This submitter and subscripts are used for joint calling the fiasco that is mendel project.

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

* PROJECT - The name of the project where the multi-sample vcf is being written to (for now it’s M\_Valle\_MendelianDisorders\_SeqWholeExome\_120511\_GATK\_3\_3-0)
* 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 variable on line 15 in the MASTER\_SUBMITTER\_MENDEL.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 here (/isilon/sequencing/Seq\_Proj/M\_Valle\_MendelianDisorders\_SeqWholeExome\_120511\_GATK\_3\_3-0/BED\_Files/BAITS\_Merged\_S03723314\_S06588914.bed)
  + This is a merged bait bed file to capture all of the regions for both bed files that have been used within the mendel project.
* When submitted, a gvcf.list file is generated from all the samples found within the sample sheet and the most previously edited gvcf list found in the $PROJECT directory. The sample sheet data is used to avoid any samples found in the GVCF directory that is not being released. This new gvcf list will be stored in the top directory as the ####.samples.ReSeq.JH2027.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
    - One for the multi-sample vcf directory
    - One for the sample directory for sample specific commands
  + We have been running 500 bed files
  + Annovar is now ran automatically on each sample using -pe slots 5
  + The VCF files and lists that Hua wants for Variant Summary Stat info are generated for her
    - They will be written out to here: /isilon/sequencing/Seq\_Proj/M\_Valle\_MendelianDisorders\_SeqWholeExome\_120511\_GATK\_3\_3-0/MULTI\_SAMPLE/VARIANT\_SUMMARY\_STAT\_VCF/
    - All the files will be written out with $PREFIX\* and the suffix will be denoted to what specific information/file it is.
    - She has been informed that these files are specifically for her so when her work is complete she can move them in a subdirectory to keep this VARIANT\_SUMMARY\_STAT\_VCF directory clean…
      * Either way there won’t be an issue overwriting unless you run the same project with the same $PREFIX variable.