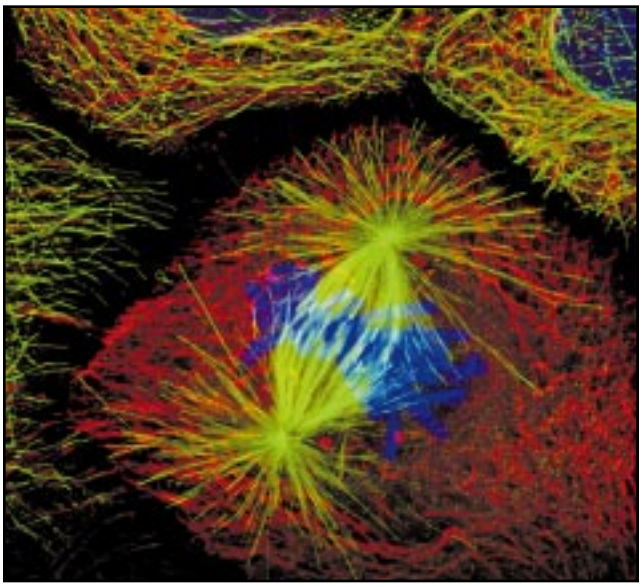
The Human Genome Project

Scott E. Walker

The focus for the front cover of *CJHP* in 2002 will be the Human Genome Project and advances in science, medicine, and technology related to this project.



Fluorescent-light microscopy shows a dividing cell in metaphase. Its duplicated chromosomes (which appear in blue) are being separated for equal distribution into 2 new cells. The spindle poles serve as centres from which microtubules (appearing in green) grow outward. The microtubules interact with special structures on the chromosomes, and the chromosomes are separated equally between the 2 new cells. Photograph by Dr Alexey Khodjakov and Dr Conly L. Rieder, Division of Molecular Medecine, Wadsworth Center, New York State Department of Health, Albany, New York.

Human chromosomes contain the DNA for thousands of individual genes. Each gene is a segment of double-stranded DNA that holds the code for making a specific molecule, usually a protein. The code is spelled out in various sequences of the 4 chemical bases in DNA: adenine (A), thymine (T), guanine (G), and cytosine (C).[1] In 1990 the Human Genome Project was initiated. Its goal was to determine the complete nucleotide sequence of DNA and to localize the estimated 50 000 to 100 000 genes within the human genome.[2] The initial 5-year research plan set specific goals for what was expected to be a 15-year project. The first and second 5-year plans outlined the intention to sequence the human genome and study genetic variation and functional analysis of the genome.[3,4] It is now anticipated that the 15-year project will be finished ahead of schedule, in 2003, just 50 years after the discovery of the double-helix structure of DNA by Watson and Crick.[5] However, it is possible that less than the totality of the genome will have been decoded by then, because some areas, such as the centromeres, are not amenable to automated assembly because of their highly repetitive sequences.

Determining the base-pair sequence of a strand of DNA is not entirely straightforward. Today's most popular method for base-pair sequencing was reported in 1977 by Sanger and others.[6] This method involves replicating each strand of DNA, adding a known base-pair sequence to each strand, and dividing the replicates into 4 groups. Nucleotide bases are then added, and base pairing begins. Each nucleotide base has a complementary base with which it can pair (A pairs with T, and G pairs with C). However, base pairing can be halted by means of a specific chain terminator — a modified nucleotide of adenine, thymine, guanine, or cytosine. Because pairing halts where the chain terminator binds, chains of various sizes are created by the chain terminator. To 1 of the original 4 groups, only an adenine chain terminator is added, such that chains in that group halt where the chain terminator of A is paired. The same process is followed for the other 3 groups and the other 3 chain terminators. All 4 groups of DNA fragments are then separated according to their size by means of electrophoresis. Starting with the smallest piece, the chain terminator is "read". If the smallest piece appears in the "A" column, then the first nucleotide in the chain is adenosine. This process continues until the entire code of the original piece is determined.

About 1 year ago, 2 groups of researchers simultaneously published their DNA sequence results.[7,8] In fact, neither group had completely sequenced all human DNA. Many small portions of the highly

continued on page 51

continued from page 4

repeated sequences between genes had not been entirely sequenced, and many will not be finished until 2003. Nevertheless, both groups agreed that humans have only about 30 000 genes, fewer than initially thought, and that genetic differences between any 2 people are small.

The sequencing of the human genome is considered the single most important project in biology and the biomedical sciences. It is anticipated that it will create an understanding of the relationship between genetic variation and the risk of developing any one of the nearly 4000 genetic diseases that affect humans. In fact, it is hoped that we can learn to prevent or treat many of these genetic diseases through knowledge gained in the Human Genome Project.

References

1. Human Genome Project information [homepage]. Oak Ridge (TN): Human Genome Program, US Department of Energy Office of Science; modified 2001 Dec 10. Available from: http://www.ornl.gov/hgmis (accessed 2002 Jan 28).
2. The Human Genome Project [homepage]. Bethesda (MD): National Human Genome Research Institute, National Institutes of Health; modified 200 Aug 23. Available from: http://www.nhgri.nih.gov/HGP/ (accessed 2002 Jan 29).
3. Collins F, Galas D. A new five-year plan for the U.S. Human Genome Program. *Science* 1993;262:43-6.
4. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L, for the members of the DOE and NIH planning groups. New goals for the U.S. Human Genome Project: 1998–2003. *Science* 1998;282:682-9.
5. Watson JD, Crick FHC. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* 1953;171:737-8.
6. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;74:5463-7.
7. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001;291:1304-51.
8. Cheung VG, Nowak N, Jang W, Kirsch IR, Zhao S, Chen XN, et al. Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature* 2001;409:953–8.

Scott Walker, MScPhm, FCSHP, is Editor of *CJHP*.