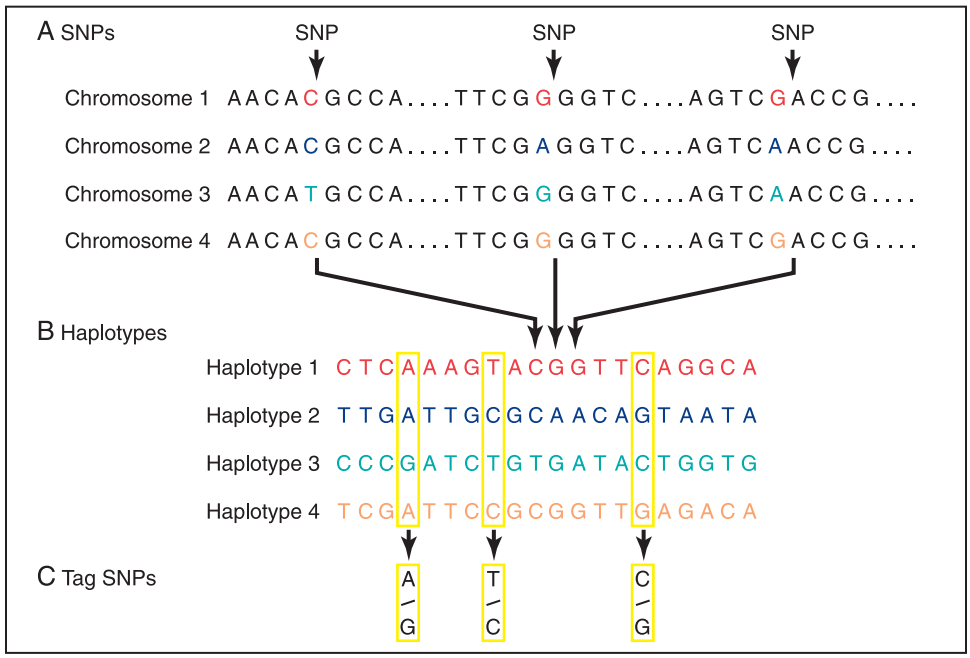

FIGURE 1-5 Single nucleotide polymorphisms (SNPs), haplotypes, and tag SNPs. Stretches of mostly identical DNA on the same chromosome are shown for four individuals. SNP refers to the variation of the three bases shown in a DNA region. The combination of nearby SNPs (**A**) defines a haplotype. Tag SNPs (**C**) are useful tools for genotyping four unique haplotypes from the 20 haplotypes (**B**). (Modified from International HapMap Consortium: The International HapMap Project, Nature 426:789–796, 2003.)

disease be discerned. In the postgenomic era, gene mapping entails the statistical association with the use of LD and high-density genetic maps that span thousands to 100,000 base pairs. To enable comprehensive association studies to become routine in clinical practice, inexpensive genotyping assays and denser maps with all common polymorphisms must be linked to all possible manifestations of the disease. Longitudinal studies of the HapMap and Perlegen cohorts can determine the effects of diet, exercise, environmental factors, and family history on future clinical events. Without similar approaches to securing adequate sample sizes and datasets, the promise of genetic population theory will not overcome the inherent limitations of linking human sequence variation with complex disease traits.

GENE MAPPING AND THE HUMAN GENOME PROJECT

The process of gene mapping involves identifying the relative order and distance of specific loci along the genome. Maps can be of two types: genetic and physical. Genetic maps identify the genomic location of specific genetic loci by a statistical analysis based on the frequency of recombination events of the locus of interest with other known loci. Physical maps identify the genomic location of specific genetic loci by direct measurement of the distance along the genome at which the locus of interest is located in relation to one or more defined markers. The precise location of genes on a chromosome is important for defining the likelihood that a portion of one chromosome will interchange, or cross over, with the corresponding portion of its complementary chromosome when genetic recombination occurs during meiosis (Fig. 1-6).

During meiotic recombination, genetic loci or alleles that have been acquired from one parent interchange with those acquired from the other parent to produce new combinations of alleles, and the likelihood that alleles will recombine during meiosis varies as a function of their linear distance from one another in the chromosomal sequence. This recombination probability (i.e., distance) is commonly quantitated in centimorgan (cM) units; 1 cM is the chromosomal distance over which there is a 1% chance that two alleles will undergo a crossover event during meiosis. Crossover events serve as the basis for mixing parental base sequences during development, promoting genetic diversity among offspring. Analysis of the tendency for specific alleles to be inherited together indicates that the recombination distance in the human genome is about 3000 cM.

Identifying the gene or genes responsible for a specific polygenic disease phenotype requires an understanding of the topographic anatomy of the human genome, which is inextricably linked to interactions with the environment. The Human Genome Project, first proposed in 1985, represented an international effort to determine the complete nucleotide sequence of the human genome, including the construction of its detailed genetic, physical, and transcript maps, with identification and characterization of all genes. This foray into large-scale biology was championed by Nobel Laureate James Watson as the defining moment in his lifetime for witnessing the path from the double helix to the sequencing of 3 billion bases of the human genome, paving the way for understanding human evolution and harnessing the benefits for human health.

Among the earliest achievements of the Human Genome Project was the development of 1-cM resolution maps, each containing 3000 markers, and the identification of 52,000 sequenced tagged sites. For functional analysis on a genome-wide scale, major technologic advances were made, including as high-throughput oligonucleotide synthesis, normalized and subtracted complementary DNA (cDNA) libraries, and DNA microarrays.