

## 1 Read Alignment

There are no questions in this section.

## 2 Performing Read Alignment

1. 145441459
2. The sam is  $\sim 157\text{M}$  and the bam is  $\sim 25\text{M}$
3. 67,461 (17.2%) or  $359 + (33551 * 2)$
4. 392,820
5. 391603/392820 (99.7%)
6. 389410
7. 0
8. 419 (mean) 113.9 (standard deviation)
9. 7,853 (2.0%)
10. First/Forward read

## 3 Alignment Visualisation

1. This exercise is just asking you to explore the genome and become familiar with navigating in IGV.
2. 23X-57X
- 3.

There is a 1bp insertion (at “T”) at position 87,483,966. This is supported by 9 reads.

There is a 28bp deletion at position 87,483,966. This is supported by 3 reads.

There is no third mutation here, apologies for the typo in the exercises.

The CRISPR-Cas9 has acted on the zygote at this locus to create Non-Homologous-End-Join-based damage around 87,483,960: that resulted in a subclonal 1bp insertion and a 28bp deletion.

- 4.

You are watching the zygote DNA-repair machinery panicking and grabbing at straws.

### 3.1 Alignment Workflows

1. -M marks shorter split hits as secondary and -R adds the read group to the header of the BAM file
2. -b means create a BAM as output and -S indicates that the input files is a SAM file. The -S option is now ignored by samtools as it can now autodetect the input file type.
3. 397506

4. 303036/397506 (76.2%)
5. 282478
6. 2239
7. 275.9 (mean) and 47.7 (standard deviation)
8. 23,789 (7.9%)
9. First
10. ~22M
11. 12399 or  $3115 + (4642 * 2)$
12. 2%

## 3.2 Exercises

1. No answer
2. The reference base is C
3. No (the reads call T)
4. The reference base is G and all reads agree
5. No answer
6. An insertion
7. No answer
8. A deletion. This is unlikely to be a true variant and may be due to misalignment due the run of T's in the flanking region.
9. The following command produces a cram file which should be ~29MB in size.

```
samtools view -C -T ../../../../ref/Saccharomyces_cerevisiae.R64-1-1.dna.toplevel.fa  
-o library1.markdup.cram library1.markdup.bam
```

-C means create a CRAM file as output

-T is the reference file to use for the compression

-o is the name of CRAM file to create