Basic Principles of Genetics

Module by: Professor Le Dinh Luong

Lecture 1. Genetics is a 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, Mendel's work was not appreciated immediately.

That's why 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.

From 1865 to now the history of genetics development is the development of human knowledge and understanding of genes. In other words, genetics is a science of the structure, function and movement of genes. Before going into the exact definition of gene, one can begin by understanding that a gene is a piece of DNA which has a function such as determining human eye color, pea seed shape or a disease.

Lecture 2. Genes are mostly located 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 or in the nucleoid region of prokaryotes.

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.

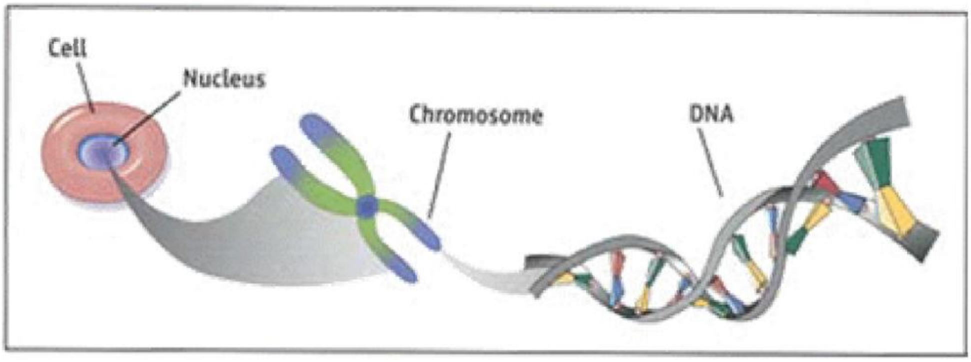


Figure 1

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.

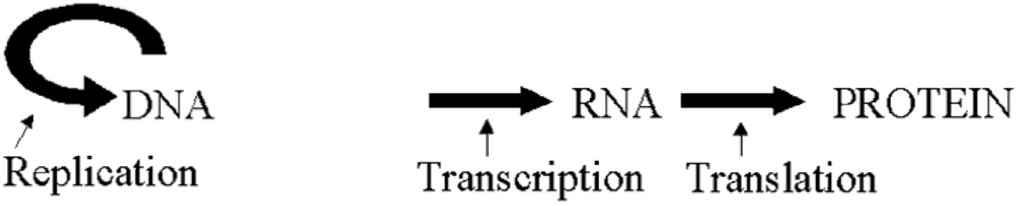


Figure 2

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.

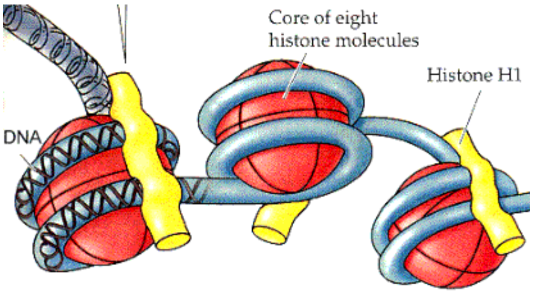


Figure 3

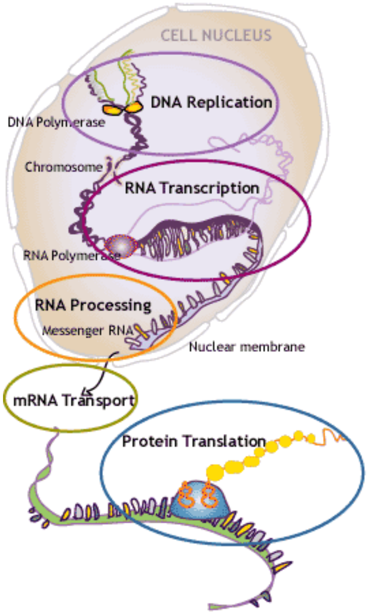


Figure 4

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.

Lecture 3. 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.

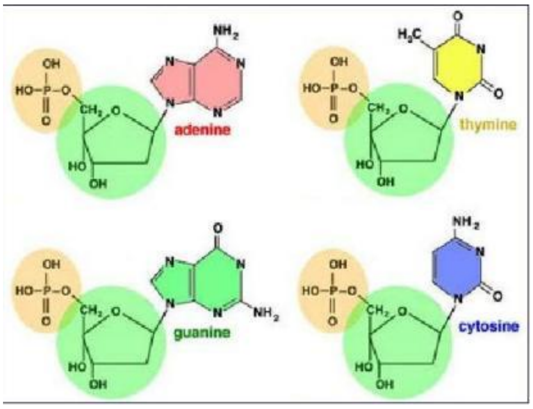


Figure 5

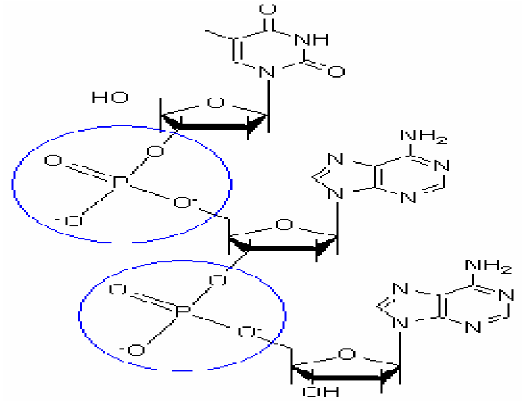


Figure 6

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 definedas 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, which is make up by the phosphodiester bonds linking between the 3' carbonatom and the 5' carbon of the sugar deoxyribose (in DNA) or ribose (in RNA).

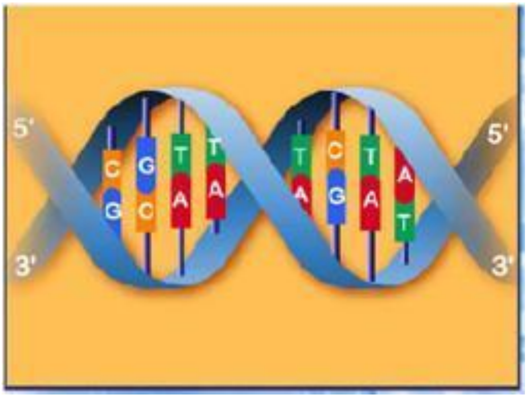


Figure 7: The two strands in a double helix are oriented in opposite directions.

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, for example, in humans, 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 in 1971 when it was discovered that some viruses' genetic information is RNA.

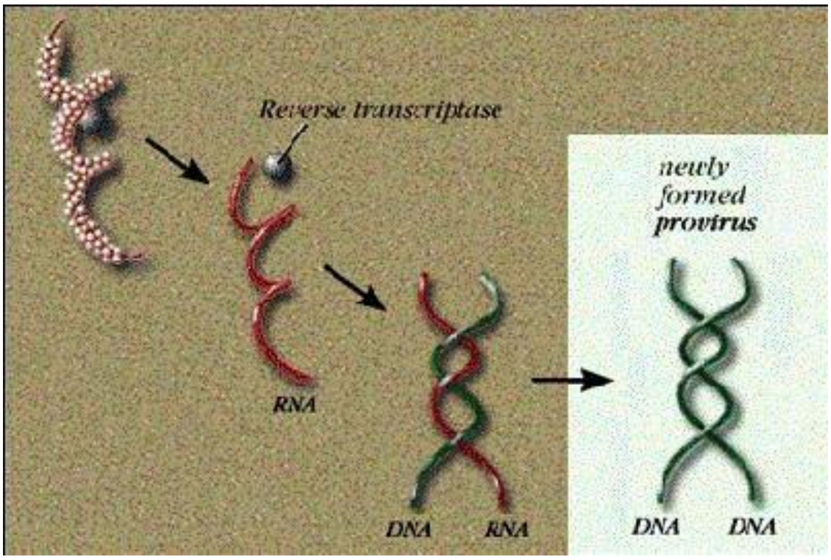


Figure 8

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 as shown in the picture.

Lecture 4. Genes can replicate themselves

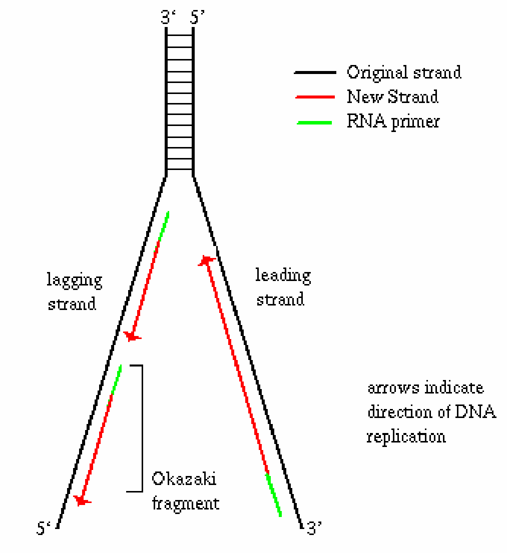


Figure 9

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 3’ direction. A primer begins any new strand, including each□strand in a 5’ 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 two DNA molecules. Each consists of an original nucleotide strand next to a new complementary strand. The two molecules are identical to each other.

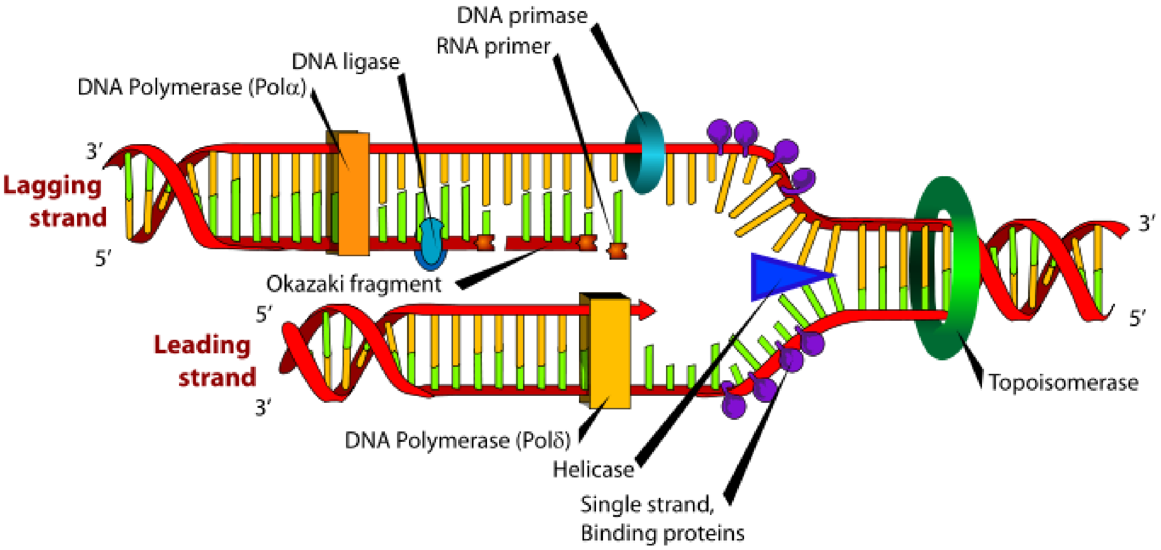


Figure 10

A detailed and clear schematic of DNA synthesis kindly provided by Prof. Douglas J. Burks is shown below:

http://upload.wikimedia.org/wikipedia/commons/thumb/9/9f/DNA_replication.svg/691px-DNA_replication.svg.png

Lecture 5. Language of genes is simple and informative

Genetic information likes 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.

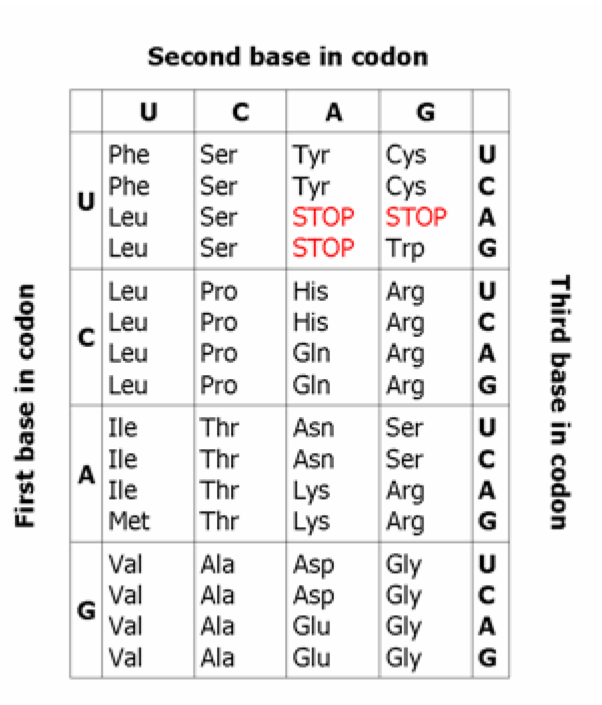
These words combine together to form sentences called genes, which encode the instruction for amino acids in a polypeptide.

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 comparisons between English Language and Genetic Language:

English Language FIXME: A LIST CAN NOT BE A TABLE ENTRY. We use 26 letters to make words. The words can be any length we need. We join words together to create sentences Each sentence starts with a capital letter. Each sentence ends with a fullstop. All the sentences combine to form a book	Genetic Language FIXME: A LIST CAN NOT BE A TABLE ENTRY. DNA uses 4 molecules to make codons. The codons can only be 3 nucleotides long. The codons join together to form genes. The gene starts with codon AUG.The gene stops at a specific stop codon. All the genes combine to form the genome



Second base in codon

First base in codon	U	C	A	G	Third base in codon
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	STOP	STOP	A
	Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Figure 11

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 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

Lecture 6. 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, and this 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.

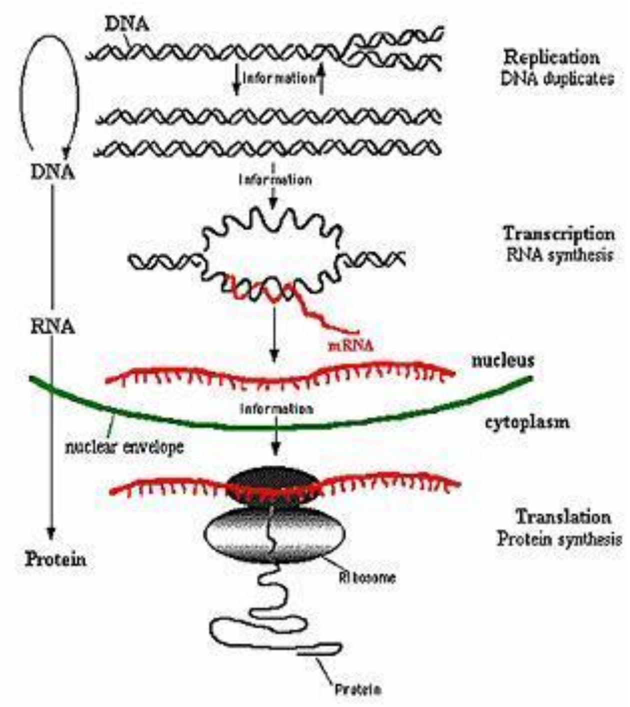
Nonsense mutation The codon for an amino acid residue within the original protein is changed to a stop codon, which leads to a premature termination of the protein resulting in a non functional protein.

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 results 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 mutations 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

Figure 12

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.

LacZ mutations* LacZ mutations are an example of particular mutations found in the LacZ gene of E.coli, which encodes the lactose hydrolyzing enzyme ß-galactosidase.

There is a special compound known as X-gal that can be hydrolyzed by ß-galactosidase to release a dark blue pigment. When X-gal is added to the growth medium in petri plates, Lac+ E. coli colonies turn blue, whereas Lac— colonies with mutations in the LacZ gene are white. By screening many colonies on such plates it is possible to isolate a collection of E. coli mutants with alterations in the LacZ gene. PCR amplification of the LacZ gene from each mutant followed by DNA sequencing allows the base changes that cause the LacZ— phenotype to be determined. A very large number of different LacZ mutations can be found, but they can be categorized into three general types: missense, nonsense and frameshift .

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.

Suppressor mutations* A powerful mode of genetic analysis is to investigate the types of mutations that can reverse the phenotypic effects of a starting mutation. Say that you start with a mi- λ phage mutant that makes small plaques. After plating a large number of these mutant phages, rare revertants can be isolated by looking for phage that have restored the ability to make large plaques. These revertants could have either been mutated such that the starting mutation was reversed, or they could have acquired a new mutation that somehow compensates for the starting mutation. The possibilities are:

1) Back mutation - true wild type

2) Intragenic suppressor - compensating mutation in same gene

3) Extragenic suppressor - compensating mutation in different gene

These possibilities can be distinguished in that a revertant that arose by suppression will still carry the starting mutation (now masked by the suppressor mutation), whereas a back mutation will produce a true wild type phage. The general test is to cross the revertant to wild type and to note whether mi- recombinants are observed. A back mutation crossed to wild type will not produce any mi- progeny, whereas a revertant that results from an extragenic suppressor will produce many mi- recombinants. Intragenic suppressors will produce an intermediate result that sometimes can be difficult to distinguish from a back mutation in practice. For example, an intragenic suppressor that lies very close to the original mi- mutation may be able to produce mi- recombinants in principle, but these recombinants may be too rare to be readily observed.

Nonsense suppressor An important class of extragenic suppressor mutations can suppress nonsense mutations by changing the ability of the cells to read a nonsense codon as codon for an amino acid. Such extragenic revertants were originally isolated by selecting for reversion of amber (UAG) mutations in two different genes. Since simultaneous back mutations at two different sites is highly improbable, the most frequent mechanism for suppression is a single mutation in the gene for a tRNA that changes the codon recognition portion of the tRNA. (For example, one of several possible nonsense suppressors occurs in the gene for a serine tRNA (tRNAser). One of six tRNAser normally contains the anticodon sequence CGA which recognizes the serine codon UCG by convention sequences which are given in the 5’ to 3’ direction.

A mutation that changes the anticodon to CUA allows the mutant tRNAser to recognize a UAG codon and insert tryptophan when a UAG codon appears in a coding sequence. Recognition of UCG (serine codon) Recognition of UAG(stop codon) by wild type tRNAser by amber suppressor mutant tRNAser (*)

mRNA: 5’———UCG———3’ 5’——UAG—————3’

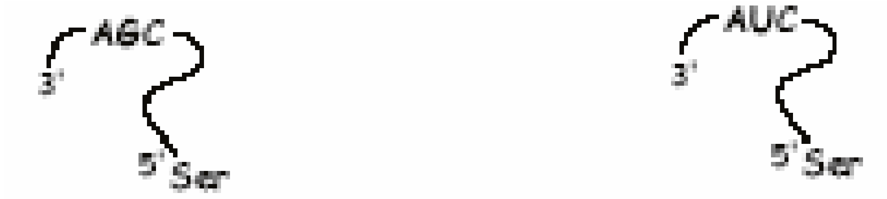


Figure 13

The presence of an amber suppressing mutation is usually designated Su+ whereas a wild-type (nonsuppressing) strain would be designated Su-.

Example: Pam designates an amber (nonsense) mutation in the λ phage P gene, which is required for λ phage DNA replication. When λ Pam phage are grown on E. coli with an amber suppressor (Su+), the phage multiply normally; but when λ Pam phage infect a nonsuppressing host (Su—), the phage DNA cannot replicate.

The combined use of amber mutations and an amber suppressor produces a conditional mutant, which is a mutant that is expressed under some circumstances but not under others. Conditional mutants are especially useful for studying mutations in essential genes. Another kind of conditional mutation is a temperature sensitive mutation for which the mutant trait is exhibited at high temperature but not at low temperature. In a sense, auxotrophic mutations are also conditional because auxotrophic mutants can be grown in the presence of the required nutrient, but the mutants will not grow when the nutrient is not provided.

Lecture 7. 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 encode for the proteins that compose our body through the processes of transcription and translation, with messenger 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 mRNA from DNA is carried out by an enzyme, RNA polymerase, 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, a mutation which leads to decreased amounts of functional hemoglobin. This condition, ð-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, the modified proteins and their functions include the following:

-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. Below is the list of 20 amino acids with their chemical formular which was kindly offered by Prof. Douglas J. Burks.

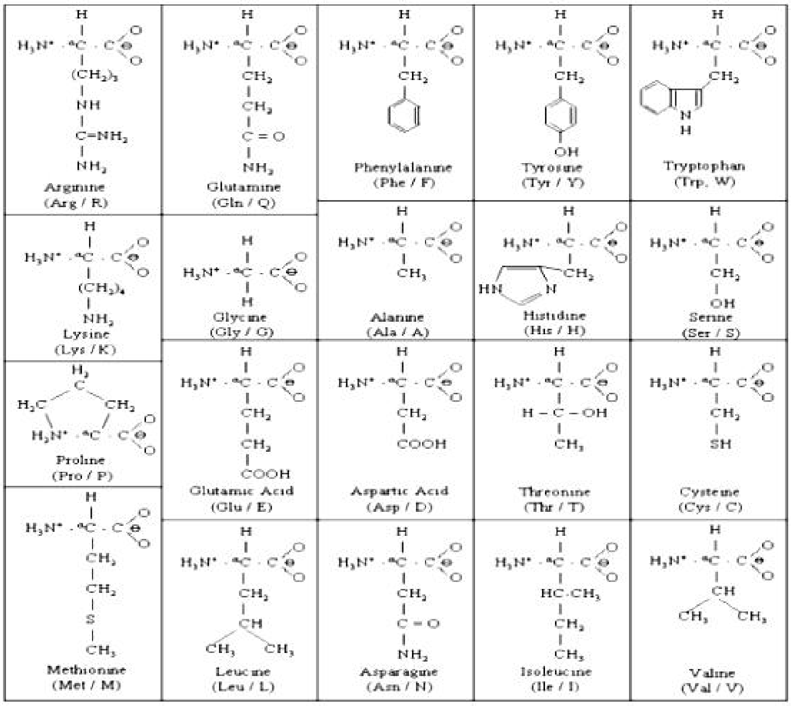


Figure 14

Protein function Proteins play an enormous variety of roles within the body. They are responsible for transport, storage and the structural framework of cells. They make up antibodies, the enzymatic machinery that catalyzes biochemical reactions responsible for metabolic activities. Finally, proteins are an important component in many hormones, and contractile proteins are responsible 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.

Lecture 8. 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, which is called differential gene expression.

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.

The details on this matter are described in the lecture 24*, where the lac operon plays a role of gene regulation unit, the schematic of which is shown below.

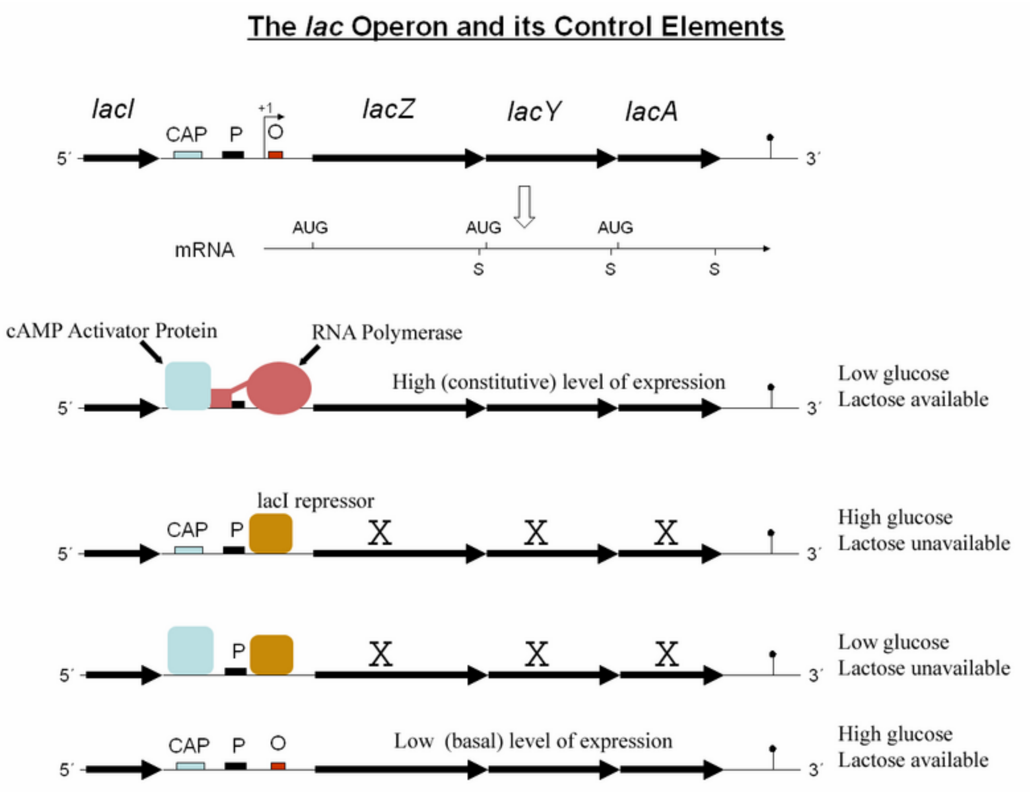


Figure 15

Lecture 9. 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, which are called differentiated cells as opposed to stem cells. 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.

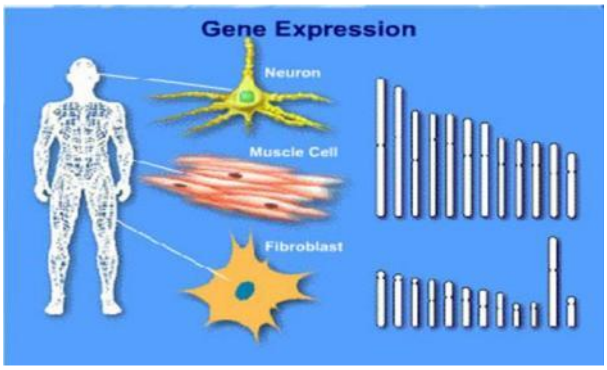


Figure 16

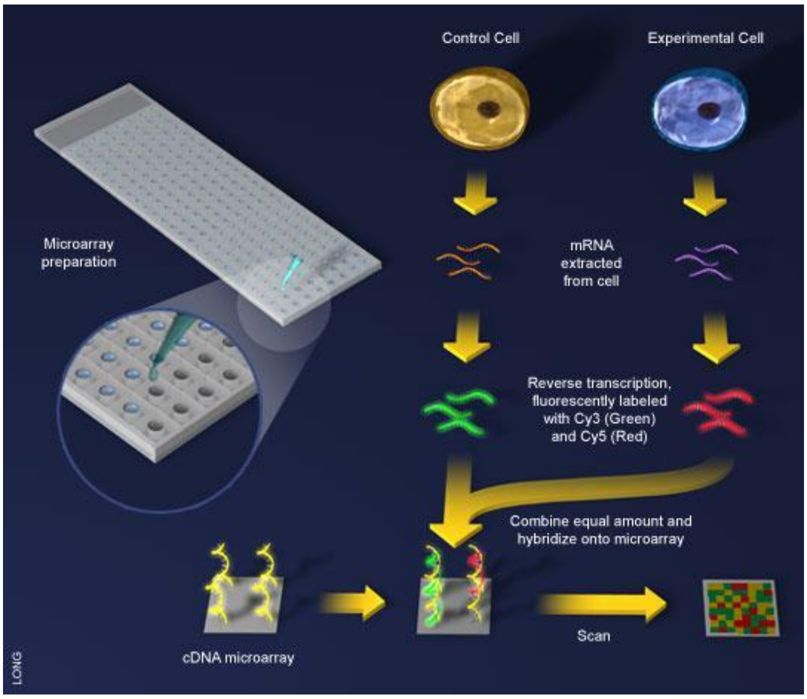

Control Cell
Experimental Cell
Microarray preparation
mRNA extracted from cell
Reverse transcription, fluorescently labeled with Cy3 (Green) and Cy5 (Red)
Combine equal amount and hybridize onto microarray
Scan
cDNA microarray
LONG

Figure 17

This notion of cell-specific expression of genes is supported 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, as shown in the figure to the left, kindly provided by Prof. Douglas J. Burks.

Lecture 10. Genes move 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, mitosis and meiosis in eukaryotic organisms.

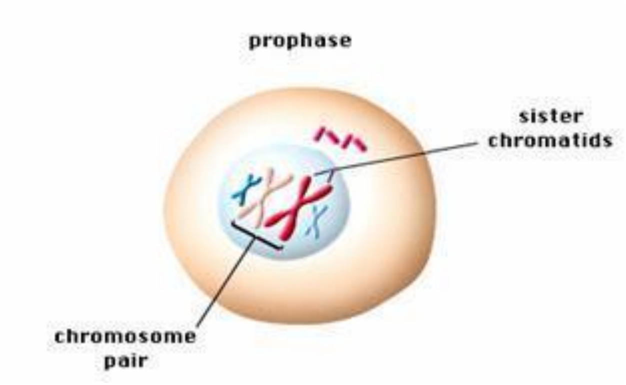


Figure 18

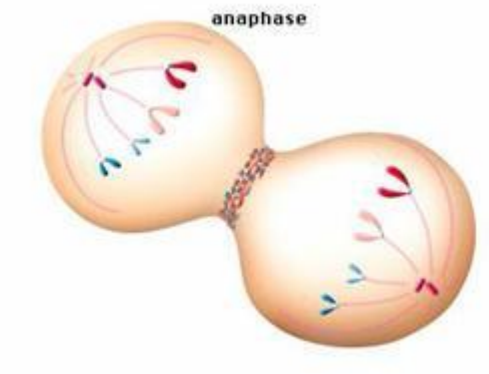


Figure 19

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.

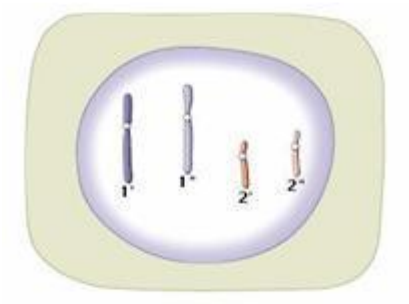


Figure 20

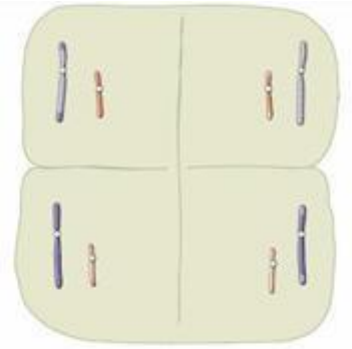


Figure 21

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, which represent Mendel's units of inheritance that show independent segregation and assortment. 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.

Lecture 11. Genes can transfer 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.

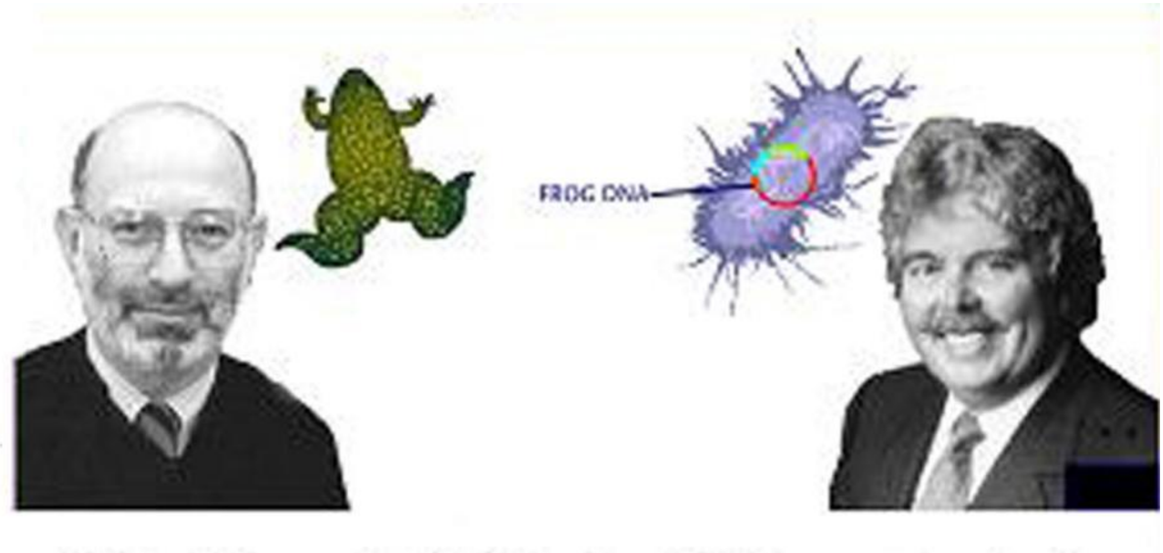


Figure 22

Stanley Cohen (on the left) and Herbert Boyer (on the right) made what would be one of the first genetic engineering experiments, in 1973. They demonstrated that the gene for frogribosomalRNA could be transferred into bacterial cells E.coli and expressed by them.

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. This transfer may occur naturally through transformation.

This is an idea that geneticists are realizing is more important than previously thought.

The techniques for gene manipulation as well as for gene transfer are described in detail in the lectures 14 and 15.

Lecture 12. A genome is an entire set of genes (http://en.wikipedia.org/wiki/Genome)

In classical genetics, the genome of a diploidorganism including eukarya refers to a full set of chromosomes or genes in a gamete; thereby, a regular somatic cell contains two full sets of genomes. In a haploidorganism, including bacteria, archaea, virus, and mitochondria, a cell contains only a single set of genome, usually in a single circular or contiguous linear DNA (or RNA for some viruses). In modern molecular biology the genome of an organism is its hereditary information encoded in DNA (or, for some viruses, RNA).

The genome includes both the genes and the non-coding sequences of the DNA. The term was adapted in 1920 by Hans Winkler, Professor of Botany at the University of Hamburg, Germany. The Oxford English Dictionary suggests the name to be a portmanteau of the words gene and chromosome; however, many related -ome words already existed, such as biome and rhizome, forming a vocabulary into which genome fits systematically.[1]

More precisely, the genome of an organism is a complete genetic sequence on one set of chromosomes; for example, one of the two sets that a diploid individual carries in every somatic cell. The term genome can be applied specifically to mean that stored on a complete set of nuclear DNA (i.e., the "nuclear genome") but can also be applied to that stored within organelles that contain their own DNA, as with the mitochondrial genome or the chloroplast genome. Additionally, the genome can comprise nonchromosomal gentic elements such as viruses, plasmids, and transposable elements[2]. When people say that the genome of a sexually reproducingspecies has been "sequenced," typically they are referring to a determination of the sequences of one set of autosomes and one of each type of sex chromosome, which together represent both of the possible sexes. Even in species that exist in only one sex, what is described as "a genome sequence" may be a composite read from the chromosomes of various individuals. In general use, the phrase "genetic makeup" is sometimes used conversationally to mean the genome of a particular individual or organism. The study of the global properties of genomes of related organisms is usually referred to as genomics, which distinguishes it from genetics which generally studies the properties of single genes or groups of genes.

Both the number of base pairs and the number of genes vary widely from one species to another, and there is little connection between the two (an observation known as the C- value paradox). At present, the highest known number of genes is around 60,000, for the protozoan causing trichomoniasis (see List of sequenced eukaryotic genomes), almost three times as many as in the human genome.

Note that a genome does not capture the genetic diversity or the genetic polymorphism of a species. For example, the human genome sequence in principle could be determined from just half the information on the DNA of one cell from one individual. To learn what variations in genetic information underlie particular traits or diseases requires comparisons across individuals. This point explains the common usage of "genome" (which parallels a common usage of "gene") to refer not to the information in any particular

DNA sequence, but to a whole family of sequences that share a biological context.

Although this concept may seem counter intuitive, it is the same concept that says there is no particular shape that is the shape of a cheetah. Cheetahs vary, and so do the sequences of their genomes. Yet both the individual animals and their sequences share commonalities, so one can learn something about cheetahs and "cheetah-ness" from a single example of either.

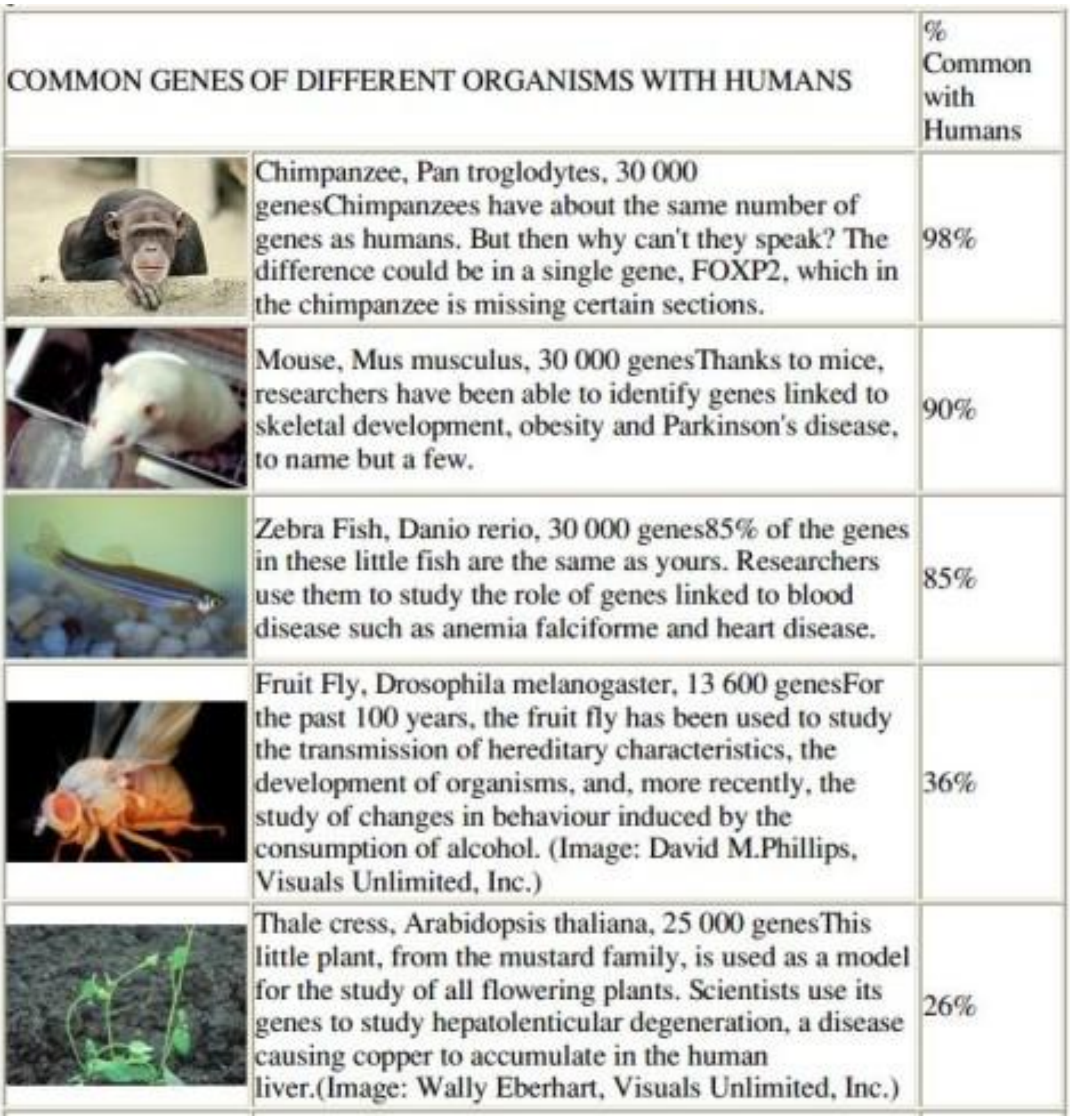
Comparison of different genome sizes

Organism	Genome size (base pairs)	Note
Virus, Bacteriophage MS2	3,569	First sequenced RNA-
		genome[3]
Virus, SV40	5,224	[4]
Virus, Phage Φ-X174;	5,386	First sequenced DNA-genome[5]
Virus, Phage λ	48,502	
Bacterium, Haemophilus influenzae	1,830,000	First genome of living organism, July
Bacterium, Carsonella ruddii	160,000	Smallest non-viral
Bacterium, Buchnera aphidicola	600,000	
Bacterium, Wigglesworthia glossinidia	700,000	
Bacterium, Escherichia coli	4,600,000	[8]
Amoeba, Amoeba dubia	670,000,000,000	Largest known
Plant, Arabidopsis thaliana	157,000,000	First plant genome sequenced, Dec
Plant, Genlisea margaretae	63,400,000	Smallest recorded flowering plant
Plant, Fritillaria assyrica	130,000,000,000	
Plant, Populus trichocarpa	480,000,000	First tree genome,
moss, Physcomitrella patens	480,000,000	First genome of a bryophyte, January
Yeast,Saccharomyces cerevisiae	12,100,000	[12]
Fungus, Aspergillus nidulans	30,000,000	
Nematode, Caenorhabditis elegans	98,000,000	First multicellular animal genome,
Insect, Drosophila melanogaster aka Fruit Fly	130,000,000	[14]

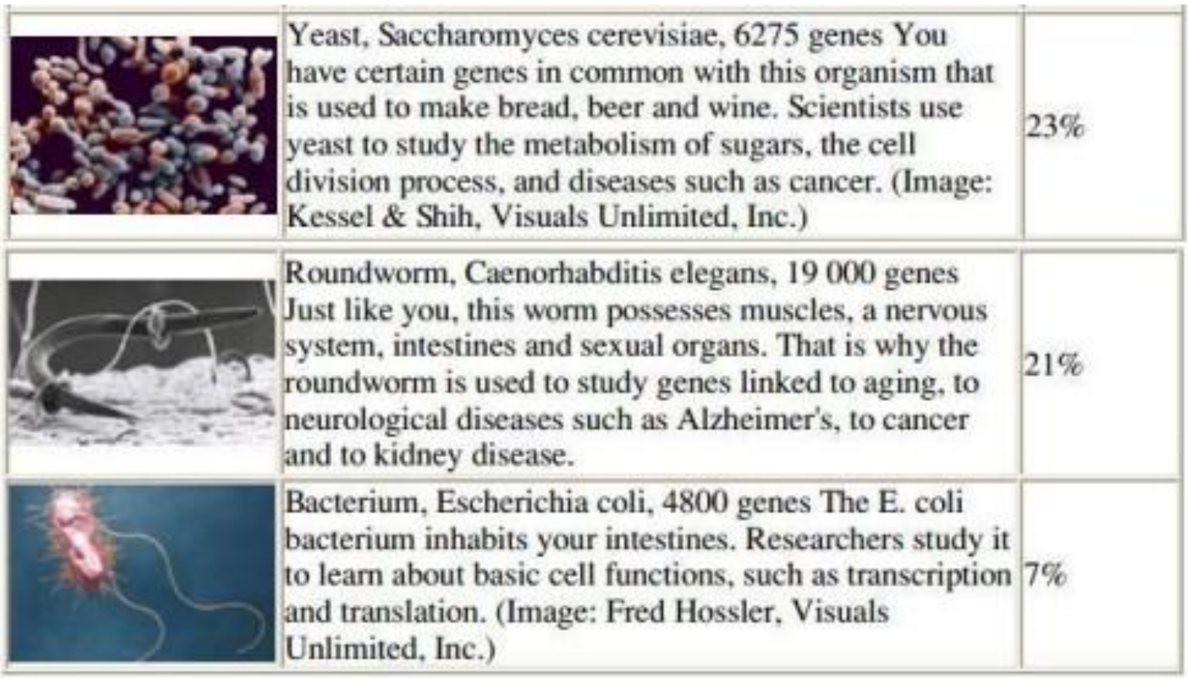
Insect, Bombyx mori aka Silk Moth	530,000,000	
Insect, Apis mellifera aka Honey Bee	1,770,000,000	
Fish, Tetraodon nigroviridis, type of Puffer fish	385,000,000	Smallest vertebrate genome known
Mammal, Homo sapiens	3,200,000,000	
Fish, Protopterus aethiopicus aka Marbled lungfish	130,000,000,000	Largest vertebrate genome known

Lecture 13. Living organisms share common genes

All organisms store genetic information in the same molecules - DNA or 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.



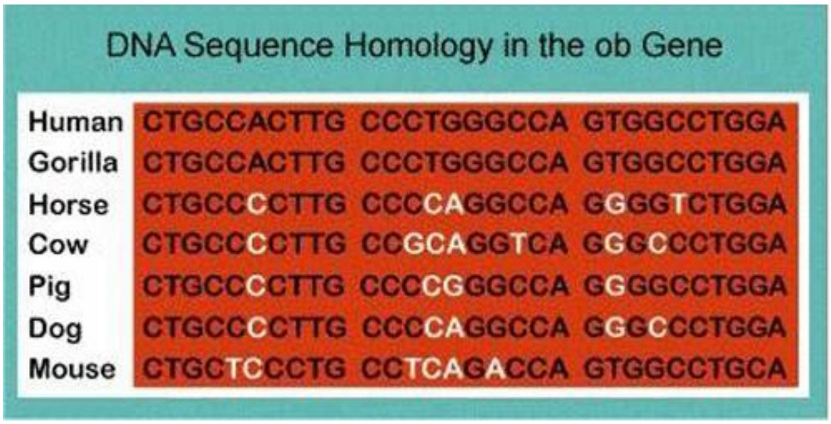
COMMON GENES OF DIFFERENT ORGANISMS WITH HUMANS		% Common with Humans
	Chimpanzee, Pan troglodytes, 30 000 genesChimpanzees have about the same number of genes as humans. But then why can't they speak? The difference could be in a single gene, FOXP2, which in the chimpanzee is missing certain sections.	98%
	Mouse, Mus musculus, 30 000 genesThanks to mice, researchers have been able to identify genes linked to skeletal development, obesity and Parkinson's disease, to name but a few.	90%
	Zebra Fish, Danio rerio, 30 000 genes85% of the genes in these little fish are the same as yours. Researchers use them to study the role of genes linked to blood disease such as anemia falciforme and heart disease.	85%
	Fruit Fly, Drosophila melanogaster, 13 600 genesFor the past 100 years, the fruit fly has been used to study the transmission of hereditary characteristics, the development of organisms, and, more recently, the study of changes in behaviour induced by the consumption of alcohol. (Image: David M.Phillips, Visuals Unlimited, Inc.)	36%
	Thale cress, Arabidopsis thaliana, 25 000 genesThis little plant, from the mustard family, is used as a model for the study of all flowering plants. Scientists use its genes to study hepatolenticular degeneration, a disease causing copper to accumulate in the human liver.(Image: Wally Eberhart, Visuals Unlimited, Inc.)	26%



	Yeast, Saccharomyces cerevisiae, 6275 genes You have certain genes in common with this organism that is used to make bread, beer and wine. Scientists use yeast to study the metabolism of sugars, the cell division process, and diseases such as cancer. (Image: Kessel & Shih, Visuals Unlimited, Inc.)	23%
	Roundworm, Caenorhabditis elegans, 19 000 genes Just like you, this worm possesses muscles, a nervous system, intestines and sexual organs. That is why the roundworm is used to study genes linked to aging, to neurological diseases such as Alzheimer's, to cancer and to kidney disease.	21%
	Bacterium, Escherichia coli, 4800 genes The E. coli bacterium inhabits your intestines. Researchers study it to learn about basic cell functions, such as transcription and translation. (Image: Fred Hossler, Visuals Unlimited, Inc.)	7%

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.

However, all living organisms do have ancient genes stemming from the beginning of time that humans share with every living organism. So, if humans have so much in common with other species, what is it that defines being human? What is it that turns humans into this complex being capable of learning, speaking, thinking and feeling? What is it that makes humans different from each other?



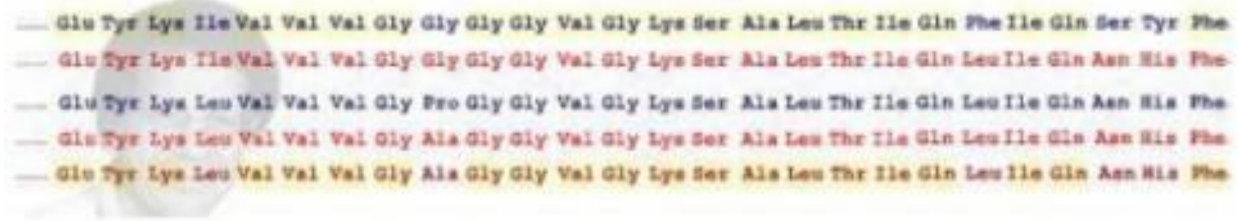
DNA Sequence Homology in the ob Gene

Human	CTGCCACTTG	CCCTGGGCCA	GTGGCCTGGA
Gorilla	CTGCCACTTG	CCCTGGGCCA	GTGGCCTGGA
Horse	CTGCCCCTTG	CCCCAGGCCA	GGGGTCTGGA
Cow	CTGCCCCTTG	CCGCAGGTCA	GGGCCCTGGA
Pig	CTGCCCCTTG	CCCCGGGCCA	GGGGCCTGGA
Dog	CTGCCCCTTG	CCCCAGGCCA	GGGCCCTGGA
Mouse	CTGCTCCCTG	CCTCAGACCA	GTGGCCTGCA

Figure 23

We have in common with a mouse or a worm more than we think!

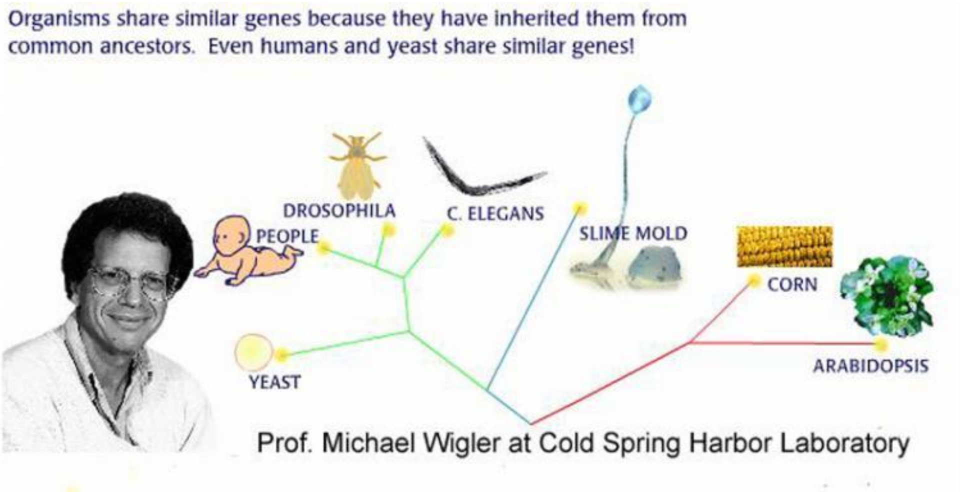
Despite appearances, we share a surprising number of genes with other species. (See above table.) Although these genes don't all have the same nucleotides in the same order, their function is similar enough for them to be considered comparable. These genes likely stem from a common ancestor, one that lived 3.5 billion years ago. Scientists theorize that through evolution this ancestor's genome became the basis for every species that we know today.



Glu Tyr Lys Ile Val Val Val Gly Gly Gly Gly Val Gly Lys Ser Ala Leu Thr Ile Gln Phe Ile Gln Ser Tyr Phe
Glu Tyr Lys Ile Val Val Val Gly Gly Gly Gly Val Gly Lys Ser Ala Leu Thr Ile Gln Leu Ile Gln Asn His Phe
Glu Tyr Lys Leu Val Val Val Gly Pro Gly Gly Val Gly Lys Ser Ala Leu Thr Ile Gln Leu Ile Gln Asn His Phe
Glu Tyr Lys Leu Val Val Val Gly Ala Gly Gly Val Gly Lys Ser Ala Leu Thr Ile Gln Leu Ile Gln Asn His Phe
Glu Tyr Lys Leu Val Val Val Gly Ala Gly Gly Val Gly Lys Ser Ala Leu Thr Ile Gln Leu Ile Gln Asn His Phe

Figure 24

That’s why composition of many genes is similar. The picture on the left shows an example for obesity (ob) gene in several different animals, where the sequences are similar. The next picture below presents even identical sequences in very different living organisms from the yeast to human beings, as shown by Dr. Michael Wigler from CSHL when stuying the yeast’s ras oncogene. He has made also a big contribution to study of molecular evolution.



Organisms share similar genes because they have inherited them from common ancestors. Even humans and yeast share similar genes!
DROSOPHILA
C. ELEGANS
PEOPLE
SLIME MOLD
CORN
YEAST
ARABIDOPSIS
Prof. Michael Wigler at Cold Spring Harbor Laboratory

Figure 25

Lecture 14. Genes can be manipulated by molecular tools I

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 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 MspI restriction site CCGG and that of EcoRI GAATTC). See the action of EcoRI.

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.

Lecture 15. Genes can be manipulated by molecular tools II

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 sequence 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 95o C), annealing (annealing of the primer to the template at 40 - 60o 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 72o 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 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); real-time PCR (detects the sequence of interest in very small quantity).

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.
7. Studying genetic diversity in species.

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.

Lecture 16. Gene and DNA analysis

PDF

As we know the knowledge of gene structure is extremely important for genemanipulation as well as for understanding basic principles of life. The common structure of a gene is shown below.

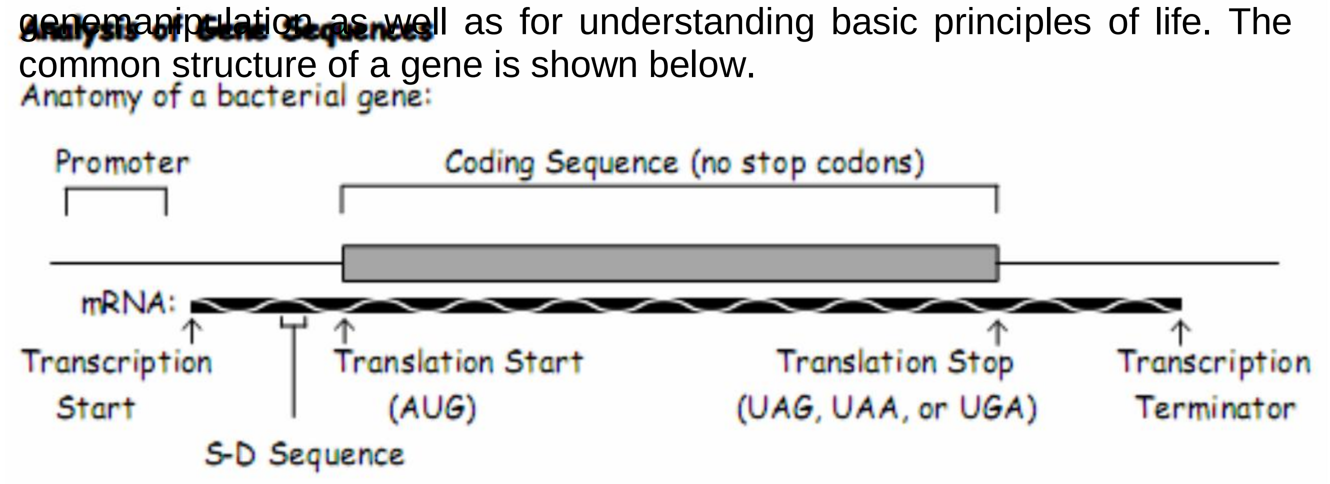

Analysis of Gene Sequences
Anatomy of a bacterial gene:
Promoter
Coding Sequence (no stop codons)
mRNA:
Transcription Start
S-D Sequence
Translation Start (AUG)
Translation Stop (UAG, UAA, or UGA)
Transcription Terminator

Figure 26

Sequence	Function
Promoter	To target RNA polymerase to DNA and to start transcription of a mRNA copy of the gene sequence.
Transcription terminator	To instruct RNA polymerase to stop transcription.
Shine-Dalgarno sequenceand	S-D sequence in mRNA will load ribosomes to begin translation. Translation almost always begins at an AUG
translation start	(an ATG in the DNA becomes an AUG in the mRNA copy). Synthesis of the protein thus begins with a methionine.
Coding Sequence	Once translation starts, the coding sequence is translated by the ribosome along with tRNAs which read three bases at a time in linear sequence. Amino acids will be incorporated into the growing polypeptide chain according to
Translation Stop	When one of the three stop codons [UAG (amber), UAA (ochre), or UGA] is encountered during translation, the polypeptide will be released from the ribosome.

Example: A gene coding sequence that is 1,200 nucleotide base pairs in length (including 1200 the ATG but not including the stop codon) will specify the sequence of a protein/= 3400 amino acids long. Since the average molecular weight of an amino acid is 110 da, this gene encodes a protein of about 44 kd, the size of an average protein. Classically, genes are identified by their function. That is, the existence of the gene is recognized because of mutations in the gene that give an observable phenotypic change. Historically, many genes have been discovered because of their effects on phenotype. Now, in the era of genomic sequencing, many genes of no known function can be detected by looking for patterns in DNA sequences. The simplest method which works for bacterial and phage genes (but not for most eukaryotic genes as we will see later) is to look for stretches of sequence that lack stop codons. These are known as open reading frames or ORFs. This works because a random sequence should contain an average of one stop codon in every 21 codons. Thus, the probability of a random occurrence of even a short open reading frame of say 100 codons without a stop codon is very small (61/ 64)100 = 8.2 x 10-3

Identifying genes in DNA sequences from higher organisms is usually more difficult than in bacteria. This is because in humans, for example, gene coding sequences are separated by long sequences that do not code for proteins. Moreover, genes of higher eukaryotes intronsintrons are interrupted by introns, which are sequences that are spliced out of the NA before intronsintrons translation. The presence of introns breaks up the open reading frames into short segments, making them much harder to distinguish from non-coding sequences. The maps below show 50 kbp segments of DNA from yeast, Drosophila, and humans. The dark grey boxes represent coding sequences and the light grey boxes represent introns. The boxes above the line are transcribed to the right and the boxes below are transcribed to the left. Names have been assigned to each of the identified genes. Although the yeast genes are much like those of bacteria (few introns and packed closely together), the Drosophila and human genes are spread apart and interrupted by many introns. Sophisticated computer algorithms were used to identify these dispersed gene sequences.

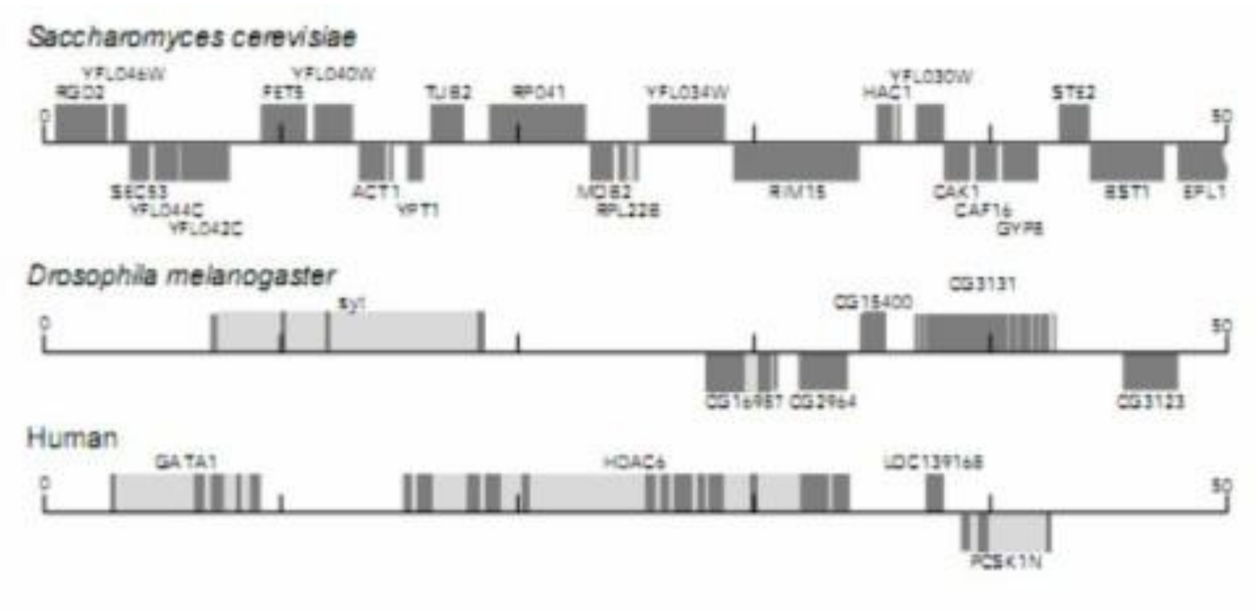


Figure 27

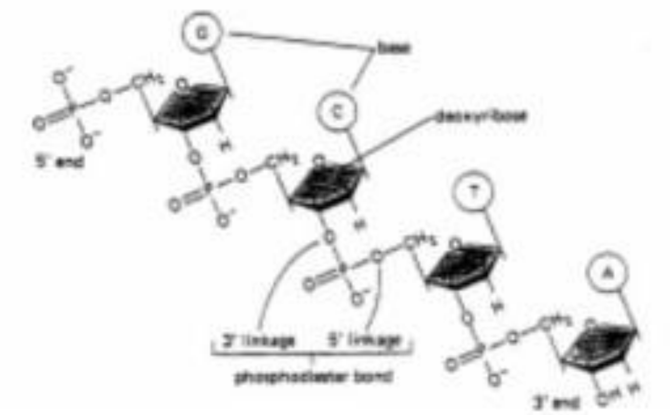


Figure 28

To see how gene sequences are actually obtained, we will first need to consider some fundamentals of the chemical structure of DNA. Each strand of DNA is directional. The different ends are usually called the 5 and 3 ends, referring to different positions on the ribose sugar ring where the linking phosphate residues attach.

In a double stranded DNA molecule the two strands run anti-parallel to one another and

the general structure can be diagramed like this:

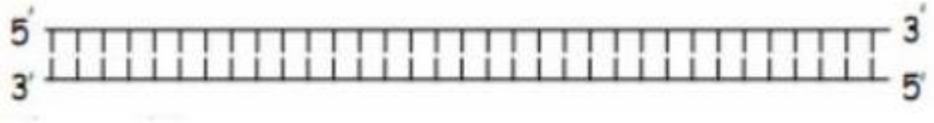


Figure 29

Note about representation of DNA sequences:

1) Single strands are always represented in direction of synthesis 5’ to 3’.

2) For double stranded DNA, usually one strand is represented in the 5’

to 3’ direction. For a gene, the strand represented would correspond to the sequence of the mRNA.

DNA polymersaes are the key players in the methods that we will be considering. The general reaction carried out by DNA polymerase is to synthesize a copy of a DNA template, starting with the chemical precursors (nucleotides) dATP, dGTP, dCTP, and dTTP (dNTPs).

All DNA polymerases have two fundamental properties in common:

(1) New DNA is synthesized only by elongation of an existing strand at its 3 end.

(2) Synthesis requires nucleotide precursors, a free 3 OH end, and a template strand.

A general substrate for DNA polymerase looks like this:

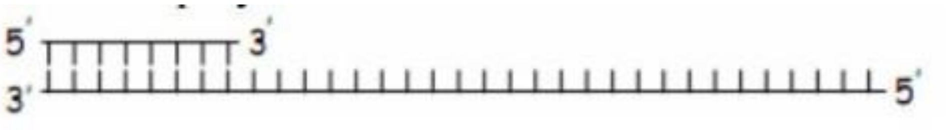


Figure 30

Note that the template strand can be as short as 1 base or as long as several thousand

bases. After addition of DNA polymerase and nucleotide precursors, this product will be readily synthesized:

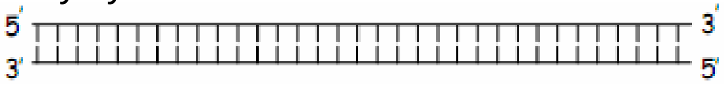


Figure 31

DNA Sequencing Consider a segment of DNA that is about 1000 base pairs long that we wish to sequence.

(1) The two DNA strands are separated. Heating to 100C to melt the base pairing hydrogen bonds that hold the strands together does this.

(2) A short oligonucleotide (ca. 18 bases) designed to be complimentary to the end of one of the strands is allowed to anneal to the single stranded DNA. The resulting DNA hybrid looks much like the general polymerase substrate shown previously.

(3) DNA polymerase is added along with the four nucleotide precursors (dATP, dGTP, dCTP, and dTTP). The mixture is then divided into four separate reactions and to each reaction a small quantity different dideoxy nucleotide precursor is added. Dideoxy nucleotide precursors are abbreviated ddATP, ddGTP, ddCTP, and ddTTP.

(4) The polymerase reactions are allowed to proceed and, using one of a variety of methods, radiolabel is incorporated into the newly synthesized DNA.

(5) After the DNA polymerase reactions are complete, the samples are melted and run on a gel system that allows DNA strands of different lengths to be resolved. The DNA sequence can be read from the gel by noting the positions of the radiolabeled fragments. The crucial element of the

sequencing reactions is the added dideoxynuclotides. These molecules are identical to the normal nucleotide precursors in all respects except that

they lack a hydroxyl group at their 3’ position (3’ OH).

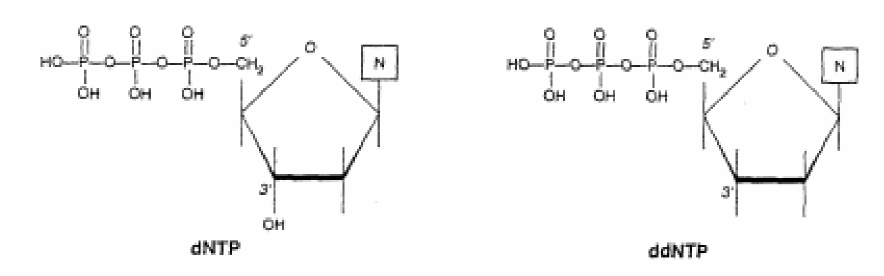


Figure 32

Thus dideoxynuclotides can be incorporated into DNA, but once a dideoxynuclotide has been incorporated, further elongation stops because the resulting DNA will no longer have a free 3 OH end. Each of the four reactions contains one of the dideoxynuclotides added at about 1% the concentration of the normal nucleotide precursors. Thus, for example, in the reaction with added ddATP, about 1% of the elongated chains will terminate at the position of each A in the sequence. Once all of the elongating chains have been terminated, there will be a population of labeled chains that have terminated at the position of each A in the sequence. A part of the final gel will look like this:

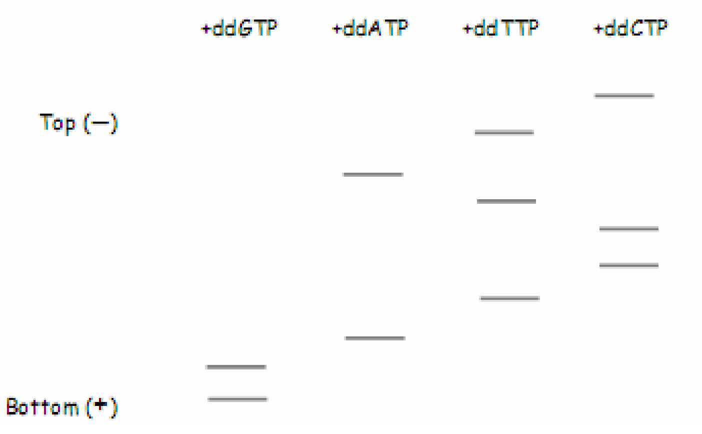


Figure 33

(Note that larger molecules migrate more slowly to the cathode on these gels)?

The deduced DNA sequence obtained from this gel is: 5’ GGATCCTATC 3’?

Polymerase Chain Reaction Now let’s consider how to obtain DNA segments that are suitable for sequencing. At first, DNA sequences were obtained from cloned DNA segments. (We will discuss some methods to clone new genes in a subsequent lecture.) Presently the entire DNA sequence for E. coli, as well as a variety of other bacterial species, has been determined. If we want to find the sequence of a new mutant allele of a known gene, we need an easy way to obtain a quantity of this DNA from a culture of bacterial cells. The best way to do this is to use a method known as PCR or polymerase chain reaction that was developed by Kary Mullis in the mid-1980s. The steps in a PCR reaction are as follows:

(1) A crude preparation of chromosomal DNA is extracted from the bacterial strain of interest.

(2) Two short oligo nucleotide primers (each about 18 bases long) are added to the DNA. The primers are designed from the known genomic sequence to be complimentary to opposite strands of DNA and to flank the chromosomal segment of interest.

(3) The double stranded DNA is melted by heating to 100C and then the mixture is cooled to allow the primers to anneal to the template DNA.

(4) DNA polymerase and the four nucleotide precursors are added, and the reaction is incubated at 370C for a period of time to allow a copy of the segment to be synthesized.

(5) Steps 3 and 4 are repeated multiple times. To avoid the inconvenience of having to add new DNA polymerase in each cycle, a special DNA polymerase that can withstand heating to 1000C is used.

The idea is that in each cycle of melting, annealing and DNA synthesis, the amount of the DNA segment is doubled. This gives an exponential increase in the amount of the specific DNA as the cycles proceed. After 10 cycles the DNA is amplified 103 fold and after 20 cycles the DNA will be amplified 106 fold. Usually amplification is continued until all of the nucleotide precursors are incorporated into synthesized DNA.

Lecture 17. Epigenetics as a way to control gene expression

Epigenetics refers to the study of heritable changes in gene expression that occur without a change in DNA sequence. Research has shown that epigenetic mechanisms provide an "extra" layer of transcriptional control that regulates how genes are expressed.

These mechanisms are critical components in the normal development and growth of cells. Epigenetic abnormalities have been found to be causative factors in cancer, genetic disorders and pediatric syndromes as well as contributing factors in autoimmune diseases and aging. This lecture note introduces the basic principles of epigenetic mechanisms and their contribution to human health as well as the clinical consequences of epigenetic errors; also the use of epigenetic pathways in new approaches to diagnosis and targeted treatments across the clinical spectrum.

This new field will have an enormous impact on medicine, specifically on the study of heritable changes in gene function that do not change the DNA sequence but, rather, provide an "extra" layer of transcriptional control that regulates how genes are expressed. This rapidly evolving field offers exciting new opportunities for the diagnosis and treatment of complex clinical disorders. Basic principles of epigenetics are DNA methylation and histone modifications.

DNA methylation and histone modifications

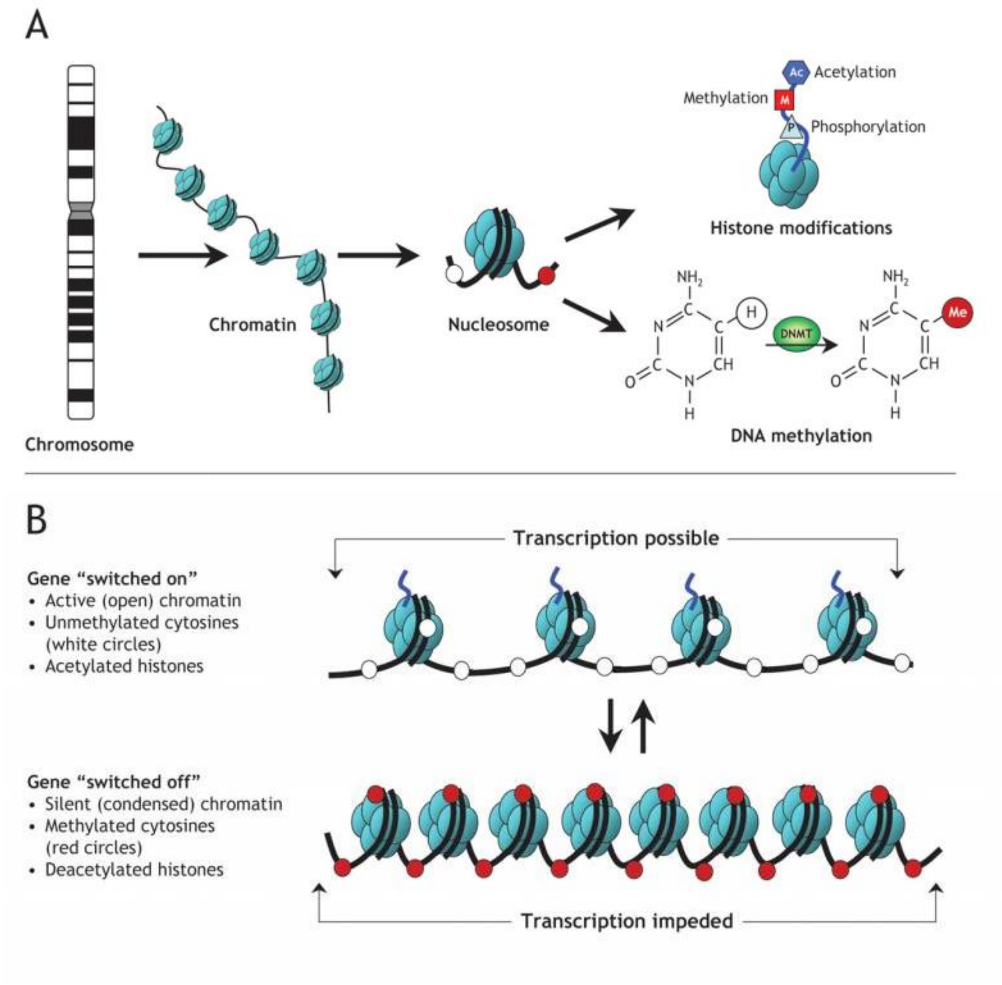


Figure 34

(A) Schematic of epigenetic modifications. Strands of DNA are wrapped around histone octamers, forming nucleosomes, which to be organized into chromatin, the building block of a chromosome. Reversible and site-specific histone modifications occur at multiple sites through acetylation, methylation and phosphorylation. DNA methylation occurs at 5-position of cytosine residues in a reaction catalyzed by DNA methyltransferases (DNMTs). Together, these modifications provide a unique epigenetic signature that regulates chromatin organization and gene expression. (B) Schematic of the reversible changes in chromatin organization that influence gene expression: genes are expressed (switched on) when the chromatin is open (active), and they are inactivated (switched off) when the chromatin is condensed (silent). White circles = unmethylated cytosines; red circles = methylated cytosines.

Clinical consequences of epigenetic errors Epigenetic mechanisms regulate DNA accessibility throughout a person's lifetime. Immediately following fertilization, the paternal genome undergoes rapid DNA demethylation and histone modifications.27 The maternal genome is demethylated gradually, and eventually a new wave of embryonic methylation is initiated that establishes the blueprint for the tissues of the developing embryo. As a result, each cell has its own epigenetic pattern that must be carefully maintained to regulate proper gene expression. Perturbations in these carefully arranged patterns of DNA methylation and histone modifications can lead to congenital disorders and multisystem pediatric syndromes or predispose people to acquired disease states such as sporadic cancers and neurodegenerative disorders.

Aging Both increases and decreases in DNA methylation are associated with the aging process, and evidence is accumulating that age-dependent methylation changes are involved in the development of neurologic disorders, autoimmunity and cancer in elderly people.88 Methylation changes that occur in an age-related manner may include the inactivation of cancer-related genes. In some tissues, levels of methylated cytosines decrease in aging cells, and this demethylation may promote chromosomal instability and rearrangements, which increases the risk of neoplasia.88 In other tissues, such as the intestinal crypts, increased global hypermethylation may be the predisposing event that accounts for the increased risk of colon cancer with advancing age.89

Cancer and epigenetic therapies Cancer is a multistep process in which genetic and epigenetic errors accumulate and transform a normal cell into an invasive or metastatic tumour cell. Altered DNA methylation patterns change the expression of cancer- associated genes. DNA hypomethylation activates oncogenes and initiates chromosome instability,78,79,80 whereas DNA hypermethylation initiates silencing of tumour suppressor genes. The incidence of hypermethylation, particularly in sporadic cancers, varies with respect to the gene involved and the tumour type in which the event occurs.

To date, epigenetic therapies are few in number, but several are currently being studied in clinical trials or have been approved for specific cancer types.1,82,83 Nucleoside analogues such as azacitidine are incorporated into replicating DNA, inhibit methylation and reactivate previously silenced genes.84 Azacitidine has been effective in phase I clinical trials in treating myelodysplastic syndrome and leukemias characterized by gene hypermethylation. The antisense oligonucleotide MG98 that downregulates DNMT1 is showing promising results in phase I clinical trials86 and in targeting solid tumours and renal cell cancer (www.methylgene.com/content.asp?node=14 [accessed 2005 Dec 22]). Similarly, small molecules such as valproic acid that downregulate HDACs are being used to induce growth arrest and tumour cell death. Combination epigenetic therapies (demethylating agents plus HDAC inhibitors) or epigenetic therapy followed by conventional chemotherapy (or immunotherapy) may be more effective since they reactivate silenced genes, including tumour suppressor genes, resensitize drug-resistant cells to standard therapies and act synergistically to kill cancer cells.1,82,87

The road ahead Our increased knowledge of epigenetic mechanisms over the last 10 years is beginning to be translated into new approaches to molecular diagnosis and targeted treatments across the clinical spectrum. With the Human Genome Project completed, the Human Epigenome Project has been proposed and will generate genome- wide methylation maps.106 By examining both healthy and diseased tissues, specific genomic regions will be identified that are involved in development, tissue-specific expression, environmental susceptibility and pathogenesis. Use of these epigenetic maps will lead to epigenetic therapies for complex disorders across the clinical spectrum. Comments, questions, feedback, criticisms?

Send feedback

E-mail the author

E-mail Vietnam OpenCourseWare

SOME RIGHTS RESERVED

More about this content: Metadata | Version History | Cite This Content

This work is licensed by Professor Le Dinh Luong. See the Creative Commons License about permission to reuse this material.

Last edited by Professor Le Dinh Luong on May 4, 2009 10:47 am GMT-5.