



Středoevropský technologický institut  
BRNO | ČESKÁ REPUBLIKA

# Masivně paralelní sekvenování

**Boris Tichý**

Sdílená laboratoř Genomika

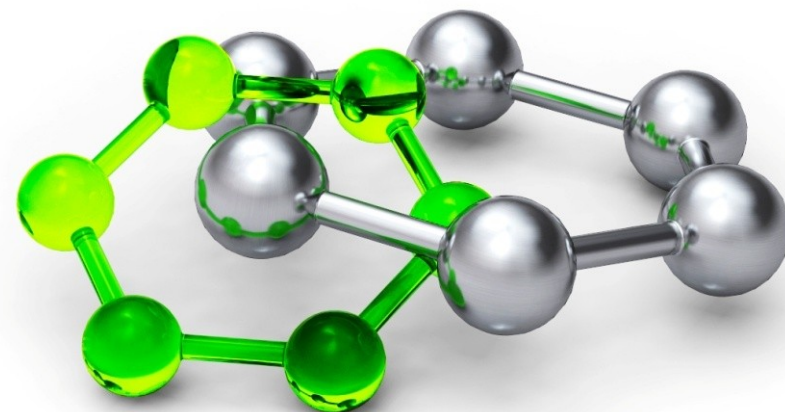
Brno, 9.10.2015



EVROPSKÁ UNIE  
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ  
INVESTICE DO VAŠÍ BUDOUCNOSTI



OP Výzkum a vývoj  
pro inovace

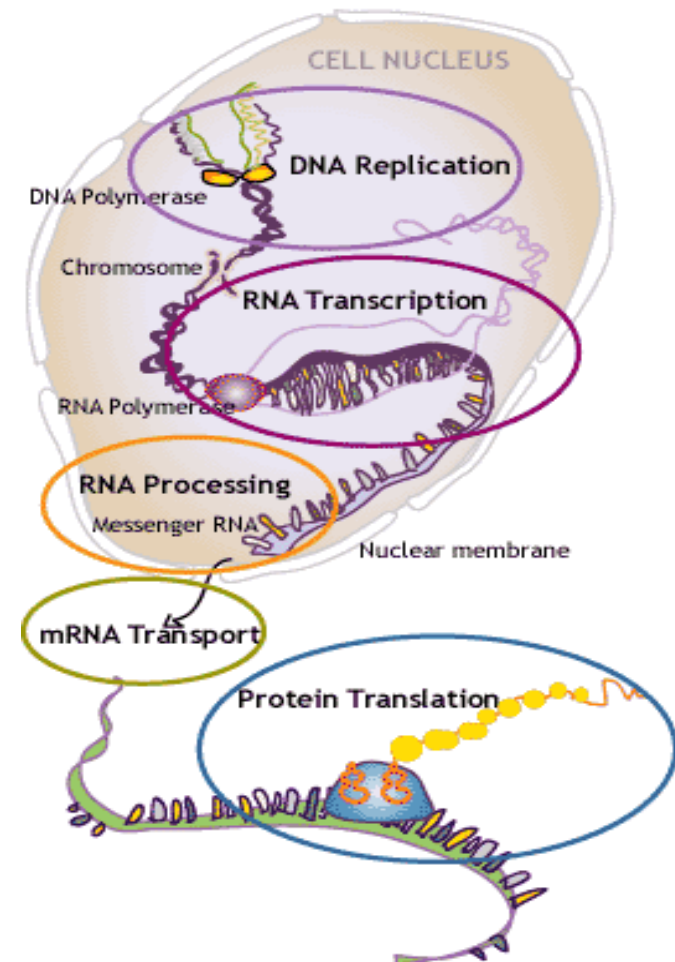
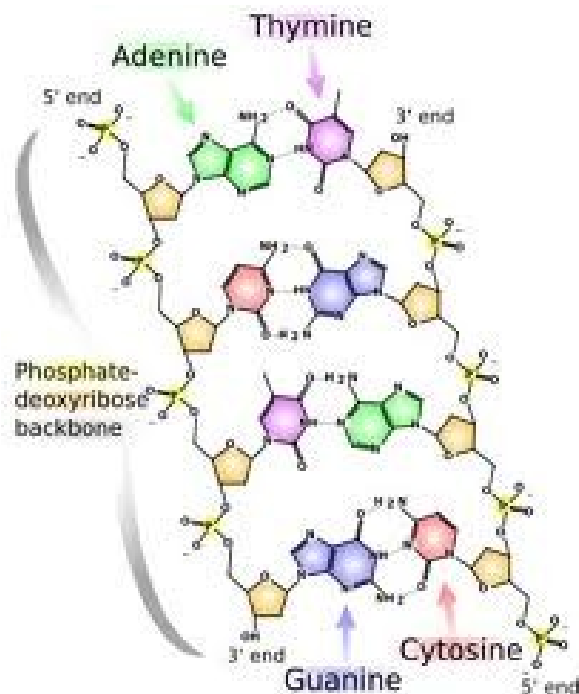


# Informace je uložena v DNA

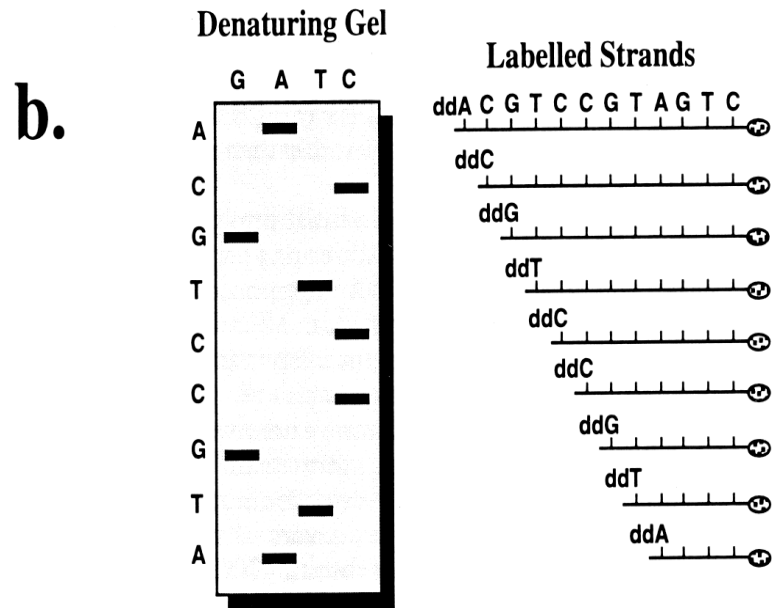
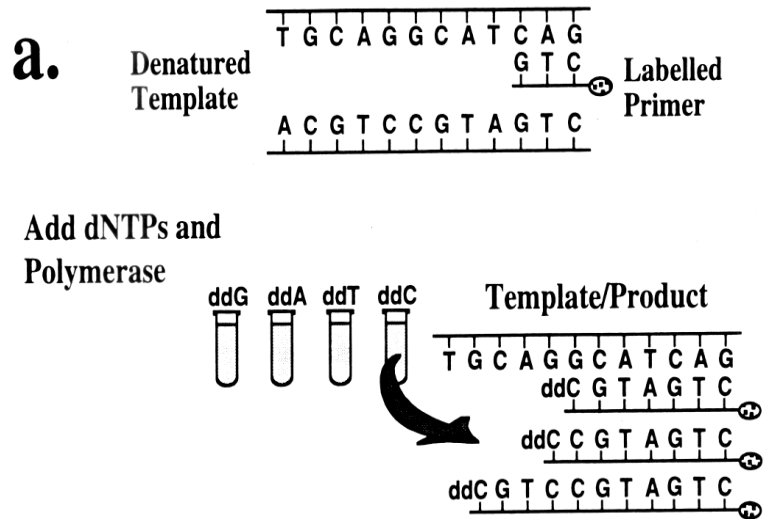
Informace uložena jako sekvence bází A, C, G, T

V každé lidské buňce je ~ 3 miliardy bází = ~ 3 metry = ~ 6.6 pikogramů

Sekvence DNA je v každé buňce stejná



# Sekvenování DNA

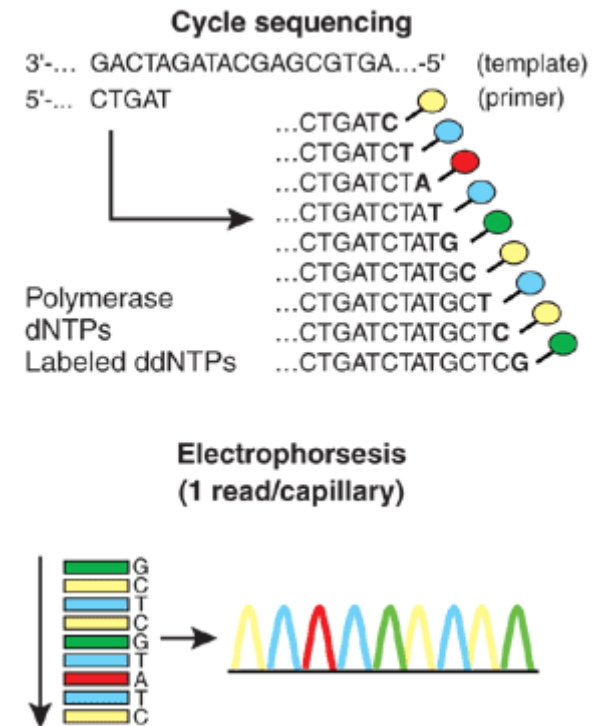


Vytvoření různě dlouhých fragmentů DNA

Dideoxy-NTPs → ukončení polymerace

Gel → kapilární elektroforéza

Značené primery nebo ddNTPs



# Masivně paralelní sekvenování

PCR amplifikace jednotlivých DNA fragmentů

nebo

Sekvenování jednotlivých DNA fragmentů

= Single molecule sequencing

Sekvence je čtena při syntéze nového řetězce

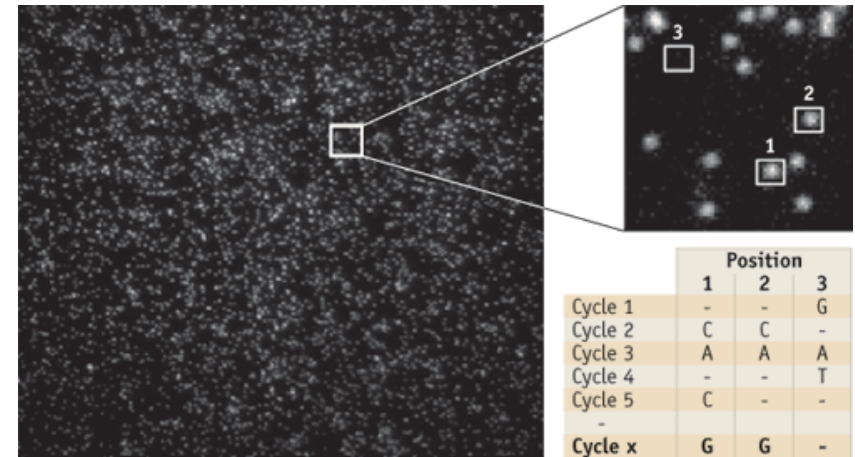
Technologie a přístroje přizpůsobeny paralelizaci

Stovky milionů jednotlivých PCR reakcí a sekvenací najednou

(běžně prodávané kapilární sekvenátory jsou max. 96-kapilární)

Většinou kratší sekvence – desítky bází

(kapilární – běžně až 1000 bází)

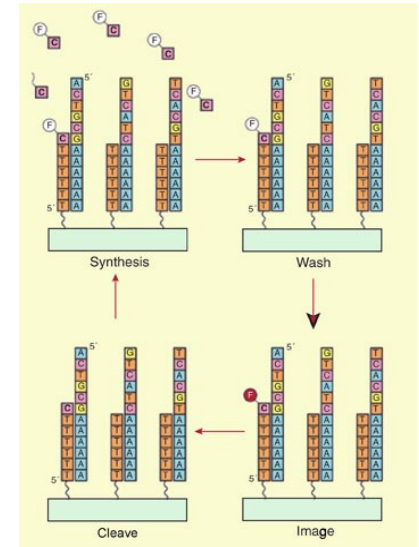


# Masivně paralelní sekvenování

Sequencing by synthesis

Polymeráza

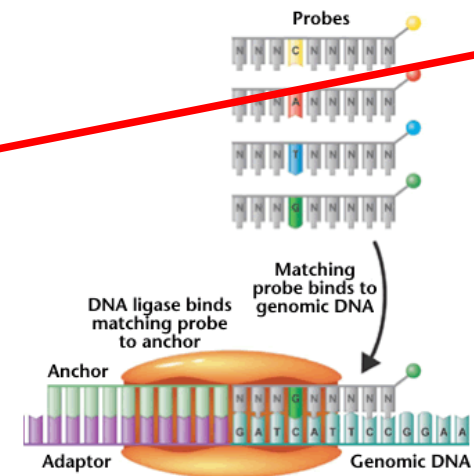
Sestavování nového řetězce z jednotlivých nukleotidů



Sequencing by ligation

Ligáza

Sestavování nového řetězce z oligonukleotidů



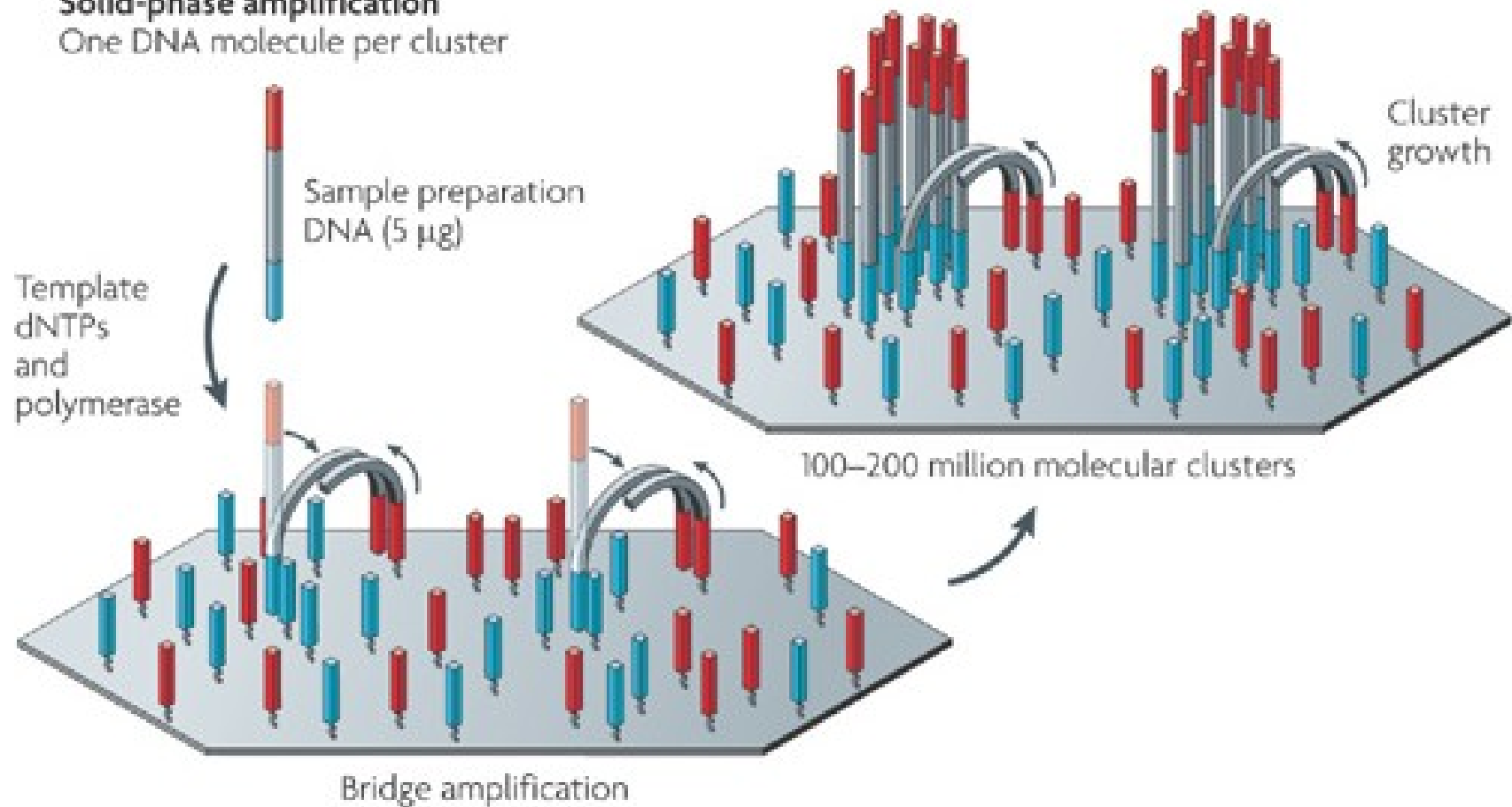
Non-enzymatic sequencing

Nanopory, elektronová mikroskopie

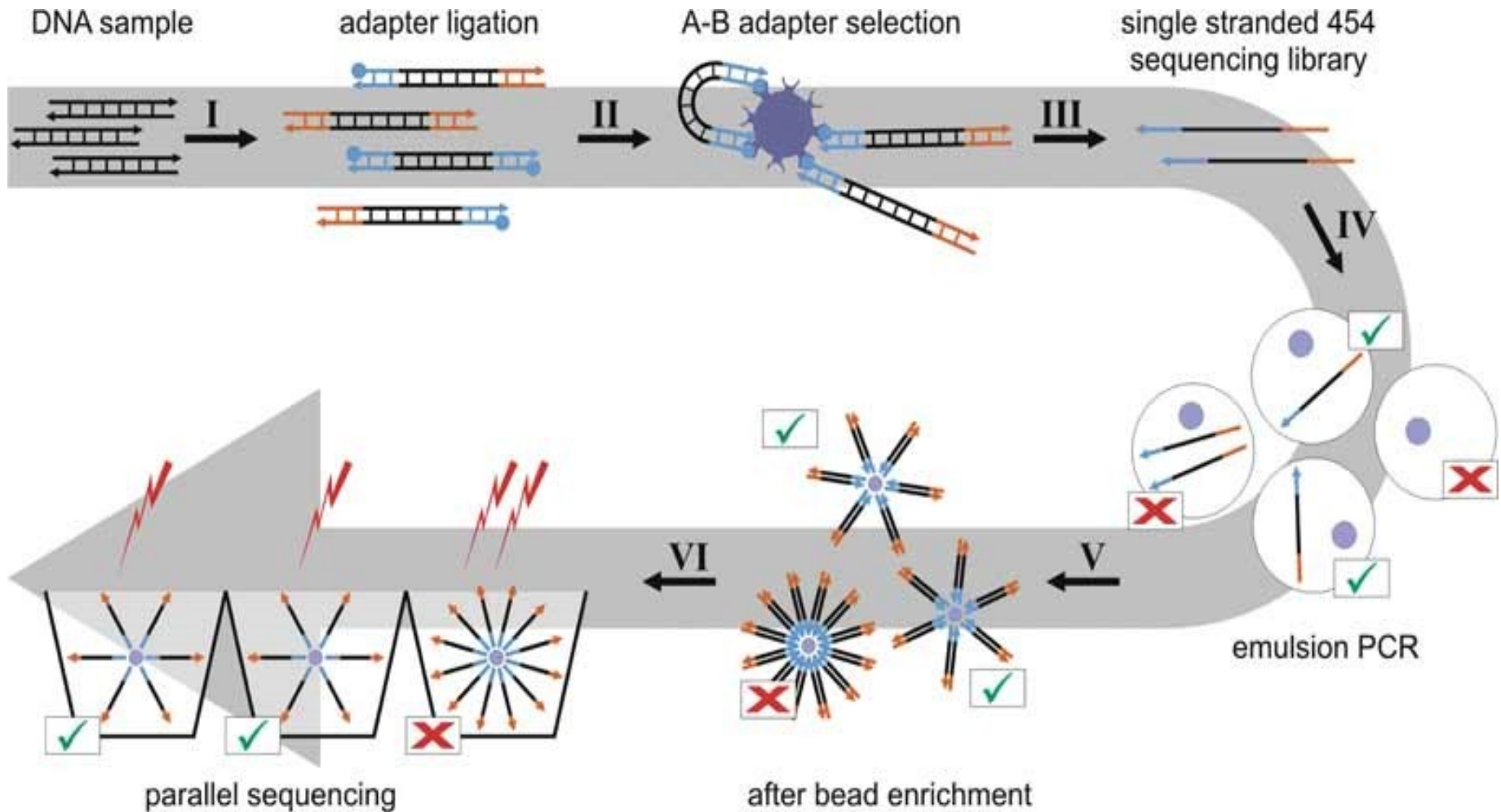
Přímé čtení sekvence

### Bridge amplifikace

**b** Illumina/Solexa  
Solid-phase amplification  
One DNA molecule per cluster

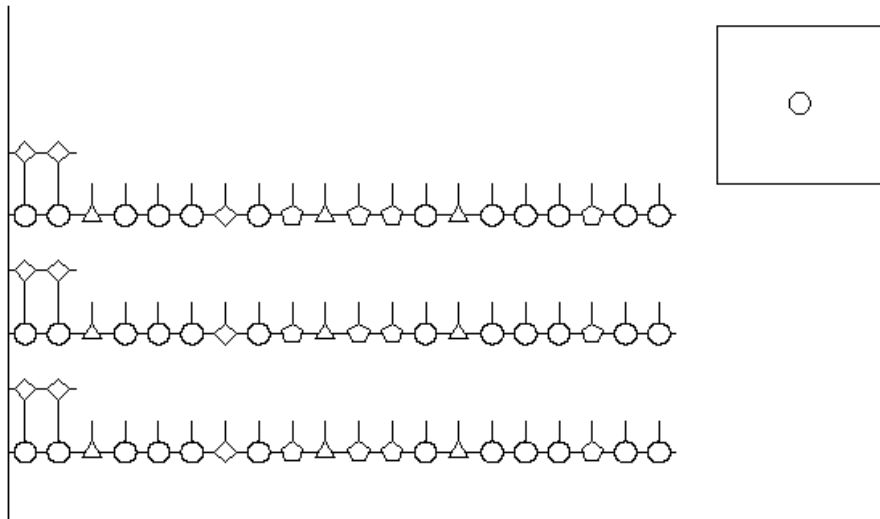


### Emulzní PCR



# Masivně paralelní sekvenování

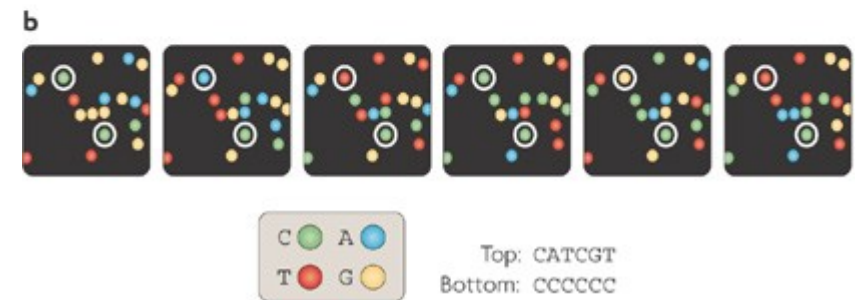
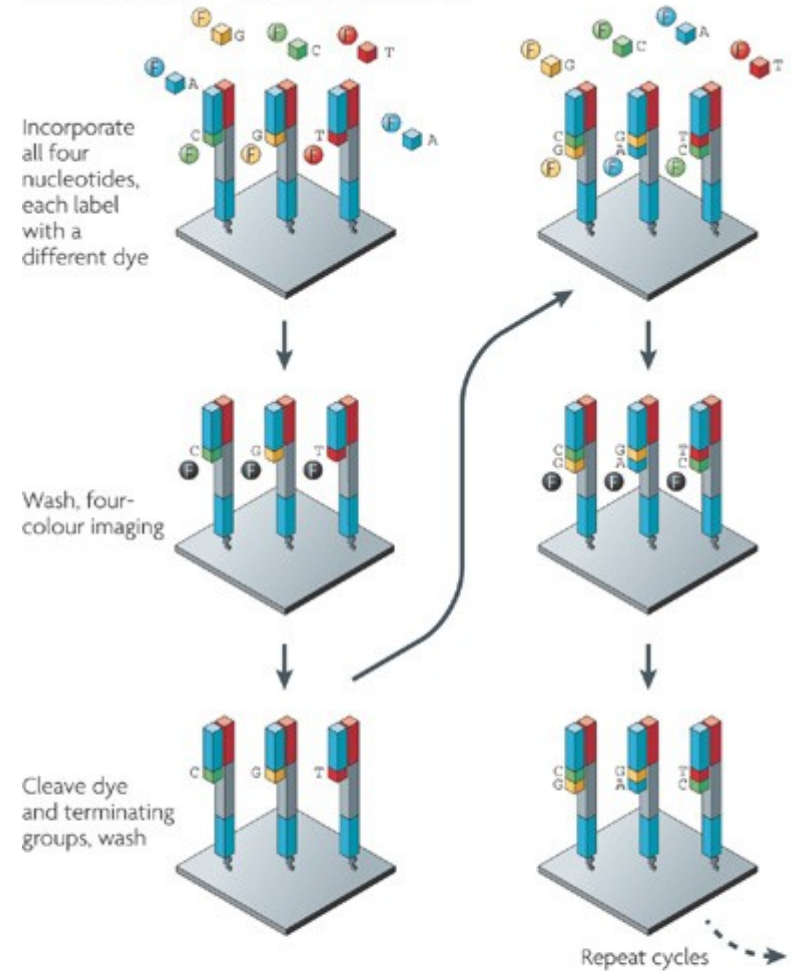
## Sekvenování s reverzibilními terminátory



[http://www.youtube.com/watch?v=Zqr8\\_KiuzHU](http://www.youtube.com/watch?v=Zqr8_KiuzHU)  
<http://www.youtube.com/watch?v=I99aKKHcxC4>

## Technologie Illumina

a Illumina/Solexa — Reversible terminators







HiSeq X Ten – kapacita 18.000 genomů ročně, cena za genom \$1.000, cena \$10M



HiSeq 4000

Kapacita 12 genomů/běh  
(24/týden), 1,5TB/běh



NextSeq 500

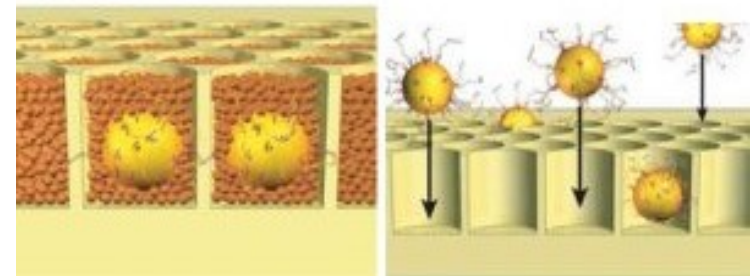
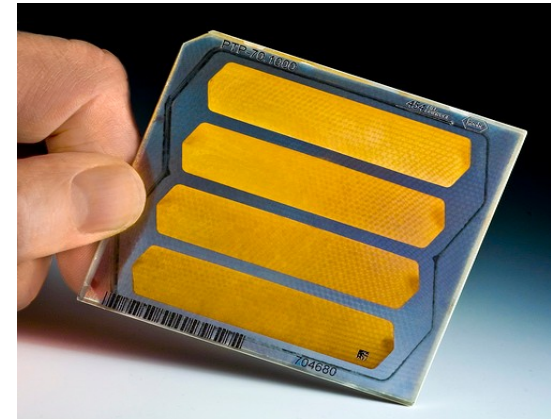
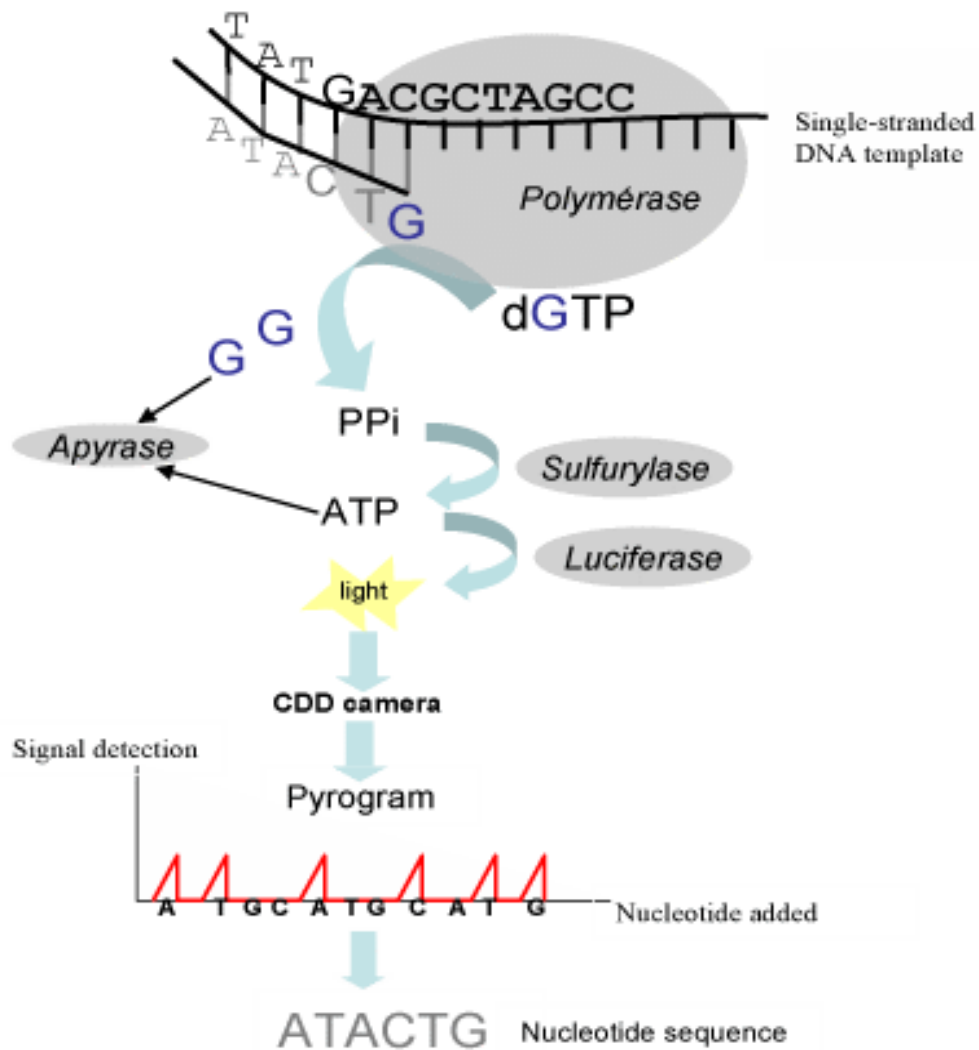
Kapacita 1 genom/běh  
(3/týden), 120GB/běh



MiSeq

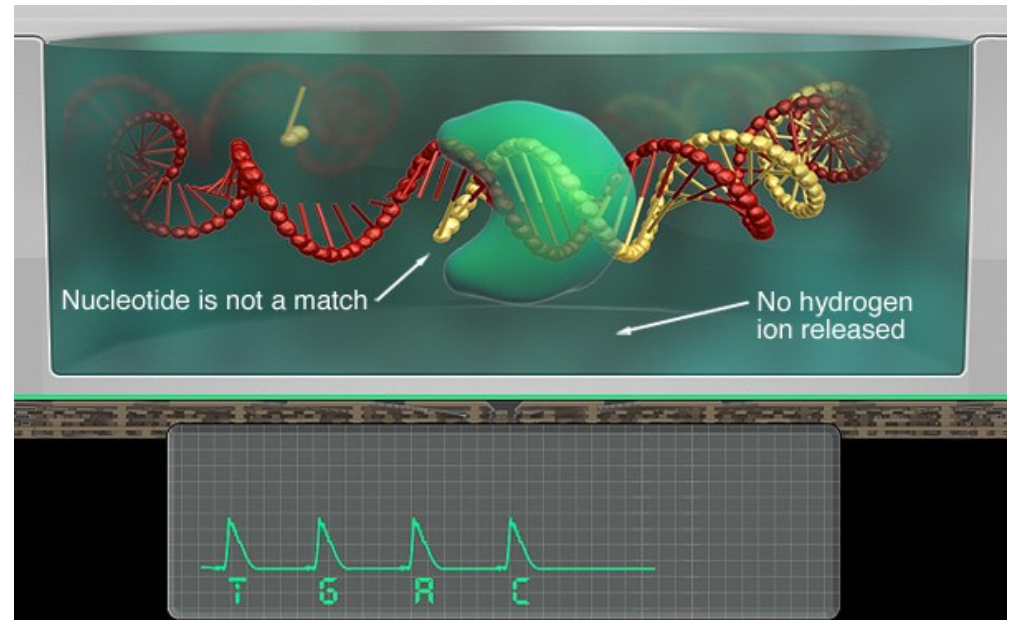
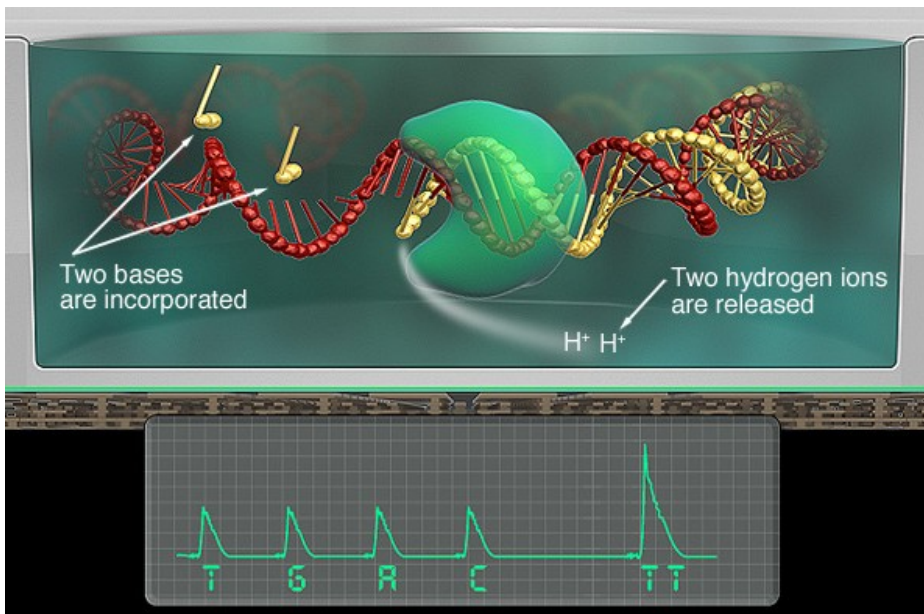
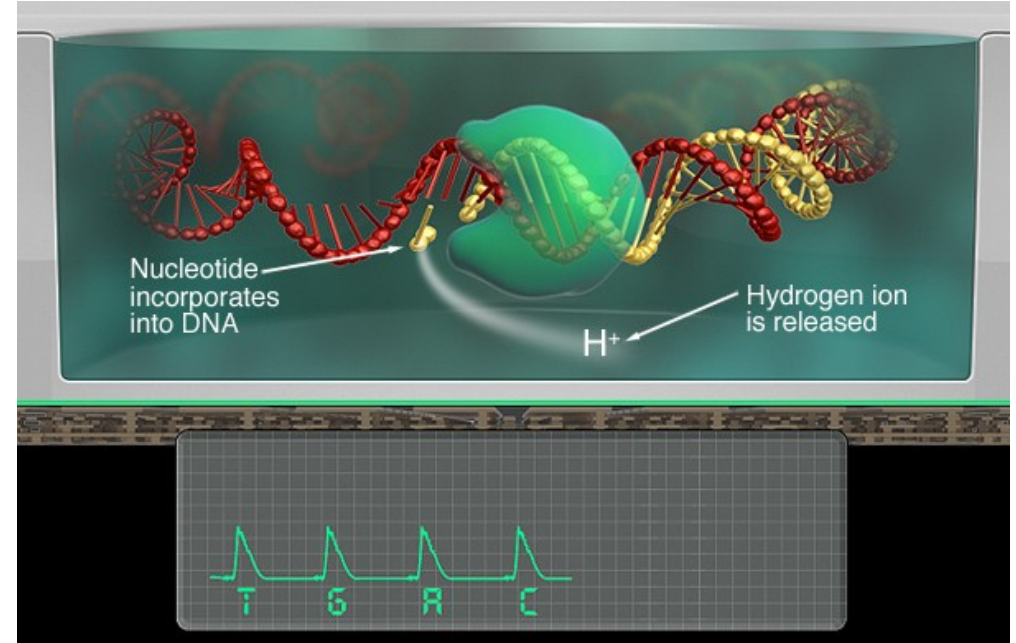
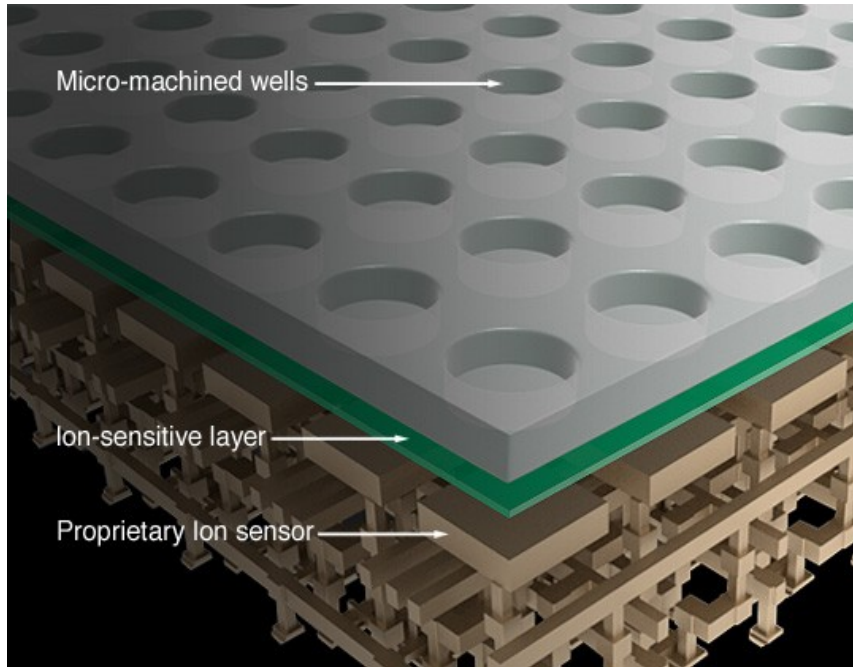
Kapacita 0,15  
genomu/běh,  
15GB/běh

### Pyrosekvenování



# Masivně paralelní sekvenování

# Technologie Ion Torrent



# Masivně paralelní sekvenování

## Technologie Ion Torrent



Ion Torrent PGM

Kapacita 2GB/běh



Ion Proton

Kapacita 10+GB/běh

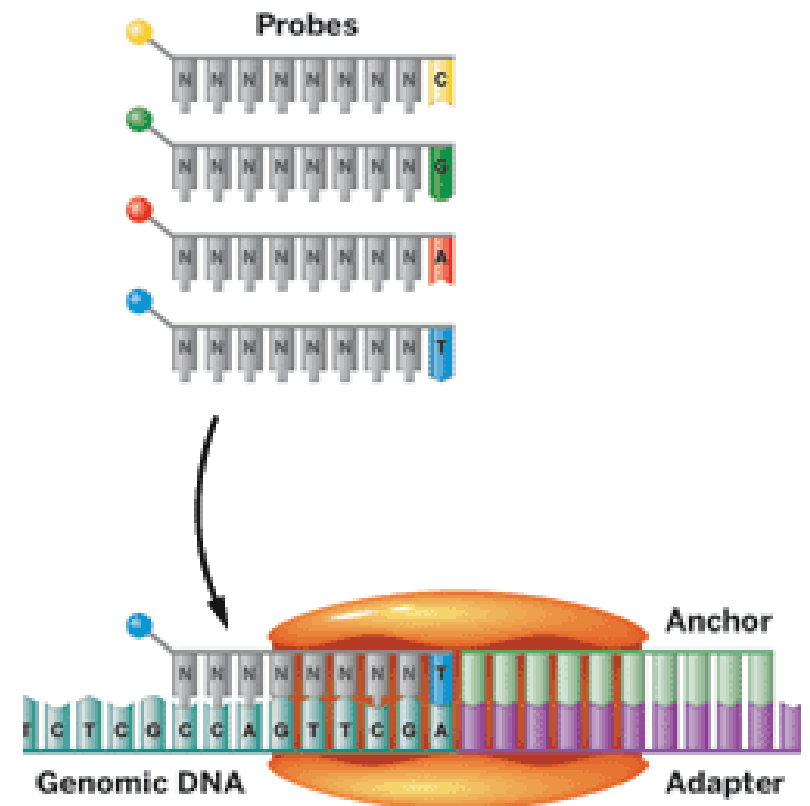
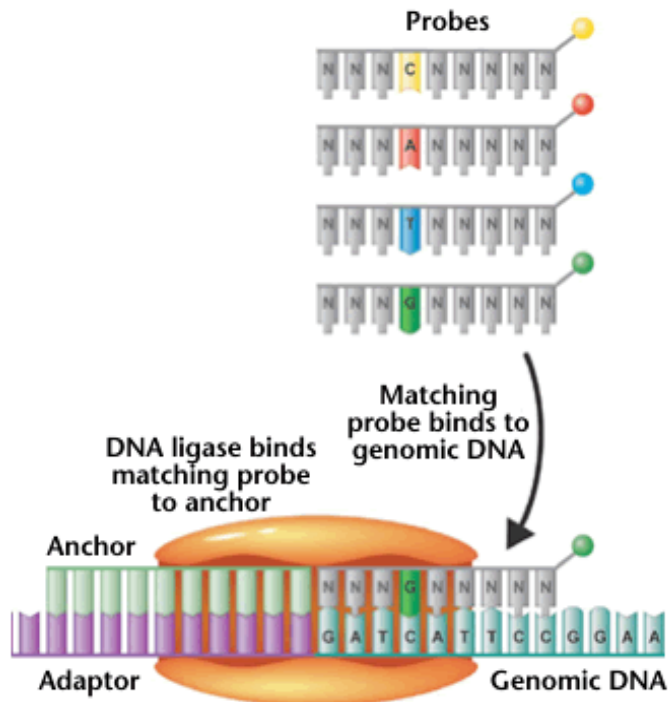


Ion S5

Kapacita 10+GB/běh

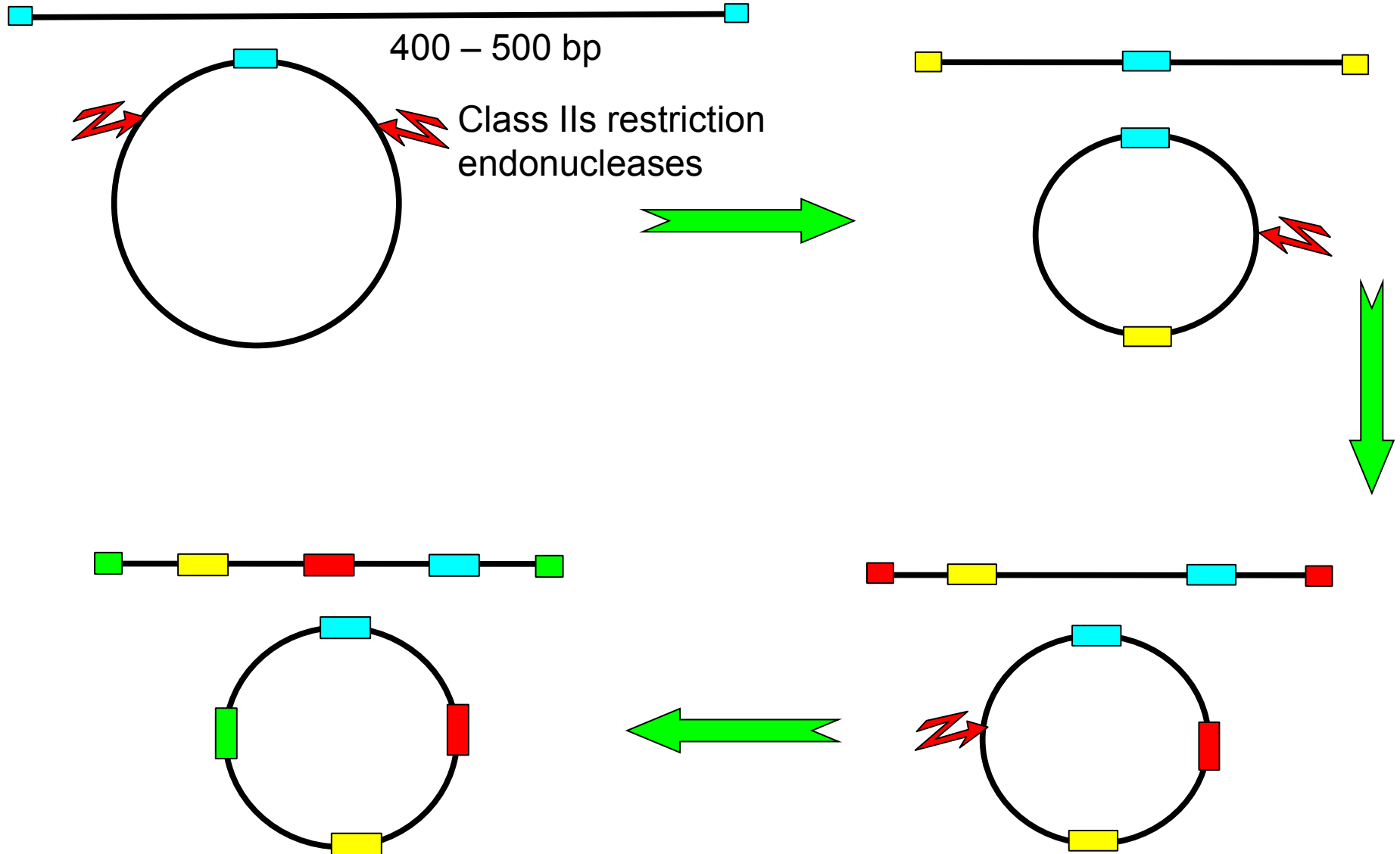
# Masivně paralelní sekvenování

## Sekvenování (hybridizací a) ligací



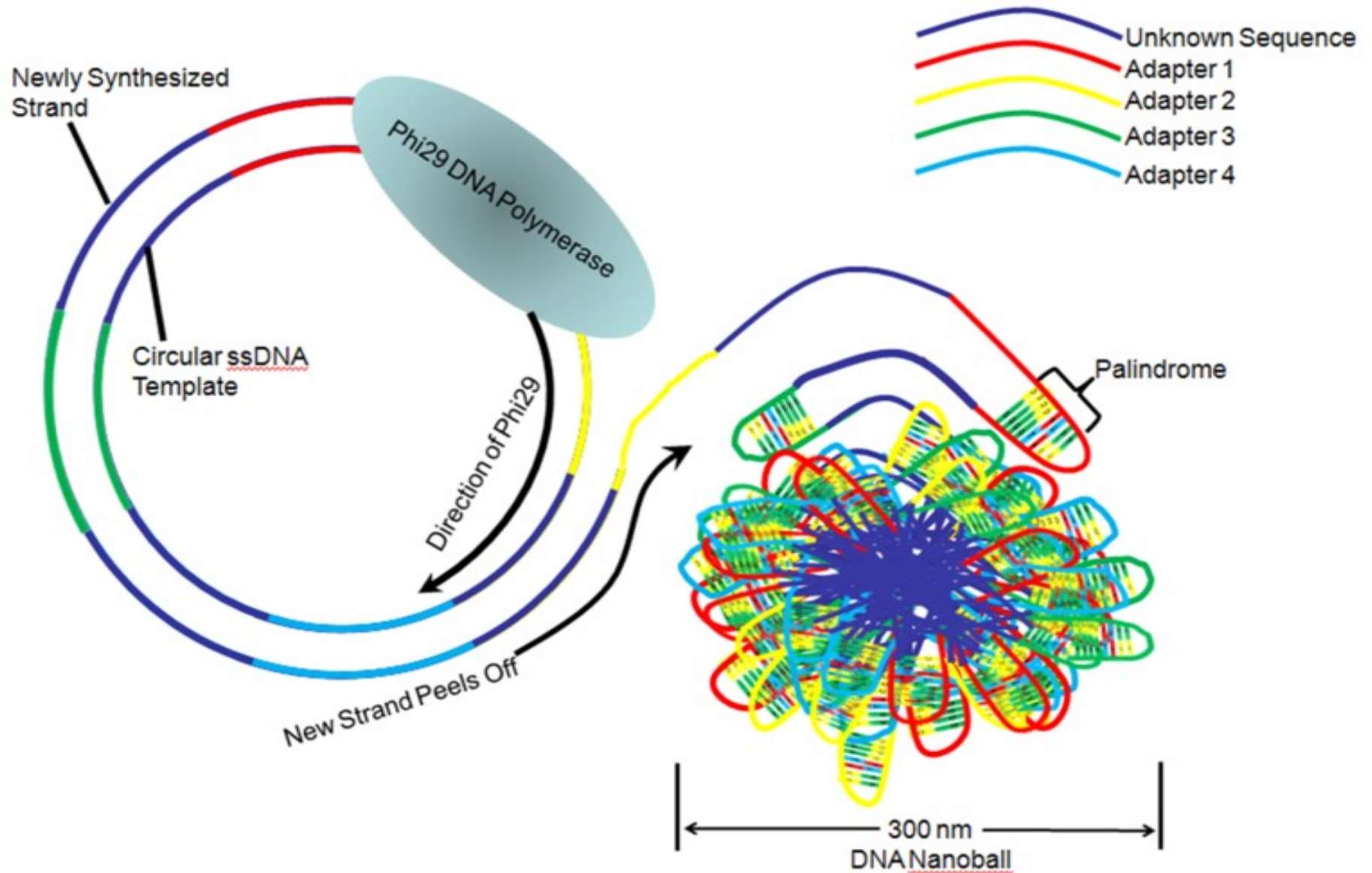
# Masivně paralelní sekvenování

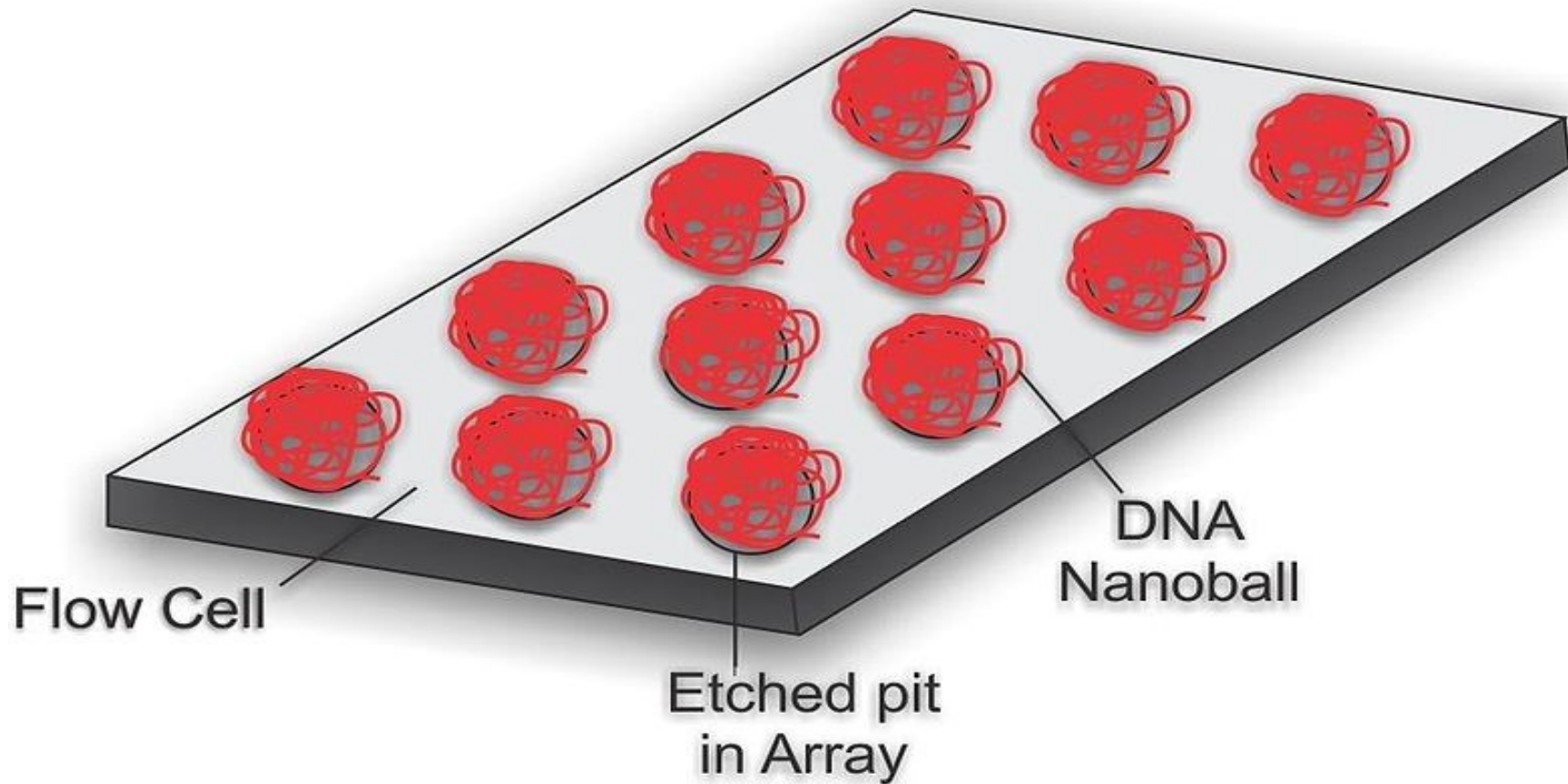
## Technologie cPAL



# Masivně paralelní sekvenování

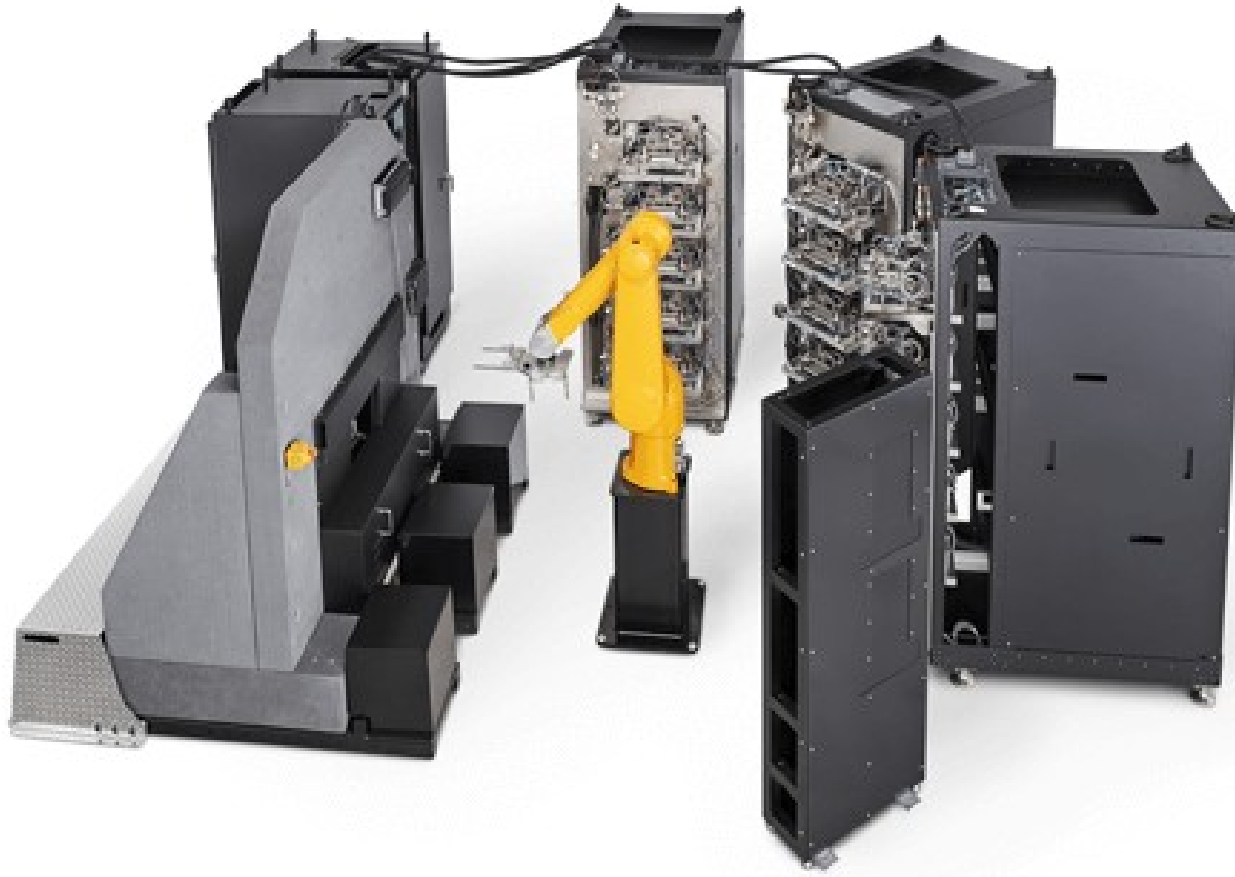
## Technologie cPAL





> 2,5 miliardy jamek na ploše velikosti podložního sklíčka (rozestup 0,7 mikrometru)





Kompletní systém pro sekvenování lidských genomů a exomů

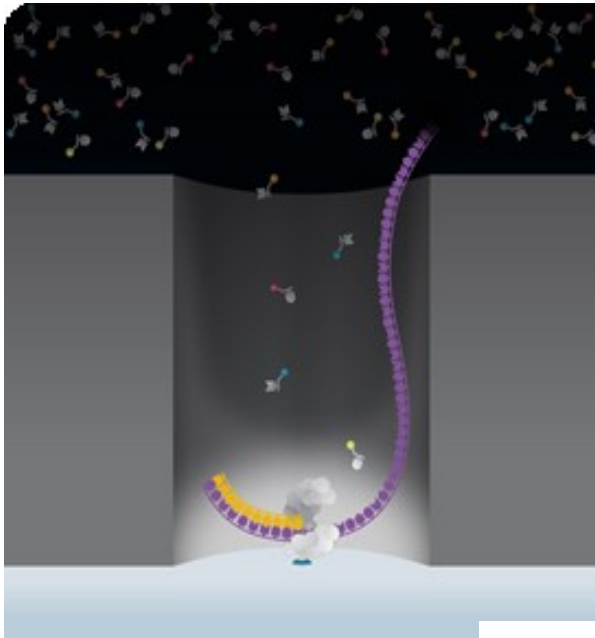
Výrobce BGI (Čína)

Kapacita 12.000 genomů/rok, cena \$ 12M

# Masivně paralelní sekvenování

## Technologie SMRT

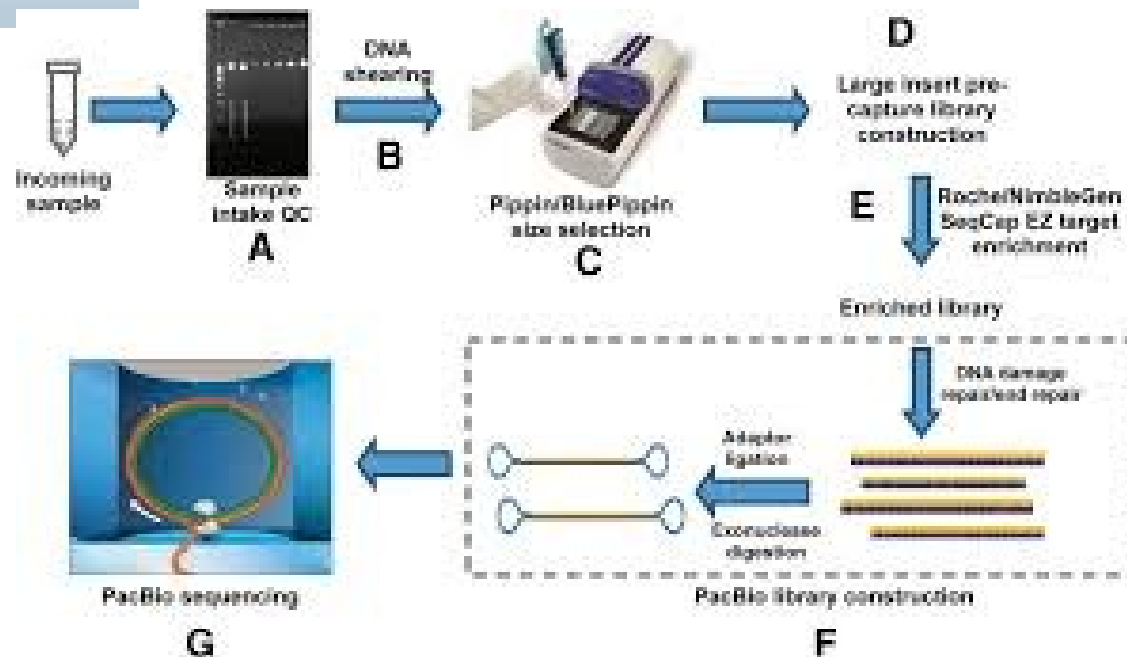
(Single Molecule Real Time)



With an active polymerase immobilized at the bottom of each ZMW, nucleotides diffuse into the ZMW chamber. In order to detect incorporation events and identify the base, each of the four nucleotides A, C, G and T are labeled with a different fluorescent color. Since only the bottom 30nm of the ZMW is illuminated, only those nucleotides near the bottom fluoresce.

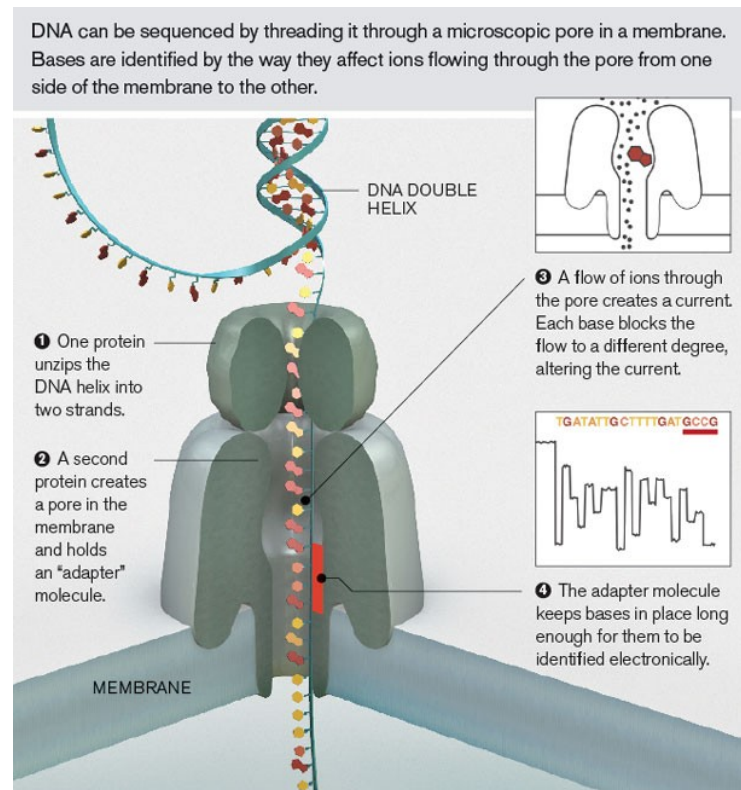
<http://www.pacb.com/smrt-science>

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





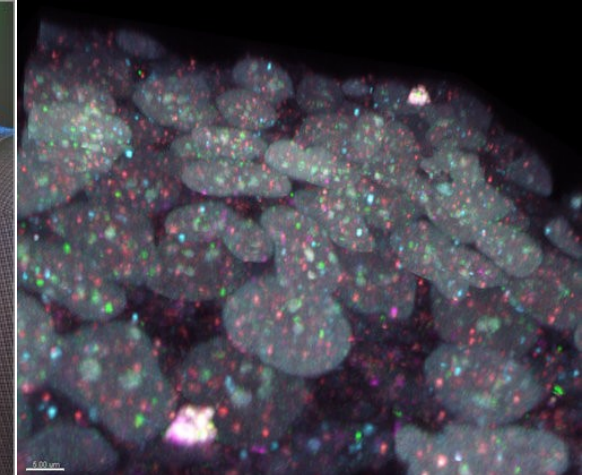
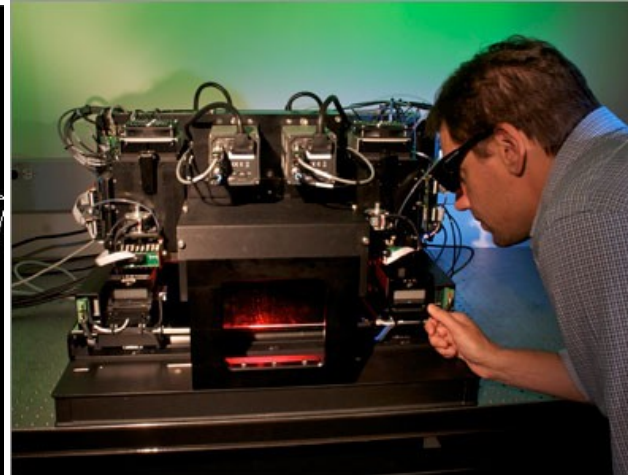
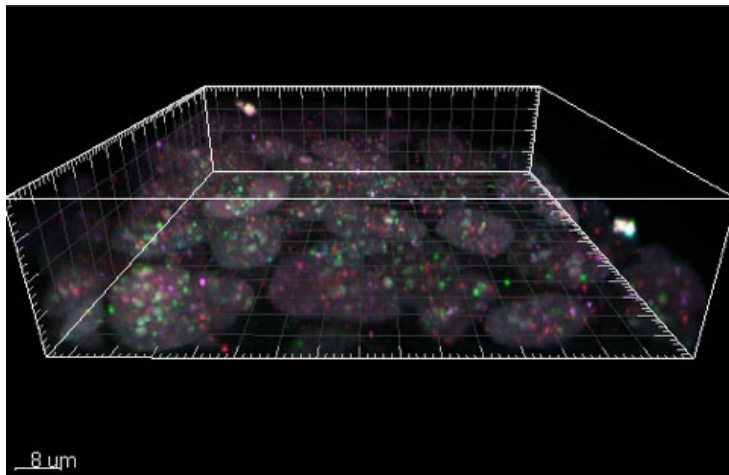
'Strand sequencing' is a technique that passes intact DNA polymers through a protein nanopore, sequencing in real-time as the DNA translocates the pore.



<https://www.nanoporetech.com>

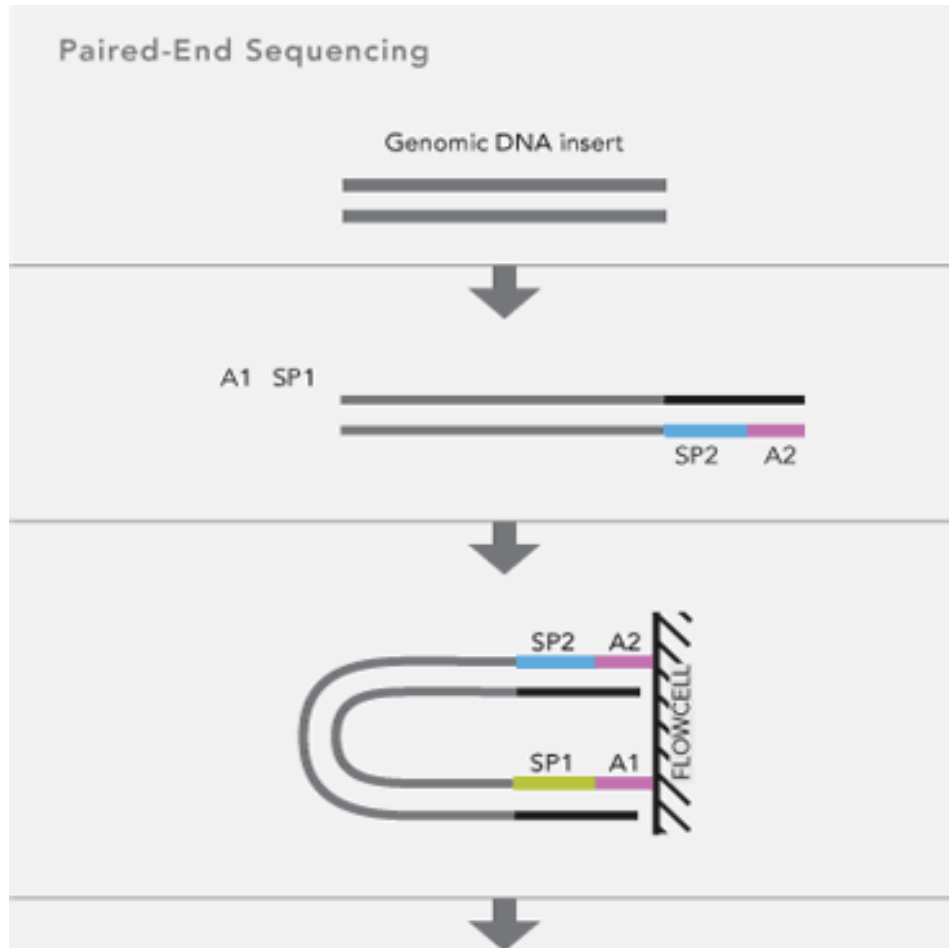
<https://www.youtube.com/watch?v=CE4dW64x3Ts>

### Fluorescent In Situ Sequencing

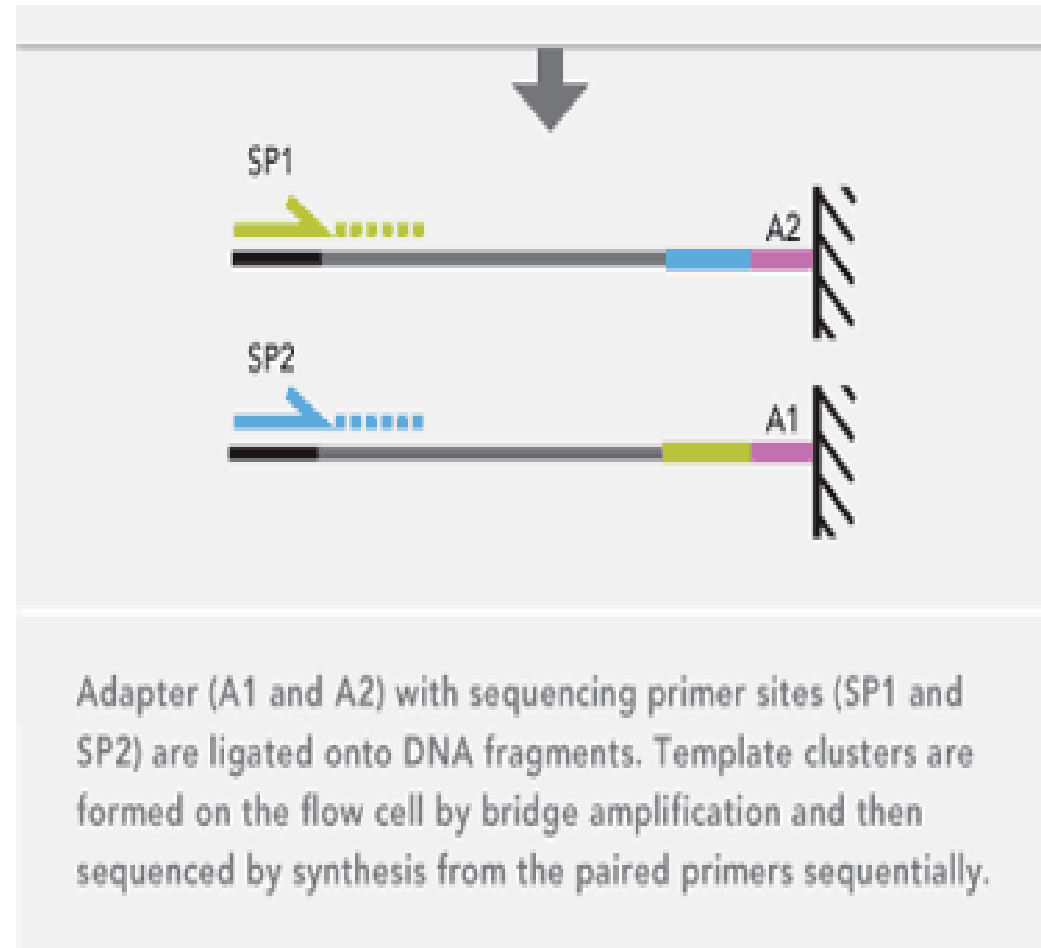


# Masivně paralelní sekvenování

## Paired-end sequencing

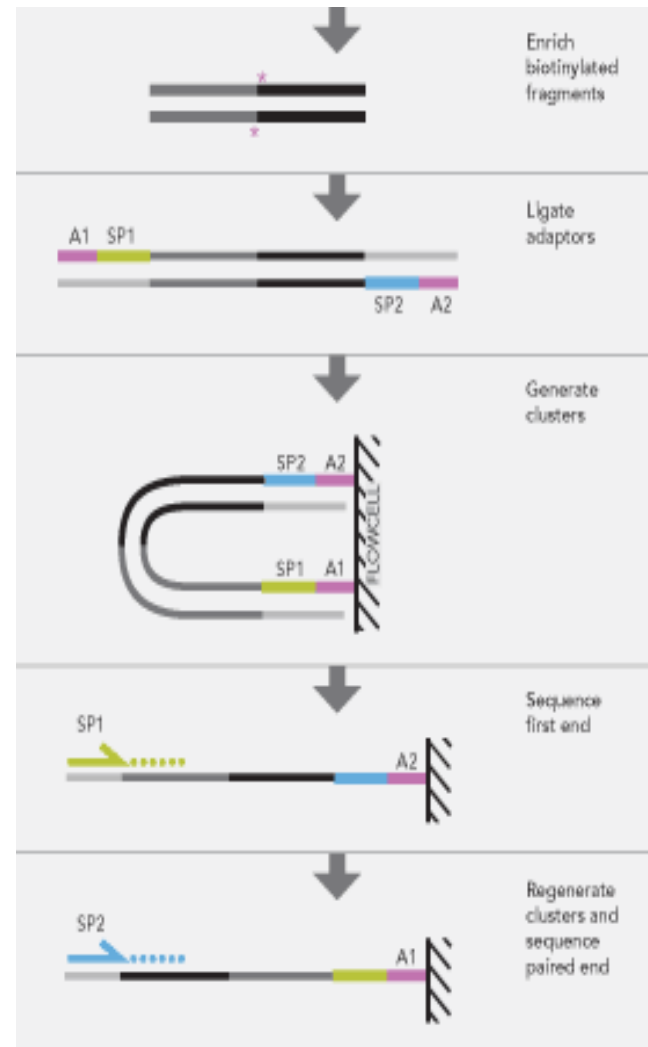
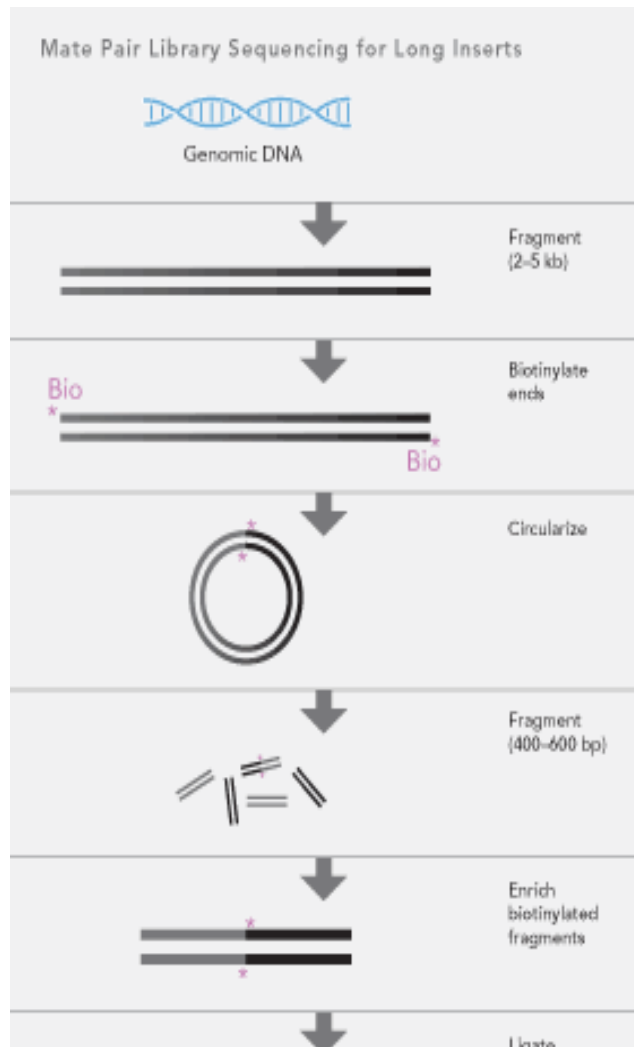


## Illumina



# Masivně paralelní sekvenování

## Mate-pair sequencing



## Illumina

Mate Pair library preparation is designed to generate short fragments that consist of two segments that originally had a separation of several kilobases in the genome. Fragments of sample genomic DNA are end-biotinylated to tag the eventual mate pair segments. Self-circularization and refragmentation of these large fragments generates a population of small fragments, some of which contain both mate pair segments with no intervening sequence. These Mate Pair fragments are enriched using their biotin tag. Mate Pairs are sequenced using a similar two-adaptor strategy as described for paired-end sequencing.

# Aplikace nových technologií

## Analýza DNA

Celogenomový screening

Sekvence

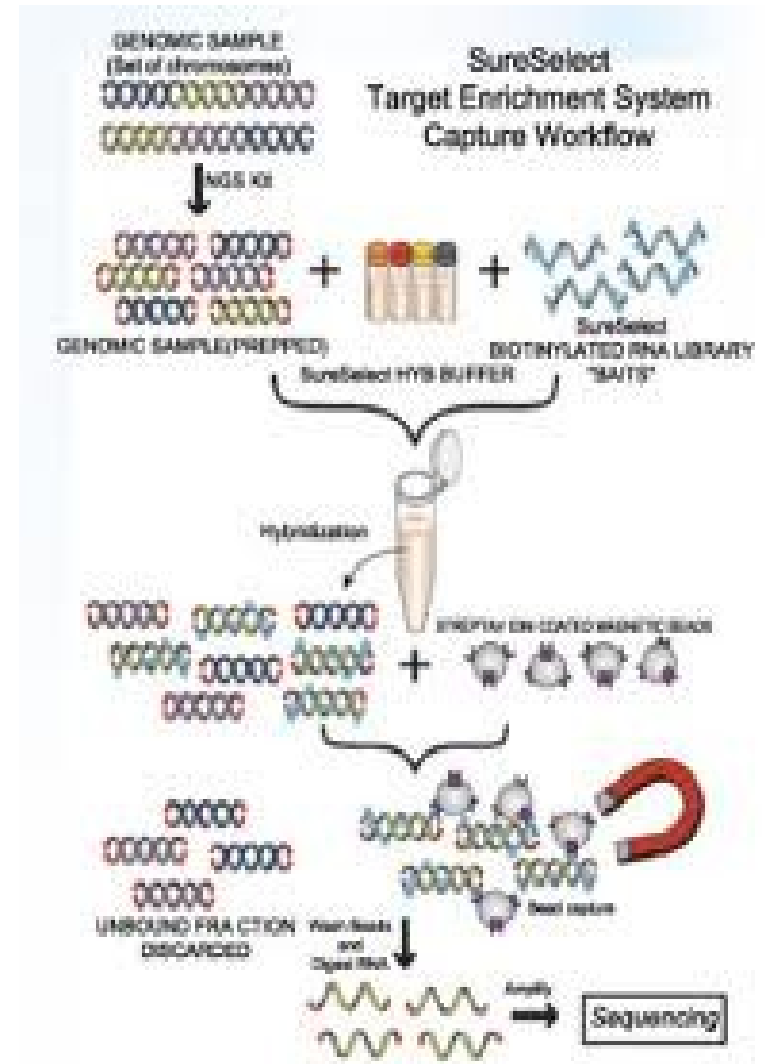
SNP

Strukturní aberace, početní aberace

Cílený screening

Sekvence

SNP



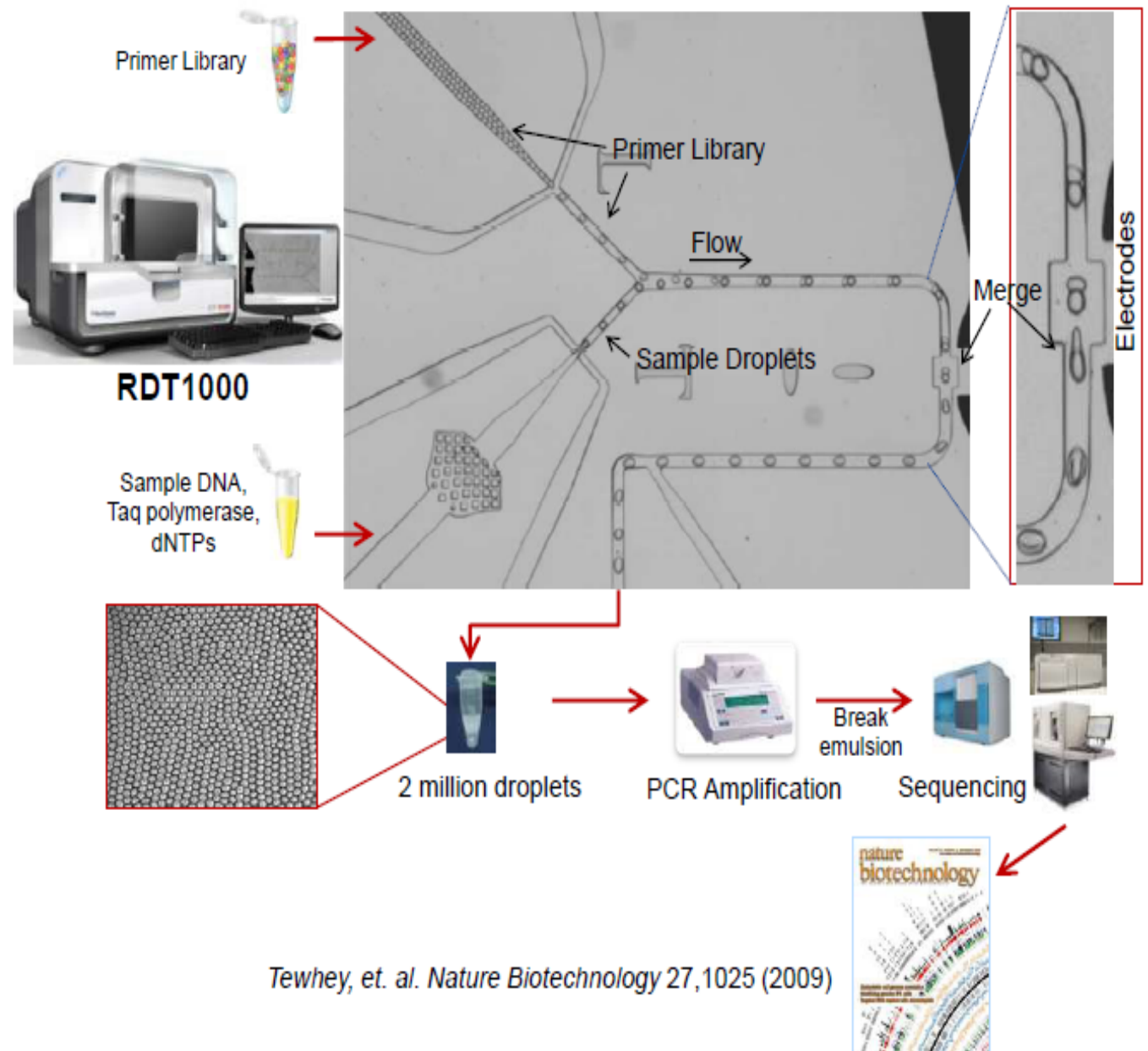
# Cílený screening

Target enrichment

Hybridizace  
na čipu  
v roztoku

PCR  
běžná  
mikrofluidní

## Targeted Sequencing Workflow Using the RDT1000



Tewhey, et. al. *Nature Biotechnology* 27,1025 (2009)



# Cílený screening

Exome sequencing

Všechny exprimované geny

Většinou včetně nekódujících

Hybridizace (v roztoku)

Gene enrichment

Jeden gen – např. dědičné poruchy

PCR, hybridizace

multiplexing

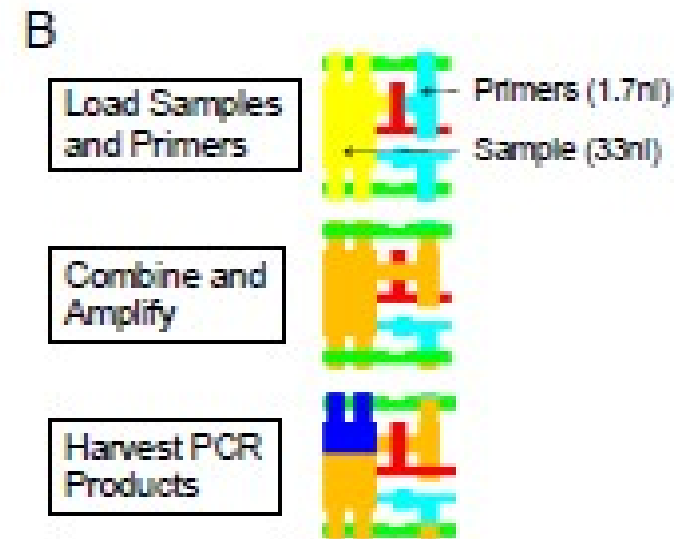
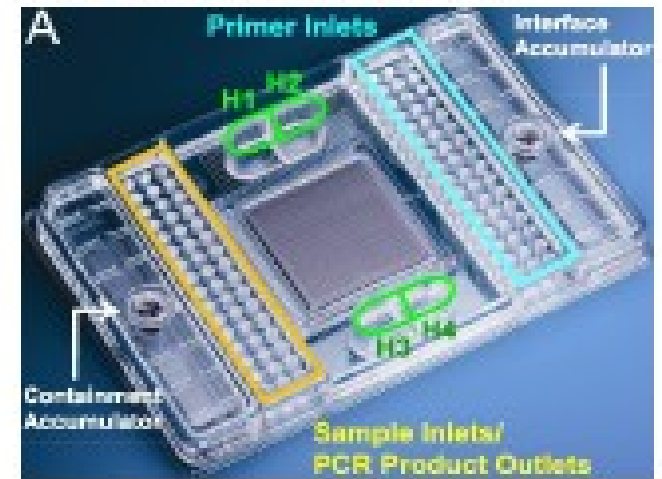
Skupiny genů – např. multifaktoriální nemoci, nádory

PCR, hybridizace

Úseky genomu – strukturní aberace

hybridizace

Figure 1: The Access Array System



# RNA-Seq

Transcriptome sequencing

Single-end – kvantifikace

Paired-end – struktura transkriptů

Tag sequencing

3' tagy, kvantifikace, bez informace o struktuře

Degradome sequencing

5' tagy, identifikace cílů microRNA

Small RNA sequencing

MicroRNA kvantifikace i de-novo identifikace

SAGE (Serial Analysis of Gene Expression)

Konkatemery tagů, původně Sanger sekvenace

RIP (RNA ImmunoPrecipitation)

Imunoprecipitace, RNA vázající proteiny

# Epigenomika/epigenetika

In biology, and specifically genetics, epigenetics is the study of heritable changes in phenotype (appearance) or gene expression caused by mechanisms other than changes in the underlying DNA sequence, hence the name epi- (Greek: επί- over, above) -genetics. These changes may remain through cell divisions for the remainder of the cell's life and may also last for multiple generations. However, there is no change in the underlying DNA sequence of the organism;<sup>[1]</sup> instead, non-genetic factors cause the organism's genes to behave (or "express themselves") differently.

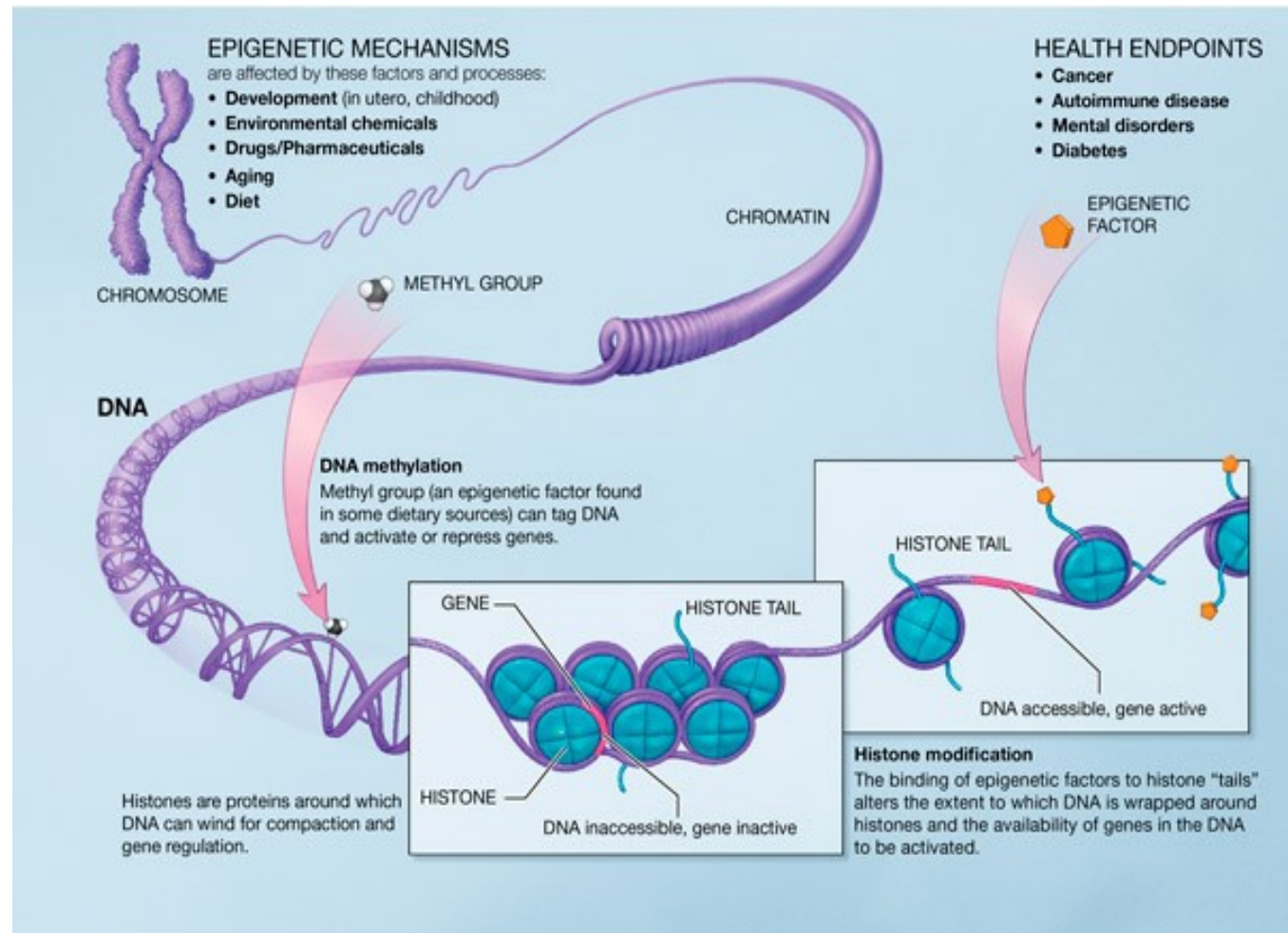
# Epigenomika

DNA metylace

C → Met-C, snížená exprese

Modifikace histonů

Aktivní I neaktivní chromatin



# Chromatinová imunoprecipitace

Modifikované histony

Acetylované, metylované

Další DNA vázající proteiny

Transkripční faktory

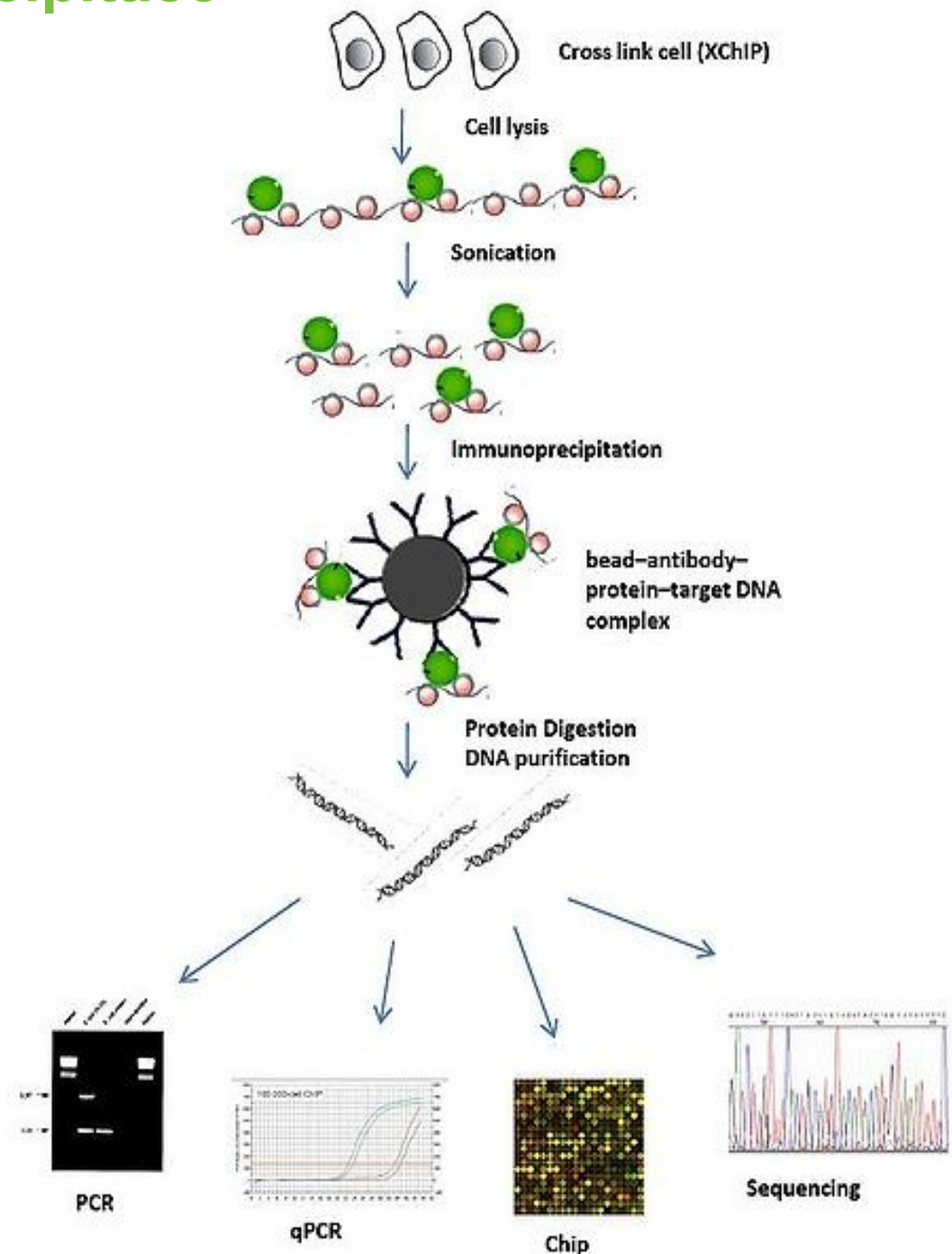
RNA polymerázy

Metylovaná DNA

Kvalita protilátky

Změna epitopu

formaldehyd



# Metylace DNA

MeDIP - Imunoprecipitace metylované DNA

Protilátka rozpoznávající Met-C

Pozice metylovaných úseků DNA

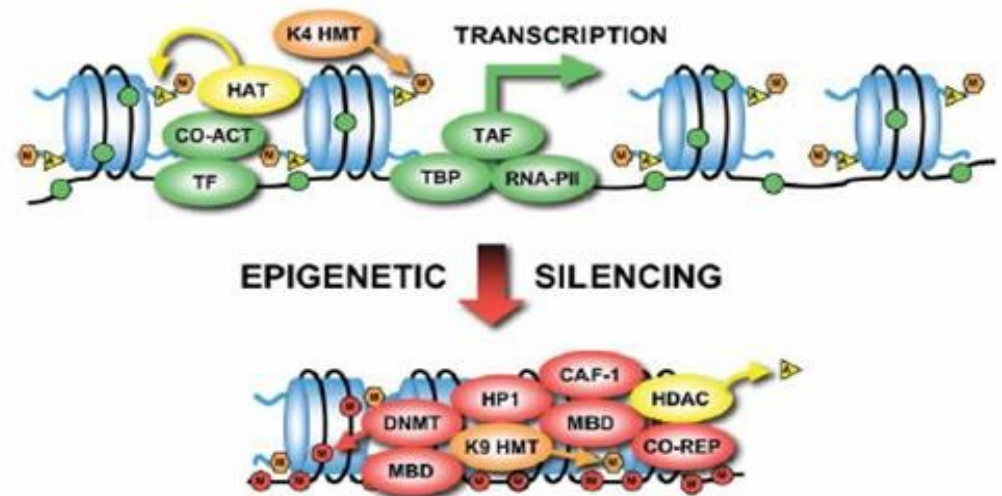
Bisulfite treatment

Konverze C → U, Met-C se nemění

Přesná identifikace jednotlivých metylovaných bazí

Některé dědičné choroby, nádory

*Human Molecular Genetics, 2005, Vol. 14, Review Issue 1*



# Analýza NGS dat

Assembling – vytvoření kontigů

De-novo

Mapování na referenční sekvenci

Identifikace variant

SNP (SNV)

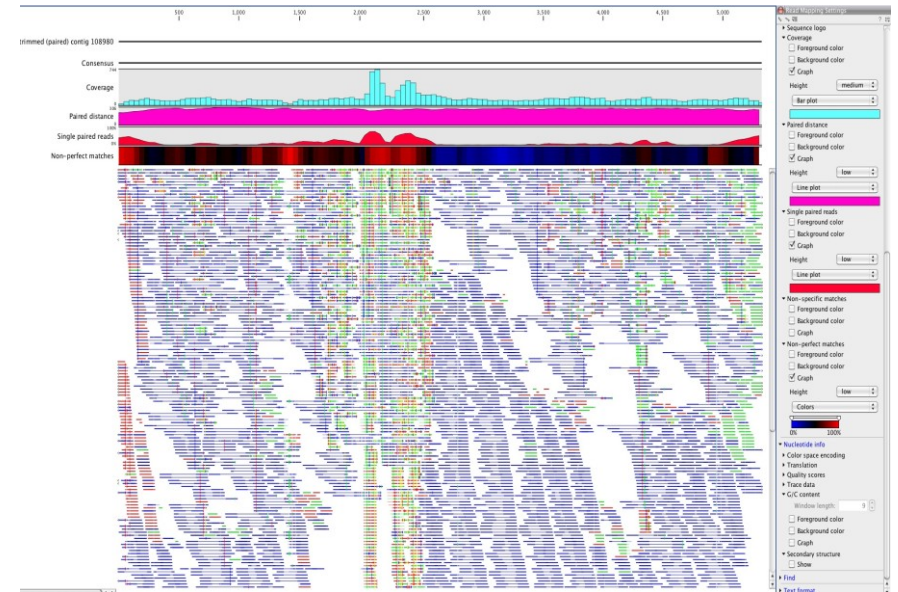
In/del

Strukturní aberace – chromosomy, transkripty

Kvantifikace

DNA – amplifikace/delece, místa vazby transkripčních faktorů

RNA – tagy, exony, transkripty



# Analýza NGS dat

Integrace dat

Databáze

Protein-protein interakce

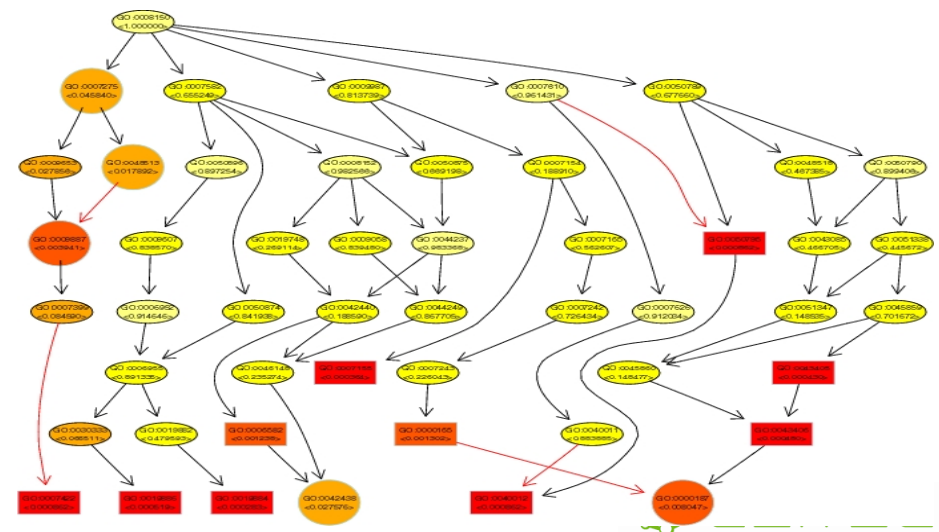
Ontologie – GO (GeneOntology), MeSH terms

Transkripční faktory, microRNA targets

Expresní profily – GSEA, profily chemických látek

Různé typy dat

ChIP + RNA + metylace + DNA copy number





# Analýza NGS dat

Statistická analýza

“klasické” statistické metody

T-test, ANOVA, neparametrické testy

Vícerozměrné analýzy

ClusteCring (HCL, k-means), PCA

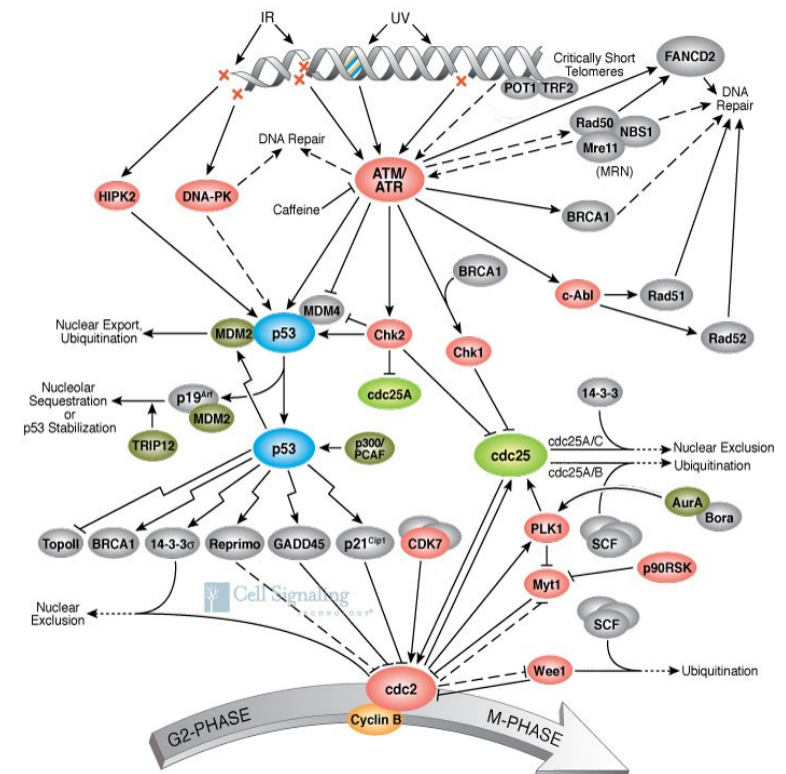
Klasifikační algoritmy

Lineární (LDA), nelineární (KNN)

Vazba s dalšími informacemi (analýza anotací)

Over-reprezentace

Sítě



# Somatické mutace u Chronické Lymfocytární Leukemie

Mutace objevující se během vývoje nemoci mohou vést ke zhoršení průběhu

Mutace TP53 u CLL

Většinou výrazně horší průběh

5-10% případů CLL

Často diagnostikovány až v průběhu nemoci

Asociace s terapií

# Somatické mutace u Chronické Lymfocytární Leukemie

Mutace TP53 diagnostikované v průběhu nemoci po podání terapie

## **Indukce nebo selekce mutací?**

Hledání mutací TP53 ve vzorcích před terapií

Potřeba vysoce citlivé metody (mutace v <1% buněk)

Ultra-deep sekvenování – pokrytí >5000x

Roche GS Junior

# Somatické mutace u Chronické Lymfocytární Leukemie

Výsledek

U 9 z 10 vzorků mutace nalezena i před terapií

0.25-5% sekvencí

## Selekce

Další plány

Rozšíření počtu vyšetřovaných genů

Větší skupina pacientů

# Děkuji za pozornost

Středoevropský technologický institut  
c/o Masarykova univerzita  
Žerotínovo nám. 9  
601 77 Brno, Česká republika

[www.ceitec.cz](http://www.ceitec.cz) | [info@ceitec.cz](mailto:info@ceitec.cz)



EVROPSKÁ UNIE  
EVROPSKÝ FOND PRO REGIONÁLNÍ ROZVOJ  
INVESTICE DO VAŠÍ BUDOUCNOSTI



OP Výzkum a vývoj  
pro inovace

