Hands-on: NGS Pipeline

Raw Reads
FASTQ
@SEQ_ID

1. Quality Control
FastQC/MultiQC

2. Alignment
BWA-MEM

3. SAM → BAM
SAM → BAM

4. Mark Duplicates

5. Variant Calling
Ref: A
Alt: G

6. Filter Variants
VCF

Filtered Variant
High-quality VCF
✓ Ready for annotati

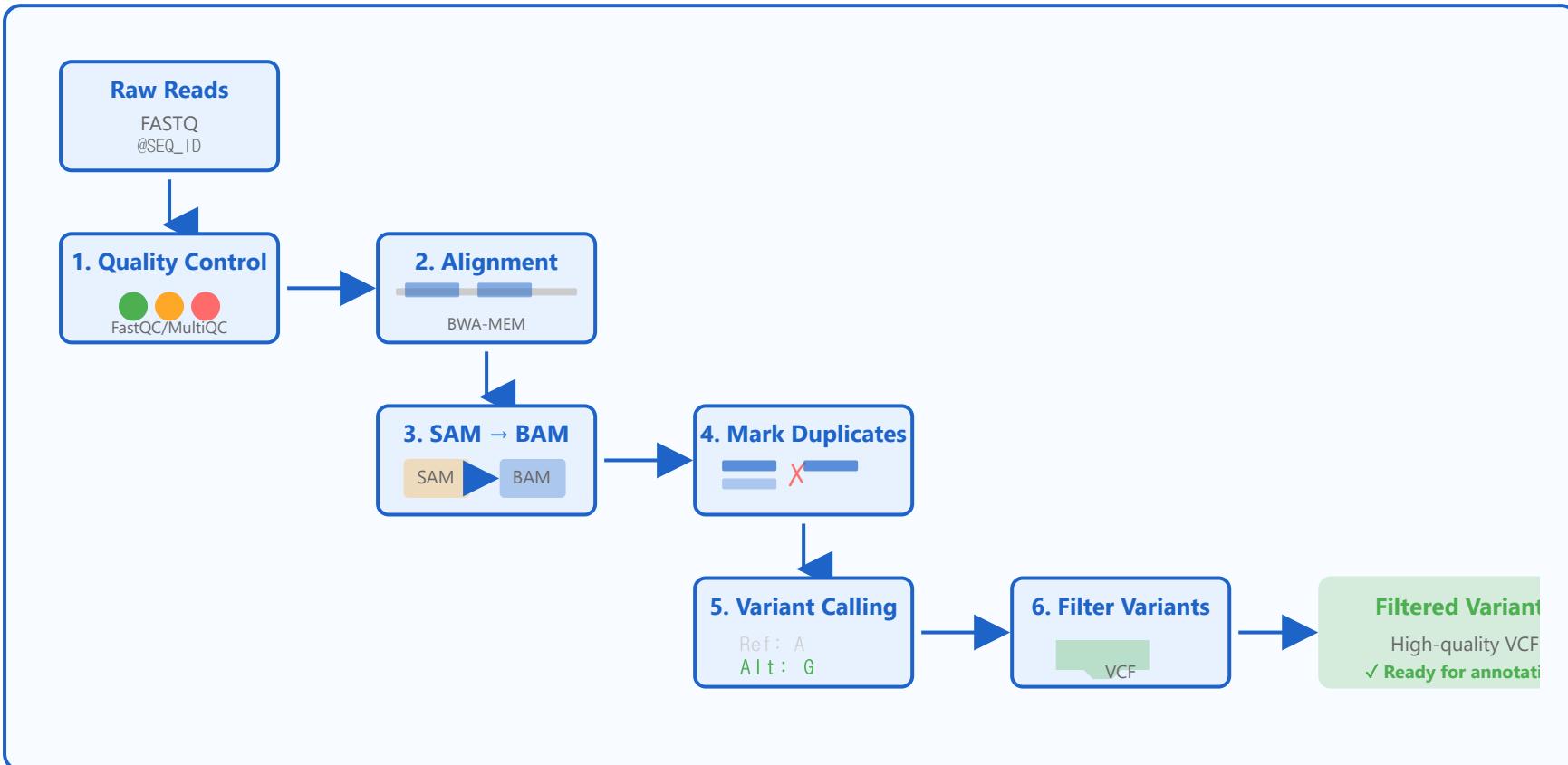## Standard NGS Analysis Pipeline

```
# 1. Quality Control

fastqc sample_R1.fastq.gz sample_R2.fastq.gz

multiqc .
```

```
# 2. Read Alignment
bwa mem -t 8 reference.fa sample_R1.fastq.gz sample_R2.fastq.gz > sample.sam

# 3. Convert SAM to BAM and Sort
samtools view -bS sample.sam | samtools sort -o sample.sorted.bam
samtools index sample.sorted.bam

# 4. Mark Duplicates
gatk MarkDuplicates -I sample.sorted.bam -O sample.dedup.bam -M metrics.txt

# 5. Variant Calling
gatk HaplotypeCaller -R reference.fa -I sample.dedup.bam -O sample.vcf

# 6. Variant Filtering
gatk VariantFiltration -R reference.fa -V sample.vcf -O sample.filtered.vcf
```

Required Software

FastQC, BWA, SAMtools, GATK, Picard

Typical Runtime

4-24 hours depending on coverage and compute resources