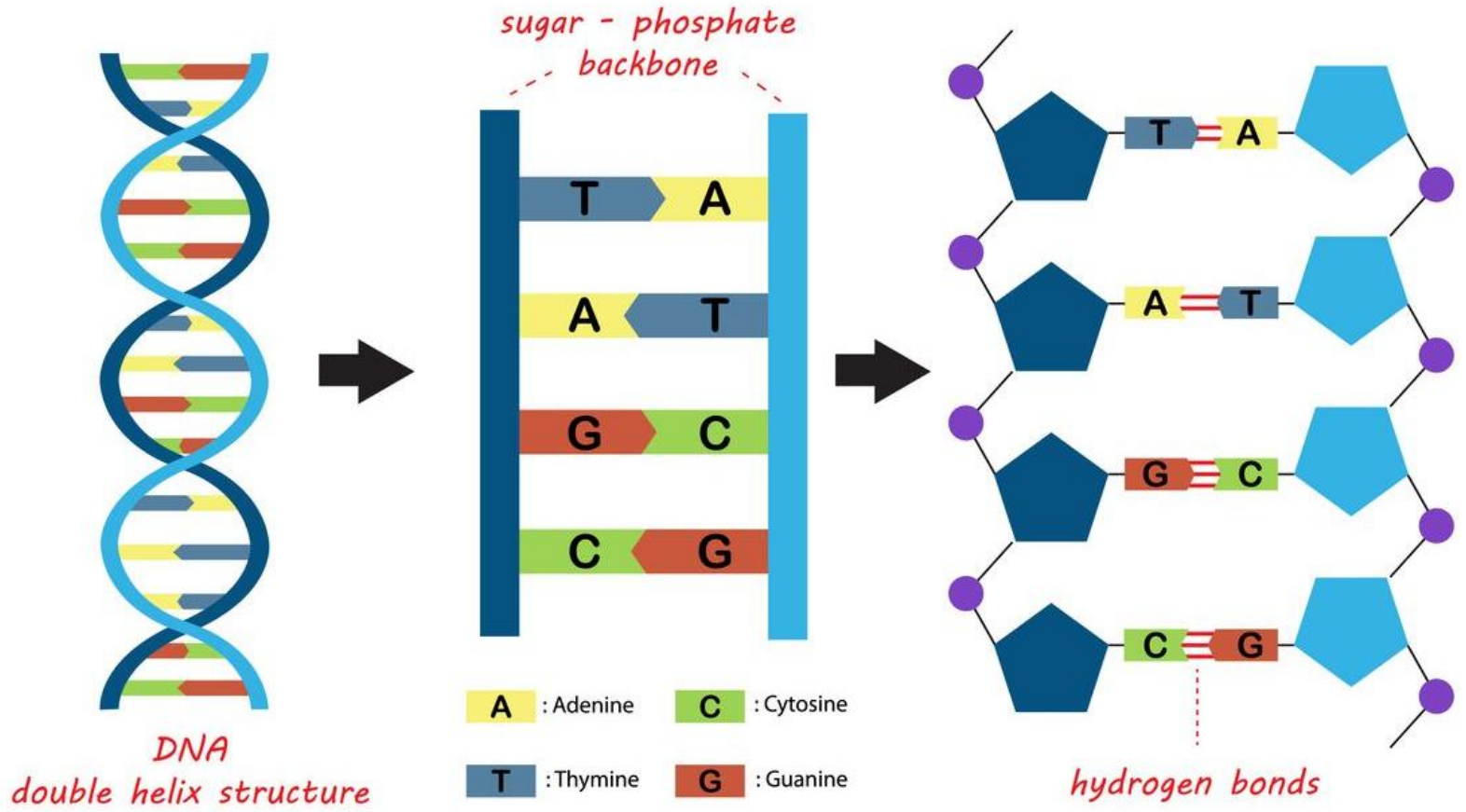

DTU

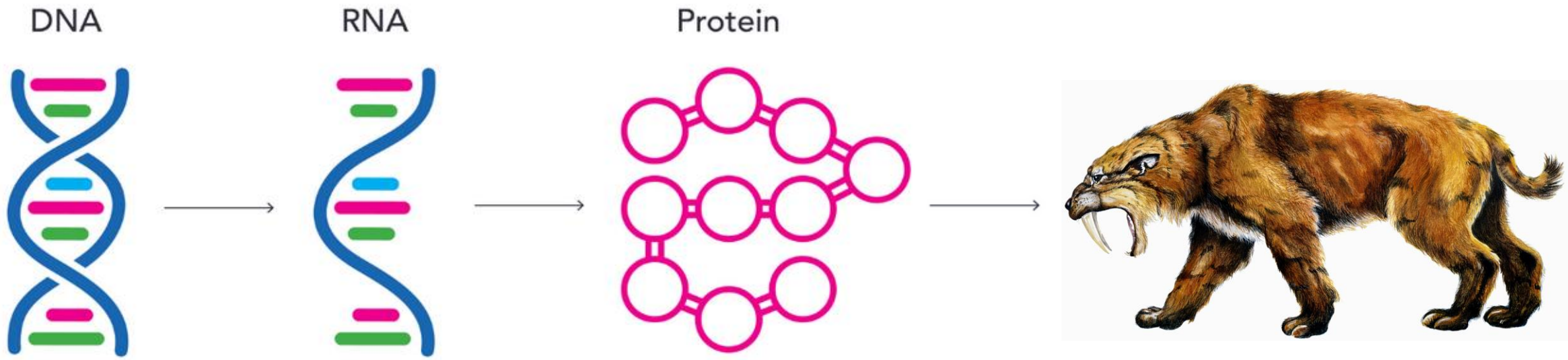
22126: Next Generation Sequencing Analysis
DTU - January 2026
Mick Westbury

Mick Westbury
Associate Professor
Section of Bioinformatics
Technical University of Denmark
micwe@dtu.dk

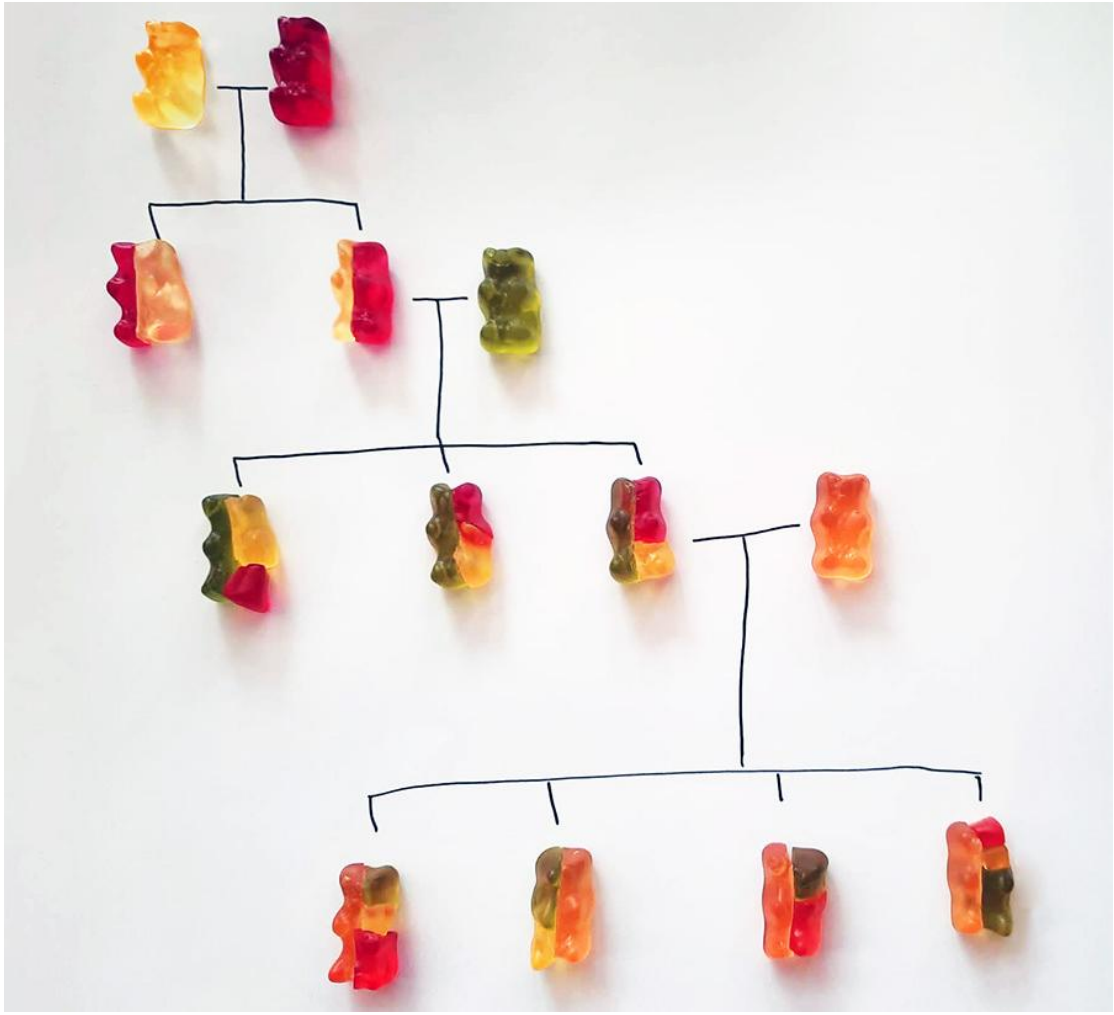
Basic genetics – composition



Basic genetics – function



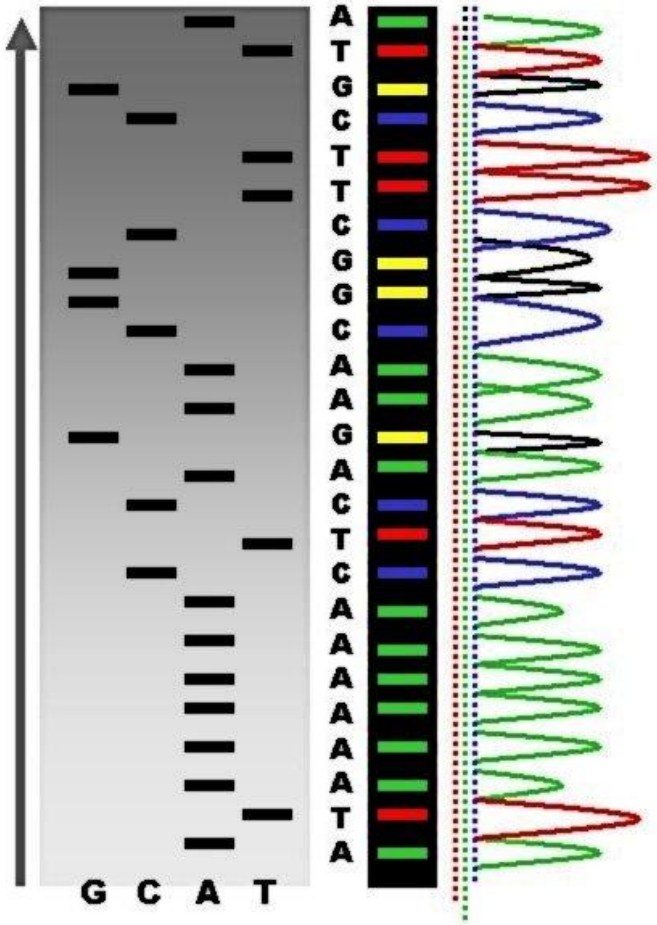
Basic genetics – inheritance




IN THE BEGINNING...
STEP ONE...
LET THERE BE LIGHT?
ACCEPT
DECLINE
deAn.

First-generation sequencing

- Sanger sequencing
- Very accurate, long reads (~800–1000 bp).
- Foundation for sequencing logic
- Relatively low throughput (96 reads/run)

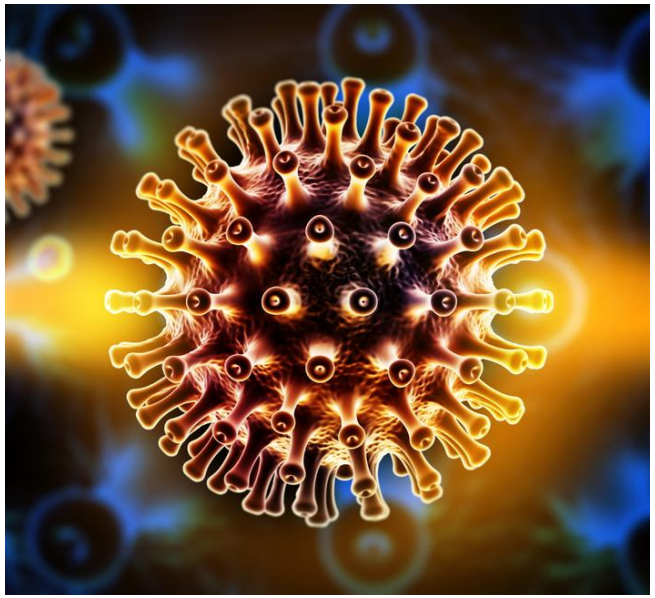


First-generation sequencing

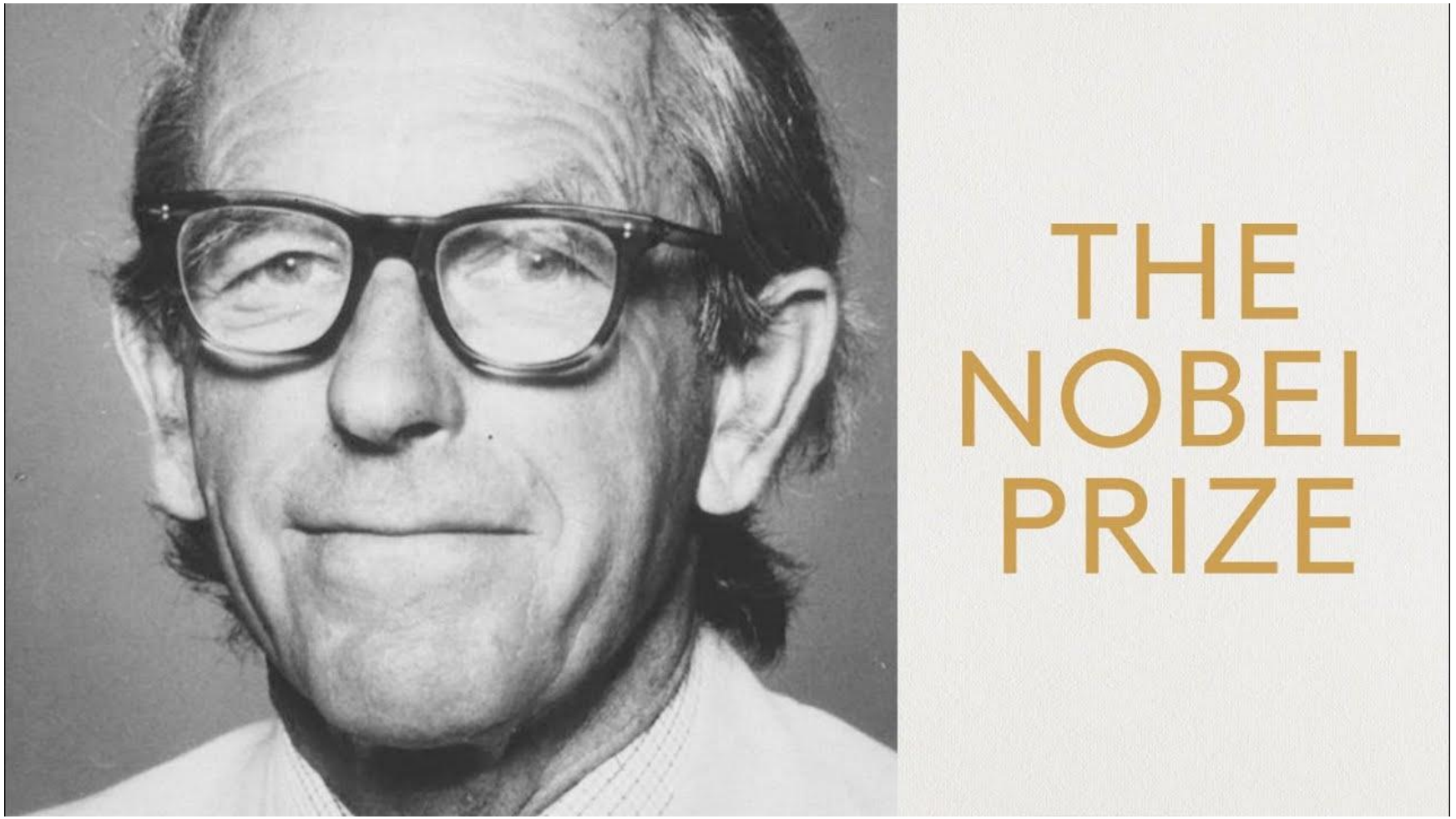
- Medical & Human Genetics
  - Diagnose disease-causing mutations
  - Targeted drug development
- Ancient DNA
- Developmental & Molecular Genetics
  - Sequencing of model organism genes
- Population & Evolutionary Genetics
  - Human migration studies
- Agriculture & Targeted Breeding
  - Sequencing of key genes controlling traits

Sanger + HIV biology

- 1985–1987: **First full HIV genomes (~9.7kb) were sequenced**
  - Reveal mutational hotspots and drug targets
  - Allowed the first antiretroviral drugs
- By the mid-1990s, genotypic drug-resistance test became part of HIV care.
  - Detect known mutations linked to treatment evasion
  - Sequencing = find right drug combination.



First-generation sequencing

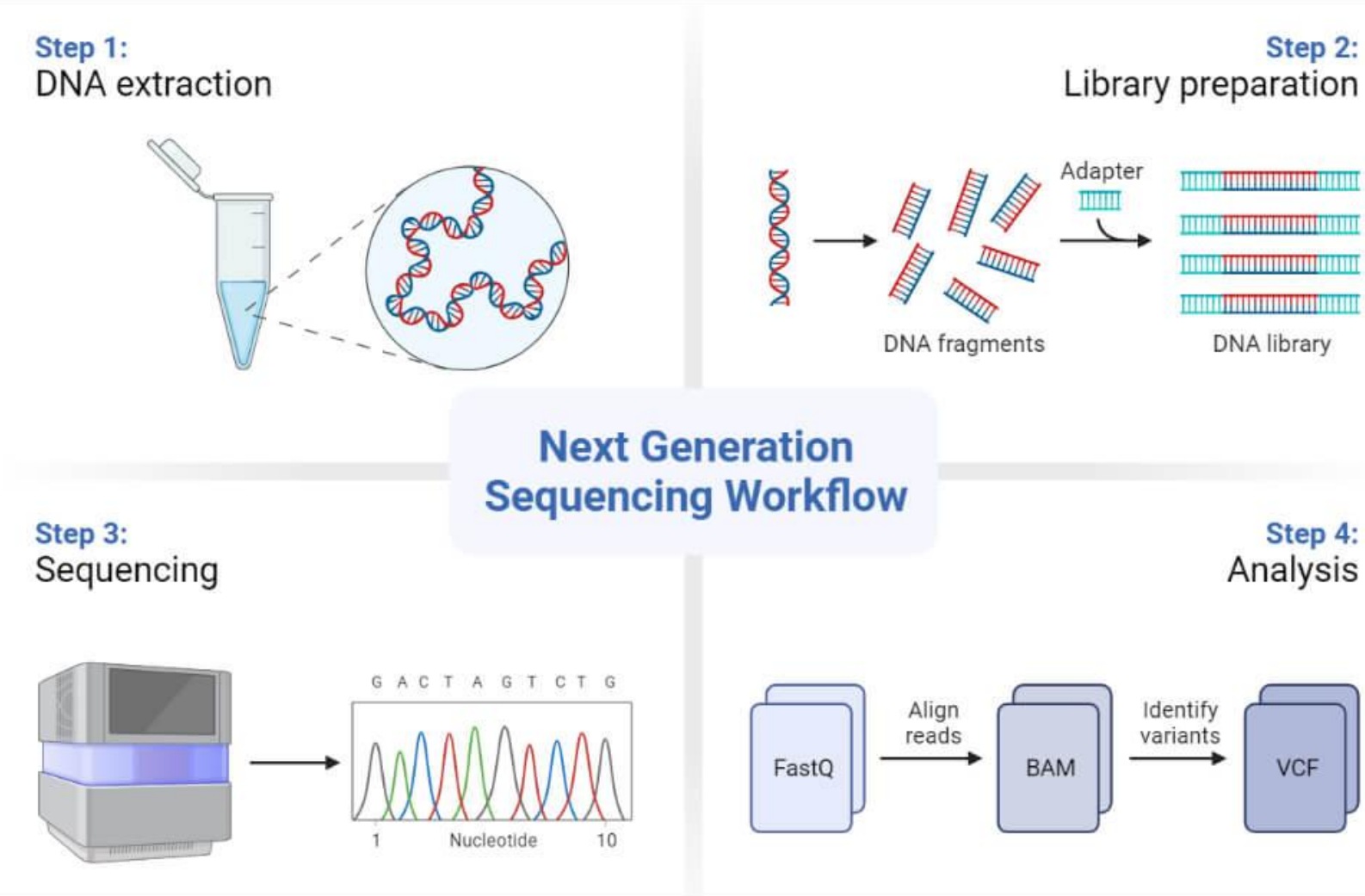


NGS basics

Next-generation sequencing

- Millions to billions of fragments sequenced in parallel
- Fluorescent imaging of each base incorporation
- PCR amplification on a flow cell (clusters)
- Advantages
  - high accuracy
  - high throughput
  - low cost
- Disadvantages
  - short reads (100–300 bp)

Short-read NGS workflow



NGS Analysis workflow

Question

Raw data

Pre-processing

Assembly – mapping or de novo

Variant calling

Post-processing

Comparison

Answer

History of NGS sequencers

- 2005 — 454
  - First massively parallel sequencing platform (pyrosequencing).
- 2007 — Illumina (Solexa)
  - Sequencing by synthesis
  - Dominant global platform
- 2010 — Ion Torrent
  - Semiconductor-based sequencing
  - Inexpensive and fast but suffers from homopolymer errors
- 2015 — BGI/DNBseq
  - DNA nanoball–based sequencing-by-synthesis
  - Cheaper, major international competitor to Illumina

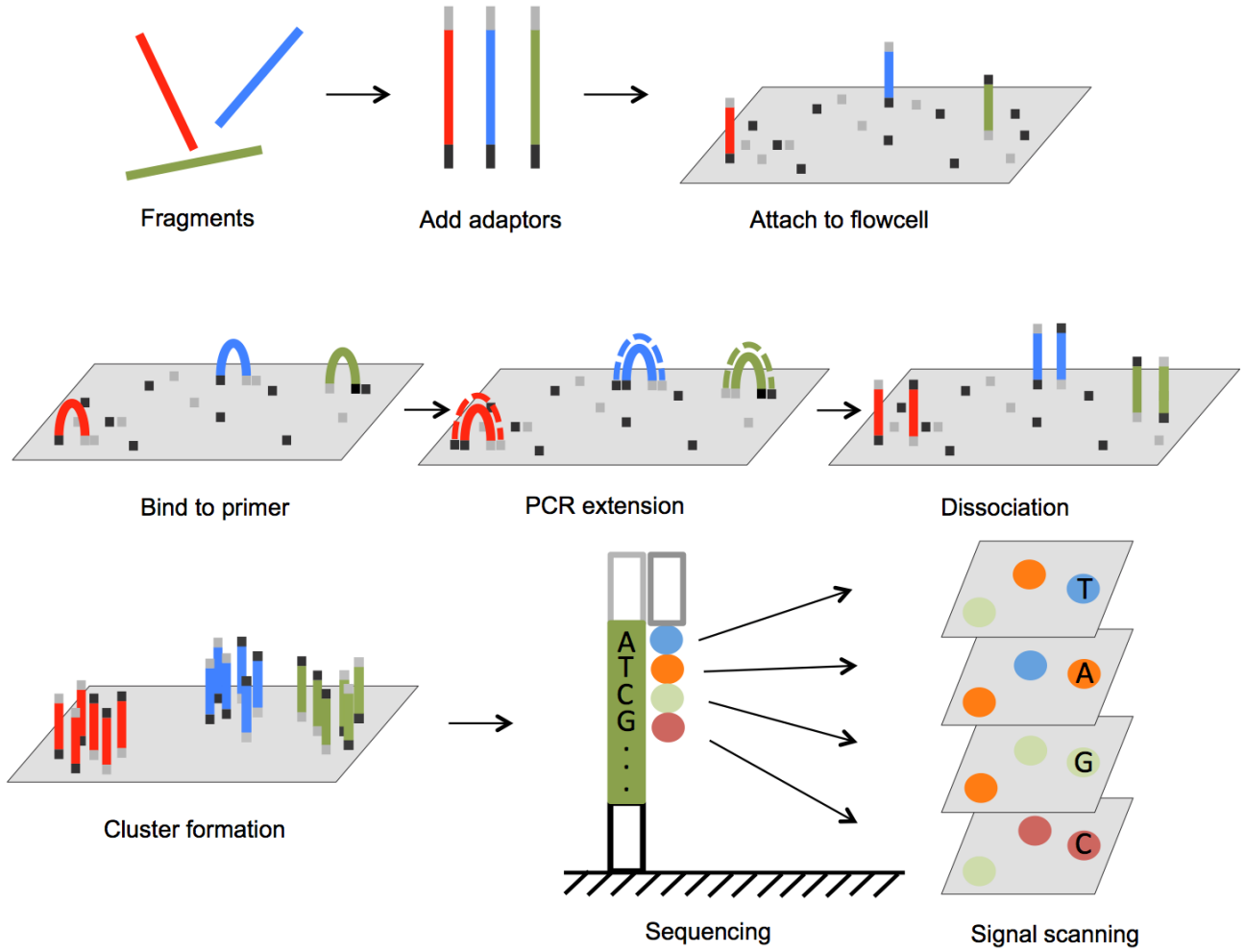
History of NGS sequencers

- 2005 — 454
  - First massively parallel sequencing platform (pyrosequencing).
- **2007 — Illumina (Solexa)**
  - **Sequencing by synthesis**
  - **Dominant global platform**
- 2010 — Ion Torrent
  - Semiconductor-based sequencing
  - Inexpensive and fast but suffers from homopolymer errors
- 2015 — BGI/DNBseq
  - DNA nanoball–based sequencing-by-synthesis
  - Cheaper, major international competitor to Illumina

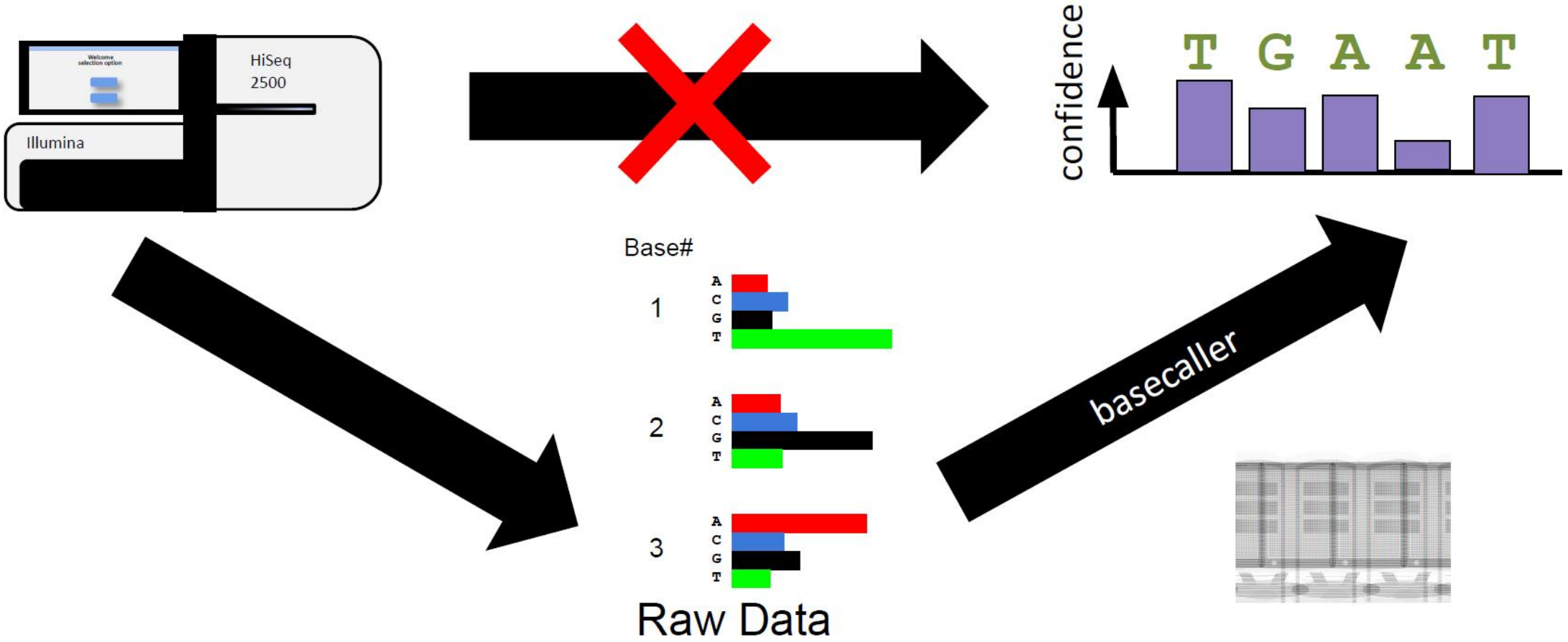
Illumina NGS

- https://www.youtube.com/watch?v=fCd6B5HRaZ8

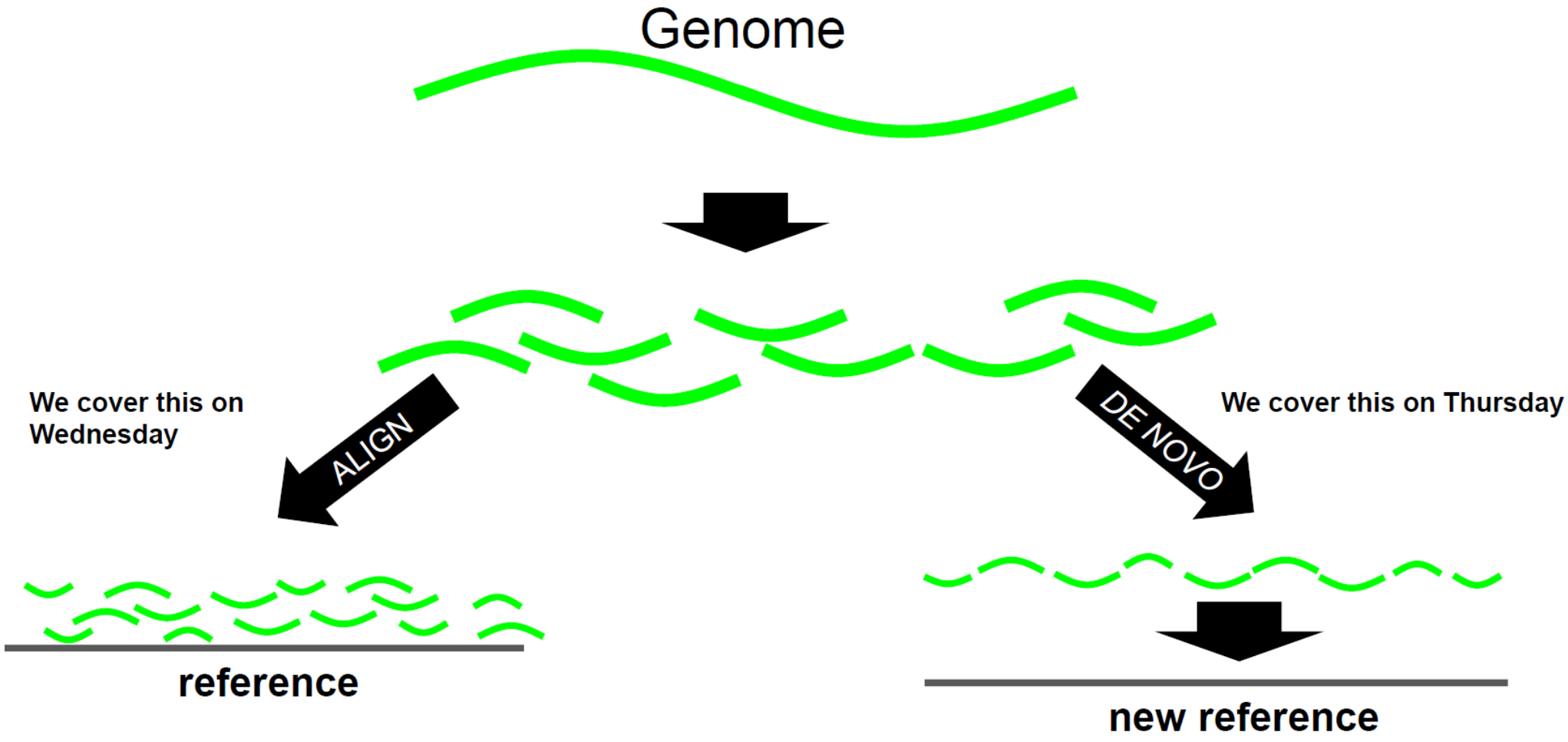
Illumina NGS



Short-read NGS



Short-read NGS workflow



Key terms

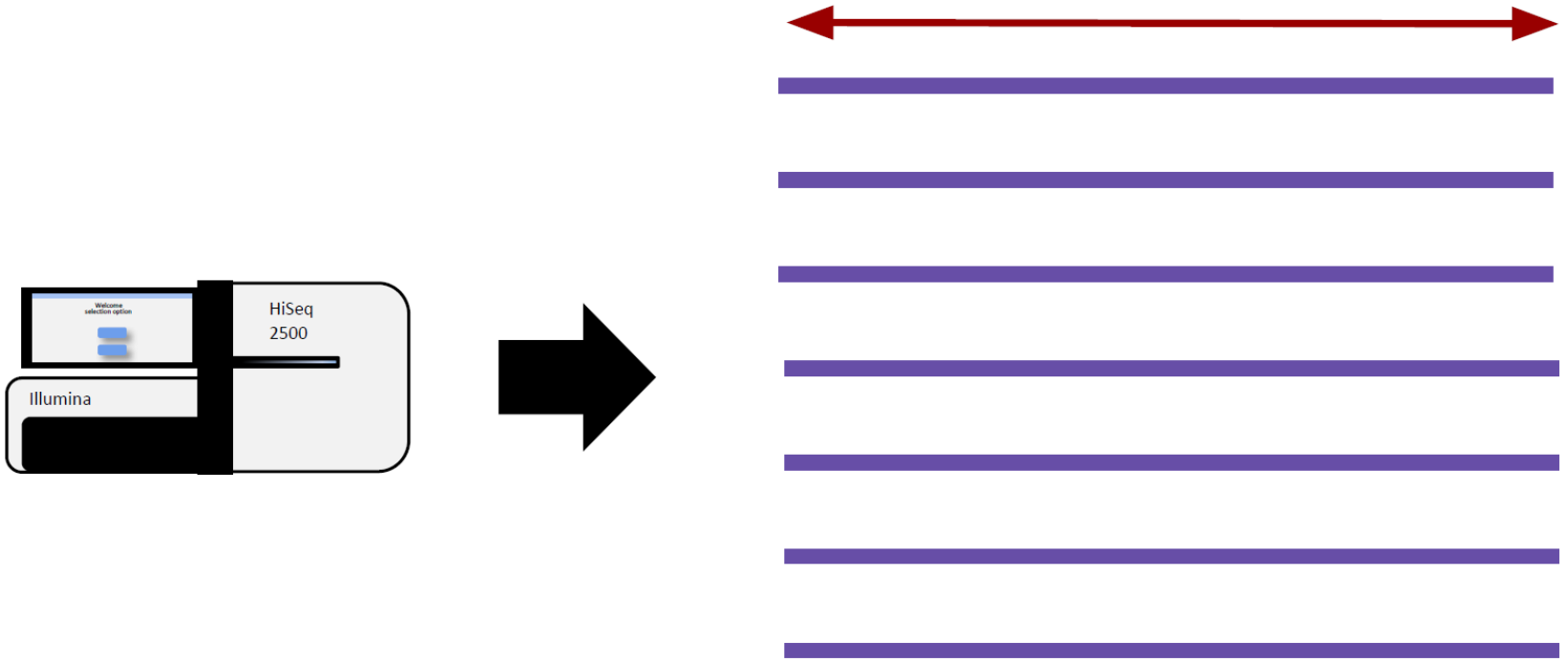
- DNA fragment
- Sequencing read
- Base call
- Errors
- Mapping/alignment
- Assembly

Key concepts

- **Read length**
  - How much of a fragment is read at once
- **Coverage**
  - How many times each base is read
- **Error types**
  - Mismatches, insertions, deletions

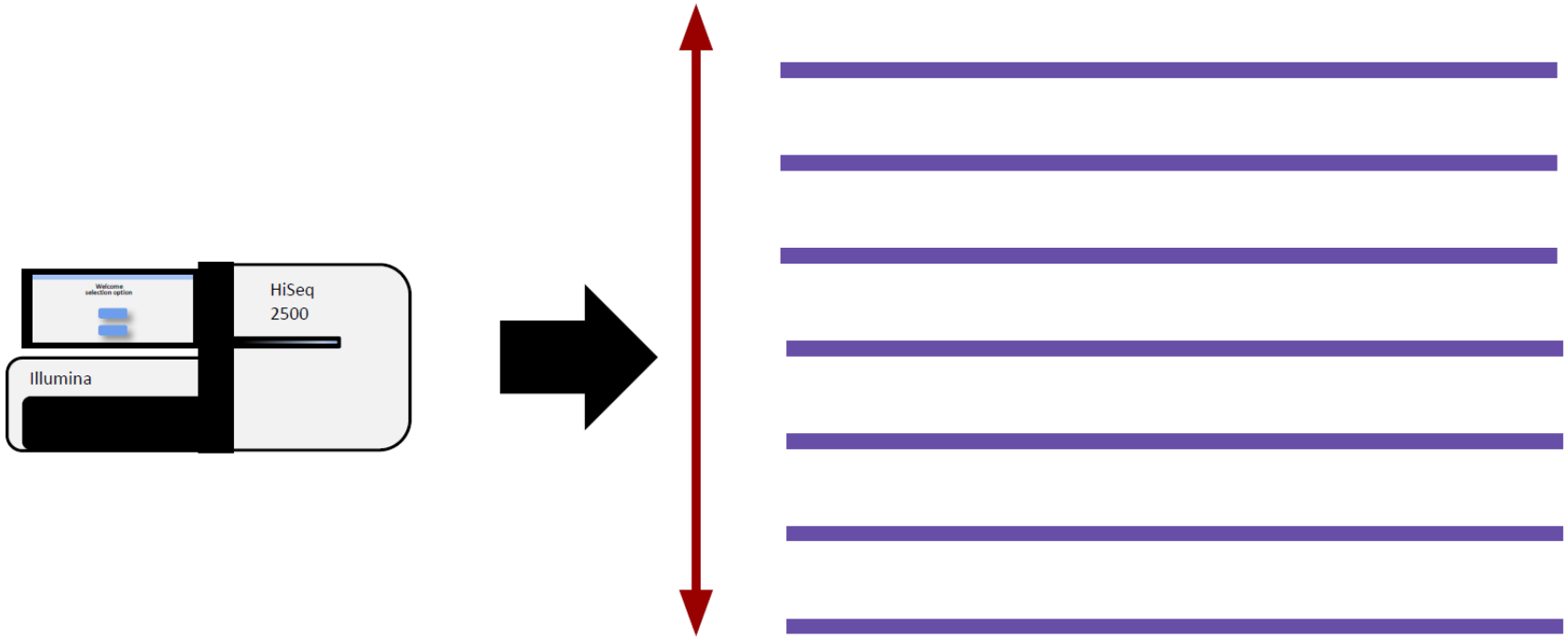
Key concepts

- **Read length**
  - –How much of a fragment is read at once



Key concepts

- **Coverage**
  - –How many times each base is read



Key concepts

- **Error types**
  - –Mismatches (wrong base)

DNA fragment

ACGTACGTACGTACGTACGGCTAGCT

ACGTACGTACGTCCGTACGGCTAGCT

Sequencing read

Key concepts

- **Error types**
  - –Insertions (extra base)

DNA fragment

ACGTACGTACGT-CGTACGCTAGCT

ACGTACGTACGTCCGTACGCTAGCT

Sequencing read

Key concepts

- **Error types**
  - –Deletions (missing base)

DNA fragment

ACGTACGTACGTACGTACGCTAGCT

ACGTACGTACGT-CGTACGCTAGCT

Sequencing read