

## 1 Variant Calling - Solutions

No exercises in this section.

## 2 Performing variant calling

Q1 66 reads

Q2 The reference allele is A and the alternate allele is G. The upper/lower case letters indicate the forward/reverse orientation of the read.

Q3 0 reads calling reference allele and 66 reads calling the alternate allele

Q Add the -v option to the command:

```
bcftools mpileup -f GRCh38_68.19.fa A_J.bam | bcftools call -mv | less -S
```

### 2.0.1 Exercises

Q1 The reference allele is A and the alternate allele is G.

Q2 Look up the tag DP in the INFO column: there were 69 raw reads at the position.

Q3 There are 0 reads calling the reference and 66 high-quality reads calling the alternate.

Q4 An indel. Five bases TGTGG were inserted after the T at position 10003649

## 3 Variant Filtering

Q1 The complete command is:

```
bcftools query -f'%POS %QUAL [%GT %AD] %REF %ALT\n' -i'QUAL>=30 && type="snp" && AD[*:1]>=25' out.vcf | head
```

Q2 The complete command is:

```
bcftools stats -i'QUAL>=30 && AD[*:1]>=25' out.vcf | grep TSTV | cut -f5
```

Q3 The complete command is:

```
bcftools stats -e'QUAL>=30 && AD[*:1]>=25' out.vcf | grep TSTV | cut -f5
```

Q4 The complete command is:

```
bcftools stats -i 'GT="het"' out.vcf | grep TSTV | cut -f5
```

Q5 The complete command is:

```
bcftools norm -f GRCh38_68.19.fa out.flt.vcf -o out.flt.norm.vcf
```

## 4 Calling Variants Across Multiple Samples

Q Use the commands:

```
bcftools mpileup -a AD -f GRCh38_68.19.fa *.bam -Ou | bcftools call -mv -Ob -o multi.bcf bcftools index multi.bcf
```

**Q** Use the commands

```
bcftools filter -s LowQual -i 'QUAL>=30 && AD[*:1]>=25' -g8 -G10 multi.bcf -Ob -o multi.filt.bcf bcftools index multi.filt.bcf
```

**Q1** Use the commands:

```
bcftools stats multi.filt.bcf | grep TSTV | cut -f5 (raw calls)
```

```
bcftools stats -i 'FILTER="PASS"' multi.filt.bcf | grep TSTV | cut -f5 (only filtered set)
```

**Q2** Use the commands:

```
bcftools stats -e 'FILTER="PASS"' multi.filt.bcf | grep TSTV | cut -f5
```

## 5 Visualising Alignments

**Q1** 75 in total, 39 on the forward and 36 on the reverse strand.

**Q2** Yes. Use the command:

```
bcftools view -H -r 19:10001946 multi.filt.bcf
```

**Q3** Yes.

**Q4** Yes. Use the command:

```
bcftools view -H -r 19:10072443 multi.filt.bcf
```

**Q5** No. It fails due to lowQual and snpGap, this means the call was removed by filtering because the quality of the call falls below the threshold set and the call is in close proximity to an indel.

**Q6** No. It is an alignment artefact, the aligner preferred two SNPs instead of a long deletion.

## 6 Variant annotation

**Q1** Use the command:

```
bcftools query -f '%BCSQ' -r 19:10088937 multi.filt.annot.bcf
```

to return

```
missense|Fads2|ENSMUST00000025567|protein_coding|-|163V>163I|10088937C>T(base)
```

**Q2** A missense mutation

**Q3** The C>T mutation changes the amino acid at position 163 in the protein sequence, from valine to isoleucine.