To generate new data in Atom on local computer:

Note:

* Data generated on cluster is under folder: …/dataOutputForPaper/Summary1\_10000/
* Data generated on computer will be in …/dataOutputForPaper/ (or another folder under this that specify for certain testing)

In Atom with src as project folder:

1. Data Generation:
   1. First set constantsForTesting.jl with RANDSEEDNUM = population seed want to run (need to restart julia in between different pop seed data generation)
   2. Then go to and run: **codeForDatasetGenerationForPaper.jl**
   3. Change constantsForTesting.jl RANDSEEDNUM to new popseed exit and restart julia and then go back to step b
2. Generate Measurements: run **generateMeasurementsForAllData.jl**
   1. \*All pop seed folders should have been created with the same SAVEDATA (generations where saving data) so can analyze correctly at the 5th step.
3. Measurement Data Collection: run **measurementDataCollection.jl** 
   1. \*Make sure line 20 is commented out and line 21 is uncommented (so you are using constants file for testing and not for the cluster) in **measureDataCollectionForCluster.jl**
4. Collected Measurement Data Concatenation: run **concatenateCollectedMeasurementData.jl**
5. Correct misnamed columns for 7 of the csv files by running: **fixConcatenatedCsvFiles.jl**
   1. Comment out line 15 and uncomment line 16 if are running on data generated on your computer instead of the cluster
6. Data Analysis of Concatenated and Corrected Data: run **dataAnalysisForPaper.jl**
   1. Functions written in **dataAnalysisFunctions.jl**
   2. User inputs:
      1. Constants file to use (either constantsUsedForDataGeneration.jl or constantsForTesting.jl 🡪 use constantsForTesting if analyzing data generated on personal computer, and constantsUsedForDataGeneration.jl when analyzing data generated on the cluster (entire dataset 280,000 different population jld files) 🡪 so comment out line 13 and uncomment line 14 if running analysis generated on computer; other way around if running analysis generated on cluster
      2. dataOutputFolderName (use “Summary1\_10000” when analyzing entire dataset generated on the cluster, and use just dataOutputForPaper folder or corresponding folder where data generated on own personal computer was stored). 🡪 so comment out line 50 and uncomment 51 if running with data generated on your computer, and other way around for data generated on the cluster
   3. Note: Comment out lines 68, 114, and 145 when running data analysis on entire dataset generated on cluster because this is the measurement of founder genes that are repressed (gene expression less than gene threshold) that are also repressed by polycomb mechanism for each generation taking measurements during evolution and do NOT have these results for entire dataset
7. Plot PCA results for specific population seed, evolutionary trajectory, and generation using **plotPcaForSinglePopAndEvolTraj.jl** (functions written in pcaPlotFunctions.jl)