Genetics

CONNEXIONS
CNX.ORG

Genetics

By: Professor Le Dinh Luong

Online: <http://cnx.org/content/col10782/1.1>

Genetics

Table of Contents

Chapter 1. Syllabus

Text

There are no assigned readings for this class although we recommend the following textbooks as valuable references:

1. Griffiths, Anthony J. F., Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, and William M. Gelbart. An Introduction to Genetic Analysis. 7th ed. New York: W. H. Freeman, 2000. ISBN: 9780716735205.

2. Egger G, Liang G, Aparicio A, et al. Epigenetics in human disease and prospects for epigenetic therapy. Nature 2004;429:457-63.

3. Principles of genetics: A textbook, with problems (McGraw-Hill publications in the agricultural and botanical sciences).

Assignments and Exams

There are seven graded problem sets for this course. Students may collaborate with classmates on the problem sets, but copying problem set solutions is not permitted. Any student who copies another problem set or allows his or her problem set to be copied will be assigned a 0 for that problem set.

There are three one-hour exams. The exams will be closed book, but students may bring one 8 1/2 x 11 sheet of notes to the exam. In addition to the exams, there will also be a final during exam week. The final will be comprehensive and will cover material from the entire course with an emphasis on material of the lecture 31 not covered by an hour exam.

Grading

Table 1.1.

Table for Grading	
ACTIVITIES	POINTS
Quiz I	100
Quiz II	100

Quiz III	100
Final	200
Seven Problem Sets	140
Total	640

Solutions

Chapter 2. Basic Principles of Genetics

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.

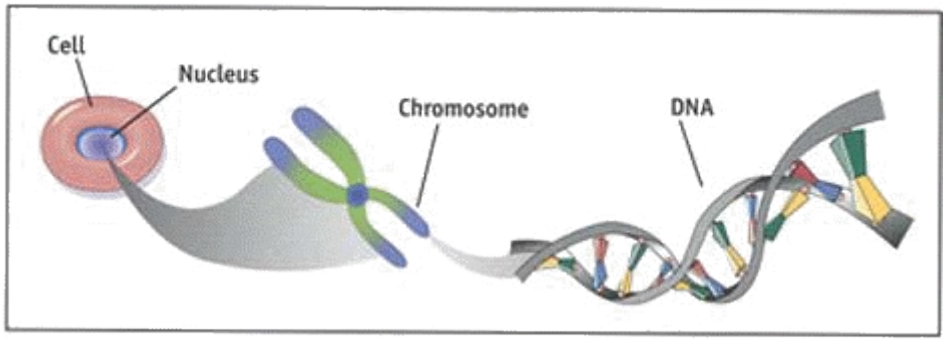


Figure 2.1.

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.

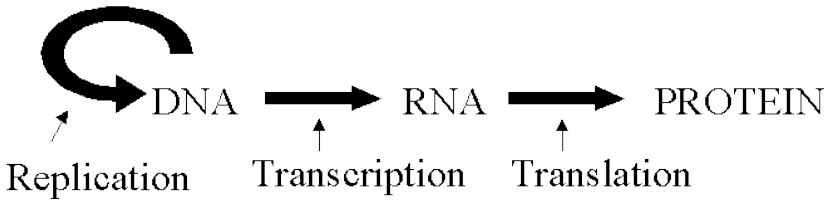


Figure 2.2.

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.

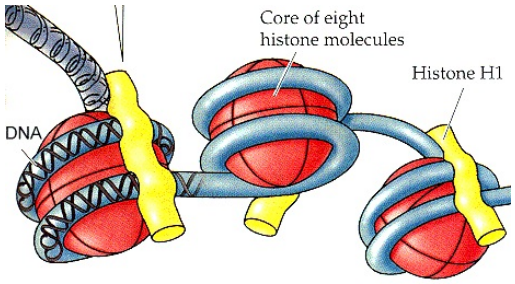


Figure 2.3.

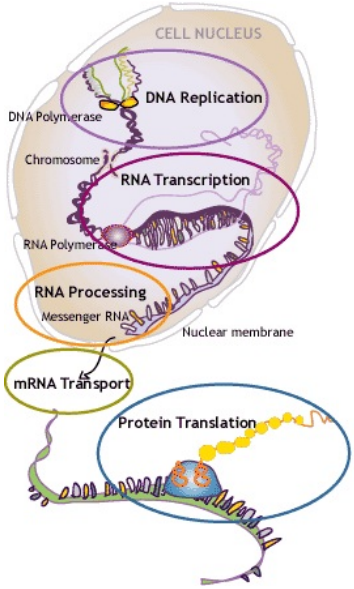


Figure 2.4.

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.

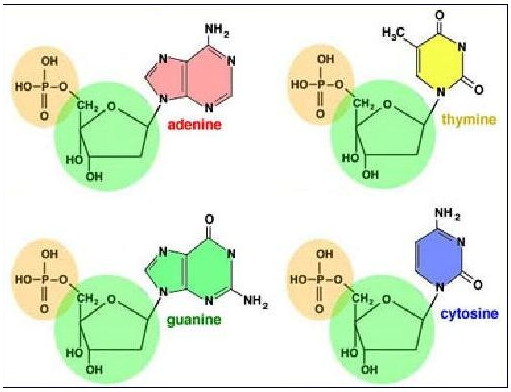


Figure 2.5.

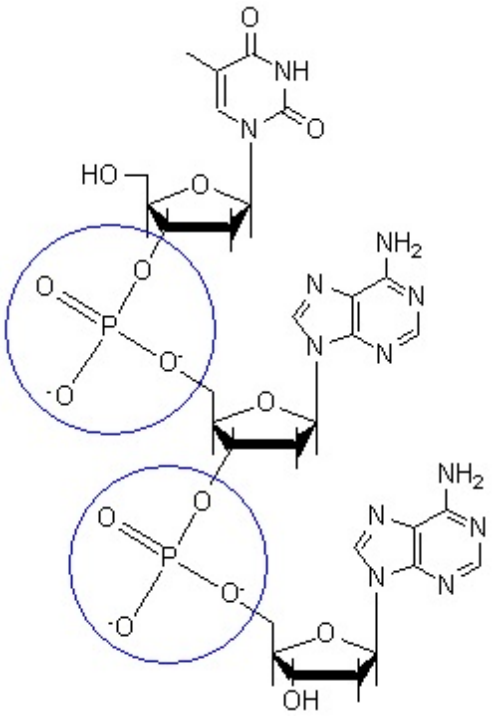


Figure 2.6.

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).

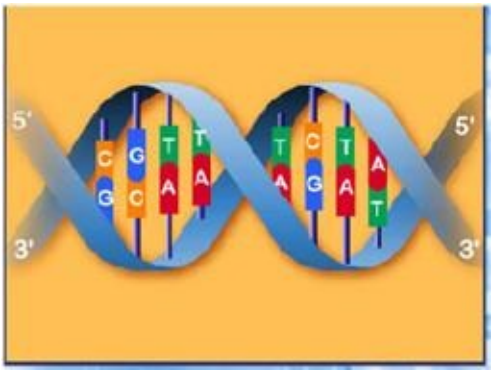


Figure 2.7.

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.

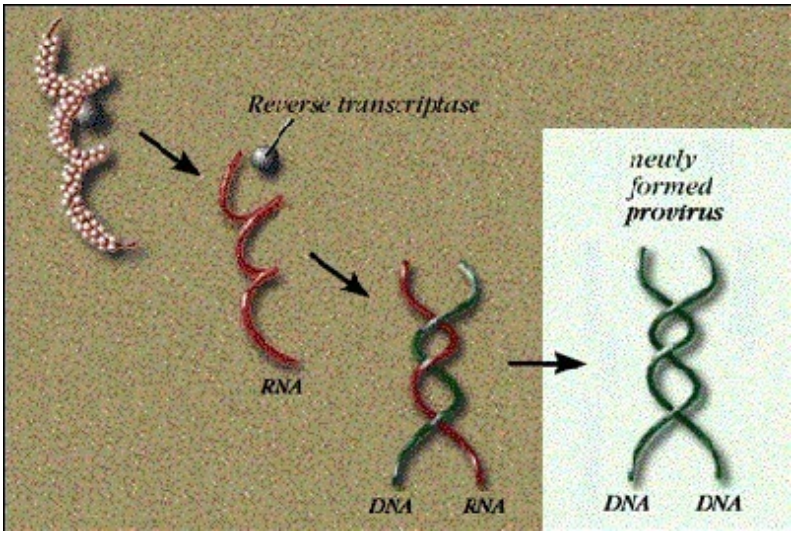


Figure 2.8.

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

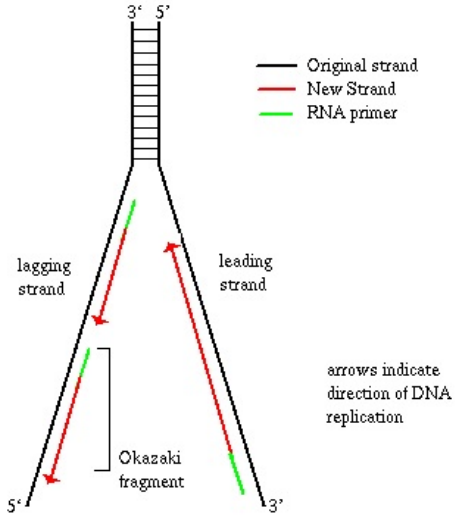


Figure 2.9.

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 strand in a 5' 3' direction. 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 two DNA molecules. Each consists of an original nucleotide strand next to a new complementary strand. The two molecules are identical to each other.

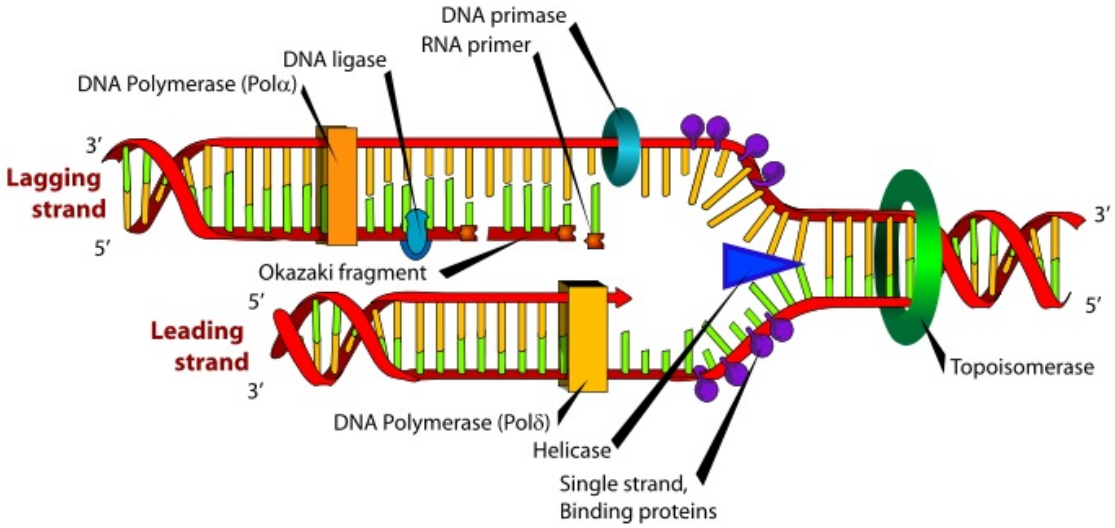


Figure 2.10.

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

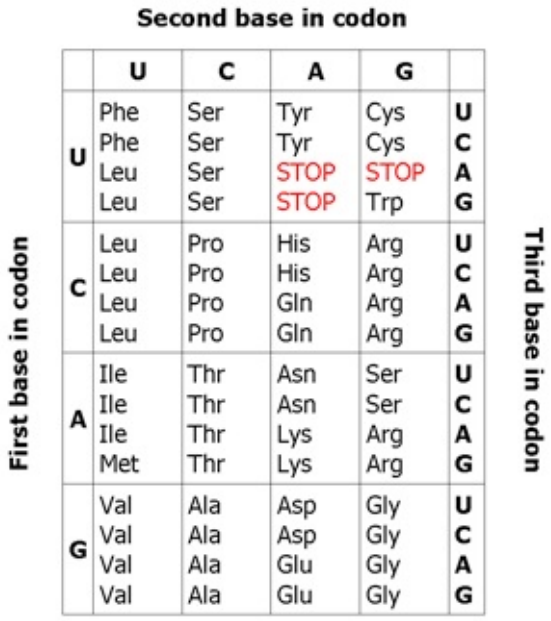
Genetic information likes a language. We use letters of the alphabet to make words and then join these words together to make sentences, paragraphs and books. In the case of **DNA**:

- 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:

Table 2.1.

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**.



	U	C	A	G	
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 2.11.

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 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.

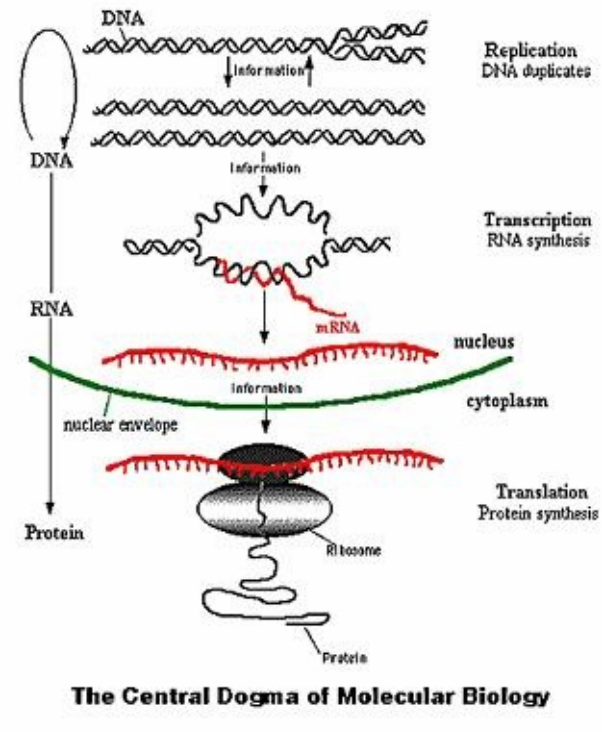


Figure 2.12.

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 2.13.

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, 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, 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.

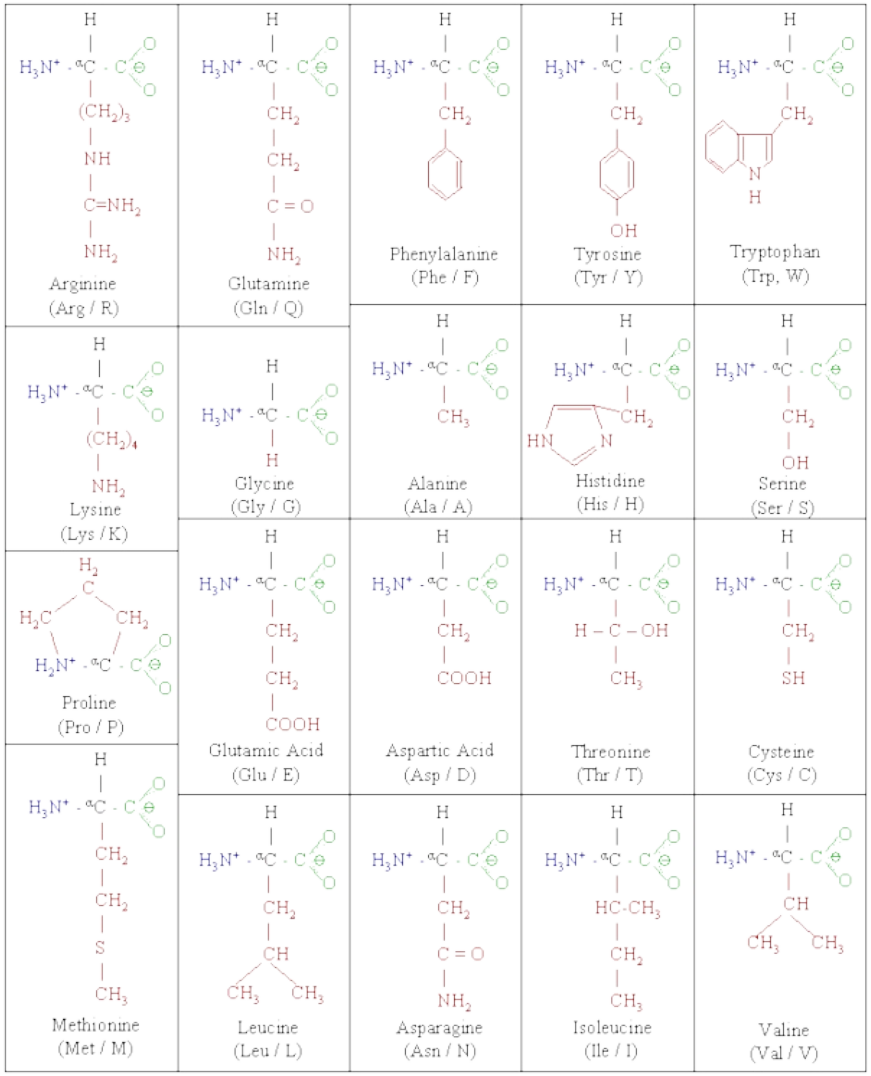


Figure 2.14.

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.

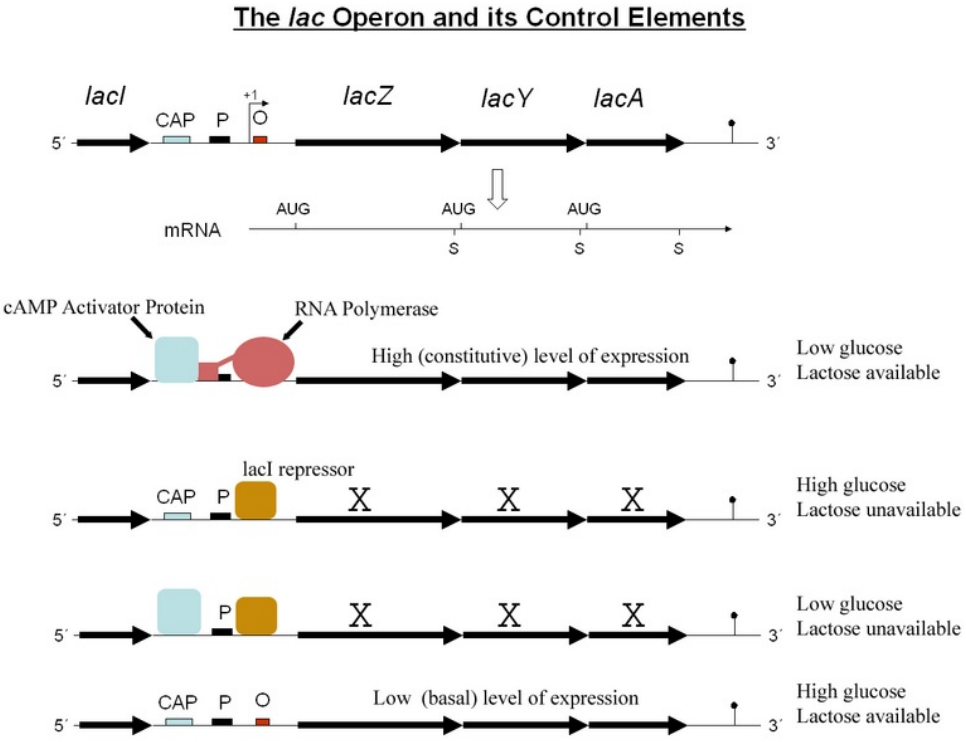


Figure 2.15.

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.

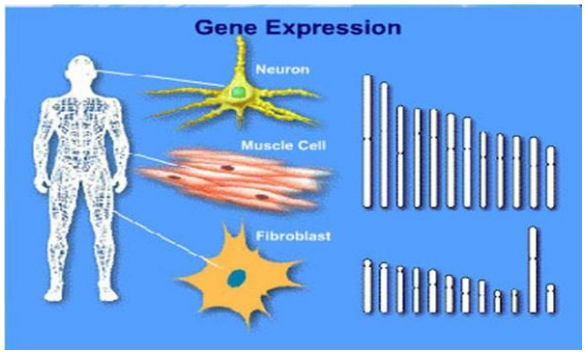


Figure 2.16.

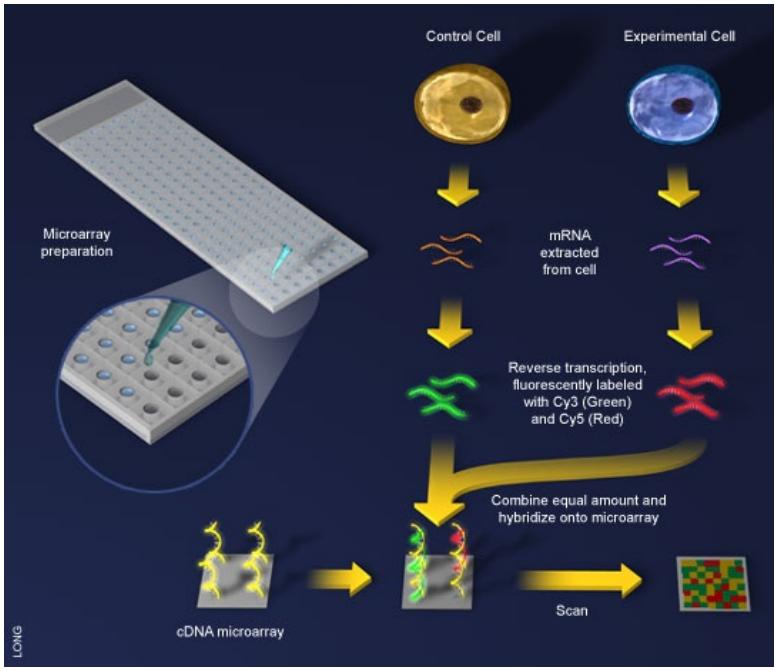


Figure 2.17.

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.

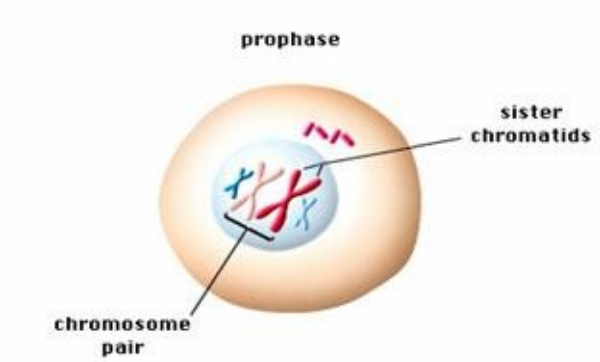


Figure 2.18.

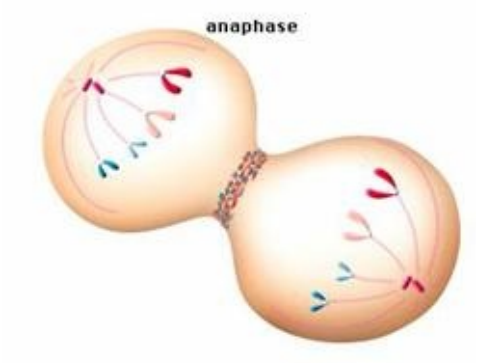


Figure 2.19.

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.

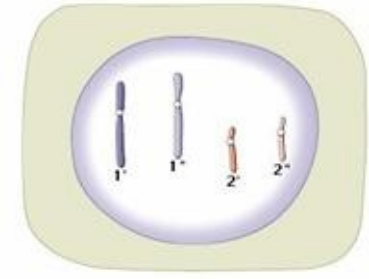


Figure 2.20.

Figure 2.21.

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.

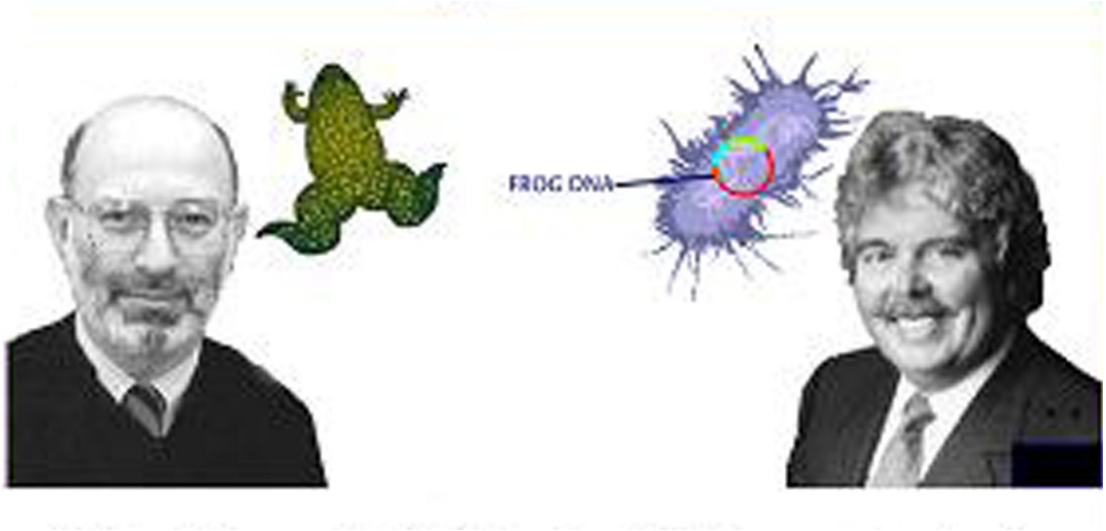


Figure 2.22.

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

Table 2.2.

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 1995[6]
Bacterium, Carsonella ruddii	160,000	Smallest non-viral genome.[7]
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 genome.[9]
Plant, Arabidopsis thaliana	157,000,000	First plant genome sequenced, Dec 2000.[10]
Plant, Genlisea margaretae	63,400,000	Smallest recorded flowering plant genome, 2006.[10]
Plant, Fritillaria assyrica	130,000,000,000	
Plant, Populus trichocarpa	480,000,000	First tree genome, Sept 2006
moss, Physcomitrella patens	480,000,000	First genome of a bryophyte, January 2008 [11]
Yeast,Saccharomyces cerevisiae	12,100,000	[12]
Fungus, Aspergillus nidulans	30,000,000	
Nematode, Caenorhabditis elegans	98,000,000	First multicellular animal genome, December 1998[13]
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	

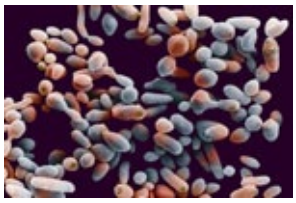
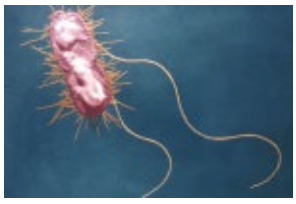
Fish, **Protopterus aethiopicus** aka **Marbled lungfish**	130,000,000,000	Largest vertebrate genome known

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.

Table 2.3.

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?

Figure 2.23.

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.

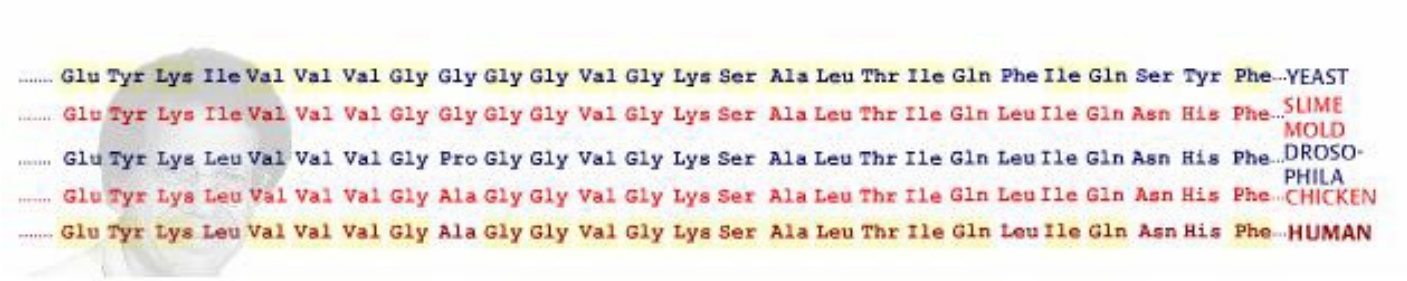


Figure 2.24.

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.

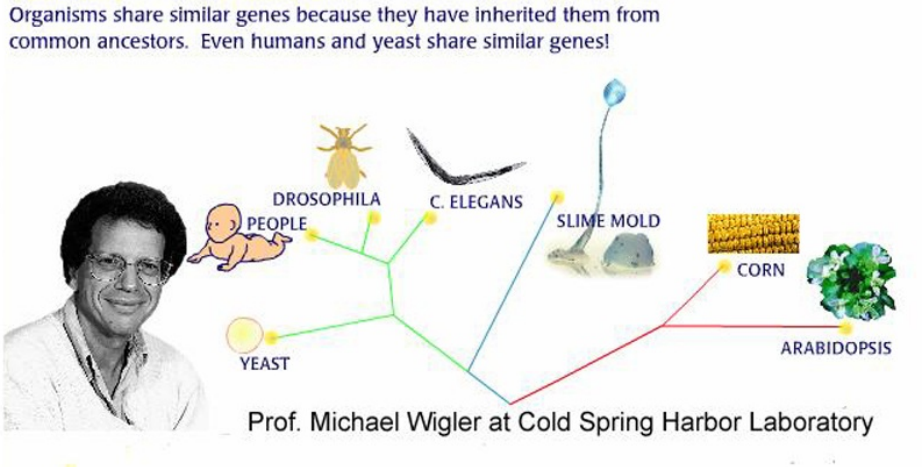


Figure 2.25.

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 gene manipulation as well as for understanding basic principles of life. The common structure of a gene is shown below.

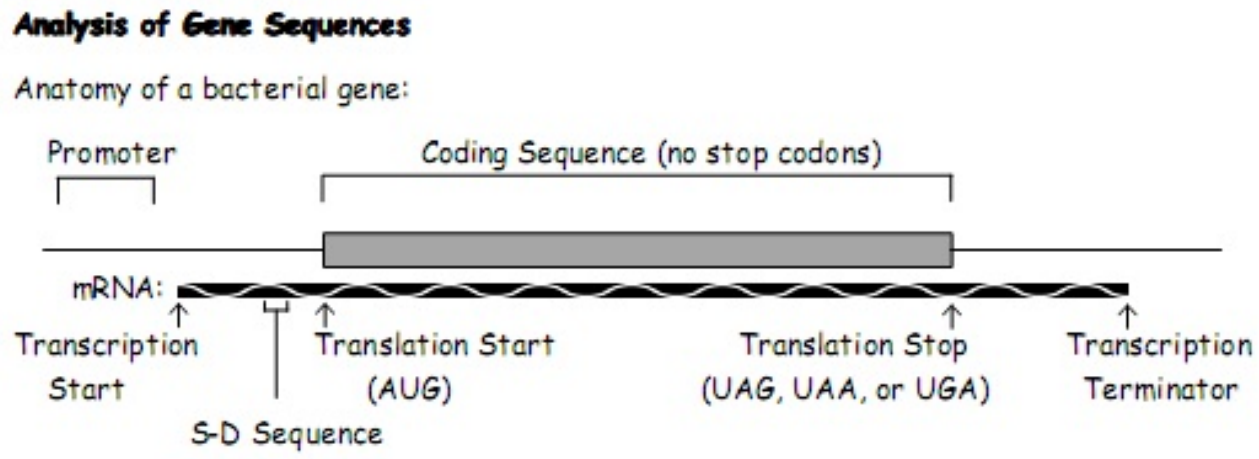


Figure 2.26.

Table 2.4.

Sequence Element	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	S-D sequence in mRNA will load ribosomes to begin translation. Translation almost

sequenceand translation start	always begins at an AUG codon in the mRNA (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 the genetic code.
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.

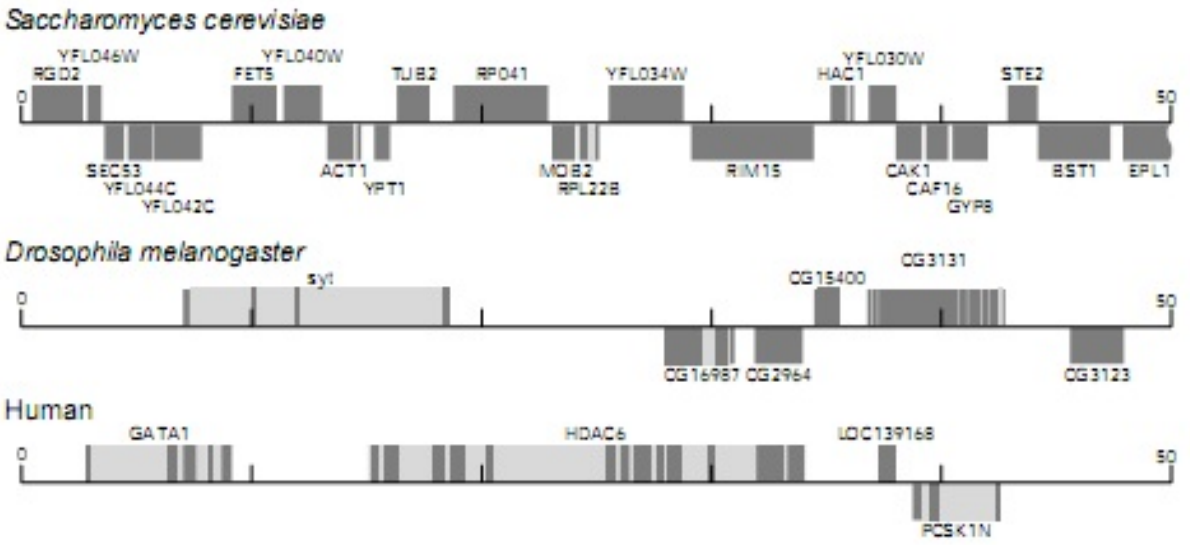


Figure 2.27.

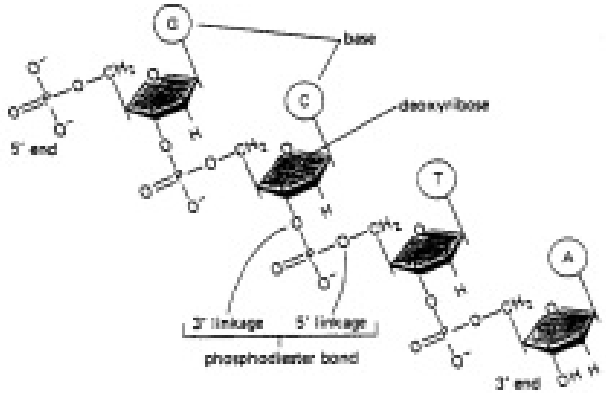


Figure 2.28.

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:

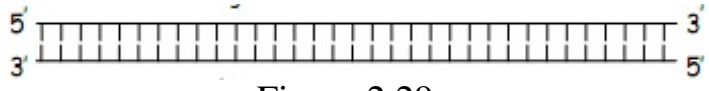


Figure 2.29.

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:

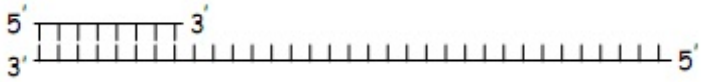


Figure 2.30.

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 2.31.

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).

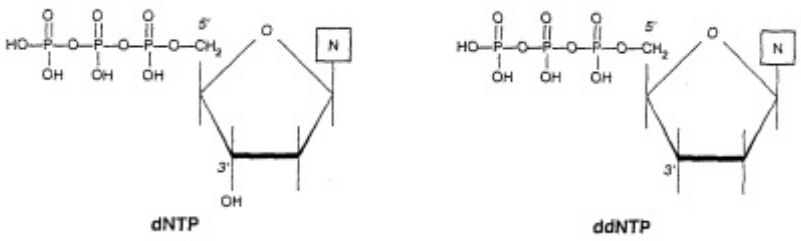


Figure 2.32.

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 2.33.

(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

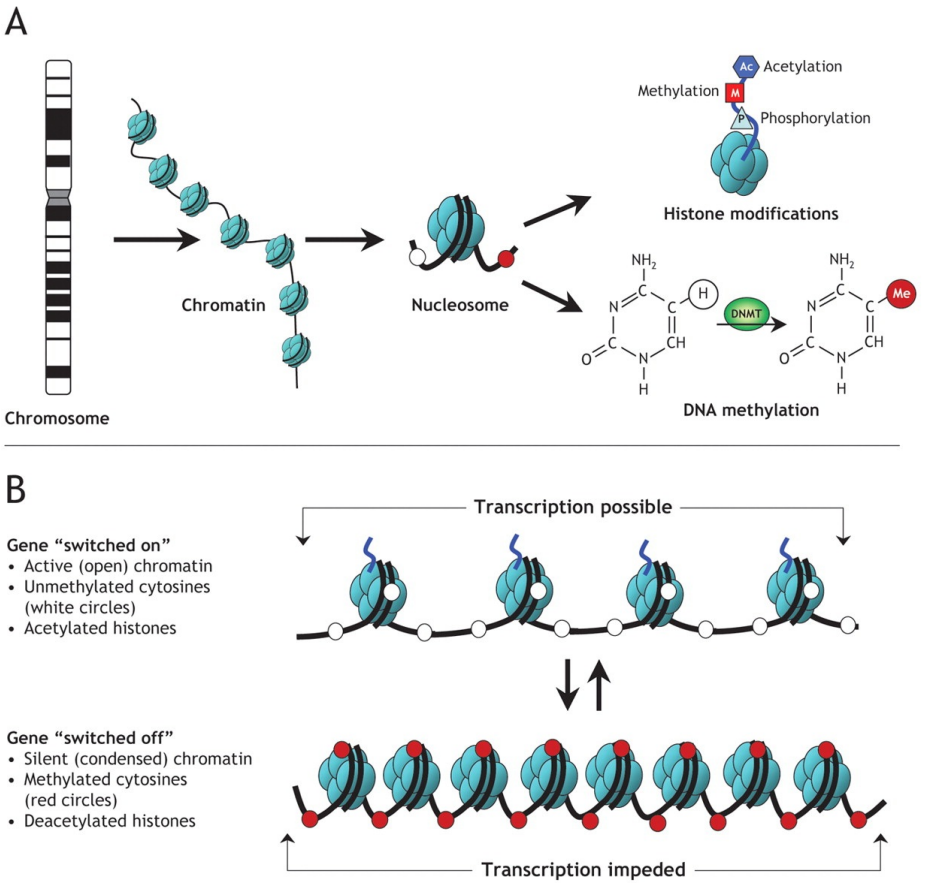


Figure 2.34.

(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 trials[86] 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.

Solutions

Chapter 3. Eukaryotic Genetics

Lecture 25. Characteristics of eukaryotic genes

Eukaryotic organisms have essential differences in cell structure compared with prokaryotic ones. Eukaryotes have typical cell structure, mitosis and meiosis. That's why their structure of gene and genome is different from prokaryotic genetic machinery.

The Differences between Eukaryotic and Prokaryotic Genes

Unlike Prokaryotes, Eukaryotes:

- have chromosomes
- contain a nucleus
- have amounts of DNA that differ between species
- have variations in the number of chromosomes between species
- genes contain introns
- (parallet structure….."have genes containing introns")
- may have multiple copies of a gene

There is great divergence of sequence between a given intron in different eukaryotic organisms. The exon sequences are much more conserved. This suggests that the actual sequence of the intron is not very important. If it were important, then any changes that occurred during evolution would be damaging, and the organisms with the changes would not be likely to survive.

RNA Splicing

The DNA in eukaryotes is organized into exons and introns. The introns do not carry any genetic information. The process of RNA splicing is responsible for removing introns from precursor RNAs to produce the final RNA product. In the process from pre-mRNA to mRNA, splicing must be extremely accurate. If splicing is off by one nucleotide, the entire coding will be messed up because all of the codons downstream of the mistake will be out of the correct reading frame--they will be out of phase.

RNA splicing is carried out by snRNPs which stands for small nuclear RNA containing ribonucleoprotein particles. The snRNPs contain both RNA and proteins. (Each snRNP contains a molecule of snRNA.) In this respect they are very similar to ribosomes, another RNP particle in the cell.

In snRNPs, the RNA carries out enzymatic duties, and the proteins hold the snRNPs in the correct configuration to stabilize them.

The role of snRNPs

The snRNAs in the snRNPs base pair with the pre-mRNA at splice junctions (and some other sites too). The snRNPs base paired at different splice junctions interact with each other to facilitate the removal of the intron between the snRNPs and to join the adjacent exons.

There is an evolutionary benefit to having introns; otherwise, the energy cost to splice would not be compensated.

Sometimes splicing skips over an exon. For example say the pre-mRNA contains A-B-C-D exons. Splicing in some tissues might lead to an A-B-D mRNA (exon C is skipped). Or the splicing could produce an A-C-D mRNA (exon B is skipped). These mRNAs would have the same end exons but different middles. They will code for different proteins. This alternative splicing uses genetic expression to facilitate the synthesis of a greater variety of proteins.

Globin Genes

Globin genes are an example of products of alternative splicing. Globins (combined with heme) bind oxygen. All globin genes have three exons and two introns. The functional protein, called hemoglobin, consists of 4 molecules of globin protein and a single molecule of heme. Human adults have two alpha-globins and two beta-globins in our hemoglobin.

Myoglobin consists of a single globin subunit plus heme and carries oxygen within muscles. Because of their similar sequence and gene organization (both have three exons in exactly the same location along the gene), it is believed that both the globin and myoglobin are derived from a common ancestor gene.

Plants called legumes have the ability to use certain kinds of bacteria as a means of getting their needed nitrogen through a process of nitrogen fixation. An example is soybeans. The roots develop a sac where bacteria can fix nitrogen. The bacteria and the plant have a symbiotic relationship; the plant provides the bacteria with food, and the bacteria fixes nitrogen for the plant. Leghemoglobin is crucial in this process because it binds oxygen within the sac which allows the bacteria to fix nitrogen. The bacteria cannot function in the presence of oxygen. The sequence of leghemoglobin is related to the sequence of the other globins, but, interestingly, the middle exon is split in leghemoglobin, giving this particular globin gene 4 exons. Since the gene organization is close to that of the rest of the globin family and the protein sequence of leghemoglobin and globin are related, it is clear that these genes all share a common ancestor. It is not known if the ancestor had three or four exons.

The characteristics of eukaryotic genes and genomes have been very well considered in MITOPENCOURSEWARE (PDF), especially in model eukaryotic organisms, the yeast Saccharomyces cerevisiae and the mouse Mus musculus.

Lecture 26. Gene regulation in eukaryotes

(http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/P/Promoter.html)

Because of essential differences in eukaryotic gene and genome structures compared with those of prokaryotes, as described in the above lecture, there are a number of ways that gene regulation in eukaryotes differs from gene regulation in prokaryotes.

Eukaryotic genes are not organized into operons. Eukaryotic regulatory genes are not usually linked to the genes they regulate. Some of the regulatory proteins must ultimately be compartmentalized to the nucleus, even when signaling begins at the cell membrane or in the cytoplasm. Eukaryotic DNA is wrapped around nucleosomes.

Now we will consider how one can use genetics to begin analysis of the mechanisms by which eukaryotic gene expression can be regulated.

The latest estimates are that a human cell, a eukaryotic cell, contains 20,000–25,000 genes.

- Some of these are expressed in all cells all the time. These so-called housekeeping genes are responsible for the routine metabolic functions (e.g. respiration) common to all cells.
- Some are expressed as a cell enters a particular pathway of differentiation.
- Some are expressed all the time in only those cells that have differentiated in a particular way. For example, a plasma cell expresses continuously the genes for the antibody it synthesizes.
- Some are expressed only as conditions around and in the cell change. For example, the arrival of a hormone may turn on (or off) certain genes in that cell.

How is gene expression regulated?

There are several methods used by eukaryotes.

- Altering the rate of transcription of the gene. This is the most important and widely-used strategy and the one we shall examine here.
- However, eukaryotes supplement transcriptional regulation with several other methods:
  - Altering the rate at which RNA transcripts are processed while still within the nucleus. [Discussion of RNA processing]
  - Altering the stability of mRNA molecules, that is, the rate at which they are degraded [Link to discussion of RNA interference].
  - Altering the efficiency at which the ribosomes translate the mRNA into a polypeptide. [Examples]

Protein-coding genes have:

- exons whose sequence encodes the polypeptide;

- introns that will be removed from the mRNA before it is translated [Discussion];
- a transcription start site;
- a promoter;
  - the basal or core promoter located within about 40 bp of the start site
  - an "upstream" promoter, which may extend over as many as 200 bp farther upstream

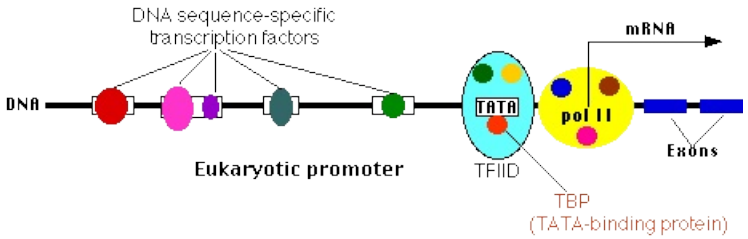


Figure 3.1.

- enhancers;
- silencers.

Adjacent genes

Adjacent genes (RNA-coding as well as protein-coding) are often separated by an **insulator** which helps them avoid cross-talk between each other's promoters and enhancers (and/or silencers).

Transcription start site

This is where a molecule of **RNA polymerase II** (pol II, also known as RNAP II) binds. Pol II is a complex of 12 different proteins (shown in the figure in yellow with small colored circles superimposed on it).

The start site is where **transcription** of the gene into RNA begins.

The basal promoter

The basal promoter contains a sequence of 7 bases (TATAAAA) called the TATA box. It is bound by a large complex of some 50 different proteins, including:

- Transcription Factor IID (TFIID) which is a complex of
  - TATA-binding protein (TBP), which recognizes and binds to the TATA box
  - 14 other protein factors which bind to TBP — and each other — but not to the DNA.
- Transcription Factor IIB (TFIIB) which binds both the DNA and pol II.

The basal or core promoter is found in all protein-coding genes. This is in sharp contrast to the upstream promoter whose structure and associated binding factors differ from gene to gene.

Although the figure is drawn as a straight line, the binding of transcription factors to each other probably draws the DNA of the promoter into a loop.

Many different genes and many different types of cells share the same transcription factors — not only those that bind at the basal promoter but even some of those that bind upstream. What turns on a particular gene in a particular cell is probably the unique combination of promoter sites and the transcription factors that are chosen.

An Analogy

The rows of lock boxes in a bank provide a useful analogy.

To open any particular box in the room requires two keys:

- your key, whose pattern of notches fits only the lock of the box assigned to you (= the upstream promoter), but which cannot unlock the box without
- a key carried by a bank employee that can activate the unlocking mechanism of any box (= the basal promoter) but cannot by itself open any box.

Link to a discussion of how the DNA sequence of promoter sites can be determined.

Transcription factors represent only a small fraction of the proteins in a cell. **Link to a discussion of how they can nonetheless be isolated and purified.**

Hormones exert many of their effects by forming transcription factors.

The complexes of hormones with their receptor represent one class of transcription factor. Hormone "response elements", to which the complex binds, are promoter sites. **Link to a discussion of these.**

Embryonic development requires the coordinated production and distribution of transcription factors.

Link to a discussion of some of the transcription factors that produce the segmented body plan in Drosophila.

Enhancers

Some transcription factors ("Enhancer-binding protein") bind to regions of DNA that are thousands of base pairs away from the gene they control. Binding increases the rate of transcription of the gene.

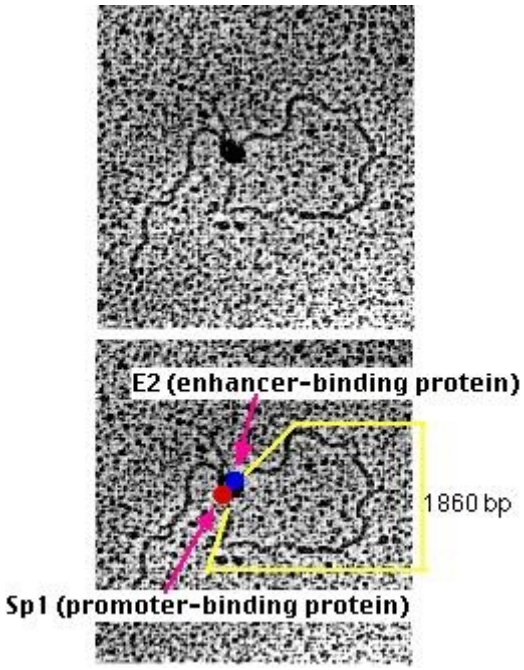


Figure 3.2.

Enhancers can be located upstream, downstream, or even within the gene they control.

How does the binding of a protein to an enhancer regulate the transcription of a gene thousands of base pairs away?

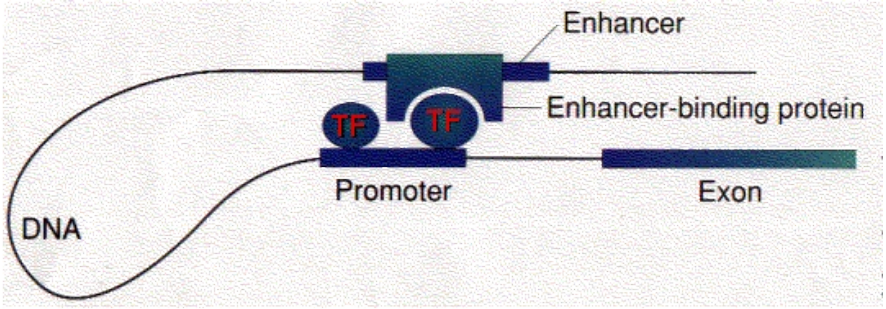


Figure 3.3.

One possibility is that enhancer-binding proteins — in addition to their DNA-binding site, have sites that bind to transcription factors ("TF") assembled at the promoter of the gene.

This would draw the DNA into a loop (as shown in the figure).

Visual evidence

Michael R. Botchan (who kindly supplied these electron micrographs) and his colleagues have produced visual evidence of this model of enhancer action. They created an artificial DNA molecule with

- several (4) promoter sites for Sp1 about 300 bases from one end. Sp1 is a **zinc-finger transcription factor** that binds to the sequence 5' GGGCGG 3' found in the promoters of many genes, especially "housekeeping" genes.

- several (5) enhancer sites about 800 bases from the other end. These are bound by an enhancer-

binding protein designated E2.

- 1860 base pairs of DNA between the two.

When these DNA molecules were added to a mixture of Sp1 and E2, the electron microscope showed that the DNA was drawn into loops with "tails" of approximately 300 and 800 base pairs.

At the neck of each loop were two distinguishable globs of material, one representing Sp1 (red), the other E2 (blue) molecules. (The two micrographs are identical; the lower one has been labeled to show the interpretation.)

Artificial DNA molecules lacking either the promoter sites or the enhancer sites, or with mutated versions of them, failed to form loops when mixed with the two proteins.

Silencers

Silencers are control regions of DNA that, like enhancers, may be located thousands of base pairs away from the gene they control. However, when transcription factors bind to them, expression of the gene they control is repressed.

Insulators

A problem:

As you can see **above**, enhancers can turn on promoters of genes located thousands of base pairs away. What is to prevent an enhancer from inappropriately binding to and activating the promoter of some other gene in the same region of the chromosome?

One answer: an insulator.

Insulators are:

- stretches of DNA (as few as 42 base pairs may do the trick)
- located between the
  - enhancer(s) and promoter or
  - silencer(s) and promoter of adjacent genes or clusters of adjacent genes.

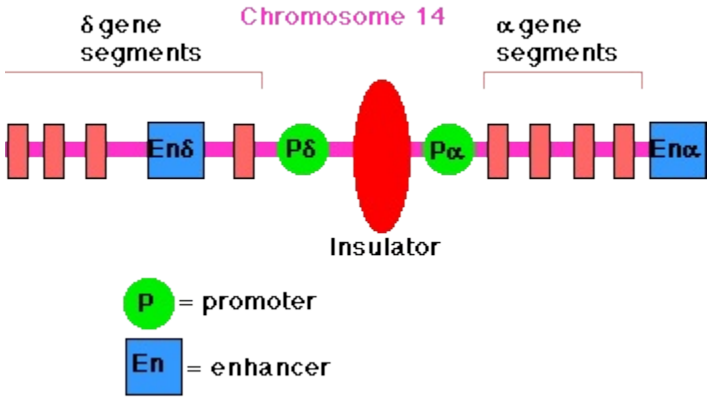


Figure 3.4.

The enhancer for the promoter of the gene for the delta chain of the gamma/delta T-cell receptor for antigen (TCR) is located close to the promoter for the alpha chain of the alpha/beta TCR (on chromosome 14 in humans). A T cell must choose between one or the other. There is an insulator between the alpha gene promoter and the delta gene promoter that ensures that activation of one does not spread over to the other.

Example: The enhancer for the promoter of the gene for the delta chain of the gamma/delta T-cell receptor for antigen (TCR) is located close to the promoter for the alpha chain of the alpha/beta TCR (on chromosome 14 in humans). A T cell must choose between one or the other. There is an insulator between the alpha gene promoter and the delta gene promoter that ensures that activation of one does not spread over to the other.

Another example: In mammals (mice, humans, pigs), only the allele for insulin-like growth factor-2 (IGF2) inherited from one's father is active; that inherited from the mother is not — a phenomenon called imprinting.

The mechanism: the mother's allele has an insulator between the IGF2 promoter and enhancer. So does the father's allele, but in his case, the insulator has been methylated. CTCF can no longer bind to the insulator, and so the enhancer is now free to turn on the father's IGF2 promoter.

Link to a discussion of imprinting.

Many of the commercially-important varieties of pigs have been bred to contain a gene that increases the ratio of **skeletal muscle** to fat. This gene has been sequenced and turns out to be an allele of IGF2, which contains a single **point mutation** in one of its introns. Pigs with this mutation produce higher levels of IGF2 mRNA in their skeletal muscles (but not in their liver).

This tells us that:

- Mutations need not be in the protein-coding portion of a gene in order to affect the **phenotype**.
- Mutations in non-coding portions of a gene can affect how that gene is regulated (here, a change in muscle but not in liver).

Mutations in non-coding portions of a gene can affect how that gene is regulated (here, a change in muscle but not in liver).

For consideration of regulation elements in detail, such as GAL genes in S. cerevisiae (**PDF**), Transcription regulation in S. cerevisiae (**PDF**), and Global transcriptional profiling (**PDF - 1.4 MB**), click PDF files from MITOPENCOURSEWARE respectively.

Lecture 27. Tetrad analysis in fungi

In general, tetrad is the products of a single meiosis in all eukaryotic diploid organisms from simplest ones such as Saccharomyces cerevisiae to complex organisms like human beings. Tetrad analysis is a genetic dissection involving tetrads and based on movement laws of chromosomes in meiosis. Theorically tetrad analysis can be carried out in all eukaryriotes. However, technically tetrad analysis can easily and Mutations in non-coding portions of a gene can affect how that gene is regulated (here, a change in muscle but not in liver).

The yeast Saccharomyces cerevisiae has been a very important genetic tool. It has been used in genetic studies for many decades as one of the best characterized eukaryotic organisms. Since it is very small and unicellular, large numbers of the yeast can be grown in culture in a very small amount of space, in much the same way that bacteria can be grown. However, yeast has the advantage of being a eukaryotic organism, so the results of genetic studies with yeast are more easily applicable to human genetics. It reproduces abundantly and quickly, producing more haploid cells. They can also mate with an appropriate strain, later undergoing karyogamy and growing as a diploid. The diploid can undergo meiosis to form ascospores, recombinant haploid progeny unlike either parent. Mitosis and meiosis can be more easily studied in these organisms. Lee Hartwell, from the Fred Hutchison Cancer Research Center in Seattle, won the Nobel Prize in Medicine in 2001 for his pioneering work on the mitosis genes in S. cerevisiae. He shared the prize with R. Timothy Hunt and Paul M. Nurse of the Imperial Cancer Research in London, who work on another yeast, Schizosaccharomyces pombe. The genes they discovered and characterized in the yeast as a model organism have led to some important discoveries in fighting cancer in humans.

There are two kinds of tetrads in fungi: ordered and unordered tetrads. Ordered tetrads contain the spores (the products of a single meiosis) inside the sac (ascus) in a linear order according to the moving behaviour of chromosomes in meiosis. The tetrads of the kind are available in Neurospora crasa, for example. Unordered tetrads contain the spores inside the ascus in a disorder without any sequence, which are available, for example, in Saccharomyces cerevisiae. Genetic analysis of ordered tetrads technically give more information than that of unordered tetrads. A demonstration of genetic analysis in ordered tetrads is given in MITOPENCOURSEWARE (**PDF**).

Lecture 28. Human DNA polymorphisms

One of the most important tools underlying the revolution in medical genetics is the ability to visualize sequence differences directly in DNA. When studied in the context of a population, these differences in DNA sequences are called polymorphisms; they may occur in coding regions (exons) or noncoding regions of genes. The ability to visualize thousands of DNA polymorphisms has made possible family studies for tracking genes of medical importance. This technique has located and identified genes for many disorders with a clear pattern of mendelian inheritance, such as cystic fibrosis, the inherited

muscular dystrophies, and neurodegenerative disorders such as Huntington's disease. Methods that exploit genetic polymorphism will also be essential for finding genes that predispose people to more common conditions in which inheritance patterns are complex, such as diabetes, atherosclerosis, and hypertension.

DNA polymorphisms are also playing a crucial part in unraveling the genetic basis of tumor formation and progression in cancer. They provide markers for the loss of specific chromosomal segments during the evolution of a tumor. DNA polymorphisms have already been crucial in the identification of genes important for susceptibility to common forms of cancer, such as colon cancer, as well as susceptibility to less common childhood tumors, such as retinoblastoma and Wilms' tumor.

The most useful DNA sequence polymorphisms have many alternative forms. The value of highly variable DNA sequences as genetic markers rests on straightforward principles. Every person carries two copies of each chromosome except the sex chromosomes. If a DNA polymorphism is to be useful in analyzing the transmission of the two chromosomes in a family or the loss of one of the chromosomes during tumorigenesis, then the DNA copies at the polymorphic site of the person under study must be different in the two chromosomes (**Figure 1A**), **Figure 1B**), **Figure 1C**), and **Figure 1D**). The likelihood that a given person will have different DNA sequences at the polymorphic site directly determines the usefulness of that site in genetic studies. Chromosomal sites at which the DNA sequences can have many alternative forms are thus ideal sites for genetic markers. At these sites, a person is most likely to carry two alternative DNA sequences, accurately marking the two alternative chromosomes.

In the human genome, the sites that have the properties most favorable to such extensive variation include a repetition of the same short DNA sequence a variable number of times. Such sequences are called tandem-repeat sequences. A DNA sequence with such variation may be as short as two base pairs or as long as several hundred base pairs. Highly variable sequences of this type are well distributed throughout the length of every human chromosome. When tandemly repeated sequences are replicated during cell division, the number of repeats can change. The frequency of this kind of replication error is high enough to make alternative lengths at the polymorphic site common, but the rate of change in the length of the site is low enough that the size of the DNA at the polymorphic site serves as a stable trait in family studies (**Figure 1A**).

Two techniques, Southern blotting and the polymerase chain reaction (PCR), can measure the length of the DNA sequence at the polymorphic site (**Figure 1B**). The one to choose depends on the length of the tandemly repeated sequence. A repeated sequence 20 to 40 base pairs in length leads to variation in DNA lengths of hundreds or even thousands of base pairs at the polymorphic site. Southern blotting is best for visualizing this degree of variation in length. Very short tandemly repeated sequences, only two, three, or four base pairs long, can also vary highly. For these, the PCR is preferred. Whichever technique is used, its goal is to assess accurately the length of the DNA segment between two fixed points on each chromosome. These two points include some DNA adjacent to the repeated sequence as well as the repeated sequence itself. In the case of Southern blotting, the position of the fixed points is defined by the location of restriction-enzyme digestion sites in the DNA flanking the repeated sequence. In the case of PCR, the positions in the flanking DNA of sequences homologous to the oligonucleotide PCR primers define the fixed points.

In Southern blotting, the DNA isolated from each patient or tumor to be typed is digested with a restriction enzyme, separated on the basis of size by agarose-gel electrophoresis, and transferred to a nylon membrane. A DNA probe can reveal directly on the nylon membrane the size of DNA fragments carrying the repeated sequence. This probe corresponds to a sequence in the DNA flanking the repeated sequence. In general, DNA from one person shows two such DNA fragments or bands (**Figure 1C**). For each chromosomal site, one of the two bands will be passed to the next generation, and the other will not, thus indicating the outcome in genetic transmission that occurred at this particular chromosomal site.

With the PCR method, the unique sites of primer binding adjacent to the repeated sequence allow specific amplification of the region that includes the repeat. The size of the amplified DNA molecules representing the polymorphic site can now be determined with the same technique that determines the DNA sequence. Precise determination of the length of the amplified DNA molecules usually shows two alternative copies of the DNA fragment, one for each of the chromosomes on which that sequence resides. The application of the two techniques has varied somewhat in human genetic studies; each has advantages and limitations. Sites of short sequence-length variation have been found to be widely distributed along the chromosomes, making them the most widely used sites in genetic-linkage studies designed to track medically important genes in families.

Studies of tumors must compare the DNA of normal cells with that of cancer cells. The normal cells usually have two bands, whereas the tumor cells often have only one. This finding is diagnostic of the loss of one copy of a chromosomal region during tumorigenesis. The problem of contamination of a tumor by normal cells presents important issues for studies of this type. Because the PCR involves an amplification process, the amount of material in the starting sample and the amount present in the final amplification product are not necessarily linearly related. Making a judgment about the loss of chromosomal material in a tumor sample contaminated with a substantial number of cells from surrounding normal tissue can be quite challenging. Unlike the results of the PCR, the signal generated by the Southern blotting procedure is directly proportional to the relative amount of each allele present in a tumor sample. Southern blotting has thus been used with particular effect in studies of the loss of chromosomal material by tumor cells (**Figure 1D**).

Genetic mapping can determine the relative positions of highly variable DNA sites on each chromosome. Well-characterized polymorphic DNA sites now number in the thousands. The availability of this large number of closely spaced genetic markers has revolutionized human genetics, because it allows the application of genetic-mapping strategies with great precision. For many medically important genes, particularly those that contribute to a predisposition to common medical conditions, the primary limitation to their identification was until recently the availability of a sufficient number of highly informative genetic markers. The techniques described here have removed this limitation. As a result, many important developments in all aspects of medicine are likely to follow.

One more important class of DNA polymorphism is single-nucleotide polymorphism

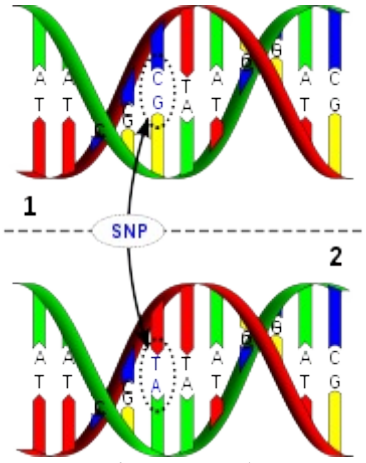


Figure 3.5.

A single-nucleotide polymorphism (SNP, pronounced snip) is a **DNA sequence** variation occurring when a single **nucleotide** — **A**, **T**, **C**, or **G** — in the **genome** (or other shared sequence) differs between members of a species (or between paired chromosomes in an individual). For example, two sequenced DNA fragments from different individuals, AAGCCTA to AAGCTTA, contain a difference in a single nucleotide. In this case we say that there are two **alleles** : C and T. Almost all common SNPs have only two alleles.

DNA molecule 1 differs from DNA molecule 2 at a single base-pair location (a C/T polymorphism).Within a population, SNPs can be assigned a **minor allele frequency** — the lowest allele frequency at a **locus** that is observed in a particular population. This is simply the lesser of the two allele frequencies for single-nucleotide polymorphisms. There are variations between human populations, so a SNP allele that is common in one geographical or ethnic group may be much rarer in another.

In the past, SNPs with a minor allele frequency of greater than or equal to 1% (or 0.5%, etc.) were given the title "SNP".**[1]** Some used "**mutation**" to refer to variations with low allele frequency. With the advent of modern **bioinformatics** and a better understanding of evolution, this definition is no longer necessary, e.g., a database such as **dbSNP** includes "SNPs" that have lower allele frequency than one percent.**[2]**

Single-nucleotide **polymorphisms** may fall within coding sequences of **genes**, **non-coding regions of genes**, or in the **intergenic regions** between genes. SNPs within a coding sequence will not necessarily change the **amino acid** sequence of the **protein** that is produced, due to **degeneracy of the genetic code**. A SNP in which both forms lead to the same polypeptide sequence is termed **synonymous** (sometimes called a **silent mutation**) — if a different polypeptide sequence is produced they are **nonsynonymous**. A nonsynonymous change may either be **missense** or **nonsense**, where a missense change results in a different amino acid, while a nonsense change results in a premature **stop codon**. SNPs that are not in protein-coding regions may still have consequences for **gene splicing**, **transcription factor** binding, or the sequence of **non-coding RNA**.

Variations in the DNA sequences of humans can affect how humans develop **diseases** and respond to **pathogens**, **chemicals**, **drugs**, **vaccines**, and other agents. SNPs are also thought to be key enablers in realizing the concept of **personalized medicine**.**[3]** However, their greatest importance in biomedical

research is for comparing regions of the genome between **cohorts** (such as with matched cohorts with and without a disease).

The study of single-nucleotide polymorphisms is also important in crop and livestock breeding programs (see **genotyping**). See **SNP genotyping** for details on the various methods used to identify SNPs.

Microsatellites

Longer DNA sequence repeats are Microsatellites, or Simple Sequence Repeats (SSRs called also STRs), which are **polymorphic** loci present in **nuclear** and **organellarDNA** that consist of repeating units of 1- 6 **base pairs** in length. **[1]** They are typically neutral, **co-dominant** and are used as **molecular markers** which have wide-ranging applications in the field of **genetics**, including **kinship** and **population** studies. Microsatellites can also be used to study gene dosage (looking for **duplications** or **deletions** of a particular genetic region).

One rare example of a microsatellite is a (CA)n repeat, where n is variable between **alleles**. These markers often present high levels of inter- and intra-specific polymorphism, particularly when tandem repeats number one hundred or greater.**[2]** The repeated sequence is often simple, consisting of two, three or four **nucleotides** (di-, tri-, and tetranucleotide repeats respectively), and can be repeated 10 to 100 times. CA nucleotide repeats are very frequent in **human** and other **genomes**, and are present in every few thousand base pairs. As there are often many alleles present at a microsatellite locus, **genotypes** within **pedigrees** are often fully informative, in that the **progenitor** of a particular allele can often be identified. In this way, microsatellites are ideal for determining paternity, population genetic studies and **recombination mapping**. It is also the only molecular marker to provide clues about which alleles are more closely related.**[3]**

Microsatellites owe their variability to an increased rate of mutation compared to other neutral regions of DNA. These high rates of mutation can be explained most frequently by **slipped strand mispairing** (slippage) during **DNA replication** on a single DNA strand. Mutation may also occur during **recombination** during **meiosis**.**[4]** Some errors in slippage are rectified by **proofreading** mechanisms within the **nucleus**, but some mutations can escape repair. The size of the repeat unit, the number of repeats and the presence of variant repeats are all factors, as well as the frequency of **transcription** in the area of the DNA repeat. Interruption of microsatellites, perhaps due to mutation, can result in reduced polymorphism. However, this same mechanism can occasionally lead to incorrect **amplification** of microsatellites; if slippage occurs early on during PCR, microsatellites of incorrect lengths can be amplified.

Microsatellites can be amplified for identification by the **polymerase chain reaction** (PCR) process, using the unique sequences of flanking regions as **primers**. DNA is repeatedly denatured at a high temperature to separate the double strand, then cooled to allow annealing of primers and the extension of nucleotide sequences through the microsatellite. This process results in production of enough DNA to be visible on **agarose** or **polyacrylamide** gels; only small amounts of DNA are needed for amplification as thermocycling in this manner creates an exponential increase in the replicated segment**[5]**.

VNTR

The longest DNA repeats are A Variable Number Tandem Repeats (or VNTR). This is a location in a genome where a short nucleotide sequence is organized as a tandem repeat. These can be found on many chromosomes and often show variations in length between individuals. Each variant acts as an inherited allele, allowing them to be used for personal or parental identification. Their analysis is useful in genetics and biology research, forensics, and DNA fingerprinting. VNTR loci are hypervariable loci or minisatellite sequences, which vary in number of repeats of short (16-300 bp) core segment.

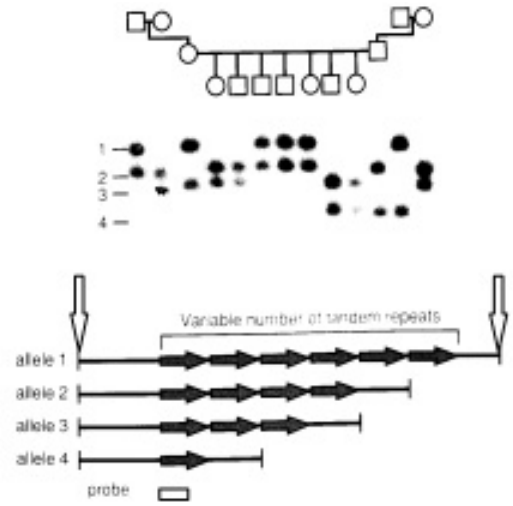


Figure 3.6.

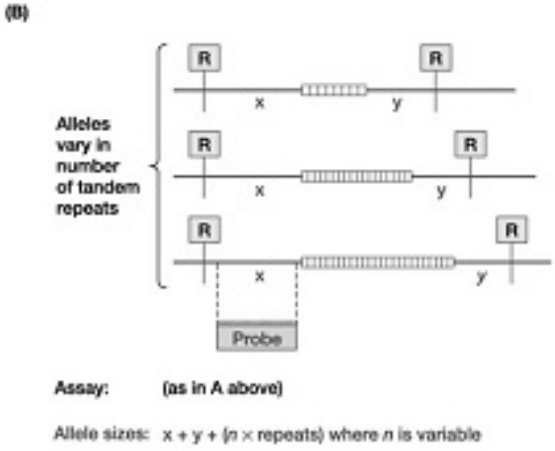


Figure 3.7.

VNTR have high levels of polymorphism, many alleles and usually can be visualized by Southern blotting or PCR as shown below. VNTR loci are applied in DNA fingerprinting, forensic paternity and linkage analysis.

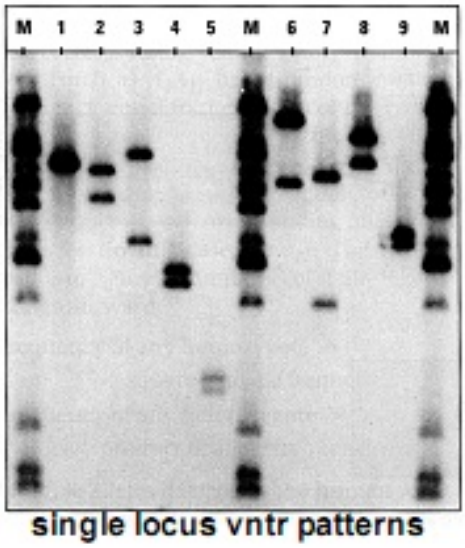


Figure 3.8.

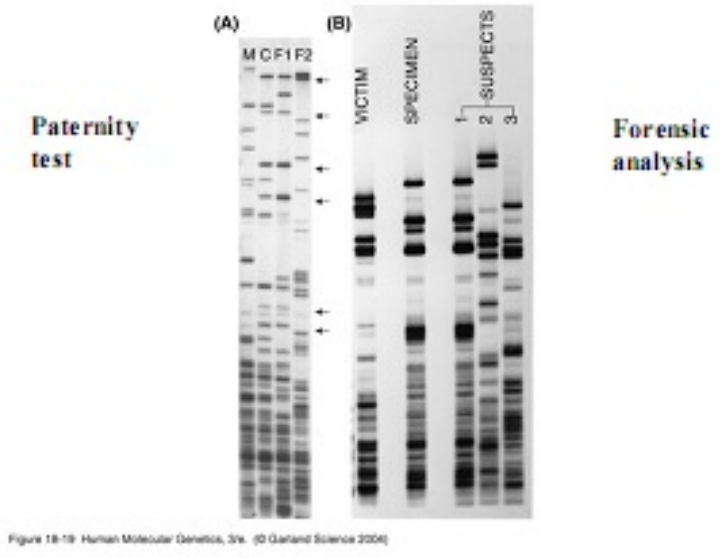


Figure 3.9.

One can see a good PowerPoint presentation describing STRs and SSRs and their applications in MITOPENCOURSEWARE (**PDF**).

Solutions

Chapter 4. Genetics in clasical understanding

Lecture 29. Mendelian discovery of genes

Mendelian inheritance (or Mendelian genetics or Mendelism) is a set of primary tenets relating to the transmission of **hereditary** characteristics from parent organisms to their children; it underlies much of **genetics**. They were initially derived from the work of **Gregor Mendel** published in 1865 and 1866 which was "re-discovered" in 1900, and were initially very controversial. When they were integrated with the **chromosome theory of inheritance** by **Thomas Hunt Morgan** in 1915, they became the core of **classical genetics**.

The laws of inheritance were derived by **Gregor Mendel**, a 19th century **[1]** monk conducting hybridization experiments in garden peas (**Pisum sativum**). Between 1856 and 1863, he cultivated and tested some 28,000 pea plants. From these experiments he deduced two generalizations which later became known as Mendel's Laws of Heredity or Mendelian inheritance. He described these laws in a two part paper, "**Experiments on Plant Hybridization**" that he read to the Natural History Society of **Brno** on **February 8** and **March 8**, **1865**, and which was published in 1866.**[2]**

The principles of heredity were written by the Augustinian monk Gregor Mendel in 1865. Mendel discovered that by crossing white flower and purple flower plants, the result was a hybrid offspring. Rather than being a mix of the two, the offspring was purple flowered. He then conceived the idea of heredity units, which he called "factors", one of which is a recessive characteristic and the other dominant. Mendel said that factors, later called genes, normally occur in pairs in ordinary body cells, yet segregate during the formation of sex cells. Each member of the pair becomes part of the separate sex cell. The dominant gene, such as the purple flower in Mendel's plants, will hide the recessive gene, the white flower. After Mendel self-fertilized the F1 generation and obtained the 3:1 ratio, he correctly theorized that genes can be paired in three different ways for each trait; AA, aa, and Aa. The capital A represents the dominant factor and lowercase a represent the recessive.

Mendel stated that each individual has two factors for each trait, one from each parent. The two factors may or may not contain the same information. If the two factors are identical, the individual is called homozygous for the trait. If the two factors have different information, the individual is called heterozygous. The alternative forms of a factor are called alleles. The genotype of an individual is made up of the many alleles it possesses. An individual's physical appearance, or phenotype, is determined by its alleles as well as by its environment. An individual possesses two alleles for each trait; one allele is given by the female parent and the other by the male parent. They are passed on when an individual matures and produces gametes, egg and sperm. When gametes from the paired alleles separate randomly, each gamete receives a copy of one of the two alleles. The presence of an allele doesn't promise that the trait will be expressed in the individual that possesses it. In heterozygous individuals, the only allele that is expressed is the dominant. The recessive allele is present but its expression is

hidden.

Mendel summarized his findings in two laws; the Law of Segregation and the Law of Independent Assortment.

Now when we know the mechanisms of meiosis, one can conclude that the two abovementioned Mendelian laws are direct consequences of the assortment laws of chromosomes in meiotic cell division, and the Mendelian “factors” are today’s genes.

A good description of Mendel’s pea crosses and his detail experiments was presented in MITOPENCOURSEWARE (**PDF**).

Lecture 30. Gene linkage

Gregor Mendel analyzed the pattern of inheritance of seven pairs of contrasting traits in the domestic pea plant. He did this by cross-breeding **dihybrids**; that is, plants that were **heterozygous** for the alleles controlling two different traits.

Mendel then crossed these dihybrids. If it is inevitable that round seeds must always be yellow and wrinkled seeds must be green, then he would have expected that this would produce a typical monohybrid cross: 75% round-yellow; 25% wrinkled-green. But, in fact, his mating generated seeds that showed all possible combinations of the color and texture traits.

- 9/16 of the offspring were round-yellow
- 3/16 were round-green
- 3/16 were wrinkled-yellow, and
- 1/16 were wrinkled-green

Finding in every case that each of his seven traits was inherited independently of the others, he formed his "second rule", the Rule of Independent Assortment:

The inheritance of one pair of factors (genes) is independent of the inheritance of the other pair. Today we know that this rule holds only if the genes are on separate chromosomes.

Mendel was lucky in that every pair of genes he studied met one requirement or the other. The table shows the chromosome assignments of the seven pairs of alleles that Mendel studied. All of these genes showed **independent assortment**, and they were inherited on separate chromosomes. With the rebirth of genetics in the 20th century, it quickly became apparent that Mendel's second rule does not apply to many matings of dihybrids. In many cases, two alleles inherited from one parent show a strong tendency to stay together as do those from the other parent. This phenomenon is called linkage.

So, gene linkage is the physical relationship of genes. Specifically, linkage means that the genes are on the same chromosome and therefore do not assort independently into gametes --in humans, ovum and

spermatozoa-- during meiosis.

Because of this co-transmittance, the traits associated with the genes do not segregate between two daughter cells, following crosses between the parental cells, as predicted by Mendelian genetics.

The genes of most organisms can exist in different forms, called alleles, in a population. If the organism has identical alleles of a gene on each of its homologous chromosomes, it is called homozygous. If the alleles are different, it is called heterozygous. During the cell division process, a separation of nuclear material into gametes occurs via meiosis. If an organism is heterozygous, two kinds of gametes are produced; if homozygous, it produces only one kind of gamete. At fertilization the male and female gametes combine and the random process that creates different units put the gametes into various combinations. The ratio of the appearance of the observed traits, or phenotypes, produced by the pattern of separation of the dominant and recessive genes for that trait was predicted by Gregor Mendel following painstaking work and observation of the crosses between pea plants.

However, early in the twentieth century, William Bateson and Reginald Crundall Punnett, two British geneticists, observed that sometimes the expected Mendelian ratio of phenotypes did not occur. Their best explanation was that in some manner the phenotypic classes, the alleles, were coupled, and so did not sort independently into gametes. Proof of their explanation was provided by Thomas Hunt Morgan, using Drosophila eye color as the examined trait.

Morgan observed that test crosses between mutants in eye color and wing development deviated from the expected Mendelian 1:1:1:1 ratio for independent assortment. The observed ratio was, rather, consistent with the non-independent segregation of two genes that were close to each other on the same chromosome.

Linked genes do not observe the genotypic or phenotypic relationships predicted by Mendelian crosses that assume independent assortment of chromosomes and genes. In a cross the parental generation is designated P1 and the first generation of offspring are designated F1(first filial generation), and the offspring resulting from the fertilization between individuals of the F1 generation are called the F2 (second filial generation). When the F1 and F2 ratios deviate from the predicted Mendelian ratios, this is evidence of gene linkage.

The linkage of genes is used to generate so-called linkage maps which give a measure of the distance between genes on a chromosome. The linkage map technique, which is based on the use of the percentage of recombinants, in which crossing over of DNA and expression of traits due to gene linkage has occurred, was devised in 1911 by Alfred Henry Sturtevant, an undergraduate student of Morgan's. The technique remains in use today as a means of producing an index of the distance between two genes.

To see concrete experiments for analyzing as well as for applying gene linkage, click over to MITOPENCOURSEWARE (**PDF**) for the complementaton test and gene function; click to (**PDF**) for tests of gene position, starting with the position of genes on chromosomes in general; click to (**PDF**) for experiments to map genes relative to one another on sex chromosomes, and, finally, click to (**PDF**) for mapping genes on autosomes by test-cross and other measures.

Lecture 31. Genetic complex traits

Genetic complex traits refer to those traits which are determined by either many genes or vice versa a single gene influences multiple phenotypictraits.

This lecture note will present the main types of the traits of the kind-- Pleiotropy, Polygenic Inheritance, Genetic Heterogeneity, Twinnings and Siblings.

Pleiotropy

Pleiotropy occurs when a single gene influences multiple phenotypictraits. Consequently, a new mutation in the gene will have an effect on all traits simultaneously. This can become a problem when selection on one trait favors one specific mutant, while the selection on the other trait favors another mutant. The underlying pleiotropic mechanism is that the gene codes for a product that is, for example, used by various cells, or has a signaling function on various targets.

A classic example of pleiotropy is the human disease PKU (phenylketonuria). This disease can cause mental retardation and reduced hair and skinpigmentation, and can be caused by any of a large number of mutations in a single gene that codes for an enzyme (phenylalanine hydroxylase) that converts the amino acidphenylalanine to tyrosine, another amino acid. Depending on the mutation involved, this results in reduced or zero conversion of phenylalanine to tyrosine, and phenylalanine concentrations increase to toxic levels, causing damage at several locations in the body. PKU is totally benign if a diet free from phenylalanine is maintained.

Antagonistic pleiotropy refers to the expression of a gene resulting in multiple competing effects, some beneficial but others detrimental to the organism.

This is central to a theory of aging first developed by G. C. Williams in 1957.[1] Williams suggested that some genes responsible for increased fitness in the younger, fertile organism contribute to decreased fitness later in life. One such example in male humans is the gene for the hormone testosterone. In youth, testosterone has positive effects including reproductive fitness but, later in life, there are negative effects such as increased susceptibility to prostate cancer. Another example is the p53 gene which suppresses cancer, but also suppresses stem cells which replenish worn-out tissue[2].

Whether or not pleiotropy is antagonistic may depend upon the environment. For instance, a bacterial gene that enhances glucose utilization efficiency at the expense of the ability to use other energy sources (such as lactose) has positive effects when there is plenty of glucose, but it can be lethal if lactose is the only available food source.

Polygenic inheritance

Polygenic inheritance is a pattern responsible for many features that seem simple on the surface. Many traits such as height, shape, weight, color, and metabolic rate are governed by the cumulative effects of many genes. Polygenic traits are not expressed as absolute or discrete characters, as was the case with Mendel's pea plant traits. Instead, polygenic traits are recognizable by their expression as a gradation of small differences (a continuous variation). The results form a bell shaped curve, with a mean value and

extremes in either direction.

Height in humans is a polygenic trait, as is color in wheat kernels. Height in humans is NOT **discontinuous**. If you line up the entire class, a continuum of variation is evident, with an average height and extremes in variation [very short (vertically challenged) and very tall (vertically enhanced)]. Traits showing continuous variation are usually controlled by the additive effects of two or more separate gene pairs. This is an example of polygenic inheritance. The inheritance of EACH gene follows Mendelian rules.

Usually polygenic traits are distinguished by

1. 1. Traits are usually quantified by measurement rather than counting.
2. 2. Two or more gene pairs contribute to the phenotype.
3. 3. Phenotypic expression of polygenic traits varies over a wide range.

Human polygenic traits include:

1. 1. Height
2. 2. SLE (Lupus). (Click **here** for an article about lupus and genetics.)
3. 3. Weight. (Click **here** for an article about obesity and genetics.)
4. 4. Eye Color. (Click **here** for an article about eye color.)
5. 5. Intelligence.
6. 6. Skin Color.
7. 7. Many forms of behavior.

Click **here** to see Genetic Heterogeneity, Twinnings and Sibblings described by MITOPENCOURSEWARE.

Lecture 32. Probability and pedigrees

In Mendel's time he used statistics to account for his observations on his experiments on peas, and, thanks to the results he abtained, he could formulate his two famous laws of genetics-- the Law of segregation and the Law of independent assortment, which were based on statistical segregation ratio 3:1, 9:3:3:1, 1:1:1:1 etc…

Nowadays in genetic research and especially in medical genetic counseling, statistics is needed for calculating the risks of genetic diseases in human pedigrees. The risks in these cases are expressed in terms of so-called probability.

The probability of an event is the chance that it will happen. The probability of tossing a coin to land heads up is roughly ½.

- The probability of an impossible event is 0.
- the probability of a certain event is 1.
- If the probability of event x is p then the probability of 'not x' is 1-p.
- The probability of two independent events ocurring at the same time is the product of their two indivdual probabilities.

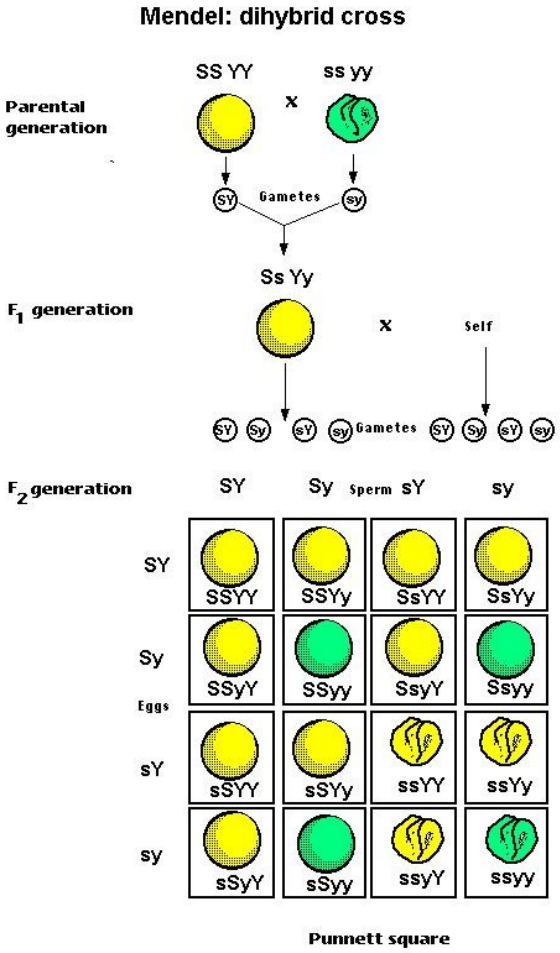


Figure 4.1.

So, for example, in the cross above, in the F2 the

- probability of a wrinkled seed is ¼; the probability of a green seed is also ¼, and the probability of being both green and wrinkled is therefore ¼ x ¼ = 1/16.
- The probability of being not wrinkled (i.e. smooth) is 1-¼ = ¾. The probability of being both smooth and green is therefore ¾ x ¼ = 3/16 and so on.
- In the **example below** about the coefficient of inbreeding of children from first cousin marriages, we considered a number of probabilities of ½ which we multiplied together to reach a final probability of 1/16 that any gene was homozygous by descent.

Autosomal recessive

A recessive trait will only manifest itself when homozygous. If it is a severe condition, it will be unlikely that homozygotes will live to reproduce, and thus most occurrences of the condition will be in matings between two heterozygotes (or carriers). An autosomal recessive condition may be transmitted through a long line of carriers before, by ill chance, two carriers mate. Then there will be a ¼ chance that any child will be affected. The pedigree will therefore often only have one 'sibship' with affected members.

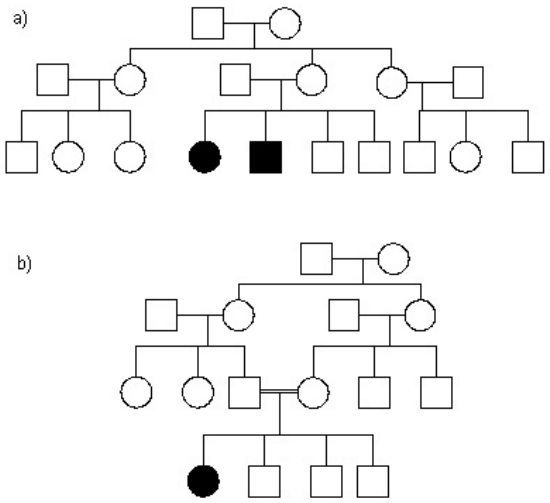


Figure 4.2.

a) A 'typical' autosomal recessive pedigree and b) An autosomal pedigree with inbreeding

If the parents are related to each other, perhaps by being cousins, there is an increased risk that any gene present in a child may have two alleles identical by descent. The degree of risk that both alleles of a pair in a person are descended from the same recent common ancestor is the degree of inbreeding of the person. Let us examine b) in the figure above.

Considering any child of a first cousin mating, we can trace through the pedigree the chance that the other allele is the same by common descent. Let us consider any child of generation IV, any gene which came from the father, III3 had a half chance of having come from grandmother II2, a further half chance of being also present in her sister, grandmother II4 a further half a chance of having been passed to mother III4 and finally a half chance of being transmitted into the same child we started from. A total risk of ½ x ½ x ½ x ½ = 1/16.

This figure, which can be thought of as either

- the chance that both maternal and paternal alleles at one locus are identical by descent or
- the proportion of all the individual's genes that are homozygous because of identity by common descent,

is known as the coefficient of inbreeding and is usually given the symbol F.

To see a compact and also clear description of using statiistics for pedigree genetic analysis, click MITOPENCOURSEWARE (**PDF**).

Lecture 33. Population genetics

If genetics is a science studying structure, function and movement rules of genes, population genetics is the third part of it: a science studying movement rules of gene carriers – chromosomes and their effects and consequences. At least from the genetic point of view, population is a unit of evolution. In terms of breeding practice, populations are plant varieties and animal breeds.

D. S. Falconer (The quote is from Introduction to Quantitative Genetics by D. S. Falconer, 1960, Ronald Press.) wrote:

"A population in the genetic sense, is not just a group of individuals, but a breeding group; and the genetics of a population is concerned not only with the genetic constitution of the individuals but also with the transmission of the genes from one generation to the next. In the transmission the genotypes of the parents are broken down and a new set of genotypes is constituted in the progeny, from the genes transmitted in the gametes. The genes carried by the population thus have continuity from generation to generation, but the genotypes in which they appear do not. The genetic constitution of a population, referring to the genes it carries, is described by the array of gene frequencies, that is by specification of the alleles present at every locus and the numbers or proportions of the different alleles at each locus." (page 6).

In fact population genetics is studying the **allele frequency** distribution and change under the influence of the four evolutionary forces: **natural selection**, **genetic drift**, **mutation** and **gene flow**. It also takes account of population subdivision and population structure in space. As such, it attempts to explain such phenomena as **adaptation** and **speciation**. Population genetics was a vital ingredient in the **modern evolutionary synthesis** whose primary founders were **Sewall Wright**, **J. B. S. Haldane** and **R. A. Fisher**, they also laid the foundations for the related discipline of **quantitative genetics**.

For humans the applications of Mendelian genetics, chromosomal abnormalities, and multifactorial inheritance to medical practice are quite evident. Physicians work mostly with patients and families. However, as important as the work of physicians may be, genes also affect populations, and in the long run their effects in populations have a far more important impact on medicine than the relatively few families each physician may serve. It is important that certain polymorphisms are maintained so that the species may survive, even at the expense of individuals. Genetic polymorphisms often are detrimental to the homozygote, but they allow others of the species to survive. Before medical intervention was possible, populations that lacked the sickle cell anemia allele could not survive in the malaria regions of West Africa. Those that had the sickle cell anemia allele survived, and the gene remains in the population at high frequency today, even though the homozygous recessive phenotype was at a severe disadvantage in the past. The high rate of thalassemia in people of Mediterranean origin, the high rate of sickle cell anemia in people of West African descent, the high rate of cystic fibrosis in people from Western Europe, and the high rate of Tay-Sachs disease in ethnic groups from Eastern Europe may all owe their origin to environmental factors that cause changes in gene frequencies in large populations by giving some advantage to heterozygotes who carry a deleterious allele. Although one may never use the

calculations of population genetics in medical practice, the underlying principles should be understood.

To have general understanding of population genetics, click (**PDF**) for studying Hardy-Weinberg Equilibrium; click (**PDF**) to see the role of Mutation and Selection in population structure, and click (**PDF**) for consideration of Inbreeding as a factor influencing the composition of populations.

Solutions

Chapter 5. About the controversal ethical issues on applications of genetics

Lecture 34. The Human Genome Project and Human cloning

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**.

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.

Lecture 35. Genetic prenatal diagnosis and Gene therapy

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.

Gene therapy: somatic and germline gene therapy

Somatic Cell Gene Therapy

Many genetic diseases may be able to be treated with gene therapy to correct the defective genes.

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 for several different diseases (including several cancers) in many countries (including the USA, EU, Canada, China, Japan, New Zealand...), and involving over 6000 patients world-wide.

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 that successfully treated adenosine deaminase deficiency (ADA) disease began in September 1990.

From 1989 until September 1999 there were thousands of patients in trials, and no one died because of the experiments. Eighteen-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.

Preimplantation genetic disease diagnosis

In **medicine** and (clinical) **genetics** preimplantation genetic diagnosis (PGD) (also known as Embryo Screening) refers to procedures that are performed on **embryos** prior to **implantation**, sometimes even on **oocytes** prior to **fertilization**. PGD is considered an alternative to **prenatal diagnosis**. Its main advantage is that it avoids selective **pregnancy termination** as the method makes it highly likely that the baby will be free of the disease under consideration. PGD thus is an adjunct to **assisted reproductive technology** and requires **in vitro fertilization** (IVF) to obtain **oocytes** or embryos for evaluation.

The term preimplantation genetic screening (PGS) is used to denote procedures that do not look for a specific disease but use PGD techniques to identify embryos at risk. PGD is a poorly chosen phrase because, in medicine, to "diagnose" means to identify an illness or determine its cause. An oocyte or early-stage embryo has no symptoms of disease. The person is not ill. Rather, he may have a genetic condition that could lead to disease. To "screen" means to test for anatomical, physiological, or genetic conditions in the absence of symptoms of disease. So both PGD and PGS should be referred to as types of embryo screening.

Ethical issues

See also: **In vitro fertilisation#Religious objections**

PGD has raised ethical issues. The technique can be used to determine the gender of the embryo and thus can be used to select embryos of one gender in preference of the other in the context of "**family balancing**." It may be possible to make other "social selection" choices in the future. While controversial, this approach is less destructive than fetal deselection during the pregnancy.

Costs are substantial and insurance coverage may not be available. Thus PGD widens the gap between people who can afford the procedure versus a majority of patients who may benefit but cannot afford the service.

PGD has the potential to screen for genetic issues unrelated to **medical necessity**. The prospect of a "**designer baby**" is closely related to the PGD technique.

By relying on the result of one cell from the multi-cell embryo, it assumed that this cell is representative of the remainder of the embryo. This may not be the case as the incidence of mosaicism is often relatively high. On occasion, PGD may result in a **false negative** result leading to the acceptance of an abnormal embryo, or in a **false positive** result leading to the deselection of a normal embryo.

Since PGD and PGH are procedures that can weed out genetically defective human pre-embryos before

they have a chance start a pregnancy, the procedure is usually requested by prospective parents who are concerned about passing a serious genetically-based disease or disorder to their child.

Typically,

- one or both partners have been genetically screened previously, and found to be a carrier; or
- one or both partners are from a human population known to have a high incidence of a genetically-based disease or disorder.

If an embryo is found to be genetically defective, it is normally destroyed. This produces a very serious concern for many pro-life supporters who believe that every pre-embryo, embryo and fetus is a human person. Destruction of a pre-embryo is considered a form of murder.

However, there are a number of arguments to support PGD:

- Scientifically, if to combine presently available DNA analysis techniques for screening samples taken both from parents at risk and from sperm/egg bank and IVF, one can produce healthy babies both phenotypically and genotypically. At the same time the disease mutation alleles can be gradually removed from human populations.
- Financially, in comparison with the costly PGD, the above-mentioned approach would considerably reduce the cost for the couples at risk.
- Ethically, it is suggested to keep and apply the ethical regulations at present used for IVF and for other human DNA analysis.

Lecture 36. Genetic Testing and Pharmacogenomics

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.

- 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.

Lecture 37. Genetic engineering and food

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.

Many kinds of GMOs have been developed for environmental purposes, for health and medicine. Genetic engineering has been particularly successfully used and applied in food and agriculture to produce genetically modified (GM) foods. Transgenic plants, created by inserting genes from various organisms, carry several enhanced characteristics. Examples include plants with increased yield, disease resistance and pest resistance (Inserted Bt genes selectively kill 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, a characteristic that is also useful during transportation to the market. Over 15 countries of the world already use GM crops for general food production.

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

Solutions

Chapter 6. Assignments and solutions

Table 6.1. Assignments cover the topics discussed in the corresponding lecture sessions

VN-LEC#	MIT-LEC #	ASSIGNMENTs	SOLUTIONS
3, 23, 29, 30, 32	1-5	Problem Set 1 (PDF)	(PDF)
18, 27, 29, 30	6-9	Problem Set 2 (PDF)	(PDF)
16, 20-23	11-15	Problem Set 3 (PDF)	(PDF)
24, 25	16-19	Problem Set 4 (PDF)	(PDF)
25-28	20-24	Problem Set 5 (PDF)	(PDF)
33, 28	25-30	Problem Set 6 (PDF)	(PDF)
31	31-35	Problem Set 7 (PDF)	(PDF)

Exercise 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

D. is present in 46 pairs in human cells

Exercise 2.

Genes:

- A. never function when they contain a mutation
- . directly produce proteins

- C. contain random pairings of nucleotides
- D. all of the above
- E. none of the above

E. none of the above

Exercise 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

A. DNA to RNA

Exercise 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

F. none of the above

Exercise 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.

A. clinical disease will develop based on the mutation alone.

Exercise 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

C. is present in all cells of one's body // is never passed from parents to offspring

Exercise 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

C. leads to substitution of an amino acid in a new place in the protein

Exercise 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

C. involves an inappropriate stop codon

Exercise 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

A. results in no change in protein structure/function

Exercise 10.

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

- True
- Flase

False

Solutions

Chapter 7. References

Textbooks - Articles

- Egger G, Liang G, Aparicio A, et al. Epigenetics in human disease and prospects for epigenetic therapy. Nature 2004;429:457-63.**[CrossRef][Medline]**
- Espino PS, Drobic B, Dunn KL, et al. Histone modifications as a platform for cancer therapy. J Cell Biochem 2005;94:1088-102.**[CrossRef**
- Griffiths, Anthony J. F., Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, and William M. Gelbart. An Introduction to Genetic Analysis. 7th ed. New York: W. H. Freeman, 2000. ISBN: 9780716735205.
- **Katrin Altmann3**, **Mark Dürr3**, **Benedikt Westermann3**. Saccharomyces cerevisiae as a Model Organism to Study Mitochondrial Biology: General Considerations and Basic Procedures
- Kidwell, M.G. (2005). "Transposable elements.". in (ed. T.R. Gregory).
- Peterson CL, Laniel MA. Histones and histone modifications. Curr Biol 2004;14:R546-51.
- Principles of genetics: A textbook, with problems (McGraw-Hill publications in the agricultural and botanical sciences) (Unknown Binding) by **Edmund Ware Sinnott**
- **The Evolution of the Genome**. San Diego: Elsevier. pp. 165–221. **ISBN 0-12-301463-8**. Notes by Amy Stone, checked by Prof. Gross (2/96).

Websites

http://ocw.mit.edu/OcwWeb/web/home/home/index.htm

http://ocw.mit.edu/OcwWeb/Biology/7-03Fall-2004/CourseHome/index.htm

http://www.dnaftb.org/dnaftb/concept_1/con1sum.swf

http://www.dnaftb.com/

http://www.dnaftb.org/dnaftb/concept_29/con29sum.swf

http://www.dnaftb.org/dnaftb/

http://www1.qiagen.com/SelectLocation.aspx

http://www.dnalc.org/ddnalc/resources/animations.html

http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/R/RFLPs.html#fingerprinting

http://www.accessexcellence.org/RC/VL/GG/restriction.html

http://www.accessexcellence.org/RC/VL/GG/H_geneLib.html

http://info.med.yale.edu/genetics/ward/tavi/Guide.html

http://biology.arizona.edu/sciconn/lessons/alongi/4pcrread.html

http://www.ornl.gov/sci/techresources/Human_Genome/medicine/medicine.shtml

http://www.ornl.gov/sci/techresources/Human_Genome/project/journals/journals.html

http://www.healthscout.com/ency/1/002220.html

http://www.stc.arts.chula.ac.th/BACKGROUND%20OF%20GENETICS.pdf

http://www.cmaj.ca/contents-by-date.0.shtml

http://en.wikipedia.org/wiki/Transformation_(genetics)

http://content.nejm.org/cgi/content/full/332/5/318#F1

http://en.wikipedia.org/wiki/Population_genetics

http://www.microbiologybytes.com/virology/Genetics.html

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2275739

http://upload.wikimedia.org/wikipedia/commons/thumb/c/c5/Amino_acids_2.png/600px-Amino_acids_2.png

http://botit.botany.wisc.edu/toms_fungi/dec2002.html

http://ocw.mit.edu/OcwWeb/Biology/7-03Fall-2004/Assignments/index.htm

http://www.stc.arts.chula.ac.th/LEMLIFE.html

Solutions

Index

Attributions

Collection: **Genetics**

Edited by: Professor Le Dinh Luong

URL: http://cnx.org/content/col10782/1.1/

Copyright: Professor Le Dinh Luong

License: http://creativecommons.org/licenses/by/3.0/

Module: **Syllabus**

By: Professor Le Dinh Luong

URL: http://cnx.org/content/m26598/1.1/

Copyright: Professor Le Dinh Luong

License: http://creativecommons.org/licenses/by/3.0/

Module: **Basic Principles of Genetics**

By: Professor Le Dinh Luong

URL: http://cnx.org/content/m26565/1.1/

Copyright: Professor Le Dinh Luong

License: http://creativecommons.org/licenses/by/3.0/

Module: **Eukaryotic Genetics**

By: Professor Le Dinh Luong

URL: http://cnx.org/content/m26561/1.1/

Module: **Genetics in clasical understanding**

By: Professor Le Dinh Luong

URL: http://cnx.org/content/m26573/1.1/

Module: **About the controversal ethical issues on applications of genetics**

By: Professor Le Dinh Luong

URL: http://cnx.org/content/m26548/1.1/

Module: **Assignments and solutions**

By: Professor Le Dinh Luong

URL: http://cnx.org/content/m26550/1.1/

Module: **References**

By: Professor Le Dinh Luong

URL: http://cnx.org/content/m26593/1.1/

About Connexions

Since 1999, Connexions has been pioneering a global system where anyone can create course materials and make them fully accessible and easily reusable free of charge. We are a Web-based authoring, teaching and learning environment open to anyone interested in education, including students, teachers, professors and lifelong learners. We connect ideas and facilitate educational communities. Connexions's modular, interactive courses are in use worldwide by universities, community colleges, K-12 schools, distance learners, and lifelong learners. Connexions materials are in many languages, including English, Spanish, Chinese, Japanese, Italian, Vietnamese, French, Portuguese, and Thai.