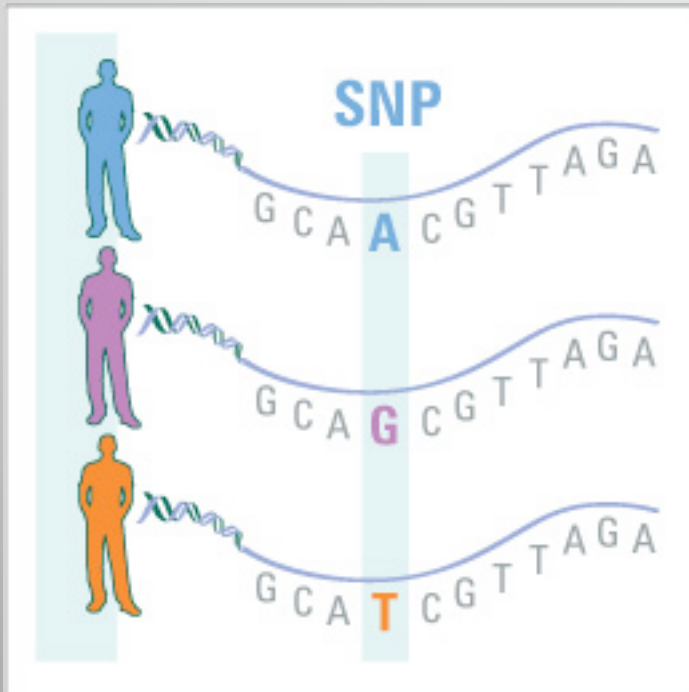


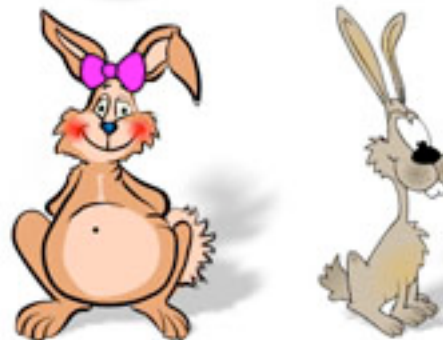
Variant calling and genotyping



InDel

TCGCGAA-----TTCCCAT
TCGCGAAACGTTTCCCAT

bb ♀ BB ♂

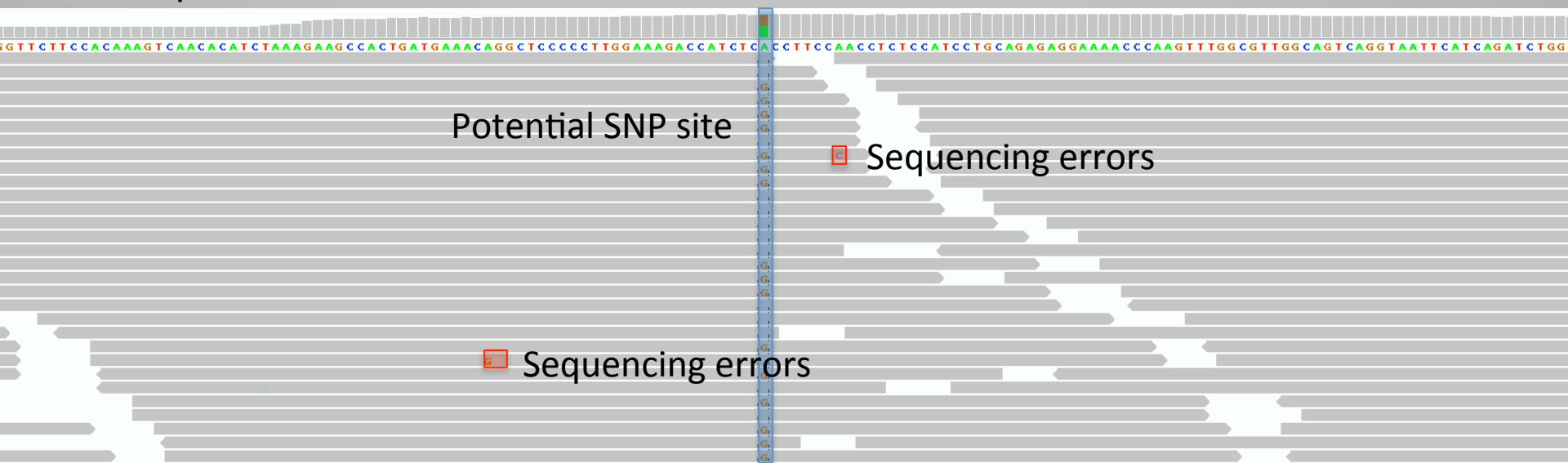


Genotype

Phenotype

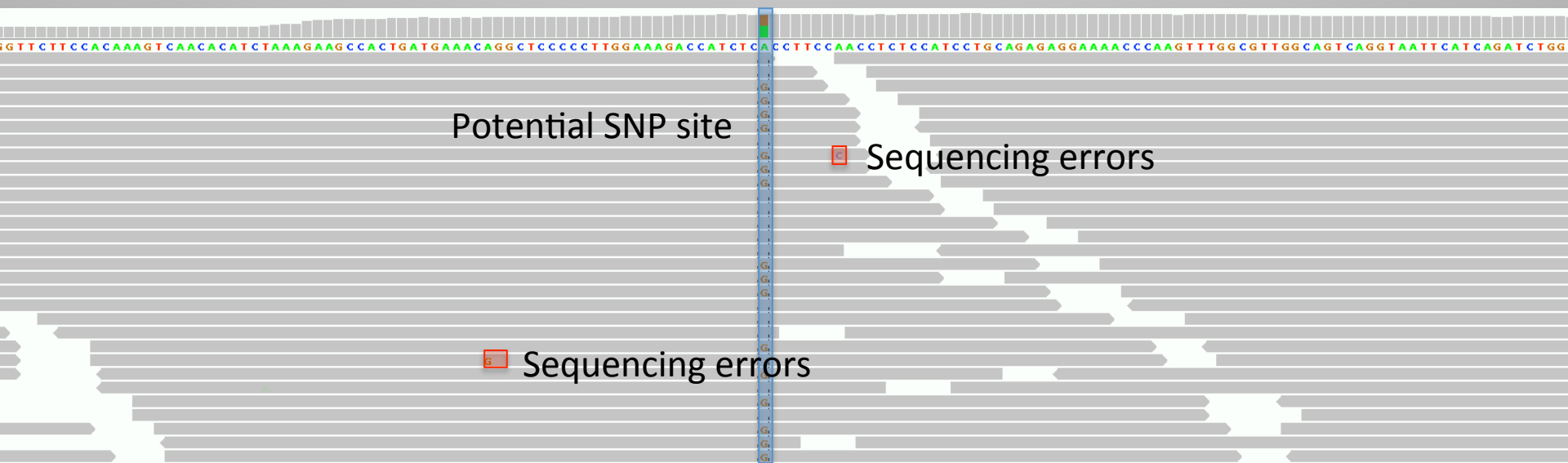
Variant detection

- Single Nucleotide Polymorphisms (SNPs) and Insertions/Deletions (InDels) can be found by searching for consistent patterns of mismatches in the alignment.
- The difficulty lies in distinguishing true SNPs from sequencing errors and alignment artifacts – extensive filtering is necessary.
- Once this is done, individuals can be genotyped at each variant site depending on the number of reads with 1 or the other base.
- Allele frequencies can also be calculated directly from pooled population samples.



Variant detection

- Both SNP calls and genotypes are assigned Phred quality scores by the SNP detection algorithm. It is very possible to have a high-quality SNP with low quality genotypes in all individuals but one.
- More individuals -> More SNPs, but proportionally fewer that can be genotyped in all individuals due to uneven coverage.
- Information on SNPs and genotypes is given in VCF file format



VCF files

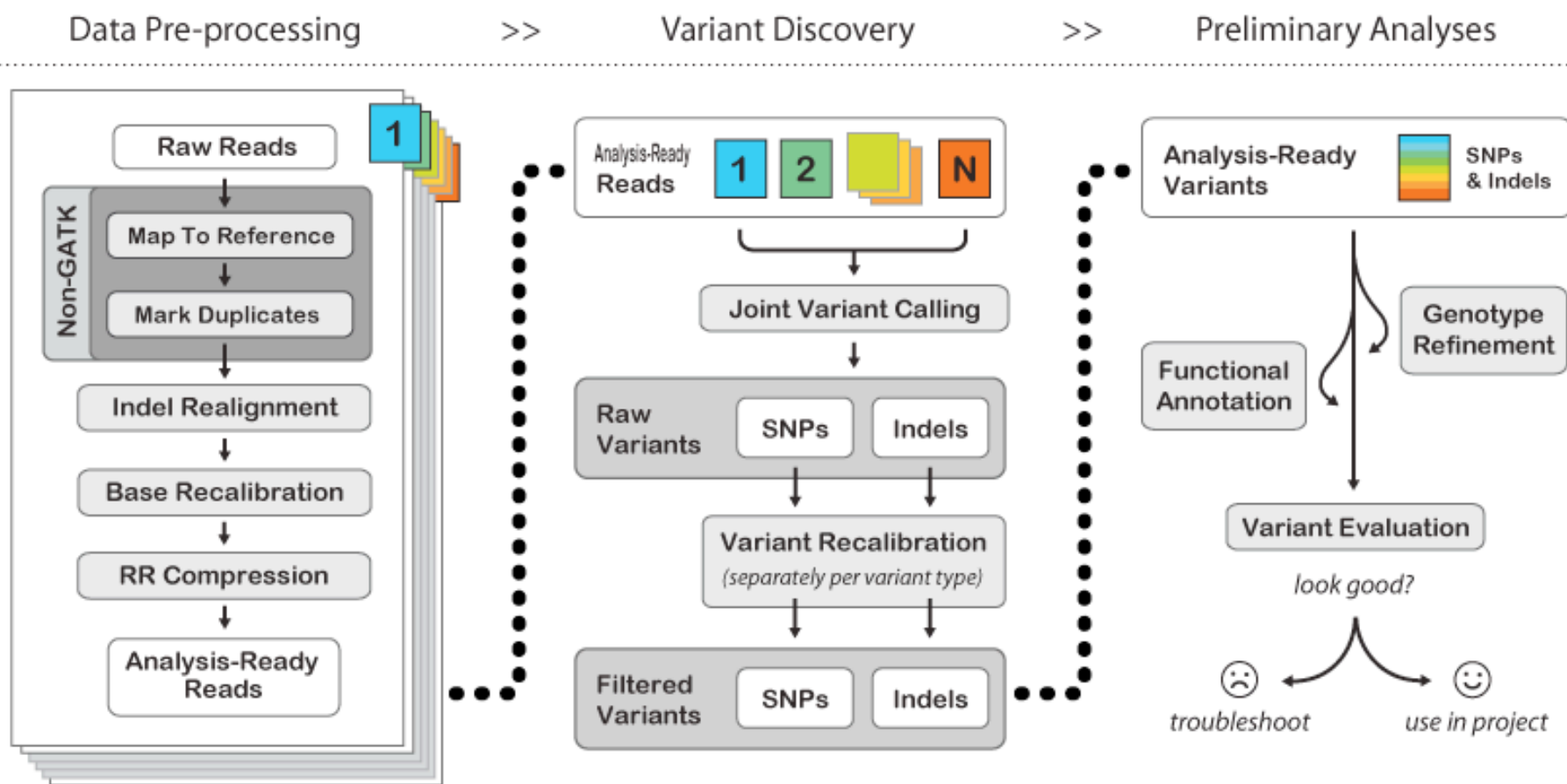
- Variant Call Format
- Contains information on the location and quality of all detected SNPs, reference and alternative allele, sequencing depth plus genotypes of all individuals in the dataset along with genotype quality.
- More info on the file format can be found at <http://www.1000genomes.org/node/101>

An example line from a VCF file:

```
contig13845 79 . C T 897.77 PASS
AB=0.462;AC=1;AF=0.013;AN=80;BaseCounts=0,1598,0,28;BaseQRankSum=2.532;DP=1626;Dels=0.00;FS=11.803;GC=47.52;HRun=1;HW=0.0;H
aplotypescore=0.3232;InbreedingCoeff=-0.0127;LowMQ=0.0000,0.0000,1626;MQ=37.00;MQ0=0;MQ0Fraction=0.0000;MQRankSum=-1.897;P
ercentNBaseSolid=0.0000;QD=17.26;ReadPosRankSum=1.229;SB=-448.65;SBD=0.94;Samples=H3;VQSLOD=16.0315
GT:AB:AD:DP:FA:GQ:MQ0:PL 0/0:..44,0:44:0.000:99:0:0,132,1741 0/0:..36,0:36:0.000:99:0:0,105,1379
0/0:..11,0:11:0.000:30.10:0:0,30,398 0/0:..42,0:42:0.000:99:0:0,126,1674
```

The GATK pipeline

<http://www.broadinstitute.org/gatk/guide/best-practices>



The GATK pipeline

<http://www.broadinstitute.org/videos/broade-overview-gatk-best-practices>

