

PŘEHLED SEKVENAČNÍCH METOD

Letní škola bioinformatiky 2014, Brno

Ing.Matej Lexa, Phd (FI MU Brno)

CO JE TO SEKVENACE A CO SE BUDE SEKVENOVAT?

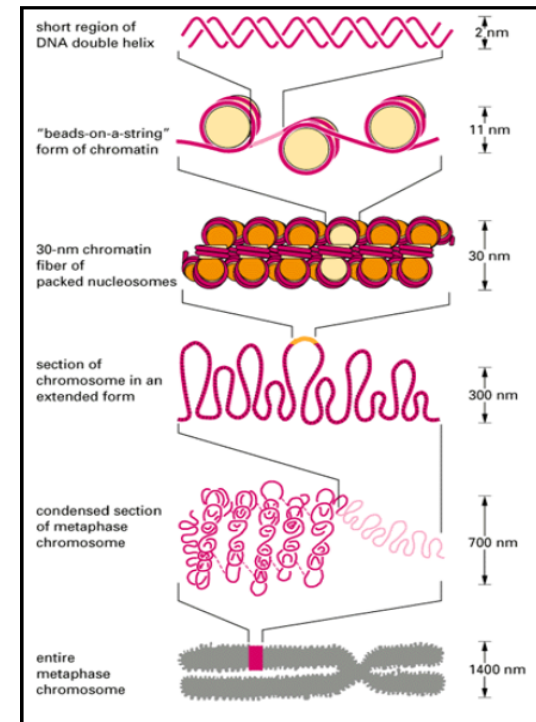
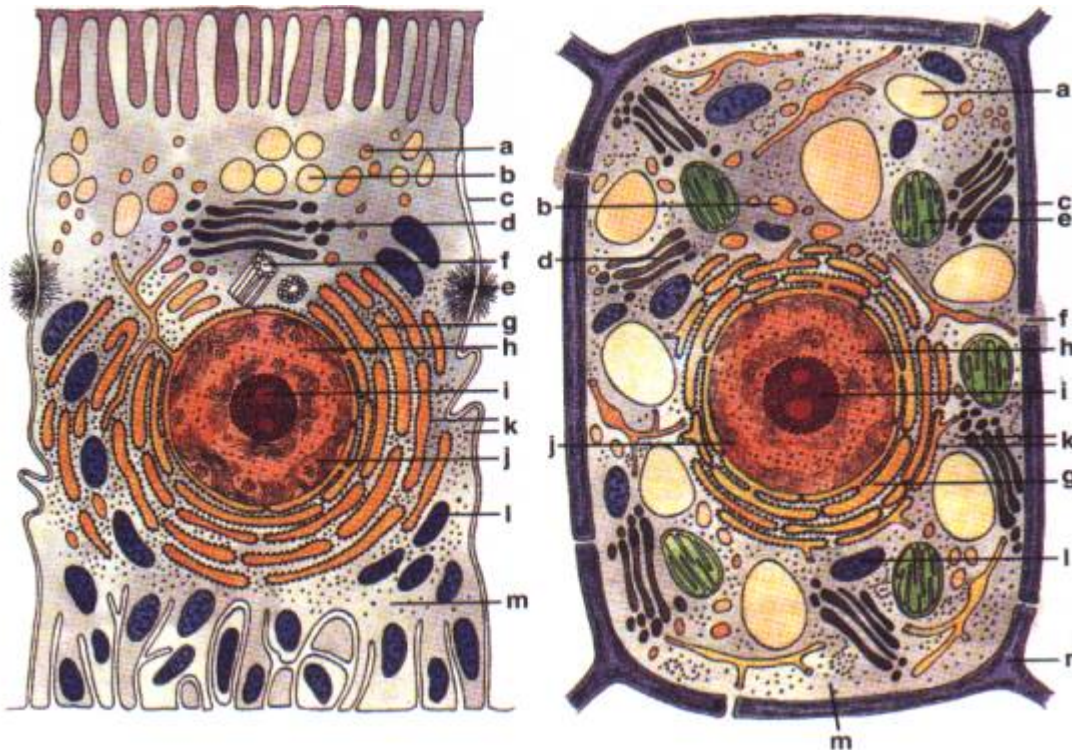
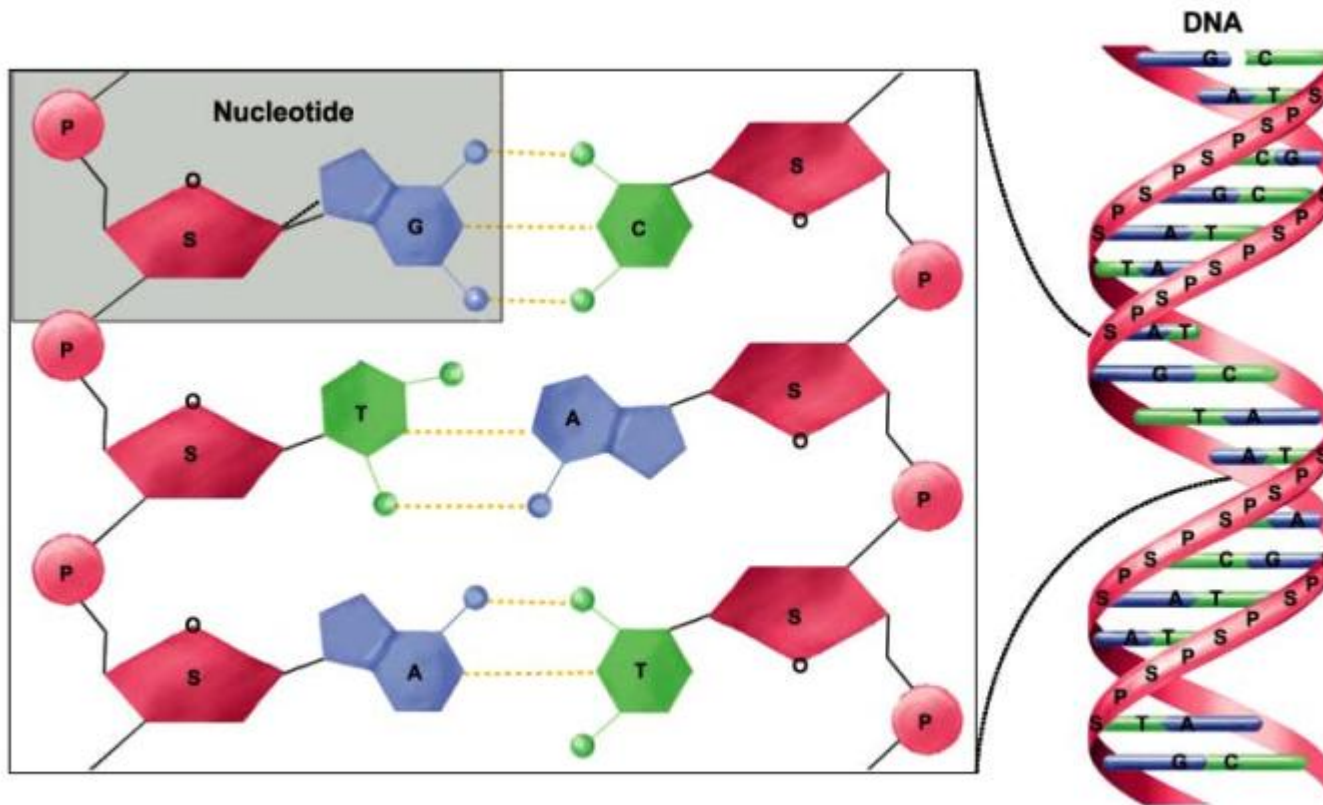


Figure 8-30. Model of chromatin packing. Alberts „Molecular Biology of the Cell”

POŘADÍ NUKLEOTIDU V DNA



SEKVENOVÁNÍ DNA

od manuálních metod po plně automatizované

1. generace (od cca 1976):

Maxam-Gilbert (čtyři štěpící reakce – G, AG, C, TC)
Sangerova metoda (terminace syntézy druhého vlákna)



2. generace (next-gen; NGS; od cca 2000-2005):

Solexa/Illumina (polymerace fluorescenčně značených oligonukleotidů)
454/Roche (pyrosekvenace, PP enzymaticky vázán na emisi světla)
SOLiD/LifeTechnologies (ligace fluorescenčně znač. oligonukleotidů)
Polonator (ligace fluorescenčně znač. oligonukleotidů)
Complete Genomics (nepoužívá PCR, ligace na shluky ssDNA)



Sekvenování jednotlivých molekul:

Helicos (optika)

SMRT/Pacific Biosciences (nanopovrch s ukotvenou polymerázou)

IonTorrent (polovodičové sekvenování, místo světla protony)

Sekvenování na membránách:

Oxford Nanopore

NABSys

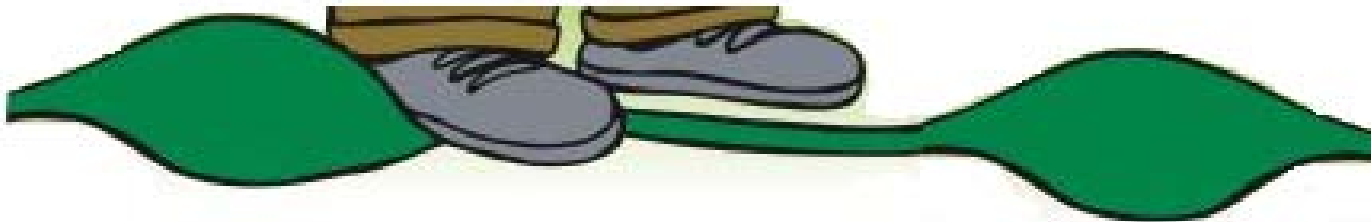
NobleGen



NOBLEGEN



KRITICKÁ FÁZE ANALÝZ 1. a 2. GENERACE



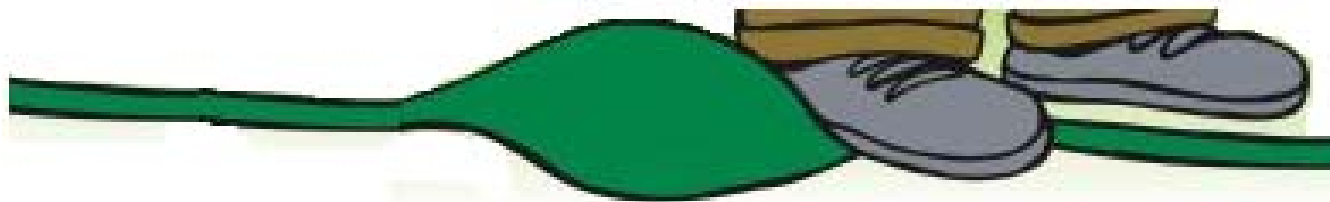
Příprava
knihovny



Sekvenace



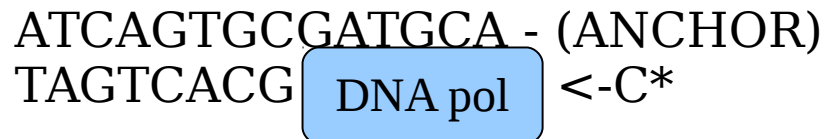
Analýza dat



PRINCIPY SEKVENACE DNA

sekvenčně-specifické štěpení DNA

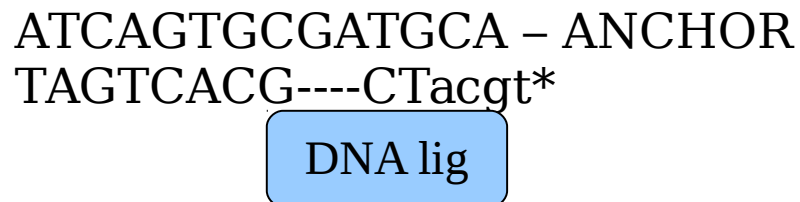
polymerace:



hybridizace:

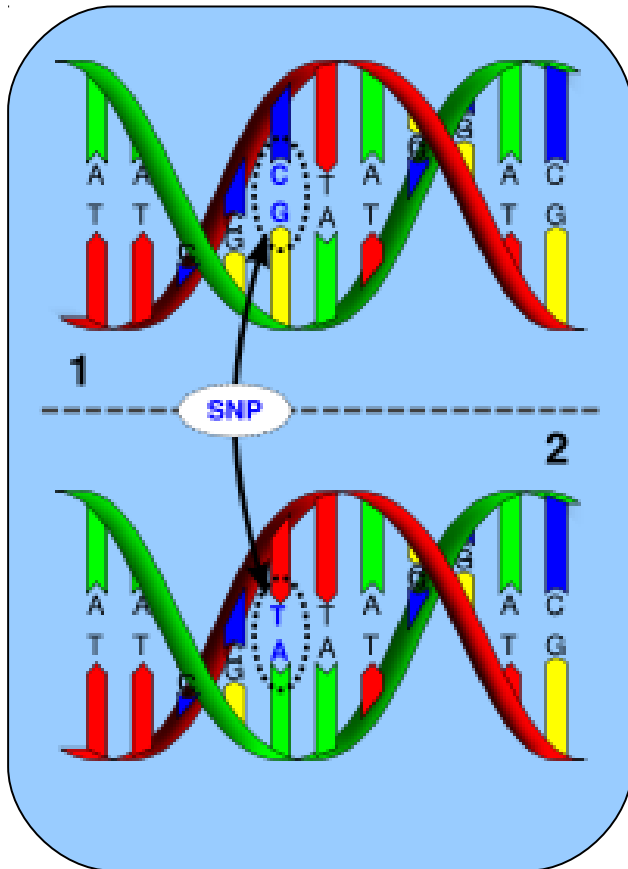


ligace:



nanopory a průchod membránou

DUVODY SEKVENACE



- de novo sekvenace
 - organismy
 - populace (metagenomika)
- resekvenování
- detekce polymorfismů
 - SNP
 - rekombinace
- zjištění stavu metylace
- měření exprese
(jako náhrada technik založených na hybridizaci)

DRUHY SEKVENACE

dle způsobu přípravy vzorků a použitých reaktantů

DNA-seq

ChIP-seq (imunoprecipitace chromatinu)

RNA-seq

sRNA-seq

Bis-seq, BisChIP-seq (metylační analýza)

CLIP-seq (“cross-linking immunoprecipitation”; RNA-prot)

Ig-seq,... (viz *Seq @ <http://liorpachter.wordpress.com/seq/>)

Klíčové kroky a parametry sekvenování DNA

- izolace DNA
- fragmentace DNA
- ligace adaptérů
- namnožení sekvencí
- typ podložky nebo média
- typ enzymů (pol, lig)
- typ detekce (světlo, proud)

Shotgun genome sequencing

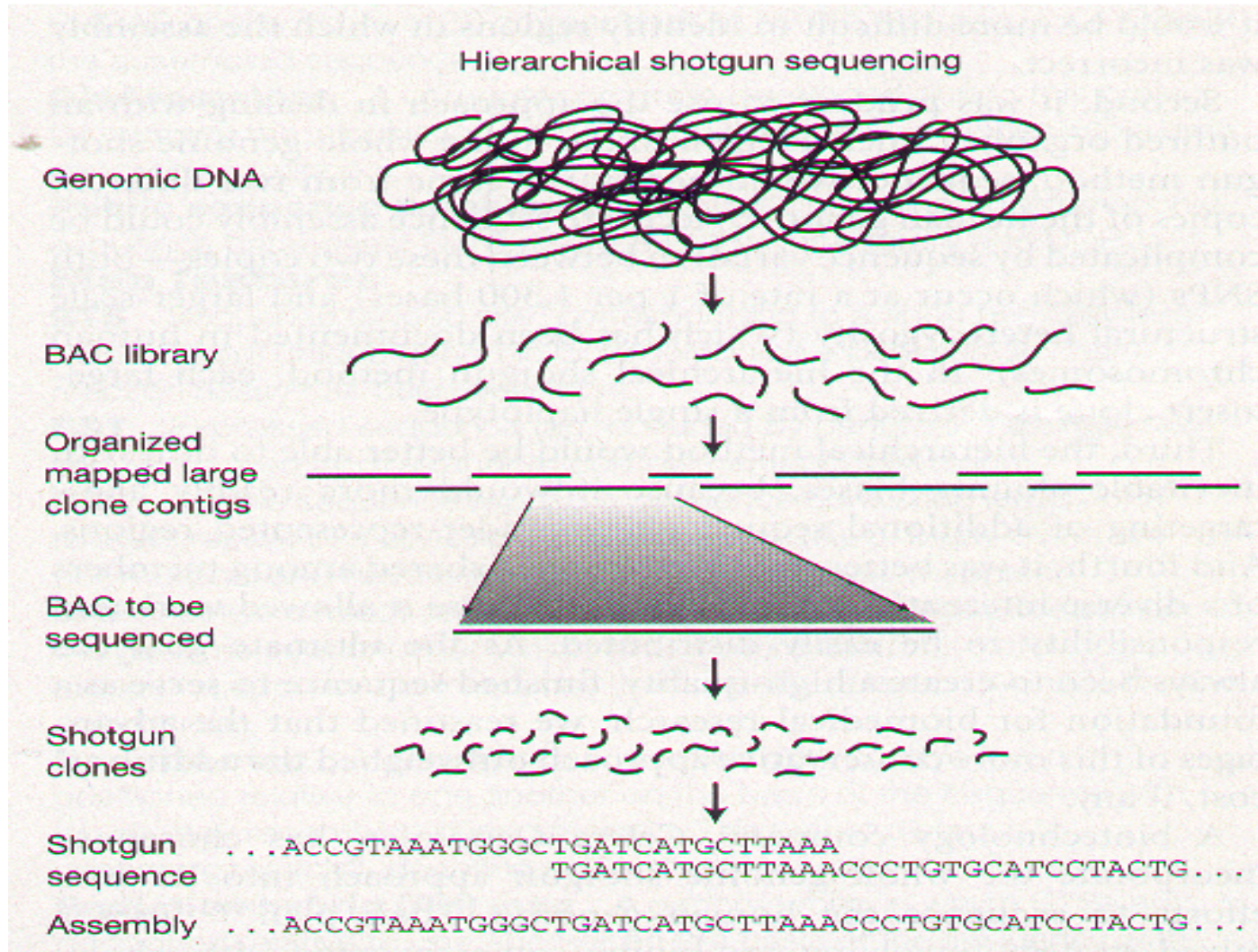
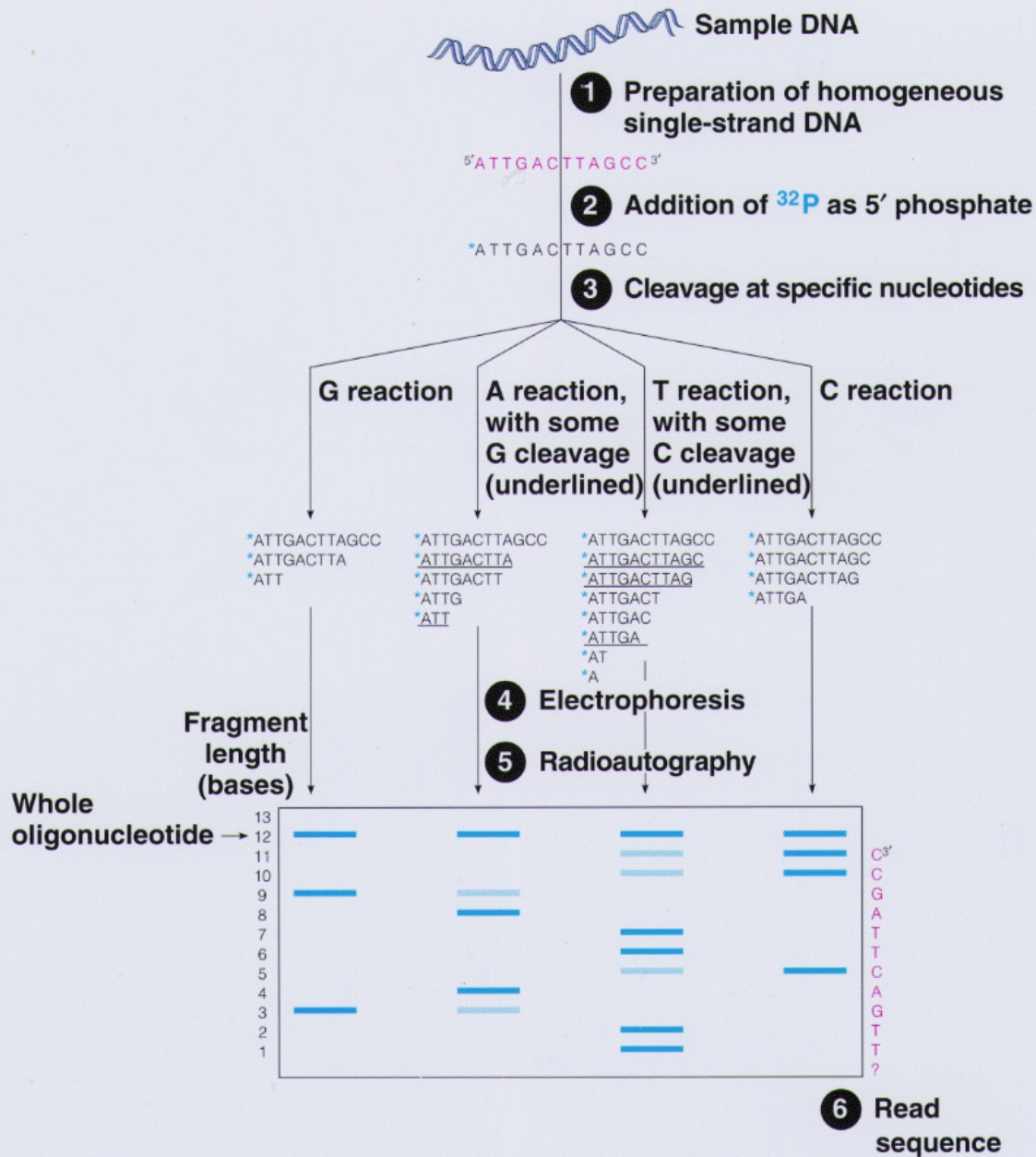
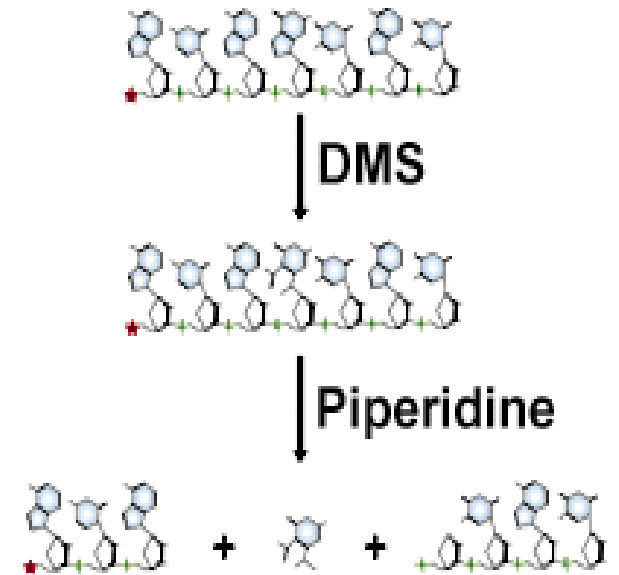
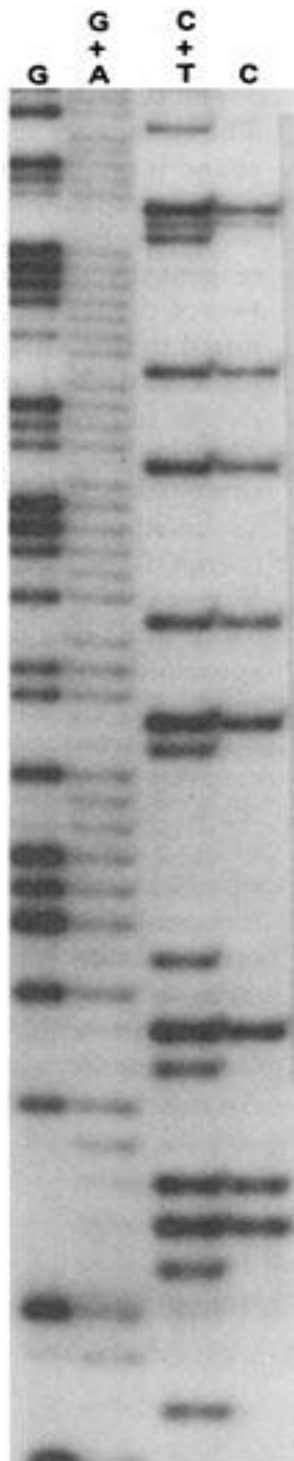


Figure 4A.4 Sequencing an oligonucleotide by the Maxam-Gilbert method



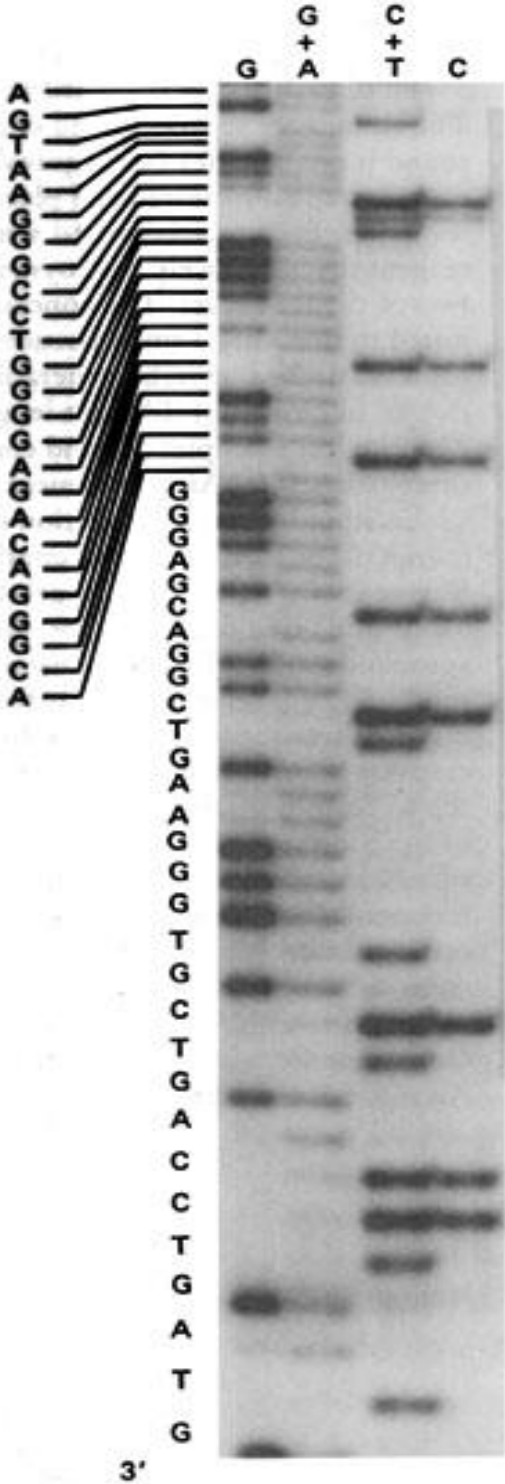
Maxam-Gilbert



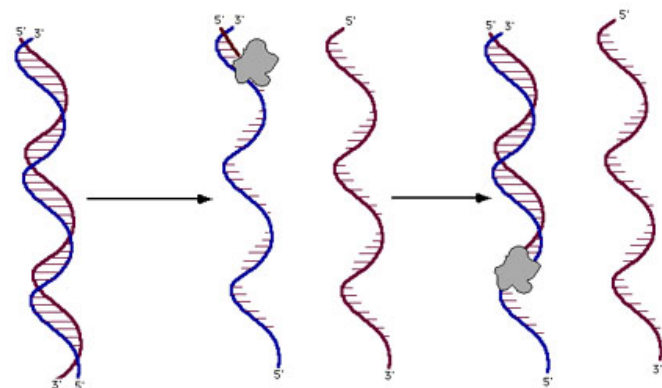


Maxam-Gilbert

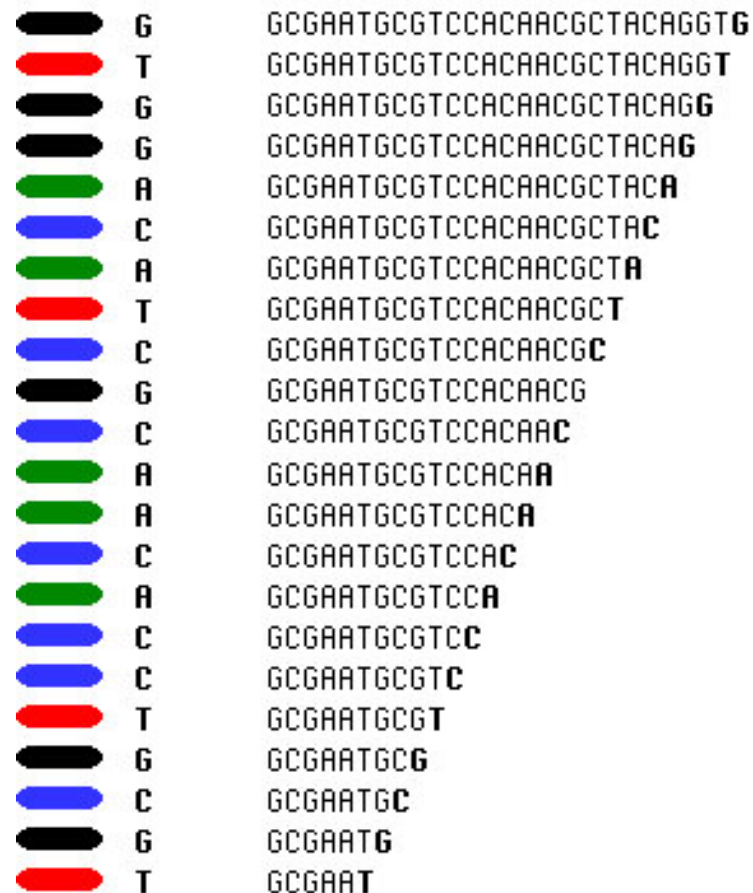
Maxam-Gilbert



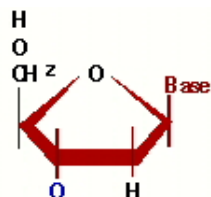
Sangerova metoda (terminace dideoxynukleotidy)



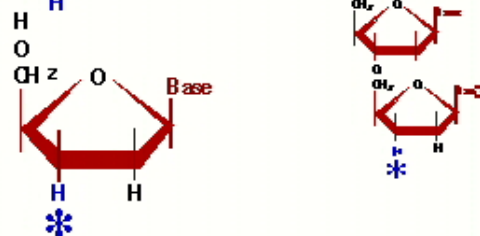
Gel:



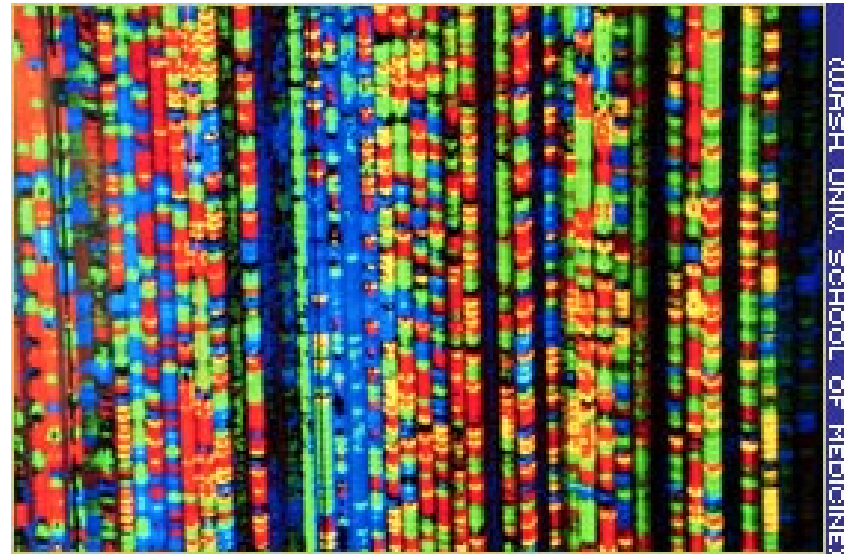
Normal nucleotides:



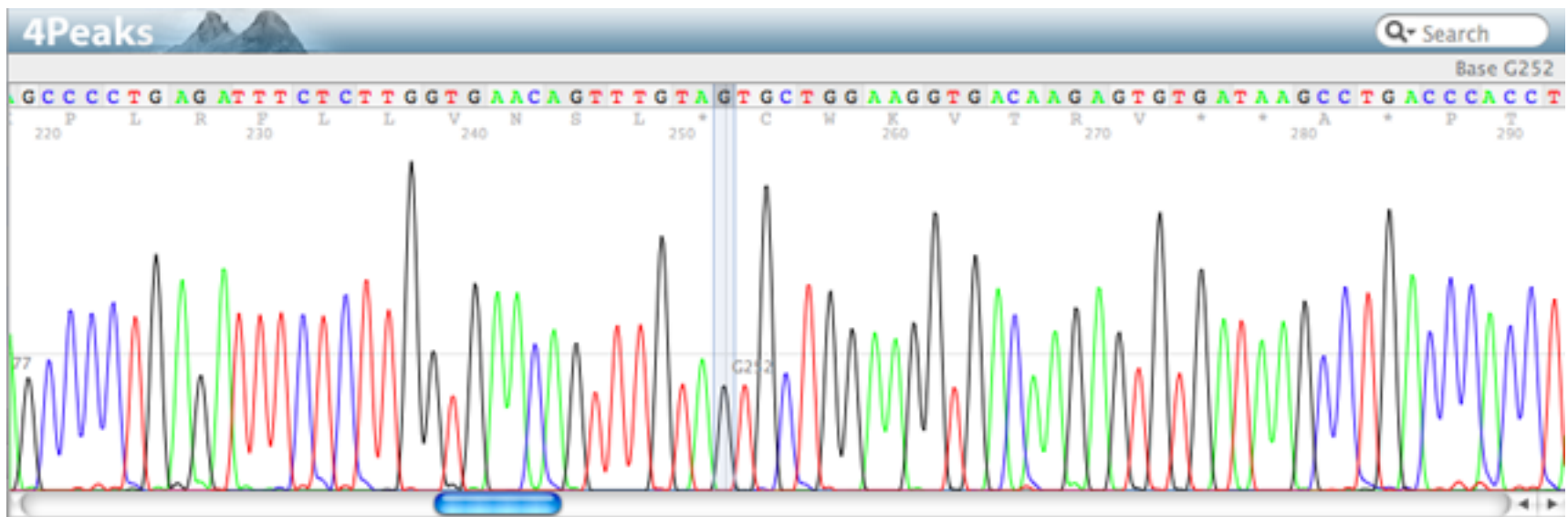
Dideoxy Chain Terminators:



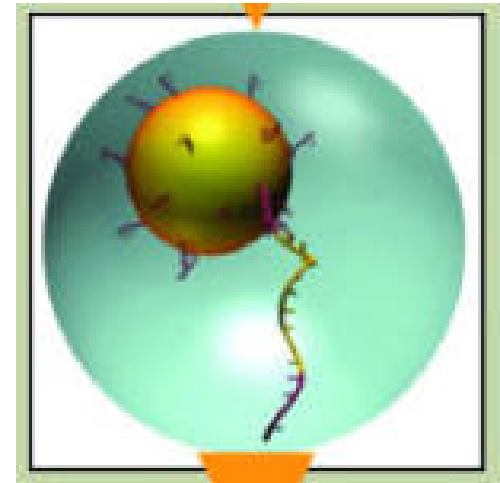
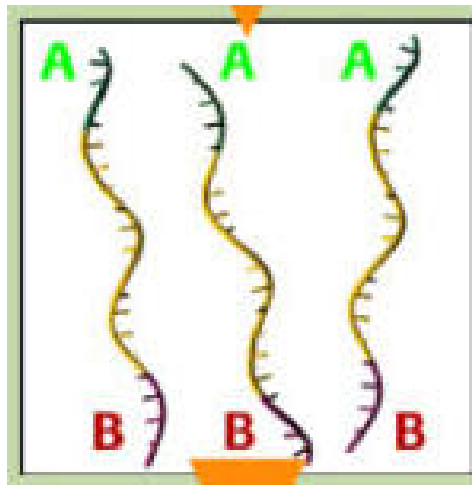
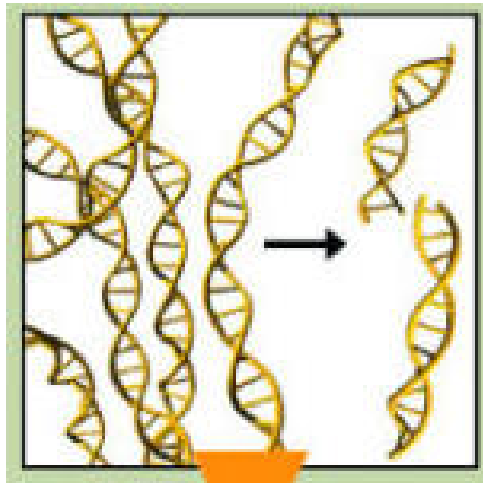
Sangerova metoda



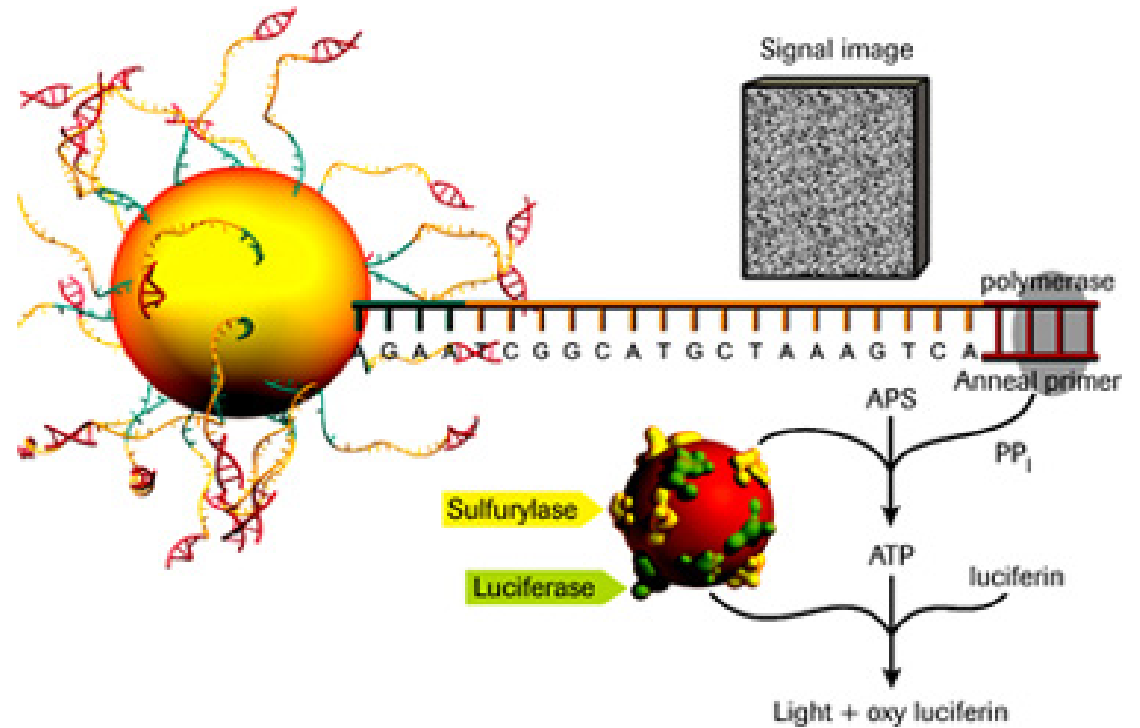
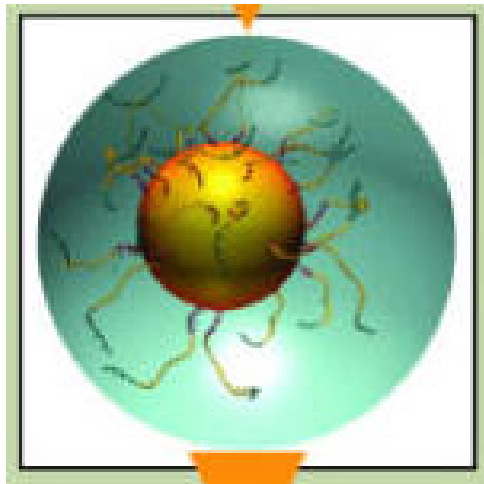
Sangerova metoda



454 (pyrosekvenování)



454 (pyrosekvenování)

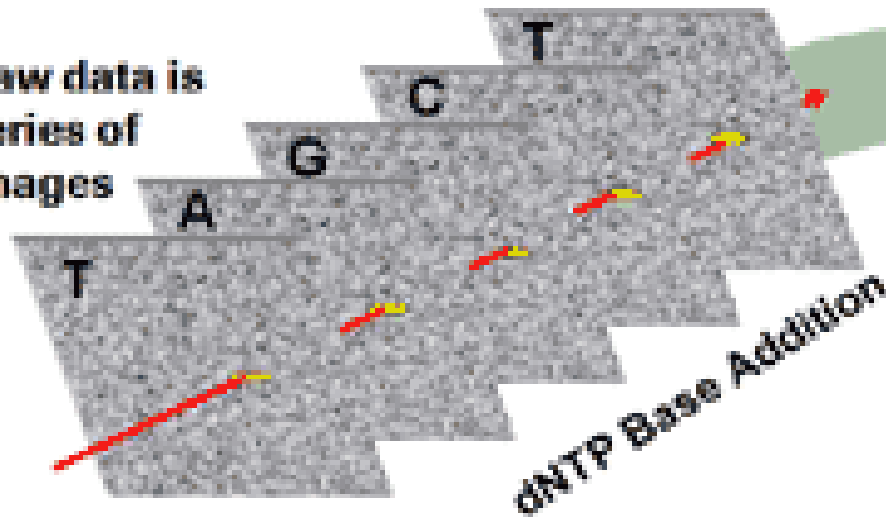


454 (pyrosekvenování)

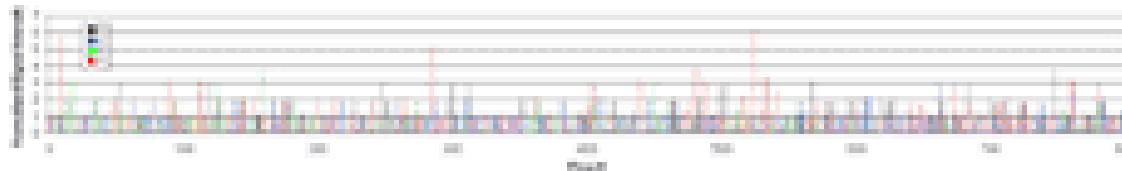
GS FLX Data

Image Processing Overview

1. Raw data is series of images

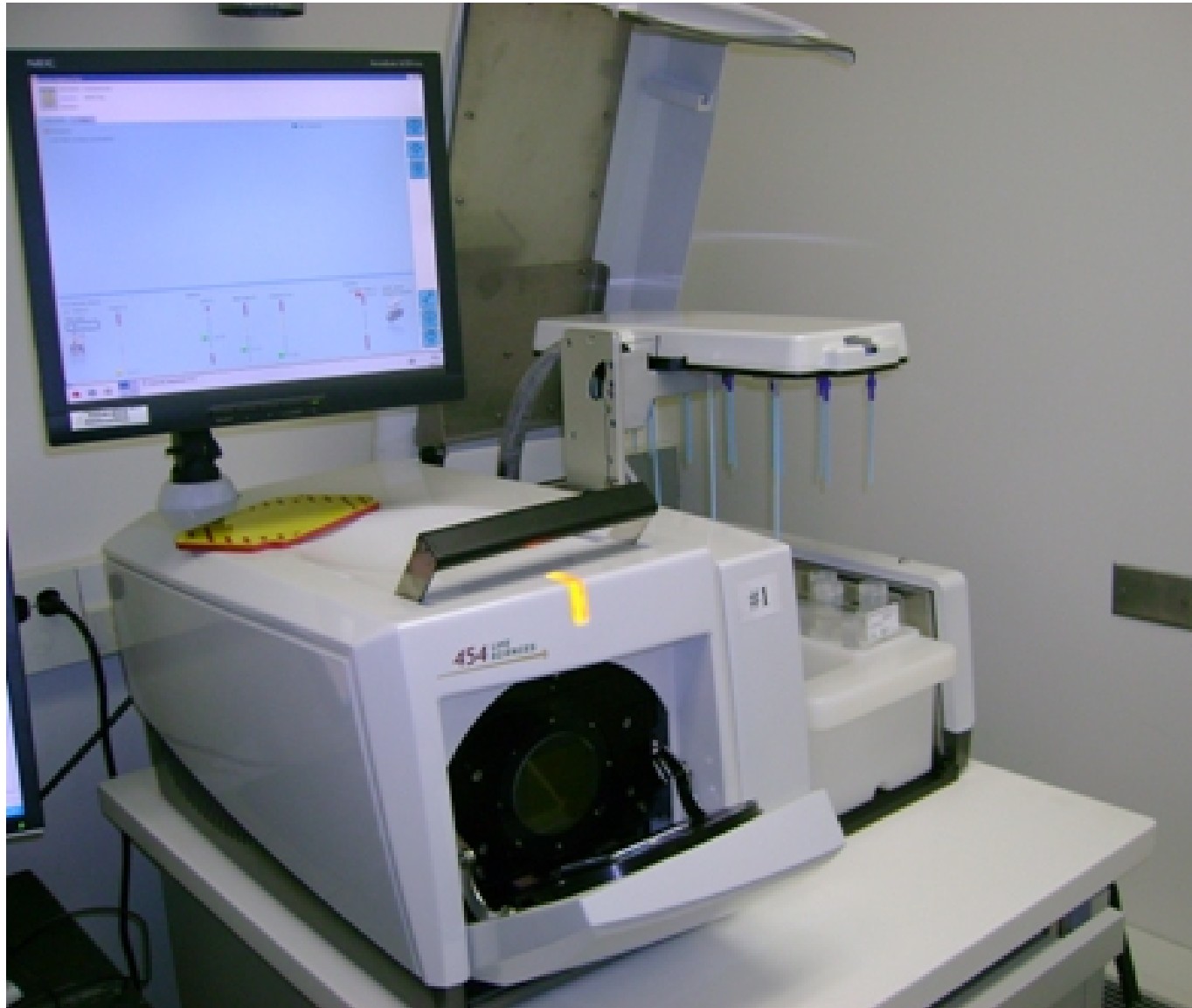


2. Each well's data extracted, quantified and normalized

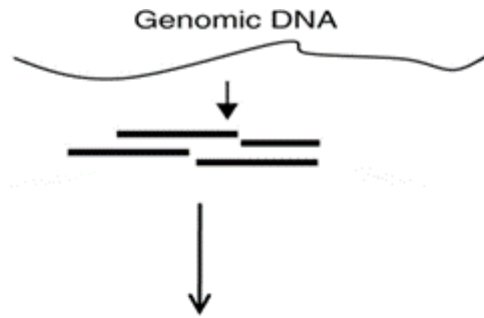


3. Read data converted into "flowgrams"

Roche 454 GS FLX

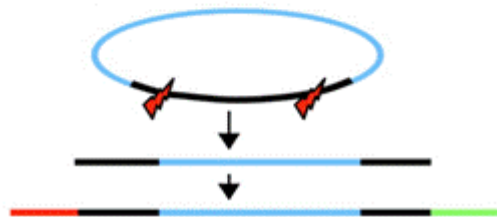


POLONATOR

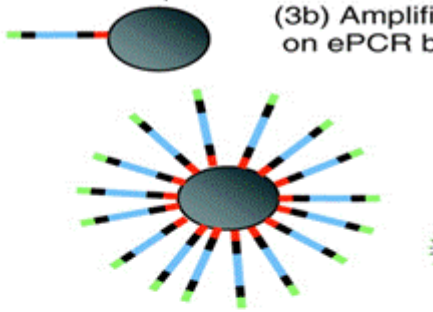


(1) Sheering to small fragments

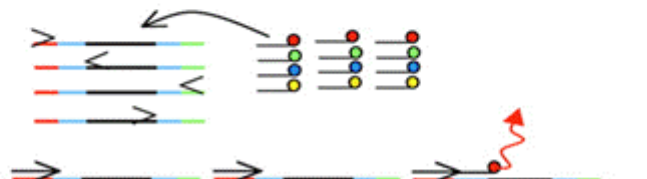
(2b) Circularization and linker attachment



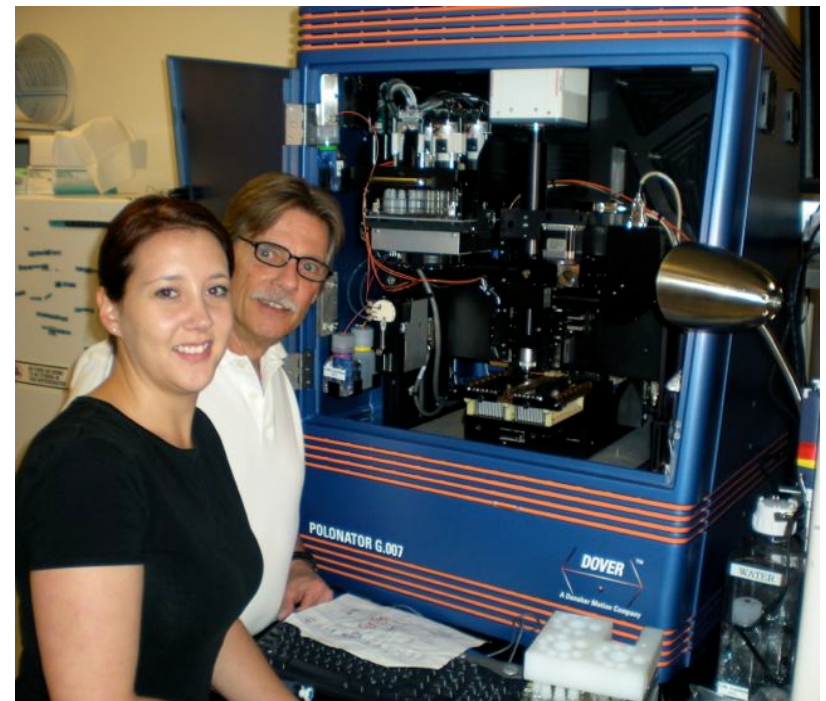
(3b) Amplification on ePCR beads



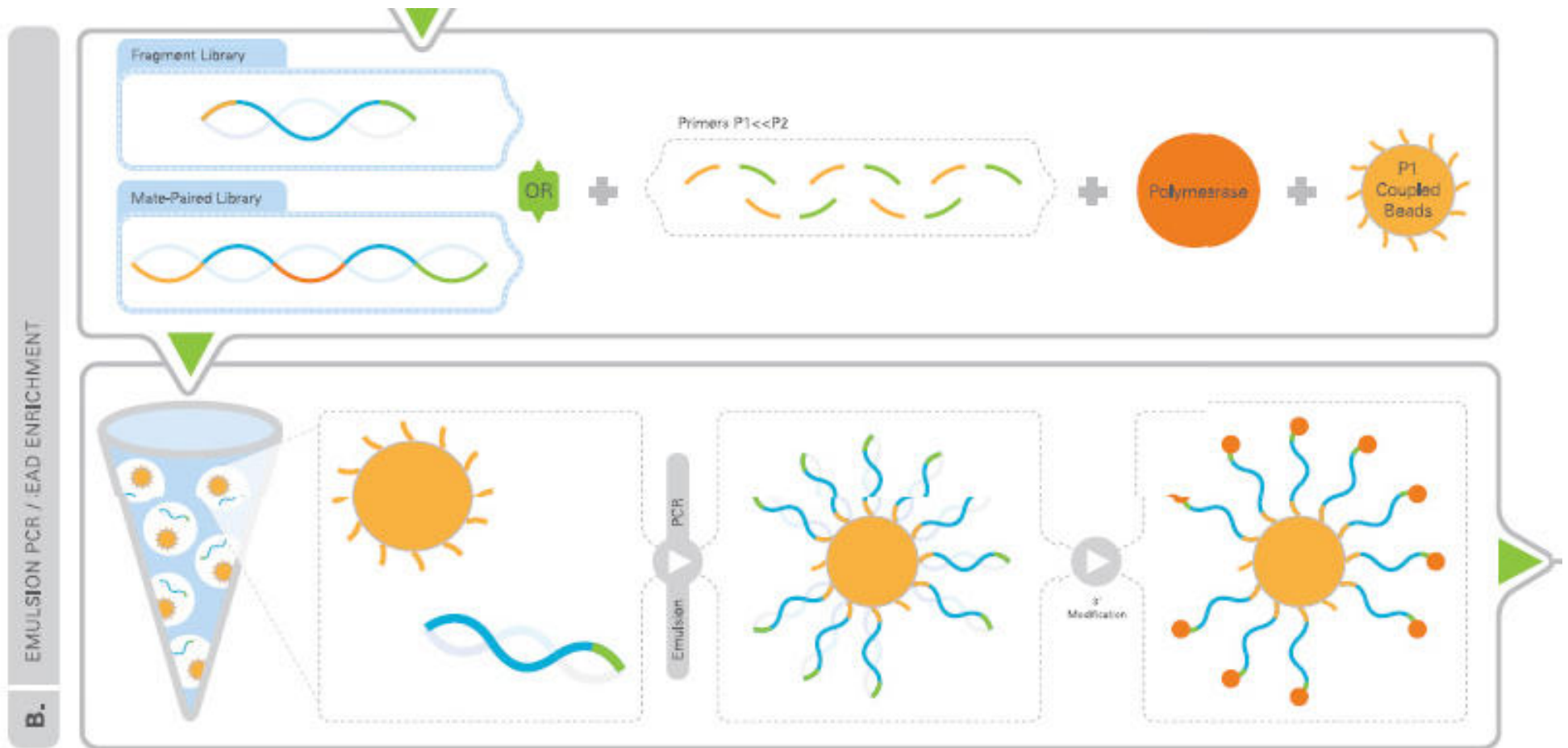
(4b) Beads immobilized in a monolayer in an acrylamide matrix



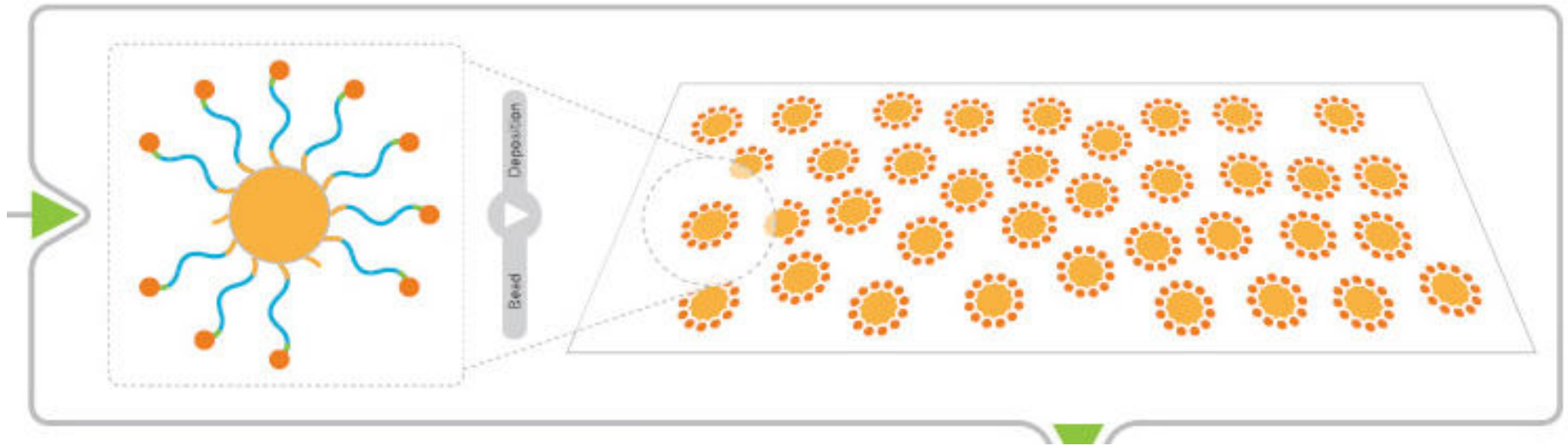
(5b) Ligation with oligo pools added in cycles and four-color imaging



SOLiD (sekvenace ligací, “polony” = polymerase colony)

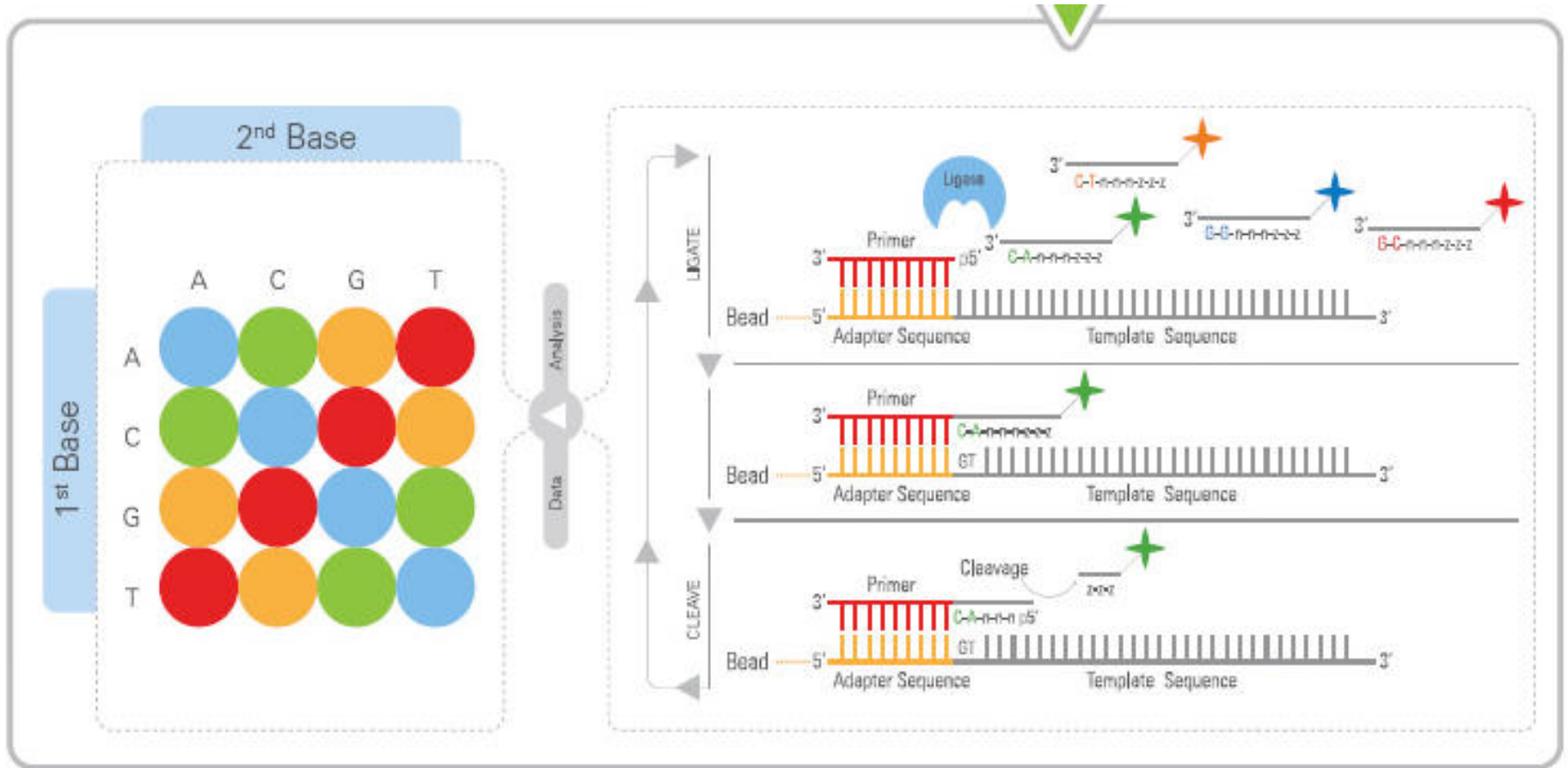


SOLiD



C.
BEAD DEPOSITION

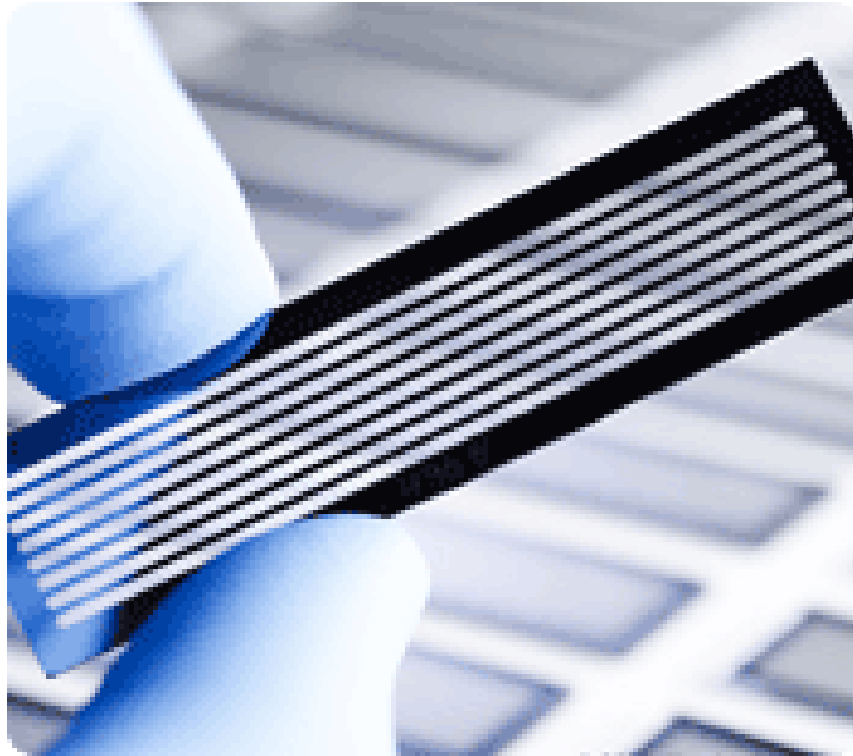
SOLiD



SOLiD 3 Plus

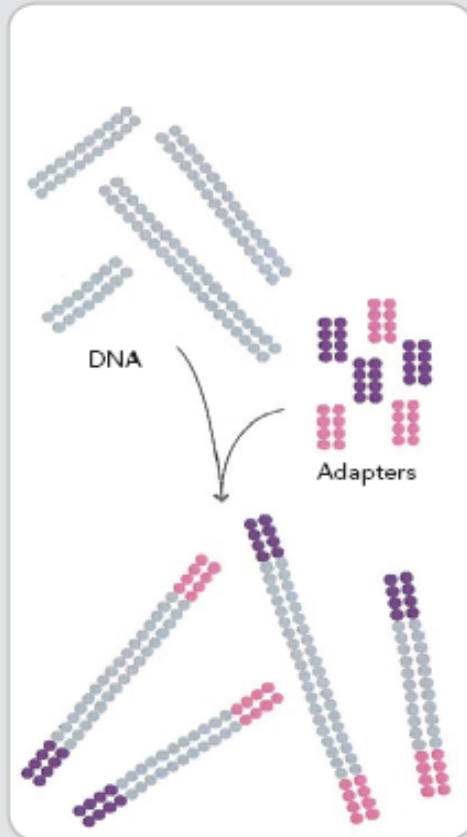


Illumina



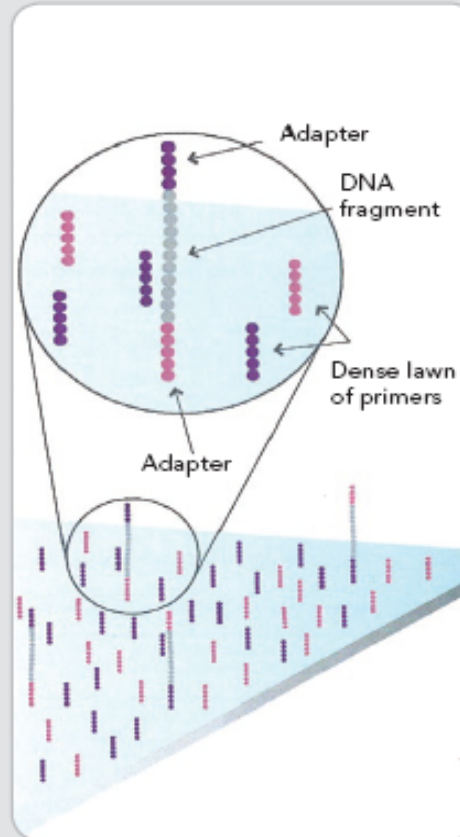
Illumina (“bridge amplification”)

1. PREPARE GENOMIC DNA SAMPLE



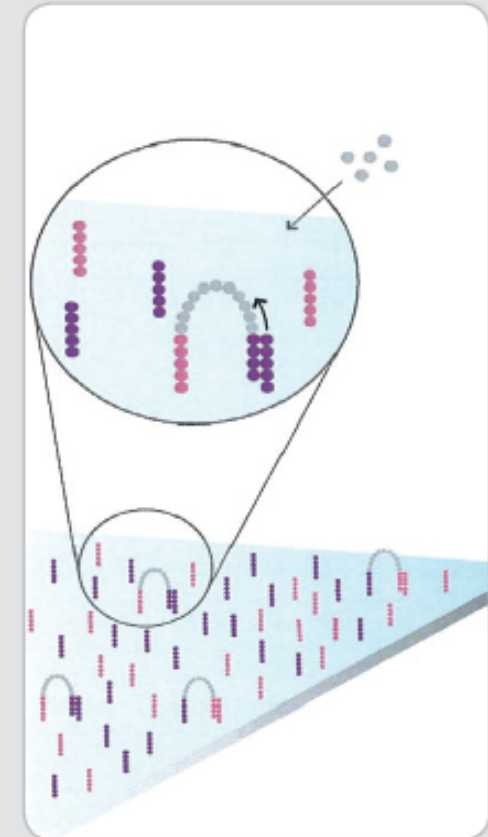
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE



Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

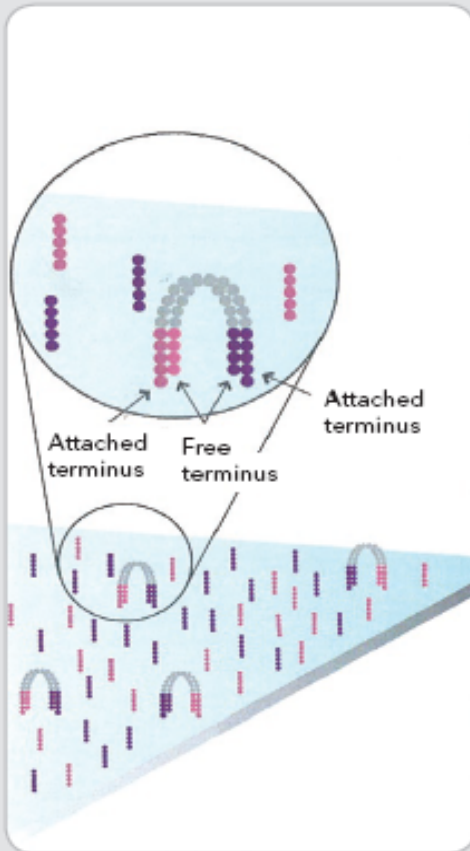
3. BRIDGE AMPLIFICATION



Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

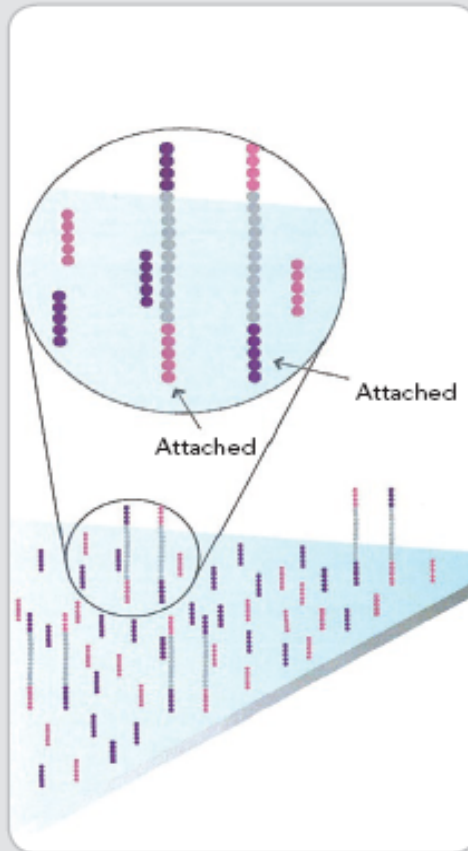
Illumina (“bridge amplification”)

4. FRAGMENTS BECOME DOUBLE-STRANDED



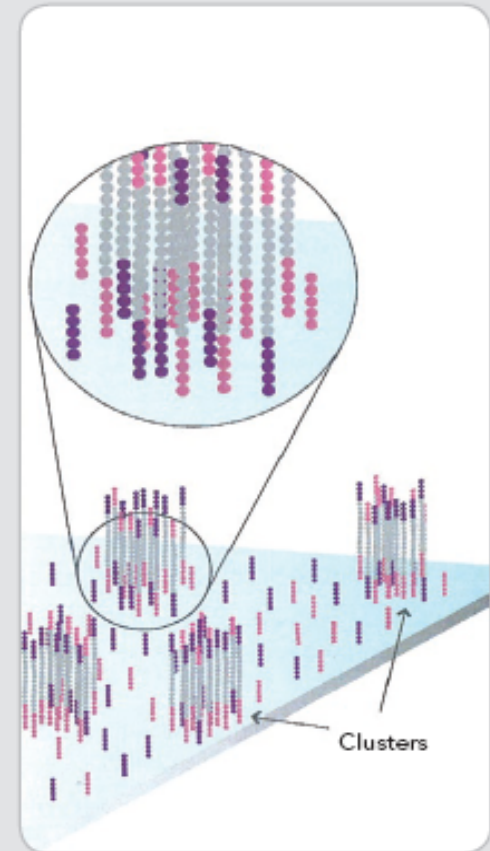
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES



Denaturation leaves single-stranded templates anchored to the substrate.

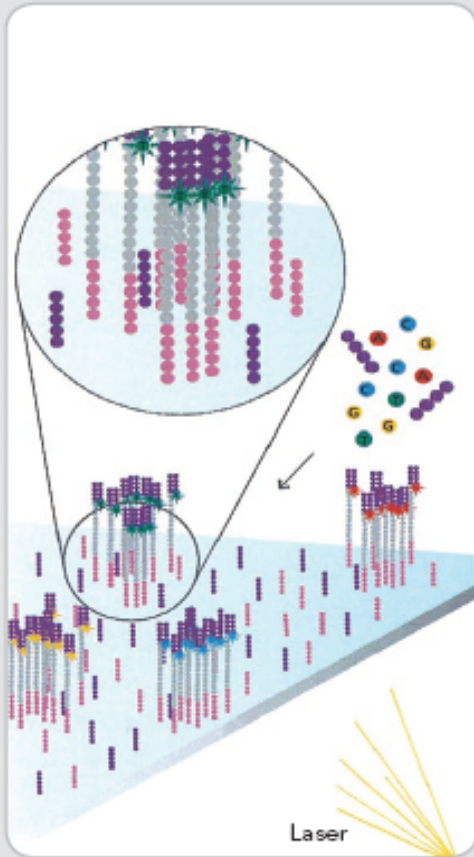
6. COMPLETE AMPLIFICATION



Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

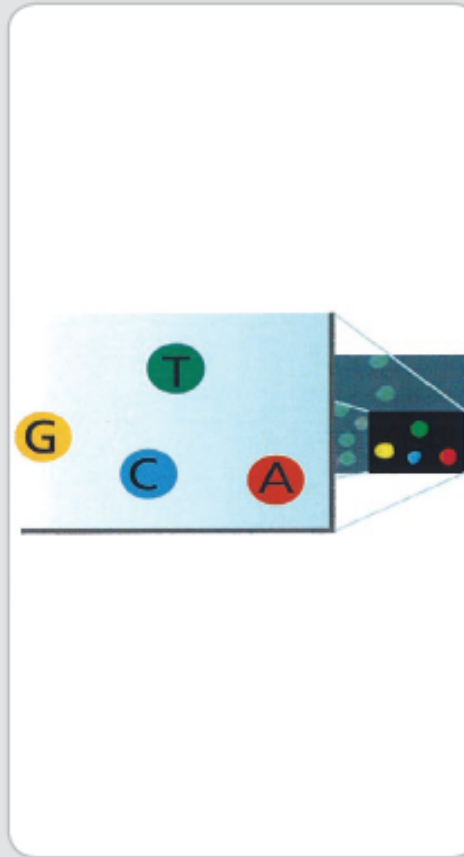
Illumina

7. DETERMINE FIRST BASE



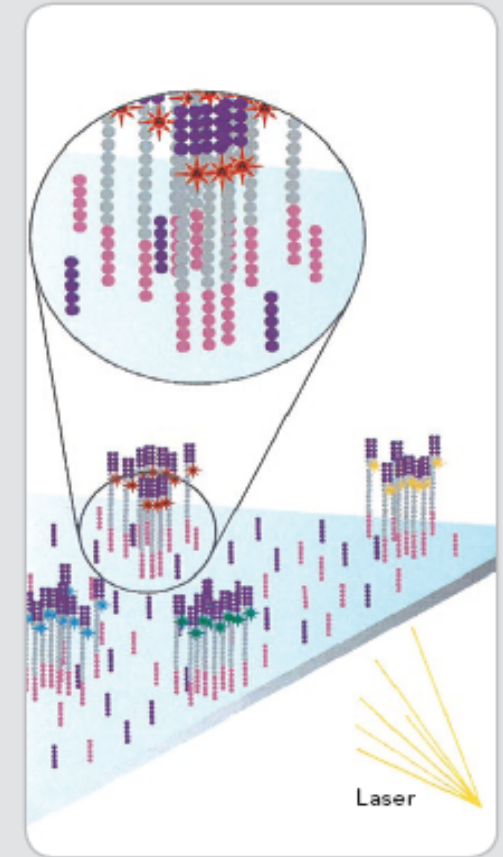
The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

8. IMAGE FIRST BASE



After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

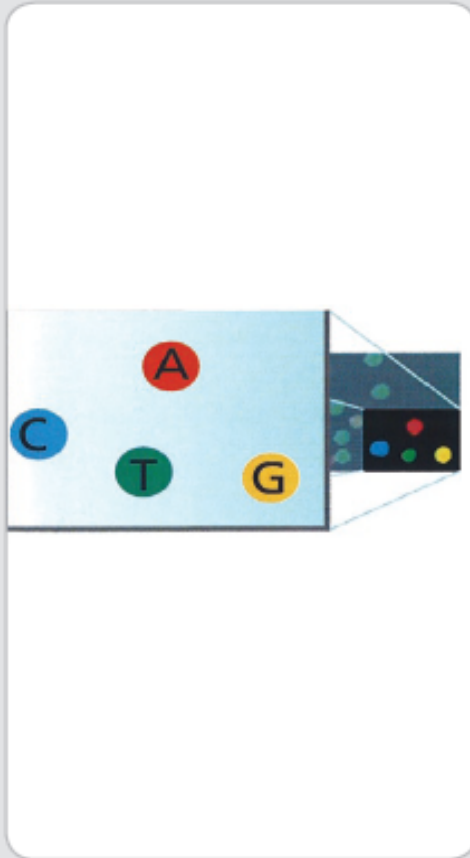
9. DETERMINE SECOND BASE



The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

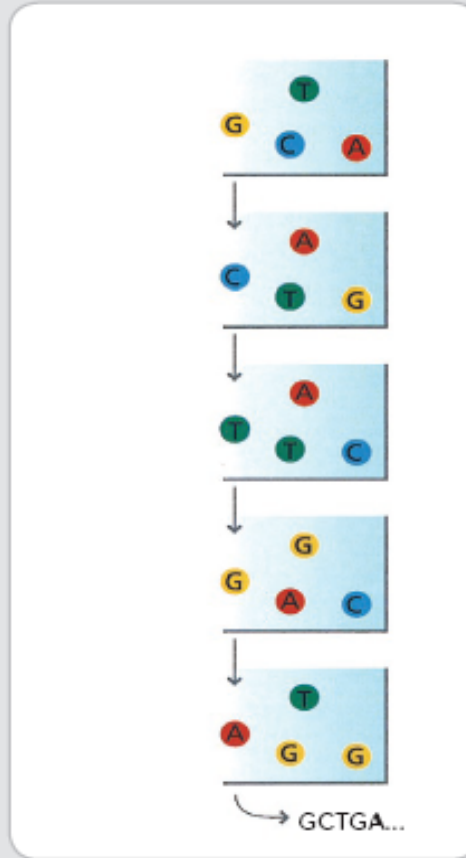
Illumina

10. IMAGE SECOND CHEMISTRY CYCLE



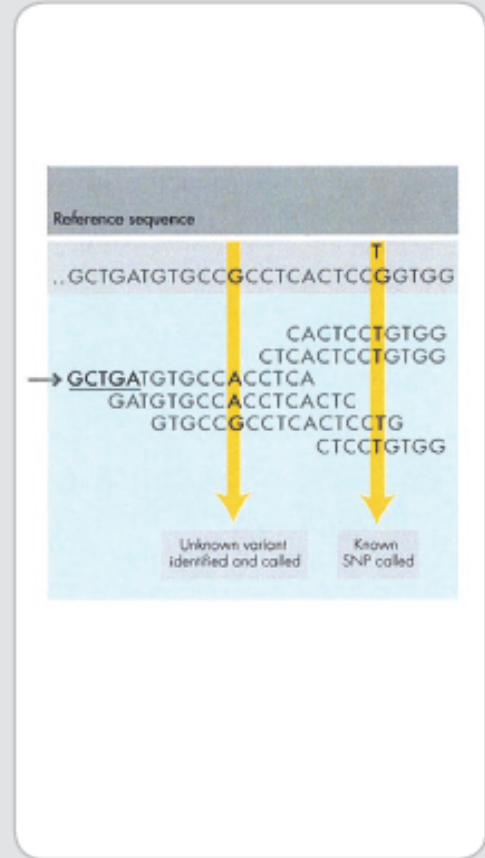
After laser excitation, the image is captured as before, and the identity of the second base is recorded.

11. SEQUENCING OVER MULTIPLE CHEMISTRY CYCLES



The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

12. ALIGN DATA

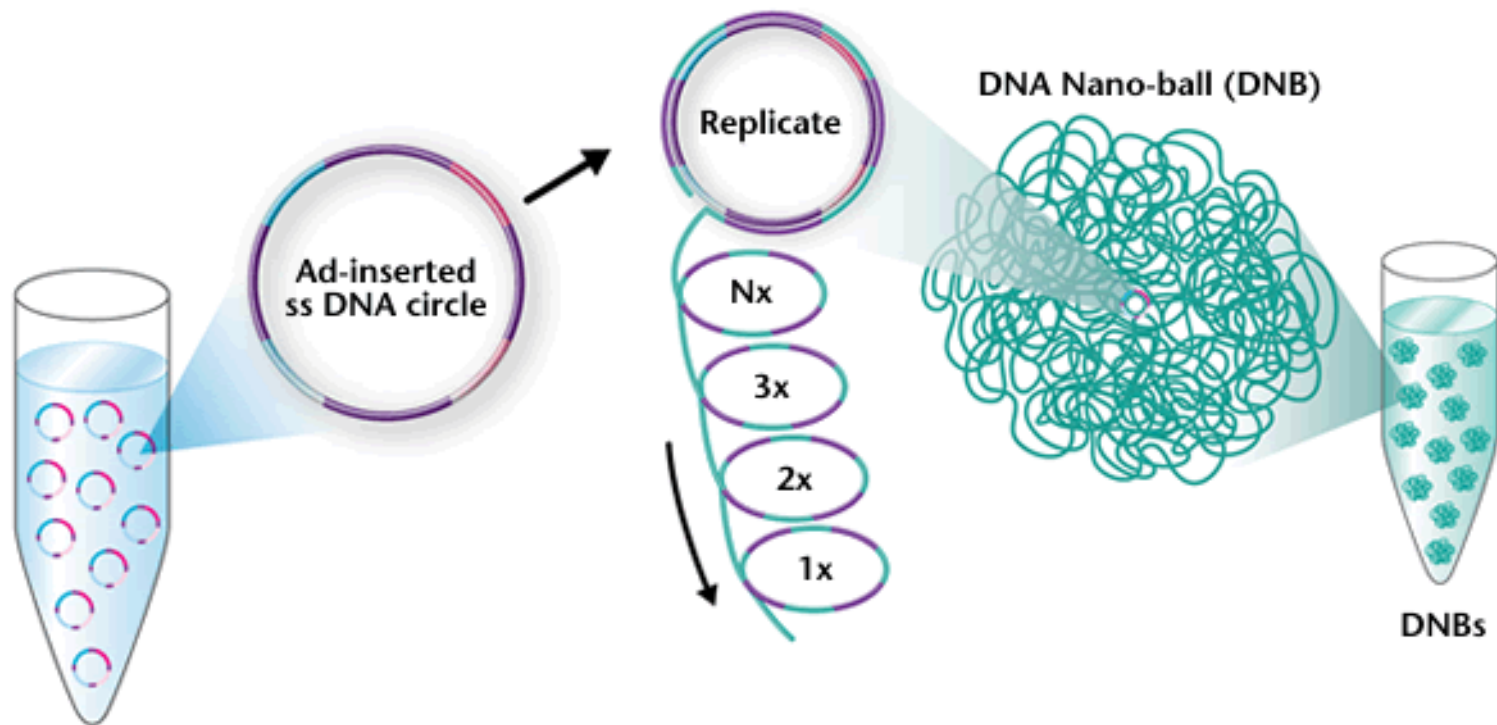


The data are aligned and compared to a reference, and sequencing differences are identified.

WT Sanger Institute Sequencing Centre - Hinxton

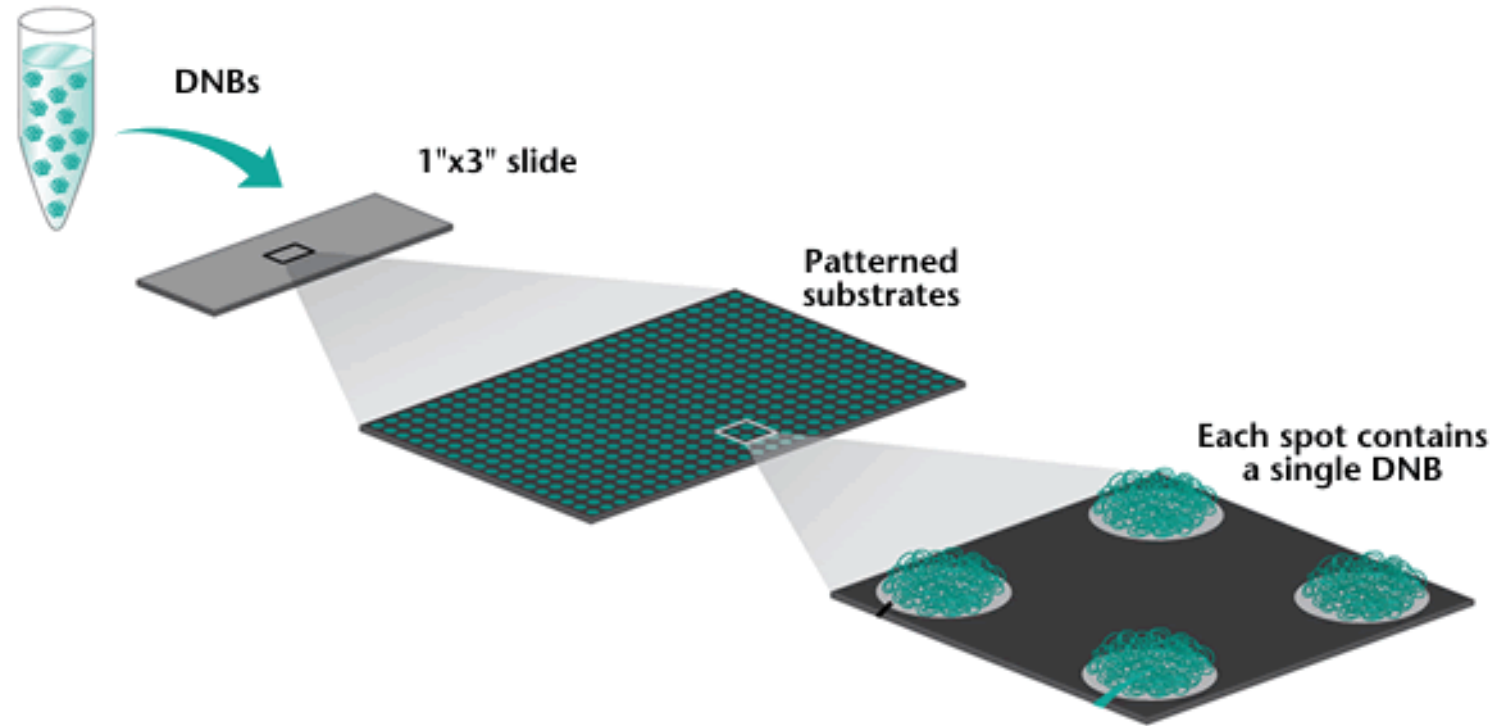


**Complete Genomics - zhluky DNA "nanoballs"
(nepoužívá PCR, ale rolling circle replication)**

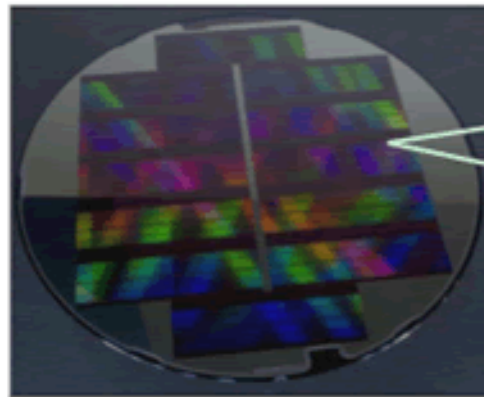


Complete Genomics

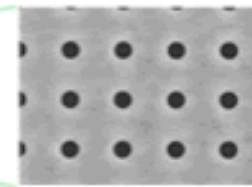
300 nm pozic, 2.8 mld na jednom sekvenovacím sklíčku



Complete Genomics



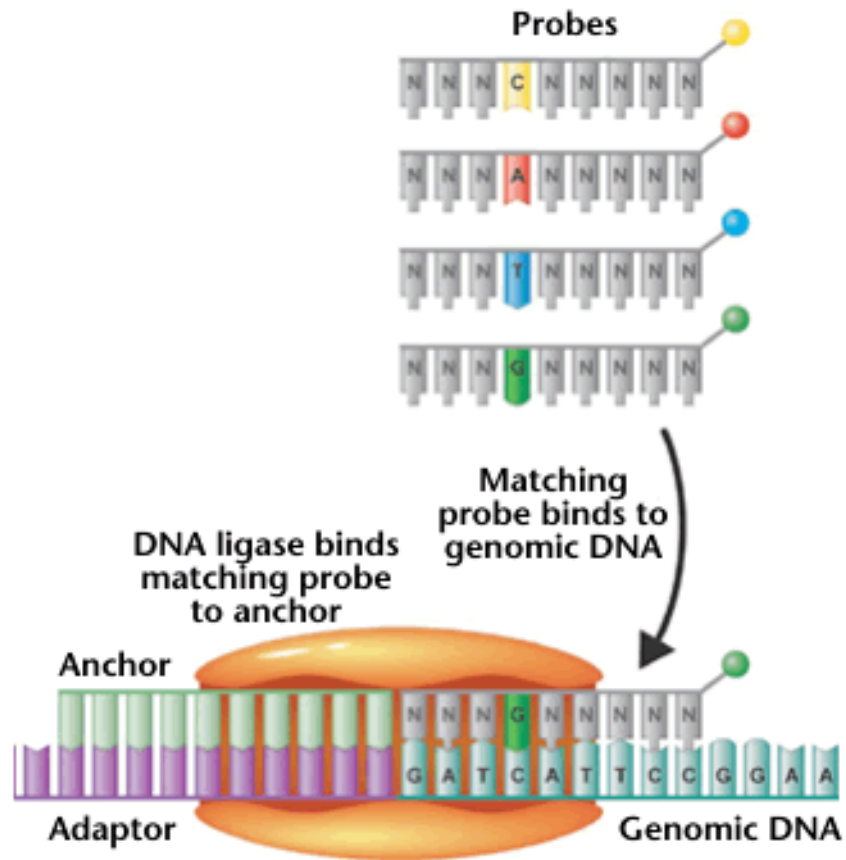
Silicon substrate consisting of 12 slides



Magnification of a section: Each spot will accommodate only a single DNB

Complete Genomics

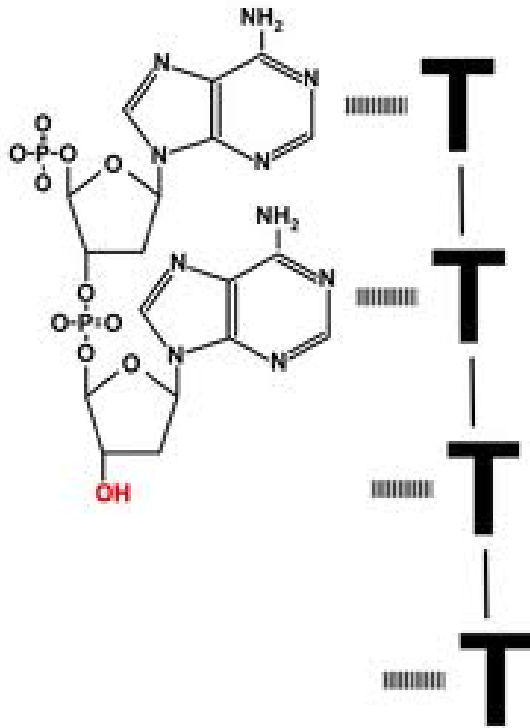
cPAL - "Probe-Anchor Ligation sequencing", 70 bází / zhluk



Ion Torrent (polovodičové sekvenování)

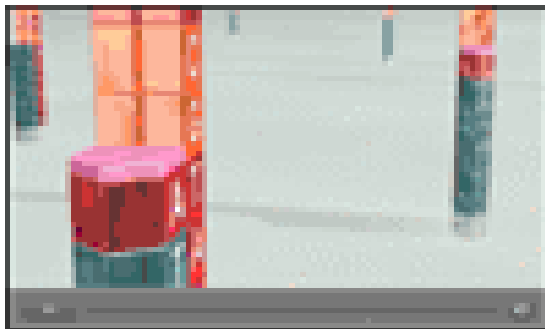
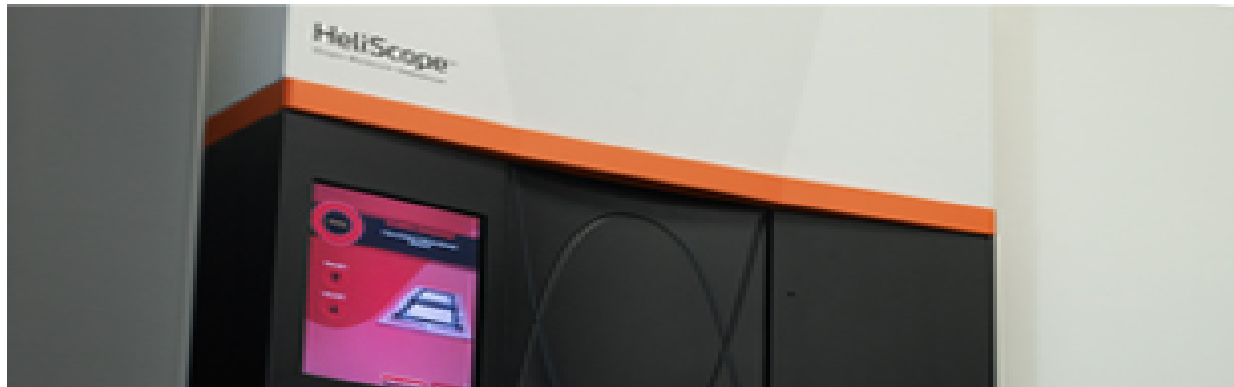
dnes součást LifeTechnologies

detekuje protony uvolněné při polymerizaci DNA



tSMS – “single molecule sequencing”

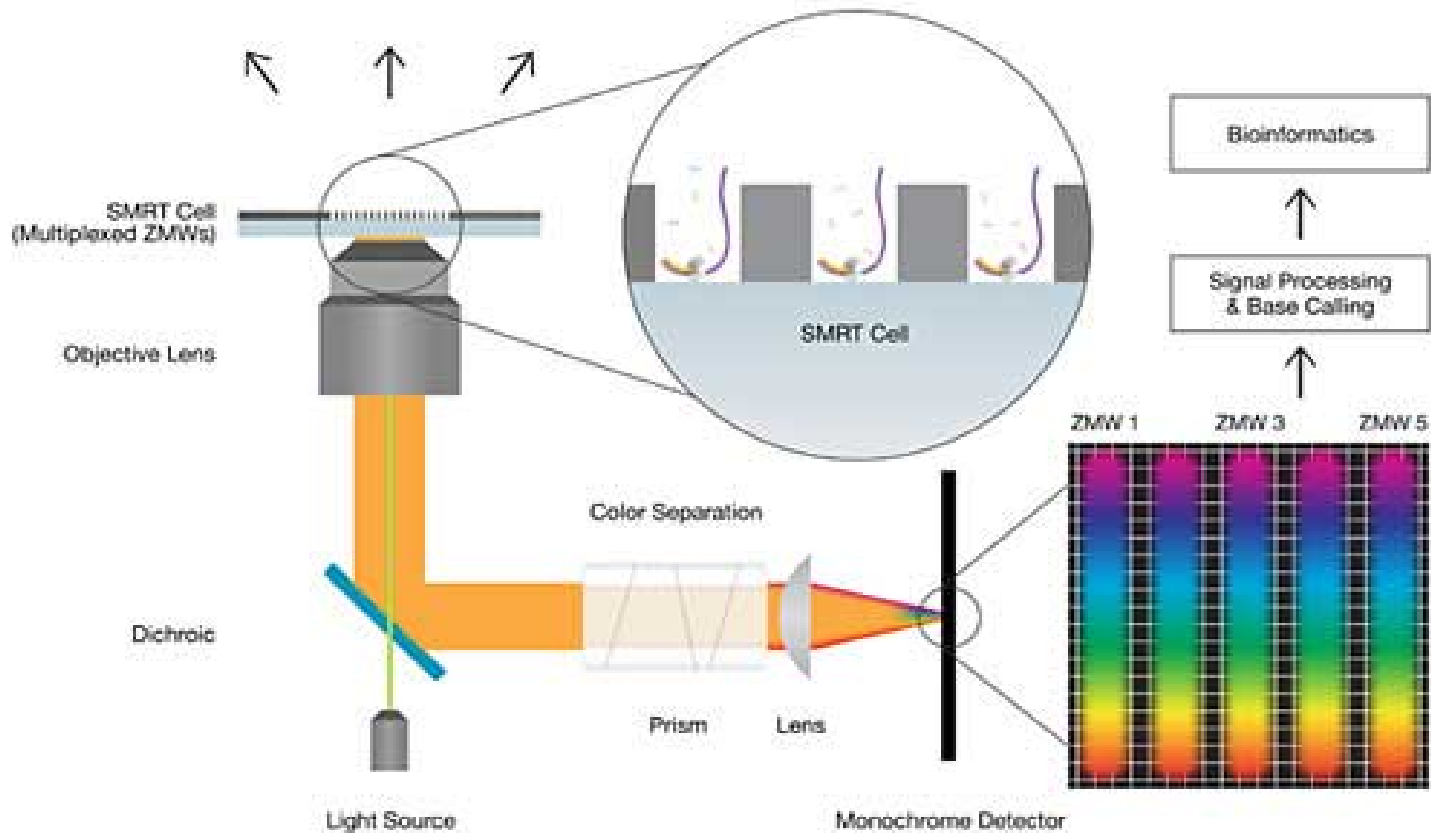
Helicos (existuje jenom několik strojů)



- 10^9 bp / day
- >35bp readlength

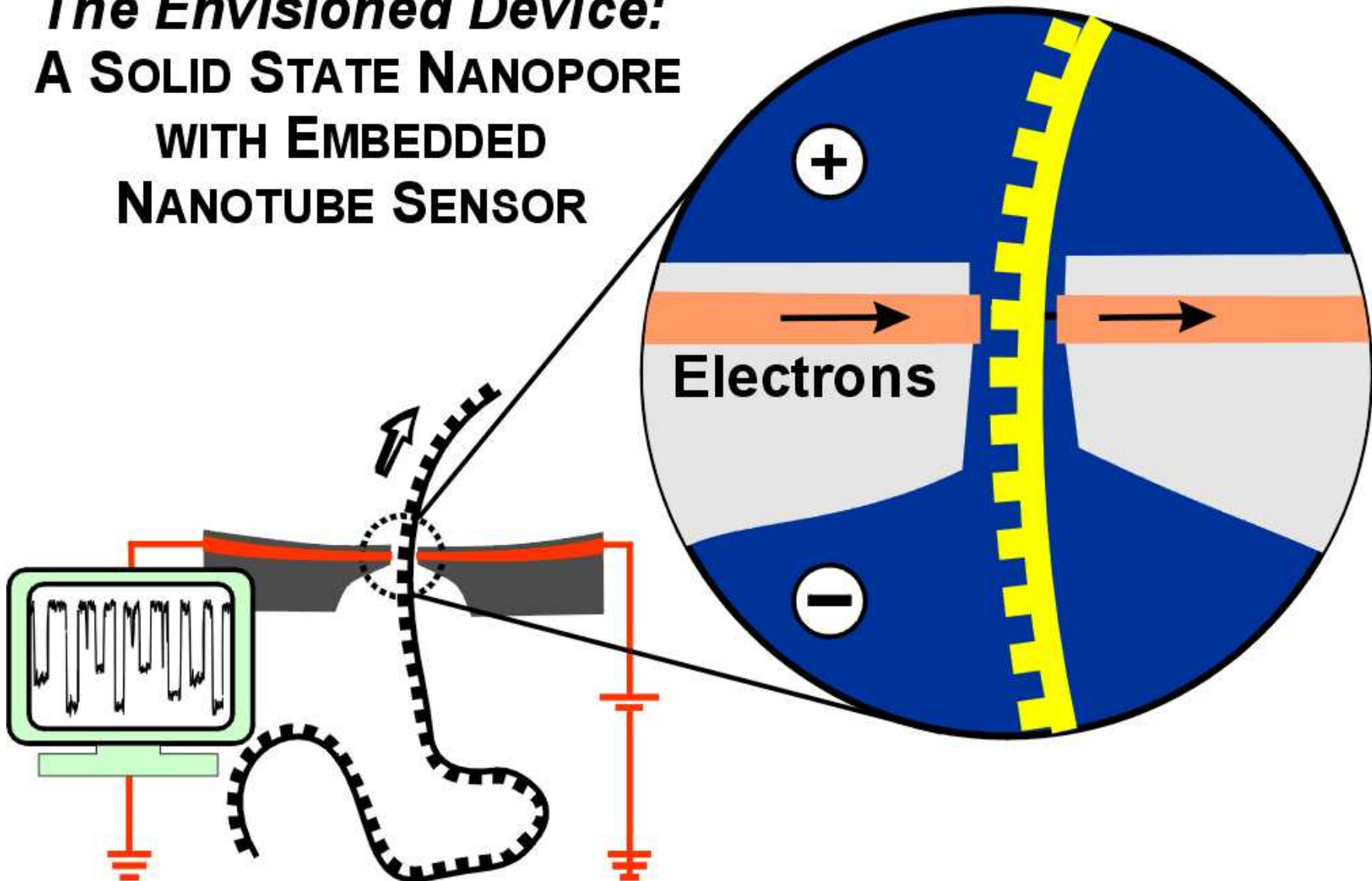
SMRT – “single molecule real time”

Pacific Biosciences

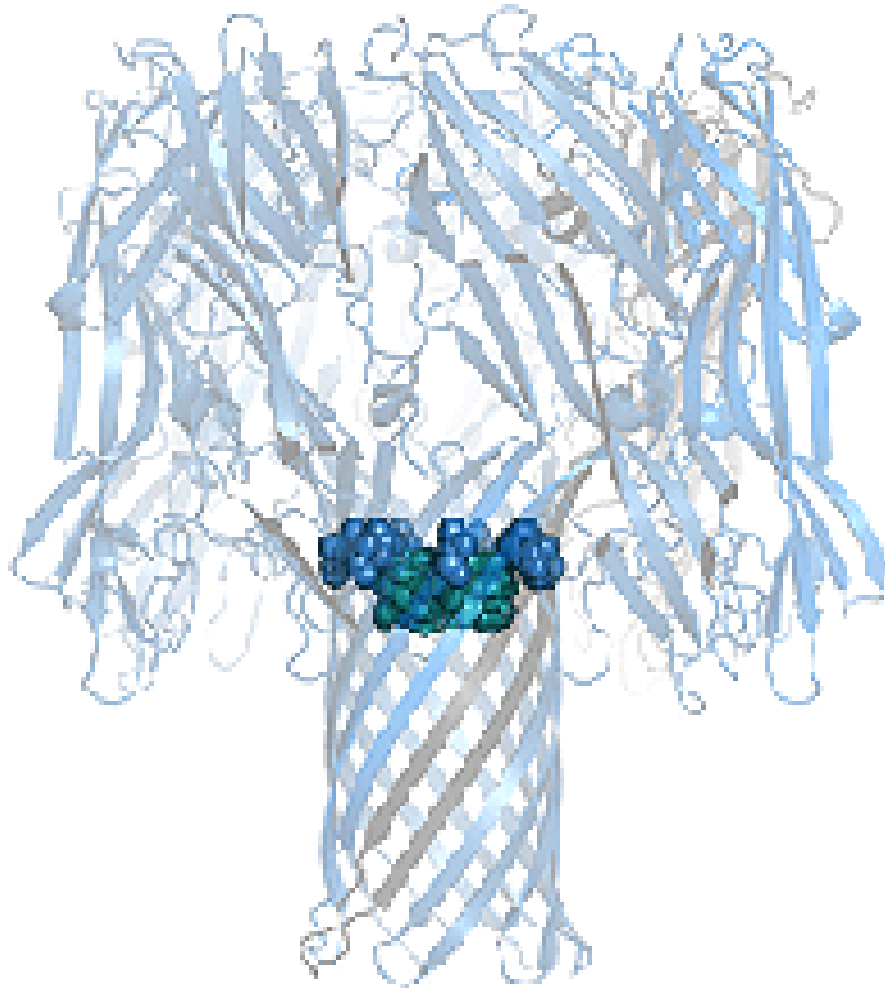


Sekvenování nanopory

The Envisioned Device:
A SOLID STATE NANOPORE WITH EMBEDDED NANOTUBE SENSOR

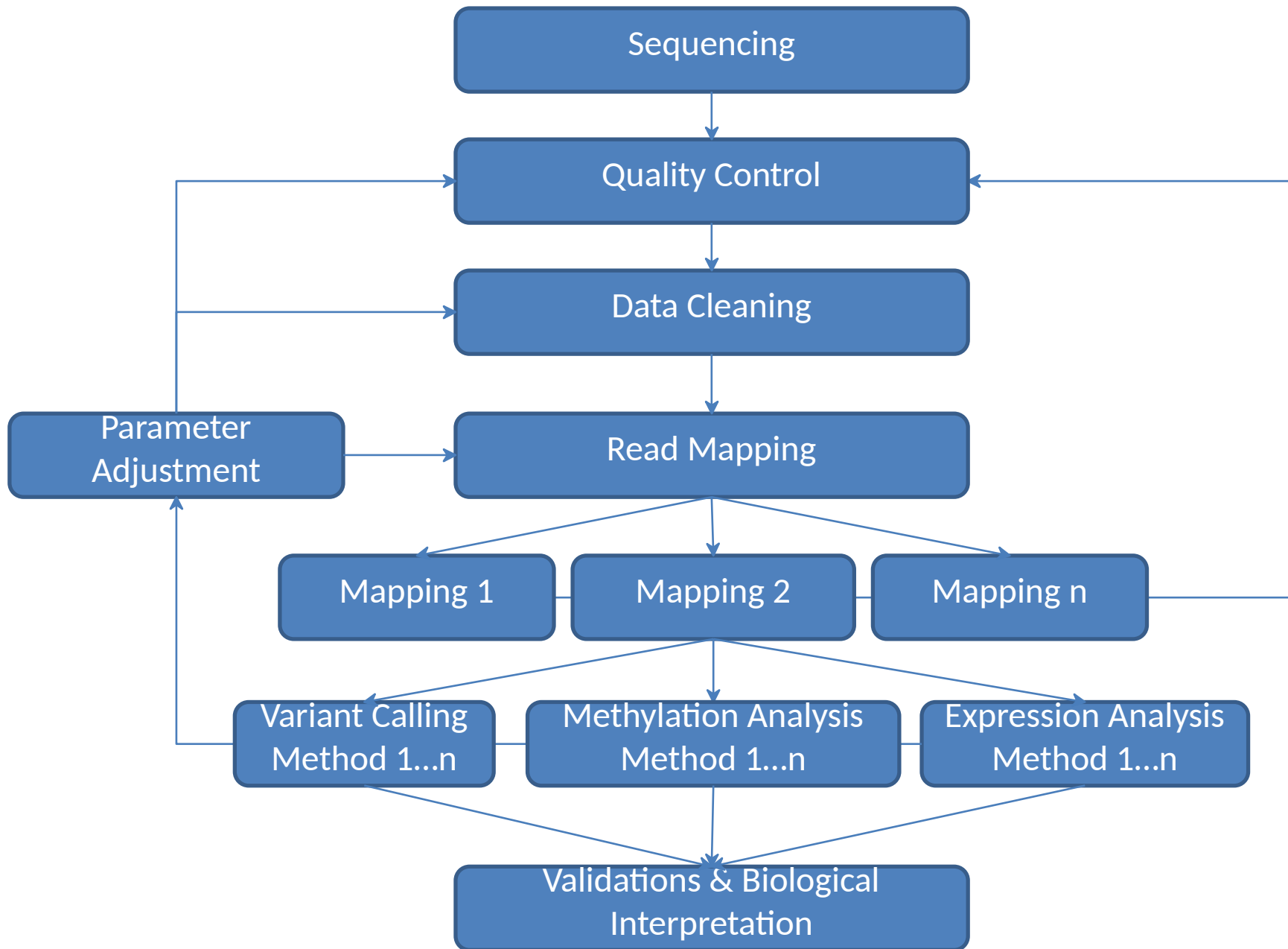


Oxford Nanopore Technologies



Nanopor tvořen proteinem alpha-hemolysin s cyclodextrinovým adapterem (místo detekce nukleotidů).

Typický postup zpracování NGS dat



Kontrola kvality a čištění dat

Zastoupení nukleotidů

Kvalita nukleotidů na pozicích (log skóre “phred”; 10 = 90%)

Kontaminace adaptéry a jinými cizími sekvencemi

Stupeň duplikca sekvencí

Zkrácení sekvencí na společnou délku

Např.:

<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>

http://hannonlab.cshl.edu/fastx_toolkit/

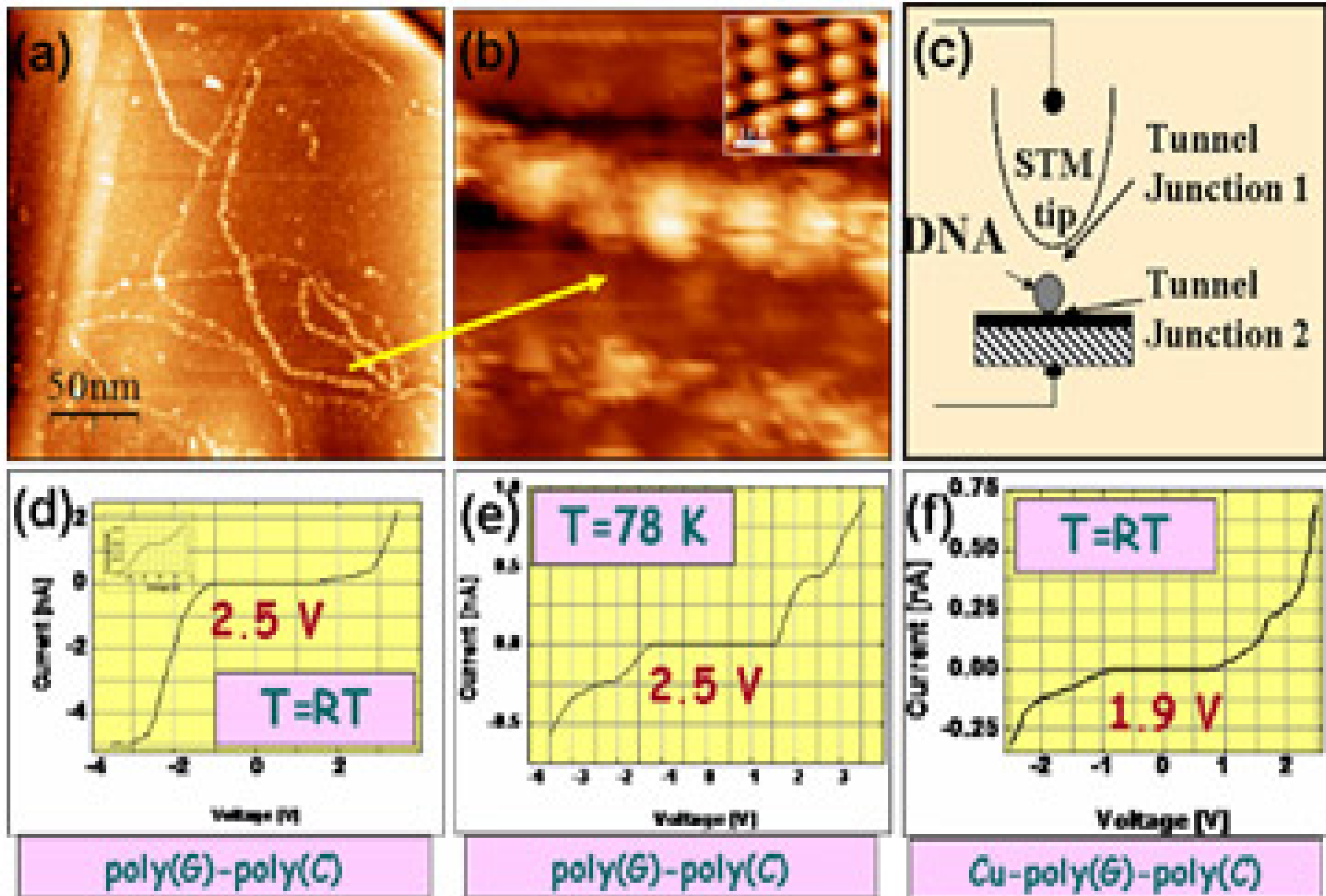
Mapování sekvencí

- Index
 - _ Tabulky slov, hešovací tabulky
 - _ Suffixové stromy a pole
 - _ Burrows-Wheelerova-Transformace (BWT) sufixového pole
 - _ FM-Index
- Filtrování
 - _ S jednou nebo více shodami
- Verifikace
 - _ prodloužení
 - _ Dynamické programování

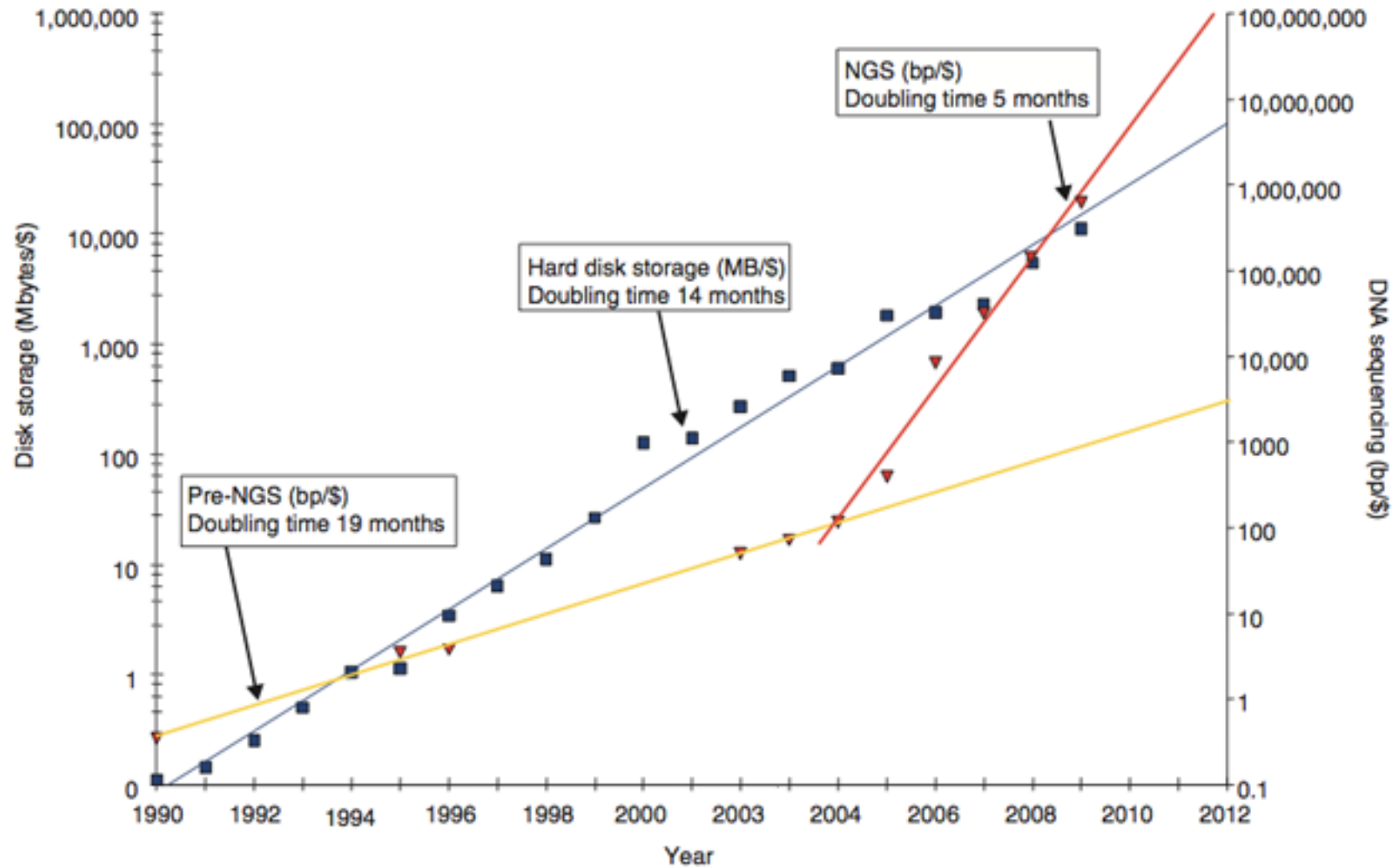
BWA, Bowtie2, Soap2 mrFAST, SHRiMP2, RazerS3, Masai
GSNAP, Stampy, BWA mem (tolerují mezery)

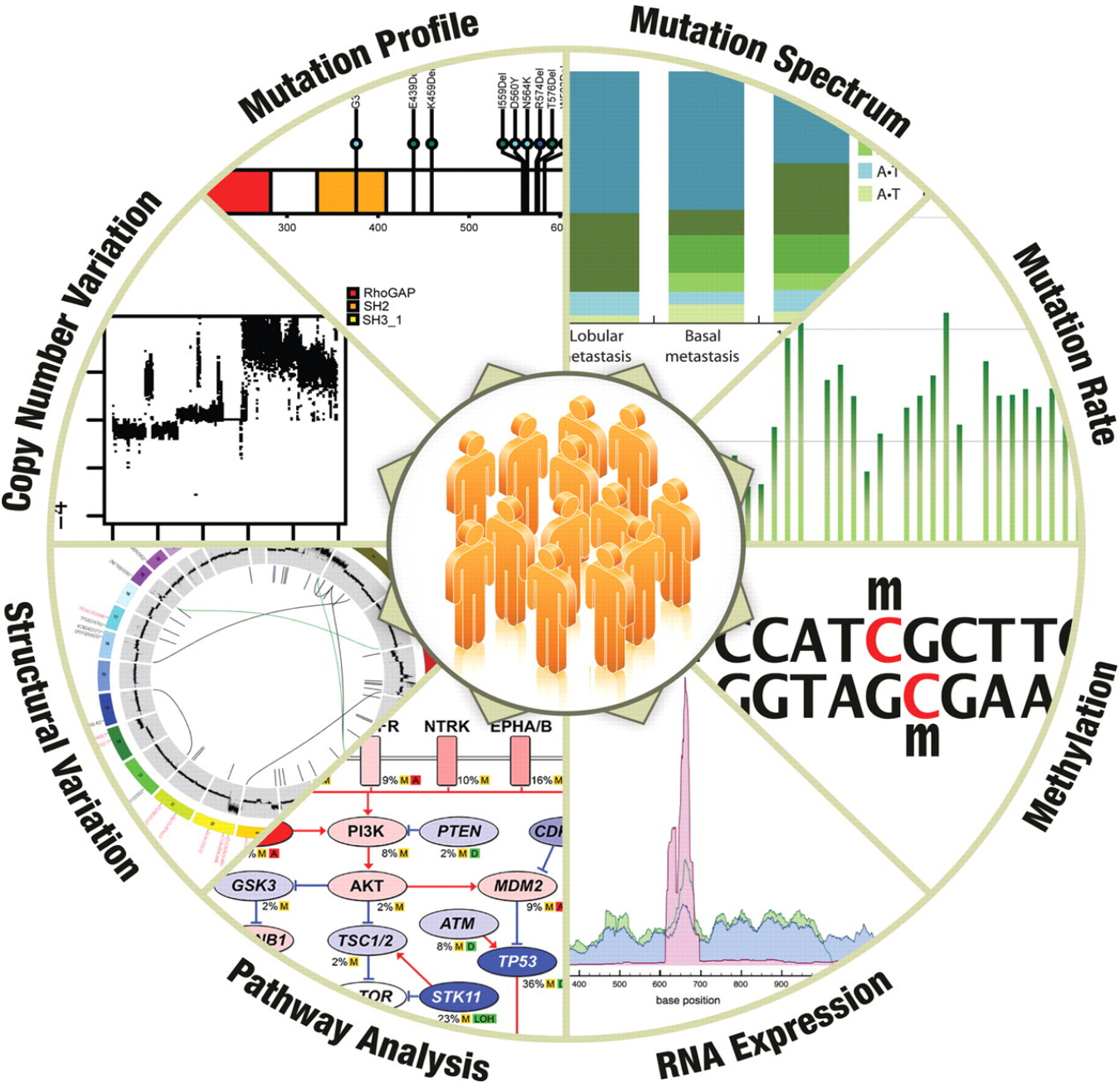
- RNA-Seq, Bis-Seq, speciální obsluha repetitivních sekvencí
TopHat, STAR, Bismark, lobSTR, ...

Analýza DNA pomoci STM (scanning tunnelling microscope)

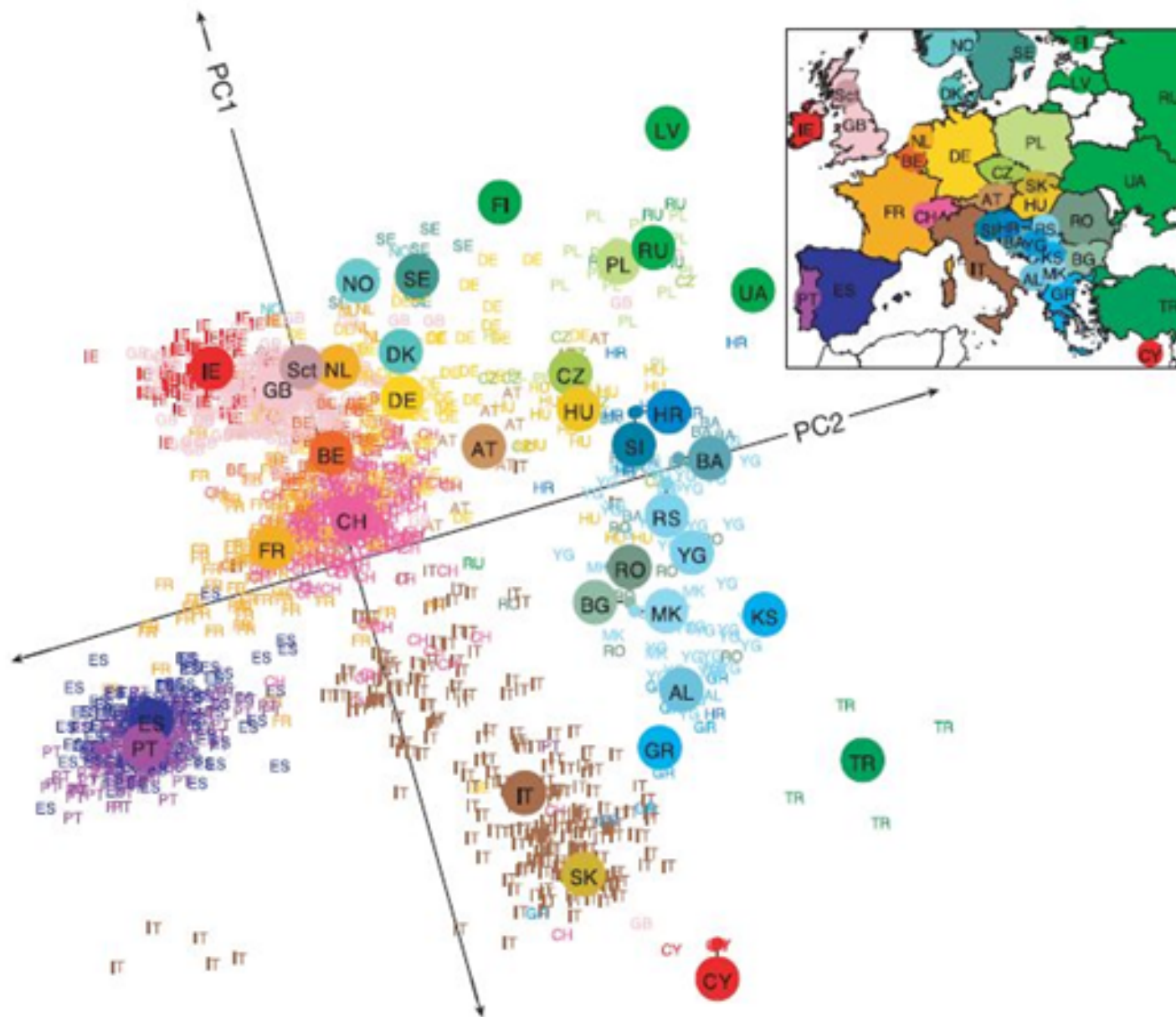


Data growth in DNA sequencing





a



<http://www.thegeneticgenealogist.com/2008/09/24/ancestral-gps-pinpointing-the-geographic-origin-of-autosomal-dna-sequences/>



Human Longevity



<http://www.fiercebiotechit.com/story/illumina-backs-venters-plan-create-worlds-largest-human-genome-sequencing-c/2014-03-05>