Session 2 Exercises

Using BWA to align Illumina reads to human chromosome 21. Call SNPs and short indels relative to the human reference sequence using the MAQ consensus model and the samtools software package. Visualize the alignment results in IGV.

Step 1. Log into the CAC Linux server. Create a session2 directory under your home directory on the CAC server. Copy all data files of the project to your session1 directory.

```
mkdir session2

cd session2

cp  /home/gfs08/jp86/ngw2010/session2/lecture1/*   ./
```

After you finish these steps, make sure you see the following files by typing "ls -l" followed by Enter key. "ls -l" command would give you the size of the file (in bytes), and last time it was modified.

chr21.fa (fasta file of the human chromosome 21)

na18507.chr21.fastq (fastq file of Illumina sequencing data)

run_bwa.sh (a script for submitting a job to run BWA)

Step 2. Submit a job to get run BWA.

- Modify the first five lines in the run_bwa.sh file.

  Modify the line "#PBS -A jp86_0004". Change "jp86_0004" to your own project name. For this project, v4 cluster (16G RAM) would work. Sometimes, the queue for the v4 cluster is very long, you might want to use v4-64g instead.

- Submit the job by typing this command line followed by Enter.

  ```
  nsub run_bwa.sh
  ```

  The job would take a few hours to finish. You can monitor the progress by "qstat" command. After the job is finished, you will see three new files created under the session2 directory: na18507.chr21.sorted.bam, na18507.chr21.sorted.bam.bai and na18507.chr21.SNP.pileup. Copy the three files to your local computer.

Step3. Use IGV to visualize the BAM file.

- Launch the IGV software. Go to the web site: http://www.broadinstitute.org/igv/ . First time users of IGV need to register, and a direct link to the software will be sent to you by email. Click one of the "Launch" buttons, depending on what computer you are using. If you are not sure, the first "Launch" button works for most computers. After the software started, you need to choose a genome. Click the genome drop-down list in the toolbar and select "Human hg19", then click the next drop-down list and select "chr21".

- Load the BAM file. From IGV menu, click "File"->"Load from file" and select the BAM file you just copied from the CAC server.

- Use the zoom slider in the upper right corner to zoom in. When zoomed in to the alignment read visibility threshold (by default, 30 KB), IGV shows the reads. The tutorial for IGV can be found at the software web site.

- Look at some of the variants in na18507.chr21.SNP.pileup relative to the human reference genome. Do these look reasonable?