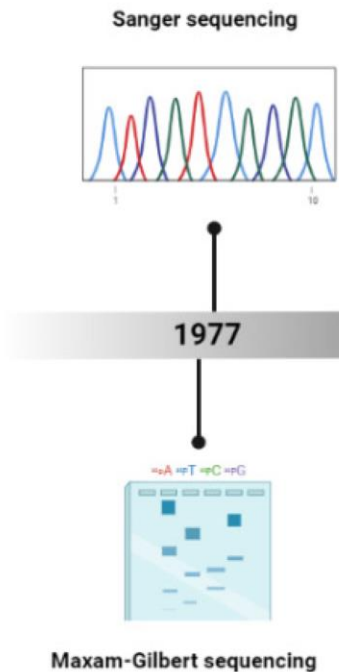


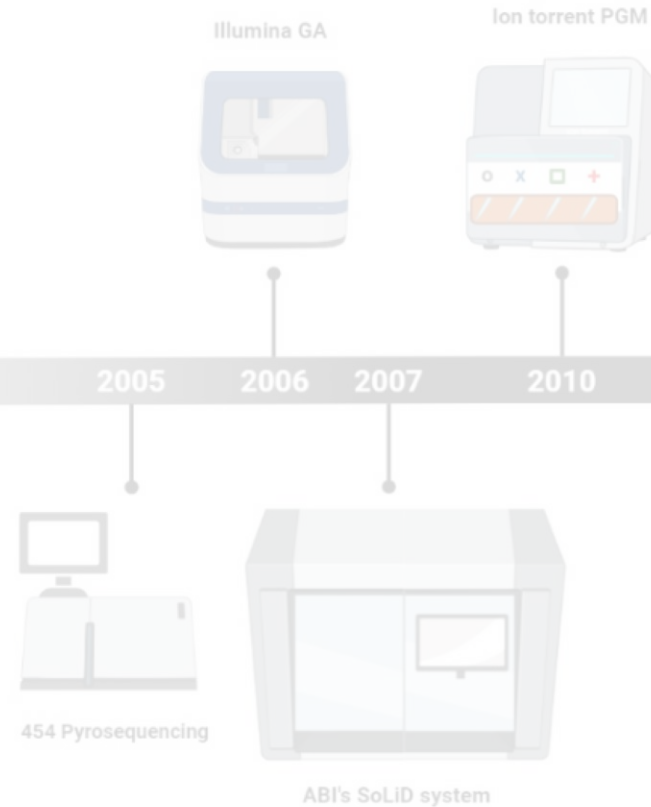
CG080 Metody v genomice: Příprava knihoven pro Next-Generation Sequencing (NGS) a Third generation Sequencing (TGS)

Přednášející: Petr Fajkus
Kontakt: fajkuspe@ibp.cz

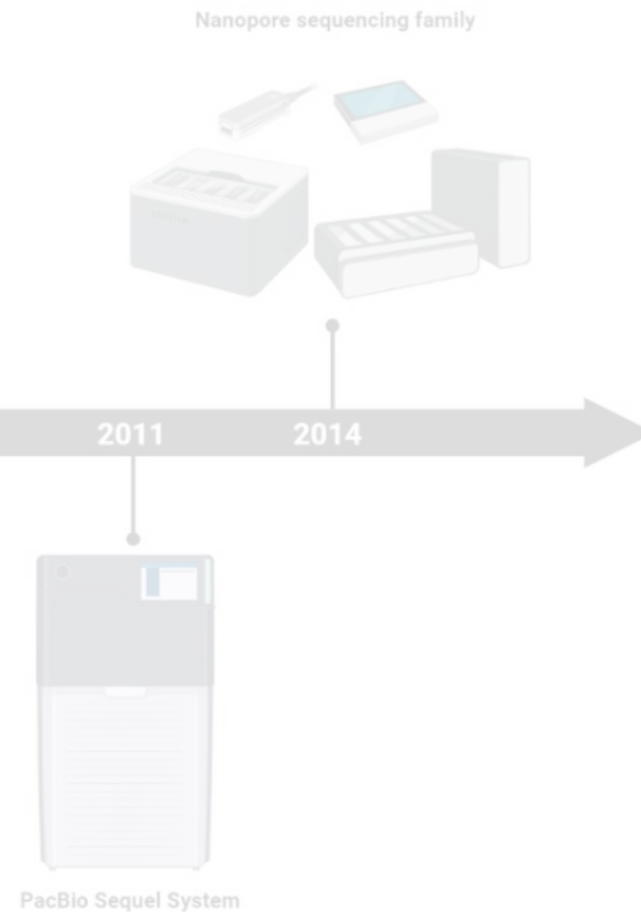
First generation sequencing



Second (Next) generation sequencing (NGS)



Third generation sequencing (TGS)



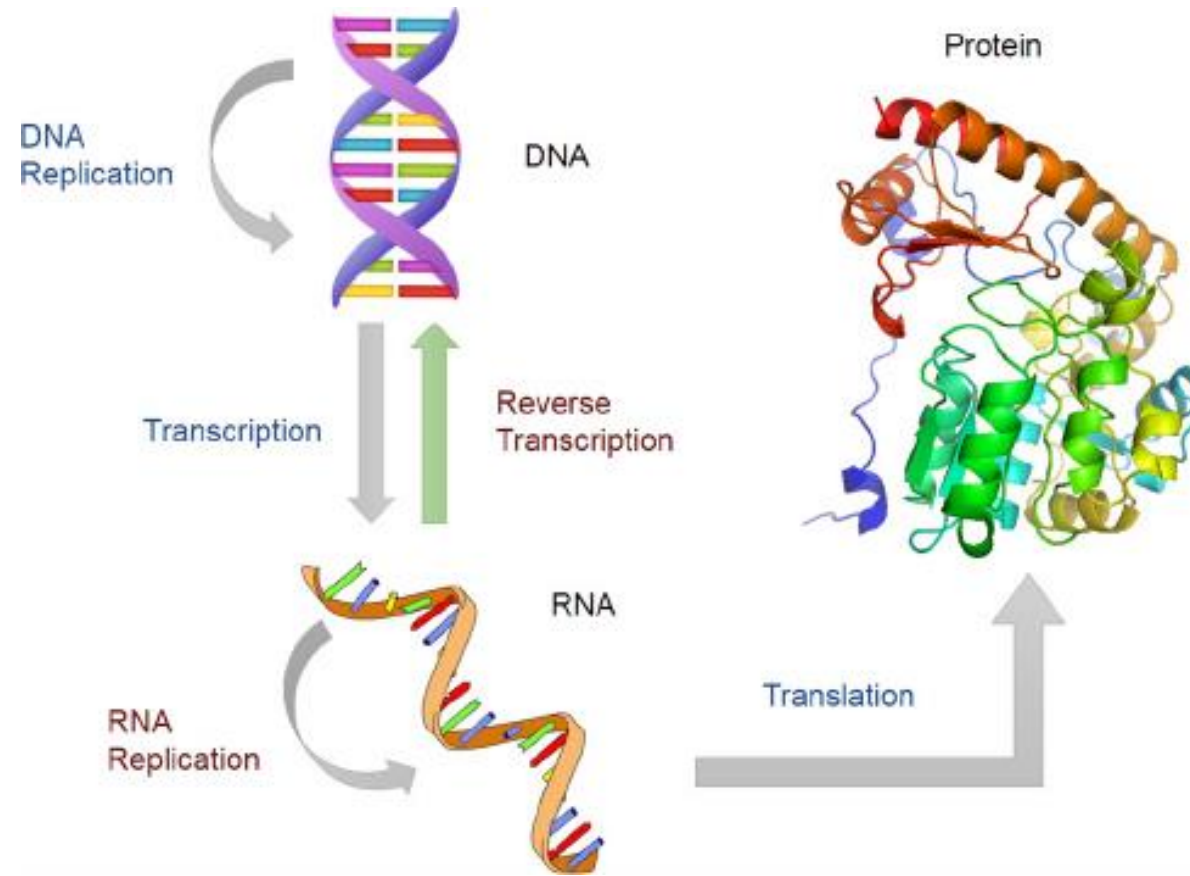
<https://doi.org/10.3390/life12010030>

Sekvenační knihovna

= soubor fragmentů DNA opatřených na koncích specifickými sekvencemi – **adaptory**.

Adaptory slouží ke specifické interakci se sekvenační platformou.

- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) Příprava NGS knihoven z RNA



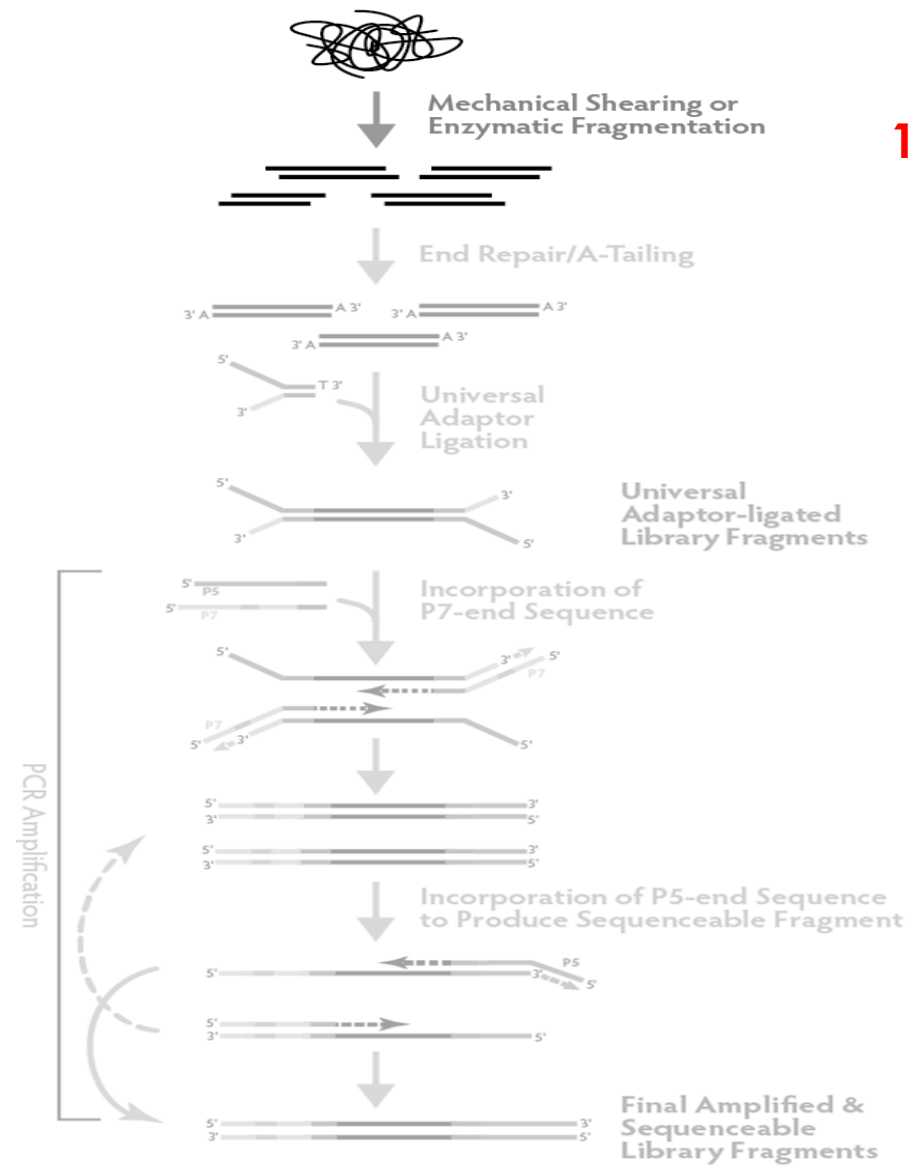
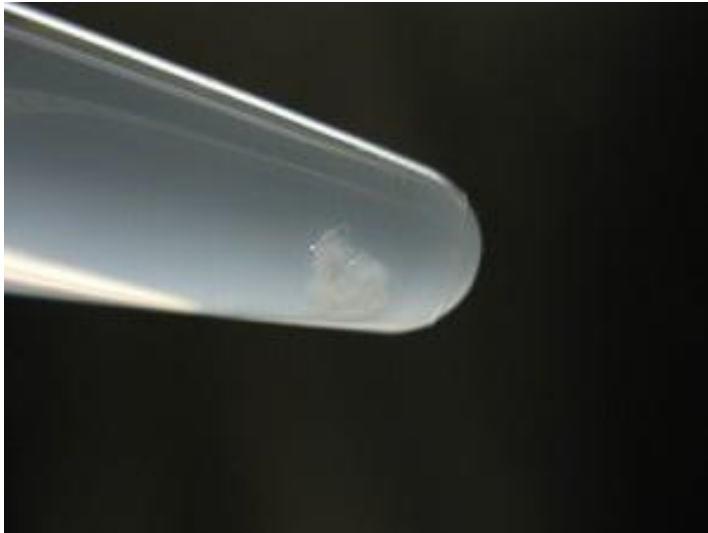
<https://doi.org/10.1016/j.bdr.2015.02.005>

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

1. Fragmentace DNA



Legend

- Starting DNA
- Universal Adaptor
- P5-end PCR Primer
- P7-end PCR Primer with Index

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA


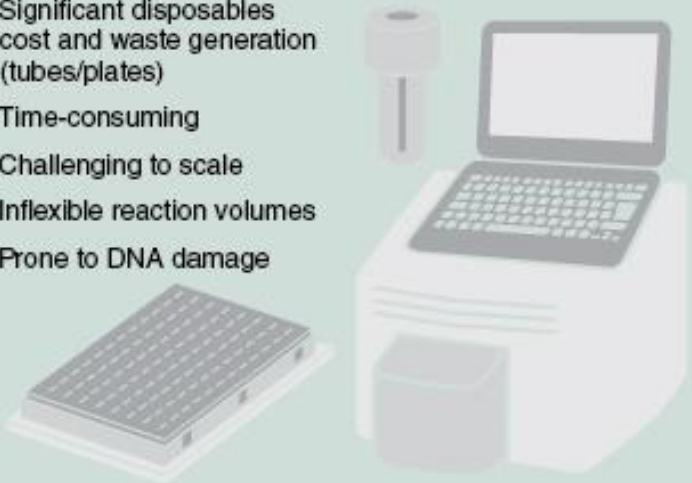
1. Fragmentace DNA

Proč fragmentovat DNA?

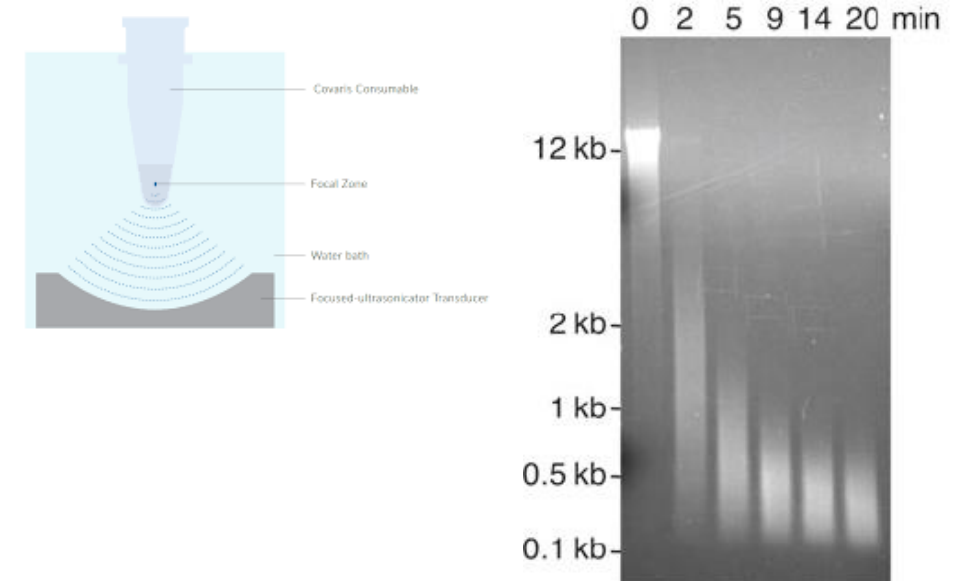
- V závislosti na použité chemii umožňují platformy Illumina sekvenovat amplikony v délce 200-800nt
- Fragmentace DNA na uniformní délku je klíčová pro dosažení maximální kvality sekvenace
- např. pro celogenomové sekvenování 350 bp fragmenty, pro Sequence Capture (pomocí hybridizačních sond) ideálně 200 bp.

Mechanická vs Enzymatická Fragmentace DNA

Focused acoustic shearing	Enzymatic fragmentation
<ul style="list-style-type: none">+ Historical "gold standard"+ Simple control of fragment size+ Same protocol regardless of GC content or input amount	<ul style="list-style-type: none">+ Scalable+ Flexible reaction volumes+ Ease of use+ Minimized DNA damage
<ul style="list-style-type: none">- Significant upfront instrument cost- Significant disposables cost and waste generation (tubes/plates)- Time-consuming- Challenging to scale- Inflexible reaction volumes- Prone to DNA damage	<ul style="list-style-type: none">- Prone to sequence bias- May be sensitive to contaminants in the DNA sample- Significant optimization may be required- Transposases can have very specific input requirements



gDNA sonication

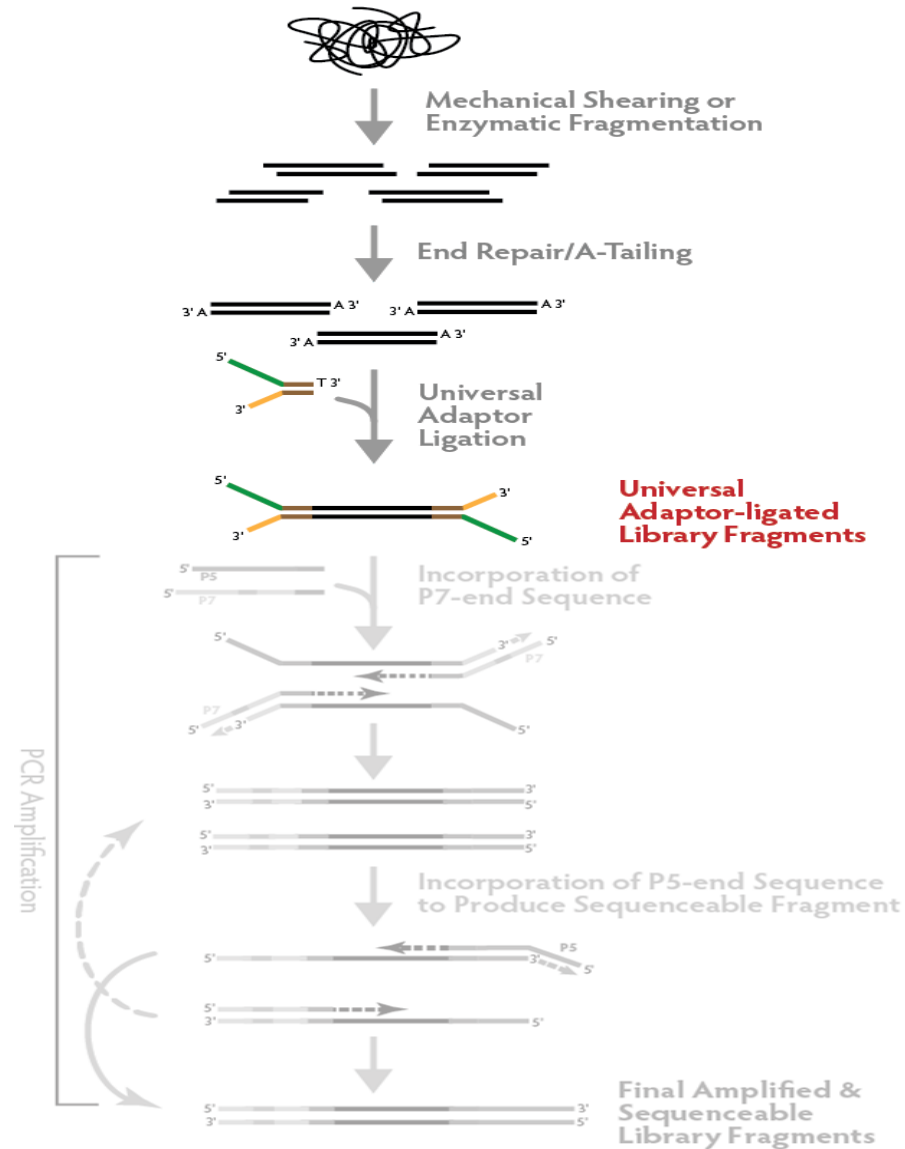


<https://www.neb.com/en/tools-and-resources/feature-articles/improving-enzymatic-dna-fragmentation-for-next-generation-sequencing-library-construction>

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA



1. Fragmentace DNA

2. Ligace adaptorů

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

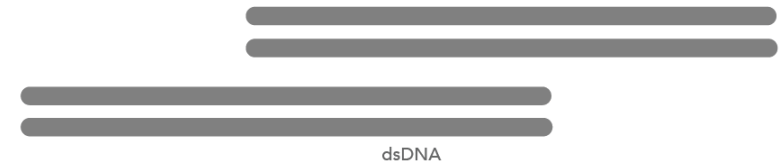
2. Ligace adaptorů

i) Enzymatická oprava konců fragmentované DNA (aby byla bez přesahů). **(End-repair)**

ii) Pro Illumina jsou pak tyto tupé konce DNA opatřeny AMP na 3' konci. **(A-Tailing)***

iii) Ligace adaptorů (s komplementárními dT přesahy)

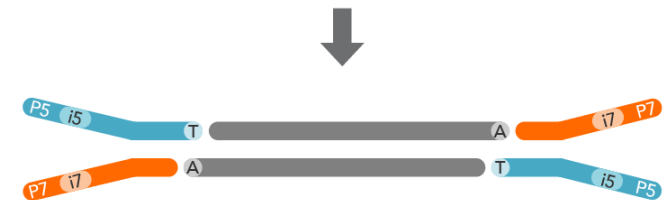
Fragmentation



End repair and A-tailing



Ligation



***A tailing** chrání před kovalentním spojováním DNA fragmentů a tím vzniku sekvenačních artefaktů

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

2. Ligace adaptorů

Single index



Unique dual index



xGen UDI-UMI adapter



 **Flow cell binding sequence:** Platform-specific sequences for library binding to instrument

 **Sequencing primer sites:** Binding sites for general sequencing primers

  **Sample indexes:** Short sequences specific to a given sample library

 **Molecular index/barcode:** Short sequence used to uniquely tag each molecule in a given sample library

 **Insert:** Target DNA or RNA fragment from a given sample library

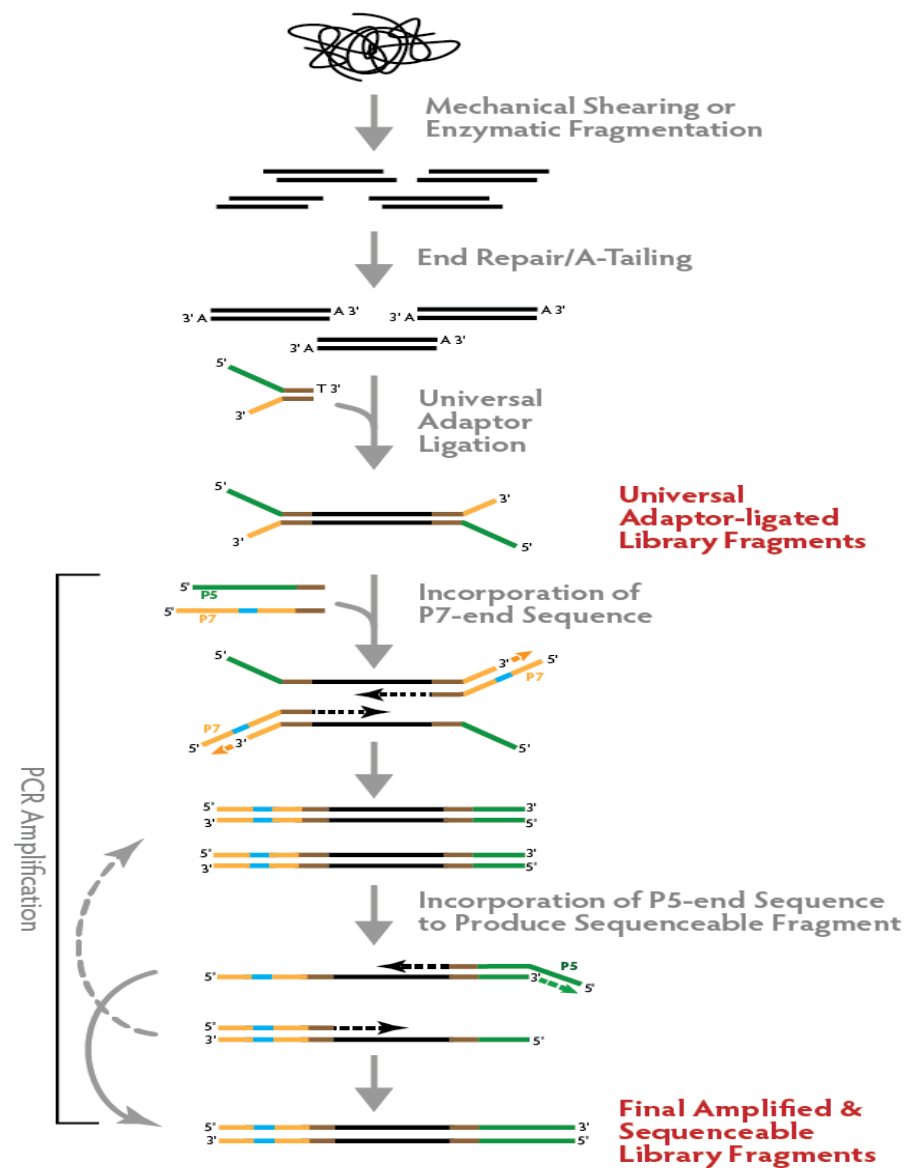
<https://sg.idtdna.com/pages/products/next-generation-sequencing/workflow/xgen-ngs-library-preparation/ngs-adapters-indexing-primers/adapters-indexing-primers-for-illumina>

Pozn. V závislosti na použitém protokolu jsou adaptory často kompletovány ve více krocích. Nejprve jsou na DNA ligovány **“univerzální“ adaptory** (v obrázku SP1 a SP2) a až v následném kroku jsou k těmto univerzálním částem připojeny (pomocí PCR) další části (např. **indexy**, **barcodey** a sekvence pro klastrování na Flowcell - **P5, P7.**)

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA



1. Fragmentace DNA

2. Ligace adaptorů

3. PCR enrichment

Legend

- Starting DNA
- Universal Adaptor
- P5-end PCR Primer
- P7-end PCR Primer with Index

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

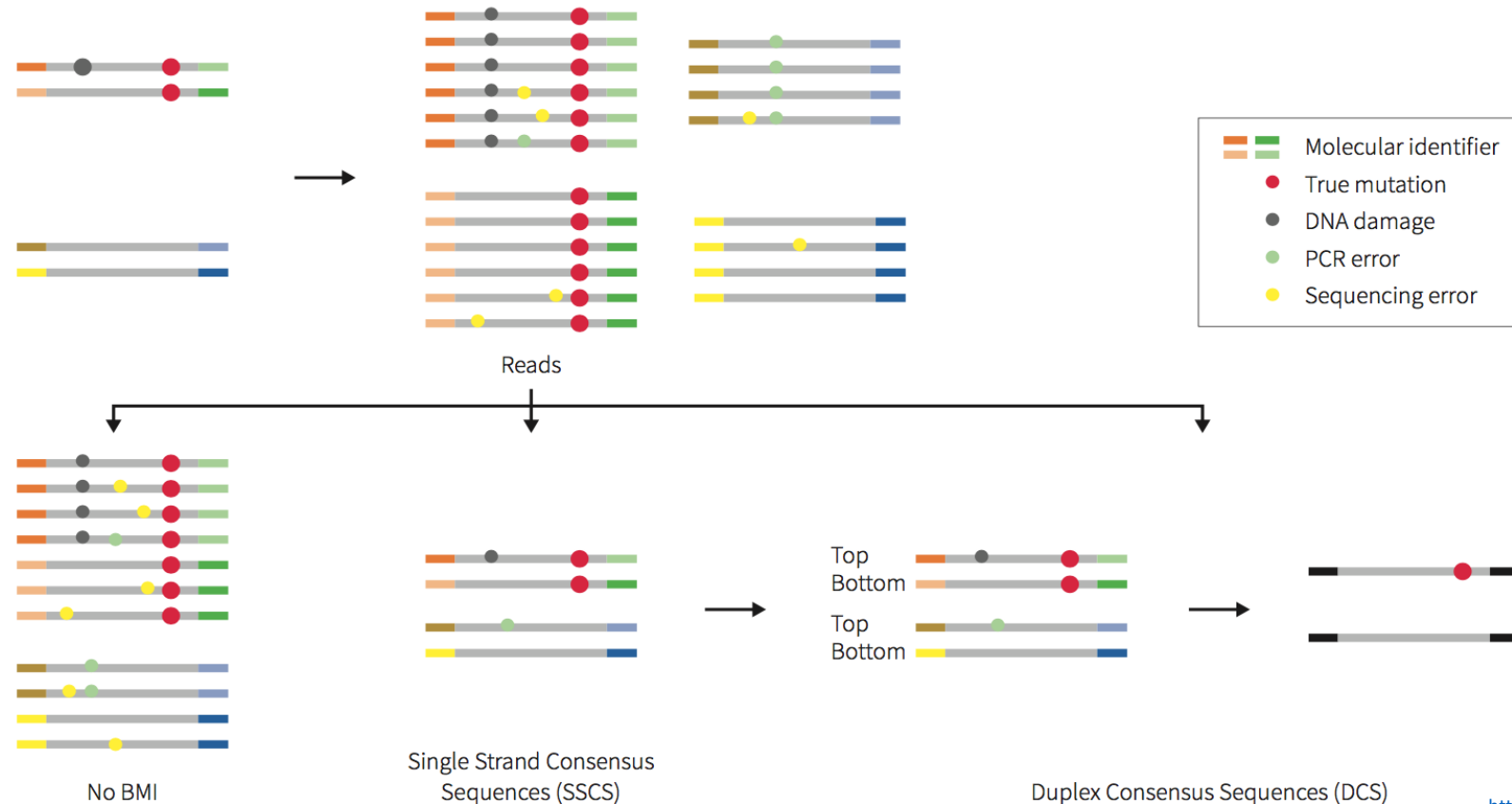
3. PCR enrichment

= namnožení fragmentů DNA opatřených adaptory*

- V rámci tohoto kroku se obvykle univerzální adaptory opatřují specifickými sekvencemi (např. **indexy** příp. **barcodey**)

Index – krátká sekvence specifická pro celou knihovnu, díky indexům lze na jeden sekvenační kit (Flow-cell) poolovat více knihoven a posléze sekvenační data rozdělit na základě sekvence indexu.

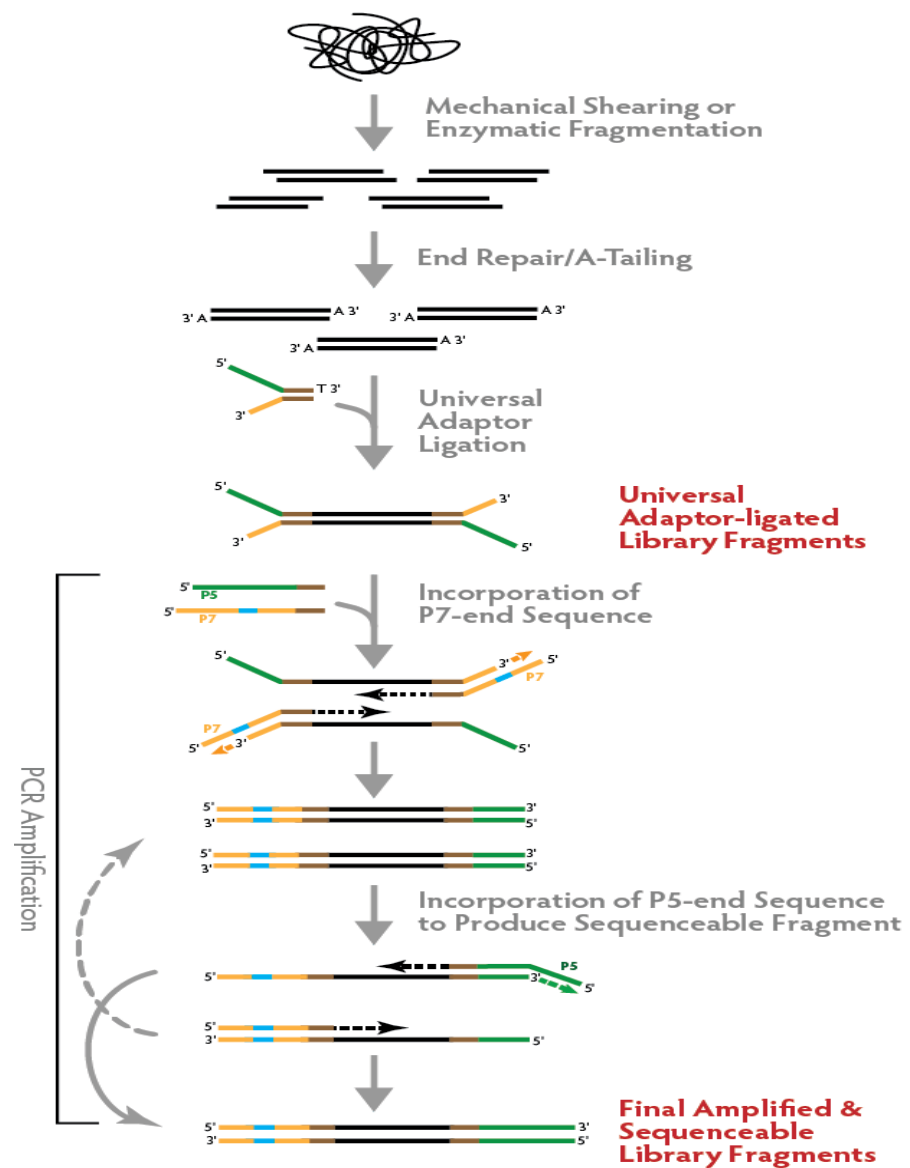
Barcode – krátká „náhodná“ sekvence, sloužící k odlišení jednotlivých amplikonů. Např. v případě identických sekvencí lze pak odlišit, zda se jedná o duplicitu téže molekuly zmnožené v PCR (= stejný Barcode) či zda se jedná o původem různé molekuly. Díky tomu je možné v sekvenačních datech spolehlivě odlišit mezi skutečnou mutací či chybou polymerázy při PCR.



a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA



1. Fragmentace DNA

2. Ligace adaptorů

3. PCR enrichment

4. Clean-up a quality check

Legend

- Starting DNA
- Universal Adaptor
- P5-end PCR Primer
- P7-end PCR Primer with Index

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

4. Clean-up a kontrola kvality

Clean-up = Purifikace DNA od všech „nepořádků“ z PCR master mixu (primery/adaptory + chemikálie), selekce fragmentů požadované délky (**size-selection**).

Na magnetických kuličkách
(SPRIselect)

Elektroforeticky (např. Pippin prep)

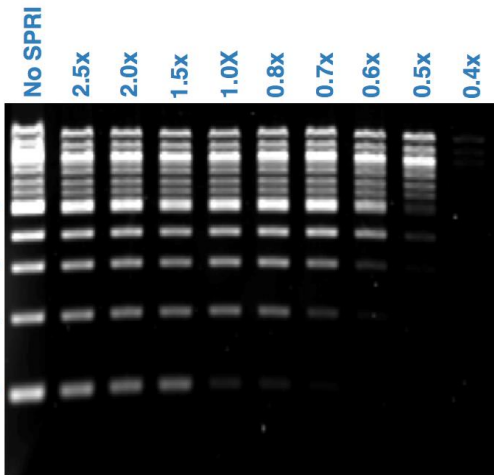
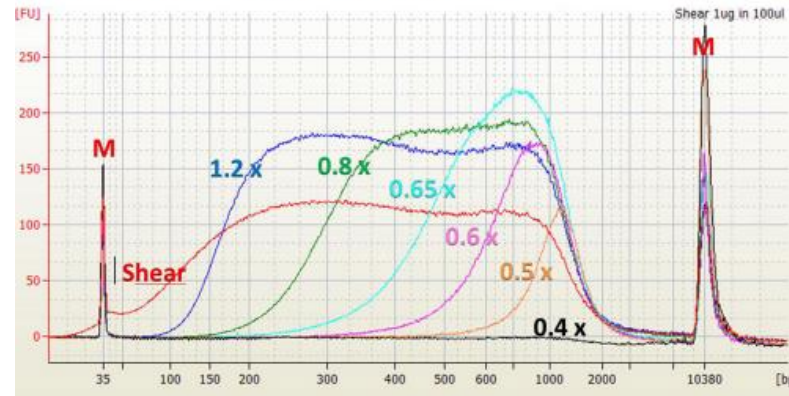


Figure 1 Agilent High Sensitivity DNA chip Electropherogram.



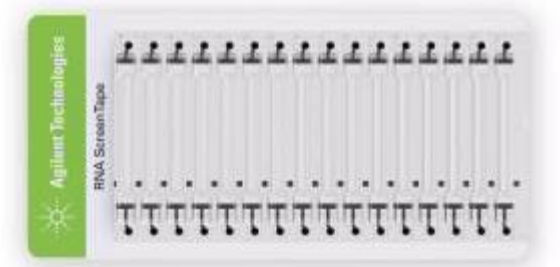
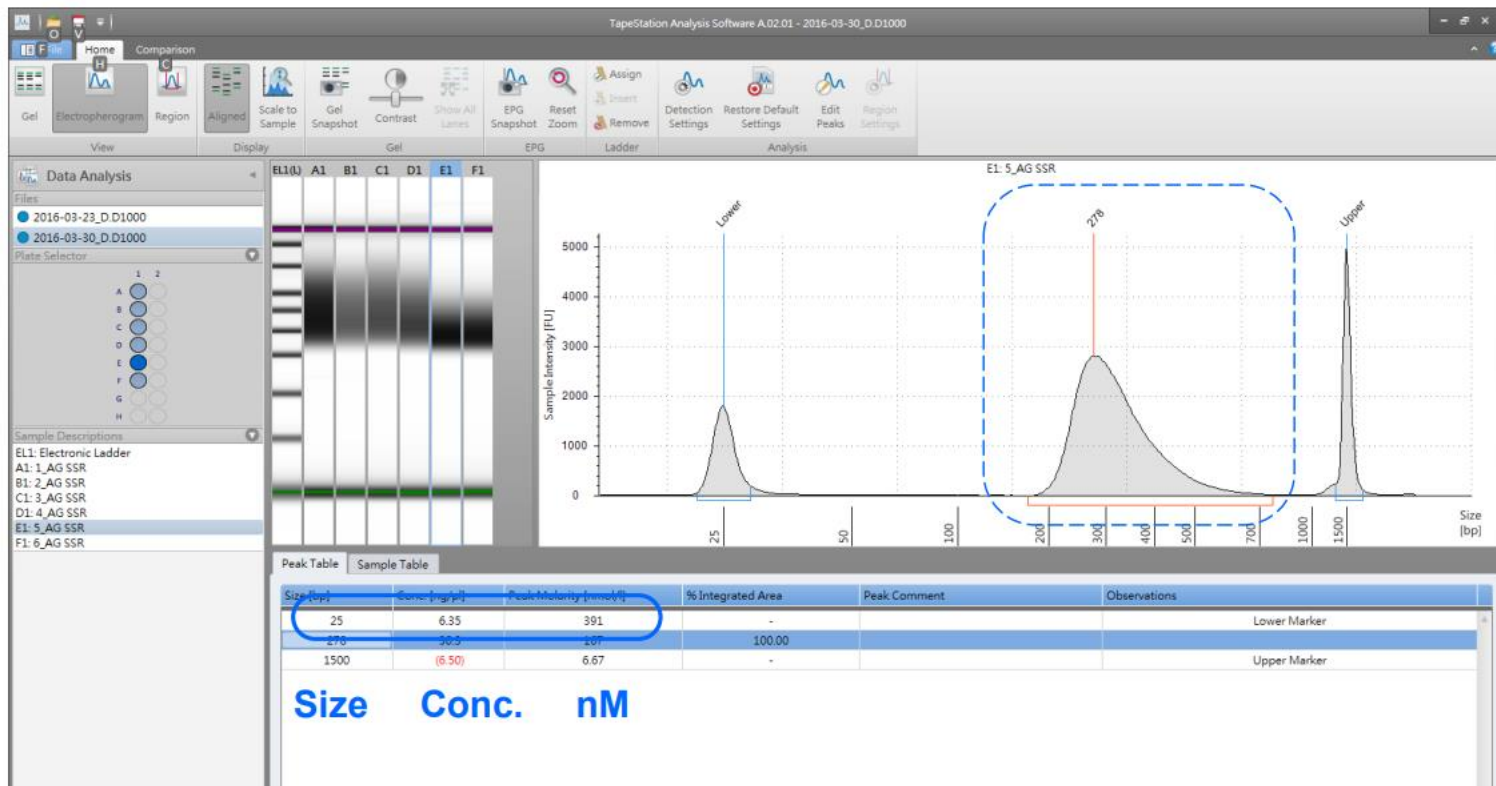
a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

4. Clean-up a kontrola kvality

Kontrola kvality knihovny = změření koncentrace knihovny a velikostního složení knihovny pomocí kapilární elektroforézy a jiných mikrofluidních systémů (např. na přístrojích Pippin prep/ Bioanalyzer/ **TapeStation**)



a) Příprava NGS knihoven z DNA

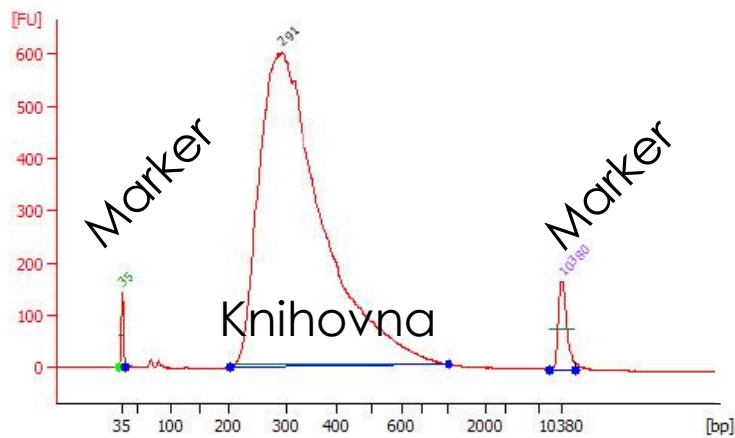
b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

4. Clean-up a kontrola kvality

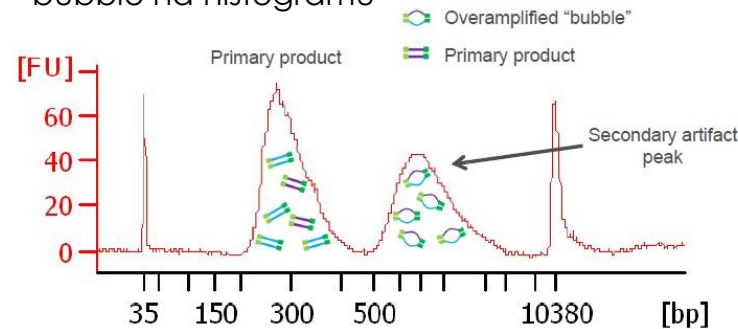
Kontrola kvality knihovny = změření koncentrace knihovny a velikostního složení knihovny pomocí kapilární elektroforózy a jiných mikrofluidních systémů (např. na přístrojích Pippin prep/ Bioanalyzer/ **TapeStation**)

V ideálním případě...

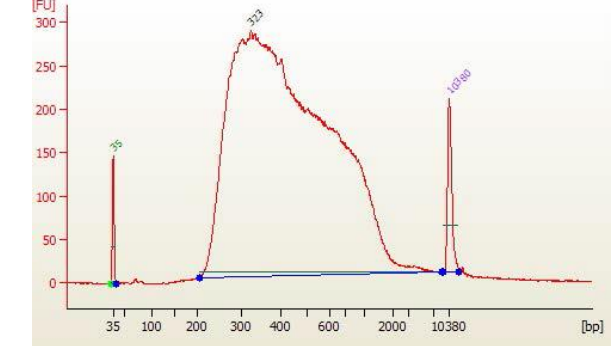


Realita...

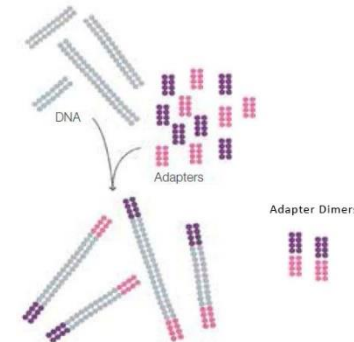
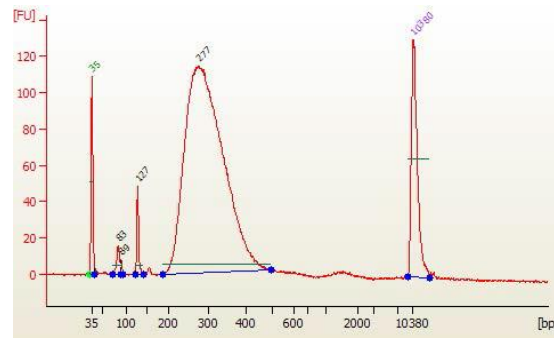
a) Přeamplifikovaná knihovna... tvorba tzv. PCR bubble na histogramu



c) Příliš dlouhé amplikony – nedostatečná fragmentace



b) Primer and Primer dimer (=Adapter dimer) kontaminace



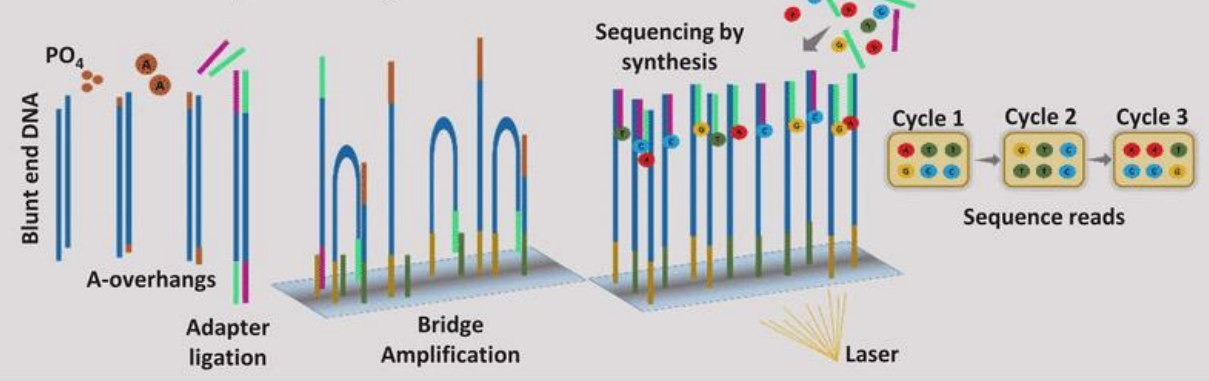
https://knowledge.illumina.com/library-preparation/general/library-preparation-general-reference_material-list/000001918

- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA**
- c) Příprava NGS knihoven z RNA

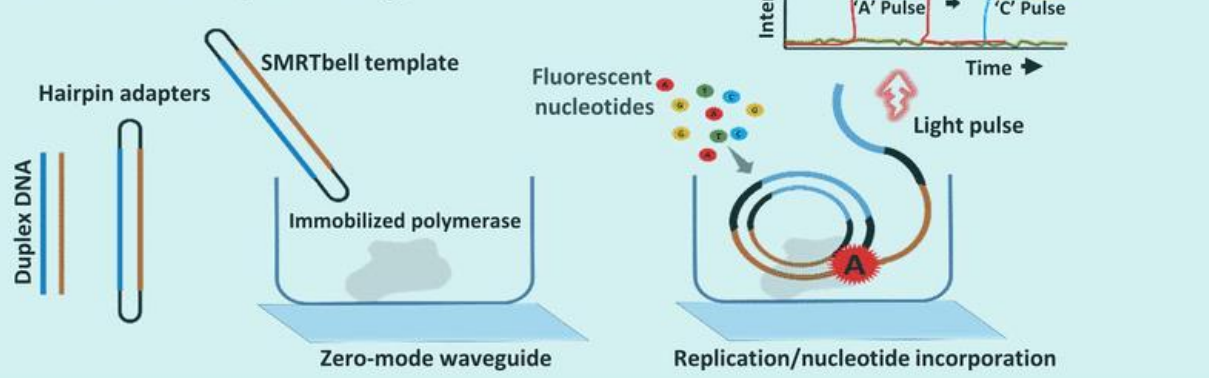
NGS

Input = fragmentovaná DNA

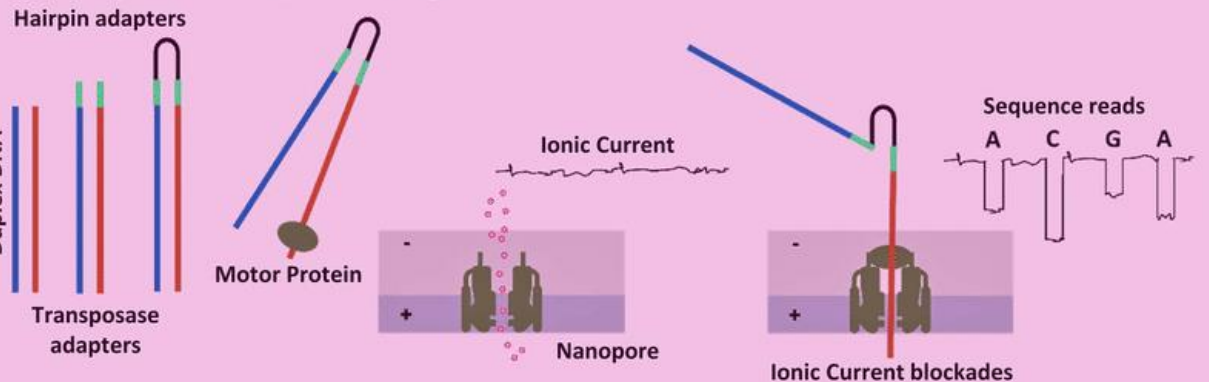
A. Illumina Sequencing



B. PacBio Sequencing



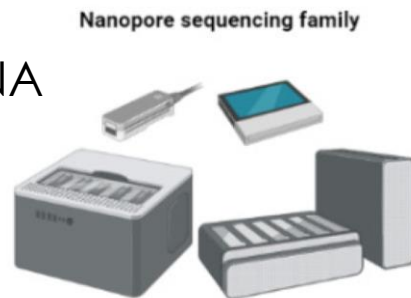
C. Nanopore Sequencing



TGS sekvenace

umožňují dlouhá čtení a detekci některých epigenetických modifikací DNA. (Nanopore rekord – 2 272 580 bází).

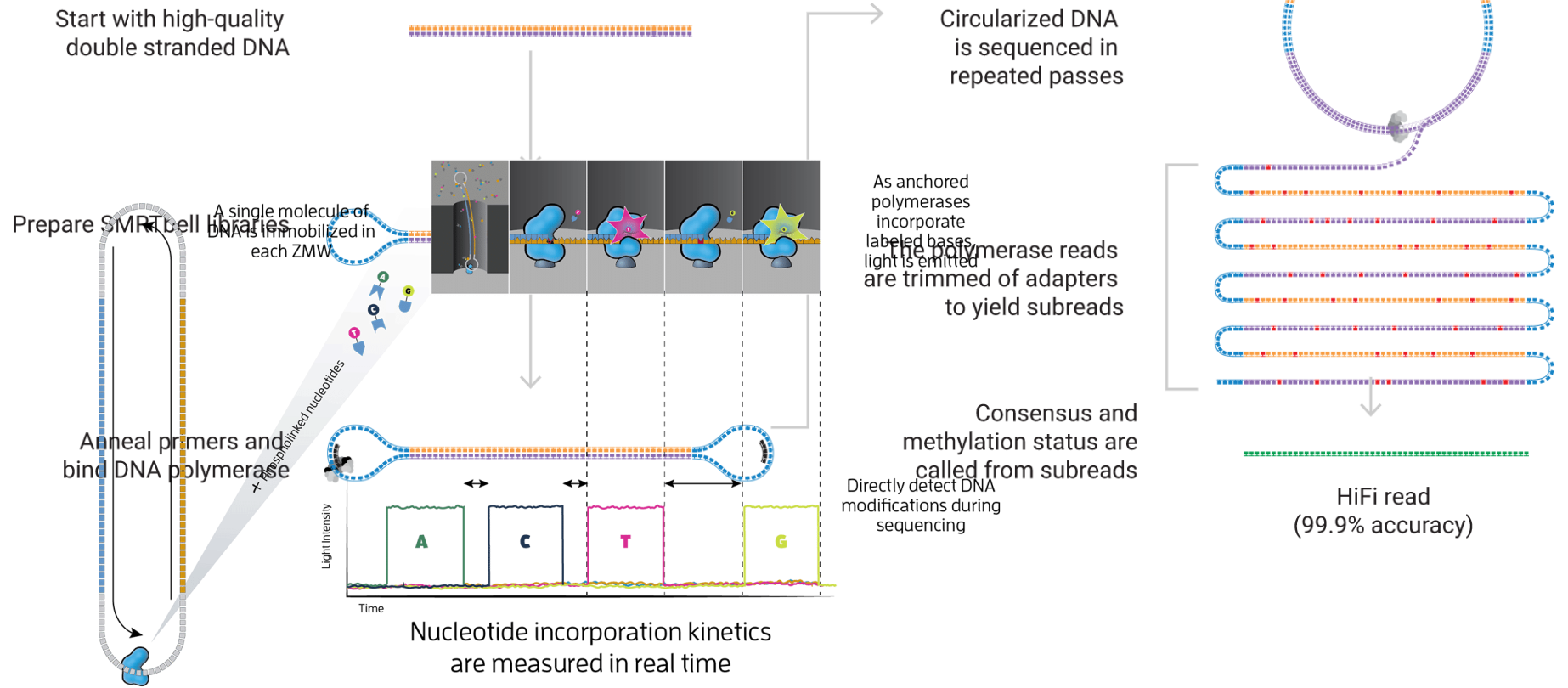
Input = vysokomolekulární DNA



a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

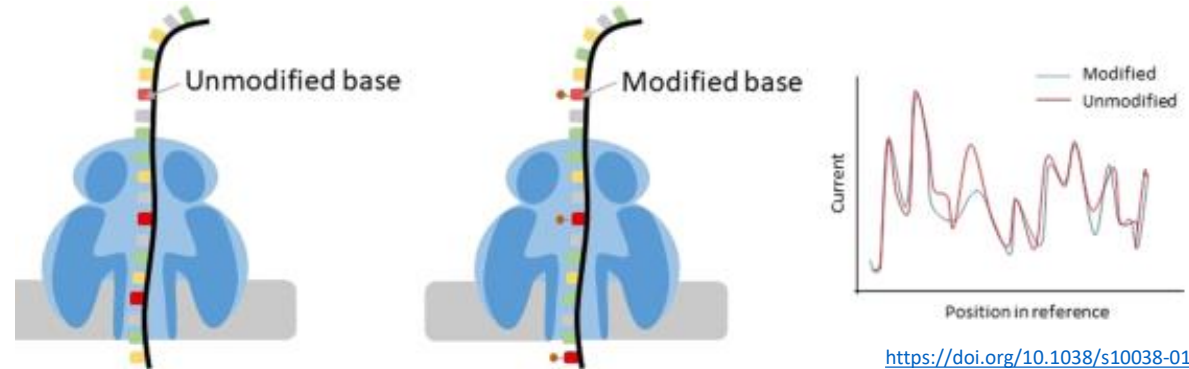
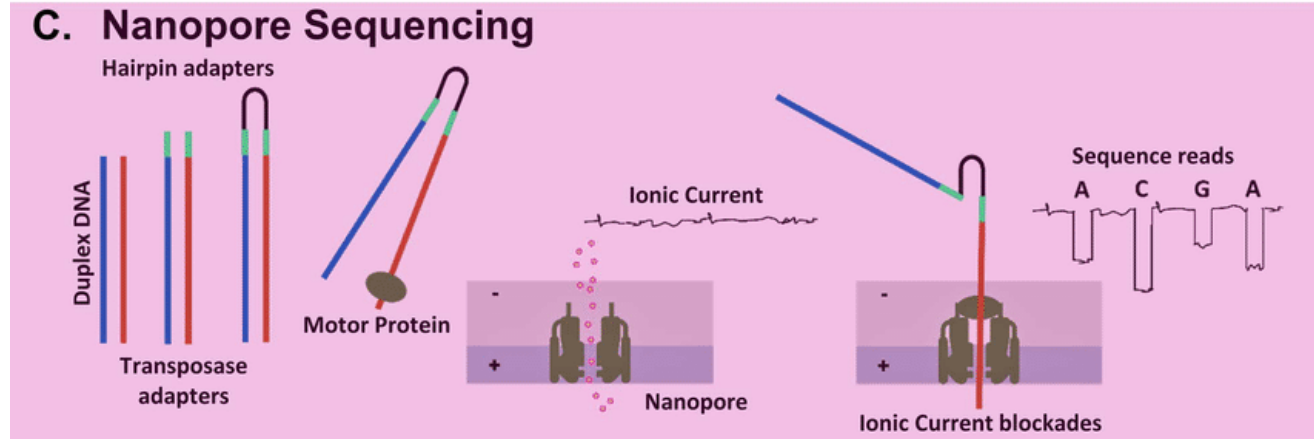
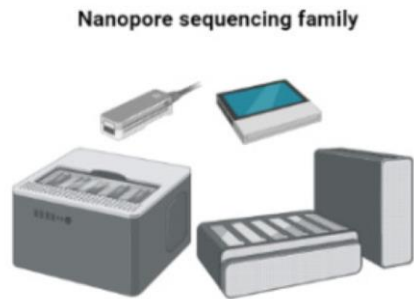


Detekce epigenetických modifikací DNA...

a) Příprava NGS knihoven z DNA

b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA



Detekce epigenetických modifikací DNA...

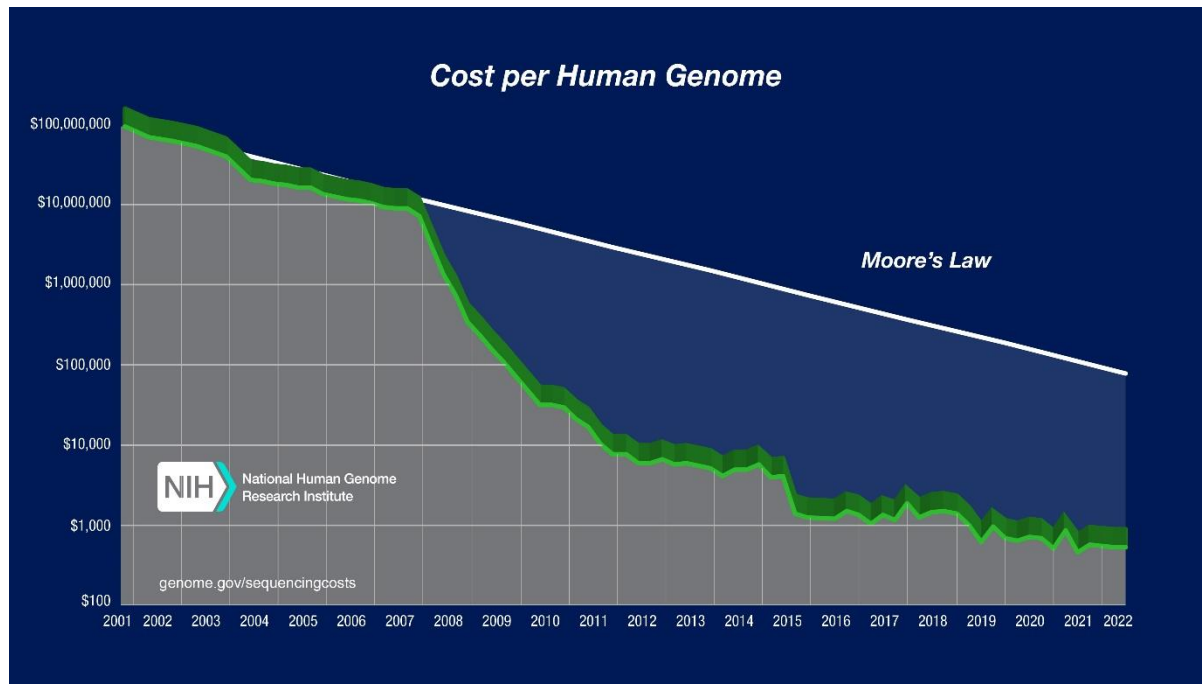
DNA-seq knihovny – časté aplikace

i) Whole genome sequencing, *de novo* assembly.

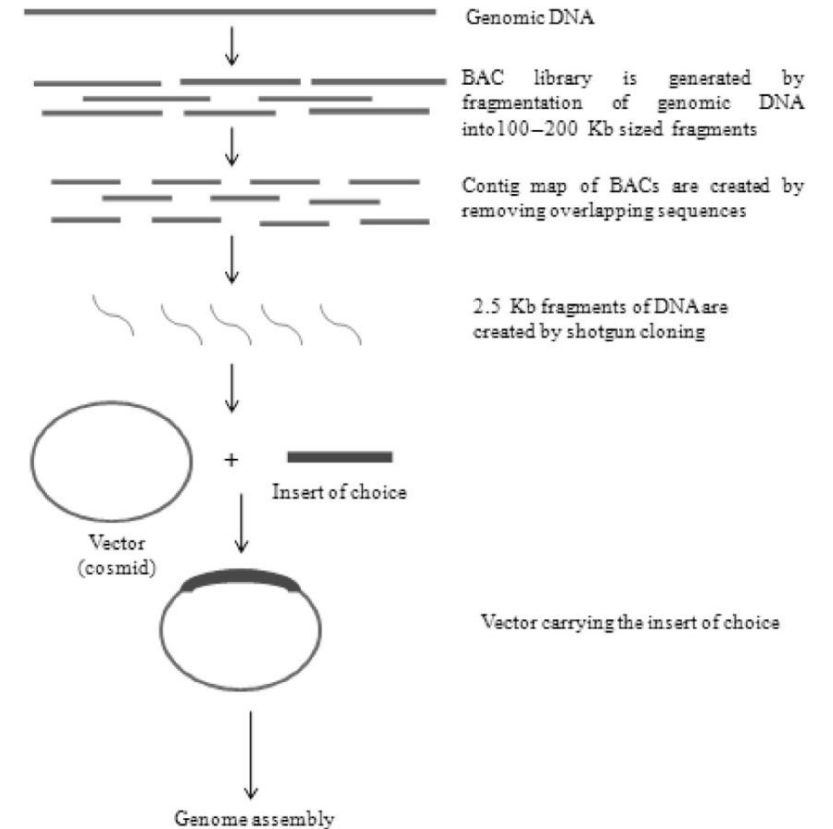
Dříve...

Human Genome Project (HGP)

- 14. 4. 2003, kompletní (~92%) genomové assembly
- Více jak 20 insitucí v projektu HGP
- přibližné celkové náklady \$2.7 mld. USD



https://en.m.wikipedia.org/wiki/File:Cost_per_Genome.png



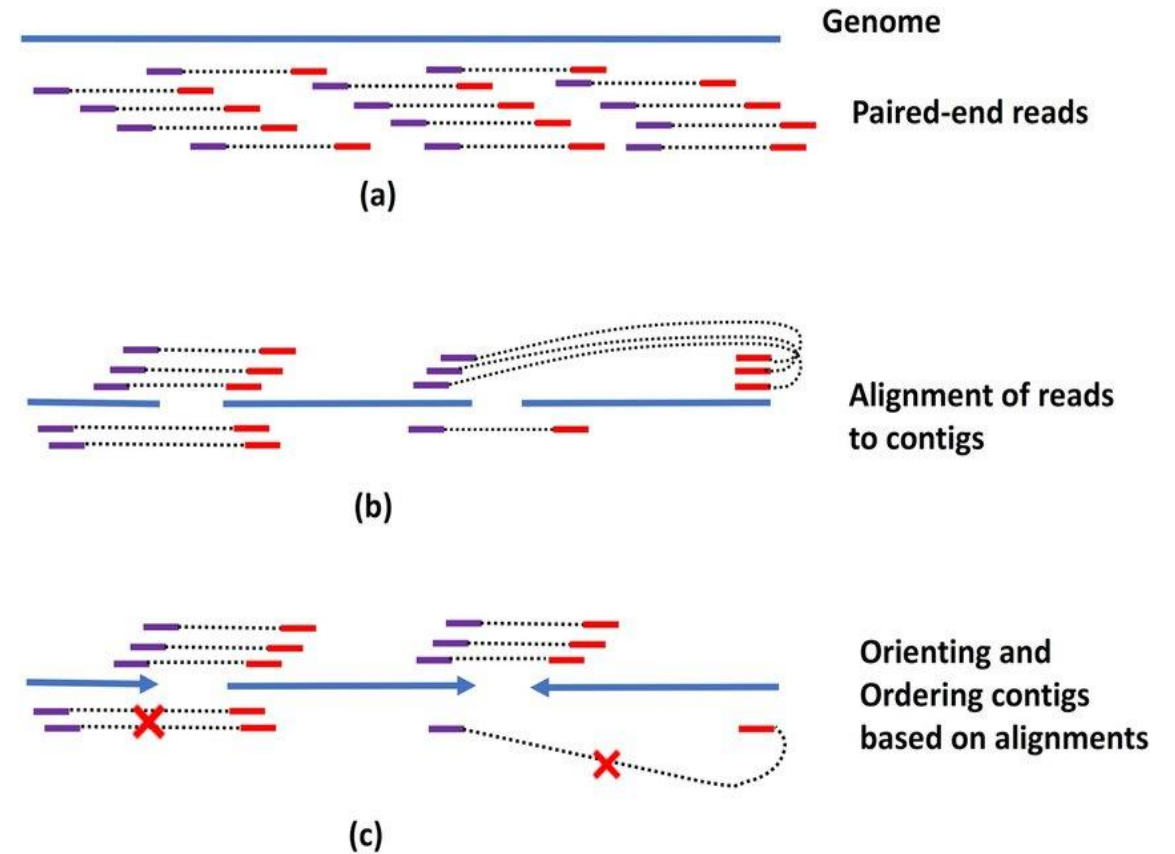
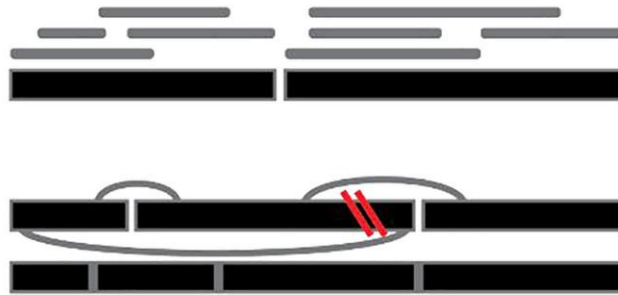
https://www.researchgate.net/figure/Outline-of-clone-by-clone-approach_fig7_327499811

DNA-seq knihovny – časté aplikace

i) Whole genome sequencing, *de novo* assembly.

Dnes: ultra-kvalitní “T2T” assembly

i) Primární assembly
dlouhých readů
(**ONT, PacBio**)



Nové sekvenační projekty např.: <https://portal.darwintreeoflife.org/tree>

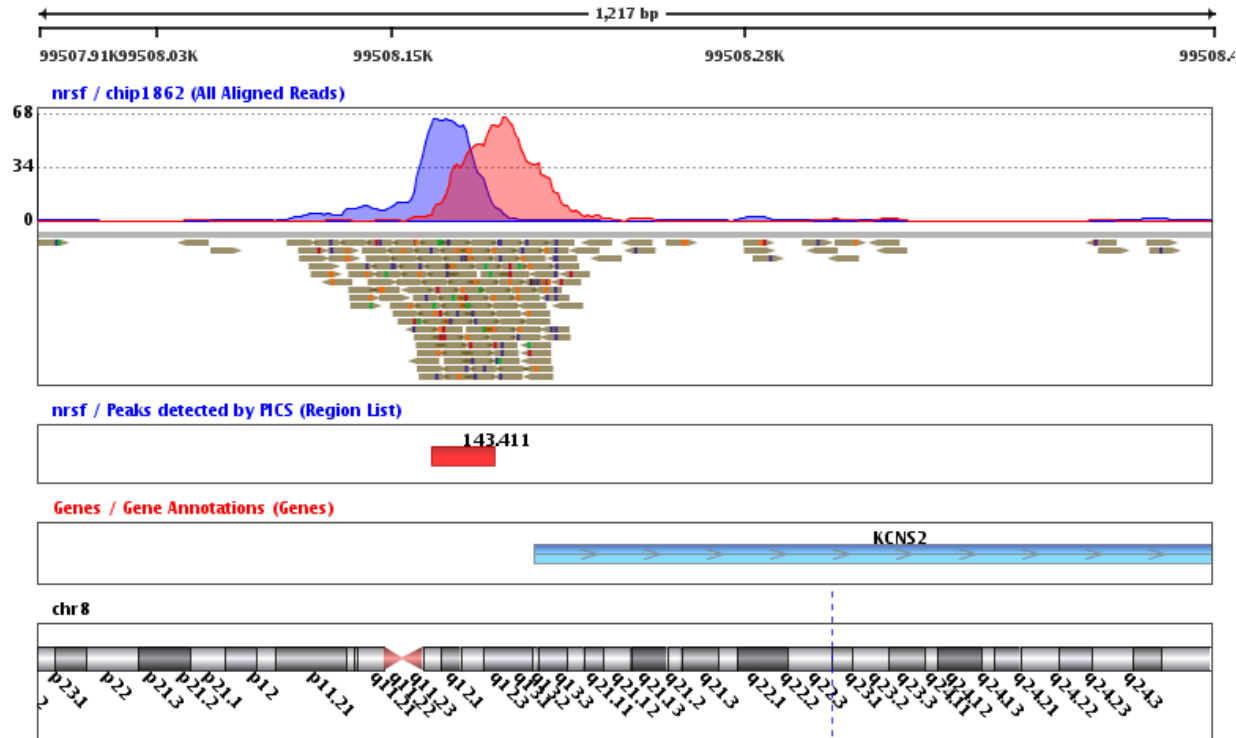
<https://doi.org/10.1371/journal.pcbi.1006994.g004>

<https://doi.org/10.1111/1755-0998.13252>

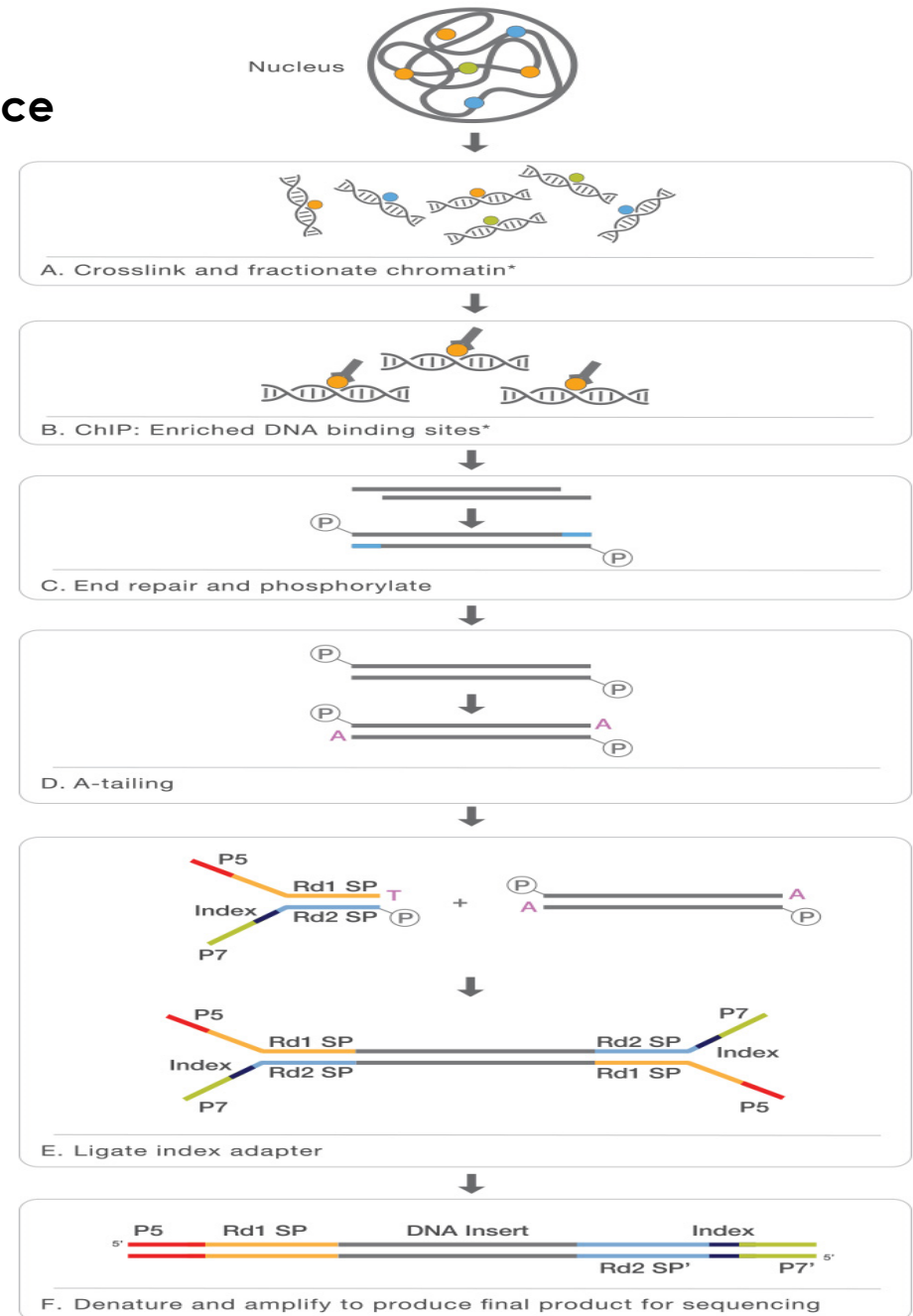
DNA-seq knihovny – časté aplikace

ii) CHIP-seq (Chromatin immunoprecipitation sequencing)

= Metoda umožňující identifikaci **DNA vazebných** míst (v celogenomovém měřítku) konkrétního DNA vazebného **proteinu**. (např. Transkripční faktory, chromatin remodelující proteiny, a jiné)



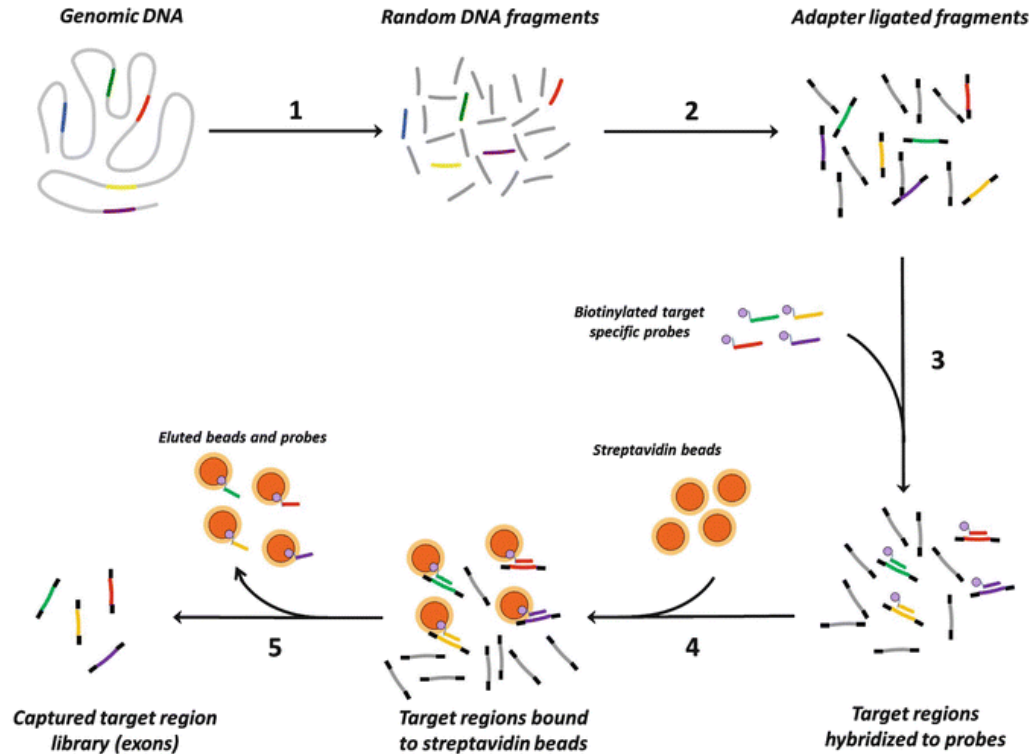
<https://www.strand-ngs.com/features/chip-seq>



<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9284502>

DNA-seq knihovny – časté aplikace

iii) Sequence capture sequencing



https://link.springer.com/chapter/10.1007/978-1-4939-0727-4_16

= metoda umožňující efektivní nabohacení sekvencí DNA (které nás zajímají) pomocí značených hybridizačních sond a jejich sekvenaci.

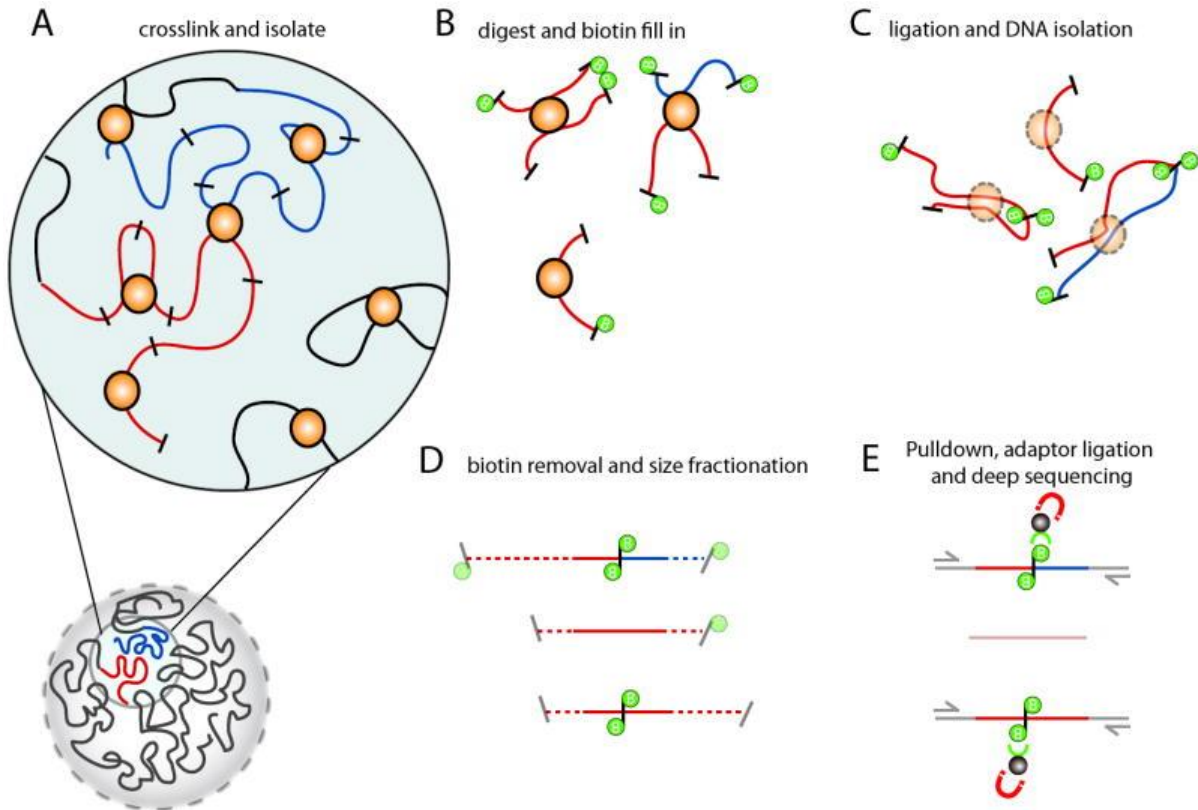
- Časté využití např. v DNA diagnostice, kdy se zkoumají případné mutace v kódujících sekvencích DNA pomocí sond vychytávající exony specifických genů (**Exome sequencing**).

Targeted panel sequencing	Whole-exome sequencing	Whole-genome sequencing
40-400 genes	22,000 genes	All genes, translocations and non-coding DNA
High coverage	Intermediate coverage	Lower coverage
Rapid (a few days), high accuracy but small number of mutations tested	Slower (a few weeks), good accuracy, many mutations tested	Slower (several weeks), all mutations tested but lower accuracy

<https://www.labroots.com/webinar/title-coming>

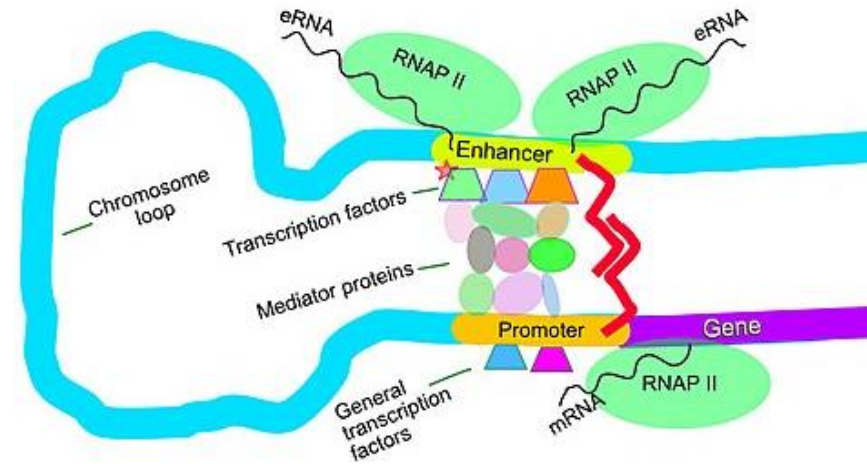
DNA-seq knihovny – časté aplikace

iv) Hi-C (Chromosome conformation capture high-throughput sequencing)



<https://doi.org/10.1016/j.ymeth.2012.05.001>

= metoda využívaná k analýze prostorového uspořádání chromatinu. Pomocí této metody se kvantifikují interakce genomických lokusů, které mohou být jinak od sebe velmi vzdálené na lineární DNA (či jiných chromozomech). Tyto interakce mohou odrážet nějakou biologickou funkci (např. **promoter-enhancer**).



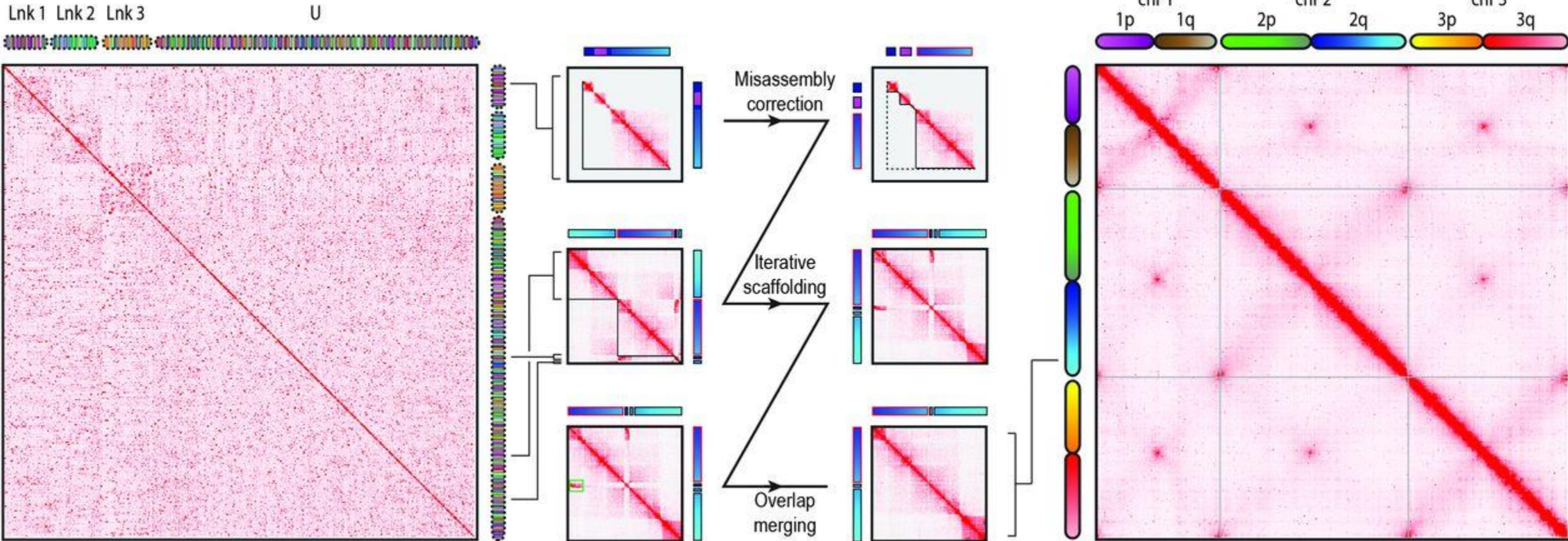
https://commons.wikimedia.org/wiki/File:Regulation_of_transcription_in_mammals.jpg

DNA-seq knihovny – časté aplikace

iv) Hi-C (Chromosome conformation capture high-throughput sequencing)

Draft assembly

End-to-end assembly

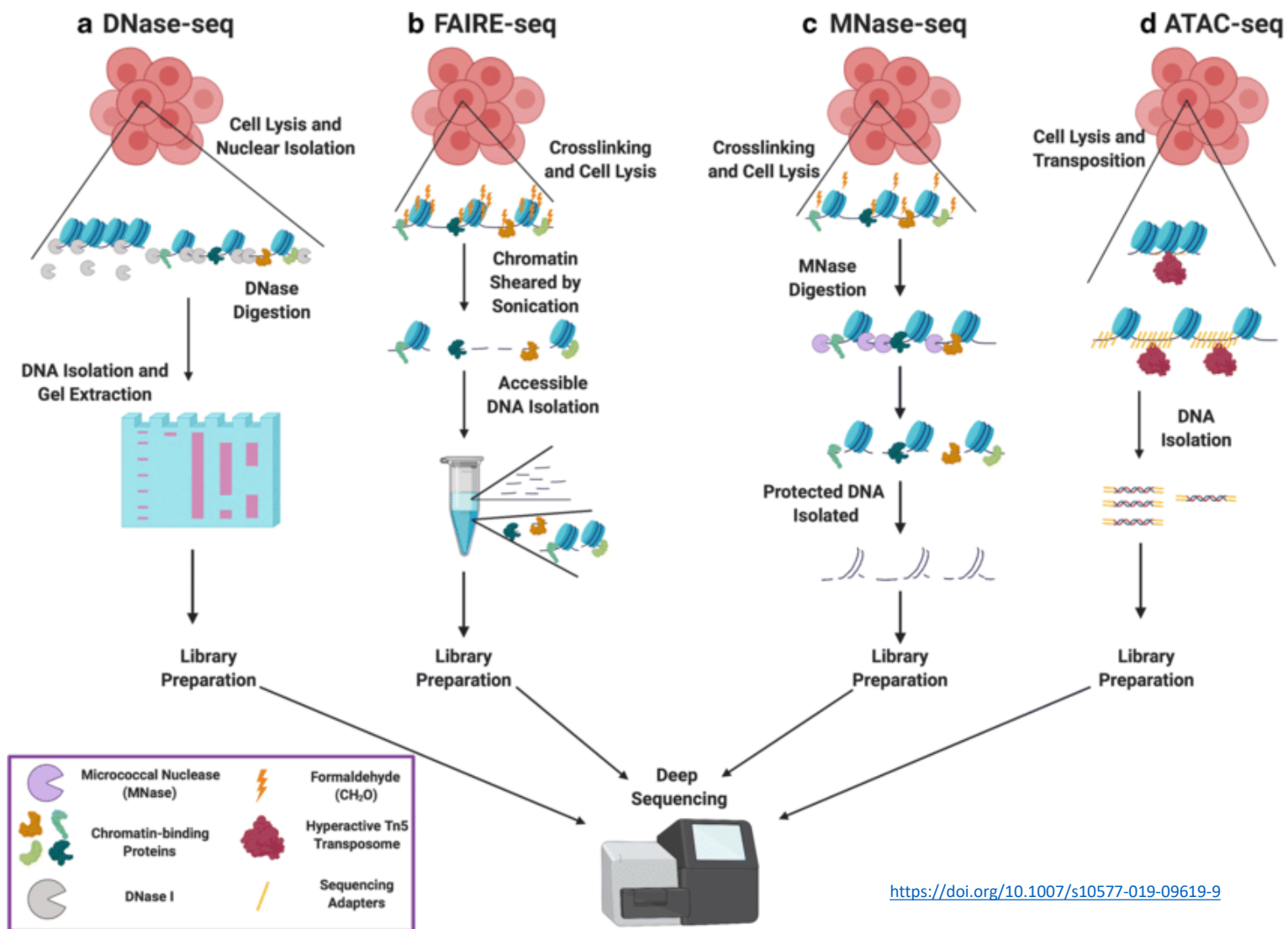


<https://www.science.org/doi/10.1126/science.aal3327>

...Optimisation of T2T assemblies

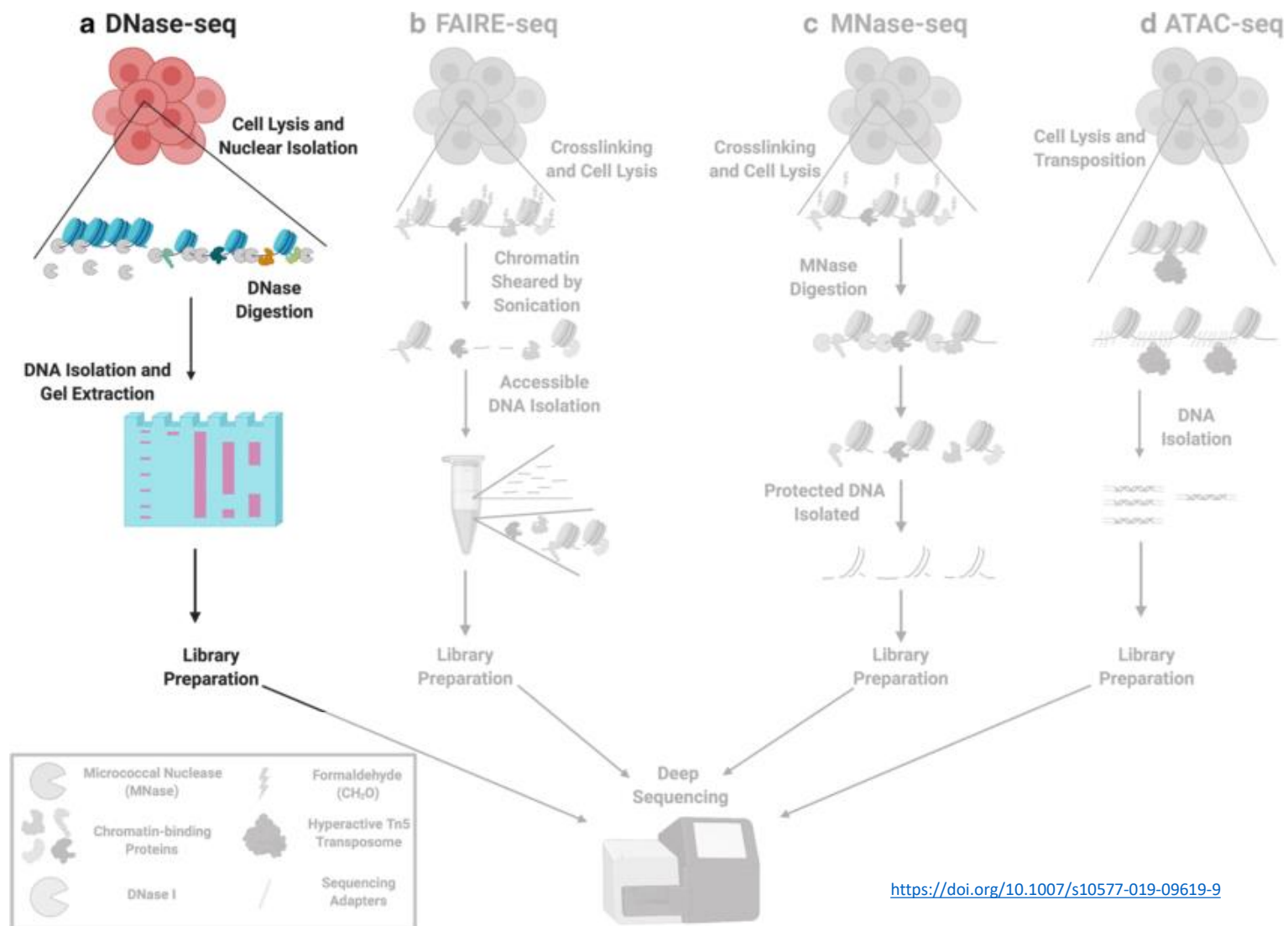
DNA-seq knihovny – časté aplikace

v) NGS metody analýzy přístupnosti chromatinu



DNA-seq knihovny – časté aplikace

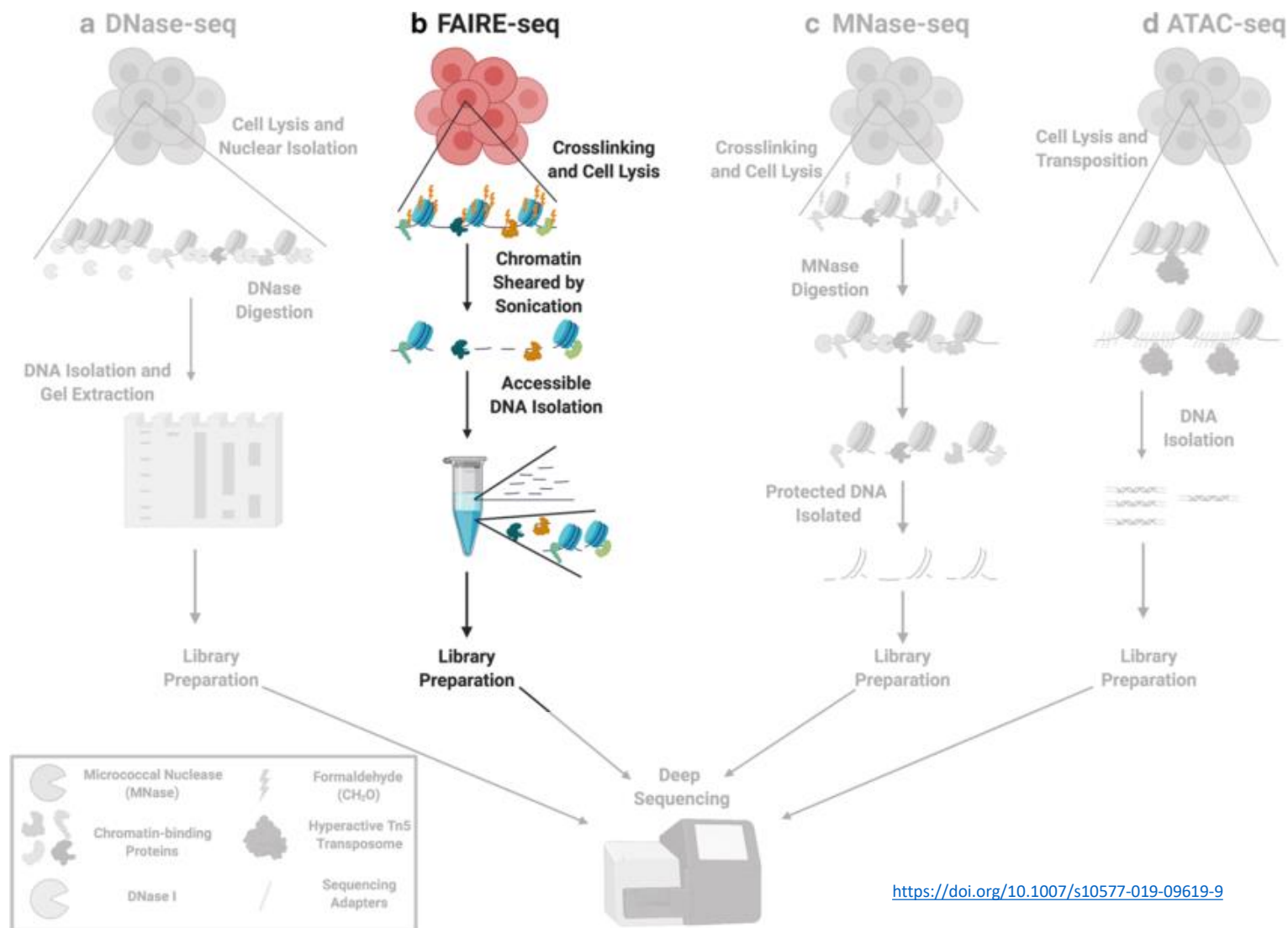
v) NGS metody analýzy přístupnosti chromatinu



DNA-seq knihovny – časté aplikace

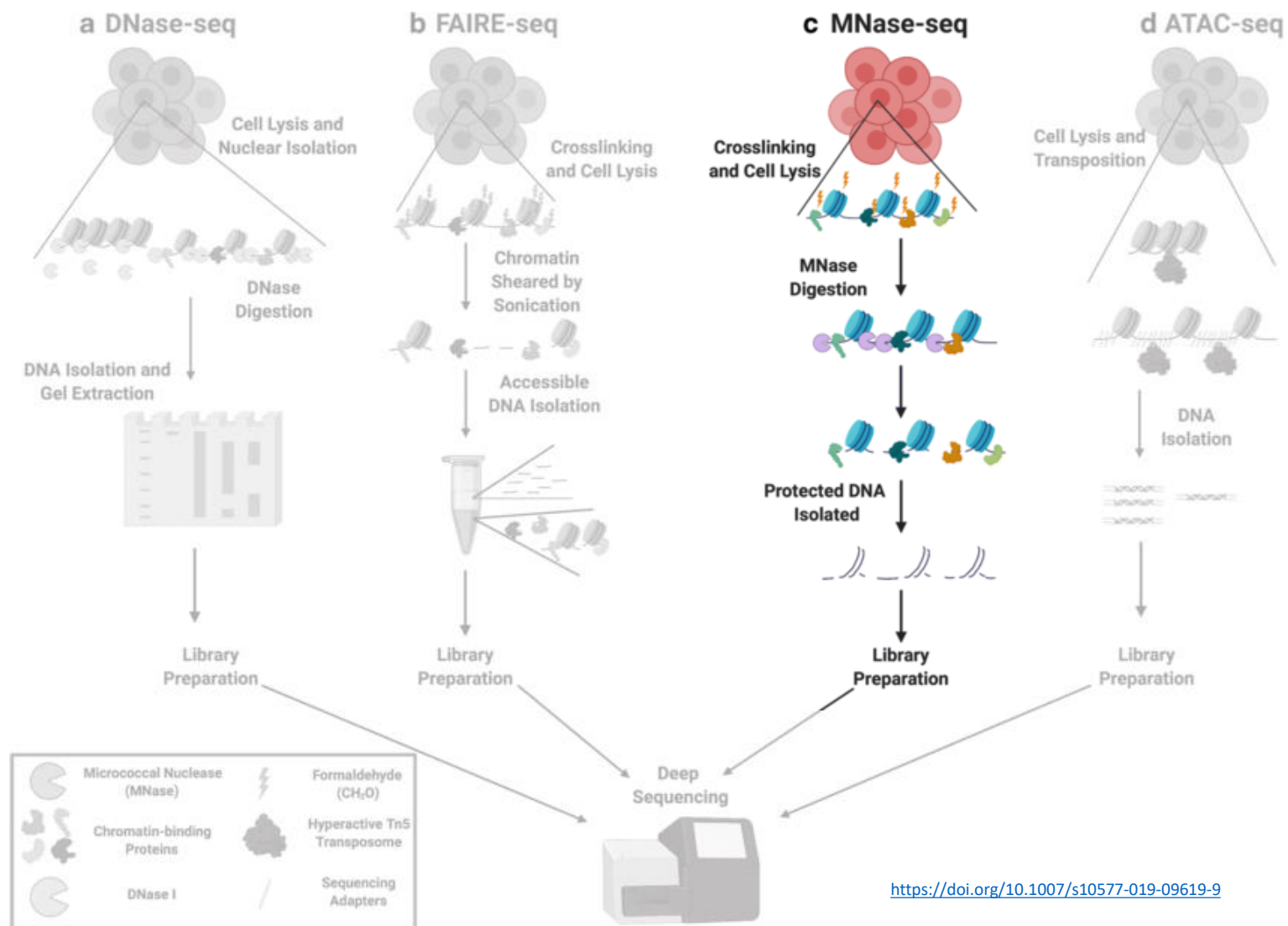
v) NGS metody analýzy přístupnosti chromatinu

Formaldehyde-Assisted Isolation of Regulatory Elements



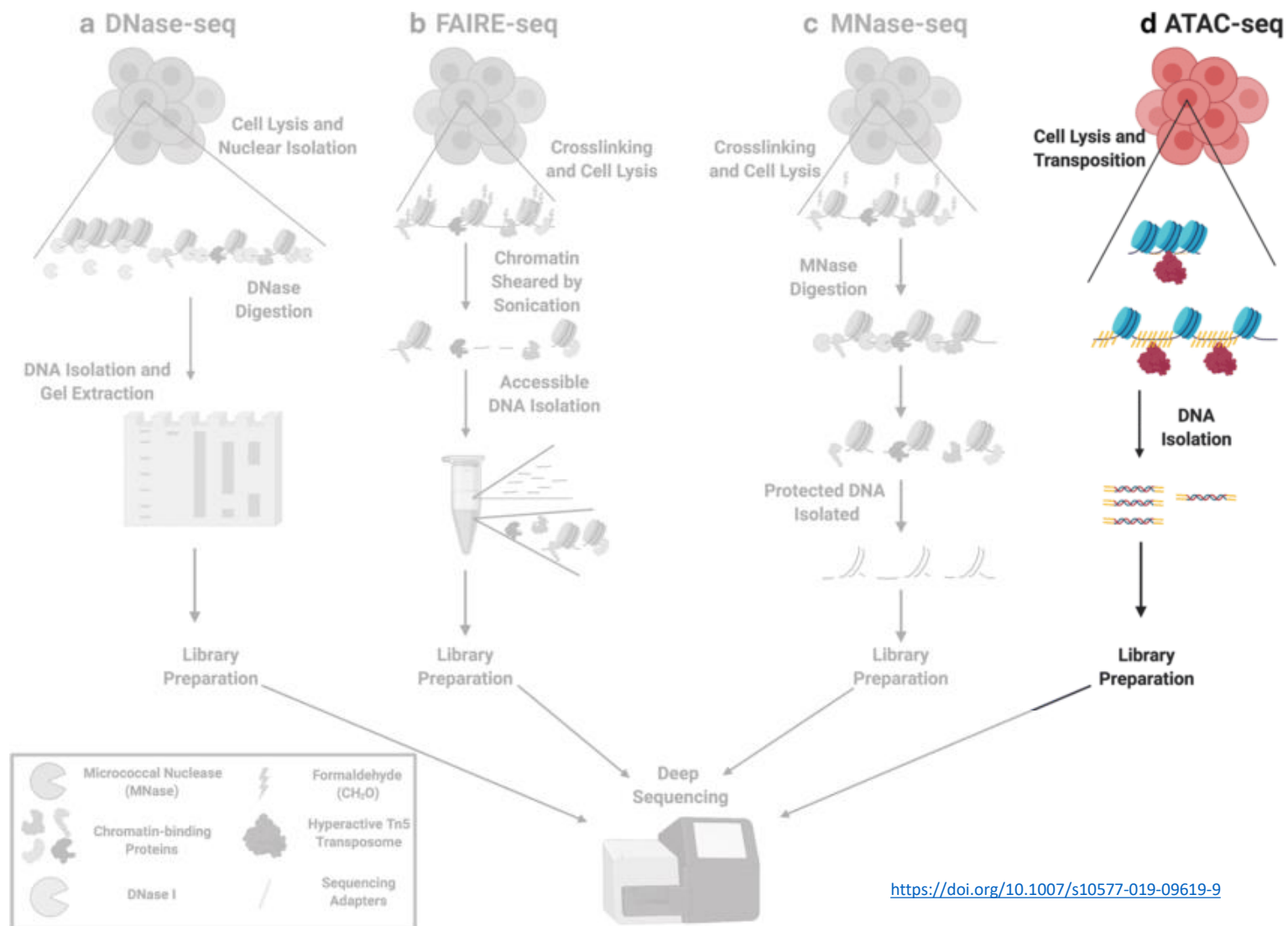
DNA-seq knihovny – časté aplikace

v) NGS metody analýzy přístupnosti chromatinu



DNA-seq knihovny – časté aplikace

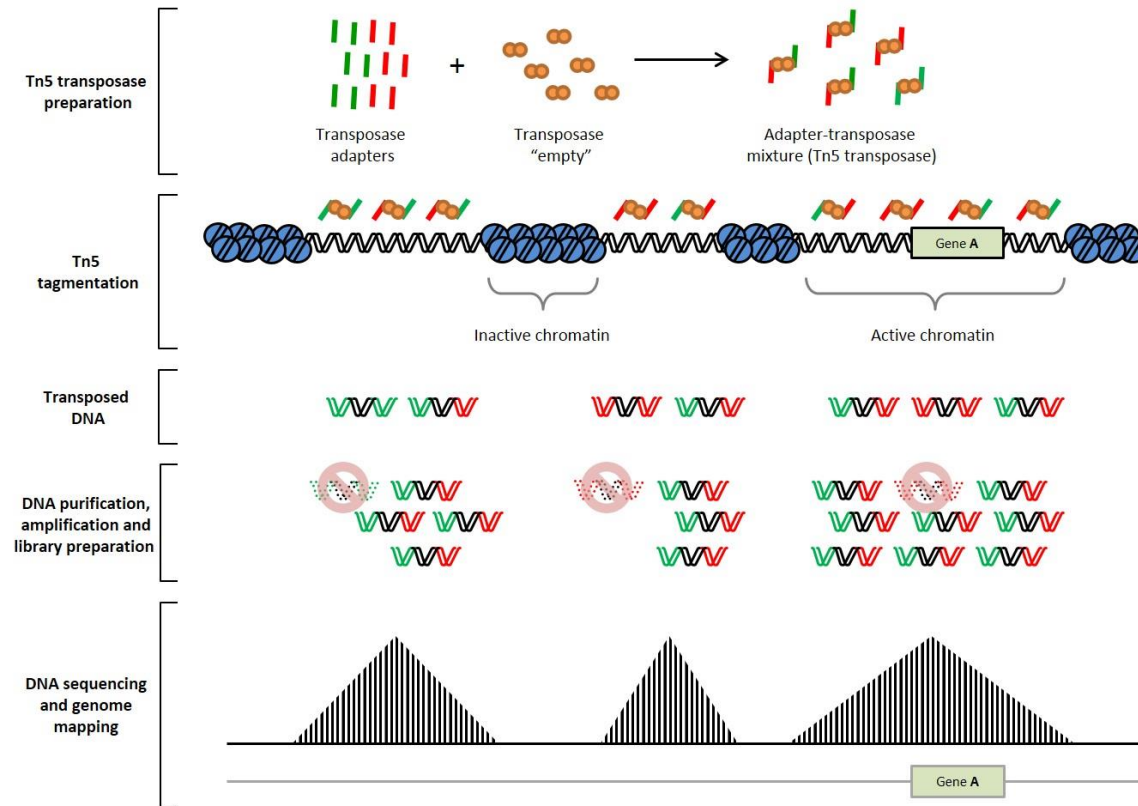
v) NGS metody analýzy přístupnosti chromatinu



DNA-seq knihovny – časté aplikace

ATAC-seq

-identifies accessible DNA regions by probing open chromatin with hyperactive mutant Tn5 Transposase that inserts sequencing adapters into open regions of the genome

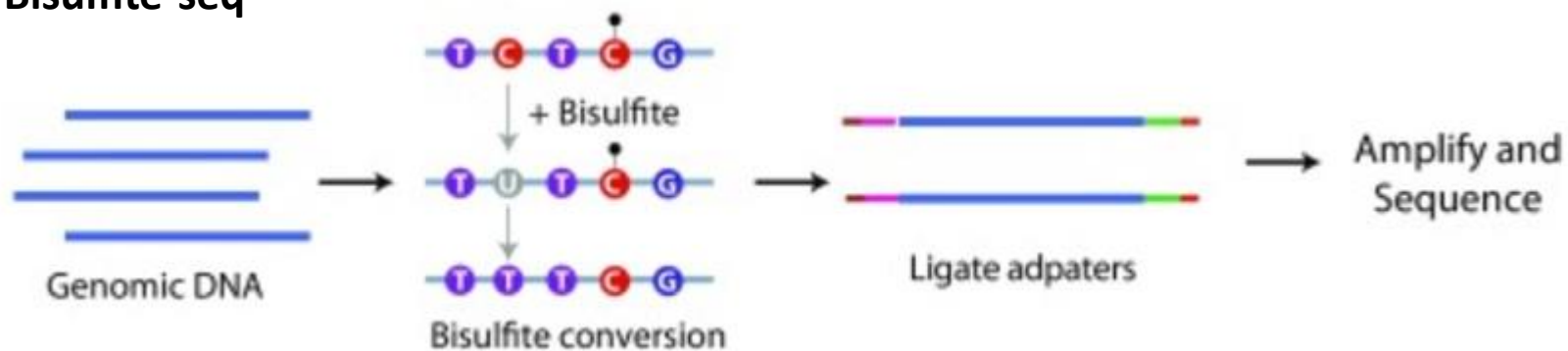


<https://eciofishr.wordpress.com/2019/04/22/technical-section-atac-seq/>

DNA-seq knihovny – časté aplikace

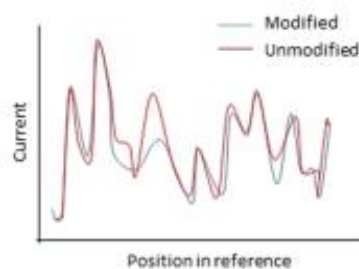
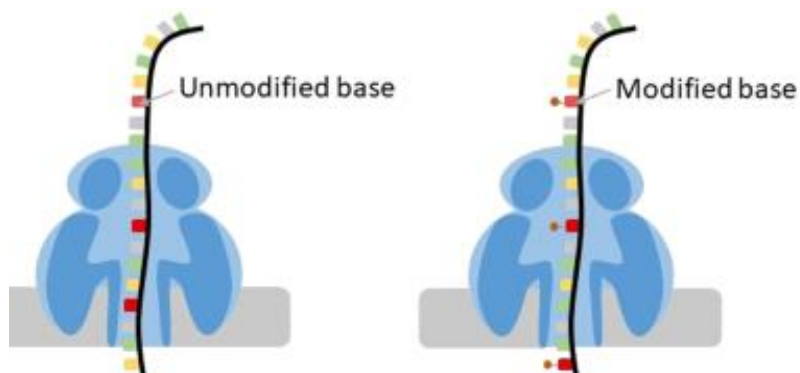
vi) DNA-seq metody analýzy chemických modifikací DNA

Bisulfite-seq

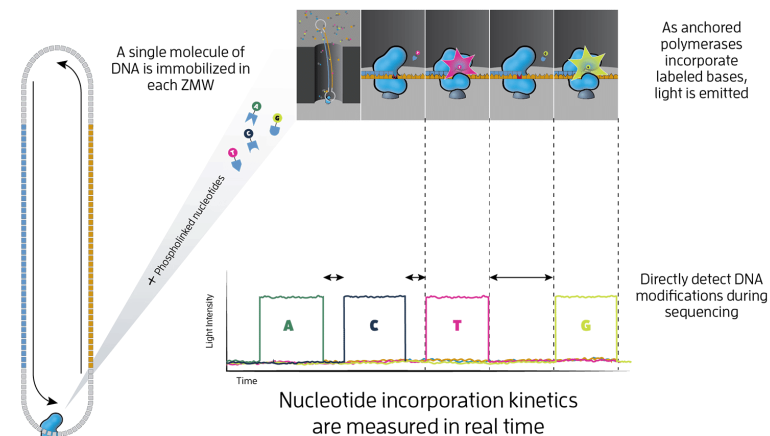


<https://www.creative-biolabs.com/suprecision/whole-genome-bisulfite-sequencing-wgbs-service.htm>

...alternativně pomocí NanoPore či PacBio.



<https://doi.org/10.1038/s10038-019-0679-0>



<https://www.pacb.com/japan/whole-genome-sequencing-jp/>

- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) Příprava NGS knihoven z RNA**

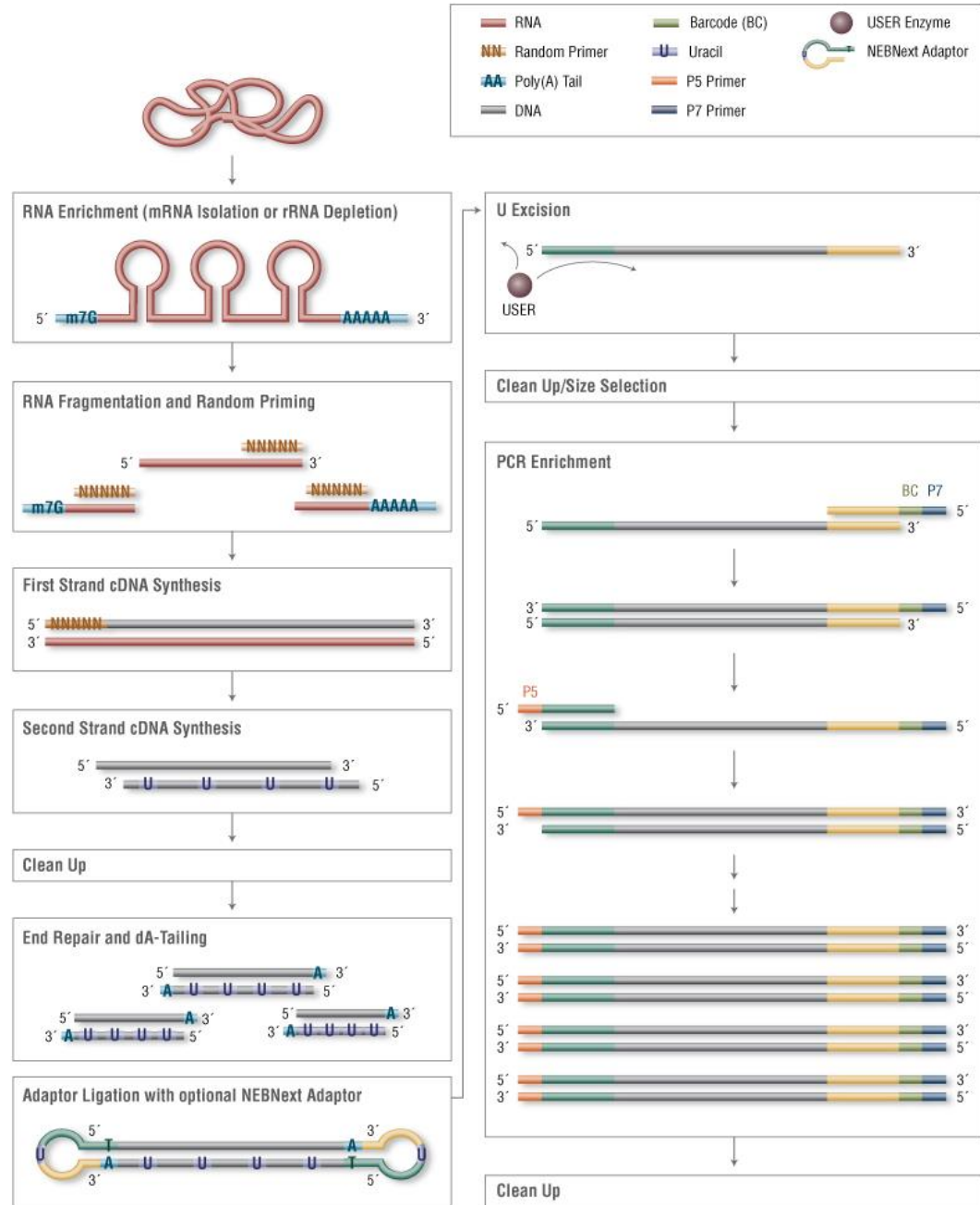
1. Kontrola kvality RNA

2. RNA enrichment

3. Fragmentace a 1st strand syntéza

4. 2nd strand syntéza, stranded/unstranded RNA seq library

...dále postup analogický jako u DNA knihoven

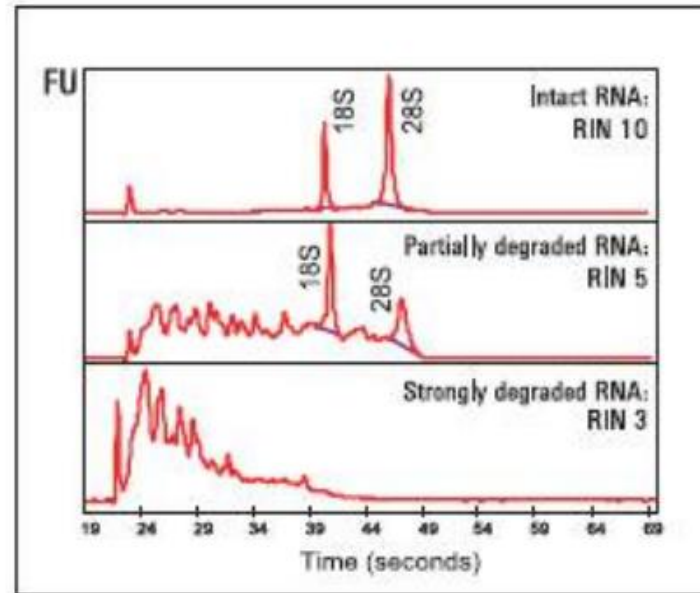
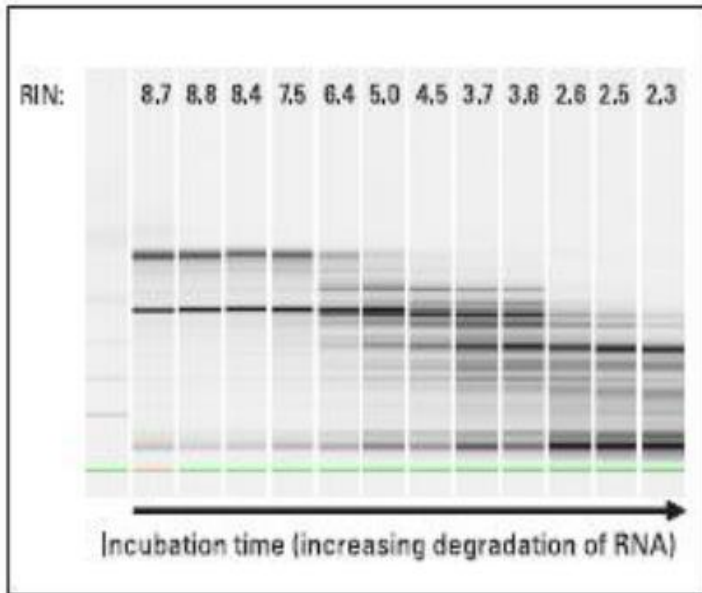


- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) **Příprava NGS knihoven z RNA**

1. Kontrola kvality RNA

Integrita RNA – elektroforeticky (např. na přístrojích TapeStation, Bioanalyzer), kvalita vstupní RNA je kvantifikována pomocí čísla **RIN (RNA integrity number)**

Koncentrace RNA – nejpřesněji pomocí fluorometru (např. Qubit) lze také měřit na TapeStation, Bioanalyzer, Nanodrop



RIN

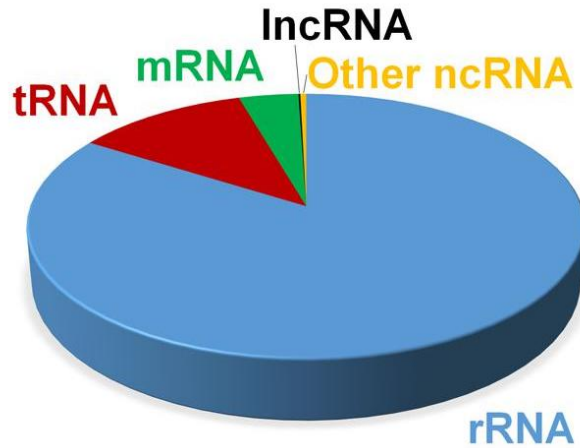
Algoritmus přiřazující k elektroforeogramu celkové RNA číslo od 1 (nejvíce degradovaná) do 10 (nejméně degradovaná). Technicky se vyhodnocuje množství signálu pod silnými peaky rRNA (18S, 28S) s vůči ostatnímu signálu pod grafem.

Pozn. V případě savčích RNA se jedná o velmi reprodučibilní postup vyhodnocení kvality RNA. V případě rostlinných a řady jiných RNA může být RIN nespolehlivý (jiné délky rRNA, interferující chloroplastové rRNA apod..).

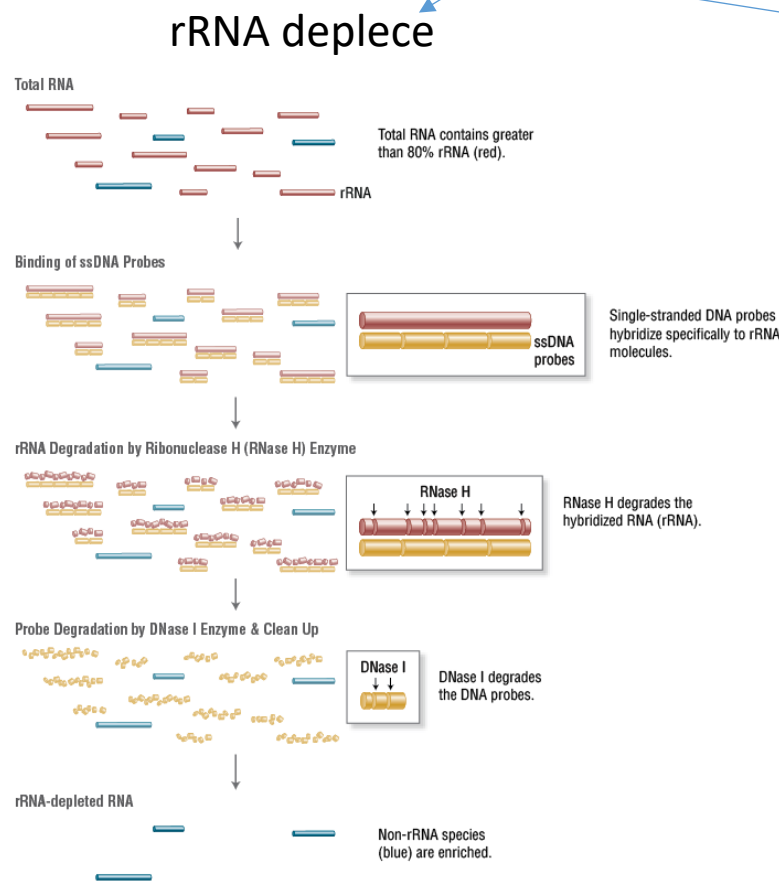
- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) Příprava NGS knihoven z RNA

2. RNA enrichment

- **rRNA** představuje více jak **80%** veškerých transkriptů...
tzn. musíme se jí **zbavit** jinak nebudeme sekvenovat nic jiného než rRNA.

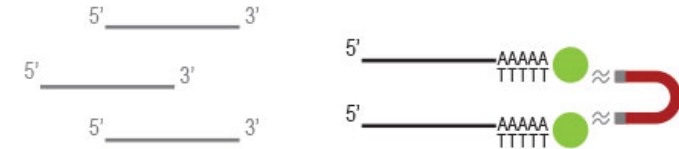


Example transcriptome



<https://www.neb.ca/e7850>

Poly-A enrichment



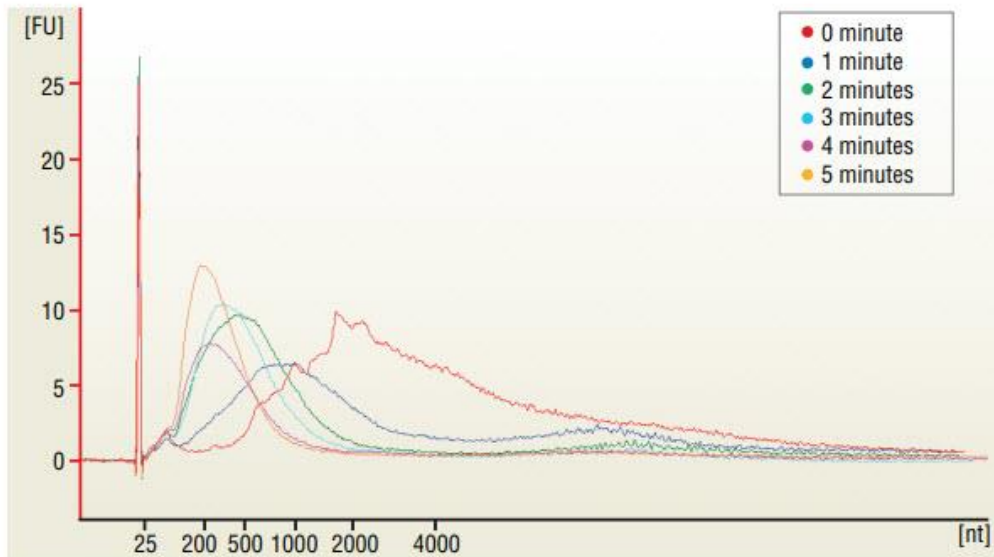
Na rozdíl od **mRNA** (RNAPII transkripty), **rRNA** (RNAPI/RNAPIII transkripty) netvoří polyA na 3' konci.

- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) Příprava NGS knihoven z RNA**

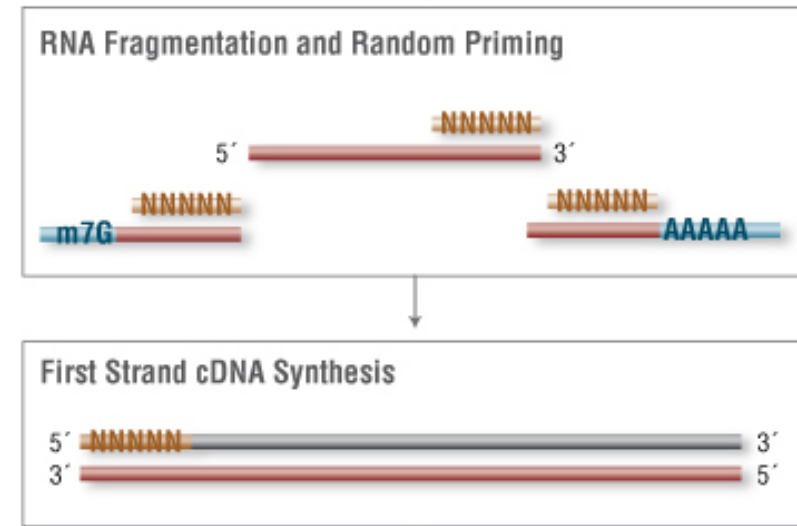
3. Fragmentace a 1st strand syntéza

Fragmentace – nejčastěji teplem (94°C) z přítomnosti Mg⁺ iontů. Délka fragmentů je kontrolována délkou inkubace.

Příklad - mRNA a délka fragmentace



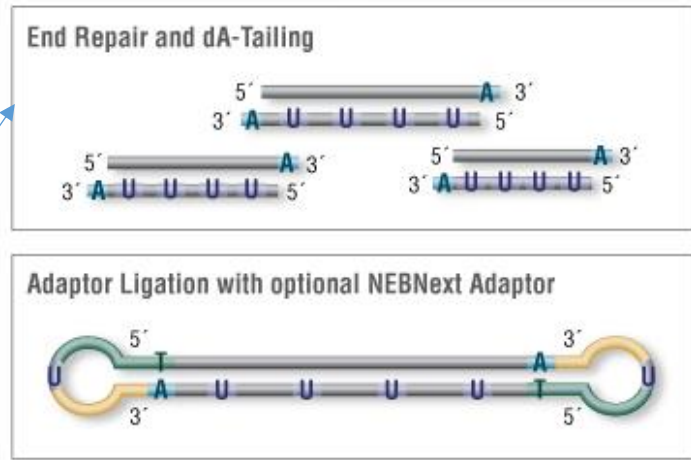
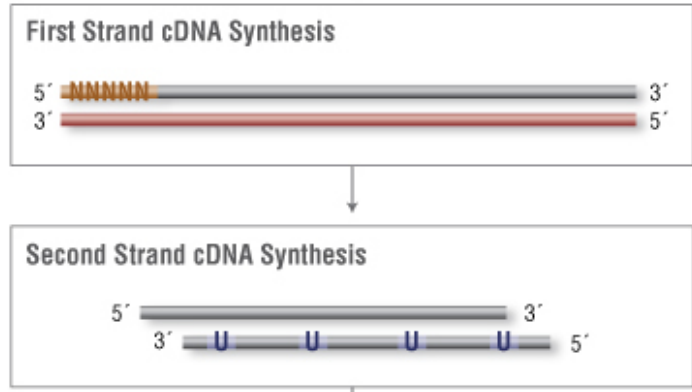
1st strand syntéza (syntéza komplementárního DNA vlákna k RNA pomocí reverzní transkriptázy a random primerů nasedajících na RNA)



- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) **Příprava NGS knihoven z RNA**

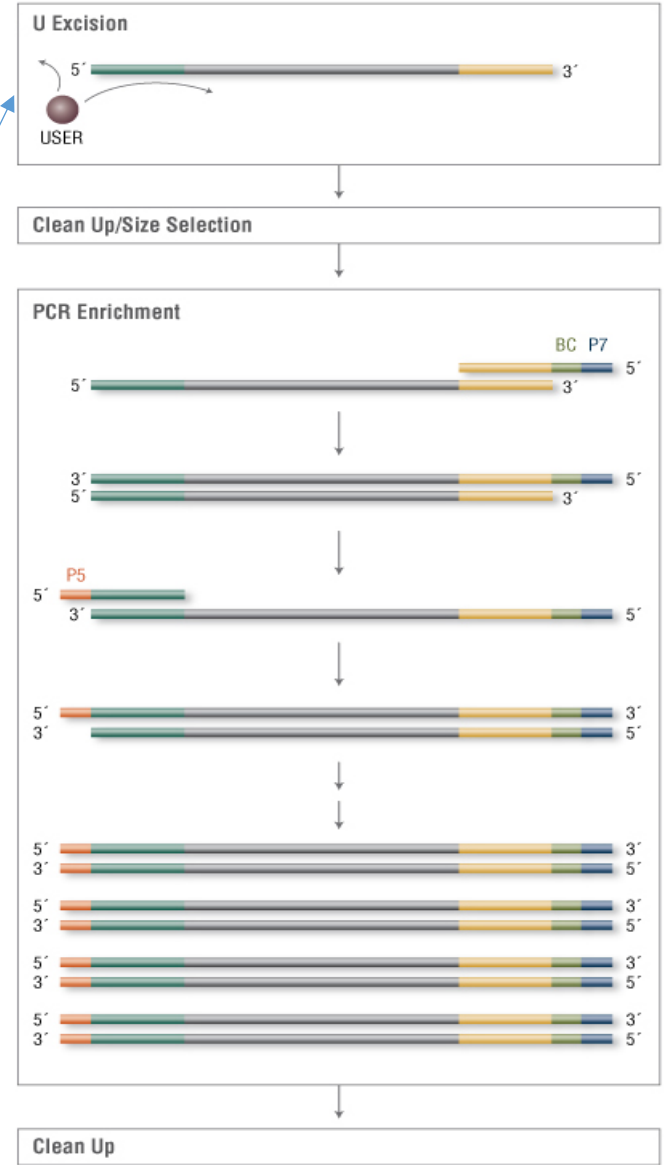
4. 2nd strand syntéza, stranded/unstranded RNA seq library

= Nahrazení původního RNA vlákna vláknem DNA



USER (Uracil-Specific Excision Reagent) degrades vlákno obsahující U.

Zbude pouze ssDNA opatřená adaptéry, který má antisense orientaci k původní RNA.



U – Uracil

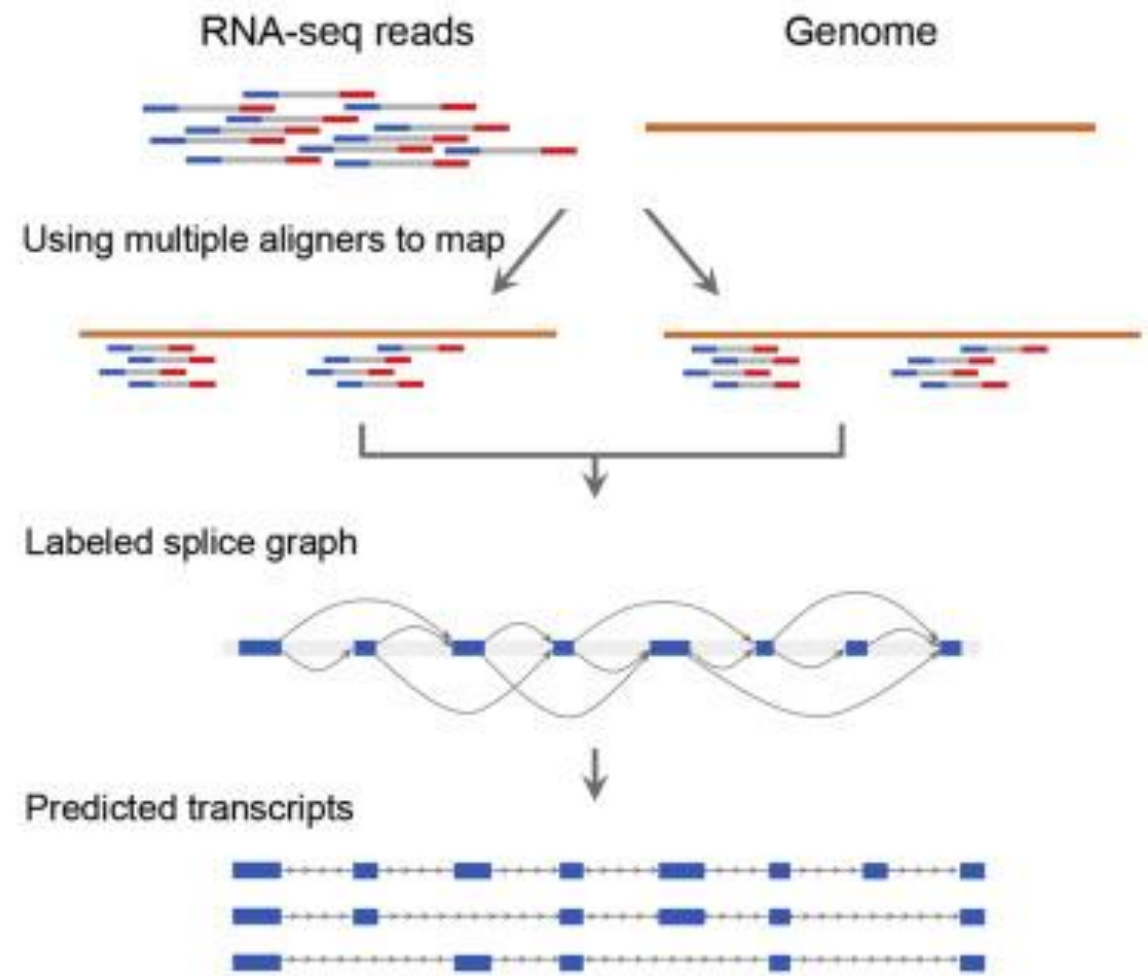
V případě tzv. **Directional RNA knihoven** je nahrazení T za U v second-strandu DNA důležité pro odlišení řetězců DNA – Tj. který řetězec má **stejnou/opačnou** orientaci jako původní RNA transkript

RNA-seq knihovny – časté aplikace

- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) Příprava NGS knihoven z RNA**

i) Transcriptome assembly

- Lze jak *de novo* tak mapováním na referenční genom

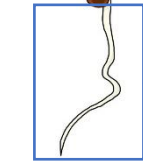
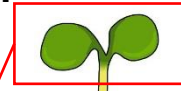


- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) Příprava NGS knihoven z RNA**

ii) Differential gene expression

RNA-seq knihovny – časté aplikace

RNA isolation:



Sample 1



Sample 2



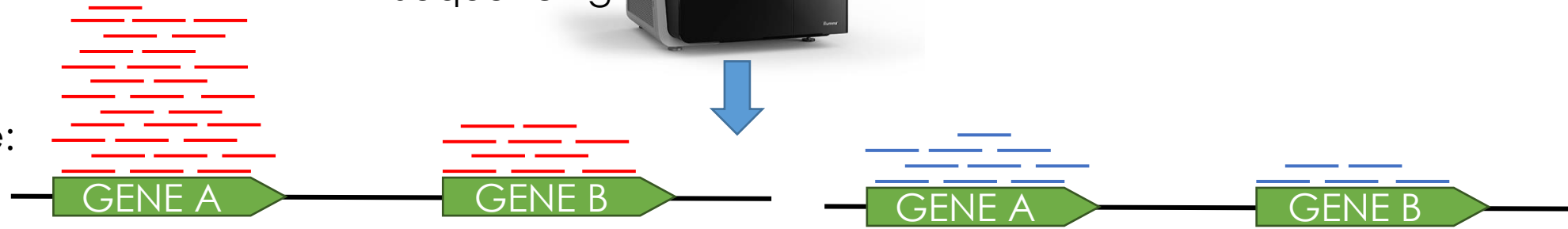
RNA-seq library preparation:



Sequencing



Read mapping to reference genome/transcriptome:



Normalizace a kvantifikace genové exprese:

-Normalizace umožňuje porovnávat vzorky mezi sebou.

Metody normalizace:

TPM (Transcripts Per Million) - normalizuje počet read namapovaných na daný transkript podle délky transkriptu a celkového počtu namapovaných readů ve vzorku

dále: FPKM (Fragments Per Kilobase Million), RPKM (Reads Per Kilobase Million),...

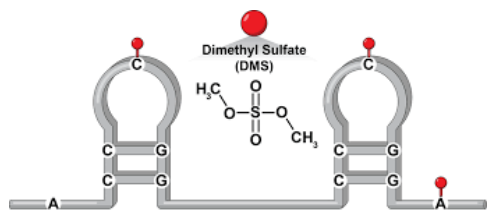
RNA-seq knihovny – časté aplikace

a) Příprava NGS knihoven z DNA

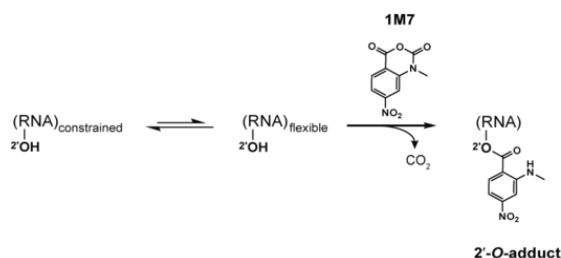
b) Příprava TGS knihoven z DNA

c) Příprava NGS knihoven z RNA

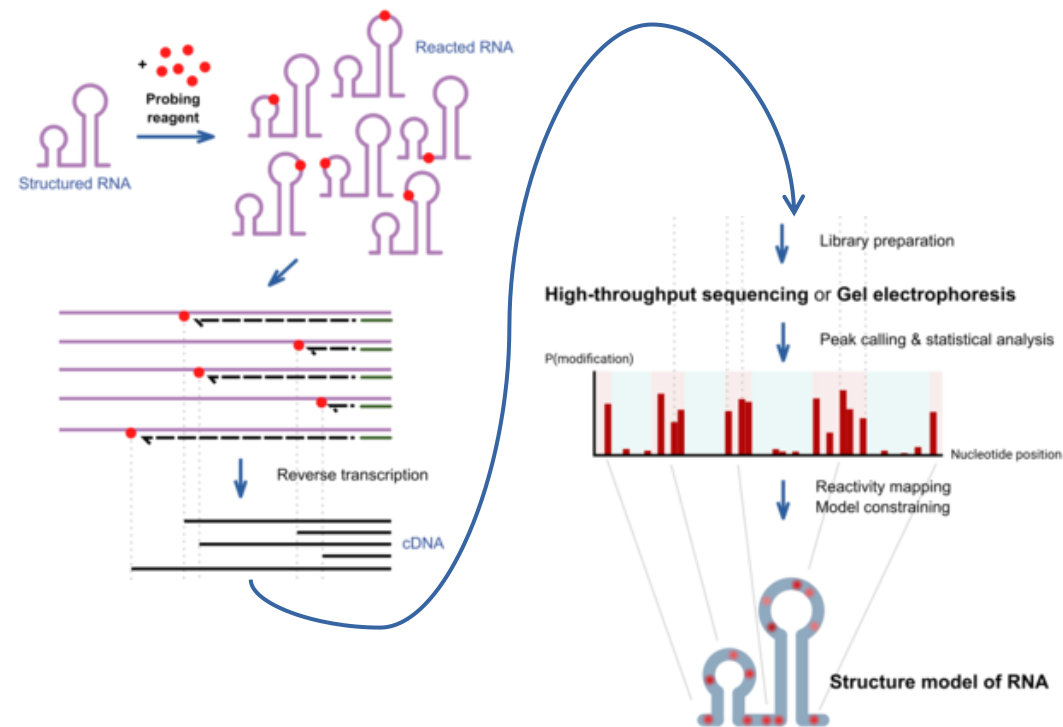
iii) Studium sekundární struktury RNA (DMS-MaPseq, SHAPE-seq)



DMS (Dimethyl sulfate) – methylates A and C ribonucleotides at the site of natural hydrogen bonds upon base-pairing, therefore modification occurs only at single-stranded nucleotides

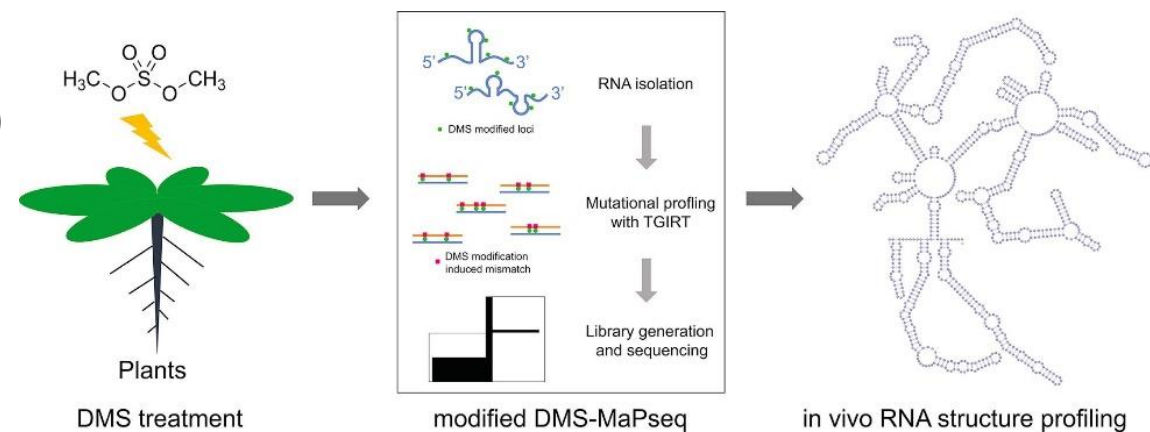


SHAPE - modify the backbone of RNA in structurally flexible regions



- SHAPE/DMS modification results in fall-off reverse transcriptase during reverse transcription – i.e. abundance of premature terminated reads indicates modified (single-stranded) positions in such RNA-seq data (obr.1)

- Thermostable group II intron reverse transcriptase (TGIRT) can go through modified bases, but generates errors in modified positions – i.e. SNPs in such RNA-seq data indicates single-stranded bases. (obr.2)

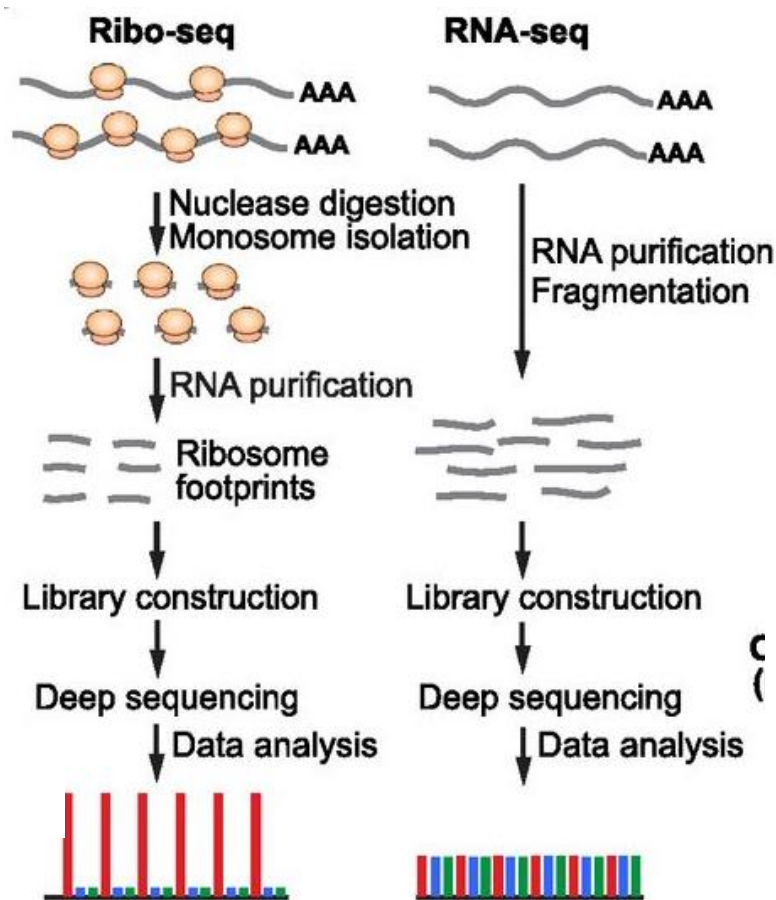


RNA-seq knihovny – časté aplikace

- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) Příprava NGS knihoven z RNA**

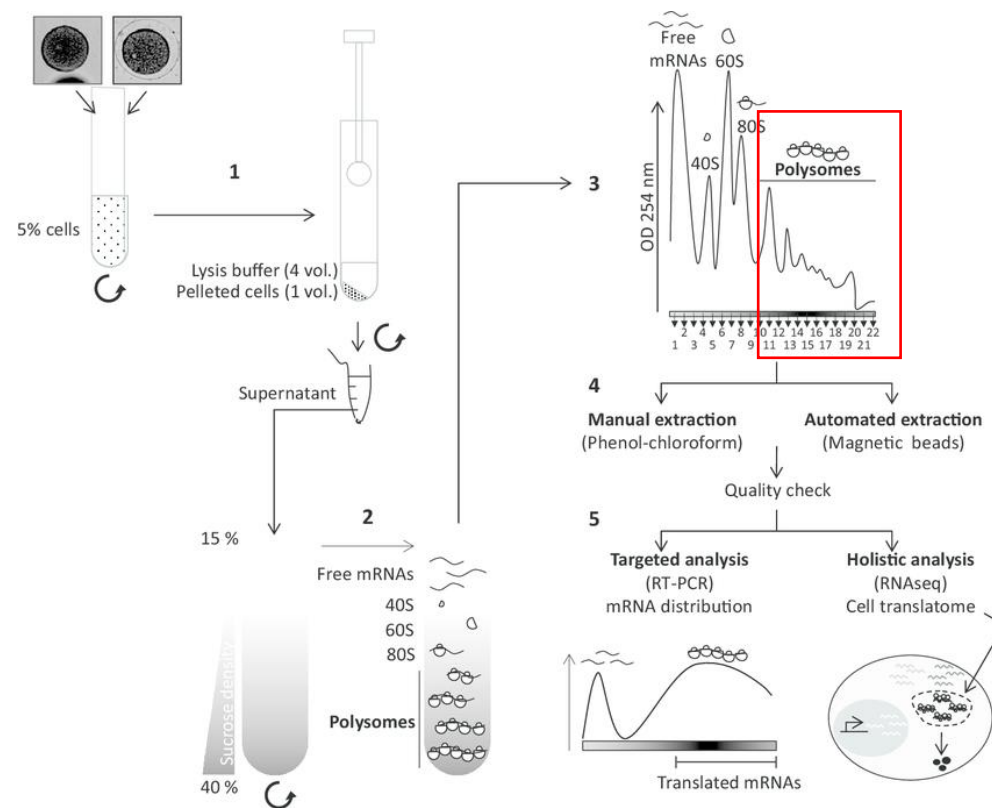
v) analysis of translatoe (actively translated RNAs)

1. Ribosome profiling (Ribo-seq)



RNA regions bounded by ribosomes are protected during Mnase digestion, free RNA is cleaved.

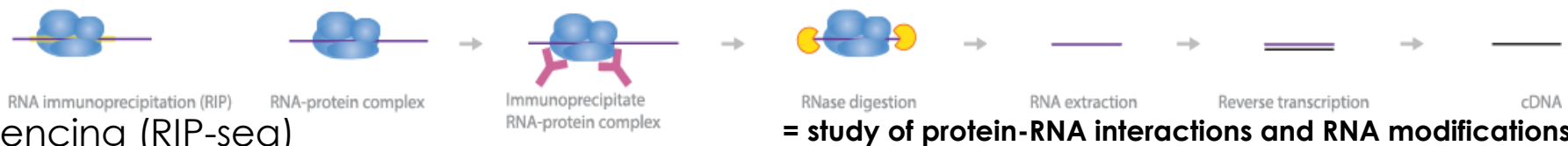
2. Polysome profiling



Compared to Ribosome profiling – total RNA associated with ribosomes (Polysomes) is extracted and sequenced.

RNA-seq knihovny – časté aplikace

- a) Příprava NGS knihoven z DNA
- b) Příprava TGS knihoven z DNA
- c) Příprava NGS knihoven z RNA**



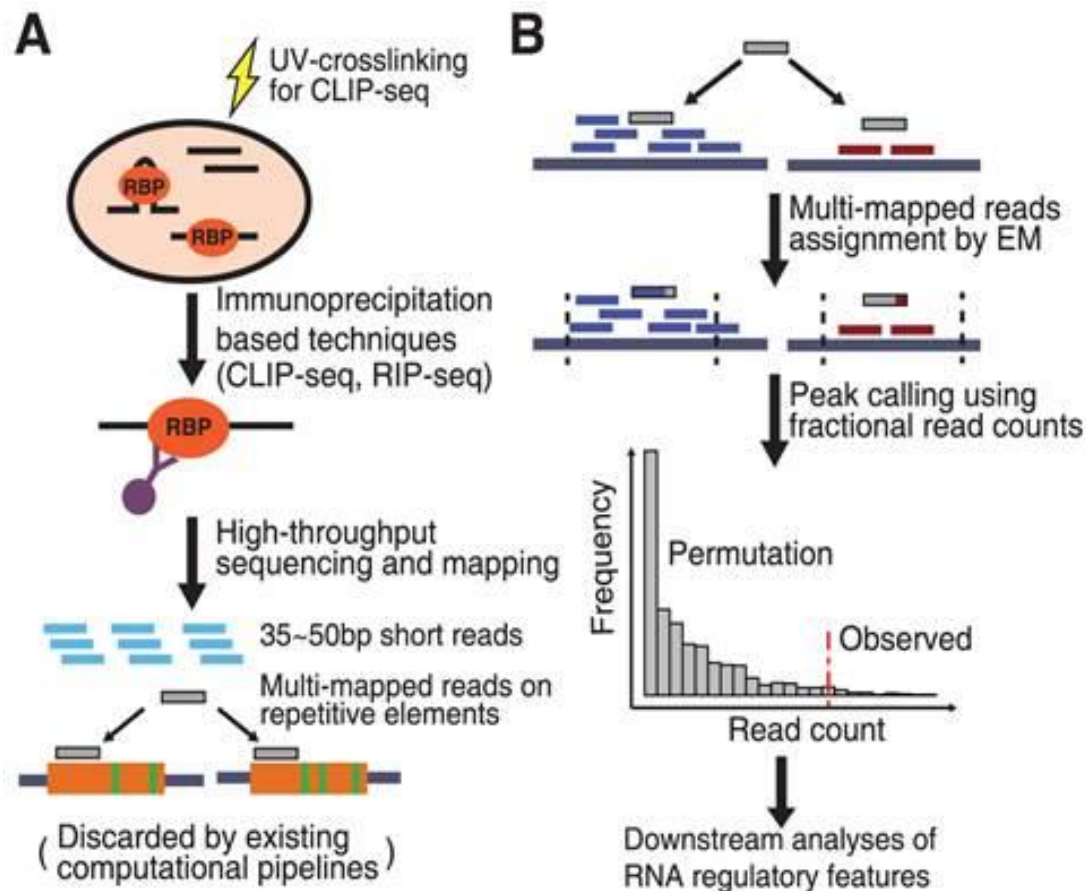
vi) RNA immunoprecipitation sequencing (RIP-seq)

CLIP-seq

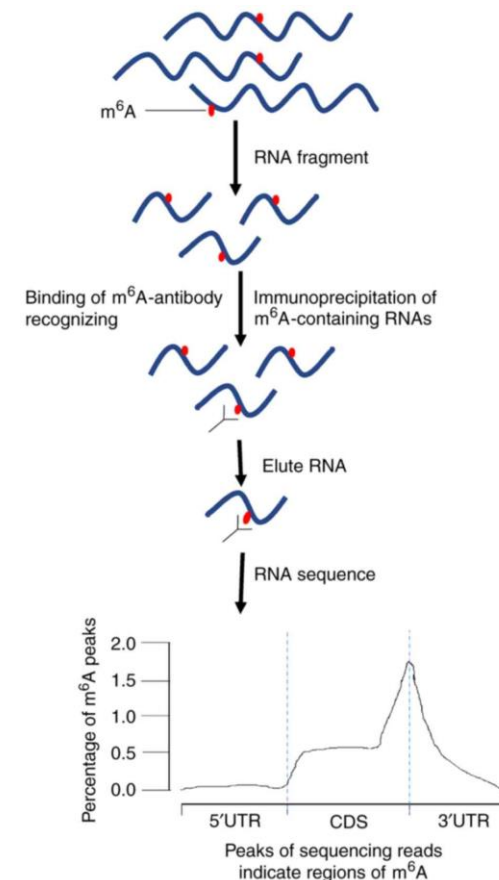
includes Cross-linking – covalent joining of RNA Binding Protein (RBP) and RNA – and subsequent immunoprecipitation with specific antibody → RNA-seq

RIP-seq

Immunoprecipitation of RBP and associated RNA(s) (without cross-linking) → RNA-seq



<https://www.semanticscholar.org/paper/CLIP-seq-analysis-of-multi-mapped-reads-discovers-Zhang-Xing/b403d8737a9f56397391232d083a351860a8215b>



<https://www.spandidos-publications.com/10.3892/ijmm.2019.4169>

Databáze s dostupnými DNA/RNA-seq daty



<https://www.ncbi.nlm.nih.gov/sra/?term=>