## Getting the Data

1. After sequencing is complete go to [sbgenomics](https://accounts.sbgenomics.com/auth/login?next=https%3A%2F%2Faccounts.sbgenomics.com%2Foauth2%2Fauthorization%3Fresponse_type%3Dcode%26client_id%3D74735b722c33498cb981b6b6e24f659f%26redirect_uri%3Dhttps%253A%252F%252Figor.sbgenomics.com%252Foauth2%252Fredirect%26state%3DUBJBxms6P86YCEdhWPVUac4cLCfOZk%26client_next%3Dhttps%253A%252F%252Figor.sbgenomics.com%252F%26scope%3Dopenid%26nonce%3D59808250767331284331701886364) website to
   1. I believe Ji will run the pipeline for you’re seq run.
   2. Make sure to BD Rhapsody™ Sequence Analysis Pipeline v >= 2.0 (Revision >= 12).
2. Make a working directory and **download the following files** in the **newly created directory.**
   1. Sample Tag Data
      1. \*\_Sample\_Tag\_Calls.csv
      2. \*\_Sample\_Tag\_ReadsPerCell.csv
      3. \*\_Sample\_Tag\_Metrics.csv
      4. A close-up of a cell

         Description automatically generated
   2. Raw Data (contains barcocde.tsv, matrix.mtx, and features.tsv)
      1. \*\_RSEC\_MolsPerCell\_MEX.zip

A group of blue text

Description automatically generated

* 1. Pipeline Output (QC’d from Seven Bridges)
     1. \*\_Seruat.rds
     2. \*.h5ad

A screenshot of a cell phone

Description automatically generated

* 1. V(D)J Data
     1. All files in URL directory

A screenshot of a computer

Description automatically generated

## Downloading and Setting Up Pipeline for Analysis

1. Install GitHub and Create an Account
   1. [On Mac](https://git-scm.com/download/mac)
   2. [On Windows](https://git-scm.com/download/win)

## Install R and RStudio

* https://posit.co/download/rstudio-desktop/

## Install Conda

* <https://conda.io/projects/conda/en/latest/user-guide/install/index.html>

## Install Packages

conda env create -f environment.yaml

## Run Jupyter and R Notebook in Numbered Order

* The jupyter files are numbered 1 – 5 in the order that they should be run.
  + Note: You may have to also run R codes as well for doublet detection, finding highly variable genes and for predicting cell types using Azimuth.

**Doublet Finder is Optional**

* If you run run dbltFinder.Rmd, then you’ll have to run 3a to remove detected doublets, otherwise you can skip to 3b.

**R Programs**

* dbltFinder.Rmd (run before using 3a)
* azPredicition.Rmd (run during 3b analysis to predict cell type)
* devianceFeatureSelection.Rmd (run during 3b to select highly variable genes)