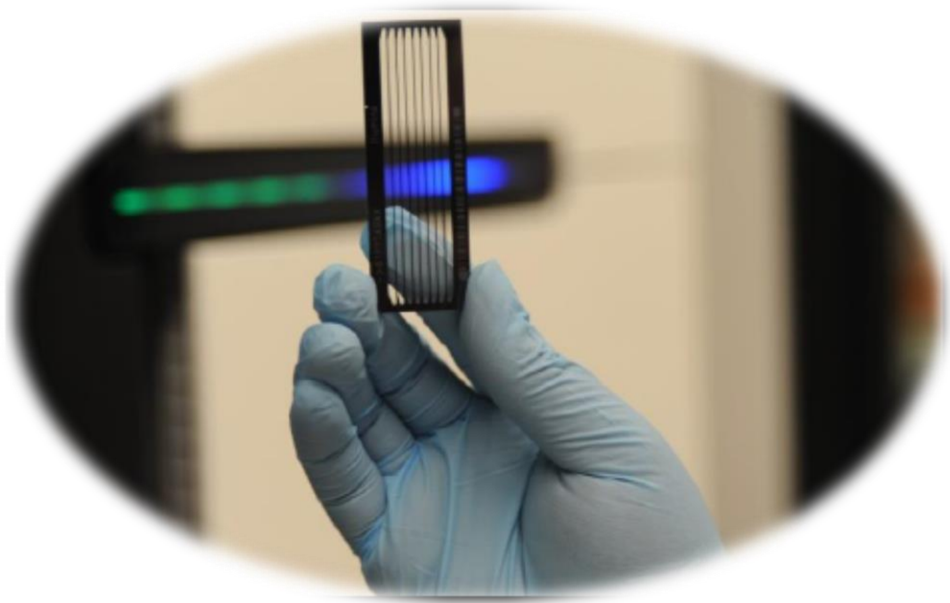


Sequencing technologies

NAHLA BAKHAMIS

- How modern DNA sequencing technologies work.
- Historical background on DNA sequencing.
- The application of the most recent sequencing technologies in both research and clinical settings.

DNA sequencing:

The determination of the precise sequence of nucleotides in a sample of DNA.

The most popular method;

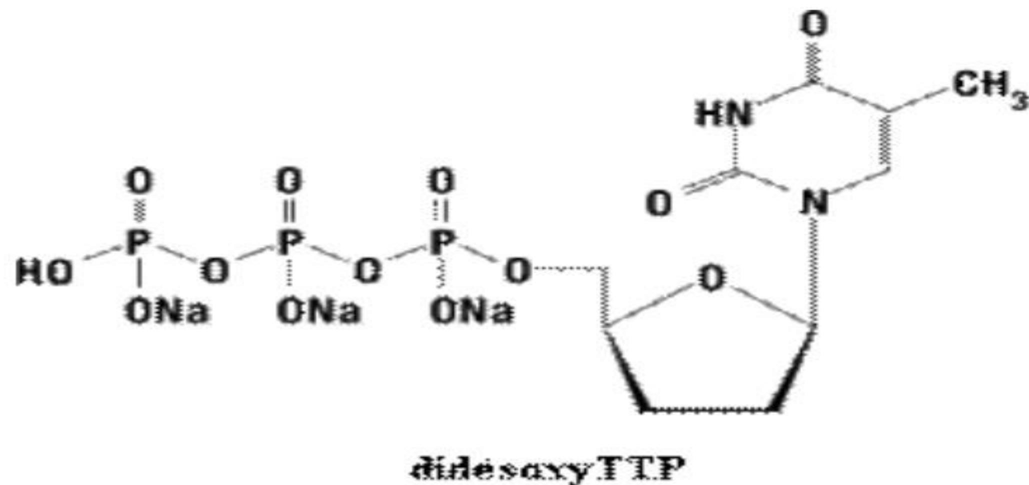
Sanger method OR dideoxy method

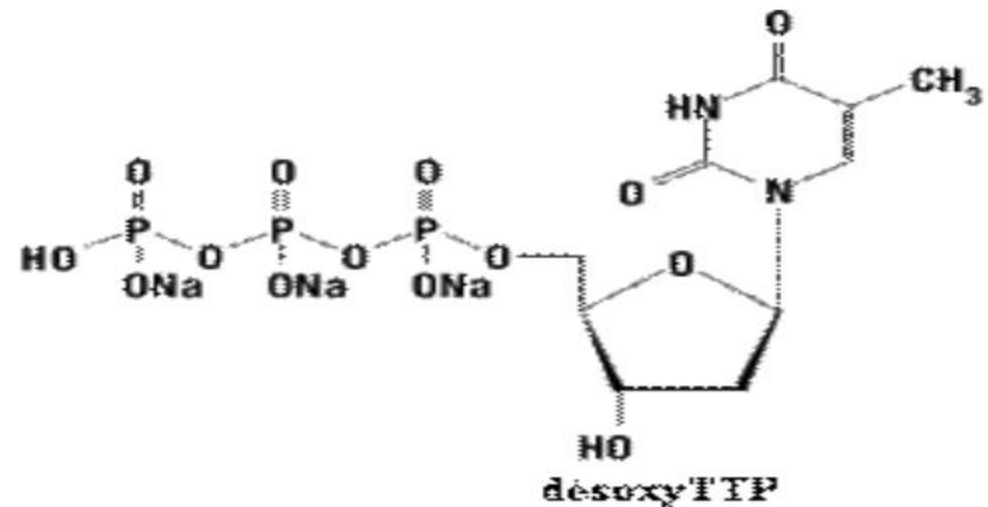
Fredrick Sanger:

- Double Nobel Laureate in Chemistry (1958) for work in protein structure, and (1980) with Walter Gilbert For DNA sequencing.
- Pioneer in DNA sequencing developed technologies still in use today.
- Sequence the first whole genome of 5,000 bases (bacteriophage).

Dideoxy method; chain termination method:

- Gets its name from the critical role played by synthetic nucleotides that lack –OH.
- ddTTP when added to the growing DNA strand, chain elongation stops as there is no 3' –OH for the next nucleotide to attach





Uses of DNA sequencing:

- De novo sequencing:

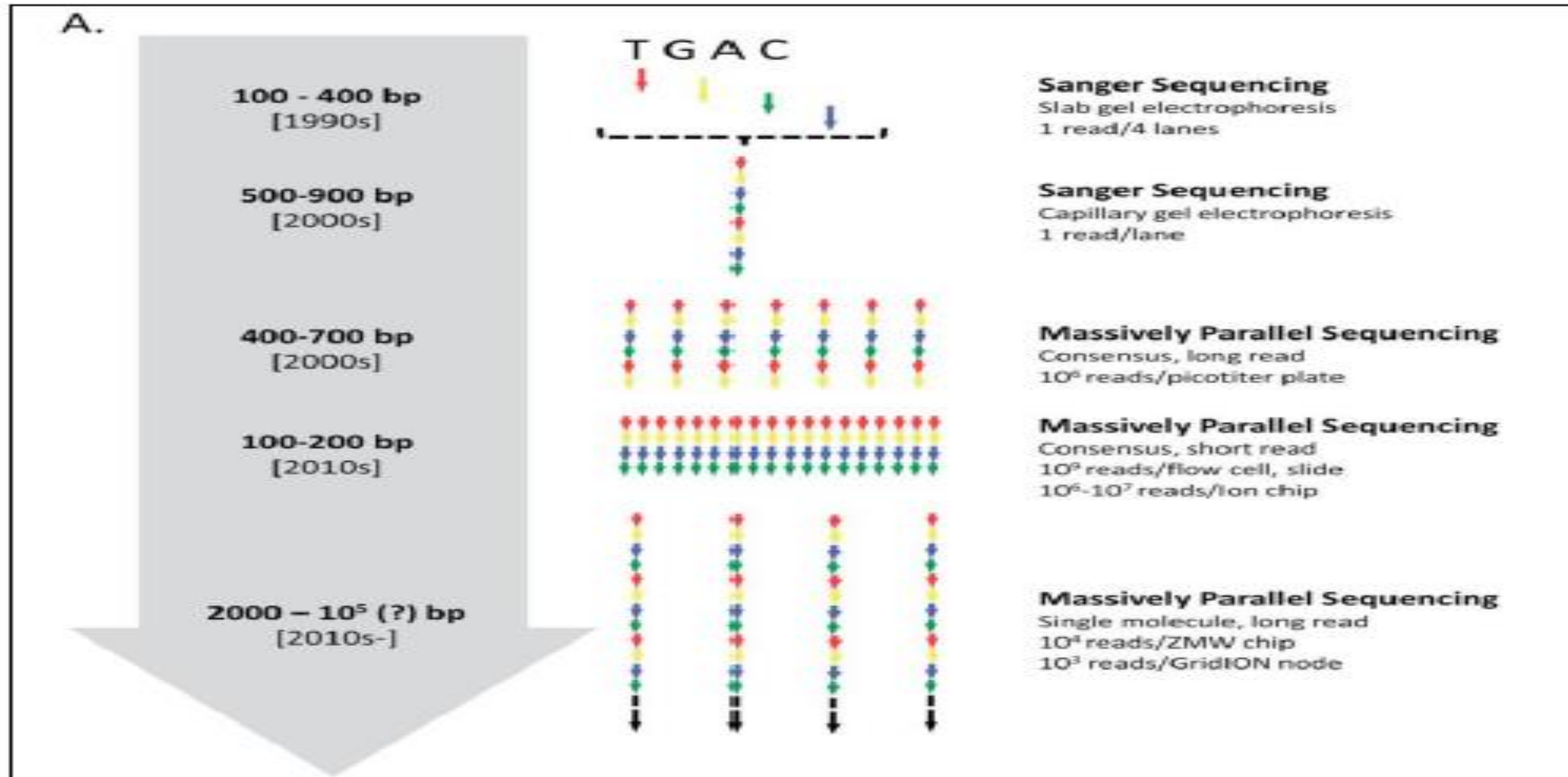
- Determine the DNA for the first time
- 1977 bacteriophage (5.368bp)
- 2001, human genome sequence draft (3,000,000,000 bp)
- Now searching the genome of other species

Re-sequencing:

- Determine variation of a kwon sequence.
- Determine normal variations (Polymorphism) within individuals.
- Discovery of diseases causing mutations.
- Gold standard for many mutation screening

The progress of DNA sequencing:

Evolved rapidly especially in the last decade



Sanger sequencing method:

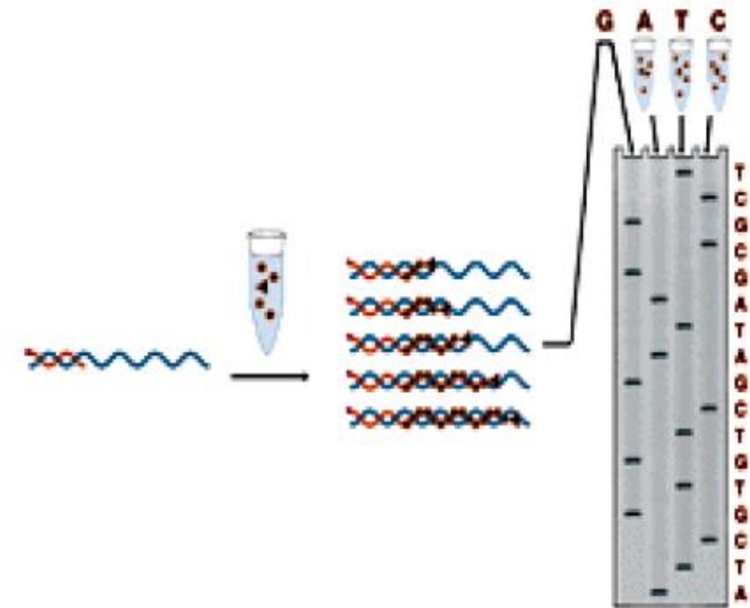
4 Tubes reaction require:

DNA template, primers, polymerase, free nucleotides (dNTPs) and ddTP.

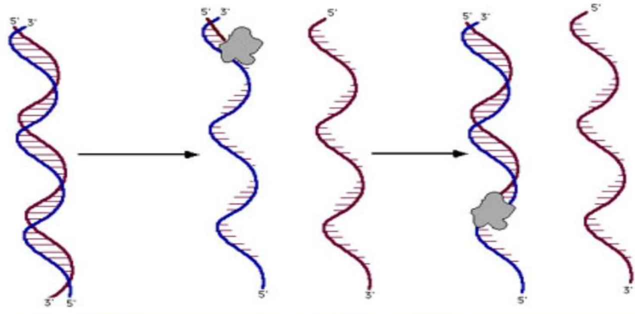
How does it work?

1. Denatured the double strands:
2. Primers annealing
3. Synthesis of new strands (adding dNTPs)
4. Dideoxynuclotides (without –OH).
5. Copied template separated by electrophoresis according to their lengths
6. The length determine the position

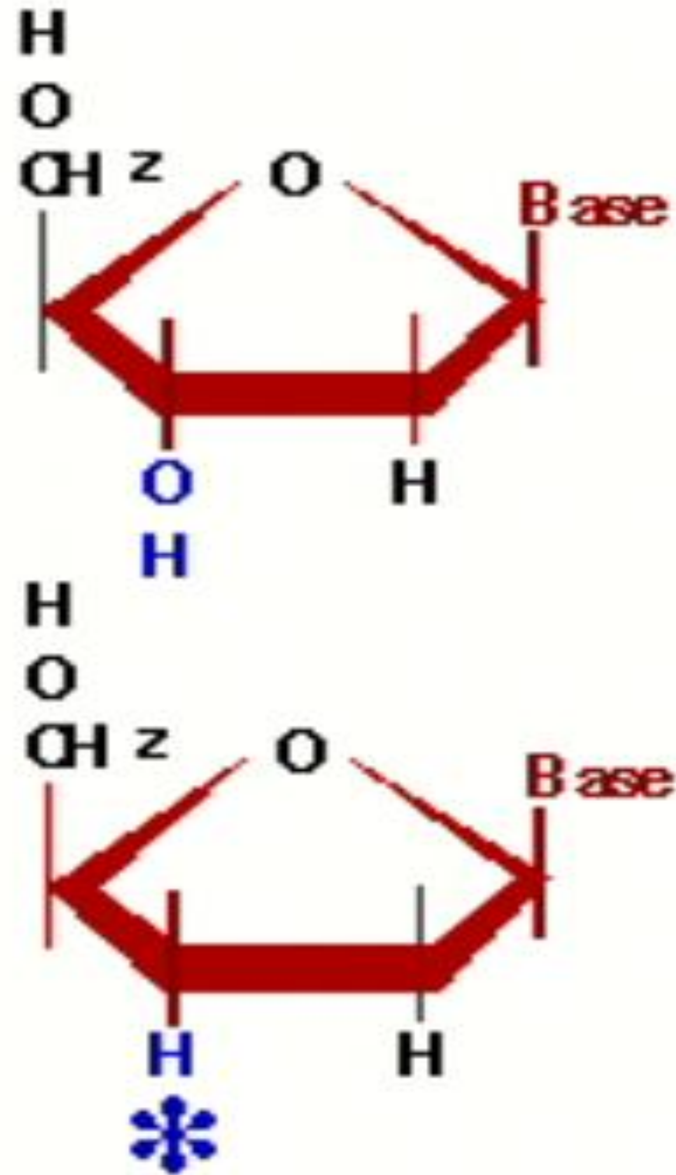
And the fluoresce determined the base

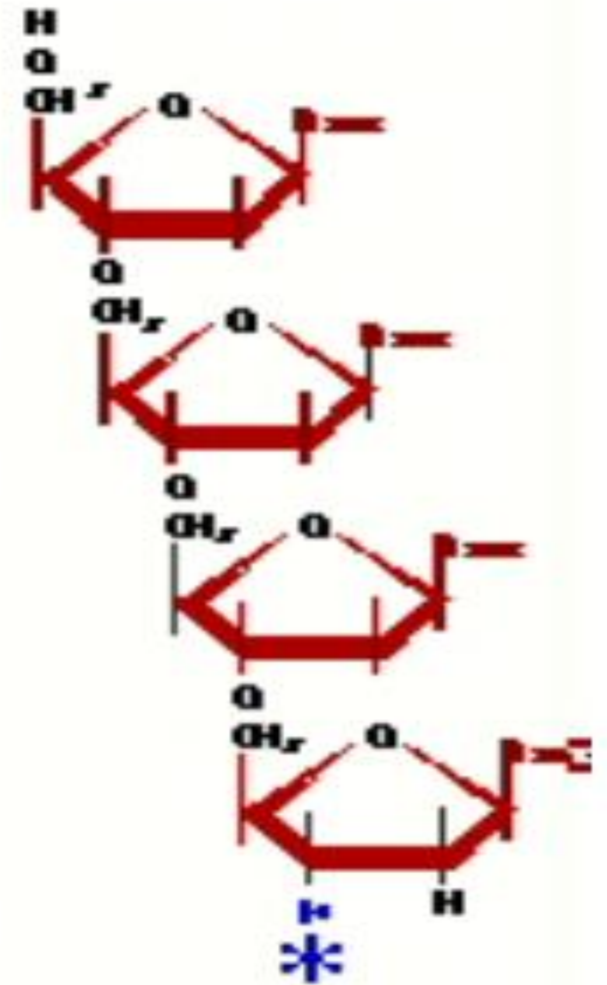


Normal nucleotides:



Dideoxy Chain Terminators:





DNA Polymerase reads the template strand and synthesizes a new second strand to match:

IF 5% of the T nucleotides are actually <u>dideoxy</u> T, then each strand will terminate when it gets a ddT on its growing end:

5' - TACGCGGTAACGGTATGTTCGACCGTTTAGCTACCGAT•

5' - TACGCGGTAACGGTATGTTCGACCGTTTAGCT•

5' - TACGCGGTAACGGTATGTTCGACCGTTT•

5' - TACGCGGTAACGGTATGTTCGACCGTT•

5' - TACGCGGTAACGGTATGTTCGACCGT•

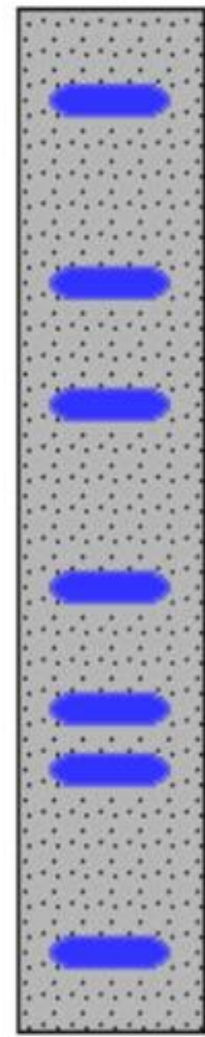
5' - TACGCGGTAACGGTATGTT•

5' - TACGCGGTAACGGTATGT•

5' - TACGCGGTAACGGTAT•

5' - TACGCGGTAACGGT•

5' - TACGCGGT•



GCGAATGCGTCCACAACGCTA**C**

GCGAATGCGTCCACAACG**C**

GCGAATGCGTCCACAA**C**

GCGAATGCGTCCA**C**

GCGAATGCGTC**C**

GCGAATGCGT**C**

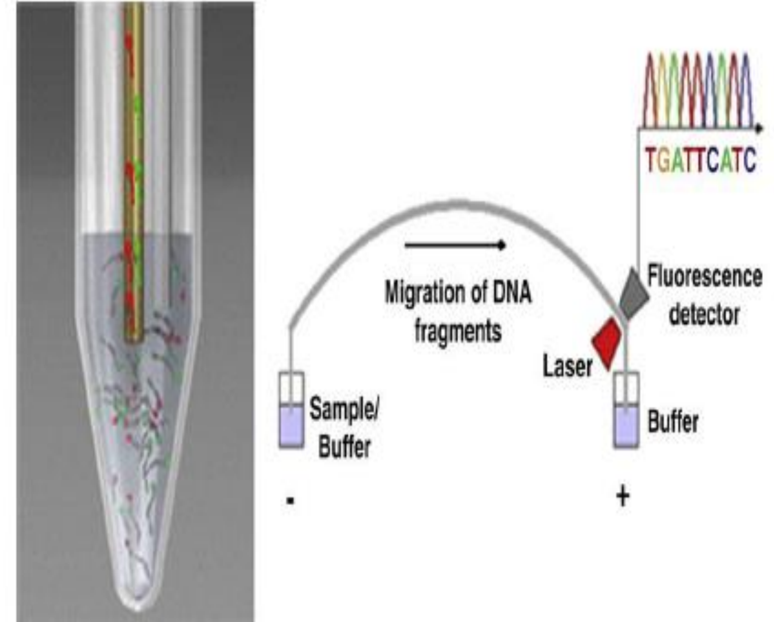
GCGAATG**C**

Gel:

Base	Sequence
G	GCGAATGCGTCCACAACGCTACAGGTG
T	GCGAATGCGTCCACAACGCTACAGGT
G	GCGAATGCGTCCACAACGCTACAGG
G	GCGAATGCGTCCACAACGCTACAG
A	GCGAATGCGTCCACAACGCTACA
C	GCGAATGCGTCCACAACGCTAC
A	GCGAATGCGTCCACAACGCTA
T	GCGAATGCGTCCACAACGCT
C	GCGAATGCGTCCACAACGC
G	GCGAATGCGTCCACAACG
C	GCGAATGCGTCCACAAC
A	GCGAATGCGTCCACAA
A	GCGAATGCGTCCACA
C	GCGAATGCGTCCAC
A	GCGAATGCGTCCA
C	GCGAATGCGTCC
C	GCGAATGCGTC
T	GCGAATGCGT
G	GCGAATGCG
C	GCGAATGC
G	GCGAATG
T	GCGAAT

Capillary electrophoresis;

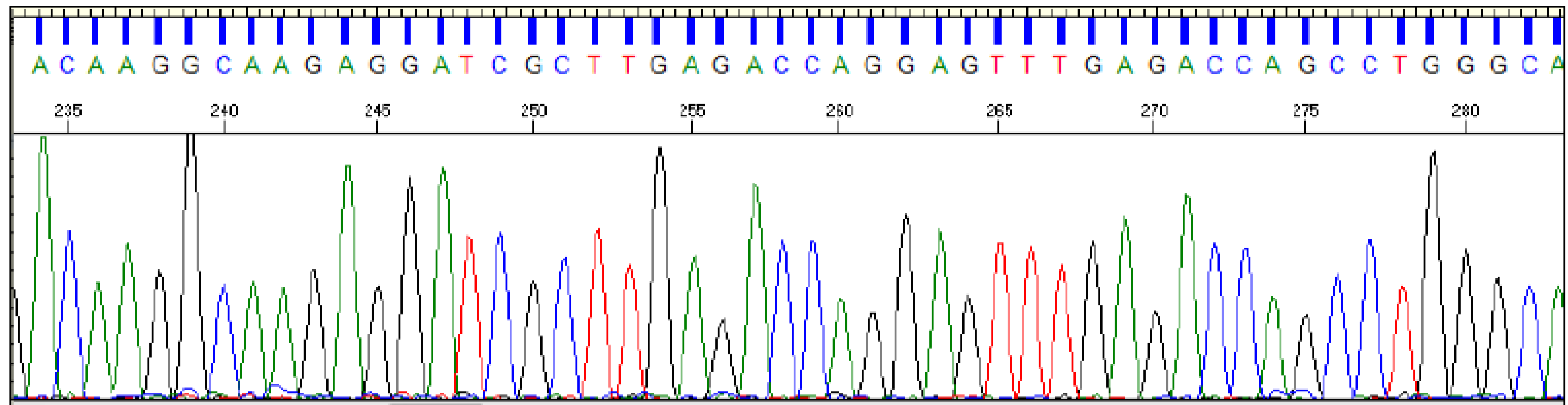
- 1 sequence reaction instead of 4
- ddNTPs labelled with different coloured florescent dye and detected by laser.
- Fragments sequenced as before and separated according to size by capillary electrophoresis



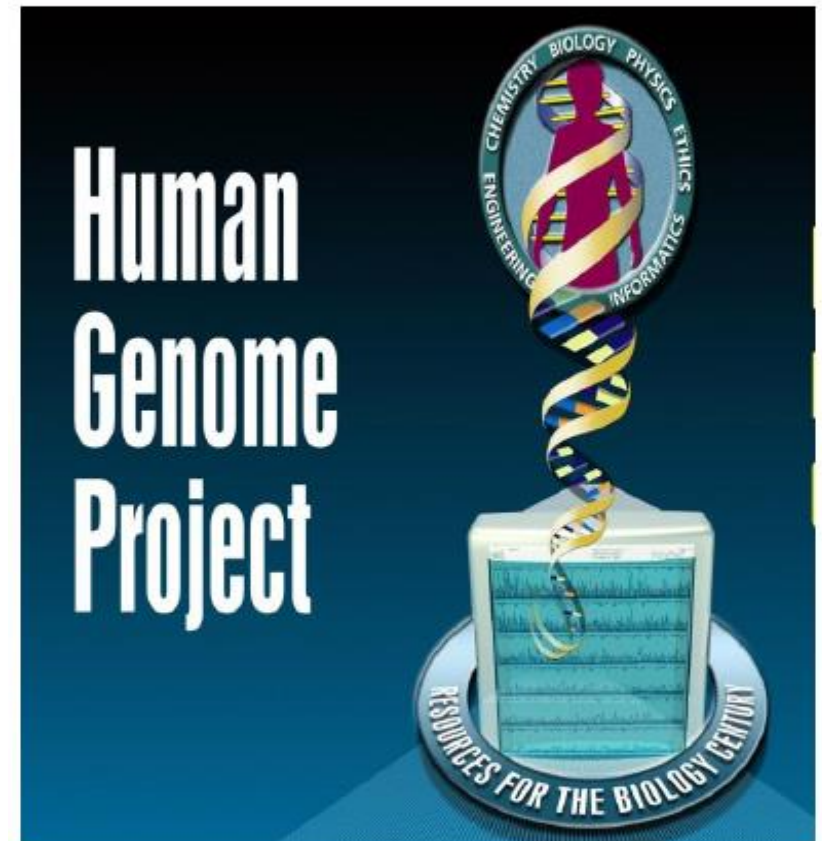
Capillary electrophoresis data analysis:

One lane of gel scanned from smallest to largest.
Red = Thymine, blue = Cytosine, black = Guanine, green = Adenine

This is just a fragment of the entire file, which would span around 900 or so nucleotides of accurate sequence



The human genome project:



- Started in 1990 in US
  1993 in UK
- The first ‘draft’ sequence published in 2001
- The first COMPLETE sequence released in 2003, further analysis published in 2011

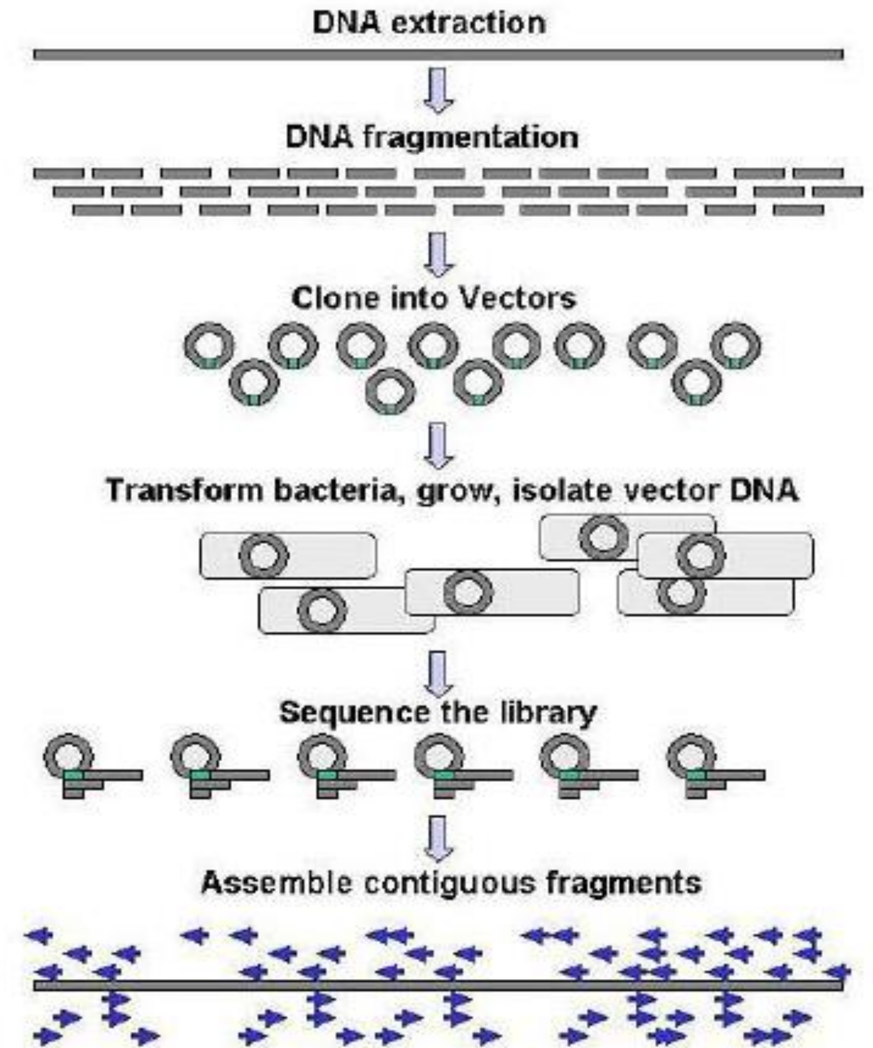
HOW was it achieved? ?

By breaking up the genome into small fragments (150kb)

Fragments cloned by inserting a vector (BACs) into a bacteria

The bacteria will replicate the fragments

DNA fragments subsequently sequenced as 'shotgun' called shotgun sequencing



Next generation sequencing (NGS):

- Also known as second generation or massively parallel sequencing.
- More than one technology exist (Iilumina HiSeq 2500, HiSeq 2000, Roche 454, SOLiD)

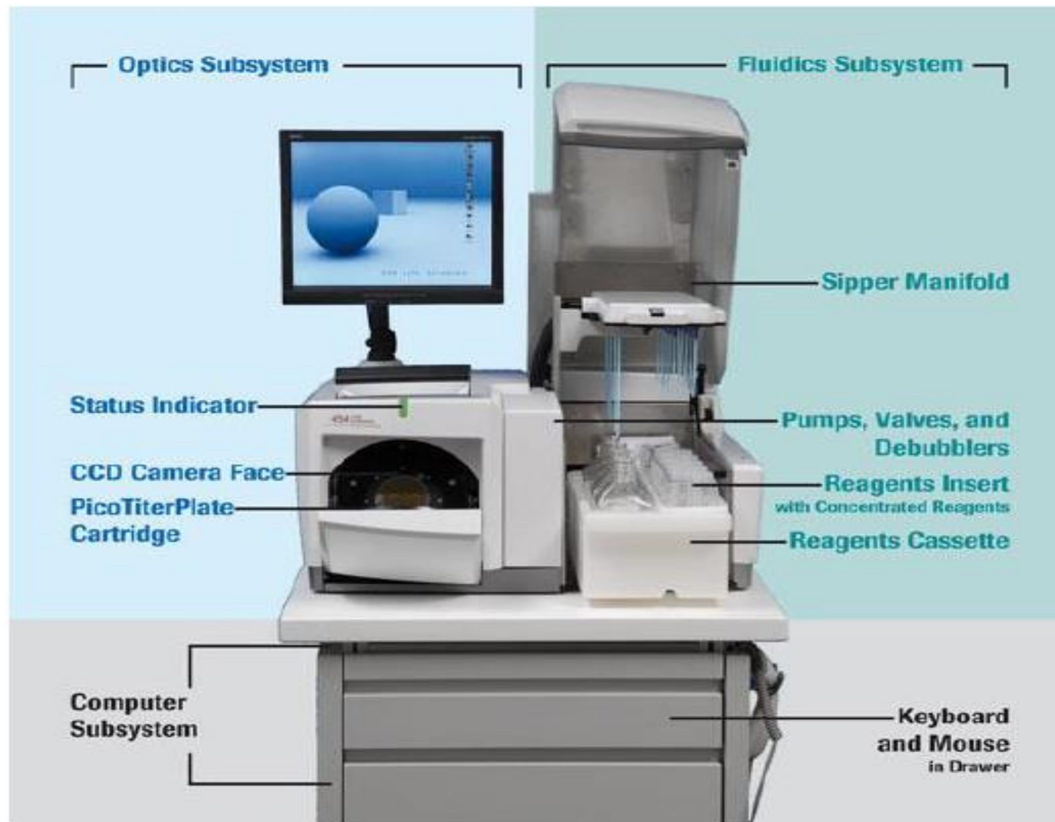
variable in:

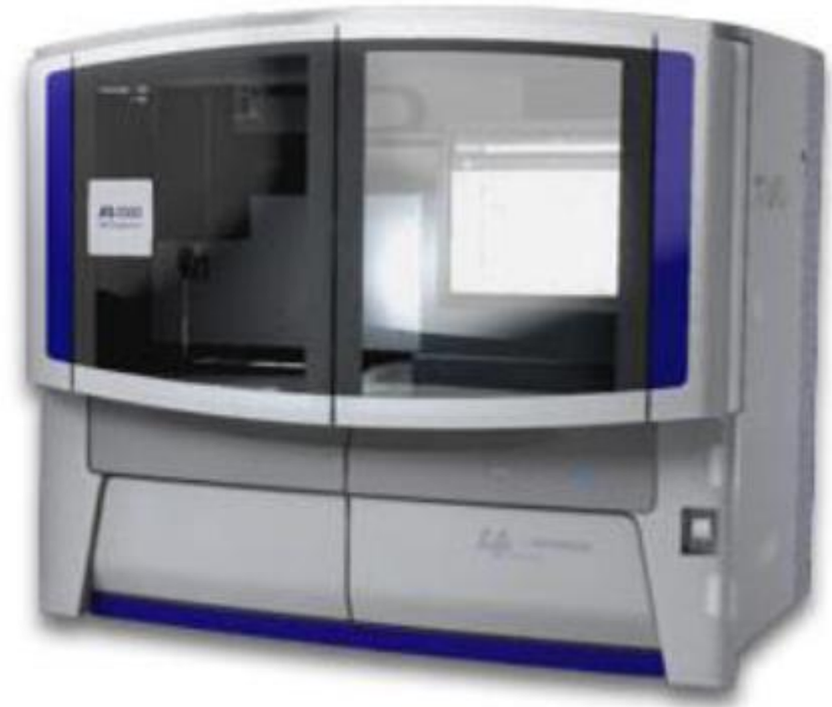
- Chemistry
- Throughput
- Cost

Similar in:

Preparation protocols

- Todays technology can sequence 5 whole human genomes in 11 days (Iilumina HiSeq2500).





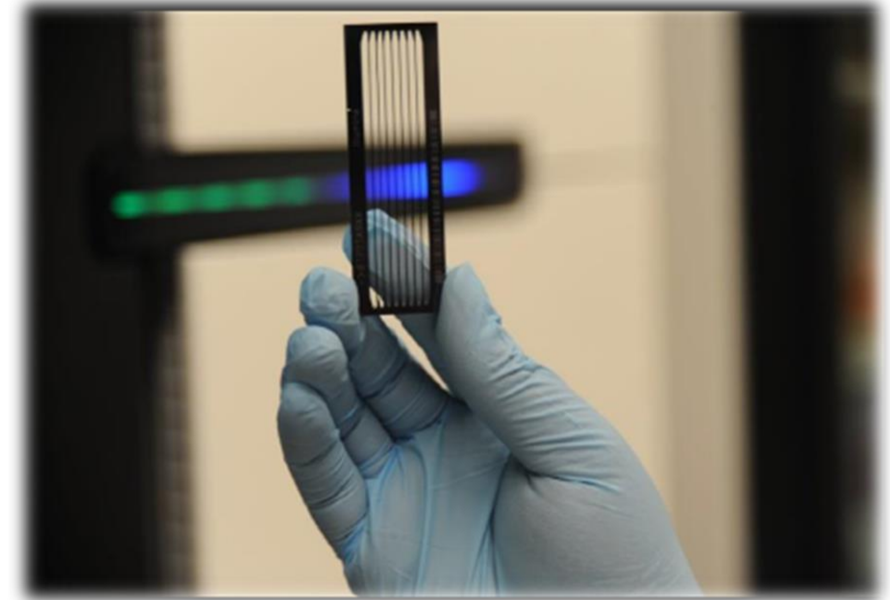
Roche 454

SOLiD

Sequencing by oligo nucleotide ligation and detection

Next generation sequencing
Illumina HiSeq2000 platform – how it works?

The flow cell is a solid substrate made of silica
And glass, 8 district channels carry millions of
Oligos with 2 types



FLOW CELL

3 main steps:

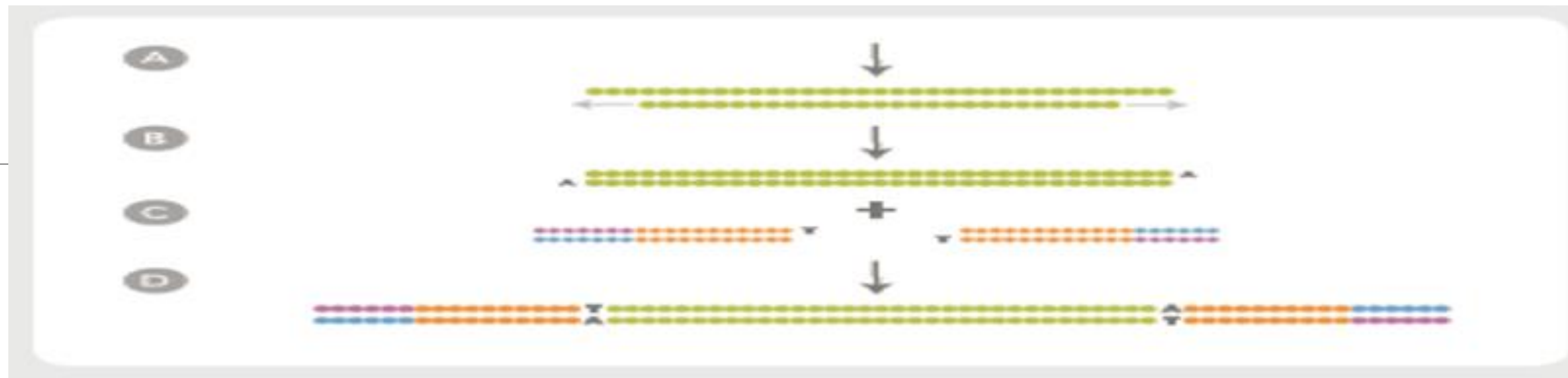
- Library preparation
- Cluster generation
- Sequencing by synthesis

1. Library preparation:

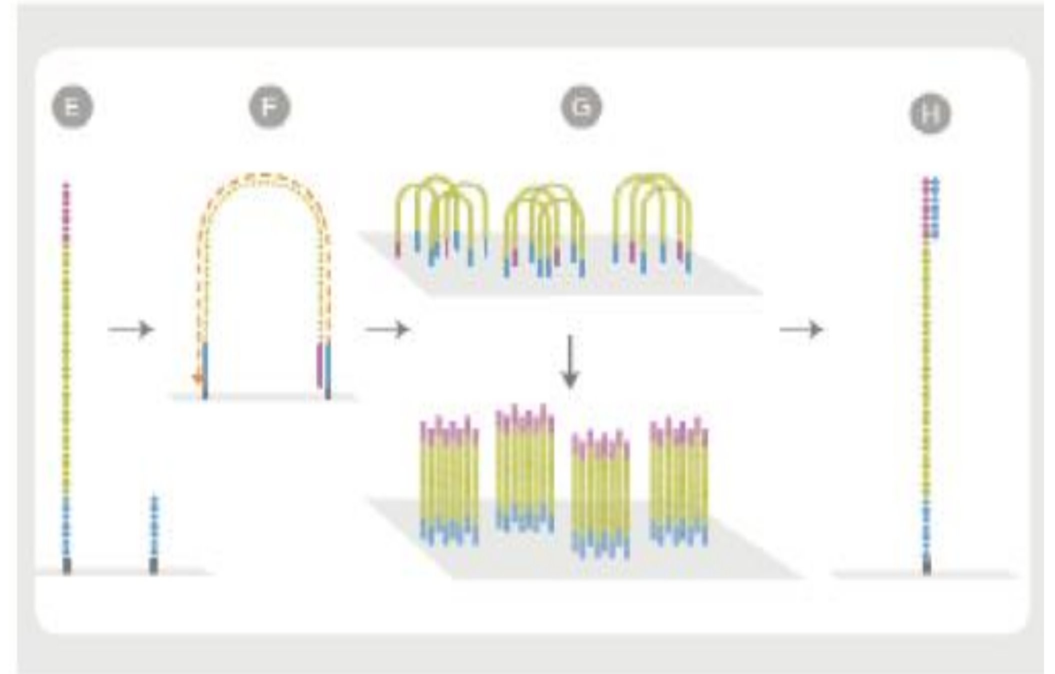
- DNA fragmented into smaller fragments (300 bases).

- 2 adapters sequence ligate to end of the fragments (adaptors complimentary to the oligos in the flow cell)

- Multiplexing



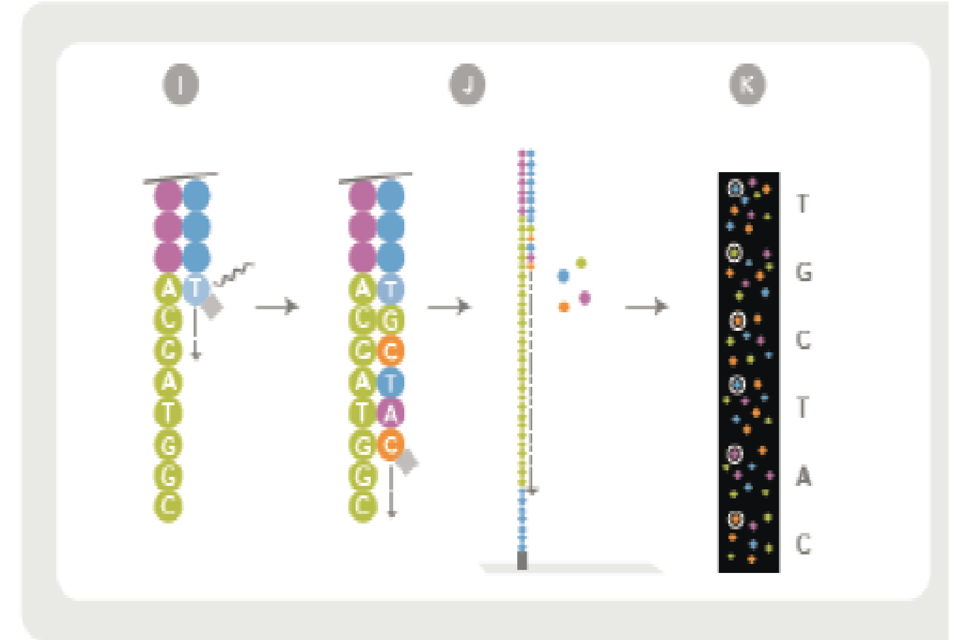
2. Cluster generation:

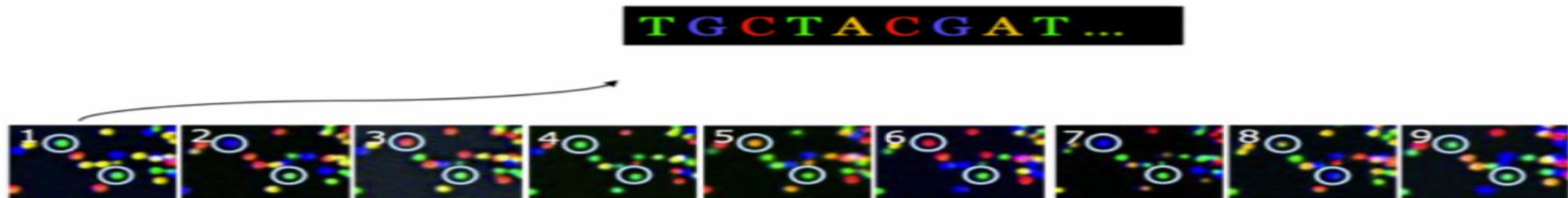


- DNA fragments denatured, attached to the flow cell
- double strand generated by polymerase
- the original strand (reverse) washed off
- the adapter of the new strand (forward) anneals to available oligo on Flow cell (bridge forming)
- Doubling strands of the bridge
- the process is repeated to create clusters
- Clusters denatured leaving only forward strands

3. Sequencing by synthesis:

- Sequencing primers and Fluorescently labelled chemically modified nucleotides are added.
- First base is synthesized, and detected via laser
- Chain is terminated bec of inactivation of 3' –OH
- Next 3; -OH de-blocking allow synthesis of the next Base.





Bioinformatics and Data analysis

- In 2 parts : read alignment and variant detection
- Illumina has its own GA pipeline program called CASAVA
- Many new software programs appear on a weekly basis

Third generation sequencing:

- ➢ A new generation of sequencing threatens to supersede NGS

- ➢ 2 main companies:
  - PacBio
  - Life Technologies

- ➢ At present time, most laboratory don’t have this technology available.

Uses of NGS in research:

- Cancer genomics
- Microbial genomics
- Forensic genomics
- Genetic disease

NGS in clinical diagnosis:

- Cancer
- Rare mendolian disease
- From single gene to multiable gene testing
- PGD

Further reading:

- International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. Nature 431:931-945
- Harismendy, O., Ng, P.C., Strausberg, R. L., Wang, X.W., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S. S., Topol, E.J., Levy, S., and Frazer, K.A. (2009). Evaluation of next generation sequencing platforms for population targeted sequencing studies. Genome Biology. Vol. 10: R32
- Munroe, D.J., and Harris, T.J.R (2010). Third-generation sequencing fireworks at Marco Island. Nature Biotechnology 28, 426-428
- Strannenheim, H. and Lundeberg, J (2012) Stepping stones in DNA sequencing. Biotechnology Journal. Vol 7 (9): 1063-1073
- Tucker, T., Marra, M., and Friedman, J.M. (2009). Massively parallel sequencing: The next big thing in genetic medicine. Am J Hum Genetics. Vol 85 142-154

Questions ??