INTRODUCTION TO GENES AND GENOMES

The word ‘gene’ was coined by W. Johannsen in 1909, but the modern concept of the gene originated with Gregor Mendel, who in the 1860s studied the inheritance of characteristics that differed sharply and unambiguously among true-breeding varieties of garden peas. Mendel in his experiments discovered that traits are governed by paired ‘factors’ that we now call genes.

What is a gene?

Genes are units of inheritance. Genes are parts of **DNA** and carry hereditary information passed from parents to off springs. A gene is a unit of inheritance which codes for the development of a particular characteristic in an organism. It is a sequence of **nucleotides** that provides cells with instruction for the synthesis of a specific protein or a particular type of RNA. The gene is the basic physical and functional unit of heredity. It consists of a specific sequence of nucleotides at a given position on a given chromosome that codes for a specific protein (or, in some cases, an RNA molecule). A gene can be approx. 1000 – 4000 nucleotides long. DNA is the cellular molecule that contains instructions for the performance of all cell functions. When a cell divides, its DNA is copied and passed from one cell generation to the next generation. **DNA** is organized into **chromosomes** and found within the nucleus of our cells. It contains programmed instructions for cellular activities. When organisms produce offspring, these instructions are passed down through DNA. DNA commonly exists as a double stranded molecule with a twisted double helix shape.

Genes determine traits, or characteristics, such as eye, skin, or hair color, plant height, size of fruit of all sexually-reproducing species. These are inherited through the DNA contained in the cells of your parents. Genes also control the proteins produced in a cell and influence how tissues and organs appear through the microscope and with the naked eye. They influence cell metabolism, behavioral and cognitive abilities such intelligence and susceptibility to certain types of genetic diseases.

Each gene consists of two alleles while some genes contain many alleles: one comes from the mother and one from the father. An allele is a particular form of a gene occupying the same locus on a chromosome as any alternative allele of the same gene. Different alleles of a gene code for the development of particular variations in the characteristic coded for by the gene. A diploid organism will always contain two alleles for a specific gene because they contain two of every chromosome. A haploid gamete (or haploid organism) will only contain one allele for a specific gene because only one of each chromosome will be present.

Some alleles are dominant which means they ultimately determine the expression of a trait. Other alleles are recessive and are much less likely to be expressed. When a dominant allele is paired with a recessive allele, the dominant allele determines the characteristic. When these traits or characteristics are visibly expressed, they are known as phenotypes. The genetic code behind a trait is known as the genotype. Some traits are controlled by a single genes or multiple protein producing genes that interact in complex ways.

When there are two copies of the same allele, for example AA or aa, it is said to be **homozygous** for that trait. If it has one copy of two different alleles, for example Aa, it is **heterozygous.**

Dominant allele, when present masks the expression of another allele of the same gene. A dominant character is the characteristic which develops because of the presence and expression of the dominant allele.

Recessive allele, if present, has its expression masked or blocked by the presence of a dominant allele of the same gene. The recessive character is the characteristic which develops because of the presence and expression of the recessive allele. This can only happen in the absence of a dominant allele for the gene.

A dominant allele of a gene is normally shown by a capital letter, for example A. The recessive allele is shown by a small (but the same) letter, example, a . If the gene has many alleles it can be shown as A_1, A_2 , A_3….. A_n, a_1, a_2, a_3 ….. a_n

Codominance exits when both alleles of a gene are present and are both expressed. The alleles are not dominant or recessive. The recessive and dominant traits appear together in the phenotype of hybrid organism example a cross between red x white ---> red & white spotted

R = allele for red flowers W = allele for white flowers
red x white ---> red & white spotted
RR x WW---> 100% RW

Incomplete dominance is shown in a cross between organisms with two different phenotypes producing offspring with a third phenotype that is a mixture of the parental traits.

Red Flower x White Flower ---> Pink Flower

To understand the functions of the genes properly we have to understand why and how DNA is organized into chromosomes and how DNA is replicated.

Chromosomes are paired long chains within a cell nucleus that are composed of genes (about 20,000 genes per chromosome pair), which are made up of the chemical substance called DNA. Genes on the chromosomes are made of segments of DNA which contain chemically coded messages resulting in the characteristics of an organism including humans.

Nucleic acids are molecules that permit organisms to transfer genetic information from one generation to the next. There are biopolymers or large biomolecules of life and are composed of monomers. Their function is to create and encode and then store information in the nucleus of every living cell of every life-form organism on Earth. They help to transmit and express that information inside and outside the cell nucleus to the interior operations of the cell and ultimately to the next generation of each living organism.

TYPES OF NUCLEIC ACIDS

A) deoxyribonucleic acid (better known as DNA) and

B) ribonucleic acid (better known as RNA).

The building blocks of DNA is the nucleotide, which is made up of three components:

phosphate molecule,
5-carbon sugar (pentose sugar),
and nitrogenous bases (adenine (A), guanine (G), Thymine and cytosine (C) or Uracil in RNA).

Adenine and Guanine are purines while Cytosine, Thymine and Uracil are pyrimidines.
If the sugar is a simple ribose, the polymer is RNA (ribonucleic acid), if the sugar is derived from ribose as deoxyribose, the polymer is DNA (deoxyribonucleic acid). Watson and Crick in 1953 discovered that nucleotides joined together to form long strands of DNA and each molecule consists of two strands that join together and wrapped around each other to form a double helix. The double helix of a DNA resembles a spiral staircase. A strand of DNA is a string of nucleotides held together by phosphodiester bonds that connect the sugar of adjacent nucleotides. A nucleotide containing A can be connected to a nucleotide containing a C, T, G or an A. The proportion of As and Ts are equivalent as are the proportions of G and C in the DNA of an organism known as Char A strand of nucleotides has a polaritiy a 5’ end and 3’ end. The polarity refers to the carbons of the deoxyribose sugar. At the 5’ end the strand the phosphate at carbon 5 is not bonded to another nucleotide and carbon 3 is involved in a phosphodiester bond. At the 3’ end, the phosphate at carbon 5 is bonded to another nucleotide but carbon 3 is not joined to another nucleotide. The two strands of DNA molecules are joined together by hydrogen bond between complementary base pairs in opposite strands. Adenine pairs with thymine and cytosine pairs with guanine. The base pairs are anti-parallel because the polarity of each strand is reversed relative to each other. The structure of DNA allows it to perform the following functions.

a) It can be copied or replicated because each strand can act as a template for the generation of the complementary DNA.
b) It can store information in the linear sequence of the nucleotides along each strand

Chromosome structure

DNA is to some extent loosely arranged and fully compacted into tightly coiled chromosomes. Before cell division, DNA is the nucleus exists in a combination with DNA- binding proteins called histones to form strings called chromatin. During cell division the chromatin is coiled into tight fibers which wrap each other to for, a highly coiled and tightly condensed package of DNA and histone protein.

The size of the chromosomes varies depending on the specie. Bacteria has a single circular chromosomes that have a few thousand genes. Human beings have two pairs or sets of chromosomes each pair from mother and father. Chromosome pairs are called **homologous** pairs or **homologues.** In humans chromosomes **1 to 22 are autosomes** while the 23rd pair are called sex chromosomes which are X and Y.

The sex gamete (egg and sperm) contain a single set of 23 chromosome called haploid number (n). Somatic cells (cells of other parts of the body) have two set of chromosomes called diploid (2n).

Therefore human somatic cell has 46 chromosomes. In the females you have 22 pairs of autosomes and two X chromosome while in males you have 22 pairs of autosomes and X and Y chromosomes.

Sex chromosomes contain genes that influence sex traits and development of reproductive organs.

In eukaryotes chromosome is made up of two thin rod like structures of DNA called **sister chromatid.** They are exact replica of each other copied during synthesis. The two sister chromatids are separated during cell division so that each newly form cell gets one.

There is a region that joins the two sister chromatid called **centromere.** The centromere demarcates the sister chromatids into short arm called **p arm** and long arm called **q arm**

Each arm of chromosome ends with a segment called **telomere.**

DNA replication

It is important that when new cells are created during cell division, the newly created cells contain equal copies of replicated DNA. Somatic cells divide through mitosis to give two daughter cell with identical copies of the DNA of the original cell. In meiosis four daughter cells are produced which can be either sperm or egg. The daughter cells are haploid containing half of the chromosomes. DNA replication occurs before either mitosis or meiosis occurs. It is a semiconservative replication. Before the beginning of the process, the complementary strands of the double helix are pulled apart in two strands. The two strands become templates for copying the 2 new strands. After the process two new double helices are formed. The process is semiconservative because each helix contains one original DNA and the newly synthesized strand. DNA replication is regulated by different proteins and it occurs differently in prokaryotes and eukaryotes.

Initiation is regulated by **DNA helicase** which is an enzyme that separates the two strands of nucleotides by unzipping them by breaking the hydrogen bond. The separated bands form a replication fork.

Single stranded binding proteins attached to each strand prevent them from base pairing and reforming a double helix which is very important.

Separation occurs at **origins of replication**. They could be many of them especially in eukaryotes which makes DNA replication fast.

Addition of RNA or primers which may be 10 to 15 nucleotides long. There are synthesized by **primase**. In eukaryotes it is done by DNa polymerase α. The primers start the process of DNA replication and serve as binding sites for DNA polymerase which is an enzyme that synthesize new DNA strand. DNA polymerase uses nucleotides present in the cell to synthesize complementary strands of DNA. It works in one direction synthesizing new strand in a 5’ to 3’ orientation and adding nucleotides to the 3’ end.

Since DNA polymerase operates only in 5’ to 3’ direction replication along it is continuous and is called the leading strand while synthesis on the other side is called lagging strand. Synthesis on the lagging side occurs in discontinuous manner because DNA polymerase will wait for the replication fork to open. Therefore the short pieces of DNA are called Okazaki fragments.

Genome

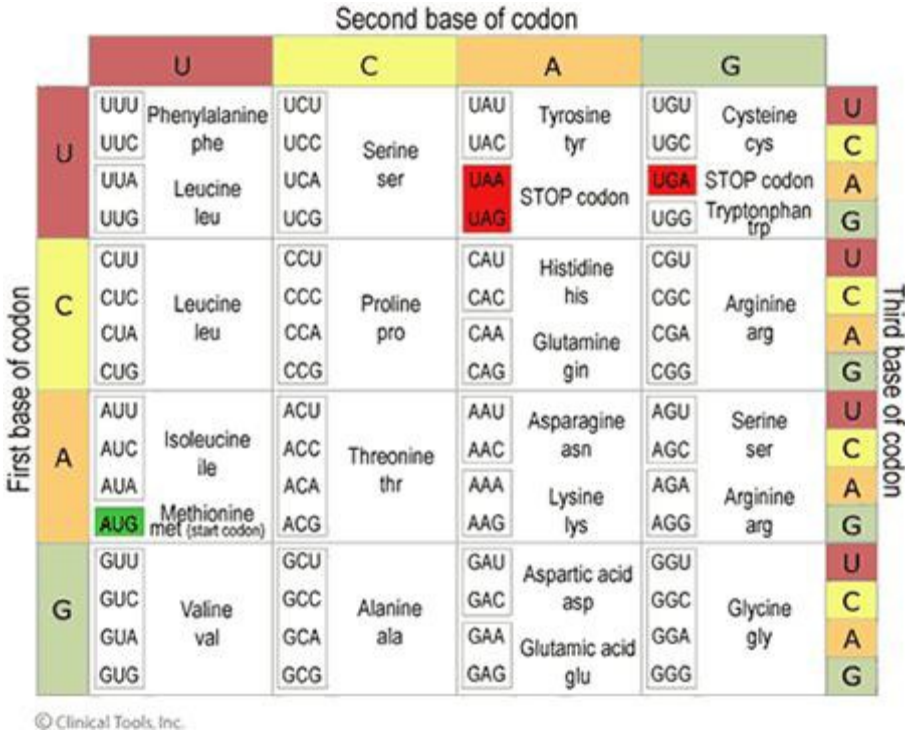
A genome is the totality of DNA in an organism. The study of genomes as a discipline is called genomics.

Genetic Code

It is the order of the bases along a single strand that constitutes the genetic code. The four-letter 'alphabet' of A, T, G and C forms 'words' of three letters called codons. Individual codons code for specific amino acids. A gene is a sequence of nucleotides along a DNA strand - with 'start' and 'stop' codons and other regulatory elements - that specifies a sequence of amino acids that are linked together to form a protein. So, for example, the codon AGC codes for the amino acid serine, and the codon ACC codes for the amino acid threonine.

Proteins which are gene products are polymers made up of chains of these formed amino acids. They constitute most of the chemical and structural machinery of the cell. There are 20 different amino acids in the proteins, and a typical protein is a string of several hundred amino acids. Genes contain the instructions for stringing together the correct sequence of amino acid for each of the thousands of proteins needed by the cell. The message contained in each gene consists of code-words for the amino acids in the protein product of the gene, written in precisely the order in which the amino acids must be connected to one another to make the protein. This message must be written in a language that uses the four ‘letters’ A, T, G and C to form words that can be read by the protein-synthesizing machinery of the cell. The structure of the genetic language is made up of three letters long called the genetic code, are three letters long. These three-letter words are called codons.

With four different letters in the alphabet, it is possible to make different codons, far more than enough to specify the names of all 20 amino acids. **Sixty four possible codons** are formed which have meanings; 61 specified the name of an amino acid and three meant ‘stop’. The three stop codons tell the protein-synthesizing machinery of the cell when the end of a protein has been reached. **The 61 codons that correspond to the names of 20 amino acids must have some synonyms**. The existence of synonyms in the code is called ‘**degeneracy’**. The synonymous codons have **the first two letters in common** and differ only in the third base. There are only two amino acids, methionine and tryptophan, that have unique codons. There is one amino acid, isoleucine, which has three codons. Leucine, serine and arginine that have six codons.



Genes consist of three types of nucleotide sequence

Coding regions, called exons, which specify a sequence of amino acids
Non-coding regions, called introns, which do not specify amino acids
Regulatory sequences, which play a role in determining when and where the protein is made (and how much is made)

Flow of genetic information

The central dogma of molecular biology describes the flow of genetic information in cells from DNA to messenger RNA (mRNA) to protein. It states that genes specify the sequence of mRNA molecules, which in turn specify the sequence of proteins. The information stored in DNA is so central to cellular function, the cell keeps the DNA protected and copies it in the form of RNA. An enzyme adds one nucleotide to the mRNA strand for every nucleotide it reads in the DNA strand.

Transcription (DNA to RNA)

Transcription is the process of creating a complementary RNA copy of a sequence of DNA. This is achieved because both RNA and DNA are nucleic acids, which use base pairs of nucleotides as a complementary language that enzymes can convert back and forth from DNA to RNA. During transcription, a DNA sequence is read by **RNA polymerase**, which produces a complementary, antiparallel RNA strand. Unlike DNA replication, transcription results in an RNA complement that substitutes the RNA uracil (U) in all instances where the DNA thymine (T) would have

occurred. Transcription is the first step in gene expression. The stretch of DNA transcribed into an RNA molecule is called a transcript.

Translation (RNA to Protein)

Translation is the process by which the **mRNA is decoded and translated to produce a polypeptide sequence**, otherwise known as a protein. This method of synthesizing proteins is directed by the **mRNA** and accomplished with the help of a ribosome, a large complex of **ribosomal RNAs (rRNAs)** and proteins. In translation, a cell decodes the mRNA's genetic message and assembles the brand new polypeptide chain. **Transfer RNA or tRNA**, then translates the sequence of codons on the mRNA strand. The main function of tRNA is **to transfer a free amino acid** from the cytoplasm to a ribosome, where it is attached to the growing polypeptide chain. tRNAs continue to add amino acids to the growing end of the polypeptide chain until they reach a stop codon on the mRNA. The ribosome then releases the completed protein into the cell.

MUTATION

A mutations are as a result of change in nucleotide sequence of DNA. Mutations can have harmful effects. They have led to evolution of species over time. It is a major cause of genetic diversity. Gene mutations can result in production of altered proteins that functions poorly or encodes a functional protein. Mutations can be due to spontaneous events such as errors during DNA replication or environmental cause’s eg chemical mutagenes

Types of mutations

Point mutation- base pair substitution

Silent mutations –changes in the codon sequence of a gene to another codon that codes for the same amino acids.

Missense change in the codon sequence to another codon that codes for a different amino acid.

Nonsense mutations- change of a codon for an amino acid into a stop codon which causes an abnormally shortened protein when translated usually creating a non-functional protein.

Frameshifts- insertions or deletions can cause reading frames of codons to be shifted to the right of the insertions changing the proteins encoded by the mRNA.

Mutations can be inherited or acquired.