DTU

DTU Health Technology
Bioinformatics

Introduction to NGS technology

Gisle Vestergaard
Associate Professor
Section of Bioinformatics
Technical University of Denmark
gisves@dtu.dk

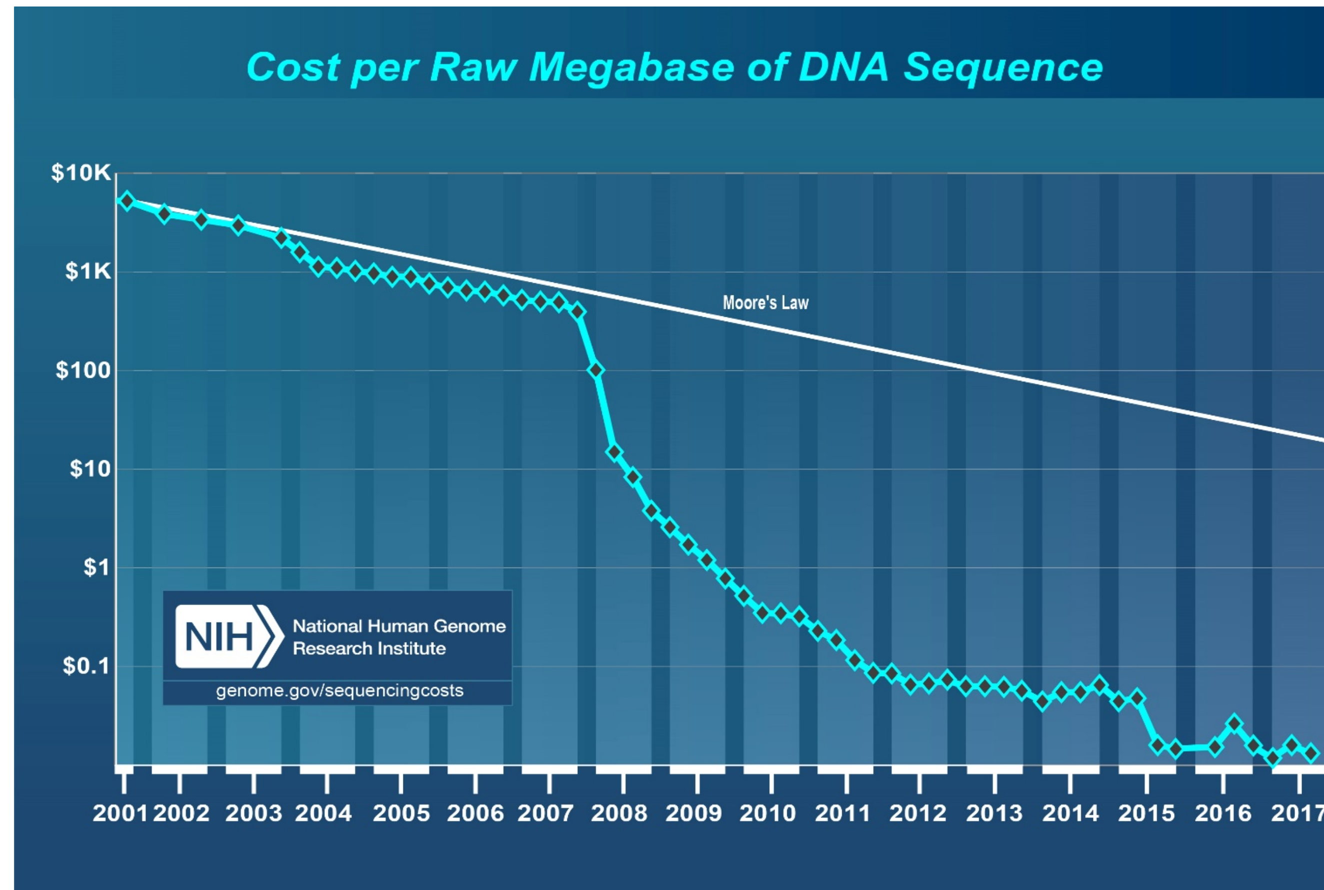
Menu

- Why NGS?
- 2nd generation NGS
- Illumina movietime!
- Your turn to basecall
- 3rd generation NGS

Development in metagenomics is linked to technological advancements

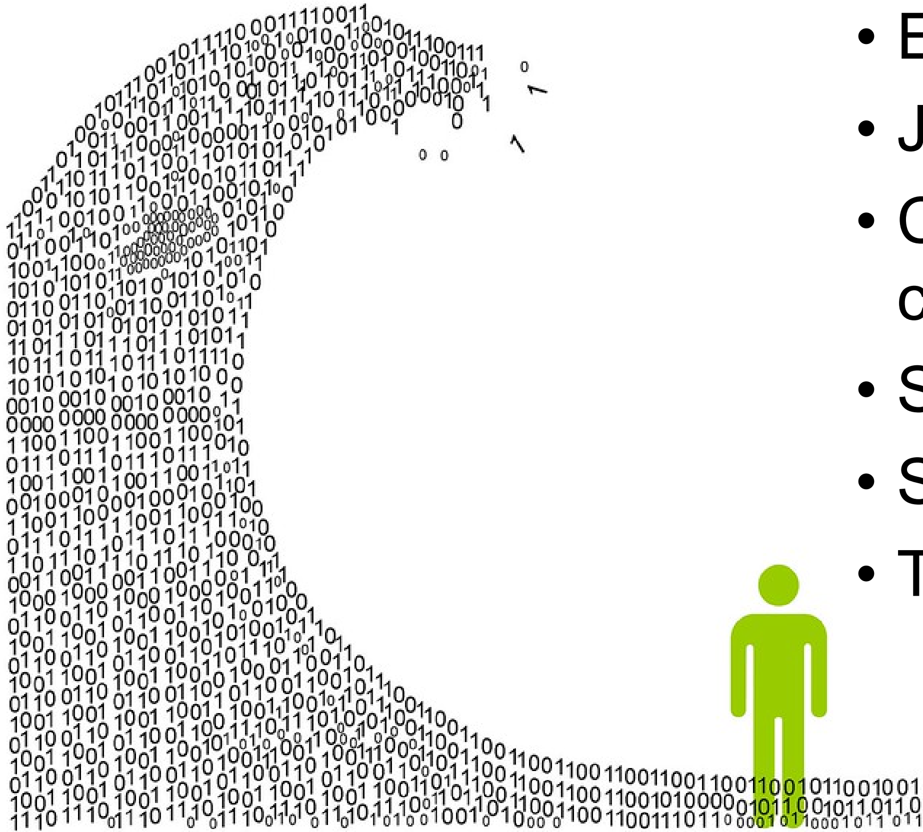
- Better, cheaper, faster

1 human Genome
3.000.000.000 USD
13 years



1 human Genome
< 1000 USD
1 day

NGS & bioinformatics



- Extreme data size causes problems
- Just transferring and storing the data
- Cost of storage and analysis is comparable to cost of sequencing
- Standard comparisons fail (N*N)
- Standard/old tools can not be used
- Think in fast and parallel programs

How it works

Second generation sequencing

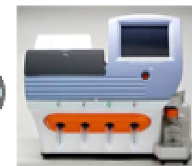
- Illumina sits on 90% of the market

Illumina

BGISEQ

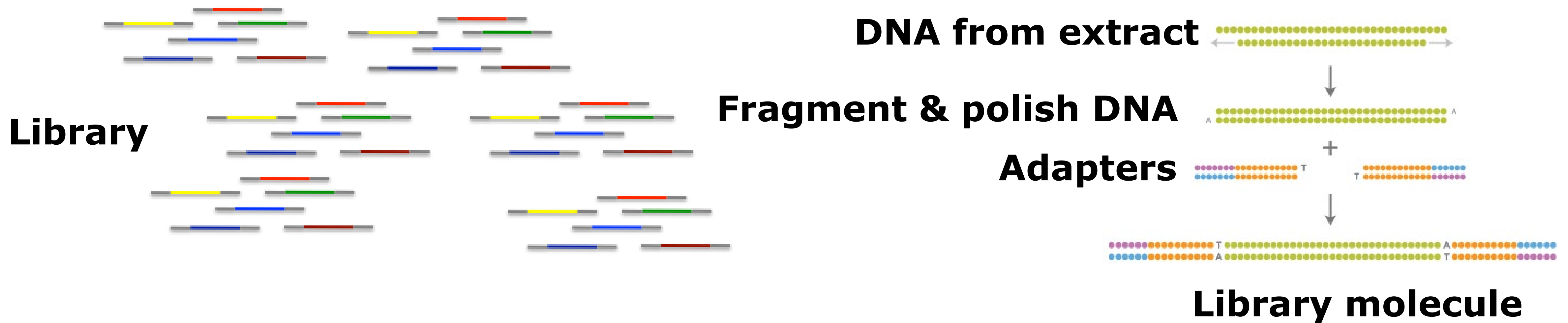
454

Ion Torrent (PGM)



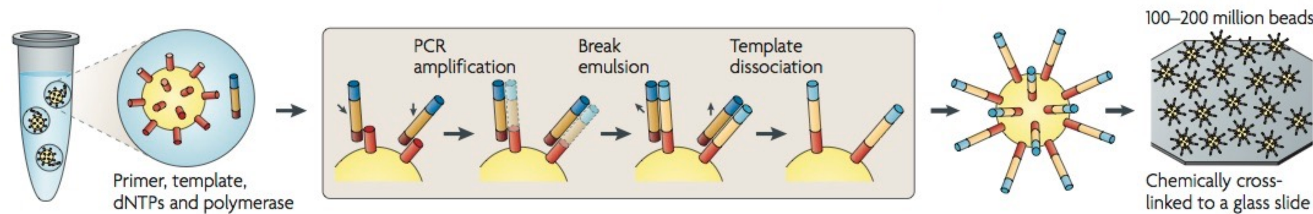
Library preparation

1. Create library molecules
2. Amplification (PCR)
3. Massive parallel sequencing

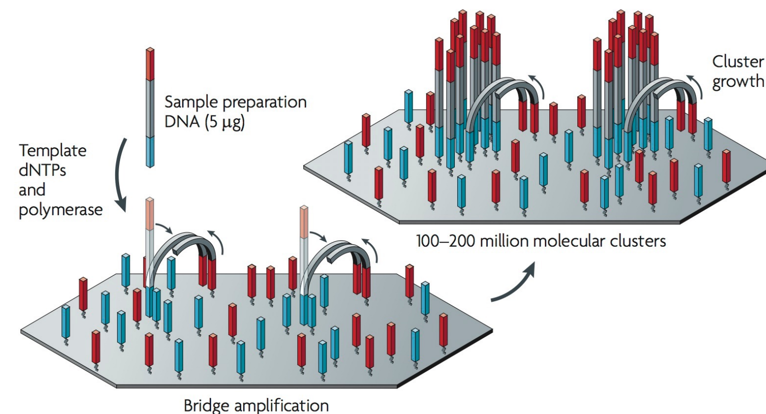


Amplification and immobilization

- Emulsion PCR (454, Solid, IonTorrent): Water, oil, beads, one DNA template/droplet



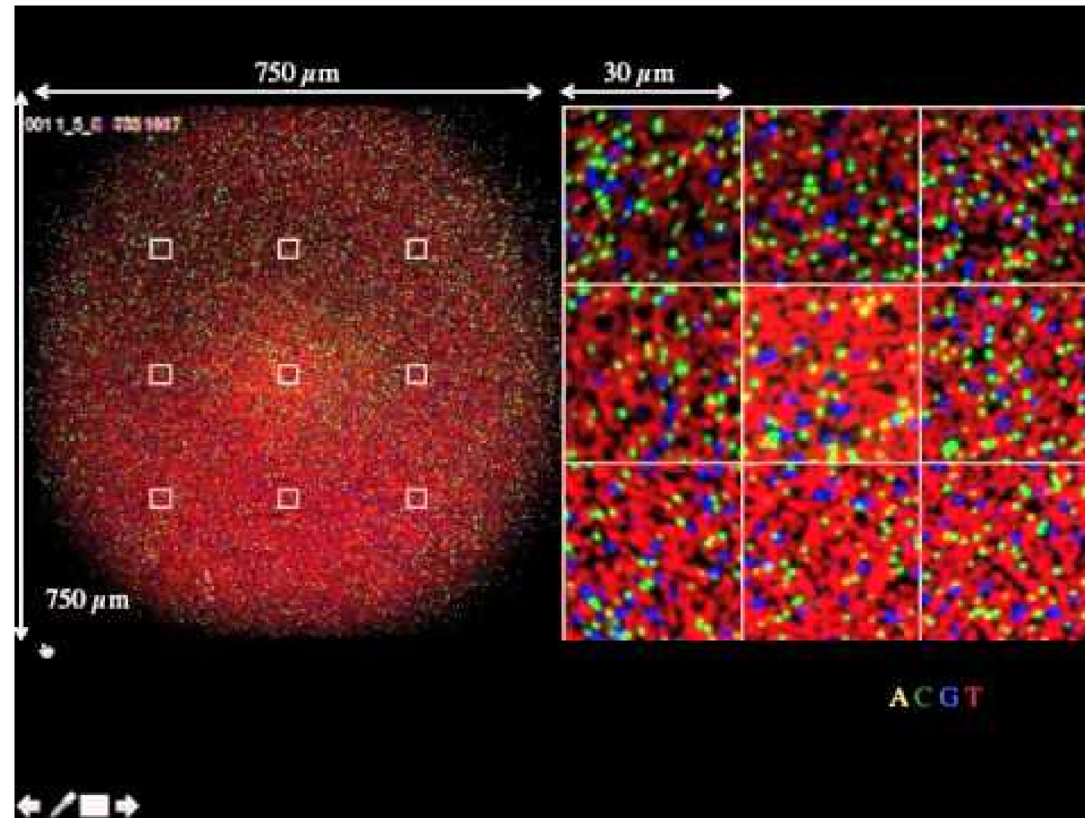
Bridge PCR (Illumina): One DNA template/cluster, primers on surface, grow by bridging primers



Illumina sequencing

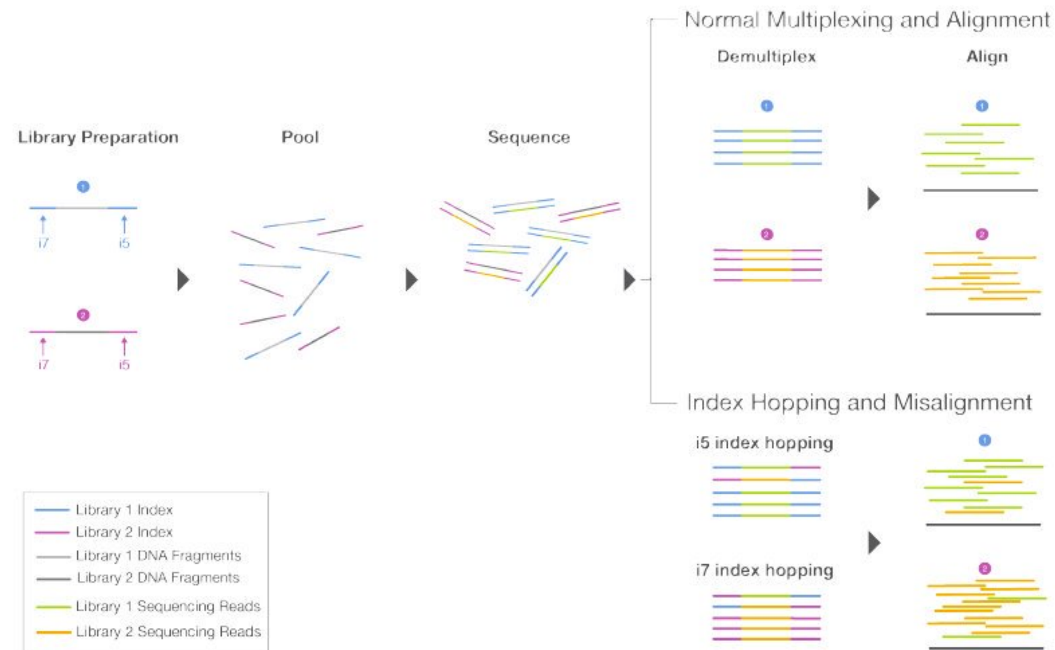
Amplicon sequencing on Illumina

- Why can't you just fill your Illumina flow cell with amplicon libraries?



Index hopping

- Index hopping in some cases creates several percent cross-talk
- Dual indexing eliminates index hopping



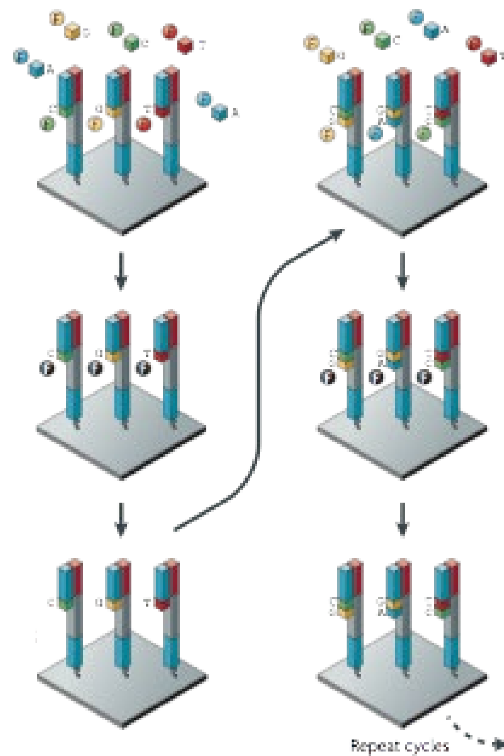
Fluorescence detection

Illumina - Cyclic reversible termination

Add all dNTPs labelled w. diff dye

Create four-color image

Cleave dye and repeat next cycle



Pyrosequencing

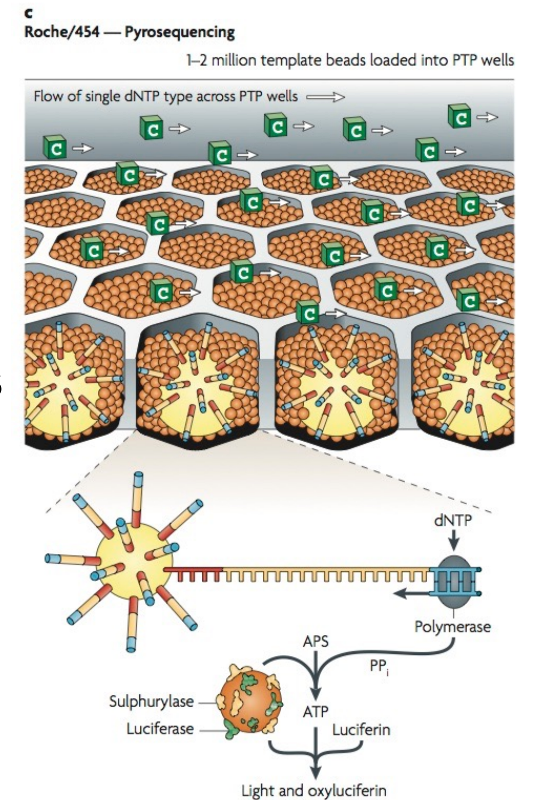
Load template beads into wells

Flow one dNTP across wells

Polymerase incorporates nucleotide

Release of PPi leads to light

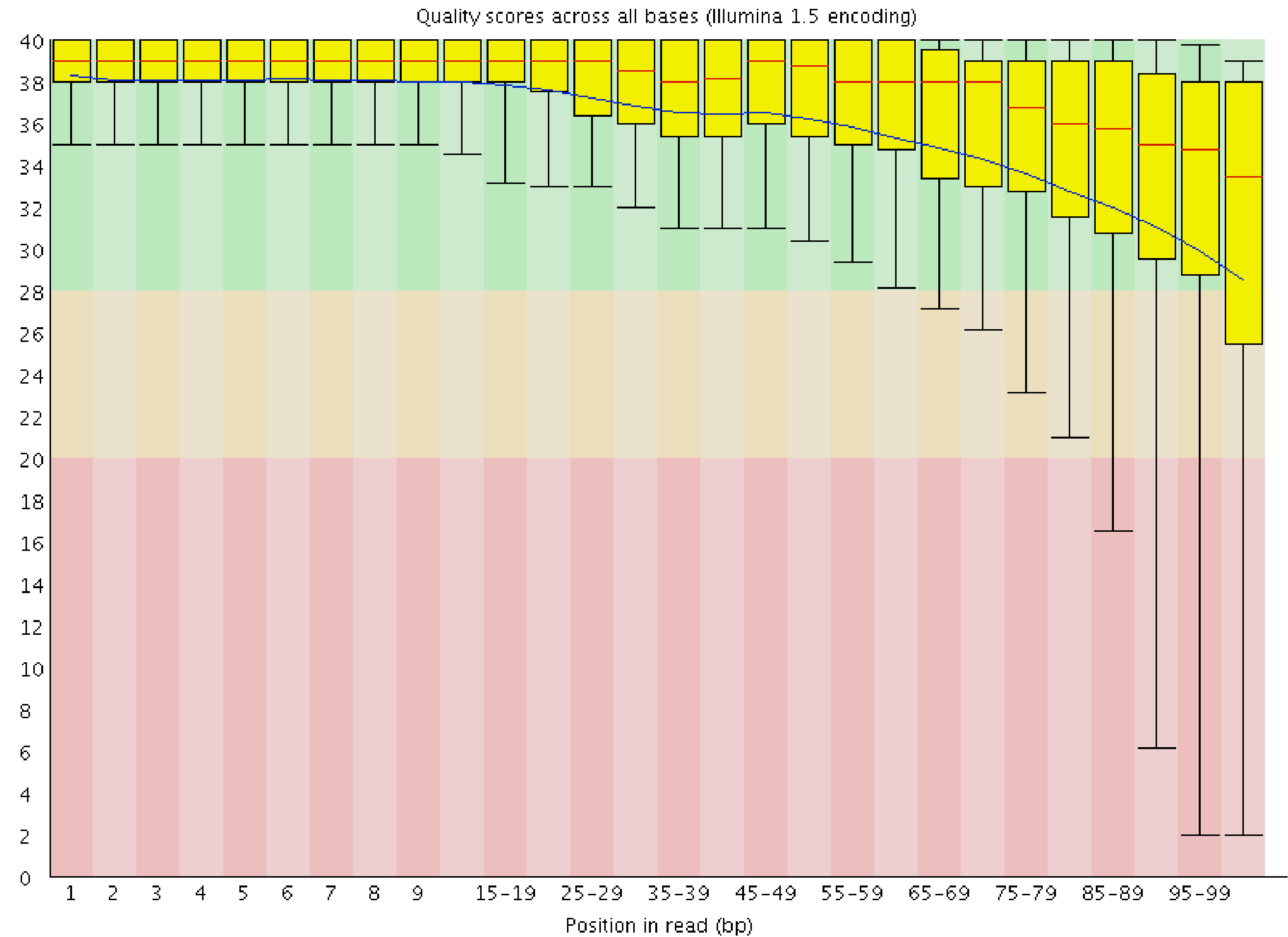
Imaging, next dNTP



Illumina: Quality deterioration

- Quality goes down
- Especially 2nd read
- Can you think of why?

- Efficiency of incorporation
- Polymerase incorporation of base
- Enzyme that cleaves the dye

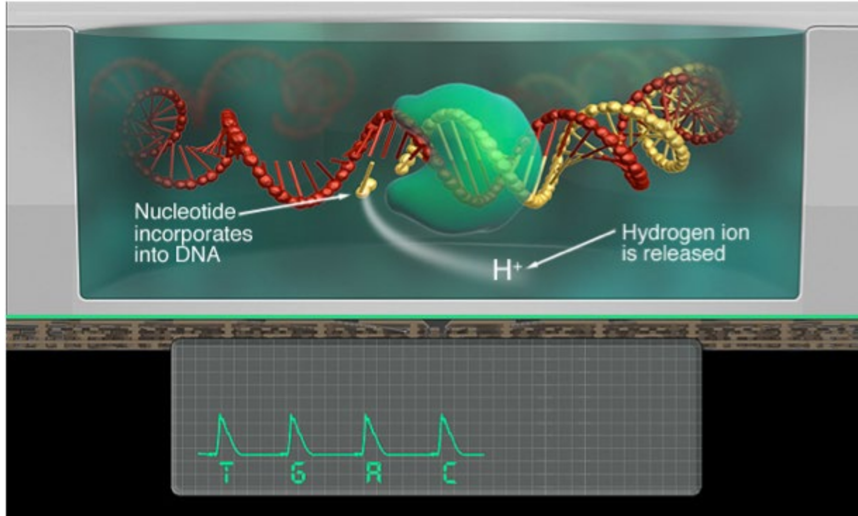


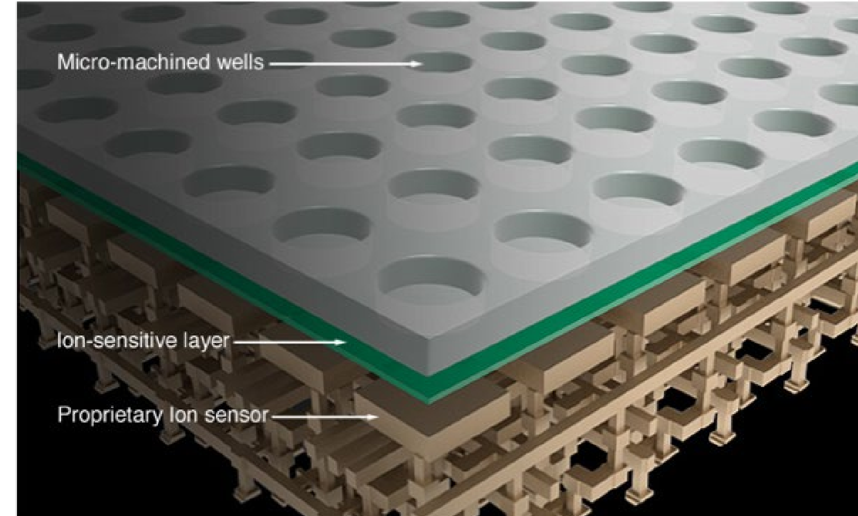
NextSeq/NovaSeq

- Chemistry is not based 4 dyes (as before) but 2 dyes
  - T (red), C (green), A (both) and G (none = “dark”)
  - Faster processing rate and cheaper reagents
  - Slightly increases error rate
  - Problem with G stretches because G is not dyed

Ion Torrent

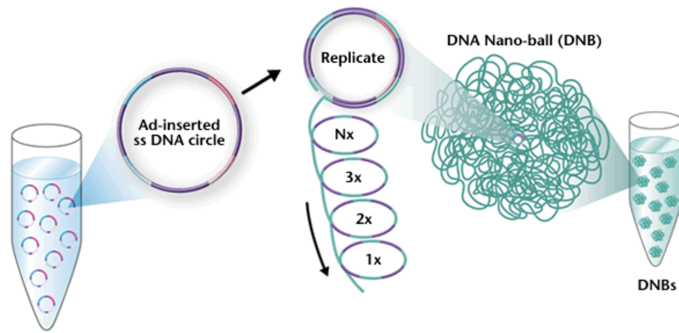
- Similar principle to 454
- Library: Emulsion PCR
- Based on semiconductors
- Detection is based on H ions (pH) changes



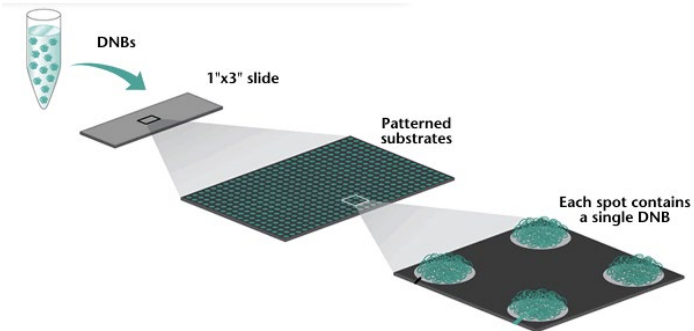


BGI-Seq

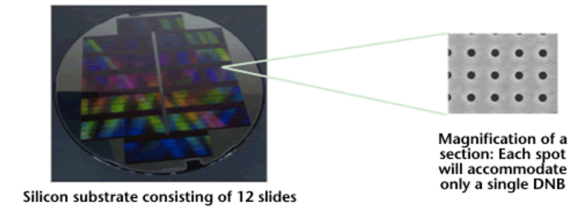
ssDNA -> DNA nanoballs



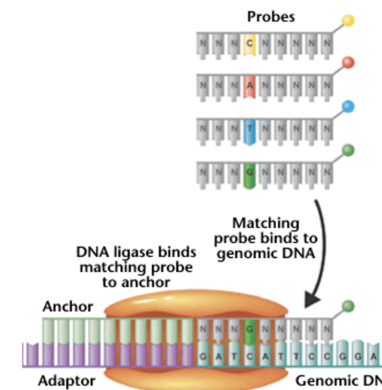
Place DNBs into each spot



Use silicon chips with sticky spots



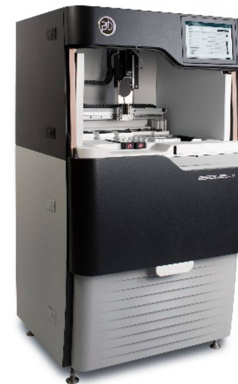
Sequence using ligase and
flourescent labeled probes



3rd generation

- Single-molecule sequencing
- No amplification -> less bias

Helicos

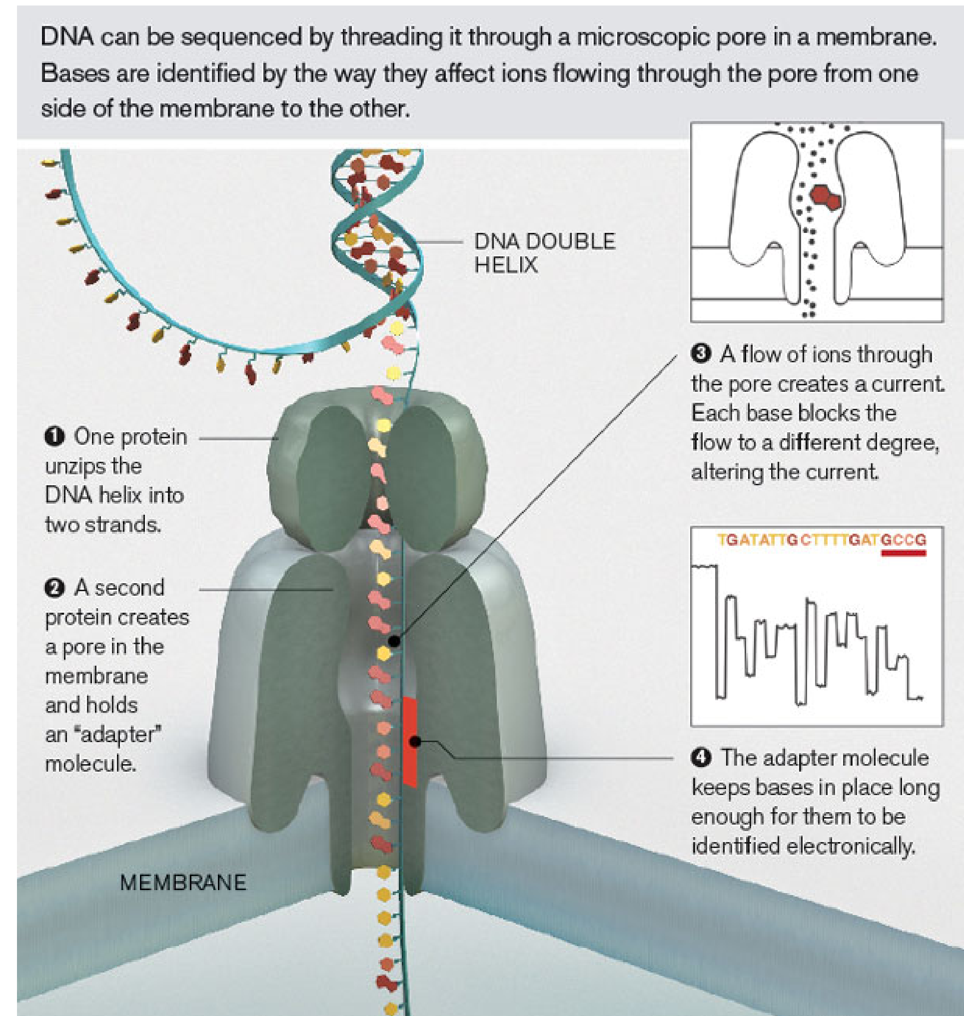


PacBio

Oxford Nanopore

Oxford Nanopore

- Litteral nanopores
- Current per base
- Non-random errors



Cheap & mobile

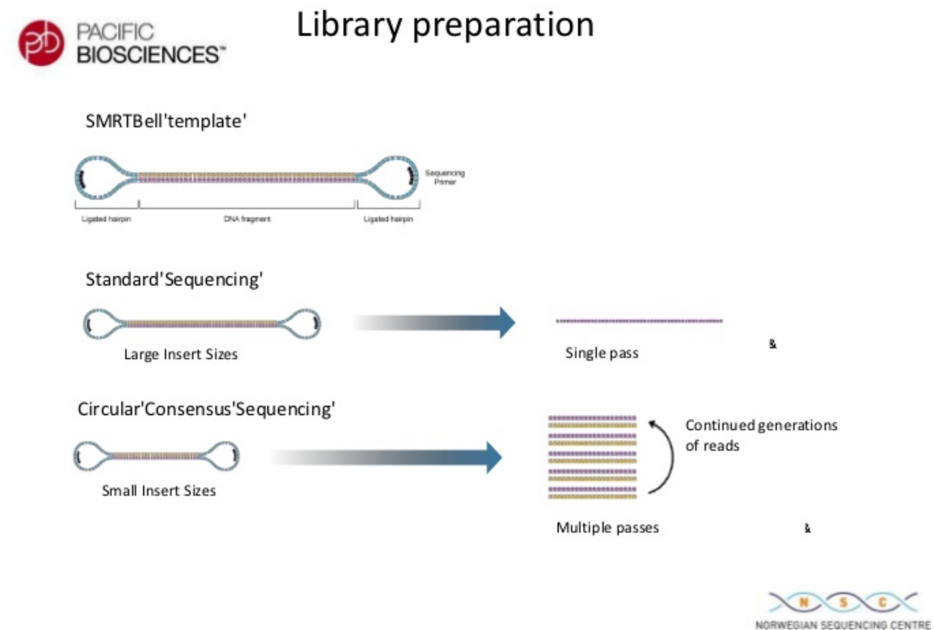
- Long reads, low quality
- Low establishment and maintainance costs
- Portability

PacBio

- Expensive machinery
- Not very portable

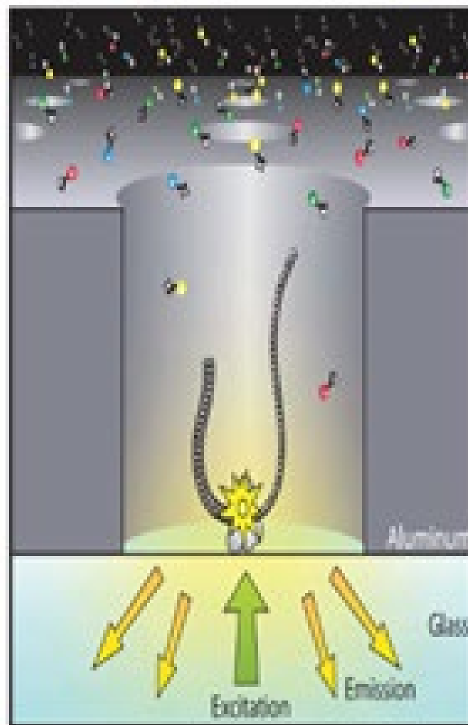
Flexible PacBio

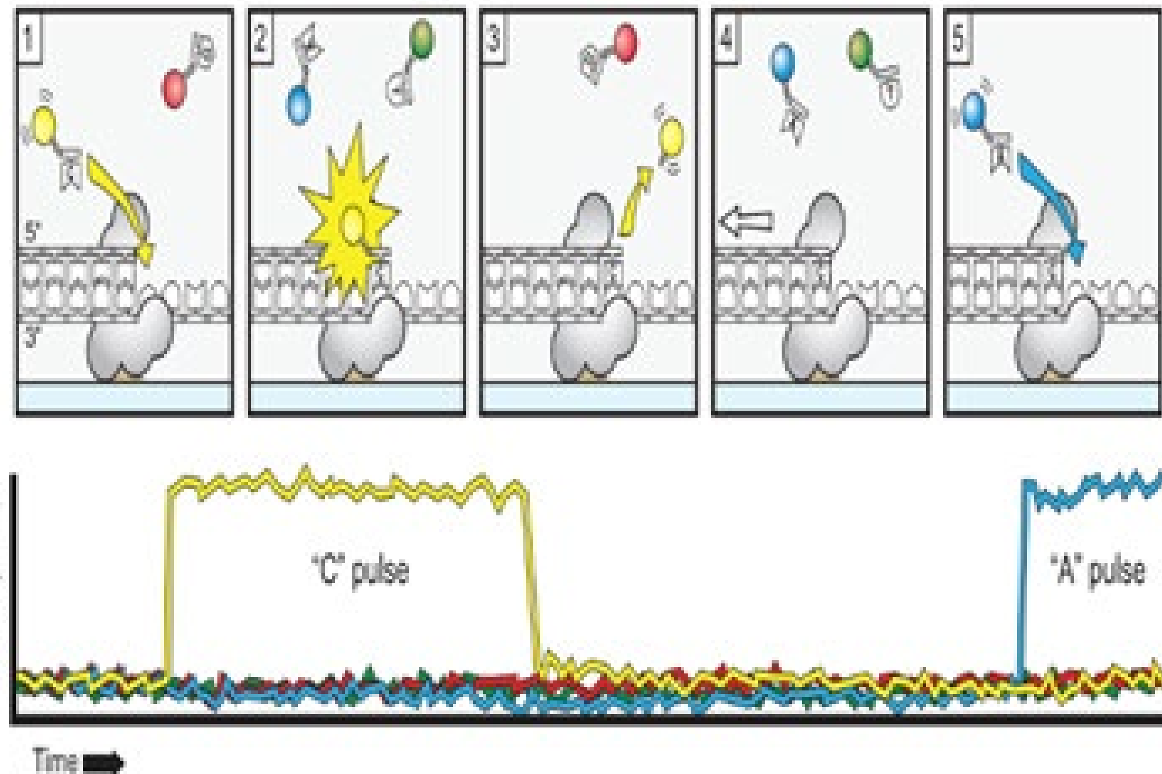
- Flexibility
  - Long but low quality or shorter but better reads
  - Robust



Tiny wells

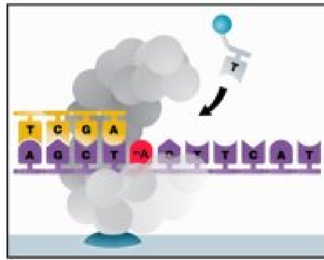
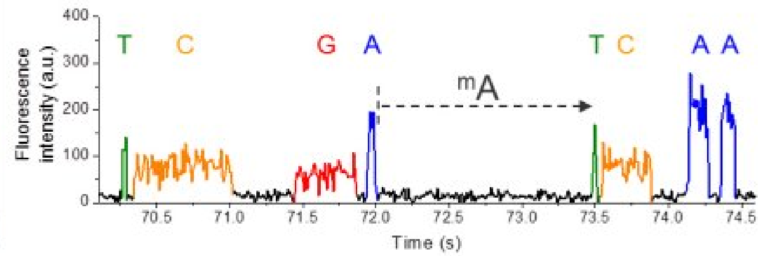
- 1 million wells per cell
- Hit the lights

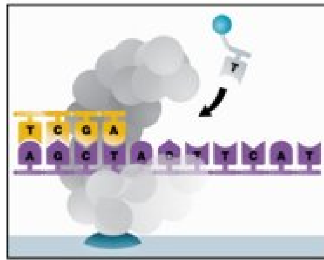
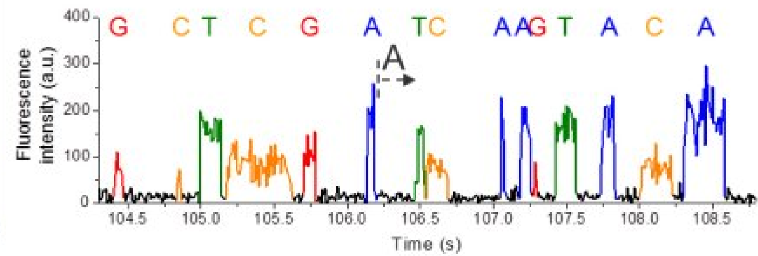




Epigenetics

- DNA modifications an be detected
- Virus & plasmid hosts can be identified

Summary

- Illumina is the current workhorse
  - Great for many applications
- Long read technology
  - Adding information
  - Resolves difficult regions during genome assembly

Deliverable IV

- Prepare a **short report (1 page)** describing the difference between generations of sequencing methods (1st, 2nd, 3rd generation sequencing methods)
- You can choose **one example per generation**
- Should include: How it works (chemistry), data output, strengths & weaknesses
- Which sequencing platforms are most useful for metagenomic analysis Include distinguished publications/references
- Hand-in next Monday 23:59 to Gisle @ gisves@dtu.dk

Inspiration: Goodwin et al. Nature Reviews 2016 (on CN)