

MBB 291

Problem Set 5

Due: 11 November 2017 11:59pm

General Instructions

- The answer sheet for this problem set should be submitted as a PDF file. You may use any word processing software to create the answer sheet. The name of the PDF file to be submitted should follow the following format: [MBB 291] < *Student Number* > – < *Last Name, First Name* > – Problem Set 5.pdf. For example: [MBB 291] 190800001 - De la Cruz, Juan - Problem Set 5.pdf.
- If you have consulted references (books, journal articles, online materials, other people), cite them as footnotes to the specific item where you used the resource/s as reference.
- Submission of the problem set answers should be done via e-mail. Attach the .pdf file, and write as the subject header of the e-mail: [MBB 291] < *Student Number* > – < *Last Name, First Name* > – Problem Set 5. For example: [MBB 291] 190800001 - De la Cruz, Juan - Problem Set 5. Send your answers to janmichaelyap@gmail.com.
- **You should receive a confirmation e-mail from me stating receipt of your deliverable within 24 hours upon your submission of the problem set.** If you have not received any, forward your previous submission using the same subject header once more.
- If you have any questions regarding an item (EXCEPT the answer and solution) in the problem set, do not hesitate to e-mail me to ask them. However, **questions regarding this problem set forwarded/received on or after 12:01am of 08 November 2017 will NOT be entertained.**

Questions

Below are the commands that we issued to perform SNP calling. Explain what each command is supposed to do, and describe what each parameter (e.g. `-a bwtsv` in the first command) is supposed to do.

- 1 `bwa index -a bwtsv -p ./data/ref.fa ./data/ref.fa`
- 2 `bwa mem -t 2 ./data/ref.fa ./data/reads1.fq ./data/reads2.fq > ./outputs/alignment.sam`
- 3 `picard SortSam INPUT=./outputs/alignment.sam OUTPUT=./outputs/alignment.sorted.sam SO=coordinate VALIDATION_STRINGENCY=SILENT`
- 4 `picard SamFormatConverter INPUT=./outputs/alignment.sorted.sam OUTPUT=./outputs/alignment.sorted.bam`
- 5 `picard MarkDuplicates INPUT=./outputs/alignment.sorted.bam OUTPUT=./outputs/alignment.dedup.bam METRICS_FILE=./outputs/mdup.metrics`
- 6 `picard AddOrReplaceReadGroups INPUT=./outputs/alignment.dedup.bam OUTPUT=./outputs/alignment.rg.bam SORT_ORDER=coordinate RGID=ref RGLB=ref RGPL=Illumina RGPU=ref RGSM=ref VALIDATION_STRINGENCY=SILENT CREATE_INDEX=True`

```
7  picard CreateSequenceDictionary REFERENCE=./data/ref.fa OUTPUT=./data/ref2.dict
8  samtools faidx ./data/ref.fa
9  GenomeAnalysisTK -T RealignerTargetCreator -R ./data/ref.fa -I ./outputs/alignment.rg.bam
   -o ./outputs/alignment.rg.bam.intervals -nt 4
10 GenomeAnalysisTK -T IndelRealigner -R ./data/ref.fa -I ./outputs/alignment.rg.bam
   -targetIntervals ./outputs/alignment.rg.bam.intervals -o ./outputs/alignment.realigned.bam
11 GenomeAnalysisTK -T HaplotypeCaller -R ./data/ref.fa -I ./outputs/alignment.realigned.bam
   -o ./outputs/snps.vcf -nct 4 -mbq 20 -out_mode EMIT_VARIANTS_ONLY
   -stand_call_conf 20 -stand_emit_conf 20
```