实验生物学

Gene Expression Analysis

胡　　苹　研究员
生化细胞所
hup@sibcb.ac.cn

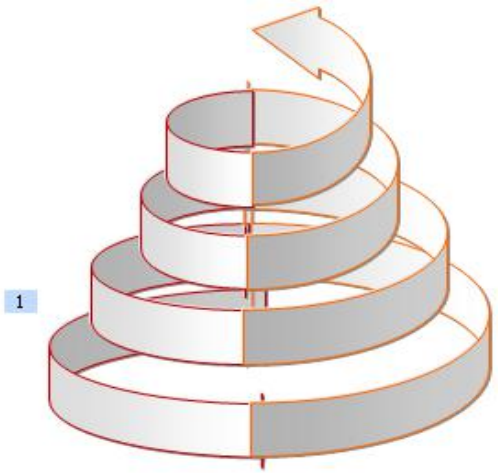
Where does science come from?

Observations
观察

Hypothesis
假说

Test
验证

Interpretation
解释



What Is Science

科学必须是能够客观验证的；
Science must be able to be tested objectively;

科学必须是能够证伪的；
Science must be able to be falsified;

科学必须是可重复的。
Science must be able to be reproduced.

实验逻辑就是科学逻辑在每一个实验中的体现

Biology is all about comparisons.

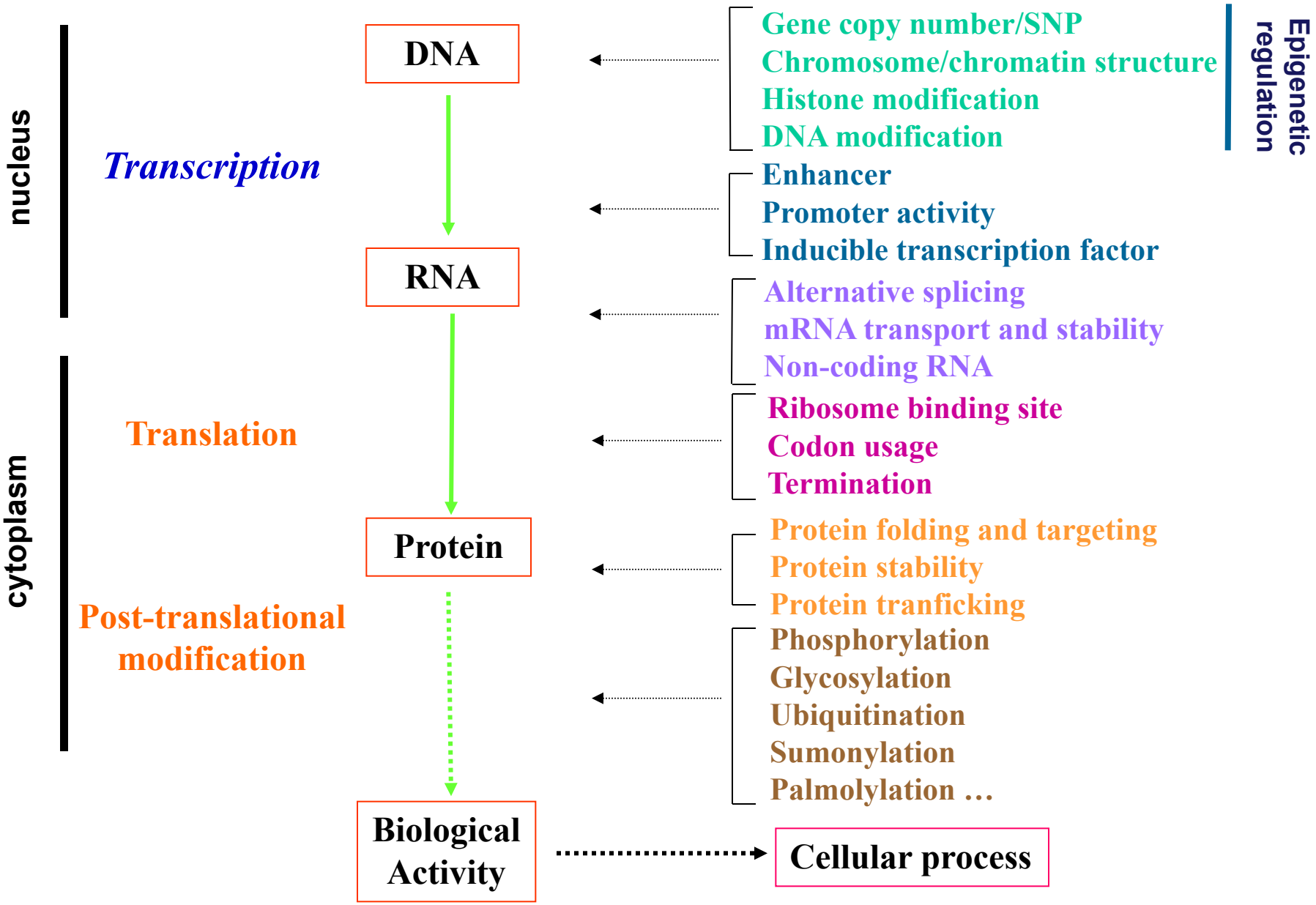
- **Control is the most important part of all experiments;**
- **Logic is the most important part of all experiments.**

Biology is a fast evolved field

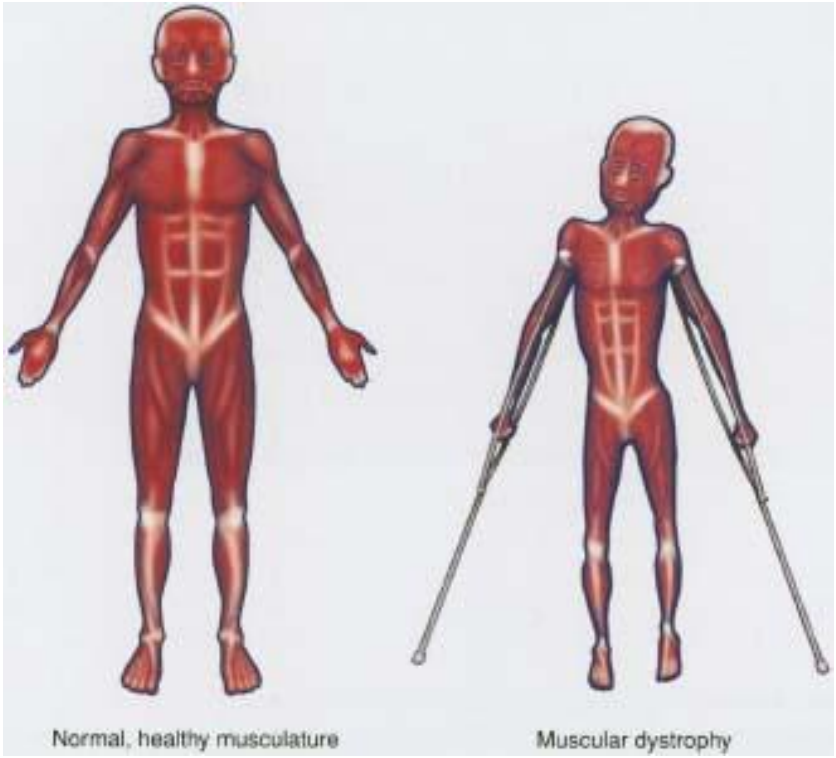
- Learning the principles of the techniques is the most important part of skill learning.
- Happy to learn all new techniques is the most valuable character of a biologist.

Gene expression changes are the driving force and consequences of all biological processes.

Gene expression is regulated at multiple levels



Observations: Muscular dystrophy passes in families



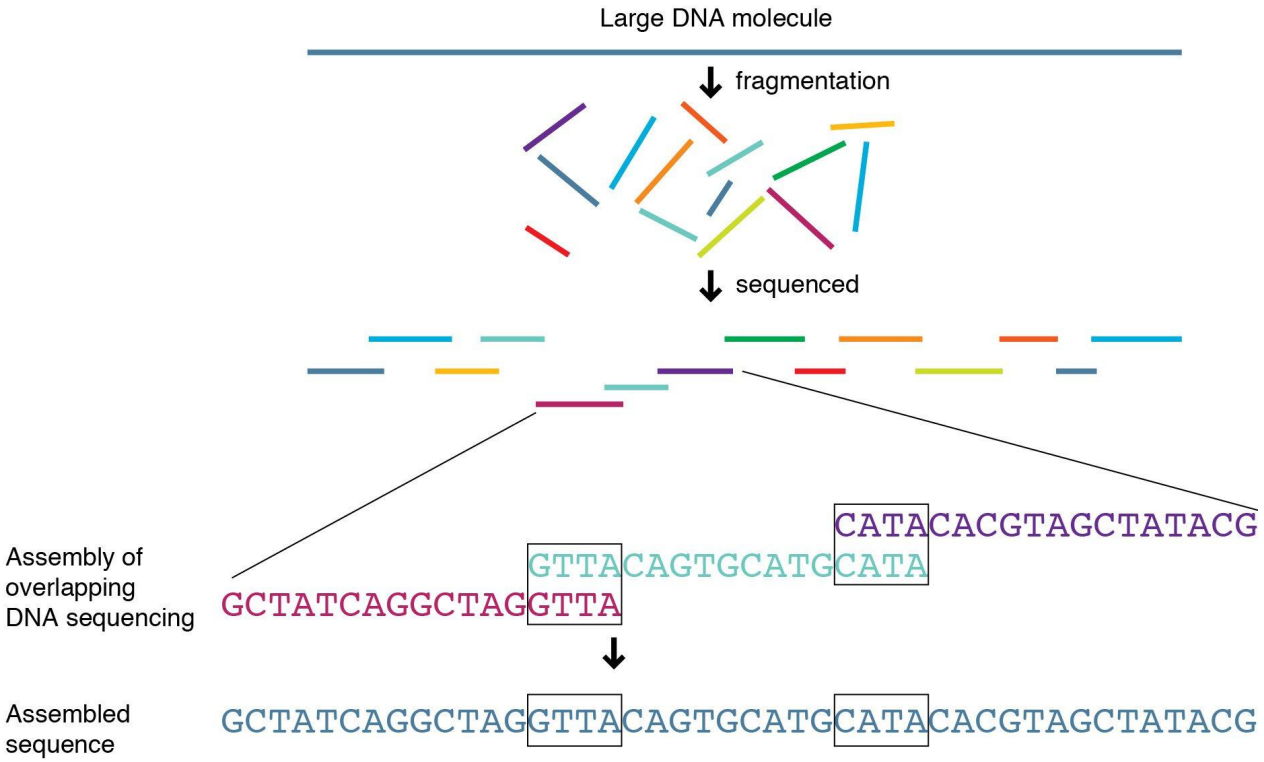
First, consistent and repeatable clinical observations (phenotype);
Second, identify the tissue/cell type affected most by the disease.

Hypothesis: There are genes controlled the muscular dystrophy phenotype.

How to test the hypothesis?

Every hypothesis must be tested by multiple methods.

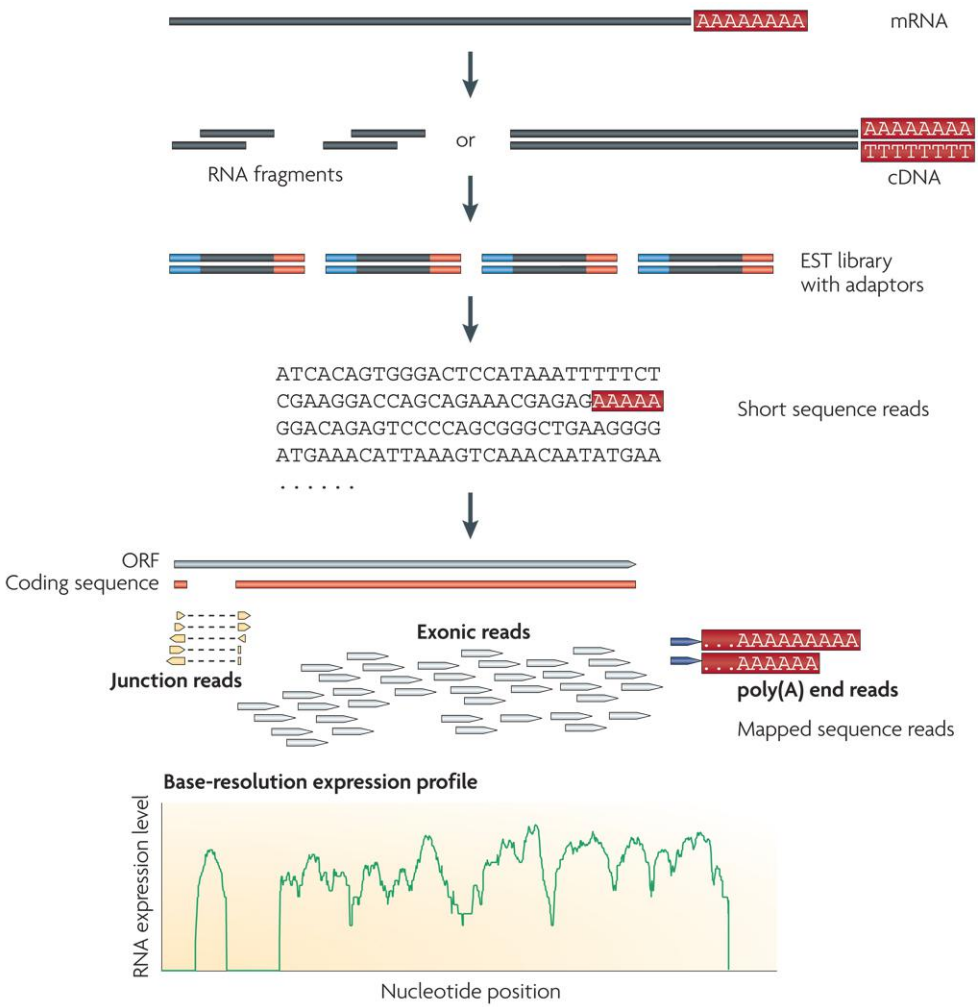
The whole genome sequencing



Compare mutants and Wild type (Patients and healthy relatives)

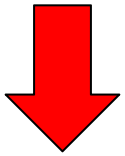
The frame shift of gene M is among another 2000 mutants identified in patients.

Identify the differentially expressed genes—RNA sequencing



The expression of M is significantly decreased.

By RNA sequencing, gene D was identified to be the key differentially expressed gene in muscular dystrophy patients.

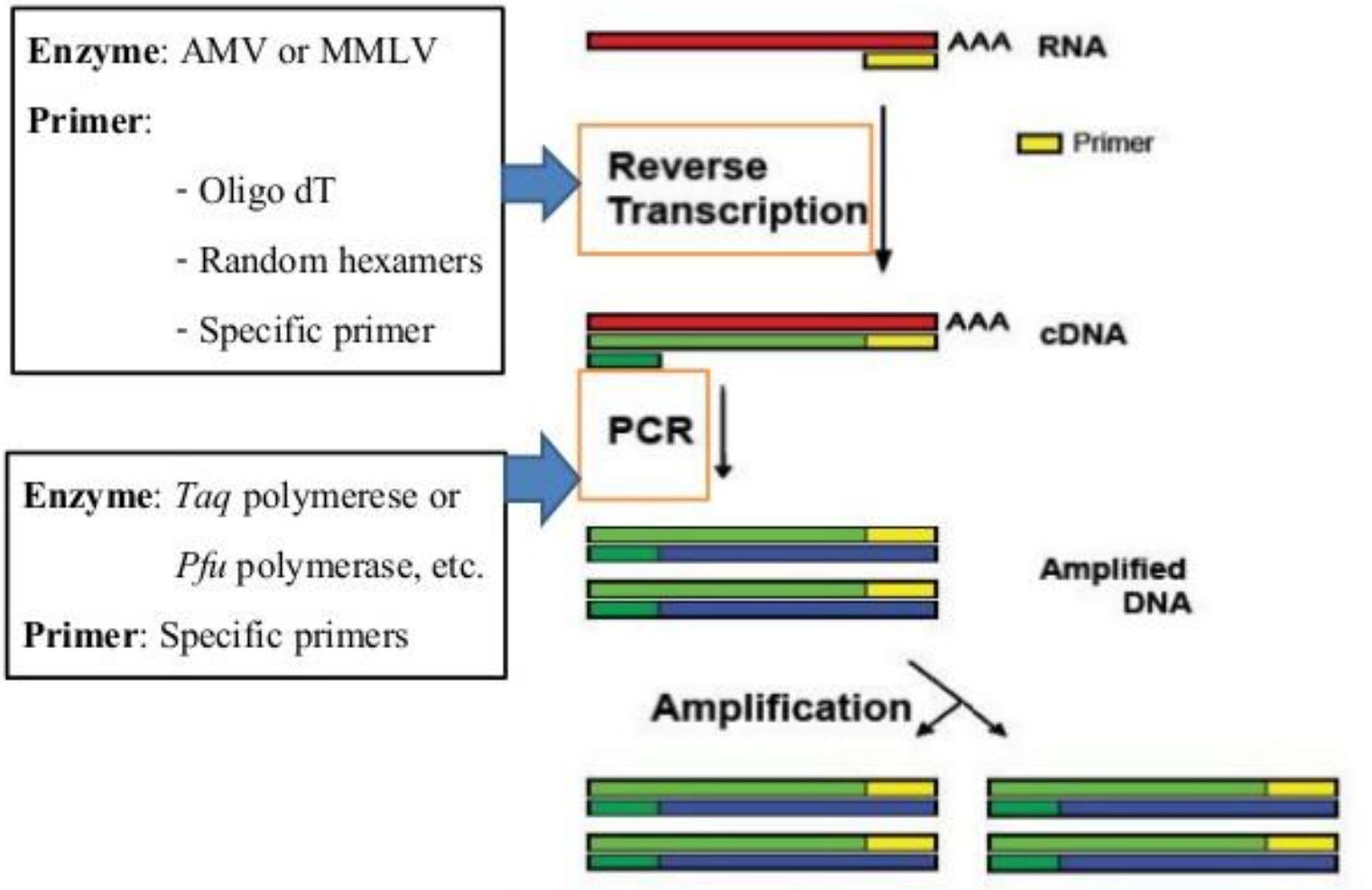


Confirm the genome wide analysis by measuring the mRNA and protein level of gene D.

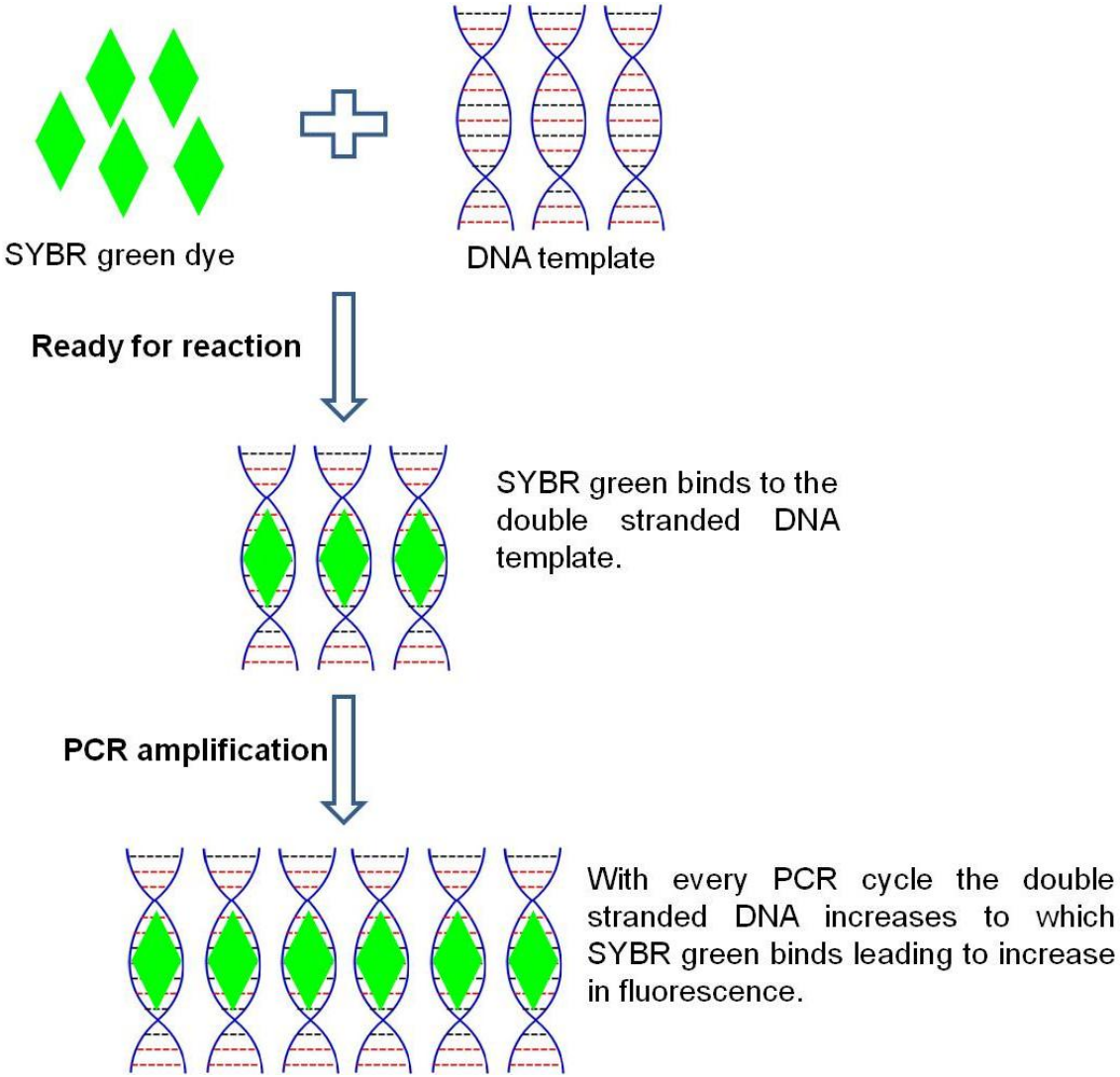
How to analyze the transcribed mRNA levels of gene M

- **RT-PCR**
- **RT-qPCR**

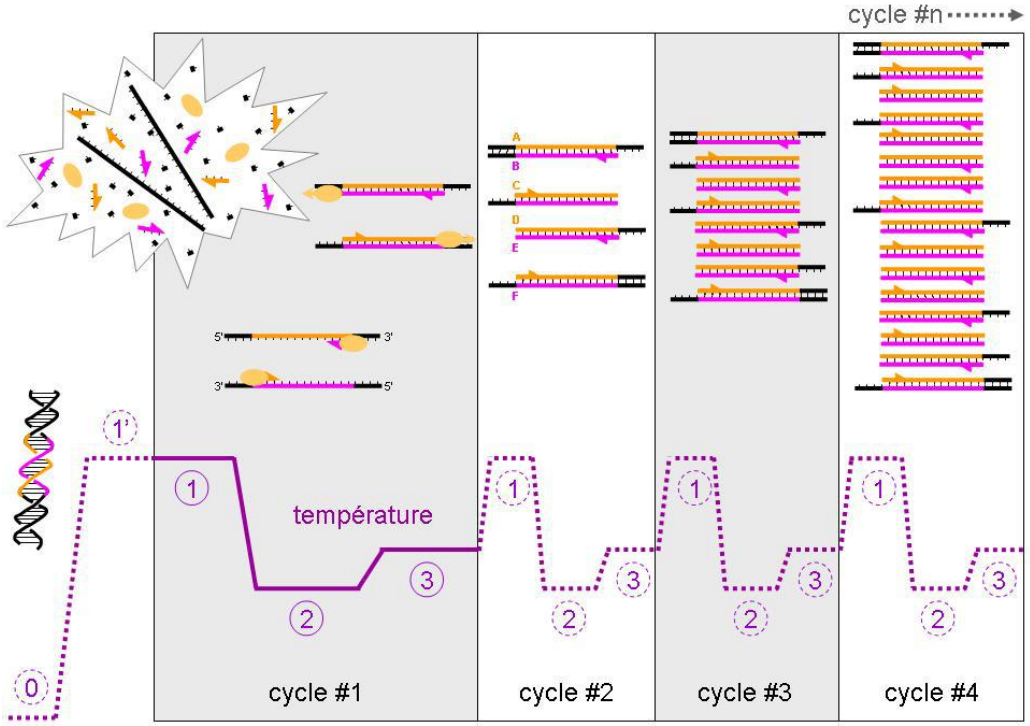
RT-PCR

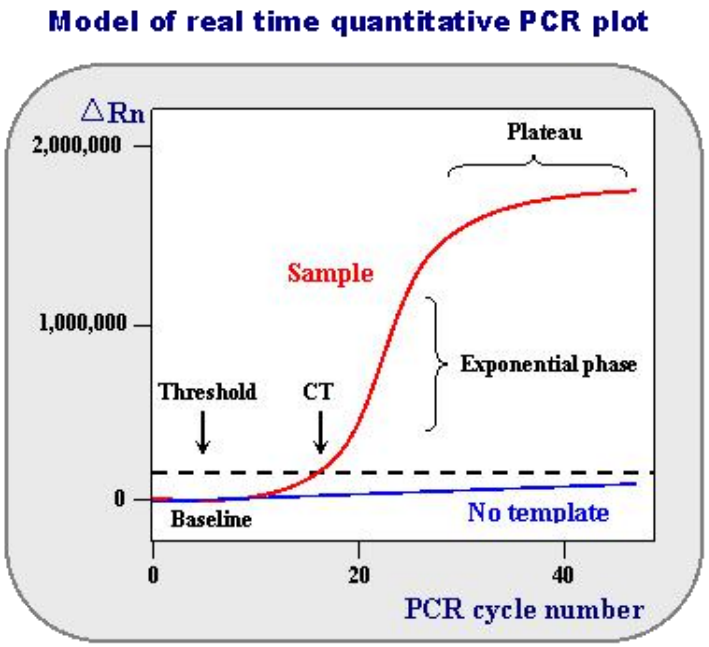

Enzyme: AMV or MMLV
Primer:
- Oligo dT
- Random hexamers
- Specific primer
Reverse Transcription
AAA
RNA
Primer
AAA
cDNA
PCR
Enzyme: Taq polymerese or
Pfu polymerase, etc.
Primer: Specific primers
Amplified DNA
Amplification

SYBR green binds double strand DNA



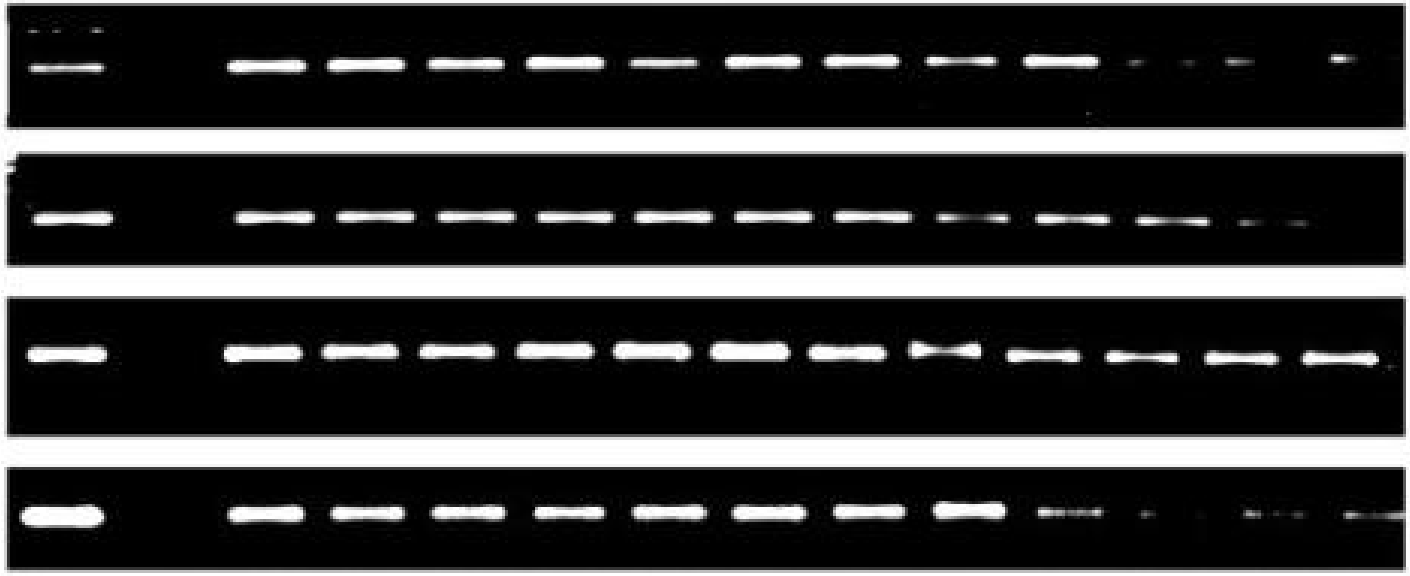
Quantitative PCR can quantificate DNA





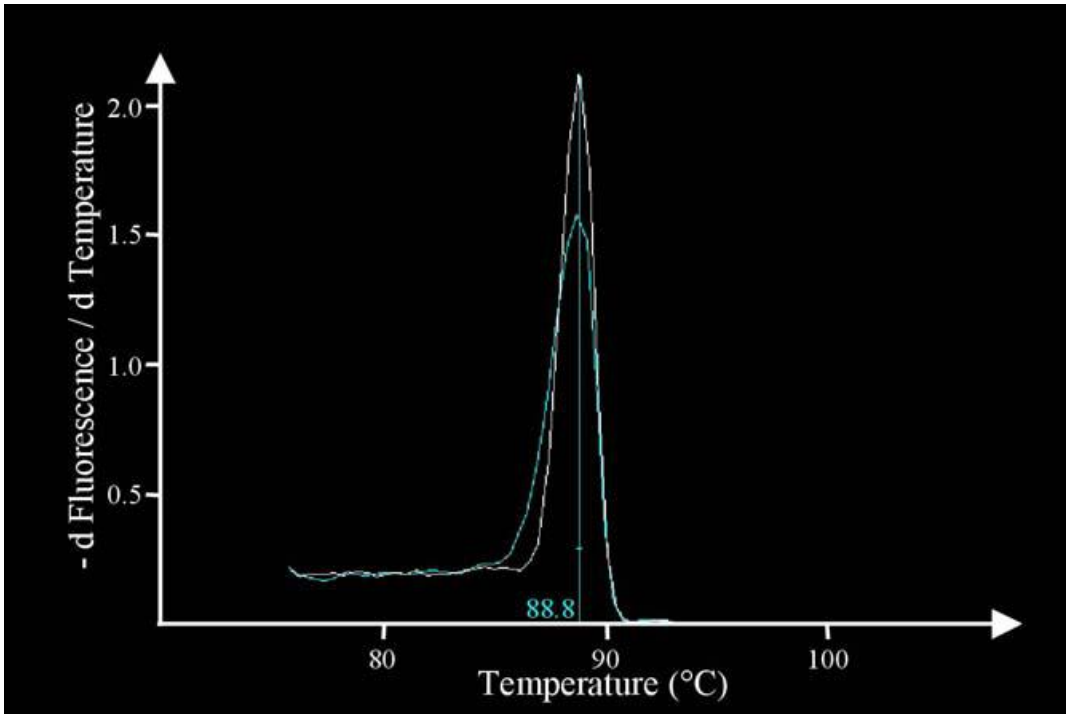
RT-qPCR can quantitate the amount of DNA obtained from reverse transcription, therefore, quantificate RNA level.

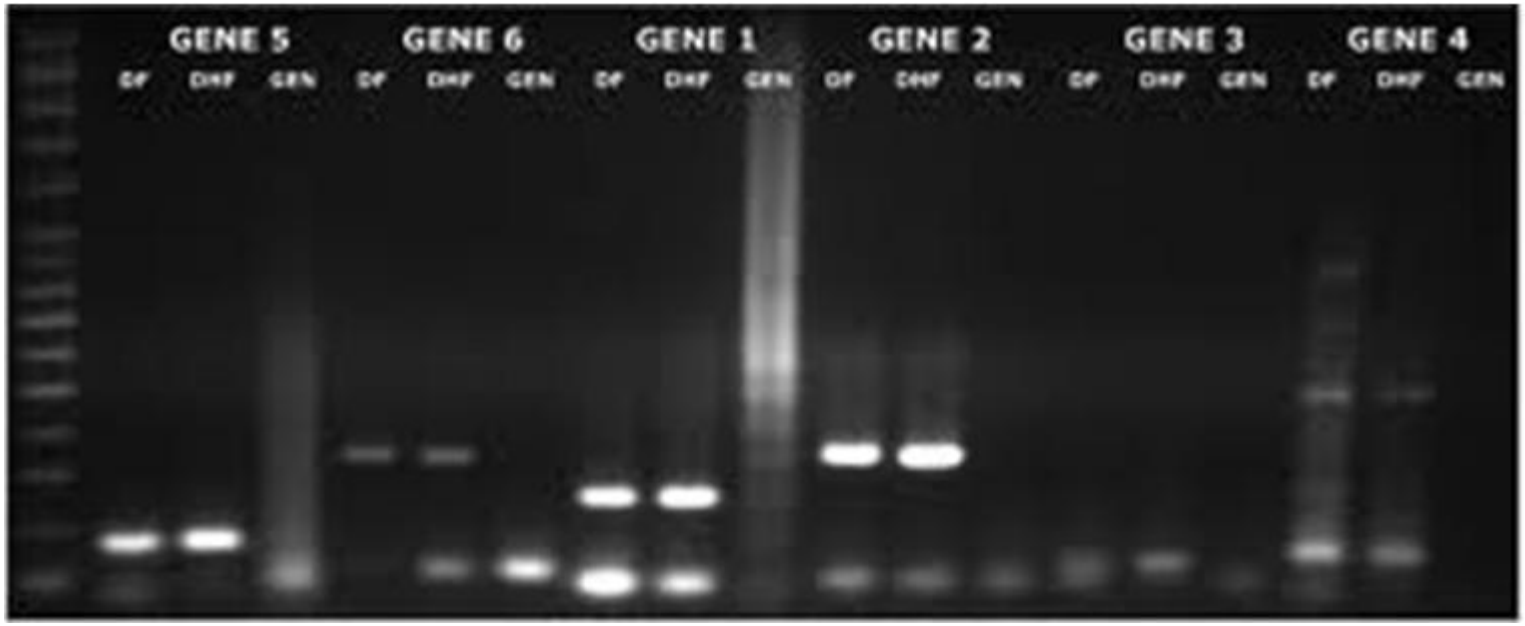
RT-PCR combined with agarose gel electrophoresis is semi-quatative



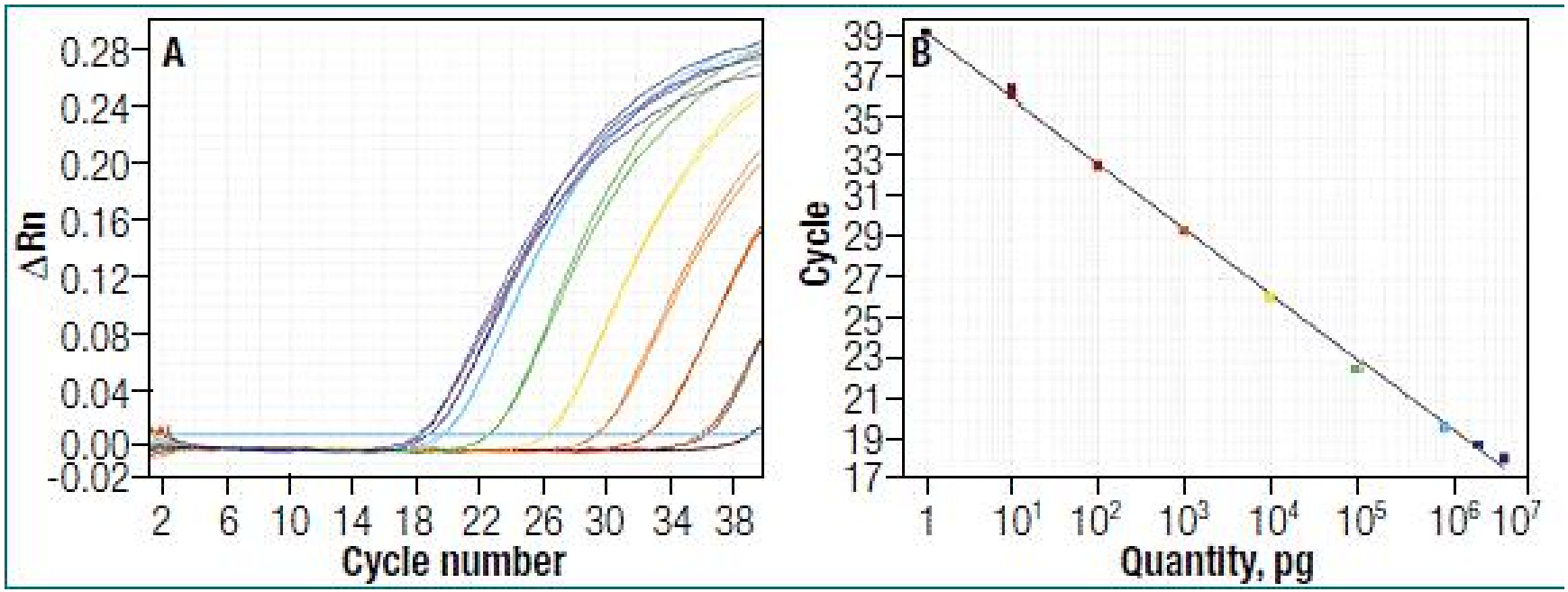
Syber green, PCR optimization & melting curve

Incoorperation of a molecule that flouresces only when bound to dsDNA. Therefore as product is made increased flourecences can be detected. The advantage of this techniques is that it is very cheap but has a detection limit depenedent on the formation of primer artifacts. As the detection system will give a signal with non-specific products and contaminated genomic DNA your PCRs must be extremely well optimizied. It also allows you to determine a melting curve for your product, as shown below:

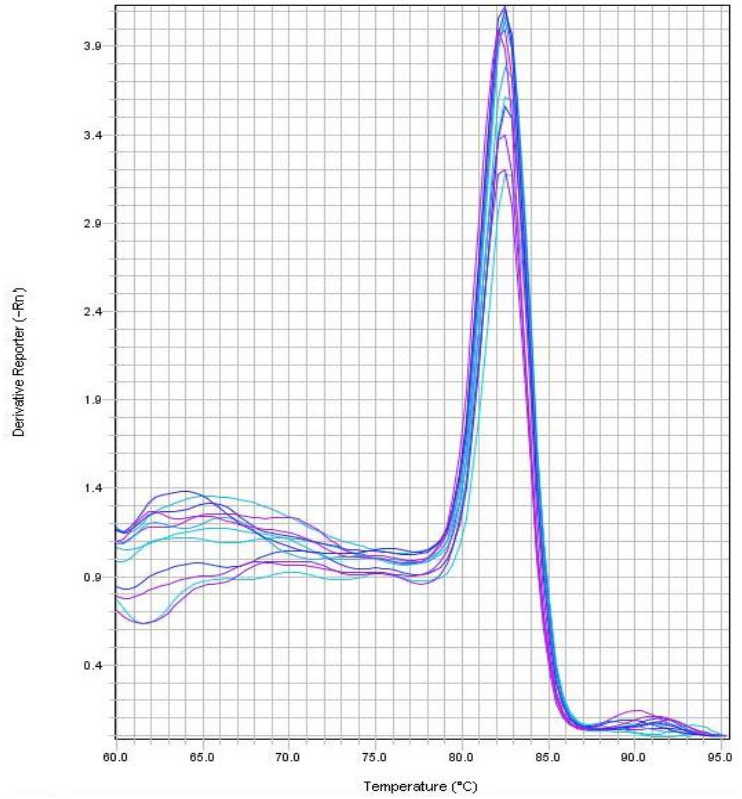


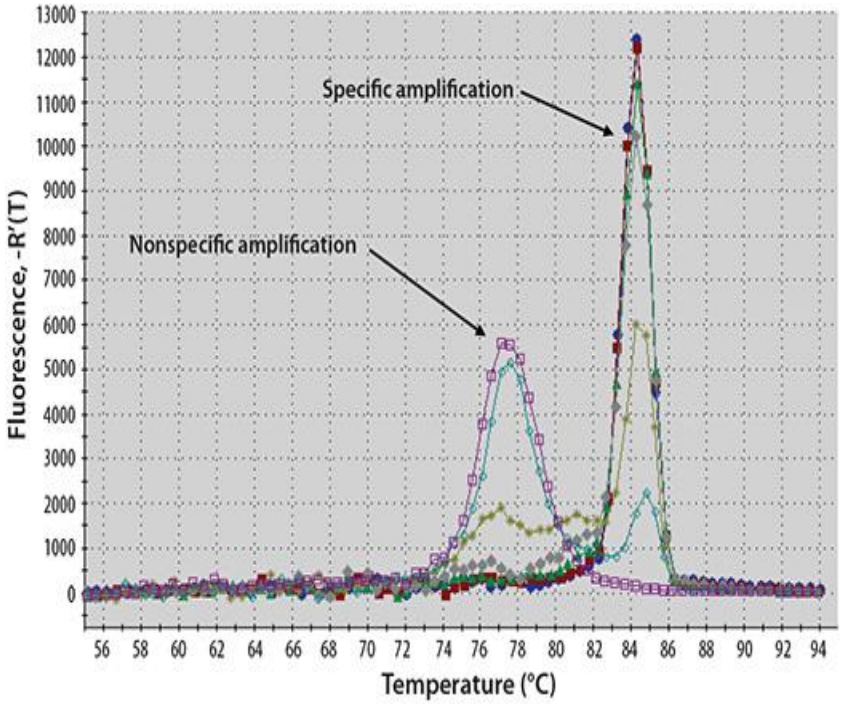

GENE 5
GENE 6
GENE 1
GENE 2
GENE 3
GENE 4
DF
DHF
GEN

Standard curve is required for a legitimate qPCR



Examples of melting curve





Common Standards for Real Time PCR

--- Glyceraldehyde-3-phosphate dehydrogenase mRNA (GAPDH)

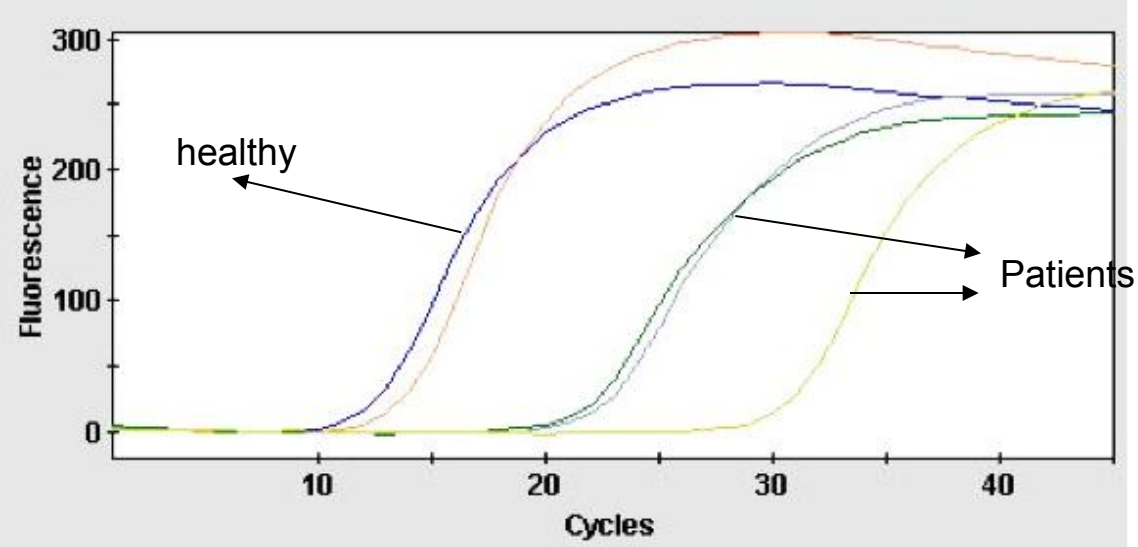
--- Beta-actin mRNA

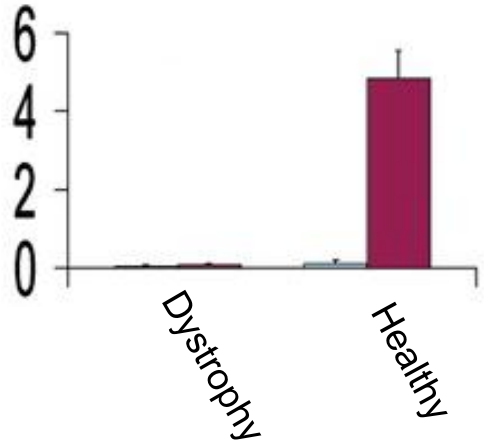
--- MHC I (major histocompatability complex I) mRNA

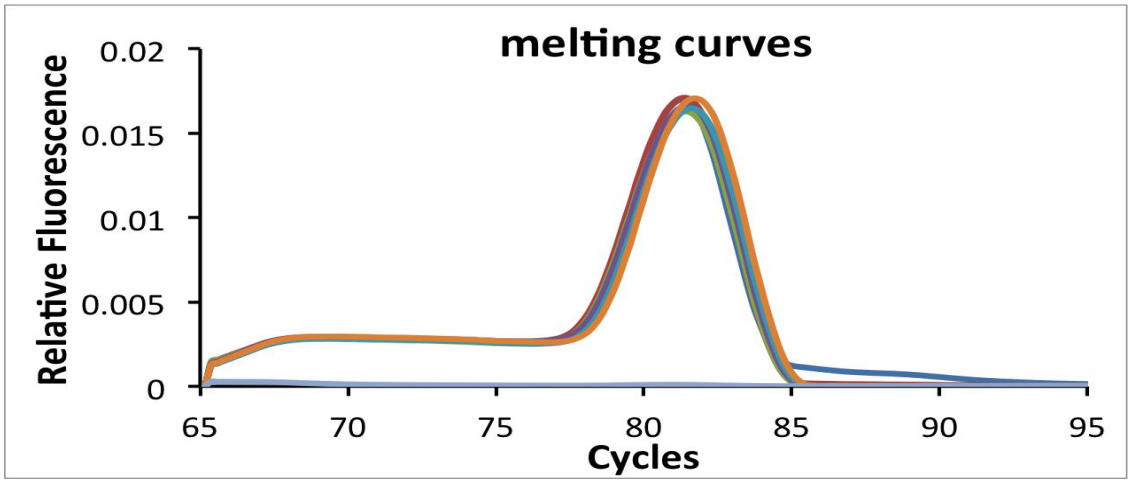
--- Cyclophilin mRNA

--- mRNAs for certain ribosomal proteins RPLP0
28S or 18S rRNA

mRNA level of gene D is decreased in dystrophy patients

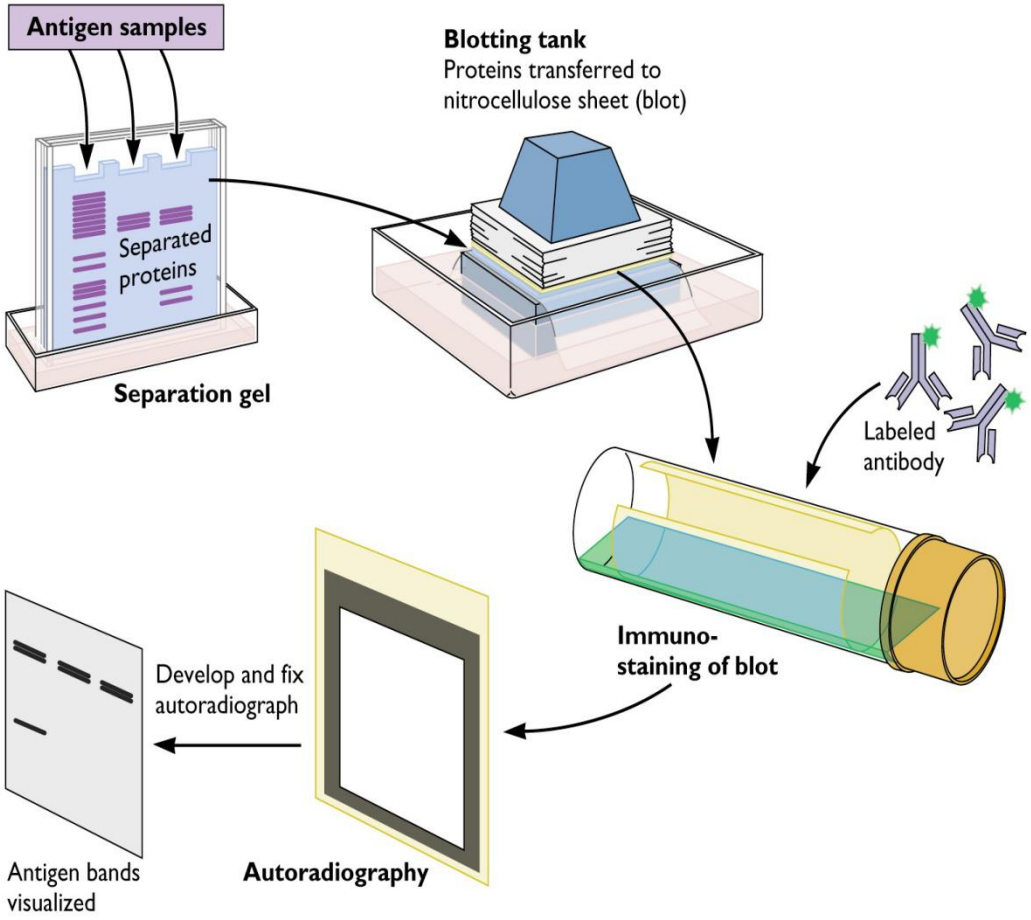


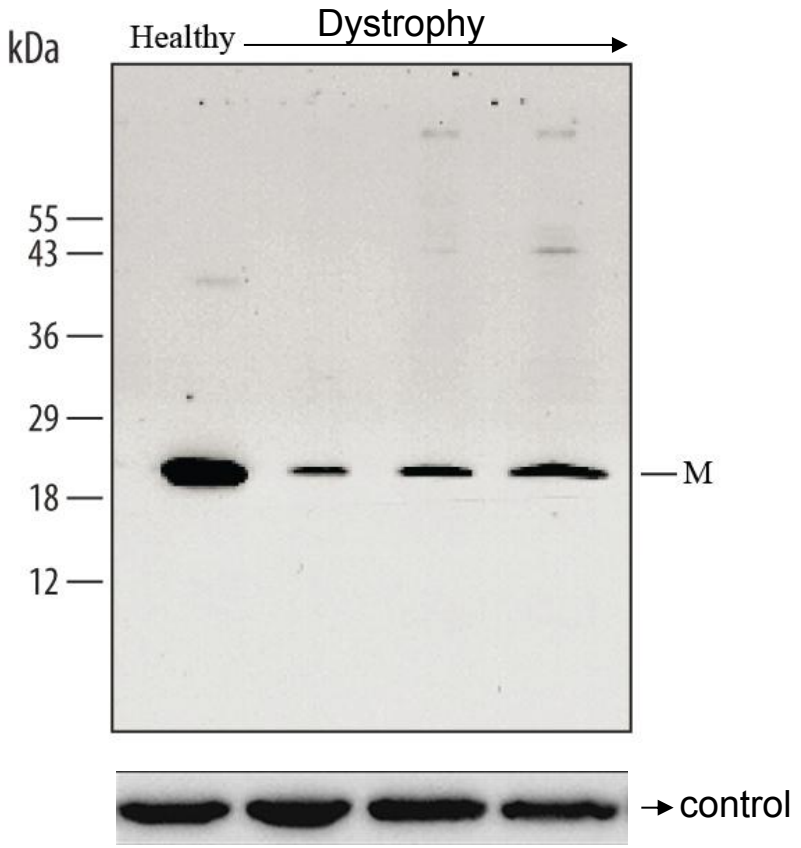




mRNA level of M decreased as indicated by RT-qPCR

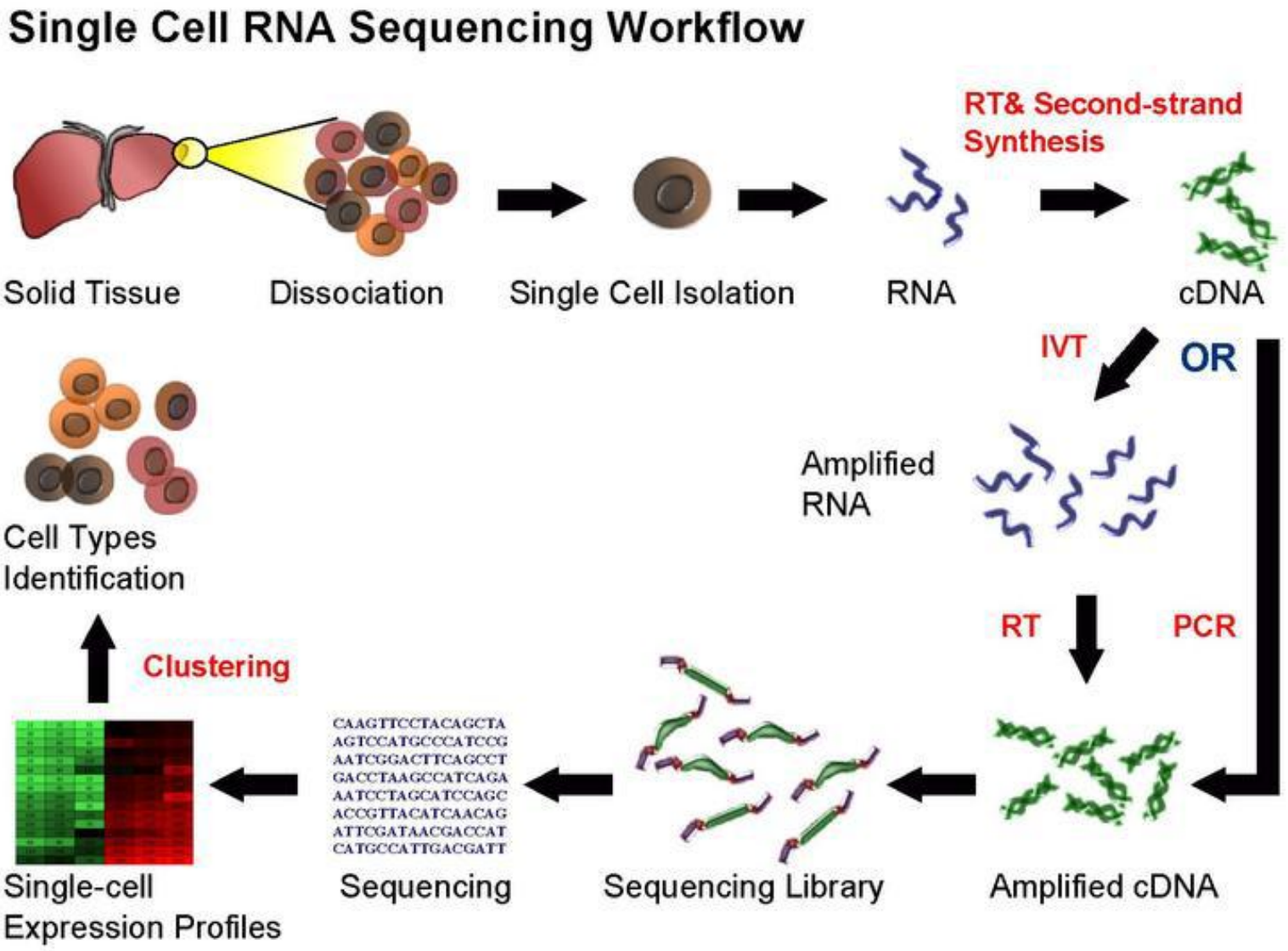
Western Blot



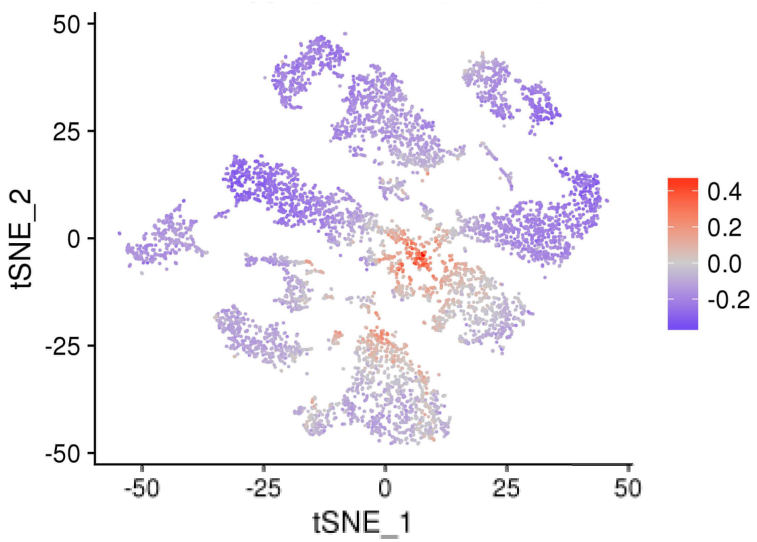


Protein level of M decreased as indicated by RT-qPCR

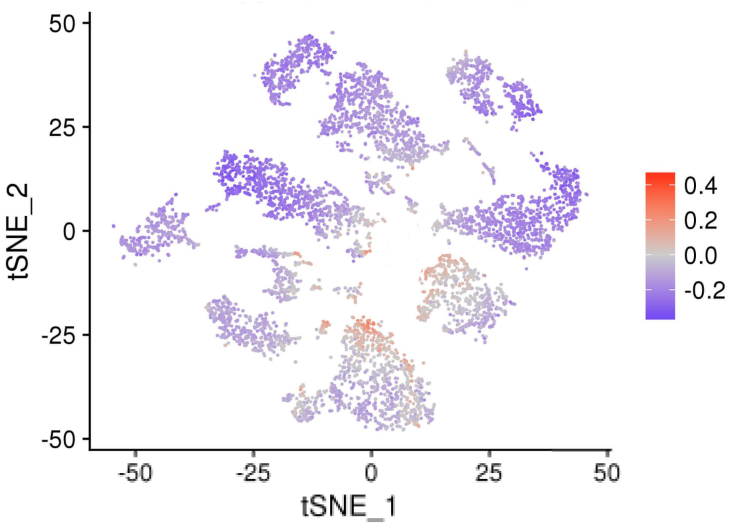
Single cell sequencing



Single cell sequencing results



Healthy



Patients

The number of the cell type expressing M decreased greatly in patients.

Results:

- M is mutated in patients.
- M is decreased at both mRNA and protein level in patients.
- Less number of cells expressing M in patients.

Conclusion 1

M is an important factor for dystrophy.

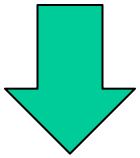
Conclusion 1

M is an important factor for dystrophy.

Next Question

Why is M important for muscle dystrophy?

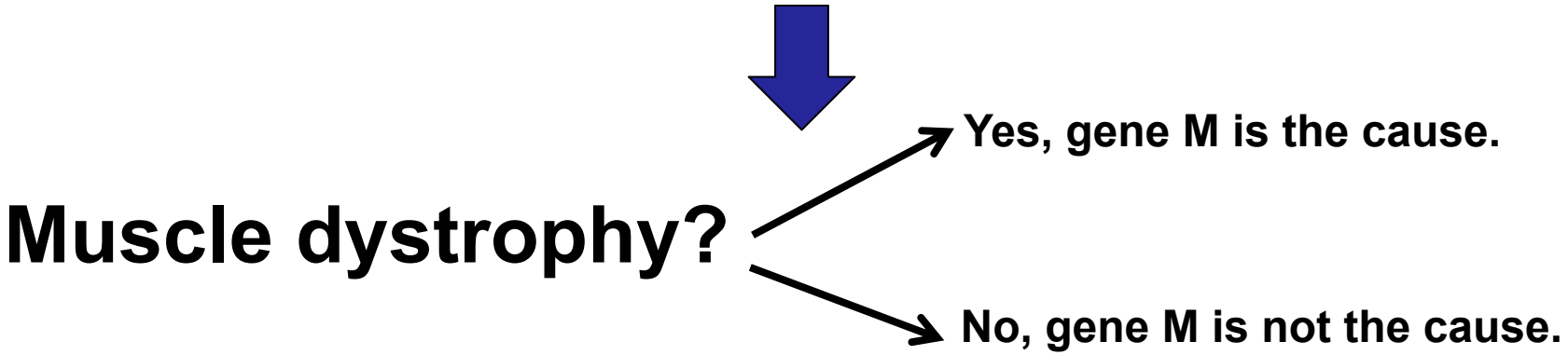
Next question:
Why is M important for muscle dystrophy?



Does the decrease of M causes muscle dystrophy?

Reduce the expression level of M in muscle cells

RNAi, CRSPR/Cas9



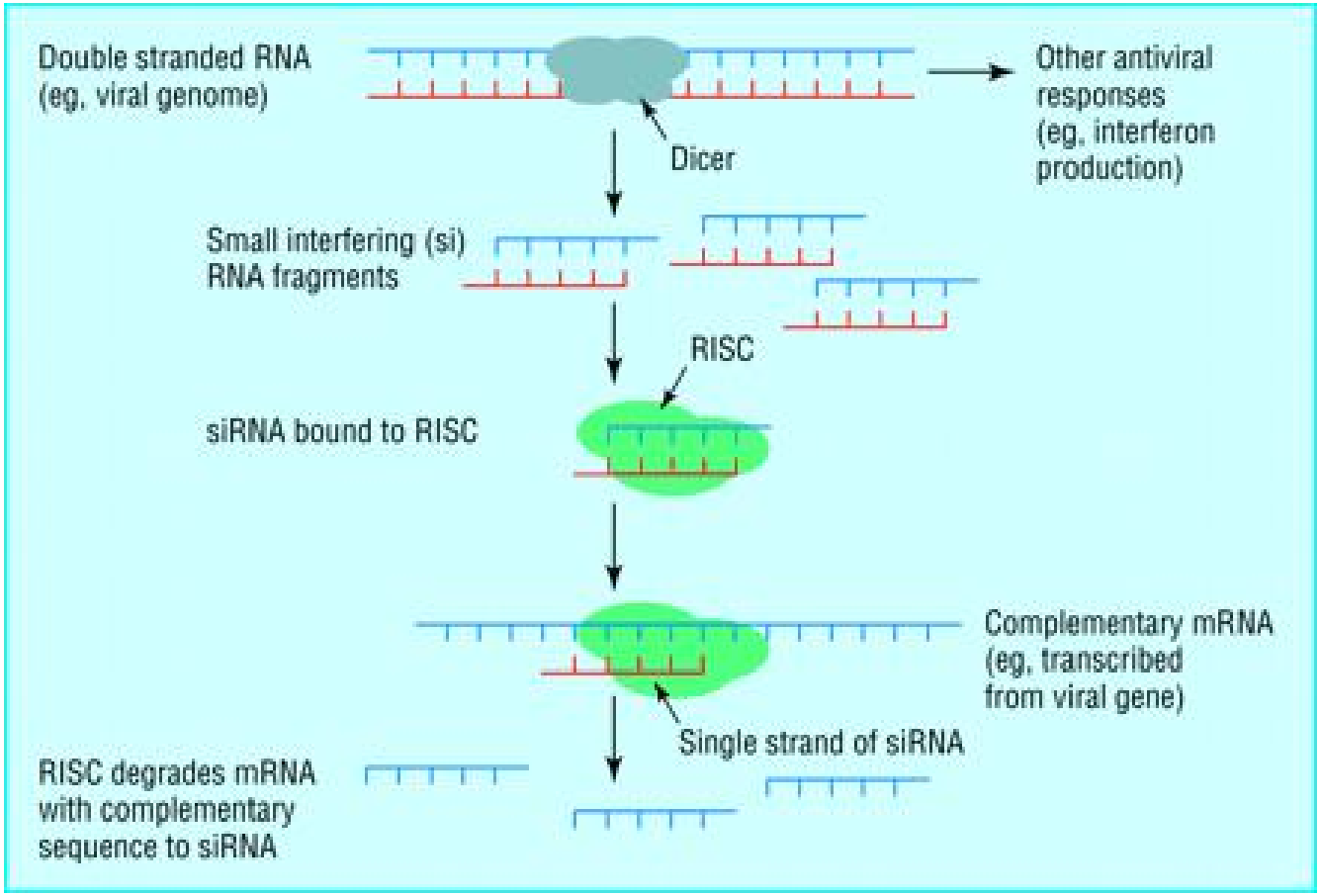
RNA interference (RNAi)

- **dsRNA, often encountered by cells during viral infection, induce robust immunological response in Mammalian organisms.**

- **The properties of RNAi demands the existence of cellular mechanisms that initiate and amplify the silencing signal, and suggest that the RNAi mechanism represents an active organismal response to foreign RNA.**

- **RNA interference has been used as a tool for reverse genetics in many different organisms including: zebrafish, planaria, hydra, fungi, *Drosophila*, and mammals**

Natural mechanism of RNA interference



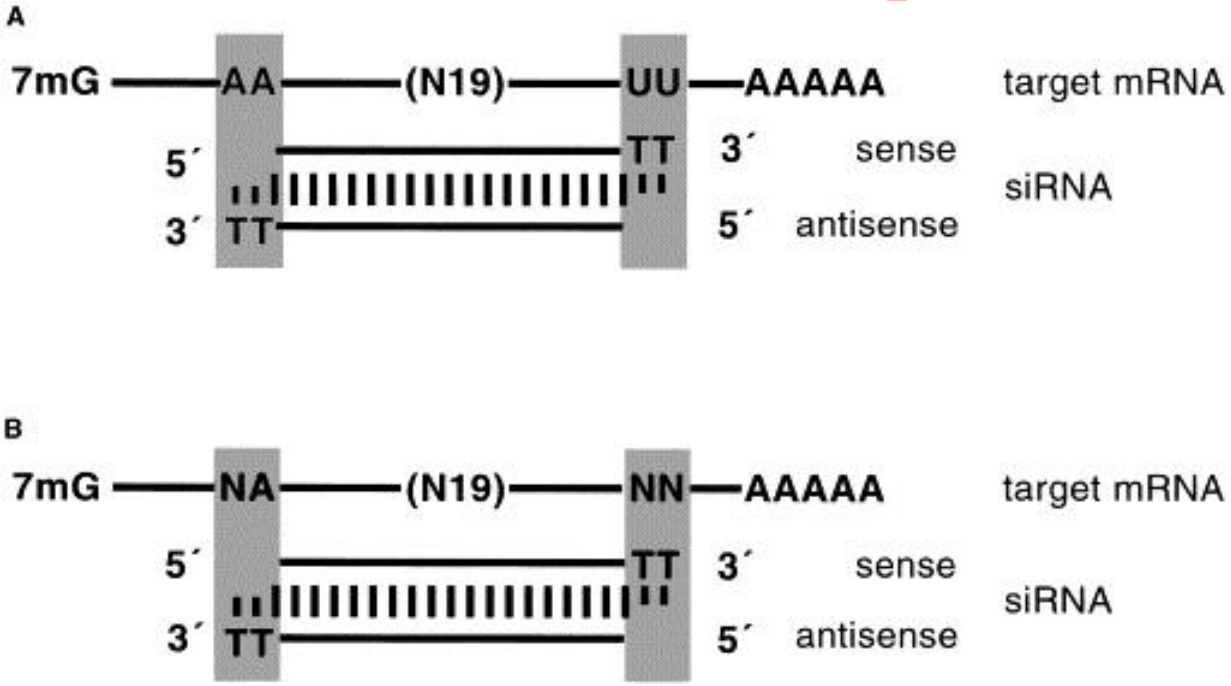
Dicer: RNase III enzyme
RISC: RNA-induced silencing complex

siRNA design

There are expanding libraries of validated, commercially available siRNAs directed toward some commonly targeted genes. These may be of use, if available. However, if the gene of interest has not been targeted using siRNA before, a novel siRNA must be developed.

Selection of the targeted region is currently a trial-and-error process, but with the likelihood of 80–90% success , given a large enough random selection of target genes

Selection of siRNA duplexes



[1]Select the target region from the open reading frame of a given cDNA sequence preferably 50 to 100 nt downstream of the start codon. Avoid 5′ or 3′ untranslated regions (UTRs) or regions close to the start codon as these may be richer in regulatory protein binding sites.

[2]Search for sequences 5′-AA(N19)UU . Choose those with approximately 50% G/C content (from 32 to 79% G/C content).

[3] Blast-search the selected siRNA sequences against EST libraries or mRNA sequences of the respective organism to ensure that only a single gene is targeted.

[4] It may be advisable to synthesize several siRNA duplexes to check the silencing effectivity. A nonspecific siRNA duplex is needed as control.

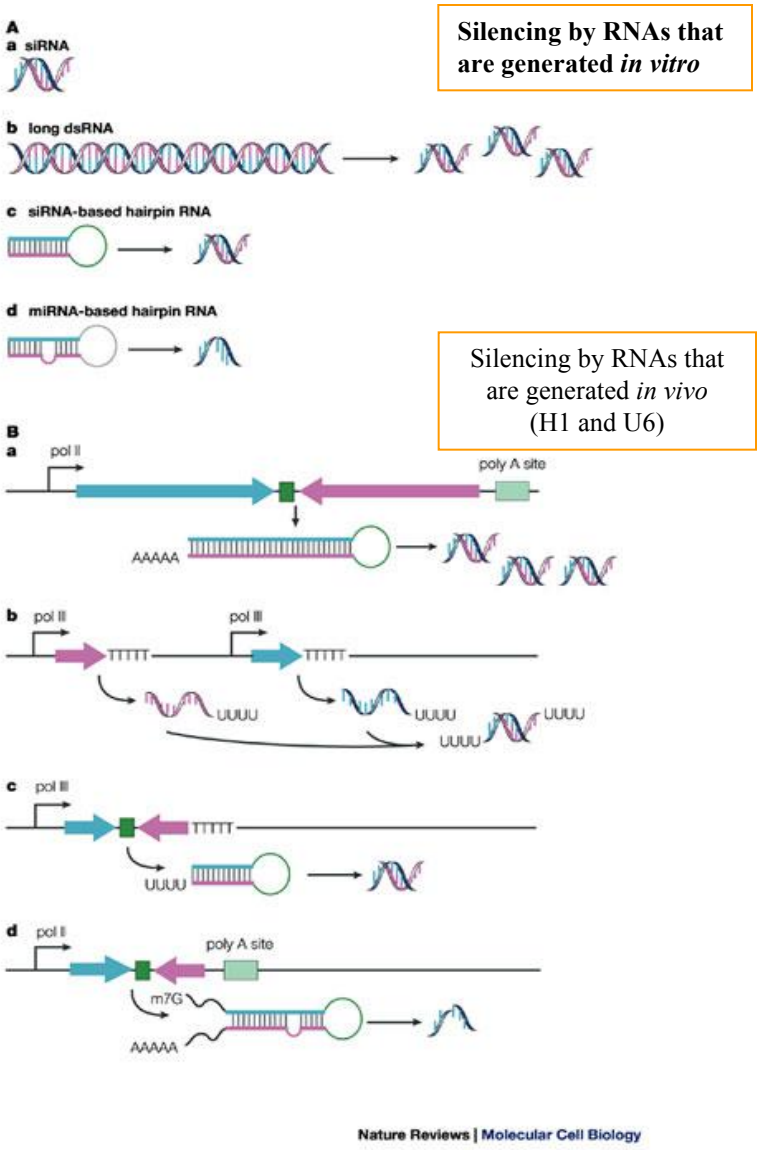
Methods to Produce siRNAs

Transient transfection

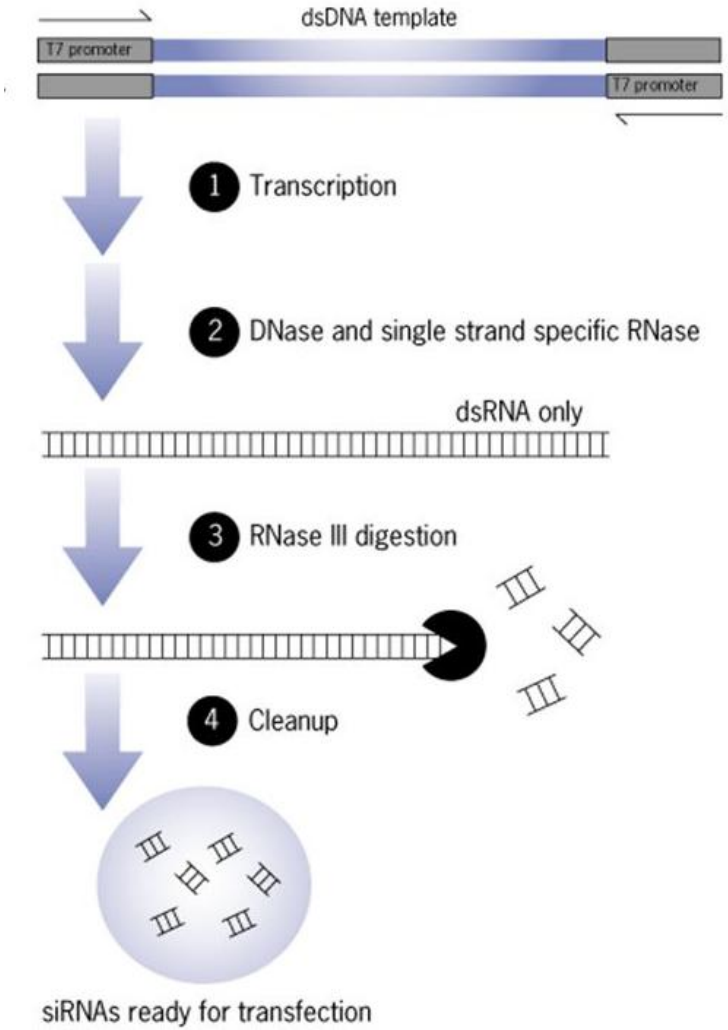
- **Chemical Synthesis siRNA**
- **In vitro transcription**

Stable transfection

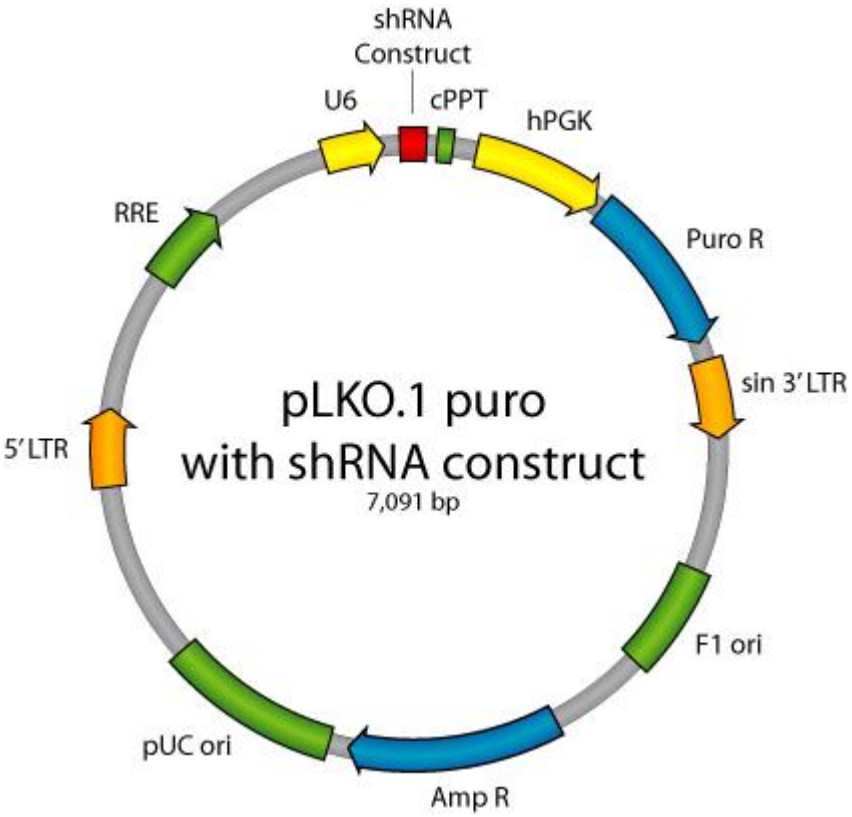
- **Expression in cells from a PCR-derived shRNA expression cassette**



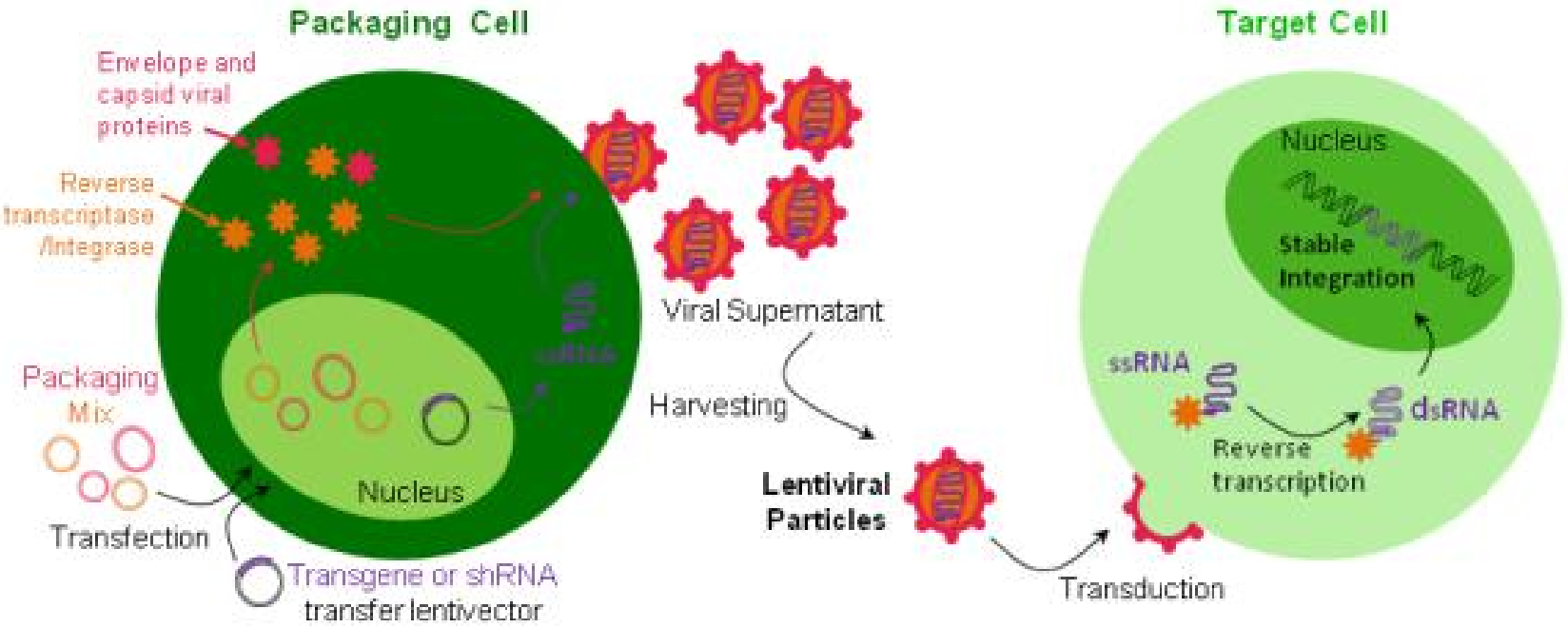
In vitro transcription to generate siRNAs



Example of shRNA viral vector



Virus packaging



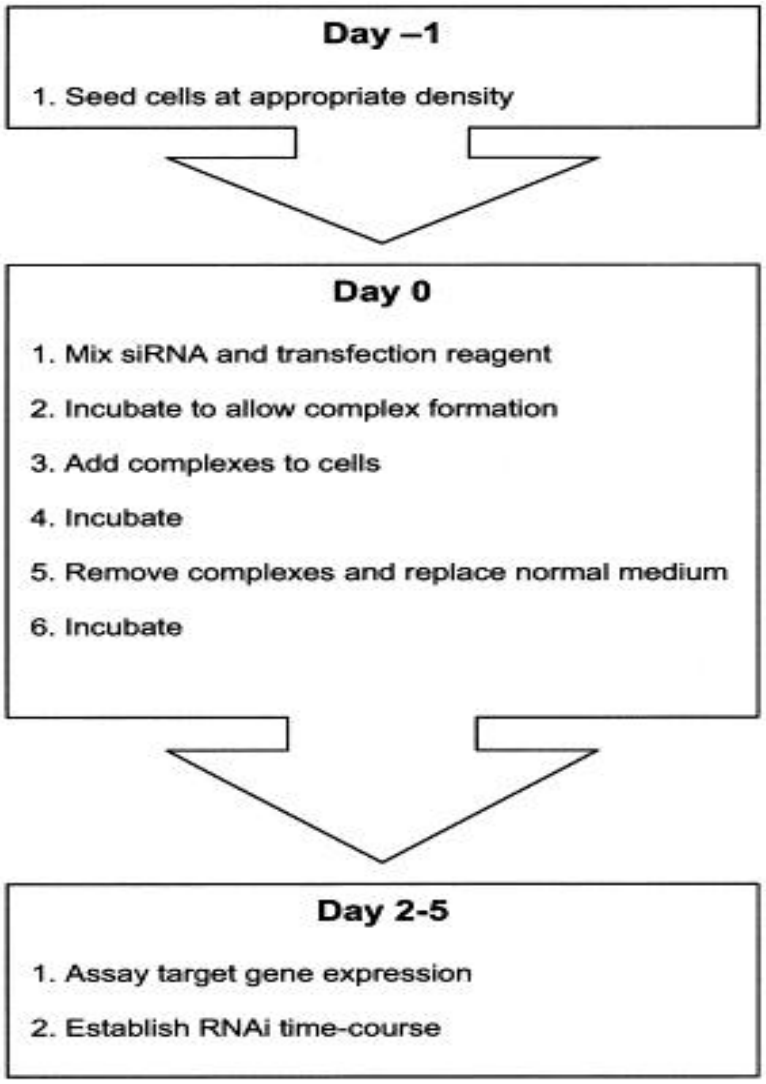
Methods of siRNA delivery

- **Introduction of siRNA is dependent upon how it is produced and the target cell type or organism**
- ***C.elegans* -- injection, soaking, or feeding**
- ***Drosophila* cells -- exposure through culture medium**
- **Mammalian cells -- transfection or electroporation or retroviral-mediated stable incorporation**

Advantages and Disadvantages of Different siRNA/shRNA Delivery Strategies

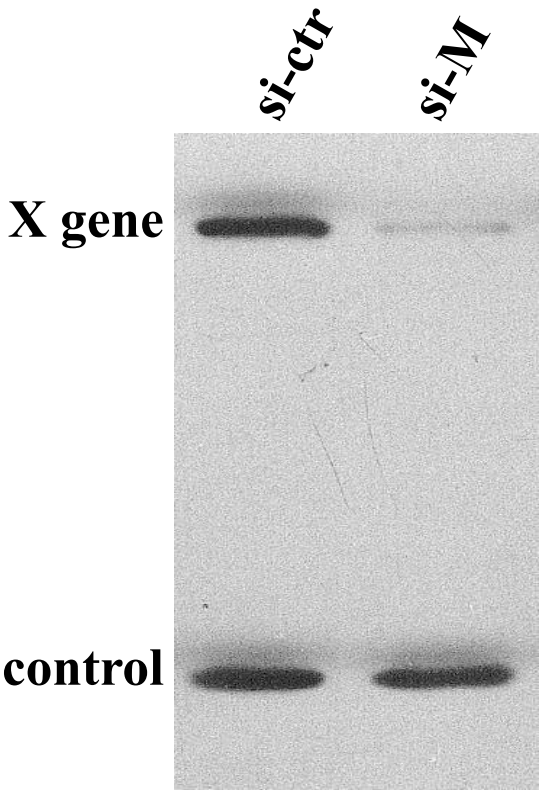
	Advantage	Disadvantage
Chemical and *in vitro* enzymatic synthesis	Rapid synthesis High purity using chemical synthesis	Transient RNAi expensive for multiple siRNA
DNA plasmid vector or cassette	More economical for multiple sequences Stable RNAi achievable using selection marker	More labor intensive to generate Transfection-dependent
Virus-mediated	May be effective in cells resistant to transfection with dsRNA and plasmids Integration produces stable RNAi even in the absence of a selection pressure	Random insertion mutation

Schematic for a typical RNAi experiment using chemically synthesized siRNA

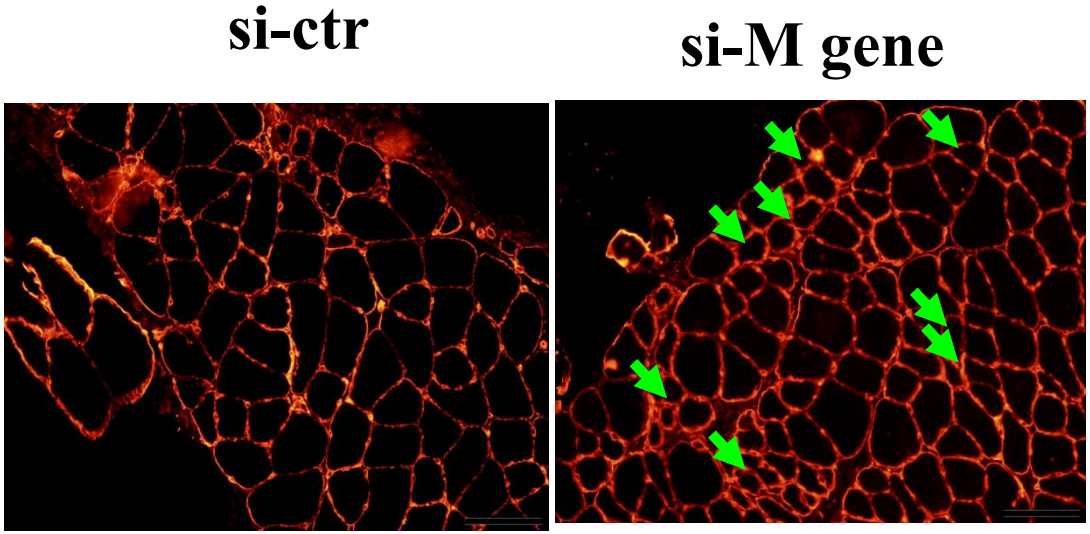


Retrovirus-mediated RNAi knockdown of M gene blocked adipogenesis

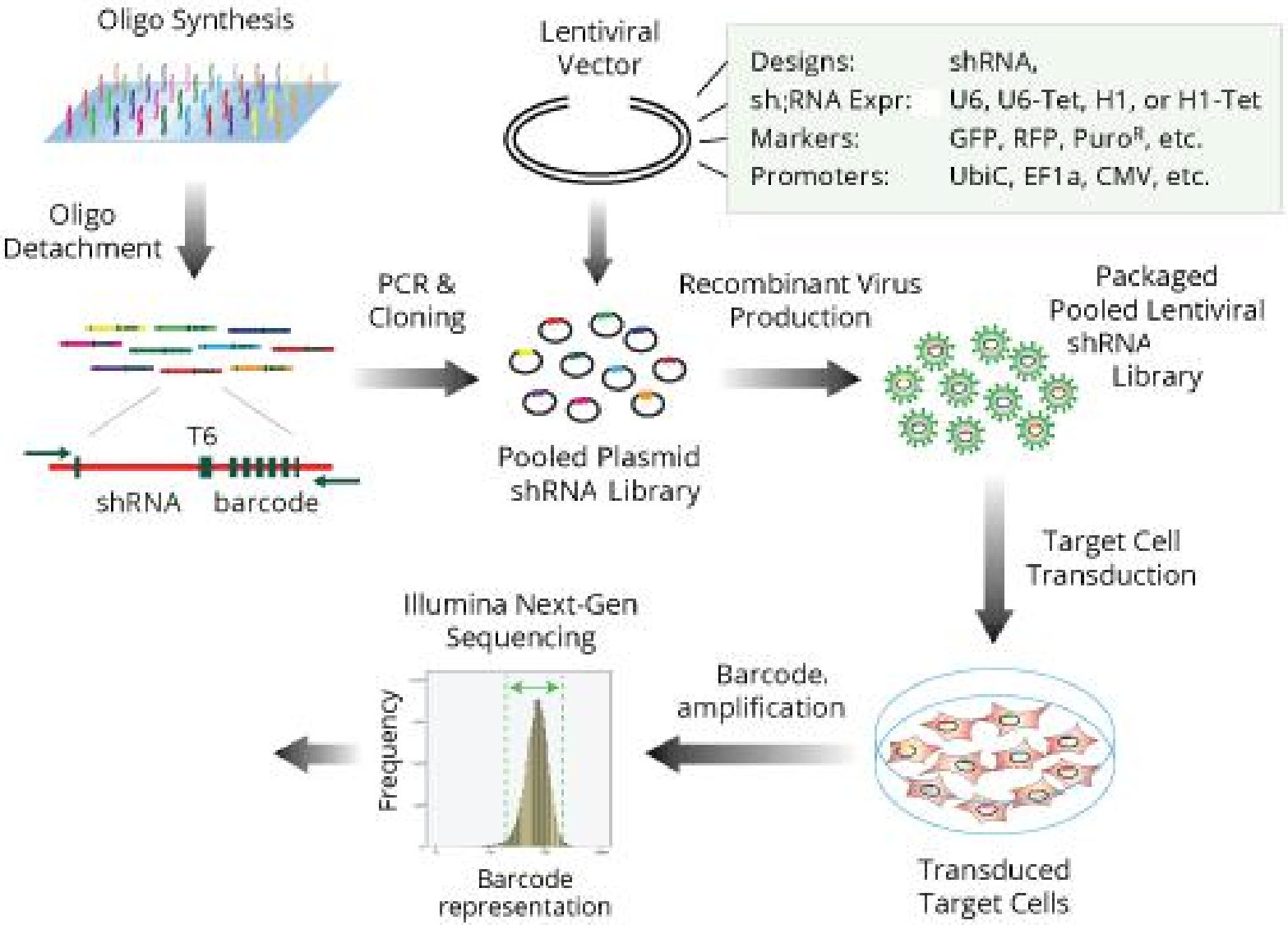
Western Blot



Oil Red O staining for adipocyte



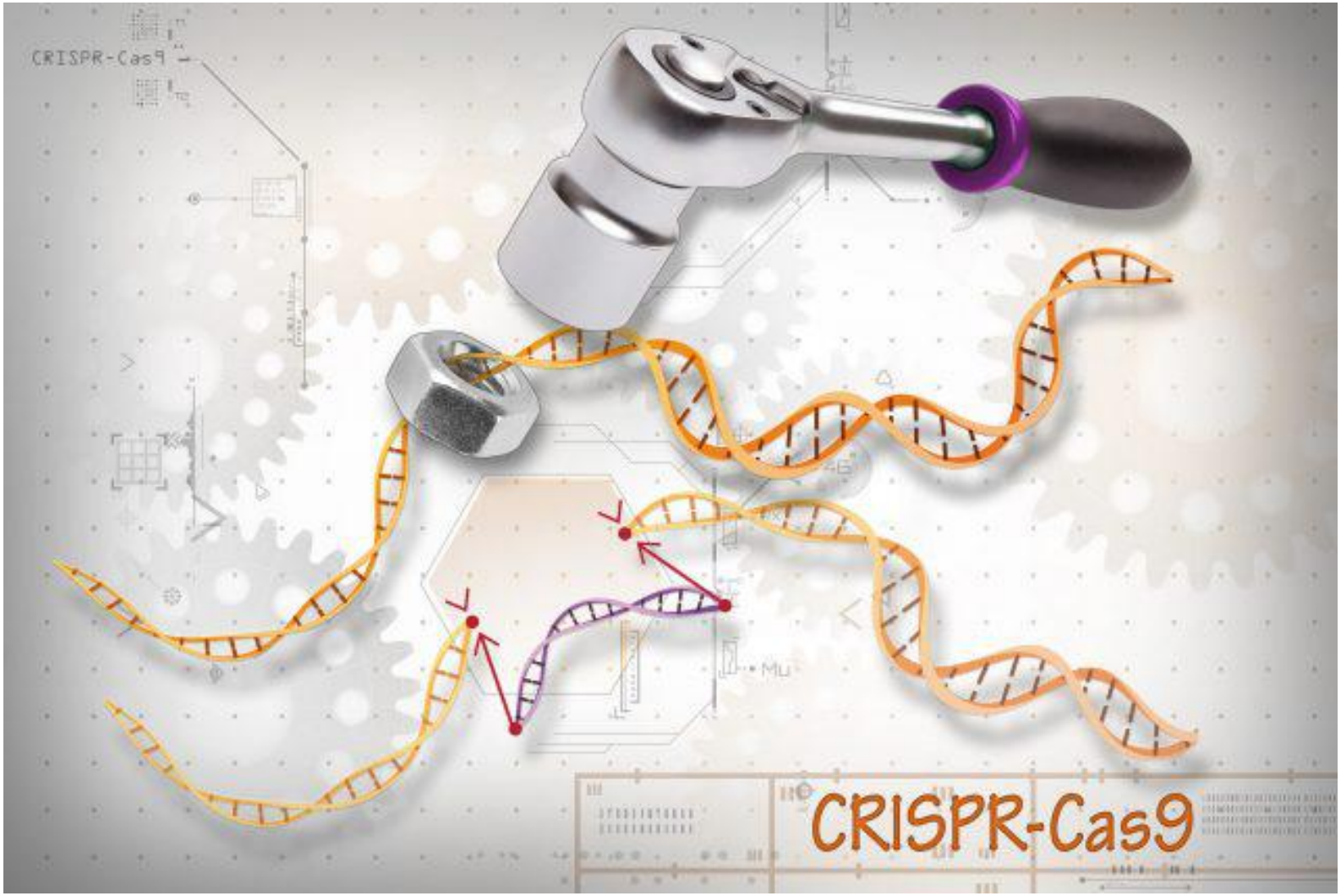
Genome wide screening of shRNA library



Disadvantage of RNAi

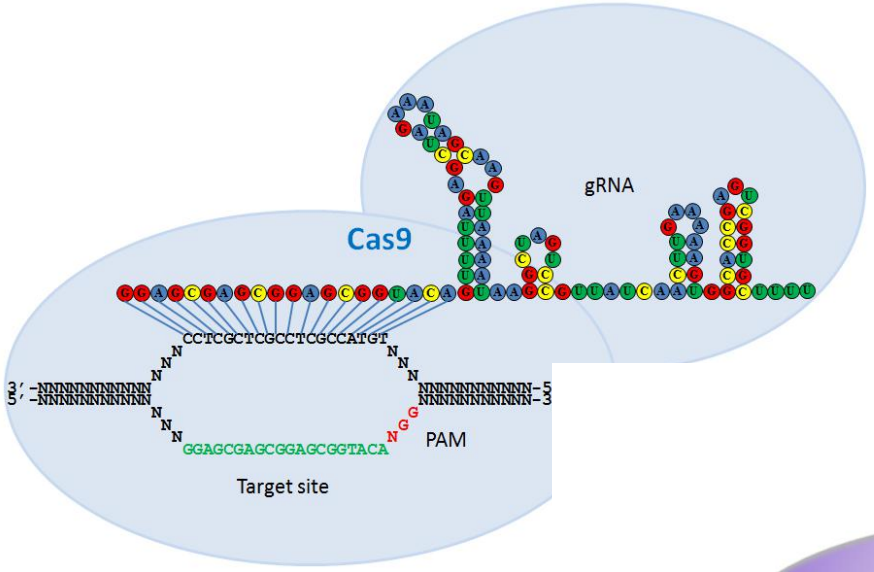
- Off target effects
- Can not completely deplete the expression
- Can not generate knock-in cells/animals
- RNAi is not the best way to knock down non-coding RNA.

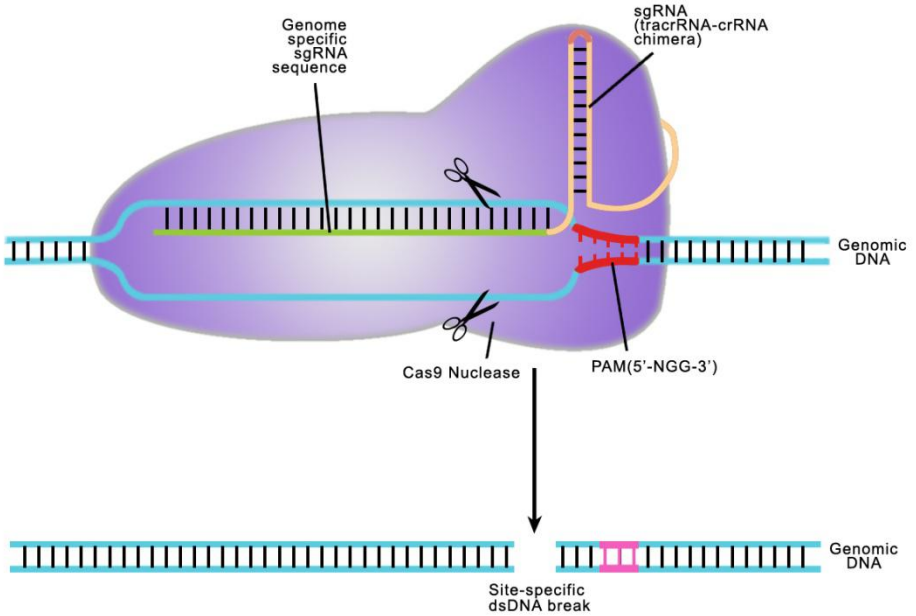
Genome editing method



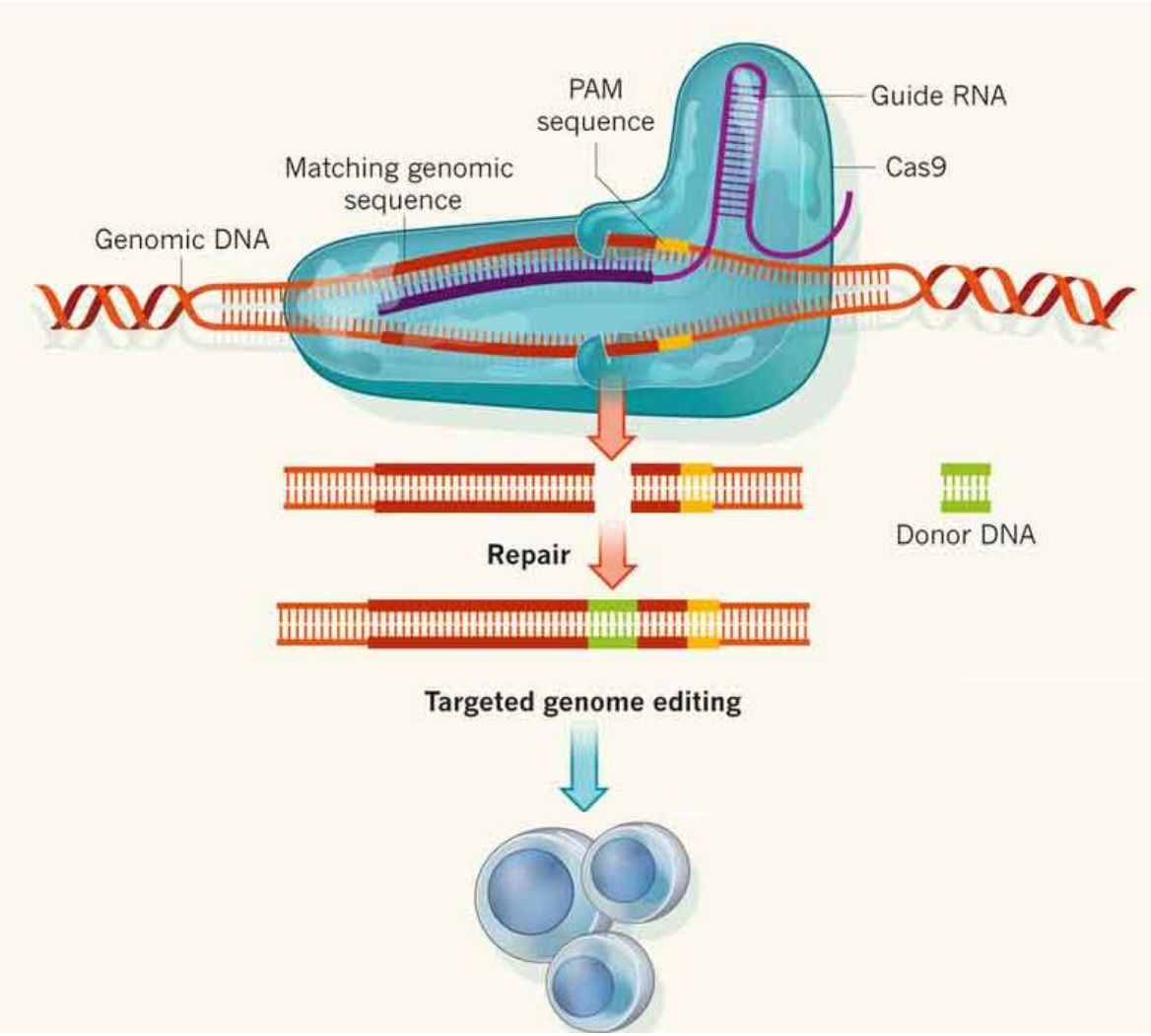
The Nobel prize winning method for 2020

CRISPR/Cas9


gRNA
Cas9
CCTCGCTCGCCTCGCCATGT
3′-NNNNNNNNNNN
5′-NNNNNNNNNNN
NNNNNNNNNNN-5
NNNNNNNNNNN-3
GGAGCGAGCGGAGCGGTACA
PAM
Target site


Genome specific sgRNA sequence
sgRNA (tracrRNA-crRNA chimera)
Genomic DNA
Cas9 Nuclease
PAM(5'-NGG-3')
Genomic DNA
Site-specific dsDNA break

CRSPR/Cas9 mediated gene editing

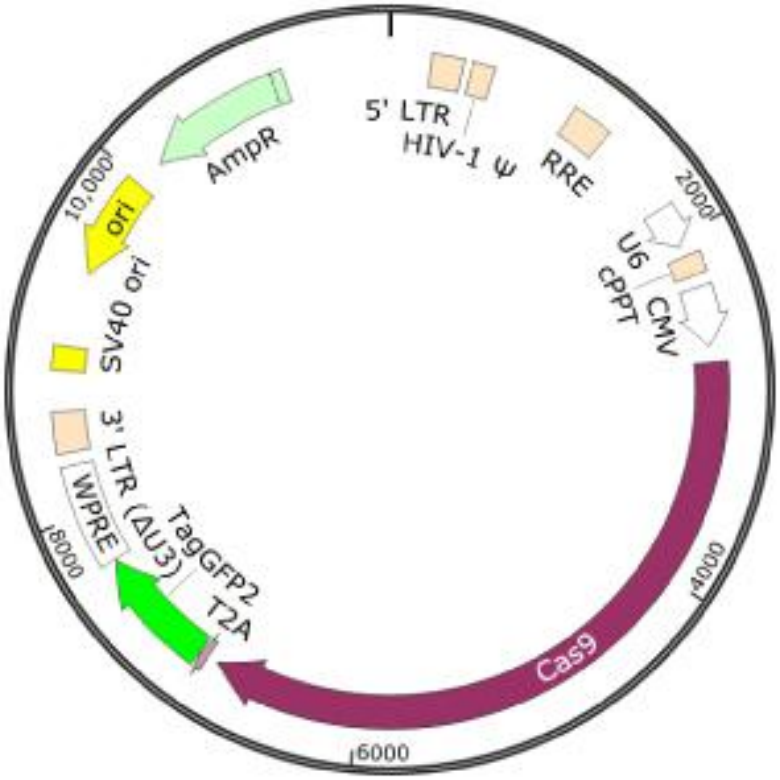


sgRNA design

- https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design
- Addgene's confirmed sgRNA dataset

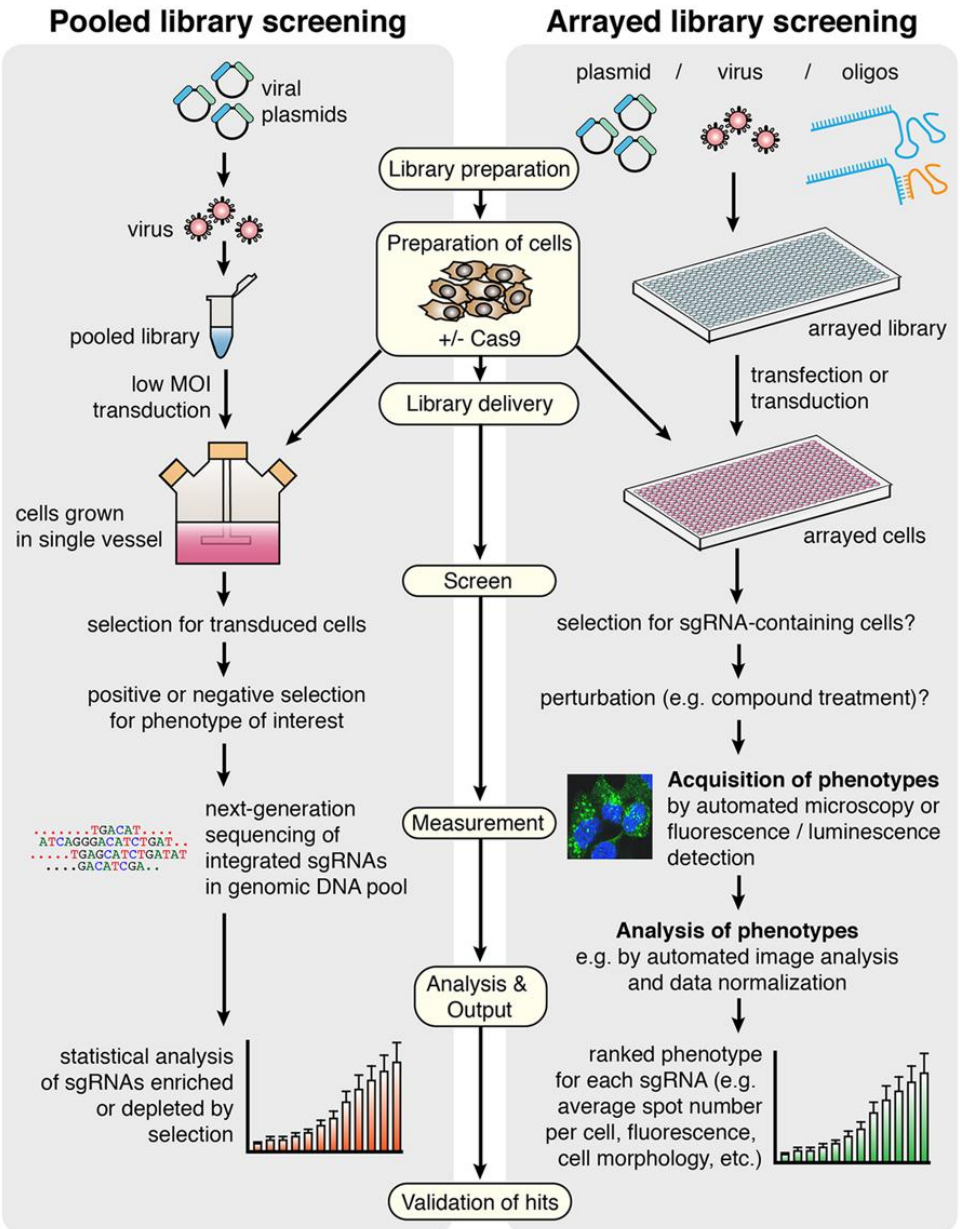
- **CRSPR/Cas9 can be used in all types of cells.**
- **It is better than RNAi.**

Viral vector to express sgRNA and Cas9



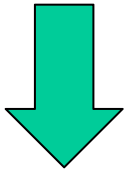
pRSGCCG-U6-(sg)-CMV-Cas9-2A-TagGFP2
11,616 bp

Genome wide sgRNA library screening



Results:

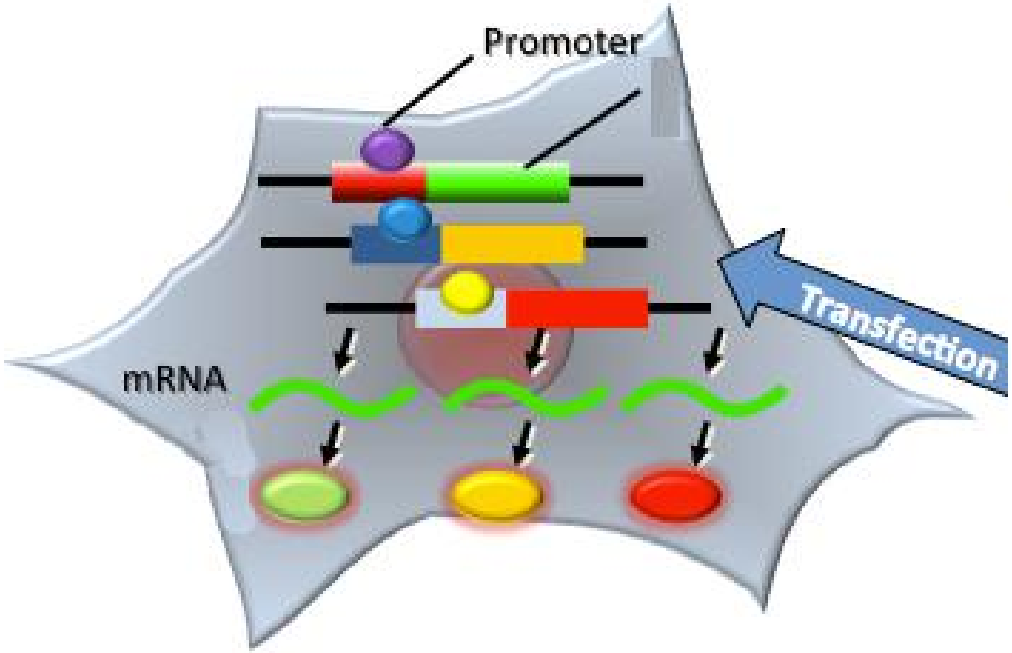
- **Reducing M expression by RNAi leads to muscle dystrophy.**
- **Knockout M expression by CRSPR/Cas9 leads to muscle dystrophy.**

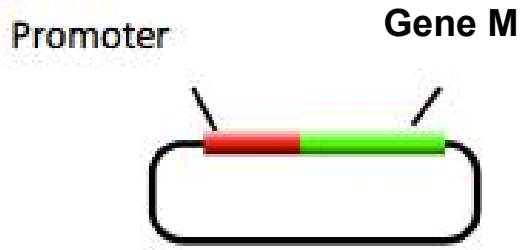


Conclusions:

The decrease of M expression causes muscle dystrophy.

Over-expression of M in muscle cells





CMV promoter etc
In frame
Proper stop
tag : Flag, His, HA etc.

Muscle hypertrophy?

Yes, gene M is the cause.

No, gene M is not the cause.

Results:

- **Reducing M expression by RNAi leads to muscle dystrophy.**
- **Knockout M expression by CRSPR/Cas9 leads to muscle dystrophy.**
- **Supplement of M rescues muscle dystrophy.**

Conclusions:

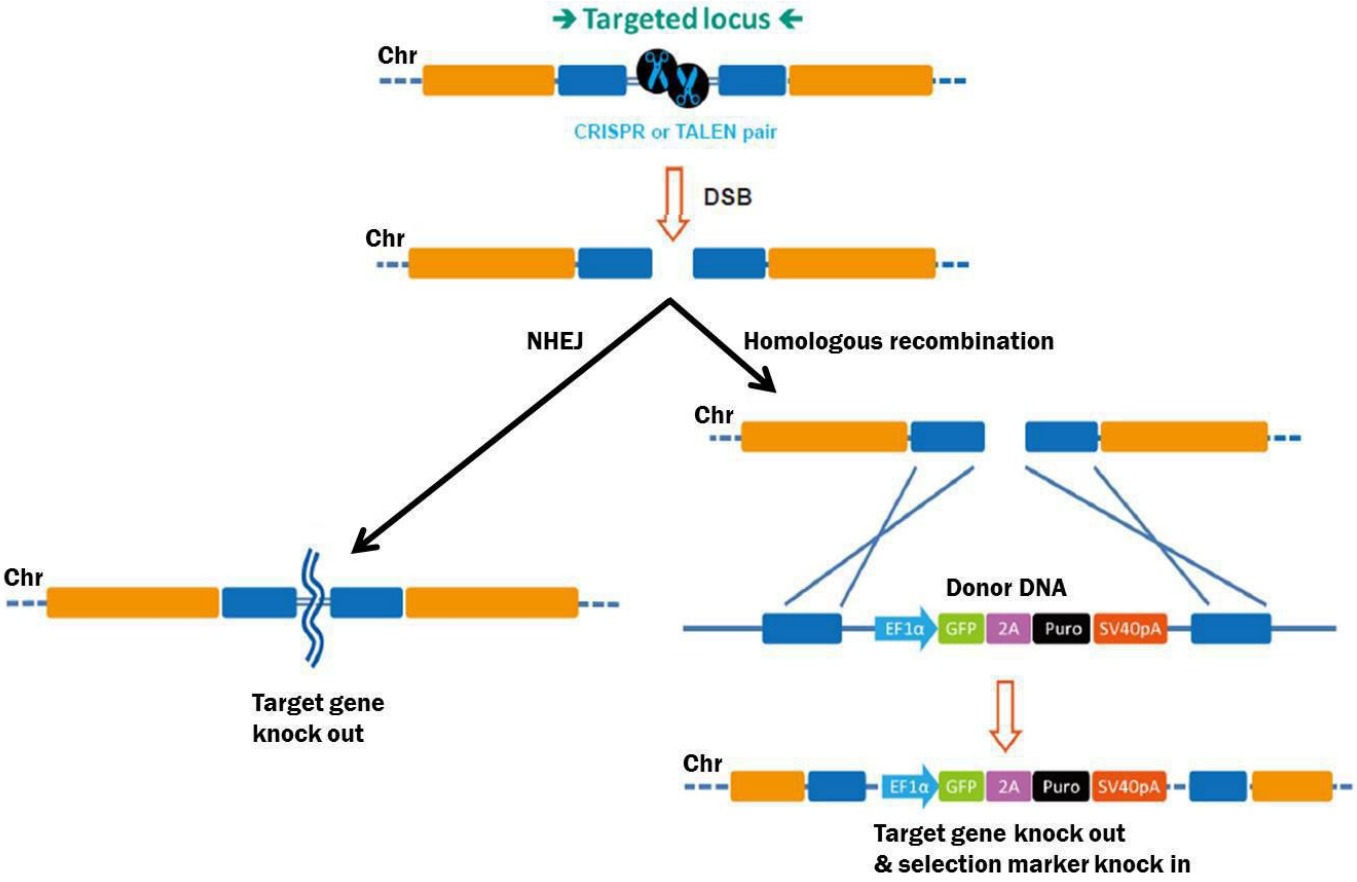
The decrease of M expression causes muscle dystrophy.

Homework thinking :

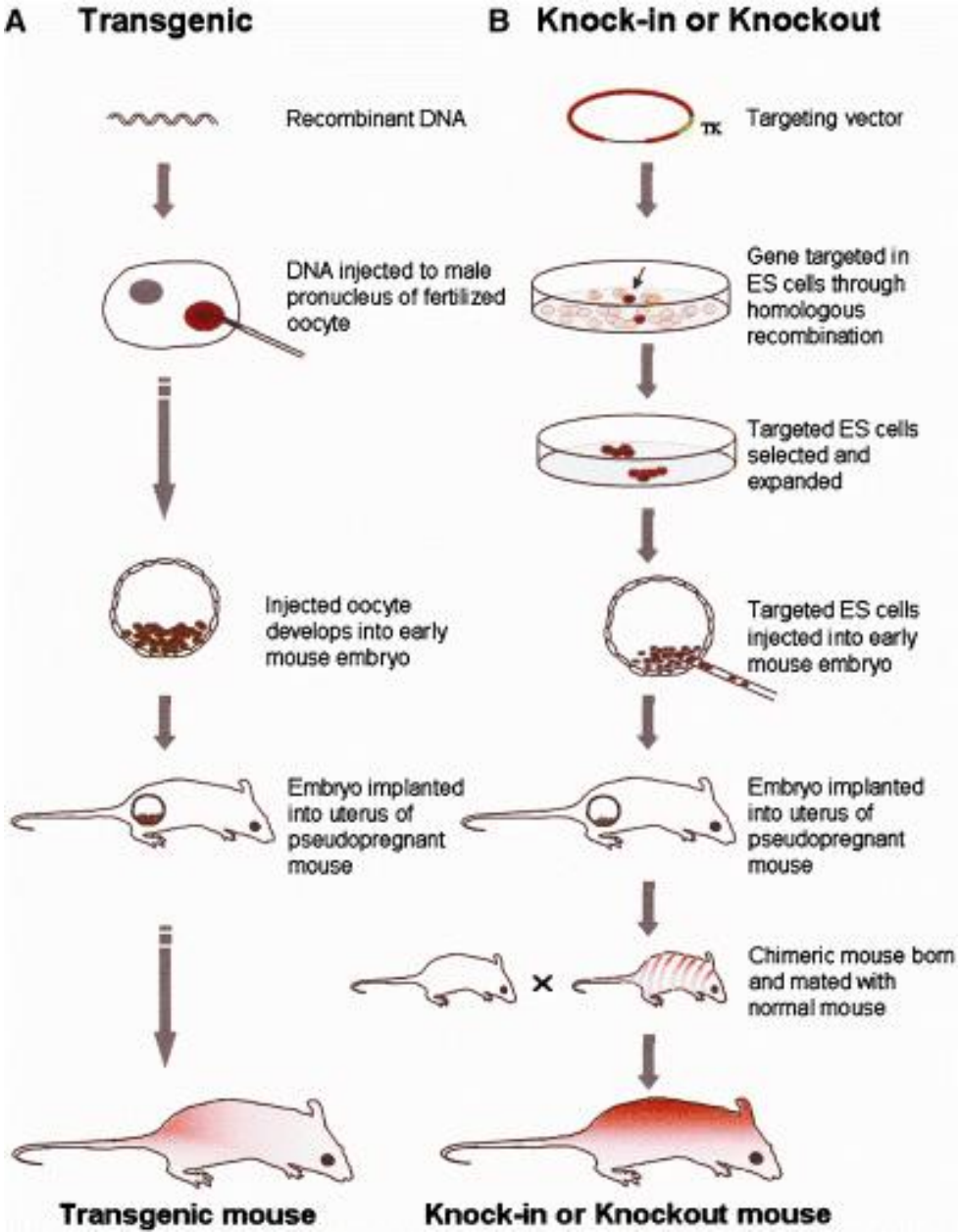
How to test whether the overexpression of a gene works?

The In vivo assays

Generation of M knock-out animals



Generation of knock-out mice



After genotyping, always do Western blot to check the protein expression level of your target gene.

M knockout animals

Whether recapitulating muscle atrophy?

If yes

M is the cause of muscular dystrophy

If no

M may not be the cause of muscular dystrophy

Results:

- **Knocking out of M leads to muscle dystrophy in model animals**

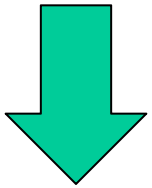
Conclusions:

The decrease of M expression causes muscle dystrophy.

Conclusion 2:

Frame shift mutation of gene M leads to decrease of gene M expression, which in turn causes muscular dystrophy.

Next Question

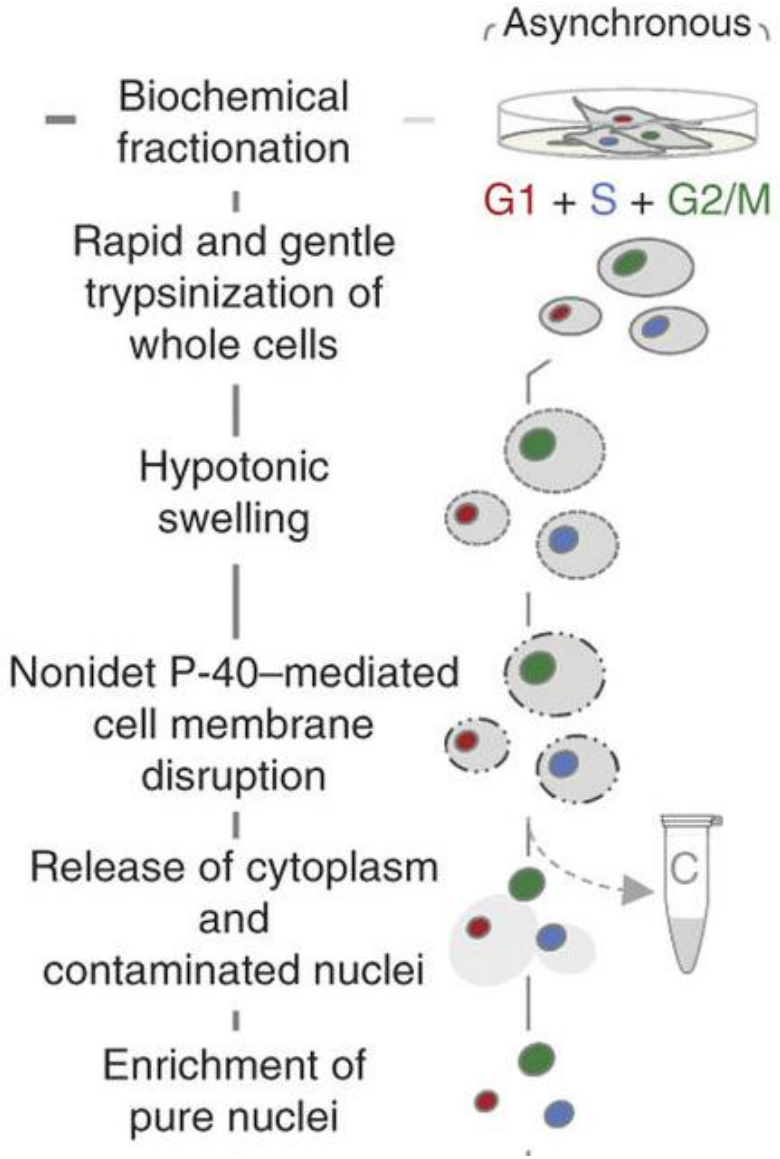


How can M cause muscle dystrophy?

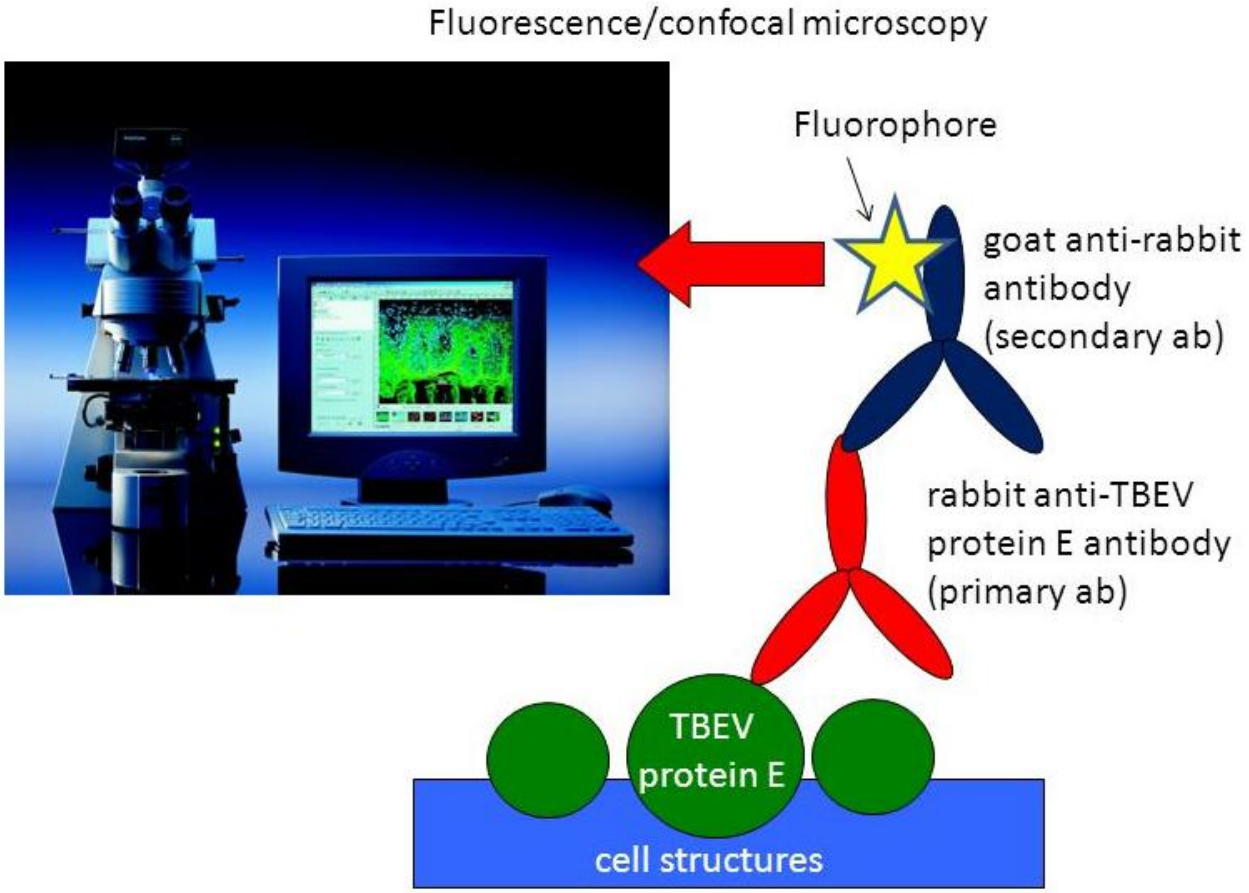
Identify the location of gene M

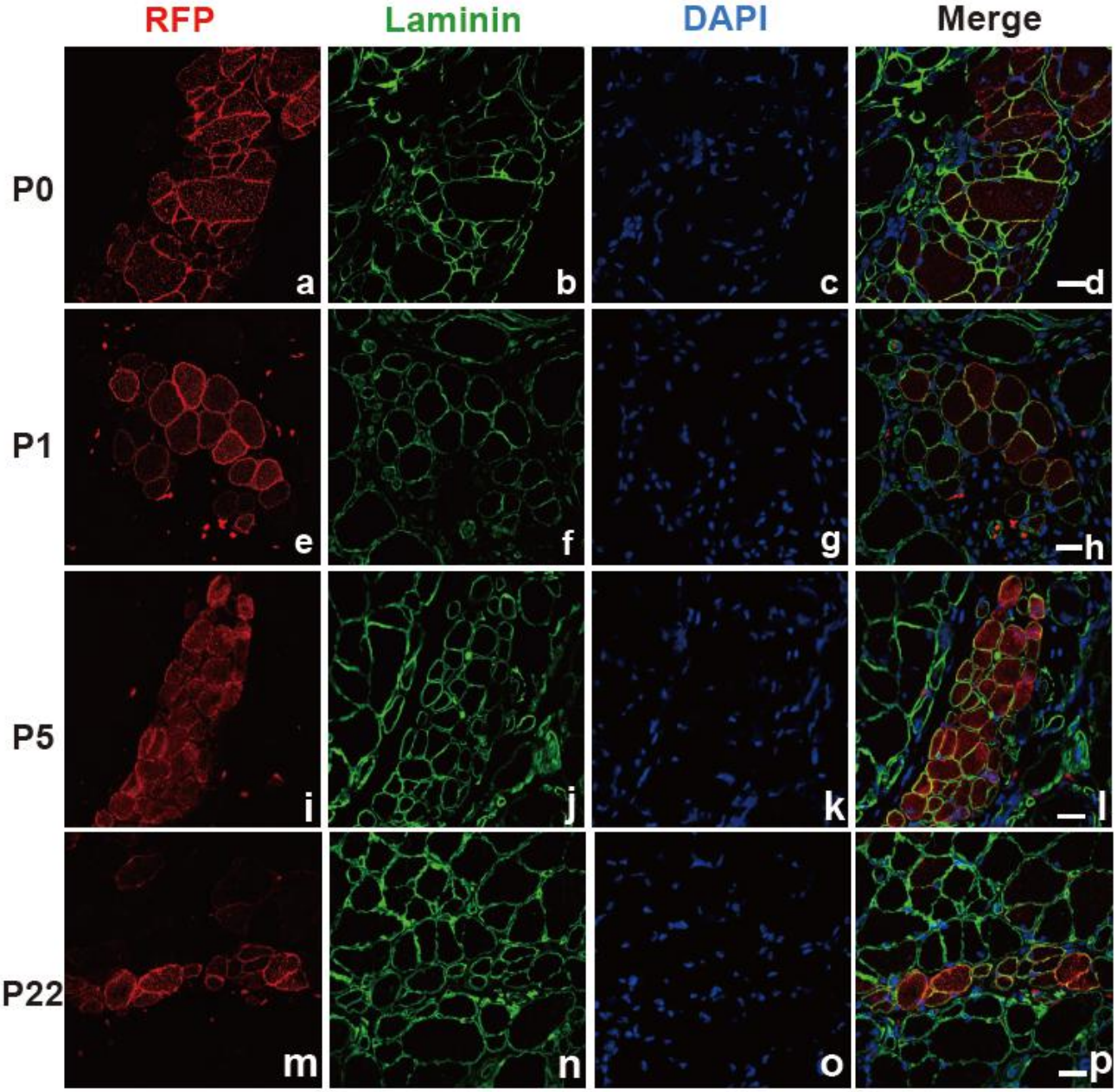
- **Isolate nuclear and cytoplasmic fraction and perform Western blot;**
- **Immuofluoresent staining;**
- **Tagged protein.**

Isolation of nuclei

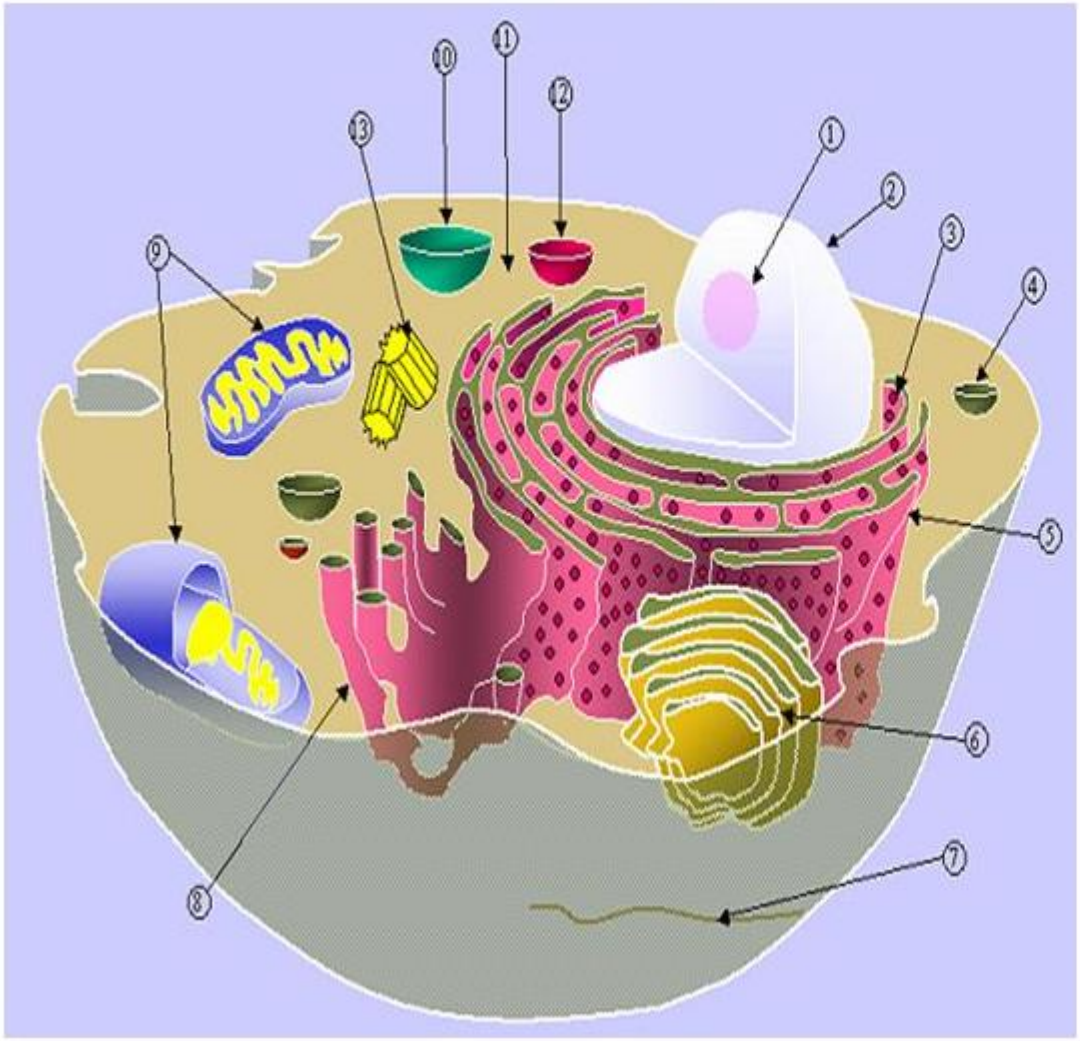


Immunofluorescence staining



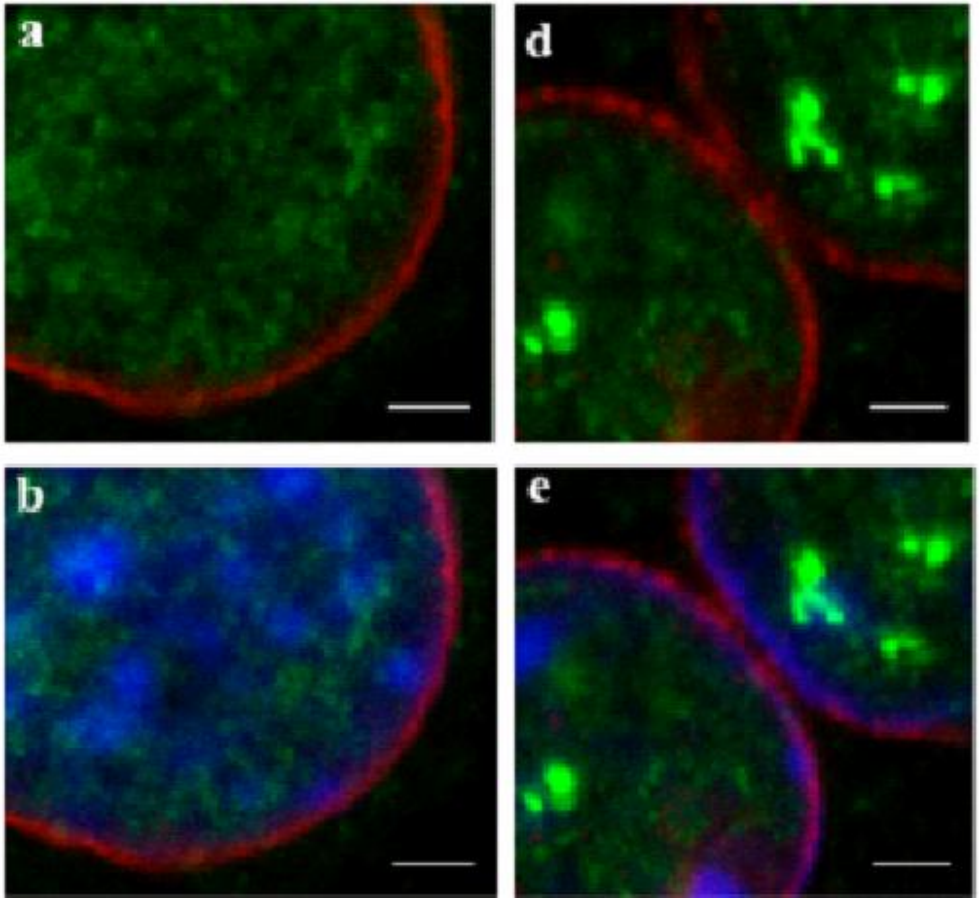

RFP
Laminin
DAPI
Merge
P0
P1
P5
P22
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p

Eukaryote Cell Structure



1. Nucleolus
2. Nucleus
3. Ribosome
4. Vesicle
5. Rough ER
6. Golgi apparatus
7. Cytoskeleton
8. Smooth ER
9. Mitochondrion
10. Vacuole
11. Cytoplasm
12. Lysosome
13. Centriole

EM, PALM, STORM, and other super resolution microscopy can determine the subnuclear localization of proteins



- **Western blot and immunofluorescence staining all highly depends on good antibodies.**

- **If no good antibody could be obtained, M should be tagged.**

Tags frequently used

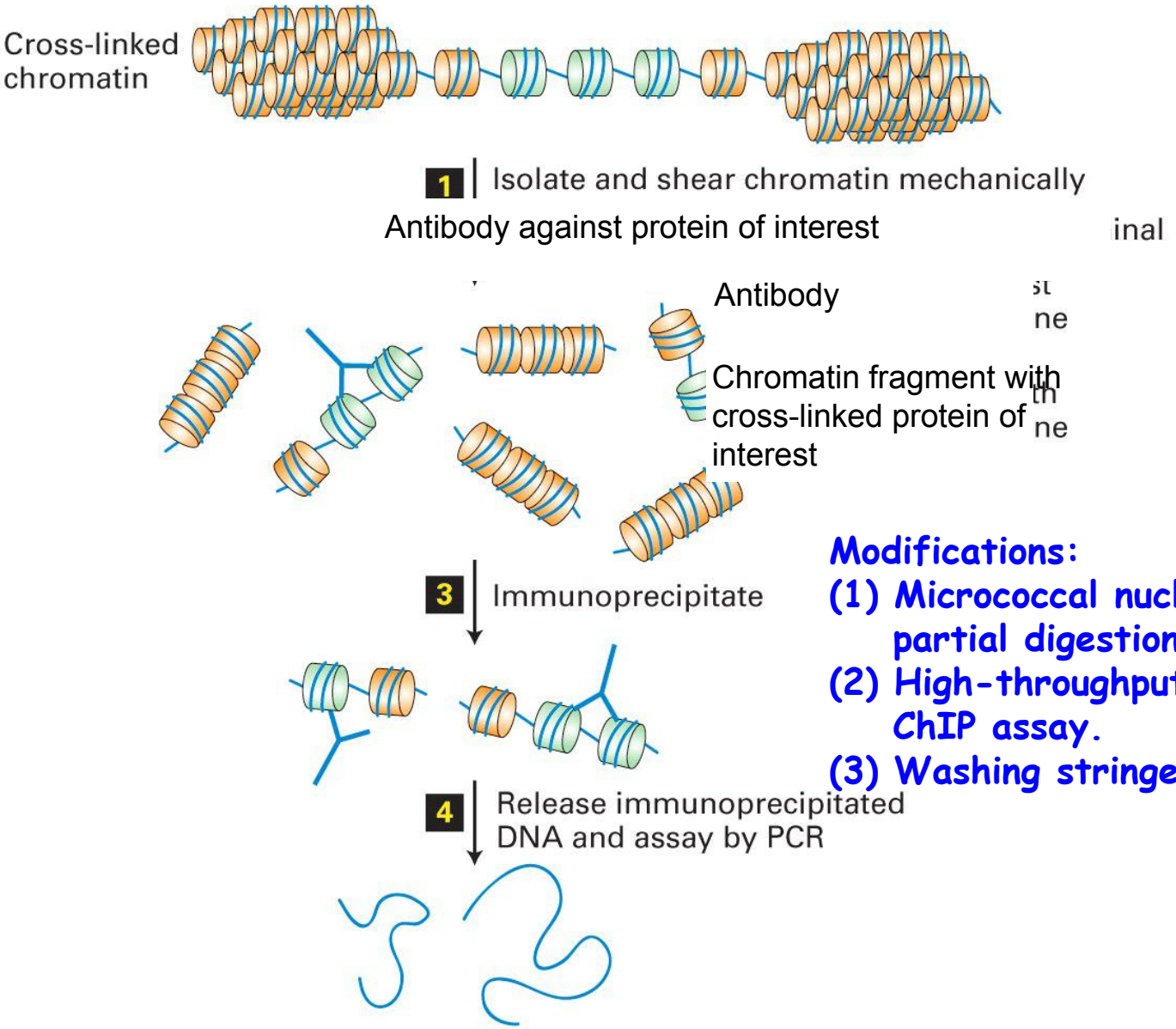
- **Flag, HA, His, GST, GFP, RFP, YFP, BFP, streptavidin, EoS, HALO etc.**
- **Tag can be added to the N- or C-terminus of a protein. Where to add the tag is a error-try process;**
- **Western blot and immunofluorescence staining can be performed with antibodies against the tag.**

M is located in nuclei.

M could be a potential transcription factor

How to prove?

Chromatin Immunoprecipitation (ChIP)



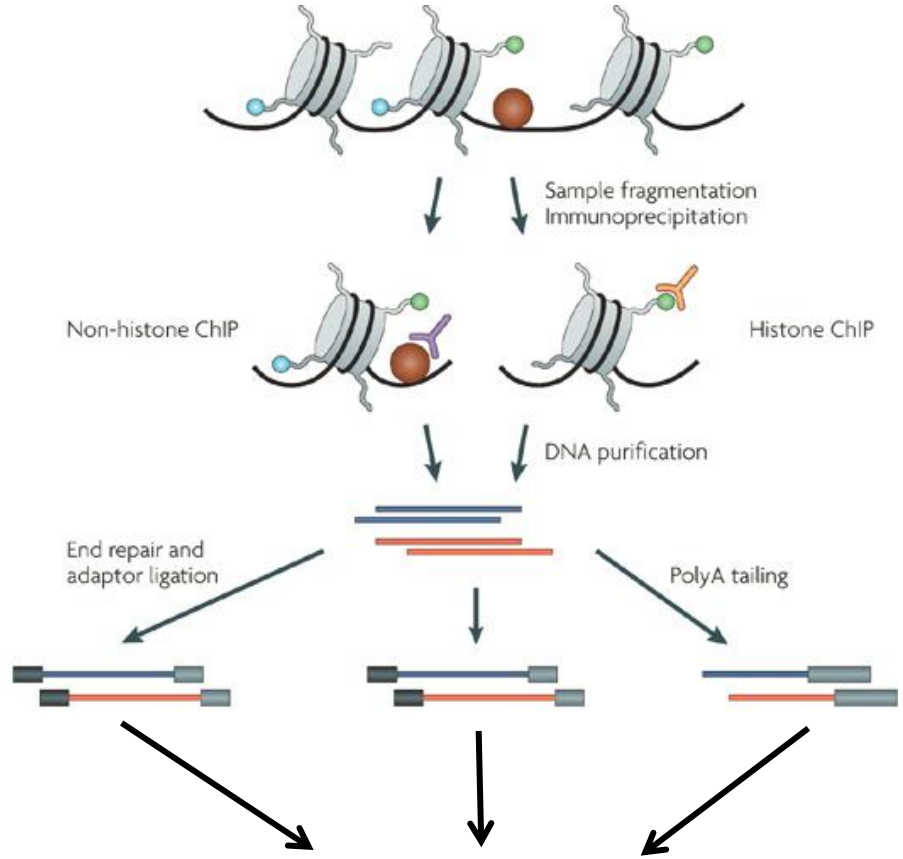
Modifications:

(1) Micrococcal nuclease partial digestion.

(2) High-throughput 96-well ChIP assay.

(3) Washing stringency

ChIP - sequencing



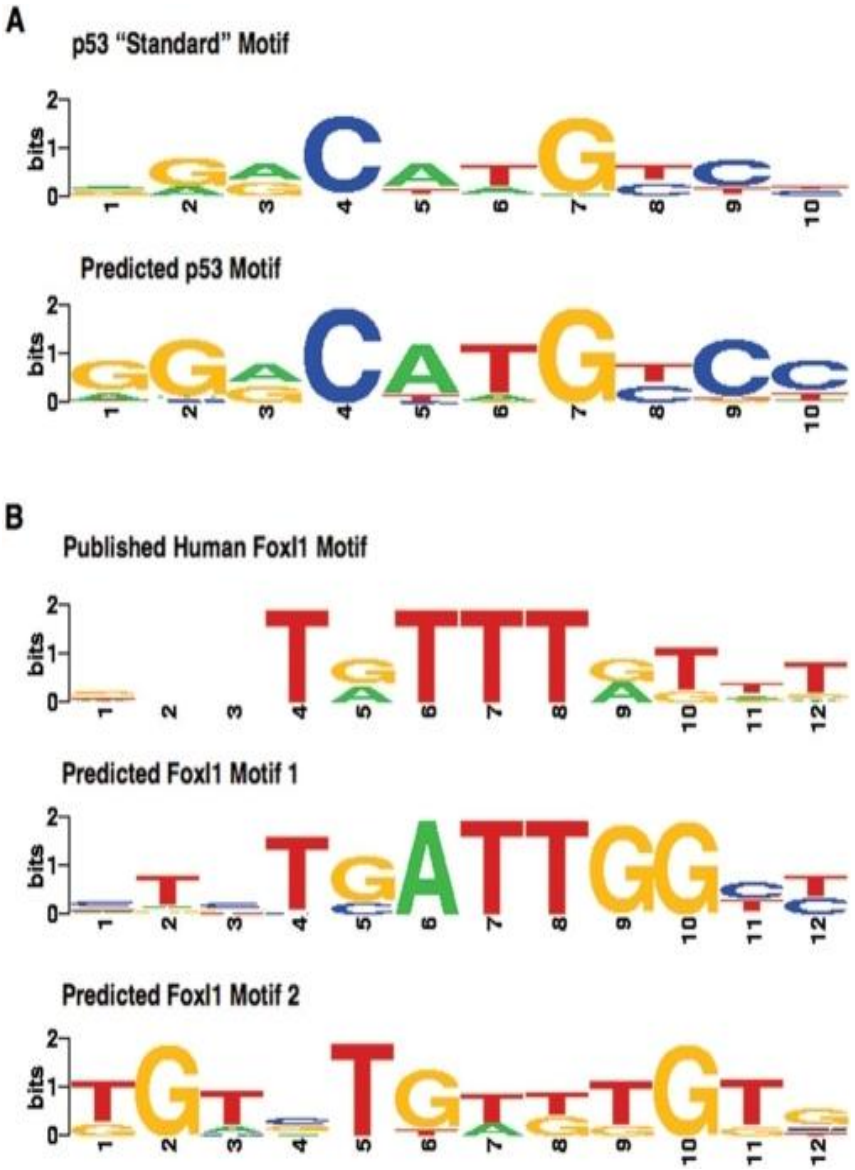
Deep sequencing

ChIP highly depends on good antibodies.

Question:

If no good antibody is available, what should you do?

Consensus sequence identification



Gel Retardation Assay
(eletrophoresis mobility assay (EMSA), gel shift assay)

To identify proteins which bind a specific DNA sequence.

To identify binding sequence of a DNA-binding protein.

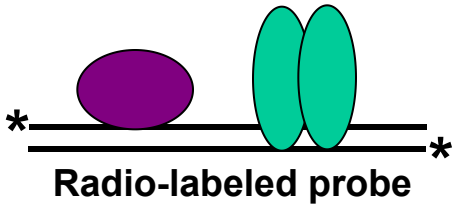
The assay exploits electrophoretic mobility differences between a rapidly migrating DNA and a more slowly migrating complex of protein bound to DNA.

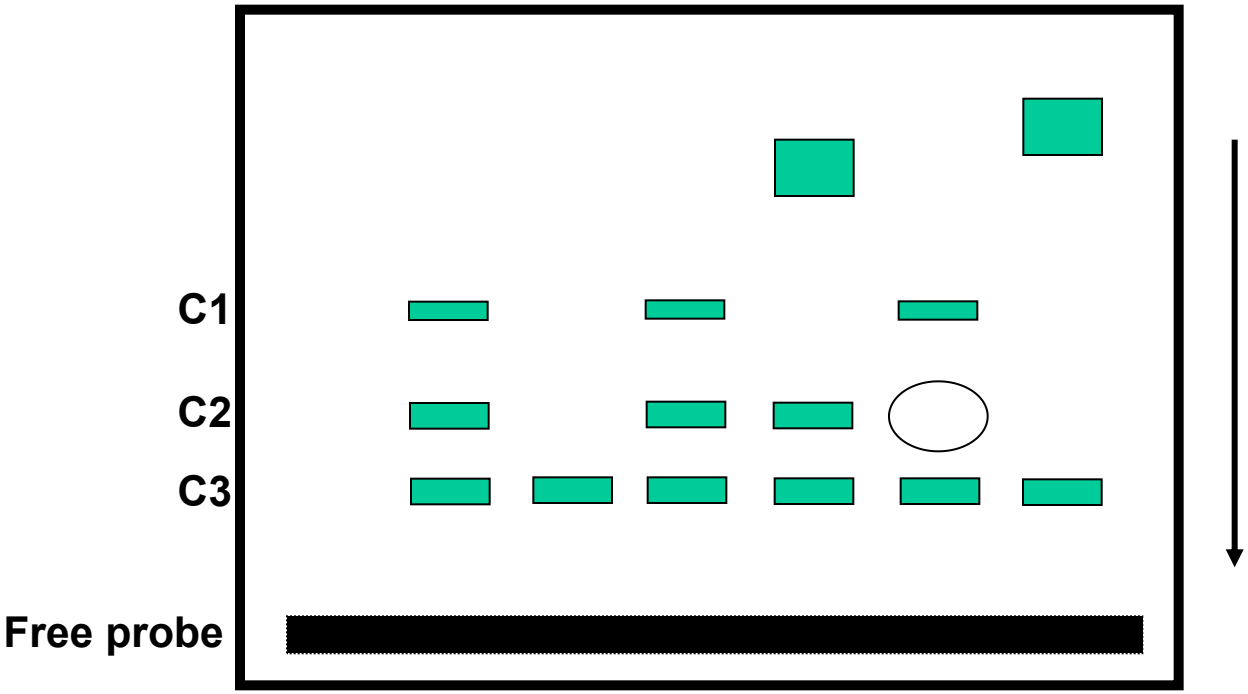
Limitations: it is an in vitro binding assay, not necessarily recapitulating what is happening in vivo.

Difference from SDS-PAGE: native gel vs. denaturing gel

Gel Shifting (EMSA) Example

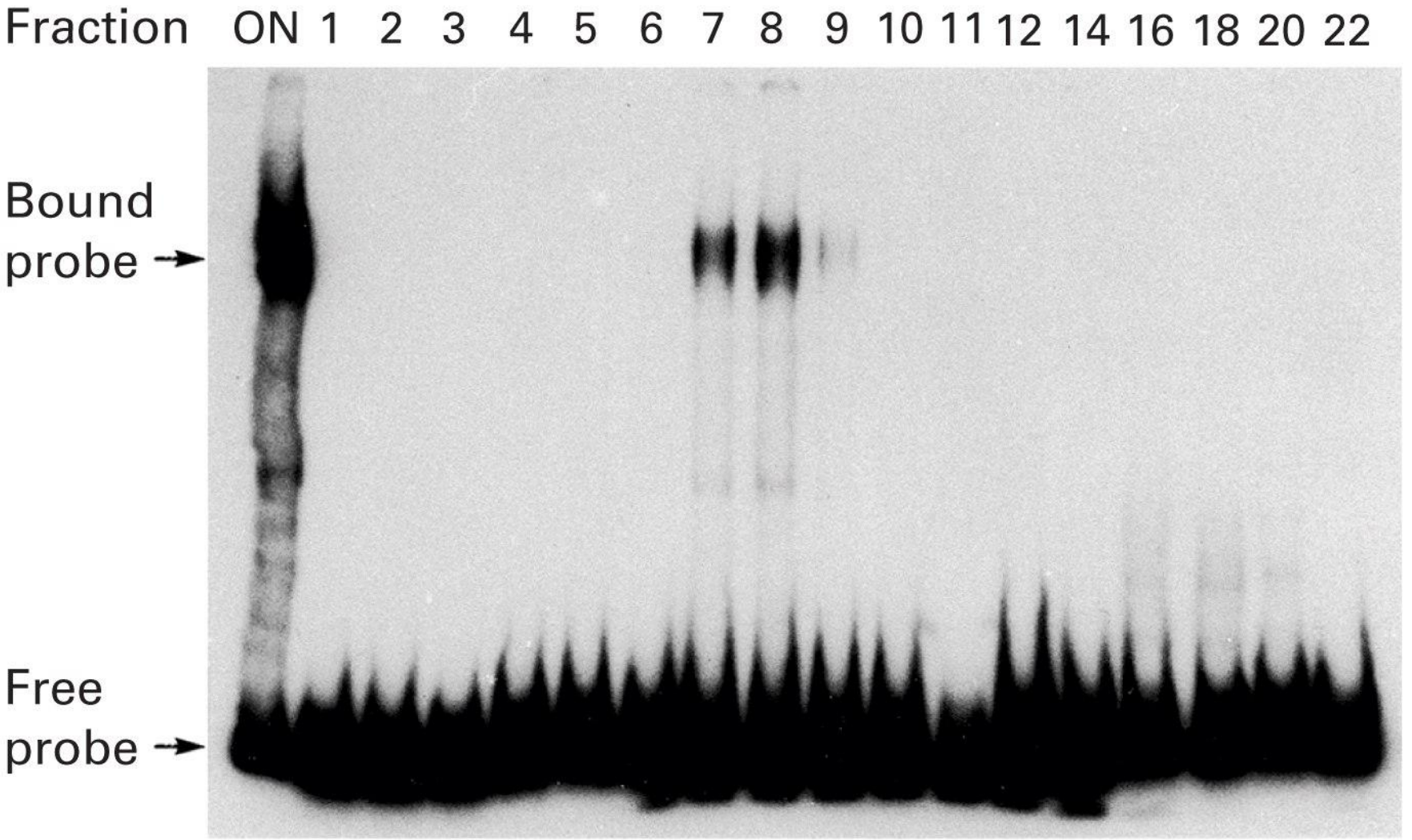
Anti-A antibody	−	−	−	−	+	−	+
Anti-B antibody	−	−	−	−	−	+	+
100X non-related DNA	−	−	−	+	−	−	−
100X cold probe	−	−	+	−	−	−	−
2 μg of nuclear extract	−	+	+	+	+	+	+
[^{32}P]-labeled DNA probe	+	+	+	+	+	+	+



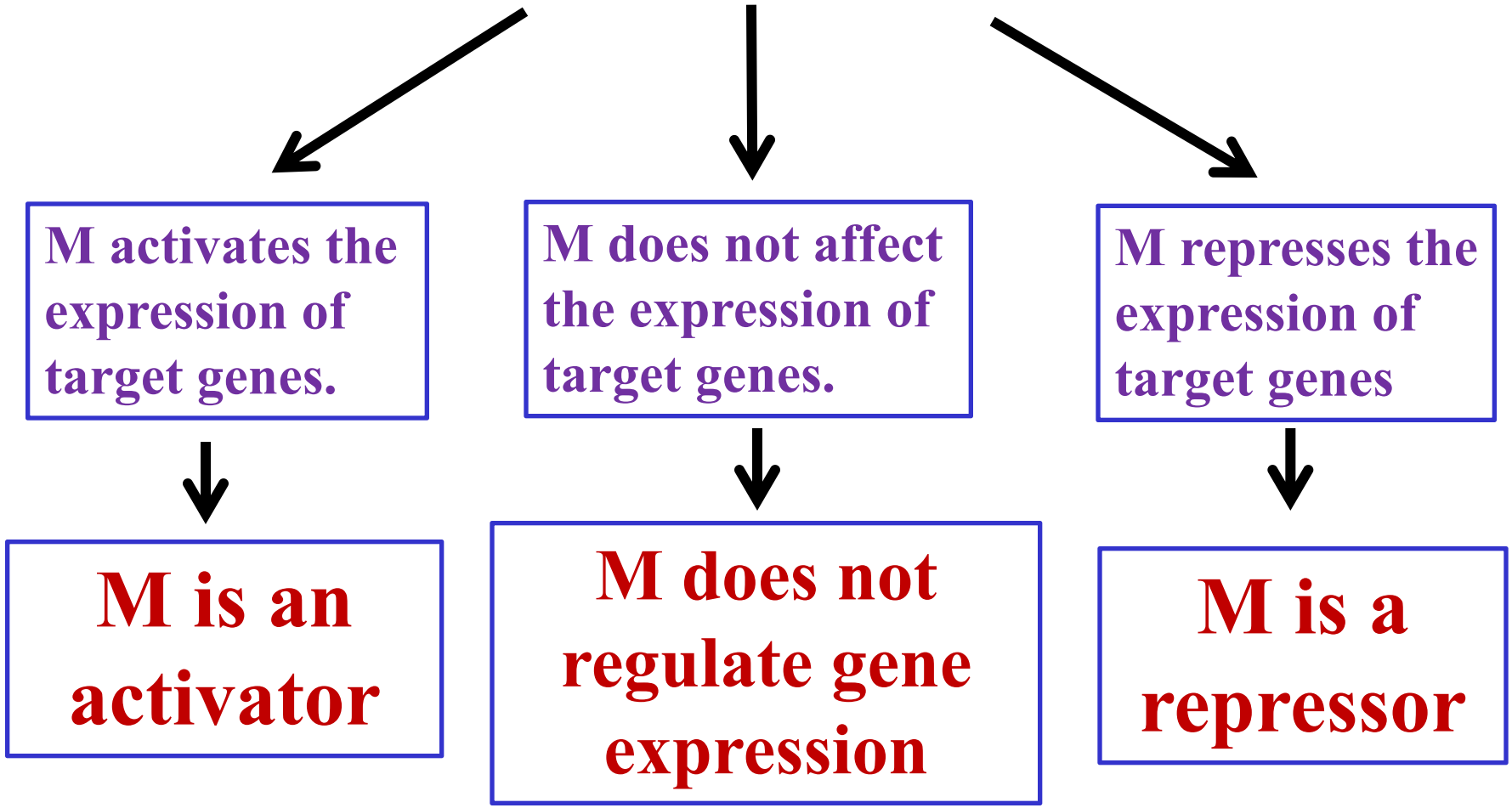


Question: Are you able to find out every protein components in a shifting complex?

An Example of Gel Shift Assay



ChIP-seq can also identify the target gene of M

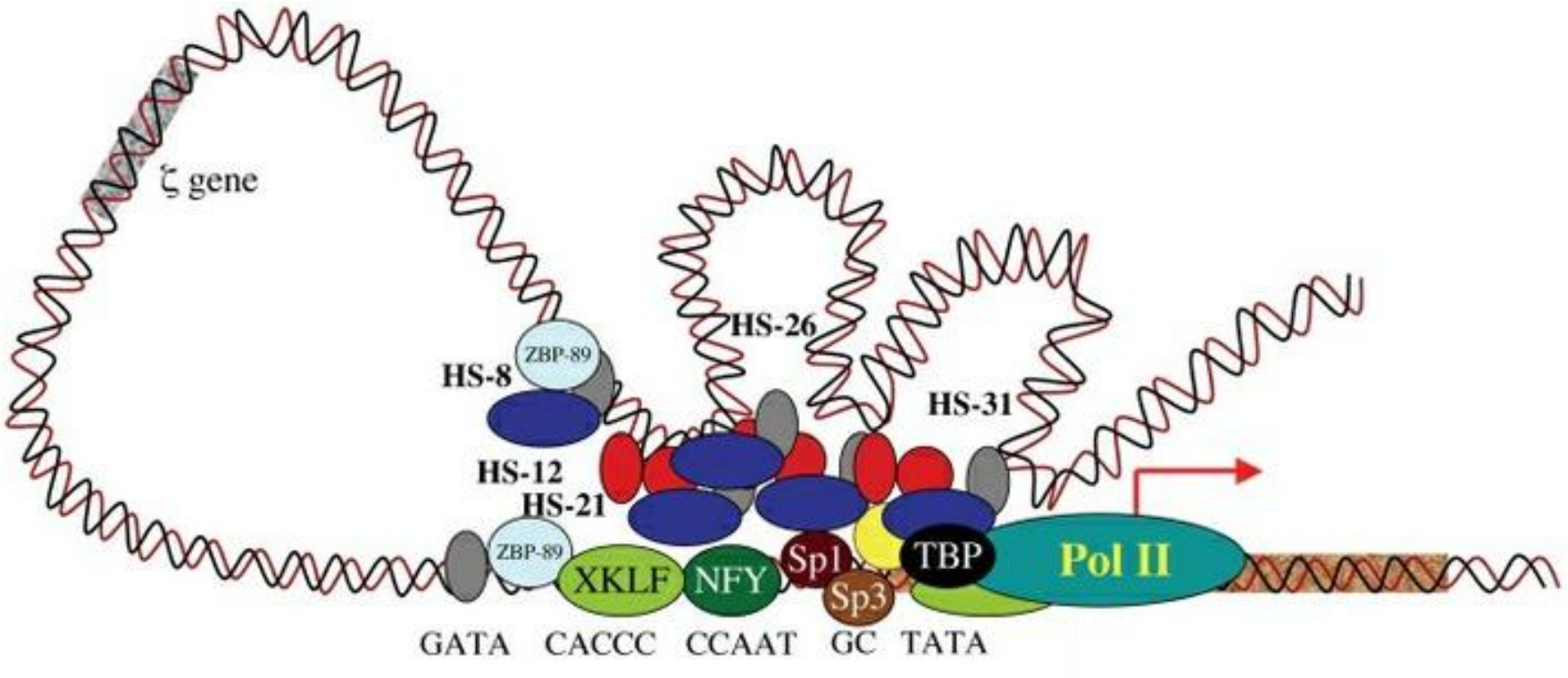


How do you test the above hypthesis?

The promoter and enhancer region of the target gene need to be identified first.

Promoter and transcription factors are critical elements to regulate gene expression

Model of RNA Polymerase II transcription



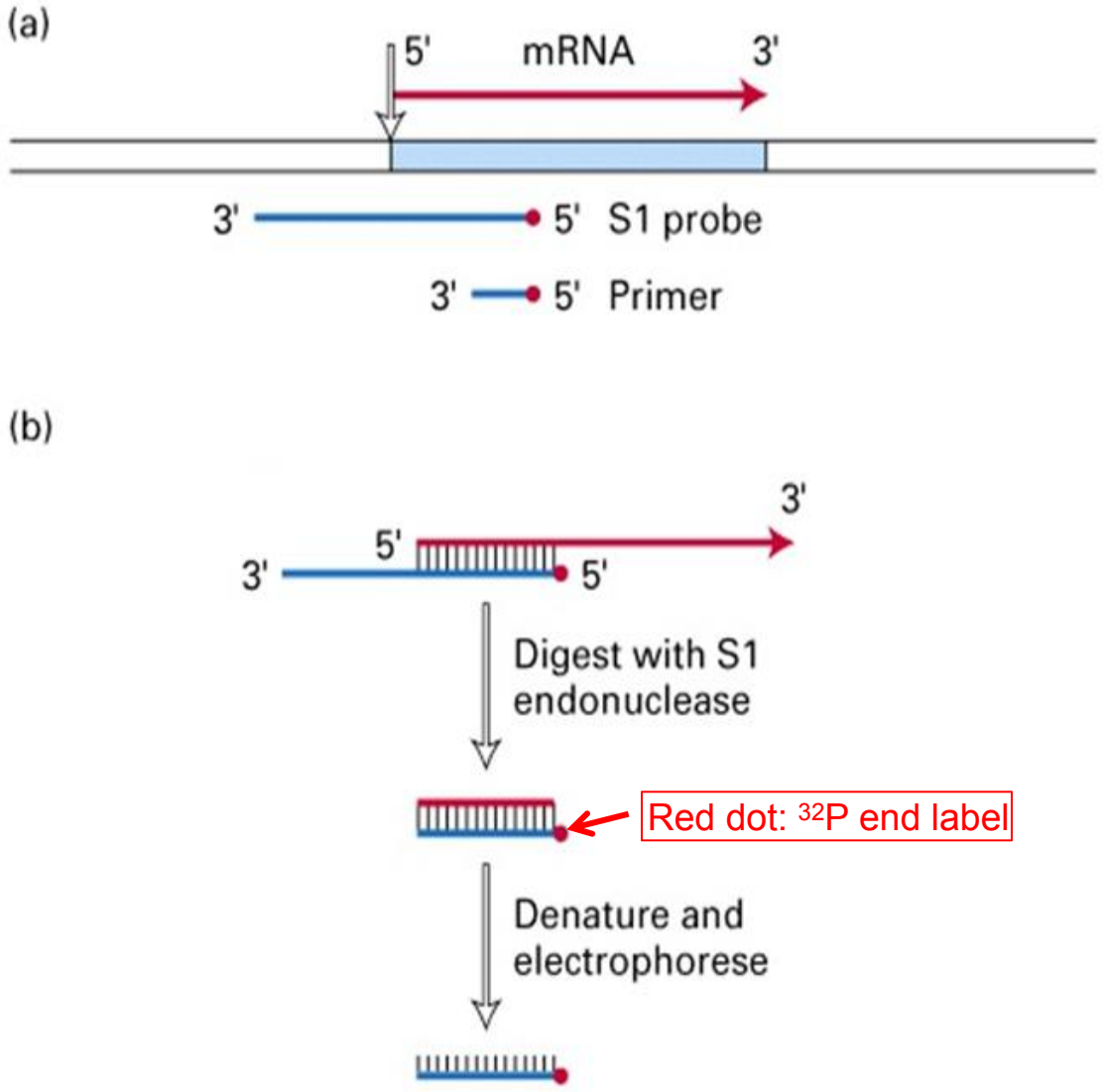
Find required promoter regions.

1) Identify transcription initiation sites.
2) Identify core promoter region.

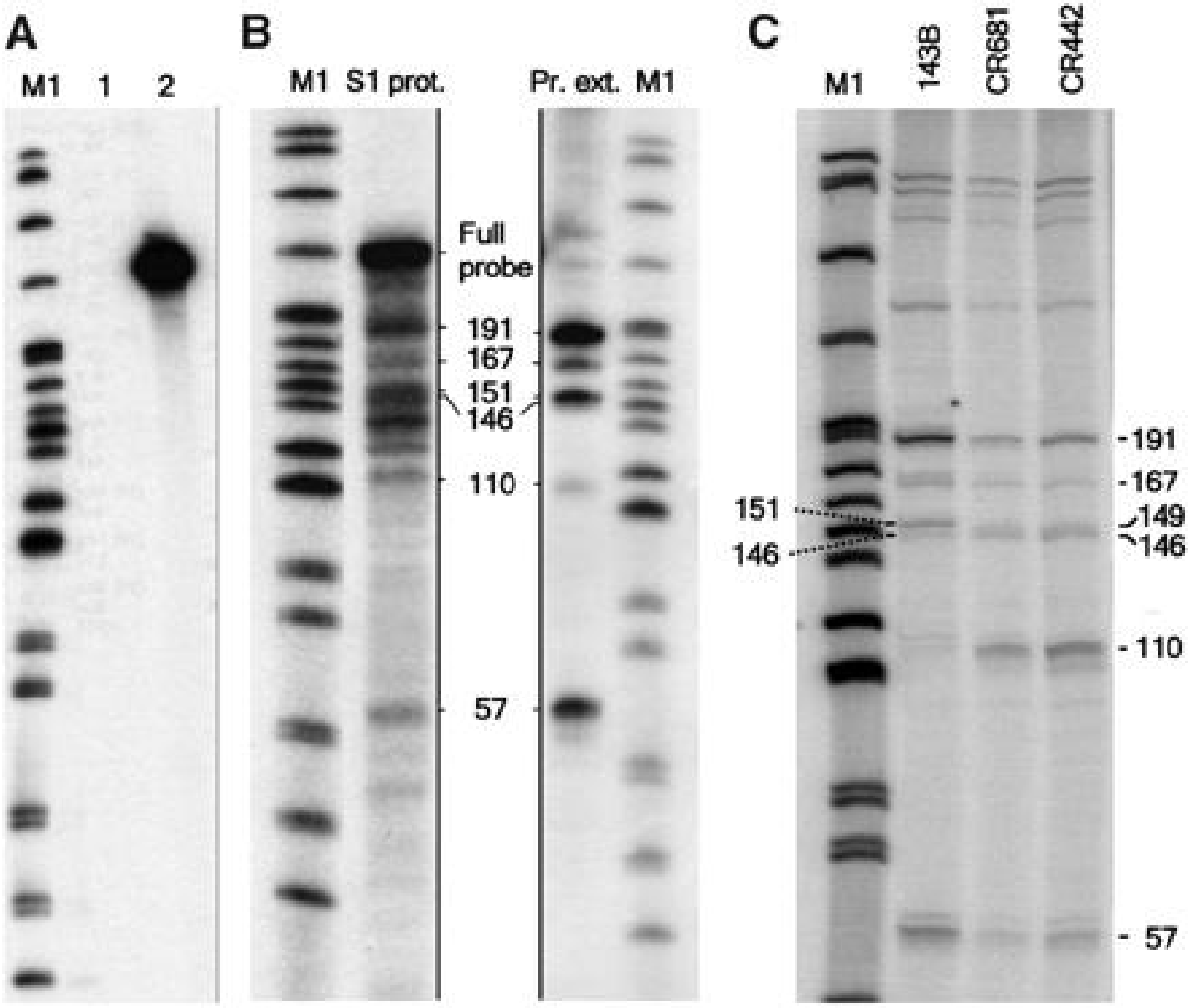
1. **Find required promoter regions.**

   1) Identify transcription initiation sites.

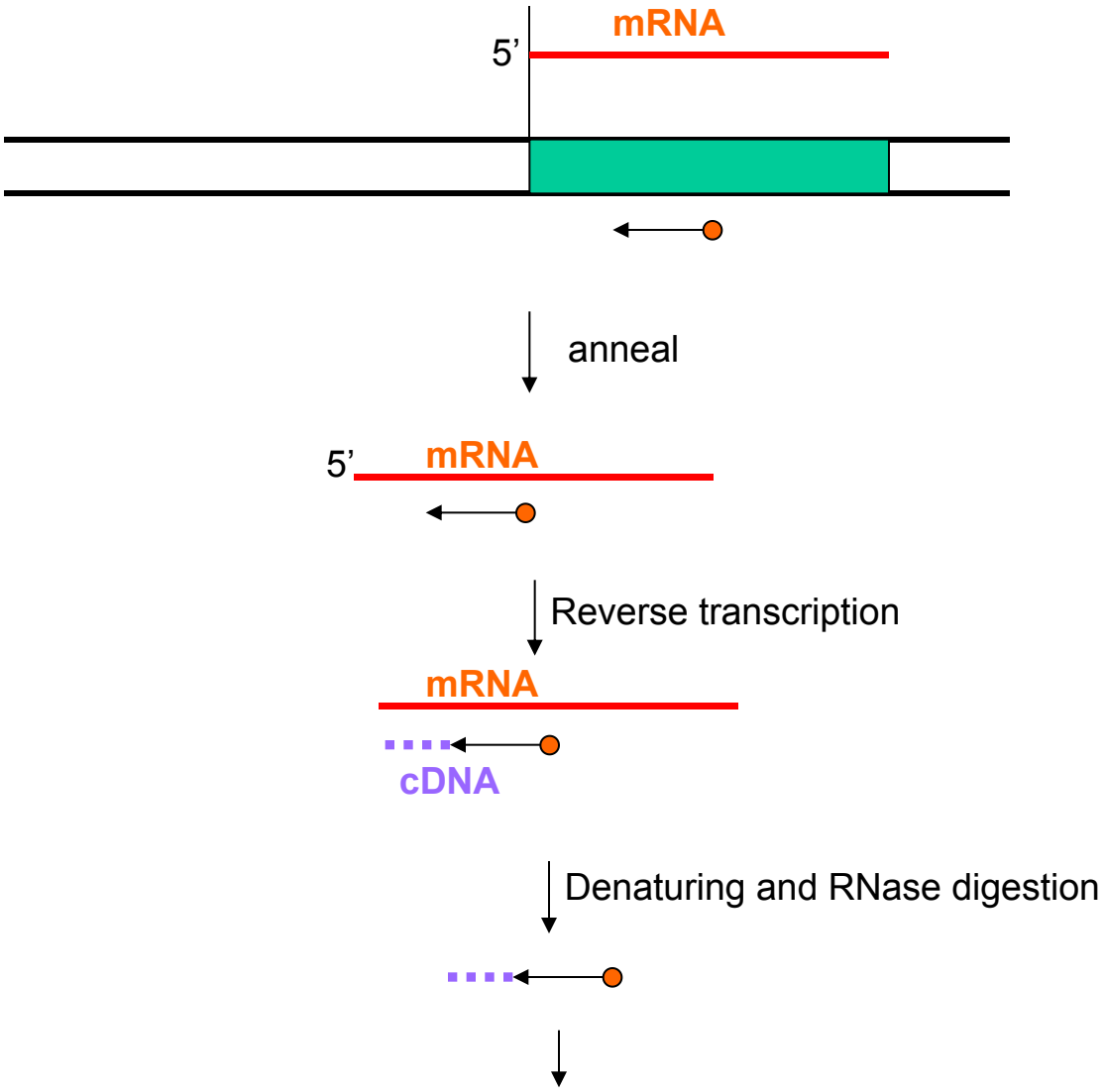
Transcription start sites can be mapped by *S1 protection* and *primer extension*

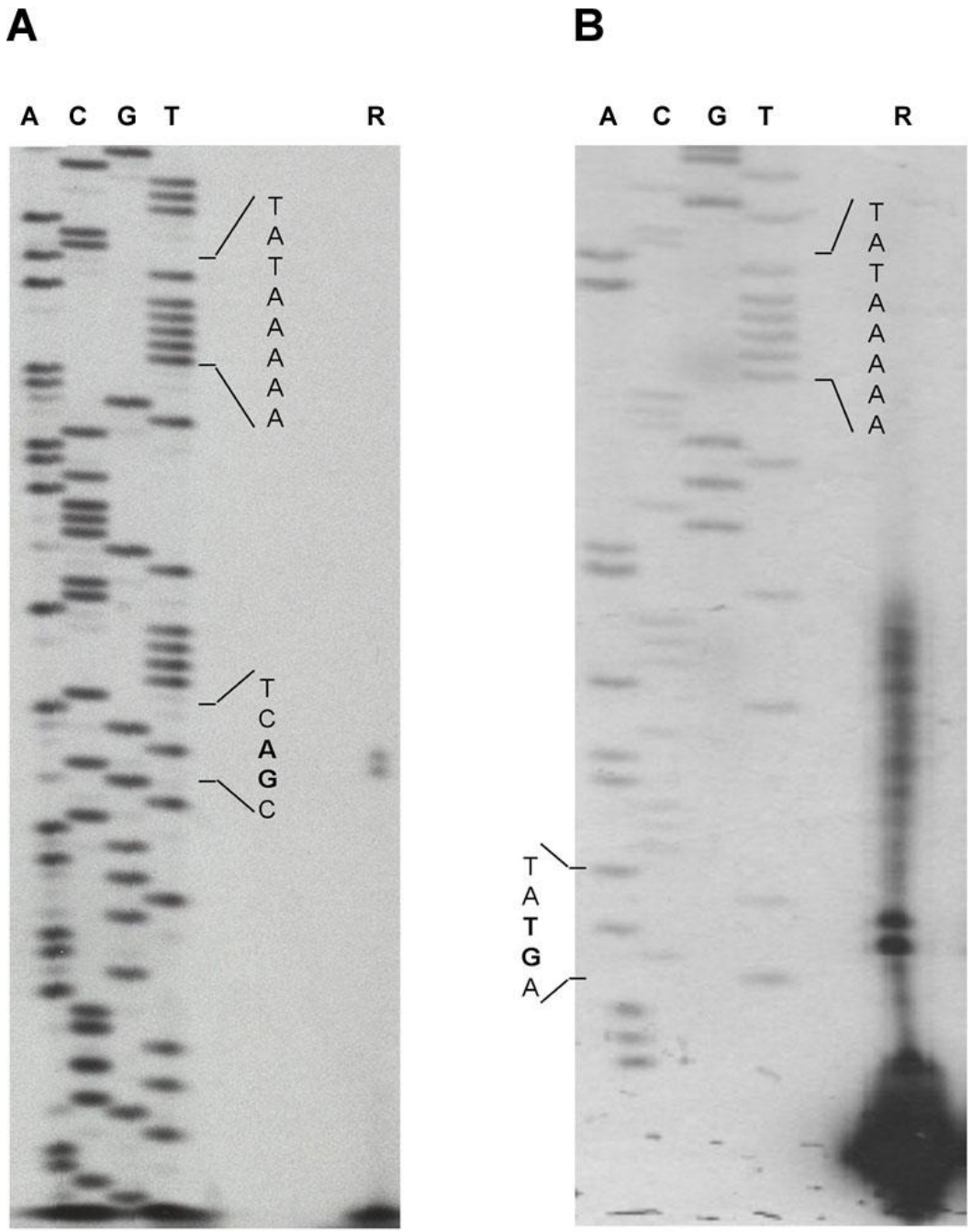


- **Mapping the start site for synthesis of a particular mRNA often helps identify the DNA regulatory sequences that control its transcription, because some of the regulatory elements are located near the start site.**
- **The position of the start site can be mapped from the length of the labeled probe segment protected from S1 digestion (S1 protection) or the resulting extension product (primer extension) on polyacrylamide-urea gel (same as DNA sequencing gel).**


A
M1
1
2
B
M1
S1 prot.
Pr. ext.
M1
Full probe
191
167
151
146
110
57
C
M1
143B
CR681
CR442
151
146
191
167
149
146
110
57

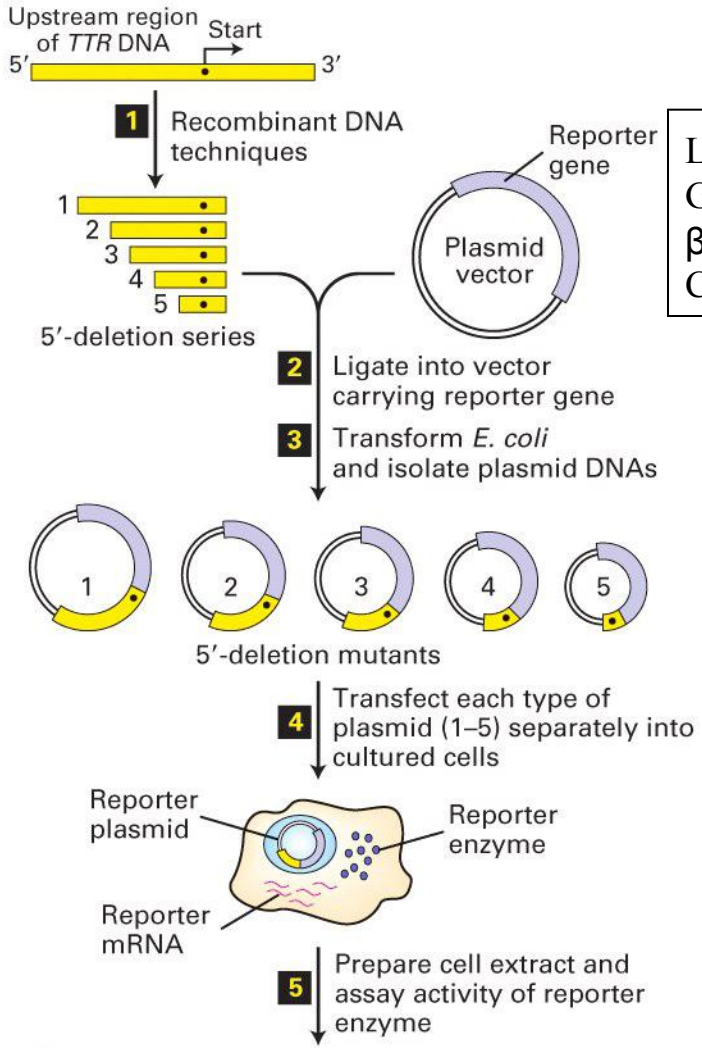
Primer extension




A
A C G T R
T A T A A A A A
T C A G C
B
A C G T R
T A T A A A A A
T A T G A

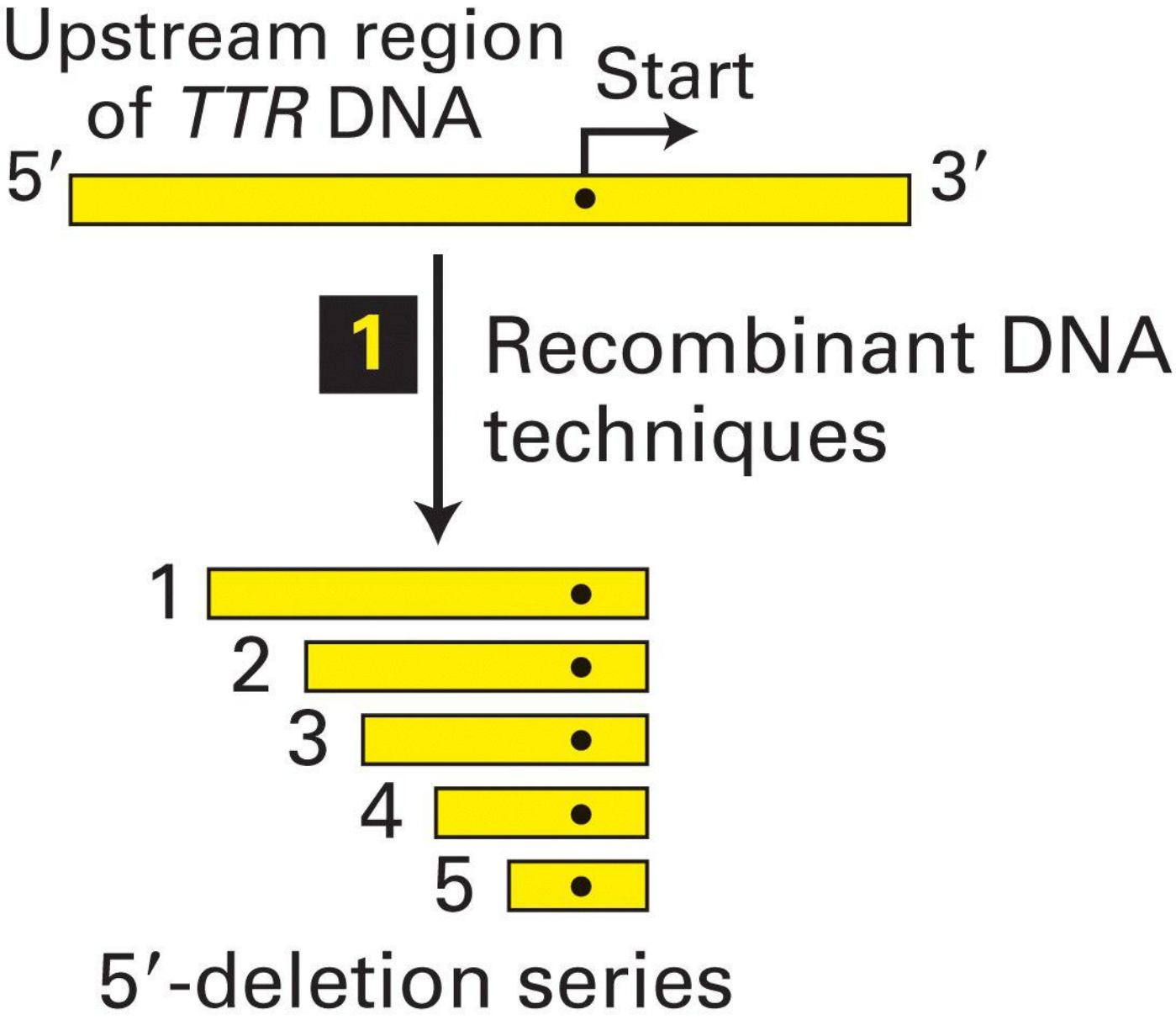
Find required promoter regions.

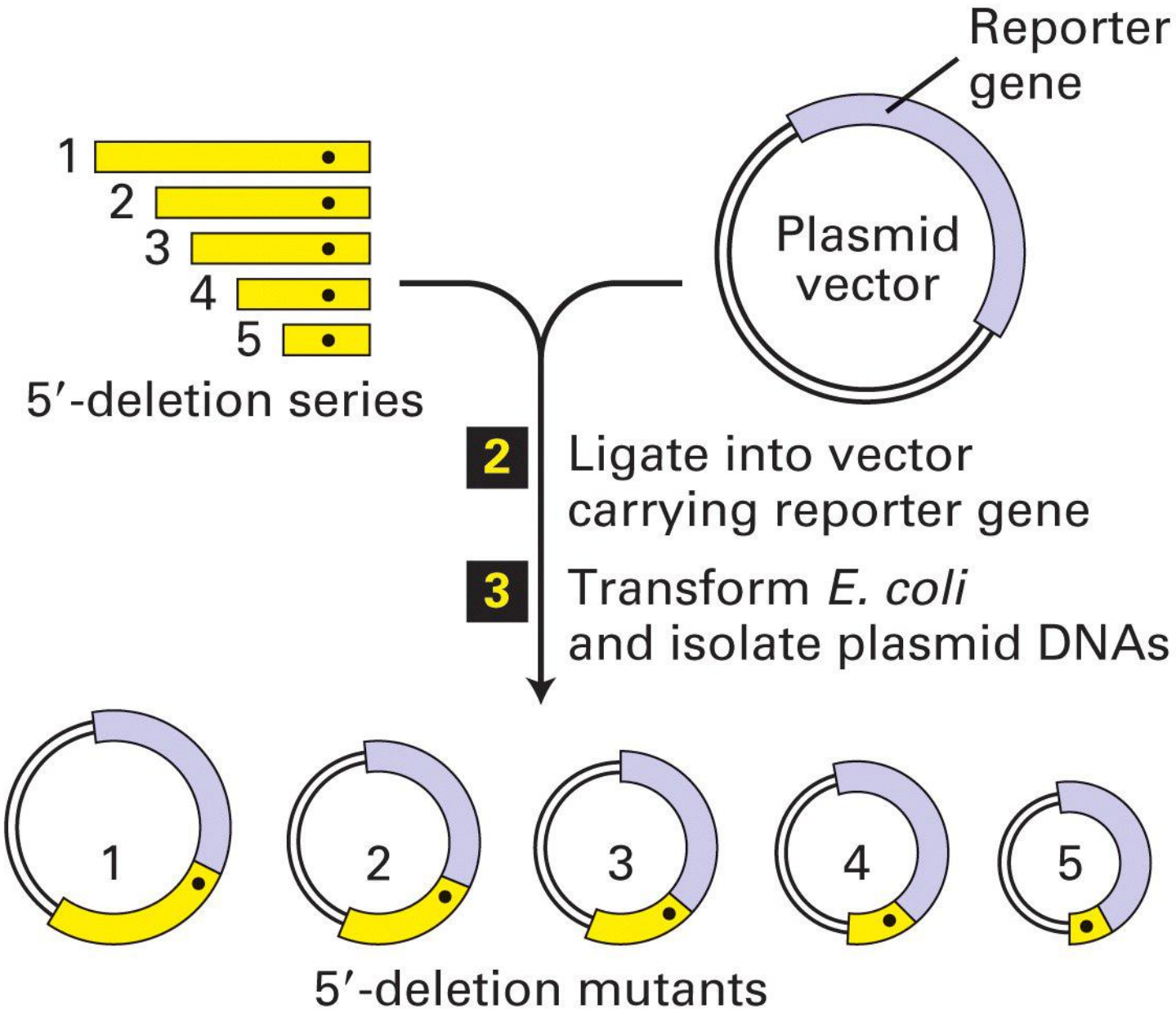
Promoter Analysis: reporter gene assay

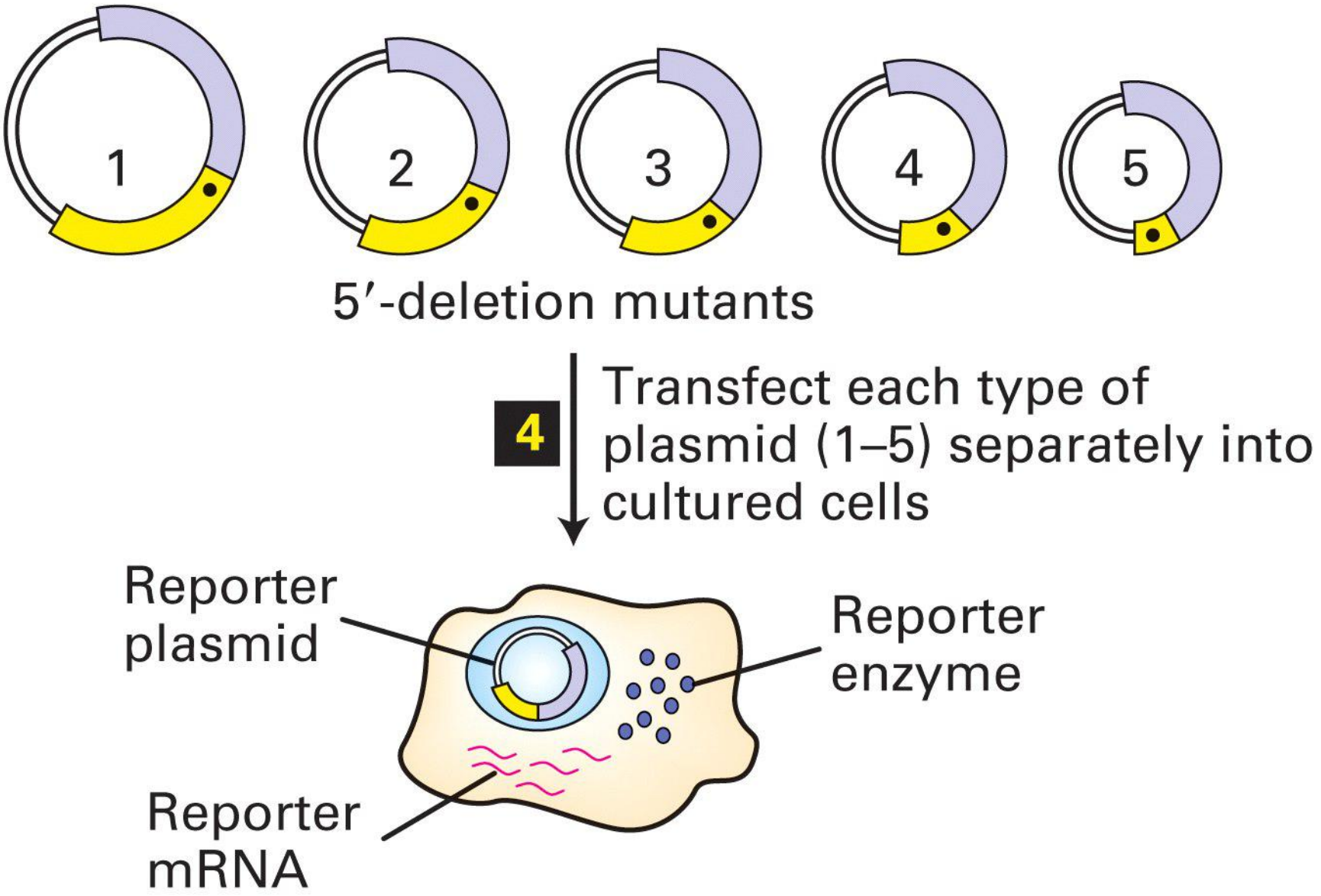


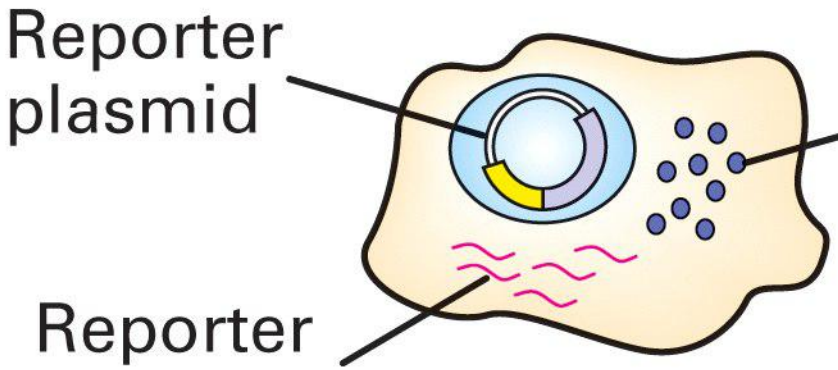
Plasmid no.	Reporter-gene expression
1	+++
2	+++
3	+
4	+
5	–

Luciferase
GFP
β-galactosidase
CAT


Upstream region
of *TTR* DNA
Start
5′
3′
1
Recombinant DNA
techniques
1
2
3
4
5
5′-deletion series

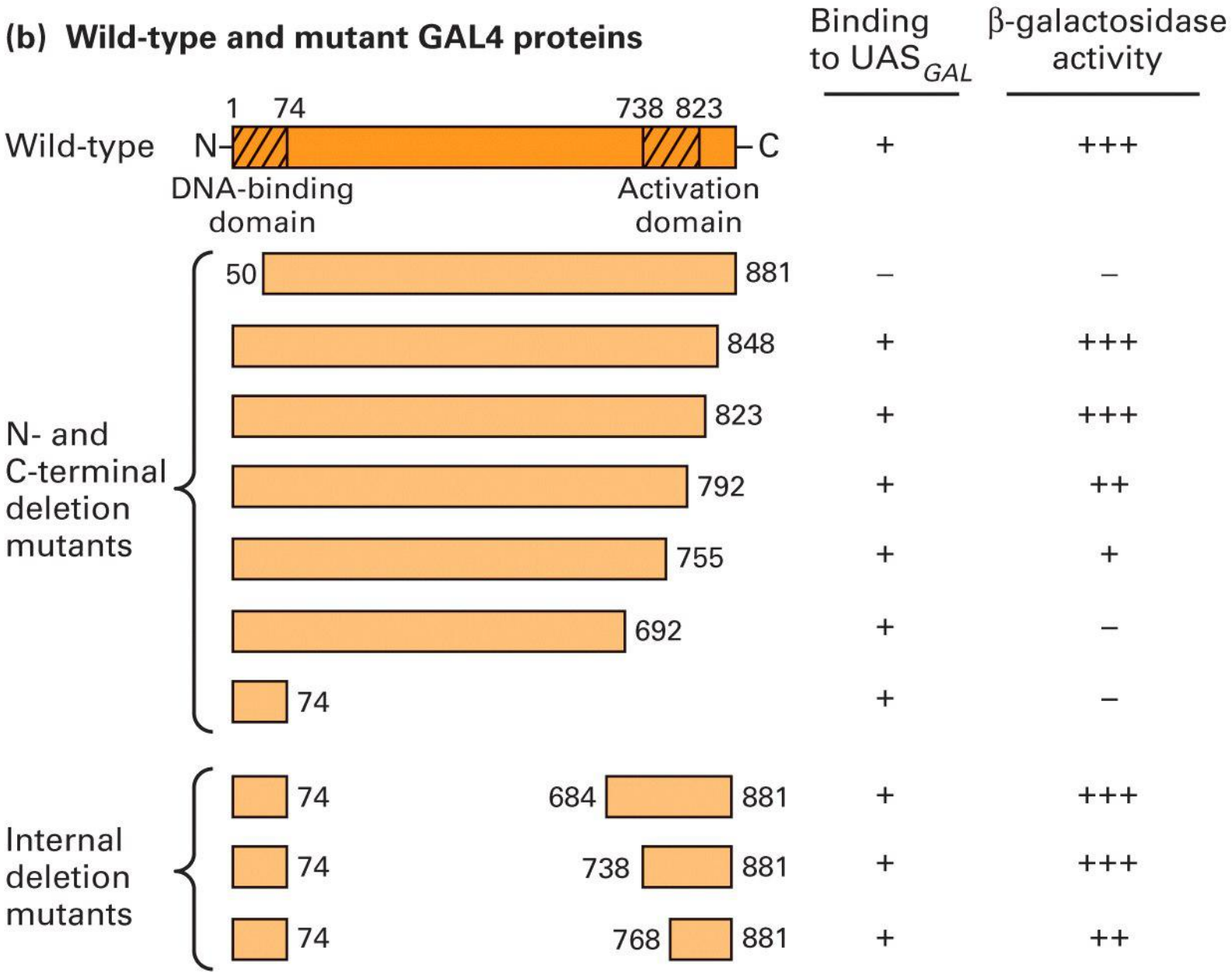

Reporter gene
1
2
3
4
5
5′-deletion series
Plasmid vector
2
Ligate into vector carrying reporter gene
3
Transform E. coli and isolate plasmid DNAs
1
2
3
4
5
5′-deletion mutants

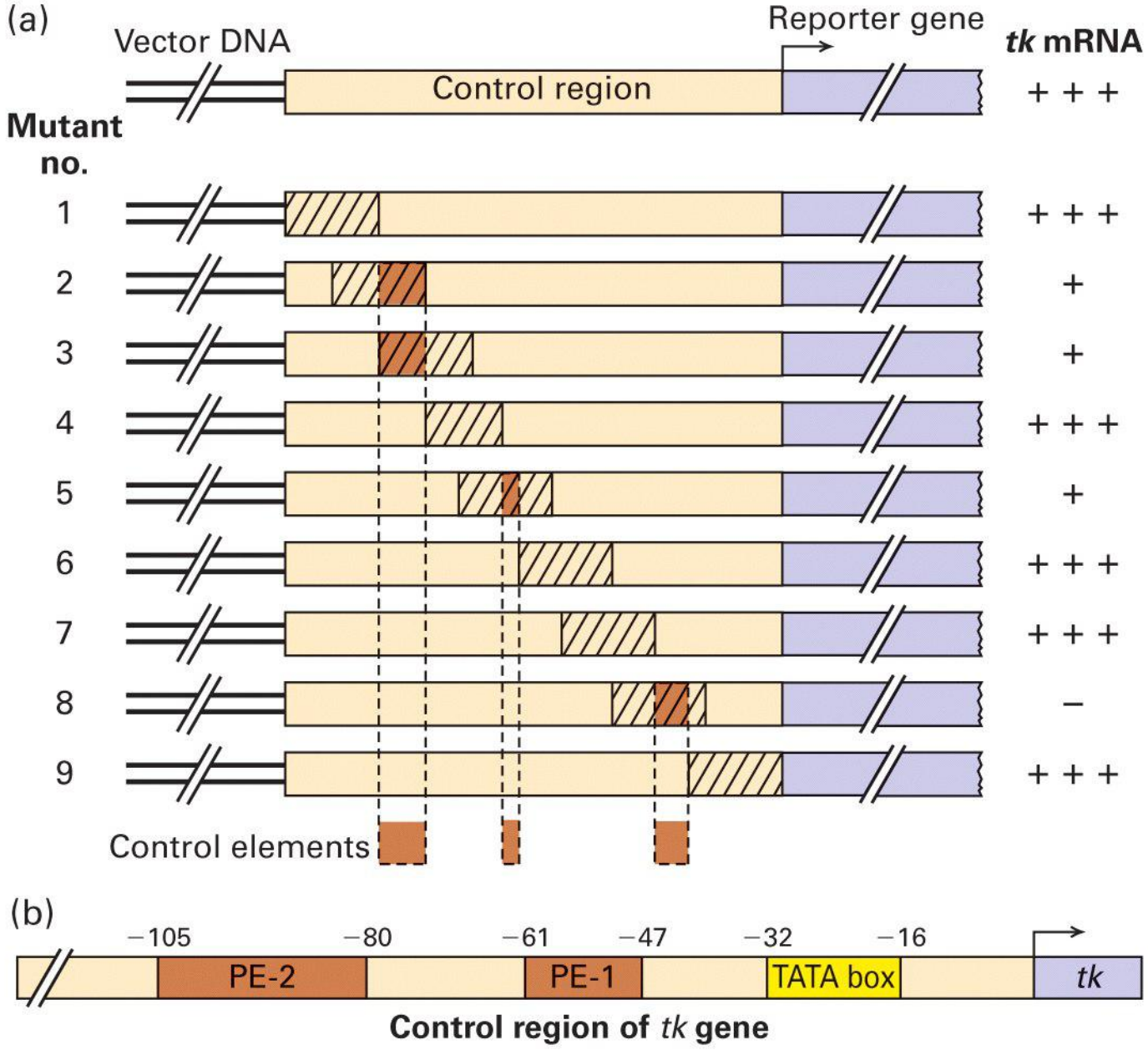

1
2
3
4
5
5′-deletion mutants
4
Transfect each type of plasmid (1–5) separately into cultured cells
Reporter plasmid
Reporter enzyme
Reporter mRNA



5 Prepare cell extract and assay activity of reporter enzyme

Plasmid no.	Reporter-gene expression	
1	+++	
2	+++	
3	+	
4	+	
5	–	

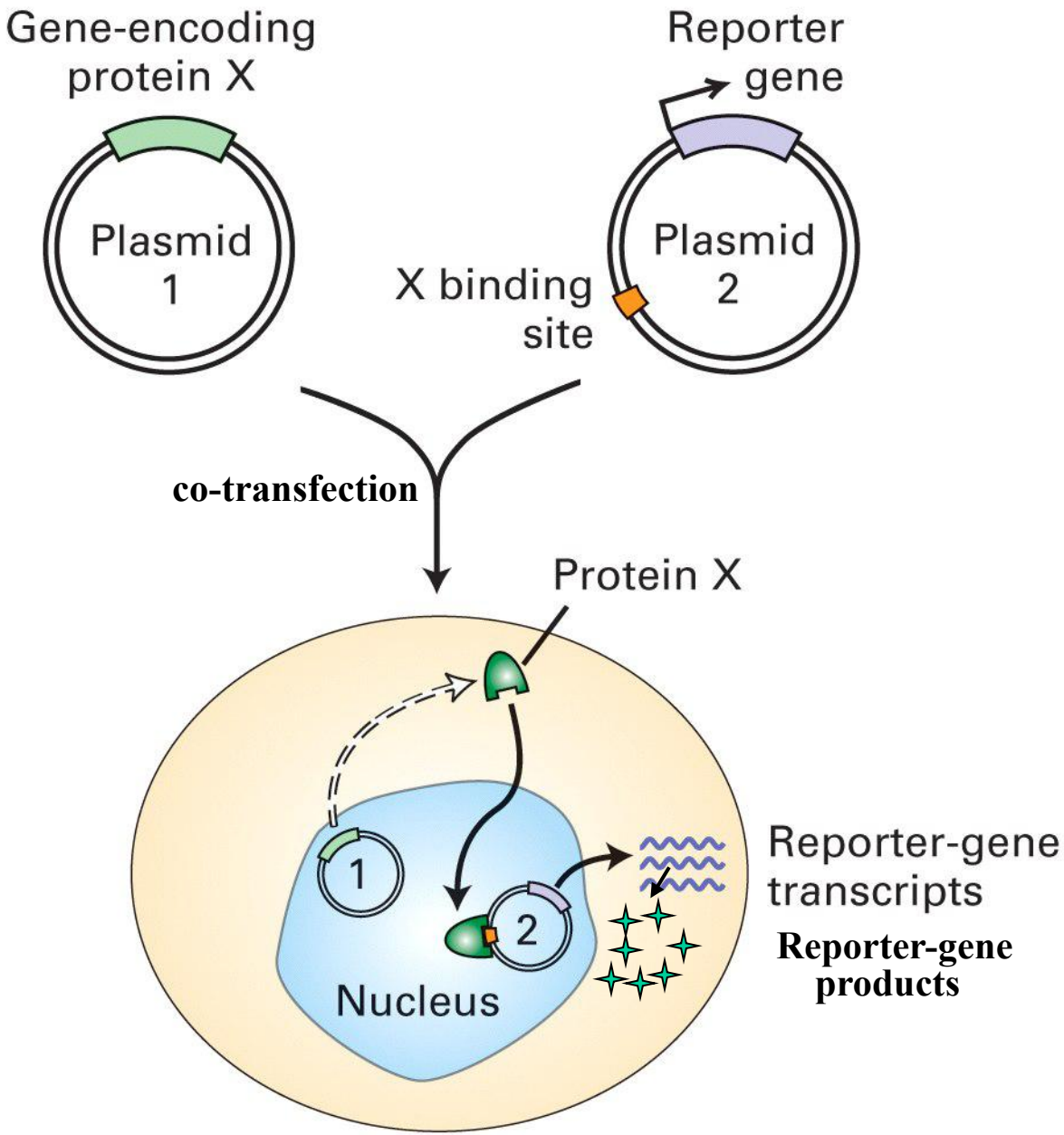

(b) Wild-type and mutant GAL4 proteins
Binding to UAS$_{GAL}$
β-galactosidase activity
Wild-type
1
74
738
823
N–
–C
DNA-binding domain
Activation domain
+
+++
N- and C-terminal deletion mutants
50
881
–
–
848
+
+++
823
+
+++
792
+
++
755
+
+
692
+
–
74
+
–
Internal deletion mutants
74
684
881
+
+++
74
738
881
+
+++
74
768
881
+
++


(a)
Vector DNA
Reporter gene
tk mRNA
Control region
+ + +
Mutant no.
1
+ + +
2
+
3
+
4
+ + +
5
+
6
+ + +
7
+ + +
8
–
9
+ + +
Control elements
(b)
–105
–80
–61
–47
–32
–16
PE-2
PE-1
TATA box
tk
Control region of tk gene

1. Find required promoter regions.

2. Identify activators or repressors.

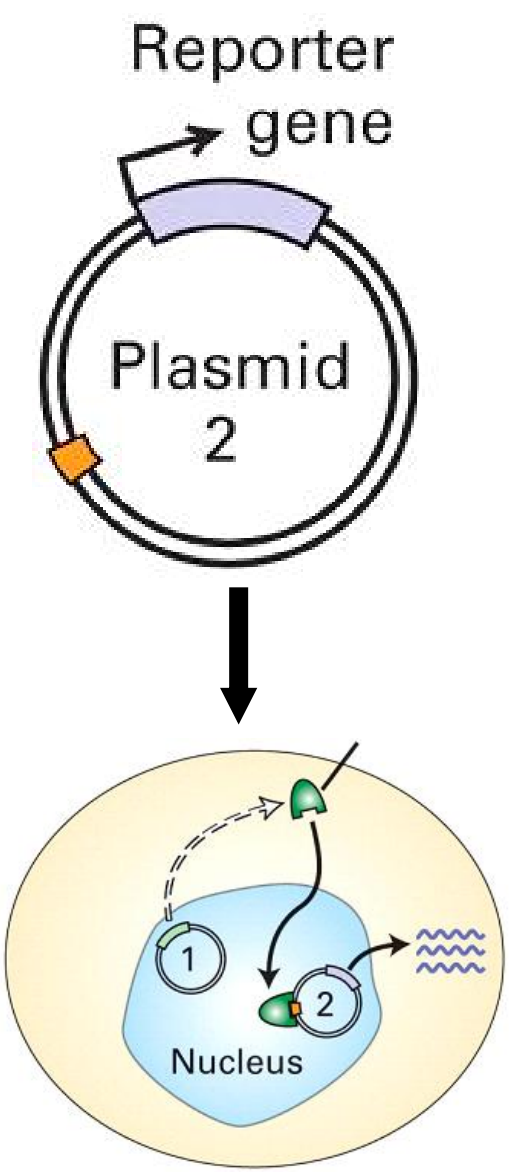
Reporter gene assay for transcription factor activity



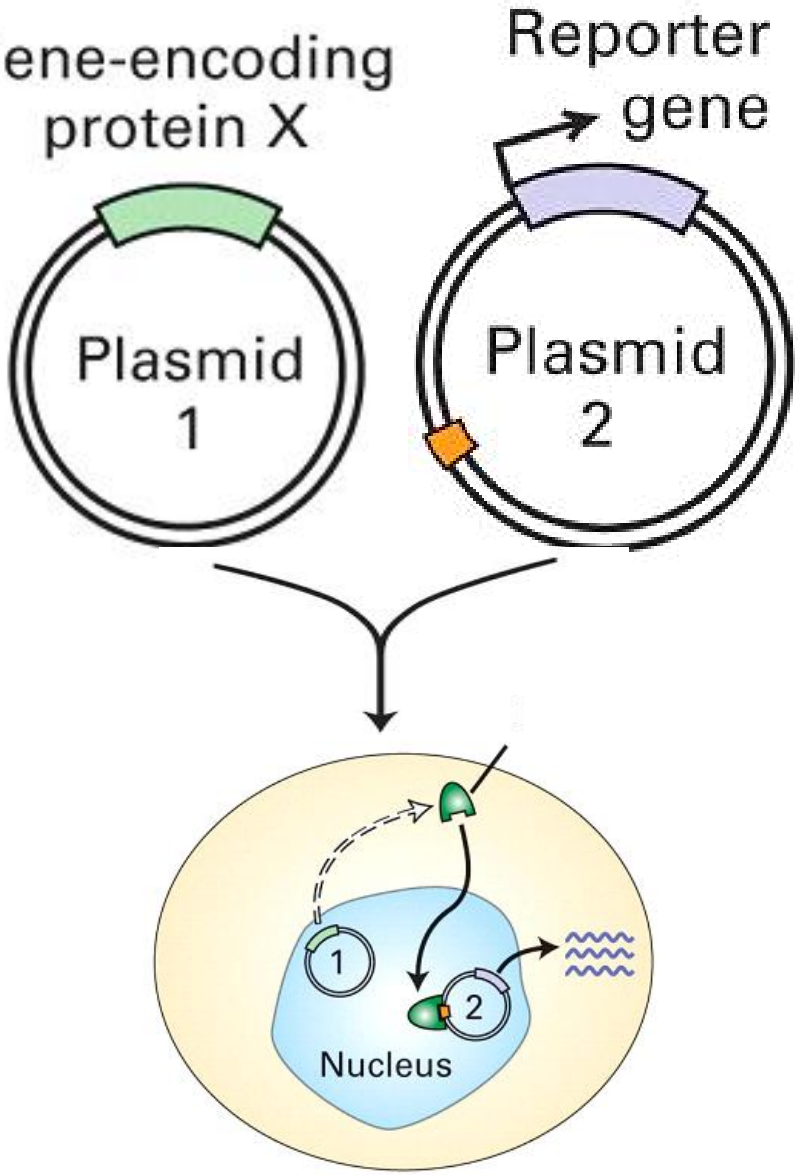
- **Host cells should lack the gene encoding protein X and the reporter protein.**

- **The production of reporter-gene RNA transcripts or the activity of the encoded protein can be assayed.**

- **If reporter-gene transcription is greater in the presence of the X-encoding plasmid, then X is an activator; if transcription is less, then X is a repressor.**



++



If ++++ X is an activator
If + X is a repressor

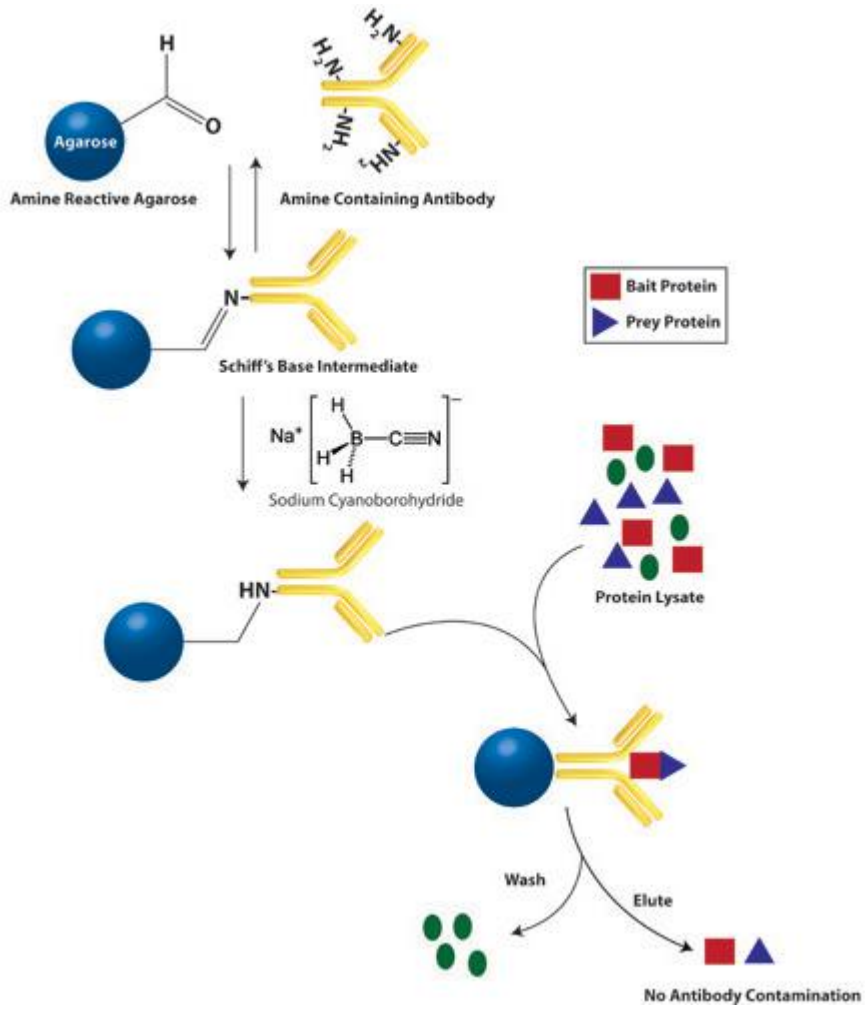
Western blot is always required to control the expression level of the ectopic proteins.

M is an activator binding core promoter and enhancer region of target genes

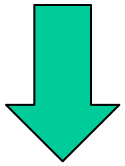
How can M activate transcription?

Find the protein partners for M

Immunoprecipitation



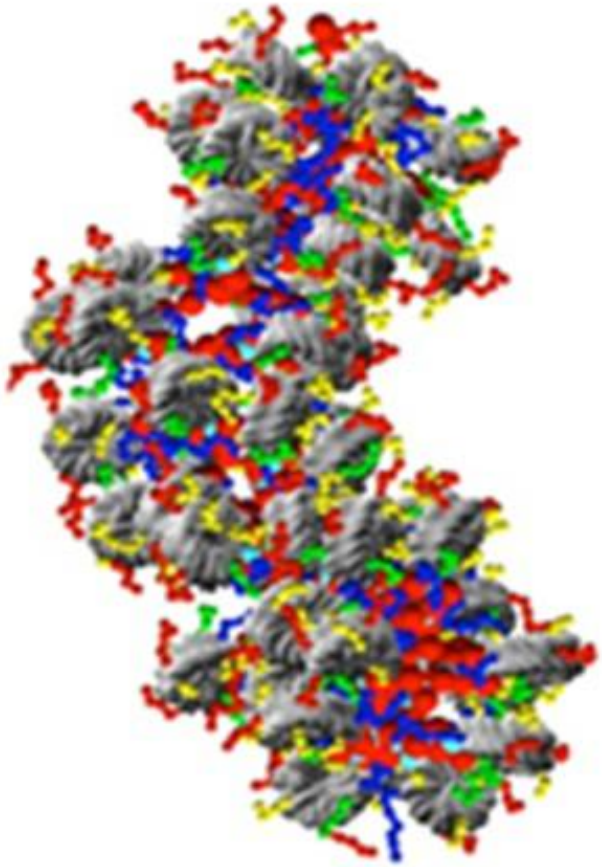
A chromatin binding protein and an DNA methlation transferase were identified as M interacting protein.

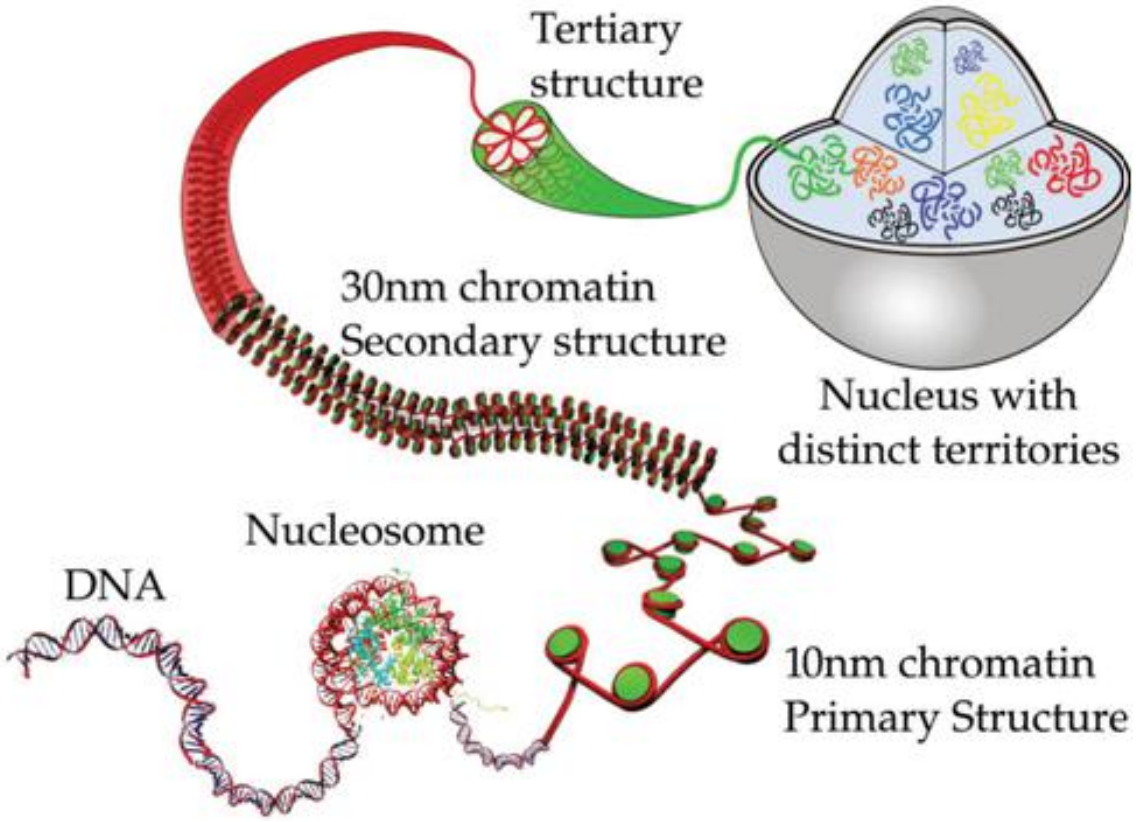


What to do next?

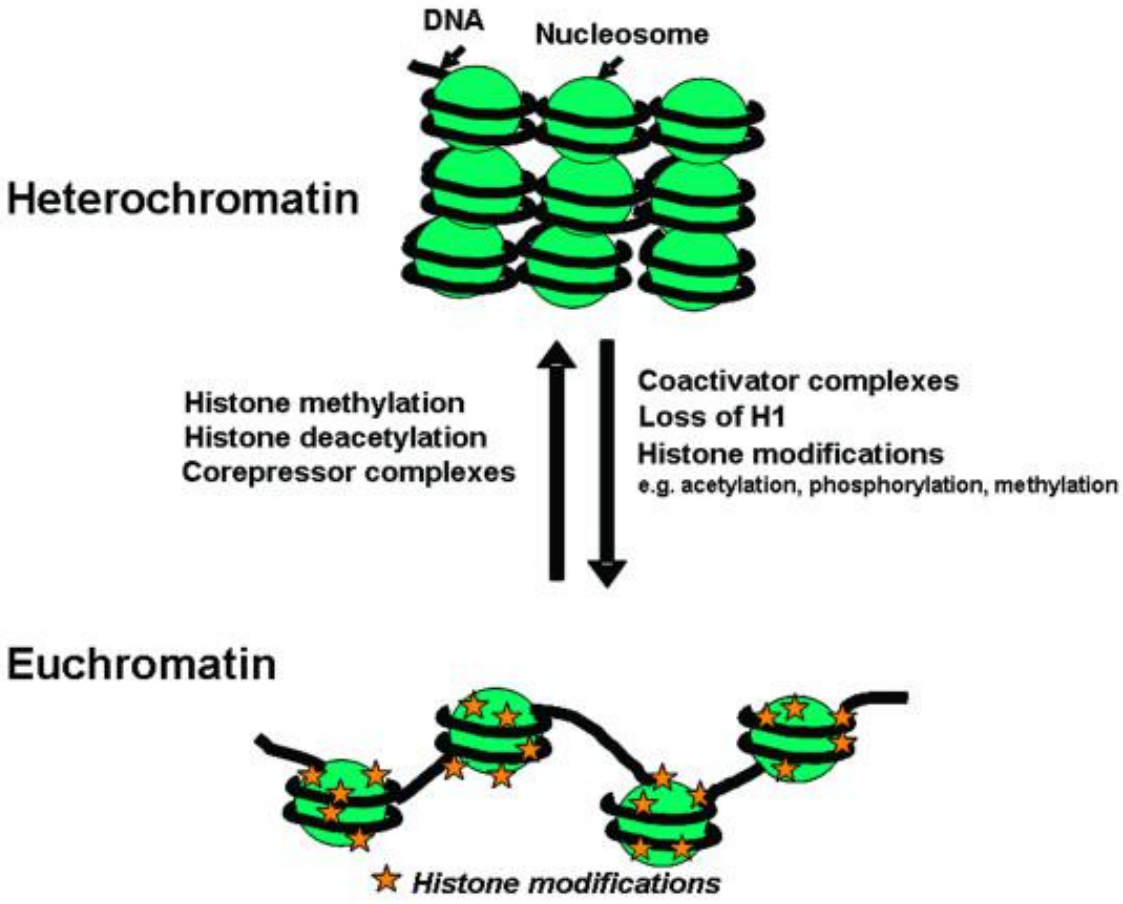
Chromatin structure is an important factor to regulate gene expression

Chromatin structure





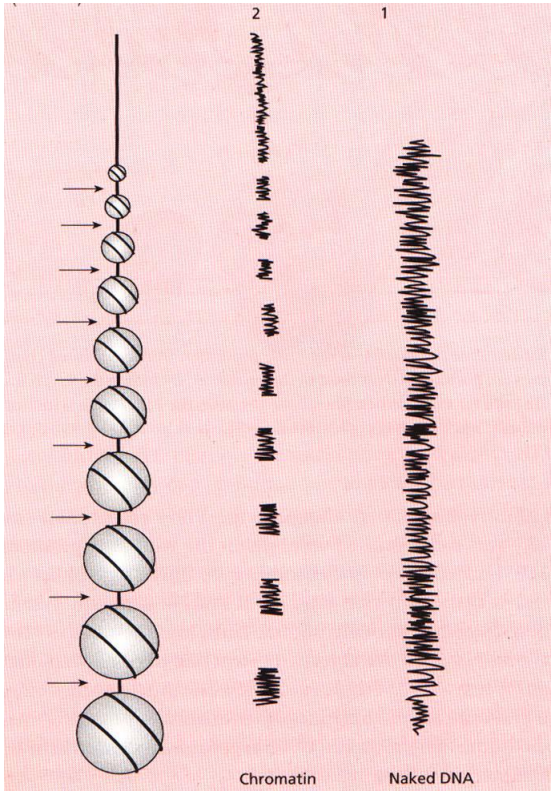
Euchromatin and heterochromatin



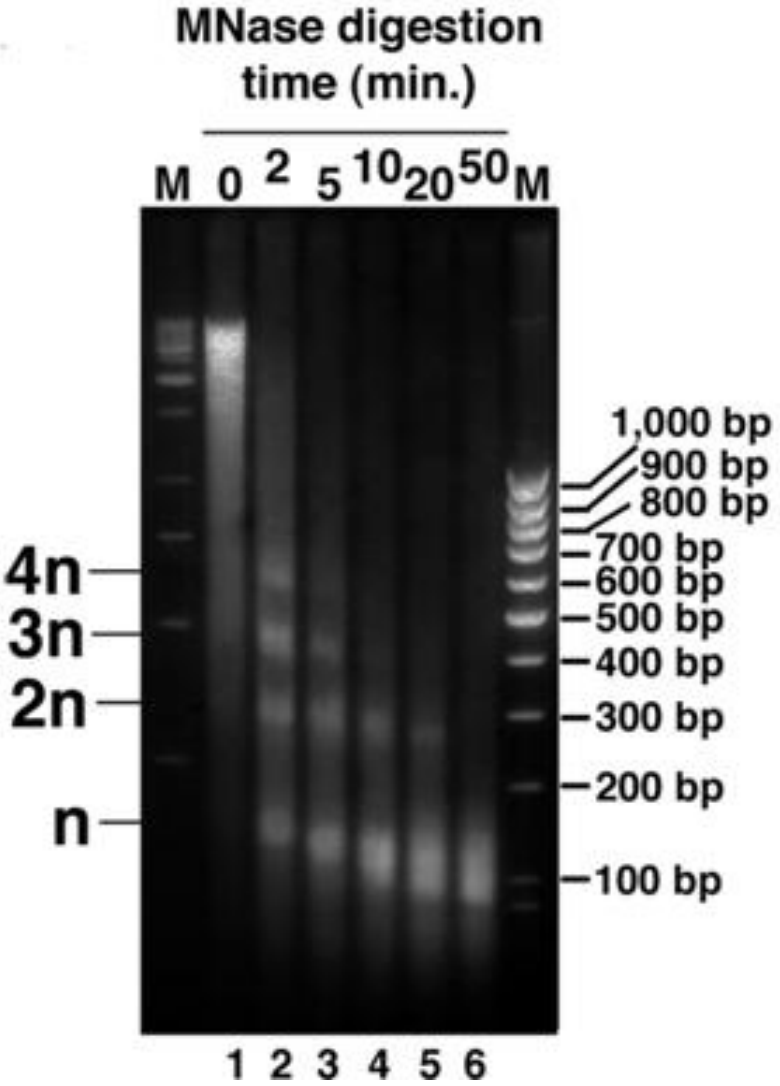
Micrococcal Nuclease Digestion of Chromatin

To map euchromatin regions

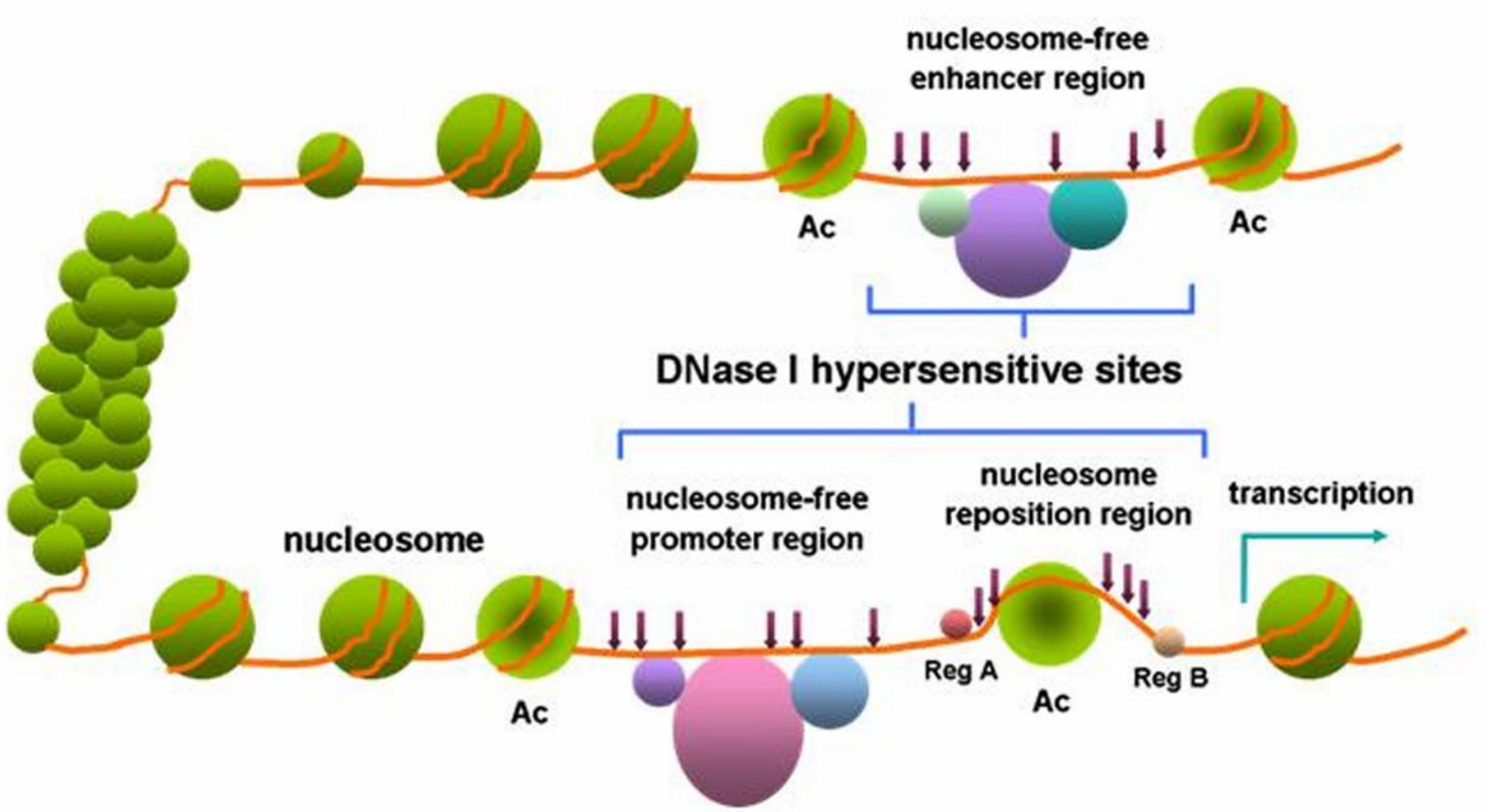
Appearance of hypersensitive sites is associated with transcriptional activation



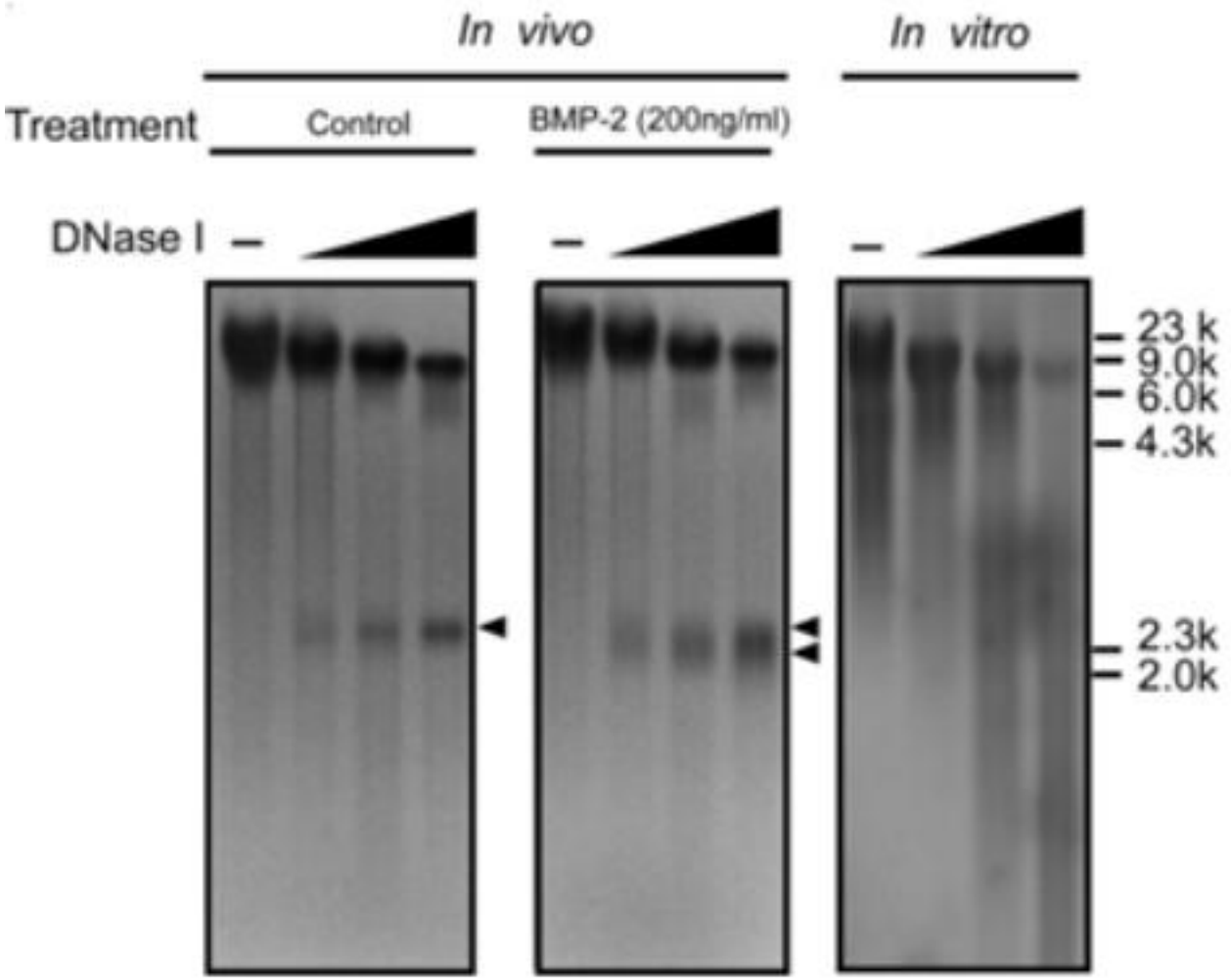
Sample result of micrococal nuclease digestion



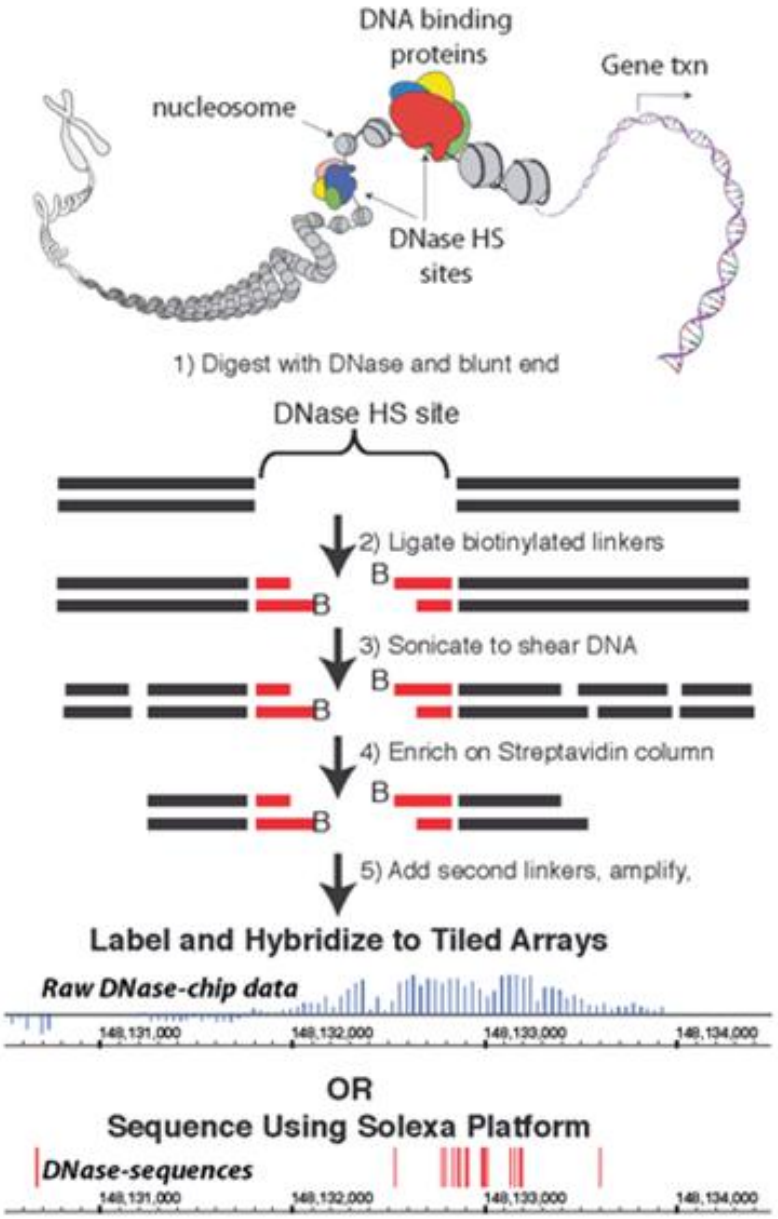
DNase I hypersensitive site mapping



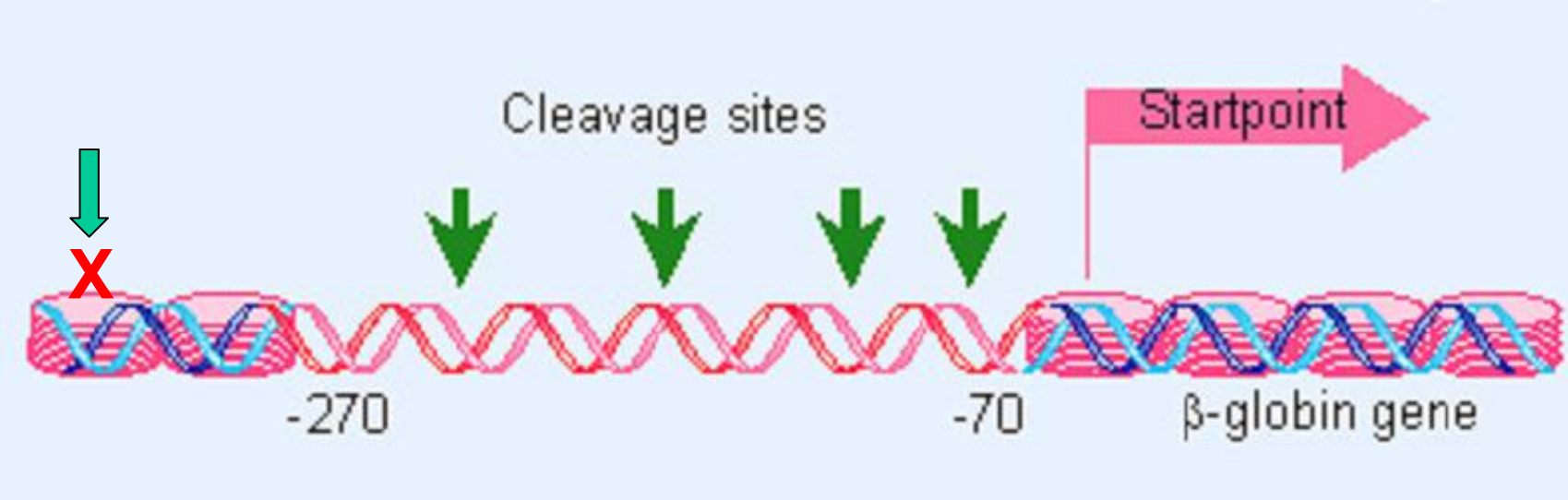
Sample result of Dnase I hypersensitive site mapping



High throughput Dnase I hypersensitive site mapping

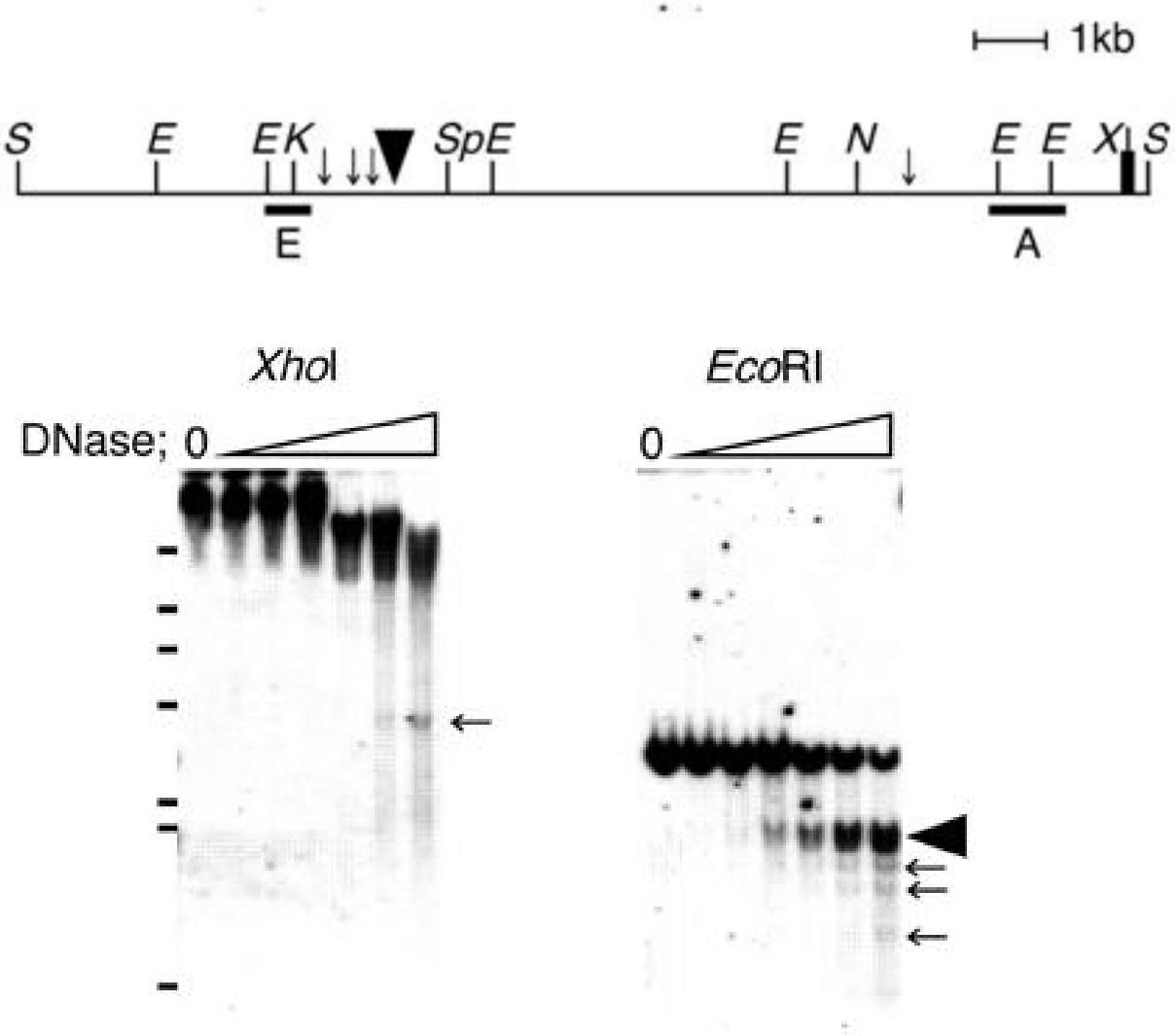


Restriction Enzyme Hypersensitive Sites Mapping



Gene VII, Fig. 19.41

Sample result of restriction enzyme hypersensitive site mapping



Micrococcal nuclease digestion

- Most commonly used
- Can be used in high throughput assays
- Some regions can not be digested by micrococcal nuclease

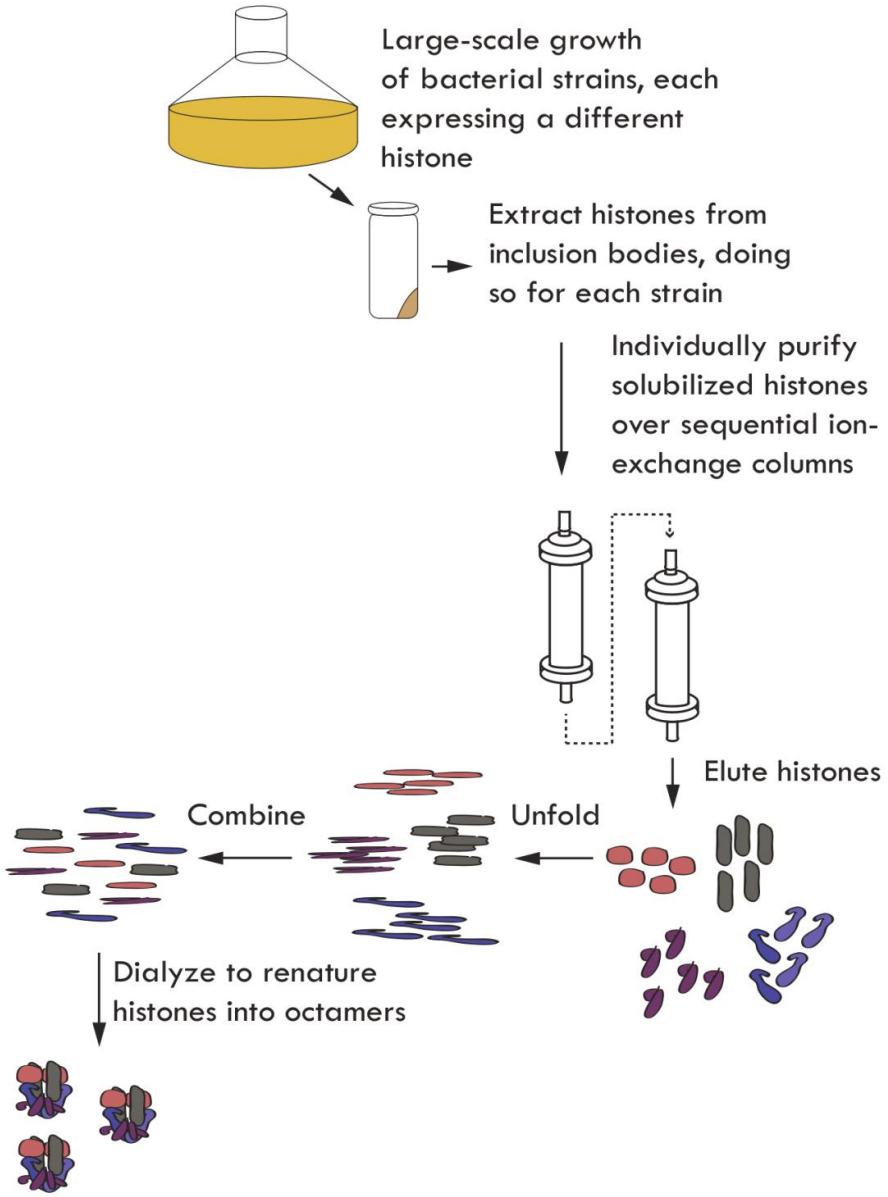
Dnase I digestion

- An alternative method of micrococcal nuclease digestion
- Can be used in high throughput assays
- More sensitive to the experimental conditions

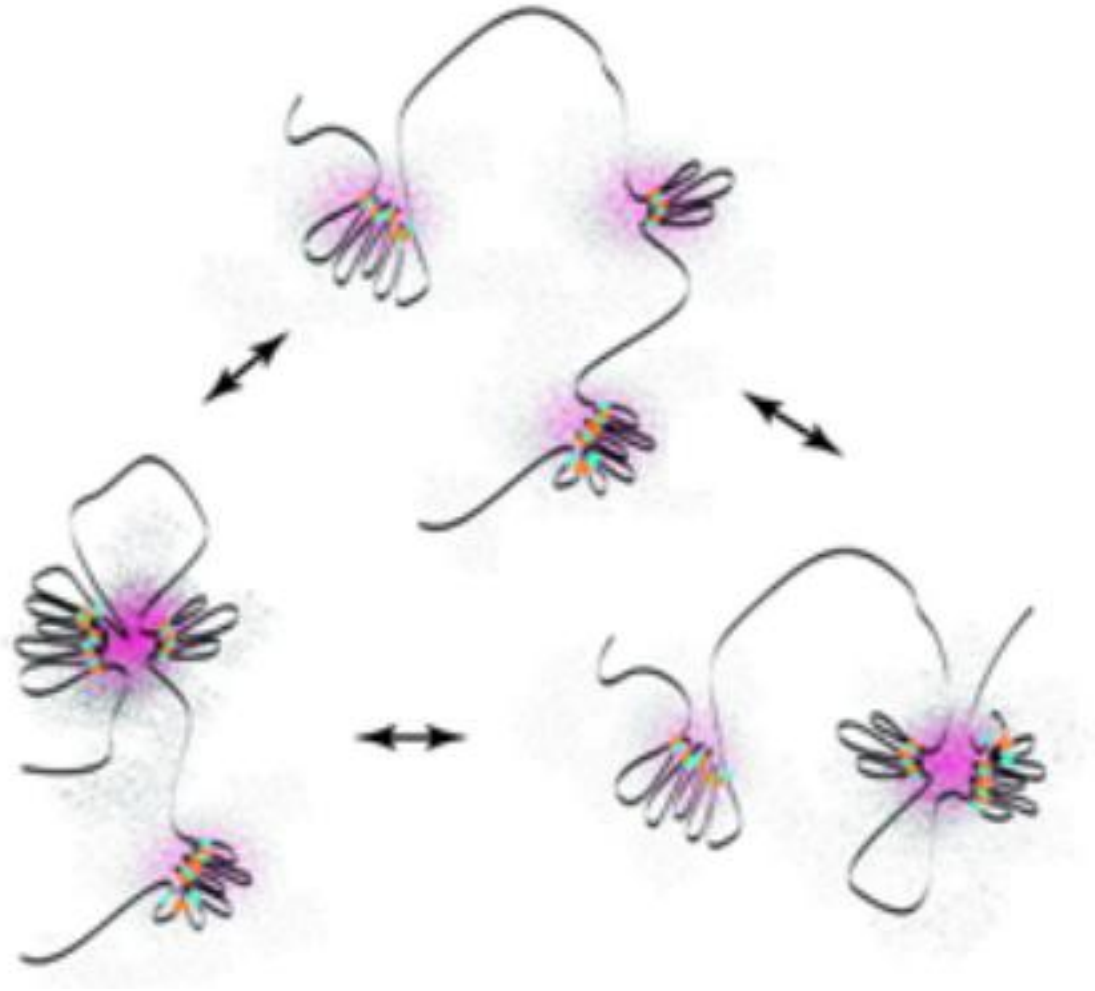
Restriction enzyme digestion

- Easy to perform
- Can not be used in high throughput assays
- False negative

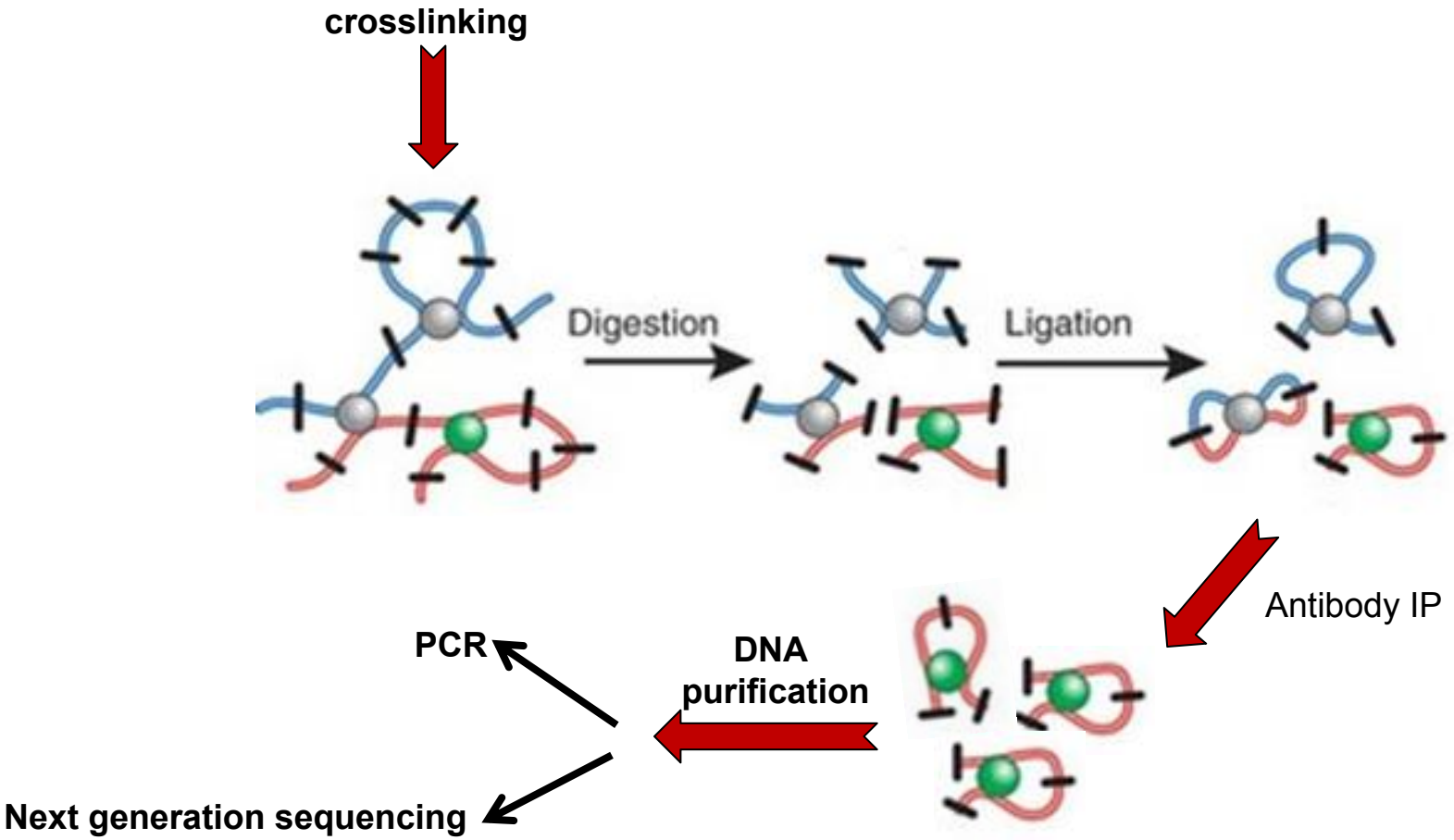
Nucleosome assembly



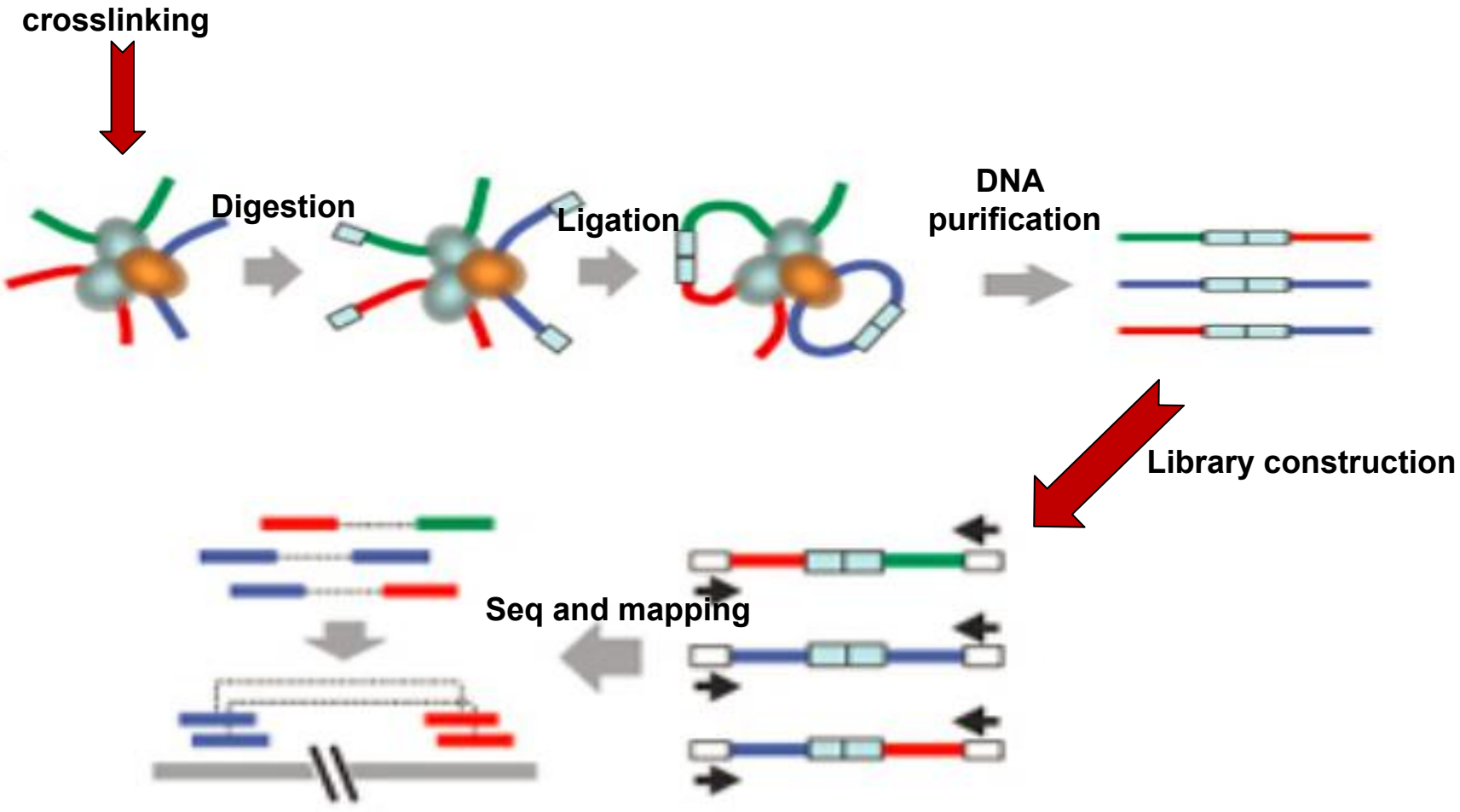
Dynamic chromatin structure



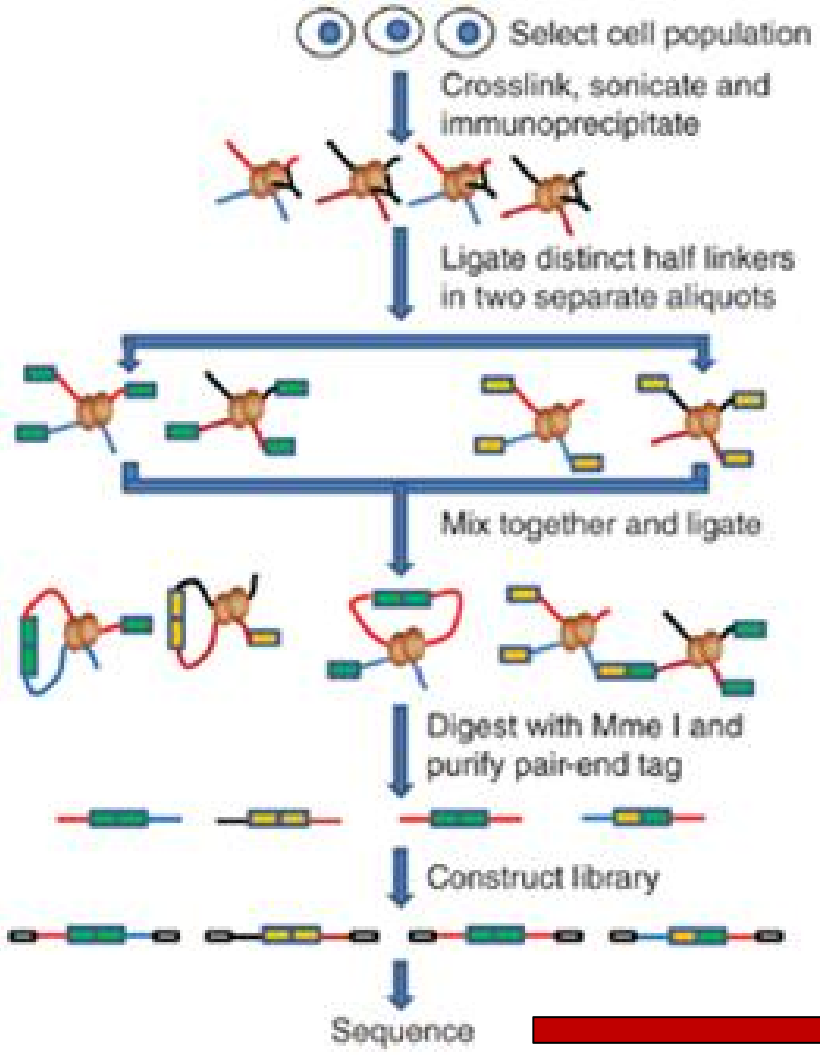
ChIP-loop

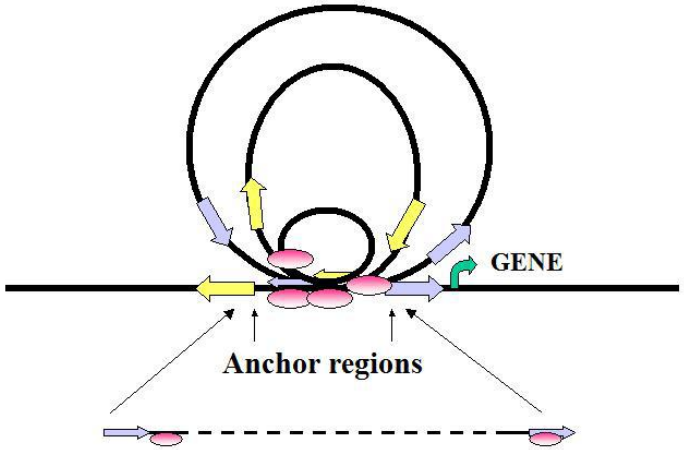

crosslinking
Digestion
Ligation
Antibody IP
PCR
DNA
purification
Next generation sequencing

Hi-C Chromatin proximity ligation

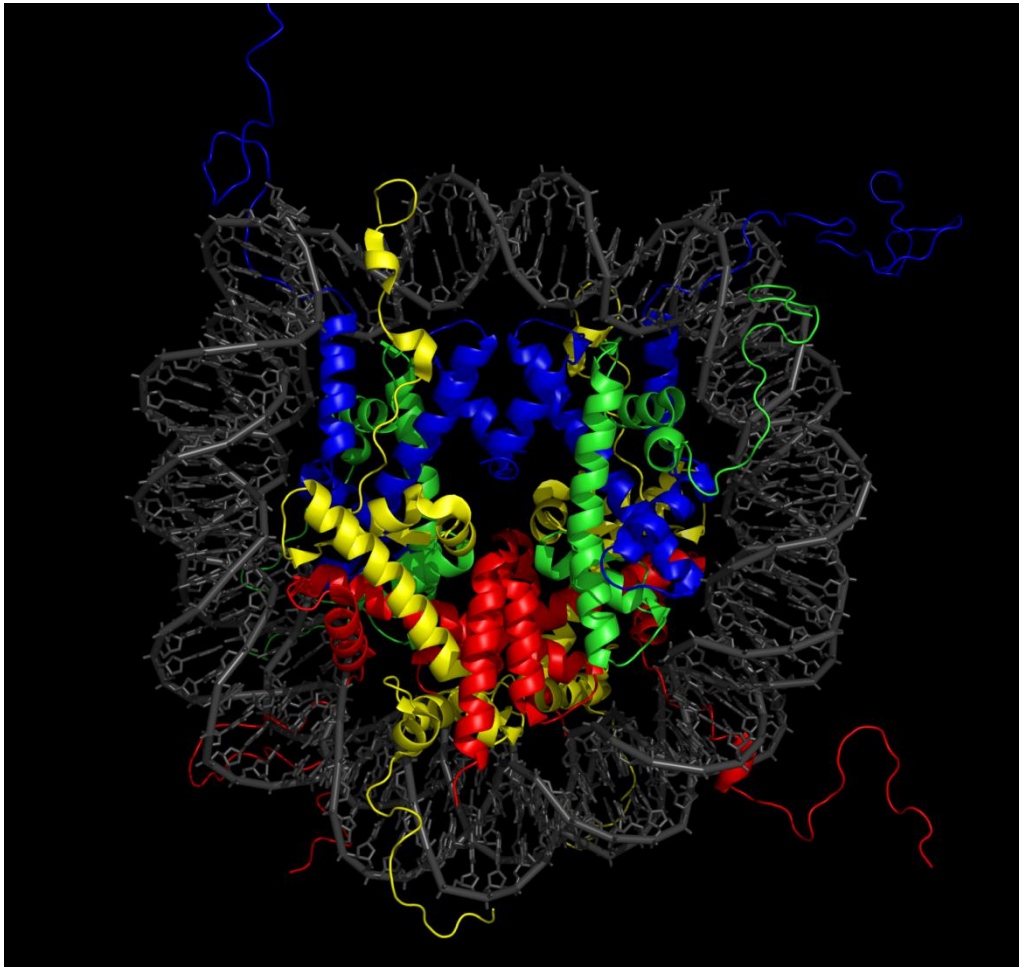


ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag Sequencing)

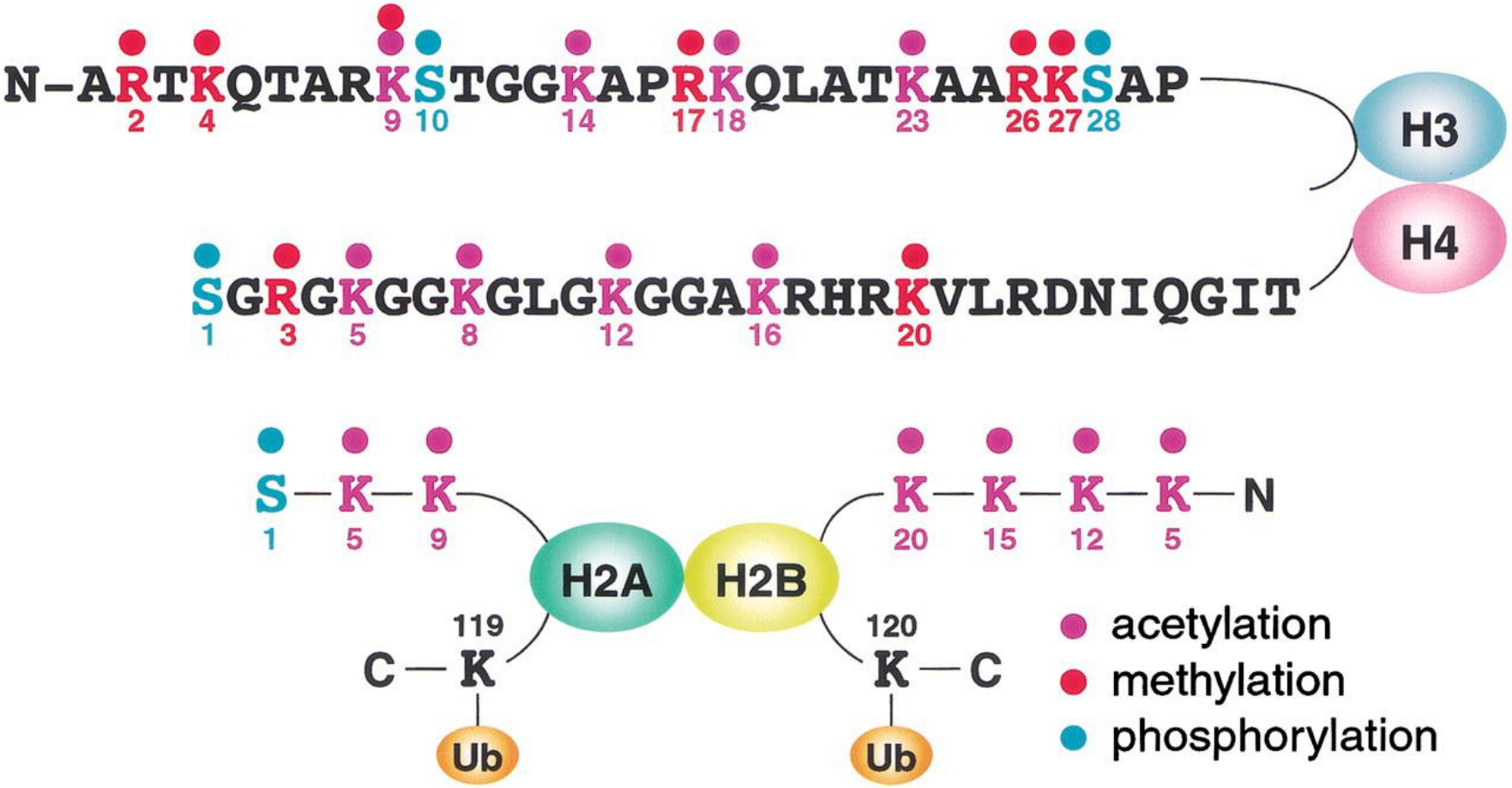




Structure of Histone Tails

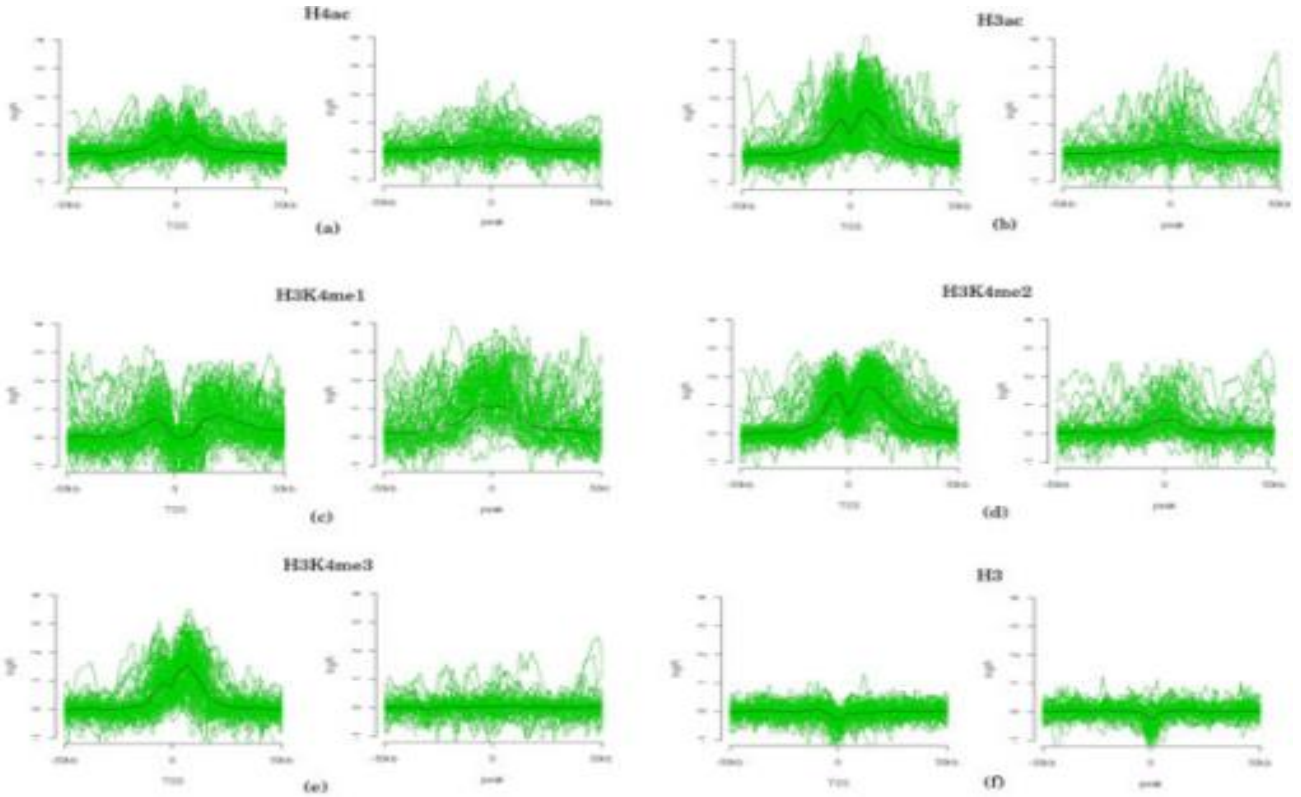


Modification of Histone tails



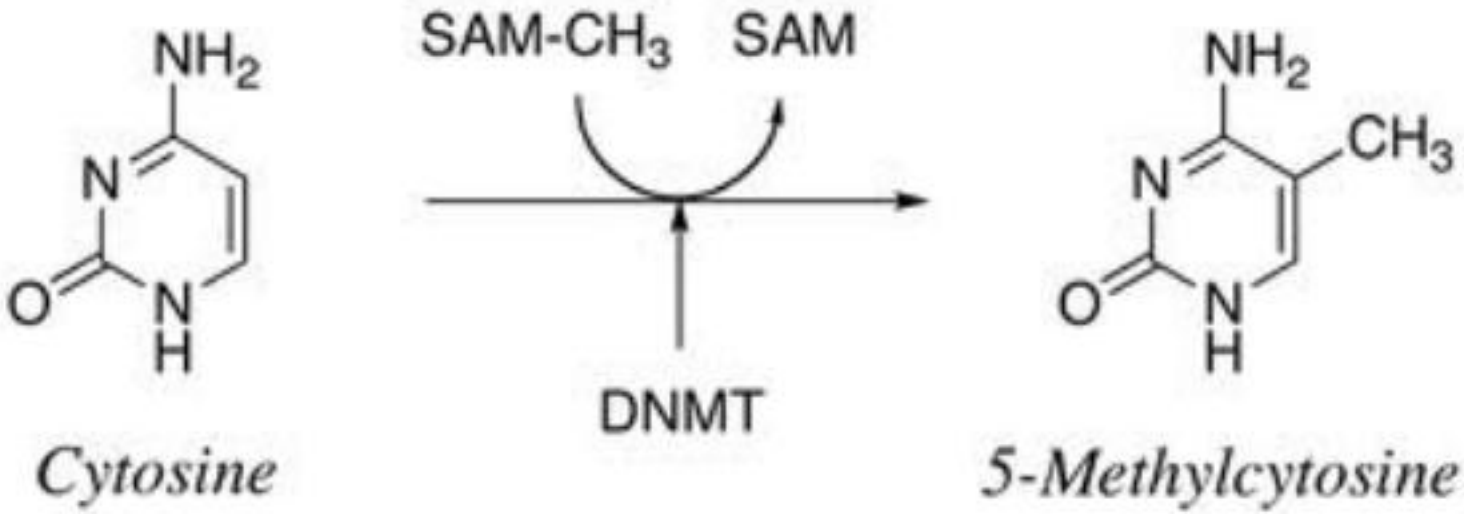
Genomewide histone code?

H3K4me1 and H3K27Ac: Active Enhancer
H3K4me3: Transcription initiation site
H3K9me3 and H3K27me3: Heterochromatin
H3K36me3: Actively transcribed gene bodies
H3K9me3 and H3K27me2: dual histone marks

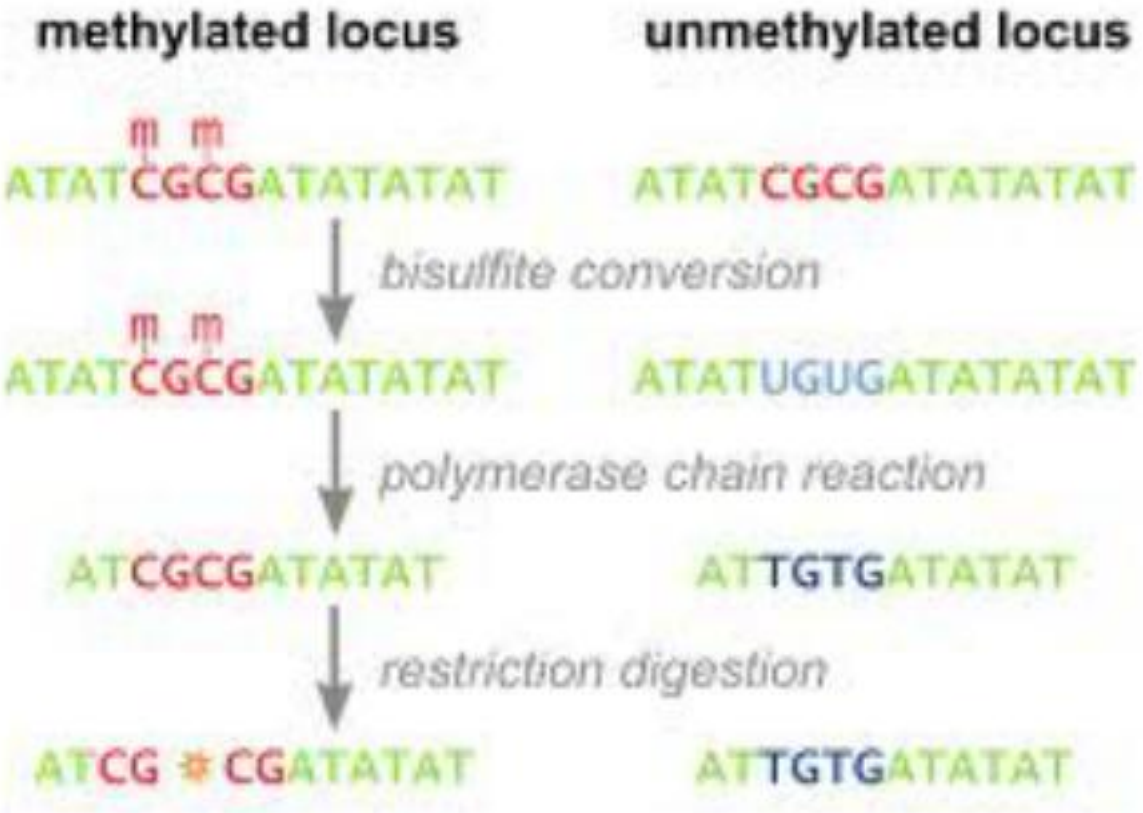


DNA modifications are important factors to regulate gene expression.

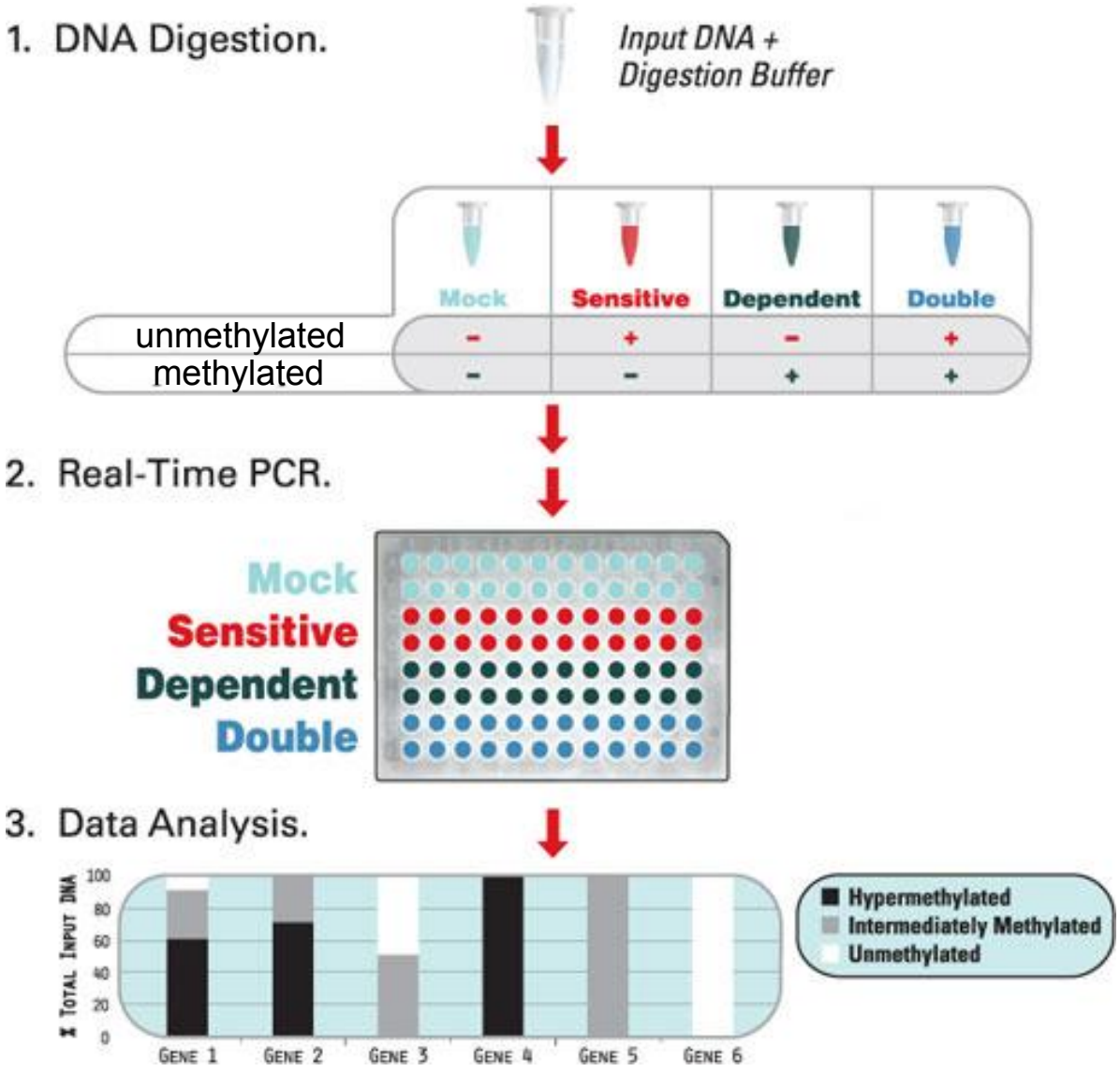
DNA methylation



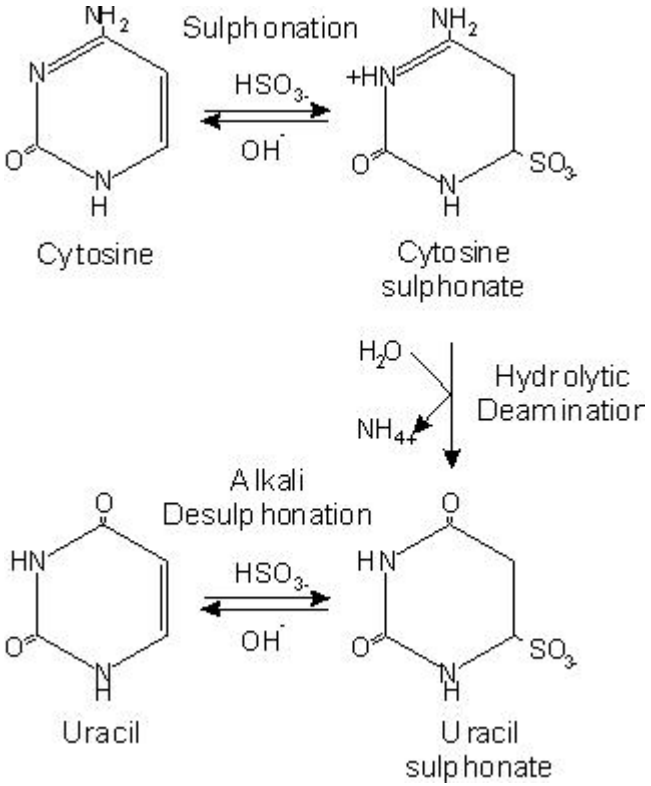
Restriction enzyme digestion to map DNA methylation

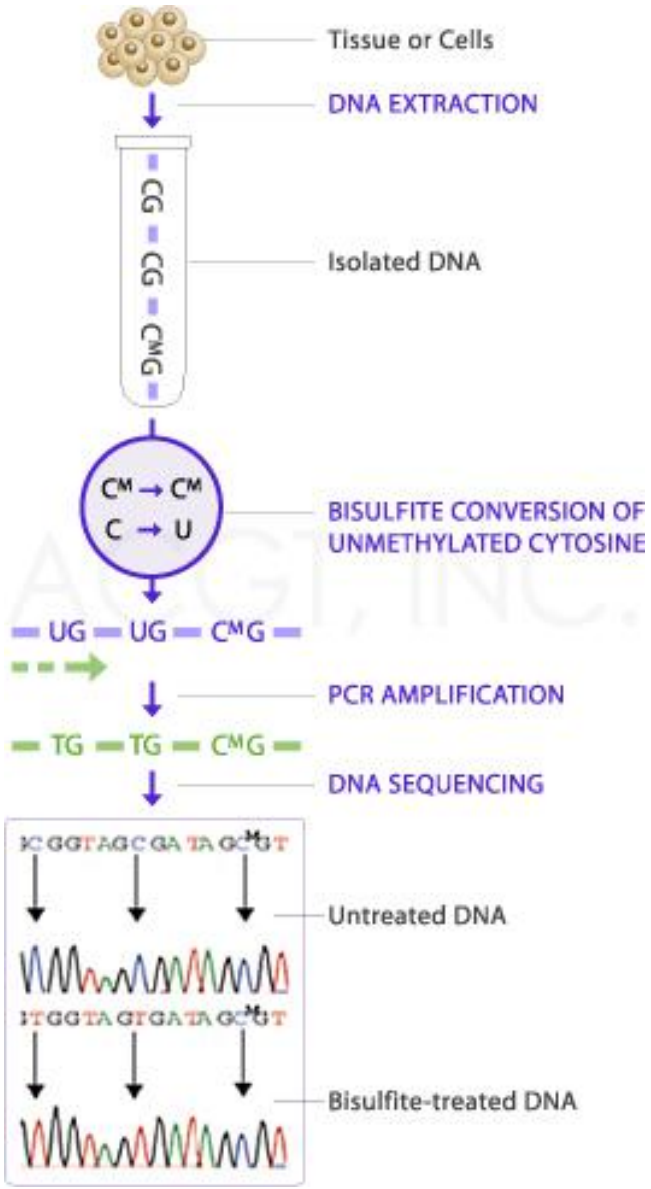


Restriction enzyme digestion coupled with qPCR to map DNA methylation

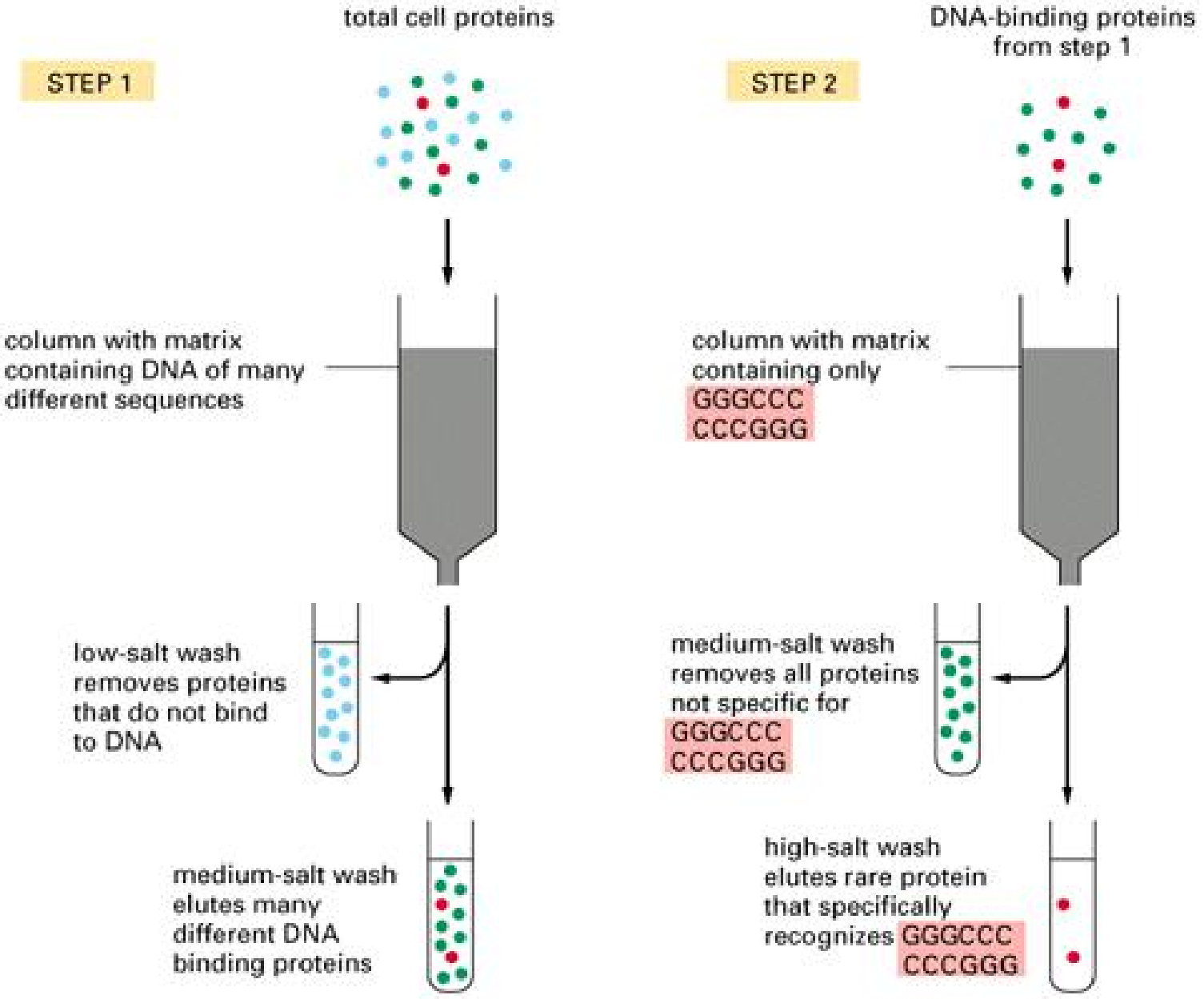


DNA bisulfite sequencing



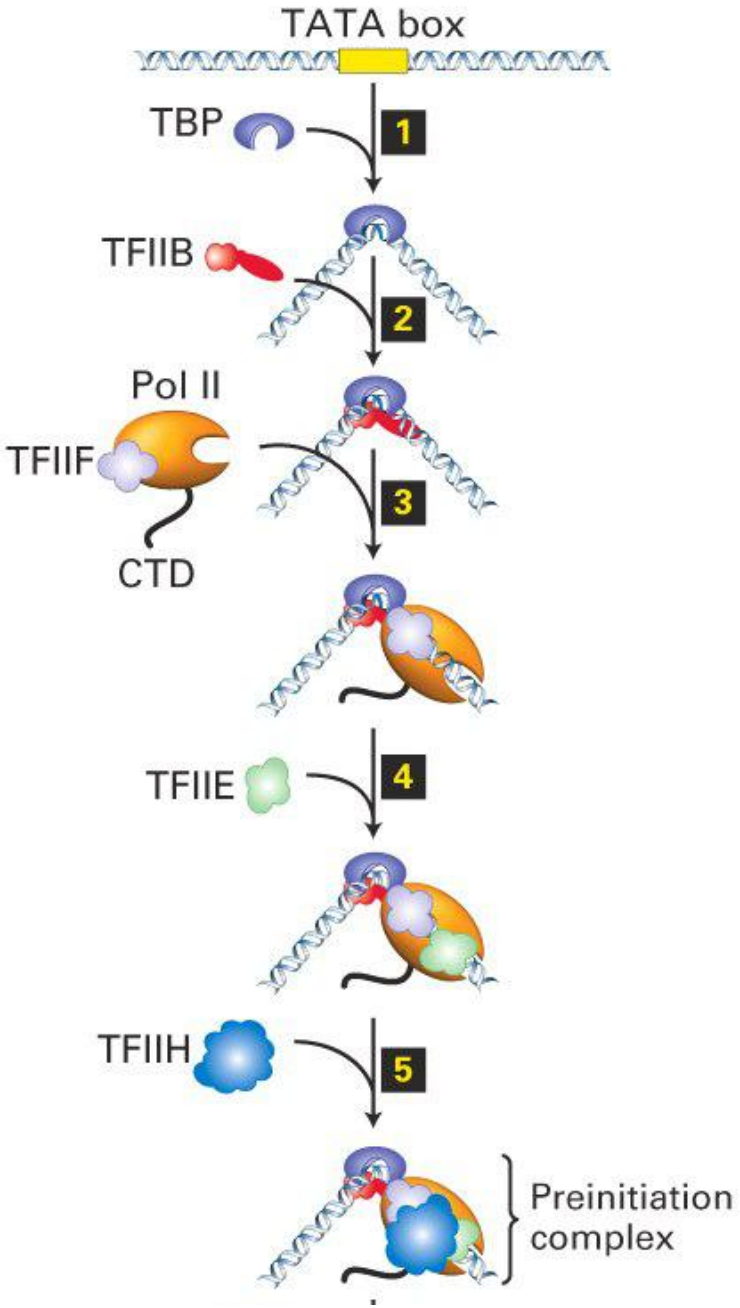


DNA affinity chromatography for purification of transcription factor



In vitro transcription

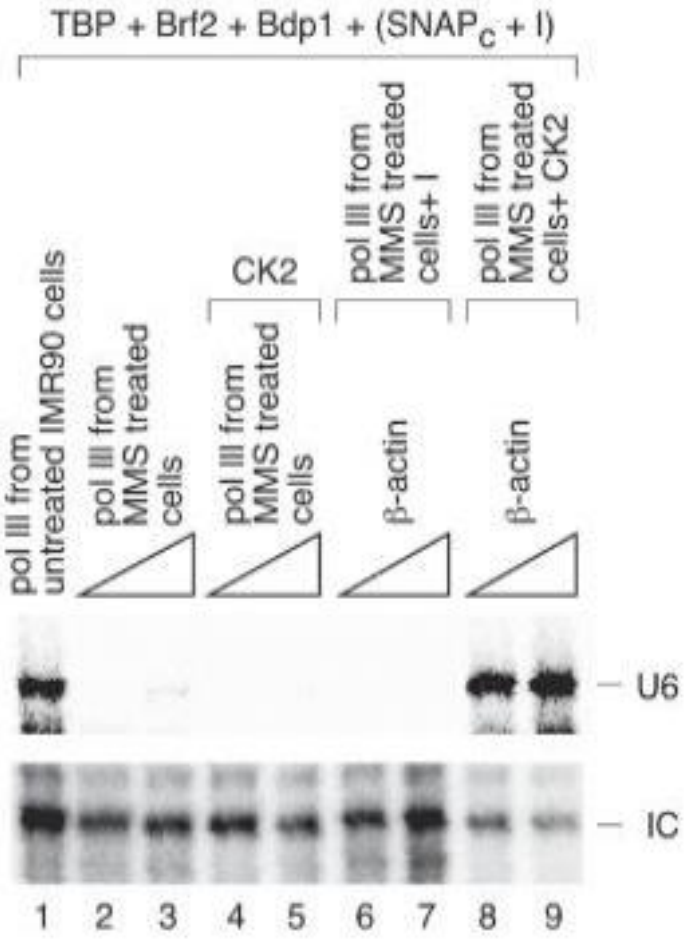
- RNA Polymerase II
- General Transcription Factors (GTFs): IID (TBP+TAFs), IIB, IIA, IIE, IIF, IIH
- Mediator Complex
- Complexes for chromatin remodeling and modifications



Lodish et al.
Molecular Cell Biology 5e

Fig. 11-27

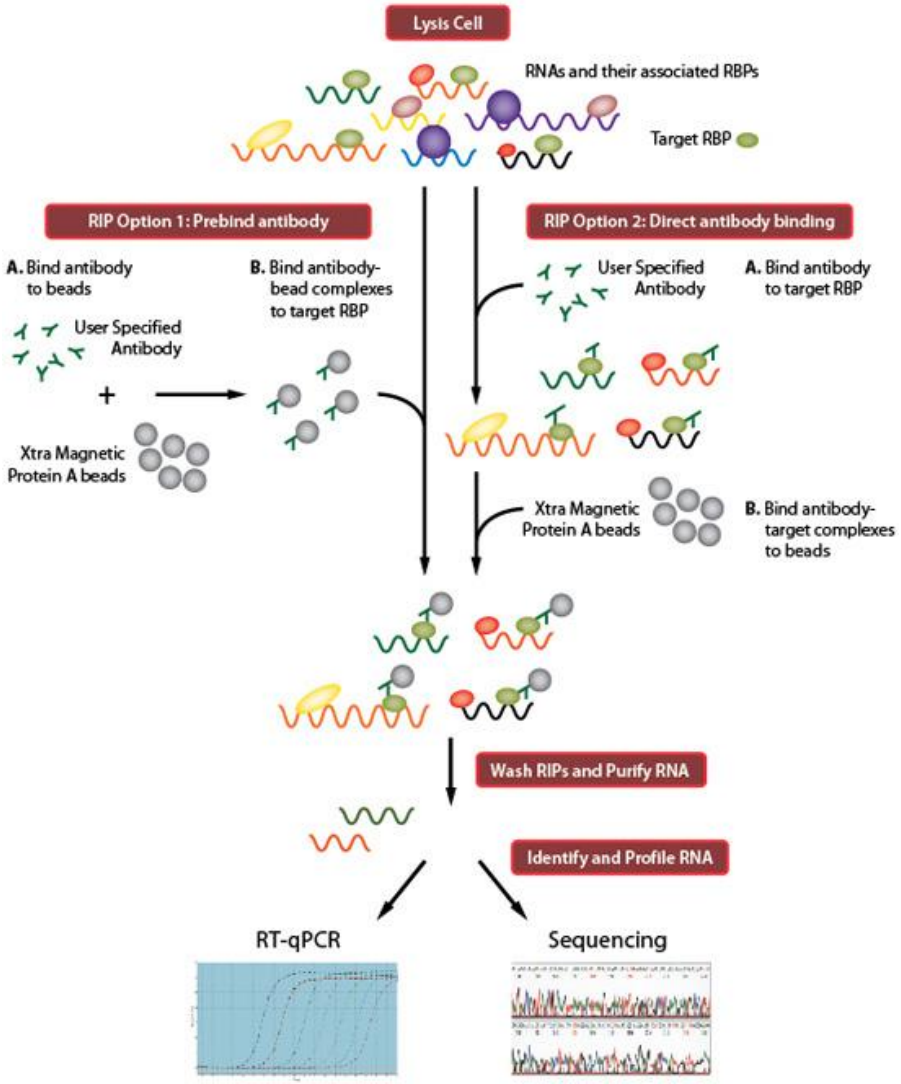
An Example of in vitro transcription experiment



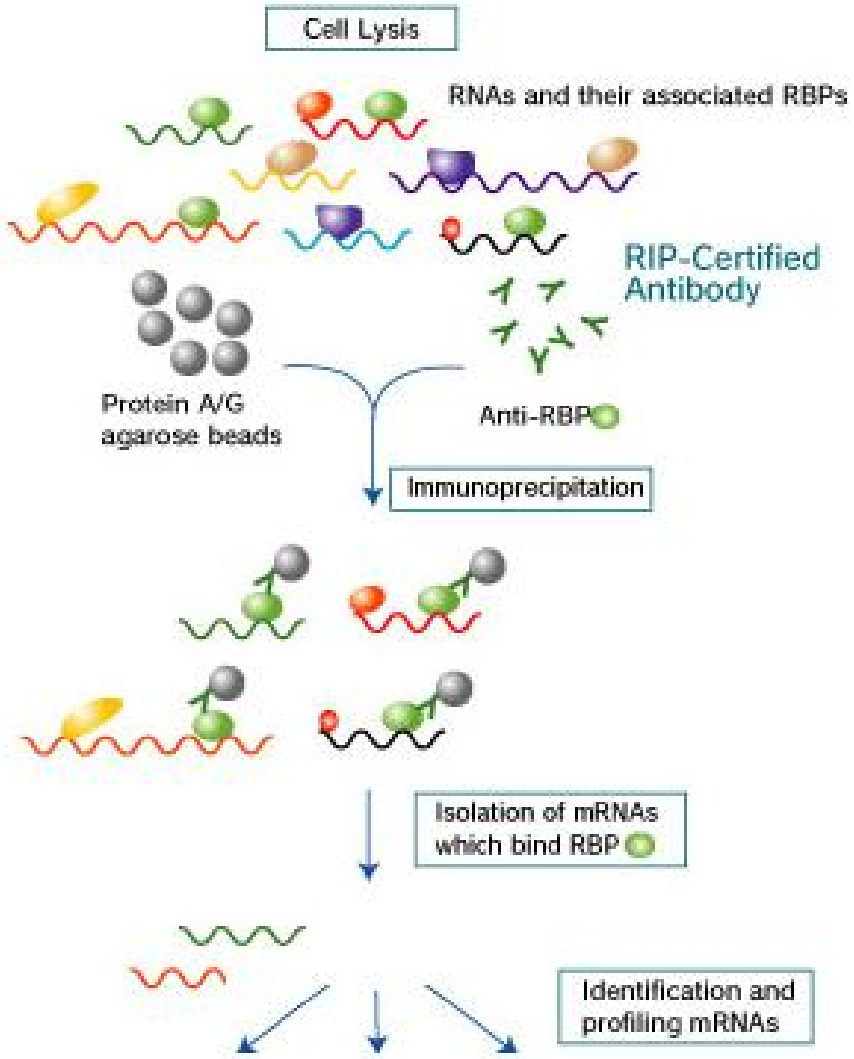
Hu et al. ***Genes and Development, 2004, 18(24): 3010-3015***

M is located in cytoplasm.

RNA IP (RIP)

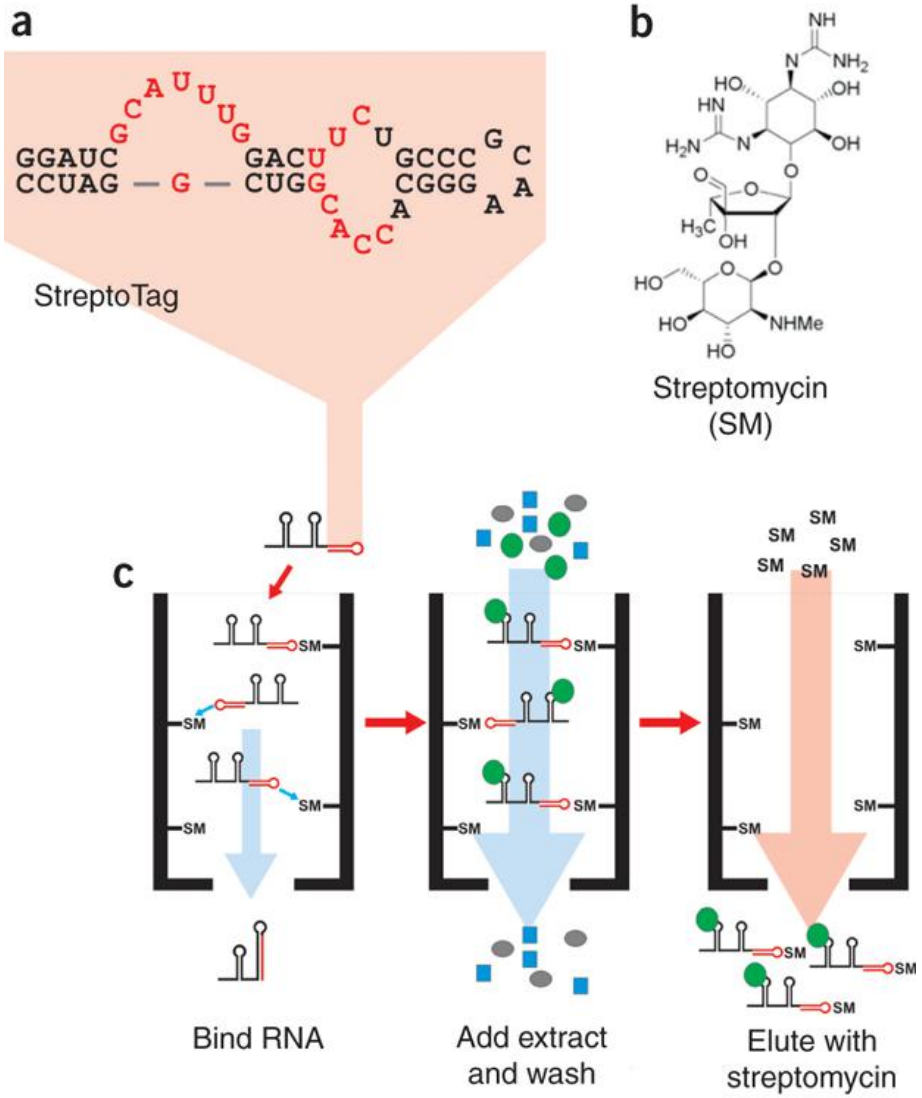

Lysis Cell
RNAs and their associated RBPs
Target RBP
RIP Option 1: Prebind antibody
A. Bind antibody to beads
User Specified Antibody
+
Xtra Magnetic Protein A beads
B. Bind antibody-bead complexes to target RBP
RIP Option 2: Direct antibody binding
User Specified Antibody
A. Bind antibody to target RBP
Xtra Magnetic Protein A beads
B. Bind antibody-target complexes to beads
Wash RIPs and Purify RNA
Identify and Profile RNA
RT-qPCR
Sequencing

RIP-seq



Deep sequencing

Isolation of specific RNA binding proteins



Techniques for studying gene expression

- Promoter analysis
- Hypersensitive site mapping/ligation mediated PCR
- Real time PCR
- siRNA knockdown of a gene
- Gel shift (or EMSA)
- Footprinting
- Chromatin IP
- RNase protection assay
- Immunoblot/Western Blot
- Immunofluorescent staining

Science is fun!