How to run:

compressor\_for\_rnd.sh (housed here: /isilon/sequencing/VITO/NEW\_GIT\_REPO/Archive\_Compressor\_2016/compressor\_for\_rnd.sh or on github)

* $1=Directory of project you want to compress
* $2=The known reference sequence used for alignment and all other steps
  + If left blank, it will default to the standard /isilon/sequencing/GATK\_resource\_bundle/1.5/b37/human\_g1k\_v37\_decoy.fasta
* This script will parse through the directory provided for any file that has a vcf/csv/txt/bam/intervals ending and compress accordingly.
* During compression, the files aren’t removed until md5sums can be validated on the newly compressed file. Once this is complete the file is removed.
* Things to note:
  + Depending on where the directory lives is if the pipeline attempts to bin the quality scores when converting to cram. The only time this is completed is when the directory is under the 02\_CIDR\_RND subdirectory.
  + Hold ID for MD5 check at the end could be a little better/more stable
    - Right now, it depends on the bam files completing the cram step and validation. These SHOULD be the longest steps, but nothing ever seems to be a guarantee…
  + Generates a TON of log files… Maybe best to consolidate some of the gzip jobs into one log file.