BEDtools:

- tools for variety of genome data manipulations, can be used on Galaxy or called via command line
- 'bedtools --help' brings up functions that can be used for data manip.
- Galaxy lists these under "Operate on Genomic Intervals" then BEDtools in the tool panel
- BEDtools intersection evaluates similarities and differences between two datasets
- BEDtools subtraction looks for non-overlapping pieces of genomic intervals

SAM(sequence alignment/map)

- file format containing sequence alignment information relative to a reference, commonly used to store results from NGS reads (through HISAT/BWA)
- tab-delimited with column headers (begin with @), each line of the file = single alignment event with 11 fields of information
    - CIGAR string: encodes variation found in the SAM file between the sequence and reference genome. M = alignment mismatch, I = insertion to the reference, D = deletion from reference
- **BAM**(binary SAM) most useful when sorted for fast retrieval and display, though needs indexing to be displayed(create .bai file). IGV looks for corresponding .bai files when .bam files are loaded.

SAMtools(v 1.11)

- installed on bfx server, used to call SNPsin genomicalignments.
- In Galaxy, SAM files are big, can convert SAM to BAM via "samtoolsview -bsfilename.sam> filename.bam"
- Sorting that new BAM file via "sort" function "samtools sortfilename.bamfilename.sorted"
    - filename.sorted will still be given .bam file extension
- Indexing sorted BAM file "samtools index filename.sorted.bam" => filename.sorted.bam.bai
    - Galaxy doesn't have an indexing tool as it is done automatically when visuals are requested for a .bam file