



Zaawansowane Metody Analizy Danych w Biologii Molekularnej

Semestr letni 2018

- 
- **Regulacja ekspresji genów**
 - **Sekwencjonowanie nowej generacji (NGS)**

Semestr letni 2018

Podumowanie transkrypcji

Animacje:

file:///Users/mjdabrowski/Movies/Animations/ch12_transcription.html

file:///Users/mjdabrowski/Movies/Animations/ch12_euk_trans_term.html



Regulacja ekspresji genów

- Jak spośród wielu tysięcy genów wyróżnić ten, który ma ulec ekspresji
- Jakie elementy informacji genetycznej ulegają różnicowaniu?

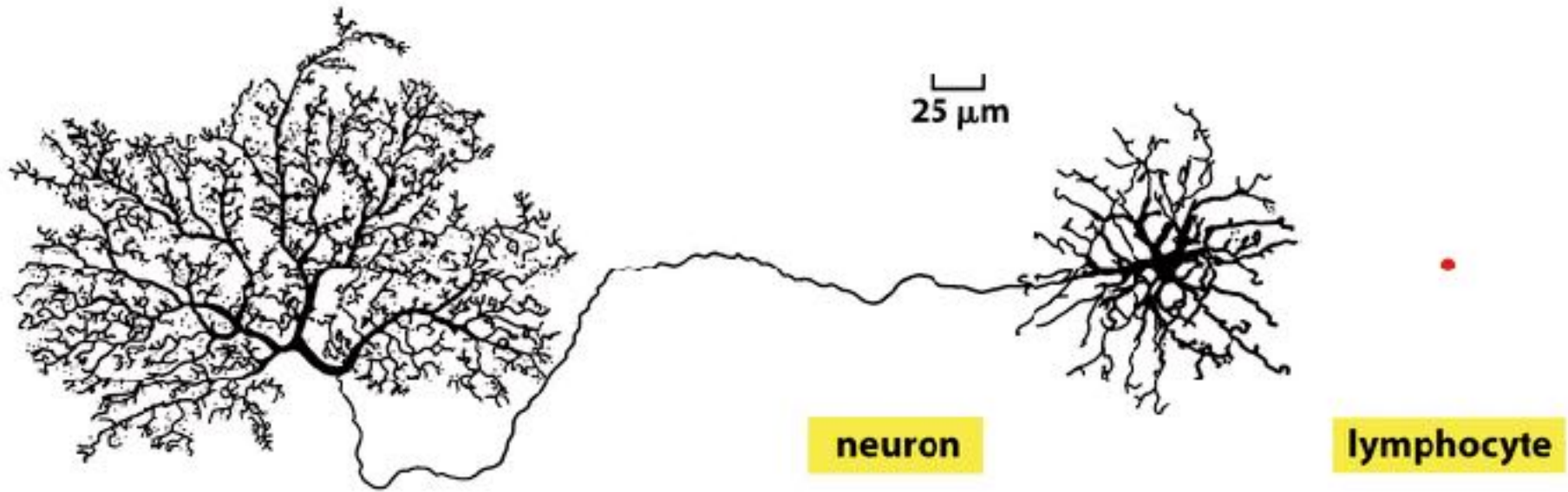


Figure 7-1 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Informacja genetyczna

Dowód na to, że każdej komórce ciała jest informacja genetyczna niezbędna do utworzenia całego organizmu.

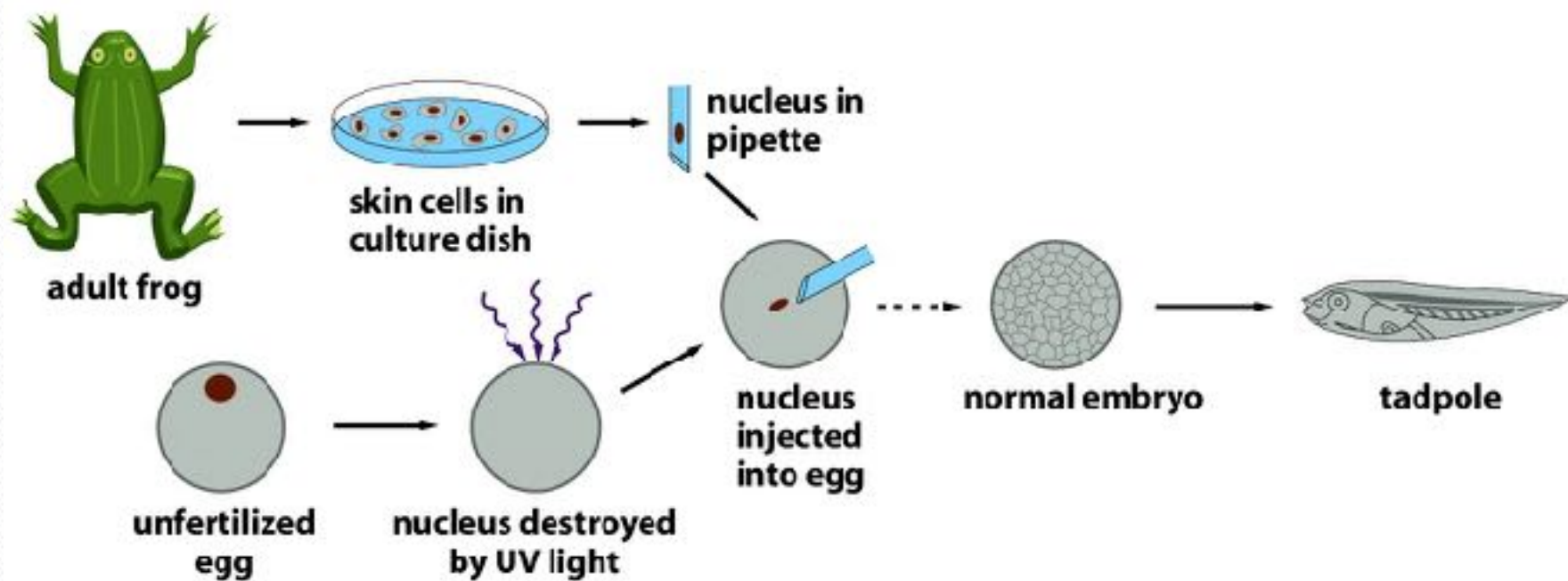


Figure 7-2a Molecular Biology of the Cell 5/e (© Garland Science 2008)

- W trakcie rozwoju dochodzi do różnicowania się – powodem jest produkcja i akumulacja różnego rodzaju RNA i białek
- Pomimo zróżnicowania, pewne cechy wspólne – geny/białka dbające o utrzymanie domu (house keeping genes/proteins)

Zróżnicowana odpowiedź komórek na sygnały z zewnątrz

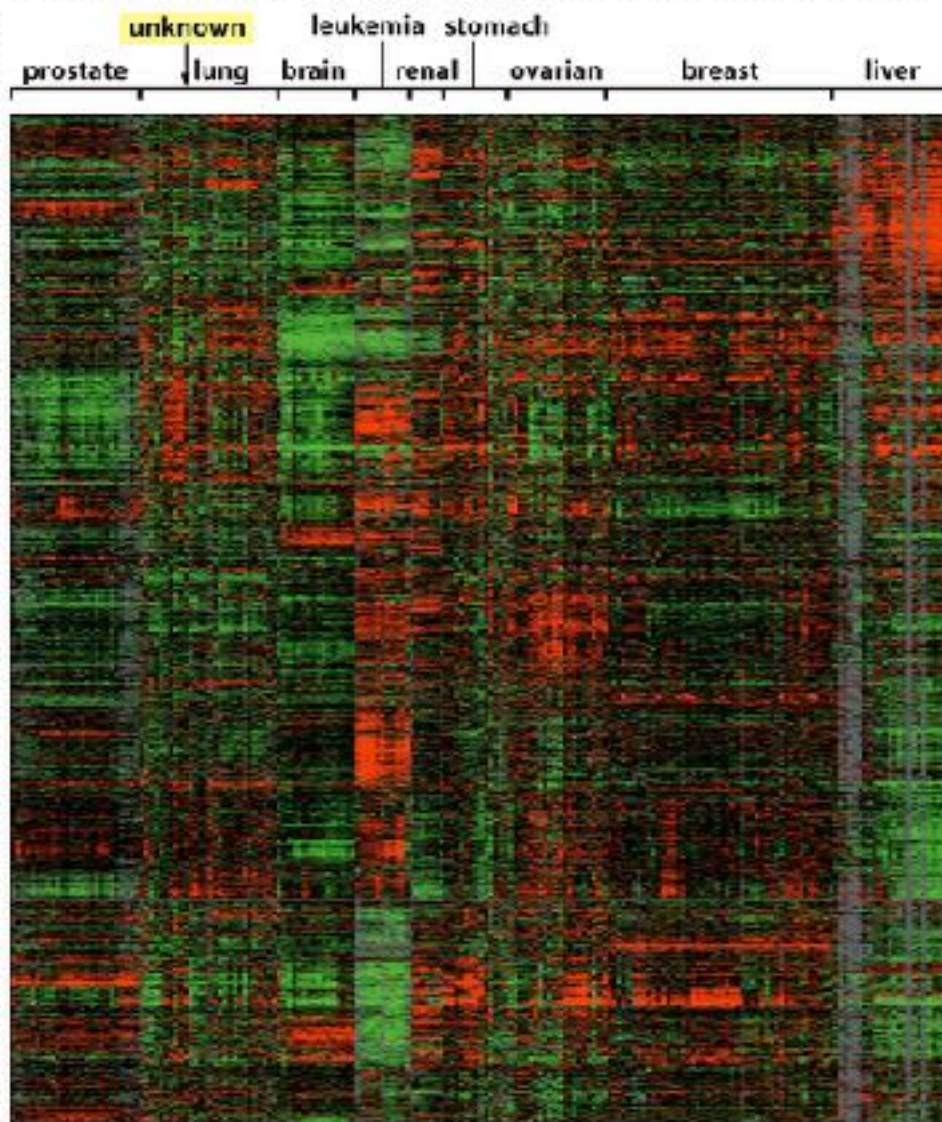


Figure 7-3 Molecular Biology of the Cell 5/e (© Garland Science 2008)

- Ekspresja w różnych typach nowotworów
- Różnice w ekspresji genów pomiędzy różnymi typami komórek (zielony – istotnie niższa ekspresja, czerwony – istotnie wyższa ekspresja, ciemny/czarny – poziom ekspresji zbliżony do średniej)



Regulacja ekspresji genów

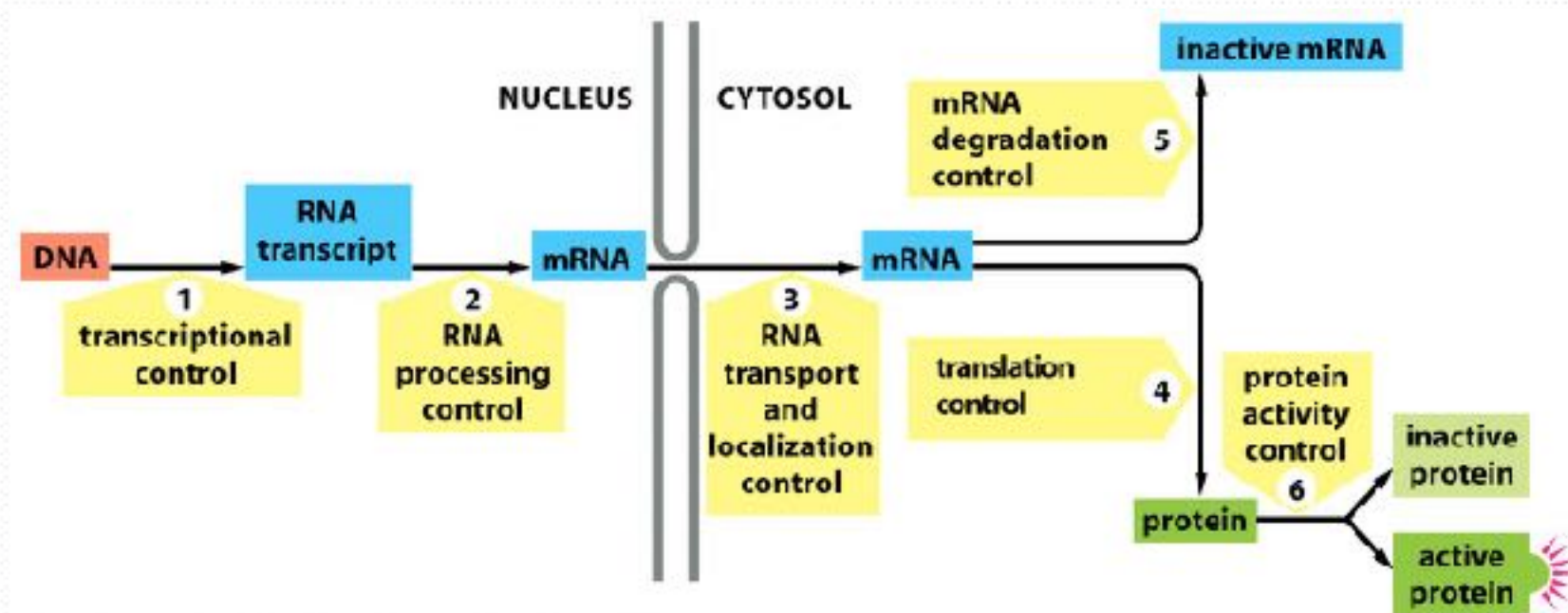
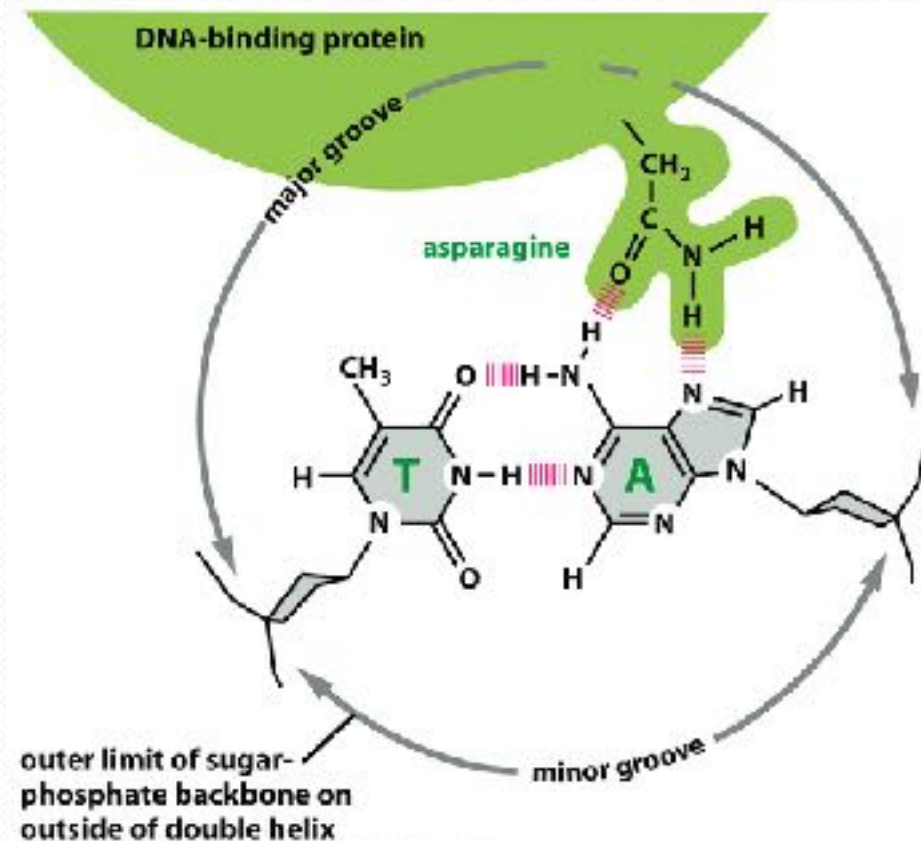


Figure 7-5 Molecular Biology of the Cell 5/e (© Garland Science 2008)

- 1-kontrola transkrypcji
- 2-kontrola dojrzewania RNA
- 3-kontrola na poziomie transportu przez por jądrowy (w otoczce jądrowej)
- 4-kontrola translacji
- 5-kontrola degradacji mRNA
- 6-kontrola aktywności białka

Regulacja transkrypcji – przede wszystkim na poziomie inicjacji

- Specyficzne sekwencje regulatorowe DNA
- Specyficzne białka rozpoznające – dopasowujące się, do krótkich sekwencji DNA (~20pz) – nazywamy czynnikami transkrypcyjnymi (TF – transcription factors)



Połączenie między białkiem a jedną parą zasad DNA. Zwykle zakres oddziaływania to 10-20 wiązań niekowalencyjnych, z których każde angażuje inny aminokwas białka.

Figure 7-9 Molecular Biology of the Cell 6e (© Garland Science 2015)



Helix-turn-helix

- Pierwszy motyw białka wiążącego się do DNA miał strukturę helix-turn-helix
- Dwie helisy alfa połączone pojedynczą nicią kilku aminokwasów
- Skręt pomiędzy helisami utrzymywany jest dzięki oddziaływaniom

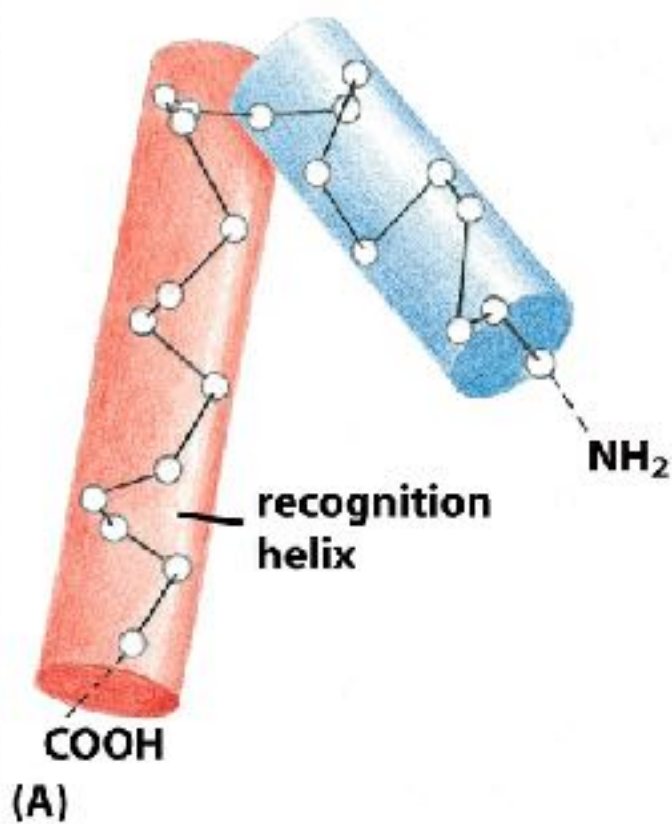


Figure 7-10 Molecular Biology of the Cell 5e (© Garland Science 2008)



- Helisa C-terminal – zwana – helisą rozpoznającą – pasuje ona do głównej bruzdy DNA (major groove)
- Dwie helisy alfa połączone pojedynczą nicią kilku aminokwasów

Palce cynkowe

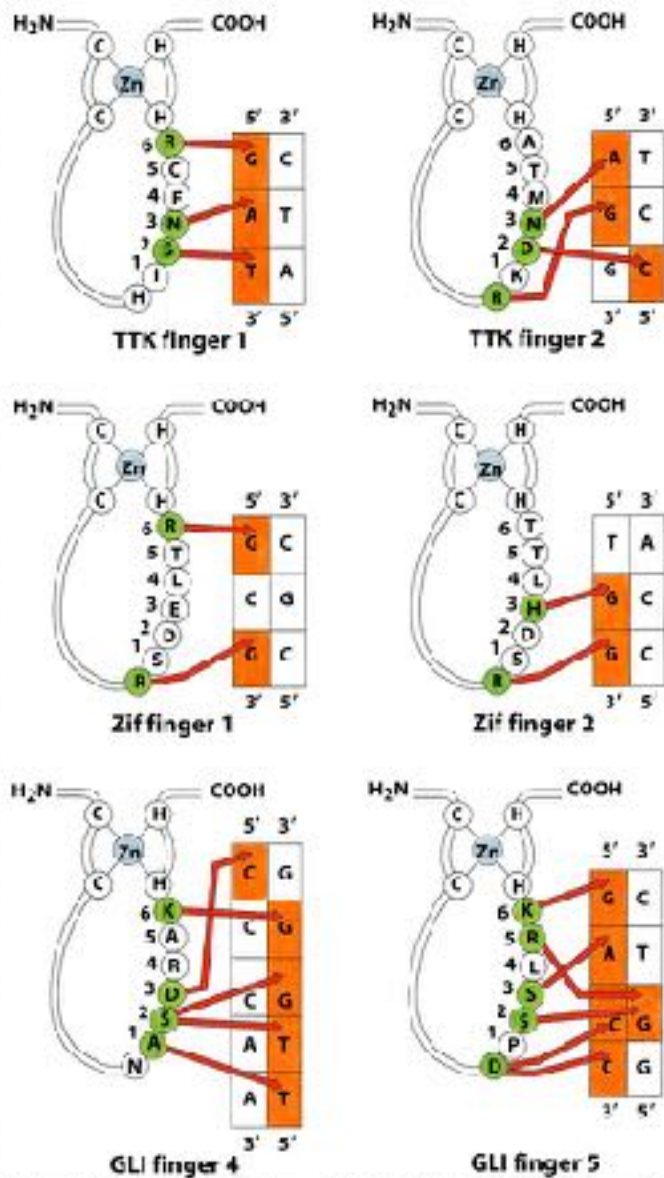


Figure 7-26 Molecular Biology of the Cell 5/e (© Garland Science 2008)

- Interakcje pomiędzy aminokwasami i specyficznymi fragmentami DNA



Poznanawanie miejsc wiązania na DNA

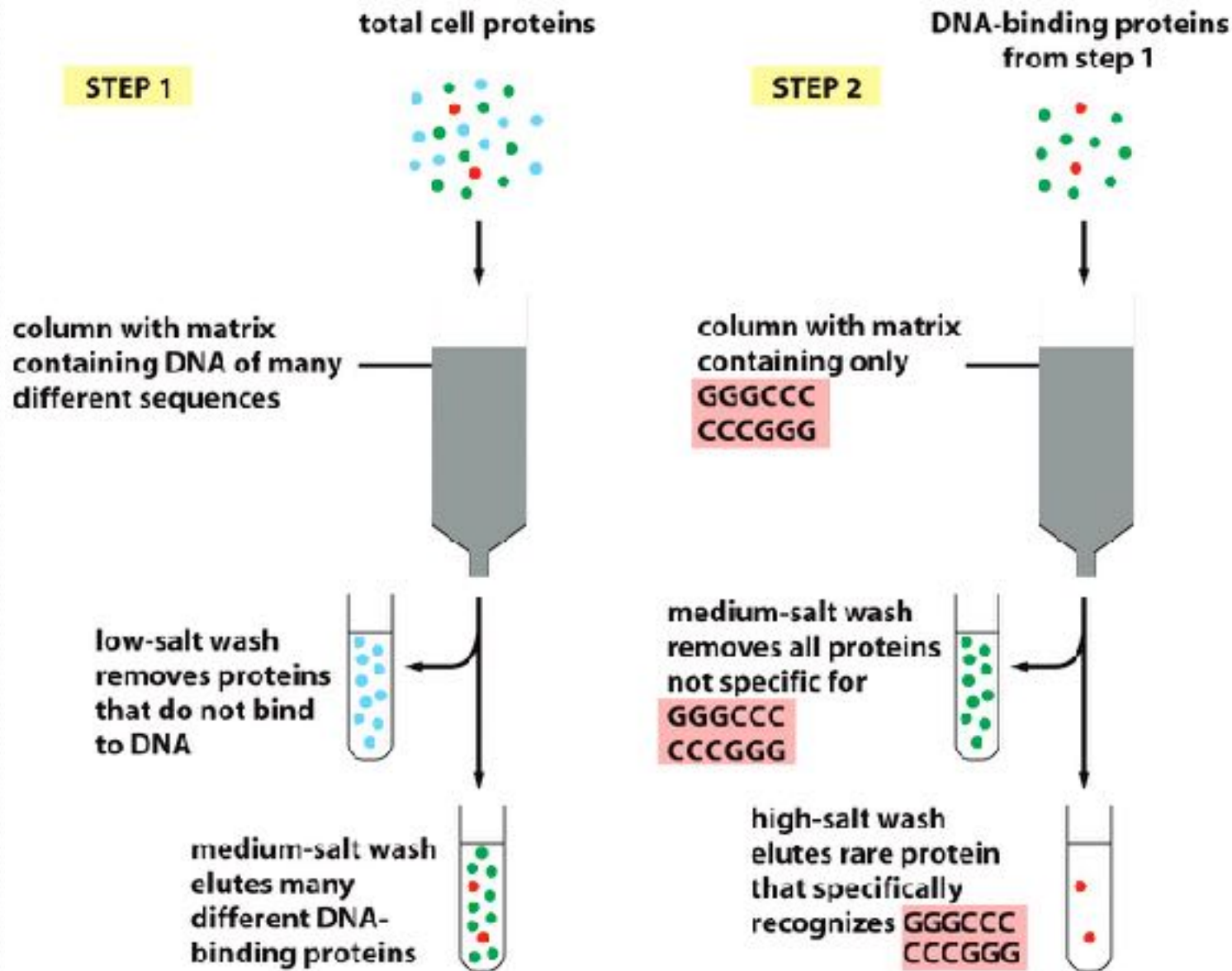


Figure 7-78 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Regulacja ekspresji genów

Geny wyłączane przez represory, a włączane przez aktywatory

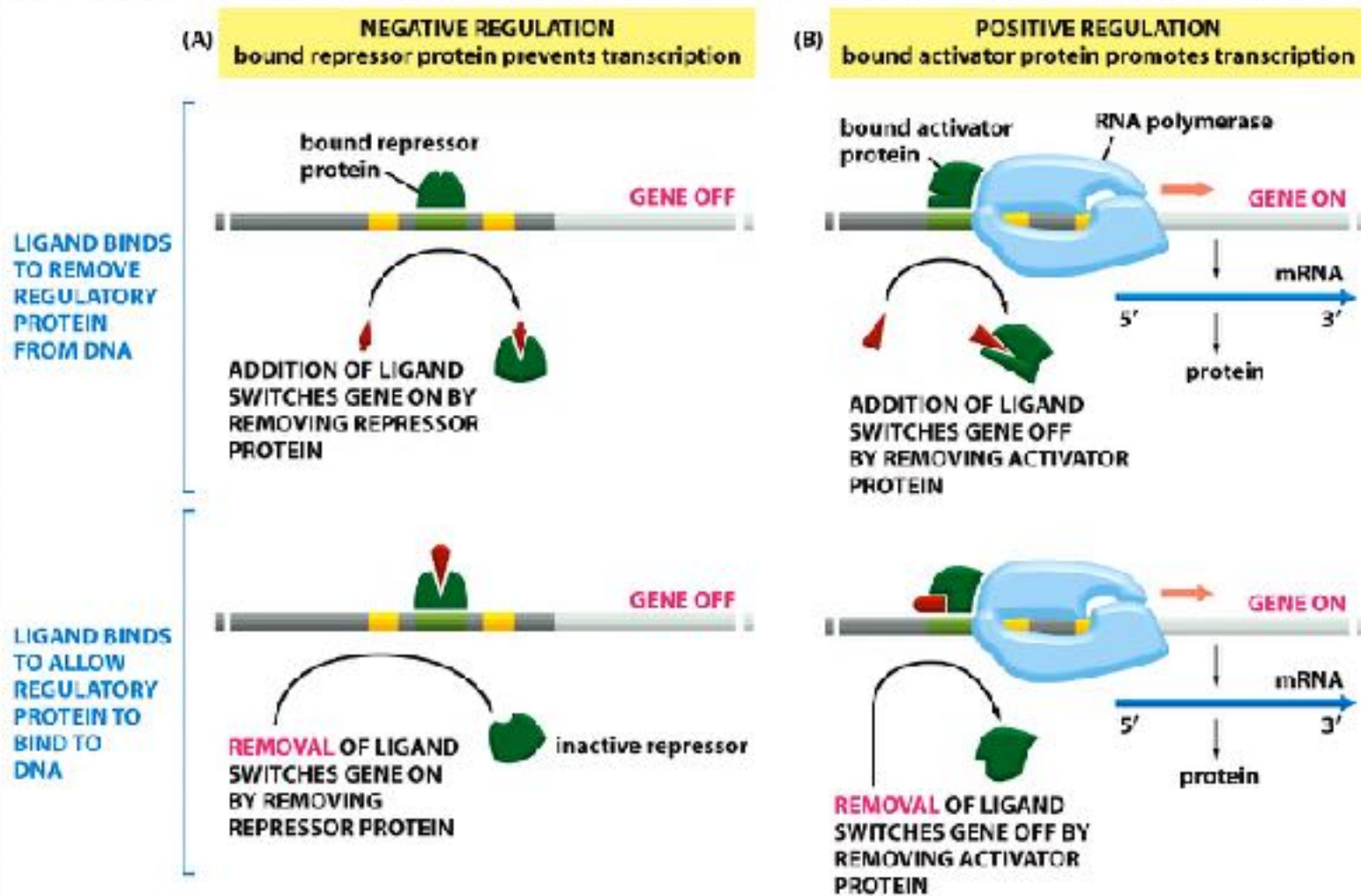


Figure 7-37 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Regulacja ekspresji genów

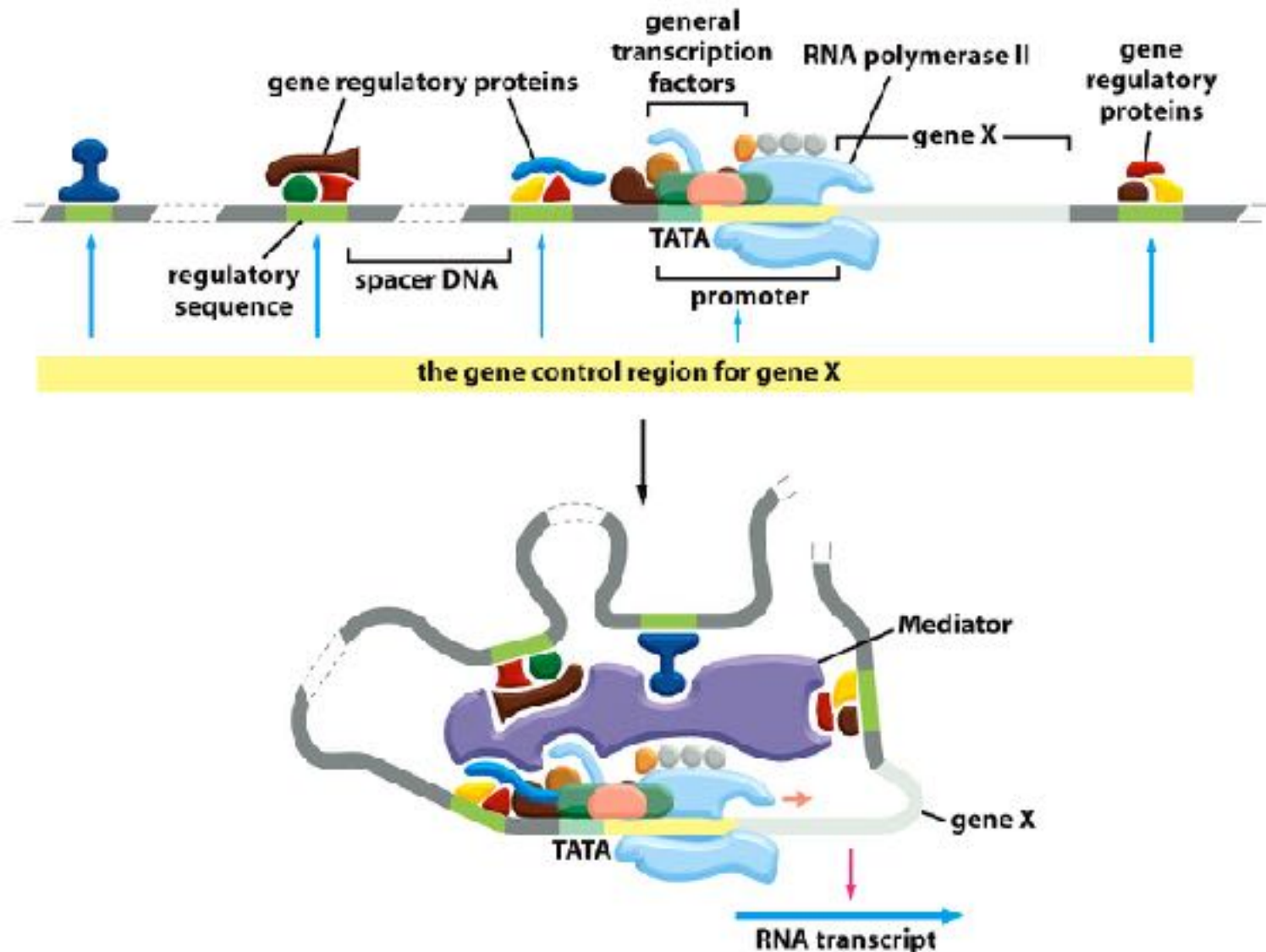


Figure 7-44 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Struktura chromatyny a ekspresja genów

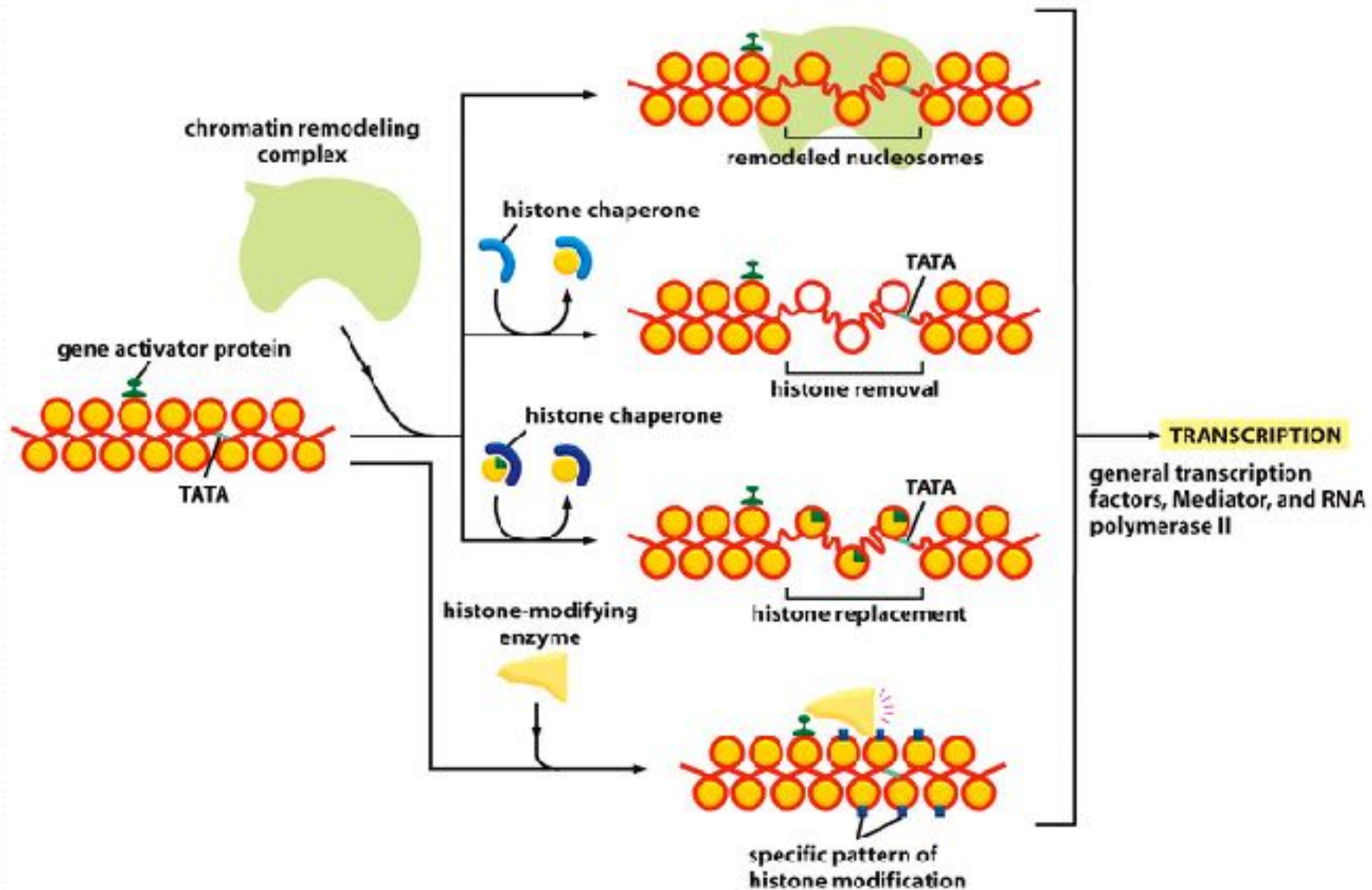


Figure 7 46 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Modyfikacje histonowe podczas transkrypcji

- Promotor ludzkiego genu: interferonu
- Białka aktywujące (aktywatory) wiążą się do DNA

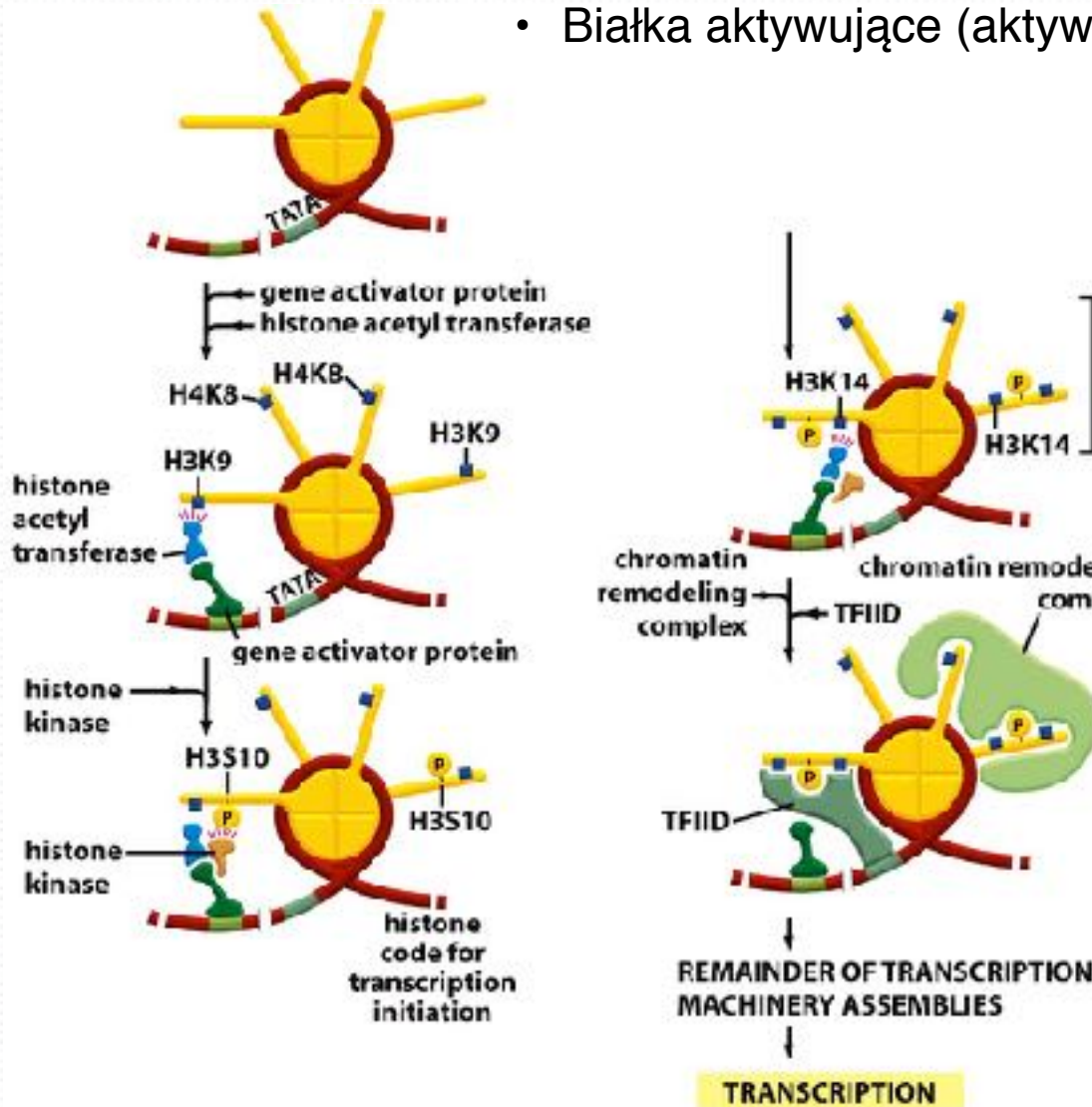


Figure 7-47 Molecular Biology of the Cell 5/e (© Garland Science 2008)

- Przyciągają enzymy acetylujące 9 lizynę histonu H3 i 8 lizynę histonu H4
- Aktywatory przyciągają enzymy fosforylujące serynę 10 histonu H3
- Dzieje się to tylko po wcześniejszej acetylacji
- Pojawienie się fosforylacji aktywuje acetylację kolejnej lizyny histonu H3
- Czynniki transkrypcyjne przyłączają się do acetylowanych fragmentów dzięki specyficznym domenom białkowym (bromodomains)

Różne typy oddziaływania represorów

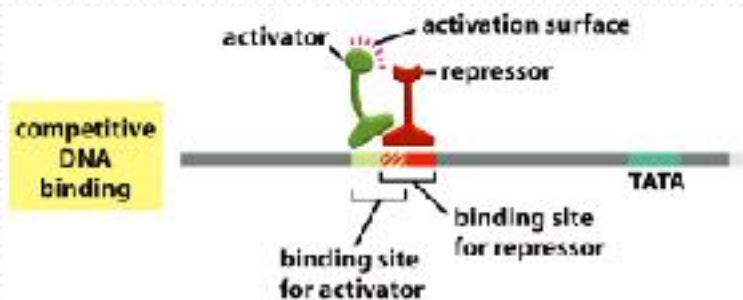


Figure 1-30a Molecular Biology of the Cell 6th Edition Garland Science 2008



Figure 1-30b Molecular Biology of the Cell 6th Edition Garland Science 2008

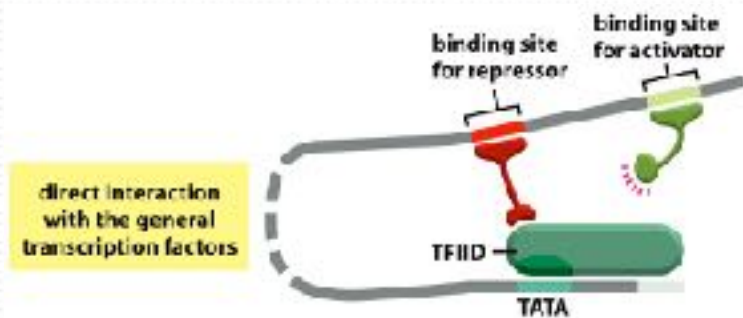


Figure 1-30c Molecular Biology of the Cell 6th Edition Garland Science 2008

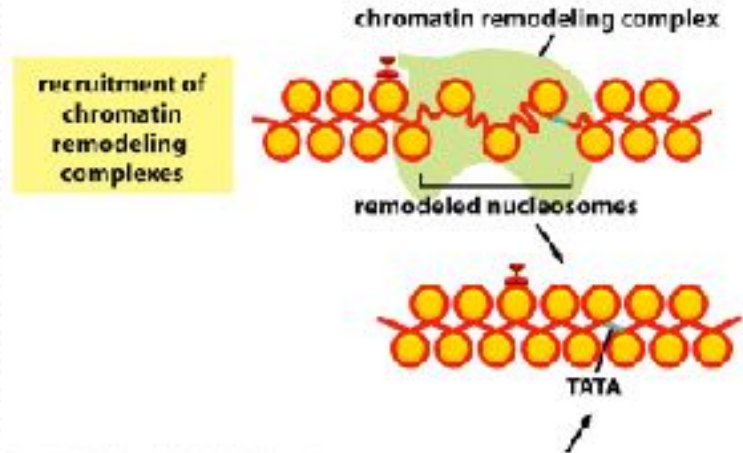


Figure 1-30d Molecular Biology of the Cell 6th Edition Garland Science 2008

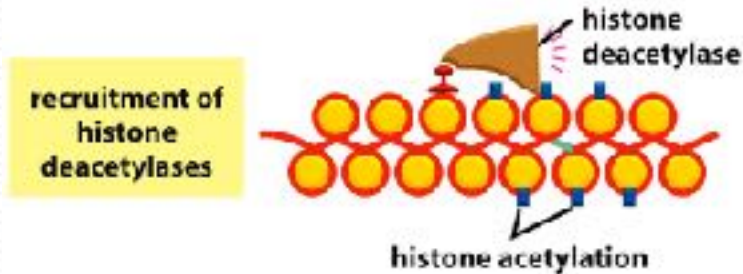


Figure 1-30e Molecular Biology of the Cell 6th Edition Garland Science 2008

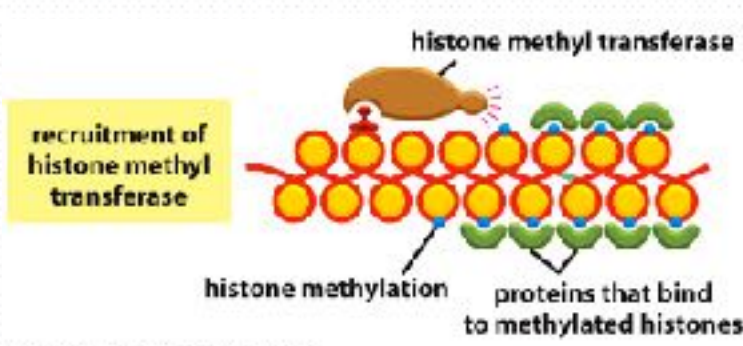


Figure 1-30f Molecular Biology of the Cell 6th Edition Garland Science 2008

Podniesienie poziomu ekspresji

Jedno białko reguluje wieloma genami

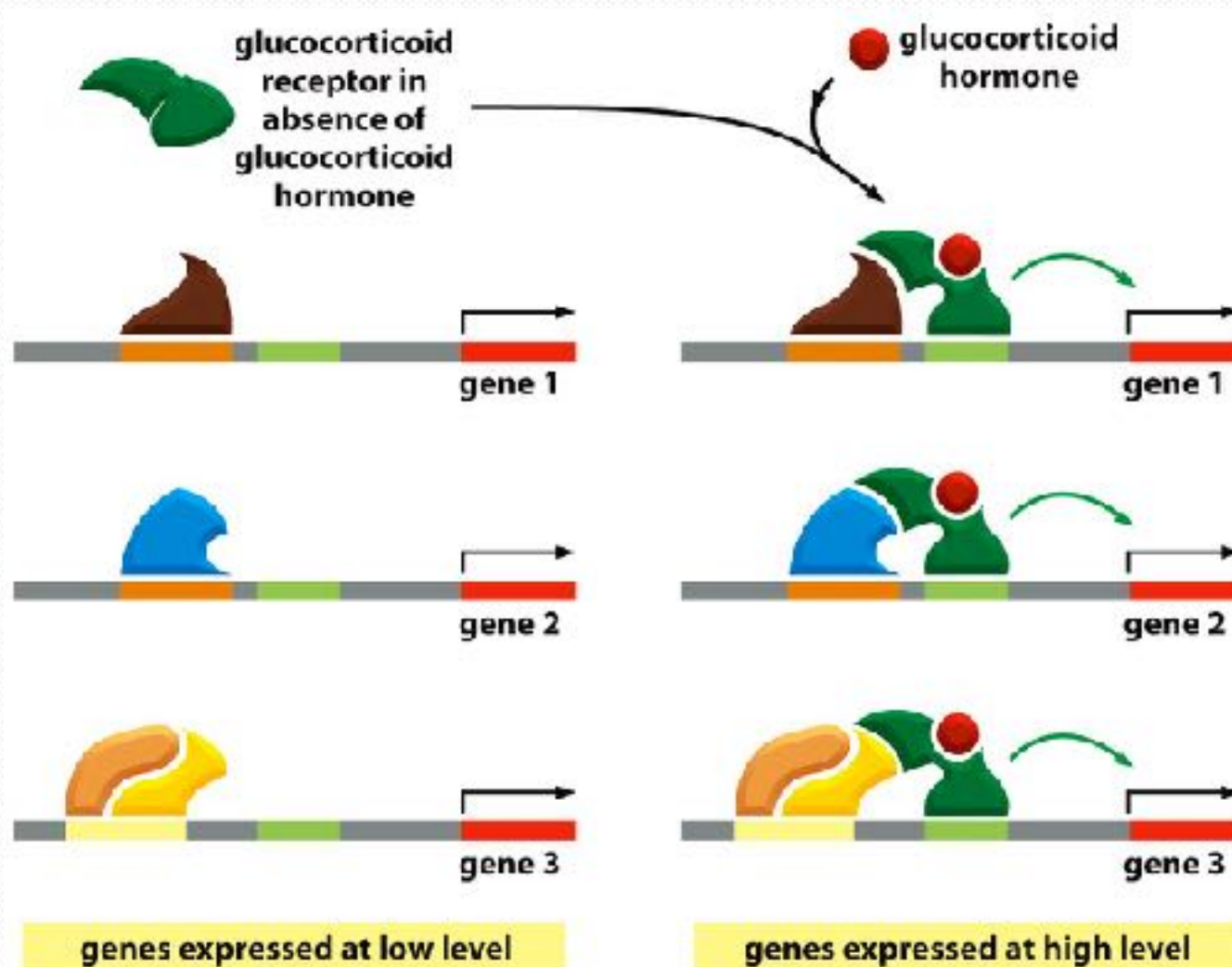


Figure 7-74 Molecular Biology of the Cell 5/e (© Garland Science 2008)

Elementy izolujące - insulator

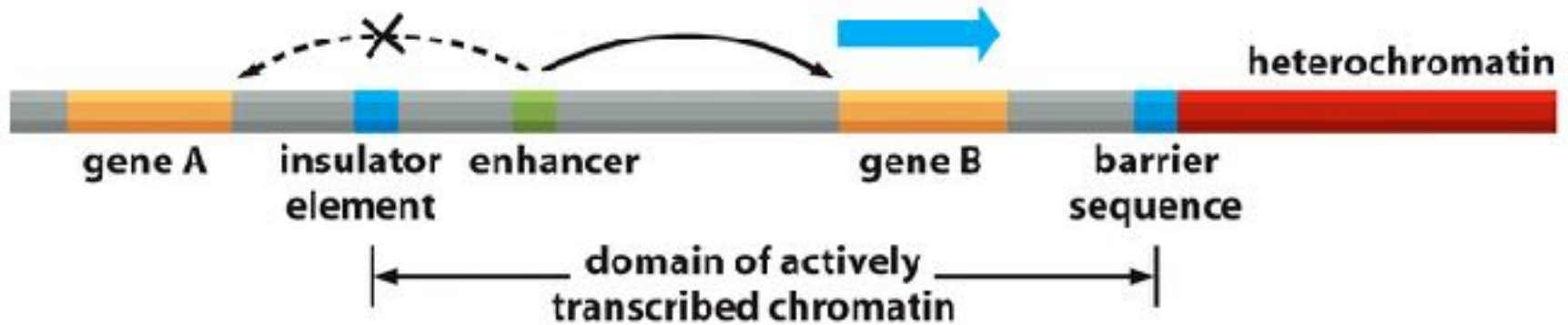


Figure 7-62 Molecular Biology of the Cell 5/e (© Garland Science 2008)



Różnicowanie się - 5 czynników 8 typów komórek

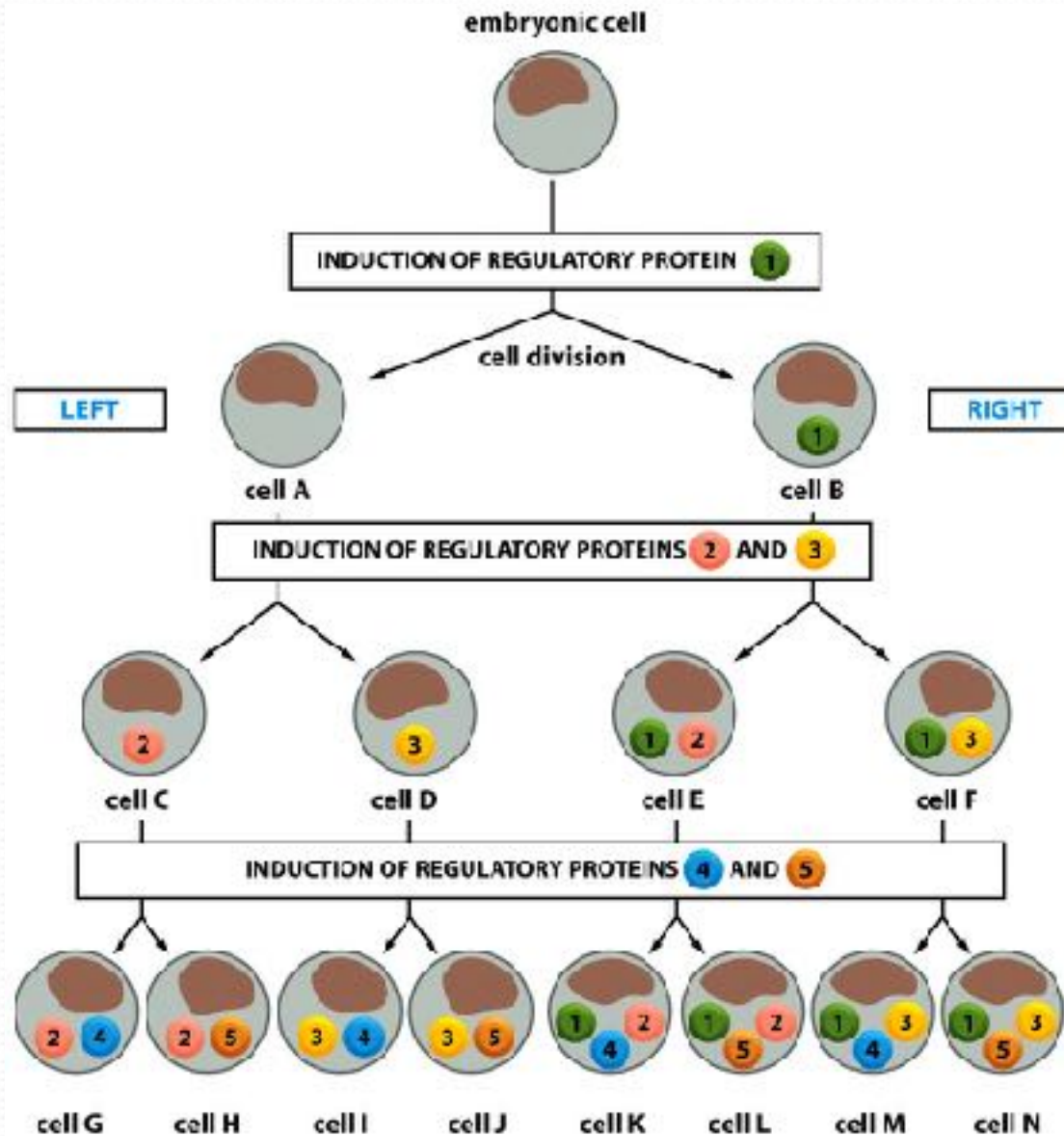


Figure 7-76 Molecular Biology of the Cell 5/e (© Garland Science 2008)

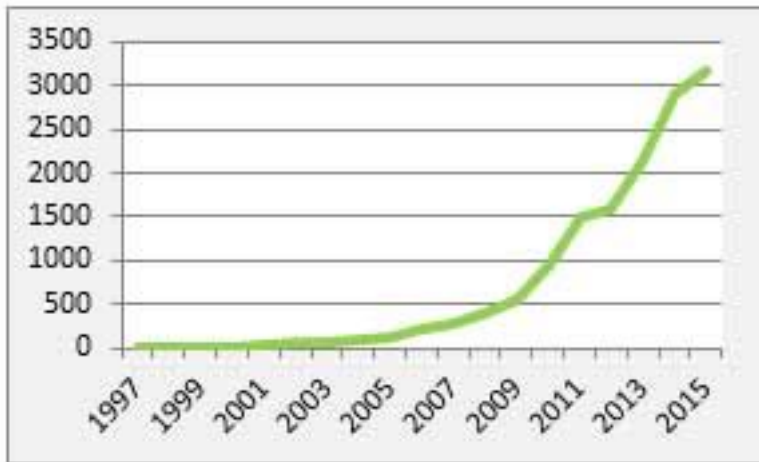


What is DNA methylation?

It is an epigenetic change

Modification that changes gene expression without affecting genetic sequence

Are these modifications important?



Epigenetics-related publications in PubMed

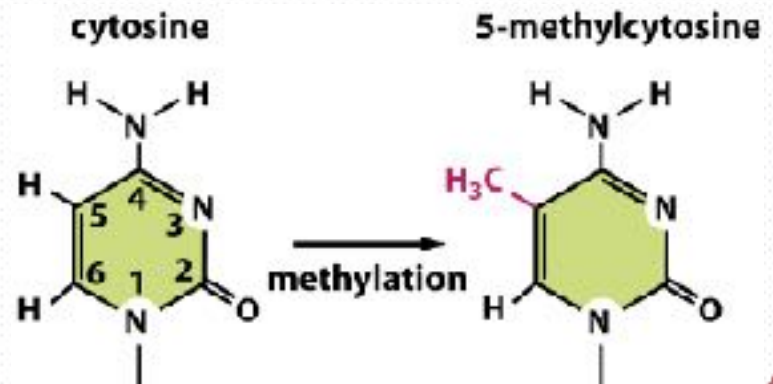
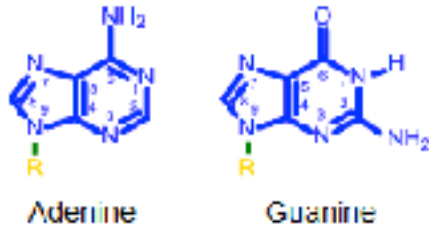


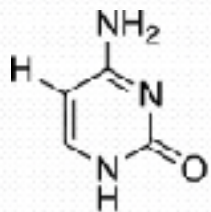
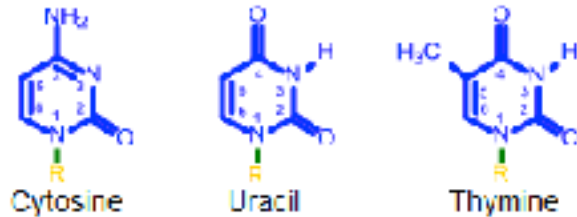
Figure 1. W. Adenauer, Biology of the Cell 10, 11 (2006)

What is DNA methylation?

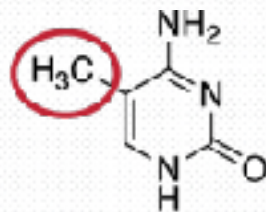
Purines



Pyrimidines



Cytosine

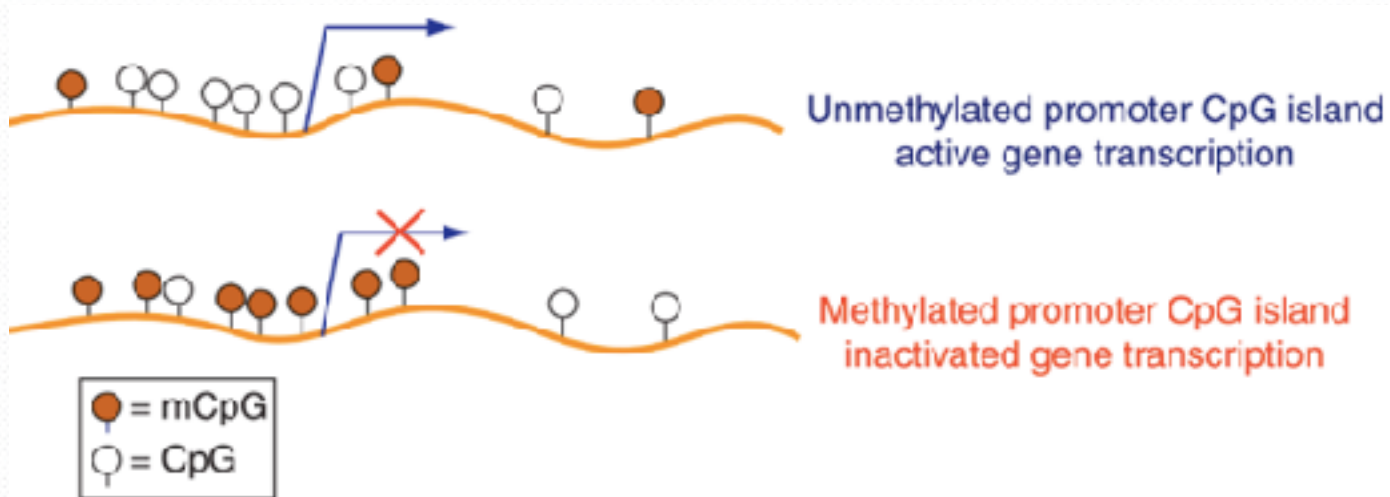
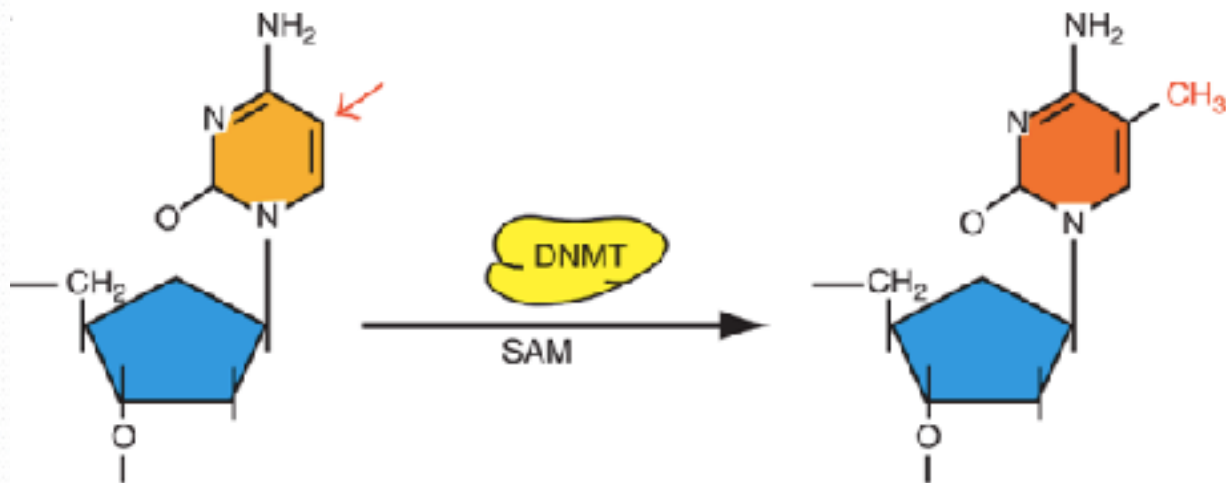


methylated Cytosine

- Addition of a methyl group to DNA
- 5-carbon of the cytosine ring
- Methyl groups project into the major groove of DNA
- Inhibit transcription



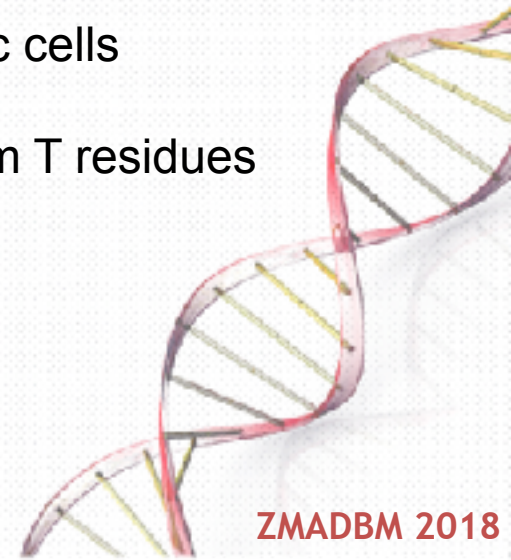
DNA methyltransferase



DNA methylation targets

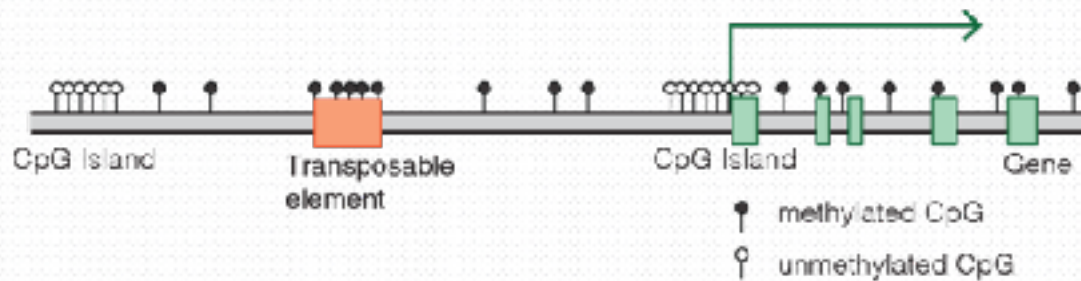


- Vertebrates – 60-80% of CpG are methylated in somatic cells
- DNA methylation appears as default state
- Methylated C residues spontaneously deaminate to form T residues
- CpG to TpG



DNA methylation targets

Typical mammalian DNA methylation landscape

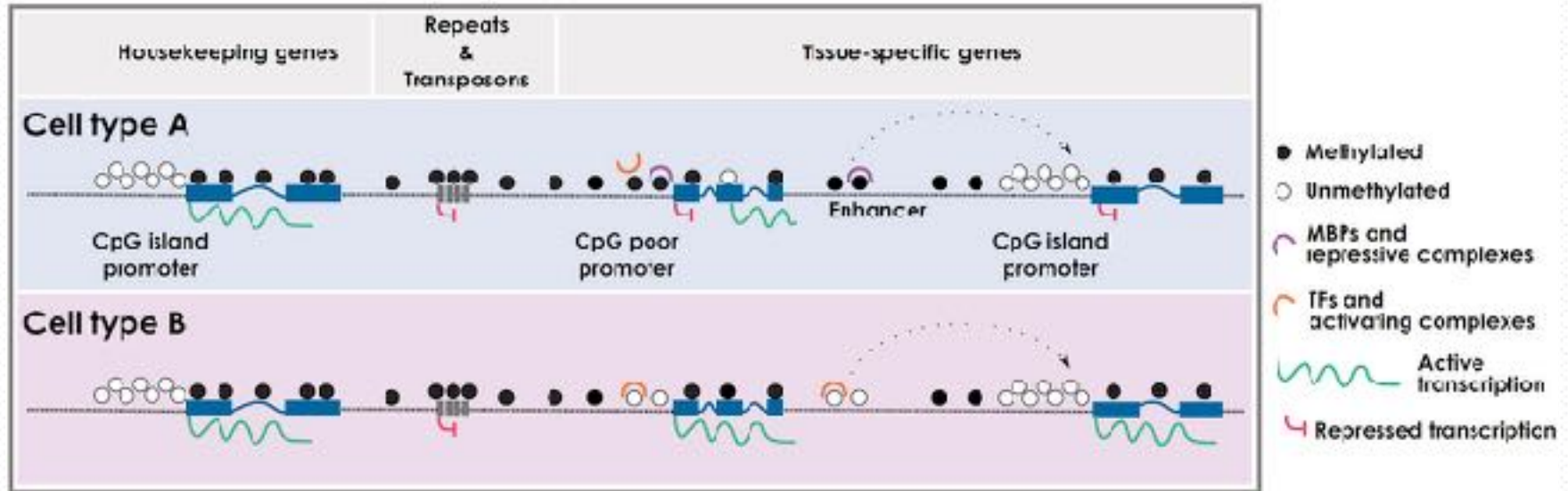


CpG islands:

- Length greater than 200bp
- G+C content above 50%
- ~ 25,000 of CpG islands
- ~75% being less than 850bp long
- ~50% being located in promoter regions
- ~25% being located within gene bodies
- 60% to 70% of genes have an island in the promoter
- Unmethylated and enriched in chromatin modification H3K4

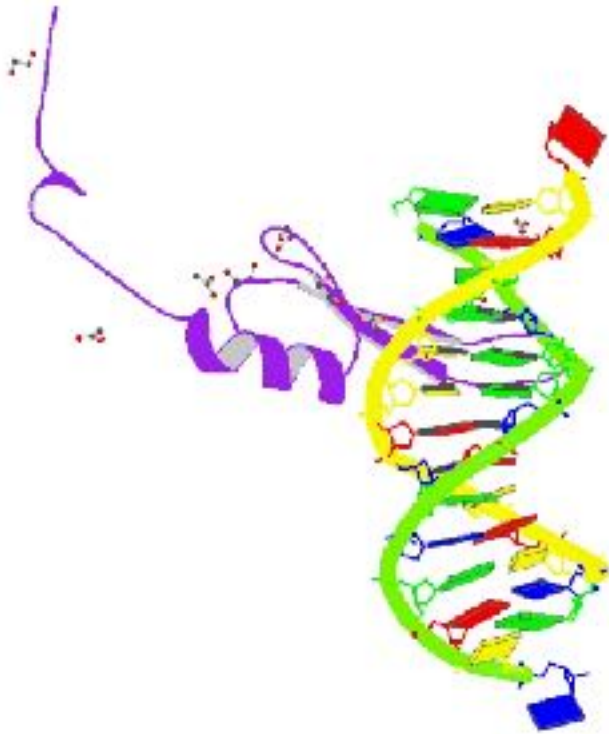


Repression of CpG-dense promoters



- CpG-dense promoters of actively transcribed genes are never methylated
- CpG-poor promoters – DNA methylation function is unclear
- Physical disturbance (TF-DNA)
- Bound by proteins known as methyl-CpG-binding domain proteins (MBDs)

Methyl-CpG-binding domain



Crystal structure of methyl CpG Binding Domain of MBD4 in complex with the 5mCG/TG sequence

- Effects of DNA methylation are mediated
- Specific domain of ~70 residues
- Linked to other domains that recruit:
 - histone deacetylase complexes
 - chromatin remodeling factors

Journal Club: Nucleic Acids Res. • v.21(21): 1069 Oct.20 • PMID:11401

Nucleic Acids Research

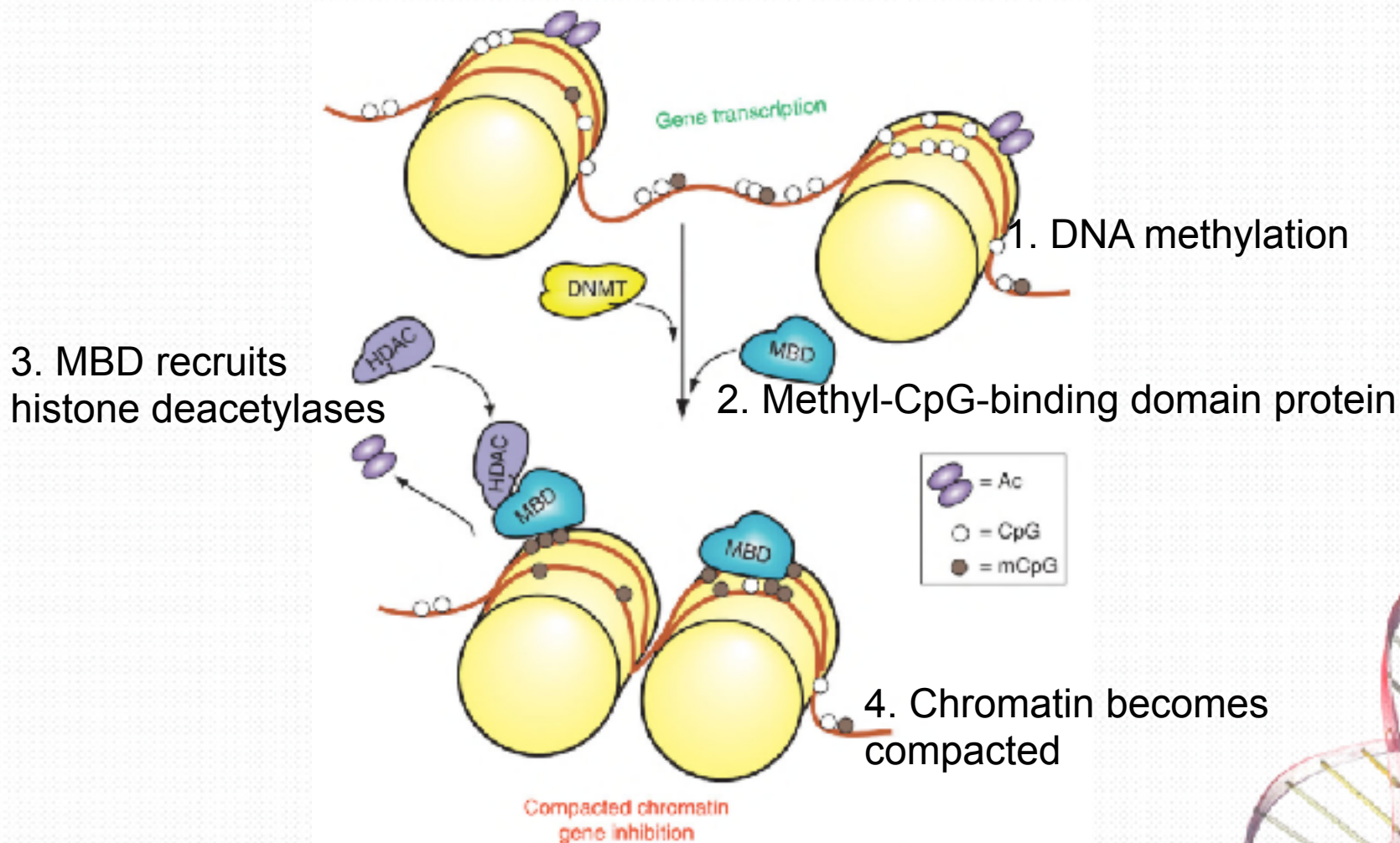
[Nucleic Acids Res. 1993 Oct 21; 21\(21\): 4006-4007.](#)

PMCID: PMC341424

Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2.










[X Nan](#), [R R Meehan](#), and [A Bird](#)

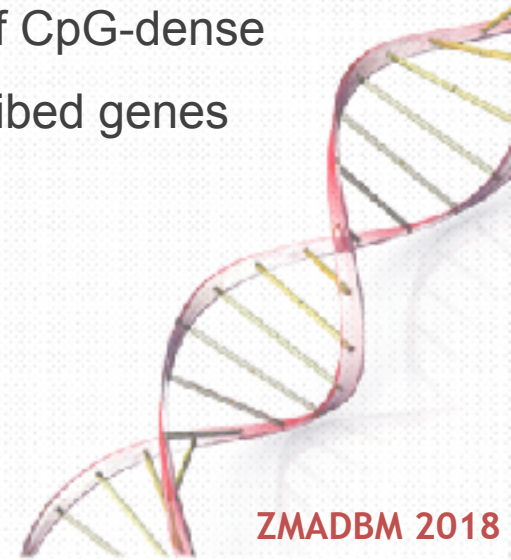
Methyl-CpG-binding domain



Kaiso-like family of proteins binds by zinc-finger motifs to methylated DNA

Functions

-  Repress gene transcription
-  Essential for normal development
-  Associated with a number of key processes:
 -  genomic imprinting,
 -  chromatin structure,
 -  X-chromosome inactivation,
 -  repression of transposable elements, repression of CpG-dense promoters, methylation of the gene body of highly transcribed genes
 -  aging,
 -  carcinogenesis.

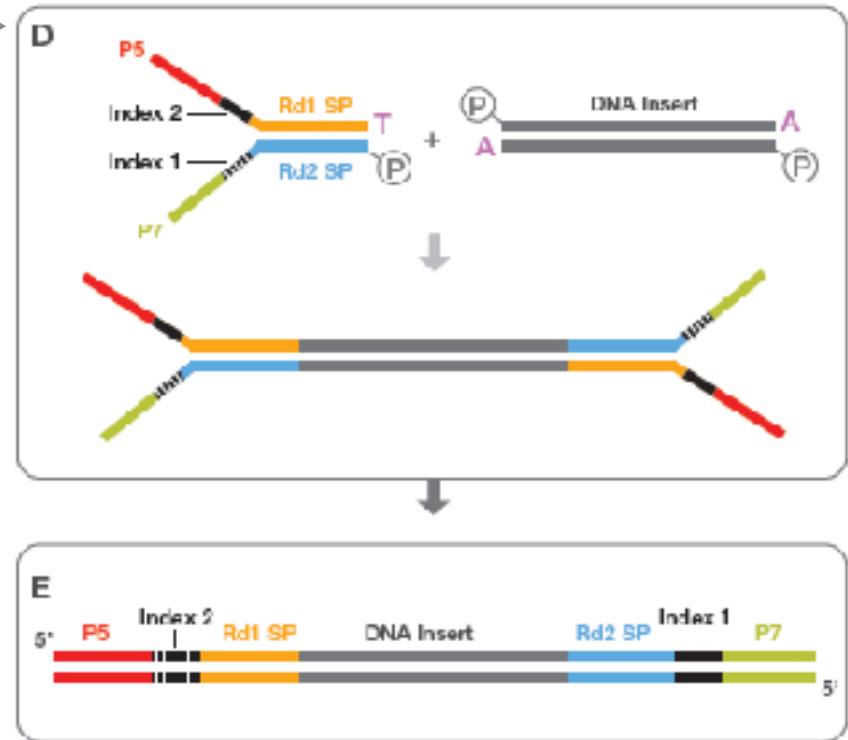
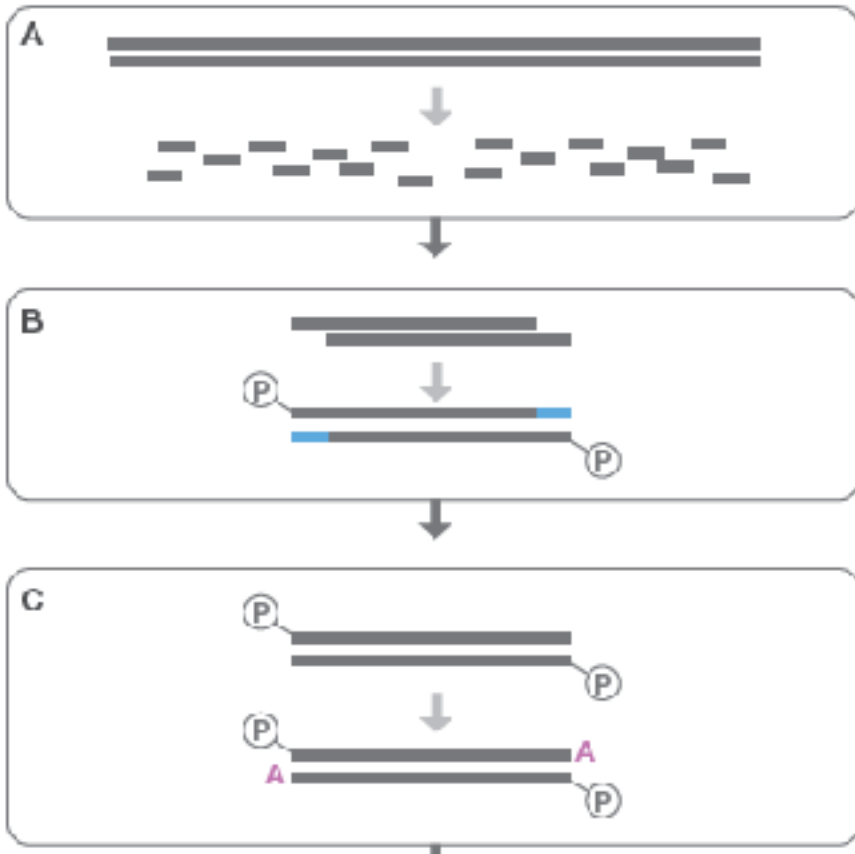


Introduction to Next Generation Sequencing - focus on Transcriptomics

Dr inż. Bartosz Wojtaś
Pracownia Neurobiologii Molekularnej,
Centrum Neurobiologii,
Instytut Biologii Doświadczalnej im. Nenckiego
PAN

DNA Sample Prep Workflow

DNA



Hybridize Fragment & Extend

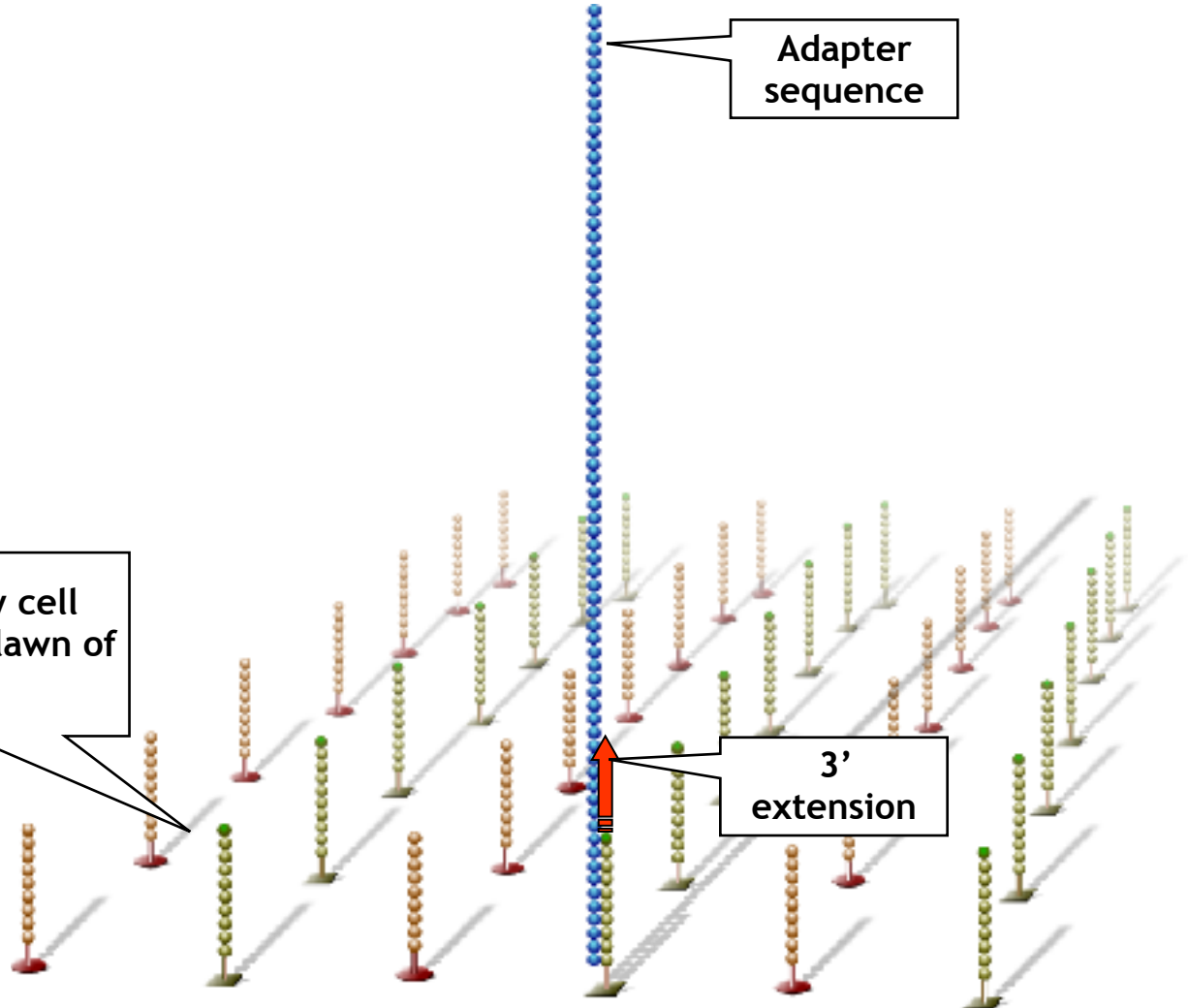
Single DNA libraries are hybridized to primer lawn

Bound libraries then extended by polymerase

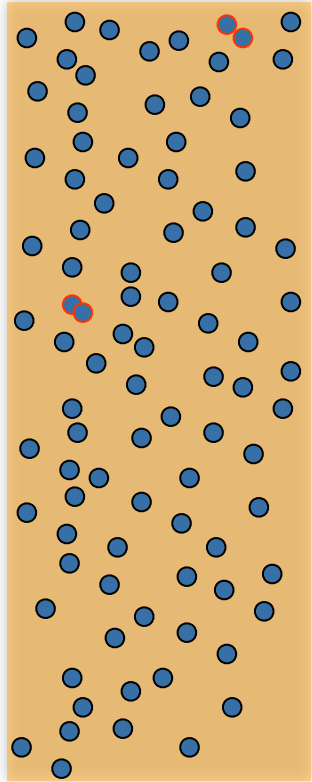
Surface of flow cell coated with a lawn of oligo pairs

Adapter sequence

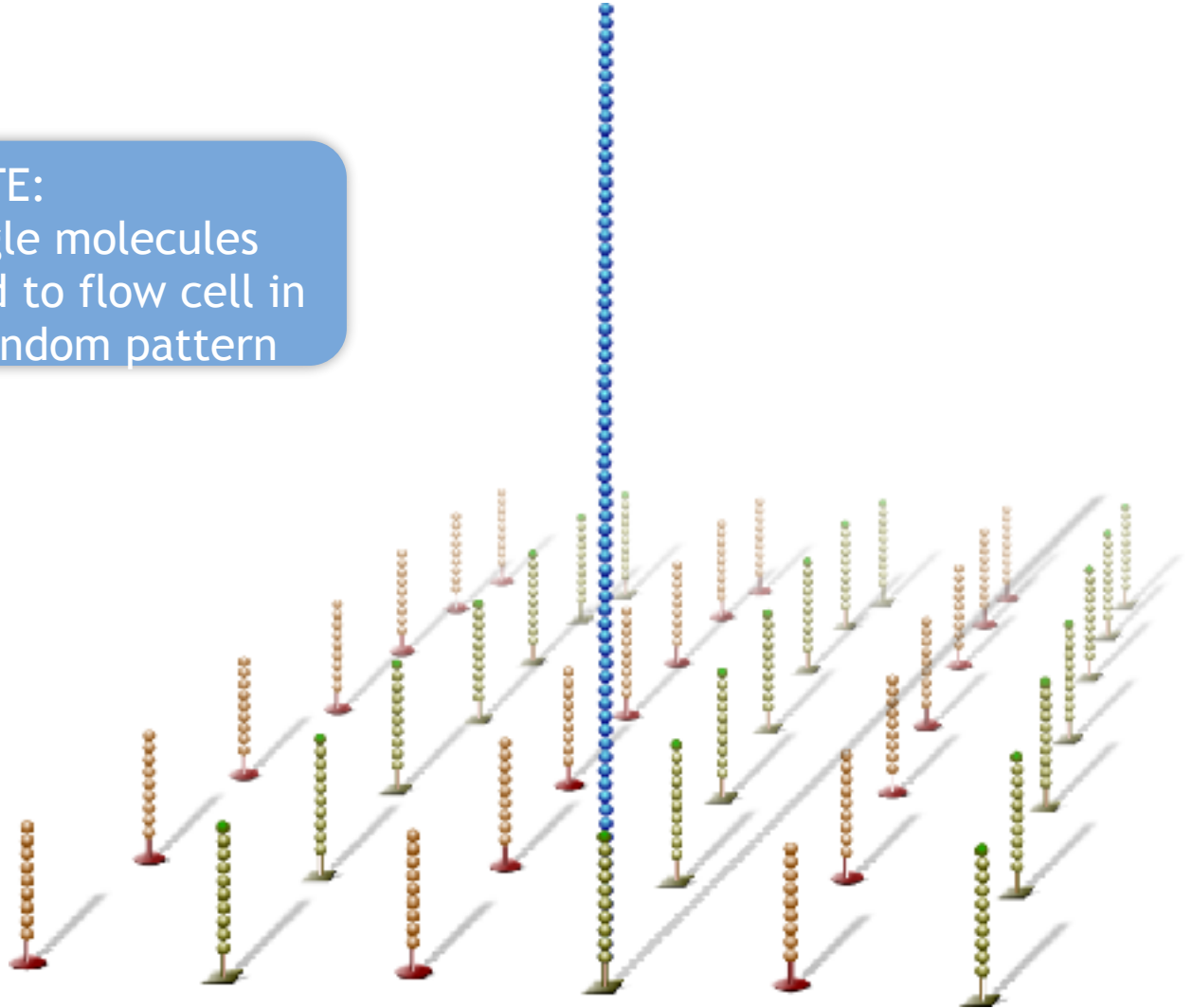
3' extension



Hybridize Fragment & Extend



NOTE:
Single molecules
bind to flow cell in
a random pattern

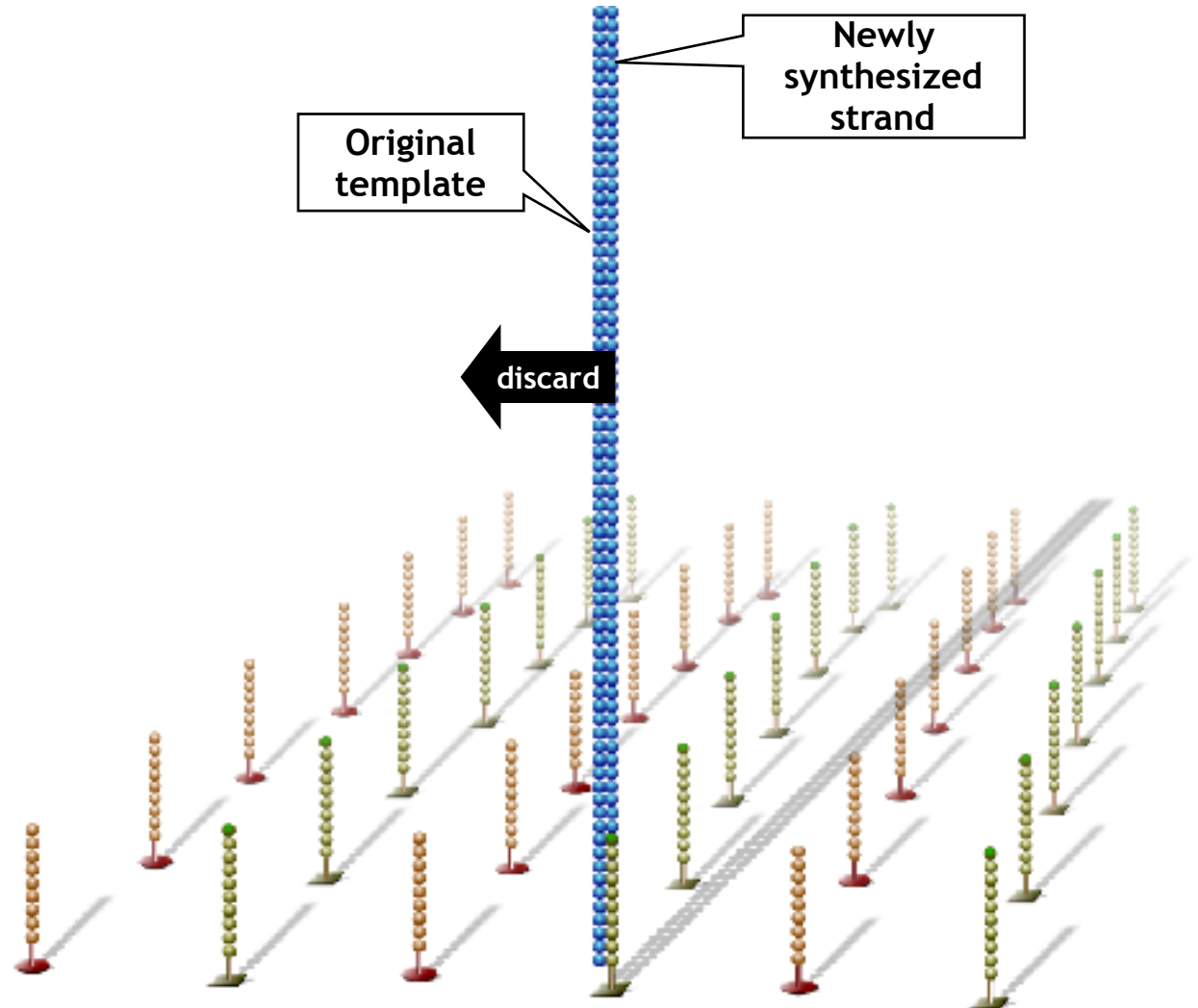


Denature Double-stranded DNA

Double-stranded molecule is denatured

Original template washed away

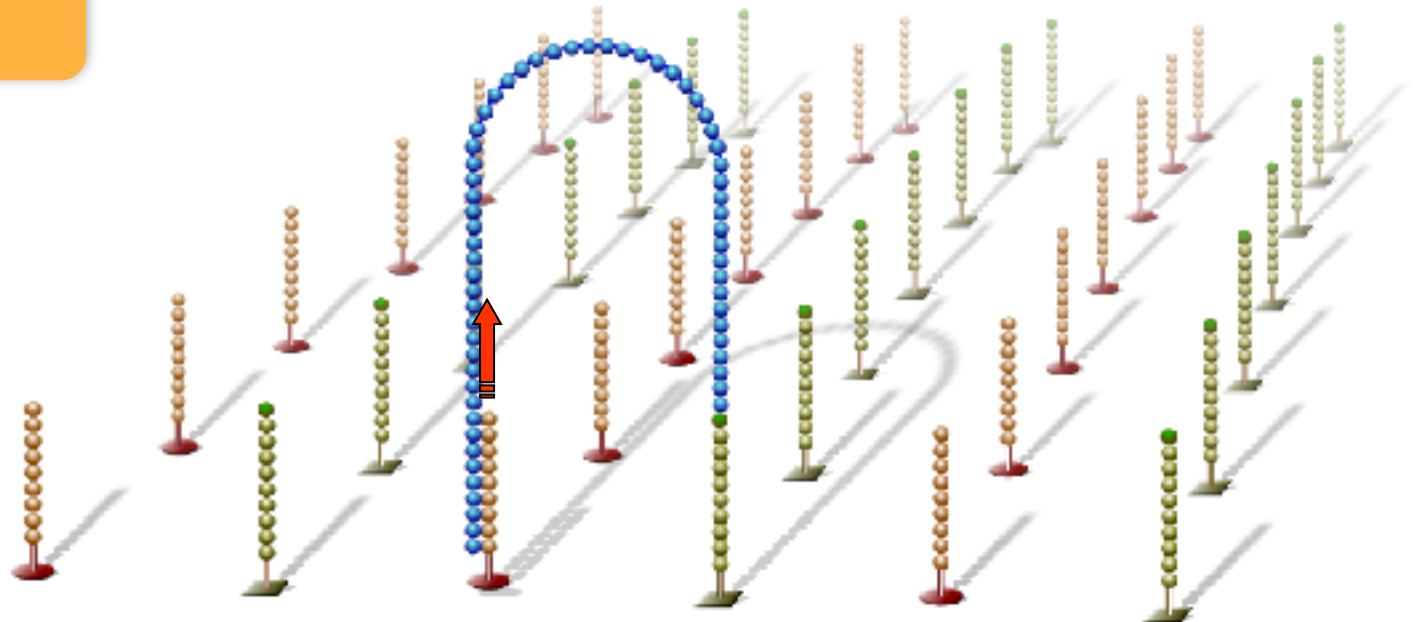
Newly synthesized strand is covalently attached to flow cell surface



Bridge Amplification

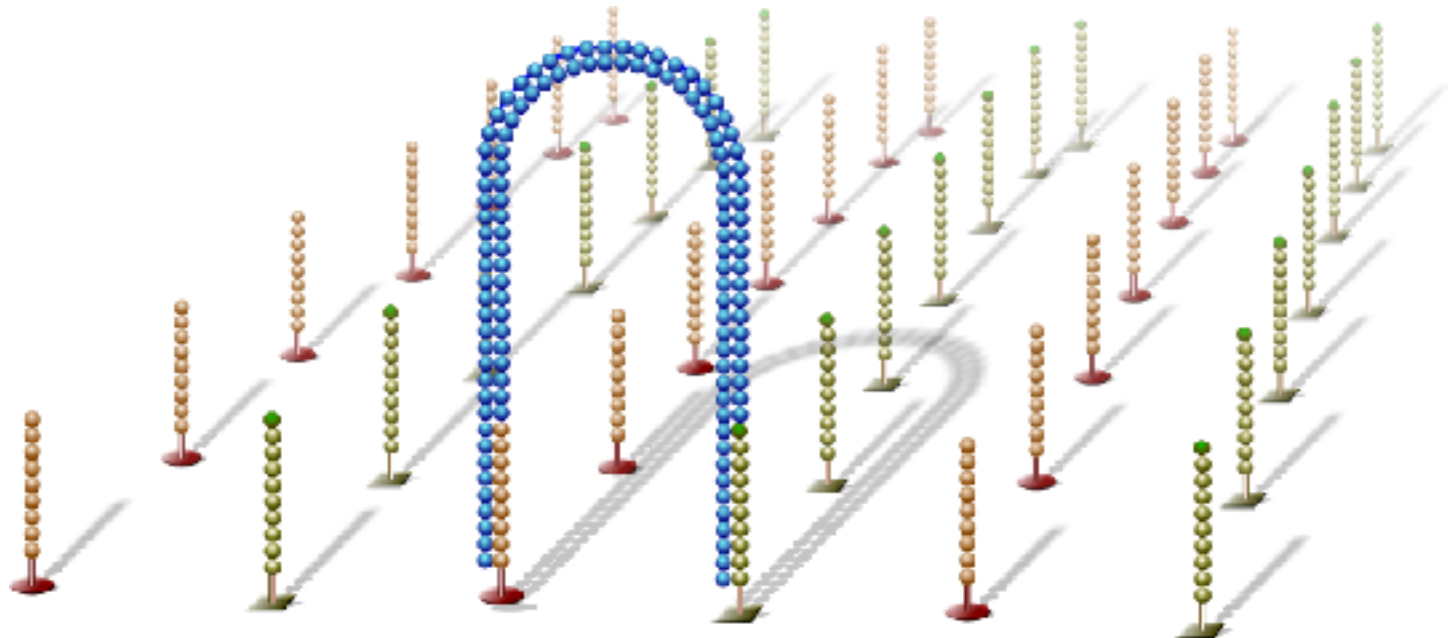
Single-stranded molecule flips over and forms a bridge by hybridizing to adjacent, complementary

Hybridized primer is extended by polymerase



Bridge Amplification

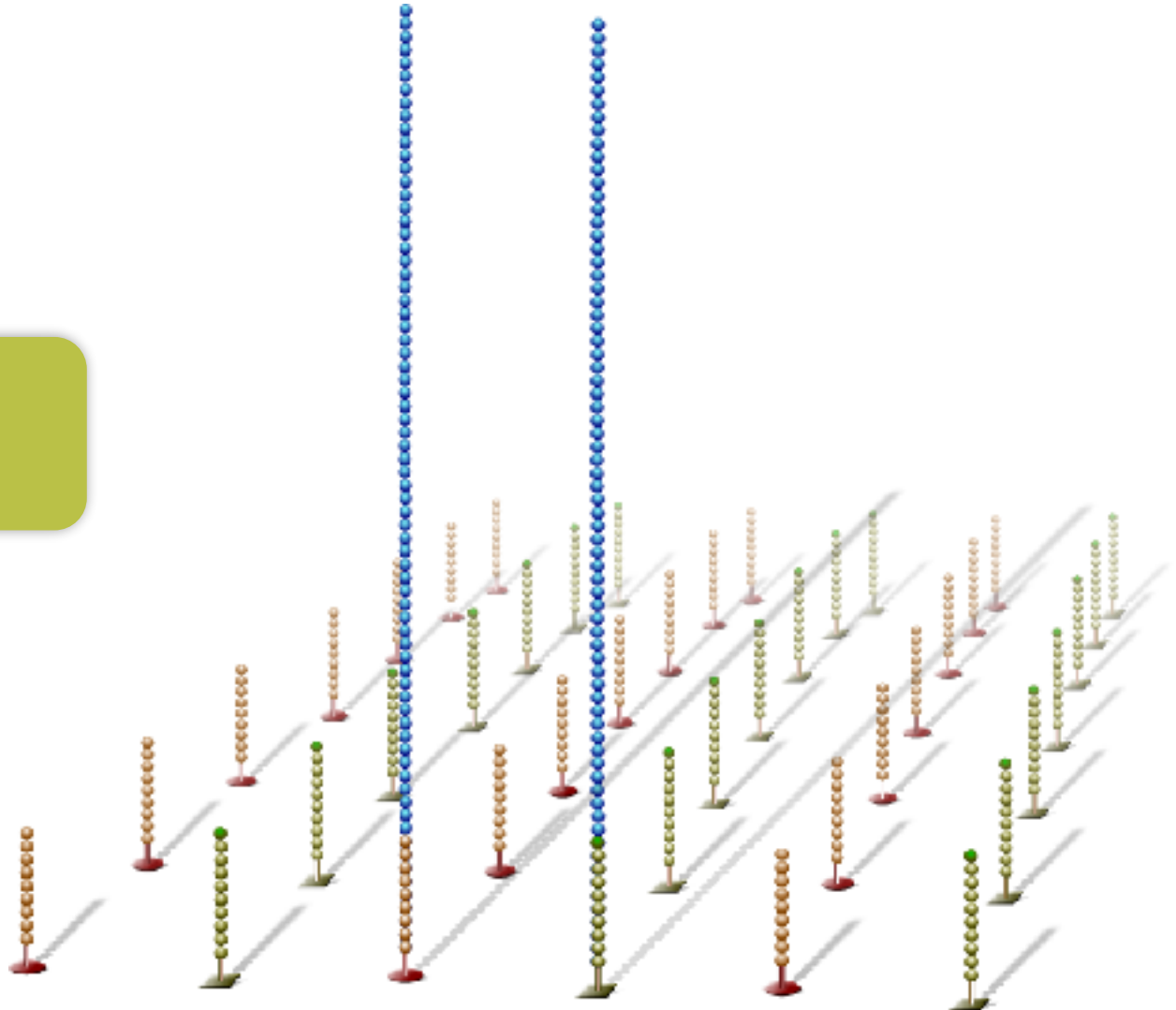
Double-stranded bridge is formed



Denature Double-stranded Bridge

Double-stranded bridge
is denatured

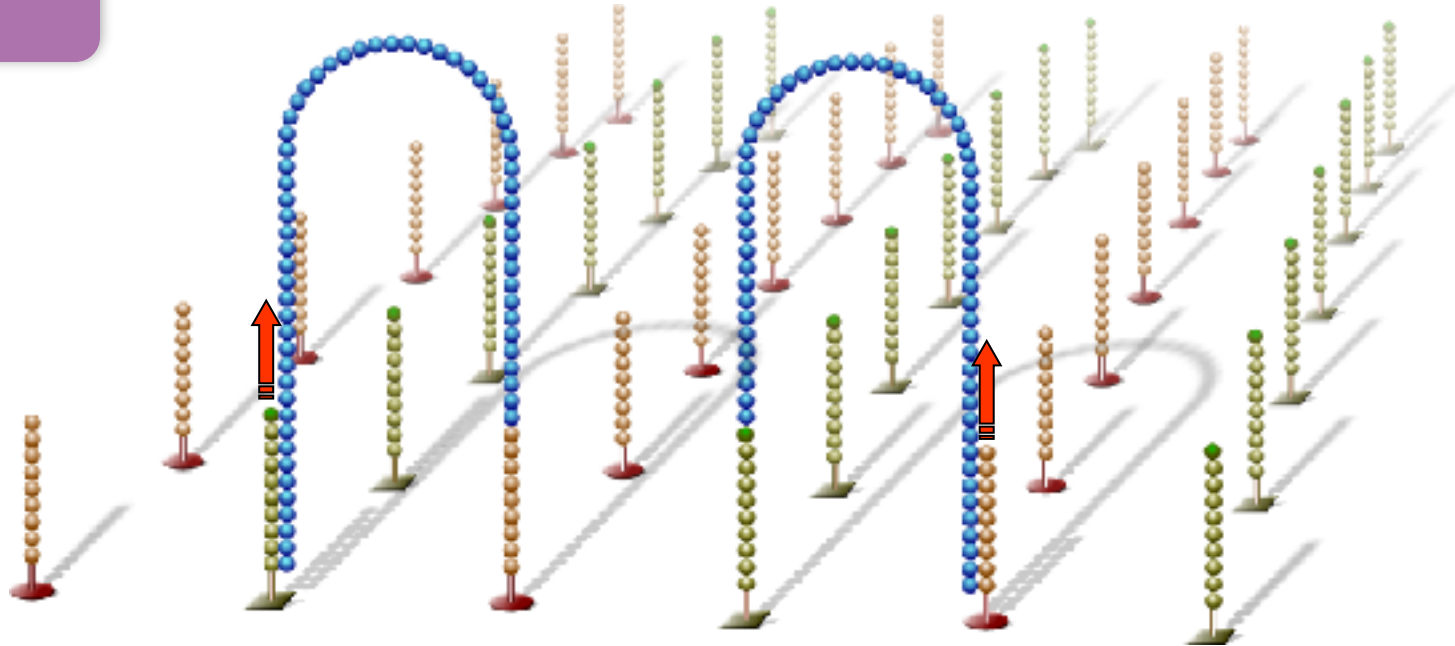
Result:
Two copies of covalently
bound single-stranded
templates



Bridge Amplification

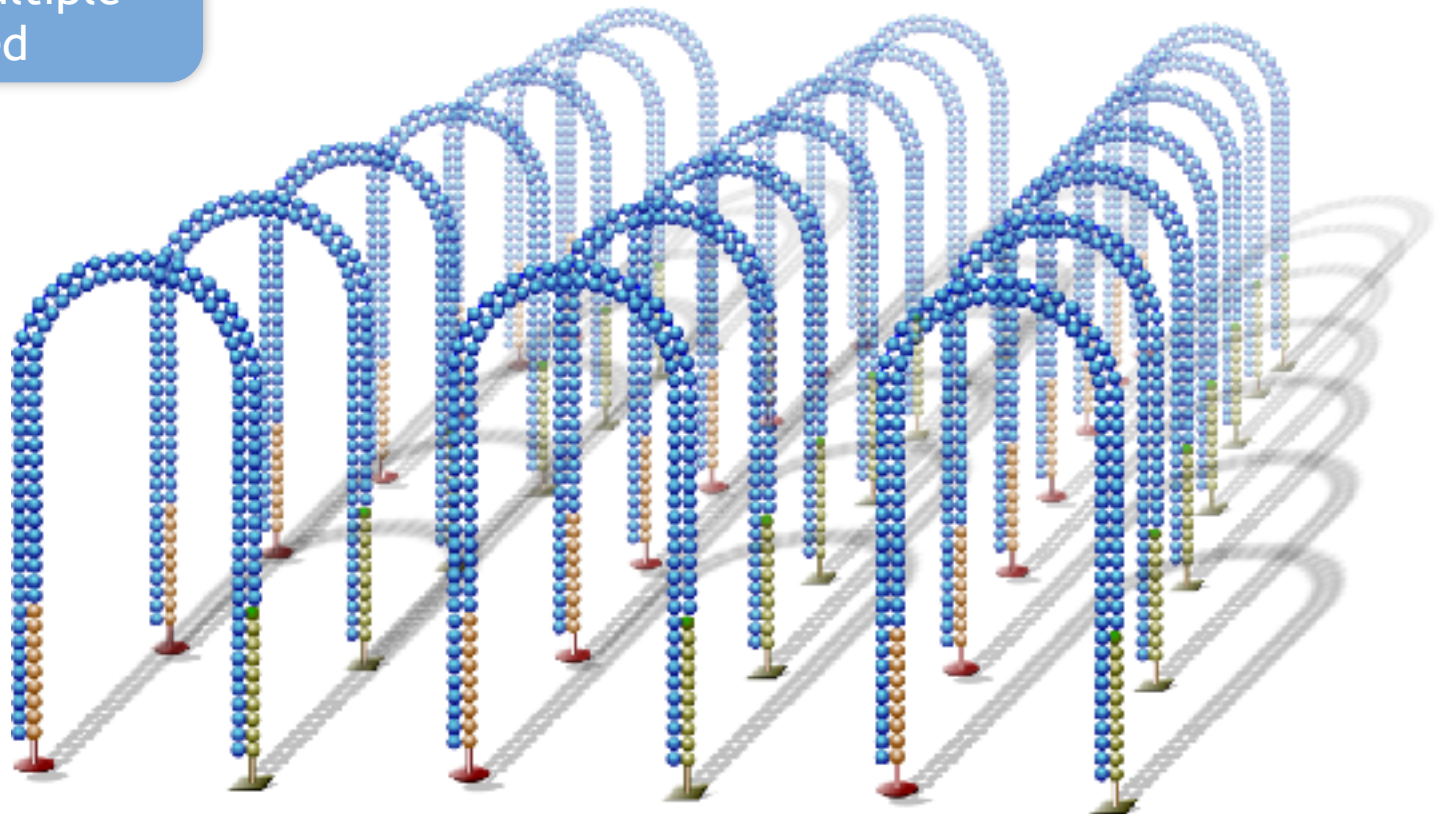
Single-stranded molecules flip over to hybridize to adjacent primers

Hybridized primer is extended by polymerase



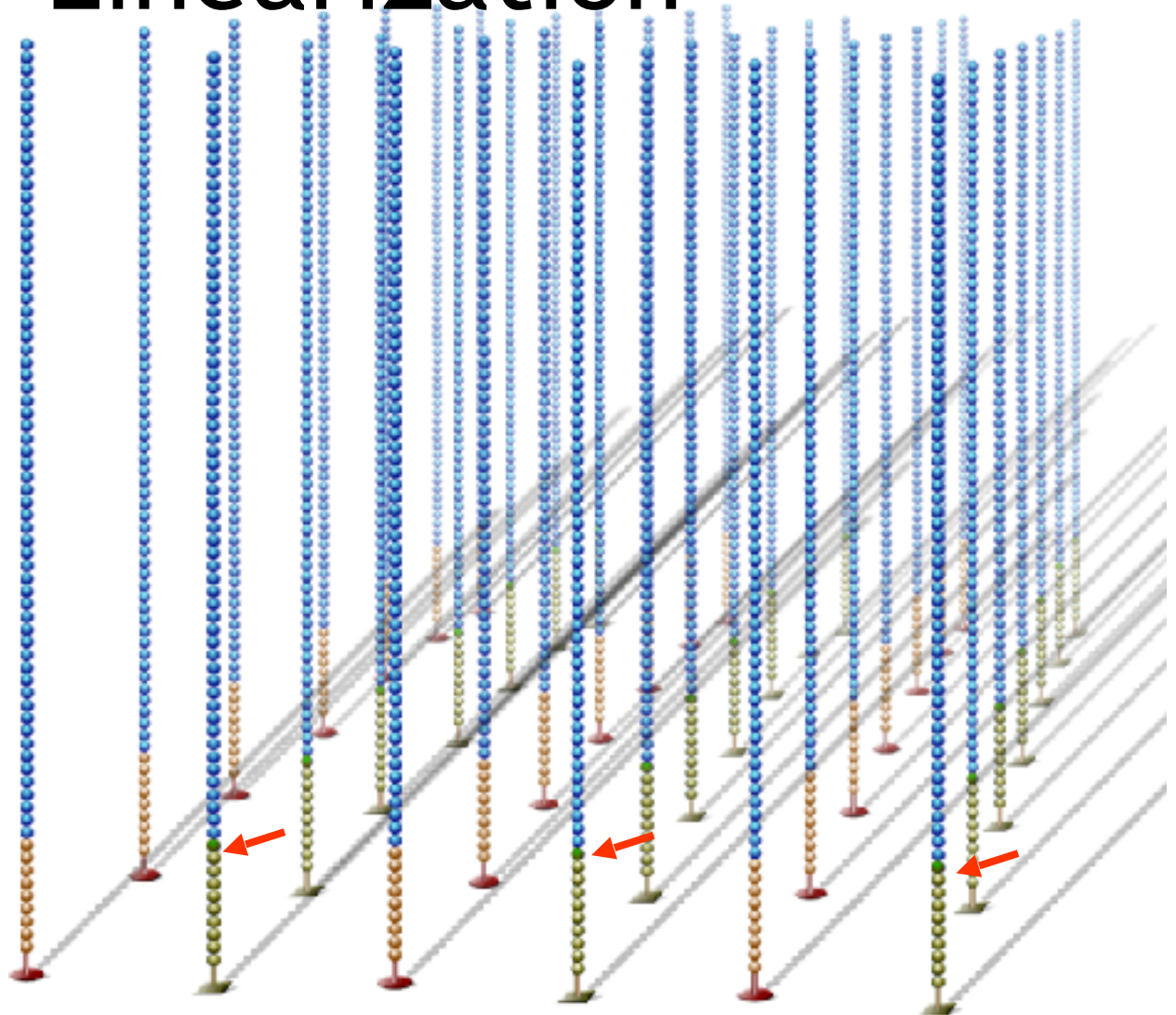
Bridge Amplification

Bridge amplification cycle repeated until multiple bridges are formed



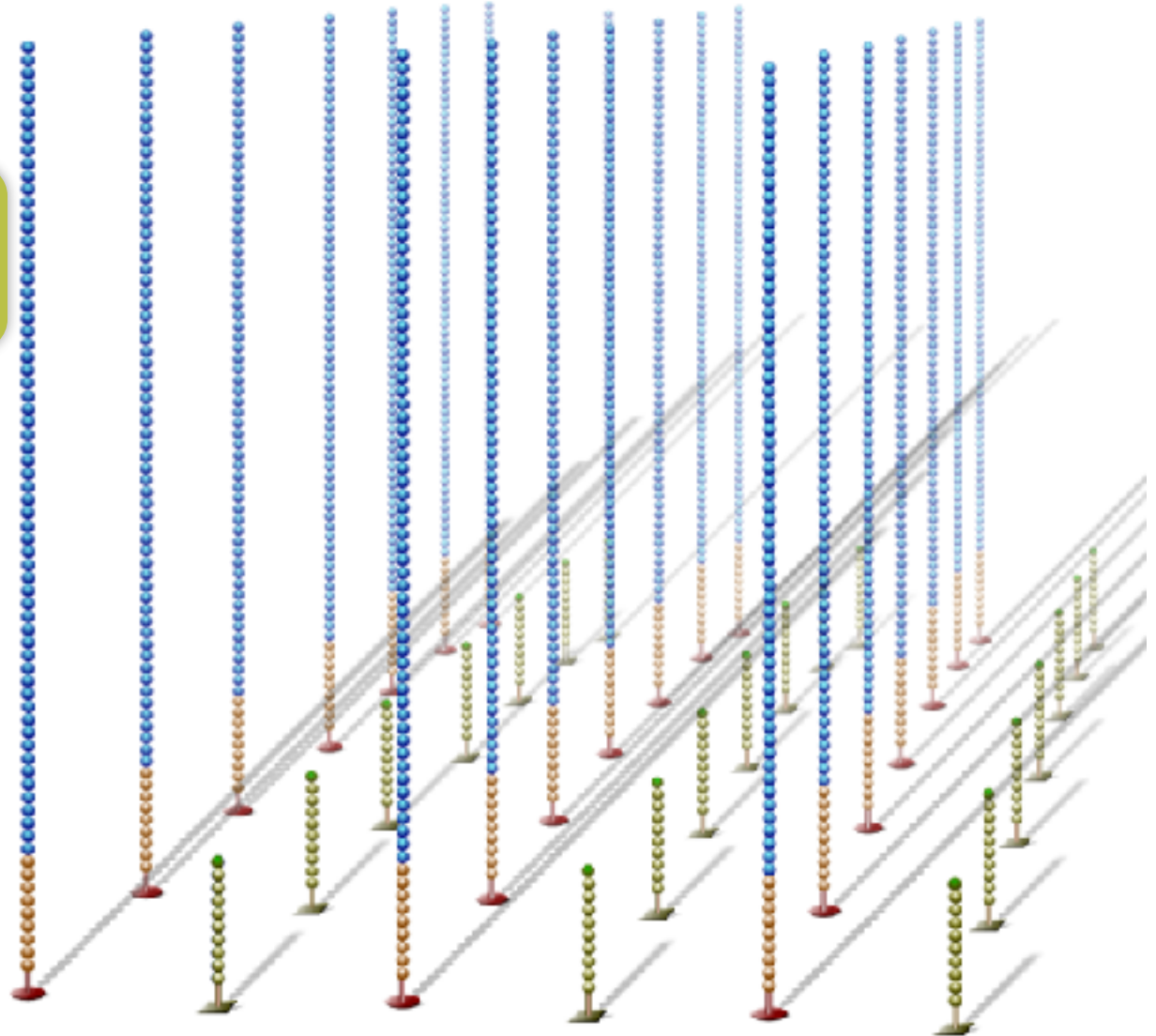
Linearization

dsDNA bridges are
denatured



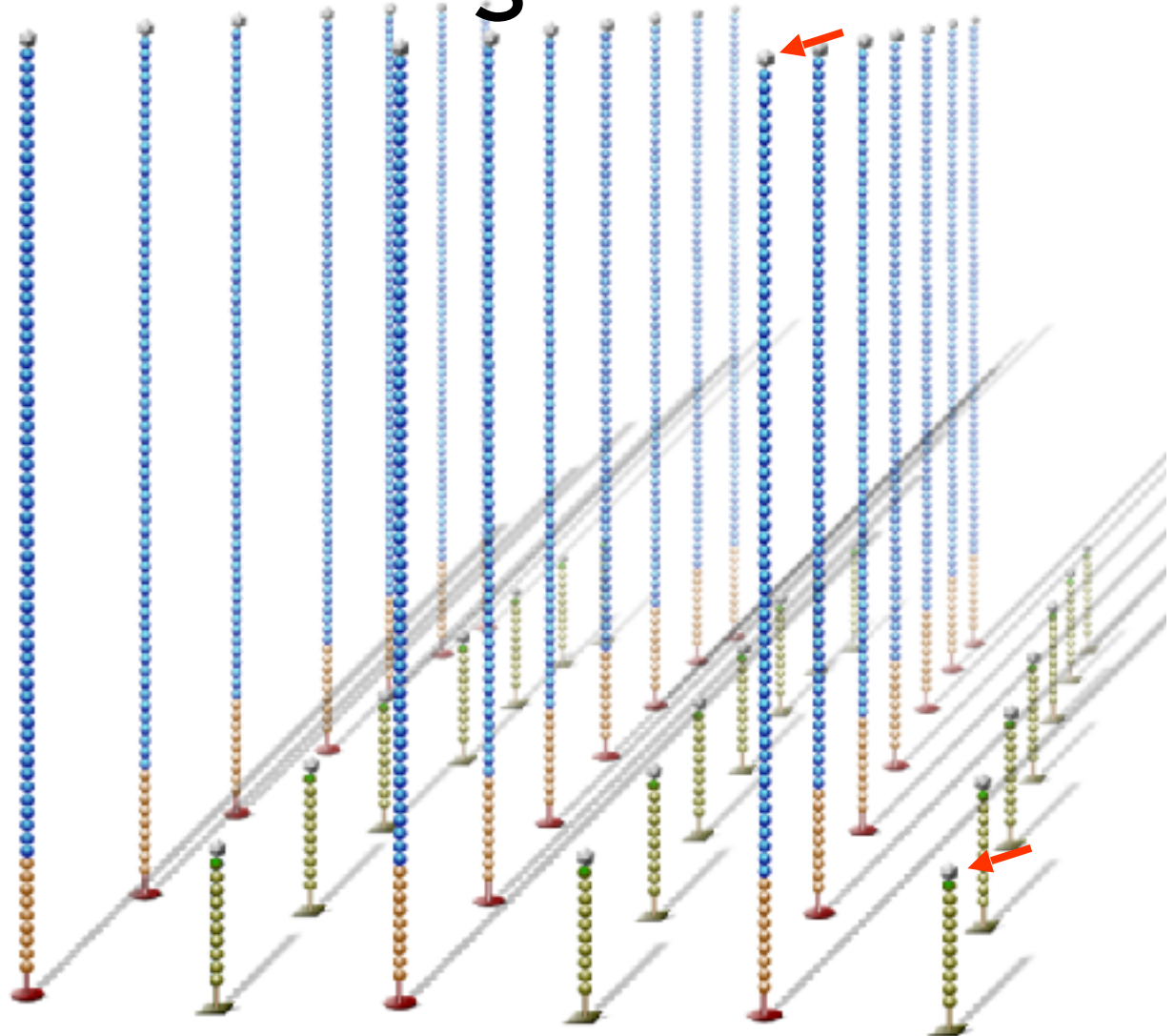
Reverse Strand Cleavage

Reverse strands cleaved and washed away, leaving a cluster with forward strands only



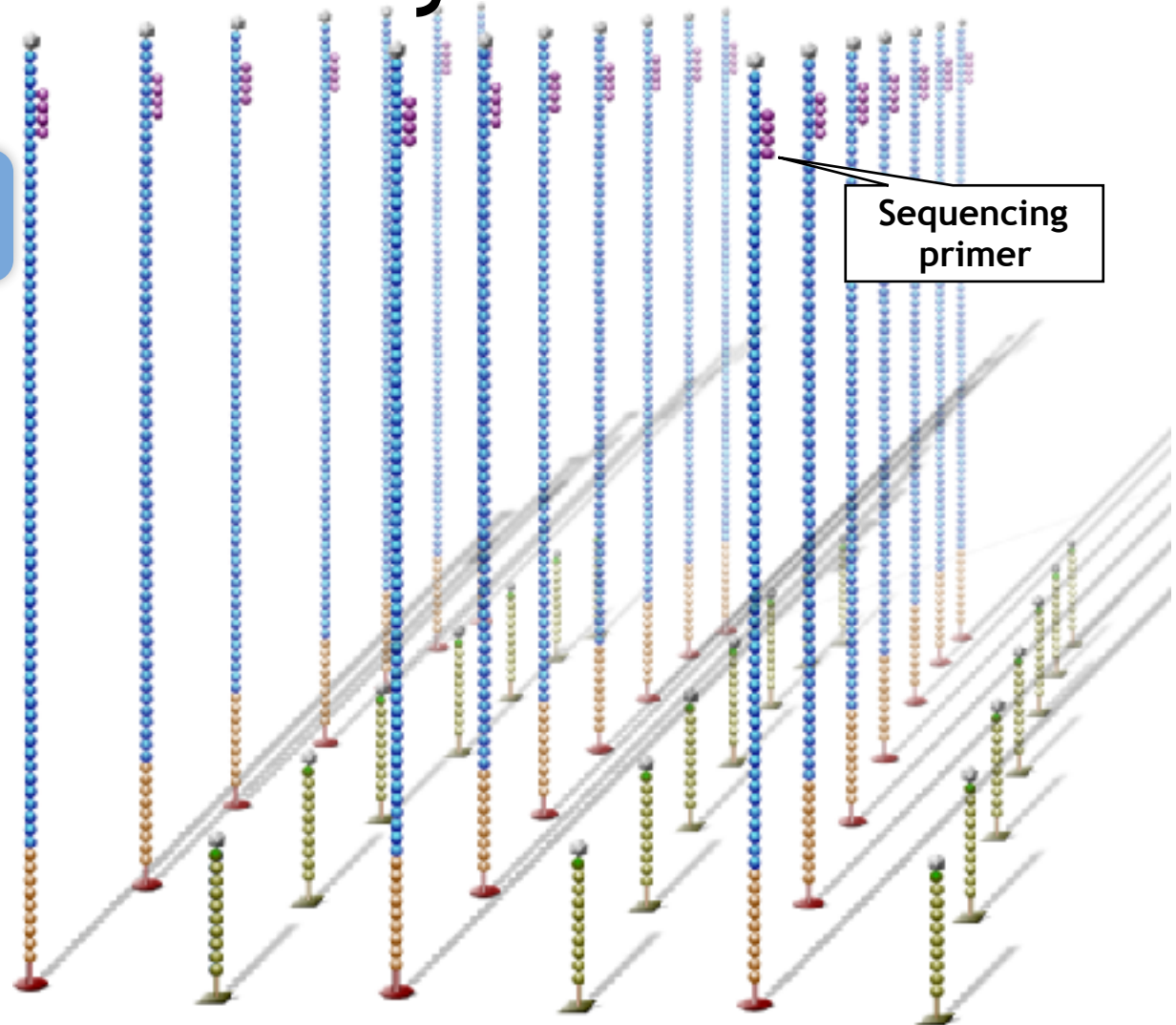
Blocking

Free 3' ends are blocked
to prevent unwanted DNA
priming

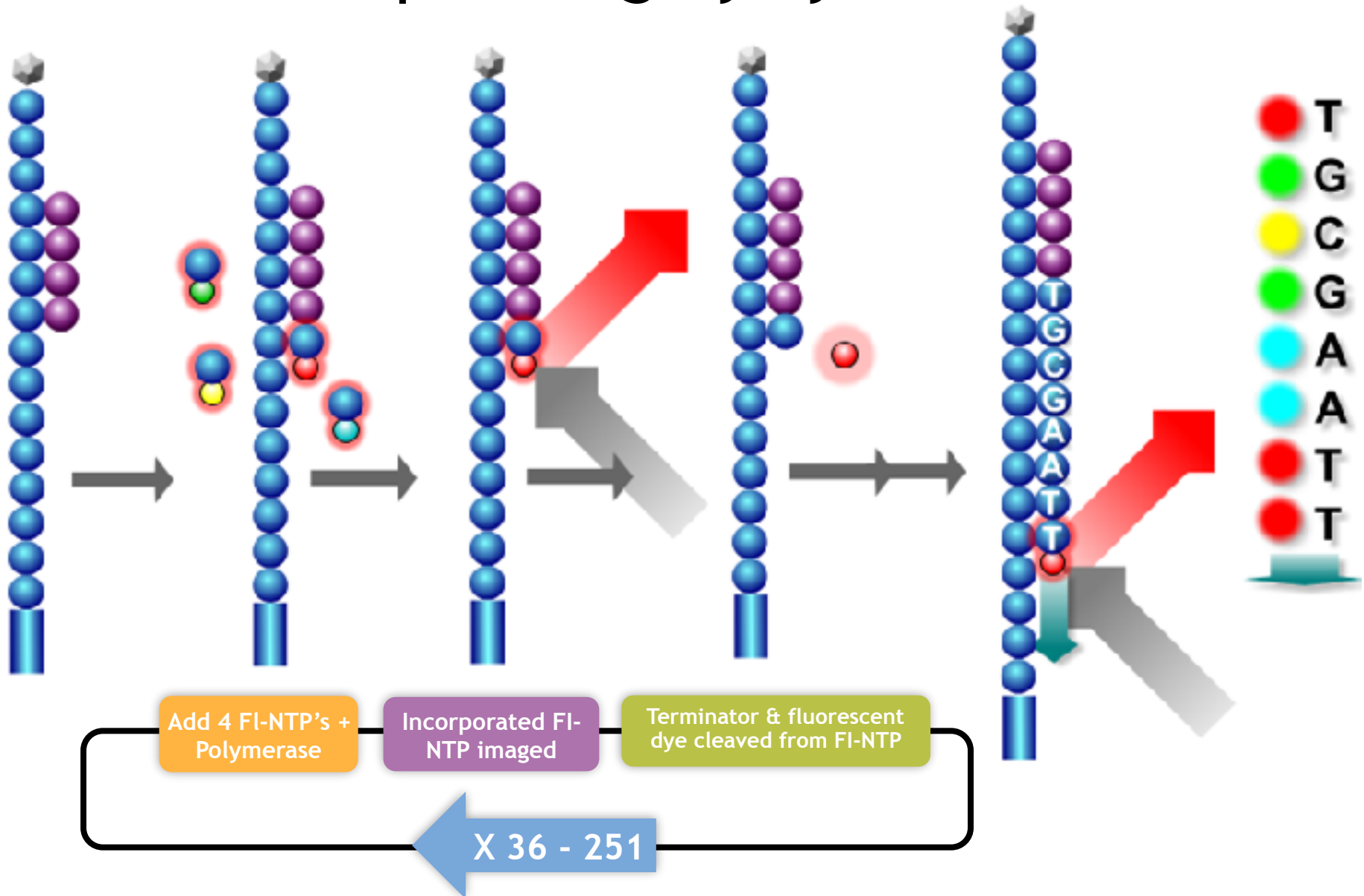


Read 1 Primer Hybridization

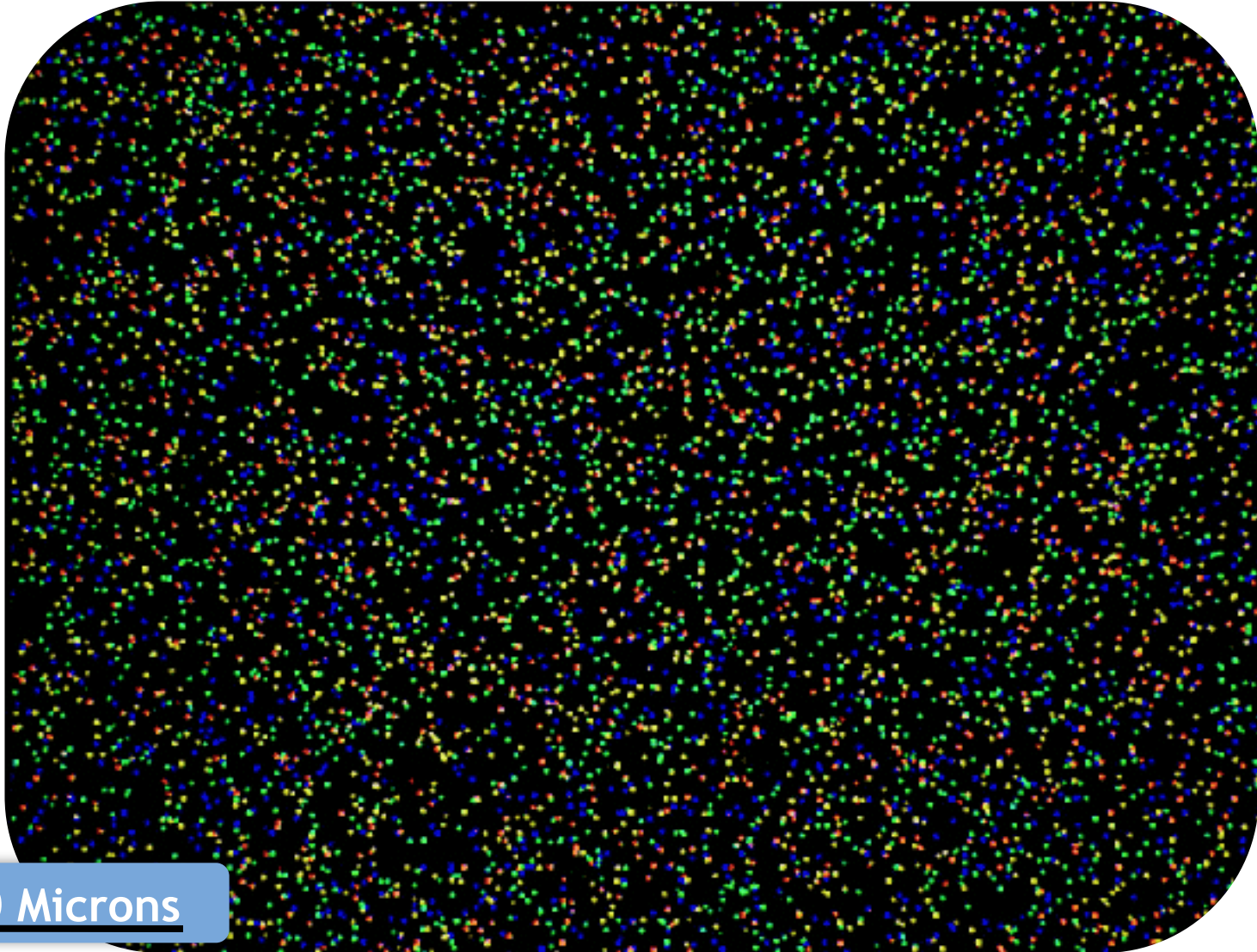
Sequencing primer is
hybridized to adapter
sequen



Sequencing By Synthesis

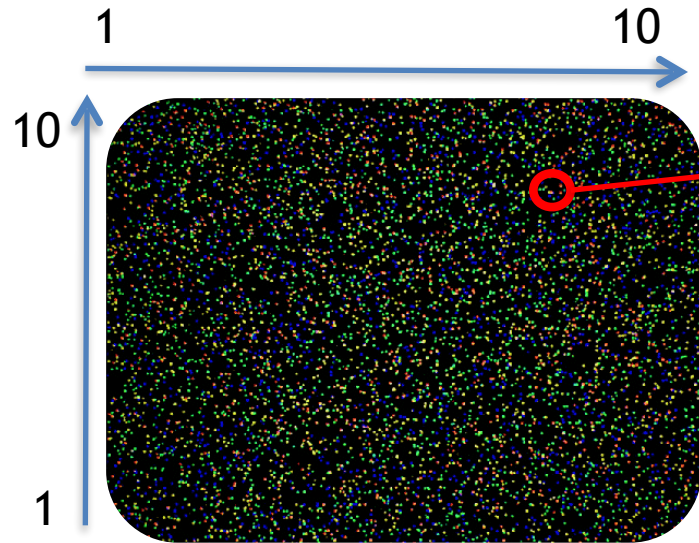


Clusters



100 Microns

Raw sequencing result is a spatial information



Cycle 1, base A read in position:

X = 8, Y = 8

Cycle 2, base T read in position:

X = 8, Y = 8

Cycle 3, base G read in position:

X = 8, Y = 8

Cycle 4, base C in position:

X = 8, Y = 8

After data transformation from
spatial to sequence (FASTQ) format
We will get sequence: ATGC

NGS summary

<https://www.youtube.com/watch?v=-kTcFZxP6kM&t=17s>

<https://www.youtube.com/watch?v=9YxExTSwgPM>



Ćwiczenia

```
tcga <- readRDS(gzcon(url("http://zbo.ipipan.waw.pl/files/data/TCGA_KM_clinical_data.rds")))
cec <- readRDS(gzcon(url("http://zbo.ipipan.waw.pl/files/data/OrigCecRNA_en.rds")))
```

- Wykonaj selekcję cech na podstawie ich zmienności. Zachowaj w df 3000 najbardziej zmiennych genów.
- Wykonaj drzewo hierarchiczne próbek z TCGA dla 3000 wybranych genów. Wartości genów mają być w dwóch wariantach: 1- jak w bazie danych; 2- zlogarytmowane (funkcja “log”). Opisz czy uzyskane wyniki są takie same. Wytlumacz dlaczego.

Wskazówki: jest wiele narzędzi do wykonania tego zadania. Do policzenia dystansu można użyć funkcji: „dist”, na policzonych dystansach funkcji „hclust” i następnie zrobić wykres „plot”. Dla chętnych polecam wypróbować funkcję „heatmap”.

- W drzewie różnymi kolorami oznacz etykiety pacjentów względem złośliwości „grade”.
- Dla wybranych 5 genów kodujących białka, które regulują proces transkrypcji RNA, sprawdź czy ich ekspresja jest istotnie różna u pacjentów o różnym stopniu złośliwości nowotworu (grade). Wyniki zobrazuj za pomocą wykresów i wykonaj właściwe testy statystyczne.

