University of Washington Genome Center

http://www.genome.washington.edu

The Human Genome Project soon will need to increase rapidly the scale at which human DNA is analyzed. The ultimate goal is to determine the order of the 3 billion bases that encode all heritable information. During the 20 years since effective methods were introduced to carry out DNA sequencing by biochemical analysis of recombinant-DNA molecules, these techniques have improved dramatically. In the late 1970s, segments of DNA spanning a few thousand bases challenged the capacity of world-class sequencing laboratories. Now, a few million base pairs per year represent state-of-the-art output for a single sequencing center.

However, the Human Genome Project is directed toward completing the human sequence in 5 to 10 years, so the data must be acquired with technology available now. This goal, while clearly feasible, poses substantial organizational and technical challenges. Organizationally, genome centers must begin building data-production units capable of sustained, cost-effective operation. Technically, many incremental refinements of current technology must be introduced, particularly those that remove impediments to increasing the scale of DNA sequencing. The University of Washington (UW) Genome Center is active in both areas.

Production Sequencing

Both to gain experience in the production of high-quality, low-cost DNA sequence and to generate data of immediate biological interest, the center is sequencing several regions of human and mouse DNA at a current throughput of 2 million bases per year. This "production sequencing" has three major targets: the human leukocyte antigen (HLA) locus on human chromosome 6, the mouse locus encoding the alpha subunit of T-cell receptors, and an "anonymous" region of human chromosome 7.

The HLA locus encodes genes that must be closely matched between organ donors and organ recipients. This sequence data is expected to lead to long-term improvements in the ability to achieve good matches between unrelated organ donors and recipients.

The mouse locus that encodes components of the T-cell–receptor family is of interest for several reasons. The locus specifies a set of proteins that play a critical role in cell-mediated immune responses. It provides sequence data that will help in the design of new experimental approaches to the study of immunity in mice—one of the most important experimental animals for immunological research. In addition, the locus will provide one of the first large blocks of DNA sequence for which both human and mouse versions are known.

Human-mouse sequence comparisons provide a powerful means of identifying the most important biological features of DNA sequence because these features are often highly conserved, even between such biologically different organisms as human and mouse. Finally, sequencing an "anonymous" region of human chromosome 7, a region about which little was known previously, provides experience in carrying out large-scale sequencing under the conditions that will prevail throughout most of the Human Genome Project.

Technology for Large-Scale Sequencing

In addition to these pilot projects, the UW Genome Center is developing incremental improvements in current sequencing technology. A particular focus is on enhanced computer software to process raw data acquired with automated laboratory instruments that are used in DNA mapping and sequencing. Advanced instrumentation is commercially available for determining DNA sequence via the "four-color–fluorescence method," and this instrumentation is expected to carry

University of Washington Genome Center
Department of Medicine
Box 352145
Seattle, WA 98195

Maynard Olson
Director
206/685-7366, Fax: -7344
mvo@u.washington.edu

For more information on research projects and investigators at the University of Washington Genome Center, see abstracts in Part 2 of this report and the center's Web site (see URL above).