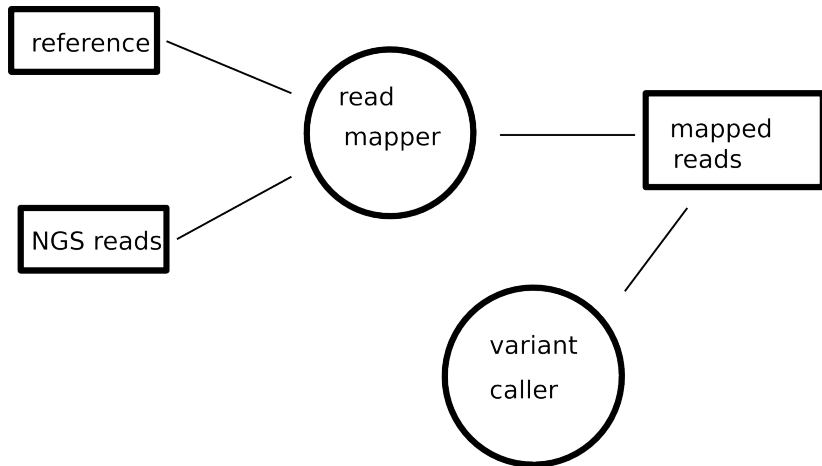


Pipeline overview



- Index reference genome (need to do it only once per reference)

```
tmap index -f reference.fasta
```

Read mapping

- Index reference genome (need to do it only once per reference)

```
tmap index -f reference.fasta
```

- Map reads against reference genome

```
tmap map3 -f reference.fasta \  
-r reads.fastq \  
-i fastq \  
-o 2 \  
-s mapped_reads.bam
```

- Let's see what is inside the bam file

```
samtools view mapped_reads.bam | head -n 20
```

[illegible]

BAM file format

The header

- Let's look at the bam header

```
samtools view -H mapped_reads.bam
```

```
@HD      VN:1.4
@SQ      SN:gi|385235550|ref|NC_017387.1|      LN:4138388
@RG      ID:NOID PG:tmap SM:NOSM
@PG      ID:tmap CL:map3 -f reference.fasta -r reads.fastq -i fastq -o 2 -s mapped_reads.bam      VN:3.4.1
```

BAM file header modifying

- Let's change the header so that it includes our sample name!

```
samtools view -H mapped_reads.bam \  
| sed 's/SM:NOSM/SM:Sample1/' \  
| samtools reheader - mapped_reads.bam \  
> mapped_reads.reheaded.bam
```

Variant calling

- Variant callers need sorted bam file

```
samtools sort mapped_reads.reheaded.bam \
mapped_reads.reheaded.sorted
```


Variant calling

```
samtools mpileup -uf reference.fasta \  
mapped_reads.reheaded.sorted.bam \  
| bcftools call -cv \  
| vcfutils.pl varFilter > variants.samtools.vcf
```

Variant call format (VCF)

```
##fileformat=VCFv4.2
##FILTER=<ID=PASS,Description="All filters passed">
##samtoolsVersion=1.0-11-geeb4b22+htslib-1.0-2-g1903fd4
##samtoolsCommand=samtools mpileup -uf reference.fasta mapped_reads.sorted.bam
##referenceFile=file://reference.fasta
##contig=<ID=gl|385235550|ref|NC_017387.1|,length=4138388>
##ALT=<ID=X,Description="Represents allele(s) other than observed.">
##INFO=<ID=INDEL,Number=0,Type=Flag,Description="Indicates that the variant is an INDEL.">
##INFO=<ID=IDV,Number=1,Type=Integer,Description="Maximum number of reads supporting an indel">
##INFO=<ID=IMF,Number=1,Type=Float,Description="Maximum fraction of reads supporting an indel">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Raw read depth">
##INFO=<ID=VDB,Number=1,Type=Float,Description="Variant Distance Bias for filtering splice-site artefacts in RNA-seq data (bigger is better)",Version=3>
##INFO=<ID=RPB,Number=1,Type=Float,Description="Mann-Whitney U test of Read Position Bias (bigger is better)">
##INFO=<ID=MQB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality Bias (bigger is better)">
##INFO=<ID=BQB,Number=1,Type=Float,Description="Mann-Whitney U test of Base Quality Bias (bigger is better)">
##INFO=<ID=MQSB,Number=1,Type=Float,Description="Mann-Whitney U test of Mapping Quality vs Strand Bias (bigger is better)">
##INFO=<ID=SQB,Number=1,Type=Float,Description="Segregation based metric.">
##INFO=<ID=MQ0F,Number=1,Type=Float,Description="Fraction of MQ0 reads (smaller is better)">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="List of Phred-scaled genotype likelihoods">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##INFO=<ID=AF1,Number=1,Type=Float,Description="Max-likelihood estimate of the first ALT allele frequency (assuming HWE)">
##INFO=<ID=AF2,Number=1,Type=Float,Description="Max-likelihood estimate of the first and second group ALT allele frequency (assuming HWE)">
##INFO=<ID=AC1,Number=1,Type=Float,Description="Max-likelihood estimate of the first ALT allele count (no HWE assumption)">
##INFO=<ID=MQ,Number=1,Type=Integer,Description="Root-mean-square mapping quality of covering reads">
##INFO=<ID=FQ,Number=1,Type=Float,Description="Phred probability of all samples being the same">
##INFO=<ID=PV4,Number=4,Type=Float,Description="P-values for strand bias, baseQ bias, mapQ bias and tail distance bias">
##INFO=<ID=G3,Number=3,Type=Float,Description="ML estimate of genotype frequencies">
##INFO=<ID=HWE,Number=1,Type=Float,Description="Chi^2 based HWE test P-value based on G3">
##INFO=<ID=DP4,Number=4,Type=Integer,Description="Number of high-quality ref-forward, ref-reverse, alt-forward and alt-reverse bases">
##bcftools_callVersion=1.0-26-g53fcd5c+htslib-1.0-6-g546bfc
##bcftools_callCommand=call -cv
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NOSM
gl|385235550|ref|NC_017387.1| 5461 . T G 20.8005 . DP=4;VDB=0.14;SGB=-0.453602;MQ0F=0;AF1=1;AC1=2;DP4=0,0,0,2;MQ=60;FQ=-32.988 GT:PL 1/1:52,6,0
gl|385235550|ref|NC_017387.1| 14453 . A C 20.8005 . DP=2;VDB=0.02;SGB=-0.453602;MQ0F=0;AF1=1;AC1=2;DP4=0,0,0,2;MQ=47;FQ=-32.988 GT:PL 1/1:52,6,0
gl|385235550|ref|NC_017387.1| 51777 . TAA T 14.1422 . INDEL;IDV=3;IMF=0.75;DP=4;VDB=0.0381671;SGB=-0.556411;MQ0F=0;AF1=1;AC1=2;DP4=0,0,0,4;MQ=25;FQ=-46.52
1:54,12,0
gl|385235550|ref|NC_017387.1| 69668 . G C 17.8363 . DP=4;VDB=0.02;SGB=-0.453602;MQ0F=0;AF1=1;AC1=2;DP4=0,0,0,2;MQ=60;FQ=-32.988 GT:PL 1/1:49,6,0
gl|385235550|ref|NC_017387.1| 108655 . G C 3.01703 . DP=4;VDB=0.1;SGB=-0.453602;RPB=0.5;MQB=0.75;BQB=0.5;MQ0F=0.25;AF1=0.500542;AC1=1;DP4=0,2,0,2;MQ=45;FQ=
=1,0.352957,1,1 GT:PL 0/1:30,0,25
```

Summary of used commands

- Reference genome indexing

```
tmap index -f reference.fasta
```

- Read mapping against reference genome

```
tmap map3 -f reference.fasta \  
-r reads.fastq \  
-i fastq \  
-o 2 \  
-s mapped_reads.bam
```

- Bam header modifying

```
samtools view -H mapped_reads.bam \  
| sed 's/SM:NOSM/SM:Sample1/' \  
| samtools reheader - mapped_reads.bam \  
> mapped_reads.reheaded.bam
```

- Mapped read sorting

```
samtools sort mapped_reads.reheaded.bam \  
mapped_reads.reheaded.sorted
```

- Variant calling

```
samtools mpileup -uf reference.fasta \  
mapped_reads.reheaded.sorted.bam \  
| bcftools call -cv \  
| vcfutils.pl varFilter > variants.samtools.vcf
```

Let's build pipeline!

```
tmap map3 -f reference.fasta \  
-r reads.fastq \  
-i fastq \  
-o 2 \  
-s mapped_reads.bam
```

```
samtools view -H mapped_reads.bam \  
| sed 's/SM:NOSM/SM:Sample1/' \  
| samtools reheader - mapped_reads.bam \  
> mapped_reads.reheaded.bam
```

```
samtools sort mapped_reads.reheaded.bam \  
mapped_reads.reheaded.sorted  
samtools mpileup -uf reference.fasta \  
mapped_reads.reheaded.sorted.bam \  
| bcftools call -cv \  
| vcfutils.pl varFilter > variants.samtools.vcf
```