

INTRODUCTION TO GENETIC EPIDEMIOLOGY

(1012GENEP1)

Prof. Dr. Dr. K. Van Steen

GENETICS REVISITED

1 Basics of molecular genetics

1.a From discrete “units” of information to DNA

The structure of cells, chromosomes, DNA and RNA

1.b How does DNA encoding work?

Reading the information, reading frames

The central dogma of molecular biology

1.c DNA mutations

Variation and sequencing

2 What can your spit tell you about your DNA?

2.a The use of saliva

2.b Genetic markers

3 What is the human epigenome?

3.a The human epigenome project

3.b Mapping the human epigenome














1 Basics of molecular genetics

1.a From discrete “units” of information to DNA

Mendel

- Many traits in plants and animals are heritable; genetics is the study of these heritable factors
- Initially it was believed that the mechanism of inheritance was a masking of parental characteristics
- Mendel developed the theory that the mechanism involves random transmission of discrete “units” of information, called genes. He asserted that,
 - when a parent passes one of two copies of a gene to offspring, these are transmitted with probability $1/2$, and different genes are inherited independently of one another (is this true?)

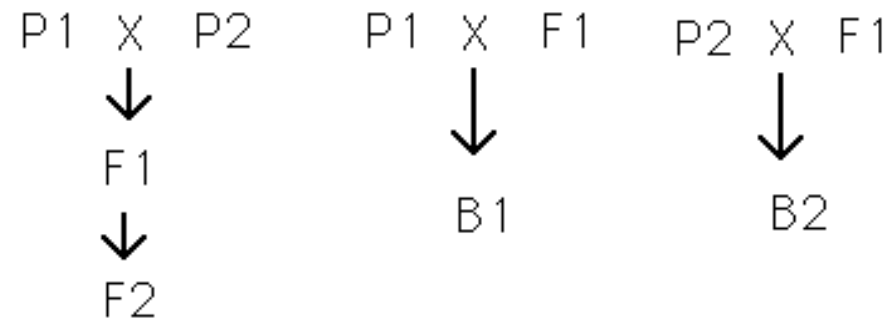
Mendel's pea traits

| Character | Dominant trait | Recessive trait | Character | Dominant trait | Recessive trait |
|--------------|--|---|-----------------|--|---|
| Seed shape |  Spherical |  Wrinkled | Flower position |  Axial |  Terminal |
| Seed color |  Yellow |  Green | | Stem height |  Tall |
| Flower color |  Purple |  White | | | |
| Pod shape |  Inflated |  Constricted | | | |
| Pod color |  Green |  Yellow | | | |

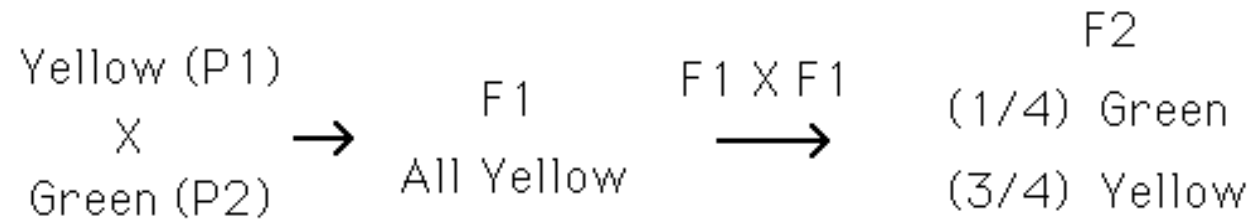
©1998 Sinauer Associates, Inc.

Some notations for line crosses

- Parental Generations (P_1 and P_2)
- First Filial Generation $F_1 = P_1 \times P_2$
- Second Filial Generation $F_2 = F_1 \times F_1$
- Backcross one, $B_1 = F_1 \times P_1$
- Backcross two, $B_2 = F_1 \times P_2$



What Mendel observed



- The F₁ were all Yellow
- Strong evidence for discrete units of heredity , as "green" unit obviously present in F₁, appears in F₂
- There is a 3:1 ratio of Yellow : Green in F₂

What Mendel observed (continued)

- Parental, F₁ and F₂ yellow peas behave quite differently

P1 Yellow
 X
Green (P2) → All Yellow

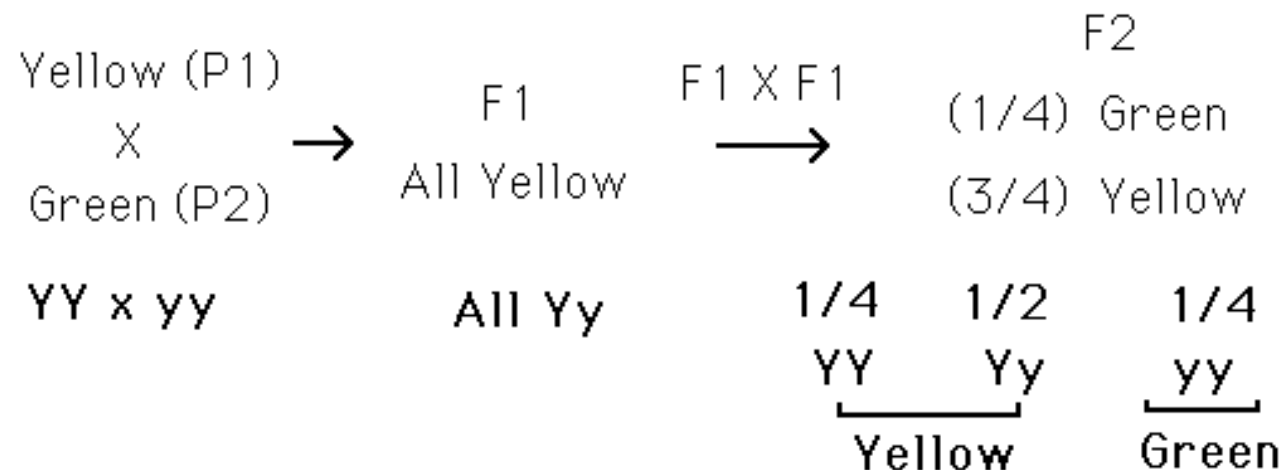
F1 Yellow
 X
Green (P2) → 1/2 Yellow
 1/2 Green

F2 Yellow
 X
Green (P2) → 2/3 Yellow
 1/3 Green

Mendel's conclusions

- **Mendel's first law** (law of segregation of characteristics)

This says that of a pair of characteristics (e.g. blue and brown eye colour) only one can be represented in a gamete. What he meant was that for any pair of characteristics there is only one gene in a gamete even though there are two genes in ordinary cells.

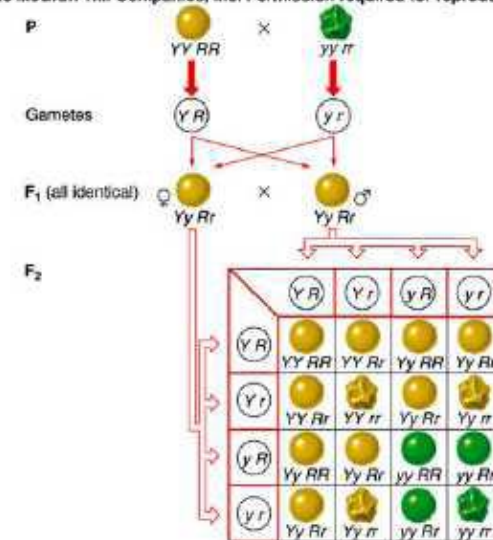


Mendel's conclusions (continued)

- **Mendel's second law** (law of independent assortment)

This says that for two characteristics the genes are inherited independently.

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



| Type | Genotype | Phenotype | Number | Phenotypic ratio |
|-------------|----------|-----------------|--------|------------------|
| Parental | Y- R- | yellow round | 315 | 9/16 |
| Recombinant | yy R- | green round | 108 | 3/16 |
| Recombinant | Y- rr | yellow wrinkled | 101 | 3/16 |
| Parental | yy rr | green wrinkled | 32 | 1/16 |

Ratio of yellow (dominant) to green (recessive) = 12:4 or 3:1
 Ratio of round (dominant) to wrinkled (recessive) = 12:4 or 3:1

Summary: Mendelian transmission in simple words

- One copy of each gene is inherited from the mother and one from the father. These copies are not necessarily identical
- Mendel postulated that mother and father each pass one of their two copies of each gene independently and at random
- At a given locus, the father carries *alleles* a and b and the mother carries c and d, the offspring may be a/c, a/d, b/c or b/d, each with probability 1/4
- Transmission of genes at two different positions, or loci, on the same *chromosome* (see later) may not be independent. If not, they are said to be *linked*

The cell as the basic unit of biological functioning

Let us take it a few levels up ...

- Each microscopic cell is as functionally complex as a small city. When magnified 50,000 times through electron micrographs, we see that a cell is made up of multiple complex structures, each with a different role in the cell's operation.

(<http://www.allaboutthejourney.org/cell-structure.htm>)

The cell as the basic unit of biological functioning

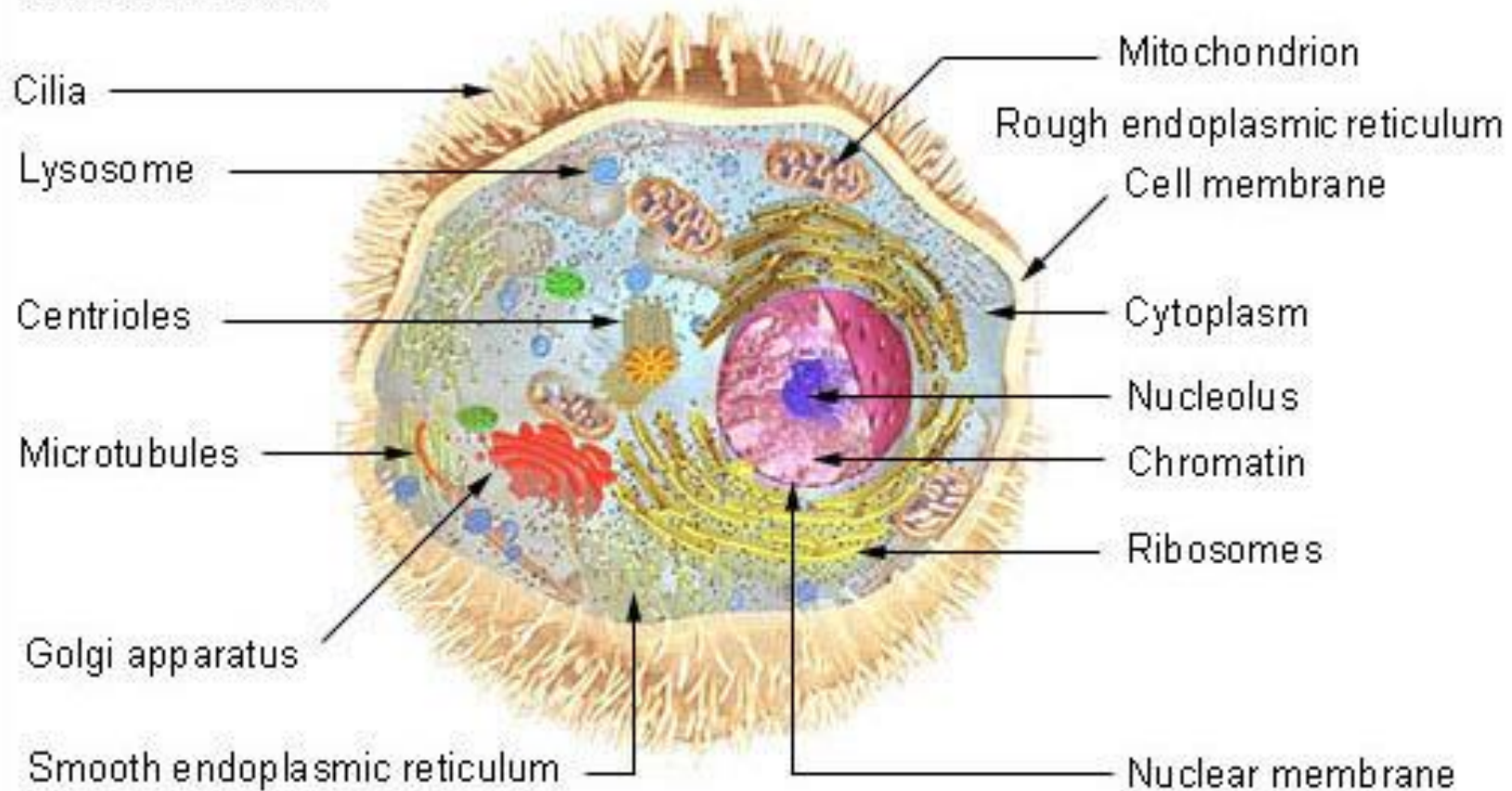
- Using the city comparison, here's a simple chart that reveals the design of a typical human cell:

| City | Cell |
|------------------|----------------------------|
| Workers | Proteins |
| Power plant | Mitochondria |
| Roads | Actin fibers, Microtubules |
| Trucks | Kinesin, Dinein |
| Factories | Ribosomes |
| Library | Genome |
| Recycling center | Lysosomes |
| Police | Chaperones |
| Post office | Golgi Apparatus |

(<http://www.allaboutthejourney.org/cell-structure.htm>)

The cell as the basic unit of biological functioning

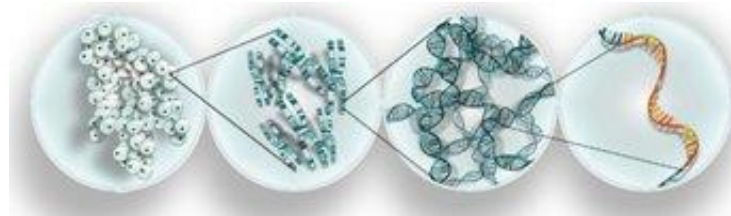
Cell Structure



(http://training.seer.cancer.gov/anatomy/cells_tissues_membranes/cells/structure.html)

DNA: the master molecule of every cell

- Three important platforms:



(VIB, Biotechnology)

- The cells of the living organism. The cells are thus the basic unit of all biological functions
 - The genetic instructions that are responsible for the properties of the cell
 - The biological mechanisms that are used by the cells to carry out the instructions.
- The genetic instructions are stored in code in the DNA (**deoxyribonucleic acid**). The collection of all possible genetic instructions in a cell is called the *genome*.

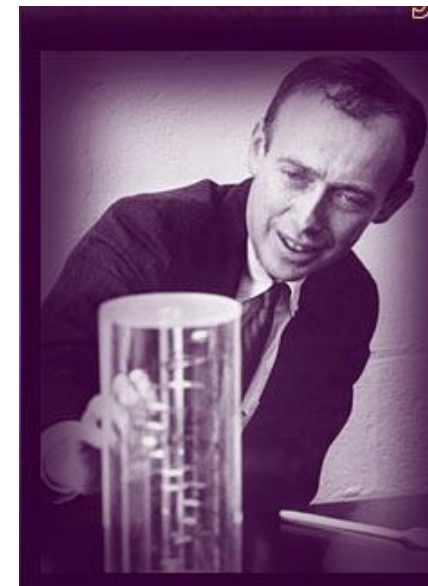
DNA: the master molecule of every cell

- It contains vital information that gets passed on to each successive generation. It coordinates the making of itself as well as other molecules (proteins). If it is changed slightly, serious consequences may result. If it is destroyed beyond repair, the cell dies.
- Changes in the DNA of cells in multicellular organisms produce variations in the characteristics of a species. Over long periods of time, natural selection acts on these variations to evolve or change the species.
- The presence or absence of DNA evidence at a crime scene could mean the difference between a guilty verdict and an acquittal. Governments have spent enormous amounts of money to unravel the sequence of DNA in the human genome in hopes of understanding /finding cures for diseases.
- Finally, from the DNA of one cell, we can clone an animal, a plant or perhaps even a human being.

From “units” to DNA

Geneticists already knew that DNA held the primary role in determining the structure and function of each cell in the body, but they did not understand the mechanism for this or that the structure of DNA was directly involved in the genetic process.

British biophysicist Francis Crick and American geneticist **James Watson** undertook a joint inquiry into the structure of DNA in 1951.

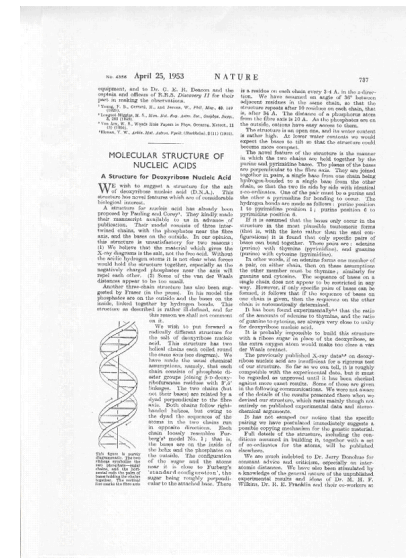


(<http://www.pbs.org/wgbh/nova/genome>)

Watson and Crick

“We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A). This structure has novel features which are of considerable biological interest.”

(Watson JD and Crick FHC. A Structure for DNA, *Nature*, 1953)



What is DNA?

Where is it found?

What makes it so special?

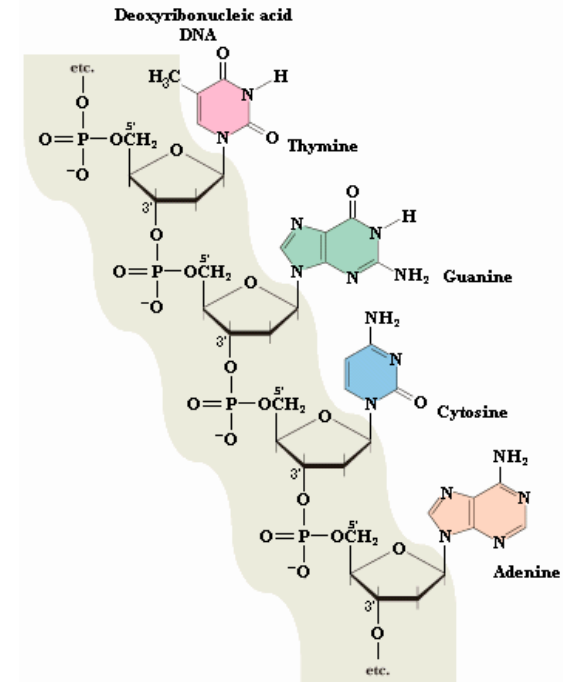
How does it work?

What is DNA?

- Deoxyribonucleic acid (DNA) is one class of molecules called “nucleic acids”. These were originally discovered in 1868 by Friedrich Meischer (isolating DNA from pus cells on bandages). At that time, he could not confirm that nucleic acids might contain genetic information.
- DNA IS the genetic information of most living organisms. In contrast, some viruses (called retroviruses) use ribonucleic acid as genetic information.
- Some interesting features of DNA include:
 - DNA can be copied over generations of cells: *DNA replication*
 - DNA can be translated into proteins: *DNA transcription* into RNA, further translated into proteins
 - DNA can be repaired when needed: *DNA repair*.
- The key to all these functions is found in the molecular structure of the DNA.

The structure of DNA

- There are 4 nucleotide *bases*, denoted A (adenine), T (thymine), G (guanine) and C (cytosine)
- A and G are called purines, T and C are called pyrimidines (smaller molecules than purines)
- The two strands of DNA in the double helix structure are complementary (sense and anti-sense strands); A binds with T and G binds with C



(Biochemistry 2nd Ed. by Garrett & Grisham)

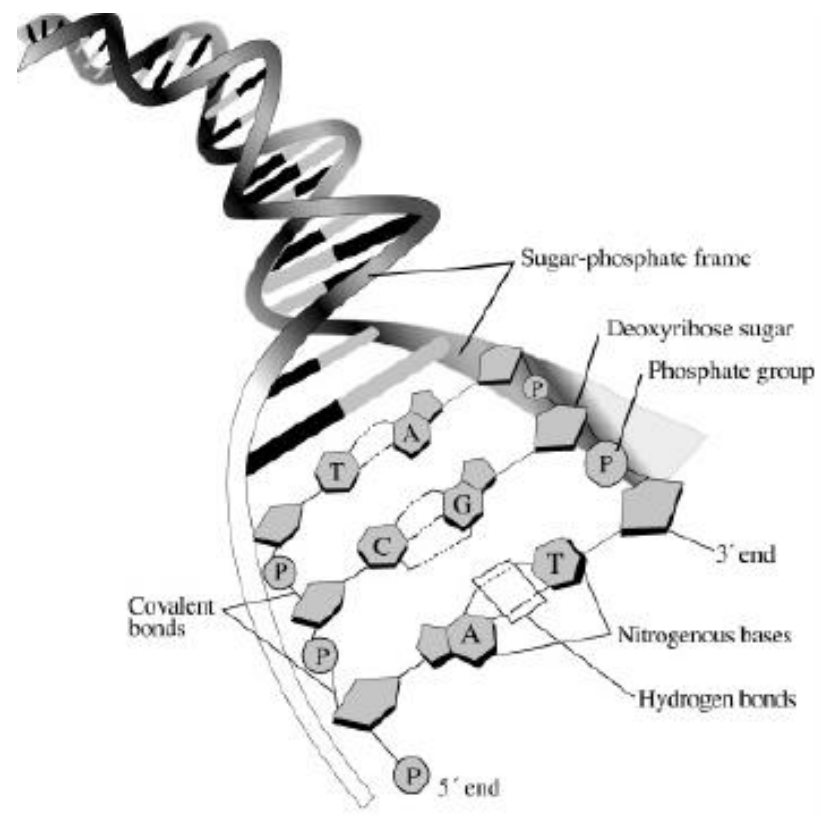
- DNA is a polymere (i.e., necklace of many alike units), made of units called nucleotides.

Primary structure of DNA

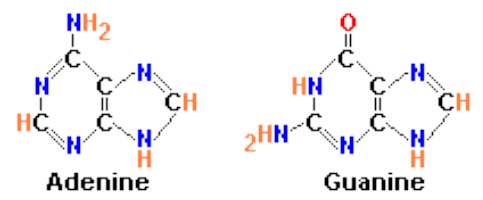
The 3 dimensional structure of DNA can be described in terms of primary, secondary, tertiary, and quaternary structure.

- The primary structure of DNA is the sequence itself - the order of nucleotides in the deoxyribonucleic acid polymer.
- A *nucleotide* consists of
 - a deoxyribose sugar, bound on one side to
 - a phosphate group, and on the other side to a
 - a nitrogenous base.
- Nucleotides can also have other functions such as carrying energy: ATP
- Note: Nucleosides are made of a sugar and a nitrogenous base...

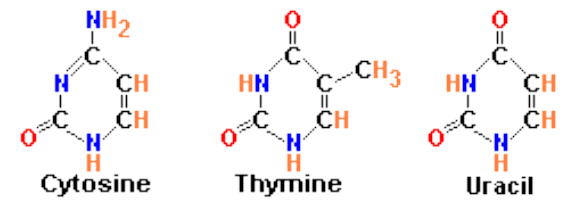
Nucleotides



Nitrogenous bases



Purines

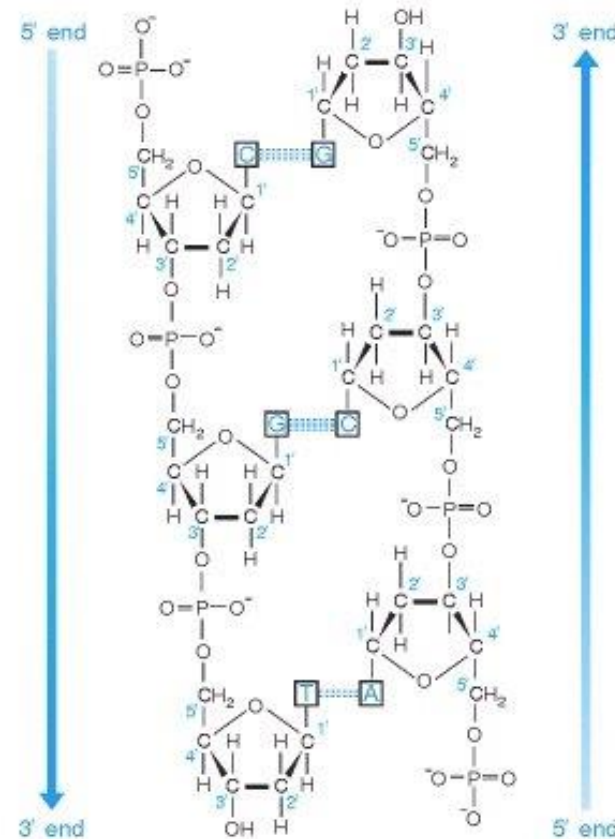


Pyrimidines

(<http://www.sparknotes.com/101/index.php/biology>)

Secondary structure of DNA

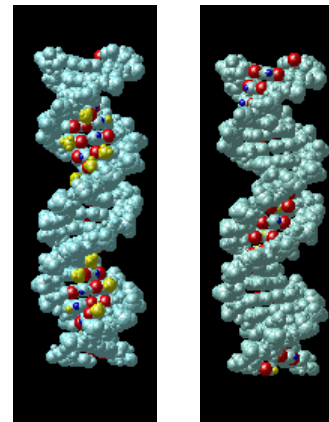
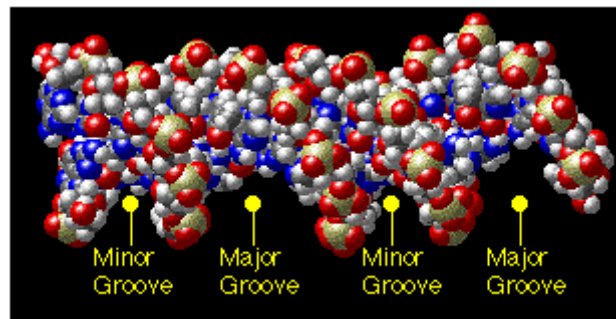
- The secondary structure of DNA is relatively straightforward - it is a *double helix*: two strands twisted together like a twisted ladder (caused by hydrogen bonds).
- The two strands are anti-parallel.
 - *The 5' end* is composed of a phosphate group that has not bonded with a sugar unit.
 - *The 3' end* is composed of a sugar unit whose hydroxyl group has not bonded with a phosphate group.



What is a base-pair?

Major groove and minor groove

- The strand backbones are closer together on one side of the helix than on the other. The *major groove* occurs where the backbones are far apart, the *minor groove* occurs where they are close together (Figure 1)
- Certain proteins bind to DNA to alter its structure or to regulate transcription (copying DNA to RNA) or replication (copying DNA to DNA). It is easier for these DNA binding proteins to interact with the bases (the internal parts of the DNA molecule) on the major groove side because the backbones are not in the way.



Tertiary structure of DNA

- This structure refers to how DNA is stored in a confined space to form the chromosomes, since the DNA needs to “fit into the cell”.
- It varies depending on whether the organisms prokaryotes and eukaryotes:
 - In prokaryotes the DNA is folded like a super-helix, usually in circular shape and associated with a small amount of protein. The same happens in cellular organelles such as mitochondria .
 - In eukaryotes, since the amount of DNA from each chromosome is very large, the packing must be more complex and compact, this requires the presence of proteins such as histones and other proteins of non-histone nature
- Hence, in humans, the double helix is itself super-coiled and is wrapped around so-called histones (see later).

- *Eukaryotes*: organisms with a rather complex cellular structure. In their cells we find *organelles*, clearly discernable compartments with a particular function and structure.
 - The organelles are surrounded by semi-permeable membranes that compartmentalize them further in the cytoplasm.
 - The Golgi apparatus (post office) is an example of an organelle that is involved in

- the transport and secretion of proteins in the cell.
- Mitochondria (power plants) are other examples of organelles, and are involved in respiration and energy production

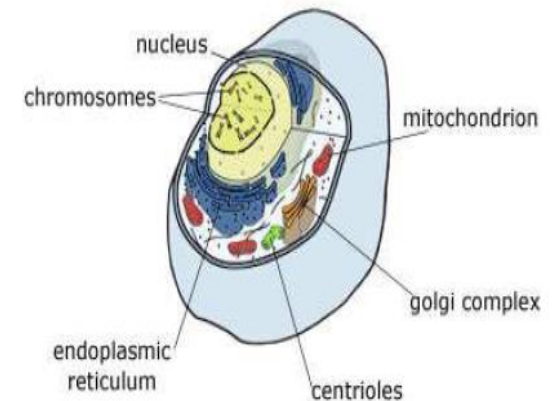
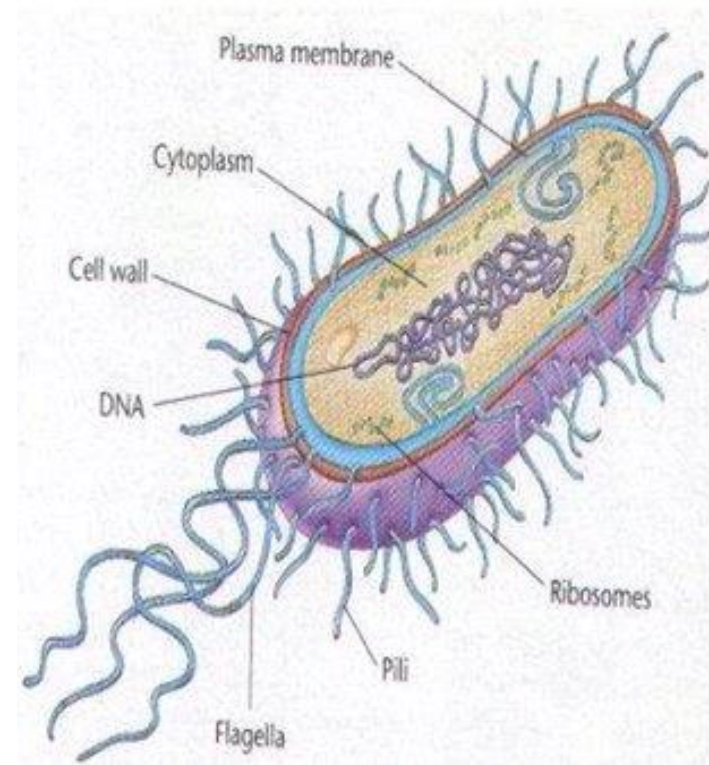


Image adapted from: National Human Genome Research Institute.

A typical eukaryotic cell.

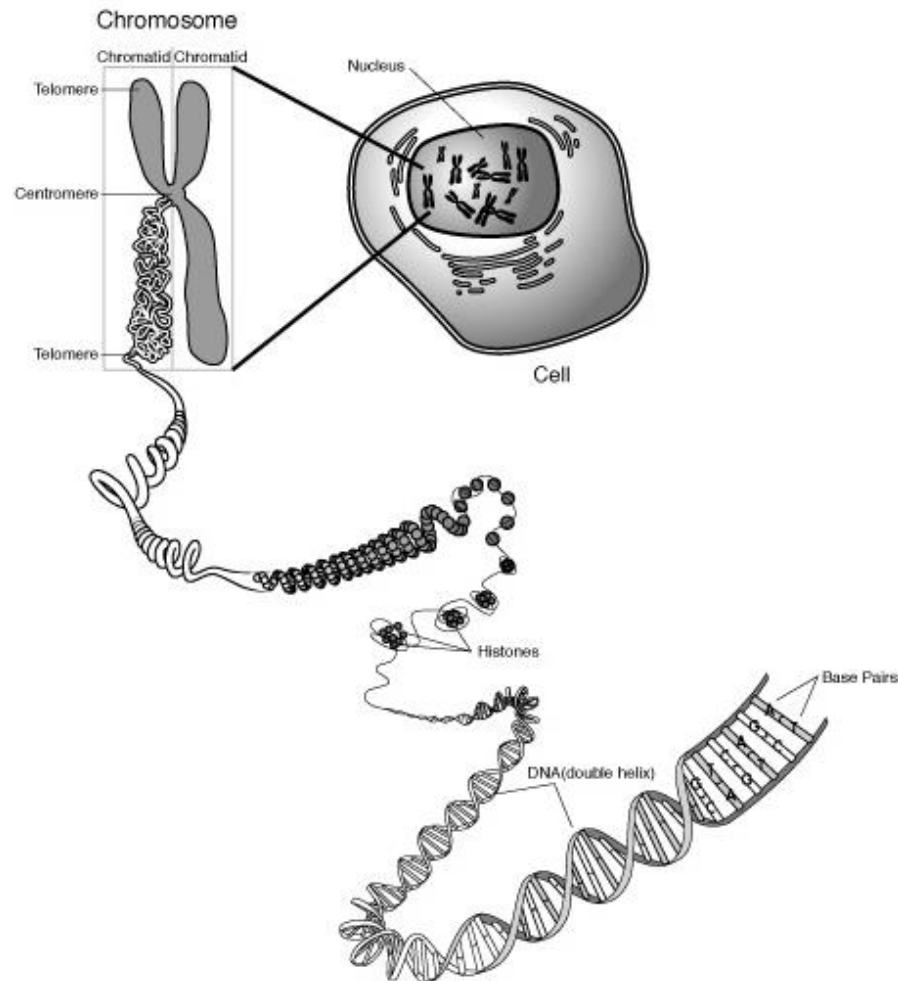
- *Prokaryotes*: cells without organelles where the genetic information floats freely in the cytoplasm



The structure of chromosomes

- 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 → Most of what researchers know about chromosomes was learned by observing chromosomes during cell division.

Histones: packaging of DNA in the nucleus

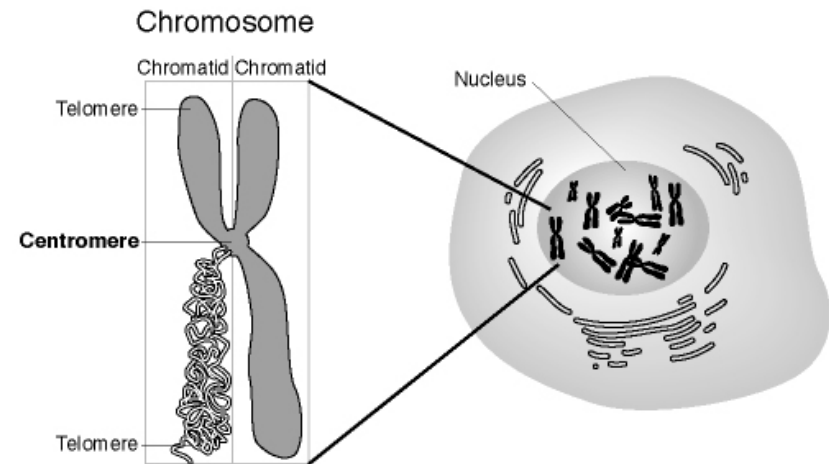


<http://www.accessexcellence.org/AB/GG/chromosome.html>

- *Histones* are proteins rich in lysine and arginine residues and thus positively-charged.
- For this reason they bind tightly to the negatively-charged phosphates in DNA.

The structure of chromosomes

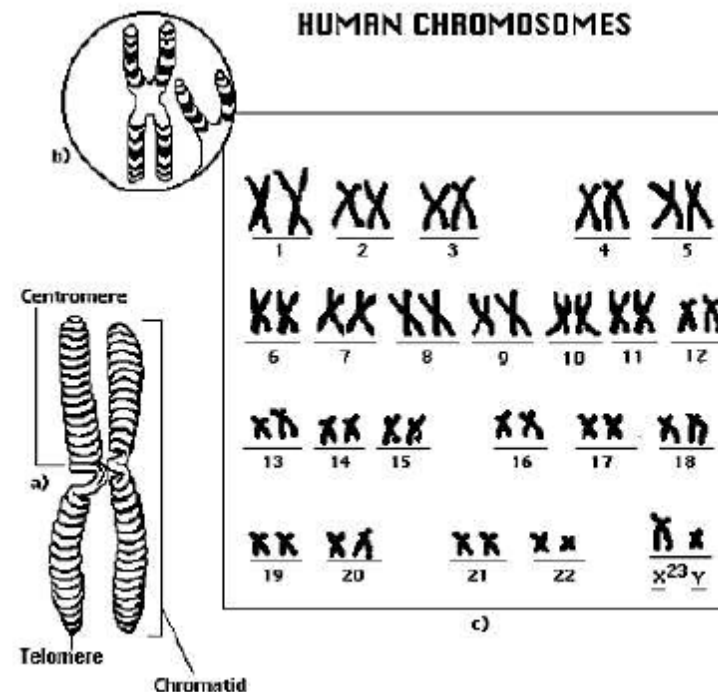
- All chromosomes have a stretch of repetitive DNA called the *centromere*. This plays an important role in chromosomal duplication before cell division.
- If the centromere is located at the extreme end of the chromosome, that chromosome is called acrocentric.
- If the centromere is in the middle of the chromosome, it is termed metacentric



(www.genome.gov)

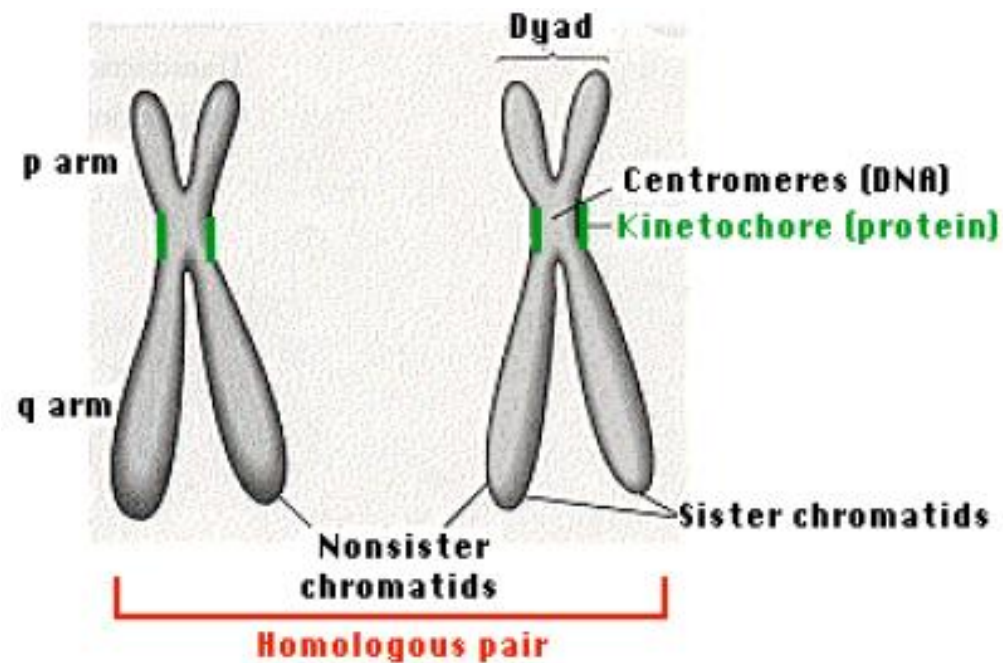
The structure of chromosomes

- The short arm of the chromosome is usually termed p for petit (small), the long arm, q , for queue (tall).



Chromosomes and chromatids

- A chromatid is one among the two identical copies of DNA making up a replicated chromosome, which are joined at their centromeres, for the process of cell division

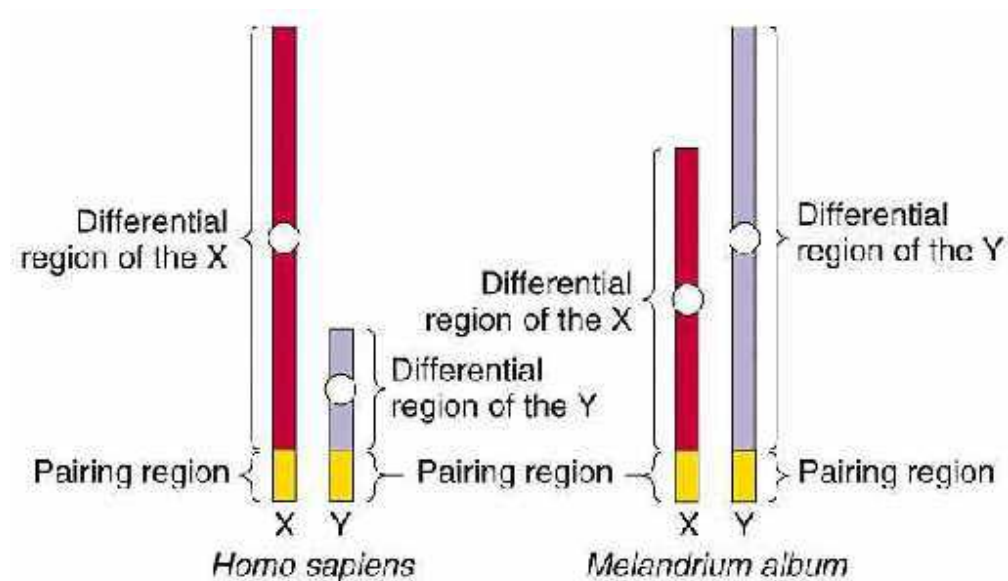


Sex chromosomes

- Homogametic sex :
that sex containing two like sex chromosomes
 - In most animals species these are females (XX)
 - Butterflies and Birds, ZZ males
- Heterogametic sex:
that sex containing two different sex chromosomes
 - In most animal species these are XY males
 - Butterflies and birds, ZW females
 - Grasshoppers have XO males

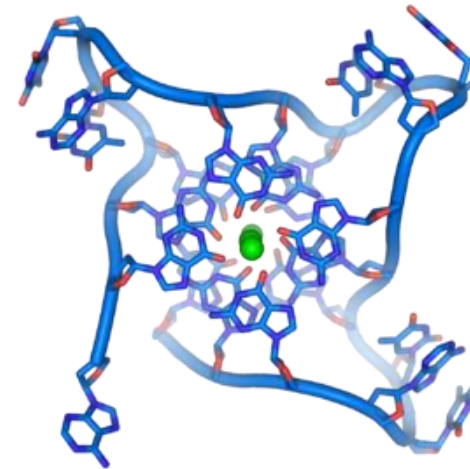
Pairing of sex chromosomes

- In the homogametic sex: pairing happens like normal autosomal chromosomes
- In the heterogametic sex: The two sex chromosomes are very different, and have special pairing regions to insure proper pairing at meiosis



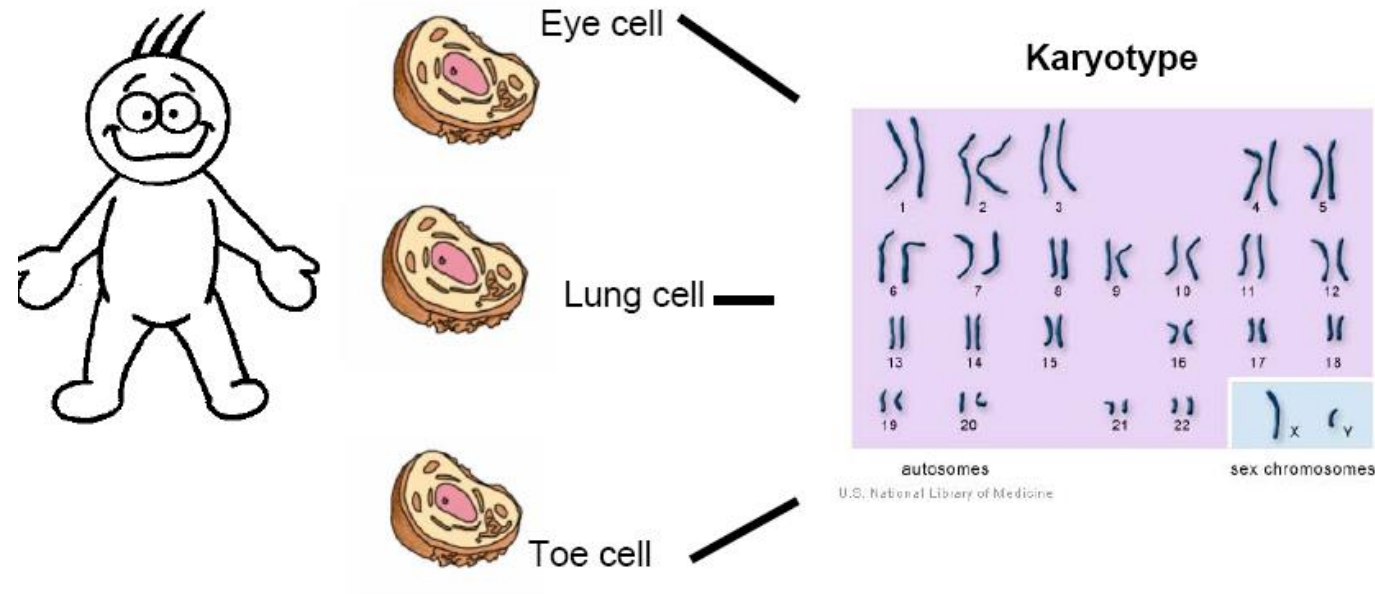
Quaternary structure of DNA

- At the ends of linear chromosomes are specialized regions of DNA called *telomeres*.
- The main function of these regions is to allow the cell to replicate chromosome ends using the enzyme telomerase, since other enzymes that replicate DNA cannot copy the 3' ends of chromosomes.
- In human cells, telomeres are long areas of single-stranded DNA containing several thousand repetitions of a single sequence TTAGGG → important role in aging.



(<http://www.boddunan.com/miscellaneous>)

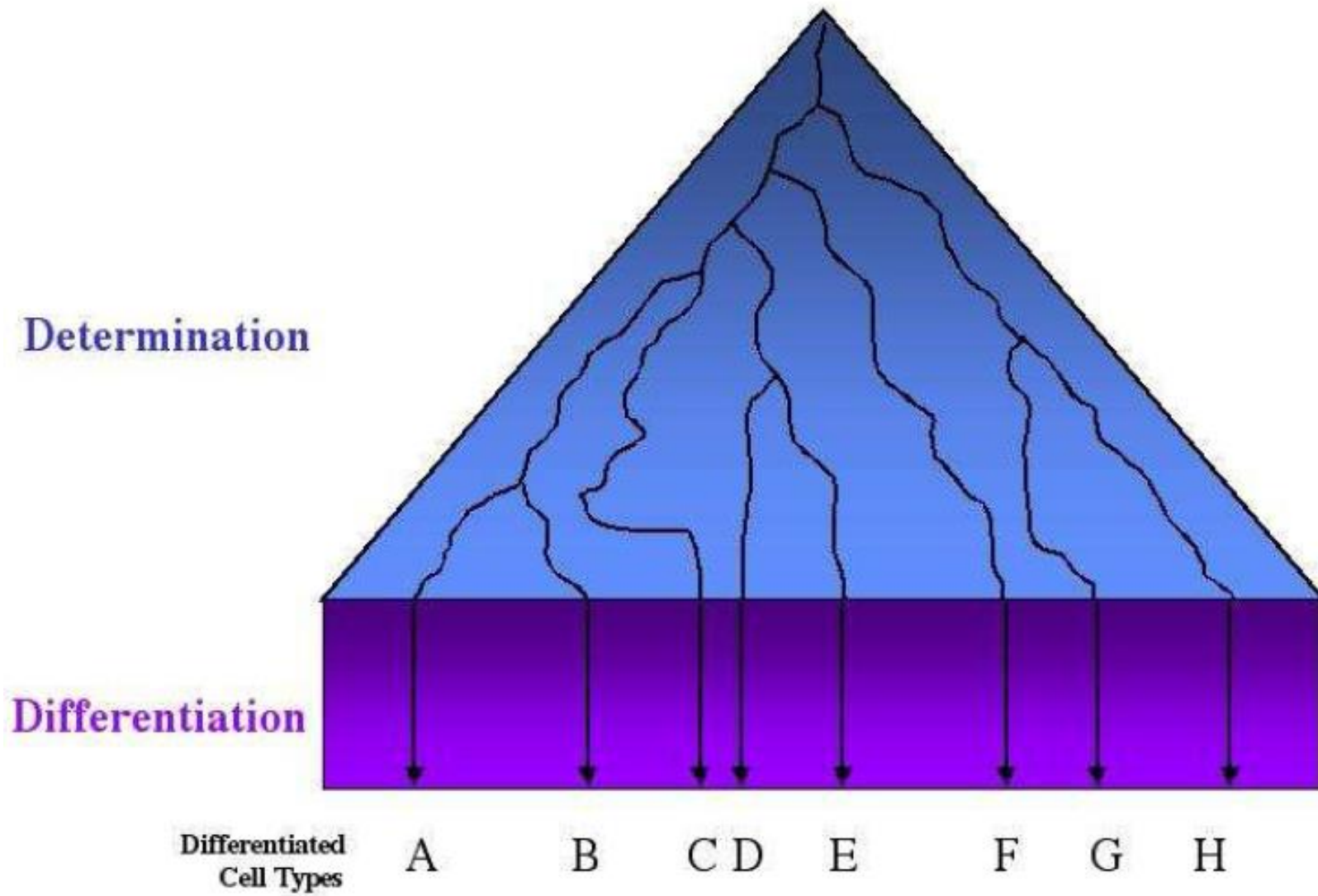
Every cell in the body has the same DNA !!!



- One base pair is 0.00000000034 meters
- DNA sequence in any two people is 99.9% identical
- The residual 0.1% leads to several million spelling differences; variations leading to dramatically higher risks of certain cancers and other diseases

Every cell in the body has the same DNA: differential expression

- The determination of different cell types (*cell fates*) involves progressive restrictions in their developmental potentials. When a cell “chooses” a particular fate, it is said to be determined, although it still “looks” just like its undetermined neighbors. Determination implies a stable change - the fate of determined cells does not change.
- *Differentiation* follows determination, as the cell elaborates a cell-specific developmental program. Differentiation results in the presence of cell types that have clear-cut identities, such as muscle cells, nerve cells, and skin cells.
- Differentiation results from differential *gene expression*



A note aside: X-inactivation

- *X-inactivation* is a process by which one of the two copies of the X chromosome present in female mammals is inactivated
- X-inactivation occurs so that the female, with two X chromosomes, does not have twice as many X chromosome gene products as the male, which only possess a single copy of the X chromosome

The ginger color of cats (known as "yellow", "orange" or "red" to cat breeders) is caused by the "O" gene. The O gene changes black pigment into a reddish pigment. The O gene is carried on the X chromosome.

The O gene is called a sex-linked gene because it is carried on a sex chromosome.

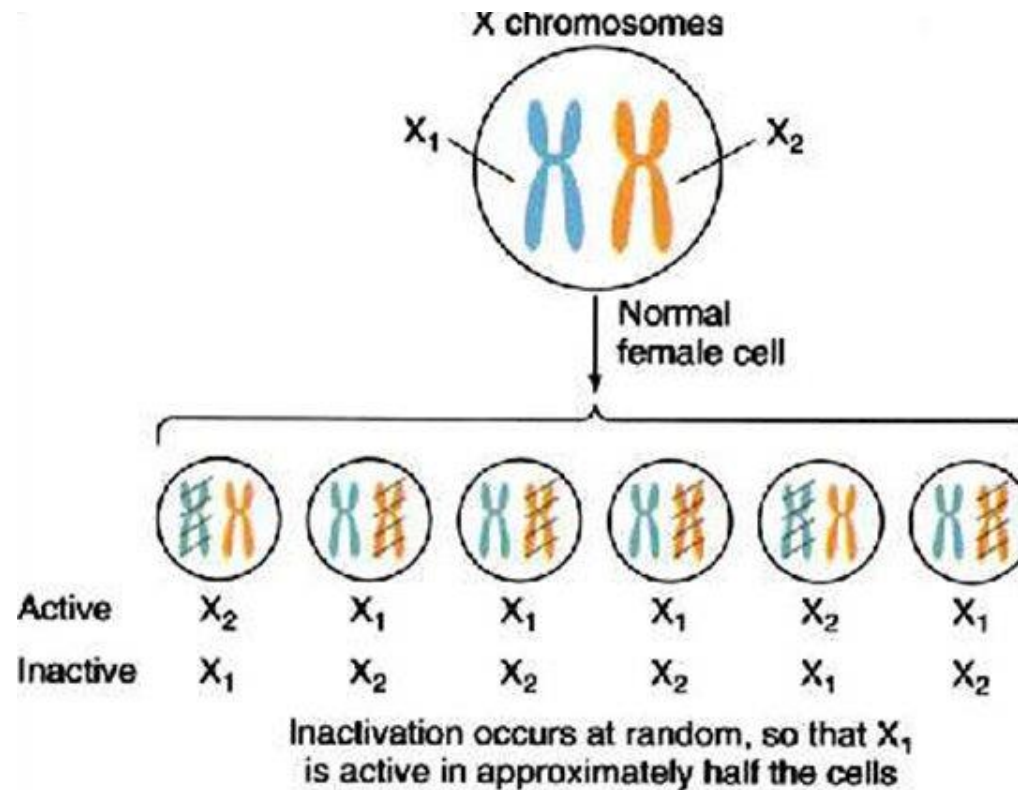
The formation of red and black patches in a female with only one O gene is through a process known as X-chromosome inactivation. Some cells randomly activate the O gene while others activate the gene in the equivalent place on the other X chromosome.



(wikipedia)

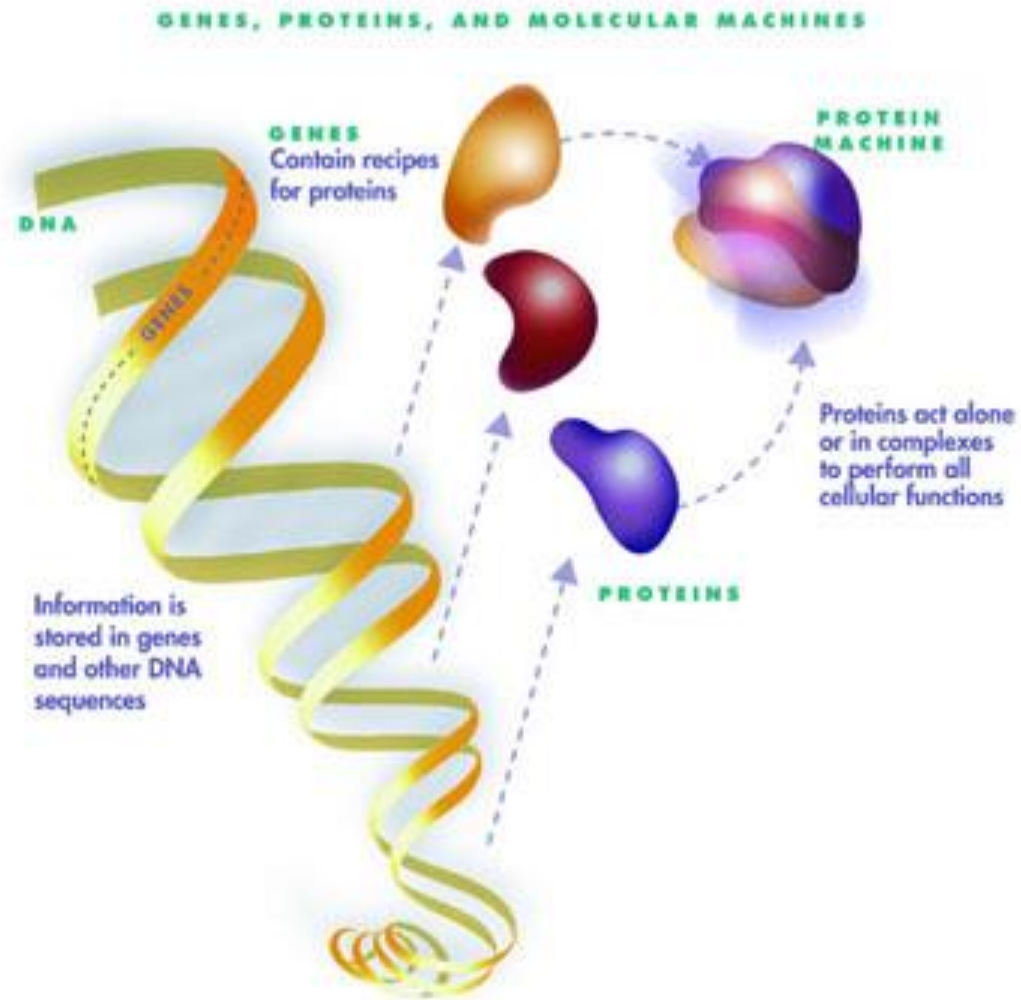
A note aside: X-inactivation

- The choice of which X chromosome will be inactivated is random in placental mammals such as mice and humans, but once an X chromosome is inactivated it will remain inactive throughout the “lifetime” of the cell.



In summary: crude composition of the human genome

- The human genome consists of about 3×10^9 base pairs and contains about 22,000 genes
- Cells containing 2 copies of each chromosome are called *diploid* (most human cells). Cells that contain a single copy are called *haploid*.
- Humans have 23 pairs of chromosomes: 22 autosomal pairs and one pair of sex chromosomes
- Females have two copies of the X chromosome, and males have one X and one Y chromosome
- DNA carries the information for making the cell's proteins. These proteins implement all functions of a living organism and determine the organism's characteristics. When the cell reproduces, it has to pass all of his information to the daughter cells.



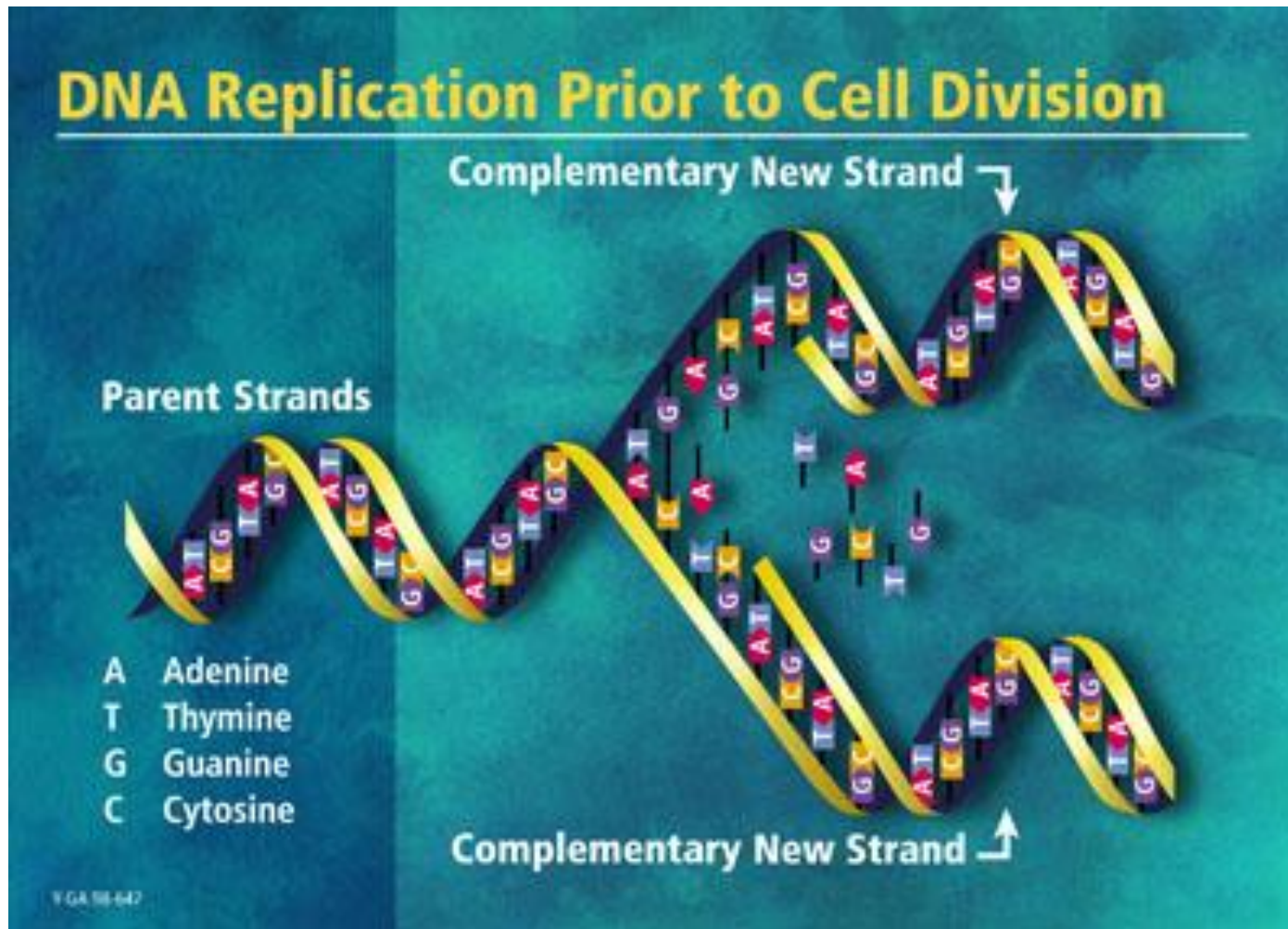
DNA replication

- We have mentioned a few times “DNA replication” ... : before a cell can reproduce, it must first *replicate* (making a copy)
- A wide variety of proteins form complexes with DNA in order to replicate it, transcribe it into RNA (the “other” nucleic acid), and regulate the transcriptional process (central dogma of molecular biology).
 - Recall that *proteins* are long chains of amino acids
 - [An *amino acids* being an organic compound containing amongst others an amino group (NH₂) and a carboxylic acid group (COOH)]
 - Think of amino acids as 3-letter words of nucleotide building blocks (the letters A,G ,T,C).
 - An example of a protein is a histone

DNA replication

- Where DNA replication occurs depends upon whether the cell is a prokaryote or a eukaryote.
- DNA replication occurs in the cytoplasm of prokaryotes and in the nucleus of eukaryotes. Regardless of where DNA replication occurs, the basic process is the same.
- The structure of DNA lends itself easily to DNA replication. Each side of the double helix runs in opposite (anti-parallel) directions. The beauty of this structure is that it can unzip down the middle and each side can serve as a pattern or template for the other side (called semi-conservative replication). However, DNA does not unzip entirely. It unzips in a small area called a replication fork, which then moves down the entire length of the molecule.

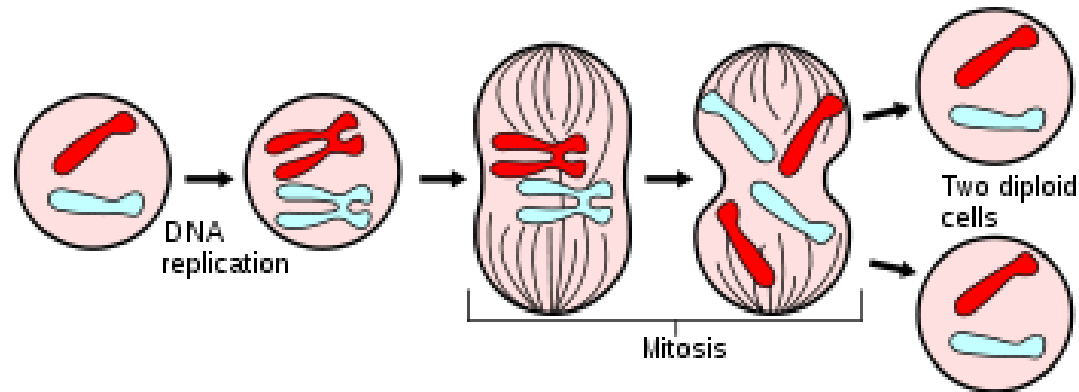
DNA replication



DNA replication

1. An enzyme called **DNA gyrase** makes a nick in the double helix and each side separates
2. An enzyme called **helicase** unwinds the double-stranded DNA
3. Several small proteins called **single strand binding proteins (SSB)** temporarily bind to each side and keep them separated
4. An enzyme complex called **DNA polymerase** "walks" down the DNA strands and adds new nucleotides to each strand. The nucleotides pair with the complementary nucleotides on the existing strand (A with T, G with C).
5. A subunit of the DNA polymerase **proofreads** the new DNA
6. An enzyme called **DNA ligase** seals up the fragments into one long continuous strand
7. The new copies **automatically wind up again**

DNA replication



- Different types of cells replicated their DNA at different rates. Some cells constantly divide, like those in your hair and fingernails and bone marrow cells. Other cells go through several rounds of cell division and stop (including specialized cells, like those in your brain, muscle and heart).
- Finally, some cells stop dividing, but can be induced to divide to repair injury (such as skin cells and liver cells).
- In cells that do not constantly divide, the cues for DNA replication/cell division come in the form of chemicals (coming from other parts of the body – hormones - or from the environment).

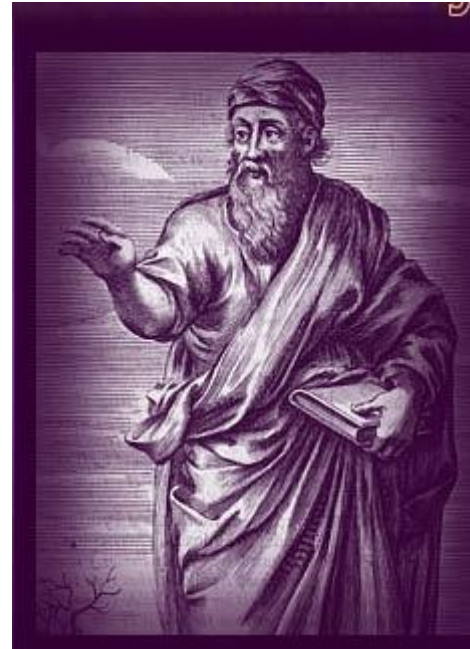
A note aside: A historical view on genetic information transmission from generation to generation

Source:

<http://www.pbs.org/wgbh/nova/genome>

Pythagoras (580-500 BC)

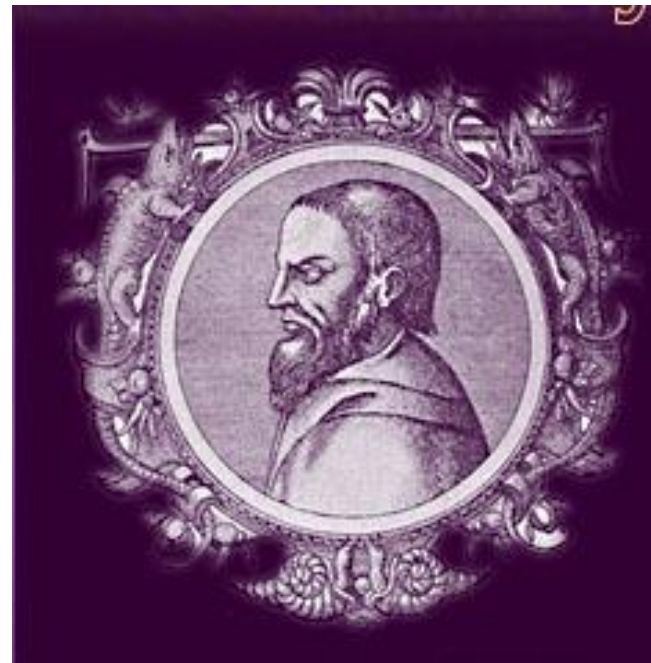
- Pythagoras surmised that all hereditary material came from a child's father. The mother provided only the location and nourishment for the fetus.
- Semen was a cocktail of hereditary information, coursing through a man's body and collecting fluids from every organ in its travels. This male fluid became the formative material of a child once a man deposited it inside a woman.



Aristotle (384-322 BC)

- Aristotle's understanding of heredity, clearly following from Pythagorean thought, held wide currency for almost 2,000 years.
- The Greek philosopher correctly believed that both mother and father contribute biological material toward the creation of offspring, but he was mistakenly convinced that a child is the

product of his or her parents' commingled blood.



De Maupertuis (1698-1759)

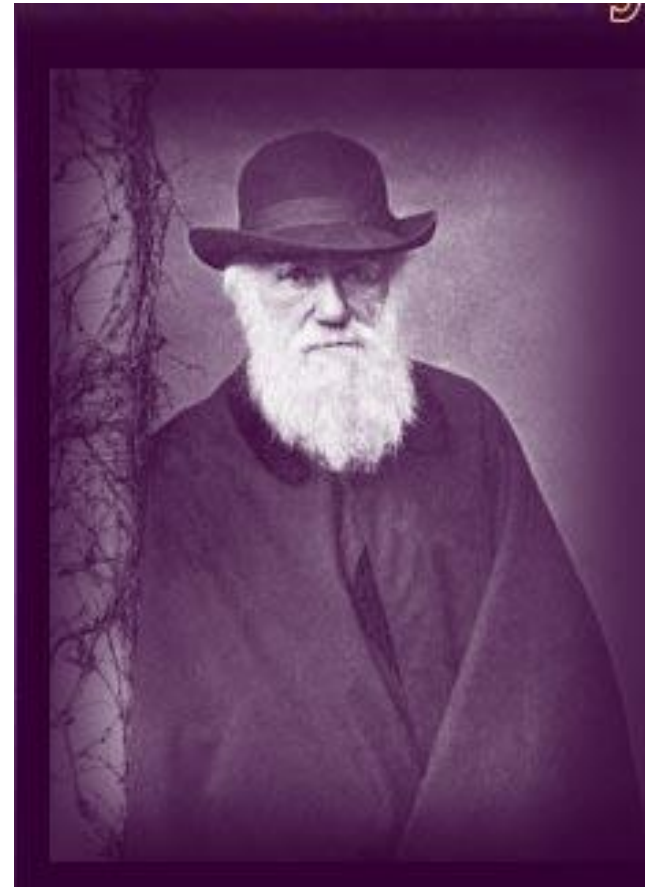
- In his 1751 book, *Système de la nature* (System of Nature), French mathematician, biologist, and astronomer Pierre-Louis Moreau de Maupertuis initiated the first speculations into the modern idea of dominant and recessive genes. De Maupertuis studied the occurrences of polydactyly (extra fingers) among several generations of one family and showed how this trait could be

passed through both its male and female members.



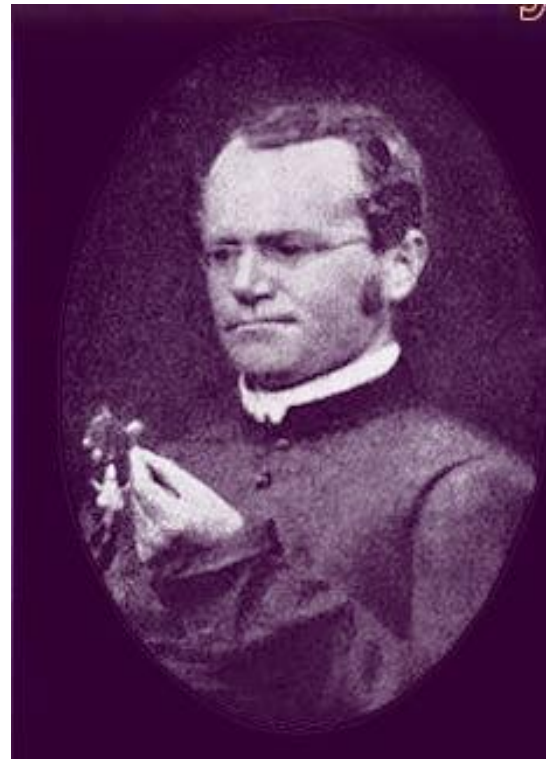
Darwin (1809-1882)

- Darwin's ideas of heredity revolved around his concept of "pangenesis." In pangenesis, small particles called pangenes, or gemmules, are produced in every organ and tissue of the body and flow through the bloodstream. The reproductive material of each individual formed from these pangenes was therefore passed on to one's offspring.



Here we meet again ... our friend Mendel (1822-1884)

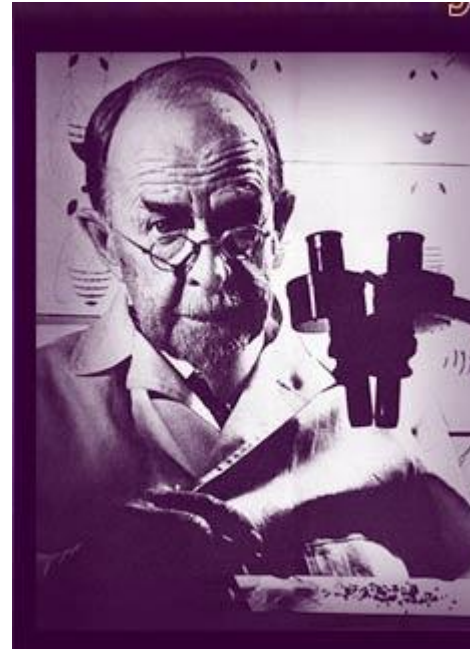
- Gregor Mendel, an Austrian scientist who lived and conducted much of his most important research in a Czechoslovakian monastery, established the basis of modern genetic science. He experimented on pea plants in an effort to understand how a parent passed physical traits to its offspring. In one experiment, Mendel crossbred a pea plant with wrinkled seeds and a pea plant with smooth seeds.
- All of the hybrid plants produced by this union had smooth seeds...



Morgan (1866-1945)

- Thomas Hunt Morgan began experimenting with *Drosophila*, the fruit fly, in 1908. He bred a single white-eyed male fly with a red-eyed female. All the offspring produced by this union, both male and female, had red eyes. From these and other results, Morgan established a theory of heredity that was based on the idea that genes, arranged on the chromosomes, carry hereditary

factors that are expressed in different combinations when coupled with the genes of a mate.



Crick (1916-2004) and Watson (1928-)

- Employing X-rays and molecular models, Watson and **Crick** discovered the double helix structure of DNA. Suddenly they could explain how the DNA molecule duplicates itself by forming a sister strand to complement each single, ladder-like DNA template.



1.b How does DNA encoding work?

Translation table from DNA building stones to protein building stones

| | | Second Base | | | | | | | |
|---|-----|-------------|-----|-----|------|-----|------|-----|---|
| | | U | C | A | G | | | | |
| U | UUU | Phe | UCU | Ser | UAU | Tyr | UGU | Cys | U |
| | UUC | | UCC | | UAC | | UGC | | C |
| | UUA | Leu | UCA | UAA | Stop | UGA | Stop | A | |
| | UUG | | UCG | UAG | Stop | UGG | Trp | G | |
| C | CUU | Leu | CCU | Pro | CAU | His | CGU | Arg | U |
| | CUC | | CCC | | CAC | | CGC | | C |
| | CUA | | CCA | | CAA | Gln | CGA | | A |
| | CUG | | CCG | | CAG | | CGG | | G |
| A | AUU | Ile | ACU | Thr | AAU | Asn | AGU | Ser | U |
| | AUC | | ACC | | AAC | | AGC | C | |
| | AUA | | ACA | | AAA | Lys | AGA | Arg | A |
| | AUG | | ACG | | AAG | | AGG | G | |
| G | GUU | Val | GCU | Ala | CAU | Asp | GGU | Gly | U |
| | GUC | | GCC | | GAC | | GGC | | C |
| | GUA | | GCA | | GAA | Glu | GGA | | A |
| | GUG | | GCG | | GAG | | GGG | | G |

(Roche Genetics)

Comparison between DNA and RNA

- Pieces of coding material that the cells needs at a particular moment, is transcribed from the DNA in RNA for use outside the cell nucleus.

| Char | DNA | RNA |
|---------------------|------------------------------------|---------------------------------------|
| Major cellular site | nucleus | cytoplasm (cell area outside nucleus) |
| Major function | genetic material; | carries out instructions |
| | directs protein synthesis; | for protein synthesis |
| | replicates itself before cell div. | |
| Sugar | deoxyribose | ribose |
| Bases | A, C, T, G | A, C, U(racil), G |
| Structure | double strand coiled | single straight or |
| | into a double helix | folded strand |

(Human Anatomy & Physiology - Addison-Wesley 4th ed)

- Note that in RNA U(racil), another pyrimidine, replaces T in DNA

Translation table from DNA building stones to protein building stones

- Because there are only 20 amino acids that need to be coded (using A, C, U or G), the genetic code can be said to be degenerate, with the third position often being redundant
- The code is read in triplets of bases.
- But depending on the starting point of reading, there are three possible variants to translate a given base sequence into an amino acid sequence. These variants are called *reading frames*

Reading the code

G U C A U G U U U A G C G C A A U C A G G A A G U G U
Val Met Phe Ser Ala Ile Arg Lys Cys

G U C A U G U U U A G C G C A A U C A G G A A G U G U
Ser Cys Leu Ala Gln Ser Gly Ser

G U C A U G U U U A G C G C A A U C A G G A A G U G U
His Val Stop Arg Asn Gln Glu Val

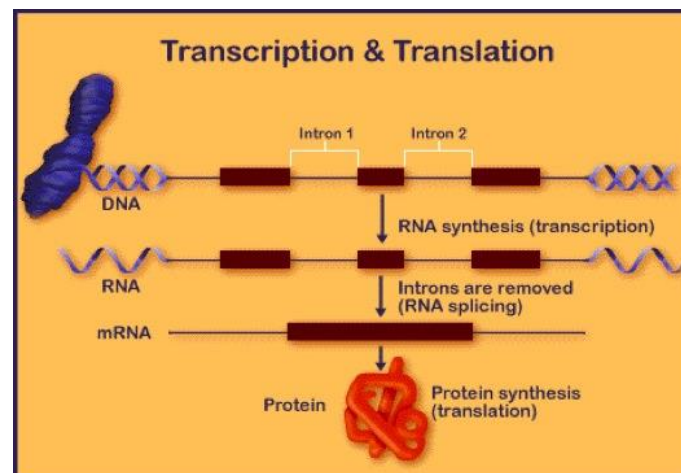
Building a protein: transcription

Building proteins is very much like building a house:

- The master blueprint is DNA, which contains all of the information to build the new protein (house).
- The working copy of the master blueprint is called **messenger RNA** (mRNA), which is copied from DNA.
- The construction site is either the **cytoplasm** in a prokaryote or the **endoplasmic reticulum** (ER) in a eukaryote.
- The building materials are amino acids.
- The construction workers are ribosomes and transfer RNA molecules.

Building a protein: transcription

- In a eukaryote, DNA never leaves the nucleus, so its information must be copied.
- In the context of building a protein, this copying process is called *transcription* and the copy is mRNA.
- Transcription takes place in the cytoplasm (prokaryote) or in the nucleus (eukaryote).
- The transcription is performed by an enzyme called RNA polymerase.

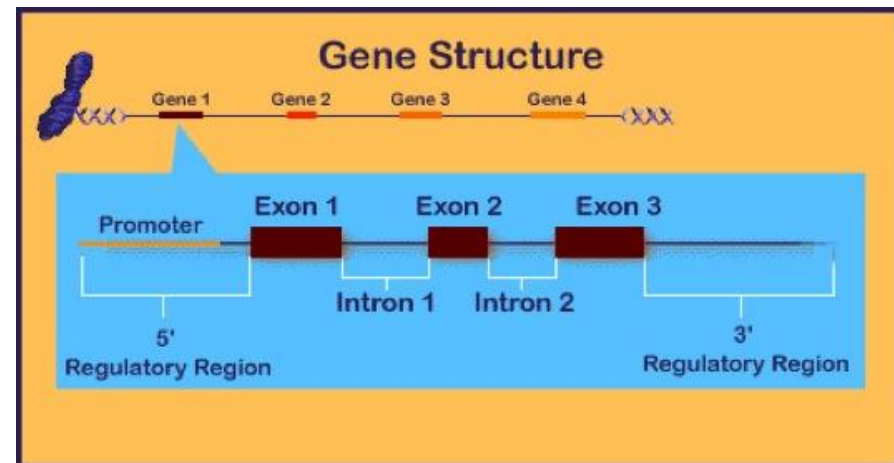


Building a protein: transcription

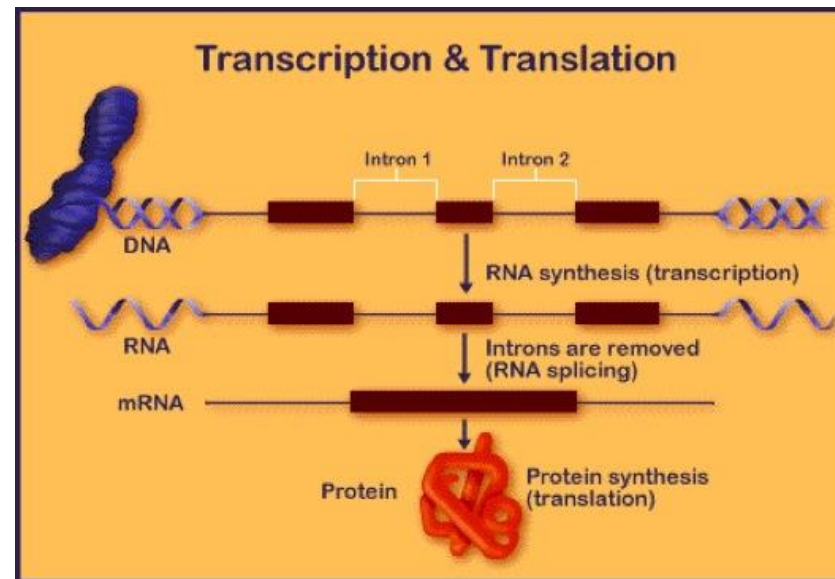
- To make mRNA, RNA polymerase:
 1. Binds to the DNA strand at a specific sequence of the gene called a promoter
 2. Unwinds and unlinks the two strands of DNA
 3. Uses one of the DNA strands as a guide or template
 4. Matches new nucleotides with their complements on the DNA strand [G with C, A with U -- remember that RNA has uracil (U) instead of thymine (T)]
 5. Binds these new RNA nucleotides together to form a complementary copy of the DNA strand (mRNA)
 6. Stops when it encounters a termination sequence of bases (stop codon)

Building a protein: transcription

- mRNA is happy to live in a single-stranded state (as opposed to DNA's desire to form complementary double-stranded helices).
- In prokaryotes, all of the nucleotides in the mRNA are part of codons for the new protein. However, in eukaryotes only, there are extra sequences in the DNA and mRNA that don't code for proteins called **introns**.
- This mRNA is then further processed:
 1. Introns get cut out
 2. The coding sequences get spliced together
 3. A special nucleotide "cap" gets added to one end
 4. A long tail consisting of 100 to 200 adenine nucleotides is added to the other end



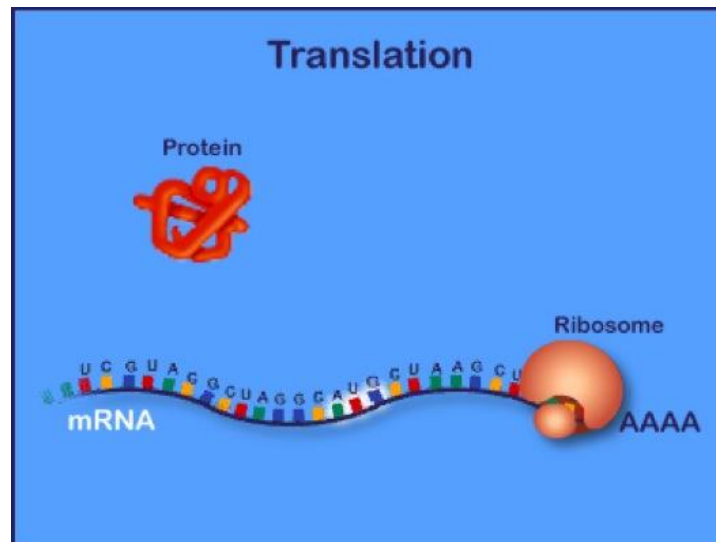
Building a protein: transcription



- The working copy of the blueprint (mRNA) must now go to the construction site where the workers will build the new protein.
- If the cell is a prokaryote such as an *E. coli* bacterium, then the site is the cytoplasm. If the cell is a eukaryote, such as a human cell, then the mRNA leaves the nucleus through large holes in the nuclear membrane (nuclear pores) and goes to the endoplasmic reticulum (ER).

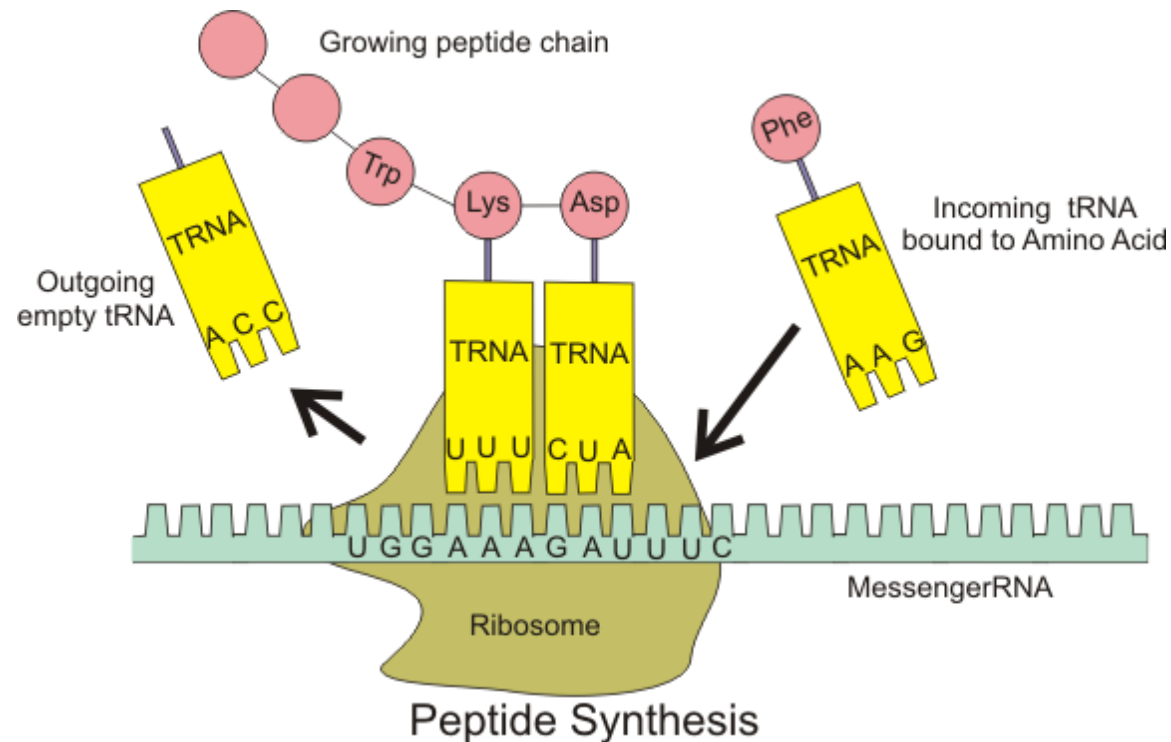
Building a protein: translation (assembly)

- To continue with our house example, once the working copy of the blueprint has reached the site, the workers must assemble the materials according to the instructions; this process is called **translation**.
- In the case of a protein, the workers are the ribosomes and special RNA molecules called **transfer RNA (tRNA)**. The construction materials are the amino acids.

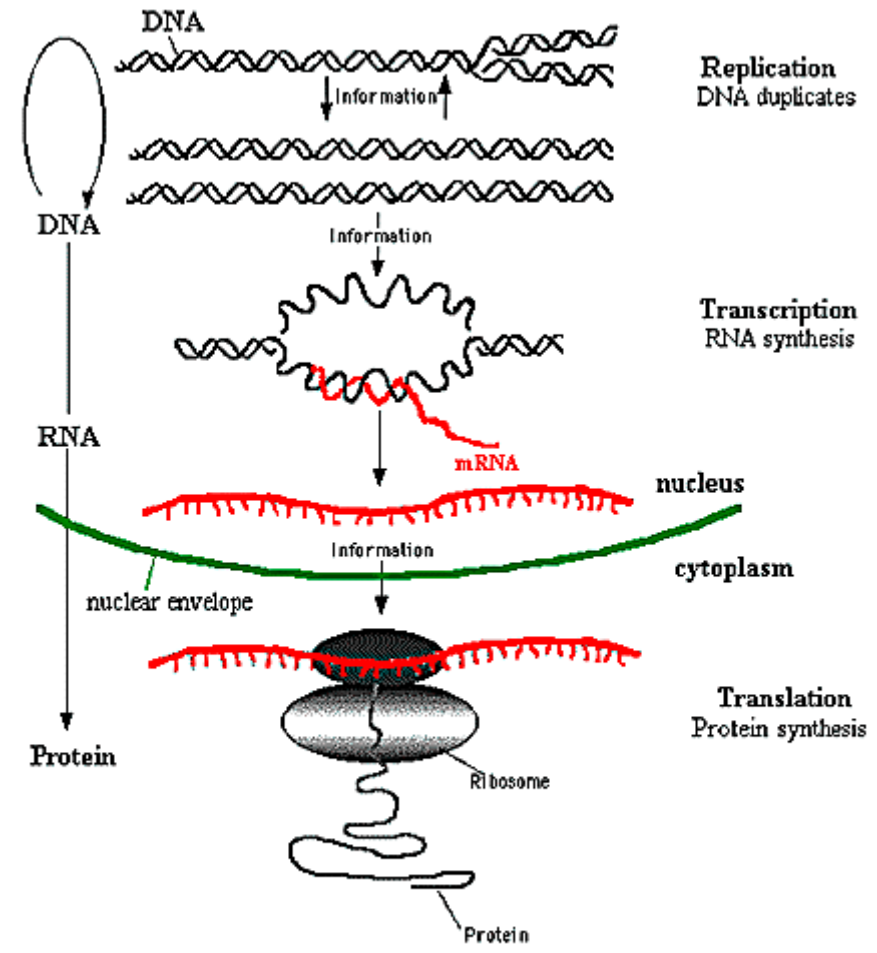


tRNA

- *Transfer RNA* (tRNA) molecules transport amino acids to the growing protein chain. Each tRNA carries an amino acid at one end and a three-base pair region, called the anti-codon, at the other end. The anti-codon binds with the codon on the protein chain via base pair matching.



The central dogma of molecular biology – in a picture



The central dogma of molecular biology – in words

- Stage 1: DNA replicates its information in a process that involves many enzymes. This stage is called the ***replication*** stage.
- Stage 2: The DNA codes for the production of messenger RNA (mRNA) during ***transcription*** of the sense strand (coding or non-template strand)
- Stage 3: In eukaryotic cells, the mRNA is ***processed*** (essentially by splicing) and migrates from the nucleus to the cytoplasm
- Stage 4: mRNA carries coded information to ribosomes. The ribosomes "read" this information and use it for protein synthesis. This stage is called the ***translation*** stage.

All living organisms share a common biomolecular basis

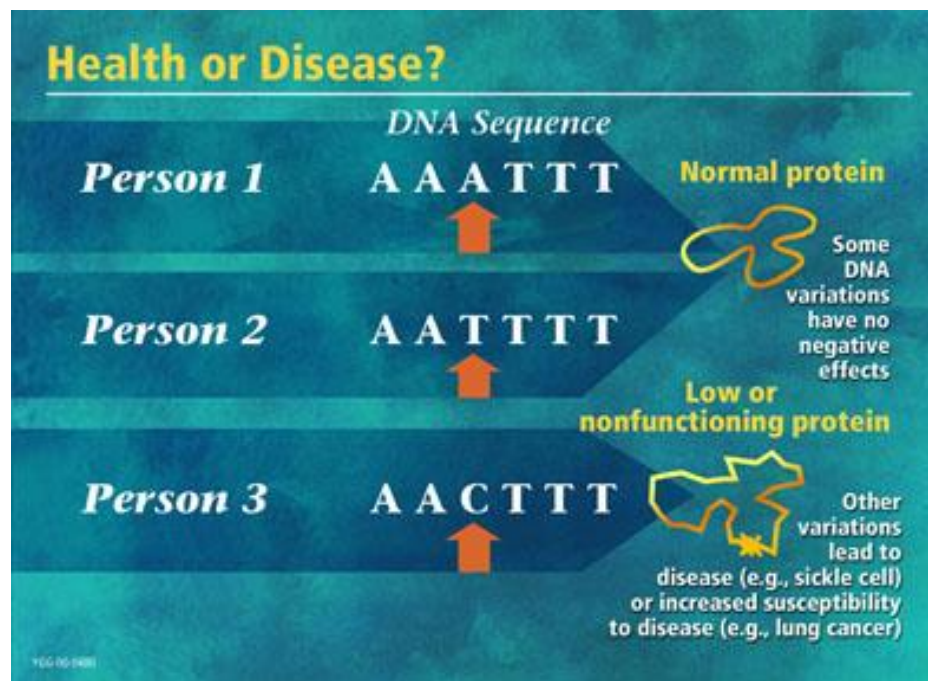
<http://videos.howstuffworks.com/discovery/28756-assignment-discovery-cell-dna-video.htm>

1.c DNA mutations

A source of variation

- As DNA polymerase copies the DNA sequence, some mistakes may occur.
- For example, one DNA base in a gene might get substituted for another. This is called a **mutation** (specifically a **point mutation**) or variation in the gene.
- Because the genetic code has built-in redundancies, this mistake might not have much effect on the protein made by the gene. In some cases, the error might be in the third base of a codon and still specify the same amino acid in the protein.

- In other cases, it may be elsewhere in the codon and specify a different amino acid. If the changed amino acid is not in a crucial part of the protein, then there may be no adverse effect. However, if the changed amino acid is in a crucial part of the protein, then the protein may be defective and not work as well or at all; this type of change can lead to disease.

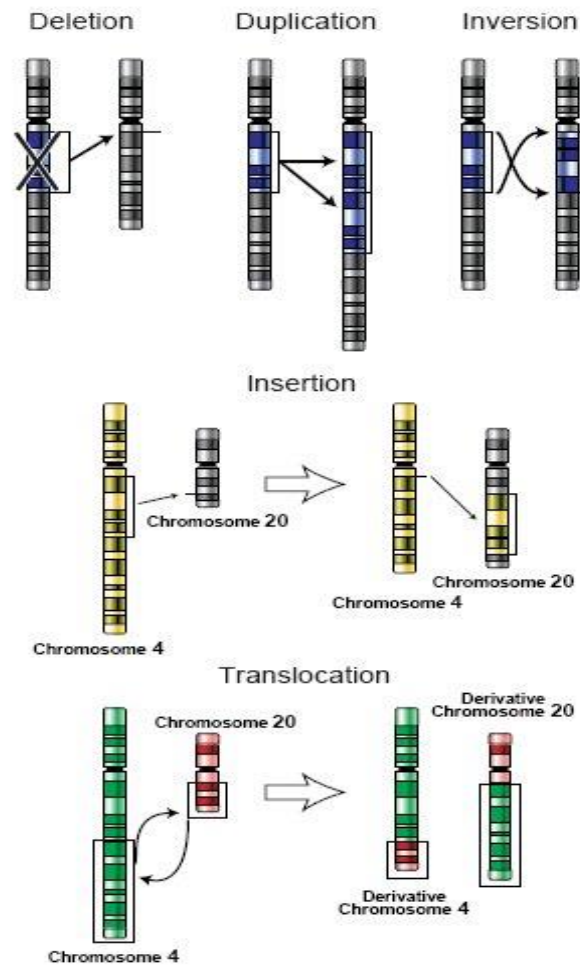


Variations in the sequence of genes during can have important consequences and cause disease.

(Photo courtesy U.S. Department of Energy Human Genome Program)

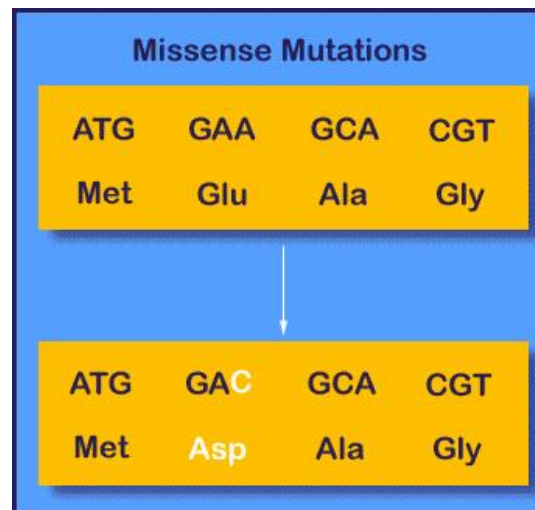
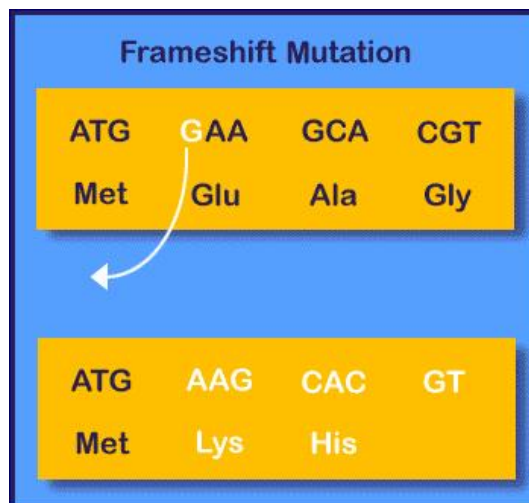
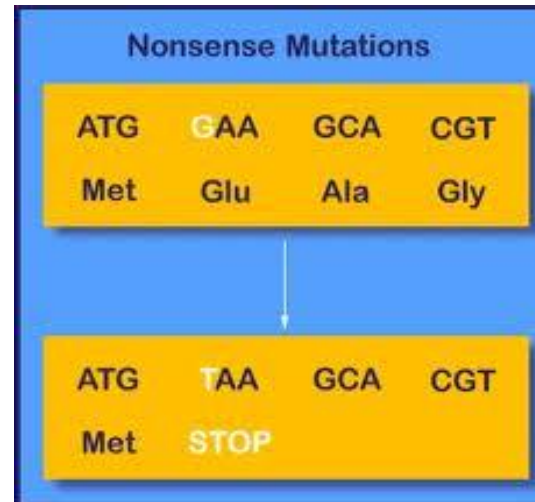
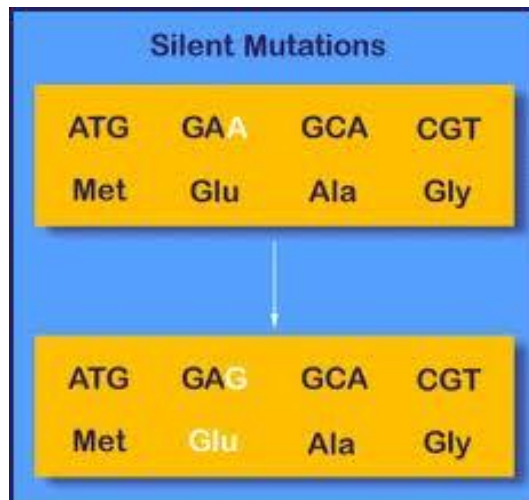
Types of mutations

- Deletion
- Duplication
- Inversion
- Insertion
- Translocation



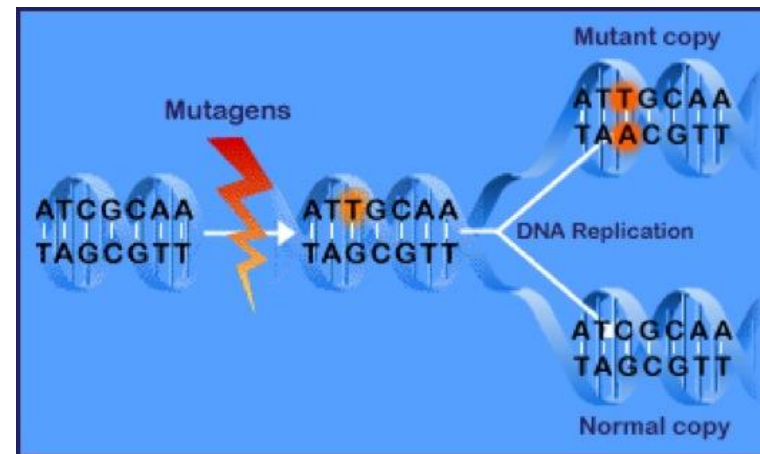
(National Human Genome Research Institute)

Types of mutations



DNA repair mechanisms

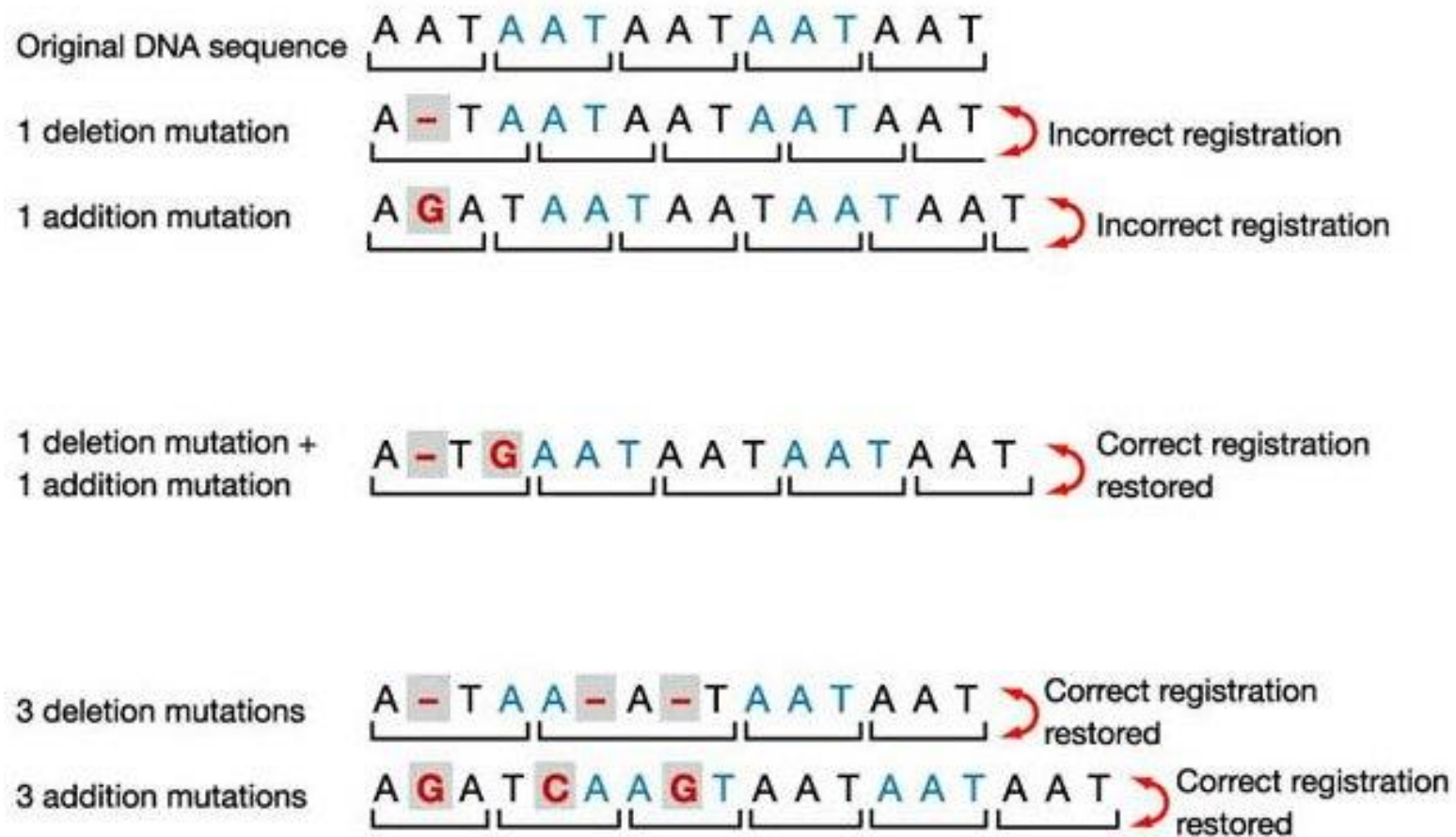
- In biology, a *mutagen* (Latin, literally origin of change) is a physical or chemical agent that changes the genetic material (usually DNA) of an organism and thus increases the frequency of mutations above the natural background level.
- As many mutations cause cancer, mutagens are typically also carcinogens.
- Not all mutations are caused by mutagens: so-called "spontaneous mutations" occur due to errors in DNA replication, repair and recombination.



(Roche genetics)

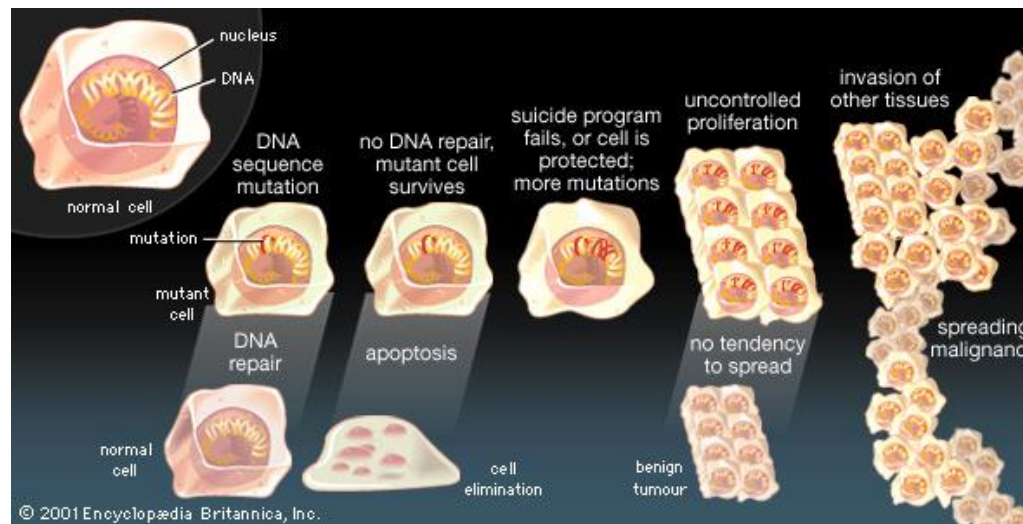
DNA repair mechanisms

- Where it can go wrong when reading the code ...



DNA repair mechanisms

- **damage reversal:** simplest; enzymatic action restores normal structure without breaking backbone
- **damage removal:** involves cutting out and replacing a damaged or inappropriate base or section of nucleotides
- **damage tolerance:** not truly repair but a way of coping with damage so that life can go on



DNA Sequencing

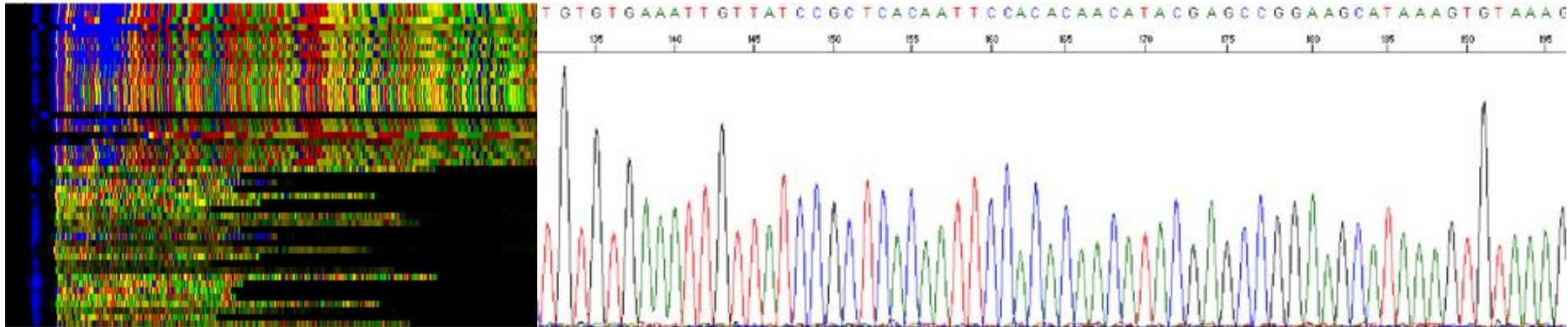
- The Human Genome Project (HGP) was initiated in the 1990s with the goal of determining the sequence of the entire human genome.
 - What genes were present?
 - Where they were located?
 - What were the sequences of the genes and the intervening DNA (non-coding DNA)?
- This task was monumental, along the order of the US Apollo Project to place a man on the Moon.
- The HGP scientists and contractors developed new technologies to sequence DNA that were automated and less expensive.

DNA Sequencing

- In one basic method to sequence DNA, you place all of the enzymes and nucleotides (A, G, C and T) necessary to copy DNA into a test tube.
- A small percentage of the nucleotides have a fluorescent dye attached to them (a different color for each type).
- You then place the DNA that you want to sequence into the test tube and let it incubate for a while.
- During the incubation process, the sample DNA gets copied over and over again (PCR reaction, mimicking real-life DNA replication). For any given copy, the copying stops when a fluorescent nucleotide gets placed into it.
- So, at the end of the incubation process, you have many fragments of the original DNA of varying sizes and ending in one of the fluorescent nucleotides.

Animation: <http://www.dnai.org/b/index.html> (go to Techniques, then Sorting and sequencing)

Sanger DNA Sequencing



The four bases are detected using different fluorescent labels. These are detected and represented as 'peaks' of different colors that can then be interpreted to determine the base sequence.

2 What can your spit tell you about your DNA?

2.a The use of saliva

- People spit for a variety of reasons. We've all employed the technique to remove a hair or some other distasteful object from our mouths. People who chew tobacco do it for obvious reasons. Ball players do it because they're nervous, bored or looking to showcase their masculinity. And people in many different cultures spit on their enemies to show disdain.
- Thanks to a phenomenon known as **direct-to-consumer genetic testing** or **at-home genetic testing**, people are spitting today for a much more productive (and perhaps more sophisticated) reason -- to get a glimpse of their own DNA.

From saliva to DNA

- Your saliva contains a veritable mother lode of biological material from which your genetic blueprint can be determined.
- For example, a mouthful of spit contains hundreds of complex protein molecules – enzymes -- that aid in the digestion of food.
- Swirling around with those enzymes are cells sloughed off from the inside of your cheek.
- Inside each of those cells lies a nucleus, and inside each nucleus, chromosomes, which themselves are made up of DNA

From saliva to DNA

- Of course, you can't look at your own spit and see sloughed-off cells, the DNA they contain or the genetic information coded in the long chain of base pairs.
- You need special equipment and scientists who know how to use it.
- You also need trained counselors who can help you interpret the data once you get it back.
- That's where companies like 23andMe, deCODEme and Navigenics come in. They give you the tools, resources and infrastructure necessary to learn more about what makes you tick at a cellular level. They each do it slightly differently, and they each reveal different aspects of your DNA profile.

Types of Genetic Tests

- Genetic tests analyze DNA present in blood and other tissue to find genetic disorders -- diseases linked to specific gene variations or mutations.
- About 900 such tests exist, ranging from more invasive procedures that require a trip to the hospital to the new generation of at-home tests that demand nothing more than spitting into a sterile, mini-sized spittoon.
- For example, **prenatal testing** may involve sampling and testing the DNA of a fetus.

One common test under this umbrella is **amniocentesis**, which requires a physician to insert a needle into the water-filled sac surrounding the fetus to withdraw a small amount of fluid. In a lab, workers culture fetal cells from the amniotic fluid to obtain a sufficient quantity of DNA. Then they analyze the DNA for chromosome abnormalities that can lead to diseases or conditions such as Down syndrome and spina bifida.

- Another approach to genetic testing is **gene sequencing**, which identifies all of the building blocks, or nucleotides, of a specific gene.
Once a person's gene has been sequenced, doctors can compare the gene against all known variations to see if it is normal or defective.
For example, inherited alterations in the genes called BRCA1 and BRCA2 (short for "breast cancer 1" and "breast cancer 2") are associated with many cases of breast cancer.
- Next up is **single nucleotide polymorphism (SNP) testing**.
Together, these nucleotides can combine in nearly infinite ways to account for much of the genetic variation we see within and between species.
Interestingly, the sequence of nucleotides in any two people is more than 99 percent identical [source: 23andMe].
Only a few nucleotides separate you from a complete stranger.
These variations are called **single nucleotide polymorphisms**, or **SNPs** (pronounced "snips").

Types of Genetic Tests

- To run a SNP test, scientists embed a subject's DNA into a small silicon chip containing reference DNA from both healthy individuals and individuals with certain diseases.
- By analyzing how the SNPs from the subject's DNA match up with SNPs from the reference DNA, the scientists can determine if the subject might be predisposed to certain diseases or disorders.
- SNP testing is the technique used by almost all at-home genetic testing companies.
- It doesn't, however, provide absolute, undisputed results!!!

From Spit to SNP: The Basic Process

- Visit the Web site of your preferred service provider. Three popular services are 23andMe, Navigenics and deCODEme.
- Next, open an account and order a test. Prices can range from \$100 to \$2,500, depending on the package you select.
- After your order is processed, the company mails a kit to you that includes any necessary equipment.
- Now comes the fun part. Using the supplied cup or tube, start collecting your spit. About 30 milliliters (2 tablespoons) of saliva are required to get a sufficient number of cheek cells. The deCODEme service actually uses a buccal DNA collector, which is a stick with rough paper on one end. You rub the paper on the inside of your cheek to collect the cells.
- Seal up your sample and place it in the conveniently provided preaddressed envelope.
- Mail it and wait patiently...

From Spit to SNP: The Basic Process

- The lab extracts DNA from your cheek cells and conducts SNP testing to see if you have any markers for certain diseases or disorders.
- When your results are ready, usually in about eight to 10 weeks, they're uploaded to your account and you're alerted by e-mail that the data is ready to be reviewed.
- What happens next depends on the service provider. Navigenics makes genetic counselors available to help you understand and interpret the data.

Can you handle the truth?

Your estimated lifetime risk

[Print this page](#)

Click anywhere on the colored boxes below to access in-depth information about each health condition, your genetic predispositions, what you can do, your specific genetic markers, and much more.

| 0 - 1% | >1 - 10% | >10 - 25% | >25 - 50% | >50 - 100% |
|--|---|--|--|-----------------------------------|
| Brain aneurysm You: 0.91% Avg: 0.90% | Alzheimer's disease You: 10% Avg: 17% | Heart attack You: 22% Avg: 25% | Osteoarthritis You: 47% Avg: 47% | You have no results in this range |
| Rheumatoid arthritis You: 0.88% Avg: 3.3% | Lung cancer You: 9% Avg: 6% | Breast cancer You: 14% Avg: 13% | Obesity You: 38% Avg: 32% | |
| Sarcoidosis You: 0.55% Avg: 1.0% | Psoriasis You: 6% Avg: 4.0% | | Diabetes, type 2 You: 37% Avg: 30% | |
| Macular degeneration You: 0.44% Avg: 3.1% | Colon cancer You: 3.6% Avg: 5% | | Atrial fibrillation You: 29% Avg: 23% | |
| Multiple sclerosis You: 0.28% Avg: 0.27% | Deep vein thrombosis You: 2.9% Avg: 3.6% | | | |

Key to your results

Condition name

Your results
Population Average

[Why orange & gray boxes?](#)
Video: [Understanding your results](#)
Tutorial: [Review the tutorial](#)
More: [How we estimate your risk](#)

Your genetic counselor

Counselors are available weekdays from 9am to 5pm PST, or you can schedule another time convenient for you.

Call (866) 522-1585

International:
+1 (650) 585-7743

[Sharing results with your doctor](#)


2.b Genetic markers

What, exactly, do a few milliliters of spit tell you?

- The most important thing you'll learn is what kind of genetic markers you carry.
- A **genetic marker** is any alteration in your DNA that may indicate an increased risk of developing a specific disease or disorder.
- Because SNPs are, by their very definition, variations in DNA, they can be used as flags or markers for nearby DNA that affects your health (more about this later – genomewide association studies).

Identifying Genetic Markers ©2009 HowStuffWorks

| Service Provider: | 23andMe | deCODEme | Navigenics |
|------------------------|---------|----------|------------|
| Arthritis | ✱ | ✱ | ✱ |
| Asthma | ✱ | ✱ | |
| Bipolar/Depression | ✱ | | |
| Cardiovascular Disease | ✱ | ✱ | ✱ |
| Multiple Sclerosis | ✱ | ✱ | ✱ |
| Osteoporosis | ✱ | | ✱ |
| Parkinson's Disease | | | |
| Schizophrenia | ✱ | | |
| Thrombosis | ✱ | ✱ | |
| Type 1/2 Diabetes | ✱ | ✱ | ✱ |



You can then use information about “increased risk for osteoporosis” to take a more proactive role in your own health care.

- You might decide to take supplements to ensure you're getting enough calcium and vitamin D.
- You might also engage in regular weight-bearing exercise and opt to have a bone density test to determine your risk for future fracture.

Question remains: how accurate / reliable are the “predictions”?

What, exactly, do a few milliliters of spit tell you?

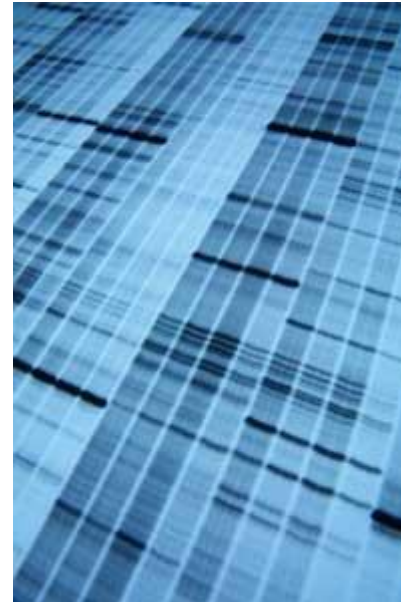
- Better health is not the only thing you “may” get out of your spit.
- You can also trace your ancestral roots.
 - This is possible because closely related individuals have more similarities in their DNA.
 - By comparing your genetic information to that of people from around the world, you can fill out a comprehensive family tree, tracing your lineage through either your mother or your father.

3 What is the human epigenome?


3.a The human epigenome project

- At a genetic level, human beings are programmed to survive.
- Deep down in our cells, in the coiled coding of our DNA, we carry all the information our bodies need to see us through this life and ensure our genetic material carries on to the next generation.
- We don't have to struggle that much anymore to carry out the necessities.
- So in our spare time, we've thrown our brains at a range of other problems: How can we secure our food supply? How can we fly through the air? How can we teach a dog to shake hands with us?

- The Human Genome Project set out to accomplish some more intimidating goals:
 - to identify human DNA's 20,000 to 25,000 genes and
 - to determine the sequences of the 3 billion chemical base pairs in DNA.
- In 2003 /2005, after 13/15 years of research, researchers completed this genomic map. Today, the project's scientists continue to analyze the stored data -- a job that will keep them busy for years to come.
- However, even with a completed genomic map, many questions remain: It's one thing to know the human genome, but another to know what factors dictate how it relates to our observable characteristics or **phenotype**.



genomics.energy.gov
Human Genome Project Information • Genomic Science Program • DOE Microbial Genomics • home



Human Genome Project Information

About the HGP



Ethical, Legal, & Social Issues

Medicine

Education

Gene Gateway

Research Archive

Completed in 2003, the Human Genome Project (HGP) was a 13-year project coordinated by the U.S. Department of Energy and the National Institutes of Health. During the early years of the HGP, the Wellcome Trust (U.K.) became a major partner; additional contributions came from Japan, France, Germany, China, and others. See our [history](#) page for more information.

Project [goals](#) were to

- *identify* all the approximately 20,000-25,000 genes in human DNA,
- *determine* the sequences of the 3 billion chemical base pairs that make up human DNA,
- *store* this information in databases,
- *improve* tools for data analysis,
- *transfer* related technologies to the private sector, and
- *address* the ethical, legal, and social issues (ELSI) that may arise from the project.

Though the HGP is finished, analyses of the data will continue for many years. Follow this [ongoing research](#) on our [Milestones](#) page. An important feature of the HGP project was the federal government's long-standing dedication to the [transfer of technology to the private sector](#). By licensing technologies to private companies and awarding grants for innovative research, the project catalyzed the multibillion-dollar U.S. biotechnology industry and fostered the development of new medical applications.

OF INTEREST

Battelle Report: \$3.8 Billion Investment in HGP Drove \$796 Billion in Economic Impact, Created Jobs, and Launched Genomic Revolution ([May 2011](#));
See [Response From Nature](#) ([May 11, 2011](#))

Gene Patent Ruling Raises Questions for Industry ([Nov. 2, 2010](#), NYT)

New Gene Tests Aimed at Reducing Colon Cancer ([Oct. 28, 2010](#), NYT)

The Once and Future Genome ([June 25, 2010](#), Seed Magazine)

Biology 2.0 Special Report on the Human Genome ([June 17, 2010](#), The Economist)


The Genome at 10: Two-part article ([June 12](#) and [June 14, 2010](#), NYT)

Human Genome at 10: 5 Breakthroughs, 5 Predictions ([Mar. 31, 2010](#), National Geographic)

Disease Cause is Pinpointed with Gene ([Mar. 10, 2010](#), NYT)

Cost of Decoding a Genome is Lowered ([Aug. 10, 2009](#), NYT)

What's next?




DOE GENOMIC SCIENCE
SYSTEMS BIOLOGY FOR ENERGY AND ENVIRONMENT
OFFICE OF SCIENCE
U.S. DEPARTMENT OF ENERGY

A systems biology program using genomic data and high-throughput technologies for studying the proteins encoded by microbial and plant genomes to develop a predictive understanding of the biological systems relevant to solving energy and environmental challenges including bioenergy production, environmental remediation, and carbon cycling.
[genomicscience.energy.gov](#)

Human Genome Project

- [FAQs](#)
- [What is the Human Genome Project?](#)
- [HGP Goals](#)
- [HGP History](#)
- [HGP Timelines](#)
- [HGP Budget](#)
- [Landmark Papers](#)



Human Genome Project Information Website Contents

Research

- [HGP Funding](#)
- [U.S. & International Research Sites](#)

Medicine & the New Genetics

- [Home](#)
- [Gene Testing](#)
- [Pharmacogenomics](#)
- [Gene Therapy](#)
- [Genetic Counseling](#)
- [CME](#)
- [Genetic Disease Information --pronto!](#)

(http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml)

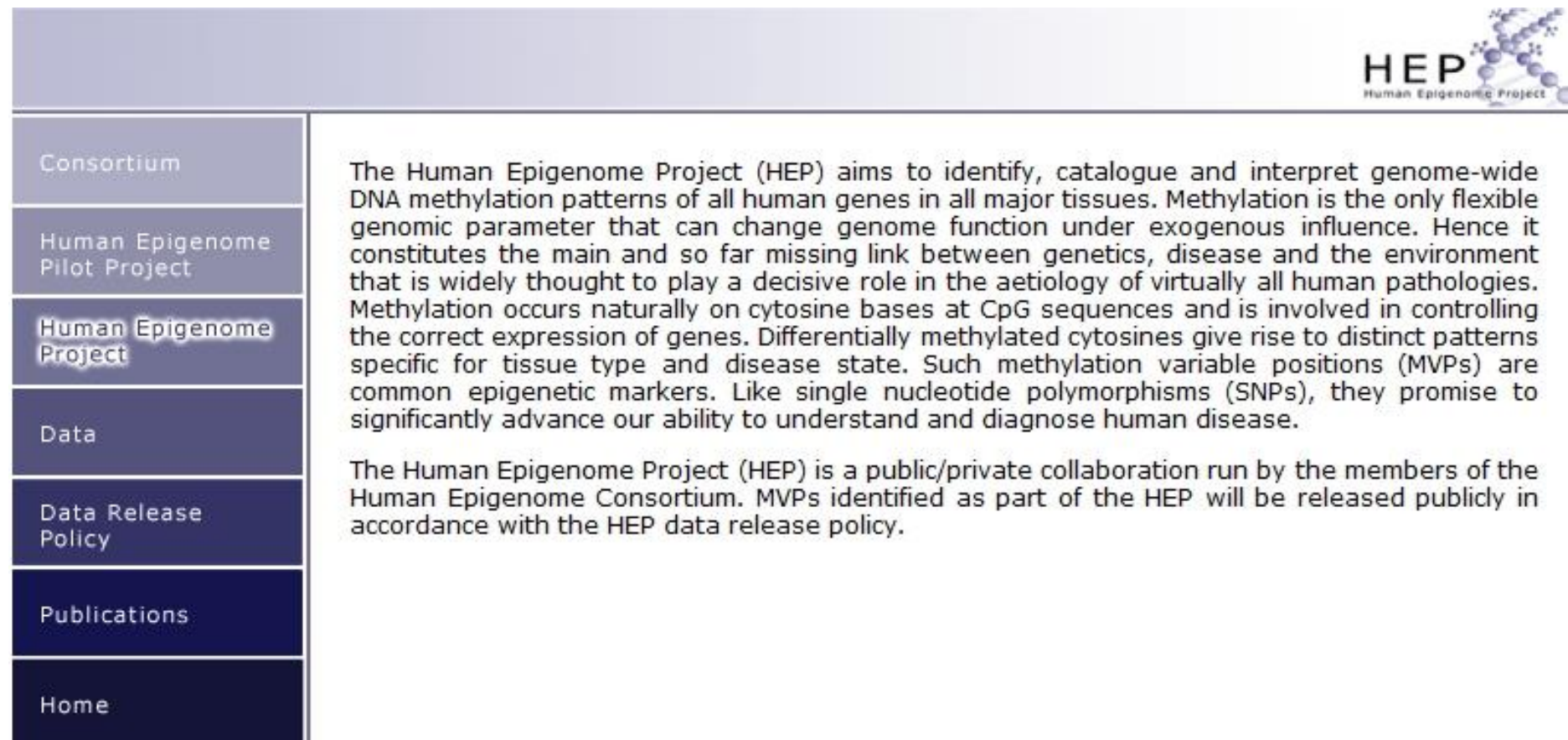
K Van Steen

95

3.b Mapping the human epigenome

- Thinking of our genes as a code that translates into a finished human being, much like a coded manuscript would translate into a readable text, imagine what that text might look like if you went in and covered up various words and phrases so they couldn't be translated.
- The finished text might be better because of this editing, but it could also be worse or even unreadable. It all depends on what words were kept out of the final copy.
- This is where **epigenetics** comes into play.
- The word literally means "above the genome" and relates to the changes that occur between the genome and the phenotype. Epigenetic changes don't alter the genes, but they do affect the way they're "expressed".

- There are several different kinds of epigenetic changes, but the one we understand the best is **methylation**.
- This process involves carbon and hydrogen bundles called methyl groups, which bind to the DNA and essentially cover up genes so they can't activate, much like the covered-up phrases in our coded manuscript.
- Some of those inactive genes could cause disease. In fact, an estimated 50 percent of the reasons for a given disease can be attributed to genetic factors.
- Others parts of the genome, such as tumor-suppressing genes, help to prevent cancer. Epigenetic changes can alter the balance, though.
- These changes can occur due to several different environmental causes, from the contents of our diet to how stressful our childhood was.



The Human Epigenome Project (HEP) aims to identify, catalogue and interpret genome-wide DNA methylation patterns of all human genes in all major tissues. Methylation is the only flexible genomic parameter that can change genome function under exogenous influence. Hence it constitutes the main and so far missing link between genetics, disease and the environment that is widely thought to play a decisive role in the aetiology of virtually all human pathologies. Methylation occurs naturally on cytosine bases at CpG sequences and is involved in controlling the correct expression of genes. Differentially methylated cytosines give rise to distinct patterns specific for tissue type and disease state. Such methylation variable positions (MVPs) are common epigenetic markers. Like single nucleotide polymorphisms (SNPs), they promise to significantly advance our ability to understand and diagnose human disease.

The Human Epigenome Project (HEP) is a public/private collaboration run by the members of the Human Epigenome Consortium. MVPs identified as part of the HEP will be released publicly in accordance with the HEP data release policy.

HEP
Human Epigenome Project

- Consortium
- Human Epigenome Pilot Project
- Human Epigenome Project
- Data
- Data Release Policy
- Publications
- Home

(<http://www.epigenome.org/index.php?page=project>)

A note aside: Epigenetic memory

- Some studies have shown that in families where there was a severe food shortage in the grandparents' generation, the children and grandchildren have a greater risk of cardiovascular disease and diabetes, which could be explained by **epigenetic memory**.
- Epigenetic memory comes in various guises, but one important form involves histones (recall: the proteins around which DNA is wrapped).
- Particular chemical modifications can be attached to histones and these modifications can then affect the expression of nearby genes, turning them on or off.
- Interestingly, these modifications can be inherited by daughter cells, when the cells divide, and if they occur in the cells that form gametes (e.g. sperm in mammals or pollen in plants) then they can also pass on to offspring.

References:

- Ziegler A and König I. *A Statistical approach to genetic epidemiology*, 2006, Wiley.
(Chapter 1, Sections 2.3.1; 3.1, 3.2.2; 5.1, 5.2.1-5.2.3)
- Burton P, Tobin M and Hopper J. Key concepts in genetic epidemiology. *The Lancet*, 2005
- Clayton D. Introduction to genetics (course slides Bristol 2003)
- URLs:
 - <http://www.rothamsted.ac.uk/notebook/courses/guide/>
 - <http://science.howstuffworks.com/>
 - <http://www.genome.gov/Education/>
 - <http://nitro.biosci.arizona.edu/courses/EEB320-2005/>
 - <http://atlasgeneticsoncology.org/GeneticFr.html>
 - <http://www.worthpublishers.com/lehninger3d/index2.html>
 - <http://www.dorak.info/evolution/glossary.html>
 - <http://www.sciencemag.org/content/vol291/issue5507/>
 - http://www.roche.com/research_and_development/r_d_overview/education.htm

Background reading / supporting document to this chapter:

- Schlötterer 2004. The evolution of molecular markers – just a matter of fashion? *Nature Reviews Genetics*, 5:63-70