Genes

Each DNA molecule contains many genes—the basic physical and functional units of heredity. A gene is a specific sequence of nucleotide bases, whose sequences carry the information required for constructing proteins, which provide the structural components of cells and tissues as well as enzymes for essential biochemical reactions. The human genome is estimated to comprise at least 100,000 genes.

Human genes vary widely in length, often extending over thousands of bases, but only about 10% of the genome is known to include the protein-coding sequences (exons) of genes. Interspersed within many genes are intron sequences, which have no coding function. The balance of the genome is thought to consist of other noncoding regions (such as control sequences and intergenic regions), whose functions are obscure. All living organisms are composed largely of proteins; humans can synthesize at least 100,000 different kinds. Proteins are large, complex molecules made up of long chains of subunits called amino acids. Twenty different kinds of amino acids are usually found in proteins. Within the gene, each specific sequence of three DNA bases (codons) directs the cell's protein-synthesizing machinery to add specific amino acids. For example, the base sequence ATG codes for the amino acid methionine. Since 3 bases code for 1 amino acid, the protein coded by an average-sized gene (3000 bp) will contain 1000 amino acids. The genetic code is thus a series of codons that specify which amino acids are required to make up specific proteins.

The protein-coding instructions from the genes are transmitted indirectly through messenger ribonucleic acid (mRNA), a transient intermediary molecule similar to a single strand of DNA. For the information within a gene to be expressed, a complementary RNA strand is produced (a process called transcription) from the DNA template in the nucleus. This

Comparative Sequence Sizes	Bases
• Largest known continuous DNA sequence (yeast chromosome 3)	350 Thousand
• *Escherichia coli* (bacterium) genome	4.6 Million
• Largest yeast chromosome now mapped	5.8 Million
• Entire yeast genome	15 Million
• Smallest human chromosome (Y)	50 Million
• Largest human chromosome (1)	250 Million
• Entire human genome	3 Billion

Fig. 3. Comparison of Known Sequence with Chromosome and Genome Sizes. *A comparison of the size of the largest DNA fragment for which the sequence has been determined, with approximate chromosome and genome sizes of model organisms and humans. A major focus of the Human Genome Project is the development of sequencing schemes that are faster and more economical.*