Verify Quality of Raw RNA-Seq using FASTQC (galaxy)
-galaxy tool for Quality Control
-Input: fastasanger.gz
-Output: .html of report and .txt of raw output

Align RNA-seq to genome using Bowtie2 (galaxy)
-Input: Toplevel FASTA, Toplevel .gff3
-Output: genome.*.bt2 File

Create VCFs using Native Variant Caller (galaxy)
-Input: Toplevel FASTA, .genome.*.bt2
-Output: NVC.vcf file (approx. 12GB)

Download VCFs, .gff3 from Galaxy to machine. The rest of the steps are in Python

Run initial Setup to establish File Structure (optional)
-based on user input; can be read in through a file, User input, OR hardcoded test data

User places downloaded Files
-(May be automated once I better understand galaxy)

NVC files Filtered
-using vcf_filter1()
-Outputs a filtered file (.txt), approx 1GB where only genes with Ref and Alt A/G counts are added
-Copies the line of information from the .vcf to the new filteredNVC.txt with A/G counts appended

Gff3 File Filtered
-using gff3_filter()
--input: Toplevel .gff3 file
-Trims down initial .gff3 file to only have lines containing genes
-Reduces to approximately 1/10 former size
-outpute: filtered.gff3

Genes in NVC matched to gene information in .gff3. [SLOWEST STEP]
-using findgene()
--input: filtered.gff3, filteredNVC.txt
-output: combined information of .gff3 gene information for the NVC gene (*_genes.txt)
-Removes Low quality counts, if: (Less than 10 reads, no ref or no alt reads, no counts for A)

Intersect genes.txt files from previous step
-intersects output files of findgene() outputs, creating merged files
-applied to all included files, and is run on its own outputs
-resulting file (comp_*.txt) contains only genes with >10 reads in in all experimental conditions
-can be run on subsets of the data as well.
-generates several intermediary intersection files

Make intermediate pre-excel file
-processes outputs of previous step (comp_*.txt) to create a pre-excel (pxl*.txt) file
-applied only to final output comparison files

Output Results as a CSV
-make_csv()
-input (pxl*.txt) files form previous step
-translates pre-excel file into a .csv
-

Analysis of Data and Statistics