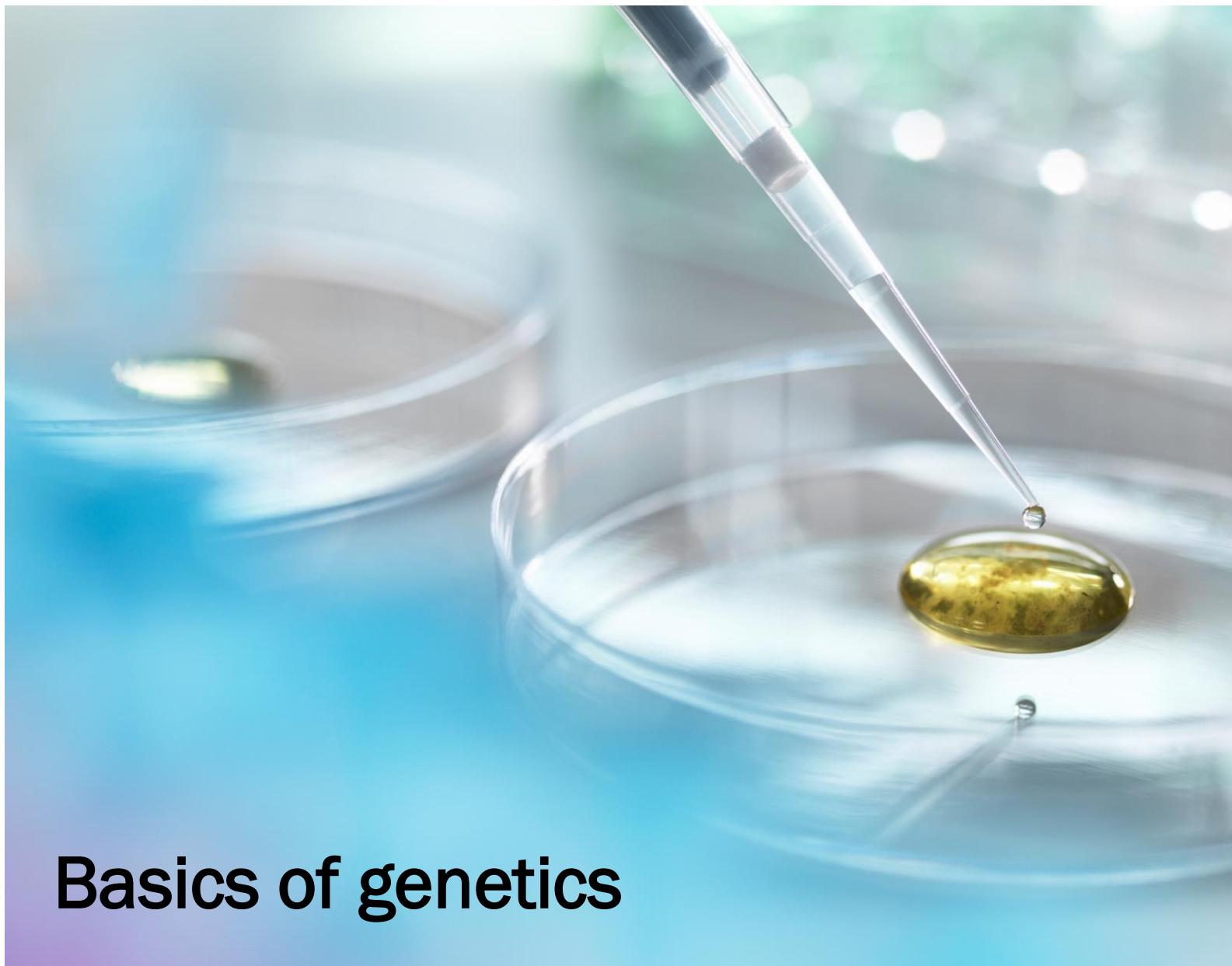


# Basics of genetics



## BIOLOGICAL FOUNDATIONS OF BIOINFORMATICS

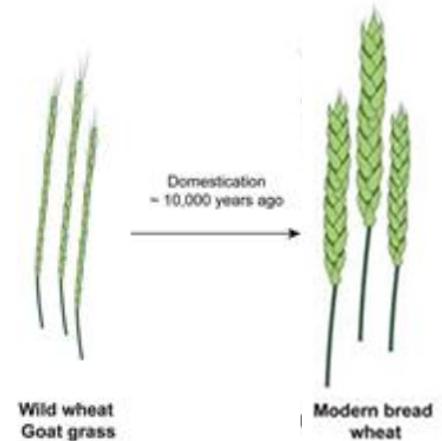


MAG. RER. NAT. ALEXANDRA HUBER  
SS 2024

# GENETICS OVERVIEW

- Cell division
- Genetic variation
- Heredity
- Mutations
- Epigenetics
- Genomics and metagenomics
- Generating data
- Genetic engineering
- Genome editing

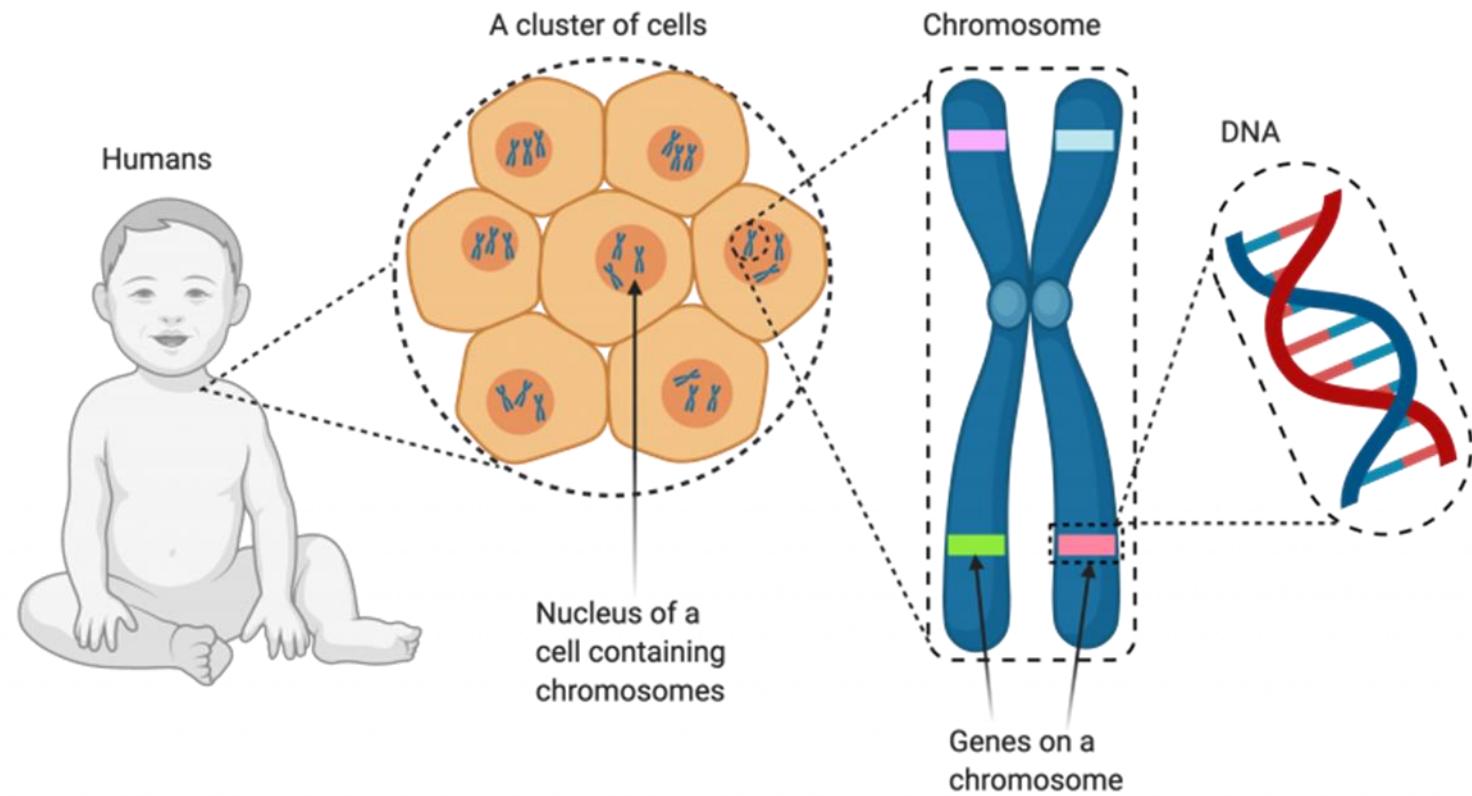
# GENETICS



refers to the **study of genes, genetic variation and heredity** in organisms.

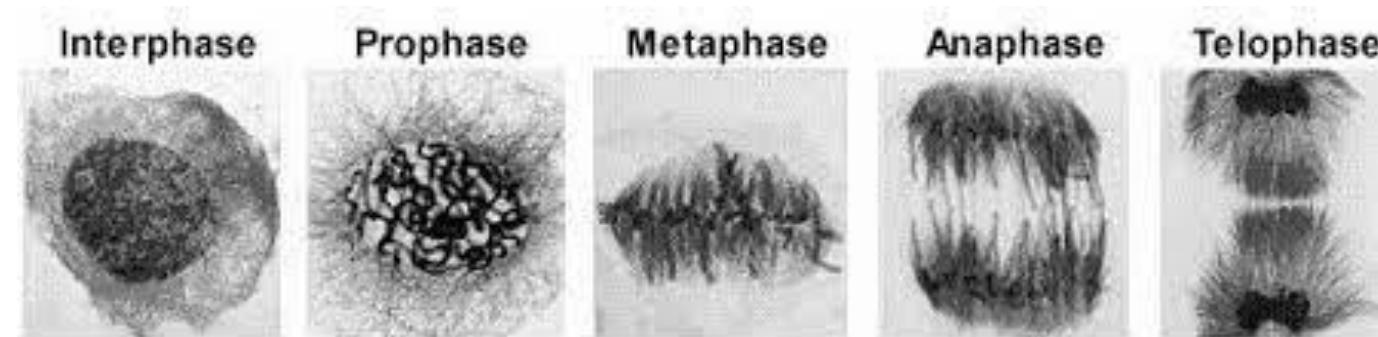
- The observation that living things inherit traits from their parents has been used since prehistoric times to improve crop plants and animals through selective breeding.
- Although heredity had been observed for millennia, **Gregor Mendel**, an Augustinian friar working in the 19th century in Brno, was the first to study genetics scientifically. Mendel discovered **patterns in the way traits are handed down** from parents to offspring. He observed that organisms (pea plants) inherit traits by way of discrete "units of inheritance" → genes.
- Modern genetics has expanded beyond inheritance to studying the function and behavior of genes.

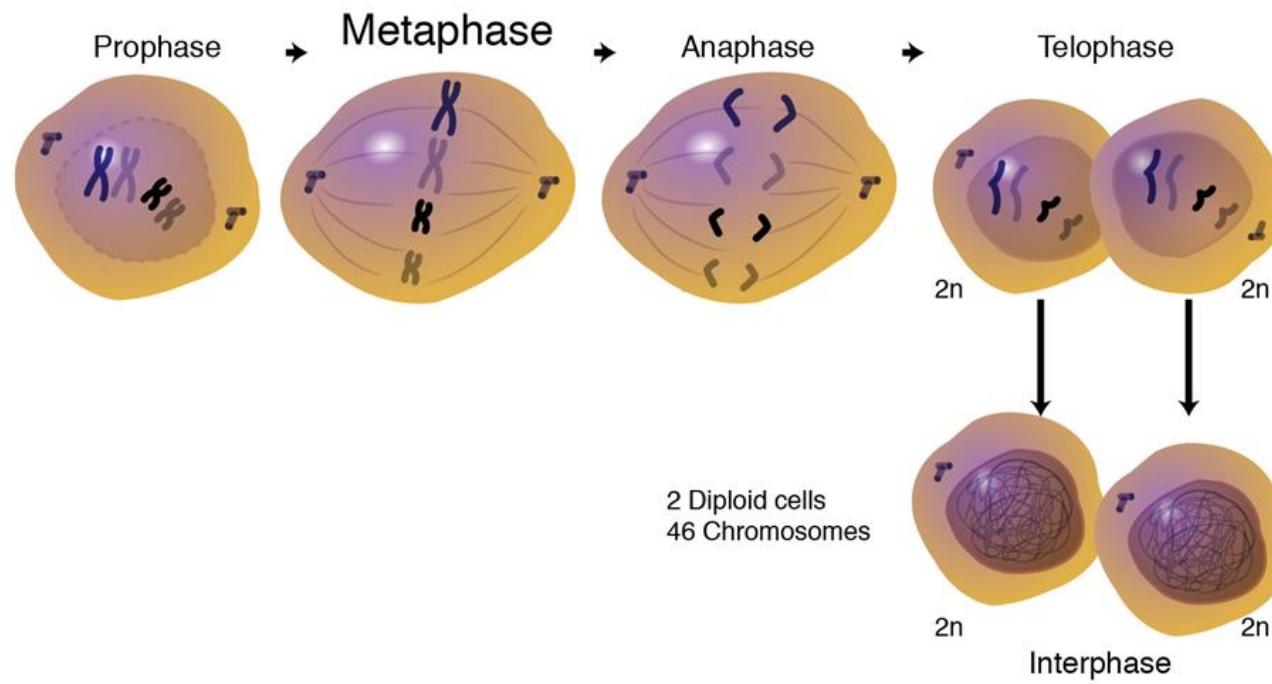
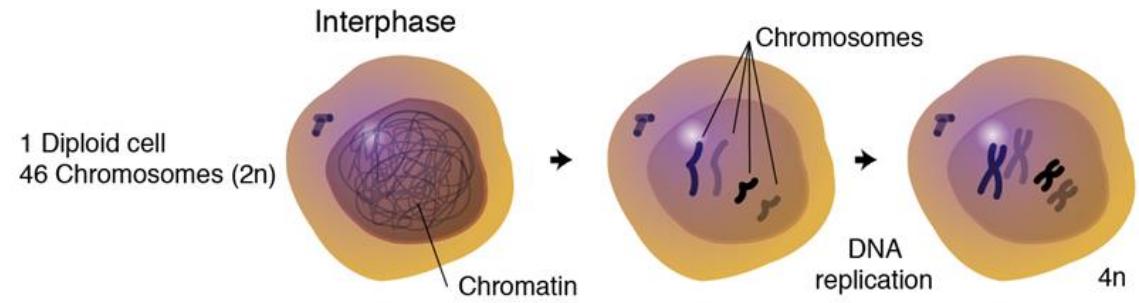
# What makes us up?



# DNA PACKAGING

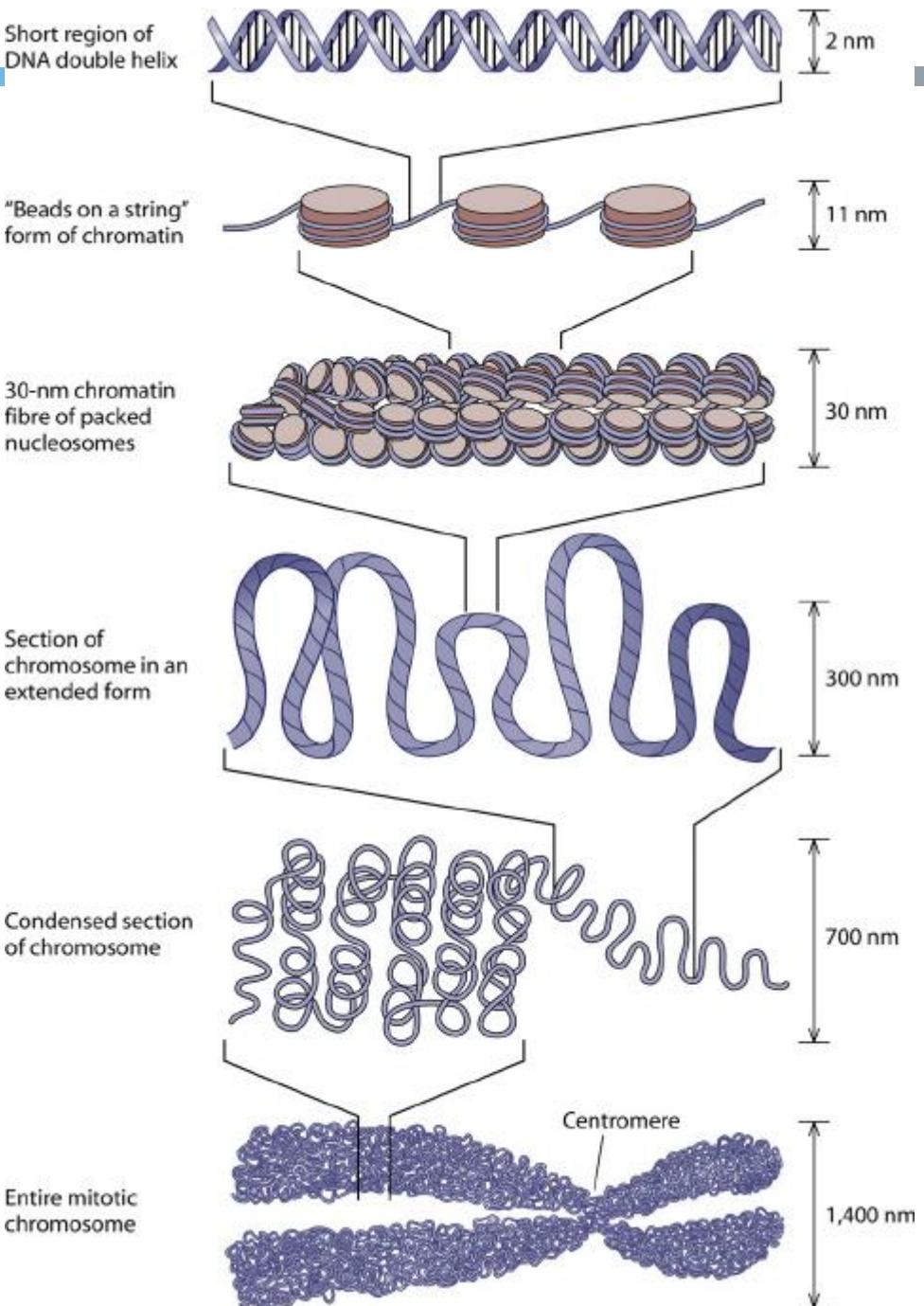
- In the nucleus of each cell, the DNA molecule is packaged into thread-like structures called chromosomes.
- Each chromosome is made up of DNA tightly coiled many times around proteins called histones that support its structure.
- Chromosomes are not visible in the cell's nucleus—not even under a microscope—when the cell is not dividing.
- However, the DNA that makes up chromosomes becomes more tightly packed during cell division and is then visible under a microscope.





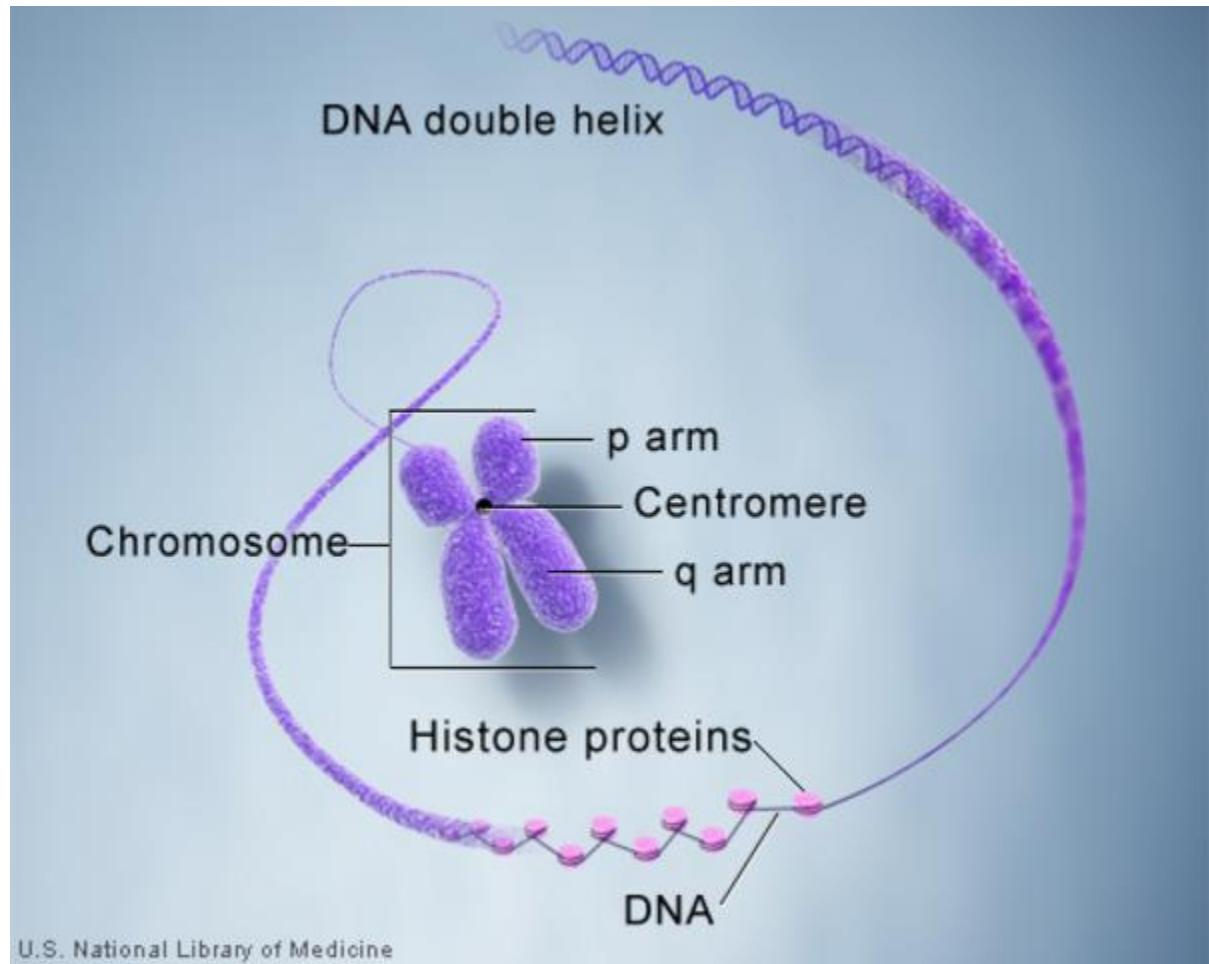
# CHROMOSOME CONDENSATION

- Chromosome condensation is the dramatic reorganisation of the long thin chromatin strands into compact short chromosomes that occurs in mitosis and meiosis.



# WHAT IS A CHROMOSOME?

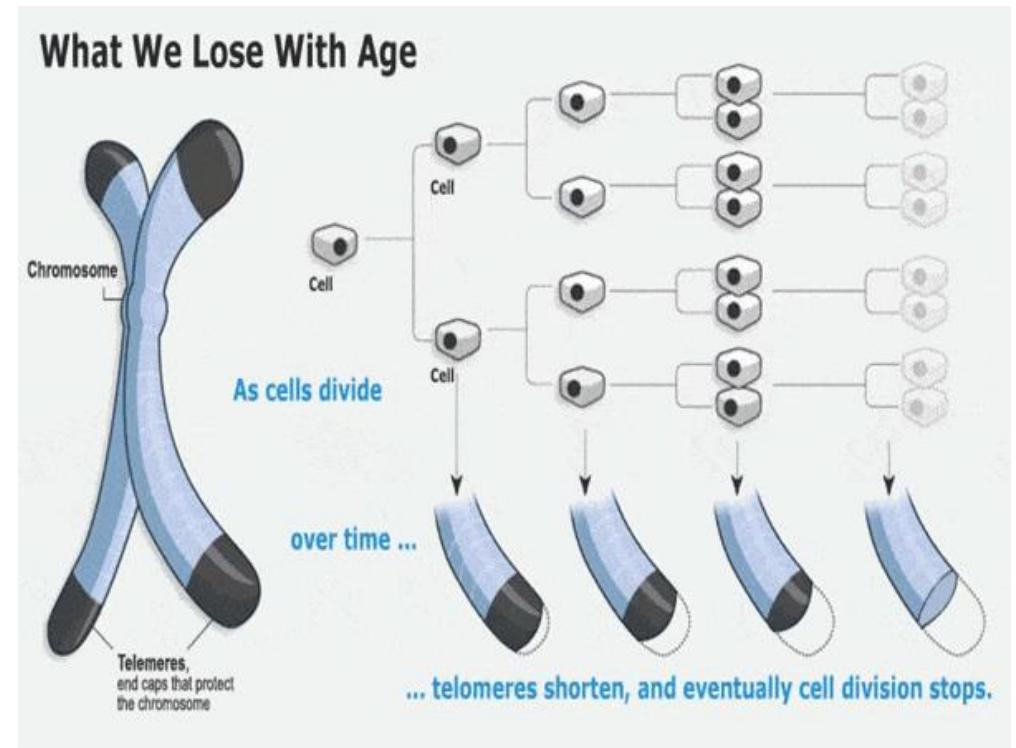
- Each chromosome has a constriction point called the centromere, which divides the chromosome into two sections, or “arms.” The short arm of the chromosome is labeled the “p arm.” The long arm of the chromosome is labeled the “q arm.” The location of the centromere on each chromosome gives the chromosome its characteristic shape and can be used to help describe the location of specific genes.
- DNA and histone proteins are packaged into structures called chromosomes.
- Chromosomes have a p arm, a q arm, and a centromere. They are made up of DNA wrapped around histone proteins.



U.S. National Library of Medicine

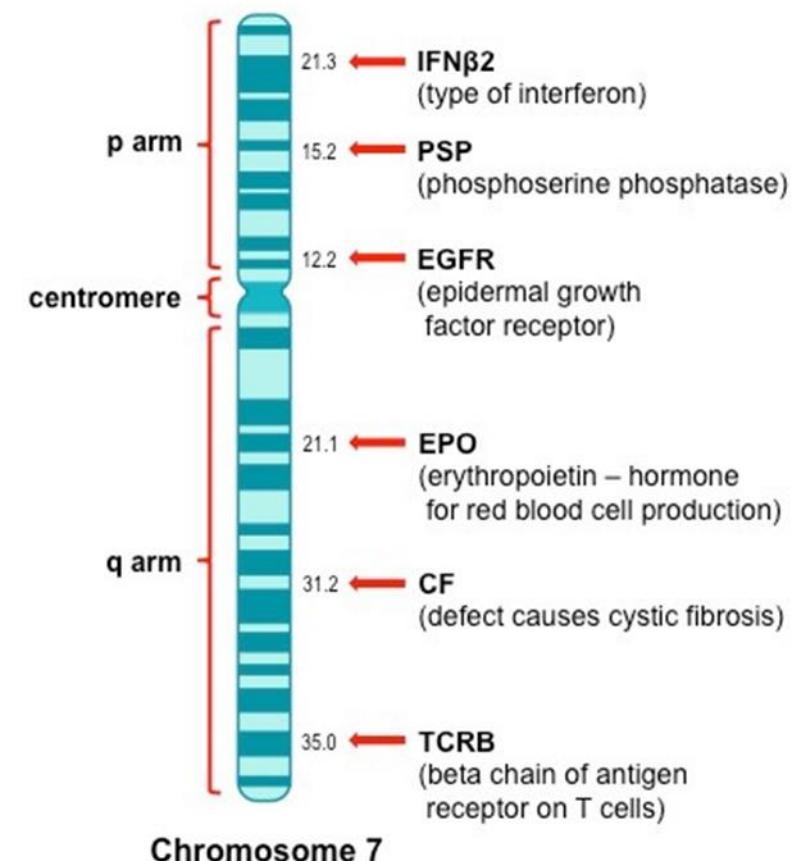
# CHROMOSOMES - TELOMERES

- Telomeres are caps at the end of each strand of DNA that protect your chromosomes, like the plastic tips at the end of shoelaces. Without this coating, shoelaces become frayed until they can no longer do their job – just as without telomeres, DNA strands become damaged and your cells can't do their job.
- Telomeres affect how your cells age. They get shorter as you age, but they can also be shortened by stress, smoking, obesity, lack of exercise and a poor diet.
- Telomeres are repetitive DNA sequences at the ends of chromosomes. Telomeres play an essential role in cellular division and are responsible for maintaining the integrity of our genes. Telomeres act like biological ticking clocks. Every time our cells divide, our telomeres get shorter. It only takes a few critically short telomeres to send a cell into a crisis state where it either stops functioning properly (senescence) or dies.



# HOW DO GENETICISTS INDICATE THE LOCATION OF A GENE?

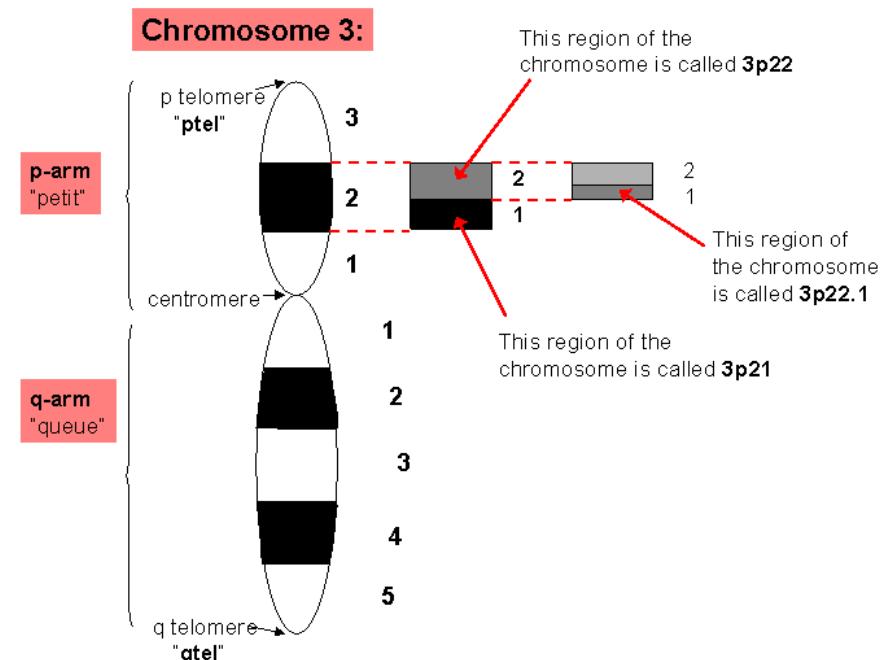
- Geneticists use maps to describe the location of a particular gene on a chromosome.
- One type of map uses the **cytogenetic location** to describe a gene's position. The cytogenetic location is based on a distinctive pattern of light and dark bands created when chromosomes are stained with certain chemicals.
- Another type of map uses the molecular location, which is a precise description of a gene's position on a chromosome. The **molecular location** is based on the sequence of DNA building blocks (nucleotides) that make up the chromosome.



# CYTOGENETIC LOCATION

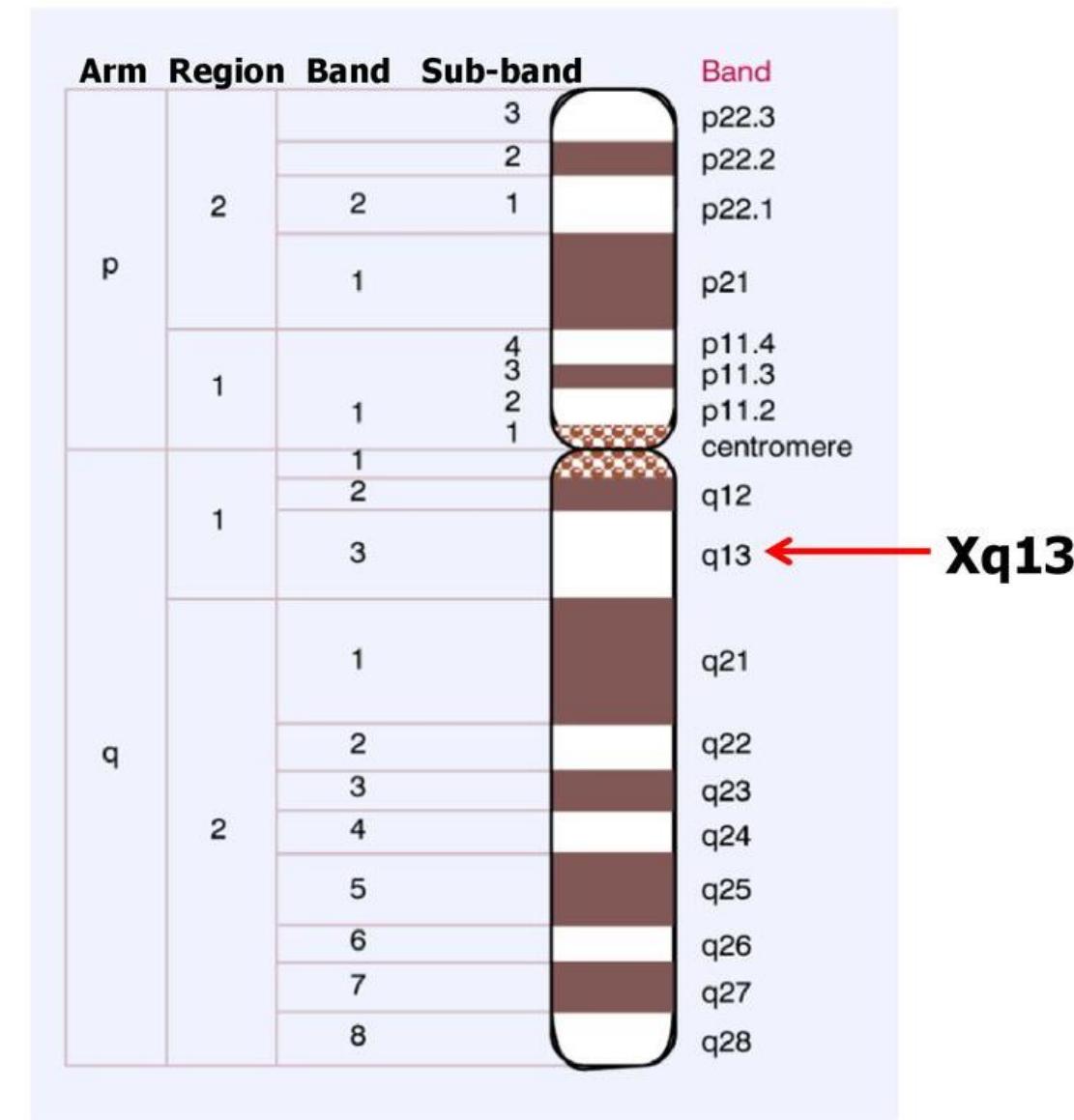
- the location describes the position of a particular band on a stained chromosome: e. g. 17q12
- It can also be written as a range of bands, if less is known about the exact location: e. g. 17q12-q21
- The combination of numbers and letters provide a gene's "address" on a chromosome. This address is made up of several parts:
  - The chromosome on which the gene can be found. The first number or letter used to describe a gene's location represents the chromosome. Chromosomes 1 through 22 (the autosomes) are designated by their chromosome number. The sex chromosomes are designated by X or Y.

## Cytogenetic Banding Nomenclature



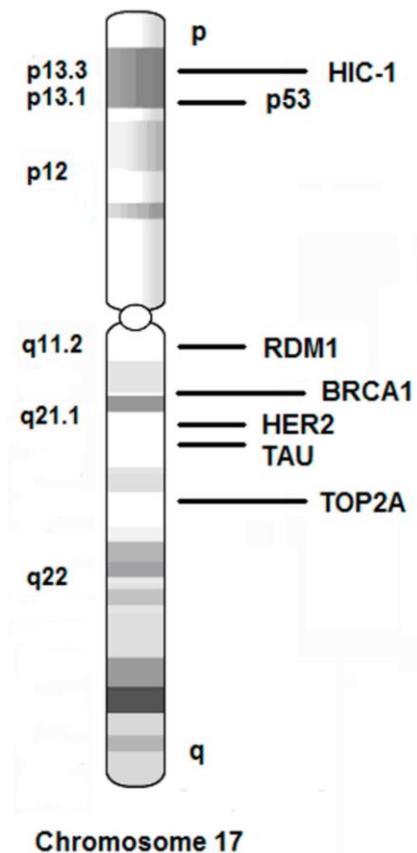
# CYTOGENETIC LOCATION

- The **arm of the chromosome**. Each chromosome is divided into two sections (arms) based on the location of a narrowing (constriction) called the centromere. By convention, the shorter arm is called p, and the longer arm is called q. The chromosome arm is the second part of the gene's address. For example, 5q is the long arm of chromosome 5, and Xp is the short arm of the X chromosome.
- The **position of the gene on the p or q arm**. The position is usually designated by two digits (representing a region and a band), which are sometimes followed by a decimal point and one or more additional digits (representing sub-bands within a light or dark area). The number indicating the gene position increases with distance from the centromere. For example: 14q21 represents position 21 on the long arm of chromosome 14. 14q21 is closer to the centromere than 14q22.
- Sometimes, the abbreviations “cen” or “ter” are also used to describe a gene's cytogenetic location. “Cen” indicates that the gene is very close to the centromere. For example, 16pcen refers to the short arm of chromosome 16 near the centromere. “Ter” stands for terminus, which indicates that the gene is very close to the end of the p or q arm. For example, 14qter refers to the tip of the long arm, or the very end, of chromosome 14.



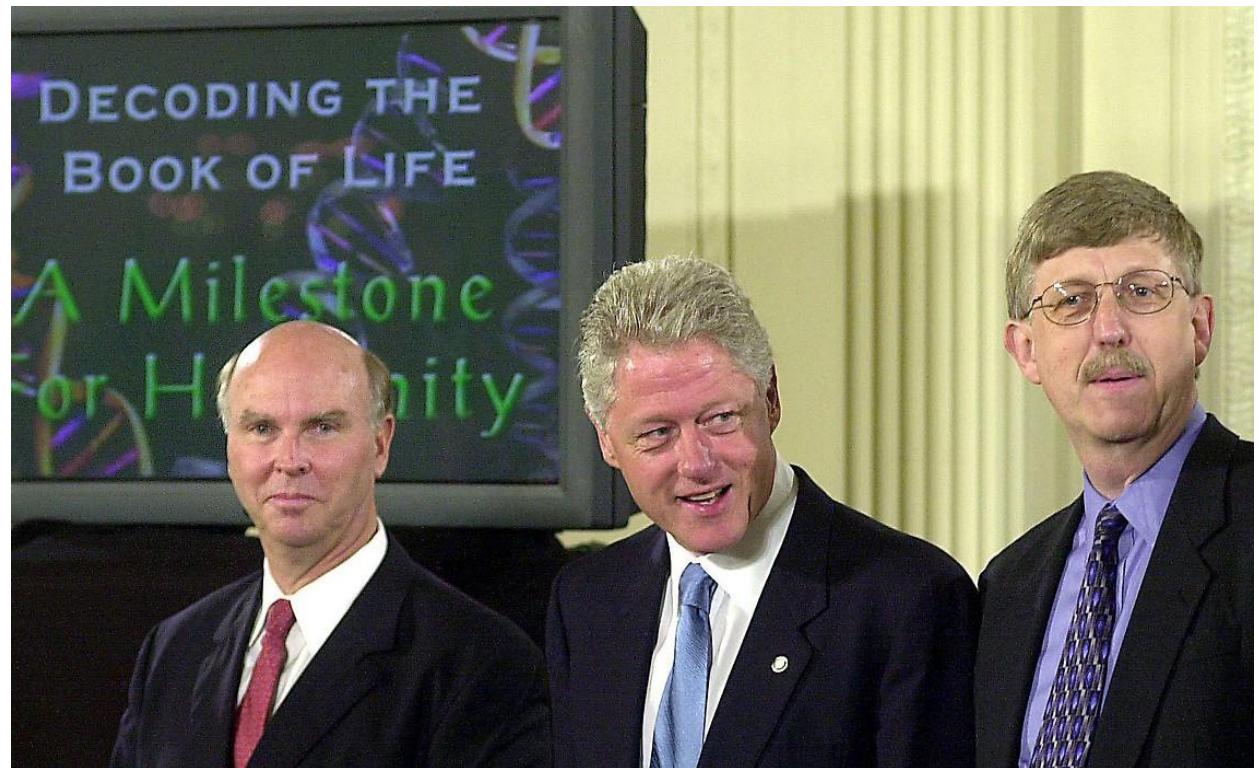
# MOLECULAR LOCATION

- The Human Genome Project, an international research effort completed in 2003, determined the sequence of nucleotides for each human chromosome.
- This sequence information allows researchers to provide a more specific address than the cytogenetic location for many genes. A gene's **molecular address** pinpoints the location of that gene in terms of **nucleotides**. It describes the gene's precise position on a chromosome and indicates the size of the gene. Knowing the molecular location also allows researchers to determine exactly how far a gene is from other genes on the same chromosome.
- The HUGO Gene Nomenclature Committee (HGNC) designates an **official name and symbol** (an abbreviation of the name) for each known **human gene**. The Committee has named more than 19,000 of the estimated 20,000 to 25,000 protein-coding genes in the human genome. In general, symbols for genes are italicized (e.g., *IGF1*), whereas symbols for proteins are not italicized (e.g., IGF1). *IGF1* stands for Insulin-like growth factor 1 (IGF-1), a hormone similar in molecular structure to insulin which plays an important role in childhood growth.



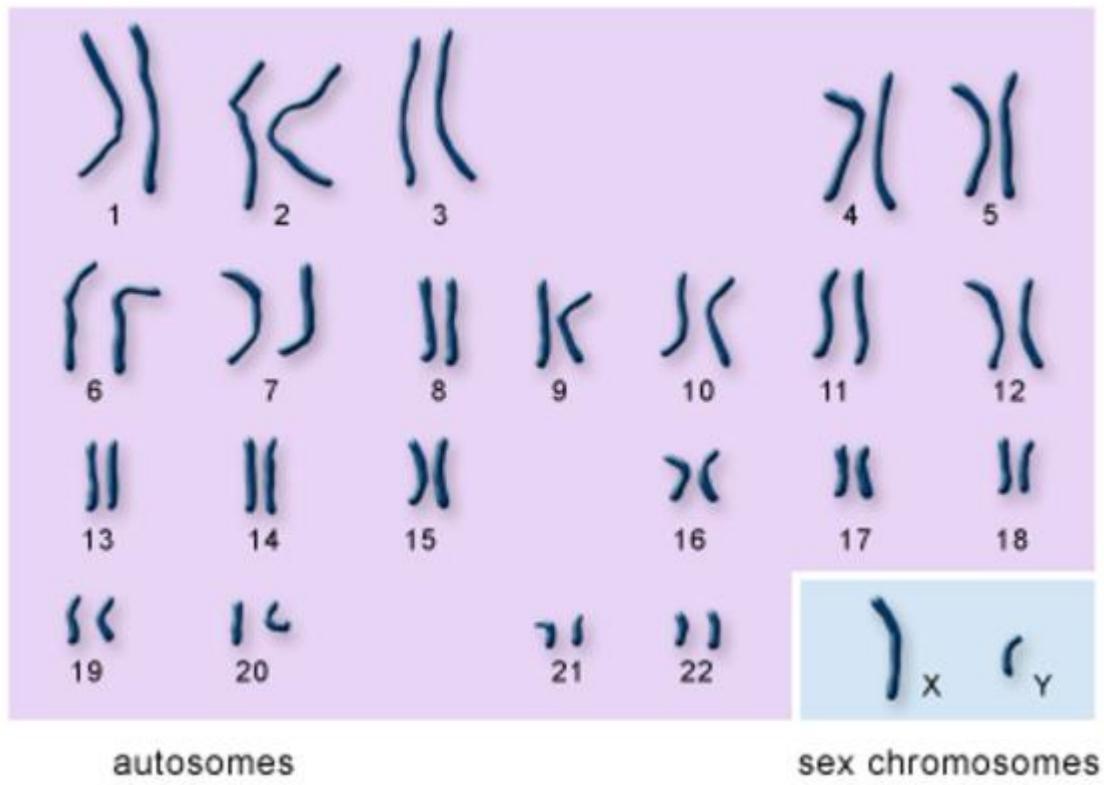
# HUMAN GENOME PROJECT

- An international research effort called the Human Genome Project (1984-2003; it remains the world's largest collaborative biological project), which worked to determine the sequence of the human genome and identify the genes that it contains, estimated that humans have between 20,000 and 25,000 genes.
- Every person has two copies of each gene, one inherited from each parent. Most genes are the same in all people, but a small number of genes (less than 1 percent of the total) are slightly different between people.
- Alleles are forms of the same gene with small differences in their sequence of DNA bases. These small differences contribute to each person's unique physical features.



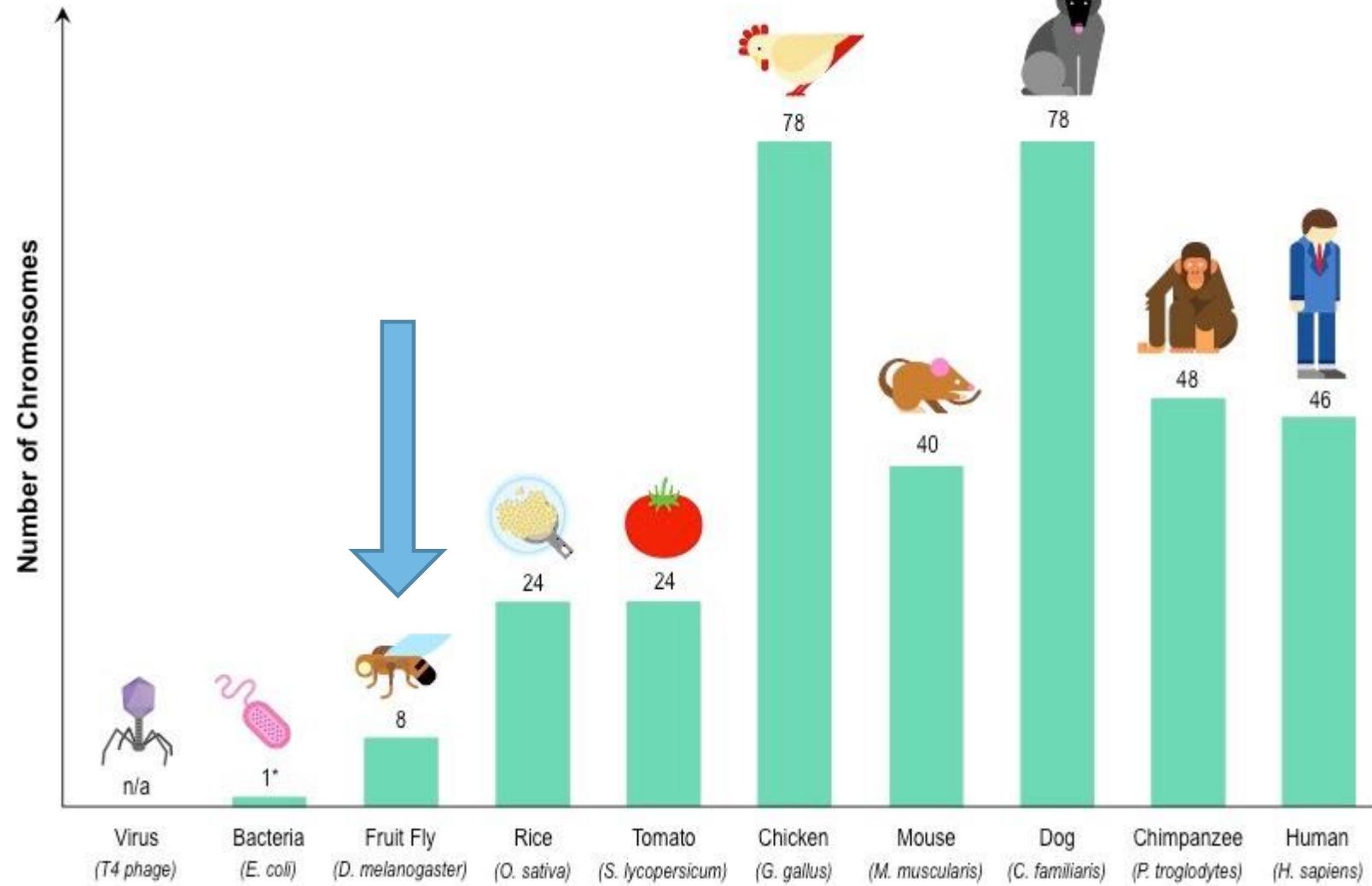
# HOW MANY CHROMOSOMES DO PEOPLE HAVE?

- In humans, each cell normally contains 23 pairs of chromosomes, for a total of 46. Twenty-two of these pairs, called autosomes, look the same in both males and females. The 23rd pair, the sex chromosomes, differ between males and females. Females have two copies of the X chromosome, while males have one X and one Y chromosome.
- The 22 autosomes are numbered by size. The other two chromosomes, X and Y, are the sex chromosomes. This picture of the human chromosomes lined up in pairs is called a karyotype.



U.S. National Library of Medicine

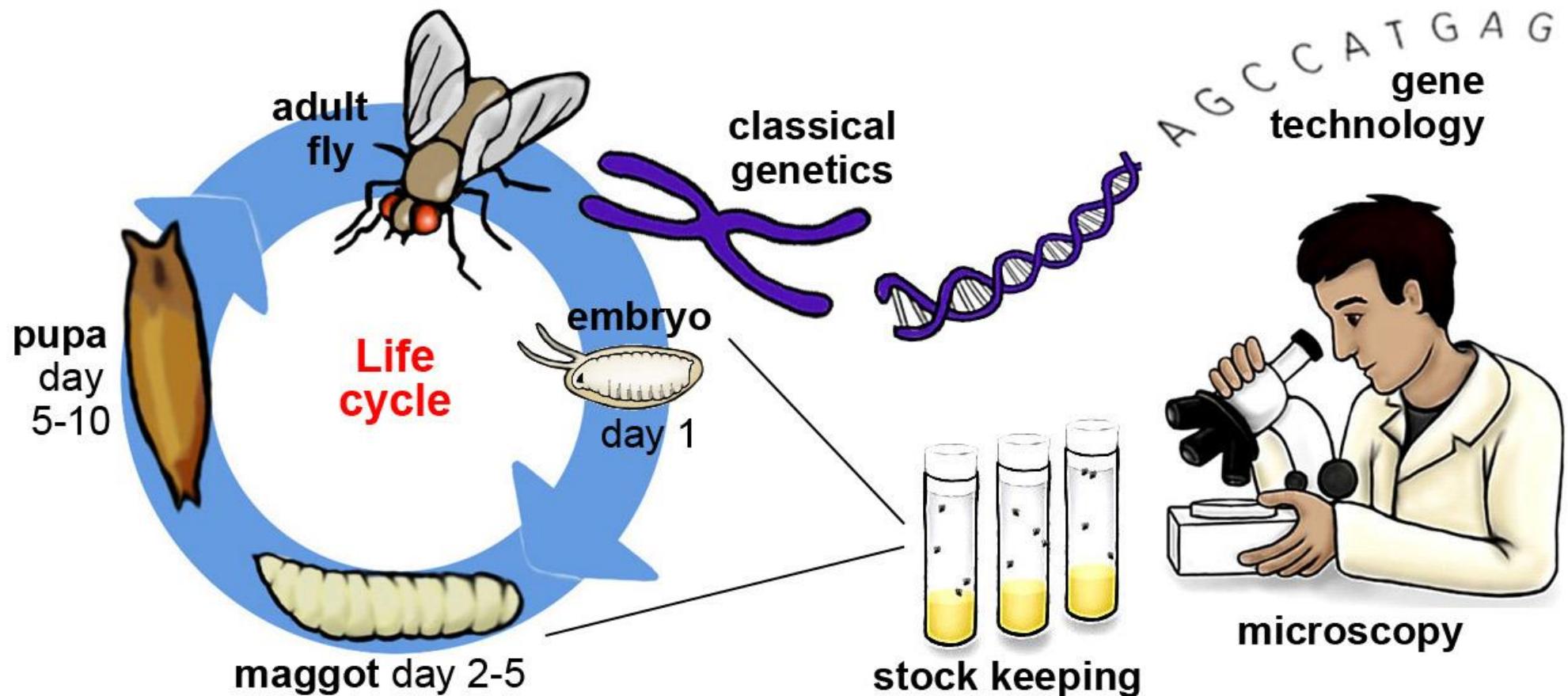
Different species have different numbers of chromosomes



## MODEL ORGANISM – DROSOPHILA MELANOGASTER



- more commonly known as the fruit fly or vinegar fly, has been used as a model for biological research for over 100 years.
- To date, Drosophila is the conceptually **best understood animal organism** in the biomedical sciences.
- Many aspects of modern medicine are based on foundations laid through fly research.
- Moreover, work in flies is advancing our understanding of the importance of nutrition for biological performance and longevity, and the fundamental biology of stem cells and cancer (cell division, cell communication).
- Drosophila is helping us to understand how certain genetic defects can lead to a range of diseases.



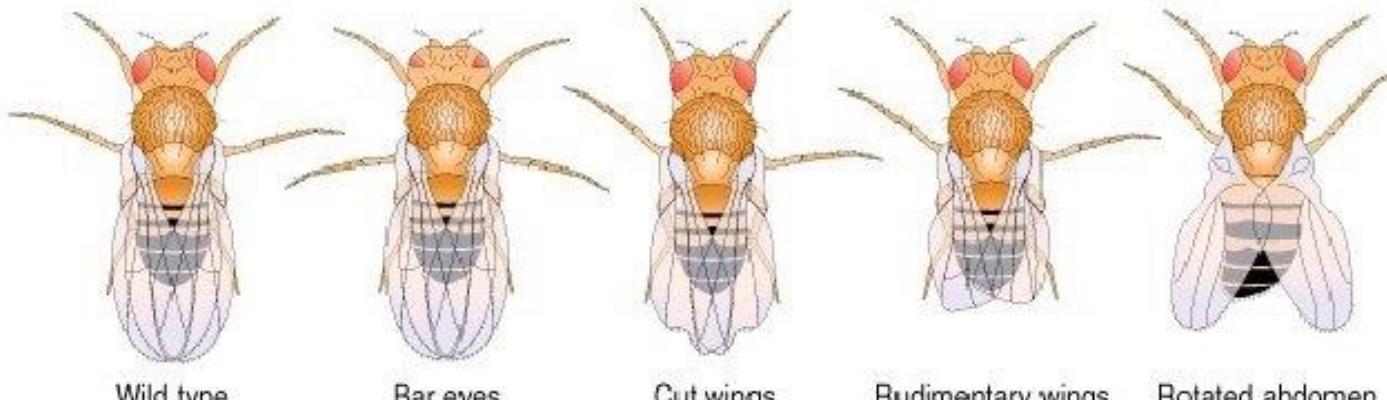
---

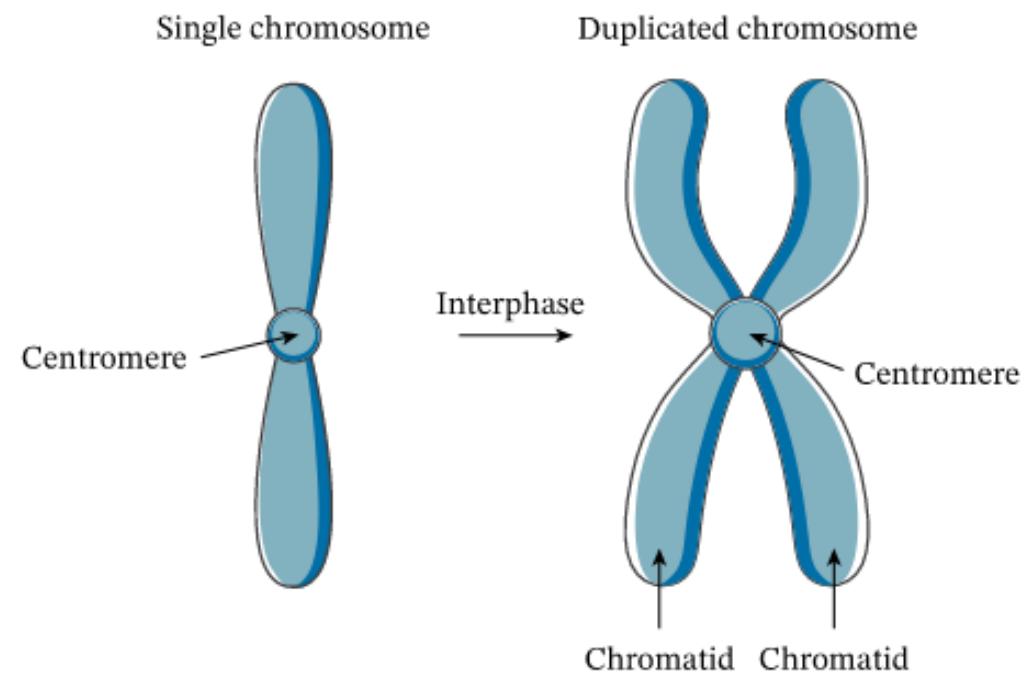
## **MODEL ORGANISM FOR GENETIC STUDIES – DROSOPHILA MELANOGASTER**

- Is easily cultured in the lab
- Has short generation time
- Can produce many offspring
- Has only few chromosomes easily visible under a microscope
- Many mutants have been created (the normal fly is called ‘wild type’ and any fly exhibiting a phenotypic mutation is called a ‘mutant’).



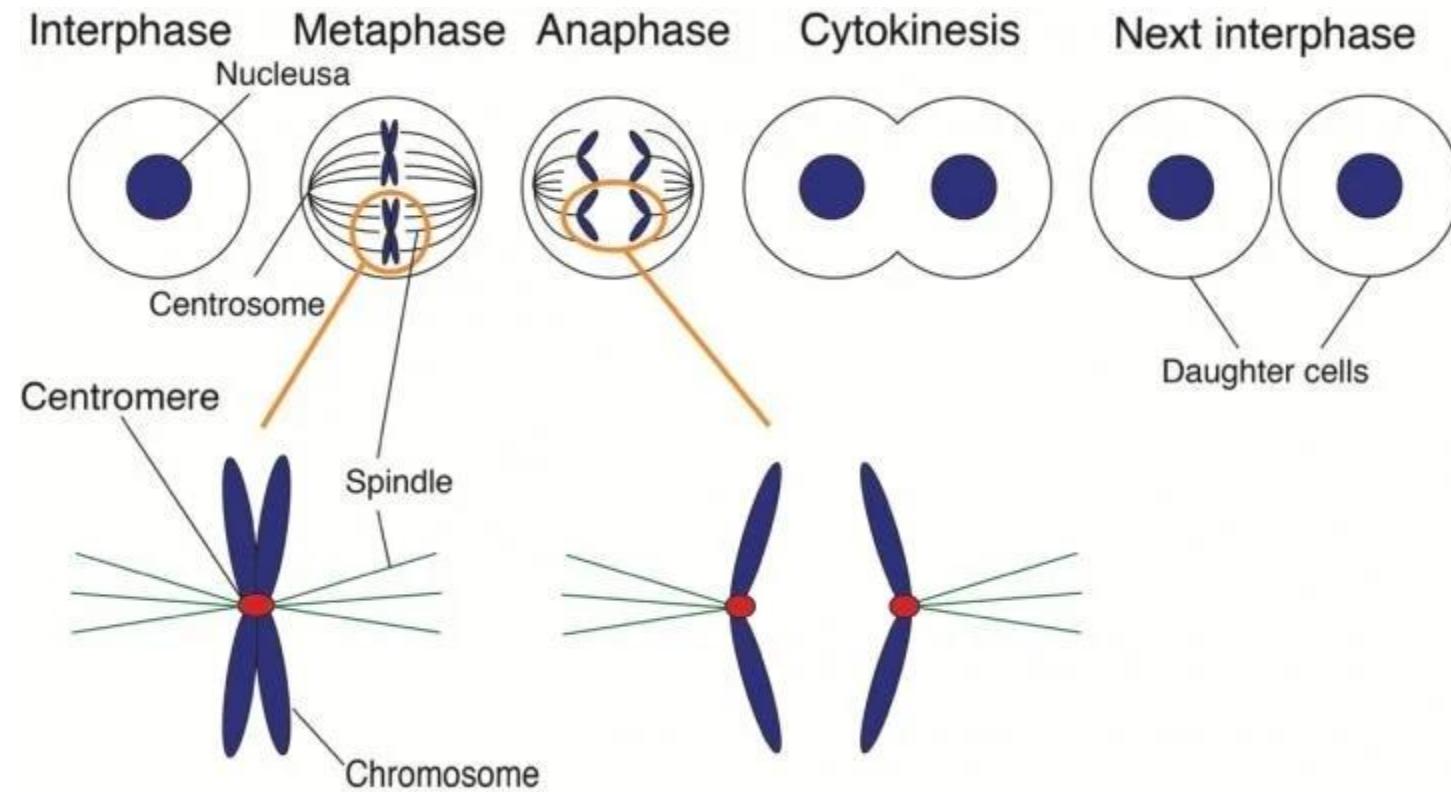
# Drosophila Melanogaster mutations

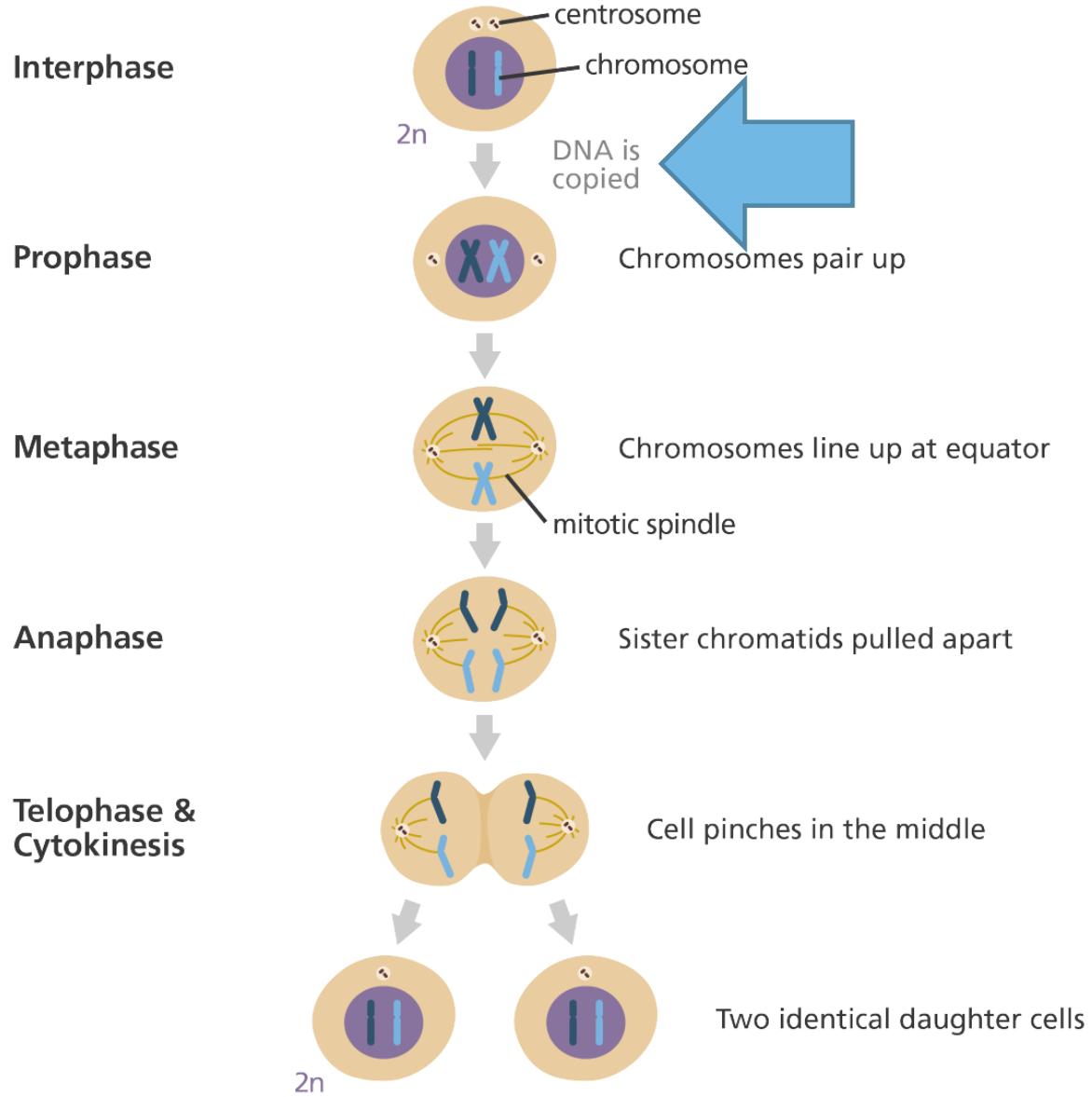


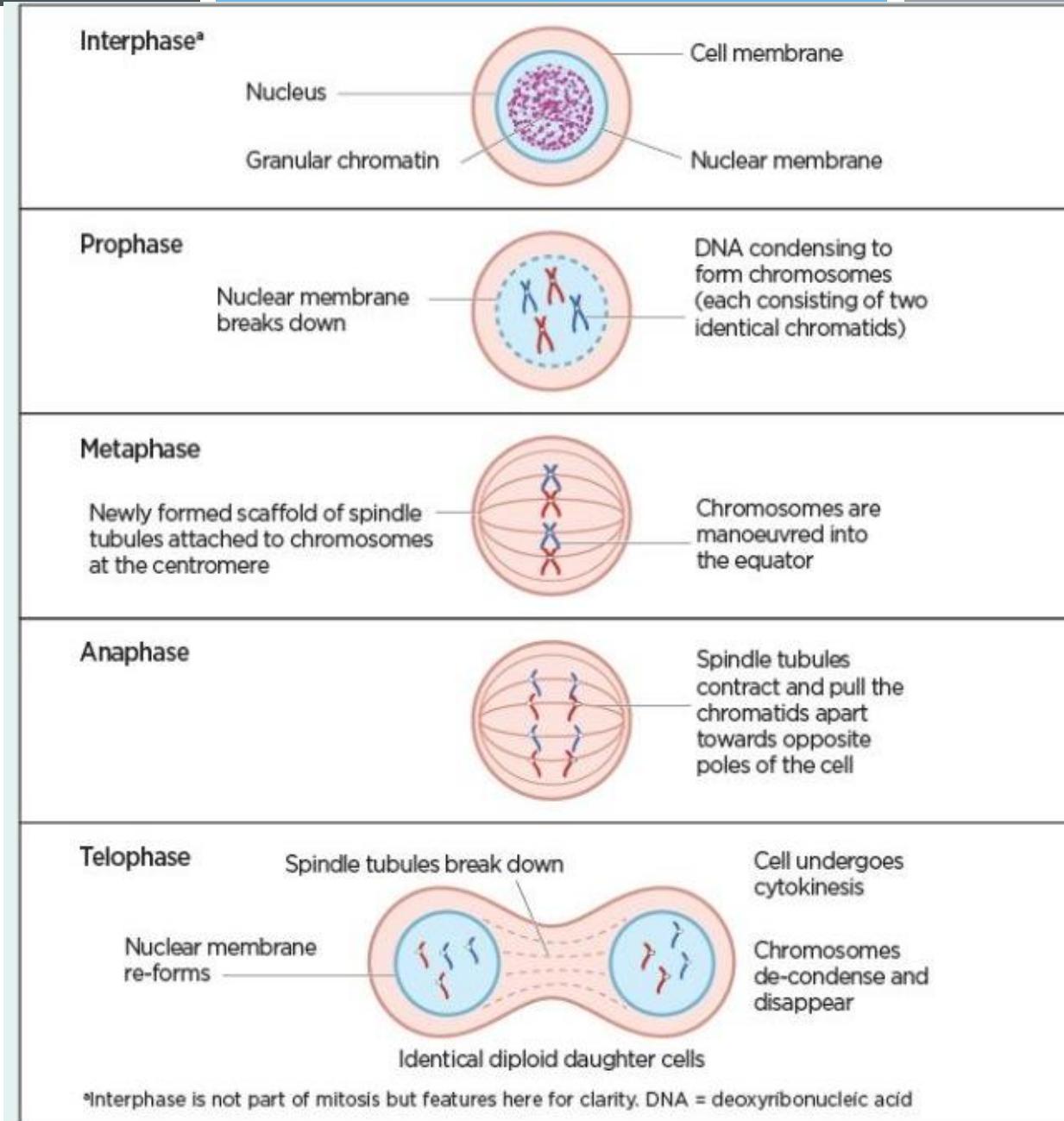


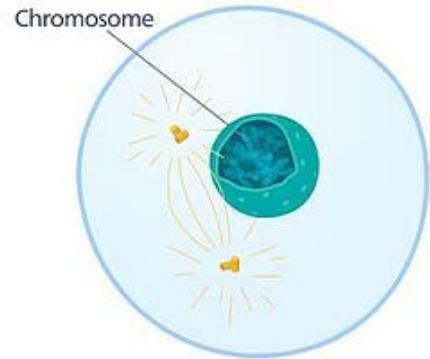
# HOW DO CELLS DIVIDE?

- There are two types of cell division: mitosis and meiosis. Most of the time when people refer to “cell division,” they mean mitosis, the process of making new body cells. Meiosis is the type of cell division that creates egg and sperm cells.
- **Mitosis** is a fundamental process for life. During mitosis, a cell **duplicates** all of its contents, including its chromosomes, and splits to form two identical daughter cells. Because this process is so critical, the steps of mitosis are carefully controlled by certain genes. When mitosis is not regulated correctly, health problems such as cancer can result.
- The other type of cell division, **meiosis**, ensures that humans have the same number of chromosomes in each generation. It is a two-step process that **reduces the chromosome number by half**—from 46 to 23—to **form sperm and egg cells**. When the sperm and egg cells unite at conception, each contributes 23 chromosomes so the resulting embryo will have the usual 46. Meiosis also allows genetic variation through a process of gene shuffling while the cells are dividing.



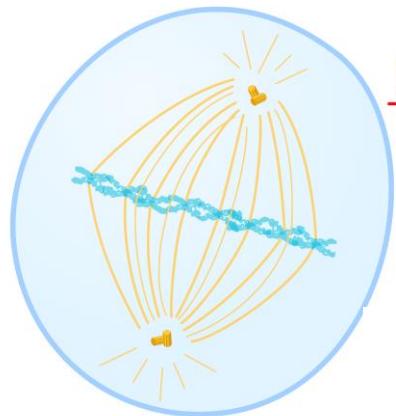






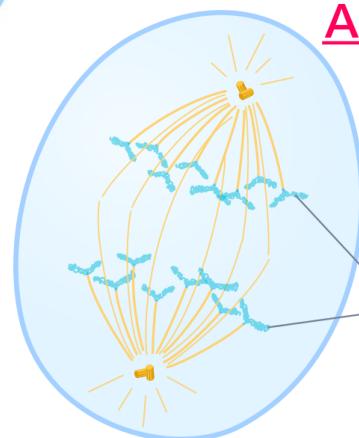
### Prophase

Chromatin condenses into chromosomes  
Nucleolus disappears



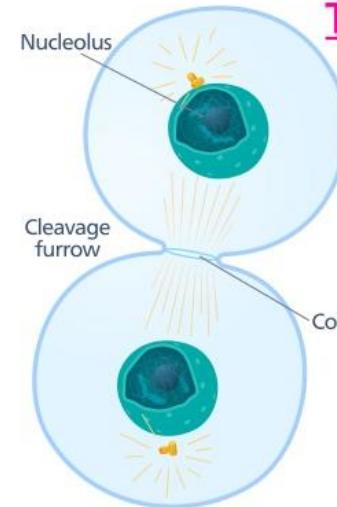
### Metaphase

Chromosomes line up along metaphase plate (imaginary plane)



### Anaphase

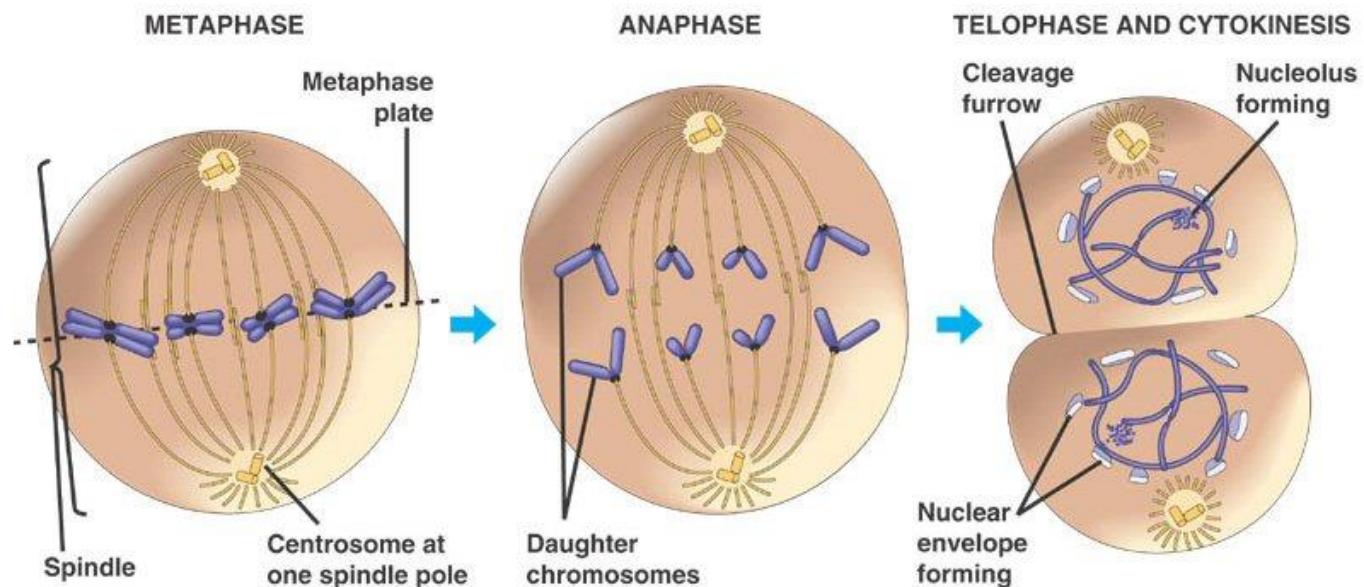
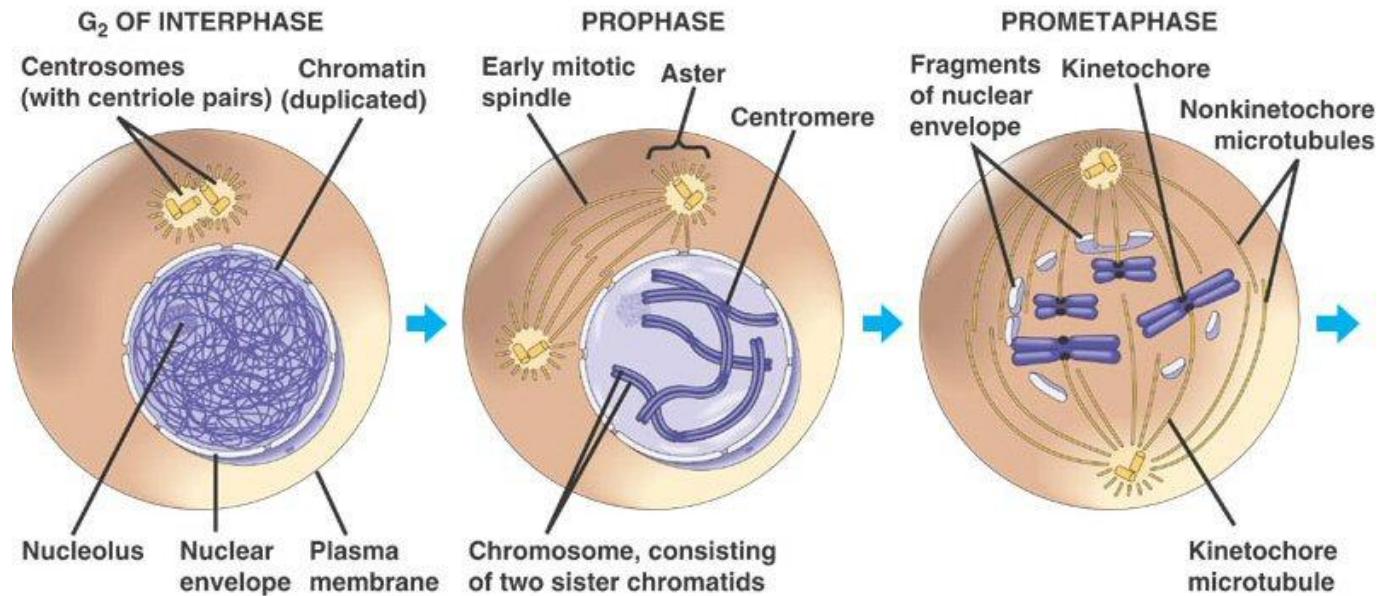
Chromosomes break at centromeres, and sister chromatids move to opposite ends of the cell



### Telophase and Cytokinesis

Nuclear membrane reforms, nucleoli reappear, chromosomes unwind into chromatin

Myosin II and actin filament ring contract to cleave cell in two

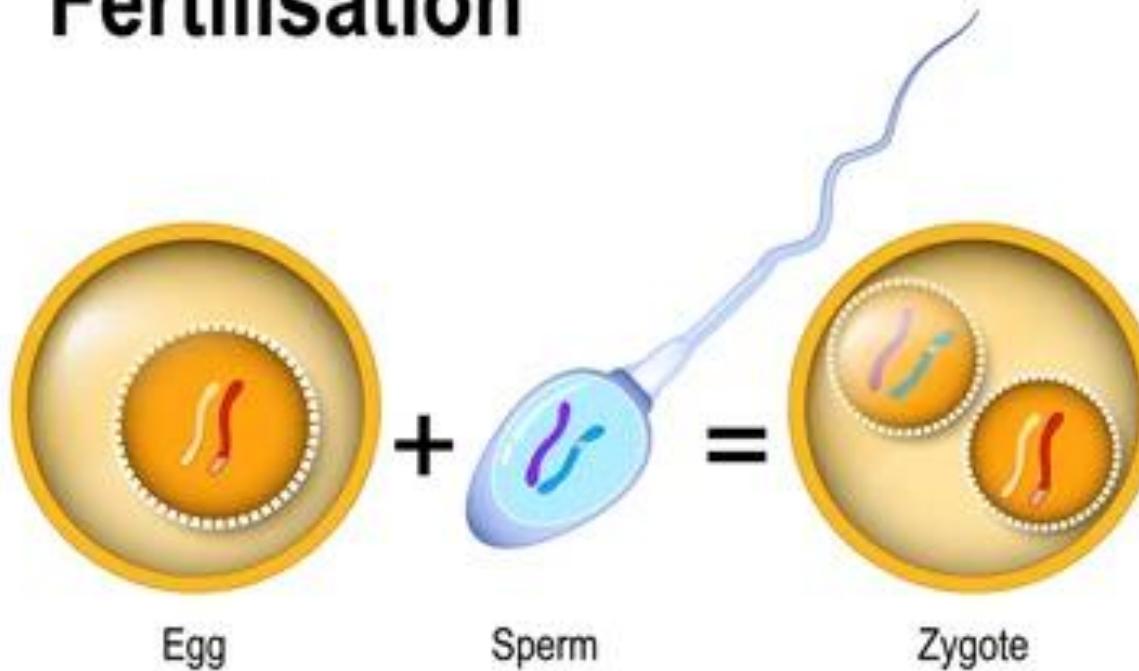


## REASONS FOR MITOSIS

Cells need to divide and duplicate for the following reasons:

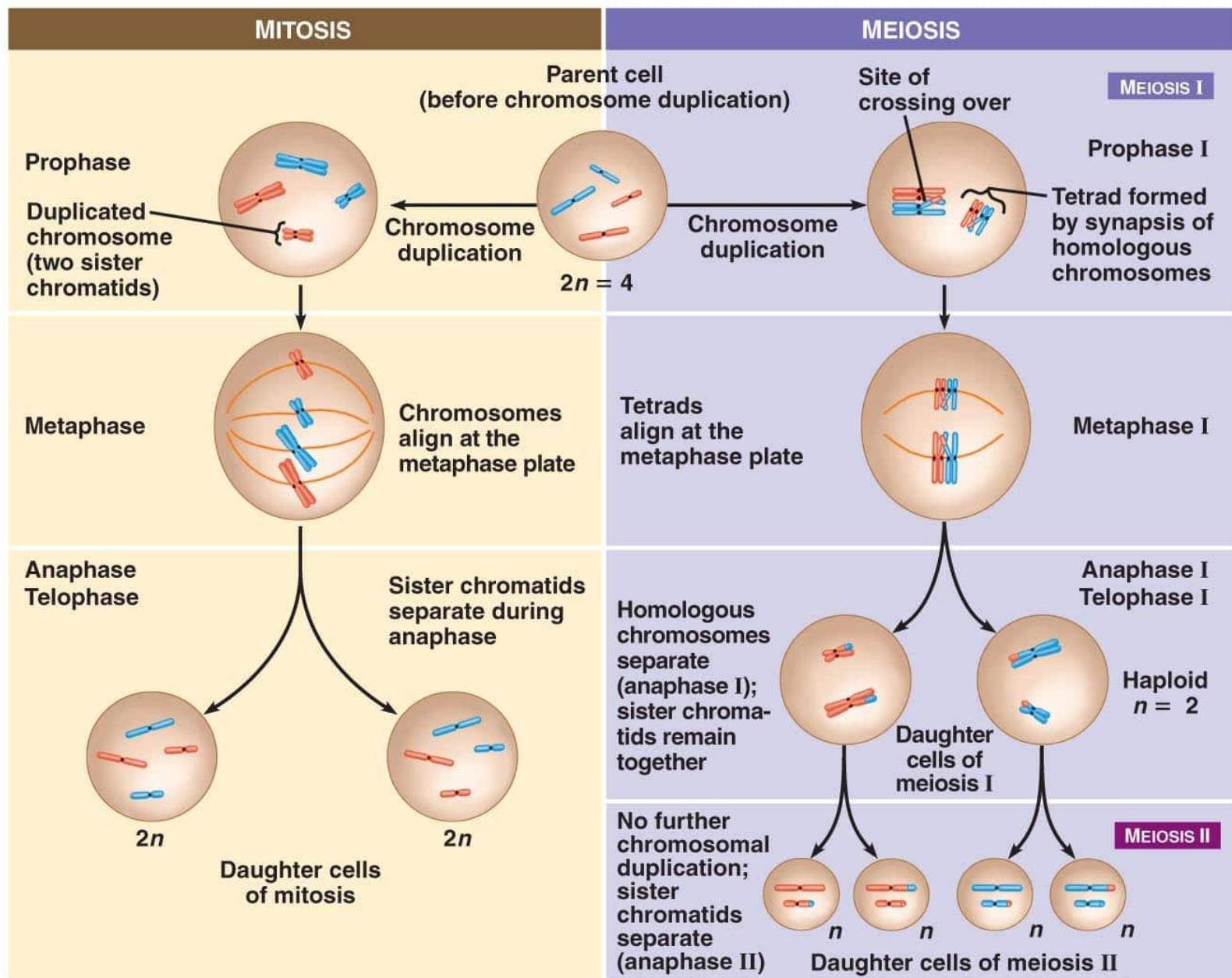
1. **Growth** – young organisms grow into adults throughout their life cycle
2. **Repair** – new cells are needed to repair damage (ie. cuts, broken bones)
3. **Replacement** – new cells need to replace old cells (ie. skin cells, lining of stomach and small intestine)

# Fertilisation



If mitosis was the only form of cell division, then new offspring would always have 2 times as many chromosomes as their parents.

Meiosis is a process of gamete formation in which diploid cells, i.e., the cells that are set aside early in animal development for sexual reproduction, yield four genetically different haploid cells.  
It occurs only in sex cells, which are eggs and sperms.



## Mitosis

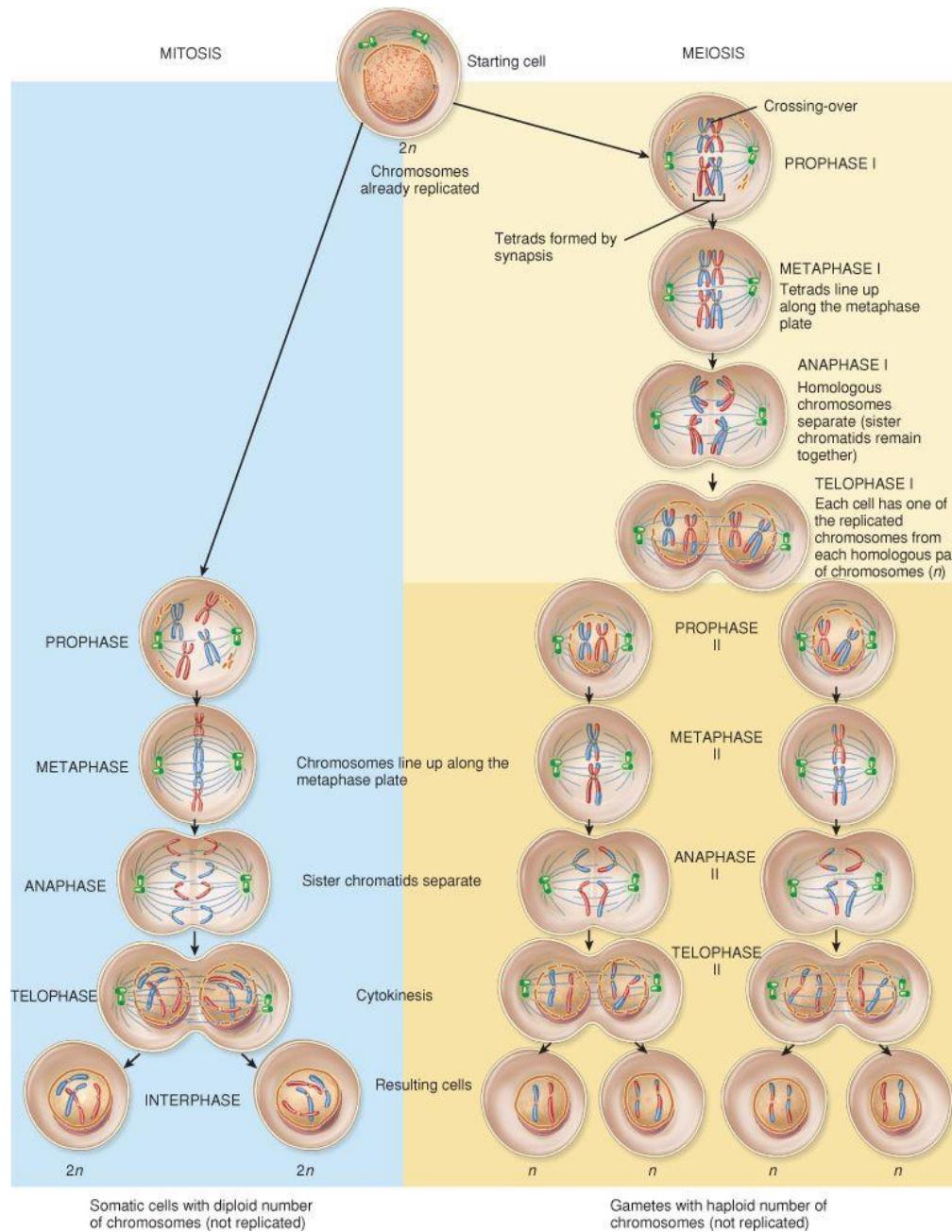
- 4 stages in total (plus interphase)
- Happens in somatic cells
- Purpose is cellular proliferation
- Produces 2 diploid daughter cells
- Chromosome number remains the same
- Genetic variation doesn't change

## Same

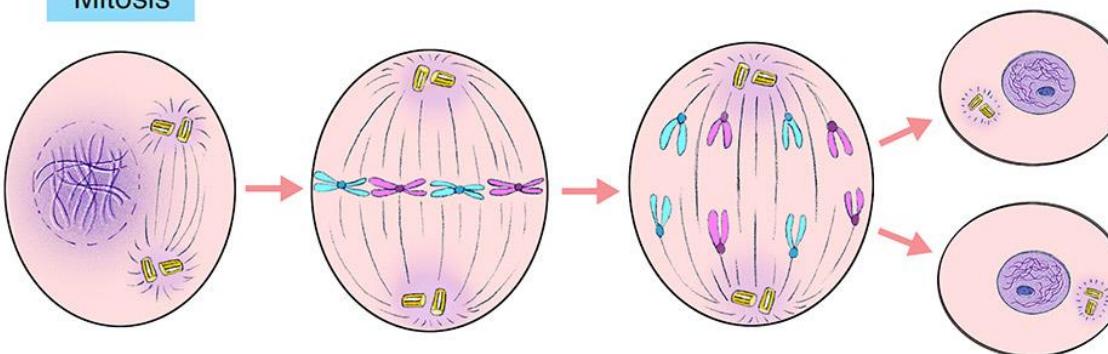
- Produce new cells
- Similar basic steps
- Start with a single parent cell

## Meiosis

- 8 stages in total (plus interphase)
- Happens in germ cells
- Purpose is sexual reproduction
- Produces 4 haploid daughter cells
- Chromosome number is halved in each daughter cell
- Genetic variation increased

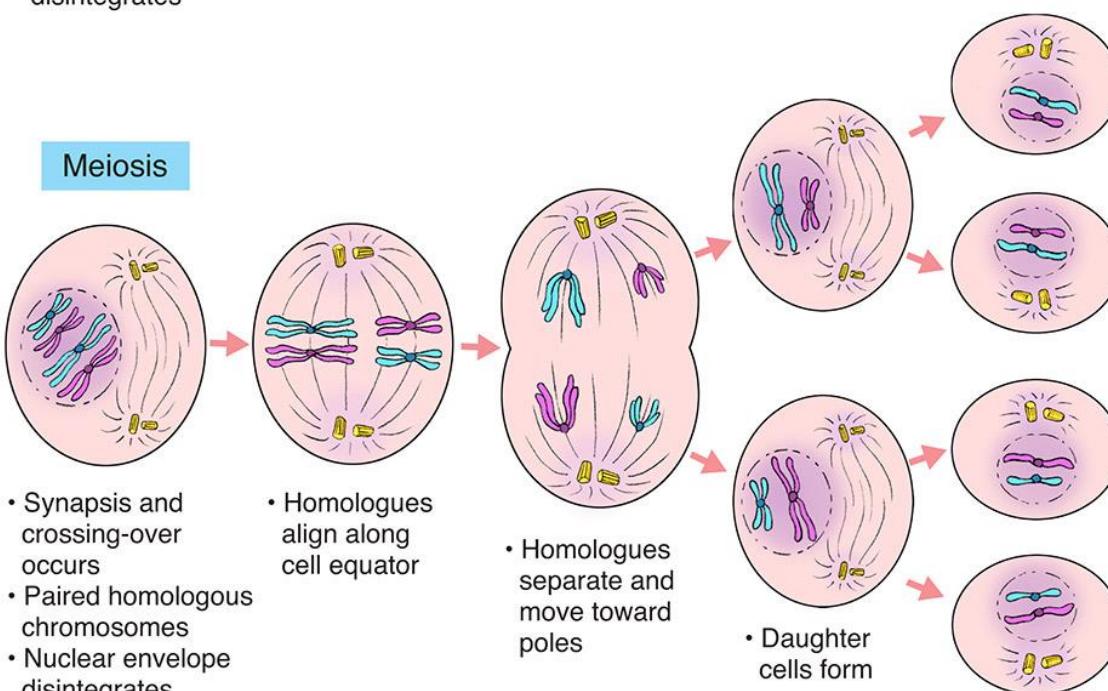


## Mitosis



- Centrioles move toward poles
- Chromatin begins to form into chromosomes
- Nuclear envelope disintegrates
- Chromosomes align along cell equator to form metaphase plate
- Sister chromatids separate and move toward poles
- Daughter cells form
- Nuclei are genetically identical to parent cell

## Meiosis



- Synapsis and crossing-over occurs
- Paired homologous chromosomes
- Nuclear envelope disintegrates
- Homologues align along cell equator
- Homologues separate and move toward poles
- Daughter cells form
- Daughter chromosomes separate to form gametes
- Nuclei are not genetically identical to parent cell

# GENETIC VARIATION - ASSORTMENTS OF CHROMOSOMES

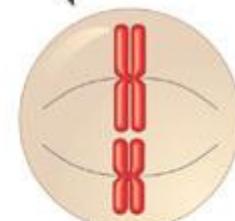
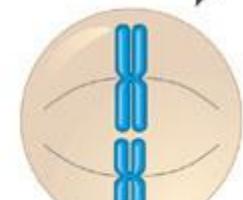
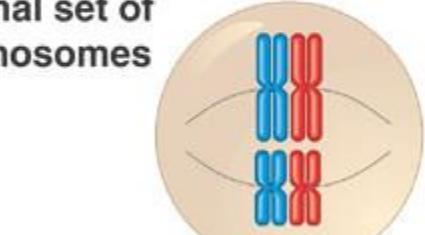
- Offspring that result from sexual reproduction are genetically different from their parents and from one another. This genetic variety in offspring is the raw material for natural selection. How can genetic variety arises through meiosis and fertilization?
- One way in which meiosis contributes to genetic variety is the way chromosomes arrange themselves during meiosis. The example is an organism with a diploid chromosome number of four ( $2n = 4$ ). How the chromosomes in each homologous pair (tetrads) line up and separate at metaphase I is a matter of chance, like the flip of a coin. So, **the assortment of chromosomes that end up in the resulting cells occurs randomly.**
- In a diploid cell with four chromosomes (two homologous pairs), there are two equally possible ways for the chromosomes inherited from the two parents to be arranged during metaphase I. This variation in the orientation of chromosomes leads to gametes with four equally possible combinations of chromosomes.
- If you know the haploid number for an organism, you can calculate the number of possible combinations in the gametes. The possible combinations are equal to  $2^n$ , where n is the haploid number. For an organism  $n = 2$ , the number of chromosome combinations is 4 ( $2^2$ ). For a human,  $n = 23$ , so there are  $2^{23}$ , about 8 million, possible chromosome combinations!

### Key

■ Maternal set of chromosomes

■ Paternal set of chromosomes

Possibility 1

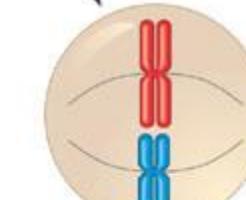
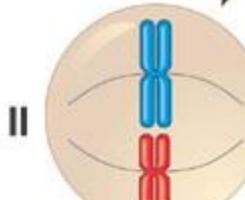
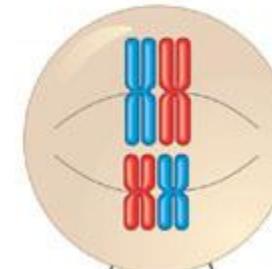


Combination 1

Combination 2

Two equally probable arrangements of chromosomes at metaphase I

Possibility 2

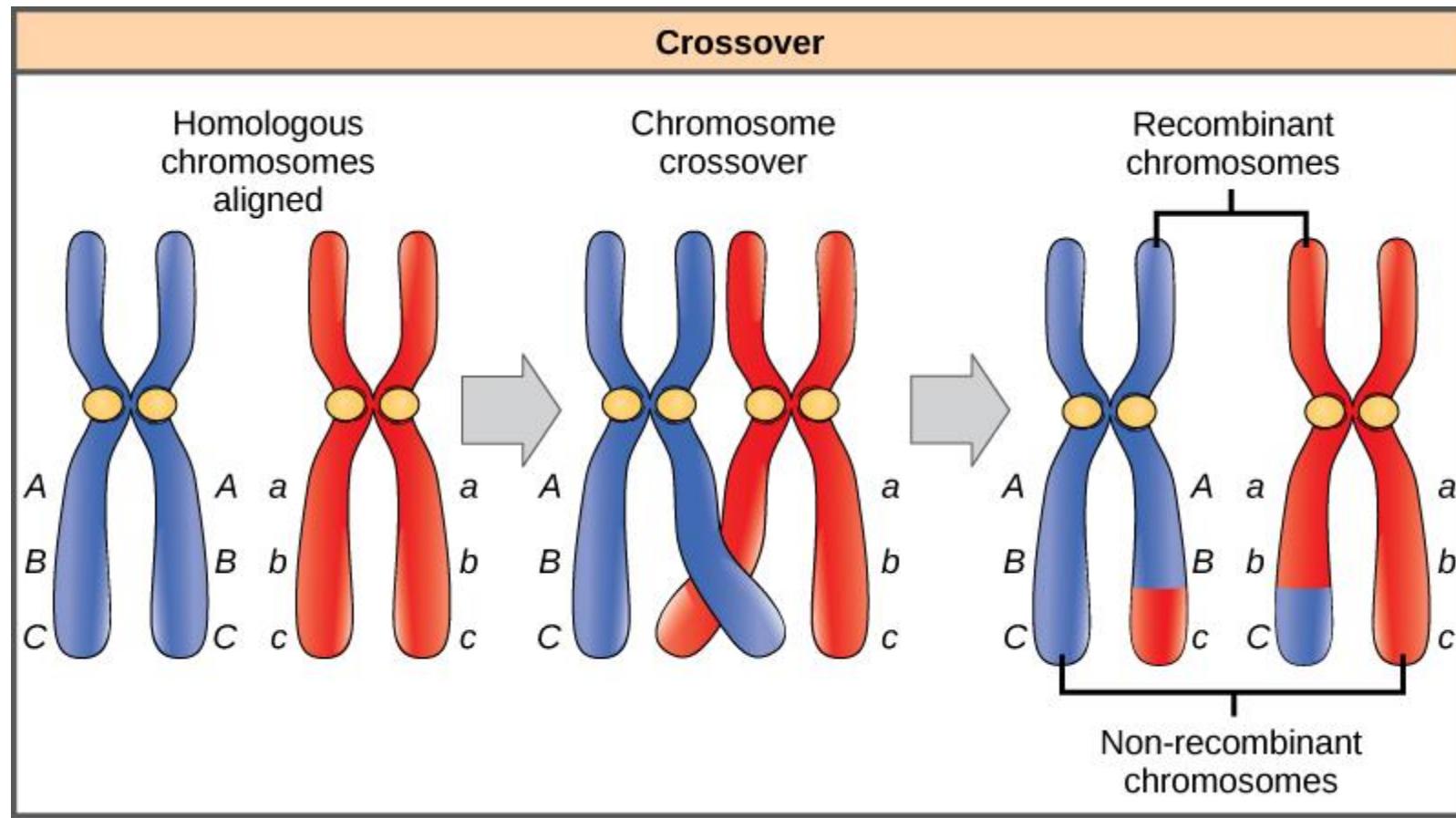


Daughter cells

Combination 3

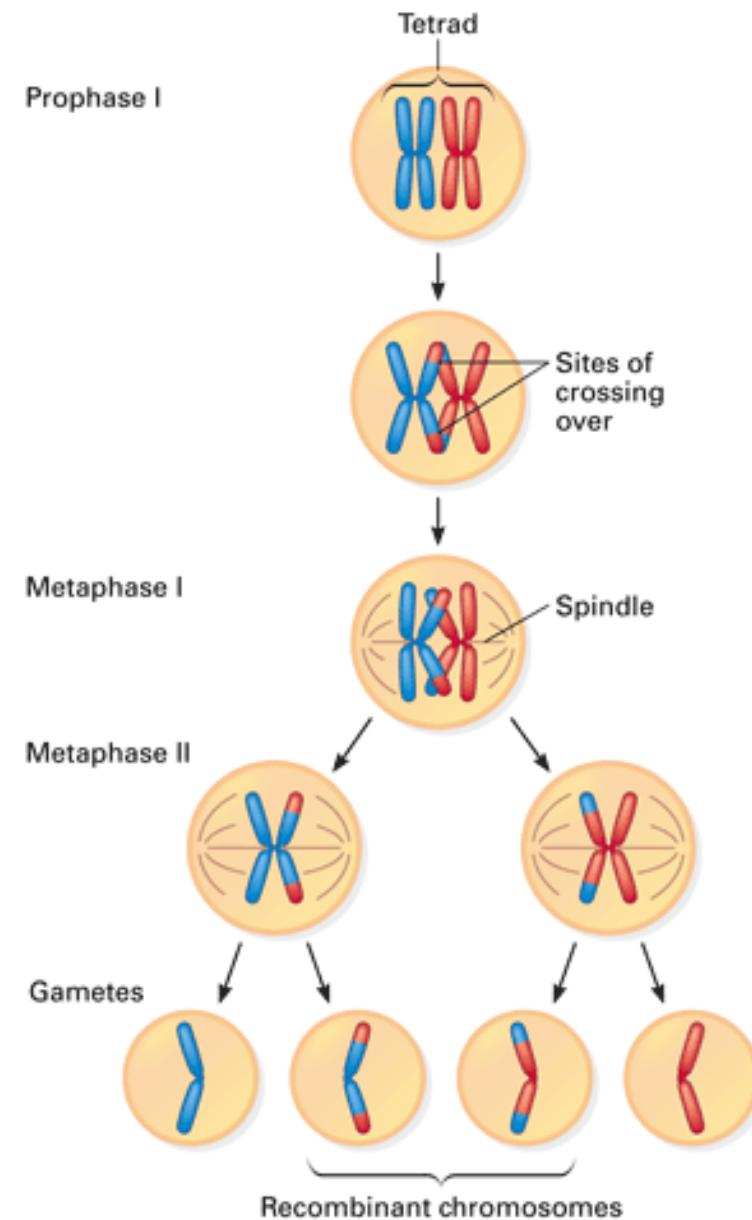
Combination 4

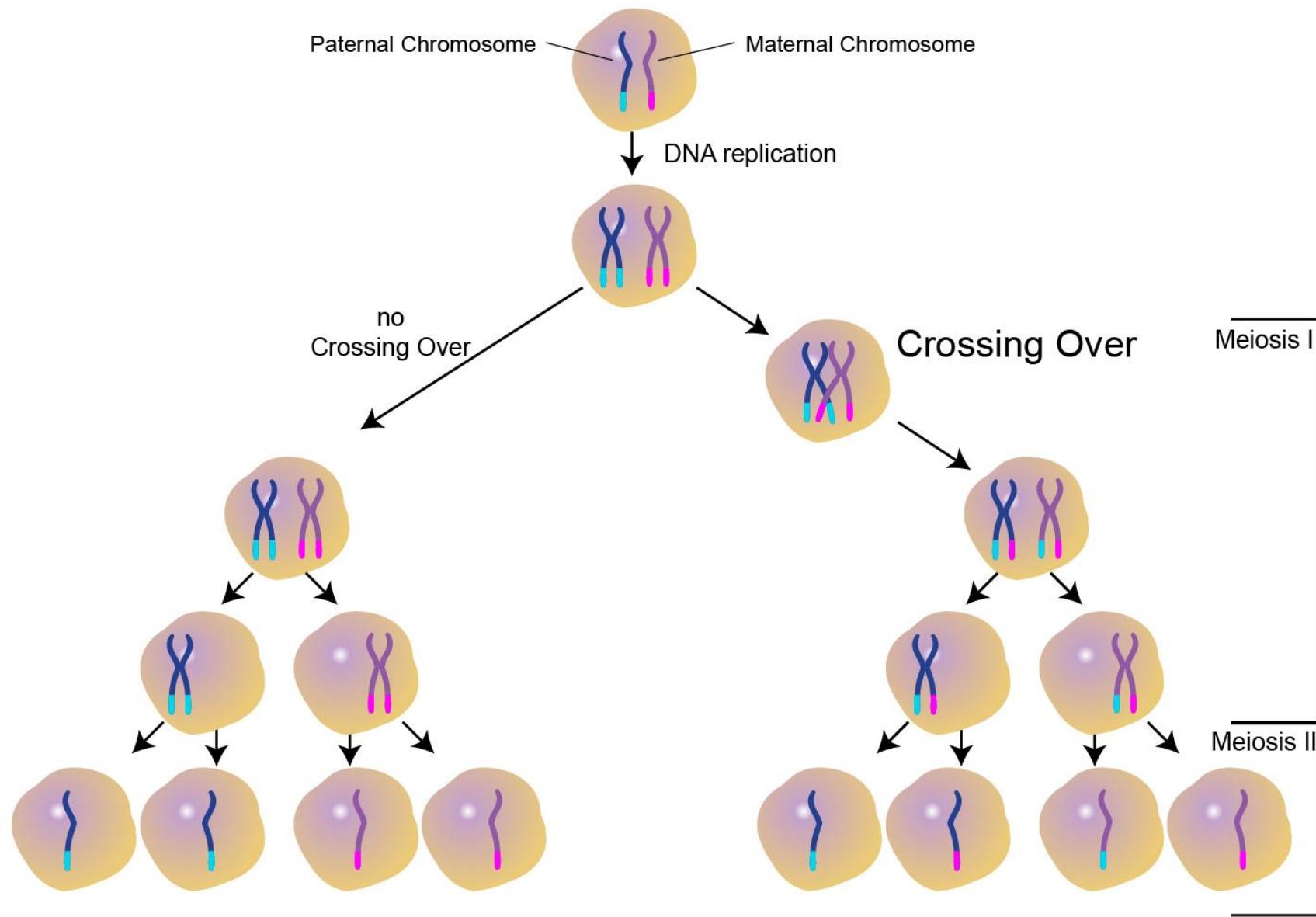
Metaphase II

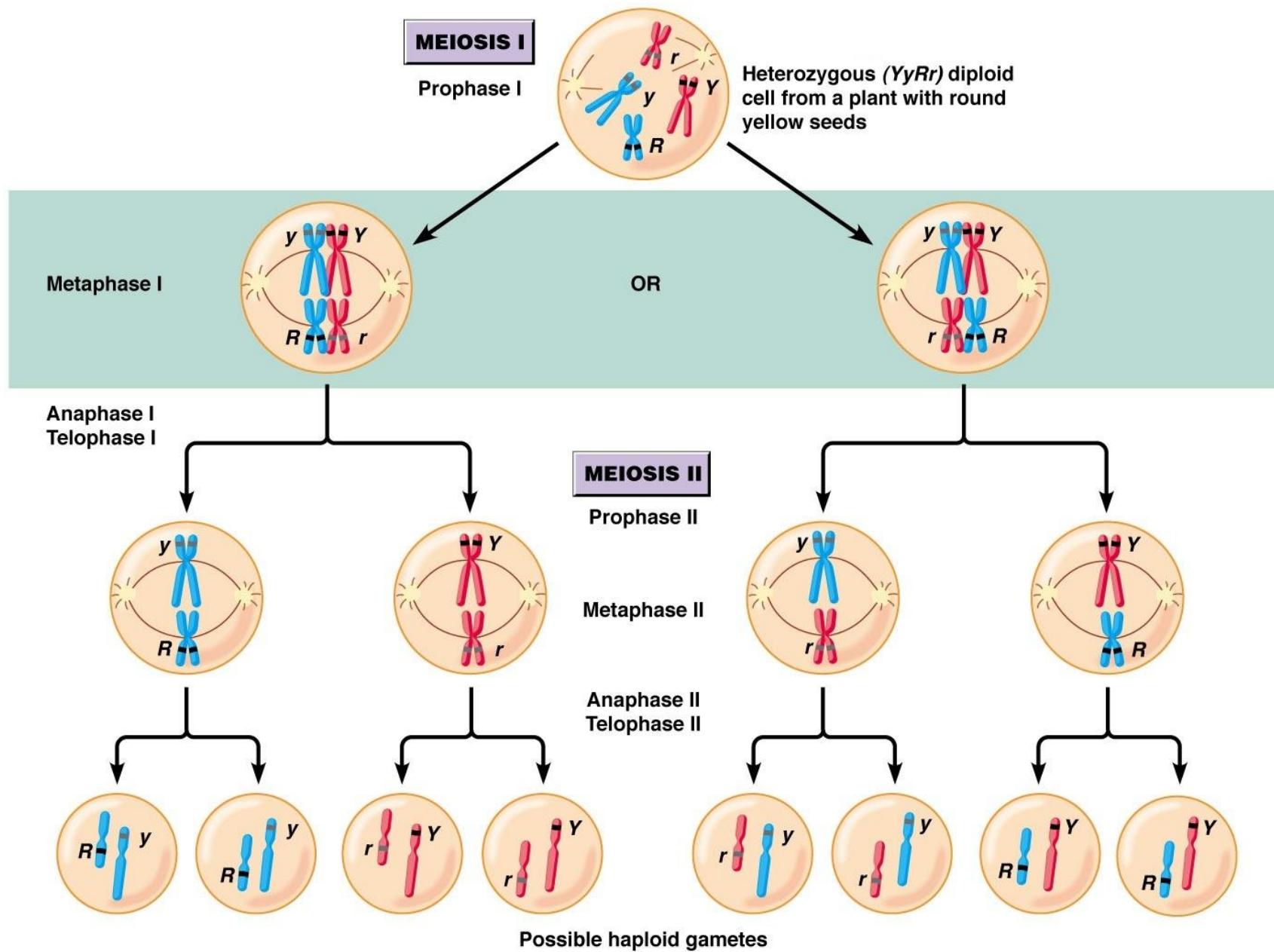


Early in prophase I, a chromatid from one chromosome exchanges a segment with the corresponding segment from the other chromosome. These altered chromosomes give rise to what are known as "recombinant chromosomes" in the gametes.

So, on top of all the possible chromosome combinations, **crossing over adds another source of variation**. Crossing over can produce a single chromosome that contains a new combination of genetic information from different parents, a result called genetic recombination.

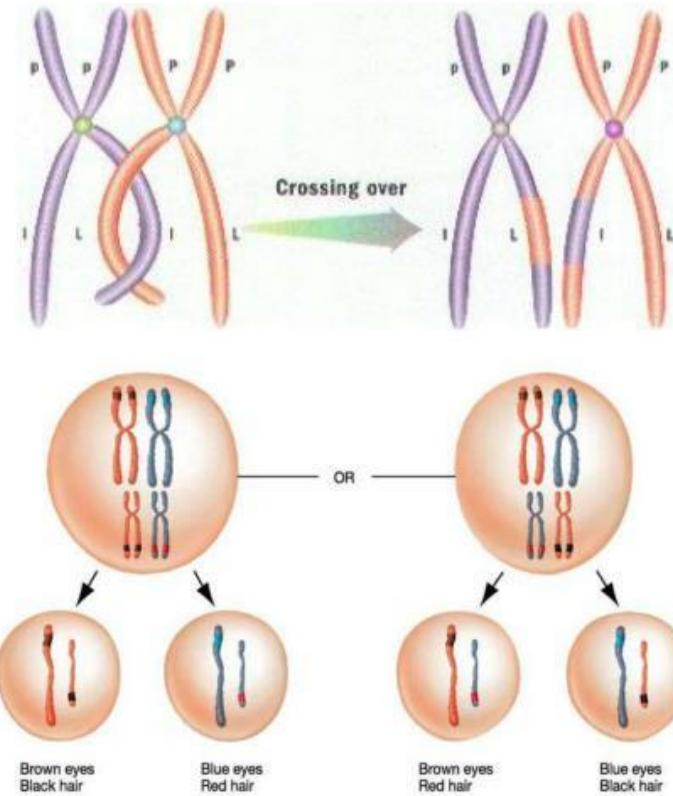






# What are the sources of Genetic Variation in Meiosis?

- 1) **Crossing Over** – crossing over of genetic information of homologous chromosomes during prophase I
- 2) **Independent Assortment** – random alignment of maternal/paternal chromosomes during metaphase I
- 3) **Fertilization** - two gametes (sex cells) coming together to form a zygote



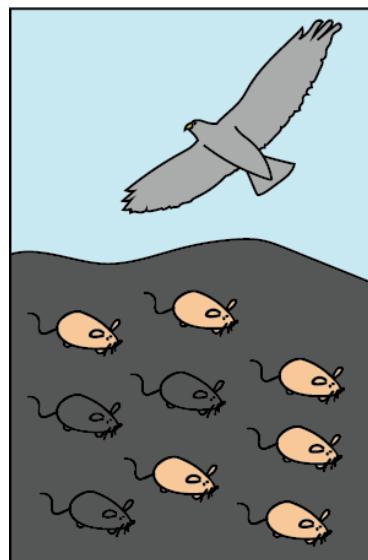
# HOW IS GENETIC DIVERSITY INCREASED BY MEIOSIS?

- Genetic diversity is increased in meiosis two ways:
  1. chromosomal crossover leading to genetic recombination during synapsis in prophase I, and,
  2. independent line up or assortment ('shuffling') before the segregation of homologous pairs of chromosomes in anaphase I, due to the random orientation of tetrads in metaphase I.
- In sexual reproduction, random fertilization of gametes resulting from meiosis also increases genetic diversity.
- **Mutation** of DNA is the only source of new genetic variation; sexual reproduction and meiosis merely rearrange it randomly into novel combinations.

# HOW ARE GENE VARIANTS INVOLVED IN EVOLUTION?

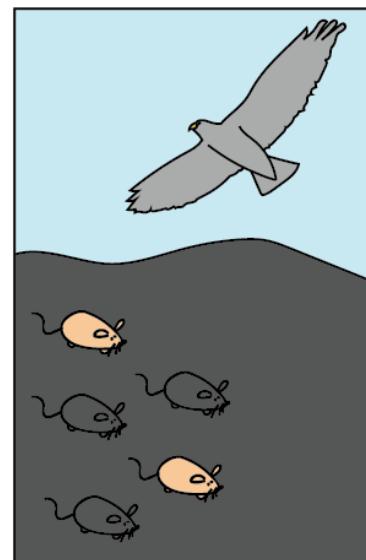
- Evolution is the process by which populations of organisms change over generations. Genetic variations underlie these changes. Genetic variations can arise from gene variants (also called mutations) or from a normal process in which genetic material is rearranged as a cell is getting ready to divide (known as genetic recombination).
- **Genetic variations** that alter gene activity or protein function **can introduce different traits** in an organism. If a trait is advantageous and helps the individual survive and reproduce, the genetic variation is more likely to be passed to the next generation (a process known as natural selection). Over time, as generations of individuals with the trait continue to reproduce, the advantageous trait becomes increasingly common in a population, making the population different than an ancestral one. Sometimes the population becomes so different that it is considered a new species.
- The environment in which a population of organisms lives is integral to the selection of traits. Some differences introduced by variants may help an organism survive in one setting, but not in another.

# VARIATION AND NATURAL SELECTION



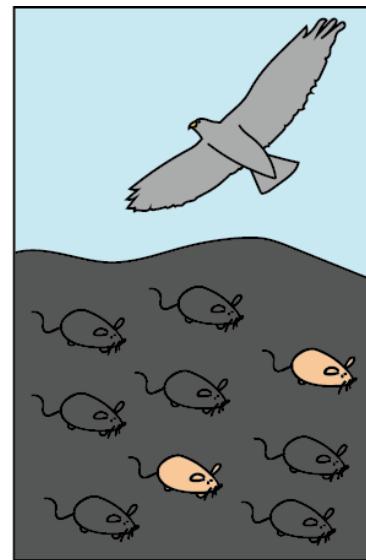
A population of mice has moved into a new area where the rocks are very dark. Due to natural genetic variation, some mice are black, while others are tan.

Some mice are eaten by birds



Tan mice are more visible to predatory birds than black mice. Thus, tan mice are eaten at higher frequency than black mice. Only the surviving mice reach reproductive age and leave offspring.

Mice reproduce, giving next generation

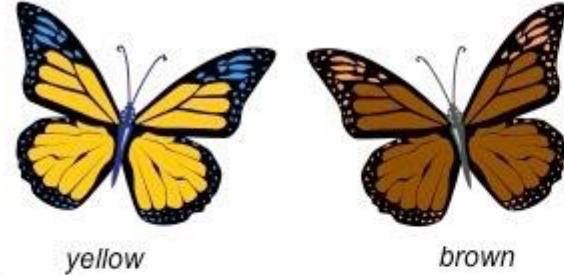


Because black mice had a higher chance of leaving offspring than tan mice, the next generation contains a higher fraction of black mice than the previous generation.

# Process of natural selection

1

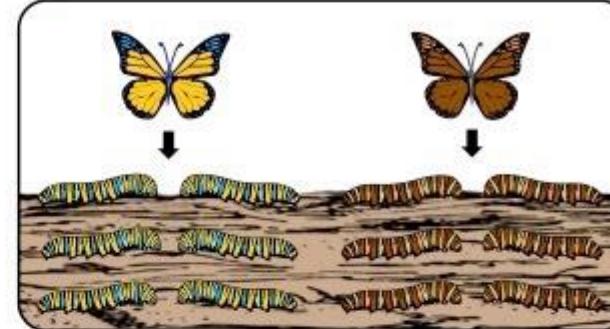
## Variation



There is genetic variation within a population which can be inherited

2

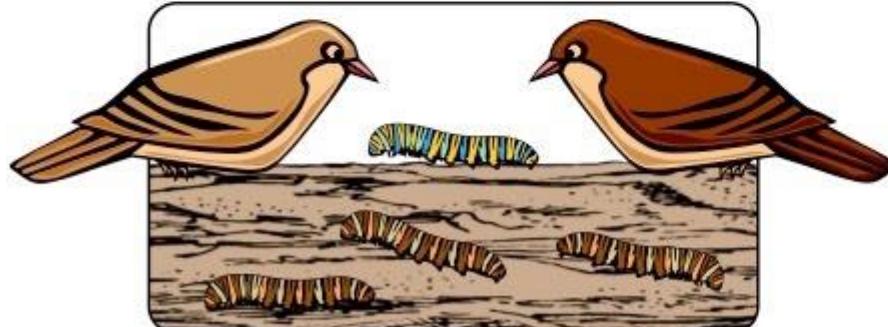
## Competition



Overproduction of offspring leads to competition for survival

3

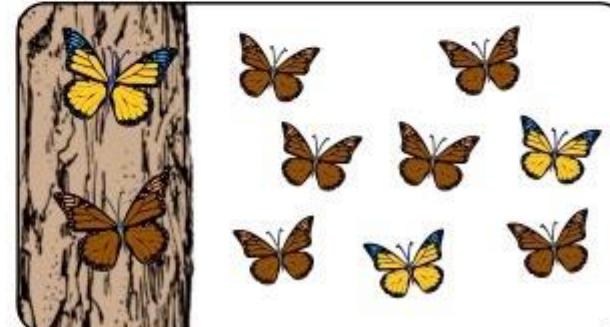
## Adaptations



Individuals with beneficial adaptations are more likely to survive to pass on their genes

4

## Selection



Over many generations, there is a change in allele frequency (evolution)

---

# HOW ARE GENE VARIANTS INVOLVED IN EVOLUTION?

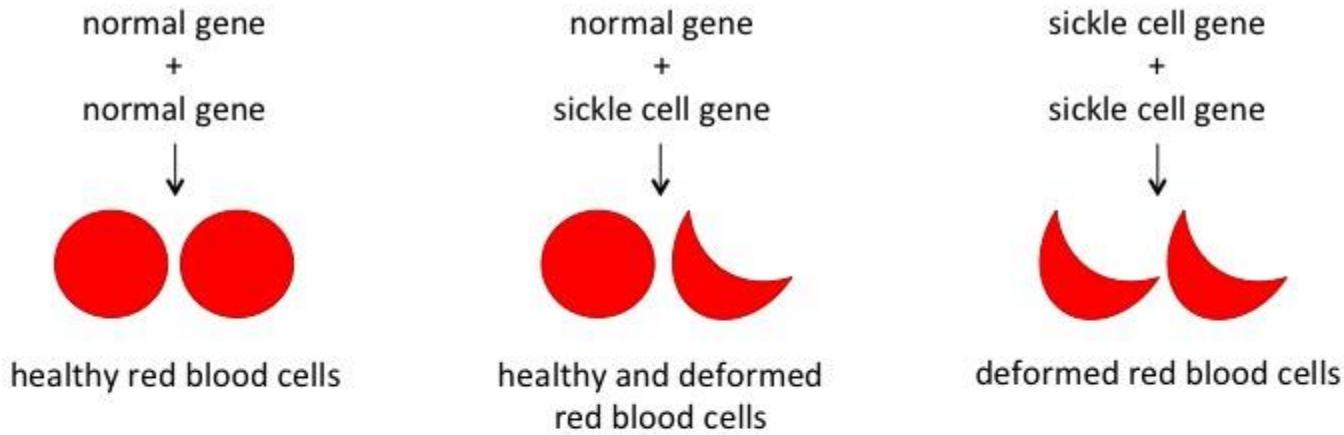
- Why do some harmful traits, like genetic diseases, persist in populations instead of being removed by natural selection?
- In many cases, the answer is not clear. For some conditions, having one altered copy of a gene in each cell is advantageous, while having two altered copies causes disease. The best-studied example of this phenomenon is sickle cell disease: Having two altered copies of the HBB gene in each cell results in the disease, but having only one copy provides some resistance to malaria. This disease resistance helps explain why the variants that cause sickle cell disease are still found in many populations, especially in areas where malaria is prevalent.

# MALARIA

*Anopheles*  
Vector for malaria



- Malaria is a disease caused by a parasite called Plasmodium. These parasites are transferred to humans through bite by female Anopheles mosquitoes. Inside humans, these parasites complete a specialized form of growth cycle and reside within the red blood cells and use the oxygen within these red cells.
- Sickle cells (heterozygous) offer protection against malaria.
- It doesn't prevent malaria; and if you have two copies of the defective gene, or full-blown sickle-cell disease, it will have a negative effect on your health and lifespan.
- However: if you have only one bad copy of the gene, giving you sickle-cell trait, you are highly resistant to the effects of malaria. Sickled-cell trait generally causes no medical problems, except under extreme conditions (high altitude, severe dehydration, scuba diving, or very high intensity physical activity).
- Since malaria kills large numbers of human beings, anything which offers some protection against malaria will confer a survival advantage. This is presumably why such a serious genetic disease is so common, and has not disappeared through natural selection - over the whole population, the benefits outweigh the harms.



When the Risk of Malaria Infection is High



get malaria



HEALTHY!



sickle cell anemia

When the Risk of Malaria Infection is Low



HEALTHY!



HEALTHY!



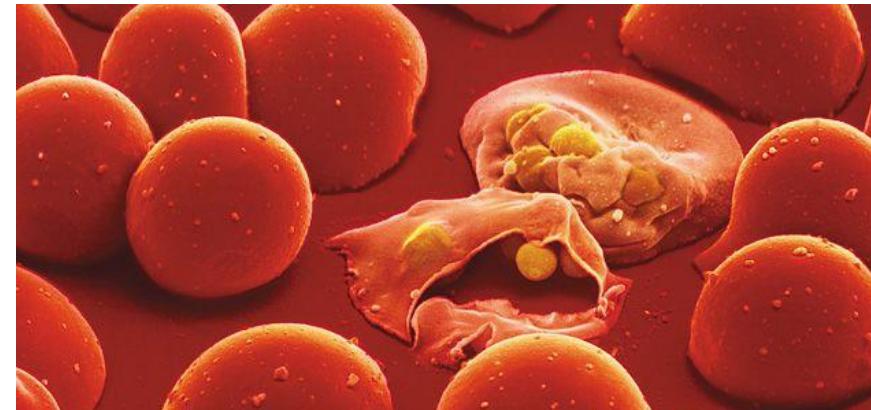
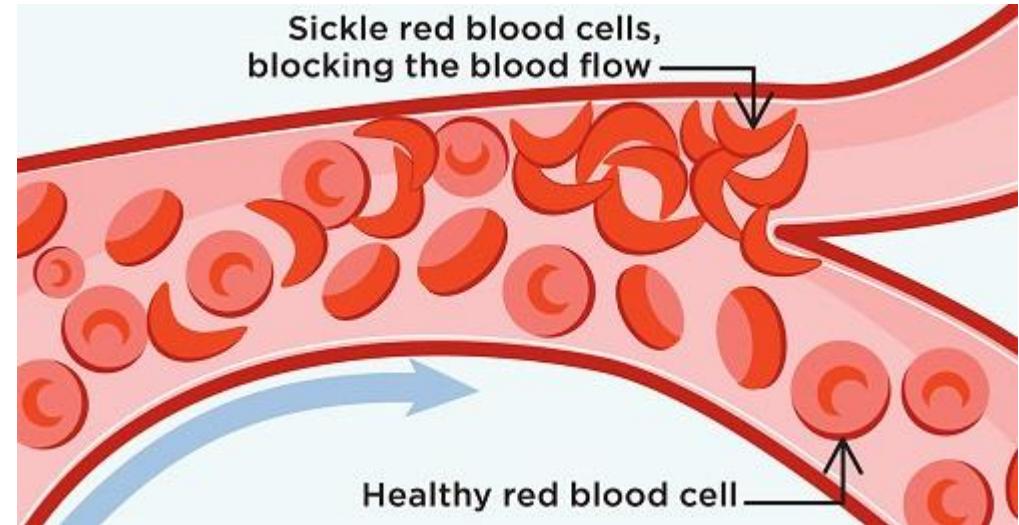
sickle cell anemia

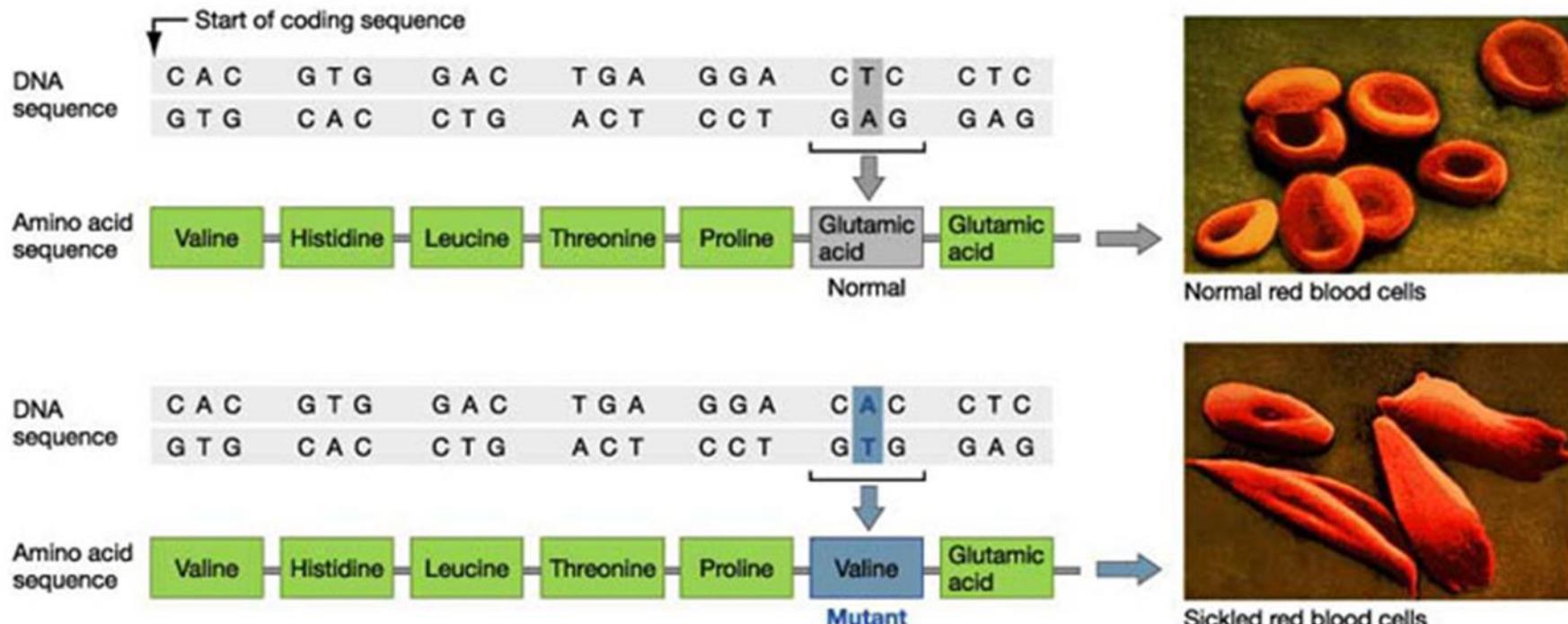
## Sickle Cell Anemia

Sickle Cell anemia is an inherited red blood cell disorder. Normal red blood cells are round like doughnuts, and they move through small blood tubes in the body to deliver oxygen.



Sickle red blood cells become hard, sticky and shaped like sickles used to cut wheat. When these hard and pointed red cells go through the small blood tube, they clog the flow and break apart. This can cause pain, damage and a low blood count, or anemia.

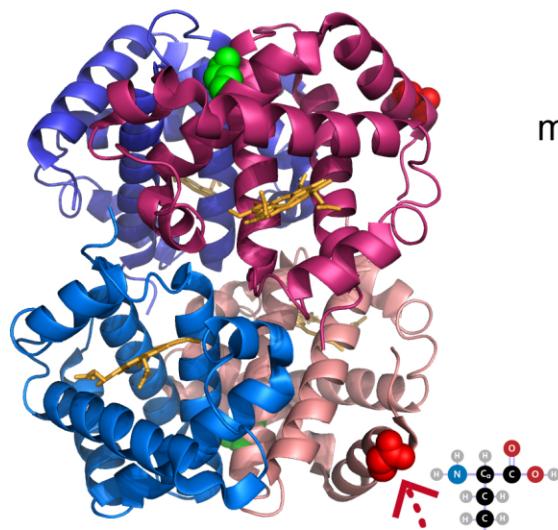




The change in amino acid sequence causes hemoglobin molecules to crystallize when oxygen levels in the blood are low. As a result, red blood cells sickle and get stuck in small blood vessels.

the  $\beta$ -globin is made based on instructions in the HBB gene on chromosome 11

mutations in this gene can lead to structural changes in the  $\beta$ -globin protein (structural hemoglobinopathies) &/or reduced  $\beta$ -globin production (thalassemias) - if you inherit 2 bad copies...



normally there's a glutamate here, which is hydrophilic (water-loving)

the sickle cell version has a valine here, which is hydrophobic (water-avoiding)

### HBB the normal form

sequence in the (double-stranded) DNA

CACCTGGACTGAGGGACTCCTC  
GTGGACCTGACTCCTGAGGAG

messenger RNA (mRNA) that gets made from it

GUGGACCUGACUCCUGAGGAG  
V H L T P E E  
Val His Leu Thr Pro Glu Glu

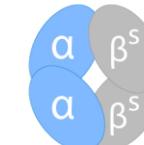
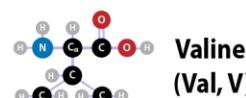
amino acids (protein letters) that get linked together to form the protein chain

### $\beta^S$ the sickle cell form

CACCTGGACTGAGGACACCTC  
GTGGACCTGACTCCTGTGGAG

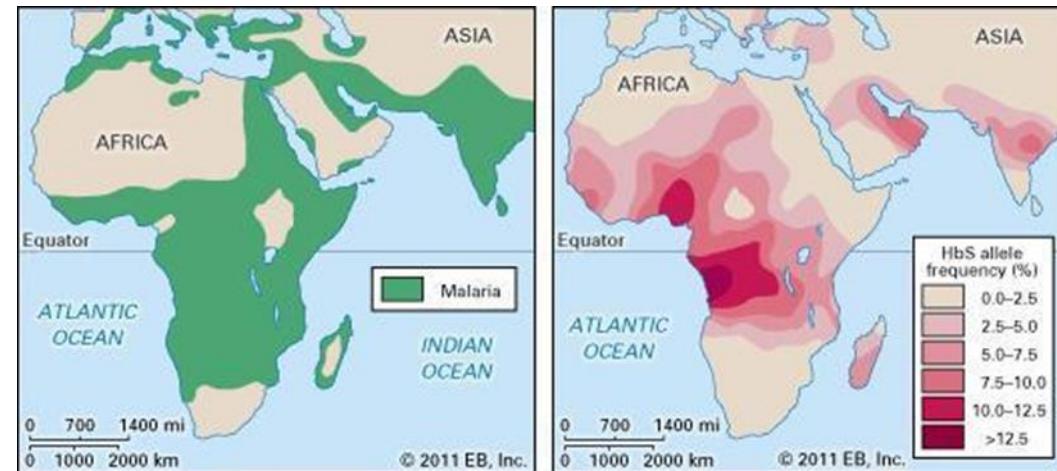
GUGGACCUGACUCCUGUGGAG  
V H L T P V E

Val His Leu Thr Pro Val Glu

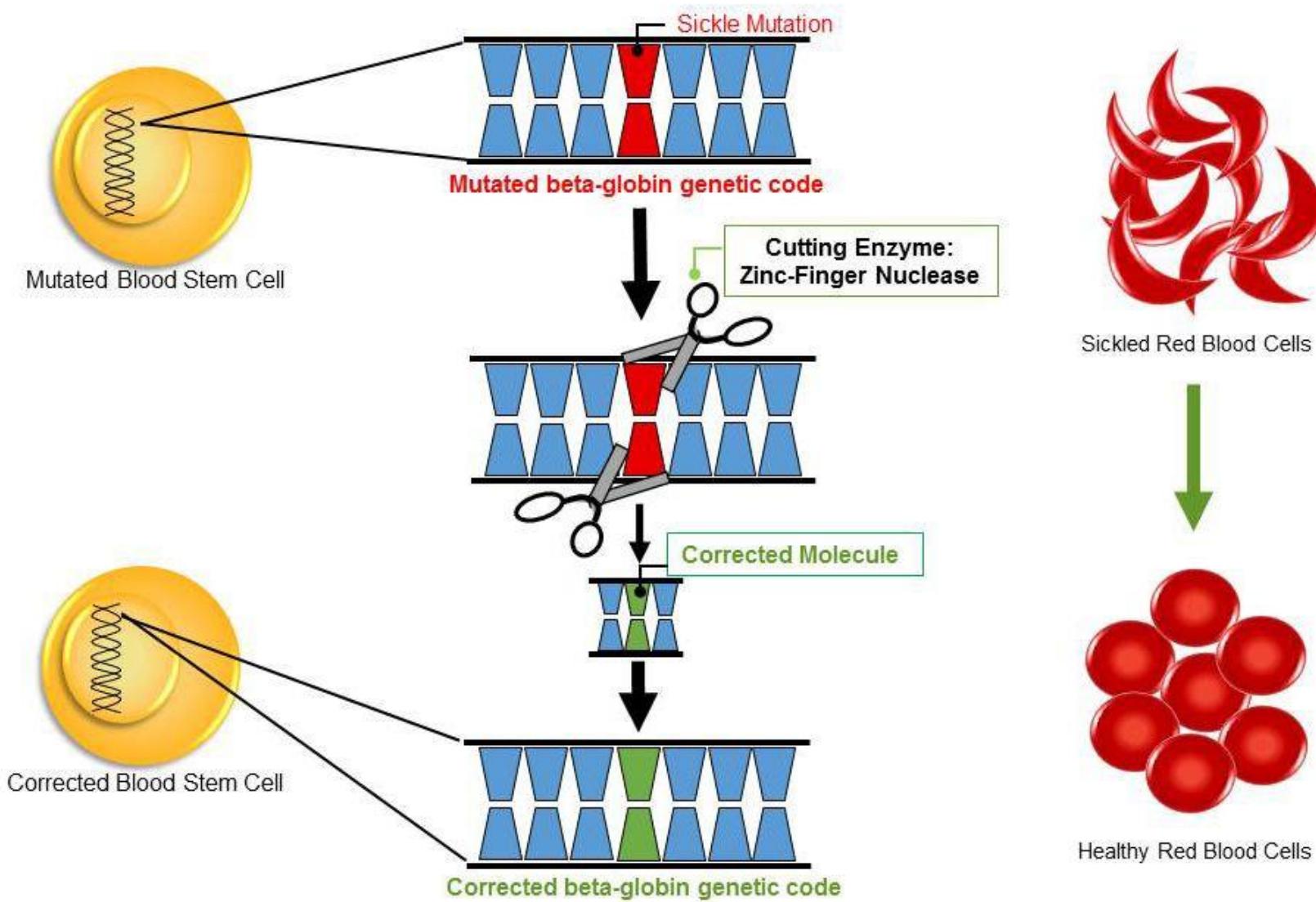


# DISTRIBUTATION OF MALARIA AND SICKLE CELL ANAEMIA

- The distribution of malaria and the distribution of sickle cell anaemia overlap in areas of Africa, Southern Asia, and the Mediterranean.
- The persistence of the HbS gene, which causes sickle cell anaemia, has been explained by the fact that heterozygous persons are resistant to malaria. (adapted from <http://media2.web.britannica.com/eb-media/84/126284>).

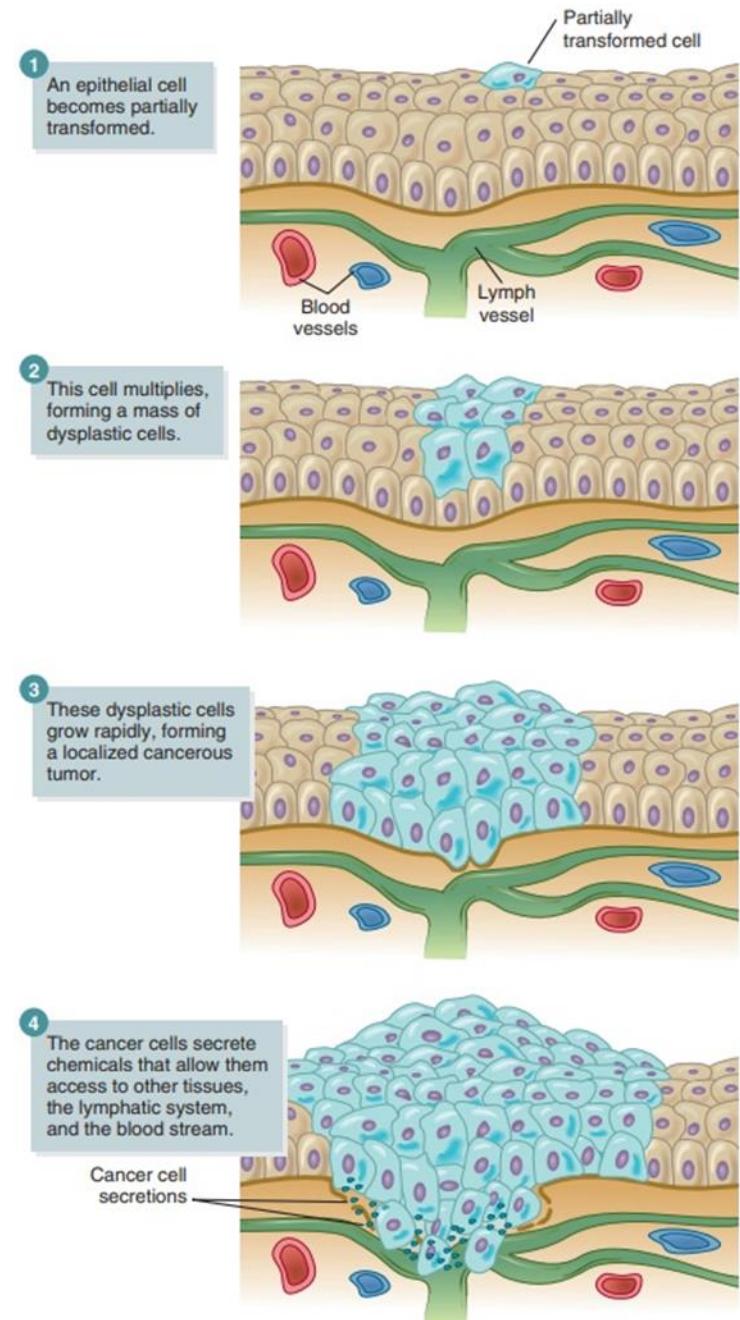


# Correction of Sickle-Cell Disease Mutation in Human Blood Stem Cells



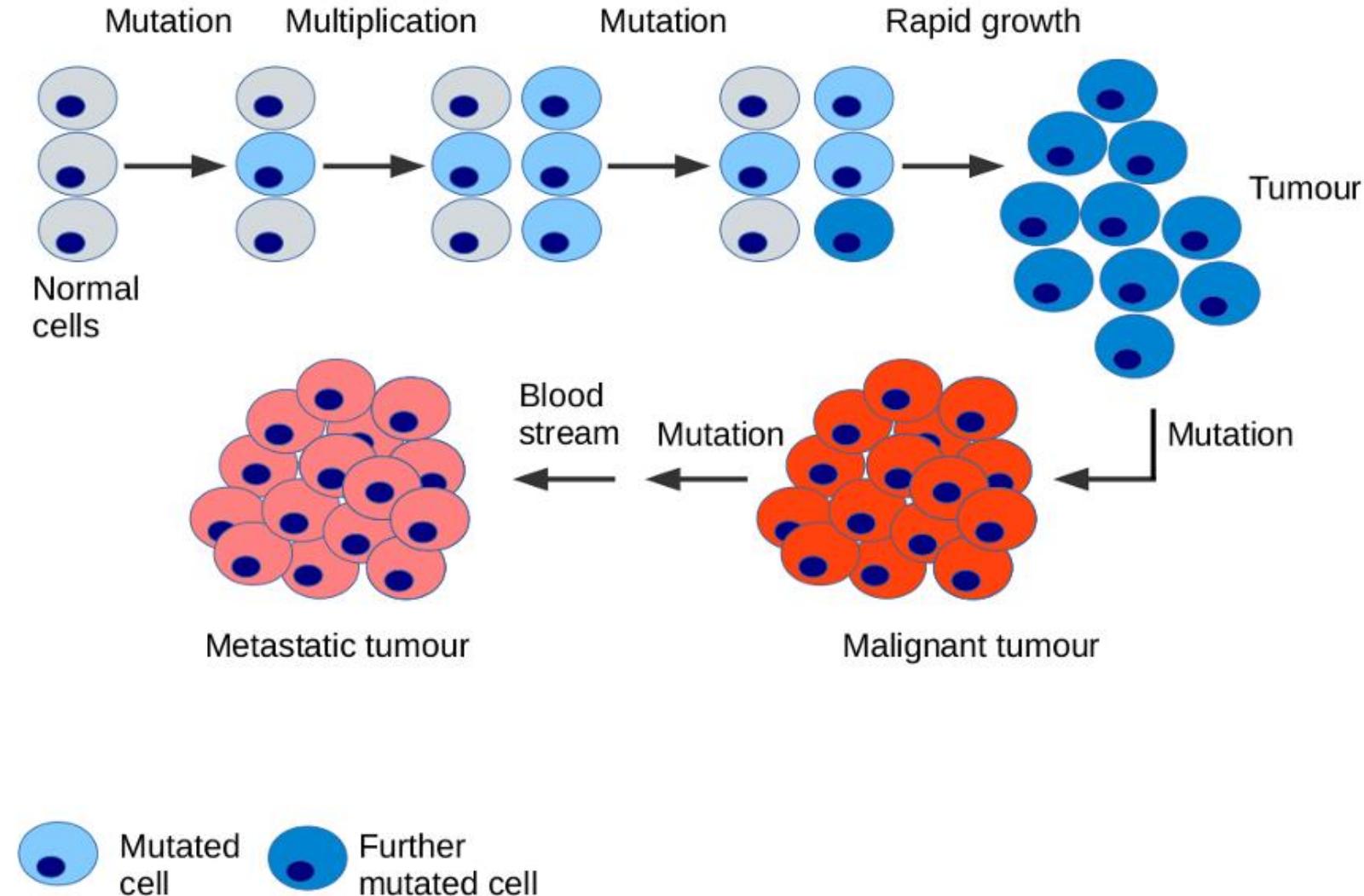
# CELL DIVISION AND CANCER

- Cell division and growth normally occur at approximately the same rate as cell death. However, when cell division and growth are higher than the cell death rate, tissues enlarge. **A tumor is a mass of tissue produced by abnormal cell growth and division.** A tumor is called **benign** when it remains within a capsule made of connective tissue. This type of tumor seldom becomes life threatening and can usually be surgically removed if it affects tissue function.
- Malignant tumors spread into surrounding tissues** in a process called invasion. The primary tumor may result in **malignant cells traveling to other organs or tissue to establish secondary tumors.** This process is called **metastasis.**
- Cancer develops mutations disrupting normal cell growth. Usually, all tumor cells are daughter cells of just one malignant cell. Malignancy often occurs when a normal gene mutates. These modified genes are called **oncogenes.** Oncogenes often code for the proteins controlling cell division.



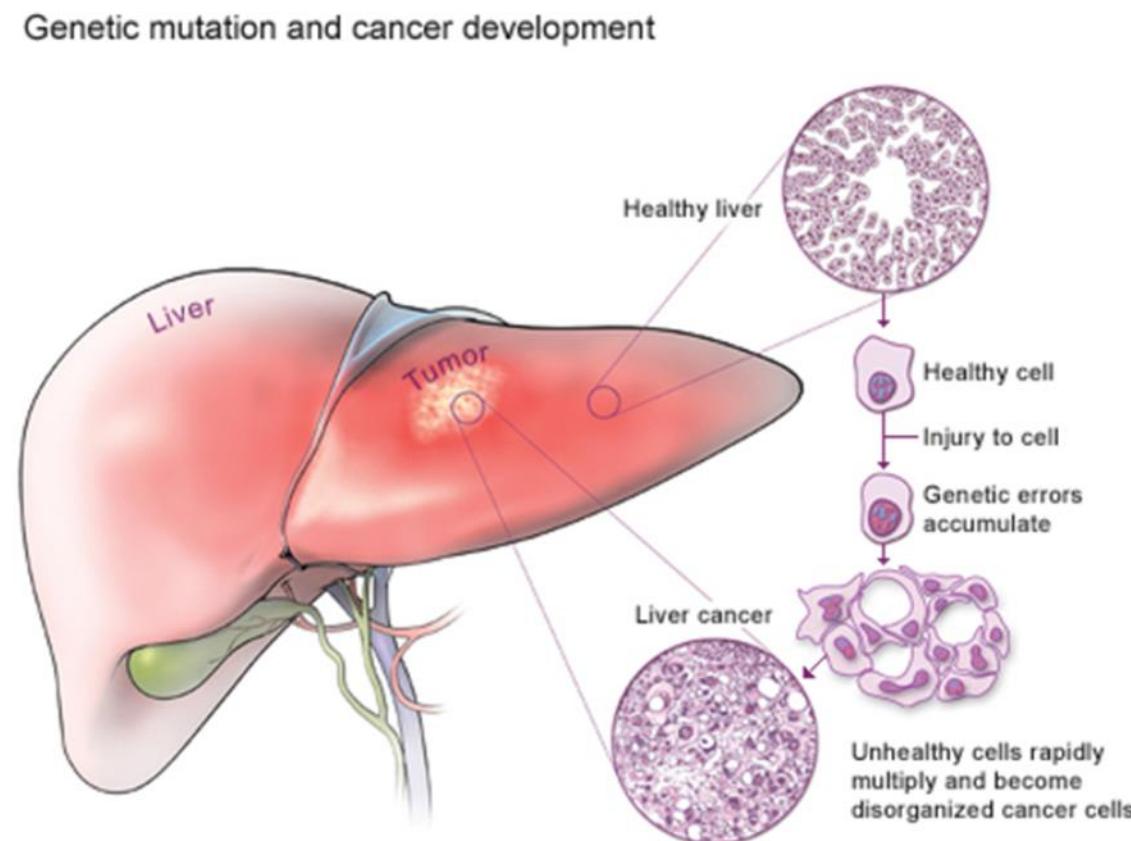
Most cancers are caused by mutations of the genes inside somatic cells occurring during cell division. Errors also occur while DNA is being replicated before cell division.

Multiple mutations, occurring over many cell generations, result in cancer. This is why older people more commonly develop cancer than younger people.



# GENES CONTROL THE GROWTH AND DIVISION OF CELLS

- Disruption of normal regulation of the cell cycle can lead to diseases such as cancer.
- When the cell cycle proceeds without control, cells can divide without order and accumulate genetic errors that can lead to a cancerous tumor.

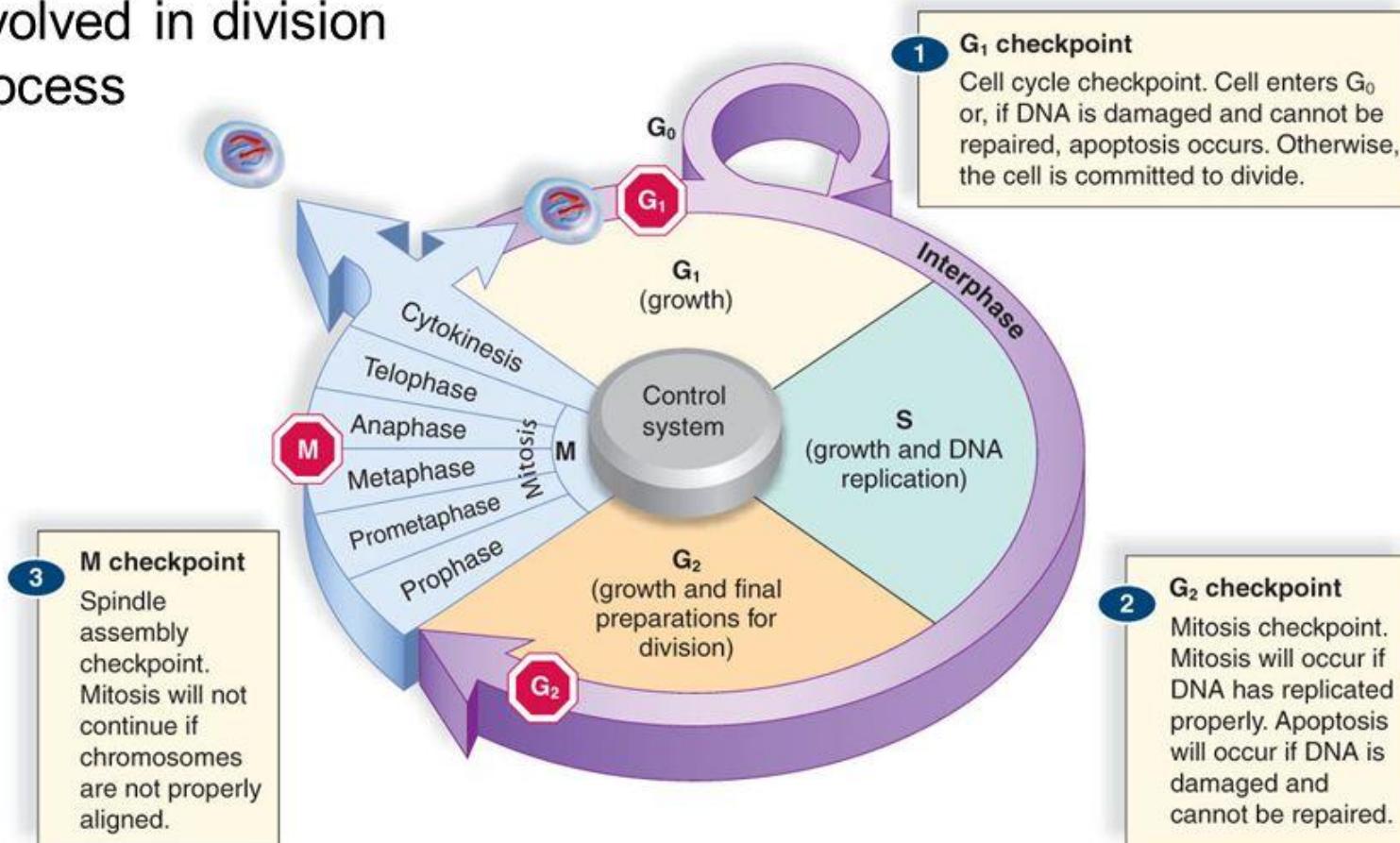


# CANCER CONTROL

- As metastasis increases, organ function changes. Cancer cells grow and multiply by taking nutrients and space from normal cells, causing weight loss in most cancer patients as the normal cells deteriorate. When cancer cells compress vital organs or have replaced healthy cells in vital organs, death may occur.
- To prevent the development of cancer in cells, the body uses two major mechanisms.
  - The first involves **DNA repair enzymes**, which detect and correct errors that occurred during replication. If these enzymes are made less effective because of their controlling genes becoming mutated, this first mechanism of fighting against cancer cannot be effective.
  - The second mechanism is **apoptosis**, which is a self-destructive process that destroys cells containing abnormal DNA. Therefore, mutated cells are made to self-destruct before cancer can develop. Apoptosis also occurs in normal cells that have a limited life span. However, mutations of genes that influence apoptosis may result in persistently mutated cells, which can continue to divide and proliferate.
- Tumor suppressor genes slow or even stop cell division. They can be affected by mutations that cause them to have reduced actions. In human cancer cells many altered tumor suppressor genes have been identified.
- Cancer therapy focuses on the confinement of malignant cells and then destruction. Cancerous tissue is killed by lasers, X-rays, and drugs (-chemotherapy) and removed by surgery. However, many cancers cannot be completely destroyed. Some of the adverse effects of cancer therapies include the destruction of normal cells in rapidly growing tissues such as bone marrow. This can lead to anemia, because of the reduced numbers of red blood cells.

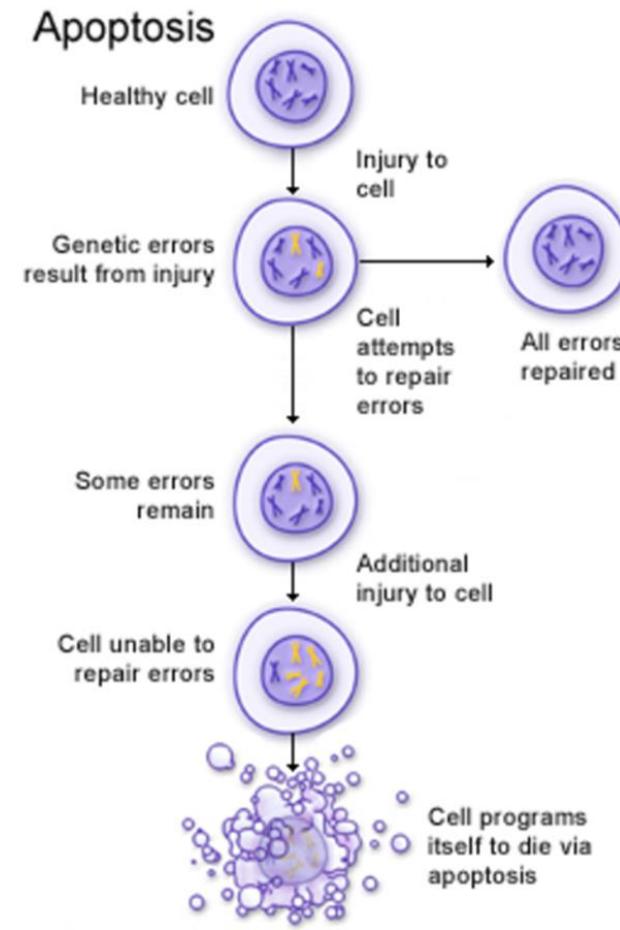
# Checkpoints in the Cell Cycle: Quality Control Systems

$G_0$  = cell not actively involved in division process



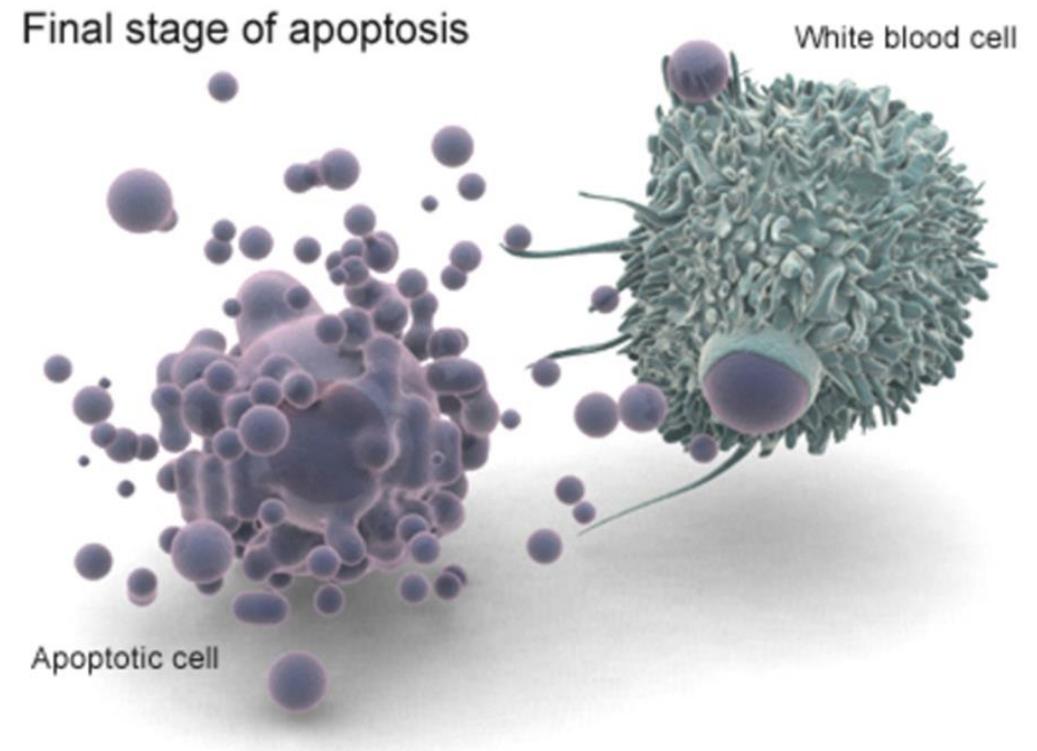
# GENES CONTROL THE GROWTH AND DIVISION OF CELLS

- A variety of genes are involved in the control of cell growth and division.
- The cell replicates itself in an organized, step-by-step fashion known as the cell cycle. Tight regulation of this process ensures that a dividing cell's DNA is copied properly, any errors in the DNA are repaired, and each daughter cell receives a full set of chromosomes. The cell cycle has checkpoints (also called restriction points), which allow certain genes to check for problems and halt the cycle for repairs if something goes wrong.
- If a cell has an error in its DNA that cannot be repaired, it may undergo **self-destruction (apoptosis)**.



# GENES CONTROL THE GROWTH AND DIVISION OF CELLS

- Apoptosis is a common process throughout life that helps the body get rid of cells that no longer work or that it doesn't need.
- Cells that undergo apoptosis break apart and are recycled by a type of white blood cell called a macrophage.
- Apoptosis protects the body by removing genetically damaged cells that could lead to cancer, and it plays an important role in the development and maintenance of tissues.

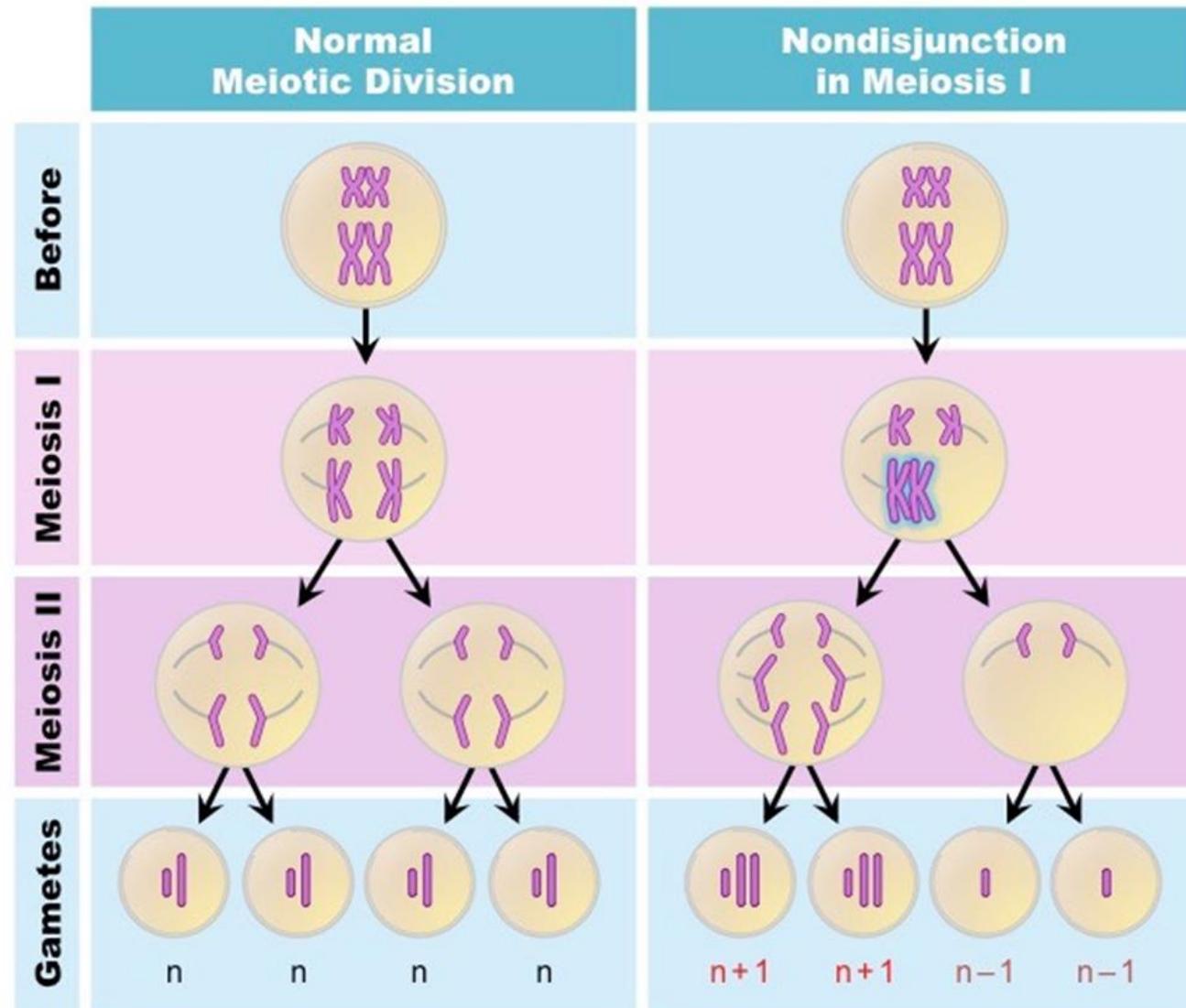


# CELL DIVISION ERRORS

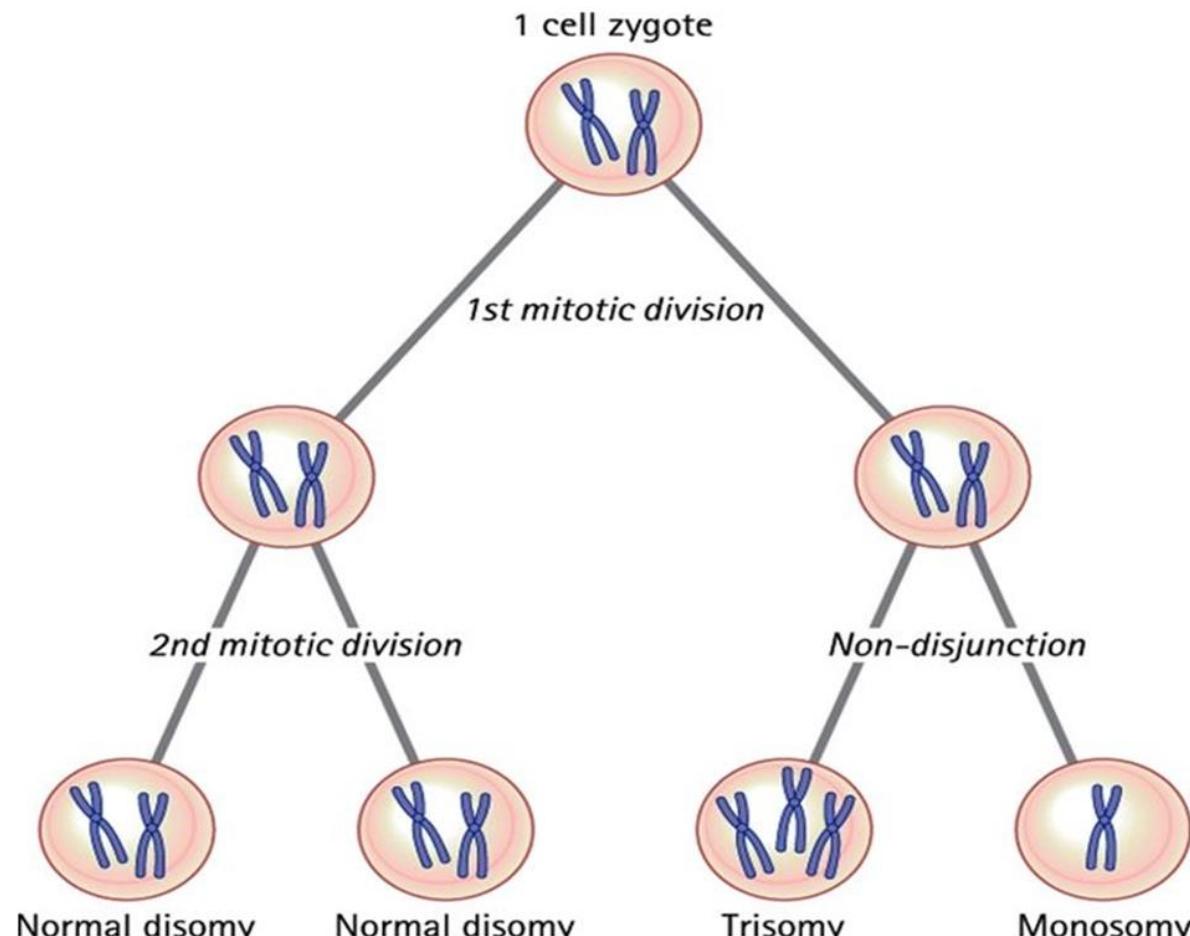
- Mistakes during meiosis lead to the production of daughter cells with too many or too few chromosomes, a feature known as aneuploidy. Nearly all aneuploidies that arise due to mistakes in meiosis or during early embryonic development are lethal, with some exceptions like trisomy 21 in humans.



# CHANGES IN THE NUMBER OF CHROMOSOMES



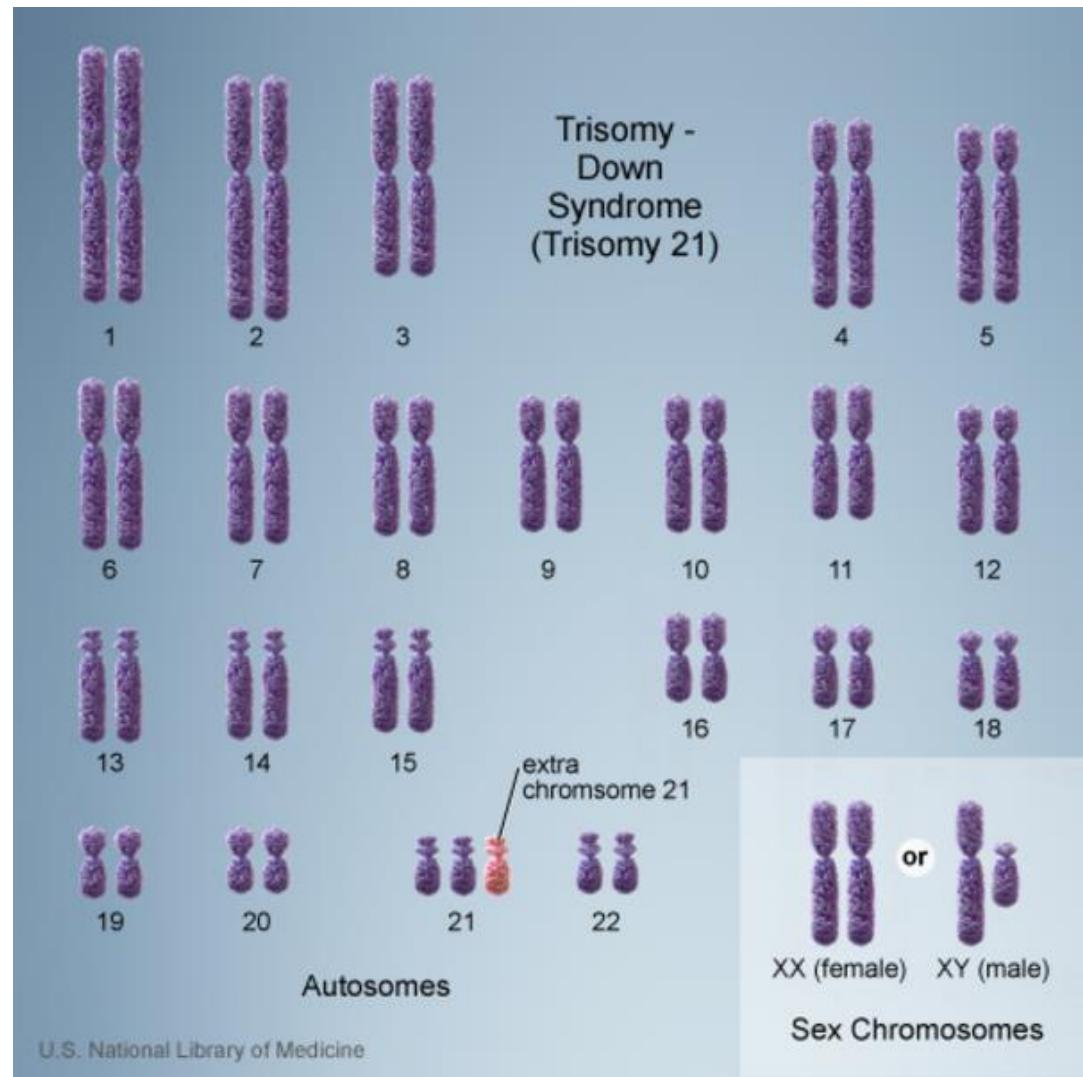
# CHANGES IN THE NUMBER OF CHROMOSOMES



Mechanism of Trisomy

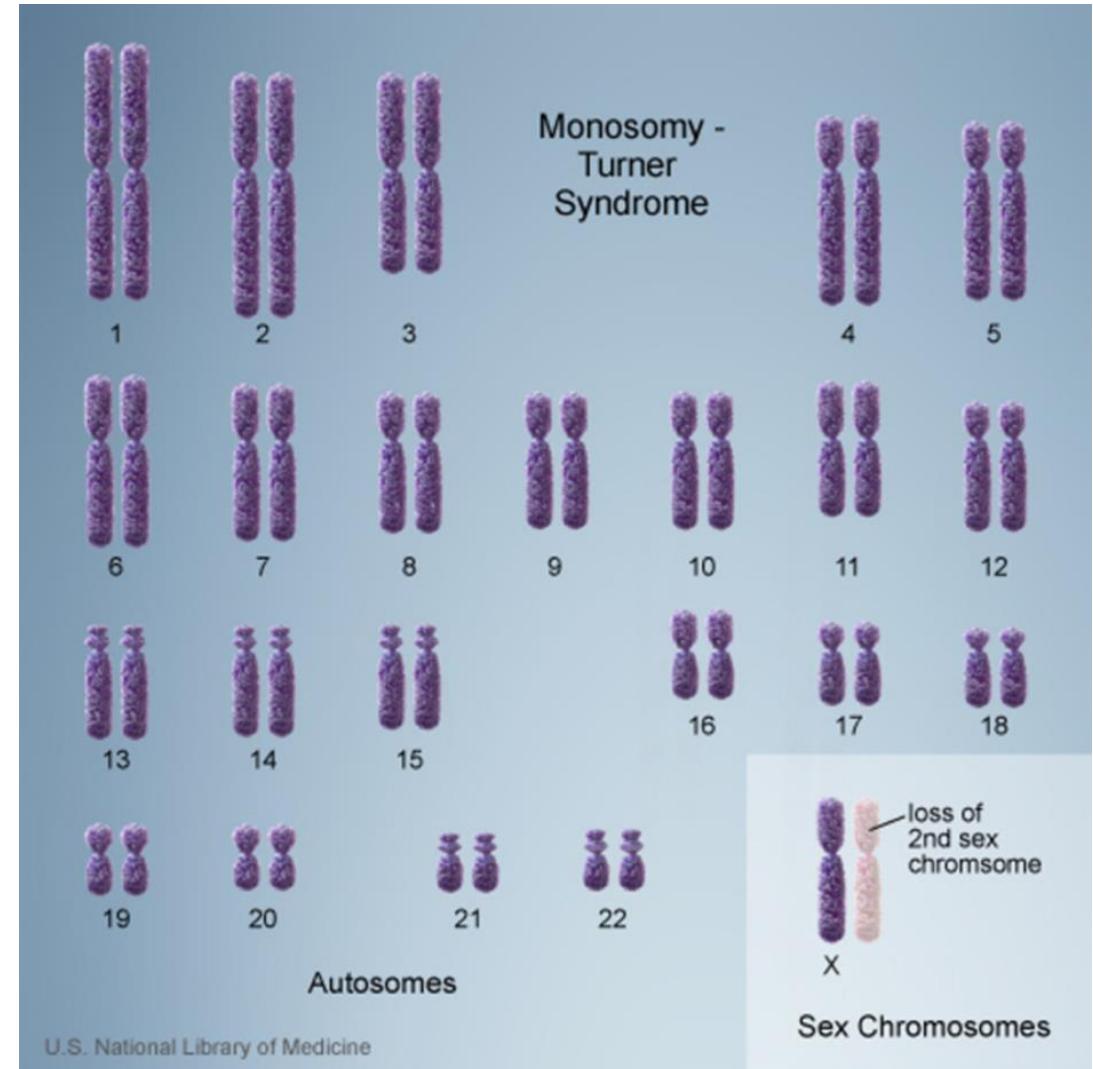
# CHANGES IN THE NUMBER OF CHROMOSOMES

- Human cells normally contain 23 pairs of chromosomes, for a total of 46 chromosomes in each cell. A change in the number of chromosomes can cause problems with growth, development, and function of the body's systems.
- These changes can occur during the formation of reproductive cells (eggs and sperm), in early fetal development, or in any cell after birth. A gain or loss in the number of chromosomes from the normal 46 is called aneuploidy.
- A common form of aneuploidy is trisomy, or the presence of an extra chromosome in cells. People with trisomy have three copies of a particular chromosome in cells instead of the normal two copies. Down syndrome (also known as trisomy 21) is an example of a condition caused by trisomy. People with Down syndrome typically have three copies of chromosome 21 in each cell, for a total of 47 chromosomes per cell.



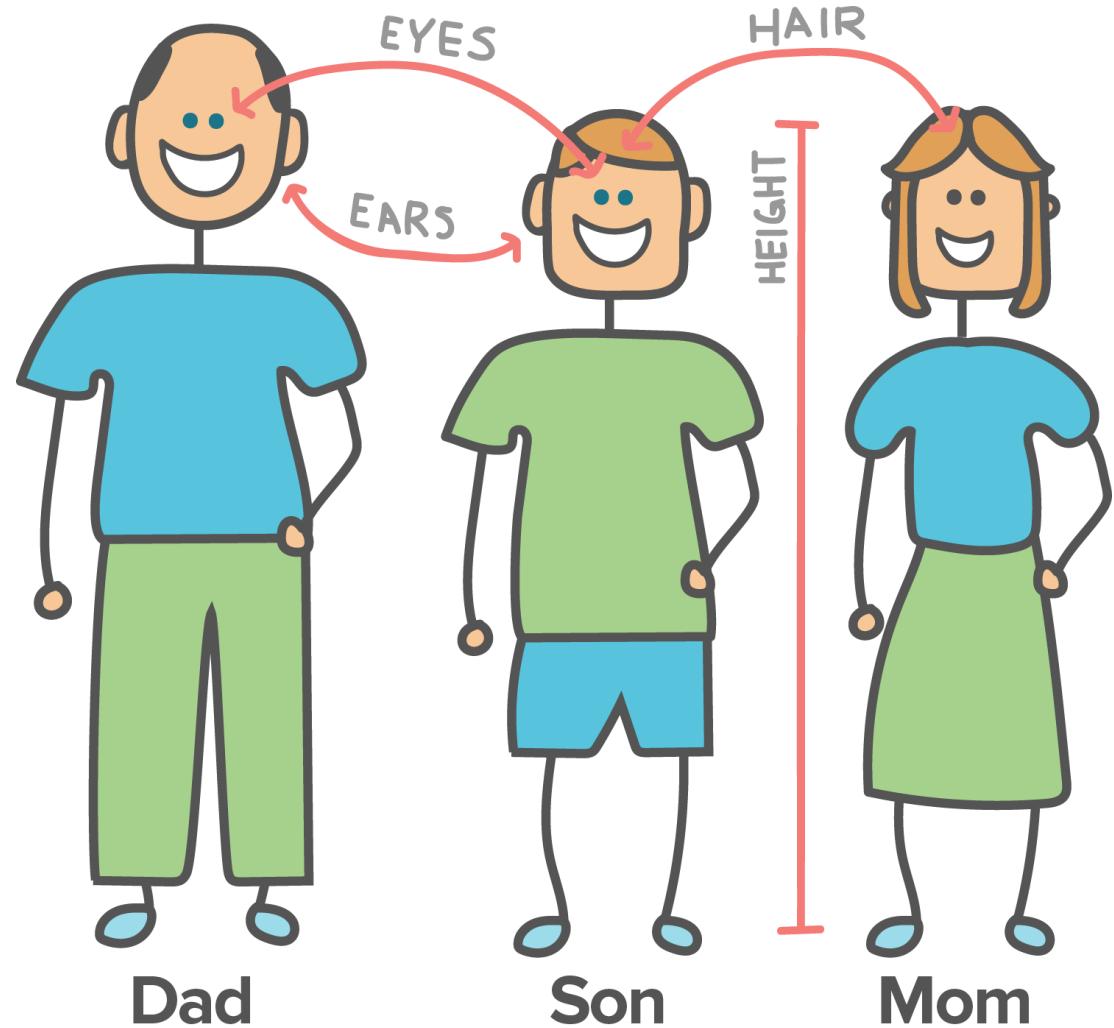
# CHANGES IN THE NUMBER OF CHROMOSOMES - MONOSOMY

- Monosomy, or the loss of one chromosome in cells, is another kind of aneuploidy. People with monosomy have one copy of a particular chromosome in cells instead of the normal two copies.
- Turner syndrome (also known as monosomy X) is a condition caused by monosomy. Women with Turner syndrome usually have only one copy of the X chromosome in every cell, for a total of 45 chromosomes per cell.
- Turner syndrome is a chromosomal condition that affects development in females. The most common feature of Turner syndrome is short stature. Most ovarian tissue degenerates before birth. Many affected girls do not undergo puberty unless they receive hormone therapy, and most are unable to conceive (infertile).



# HEREDITY

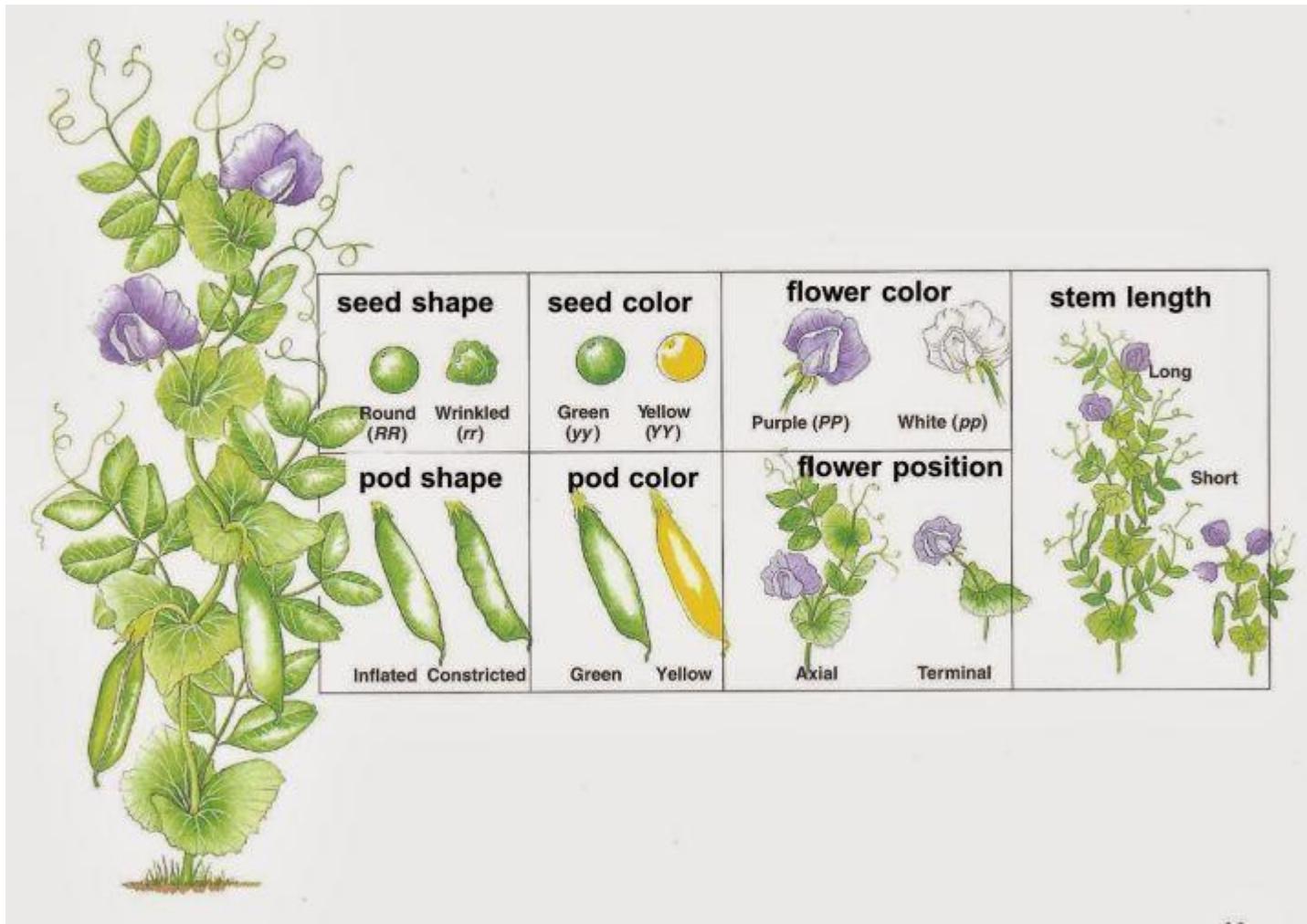
- Heredity refers to the genetic heritage passed down by the biological parents.
- More specifically, it is the transmission of traits from one generation to the next.

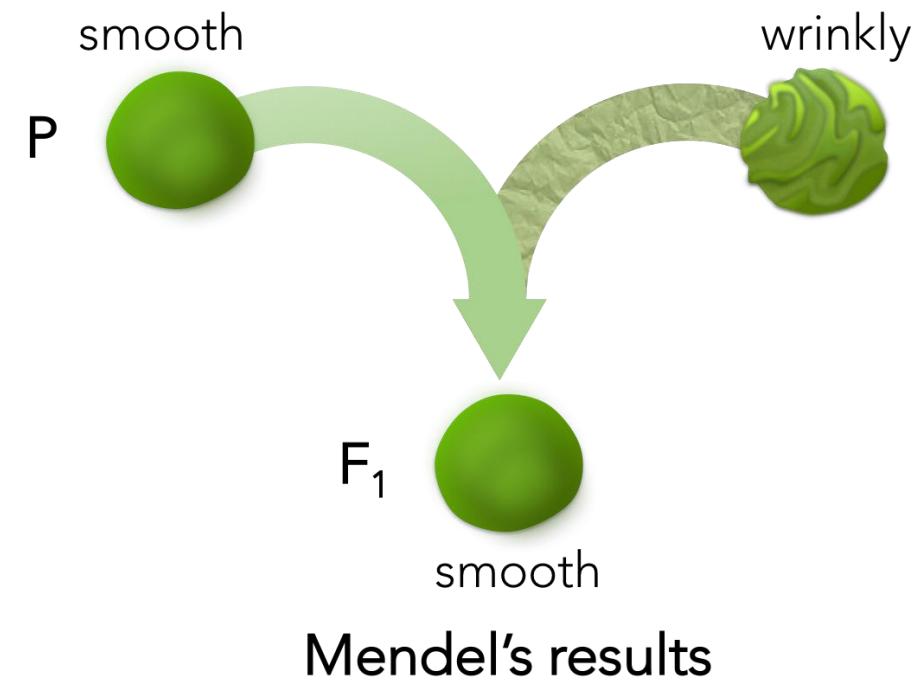
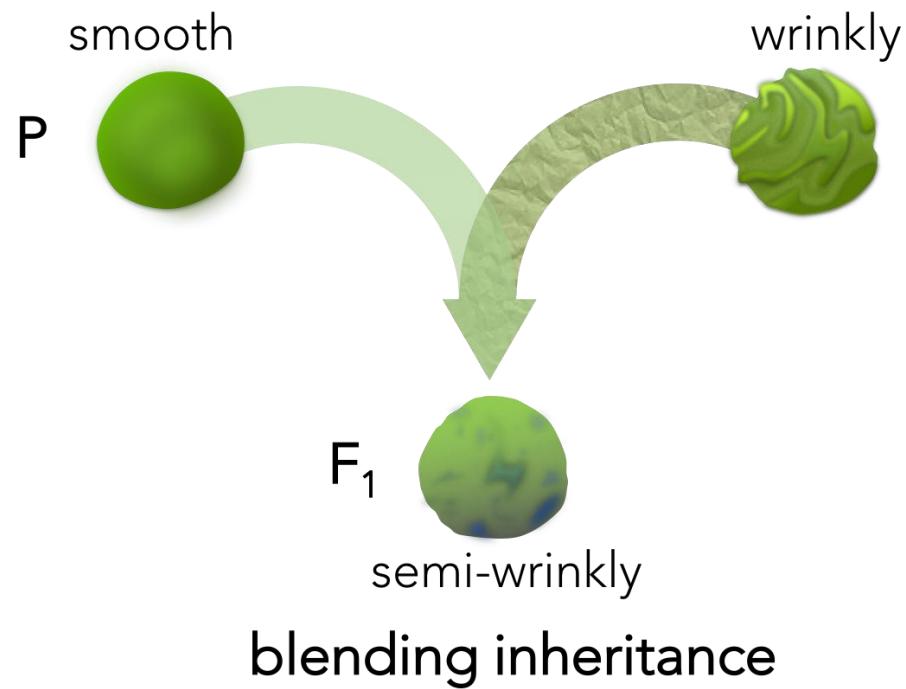


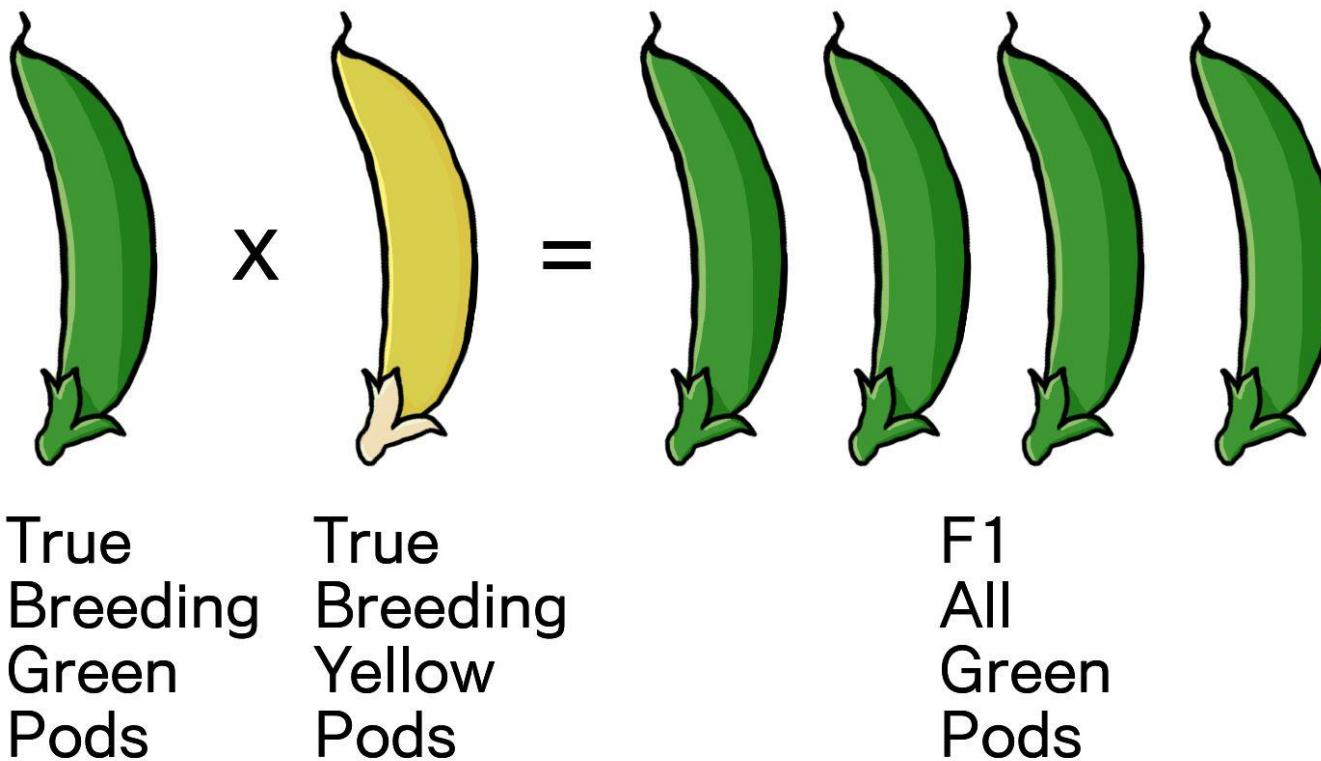
# Who was Gregor Mendel?



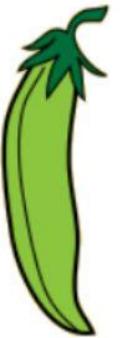
- Gregor Mendel was an Austrian monk, who lived in the 1800's.
- Mendel conducted thousands of experiments on pea plants to see how traits (shape, color) were passed from generation to generation.
- Mendel is known as the "Father of Genetics" for figuring out the basic rules of how traits are inherited.







True  
Breeding  
Green  
Pod

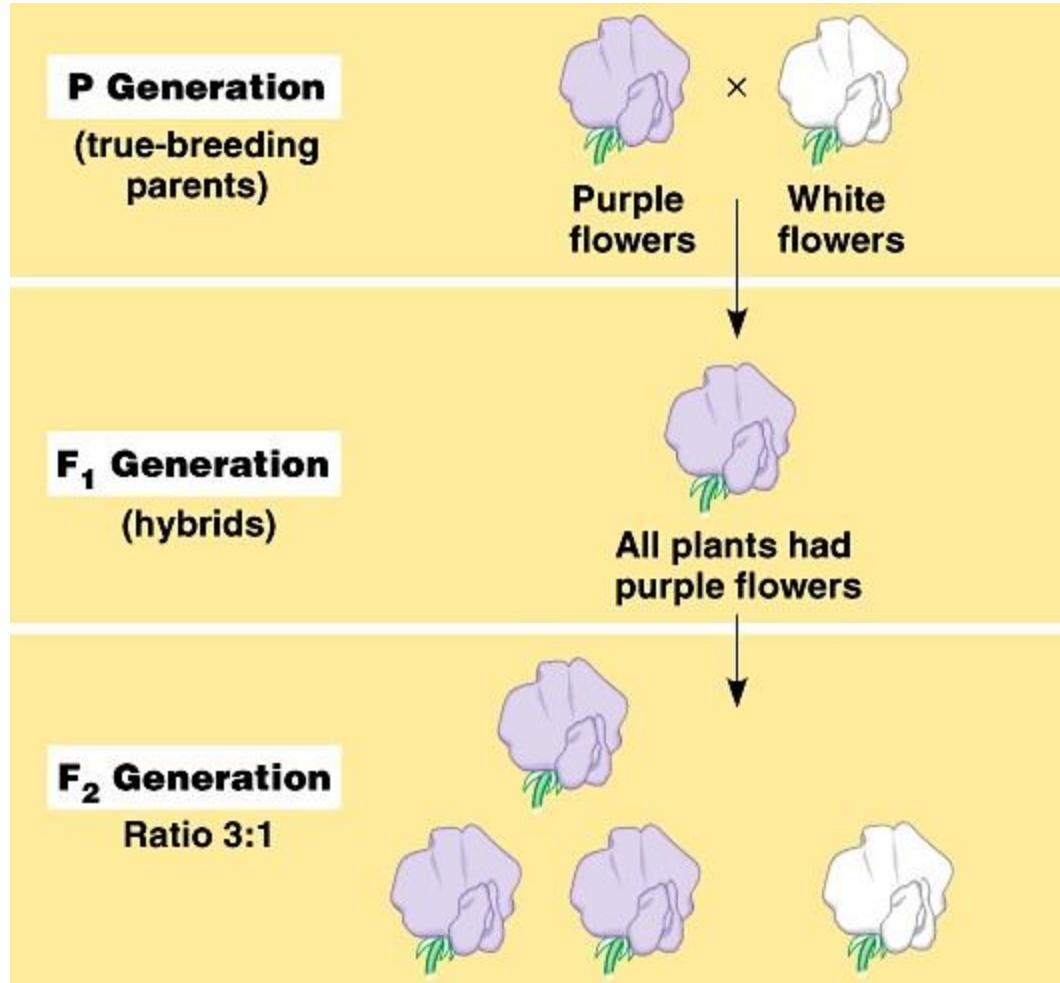


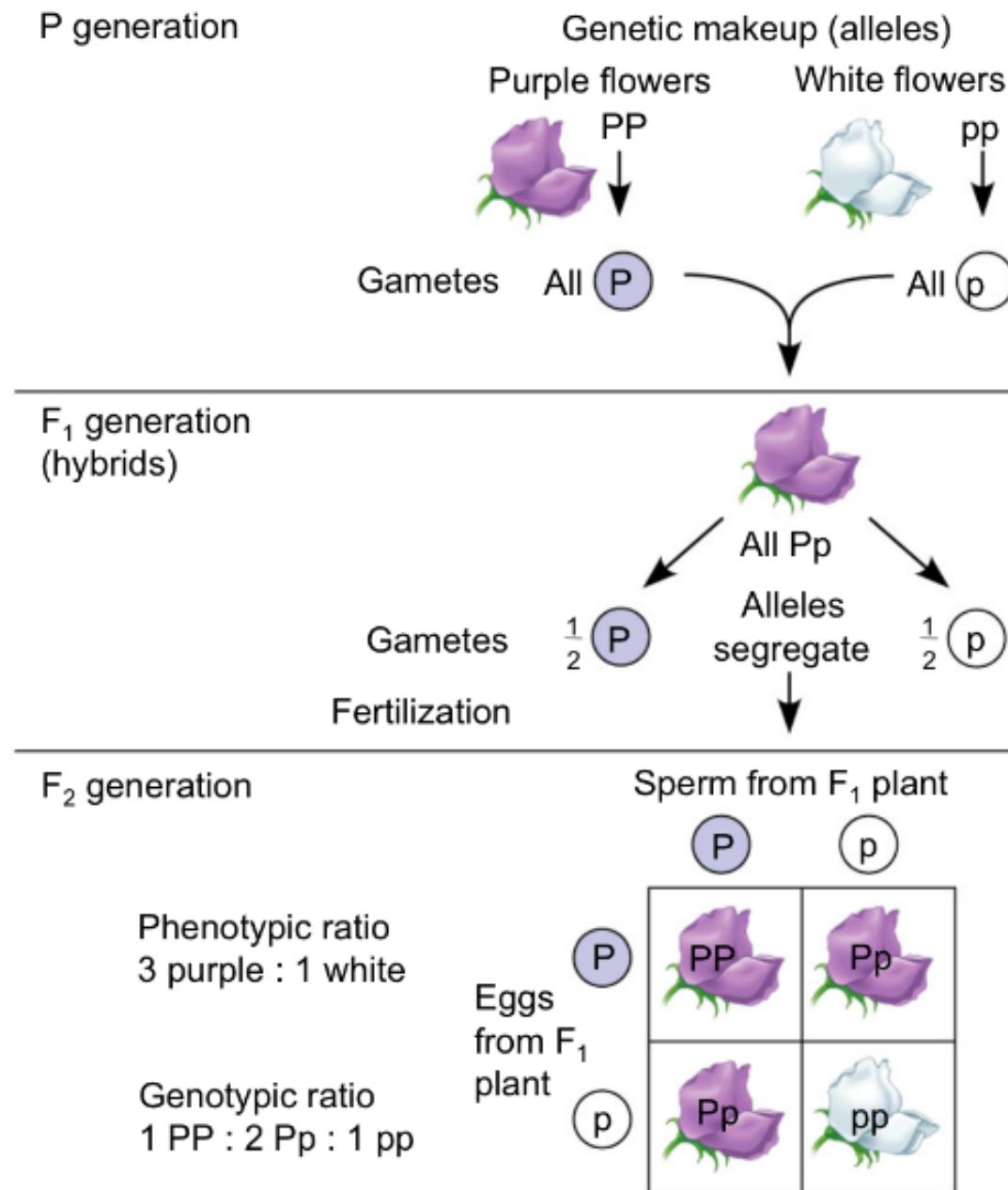
G

	g	g
G	Gg	Gg
G	Gg	Gg



Breeding  
Yellow Pod

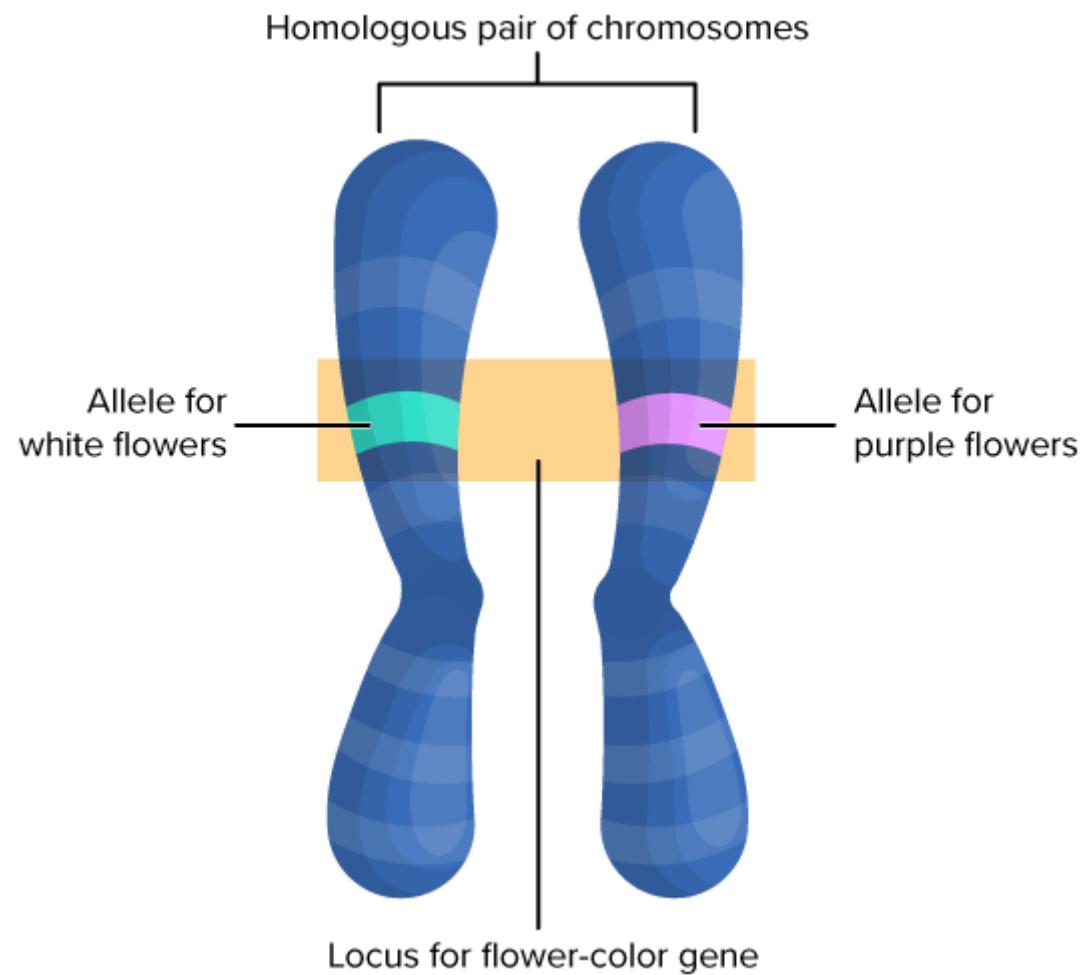


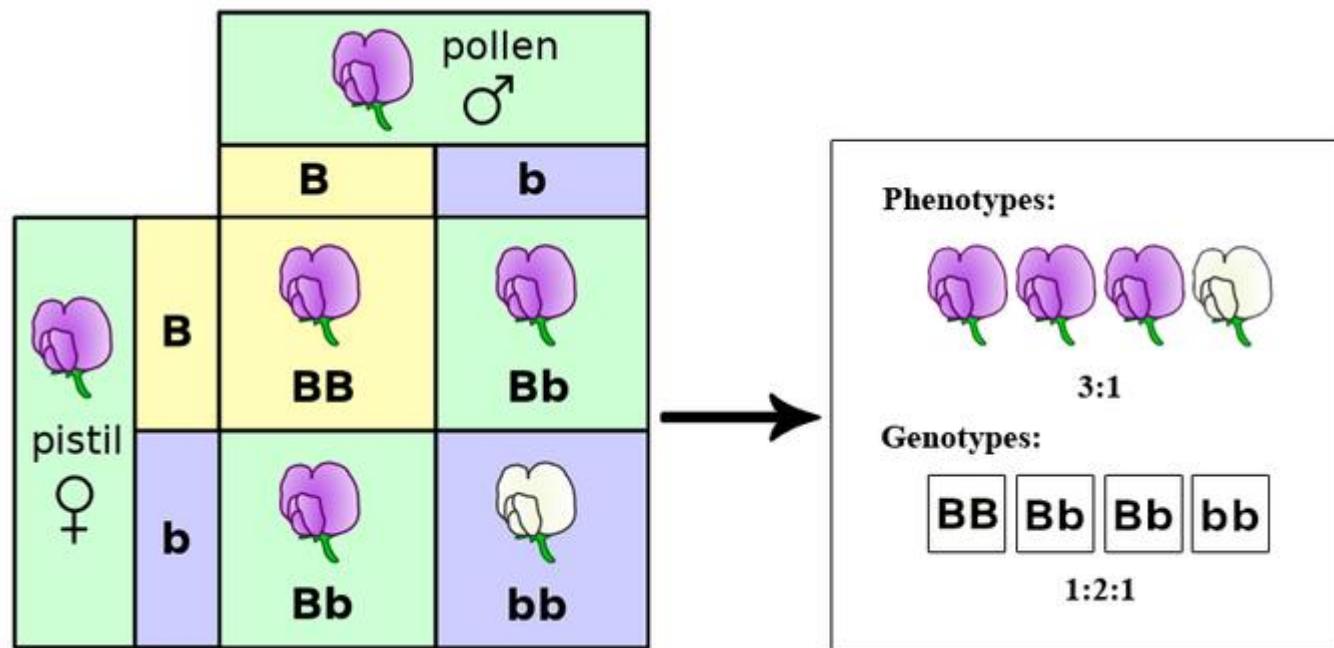


## ALLELES

- In his experiments studying the trait for flower color, Mendel observed that the flowers of each pea plant were either purple or white—but never an intermediate between the two colors. These different, but discrete versions of the same gene are called **alleles**.
- These alternative versions are usually very similar and differ only by a few bases. However, this small difference can be enough to effect the associated trait.

		pollen ♂	
		B	b
pistil ♀	B	BB	Bb
	b	Bb	bb





## Genotype vs. Phenotype

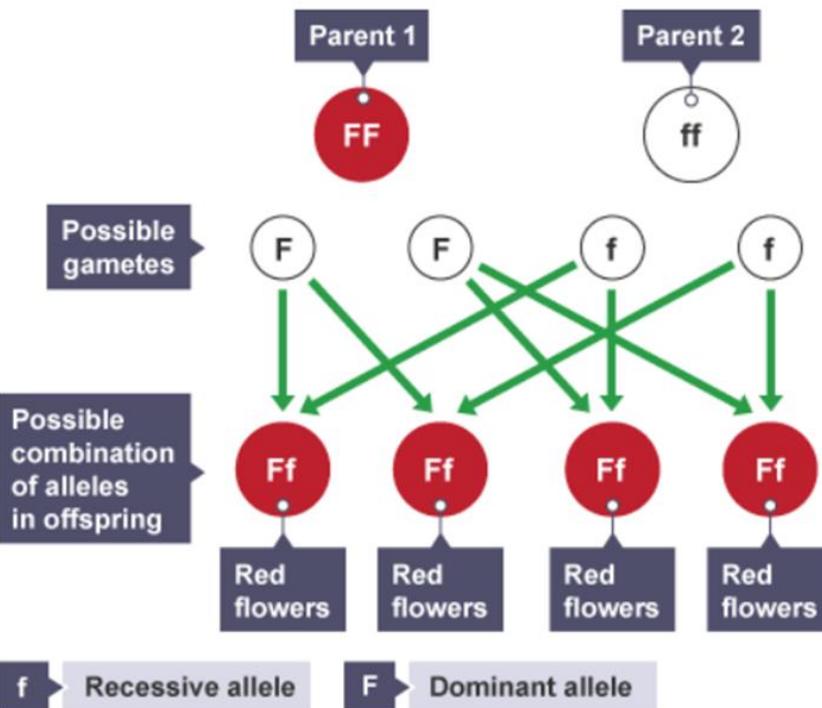
Genotype	Phenotype
BB <b>Homozygous dominant</b>	
Bb <b>Heterozygous</b>	
bb <b>Homozygous recessive</b>	

	Flower Colour	Plant Height	Seed Color	Seed Shape	Pod Colour	Pod Shape	Flower Position
Dominant Trait							
Recessive Trait							
	Purple	Tall	Yellow	Round	Green	Inflated (full)	Axial
	White	Short	Green	Wrinkled	Yellow	Constricted (flat)	Terminal

# LAW OF DOMINANCE

The genetic diagram shows all of the possible alleles for a particular characteristic.

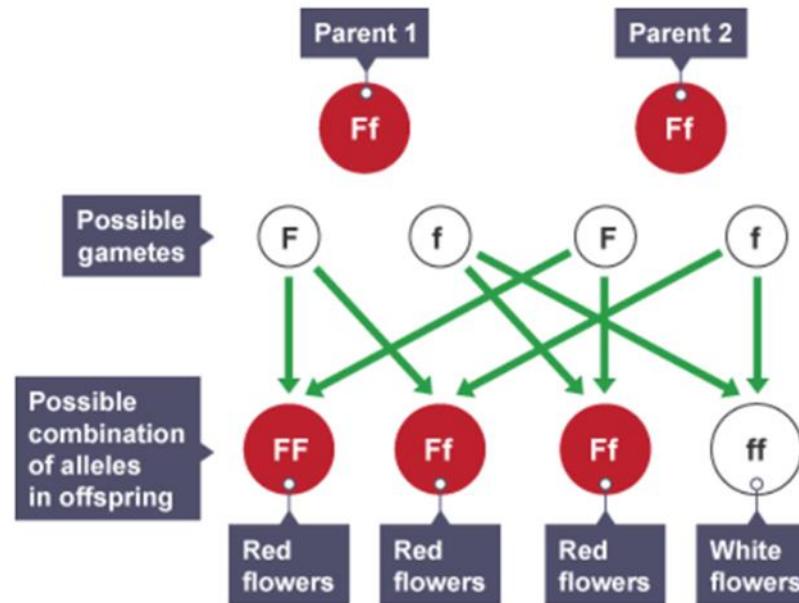
Dominant alleles are capital letters, while the recessive alleles are lower-case letters.



This genetic diagram shows the outcome of Mendel's first cross. All the offspring have red flowers (100%), even though they are heterozygotes and carry the recessive allele for white flowers (Ff).

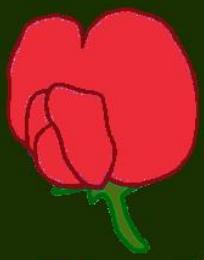
# LAW OF SEGREGATION

- In the following generation, the white flowers re-emerge in plants where are no dominant F genes, and both members of the pair for flower colour are the recessive, ff!



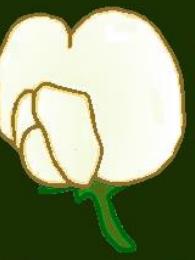
### Principle of Uniformity

P - Generation



Phenotype

Genotype



F1 - Generation

Phenotype



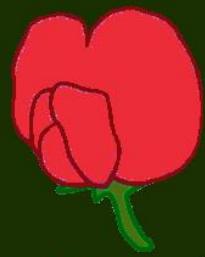
Genotype

● = dominant allele for red flower colour

○ = recessive allele for white = missing pigmentation

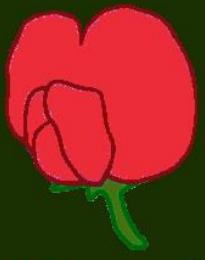
### Segregation of alleles

F1 - Generation



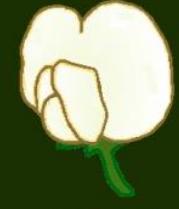
Phenotype

Genotype



F2 - Generation

Phenotype



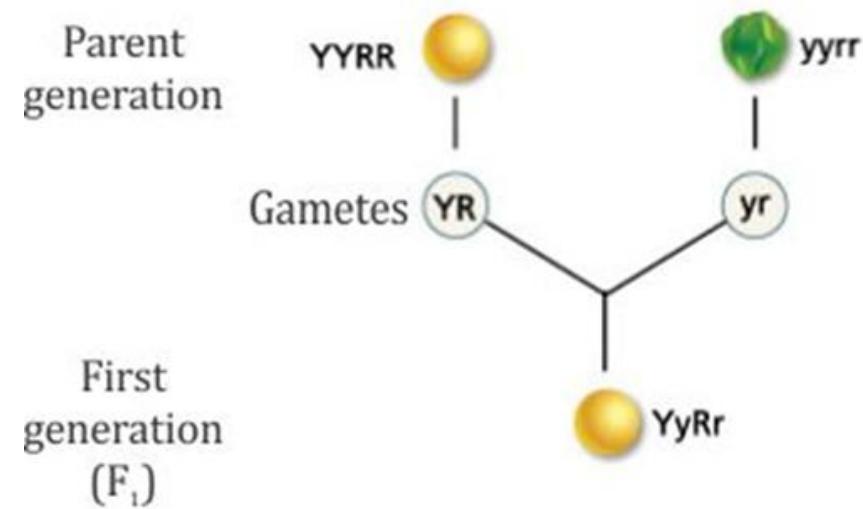
Genotype

The recessive trait from the P-Generation phenotypically reappears in the individuals that are homozygous in the recessive genetic trait.

# MENDEL'S LAWS

- In the case of the pea, which is a **diploid species**, each individual plant has **two copies of each gene**, one copy inherited from each parent.
- Many species, including humans, have this pattern of inheritance. Diploid organisms with **two copies of the same allele** of a given gene are called **homozygous** at that gene locus, while organisms with **two different alleles** of a given gene are called **heterozygous**.
- The set of **alleles** for a given organism is called its **genotype**, while the observable **traits** of the organism are called its **phenotype**. When organisms are heterozygous at a gene, often one allele is called **dominant** as its qualities dominate the phenotype of the organism, while the other allele is called **recessive** as its qualities recede and are not observed. Some alleles do not have complete dominance and instead have incomplete dominance by expressing an intermediate phenotype, or codominance by expressing both alleles at once.
- When a pair of organisms reproduce sexually, their offspring randomly inherit one of the two alleles from each parent. These observations of discrete inheritance and the segregation of alleles are collectively known as Mendel's first law.

# LAW OF INDEPENDENT ASSORTMENT



Second generation  
( $F_2$ )

YR 1/4	Yr 1/4	yR 1/4	yr 1/4
YYRR	YYRr	YyRR	YyRr
YYRr	YYrr	YyRr	Yyrr
YyRR	YyRr	yyRR	yyRr
YyRr	Yyrr	yyRr	yyrr

9/16 Yellow-round  
3/16 Green-round  
3/16 Yellow-wrinkled  
1/16 Green-wrinkled

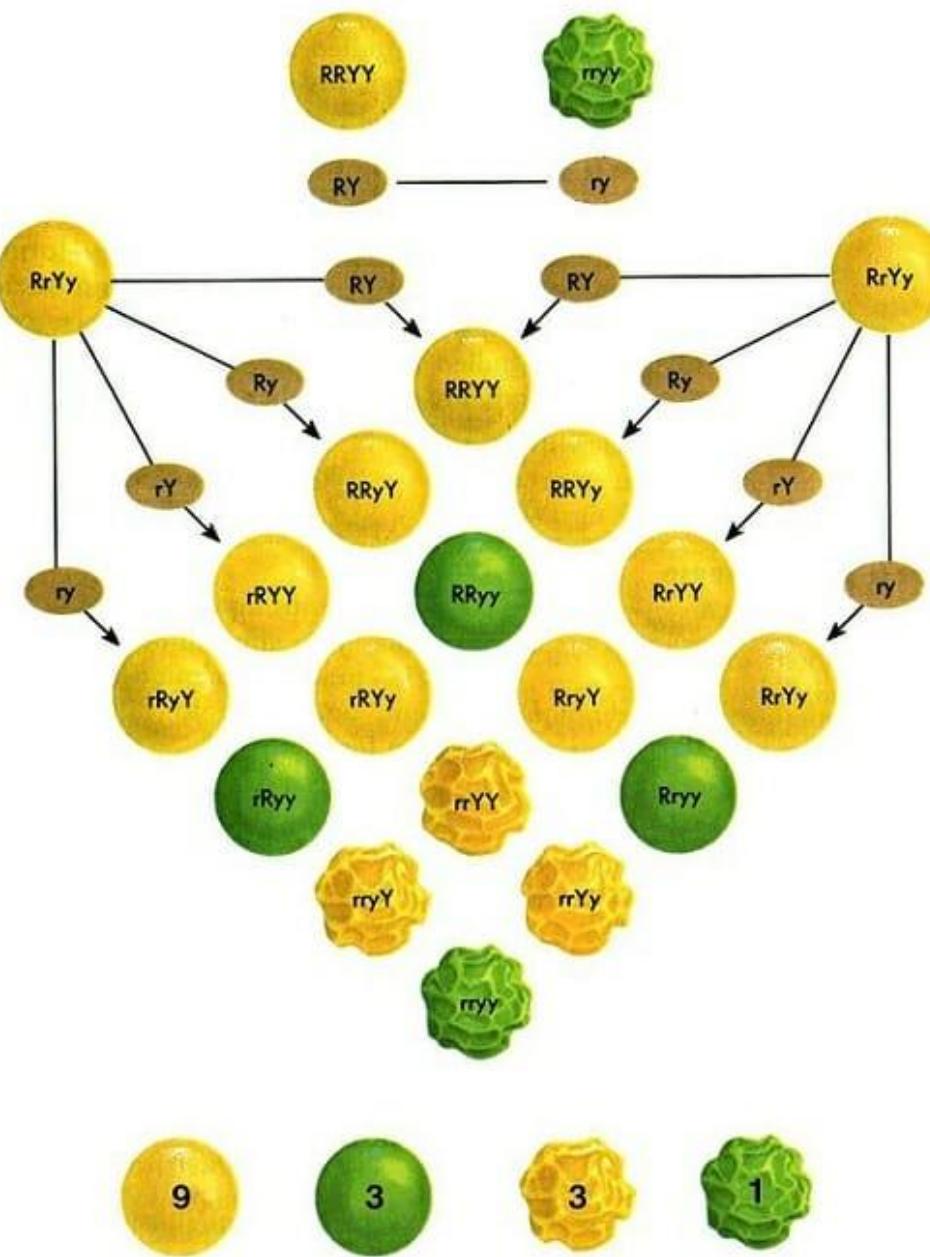
Y = dominant allele for seed colour (yellow)

y = recessive allele for seed colour (green)

R = dominant allele for seed shape (round)

r = recessive allele for seed shape (wrinkled)

An illustration of dihybrid cross



## MENDEL'S LAWS

**Law of Dominance:** if the two alleles at a locus differ, then one, the **dominant allele**, determines the organism's appearance; the other, the **recessive allele**, has no noticeable effect on the organism's appearance

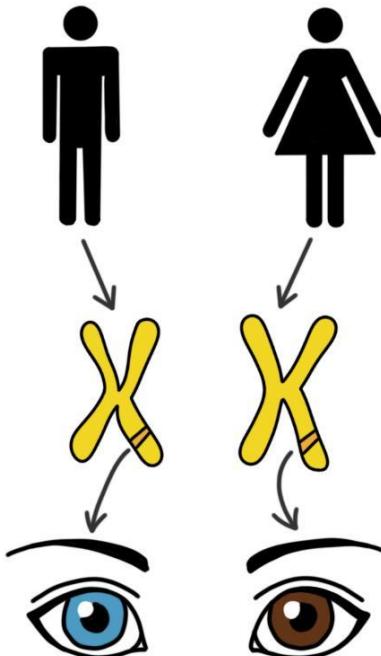
**Law of Segregation:** the two alleles for a heritable character separate (segregate) during gamete formation and end up in different gametes

**Law of Independent Assortment:** each pair of alleles segregates independently of other pairs of alleles during gamete formation

**humans get 2 copies of every gene, one copy from each parent**

**the 2 copies don't have to be identical**

**For example:  
if a gene contains  
information for eye color,  
one allele might code for  
blue eyes and another  
might code for brown eyes**

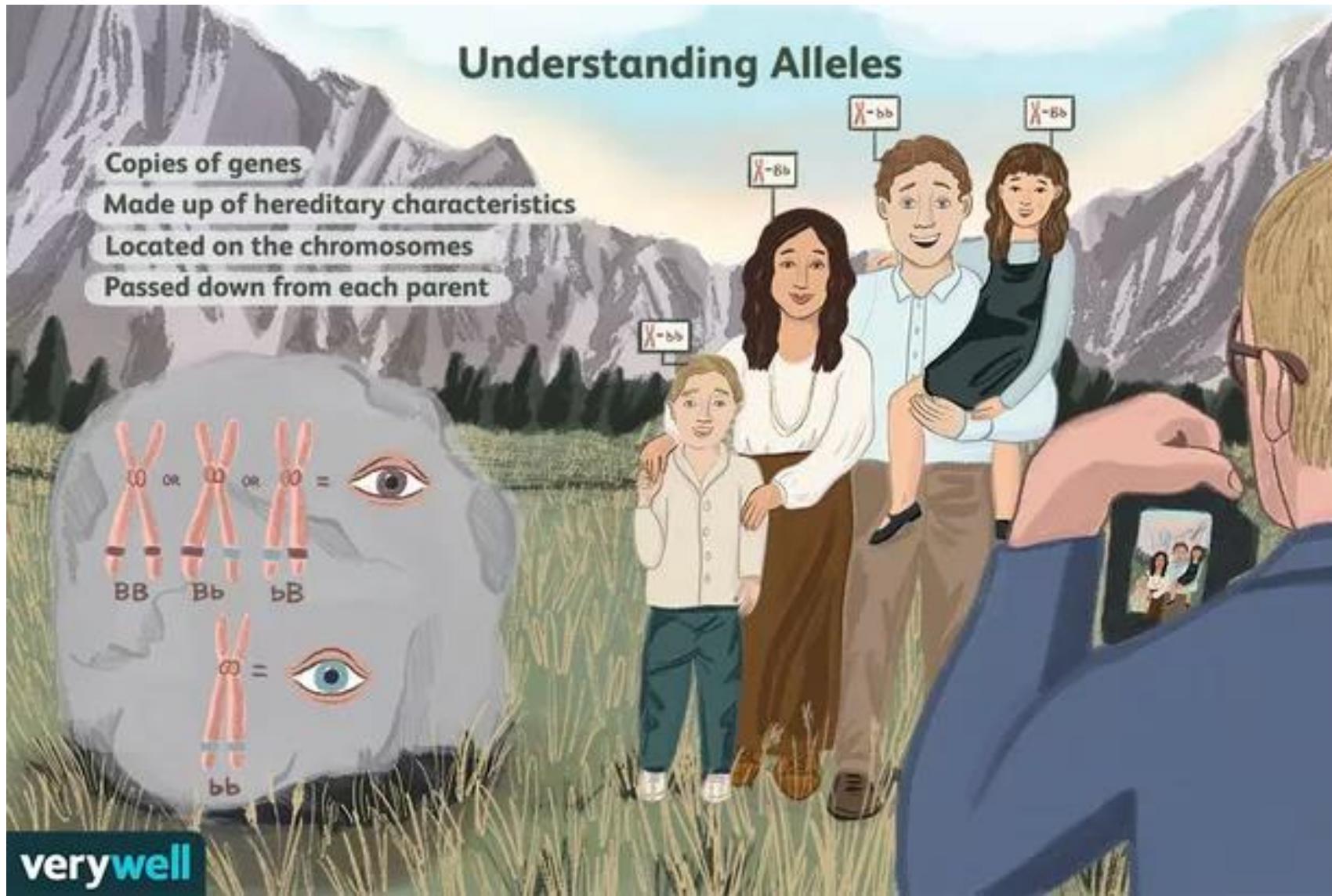


**each variation of a gene is called an allele**

**alleles create diversity**

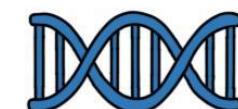
## Understanding Alleles

Copies of genes  
Made up of hereditary characteristics  
Located on the chromosomes  
Passed down from each parent



# WHAT ARE ALLELES?

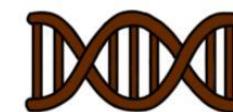
for every gene you can have 2 of the same allele



or you can have 2 different alleles



RECESSIVE ALLELE



DOMINANT ALLELE

a recessive allele will only produce a phenotype if there is no dominant allele present

in other words, you need 2 copies of the recessive allele in order to have the trait that it codes for

recessive alleles are symbolized by lowercase letters

	A	a
A	AA	Aa
a	Aa	aa

only the person with this genotype will have blue eyes



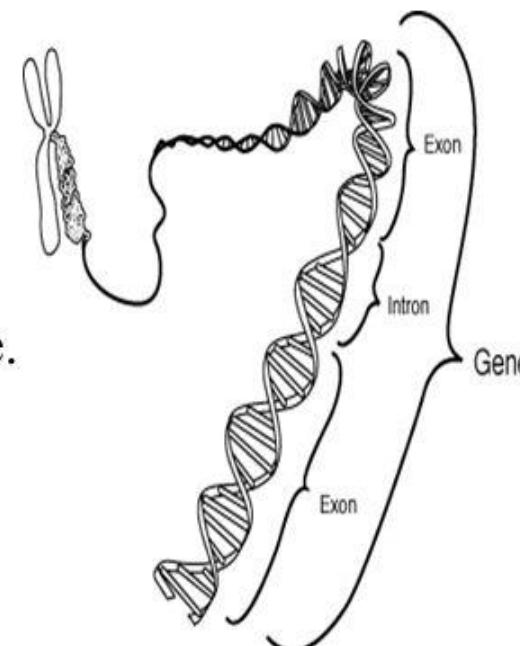
# GENES vs. ALLELES

**Genes:** The section of DNA on the chromosome that codes for the protein production resulting in a specific trait.

Ex. Hair color, Eye color

**Allele:** The different forms of the gene.

Ex.      Blonde or Brown hair  
          Brown or Blue eyes

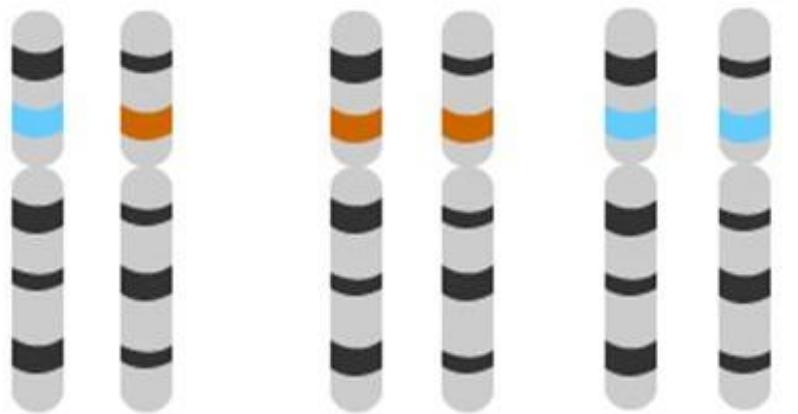


**Allele:** different forms of a gene (for a trait).

**Homozygous:** purebred; identical set of alleles for a trait

**Heterozygous:** hybrid; nonmatching alleles

 = allele for blue eyes (recessive)  
 = allele for brown eyes (dominant)



Individual A:  
Heterozygous  
(will have brown eyes)  
© ABPI 2007

Individual B:  
Homozygous  
(for brown eyes)

Individual C:  
Homozygous  
(for blue eyes)

# PRINCIPLE OF MENDELIAN INHERITANCE



## **Law of Segregation**

The two alleles for each gene are placed in different gametes.

## **Law of Independent Assortment**

The inheritance of one gene doesn't affect the inheritance of any other gene.

## **Law of Dominance**

When two different alleles are present, only one is dominant and will be expressed.

# EXCEPTIONS

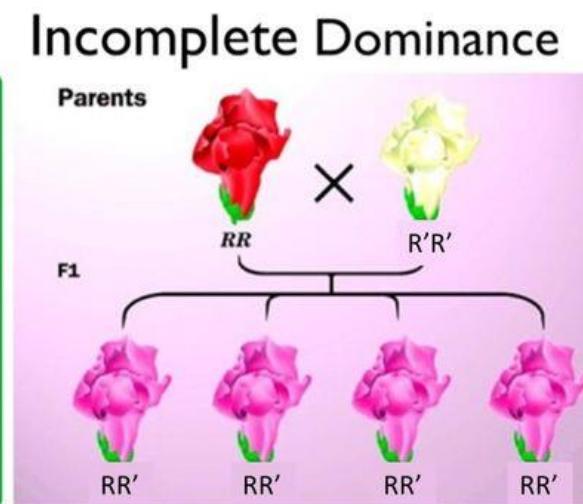
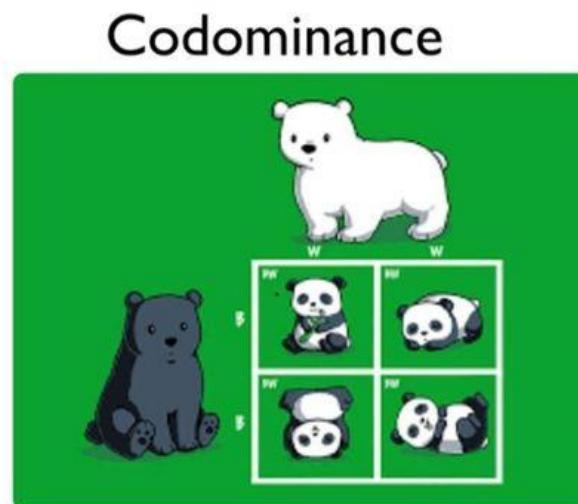
- Intermediate alleles:  
two alleles are  
mixed, blending of  
traits, one allele is  
not completely  
dominant over  
another, phenotype  
is a combination of  
the two alleles
- Codominant alleles:  
are equally  
dominant

Incomplete Dominance –

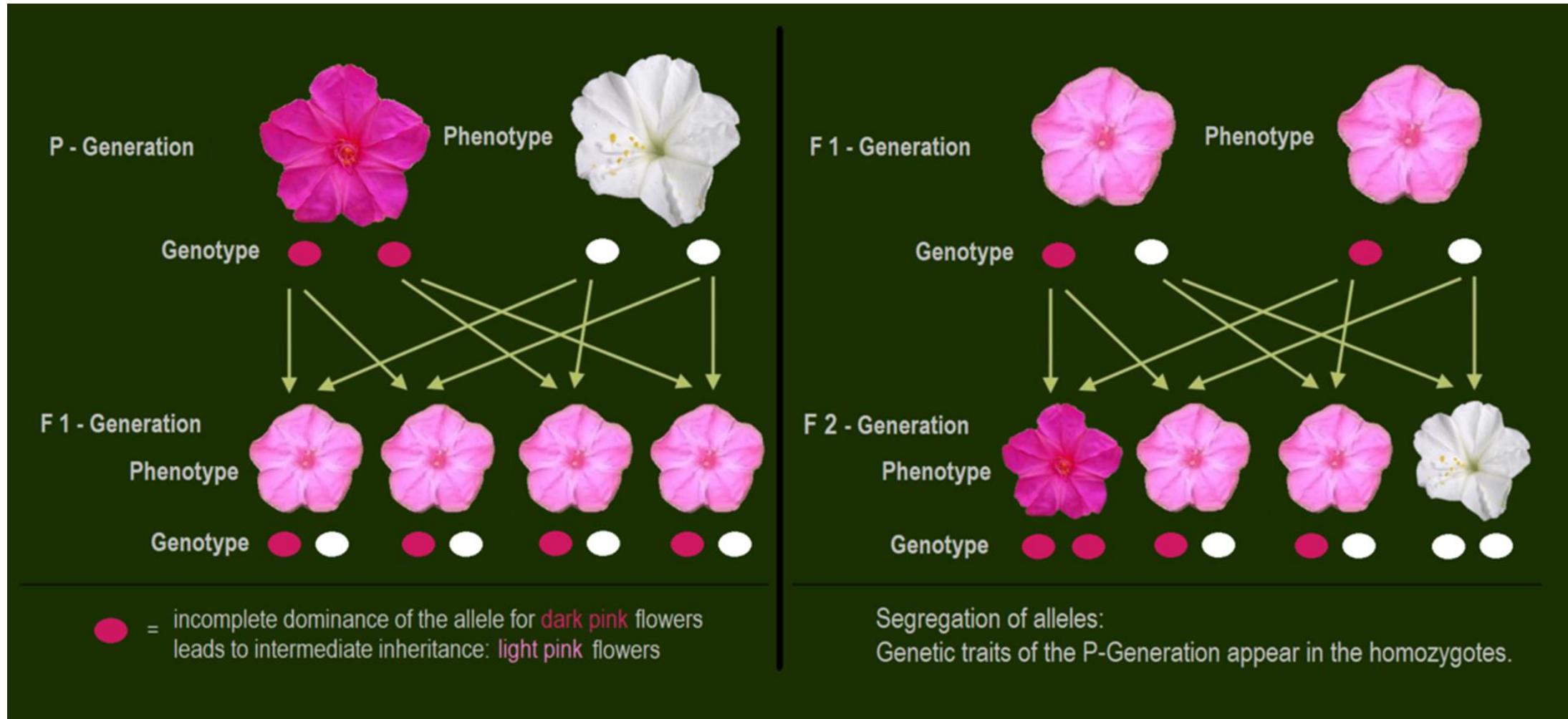
Both dominant genes are expressed as a “Mixture” of the two traits.

Codominance –

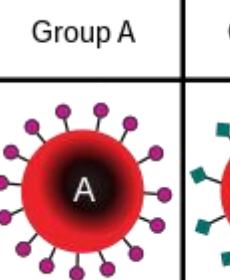
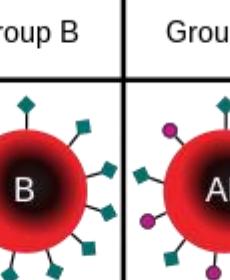
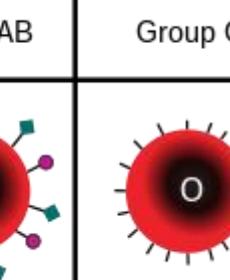
Both dominant genes are expressed fully in the organism



# INTERMEDIATE ALLELES



# MULTIPLE/CODOMINANT ALLELES – HUMAN BLOOD GROUPS

	Group A	Group B	Group AB	Group O
Red blood cell type				

- Some genes have multiple alleles
- Allel A and B are codominant, O is recessive

ABO Blood Group	
Genotype	Phenotype(blood type)
AA	A 
AO	A 
BB	B 
BO	B 
OO	O 
AB	AB 

## **Codominance in Human Blood**

Human blood types are determined by genes that follow the codominance pattern of inheritance.

There are two dominant alleles A and B and one recessive allele O.

**Possible blood types are:**

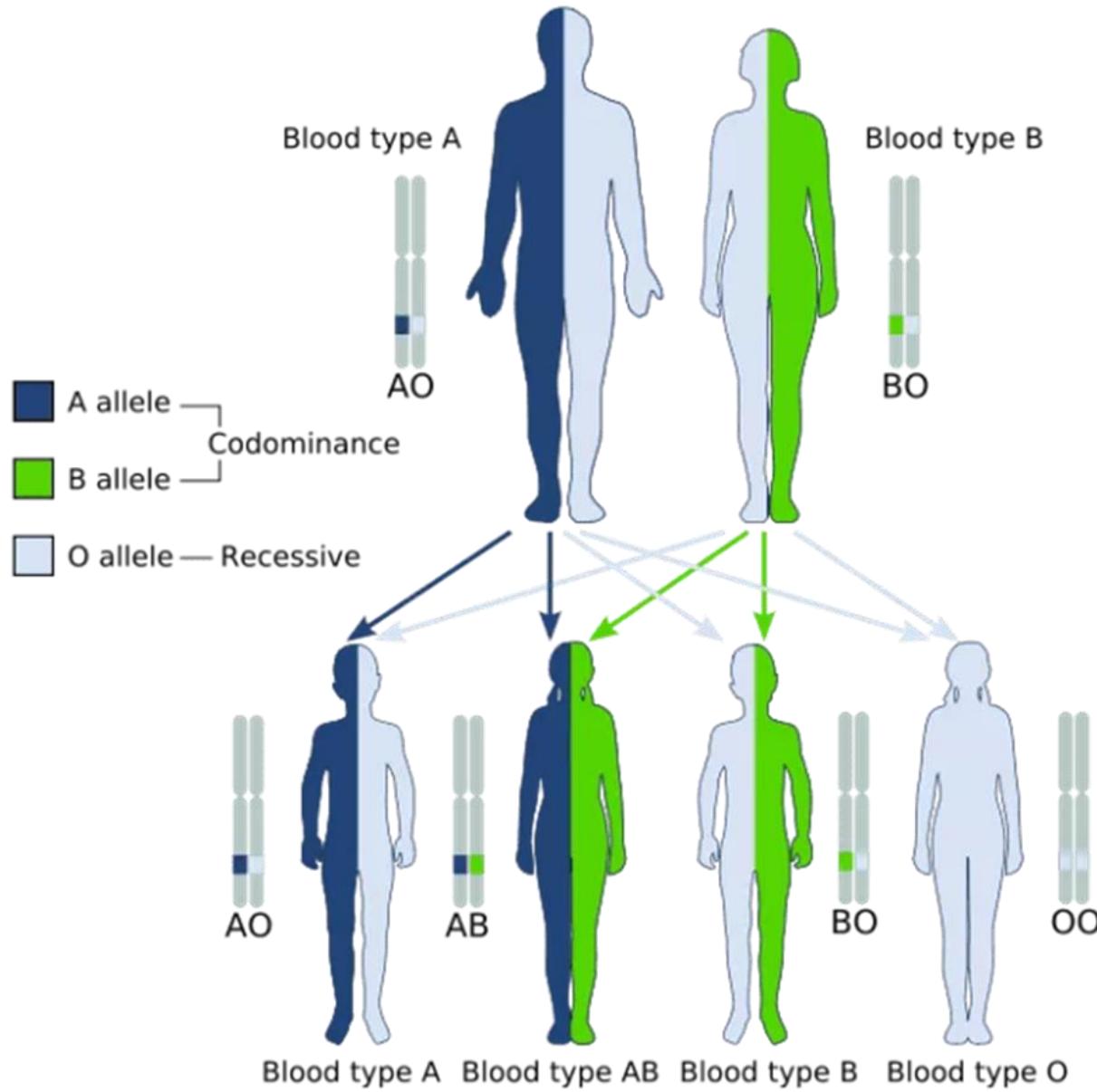
**Type O = OO**

**Type AB = AB**

**Type A = AA (Homozygous) or AO (Heterozygous)**

**Type B = BB (Homozygous) or BO (Heterozygous)**

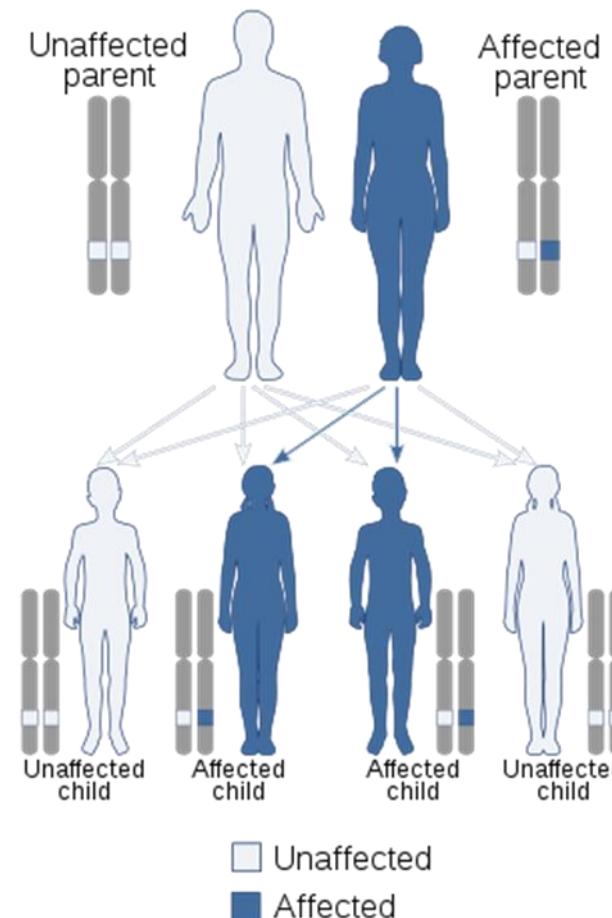
# CODOMINANT ALLELES



# HUMAN GENETIC DISEASES

- A disease trait that is inherited in an autosomal dominant manner can occur in either sex and can be transmitted by either parent.
- A parent with an autosomal dominant condition has a 50% chance of having a child with the condition. This is true for each pregnancy.
- It means that each child's risk for the disease does not depend on whether their sibling has the disease.
- Children who do not inherit the abnormal gene will not develop or pass on the disease.
- Example: Chorea Huntington

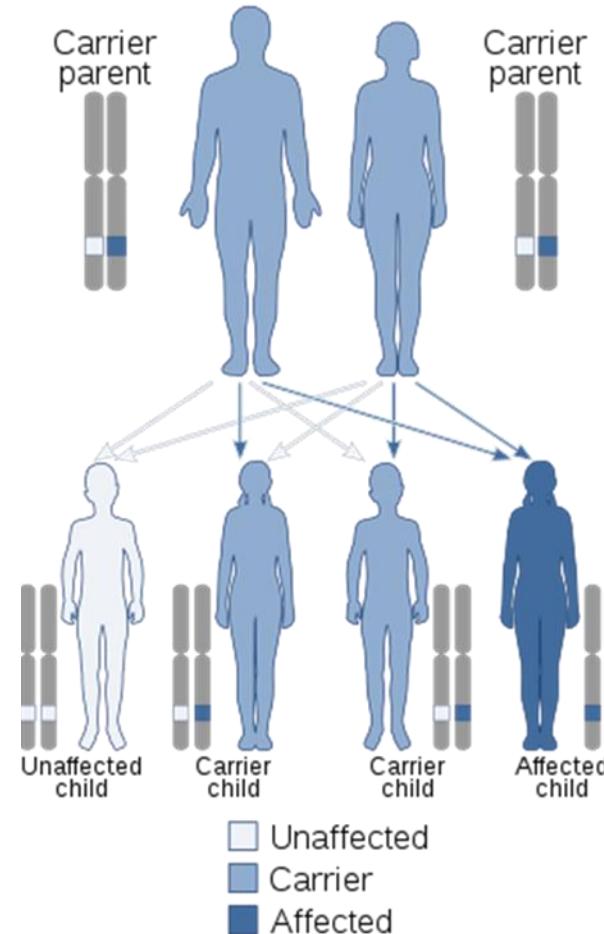
Autosomal dominant



# HUMAN GENETIC DISEASES

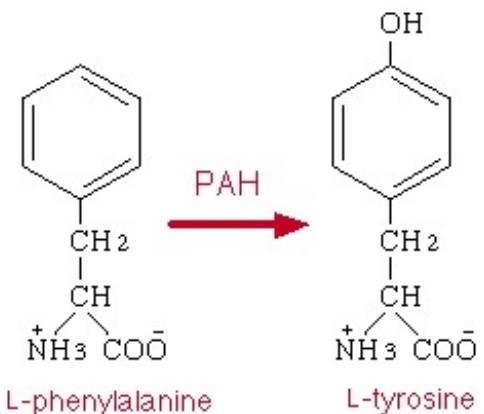
- An autosomal recessive disorder means two copies of an abnormal gene must be present in order for the disease or trait to develop.
- Example:
  - Cystic fibrosis (CF), people with CF produce abnormally thick and sticky mucus that can damage body organs.
  - Phenylketonuria (PKU), aminoacid phenylalanine cannot be metabolized and transformed to tyrosin (leads to intellectual disabilities, newborn screening, diet)

## Autosomal recessive



# PKU

Newborn screening –  
Phenylketonuretic babies  
have mutations in the  
PAH gene (1:12 000)



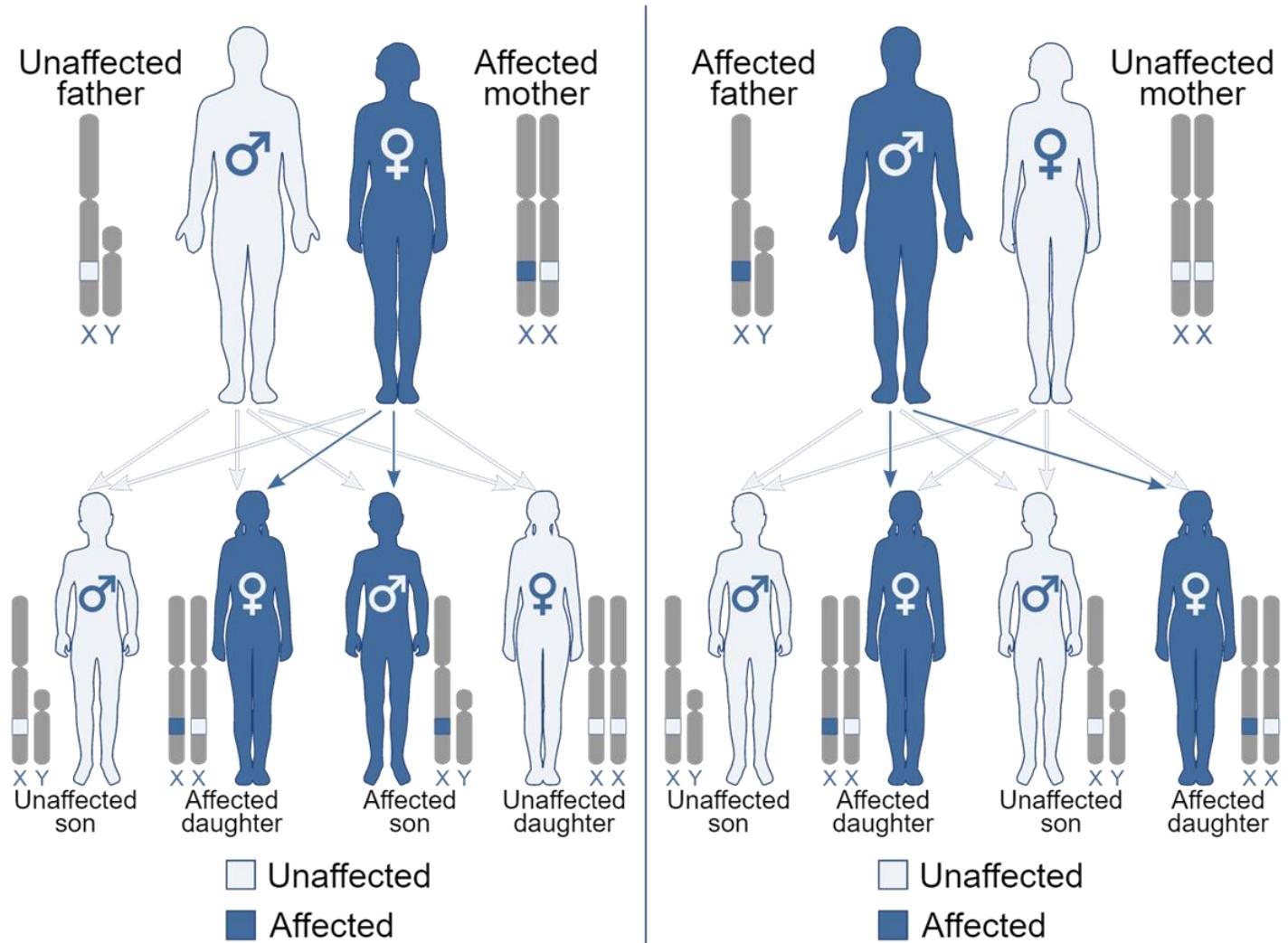
The enzyme phenylalanine hydroxylase converts the amino acid phenylalanine to tyrosine.



# HUMAN GENETIC DISEASES

- X-linked dominant disorders are caused by variants in genes on the X chromosome. In males (who have only one X chromosome), a variant in the only copy of the gene in each cell causes the disorder. In females (who have two X chromosomes), a variant in one of the two copies of the gene in each cell is sufficient to cause the disorder. Females may experience less severe symptoms of the disorder than males.
- A characteristic of X-linked inheritance is that fathers cannot pass X-linked traits to their sons (no male-to-male transmission).
- Example: Rett syndrome (impaired development), male babies mostly die after birth

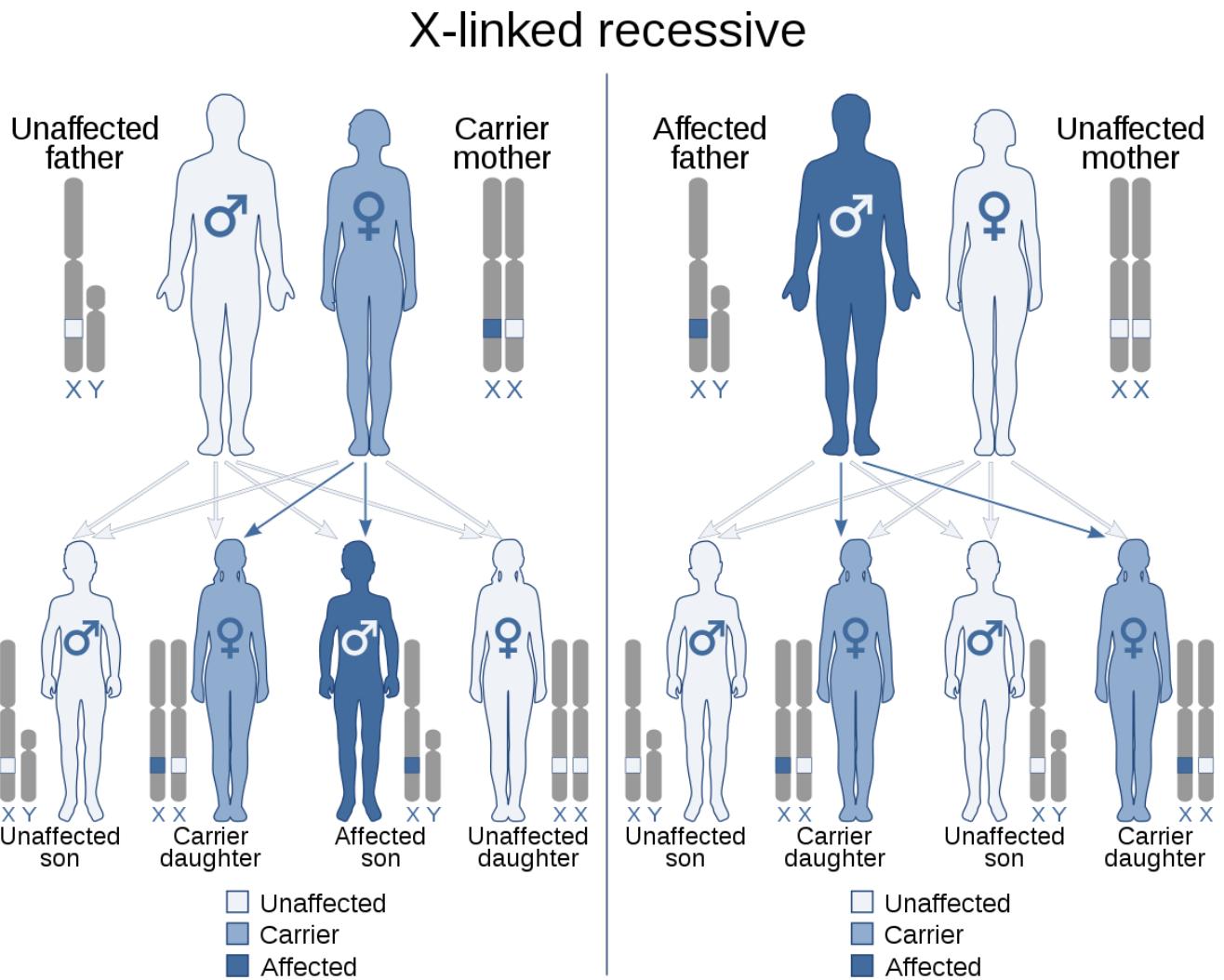
## X-linked dominant



Note: some X-linked dominant disorders are embryonic lethal in males, and most affect females less severely.

# HUMAN GENETIC DISEASES

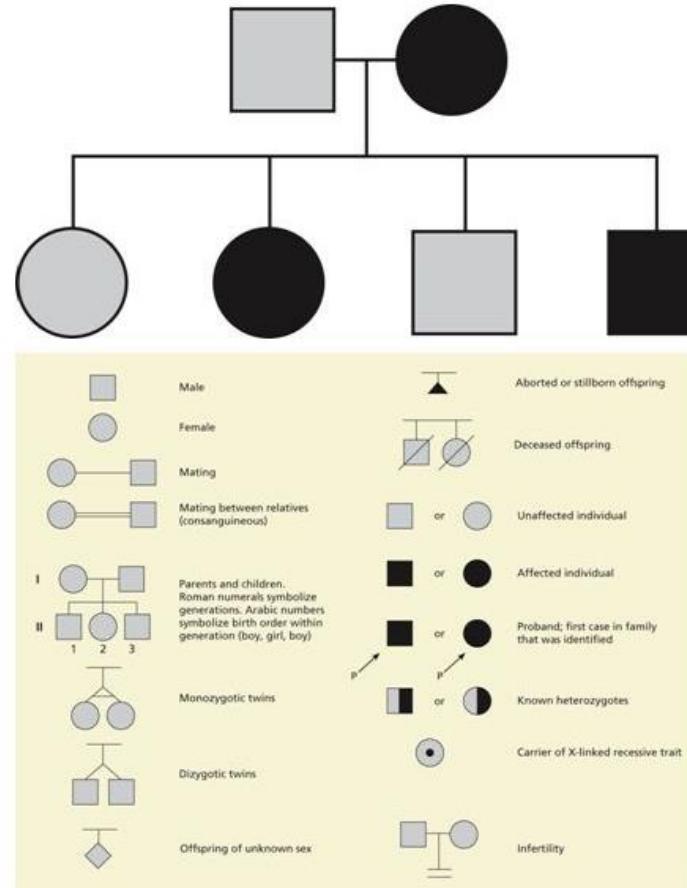
- X-linked recessive disorders are also caused by variants in genes on the X chromosome.
- In males (who have only one X chromosome), one altered copy of the gene in each cell is sufficient to cause the condition.
- In females (who have two X chromosomes), a variant would have to occur in both copies of the gene to cause the disorder. Because it is unlikely that females will have two altered copies of this gene, males are affected by X-linked recessive disorders much more frequently than females.
- A characteristic of X-linked inheritance is that fathers cannot pass X-linked traits to their sons (no male-to-male transmission).
- Example: Hemophilia, a bleeding disorder that slows the blood clotting process; red-green color blindness



Note: a few carriers may be mildly affected due to skewed X-inactivation.

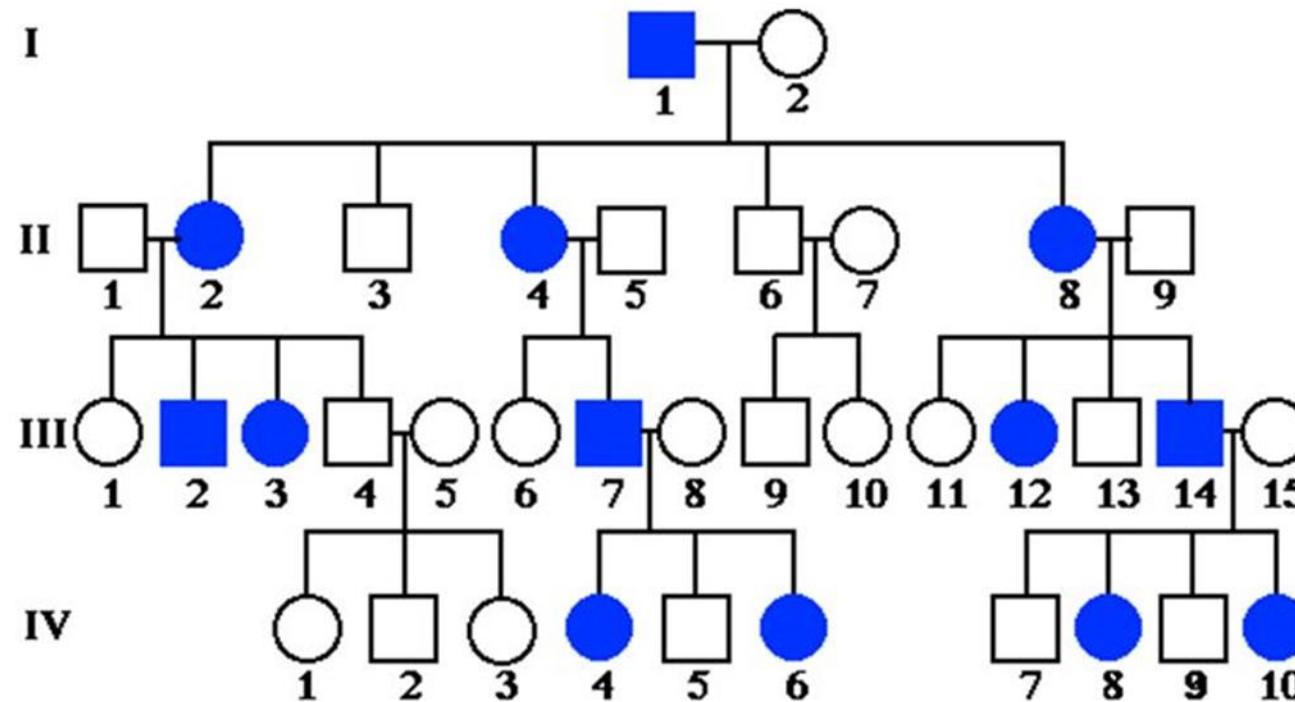
# Pedigree Symbols

- A key is typically provided
- If it is not, these are the standards:
  - Male is square, female is circle
  - Age left-right
  - Marriage is horizontal line
  - Offspring is vertical branched line

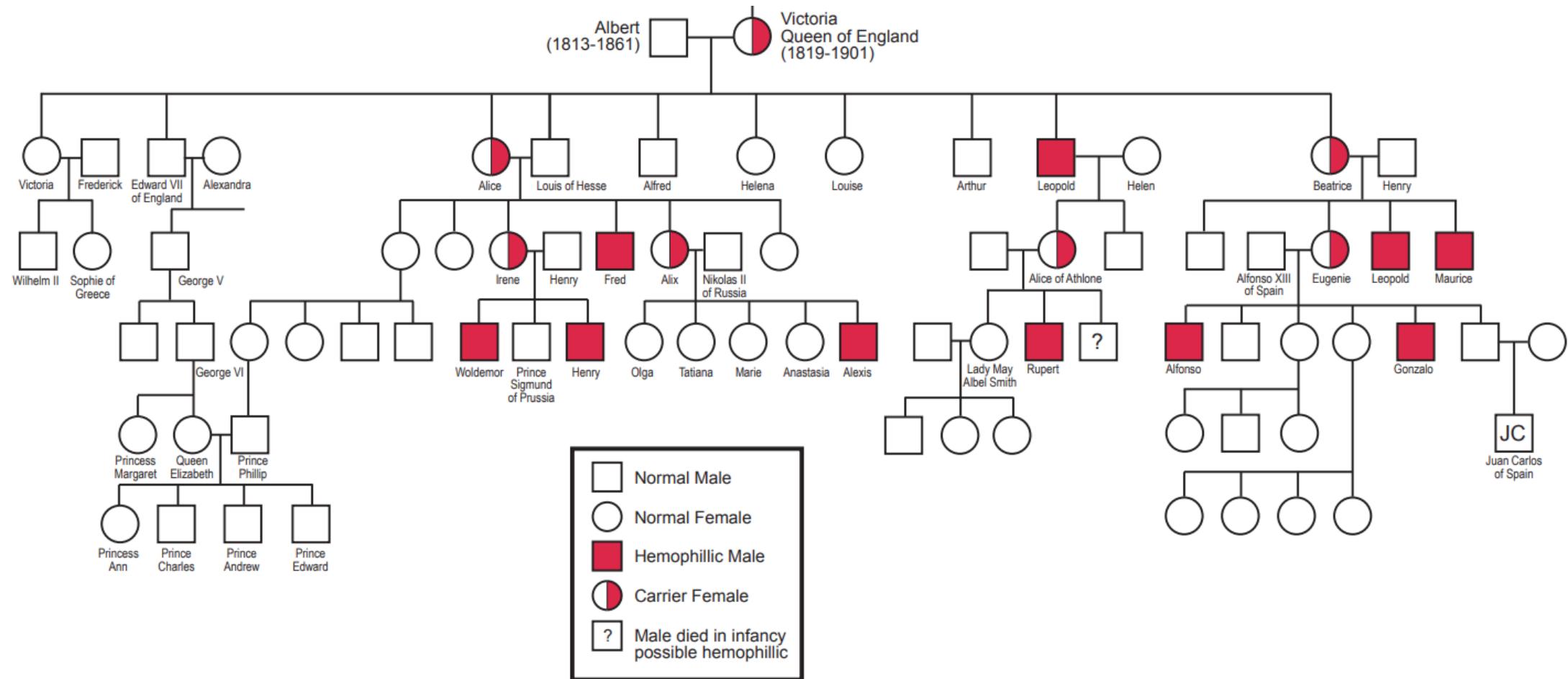


## PEDIGREES SHOW PATTERNS OF INHERITANCE

- X-linked dominant inheritance

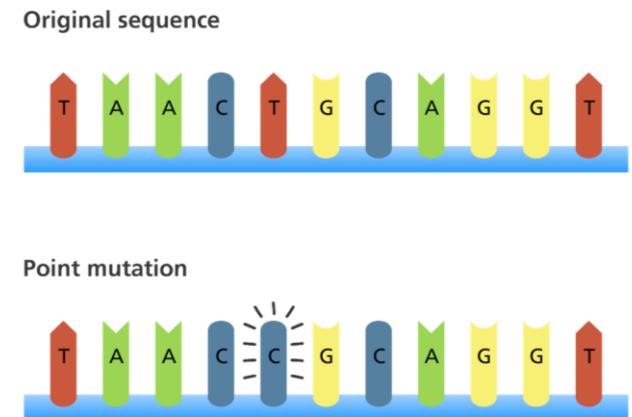


# A pedigree showing haemophilia in Queen Victoria's family



# POINT MUTATIONS

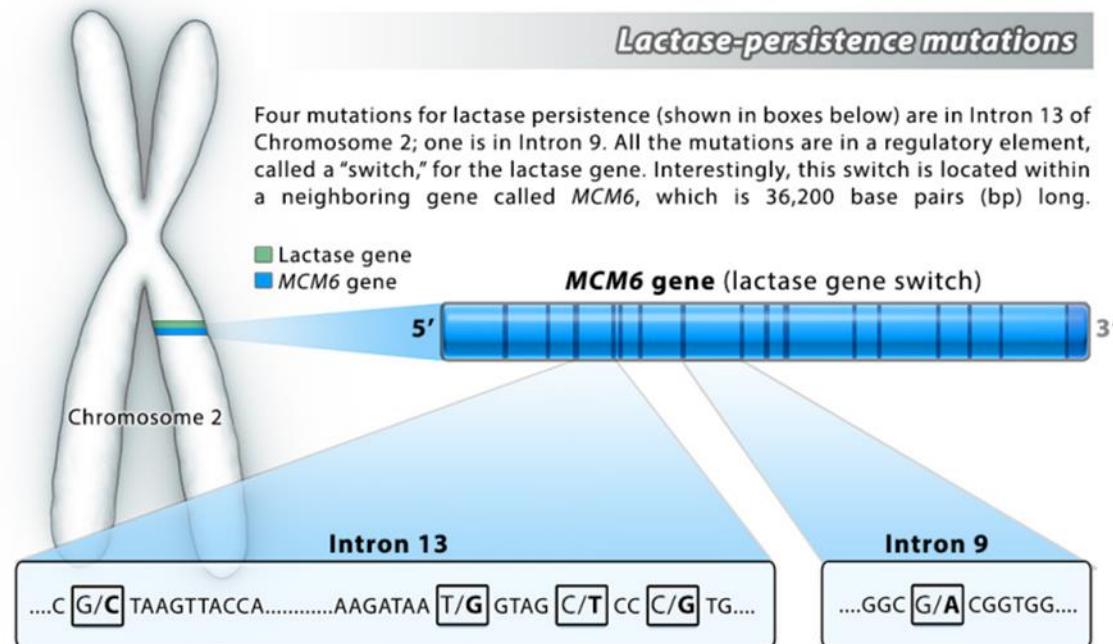
- During DNA replication and gene expression, errors can occur that change the structure and function of the polypeptide product. Errors that might also occur resulting from harmful agents in the environment.
- An organism's genome can be altered through a permanent and heritable change to one or more nucleotide bases.
- Such mutations usually involve a change in or disruption to a base sequence in the DNA.
- The result is the synthesis of a miscoded mRNA and ultimately a change in one or more amino acids found in the polypeptide during translation.
- Many mutations are harmful because they alter some aspect of cellular activity in a negative way, for example, perhaps a mutation causes an enzyme to fold incorrectly, shutting down an important metabolic pathway in a cell.
- However, a mutation can also be beneficial and give the organism a novel property.



# LACTASE PERSISTENCE MUTATIONS

## **Lactase-persistence mutations**

Four mutations for lactase persistence (shown in boxes below) are in Intron 13 of Chromosome 2; one is in Intron 9. All the mutations are in a regulatory element, called a "switch," for the lactase gene. Interestingly, this switch is located within a neighboring gene called *MCM6*, which is 36,200 base pairs (bp) long.



**Figure 3.** A diagram of the five mutations that cause lactase persistence; the prevalence of these mutations varies in different populations. All the mutations occur in *MCM6*, the genetic “switch” for the lactase gene, on Chromosome 2. The introns in *MCM6* are colored light blue, and the exons are colored dark blue.

Gene-Culture Co-evolution

Early humans, which evolved around 200,000 years ago, survived primarily by hunting animals and gathering plant-based foods. Between 10,000 and 7,500 years ago, during the Neolithic Period, populations in both Africa and Europe began domesticating other mammals like goats, sheep, and cows. This practice of keeping and using domesticated animals is called **pastoralism**, and it gave these cultures a reliable source of meat, milk, and milk products — such as butter, cheese, and yogurt. A practice like pastoralism is considered **cultural** because it is taught and learned, rather than an instinct dependent on the transmission of genetic variations.

Genetic analyses suggest that the mutations associated with lactase persistence arose less than 10,000 years ago, and that they became more common specifically in populations that practiced pastoralism. In these populations, individuals with the lactase-persistence mutations were more likely to survive, reproduce, and pass on their alleles than were people without the mutations — especially in times of famine, when other food sources were scarce.

# MUTATIONS CAN BE SPONTANEOUS OR INDUCED

- Spontaneous mutations are heritable, random changes to the base sequence in the DNA.
- These changes could result from errors made and not corrected by DNA polymerase during replication or from physical or chemical agents in the environment.
- It has been estimated that one such mutation can occur for every 10 to 1000 divisions in a bacterial population.
- A mutant cell arising from a spontaneous mutation usually is masked by the much larger population of normal cells.
- However, should the environment favor the mutant, it will multiply and emerge as the predominant form. For example, for many decades doctors used penicillin to treat gonorrhea, which is caused by *Neisseria gonorrhoeae*.
- Then, in 1976, a penicillin-resistant strain of *N. gonorrhoeae* emerged.
- Many investigators suggested that the penicillin-resistant strain had existed in the *N. gonorrhoeae* population for centuries, but only with heavy penicillin use in the 1970s could the penicillin-resistant strain arise and surpass the penicillin susceptible forms.
- The majority of our knowledge of mutations has occurred from researches involving induced mutations, which are developed by external physical and chemical agents called mutagens.

# MUTAGENS

## Radiation



**UV Radiation**  
Both natural sunlight and tanning beds



**X-Rays**  
Medical, dental, airport security screening

## Chemicals



**Smoking or Vaping**  
Contains dozens of mutagenic chemicals



**Nitrate and Nitrite Preservatives**  
In hot days and other processed meats  
**Barbecuing**  
Creates mutagenic chemicals in foods



**Benzoyl Peroxide**  
Common ingredient in acne products

## Infectious Agents



**Human Papillomavirus (HPV)**  
Sexually transmitted virus



**Helicobacter Pylori**  
Bacteria spread through contaminated food

Mutagen	Risk factors	Health behavior that prevents exposure
Sunlight 	Increased exposure to sunlight leading to thymine dimers in DNA	<ul style="list-style-type: none"> <li>• Wear sunscreen</li> <li>• Avoid excess sun exposure</li> <li>• Reapply sunscreen every 2 h</li> <li>• Wear clothes/hat that cover skin</li> </ul>
X-ray radiation 	<ul style="list-style-type: none"> <li>• Multiple X-rays</li> <li>• Close proximity to X-rays</li> </ul>	<ul style="list-style-type: none"> <li>• Limit the number of X-rays to only ones that are necessary</li> <li>• Wear protective lead vest over areas not being X-rayed</li> <li>• Use alternative imaging devices</li> </ul>
Tobacco products 	<ul style="list-style-type: none"> <li>• Smoking</li> <li>• Chewing tobacco</li> <li>• Second-hand smoke</li> </ul>	<ul style="list-style-type: none"> <li>• Avoid smoking/tobacco products</li> <li>• Avoid being near smoke</li> </ul>
Chemicals 	Increased exposure to carcinogenic/mutagenic chemicals	<ul style="list-style-type: none"> <li>• Limit exposure</li> <li>• Dispose of chemicals properly</li> <li>• Wear protective gear</li> </ul>
Nitrites	Increased consumption of processed meats, such as hot dogs, sausages, and bacon	Limit consumption of processed meats and grilled meats

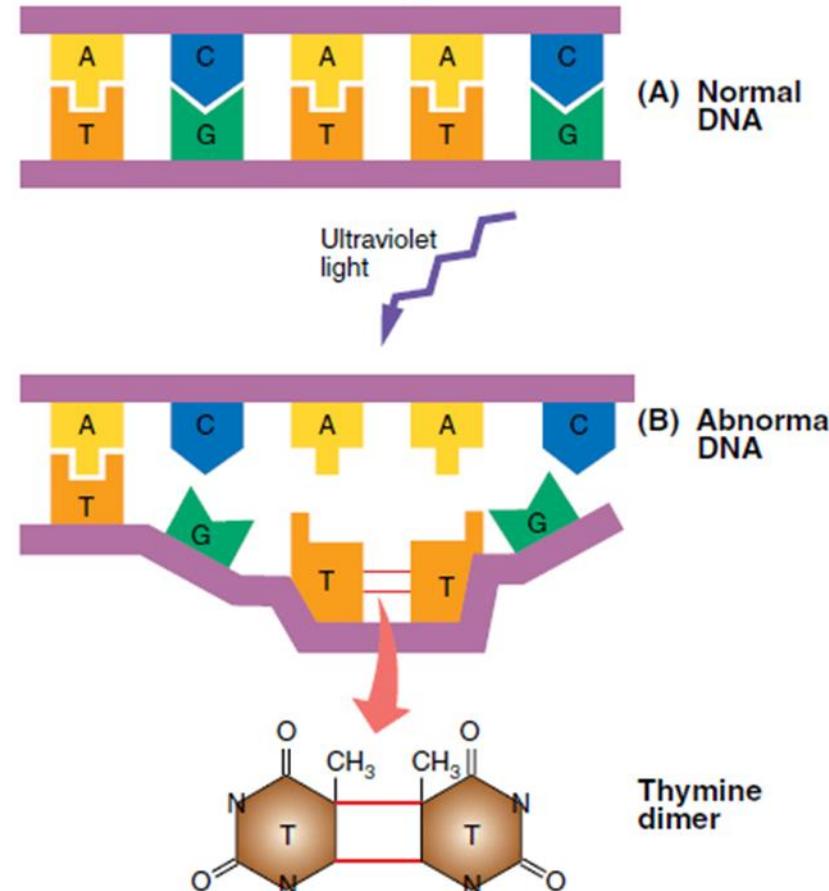
# PHYSICAL MUTAGENS

Ultraviolet (UV) light is a physical mutagen whose energy causes adjacent thymine (or cytosine) bases in the DNA to covalently link together forming dimers.

- (A) When cells are irradiated with ultraviolet (UV) light either naturally or through experiment, the radiations might affect the cell's DNA. (B) UV light can cause adjacent thymine molecules to covalently pair (red lines) within the DNA strand to form a thymine dimer.
- If such type of dimers appear in a protein-coding gene, the RNA polymerase can not place the correct bases (A-A) in mRNA molecules where the dimers are situated

In addition, ionizing radiations, such as gamma rays and X-rays, can cause physical breaks in the double-strand DNA. Loss of cellular function usually results.

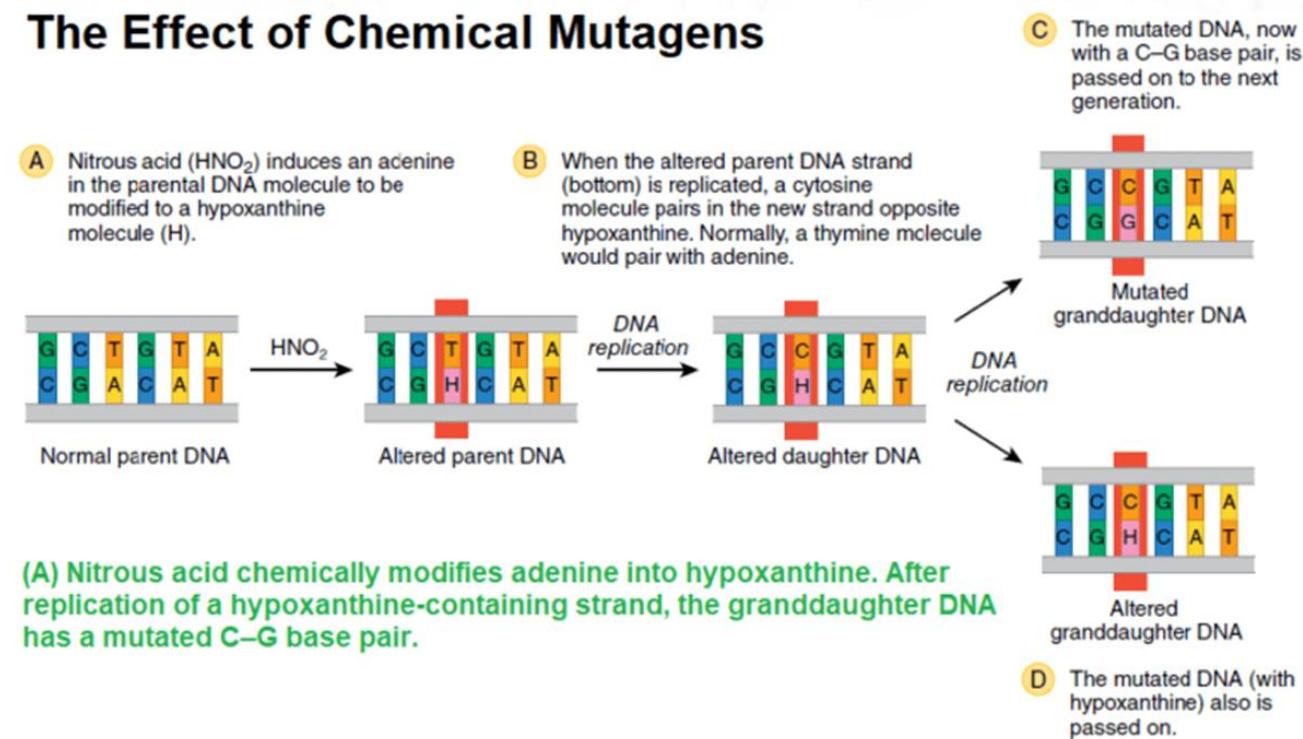
## Ultraviolet Light and DNA



# CHEMICAL MUTAGENS

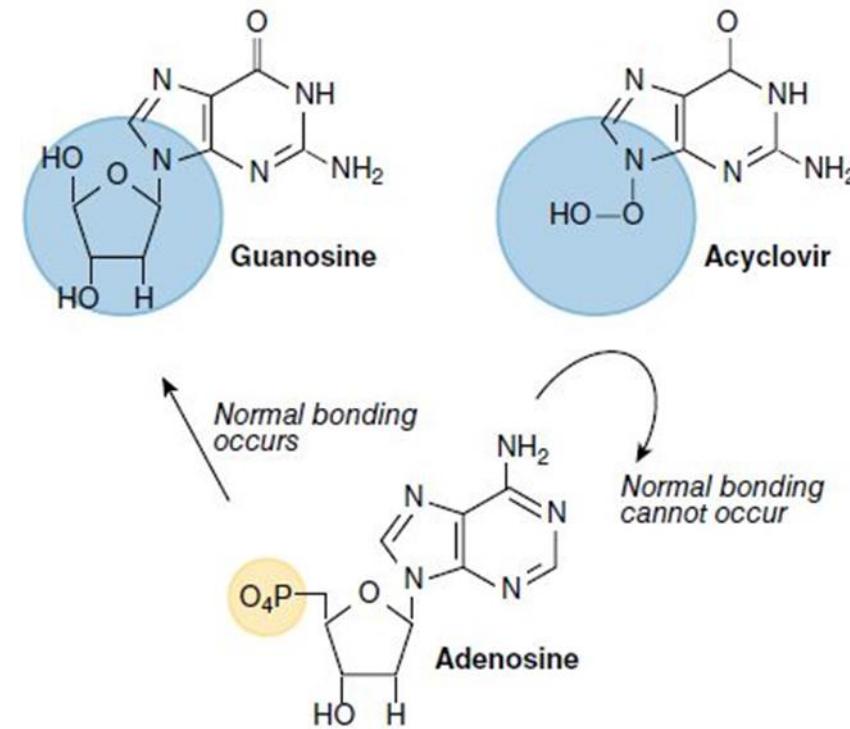
- Many chemicals are mutagenic; that is, they can cause mutations.
- Nitrous acid is an example of a chemical mutagen that converts DNA's adenine bases to hypoxanthine bases.
- Adenine generally base pair with thymine, but the existence of hypoxanthine creates a base pairing with cytosine while replication.
- After, if replication happens from the gene with the cytosine mutation, the mRNA will have a guanine base instead of an adenine base.

## The Effect of Chemical Mutagens



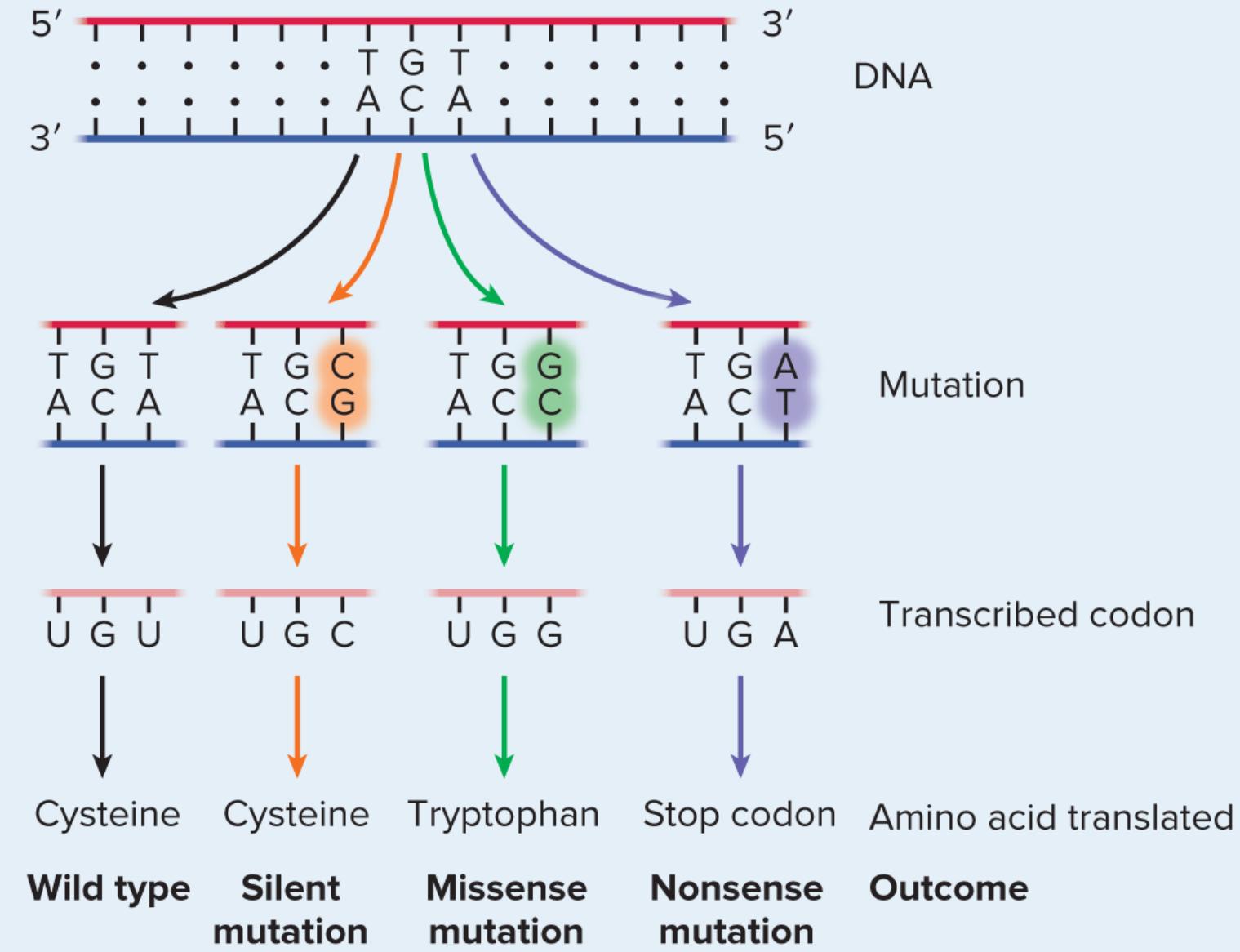
# BASE ANALOGS

- Mutations also are induced by base analogs, which bear a close chemical similarity to a normal nucleotide.
- Base analogs induce mutations by substituting for nitrogenous bases in the synthesis of DNA. Note the similarity in chemical structure between guanosine and the base analog acyclovir.
- For example, acyclovir is a base analog that can substitute for guanine during virus replication.
- However, incorporation of acyclovir blocks further viral replication because the analog lacks the deoxyribose sugar needed for the addition of the next nucleotide. As a result, new virus particles cannot be produced.
- Acyclovir, therefore, is an effective treatment to limit infections caused by different viruses, e. g. the herpes simplex 1 and 2 viruses (cold sores and genital herpes).



# POINT MUTATIONS CAN AFFECT GENE EXPRESSION

- Regardless of the cause of the mutation, one of the most common results is a point mutation, which usually affects just one point (a base pair) in a gene.
- Such mutations might be a change to or substitution of a different base pair or a deletion or addition of a base.
- Base-Pair deletion or insertion
- If a point mutation creates a base-pair substitution, the transcription of that gene might have one wrong base in the mRNA sequence of codons.
- Perhaps one way to see the effects of such changes is using an English sentence made up of three-letter words (representing codons) in which one letter has been changed.
- As three-letter words, the letter substitution still reads correctly, but the sentence makes less sense.
- Normal sequence: THE FAT CAT ATE THE RAT
- Substitution (H for C): THE FAT HAT ATE THE RAT
- If the substitution, because of redundancy in the genetic code, does not alter the amino acid sequence, the change is referred to as a silent mutation because there is no change in protein function.
- If a base-pair substitution does result in a wrong amino acid in a polypeptide, the change is referred to as a missense mutation because the polypeptide is still assembled but might not have the correct shape.
- Our sentence analogy presented above is an example of a missense mutation.
- Finally, if the substitution causes a codon to become a stop codon, the change is called a nonsense mutation because translation is terminated prematurely, and any polypeptide produced probably will be nonfunctional.

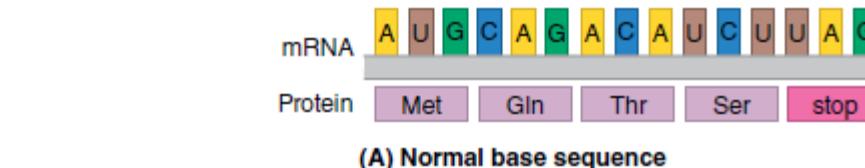


# POINT MUTATIONS CAN AFFECT GENE EXPRESSION

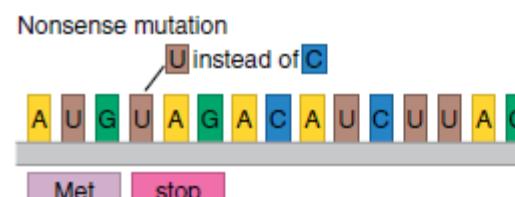
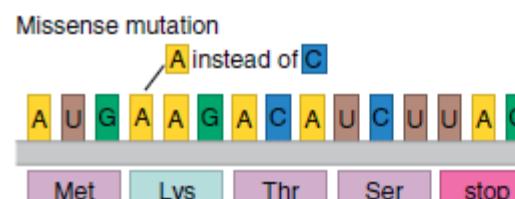
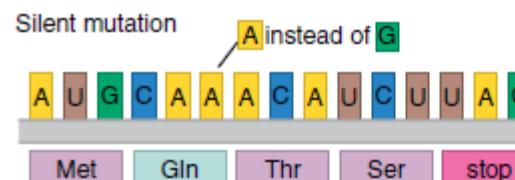
- Base-Pair deletion or insertion
- Point mutations also can lead to the loss or addition of a base in a gene, producing in an improper quantity of bases.
- Again, using our English sentence analogy, we can see how a deletion or insertion of one letter affects the reading frame of the three letter word sentence.
- Normal sequence: THE FAT CAT ATE THE RAT
- Deletion: THE F\_T CAT ATE THE RAT
- Insertion: THE FAT ACA TAT ETH E RAT
- As you can see, the “sentence mutations” are nonsense when reading the sentence as three-letter words. The same is true in a cell.
- Ribosomes always read three letters (one codon) at one time, producing potentially extensive issues in the amino acid sequence if there are too few or too many letters.
- Thus, like our English sentence, the deletion or addition of a base will cause a “reading frameshift” because the ribosome always reads the genetic code in groups of three bases.
- Therefore, loss or addition of a base also represents a frameshift mutation because it shifts the reading of the code by one base.
- The result is serious sequence errors in the amino acids, which will probably produce an abnormal protein (nonsense) unable to carry out its functional role in the cell.

(A) The normal sequence of bases in a gene.

(B) Base-pair substitutions can produce silent, missense, or nonsense mutations.

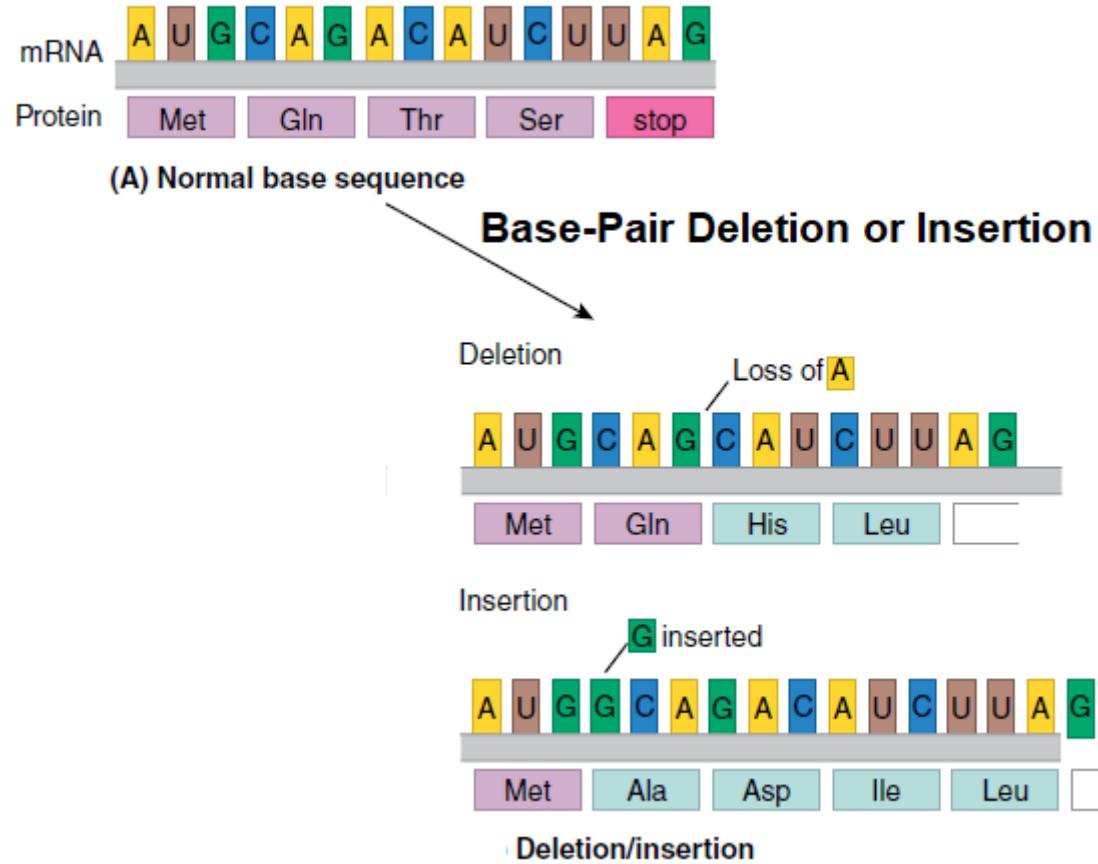


## Point Mutations



(B) Base-pair substitutions

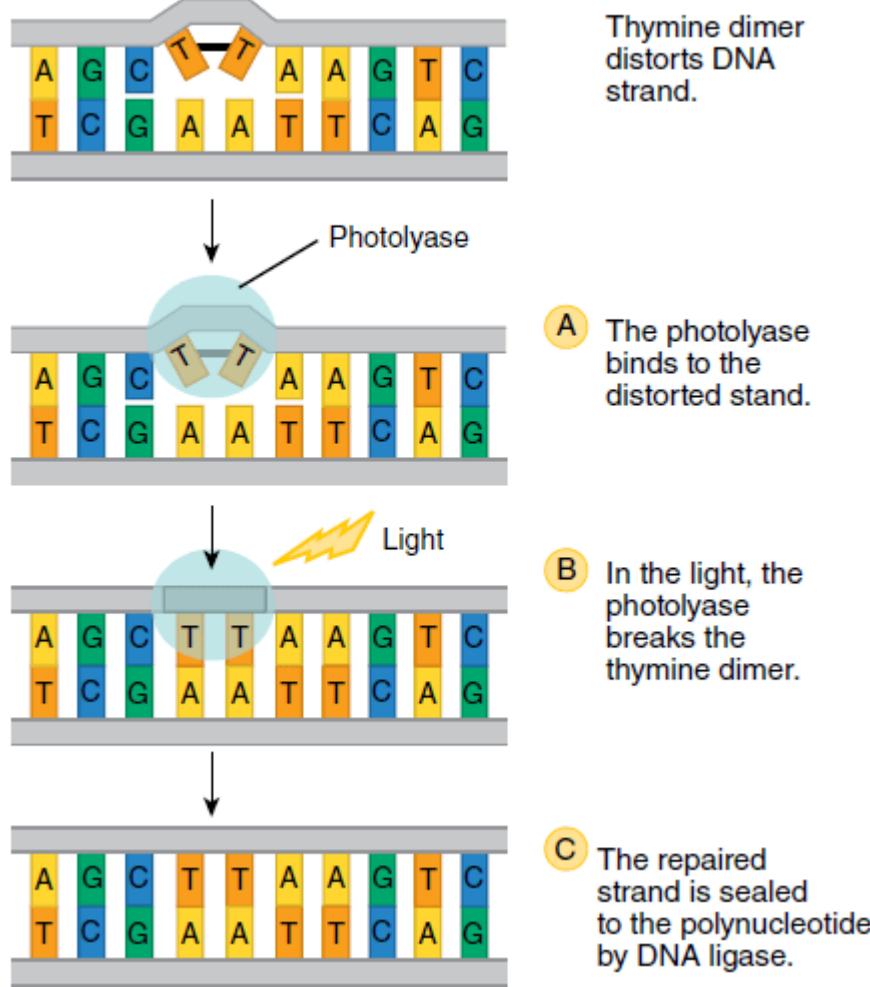
Deletions or insertions shift the reading frame of the ribosome.



# DNA REPAIR

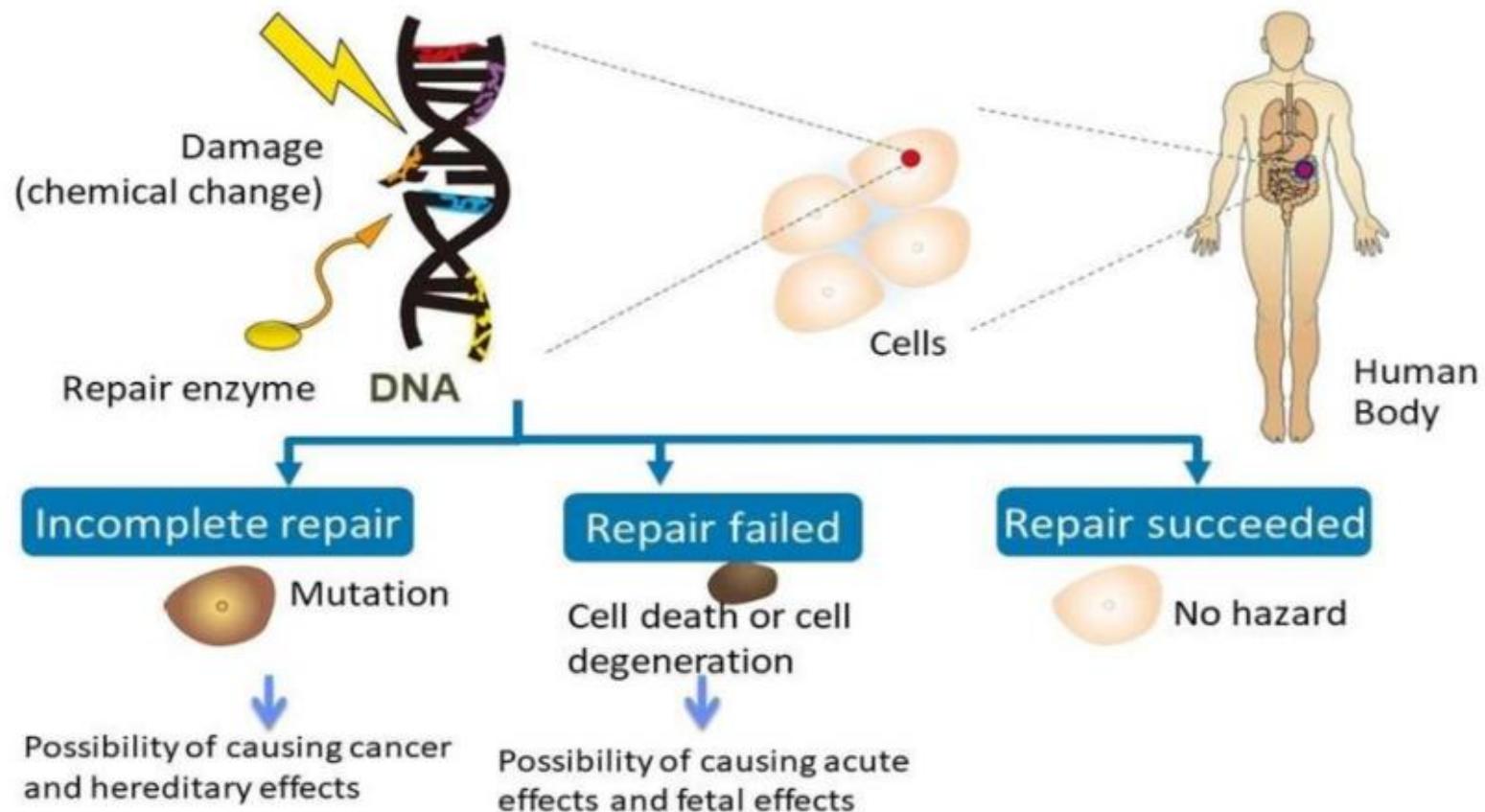
- During the life of a microbial cell cellular DNA endures thousands of potentially damaging events resulting from DNA replication errors.
- The fact that DNA is double stranded is not an accident.
- By being double stranded, one polynucleotide strand can act as the master copy or template to repair mismatches between strands or distortions within one strand.
- Therefore, if the damaged DNA is repaired before the cell divides, no mutation will result.
- Even though DNA polymerase is very accurate in proofreading during replication, errors (mismatched nucleotides) are sometimes missed. Mismatch repair can be used to detect and repair these errors.
- First, mismatch correction enzymes scan newly synthesized DNA for any mismatched pairs.
- Finding such a pair, the enzymes cut out (excise) a small segment of the polynucleotide strand containing the mismatched nucleotides.
- A DNA polymerase then uses the old strand as a template to replace the excised nucleotide segment with the correct set of bases, which is sealed in place by a DNA ligase.
- DNA repair is seldom 100% perfect, sometimes “shoddy repairs” fail to correct an error and the mutation becomes “locked in” and is inheritable, because it now exists as part of the master copy that can be transferred to future generations of cells during cell division.

## Nucleotide Excision Repair



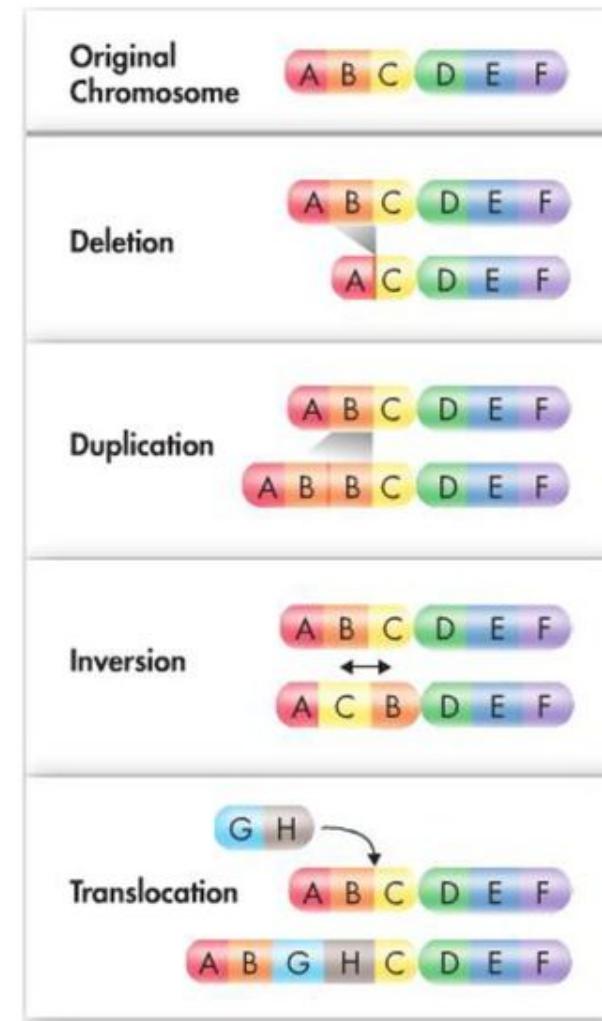
Thymine dimer distortion triggers photolyase enzymes to repair the damaged DNA.

The enzyme binds to the dimer and, when exposed to visible light (photoreactivation), the enzyme breaks the bond holding the thymine dimer together.



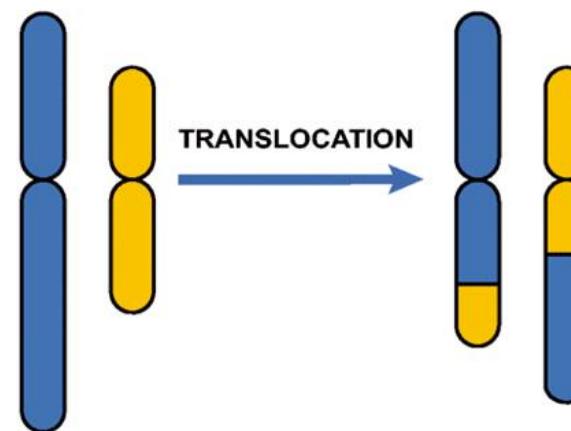
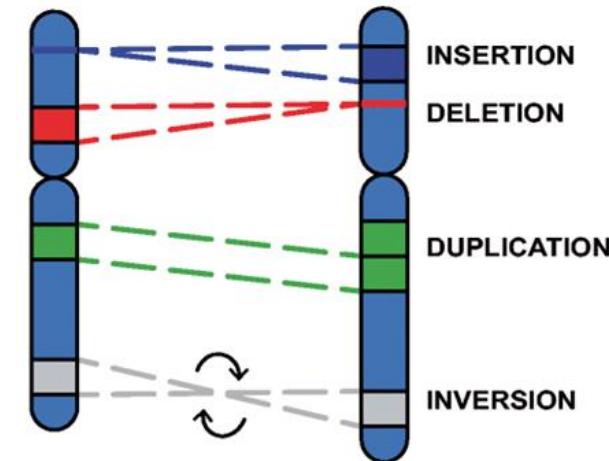
# Chromosomal mutation

- A chromosomal mutation involves a change in the structure or number of chromosomes
- 4 types of chromosomal mutations:
  - Deletion: loss of all or part of a chromosome
  - Duplication: extra copy of all or part of a chromosome
  - Inversion: reverses the direction of parts of a chromosome
  - Translocation: part of one chromosome breaks off and attaches to another chromosome



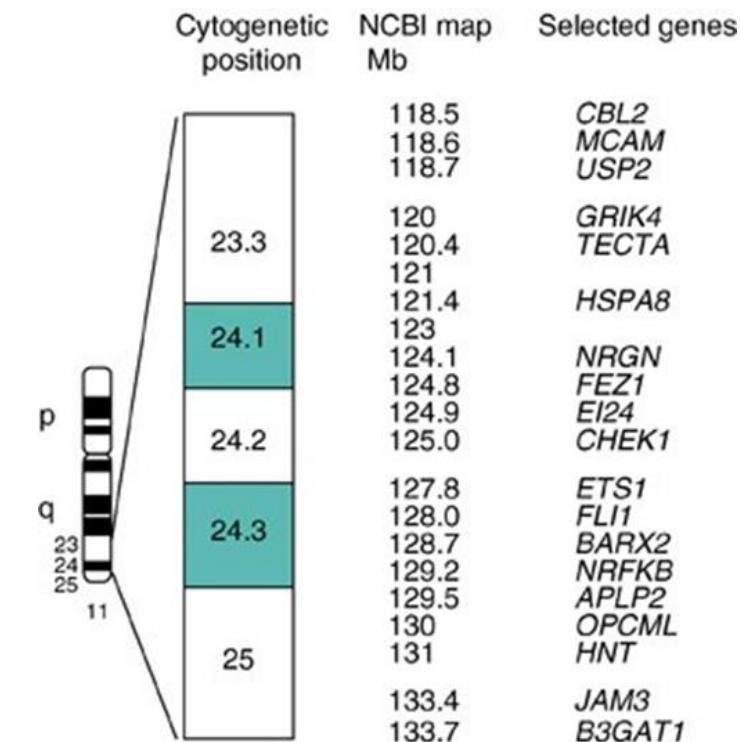
# CHROMOSOMAL MUTATION

- A chromosomal mutation is a missing, extra, or irregular portion of chromosomal DNA. These can occur in the form of structural abnormalities, where one or more individual chromosomes are altered. Chromosome anomalies usually occur when there is an error in cell division following meiosis or mitosis.



# DELETION - JACOBSEN SYNDROME

- results from deletion of genes from chromosome 11 that includes band 11q24.1.
- The deletion may range from 5 million to 16 million deleted DNA base pairs. The severity of symptoms depends on the number of deletions; the more deletions there are, the more severe the symptoms are likely to be.
- People with Jacobsen syndrome have serious intellectual disabilities, dysmorphic features, delayed development and a variety of physical problems including heart defects.



# ENVIRONMENTAL MODIFICATION

- Genetic processes work in combination with an organism's environment and experiences to influence development and behavior. The intracellular or extracellular environment of a living cell or organism may switch gene transcription on or off.
- A classic example is two seeds of genetically identical corn, one placed in a temperate climate and one in an arid climate (lacking sufficient waterfall or rain). While **the average height of the two corn stalks** may be **genetically determined** to be equal, the one in the arid climate only grows to half the height of the one in the temperate climate due to lack of water and nutrients in its environment.

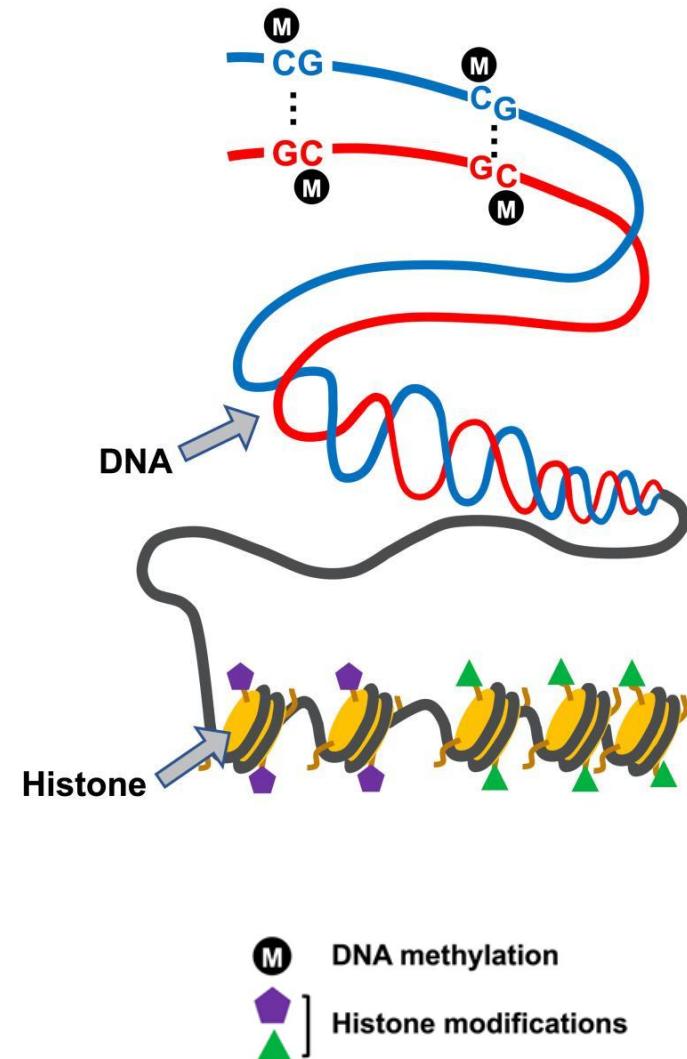


# WHAT IS EPIGENETICS?

- Epigenetics is the study of how **cells control gene activity without changing the DNA sequence**.
- "Epi-" means on or above in Greek, and "epigenetic" describes factors beyond the genetic code.
- Epigenetic changes are modifications to DNA that regulate whether genes are turned on or off. These modifications are attached to DNA and do not change the sequence of DNA building blocks. Within the complete set of DNA in a cell (genome), all of the modifications that regulate the activity (expression) of the genes is known as the epigenome.
- Epigenetic changes help determine whether genes are turned on or off, so they influence the production of proteins in cells.
- This regulation helps ensure that each cell produces only proteins that are necessary for its function. For example, proteins that promote bone growth are not produced in muscle cells.
- Patterns of epigenetic modification vary among individuals, in different tissues within an individual, and even in different cells within a tissue. Environmental influences, such as a person's diet and exposure to pollutants, can impact the epigenome. Epigenetic modifications can be maintained from cell to cell as cells divide and, in some cases, can be inherited through the generations.

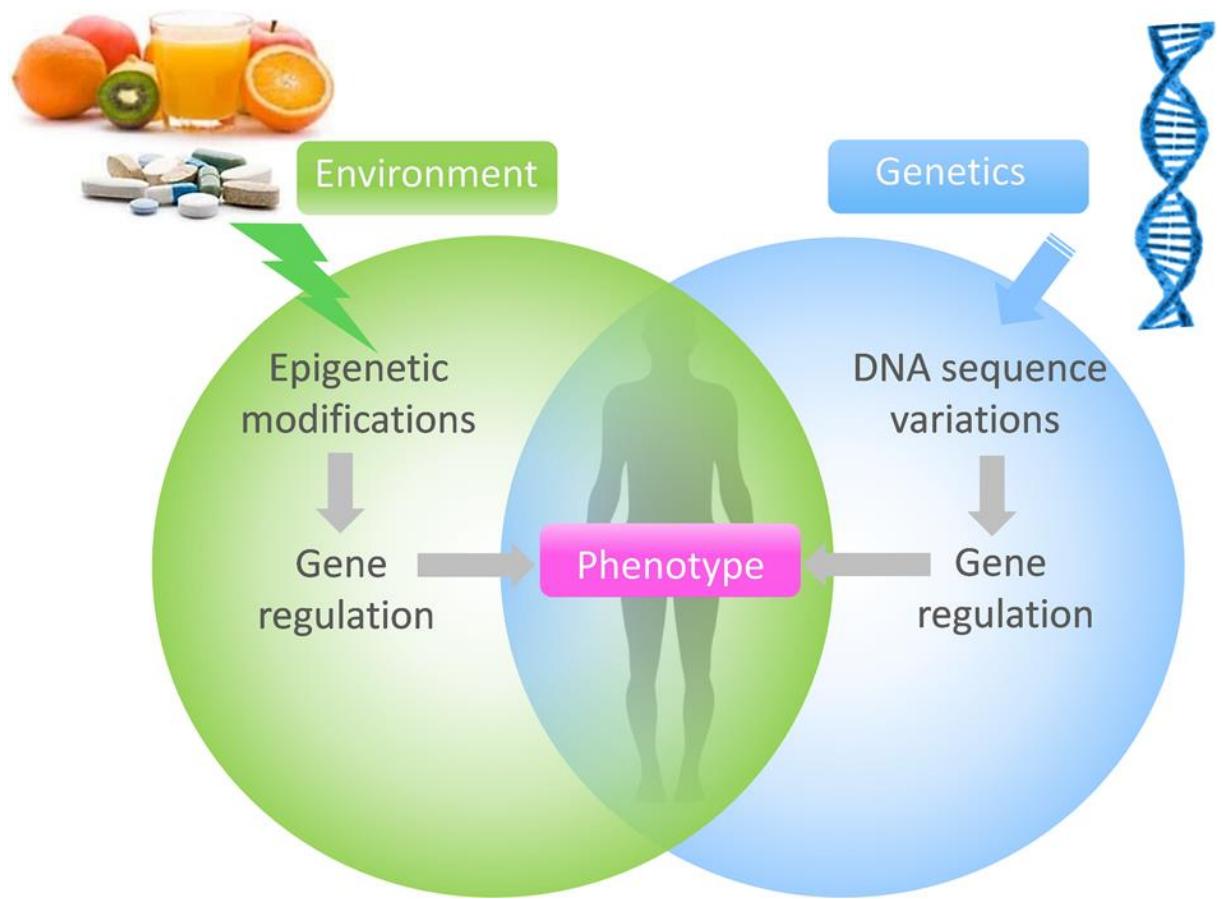
# EPIGENETIC MODIFICATION

- A common type of epigenetic modification is called **DNA methylation**. DNA methylation involves the attachment of small chemical groups called methyl groups (each consisting of one carbon atom and three hydrogen atoms) to DNA building blocks. When methyl groups are present on a gene, that gene is turned off or silenced, and no protein is produced from that gene.
- Another common epigenetic change is **histone modification**. Histones are structural proteins in the cell nucleus. DNA wraps around histones, giving chromosomes their shape. Histones can be modified by the addition or removal of chemical groups (methyl or acetyl groups) which influence how tightly the DNA is wrapped around histones, which affects whether a gene can be turned on or off.



# EPIGENETIC MODIFICATION

- Errors in the epigenetic process, such as modification of the wrong gene or failure to add a chemical group to a particular gene or histone, can lead to abnormal gene activity or inactivity. Altered gene activity, including that caused by epigenetic errors, is a common cause of genetic disorders. Conditions such as cancers, metabolic disorders, and degenerative disorders have been found to be related to epigenetic errors.
- Scientists continue to explore the relationship between the genome and the chemical compounds that modify it. In particular, they are studying the effects that epigenetic modifications and errors have on gene function, protein production, and human health.



# BIOINFORMATICS

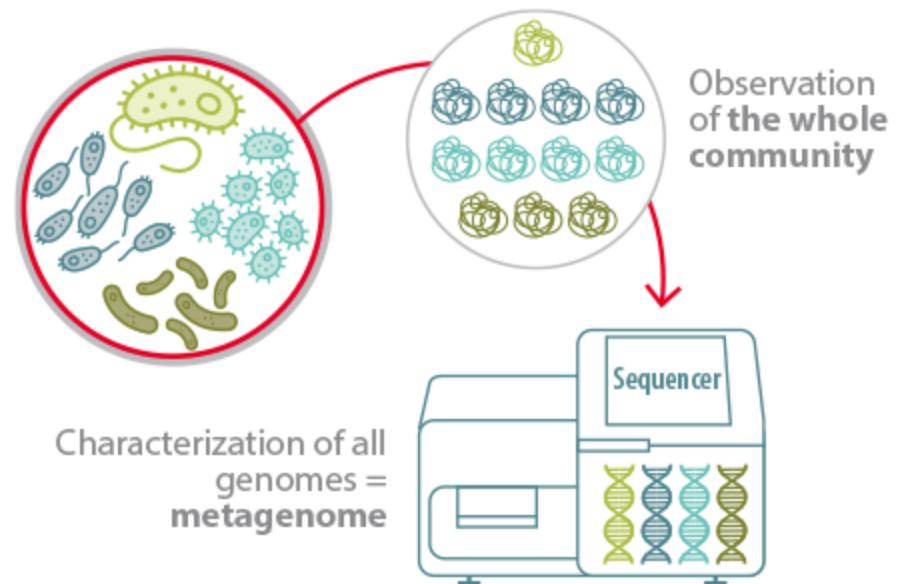
is an interdisciplinary field that develops methods for storing, retrieving, organizing and analyzing biological data. A major activity in bioinformatics is to develop software tools to generate biological knowledge (e. g. protein expression and regulation). Bioinformatics tools aid in the comparison of genetic and genomic data.

- **Systems Biology** is a biology-based inter-disciplinary field of study that focuses on **complex interactions within biological systems**, using a more holistic perspective (instead of the more traditional reductionism) approach. One of the aims is to discover emergent properties of cells, tissues and organisms functioning as a system.
- **Genomics** applies DNA sequencing methods to analyze the function of **genomes (the complete set of DNA within a single cell of an organism)**. The field includes efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping. In contrast, the investigation of single genes is a primary focus of genetics.
- **Metagenomics** is the study of metagenomes, genetic material recovered directly from environmental samples. It may also be referred to as environmental genomics, ecogenomics or community genomics. While traditional microbiology and microbial genomics rely upon cultivated clonal cultures, early environmental gene sequencing cloned specific genes (often the 16S rRNA gene) to **produce a profile of diversity in a natural sample**.

# METAGENOMICS

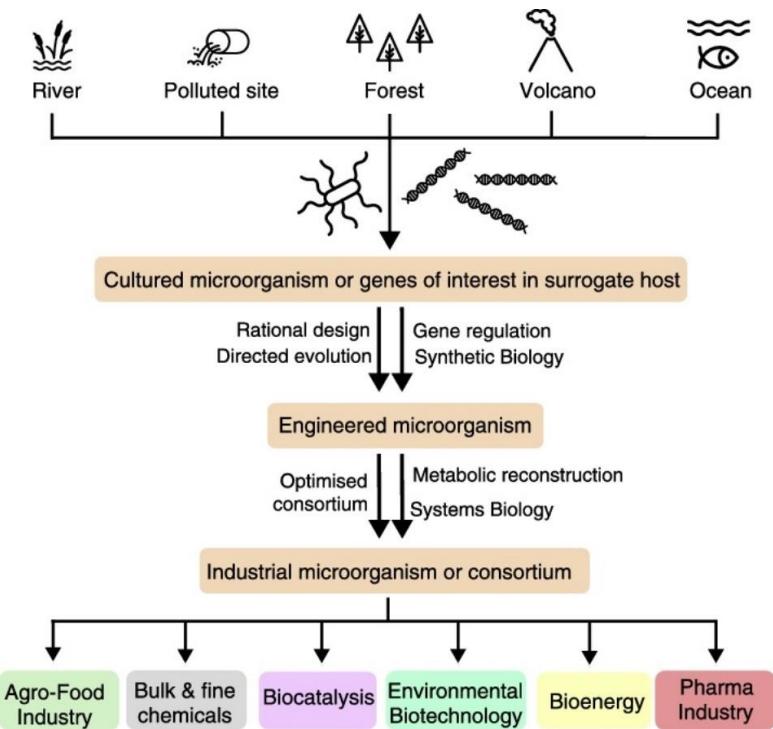
- revealed that the previously hidden vast majority of microbial biodiversity had been missed by cultivation-based methods
- metagenomics offers a powerful lens for viewing the microbial world that has the potential to **revolutionize understanding** of the entire living world

## What is metagenomics?



# HARNESSING THE POWER OF MICROBIAL METABOLISM

- The untapped diversity of microbial metabolisms has huge biotechnological potential.
- Microorganisms are rich repositories of genetic material encoding many activities of potential interest.
- Improved metagenomic screening techniques make it easier to identify activities of interest.
- Synthetic biology and efficient genome editing techniques allow microbial genomes to be modified.
- Computational approaches make it possible to predict the outcome of the metabolic processes and modifications required for optimization.
- Together these advances represent a major breakthrough in microbial biotechnology that is expected to yield new generations of tailor-made biocatalysts suitable for multiple biotechnological applications.



## Genome annotation

## STRUCTURAL ANNOTATION

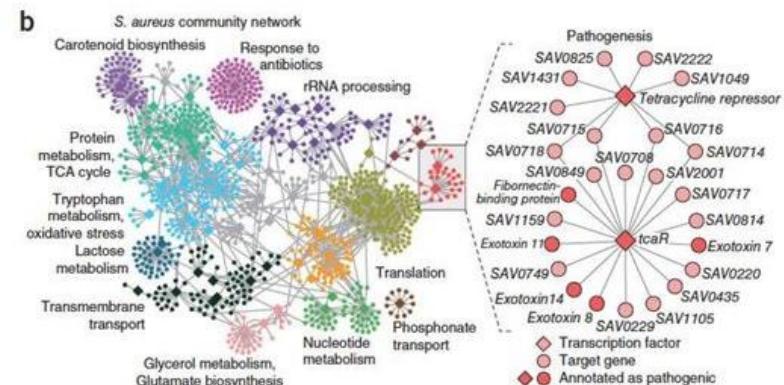
- Open reading frame and their localization
  - Exons, introns, UTRs
  - Start/Stop
  - Location of regulatory motifs
  - Splice Sites
  - Non coding Regions
  - Transposable elements
  - tRNA, miRNA, rRNA, ncRNA



## FUNCTIONAL ANNOTATION

Gene function prediction: attaching biological information to these elements

- Biochemical function
  - Biological function
  - Involved regulation and interactions

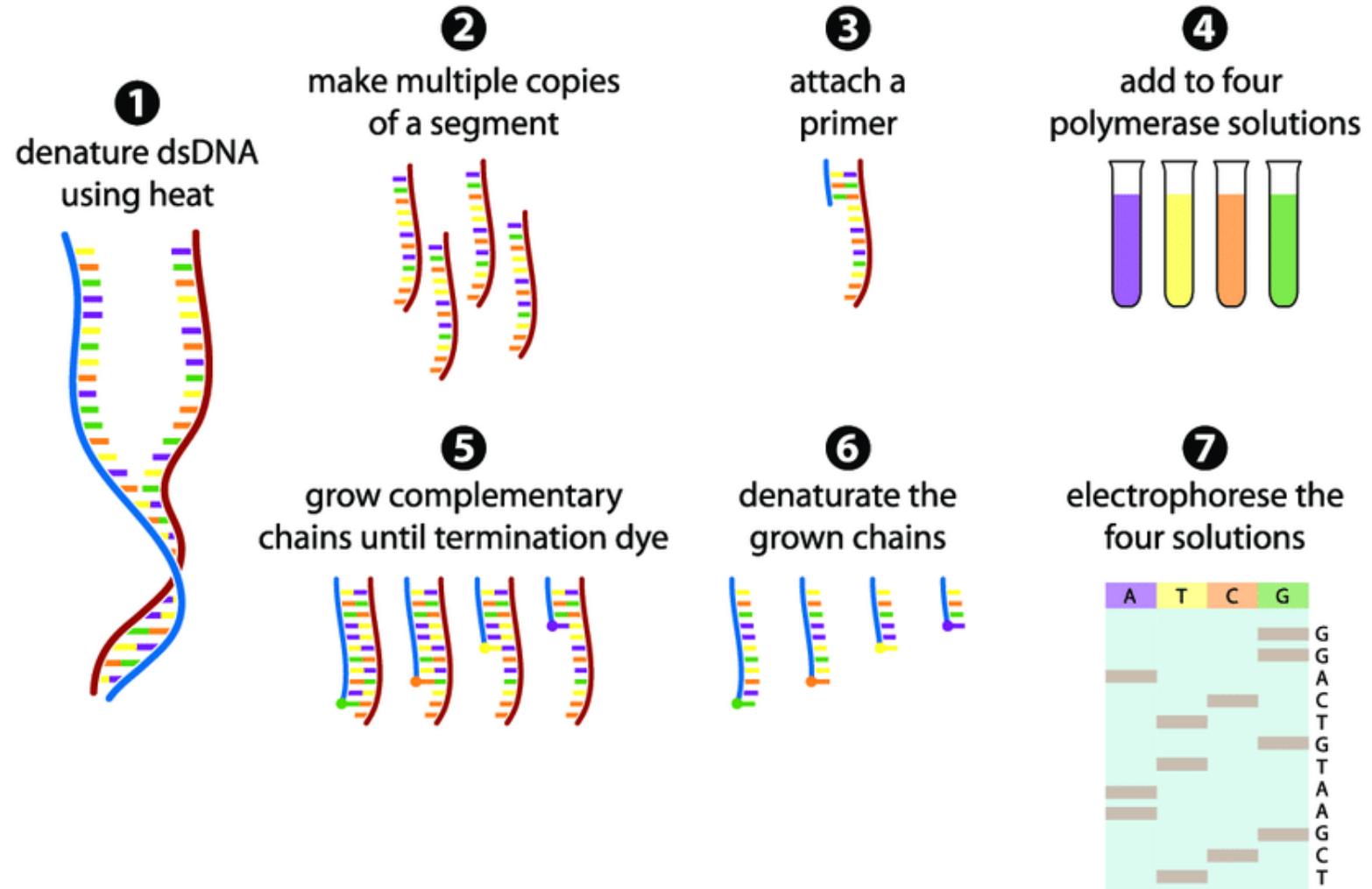


## IN SILICO ANALYSES

- Bioinformatics has been used for **in silico analyses** of biological queries using mathematical and statistical techniques. In biology and other experimental sciences, an in silico experiment is one **performed on computer or via computer simulation**. The phrase is pseudo-Latin for 'in silicon' (in Latin it would be in silicio), referring to silicon in computer chips.
- It was coined as an allusion to the Latin phrases **in vivo**, **in vitro**, and **in situ**, which are commonly used in biology (especially systems biology). The latter phrases refer, respectively, to experiments done in living organisms, outside living organisms, and where they are found in nature.

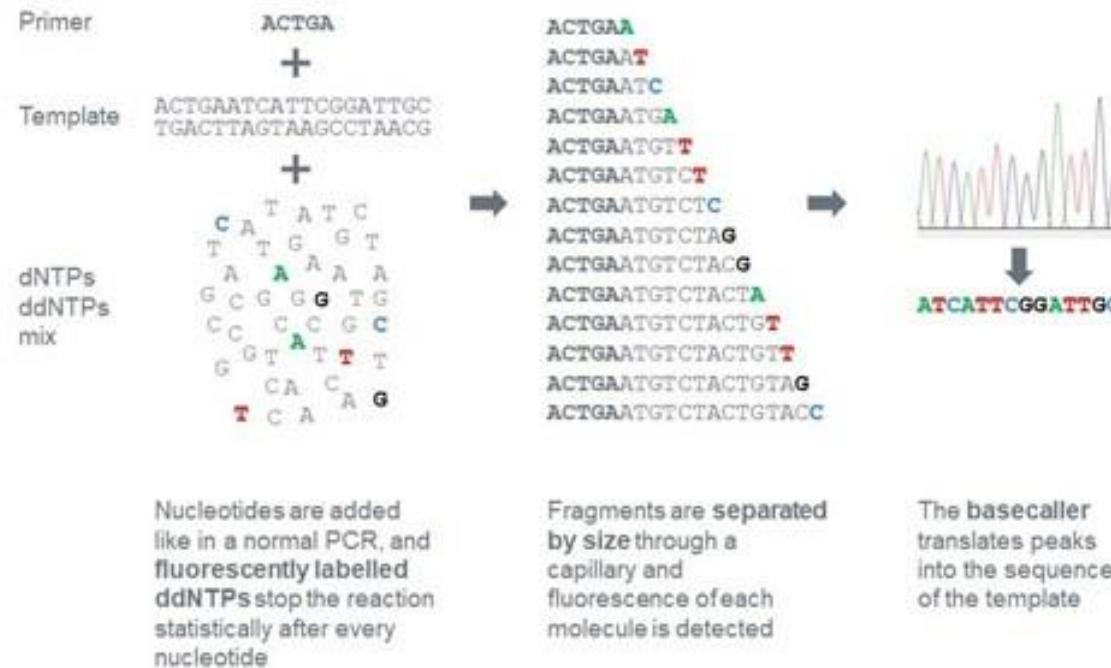
# SANGER SEQUENCING

- Sanger sequencing, also known as the “chain termination method”, is a method for determining the nucleotide sequence of DNA.
- The method was developed by Frederick Sanger in 1977.



# SANGER SEQUENCING

- Sanger sequencing, also known as the “chain termination method”, is a method for determining the nucleotide sequence of DNA.
- The method was developed Frederick Sanger in 1977.

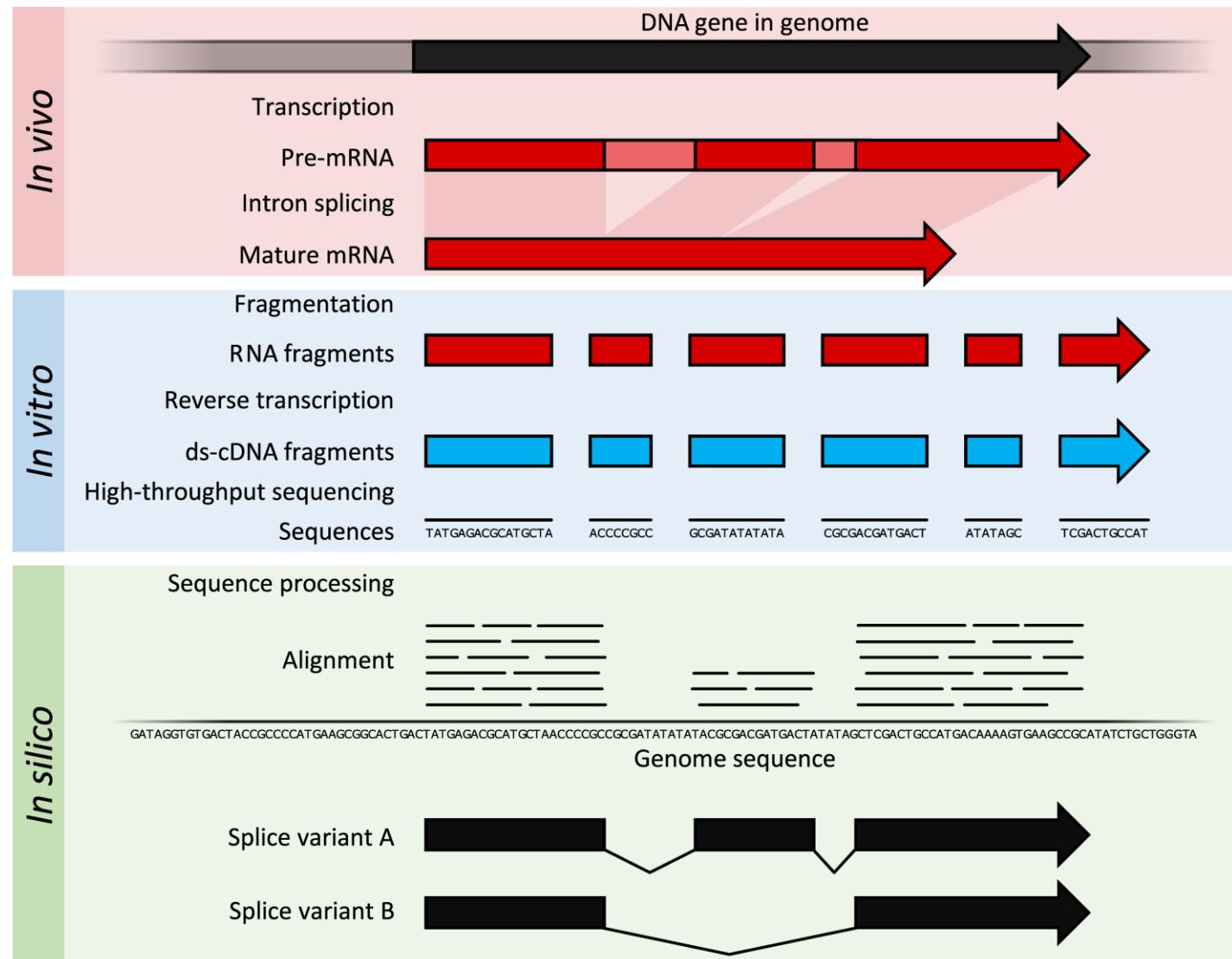


## RNA sequencing

Within the organisms, genes are transcribed and spliced (in eukaryotes) to produce mature mRNA transcripts (red).

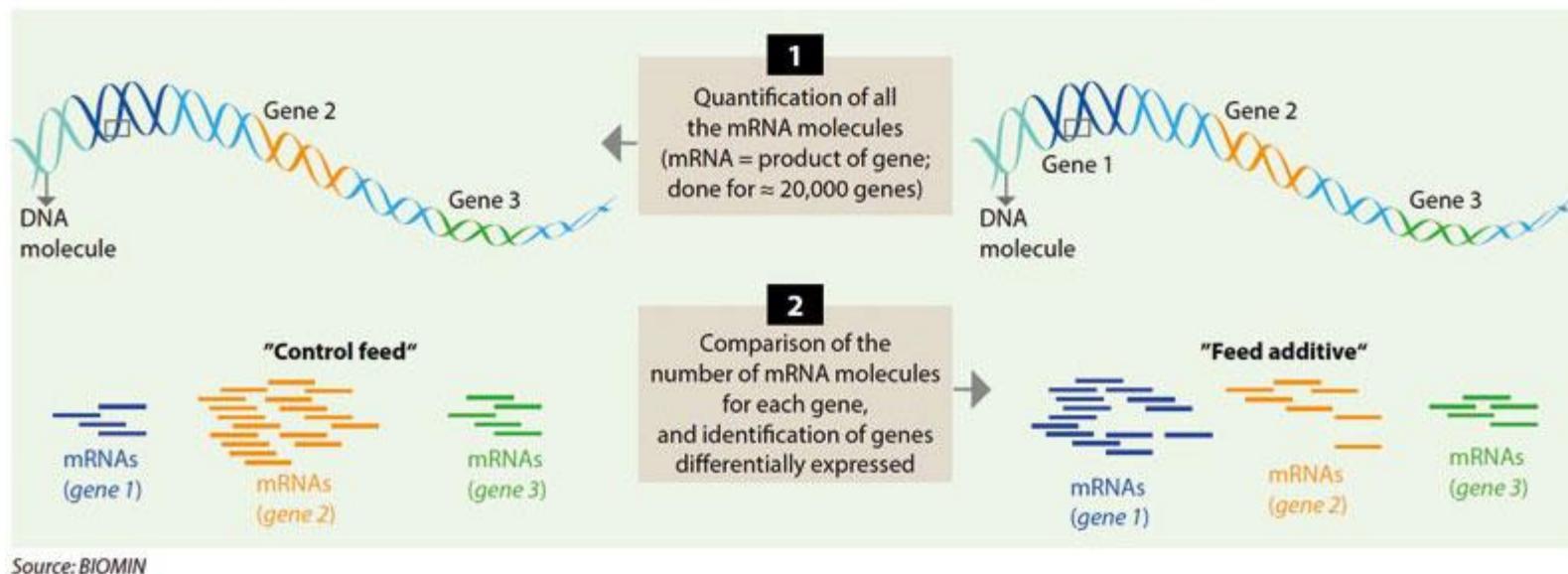
The mRNA is extracted from the organism, fragmented and copied into stable double-stranded-cDNA (ds-cDNA; blue). The ds-cDNA is sequenced using high-throughput, short-read sequencing methods.

These sequences can then be **aligned to a reference genome** sequence to reconstruct which genome regions were being transcribed. These data can be used to annotate where expressed genes are, their relative expression levels, and any alternative splice variants.

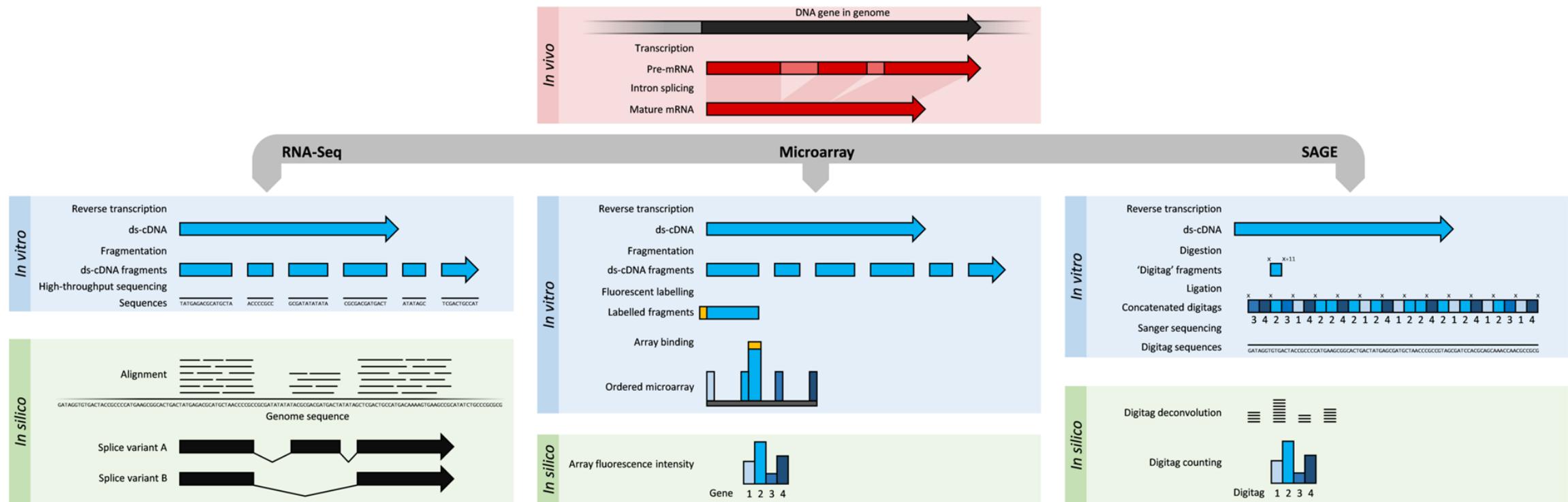


# NEXT GENERATION SEQUENCING AND NUTRIGENOMICS

- Next generation sequencing (NGS) reveals links between nutrition and genes
- Rapidly developing NGS technology (massively parallel sequencing technology that offers ultra-high throughput and speed) will play an important role in increasing our understanding of how nutrition influences metabolic and immunity pathways and enhances health and well-being—a field called nutrigenomics.



# GENERATING DATA



# GENE EXPRESSION PROFILES

- Nearly all the biological events such as cell division and differentiation, responsiveness to hormones or growth factors and ultimately cell death are associated with changes in expression of key genes.
- During the onset and progression of disease, extensive changes take place in gene expression. By **comparing gene expression profiles** under different conditions, individual genes or group of genes can be identified that play an important role in a particular signaling cascade or process or in disease etiology.
- The methods for **assaying gene expression** can be grouped into two major types:
  - (1) closed method, which measures expression levels of already collected clones or sequences, e. g. Microarrays
  - (2) open method, which does not require prior knowledge of the genes being measured, e. g. SAGE
- The sequencing of cDNA clones provided the first glimpse of gene expression patterns in different human tissues.

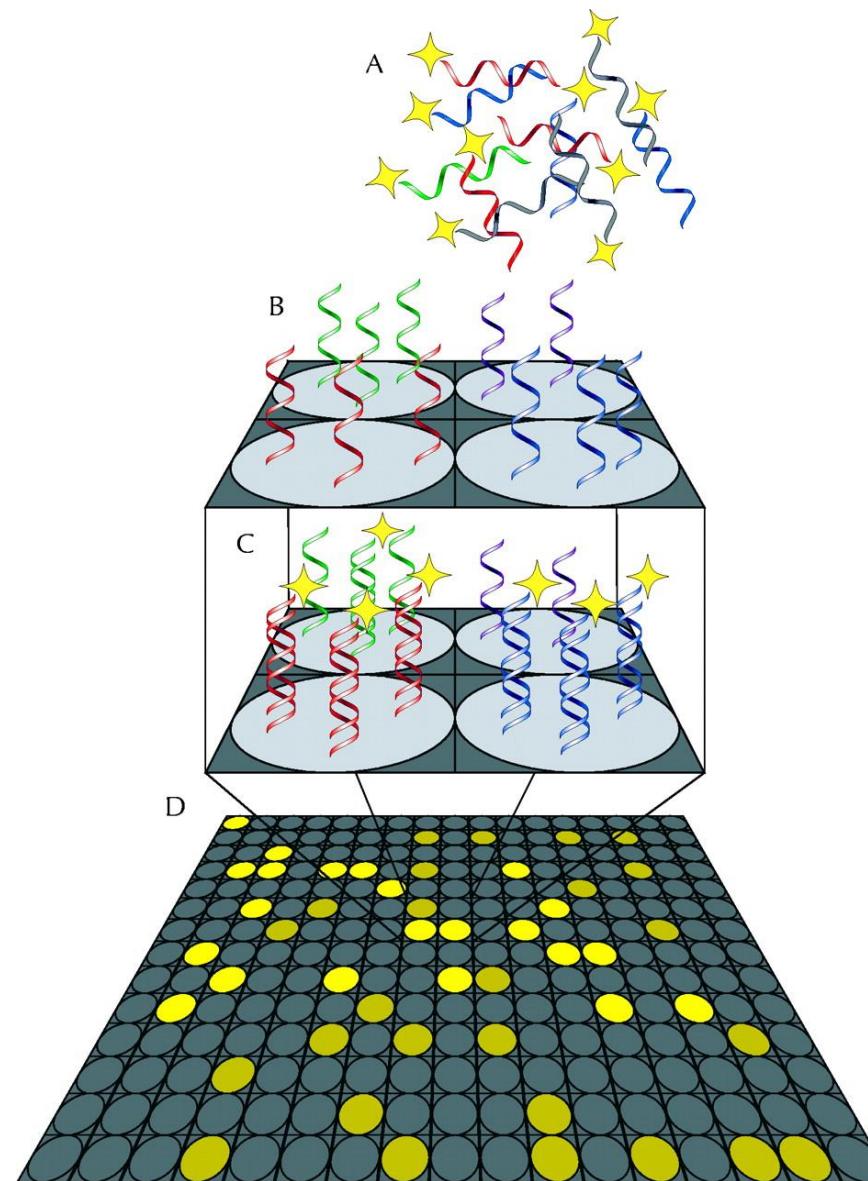
# MICROARRAYS

Is it possible to test for many mutations at once?

- DNA “chips” or microarrays are a possible solution to the problem of testing for multiple mutations.
- In a DNA chip, single DNA strands containing sequences of different human genes are fixed to a solid support (such as a glass slide) in an array pattern, one gene per array point. Denatured DNA prepared from a biopsy or laboratory specimen and prelabelled with fluorescent tags is then applied to the chip. Similarly, DNA that represents all the active genes in a cell can be prepared by (enzymatically) making a DNA copy of the messenger RNA in the cell. During incubation at an appropriate temperature, any **DNA in the sample that is complementary to DNA on the chip binds to the fixed DNA strands** that are its match. A laser is then used to detect the presence of **fluorescence**: a signal at any position on the array indicates that DNA for the gene fixed to the chip at that position is to be found in the sample.

## DNA microarray

- (A) A mixture of DNA or RNA from a diagnostic **sample** is labelled with a fluorescent probe and layered onto a chip.
- (B) The chip consists of a solid support spotted in a grid pattern with thousands to tens of thousands of **different single-stranded DNA sequences**.
- (C) During incubation under carefully chosen conditions, only **sequences complementary to each individual probe bind to that spot on the array**.
- (D) When the array is read, the pattern of **fluorescence** indicates the sequences present in the sample and their quantities.



# MICROARRAYS

The power of this method is in the number of genes that can be examined simultaneously. Present-day chips can hold portions of tens of thousands of individual genes – enough to represent all the genes active in a particular cell, a lymphocyte, for instance. Chips carrying the entire set of human genes are close to realization.

Applications:

- Clinical application of DNA chips is still in the experimental stage. The pattern of gene expression observed by microarray analysis has been used to discriminate between subtypes of leukemia, lymphoma, melanoma and breast cancer, with particular emphasis on patterns of gene expression that mark tumour progression and **treatment resistance**.
- Other potential applications include genetic screening, particularly for diseases that can arise as a result of multiple possible mutations within one gene.
- Related applications are **human genotyping**, to examine multiple **disease markers** simultaneously to determine disease susceptibility, or the emerging field of pharmacogenomics, to **predict responses to medications resulting from variation in types of target receptors** or in enzymes used in the uptake or metabolism of drugs.
- In infectious disease, DNA chips might be used for the rapid identification and **subtyping of bacteria, viruses and parasites**, including antibiotic resistance.

# MICROARRAYS

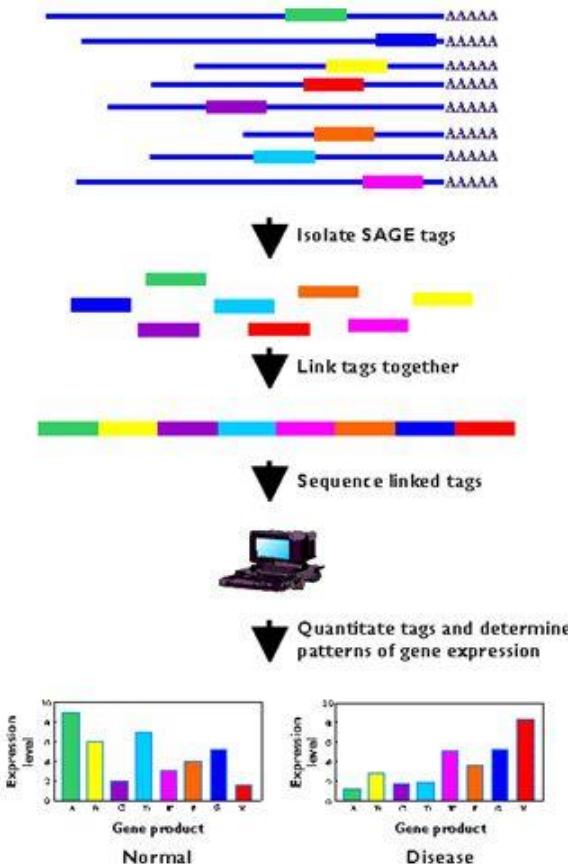
- There are barriers to immediate wide application of the technique.
  - The equipment is **expensive**, as are the arrays themselves, which are produced on a commercial scale by only a limited number of manufacturers.
  - Relatively **large amounts of DNA** are used for analysis, so an amplification step is required, which must in turn amplify all sequences in the DNA proportionately.
  - **Substantial computing power** is required for the analysis of thousands of data points, and there are statistical pitfalls in the analysis of a large number of data points collected from a small number of samples.
  - In addition, gene-level examination is liable to reveal previously unappreciated variations in gene expression between normal individuals and **wide heterogeneity** in disease, which will lead to challenges in **identifying the limits of “normal”** and defining disease.

# SERIAL ANALYSIS OF GENE EXPRESSION (SAGE)

- is a method for identifying and quantifying **transcripts** from eukaryotic genomes.
- SAGE has been widely applied to **analyzing gene expression** in many biological and medical studies.
- The principle of SAGE has been developed to address specific issues such as determination of normal gene structure and identification of abnormal genome structural changes.
- SAGE was designed to gain a direct and quantitative measure of gene expression. The SAGE method is based on the **isolation of unique sequence tags (9-10 bp in length)** from individual mRNAs and concatenation of tags serially into long DNA molecules for a lump-sum sequencing.
- The SAGE method can be applied to the studies exploring virtually any kinds of biological phenomena in which the changes in cellular transcription are responsible. SAGE is a highly competent technology that can not only give a global **gene expression profile** of a particular type of cell or tissue, but also help us identify a set of specific genes to the cellular conditions by comparing the profiles constructed for a pair of cells that are kept at different conditions. The SAGE technique can be used in a wide variety of applications including analysis of the **effect of drugs** on tissues or identification of disease-related genes.

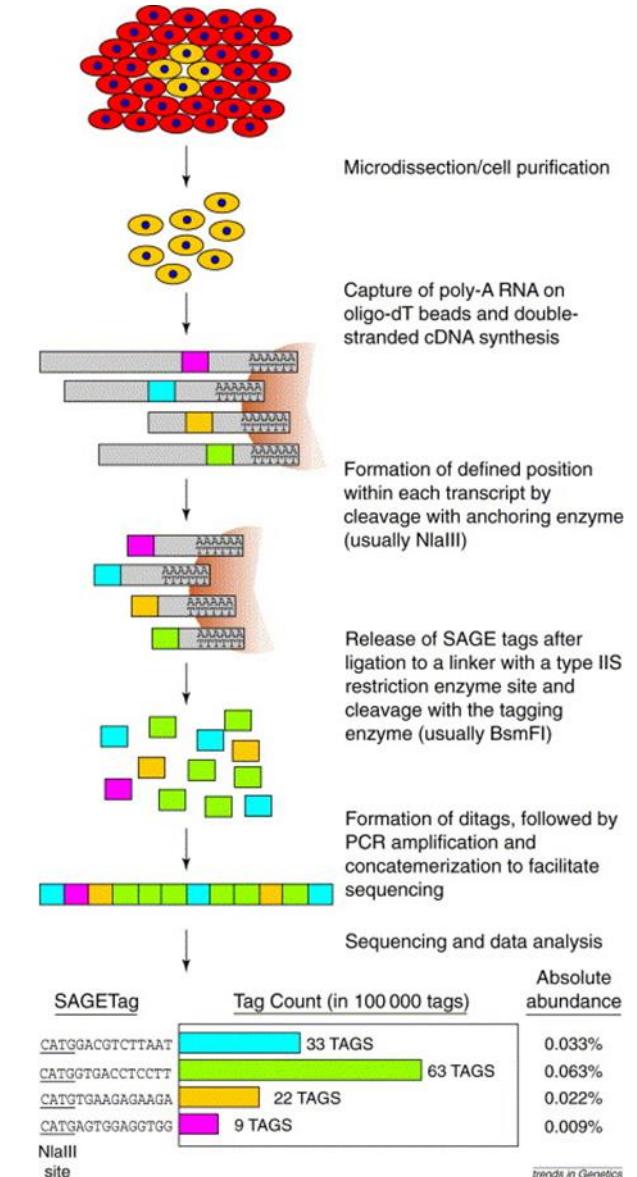
# SAGE

- Serial Analysis of Gene Expression
- The basis of this technique is that a gene can be uniquely identified using only a small (10-30 nt) piece from the 3' end (which is not translated)
- These tags are extracted (from cDNA), then concatenated into long molecules that are amplified with PCR (or cloned) and sequenced.
- The number of times each tag appears is proportional to the amount of its mRNA present.
- Much SAGE data in NCBI.



# SERIAL ANALYSIS OF GENE EXPRESSION (SAGE)

- Serial analysis of gene expression (SAGE) is a powerful technique that can be used for global analysis of gene expression.
- Its chief advantage over other methods is that it does not require prior knowledge of the genes of interest and provides qualitative and quantitative data of potentially every transcribed sequence in a particular cell.
- This is a technique of expression profiling, which **permits simultaneous, comparative and quantitative analysis** of gene-specific, 9- to 13-basepair sequences. These short sequences, called SAGE tags, are linked together for efficient sequencing.
- The sequencing data are then analyzed to identify each gene expressed in the cell and the levels at which each gene is expressed.



# DETECTING MUTATIONS IN HUMAN GENES

How are mutations identified?

- Until the 1980s, the only directly observed DNA mutations were those abnormalities large enough to be detected by **karyotyping**, the process of examining an array of chromosomes seen under a microscope.
- Mutations that resulted in missing or altered protein function, such as phenylketonuria, could be inferred from the results of clinical **biochemistry**.
- Since then, innovations in the manipulation of **DNA** have given rise to a panoply of methods for detecting genetic abnormalities, the appropriate method depending upon the size and nature of the mutation. Some techniques are applied to the chromosomal DNA itself, some to the RNA copies produced by transcription of active DNA and some to the protein product of the gene. All exploit one or more of the basic properties of DNA or the enzymes that act upon it.

# DETECTING MUTATIONS IN HUMAN GENES

Single base pair mutations can be identified by any of the following methods:

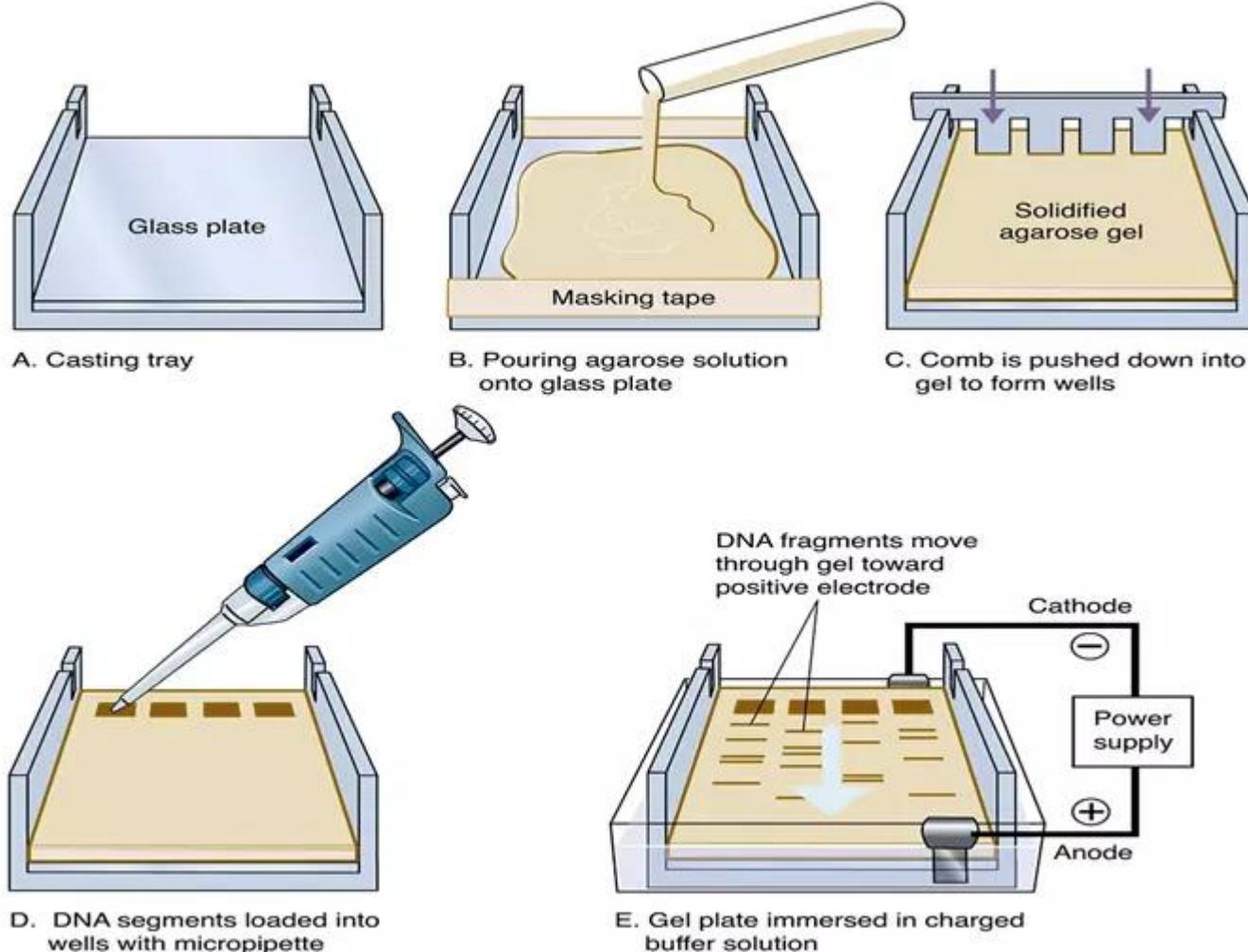
- Direct sequencing, which involves identifying each individual base pair, in sequence, and comparing the sequence to that of the normal gene. This is a labour-intensive method reserved for previously unidentified mutations or rare mutations of a common disease (such as cystic fibrosis), when other methods do not detect the disease that is clinically suspected.
- DNA hybridization methods, which make use of the strong binding of a lone DNA strand to a strand whose sequence is the perfect complement of the first and which can thus discriminate between DNA that contains mutations and DNA that contains the normal sequence.
- Restriction enzyme digestion: Restriction enzymes are specialized enzymes that recognize and cut the DNA double helix wherever they encounter a specific very short sequence (the particular sequence depending upon the enzyme). Single base pair mutations may remove or, alternatively, create one of these sequences and thus alter the sizes of the DNA fragments that result.

# DETECTING MUTATIONS IN HUMAN GENES

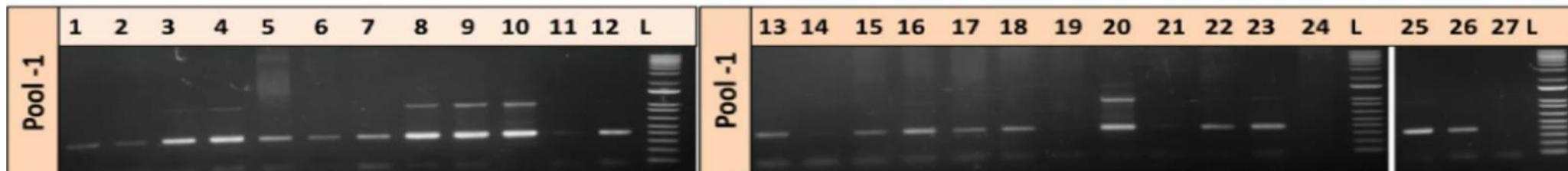
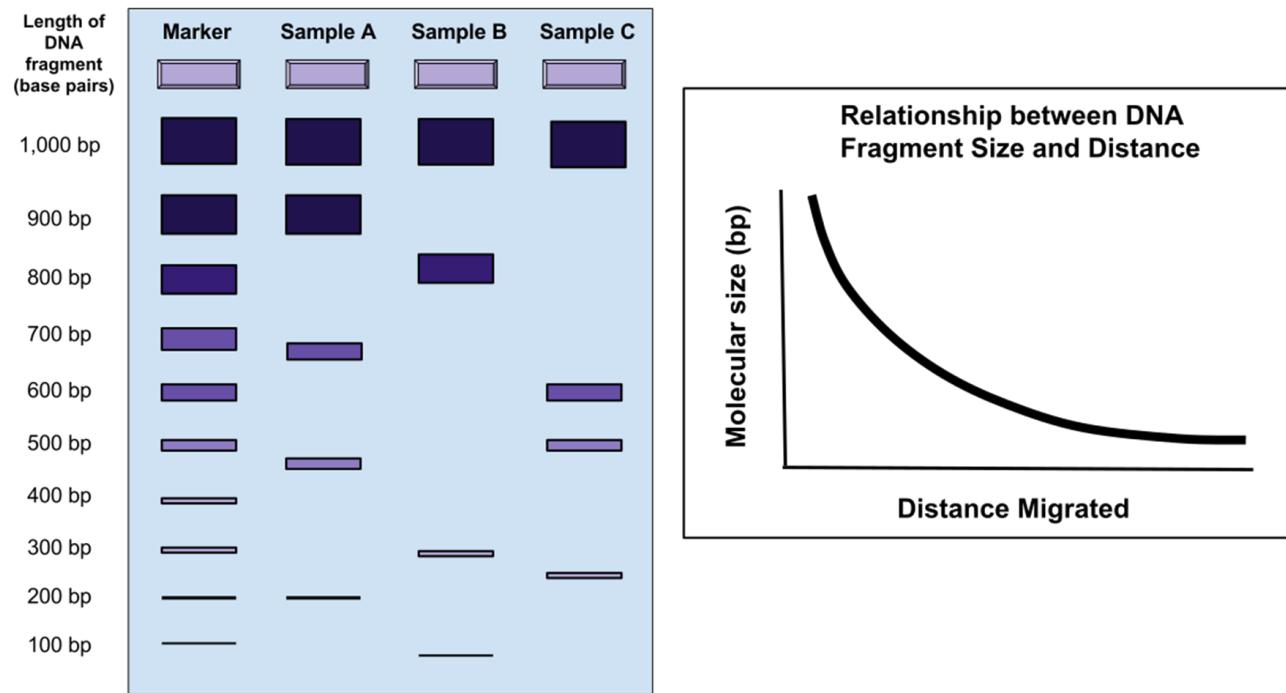
Larger mutations involve the deletion, rearrangement, expansion or duplication of parts of genes, entire genes or multiple genes:

- The presence or absence of a gene or substantial part of a gene can be determined by hybridization, with a labelled “probe” containing the gene sequence. If the gene is present, there will be a signal; if not, there is no signal.
- By changing the distance between sites or removing sites entirely, large mutations alter the characteristic pattern of cuts in the genome that are made by restriction enzymes. For example, hereditary diseases caused by duplication of a gene; the mutated chromosome is cut and compared with the normal chromosome.
- A number of strategies use the polymerase chain reaction to amplify specifically the region involving the mutation.
- Some assays detect not the mutation itself, but the altered protein it produces. Current assays for BRCA1 and BRCA2 mutations are based on detection of the truncated (shortened and nonfunctional) protein produced by the mutated genes.

# GEL ELECTROPHORESIS

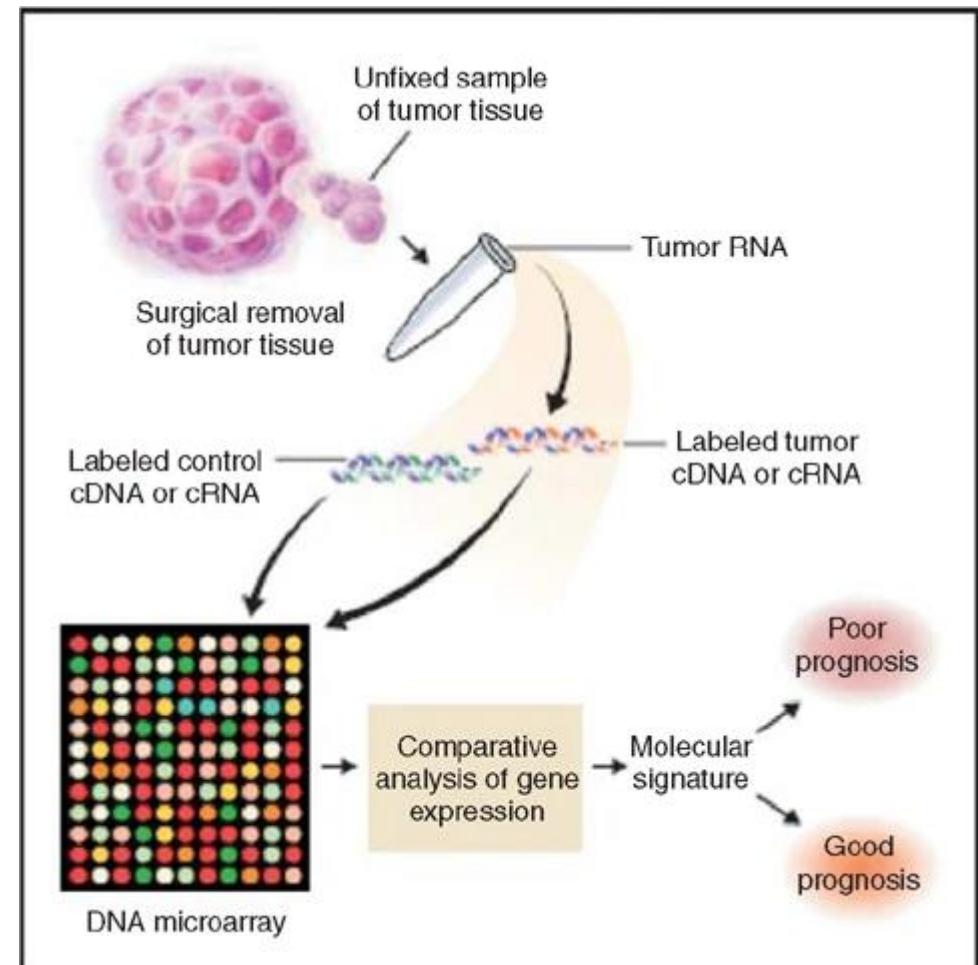


# GEL ELECTROPHORESIS



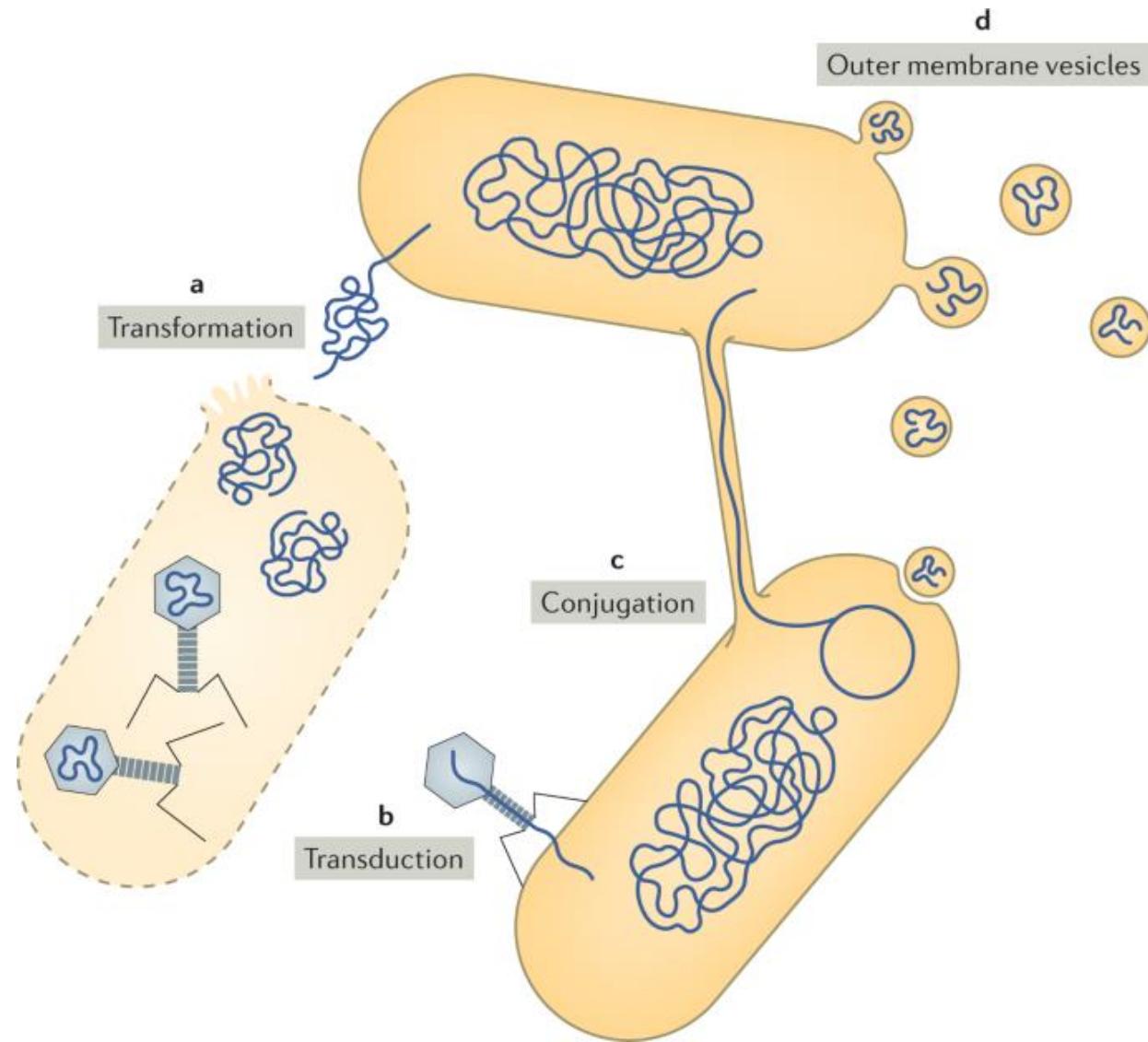
# MASSARRAY-BASED GENOMIC TESTING IN HUMAN GENES

- Colorectal cancer (CRC) is one of the leading causes of cancer death around the world. A mass of clinical researches, through comprehensive genomic approaches, had highlighted the CRC genomic landscape complexity.
- Mutations and variants of these genomes are identified as prognostic and/or predictive makers for CRC.
- A robust and cost-effect method to simultaneously characterize all clinical relevant mutations and genomic variants is urgently needed for realization of personalized medicine (e. g. DNA microarrays)

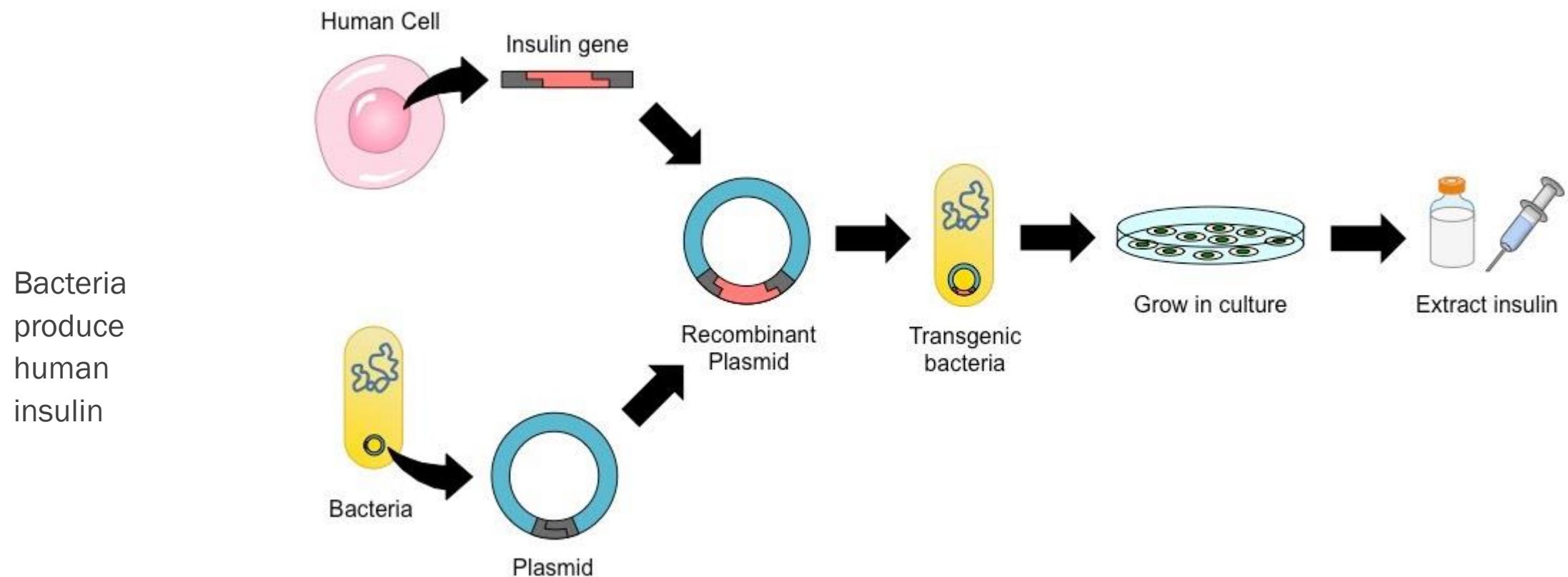


# GENE TRANSFER

horizontal gene transfer  
in microbial communities



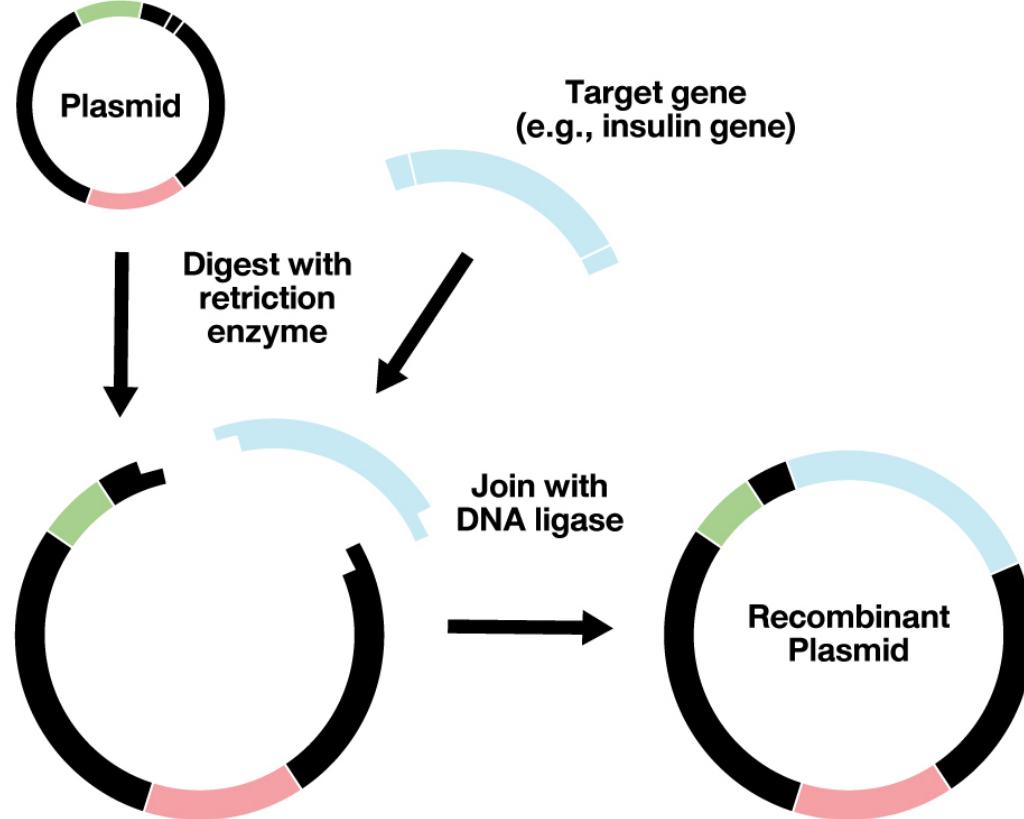
# GENE TRANSFER



# GENETIC ENGINEERING

Genetic engineering is the process of using recombinant DNA (rDNA) technology to alter the genetic makeup of an organism which is done by:

- Extracting DNA from human cells.
- Restriction enzymes, as known as molecular scissors cut DNA into smaller pieces, leaving unpaired bases called 'sticky ends'
- The same restriction enzyme is used to cut plasmids in bacteria. This opens up their circular shape and leaves sticky ends as well. The sticky ends at the cut human DNA and the sticky ends of the plasmid have complementary bases.
- The human DNA and the cut plasmid are mixed together. Their complementary bases and ligase enzymes (aka. 'The glue') join the unpaired bases of the human DNA and the cut plasmid together.
- A recombinant DNA is formed; containing DNA from humans and bacteria.
- The recombinant DNA is put into the bacteria which is then fermented. This allows them to reproduce offspring which contain copies of recombinant DNA.



# How Crops are Genetically Modified

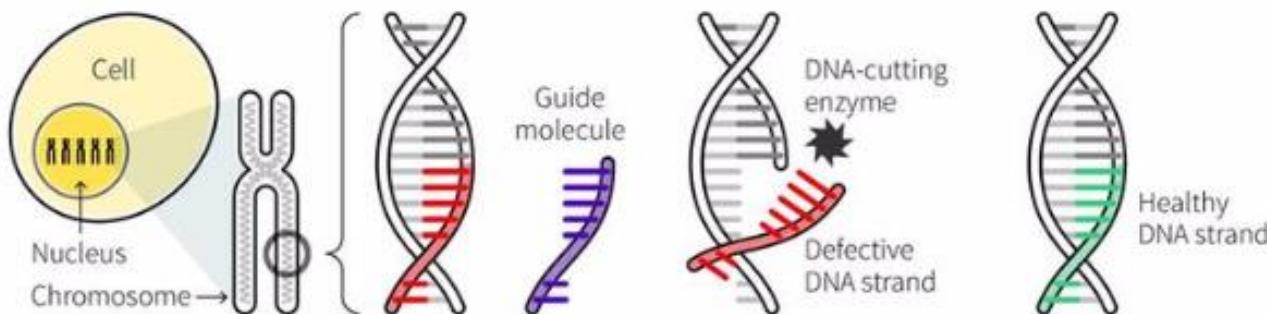
Traditional Breeding	Mutagenesis	RNA Interference	Transgenics	Gene Editing
Crossing plants and selecting offspring  Desired gene(s) inserted with other genetic material	Exposing seeds to chemicals or radiation  Random changes in genome, usually unpredictable 	Switching off selected genes with RNA  Targeted gene(s) switched off or 'silenced' 	Inserting selected genes using recombinant DNA methods  Only gene(s) inserted at desired locations selected 	When used to delete genes using engineered nucleases (CRISPR, TALENs, ZFNs, etc.)  Desired gene(s) deleted only at known locations 
Almost all crops				
<b>Number of genes affected:</b> few genes to whole genomes	100s - 1,000s	1 – dozens	1 – 8	1 or more

# GENE EDITING

## DNA editing

A DNA editing technique, called CRISPR/Cas9, works like a biological version of a word-processing programme's "find and replace" function.

### HOW THE TECHNIQUE WORKS



A cell is transfected with an enzyme complex containing:

- Guide molecule
- Healthy DNA copy
- DNA-cutting enzyme

A specially designed synthetic guide molecule finds the target DNA strand.

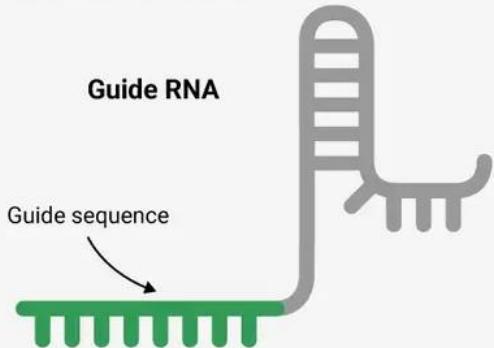
An enzyme cuts off the target DNA strand.

The defective DNA strand is replaced with a healthy copy.

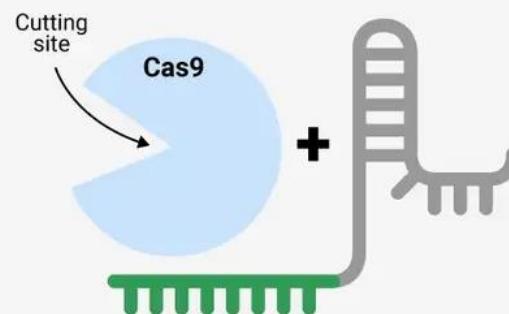
Sources: Reuters; Nature; Massachusetts Institute of Technology

# EDITING A GENE USING THE CRISPR/CAS9 TECHNIQUE

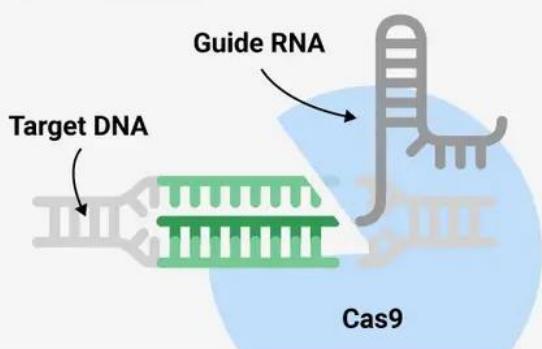
- 1** Scientists create a genetic sequence, called a "guide RNA," that matches the piece of DNA they want to modify.



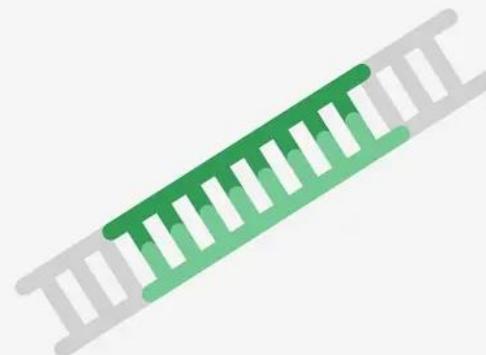
- 2** This sequence is added to a cell along with a protein called Cas9, which **acts like a pair of scissors** that cut DNA.



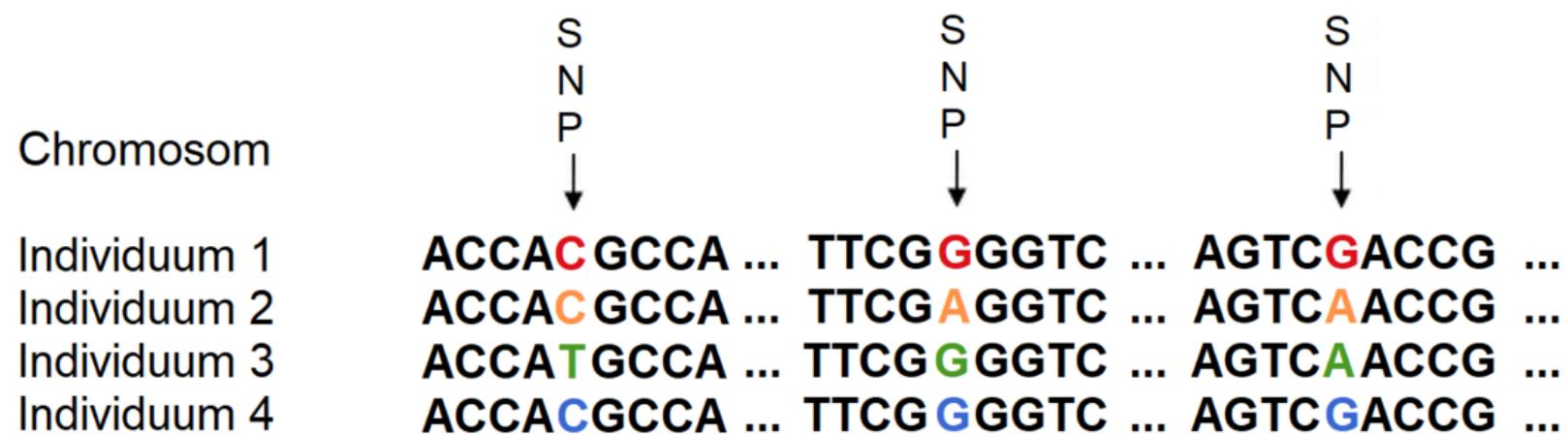
- 3** The guide RNA homes in on the target DNA sequence, and Cas9 **cuts it out**. Once their job is complete, the guide RNA and Cas9 leave the scene.



- 4** Now, another piece of DNA is swapped into the place of the old DNA, and **enzymes repair the cuts**. Voilà, you've edited the DNA!

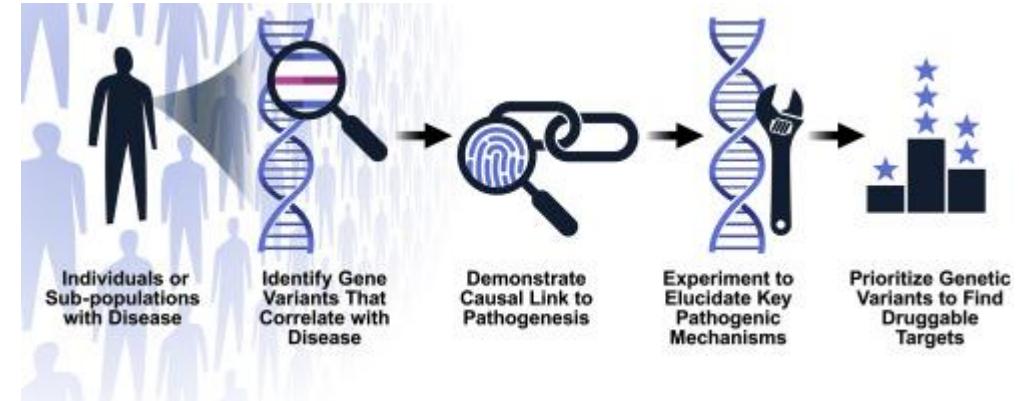


# SINGLE NUCLEOTIDE POLYMORPHISM



- SNPs are common variants in DNA that can have one of the 4 DNA bases at a single site. This leads to variations between different individuals.
- SNPs are important for predicting response to drugs

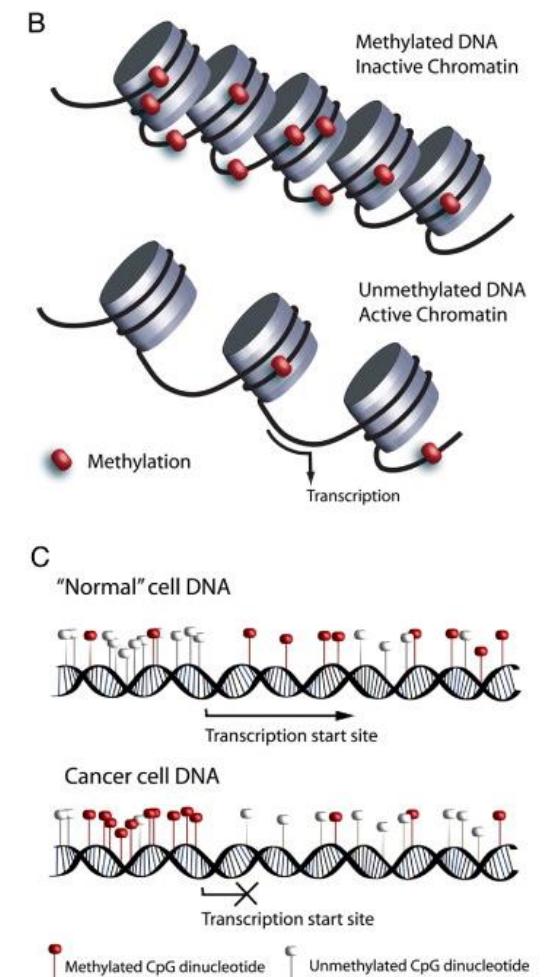
# WHAT DOES IT MEAN TO HAVE A GENETIC PREDISPOSITION TO A DISEASE?



- A genetic predisposition (sometimes also called genetic susceptibility) is an increased likelihood of developing a particular disease based on a person's genetic makeup. A genetic predisposition results from specific genetic variations that are often inherited from a parent. These **genetic changes contribute to the development of a disease but do not directly cause it**. Some people with a predisposing genetic variation will never get the disease while others will.
- Genetic variations can have large or small effects on the likelihood of developing a particular disease. For example, certain variants (also called mutations) in the BRCA1 or BRCA2 genes greatly increase a person's risk of developing breast cancer and ovarian cancer. Particular variations in other genes, such as BARD1 and BRIP1, also increase breast cancer risk, but the contribution of these genetic changes to a person's overall risk appears to be much smaller.
- Current research is focused on identifying genetic changes that have a small effect on disease risk but are common in the general population. Although each of these variations only slightly increases a person's risk, having changes in several different genes may combine to increase disease risk significantly. Changes in many genes, each with a small effect, may underlie susceptibility to many common diseases, including cancer, obesity, diabetes, heart disease, and mental illness. Researchers are working to **calculate an individual's estimated risk for developing a common disease based on the combination of variants in many genes** across their genome. This measure, known as the **polygenic risk score**, is expected to help guide healthcare decisions in the future.
- In people with a genetic predisposition, the risk of disease can depend on multiple factors in addition to an identified genetic change. These include other genetic factors (sometimes called modifiers) as well as lifestyle and environmental factors. Diseases that are caused by a combination of factors are described as multifactorial. Although a person's genetic makeup cannot be altered, some lifestyle and environmental modifications (such as having more frequent disease screenings and maintaining a healthy weight) may be able to reduce disease risk in people with a genetic predisposition.

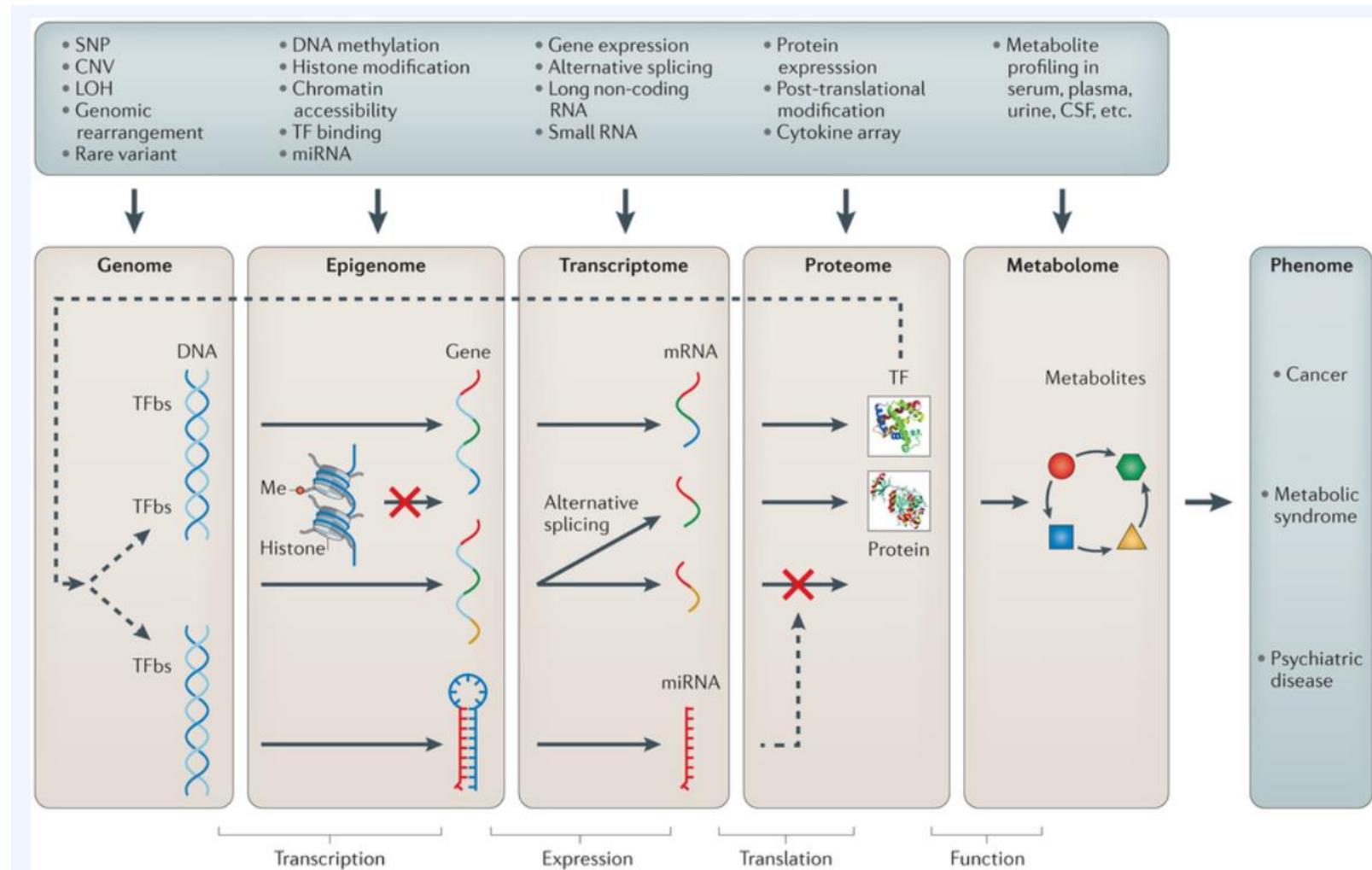
# METHYLOMIC DEEP SEQUENCING OF OVARIAN CANCER REVEALS DIAGNOSTIC AND PROGNOSTIC SIGNATURES

- The heterogeneity of ovarian cancer requires biomarkers for personalized medicine. **DNA methylation aberration is a hallmark of cancer.** The attempt of using DNA methylation as biomarkers for prognosis prediction is appealing.
- With the advancement of high throughput technology, we are now able to assess global DNA methylation of ovarian cancer using methyl-DNA capture coupled with next-generation sequencing (MethylCap-seq) technology.
- Methylomic analysis provides a methylomic signature for prognosis prediction, which will be of great help for the personalized chemotherapy especially the use of demethylation agents of ovarian cancer patients.



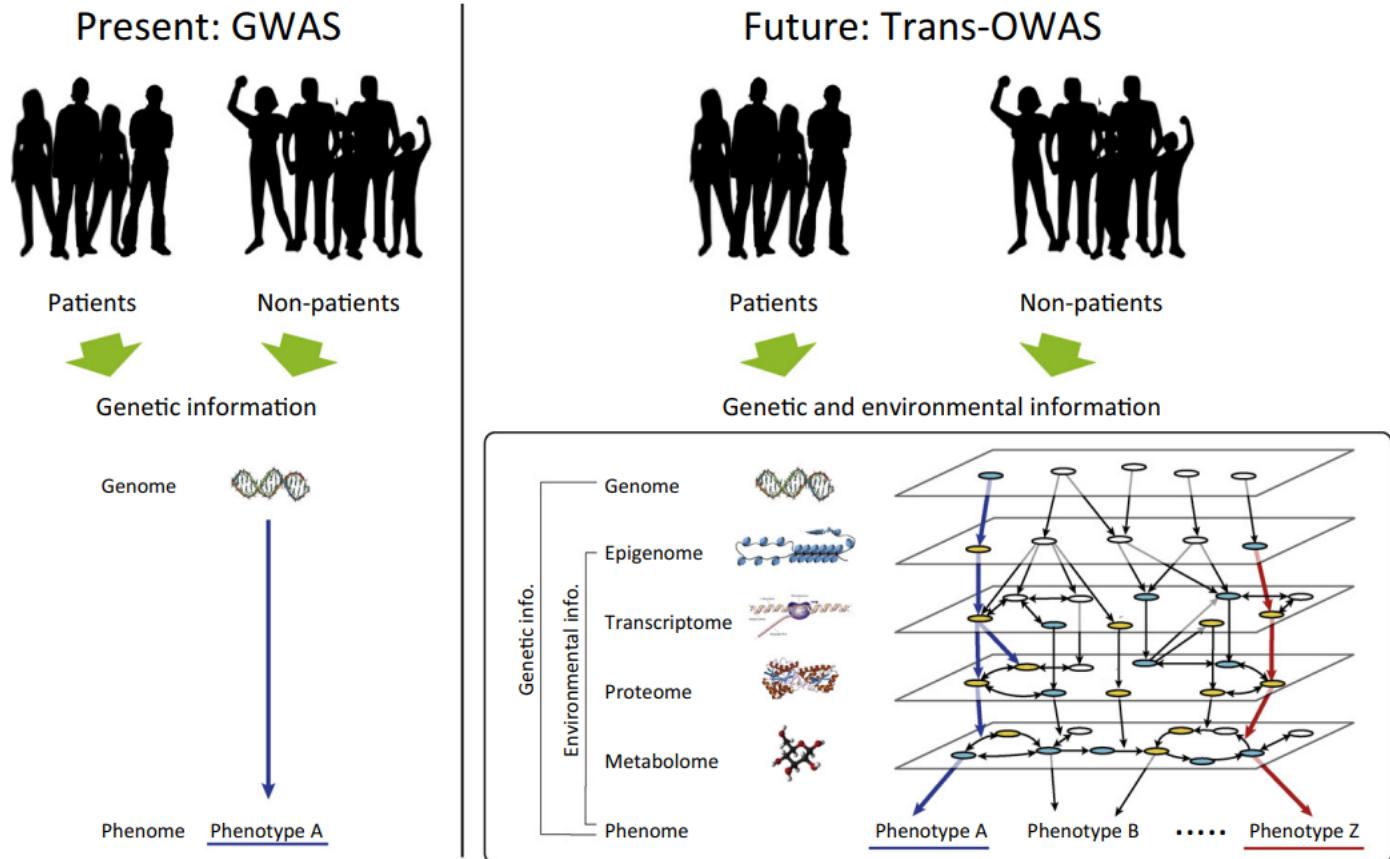
# INTEGRATING DATA TO UNCOVER GENOTYPE-PHENOTYPE INTERACTIONS

- Recent technological advances have expanded the breadth of available omic data, from whole-genome sequencing data, to extensive transcriptomic, methylomic and metabolomic data.
- A key goal of analyses of these data is the identification of effective models **that predict phenotypic traits and outcomes**, elucidating biomarkers and generating insights into the genetic underpinnings of the heritability of complex traits.
- There is still a need for powerful analysis strategies to fully harness the utility of these comprehensive data, identifying true associations.
- Emerging approaches for data integration - including meta-dimensional and multi-staged analyses - aim to deepen our understanding of the role of genetics and genomics in complex outcomes. By developing these approaches, an **improved understanding of the relationship between genomic variation and human phenotypes** may be revealed.



# TRANS-OMIC STUDIES

- e. g. case studies of metabolism-centric trans-omic studies to show how to reconstruct a biochemical trans-omic network by connecting multi-omic data and how to analyze it
- Lifestyle diseases like type 2 diabetes are largely elicited by multiple factors belonging to multiple omic layers that are influenced not only by genetic factors but also by environmental factors.
- GWAS (genome wide association studies) can associate phenotypes only with genetic factors. In **trans-OWAS** (ome-wide association studies), the individual network is reconstructed from the multiple omic data.



# MULTI-OMIC APPROACH

- While the genome remains mostly stable over time, other “omes” change based on what genes are turned on and off at particular moments in particular places in the body. The **proteome** (all an organism’s proteins) and the **metabolome** (all the metabolites, or small molecules that are the outputs of biological processes) are two of several powerful datasets that become more informative when used together in a multi-omic approach. They show how that genomic instruction manual is actually being applied.
- “The genome tells you what can happen, the proteome and the metabolome can show what’s actually going on”. Just as city planners use data about traffic patterns to figure out where to widen roads and how to time stoplights, biologists can use those entwined networks to predict at a molecular level how individual organisms will respond under specific conditions.
- By linking these layers and others to expand from genomics to multi-omics, scientists might be able to meet the goals of **personalized medicine**: to figure out, for example, **what treatment a particular cancer patient will best respond to**, based on the network dynamics responsible for a tumor. Or predict whether an experimental vaccine will work before moving into expensive clinical tests. Or help crops grow better during a drought.

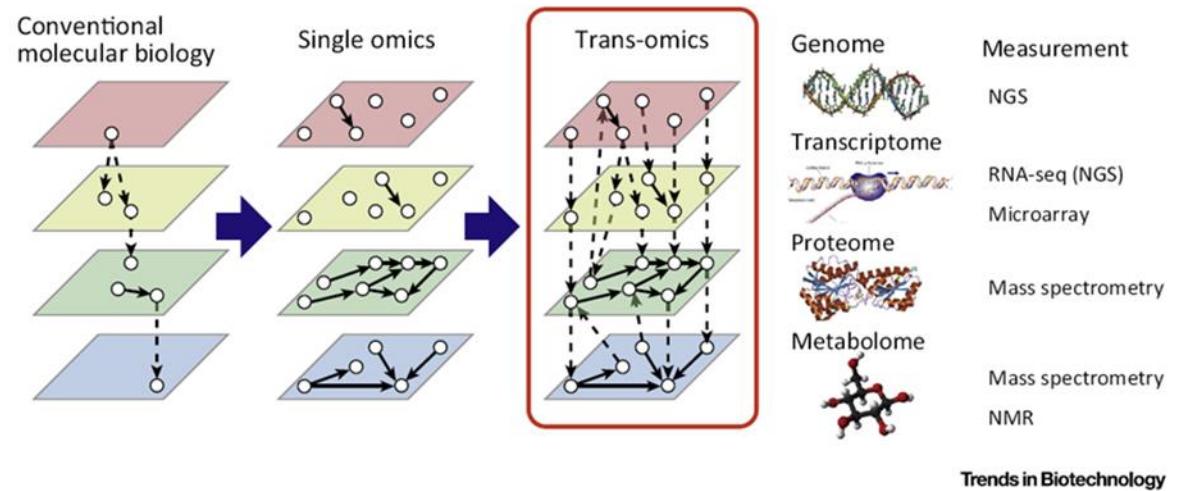
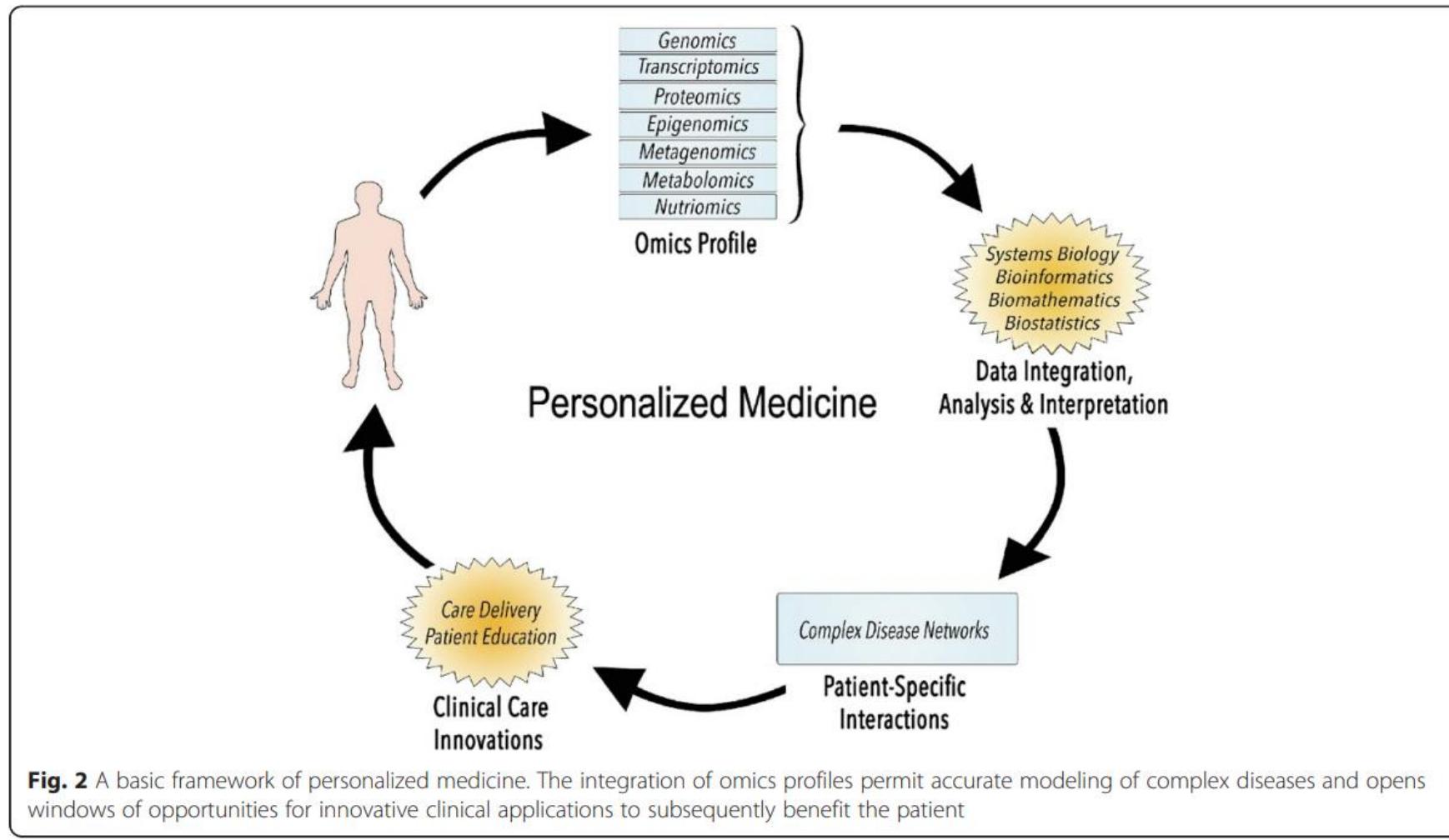
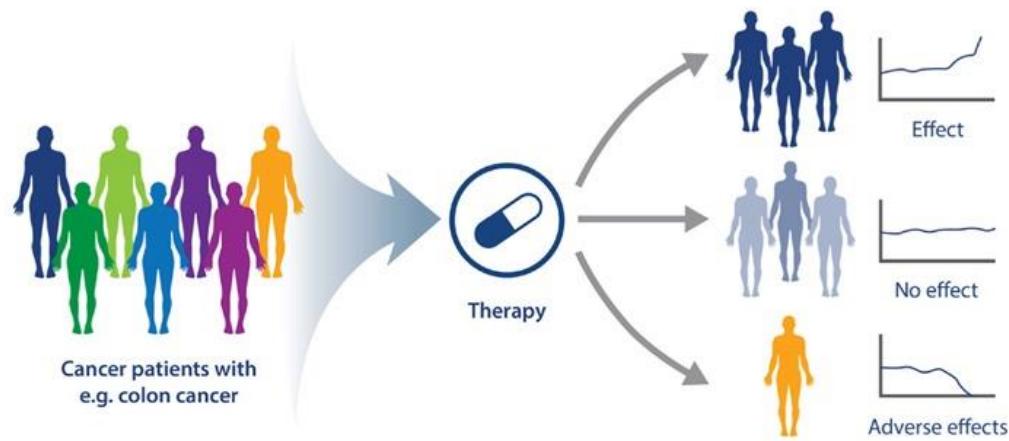


Figure 1. Trans-Omic Network Across Multiple Omic Layers (from Left to Right). Conventionally, a network has been identified by accumulating literature on specific molecules. Measurement of a single omic layer has now become available. Trans-omics is becoming available by connecting multi-omic measurements. A group of molecules with similar chemical properties, such as genome, transcriptome, proteome, and metabolome, is called an ‘omic’ layer, which can be measured by next-generation sequencers (NGS), microarray, mass spectrometry, and NMR. This figure partly includes ‘Process of transcription’ by NHS National Genetics and Genomics Education Centre licensed under CC BY 2.0/modified from the original ([www.flickr.com/photos/119980645@N06/13080846733/in/photostream/](http://www.flickr.com/photos/119980645@N06/13080846733/in/photostream/)).



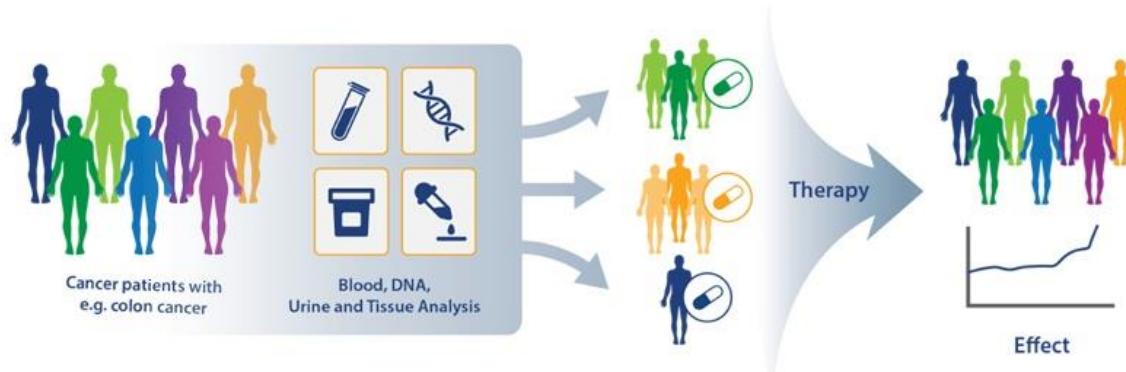
## Current Medicine

One Treatment Fits All



## Future Medicine

More Personalized Diagnostics



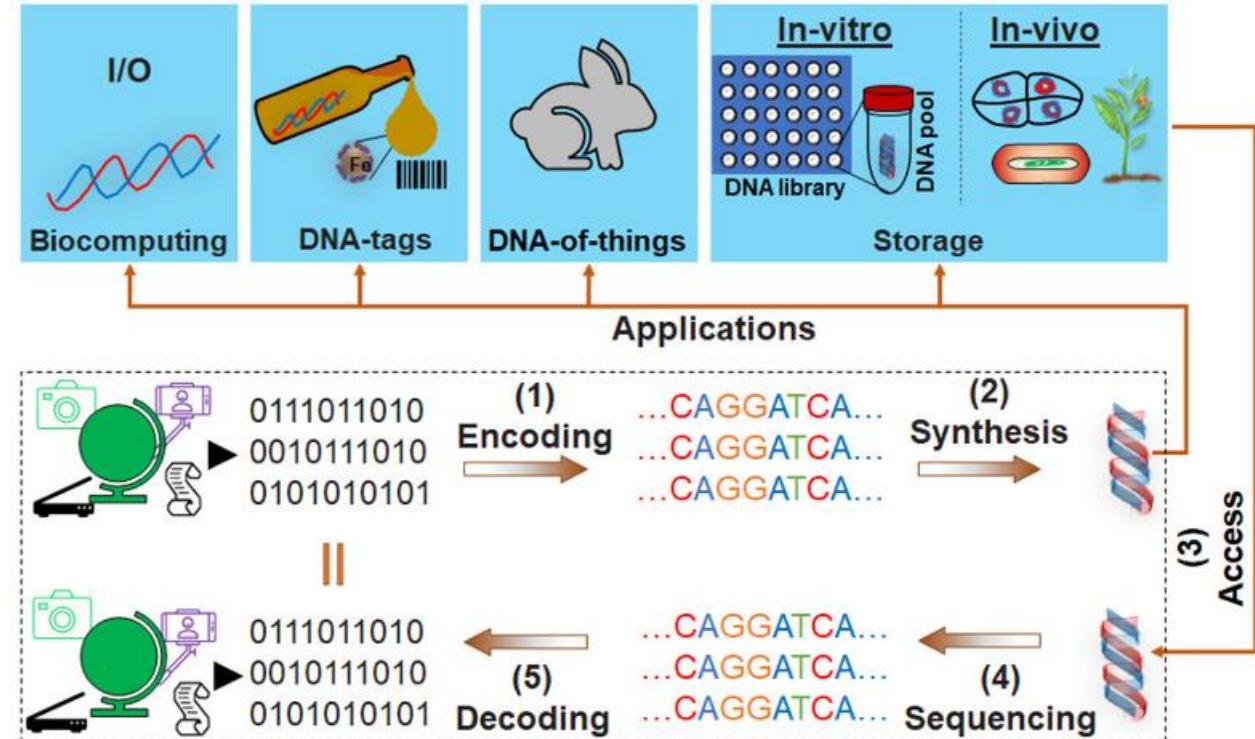
# DNA STORAGE

- data scientists are continually looking for better, more stable, and space-efficient alternatives to store these huge datasets. Because of its unique biological properties, highly condensed DNA has great potential to become a storage material for the future. Indeed, DNA-based data storage has recently emerged as a promising approach for long-term digital information storage.
- Four unique biological features make DNA the focus of the next generation of digital information storage.
- First, DNA is remarkably stable compared with other storage media. With its double-helix structure and base-stacking interactions, DNA can persist 1,000 times longer than a silicon device, and survive for millennia, even in harsh conditions.
- Second, DNA possesses a high storage density. Theoretically, each gram of single-stranded DNA can store up to 455 exabytes of data. As storage strategies continue to improve, scientists have now achieved a density that could reach this theoretical limit.
- Third, DNA can be easily and rapidly replicated through the PCR, thereby providing the possibility for large-scale data backup. It should not be neglected that living cells are also perfect tools for *in vivo* information replication and backup.
- Last but not least, the biological properties of DNA enable current sequencing and chemical synthesis technologies to read and write the information stored in DNA, thereby making it an excellent material to store and retrieve data.

# ENCODING - DECODING

Basics: Each nucleotide is 2 bits	
A	00
C	01
G	10
T	11

Converting binary 0 and 1  
into a sequence of bases

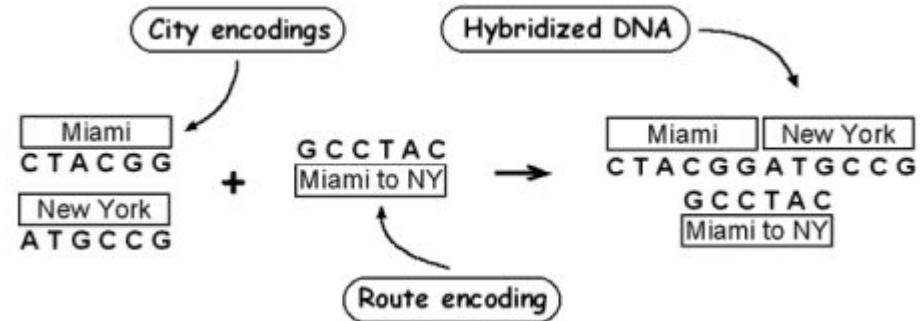


# DNA COMPUTING



- There's no denying the fact that scientists need to look for silicon alternatives as the silicon-based computers have a finite number of capabilities.
- The concept of DNA computing was first introduced in 1994. It deals with “biochips” made of DNA that are able to **perform billions of calculations at once** by multiplying themselves in number. DNA’s ability to replicate can execute an exponential number of paths. In other words, a DNA computer grows as it computes. Conventional computers perform calculations one after another, a DNA computer does those calculations at once by making multiple copies of itself.
- In 1994 the concept of DNA computing was invented by the famous cryptographer Leonard Adleman who used DNA to solve the “traveling salesman” problem. The problem aimed at finding out the shortest route between a number of cities by going through each city only once. Adleman showed that billions of molecules in a drop of DNA had so much computational power that can simply overpower silicon computers.

# WORLD'S FIRST DNA COMPUTER



- Adleman's first DNA computation solved a traveling salesman problem of seven cities.
- The method of doing so is surprisingly simple. Each city is represented by a unique sequence of bases. Connections between two cities are created from a combination of the complement of the first half of the sequence of one city, and the complement of the second half of the sequence of a connected city. In this way DNA representing the trip will be created with one strand representing a sequence of cities and the complementing strand representing a series of connections.
- The above will create DNA representing all sorts of trips. Some trips may not visit all cities, while other trips may visit some cities more than once. The next step is to process the data until the correct trip remains (one that starts at a certain city and ends in another city after visiting all the others once). The correct answer is obtained by filtering out trips that start and end in the correct cities, then filtering trips with the correct number of cities, and finally filtering out trips that contain each city only once.

# DNA COMPUTING

- DNA molecules are able to store billions of times more data as compared to tradition storage devices.
- Due to an abundance availability of DNA, it's a **cheap resource**. Also, a DNA computer will be **environment-friendly and compact in size**.
- A single gram of dried DNA is capable of storing the same amount of information as could fit on one trillion CDs. The
- This, along with the **benefits of parallel processing** and the negligible power required, guarantee that the DNA computer, or nanocomputer, will continue to be refined and perfected (disadvantage: rather slow).
- Currently, the field of DNA computing is in its nascent stage, and it'll take a long time to develop a working DNA computer. Whatever might be the rate of progress, the concept surely sounds exciting.