

# Sekvenování DNA, genomové projekty

- stanovení pořadí nukleotidů v molekule DNA (primární struktura)

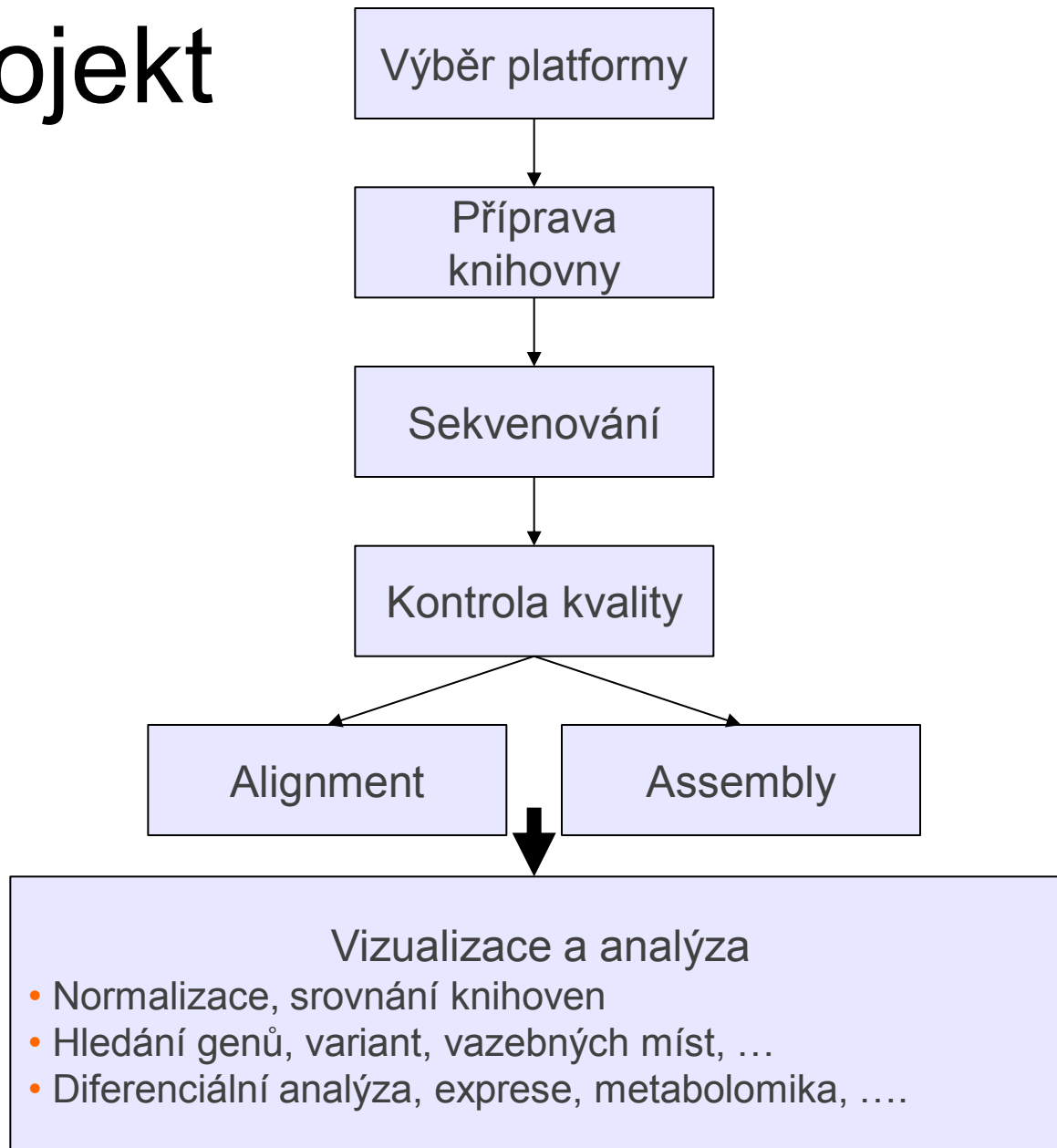
# Důvody řešení genomových projektů

- ◆ Genomika modelových organismů (*de novo* sekvenování)
- ◆ Analýza mutací a SNP, výzkum genetických onemocnění (resekvenování genomů)
- ◆ Identifikace mikroorganismů (*de novo* sekvenování a resekvenování)
- ◆ Analýza transkriptomu (RNAseq)
- ◆ DNA metylační studie (medip-seq, bisulfite treated DNA)
- ◆ Studie transkripčních faktorů, proteinů vázajících se na DNA (ChIPseq)
- ◆ miRNAs, siRNA, piRNA, tRF, aj... (small RNA seq)
- ◆ Metagenomika (16S rRNA, celé metagenomy)

# Vývoj sekvenačních metod

- **Klasické sekvenování kapilární elektroforézou**
  - Vyžaduje velký počet kopií vstupního materiálu DNA jako templátu pro přípravu jednořetězců
- **Sekvenování nové generace**
  - Jako templát slouží jediná molekula, která je amplifikována pro získání dostatečného signálu, produkuje **krátká** čtení, možnost kvantitativní analýzy (SNP)
- **Sekvenování třetí generace**
  - Jako templát slouží jediná molekula. Nevyužívá amplifikace za účelem zvýšení signálu, produkuje **dlouhá** čtení

# Projekt

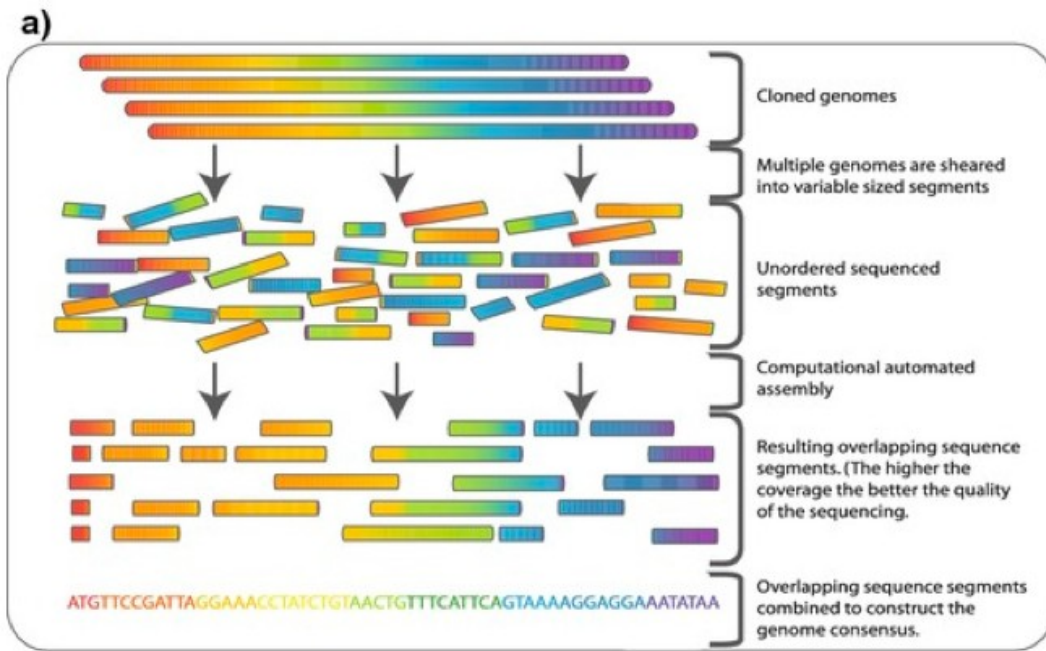


# Základní požadavky pro úspěšné stanovení sekvence Sangerovým sekvenováním:

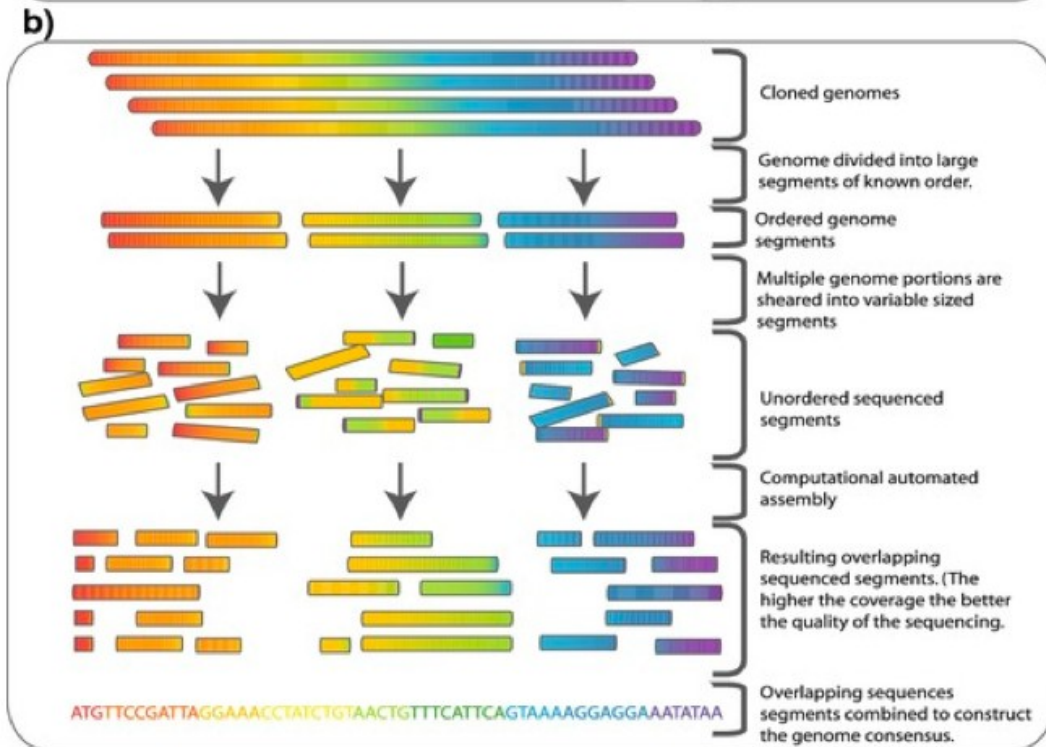
- ❖ **Příprava knihovny pro sekvenování - DNA**  
s přesně definovanými konci
  - s místem pro vazbu primeru
  - s koncovou značkou

Zajistit dostatečné pokrytí při čtení genomu

- ❖ Příprava variant fragmentů ssDNA lišících se svou délkou o jeden nukleotid
- ❖ Přesná separace těchto fragmentů na základě jejich délky, detekce připojené koncové báze, kontinuální čtení a detekce značených bází



Celogenomové  
shotgun (náhodné)  
sekvenování



versus

postupné shot-gun  
sekvenování

# Příprava knihovny pro Sangerovo sekvenování



Izolace DNA



Fragmentace 1300 – 2000 bp

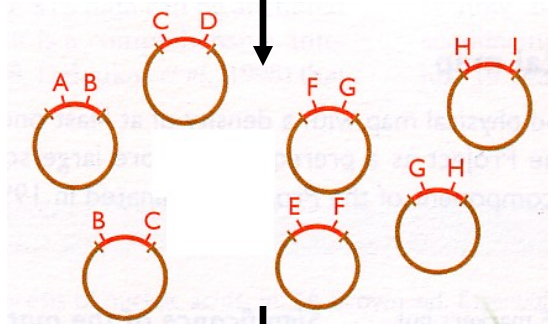


+



Vektor

Úprava konců a klonování



Sestavení překrývajících se klonů



# Příprava knihovny pro náhodné sekvenování genomů

- Při náhodném sekvenování genomu jsou nejdříve připraveny mechanickým stříháním genomové DNA fragmenty s optimální velikostí (1 300 – 2 000 bp) a po úpravě jejich konců jsou nahodile naklonovány do vhodného vektoru za vzniku velkého počtu klonů.
- Ke štěpení DNA se využívá sonikace nebo pankreatická DNáza I za přítomnosti  $Mn^{2+}$ .



- Předpokládá se, že celková informace o sekvenci obsažená v připravených malých klonech odpovídá původní DNA a sekvence fragmentů jednotlivých klonů se vzájemně překrývají.
- Pomocí univerzálních sekvenačních primerů připojujících se k vektoru poblíž klonovacího místa jsou stanoveny sekvence krátkých úseků na koncích klonovaných fragmentů (minimálně 500 bází).
- Stanovené sekvence jsou pak použity k uspořádání klonovaných fragmentů z jednotlivých vektorů do souvislých sekvencí genomové DNA - kontigů.

# Techniky sekvenování

- Klasické techniky:
  - Chemická (Maxamova-Gilbertova)
    - Již se nepoužívá
  - **Enzymová (Sangerova)**
    - V současnosti se používá automatická varianta
- Metody sekvenování nové generace (next generation sequencing, NGS)
  - **454, IonTorrent, Illumina, SoLiD**
- Metody sekvenování třetí generace
  - **PacBio, NanoPore, Helicos, etc.**

# Postup enzymatického sekvenování DNA

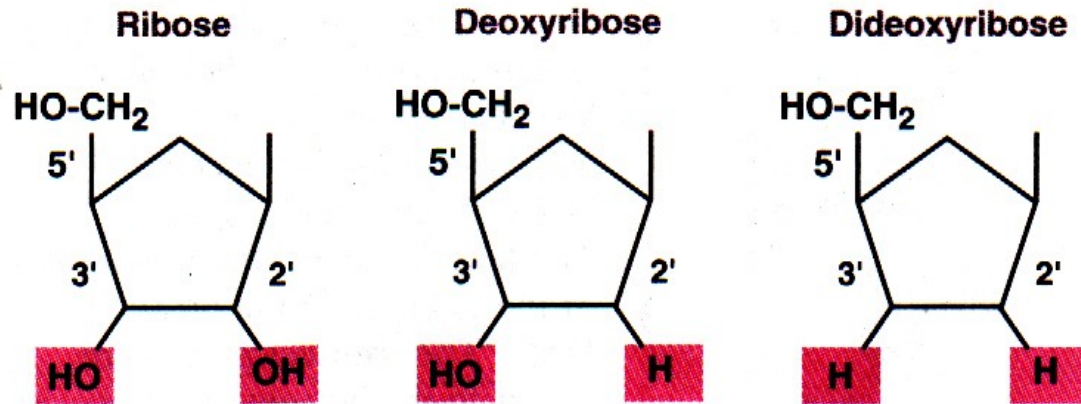
1. Příprava ssDNA jejíž sekvence má být stanovena
2. Rozdělení do 4 vzorků
3. Reakce s DNA polymerázou při níž se začlení do syntetizované DNA místo normálního deoxyribonukleosidtrifosfátu jeho analog, který působí jako koncový inhibitor syntézy DNA - dideoxynukleosidtrifosfát (ddNTP). V každém ze 4 vzorků vzniknou fragmenty, které jsou zakončeny příslušným ddNTP (ddCTP, ddGTP, ddATP a ddTTP).

## Reakce obsahuje:

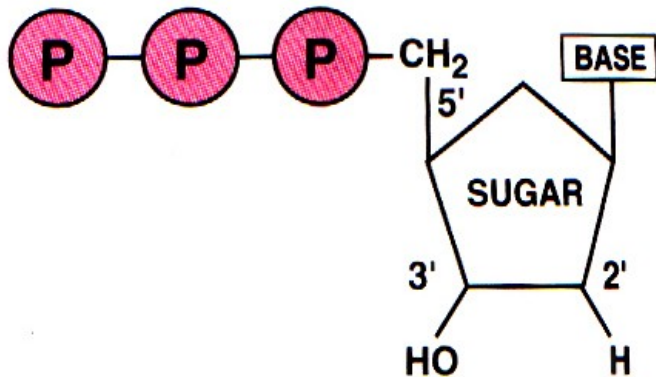
- molekulu DNA
- značený primer při pojující se k části molekuly DNA (místo odkud začínáme stanovovat sekvenci)
- směs obsahující 4 normální nukleotidy
- jeden ddNTP (v každém ze 4 vzorků je jiný)
- DNA polymerázu

4. Denaturace produktů
5. Elektroforetická separace umožňující oddělení fragmentů lišících se délkou o jednu bázi
6. Detekce fragmentů, které nesou označený konec

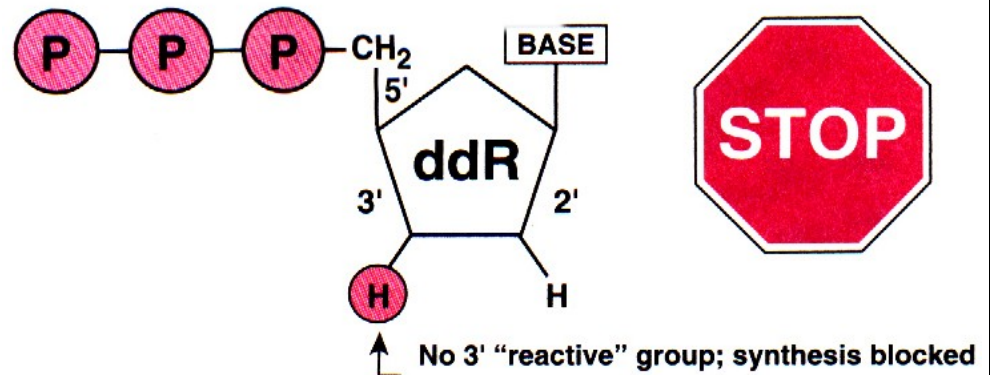
# Dideoxynukleotidy nemají hydroxyl na 3'-konci



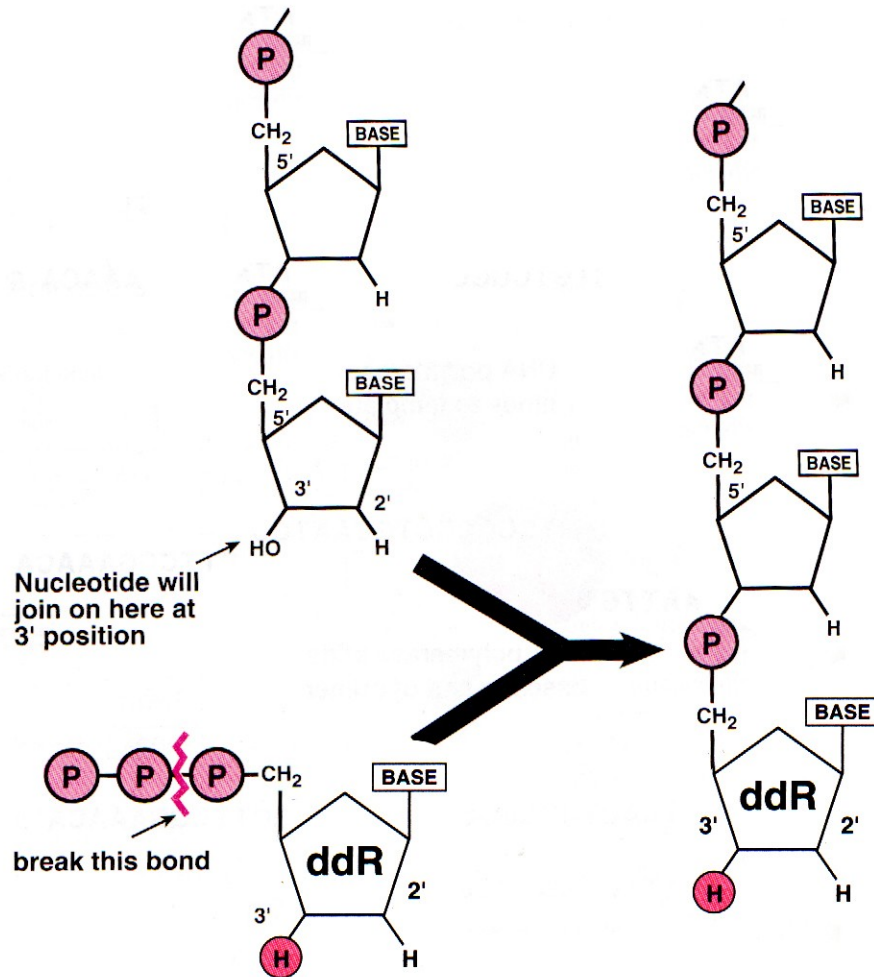
## 23.4 DEOXYNUCLEOSIDE TRIPHOSPHATE



## 23.7 DIDEOXYRIBOSE BLOCKS ELONGATION



Pokud je  
dideoxynukleotid  
inkorporován do  
syntetizujícího se  
řetězce, působí  
jako terminátor  
reakce



# Replikace DNA s normálními deoxynukleotidy

## Původní sekvence

A G C C T G G C G A C C A T C G T  
T C G G A C C G C T G G T A G C A

## Denaturovaný templátový řetězec

3' A G C C T G G C G A C C A T C G T 5'

## Prfektní kopie

A G C C T G G C G A C C A T C G T  
T C G G A C C G C T G G T A G C A

Co se stane pokud při reakci použijete směs dGTP a ddGTP?

Původní sekvence

A G C C T G G C G A C C A T C G T  
T C G G A C C G C T G G T A G C A

Templátový řetězec

3' A G C C T G G C G A C C A T C G T 5'

Perfektní kopie

~~A G C C T G G C G A C C A T C G T  
T C G G A C C G C T G G T A G C A~~

Směs řetězců ukončených GUANINEM  
při použití směsi dGTP a ddGTP

T C G  
T C G G  
T C G G A C C G  
T C G G A C C G C T G  
T C G G A C C G C T G G  
T C G G A C C G C T G G T A G

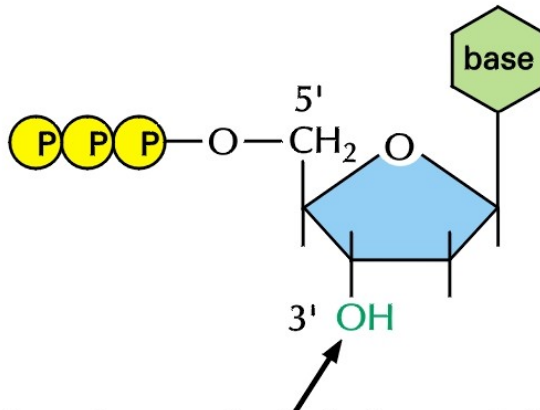


# Enzymová metoda sekvenování DNA (Sanger)

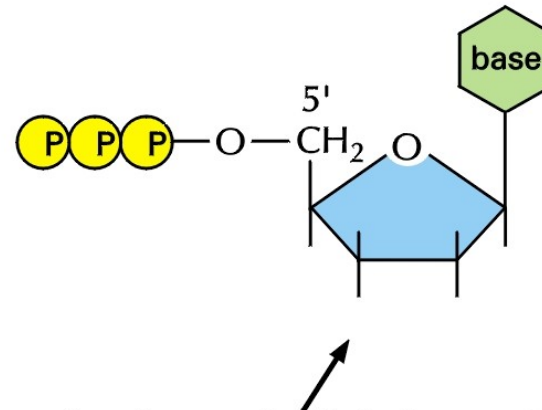
(A)

deoxyribonukleosidtrifosfát

dideoxyribonukleosidtrifosfát



je možno prodloužit řetězec na 3' -konci

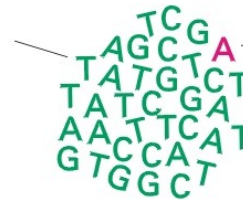


není možno prodloužit řetězec na 3' -konci

(B)

normální prekursorý  
deoxyribonukleosidtri-  
fosfátů (dATP, dCTP,  
dGTP, dTTP)

malé množství jednoho  
dideoxyribonukleosidtrifosfátu  
(ddATP)



oligonukleotidový primer  
pro DNA-polymerázu



vzácná inkorporace  
dideoxyribonukleosidtrifosfátu  
DNA-polymerázou zastaví další  
růst molekuly DNA

jednořetězcová DNA,  
která bude sekvenována

(C)

5' GCATATGTCAGTCCAG 3'  
3' CGTATACAGTCAGGTC 5' } dvouřetězcová DNA

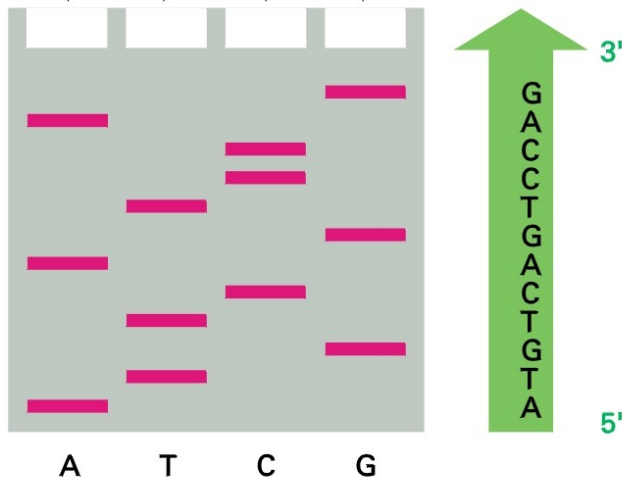
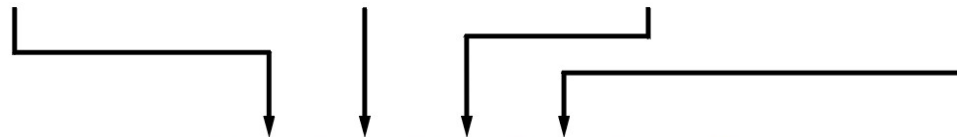
označený primer

5' GCAT 3'  
3' CGTATACAGTCAGGTC 5' } jednořetězcová DNA

+ nadbytek dATP  
dTTP  
dCTP  
dGTP

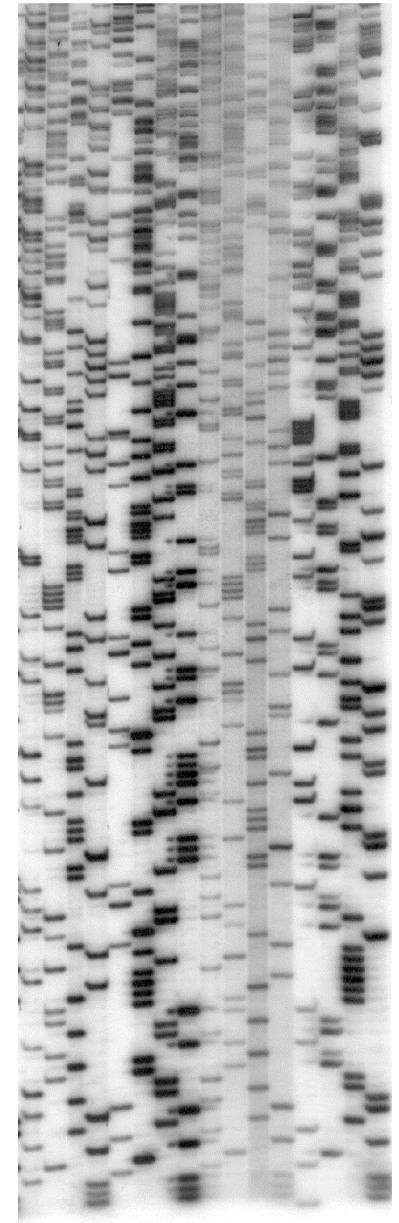
+ ddATP + DNA-polymeráza    + ddTTP + DNA-polymeráza    + ddCTP + DNA-polymeráza    + ddGTP + DNA-polymeráza

GCAT A                      GCAT AT                      GCAT ATGTC                      GCAT ATG  
GCAT ATGTCA                      GCAT ATGT                      GCAT ATGTCAGTC                      GCAT ATGTCAG  
GCAT ATGTCAGTCCA                      GCAT ATGTCAGT                      GCAT ATGTCAGTCC                      GCAT ATGTCAGTCCAG



[seq4.mov](#)

Obraz autoradiogramu ze sekvenačního gelu



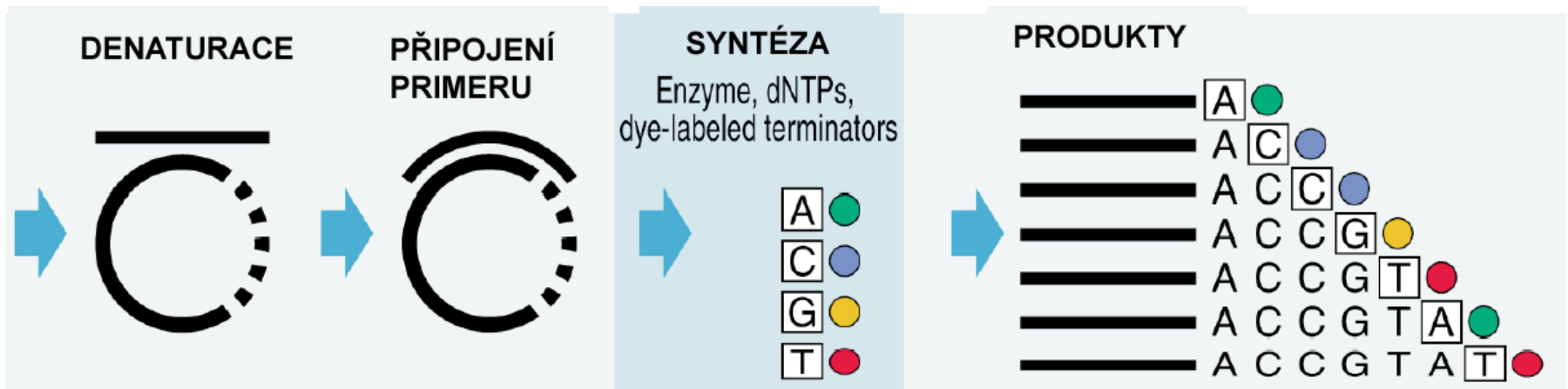
ATGC ATGC ATGC ATGC

# Automatické sekvenování DNA

- Je variantou enzymatického sekvenování DNA.
- Syntéza DNA probíhá v jedné reakci
- Ke značení produktů se používají čtyřmi různými fluorescenčními značkami označené
  - primery
  - dideoxyribonukleotidy

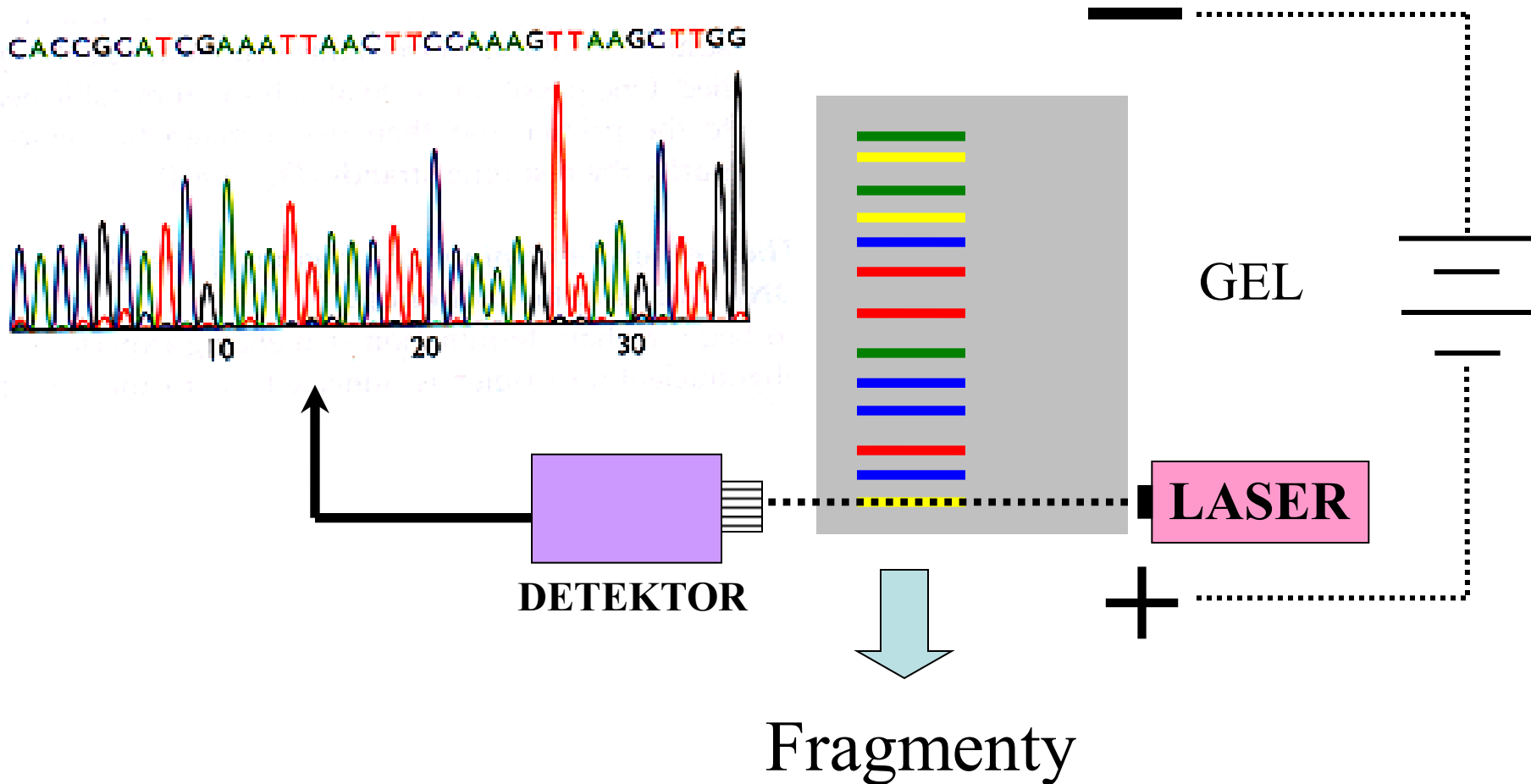
# Strategie barevných terminátorů

asymetrická PCR



# Princip detekce produktů

- Detekce produktů probíhá během elektroforézy pomocí laserového detektoru (laserem indukovaná fluorescence, LIF) napojeného na počítač, který vyhodnocuje sekvenci.

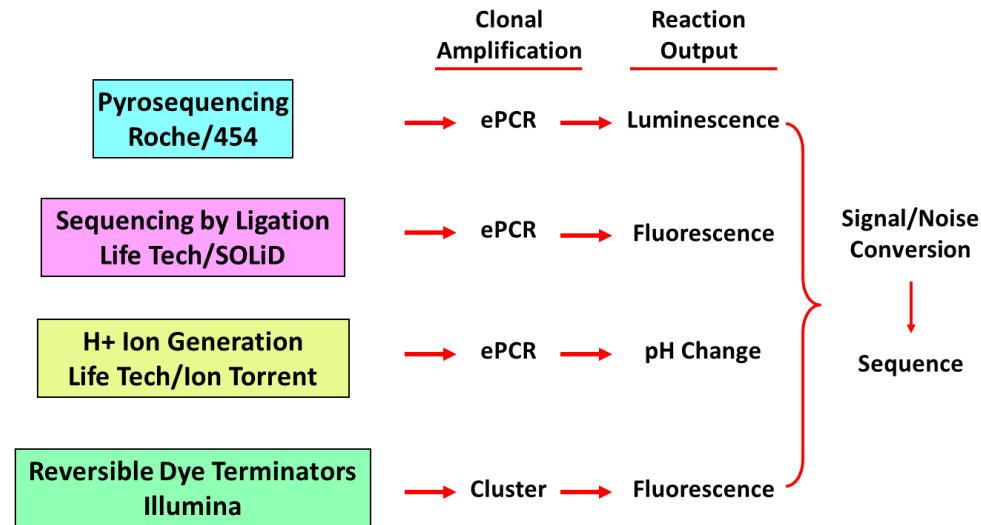


# Uspořádané sekvenování

- Pro doplnění sekvence mezer zbývajících po náhodném sekvenování
- Pro sekvenování malých fragmentů DNA
- Dva odlišné přístupy:
  - **procházení primerem**
    - vyžaduje znalost sekvence, ke které se připojuje primer, který umožní prodloužení řetězce DNA-polymerázou.
    - získaná sekvence z první reakce je použita pro návrh primeru pro další reakci a tento krok se opakuje, dokud není dosaženo kompletního stanovení sekvence.
    - Tento proces nevyžaduje další klonování a minimalizuje stanovení nadbytečných sekvencí, ale správnost stanovené sekvence by měla být ověřena sekvenováním **obou řetězců**.
  - **postupné zkracování fragmentu exonukleázou a tvorba sousedících delecí**

# Metody sekvenování nové generace (next generation sequencing, NGS)

- Liší se v provedení a chemii
- V základu se podobají v sekvenování tisíců až miliónů klonálně amplifikovaných molekul (masivní paralelní sekvenování)
- Probíhá vývoj technik
  - Více čtení
  - Delší čtení
  - Rychlejší sekvenování
  - Nižší cena
- Nové aplikace v klinické oblasti, vývoj kitů cílených na určité části genomů (dědičná onemocnění, nádorová onemocnění, metagenomika, 16S rRNA)

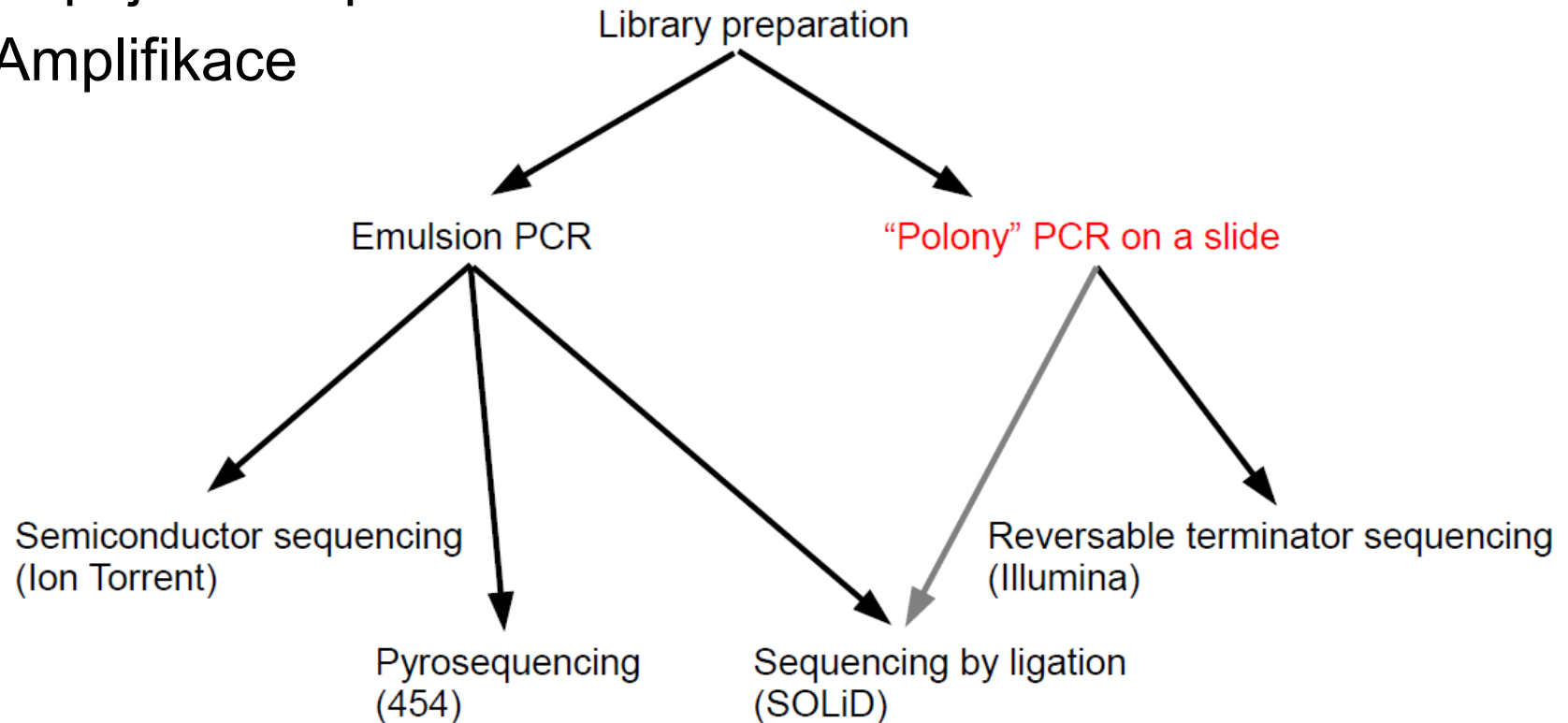


## • Sekvenování třetí generace

- Pacific Biosciences
- Helicos Biosciences
- NABsys
- VisiGen Biotechnologies
- Oxford Nanophore Technologies

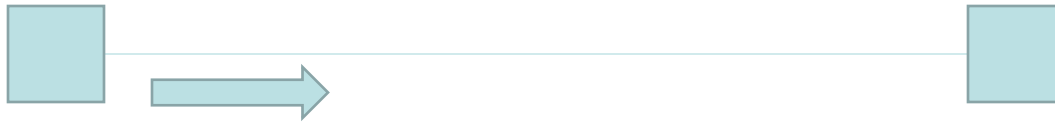
# Metody přípravy knihovny

- Fragmentace
  - Enzymatická
  - Fyzikální (sonikace, nebulizace, Hydroshear)
- Připojení adaptorů
- Amplifikace





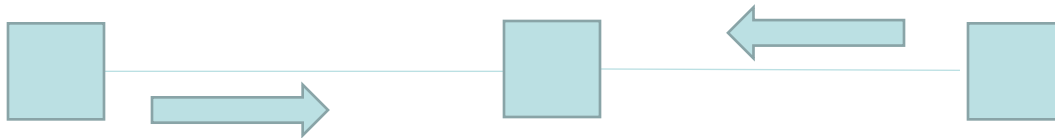
# Typy uspořádání NGS sekvenování



Single Read



Paired-end read



Mate-pair read

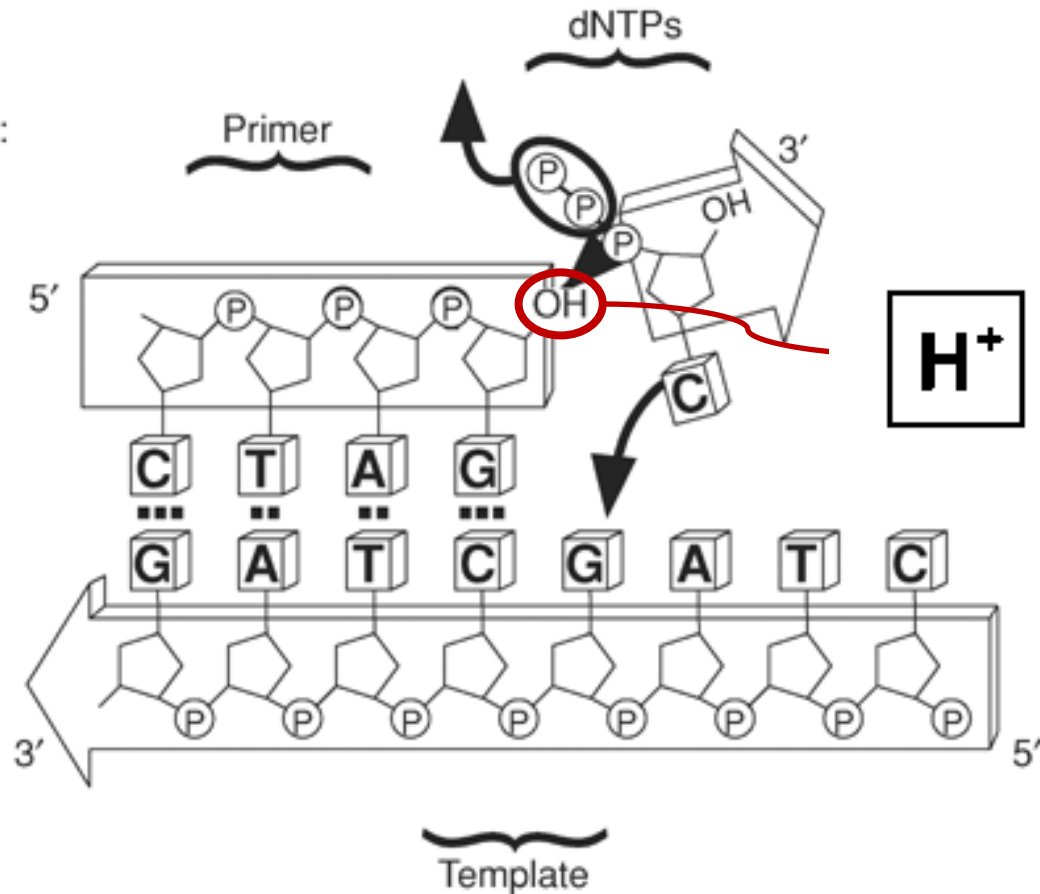
# Ion Torrent polovodičové sekvenování

- prvním sekvenátor DNA, který **nepoužívá pro detekci sekvenovaných bází DNA světelný signál, který by byl uvolňovaný při enzymatických reakcích s fluorescenčními nebo chemiluminescenčními substráty.**
- Sekvenování Ion Torrent je založené na elektrochemické detekci vodíkových iontů (pH) které jsou uvolněny při replikaci sekvenovaného templátu na speciálním polovodičovém čipu.

# Uvolnění vodíku v nepufřujících podmínkách

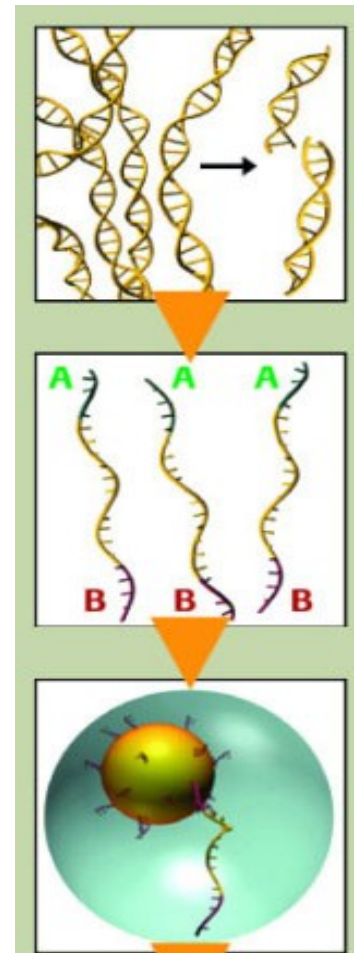


Example:



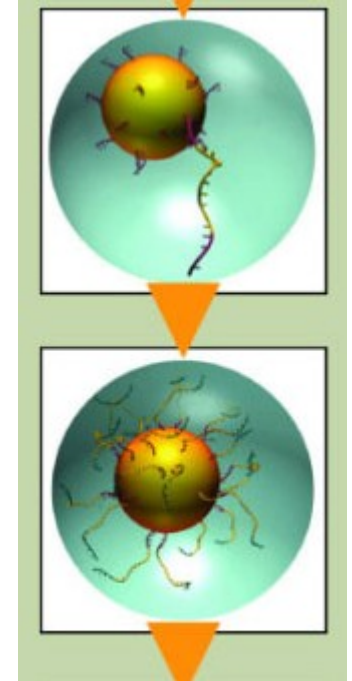
# Knihovna

- Příprava jednořetězcové DNA knihovny
  - Umožňuje zpracovat různé typy vzorků
    - genomové DNA
    - produkty PCR
    - BACs
    - cDNA
  - Frakcionace dlouhých úseků na 200- až 400-bp dlouhé fragmenty
  - Vazba dvou adaptorů specifických pro 3'- a 5'-konce na každý fragment
    - Adaptor A obsahuje vazebné místo pro primer
    - Adaptor B B obsahuje 5'-biotinovou značku, která slouží k imobilizaci DNA knihovny na mikrosféry pokryté streptavidinem



# Klonální amplifikace knihovny emulzní PCR

- Klonální amplifikace knihovny
  - **emulzní PCR**, amplifikační reagensie ve směsi voda-olej → mikroreaktory obsahující jedinou kuličku s unikátním fragmentem DNA z knihovny - ( $10^7$ ) kopií
- Separace kuliček a výběr pouze těch, které obsahují amplifikovanou DNA (enrichment)
- Sekvenační reakce
- Analýza dat a assembly



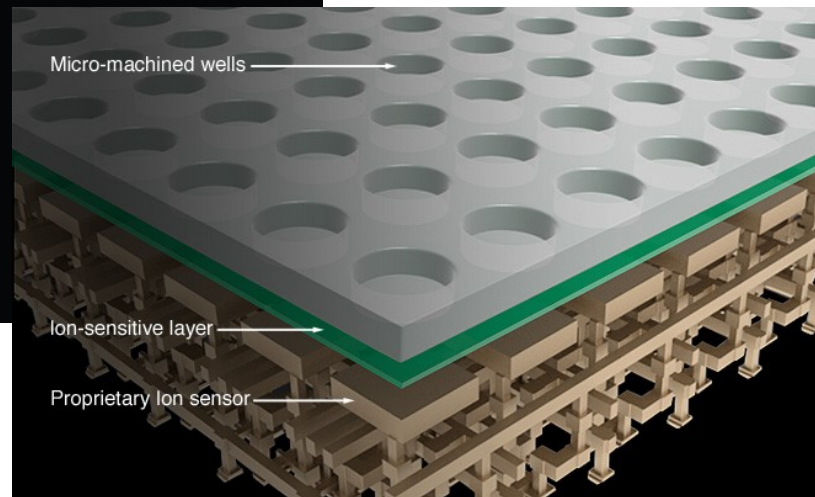
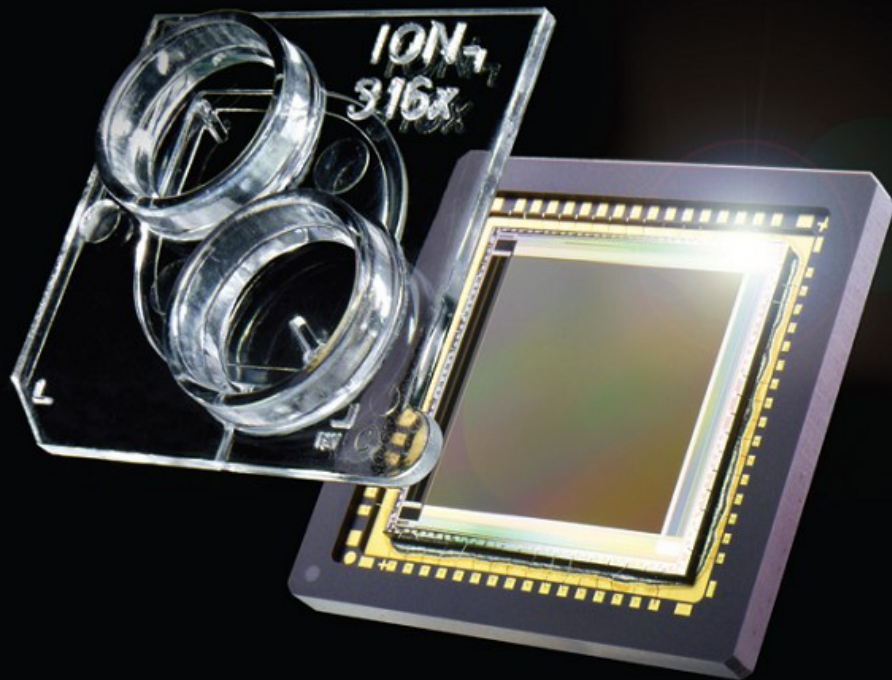
# IonTorrent polovodičové sekvenování

- Na čipu se nachází milióny jednotlivých reakčních cell (reaktorů), ve kterých probíhá sekvenační analýza individuálních kopií DNA na základě kontinuálního střídání reakčních složek (jednotlivých typů nukleotidů) v mikrofluidním systému.
- DNA imobilizovaná na mikrosférách (po emulzní PCR)
- Kapacita závisí na typu použitého polovodičového chipu
  - 10 – 65 miliónů
  - 100 – 600 miliónů
  - > 1000 miliónů

# Ion Torrent polovodičové sekvenování

- Technologie Ion Torrentu je otevřená pro sekvenování DNA de-novo i cílené sekvenování vybraných oblastí genomu
- Minimální délka sekvenačního čtení
  - V jedno směru: 400bp
  - Obousměrné (Paired end)
  - Počet vzorků analyzovaných v jednom běhu
  - až 96 vzorků – pomocí označení tzv. barkodem
- Doba sekvenační analýzy
  - 2 - 3,5 hodiny – podle použitého čipu

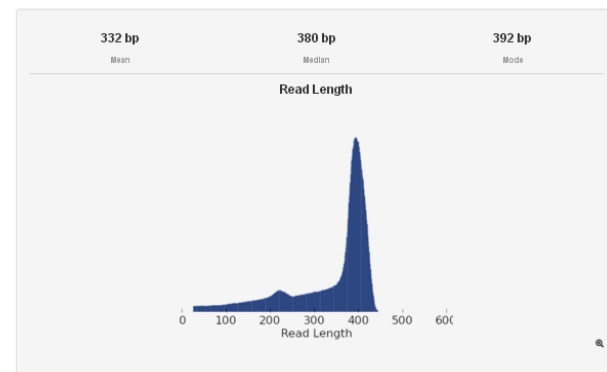
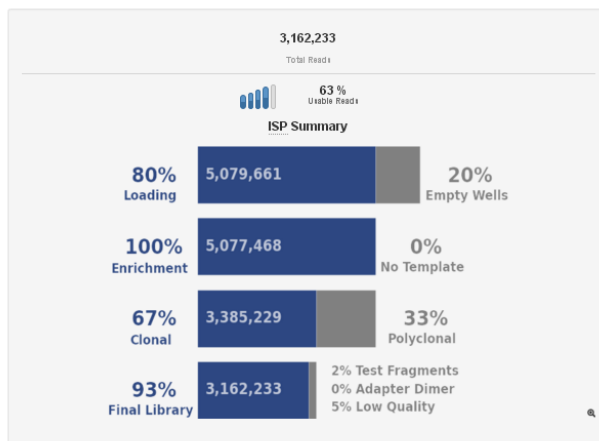
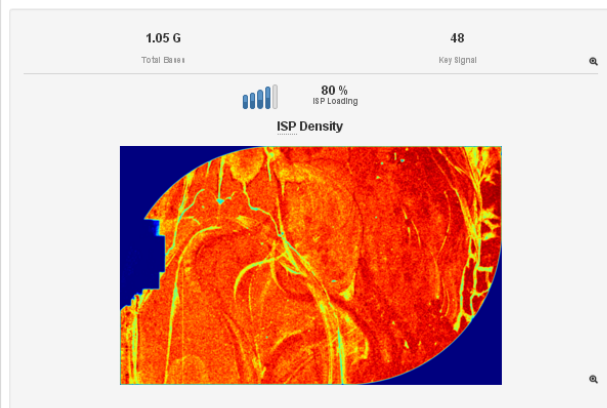
# IonTorrent čip



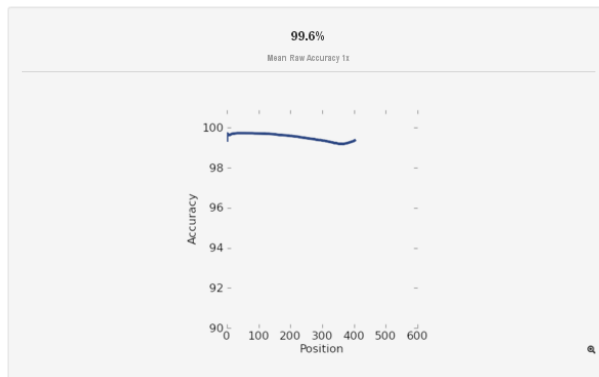
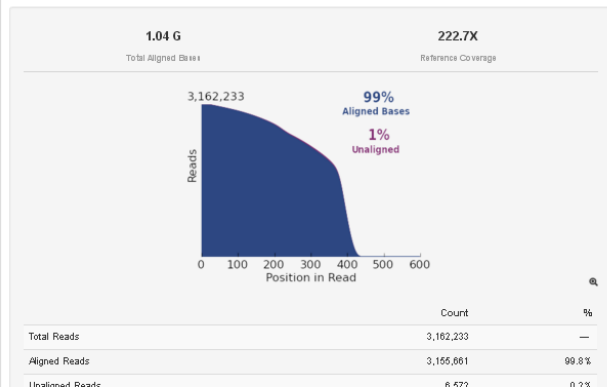


# Výstup z Torrent server

## Read Summary: Unaligned



## Aligned to E. coli DH10B



1.03 G AQ17 Total Bases

Alignment Quality

	AQ17	AQ20	Perfect
Total Number of Bases [bp]	1.03 G	1 G	780 M
Mean Length [bp]	332	325	259
Longest Alignment [bp]	499	499	472
Mean Coverage Depth [x]	220.3	213.8	166.4

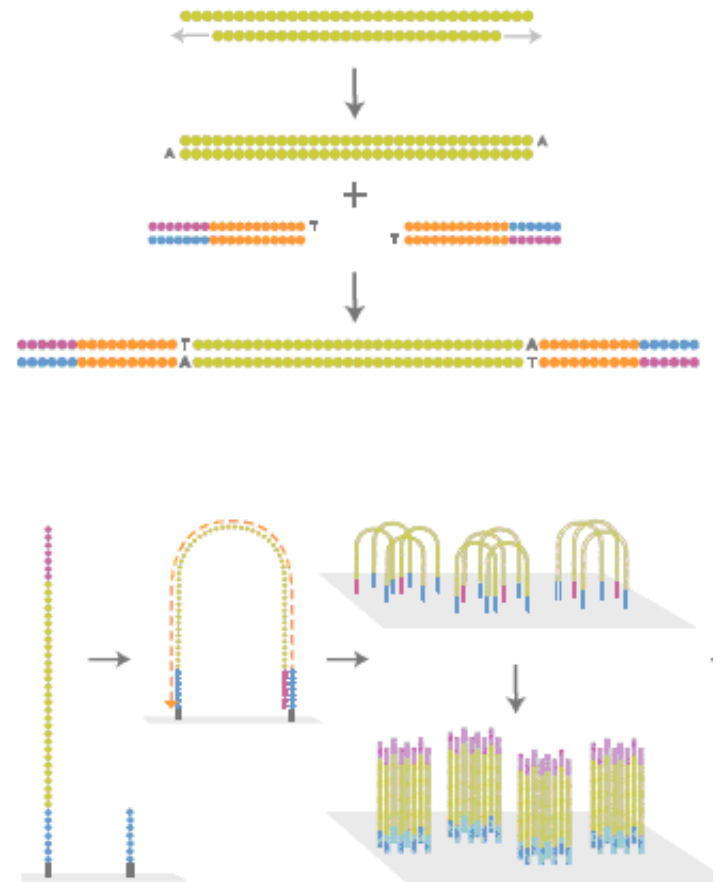


# Technologie firmy Illumina/Solexa

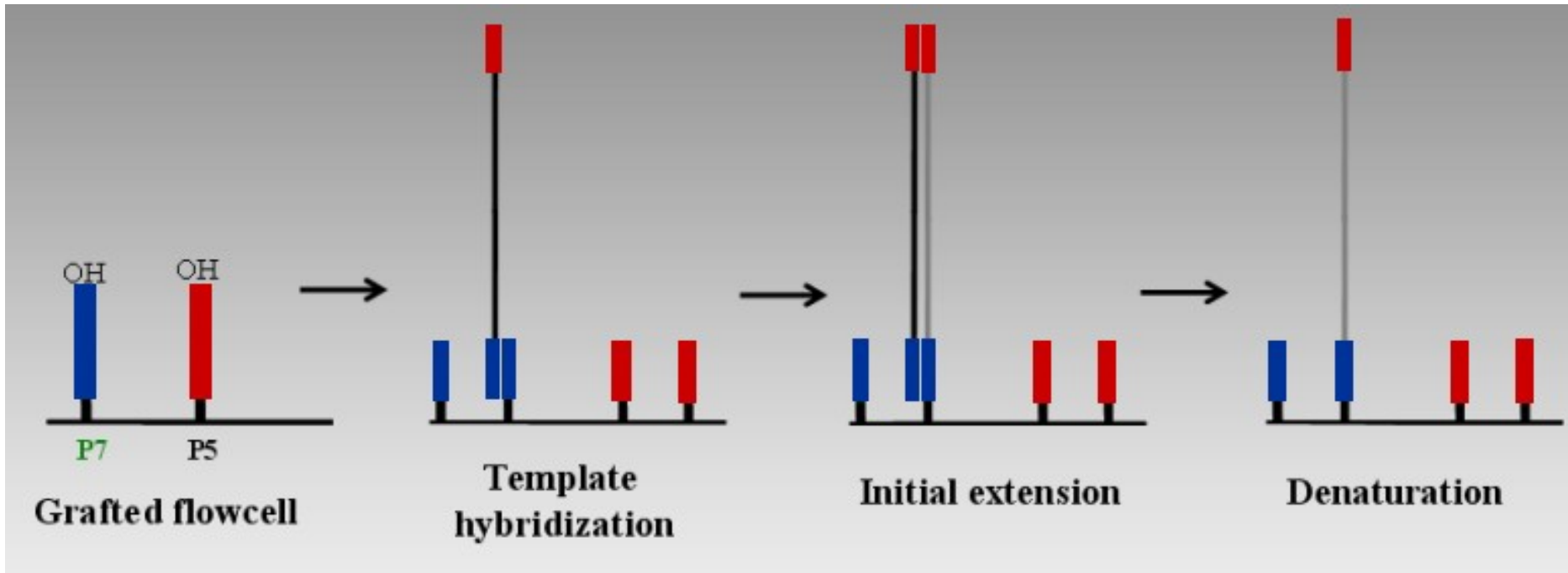
- Uvedena na trh 2006
- Prošla řadou inovací, zejména délky čtení
- Masivní paralelní sekvenování milionů fragmentů DNA („polony“ sequencing)
- Sekvenacování syntézou na čipu (flow cell), které využívá strategii reverzibilních terminátorů
- Nevyžaduje klonování ani přípravu knihovny na mikrosférách
- Dosahuje 99,9 % přesnosti

# Příprava knihovny – polony PCR

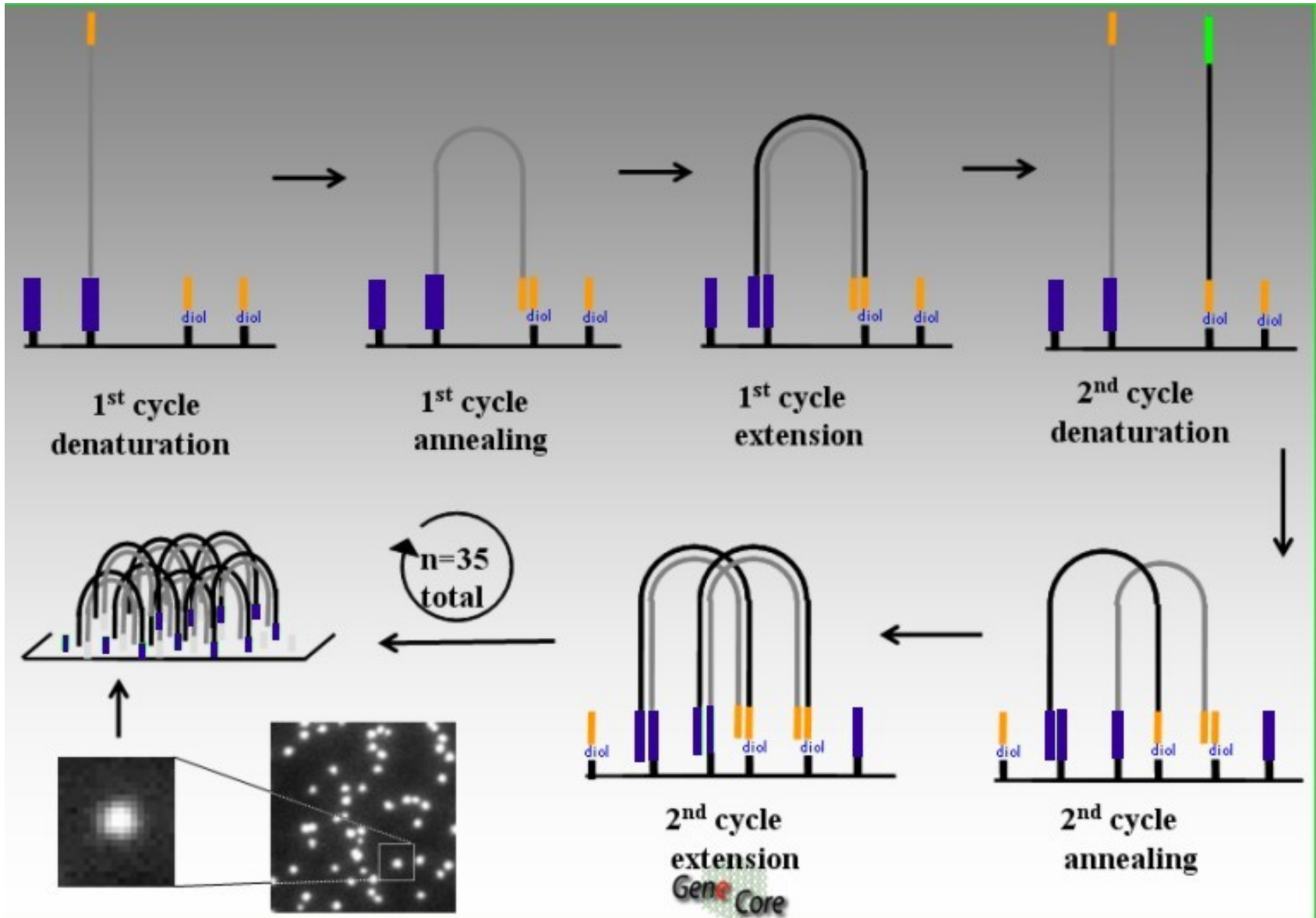
- Náhodná fragmentace DNA na krátké úseky 200 – 500 bp
- Zatupení konců a vytvoření 1 nt A-přesahu na 3'-konci (T4 DNA polymeráza, Klenow a polynukleotid-kináza T4)
- Navázání adaptorů umožňujících kovalentní vazbu na opticky transparentní povrch pro sekvenování
- Každý fragment je na povrchu uchycen pouze jedním koncem, po přidání enzymů pro amplifikaci dochází k ohnutí fragmentu do „mostu“.
- Výsledkem amplifikace jsou dva řetězce, každý s jedním volným a jedním pevným koncem
- Po denaturaci jsou fragmenty narovnány a uspořádány do shluků, ve kterých je dosažena značná hustota, až 1000 kopií fragmentu na  $\mu\text{m}^2$  povrchu
- Na celém povrchu je dosaženo hustoty deseti milionů shluků na  $\text{cm}^2$



Bridge amplification:  
initiation

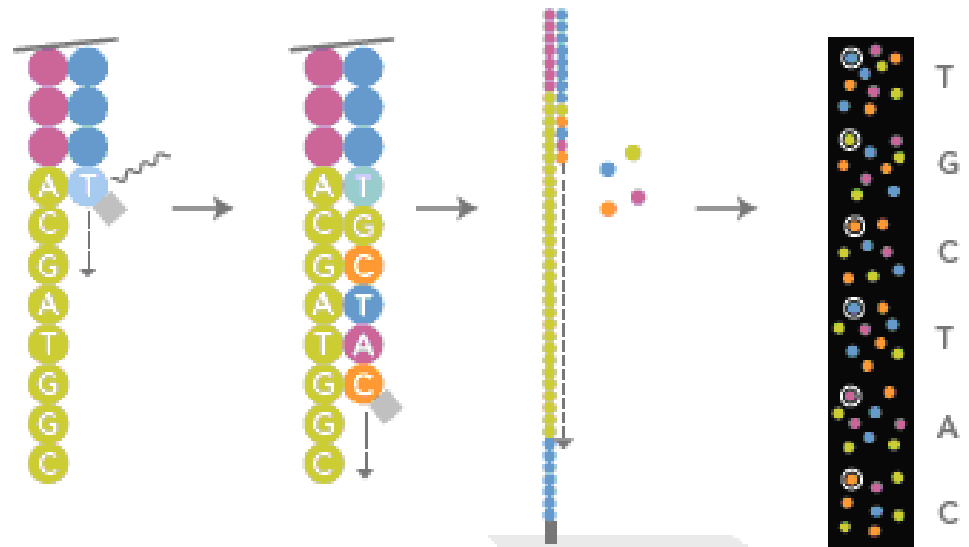


On the surface: complementary oligos



# Illumina

- **Průběh sekvenační reakce**
- Přidání primerů ve směsi s DNA polymerázou a fluorescenčně značenými dNTP
- Nukleotidy jsou na 3'-konci modifikovány tak, že umožňují **reverzibilní ukončení** prodlužujícího se řetězce DNA
- Je tak zajištěno, že se řetězec v každém cyklu prodlouží právě o jednu bázi.
- Po zachycení obrazu připojené báze následuje odblokování 3'-konce



# Illumina

- Délka čtených úseků:
  - HiSeq : 100 nt or 150 nt
  - MiSeq : 250 nt
- Mnohonásobné pokrytí sekvence (30 x – 100 x)
- Možnost mnohonásobného sekvenování
  - Povrch sklíčka rozdělen do částí
  - Vzorke označeny pomocí **dvanácti** různých oligonukleotidů
  - Současně lze analyzovat až 96 různých vzorků.
- Produkuje desítky až stovky Gb za 1 běh (2 - 8 dnů)
- Aplikace
  - Resekvenování
  - Sekvenování RNA
  - Stanovení metylací

## MiSeq





# Sekvenování třetí generace

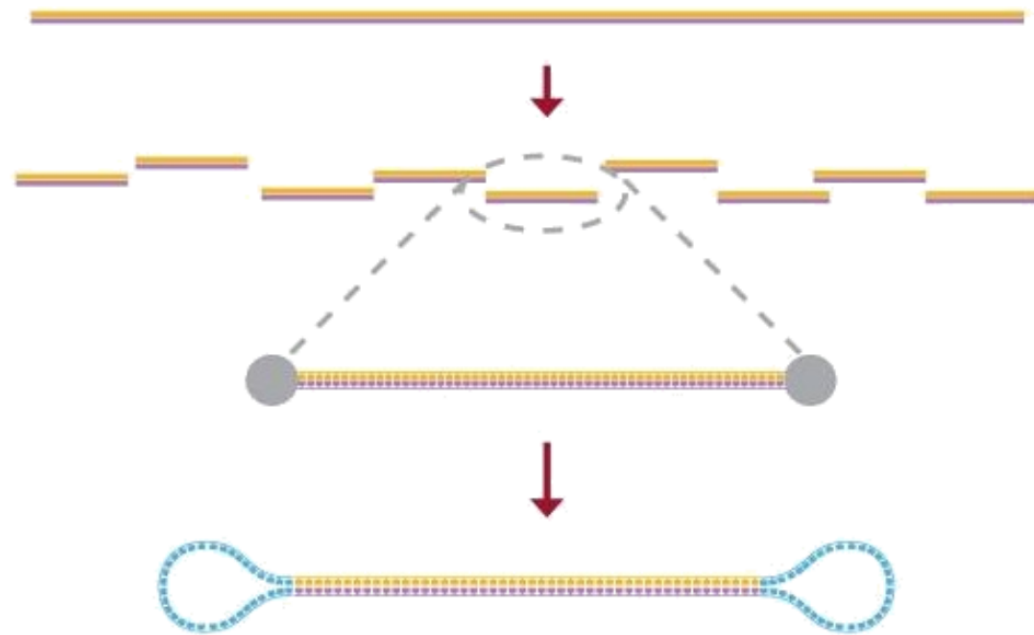
- Analýza jedné molekuly
- Potřeba malého množství vzorku
- Bez rizika kontaminace
- Přesné čtení homopolymerních úseků
- Analýza jakékoli NK (DNA/RNA)
- Analýza poškozených NK (archaické, muzejní, forenzní)
- Analýza DNA z nekultivovatelných organismů
- Absolutní kvantifikace

# PacBio

- **Kružnicové kontinuální sekvenování**
- SMRT templát obsahuje **dvouřetězcovou oblast** (inzert) na obou koncích **uzavřenou jednořetězcovými vlásenkovými smyčkami**
- Vlášenkové smyčky představují jednořetězcovou oblast, na kterou se může vázat sekvenační primer.
- Polymeráza prodlužuje primer z jedné vlásenkové struktury využívá jedno vlákno DNA jako templát a druhé vytěsňuje
- Když se polymeráza dostane zpět k 5'konci primeru, začne vytlačovat již nasyntetizovaný řetězec a pokračuje v syntéze DNA
- Výsledná sekvence je odvozená z obou vláken

# Příprava knihovny pro PacBio

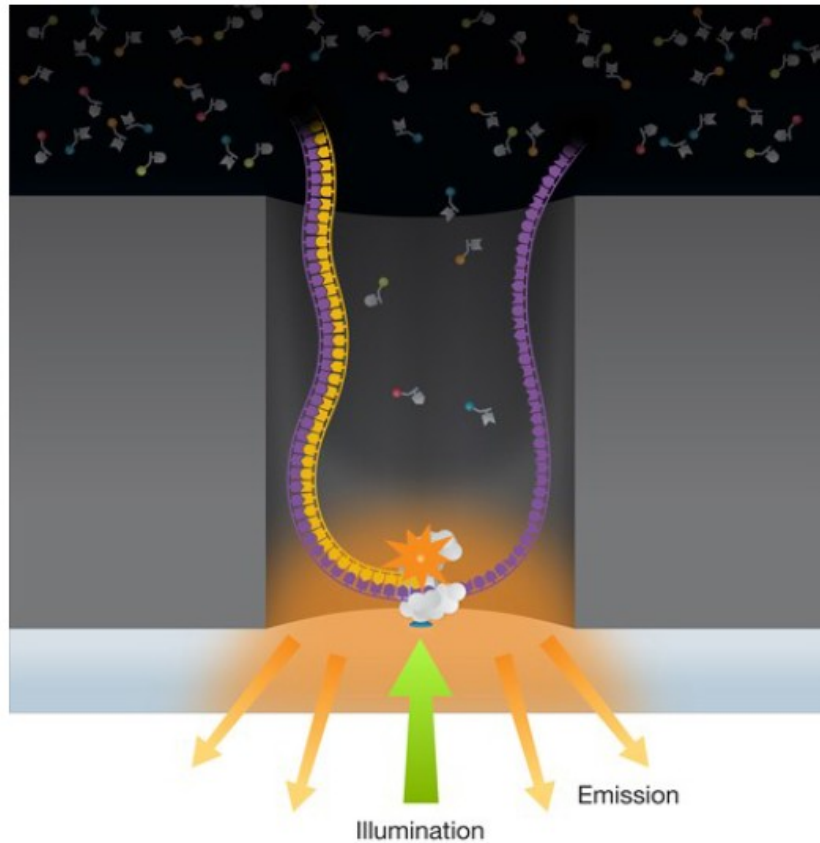
## Template Preparation



**SMRTbell™** Template preparation can be used to create libraries of various insert sizes from 250 bp to 20,000 bp depending on the needs of the application.

# Sekvenování třetí generace – PacBio RS

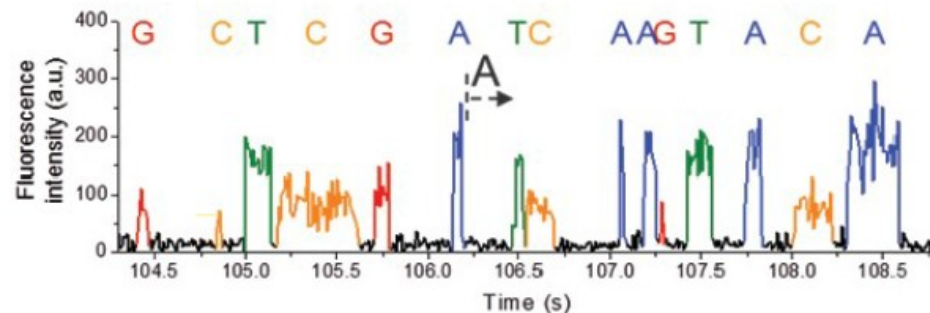
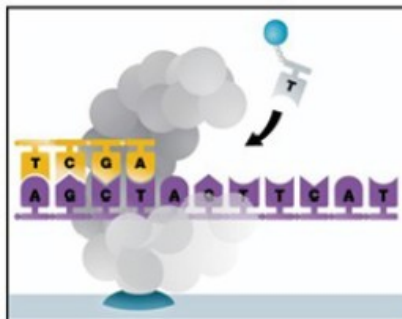
Pacific Biosciences



4 nucleotides with different fluorescent dye simultaneous present

2-3 nucleotides/sec  
2-3 Kb (up to 50) read length  
6 TB data in 30 minutes

laser damages polymerase



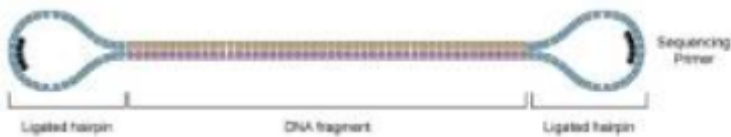
# PacBio

## High-throughput sequencing

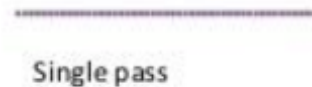


### Library preparation

SMRTBell 'template'

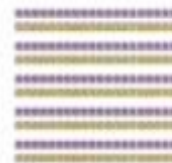


Standard 'Sequencing'



&

Circular 'Consensus' Sequencing'



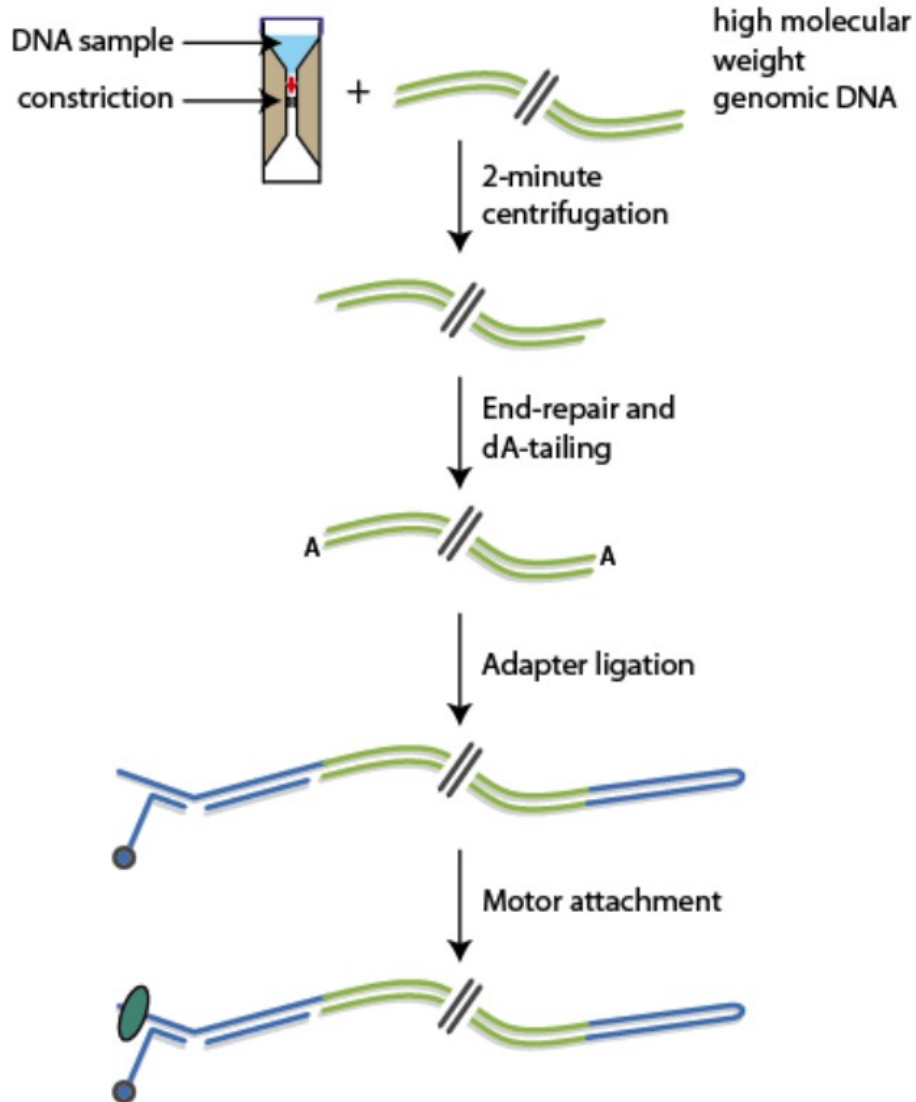
Continued generations of reads

&

# NanoPore (1995)

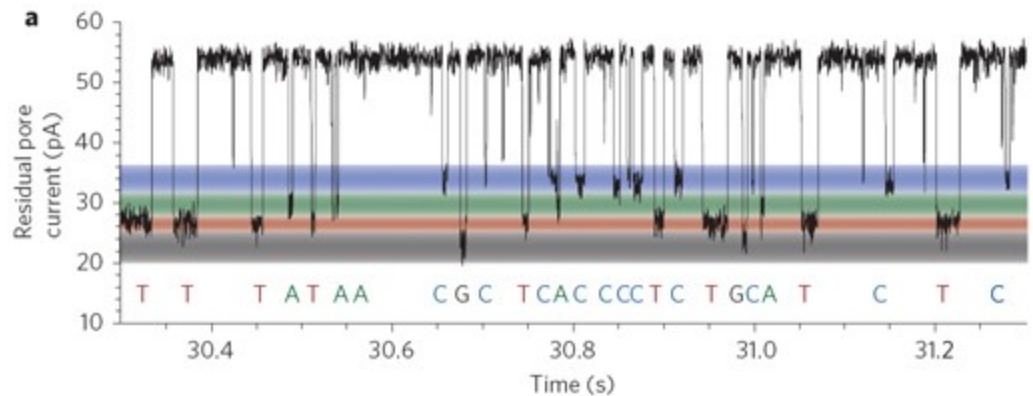
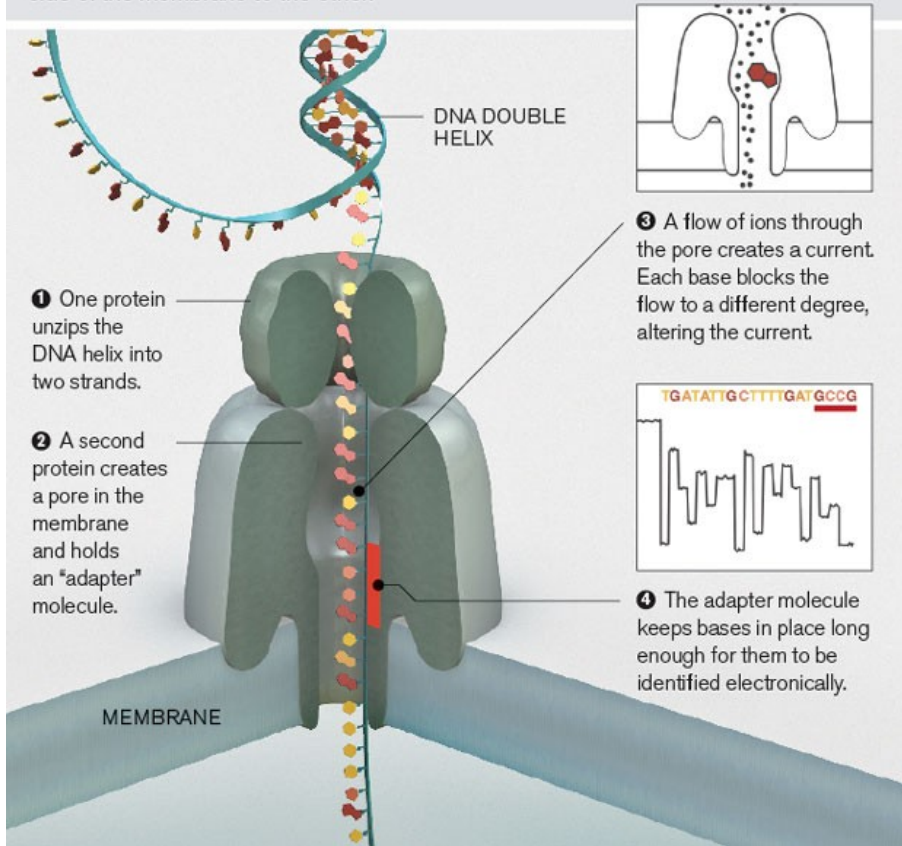
- **Využívá měření elektrického potenciálu přes membránu**
- Elektricky odolná polymerní membrána (synтетická lipidová dvouvrstva) s transmembránovým porózním proteinem
- Modifikovaný  $\alpha$  hemolysin ( $\alpha$ HL) nebo porin A (MspA) *Mycobacterium smegmatis*
- Nanopore je ponořený ve vodivém roztoku
- Průchod bází na sekvenovaném řetězci přerušuje proud, který je **specifický podle procházející báze**

## G-tube sense/antisense



# Příprava knihovny MinION

DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.





# the MinION device



the MinION device