2nd example

2nd example: low clonality somatic variant

For this example, we extracted another 1000 base-region of a whole genome paired end sequencing experiment from the same sample pair as in the first example, but with higher coverage. This region contains only one variant that is called by the variant calling program. However, we have evidence (let's say from previous experiments or literature) that there may be another somatic variant in the region, which has lower than normal "heterozygous" frequency. This may happen in heterogeneous tumors where only some cells harbour the mutation.

There will be less tasks for this example than for the first one, we will focus on the pileup format of the region and the position of interest, and try a different approach to distinguish the variant position from the rest.

Files from Variant_calling/1_realigned:

- Open the Integrative Genomics Viewer

```
java -jar igv.jar
```

- Load the samplepair bam file
- Select the correct genome to be displayed: Human (b37)
- Select the region covered by the bam file: chr16:81077000-81078000
- Inspect the variant position: chr16:81077733 ; in which gene falls this position? Which bases are read in the control and tumor samples, respectively?
- Leave the IGV open with the bam file loaded

Files from Variant_calling/2_pileup:

- The files in this folder were generated running samtools mpileup on the whole genomic region covered by the bam files. Use them to control your own results. We will use the control.RG and tumor.RG files for excluding specific read groups in one of he exercises, though.
- Inform yourself on the options of samtools mpileup:

```
samtools mpileup
```

- For practise, use the samplepair bam file as input and run samtools mpileup for the genomic position of interest:

```
samtools mpileup -f human37.fa -r chr16:81077733-81077733 samplepair.bam
```

- What is the total coverage at our variant position of interest?
- Run samtools mpileup for the same position, but exclude the tumor or the control sample, respectively:

```
samtools mpileup -f human37.fa -r chr16:81077733-81077733 -R tumor.RG samplepair.bam
samtools mpileup -f human37.fa -r chr16:81077733-81077733 -R control.RG samplepair.bam
```

- What is the coverage at our variant position of interest in the control and tumor sample, respectively, using these commands?
- What is the coverage at our variant position of interest in the control and tumor sample, respectively, setting the base quality cutoff to 20? And setting the base quality cutoff to 30?

```
samtools mpileup -f human37.fa -r chr16:81077733-81077733 -Q 20 -R tumor.RG
samplepair.bam
samtools mpileup -f human37.fa -r chr16:81077733-81077733 -Q 20 -R control.RG
samplepair.bam
```

- With each of these base quality cutoff values, what is the proportion of alternative bases of all bases covering this position?

Files from Variant_calling/3_vcf_samtools:

- Generate the vcf file in this folder yourself running samtools mpileup piped to bcftools view:

```
samtools mpileup -DSug -f human37.fa samplepair.bam | bcftools view -vcg - >
variants.vcf
```

- Look inside the vcf file.
- Take another look at the options of samtools mpileup and bcftools view. Can you think of tuning any of the parameters in order to be less strict at calling or outputting our variant of interest in the vcf file?

Files from Variant_calling/4_vcf_GATK:

- Generate the vcf file in this folder yourself running GATK UnifiedGenotyper:

```
java -jar GenomeAnalysisTK.jar -R human37.fa -T UnifiedGenotyper -I samplepair.bam -o
variants_GATK.vcf -L chr16:81077733-81077733 -glm BOTH -dt NONE
```

- Look inside the vcf file.
- Take another look at the options of GATK UnifiedGenotyper. Hint: Interesting options to include may be: -ploidy, -out_mode .