# Evolution of DNA sequencing technologies

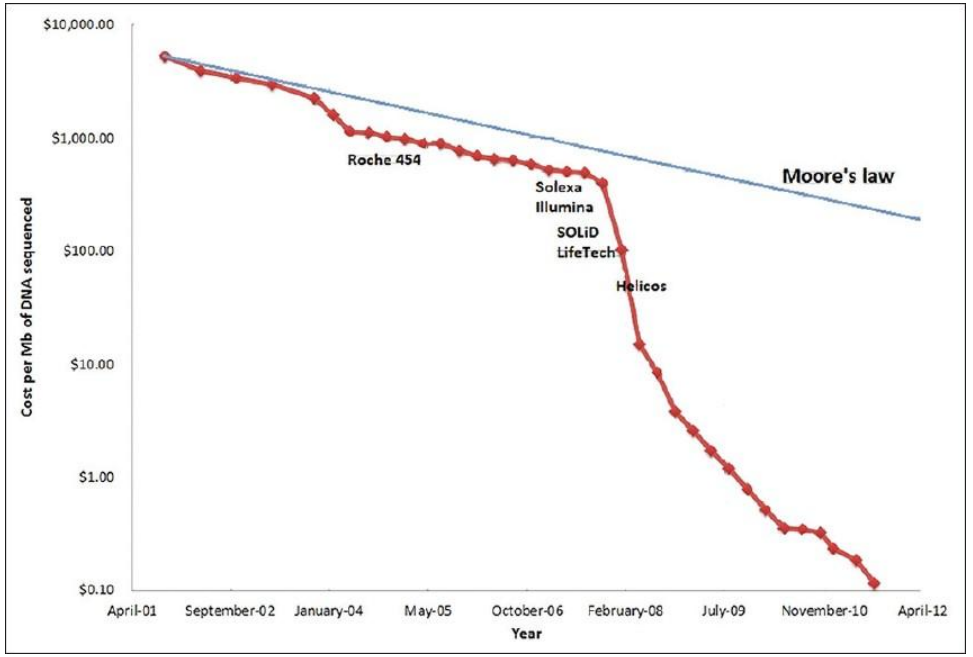
Andrew Tolonen

Genoscope-CEA et l'Université d'Évry

atolonen “at” genoscope.cns.fr
@andrew_tolonen
http://www.tolonenlab.org

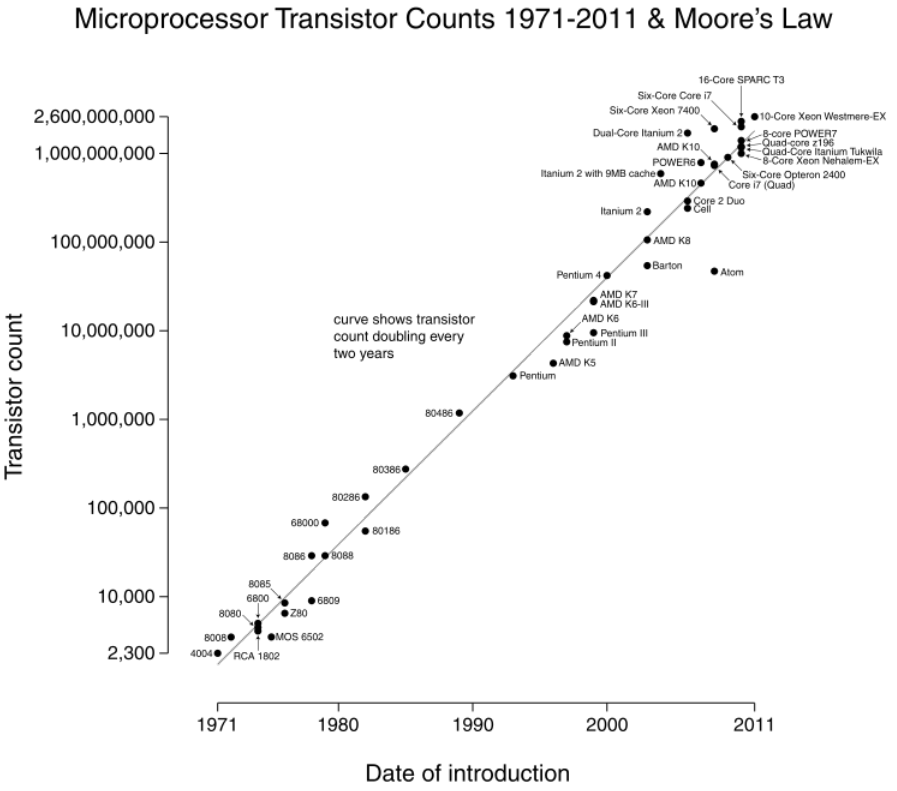
# The DNA sequencing revolution: sequencing cost has fallen faster than Moore's law

Cost per megabase of DNA sequenced




Gullapalli et al, 2012 (PMID 23248761)

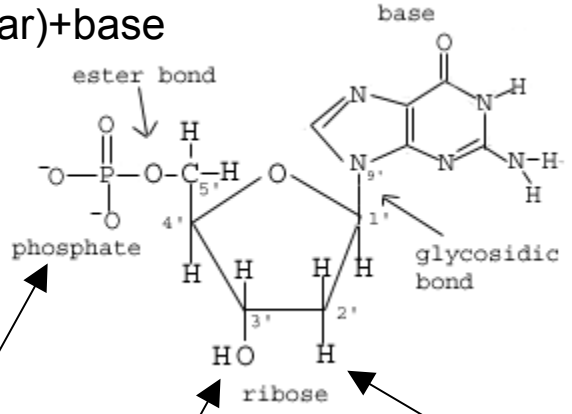
**Moore's law** describes increases in computing speed: the number of transistors in an integrated circuit doubles every 2 years.




http://en.wikipedia.org/wiki/Moore's_law

# DNA structure and polymerization

DNA nucleotide:phosphate+deoxyribose (sugar)+base



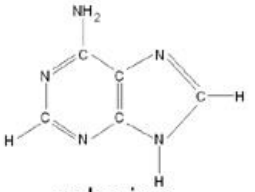

5' phosphate

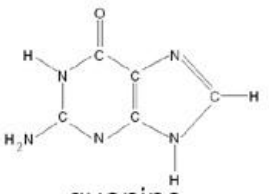
3' hydroxyl

2' deoxy (no hydroxyl group)

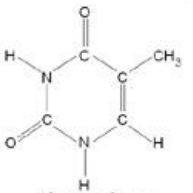
4 bases in DNA

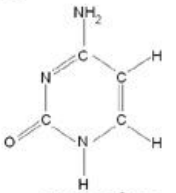
**Purines**


adenine

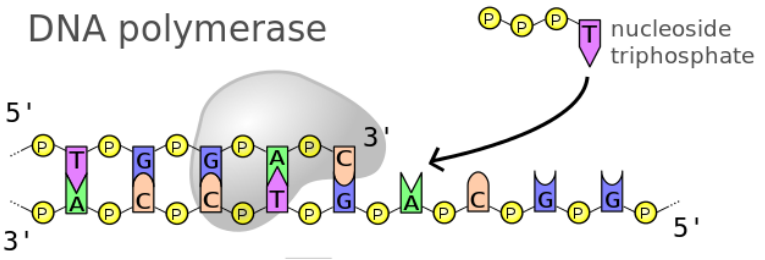

guanine

**Pyrimidines**


thymine

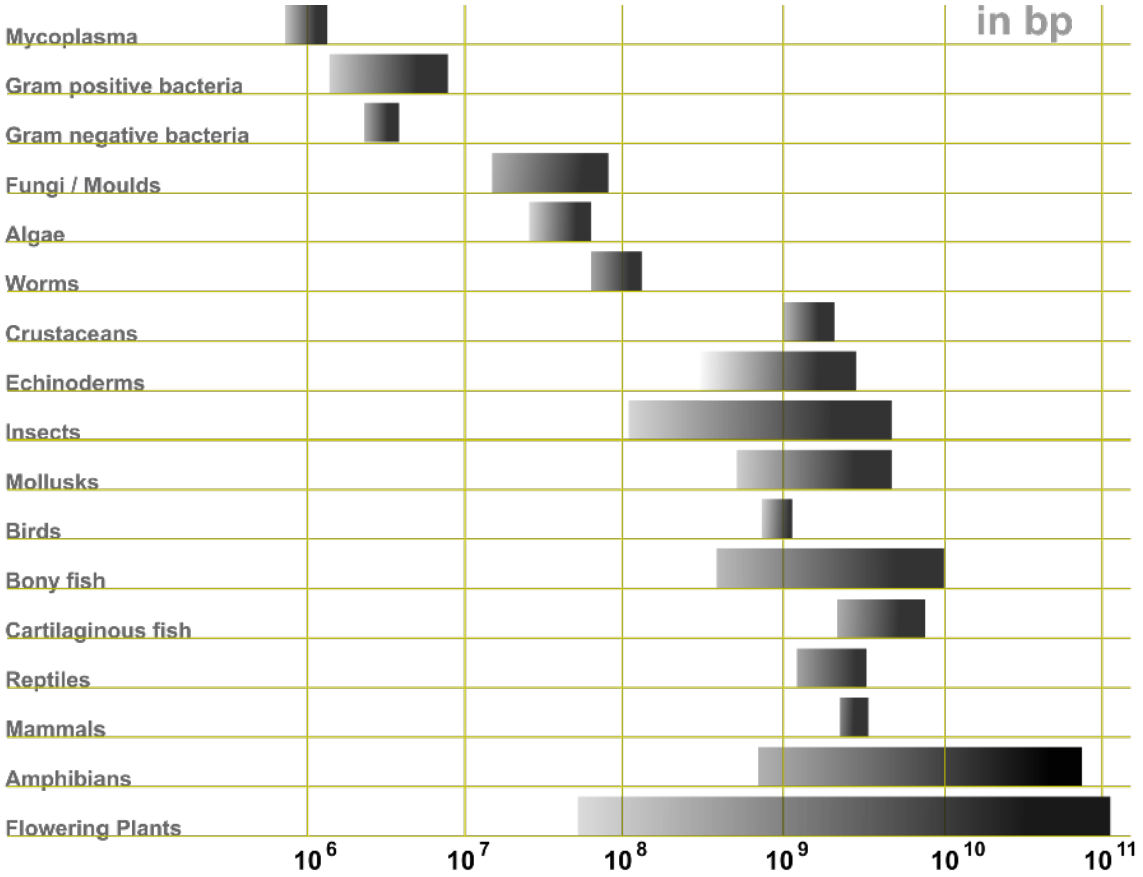

cytosine

DNA polymerase is an enzymes that replicates DNA in the 5' to 3' direction by forming a bond between the 5' triphosphate of the free nucleotide (dNTP) and the 3' OH of the DNA strand.




nucleotide incorporate liberates PPi and H+

# genome sizes vary among organisms




http://en.wikipedia.org/wiki/Genome_size

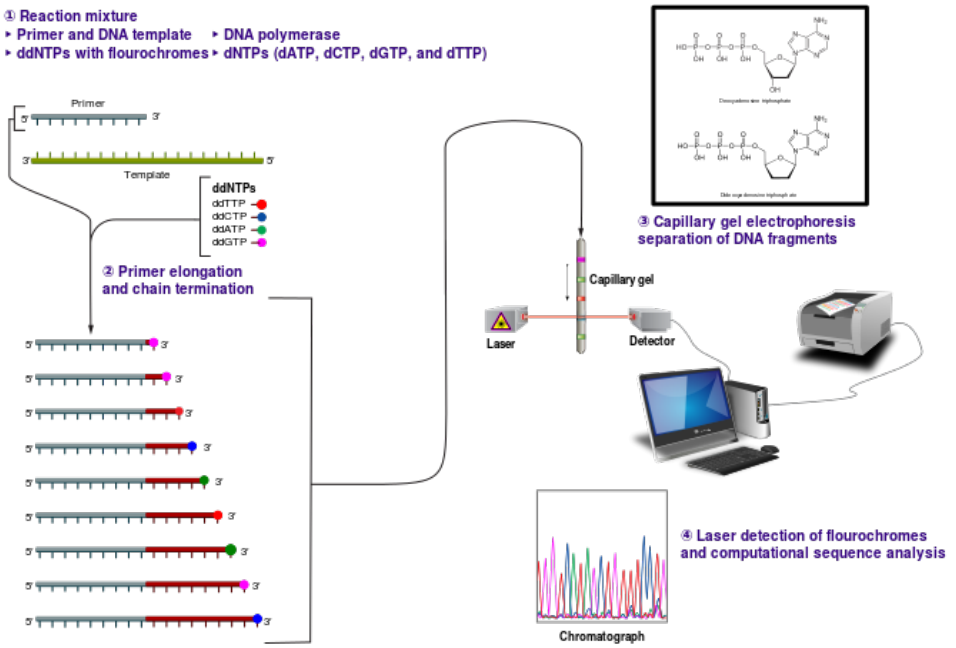
bacteria=a few million base pairs
humans=a few billion base pairs
plants=can be hundreds of billions of base pairs

# Sanger sequencing (1977)

**Reagents**
-DNA template
-ssDNA primer
-DNA polymerase
-dNTPs (lack 2' OH)
-ddNTPs, fluorescent or radioactive
(lack 2' and 3' OH -> stop elongation)

**Procedure**
1 Anneal primer to single-stranded DNA
2 DNA polymerase adds dNTP 5'->3'
3 if polymerase adds a ddNTP, elongation stops
4 size separate DNA molecules on gel
5 visualize gel to determine DNA sequence

http://en.wikipedia.org/wiki/Sanger_sequencing



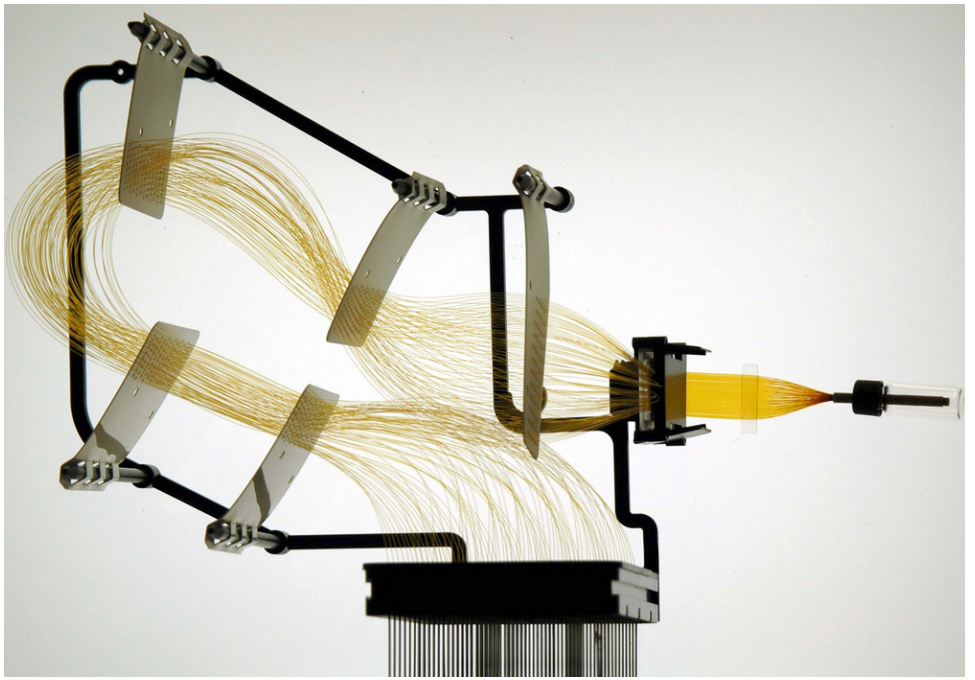

Tube contains a mix of DNA molecules, each with a different length

DNA of each length shown as chromatogram

# ABI 3730: workhorse of the human genome project

The ABI 3730 does Sanger sequencing of 96 reactions in parallel using capillaries

http://museum.mit.edu



# Sanger sequencing

Two 3730 sequencers at Genoscope: reduced from 19 machines.

Advantages

-accurate and long sequences (800-1000 bp).

-good when only a few sequences are needed (i.e. confirming inserts for cloning).

Disadvantages

-low throughput (single sequences).

-expensive on a bp basis

# 454 Sequencing (2005)
## (pyrosequencing)

**Procedure**

1 fragment DNA to 400-600 bp

2 ligate adapters to attach to bead

3 attach ssDNA to micron-sized agarose beads

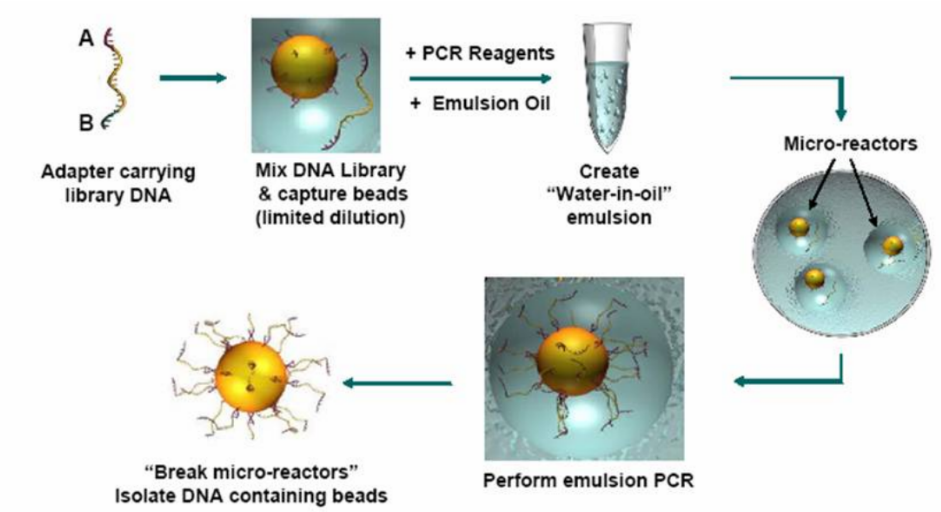
4 isolate single beads in oil droplets

5 amplify DNA on bead by emulsion-based PCR (emPCR)

6 transfer beads to pico-titer plates
(1.6 million wells with 44 μm diameter
Contain 75 picoliters)

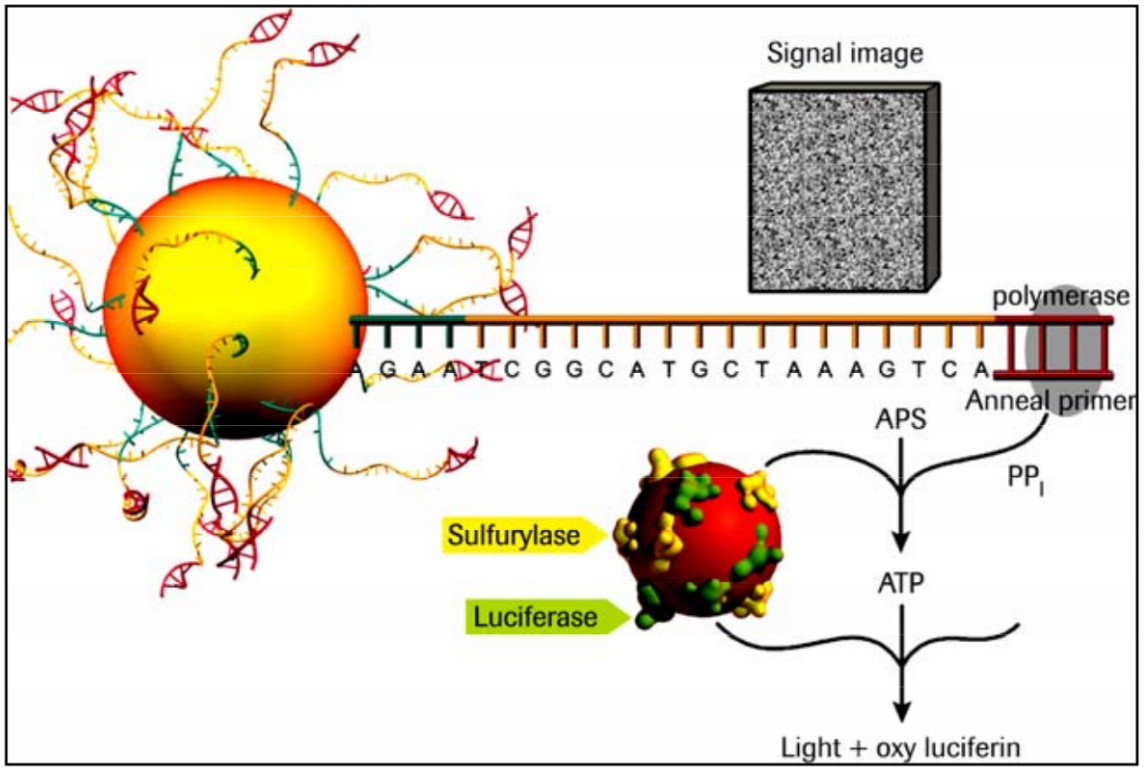
7 454 sequencing in pico-titer plate

Emulsion PCR on individual beads




454.com

# 454 sequencing chemistry




454.com

## Sequencing procedure

1 add 1 type of dNTP to well (ie dATP)

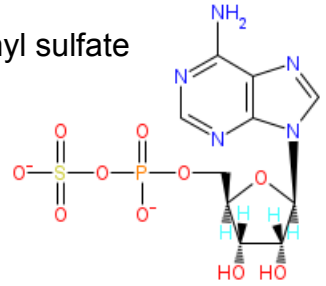
2 if polymerase incorporates nucleotide, PPi released.

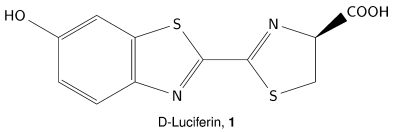
3 Sulfurylase reaction produces ATP
PPi + adenyl sulfate → ATP + sulfate

4 Luciferase reaction produces light
Luciferin + ATP → oxyluciferin + AMP + light

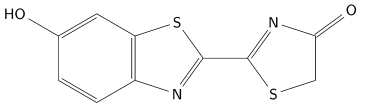
5 CCD camera measures light emission in each well (light only emitted if dATP was incorporated.

Adenyl sulfate





D-Luciferin, **1**



Oxyluciferin, **2**

# 454 sequencing

Three 454 sequencers at Genoscope: Machines are no longer sold in France. Reagents are available until 2016.

Advantages

-long sequences (800 bp).

-Higher throughput than Sanger: 1 million reads per run, run takes 1 day= 1Gbp per day.

Disadvantages

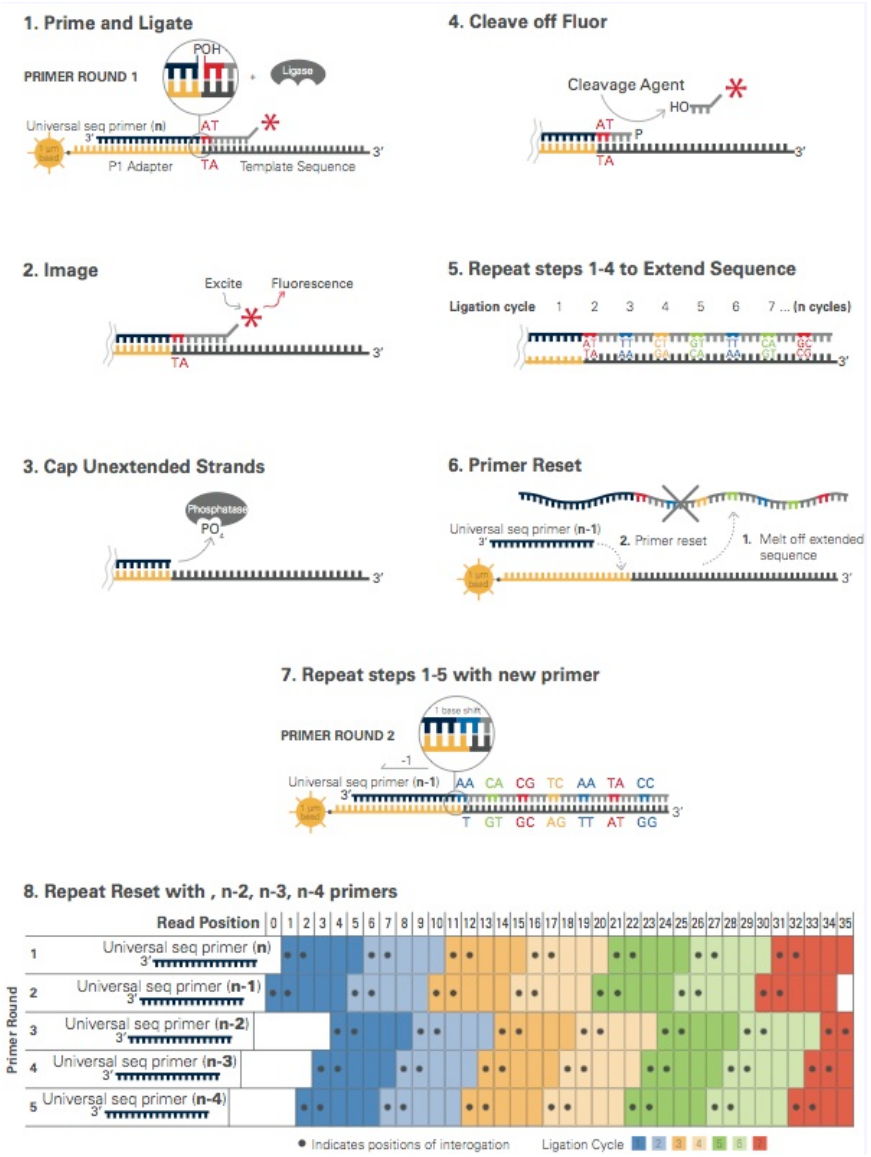
-Sample preparation is difficult (esp. em PCR) and takes at least 4 days.

-problems reading homopolymers (i.e A-A-A-A-A doesn’t produce 5x light).

-'large' wells and multiple enzymes make it more expensive than other high-throughput sequencing methods

# ABI SOLiD sequencing (2004)

(Sequencing by Oligonucleotide Ligation and Detection)




http://seqanswers.com/forums/showthread.php?t=10

## Procedure

1 fragment DNA and ligate adapters.

2 attach single DNA molecules to agarose beads.

3 isolate beads in oil emulsion.

4 amplify DNA on beads by PCR.

5 covalently attach beads to glass slide.

6 Anneal primer, hybridize, ligate a mixture of fluorescent oligos (8-mers) whose 1st & 2nd 3' bases match template.

7 image fluorescence, cleave fluor.

8 repeat step 6 to extend sequencing.

9 repeat steps 6-8 with n-1 primer.

# SOLiD sequencing

Video of SOLiD sequencing: https://www.youtube.com/watch?v=nlvyF8bFDwM

no SOLiD sequencers Genoscope: there were 2 machines.

Advantages

-Higher throughput than 454.

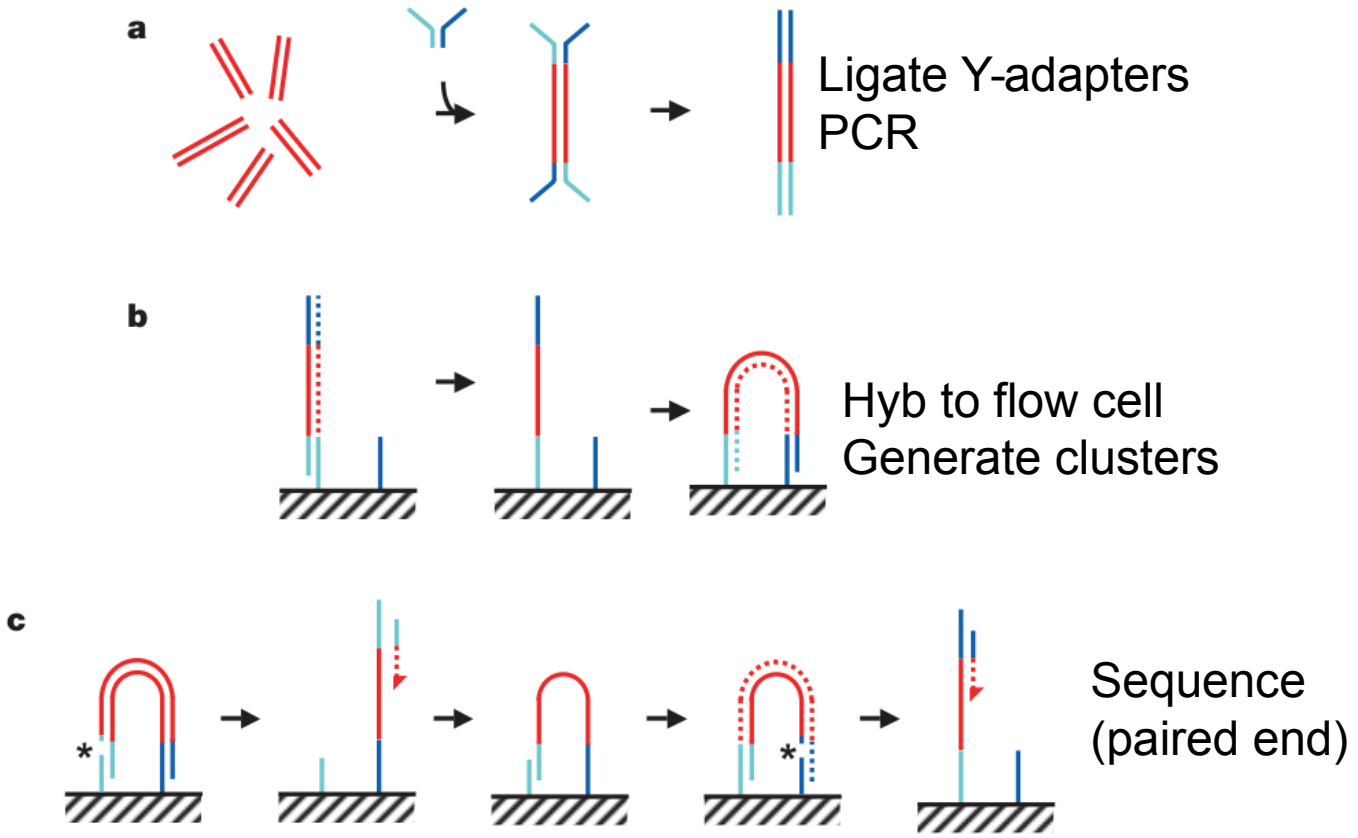
Disadvantages

-Sample prep (emPCR) is difficult.

-Read length limited to 35 bp.

# Illumina sequencing

Video of Illumina sequencing:
https://www.youtube.com/watch?v=womKfikWlxM




Bentley et al, 2008

**Procedure**

**A** shear DNA ~500 bp and Ligate adapters

**B** hybridize DNA to flow cell surface using adapters. Amplify clusters by bridge PCR

**C** linearize DNA, hybridize DNA sequencing primer

**D** sequence by adding all four reversible dye terminators and imaging at each cycle

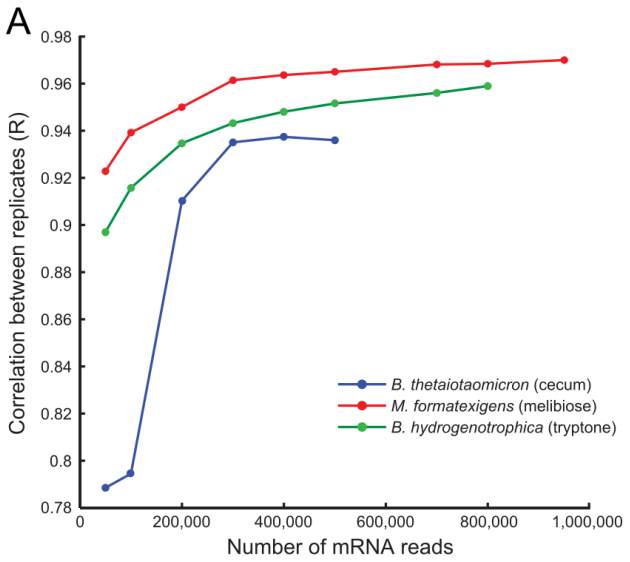
Reversible dye terminators are the key to this method of sequencing by synthesis

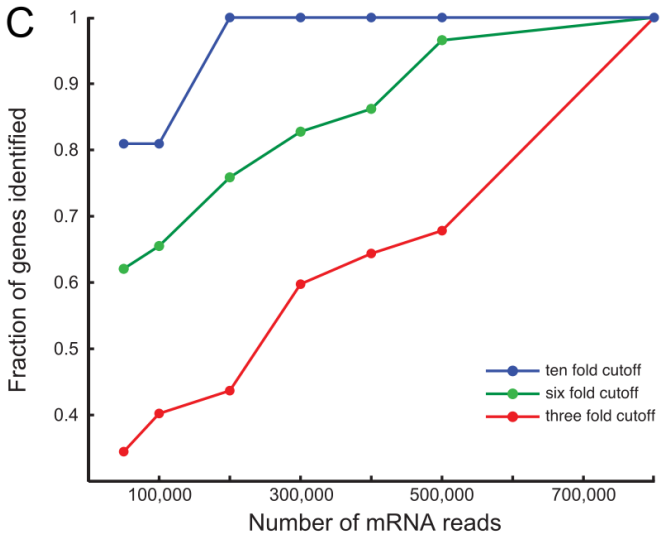
# Sequence capabilities of different Illumina machines

| | Read Length | Run time | Output range | Expected Reads |
|---|---|---|---|---|
| MiSeq | 2 x 25 bp | ~ 5.5 hrs | 0,8 Gb | 15 M |
| | 2 x 150 bp | ~ 24 hrs | 5 Gb | 15 M |
| | 2 x 250 bp | ~ 39 hrs | 8 Gb | 15 M |
| | 2 x 300 bp | ~ 65 hrs | 15 Gb | 25 M |
| HiSeq 2000 High-Output | 2 x 100 bp | ~ 11 days | 300 Gb (x 2 FC) | 1 500 M |
| HiSeq 2500 High-Output | 2 x 125 bp | ~ 6 days | 500 Gb (x 2 FC) | 2 000 M |
| HiSeq 2500 Rapid Run | 2 x 100 bp | ~ 27 hrs | 60 Gb (x 2 FC) | 300 M |
| | 2 x 150 bp | ~ 40 hrs | 90 Gb (x 2 FC) | 300 M |
| | *2 x 250 bp* | *~ 90 hrs* | *150 Gb (x 2 FC)* | *300 M* |
| HiSeq Xten | 2 x 150 bp | < 3 days | 900 Gb (x 2 FC) | 3 000 M |

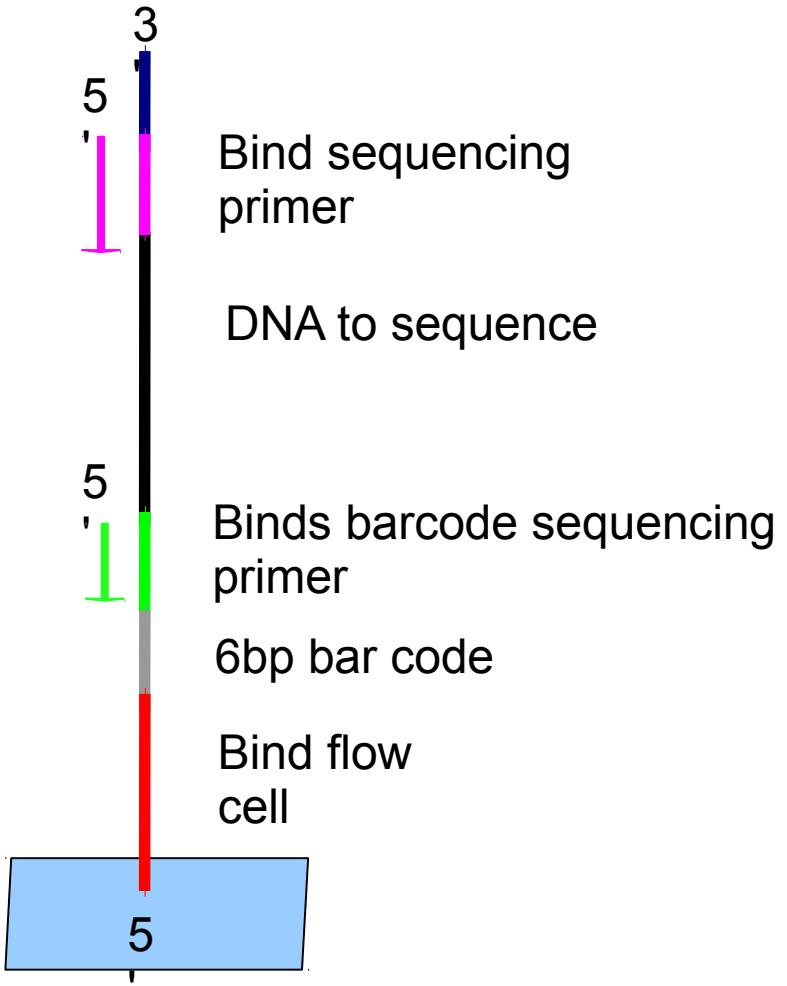
source Karine Labadie, Genoscope

# multiple samples can be sequenced in the same reaction using bar codes in the adapters




Rey et al, 2010



# Illumina sequencing

Machines at Genoscope: 4 hiseq 2000, 2 hiseq 2500, 2 miseq

Advantages

-high-throughput: Hiseq has 2 flow cells, each flow cell has 8 lanes, 250 million reads per lane. 16 X 250 million reads= 4 billions reads par run. Read is 150 bp=$6x10^{11}$ bp per run.

-384 multiplexing of samples.

Disadvantages

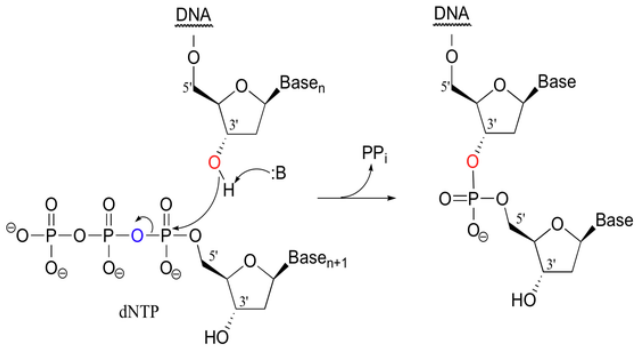
-shorter sequences (now up to 300 bp on Miseq).

-reversible terminators are expensive.

-a Hiseq run takes 1 week (scanning of flow cell is slow).

# Ion torrent sequencing (2010)
## (ion semiconductor sequencing)

DNA polymerization releases PPI and H+




### Procedure

1 many copies of specific DNA Template are added to microwells in semiconductor chip.

2 DNA polymerase and a single Type of dNTP (A, C, G, or T) are added to each well.

3 If dNTP is incorporated, H+ released and the pH is reduced.

4 ISFET sensor (ion sensitive Transistor) detects pH drop.

5 unincorporated dNTPs washed away before next cycle

# Ion torrent sequencing

Ion Torrent sequencing:
https://www.youtube.com/watch?v=WYBzbxIfuKs

No Ion Torrent sequencers at Genoscope: 1 machine at CNG

Advantages

-no modified nucleotides, special enzymes, or optics required.

-fast: sequencing occurs in real time (15 seconds per cycle, run takes 1h), which is much faster than Illumina.

Disadvantages

-reads are 100 bp with $10^8$ bp per run, which is at least 1000x less bp than Illumina.

-less accurate for homopolymers (multiple H+ released).

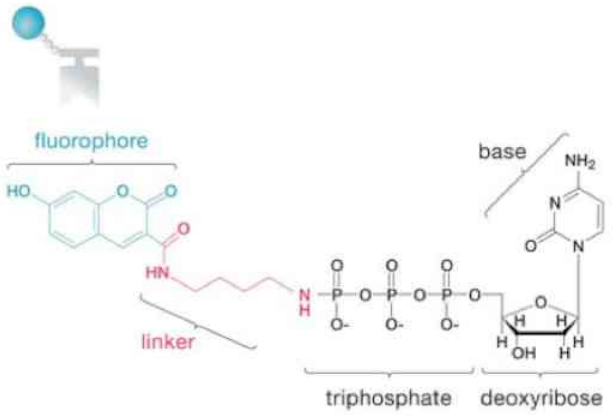
-requires emPCR for sample prep. Sample prep reagents are expensive.

# Pacific Biosciences sequencing (2004)

https://www.youtube.com/watch?v=v8p4ph2MAvI

visualize dNTP incorporation into a single molecule by DNA polymerase

**Phospholinked nucleotides**:Each type of nucleotide has a different fluorophore that is released when incorporated into DNA chain.




http://www.csulb.edu/~cohlberg/DNAsequencing.html

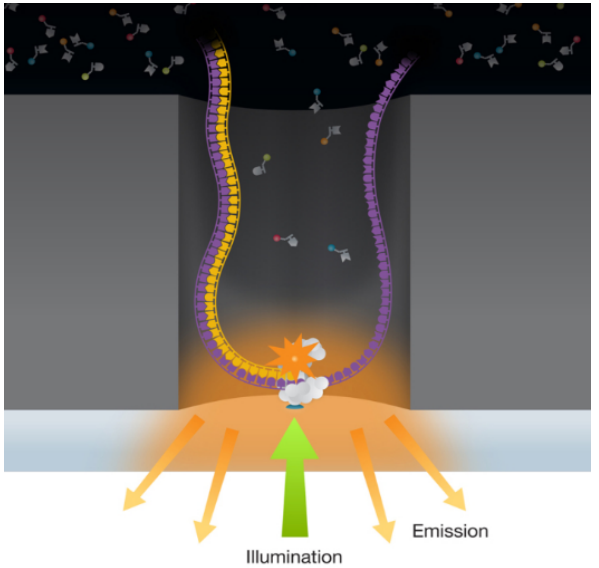
**Zero-mode wave guide**: nano-structures allow individual molecules to be isolated for optical analysis

## Procedure

1 insert single template DNA molecule and DNA polymerase into 'well' of zero-mode wave guide

2 add mix of phospholinked dNTPs

3 DNA polymerase incorporates a dNTP, fluorescence observed.




http://opfocus.org

# Pacific Biosciences sequencing

No PacBio machines at Genoscope or elsewhere in France.

Advantages
-single molecule sequencing (no PCR bias).

-rapid sequencing can be used for disease outbreaks (cholera in Haiti 2010).

-long reads: avg 5,000 bp, up to 30 kb.
Disadvantages
-high error rate (13%).

-low throughput: 60,000 reads per run = $3x10^8$ bp in 3h.

-requires 1 ug DNA for sample, so often need to amplify anyways.

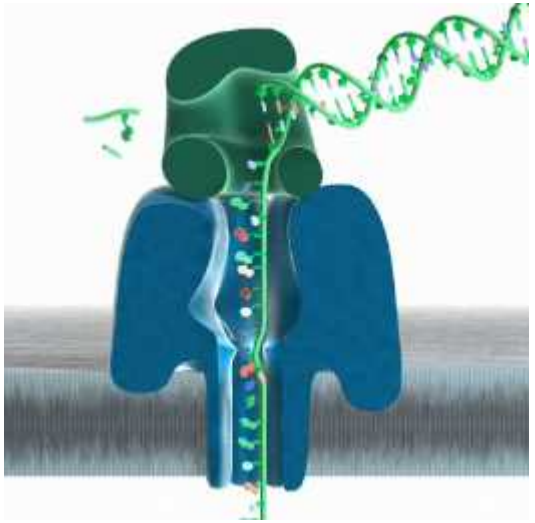
-Polymerase is photo-degraded. Need more stable polymerase.

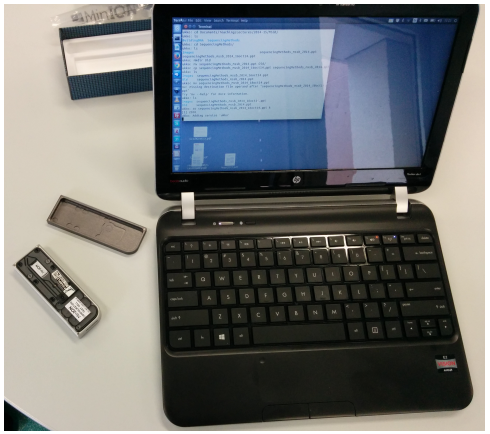
Source http://nextgenseek.com/2013/04/pacbio-launches-pacbio-rs-ii-sequencer/

# Nanopore sequencing (2014)

(strand sequencing)

DNA sequenced as it passes through a nanopore





Oxford Nanopore MinION USB key sequencer

## Procedure

1 embed protein nanopore in resistant membrane.

2 put nanopore in conducting fluid.

3 apply voltage.

4 electrical current due to ion conduction through pore.

5 passage of DNA through pore changes current.

6 current change through pore as DNA passes is sequence readout.

DNA sequenceread in in 5 bp intervals. Each 5 bp gives a different signal (1000 possibilites).

# Nanopore sequencing

Video of Oxford Nanopore sequencing:
https://www.youtube.com/watch?v=3UHw22hBpAk

## 6 Oxford Nanopore sequencers at Genoscope

### Advantages

-inexpensive: no PCR, modified nucleotides, special enzymes, or optics required.

-fast: DNA passes through pore at 20-100 bp per second.

-very long reads are possible (8-9 kb per read).

-possible to sequence other polymers (proteins).

### Disadvantages

-low throughput: 512 pores per flow cell.

-<80% of pores are active. Pores can get blocked. Forward run is better than reverse.

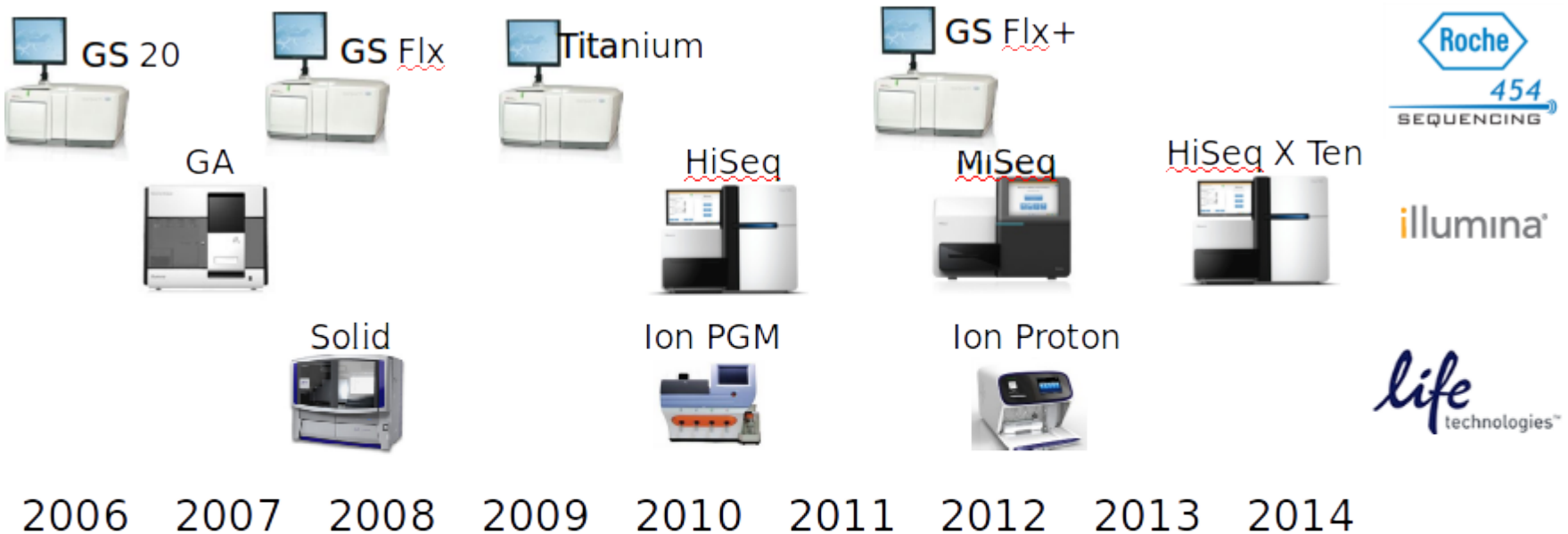
-DNA gets stuck and does not pass through pore at constant rate.

# Illumina is currently 80% of the DNA sequencing market. What could change that?

-**Ion Torrent**: eliminate emPCR? Increase chip density (more ads per run)?

-**PacBio**: reduce sample prep to enable real-time sequencing?

-**Oxford Nanopore**: Control passage of DNA through pore? Increase pores per chip?




source Karine Labadie, Genoscope