Coordinated Mapping and Sequencing

A simple strategy was proposed in 1996 for choosing BACs or PACs to elongate sequenced regions most efficiently [*Nature* **381**, 364–66 (1996)]. The first step is to develop a BAC end sequence database, with each entry having the BAC clone name and the sequences of its human insert ends. In toto, the source BACs should represent a 15- to 20-fold coverage of the human genome. Then for any BAC or chromosomal region sequenced, a comparison against the database will return a list of BACs (or PACs) that overlap it. Optimal choices for the next BACs (or PACs) to be sequenced can then be made, entailing minimal everlap (and therefore minimal redundancy of sequencing).

Two pilot BAC-PAC end-sequencing projects were initiated in September of 1996 to explore feasibility, optimize technologies, establish quality controls, and design the necessary informatics infrastructure. Particular benefits are anticipated for small laboratories that will not have to maintain large libraries of clones and can avoid preliminary contig mapping (see abstracts of Glen Evans; Julie Korenberg; Mark Adams, Leroy Hood, and Melvin Simon; and Pieter de Jong in Part 2 of this report).

Updated information on BAC-PAC resources can be found on the Web (*http://www.ornl.gov/meetings/bacpac/95bac.html*). [See Appendix C: Human Subjects Guidelines, p. 77 or *http://www.ornl.gov/hgmis/archive/nchgrdoe.html* for DOE-NIH guidelines on using DNA from human subjects for large-scale sequencing.]

cDNA Libraries

In 1990, DOE initiated projects to enrich the developing chromosome contig maps with markers for genes. Although the protein-encoding messenger RNAs are good representatives of their source genes, they are unstable and must be converted to complementary DNAs (cDNAs) for practical applications. These conversions are tricky, and artifacts are introduced easily. The team led by Bento Soares (University of Iowa) has optimized the steps and continues to produce cDNA libraries of the highest quality. At LLNL, individual cDNA clones are put into standard arrays and then distributed worldwide for characterization by the international IMAGE (for Integrated Molecular Analysis of Gene Expression) Consortium (see box, p. 13).

Initially supported under a DOE cDNA initiative, Craig Venter's team (now at The Institute for Genomic Research) greatly improved technologies for reading sequences from cDNA ends (expressed sequence tags, called ESTs). Together with complementary analysis software, ESTs were shown to be a valuable resource for categorizing cDNAs and providing the first clues to the functions of the genes from which they are derived. This fast EST approach has attracted millions of dollars in commercial investment. Mapping the cDNA onto a chromosome can identify the location of its corresponding gene. Many laboratories worldwide are contributing to the continuing task of mapping the estimated 70,000 to 100,000 human genes.

HAECs

All the previously described DNA clones are maintained in bacterial host cells. However, for unknown reasons, some regions of the human genome appear to be unclonable or unstable in bacteria. The team led by Jean-Michel Vos (University of North Carolina, Chapel Hill) has developed a human artificial episomal chromosome (HAEC) system based on the Epstein-Barr virus that may be useful for coverage of these especially difficult regions. In the broader biomedical community, HAECs also show promise for use in gene therapy.