

# GENOVÉ TECHNOLOGIE

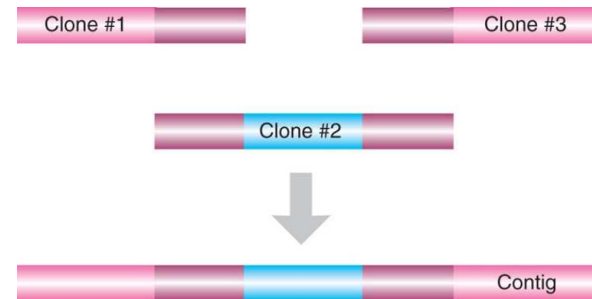
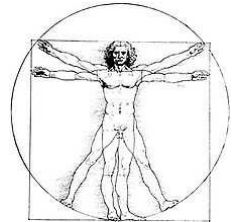
## **Genomika a genová exprese**

techniky mapování genů, nekódující části genomu, bioinformatické nástroje, farmakogenetika, DNA mikroarrays, RNA-seq techniky, metagenomika, epigenetika

# Sekvenace lidského genomu

- v době počátku (rok 1990) monumentální úkol
- započat v roce 1990 za účasti DOE and NIH
- sekvenace prováděna pomocí kontigových map a BACs
- prvotní plán počítal s dobou trvání 15 let
- nakonec sekvenace pomocí Sangerovy metody téměř dokončena již v roce 2000
- výsledná sekvenční mapa publikována 14. dubna 2003, 99.99% přesnost (National Human Genome Research Institute)
- celkové náklady projektu 3 miliardy dolarů
- v roce 2000 prezident Bill Clinton ujistil o nepatentovatelnosti lidské DNA

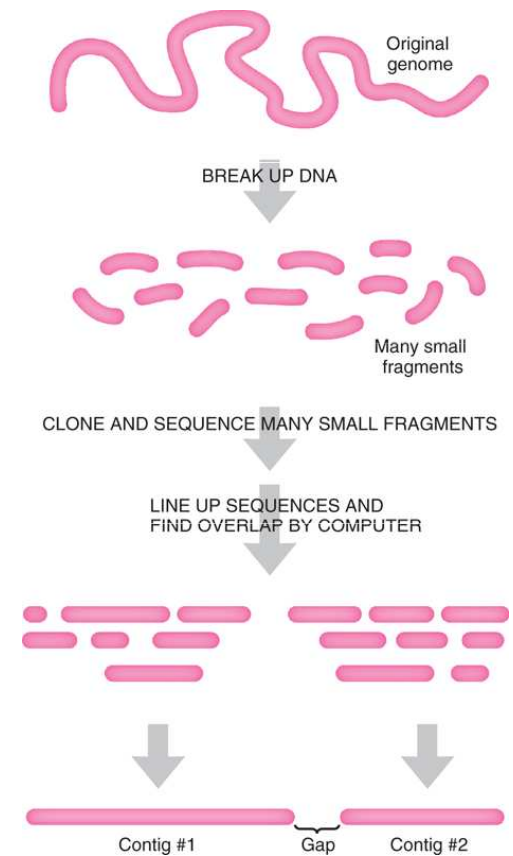
<https://www.genome.gov/25019885/online-education-kit-how-to-sequence-a-human-genome//>



# Celera Genomics Project



- založena vědcem Craig Venterem a v roce 1998 započala sekvenační projekt
- celkové náklady 300 mil. dolarů byly hrazeny plně s privátních zdrojů
- poprvé použita metoda „whole genome shotgun sequencing“
- k analýze sekvenačních dat použit přístup vyvinutý Gene Myersem
- tento přístup však vyžadoval extrémní výpočtové požadavky
- finální výpočet prováděn na 7000 procesorech k získání 1000-násobné rychlosti oproti Pentium počítačům
- tento inovativní přístup dovolil dokončit sekvenaci již za 9 měsíců



# Silná role diplomacie

**It is hard to imagine today's politicians reminding scientists that cooperation has as much value as competition.**

In 26 June 2000, US President Bill Clinton and UK Prime Minister Tony Blair presided over a carefully choreographed piece of scientific theatre. Through a video link connecting Washington DC and London, they announced to the world that scientists had completed a rough first draft of the human genome sequence.



Craig Venter (vlevo), Francis Collins (uprostřed), Bill Clinton (vpravo)

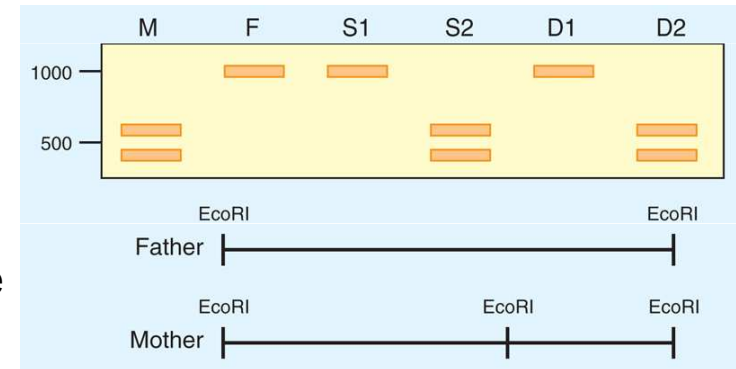
# Mapovací techniky

- Genomové mapy poskytují lineární sérii značek pro skládání sekvenčních dat
- Tvorba genomové mapy:
  - genetické mapy (křížení, analýza rodokmenu, přenos genů) – vazebné mapy
  - fyzikální mapy (radiační hybridní panel, FISH)
- Genetické mapy založeny na vazbě = pravděpodobnosti, že se dva mapované znaky od sebe v rámci křížení oddělí
- Pro určení relativní vzdálenosti markerů zásadní procento případů, kdy jsou nalezeni spolu
- Dnes je používána celá řada markerů

Typ mapování	Markery	Metody lokalizace
Genetické	Gen, biochemická vlastnost, DNA markery (RFLP, VNTRs, mikrosatelity, SNPs)	Vazebná analýza využívající křížení nebo páření, Analýza příbuznosti
Fyzikální	STSs, EST, VNTRs, mikrosatelity	Restrikční analýza, Radiační hybridní panel FISH, Cytogenetické mapování

# Genetické markery

- RFLP analýza příbuzných jedinců, snadná identifikace
- Variable Number Tandem Repeat (VNTR, minisatelite) – tandemové repetice o délce 9-80bp (forenzní testování, paternitní testy)
- Mikrosatelitní polymorfismus – tandemová repetice o délce 2 – 5bp
- Single Nucleotide Polymorphism (SNP)
- SNPs, VNTRs RFLPs jsou rovněž používány ve fyzikálním mapování
- Pro velké genomy potřebujeme další markery
  - STSs (Sequence Tagged Sites) – unikátní sekvence 100-500 bp
  - ESTs (Expressed Sequence tags) – identifikace cDNA knihovnách
- Digesce gDNA pomocí restričních enzymů – metoda fyzikálního mapování

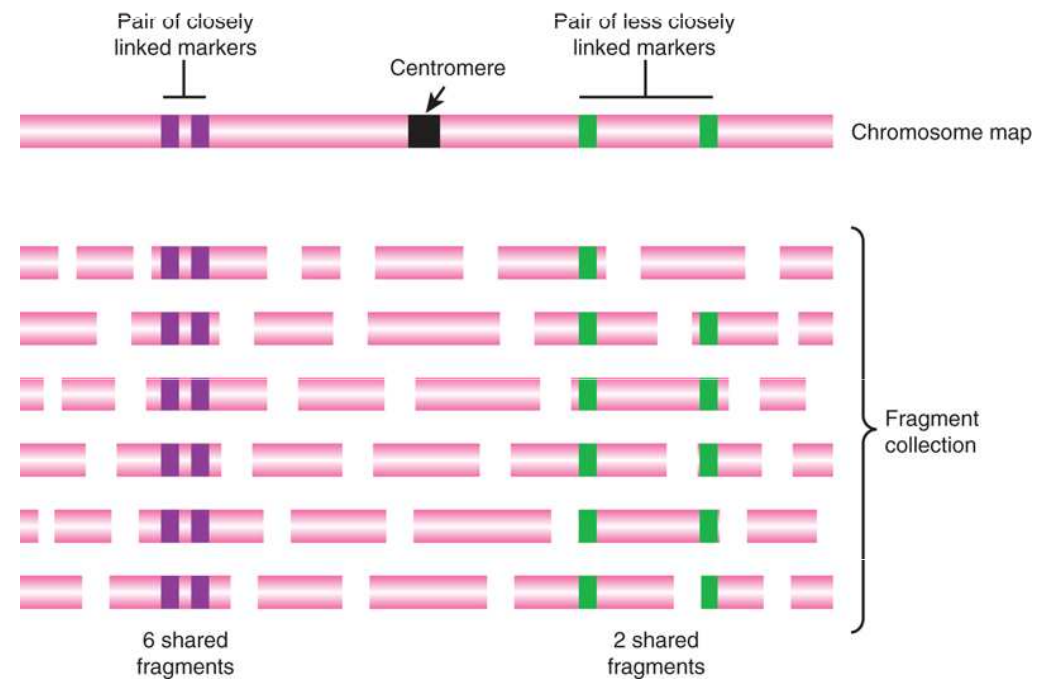
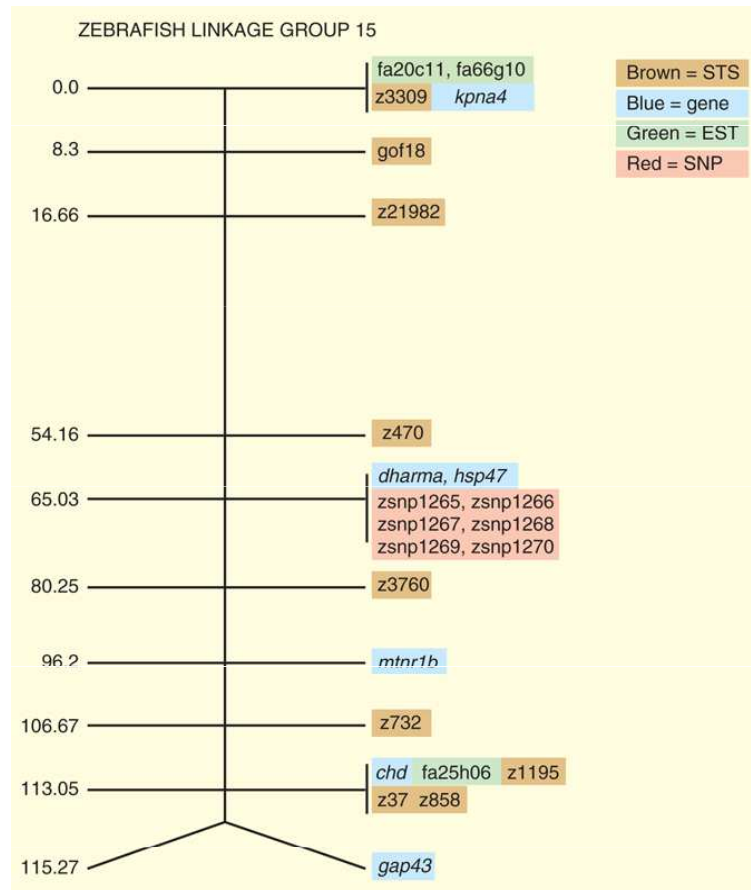


5' GTACTAGACTTA GTACTAGACTTA  
GTACTAGACTTA GTACTAGACTTA 3'

5' AAG G TAT 3' to 5' AAG C TAT 3'

Clark and Pazdernik, 2016

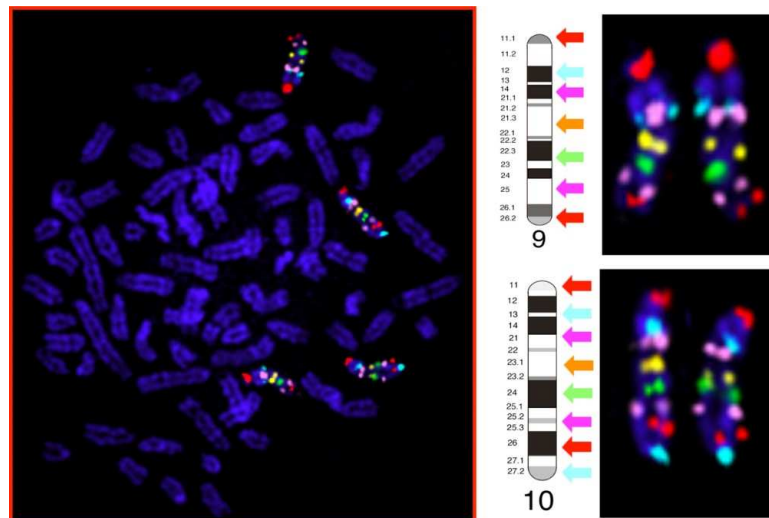
# Genetické markery



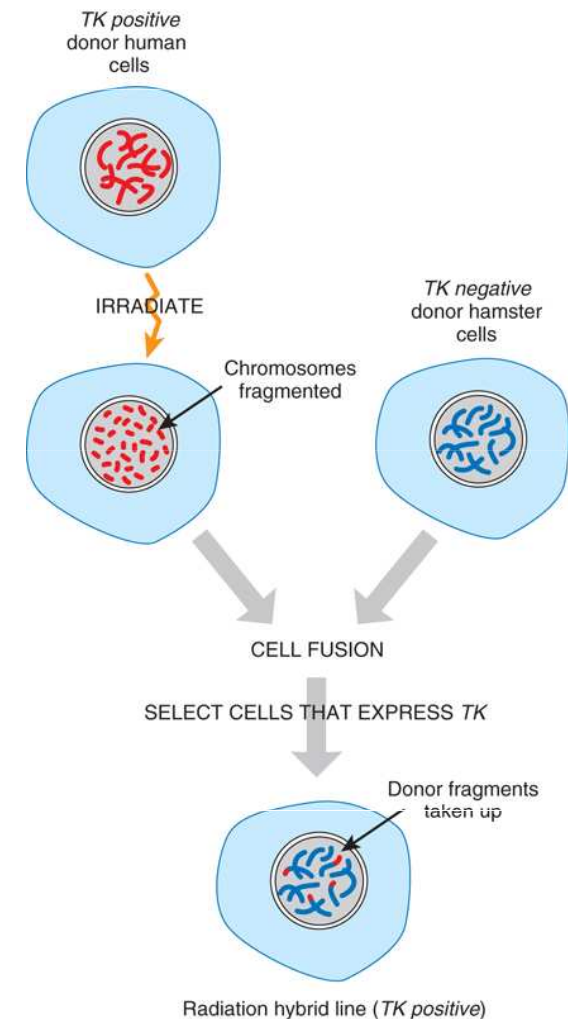
Clark and Pazdernik, 2016

# Fyzikální mapovací techniky

- FISH (Fluorescence in-situ hybridization) – umístění konkrétní DNA próby u chromozómů v metafázi vzhledem k proužkování (chromosome painting)



- radiační hybridní mapování – velké segmenty klonovaného genomu mohou obsahovat dva fragmenty z různých částí genom

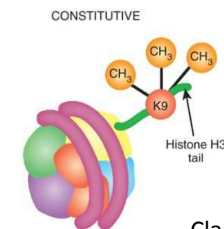
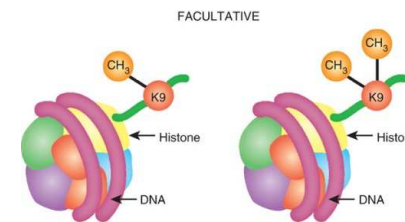
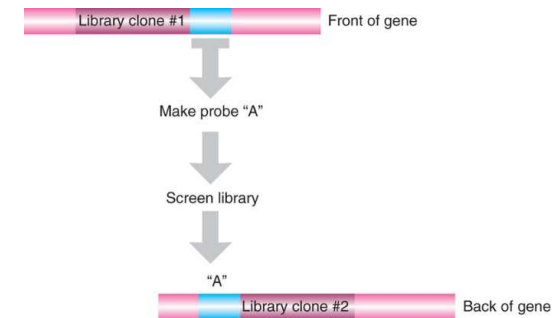


Clark and Pazdernik, 2016



# Mezery v rámci lidského genomu

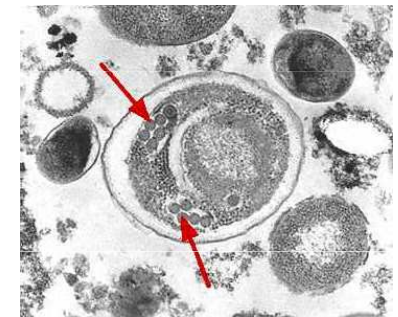
- Pro doplnění mezer použití techniky „chromosome walking“
- První klon je většinou lokalizován známého markeru (STS nebo RFLP)
- Nejvíc mezer je v oblastech repetitivní DNA – heterochromatin
  - hypoacetylace
  - metylace H3 na specifických Lys
  - methylace CpG nebo CpNpG
- Fakultativní X Konstitutivní heterochromatin



Clark and Pazdernik, 2016

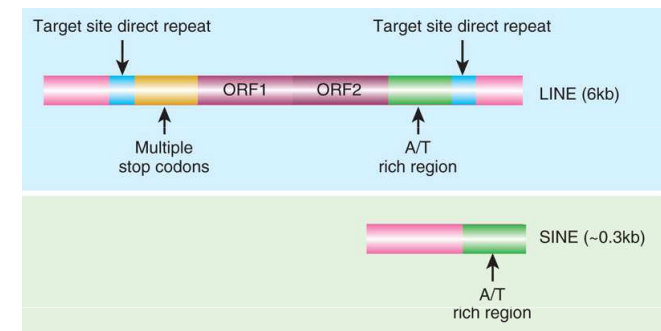
# Počet genů X Genom

Organismus	Velikost genomu (Mbp)	Počet protein-kódujících genů
Pšenice	17 000	95 000
Rýže	520	45 000
<i>Paris Japonica</i> (Pieris japonský)	149 000	26 000
<i>Trichomonas vaginalis</i>	160	46 000
<i>Encephalozoon intestinalis</i>	2.25	1833
Bahník východoafrický	130 000	?
Člověk	3200	21 850
Háďátko	97	20 493
Octomilka	180	13 600
<i>Streptomyces coelicolor</i>	8.7	7800
<i>E. coli</i>	4.6	4300
<i>Mycoplasma genitalium</i>	0.58	470



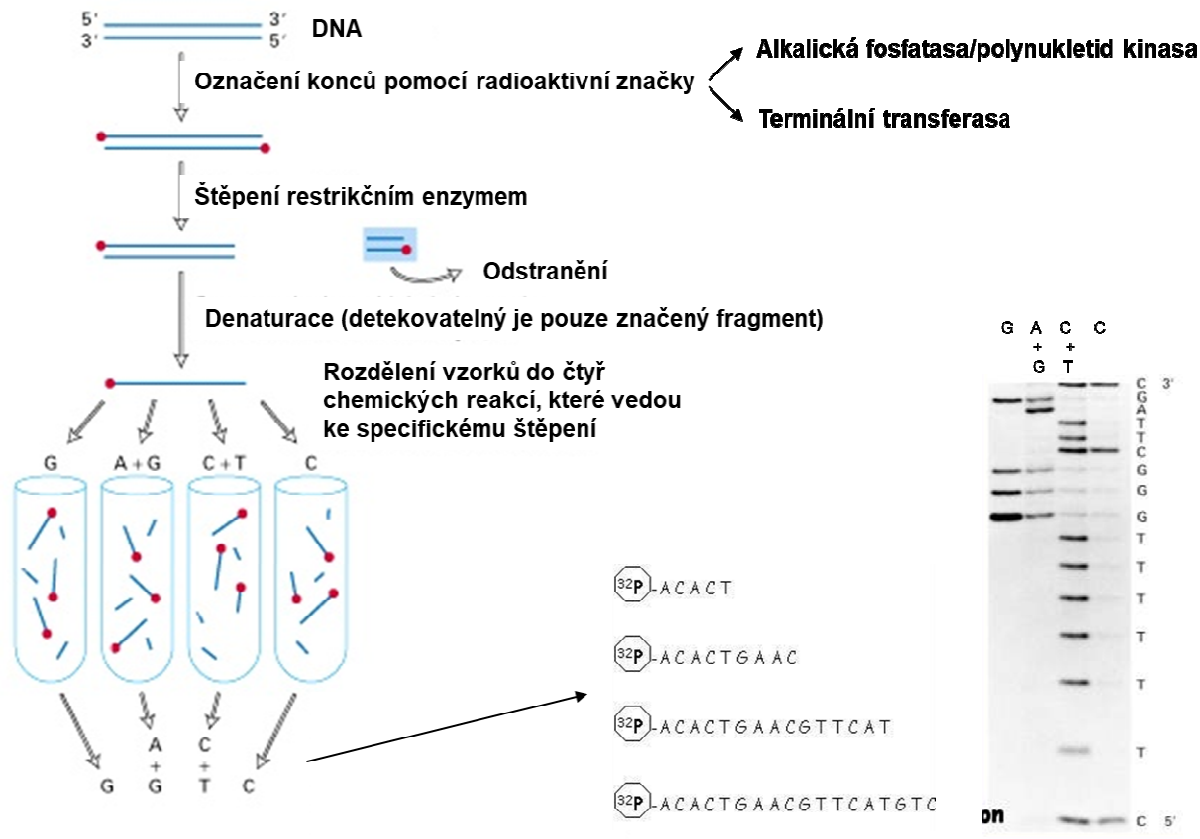
# Nekódující genom

- Protein-kódující geny tvoří pouze malou část genomu člověka
- V rámci různých druhů celá řada CNE (conserved noncoding elements)
- Možné zesilovače transkripce nebo izolační sekvence
- 25% lidského genomu tvoří geny pro proteiny, z toho ale jen 1% CDS
- Zbývá objasnit roli intronů
- Další velká část = repetitivní sekvence
  - ribozomální RNA geny (kódující)
  - LINE (long interspersed element) (nekódující) – až 20% genomu, retrovirus-like elementy
  - SINE (short interspersed element) – až 13% genomu, Alu elementy



Clark and Pazdernik, 2016

# Maxam-Gilbertova sekvenační metoda



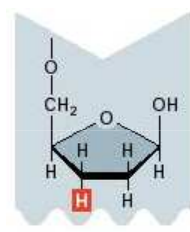
- Touto metodou jsme schopni sekvenovat přibližně 250-300 bp dlouhé fragmenty
- Musíme pracovat s velkým množstvím DNA
- Velká pracnost metody (několik purifikačních kroků), nemožnost plné automatizace, práce s mutagenními chemikáliemi
- Používá se stále pro "footprinting"

# Sekvenační metoda dle Sangera

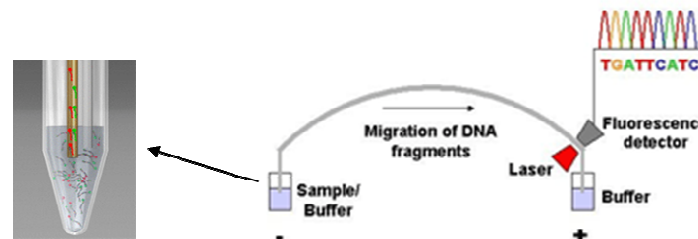
- Syntéza DNA in-vitro za použití “terminátorů” – dideoxynukleotidů zabraňujících po svém začlenění do DNA její další elongaci.



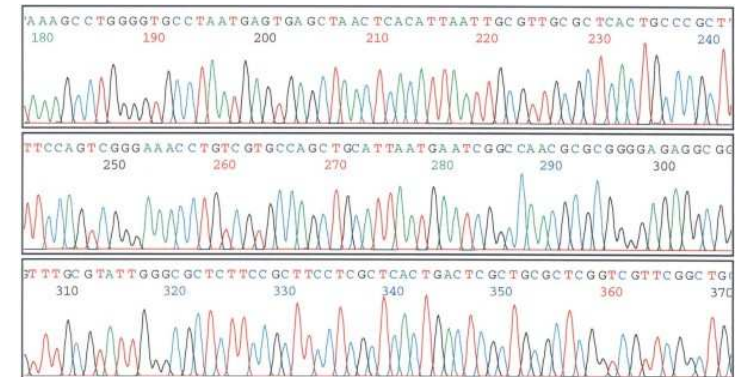
Deoxyribose



Dideoxyribose



- Vyžaduje použití iniciálního primeru, DNA polymerázy a směsi dNTPs se značenými ddNTPs
- Nasyntetizované řetězce jsou poté separovány pomocí polyakrylamidové gelové elektroforézy nebo kapilární elektroforézy
- Možnost plně automatizované separace za použití fluorescenčně značených ddNTPs



# Sekvenační metoda dle Sanger



## Throughput/Performance by Run Module

XLRseq: 768 samples per day (690 Kbases)

LongSeq: 1152 samples/day (980 Kbases)

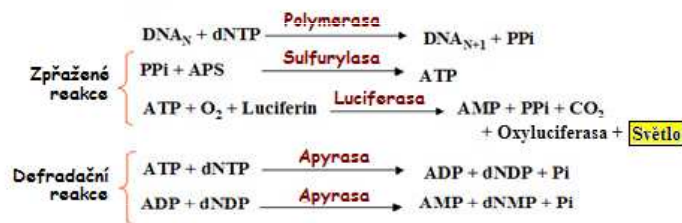
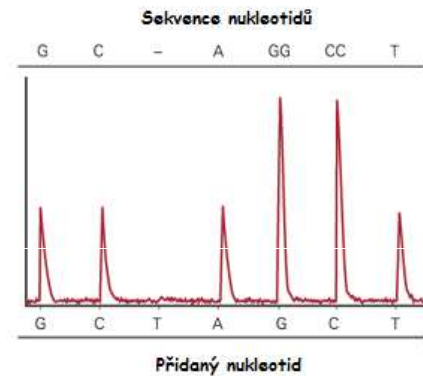
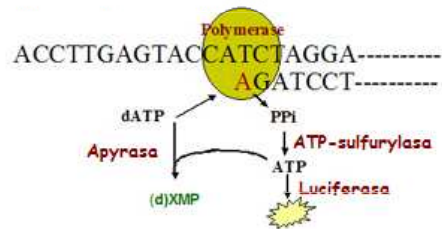
StdSeq: 2304 samples/day (1550 Kbases)

FastSeq: 2304 samples/day (1600 Kbases)

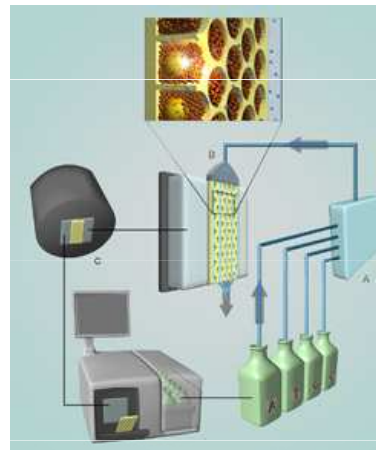
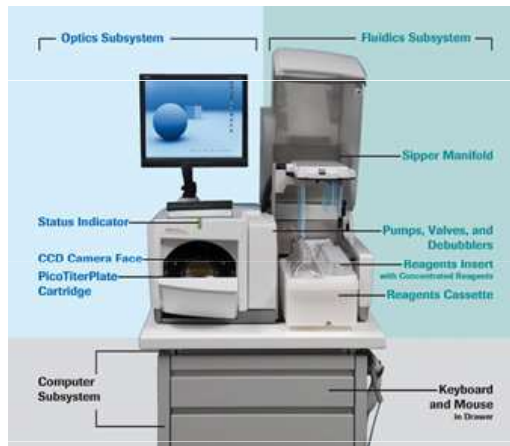
**RapidSeq: 3840 samples per day (2100 Kbases)**

# Pyrosekvenování (1990)

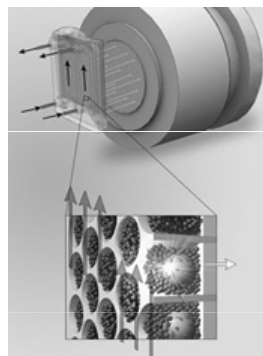
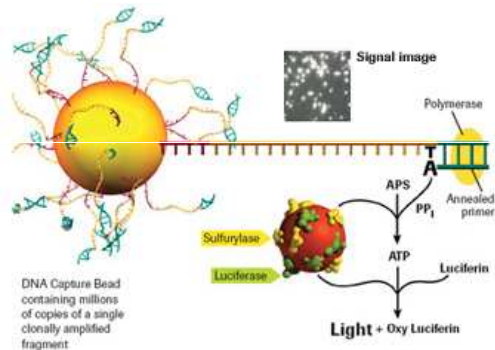
- umožňuje rychlou sekvenaci krátkých úseků DNA - sekvenace 30 až 50 bazí trvá přibližně 30 až 45 minut.
- Jedná se o bio-luminometrické sekvenování DNA založené na detekci anorganického pyrofosfátu (PPi) uvolněného během inkorporace nukleotidů.



# 454 a GS Junior systém



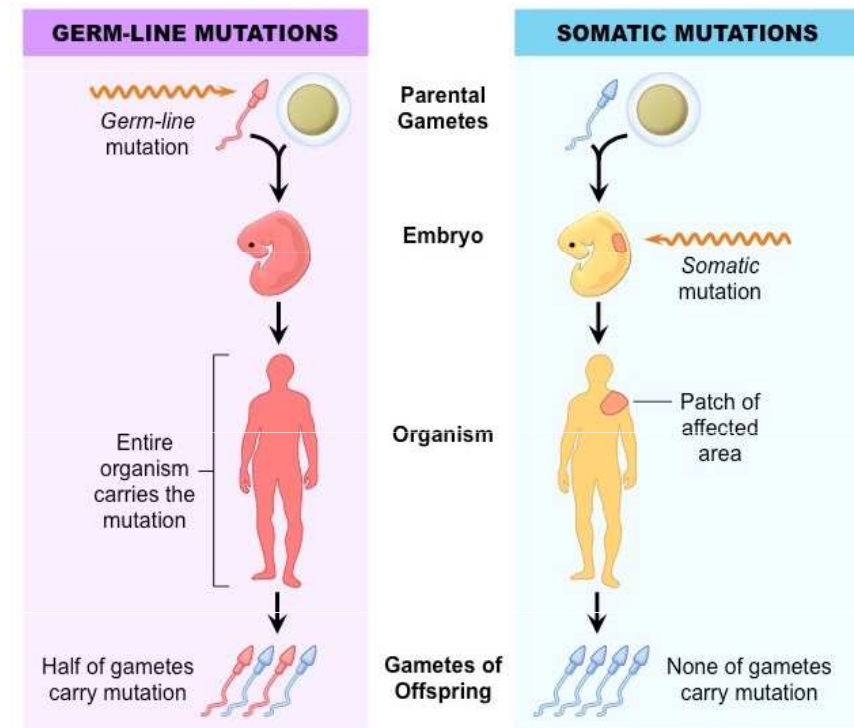
Průchodnost	1 miliarda bazí za den
Doba analýza	10.0 hodin
Délka čtení	400
Počet čtení/analýzu	1 000.000
Správnost	>99.0% správnost jednoho čtení na 400 bazích
Potřebné množství DNA	Méně než 100 ng DNA
Multiplexování	Až 192 vzorků/běh





# Akumulace mutací v DNA

- Projekt HUGO otevřel mnoho nových možností
- V rámci miliónů let dochází k hromadění mutací v rámci genomu
- Většina defektních buněk umírá procesem apoptózy (řízené buněčné smrti)
- Mutace
  - v somatických buňkách – nepřenáší se na potomstvo
  - v zárodečných buňkách – přenáší se na potomstvo

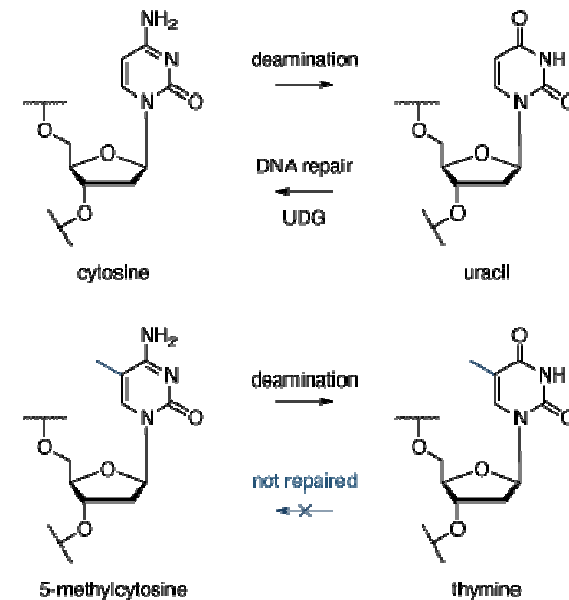


# Typy mutací

<b>Tranzice</b>	GA <u>A</u> CGT → GA <u>G</u> CGT
<b>Transverze</b>	GA <u>A</u> CGT → GA <u>T</u> CGT
<b>Missense mutace</b>	GA <u>A</u> CGT → GA <u>T</u> CGT Glu Arg Asp Arg
<b>Konzervativní substituce</b>	<u>A</u> CTCGT → <u>T</u> CTCGT Thr Arg Ser Arg
<b>Radikální nahrazení</b>	GA <u>A</u> TCGT → G <u>C</u> TCGT Asp Arg Ala Arg
<b>Nonsense mutace</b>	<u>G</u> AACGT → <u>T</u> AACGT Asp Arg STOP
<b>Inzerce</b>	GAACGT → GAA <u>A</u> CGT
<b>Delece</b>	GAACGT → GACGT

**Mutační hot-spoty** – oblasti s vyšším výskytem mutací (metylace cytosinu, prokluz v repetitivních oblastech)

Genetická variabilita zohledňuje **rekombinační hot-spoty** = haplotypové bloky (haplobloky)



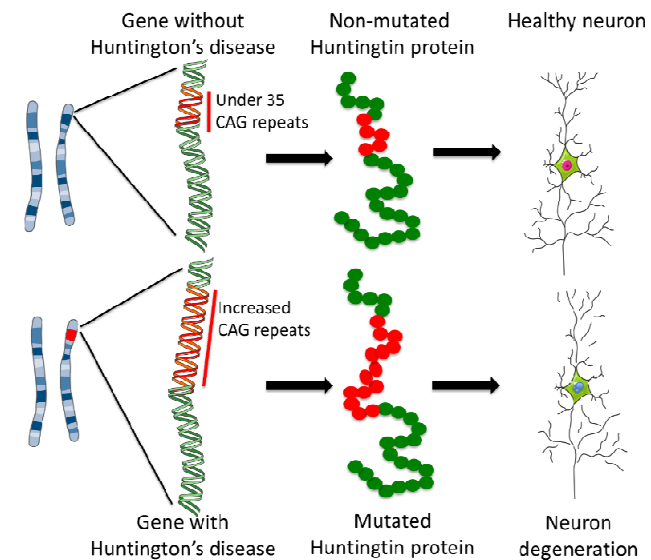
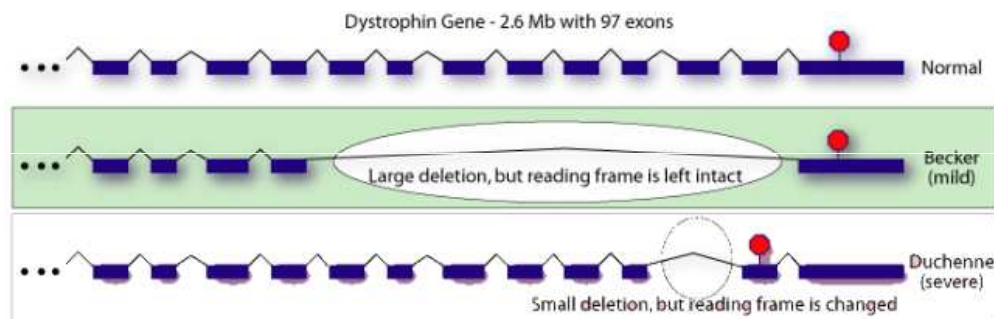
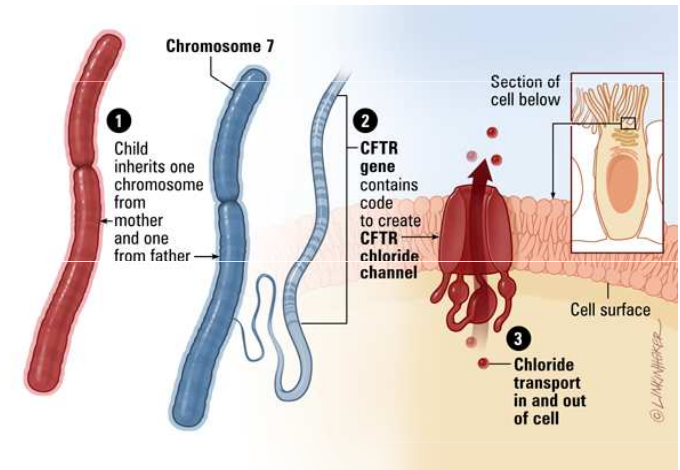
# Míra mutací v genomech

**Table 8.7** Mutation Rates in DNA Genomes

Organism	Genome Size (Kilobases)	Mutation Rate per Generation		
		Per kb	Per Genome (Uncorrected)	Per Effective Genome
Bacteriophage M13	6.4	$7.2 \times 10^{-4}$	0.005	0.005
Bacteriophage Lambda	49	$7.7 \times 10^{-15}$	0.004	0.004
<i>Escherichia coli</i>	4,600	$5.4 \times 10^{-7}$	0.003	0.003
<i>Saccharomyces cerevisiae</i>	12,000	$2.2 \times 10^{-7}$	0.003	0.003
<i>Caenorhabditis elegans</i>	80,000	$2.3 \times 10^{-7}$	0.018	0.004
<i>Drosophila</i>	170,000	$3.4 \times 10^{-7}$	0.058	0.005
Human	3,200,000	$5.0 \times 10^{-8}$	0.16	0.004

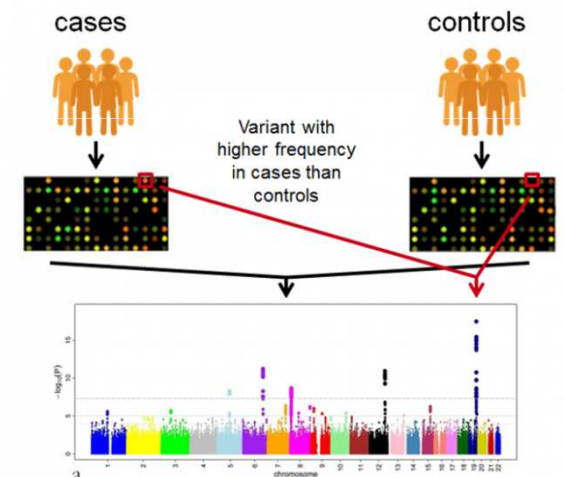
# Genomika v medicíně

- Největší aplikace genomických dat v diagnostice onemocnění
- Genetické testování – stanovení přítomnosti genu asociovaného s onemocněním:
  - svalová dystrofie (gen pro dystrophin)
  - cystická fibróza (gen CFTR)
  - Huntingtonova choroba (HTT gen)

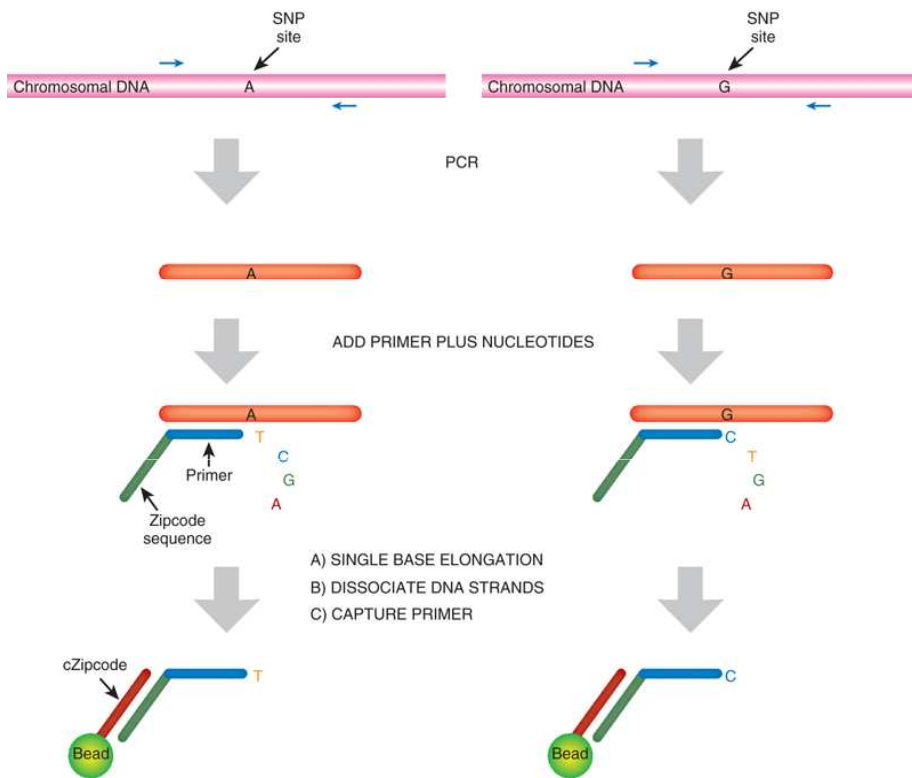


# Genomika v medicíně

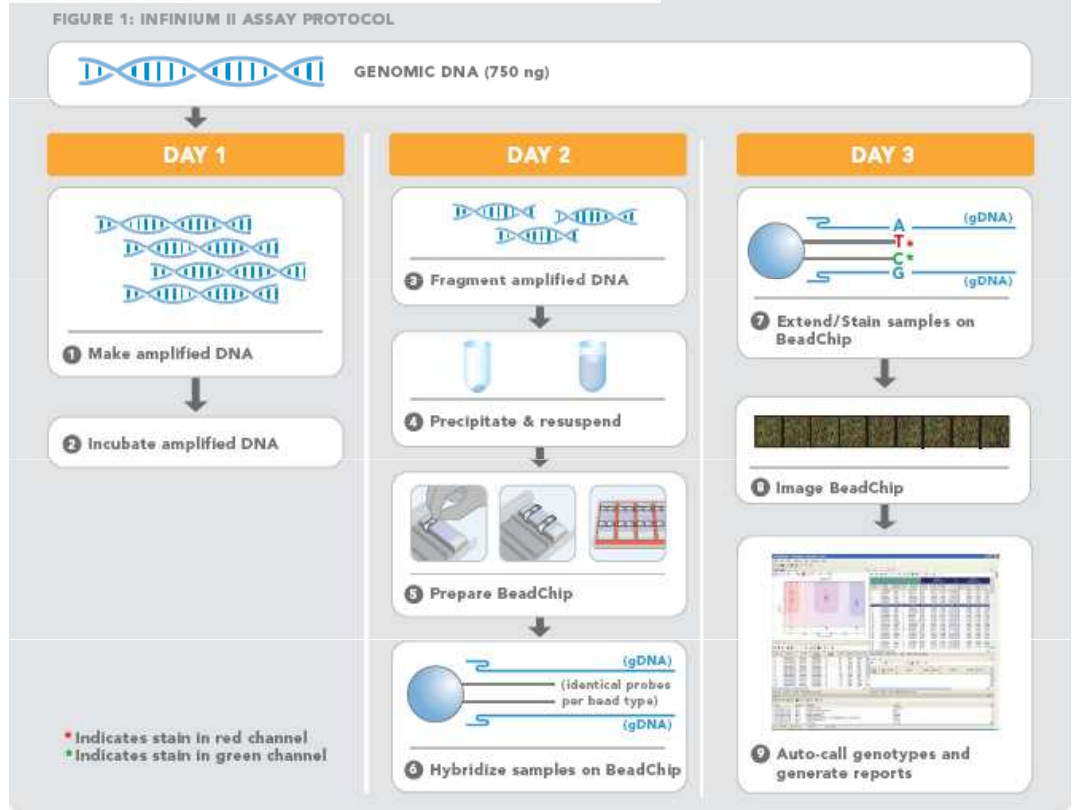
- Pro identifikaci kauzálních mutací výhodnější sekvenovat exom (2%) než genom
- V současné době více než 3000 onemocnění identifikováno pomocí genomiky a rodokmenové analýzy
  - tzv. Mendelovy choroby (mutace v jednom genu vede k onemocnění)
- Celá řada onemocnění je polygenních (příspěvek více genů pro rozvoj onemocnění)
  - Crohnova choroba
  - autoimunitní onemocnění
  - psychiatrické poruchy (schizofrenie, AD, mírná kognitivní porucha)
- V rámci těchto onemocnění využití GWAS (genome-wide association study)
  - analýza jednobodových polymorfismů (SNPs) – frekvence nižší než 1%
  - vliv genotypu a prostředí na rozvoj onemocnění



# Zipcode analysis



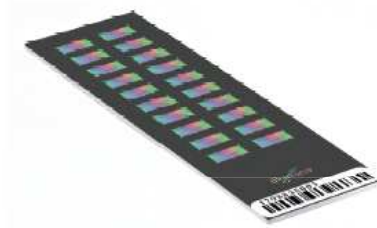
Clark and Pazdernik, 2016



# Metoda Infinium

- Celogenomové typizace SNPs
- Možnost provádění CNV analýz (copy number variation)
- Průměrná call rate > 99%
- 1 072 820 SNPs/vzorek
- 4 až 48 vzorků najednou

illumina®

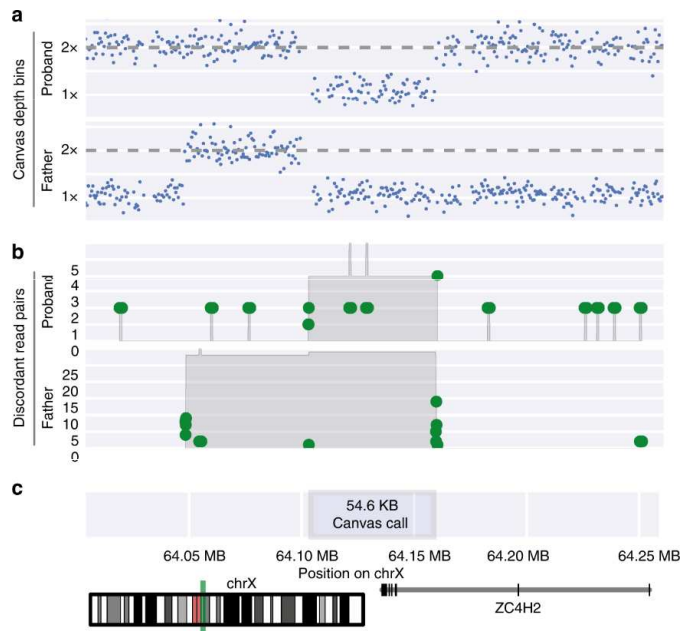


Infinium HD BeadChips	Samples per BeadChip	Markers per Sample
HumanOmni1-Quad	4	> 1 million*
Human1M-Duo	2	> 1 million
HumanOmniExpress	12	> 700,000
Human660W-Quad	4	> 658,000
HumanCytoSNP-12	12	~ 300,000
Semi-Custom Human1M-Duo+, and HumanHap550-Quad+	2 / 4	standard content and up to 60,800 customized SNPs per sample

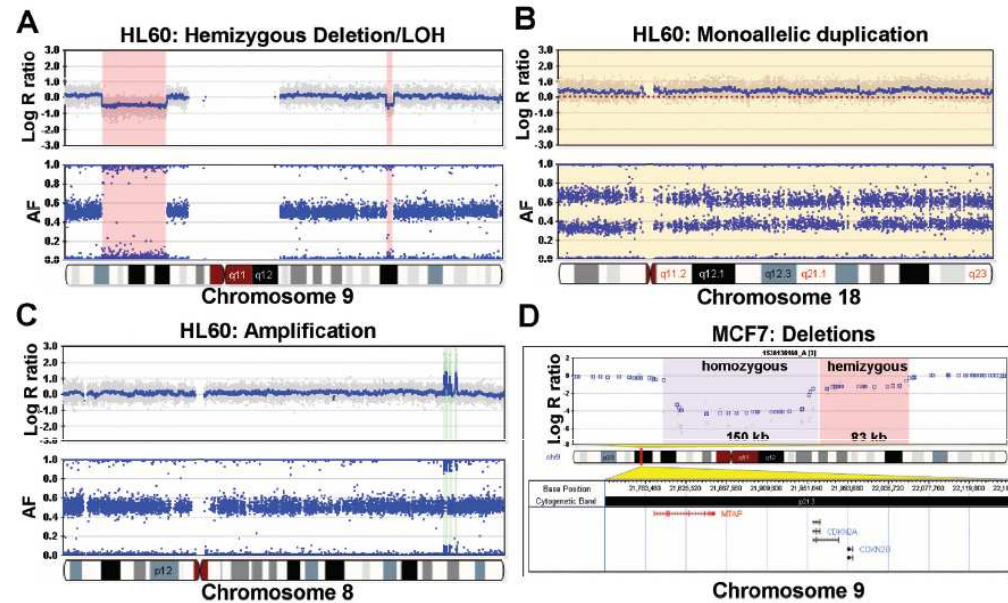


# CNVs (copy number variants) analýza

- Pomůcka pro detekci celé řady geneticky-podmíněných onemocnění
- Analýza chromozomálních aberací (delece, inzerce, multiplikace, přesupení)



Gross et al. 2018





# Farmakogenetika a genomika

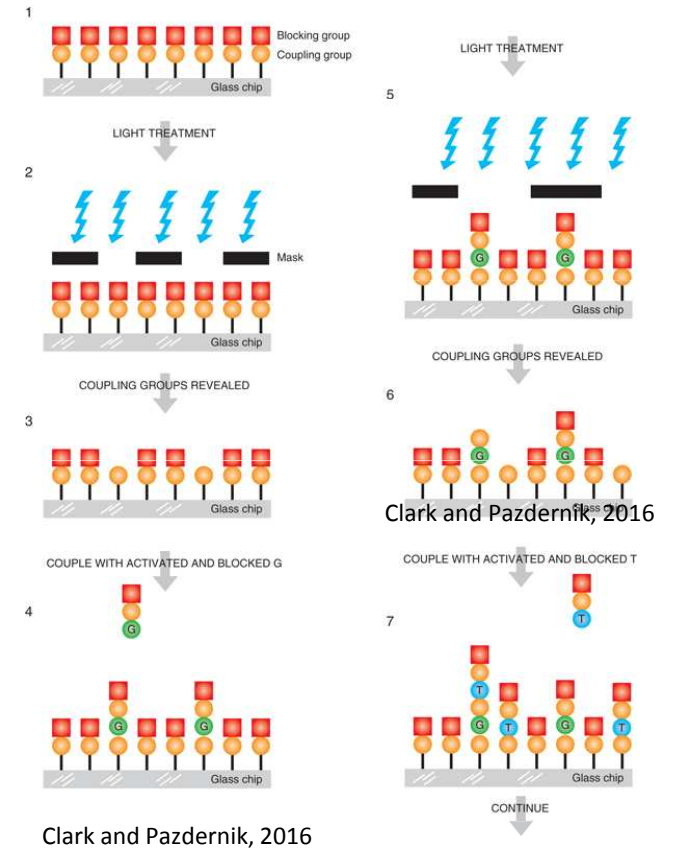
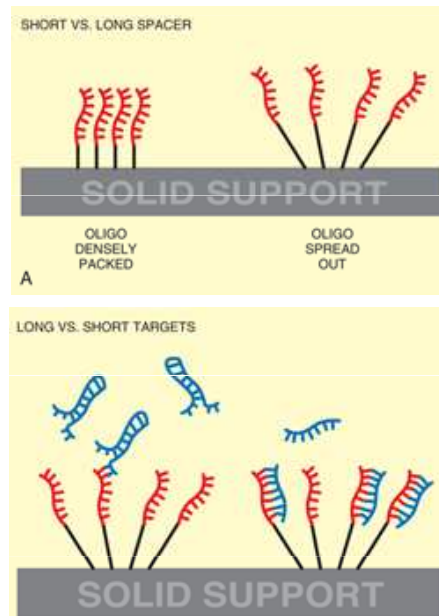
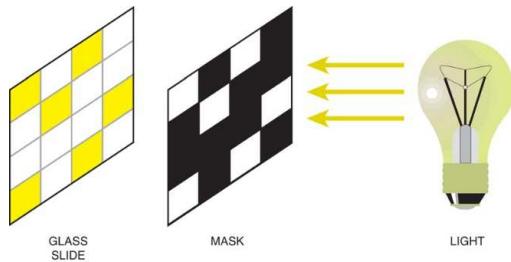
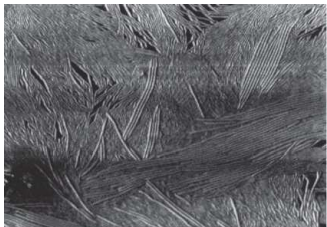


- Vývoj celé řady léčiv byl řízen omylem (penicilin, paralen, viagra, atd.)
- Velký problém při vývoji léčiv je ADR (adverse drug reaction) – přibližně u 7% lidí
- Farmakogenetika – studium dědičných rozdílů v metabolismu léků
- Farmakogenomika – studium všech genů ovlivňujících odpověď na léčivo
- Roche Diagnostics' AmpliChip P450 2D6/2C19 – první diagnostická analýza SNPs spojených s metabolismem léčiv pomocí DNA čipu

Determined Genotype	OR	OR	OR	OR
Predicted Phenotype	<b>Ultrarapid Metabolizers:</b> Too rapid drug Metabolism, no Drug response	<b>Extensive Metabolizers:</b> Expected drug response	<b>Intermediate Metabolizers:</b> Slow Metabolism, higher drug levels Than expected	<b>Poor Metabolizers:</b> Very slow or No metabolism, High levels of drug, High risk of ADR
An Example of Population ratio				
An Example of Drug Dosage				
An Example of Pro-Drug Dosage				

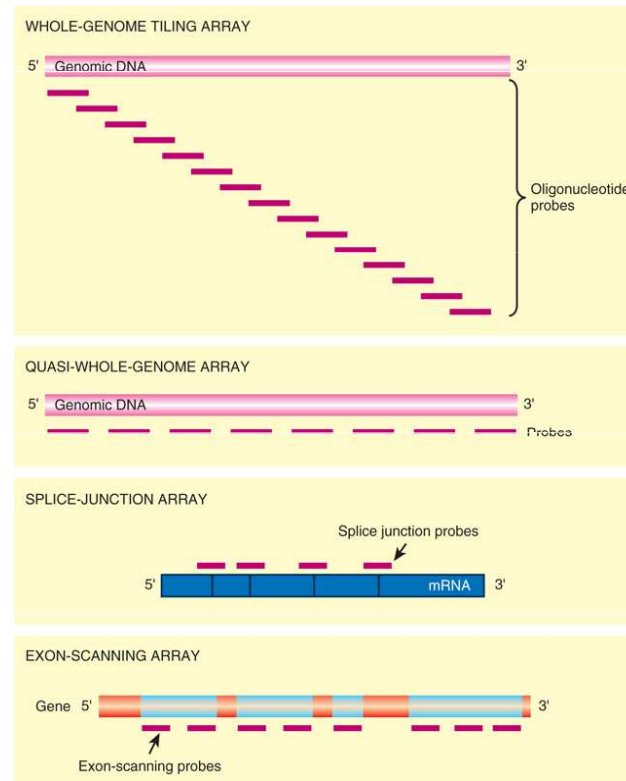
# Genová exprese - Microarrays

- cDNA microarrays:
  - nanesení části cDNA knihoven, PCR produktů na čip (25tis. genů u člověka)
  - přímá syntéza oligonukleotidů na čipu pomocí fotolitografie
- Výběr vhodných oblastí pro sondy (GC poměr, sek. struktury)

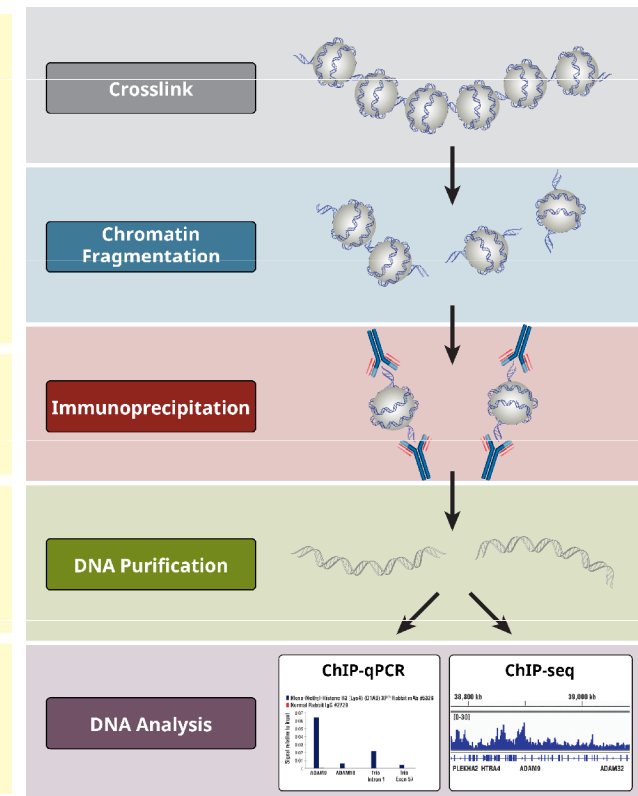


# Genová exprese – WGAs, ChIP

- WGAs (whole-genome tiling arrays pokrývají celý genom
- Poprvé u Arabidopsis (25-merní oligonucleotidy)
- Objevení nových genů, sestřihových variant
- ChIP (chromatin immunoprecipitation):
  - analýza DNA oblastí jednotlivých transkripčních faktorů
  - analýza DNA oblastí asociovaných s PTMs histonů

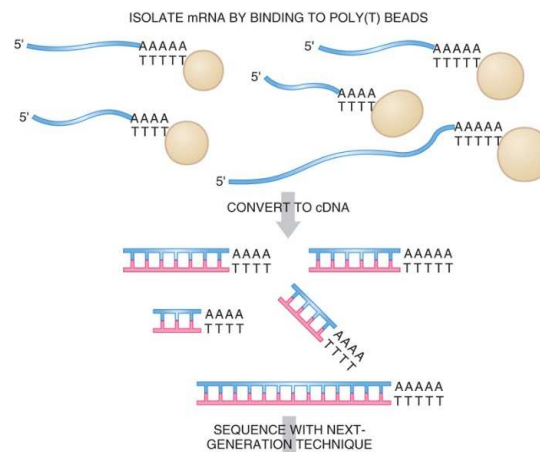
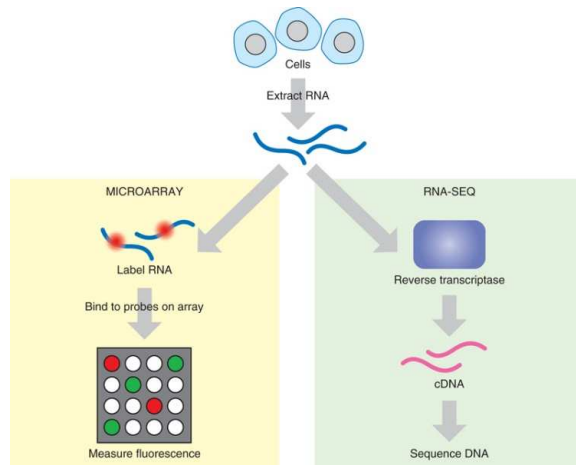


Clark and Pazdernik, 2016

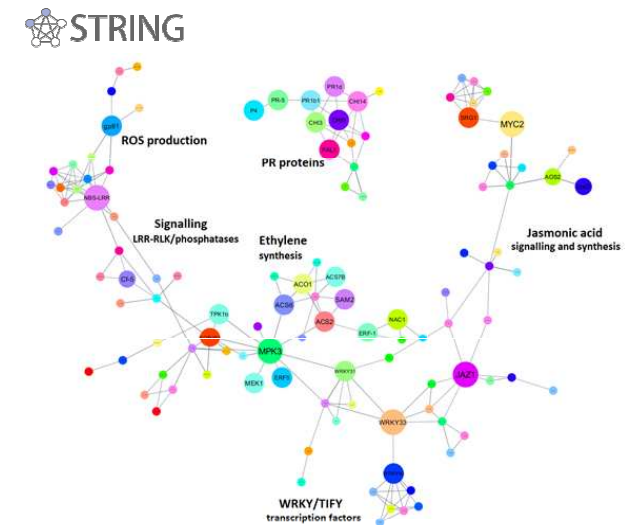
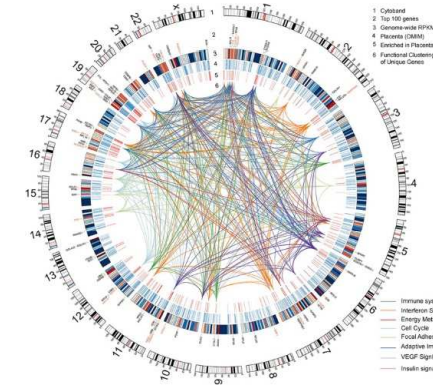


# Genová exprese - RNAseq

- Výhody RNAseq metody:
  - není závislý na sondách (správnější kvantifikace daných RNA molekul)
  - velký dynamický rozsah
  - detekce alternativních sestřihů a možnost jejich kvantifikace
  - možnost provedení i bez znalosti genomové sekvence
  - možnost provedení i z jedné buňky

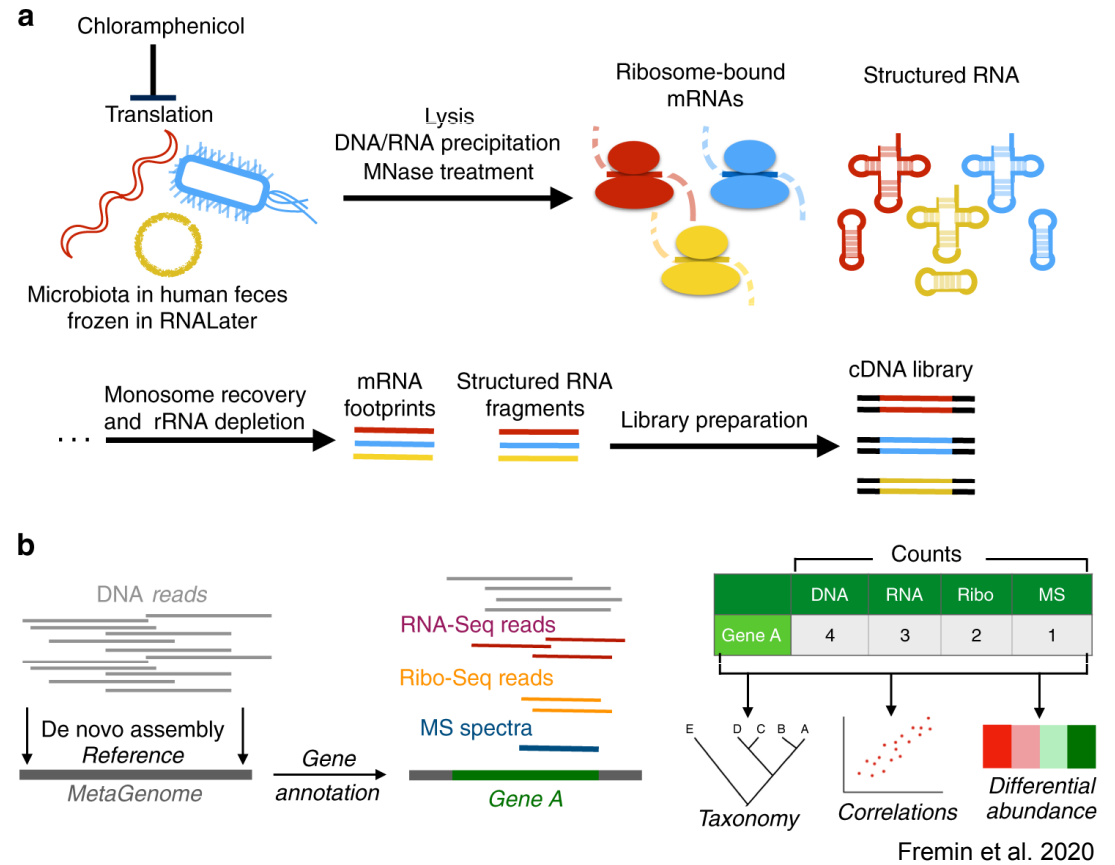
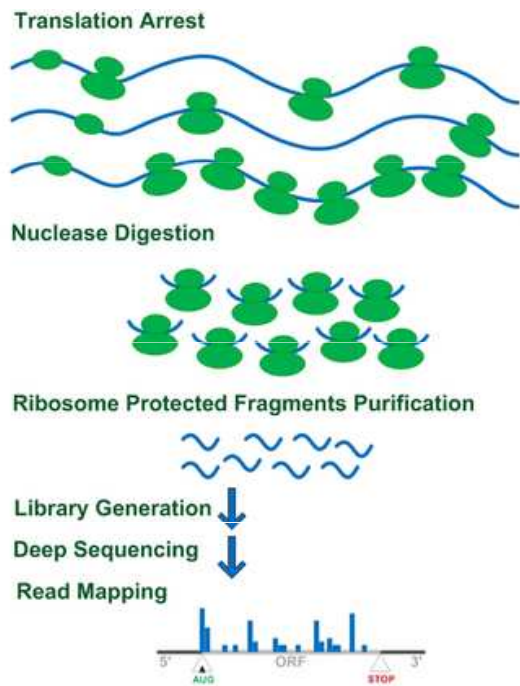


Clark and Pazdernik, 2016



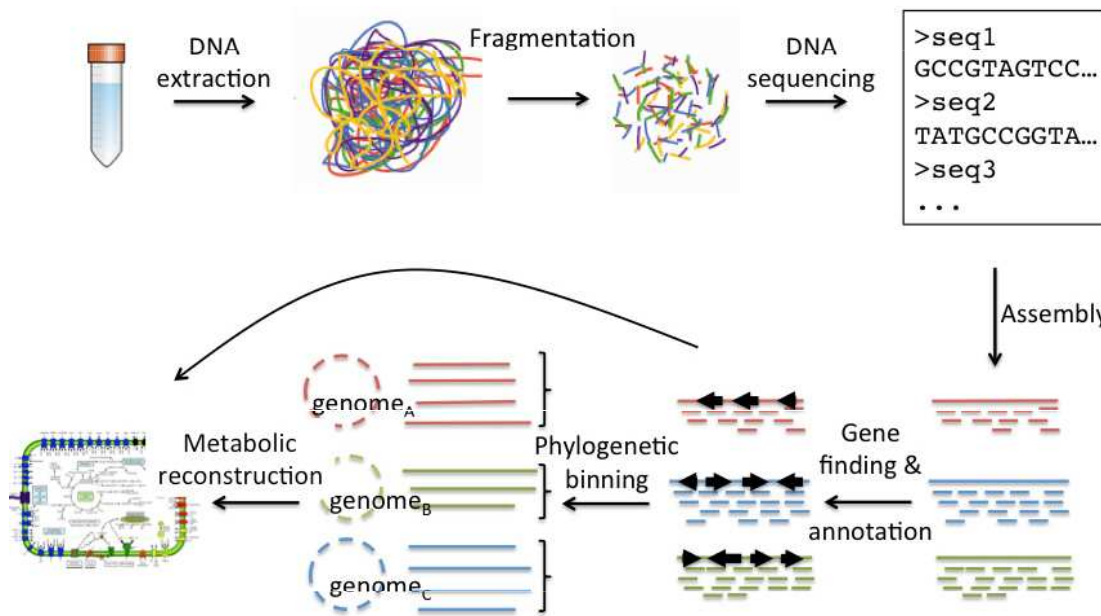
# MetaRibo-Seq

- Riboseq – zastavení translace a následná sekvenace translatomu

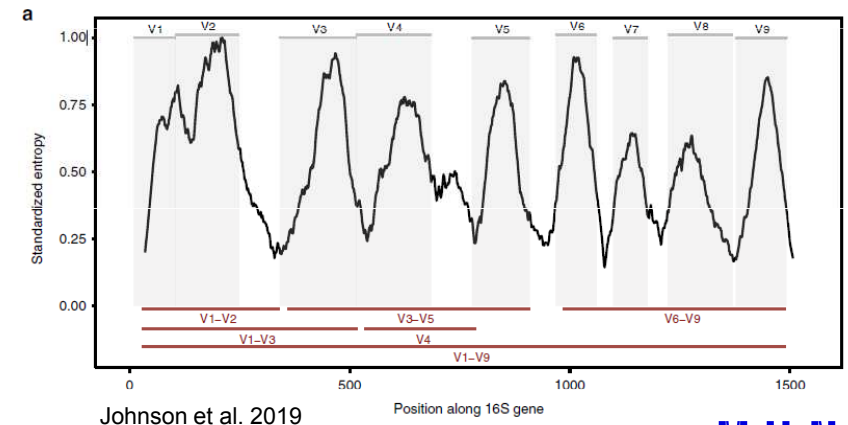
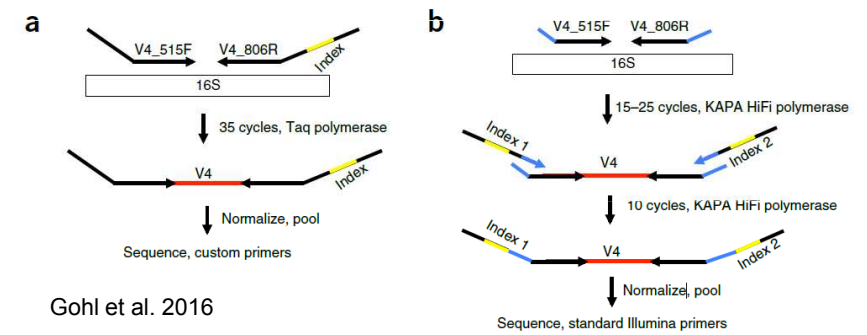


# Metagenomika

- Studium genetického materiálu obsaženého ve vzorku
- Shotgun přístup X sekvenace specifických fylogenetických oblastí (16S, 18S, ITS, mcrA)



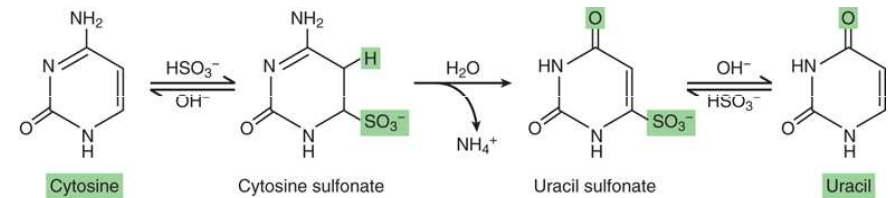
Microbiome	Microbiota	Metagenome
Microorganisms (and their genes) living in a specific environment	Microorganisms (by type) living in a specific environment	The genes of microorganisms in a specific environment



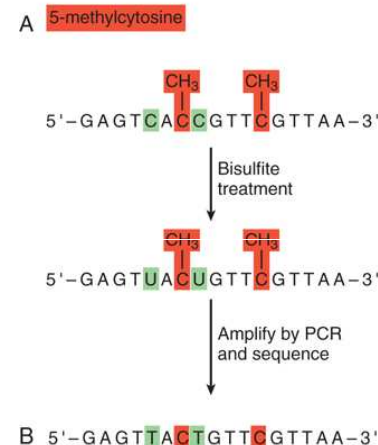
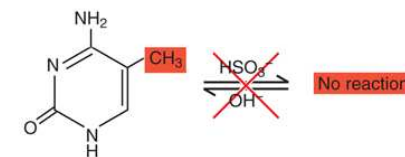


# Sledování methylomu

- Analýza míst metylace gDNA
- Metylací většinou umlčení transpozonové elementy
- Umlčení jedné kopie chromozomu X u žen
- Analýza pomocí bisulfitové metody
  - přidání siřičitanu sodného vede k konverzi nemetylovaných cytosinů na uracil
  - následná sekvenace bez a s přidavkem siřičitanu vede k odhalení metylačních míst
- sekvenátory 3. generace (Nanopores, PacBIO) jsou schopny přímo číst metylaci na cytosinu



Clark and Pazdernik, 2016

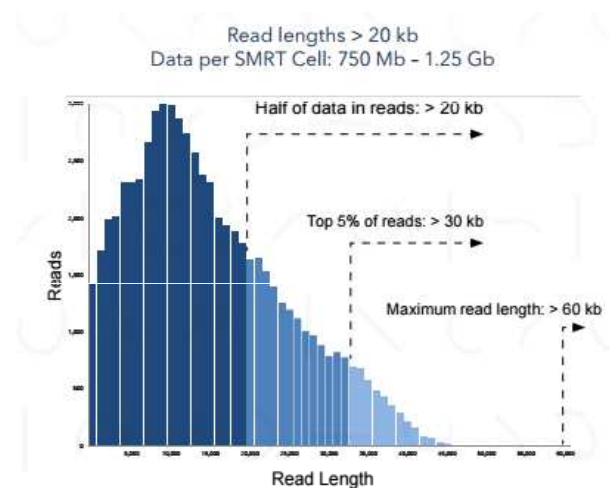




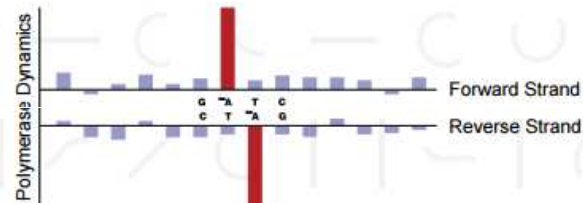
# Sekvenátor PacBio

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

- Sekvence založena na Single Molecule, Real-Time (SMRT®) technologii
- Vyžívá tzv. Zero-Mode Waveguides (ZMWs) umožňující osvětlení pouze spodní části jamky, ve které je dole imobilizována DNA polymeráza
- Hlavní výhoda je možnost dlouhého čtení (až 20 kb)
- Další výhoda je možnost přímé detekce metylovaných bazí (epigenom)



Directly detect DNA modifications  
using polymerase kinetics

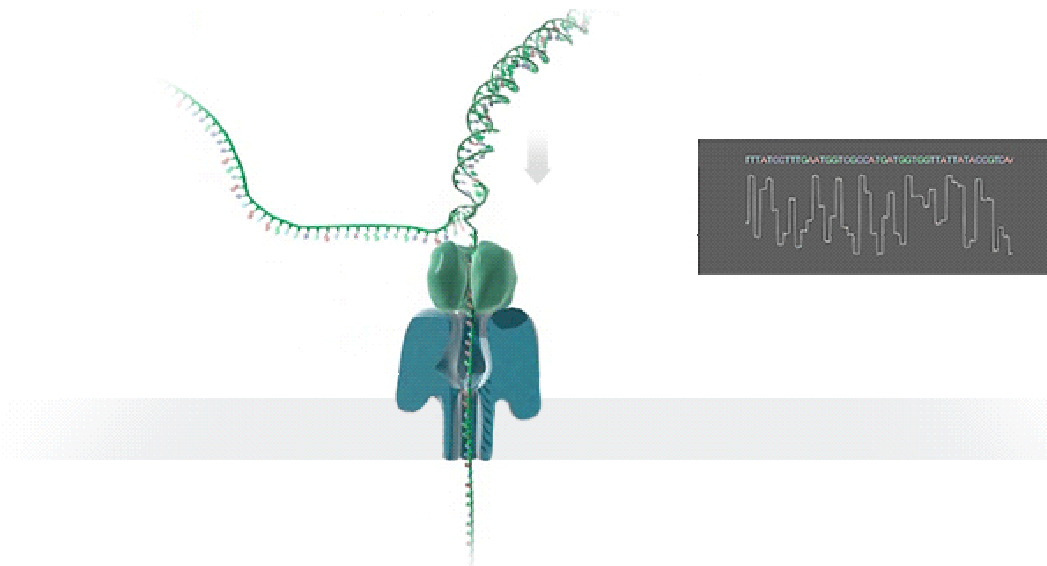


## PacBio RS II



# Sekvenátor Oxford Nanopores

- Základem technologie jsou nanopóry (nanodíry)
- Na začátku sekvenace je NK navázána na nanopór tvořený proteinem
- Poté je rozpletena a prochází přes nanopór, což generuje změnu proudu
- Na základě pozorované změny jsou odečítány v reálném čase jednotlivé báze
- Umožňuje sekvenaci velmi dlouhých úseků (desítky až stovky kilobází)
- Nevýhodou je vyšší chybovost, správnost >90%

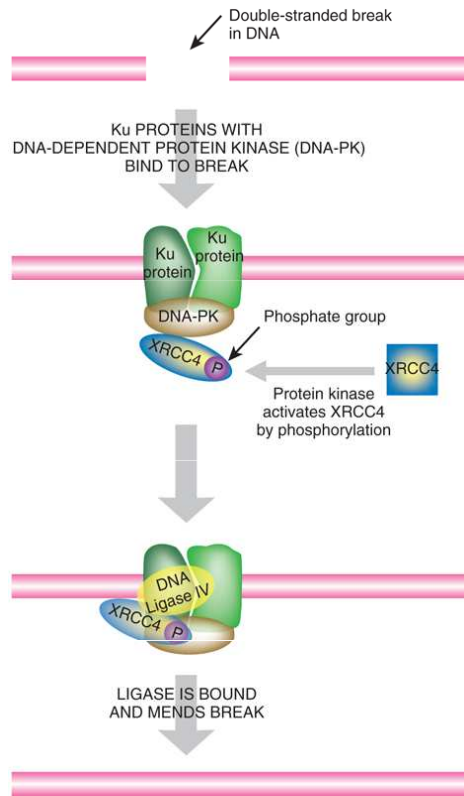


Minion

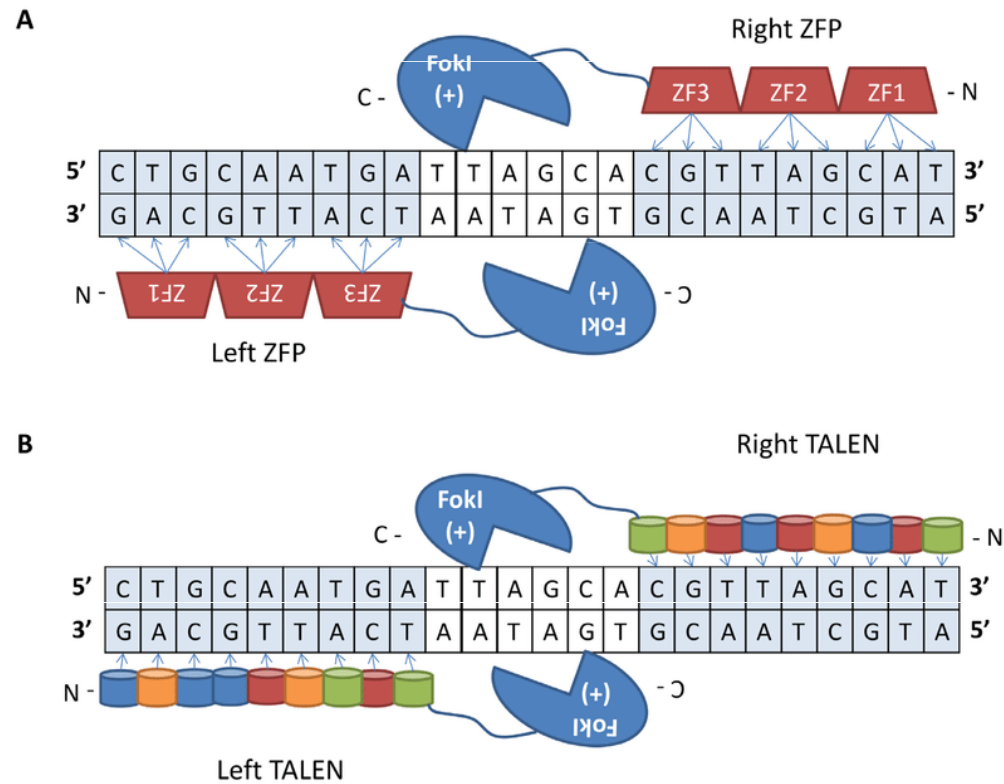
# Editace genomu

- Všechny techniky pracují na základě tvorby dvou-řetězcových zlomů
- Tyto zlomy jsou následně opraveny nehomologním párováním konců (NHEJ) nebo homologní rekombinací (HR)
- V rámci procesu může dojít k začlenění nového genu nebo vnesení krátké inserce/delece inaktivující daný gen
- Dvě základní metody pro tvorby dvou-řetězcových zlomů:
  - endonukleázy nebo restriční enzymy s dlouhou rozpoznávací sekvencí (až 40 bp)
  - použití CRISPR/Cas9 systému
- ZFNs (Zinc Finger Nucleases) – doména zinkového prstu rozpoznává sekvenci, DNA je štěpena *FokI* restriktázou.
- TALE nukleázy (TALENs) – rozpoznávací doména pochází z TALE (Transcription Activator-Like Effector) proteinu, DNA je štěpena *FokI* restriktázou

# ZFNs a TALENs

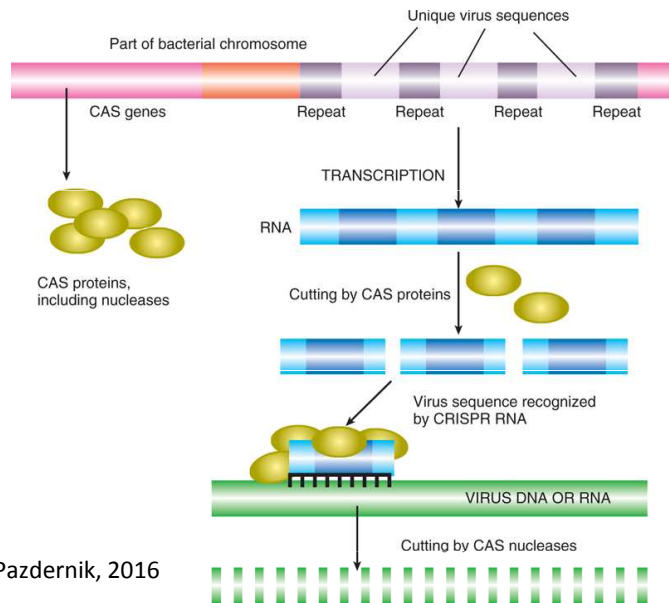


Clark and Pazdernik, 2016

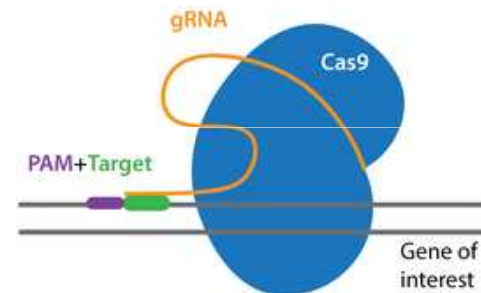


# CRISPR/Cas9

- Co-exprese endonukleázy Cas9 nebo Cas12a a gRNA specifické k cílovému genu (ropoznávaná sekvence ~20 nukleotidů)
- Nutno splnit dvě podmínky:
  1. Sekvence je unikátní v genomu
  2. Cílová sekvence je přítomna hned za Protospacer Adjacent Motif (PAM) – vazba Cas9 endonukleázy



Clark and Pazdernik, 2016



Genome Engineering  
Transcriptional Regulation  
Other Applications

