Sequence analysis practical

UZH-BIO392 day 08

Max Verbiest 2023-09-28

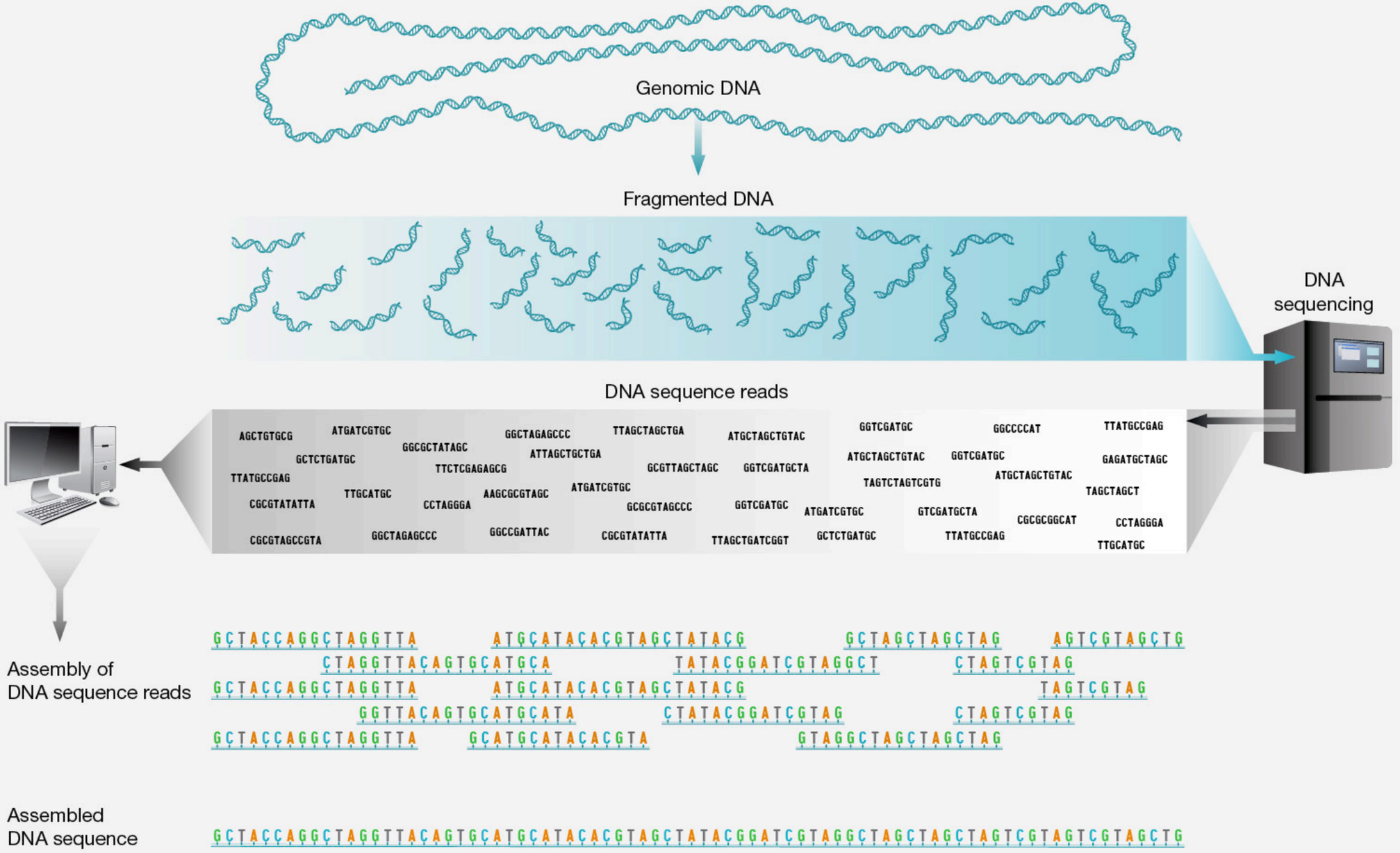


Image courtesy of National Human Genome Research Institute (https://www.genome.gov/genetics-glossary/Shotgun-Sequencing)

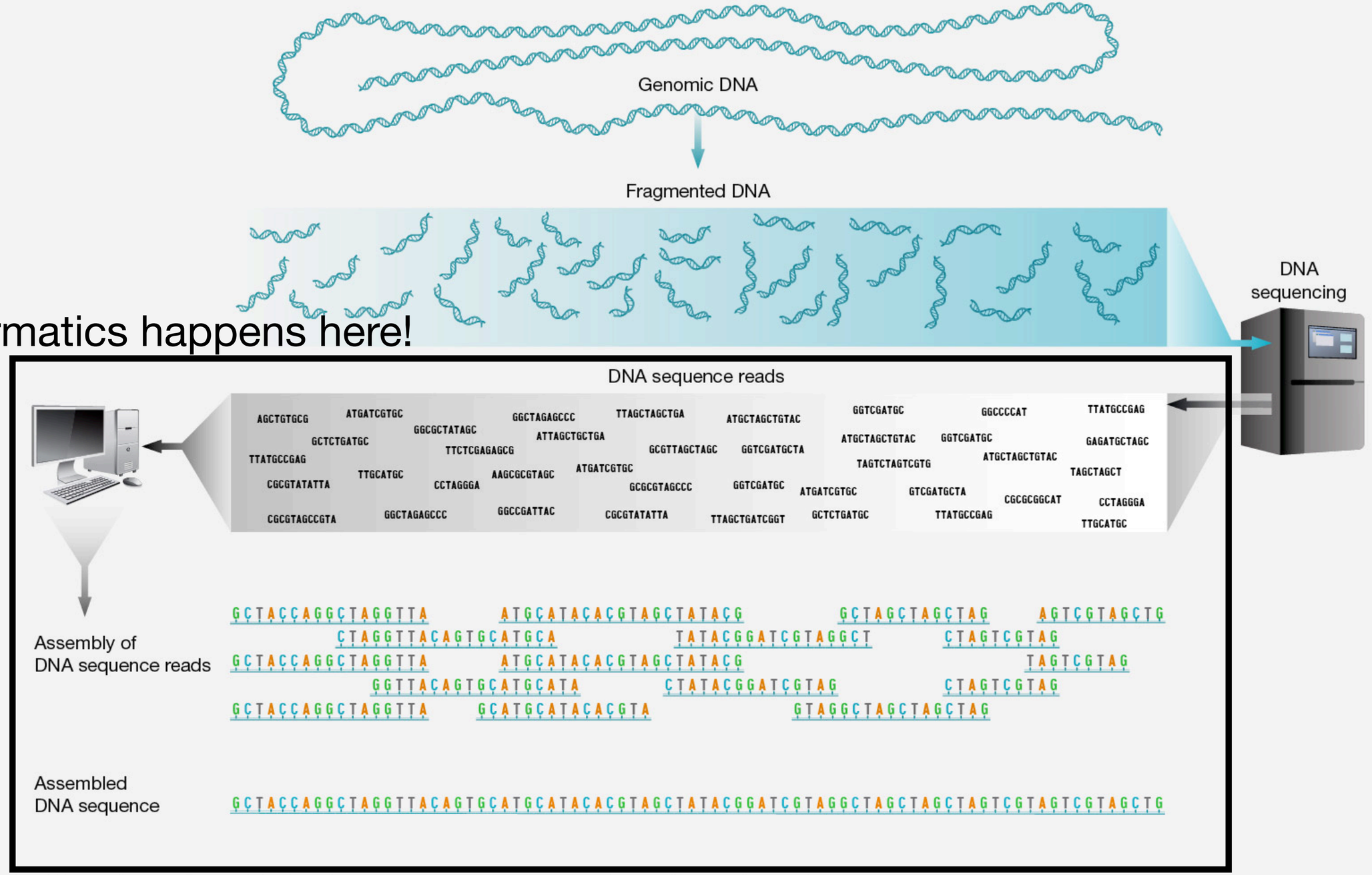
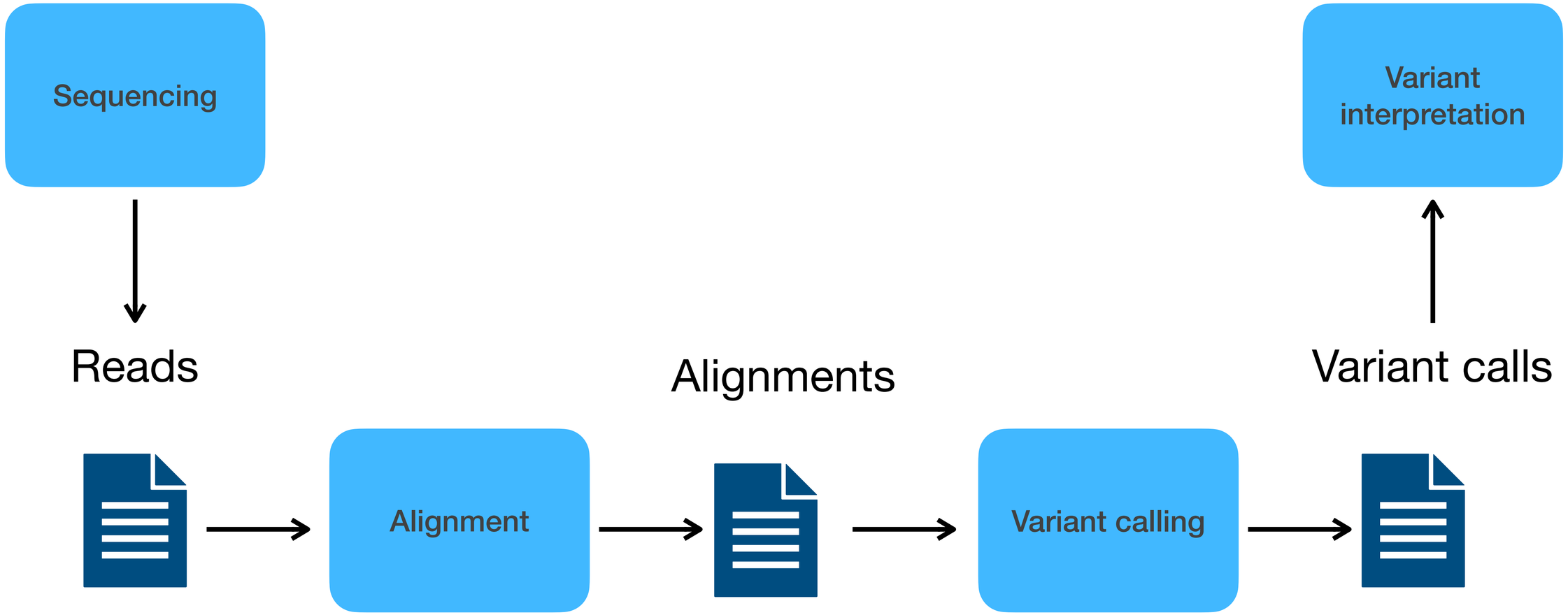
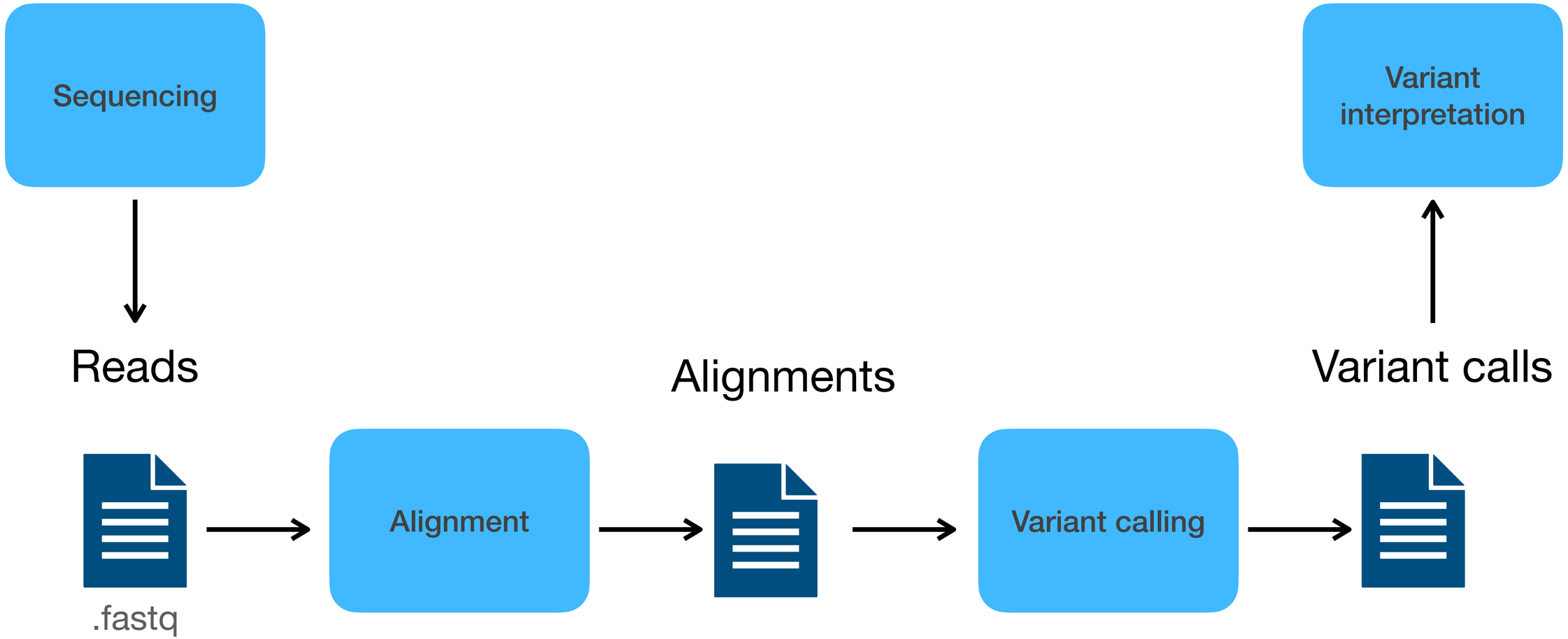


Image courtesy of National Human Genome Research Institute (https://www.genome.gov/genetics-glossary/Shotgun-Sequencing)

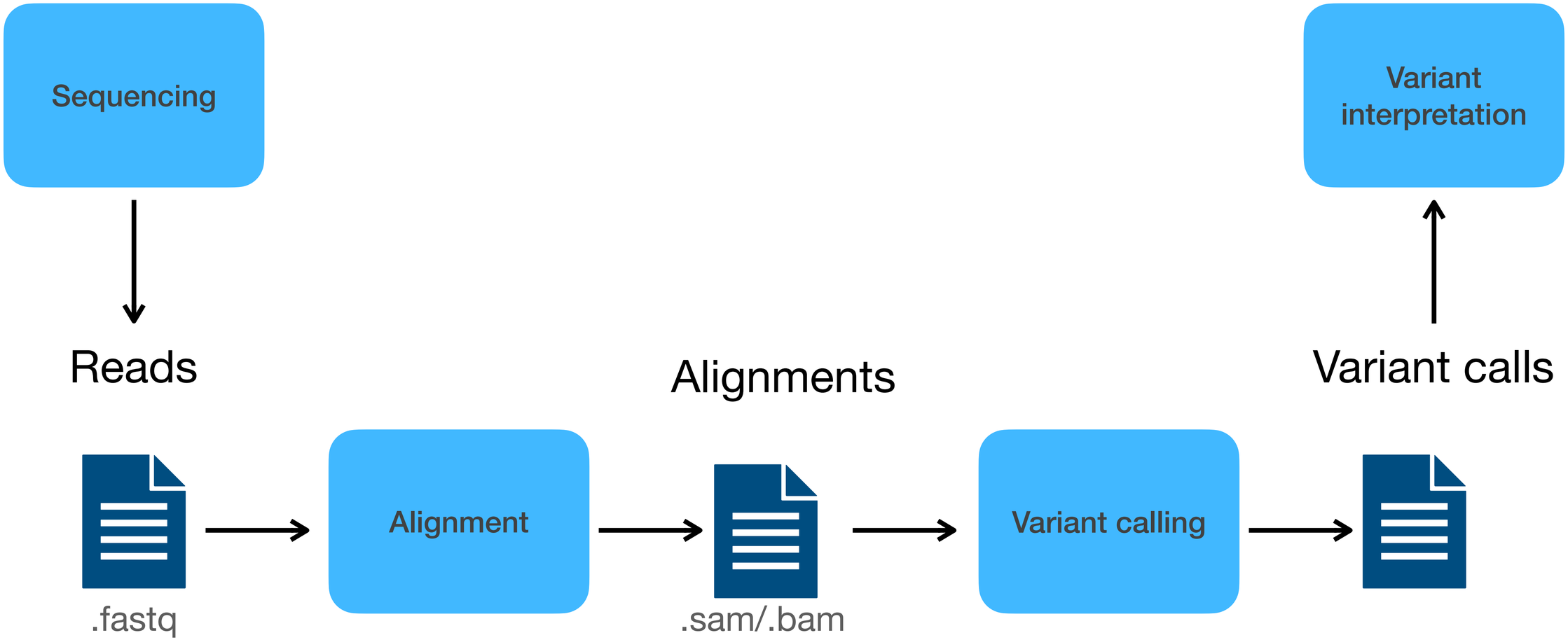
Generic bioinformatics pipeline



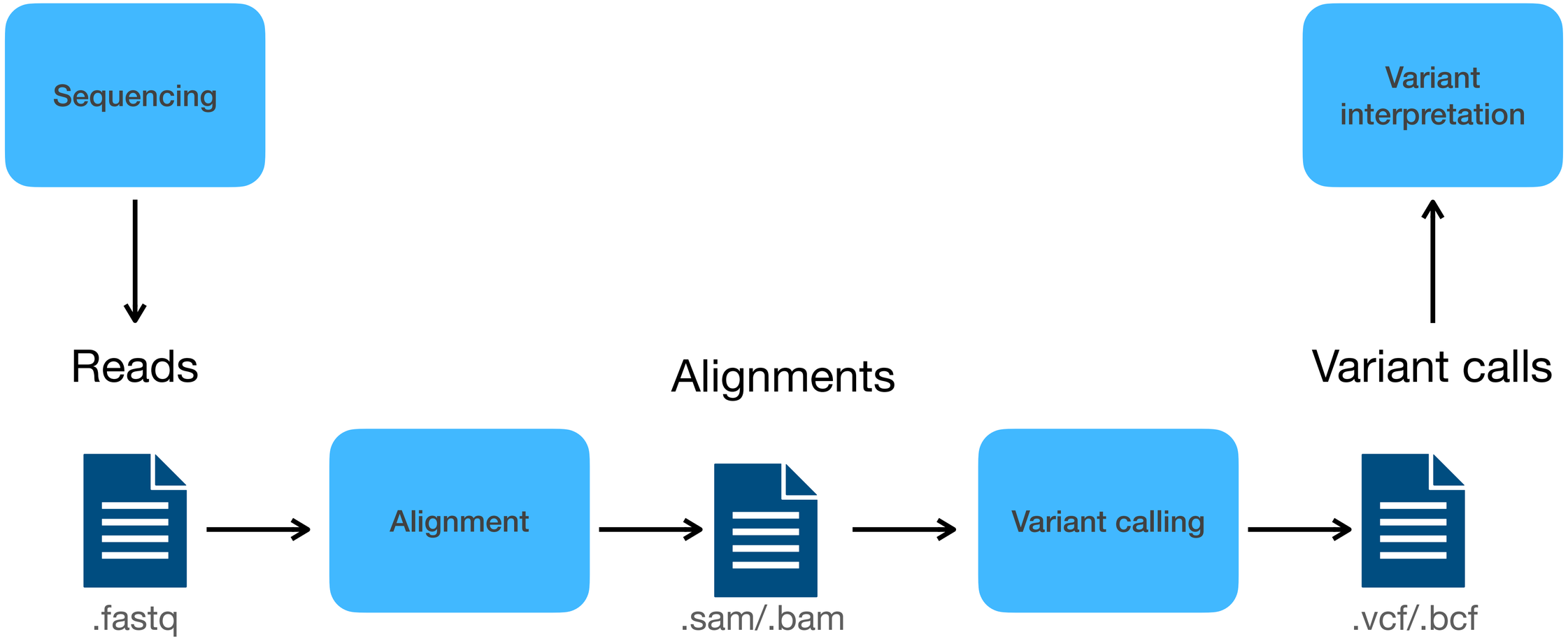
Generic bioinformatics pipeline



Generic bioinformatics pipeline



Generic bioinformatics pipeline



Program

Program

- Background: short tandem repeats

Program

- Background: short tandem repeats
- Shotgun sequencing

Program

- Background: short tandem repeats
- Shotgun sequencing
- Bioinformatics pipeline
  - Processing reads
  - Alignment
  - Variant calling
  - Variant interpretation

Short tandem repeats

Repetitions of 1-6 bp DNA units

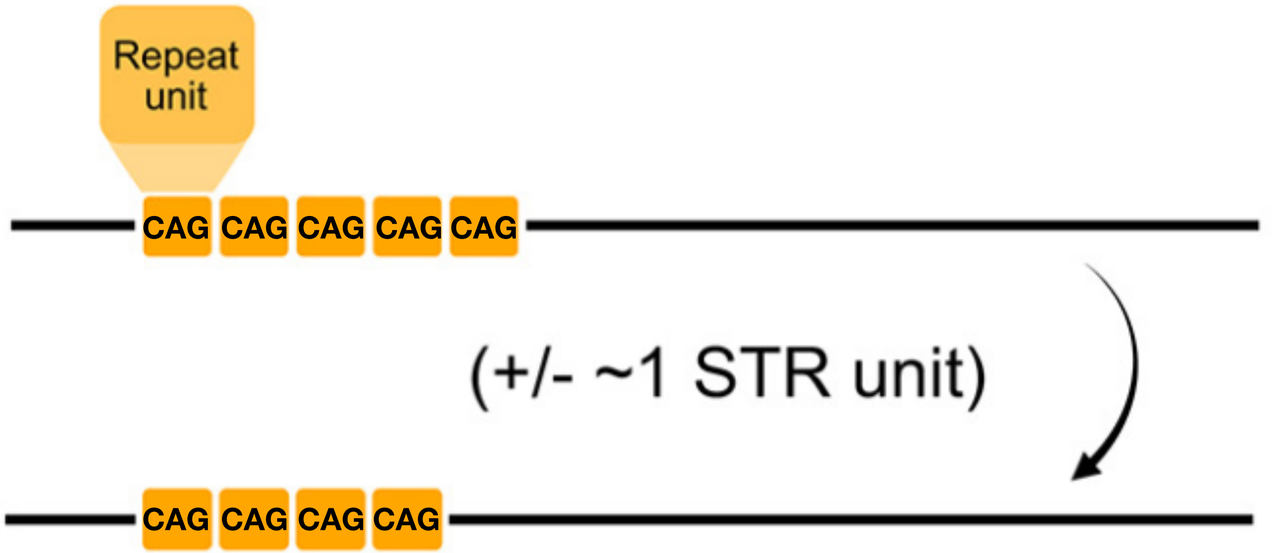
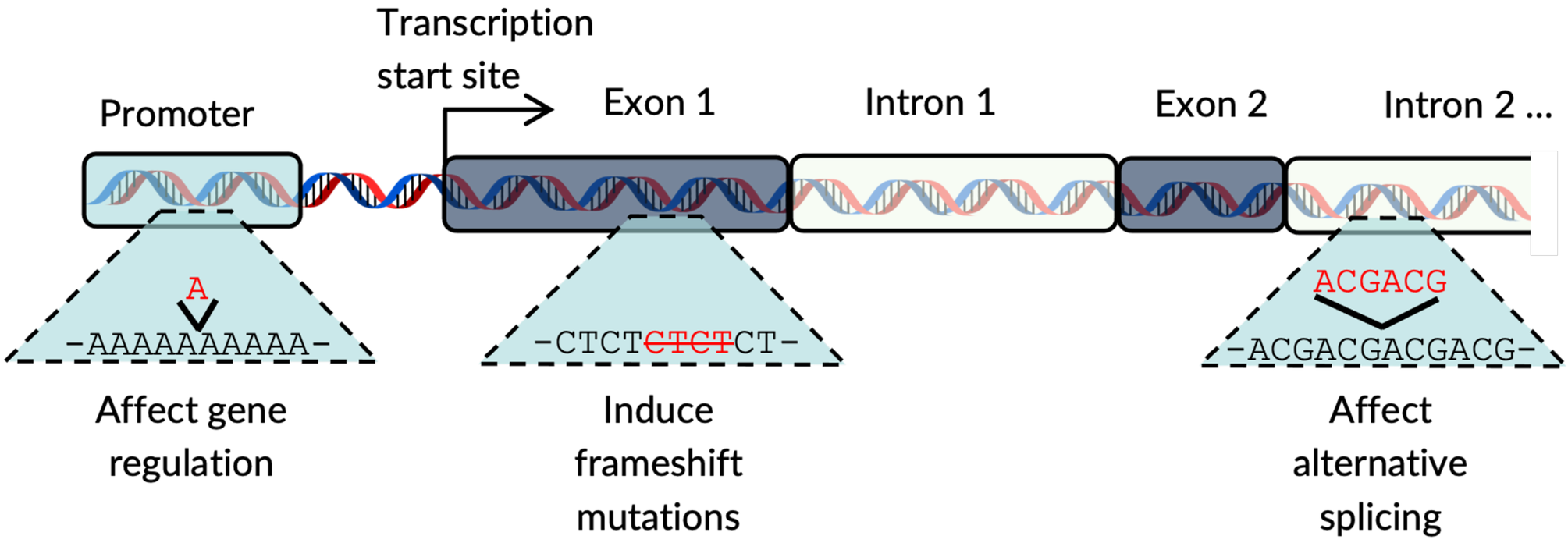


Image adapted from: Verbiest *et al*., 2023


Transcription start site
Promoter
Exon 1
Intron 1
Exon 2
Intron 2 ...
A
-AAAAAAAAAA-
-CTCTCTCTCT-
ACGACG
-ACGACGACGACG-
Affect gene regulation
Induce frameshift mutations
Affect alternative splicing

Effects of STR variation still poorly understood

Image courtesy of National Human Genome Research Institute (https://www.genome.gov/genetics-glossary/Cell)

Shotgun sequencing

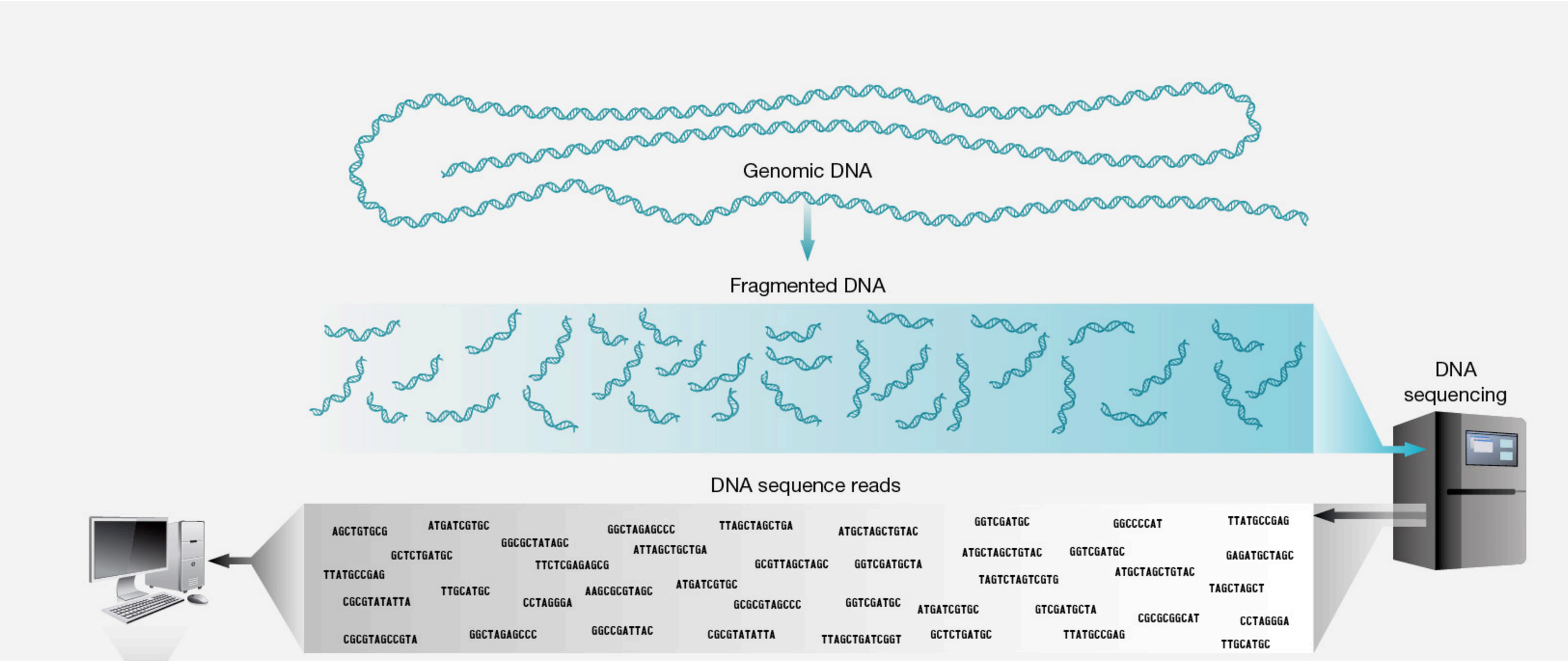
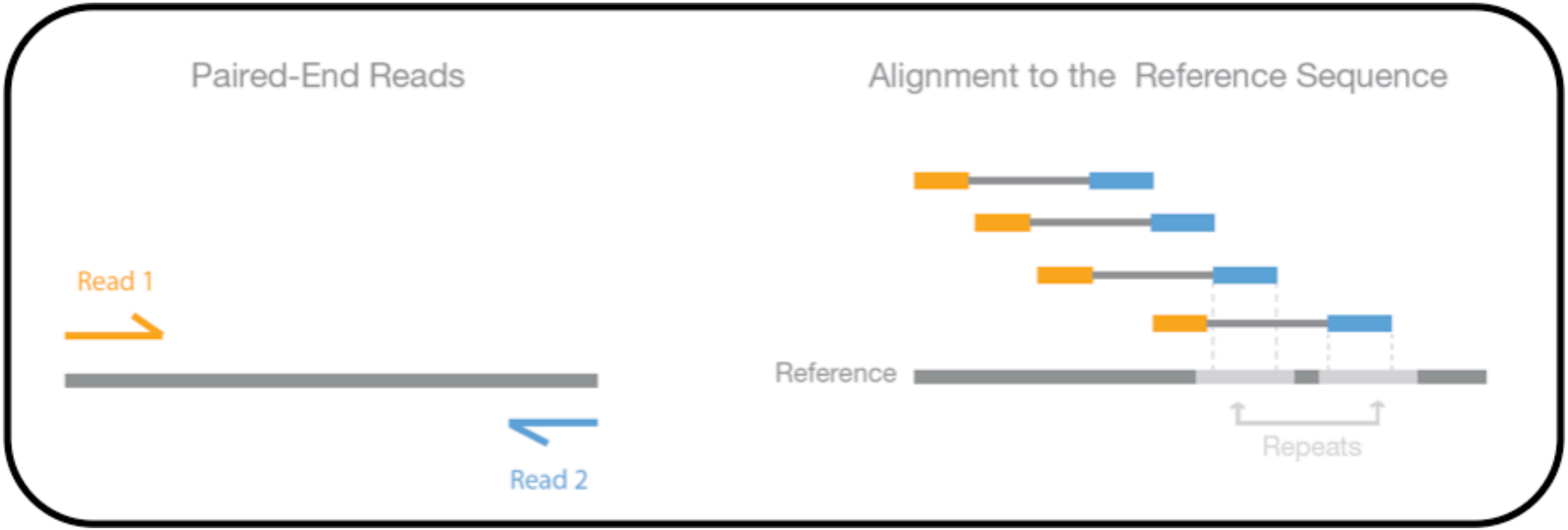


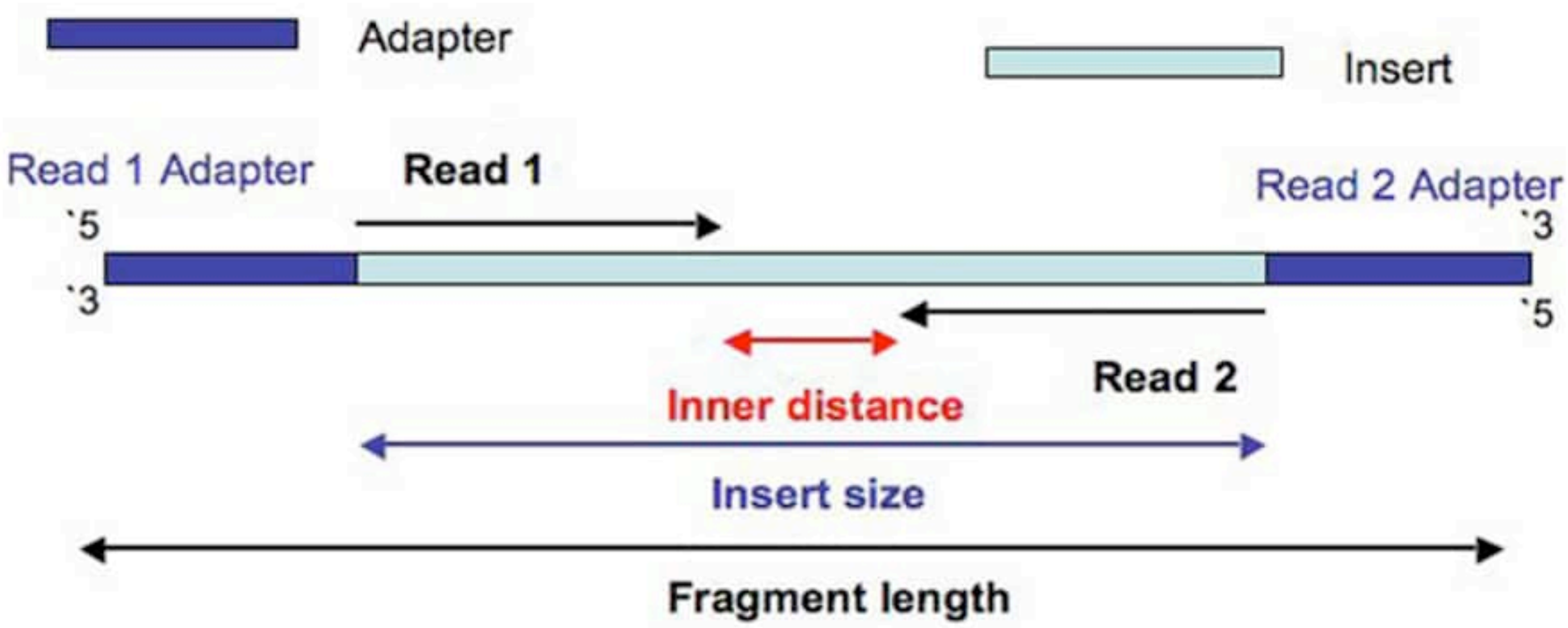
Image courtesy of National Human Genome Research Institute (https://www.genome.gov/genetics-glossary/Shotgun-Sequencing)

Paired-end reads



https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/rna-sequencing/performing-a-rna-seq-experiment/design-considerations/

Paired-end reads: more detail

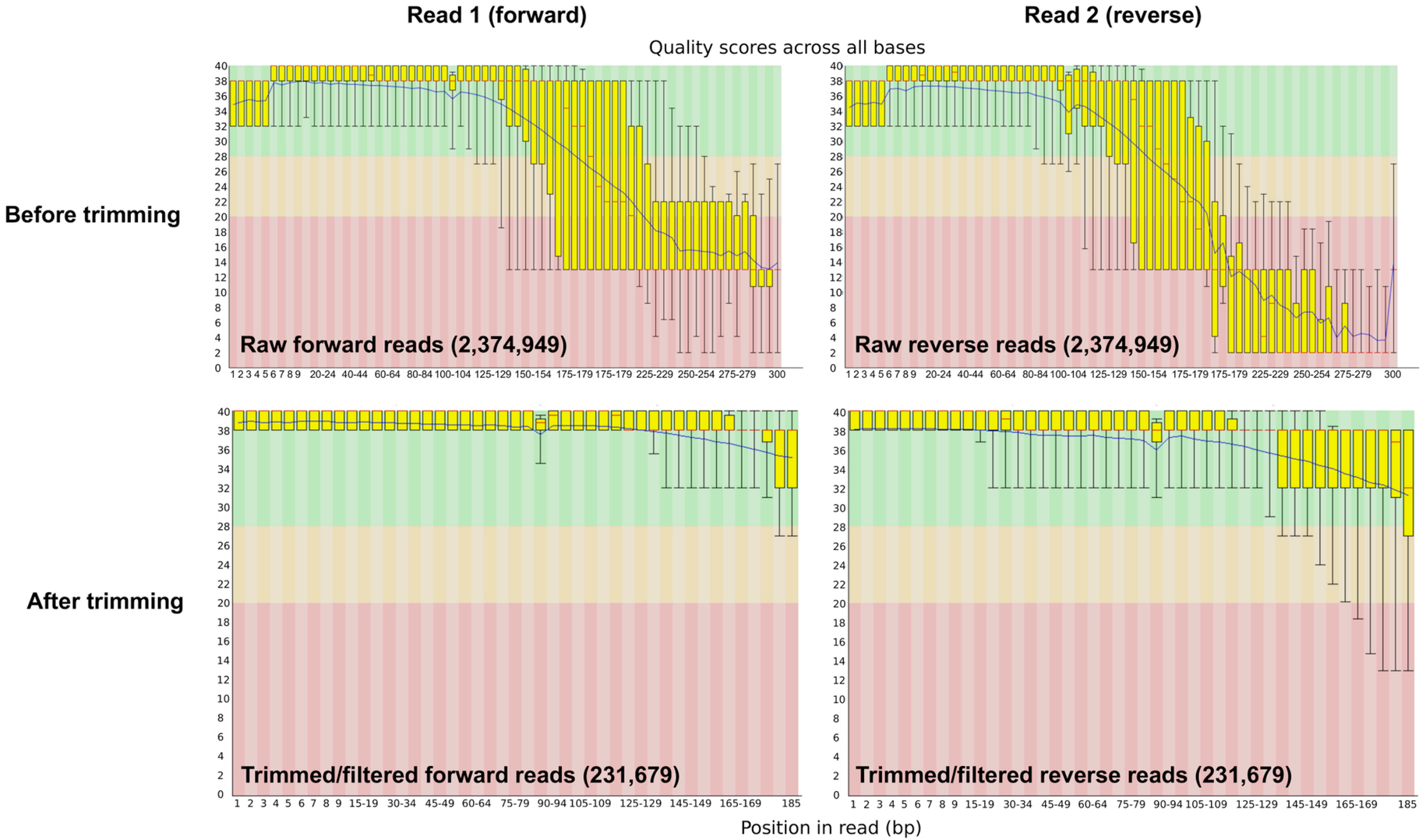


https://thesequencingcenter.com/knowledge-base/what-are-paired-end-reads/

Reads are processed before analysis

- Typical read processing steps:
    - Adapter trimming (remove leftovers from sequencing process)
    - Quality trimming (remove low Phred-score bases)
    - Length filter (remove too-short reads)

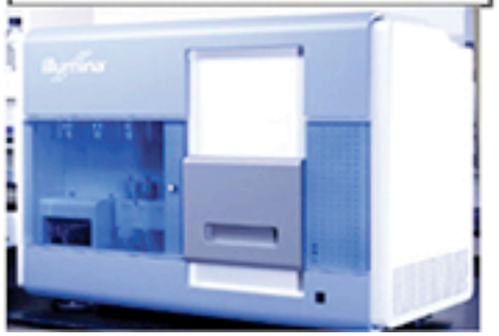
Read processing before/after



https://www.jove.com/t/54936/metagenomic-analysis-of-silage

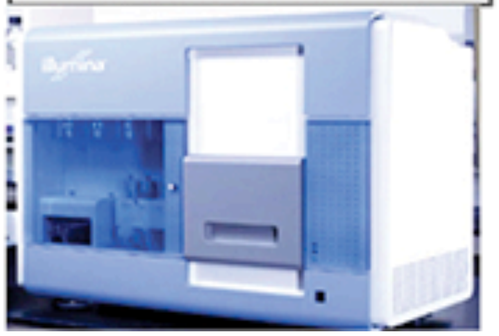
Short reads

.....
GGTCTGGATGC
CGGTCTGGATGC
GCGGTCTGGATG
GCGGTCTGGAT
GGCGGTCTGGAT
GGCGGTCTGGA
TCTATGCGGGCCCCT
TCTATGCGGGCCCC
ATCTATGCGGGCC
TATCTATGCGGGC
TTATCTATGCGGG
CTTATCTATGCGGG



Short reads

.....
GGTCTGGATGC
CGGTCTGGATGC
GCGGTCTGGATG
GCGGTCTGGAT
GGCGGTCTGGAT
GGCGGTCTGGA
TCTATGCGGGCCCCT
TCTATGCGGGCCCC
ATCTATGCGGGCC
TATCTATGCGGGC
TTATCTATGCGGG
CTTATCTATGCGGG

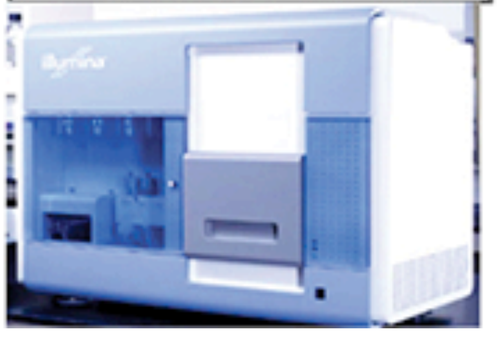


GGCGGTCTAGATGCTTATCTATGCGGGCCCCT

Reference genome sequence

Short reads

.....
GGTCTGGATGC
CGGTCTGGATGC
GCGGTCTGGATG
GCGGTCTGGAT
GGCGGTCTGGAT
GGCGGTCTGGA
TCTATGCGGGCCCCT
TCTATGCGGGCCCC
ATCTATGCGGGCC
TATCTATGCGGGC
TTATCTATGCGGG
CTTATCTATGCGGG



Alignment of reads to the reference genome and SNP calling

SNP:A->G

GTCTGGATGCT TCTATGCGGGCCCCT
GGTCTGGATGC TCTATGCGGGCCCC
CGGTCTGGATGC ATCTATGCGGGCC
GCGGTCTGGATG TATCTATGCGGGC
GCGGTCTGGAT TTATCTATGCGGG
GCGGTCTGGAT CTTATCTATGCGGG
GGCGGTCTGGAT CTTATCTATGCGG
GGCGGTCTGGA CTTATCTATGCGG

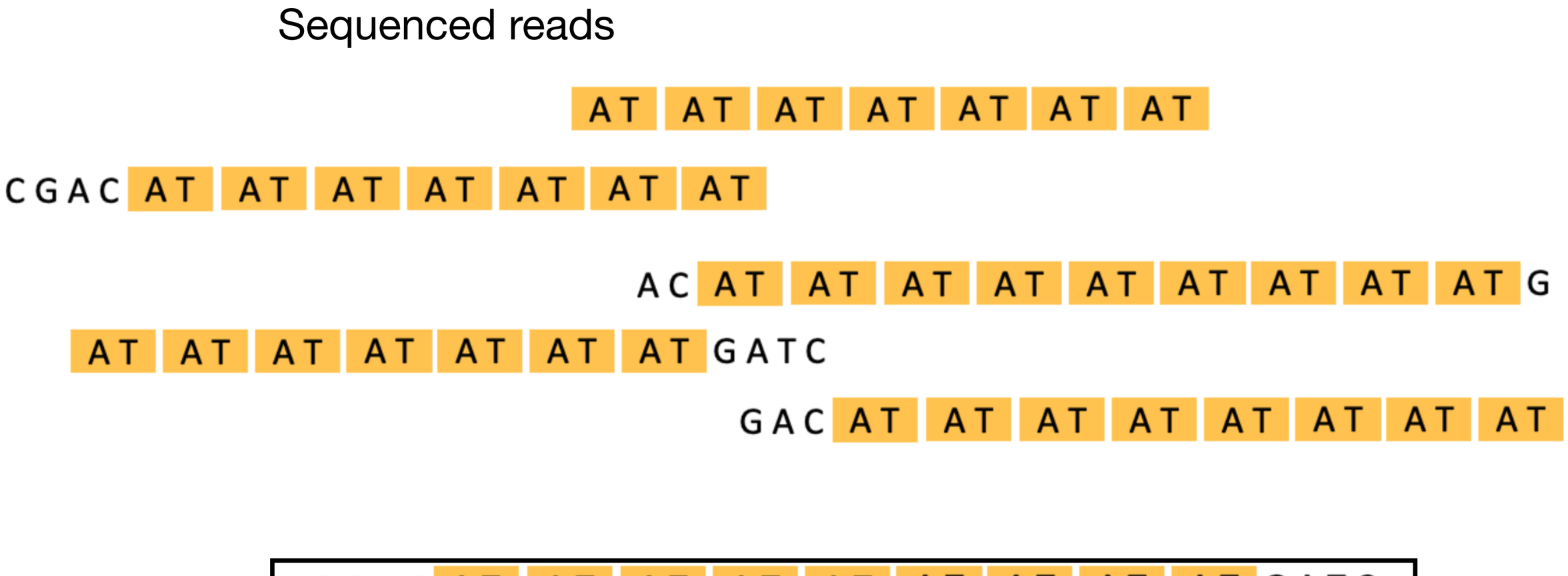
GGCGGTCTAGATGCTTATCTATGCGGGCCCCT

Reference genome sequence

Aligning reads to STRs

Reference genome

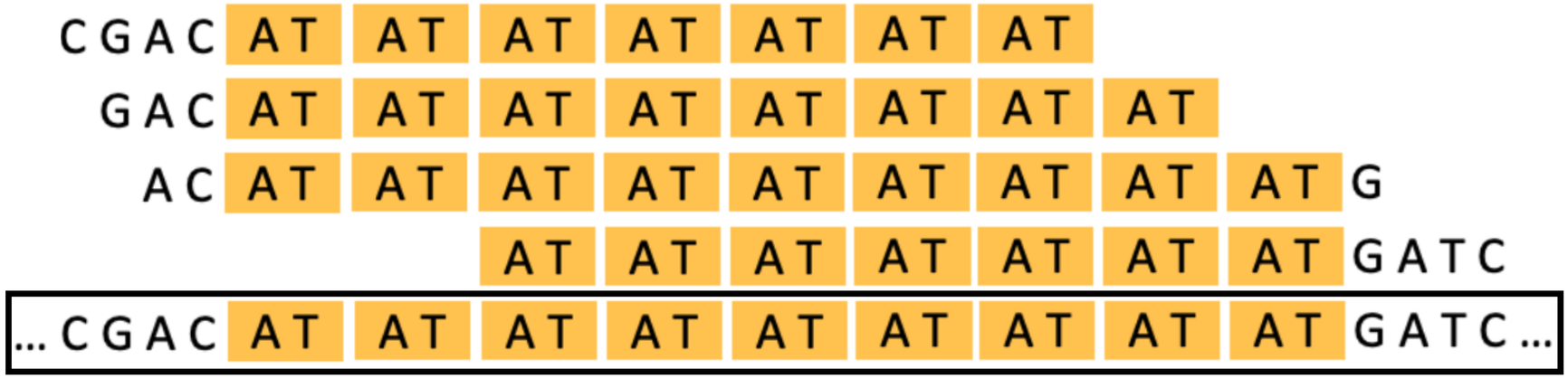
Aligning reads to STRs



Aligning short reads to STRs

Sequenced reads

Aligned reads

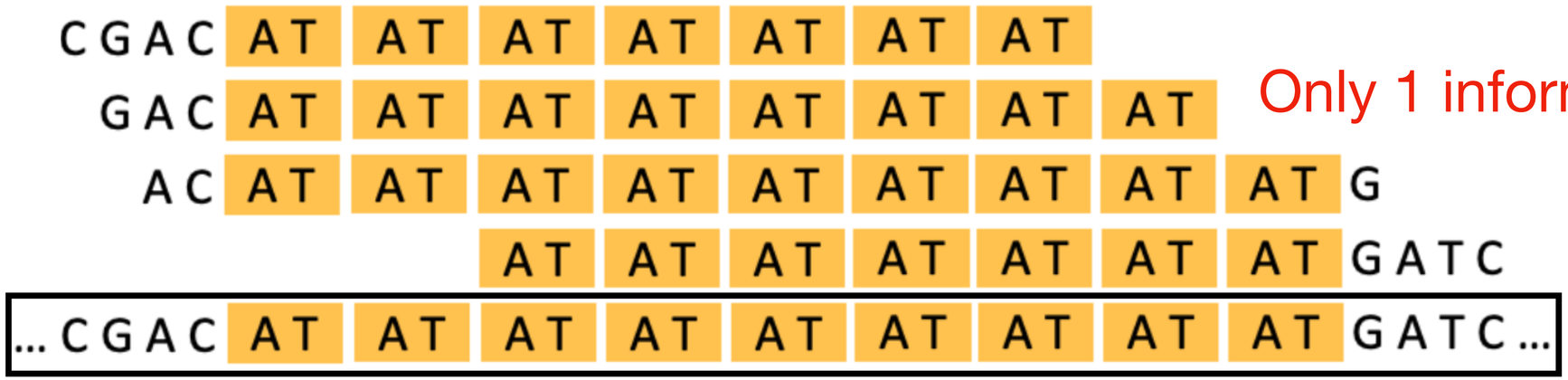


Reference genome

Aligning short reads to STRs

Sequenced reads

Aligned reads



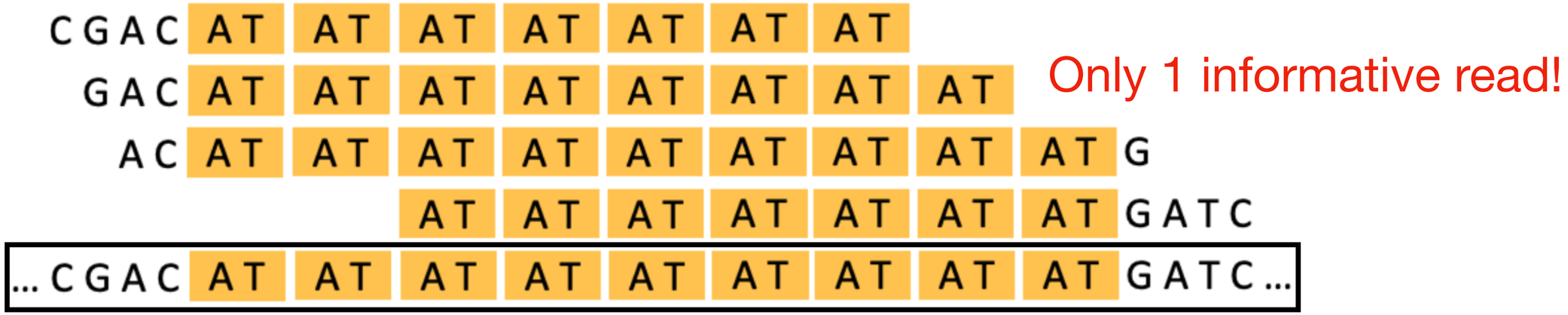
Only 1 informative read!

Reference genome

Aligning short reads to STRs

Sequenced reads

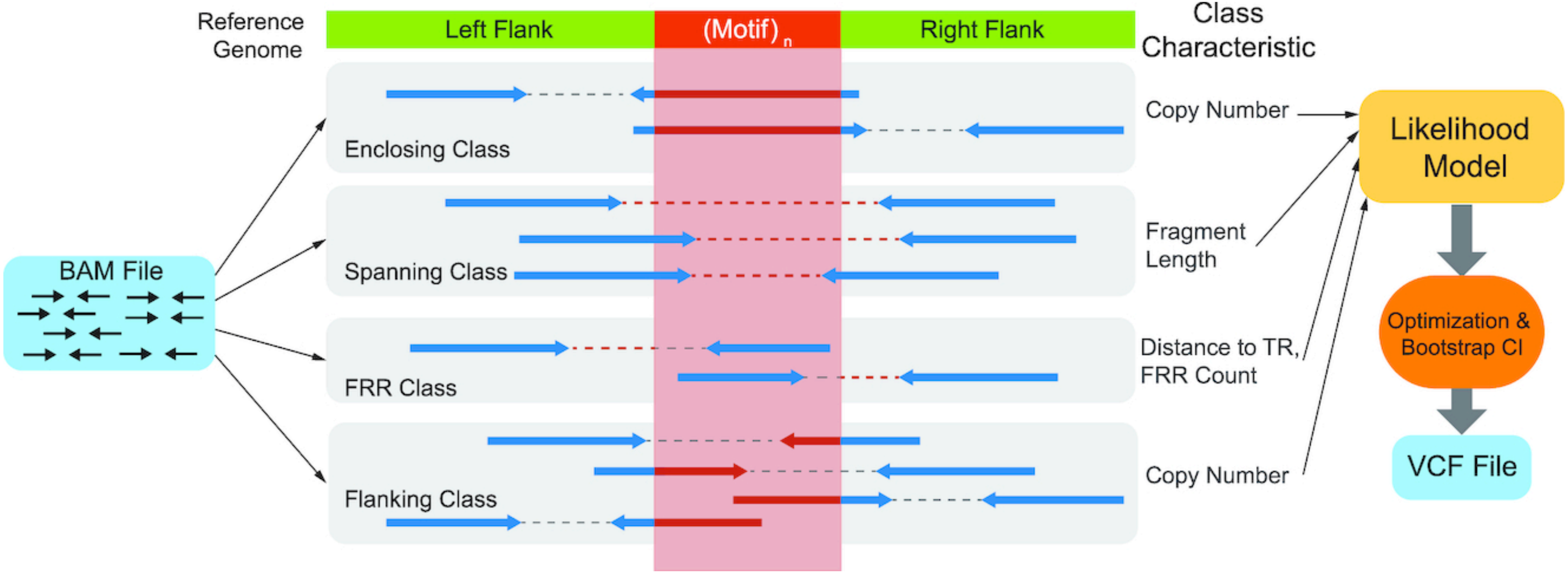
Aligned reads



Reference genome

Specialised tools are needed!

E.g., GangSTR



https://academic.oup.com/nar/article/47/15/e90/5518310#140580783

End product: VCF

```
#CHROM  POS     ID      REF     ALT     QUAL    FILTER  INFO    FORMAT  mut
chr5    298     .       aaaaaaaaaaaa    .       .       .       END=309;RU=a;PERIOD=1;REF=12;GRID=9,15;STUTTERUP=0.05;STUTTERDOWN=0.05;STUT
chr5    7241    .       aaaaaaaaa       .       .       .       END=7249;RU=a;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STUT
chr5    9390    .       aaaaaaaaaa      .       .       .       END=9399;RU=a;PERIOD=1;REF=10;GRID=7,13;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    10062   .       ttttttttttttttt .       .       .       END=10077;RU=t;PERIOD=1;REF=16;GRID=13,19;STUTTERUP=0.05;STUTTERDOW
chr5    10673   .       aaaaaaaaaaaaaaaa        .       .       .       END=10688;RU=a;PERIOD=1;REF=16;GRID=13,19;STUTTERUP=0.05;STUTTERDOW
chr5    15411   .       ttttttttttttttttttttttttttttt   .       .       .       END=15439;RU=t;PERIOD=1;REF=29;GRID=26,32;STUTTERUP=0.05;ST
chr5    15503   .       ttttttttt       .       .       .       END=15512;RU=t;PERIOD=1;REF=10;GRID=7,13;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    16887   .       tttttttttt      .       .       .       END=16897;RU=t;PERIOD=1;REF=11;GRID=8,14;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    17044   .       ttttttttttttttt .       .       .       END=17058;RU=t;PERIOD=1;REF=15;GRID=12,18;STUTTERUP=0.05;STUTTERDOWN=0.05;S
chr5    19394   .       aaaaaaaaa       .       .       .       END=19402;RU=a;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    20647   .       aaaaaaaaa       .       .       .       END=20655;RU=a;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    20965   .       aaaaaaaaaa      .       .       .       END=20974;RU=a;PERIOD=1;REF=10;GRID=7,13;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    21005   .       atatatatat      .       .       .       END=21014;RU=at;PERIOD=2;REF=5;GRID=2,8;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    22490   .       tttttttt        .       .       .       END=22498;RU=t;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    24686   .       ttttttttttt     .       .       .       END=24697;RU=t;PERIOD=1;REF=12;GRID=9,15;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    25452   .       aaaaaaaaaaa     .       .       .       END=25462;RU=a;PERIOD=1;REF=11;GRID=8,14;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    32035   .       aaaaaaaaa       .       .       .       END=32043;RU=a;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    32296   .       ttttttttttttttt .       .       .       END=32310;RU=t;PERIOD=1;REF=15;GRID=12,18;STUTTERUP=0.05;STUTTERDOWN=0.05;S
chr5    32477   .       tttttttt        .       .       .       END=32485;RU=t;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    32685   .       atatatatat      .       .       .       END=32694;RU=at;PERIOD=2;REF=5;GRID=2,8;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    37544   .       tacatacataca    .       .       .       END=37555;RU=taca;PERIOD=4;REF=3;GRID=1,6;STUTTERUP=0.05;STUTTERDOWN=0.05;S
chr5    41515   .       tttttttt        .       .       .       END=41523;RU=t;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    46982   .       ttttttttttttttttttttttt .       .       .       END=47004;RU=t;PERIOD=1;REF=23;GRID=20,26;STUTTERUP=0.05;STUTTERDOW
chr5    47414   .       tttttttttttt    .       .       .       END=47426;RU=t;PERIOD=1;REF=13;GRID=10,16;STUTTERUP=0.05;STUTTERDOWN=0.05;S
chr5    47459   .       tttttttttt      .       .       .       END=47469;RU=t;PERIOD=1;REF=11;GRID=8,14;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    47471   .       ttttttttttt     .       .       .       END=47482;RU=t;PERIOD=1;REF=12;GRID=9,15;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    47958   .       tttttttttt      .       .       .       END=47968;RU=t;PERIOD=1;REF=11;GRID=8,14;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    53447   .       aaaaaaaaaaaaaaa .       .       .       END=53461;RU=a;PERIOD=1;REF=15;GRID=12,18;STUTTERUP=0.05;STUTTERDOWN=0.05;S
chr5    53475   .       aaaaaaaaaaa     .       .       .       END=53485;RU=a;PERIOD=1;REF=11;GRID=8,14;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    57340   .       tttttttt        .       .       .       END=57348;RU=t;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    58315   .       ttttttttttttt   .       .       .       END=58328;RU=t;PERIOD=1;REF=14;GRID=11,17;STUTTERUP=0.05;STUTTERDOWN=0.05;S
chr5    60745   .       tttttttt        .       .       .       END=60753;RU=t;PERIOD=1;REF=9;GRID=6,12;STUTTERUP=0.05;STUTTERDOWN=0.05;STU
chr5    60865   .       ttttttttt       .       .       .       END=60874;RU=t;PERIOD=1;REF=10;GRID=7,13;STUTTERUP=0.05;STUTTERDOWN=0.05;ST
chr5    61581   .       aaaaaaaaaaaaaaaaa       .       .       .       END=61598;RU=a;PERIOD=1;REF=18;GRID=15,21;STUTTERUP=0.05;STUTTERDOW
chr5    61748   .       aaaaaaaaaaaaaaaaa       .       .       .       END=61765;RU=a;PERIOD=1;REF=18;GRID=15,21;STUTTERUP=0.05;STUTTERDOW
chr5    65546   .       ttttttttttttttt .       .       .       END=65562;RU=t;PERIOD=1;REF=17;GRID=14,20;STUTTERUP=0.05;STUTTERDOW
chr5    65866   .       aaaaaaaaaaaa    .       .       .       END=65878;RU=a;PERIOD=1;REF=13;GRID=10,16;STUTTERUP=0.05;STUTTERDOWN=0.05;S
chr5    66016   .       ttttttttttttttttt       .       .       .       END=66032;RU=t;PERIOD=1;REF=17;GRID=14,20;STUTTERUP=0.05;STUTTERDOW
chr5    66553   .       tttttttttttttttttttt    .       .       .       END=66572;RU=t;PERIOD=1;REF=20;GRID=17,23;STUTTERUP=0.05;STUTTERDOW
```

For all things genome: Ensembl

For all things genome: Ensembl

Variant Effect Predictor (VEP)

Web Tools
- Web Tools
  - BLAST/BLAT
  - Variant Effect Predictor
  - Linkage Disequilibrium Calculato
  - Variant Recoder
  - File Chameleon
  - Assembly Converter
  - ID History Converter
  - VCF to PED Converter
  - Data Slicer

Configure this page

Custom tracks

Export data

Share this page

Bookmark this page

Variant Effect Predictor

New job

Species:

Homo_sapiens X

Assembly: GRCh38.p14

Change species

If you are looking for VEP for Human GRCh37, please go to GRCh37 website.

Name for this job (optional):

Input data:

Either paste data:

```
1 65568 . A C . . .
2 265023 . C T . . .
3 319780 . GA G . . .
```

Run instant VEP for current line ›

Examples: Ensembl default, VCF, Variant identifiers, HGVS notations, SPDI

Or upload file: Browse... No file selected.

Or provide file URL:

Transcript database to use:

- (•) Ensembl/GENCODE transcripts
- ( ) Ensembl/GENCODE basic transcripts
- ( ) RefSeq transcripts