These series of scripts/python code will download BAM files from the 1kgp, run qmotif on them (sequentially), estimate telomere motif counts and save them.

Sequence to run the files:  
1.) download\_files.sh (requires: files\_list.txt)

2.) hope2\_bblah.py

3.) stage2.py (requires: log files generated by hope2\_bblah.py)

4.) coverage.sh (requires: chrnames)

5.) totaltelomere.py (requires: stage2coverage)

1. download\_files.sh, along with files\_list.txt will download the BAM and its corresponding BAI index file from NCBI’s 1000 Genomes Project server (<https://www.ncbi.nlm.nih.gov/projects/faspftp/1000genomes/>)
2. Before running hope2\_bblah.py (will rename the file later), make sure qmotif is installed and invoke java with the qmotif tar file. You’ll also need to set the path to the tool-qmotif and where your input files- BAM and BAI are. (more info on this-<https://adamajava.readthedocs.io/en/latest/qmotif/qmotif_1_0/>)
3. stage2.py requires the log files generated by hope2\_bblah.py (check what working directory you’re in before proceeding.) This script will give you the chromosome number and the final telomere motif counts for each chromosome after stage 2 -> this output will be stored in “{sequence\_name}*\_*stage2\_coverage.txt file
4. coverage.sh is to get a tally of telomere motifs for each chromosome. This script uses “chrnames” file, make sure they’re in the same folder. This script creates only one output file “stage2coverage” which has all the info on all the files you ran.
5. totaltelomere.py gives total telomere motif counts for the sequence across all chromosomes. This uses the “stage2coverage” file we got as an output in the previous step and generates a file named “total\_telomere\_motif\_counts”