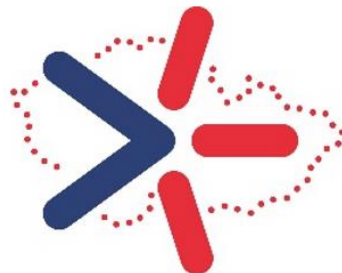




Financováno
Evropskou unií
NextGenerationEU



Národní
plán
obnovy

MSMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Analýza epigenetických značek a struktury chromatinu

Jiří Fajkus

Molekulární komplexy chromatinu, Mendelovo centrum genomiky a proteomiky rostlin, CEITEC MU

Národní centrum pro výzkum biomolekul, Lab. FGP, PřF MU

fajkus@sci.muni.cz

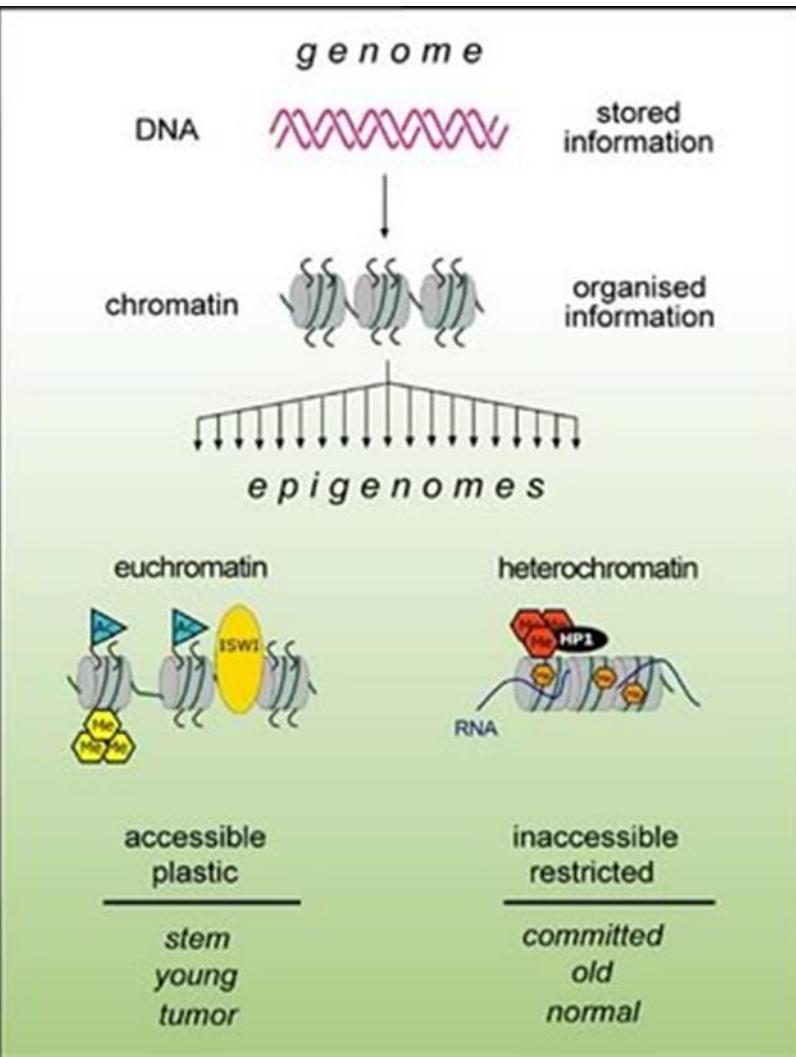


MUNI
SCI

Co je epigenetika a proč by nás měla zajímat (podrobně jste slyšeli v přednášce doc. Fojtové)

Mnoho definic, od Aristotela po současnost, dobrá je např. tato: Arthur D. Riggs et al., 1996:

Studium mitoticky nebo meioticky děditelných změn genových funkcí, které nemohou být vysvětleny změnami v sekvenci DNA.



Během vývoje mnohobuněčného organismu dochází k divergenci fenotypů u diferencujících se buněk. Jakmile jsou tyto odlišné/specifické rozdíly „v naprogramování“ buněk nastaveny, mohou být klonálně přenášeny do dceřiných buněk – jako **epigenetická informace**.

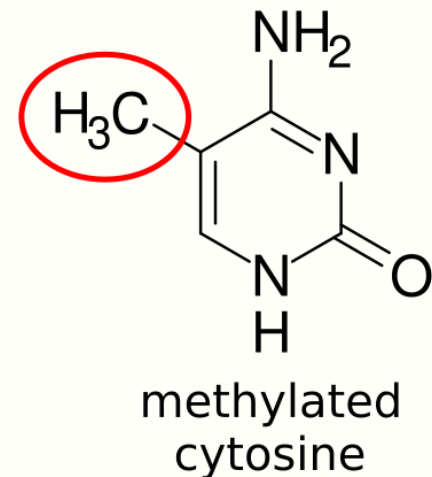
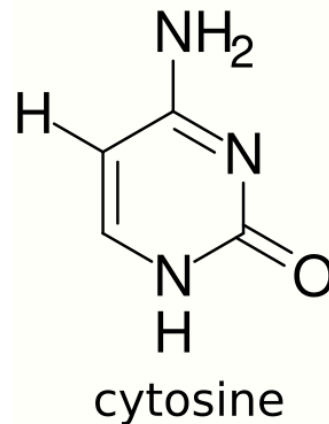
Stejná genetická výbava (genotyp) může být zdrojem různého fenotypu



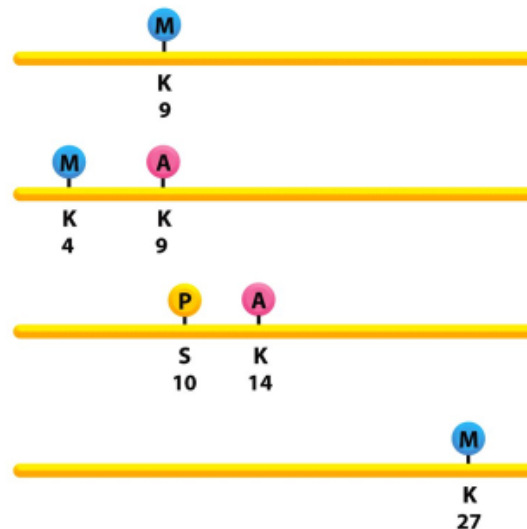
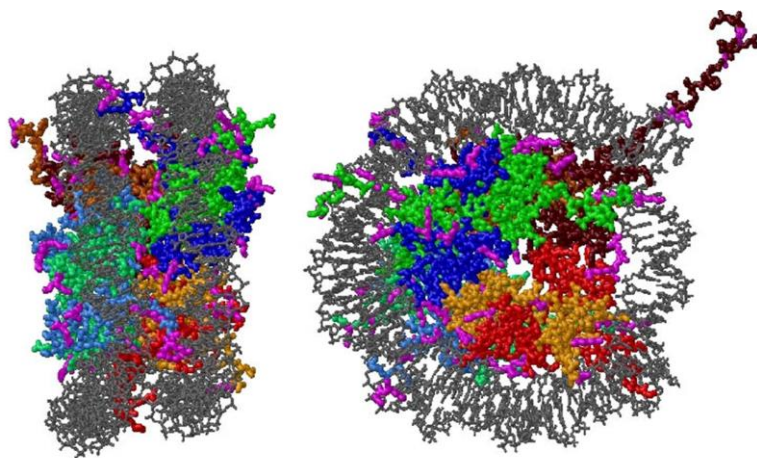
Známé epigenetické mechanismy zahrnují:

1. Methylaci DNA

U eukaryot dochází k methylaci cytosinu v 5' poloze – působením DNA methyltransferáz. Methylace promotorů vede k umlčení genů



2. Posttranslační modifikace histonů v nukleozómech (methylace, acetylace, fosforylace, ubiquitinace), náhrada histonů jejich specifickými variantami (např. CENH3)



Význam

Heterochromatin,
umlčení genů

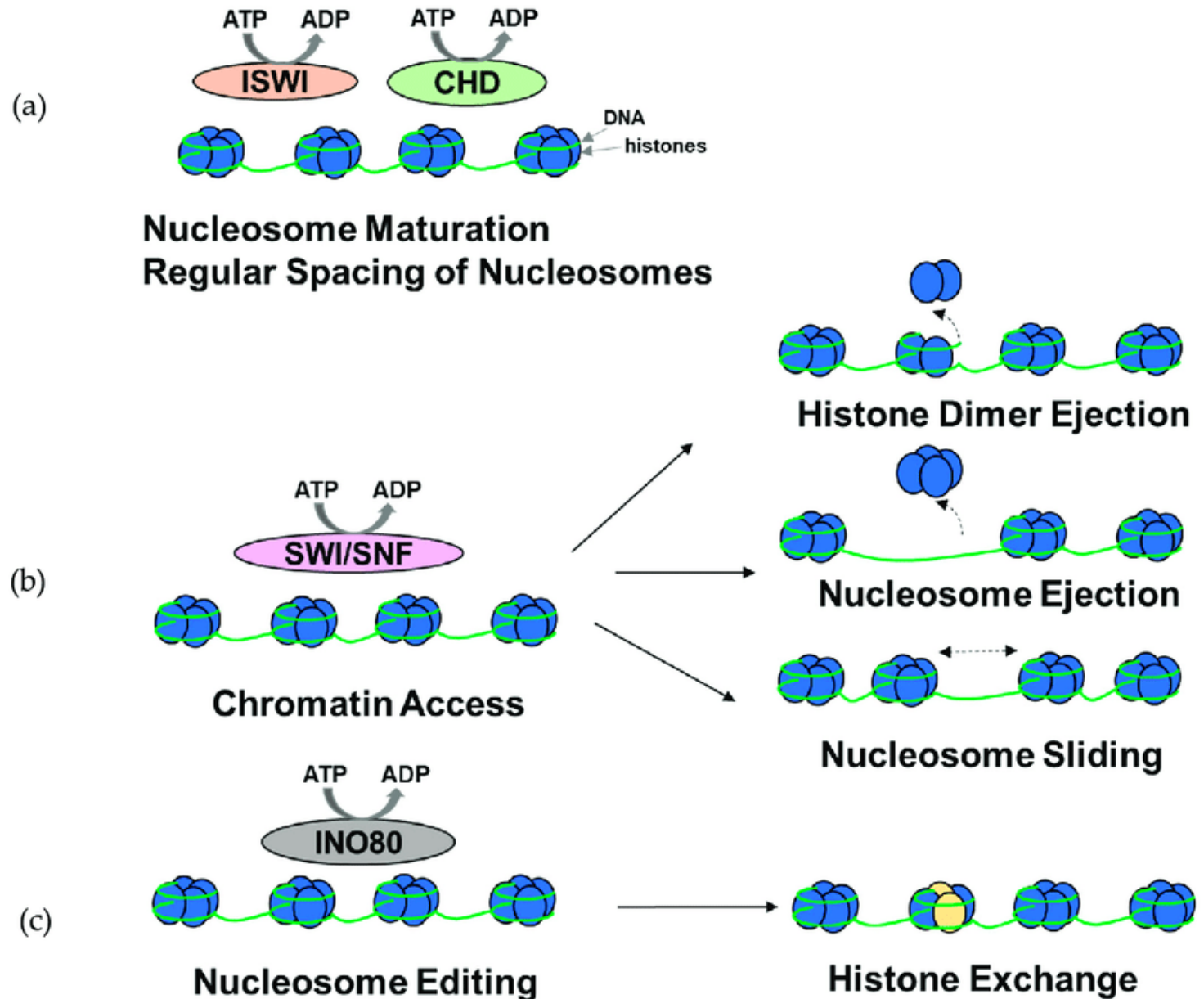
Expres genů

Expres genů

Umlčování Hox genů,
inaktivace X-chr

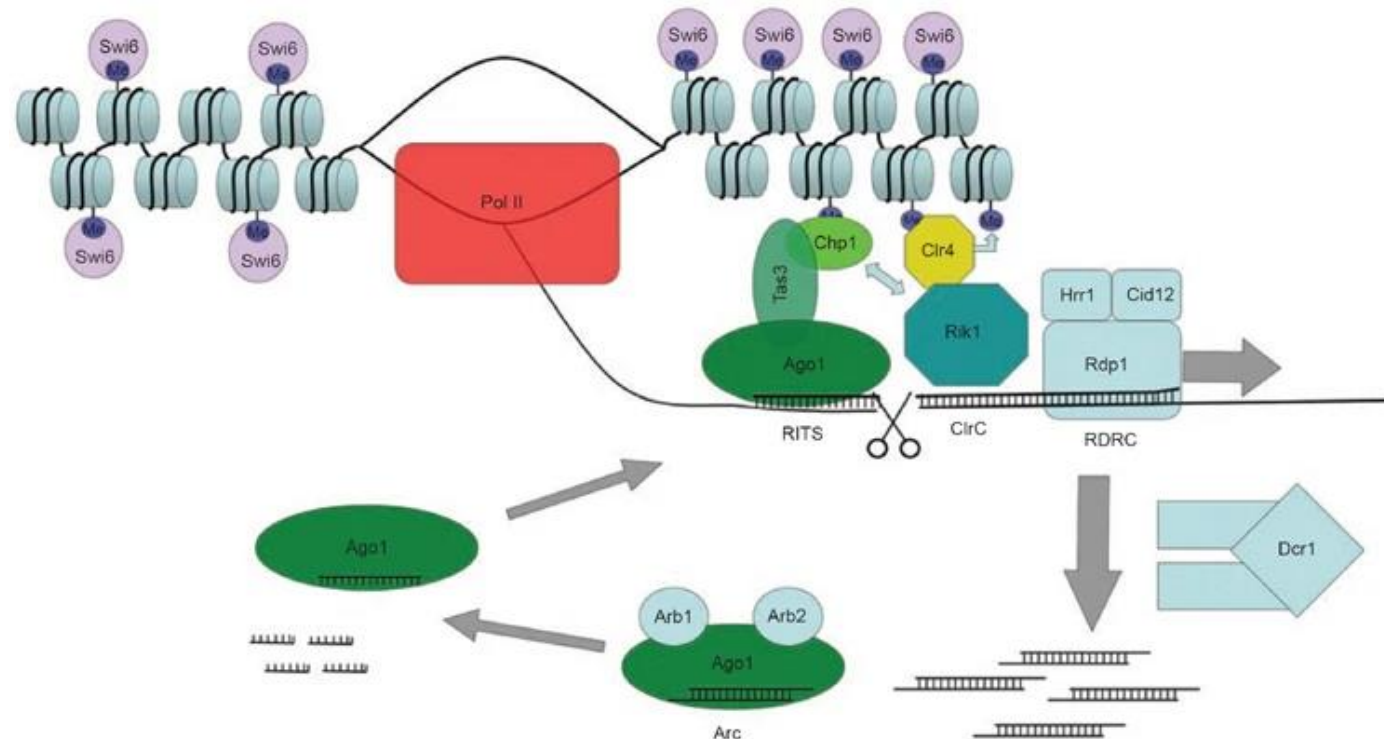
Další epigenetické mechanismy/ procesy

Remodelace chromatinu



Další epigenetické mechanismy/ procesy

RNA interference, TGS, PTGS

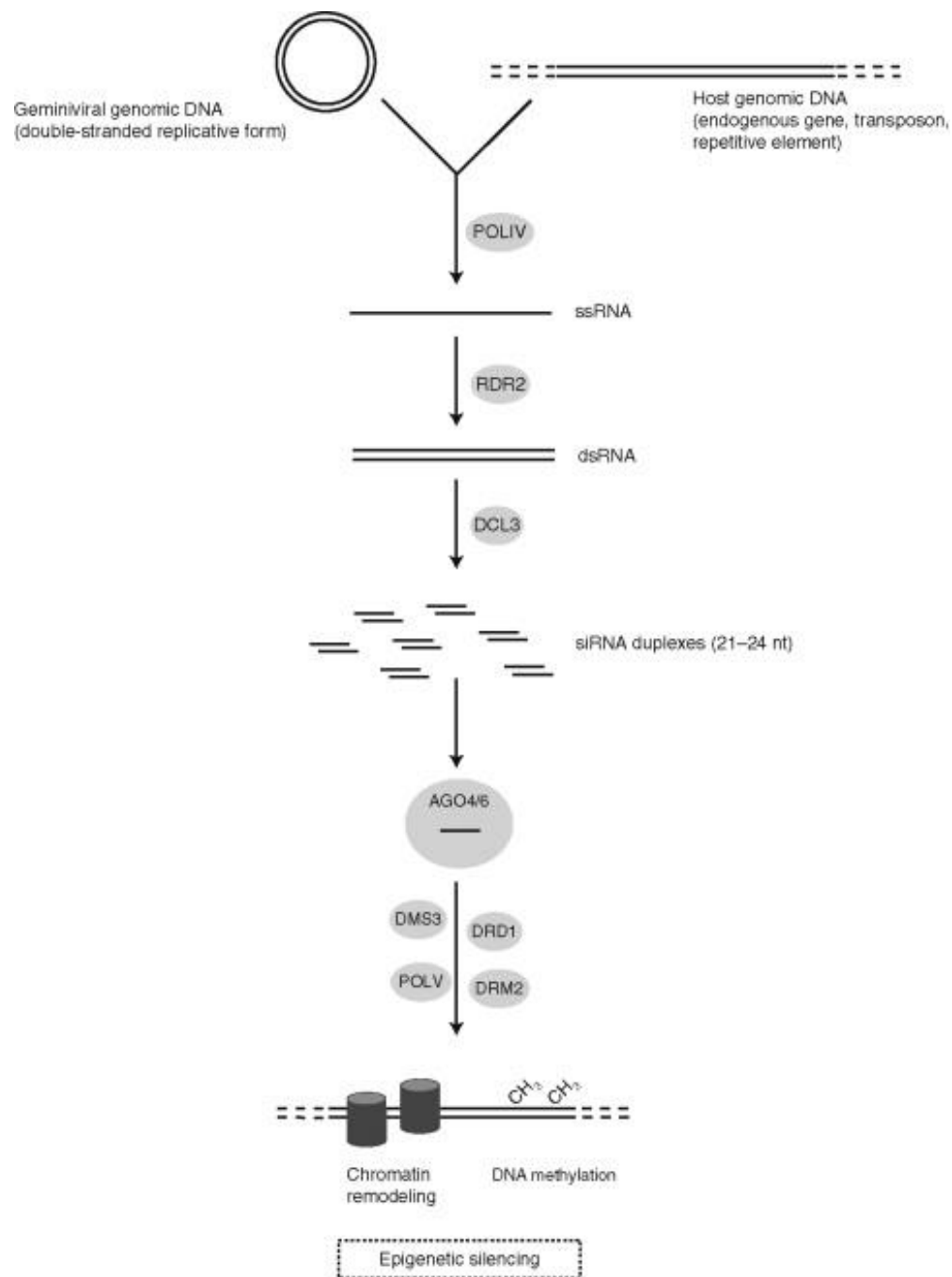


RNAi a heterochromatinová zpětnovazební smyčka *S. pombe*. Pol II transkripce centromerických repetitiv vede k zachycení RITS prostřednictvím komplementární siRNA a rozpoznání H3K9me. Ago1 štěpí vznikající RNA a rekrutuje ClrC a RDR. Clr4 z ClrC se váže a vytváří H3K9me (modré kroužky histonů), Rik1 interaguje s Chp1 a potenciálně s nascentní RNA. RDR syntetizuje dsRNA templátovou nascentní RNA, produkující buď transkripty plné délky, které jsou substráty Dcr1, nebo sekundární siRNA, která může být vázána přímo do RITS. Primární duplexní siRNA generované Dcr1 jsou inkorporovány do komplexu Arc a následně Ago opouští Arc, štěpí siRNA a spolu s Chp1 a Tas3 tvoří další komplex RITS, připravený pro cílení homologní RNA.

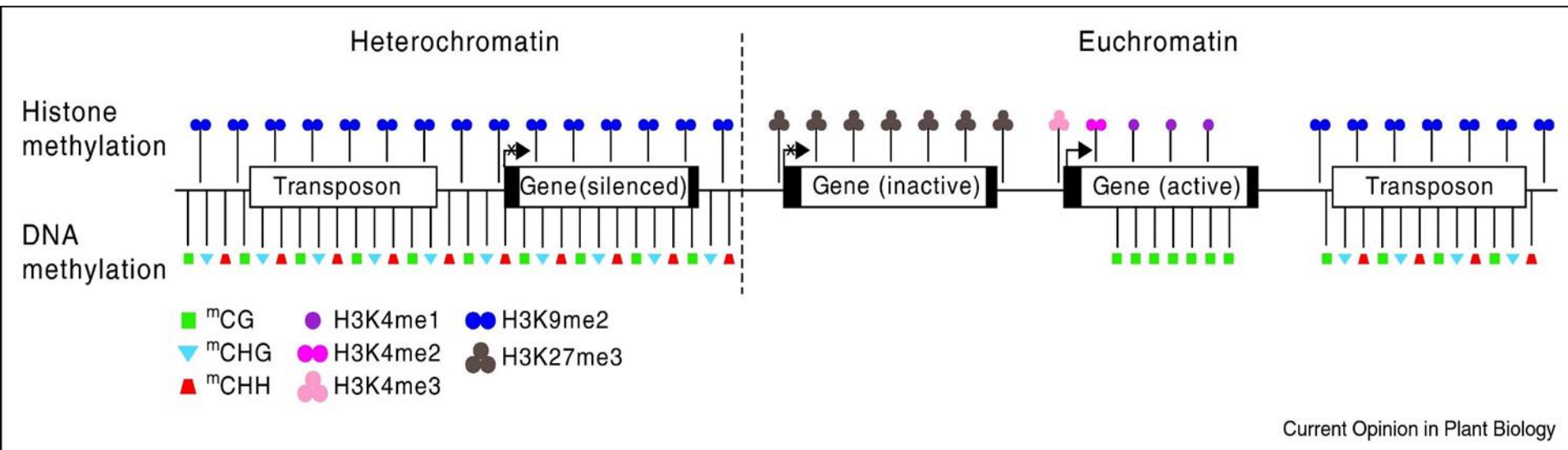
Další epigenetické mechanismy/ procesy

RNA Directed DNA Methylation (RdDM)

RNA funguje jako primární signál pro
umlčování prostřednictvím methylace
DNA.



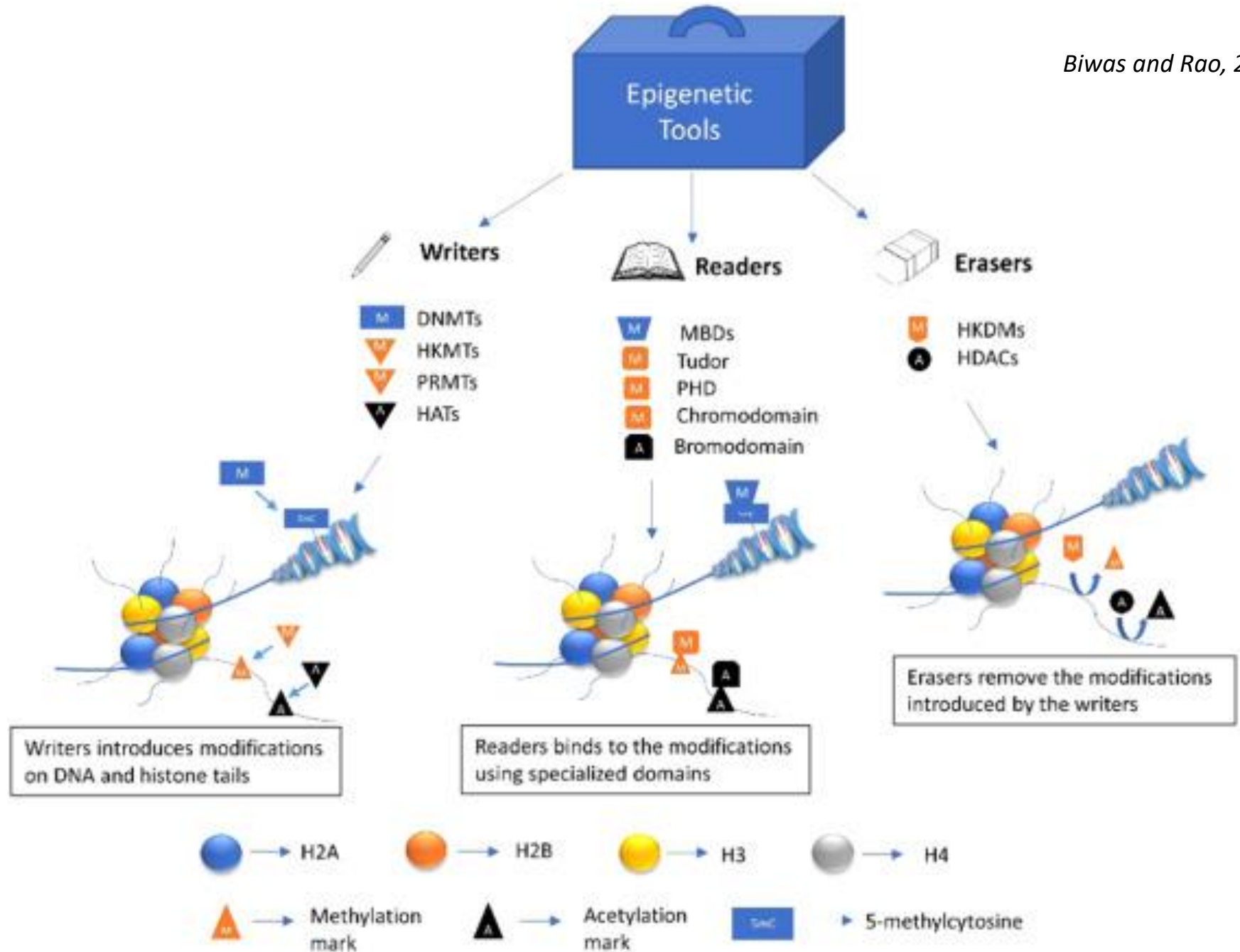
Příklad: uspořádání epigenetických značek v rostlině *Arabidopsis thaliana*



Genom *A. thaliana* má heterochromatickou část v oblastech blízko centromerám, zatímco chromozomová ramena jsou euchromatická. Pericentromerický heterochromatin obsahuje hodně transpozonů a umlčených genů a je charakteristický vysokým obsahem H3K9me2 a methylace cytosinu ve všech třech možných kontextech (CG, CHG a CHH). Transpozony v euchromatinu obsahují také H3K9me2 a methycytosiny ve všech třech kontextech, ale tvoří jen krátké úseky limitované délkou transpozonu. Některé neaktivní geny postrádají metylaci DNA, ale jsou značeny H3K27me3. Aktivní geny jsou značeny H3K4me3 a H3K4me2 na promotorech a H3K4me1 uvnitř genu. Geny s nižší úrovní exprese mohou mít metylaci typu CG uvnitř genu.

Současné pojetí epigenetických nástrojů: Writers, Readers, Erasers..... (+ Modulators)

Biwas and Rao, 2018



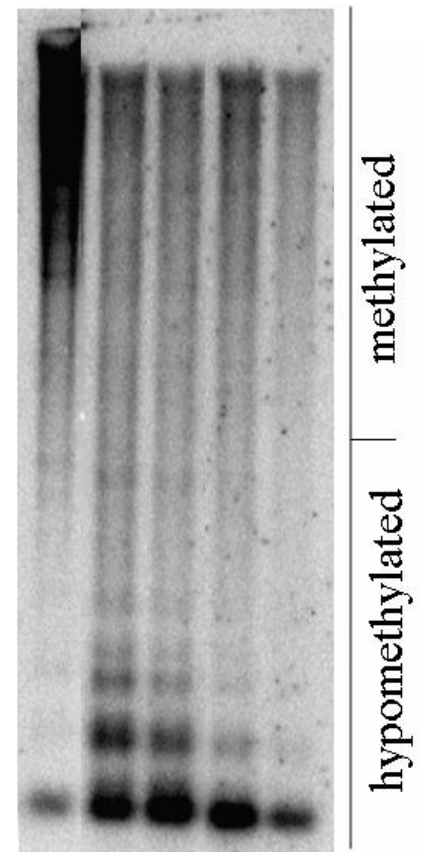
METODY STUDIA METHYLACE DNA

1. Štěpení methylačně citlivými restričními endonukleázami
 - v rozpoznávací sekvenci mají cytosin, neštěpí, pokud je methylován

CG: *Hpa*II mC^mCGG
*Cfo*I G^mCGC
*Sma*I CC^mCGGG
*Tai*I A^mCGT

CNG: *Msp*I $mCCGG$

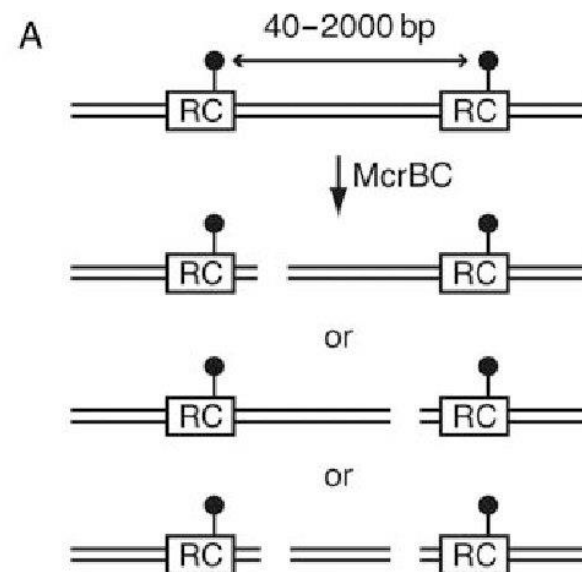
CHH: *Sau*96I $GG(A/T)^mC^mC$
(záleží na tom, co v sekvenci následuje)



Současná **inovace methylačně-specifického štěpení:**

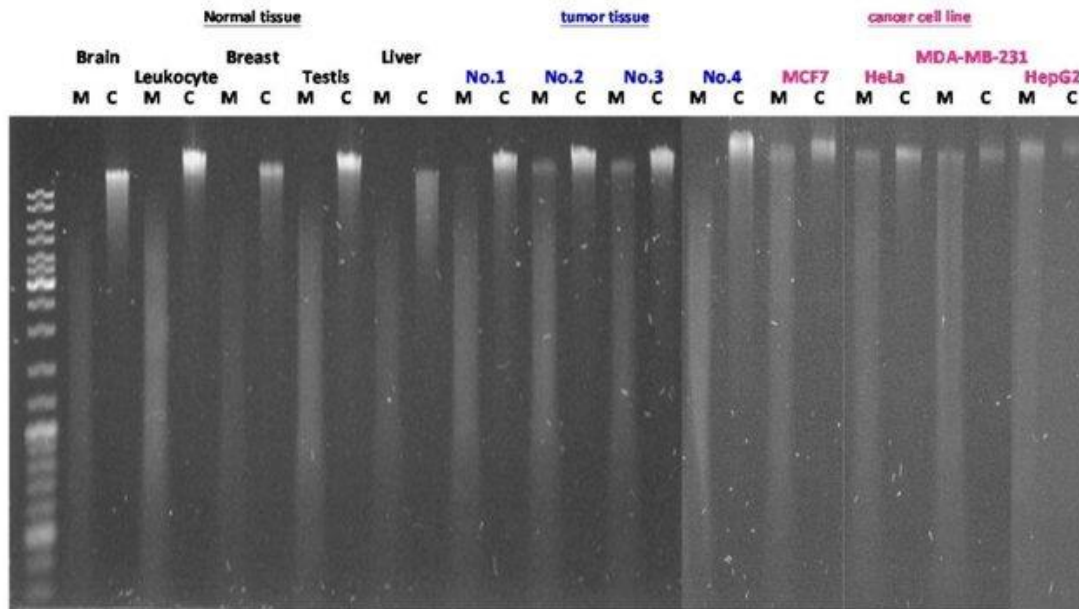
McrBC restriční endonukleáza $5' \dots \text{Pu}^{\text{m}}\text{C} (\text{N}_{40-3000}) \text{Pu}^{\text{m}}\text{C} \dots 3'$

McrBC štěpí DNA obsahující methylcytosin nebo hydroxymethylcytosin v jednom nebo obou řetězcích. Cílová místa na DNA rozpoznávaná McrBC se skládají ze dvou polovin (G/A)mC. Tato místa mohou být od sebe vzdálena až 3 kb, optimálně 55-103 bp.

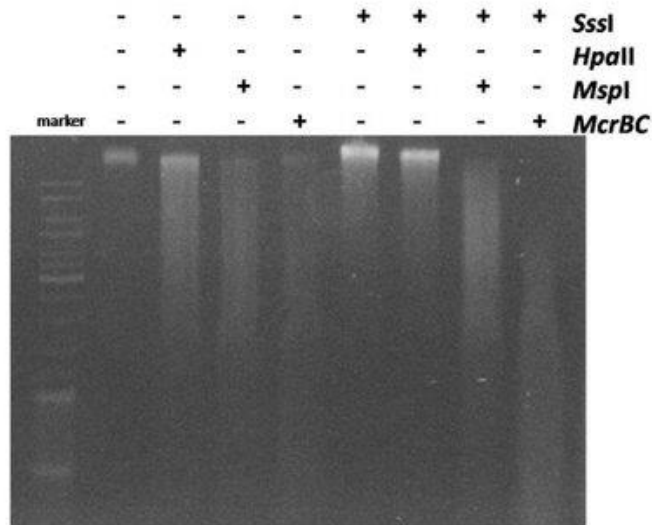


M: McrBC digest; C: control genomic DNA

(a)



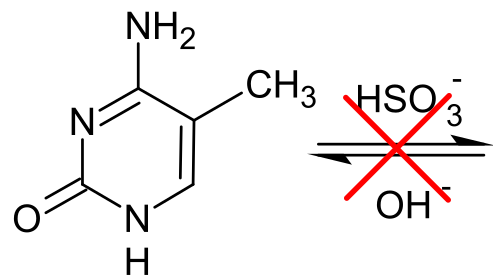
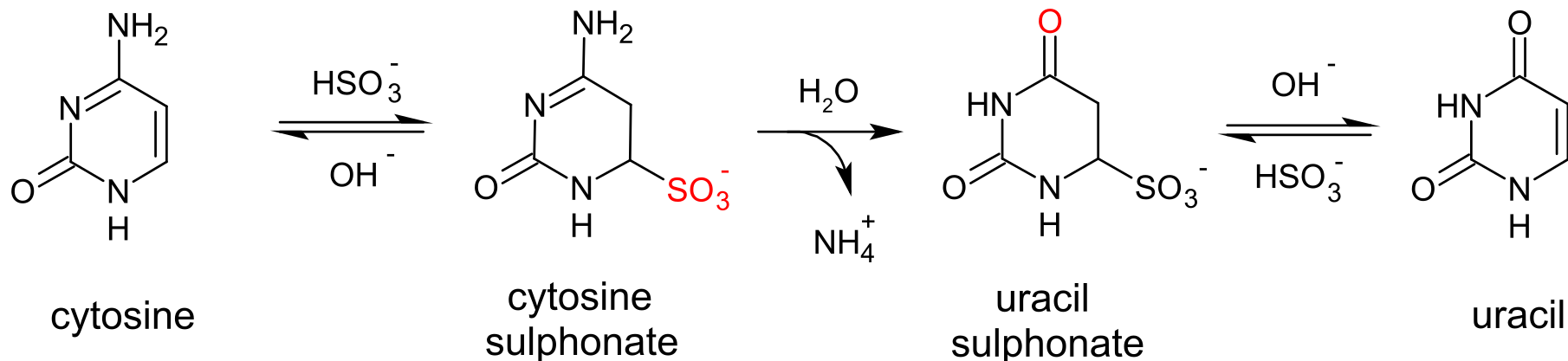
(b)



SssI: CpG methyltransferase
HpaII: methylation sensitive restriction enzyme
MspI: an isoschizomer of *HpaII*

Genomová DNA štěpená McrBC. (a) Štěpení DNA z normální a nádorové tkáně enzymem McrBC. DNA z primárních nádorů prsu a z různých nádorových buněčných linií vykazuje přítomnost vysokomolekulární DNA odolné proti McrBC, zatímco DNA z normálních tkání jsou štěpeny na malé fragmenty. (b) Velké kusy McrBC-rezistentní MCF-7 DNA se staly citlivými na štěpení McrBC po *in vitro* CpG methylaci *SssI* metylázou (srovnej 4. dráhu s 8. dráhou).

Konverze cytosinu na uracil pomocí „bisulfitu“ (hydrogensířičitanu)



2. Modifikace DNA bisulfitem

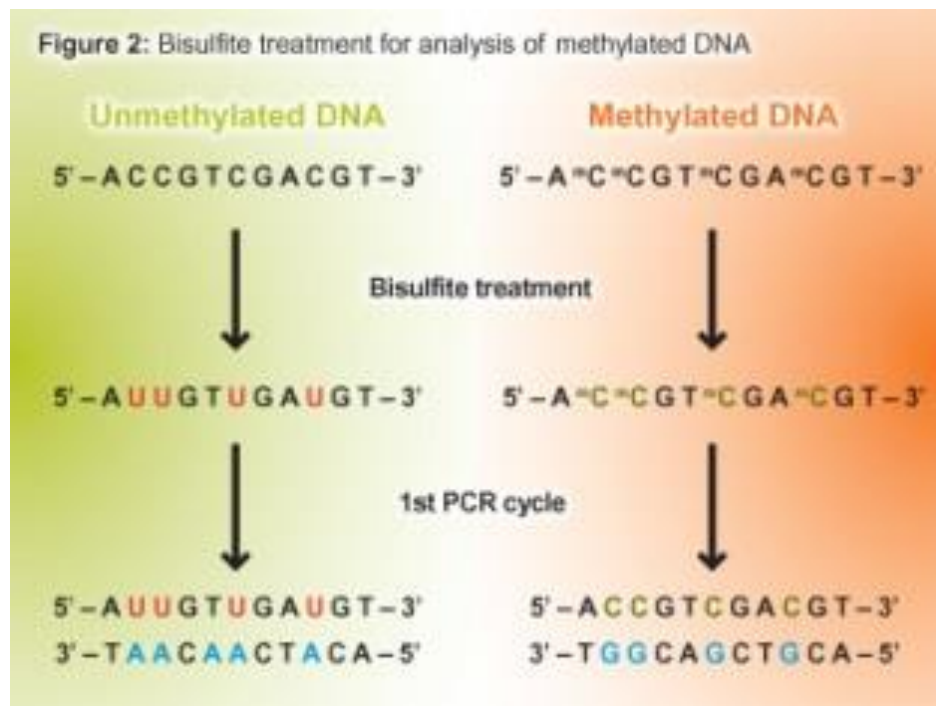
cytosiny kovertovány na uracily,
^mC nereagují.

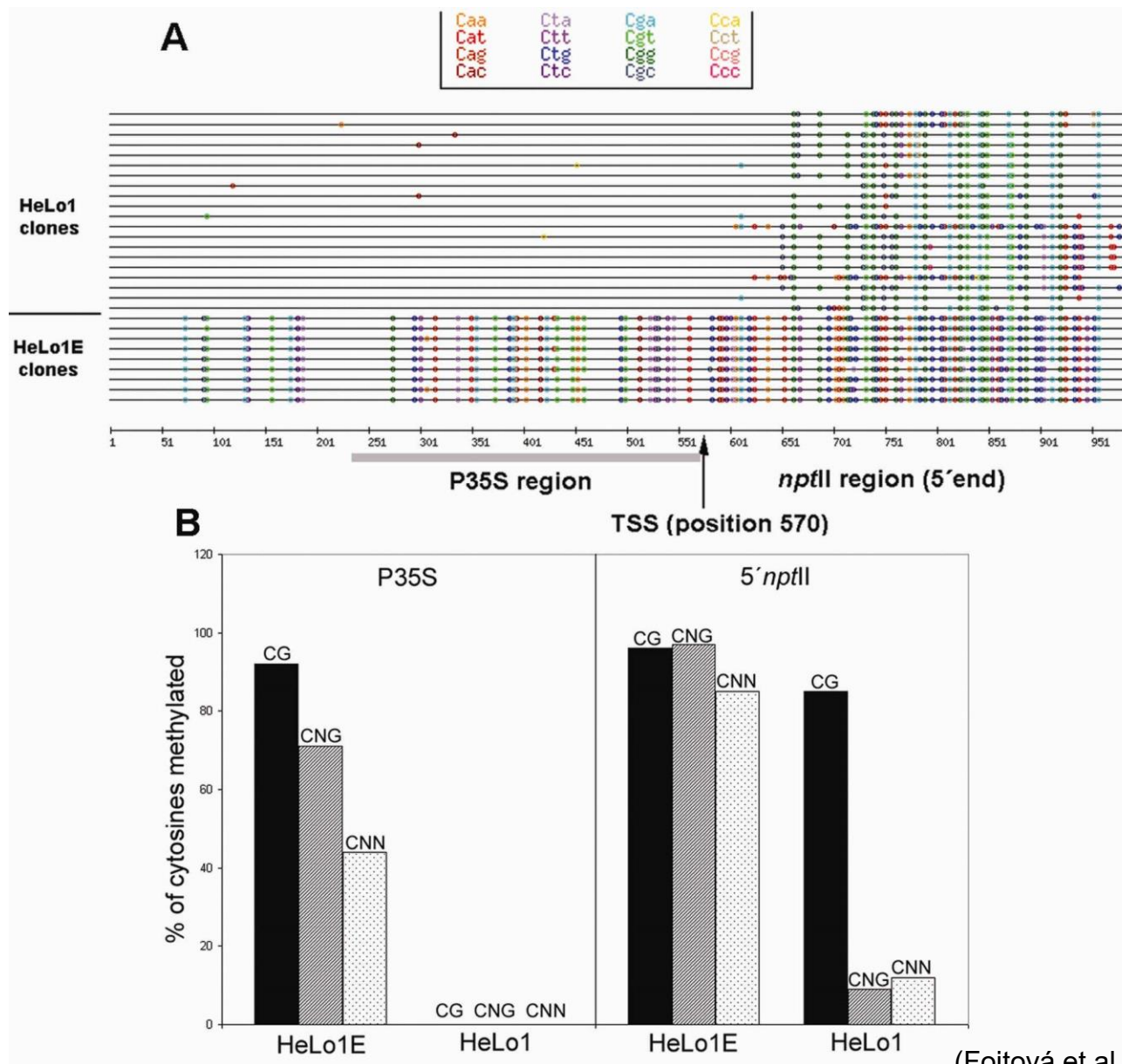
Modifikovaná DNA je namnožena pomocí PCR, uracily se párují s adeninem jako thyminy. **Primery musí amplifikovat modifikovaný i nemodifikovaný templát;** amplifikuje se každé vlákno zvlášť – nejsou již po modifikaci komplementární.

PCR produkt se klonuje a sekvenuje – cytosiny jsou pouze tam, kde byly původně ^mC.

Výhoda analýzy:

informace o lokalizaci methylovaných cytosinů v celé sekvenci, ne pouze v konkrétním restrikčním místě





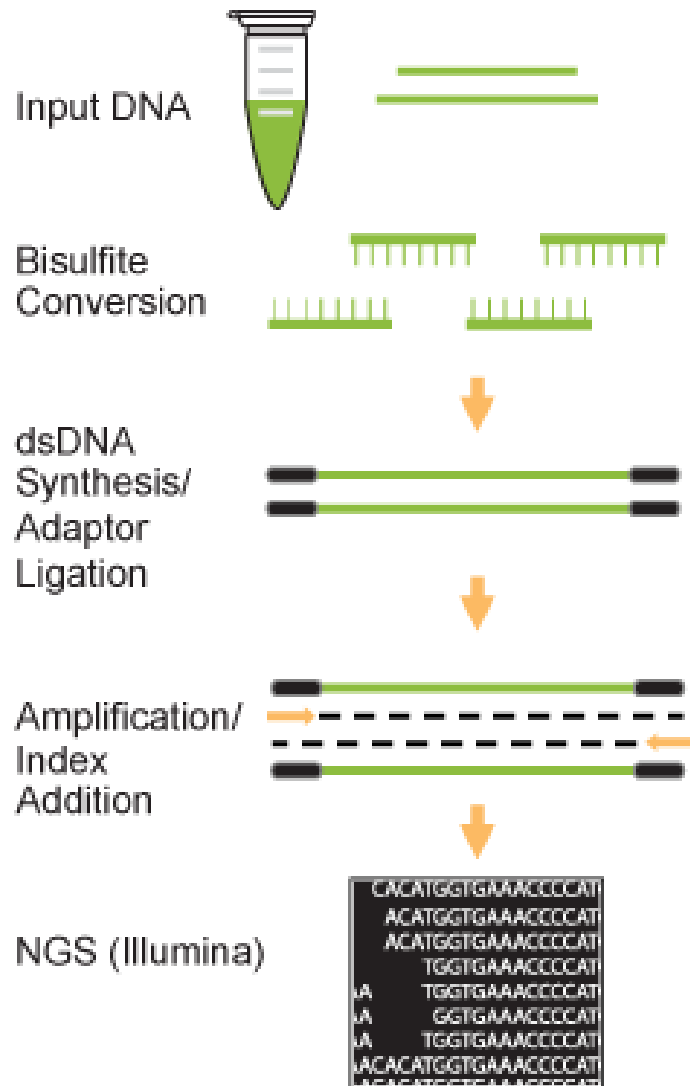
Celogenomová analýza methylace DNA:

- 1. Sekvenování DNA po modifikaci bisulfitem**
- 2. Imunoprecipitace pomocí protilátek proti ^mC nebo afinitní purifikace pomocí ^mC vazebného proteinu**
- 3. Oxford Nanopore Technology**

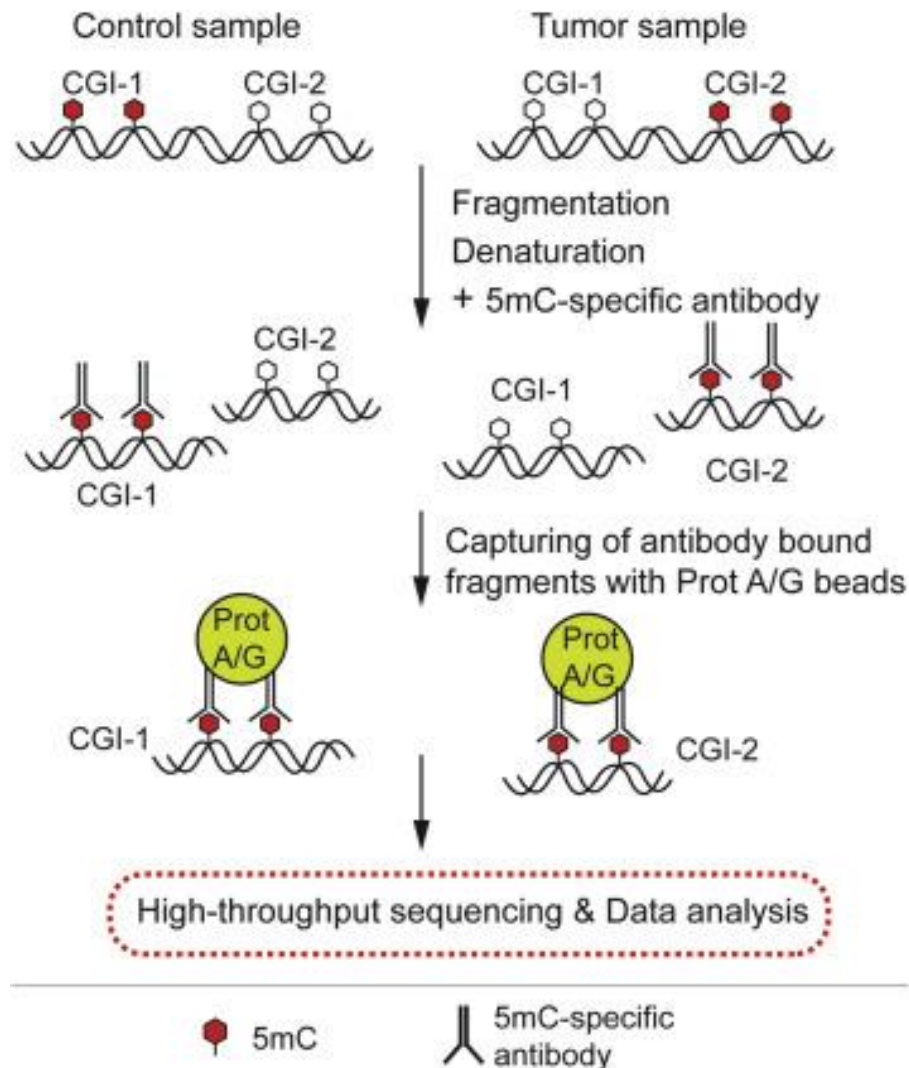
Výsledky analýzy genomu *Arabidopsis*:

- cca 20% cytosinů v genomu je methylovaných
- nejvíce ^mC je v transpozonech a repetitivních sekvencích
- nejméně methylované jsou promotory endogenních genů
- asi 1/3 genů obsahuje „body methylation“ (CG místa na 3'konci kódující oblasti)

Ad 1) Sekvenování DNA po modifikaci bisulfitem



Ad 2) Imunoprecipitace protilátkami proti 5mC nebo proti MBD proteinu, který se na 5mC váže

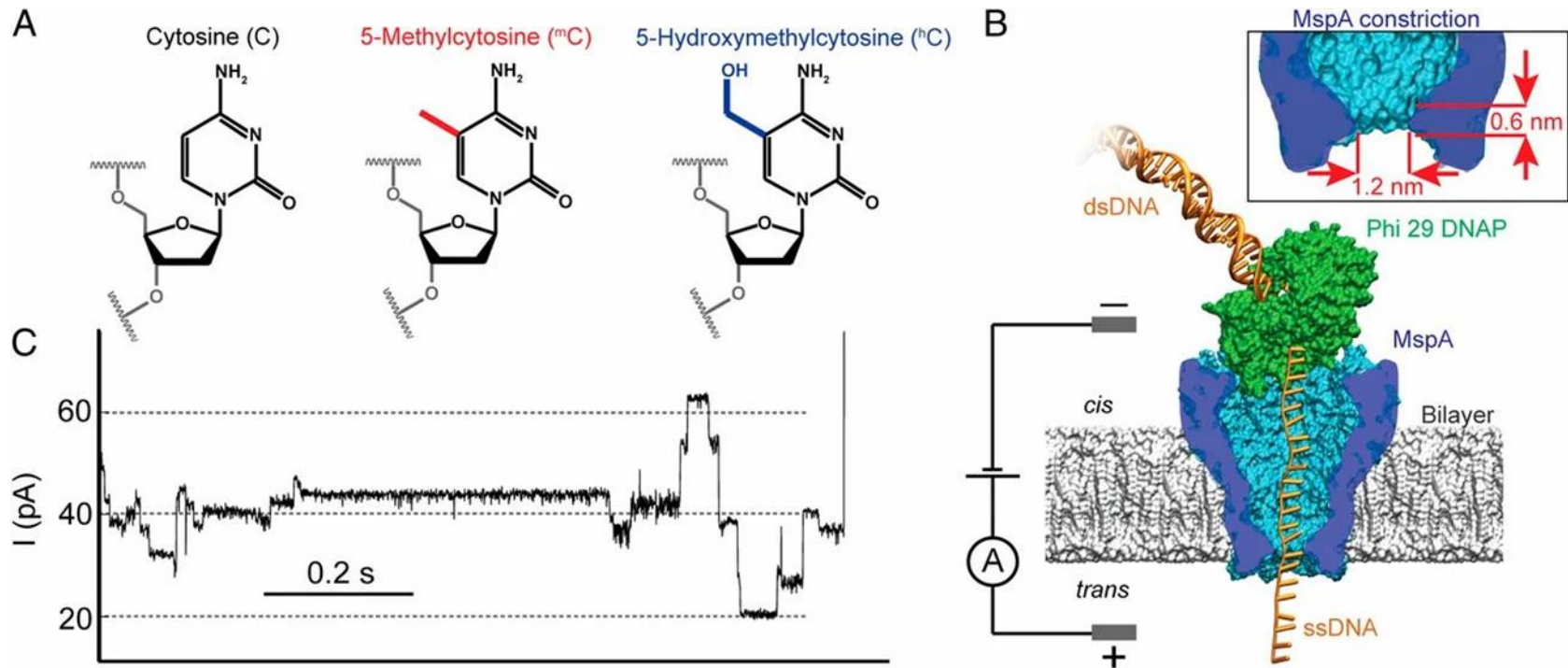


Nevýhoda: přesná poloha 5mC není takto zjistitelná, přesnost je limitována velikostí fragmentů po sonikaci.

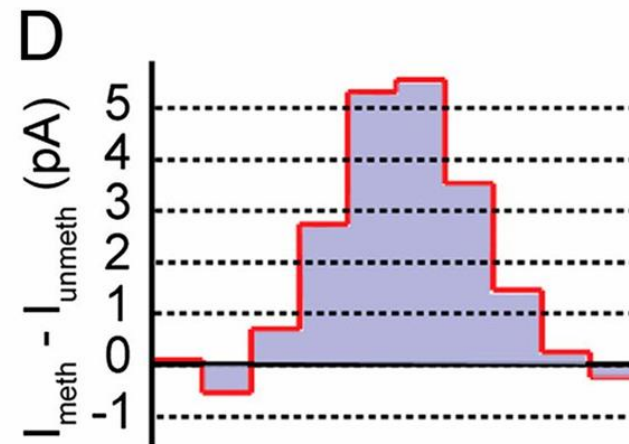
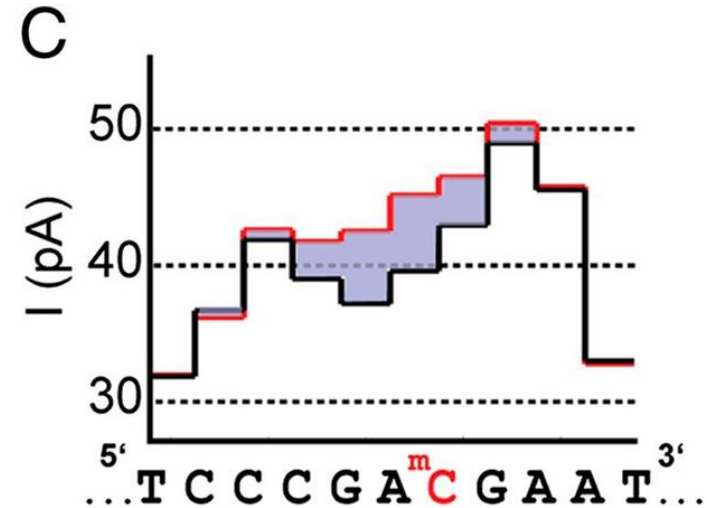
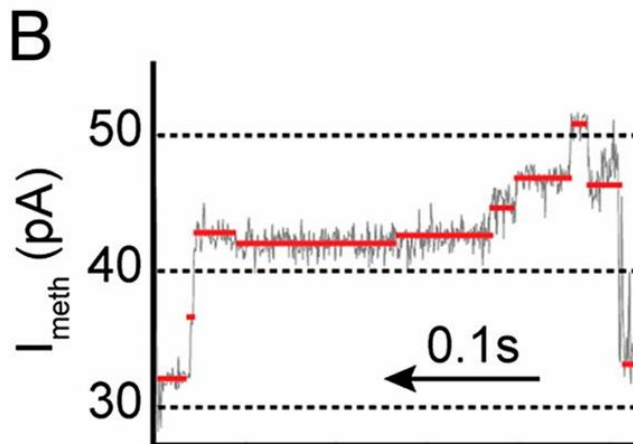
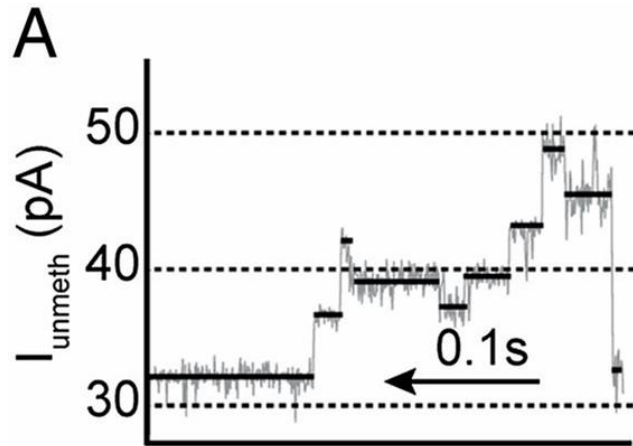
Ad 3) Přímé čtení ^mC a ^hC pomocí ONT nebo PacBio SMRT sekvenování

Měří se změny vodivosti / proudu při průchodu vlákna DNA nanoporem (*Mycobacterium smegmatis* porin A (MspA) (Laszlo et al., PNAS 2013)

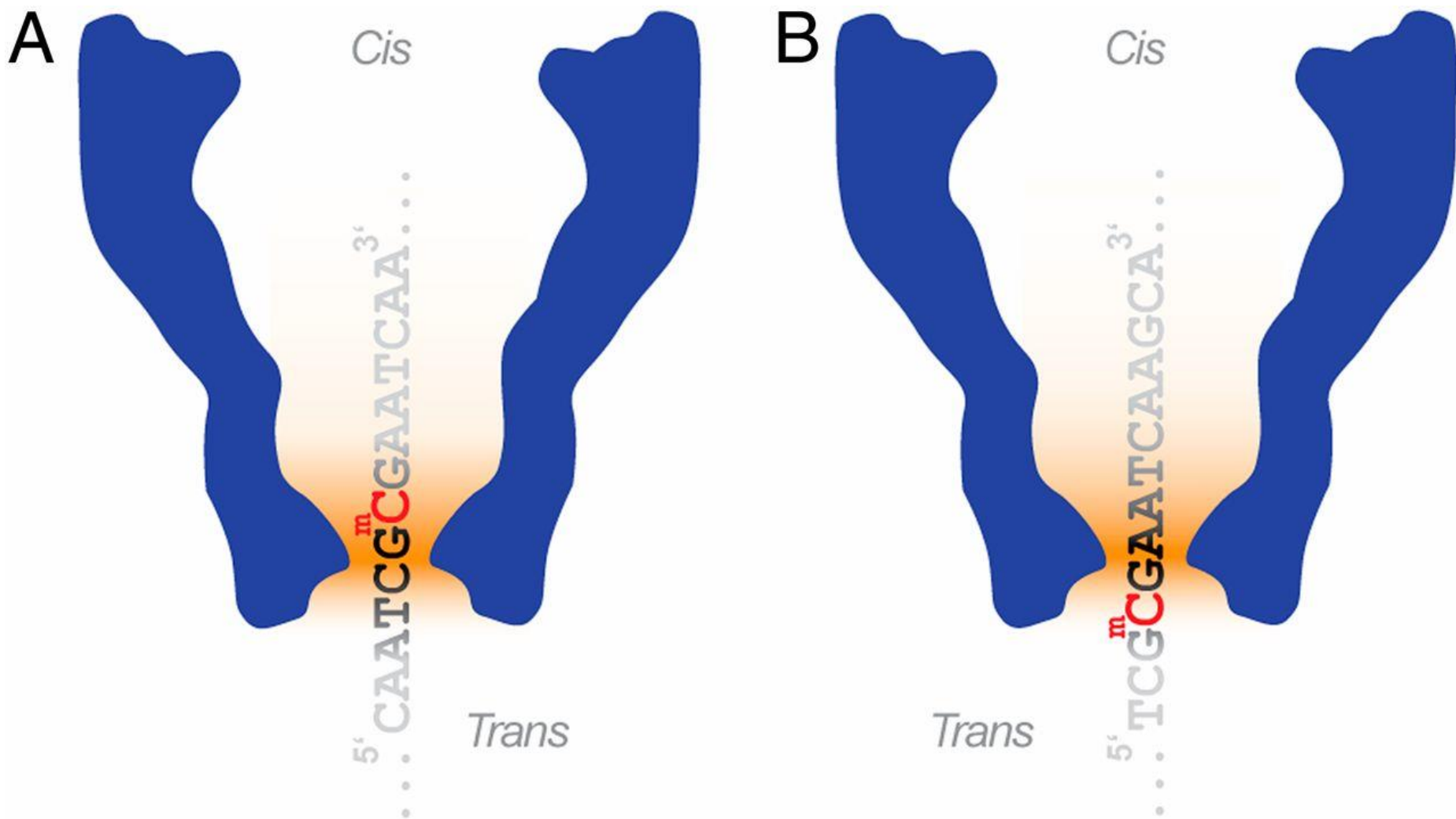
MspA je zanořen v membráně, na kterou je vloženo napětí, které působí protahování ssDNA přes nanopore. V tomto konkrétním uspořádání Phi29 DNAP katalysuje definovaný posun DNA v jednonukleotidových krocích přes MspA.



Sleduje se proud při průchodu nanoporem. Přítomnost ^mC nebo ^hC způsobí změny **vodivosti** detekované při průchodu asi 4nt úseku nejužším místem nanoporu. Přítomnost ^mC proud zvyšuje oproti C, přítomnost ^hC naopak snižuje.

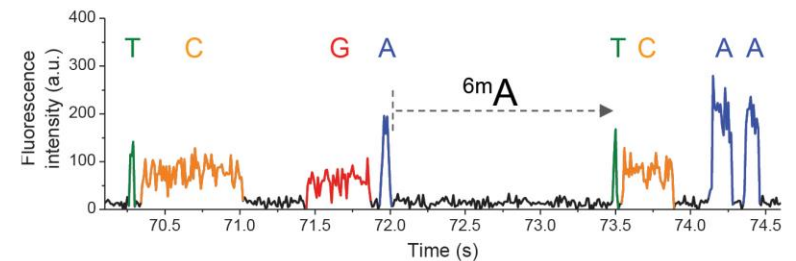
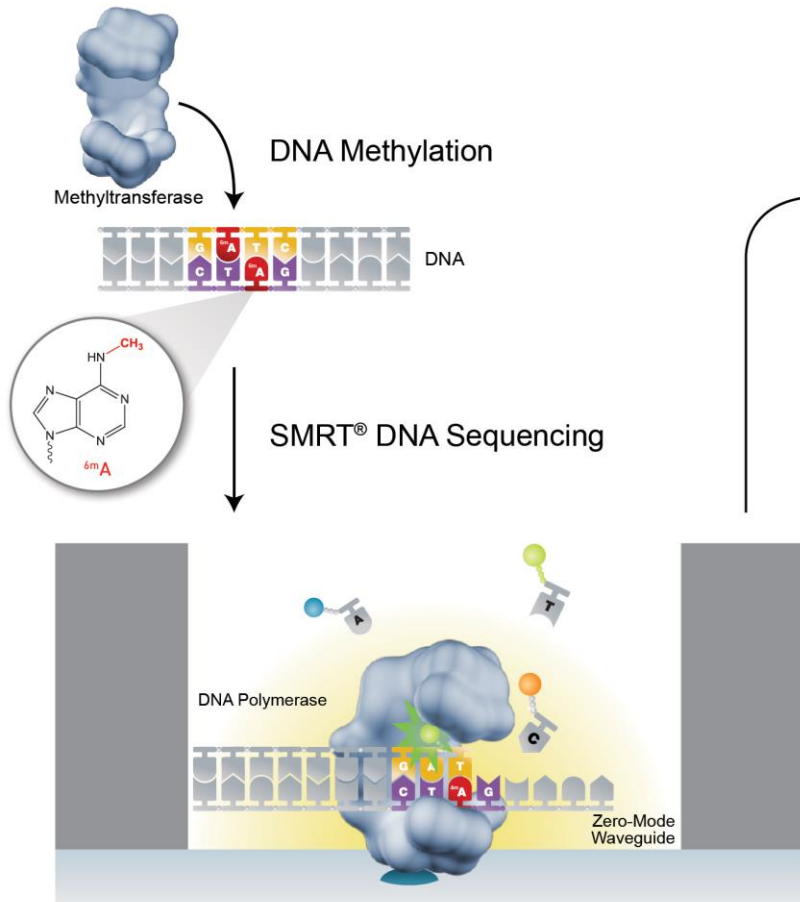


Oranžově vyznačená oblast v největším zúžení uvnitř nanoporu má největší proudovou hustotu, v tomto úseku DNA je tedy vliv bází a jejich modifikací na velikost el. proudu největší.

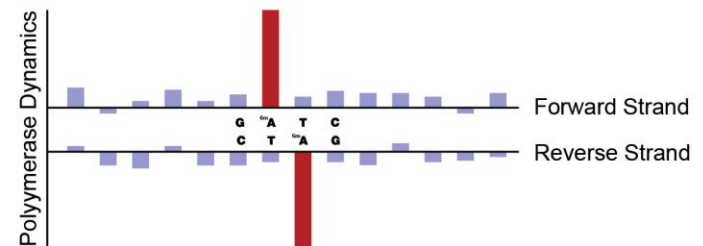


Vyvíjejí se ONT aplikace na **přímou detekci modifikací RNA** – těch je asi 170 !!! – klíčový nástroj pro rozvoj **EPITRANSKRIPТОMIKY**.

Při použití PacBio SMRT sekvenování je detekce modifikovaných nukleotidů založena na zpomalení kinetiky - delším čase, který potřebuje modifikovaný nukleotid pro průchod DNA polymerázou



Analysis of Polymerase Kinetics



Detekce posttranslačních modifikací (PTM) histonů – methylace, acetylace, fosforylace aj.

Methylace – např. lysin v poloze 9 na histonu H3 (H3K9)

Distribuce euchromatinových a heterochromatinových značek v *Arabidopsis thaliana* a myši (podle Fransz *et al.*, 2006)

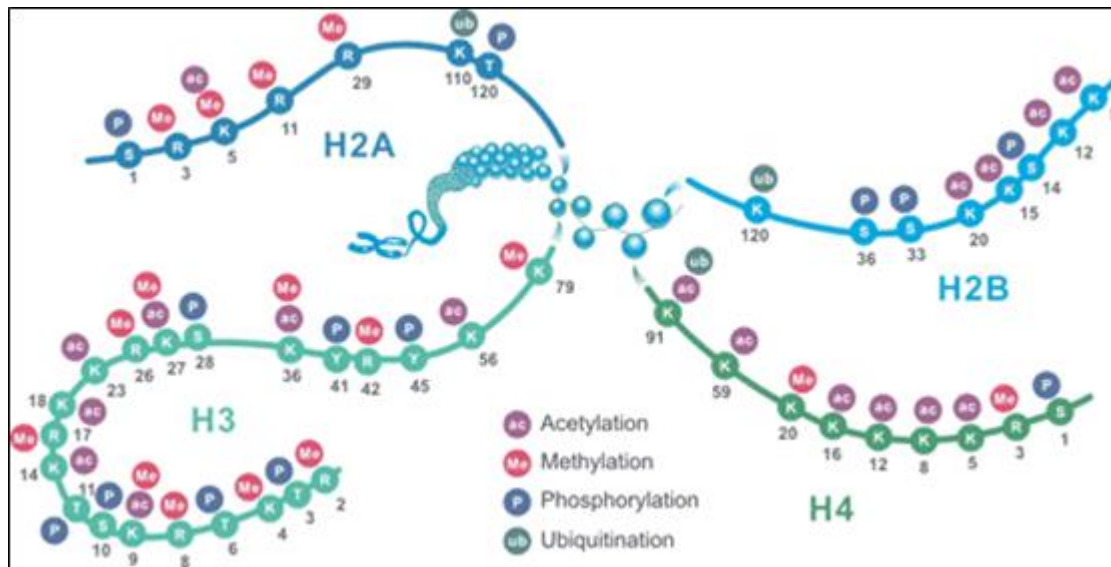
brněnské názvosloví



Brno nomenclature – established in 2004 at the first meeting of the Epigenome Network of Excellence

Modifikace	Stupeň	<i>A. thaliana</i>		myš	
		euchromatin	heterochromatin	euchromatin	heterochromatin
H3K9	mono	-	+	+	-
	di	-	+	+	-
	tri	+	-	-	+
H4K20	mono	-	+	+	-
	di	+	-	+	-
	tri	+	-	-	+
5m-C		-	+	-	+

Jde o druhově- a dokonce lokus-specifické, dynamické modifikace



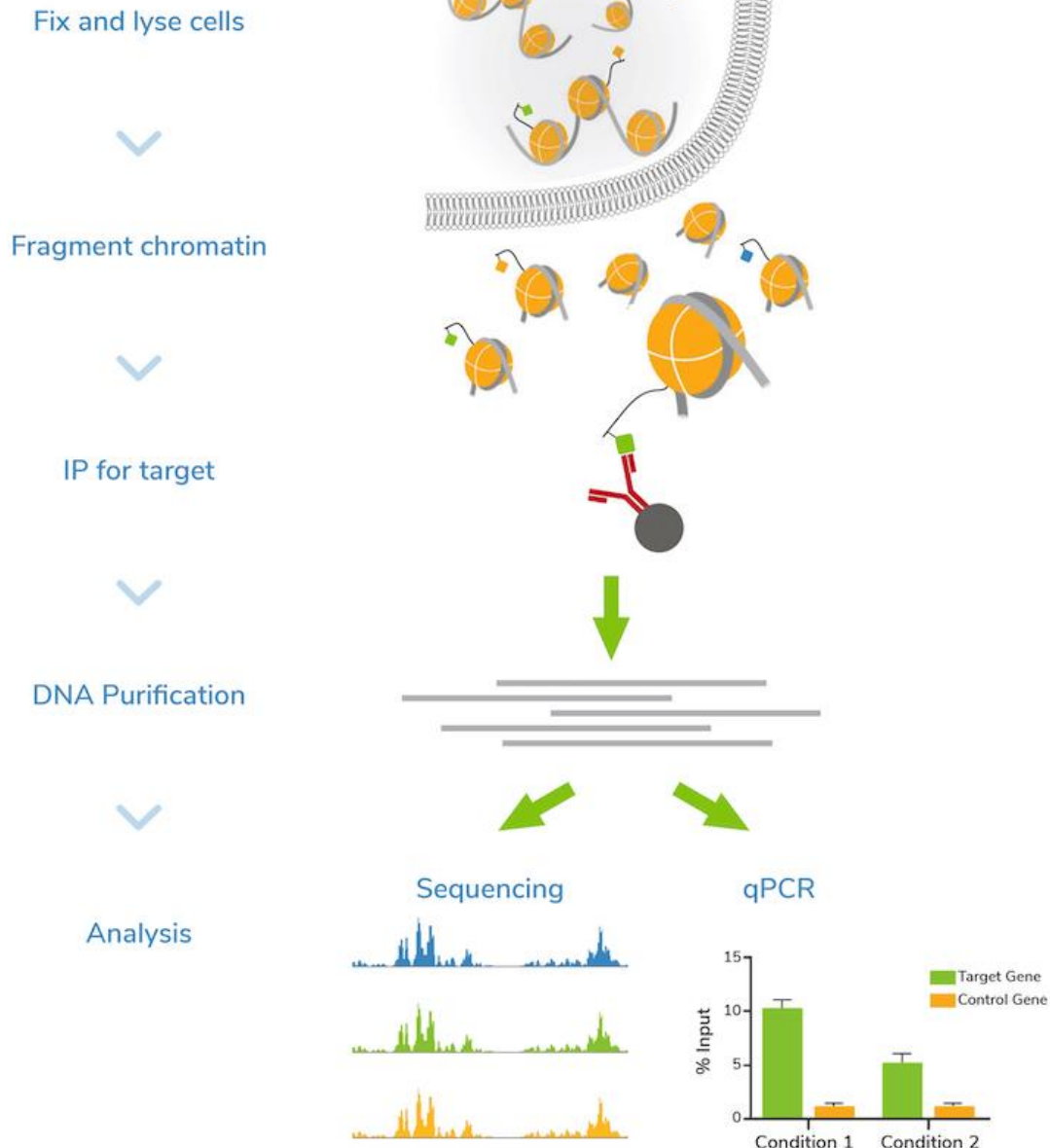
Globální úroveň např. různě acetylovaných forem lysinu v jednotlivých histonech lze určit např. pomocí **hmotnostní spektrometrie (MS)**. Avšak ta nám neřekne, **kde na chromozómech** se příslušná modifikace nachází – a to nás zajímá z hlediska regulačních mechanismů nejvíc.

Jak na to? Analýza chromatinu pomocí imunoprecipitace - ChIP

ChIP – PCR (pro kvantitativní analýzu konkrétní oblasti chromatinu)

ChIP – Seq (pro kvantitativní analýzu přítomnosti dané PTM v chromatinu na celogenomové úrovni)

ChIP Workflow



Po **fixaci** a **lyzi** buněk se chromatin **fragmentuje** (ultrazvukem, enzymaticky), optimálně na mononukleozomální fragmenty

Fragmenty chromatinu se **inkubují s protilátkou** specifickou pro hledanou modifikaci (např. H3K4me3). Protilátka je ukotvena k magnetickým kuličkám pro snadnější separaci.

Ze separovaného chromatinu se odstraní proteiny působením proteinázy K a crosslinky se revertují působením vysoké koncentrace solí za zvýšené teploty

Následuje analýza pomocí NGS nebo qPCR. Současně se analyzuje **input** (původní vzorek před IP) a vzorek zpracovaný stejným postupem bez protilátek (**nespecifické pozadí**)

Pro ChIP-seq se provádějí následující kroky:

1. Příprava knihovny pro NGS a indexování – DNA získaná z ChIP DNA a input DNA se opravují a amplifikují pro NGS. Během PCR (9-16 cyklů) se přidávají různé indexy (barcodes), aby bylo možno odlišit jednotlivé knihovny v multiplexní NGS.
2. Připravené knihovny se kvantifikují, potvrzuje se velikostní profil – typicky kapilární elfo nebo mikrofluidika
3. Pro multiplexní sekvenování se knihovny opatřené barcody spojují v ekvimolárních poměrech (na základě předchozí kvantifikace) a nanáší na NGS platformu (zpravidla Illumina).

Důležité parametry:

Počet buněk – zpravidla 500000 - 1000000 buněk

Počet replik – zpravidla 3 biologické repliky na každý vzorek, technické repliky se dělají při optimalizaci, u optimalizované techniky nejsou nutné.

Crosslinking - volba činidla (např. formaldehyd, glutaraldehyd, disukcinimidyl-glutarate [DSG]) a optimalizace podmínek. Při nativní ChIP lze crosslink vynechat.

Fragmentace chromatinu – žádoucí je 150-300 bp (sonikátory Diagenode Bioruptor, Covaris). Při nativní ChIP se upřednostňuje enzymatické štěpení – MNáza.

Výběr vhodné protilátky (měla by být deklarována jako ChIP-grade antibody nebo Ab pro IHC

Optimalizace podmínek IP.....

Optimalizace sonikace (Covaris)

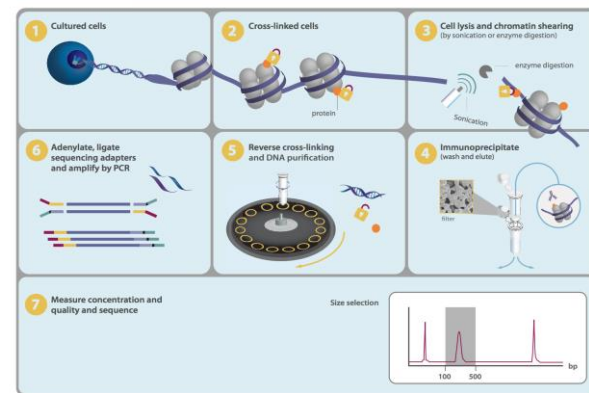
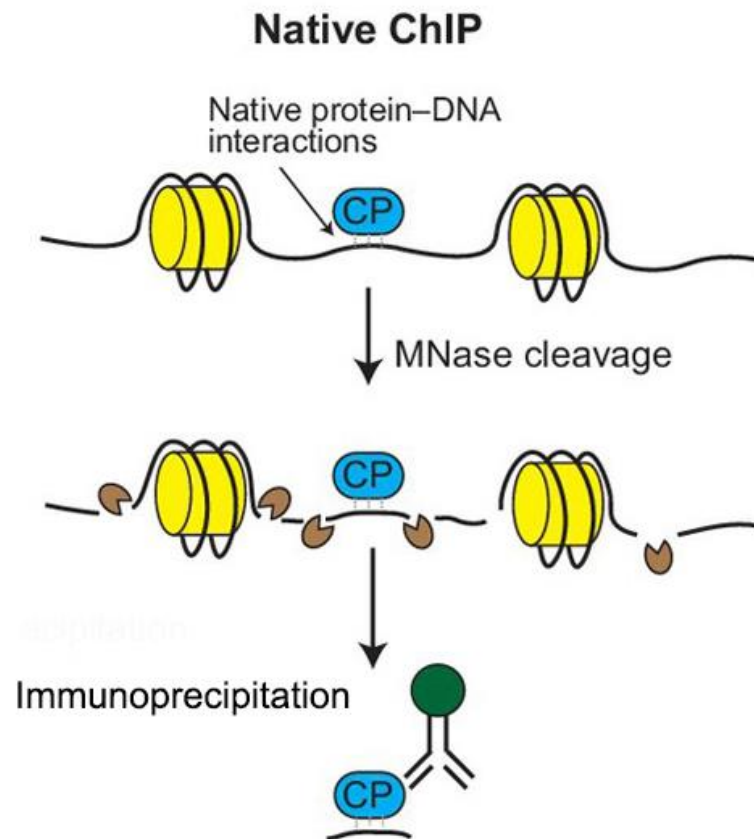
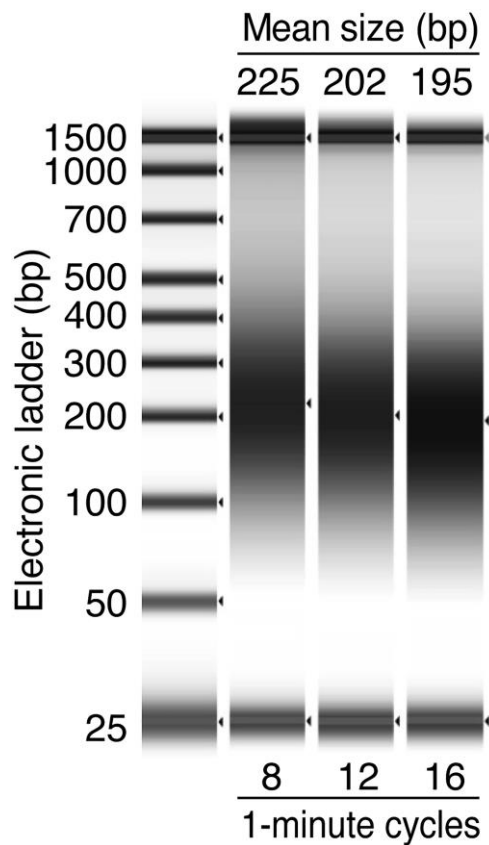
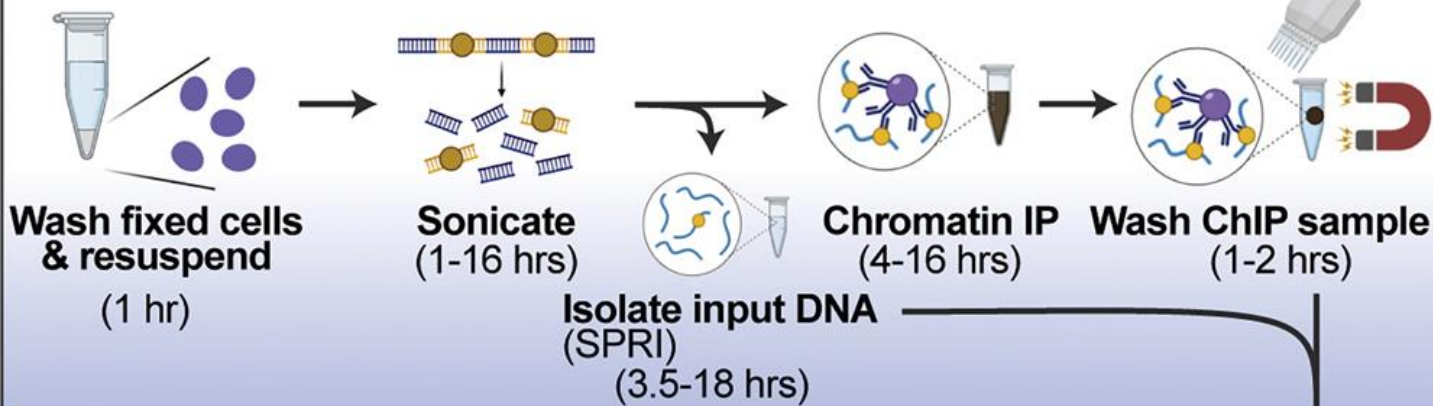


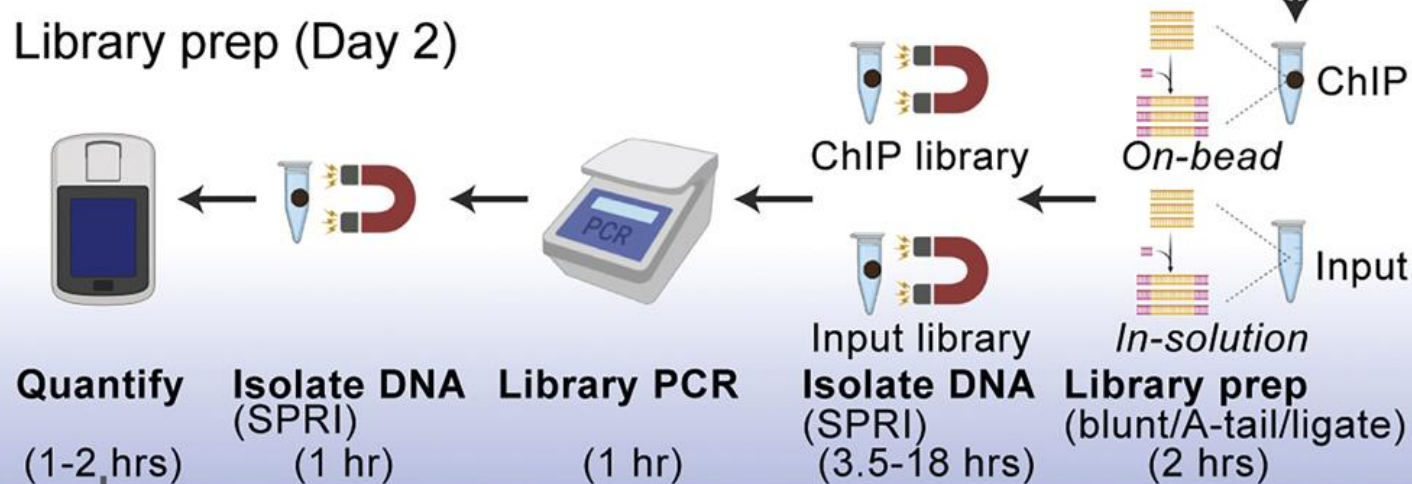
Figure 1: Overview of ChIP-seq process

Příklad reagenčních souprav- ChIP-Seq kit Chromatrap

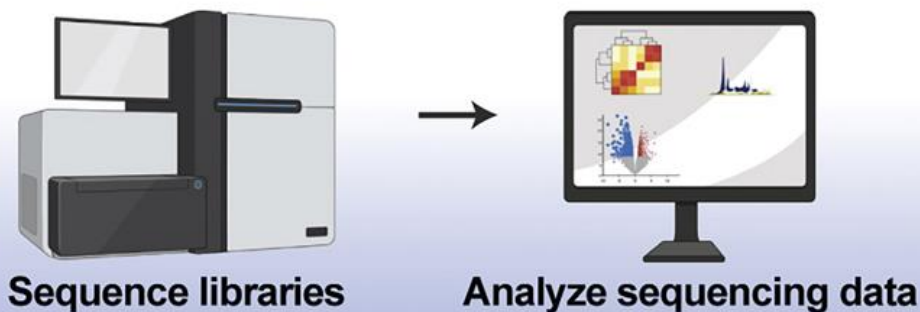
ChIP (Day 1-2)



Library prep (Day 2)



Sequence/Bioinformatics

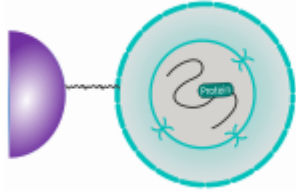


Detailní protokol
např. zde:
<https://star-protocols.cell.com/protocols/491>

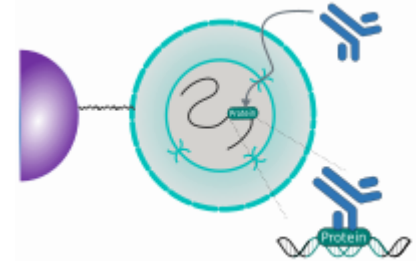
Vylepšené přístupy (nativní):

CUT&RUN přístup: Extrakce bez fragmentace, k fragmentaci dochází uvnitř permeabilizovaných buněk

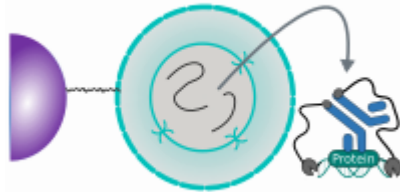
1. Imobilizace
(magnetické kuličky
s Concanavalinem
A) a permeabilizace
buněk (digitonin)



2. Vazba primární
protilátky proti
hledanému
proteinu



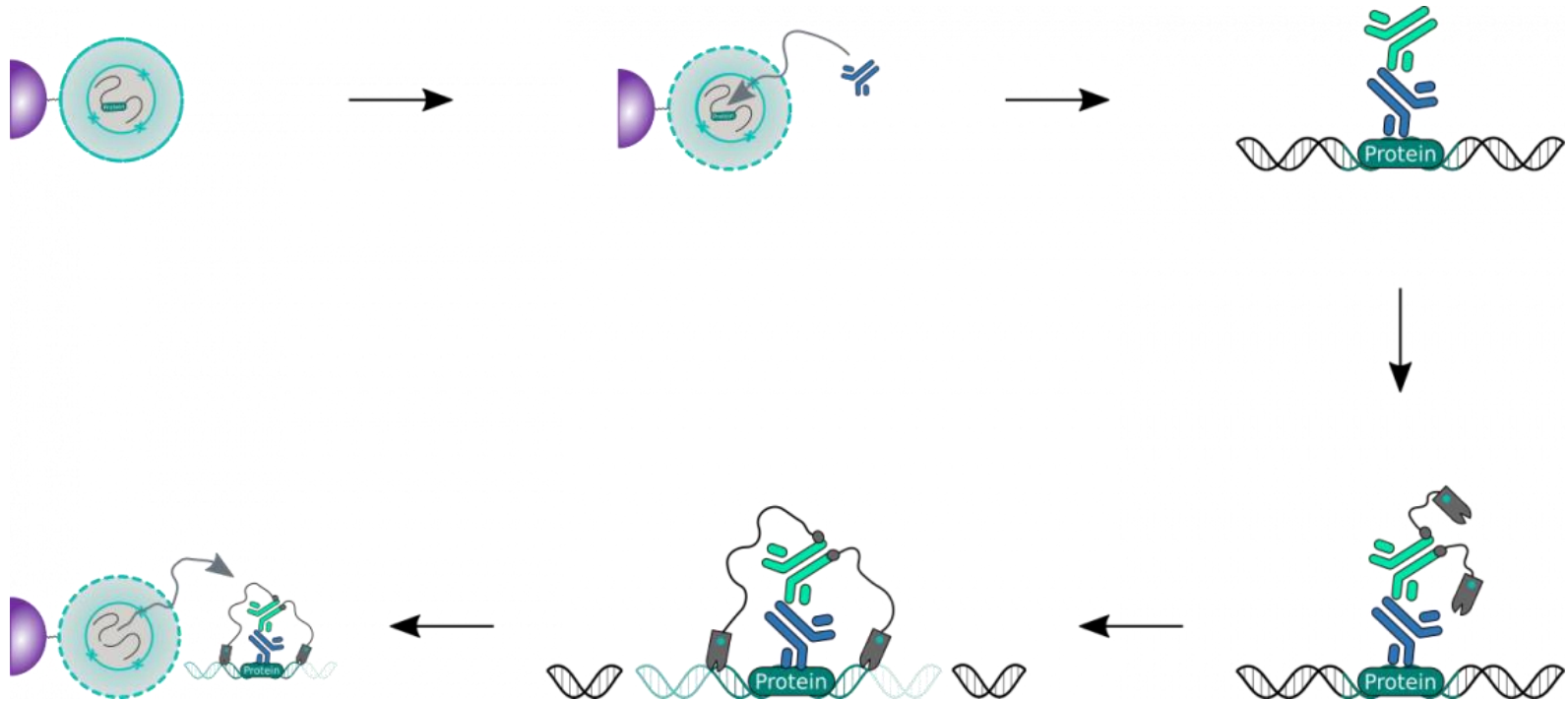
4. Uvolnění
chromatinu – lyze
močovinou



3. Přidání Mnázy
fúzané s
proteinem A/G –
pAG-Mnase (vazba
na Fc fragment
primární protilátky



CUT&Tag – místo MNázy se používá pA/G-Tn5 (**Tn5 je hyperaktivní transponáza**) s duplexními sekvenačními adaptory. K transpozici dochází na duplexní DNA



Postup: 1. imobilizace buněk 2. Permeabilizace buněk 3. Vazba primární protilátky, vazba sekundární protilátky 4. Vazba pA-Tn5 5. „Tagmentace“ 6. Uvolnění chromatinu/lyze buněk

Mezi významné epigenetické charakteristiky patří také **POLOHA NUKLEOZÓMŮ** (zejm. v promotorových oblastech) – limituje dostupnost promotorů pro transkripční faktory. K její změně dochází při remodelaci chromatinu. Základní metodou mapování polohy nukleozómů je **štěpení MNázou** v mezerkách mezi nukleozómy

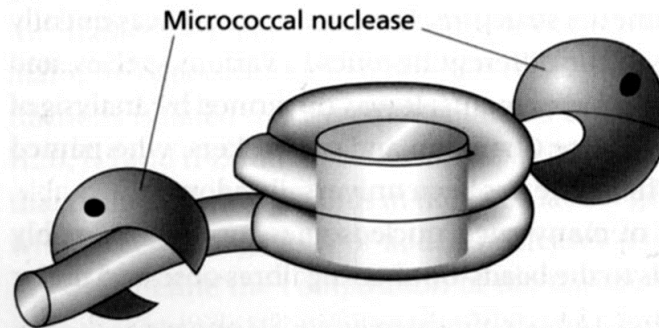


Fig. 3.4 Micrococcal nuclease must surround DNA in order to cut it, so DNA in the nucleosome core is protected, even though it lies on the external surface.

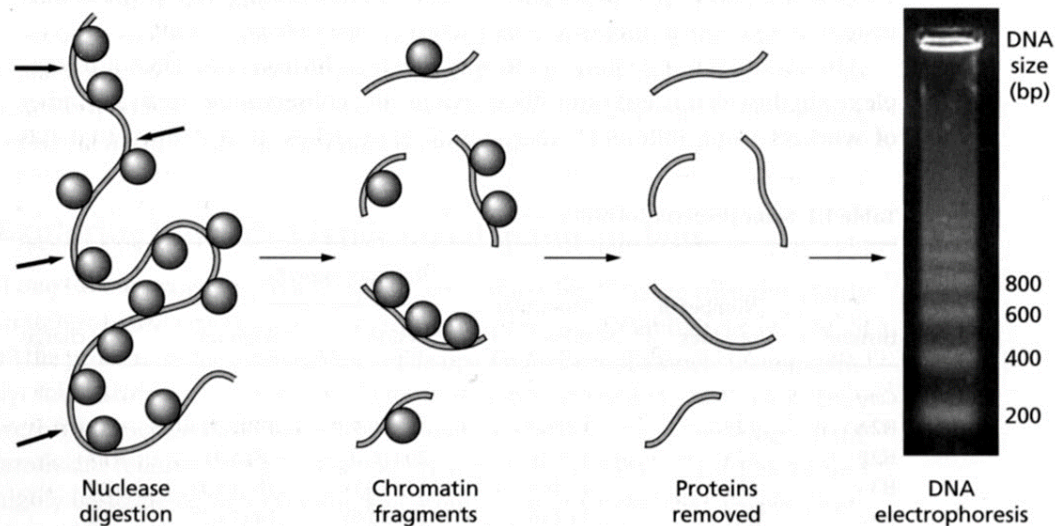
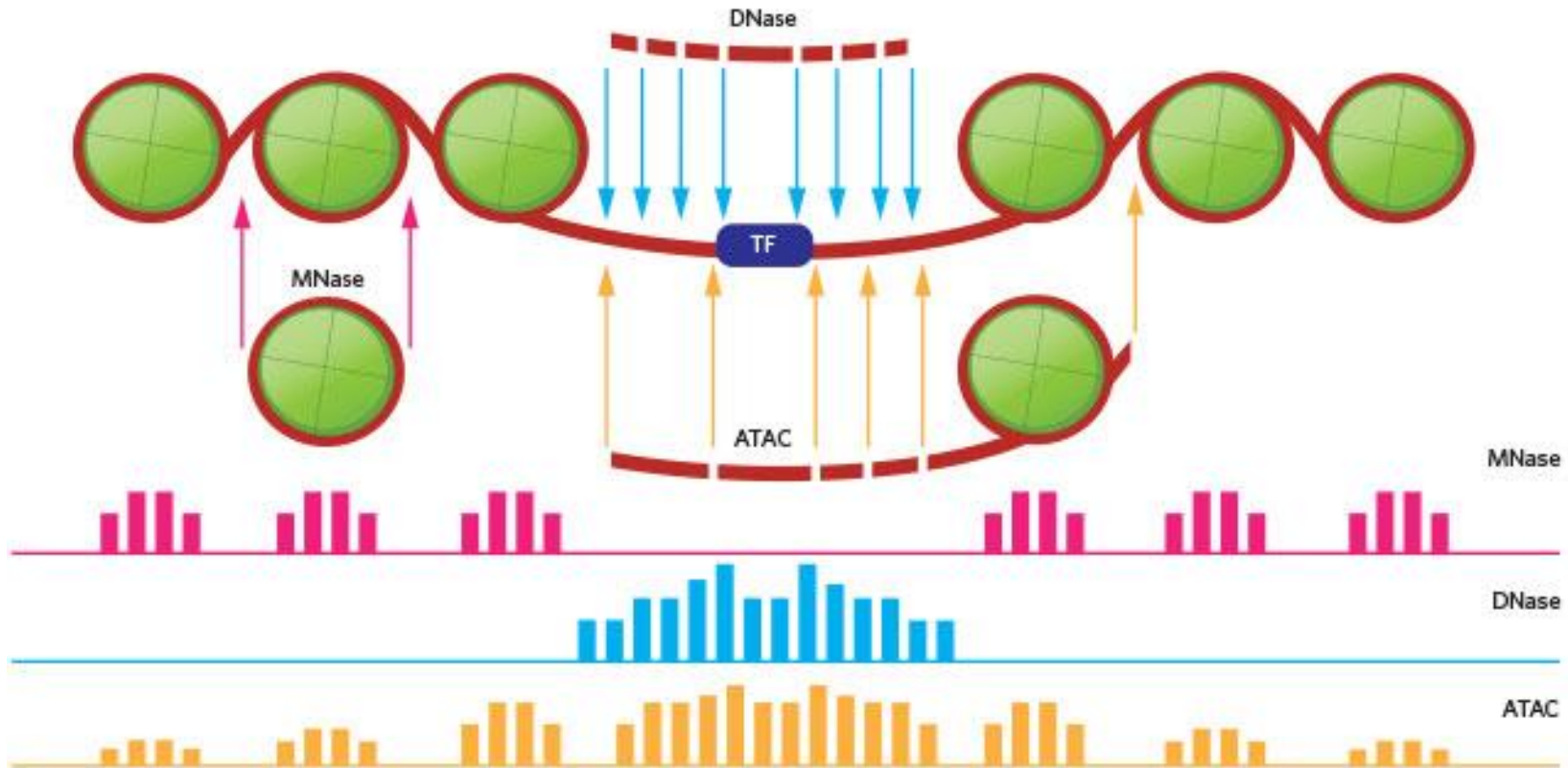


Fig. 3.1 DNA fragments generated by exposure of chromatin to nucleases have sizes that are multiples of about 200 bp. Thick arrows illustrate possible nuclease cutting sites.

Na celogenomové úrovni se pro analýzu struktury chromatinu používají již uvedené CUT&RUN, CUT&TAG přístupy– **mapování polohy nukleozómů**, míst bez nukleozómů (DNase-sensitive míst) apod...

V současnosti nejužívanější způsob mapování polohy nukleozómů na celogenomové úrovni je **MNase-Seq** (na mononukleozomové frakci) a **ATAC-Seq** (využití Tn5)



ATAC-seq (Assay for Transposase-Accessible Chromatin using **sequencing**)

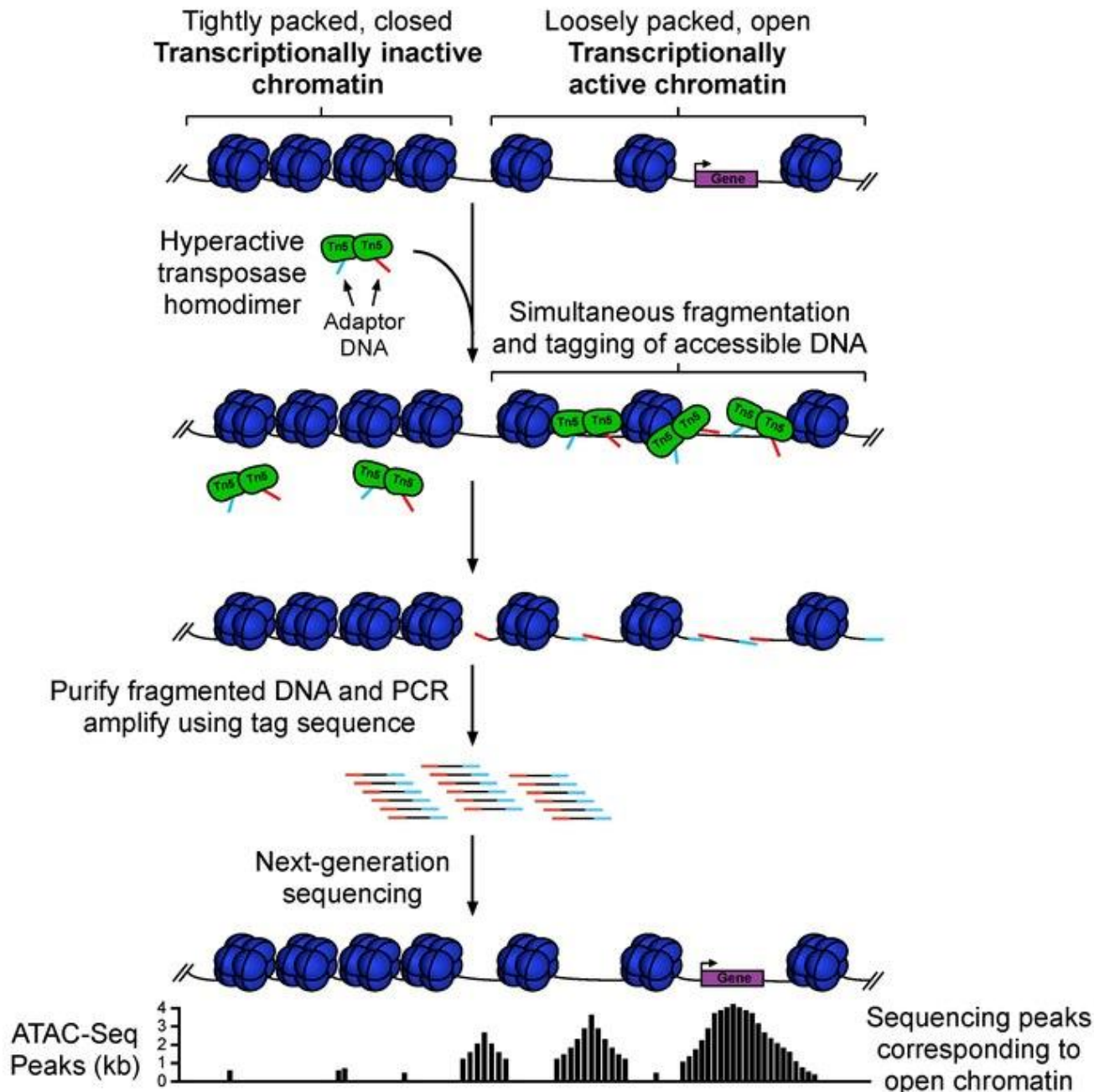
Princip ATAC-Seq

Buenrostro et al, Nature Meth 2013

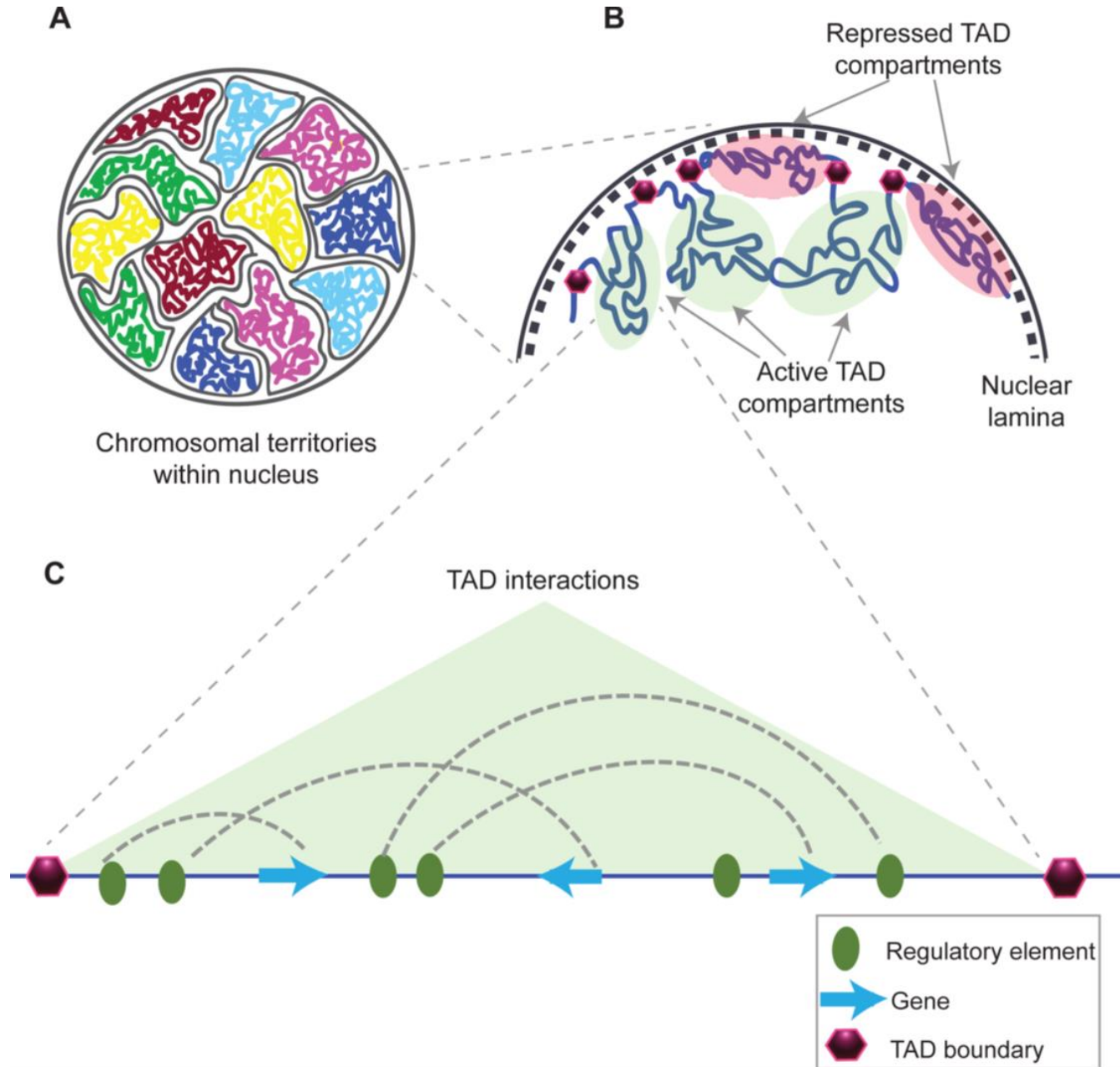
Assay for Transposase-Accessible Chromatin using **sequencing**)

ATAC-seq identifikuje přístupné oblasti DNA tak, že sonduje otevřený chromatin pomocí hyperaktivní mutantní **Tn5** transponázy, která vkládá sekvenační adaptory do otevřených oblastí chromatinu (*tagmentace*).

Tn5 štěpí a vkládá tagy z dvojvláknové DNA obsahující sekvenační adaptory



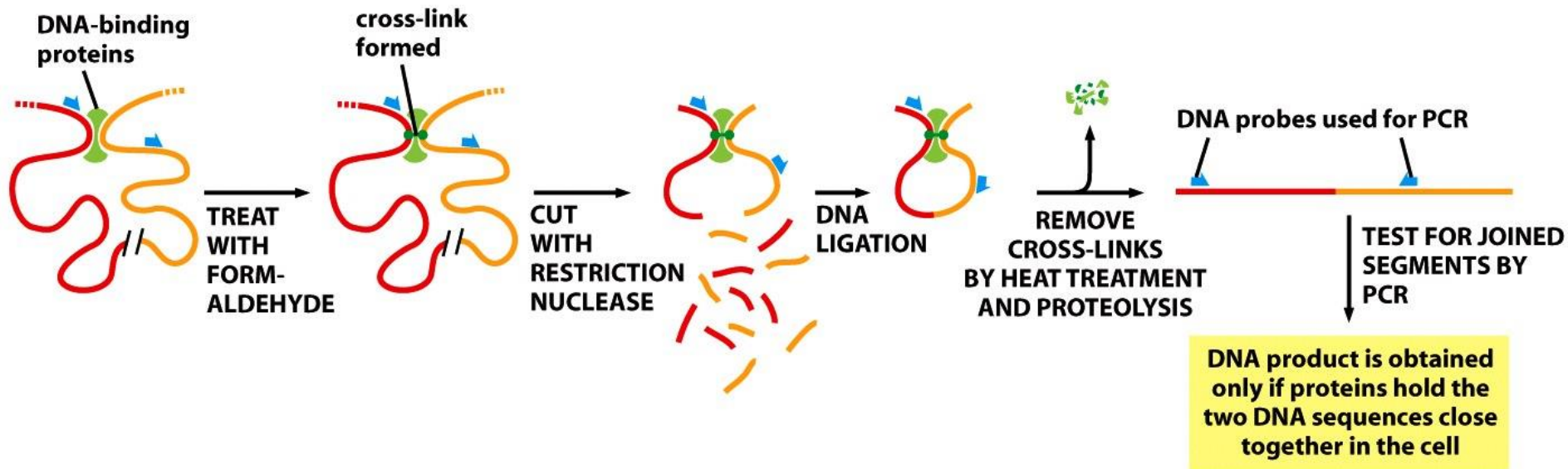
Mapování vyššího uspořádání chromatinu (topologicky asociované domény - TAD, kompartmenty a chromozomová teritoria)



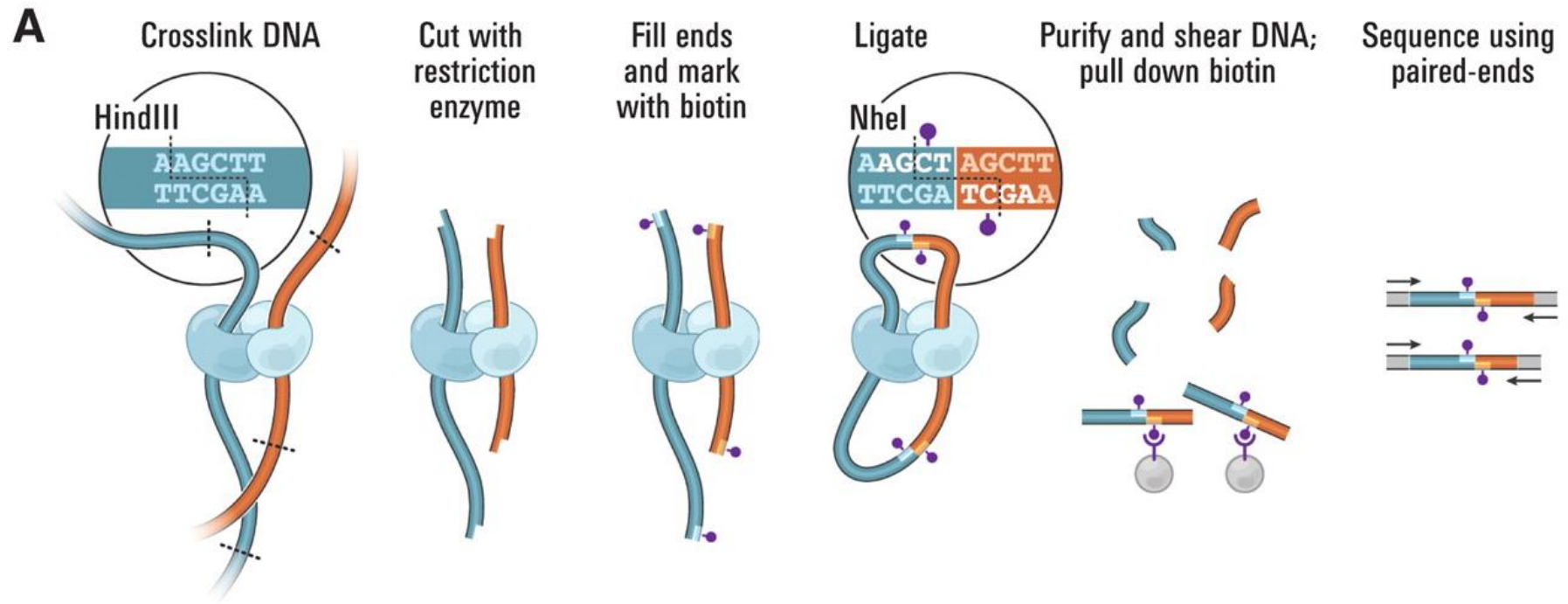
Metody analýzy 3D uspořádání chromatinu

Chromosome conformation capture assay (3C assay) –

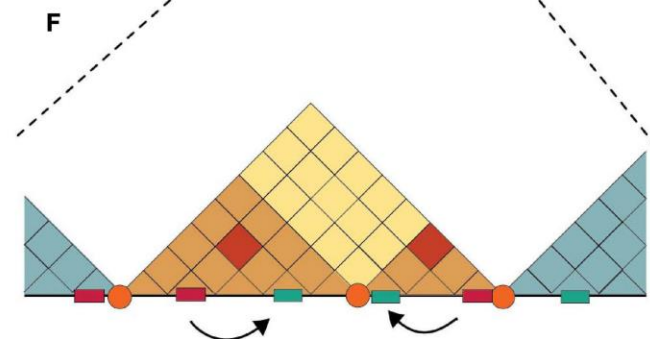
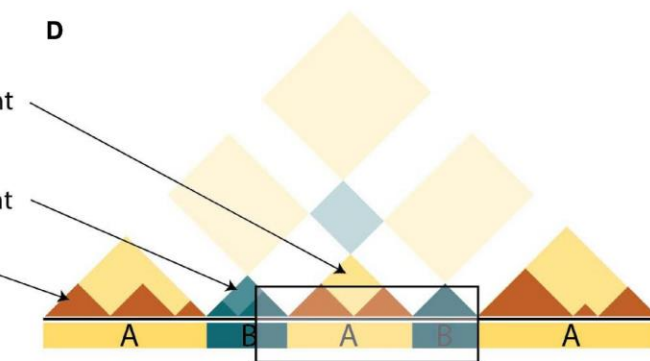
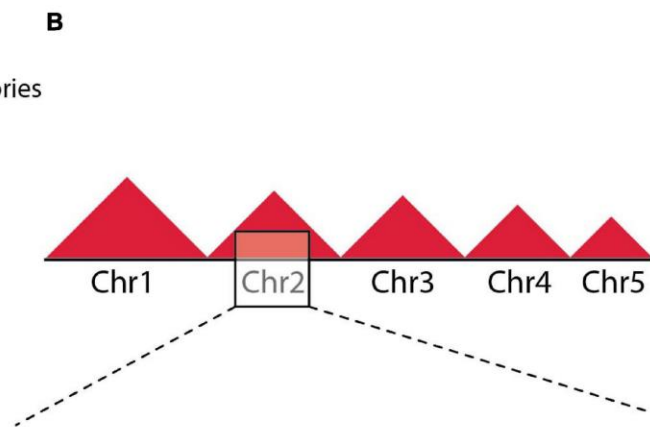
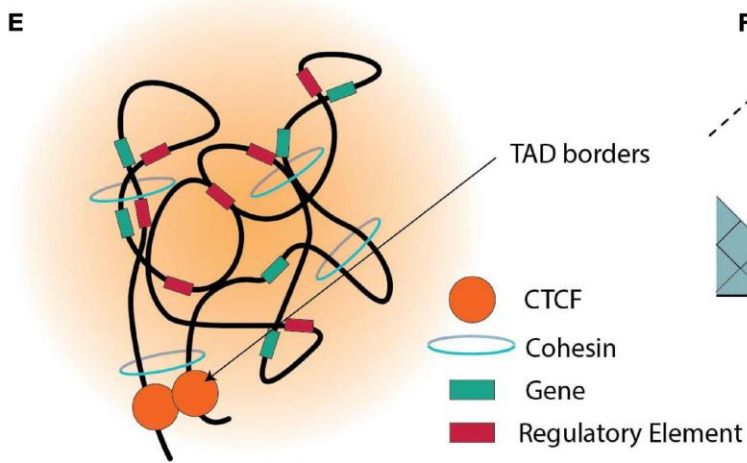
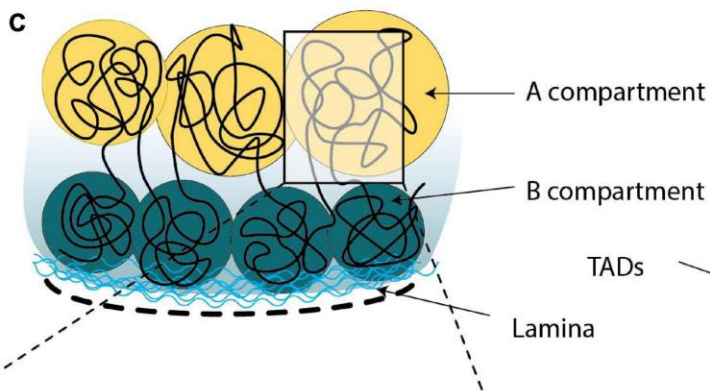
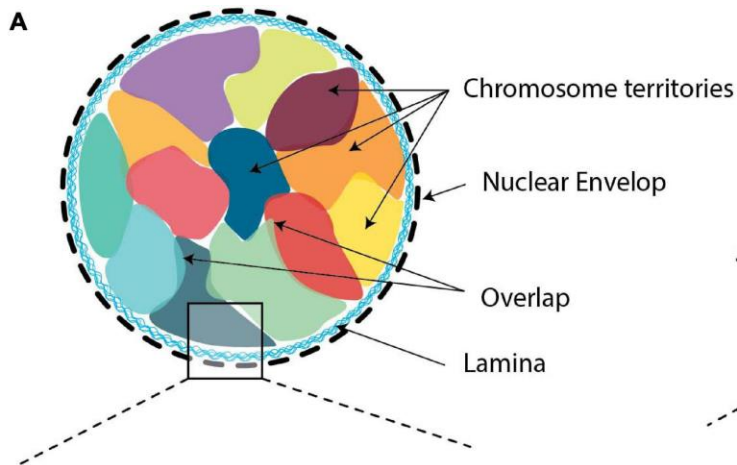
- Zjišťování interakcí mezi vzdálenými oblastmi chromatinu. V původní verzi (3C) umožňuje zjišťovat interakce jedné oblasti s jinou konkrétní oblastí. V HiC verzi umožňuje zjistit interakce „všech se všemi“



Hi-C verze CCC: analýza všech kontaktů/blízkostí



Overview of Hi-C -Lieberman-Aiden et al., Science 2009. (A) Buňky jsou crosslinkovány formaldehydem, vznikají kovalentní vazby mezi prostorově blízkými oblastmi chromatinu. Chromatin se štěpí restriční enzymem (zde HindIII), a výsledné kohezní konce jsou doplněny nukleotidy, z nichž jeden je biotinylovaný (fialová tečka). Ligace se provádí za velkého zředění a vznikají chimérické molekuly; HindIII místo je zničeno a NheI místo vytvořeno. DNA se čistí a láme/sonikuje na krátké fragmenty. Biotinylované spoje se izolují pomocí streptavidinových kuliček a identifikují pomocí paired-end sekvenování.



3D organizace chromatinu

A) schema uspořádání chromosomů v jádře, chromosomy jsou v kontaktu s jadernou obálkou (jadernou laminou). Každý chromosom sedí ve svém teritoriu, které je v kontaktu nebo i překryvu s jinými. **(B)** Schema Hi-C map chromosomů v genomové škále. Převažují intra-chromosomální kontakty nad inter-chromosomálními. **(C)** Chromatin je uspořádán do kompartmentů “A” (žlutá) a “B” (zelená). “B” kompartmenty jsou u jaderné laminy. **(D)** Schema Hi-C map na úrovni kompartmentů. Vzdálené chromatinové kontakty vytvářejí výrazný kostkovaný vzor s A and B kompartmenty. **(E)** Topologicky asociované domény (TADs) se vytvářejí smyčkami chromatinu, a u TAD hranic se nalézají architektonické proteiny (SMC, CTCF). Uvnitř každé TAD je sbalování chromatinu zprostředkováno prstenci **kohezinu** **(F)** Schema Hi-C map na sub-Mb škále, TADs se jeví jako interakčně-bohaté trojúhelníky oddělené hranicemi **TAD**. Díky smyčkovému uspořádání jsou enhancery přeneseny do blízkosti promotorů, které ovlivňují.

EPITRANSKRIPТОMIKA – studuje biochemické modifikace RNA (transkriptomu) v buňce. Modifikace nekódujících RNA (ncRNA, lncRNA, rRNA, tRNA) i kódujících (mRNA)

Mnohem širší repertoár - asi 170 známých modifikací RNA!

Nejběžnější je N6-Methyladenosin (m^6A) – asi 3x v každé mRNA, často v posledním exonu většinou v 3'UTR

Podobně jako v epigenetice, i zde lze aplikovat koncept „writers, readers, erasers“.

V případě m^6A je *writerem* komplex proteinů m^6A methyltransferáza. Eraserem jsou nejméně dvě demethylázy (ALKBH5, FTO). Vyřazení FTO u myši způsobuje poruchy v kontrole hmotnosti těla, v případě Alkbh5 neplodnost.

Přítomnost m^6A je potřebná pro export mRNA z jádra (závisí i na dalších značkách, např. m^5C), ovlivňuje účinnost a přesnost translace na ribozomu, zejm. interakci s tRNA.

m^6A Readers - YTHDF1 – indukuje vazbu mRNA k ribozómu; eIF3 – indukuje iniciaci translace a vazbu k 40S podjednotce ribozómu; YTHDF2 – svou vazbou k m^6A snižuje její stabilitu, předává mRNA do P-tělísek k degradaci. Proces je významný pro rychlou degradaci transkriptů transkripčních faktorů zajišťujících pluripotenci.

Další role m^6A : zvýšení alternativního sestřihu – zvyšuje vazbu hnRNPC a tím alternativní vystřižení intronu.

Role m^6A v umlčování X-chromozomu – XIST lncRNA je vysoce methylovaná, váže YTHDC1 protein, který je potřebný k umlčení X

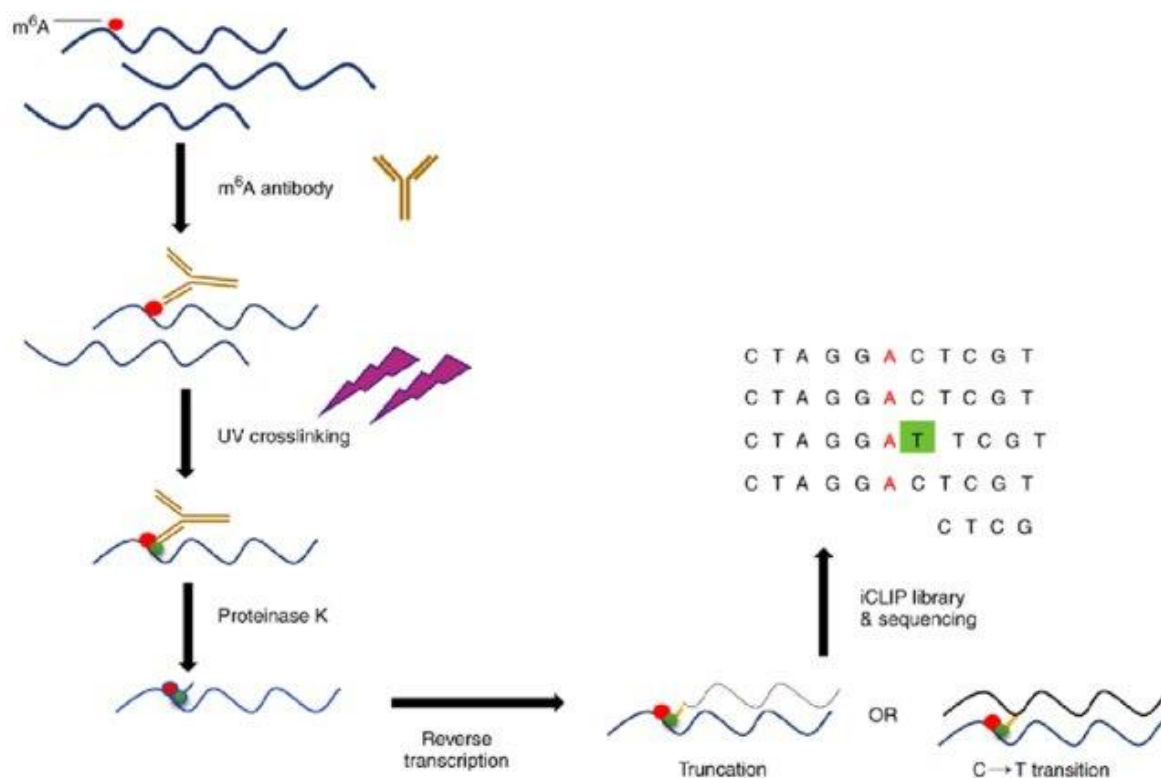
Metody epitranskriptomiky

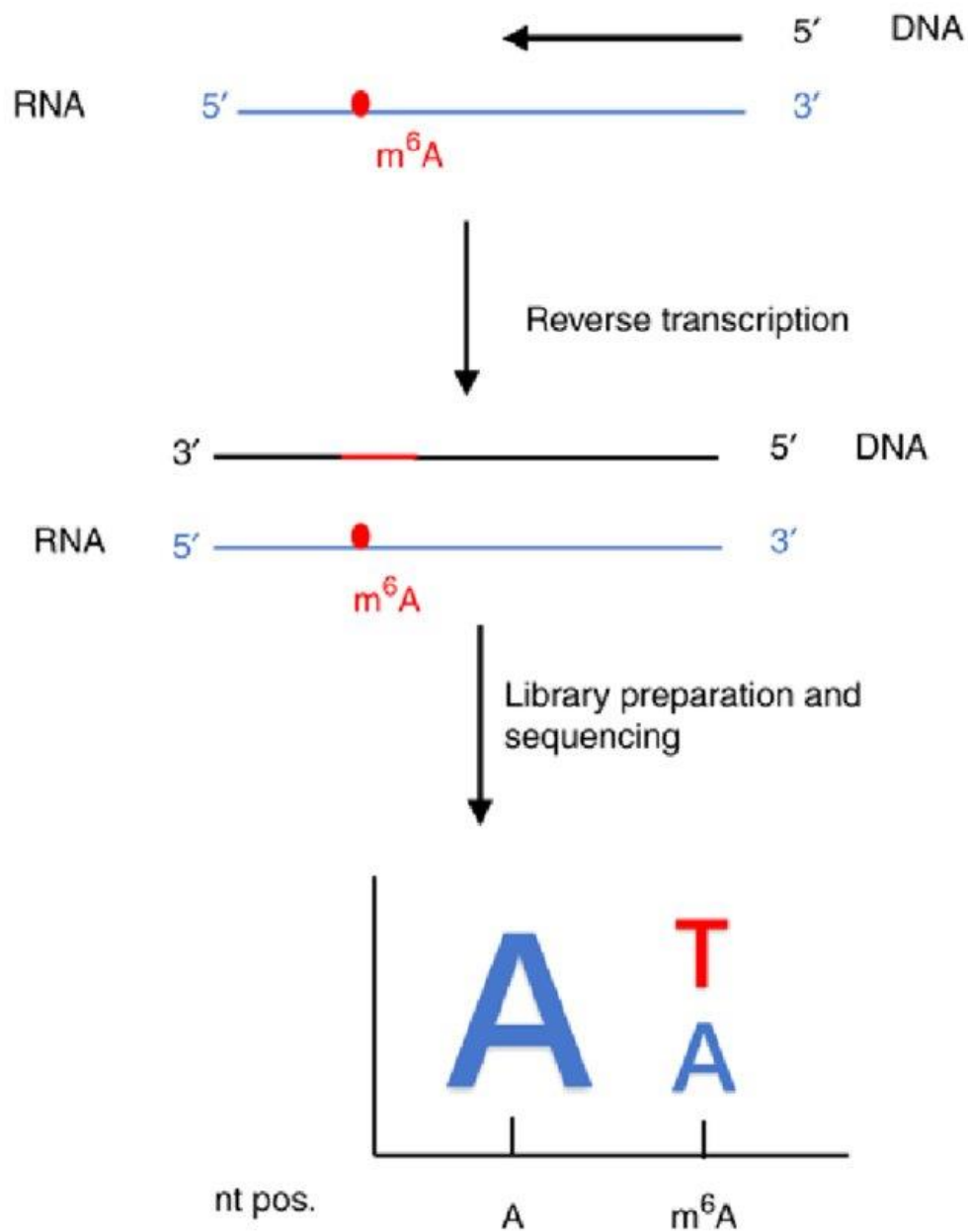
HPLC spřažená s triple-quadrupole hmotnostní spektrometrií (**LC-MS/MS**) – poskytuje **kvantitativní info o zastoupení různých modifikací**, ale není vhodná pro identifikaci a lokalizaci jednotlivých modifikací.

IP asociovaná s TGS: **m⁶A-seq** neboli **MeRIP-seq**, obecně i pro jiné modifikace. Nevýhoda: neurčí přesnou polohu modifikace. Lze provést i ve variantě dot-blot.

Vylepšení: **miCLIP** – UV-indukovaný crosslink protilátky k IP fragmentům RNA.

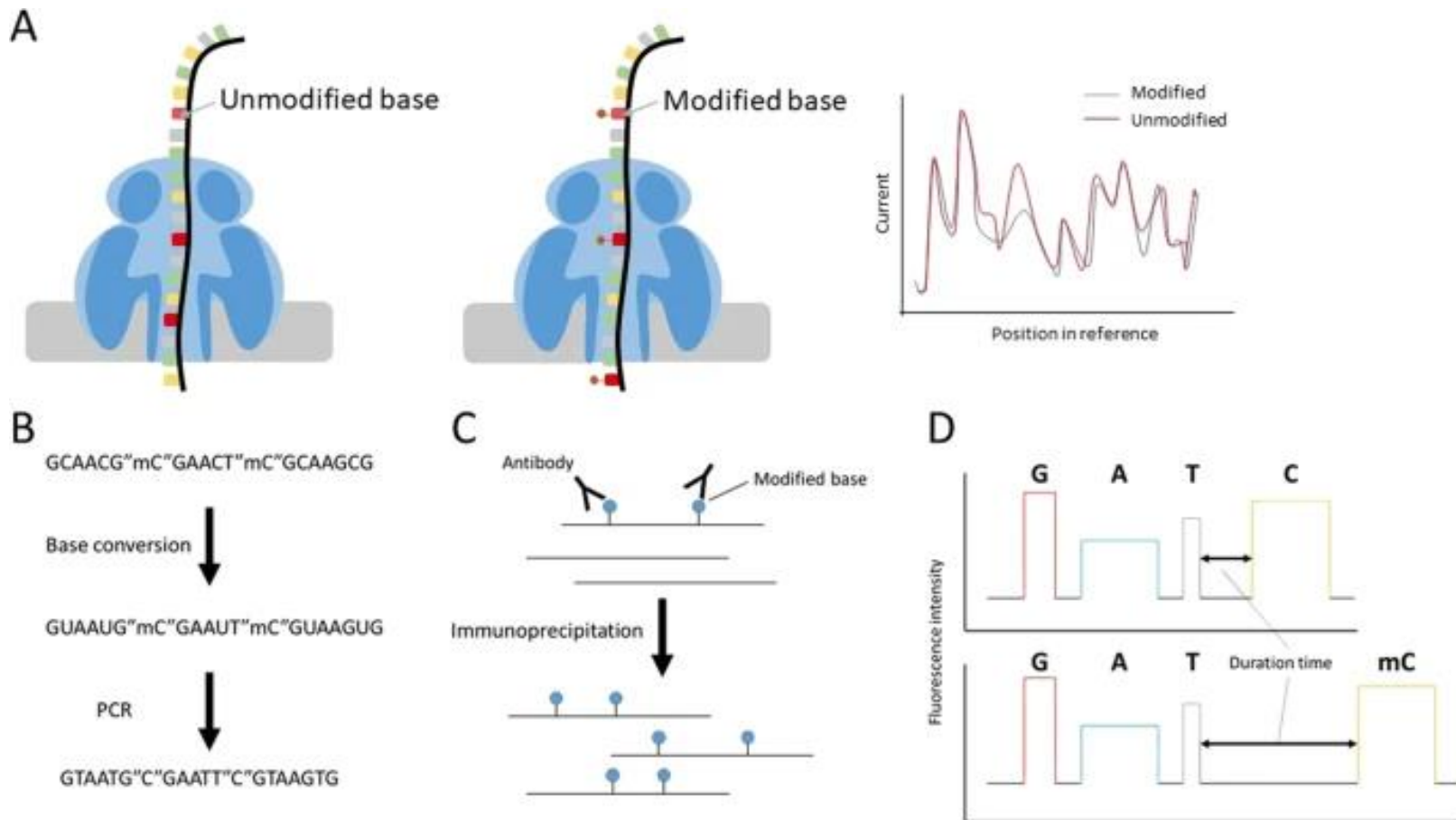
Po částečném štěpení komplexu proteinázou K se provádí reverzní transkripce, která se v místě modifikace buď zastaví, nebo dojde k zařazení chybné báze.





Přímá detekce bez IP, speciální reverzní transkriptáza zařadí proti m⁶A buď A nebo T, připraví se sekvenační knihovna a hodnotí pozice mutací.

Nejslibnější se v současnosti jeví využití přímého **ONT sekvenování nativní RNA**



Detekce modifikovaných bází s využitím Nanopore sekvenování (**A**), pomocí bisulfitové konverze a PCR (a následně sekvenování) (**B**), imunoprecipitace NA (**C**), a SMRT sekvenování (**D**)

Závěr:

Seznámili jsme se s metodami analýzy struktury chromatinu a jeho epigenetických značek (methylace DNA, PTM histonů, poloha nukleozomů, 3D organizace chromatinu)

Mezi významné epigenetické hráče patří dále **exprese nekódujících RNA**, která např. spouští RDDM (*RNA-directed DNA methylation*) a další procesy transkripční a posttranskripční regulace genů. Metodicky se ale jedná o RNA-Seq (viz přednáška o sekvenačních metodách).

Tyto analýzy od 90. let setrvale nabývají na významu vzhledem k možnosti objasnění mechanismů regulace genové exprese, vzniku chromozomálních translokací, objasnění „přeprogramování“ aktivity genů během vývoje organismu/diferenciace buněk, při patologických procesech....

Klesající ceny a možnost NGS přístupů formou kvalifikovaného servisu firem nebo CF činí celogenomové analýzy podstatně dostupnější, úzkým hrdlem je ale stále **bioinformatická analýza dat**.

Zběžně jsme se seznámili s expandující oblastí EPITRANSKRIPTOMIKY – metodami analýzy modifikací RNA (asi 170 různých), které ovlivňují např. její stabilitu, lokalizaci, 3D strukturu, sestřih...