Input Files

1. **Reference Genome:**
   a. `Ref.fasta`
2. **Reads:**
   a. `kanin-3_1-cln.fq.gz`
   b. `kanin-3_2-cln.fq.gz`
   c. `kanin-4_1-cln.fq.gz`
   d. `kanin-4_2-cln.fq.gz`

Create Directory for Output Files

```
mkdir output
```

Reference Data Preparation

```
bwa index Ref/Ref.fasta
samtools faidx Ref/Ref.fasta
java -jar $PICARD CreateSequenceDictionary -REFERENCE Ref/Ref.fasta -OUTPUT Ref/Ref.dict
```

Raw Data Preparation

QC

```
fastqc kanin/kanin-3_1.fq.gz kanin/kanin-3_2.fq.gz kanin/kanin-4_1.fq.gz kanin/kanin-4_2.fq.gz
```

Cleaning

```
fastp -i kanin/kanin-3_1.fq.gz -I kanin/kanin-3_2.fq.gz -o kanin/kanin-3_1-cln.fq.gz -O kanin/kanin-3_2-cln.fq.gz

fastp -i kanin/kanin-4_1.fq.gz -I kanin/kanin-4_2.fq.gz -o kanin/kanin-4_1-cln.fq.gz -O kanin/kanin-4_2-cln.fq.gz
```

Alignment to Reference Genome

```
bwa mem -t 8 Ref/Ref.fasta kanin/kanin-3_1-cln.fq.gz kanin/kanin-3_2-cln.fq.gz > Output/kanin_3.sam
bwa mem -t 8 Ref/Ref.fasta kanin/kanin-4_1-cln.fq.gz kanin/kanin-4_2-cln.fq.gz > Output/kanin_4.sam
```

Sorting and Compression

```
java -jar $GATK SortSam --INPUT Output/kanin_3.sam --OUTPUT Output/kanin_3.sorted.bam -SORT_ORDER
coordinate

java -jar $GATK SortSam --INPUT Output/kanin_4.sam --OUTPUT Output/kanin_4.sorted.bam -SORT_ORDER
coordinate
```

Fix mate information

```
java -jar $GATK FixMateInformation --INPUT Output/kanin_3.sorted.bam --OUTPUT Output/kanin_3.fxmt.bam --
CREATE_INDEX TRUE --VALIDATION_STRINGENCY LENIENT

java -jar $GATK FixMateInformation --INPUT Output/kanin_4.sorted.bam --OUTPUT Output/kanin_4.fxmt.bam --
CREATE_INDEX TRUE --VALIDATION_STRINGENCY LENIENT
```

Add read group information

```
java -jar $GATK AddOrReplaceReadGroups --INPUT Output/kanin_3.fxmt.bam --OUTPUT Output/kanin_3.addrep.bam -
-RGID kanin_3 --RGPU unit1 --RGLB kanin_3 -PL Illumina -SM kanin -CN CN --VALIDATION_STRINGENCY LENIENT --
CREATE_INDEX TRUE

java -jar $GATK AddOrReplaceReadGroups --INPUT Output/kanin_4.fxmt.bam --OUTPUT Output/kanin_4.addrep.bam -
-RGID kanin_4 --RGPU unit1 --RGLB kanin_4 -PL Illumina -SM kanin -CN CN --VALIDATION_STRINGENCY LENIENT --
CREATE_INDEX TRUE
```

Mark Duplicates

```
java -jar $GATK MarkDuplicates --INPUT Output/kanin_3.addrep.bam --OUTPUT Output/kanin_3.mkdup.bam --
VALIDATION_STRINGENCY LENIENT --CREATE_INDEX TRUE --METRICS_FILE Output/kanin_3.mkdup.metrics

java -jar $GATK MarkDuplicates --INPUT Output/kanin_4.addrep.bam --OUTPUT Output/kanin_4.mkdup.bam --
VALIDATION_STRINGENCY LENIENT --CREATE_INDEX TRUE --METRICS_FILE Output/kanin_4.mkdup.metrics
```

Merging BAM files

```
samtools merge Output/kanin.mkdup.bam Output/kanin_3.mkdup.bam Output/kanin_4.mkdup.bam
```

Indexing BAM Files

```
samtools index Output/kanin.mkdup.bam
```

Variant Calling

```
java -jar $GATK HaplotypeCaller \
-R Ref/Ref.fasta \
-I Output/kanin.mkdup.bam \
-O Output/kanin.g.vcf \
-ERC GVCF \
--bam-output Output/kanin.bamout.bam
```