

Human Genome Project Goals

Area	Goal	Achieved	Date
Genetic Map	2- to 5-cM resolution map (600 – 1,500 markers)	1-cM resolution map (3,000 markers)	September 1994
Physical Map	30,000 STSs	52,000 STSs	October 1998
DNA Sequence	95% of gene-containing part of human sequence finished to 99.99% accuracy	99% of gene-containing part of human sequence finished to 99.99% accuracy	April 2003
Capacity and Cost of Finished Sequence	Sequence 500 Mb/year at < \$0.25 per finished base	Sequence >1,400 Mb/year at <\$0.09 per finished base	November 2002
Human Sequence Variation	100,000 mapped human SNPs	3.7 million mapped human SNPs	February 2003
Gene Identification	Full-length human cDNAs	15,000 full-length human cDNAs	March 2003
Model Organisms	Complete genome sequences of <i>E. coli</i> , <i>S. cerevisiae</i> , <i>C. elegans</i> , <i>D. melanogaster</i>	Finished genome sequences of <i>E. coli</i> , <i>S. cerevisiae</i> , <i>C. elegans</i> , <i>D. melanogaster</i> , plus whole-genome drafts of several others, including	April 2003

		<i>C. briggsae</i> , <i>D. pseudoobscura</i> , mouse and rat	
Functional Analysis	Develop genomic-scale technologies	High-throughput oligonucleotide synthesis DNA microarrays Eukaryotic, whole-genome knockouts (yeast) Scale-up of two-hybrid system for protein-protein interaction	1994 1996 1999 2002

Key Definitions

cDNA: cDNA stands for **complementary DNA**, a synthetic type of DNA generated from messenger RNA, or mRNA, the molecule in the cell that takes information from protein-coding DNA - the genes - to the protein-making machinery and instructs it to make a specific protein. By using mRNA as a template, scientists use enzymatic reactions to convert its information back into cDNA and then clone it, creating a collection of cDNAs, or a cDNA library. These libraries are important to scientists because they consist of clones of all protein-encoding DNA, or all of the genes, in the human genome.

cM: cM stands for **centiMorgan**, a unit of genetic distance. Generally, one centiMorgan equals about 1 million base pairs.

Eukaryotic: A eukaryote is a single-celled or multicellular organism whose cells contain a distinct membrane-bound nucleus. If something is described as “eukaryotic,” it means that it has cells with membrane-bound nuclei.

Mb: Mb stands for **megabase**, a unit of length equal to 1 million base pairs and roughly equal to 1 cM.

Microarray: Microarrays are devices used in many types of large-scale genetic analysis. They can be used to study how large numbers of genes are expressed as messenger RNA in a particular tissue, and how a cell's regulatory networks control vast batteries of genes simultaneously. In microarray studies, a robot is used to precisely apply tiny droplets containing functional DNA to glass slides. Researchers then attach fluorescent labels to complementary DNA (cDNA) from the tissue they are studying. The labeled cDNA binds to its matched DNA sequence at a specific location on the slide. The slides are put into a scanning microscope that can measure the brightness of each fluorescent dot. The brightness reveals how much of a specific cDNA fragment is present, an indicator of how active a gene is.

Scientists use microarrays in many different ways. For example, microarrays can be used look at which genes in cells are actively making products under a specific set of conditions, as well as to detect and/or examine differences in gene activity between healthy and diseased cells.

Oligonucleotide: A short polymer of 10 to 70 nucleotides. A nucleotide is one of the structural components, or building blocks, of DNA and RNA. A nucleotide consists of a base chemical – either adenine (A), thymine (T), guanine (G) or cytosine (C) – plus a sugar-phosphate backbone. Oligonucleotides are often used as probes for detecting complementary DNA or RNA because they bind readily to their complements.

SNP: SNP stands for **single nucleotide polymorphism**. SNPs – pronounced “snips” – are common, but minute, variations that occur in the human genome at a frequency of one in every 300 bases. That means 10 million positions out of the 3 billion base-pair human genome have common variations. These variations can be used to track inheritance in families and susceptibility to disease, so scientists are working hard to develop a catalogue of SNPs as a tool to use in their efforts to uncover the causes of common illness like diabetes or heart disease.

STS: STS stands for **sequence tagged site**, a short DNA segment that occurs only once in a genome and whose exact location and order of bases is known. Because each is unique, STSs are helpful in chromosome placement of mapping and sequencing data from many different laboratories. STSs serve as landmarks on the physical map of a genome