

# *Introduction to high-throughput sequencing and its application in high-throughput sequencing and single-nucleotide variant analysis in cancer*

Cancer genomics and transcriptomics course, 2023



**Francesc Muyas Remolar, PhD**  
EMBL-EBI  
10 – 07 - 2023

# Outline

---

1. Introduction
2. Somatic variant calling
3. Workflow for cancer genome analysis
4. Library preparation and sequencing
5. Fastq files
6. Alignment
7. Variant detection
8. Variant annotation

# What do you know about these concepts?

---

- Next Generation Sequencing (NGS)
- Paired-read sequencing
- Fastq file
- BAM/SAM file
- VCF file
- Variant Allele Frequency (VAF)
- Depth of coverage
- Variant annotation

# Importance of understanding the somatic landscape in cancer

## Comprehensive Mutational Landscape

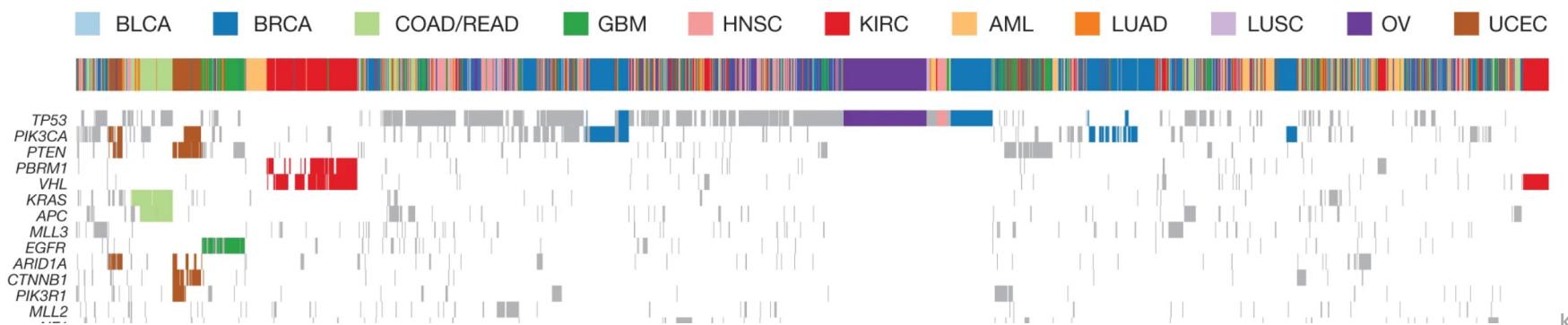
High-throughput sequencing technologies enable the analysis of cancer genomes, revealing a comprehensive view of mutational events (SNVs, SVs, CNAs...)

## Research and Clinical Applications

Exome sequencing (WES) and whole-genome sequencing (WGS) are widely used in research to characterize genomic alterations

## Precise and Informed Interventions

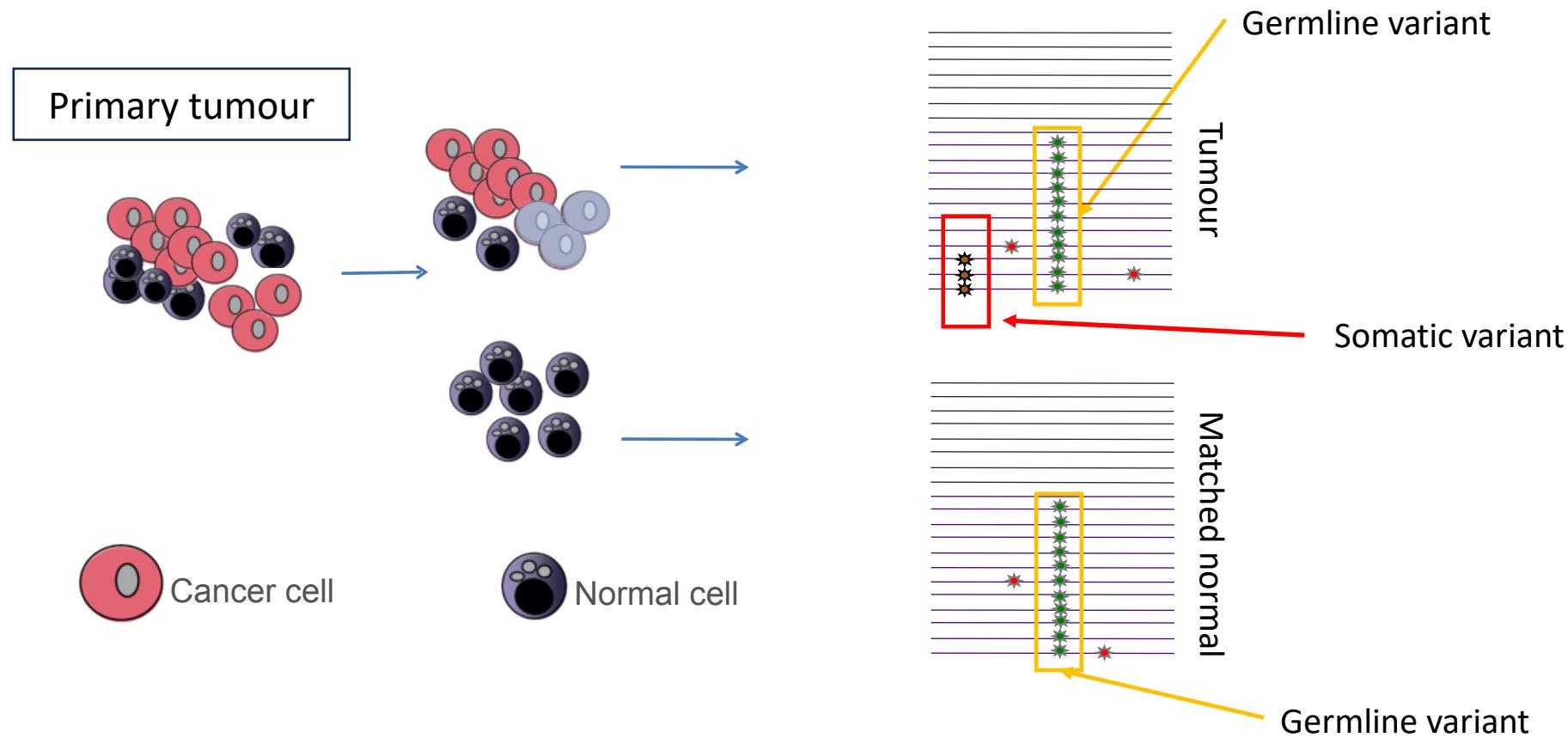
The information derived from sequencing cancer genomes allows for more precise interventions, including personalized treatment strategies and the identification of potential therapeutic targets, ultimately improving patient outcomes.



Kandoth et al 2013

# Somatic variant calling

- Somatic variant calling typically involves sequencing both tumour and matched normal samples from an individual.
- The normal sample serves as a control, representing the non-mutated or germline cells, while the tumour sample contains genetic alterations specific to the tumour cells.

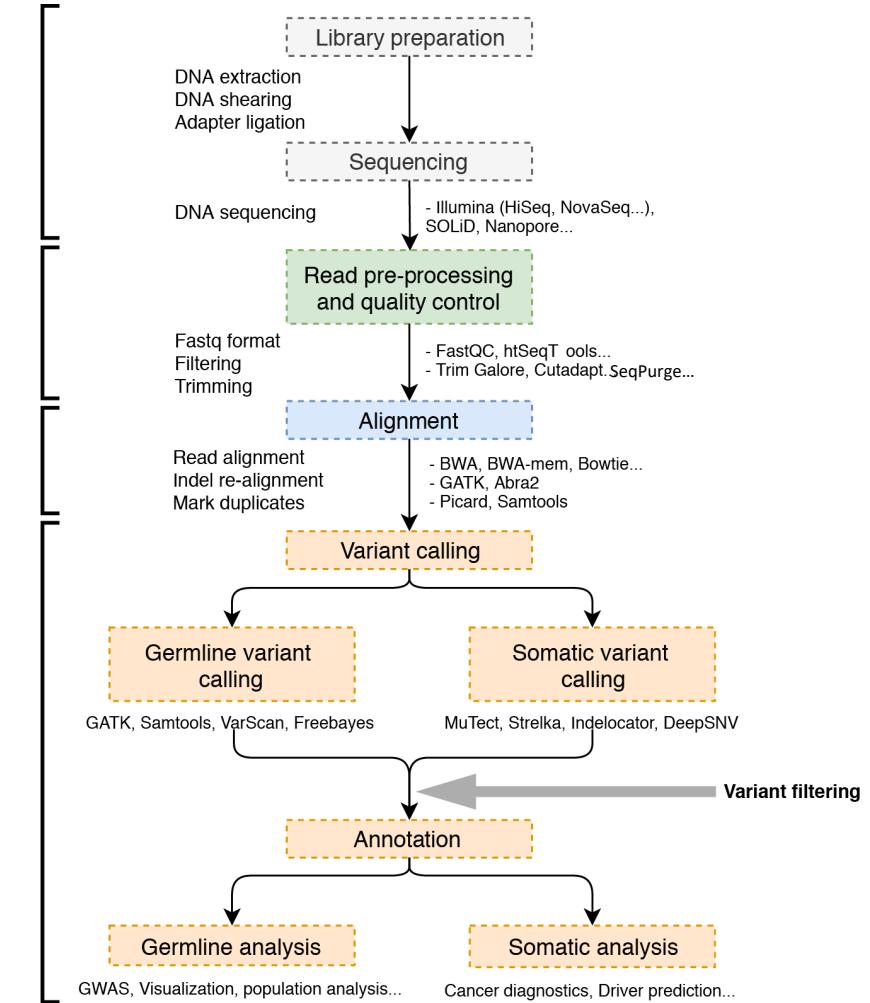
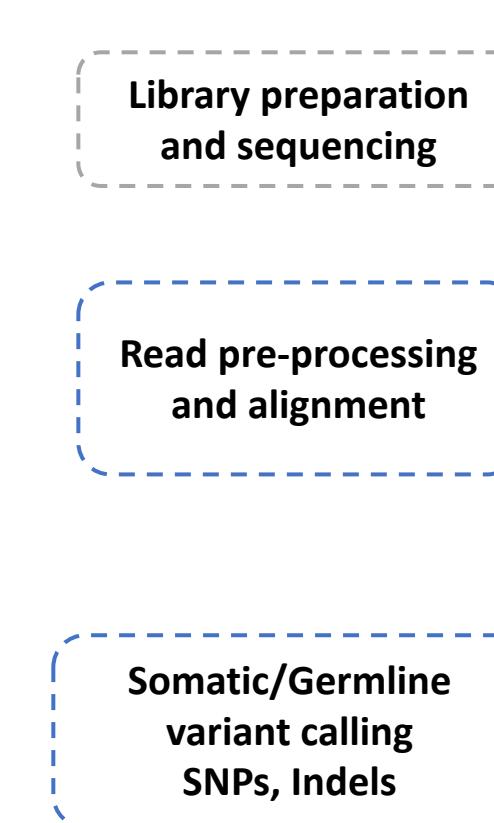
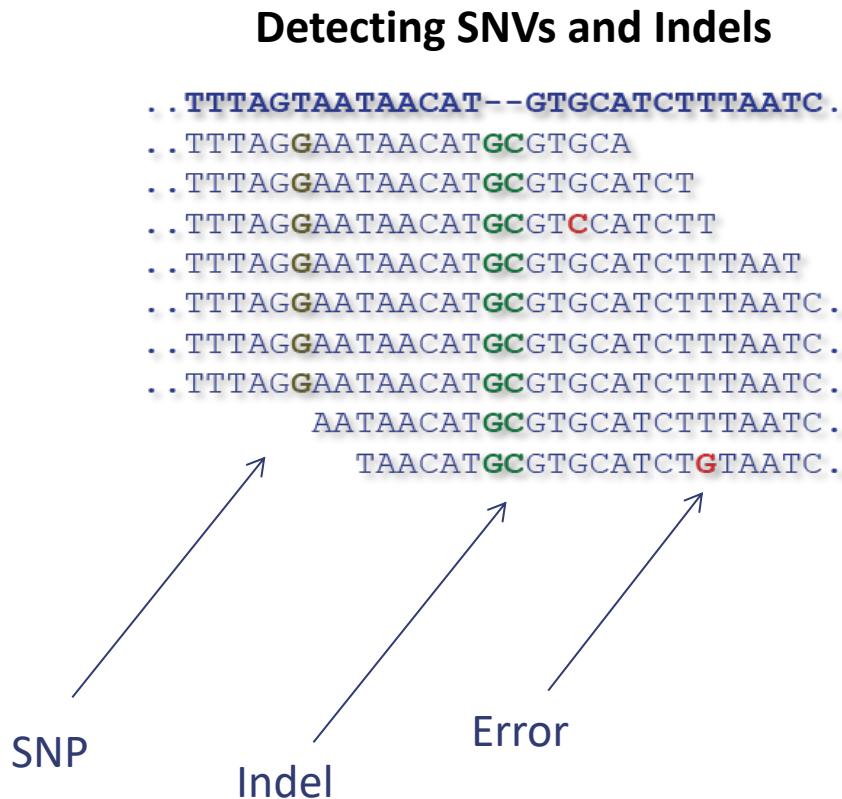


# Workflow for cancer genome analysis

- Detection of somatic mutations in human cancers involves the extraction of DNA from a tumour and, ideally, a matched normal sample.
- Sequencing is usually performed in the paired-end mode with read lengths of 100–150 bp.
- Quality control is performed on the sequencing reads by assessing several metrics before and after alignment to the reference genome.



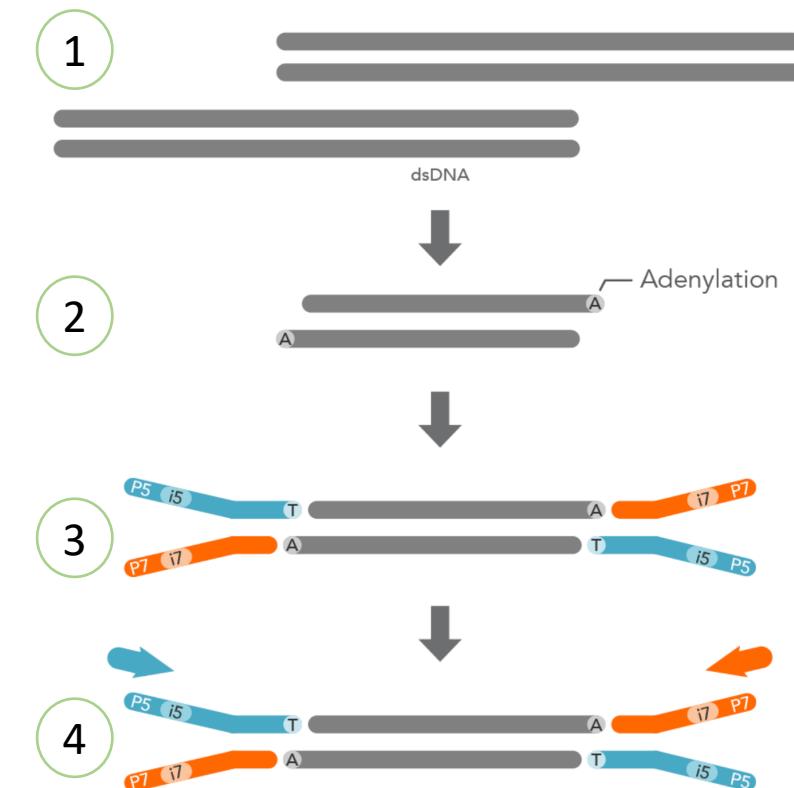
# Workflow for cancer genome analysis



# Library preparation

Library preparation is a critical step in Next-Generation Sequencing (NGS) that involves the conversion of DNA or RNA samples into a format suitable for sequencing.

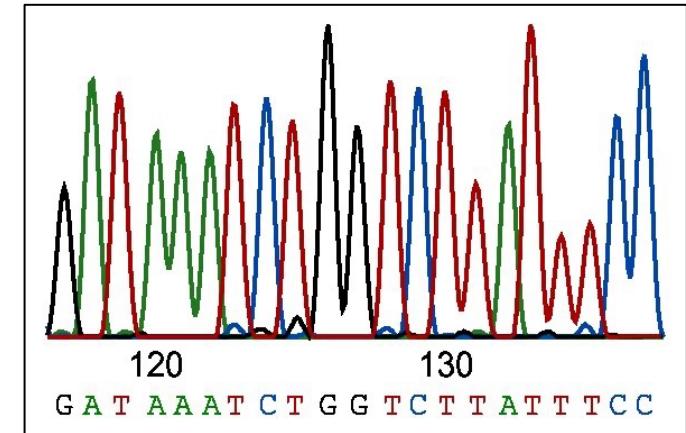
- 1. Fragmentation:** The DNA or RNA sample is fragmented into smaller pieces to create a more manageable size range for sequencing (e.g., sonication, enzymatic methods)
- 2. End Repair and A-Tailing:** The fragmented DNA ends are repaired to generate blunt ends, or in some cases, overhangs with specific adapters
- 3. Adapter Ligation:** Short, synthetic DNA adapters containing sequences complementary to the sequencing platform are ligated to the repaired DNA fragments
- 4. Amplification:** The ligated DNA fragments, now with adapters attached, are amplified using polymerase chain reaction (PCR)
- 5. Quality Control:** The resulting DNA library is evaluated to assess its quality and quantity



# Sequencing technologies

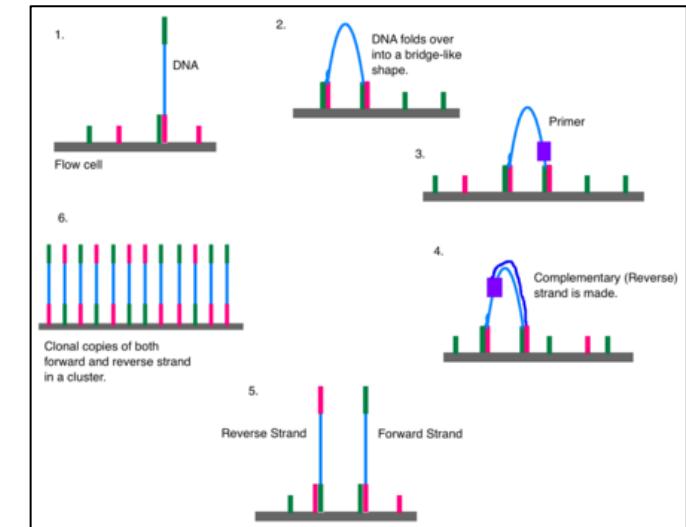
## Sanger Sequencing

- Known as the chain termination method (first widely adopted sequencing technology)
- It relies on enzymatic replication of DNA and incorporates chain-terminating dideoxynucleotides labelled with fluorescent dyes.
- The resulting fragments are separated by capillary electrophoresis, allowing the determination of the DNA sequence.
- Often used for targeted sequencing and validation of genetic variants.



## Next-Generation Sequencing (NGS)

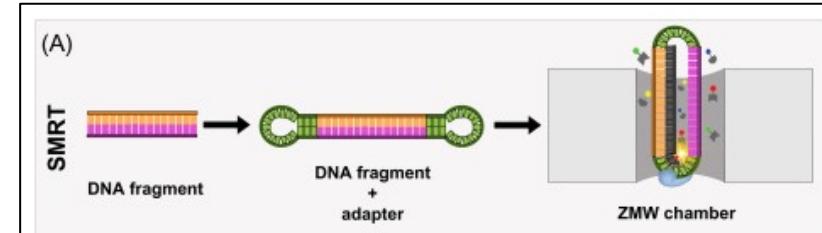
- Encompasses a group of high-throughput sequencing technologies that revolutionized genomic research.
- Illumina sequencing (most used) → massively parallel sequencing by synthesis
- They generate millions to billions of short DNA fragments simultaneously (low error rate)
- NGS offers high-throughput and cost-effective sequencing



# Sequencing technologies

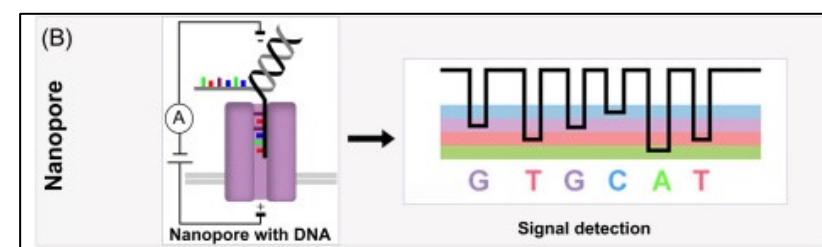
## PacBio Single Molecule Real-Time Sequencing (SMRT)

- Employs a different technology known as single-molecule real-time sequencing.
- Uses specialized DNA polymerases and detects the incorporation of fluorescently labelled nucleotides in real-time as the DNA is being synthesized.
- Enables long-read sequencing → Useful for resolving complex genomic regions and structural variants



## Oxford Nanopore Sequencing

- Portable and real-time sequencing technology.
- It involves passing DNA strands through nanopores, and the changes in electrical current as the DNA molecule passes through the pore are recorded.
- The current signal is then decoded to infer the DNA sequence.
- Nanopore sequencing offers long-read capabilities, allowing the sequencing of long DNA fragments without the need for extensive library preparation.
- It has applications in genomics & metagenomics



# Sequencing technologies

## Summary of the sequencing technologies

|                | 454   | Next generation   |            |             |            | Third generation*  |  |  |
|----------------|---|---|------------|-------------|------------|--|--|--|
|                |   | Illumina  |            |             |            | Ion Torrent  | PacBio   | Oxford nanopore  |
| Platform       | GS FLX+   | HiSeq 2500  | MiSeq      | NextSeq 500 | NovaSeq S4 | PGM 318  | Sequel II System   | MinION R9.4  |
| Run time       | ~ 24 h  | ~ 6 days  | 2-3 days   | 12-30 h     | ~ 44 h     | 4-7 h  | ~ 30 h   | 24 h**   |
| Output / run   | 700 Mb  | 1 Tb  | 15 Gb      | 120 Gb      | 6 Tb       | 2 Gb   | 300 Gb   | 50 Gb  |
| Read length    | 1 kb  | 2 x 125 bp  | 2 x 300 bp | 2 x 150 bp  | 2 x 150 bp | 400 bp   | 10 -16 kb (average)  | 18 kb**  |
| Error rate     | ~ 1 %   | ~ 0.1 %   |            |             |            | ~ 1 %  | ~ 14 %   | ~ 10 % **  |
| Primary errors | Indels  | SNVs  |            |             |            | Indels   | Indels   | Indels   |
| Advantages     | - Long reads<br>- Relative fast run time  | <ul style="list-style-type: none"> <li>- Highest throughput of all platforms and lowest per-base cost.</li> <li>- Low per-base cost</li> <li>- Low error rate</li> </ul>  |            |             |            | <ul style="list-style-type: none"> <li>- Unmodified nucleotides</li> <li>- No optical scanning necessary</li> <li>- Fast run time</li> </ul> | <ul style="list-style-type: none"> <li>- Very long reads</li> <li>- Does not require PCR amplification before sequencing</li> <li>- No bias based on GC content</li> </ul> | <ul style="list-style-type: none"> <li>- Very long reads</li> <li>- Does not require PCR amplification before sequencing</li> <li>- No bias based on GC content</li> <li>-Portable and easy use</li> </ul> |
| Limitations    | <ul style="list-style-type: none"> <li>- High error rate in homopolymers.</li> <li>- Low throughput</li> <li>- High cost</li> <li>- Cumbersome emPCR</li> </ul> | <ul style="list-style-type: none"> <li>-Short reads</li> <li>- Overloading results in overlapping clusters and poor sequence quality</li> <li>- Requirement for sequence complexity. Problems in low-complex regions</li> <li>- Bias in coverage correlated with GC content.</li> </ul> |            |             |            | <ul style="list-style-type: none"> <li>- High error rate in homopolymers</li> <li>- Cumbersome emPCR</li> </ul>                              | <ul style="list-style-type: none"> <li>- High cost</li> <li>- High error rates</li> </ul>  | <ul style="list-style-type: none"> <li>- High error rates, specially in homopolymers and low-complexity regions</li> </ul>   |

\*Many updates in the last years

Adapted from Pfeifer et al 2017 and Ardui et al 2018

# Single-end and paired-end sequencing

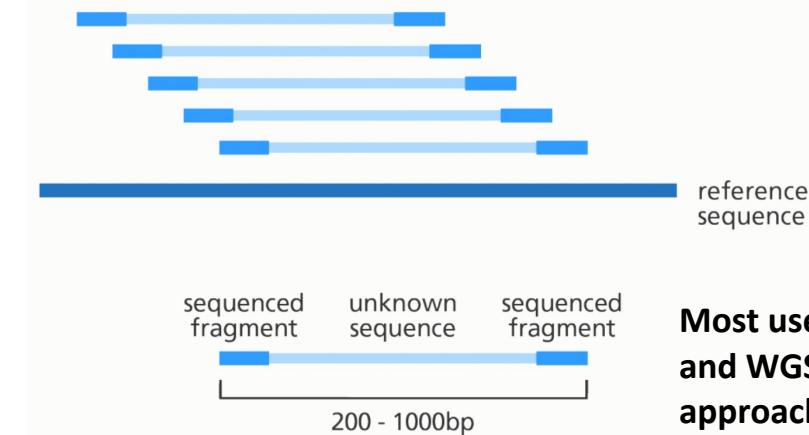
During short read sequencing using NGS platforms such as Illumina, there are two sequencing strategies available for implementation.

## Single-end reads



- Single-read sequencing involves sequencing DNA from only one end → simplest way to utilize Illumina sequencing
- **Cost-effective uses:** This solution delivers large volumes of high-quality data, rapidly and economically
- **Specific applications:** good choice for certain methods like small RNA-Seq or ChIP-Seq

## Paired-end reads



**Most used in WES and WGS approaches**

- Paired-end sequencing allows to sequence both ends of a fragment and generate high-quality, alignable sequence data.
- Paired-end DNA sequencing also detects common DNA rearrangements such as insertions, deletions, and inversions

# Limitations at different steps of the variant sequencing workflow

Each step of this workflow has particular limitations and issues, which might affect the downstream analysis and interpretation of the calls

## Library preparation and sequencing

### Oxidative damage during acoustic shearing

C to A  
G to T

## Read pre-processing and alignment

### PCR error rate *Taq Polimerase*

$1-20 \times 10^{-5}$

## Somatic/Germline variant calling SNPs, Indels, SVs, CNVs

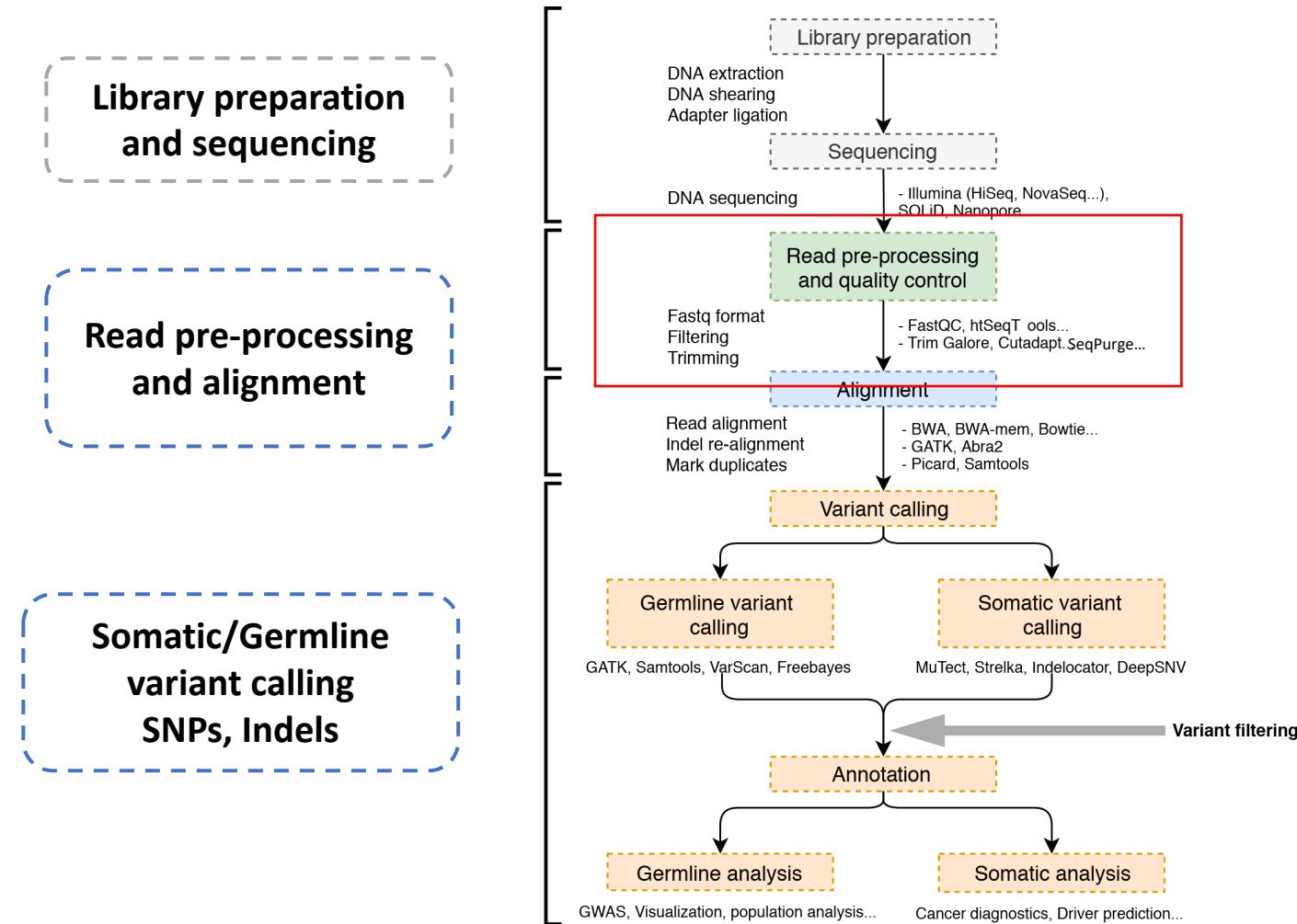
### Sequencing errors

Crosstalk  
A to C | G to T

Dephasing  
Indels

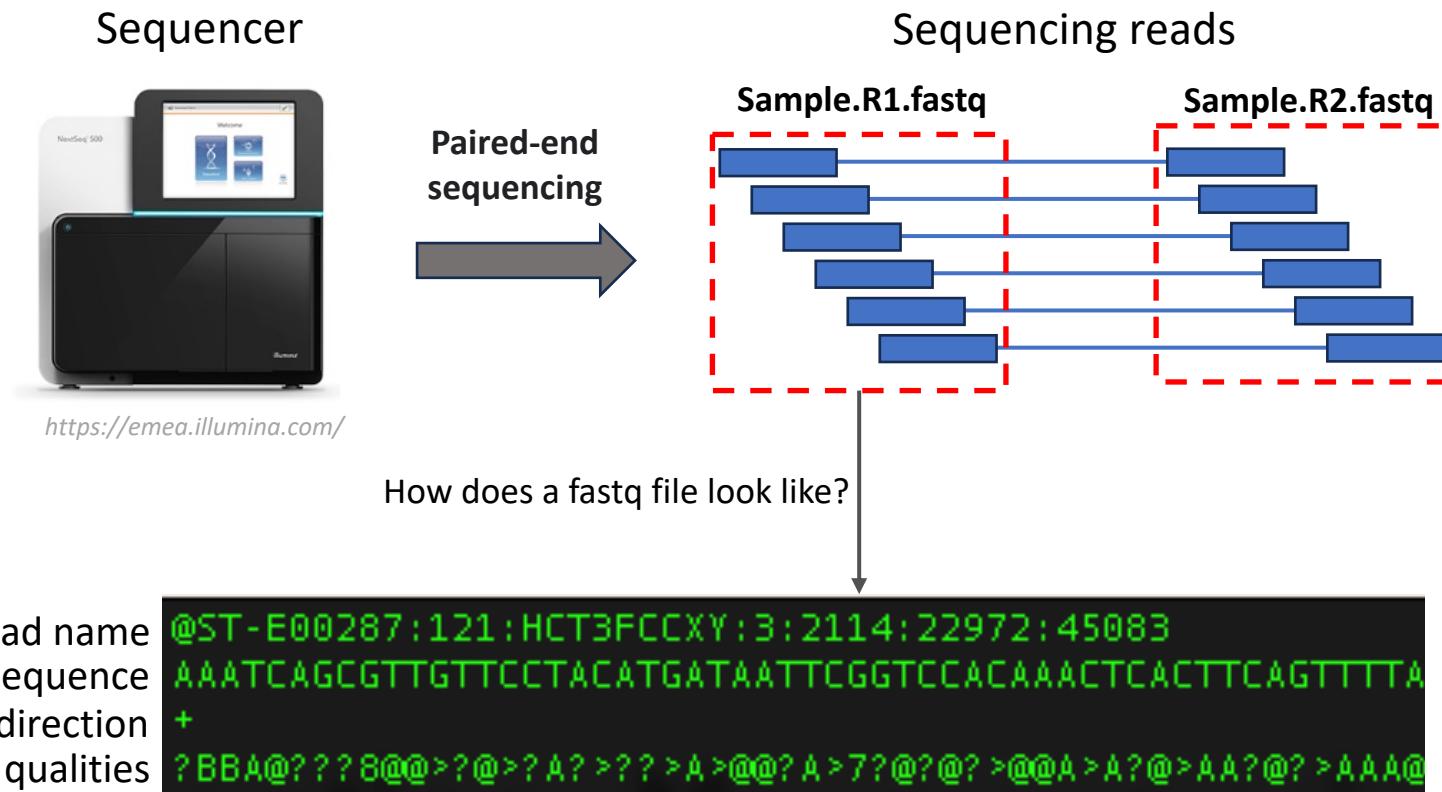
Low coverage  
GC content

# Read processing



# Fastq files

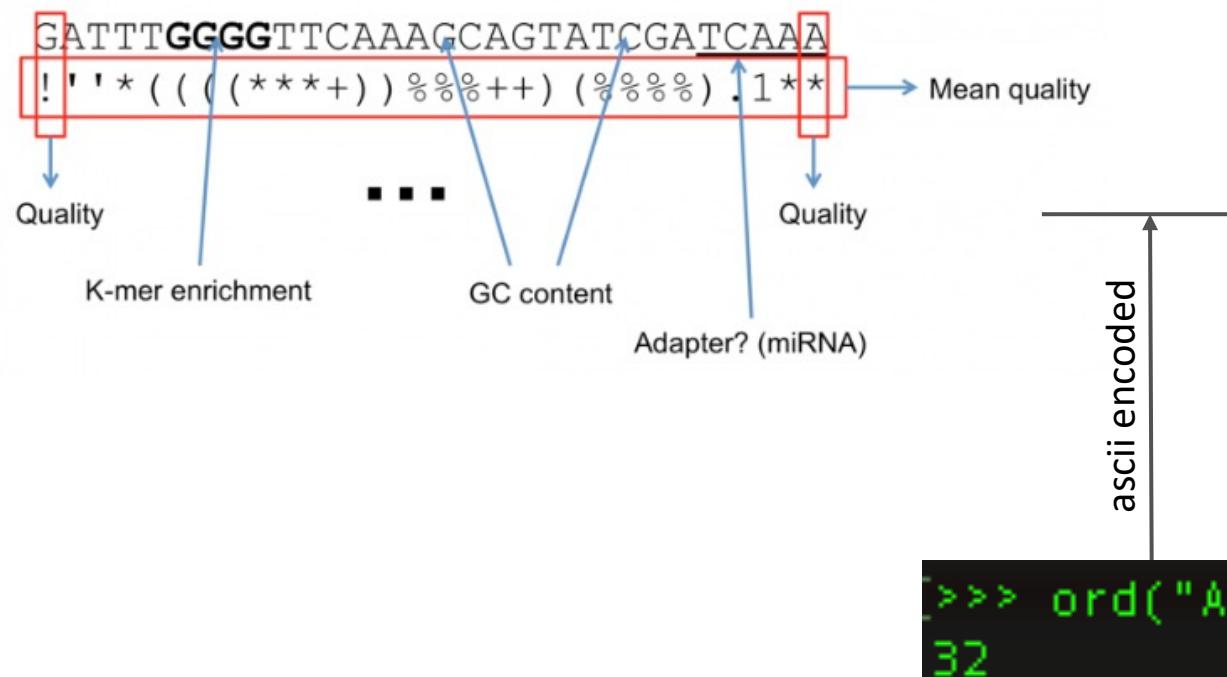
- A FASTQ file is a commonly used file format in bioinformatics for storing DNA or RNA sequencing data
- It contains both the sequence information and corresponding quality scores for each nucleotide in the sequence.



Each sample will generate (depending on the read length) ~ 5Gb (100X WES) to ~ 300 Gb (100X WGS) of data

# Fastq files

**Base quality (Phread Quality Score):** The quality value ( $Q$ ) is a representation of the probability ( $p$ ) that a base call is incorrect



Example in python to transform letter quality to numeric quality value

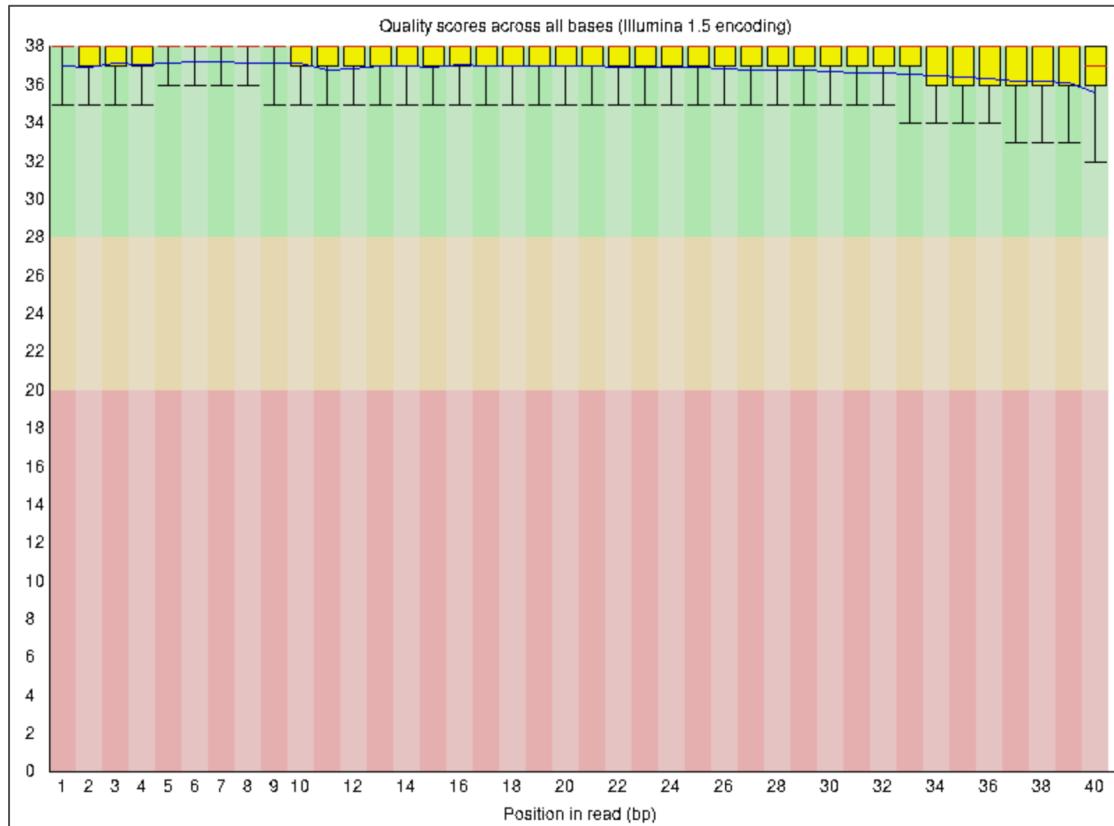
$$Q = -10 \log_{10} P \rightarrow P = 10^{\frac{-Q}{10}}$$

| Phred Quality Score | Probability of incorrect base call | Base call accuracy |
|---------------------|------------------------------------|--------------------|
| 10                  | 1 in 10                            | 90%                |
| 20                  | 1 in 100                           | 99%                |
| 30                  | 1 in 1000                          | 99.9%              |
| 40                  | 1 in 10000                         | 99.99%             |
| 50                  | 1 in 100000                        | 99.999%            |

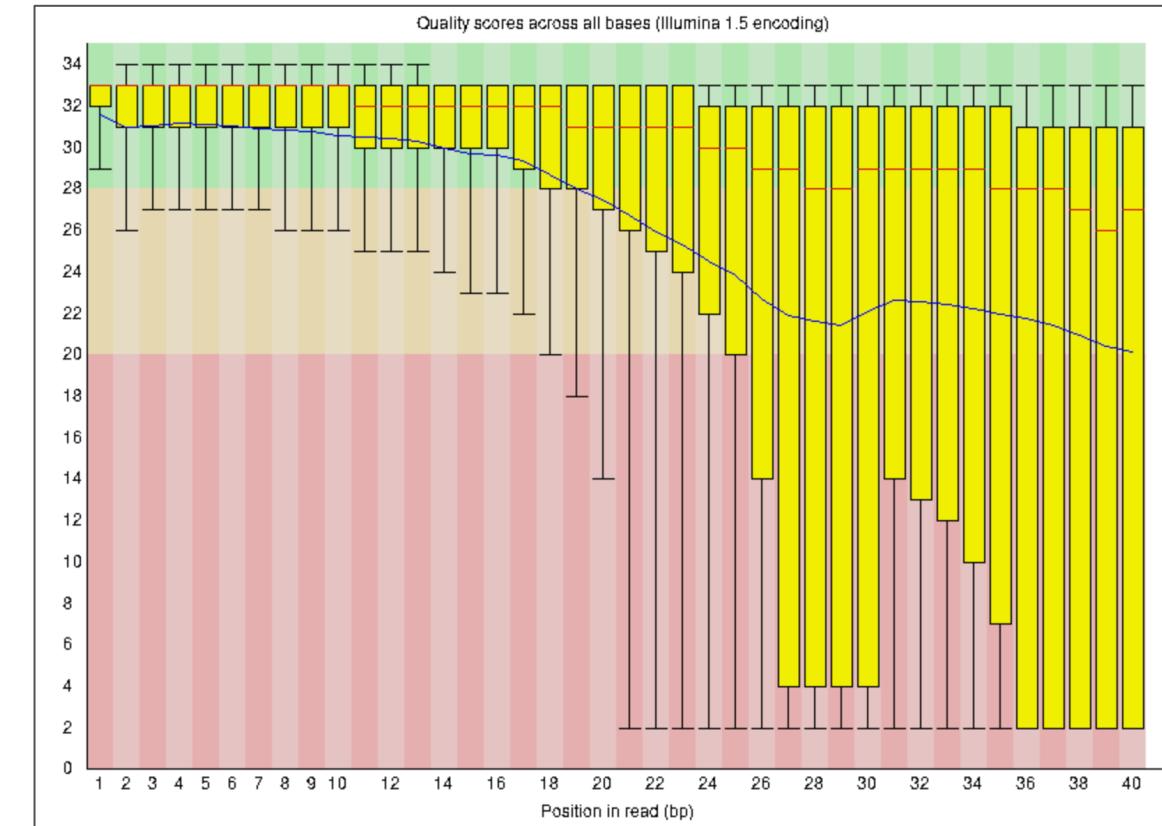
# QC of raw sequences

- Low quality bases or low-quality reads can bias all the downstream analysis (alignment, SNV and indel calling...)
- This step can be performed with the **FastQC software**

Good quality illumina data



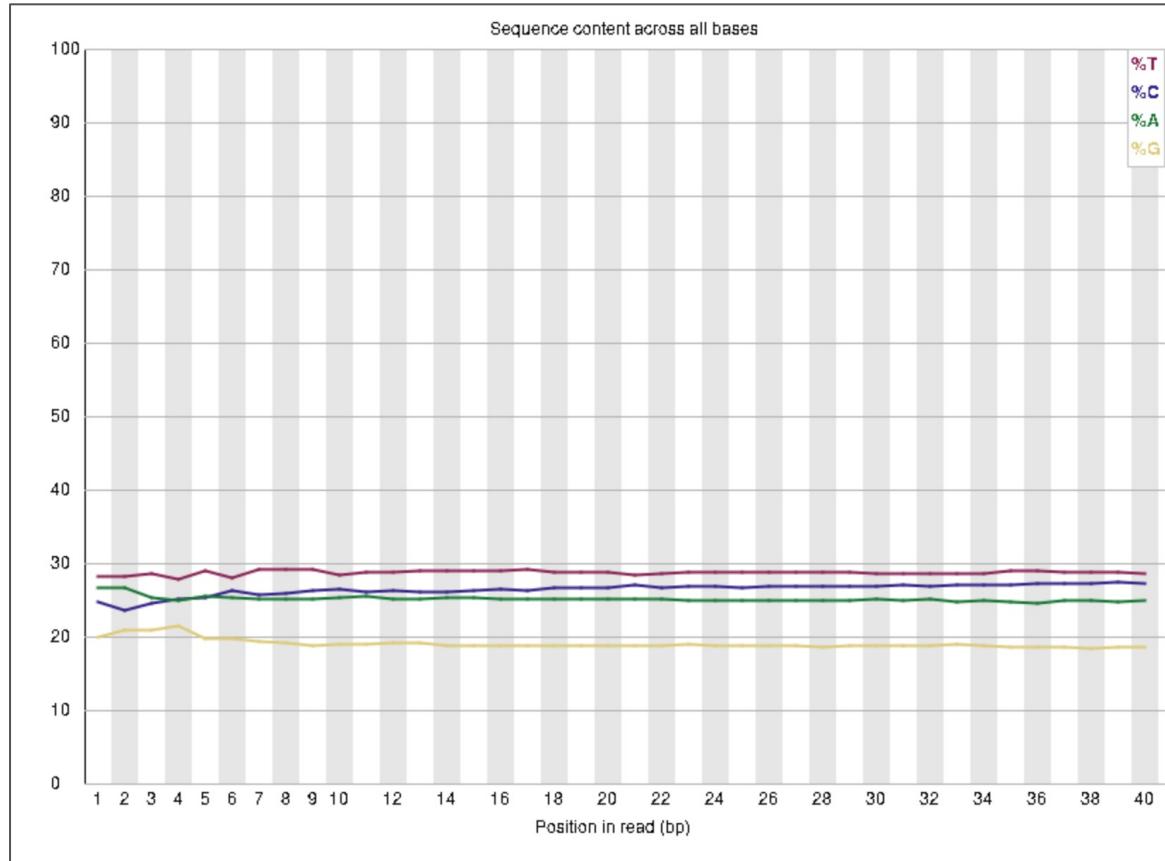
Bad quality illumina data



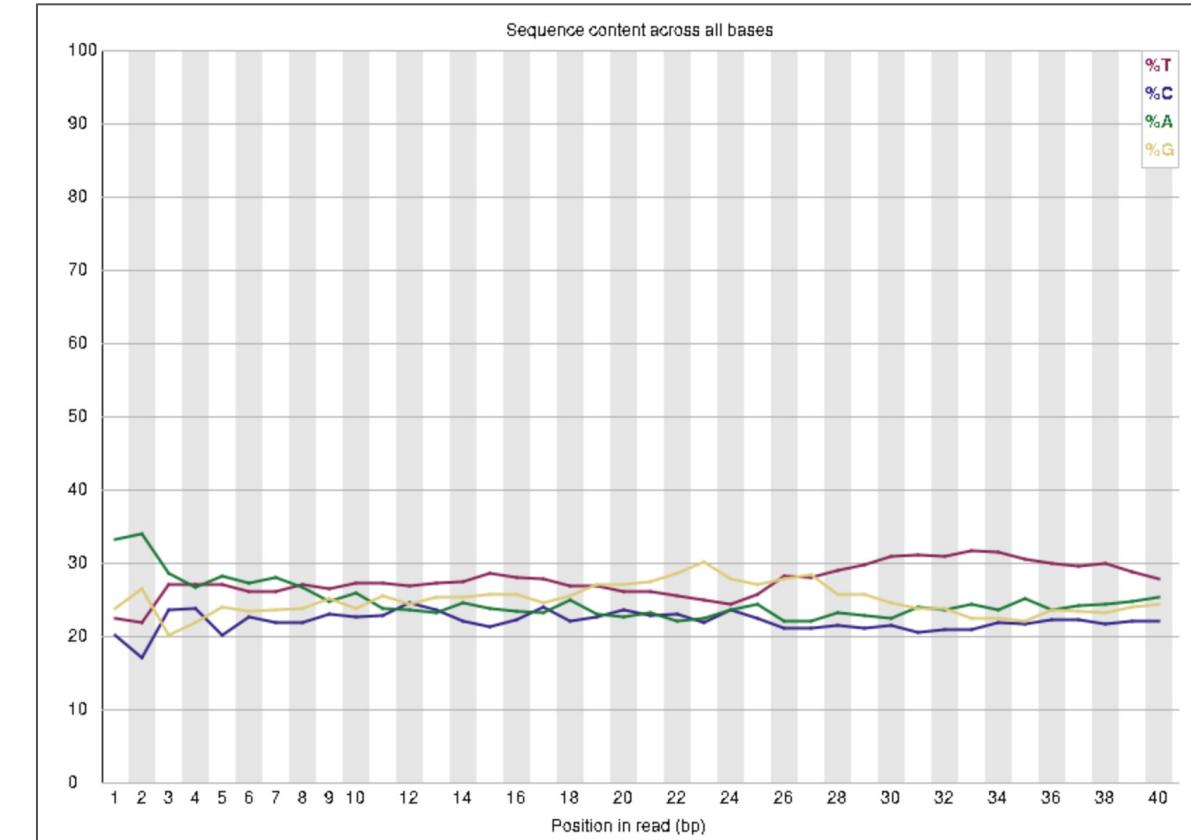
# QC of raw sequences

Position base-content is usually consistent and uniform across the read

Good quality illumina data



Bad quality illumina data



# QC of raw sequences

Some samples have Illumina adapter contamination

## ✖ Overrepresented sequences

| Sequence                                 | Count | Percentage          | Possible Source  |
|--|-------|---------------------|--|
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCT | 8122  | 8.122               | Illumina Paired End PCR Primer 2 (100% over 40bp)        |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGATCGGAAG | 5086  | 5.086               | Illumina Paired End PCR Primer 2 (97% over 36bp)         |
| AATGATAACGGCGACCACCGAGATCTACACTCTTCCCTAC | 1085  | 1.085               | Illumina Single End PCR Primer 1 (100% over 40bp)        |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGGAAG | 508   | 0.508               | Illumina Paired End Sequencing Primer 2 (100% over 36bp) |
| AATTATAACGGCGACCACCGAGATCTACACTCTTCCCTAC | 242   | 0.242               | Illumina Single End PCR Primer 1 (97% over 40bp)         |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAAGATCGGAA | 235   | 0.23500000000000001 | Illumina Paired End Adapter 2 (96% over 31bp)            |
| GATCGGAAGAGCGGTTCAGCAGGAATGCGAGATCGGAAGA | 228   | 0.22799999999999998 | Illumina Paired End Adapter 2 (96% over 28bp)            |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGGACG | 205   | 0.20500000000000002 | Illumina Paired End Sequencing Primer 2 (100% over 36bp) |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGGTGCGAAG | 183   | 0.183               | Illumina Paired End Adapter 2 (100% over 32bp)           |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGGATCGGAA | 183   | 0.183               | Illumina Paired End Adapter 2 (100% over 32bp)           |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGAACT | 164   | 0.164               | Illumina Paired End PCR Primer 2 (97% over 40bp)         |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGGTCT | 129   | 0.129               | Illumina Paired End PCR Primer 2 (97% over 40bp)         |
| AATTATACTTCTACCAACCTATATCTACACTCTTCCCTAC | 123   | 0.123               | No Hit   |
| GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGGACT | 122   | 0.122               | Illumina Paired End Sequencing Primer 2 (100% over 36bp) |
| CGGTTCAGCAGGAATGCCGAGATCGGAAGAGCGGTTCA   | 113   | 0.11299999999999999 | Illumina Paired End PCR Primer 2 (96% over 25bp)         |

These regions need to be trimmed before the alignment

Forward read →



← Reverse read

Adapter      Original sequence

Many tools available that can do this trimming:

- Cutadapt
- SeqPurge
- Trimmomatic
- ...

# Alignment

Read mapping, also known as alignment or mapping, is a process in computational genomics that involves aligning or matching short DNA/RNA sequences (reads) to a reference genome (in our case, the human reference genome **Hg38**)

## Mapping problems

- Mapping millions of short reads to a genome
- The genome is very large
- The human genome is highly repetitive (~45%) → many possible mapping locations
- Highly affected by sequencing errors



## Many algorithms (tools)

- **BWA\***
- Novoalign
- Bowtie2
- GEM

## Why do we use BWA(-mem)?

- It uses indexing technique called the Burrows-Wheeler transform to create a compressed version of the reference genome
- When aligning reads to the reference genome using BWA, it breaks each read into smaller fragments called seeds.
- Faster, less memory usage and same performance

# Alignment

- **SAM** (Sequence Alignment/Map) and **BAM\*** (Binary Alignment/Map) files are commonly used file formats in bioinformatics for storing DNA/RNA sequencing alignment data (between 10Gb and 500Gb per file)
- We obtain (in general) one for the tumour and another for the matched normal sample

| @HD VN:1.5 SO:coordinate |      |     |    |    |            |   |    |     |                   | Header section             |
|--------------------------|------|-----|----|----|------------|---|----|-----|-------------------|----------------------------|
| @SQ SN:ref LN:45         |      |     |    |    |            |   |    |     |                   |                            |
| r001                     | 99   | ref | 7  | 30 | 8M2I4M1D3M | = | 37 | 39  | TTAGATAAAGGATACTG | *                          |
| r002                     | 0    | ref | 9  | 30 | 3S6M1P1I4M | * | 0  | 0   | AAAAGATAAGGATA    | *                          |
| r003                     | 0    | ref | 9  | 30 | 5S6M       | * | 0  | 0   | GCCTAACGCTAA      | * SA:Z:ref,29,-,6H5M,17,0; |
| r004                     | 0    | ref | 16 | 30 | 6M14N5M    | * | 0  | 0   | ATAGCTTCAGC       | *                          |
| r003                     | 2064 | ref | 29 | 17 | 6H5M       | * | 0  | 0   | TAGGC             | * SA:Z:ref,9,+,5S6M,30,1;  |
| r001                     | 147  | ref | 37 | 30 | 9M         | = | 7  | -39 | CAGCGGCAT         | * NM:i:1                   |

Optional fields in the format of TAG:TYPE:VALUE

**QUAL:** read quality; \* meaning such information is not available

**SEQ:** read sequence

**TLEN:** the number of bases covered by the reads from the same fragment. Plus/minus means the current read is the leftmost/rightmost read. E.g. compare first and last lines.

**PNEXT:** Position of the primary alignment of the NEXT read in the template. Set as 0 when the information is unavailable. It corresponds to POS column.

**RNEXT:** reference sequence name of the primary alignment of the NEXT read. For paired-end sequencing, NEXT read is the paired read, corresponding to the RNAME column.

**CIGAR:** summary of alignment, e.g. insertion, deletion

**MAPQ:** mapping quality

**POS:** 1-based position

**RNAME:** reference sequence name, e.g. chromosome/transcript id

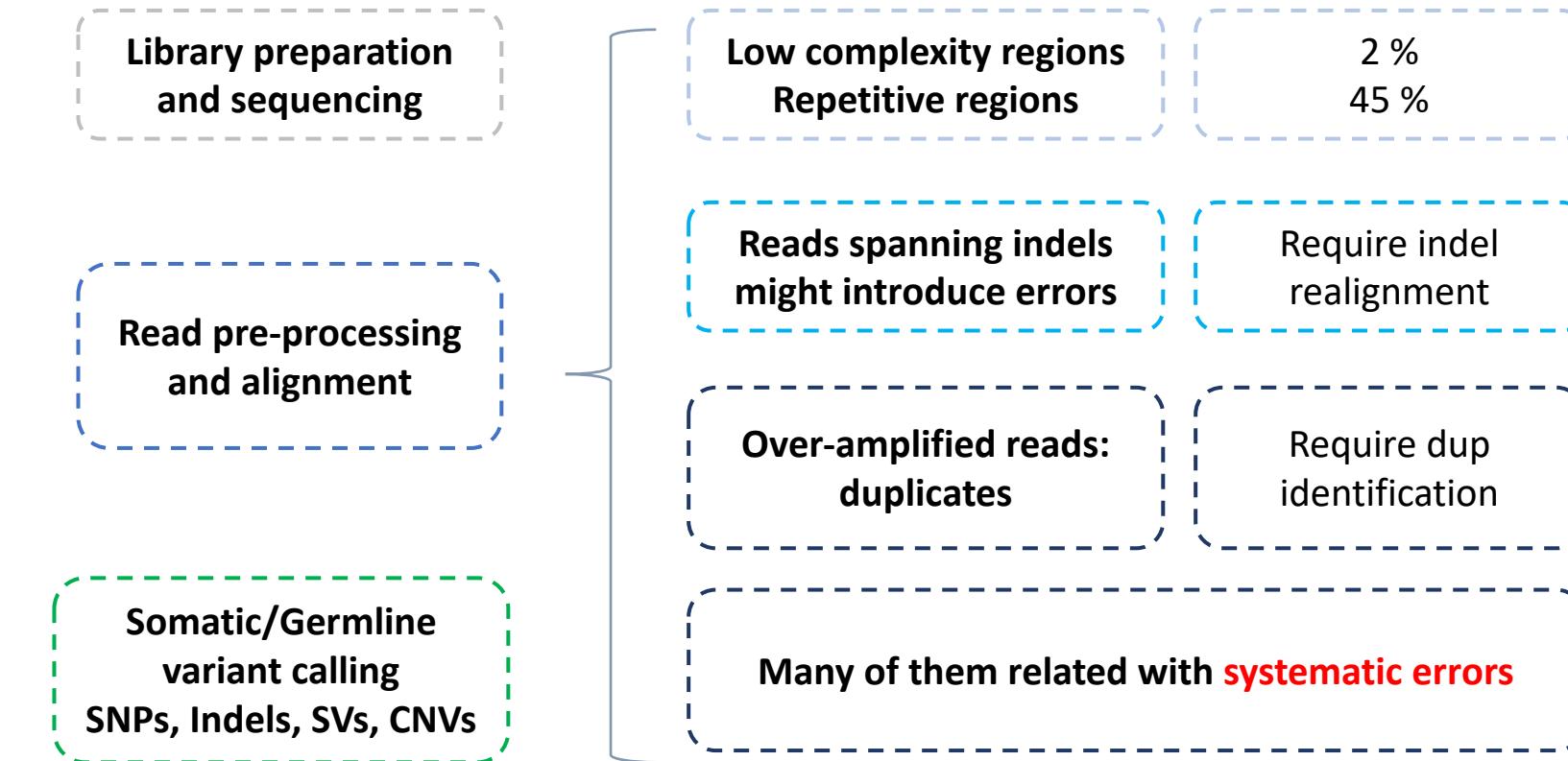
**FLAG:** indicates alignment information about the read, e.g. paired, aligned, etc.

**QNAME:** query template name, aka. read ID

\*It also exists CRAM:  
compressed version of BAM  
(0.2-0.8 of bam size)

# Limitations at different steps of the variant sequencing workflow

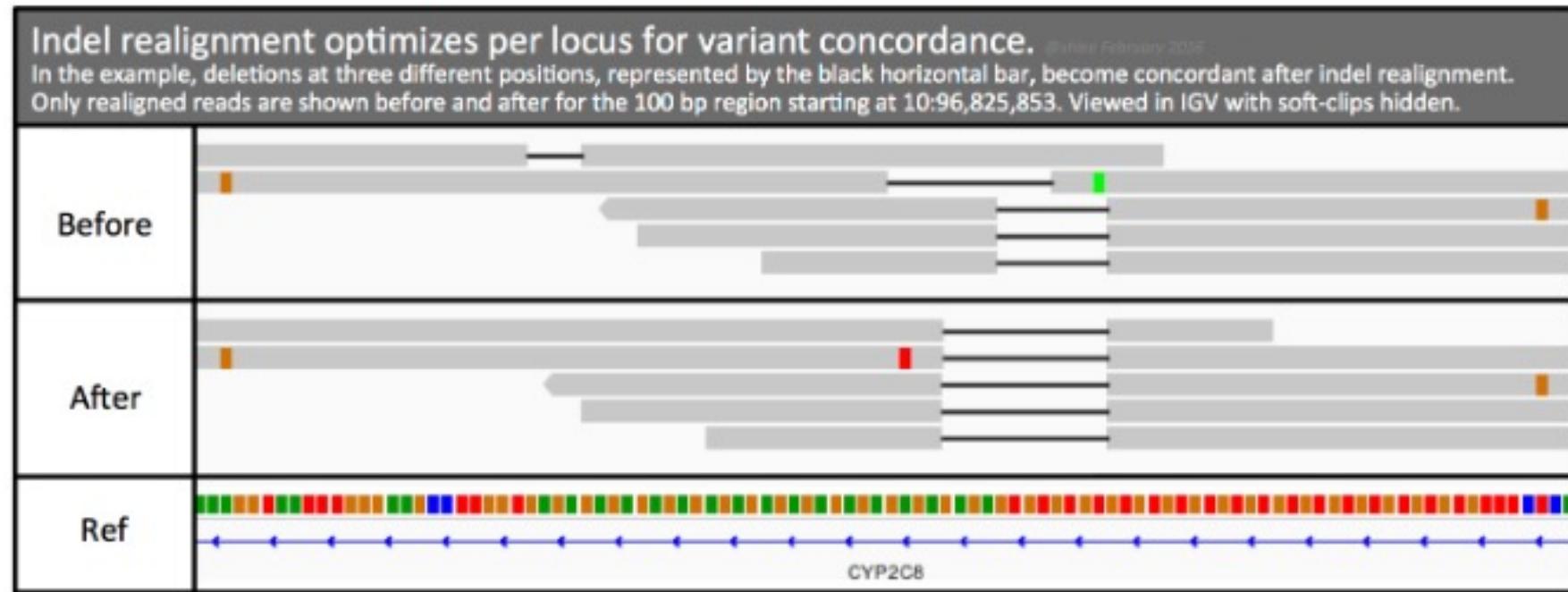
Each step of this workflow has particular limitations and issues, which might affect the downstream analysis and interpretation of the calls



# Processing bam files for variant calling

## Indel realignment

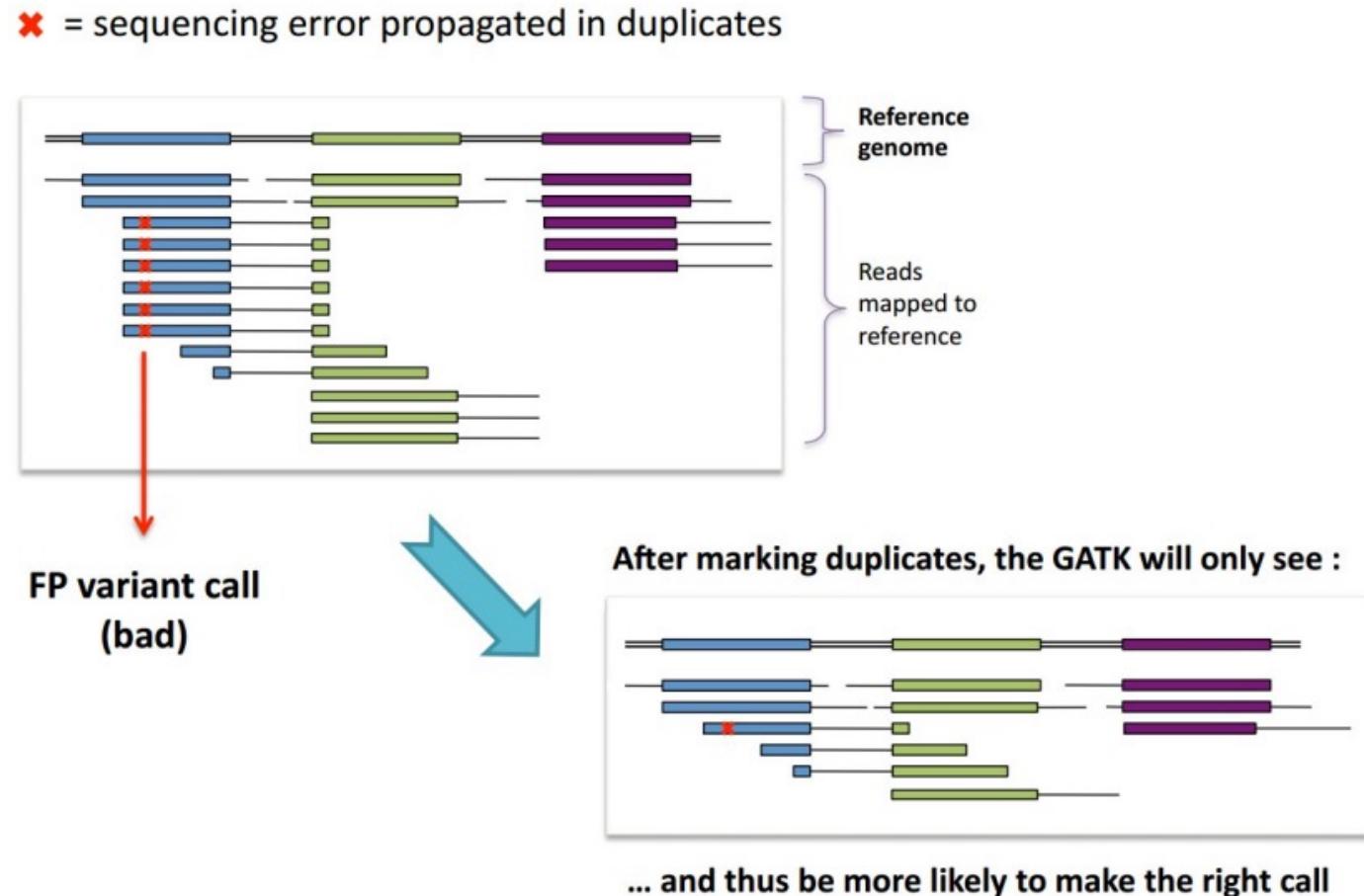
- During the sequencing process, errors can occur, leading to misalignments around indel regions.
- These misalignments can result in false-positive or false-negative variant calls.
- Indel realignment aims to correct these misalignments by identifying and repositioning reads that span indel sites.
- In the past, it was an individual step in GATK3. In the new versions of GATK4, it is already implemented in the variant calling algorithm



# Processing bam files for variant calling

## Marking duplicates

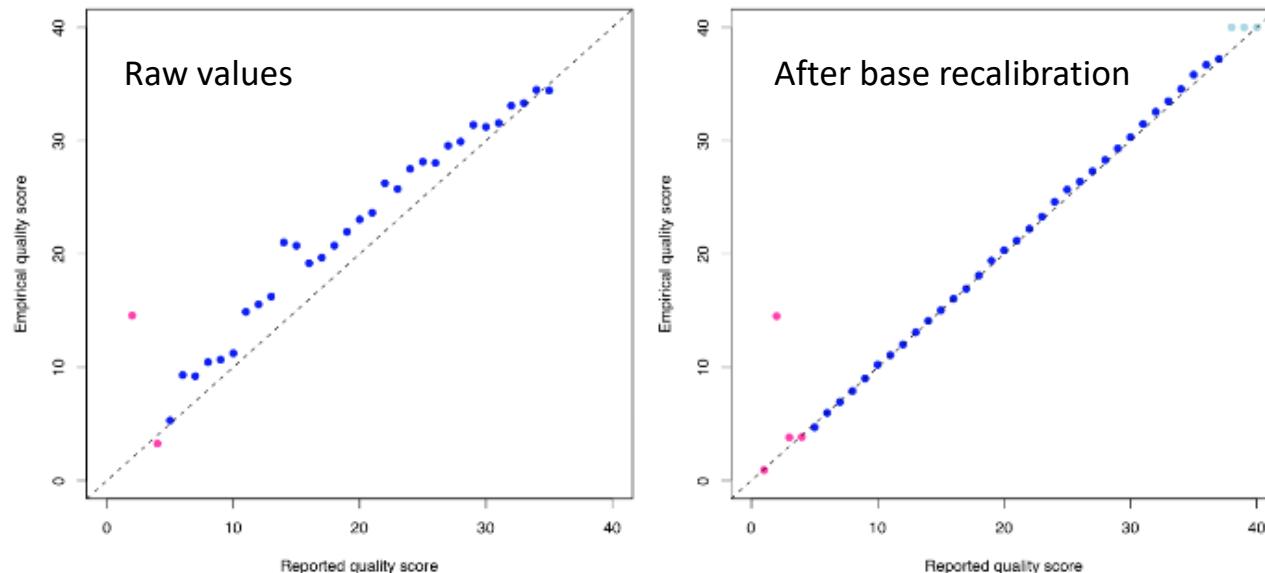
- Read duplicates refer to multiple identical or nearly identical DNA sequencing reads that arise from the same DNA fragment rather than representing distinct biological molecules.
- These duplicates can occur due to biases in library preparation and amplification steps (PCR).
- Understanding and identifying read duplicates is important in order to accurately interpret sequencing data and avoid potential biases in downstream analyses.
- Available tools: Picard, GATK4, SAMtools...



# Processing bam files for variant calling

## Base recalibration

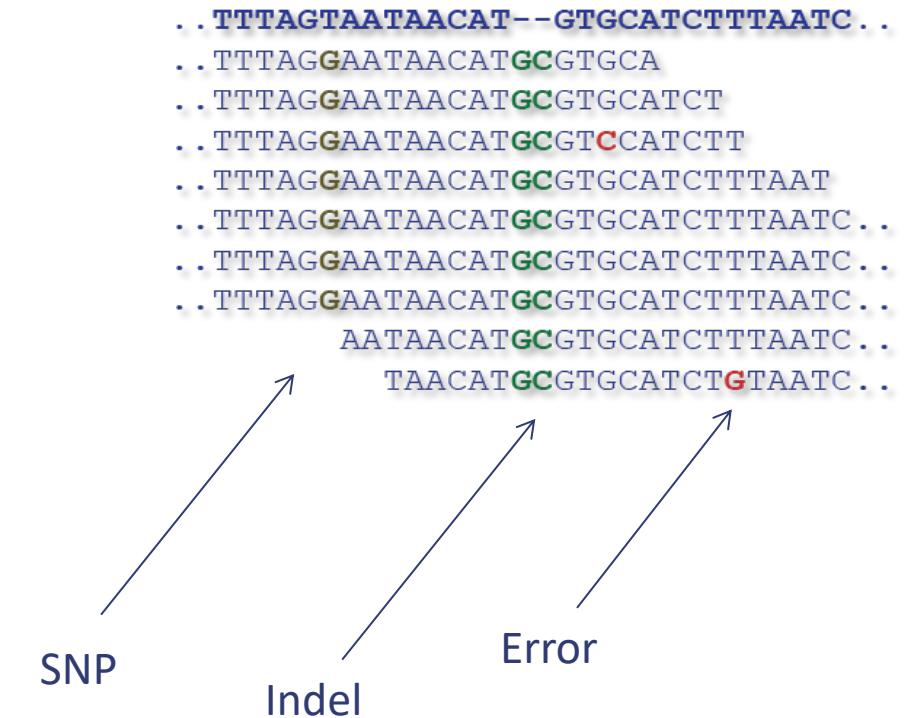
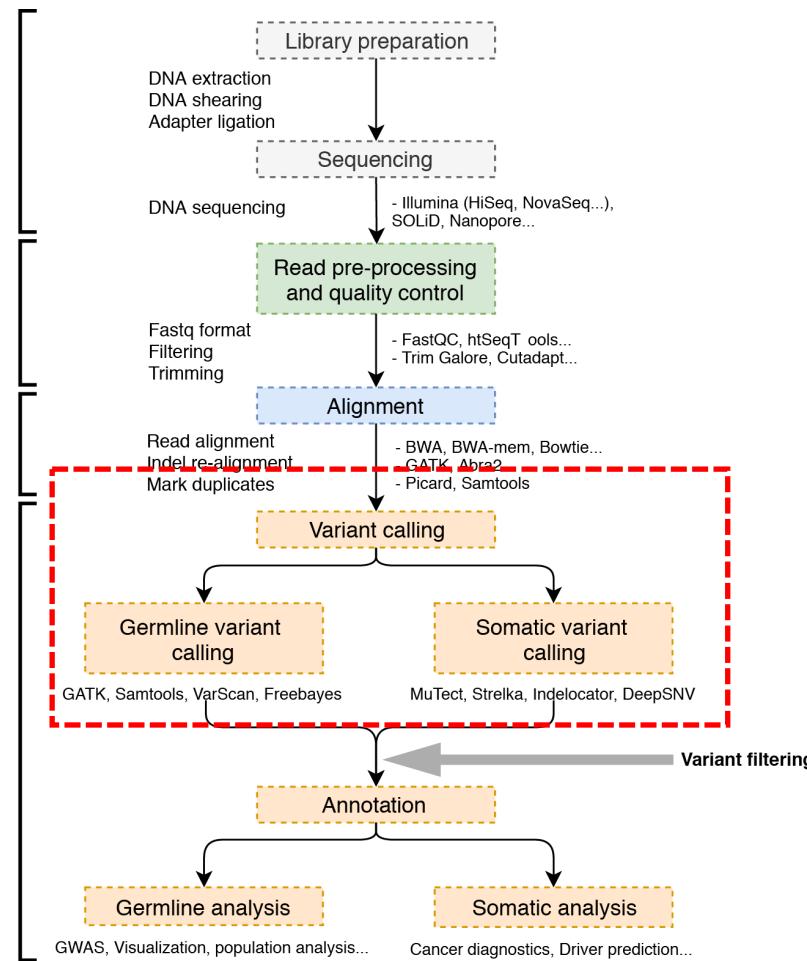
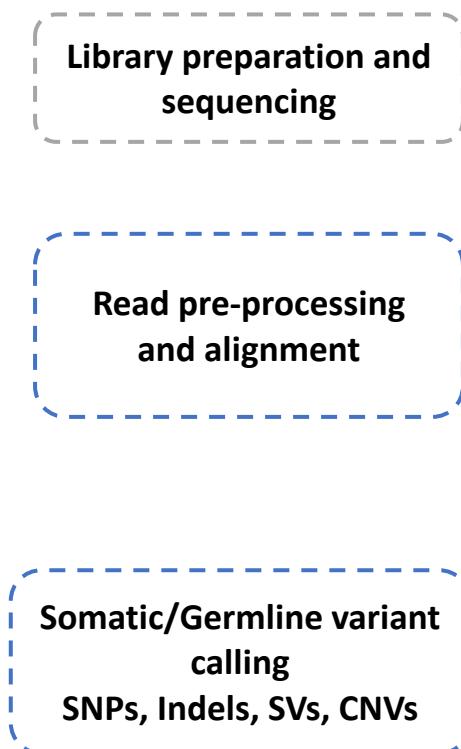
- The scores produced by the machines are subject to various sources of systematic (non-random) technical error, leading to over- or under-estimated base quality scores in the data.
- Base recalibration is a data pre-processing step that detects systematic errors made by the sequencing machine when it estimates the accuracy of each base call
- Base quality score recalibration (BQSR, GATK) is a process in which we apply machine learning to model these errors empirically and adjust the quality scores accordingly. (Recommended for GATK variant calling pipelines)



Source: <https://gatk.broadinstitute.org/>

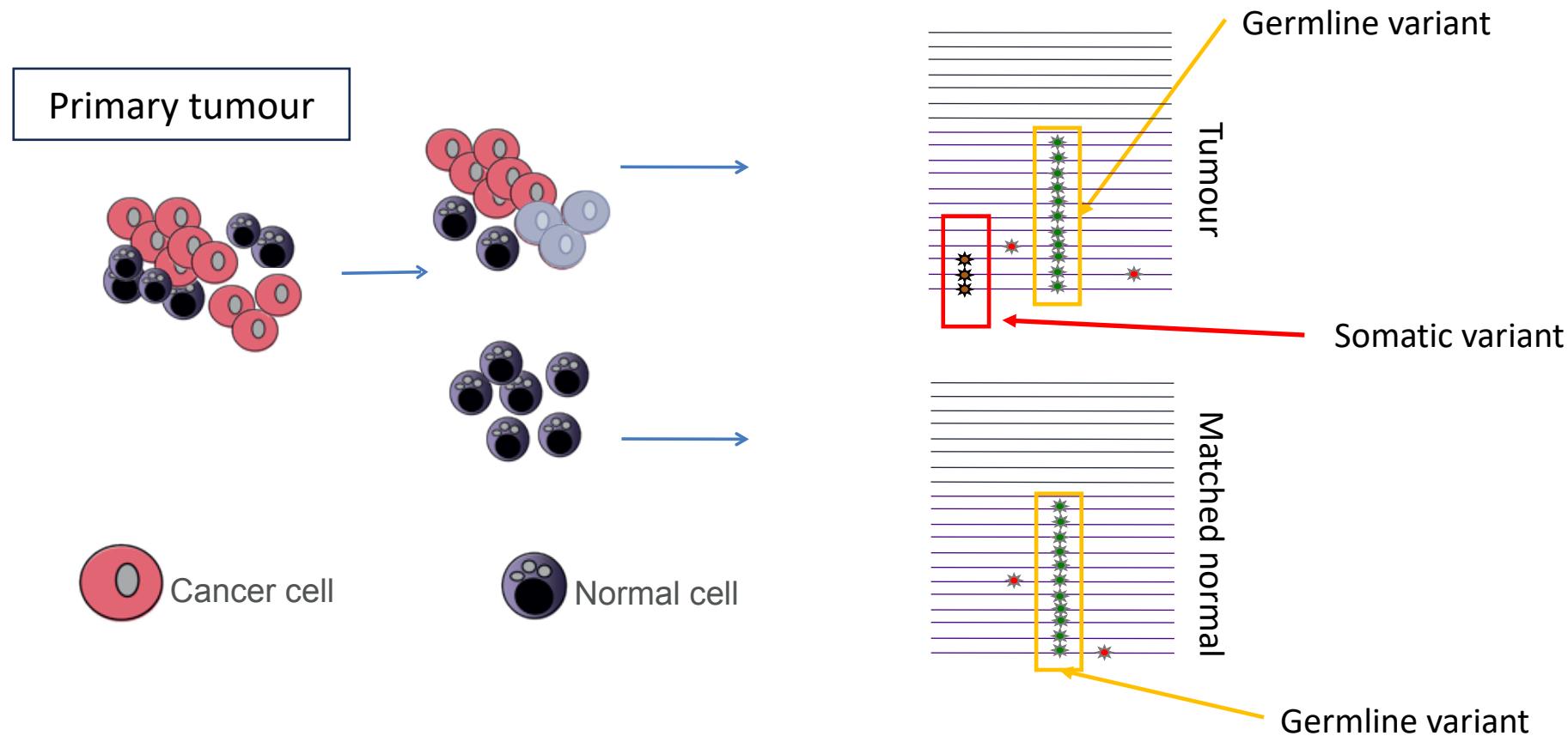
# Variant calling (VC)

- VC refers to the process of identifying genetic variations or variants present in an individual's DNA sequence.
- VC compares the sequenced DNA reads to a reference genome and determines where the individual's sequence differs from the reference.



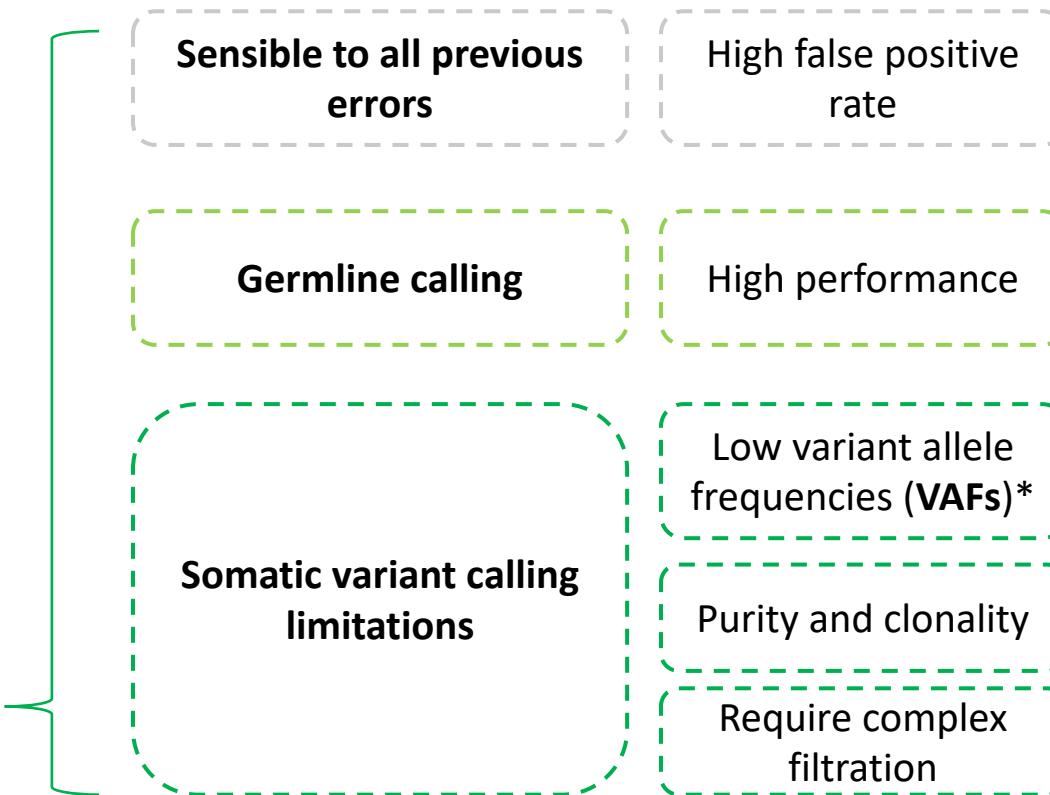
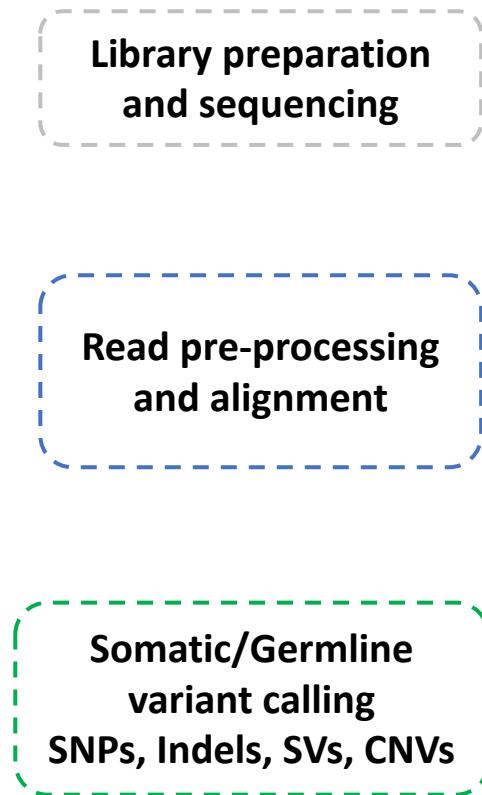
# Somatic variant calling

- Somatic variant calling typically involves sequencing both tumour and matched normal samples from an individual.
- The normal sample serves as a control, representing the non-mutated or germline cells, while the tumour sample contains genetic alterations specific to the tumour cells.

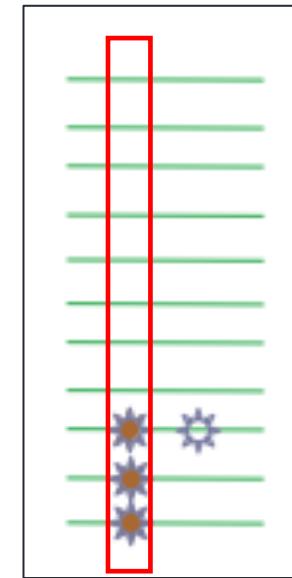


# Somatic variant calling

## Limitations



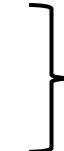
\***VAF (Variant allele frequency)**  
Fraction of reads supporting the variant



Huge importance of **read coverage**

# Somatic variant calling

## Available tools for somatic variant calling

- MuTect2 (part of GATK)
  - Strelka2
  - MuSe2
  - VarScan2
  - SAMtools
  - VarDict
  - Somatic Sniper
  - Scalpel (indels in WES data)
  - ...
- 
- Tools we will learn  
to run today

Please! Don't use germline  
callers for somatic variant  
detection

# Somatic variant calling

## Variant filtering

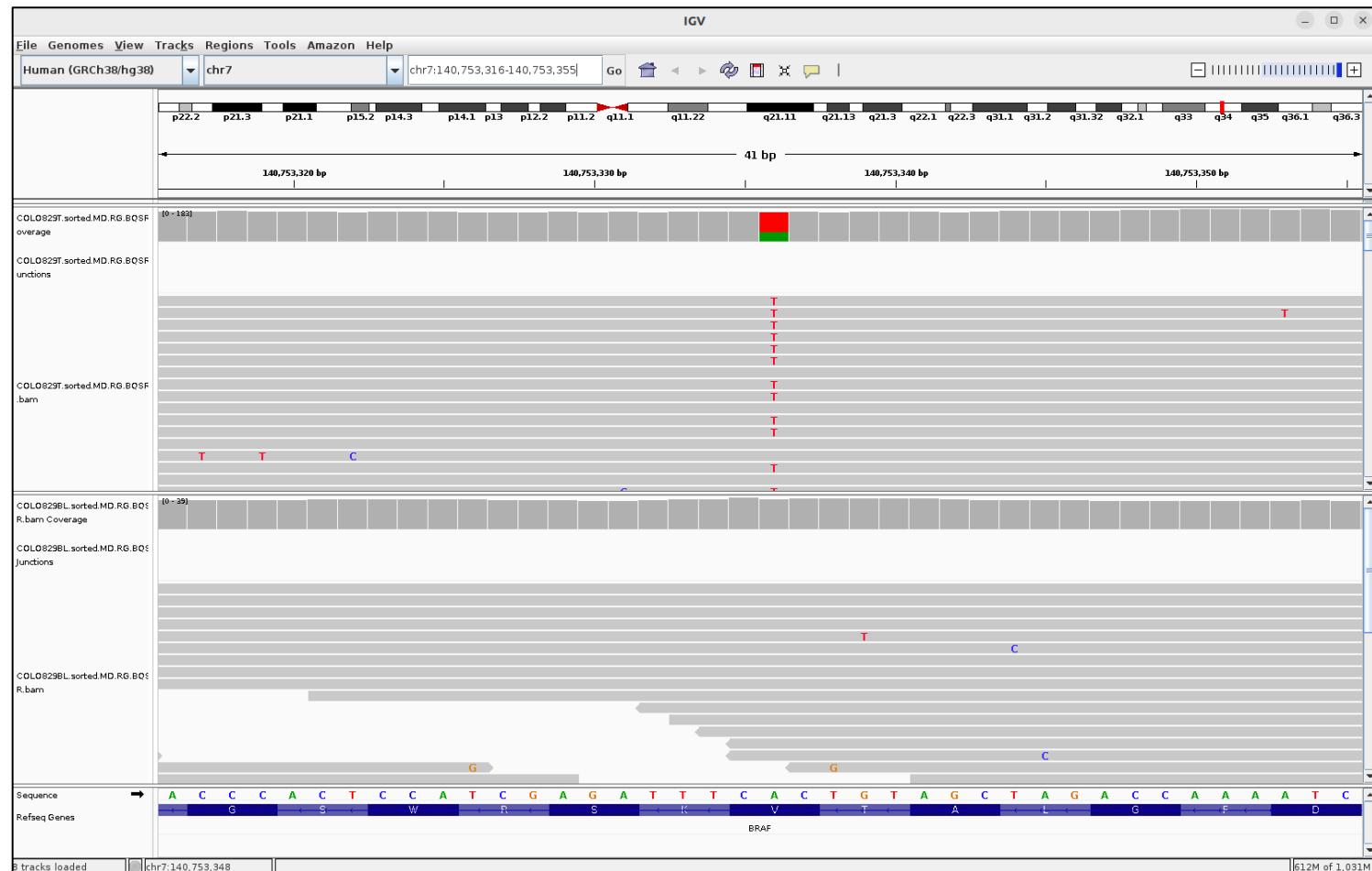
- Raw variant calls usually have lots of false positive calls
- Many algorithms have their own set of internal filters, but some post-processing is sometimes required
- Filters like minimum **VAF**, **coverage**, blacklist regions are sometimes useful
- Some tools like GATK (*GATK VariantRecalibrator*) can rank the mutations based on the quality

Some of these filters are usually applicable if sufficient number of samples can be analysed in the same batch (It will not be performed today)

# Somatic variant calling

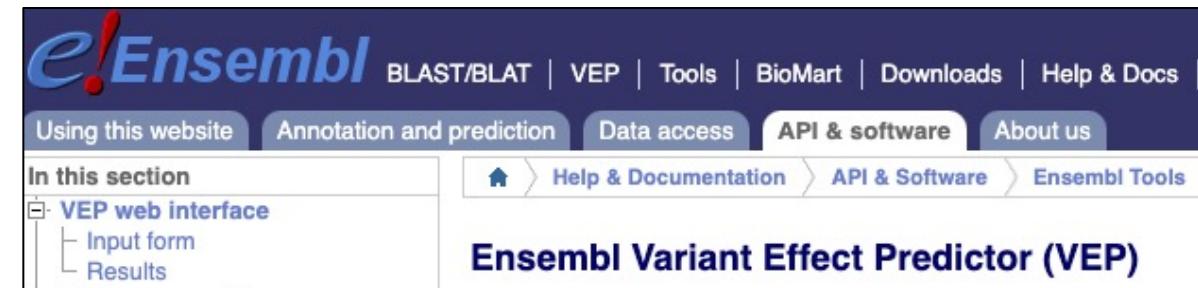
## Variant visualization

IGV (Integrative Genomics Viewer) software is very useful to check how the detected variants look in the alignment file (bam)



# Variant annotation

- Detected variants can be annotated with additional information, such as functional consequences (e.g., synonymous or nonsynonymous mutations), predicted effects on proteins, and known associations with diseases.
- This annotation provides insights into the potential impact and significance of the variants.
- Some available tools:
  - **ANNOVAR** (ANNOtate VARiation)
  - **VEP** (Variant Effect Predictor)
  - **SnpEff** (SNP Effects)



ANNOVAR Documentation ANNOVAR User Guide ▾ Misc ▾ Articles ▾ Q Search

[ANNOVAR Documentation](#) ANNOVAR Documentation

# SnpEff & SnpSift

Genomic variant annotations and functional effect prediction toolbox.

# Variant annotation

## ANNOVAR output

- It provides insane amount of information.
  - Very flexible with the type of datasets you can use annotate variants
  - Gene affected, location, type of mutation, amino acid change, band location, variant population frequency (ExAC, GnomAD...), impact ...

This is a small part of the columns of the file we will generate today...

# Variant annotation

## ANNOVAR output (mutation impact)

- It also provides the **mutation impact** obtained by LJB dbNSFP (on non-synonymous variants annotation) included in the ANNOVAR output
- The LJB databases include several impact scores\*
- It also can include *ClinVar*\*: check if mutation is known to have a clinical impact (<https://www.ncbi.nlm.nih.gov/clinvar/intro/>) (Not used today)
- Check *Intogen* (<https://www.intogen.org/>) to see how frequent the genes are affected in each cancer type
- Check Cosmic (<https://cancer.sanger.ac.uk/census>) to see the list of Cancer Gene Census (CGC).

| Score (dbtype)             | # variants in LJB23 build hg19 | Categorical Prediction  |
|----------------------------|--------------------------------|---|
| SIFT (sift)                | 77593284                       | D: Deleterious (sift<=0.05); T: tolerated (sift>0.05)   |
| PolyPhen 2 HDIV (pp2_hdiv) | 72533732                       | D: Probably damaging (>=0.957), P: possibly damaging (0.453<=pp2_hdiv<=0.956); B: benign (pp2_hdiv<=0.452)      |
| PolyPhen 2 HVar (pp2_hvar) | 72533732                       | D: Probably damaging (>=0.909), P: possibly damaging (0.447<=pp2_hdiv<=0.909); B: benign (pp2_hdiv<=0.446)      |
| LRT (lrt)                  | 68069321                       | D: Deleterious; N: Neutral; U: Unknown  |
| MutationTaster (mt)        | 88473874                       | A" ("disease_causing_automatic"); "D" ("disease_causing"); "N" ("polymorphism"); "P" ("polymorphism_automatic") |
| MutationAssessor (ma)      | 74631375                       | H: high; M: medium; L: low; N: neutral. H/M means functional and L/N means non-functional                       |
| FATHMM (fathmm)            | 70274896                       | D: Deleterious; T: Tolerated  |
| MetaSVM (metasvm)          | 82098217                       | D: Deleterious; T: Tolerated  |
| MetaLR (metalr)            | 82098217                       | D: Deleterious; T: Tolerated  |
| GERP++ (gerp++)            | 89076718                       | higher scores are more deleterious  |
| PhyloP (phylop)            | 89553090                       | higher scores are more deleterious  |
| SiPhy (siphy)              | 88269630                       | higher scores are more deleterious  |

\*Check here for a more detailed information of each score:  
<https://annovar.openbioinformatics.org/en/latest/user-guide/filter/>

# Conclusions

---

1. We learnt the different steps of a somatic variant calling workflow:
  - Quality control of raw reads
  - Alignment
  - How to prepare the BAM files for the variant calling
  - Somatic variant calling
  - Variant annotation
2. We explained the different source of artefacts in the somatic variant calling workflow
3. We learnt the challenges of SNV detection and how to solve them (sequencing/mapping errors, VAF limitations, coverage...)

# What is next?

---

## Day one – Monday 10

July 2023

09:00 – 09:30

Arrival and registration

09:30 – 09:45

Welcome

Ajay Mishra

09:45 – 10:45

Course introduction

Isidro Cortes-  
Ciriano

10:45 – 11:00

Break

11:00 – 12:30

Introduction to cancer genomics and its caveats

Tobias Rausch

12:30 – 13:30

Lunch

13:30 – 14:30

High-throughput sequencing and single-nucleotide  
variant analysis in cancer

Francesc Muyas  
Remolar

14:30 – 15:00

Break

15:00 – 17:30

Practical: Alignment and SNV analysis

Francesc Muyas  
Remolar

17:30 – 18:30

Accommodation check-in and free time

18:30

Dinner at Hinxton Hall Conference Centre

# Acknowledgements

---

**Cortes-Ciriano group (EMBL-EBI)**

**Course organisers**

**All trainers**

Isidro Cortes Ciriano, EMBL-EBI

Ajay Mishra, EMBL-EBI

Tobias Rausch, EMBL

Sophie Spencer, EMBL-EBI

