Short Read Alignment
&
Variant Calling

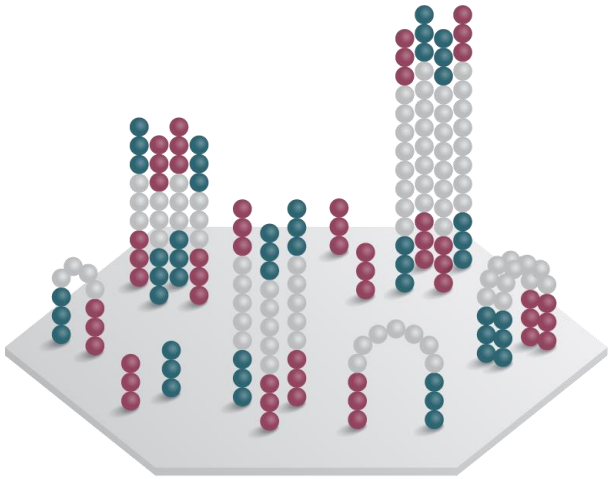
- Loo...ts of **Google/ChatGPT** search
- Proper **data** & **directory** management
- Perform **background study**
- Be **persistent**

Workflow


Short Read Alignment
Refine Alignment
Variant Calling
Variant Filtration
Annotation

Short Read Alignment

Basic Concepts and Common Terminologies

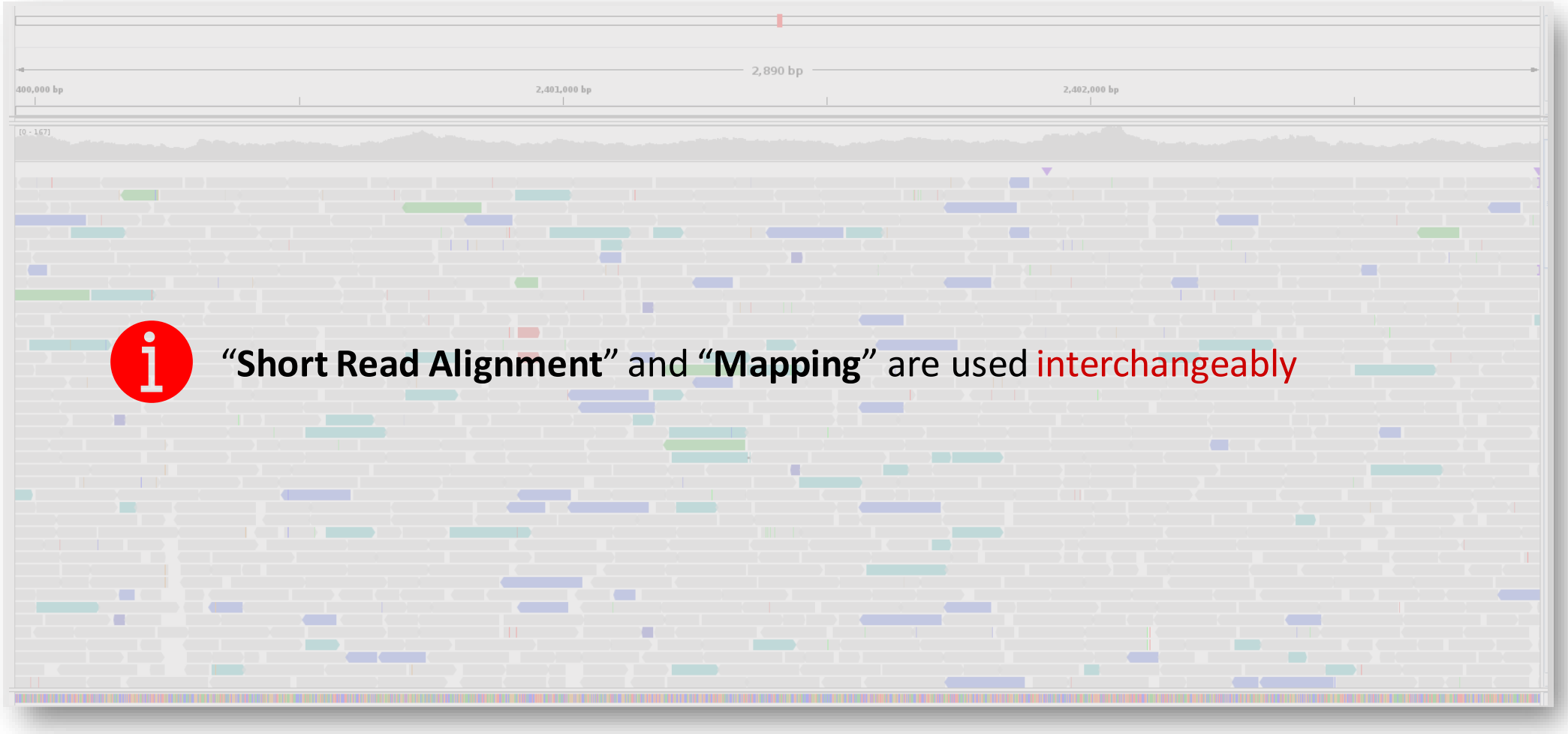


Child Health Research Foundation

Prevent Infections, Save Lives

Short Read Alignment / Mapping

Determine the most likely **location** and **orientation** of each read within the reference sequence.



Commonly Used Tools...

bwa

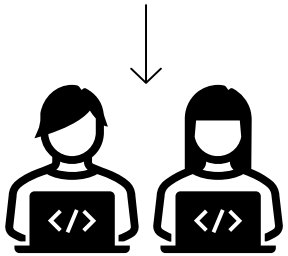
minimap2

BWA-MEM2

bowtie2

HISAT2

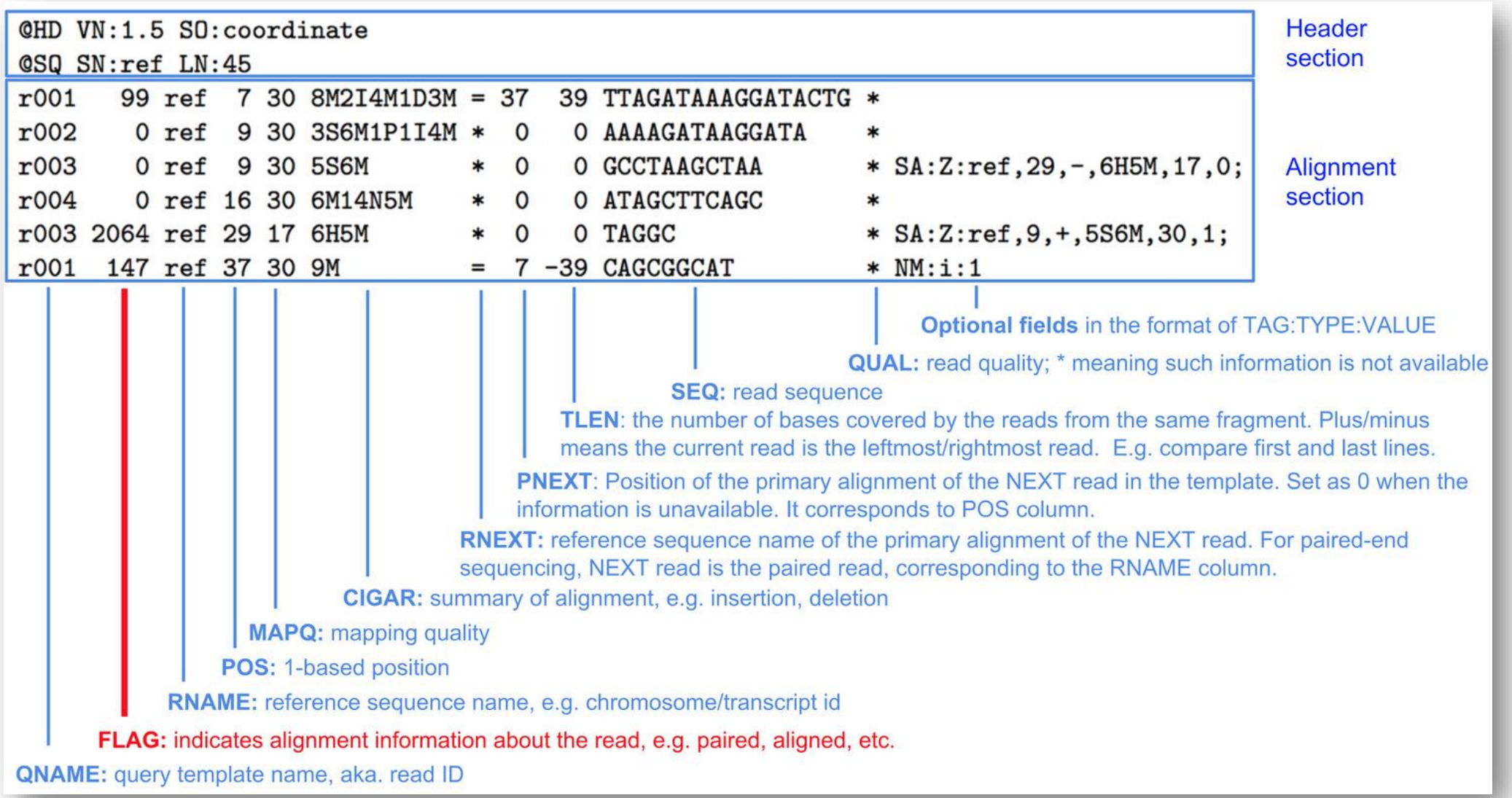
STAR



SAM File

```
@SQ     SN:NC_003198.1  LN:4809037
@PG     ID:bwa  PN:                                                    .fastq.gz data/STY_0014_R
2.fastq.gz
VH00293:1:AAAM7FVM5                                                     2329299 -168    CTGGCCGC
AAATGATTTGCGTCAGGCC                                                    GTGCGCAGAATACAGATCCCCACCG
GCGCGGAGGCAACGATTTT                                                    CCCCCCCCCCCCCCCCCCCCCCCCC
CCCCCCCCCCCCCCCCCC                                                      AS:i:150        XS:i:0
VH00293:1:AAAM7FVM5                                                     2329317 168     ATCAACAA
AGTTGACCTGCTGGCCGCA                                                    ATATTGACACCATCAATAGTGCGCA
GAATACAGATCCCCACCGC                                                    CCCCCCCCCCC;C-CCCCCCCCCCC
CCCCCCC;CCCCCCCC;CC                                                     AS:i:151        XS:i:0
VH00293:1:AAAM7FVM5                                                     2873318 -35     CCCCCCCC
CCCCCCCCCCACACCACCC                                                    CCAGAAATTCTCTTATTGTCAGAAT
AGTGAGCGTGCCTTACCGA                                                    CCCCCCCCCCCCCCCCCCCCCCCCC
CCCCCCCC;CCCCCCCCCC                                                     XS:i:0
VH00293:1:AAAM7FVM5                                                     =       2873318 35      G
TTCACTTTCCAGAAATTCT                                                    CTATAGGTGTAGATCTCCGTGGTCG
CCGTTTGTGTTATAACAGC                                                    CCCCCCCCCCCCC---CC--C---C
C;C--CC-C-CC-;C;CC;                                                     AS:i:36 XS:i:0
VH00293:1:AAAM7FVM5                                                     3459458 -75     CTTTTTCA
AGCAGAAGACGGCATACGA                                                    GGTAGCATATTGTCATGCTAACGAC
GTAACCCAACATATGGATC                                                    CCCCCCCCCCC-CCCCCCCCCC-CC
CCCCCCCCCCCCCC;CCCC                                                     XS:i:0
VH00293:1:AAAM7FVM5                                                     3459458 75      AAAAAATA
TGCCTCCCCTTGTAAGGGC                                                    TTTGTTAACCACTATCGGTGTAGAT
CTCGGTGGGGGCCGTATCATTAA CCCCCCCCCCCCCCCCCCCC;-CCCCCCC;CCCCCCCCCCCCCCCCCCCCCCCCCCCC;CCCCC-CCCCCCCCCCCC-CCCCCCCCCCCCCCCC-;
```

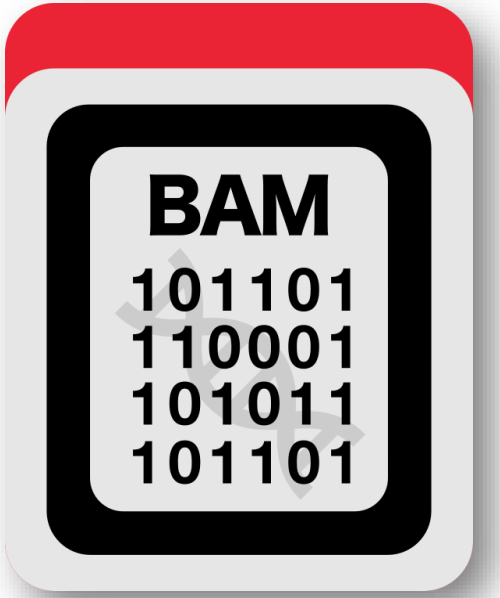
SAM (Sequence Alignment/Map) File



Source: https://www.samformat.info/sam-format-flag

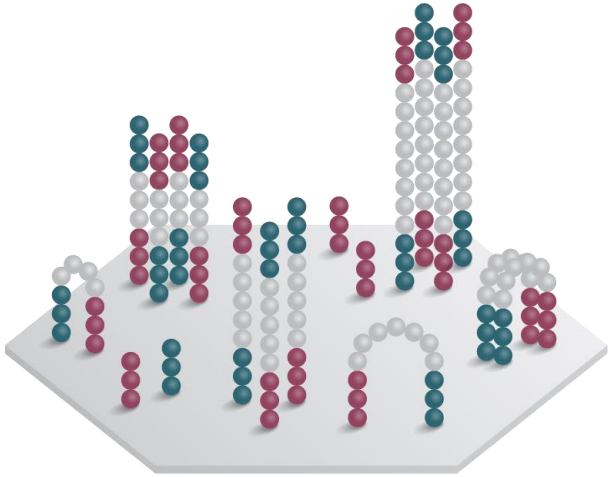
BAM (Binary Alignment/Map file) File

- **Binary, compressed (and almost always sorted) representation of the SAM information**
  - Sorting by coordinate allows fast query on the information by location.



Variant Calling

Basic Concepts and Common Terminologies

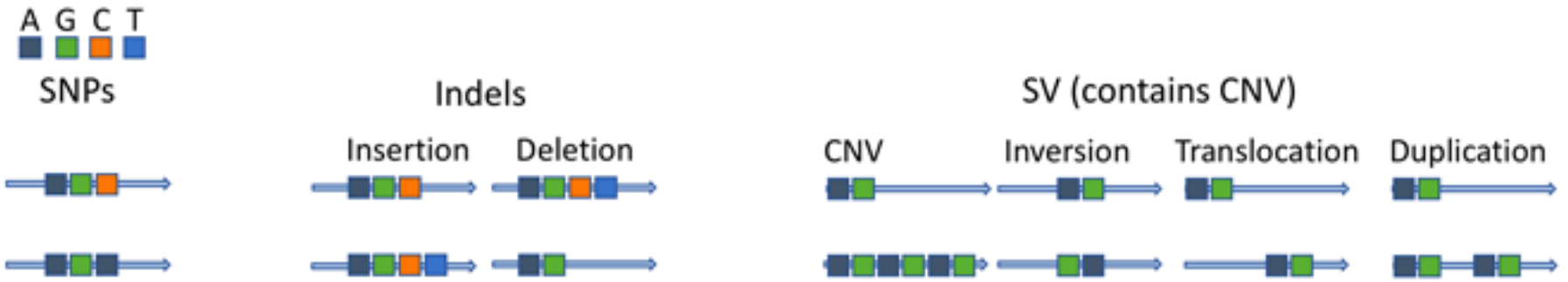


Child Health Research Foundation

Prevent Infections, Save Lives

Variant Calling

Variant calling is the process of identifying genetic variations or differences, such as single nucleotide polymorphisms (**SNPs**) or insertions/deletions (**indels**), in a subject's genome compared to a reference genome.



Source: https://www.csc.fi/-/crunching-ngs-data-on-pouta-cloud

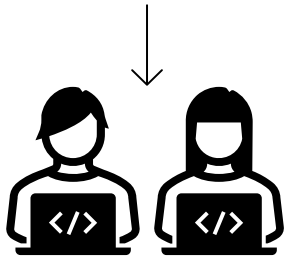
Commonly Used Variant Callers

bcftools

FreeBayes

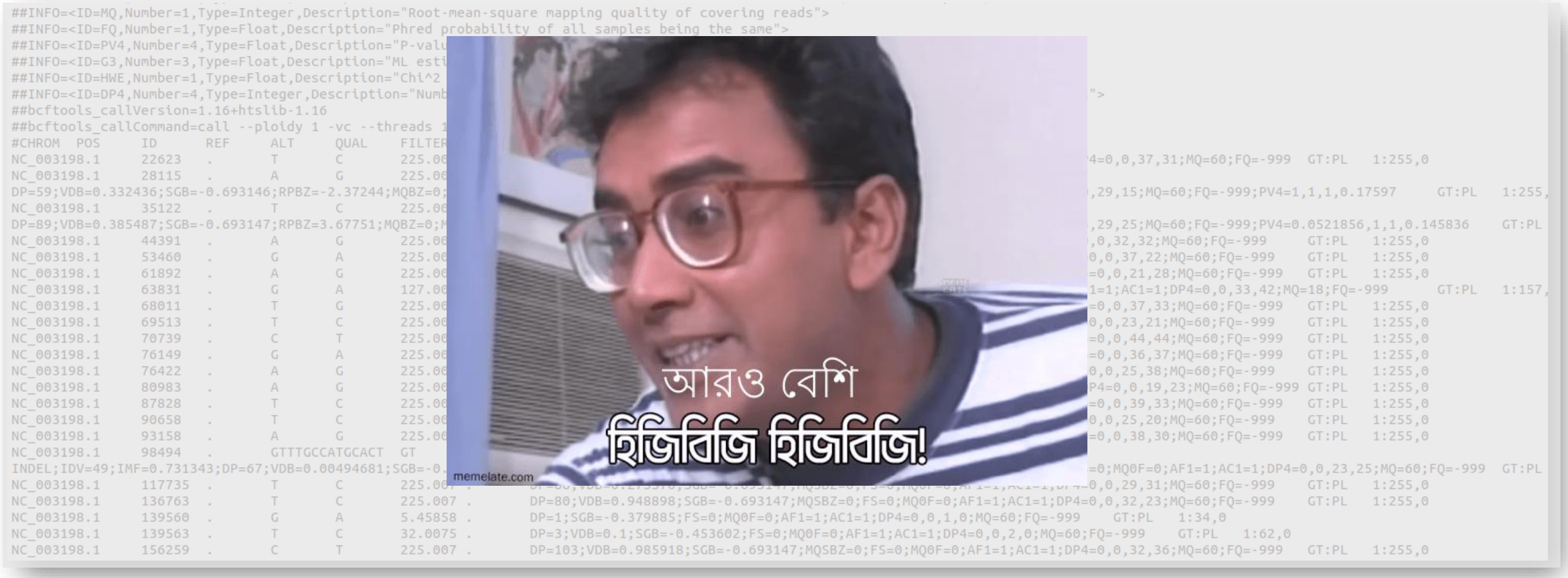
VarScan2

GATK



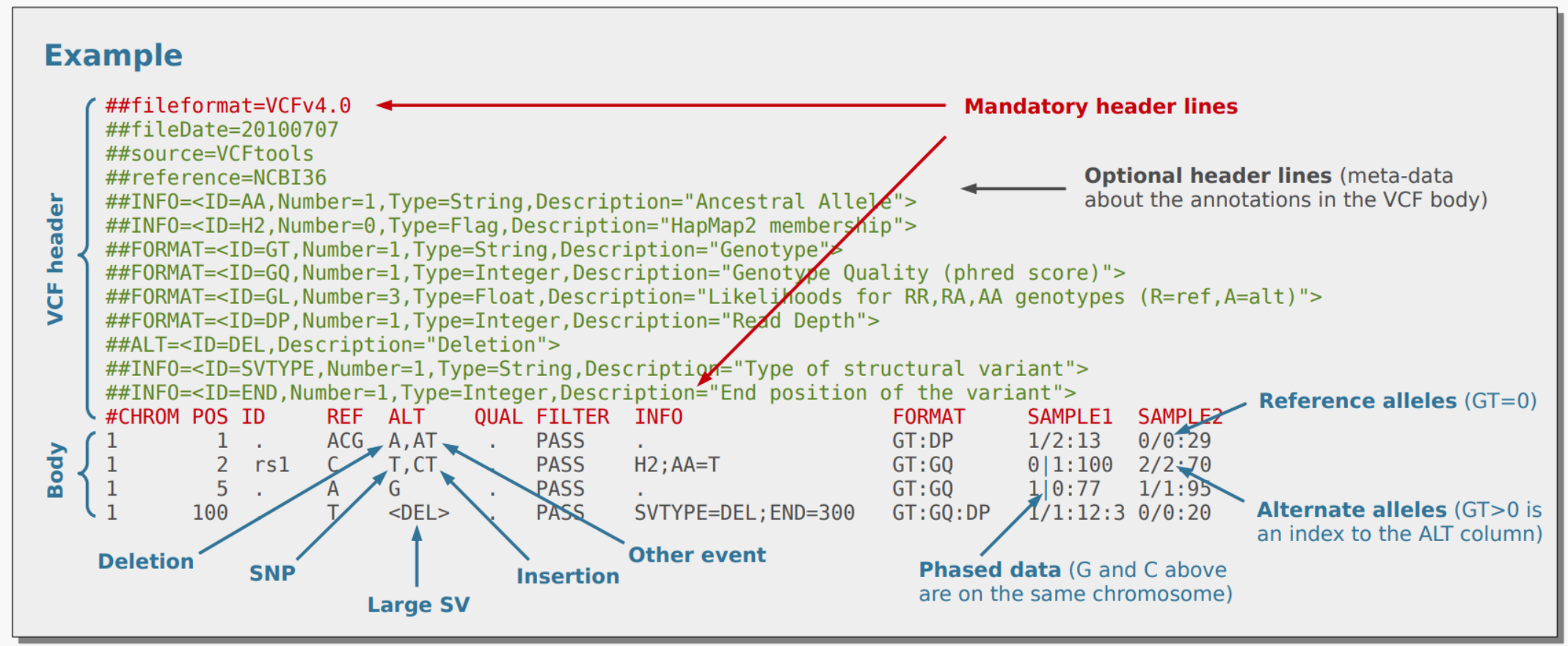
VCF file format

Data representation format used to describe **variations** in the genome



VCF file format

Data representation format used to describe **variations** in the genome



Child Health Research Foundation

Prevent Infections, Save Lives

Source: https://vcftools.sourceforge.net/VCF-poster.pdf

Happy Learning