Workshop 2024 commands

Make a directory for output

```
mkdir output
```

0. Quality Control

```
fastqc ${sequences}/24NGS775-B1_S55_R1_001.fastq.gz ${sequences}/24NGS775-
B1_S55_R2_001.fastq.gz -o output/
```

1. Trimming

```
trimmomatic PE ${sequences}/24NGS775-B1_S55_R1_001.fastq.gz ${sequences}/24NGS775-
B1_S55_R2_001.fastq.gz -baseout output/24NGS775-B1.fq.gz
ILLUMINACLIP:$adapters:2:30:10:2:keepBothReads LEADING:3 SLIDINGWINDOW:4:15
MINLEN:40
```

2. Paired end assembly

```
flash output/24NGS775-B1_1P.fq.gz output/24NGS775-B1_2P.fq.gz --cap-mismatch-quals
-O -M 250 -o output/24NGS775-B1
```

3. Mapping to reference genome

```
bwa mem -R "@RG\tID:AML\tPL:ILLUMINA\tLB:LIB-MIPS\tSM:24NGS775-B1\tPI:200" -M -t
20 ${genome} output/24NGS775-B1.extendedFrags.fastq > output/24NGS775-B1.sam
```

4. Sam conversion

```
samtools view -b output/24NGS775-B1.sam > output/24NGS775-B1.bam
samtools sort output/24NGS775-B1.bam > output/24NGS775-B1.sorted.bam
samtools index output/24NGS775-B1.sorted.bam > output/24NGS775-B1.sorted.bam.bai
```

5. GATK Best Practices for data pre-processing

Details of this step can be found here : https://gatk.broadinstitute.org/hc/en-us/articles/360035535912-Data-pre-processing-for-variant-discovery

```
java -Xmx8G -jar ${GATK38_path} -T RealignerTargetCreator -R ${genome} -nt 10 -I
output/24NGS775-B1.sorted.bam --known ${site1} -o output/24NGS775-B1.intervals

java -Xmx8G -jar ${GATK38_path} -T IndelRealigner -R ${genome} -I output/24NGS775-
B1.sorted.bam -known ${site1} --targetIntervals output/24NGS775-B1.intervals -o
output/24NGS775-B1.realigned.bam

java -Xmx8G -jar ${GATK38_path} -T BaseRecalibrator -R ${genome} -I
output/24NGS775-B1.realigned.bam -knownSites ${site2} -knownSites ${site3} -
maxCycle 600 -o output/24NGS775-B1.recal_data.table

java -Xmx8G -jar ${GATK38_path} -T PrintReads -R ${genome} -I output/24NGS775-
B1.realigned.bam --BQSR output/24NGS775-B1.recal_data.table -o output/24NGS775-
B1.final.bam
```

6. Coverage calculation

```
bedtools bamtobed -i output/24NGS775-B1.final.bam > output/24NGS775-B1.final.bed
bedtools coverage -counts -a ${bedfile}.bed -b output/24NGS775-B1.final.bed >
output/24NGS775-B1.counts.bed
```

7. Variant calling

```
java -Xmx10G -jar ${GATK38_path} -T MuTect2 -R ${genome} -I:tumor output/24NGS775-
B1.final.bam -o output/24NGS775-B1_mutect.vcf -L ${bedfile}.bed
```

8. Variant annotation

```
convert2annovar.pl -format vcf4 output/24NGS775-B1_vardict.vcf --outfile
output/24NGS775-B1.avinput --withzyg --includeinfo

table_annovar.pl output/24NGS775-B1.avinput --out output/24NGS775-B1_final --
remove --protocol refGene,cosmic84,exac03 --operation g,f,f --buildver hg19 --
nastring '-1' --otherinfo --csvout ${database}
```

9. Format output

```
python3 ${formatMutect_script_path} output/24NGS775-B1_final.hg19_multianno.csv
24NGS775-B1 output/
```

10. KDMdb

```
python3 ${KDMdb_script_path} output/24NGS775-B1_mutect.csv output/ 24NGS775-B1
```