Bioinformatics

Lecture 02

Course Code: CSE 469

Credit: 3.0

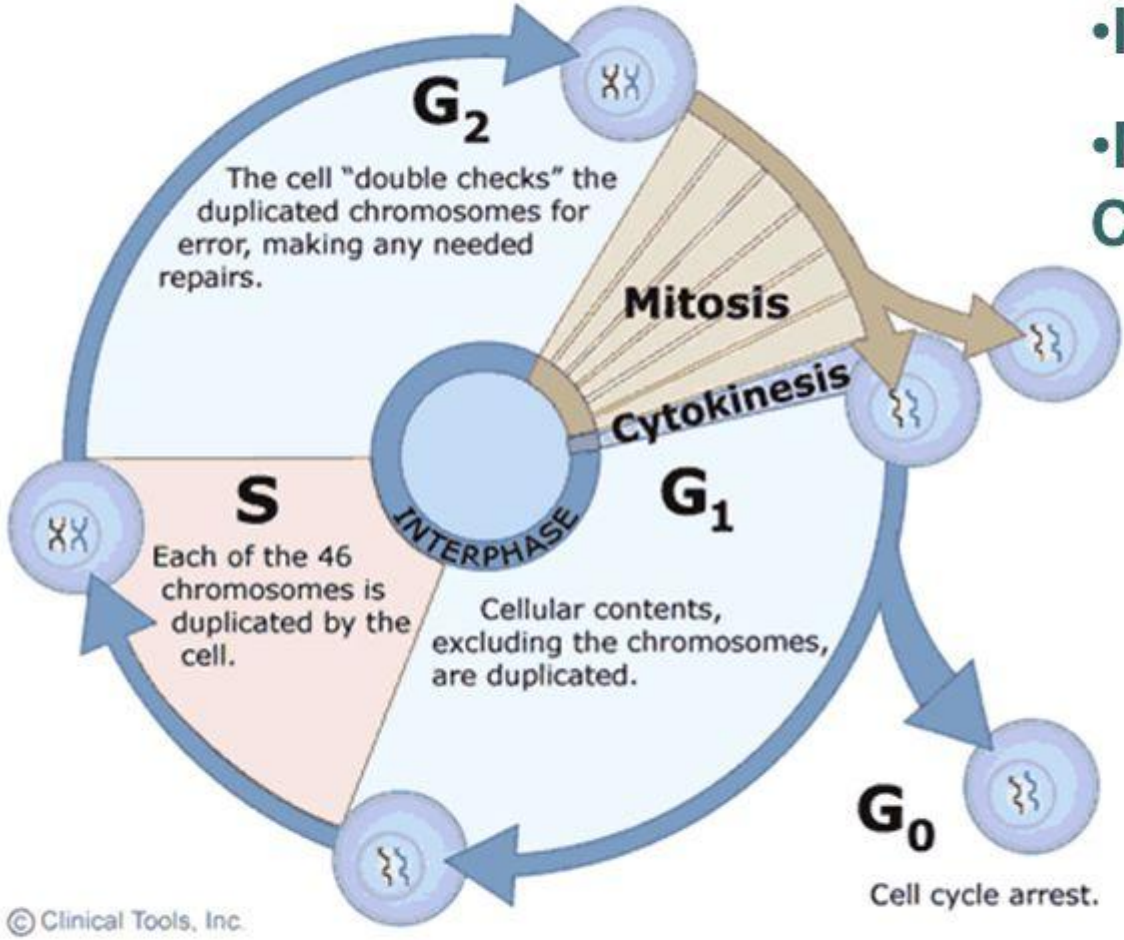
Total Course Hour: 36

Cell Cycle

- Living cells go through a series of stages known as the **cell cycle**. The cells grow, copy their chromosomes, and then divide to form new cells.

1. **G1 phase**: The cell grows.
2. **S phase**: The cell makes copies of its chromosomes. Each chromosome now consists of two sister **chromatids**.
3. **G2 phase**: The cell checks the duplicated chromosomes and gets ready to divide.
4. **M phase**: The cell separates the copied chromosomes to form two full sets (mitosis) and the cell divides into two new cells (cytokinesis).
5. **Interphase:** The period between cell divisions is known as **'interphase**'.
6. **G0 phase:** Cells that are not dividing leave the cell cycle and stay in **G0**.

The Cell Cycle



Consists of:

- **Interphase**
- **Mitosis/ Cytokinesis**

Mitosis is a small part of the cycle; a cell spends almost all of it's time in interphase.

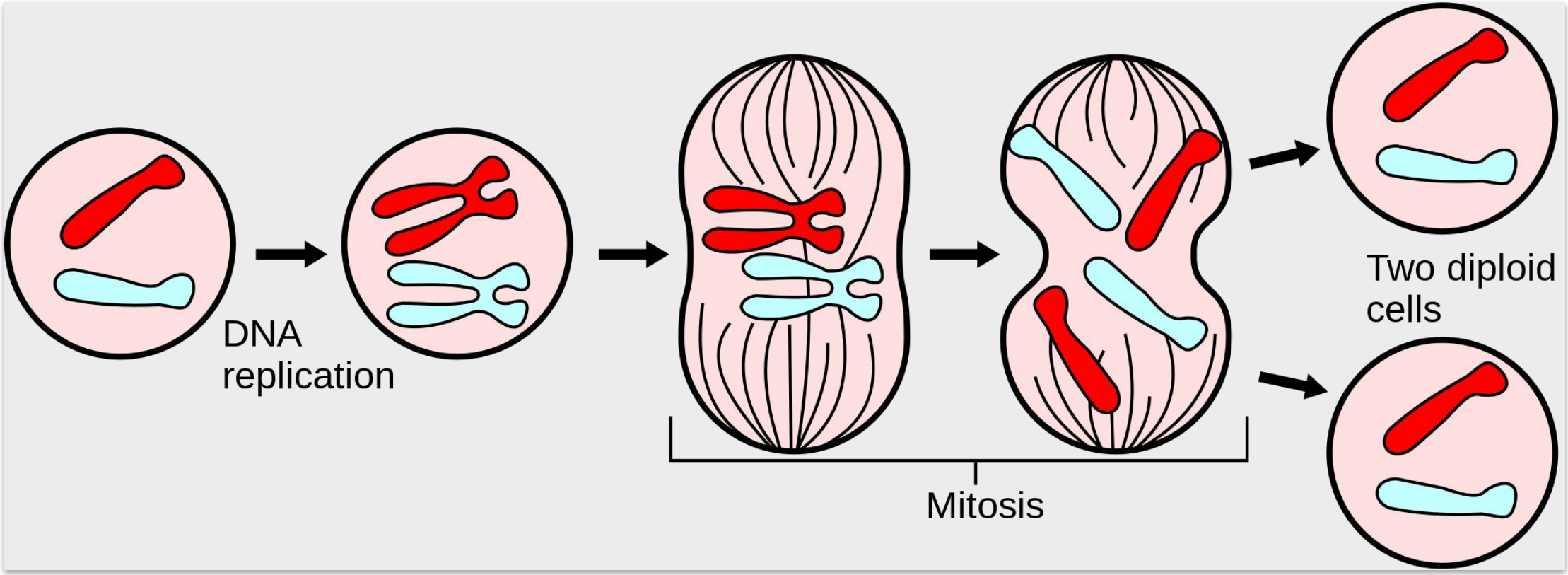
Most cells don't continually go through the cell cycle; many stay in interphase, without preparing to divide, for long periods (G_0).

Cell Division

- All cells are derived from **pre-existing** cells
- New cells are produced for **growth** and to **replace** damaged or old cells
- Differs in prokaryotes (bacteria) and eukaryotes (plants & animals)
- DNA must be copied or replicated before cell division
- Each new cell will then have an **identical copy** of the DNA
- Two types **Mitosis** & **Meiosis**

Mitosis

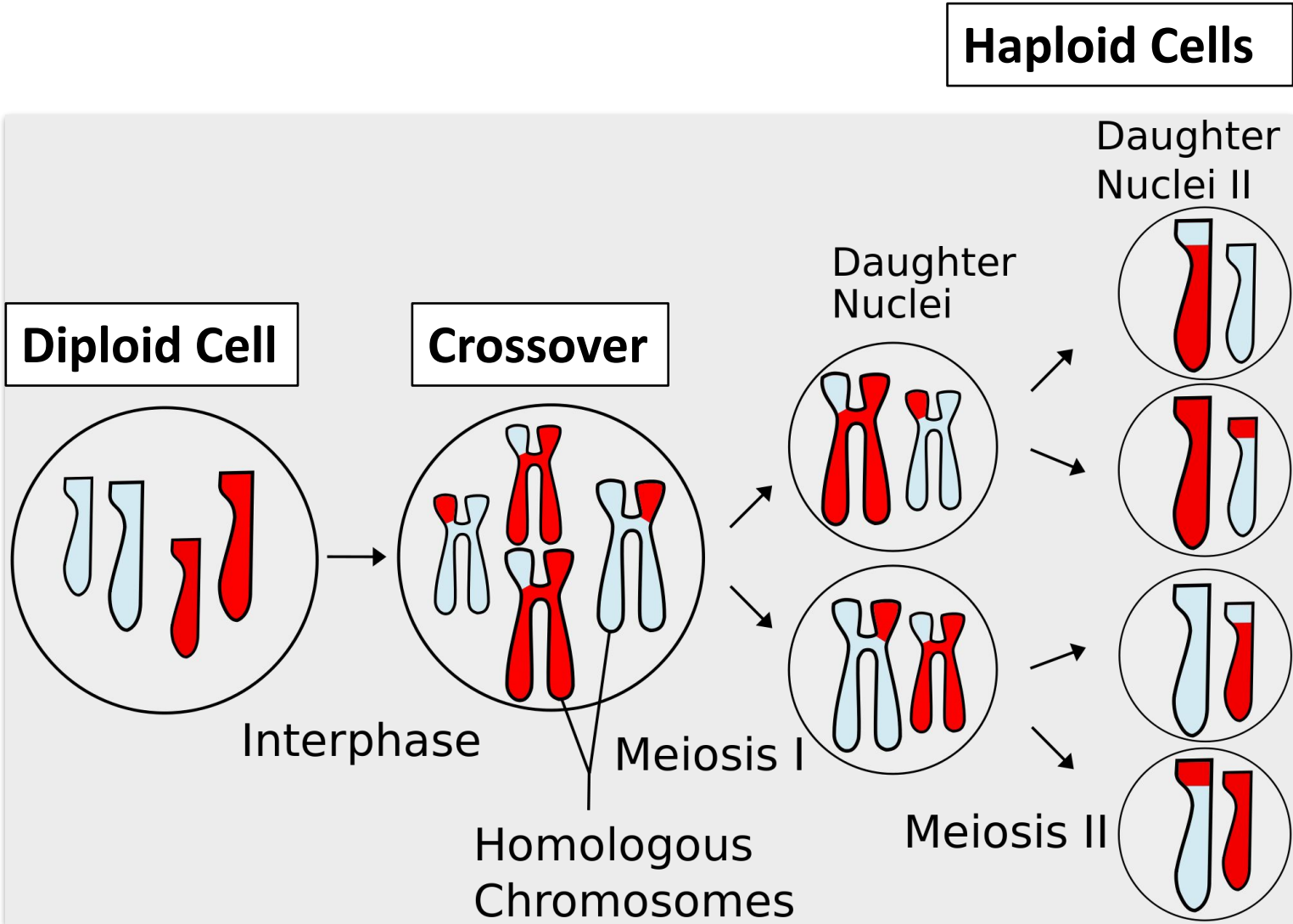
Mitosis is a process of nuclear division in eukaryotic cells that occurs when a **parent cell** divides to produce **two identical daughter cells**. During cell division, mitosis refers specifically to the separation of the duplicated genetic material carried in the nucleus. Both **haploid** and **diploid** cells can undergo **mitosis**.



Meiosis

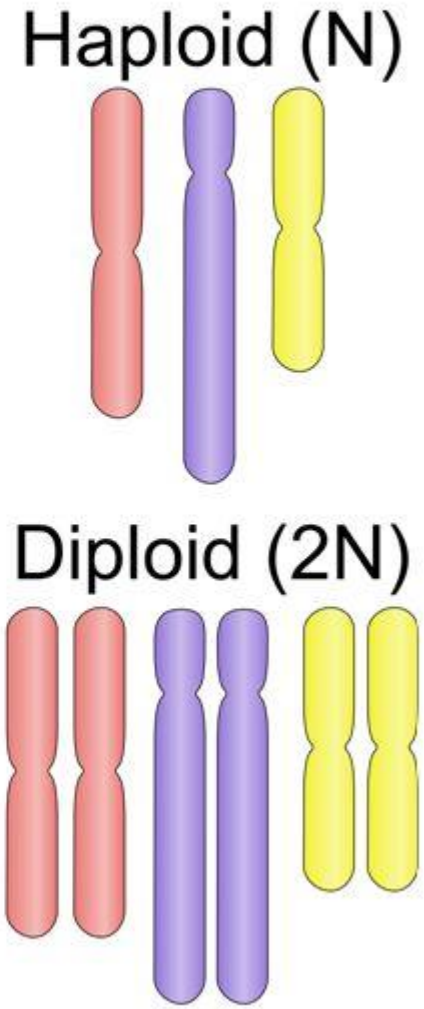
Meiosis is a type of cell division that **reduces** the number of **chromosomes** in the parent cell by **half** and produces **four gamete** cells. This process is required to produce egg and sperm cells for reproduction. During reproduction, when the sperm and egg unite to form a single cell, the number of chromosomes is restored in the offspring.

Meiosis begins with a parent cell that is **diploid**, meaning it has two copies of each chromosome. The process results in **four** daughter cells that are **haploid**, which means they contain half the number of chromosomes of the diploid parent cell.



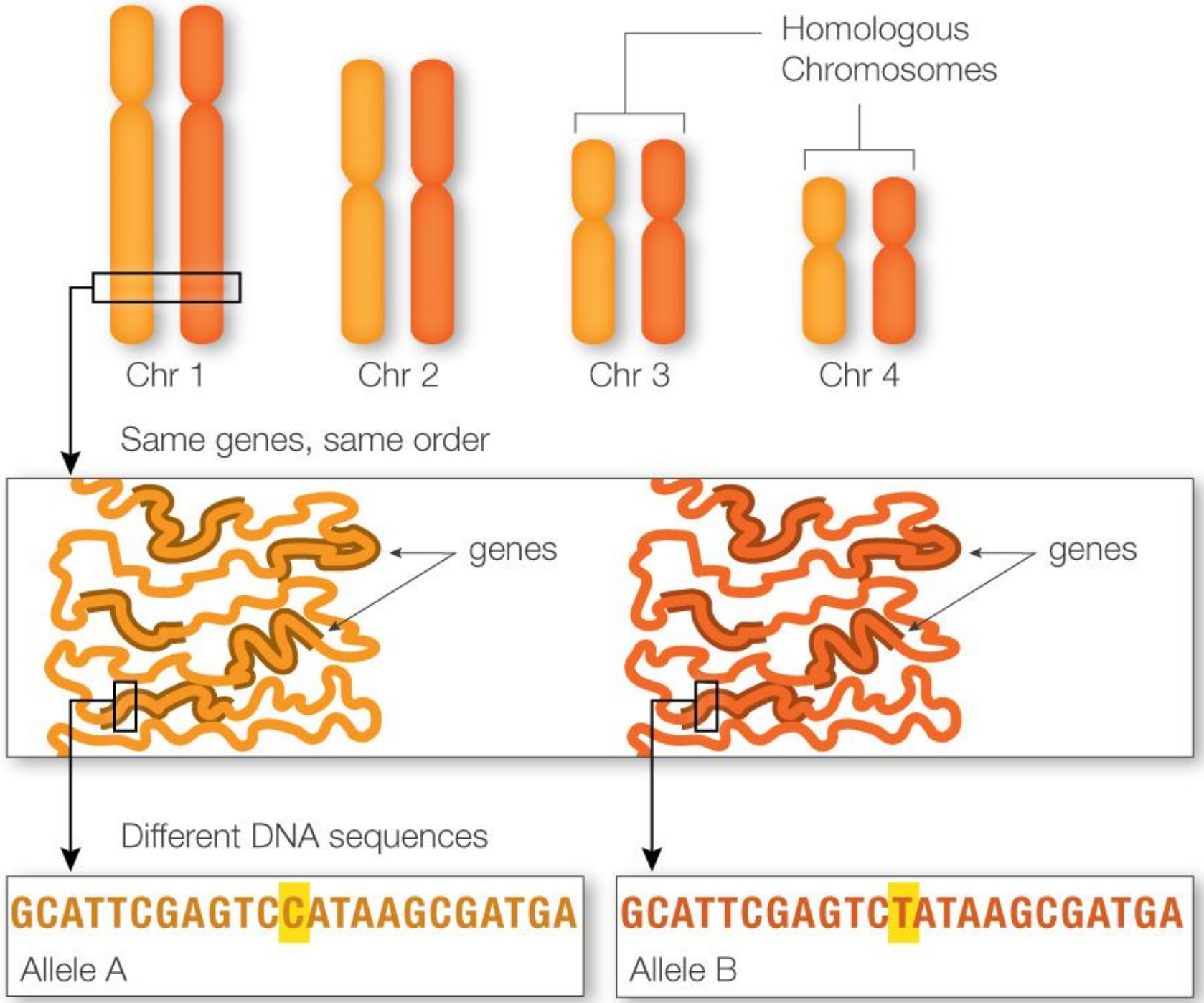
Haploid vs Diploid Cells

- A **diploid cell** has **two** of each **chromosome**, one from each parent. The body (AKA somatic) cells of most organisms are diploid. **Diploid** human cells contain **46 chromosomes** (22 **autosomal pairs** plus XX or XY) with a total of **6 billion base pairs** of DNA.
- A **haploid cell**, which only has **one** copy of every **chromosome**. Examples of haploid cells are eggs and sperm (also called gametes), Pollen (the reproductive cells of male plants), Spores (the reproductive cells of fungi, algae, and plants). The **haploid** human genome size is **3 billion base pairs**.
- If a haploid cell has **n** chromosomes, a diploid cell has **2n** (n represents a number, which is different for every species – in humans, for example, **n = 23** and **2n = 46**).



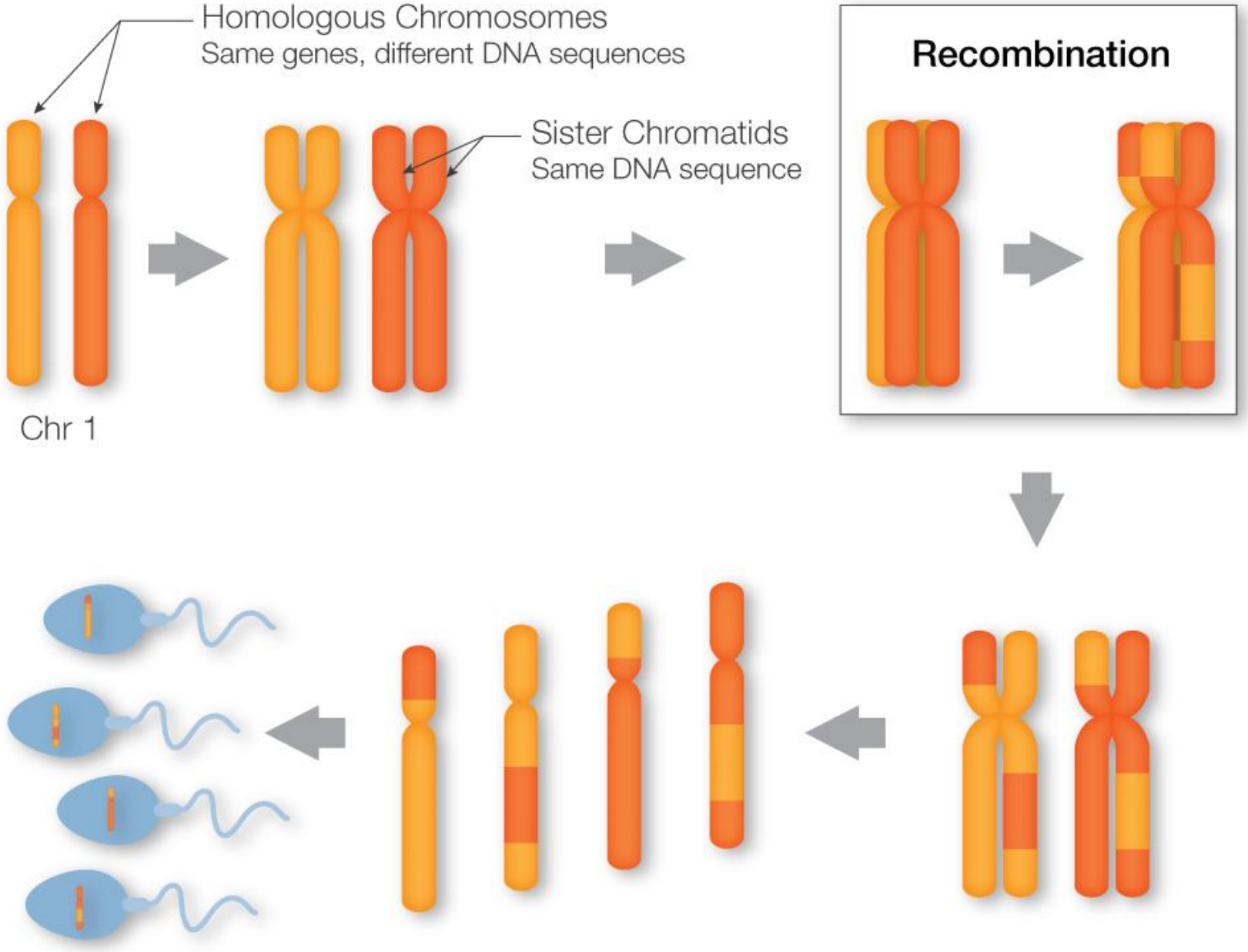
Homologous Chromosomes

Homologous chromosomes are **paired** chromosomes with two pieces of DNA within a **diploid** organism which carry the **same genes** in the **same order**, one from each parental source. But there may be variation between them, resulting in different **alleles**.



Recombination or Crossover

Homologous chromosome **recombination** occurs in meiosis and plays an important role in genetic **diversity**.



Genetics

- **Gene -** a unit of **heredity**; a section of DNA sequence **encoding** a **single protein**; A gene is the basic physical and functional **unit** of **heredity**. Genes, which are made up of DNA, act as instructions to make molecules called proteins. In humans, genes vary in size from a few **hundred** DNA bases to more than **2 million** bases. Research has estimated that humans have as few as **45,000** or as many as **140,000** distinct genes. Every person has **two** copies of each gene, one **inherited** from each parent.
- **Genome -** A genome is an organism's **complete set** of DNA, including all of its **genes**. Each genome contains all of the information needed to build and maintain that organism. In humans, a copy of the entire genome more than **3 billion** DNA base pairs is contained in all cells that have a nucleus.
- **Genetics** is the study of **genes** or **heredity**, the process in which a parent passes certain **genes** onto their children.

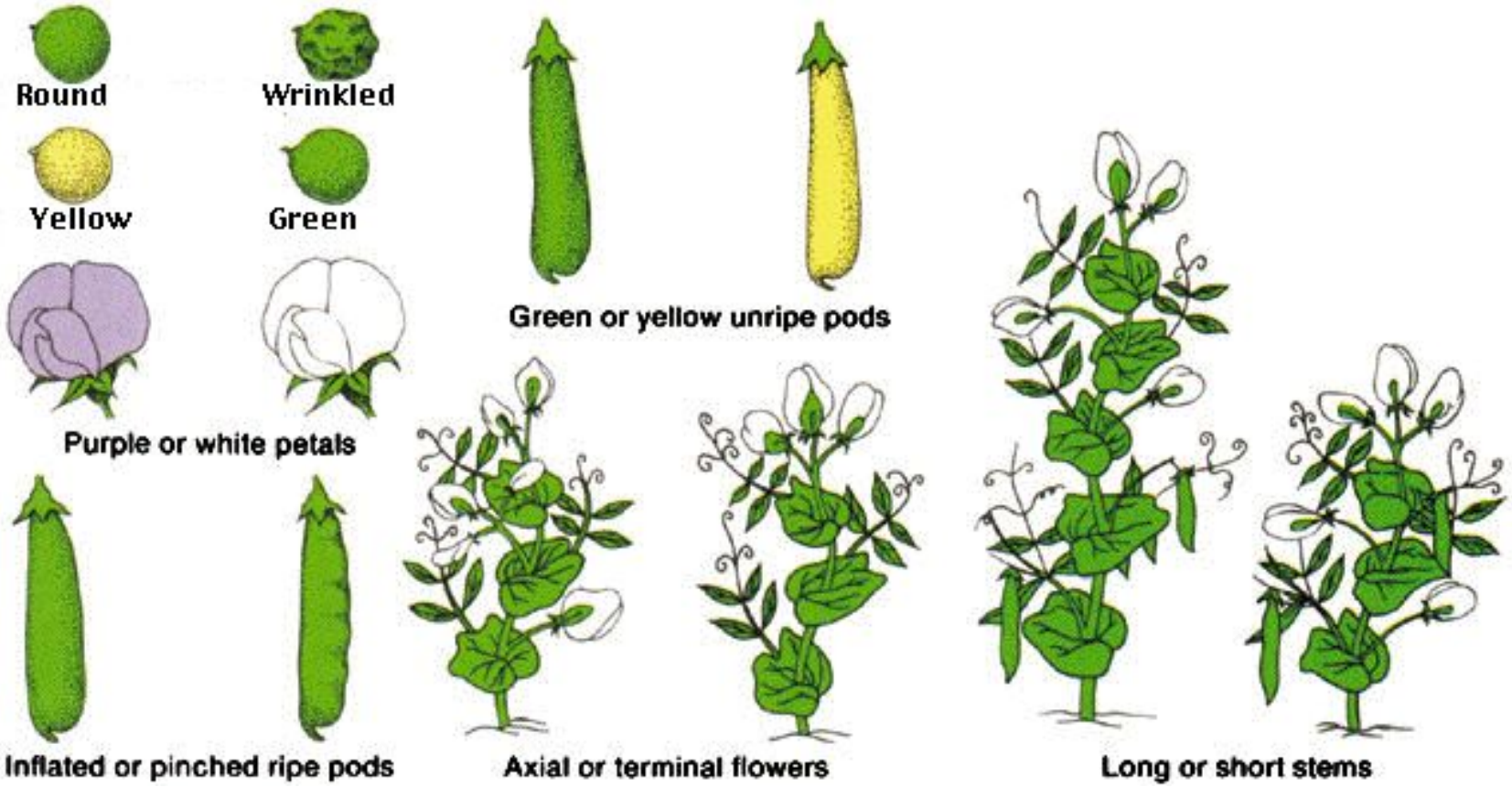
Gregor Johann Mendel


Gregor Mendel

- Is considered “**The Father of Genetics**“, Austrian Monk, born in what is now Czech Republic in 1822.
- Son of **peasant farmer**, studied Theology and was ordained priest Order St. Augustine.
- Went to the university of Vienna, where he studied botany and learned the Scientific Method. Worked with pure lines of **peas** for **eight years**
- Prior to Mendel, heredity was regarded as a "**blending**" process and the offspring were essentially a "**dilution**“ of the different parental characteristics.
- Work was largely ignored for **34 years**, until 1900, when 3 independent botanists rediscovered Mendel’s work.
- Mendel was the first biologist to use Mathematics – to explain his results quantitatively.
- Mendel predicted
  - The **concept of genes**
  - That **genes occur in pairs**
  - That **one gene** of each pair is present in the **gametes**

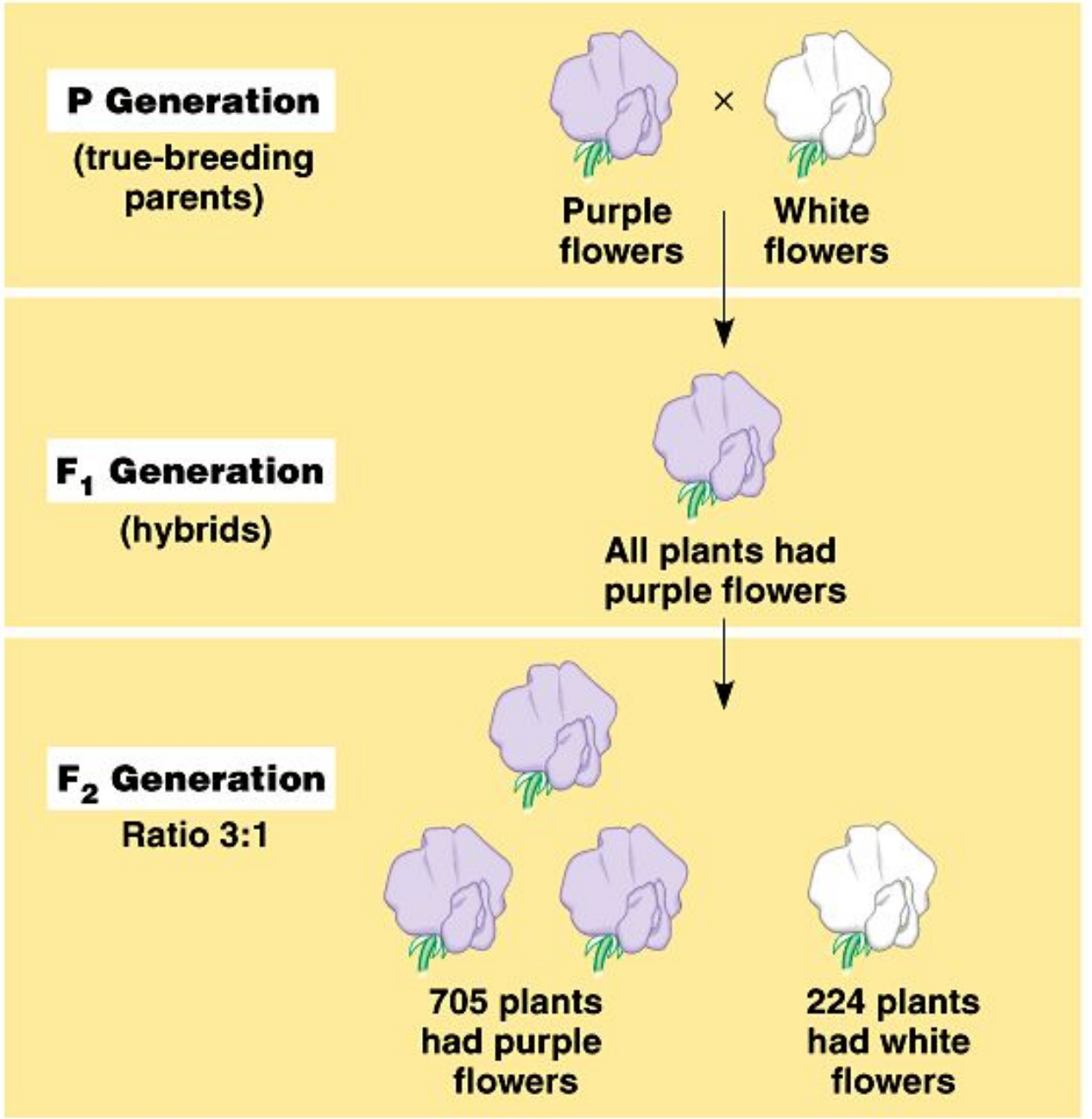
Mendel's peas

- **Mendel looked at seven traits or characteristics of pea plants:**



Mendel's Laws

- The first law, ***The Law of Segregation***, states that during fertilization each parent passes on one allele for each trait. Which allele the offspring would get from the parents is **random**.
- The second law, ***The Law of Independent Assortment***, states that transmission of one trait does not affect the transmission of other traits.
- The third law, ***The Law of Dominance***, states that one type of **allele** (the **dominant**) could mask the other (the **recessive**).



Genetic Terms

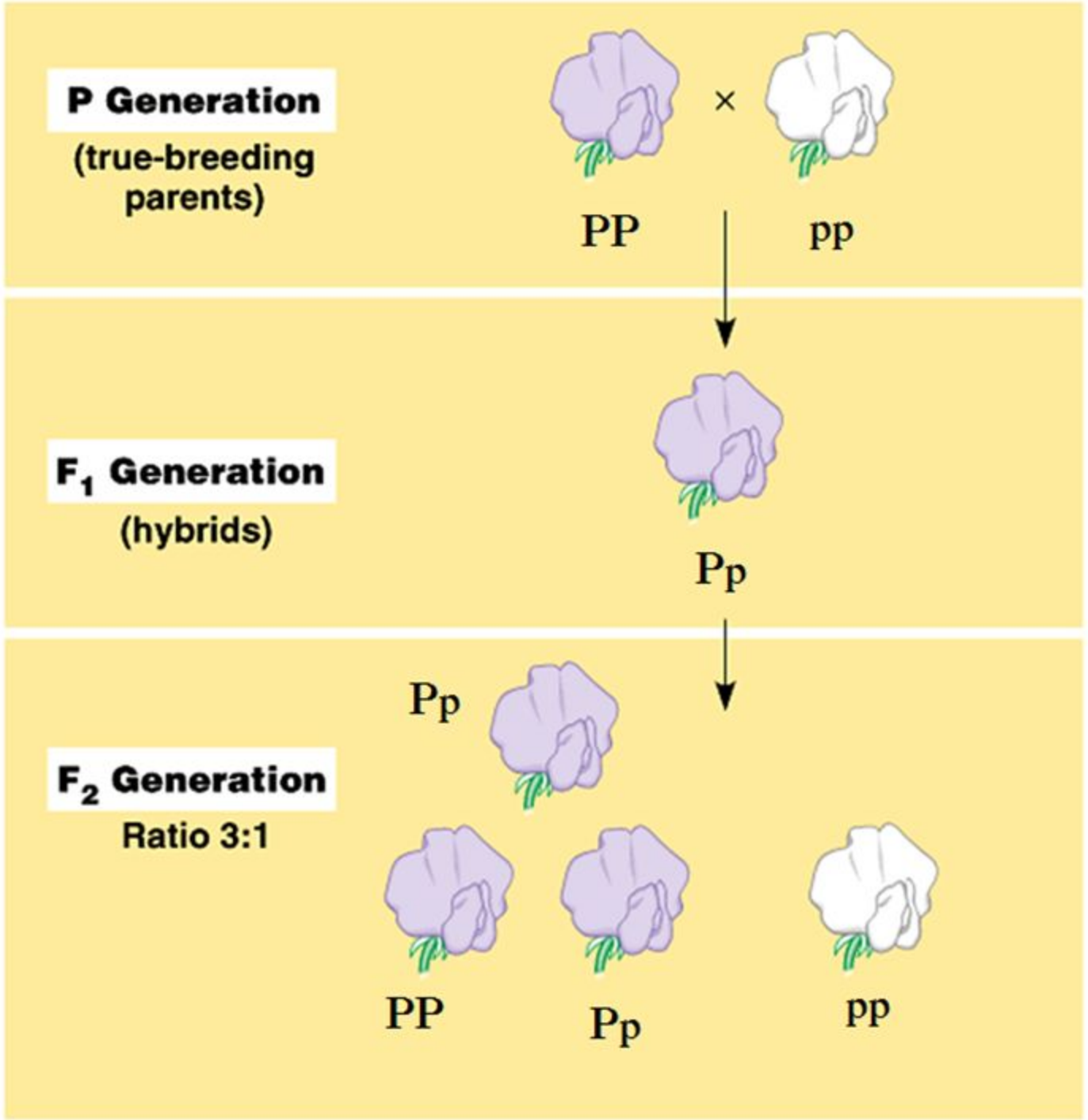
- **Homologous Chromosome -** The two chromosomes from a particular pair, normally one inherited from the mother and one from the father, containing the same genetic loci in the same order
- **Allele -** two genes that occupy the same position on homologous chromosomes and that cover the same trait (like ‘flavors’ of a trait).
- **Homozygous** – having identical genes (one from each parent) for a particular characteristic.
- **Heterozygous** – having two different genes for a particular characteristic.
- **Dominant** – the allele of a gene that masks or suppresses the expression of an alternate allele; the trait appears in the heterozygous condition.
- **Recessive** – an allele that is masked by a dominant allele; does not appear in the heterozygous condition, only in homozygous.

More Genetic Terms

- **Genotype** – the genetic makeup of an organisms
- **Phenotype** – the physical appearance of an organism (Genotype + environment)
- **Monohybrid cross**: a genetic cross involving a single pair of genes (one trait); parents differ by a single trait.
- **Dihybrid cross**: a genetic cross involving two pair of genes (two trait); parents differ by a two trait.
- **P** = Parental generation
- **F_1** = First filial generation; offspring from a genetic cross.
- **F_2** = Second filial generation of a genetic cross

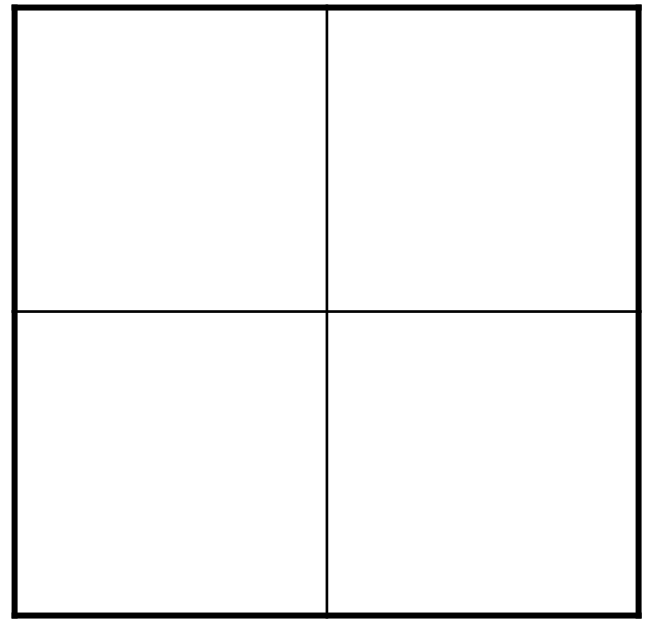
Genetic Terms

- P = Allele for **Purple** (dominant)
- P = Allele for **White** (recessive)
- PP = Homozygous Purple Flower Pea Plant
- pp = Homozygous White Flower Pea Plant
- Pp = Heterozygous Purple Flower Pea Plant



Punnett Square

- A useful tool to do genetic crosses
- For a monohybrid cross, you need a square divided by four….
- Looks like a window pane…
- We use the Punnett square to **predict** the **genotypes** and **phenotypes** of the **offspring**.



Punnett Square for Parent Generation (P)

		Homozygous Purple (PP)	
		P	P
Homozygous White (pp)	p	Pp Heterozygous Purple	Pp Heterozygous Purple
	p	Pp Heterozygous Purple	Pp Heterozygous Purple

Genotypes	
Pp	100%

Phenotypes	
Purple	100%

Punnett Square for Filio Generation (F1)

		Heterozygous Purple (Pp)	
		P	p
Heterozygous Purple (Pp)	P	PP Homozygous Purple	Pp Heterozygous Purple
	p	Pp Heterozygous Purple	pp Homozygous White

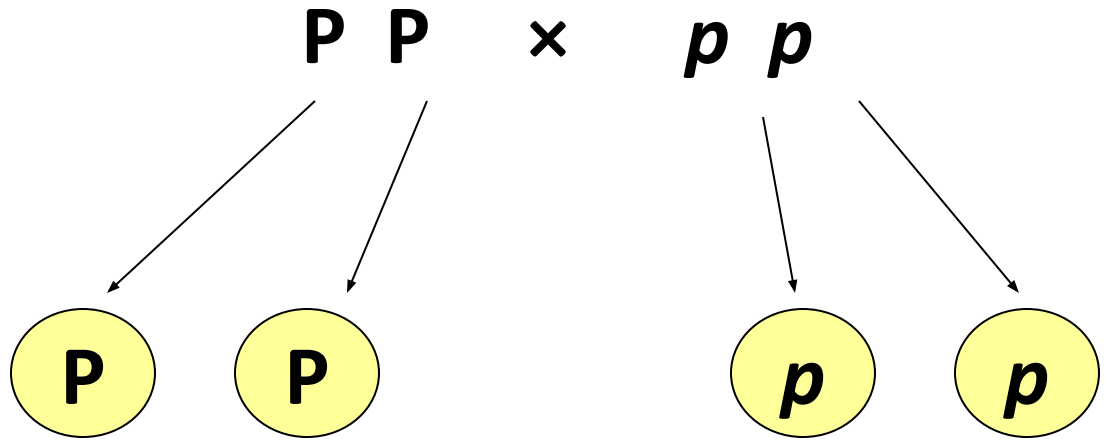
Genotypes	
PP	1/4
Pp	1/2
pp	1/4

Phenotypes	
Purple	3/4
White	1/4

Secret of the Punnett Square

- Key to the Punnett Square:
- Determine the gametes of each parent…
- How? By “splitting” the genotypes of each parent:

If this is your cross



The gametes are:

Punnett Square for Parent Generation (P) Shortcut

		Homozygous Purple (PP)
		P
Homozygous White (pp)	p	Pp Heterozygous Purple

Genotypes	
Pp	100%

Phenotypes	
Purple	100%

Dihybrid cross: flower color and stem length

		Heterozygous Tall,Purple (TtPp)			
		TP	Tp	tP	tp
Heterozygous Tall,Purple (TtPp)	TP	TTPP Tall,Purple	TTPp Tall,Purple	TtPP Tall,Purple	TtPp Tall,Purple
	Tp	TTPp Tall,Purple	TTpp Tall,White	TtPp Tall,Purple	Ttpp Tall,White
	tP	TtPP Tall,Purple	TtPp Tall,Purple	ttPP Short,Purple	ttPp Short,Purple
	tp	TtPp Tall,Purple	Ttpp Tall,White	ttPp Short,Purple	ttpp Short,White

Genotypes	
TTPP	1/16
TTPp	2/16
TtPP	2/16
TtPp	4/16
TTpp	1/16
Ttpp	2/16
ttPP	1/16
ttPp	2/16
ttpp	1/16

Phenotypes	
Tall,Purple	9/16
Tall,White	3/16
Short,Purple	3/16
Short,White	1/16

Sample Problems

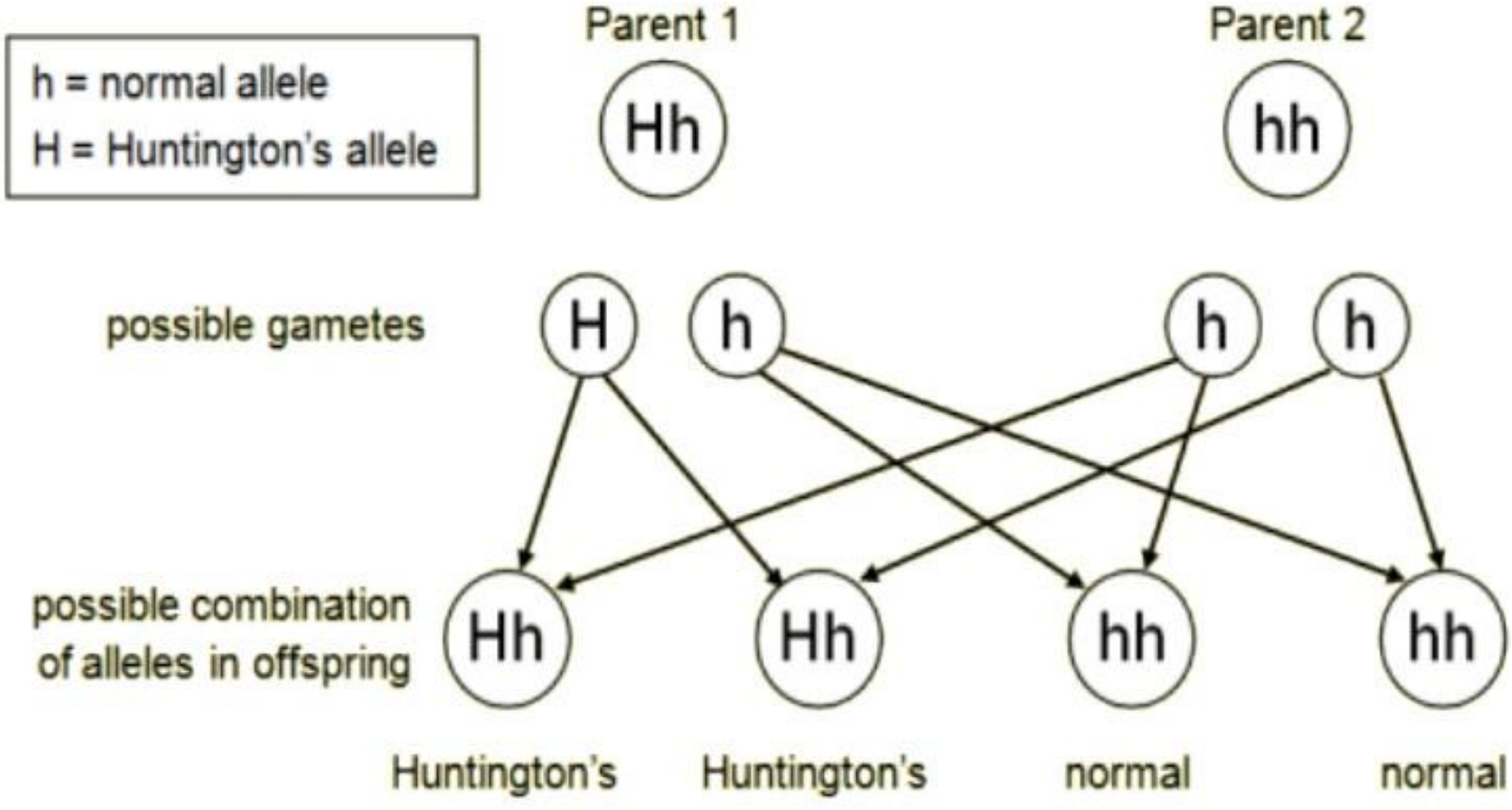
Round (R) is dominant to wrinkled (r). Yellow (Y) is dominant to green (y)

- Classify the following as **heterozygous** or **homozygous**: RR, Rr, yy, YyRR?
- What is the **phenotype** of the following: Yy, Rr, yy, YyRr?
- What is the probability of Rr x Rr producing **wrinkled** seeds?
- What is the probability of Yy x yy producing **green** seeds?
- What is the probability that RrYy x RRYy would produce RrYy?

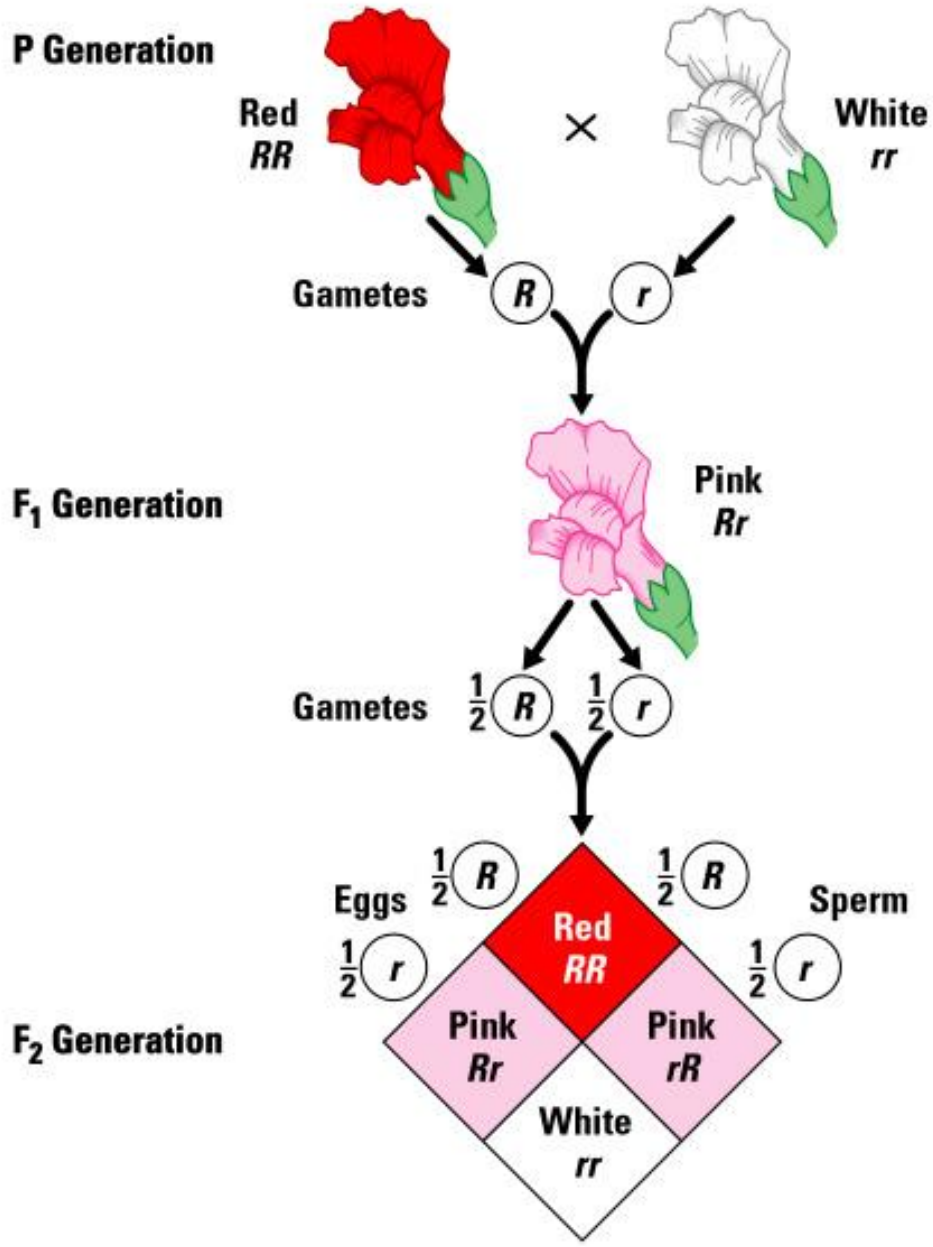
Huntington's Disease

Huntington's disease

Huntington's disease is an inherited disorder that affects the nervous system. It is caused by a **dominant** allele. This means that it can be passed on by just one parent if they have the disorder. The genetic diagram shows how this can happen.



Snapdragons (flower)

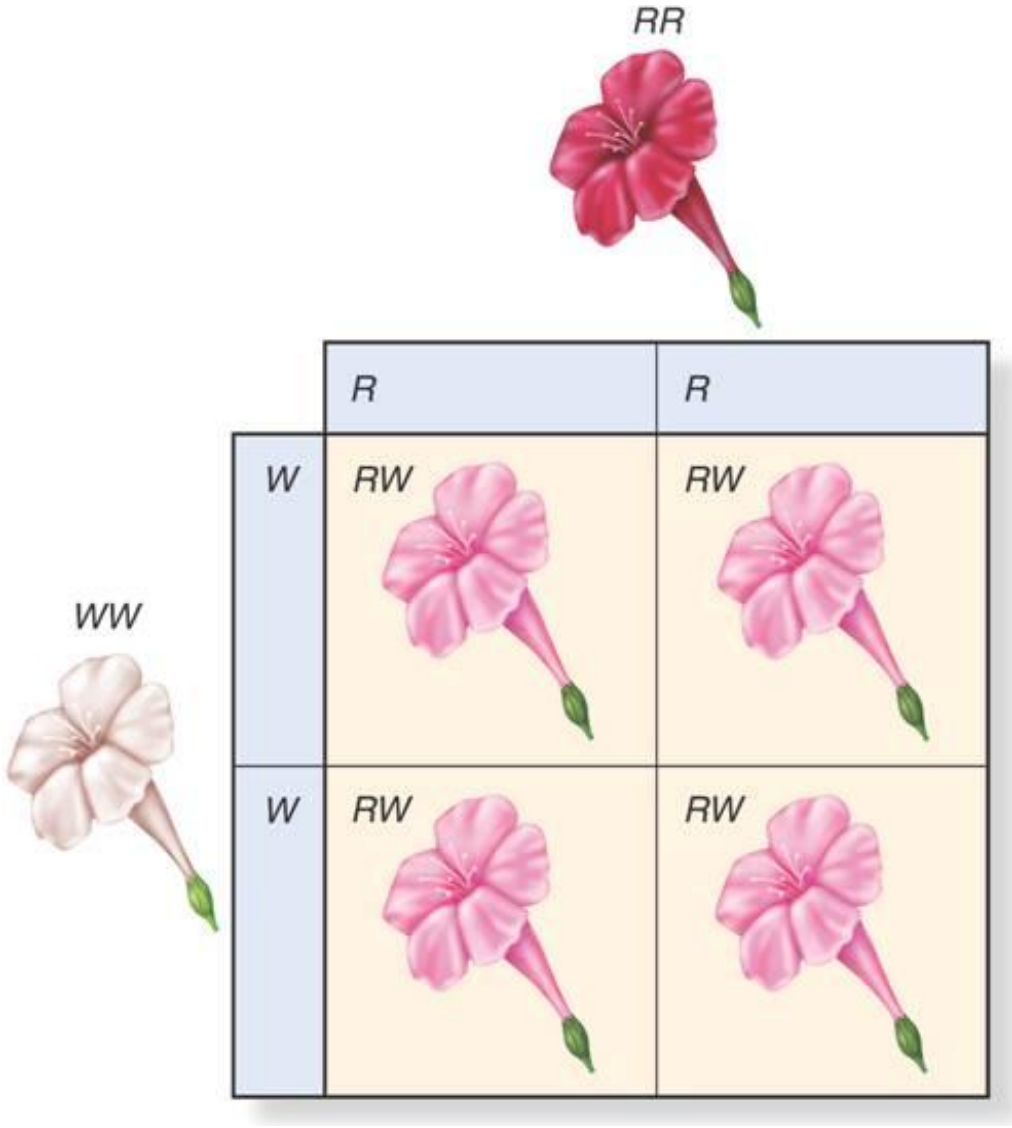


Beyond Dominant and Recessive Alleles

- Not all traits are controlled by **single genes** with simple **dominant - recessive** alleles
- Some alleles are **neither dominant nor recessive**, and many traits are controlled by **multiple alleles or multiple genes**.

Incomplete Dominance

- When one allele is not completely dominant over another.
- Heterozygous phenotype = **BLEND**
- Since one allele is not completely dominant over the other allele, we use **two different capital letters** to represent the alleles



Example: A cross between **red (RR)** and **white (WW)** snapdragons produces **pink** flowers (**RW**).

Codominance

- **both** alleles **contribute** to the phenotype.
- Heterozygous phenotype will have **both** phenotypes **visible**
- Since one allele is not completely dominant over the other allele, and both traits appear together, we use **two different capital letters** to represent the alleles

Example: **Shorthorn Cattle**
The offspring of Homozygous **red (RR)** and Homozygous **white (WW)** will have both **red and white** hairs (**RW**)
The offspring are heterozygous and called "roan"

Multiple Alleles

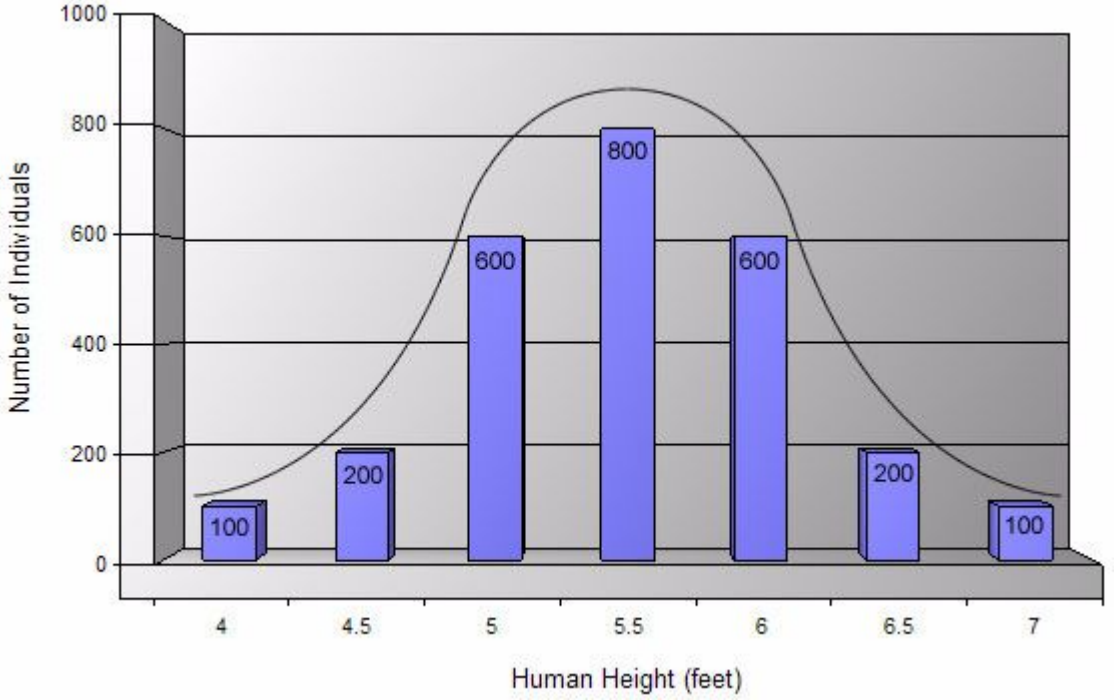
- Multiple alleles -genes that are controlled by more than two alleles.
- **An individual can't have more than two alleles** (Remember you get one copy from each parent). However, more than two possible alleles can **exist** in a **population**.

Phenotype (Blood Type)	Genotype
Type A	I^AI^A or I^Ai
Type B	I^BI^B or I^Bi
Type AB	I^AI^B
Type O	ii

Example: A human's blood type is determined by a single gene that has three different alleles.

Polygenic Traits

- Polygenic traits -traits controlled by **two or more genes**.
- **Skin color** in humans is a polygenic trait controlled by more than **four** different **genes**.
- Human **height** and **eye color** is also polygenic trait



Polygenic traits tend to result in a distribution that resembles a **bell-shaped curve**, with few at the extremes and most in the middle.

Height Gene in Punnett Square

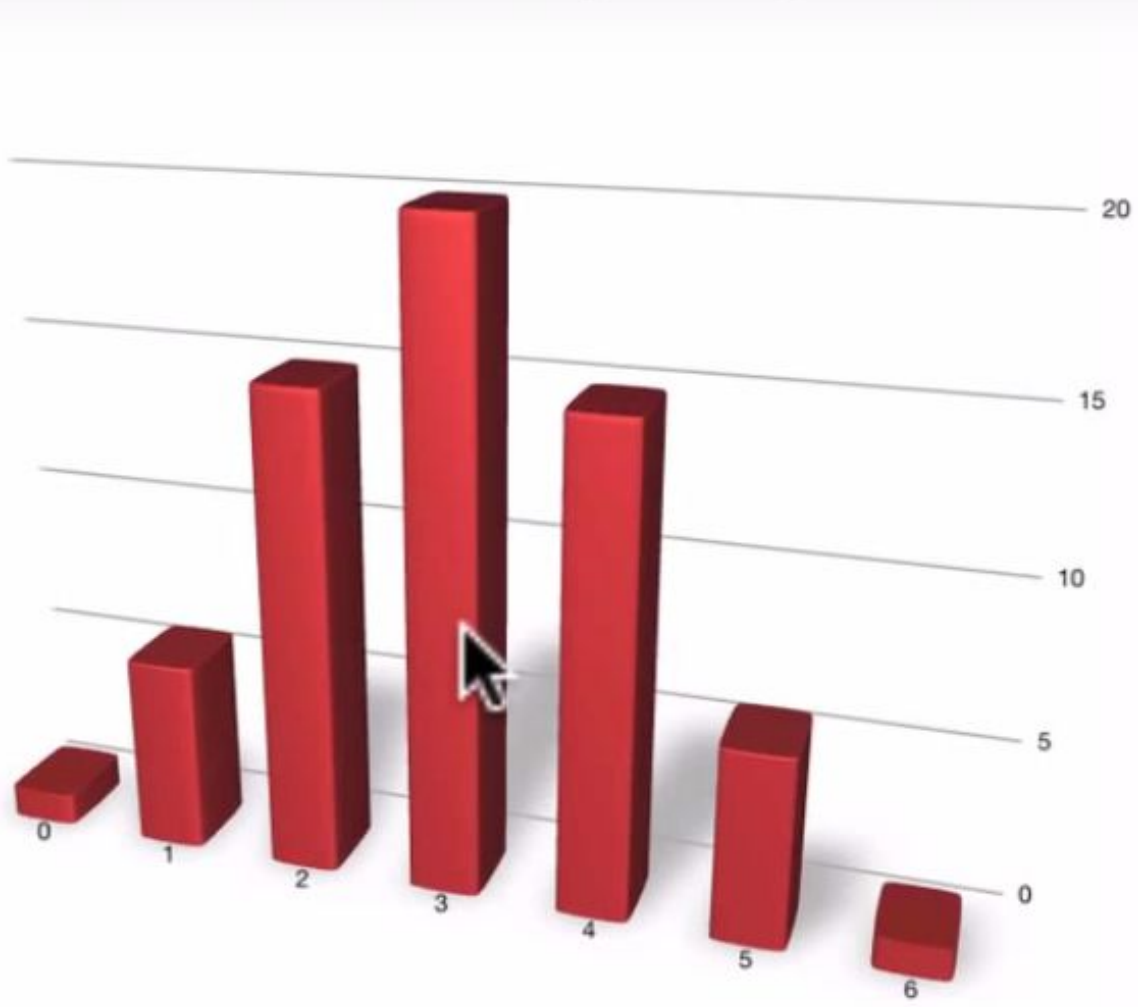
AaBbCc
x
AaBbCc

	ABC	ABc	AbC	Abc	aBC	aBc	abC	abc
ABC	6	5	5	4	5	4	4	3
ABc	5	4	4	3	4	3	3	2
AbC	5	4	4	3	4	3	3	2
Abc	4	3	3	2	3	2	2	1
aBC	5	4	4	3	4	3	3	2
aBc	4	3	3	2	3	2	2	1
abC	4	3	3	2	3	2	2	1
abc	3	2	2	1	2	1	1	0

Height	Total
0	1
1	6
2	15
3	20
4	15
5	6
6	1

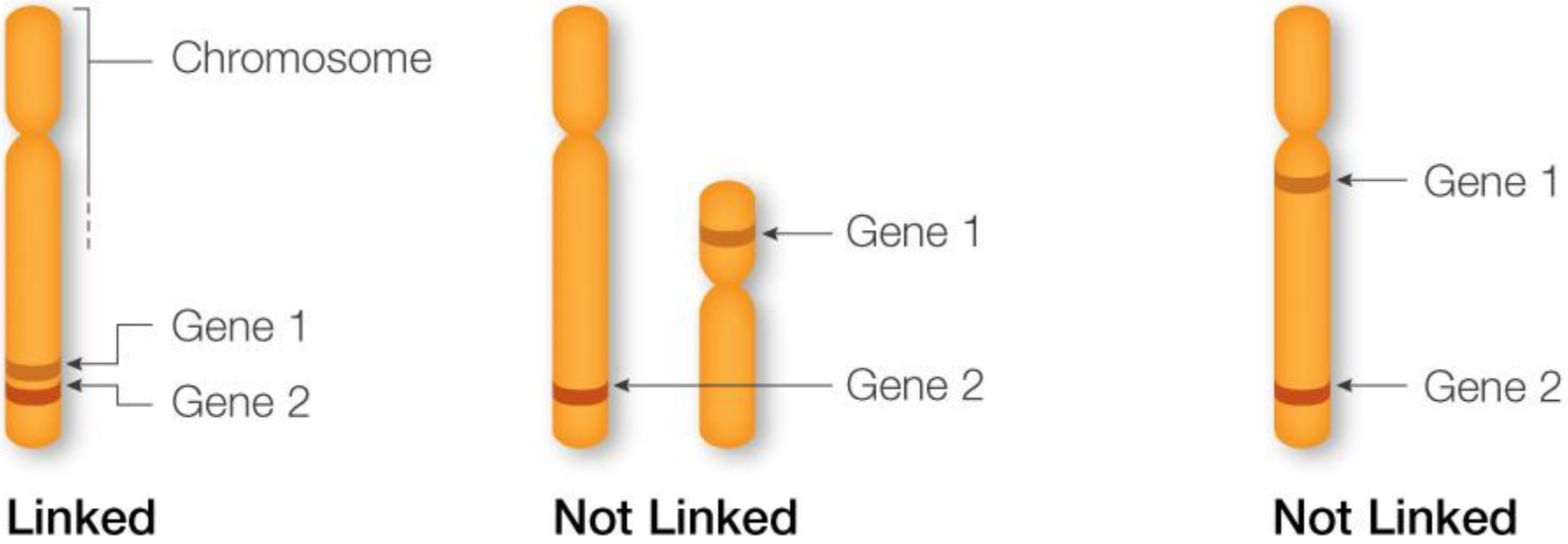
Height Gene as Bell Curve

AaBbCc
x
AaBbCc



Height	Total
0	1
1	6
2	15
3	20
4	15
5	6
6	1

Genetic Linkage

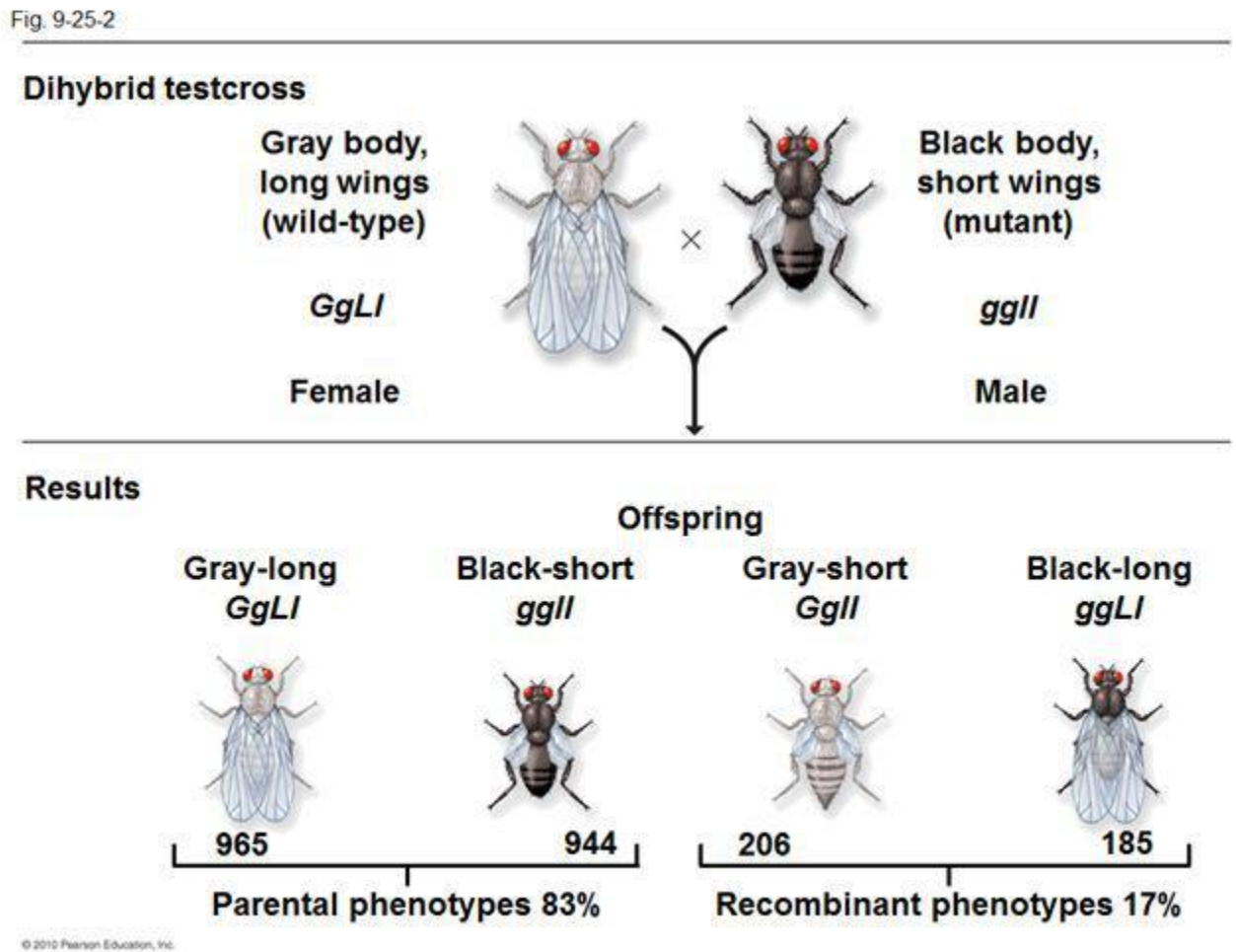


- **Linked genes** sit close together on a chromosome, making them likely to be inherited together (left).
- Genes on **separate chromosomes** are **never linked** (center).
- But not all genes on a chromosome are linked. Genes that are farther away from each other are more likely to be separated during a process called homologous recombination (right).
- For More Info: http://learn.genetics.utah.edu/content/pigeons/geneticlinkage/

Linked Genes

Example

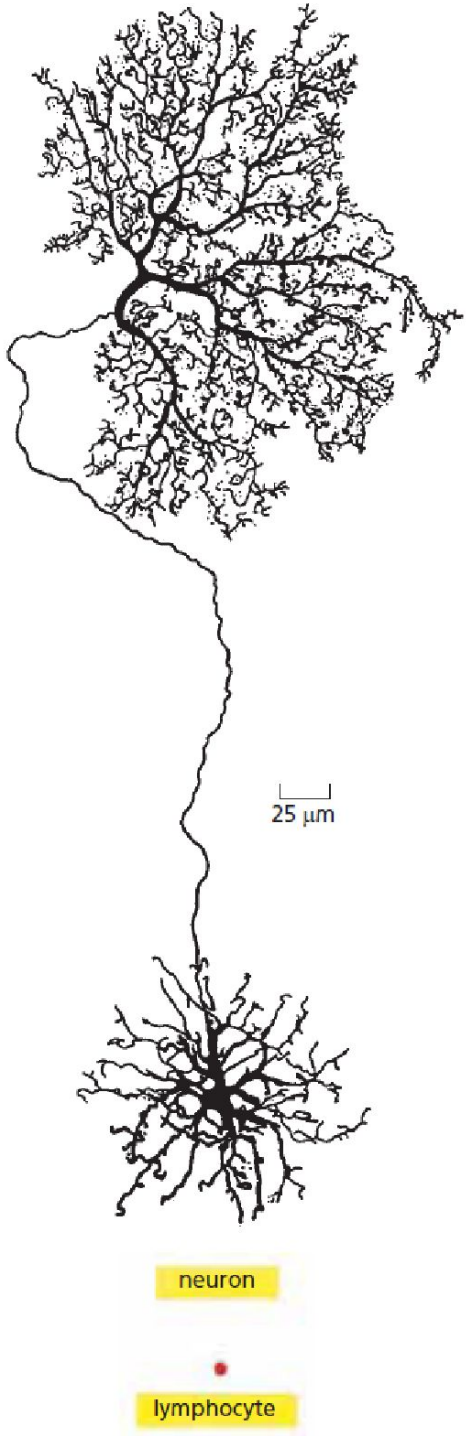
- **Thomas Hunt Morgan:** In 1916, published a paper on genes in *Drosophila melanogaster* (fruit fly)
  - Found the **recombinant frequency** to be 17%, not 50%



The Different Cell Types of a Multicellular Organism Contain the Same DNA?

A neuron and a lymphocyte

- A **neuron** and a **lymphocyte** share the **same genome**.
- The long branches of this **neuron** from the retina enable it to **receive electrical signals** from many cells and **carry** them to many neighboring cells.
- The **lymphocyte**, a **white blood cell** involved in the immune response to infection (drawn to scale), moves freely through the body.
- Both of these mammalian cells contain the **same genome**, but they **express different RNAs and proteins**.



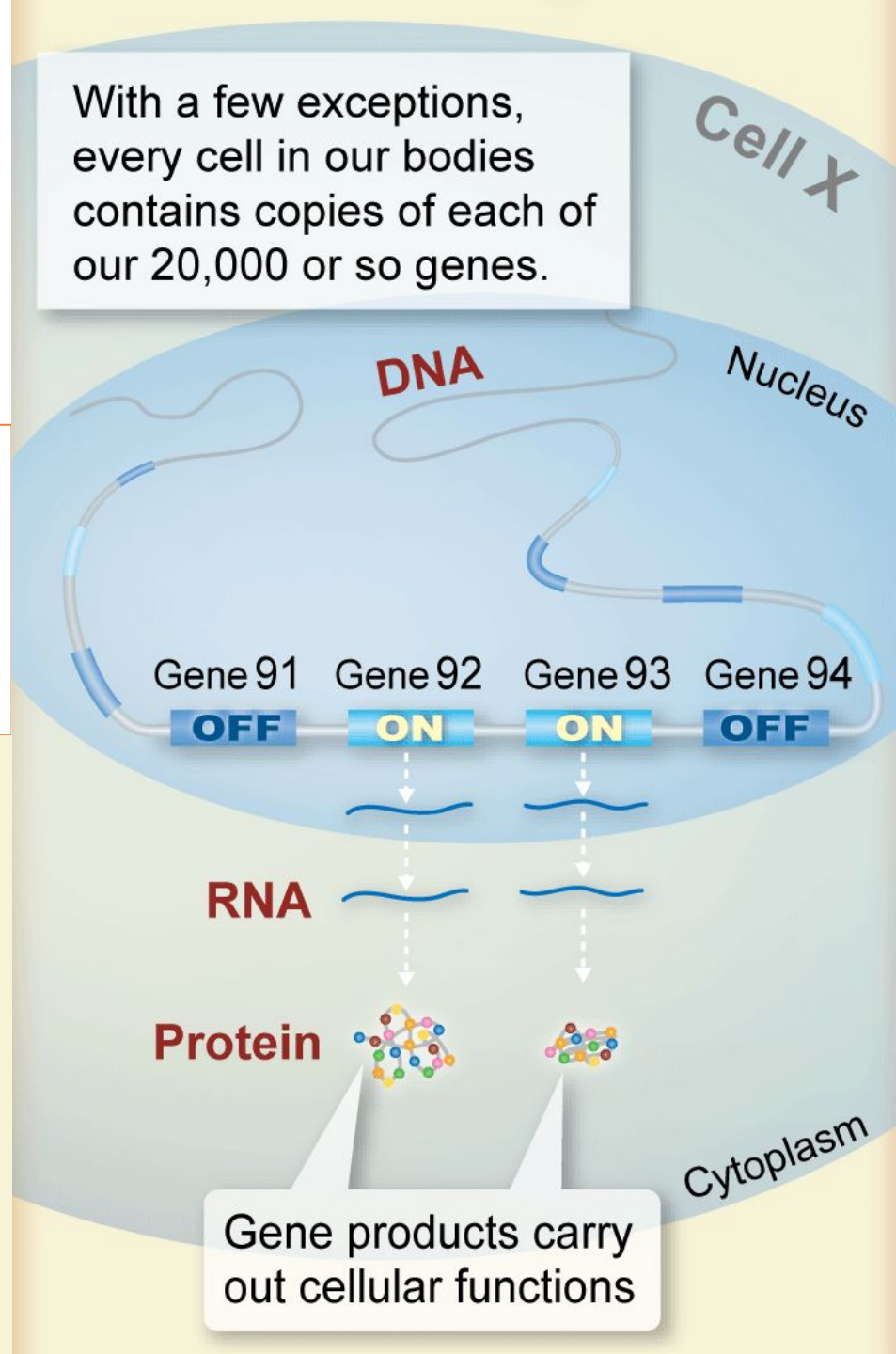
Gene Expression

- **Each cell** contains **all genes**!!! But **not all genes are expressed** in all cells (of a multicellular organism) or at all times
- **Gene expression** is the process by which the instructions in our **DNA** are **converted** into a functional product, such as a **protein or rna**.

- When a gene is "on" and its protein or RNA product is being made, scientists say that the gene is being expressed.
- The on and off states of all of a cell's genes is known as a **gene expression profile**.
- Each cell type has a unique gene expression profile.

CELL X's
GENE EXPRESSION PROFILE:

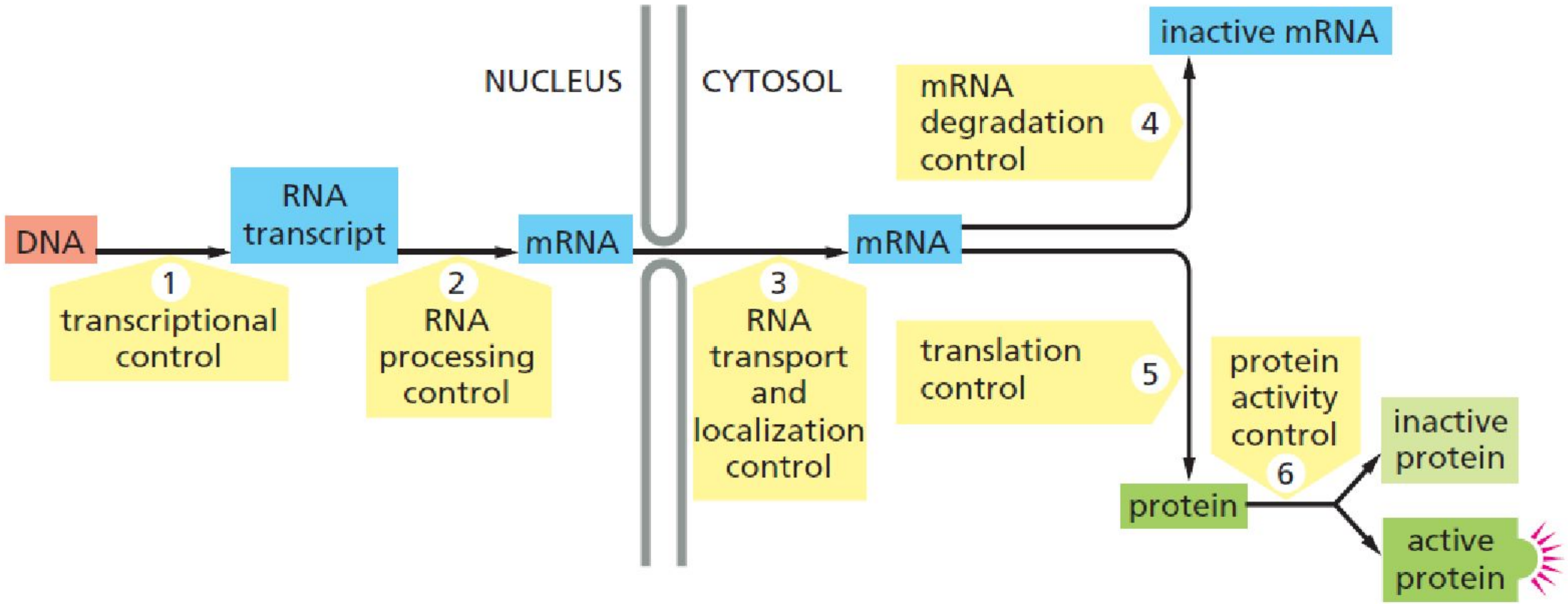
Gene 90	OFF
Gene 91	OFF
Gene 92	ON
Gene 93	ON
Gene 94	OFF
Gene 95	ON

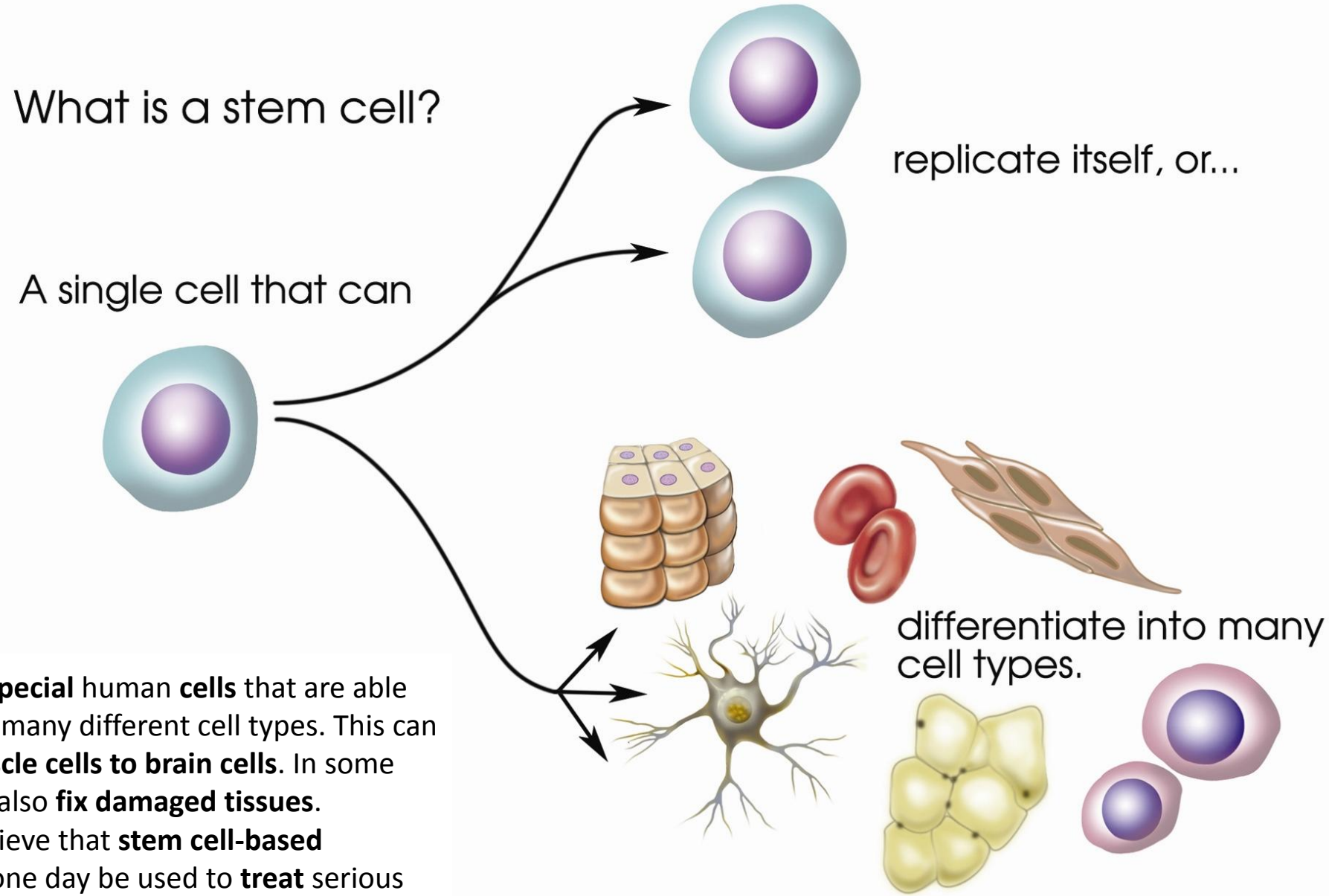


Regulation of Gene Expression

- **Gene Expression** Can Be **Regulated** at Many of the Steps in the Pathway from **DNA to RNA to Protein**
- Levels of control of gene expression
  - **Short term control** (to meet the daily needs of the organism)
  - **Long term control** (gene regulation in development/differentiation)

Eukaryotic gene expression can be controlled at several different steps.

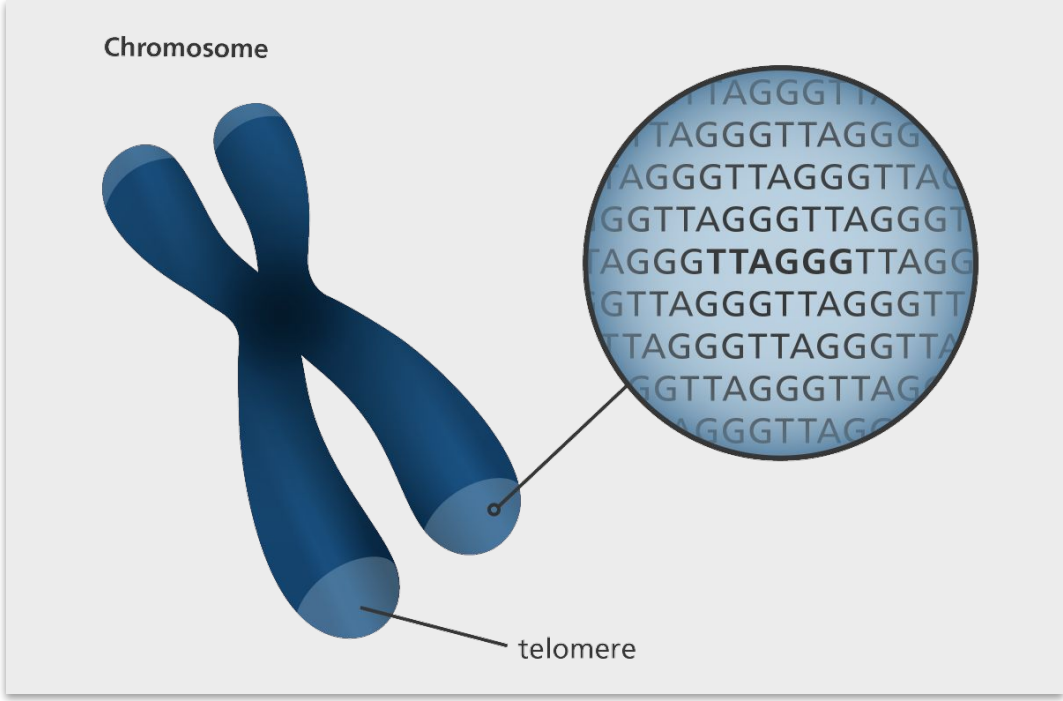




Stem cells are **special** human **cells** that are able to **develop** into many different cell types. This can range from **muscle cells to brain cells**. In some cases, they can also **fix damaged tissues**. Researchers believe that **stem cell-based therapies** may one day be used to **treat** serious **illnesses** such as paralysis and Alzheimer disease.

Telomere

- **Telomeres** are distinctive **structures** found at the **ends** of our **chromosomes**. They consist of the same **short DNA sequence repeated over and over** again.



- Telomeres are sections of DNA found at the ends of each of our chromosomes
- They consist of the **same sequence** of bases **repeated** over and over.
- In **humans** the telomere sequence is **TTAGGG**.
- This sequence is usually **repeated** about **3,000 times** and can reach up to **15,000 base pairs in length**.

What do telomeres do?

- **Telomeres** serve **three** major **purposes**:

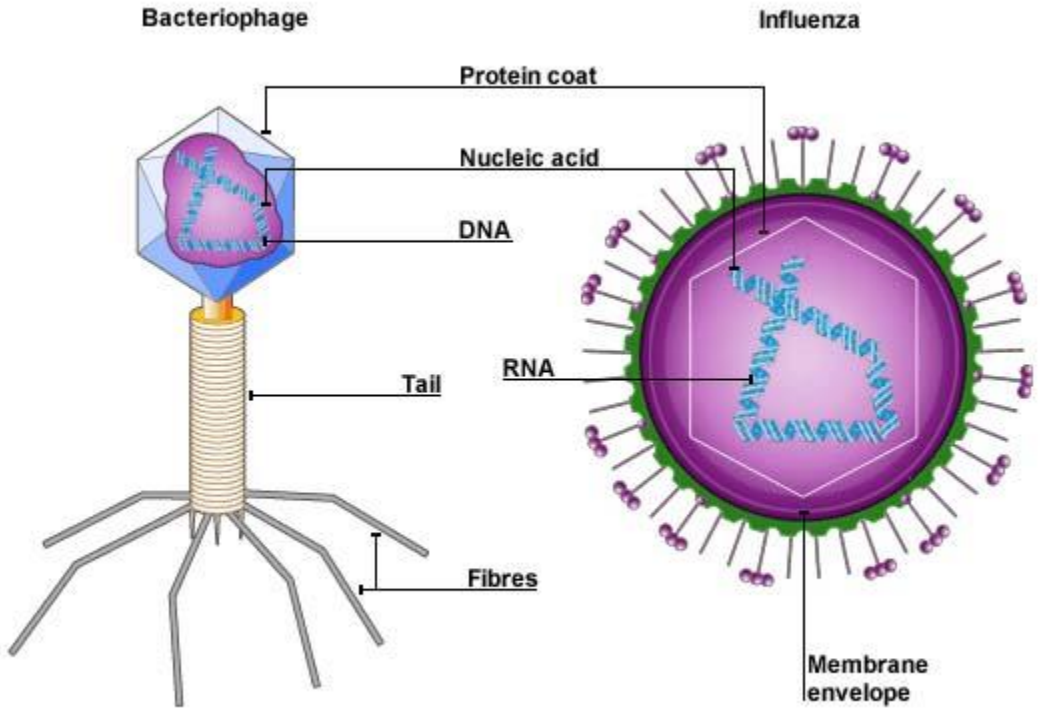
1. They help to **organise** each of our 46 **chromosomes** in the nucleus (control centre) of our cells
2. They **protect** the **ends** of our **chromosomes** by **forming a cap**, much like the **plastic tip on shoelaces**. If the telomeres were not there, our chromosomes may end up sticking to other chromosomes.
3. They allow the chromosome to be **replicated properly** during cell division:
    - Every time a cell carries out DNA replication the **chromosomes** are **shortened** by about **25-200 bases** (A, C, G, or T) **per replication**.
    - However, because the ends are protected by telomeres, the only part of the chromosome that is lost, is the telomere, and the DNA is left undamaged.
    - Without telomeres, important DNA would be lost every time a cell divides (usually about 50 to 70 times). This would eventually lead to the loss of entire genes
    - When the **telomere** becomes **too short**, the **chromosome** reaches a ‘critical length’ and can **no longer be replicated**.

How is telomere length maintained?

- **Telomerase** is an **enzyme** that **adds** the **telomere sequence (TTAGGG)** to the **ends** of **chromosomes**.
- **Telomerase** is only found in **very low** concentrations in our **somatic** cells. Because **these cells** do not regularly use telomerase they **age** leading to a reduction in normal function.
- **Telomerase** is found in **high levels** in **germline cells** (egg and sperm) and **stem cells**. In **these cells** telomere length is maintained after DNA replication and the cells **do not** show signs of **ageing**.
- **Telomerase** is also found in **high** levels in **cancer cells**. This enables **cancer cells** to be **immortal** and **continue replicating** themselves. If telomerase activity was switched off in cancer cells, their telomeres would shorten until they reached a ‘critical length’. This would, prevent the cancer cells from dividing uncontrollably to form tumours.

Virus

A **virus** is a small **parasite** that **cannot reproduce by itself**. **Once** it **infects** a susceptible cell, however, a **virus** can direct the cell machinery to **produce** more **viruses**. Most viruses have either **RNA or DNA** as their genetic material. The nucleic acid may be **single or double-stranded**. The **simplest** viruses contain only enough RNA or DNA to encode **four proteins**. The most **complex** can encode **100 – 200 proteins**.

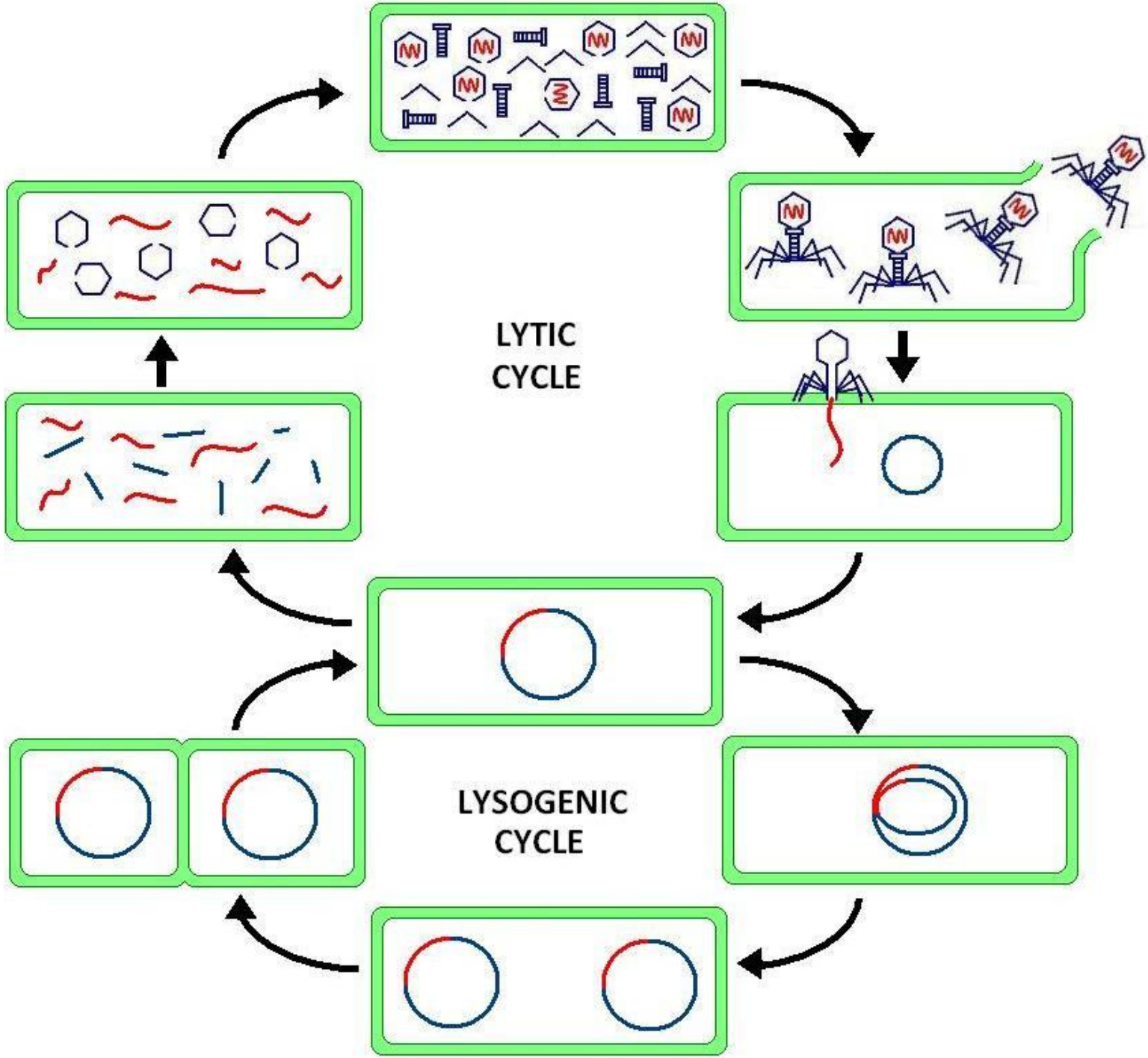


- **Viruses** are often considered **non-living** as they exist in an **inert state outside** of a **host** cell. They consist of a strand of nucleic acid, either **DNA or RNA**, surrounded by a **protective protein coat** (the capsid). Sometimes they have a further **membrane** of lipid, referred to as an envelope, surrounding the protein. They may also have a **tail** section.
- **Viruses** are **species specific**, but almost every species on Earth can be affected by some form of virus.

Lytic vs Lysogenic Cycle of Virus

- The **lytic cycle** involves the reproduction of viruses using a host cell to manufacture more viruses; the viruses then burst out of the cell.

- The **lysogenic cycle** involves the incorporation of the viral genome into the host cell genome, infecting it from within.



Immune System (T-Cells) and HIV Virus

- **T-cells** are a type of **white blood cell** that circulate around our bodies, scanning for cellular abnormalities and infections. Two types, killer T-cells to hunt down and destroy cells that are **infected** with germs or that have become cancerous. The other main type of T-cells are called helper T-cells. **Helper T-cells** orchestrate an **immune response** and play important roles in all arms of **immunity**.
- **HIV** stands for **human immunodeficiency virus**. HIV **infects** vital cells in the human immune system such as **helper T cells** and the body becomes progressively more susceptible to opportunistic infections. The body can **no longer defend** itself against **infection**.

Gene Therapy

- **Gene therapy** is a **medical** approach that **treats or prevents disease** by correcting the underlying **genetic problem**. Researchers are testing **several approaches to gene therapy**, including:
  - **Replacing** a **mutated gene** that causes disease with a **healthy** copy of the **gene**.
  - **Inactivating**, or “Turn off” a **mutated gene** that is functioning improperly.
  - **“Turn on”** a **gene** to help **fight** a disease.
  - Introducing a **new gene** into the body to help **fight** a **disease**.
  - **Remove** a piece of **DNA** that is impairing gene function and **causing disease**.
- This approach is **different** from **traditional** drug-based approaches, which may treat symptoms but not the underlying genetic problems. Most commonly, **gene therapy** uses a **vector**, typically a **virus**, to deliver a gene to the cells where it's needed. Once it's inside, the cell's gene-reading machinery uses the information in the gene to build RNA and protein molecules. The proteins (or RNA) can then carry out their job in the cells.
- Although **gene therapy** is a **promising** treatment option for a number of diseases (including inherited disorders, some types of cancer, and certain viral infections), the technique remains **risky** and is still under study to make sure that it will be safe and effective. Gene therapy is currently only being tested for the treatment of diseases that have no other cures.