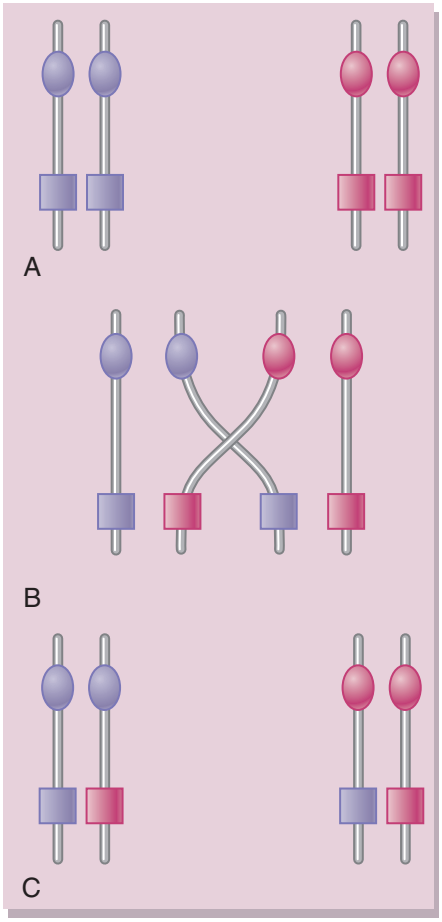

FIGURE 1-6 Crossing over and recombination. **A,** Two haploid chromosomes are shown, one from each parent (*red* and *blue*) with two genomic loci denoted by the *circles* and *squares*. **B,** Crossing over of one haploid chromosome from each parent. **C,** Resulting recombination of chromosomal segments redistributes one haploid locus *(squares)* from one diploid pair to another.

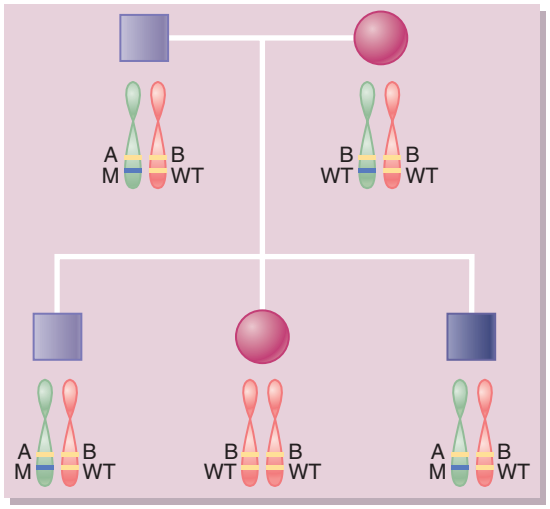

FIGURE 1-7 Linkage analysis. Analysis of the association (i.e., genomic contiguity) of a mutation *(M)* and a polymorphic allelic marker *(A)* shows close linkage in that the mutation segregates with the *A* allele, whereas the wild-type gene locus *(WT)* associates with the *B* allele.

In 1998, the Celera private venture proposed a goal similar to that of the Human Genome Project using a revolutionary approach, called *shotgun sequencing*, to determine the sequence of the human genome (http://www.dnai.org/c/index.html). The shotgun sequencing method was designed for random, large-scale sequencing and subsequent alignment of sequenced segments using computational and mathematic modeling. In the end, the Human Genome Project in collaboration with the Celera private venture produced a refined map of the entire human genome in 2001.

Because of the differences in genomic sequence that arise as a consequence of normal biologic variations or sequence polymorphisms, the resulting restriction fragment length polymorphisms (RFLPs) differ among individuals and are inherited according to mendelian principles. These polymorphisms can serve as genetic markers for specific loci in the genome. One of the most useful types of RFLP for localization of genetic loci in the genome is that produced by tandem repeats of sequence. Tandem repeats arise through *slippage* or stuttering of the DNA polymerase during replication in the case of STRs; longer variations arise through unequal crossover events. STRs are distributed throughout the genome and are highly polymorphic. These markers have two different alleles at each locus that are derived from each parent; the origins of the two chromosomes can be discerned through this analysis.

The use of highly polymorphic tandem repeats that occur throughout the genome as genomic markers has provided a basis for mapping specific gene loci by establishing the association or linkage with select markers. Linkage analysis is predicated on a simple principle: the likelihood that a crossover event will occur during meiosis decreases the closer the locus of interest is to a given marker. The extent of genetic linkage can be ascertained for any group of loci, one of which may contain a disease-producing mutation (Fig. 1-7).

IDENTIFYING MUTANT GENES

Deducing the identity of a specific gene sequence thought to cause a specific human disease requires identification of mutations in the gene of interest. If the gene suspected to be responsible for the disease phenotype is known, its sequence can be determined by conventional cloning and sequencing strategies, and the mutation can be identified. A variety of techniques are available for detecting mutations. Mutations that involve insertion or deletion of large segments of DNA can be detected by Southern blot, in which the isolated DNA is annealed to a radioactively labeled fragment of a cDNA sequence. Prior incubation of the DNA with a specific restriction endonuclease cleaves the DNA sequence of interest at specific sites to produce smaller fragments that can be monitored by agarose gel electrophoresis. Shifts in mobility on the gel in comparison with wild-type sequence become apparent as a function of changes in the molecular size of the fragment. Alternatively, the polymerase chain reaction (PCR) can be used to identify mutations.

In the PCR approach, small oligonucleotides (20 to 40 bases long), which are complementary to regions of DNA that bracket the sequence of interest and are complementary to each strand of the double-stranded DNA, are synthesized and used as primers for the amplification of the DNA sequence of interest. These