Late May – June 2nd

* Complete several accredited R courses on Dataquest
* Living Github of all code and files created during internship
* Initial scRNASeq vignettes with Seurat
* Initial testing and familiarity with SingleR
  + <https://bioconductor.org/packages/release/bioc/vignettes/SingleR/inst/doc/SingleR.html>
* Familiarity with SingleCellExperiment data type
  + Conversion between datatypes (Seurat <-> SingleCellExperiment)
  + <https://satijalab.org/seurat/archive/v3.1/conversion_vignette.html>

Week of June 5th

* Continued probability and stats studies: Bayes Thm & Naïve Bayes Algorithm
  + Competent understanding of fundamentals, notes put into word file
* Continued exploration of SingleR and SingleCellExperiment data structure
  + Completed guided walkthroughs with commented code for explanations of each line; uploaded to Github
* Tabula sapiens/Tabula muris familiarity, begin processing for easy access in SingleCellExperiment structure (Step 3 in SingleR vignette)
  + Downloaded pancreas code (to test against references)
* Start writing code to run SingleR testing of Tabula Muris/Sapiens datasets; Likely using references from “scRNAseq” package; Example run: Test Muraro pancreas data from “scRNAseq” package in R against Tabula Sapiens pancreas reference data.
  + Code is written for comparisons and making graphs but the ReadH5AD function is not reading in data properly

Week of June 12th

* Debugging code and reformatting reference and test pancreas datasets for SingleR comparison
* Create repository in external Biowulf supercomputer using Terminal

Week of June 19th

* Use ensembl IDs from Biomart to remap genes with more standard identifiers
* First SingleR comparison with label output complete
* Heatmap, delta distributions, and violin plots created with labels and scores
* Conversion back to Seurat object; UMAP analysis

June 26th – July 14th

* Mitochondrial percentage QC
* Created dimension plots of Tabula Sapiens vs Muraro pancreas SingleR comparisons
* Created heatmaps, delta distributions, dimension plots of Tabula Sapiens vs Grun pancreas comparisons

Week of July 17th

* Utilize Biowulf supercomputer for larger SingleR comparison jobs
* Develop interactive R Markdown file for user-input comparison (beginning of Universal Pipeline)

Week of July 24th

* Wrote first bash file to run future Biowulf jobs in background
  + Successful Zilionis lung data comparison (with Tabula Sapiens)
* Successful Tabula Muris vs Wu kidney data comparison
  + Created labels, heatmap, dimension plot
* Successful Zilinois lung data comparison (vs Tabula Muris)

Week of July 31st

* Created first swarm file to process and clean downloaded data all within Terminal
* Created repository of full Tabula Sapiens and Tabula Muris data converted and processed into SingleR-ready references (.rds)

Week of August 7th

* Created \_\_\_filetype\_\_\_ to batch process SingleR comparisons
  + Takes in user input of reference and test files, including external or internal
  + Data is cleaned and annotated without further user input
  + Annotations added to experimental object
  + Generates annotation report in HTML format
* Created master list of unique genes for Tabula Sapiens and Tabula Muris by each cell type (104 TM; 480 TS)
* Working on pipeline involving bash call -> Rscript -> RMarkdown call