

MOLECULAR METHODS AND BIOINFORMATICS

LM Evolutionary Biology, University of Padua

Dr. Enrico Gaffo, Dr. Silvia Orsi,

Prof. Stefania Bortoluzzi

Practical lesson 5

“NGS data analysis: Variant calling”

Padua, November 18, 2024

COMMAND GUIDE

Fields preceded by the "\$" symbol mean that you have to insert the file yourself.

Alignment to the reference genome

```
bwa mem -R
```

```
"@RG\tID:sample\tLB:exome\tSM:sample\tPL:ILLUMINA"
```

```
$genome $first_fastq $second_fastq > $reads_aligned_SAM
```

SAM to BAM Conversion

```
samtools view -b $reads_aligned_SAM > $reads_aligned_BAM
```

Sorting of aligned reads

```
samtools sort $reads_aligned_BAM > $reads_sorted_BAM
```

BAM Indexing

```
samtools index $reads_sorted_BAM
```

PCR Duplicate Removal

```
gatk MarkDuplicates -I $reads_sorted_BAM -O  
$no_duplicates_BAM -M metrics.txt -REMOVE_DUPLICATES  
true -CREATE_INDEX true
```

Recalibration of the quality of the bases

```
gatk BaseRecalibrator -I $no_duplicates_BAM -R $genome
```

```
--known-sites $dbSNP -O model.grp
```

```
gatk ApplyBQSR -R $genome -I $no_duplicates_BAM -bqsr  
model.grp -O $recalibrated_alignment_BAM
```

Calling of germline variants

```
gatk HaplotypeCaller -R $genome -I  
$recalibrated_alignment_BAM -O $VCF_germline_variants
```

```
gatk VariantFiltration -V $VCF_germline_variants -filter  
"QUAL < 30.0" --filter-name "QUAL30" -filter "MQ < 40.0"  
--filter-name "MQ40" -filter "DP < 30" --filter-name  
"DP30" -O $filtered_variants_VCF
```

```
gatk SelectVariants -R ref.fa -V $filtered_variants_VCF  
--exclude-filtered -O $selected_variants_VCF
```

Annotations of germline variants

```
gatk VariantAnnotator -R $genome -V  
$selected_variants_VCF -O $annotated_variants_VCF  
--dbSNP $dbSNP
```