This submitter and subscripts are used for joint calling exome projects.

The MASTER\_SUBMITTER.sh takes in the following arguements:

* PROJECT - The name of the project directory found in Seq\_Proj (no / marks)
* SAMPLE\_SHEET - The sample sheet with the samples to be joint called (fully qualified path)
* PREFIX - The prefix of the multi-sample vcf files that will be generated
* NUMBER\_OF\_BED\_FILES - The number of bed files to split on. If left blank, the default will be 100

\*\*If these scripts are used you must change the SCRIPT\_DIR on line 14 in the MASTER\_SUBMITTER.sh to where the SCRIPTS folder lives.\*\*

The setup splits based on bed file as opposed to chromosome. This allows for each step to run more efficiently and take up less time per job slot. Utilizing a larger number of job submissions that run faster allows us to use the cluster more efficiently.

* The bait bed file used to split on is found from the sample sheet and is checked for formatting and to ensure the file is sorted properly. Then this bed file is split and stored in the TEMP/BED\_FILE\_SPLIT directory.
* When submitted, a gvcf.list file is generated for all the samples found within the sample sheet. The sample sheet data is used to avoid any samples found in the GVCF directory that is not being released. This gvcf list will be stored in the top directory as the ##.samples.gvcf.list
* The mini-bed files are used to create small jobs across three steps (CombineGvcfs, GenotypeGvcfs, and VariantAnnotator)
* After they are combined using CatVariants, the pipeline proceeds as it did before
* TO NOTE:
  + Commands are written out to the top project directory in a file named command\_lines.txt