TOPIC 1. DNA-seq: techniques

Sanger sequencing. 2nd and 3rd generation sequencing techniques: 454, Illumina, SOLiD, IonTorrent, Oxford-Nanopore, PacBio.

Omics Techniques

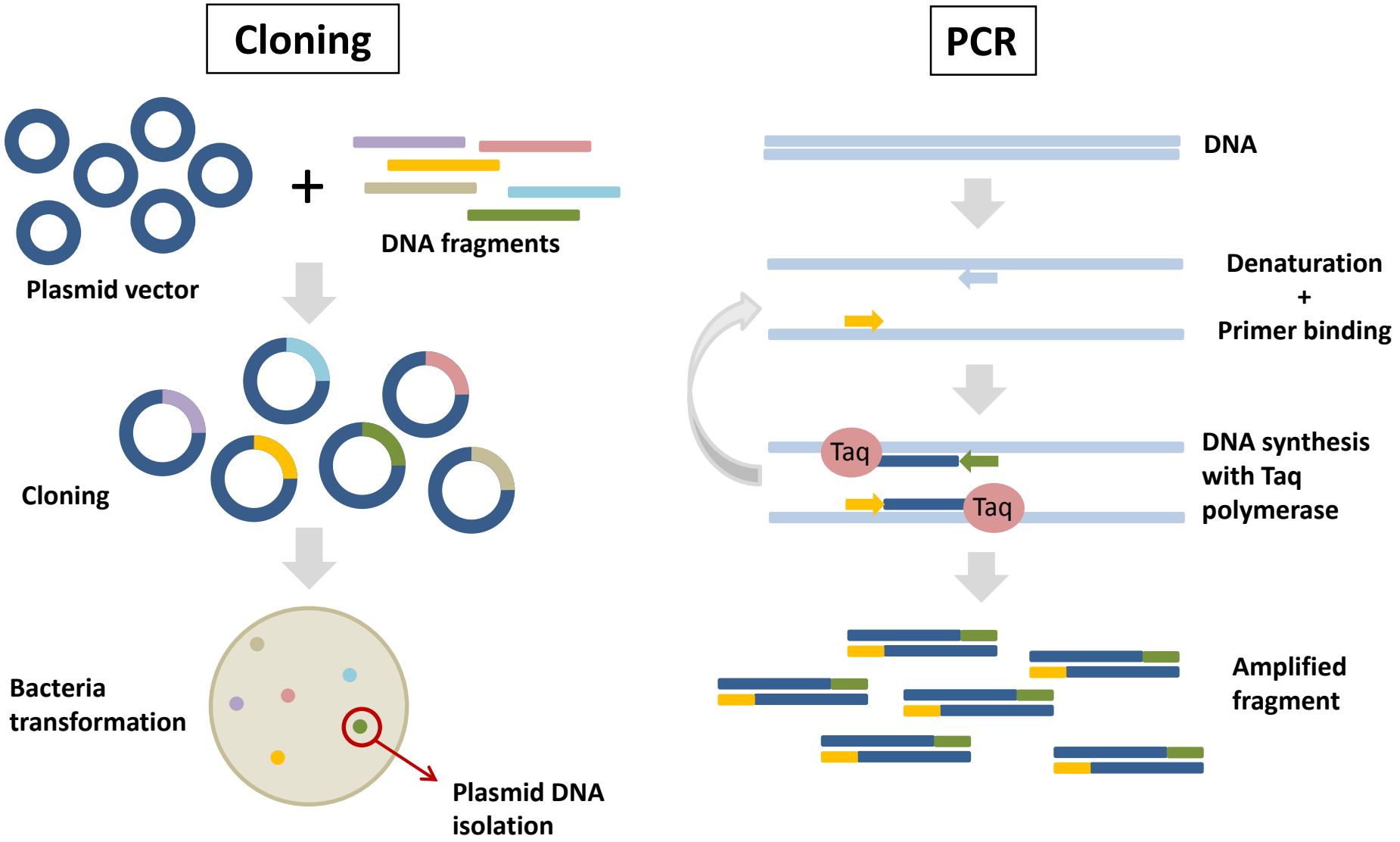
Bachelor's Degree in Bioinformatics

Sònia Casillas, UAB

Name different sequencing platforms

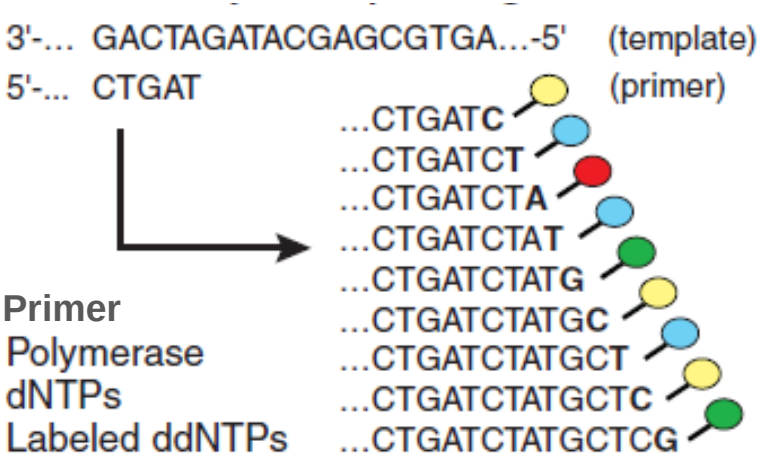
Sanger sequencing method

1. DNA amplification

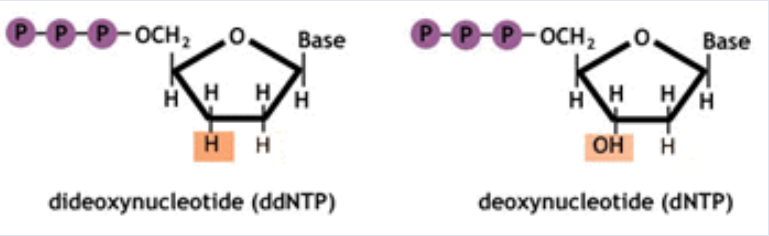


Sanger sequencing method

2. Sequencing reaction



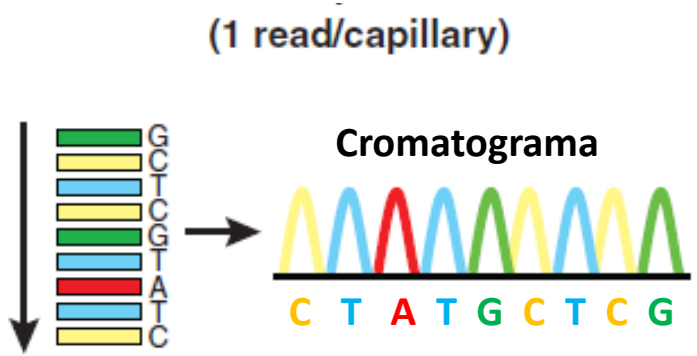
Dideoxynucleotides (ddNTPs)



No additional nucleotide can be incorporated

Each ddNTP is labelled with a different fluorochrome

3. Capillary electrophoresis

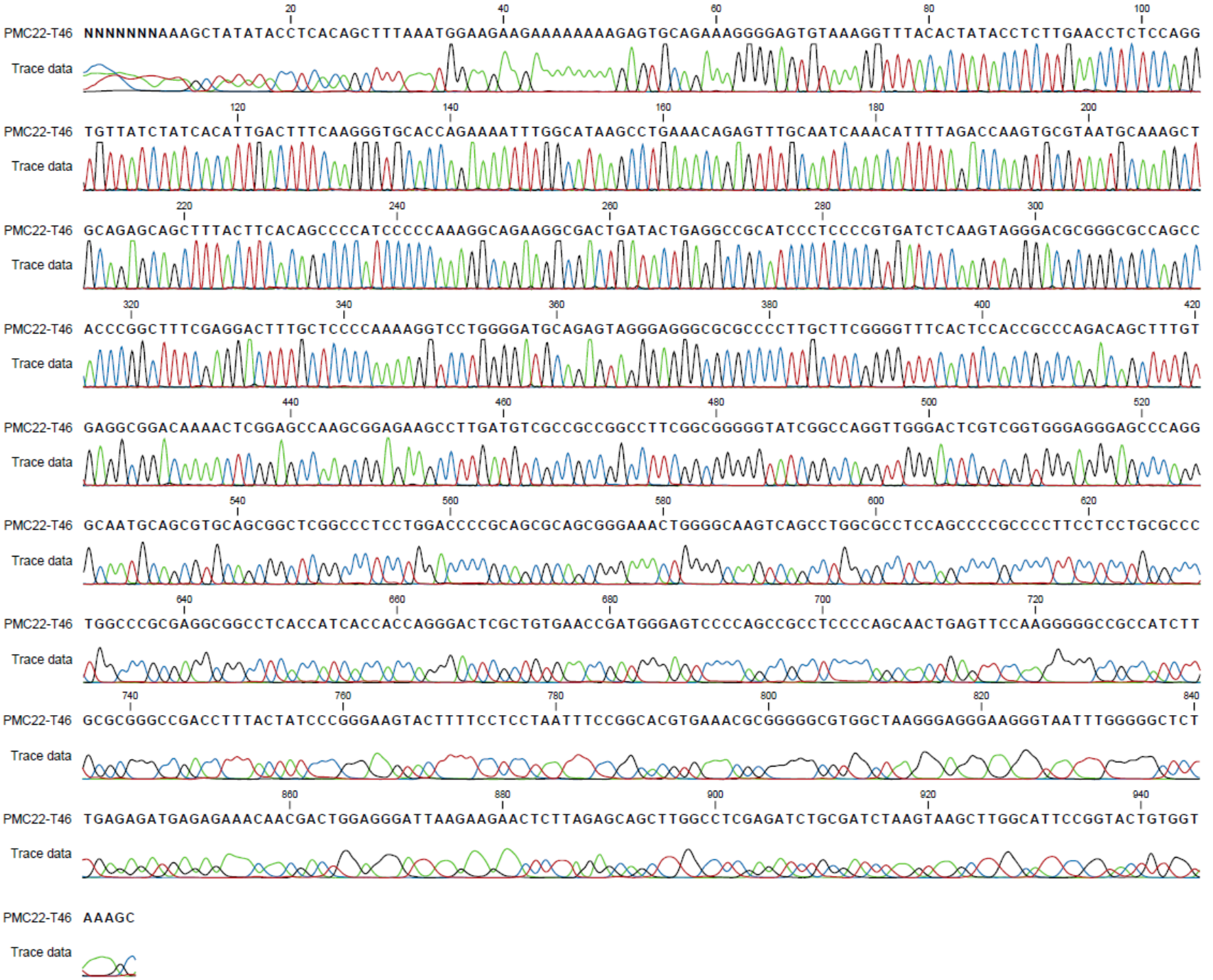


Size fragment separation in polyacrilamide gen + **Fluorescence detection**

RESULT OF SANGER SEQUENCING

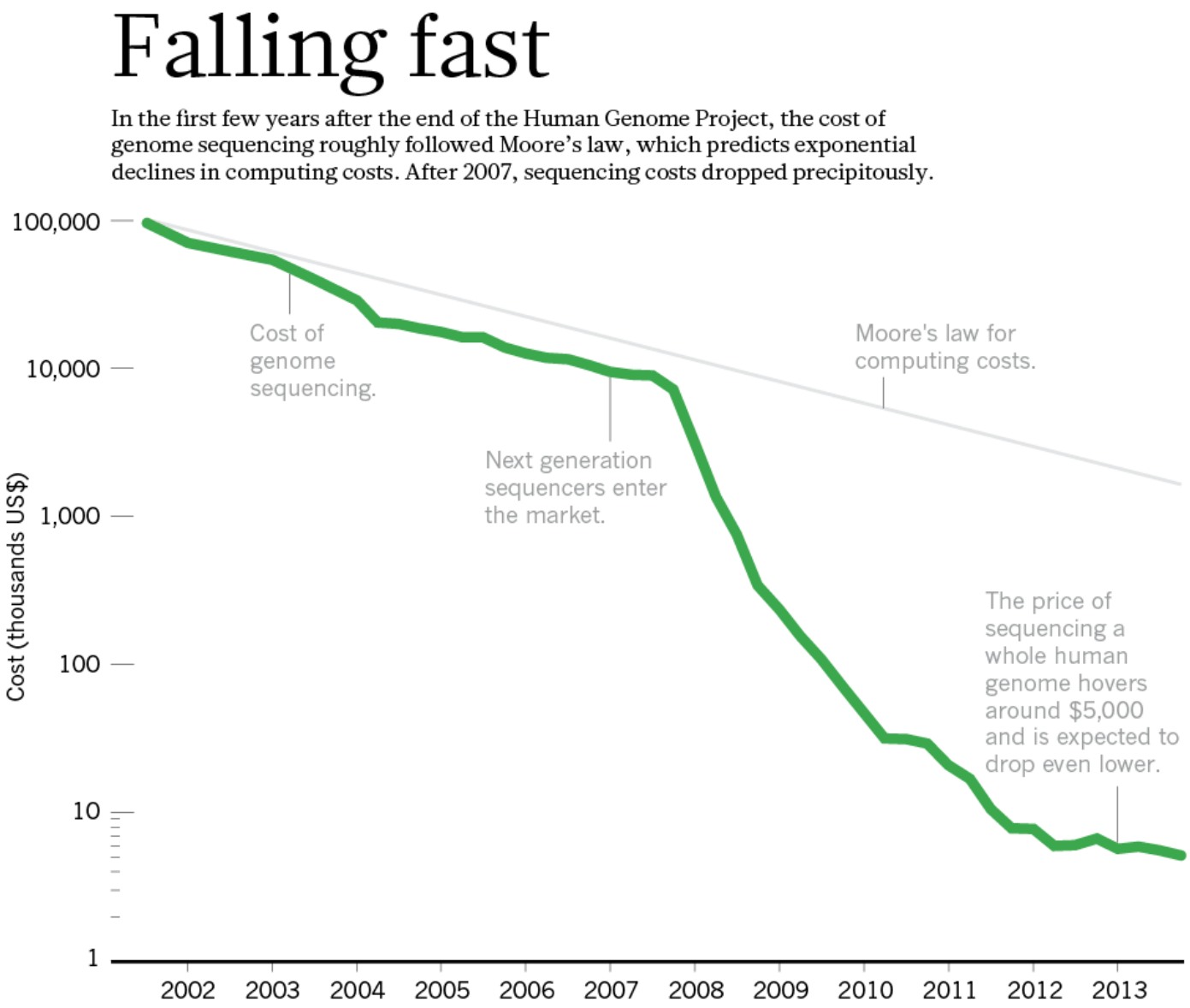
- Long reads (500-1000 bp)
- Low throughput (96 reactions/run)

Figure 1. Shendure and Ji (2008) *Nature Biotechnology* 26: 1135-1145.


PMC22-T46 NNNNNNNAAAGCTATATACCTCACAGCTTTAAATGGAAGAAGAAAAAAAAGAGTGCAGAAAGGGGAGTGTAAAGGTTTACACTATACCTCTTGAACCTCTCCAGG
Trace data
PMC22-T46 TGTTATCTATCACATTGACTTTCAAGGGTGCACCAGAAAATTTGGCATAAGCCTGAAACAGAGTTTGCAATCAAACATTTTAGACCAAGTGCGTAATGCAAAGCT
Trace data
PMC22-T46 GCAGAGCAGCTTTACTTCACAGCCCCATCCCCCAAAGGCAGAAGGCGACTGATACTGAGGCCGCATCCCTCCCCGTGATCTCAAGTAGGGACGCGGGCGCCAGCC
Trace data
PMC22-T46 ACCCGGCTTTCGAGGACTTTGCTCCCCAAAAGGTCCTGGGGATGCAGAGTAGGGAGGGCGCGCCCCTTGCTTCGGGGTTTCACTCCACCGCCCAGACAGCTTTGT
Trace data
PMC22-T46 GAGGCGGACAAAACTCGGAGCCAAGCGGAGAAGCCTTGATGTCGCCGCCGGCCTTCGGCGGGGGTATCGGCCAGGTTGGGACTCGTCGGTGGGAGGGAGCCCAGG
Trace data
PMC22-T46 GCAATGCAGCGTGCAGCGGCTCGGCCCTCCTGGACCCCGCAGCGCAGCGGGAAACTGGGGCAAGTCAGCCTGGCGCCTCCAGCCCCGCCCCTTCCTCCTGCGCCC
Trace data
PMC22-T46 TGGCCCGCGAGGCGGCCTCACCATCACCACCAGGGACTCGCTGTGAACCGATGGGAGTCCCCAGCCGCCTCCCCAGCAACTGAGTTCCAAGGGGGCCGCCATCTT
Trace data
PMC22-T46 GCGCGGGCCGACCTTTACTATCCCGGGAAGTACTTTTCCTCCTAATTTCCGGCACGTGAAACGCGGGGGCGTGGCTAAGGGAGGGAAGGGTAATTTGGGGGCTCT
Trace data
PMC22-T46 TGAGAGATGAGAGAAACAACGACTGGAGGGATTAAGAAGAACTCTTAGAGCAGCTTGGCCTCGAGATCTGCGATCTAAGTAAGCTTGGCATTCCGGTACTGTGGT
Trace data
PMC22-T46 AAAGC
Trace data

Large sequencing facilities

The quest for the $1,000 genome



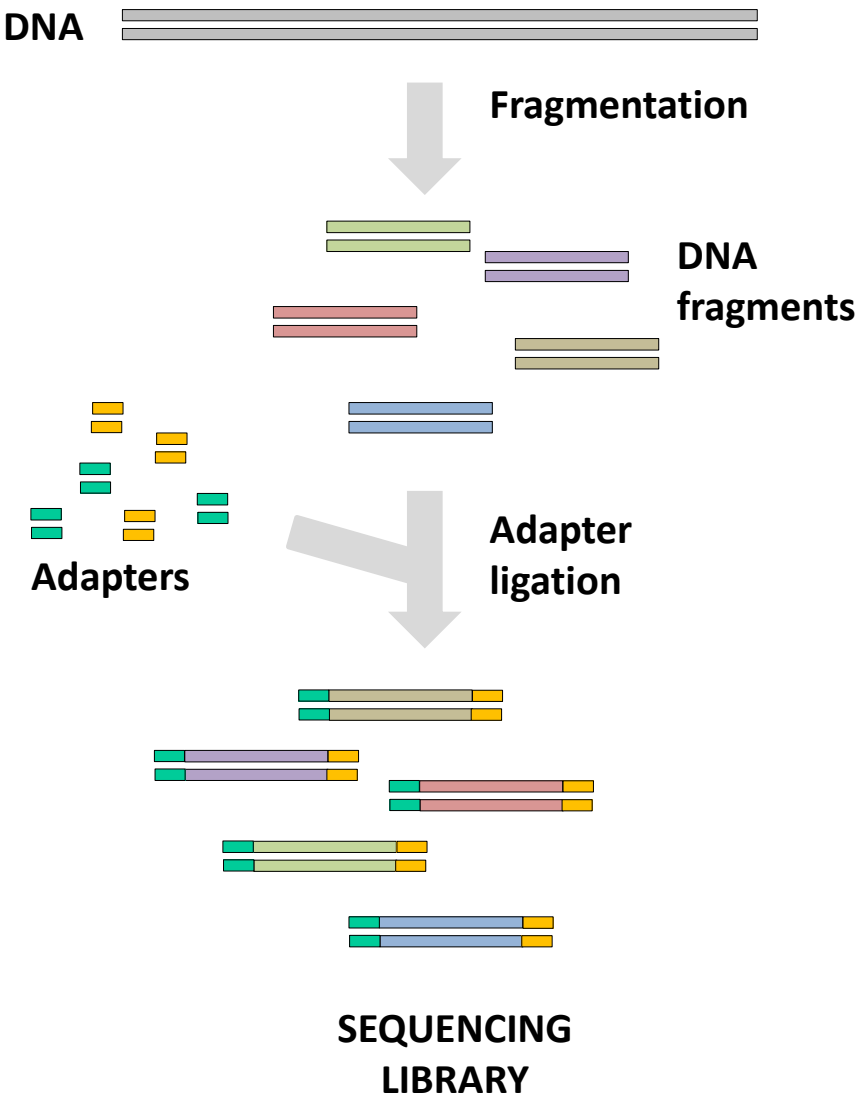
Next generation sequencing (NGS) methods

Next generation sequencing instruments can generate as much data in one day as several hundred Sanger DNA capillary sequencers!

Common characteristics of NGS methods

1. Cell-free preparation of sequencing library
2. Solid-phase amplification
3. Massively parallel sequencing reaction of each DNA fragment independently
4. Direct sequencing without need of electrophoresis

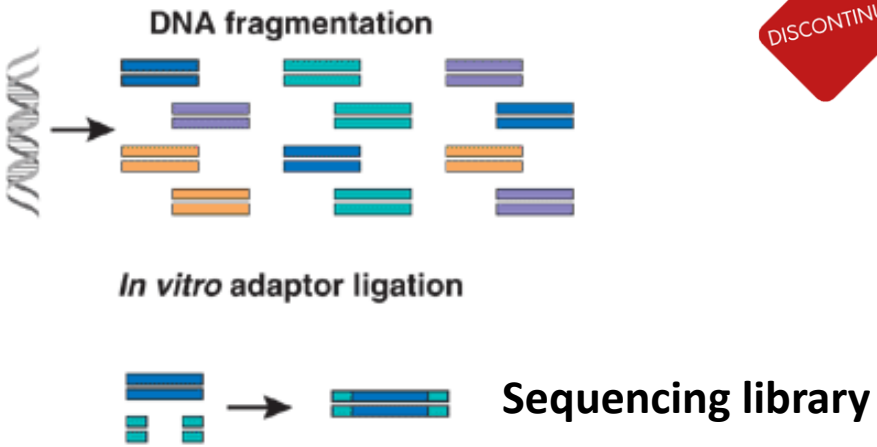
MASSIVELY PARALLEL SEQUENCING



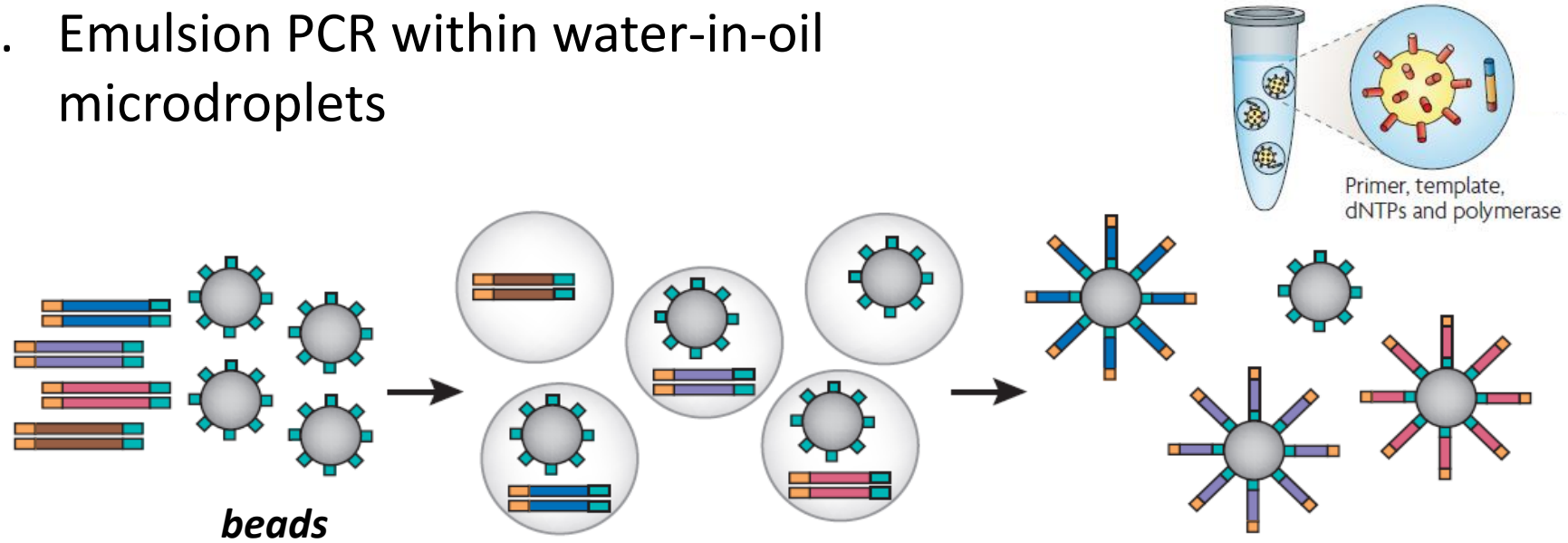
454/Roche – Pirosequencing

DISCONTINUED

1. DNA fragmentation and adapter ligation



2. Emulsion PCR within water-in-oil microdroplets



Figures 1 and 2. Shendure and Ji (2008) *Nature Biotechnology* 26: 1135-1145.

454/Roche – Pirosequencing

DISCONTINUED

3. Distribution in individual wells

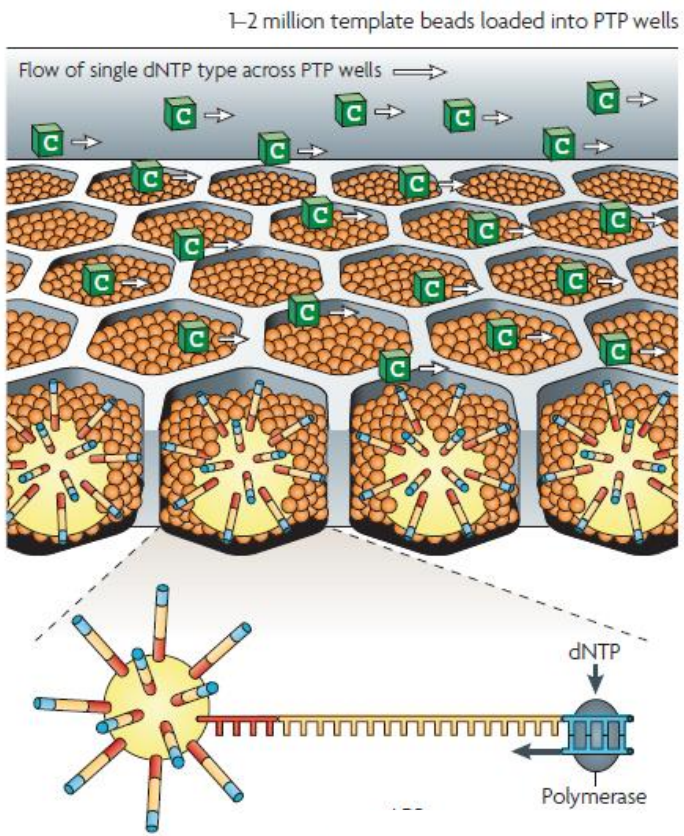
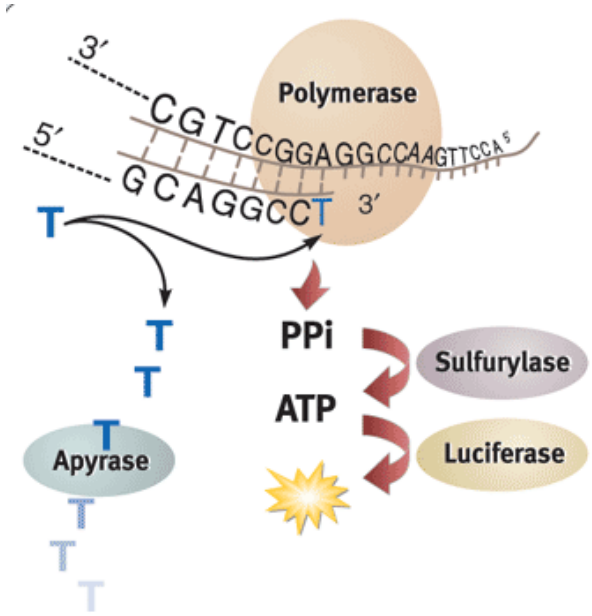


Figure 3. Metzker (2010) *Nature Reviews Genetics* 11: 31-46.

4. Pirosequencing



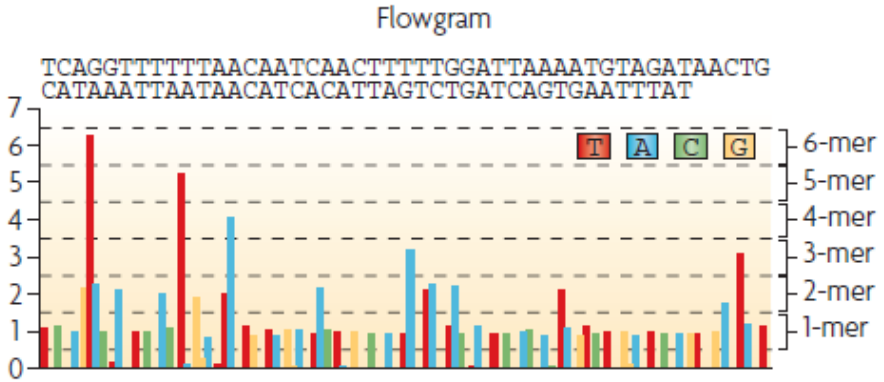
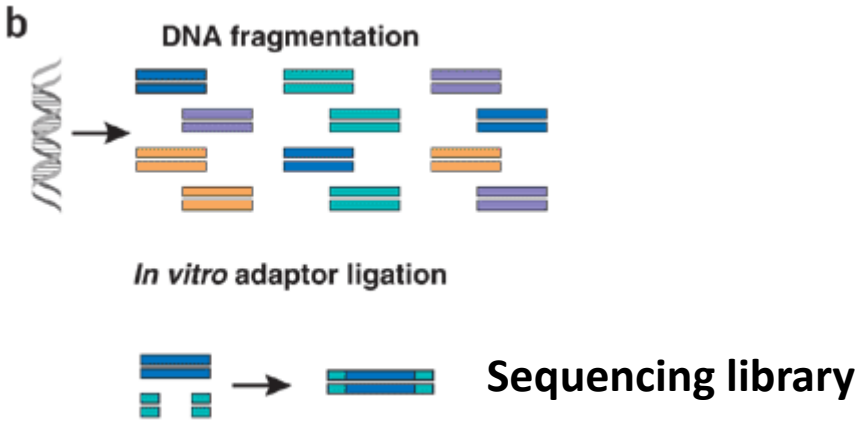


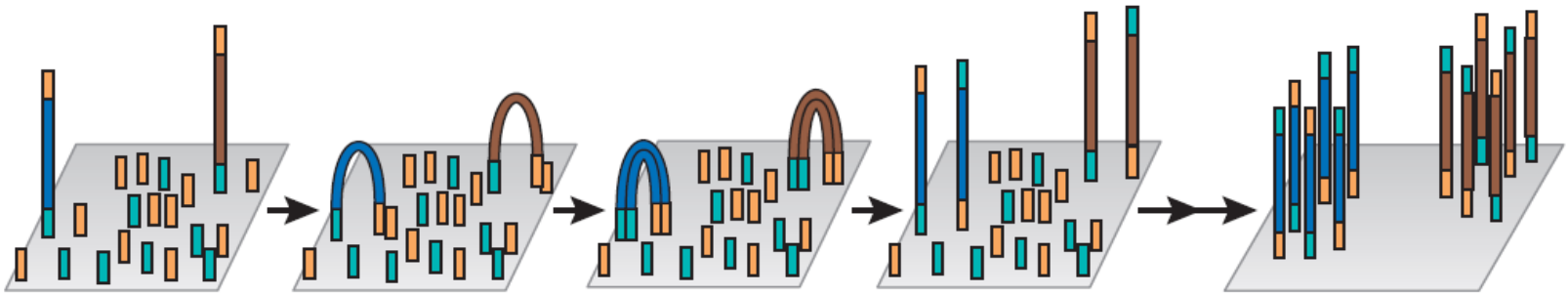
Figure 1. England and Pettersson (2005) *Nature Methods 2*: Application Note

Illumina – Reversible terminators

1. DNA fragmentation and adapter ligation

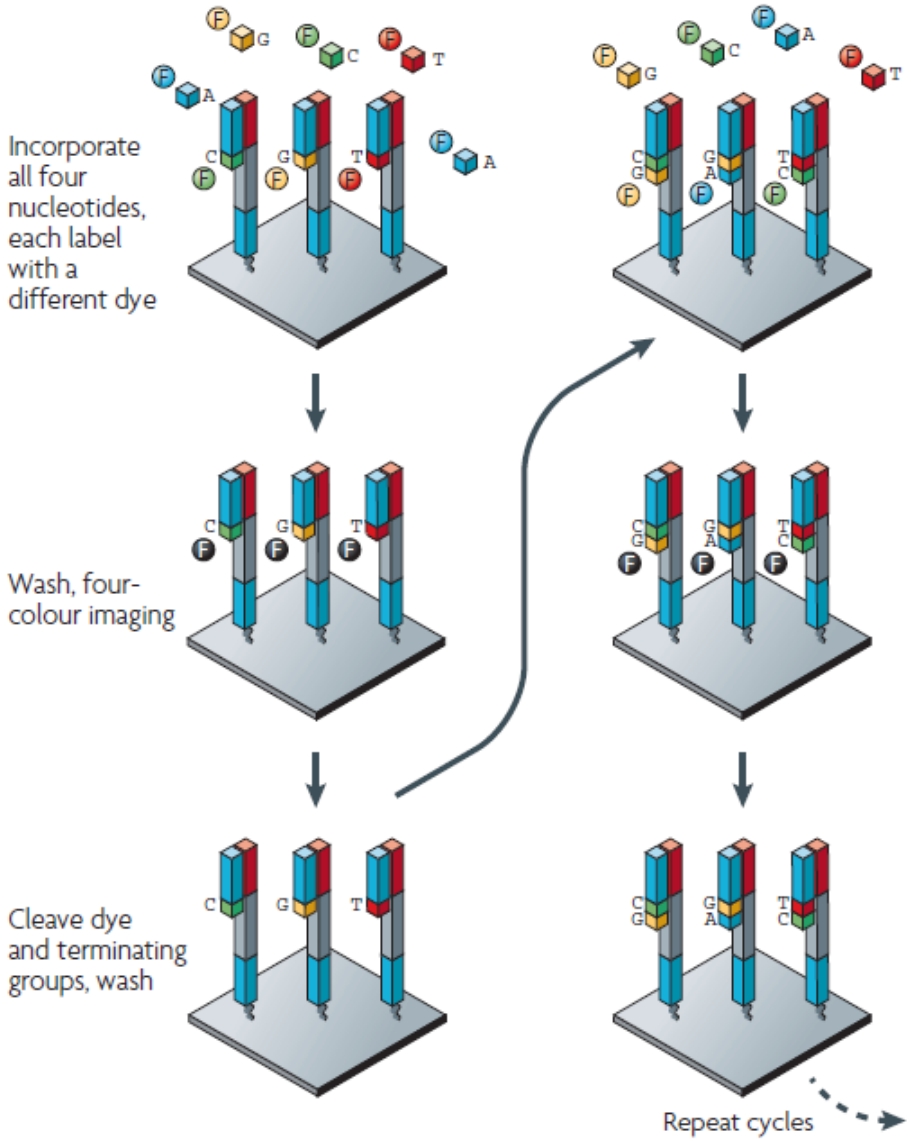


2. Solid-phase amplification and cluster generation by bridge PCR



Figures 1 and 2. Shendure and Ji (2008) *Nature Biotechnology* 26: 1135-1145.

Illumina – Reversible terminators



3. Flowing of fluorescent reversible terminator dNTPs and incorporation of a single base per cycle

4. Read identity of each base of a cluster from sequential images

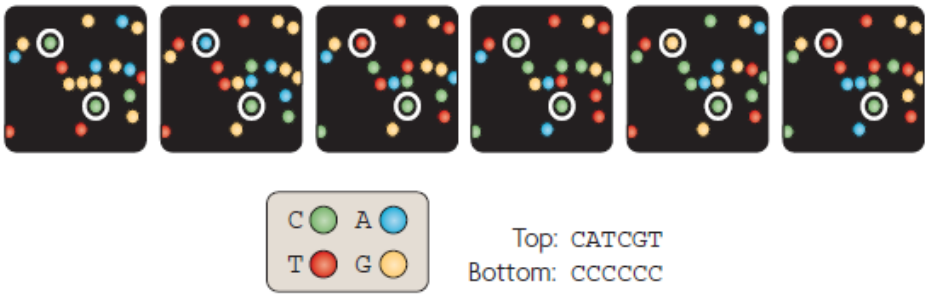
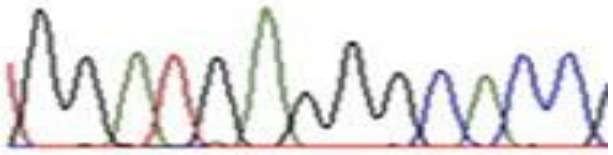


Figure 2. Metzker (2010) *Nature Reviews Genetics* 11: 31-46.

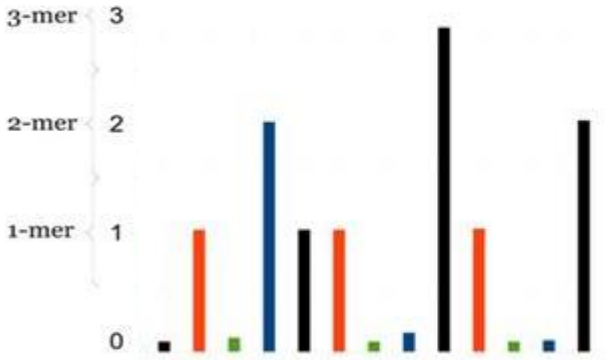
QUIZ

1



Illumina

2



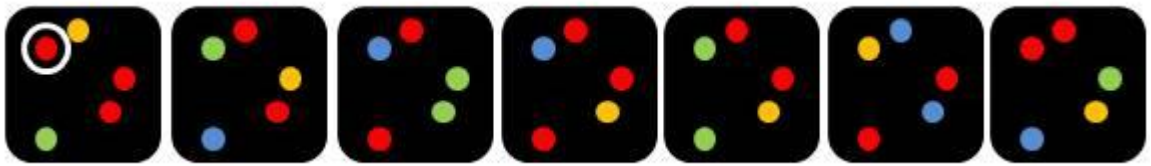
454/Roche

TCCGTGGGTGG

Sanger

GGATGAGGGCACC

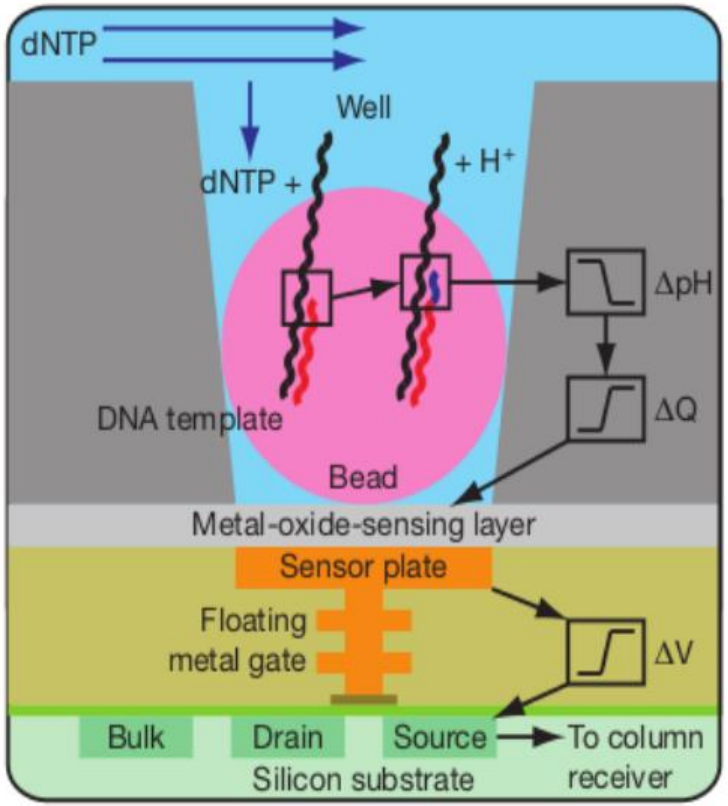
3



TACCAGT

Ion Torrent – Proton detection

- Emulsion PCR within water-in-oil microdroplets
- Real-time sequencing by using a semiconductor plate to count proton release during DNA synthesis



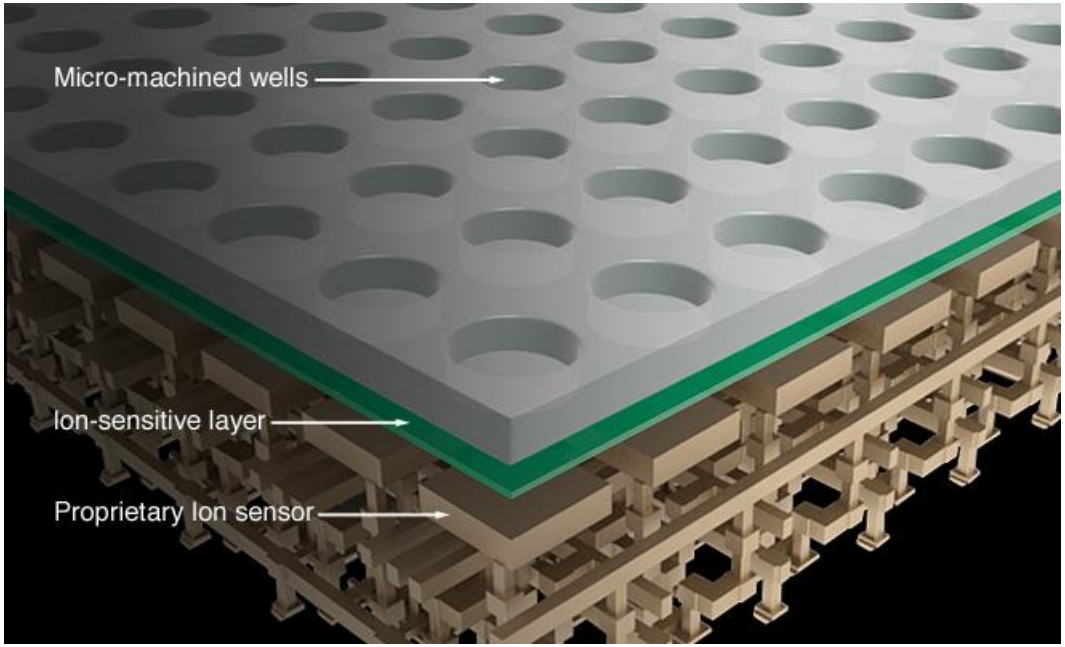
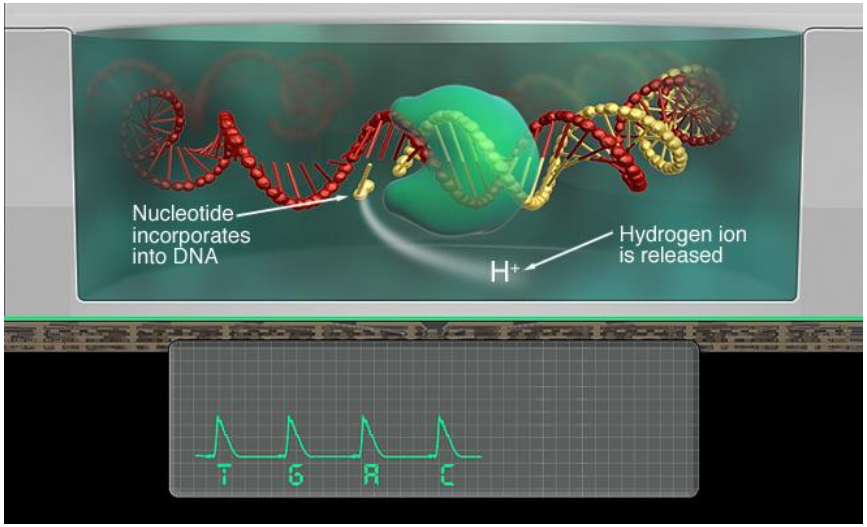
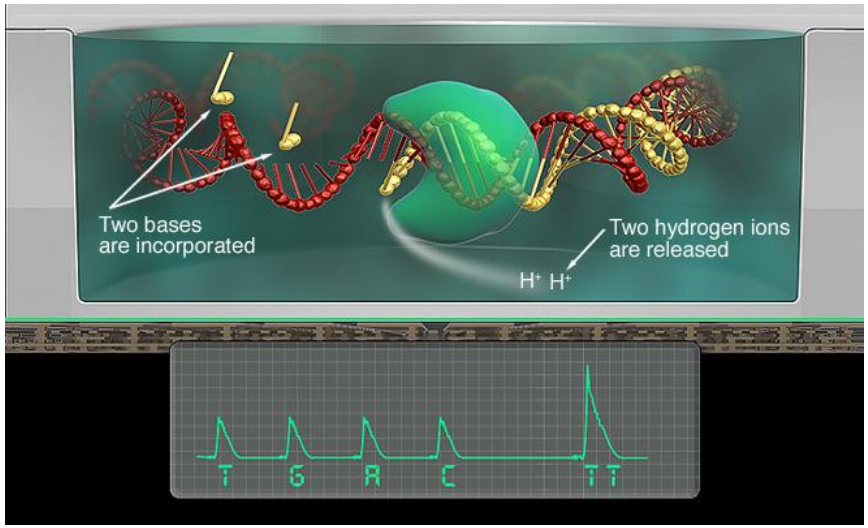


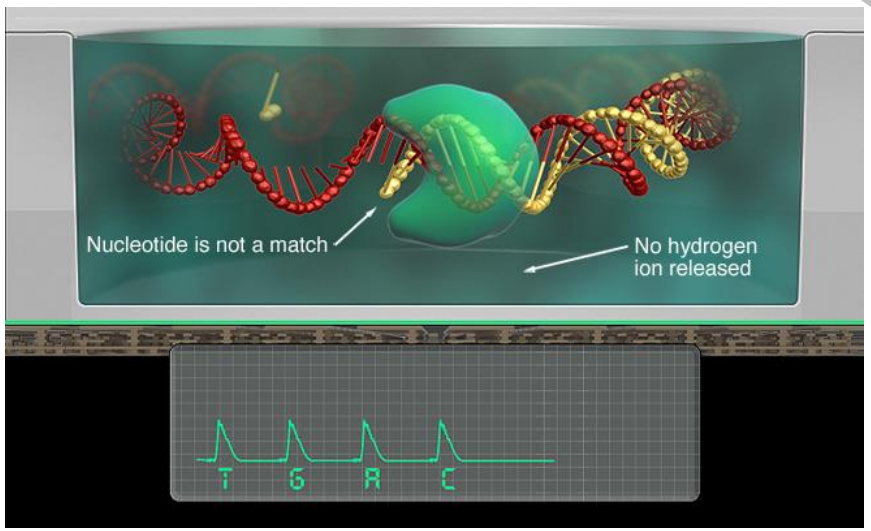
Figure 1. Rothberg *et al*. (2011) *Nature* 475:348-52.

http://www.iontorrent.com/technology-how-does-it-work-more/

Ion Torrent – Proton detection







- Normal nucleotides (not labelled) flow sequentially through the chip.
- The incorporation of one nucleotide released one proton (pH change). This is detected by the semi-conductor plate, which converts the chemical information into digital information.
- No optical machines are needed (no scanning, fluorescence, laser excitation, …)

Comparison of different sequencing methods

Table 1. Comparison of high-throughput sequencing technologies available

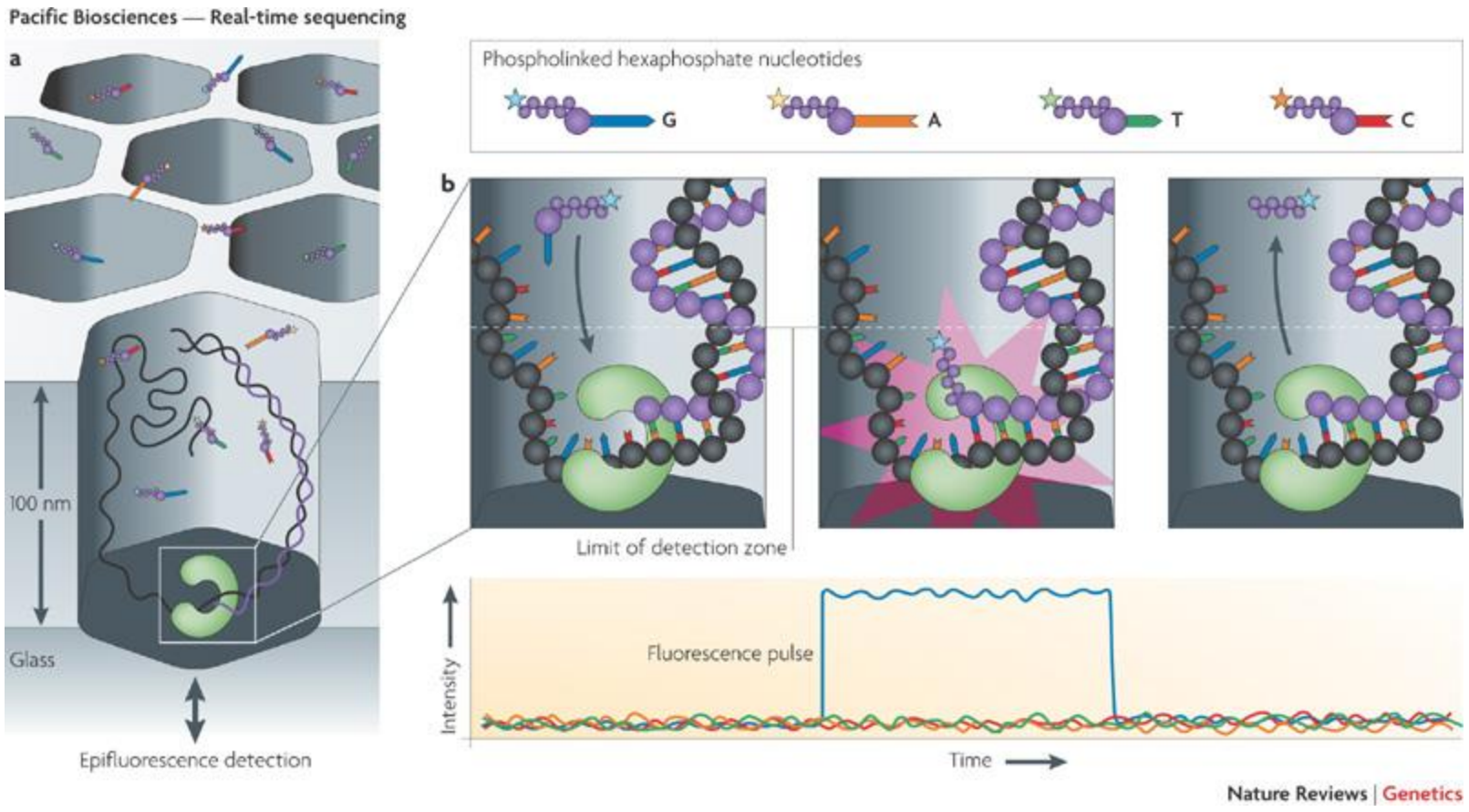
	Throughput	Length	Quality	Costs	Applications	Main sources of errors
Sanger	6 Mb/day	800 nt	10^{-4}–10^{-5}	~500$/Mb	Small sample sizes, genomes/ scaffolds, InDels/SNPs, long haplotypes, low complexity regions, etc.	Polymerase/amplification, low intensities/missing termination variants, contaminant sequences
454/Roche	750 Mb/day	400 nt	10^{-3}–10^{-4}	~20$/Mb	Complex genomes, SNPs, structural variation, indexed samples, small RNA^{+}, mRNAs^{+}, etc.	Amplification, mixed beads, intensity thresholding, homopolymers, phasing, neighbor interference
Illumina	5,000 Mb/day	100 nt	10^{-2}–10^{-3}	~0.50$/Mb	Complex genomes, counting (SAGE, CNV ChIP, small RNA), mRNAs, InDels/homopolymers, structural variation, bisulfite data, indexing, SNPs^{+}, etc.	Amplification, mixed clusters/neighbor interference, phasing, base labeling
SOLiD	5,000 Mb/day	50 nt	10^{-2}–10^{-3}	~0.50$/Mb	Complex small genomes, counting (SAGE, ChIP, small RNA, CNV), SNPs, mRNAs, structural variation, indexing, etc.	Amplification, mixed beads, phasing, signal decline, neighbor interference
Helicos	5,000 Mb/day	32 nt	10^{-2}	<0.50$/Mb	Non-amplifiable samples, counting (SAGE, ChIP, small RNA), etc.	Polymerase, low intensities/thresholding, molecule loss/termination

Figure 1. Kircher and Kelso (2010) *Bioessays* 32: 524-536.

Challenges of NGS methods

- Increase read length
- Improve sequence accuracy
- Single-molecule sequencing (no amplification)
- De-novo assembly of complex genomes
- Sequencing of complex regions

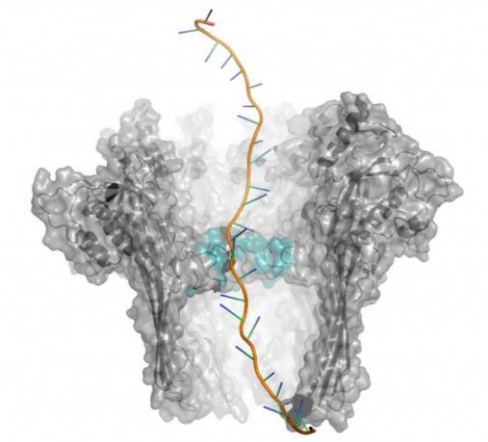
Pacific Biosciences – Real-time sequencing

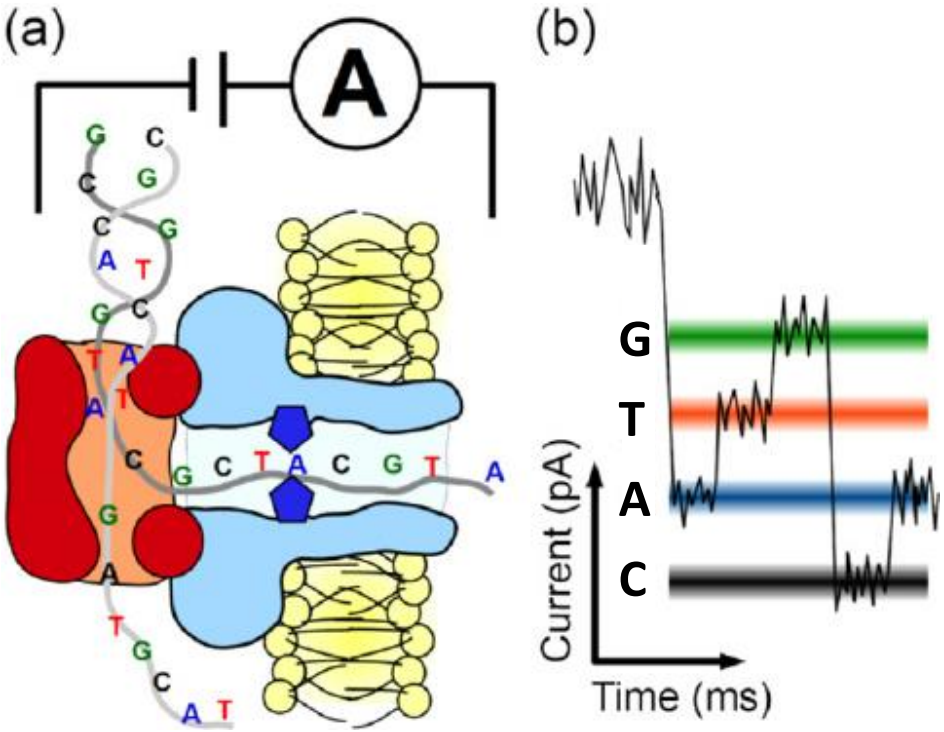


- **Real-time** monitoring of **single-molecule** sequencing as it occurs with an immobilized DNA polymerase (*Single Molecule Real Time* –SMRT-technology)
- Read length avg. ~**10-15 kb**, up to ~**50 kb**, but error rate is ~**15%**

Figure 4. Metzker (2010) *Nature Reviews Genetics* 11: 31-46.

Oxford Nanopore technology





Base identification through differences in conductance of nucleic acid molecules driven through a nanopore

- Single-molecule sequencing (no amplification needed)
- Very long read lengths
- High throughput
- Electrical base detection (no optics)

MinION Mobile Sequencing

Comparison of different NGS platforms