Documentation for Using the New Pipeline for Analysis

Step 0: Make sure you have the following files in the same folder:

Python files

* Analyzer.py
* v12-q50.py
* merger.py
* zeroscreator.py

Text Files

* genomeADalts.txt
* genomeADposns.txt
* genomewithinst.txt

Matlab Files

* Matlab\_commands\_3\_16\_GenomADupdate.m
* Matlab\_commands\_3\_16\_GenomADupdate.m~
* matlab\_input\_genomeADcoord.mat

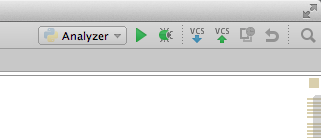
Other files

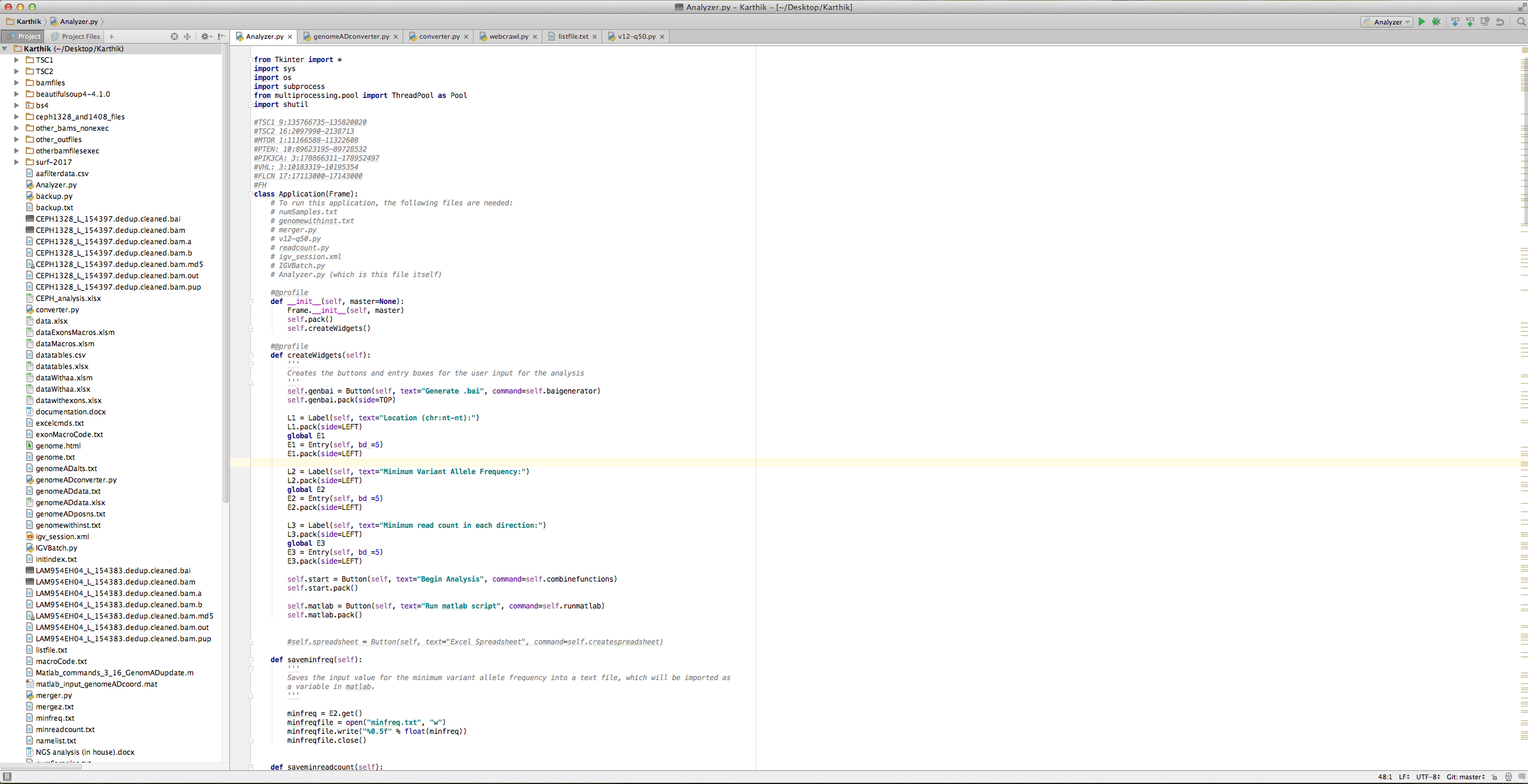
* dataWithaa.xlsm
* All .bam and .bai and .md5 files for the samples being analyzed

Now, open the Analyzer.py file using the program PyCharm CE.

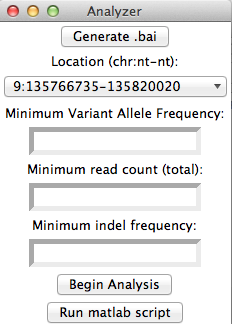
Step 1: Click the “run” button in the PyCharm Window. (Alternatively, type

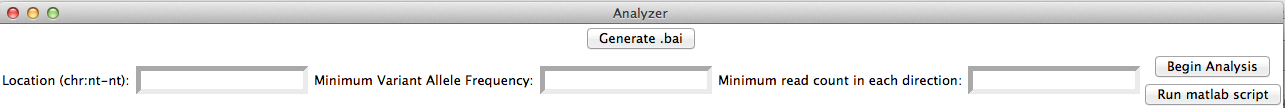
“python Analyzer.py” in the Terminal.



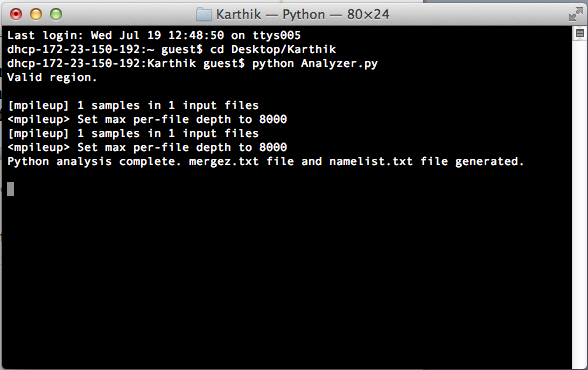


Step 2: Now, the following window should appear. Type in the location of interest for the analysis (in the form chr:nt-nt), the minimum variant allele frequency (as a decimal), and the minimum read count in each direction in the fields shown below.

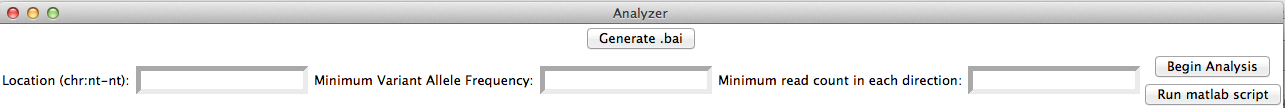


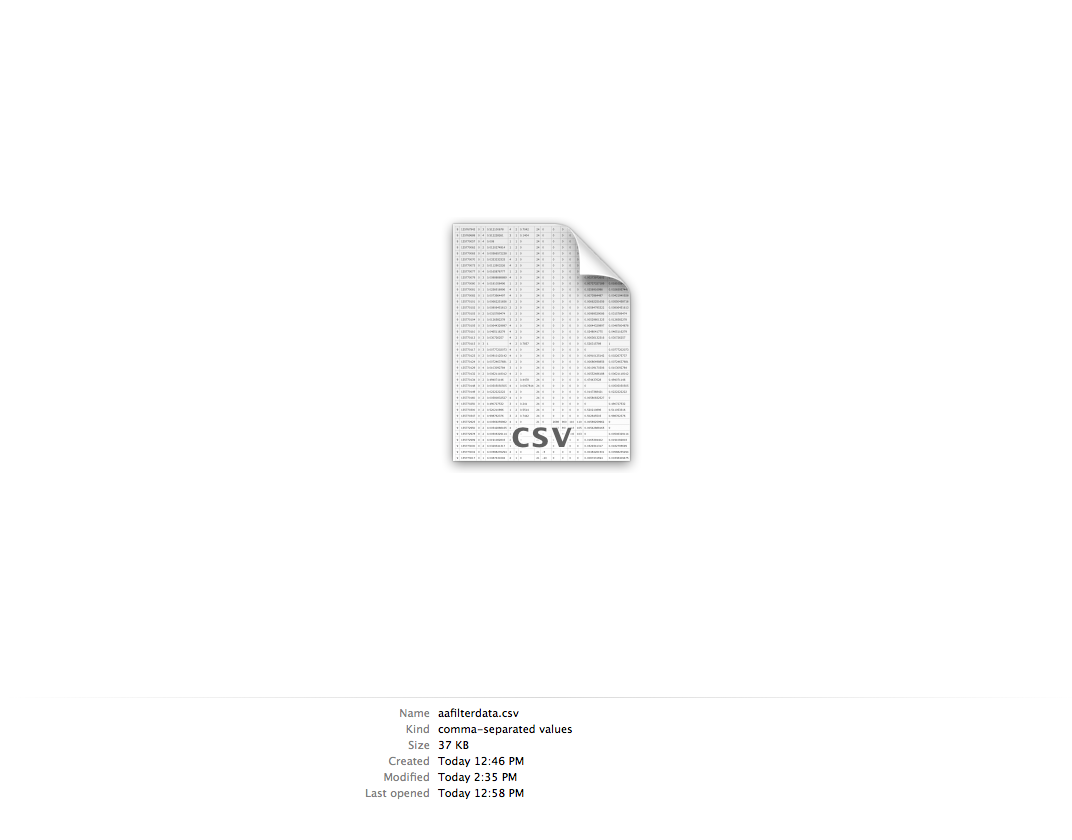


Step 3: Click “Begin Analysis.” Now the mergez.txt and namelist.txt files are being generated. When these are generated, the following will be displayed:



Step 4: Now, click “Run matlab script” in the previous window. Then, when the Matlab script has completed, then the filtered data of interest are in the files “aafilterdata.csv” and “formatfilterindels.csv”, which can be opened in Excel.





Step 5: Open the file “aafilterdata.csv” in Excel and click on the tab “Developer.” Then select “Macros” and select “dataWithaa.xlsm!FormatAA.” Then, click “Run.” The resulting file should be formatted as desired.