MOLECULAR MEDICINE

HUMAN DNA POLYMORPHISM

DAVID HOUSMAN, PH.D.

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 (Fig. 1). 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 (Fig. 1A).

Two techniques, Southern blotting and the polymerase chain reaction (PCR), can measure the length of the DNA sequence at the polymorphic site (Fig. 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

Figure 1. Visualization of Highly Variable DNA Sequences in Human DNA.

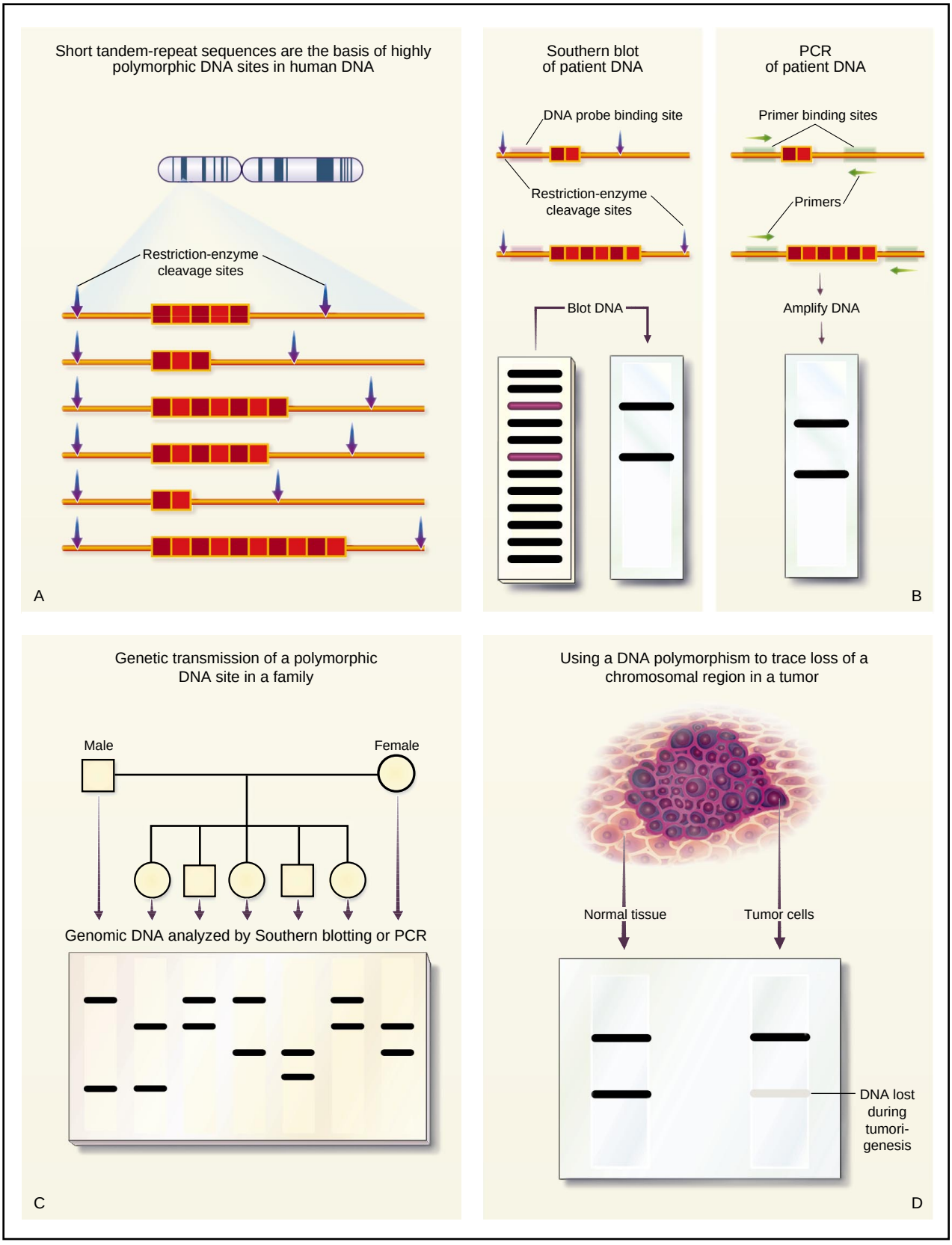
Variable-length sequences in human DNA can be created by variations in the number of copies of a tandem-repeat DNA sequence (Panel A). Each line in the figure represents a copy of a human DNA sequence. The copies are identical in sequence except for the tandemly repeated DNA sequence indicated by the boxes. The number of copies of the tandemly repeated DNA sequence is indicated by the number of boxes. The size of the DNA fragment that includes the tandem-repeat sequence is measured between two fixed points. In Southern blotting the sites of restriction-enzyme digestion are the fixed points that determine the ends of the DNA fragment.

In Panel B, the number of copies of the tandem-repeat sequence determines the total length of the fragment that can be visualized by Southern blotting. In the polymerase chain reaction (PCR) the binding sites of oligonucleotide primers are the fixed points that determine the ends of the DNA fragment. The number of copies of the tandem-repeat sequence determines the total length of the fragment that can be visualized by separating the products of the PCR reaction on a polyacrylamide gel.

In Panel C, a family in which a highly polymorphic marker is used for genetic analysis is illustrated. Note that the two copies of the DNA fragment from the offspring can be distinguished from the two copies of the fragment from the father, making the inheritance pattern from each parent clear for this chromosomal site. Detection may be carried out by Southern blotting or PCR, depending on the size of the tandem-repeat sequence.

In Panel D, a highly polymorphic site distinguishes the two copies of a chromosomal region in a person's normal cells. During tumorigenesis, one copy of the chromosomal region is lost. Only one of the variable DNA fragments (the one marking the chromosome that is not lost during tumorigenesis) remains in the tumor. A light band is evident in the tumor DNA sample at the position characteristic of the chromosomal region that has been lost during tumorigenesis, perhaps indicating contamination of the tumor sample by normal tissue. Detection may be carried out by Southern blotting or PCR, depending on the size of the tandem-repeat sequence.

From the Center for Cancer Research, Massachusetts Institute of Technology, Bldg. E17, Rm. 543, 40 Ames St., Cambridge, MA 02139, where reprint requests should be addressed to Dr. Housman.


Short tandem-repeat sequences are the basis of highly polymorphic DNA sites in human DNA
Restriction-enzyme cleavage sites
A
Southern blot of patient DNA
DNA probe binding site
Restriction-enzyme cleavage sites
Blot DNA
PCR of patient DNA
Primer binding sites
Primers
Amplify DNA
B
Genetic transmission of a polymorphic DNA site in a family
Male
Female
Genomic DNA analyzed by Southern blotting or PCR
C
Using a DNA polymorphism to trace loss of a chromosomal region in a tumor
Normal tissue
Tumor cells
DNA lost during tumori-genesis
D

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 (Fig. 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 (Fig. 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.

Recommended Reading

Nakamura Y, Leppert M, O'Connell P, et al. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 1987; 235:1616-22.

Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 1989;44:388-96.

Gyapay G, Morissette J, Vignal A, et al. The 1993–94 Généthon human genetic linkage map. Nat Genet 1994;7:246-339.