SCIENTIFIC BACKGROUND OF GENETICS

A - Basic Principles of Genetics

Le Dinh Luong
Vietnam National University

Genetics is science of genes

Since the beginning of human history, people have wondered how traits are inherited from one generation to the next. Although children often look more like one parent than the other, most offspring seem to be a blend of the characteristics of both parents. Centuries of breeding of domestic plants and animals had shown that useful traits - speed in horses, strength in oxen, and larger fruits in crops - can be accentuated by controlled mating. However, there was no scientific way to predict the outcome of a cross between two particular parents.

It wasn't until 1865 that an Augustinian monk named Gregor Mendel found that individual traits are determined by discrete "factors," later known as ***genes***, which are inherited from the parents. His rigorous approach transformed agricultural breeding from an art to a science.

However, the science of genetics really began with the rediscovery of Gregor Mendel's work at the turn of the 20th century, and the next 40 years or so saw the elucidation of the principles of inheritance and genetic mapping. Microbial genetics emerged in the mid 1940s, and the role of DNA as the genetic material was firmly established. During this period great advances were made in understanding the mechanisms of gene transfer between bacteria, and a broad knowledge base was established from which later developments would emerge.

The discovery of the structure of DNA by James Watson and Francis Crick in 1953 provided the stimulus for the development of genetics at the molecular level, and the next few years saw a period of intense activity and excitement as the main features of the gene and its expression were determined. This work culminated with the establishment of the complete genetic code in 1966 - the stage was now set for the appearance of the new genetics.

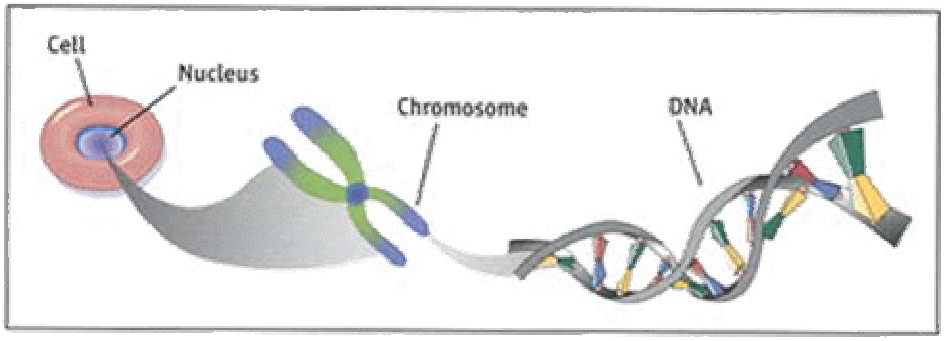
Since 1865 year up to now one can say the history of genetics development is the development of the human knowledge and understanding of genes.

Genes are mostly located in cell nucleus on chromosomes

All living organisms are composed of cells. Many of the chemical reactions of an organism, its metabolism, take place inside of cells. The genetic information required for the maintenance of existing cells and the production of new cells is stored within the membrane-bound nucleus in eukaryotic cells. This genetic information passes from one generation to the next.

The nucleus, which contains the genetic information (DNA), is the control center of the cell. DNA in the nucleus is packaged into chromosomes. DNA replication and RNA transcription of DNA occur in the nucleus. Transcription is the

first step in the expression of genetic information and is the major metabolic activity of the nucleus.



A gene, a unit of hereditary information, is a stretch of DNA sequence, encoding information in a four-letter language in which each letter represents one of the nucleotide bases. Much of the information stored in stretches of DNA sequence is subsequently expressed as another class of biopolymers, the proteins.

Work on cytology in the late 1800s had shown that each living thing has a characteristic set of chromosomes in the nucleus of each cell. During the same period, biochemical studies indicated that the nuclear materials that make up the chromosomes are composed of DNA and proteins. In the first four decades of the 20th century, many scientists believed that protein carried the genetic code, and DNA was merely a supporting "scaffold."

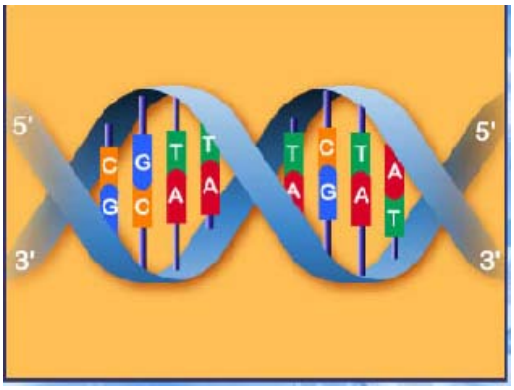
Just the opposite proved to be true. Work by Avery and Hershey, in the 1940s and 1950s, proved that DNA is the genetic molecule. Work done in the 1960s and 1970s showed that each chromosome is essentially a package for one very long, continuous strand of the DNA. In higher organisms, structural proteins, some of which are histones, provide a scaffold upon which DNA is built into a compact chromosome. The DNA strand is wound around histone cores, which, in turn, are looped and fixed to specific regions of the chromosome.

Genes are made of DNA or RNA

Structure of DNA

Deoxyribonucleic acid (DNA) is composed of building blocks called nucleotides consisting of a deoxyribose sugar, a phosphate group, and one of four nitrogen bases - adenine (A), thymine (T), guanine (G), and cytosine (C). Phosphates and sugars of adjacent nucleotides link to form a long polymer. It was showed that the ratios of A - to T and G – to - C are constant in all living things. X-ray crystallography provided the final clue that the DNA molecule is a double helix, shaped like a twisted ladder.

In 1953, the race to determine how these pieces fit together in a three-dimensional structure was won by James Watson and Francis Crick at the Cavendish Laboratory in Cambridge, England. They showed that alternating deoxyribose and phosphate molecules form the twisted uprights of the DNA ladder. The rungs of the ladder are formed by complementary pairs of nitrogen bases - A always paired with T and G always paired with C.



Base pairs bond the double helix together. The "beginning" of a strand of a DNA molecule is defined as 5'. The "end" of the strand of A DNA molecule is defined as 3'. The 5' and 3' terms refer to the position of the nucleotide base, relative to the sugar molecule in the DNA backbone. The two strands in a double helix are oriented in opposite directions.

Each chromosome is composed of a single DNA molecule. Our DNA contains greater than 3 billion base pairs--an enormous amount by any measure. All of this information must be organized in such a manner that it can be packaged inside the nucleus of the cell. To accomplish this, DNA is complexed with histones to form chromatin. Histones are special proteins that the DNA molecule coils around to become more condensed. The chromatin then becomes coiled upon itself, which ultimately forms chromosomes.

When one cell divides into two daughter cells, the DNA, all 46 chromosomes, must be replicated. The specificity of base pairing between A/T and C/G is essential for the synthesis of new DNA strands that are identical to the parental DNA. Each strand of DNA serves as a template for DNA synthesis. Synthesis occurs by adding bases that exactly mirror the template strand. So, as each strand is copied, two sets of DNA are made that are identical to the original two strands. The order of nucleotide bases along a DNA strand is known as the ***sequence***.

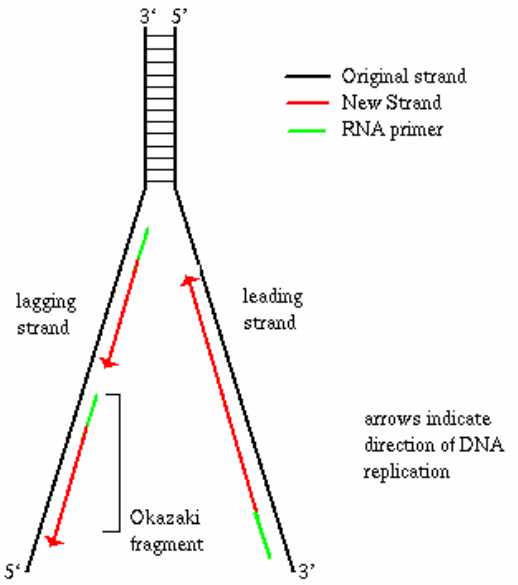
If a problem occurs during DNA replication, this can lead to a disruption of gene function. For example, if the wrong base is inserted during replication (a mutation) and this mistake happens to be in the middle of an important gene, it could result in a non-functional protein. Fortunately, we have evolved various mechanisms to ensure that such mutations are detected, repaired, and not propagated. However, these mechanisms sometimes fail and uncorrected mutations will occur. If the resulting alteration in gene function, through its interplay with the environment, sufficiently disrupts metabolism or structure, clinical disease can result.

Some viruses store genetic information in RNA

DNA was believed to be the sole medium for genetic information storage. Furthermore, Watson and Crick's central dogma assumed that information flowed "one-way" from DNA to RNA to protein. So it came as a surprise when in 1971, it was discovered that some viruses shift their genetic information from RNA to DNA.

Even so, these viruses ultimately make proteins in the same way as higher organisms. During infection, the RNA code is first transcribed "back" to DNA - then to RNA to protein, according to the accepted scheme. The initial conversion of RNA to DNA - going in reverse of the central dogma - is called reverse transcription, and viruses that use this mechanism are classified as retroviruses. A specialized polymerase, reverse transcriptase, uses the RNA as a template to synthesize a complementary and double stranded DNA molecule.

Genes can make themselves



As genes are made of DNA, they can make themselves when DNA is replicated. The specificity of base pairing between A/T and C/G helps explain how DNA is replicated prior to cell division. Enzymes unzip the DNA by breaking the hydrogen bonds between the base pairs. The unpaired bases are now free to bind with other nucleotides with the appropriate complementary bases. The enzyme primase begins the process by synthesizing short primers of RNA nucleotides complementary to the unpaired DNA. DNA polymerase now attaches DNA nucleotides to one end of the growing complementary strand of nucleotides. Replication proceeds continuously along one strand, called the leading strand, which is shown here on the right. The process occurs in separate short segments called Okazaki fragments next to the other, or lagging, strand on the left. This difference is due to the fact that DNA polymerase can only add new nucleotides to the 3 prime end of a nucleotide strand. A primer begins any new strand, including each Okazaki fragment. An enzyme replaces the RNA primer with DNA nucleotides. Then an enzyme called DNA ligase binds the fragments to one another.

There are now 2 DNA molecules. Each consists of an original nucleotide strand next to a new complementary strand. The two molecules are identical to each other.

Language of genes is simple and informative

Genetic information is like a language. We use letters of the alphabet to make words and then join these words together to make sentences, paragraphs and books. In the case of DNA:

- the alphabet is only 4 letters (A,T,G and C) long
- each letter represents a chemical compound called a ***base*** or ***nucleotide***
- these 4 letters are used to form the genetic words called ***codons***
- unlike a normal language all genetic words are only three letters long
- these words combine together to form sentences called ***genes***
- at the end of each sentence is a special word or full stop called a ***stop codon***
- all the sentences join together to form a book that contains all the genetic information about you called your ***genome***

Let's make some comparison between English Language and Genetic Language:

English Language	Genetic Language
• we use ***26 letters*** to make ***words***	• DNA uses ***4 molecules*** to make ***codons***
• the ***words*** can be ***any length*** we need	• the ***codons*** can only be ***3 molecules*** long
• we join words together to create ***sentences***	• the codons join together to form ***genes***
• each sentence ends with a ***fullstop***	• the gene stops at a specific ***stop codon***
• all the sentences combine to form a ***book***	• all the genes combine to form the ***genome***

The Genetic Language of DNA provides the information needed to produce ***proteins***. It is these proteins that carry out the biochemical processes (metabolism) to ensure an organism's ongoing survival. Proteins have their own language that has an "alphabet" of 20 "letters". These letters are the ***amino acids***. RNA is used to "translate" genetic language into protein language. It takes the information from a ***gene*** of the DNA strand and creates the proteins necessary for life.

Along the gene (and DNA itself) the information for the amino acids that will make up the gene is stored in three-letter words called ***codons***. Each codon specifies a particular amino acid. By "reading" this set of codons the specific protein can be generated from this chunk of genetic code. The codons on ***DNA*** code for a specific ***amino acid***. There are 20 amino acids commonly found in natural ***proteins***.

Below is a "paragraph" of gene language:

CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG ACG TCC GAA GAG TGA CCG

Altered genes are mutations

The DNA sequences from two individuals of the same species are highly similar - differing by only about one nucleotide in 1,000. A mutation is, most simply, an alteration in a DNA sequence. This change may or may not lead to a change in the protein coded by the gene. A change that has no effect on protein sequence or function is termed a ***polymorphism*** and is a part of the normal variation present in the human genome. Often, however, a change in a DNA sequence will result in the disruption of gene function that we term "Clinical Manifestations" in the ***Clinical Integration Model***. The altered protein that results from a mutation can disrupt the way a gene functions, which can lead to clinical disease. How these mutations

manifest themselves depends on each individual's unique genetic endowment and interactions with their environment.

Furthermore, the change may or may not be passed on to subsequent generations. If, as in non-familial cancer, the mutation occurs in isolated ***somatic cells***, it will not be passed on to subsequent generations. Only those mutations occurring to the DNA in the ***gametes*** (egg or sperm) will potentially be passed on to offspring. If the mutation is passed on to the offspring, they will carry this mutation in all of the cells in their body.

Following is a brief review of different types of mutations:

Base pair substitution: Replacement of one DNA base by another in the DNA sequence. Replacement of nucleotide bases can have several possible consequences.

Missense mutation: An amino acid residue in the original protein may be replaced by a different one in the mutated protein. (Figure 1)

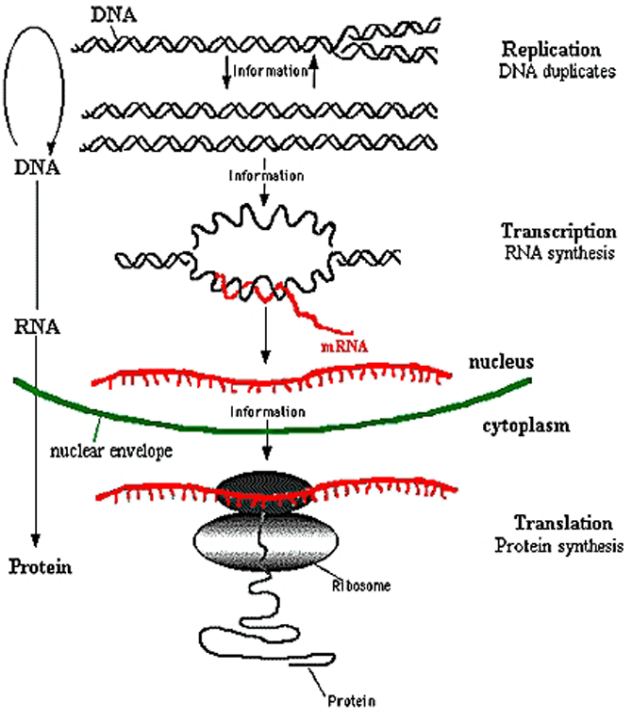
Nonsense mutation: The codon for an amino acid residue within the original protein is changed to a stop codon, which leads to a premature termination of the protein resulting into a non functional protein. (Figure 1)

Silent Mutation: The codon for an amino acid is changed, but the same amino acid is still coded for. This is possible because some amino acids are coded for by multiple codons. For example, the sequences UGC and UGU both code for Cysteine.

Frameshift mutation: A deletion or insertion of any number of bases other than a multiple of three bases has a much more profound effect. Such frameshift mutation result in a complete change in the amino acid sequence downstream from the point of mutation, instead of simply a change in the number of amino acids.

Deletions, Insertions, and Duplications: Deletions or insertions may be large or small. Large insertions and deletions in coding regions almost invariably prevent the production of useful proteins. The effect of short deletions or insertions depends on whether or not they involve multiples of three bases. If one, two, or more whole codons (three base pairs or any multiple of three) are removed or added, the consequence is the deletion or addition of a corresponding number of amino acid residues. Sometimes, an entire gene can be inserted (duplicated) or deleted. The effects of these types of mutation depend on where in the genome they occur and how many base pairs are involved.

Normal
THE BIG RED DOG RAN OUT.
Missense
THE BIG RAD DOG RAN OUT.
Nonsense
THE BIG RED.
Frameshift - deletion
THE BRE DDO GRA.
Frameshift - insertion
THE BIG RED ZDO GRA.



The Central Dogma of Molecular Biology

Inversions: This type of mutation occurs when a chromosomal section is separated from the chromosome, rotates 180 degrees, and rejoins the chromosome in an opposite orientation. This type of mutation can affect a gene at many levels. If an inversion disrupts a promoter region, the gene may not be transcribed at all. If the coding sequence is disrupted, a non-functional gene product (protein) may result. **Translocations:** This type of chromosomal aberration results when one portion of a chromosome is transferred to another chromosome. This can be a very harmful event if it leads to a subsequent gain or loss of genetic material. Additionally, when a gene from one chromosome moves to another chromosome, large changes in the ability to regulate expression of the gene may occur. Some forms of leukemia result from translocations. In these cases, various genes controlling growth of white blood cells are constantly turned on, leading to an uncontrolled proliferation of these cells and the various clinical manifestations of leukemia.

Causes of mutations: Mutations are caused by substances that disrupt the chemical structure of DNA or the sequence of its bases. Radiation, various chemicals, and chromosome rearrangements are some of the many sources of mutation.

Mutation rates: All of us are subjected to mutagenic events throughout our lifetime. Depending upon the type of mutation, the frequency ranges from 10^{-2}/cell division to 10^{-10}/cell division. Our cells have numerous mechanisms to repair and/or prevent the propagation of these mutations.

The way from genes to traits

The following is an overview of the processes involved in turning the genes coded for in your DNA into the proteins that make up your body. This is sometimes referred to as the "Central Dogma" of genetics.

- *Replication is the process by which DNA copies itself in order to be passed on to a new cell during cell division.*

- ***Transcription** is the process by which the DNA sequence of a gene is used to form an identical strand of mRNA, which will be used to guide protein synthesis.*

- ***Translation** is the process by which the mRNA sequence is used to guide construction of a protein from its constituent amino acids.*

Problems during any one of these processes can lead to a disruption of normal gene function, which can manifest itself as clinical disease. How this can occur will be discussed in the following sections.

The genes in our DNA become the proteins that compose our body through the processes of transcription and translation, with RNA being the intermediary.

Transcription

Transcription is the process whereby DNA is used as the template for the production of molecules of RNA. RNA has different forms, including messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). Each type of RNA is involved in the process of constructing a protein based on the DNA sequence of a gene.

The process of constructing RNA from DNA is carried out by an enzyme, RNA polymerase II, and is controlled through sequences in the genome termed promoters. This process requires many different proteins and is tightly regulated to ensure proper gene expression. Mutations in the proteins that are involved in replication, or mutations in the DNA promoter sequences themselves, can lead to improper expression and function of a gene. A mutation in a promoter sequence that makes it non-functional would lead to decreased expression of the gene and, therefore, decreased amounts of a protein. An example of this is a mutation in the promoter sequence for a component of hemoglobin, which leads to decreased amounts of functional hemoglobin. This condition, ß-Thalassemia, leads to severe anemia and death by the mid-20's. Transcription and the proteins regulating it are a vital part of gene function.

Transcription occurs in the cell nucleus. Once the RNA is made, it is transported out of the nucleus to the cytoplasm, the location of translation.

Translation

Translation is the process that turns a gene sequence, via a transcribed RNA molecule, into a protein. The various types of RNA play different roles in this process. mRNA provides the sequence that is translated, rRNA helps to direct the orderly translation of this sequence, and tRNA is the direct link between the sequence of bases and the amino acids that they code for. These amino acids are joined together to form proteins.

Once formed, these proteins are sorted and modified to carry out their ultimate functions. These include:

- Enzymes, such as those in the digestive system
- Structural components, such as the collagen in ligaments and tendons
- Protection, including antibodies and components of the blood clotting cascade
- Regulatory hormones, including Insulin and Growth hormone
- Movement, due to the actin and myosin in our muscles
- Transport, carried out by hemoglobin and albumin in our blood

Proteins and amino acids

All proteins are linear polymers and are made up of basic building blocks called amino acids. Translation, or protein construction, takes place in the cytoplasm. RNA codes for 20 different amino acids that are then incorporated into proteins. These 20 different amino acids contain 20 different side chains, a remarkable collection of diverse chemical groups, which allow proteins to exhibit such a great variety of structures and properties. The conformation (3-D structure) and function of a protein are determined by its amino acid composition, by the sequence in which these amino acids are strung together, and by interactions with other proteins.

Protein function

Proteins play an enormous variety of roles such as transport, storage, the structural framework of cells, antibodies, the enzymatic machinery that catalyzes biochemical reactions essential for metabolic activities, many hormones, and contractile proteins for muscle contraction and cell motility. Examples of proteins include hemoglobin, collagen, thyroid hormone, insulin, and myosin. Disease is often a manifestation of improper protein function, which can result from genetic and/or environmental influences.

Centrality of gene function

It is essential that the flow of information from genes to proteins functions properly. Each cell is endowed with the essential proteins for its proper function through selective activation of many genes. Mutations in essential genes will disrupt the life cycle of an individual cell and potentially the entire organism. Loss of control over proteins controlling cell growth results in cancer. Improper function of certain proteins in the cell membrane is the cause of cystic fibrosis. A proper understanding and management of clinical manifestations (phenotype) seen in those with a given disease requires a thorough grounding in normal and abnormal gene function.

Genes can be turned on and off

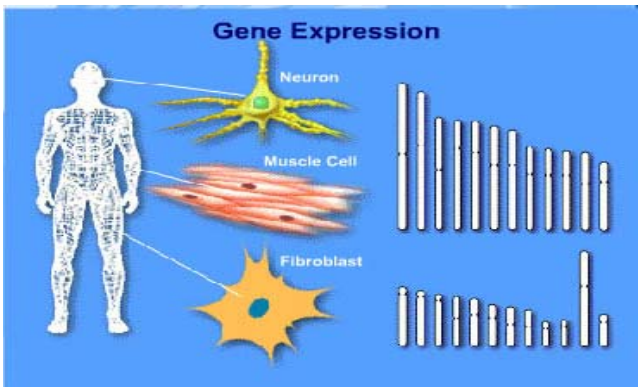
As researchers untangled the genetic code and the structure of genes in the 1950s and 60s, they began to see genes as a collection of plans, one plan for each protein. But genes do not produce their proteins all the time, suggesting that organisms can regulate gene expression. French researchers first shed light on gene regulation using bacteria.

When lactose is available, *E. coli* turn on an entire suite of genes to metabolize the sugar. Researchers tracked the events lactose initiates and found that lactose removes an inhibitor from the DNA. Removing the inhibitor turns on gene production.

The gene that produces the inhibitor is a regulatory gene. Its discovery altered perceptions of development in higher organisms. Cells not only have genetic plans for structural proteins within their DNA, they also have a genetic regulatory program for expressing those plans.

Different genes are active in different cells

All cells in the body carry the full set of genetic information but only express about 20% of the genes at any particular time. Different proteins are expressed in different cells according to the function of the cell. Gene expression is tightly controlled and regulated.



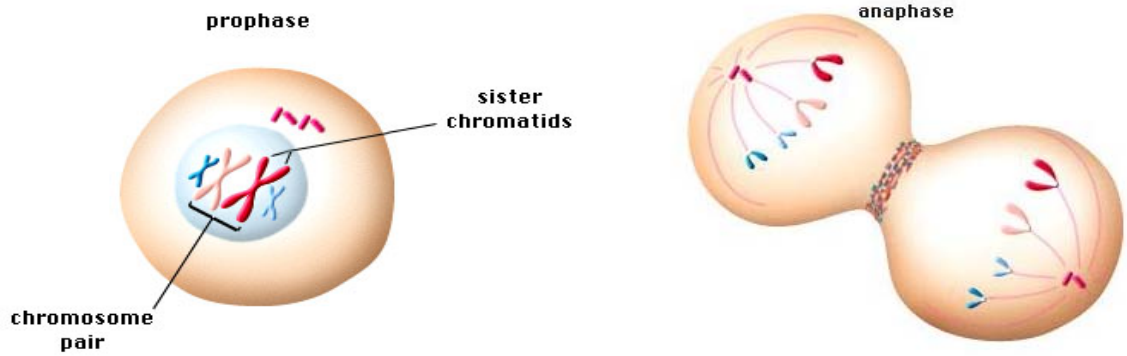
Most living organisms are composed of different kinds of cells specialized to perform different functions. A liver cell, for example, does not have the same biochemical duties as a nerve cell. Yet every cell of an organism has the same set of genetic instructions, so how can different types of cells have such different structures and biochernical functions? Since biochemical function is determined largely by specific enzymes (proteins), different sets of genes must be turned on and off in the various cell types. This is how cells differentiate.

This notion of cell-specific expression of genes is upheld by hybridization experiments that can identify the unique mRNAs in a cell type. More recently, DNA arrays and gene chips offer the opportunity to rapidly screen all gene activity of an organism. Co-expression of genes in response to external factors can thus be explored and tested.

Genes move from generation to generation mostly together with chromosomes

The inheritance of genes is based on the behavior of chromosomes, on which genes are located, and how the chromosomes are distributed during cell divisions, which mostly include mitosis and meiosis among eukaryotic organisms.

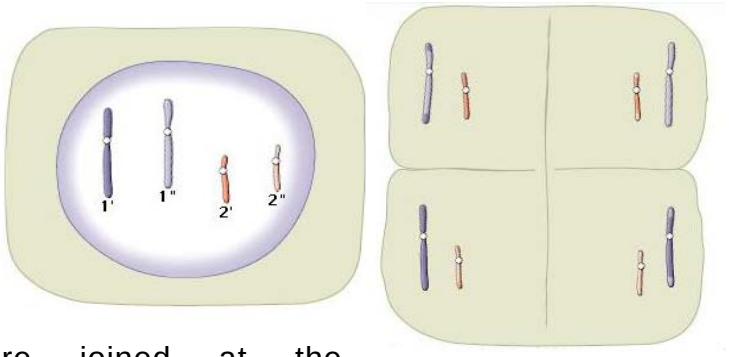
Mitosis produces genetically identical cells; meanwhile products of



meiosis are genetically distinct because of independent assortment and crossing-over.

Mitosis is the process by which the contents of the eukaryotic nucleus are separated into 2 genetically identical packages. The result is 2 cells, each with an identical set of chromosomes.

Meiosis is the process by which a diploid nucleus divides twice to produce 4 haploid nuclei. The divisions are called meiosis I and meiosis II. In the life cycles of diploid organisms meiosis precedes sexual reproduction. Among animals, the products of meiosis are gametes - eggs or sperm.



DNA is replicated prior to the start of meiosis. The identical sister chromatids are joined at the centromere as in mitosis. Unlike in mitosis, homologous chromosomes pair with one another. The result is 4 haploid cells.

Genetic information is reshuffled during meiosis, producing genetic diversity in populations. A diploid cell contains two sets of chromosomes. The maternal set was contributed by the mother and the paternal set was contributed by the father. A pair of homologous chromosomes consists of one maternal and one paternal chromosome. Homologous chromosomes carry the same genes but may have different forms or alleles of the genes. At the beginning of meiosis, homologous chromosomes pair and non-sister chromatids exchange sections of DNA through the process known as crossing-over or recombination.

The resulting chromosomes may now contain different combinations of alleles than were found in the chromosomes inherited from the parents. At the middle of meiosis I, the maternal and paternal chromosomes of one homologous pair align independently of the maternal and paternal chromosomes of the other homologous pairs. Genes that are located on different chromosomes undergo independent assortment because of the random alignment of the maternal and paternal chromosomes. Gametes produced by meiosis have different combinations of alleles as a result of both recombination and independent assortment.

Genes can be moved between species

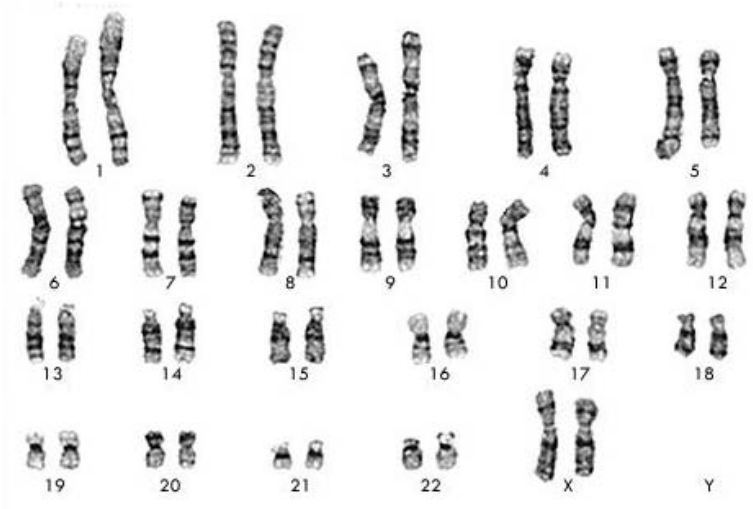
Because of the universality of the genetic code, the polymerases of one organism can accurately transcribe a gene from another organism. For example, different species of bacteria obtain antibiotic resistance genes by exchanging small chromosomes called plasmids. In the early 1970s, researchers in California used this type of gene exchange to move a "recombinant" DNA molecule between two different species. By the early 1980s, other scientists adapted the technique and spliced a human gene into E. coli to make recombinant human insulin and growth hormone.

Recombinant DNA technology - genetic engineering - has made it possible to gain insight into how genes work. In cases where it is impractical to test gene function

using animal models, genes can first be expressed in bacteria or cell cultures. Similarly, the phenotypes of gene mutations and the efficacy of drugs and other agents can be tested using recombinant systems.

A genome is an entire set of genes

Each organism has a defining set of chromosomes that contain all of its genetic information. An organism's total DNA content is known as its ***genome***. The human genome, for example, is the set of genetic information encoded in 46 chromosomes found in the nucleus of each cell. The chromosomes are organized into 23 pairs - one chromosome of each pair is inherited from the mother and one from the father. One pair of chromosomes - X and Y - determines sex; the other 22 pairs are called ***autosomes***.



So, the human genorne is made up of a set of very long DNA molecules, one corresponding to each chromosome. Arrayed along these molecules are an estimated 35,000 genes. The object of the Human Genome Project is to determine the entire nucleotide sequence of each of these DNA molecules - and the location and identity of all the genes. Sequencing the human genome has relied mainly on automated machines that sequence the DNA and computer programs that search and identify genes. A "working draft" DNA sequence of the human genome was completed in June 2000. Initial analyses of this working draft were published in February 2001.

Living organisms share common genes

All living organisms store genetic information using the same molecules - DNA and RNA. Written in the genetic code of these molecules is compelling evidence of the shared ancestry of all living things. Evolution of higher life forms requires the development of new genes to support different body plans and types of nutrition. Even so, complex organisms retain many genes that govern core metabolic functions carried over from their primitive past.

Genes are maintained over an organism's evolution; however, genes can also be exchanged or taken from other organisms. Bacteria can exchange plasmids carrying antibiotic resistance genes through conjugation, and viruses can insert their genes into host cells. Some mammalian genes have also been adopted by viruses and later passed onto other mammalian hosts. Regardless of how an organism gets and retains a gene, regions essential for the correct function of the protein are

always conserved. Some mutations can accumulate in non-essential regions; these mutations are an overall history of the evolutionary life of a gene.

Genes can be manipulated and modified with molecular tools

Progress in any scientific discipline is dependent on the availability of techniques and methods that extend the range and sophistication of experiments which may be performed. Over the last 30 years or so this has been demonstrated in spectacular fashion by the emergence of molecular genetics. This field has grown rapidly to the point where, in many laboratories around the world, it is now routine practice to isolate a specific DNA fragment from the genome of an organism, determine its base sequence, and assess its function. What is particularly striking is that this technology is readily accessible by individual scientists, without the need for large-scale equipment or resources outside the scope of a reasonably well-found research laboratory.

Although there are many diverse and complex techniques involved, the basic principles of genetic manipulation are reasonably simple. The premise on which the technology is based is that genetic information, encoded by DNA and arranged in the form of genes, is a resource which can be manipulated in various ways to achieve certain goals.

DNA extraction: Depending on the cell characteristics, DNA extraction from animal cells differs from DNA extraction from plant or prokaryotic cells. Link to **DNA extraction** protocols, **Roche Manual**, **Amersham Manual** (see also **Gentra Puregene Protocols** for technical reports on DNA extraction).

Hybridization techniques: Southern blotting, Northern blotting and in situ hybridization (including fluorescent in situ hybridization - FISH). Hybridization techniques allows picking out the gene of interest from the mixture of DNA/RNA sequences. Hybridization only occurs between single stranded and complementary nucleic acids. The level of similarity between the probe and target determines the hybridization temperature. See the overview of **blotting techniques** from the Biology Hypertextbook, an animation of **Southern blotting**, and an example of **DNA fingerprinting**.

Enzymatic modification of DNA: DNA ligase and restriction enzymes (sticky ends, blunt ends). Most restriction enzymes recognize palindromic sequences. These are short sequences which are the same on both strands when read 5' to 3' (such as he *Msp*I restriction site CCGG and that of *Eco*RI GAATTC). See the action of ***Eco*RI**.

Cloning into a vector: vectors can be a plasmid (pBR322, pUC including Blue Script), lambda (λ) bacteriophage, cosmid, PAC, BAC, YAC, expression vectors. The Ti plasmid is the most popular vector in agricultural biotechnology. Plasmids can accommodate up to 10 kb foreign DNA, phages up to 25 kb, cosmids up to 44 kb, YACs usually several hundred kb but up to 1.5 Mb. Gene cloning contributed to the following areas: identification of specific genes, genome mapping, production of recombinant proteins, and the creation of genetically modified organisms. Link to examples of **plasmids**.

Gene libraries: Genomic (restriction digestion, sonication) or cDNA libraries are made to identify a gene. See the construction of a **human genomic library**.

Polymerase Chain Reaction (PCR): Using the thermostable DNA polymerase obtained from Thermophilus aquaticus (briefly *Taq*), the PCR amplifies a desired sequences millions-fold. It requires a primer pair (18-30 nucleotides) to get the DNA polymerase started, the four nucleotides (dNTPs), a template DNA and certain chemicals including magnesium chloride (as a cofactor for *Taq* polymerase). The three steps in a cycle of the PCR - denaturation (the separation of the strands at 95^{o} C), annealing (annealing of the primer to the template at 40 - 60^{o} C), and elongation (the synthesis of new strands) - take less than two minutes. *Taq* polymerase extends primers at a rate of 2 - 4 kb/min at 72^{o} C (the optimum temperature for its activity). Each cycle consisting of these three steps is repeated 20 - 40 times to get enough of the amplified segment. Annealing temperature of each primer is calculated using its base composition. For primers less than 20 base-long: Tm = 4(G+C) + 2(A+T).

The conventional PCR is able to amplify DNA sequences up to 3 kb but the newer enzymes allow amplification of DNA fragments up to 30 kb-long. Nanogram levels of template DNA (even from a single cell) is enough to obtain amplification. The more recent '**real-time PCR**' techniques are able to detect the sequence of interest in 20 picogram of total RNA. *Taq* polymerase has a relatively high misincorporation rate. It has been genetically modified to reduce the misincorporation rate.

See an overview of **PCR**, an article on **PCR,** an animation of **PCR**, and a technical guide to **PCR**.

Different versions of PCR: Nested PCR (for increased sensitivity and specificity), reverse transcriptase (RT) PCR (starts with mRNA instead of genomic DNA), amplified fragment length polymorphism (AFLP) (replaced Southern blotting) , overlap PCR (joins two PCR products together), inverse PCR (amplifies an unknown DNA sequence flanking a region of known sequence).

Applications of PCR

1. Diagnostic use in medical genetics, medical microbiology and molecular medicine
2. HLA typing in transplantation
3. Analysis of DNA in archival material
4. Forensic analysis
5. Preparation of nucleic acid probes
6. Clone screening and mapping

DNA sequencing: The new technology allows direct sequencing of DNA fragments rather than trying to figure out the gene order, DNA mutations and new genes by traditional methods such as RFLP analysis, chromosomal walking or even transduction and conjugation experiments in bacteria. DNA sequencing has now reached the automated stage and is routinely used in many laboratories even for HLA typing. In automated sequencing, a single sequencing reaction is carried out in which the four ddNTPs are labeled with differently colored dyes. At the end of the reaction, the mixture is run in a polyacrylamide gel and the colored chains are detected as they migrate through the gel. The detection system identifies the terminal base from the

wavelength of the fluorescence emitted upon excitation by a laser. The DNA polymerase used in a sequencing reaction is usually part of the E.coli polymerase known as the Klenow fragment or a genetically modified DNA polymerase from the phage T7 (Sequenase). The usual *Taq* DNA polymerase can also be used for this purpose.

B – The genetic issues directly associated with bioethics

Human cloning: reproductive and therapeutic cloning

Cloning is the process of asexually producing a group of cells (clones), all genetically identical to the original ancestor. The word is also used in recombinant DNA manipulation procedures to produce multiple copies of a single gene or segment of DNA. It is more commonly known as the production of a cell or an organism from a somatic cell of an organism with the same nuclear genomic (genetic) characters - without fertilization. A clone is a collection of cells or organisms that are genetically identical. Some vegetables are made this way, like asparagus, or flowers like orchids.

Human **reproductive cloning** is the production of a human fetus from a single cell by asexual reproduction. In 2001 a cloned embryo was reported made by nuclear transfer, though in 1993 cloned embryos were made by splitting human embryos. In the late 1990s reproductive cloning was used to produce clones of the adults of a number of mammalian species, including sheep, mice and pigs. The most famous of these was Dolly, the sheep. Many countries rushed to outlaw the possibility of reproductive cloning in humans. Most mammalian embryos can only be split into 2-4 clones, after that the cells lack the ability to start development into a human being.

Therapeutic cloning is the cloning of embryos containing DNA from an individual's own cell to generate a source of embryonic stem (ES) cell-progenitor cells that can differentiate into the different cell types of the body. ES cells are capable of generating all cell types, unlike multipotent adult-derived stem cells which generate many but not all cell types. The aim is to produce healthy replacement tissue that would be readily available. Since it is from the same body it is immunocompatible so that the recipients would not have to take immunosuppressant drugs for the rest of their lives, as they do if they receive an organ from another person.

The Human Genome Project

Begun formally in 1990, the U.S. Human Genome Project was a 13-year effort coordinated by the U.S. Department of Energy and the National Institutes of Health. The project originally was planned to last 15 years, but rapid technological advances accelerated the completion date to 2003. 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.

To help achieve these goals, researchers also studied the genetic makeup of several nonhuman organisms. These include the common human gut bacterium *Escherichia coli*, the fruit fly, and the laboratory mouse.

A unique aspect of the U.S. Human Genome Project is that it was the first large scientific undertaking to address potential ELSI implications arising from project data.

Another important feature of the 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.

Sequence and analysis of the human genome working draft was published in February 2001 and April 2003 issues of *Nature* and *Science*. See an index of these papers and learn more about the insights gained from them.

Gene therapy: somatic and germline gene therapy

Somatic Cell Gene Therapy

Many genetic diseases may be able to be treated by correcting the defective genes, by gene therapy. Gene therapy is a therapeutic technique in which a functioning gene is inserted into the cells of a patient to correct an inborn genetic error or to provide a new function to the cell. It means the genetic modification of DNA in the body cells of an individual patient, directed to alleviating disease in that patient.

There have been several hundred human gene therapy clinical trials in many countries (including USA, EU, Canada, China, Japan, New Zealand...), involving over 6000 patients world-wide, for several different diseases including several cancers.

Somatic cell gene therapy involves injection of 'healthy genes' into somatic (body) cells of a patient. The DNA change is not inherited to children. The first human gene therapy protocol began in September 1990 that successfully treated adenosine deaminase deficiency (ADA) disease.

From 1989 until September 1999 there were thousands of patients in trials and no one died because of the experiments. 18 year-old Jesse Gelsinger died at the University of Pennsylvania (USA) on 17 September 1999, four days after receiving a

relatively high dose of an experimental gene therapy. His death was the result of a large immune reaction to the genetically engineered adenovirus that researchers had infused into his liver. There was much review of the procedures for safety following that case.

Gene therapy is still an experimental therapy, but if it is safe and effective, it may prove to be a better approach to therapy than many current therapies, because gene therapy cures the cause of the disease rather than merely treating the symptoms of a disease. Also, many diseases are still incurable by other means, so the potential benefit is saving life.

Germ-line gene therapy

At the present gene therapy is not inheritable. Germ cells are cells connected with reproduction, found in the testis (males) and ovary (females), i.e. Egg and sperm cells and the cells that give rise to them. Germ-line gene therapy targets the germ cells. This type of therapy may also mean injecting DNA to correct, modify or add DNA into the pronucleus of a fertilized egg. The latter technology would require that fertilization would occur *in vitro* using the usual IVF procedures of super-ovulation and fertilization of a number of egg cells prior to micromanipulation for DNA transfer and then embryo transfer to a mother after checking the embryo's chromosomes.

Genetic counseling and prenatal diagnosis

Present-day medicine recognizes that genetic diseases are inherited based on the nature of DNA, genes, and chromosomes. Now that the human genome has been completely sequenced, scientists are better able to study how changes in DNA cause human disease. This will ultimately help in diagnosing and treating genetic disorders.

However, until science has the knowledge to treat some of the more serious, sometimes fatal genetic disorders, the best option is prevention. Prevention of genetically transmitted diseases can consist of major choices: abstinence from pregnancy, egg or sperm donation, preimplantation or prenatal diagnosis and termination, or early treatment of affected pregnancies.

Prenatal diagnosis involves testing fetal cells, amniotic fluid, or amniotic membranes to detect fetal abnormalities. Preimplantation diagnosis is a new technique only available in specialized centers. It involves *in vitro* fertilization and genetic testing of the resulting embryos prior to implanting only those embryos found not to have the abnormal gene.

Genetic counseling and prenatal diagnosis provides parents with the knowledge to make intelligent, informed decisions regarding possible pregnancy and its outcome. Based on genetic counseling, some parents (in the face of possibly lethal genetic disease) have forgone pregnancy and adopted children while others have opted for egg or sperm donation from an anonymous donor who is not likely to be a carrier of the specific disease.

Many diseases transmitted as a single gene defect can now be diagnosed very early in pregnancy. Because of this some parents choose to become pregnant and

have the disease status of the fetus determined early in the pregnancy. The pregnancy is continued if the fetus is disease-free. Parents who decide to continue the pregnancy with a defective fetus may be able to better prepare to care for the infant by being informed about the disease in advance. For example, genetic diseases that have a diet intolerance component may be treated with specialized diets for the mother and newborn baby.

Genetic Testing

Genetic tests, also called Gene tests or DNA-based tests, the newest and most sophisticated of the techniques used to test for genetic disorders, involve direct examination of the DNA molecule itself. Other genetic tests include biochemical tests for such gene products as enzymes and other proteins and for microscopic examination of stained or fluorescent chromosomes. Genetic tests are used for several reasons, including:

- carrier screening, which involves identifying unaffected individuals who carry one copy of a gene for a disease that requires two copies for the disease to be expressed
- preimplantation genetic diagnosis
- prenatal diagnostic testing
- newborn screening
- presymptomatic testing for predicting adult-onset disorders such as Huntington's disease
- presymptomatic testing for estimating the risk of developing adult-onset cancers and Alzheimer's disease
- confirmational diagnosis of a symptomatic individual
- forensic/identity testing

In gene tests, scientists scan a patient's DNA sample for mutated sequences. A DNA sample can be obtained from any tissue, including blood. For some types of gene tests, researchers design short pieces of DNA called probes, whose sequences are complementary to the mutated sequences. These probes will seek their complement among the three billion base pairs of an individual's genome. If the mutated sequence is present in the patient's genome, the probe will bind to it and flag the mutation. Another type of DNA testing involves comparing the sequence of DNA bases in a patient's gene to a normal version of the gene. Cost of testing can range from hundreds to thousands of dollars, depending on the sizes of the genes and the numbers of mutations tested.

Gene testing already has dramatically improved lives. Some tests are used to clarify a diagnosis and direct a physician toward appropriate treatments, while others allow families to avoid having children with devastating diseases or identify people at high risk for conditions that may be preventable. Aggressive monitoring for and removal of colon growths in those inheriting a gene for familial adenomatous polyposis, for example, has saved many lives. On the horizon is a gene test that will

provide doctors with a simple diagnostic test for a common iron-storage disease, transforming it from a usually fatal condition to a treatable one.

Genetic DNA testing to evaluate paternity/parentage or forensic/identity testing is possible because our biological characteristics are passed from generation to generation following the basic rules of inheritance. These rules have been known for more than a century. Deoxyribonucleic acid (DNA), which is a very stable and strictly inherited molecule, encodes all genetic information and determines our biological characteristics. Modern DNA paternity testing relies on the fact that we can detect and study "DNA markers" at specific structural regions of the DNA. Many different DNA markers exist in the general population. However, only two such DNA markers exist in any one individual. A child inherits one DNA marker from the mother and one from the father. A DNA test begins by learning which DNA markers are present in the child and the mother. It is then possible to determine which of the child's DNA markers was inherited from the mother and which was inherited from the biological father. To evaluate paternity and complete a paternity test, a series of DNA tests is performed on the biological specimens provided by the mother, child, and alleged father. When the DNA Profiles™ of this trio are compared to each other, the paternity test will provide two possible results; the alleged father will be either included or excluded as the biological father of the child.

Pharmacogenomics

Pharmacogenomics is the study of how an individual's genetic inheritance affects the body's response to drugs. The term comes from the words pharmacology and genomics and is thus the intersection of pharmaceuticals and genetics.

Pharmacogenomics holds the promise that drugs might one day be tailor-made for individuals and adapted to each person's own genetic makeup. Environment, diet, age, lifestyle, and state of health all can influence a person's response to medicines, but understanding an individual's genetic makeup is thought to be the key to creating personalized drugs with greater efficacy and safety.

Pharmacogenomics combines traditional pharmaceutical sciences such as biochemistry with annotated knowledge of genes, proteins, and single nucleotide polymorphisms.

One can anticipate the benefits of Pharmacogenomics, which are as follows:

- **More Powerful Medicines**

Pharmaceutical companies will be able to create drugs based on the proteins, enzymes, and RNA molecules associated with genes and diseases. This will facilitate drug discovery and allow drug makers to produce a therapy more targeted to specific diseases. This accuracy not only will maximize therapeutic effects but also decrease damage to nearby healthy cells.

- **Better, Safer Drugs the First Time**

Instead of the standard trial-and-error method of matching patients with the right drugs, doctors will be able to analyze a patient's genetic profile and prescribe the best available drug therapy from the beginning. Not only will this take the guesswork

out of finding the right drug, it will speed recovery time and increase safety as the likelihood of adverse reactions is eliminated. Pharmacogenomics has the potential to dramatically reduce the estimated 100,000 deaths and 2 million hospitalizations that occur each year in the United States as the result of adverse drug response (1).

- **More Accurate Methods of Determining Appropriate Drug Dosages** Current methods of basing dosages on weight and age will be replaced with dosages based on a person's genetics --how well the body processes the medicine and the time it takes to metabolize it. This will maximize the therapy's value and decrease the likelihood of overdose.

- **Advanced Screening for Disease**

Knowing one's genetic code will allow a person to make adequate lifestyle and environmental changes at an early age so as to avoid or lessen the severity of a genetic disease. Likewise, advance knowledge of particular disease susceptibility will allow careful monitoring, and treatments can be introduced at the most appropriate stage to maximize their therapy.

- **Better Vaccines**

Vaccines made of genetic material, either DNA or RNA, promise all the benefits of existing vaccines without all the risks. They will activate the immune system but will be unable to cause infections. They will be inexpensive, stable, easy to store, and capable of being engineered to carry several strains of a pathogen at once.

- **Improvements in the Drug Discovery and Approval Process**

Pharmaceutical companies will be able to discover potential therapies more easily using genome targets. Previously failed drug candidates may be revived as they are matched with the niche population they serve. The drug approval process should be facilitated as trials are targeted for specific genetic population groups -- providing greater degrees of success. The cost and risk of clinical trials will be reduced by targeting only those persons capable of responding to a drug.

- **Decrease in the Overall Cost of Health Care**

Decreases in the number of adverse drug reactions, the number of failed drug trials, the time it takes to get a drug approved, the length of time patients are on medication, the number of medications patients must take to find an effective therapy, the effects of a disease on the body (through early detection), and an increase in the range of possible drug targets will promote a net decrease in the cost of health care.

Genetic engineering and Food

Genetic engineering or genetic modification is to alter the genetic constitution of organisms by mixing the DNA of different genes and species together. The living organisms with altered DNA are called Genetically Modified Organisms (GMOs). Genetic engineering is considered special because often the techniques involves manipulating genes in a way that is not expected to occur ordinarily in nature, that characters can be changed between the species.

Many kinds of GMOs have been developed for environmental purposes and for health and medicine. Genetic engineering has been particularly successfully used and applied in food and agriculture to produce genetically modified (GM) foods. Such as transgenic plants that carry several enhanced characteristics by inserting genes from various organisms, for example, plants with increased yield, disease resistance, and pest resistance with inserted Bt genes kill selectively pests that eat crops. There have also been fruits and vegetables modified for long term storage or delayed ripening that remain fresh for a long time, which is useful also during transportation to the market. Over 15 countries of the world use GM crops for the general food production already.

The second wave of GM plants are those with high nutritional content and improved food quality like golden rice, plants that can tolerate high salt levels in the land or those modified so that they can grow in harsh conditions like drought.

To consolidate the knowledge presented on this article

1. A chromosome:

A. is composed of amino acids
B. is organized in the nucleus by histones
C. is produced from RNA
D. is present in 46 pairs in human cells

ANSWER

2. Genes:

A. never function when they contain a mutation
B. directly produce proteins
C. contain random pairings of nucleotides
D. all of the above
E. none of the above

ANSWER

3. During the process of transcription, genetic information is transferred from:

A. DNA to RNA
B. RNA to DNA
C. DNA to protein
D. Protein to RNA

ANSWER

4. A mutation that _____ production of a given _____ can manifest as clinical disease.

A. increases/protein
B. decreases/mRNA
C. decreases/ protein
D. increases/mRNA
E. all of the above

F. none of the above
ANSWER

5. A mutation occurs that disrupts the normal structure and function of hemoglobin. Which of the following is true?

A. clinical disease will develop based on the mutation alone.
B. environmental factors can play a large role in the development of clinical disease.
C. each person with the same mutation will follow the same clinical course.
D. family members should be tested for this hereditary condition.
ANSWER

6. A germline mutation _____ while a somatic mutation _____.

A. is never passed from parents to offspring // is present in all cells of one's body
B. is always passed from parents to offspring // is present in all cells of one's body
C. is present in all cells of one's body // is never passed from parents to offspring
D. is responsible for non-hereditary cancers // is not often a direct cause of inherited disease
ANSWER

7. A missense mutation

A. does not affect protein structure
B. does not affect protein function
C. leads to substitution of an amino acid in a new place in the protein
D. all of the above
E. none of the above
ANSWER

8. A nonsense mutation

A. does not affect protein structure
B. may not lead to clinical disease
C. involves an inappropriate stop codon
D. A and B
E. A and C
F. All of the above
ANSWER

9. A silent mutation

A. results in no change in protein structure/function
B. can sometimes lead to clinical disease
C. involves substitution of one amino acid for another
D. A and C
E. A and B
ANSWER

10. A polymorphism is a form of mutation that leads to clinical disease.

True
False
ANSWER