**Use of python to parse VCF file** (<http://www.internationalgenome.org/wiki/Analysis/vcf4.0/>) and subsequent filtering and statistics generation. This is an ongoing project where I am working on taking a variant call format file and then filtering variants based on attributes as well as generating summary statistics. It has been inspired by the work of Petr Danecek, Adam Auton et al. (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3137218/>) The objective of this is to use sorting and searching algorithms such as merge sort and binary search to sort and filter VCF file based on specific attributes and retrieve information to suit various needs. Subsequently there will be object oriented implementation of this tool

**Type: python3 vcfAnalyser.py -h**

Above command gives all possible options. As of now only one at a time can be used except for the boolean --keepChrom and --chrom together or the boolean --lociList and --keepLoci together

**Example Usage:**

Lets's say we want to filter by quality score. Something like this will work:

**python3 vcfAnalyser.py --vcf myvcf.vcf --qualScore 2500 --out keepingVarsWithQualAbove2500 where argument with out is the filename for the new vcf**

Or

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --chrom 1 --keepChrom False --out RemovingChr1Vars**

The above command will get rid of all chrom 1 variants and rewrite remaining variants to a new vcf.

Alternatively:

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --chrom 1 --keepChrom True --out KeepChr1Vars**

The above command will keep only chr1 variants and rewrite remaining variants to a new vcf.

Or

If there is a list of variants you want to retain in your vcf, then something like this should work:

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --lociList chrLociToExtractVar.txt --keepLoci True --out testingKeepingSelectedVariants**

The loci list should be tab delimited header less file with first column being chromosome number and second column being position on the chromosome. Or if you want to get rid of certain variants

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --keepLoci False --lociList variantRemovalLoci.txt --out testingRemovingVariants**

If you want keep variants in certain genomic bed regions something like this should work.

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --bedFile shortBed.txt --out fileWithVariantsForBed**

The bed file should be a header less file with three columns that are tab delimited with the first column being chromosome number, the second column being start position of the bed region and the third column being the end position for the bed region.

If you want to filter by quality score of variants, code of following nature should work:

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --qualScore 2500 --out keepingVarsWithQualAbove2500**

If you want to get transition count following should work,

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --getTs Ts --out transitionCountTest**

If you want to get transversion count, try the code below.

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --getTv Tv --out transversionCountTest**

If you want both transiton and transversion, following should work

**python3 vcfAnalyser.py --vcf sampleChr1Chr10.vcf --getTsTv TsTv --out tstvCount**

If you want to keep sites with a certain minimum DP, try:

**python3 vcfAnalyser.py –vcf sampleChr1Chr10.vcf –minDP 5700 –out myNew.vcf**