**Variant calling workshop: question sheet**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q1** | What is the mean coverage? | | | |
|  | | | |
| **Q2** | What percentage of target bases are covered by 15 or more reads? | | | |
|  | | | |
| **Q3** | What is the mean insert size and how does it compare with the Whole-Exome sequence data? | | | |
|  | | | |
| **Q4** | Use the pileup file to complete the table | | | |
| No. reference reads | | No. alternate reads | |
| Forward strand (.) | Reverse strand (,) | Forward strand (ACGTN) | Reverse strand (acgtn) |
|  |  |  |  |
| **Q5** | How many variants are called by GATK unified genotyper? | | | |
|  | | | |
| **Q6** | How many variants are called by Varscan? | | | |
|  | | | |
| **Q7** | Which variant caller (MPileup/bcftools, GATK and Varscan) has the highest true positive rate/sensitivity? Sensitivity = (true positive/[true positive + false negative]) | | | |
|  | | | |
| **Q8** | Which variant caller (MPileup/bcftools, GATK and Varscan) has the lowest percentage of false positives? False positive % = 100\*(false positive/[false positive + true positive]) | | | |
|  | | | |
| **Q9** | How does the rate of variation per bp, het:hom ratio and tiTvRatio compare with the expected genome wide values from Platinum Genomes (1 variant per 650bp, 1.6, and 2 respectively)? | | | |
|  | | | |
| **Q10** | Considering that the targeted region on chromosome 22 spans 1,183,396 bp what is the rate of variation and does it come close to the prediction from Platinum Genomes for coding regions (1 variant per 1400bp)? | | | |
|  | | | |
| **Q11** | Comment on the amount of variation that is present in dbSNP129 and how well it agree with expectation (~83% of variation is usually present in dbSNP version 129)? | | | |
|  | | | |
| **Q12** | What is the ratio of transitions to transversions and is it in line with the predicted value of 3? | | | |
|  | | | |