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
* nt\_coding.txt (do not include for TSC1/TSC2)
* genomeADfreqs.txt (do not include for TSC1/TSC2)
* exonCoordData.txt (do not include for TSC1/TSC2)

Matlab Files

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

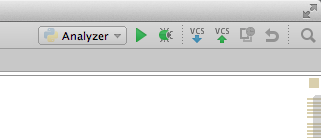
Other files

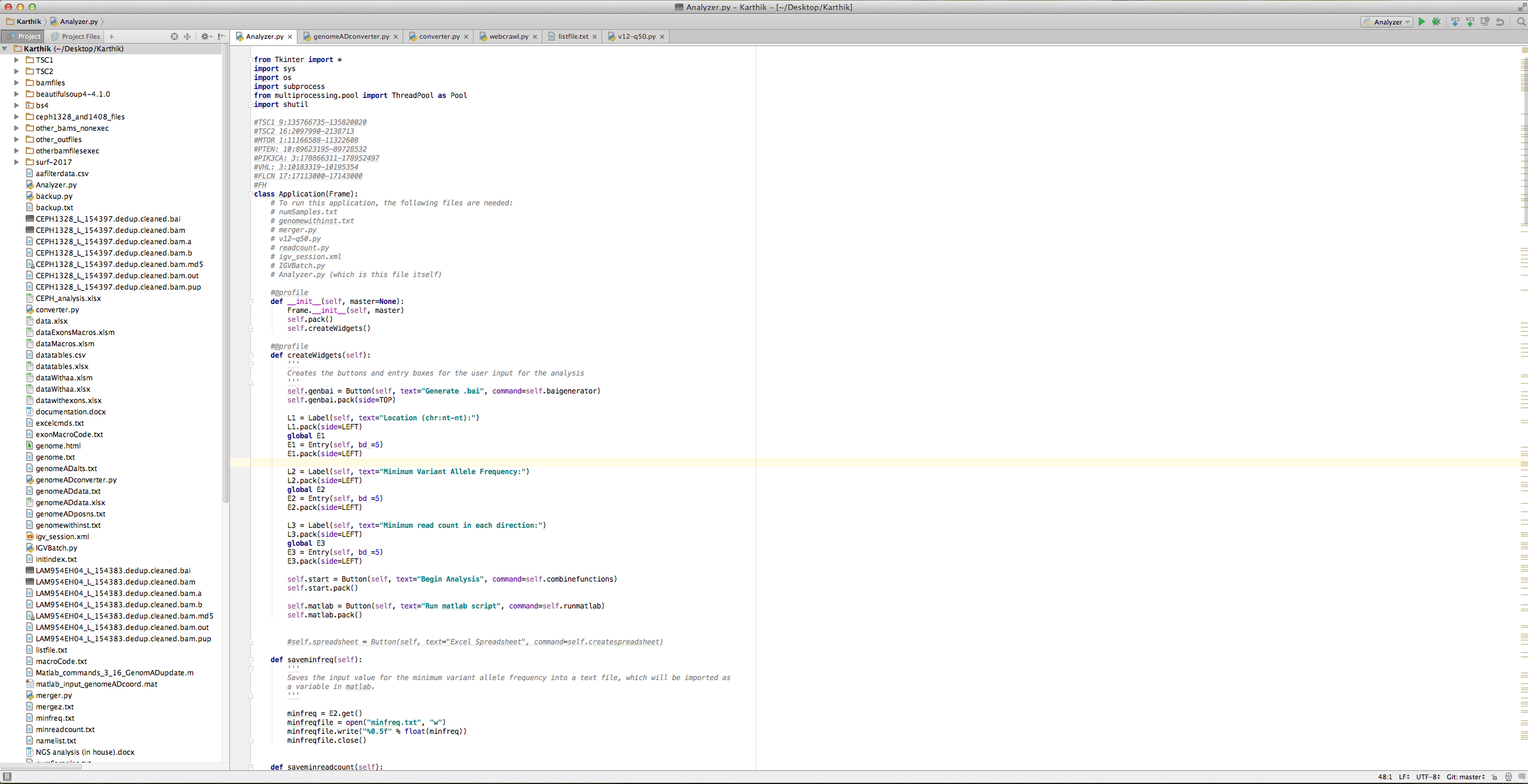
* dataWithaa.xlsm
* indelmacro.xlsm
* All .bam and .bai and .md5 files for the samples being analyzed

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

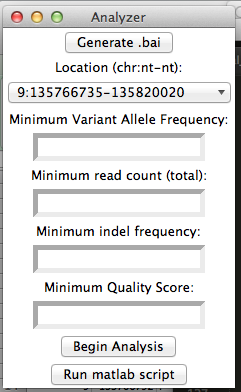
Step 1: Click the “run” button in the PyCharm Window. (Alternatively, navigate to the location of the Analyzer.py file for the gene in this analysis in Terminal and type

“python Analyzer.py.”

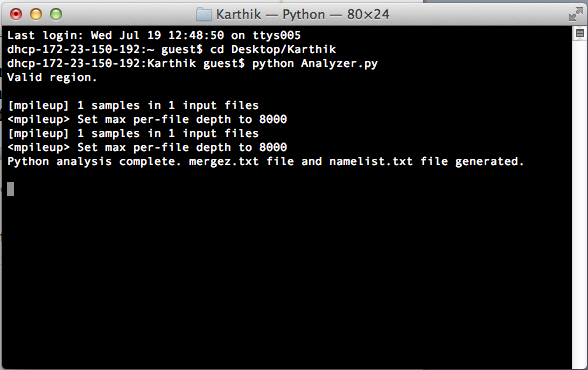




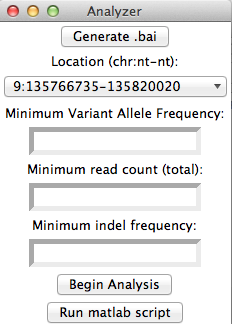
Step 2: Now, the following window should appear. Type in the location of interest for the analysis (in the form chr:nt-nt), the minimum variant allele frequency (as a decimal), and the minimum read count in each direction in the fields shown below. You can leave the Minimum Quality Score section blank if you don’t know what to put as a value here. Currently, the default is 46. Subtracting 33 from the quality score given here will give you the Phred quality score seen in IGV.



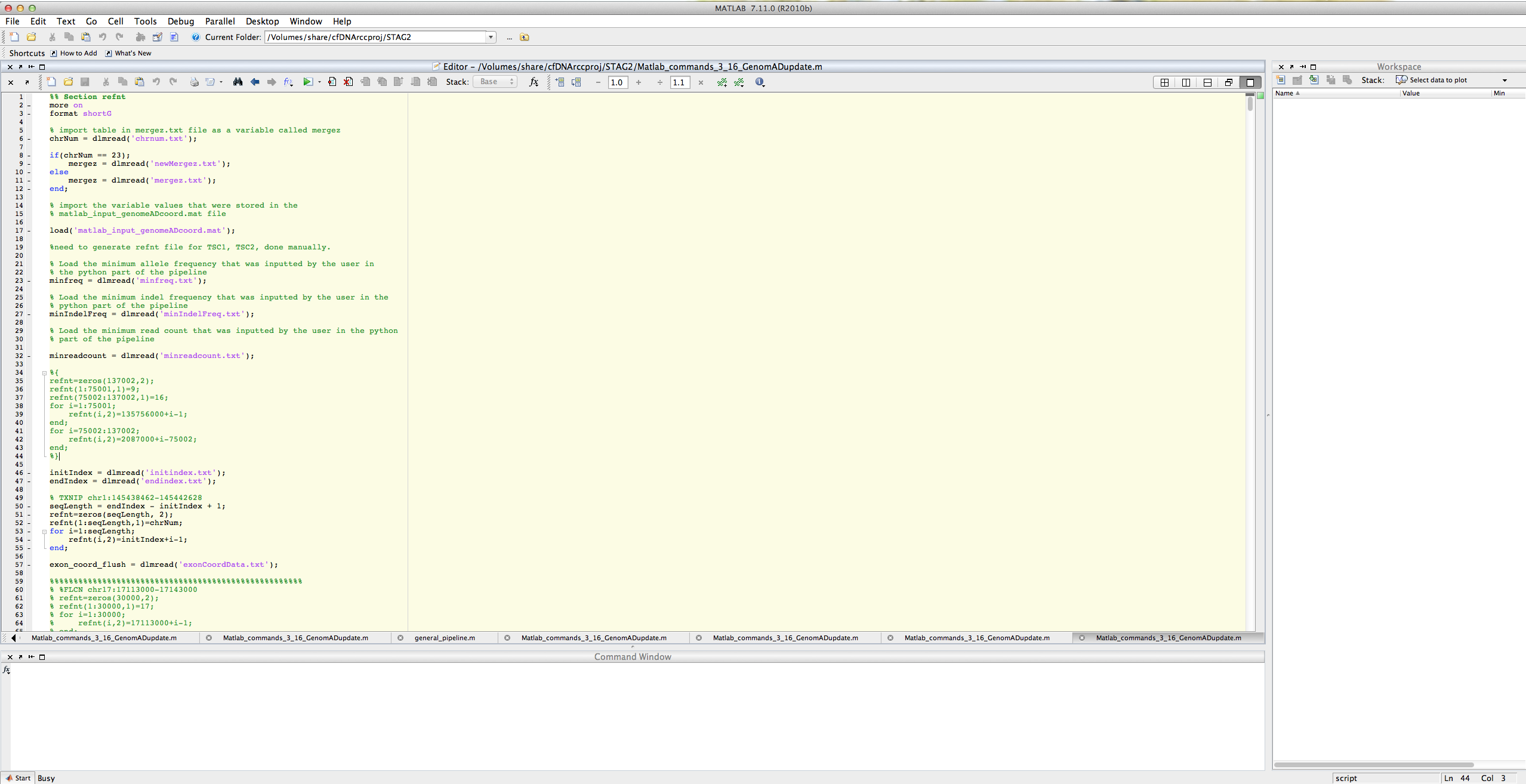
Step 3: Click “Begin Analysis.” Now the mergez.txt and namelist.txt files are being generated. When these are generated, the following will be displayed:



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 “snvdata.csv” and “formatfilterindels.csv”, which can be opened in Excel.



Step 4 (Alternate): Open the file “Matlab\_commands\_3\_16\_GenomADupdate.m” in Matlab. Then, click the “Run” button at the top of the screen. Then, click “Change Folder” if a popup comes up after clicking the run button. Wait until the bottom left-hand corner of the Matlab window does not say “Busy.”



Step 5: Open the file “snvdata.csv” in Excel. Create a new sheet in this workbook and title this sheet “aa codes.” Copy the data in the “aa codes” sheet from the dataWithaa.xlsm file and paste it in the same sheet in this snvdata.csv file. Then, click on the tab “Developer.” Select “Macros” and select “dataWithaa.xlsm!FormatAA.” Then, click “Run.” The resulting file should be formatted as desired. Now, save this file with the name “snvdata” as an xlsx file (Excel workbook).

Step 6: Open the file “formatfilterindels.csv” in Excel and click on the tab “Developer.” Then select “Macros” and select “indelmacro.xlsm!formatIndels.” Then, click “Run.” The resulting file should be formatted as desired. Now, save this file with the name “indeldata” as an xlsx file (Excel workbook).

Note: The order of the samples as displayed in the snvdata and formatfilterindels files is the same as the order in the namelist.txt file.