# BMEG3102 Bioinformatics

Huating Wang, Ph.D., CUHK
huating.wang@cuhk.edu.hk

# Genetics and Epigenetics

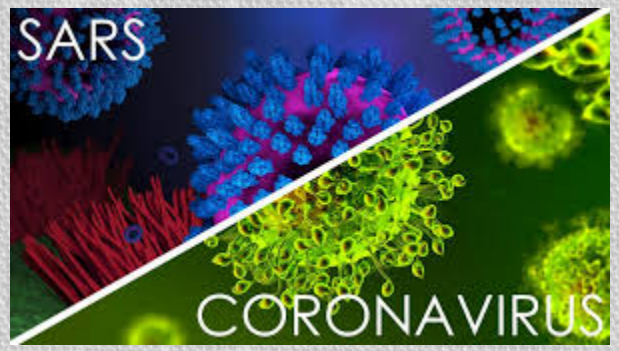
- Genetics and epigenetics
- High throughput sequencing (HTS), genomics and epigenomics
- HTS applications in clinic

# Part I: Genetics and Inheritance

# What is genetics?

- **Genetics** is the study of genes, and studies what genes are and how they work.
  - Study of human genetics can be useful as it can answer questions about human nature,
  - Understand genetics of human life,
  - Understand the diseases and development of effective disease treatment.


Angelina's
DOUBLE MASTECTOMY
Inside
HER BRAVE CHOICE


SARS
CORONAVIRUS

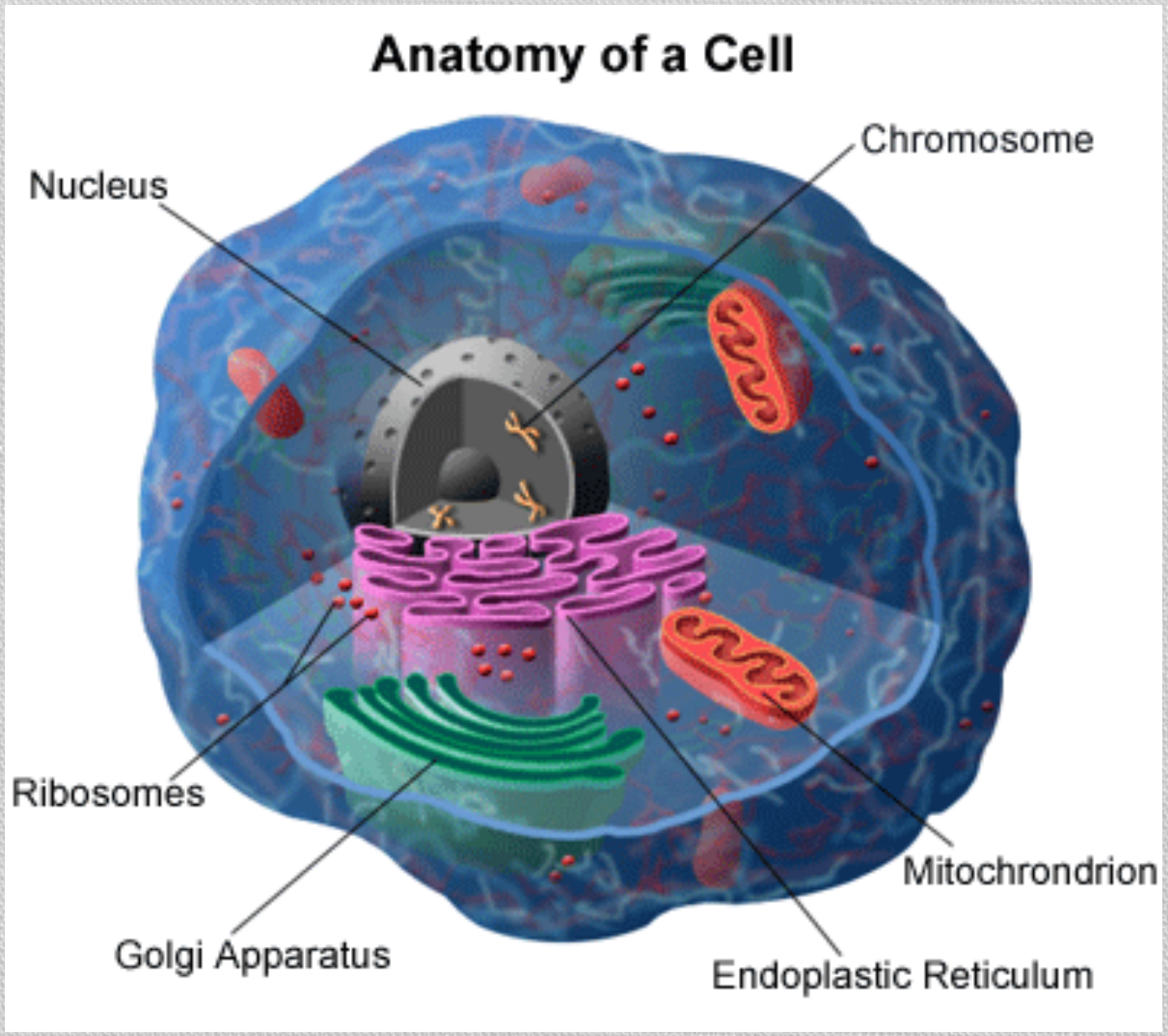
# Traits

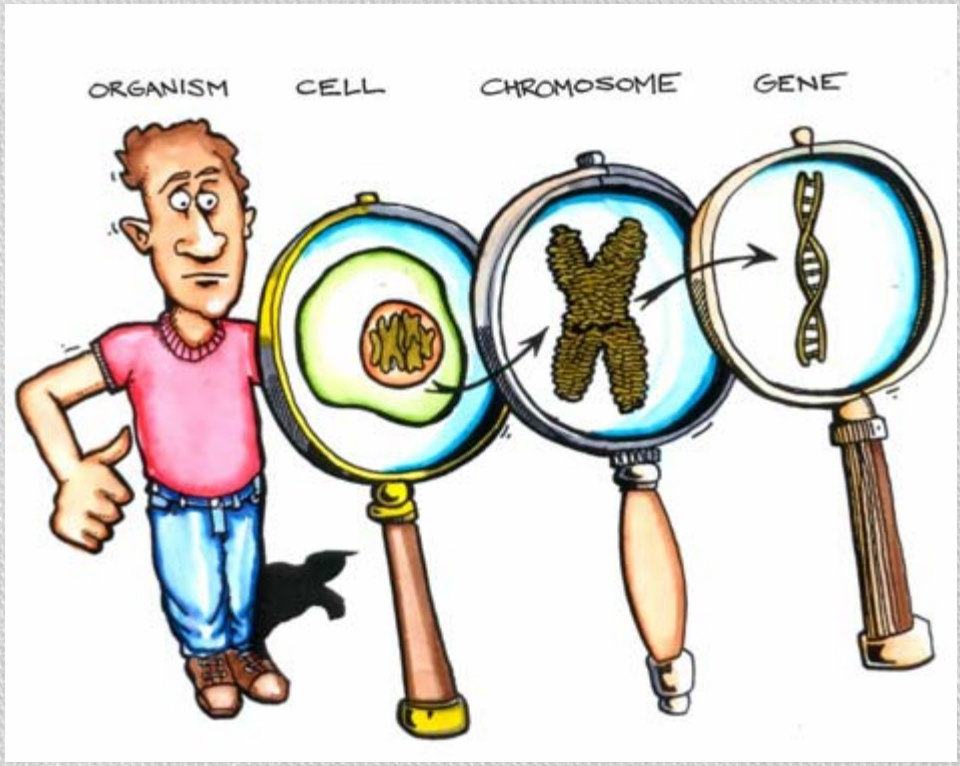
- In genetics, a feature of a living thing is called a “trait".
- Some traits are part of an organism's physical appearance; such as a person's eye-color, height or weight. Other sorts of traits are not easily seen and include blood types or resistance to diseases.
- The way our genes and environment interact to produce a trait can be complicated.

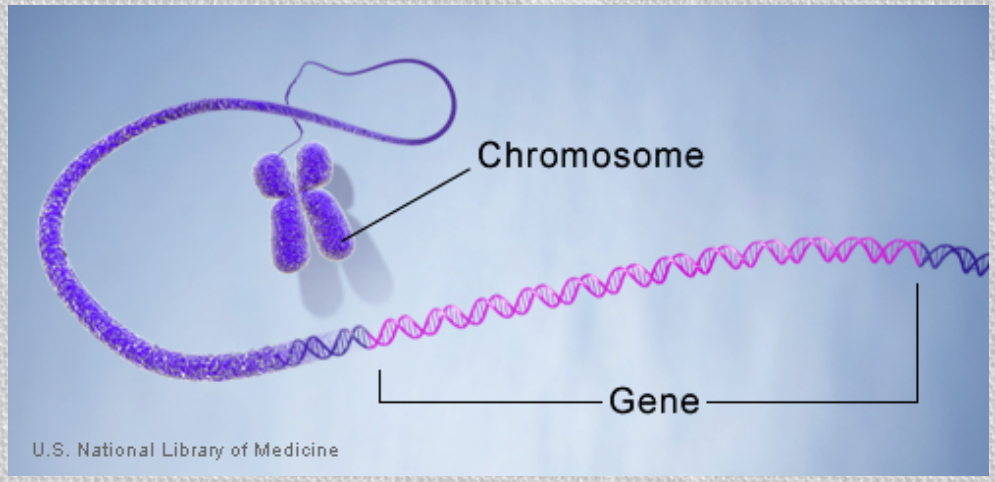
# Genes and Inheritance

- Genes are how living organisms inherit features from their ancestors; for example, children usually look like their parents because they have inherited their parents' genes.

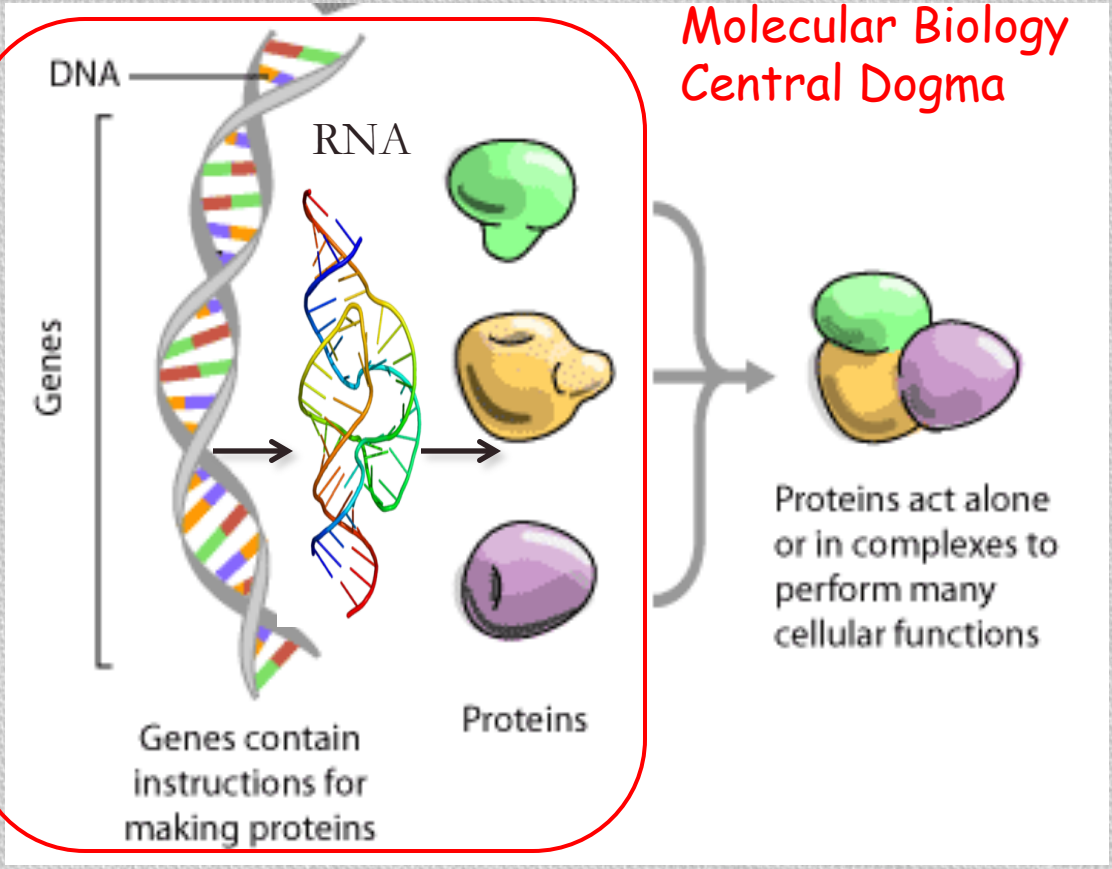
- Genetics tries to identify which features are inherited, and explain how these features pass from generation to generation.


Anatomy of a Cell
Chromosome
Nucleus
Ribosomes
Mitochrondrion
Golgi Apparatus
Endoplastic Reticulum


ORGANISM
CELL
CHROMOSOME
GENE


Chromosome
Gene
U.S. National Library of Medicine

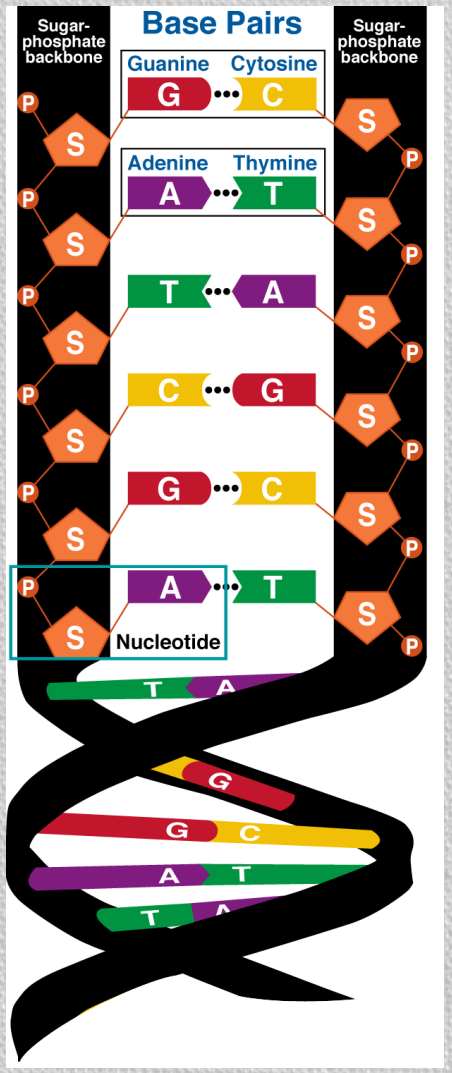
# The function of genes is to provide the information needed to make molecules called protein in cells



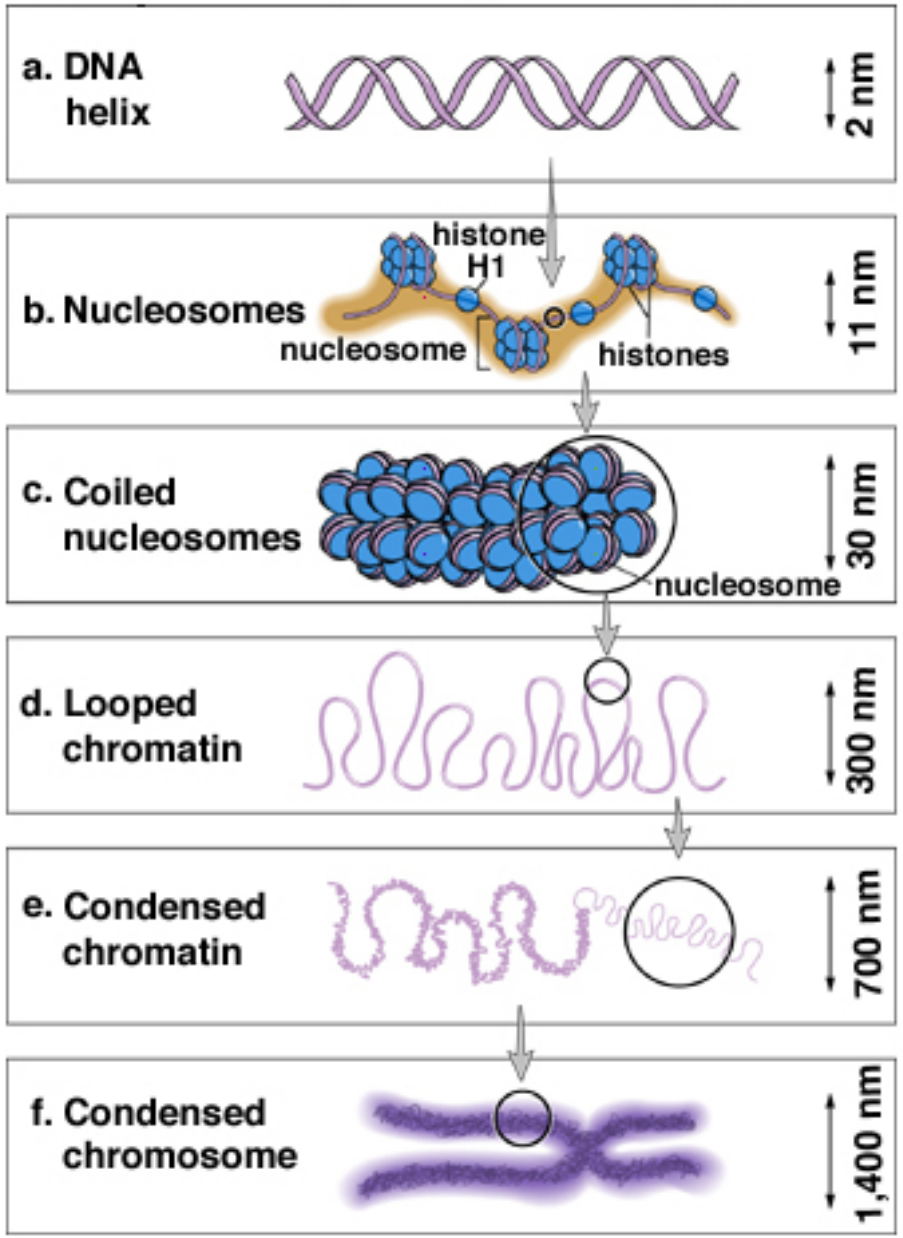
# The Genetic Code

| | | | |
|---|---|---|---|
| UUU, UUC — phenyl alanine<br>UUA, UUG — leucine | UCU, UCC, UCA, UCG — serine | UAU, UAC — tyrosine<br>UAA, UAG — stop | UGU, UGC — cysteine<br>UGA — stop<br>UGG — tryptophan |
| CUU, CUC, CUA, CUG — leucine | CCU, CCC, CCA, CCG — proline | CAU, CAC — histidine<br>CAA, CAG — glutamine | CGU, CGC, CGA, CGG — arginine |
| AUU, AUC, AUA — isoleucine<br>AUG — methionine | ACU, ACC, ACA, ACG — threonine | AAU, AAC — asparagine<br>AAA, AAG — lysine | AGU, AGC — serine<br>AGA, AGG — arginine |
| GUU, GUC, GUA, GUG — valine | GCU, GCC, GCA, GCG — alanine | GAU, GAC — aspartic acid<br>GAA, GAG — glutamic acid | GGU, GGC, GGA, GGG — glycine |

The structure of the **double helix** is somewhat like a ladder, with the base pairs forming the ladder's rungs and the sugar and phosphate molecules forming the vertical sidepieces of the ladder.



# Levels of chromosome structure



# The organization of genes of a human chromosome

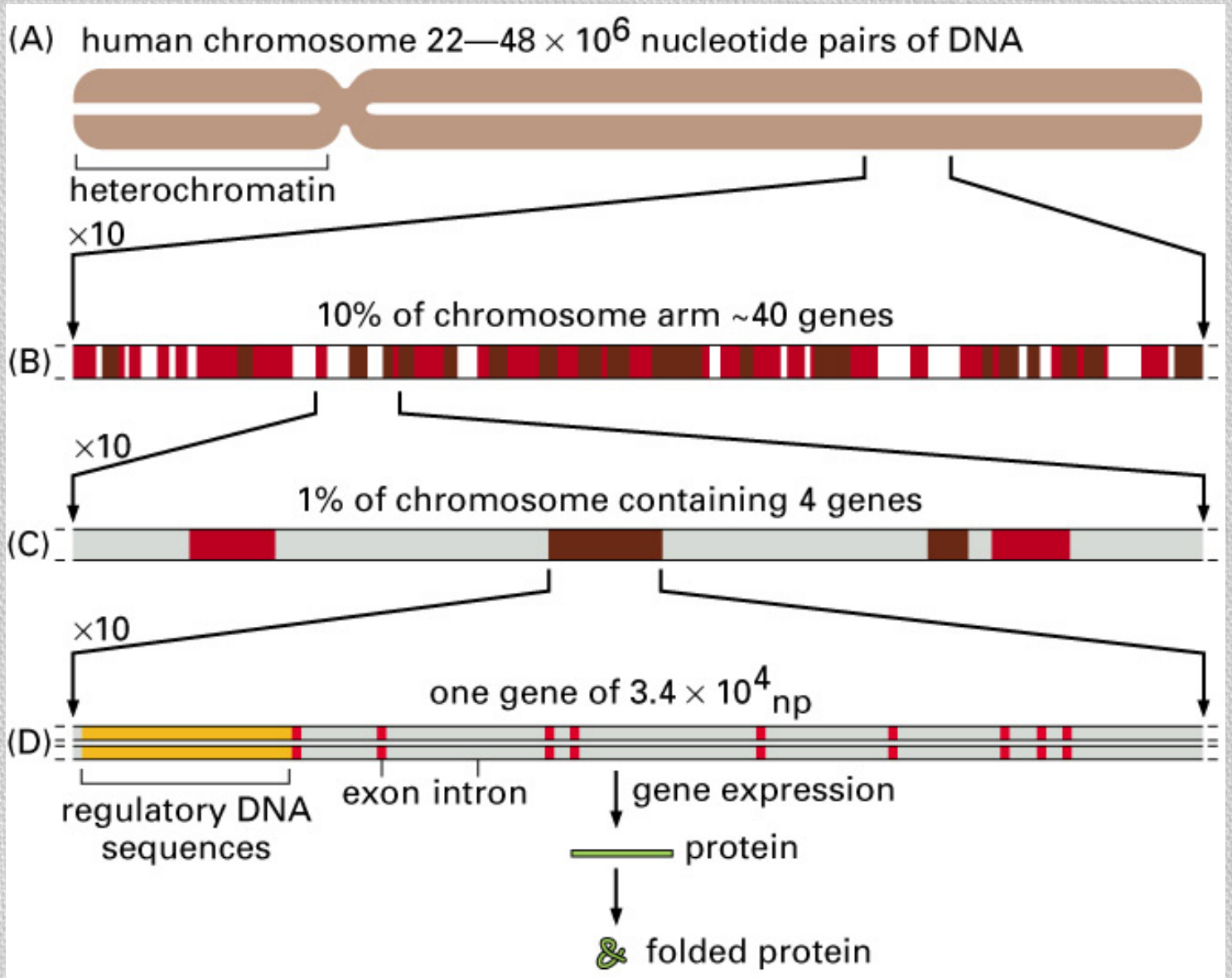
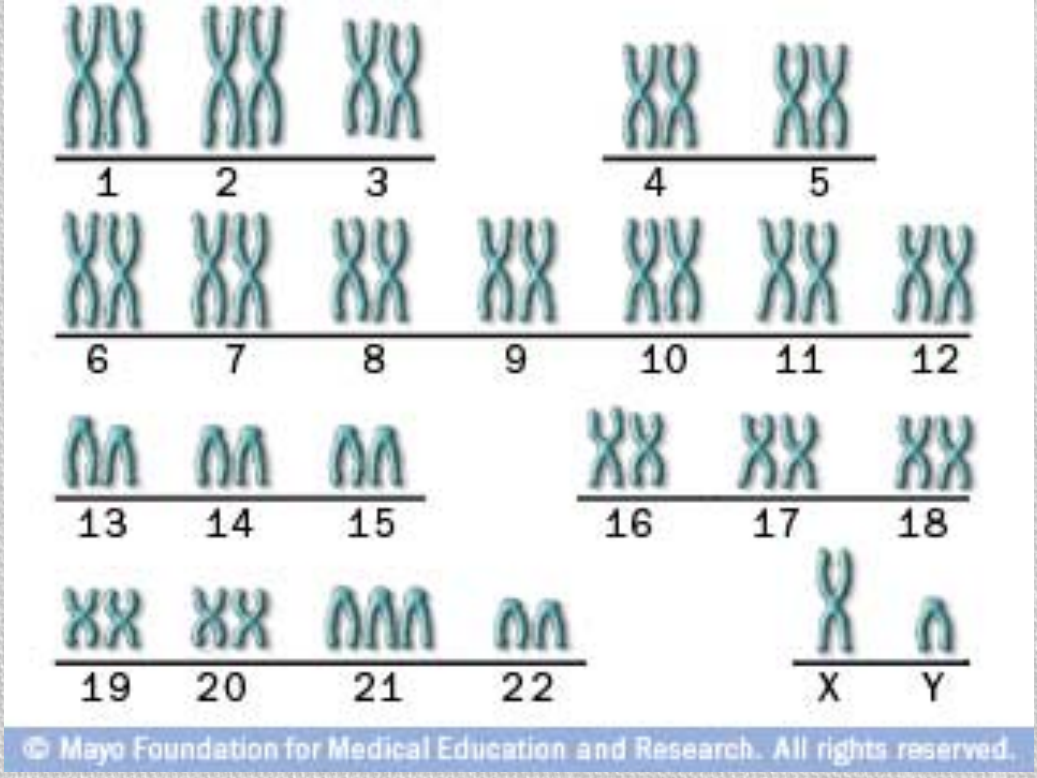


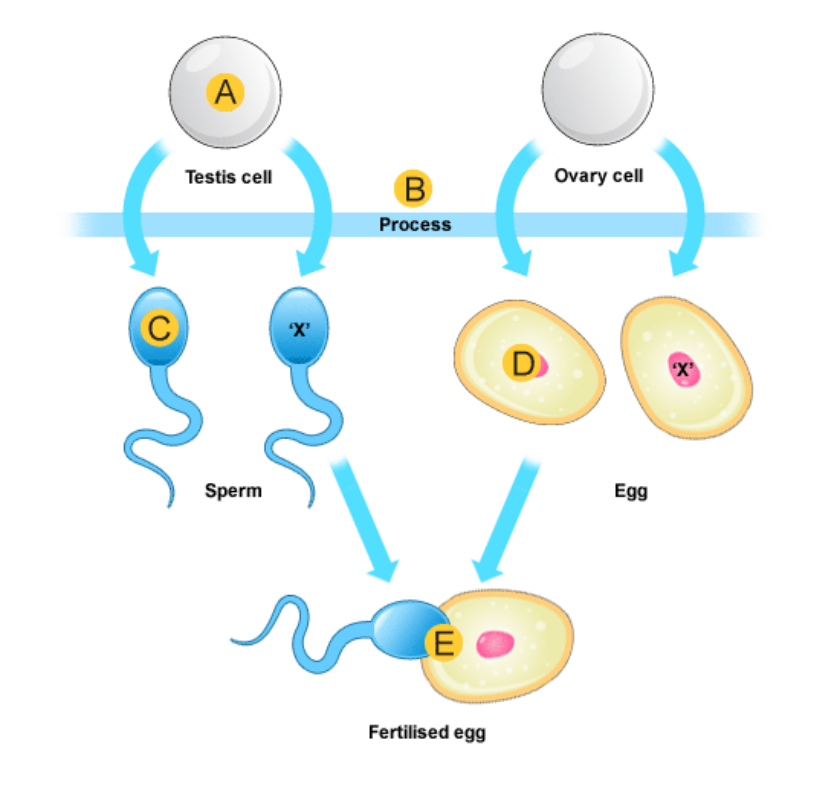

Figure 4–15. Molecular Biology of the Cell, 4th Edition.

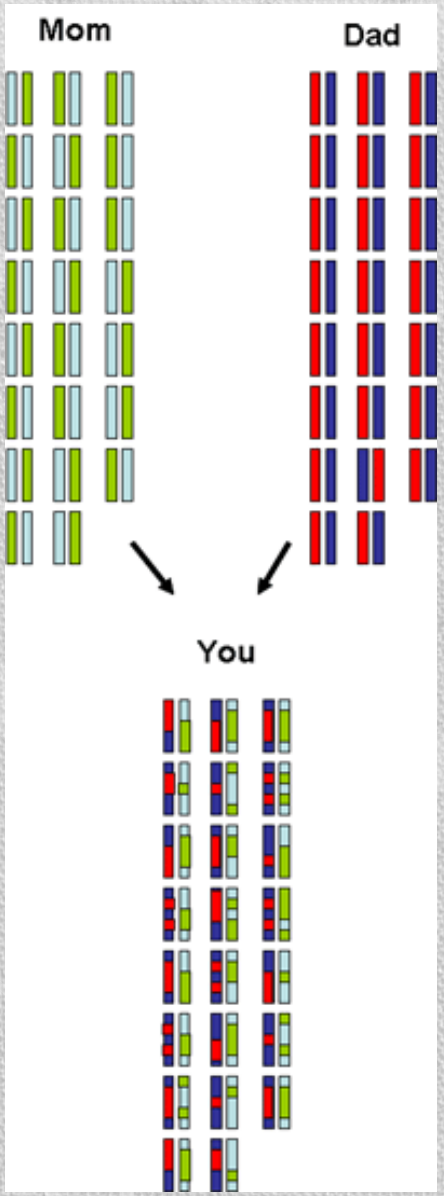
**Human Chromosome:**

- 44 homologous chromosomes and 2 sex chromosomes
- The arrangement of the full chromosome set is called **karyotype**



- Genes are inherited as units, with two parents dividing out copies of their genes to their offspring.

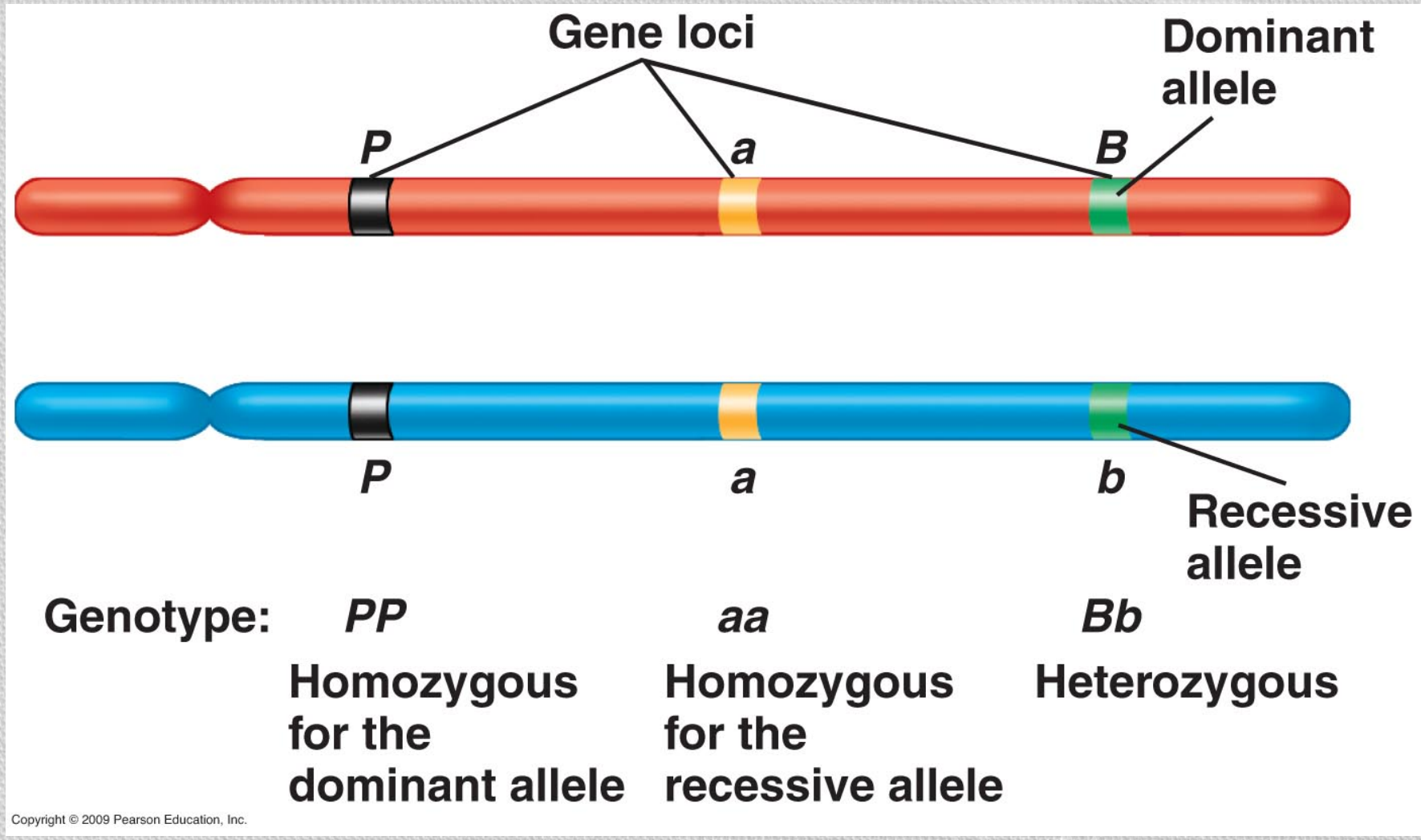


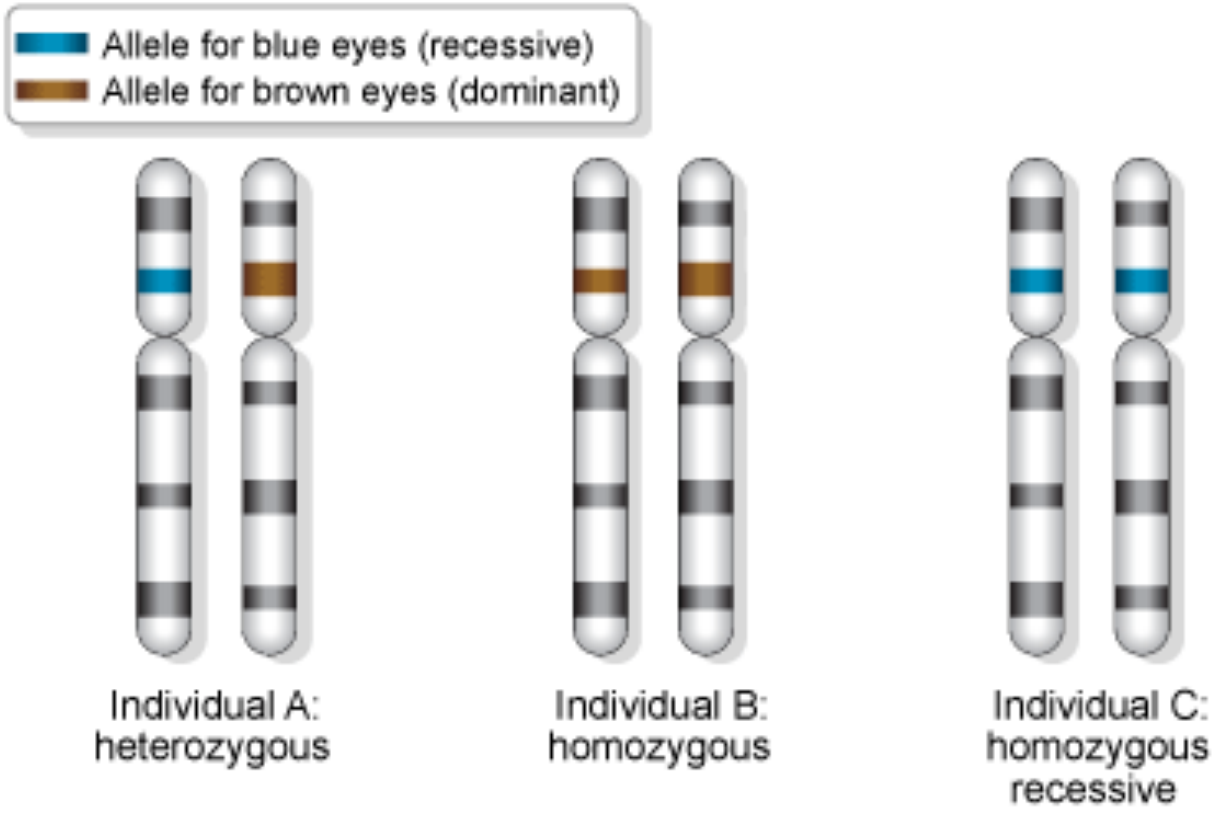



# Allele

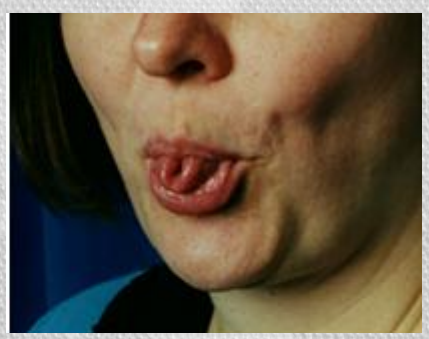
- The information within a particular gene is not always exactly the same between one organism and another, so different copies of a gene do not always give exactly the same instructions. Each unique form of a single gene is called an allele.

- The effects of this mixing depend on the types (the alleles) of the gene If one allele overrides the instructions from another, it is called the **dominant** allele, and the allele that is overridden is called the **recessive** allele.

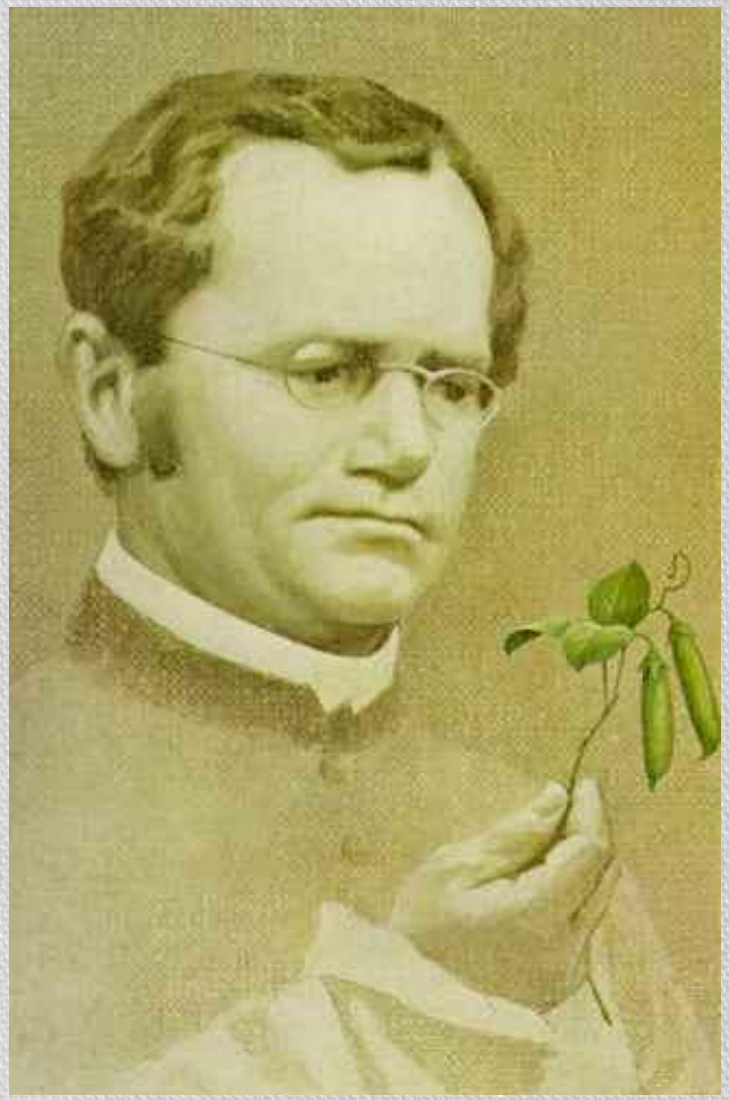


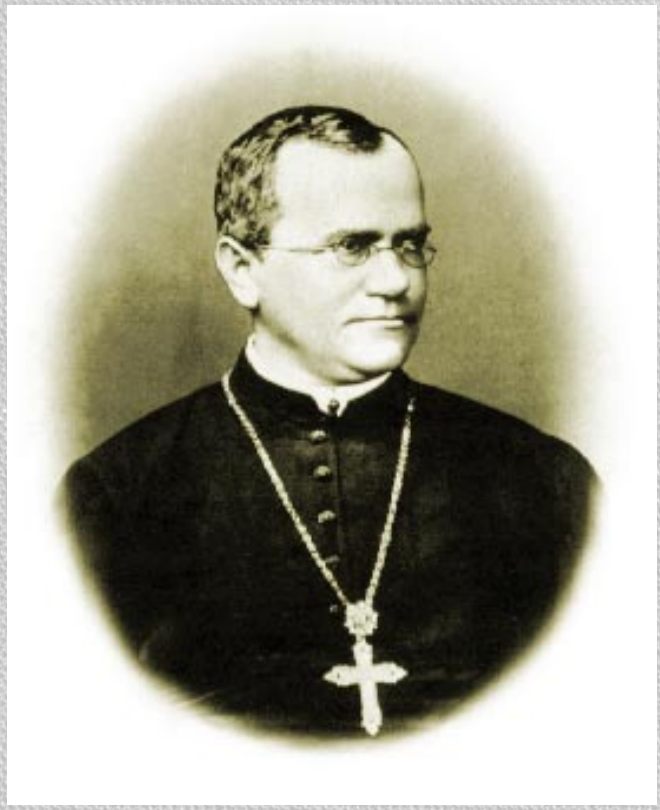

Allele for blue eyes (recessive)
Allele for brown eyes (dominant)
Individual A: heterozygous
Individual B: homozygous
Individual C: homozygous recessive

| Dominant | Recessive |
|---|---|
| Low heart rate | High heart rate |
| Widow's peak | straight hair line |
| ocular hypertelorism | Hypotelorism |
| normal digestive muscle | POLIP syndrome |
| Facial dimples | No facial dimples |
| Able to taste PTC | Unable to taste PTC |
| Unattached (free) earlobe | Attached earlobe |
| Clockwise hair direction (left to right) | Counter-Clockwise hair direction (right to left) |
| Cleft chin | smooth chin |
| straight nose | turned up nose |
| no progressive nerve damage | Friedreich's ataxia |
| Ability to roll tongue (Able to hold tongue in a U shape) | No ability to roll tongue |
| extra finger or toe | Normal five fingers and toes |
| straight pinkies | Crooked pinkies |
| Straight Thumb | Hitchhiker's Thumb |
| Freckles | No freckles |
| Wet-type earwax | Dry-type earwax |

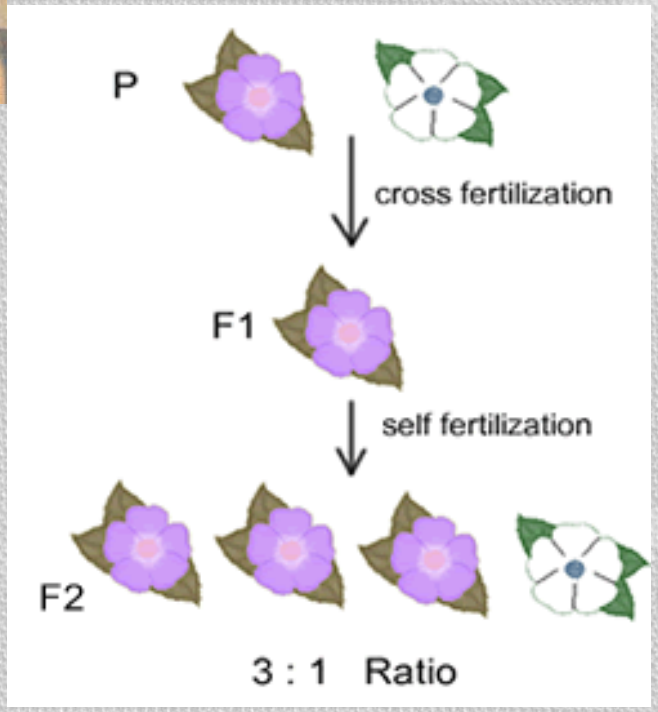



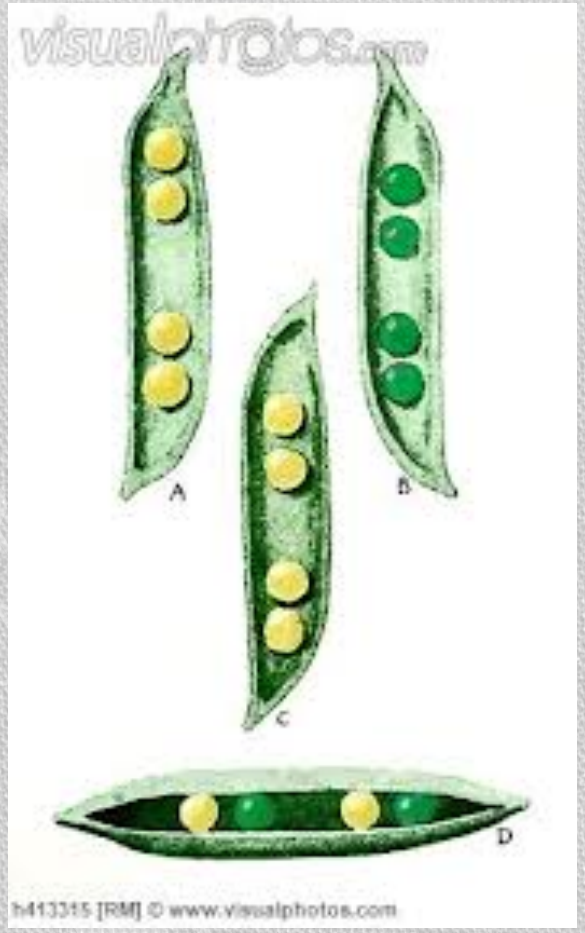

# Genetics: History and Mendel





In the mid-1800s, the rules underlying patterns of inheritance were uncovered in a series of experiments performed by an Austrian monk named Gregor Mendel.


P
cross fertilization
F1
self fertilization
F2
3 : 1 Ratio


A
B
C
D
h413315 [RM] © www.visualphotos.com

| Character | Dominant Trait | × | Recessive Trait | $F_2$ Generation Dominant:Recessive | Ratio |
|---|---|---|---|---|---|
| Flower color |  Purple | × |  White | 705:224 | 3.15:1 |
| Flower position |  Axial | × |  Terminal | 651:207 | 3.14:1 |
| Seed color |  Yellow | × |  Green | 6022:2001 | 3.01:1 |
| Seed shape |  Round | × |  Wrinkled | 5474:1850 | 2.96:1 |
| Pod shape |  Inflated | × |  Constricted | 882:299 | 2.95:1 |
| Pod color |  Green | × |  Yellow | 428:152 | 2.82:1 |
| Stem length |  Tall | × |  Dwarf | 787:277 | 2.84:1 |

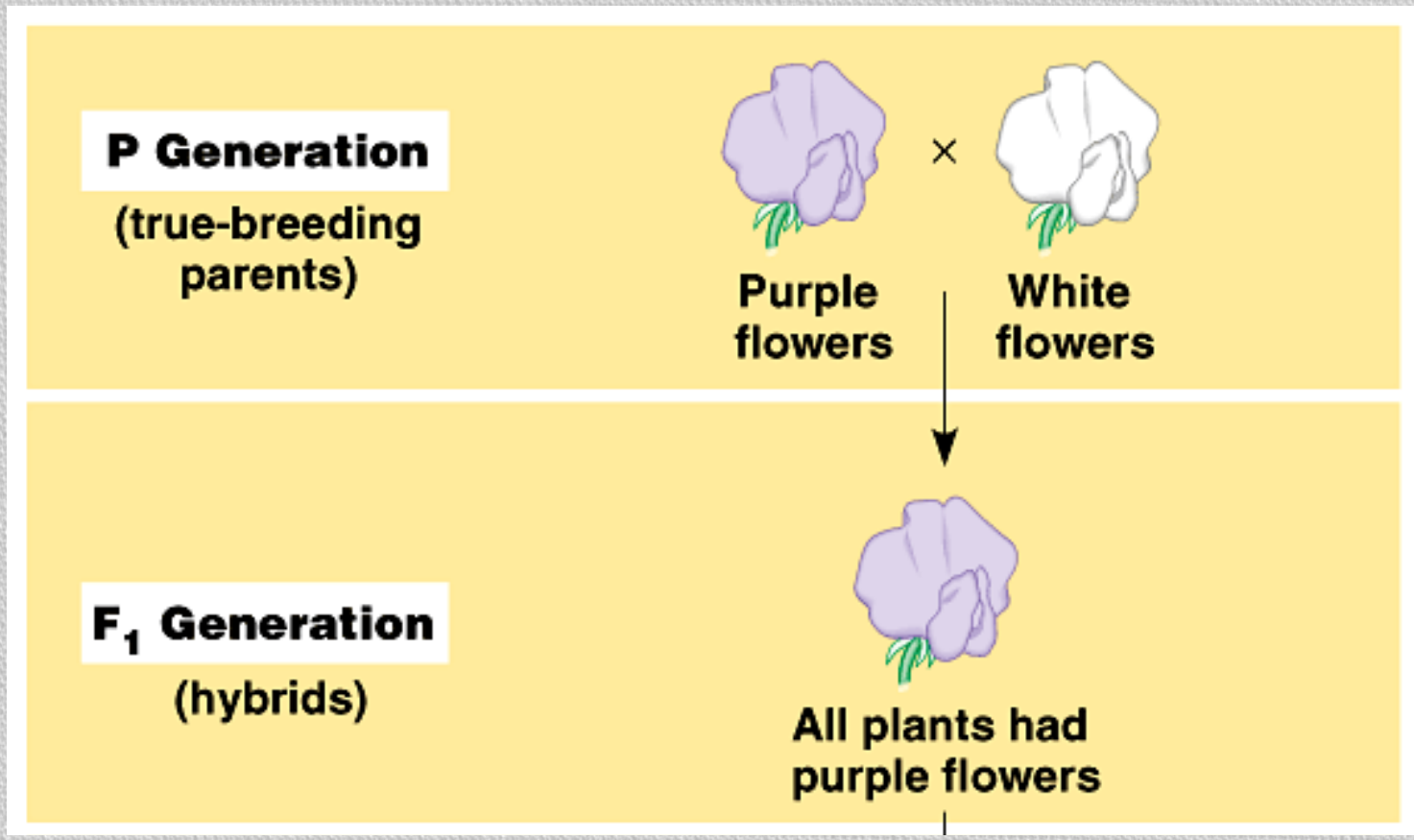
Statistics indicated a pattern.

# Mendel's Law

the Law of Dominance: In a cross of parents that are pure for contrasting traits, only one form of the trait will appear in the next generation.

Offspring that are hybrid for a trait will have only the dominant trait in the phenotype.

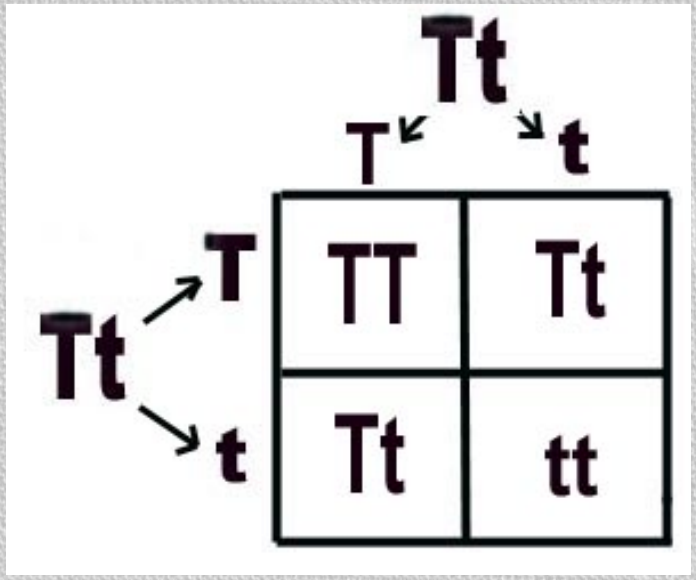
| Parent Pea Plants | F1 Pea Plants |
|---|---|
| tall stem x short stem | all tall stems |
| yellow seeds x green seeds | all yellow seeds |
| green pea pods x yellow pea pods | all green pea pods |
| round seeds x wrinkled seeds | all round seeds |
| axial flowers x terminal flowers | all axial flowers |


P Generation
(true-breeding parents)
Purple flowers
×
White flowers
F1 Generation
(hybrids)
All plants had purple flowers

The Law of Segregation: During the formation of gametes (eggs or sperm), the two alleles responsible for a trait separate from each other.

Alleles for a trait are then "recombined" at fertilization, producing the genotype for the traits of the offspring.

Punnet square

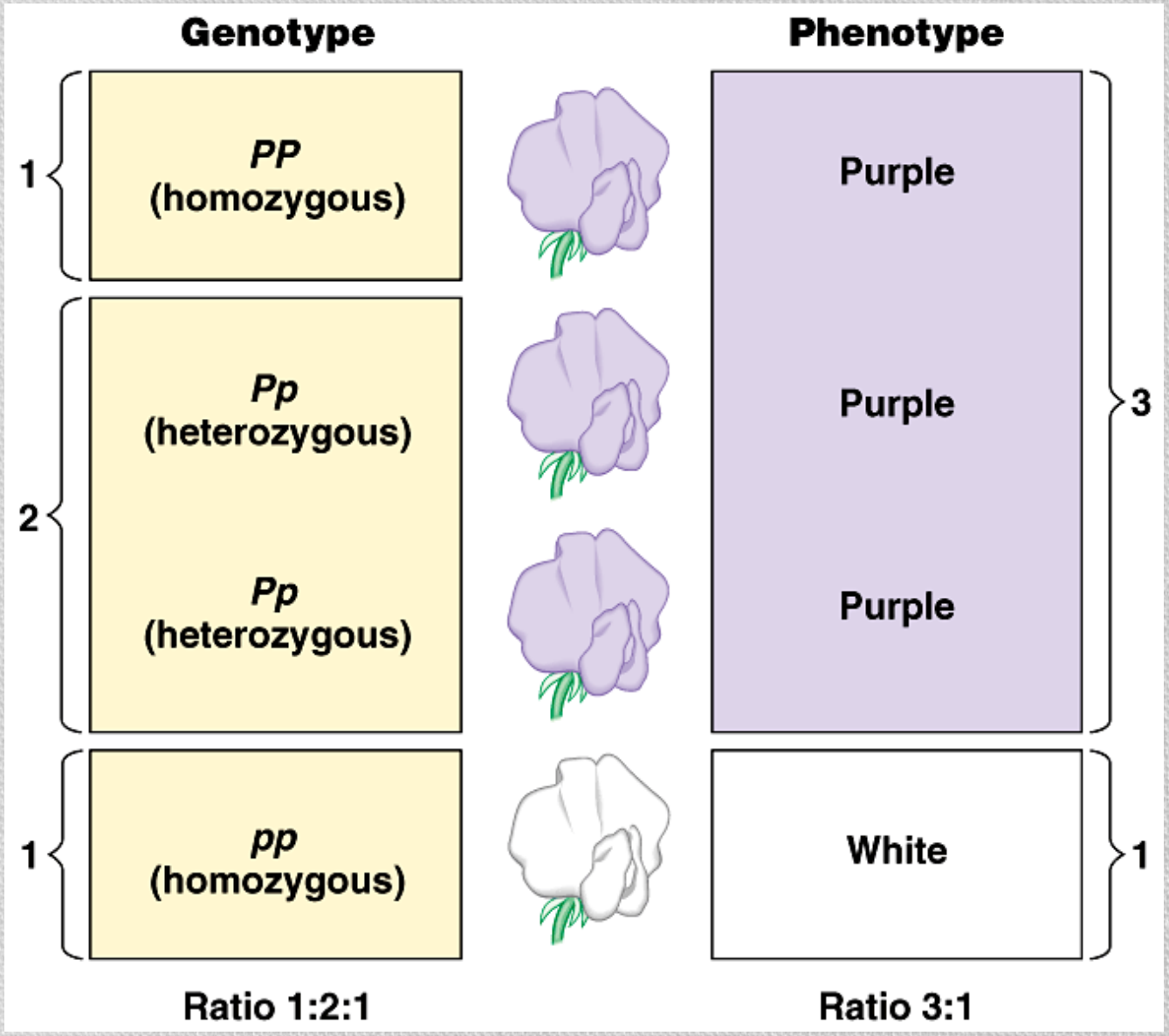


The Law of Independent Assortment: Alleles for different traits are distributed to offspring independently of one another.

## RrGg x RrGg

| | RG | Rg | rG | rg |
|---|---|---|---|---|
| RG | RRGG round | RRGg round | RrGG round | RrGg round |
| Rg | RRGg round | RRgg round | RrGg round | Rrgg round |
| rG | RrGG round | RrGg round | rrGG wrinkled | rrGr wrinkled |
| rg | RrGg round | Rrgg round | rrGg wrinkled | rrgg wrinkled |

# Genotype versus phenotype.




genotype ratio differs from the phenotype ratio

# Part II: Human Genetic Disorders

# Genetic disorder

- A genetic disorder is an illness caused by abnormalities in the genome, especially a condition that is present from birth (congenital).

- Most genetic disorders are quite rare and affect one person in every several thousands or millions.

- Genetic disorders are heritable, and are passed down from the parents' genes.

# Mutations

- Defined as a permanent change in the DNA
- Origin
  - germ cells – transmitted to progeny
  - somatic cells – cancer and some congenital malformations
- Types of mutation
  - Chromosome mutation – structural changes within the chromosome – translocations, deletions, etc
  - Genome mutation – loss or gain of whole chromosomes: monosomy and trisomy
  - Gene mutation – alterations at the level of the gene

# Genetic Code

| | | | |
|---|---|---|---|
| UUU, UUC — phenylalanine<br>UUA, UUG — leucine | UCU, UCC, UCA, UCG — serine | UAU, UAC — tyrosine<br>UAA, UAG — stop | UGU, UGC — cysteine<br>UGA — stop<br>UGG — tryptophan |
| CUU, CUC, CUA, CUG — leucine | CCU, CCC, CCA, CCG — proline | CAU, CAC — histidine<br>CAA, CAG — glutamine | CGU, CGC, CGA, CGG — arginine |
| AUU, AUC, AUA — isoleucine<br>AUG — methionine | ACU, ACC, ACA, ACG — threonine | AAU, AAC — asparagine<br>AAA, AAG — lysine | AGU, AGC — serine<br>AGA, AGG — arginine |
| GUU, GUC, GUA, GUG — valine | GCU, GCC, GCA, GCG — alanine | GAU, GAC — aspartic acid<br>GAA, GAG — glutamic acid | GGU, GGC, GGA, GGG — glycine |

# Point mutation

- result from substitution of a single base in the DNA

Coding portion of gene

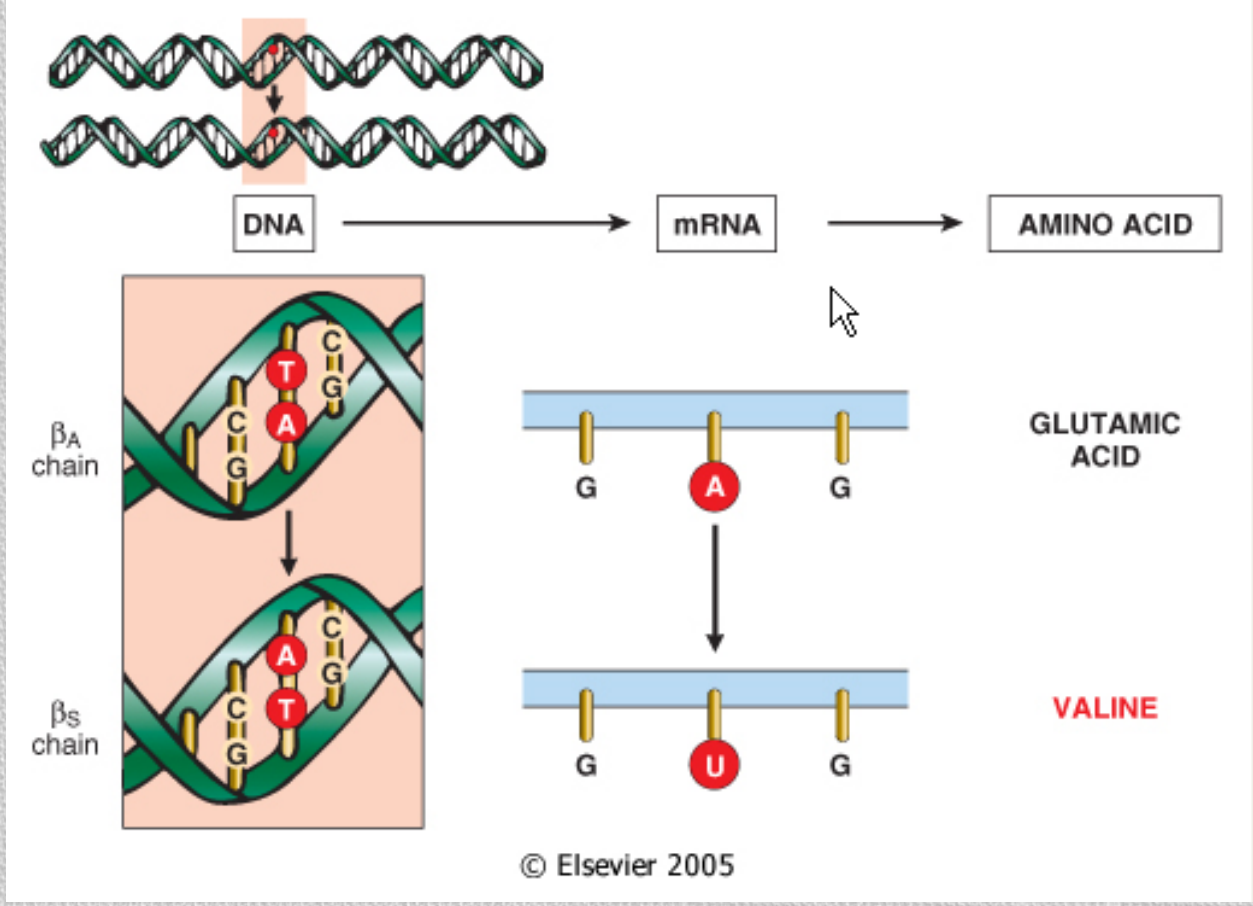
- Missense– result in substitution of one amino acid for another in the coded protein
  - conservative – function of protein is not affected
  - nonconservative – function of protein altered
- Nonsense
  - stop codon – results in truncated protein
- Noncoding portion of gene
  - promoter and enhancer regions
  - posttranslational processing – defective splicing

# Deletions and Insertions

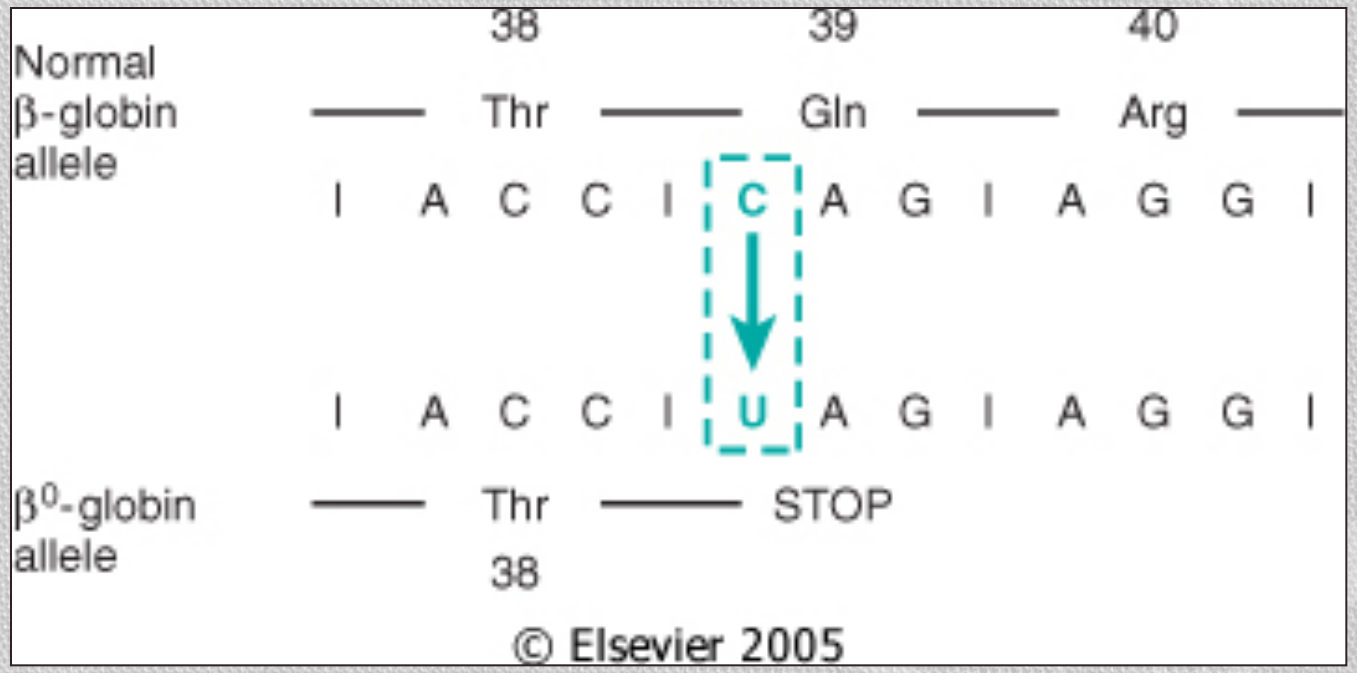
- Deletion of multiple of 3 bases
  - three bases code for one amino acid
  - abnormal protein missing one or more amino acids

- Frameshift mutation
  - will result in a different sequence of base triplets
  - meaning of genetic code is altered distal to the mutation
  - usually leads to stop codon and truncated protein

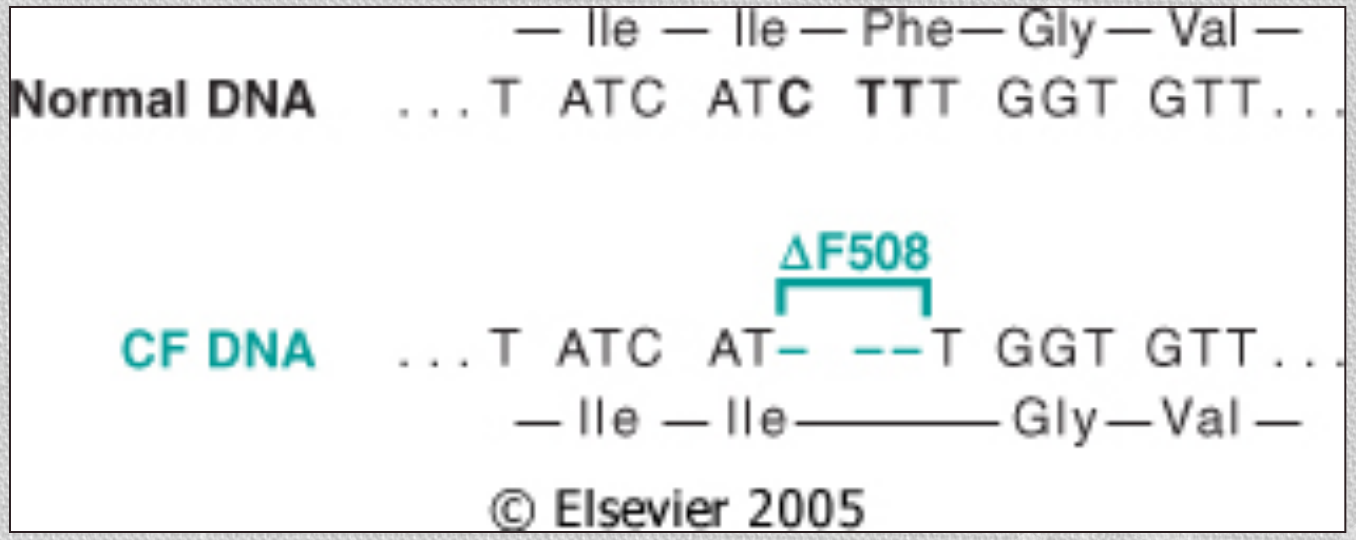
# Hemoglobin S: Point Mutation Resulting From A Single Base Pair Change In The DNA (Sickle Cell Anemia)



# $\beta^0$ Thalassemia: Point Mutation Leading To Premature Chain Termination



# Three-base Deletion In The Common Cystic Fibrosis (CF) Allele



# I. Single Gene Disorder

- Monogenic disease; Mendelian disorder
- The result of a single mutated gene
- Over 4000 human diseases are caused by single gene defects.
- Single gene disorders can be passed on to subsequent generations in several ways.
- Most follow pattern of Mendelian inheritance
- Main types
  - Autosomal dominant
  - Autosomal recessive
  - X-linked dominant
  - X-linked recessive
  - Y-linked

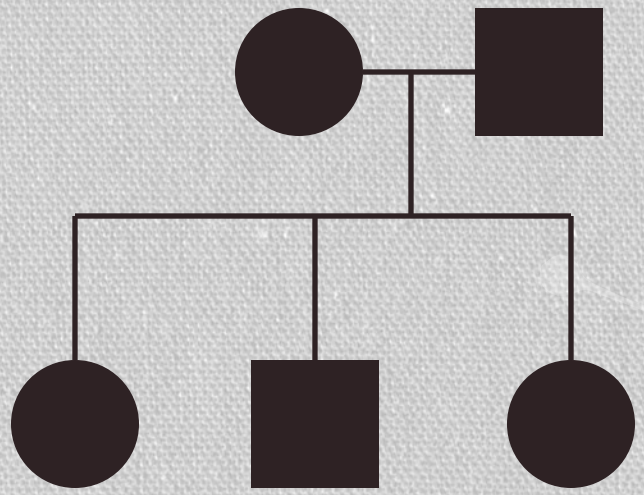
# Pedigree

- Pedigree:  a family history that shows how a trait is inherited over several generations.

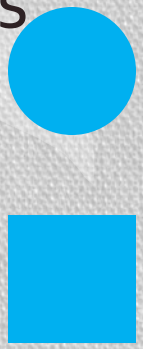
- Pedigrees are usually used when parents want to know if they are carriers of a particular disorder

# Making a Pedigree

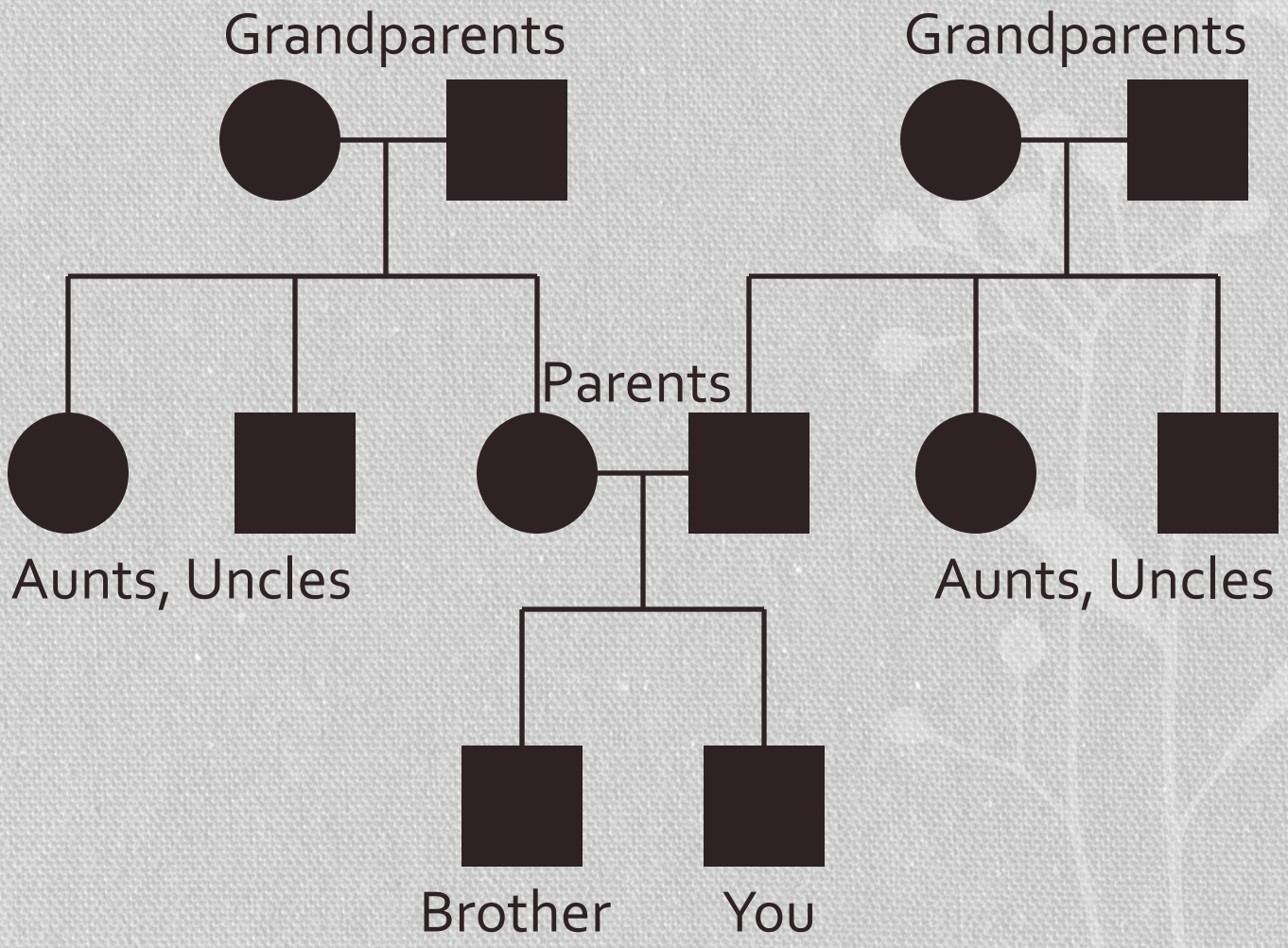
- Female
- Male
- Married Couple
- Siblings



Filled in symbols indicate individual is affected with a disorder



# Example of a Pedigree

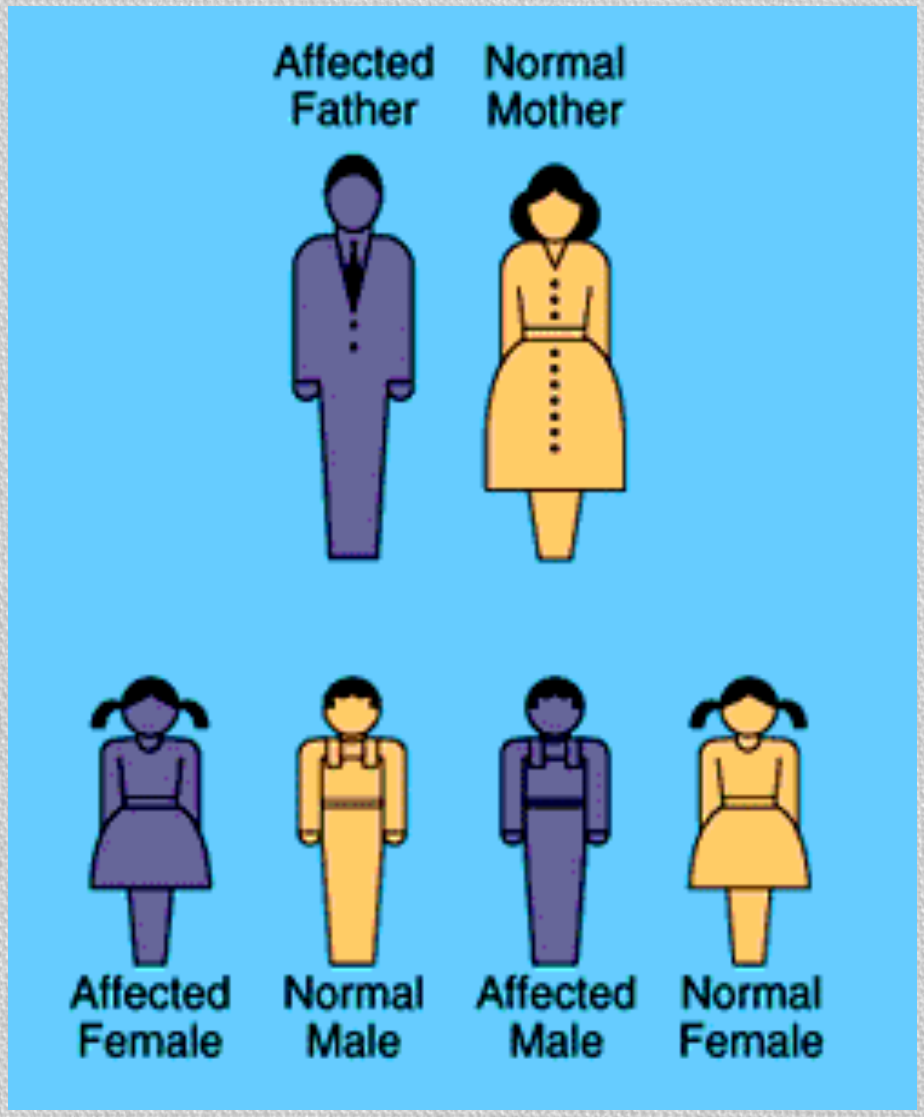


# Autosomal Dominant

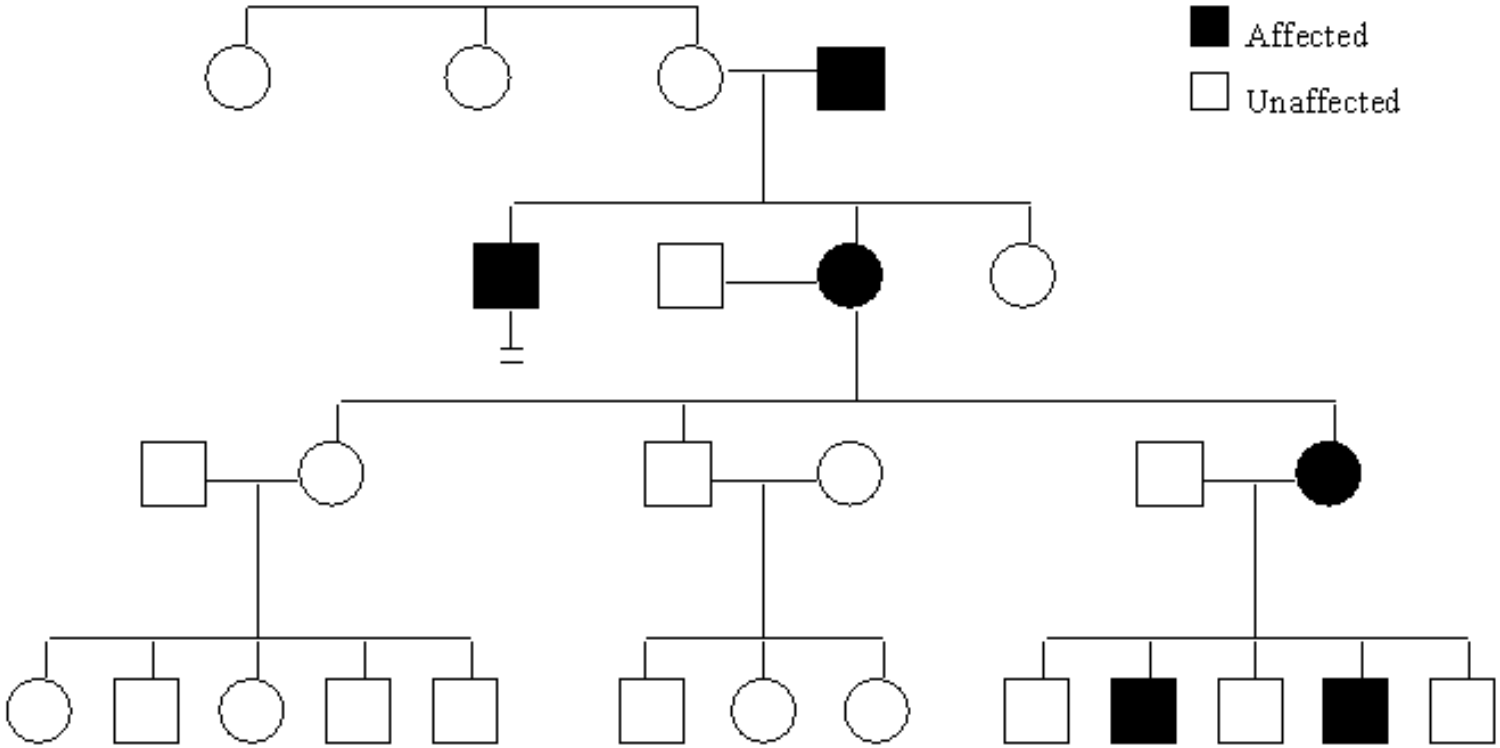
- Only one mutated copy of the gene will be necessary for a person to be affected by an autosomal dominant disorder.

- Each affected person usually has one affected parent. The chance a child will inherit the mutated gene is 50%.

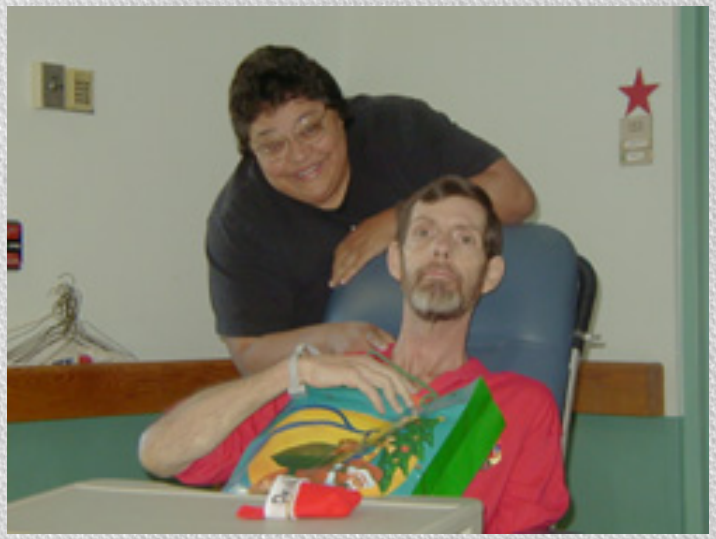
- Examples of autosomal dominant traits and disorders are Huntingtons disease and achondroplasia.


Affected Father
Normal Mother
Affected Female
Normal Male
Affected Male
Normal Female

# Autosomal Dominant



# Huntington’s Disease



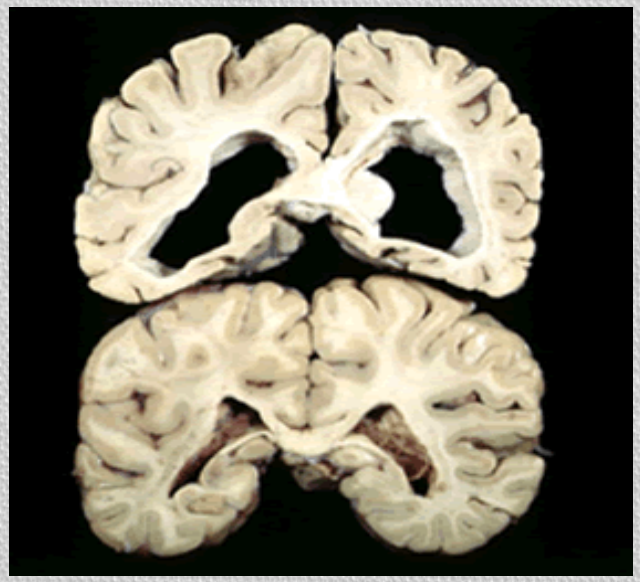
- Huntington's disease (HD) is an inherited, degenerative brain disorder which results in an eventual loss of both mental and physical control. The disease is also known as Huntington's chorea. Chorea means "dance-like movements" and refers to the uncontrolled motions often associated with the disease.

# Huntington's Disease.

- An autosomal dominant neurodegenerative disorder first clinically described by Huntington in 1872.
- It is rare, with death rates of 1.6 per million, and is commonest among white Europeans.
- Initial symptoms occur in those aged 30 – 50.
- Sometimes it can strike in the 20's, when suicide is often the result.
- Death comes on average 12 years after the onset.

# Huntington gene

```
  1 ttg ctg tgt gag gca gaa cct gcg ggg gca
    ggg gcg ggc tgg ttc cct ggc cag cca ttg
 61 gca gag tcc gca ggc tag ggc tgt caa tca
    tgc tgg ccg gcg tgg ccc cgc ctc cgc cgg
121 cgc ggc ccc gcc tcc gcc ggc gca cgt ctg
    gga cgc aag gcg ccg tgg ggg ctg ccg gga
181 cgg gtc caa gat gga cgg ccg ctc agg ttc
    tgc ttt tac ctg cgg ccc aga gcc cca ttc
241 att gcc ccg gtg ctg agc ggc gcc gcg agt
    cgg ccc gag gcc tcc ggg gac tgc cgt gcc
301 ggg cgg gag acc gcc atg gcg acc ctg gaa
    aag ctg atg aag gcc ttc gag tcc ctc aag
361 tcc ttc cag cag cag cag cag cag cag cag
    cag cag cag cag cag cag cag cag cag cag
421 cag cag cag caa cag ccg cca ccg ccg ccg
    ccg ccg ccg ccg cct cct cag ctt cct cag
```

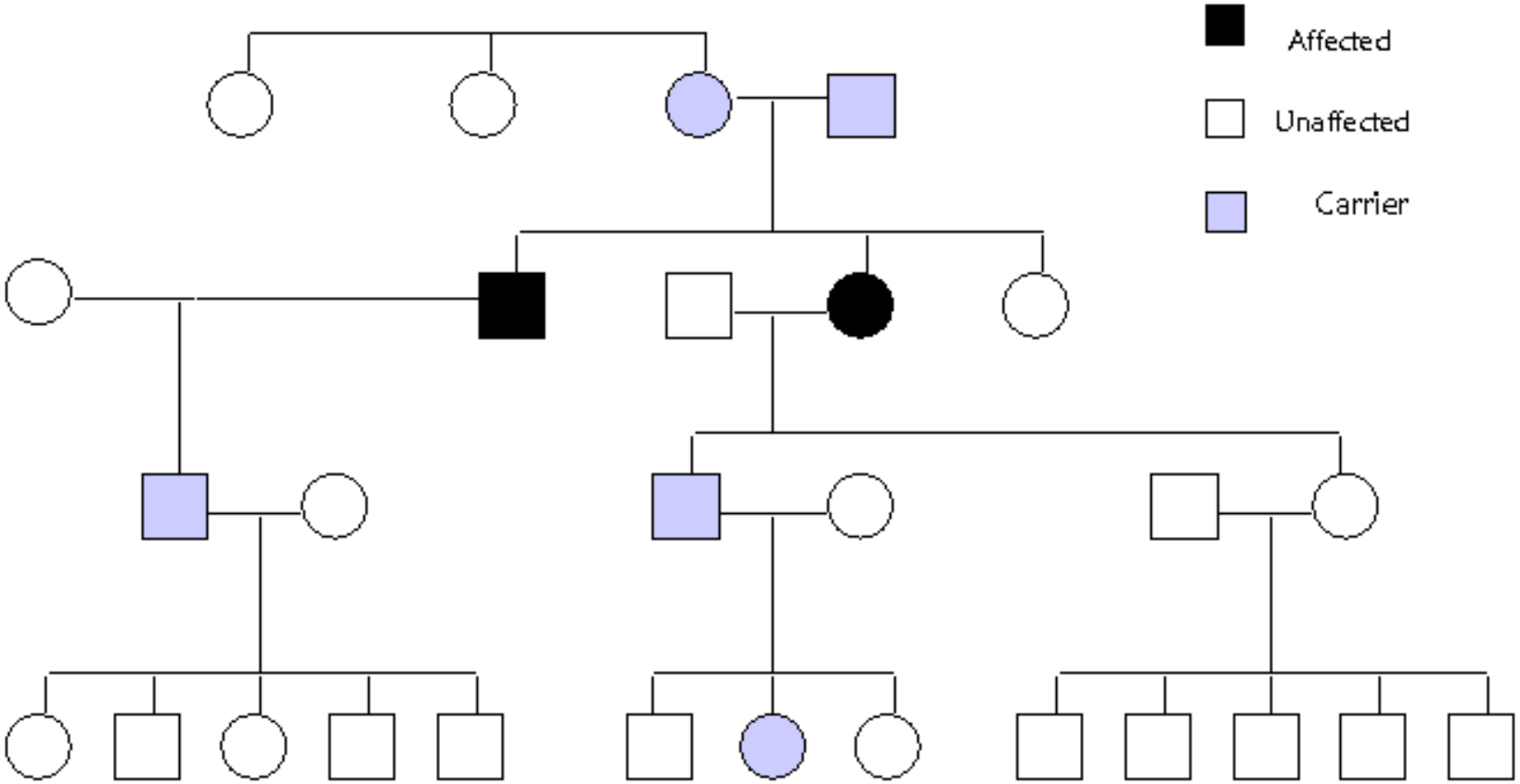


Encodes a run of 11-34 glutamine amino acid residues in the HD protein. A run of > 34 glutamine residues causes the protein to aggregate in the brain cells and cause progressive cell death.

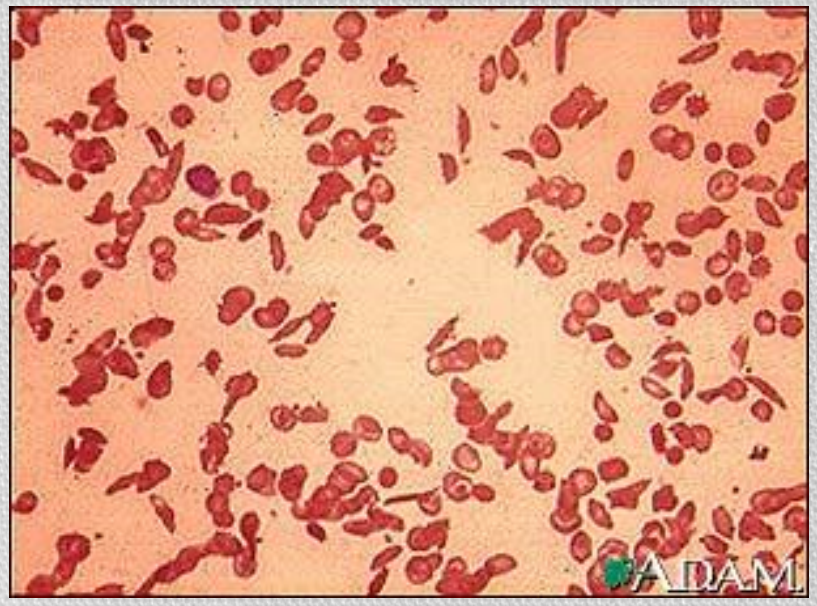
# Autosomal Recessive

- Two copies of the gene must be mutated for a person to be affected by an autosomal recessive disorder.
- An affected person usually has unaffected parents who each carry a single copy of the mutated gene (and are referred to as carriers). Two unaffected people who each carry one copy of the mutated gene have a 25% chance with each pregnancy of having a child affected by the disorder.
- Examples of this type of disorder are cystic fibrosis, sickle cell disease, Tay Sachs disease, Spinal muscular atrophy and Roberts syndrome.

# Autosomal Recessive



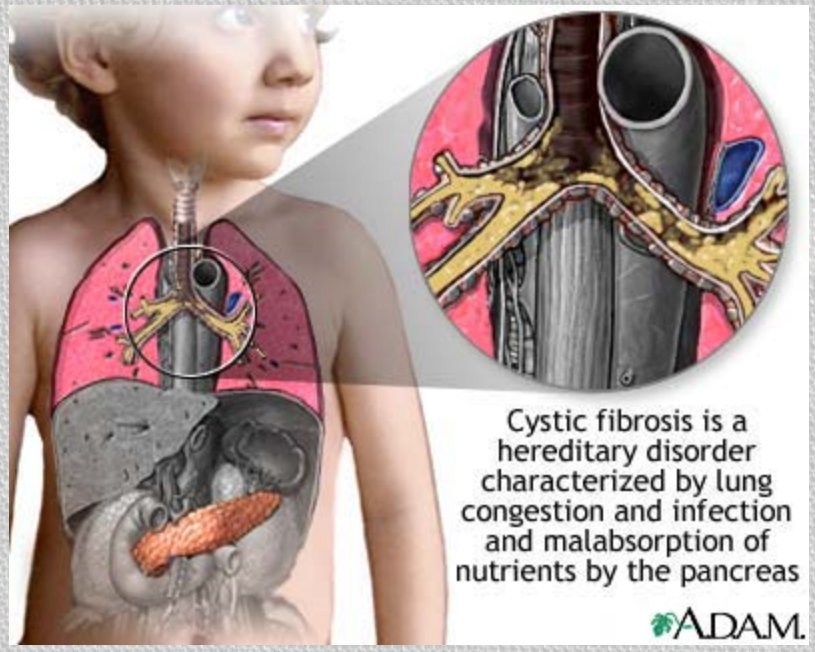
# Sickle Cell Anemia


- Disorder where abnormal hemoglobin (a protein inside red blood cells) is produced and warps red blood cells
- Sickle cells deliver less oxygen to body's tissues and can get stuck in small blood vessels
- tends to be seen in people of African or Mediterranean descent

## Cystic Fibrosis

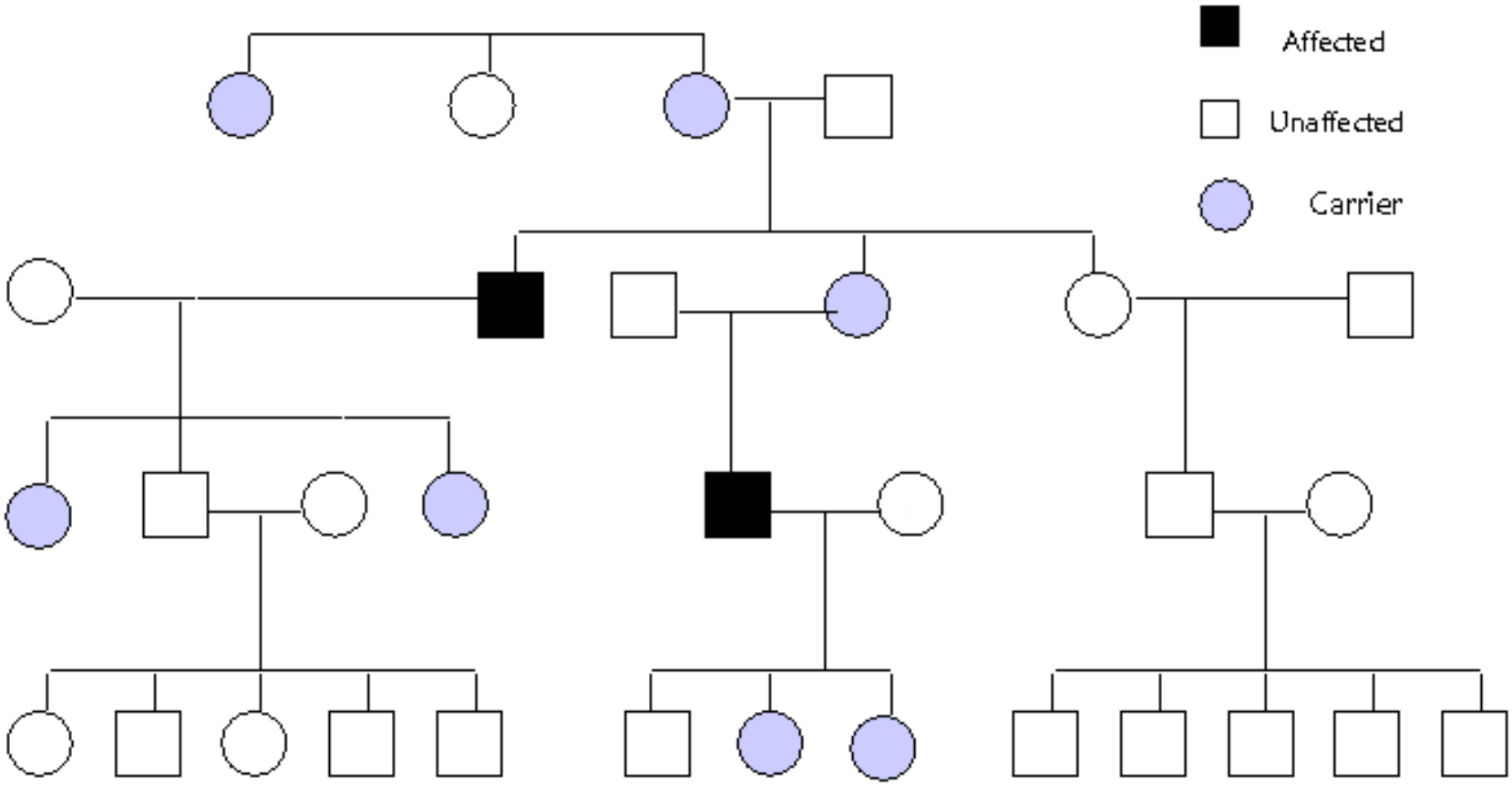
- Life threatening, causes thick mucus to build up in various areas of the body (lungs, digestive tract, etc).
- Tends to run in Caucasians, of Northern/Central European descent  (1 in 29 Americans carry the allele)
- Average life span in US for people with CF is 37, death usually caused by lung complications



# X-linked recessive

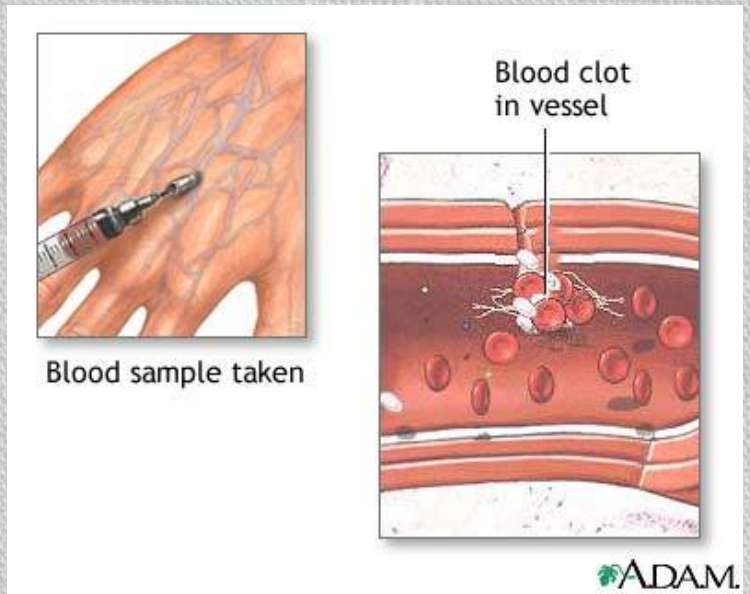
- X-linked recessive conditions are also caused by mutations in genes on the X chromosome. Males are more frequently affected than females, and the chance of passing on the disorder differs between men and women.
- The sons of a man with an X-linked recessive disorder will not be affected, and his daughters will carry one copy of the mutated gene. A woman who is a carrier of an X-linked recessive disorder ($X^RX^r$) has a 50% chance of having sons who are affected and a 50% chance of having daughters who carry one copy of the mutated gene and are therefore carriers.
- hemophilia A, Duchenne muscular dystrophy, and Lesch-Nyhan syndrome, male pattern baldness and red-green color blindness.

# X-linked Recessive

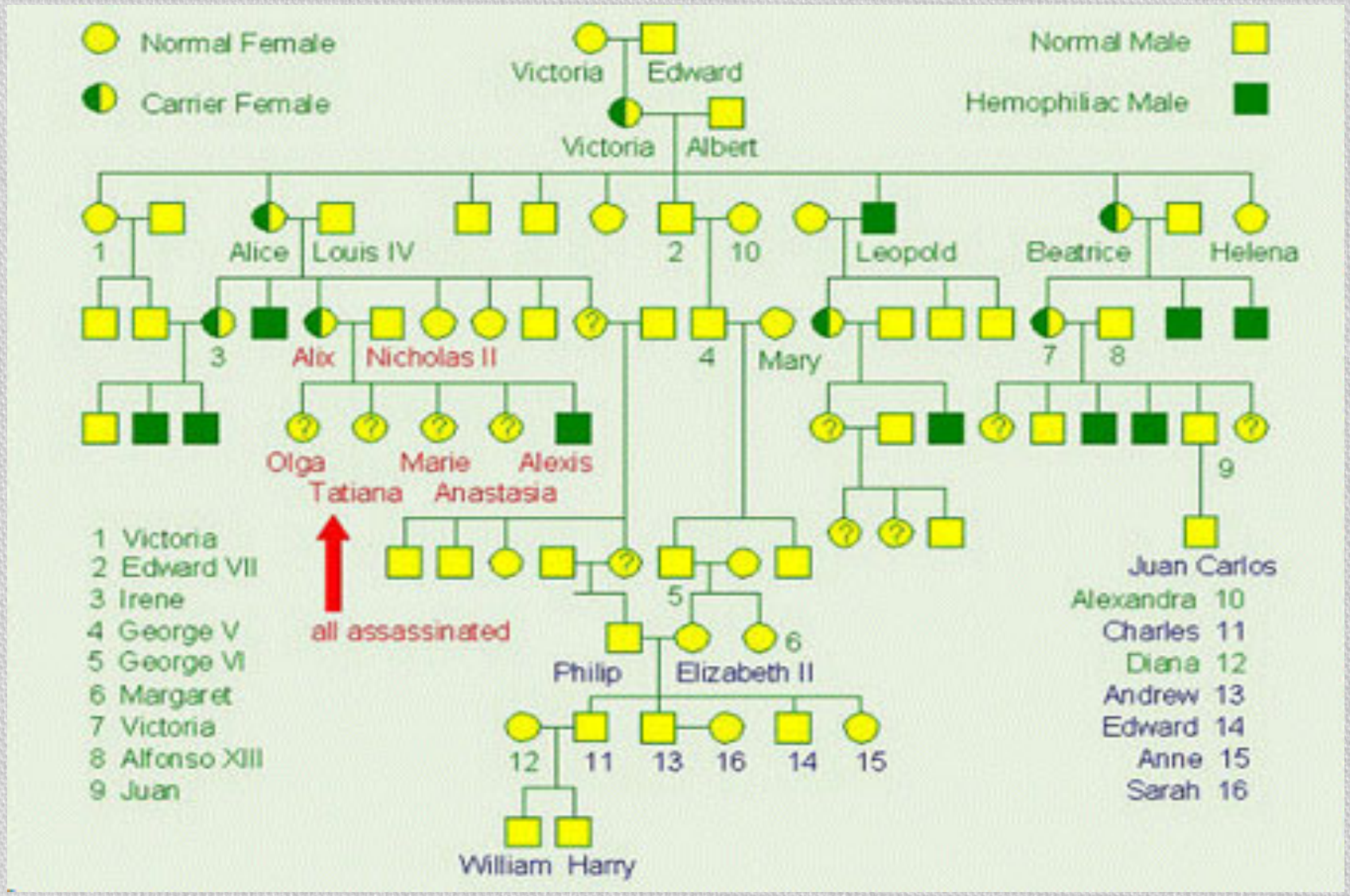


# Hemophilia

- Bleeding disorder, where it takes a long time for blood to clot (body lacks protein FVIII involved in clotting)
- Sex-linked (carried on the X chromosome)
- Treatment involves injection with missing clotting protein.
- the royal disease.



# X-linked Inheritance pedigree chart



# X-linked dominant

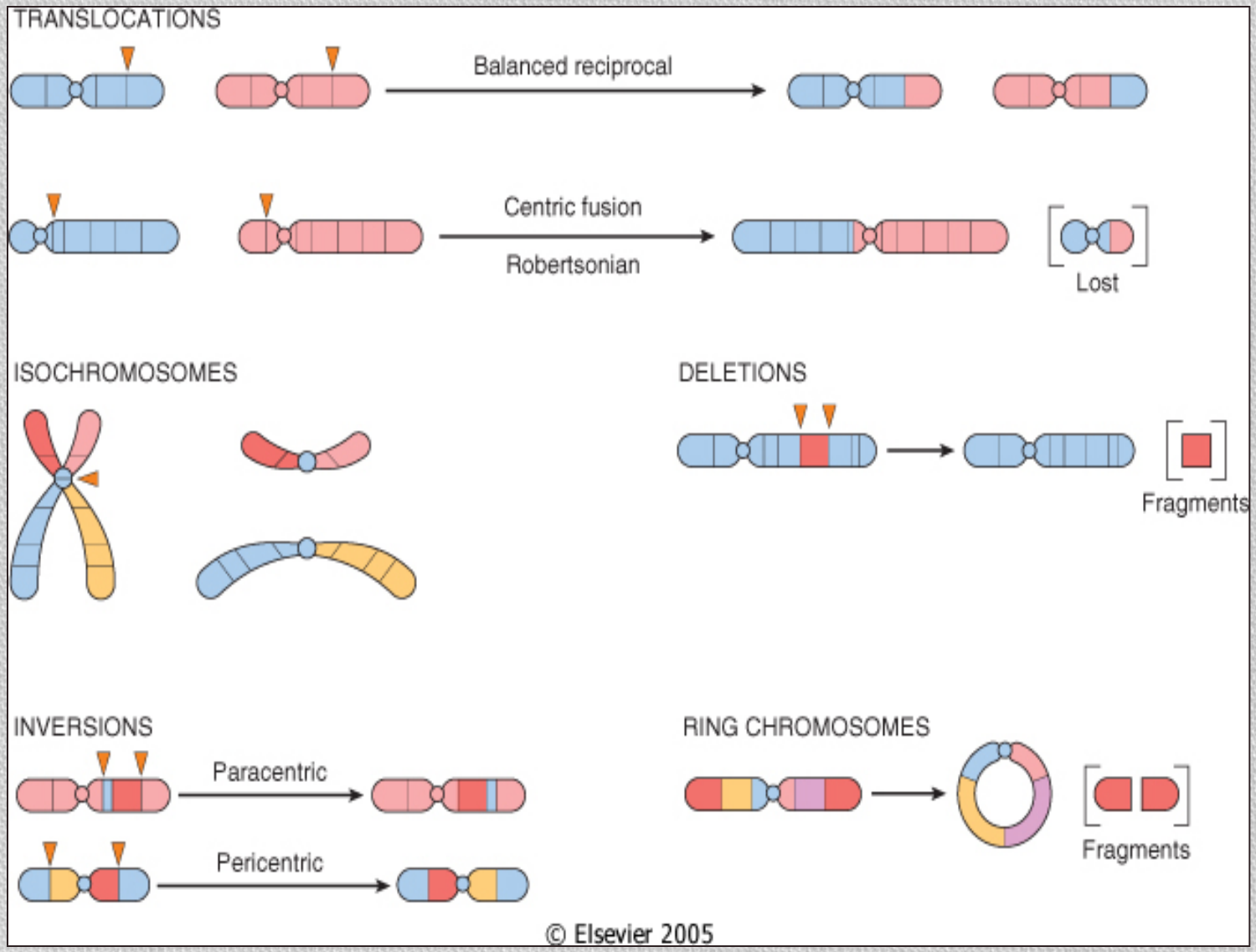
- X-linked dominant disorders are caused by mutations in genes on the X chromosome. Only a few disorders have this inheritance pattern, with a prime example being X-linked hypophosphatemic rickets.
- Males and females are both affected in these disorders, with males typically being more severely affected than females.
- Some X-linked dominant conditions, such as Rett syndrome, incontinent pigment type 2 and Aicardi syndrome, are usually fatal in males either *in utero* or shortly after birth, and are therefore predominantly seen in females.

# II. Chromosomal disease

- Human disorders due to chromosome alterations.

- Most chromosome abnormalities occur as an accident in the egg or sperm, and therefore the anomaly is present in every cell of the body. Some anomalies, however, can happen after conception. If the parents do not possess the abnormality it was not initially inherited; however it may be transmitted to subsequent generations.

# Types Of Chromosomal Rearrangements



# Aneuploidy

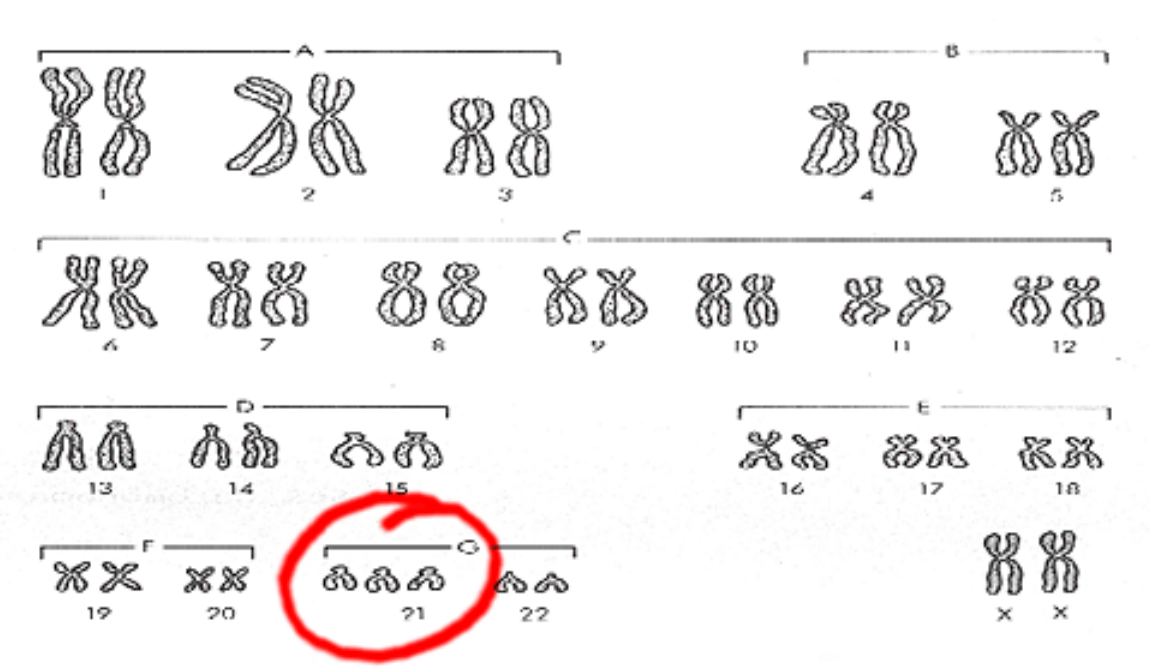
- Abnormal number of chromosomes
- Autosomal:
  - Trisomy 21 (Down syndrome)
  - Trisomy 18 (Edward syndrome)
  - Trisomy 13 (Patau syndrome)

- Sex chromosome:
  - 47XXY (Klinefelter syndrome)
  - 45X (Turner syndrome)

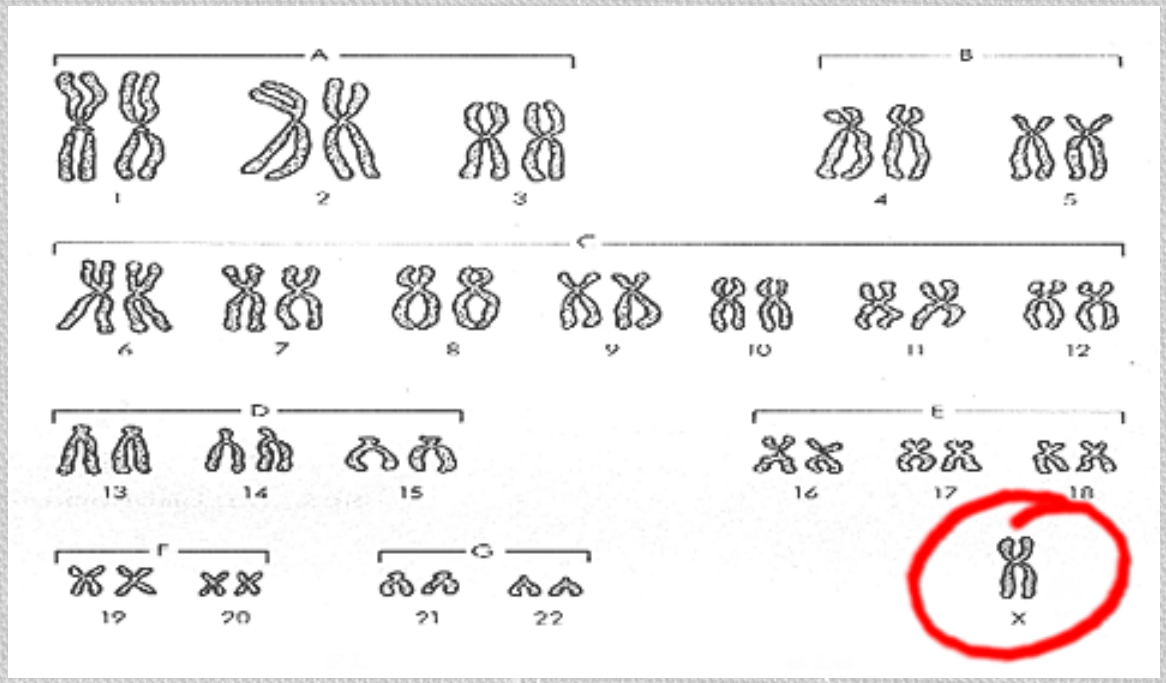
# Down's Syndrome

- Caused by non-disjunction of the 21st chromosome.
- Most common chromosomal disorder
- Affects 1 in 750 newborns overall, but is related to maternal age
  - 1 in 1550 live births of mothers > 20 years
  - 1 in 25 live births of mothers > 45 years

# Down's Syndrome or Trisomy 21

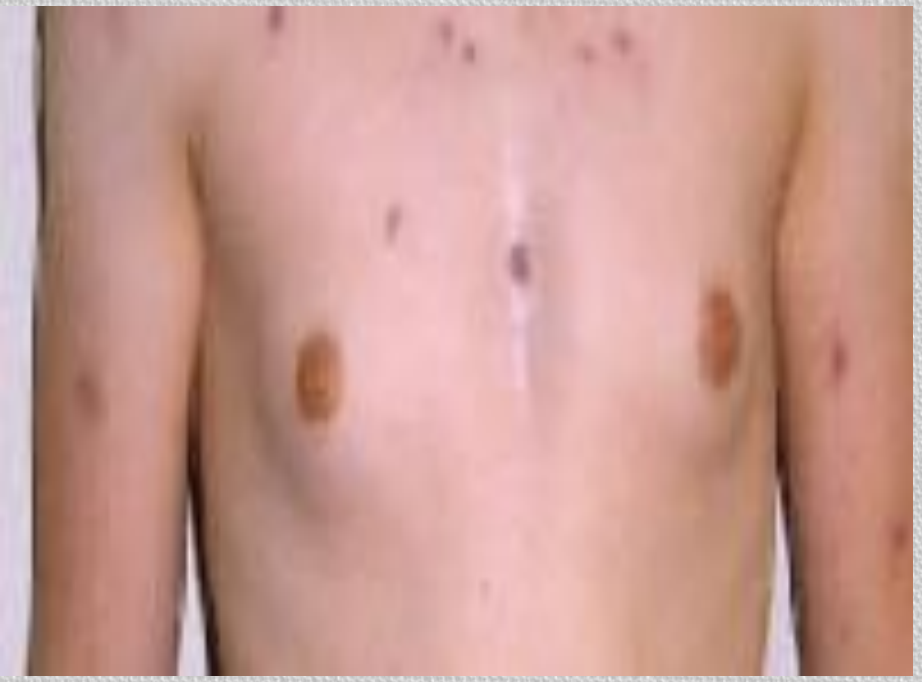


# Turner's Syndrome



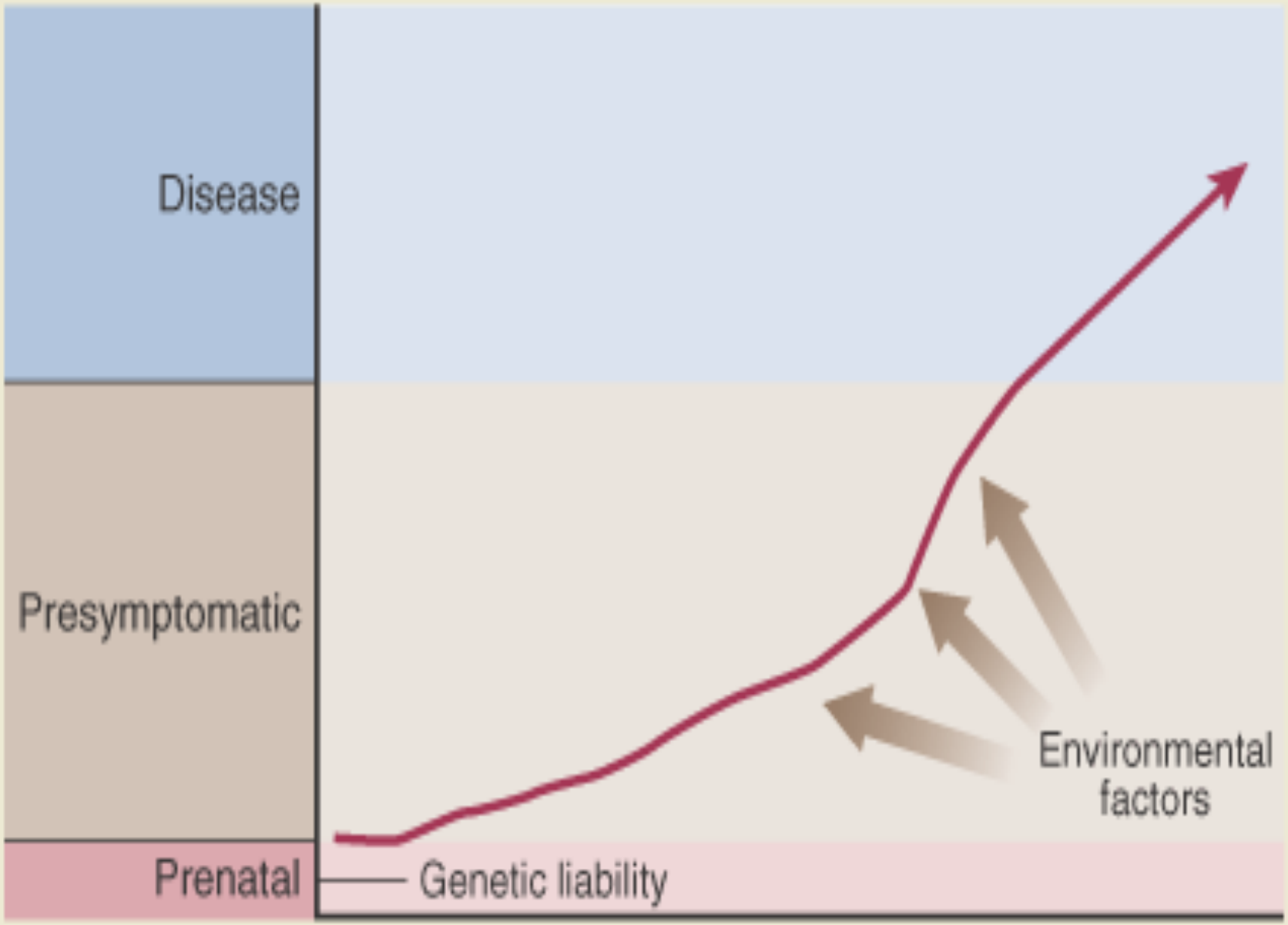
- Turner syndrome is associated with underdeveloped ovaries, short stature, webbed, and is only in women.
- Bull neck, and broad chest. Individuals are sterile, and lack expected secondary sexual characteristics.
- Mental retardation typically not evident.

# Klinefelter syndrome 47, XXY



# III. Multifactorial and polygenic (complex) disorders

- Associated with the effects of multiple genes in combination with lifestyles and environmental factors.
- heart disease, most cancers, and behavioral disorders such as alcoholism, obesity, mental illness, and Alzheimer's disease are examples
- Although complex disorders often cluster in families, they do not have a clear-cut pattern of inheritance. This makes it difficult to determine a person's risk of inheriting or passing on these disorders.



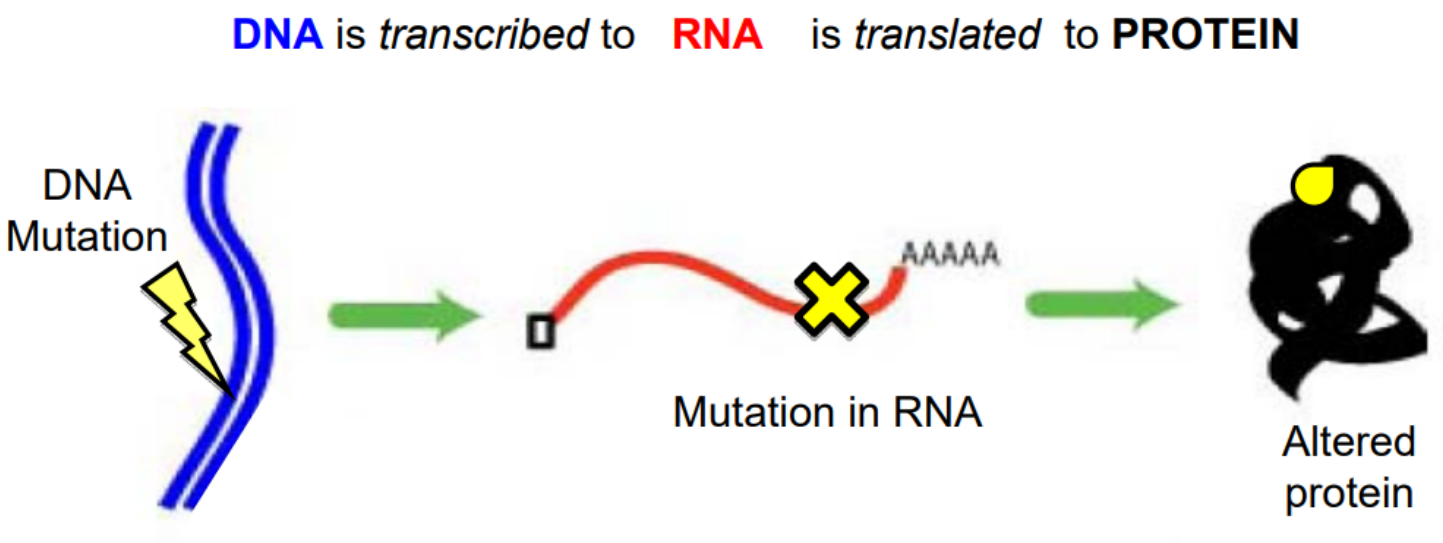

Goldman: Cecil Textbook of Medicine, 22nd ed.

# Part II: Epigenetics and Inheritance

# Epigenetics

- Epigenetics literally means “above genetics”, and refers to the heritable information in gene expression coded something other than DNA sequence.
- The study of epigenetics at the genome wide scale.

# Central Dogma

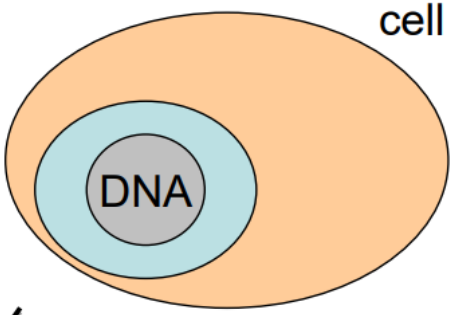


# Same DNA, different looks

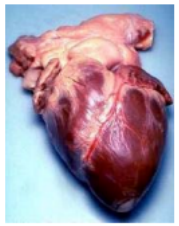
- We are made up of over 200 cell types.
- Each cell has the same DNA!
- How can they look so different?

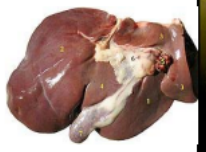
Epigenetics!

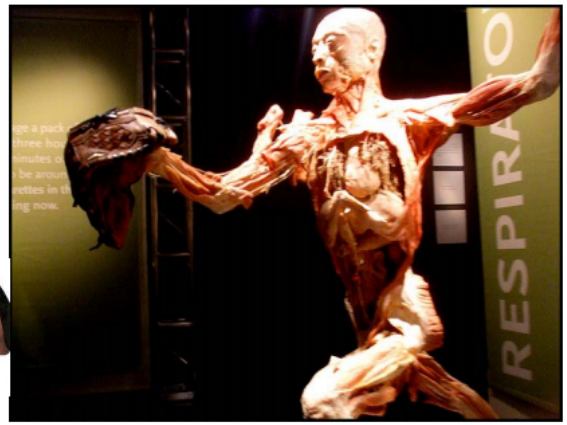
- Genes turned on or off

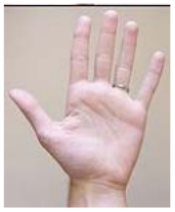


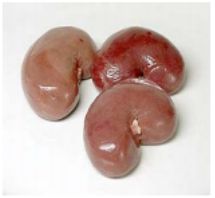

?



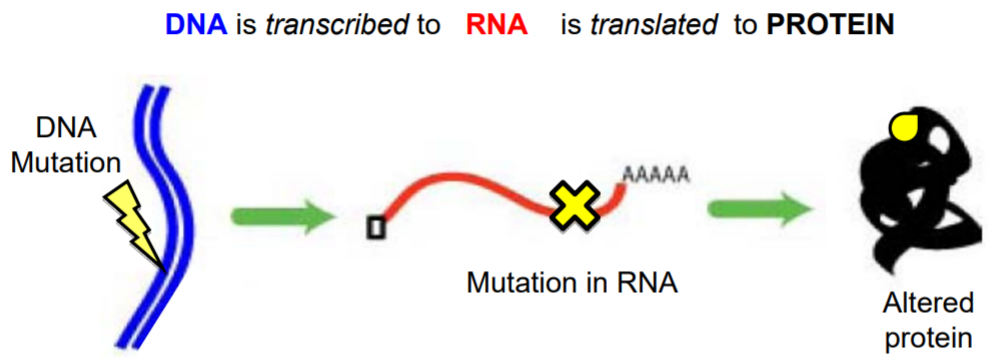


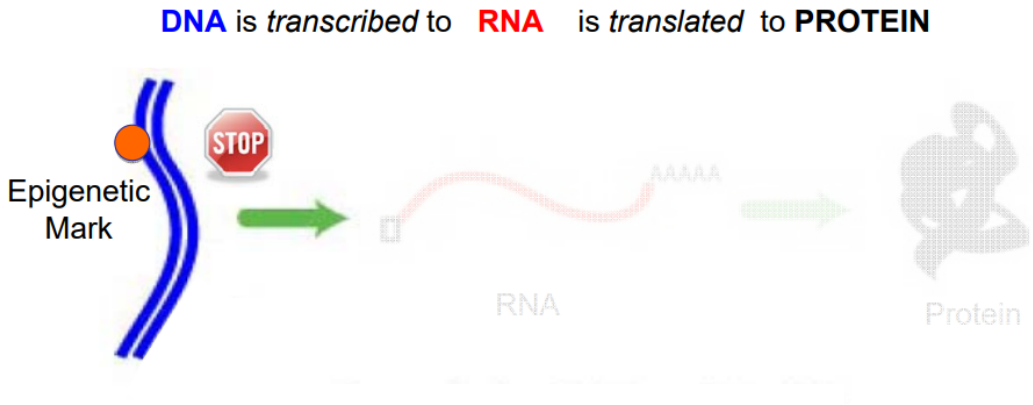




# Epigenetic marks and gene expression



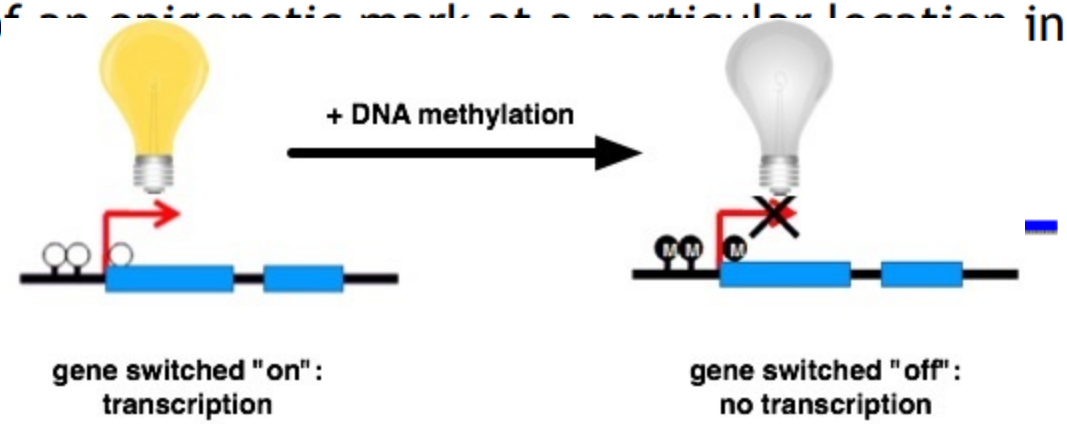
# What is Epigenetics? formal definition

- Epigenetics is the study of inherited traits caused by mechanisms other than changes in the underlying DNA sequence.
- Epi- = “above” or “upon”
- Framework for answering:
  - How can the same DNA sequence lead to different outcomes?

# Each cell has an epigenetic signature

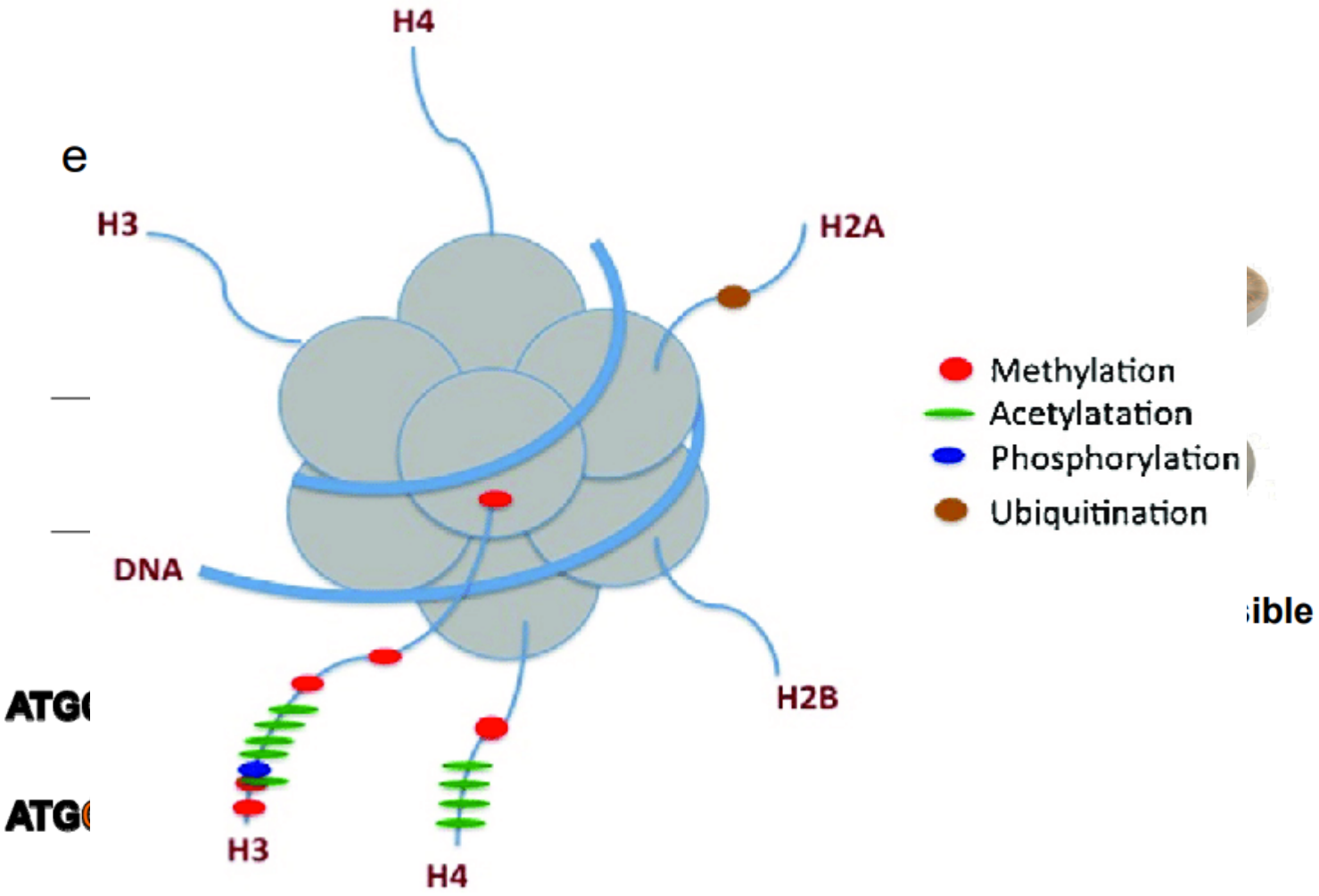
- Epigenetic state is like a barcode:


Presence of an epigenetic mark at a particular location in the DNA

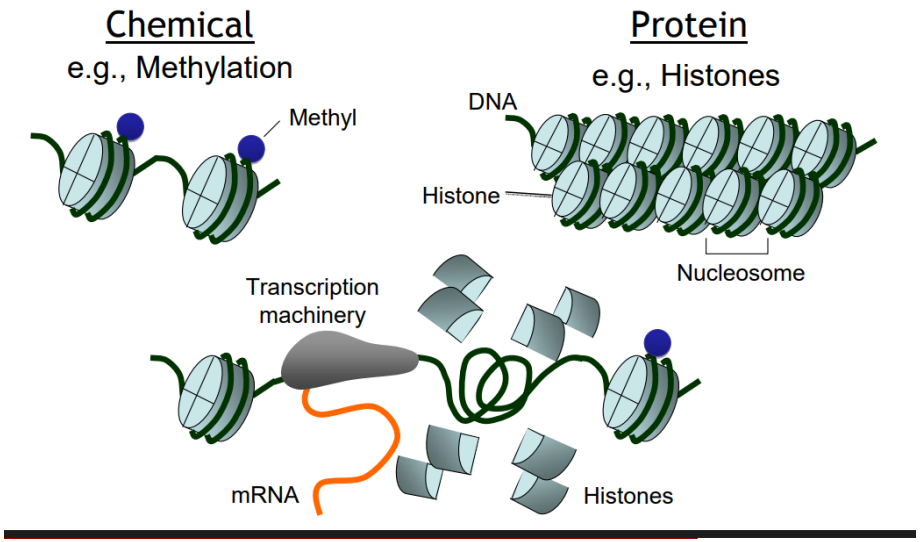


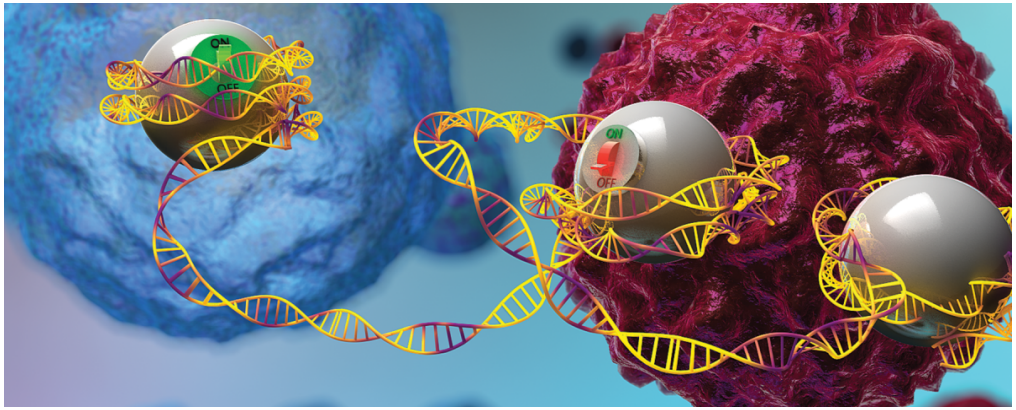

- Epigenome = complete epigenetic state of a cell

# Two types of epigenetic marks



# Combinations of Epigenetic Marks



# What do epigenetic marks do?

If DNA is like the alphabet, epigenetic marks are like the **accents** and **punctuation**

**DNA sequence**

TAG CAT ACT

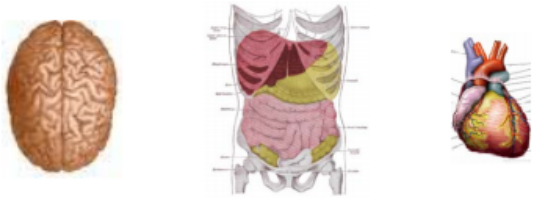
TAG! CAT? ~~ACT~~

**Epigenetic marks**

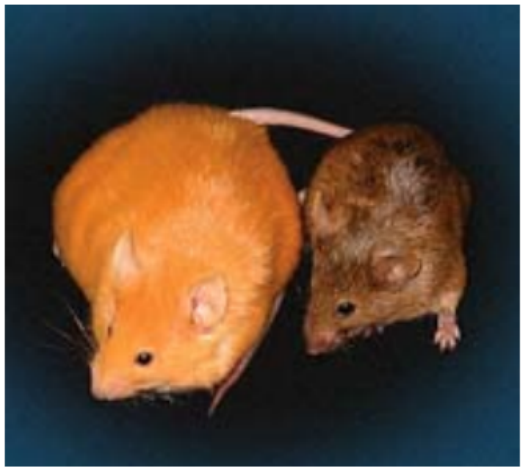
If DNA is like a book, epigenetic marks are like **sticky notes**


**Epigenetic marks tell our cells whether and how to read the genes**

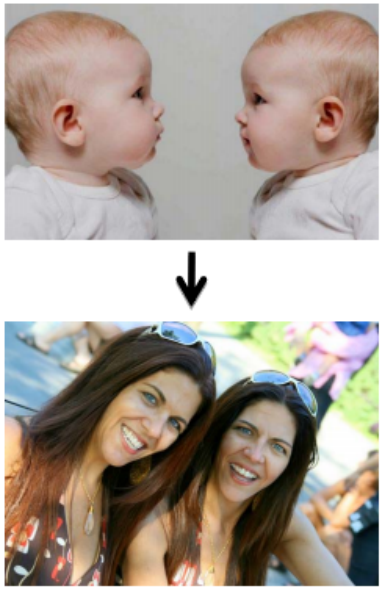
# How does Epigenetics Affect Us?



Tissues have different epigenetic states



Sisters? Actually, genetically identical.
Mothers were on different diets!



Diet and environment impact
your epigenetic state

# Sources of Epigenetic Variation

## Nature vs. Nurture

- **Nature:** Epigenetic marks can be inherited.
- **Nurture:** Epigenetic marks can be accumulated.
  - Only those in germ line will be passed down.

Nature

Nurture

The environment such as diet, toxins, vitamins, stress, affection (licking, hugging), etc. etc. can affect your epigenetics.

# Summary

- Epigenetics is the study of heritable changes in gene activity that do not involve alterations to the genetic code
- Epigenetic marks tell your genes to switch on or off
- Two types of marks: chemical (e.g., methylation) or protein (e.g., histones)
- Tissues have specific patterns of epigenetic modification

# Epigenetic inheritance

- Through epigenetic marks, environmental factors like diet, stress and prenatal nutrition can make an imprint on genes passed from on generation to the next.
- Epigenetic marks can be inherited or accumulated.

# Lamarckism and epigenetic inheritance

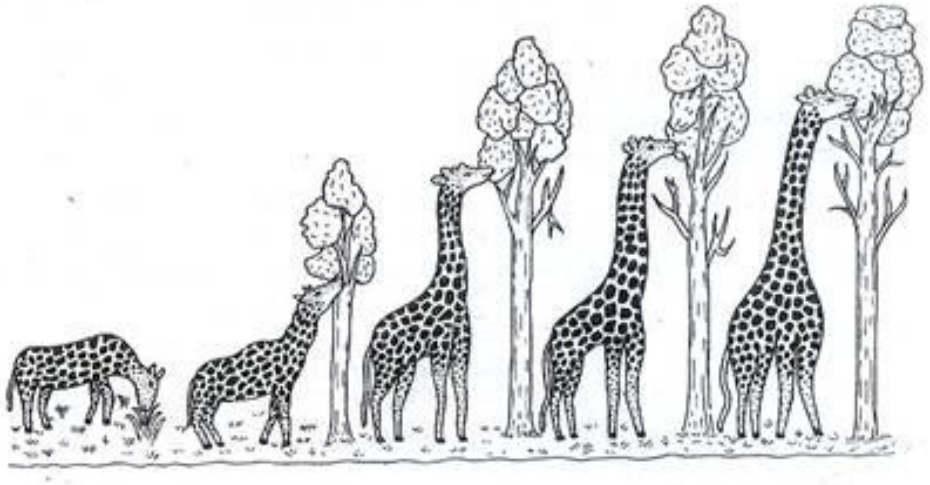
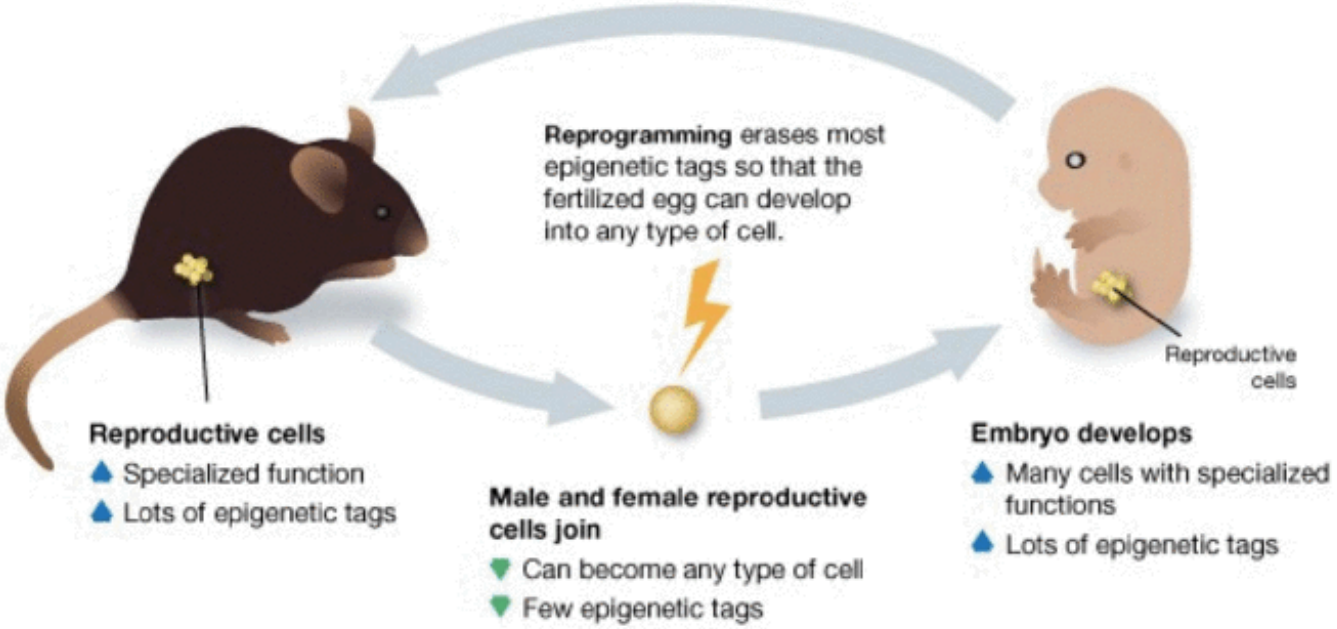


Diagram showing elongation of neck in giraffe according to Lamarck.

- Lamarck believed was that every living organism was the product of the accumulated responses that all its ancestors had made to the environments they had inhabited.
- The changes that occur to an organism during its lifetime, as a result of behavior, diet, climate or illness, could (though not necessarily would) be transmitted to any subsequent offspring.

# Epigenetic inheritance

- Most epigenetic modifications, i.e. epigenetic marks from the previous generation are typically removed in the germline, to enable totipotency of cells in early embryo.
- Occasionally the reprogramming is bypassed and some epigenetic marks get passed on, but where the DNA itself remains unchanged.




DNA methylation is typically removed during zygote formation

- Only those epigenetic modifications that are heritable *via* the gametes (germ line) → lead to **epigenetic inheritance**.

# Video learning time

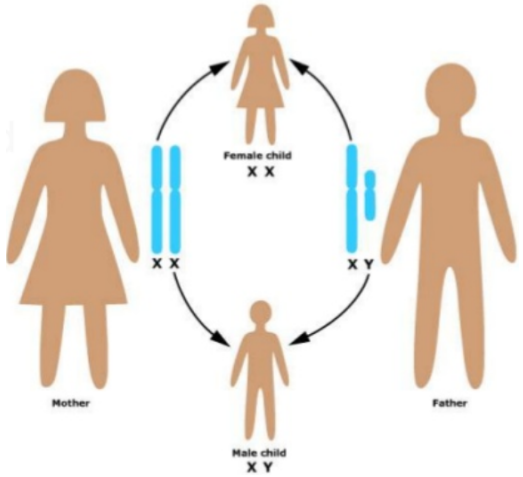
- (PBS: Is Inheritance Really All In Our Genes?)

# Prime Examples of Epigenetic Inheritance

- Genome imprinting
- X-inactivation

# X-inactivation

- The sex chromosomes in male and female mammals differ genetically – XX in females and XY in males. This leads to a potential imbalance, as more than a thousand genes on the X chromosome would be expressed in a double dose in females compared to males.
- To avoid this imbalance, which has been shown to lead to early embryonic lethality, female embryos shut down the expression of genes on one of their two X chromosomes.
- The inactivated X chromosome then condenses into a compact structure called a Barr body, and it is stably maintained in a silent state.

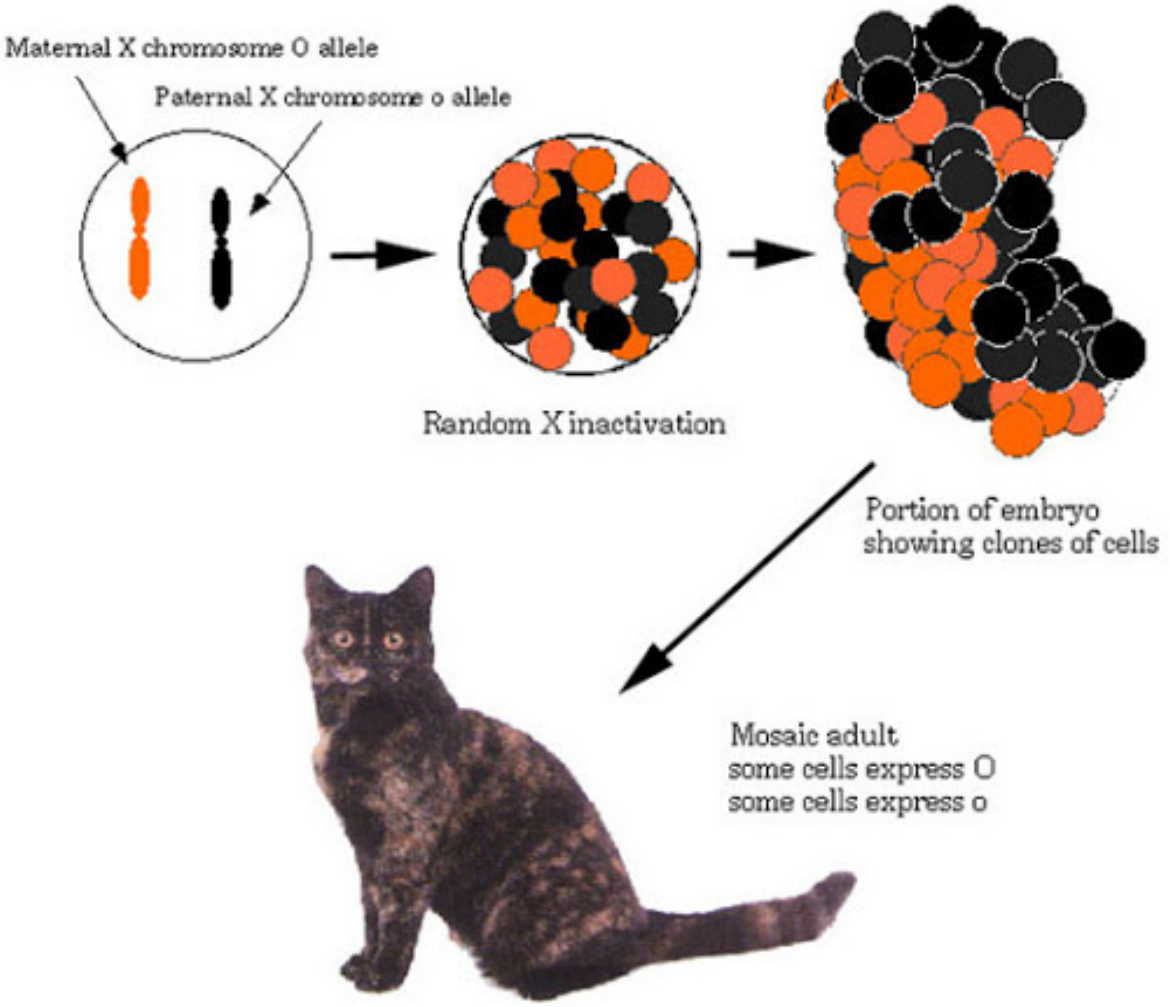


# Video learning time

- (X-inactivation and Epigenetics 5:54)

# Calico (Tortoiseshell) cats
# A prime example of X inactivation

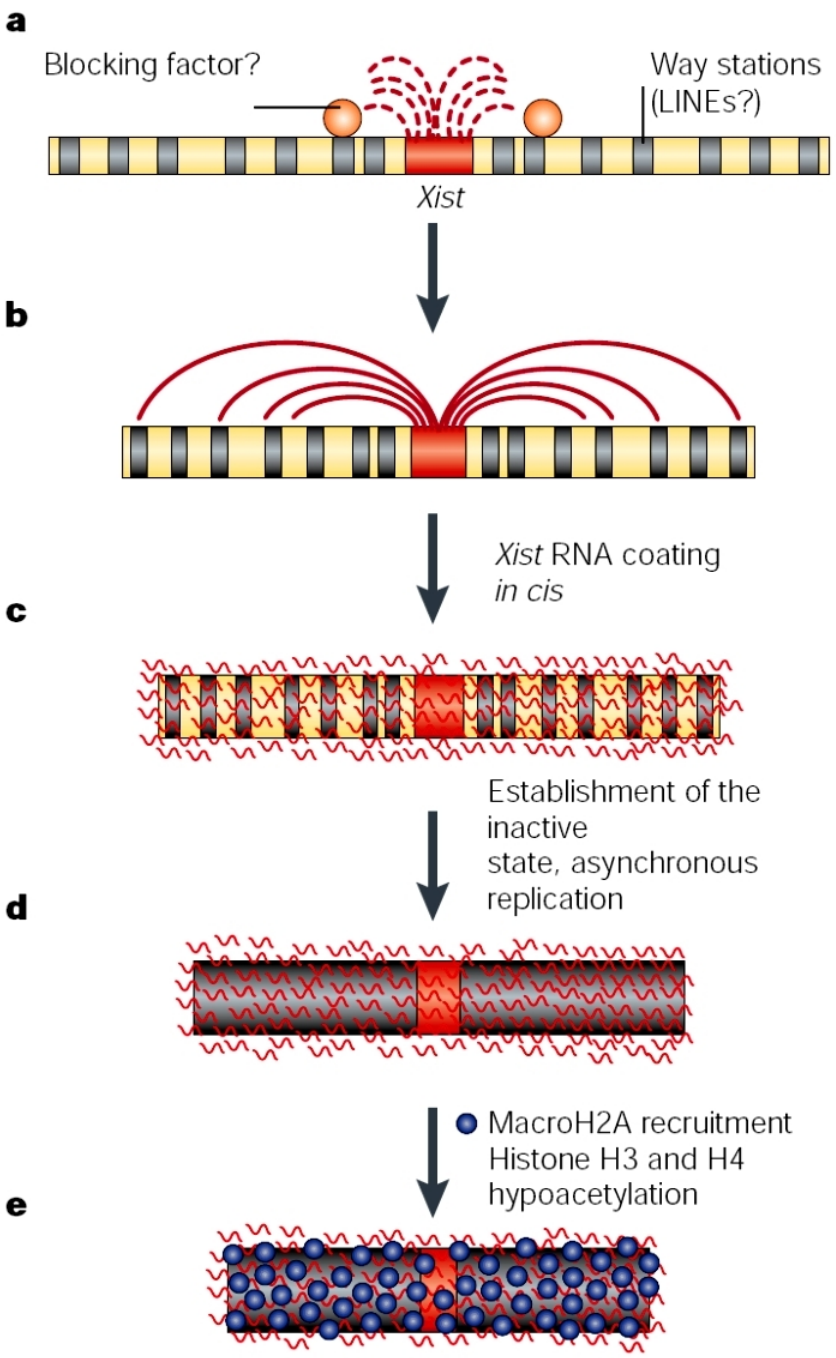
https://www.biointeractive.org/classroom-resources/x-inactivation



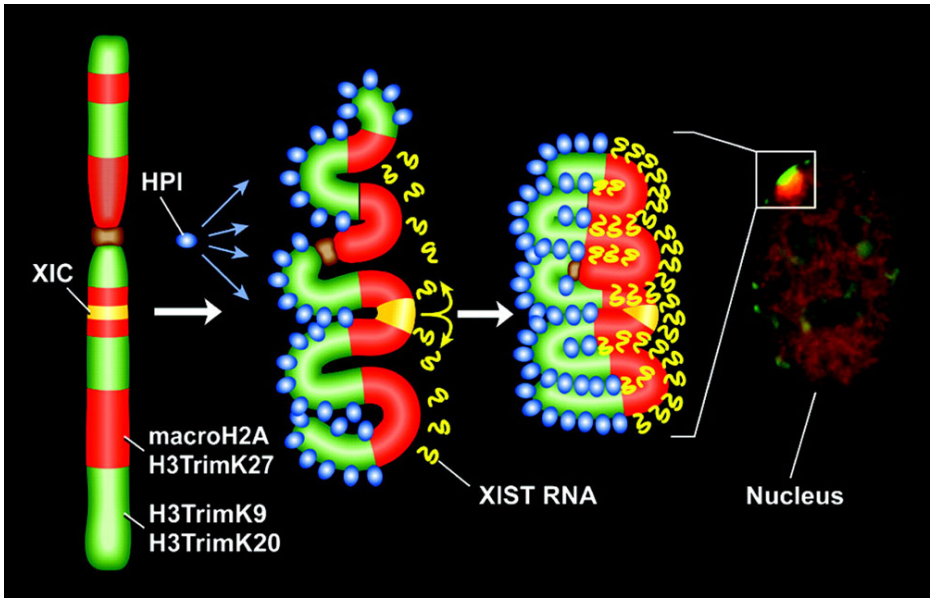

- In cats, the fur pigmentation gene is X-linked, and depending on which copy of the X chromosome each cell chooses to leave active, either an orange or black coat color results. X inactivation only occurs in cells with multiple X chromosomes, which explains why almost all calico cats are female.

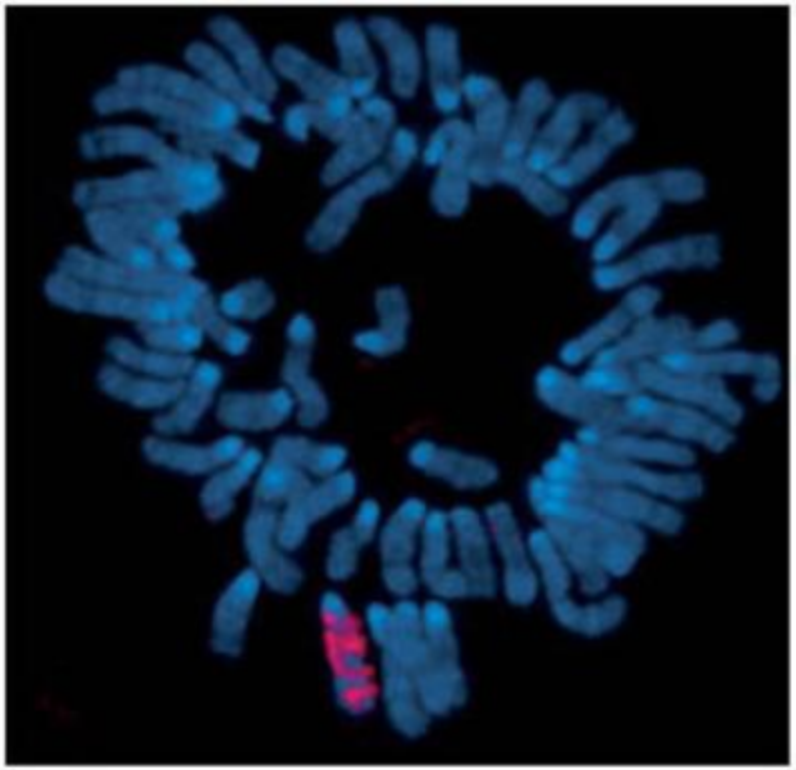
# XI mechanism

- XI starts at the XIC (X-inactivation center).
- XIST RNA is expressed from XIC and coats the inactive X chromosome
- High methylation of DNA and histone hypoacetylation are established to maintain the gene silencing.



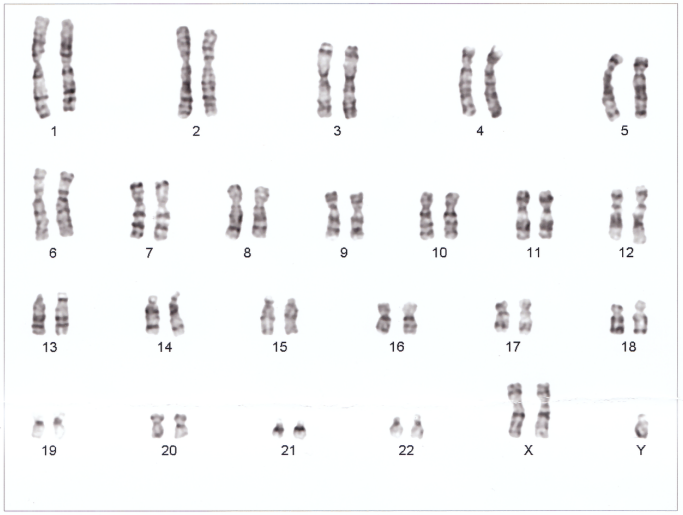
# Xist RNA coats the inactive X (FISH)



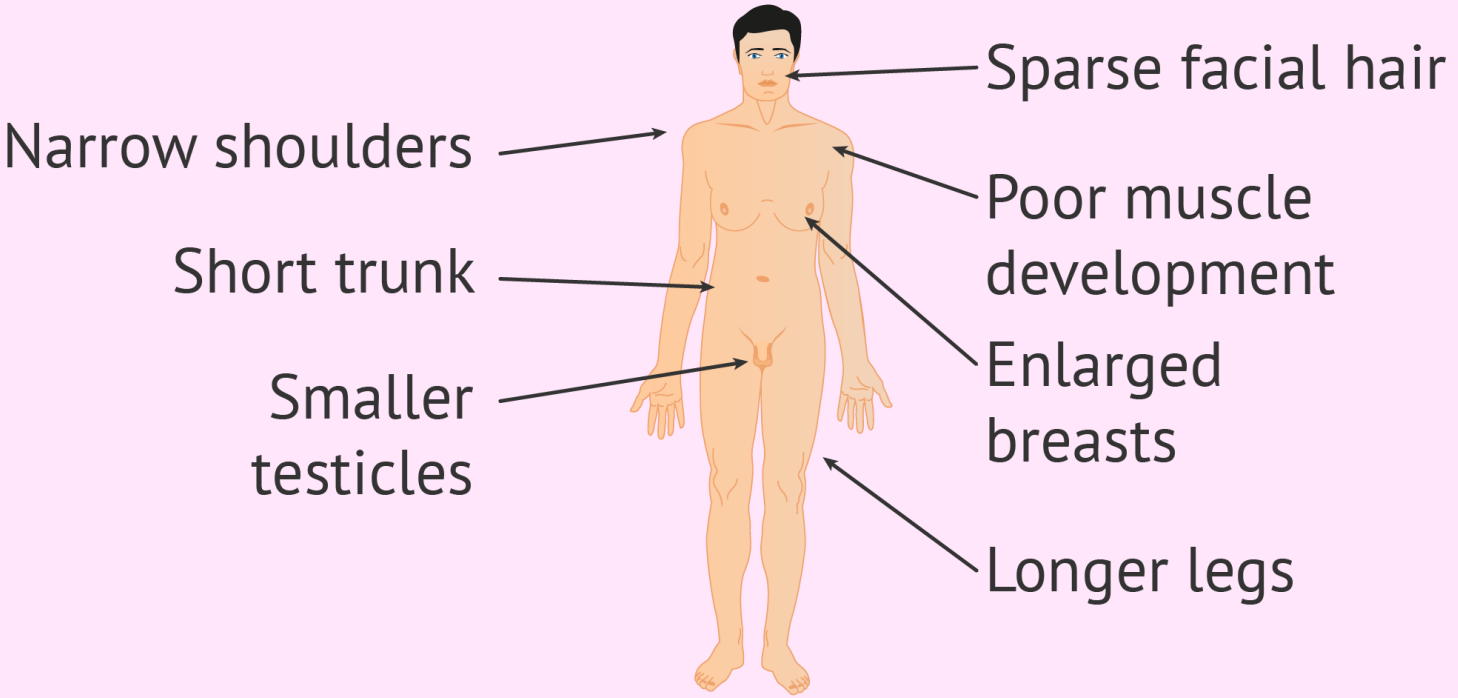



# Skewed X-inactivation

- Skewed X-chromosome inactivation occurs when the inactivation of one X chromosome is favored over the other, leading to an uneven number of cells with each chromosome inactivated. It is usually defined as one allele being found on the active X chromosome in over 75% of cells, and extreme skewing is when over 90% of cells have inactivated the same X chromosome
- Klinefelter syndrome (KS), also known as 47, XXY is the set of symptoms that result from two or more X chromosomes in males.




核型：47, XXY　　Cell No.：003



# Interesting videos

- https://www.youtube.com/watch?v=81rFpRsF80c (PBS: Is Inheritance Really All In Our Genes?)
- https://www.youtube.com/watch?v=mHak9EZjySs (X-inactivation and Epigenetics 5:54)
- https://www.youtube.com/watch?v=nTQxJRyr6GM (Imprinting)
- https://www.youtube.com/watch?v=RGoAX_4QW1U&t=44s (Genomic imprinting)