd  Average Speed   Time    Time     Time  Current
                                 Dload  Upload   Total   Spent    Left  Speed
  2 40.9G    2 1045M    0     0  15.8M      0  0:44:06  0:01:05  0:43:01 13.7M
```

6. Now, run the sbatch script

```
-bash-4.2$ sbatch cellrangerCount_sbatch
```

7. Check and see if the job is running for a minute or so (no errors)

```
-bash-4.2$  squeue -u ozeroff
```

8. Cancel your job after it has been running for a minute or so, and move onto next worksheet

```
-bash-4.2$ scancel YOURJOBID
```