Genetic Variation

All genetic variation originates from the process known as mutation, which is defined as a change in DNA sequence.

Mutations can affect either somatic or germline cells (cells that produce gametes).

Mutations in somatic cells (all cells other than germline cells) can lead to cancer and are thus of significant concern.

Our attention will be directed to germline mutations because these can be transmitted from one generation to the next.

As a result of mutations, a gene may differ among individuals in terms of its DNA sequence. The differing sequences are referred to as alleles.

A gene location on a chromosome is termed a locus (from the Latin for ”place”). Thus, we might say that an individual has a certain allele at the B-globin locus on chromosome 11.

If an individual has the same allele on both members of a chromosome pair, he or she is said to be a homozygote. If the alleles differ in DNA sequence, the individual is heterozygote.

The alleles that are present at a given locus are referred to as the individual's genotype.

Some loci vary considerably among individuals. If a locus has two or more alleles whose frequencies each exceed 1% in a population, the locus is said to be polymorphic (many forms). The polymorphic locus is often termed a polymorphism.

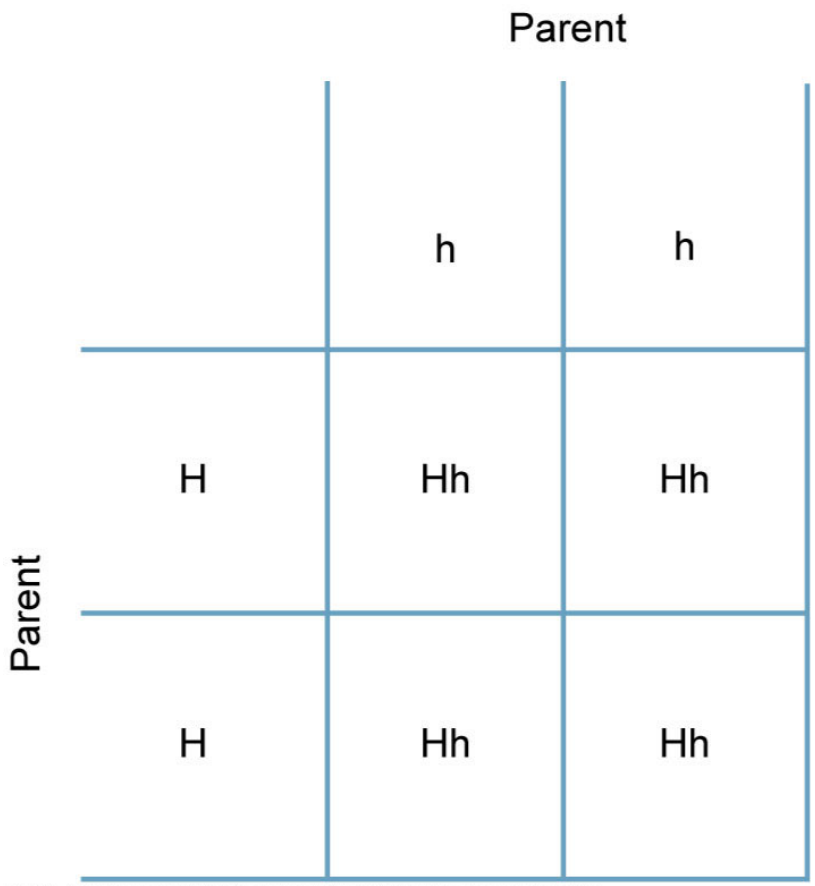


Fig. 3-1. Punnett square illustrating a cross between *HH* and *hh* homozygote parents.

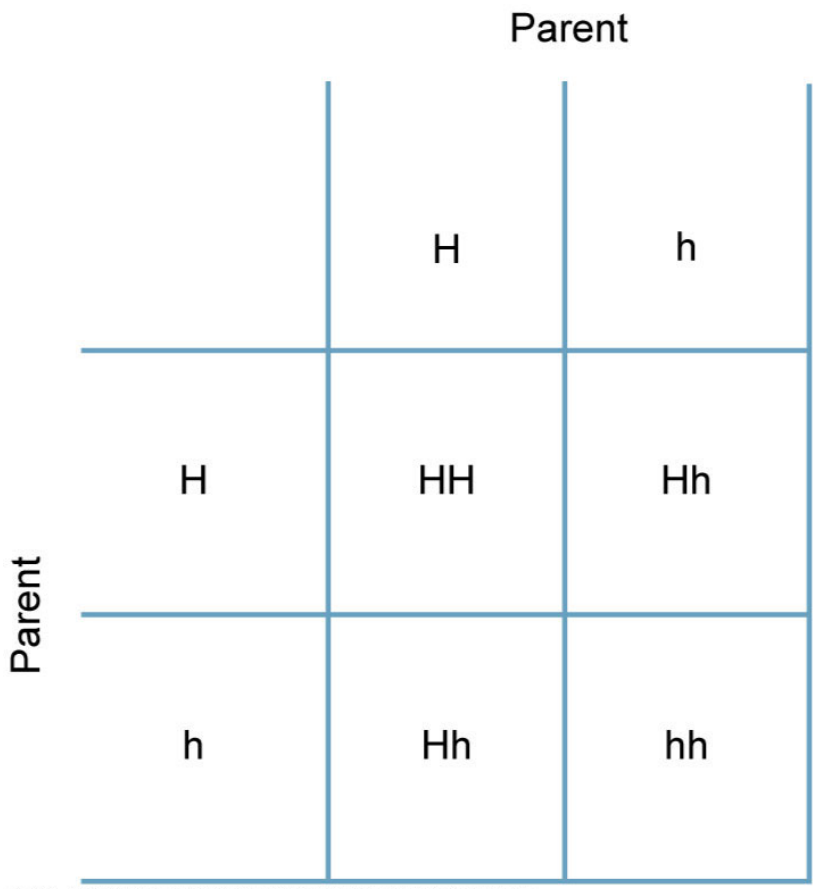


Fig. 3-2. Punnett square illustrating a cross between two *Hh* heterozygotes.

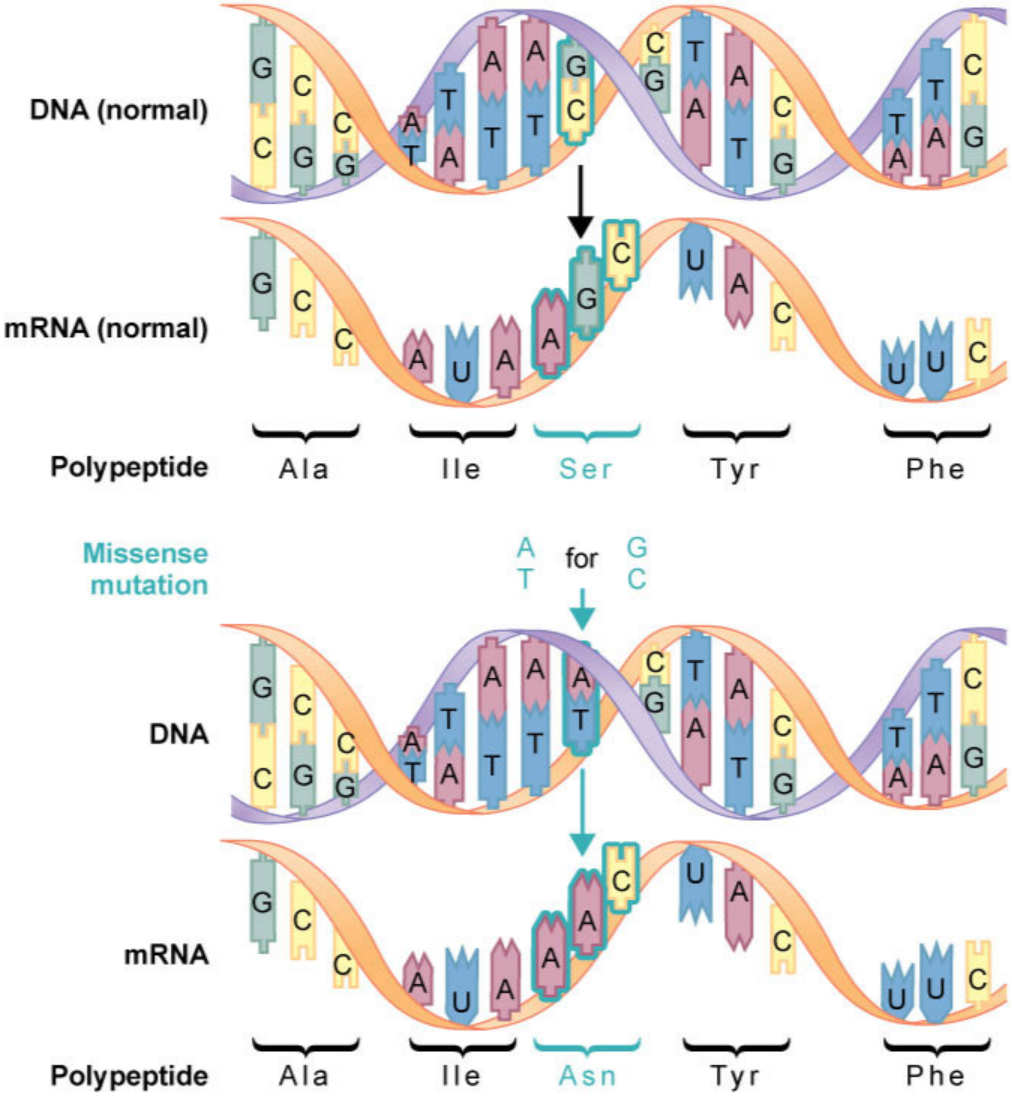
MUTATI0N—THE SOURCE OF GENETIC VARIATION
Types of Mutation

- **Base pair substitution:**
- **Silent substitution**: do not change the aa sequence
- **Non silent**:
  - **Missense mutations**: produce a change in a single aa
  - **Nonsense mutations**: produce one of the three stop codons in the mRNA (UAA, UAG, or UGA)

- **Deletions or Insertions** of one or more base pairs
  - **Frameshift mutations**

5'-ACT GAT TGC GTT-3' to 5'-ACT GA A TTG CGT-3'

Thr-Asp-Cys-Val to Thr-Glu-Leu-Arg



A

Fig. 3-3. Base pair substitution. Missense mutations **(A)** produce a single amino acid change, whereas nonsense mutations **(B)** produce a stop codon in the mRNA. Stop codons terminate translation of the polypeptide.

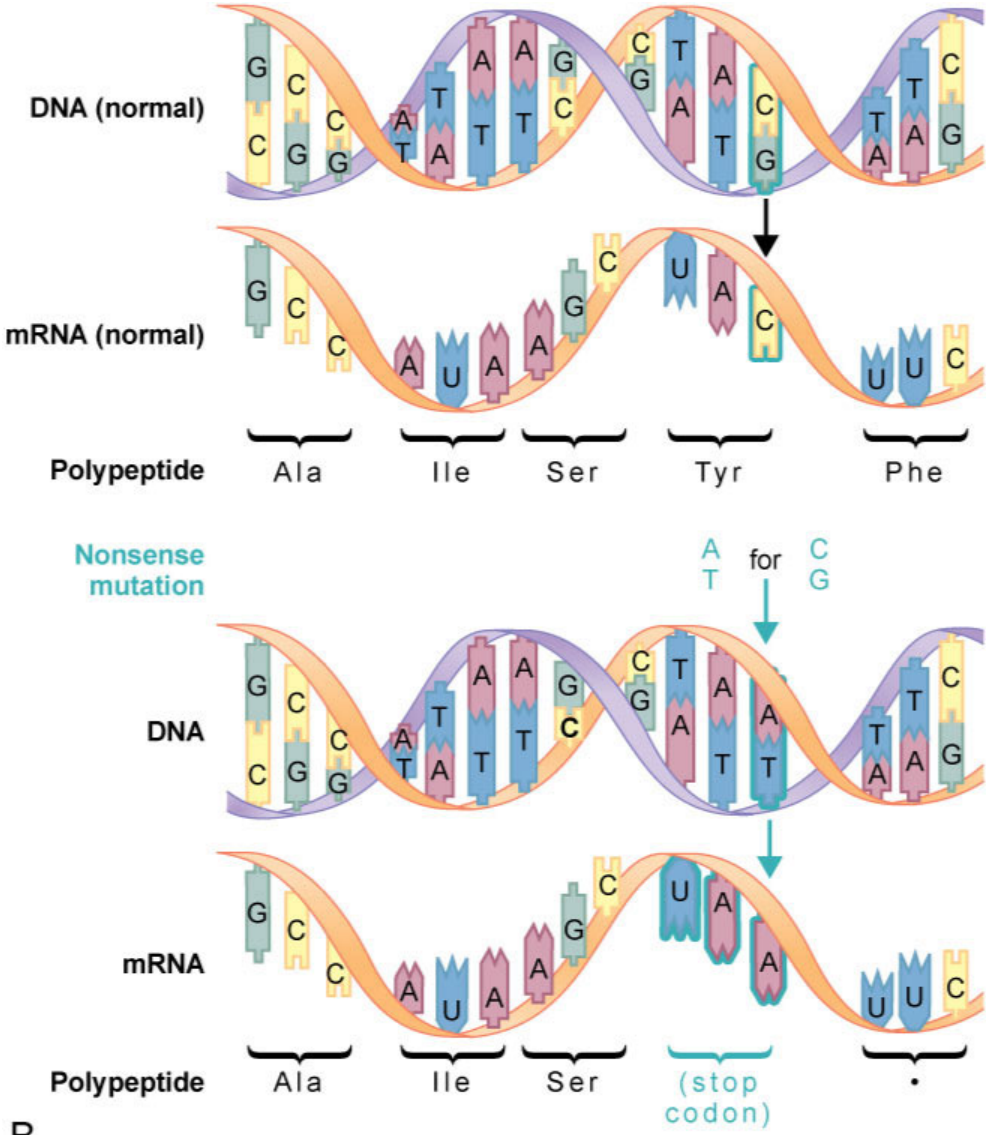


Fig. 3-3. Base pair substitution. Missense mutations **(A)** produce a single amino acid change, whereas nonsense mutations **(B)** produce a stop codon in the mRNA. Stop codons terminate translation of the polypeptide.

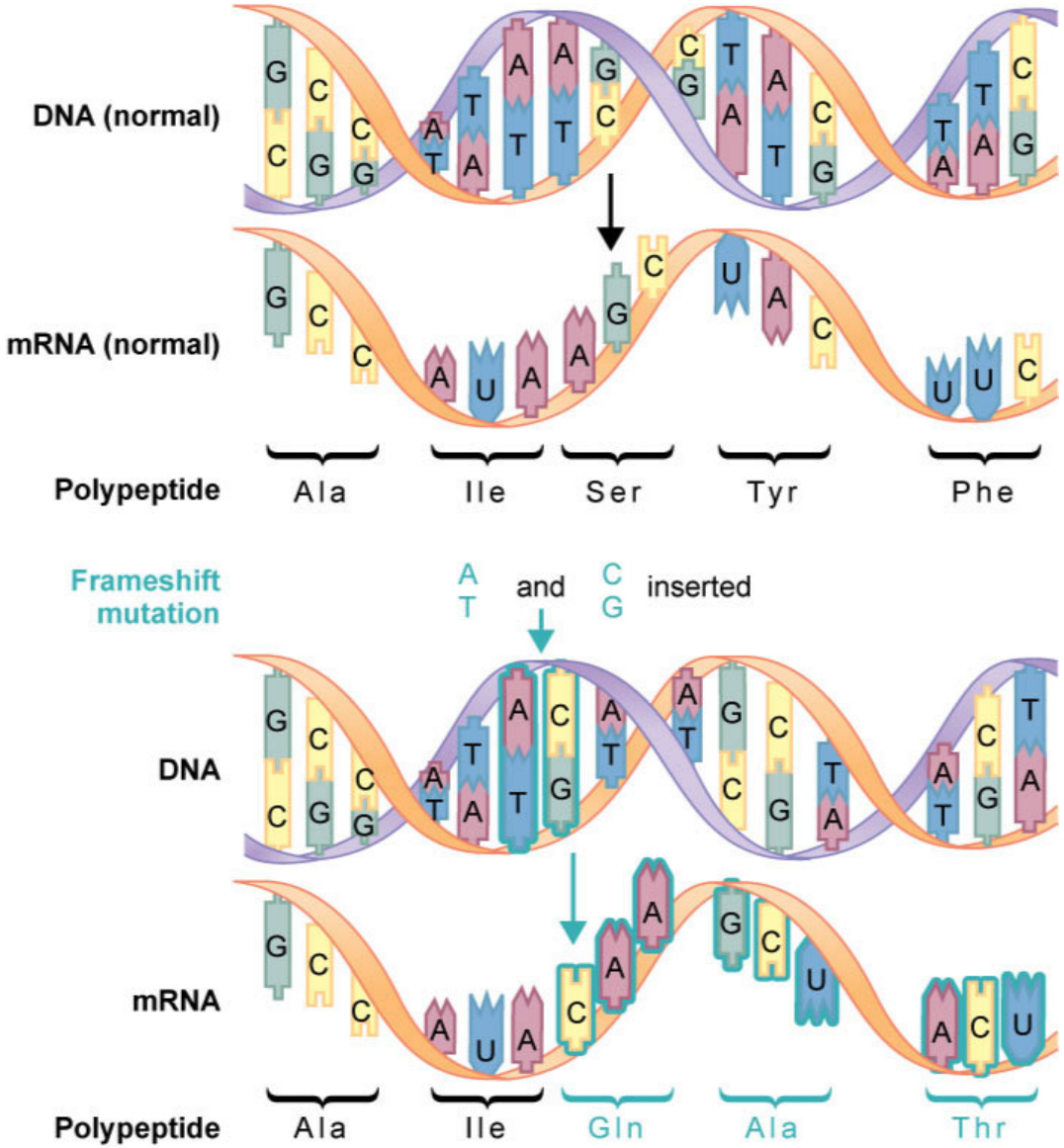


Fig. 3-4. Frameshift mutations result from the addition or deletion of a number of bases that is not a multiple of three. This alters all of the codons downstream from the site of insertion or deletion.

Types of Mutation

- **Duplications** of whole genes also lead to genetic disease

- Charcot-Marie-Tooth disease

→ Type I: 1.5 m bp duplication on one copy of Chrom. 7

→ Dosage sensitivity: peripheral myelin protein (PMP22 gene)

Other types of mutations can alter the regulation of transcription or translation.

A **promoter mutation** can decrease the affinity of RNA polymerase for a promoter site, often resulting in reduced production of mRNA. The final result is decreased production of a protein.

Mutations of transcription factor genes or enhancer sequences can have similar effects.

Mutations may also interfere with the splicing of introns as mature mRNA is formed from the primary mRNA transcript.

Splice site mutations, occurring at intron-exon boundaries, alter the splicing signal necessary for proper excision of an intron.

These may occur at the GT sequence that always defines the 5' donor site or at the AG sequence that defines the 3' acceptor site.

They may also take place in the sequences that lie near the donor and acceptor sites.

When such mutations occur, the excision will often be made within the next exon, at a splice site located in the exon.

Splice site mutations can also result in the abnormal inclusion of part or all of an intron in the mature mRNA.

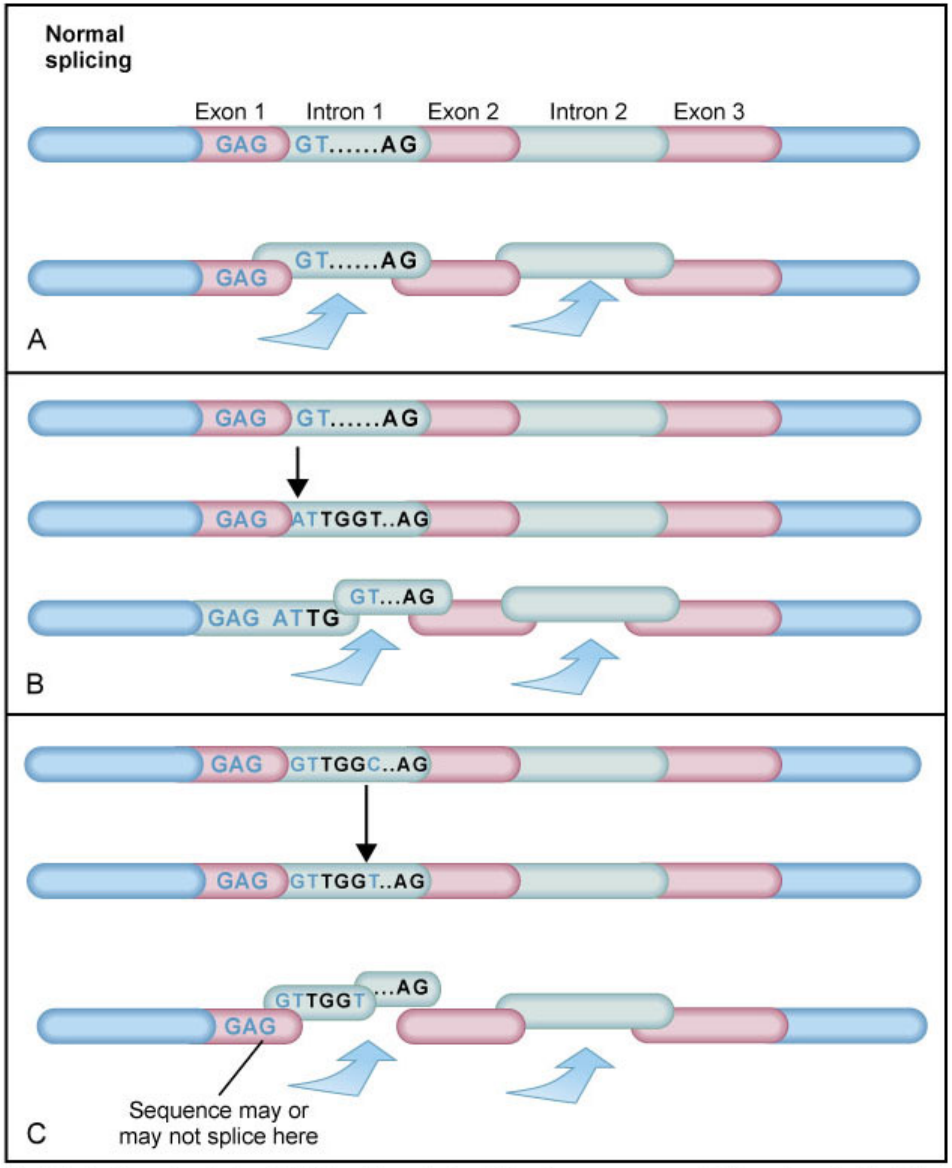


Fig. 3-5. **A,** Normal splicing. **B,** Splice-site mutation. The donor sequence, GT, is replaced with AT. This results in an incorrect splice that leaves part of the intron in the mature mRNA transcript. In another example of splice-site mutation **(C),** a second GT donor site is created within the first intron, resulting in a combination of abnormally and normally spliced mRNA products.

Several types of DNA sequences are capable of propagating copies of themselves; these copies are then inserted in other locations on chromosomes.

The insertion of these transposons (or mobile elements) can cause frameshift mutations. Until recently, it was not clear whether this phenomenon, which has been well documented in experimental animals such as fruit flies, occurred in humans.

The insertion of mobile elements has now been shown to cause isolated cases of type 1 neurofibromatosis, familial breast cancer, familial polyposis (colon cancer), and hemophilia A and B (clotting disorders) in humans.

- The final type of mutation has been discovered quite recently and affects tandem repeated DNA sequences that occur within or near certain disease genes.

- The repeat units are three bases long, so a typical example would be CAGCAGCAG
- A normal individual will have a relatively small number of these tandem repeats (eg, 20 to 30) at a specific chromosome location.
- For reasons that are not yet understood, the number of repeats can increase dramatically during meiosis or possibly during early fetal development, so that a newborn may have hundreds or even thousands of repeats.
- When this occurs in certain regions of the genome, it causes genetic disease.

- Like other mutations, these expanded repeats can be transmitted to the patients offspring. More than a dozen genetic diseases are now known to be caused by expanded repeats.

Molecular Consequences of Mutation

Mutations can result in either a gain of function or loss of function of the protein product.

Gain of function mutations occasionally result in a completely novel protein product. They result in over-expression of the product or inappropriate expression (i.e., in the wrong tissue or in the wrong stage of development).

Gain-of-function mutations produce a dominant disorder. Charcot-Marie-Tooth disease, results from overexpression of the protein product and is considered a gain-of-function mutation. Huntington disease is another example.

Loss-of-function mutations are often seen in recessive diseases. Here, a mutation results in the loss of 50% of the protein product (e.g., a metabolic enzyme) but the 50% that remains is sufficient for normal function. The heterozygote is thus unaffected.

In some cases, however, 50% of the protein product is not sufficient for normal function (haploinsufficiency), and a dominant disorder can result. Haploinsufficiency is seen, for example, in the autosomal dominant disorder familial hypercholesterolemia.

Both haploinsufficiency and gain-of-function mutations are examples of dosage sensitivity.

Another type of loss of function is seen in **the dominant negative mutation**. This type of mutation results in a protein product that is not only nonfunctional but also inhibits the function of the protein produced by the normal allele in the heterozygote.

Dominant negative mutations are seen in genes that encode multimeric proteins (i.e., proteins composed of two or more subunits).

Type 1 collagen, which is composed of three helical subunits, is an example of such a protein. An abnormal helix, created by a single mutation, may combine with the other helices, distorting them and producing a seriously compromised triple helix protein.

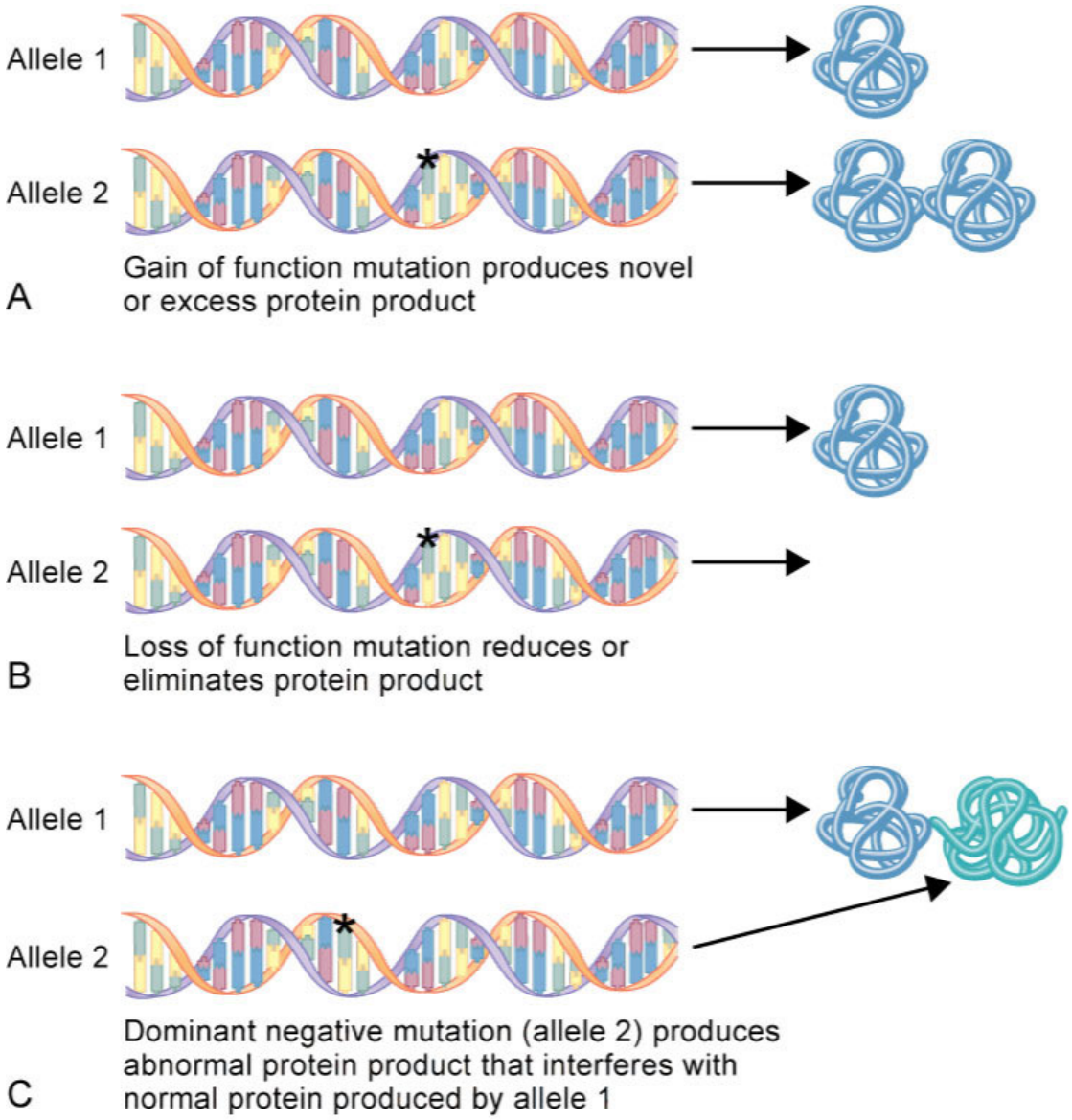


Fig. 3-6. **A,** Gain-of-function mutations produce a novel protein product or an increased amount of protein product. **B,** Loss-of-function mutations decrease the amount of protein product. **C,** Dominant negative mutations produce an abnormal protein product that interferes with the otherwise normal protein product of the normal allele in a heterozygote.

Causes of Mutation

Induced mutations. attributed to a known environmental cause. **Spontaneous mutations**, arise naturally during the process of DNA replication.

Agents that cause induced mutations are known collectively as mutagens.

Animal studies have shown that radiation is an important class of mutagen.

Ionizing radiation, such as that produced by x-rays and nuclear fallout, can eject electrons from atoms, forming electrically charged ions.
This form of radiation can reach all cells of the body, including the germ line.

Nonionizing radiation does not form charged ions but can move electrons from inner to outer orbits in the atom. The atom becomes chemically unstable.

Ultraviolet (UV) radiation, which occurs naturally in sunlight, is an example of nonionizing radiation.

UV radiation causes the formation of covalent bonds between adjacent pyrimidine bases (i.e., cytosine or thymine]. These pyrimidine dimers are unable to pair ‘properly with purines during DNA replication; this results in a base pair substitution.

Because UV radiation is absorbed by the epidermis, it does not reach the germ line but can cause skin cancer.

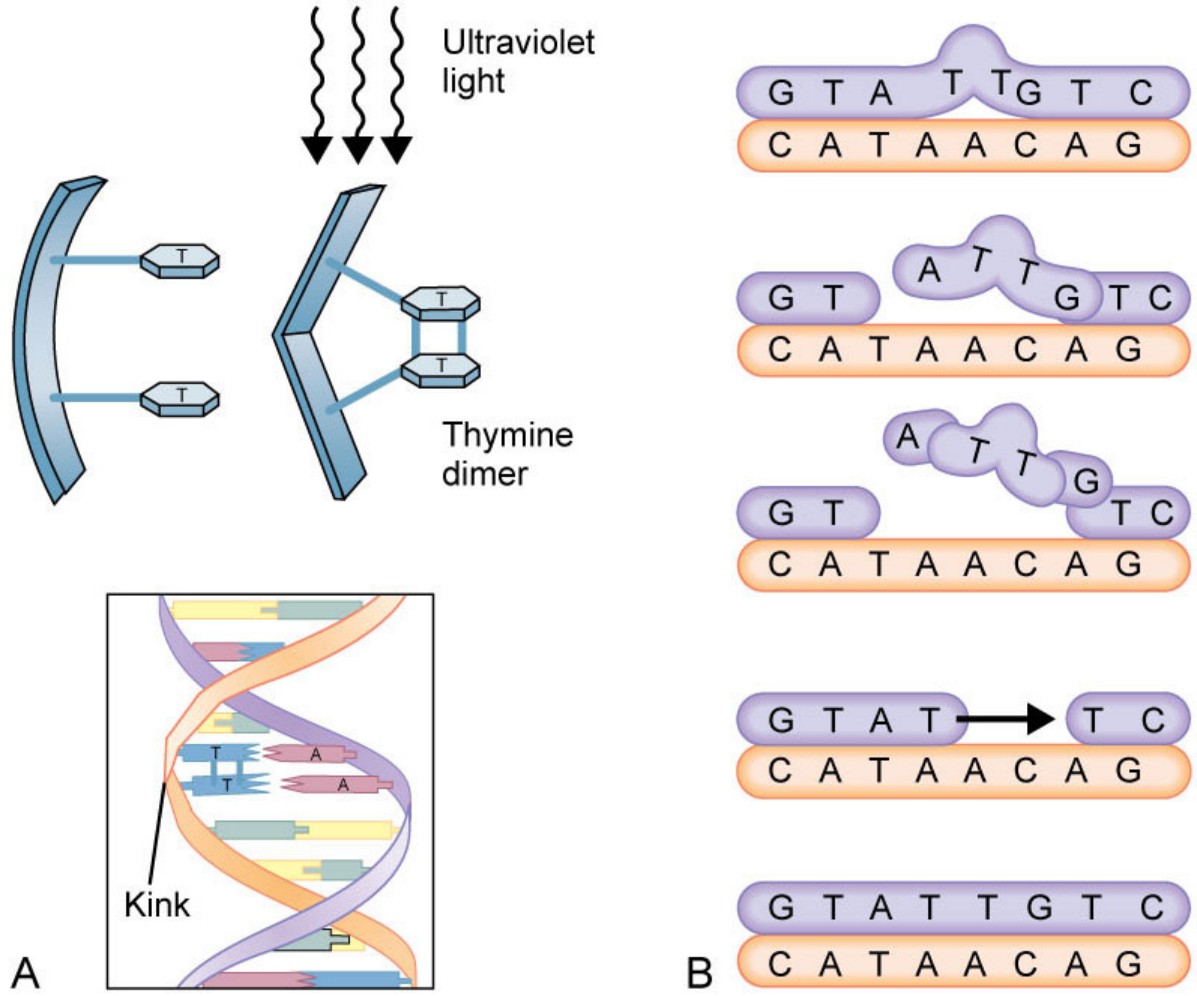


Fig. 3-11. **A,** Pyrimidine dimers originate when covalent bonds form between adjacent pyrimidine (cytosine or thymine) bases. This deforms the DNA, interfering with normal base pairing. **B,** The defect is repaired by removal and replacement of the dimer and bases on either side of it, with the complementary DNA strand used as a template.

A variety of **chemicals** can also induce mutations, sometimes because of their chemical similarity to DNA bases. Because of this similarity, these **base analogs**, such as 5-bromouracil, can be substituted for a true DNA base during‘ replication.

→ The analog is not exactly the same as the base it replaces, so it can cause pairing errors during subsequent replications.

Other chemical mutagens, such as acridine dyes, can physically insert themselves between existing bases, distorting the DNA helix and causing frameshift mutations.

Other mutagens can directly alter DNA bases, causing replication errors, like nitrous acid, which removes an amino group from cytosine, converting it to uracil.

→ Although uracil is normally found in RNA, it mimics the pairing action of thymine in DNA. So, it pairs with adenine, instead of pairing with guanine, as the original cytosine would have done. The end result is a base pair substitution.

DNA Repair

Considering that 3 billion DNA base pairs must be replicated in each cell division, and considering the large number of mutagens to which we are exposed, DNA replication is surprisingly accurate.

A primary reason for this is **DNA repair**, which takes place in all normal cells of higher organisms. Several dozen different enzymes are involved in repairing damaged DNA.

They collectively recognize an altered base, excise it by cutting the DNA strand, replace it with the correct base (determined from the complementary strand), and reseal the DNA.

It is estimated that these repair mechanisms correct 99.9% of initial errors.

Because DNA repair is essential for the accurate replication of DNA, defects in DNA repair systems can lead to many types of disease.

For example, inherited mutations in genes responsible for DNA mismatch repair result in the persistence of cells with replication errors (i.e., mismatches) and can lead to some types of cancers.

A compromised ability to repair double-stranded DNA breaks can lead to ovarian and breast cancer.

Nucleotide excision repair is necessary for the removal of larger changes in the DNA helix (e.g., pyrimidine dimers). Defects in excision repair lead to a number of diseases like xeroderma pigmentosum.

Examples of Diseases That Are Caused by a Defect in DNA Repair

TABLE 3-2 Examples of Diseases That Are Caused by a Defect in DNA Repair

DISEASE	FEATURES	TYPE OF REPAIR DEFECT
Xeroderma pigmentosum	Skin tumors, photosensitivity, cataracts, neurological abnormalities	Nucleotide excision repair defects, including mutations in helicase and endonuclease genes
Cockayne syndrome	Reduced stature, skeletal abnormalities, optic atrophy, deafness, photosensitivity, mental retardation	Defective repair of UV-induced damage in transcriptionally active DNA; considerable etiological and symptomatic overlap with xeroderma pigmentosum and trichothiodystrophy
Fanconi anemia	Anemia; leukemia susceptibility; limb, kidney, and heart malformations; chromosome instability	As many as eight different genes may be involved, but their exact role in DNA repair is not yet known
Bloom syndrome	Growth deficiency, immunodeficiency, chromosome instability, increased cancer incidence	Mutations in the reqQ helicase family
Werner syndrome	Cataracts, osteoporosis, atherosclerosis, loss of skin elasticity, short stature, diabetes, increased cancer incidence; sometimes described as "premature aging"	Mutations in the reqQ helicase family
Ataxia-telangiectasia	Cerebellar ataxia, telangiectases,* immune deficiency, increased cancer incidence, chromosome instability	Normal gene product is likely to be involved in halting the cell cycle after DNA damage occurs
Hereditary nonpolyposis colorectal cancer	Proximal bowel tumors, increased susceptibility to several other types of cancer	Mutations in any of six DNA mismatch-repair genes

*Telangiectases are vascular lesions caused by the dilatation of small blood vessels. This typically produces discoloration of the skin.

Mutation Rates

At the nucleotide level, the mutation rate is usually estimated to be about 10^{-9} per base pair per cell division.

At the level of the gene, the mutation rate is quite variable, ranging from 10^{-4} to 10^{-7} per locus per cell division. There are at least two reasons for this large range of variation.

1. Genes vary tremendously in size. The somatostatin gene, for example, is quite small, containing 1,480 bp. In contrast, the gene responsible for Duchenne muscular dystrophy (DMD) spans more than 2 million bp.
2. It is well established that certain nucleotide sequences are especially susceptible to mutation. These are termed mutation hot spots.

The best-known example is the two-base (dinucleotide) sequence CG. In mammals, about 80% of CG dinucleotides are methylated: a methyl group is attached to the cytosine base.

The methylated cytosine, 5-‘methylcytosine, easily loses an amino group, converting it to thymine. The end result is a mutation from cytosine to thymine.

Surveys of mutations in human genetic diseases have shown that the mutation rate at CG dinucleotides is about 12 times higher than at other dinucleotIde sequences.

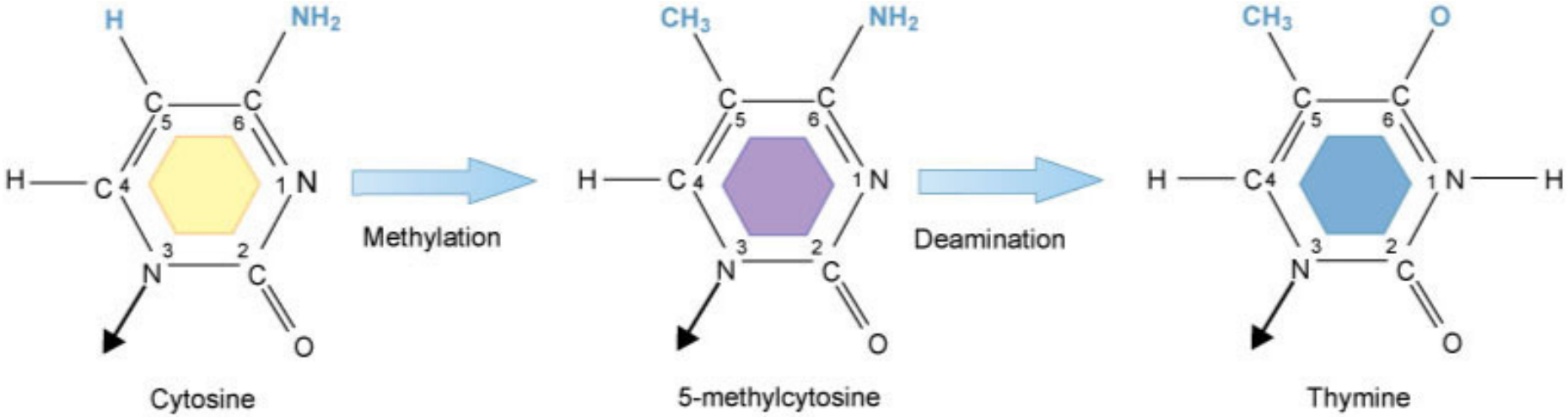


Fig. 3-12. Cytosine methylation. The addition of a methyl group (CH_3) to a cytosine base forms 5-methylcytosine. The subsequent loss of an amino group (deamination) forms thymine. The result is a cytosine → thymine substitution.

Mutation rates also vary considerably with the age of the parent, Some chromosome abnormalities increase dramatically with elevated maternal age. In addition, single-gene mutations can increase with elevated paternal age.

This increase is seen in several single-gene disorders, including Marfan syndrome and achondroplasia.

The risk of producing a child with Marfan syndrome is approximately five times higher for a male over age 40 than for a male in his 20's.

This paternal age effect is usually attributed to the fact that the stem cells giving rise to sperm cells continue to divide throughout the life of males, allowing a progressive buildup of DNA replication errors.

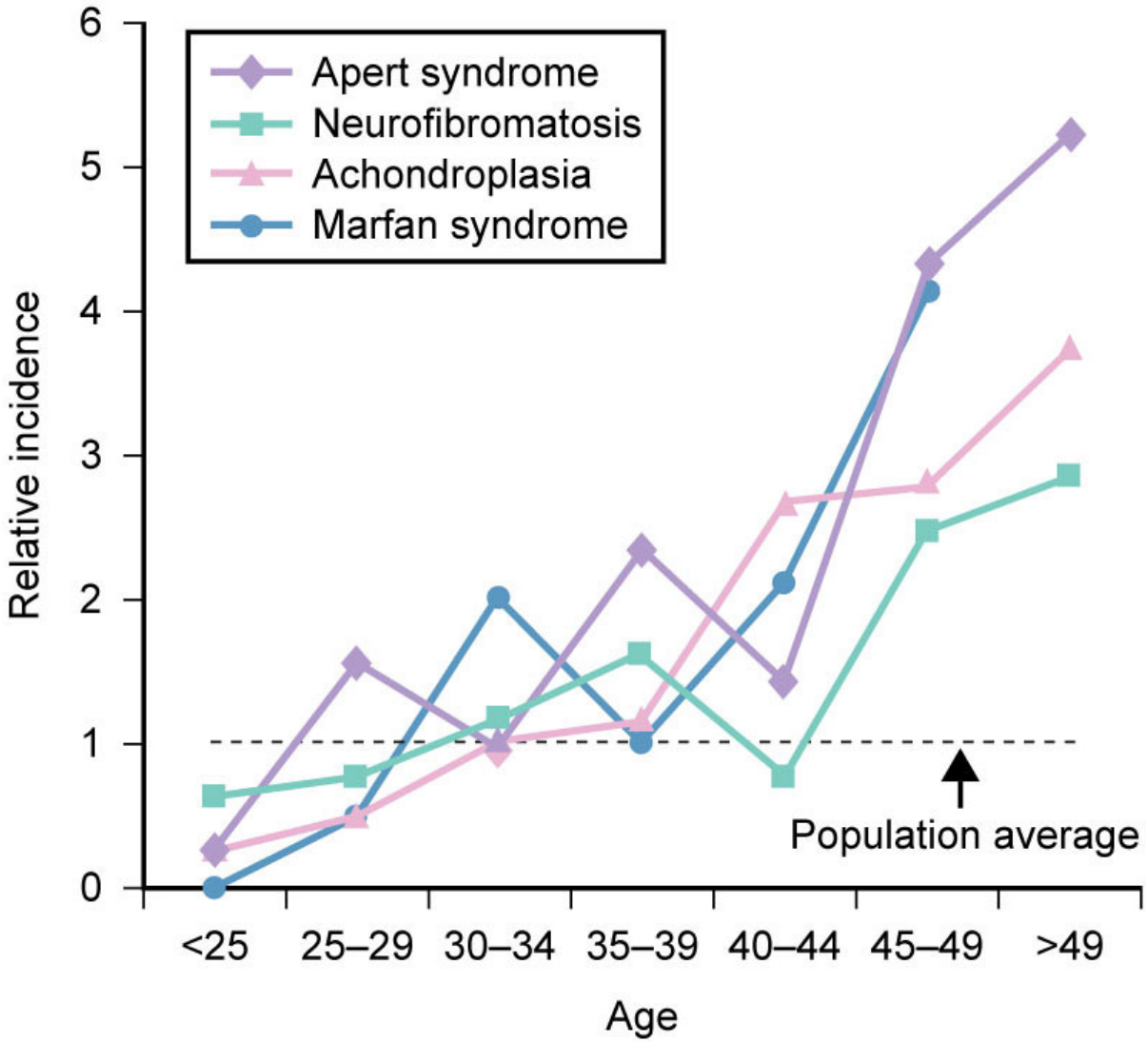


Fig. 3-13. Paternal age effect. For some single-gene disorders, the risk of producing a child with the condition (*y-axis*) increases with the father's age (*x-axis*).

Clinical Consequences of Mutation: The Hemoglobin Disorders

Genetic disorders of human hemoglobin are the most common group of single-gene disorders

An estimated 5% of the world's population carries one or more mutations of the genes involved in hemoglobin synthesis.

The hemoglobin molecule is a tetrarmer composed of four polypeptide chains, two of which are labeled α and two of which are labeled β.

The β chains are encoded by a gene on chromosome 11, and the α chains are encoded by two genes on chromosome 16 that are very similar to one another.

A normal individual would thus have two normal β genes and four normal α genes.

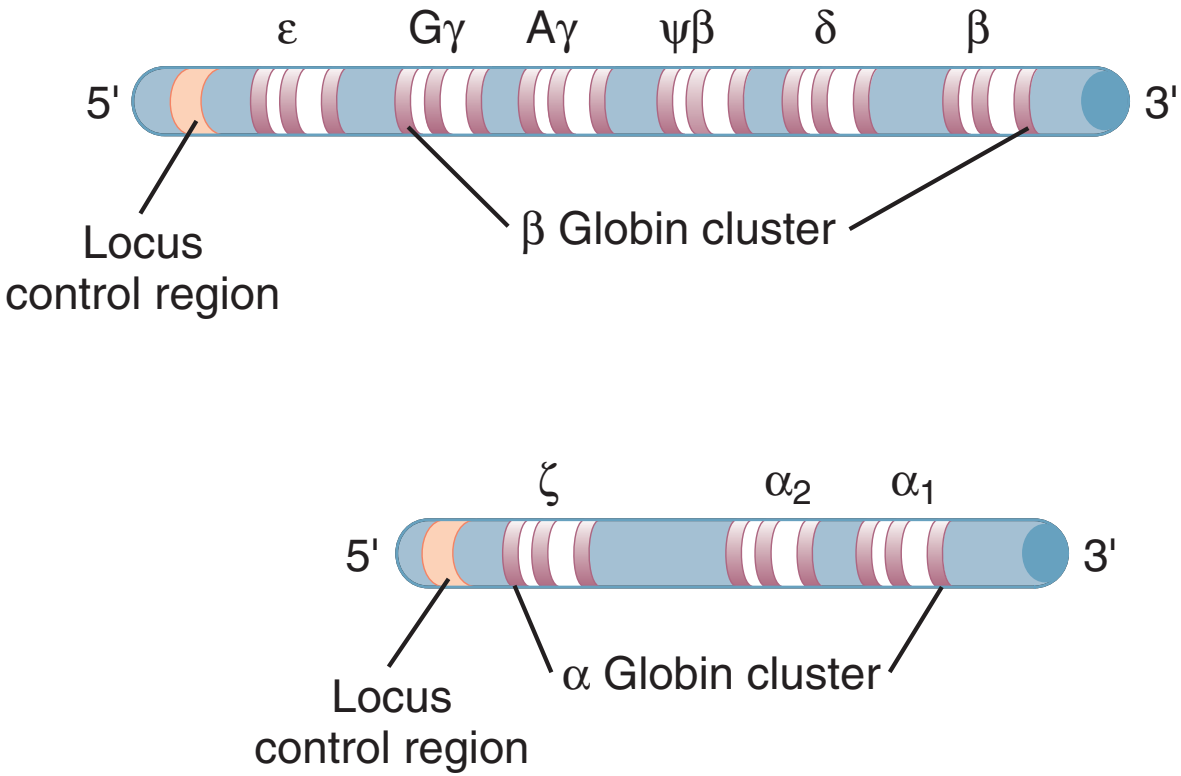


FIG 3-7 The α-globin gene cluster on chromosome 16 and the β-globin gene cluster on chromosome 11. The β-globin cluster includes the ε-globin gene, which encodes embryonic globin, and the γ-globin genes, which encode fetal globin. The ψβ gene is not expressed. The α-globin cluster includes the ζ-globin gene, which encodes embryonic α-globin.

Each of these globin chains is associated with a heme group, which contains an iron atom and binds with oxygen.

This property allows hemoglobin to perform the vital function of transporting oxygen in erythrocytes.

The hemoglobin disorders can be classified into two broad categories:

structural abnormalities, in which the hemoglobin molecule is altered,

and thalassemias, a group of conditions in which the hemoglobin is structurally normal but reduced in quantity.

Summary of the major hemoglobin disorders

TABLE 3-1 Summary of the Major Hemoglobin Disorders

DISEASE	MUTATION TYPE	MAJOR DISEASE FEATURES
Sickle cell disease	β-globin missense mutation	Anemia, tissue infarctions, infections
HbH disease	Deletion or abnormality of three of the four α-globin genes	Moderately severe anemia, splenomegaly
Hydrops fetalis (Hb Barts)	Deletion or abnormality of all four α-globin genes	Severe anemia or hypoxemia, congestive heart failure; stillbirth or neonatal death
β^0-Thalassemia	Usually nonsense, frameshift, or splice-site donor or acceptor mutations; no β-globin produced	Severe anemia, splenomegaly, skeletal abnormalities, infections; often fatal during first decade if untreated
β^+-Thalassemia	Usually missense, regulatory, or splice-site consensus sequence or cryptic splice-site mutations; small amount of β-globin produced	Features similar to those of β^0-thalassemia but often somewhat milder

Sickle Cell Disease

The most important of the structural hemoglobin abnormalities is sickle cell disease, a disorder that affects approximately 1/400 to 1/600 African-American births.

It is even more common in parts of Africa, where it can affect up to one in 50 births, and it is also seen occasionally in Mediterranean and Mideastern populations.

Sickle cell disease is caused by a single missense mutation that effects a substitution of valine for glutamic acid at position six of the B-globin polypeptide chain.

In homozygous form, this amino acid substitution alters the characteristics of the hemoglobin molecule such that the erythrocytes assume a characteristic ”sickle" shape under conditions of low oxygen tension.

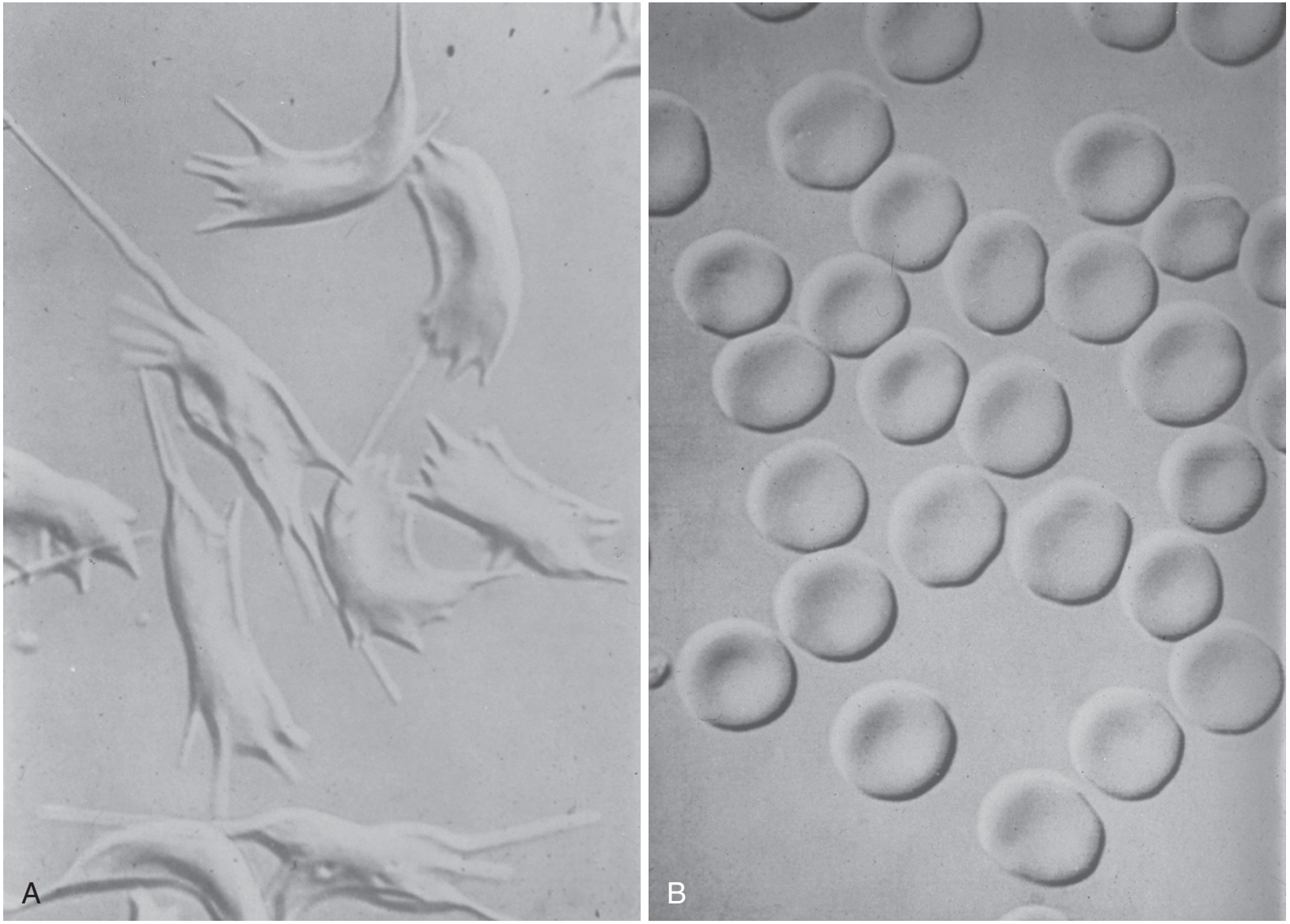


FIG 3-8 A, Erythrocytes from patients with sickle cell disease assume a characteristic shape under conditions of low oxygen tension. **B,** Compare with normal erythrocytes.

Normal erythrocytes can squeeze through capillaries, but sickled erythrocytes are less flexible and are unable to do so.

The resultant vascular obstruction produces localized hypoxemia, painful sickling ”crises," and infarctions of various tissues, including bone, spleen, kidneys, and lungs.

Premature destruction of the sickled erythrocytes decreases the number of circulating erythrocytes and the hemoglobin level, producing anemia.

The spleen becomes enlarged (splenomegaly), but infarctions eventually destroy this organ, producing some loss of immune function.

This contributes to the recurrent bacterial“ infections (especially pneumonia) that are commonly seen in individuals with sickle cell disease and that frequently cause death. In North America, it is estimated that about 15% of children with sickle cell disease die before the age of 5 years.

Thalassemia

The term ”thalassemia” is derived from the Greek word for ”sea," thalassa.

→ thalassemia was first described in populations living near the Mediterranean Sea. It is also common in portions of Africa, the Mideast, India, and Southeast Asia.

Divided into two major groups, α-thalassemia and β-thalassemia, depending on the globin chain that is reduced in quantity.

When one type of chain is decreased in number, the other chain type, unable to participate in normal tetramer formation, tends to form molecules consisting of four chains of the excess type only [these are termed homotetramers, in contrast to the heterotetramers normally formed by α and β chains).

In α-thalassemia, the α-globin chains are deficient, so the β chains (or γ chains in the fetus) are found in excess. They form homotetramers that have a greatly reduced oxygen-binding capacity, producing hypoxemia.

In β-thalassemia, β-chains are reduced or absent, so the excess α-chains form homotetramers that precipitate and damage the cell membranes of red cell precursors. This leads to premature erythrocyte destruction and anemia.

Most cases of α-thalassemia are caused by deletions of the α-globin genes. A loss of one or two of these genes has no clinical effect.

The loss or abnormality of three of the α genes produces moderately severe anemia and splenomegaly (Hb H disease).

Loss of all four α genes, a condition seen primarily among Southeast Asians, produces hypoxemia in the fetus and hydrops fetalis. Severe hypoxemia invariably causes Stillbirth or neonatal death.

Individuals with a β-globin mutation in one copy of chromosome 11 (heterozygotes) are said to have B-thalassemia minor, a condition that involves little or no anemia and does not ordinarily require clinical management.

Those in whom both copies of the chromosome carry a β-globin mutation develop B-thalassemia major (also called Cooley's anemia) or the less serious condition, β-thalassemia intermedia.

β-globin may be completely absent (β°-thalassemia) or reduced to about 10% to 30% of the normal amount (β^{+}-thalassemia).

Because β-globin is not produced until after birth, the effects of β-thalassemia major are not seen clinically until the age of 2 to 6 months.

These patients develop severe anemia. If the condition is left untreated, substantial growth retardation can occur.

The anemia causes bone marrow expansion, which in turn produces skeletal changes, including a protuberant upper jaw and cheekbones and thinning of the long bones (making them susceptible to fracture).

Splenomegaly and infections are common, and untreated β-thalassemia major patients often die during the first decade of life.

B-thalassemia can vary considerably in severity, depending on the precise nature of the responsible mutation.

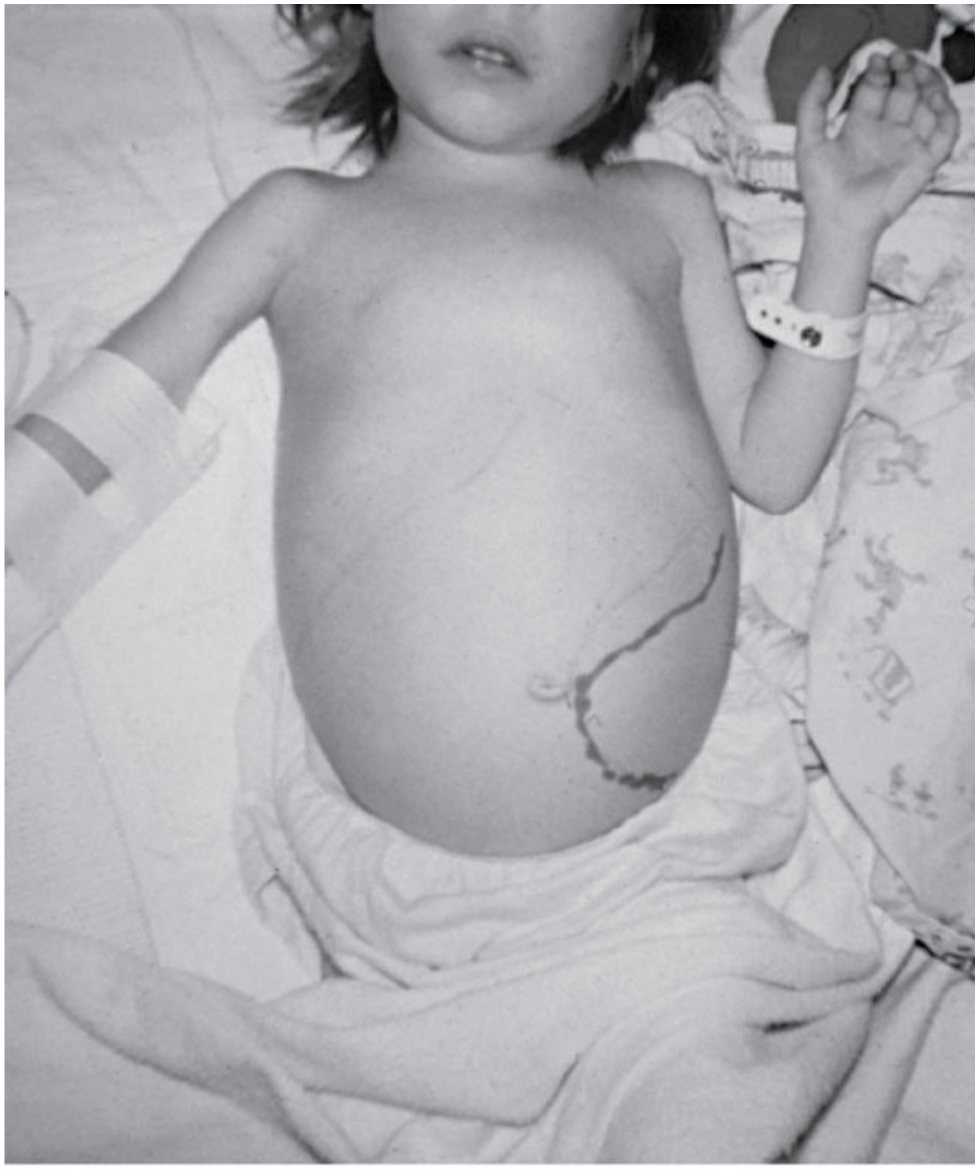


Fig. 3-9. A child with β-thalassemia major who has severe splenomegaly.

In contrast to α-thalassemia, gene deletions are relatively rare in β-thalassemia. Instead, most cases are caused by single-base mutations.

Nonsense mutations, which result in premature termination of translation of the β-globin chain, usually produce β°-thalassemia. Frameshift mutations also typically produce the β° form.

In addition to mutations in the β-globin gene itself, alterations in regulatory sequences are often seen.

β-globin transcription is regulated by a promoter, two enhancers, and an upstream region known as the locus control region (LCR).

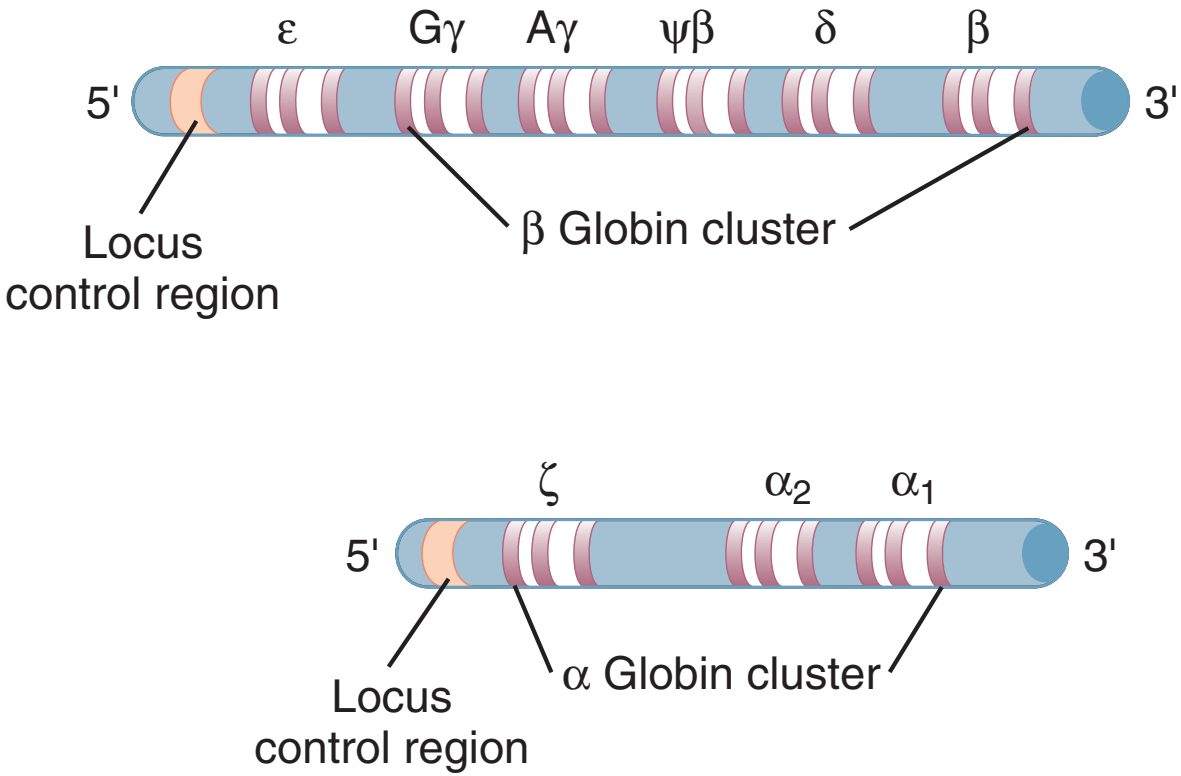


FIG 3-7 The α-globin gene cluster on chromosome 16 and the β-globin gene cluster on chromosome 11. The β-globin cluster includes the ε-globin gene, which encodes embryonic globin, and the γ-globin genes, which encode fetal globin. The ψβ gene is not expressed. The α-globin cluster includes the ζ-globin gene, which encodes embryonic α-globin.

Mutations in the regulatory regions usually result in reduced synthesis of mRNA and a reduction, but not complete absence, of β-globin (β^+-thalassemia).

Several types of splice site mutations have also been observed. When a point mutation occurs at the donor or acceptor sites, normal splicing is destroyed completely, producing B°-thalassemia.

Mutations in the surrounding consensus sequences usually produce B^+-thalassemia.

Mutations also occur in the cryptic splice sites found in introns or exons of the B-globin gene, causing these sites to be available to the splicing mechanism.

These additional splice sites then compete with the normal splice sites, producing some normal and some abnormal β-globin chains. The result is usually β^+-thalassemia.

More than 300 different β-globin mutations have been reported.

Consequently, most β-thalassemia patients are not ”homozygotes" in the strict sense: they usually have a different β-globin mutation on each copy of chromosome 11 and are termed compound heterozygotes.

Even though the mutations differ, each of the two β-globin genes is altered, producing a disease state. It is quite common to apply the term ”Homozygote” loosely to compound heterozygotes.

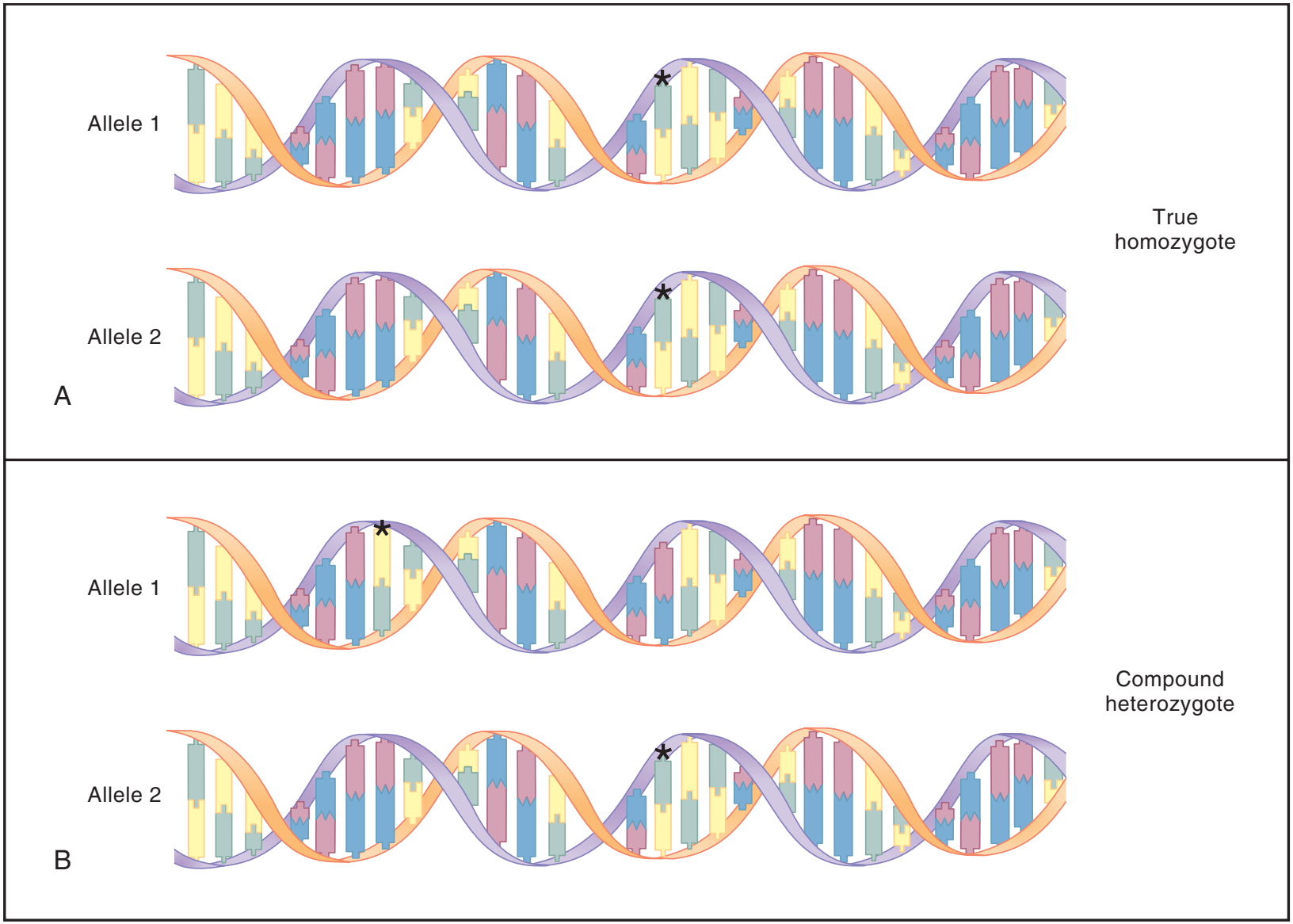


FIG 3-10 **A,** True homozygotes have two alleles that are identical in DNA sequence. Here, the homozygote has two copies of a single-base mutation, shown by the *asterisk* in the same position in the DNA sequence. Both mutations (alleles 1 and 2) have a loss-of-function effect, giving rise to a recessive disease. **B,** The same effect is seen in a compound heterozygote, which has two different mutations *(asterisks)* in two different locations in the gene's DNA sequence. Each allele has a loss-of-function effect, again causing a recessive disease.

Sickle cell and B-thalassemia major patients are sometimes treated with blood transfusions and with chelating agents that remove excess iron introduced by the transfusions.

Prophylactic administration of and antipneumococcal vaccination are used to diminish bacterial infections in sickle cell disease patients, and analgesics are administered for pain relief during sickling crises.

Bone marrow transplants, which provide donor stem cells that produce genetically normal erythrocytes, have been performed on patients with severe B-thalassemia and sickle cell disease.

It is often impossible to find a suitably matched donor, and the mortality rate from these transplants is still fairly high (approximately 5% to 30%, depending on the severity of disease and the age of the patient).

The hemoglobin disorders are a possible candidate for gene therapy