

MOŻLIWOŚCI SEKWENCJONOWANIA NASTĘPNEJ GENERACJI (NGS)

Monika Gos

Zakład Genetyki Medycznej

Instytut Matki i Dziecka

Warsztaty: Sekwencjonowanie następnej generacji – warsztaty
dla pracowników laboratoriów, 23-24.03.2022



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska



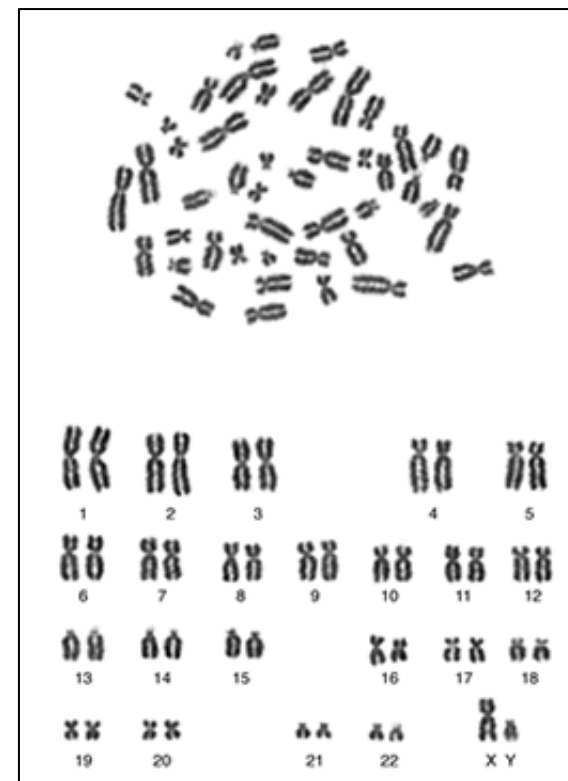
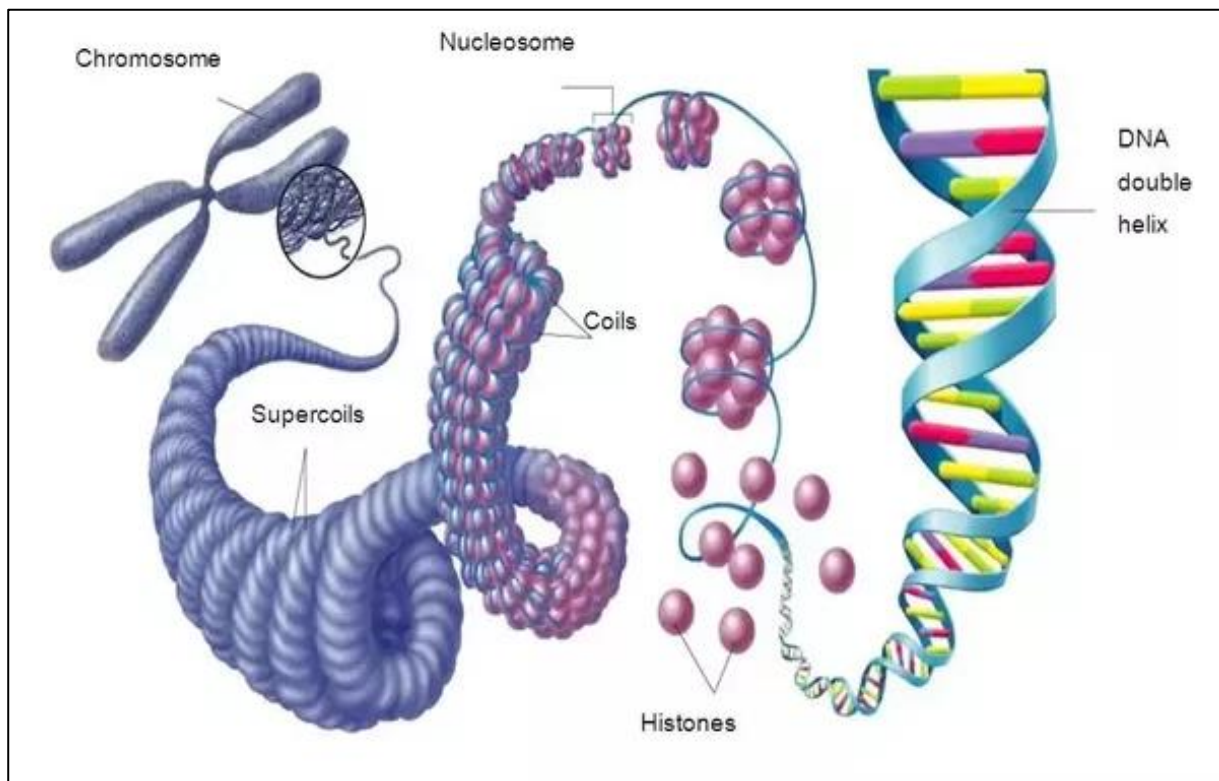
Instytut
Matki i Dziecka

Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl

W każdej (prawie) komórce znajduje się taka sama informacja genetyczna (DNA)



GENOM: 3 miliardy par zasad (pz) - 46 chromosomów (22 autosomy + 2 chr. płciowe)
≈ 2m DNA / komórka

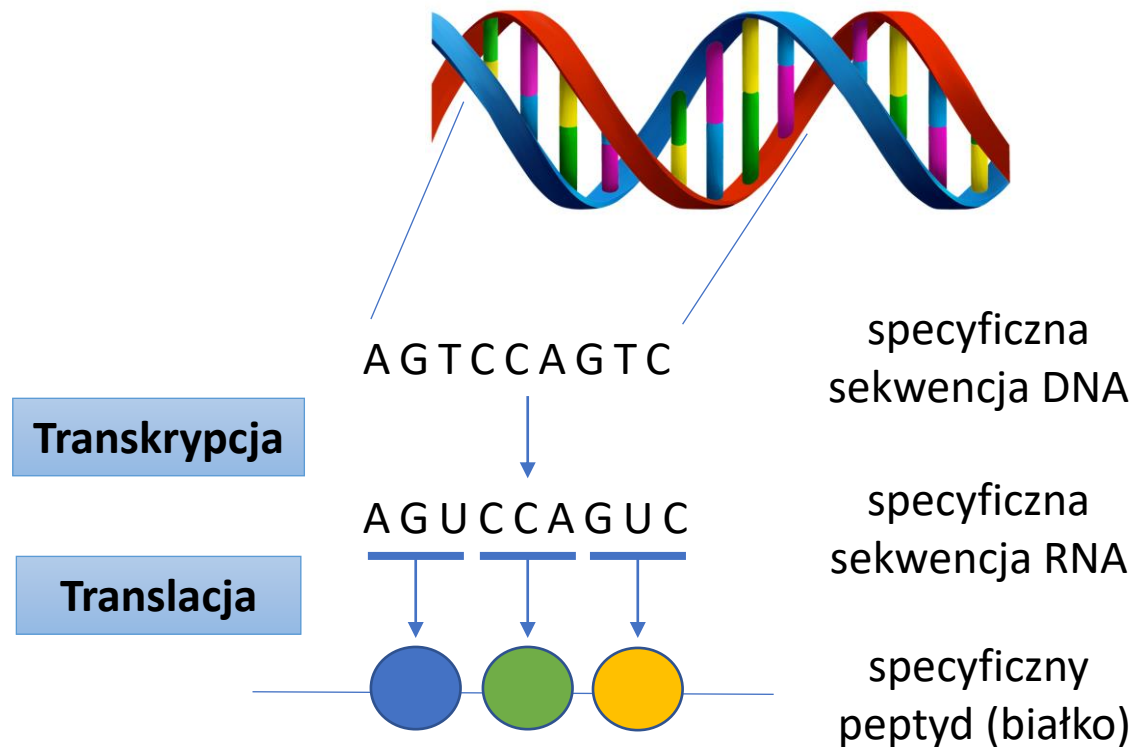
2.0×10^{13} m DNA we wszystkich komórkach (70x słońce-ziemia-słońce)

Upakowanie DNA – rdzeń histonowy – 146bp

GEN – fragment DNA (lub u niektórych wirusów RNA) kodujący określone białko lub RNA.
W bardziej potocznym rozumieniu gen jest odcinkiem DNA warunkującym wystąpienie określonej cechy organizmu.

W jaki sposób 4 zasady azotowe A, T, G, C tworzą
organizm składający się z 3.72×10^{13} komórek?

KOD GENETYCZNY – system zapisu informacji genetycznej
oparty na kombinacji 3 zasad azotowych (64 potencjalne kombinacje)

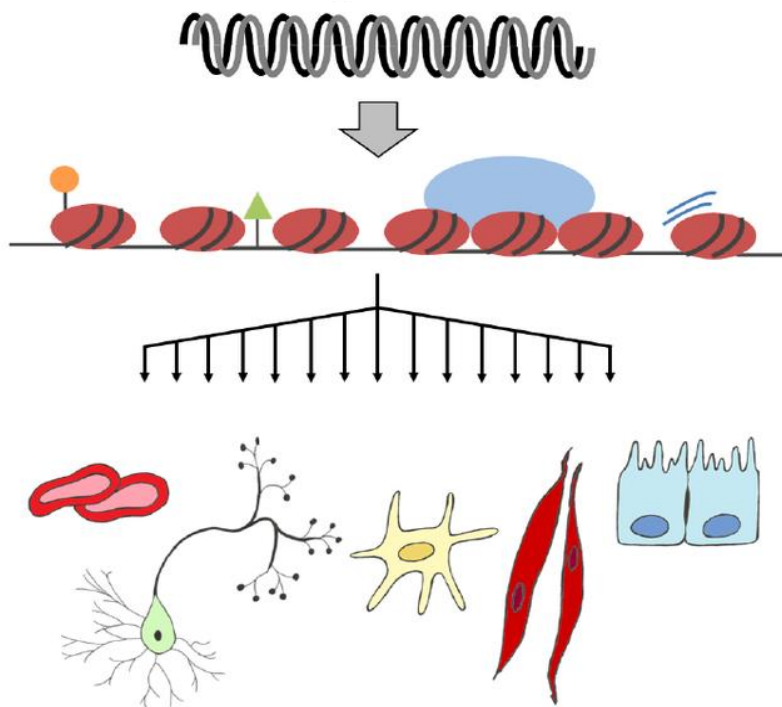


Jeśli w każdej komórce mamy taką samą informację genetyczną,
to dlaczego każda komórka jest inna?

Różna ekspresja genów (transkrypcja) – czasowo- i tkankowospecyficzna

Regulacja epigenetyczna – stopniowe wyciszanie ekspresji genów w trakcie rozwoju

EPIGENOM – metylacja DNA, modyfikacje białek histonowych, chromatyna (eu- i hetetro-)



Omiki

dziedziny nauki zajmujące się zbiorczą analizą jakościową i ilościową cząstek biologicznych, co ma przełożyć się na stworzenie całościowego obrazu działania organizmu (jego struktury, funkcji i dynamiki funkcjonowania)

EPIGENOM

modyfikacje
epigenetyczne



GENOM

TRANSKRYPTOM

PROTEOM

METABOLOM

DNA

RNA

białko

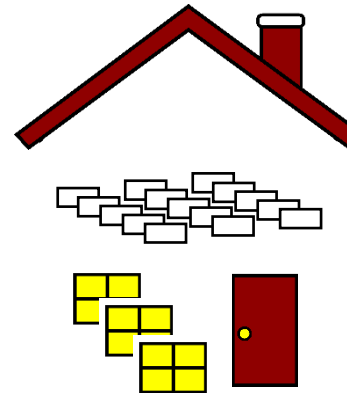
organizm



PROJEKT
„dom”

- cegła
- dach
- okna
- drzwi

ZAMÓWIENIE



PRODUKCJA



PRODUKT



Fundusze
Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska



Instytut
Matki i Dziecka

Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl

Genom: 3 000 000 000 pz (3×10^9) – 25 000 genów

Średnia wielkość genu: $16,3 \times 10^3$ pz

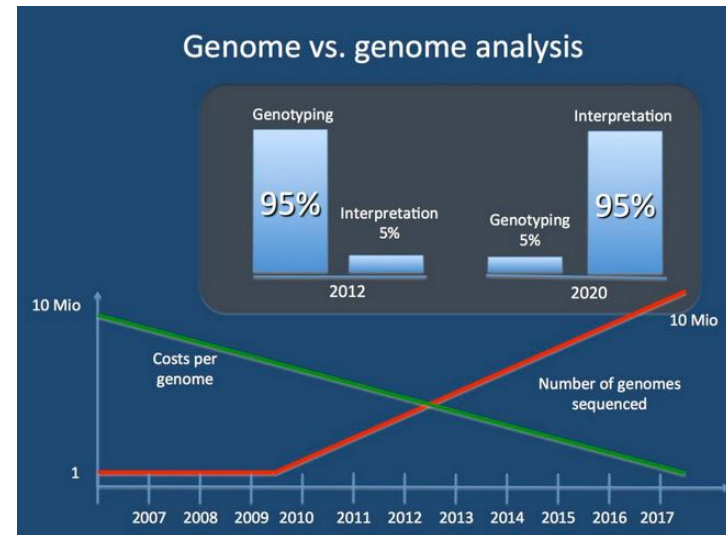
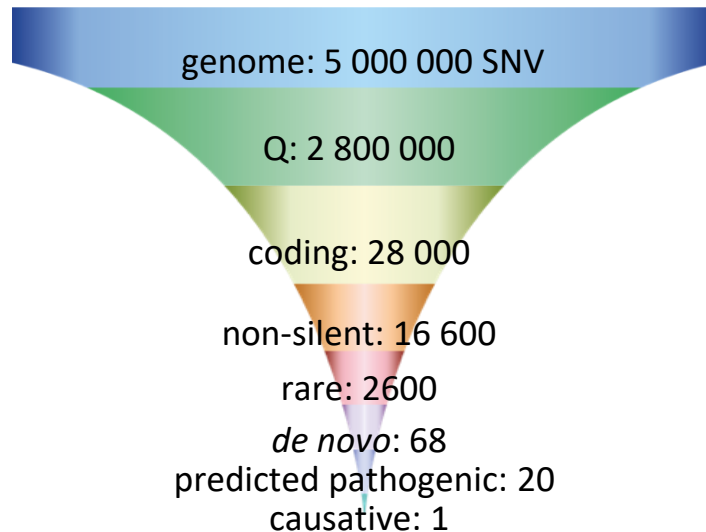
x liczba genów = $4,07 \times 10^8$

1.5% genomu – geny unikatowe, kodujące białka, RNA (**EKSOM**)

>95% genomu – DNA niekodujący (m.in. sekwencje powtórzone)

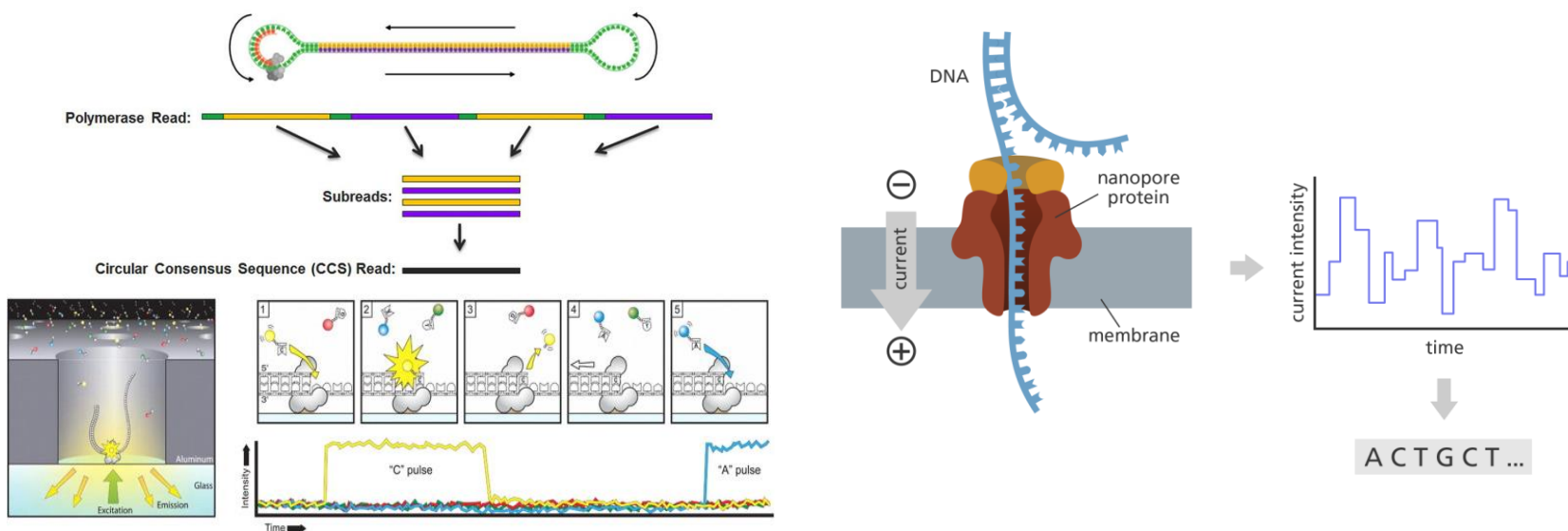
GENOMIKA – najbardziej statyczny materiał
(nie dotyczy mutacji somatycznych)

Sekwencjonowanie genomów *de novo* (*de novo assembly*)
Resekwencjonowanie (mapowanie do sekwencji referencyjnej)
→ Identyfikacja wariantów powiązanych z fenotypem



Wady „krótkich” odczytów

- Problemy z sekwencjonowaniem regionów bogatych w pary GC
- Słaba rozdzielczość w regionach „kompleksowych” np. MHC
- Regiony powtórzone / paralogi – problemy z uliniowaniem sekwencji
- Brak możliwości identyfikacji rearanżacji chromosomowych (rozległe delecje/duplikacje; warianty strukturalne)
- Brak możliwości „fazowania” wariantów



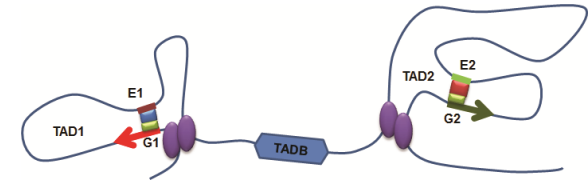
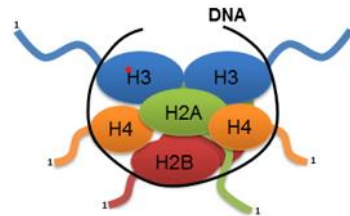
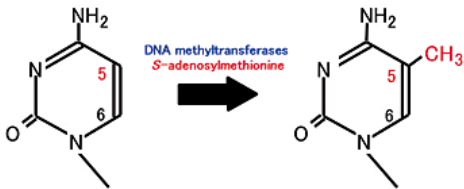
Epigenomika

MECHANIZMY REGULACJI EPIGENETYCZNEJ

metylacja:
(wyspy CpG, zahamowanie ekspresji)

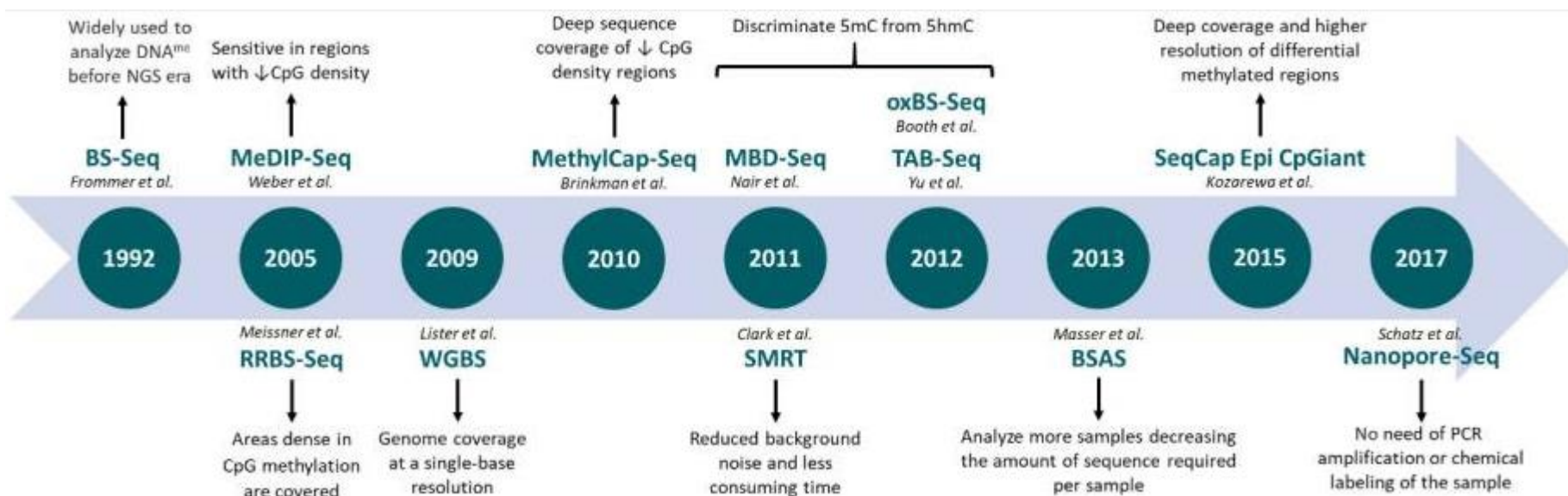
modyfikacje białek histonowych:
acetylacja, metylacja,
fosforylacja, ubikwitynacja (efekt
zależny od miejsca i charakteru
modyfikacji)

struktura chromatyny:
gęstość upakowania chromatyny
(eu- i heterochromatyna –
nieaktywna transkrypcyjnie)
struktury przestrzenne tworzone
przez chromatynę (pętle)
oddziaływanie z kompleksami
białkowymi



Analiza metylacji

wdrożenie NGS – możliwość rezygnacji z macierzy DNA



Genes (Basel). 2018 Sep; 9(9): 429.



Fundusze Europejskie
Wiedza Edukacja Rozwój



Rzeczpospolita
Polska



Instytut
Matki i Dziecka

Unia Europejska
Europejski Fundusz Społeczny



edu-metgen.imid.med.pl

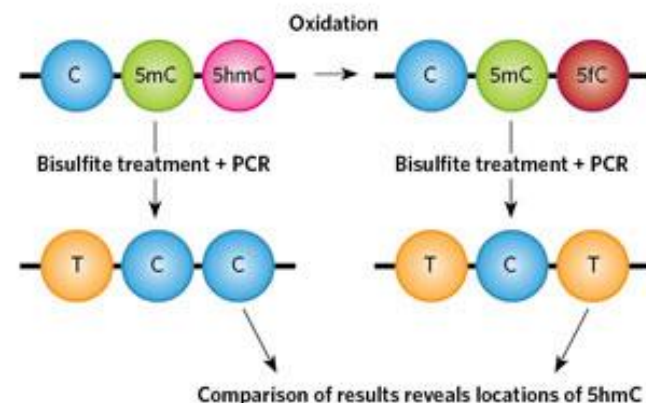
Analiza metylacji

- Metody oparte o powinowactwo do metylowanego DNA – wykorzystanie przeciwciał (methylated DNA immunoprecipitation / MeDIP-Seq) lub białek wiążących metylowane DNA (MBD-Seq) do selekcji metylowanych fragmentów DNA – dobre do analizy regionów bogatych w pary GC (wyspy CpG, promotory); brak możliwości oceny ilościowej
- metody oparte o enzymy restrykcyjne np. MspI (CCGG) – do sekwencjonowania nadają się tylko krótkie (40-220pz) fragmenty DNA (zależność od sekwencji) - RRBS
- metody oparte o zastosowanie białek mających powinowactwo do metylowanego DNA – MethylCap-Seq – MBD-MECP2

Genes (Basel). 2018 Sep; 9(9): 429.

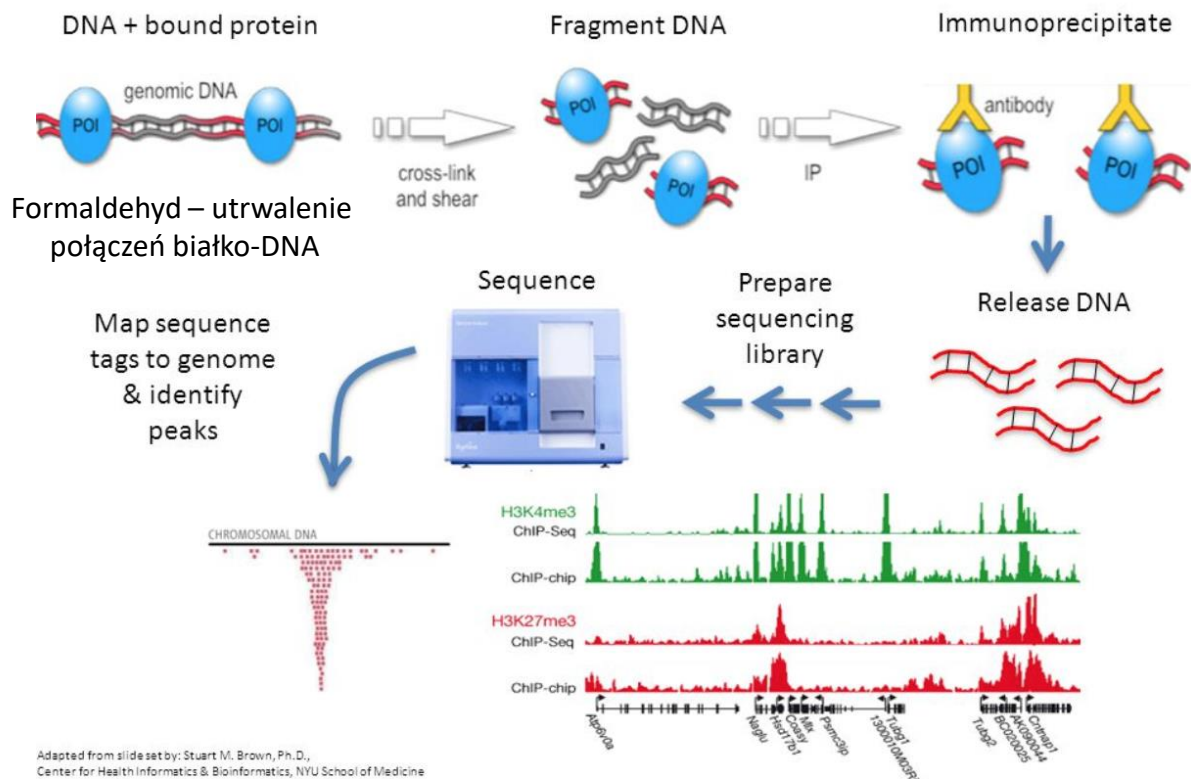
Analiza metylacji

- Metody oparte o modyfikację DNA kwaśnym dwusiarczynem sodowym – BS-Seq – konwersja C→T; konieczność wykorzystania innych narzędzi do przyrównania sekwencji; najbardziej informacyjna i najdroższa technologia (WGBS); modyfikacja: bisulfite amplicon sequencing (BSAS)
- Metody oparte o modyfikację DNA dwusiarczynem sodowym z dodatkowym etapem utleniającym – Ox-BS
 - możliwość odróżnienia 5mC od 5hmC
- Metody BS-Seq z enzymem TET (TAB-Seq)
 - 5mC → 5caC → BS → U/T
 - 5hmC → BS → C



ChIP-Seq

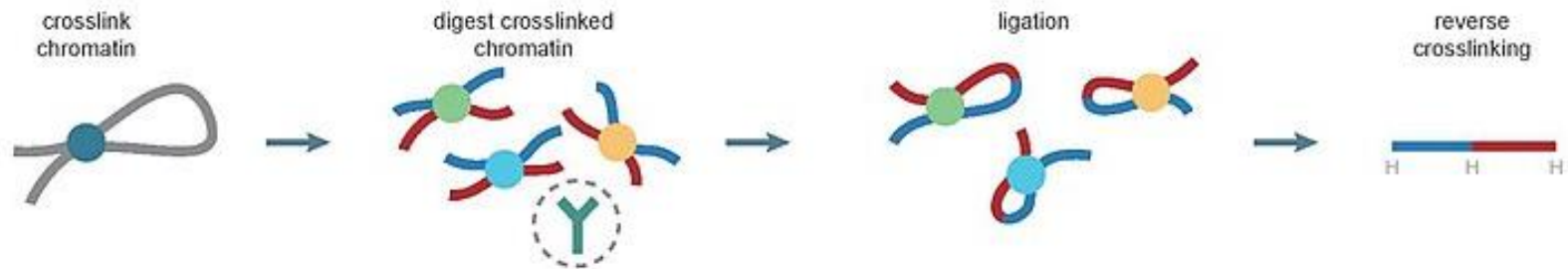
Badanie
oddziaływań
DNA-białko –
mapowanie
modyfikacji
białek
histonowych
i miejsc wiązania
czynników
transkrypcyjnych



Adapted from slide set by: Stuart M. Brown, Ph.D.,
Center for Health Informatics & Bioinformatics, NYU School of Medicine

<https://noorsiddiqui.com/what-is-chip-seq-atac-seq/>

Badanie konformacji chromatyny

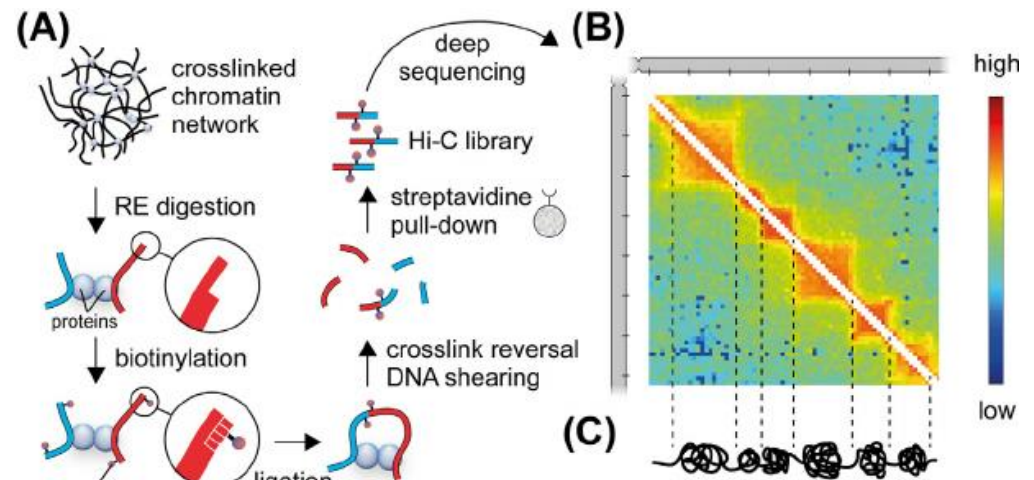


3C – badanie oddziaływań pomiędzy 2 loci np. promotor + enhancer

4C – badanie oddziaływań pomiędzy 1 locus a innymi

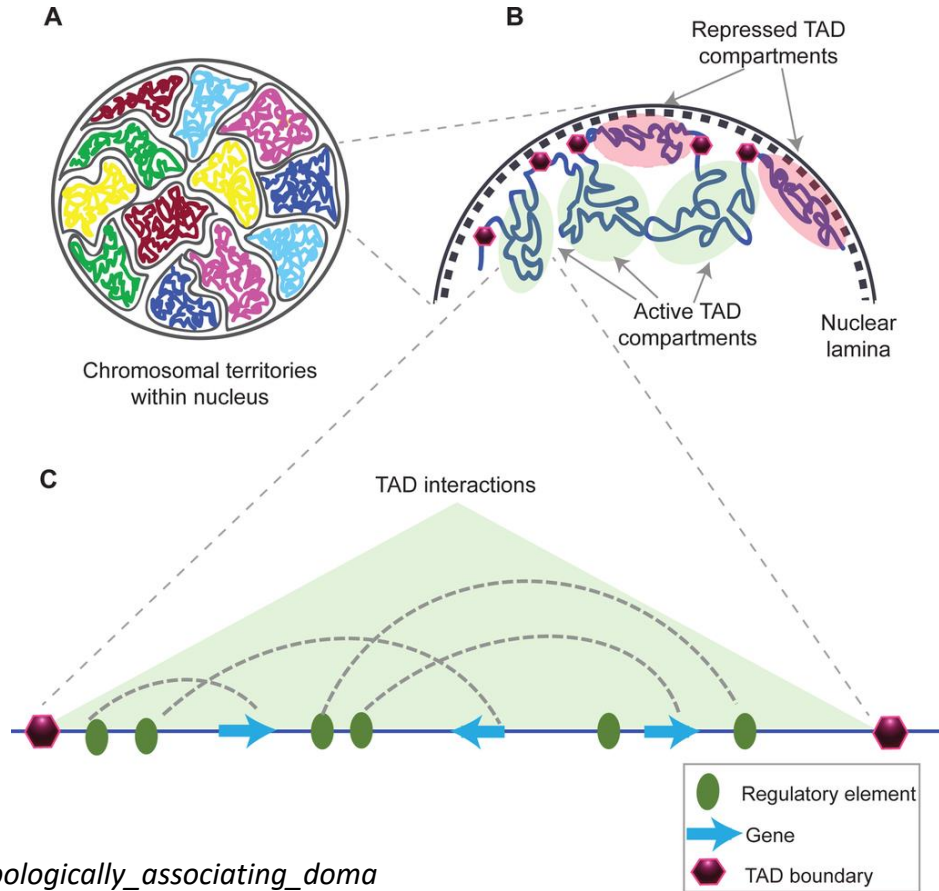
5C – badanie wszystkich oddziaływań w danym regionie

HiC – badanie wszystkich oddziaływań za pomocą NGS



DOI: 10.1016/bs.ircmb.2014.11.004

Badanie konformacji chromatyny



https://en.wikipedia.org/wiki/Topologically_associating_domains#/media/File:Structural_organization_of_chromatin.png

ATAC-Seq, DNase-seq, FAIRE-Seq

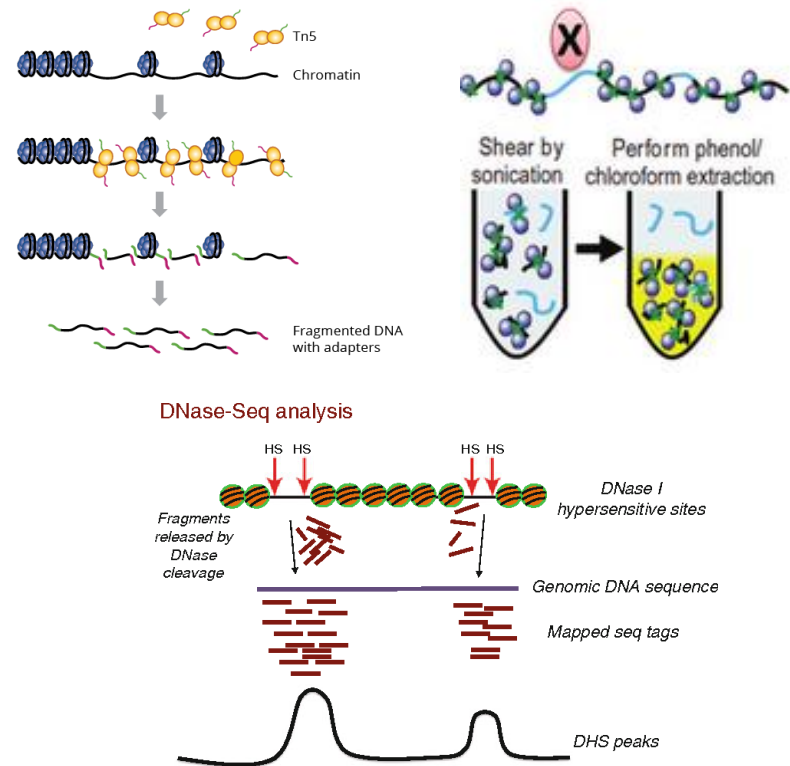
Badanie struktury chromatyny pod względem jej dostępności dla białek (regiony promotorowe, enhancery, elementy regulatorowe *in cis*)

ATAC-Seq – transpozaza

DNase-Seq – DNAza I

FAIRE-Seq – formaldehyd

Mnase-Seq - nukleosomy

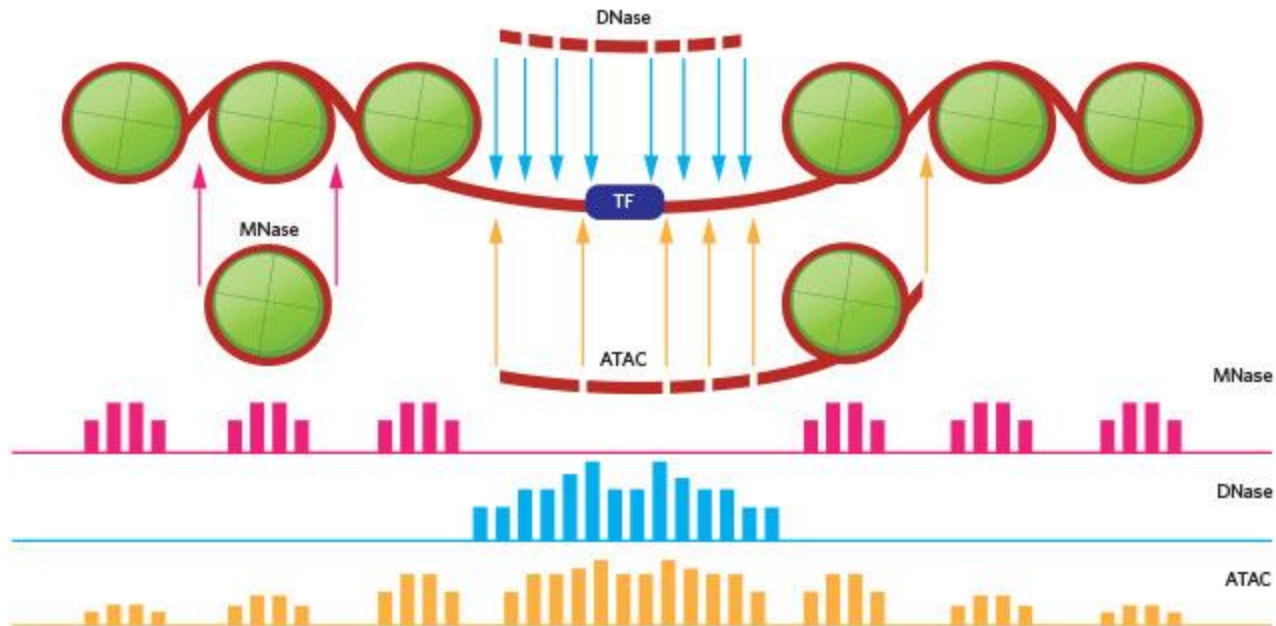


<https://www.genewiz.com/en-GB/Public/Services/Next-Generation-Sequencing/Epigenomics/ATAC-Seq/>

DOI: 10.1101/sqb.2012.76.010587

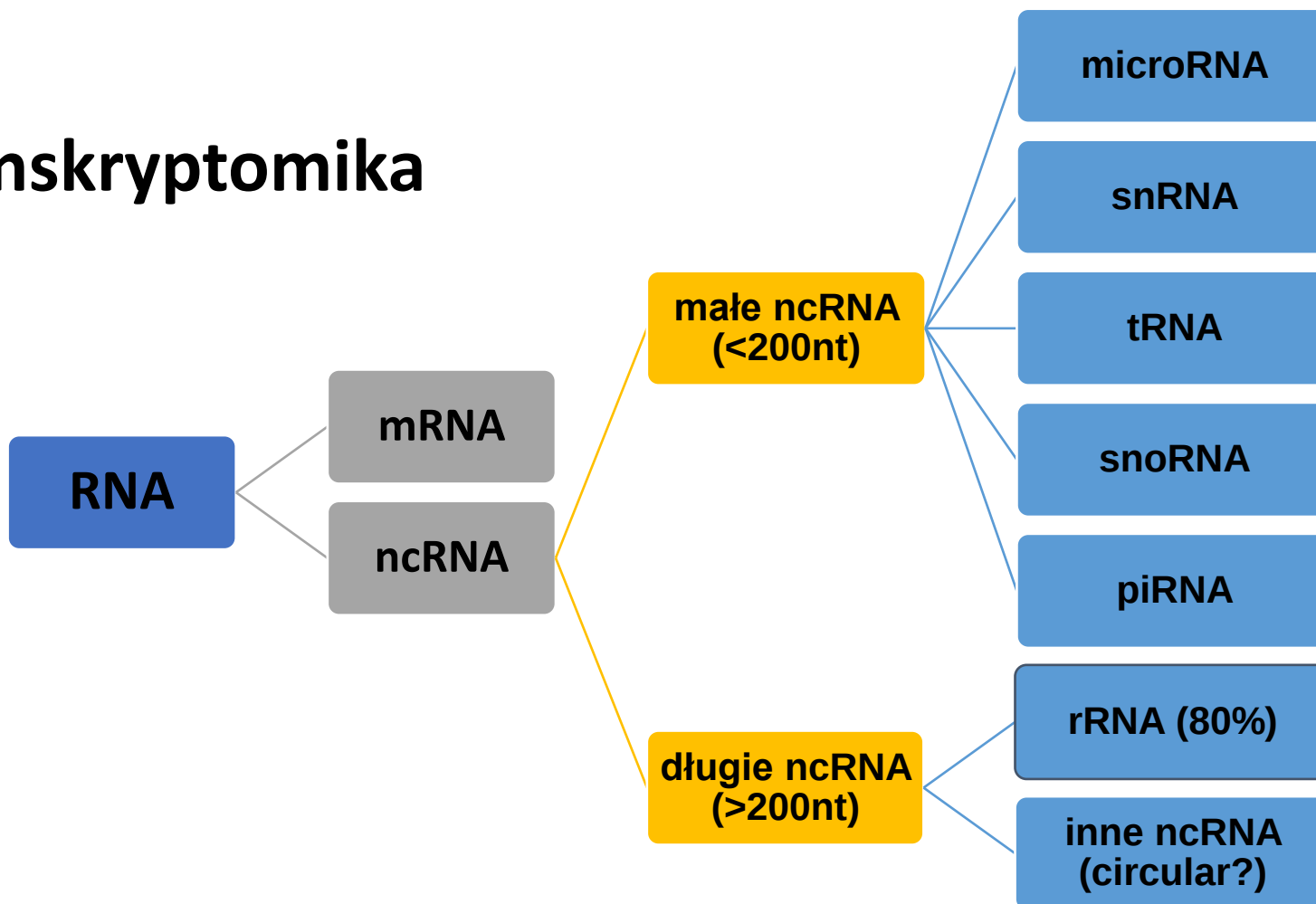
doi: 10.1007/978-1-62703-284-1_2.

Integracja danych



<https://www.the-scientist.com/lab-tools/reveling-in-the-revealed-34261>

Transkryptomika



Podejścia umożliwiające analizę poszczególnych klas RNA

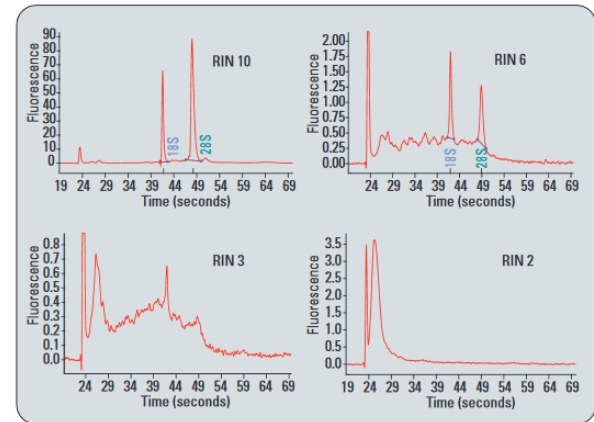
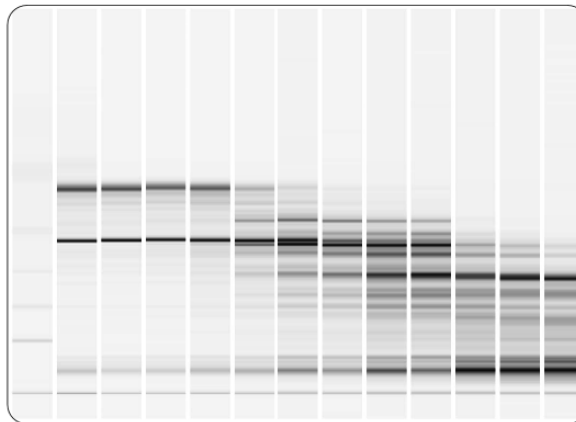
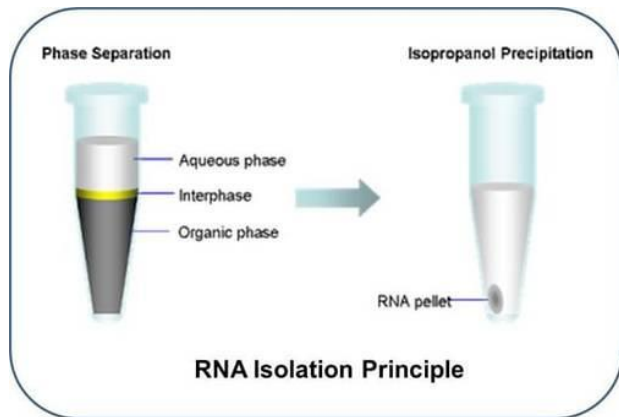
- usunięcie rRNA z puli analizowanych cząsteczek / selekcja fragmentów
- usunięcie mRNA, które występują w nadmiarze np. b-globina

Transkryptomika

- Badanie RNA (poszczególnych klas, lub wszystkich cząsteczek RNA)
- Badania jakościowe (obecność alternatywnych transkryptów)
- Badania ilościowe (określenie poziomu ekspresji poszczególnych genów)
- Wcześniej: analiza ekspresji z wykorzystaniem mikromacierzy – głównie analiza mRNA (specyficzne oligonukleotydy)
- Aktualnie RNA-seq z wykorzystaniem sekwencjonowania następnej generacji
- Zalety: możliwość analizy wszystkich klas RNA lub wybranych, możliwość analizy izoform (również zaburzeń składania transkryptów i genów fuzyjnych), możliwość analizy cząsteczek RNA niezależnie od poziomu ich ekspresji; mniejsza ilość RNA do analizy

Transkryptomika

- Izolacja RNA – kolumny, izolacja na piechotę
- RNA szybciej ulega degradacji – konieczność stosowania preparatów zapobiegających jego degradacji (RNAlater, PAX blood tubes)
- Jakość RNA – dobra (RNA Integrity Number – RIN, Agilent)

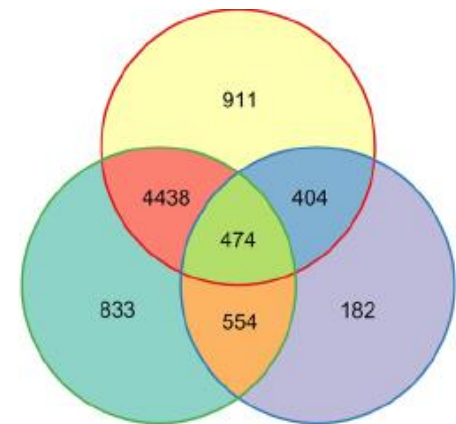
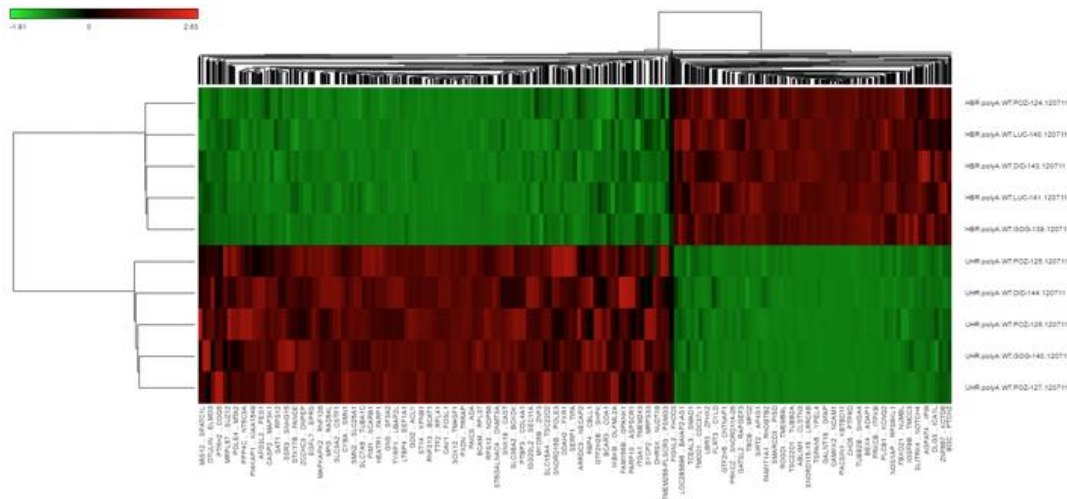


Ze względu na niestabilność RNA...

- Obligatoryjny etap przygotowania bibliotek: odwrotna transkrypcja – przepisanie RNA na cDNA (heksamery, oligo dT)
- Otrzymujemy stabilny materiał, który poddawany jest dalszemu procesowi (sekwencjonowanie, analiza bioinformatyczna) w taki sam sposób jak przy sekwencjonowaniu DNA
- Możemy sekwencjonować wszystkie cząsteczki RNA lub zastosować podejście celowane (specyficzne startery do amplifikacji lub sondy do hybrydyzacji)
- Procedury wzbogacania:
 - 1) selekcja transkryptów zawierających poliA
 - 2) deplecja rRNA (80%)
 - 3) deplecja beta-globiny

Czemu ma służyć prowadzona analiza?

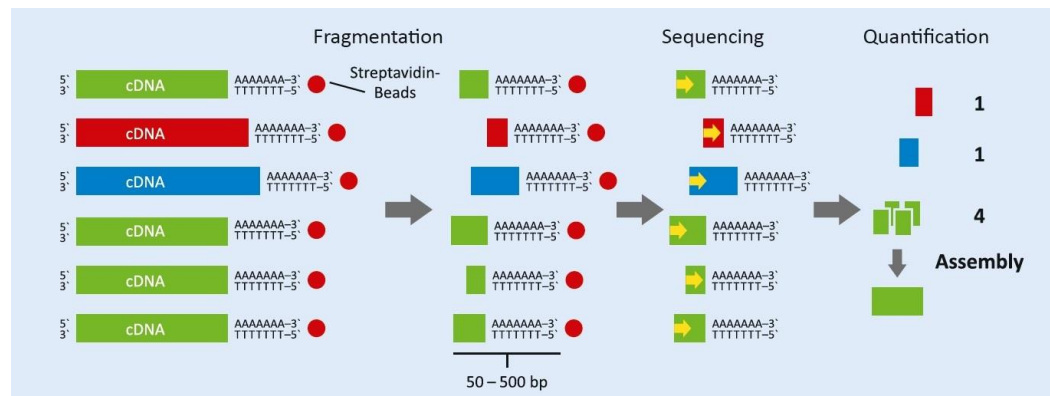
- Ilościowa – badanie różnic w ekspresji genów
- Jakościowa – identyfikacja nowych transkryptów



Techniki analizy RNA i ich specyficzne wykorzystanie

Metody do analizy ekspresji genów

- jeden fragment analizowany reprezentuje ekspresję całego genu
- metoda podatna na wpływ takich czynników jak ligacja adapterów, sposób fragmentacji; duża zmienność
- ekonomiczne – mniej odczytów niezbędnych do przeprowadzenia analizy
- brak możliwości analizy izoform / wariantów splicingowych



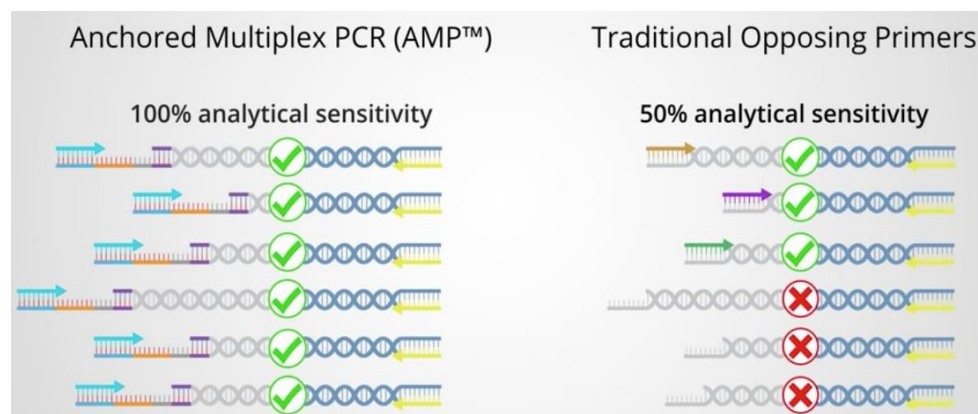
Techniki analizy RNA i ich specyficzne wykorzystanie

Metody do analizy alternatywnego splicingu i obecności genów fuzyjnych

Geny wieloeksonowe podlegają alternatywnemu splicingowi (izoformy) – zaburzenia procesu mogą skutkować wystąpieniem określonego fenotypu

Geny fuzyjne – powstające w wyniku rearanżacji chromosomowych – 20% nowotworów (konieczność identyfikacji tzw. fusion junctions, co jest możliwe jedynie w długich odczytach lub na poziomie RNA)

- Możliwa jest analiza celowana, ale trzeba znać sekwencje eksonów / intronów lub wiedzieć gdzie jest punkt złamania



ArcherDX

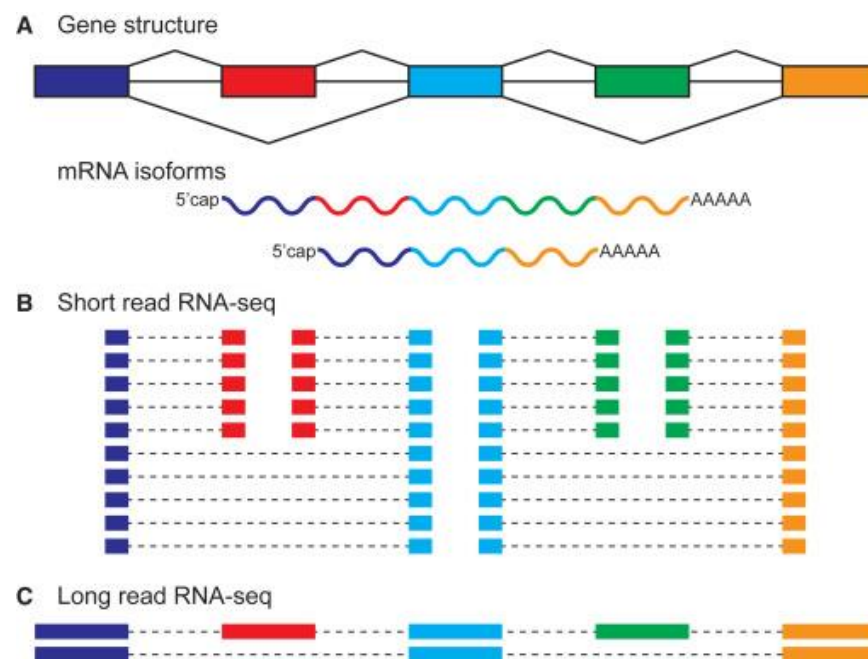
Techniki analizy RNA i ich specyficzne wykorzystanie

Metody do analizy alternatywnego splicingu i obecności genów fuzyjnych

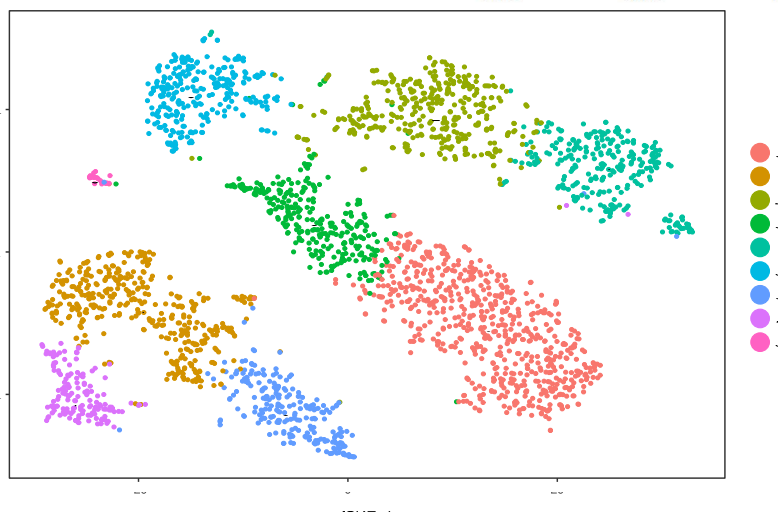
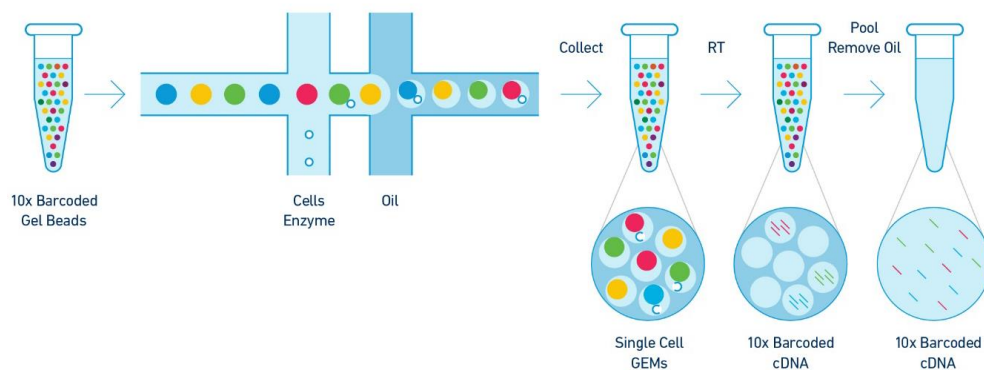
Alternatywy:

sekwencjonowanie transkryptów
od początku do końca
PacBio (SMRT) lub Oxford Nanopore

badanie obecności zaburzeń składania
RNA na podstawie danych uzyskanych
z krótkich odczytów: specjalne progra-
my do analizy wykorzystujące genom/
transkryptom referencyjny, w którym
określone są konstytutywne miejsca
składania RNA



Single cell sequencing (10x Genomics)



Integracja danych

