



# Eukaryotic Genetics

Bởi:

Dinh Luong Le

## 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
- (parallel 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?

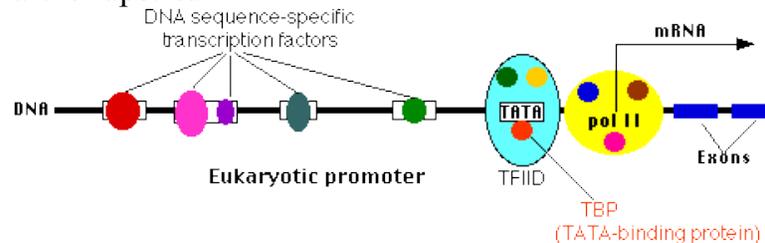
There are several methods used by eukaryotes.

- 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



- [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.](#)

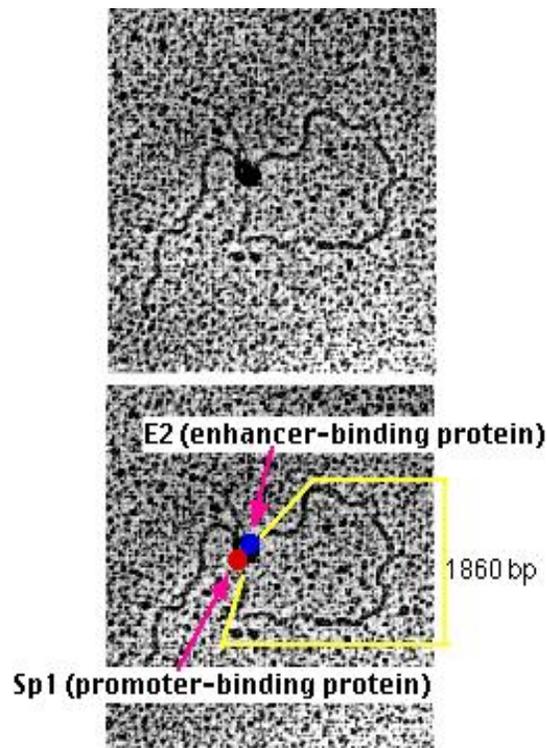
Hormones exert many of their effects by forming transcription factors.

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.

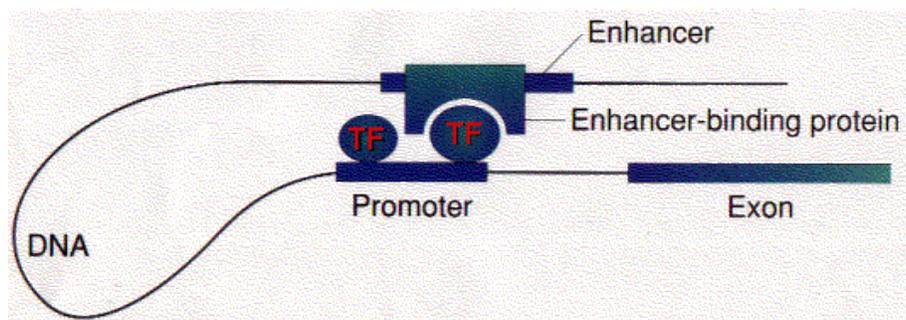
[Link to a discussion of some of the transcription factors that produce the segmented body plan in Drosophila.](#)

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.



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?



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**

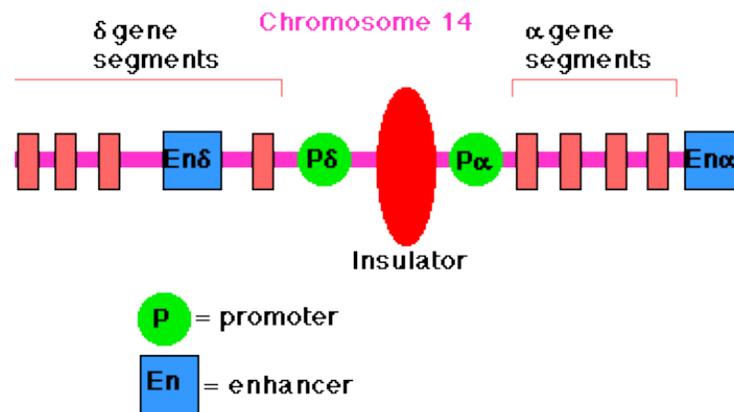
A problem:

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.



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. Theoretically tetrad analysis can be carried out in all eukaryotes. 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 Hutchinson 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 crassa*, 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.

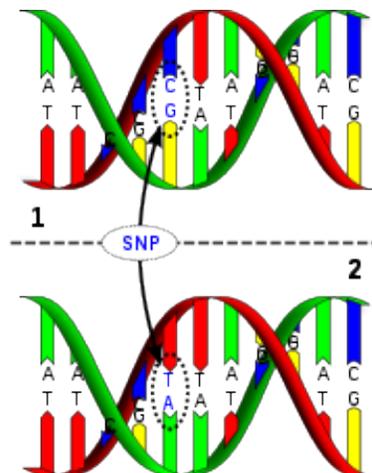
In Southern blotting, the DNA isolated from each patient or tumor to be typed is digested with a restriction enzyme, separated on the basis of size by agarose-gel electrophoresis, and transferred to a nylon membrane. A DNA probe can reveal directly on the nylon membrane the size of DNA fragments carrying 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



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".<sup>[1]</sup> 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.<sup>[2]</sup>

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](#).<sup>[3]</sup> 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. <sup>[1]</sup> 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).

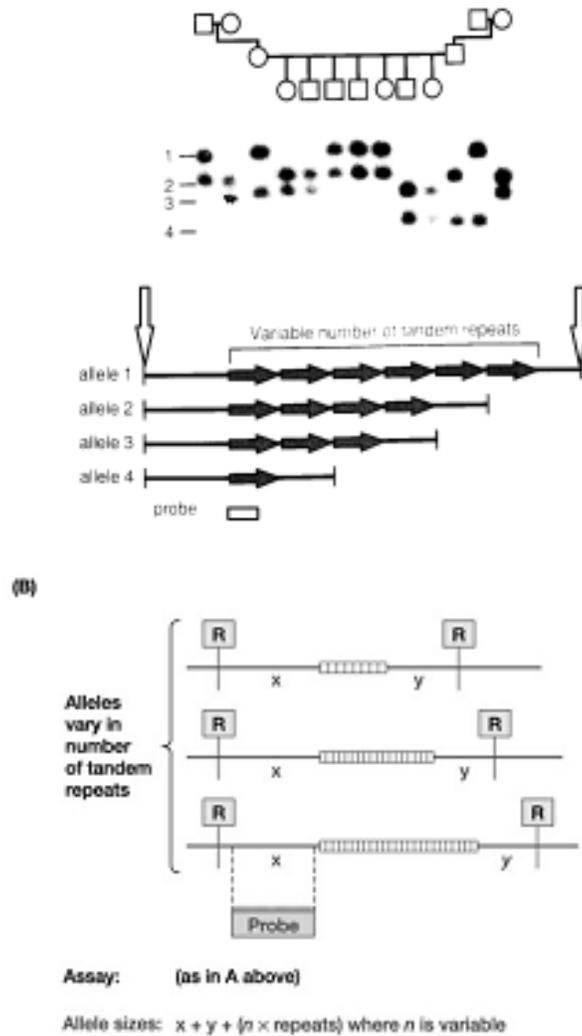
One rare example of a microsatellite is a (CA)<sub>n</sub> repeat, where n is variable between [alleles](#). These markers often present high levels of inter- and intra-specific polymorphism, particularly when tandem repeats number one hundred or greater.[2] The repeated sequence is often simple, consisting 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.



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.

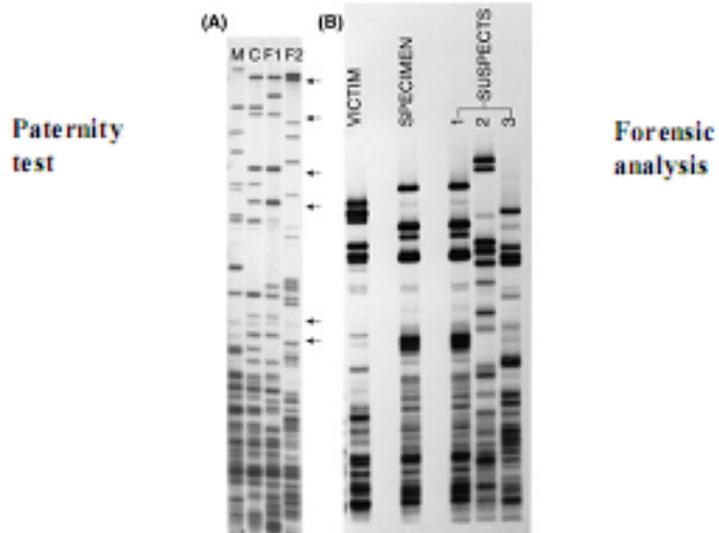
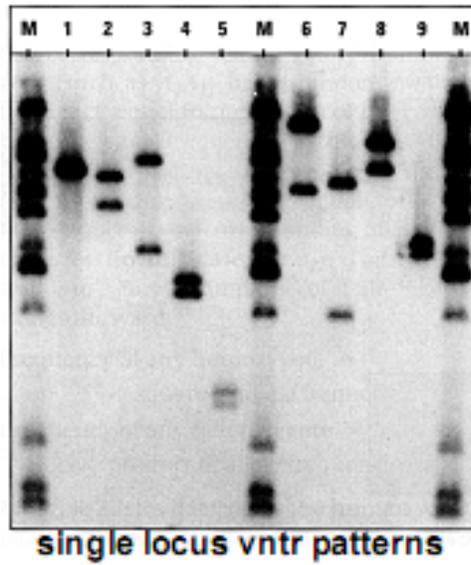


Figure 18-19: Human Molecular Genetics, 5/e. © Garland Science 2004

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