TABLE 1-1 MOLECULAR BASIS OF MUTATIONS

TYPE	EXAMPLES
POINT MUTATIONS	
Deletion	α-Thalassemia, polycystic kidney disease
SUBSTITUTIONS	
Silent	Cystic fibrosis
Missense	Sickle cell anemia, polycystic kidney disease, congenital long QT syndrome
Nonsense	Cystic fibrosis, polycystic kidney disease
LARGE MUTATIONS (GENE OR GENE CLUSTER)	
Deletion	Duchenne's muscular dystrophy
Insertion	Factor VIII deficiency (hemophilia A)
Duplication	Duchenne's muscular dystrophy
Inversion	Factor VIII deficiency
Expanding triplet	Huntington's disease
VERY LARGE MUTATIONS (CHROMOSOMAL SEGMENT OR CHROMOSOME)	
Deletion	Turner's syndrome (45,X)
Duplication	Trisomy 21
Translocation	XX male [46,X; t(X;Y)]*

*Translocation onto an X chromosome of a segment of a Y chromosome that bears the locus for testicular differentiation.

Differences in DNA sequences studied within the context of a population are referred to as *genetic polymorphisms*, and these polymorphisms underlie the diversity observed within a given species and among species.

Despite the high prevalence of benign polymorphisms in a population, the occurrence of harmful mutations is rare because of selective pressures that eliminate the most harmful mutations from the population (i.e., lethality) and the variability within the genomic sequence in response to polymorphic change. Some portions of the genome are remarkably stable and free of polymorphic variation, whereas other portions are highly polymorphic, the persistence of variation within which is a consequence of the functional benignity of the sequence change. In other words, polymorphic differences in DNA sequence between individuals can be categorized as those producing no effect on phenotype, those causing benign differences in phenotype (i.e., normal genetic variation), and those producing adverse consequences in phenotype (i.e., mutations). The latter group can be further subdivided into the polymorphic mutations that alone are able to produce a functionally abnormal phenotype such as monogenic disease (e.g., sickle cell anemia) and those that alone are unable to do so but in conjunction with other mutations can produce a functionally abnormal phenotype (i.e., complex disease traits such as essential hypertension).

Polymorphisms are more common in noncoding regions of the genome than they are in coding regions, and one common type involves the tandem repetition of short DNA sequences a variable number of times. If these tandem repeats are long, they are called *variable number tandem repeats* (VNTRs); if the repeats are short, they are called *short tandem repeats* (STRs). During mitosis, the number of tandem repeats can change, and the frequency of this kind of replication error is high enough to make alternative lengths of the tandem repeats common in a population. However, the rate of change in length of the tandem repeats is low enough to make the size of the polymorphism useful as a stable genotypic trait in families, and polymorphic tandem repeats are useful in determining the familial heritability of specific genomic loci.

Polymorphic tandem repeats are sufficiently prevalent along the entire genomic sequence, enabling them to serve as genetic markers for specific genes of interest by analysis of their linkage to those genes during crossover and recombination events. Analyses of multiple genetic polymorphisms in the human genome (i.e., genotyping) reveal that a remarkable variation exists among individuals at the level of the sequence of genomic DNA. A single-nucleotide polymorphism (SNP), the most common variant, differs by a single base between chromosomes on a given stretch of DNA sequence (Fig. 1-5). From genotyping of the world's representative population, 10 million variants (i.e., one site per 300 bases) are estimated to make up 90% of the common SNP variants in the population, with the rare variants making up the remaining 10%. With each generation of a species, the frequency of polymorphic changes in a gene is 10^{-4} to 10^{-7}. In view of the number of genes in the human genome, between 0.5% and 1.0% of the base sequence of the human genome is polymorphic. In this context, the new variant can be traced historically to the surrounding alleles on the chromosomal background present at the time of the mutational event.

A haplotype is a specific set or combination of alleles on a chromosome or part of a chromosome (see Fig. 1-5). When parental chromosomes undergo crossover, new *mosaic* haplotypes that contain additional mutations are created from the recombination. SNP alleles within haplotypes can be co-inherited with other alleles in the population, a mechanism called *linkage disequilibrium* (LD). The association between two SNPs declines with increasing distance, enabling patterns of LD to be identified from the proximity of nearby SNPs. Conversely, a few well-selected SNPs are often sufficient to predict the location of other common variants in the region.

Haplotypes associated with a mutation are expected to become common by recombination in the general population over thousands of generations. In contrast, genetic mapping with LD departs from traditional mendelian genetics by using the entire human population as a large family tree without an established pedigree. Of the possible 10 million variants, the International HapMap Project and the Perlegen private venture have deposited more than 8 million variants comprising the public human SNP map from more than 341 people representing different population samples. The SNPs distributed across the genome of unrelated individuals provide a sufficiently robust sample set for statistical associations to be drawn between genotypes and modest phenotypes. A mutation is now defined as a specific type of allelic polymorphism that causes a functional defect in a cell or organism.

The causal relationship between monogenic diseases with well-defined phenotypes that co-segregate with the disease requires only a small number of affected individuals compared with unaffected control individuals. In contrast, complex disorders (e.g., diabetes, hypertension, cancer) necessitate the combinatorial effects of environmental factors and genes with subtle effects. Only by searching for variations in genetic frequency between patients and the general population can the causation of