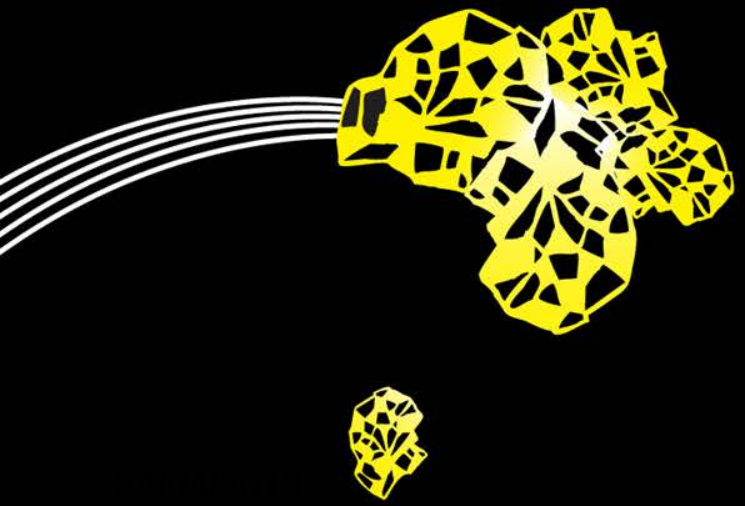
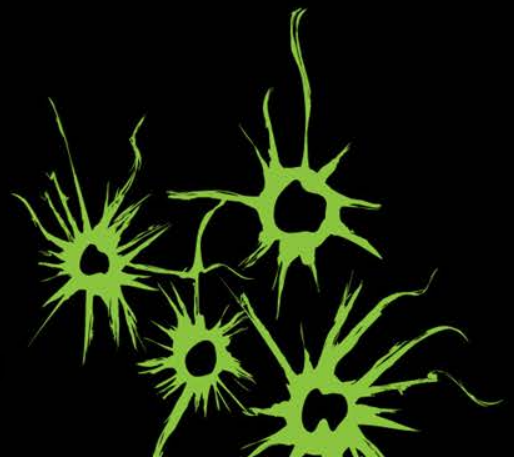
UNIVERSITY OF TWENTE.

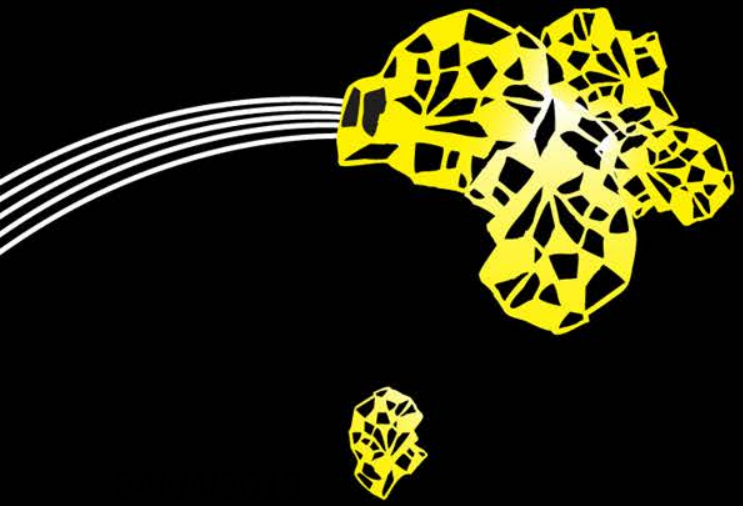


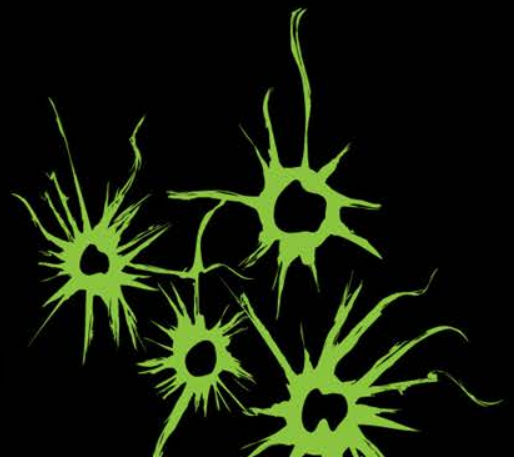


Next Gen Sequencing

Summary of the short course “Next Gen Sequencing” at Avans hogeschool, Breda

UNIVERSITY OF TWENTE.





2nd Gen Sequencing

Summary of the short course “Next Gen Sequencing” at Avans hogeschool, Breda

Generations of sequencing technologies

- 1st gen: 1977-2007 Sanger sequencen
- 2nd gen 2007-2013 (Next gen) massive parallel sequencing
- 3rd gen 2012-? Single molecule sequencing

Overview

- History
- NGS (2nd gen) platforms
  - Illumina (former: Solexa)
  - SoLID: Life technlogies (former: Applied Biosystems) (SOLID)
  - Roche (Pyrosequencing)
  - Ion Torrent (semiconductor hydrogen ion detection)
- 3rd gen platforms
  - Helicos
  - Pac bio
  - Oxford Nanopore
- Sequencing strategies
- Not covered NGS subjects

1973 First sequence published

Proc. Nat. Acad. Sci. USA
Vol. 70, No. 12, Part I, pp. 3581–3584, December 1973

The Nucleotide Sequence of the *lac* Operator

(regulation/protein–nucleic acid interaction/DNA–RNA sequen

WALTER GILBERT AND ALLAN MAXAM

Department of Biochemistry and Molecular Biology, Harvard University, Caml

Communicated by J. D. Watson, August 9, 1973

ABSTRACT The *lac* repressor protects the *lac* operator against digestion with deoxyribonuclease. The protected fragment is double-stranded and about 27 base-pairs long. We determined the sequence of RNA transcription copies of this fragment and present a sequence for 24 base pairs. It is:

5′--T G G A A T T G T G A G C G G A T A A C A A T T 3′
3′--A C C T T A A C A C T C G C C T A T T G T T A A 5′

The sequence has 2-fold symmetry regions; the two longest are separated by one turn of the DNA double helix.

1977

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463–5467, December 1977
Biochemistry

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕX174)

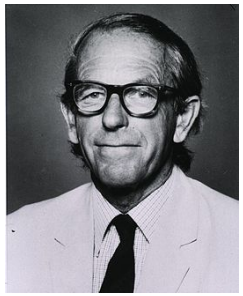
F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

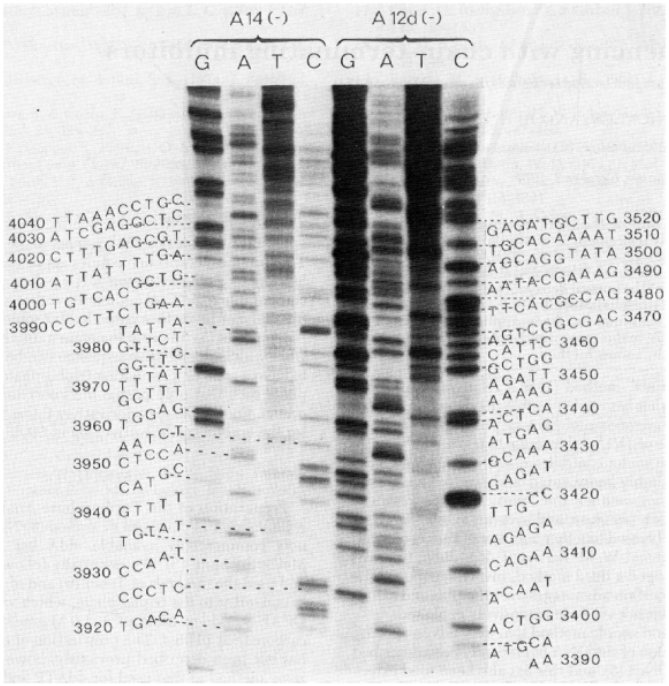
Contributed by F. Sanger, October 3, 1977

ABSTRACT **A new method for determining nucleotide sequences in DNA is described. It is similar to the "plus and minus" method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* 94, 441–448] but makes use of the 2′,3′-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage ϕX174 and is more rapid and more accurate than either the plus or the minus method.**

a stereoisomer of ribose in which the 3′-hydroxyl group is ori
ented in *trans* position with respect to the 2′-hydroxyl group
The arabinosyl (ara) nucleotides act as chain terminating in
hibitors of *Escherichia coli* DNA polymerase I in a manne
comparable to ddT (4), although synthesized chains ending i
3′ araC can be further extended by some mammalian DNA
polymerases (5). In order to obtain a suitable pattern of band
from which an extensive sequence can be read it is necessary



Frederick Sanger

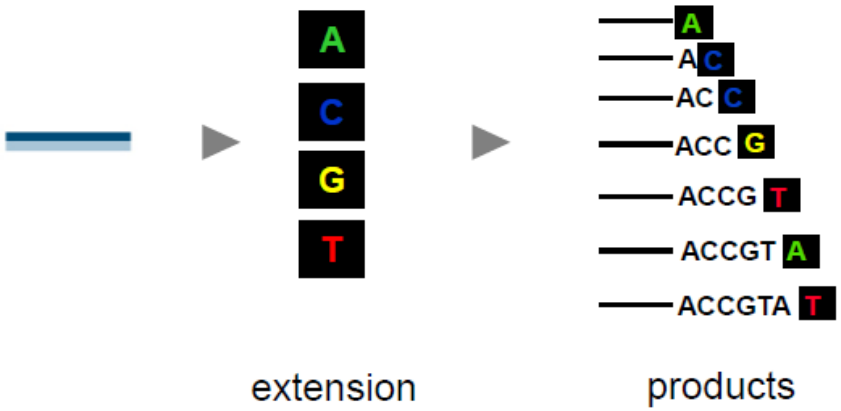


1985 Human genome project proposed

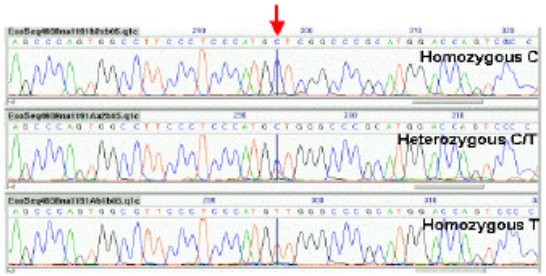
- Start of the project in 1987

1987 first automated sequencer ABI370

DNA polymerase, dNTP's and dye-labeled dideoxy-nucleotides (terminators)



extension products



1997 first “next gen” patents filed

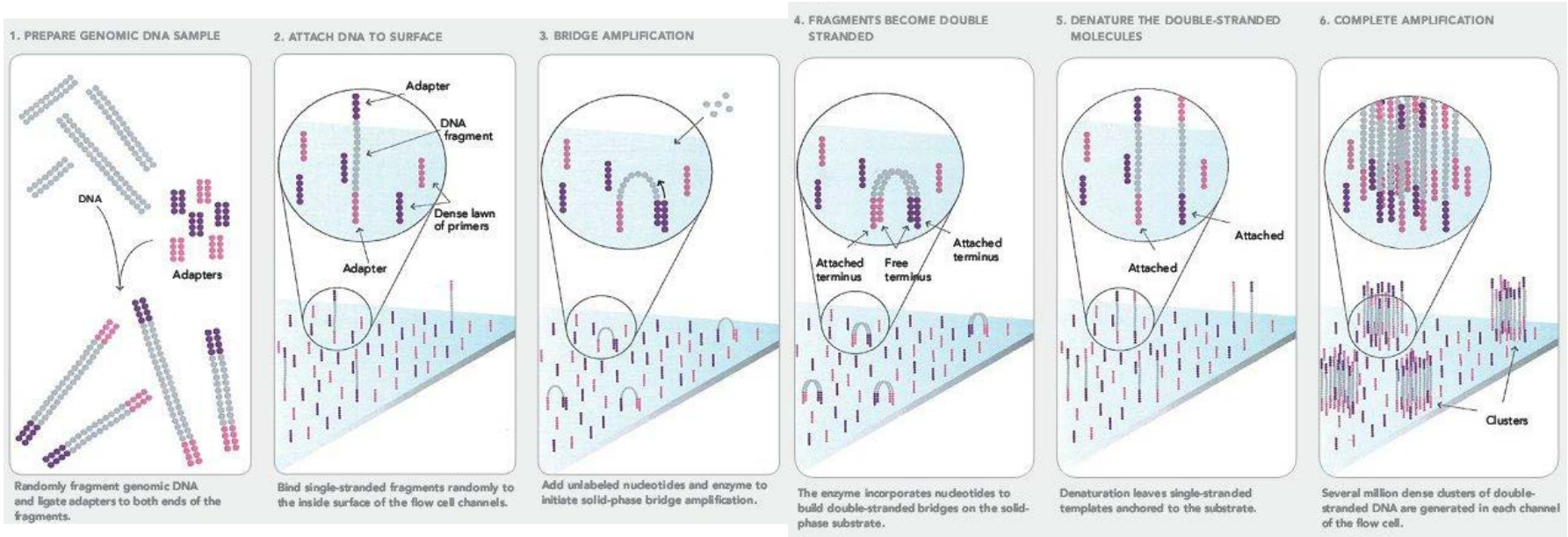
- 2003 Human genome project finished
- 2005 first next gen sequencer available
- 2012 first “2nd gen” sequencer available

Illumina

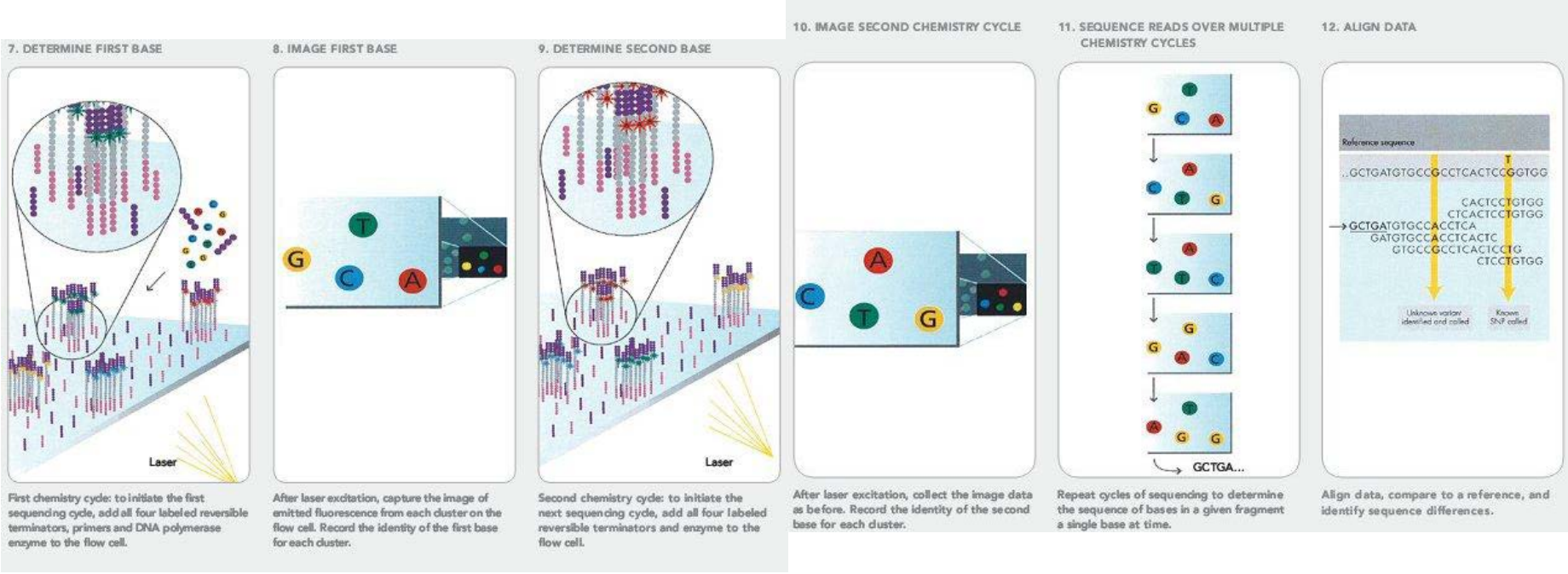
The illumina sequencer family:

- *Genome AnalyzerIIx, HiSeq 2500, HiSeq 2000, HiSeq 1500, HiSeq 1000, HiScanSQ instruments, MiSeq*
- TruSeq chemistry: sequencing by synthesis.

Illumina sample prep



Illumina sequencing

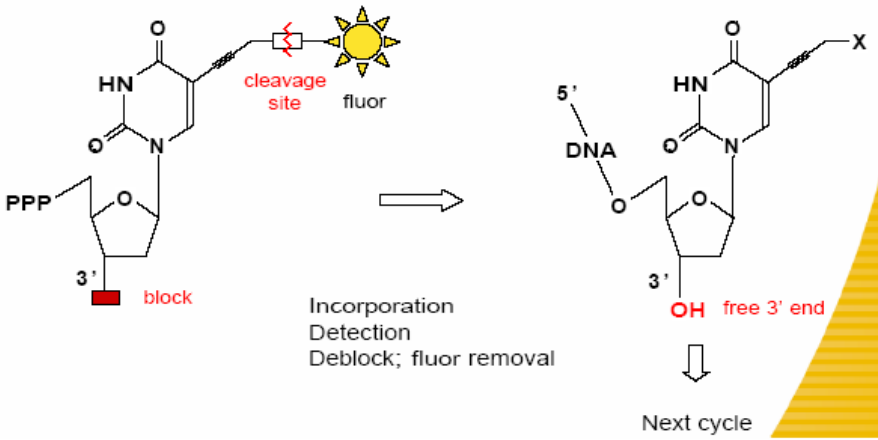


Illumina sequencing

Reversible Terminator Chemistry

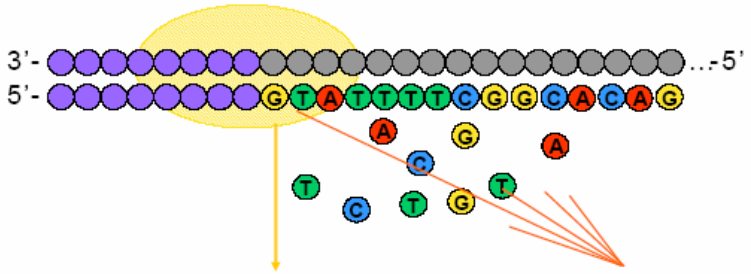
Solexa

- All 4 labelled nucleotides in 1 reaction
- Higher accuracy
- No problems with homopolymer repeats



Sequencing By Synthesis (SBS)

Solexa



Cycle 1:	Add sequencing reagents
	First base incorporated
	Remove unincorporated bases
	Detect signal
Cycle 2-n:	Add sequencing reagents and repeat

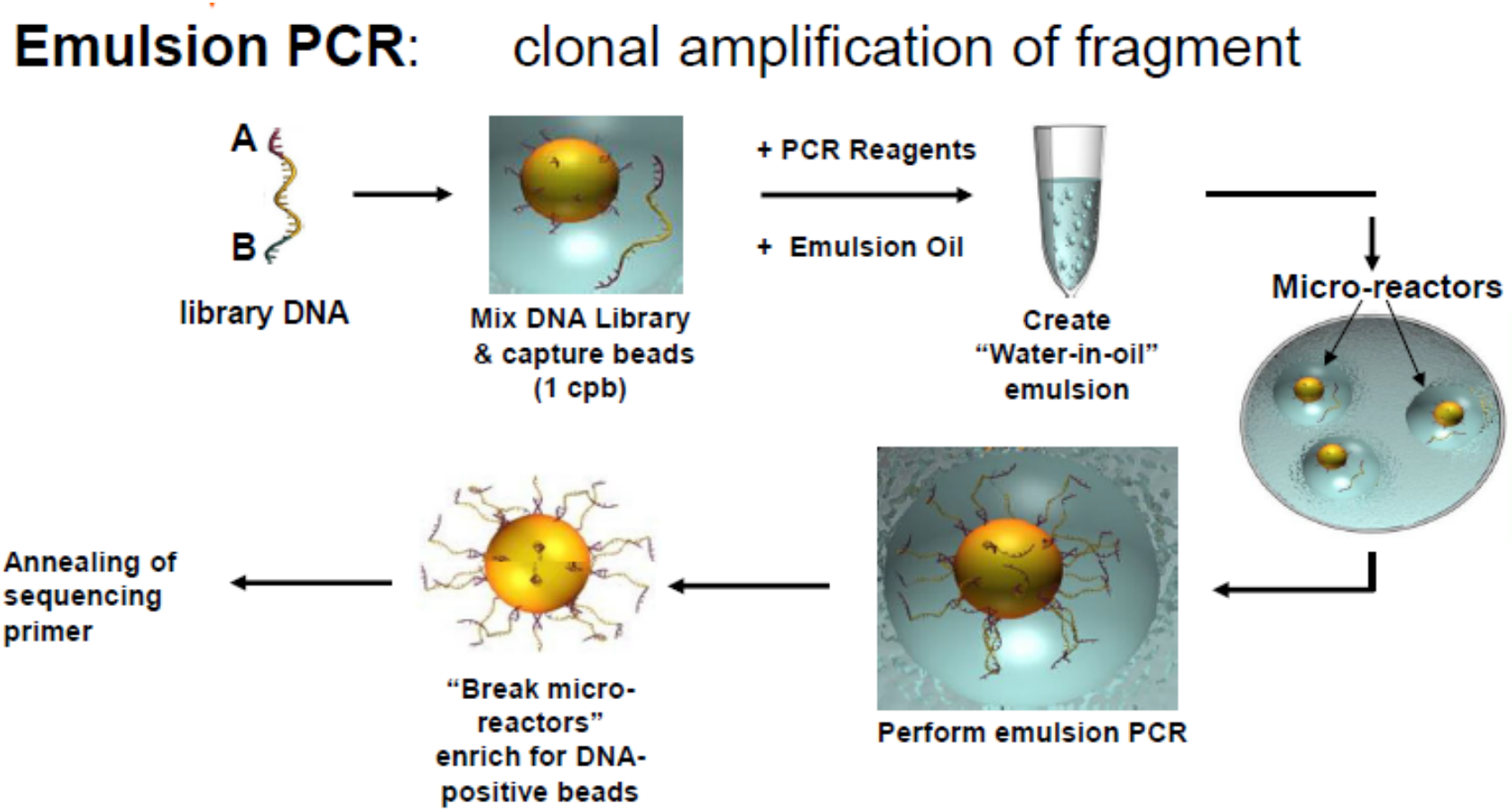
SOLiD (Life technologies)

- Sequencing by ligation

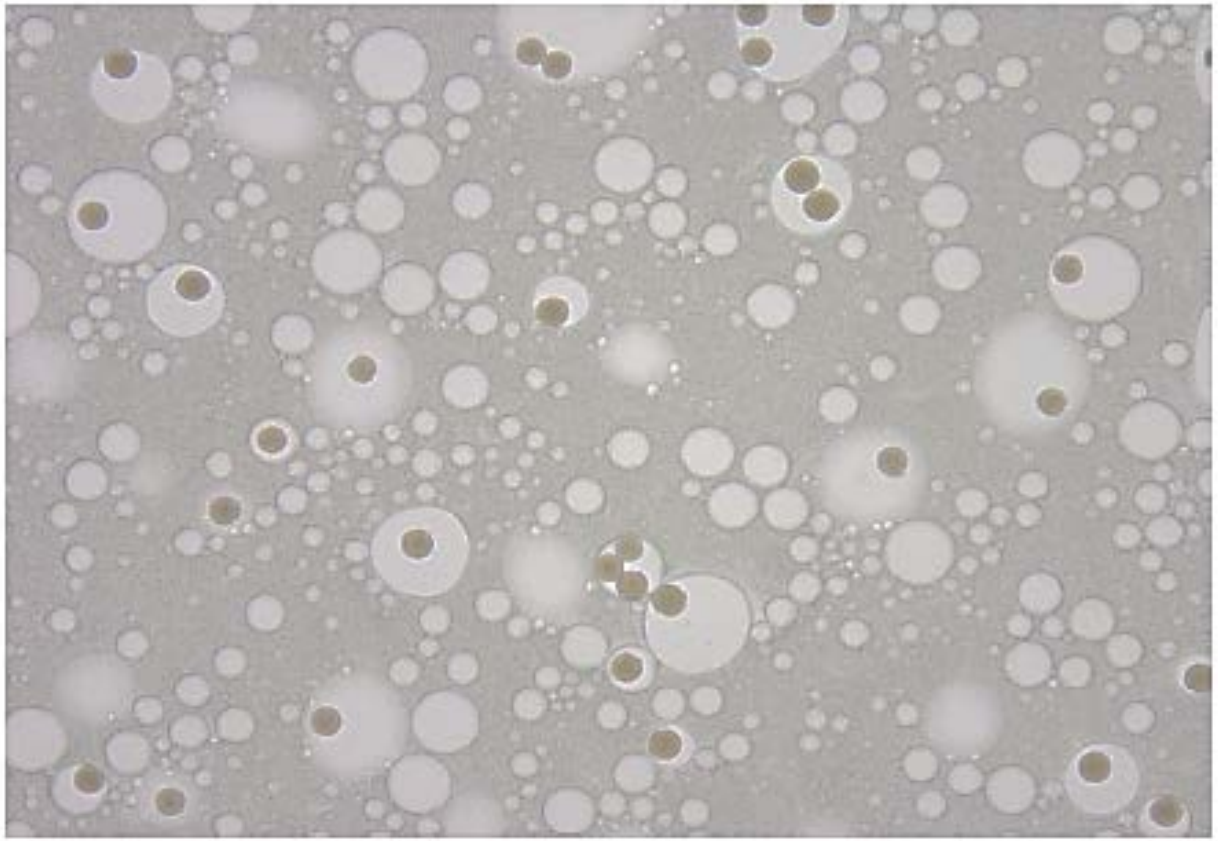
SOLiD sample prep

- Ligate biotinilated adapters to fragmented DNA
- Bind fragments to beads in a 1:1 ratio
- Emulsion PCR or on chip amplification

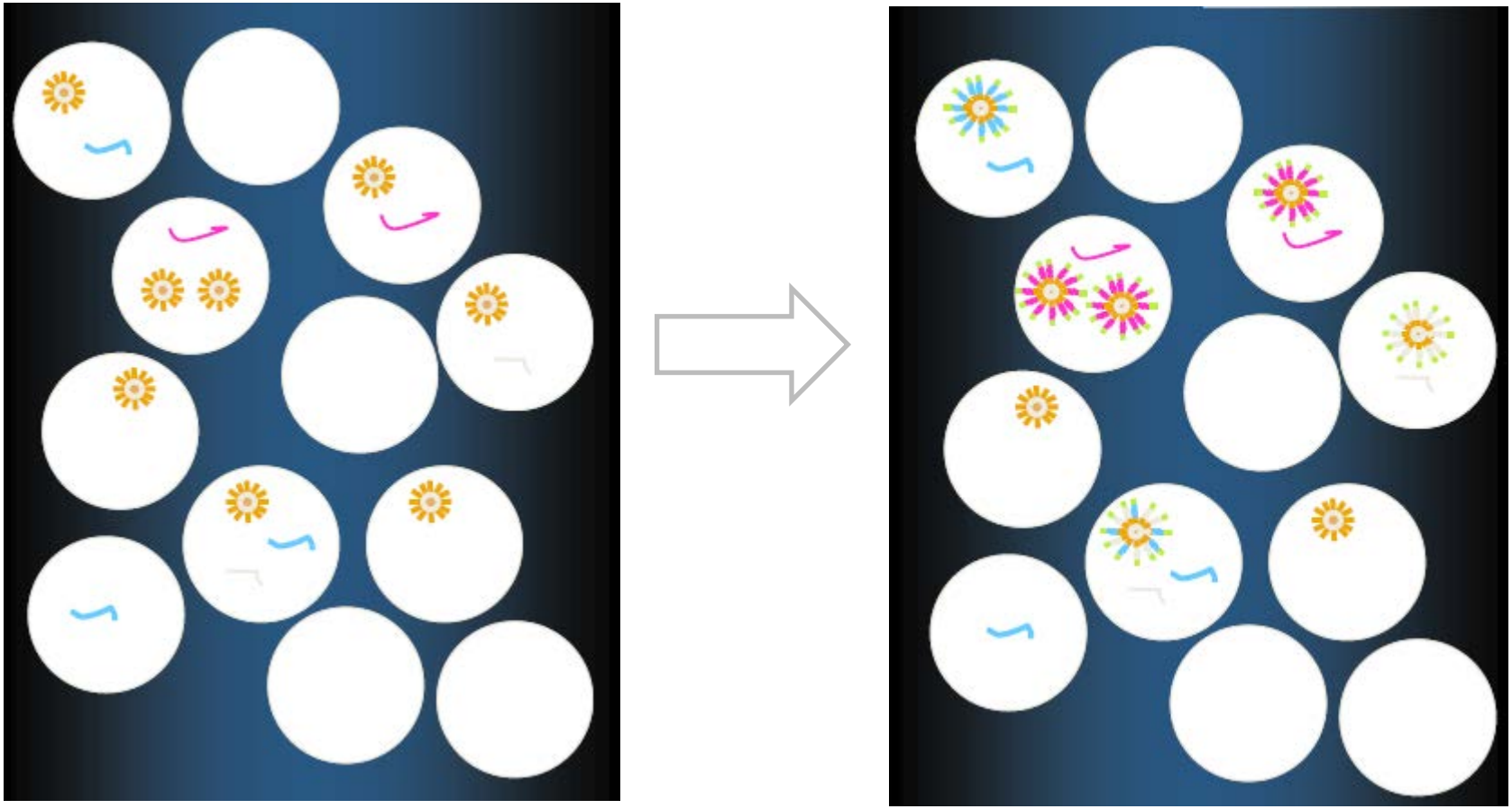
Emulsion PCR



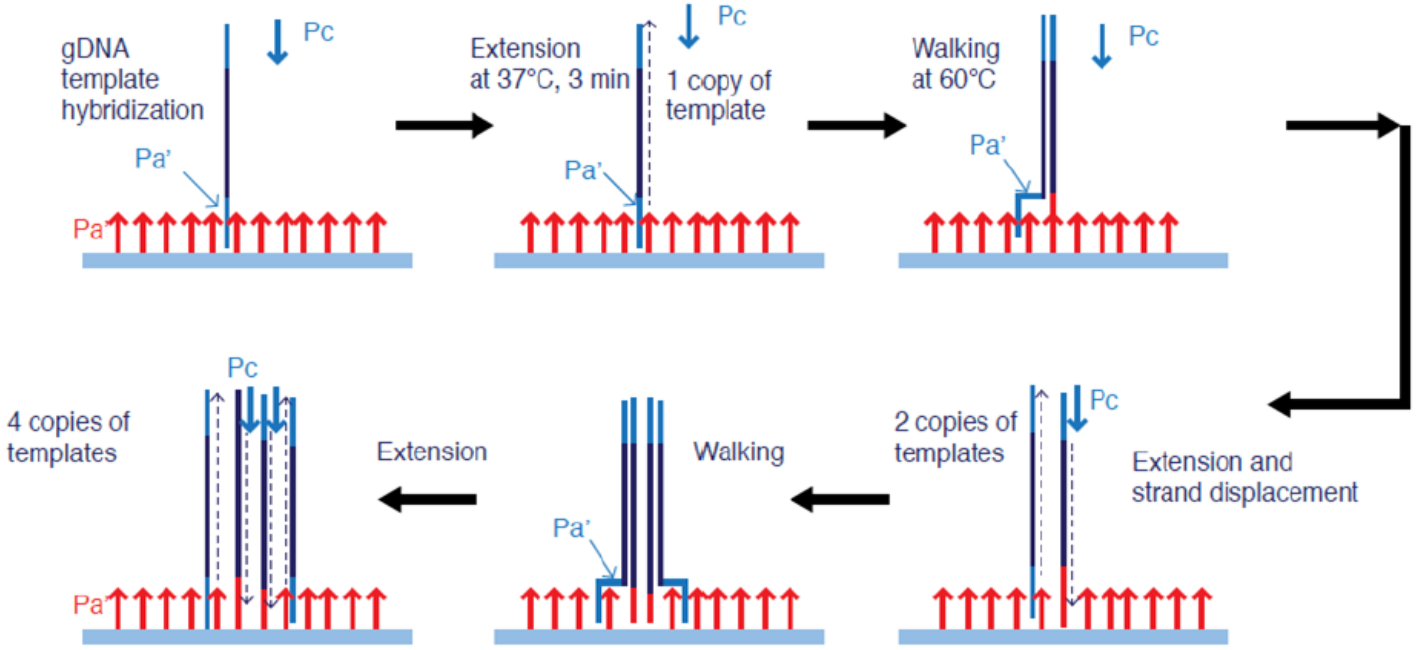
Emulsion PCR



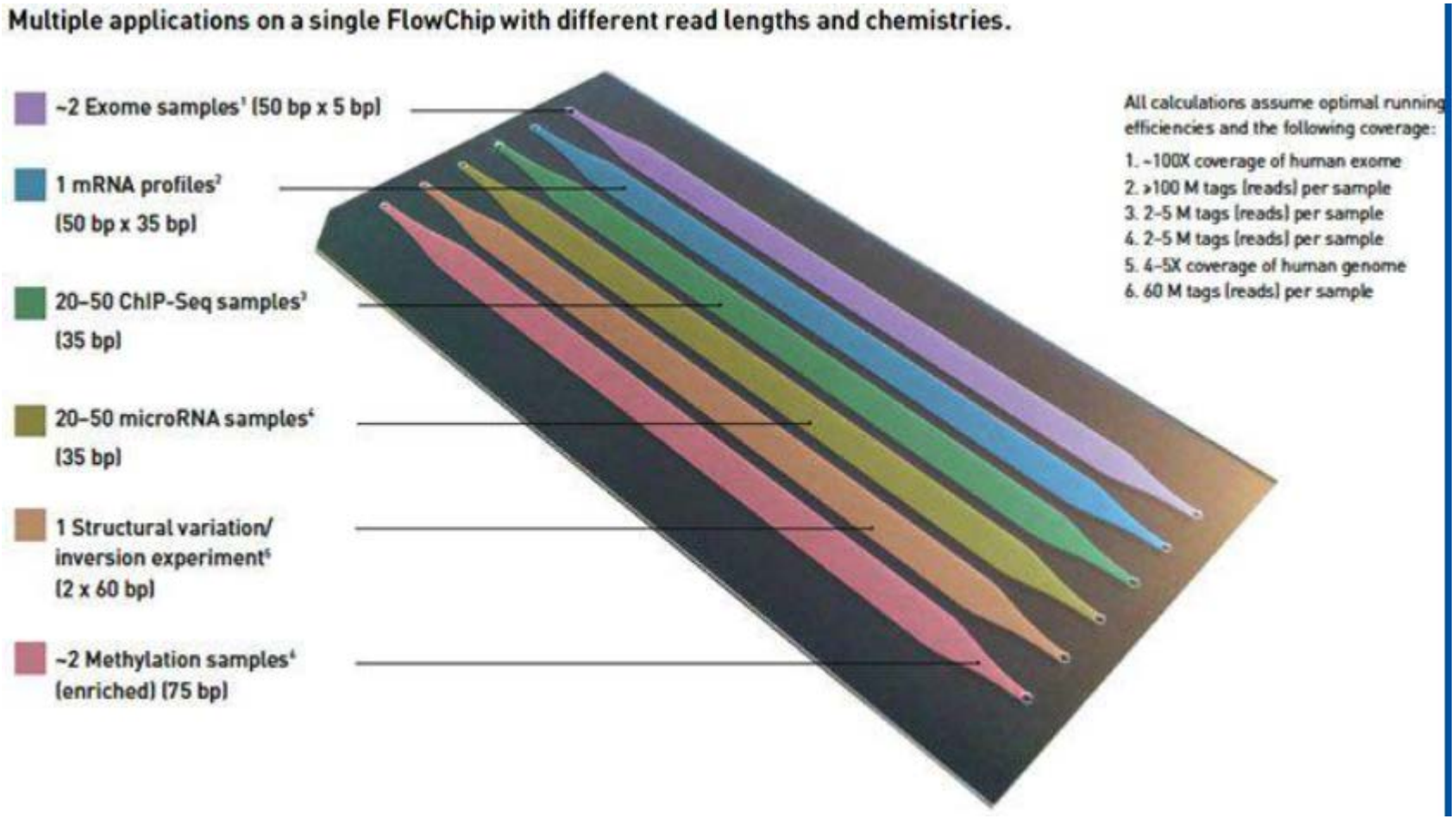
Emulsion PCR



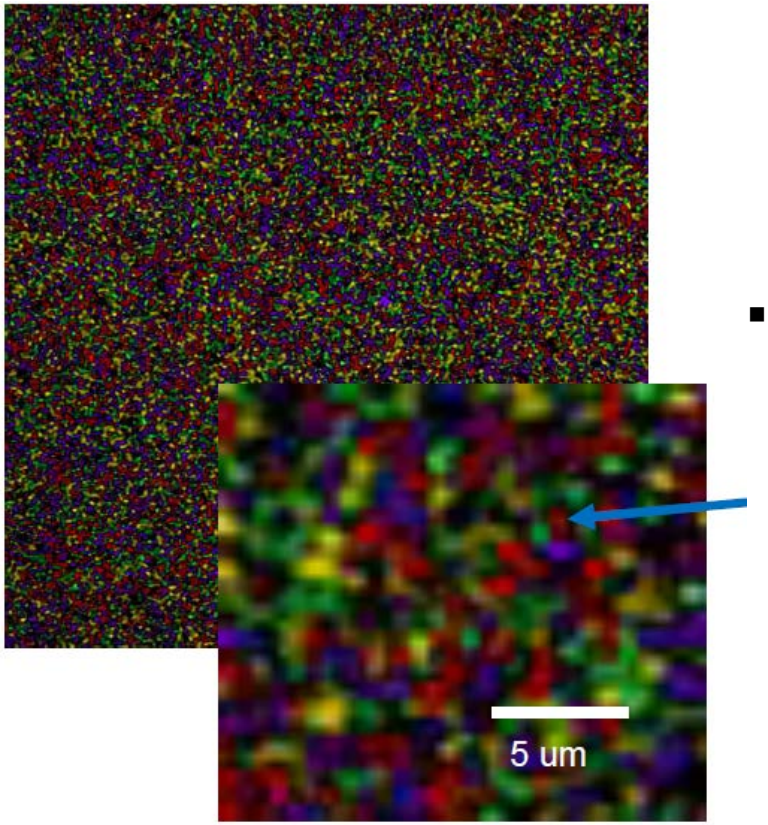
NEW: In stead of emulsion PCR: The wildfire chemistry



SOLiD sequencing chip

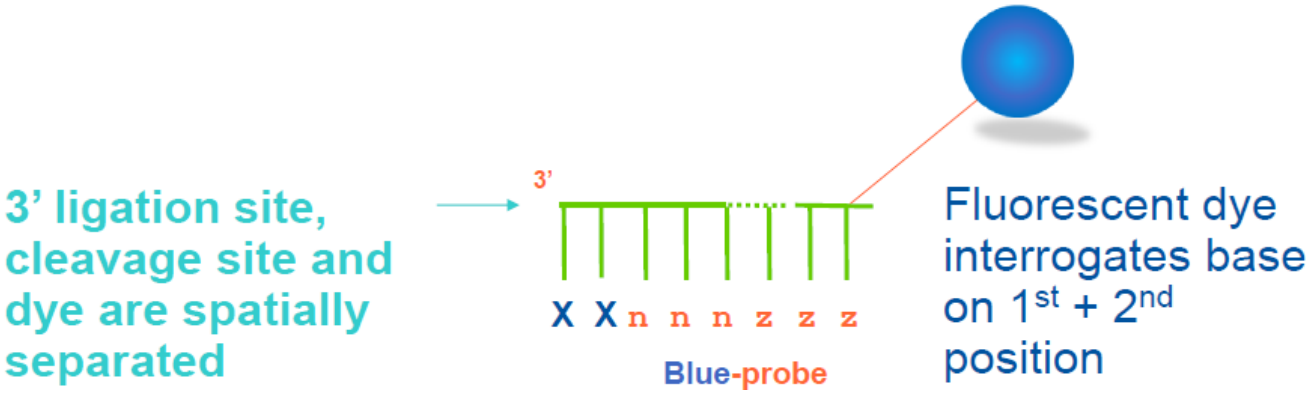


SOLiD sequencing surface

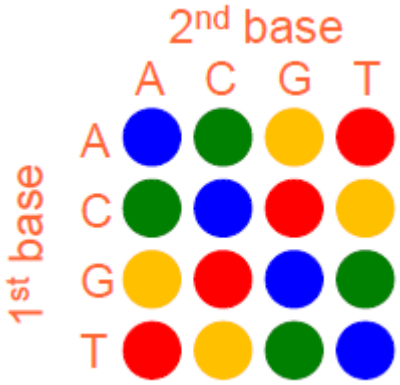


- Individual sequence colonies

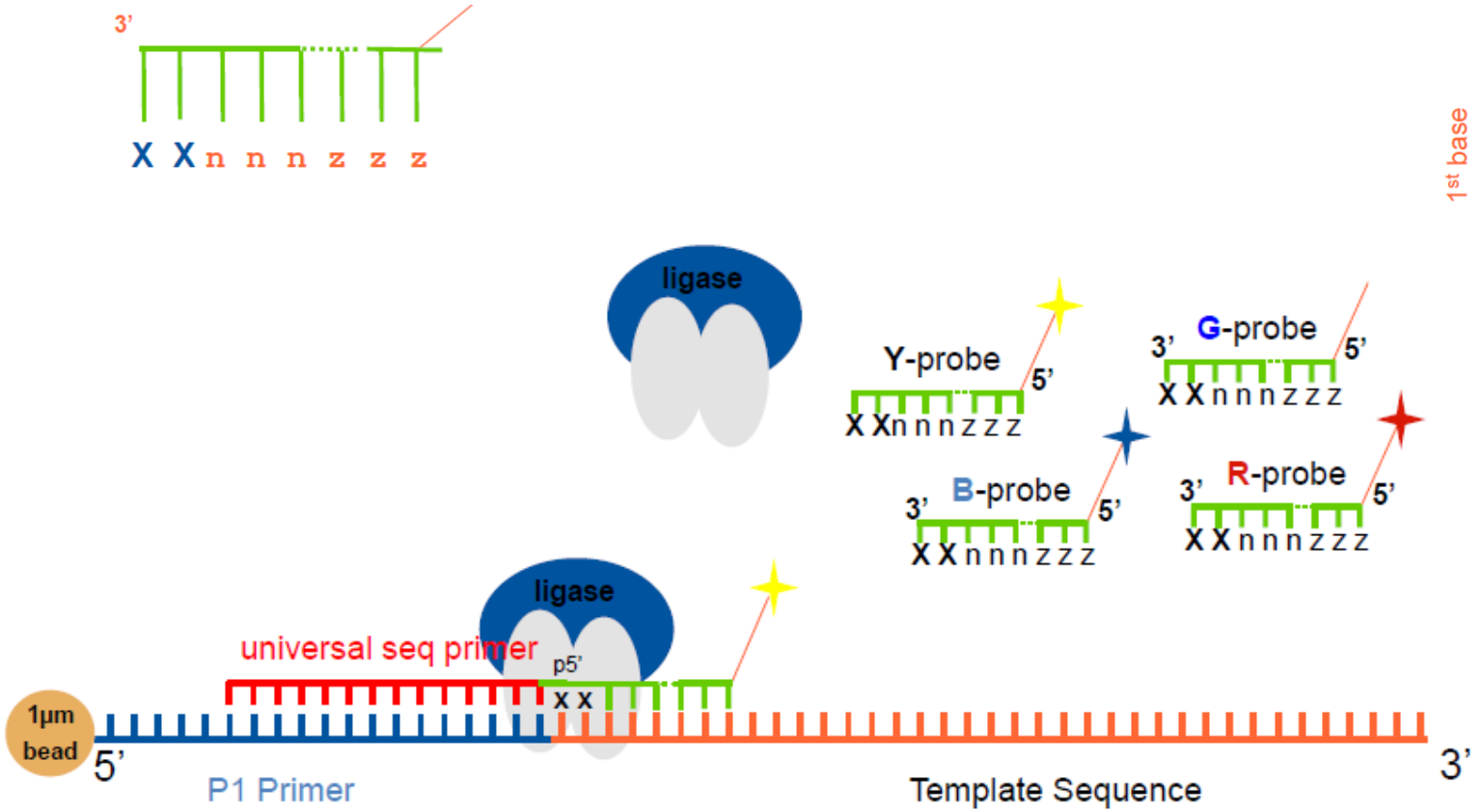
SOLiD sequencing: sequencing by ligation



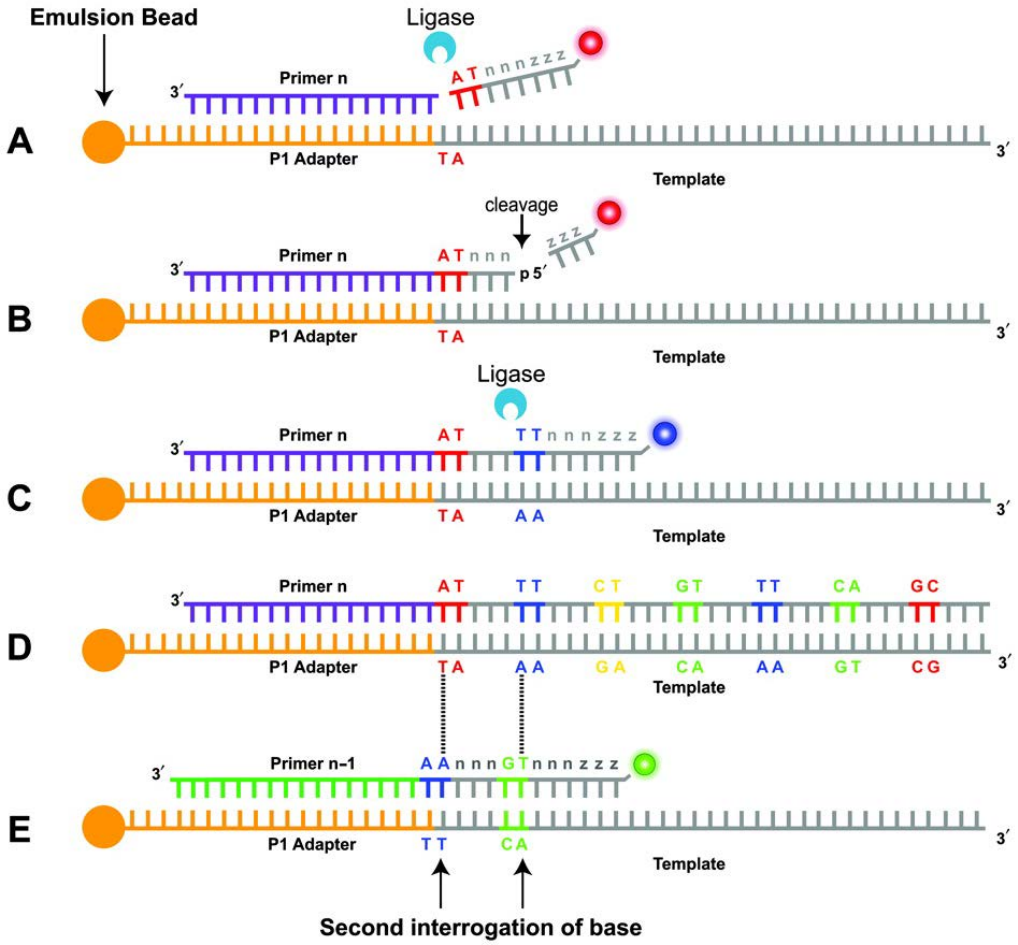
n = random A C T or G
z = base that fits all

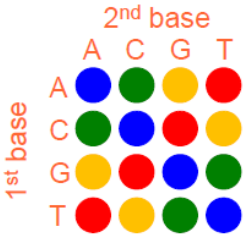


SOLiD sequencing: sequencing by ligation

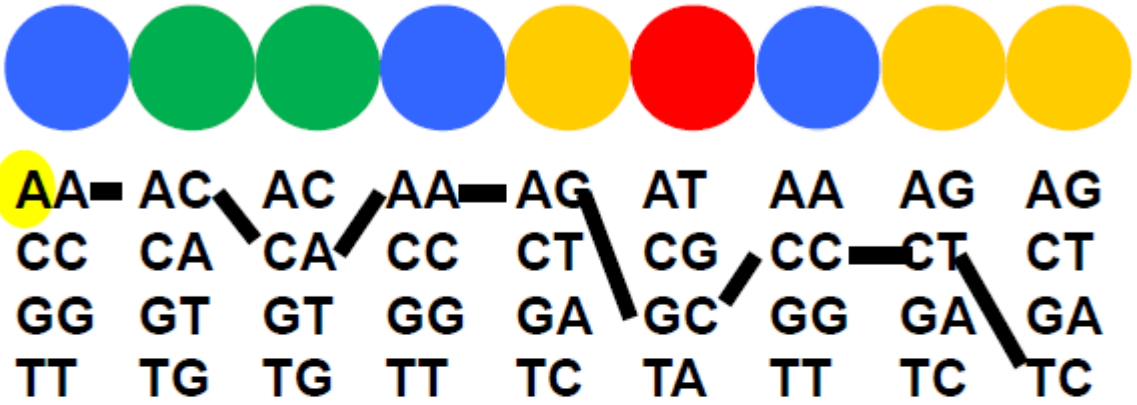


SOLiD sequencing: sequencing by ligation





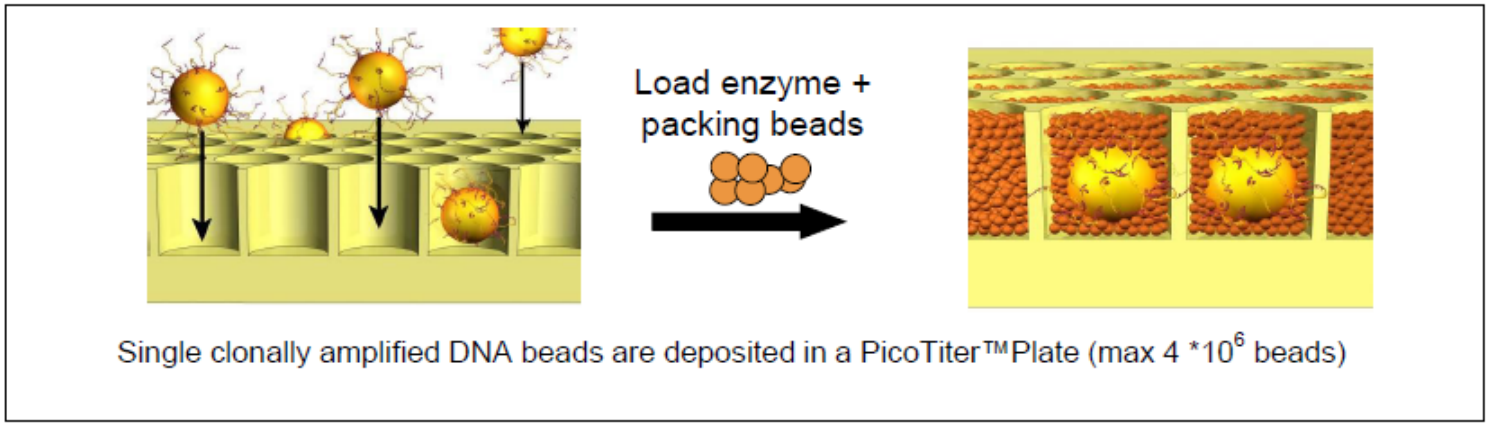
SOLiD sequencing: sequencing by ligation

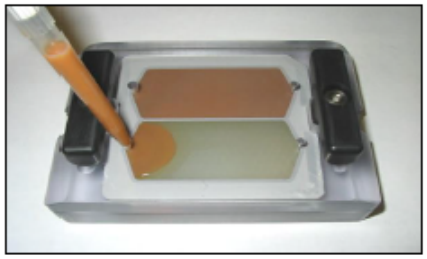


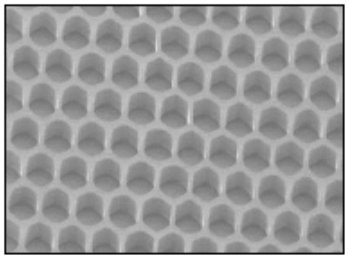
If know first base is an A then immediately it decodes 2nd base. This must be an A as Blue translates 2nd base A if first base A

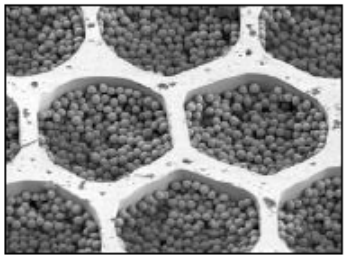
AACAAGCCTC

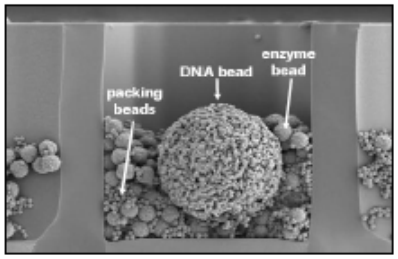
Roche Pyrosequencing







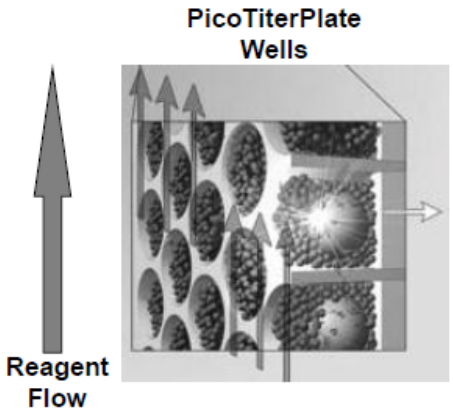


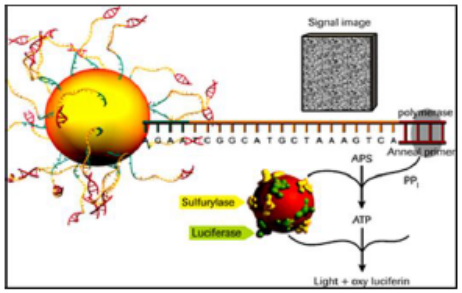


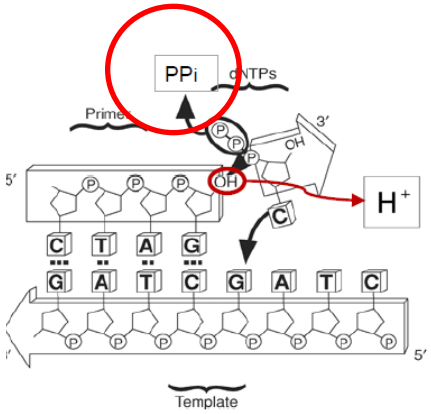
Roche Pyrosequencing

Bases (TACG) are flown sequentially across the PicoTiterPlate device

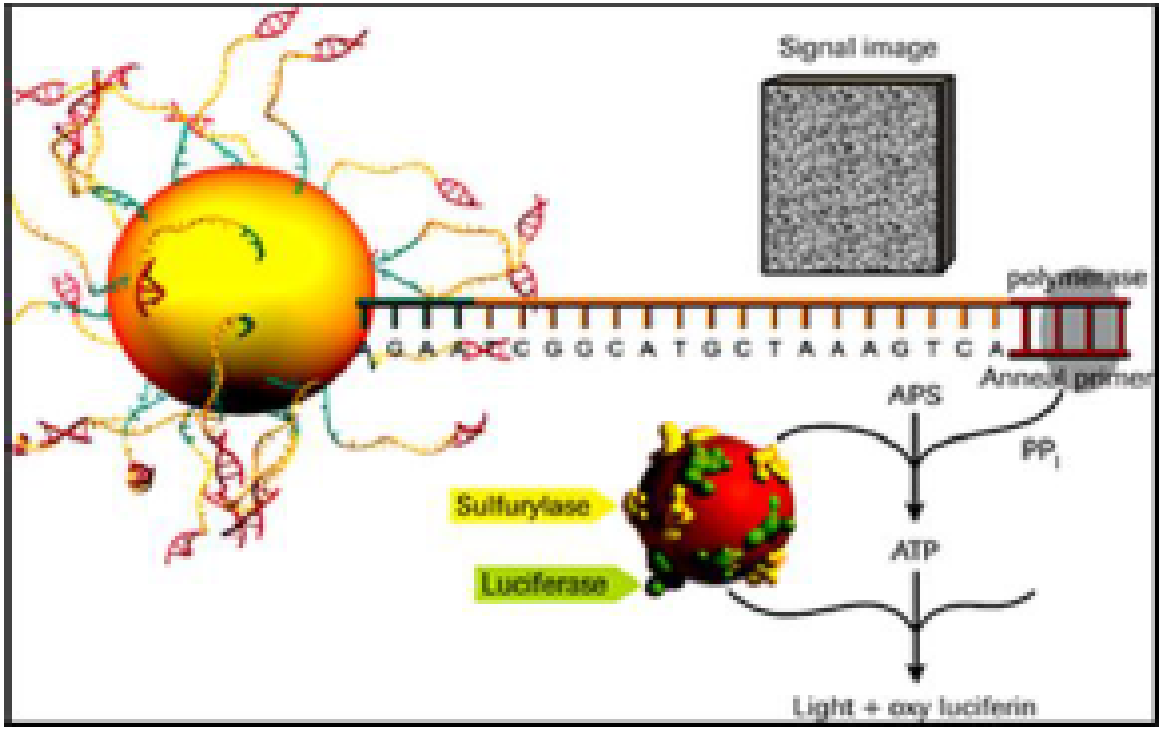
Build in complementary nucleotides generate light signals





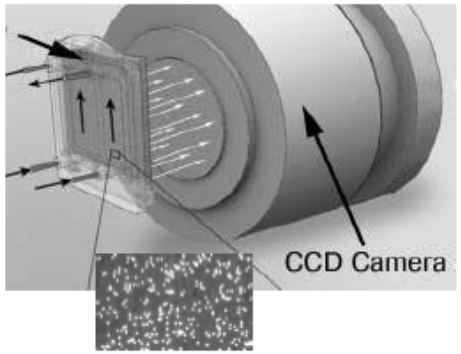


Roche Pyrosequencing



Roche Pyrosequencing

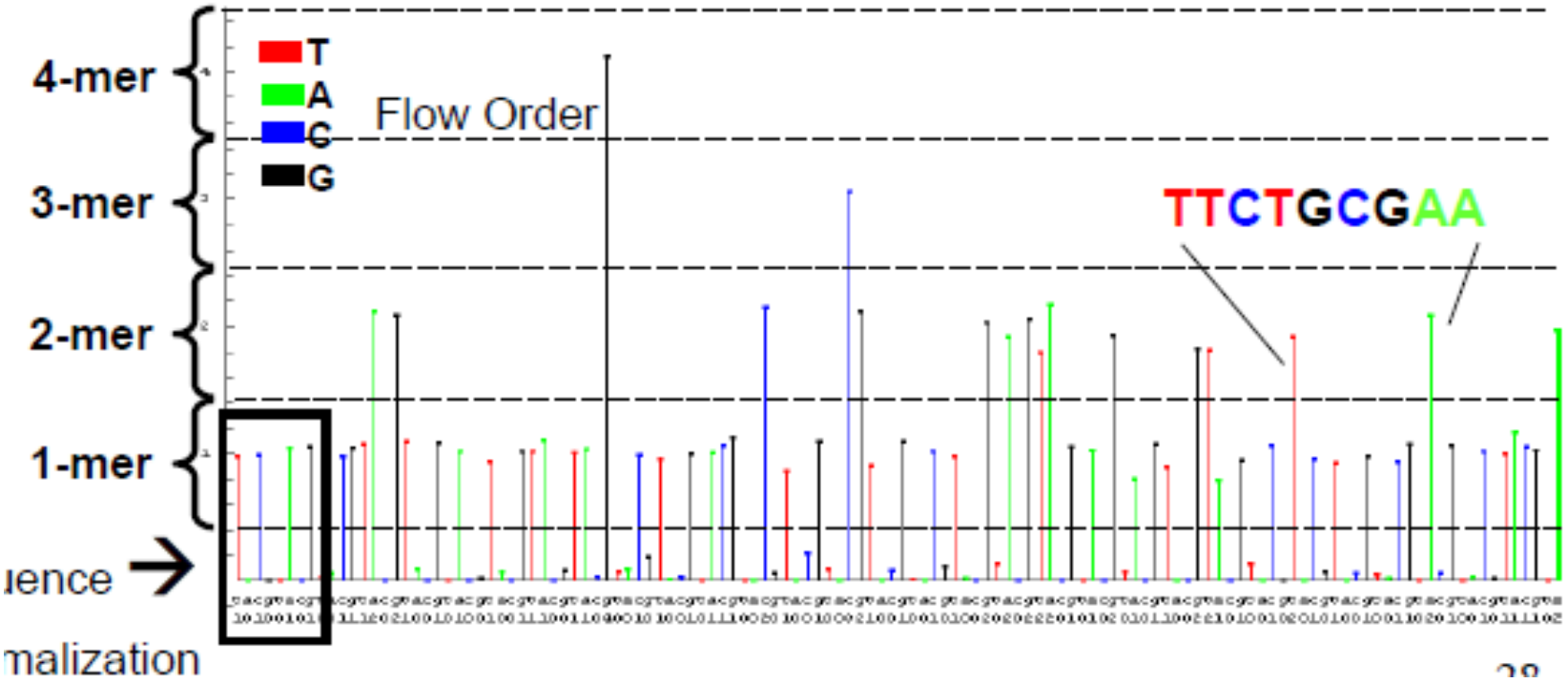
The light signal is recorded by a CCD camera.



The signal strength is proportional to the number of nucleotides incorporated

TCAG for signal calibration and normalization

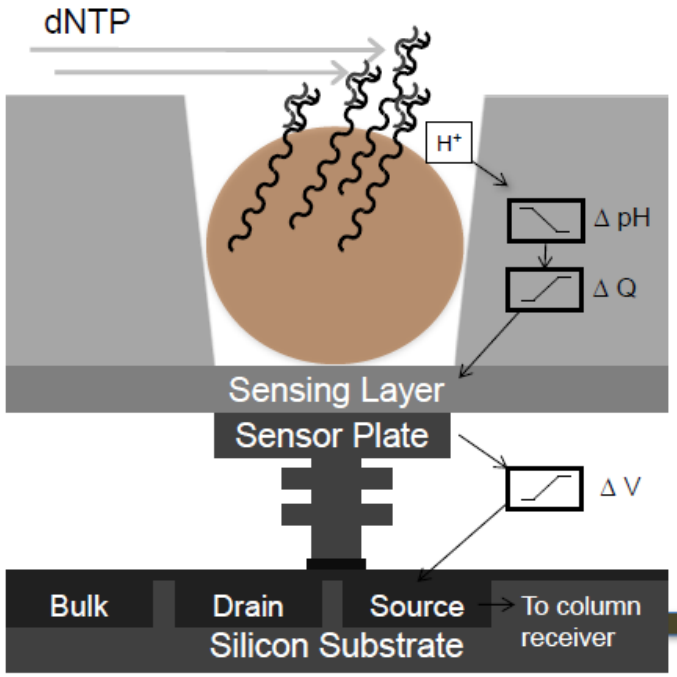
Roche Pyrosequencing



Ion Torrent

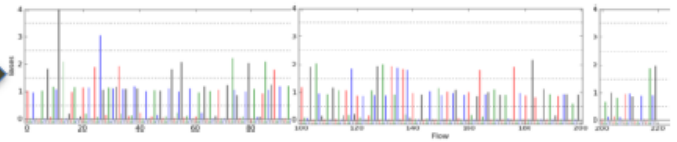
- Semiconductor detection of H+
- Detection of clonally amplified sequences on beads (emulsion PCR)

Ion Torrent

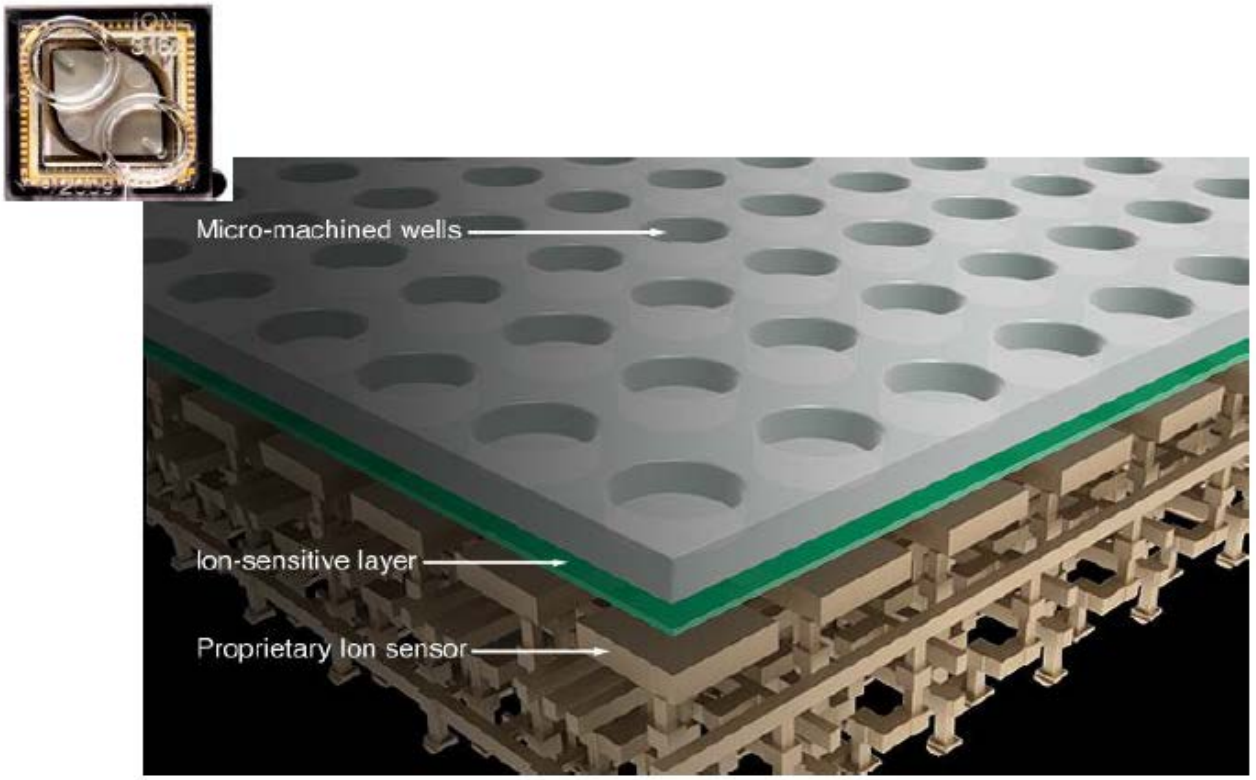


DNA → Ions → Sequence

- Nucleotides flow sequentially over Ion semiconductor chip
- One sensor per well per sequencing reaction
- **Direct** detection of natural DNA extension, **no** camera's
- Millions of sequencing reactions per chip
- Fast detection, fast cycle time, real time detection



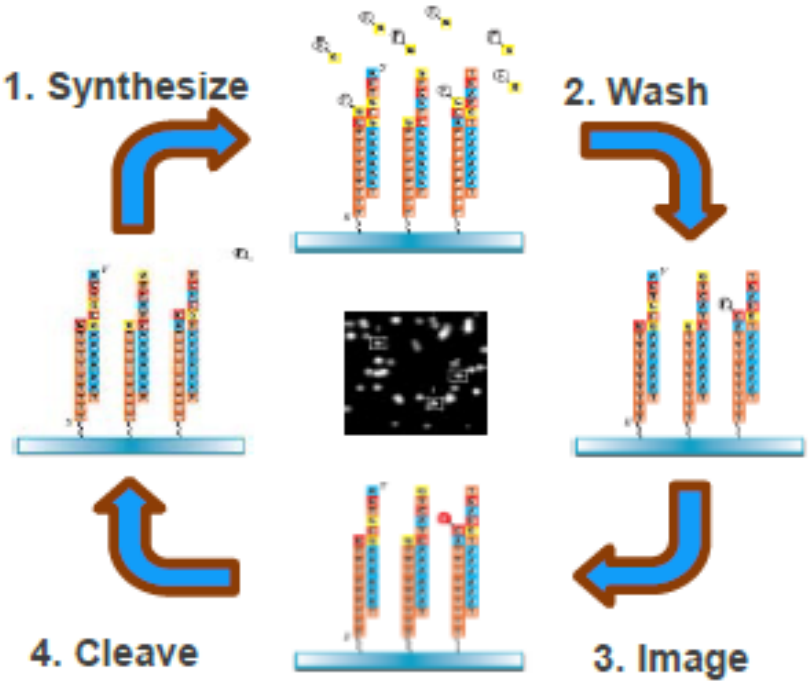
Ion Torrent


Micro-machined wells
Ion-sensitive layer
Proprietary Ion sensor

Single molecule sequencers (3rd gen)

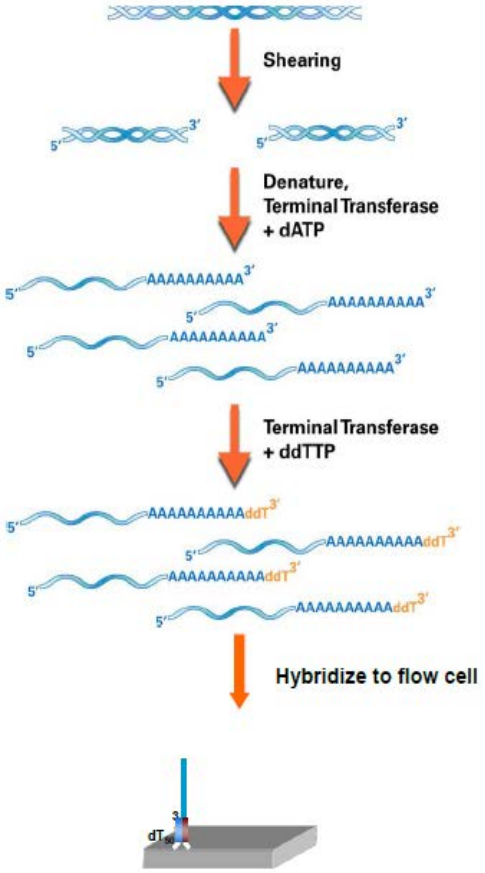
- Helicos
- PacBio
- Oxford Nanopore

Helicos



Helicos

Sampleprep Helicos:

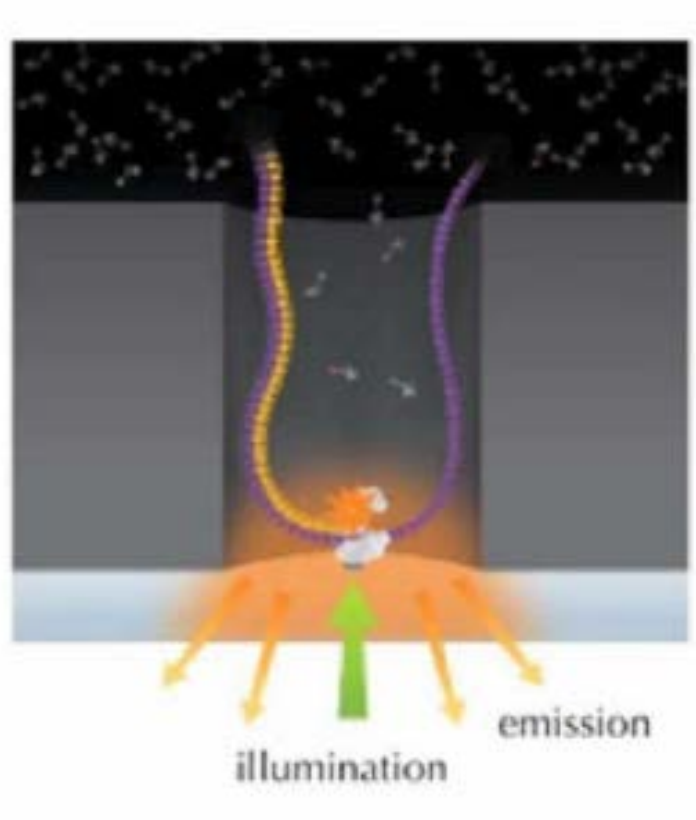


Disadvantage Helicos:

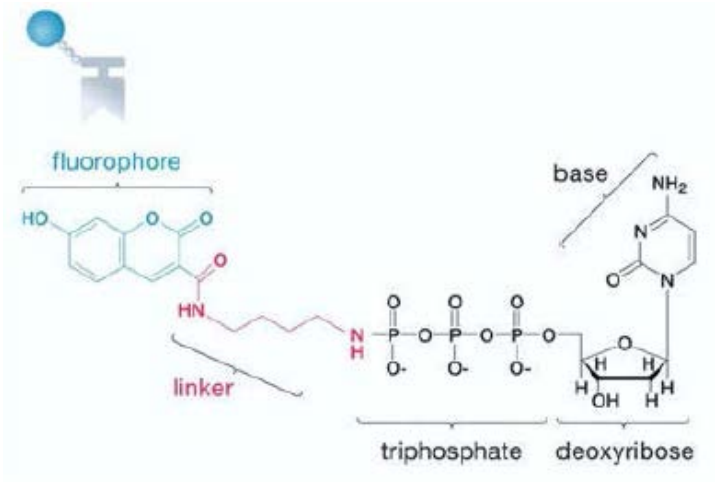
Dark nucleotides

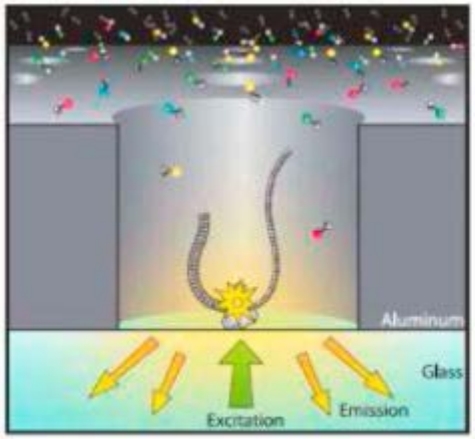
```
Reference:  ACGCTAGCAAGCT
Aligned Reads: ACGCTAGCA-GC
                CGCTAGCAAGC
               AC-CTATCAA
               ACGCTATCA-GC
               ACG-TATC--GC
```

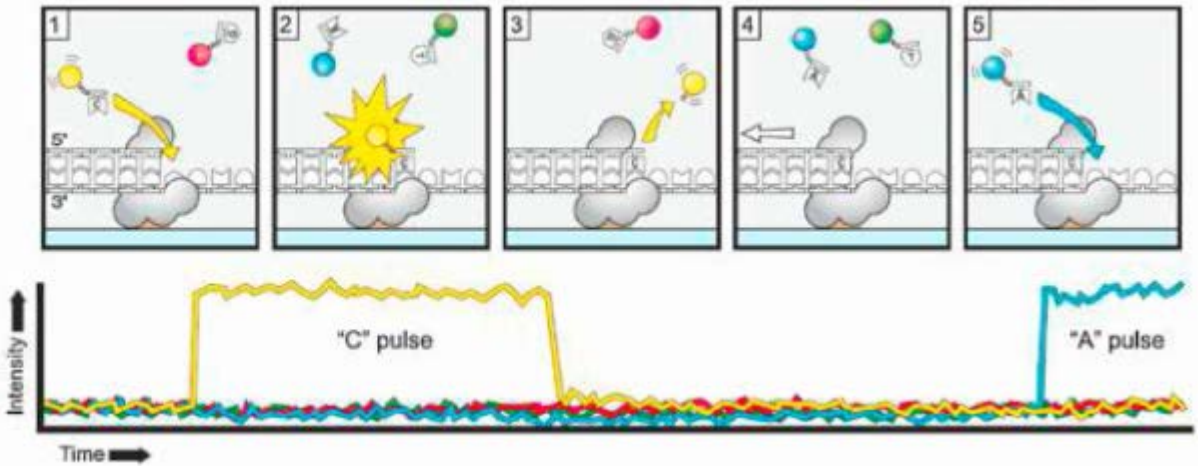
PacBio (Pacific Biosystems)



PacBio

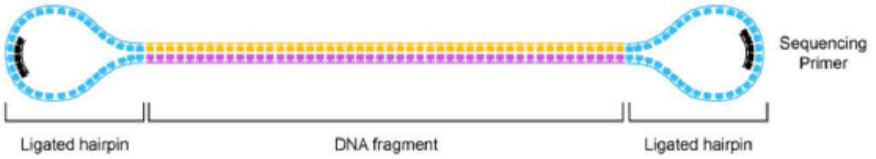

fluorophore
base
linker
triphosphate
deoxyribose


Aluminum
Glass
Excitation
Emission


1
2
3
4
5
5'
3'
Intensity
"C" pulse
"A" pulse
Time

PacBio SMRTbell

Creating a SMRTbell Template



Key Advantages:
- Structurally linear
- Topologically circular
- Structural homogeneity of templates
- BaseModification preserved
- Provides sequences of both forward and reverse strands in the same trace*

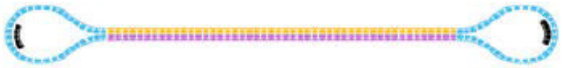
*)depending on the size of the Template

The bigger the desired insert size, the more DNA input needed...

Library insert size (Kb)	Input DNA (ng)
0.5	1000
2	3000
8	8000

Pac Bio SMRTbell sequencing strategies

Large Insert Sizes (3-10Kbp)



Continuous Long Read (**CLR**)

Generates one pass on each molecule Sequenced
(1 molecule: 1 subread)

Small Insert Sizes (250-1000bp)

Circular Consensus Sequencing (**CCS**)

Generates multiple passes on each molecule sequenced
(1 molecule: 2 or more subreads)

Oxford Nanopore (NO PRODUCT YET AVAILABLE)

Nanopore sequencing

Compound detection

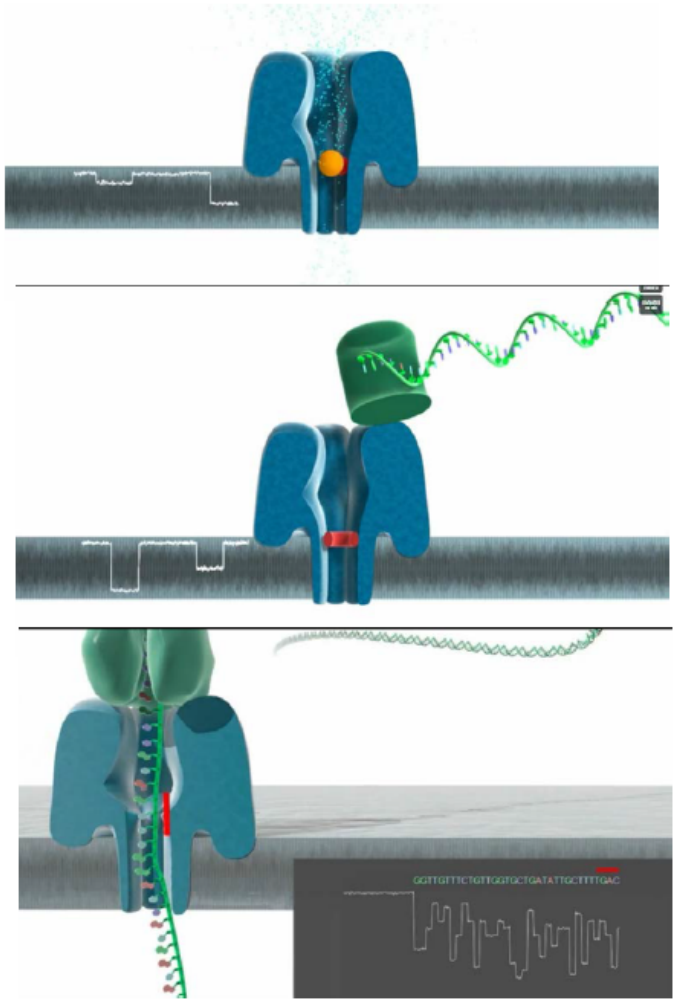
- protein sensing

DNA sequencing ***per base***

- 1 strand sequence *per base*
- *error model?*

DNA sequencing ***per triplet***

- Both strands are sequenced
- *Triplets of bases* are measured
- 1 base is measured 3x on each strand
- *error model?*



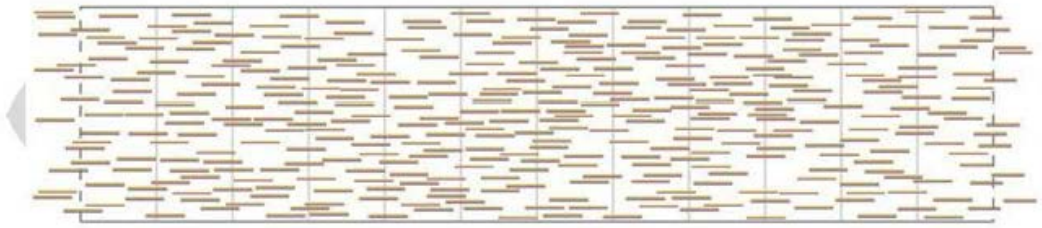
Comparison of sequencing methods

method	Pac Bio	Ion torrent	Pyroseq	Seq by synthesis	Seq by ligation	Sanger
Company	Pac Bio	Life tech	Roche	Illumina	Applied bio	
Read length	2900 bp	200 bp	700 bp	150 bp	50 bp	800 bp
Reads/run	$75 * 10^5$	$5 * 10^6$	$1 * 10^6$	$3 * 10^9$	$1,2 * 10^9$	1
Time /run	2 hours	2 hours	24 hours	10 days	14 days	2 hours
$ $1*10^6$bases	$2	$1	$10	$0.1	$0.13	$2400

Read length

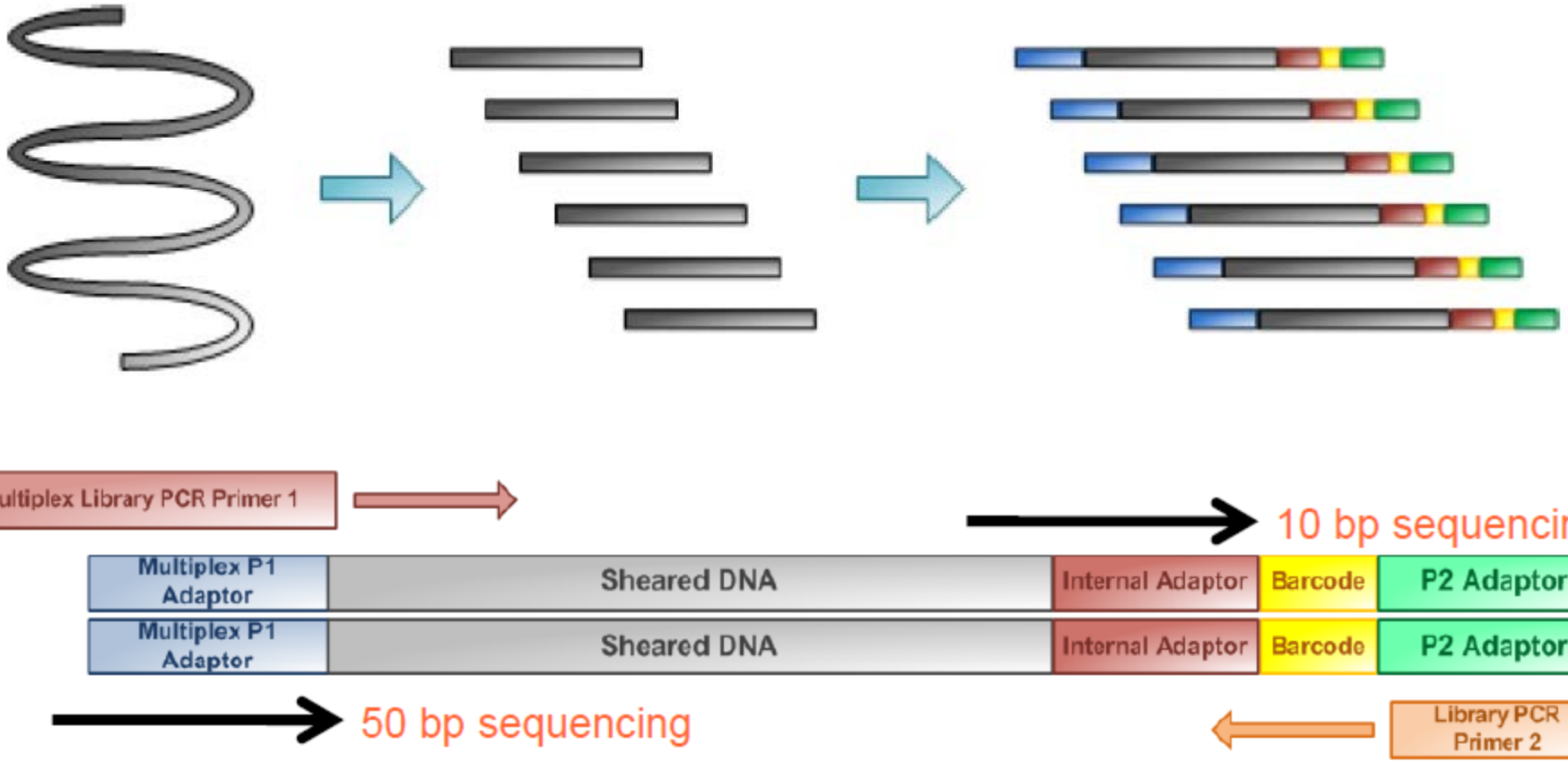
Why longer Reads?

Short reads:

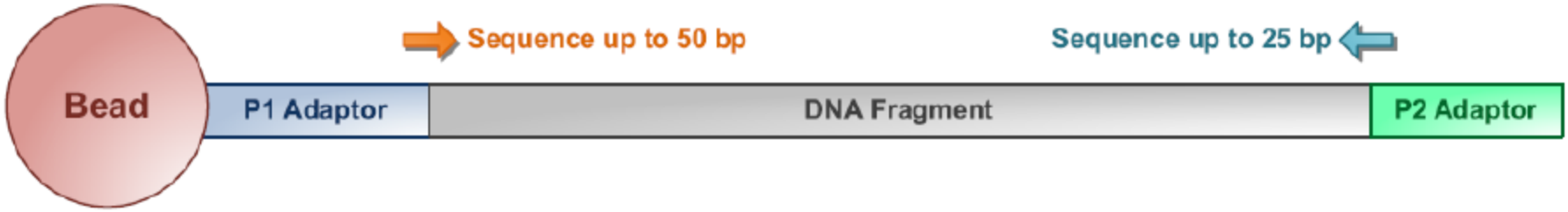


Long reads:

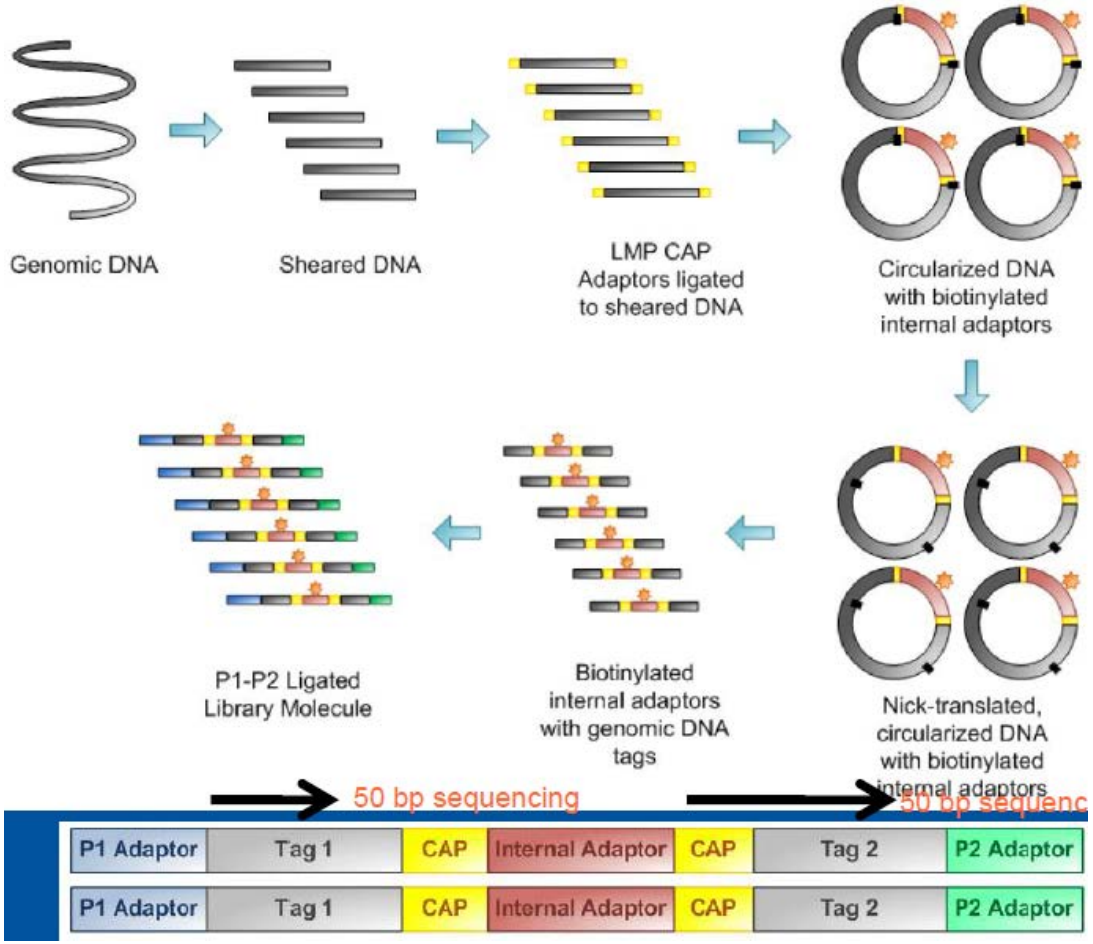
Sequencing aproaches: barcoding



Sequencing aproaches: Paired end



Sequencing aproaches: mate paired



For large structural variations

NGS Topics not covered in this presentation

- De novo seq → Sequencing and building a new genome
- Resequencing → looking for structural and functional mutations
- cDNA seq → RNA expression and mutations and splice variants
- MiRNA seq → seq of 20-25bp non coding regulating RNA's
- ChIP seq→ Chromatin Immunoprecipitation, discover regulation sequences
- Methylation sequencing → Epigenetic DNA regulation
- Metagenomics → sequencing of a pool of organisms (bacteria/virussus)
- Targeted sequencing→ Fish out a pool of sequences you like to sequence
- NGS data analysis

acknowledgements

All speakers of the Next gen sequencing short course.