INTRODUCTION

Genomics refers to the study or science of genomes, which contain all of the DNA that codes for an organism. Knowing the human genome has implications for biology and mankind that are unparalleled, in part because having such an abundance of information regarding the inner workings of humans is unprecedented. Because a genomic sequence is a blueprint of the biological activities of an organism, determining what regulates a genome has implications that stimulate the imagination. For example, there are probably only 200–300 genes that provide susceptibility for the 20 diseases that account for 80% of all deaths globally, and most of these genes will be identified in the next 5 years. The genomes and genes of the organisms Haemophilus influenzae and Treponema pallidum have been identified (1). These genes and their end products provide novel targets for the development of antibiotics and vaccines. The cloning of growth factors that selectively modulate cardiac myocyte growth or inhibit fibrosis is to be expected.

In the 1940s, it was established that DNA—not proteins—passes on inherited characteristics. In humans, the genome consists of the DNA contained in the 23 pairs of chromosomes that are enclosed in the nucleus. Twenty-two of these chromosomes are paired with a homologous chromosome and are referred to as autosomes. The remaining two chromosomes are the sex chromosomes, of which females have two X chromosomes and males have an X and a Y chromosome. The X and Y chromosomes share only a small area of homology. Homologous chromosomes share the same set of genes, but because each set came from a different parent, there are minor differences between the DNA sequences. These differences may or may not alter gene function, although they are very important genetic markers and are frequently used for mapping and identifying genes.

Each chromosome is a single, long, linear molecule of DNA. The three-dimensional structure of DNA consists...
of two long helical strands coiled around a common axis, forming a double helix. Each strand of DNA is composed of four different monomers called nucleotides. The two strands of DNA are coiled antiparallel to each other and are held together by hydrogen bonds formed between the bases. A nucleotide consists of a phosphate group linked by a phosphodiester bond to a pentose (five-carbon sugar molecule) that in turn is linked to an organic base. In DNA, the pentose is deoxyribose; in RNA it is ribose. The number of organic bases found in DNA is limited to four: adenine (A), guanine (G), thymine (T), and cytosine (C). The same is true for RNA, except it contains uracil (U) instead of thymine. Thus, each chromosome is a macromolecule of repeat sequences of various combinations of the four bases—A, C, G, and T. DNA sequences are read from left to right, with the proximal sequence (left) referred to as the 5' end and the distal sequence (right) as the 3' end. The hydrogen bonds that hold the two strands of DNA together form between the organic bases in a precise and obligatory manner. Adenine can only pair with thymine, and it does so through two hydrogen bonds. Guanine can only pair with cytosine, and it does so through three hydrogen bonds. This complementary base pairing between the two strands contributes to the stability of the double helix and is also a unique feature of DNA. When DNA is denatured—for example, by increased temperature—the hydrogen bonds break and the helix separates into two single strands. Restoration of the lower temperature induces annealing, and the two DNA strands come together in the same precise order as prior to separation. Practically all DNA diagnostic tests capitalize upon DNA's ability to denature and reanneal.

STRUCTURE OF THE HUMAN GENOME

If joined head to tail, the 46 human chromosomes would span about 3 m. Yet the average cell only spans about 0.0015 cm, and the nucleus containing the chromosomes occupies less than 10% of the volume of the cell. The chromosomes, compacted to occupy very little space, are wound around proteins—primarily around histone proteins—in a specific structure that has important implications for gene regulation. The combination of DNA and protein is referred to as chromatin. The family of histones that comprise the bulk of chromatin are rich in lysine, a positively charged amino acid that binds to negatively charged phosphate groups in DNA. As viewed using an electron microscope, chromatin exhibits bead-like structures referred to as nucleosomes. Nucleosomes are about 30 nm in diameter and are the primary structural unit of chromatin. The DNA, wound around the surface of the protein core, makes slightly less than two turns for every 146 basepairs (bp) of DNA. DNA that is not undergoing transcription is more compact—and probably more tightly bound to histones—than DNA that is undergoing transcription. Acetylation and deacetylation of histone proteins is now recognized as fundamental to the regulation of gene expression (2). Histone acetyltransferases (HAT) acetylate the lysine groups on histones, which disrupts histone binding to DNA and results in relaxation of the chromatin and increased transcription. Histone deacetylases (HDAC) antagonize this action and repress transcription. There are 16 known HDACs in humans. Class II HDACs are known to repress transcription of myocyte enhancer factor (MEF-2), one of the pivotal transcription factors in myogenesis (3). When chromosomes are fully compacted, access of RNA polymerase and possibly other proteins required for transcription is blocked. Thus, chromatin compaction and decompaction determines whether a gene will be transcribed and expressed into its functional unit—the protein.

GENE STRUCTURE AND REGULATION

The total human genome has about 3 billion bp. The average chromosome has about 135 million bp. The longest chromosome (chromosome 1) has over 250 million bp, and the smallest chromosome (chromosome 21) has only 50 million. The hereditary characteristics of an individual are determined by the linear sequence of the four bases and are passed on by genes. A gene is a distinct unit of heredity whose transcript, or messenger RNA (mRNA), encodes a single polypeptide. Genes have a start and stop site, and they vary in size from 10,000 to 2 million bp. It is estimated that an average human gene is 20,000 bp long. Genes themselves do not participate in cellular function, but through their intermediary mRNA they code for a polypeptide with specific amino acids. The precise linear sequence of the amino acids in a polypeptide corresponds to the linearity of the codons, which are triplets of bases, in the gene. Each codon encodes a specific amino acid. It is estimated that the human genome has about 40,000 genes. Thus, genes represent only about 2% of human DNA (Table 1). About 20–30% of the DNA is transcribed into mRNA, but most of it is spliced out in the process of making mature mRNA that will exit the nucleus and serve as the polypeptide template. Those sequences that form mRNA for proteins