

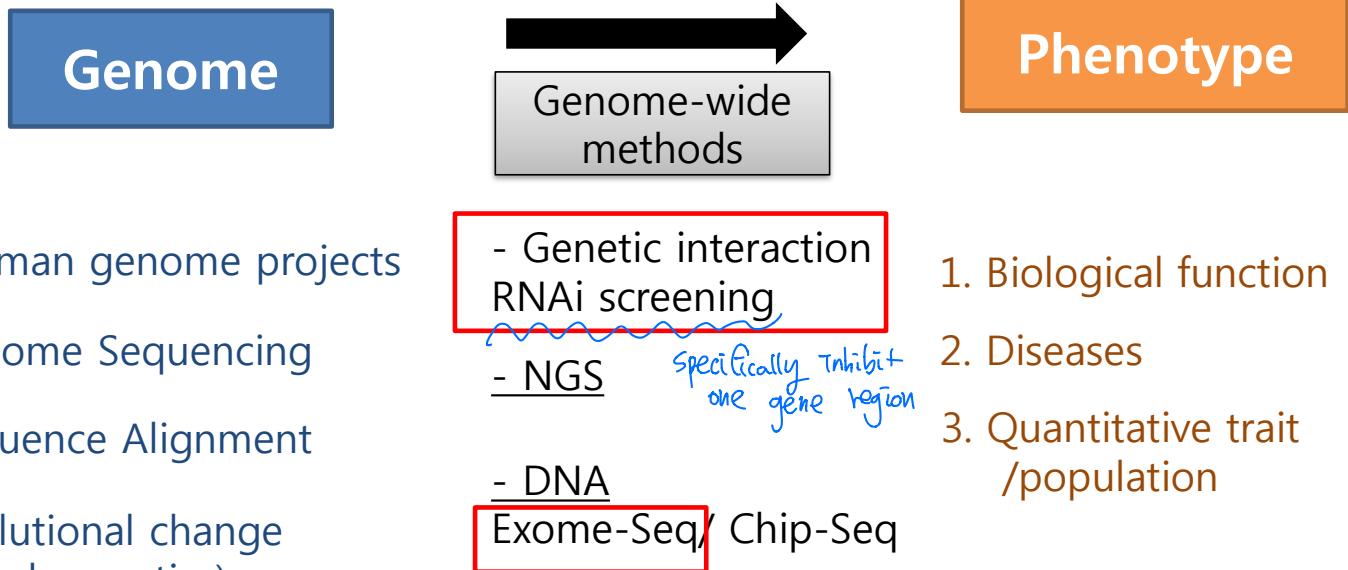
# Summary & Review

:Exome-Seq, functional genetics

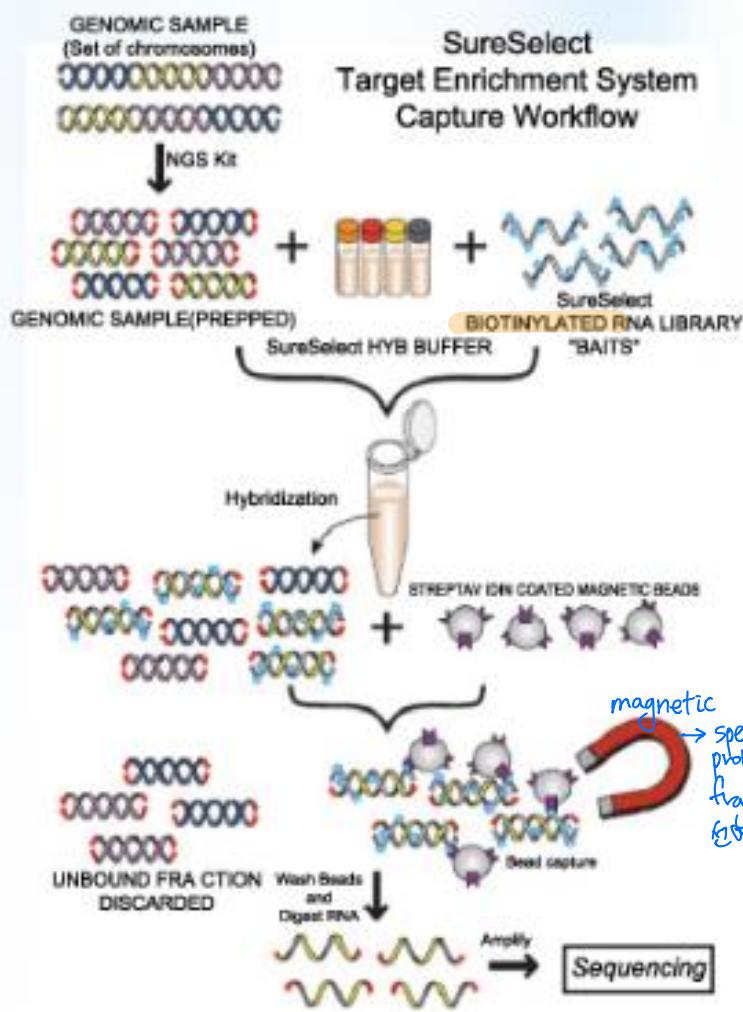
Sung Wook Chi  
Division of Life Sciences, Korea University

# Functional Genomics

## Gene expression



# Exome-Seq



$\sim 1\%$ ,  $2\%$   $\rightarrow$  exon

## Exome Capture by hybridizing with probes from exons

↓  
only hybridize to exon part  
(NGS를 통해 찾는 exon 인지 알기 어렵다)

### Advantages

- Higher coverage  
(More confident in variations)
- Less raw sequences and cost
- Elimination of background noise  
↳ contaminated DNA가 많을 때 필요할 것인 select 가능

### Disadvantages

- Incomplete exome  
(depending on annotation)
- Limited to mutations in only CDS
- Difficult to detect structural variations  
coding region의 mutation은 detect (Intron/exon)
- Intrinsic variations caused by probe hybridization

↳ probe sequence  $\approx \frac{1}{2}$  (G,C content  $\approx \frac{1}{2}$ )  
 $\Rightarrow$  not evenly purified

# Ancient human genome (Phylogenetic tree)

왜 중요한지 설명  
하는 슬라이드

## ARTICLE

99%는 from micro organism  
1%는 our ancestor

→ 지금은 from seq. sample X

10.1038/nature12886

highthroughput seq. technology 이용해  
유전체 seq

## The complete genome sequence of a Neanderthal from the Altai Mountains

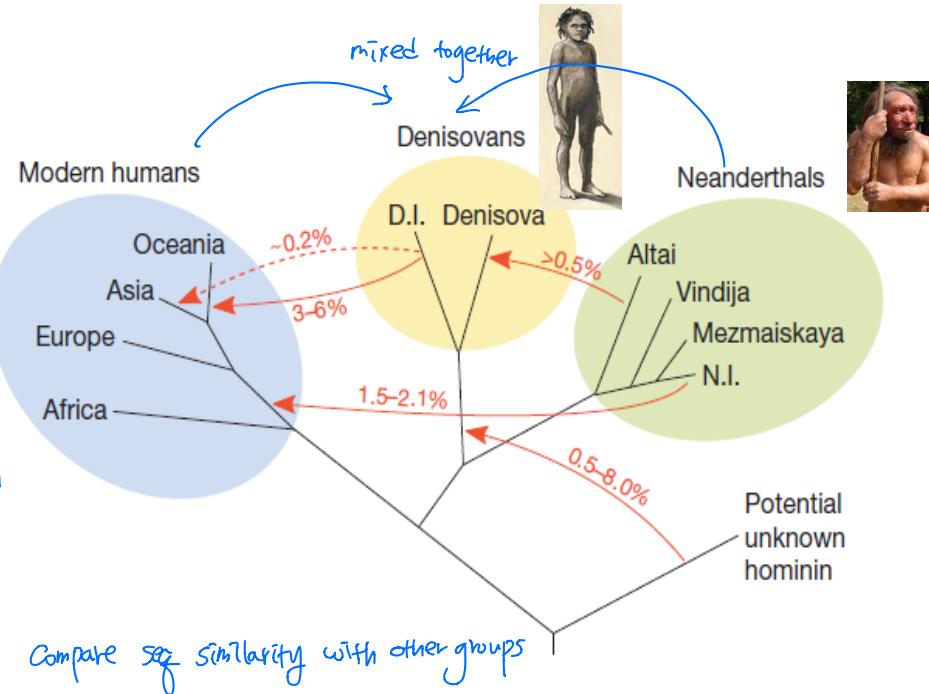
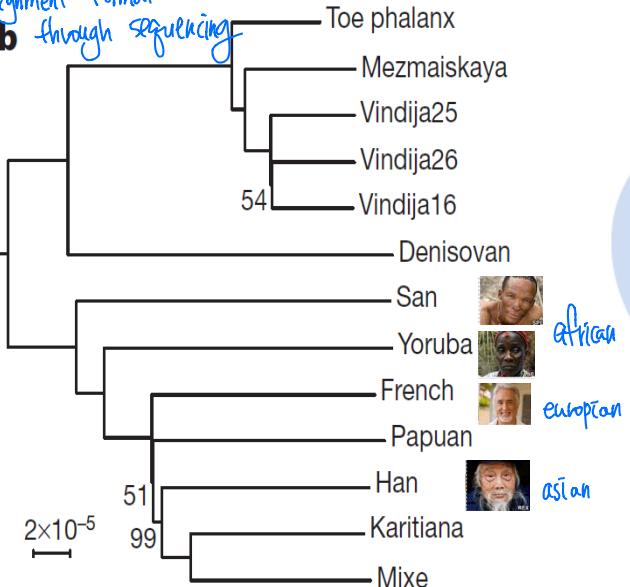
2014, Nature



A neighbour-joining tree (Fig. 2b) based on transversions, that is, purine–pyrimidine differences, among 7 present-day humans

alignment format

b through sequencing



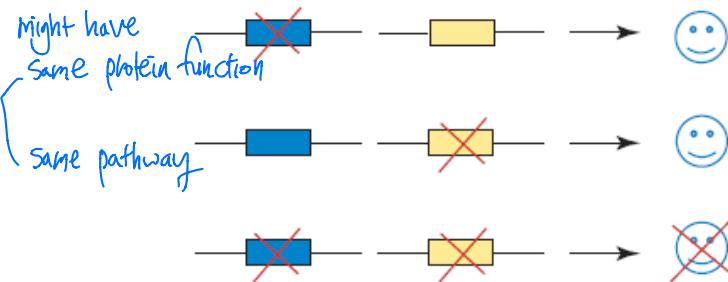
# Genetic interaction

- **Forward genetics** (phenotype → gene) : random mutagenesis
- **Reverse genetics** (gene → phenotype) : targeted mutagenesis  
→ screening sequencing

K.O 등으로서는基因을  
targeted mutagenesis로  
생성하는 방법

## Genetic interaction

(e.g. synthetic sick or lethal interaction)



## Epistasis

Gene for blond hair	Gene for red hair
Blond hair	Red hair
+ Gene for baldness	
Bald	Bald

regardless of genotype  
the phenotype depends on  
gene (제어권を持つ)

Initial enzyme is  
controlled by  
gene

Type of screen
Loss-of-function
Gain-of-function
Dominant-negative
Modifier

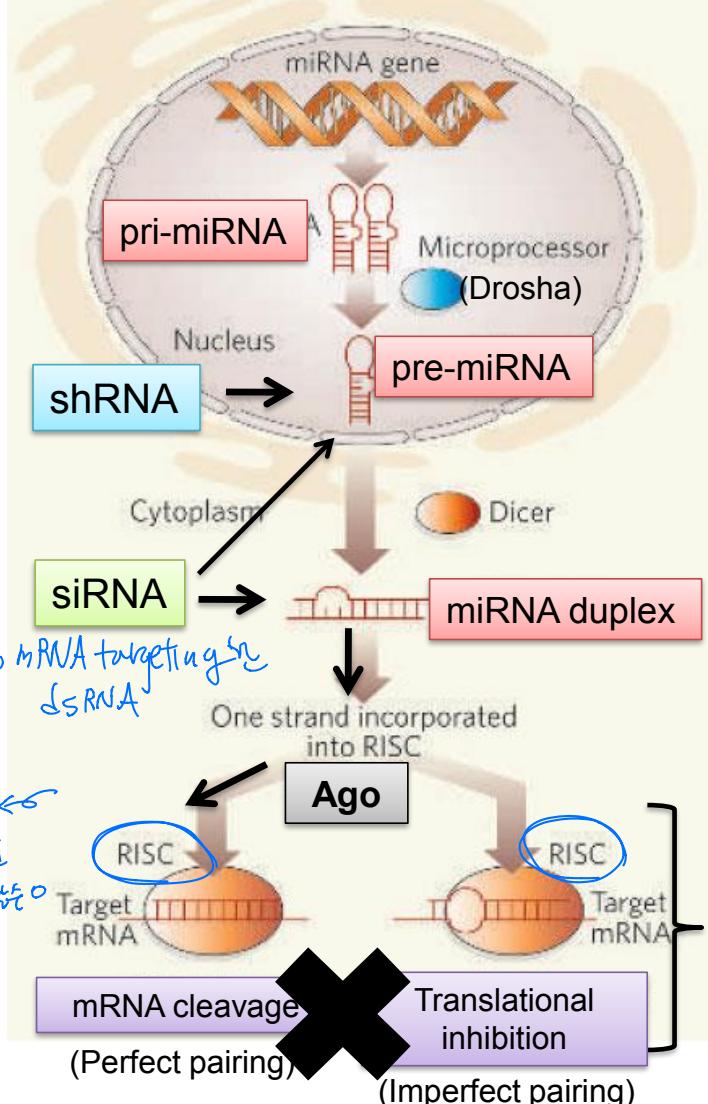
① (Loss-of-function)  
Dominant-negative → artificially induce expression  
Non-Functional



같은 gene product가 두 가지(기능과 비기능)  
한 가지 mutation 생기면 가능 X 일정

→ dominantly affect  
→ non-functional

# RNAi screening : functional genomics



RNA screening

Genome

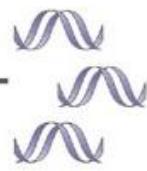
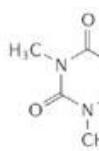
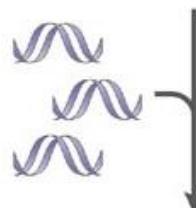
RNAi : loss of function

design library

siRNA library

targeted genes / drug on  
ratio or sensitivity test

Drug + siRNAs



Phenotype

# Selective gene inactivation and modification

- Study of gene function in cultured cells or using cell extracts has limitations
- Defining gene function in this wider context requires the genetic manipulation of model organisms

virus vector 등을 이용해  
artificially over expression  
↓

(A)

directly alter endogenous gene

## Loss of function

(B)

add exogenous gene copy

## Gain of function

gene modification

gene inactivation

at DNA level

at RNA level

at protein level

mutate gene

to change amino acid or motif, or swap domains, etc.

gene knockout

by mutating germ-line DNA to create a null allele

base pairing with specific transcripts to destroy RNA or block gene expression

protein block

binding to proteins to disrupt function

transgene expression

over-expression

coupling to a very strong promoter

ectopic expression

in unusual cells or time points

random gene knockout

by random mutagenesis, e.g., using chemical mutagens

targeted gene knockout

inactivation of a specific predetermined gene

RNA interference

see Figures 12.3 and 12.4

morpholino antisense oligos

designed to block translation or to induce alternative splicing

dominant negatives

mutant proteins that bind to normal proteins

aptamers

synthetic nucleic acids (or peptides) selected to bind to a protein

intrabodies

antibodies designed to work *within* cells

# Functional genomics : Genetic screening & ChIP-Seq

Sung Wook Chi

Division of Life Sciences, Korea University

# Saturation Forward Genetics

## Forward Genetic Screens



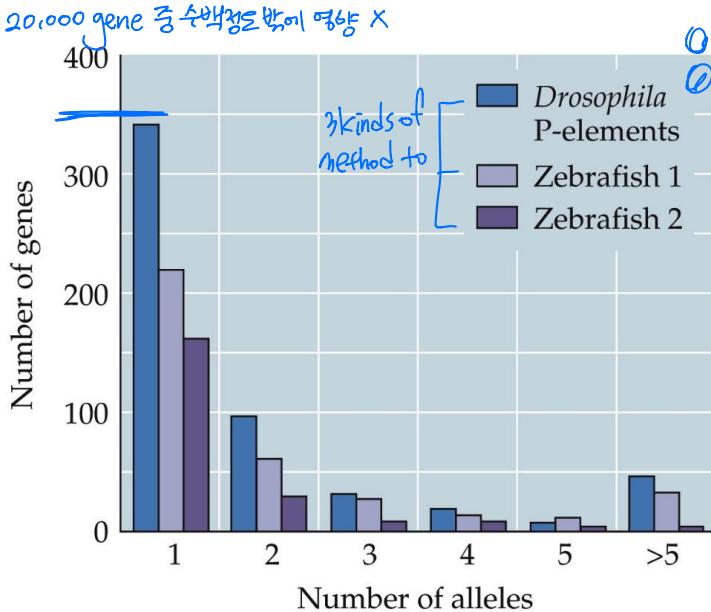
Discover  
Gene  
underlying  
Phenotype

Gene?  
what genes are  
acquired by this mutation?

Phenotype -> Gene

Screening of comprehensive 히지 암을  
Uneven 한정 Saturation 팀

① ASMO  
② NEMO

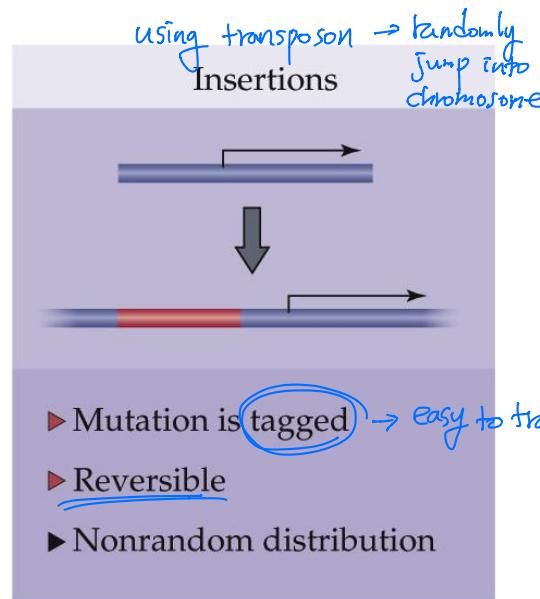
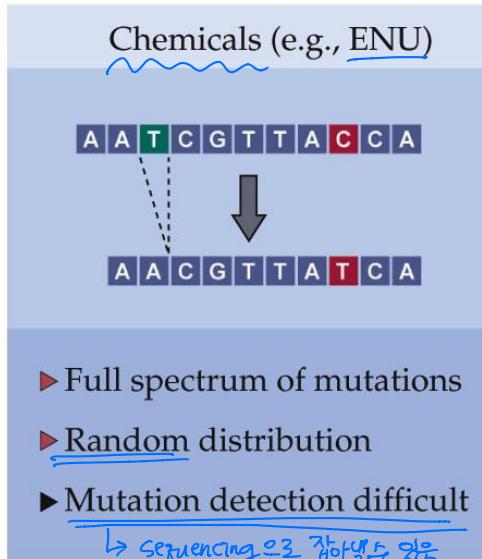
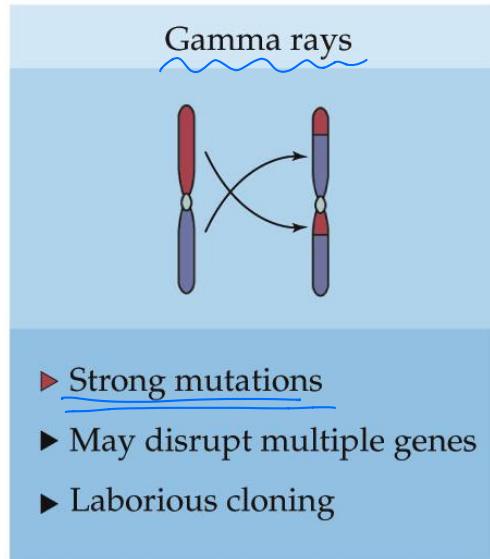


genetic screens of such large scale that a point is reached where most new mutations represent second or multiple hits of previously identified loci.

How we can

→ limitation

# The three major types of mutagen (*induce mutation*)



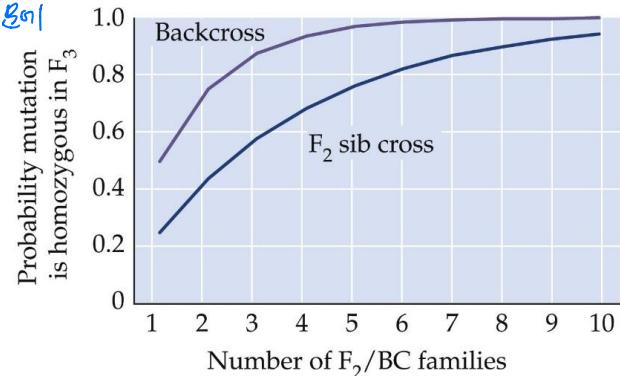
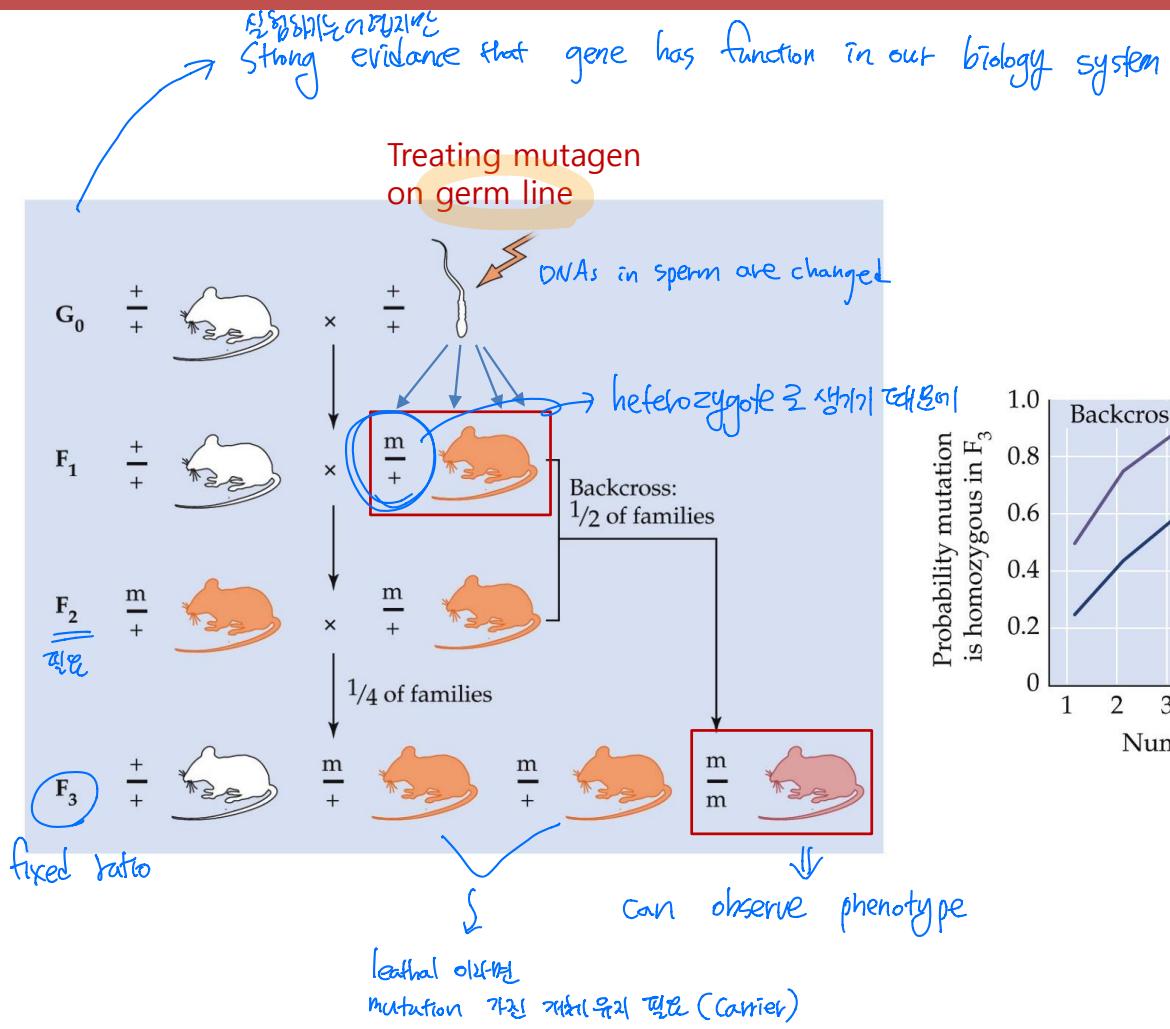
단점) Double - strand  
breakage of DNA  
 $\rightarrow$  cell or organism  
 $\rightarrow$  repair  $\rightarrow$  mutation  
recombination 가능성이  
 $\rightarrow$  gamma-ray  $\rightarrow$  random  
recombination 가능

ENU, also known as N-ethyl-N-nitrosourea  
- Alkylating agent  $\rightarrow$  base alkylating agent  
: transferring the ethyl group of ENU to nucleobases (usually thymine)

efficient way to give tiny mutation

P element in drosophila  
(transposon)

# F1 and F3 genetics screens

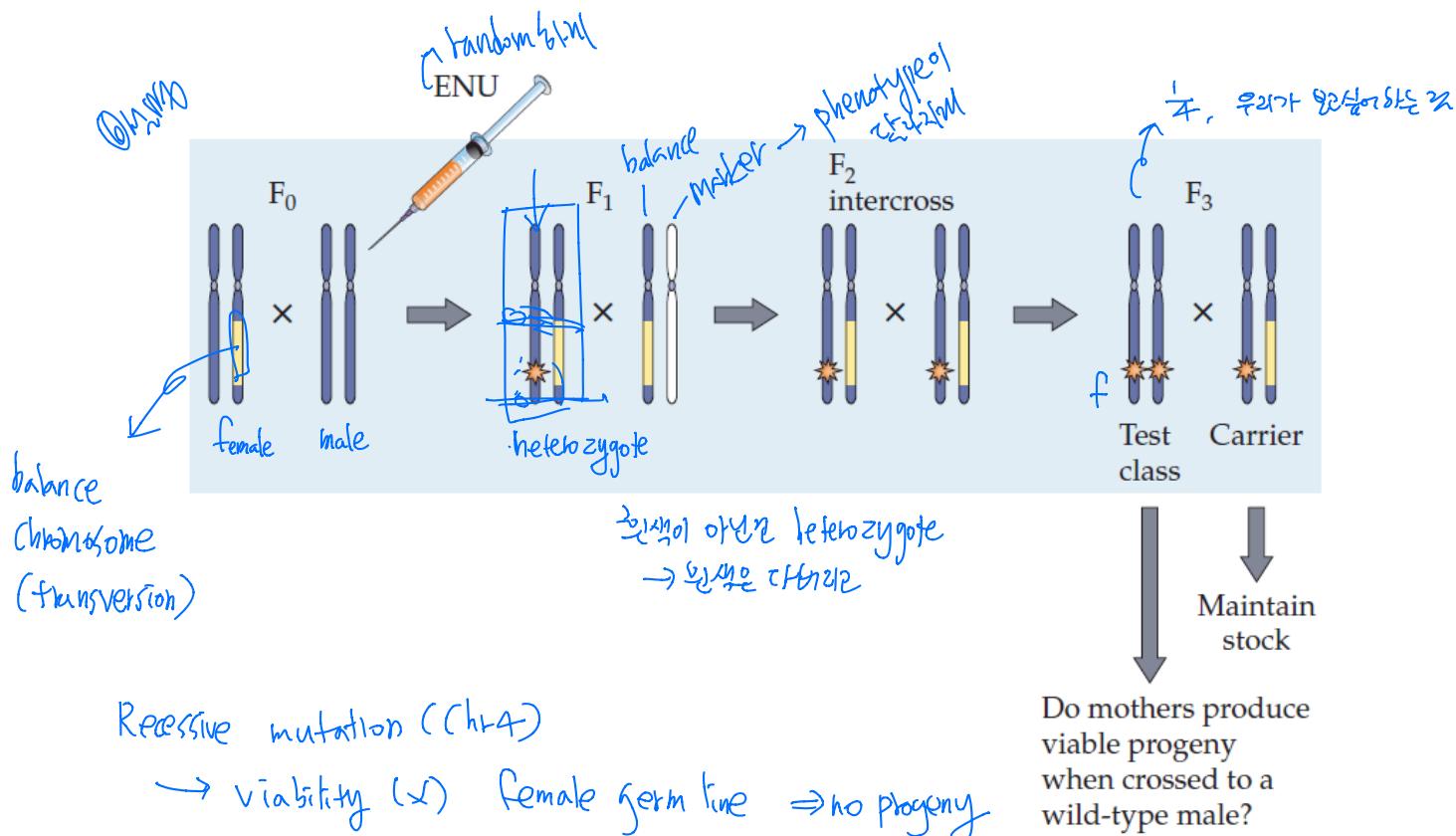


## **EXERCISE 5.4** *Designing a genetic screen*

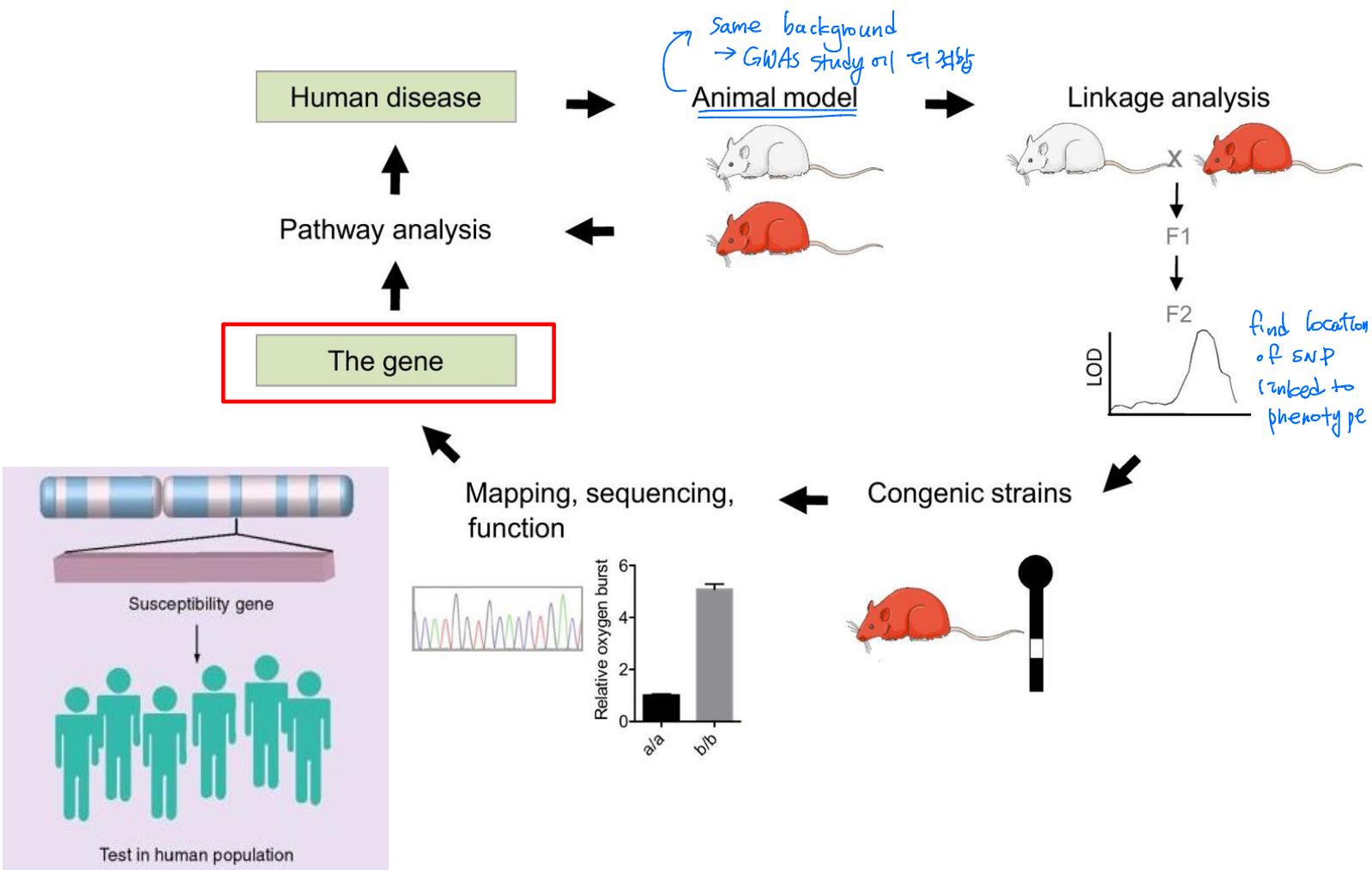
Create a genetic screen designed to detect recessive mutations on chromosome 4 that are required in the female germline for viability of mice.

Assume that you have available a strain of mice that carry a dominantly marked chromosome 4 with a large inversion (namely, a balancer chromosome).

\* leafhail o1 female  
germ line o1ng  
Can not child birth

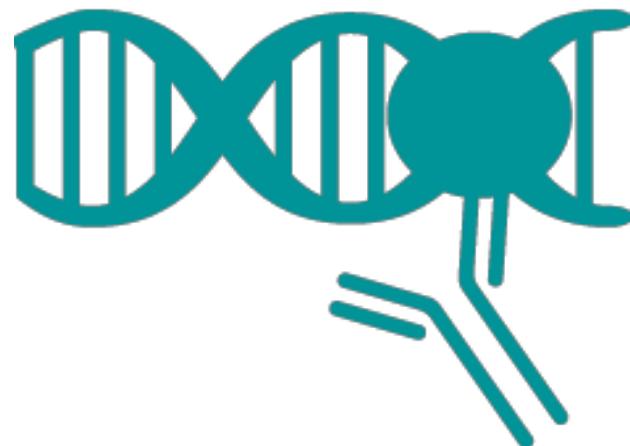


# Functional genomics studies : Forward genetics & GWAS



# ChIP-Seq

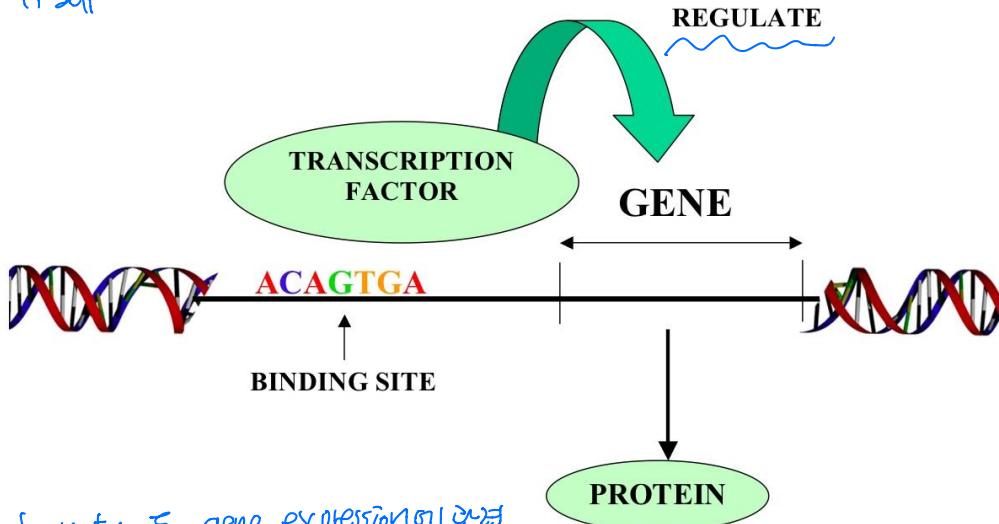
(Chromatin immunoprecipitation Sequencing)



only see region bounded by transcription factors  $\Rightarrow$  regulation 관계

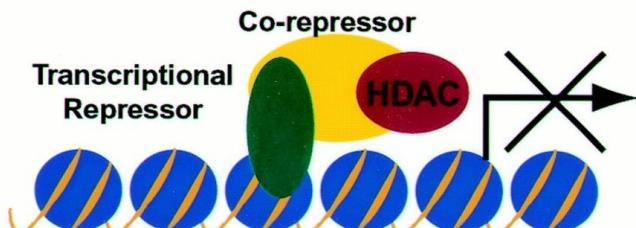
# Transcription Factor & Chromatin modification

① sequence it self



② Structure of chromatin  $\Rightarrow$  gene expression

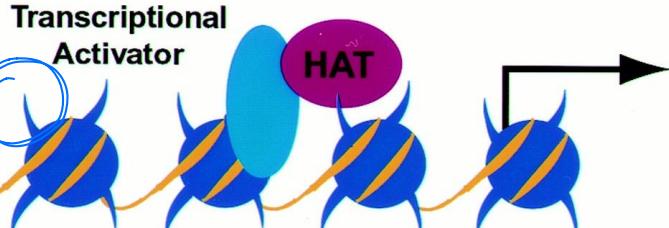
**Repression** condensed



Hypoacetylated, condensed chromatin

Modification of histone is key of structure

**Activation** decondensed



Hyperacetylated, accessible chromatin

open structure  $\rightarrow$  more accessible

# Histones & Chromatin formation

lysine 9

H3K9me3, H3K27me3 : repression

histone 3 tri methylated

H3K4me3: promoters during transcription

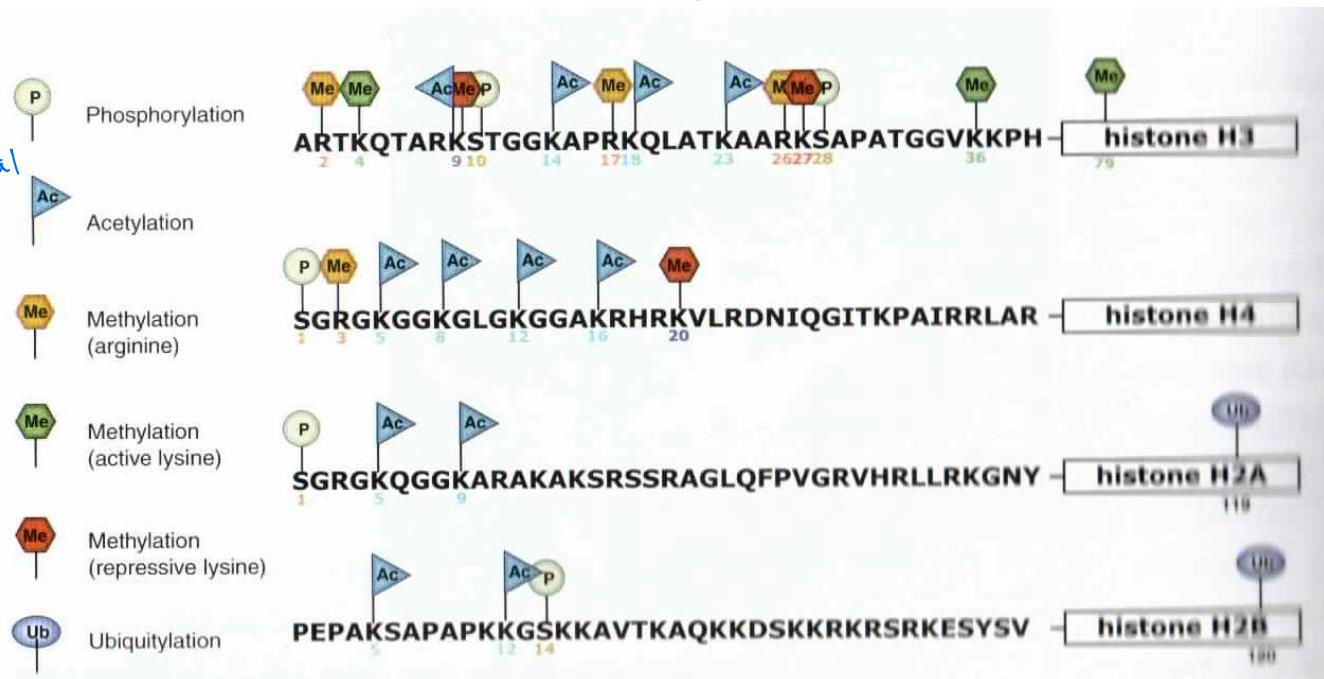
H3K9ac : transcriptionally active

H3K36me3 : entire transcribed regions

Histone 단백질의  
N terminal은 바깥을 향함  
⇒ Can gain modification

histone: H2A, H2B, H3, H4  
octamer

post translational  
modification  
that can  
be attached  
to N-terminal  
peptide

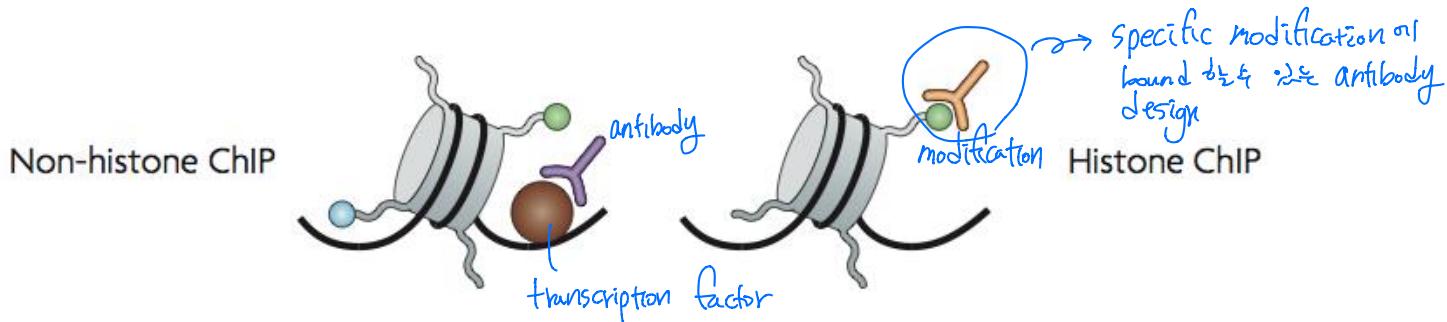


"Histone code" : Epigenetics

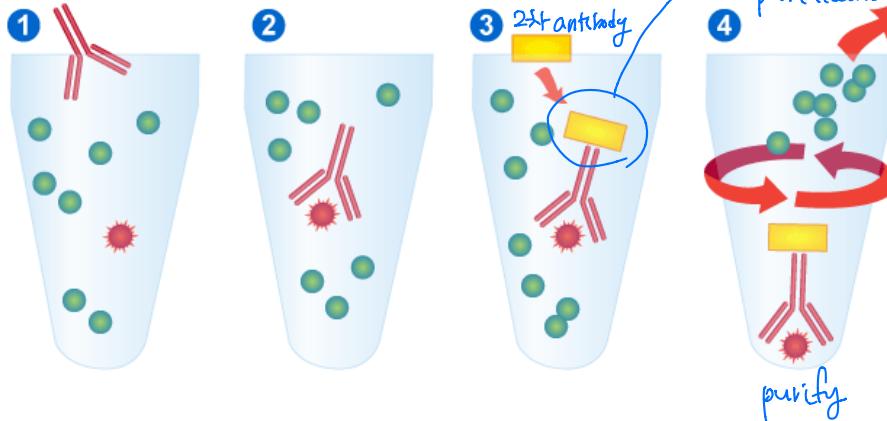
# Chromatin immunoprecipitation (ChIP)

to identify

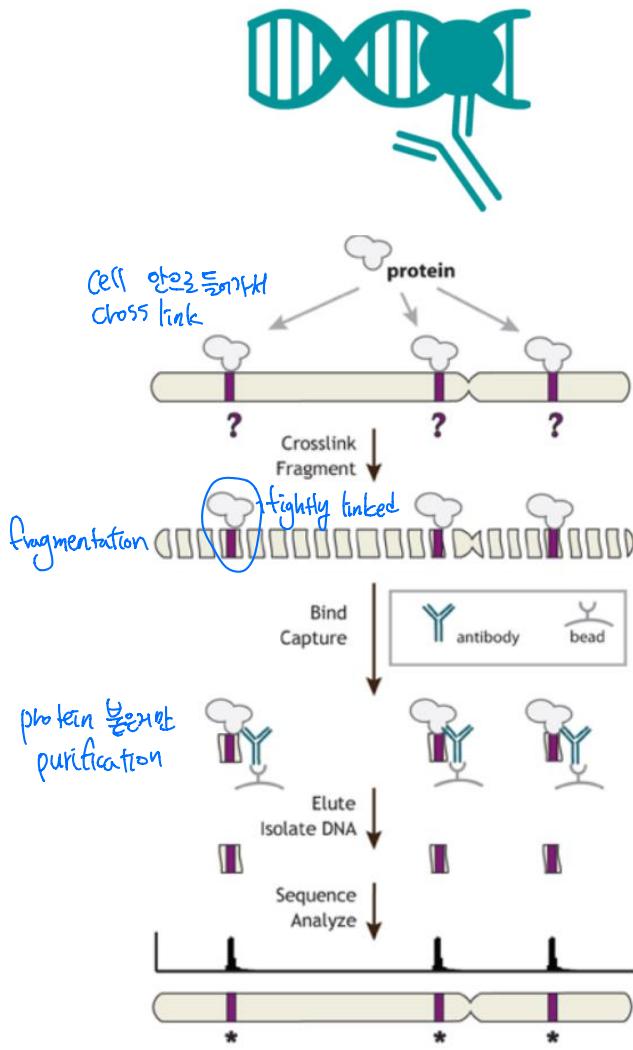
- Protein-DNA interactions
- Chromatin States
- Transcriptional regulation



\* Immunoprecipitation

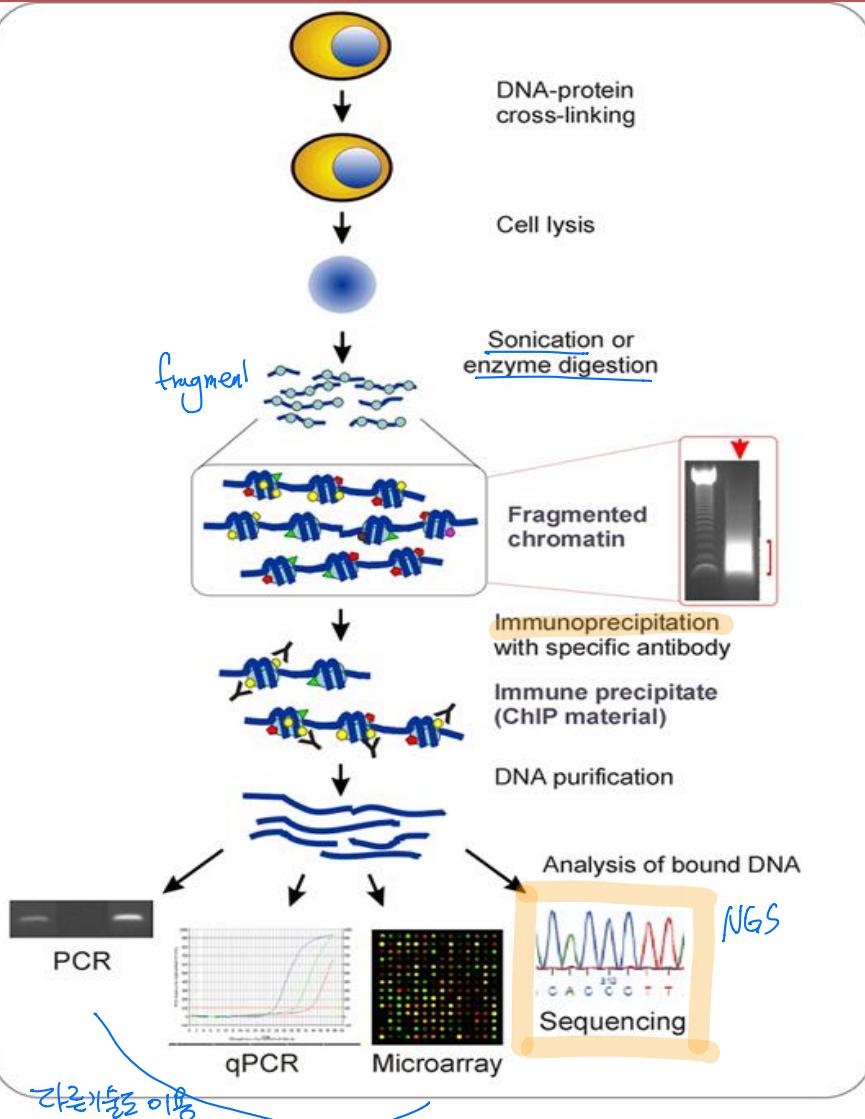


### III-3 Chromatin immunoprecipitation sequencing (ChIP-Seq)



- ★ **DNA-protein** interactions
  - Regulatory element - **Transcription factors**
  - Genomic regions – **Modified Histones**
- 1. **Crosslink** the cells with *make covalent bond* formaldehyde to “fix” factors in place.  
Exception: Native ChIP with histone antibodies.
- 2. **Shear chromatin** to smaller pieces.  
Shear size determines resolution.  
Note: ChIP-exo uses an exonuclease at a later step to increase resolution.
- 3. **Enrich** target using an antibody.  
Enrichment is only as good as the antibody.

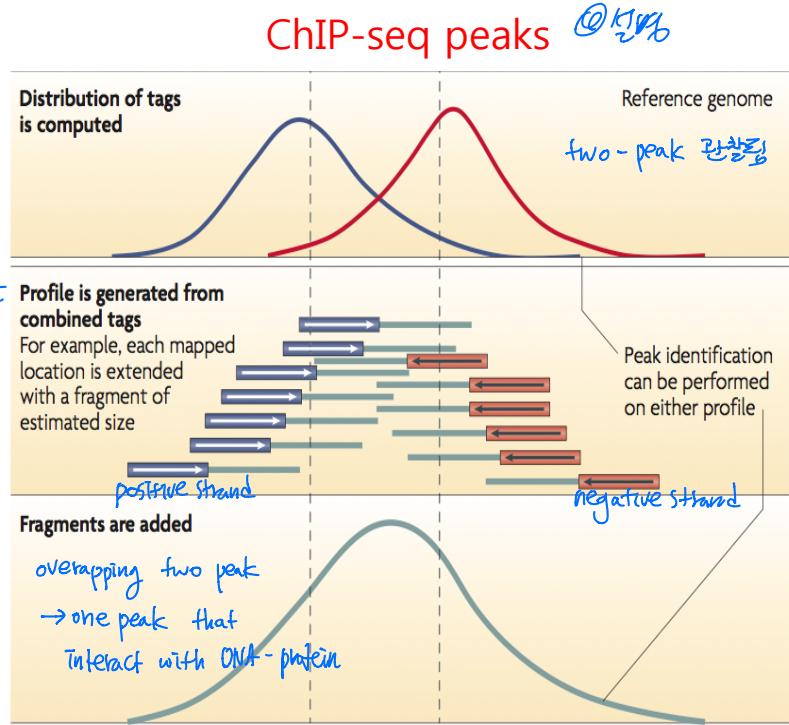
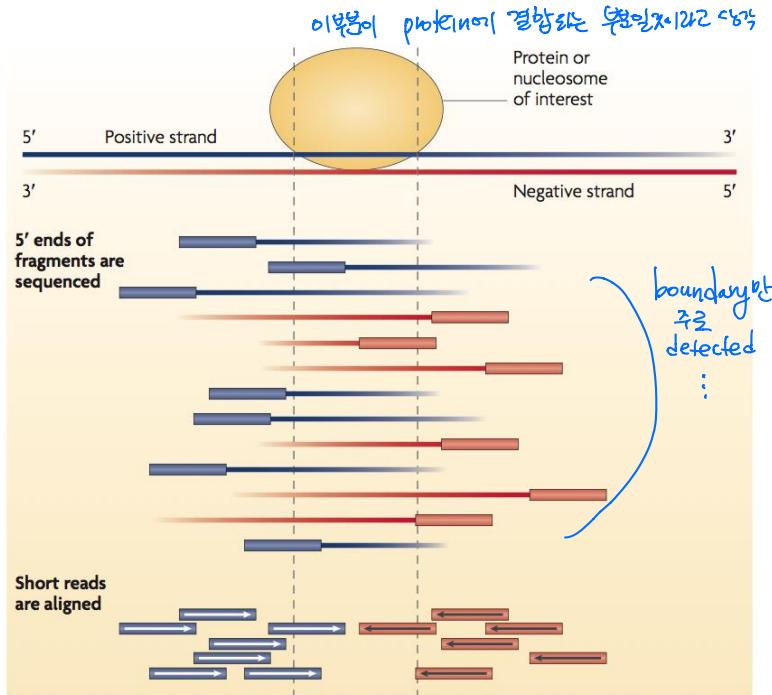
# Chromatin immunoprecipitation (ChIP)



## In Nutshell

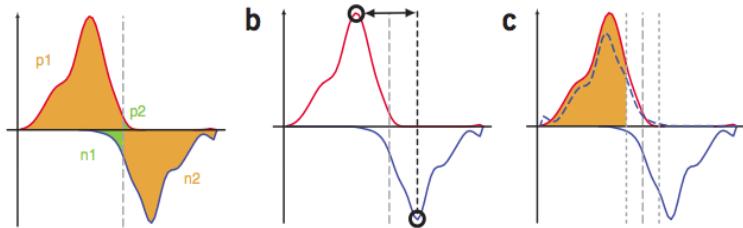
- Protein cross-linked to DNA *in vivo* by treating cells with formaldehyde
- Shear chromatin (sonication)
- IP with specific antibody
- Reverse cross-links, purify DNA
- PCR amplification\*
- Identify sequences
- Genome-wide association map

# ChIP-Seq data analysis



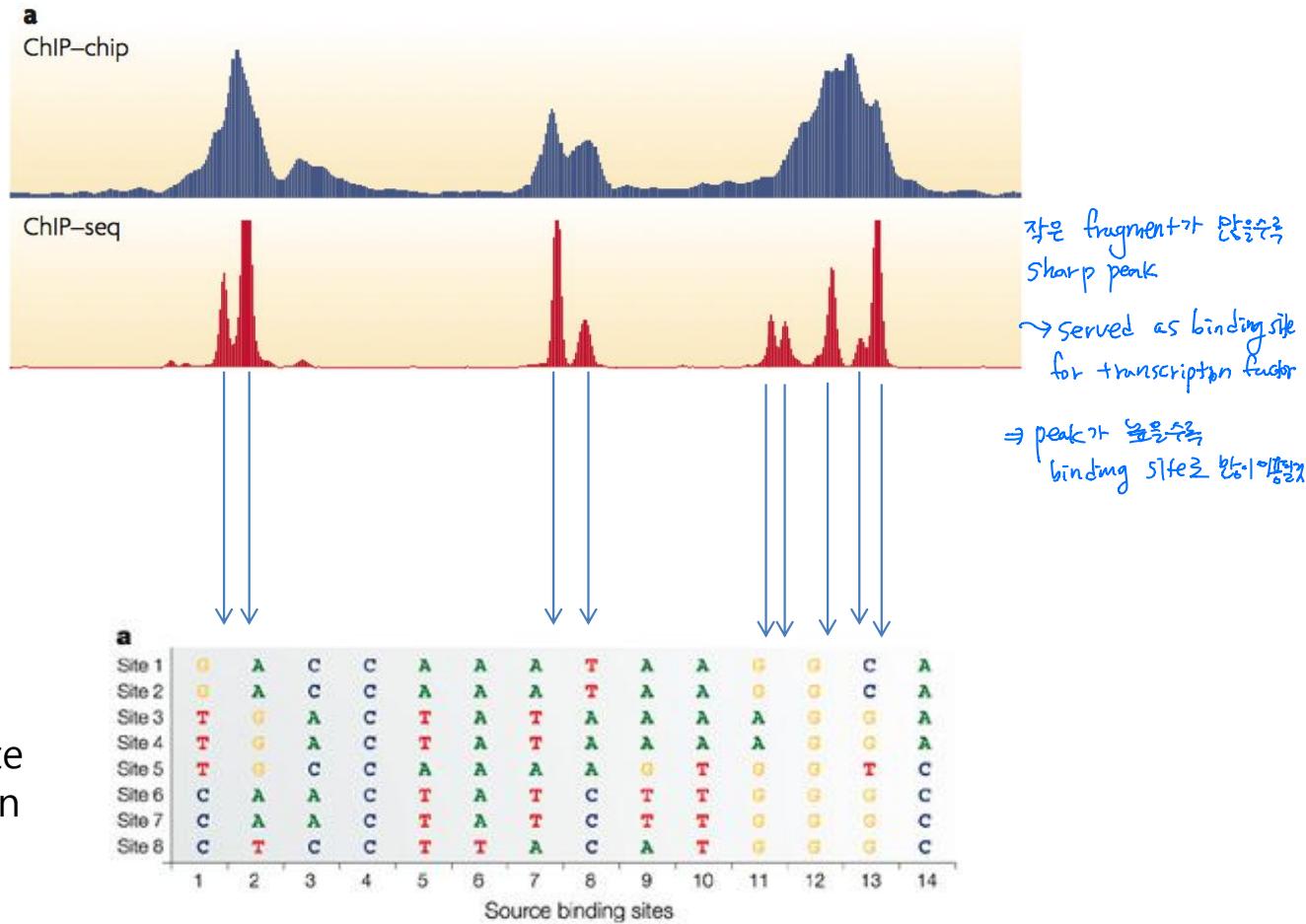
- Only 5' end of fragments are sequenced
- Tags from both + and - strand aligned to reference genome

## Binding Site Detection



# ChIP-Seq data analysis

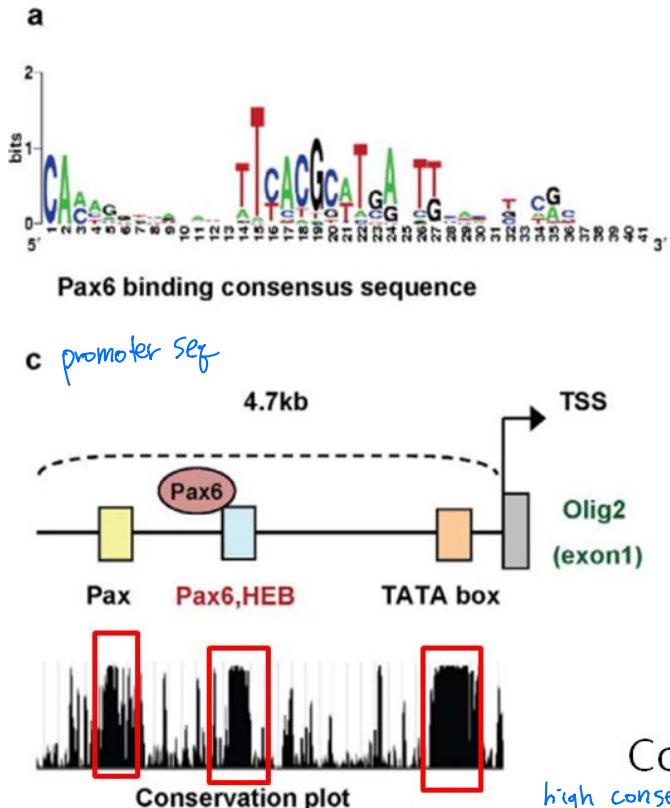
Transcription factor A



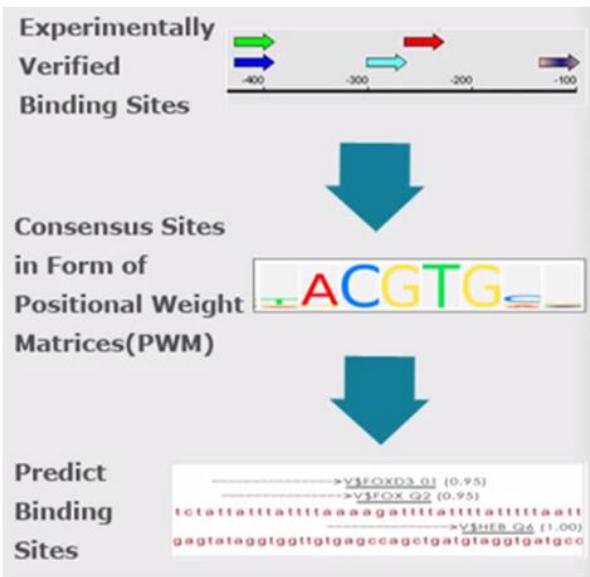
# Analysis of transcription factor binding sites

Sequence Conservation  
→ Evolution Footprint

Multiple sequence alignment → Pattern, Profile



## TRANSFAC Database



d profileol global bioinformatics 2014 binding site

Site name	Sequence	Core match	Matrix match
Pax	CAGGAa	0.851	0.868
Pax6	tTAACGg	0.93	0.811
Heb	CAGCTG	1	1
TATA box	TATAAAa	1	0.947

## Conservation Score (PhastCons)

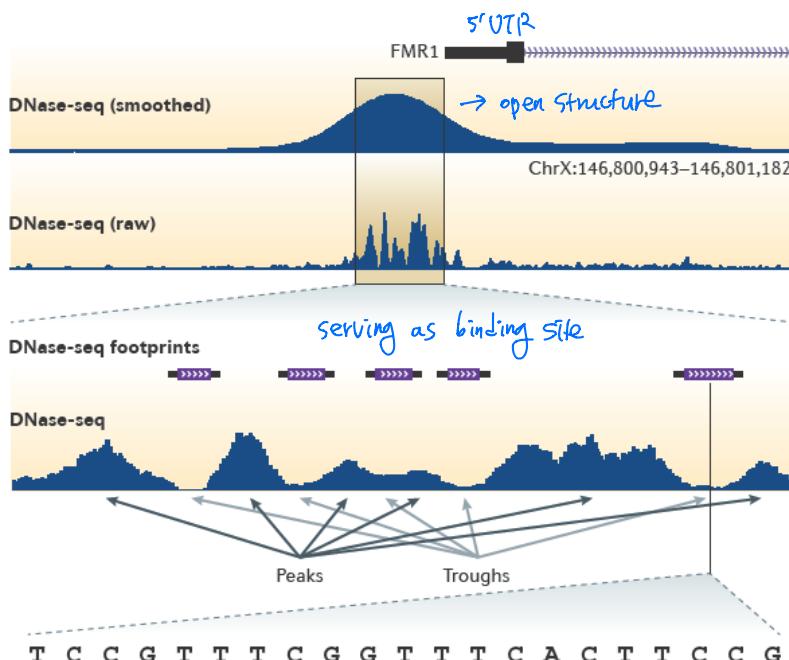
high conservation sequence through species  
→ 2014

# Fine scale information from DNase-seq



Sequencing the ends of DNase cuts identifies regions of bare DNA.

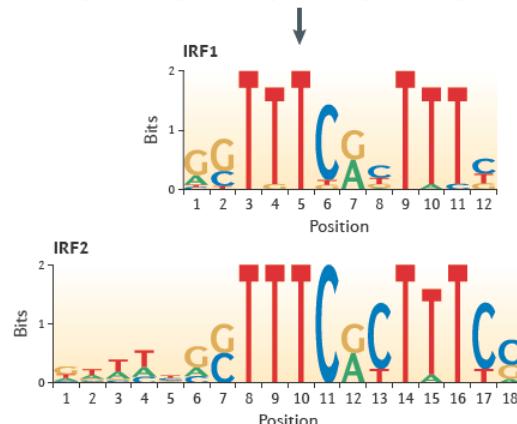
Fine scale analysis of this data can identify individual TF binding sites.



Motifs from JASPAR database

Model name	Score	Relative score	Start	End	Strand	Predicted site sequence
IRF1	12.986	0.904279917181229	3	14	-1	GAAACGAAACG
IRF2	17.216	0.907706906384892	4	21	-1	CGGAAGTCAAACGAAAC
SPIB	4.820	0.806987596140569	5	11	-1	ACCGAAA
BRCA1	4.228	0.802287513481405	8	14	-1	GAAACCG

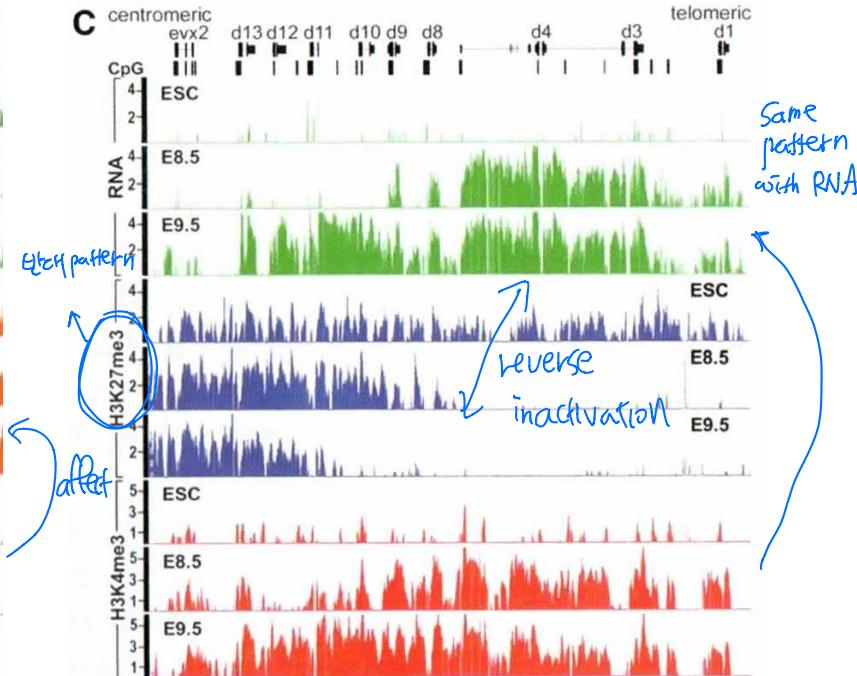
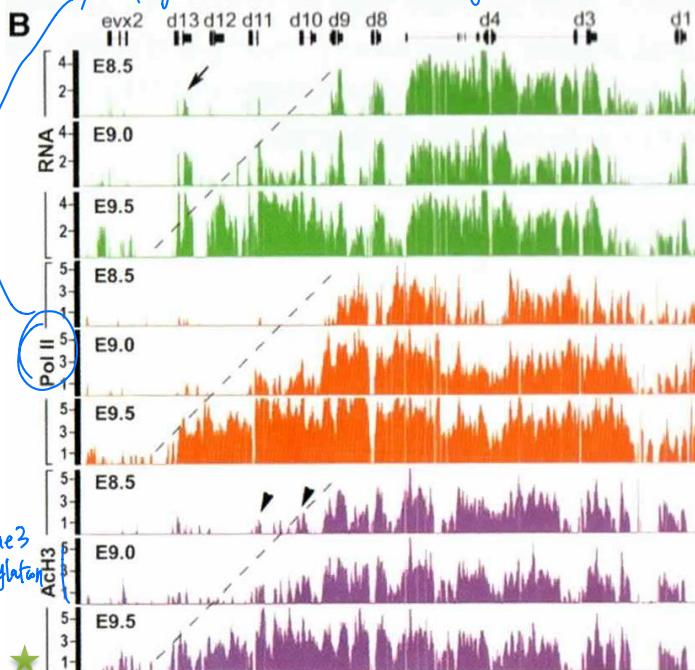
Sequence를 일관되게 binding site인지 알수도



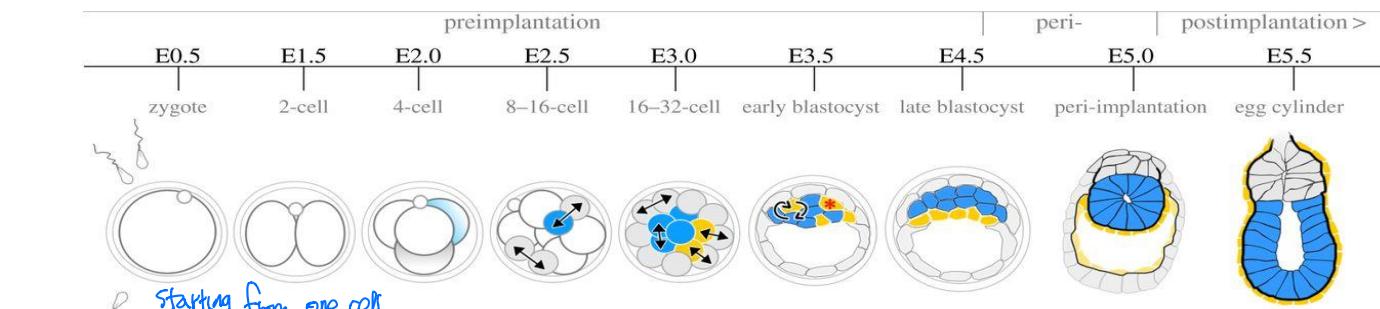
# Dynamic landscape of histone modification during development (elucidated by ChIP-Seq)

(M) M6

## Association of acetylation and methylation of histone H3 with transcription



histone 3 acetylation

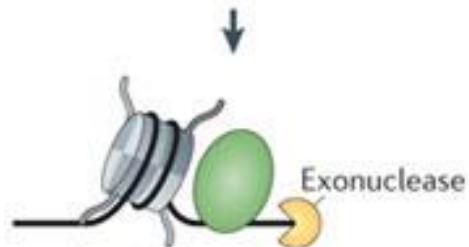


# ChIP-Seq: Chromatin Immunoprecipitation

## a DNA-binding protein ChIP-seq

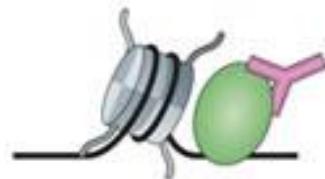


Crosslink proteins and DNA



Exonuclease

- Sample fragmentation
- Sonication
- Endonuclease (ChIP-exo)

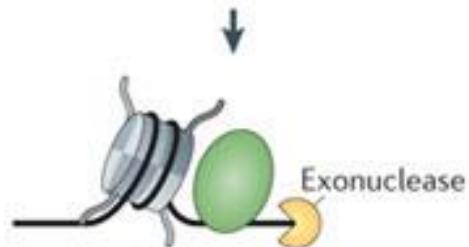


Immunoprecipitate and then purify DNA

## b Histone modification ChIP-seq



Crosslink proteins and DNA



Sample fragmentation

- MNase digestion



Immunoprecipitate and then purify DNA

## DNA-protein interactions

- Regulatory element - Transcription factors
- Genomic regions – Modified Histones

## DNA fragmentation



## Immunoprecipitation



## Sequencing



# Functional Genomics Class : part I

~ 10 questions will be in the middle-term exam !!!

1. Introduction to genomics
2. Mapping genome
3. Human genome project and sequencing
4. Next-generation sequencing
5. Sequence alignment, haplotype (LD analysis)
6. Genomic variation
7. Whole genome sequencing (NGS)
8. Genome-wide association study (GWAS)
9. Exome-seq and functional genomics
10. Phylogenetics, ChIP-Seq