

## How to update CpG Project

### 1<sup>st</sup> update excel sheet, Final\_CpG\_list.csv with virus info

Location: data/list/Final\_CpG\_list.csv

Need: exact name of file, place file is located, number of sequences, number of nucleotides and a presentable name (nice name)

Objective: New/update virus info or virus file

### 2<sup>nd</sup> Run connect.R file, makes the csvs by using the rscripts from the bioinformatics class

Location: R\_scripts/Basic\_Bioinfo\_Rscripts/connect.R

Need: Final\_CpG\_list.csv

Objective: create csv files with information about frequency, wild type nucleotide, wild type & mutated amino acid (AA), big AA change, if a CpG site was made and the type of mutation

Output: Csv files placed in data/*place where your data is from*/data\_used/Csv

### 3<sup>rd</sup>/4<sup>th</sup> Frequency graphs & Wilcox Tables, run MakesTables\_Graphs\_other.R

Location: R\_scripts/graphs/MakesTables\_Graphs\_other.R

Need: Final\_CpG\_list.csv, Csv files

Objective: create csv files with information about frequency, wild type nucleotide, wild type & mutated amino acid (AA), big AA change, if a CpG site was made and the type of mutation

Output: graphs placed in output/WilcoxTables or output/M\_frequency\_graphs

### 5<sup>th</sup> Costly graph, Costly\_Graph.R,

Location: R\_scripts/graphs/Costly\_Graph.R

Need: Final\_CpG\_list.csv, Csv files

Change: files for viruses (if new files are made or name change in file), my.list, name.list (must match nice name), rename file png("change the date"), update the colored box numbers, and ablines as viruses are added or taken out, potentially update limits and added text. Repeat number changes for second figure

Output: costly graph in output/All\_Data/Costly

### 6<sup>th</sup> Update CpG\_Tally\_923.csv

Location: output/All\_Data/Data\_wilcox\_tally\_results/CpG\_Tally\_923.csv

Need: Look at wilcox table results input 1 for significant <0.05 and 0 for insignificant

Output: Updated tally file

### 7<sup>th</sup> Re-run the points graph, all\_data\_points.R

Location: R\_scripts/graphs/all\_data\_points.R

Need: updated CpG\_Tally\_923.csv

Output: graph of # data points and significance, located in

output/All\_Data/Num\_of\_data\_points/ & basic % statistics of what is costly