٤ علم الجينومات وتحديد تتابعات الجينات

التقرير الأول

المرجع (٢) 272-274 ص

15.5 Genomics and Gene Sequencing Have Enhanced Our Understanding of Mutations in Humans

من هنا

عززت معرفتنا للطفرات في الإنسان

So far, we have discussed the molecular basis of mutation largely in terms of nucleic acid chemistry. Historically, scientists learned about mutations by analyzing the amino acid sequences of proteins within populations that show substantial diversity. This diversity, which arises during evolution, is a reflection of changes in the triplet codons following substitution, insertion, or deletion of one or more nucleotides in the DNA sequences of protein-coding genes.

As our ability to analyze DNA more directly increases, we are able to examine the actual nucleotide sequence of genes

الصفحة التالية →

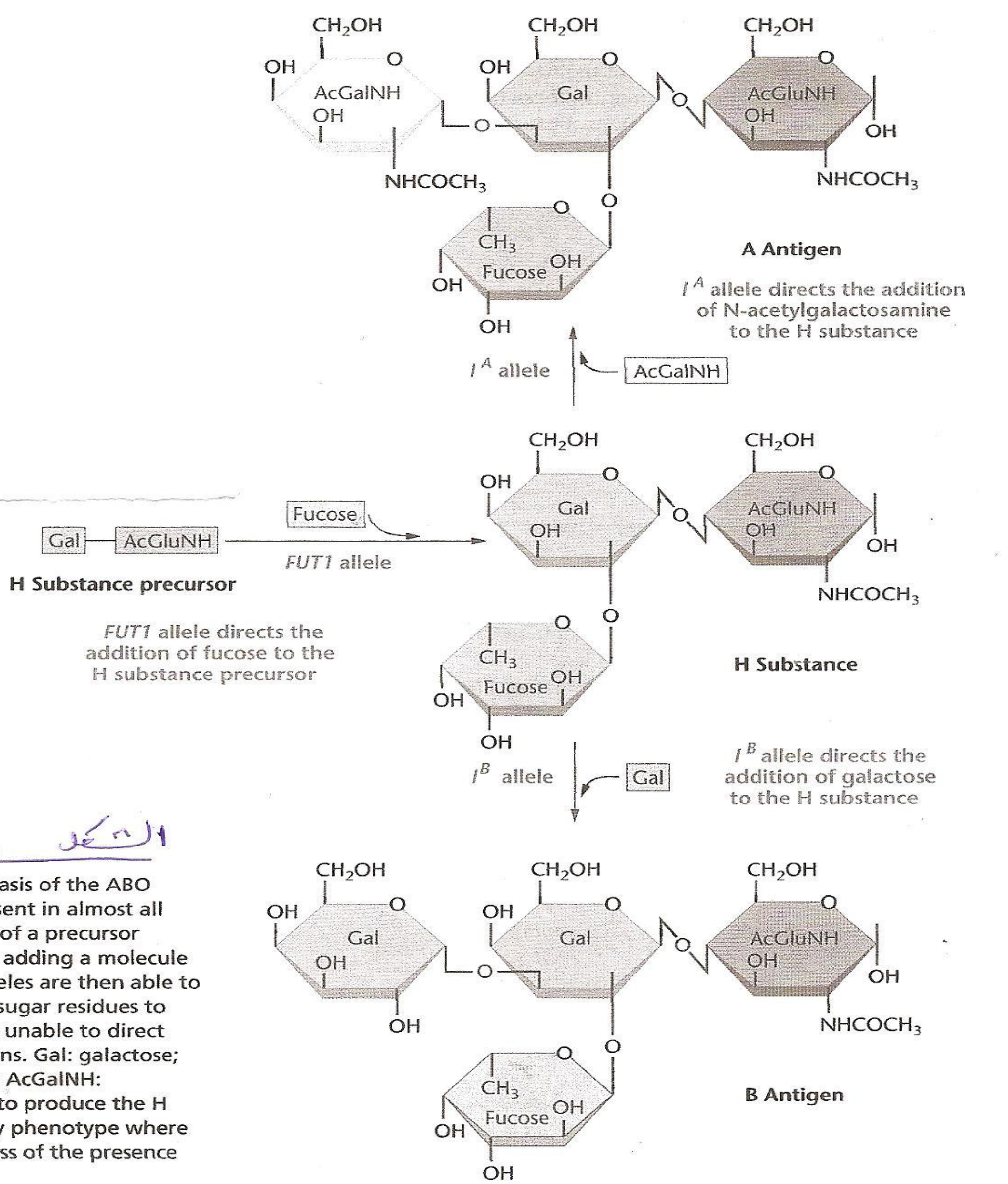

Fig. 4-2 الشكل ١

ıe biochemical basis of the ABO ... The *H* allele, present in almost all ... s the conversion of a precursor ... e H substance by adding a molecule ... The I^A and I^B alleles are then able to ... tion of terminal sugar residues to ... e. The I^O allele is unable to direct ... terminal additions. Gal: galactose; ... etylglucosamine; AcGalNH: ... osamine. Failure to produce the H ... its in the Bombay phenotype where ... type O, regardless of the presence ... llele.

and to gain greater insights into the nature of mutations. Several techniques capable of accurate, rapid sequencing of DNA (see Chapter 19) have greatly extended our knowledge of mutations. In this section, we examine the results of a number of studies that have investigated the actual gene sequences of various mutations that affect human health.

ABO Blood Types

The ABO system is based on a series of antigenic determinants found on erythrocytes and other cells, particularly epithelial cells. As we discussed in Chapter 4, there are three alleles of a single gene that encodes the glycosyltransferase enzyme. The H substance is modified to either the A or B antigen, as a result of the product of the I^A or I^B allele, respectively. Failure to modify the H substance results from the null I^O allele. (See Figure 4–2.)

The glycosyltransferase gene has been sequenced in 14 people with different ABO status. When the DNAs of the I^A and I^B alleles are compared, four consistent nucleotide substitutions are found. It is assumed that the resulting changes in the amino acid sequence of the glycosyltransferase gene product lead to the different modifications of the H substance.

The I^O allele situation is unique and interesting. Individuals homozygous for this allele have type O blood, lack glycosyltransferase activity, and fail to modify the H substance. Analysis of the DNA of this allele shows one consistent change that is unique compared with the sequences of the other alleles—the deletion of a single nucleotide early in the coding sequence, causing a frameshift mutation. A complete messenger RNA is transcribed, but at translation, the reading frame shifts at the point of the deletion and continues out of frame for about 100 nucleotides before a stop codon is encountered. At this point, the polypeptide chain terminates prematurely, resulting in a nonfunctional product.

These findings provide a direct molecular explanation of the ABO allele system and the basis for the biosynthesis of the corresponding antigens. The molecular basis for the antigenic phenotypes is clearly the result of mutations within the gene encoding the glycosyltransferase enzyme.

Muscular Dystrophy

Muscular dystrophies are a group of genetic diseases characterized by progressive muscle weakness and degeneration. There are many types of muscular dystrophy which differ in their severities, onsets and genetic causes. Two related forms of muscular dystrophy—**Duchenne muscular dystrophy (DMD)** and **Becker muscular dystrophy (BMD)**—are recessive, X-linked conditions. DMD is the more severe of the two diseases, with a rapid progression of muscle degeneration and involvement of the heart and lungs. Males with DMD usually lose the ability to walk by the age of 12 and may die in their early 20s. Because the condition is recessive and X-linked, and affected males usually die before they can reproduce, females are rarely affected by the disorder. The incidence of 1 in 5000 live male births makes DMD one of the most common life-shortening hereditary diseases. In contrast, BMD does not involve the heart or lungs and progresses slowly, from adolescence to the age of fifty or more.

The gene responsible for DMD and BMD—the *dystrophin* gene—is unusually large, consisting of about 2.5 million base pairs. In normal (unaffected) individuals, transcription and subsequent processing of the *dystrophin* initial transcript results in a messenger RNA containing only about 14,000 bases (14 kb). It is translated into the protein **dystrophin**, which consists of 3685 amino acids. In most cases of the less severe BMD, the dystrophin protein can be detected in the sarcolemma of smooth, striated, and cardiac myofibers. However, it is rarely detectable in the muscles of DMD patients. These observations led to the hypothesis that mutations causing BMD may not alter the reading frame of the gene, but mutations causing DMD may change the reading frame early in the gene, resulting in premature termination of dystrophin translation. This hypothesis would be consistent with the observed differences in severity of these two forms of the disorder.

In an extensive study performed in 1989, J. T. Den Dunnen and associates analyzed the DNA of 194 patients (160 DMD and 34 BMD). In most cases, the results were consistent with the "reading frame" hypothesis. With few exceptions, DMD mutations change the reading frame of the dystrophin gene, whereas BMD mutations usually do not. Subsequent studies have shown that about two thirds of mutations in the dystrophin gene that lead to DMD and BMD are deletions and duplications. Only one third are point mutations, and of these, the majority are small insertions or deletions. The majority of DMD mutations, both rearrangements and point mutations, lead to premature termination of translation. This, in turn, leads to degradation of the improperly translated dystrophin transcript and low levels of dystrophin protein. In contrast, the majority of BMD gene rearrangements and point mutations may alter the internal sequence of the dystrophin transcript and protein, but do not alter the translation reading frame.

These observations reflect the fact that a mutation caused by a random single-nucleotide substitution is more likely to be tolerated without a devastating phenotypic effect than the addition or loss of nucleotides that may alter the translational reading frame. There are three reasons for this:

1. A nucleotide substitution may not change the encoded amino acid, since the genetic code is degenerate.
2. If an amino acid substitution does result, the change may not be present at a location within the protein that is critical to its function.
3. Even if the altered amino acid is present at a critical region, it may still have little or no effect on the function of the protein. For example, an amino acid might be changed to another with nearly identical chemical properties or to one with very similar physical properties, such as shape.

Trinucleotide Repeats in Fragile X Syndrome, Myotonic Dystrophy, and Huntington Disease

Beginning about 1990, molecular analysis of the genes responsible for a number of inherited human disorders provided a remarkable set of observations. Researchers discovered that some mutant genes contained expansions of **trinucleotide repeat sequences**, usually from fewer than 15 copies in normal individuals to a large number in affected individuals.

For example, the genes responsible for fragile X syndrome, myotonic dystrophy, and Huntington disease contain a specific trinucleotide DNA sequence repeated many times. While these

types of repeated sequences are also present in the nonmutant (normal) allele of each gene, the mutations found in individuals with these disorders consist of significant increases in the number of times the trinucleotide is repeated. Table 15.4 summarizes the cases discussed in this section, with particular emphasis on the size of the repeats in normal and diseased individuals.

In Chapter 4 we discussed the onset of the expression of various phenotypes. In several cases, a correlation exists between the number of repeats and the age of manifestation of mutant phenotypes. The greater the number of repeats, the earlier disease onset occurs. Further, in affected individuals, the number of repeats may increase in each subsequent generation. This general phenomenon is known as **genetic anticipation**.

We begin with a short discussion of **fragile X syndrome**, described in detail in Chapter 8. The responsible gene, *FMR-1*, may have several hundred to several thousand copies of the trinucleotide sequence CGG, located in the 5′ untranslated region of the gene. Individuals with up to 54 copies are normal and do not display the mental retardation associated with the syndrome. Individuals with 54 to 230 copies are considered carriers. Although they are normal, their offspring may contain even more copies and express the syndrome. The large regions of CGG repeats in the gene's regulatory region result in loss of expression of the FMRP protein, thought to be an RNA-binding protein affecting brain cell function.

Myotonic dystrophy, or DM (short for its original name, dystrophia myotonica), is the most common form of adult muscular dystrophy, affecting 1 in 8000 individuals. The dominantly inherited disorder is not as severe as DMD and it is highly variable both in symptoms and age of onset. Mild myotonia (atrophy and weakness) of the musculature of the face and extremities is most common. Cataracts, reduced cognitive ability, and cutaneous and intestinal tumors are also part of the syndrome.

The affected gene, *MDPK*, is located on the long arm of chromosome 19. It encodes a serine–threonine protein kinase, MDPK. This protein is the product of 15 exons of the gene, the last of which encodes the 3′-untranslated RNA sequence of the mRNA. It is this sequence that houses the multiple copies of the trinucleotide CTG. Individuals with 5 to 37 copies of the CTG repeat are normal, and the number of copies is stable from generation to generation. Individuals with more than 37 copies exhibit symptoms ranging from mild to severe, with onset occurring anywhere between birth and age 60. Both the severity and onset are directly related to the size of the repeated sequence. Minimally affected patients have up to 150 repeats, while severely affected patients have up to 1500 copies of the CTG triplet. Genetic anticipation is exhibited in the offspring. The mechanism by which these repeated regions cause DM is still uncertain.

Huntington disease (HD) is a fatal neurodegenerative disease, inherited as an autosomal dominant. (Also see Chapter 4.) The gene responsible for HD is located on chromosome 4. The gene contains the trinucleotide CAG sequence, repeated 10 to 35 times in normal individuals. The CAG repeat is located within the coding region of the gene, and encodes a polyglutamine tract. The CAG repeat sequence exists in significantly increased numbers (up to 120) in diseased individuals. Much earlier onset occurs when the number of copies is closer to the upper range. Interestingly, in still another disorder, **spinobulbar muscular atrophy** (Kennedy disease), the involved gene (different from that in Huntington disease) also contains repeated copies of the CAG triplet. However, only 35 to 60 copies of the sequence cause individuals to be affected.

The role of such repeated sequences in normal and mutant genes remains a mystery. Their locations within the gene vary in each case. In Huntington and Kennedy diseases, the repeats lie within the coding portion of the gene. In the case of Huntington disease, this causes the mutant **huntingtin** protein to contain an excess of glutamine residues. Such is not the case in the other two disorders, however. In the gene responsible for fragile X syndrome, the repeat is upstream (the 5′ end) of the gene's coding region, in an area that is most often involved in regulating gene expression. In the case of myotonic dystrophy, the repeat is downstream of the coding region (the 3′ end).

The mechanism by which the repeated sequence expands from generation to generation is of great interest. Present-day thinking is that expansion can result from both errors during replication as well as in mechanisms responsible for repairing damaged DNA. Whatever the cause may be, this general instability of these short repeats seems to be more prevalent in humans than in many other organisms.

Now solve this

Problem 15.24 on page 389 asks you to speculate on why the *dystrophin* gene appears to suffer a large number of mutations.

Hint: In answering this question, you may want to think about the structure and function of this gene. Consider what features of *dystrophin* gene structure might make it susceptible to the accumulation of many mutations.

TABLE 15.4 SUMMARY OF TRINUCLEOTIDE-REPEAT DISORDERS

	Trinucleotide Repeat	Number in Normal Individuals	Number in Affected Individuals
Huntington disease	CAG	6–35	36–120
Myotonic dystrophy	CTG	5–37	37–1500
Fragile X syndrome	CGG	6–230	>230
Spinobulbar muscular atrophy	CAG	10–35	35–60