GENETICS

The Human Gene Hunt Scales Up

Scientists who have been fishing for human genes with the equivalent of a bamboo pole may soon be able to use a trawler with a big net. The reason: Public and private institutions this fall are investing in ways to speed up the pace at which full-length human genes are trapped and analyzed, and they are promising to release their data quickly into public data banks.

RICK KOZAK

Free data. Venter will open TIGR's database.

The new impetus is coming mainly from Richard Klausner, director of the National Cancer Institute (NCI). On 22 November, NCI's Board of Scientific Advisers authorized Klausner to spend $19 million in extramural funding for what he's calling the "Cancer Genome Anatomy Project." The Department of Energy is contributing $1 million as well, and NCI's intramural program may spend another $7 million on the project. In the next few years, NCI may pour tens of millions of dollars into technologies for high-volume gene analysis and sequencing full-length complementary DNAs (cDNAs)—the DNA of expressed genes. NCI will start with genes from common cancer sites: the breast, colon, and prostate.

This new effort will aim to do for academic researchers what some private companies have already begun doing for themselves—using high-volume methods to identify genes expressed in tissue and sequence them completely. This is a less daunting task than what the Human Genome Project aims to do, which is to sequence the entire genome, including all the noncoding regions. But it will produce a wealth of information on genes that are active in specific tissues—information that researchers can use to hunt down disease genes, use genes in biological experiments, and look for similarities across species.

NCI's new program will build on a massive effort to identify cDNAs and "tag" them by sequencing a small fragment called an expressed sequence tag (EST). Tagging genes was pioneered by J. Craig Venter, a former National Institutes of Health researcher who founded The Institute for Genomic Research (TIGR), a nonprofit DNA sequencing center in Gaithersburg, Maryland. TIGR and other firms have built up huge private databases of ESTs. More recently, Robert Waterston's lab at Washington University in St. Louis—with funding from Merck and Co.—has built up a public collection of 450,000 ESTs called the Merck Gene Index.

The Merck Gene Index was launched in part because private firms restrict access to their data—in TIGR's case, at the insistence of its backers, Human Genome Sciences (HGS) of Rockville, Maryland, and the SmithKline Beecham Corp. of Philadelphia. But now, says Venter, the institute is ending those agreements and setting a new course: "We are completely opening up our database" of cDNA fragments. On 18 April, TIGR will simply "throw away" its system of user accounts and passwords and make its data accessible through its Web site, Venter says. TIGR is also giving out its clones. The institute will forfeit some funding from HGS and SmithKline, but expects to continue receiving HGS support for other work.

And TIGR is not just making its existing data public. It is hiring John Quackenbush of Stanford University's genome center to head a new, 10-person sequencing project focusing on full-length human cDNA sequences. Venter says all the data will be released to the National Library of Medicine's GenBank. Quackenbush says "I'm really excited" to run one of the first high-volume, public efforts to sequence complete human genes. TIGR is launching this project with a $1 million commitment of its own, says Venter, but it will seek grants, too—including from NCI's new program.

Waterston's lab also intends to begin full-length cDNA sequencing, which Waterston calls "the obvious next thing"—and something his group "always planned to do." Merck Vice President Alan Williamson says the company would like to help Waterston move to the next stage of gene identification. "There's no point in shouting about it until you've done it," says Williamson, adding, "We're going to do it."

All these projects will channel lots of high-quality genetic information into the public databases, and that makes GenBank's director, David Lipman, "superexcited." A few years ago, Lipman says, it wouldn't have been practical to try to collect full-length cDNA sequences from the entire human genome. But now, thanks to the Merck Gene Index and advancing technology, it looks feasible. Lipman says he's rushing to put NCI's initial cancer gene data—including images of source tissue—on the Web, perhaps in a week or two. And he hopes to get TIGR's data on line before long.

–Eliot Marshall

U.K. UNIVERSITIES

TOP-RANKED DEPARTMENTS, 1991–95
Ranked by citations per paper (in parentheses)

Field	1	2	3
Physics	U. Glasgow (6.85)	U. Lancaster (6.77)	U. Sussex (5.87)
Chemistry	U. Cambridge (6.35)	U. Sussex (6.17)	U. Durham (4.97)
Materials Science	U. Hull (3.90)	Q. Mary & Westfield (2.94)	U. Cambridge (2.86)
Geosciences	Open U. (5.21)	U. Oxford (4.83)	U. East Anglia (4.74)
Mathematics	U. Durham (3.38)	U. Bristol (3.21)	Imperial College (2.54)
Clinical Medicine	Imperial College (6.71)	U. Oxford (5.87)	U. Coll. London (5.30)
Immunology	U. Oxford (12.77)	U. Birmingham (11.88)	Imperial College (9.43)
Microbiology	U. Edinburgh (8.49)	U. East Anglia (7.46)	U. Cambridge (7.13)
Biology & Biochemistry	U. Dundee (10.01)	U. Oxford (8.05)	U. Coll. London (7.40)
Molecular Biology	U. Dundee (17.74)	U. Oxford (15.17)	U. Cambridge (13.41)
Neuroscience	U. Oxford (7.91)	U. Coll. London (7.86)	U. Cambridge (7.10)

SOURCE: ISI

The Universities of Oxford and Cambridge still have gold-plated reputations in British science, but they do not have a monopoly on the hot papers, according to figures to be published in January by the Philadelphia-based Institute for Scientific Information. ISI analyzed statistics on published papers and their citations for more than 100 U.K. universities over the period 1991 to '95. The result is a table of the universities' "citation impact"—total number of citations divided by the number of papers published—in 21 fields of research.* This analysis can let small departments shine, because if a department publishes few papers, one or two highly cited ones can make a big impact.

Oxford or Cambridge generated the most papers in 13 of the 21 fields surveyed, and Oxford achieved the highest impact in four. But London's Imperial College and the Universities of Glasgow, Durham, and Dundee were not far behind, all topping two fields, while Cambridge managed only one. "This does come as a surprise," says Edward Corrigan, head of Durham University's Department of Mathematical Science, which, with 36 research staff, was tops in math. "But it is possible to pick isolated papers with several hundred citations." David Finnegan, head of the University of Edinburgh's Institute of Cell and Molecular Biology, which came on top in microbiology, was bemused. "I'd say we were stronger in genetics. ... These lists are a bit hard to fathom." **–D.C.**

* For the complete list, see http://www.sciencemag.org/science/scripts/display/full/274/5292/1456b.html