15/01/2024

Next-Generation Sequencing

- → Roche 454 (2006-2013)
- → Illumina → most popular (90%)
- → PacBio SMRT → single-molecule real-time
- → Ion Torrent
- → Oxford Nanopore → latest technology

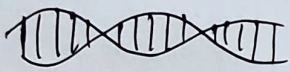
→ Data Analysis
→ Genomics
→ Transcriptomics
→ Assembly Algorithms

Sanger sequencing method:

di-deoxy termination method.

↓

capillary electrophoresis



- ✓ one DNA fragment at a time → Drawbacks
- ✓ we get around 1 kb (10^3 bp)
- ✓ took 13 years to sequence the entire human genome!

Human Genome Project

Shotgun sequencing

NGS Technologies:

- ✓ direct sequencing
- ✓ high-throughput, accurate ~ 96 fragments together
- ✓ cost-effective ~ $1000 & can go down

Atlas Project

↳ expression of genes changes during diseases → The Cancer Genome Atlas (TCGA). ← dataset to correlate genetics & drugs

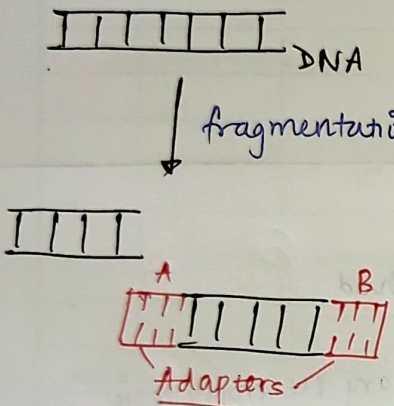
Precision Medicine ← useful in patient-specific therapy.

Reads:

- ✓ output sequences from sequencing platforms
- ✓ of different lengths

Roche 454 Sequencing:

↳ Pyrosequencing



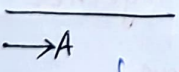
↳ synthetically-designed DNA sequences of known sequences → primer-binding sites → as we do not know the seq. of fragment yet

↳ Index sequence/ DNA barcode → to identify each sample differently

Multiplexing → mixing multiple samples & sequencing them together

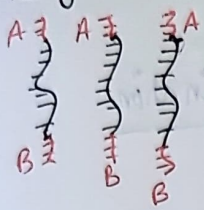
Pyrosequencing:

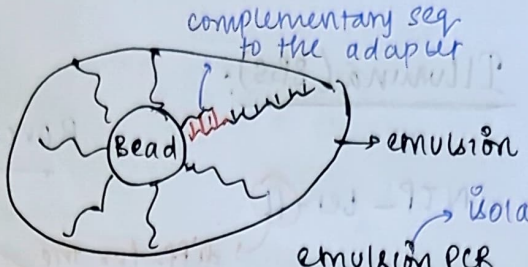
↳ relies on synthesizing the complementary strand



↳ every time there is an addition of a base, a light signal is emitted → if two same bases simultaneously → double the intensity

fragment



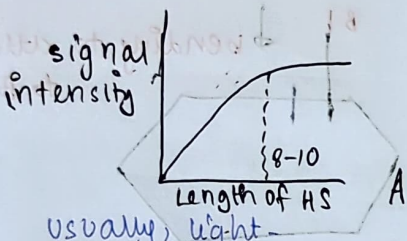


emulsion PCR → isolation of fragments → clonally amplified.

picotitre plates → 100,000 fragments → put in sequencer

Drawbacks:

- ✓ many enzymes & substrates required
- ✓ expensive compared to Illumina, etc.



ATTTTTTTT...

homopolymeric stretches ~5-20

usually, light intensity increases proportionally with repeated bases

after ~8 repeats → signal is saturated → no more proportional increase

might appear as indels (insertion/deletions)

Illumina – Sequencing by Synthesis (SBS):

Reversible termination (sanger has irreversible termination)

modified dNTP

dNTP – ter – fl

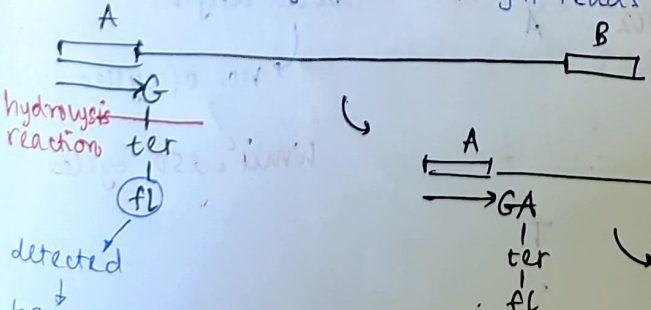
fl: fluorophore

ter: terminator

can be removed by a hydrolysis reaction

one base at a time

n cycles → n length reads

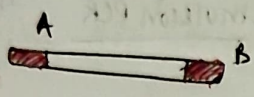


29/1/24

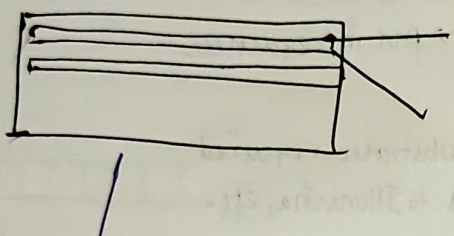
Illumina (SBS):

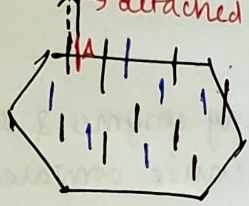
Reversible Termination

dNTP – ter-fl → diff. for the four bases: fl1, fl2, fl3, fl4

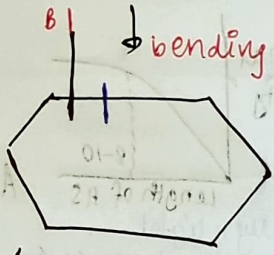


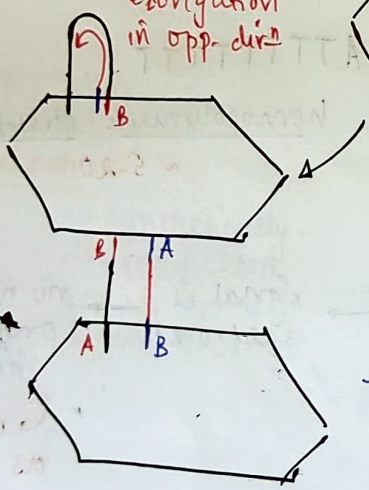
B → detached





Flow cell

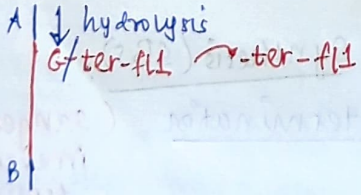




Bridge Amplification

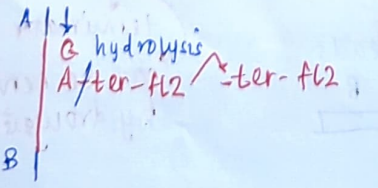
→ after many cycles, we get a cluster of one type of sample

camera



G

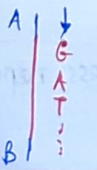
cycle 1



cycle 2

A

cluster has same bead length → no. of cycles → limit ~ 250 cycles



nth cycle

T

so on

in a cluster → ~100 molecules that are identical → detector reads average signal from a cluster

so after 250 ~~signals~~/bases, these avg. signals can be confusing & hard to decipher

in the adaptor seq:
- we add index sequences eg: ATTGCC
  - to identify sequences in a mixed sample

Multiplexing

Single-end (SE) and Paired-end (PE) sequencing:

↳ Illumina Data Processing.

↑ Throughput : high reads / lot of data

150 bp : SE seq.
2X150 bp : PE seq. (like in Illumina video).

Advantages:

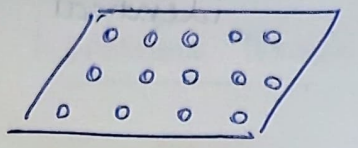
- cost-effective (not many enzymes are required).
- polymeric stretch → seq. one by one, not all at a time like Roche 454.
- highly accurate

Drawbacks:

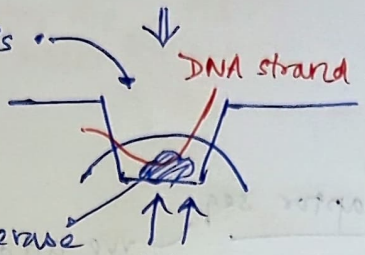
- expensive compared to the newer NGS ← even a μl costs in thousands.
- short reads (upto 2X250 bp)

Single Molecule Real Time Sequencing (SMRT): → [SBS]

↳ sequencing happens inside ZMWs.



SMRT-cell



everything else is similar to Illumina ←

immobilised at the bottom

light source → does not pass through, just illuminate the bottom

→ no amplification step is required.

→ no cluster formation

→ no averaging of signals → that's why 'single molecule'

↳ more sensitive detec[n]

as λ of light too large

signal detec[n] only at illumination

so incoming DNTPs do not produce signal

Advantages:

- long reads
- avg read length of 2–3 kb, some 10–20 kb
- single-molecule resolution

Drawbacks:

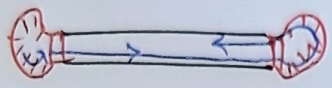
- high error rate (~14%)

Circular Consensus Sequence (CCS) reads → 99% accuracy

1. ATGCCTATT
2. ATCCCTATT
3. ATGCCAATT
4. ATGCATATT

take the most freq. at a place ↑

ATGCCTATT } Consensus sequence



multiple repeats
↓
circular synthesis.

polymerase at the bottom → always illuminated by light → leads to adverse effect on fidelity
↳ high energy state

use of Illumina & SMRT together → Hybrid Sequencing
- Illumina → short reads but high accuracy
- SMRT → large reads but low accuracy.