GENOMICS

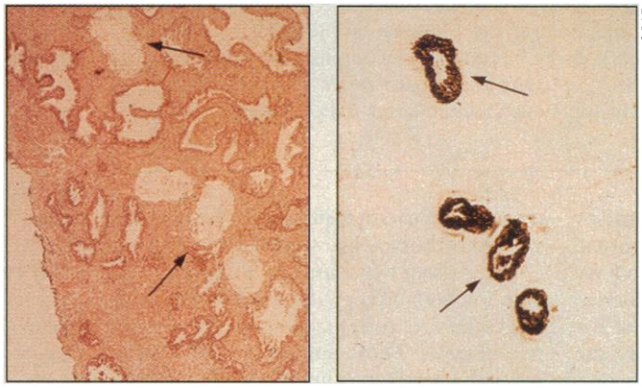
A Catalog of Cancer Genes At the Click of a Mouse

Webster Cavenee can't wait to start doing cancer research on the Web. Like thousands of other biologists, Cavenee, a cancer geneticist at the Ludwig Institute for Cancer Research in San Diego, has already come to depend on a burgeoning array of online databases to help search for new genes and understand what they do. But sometime next month, Cavenee and his colleagues will be able to take their computer queries to a new level, thanks to the Cancer Genome Anatomy Project (CGAP)—an ambitious effort that aims at nothing less than a complete catalog of all the genes expressed in cancer cells.

Starting with five big killers—breast, colon, lung, prostate, and ovarian cancers—CGAP will classify tumor genes by the type of cancer cell they came from and the degree of the cell's malignancy. "We want to achieve a comprehensive molecular characterization of cancer and precancerous cells," says National Cancer Institute (NCI) molecular biologist Robert Strausberg, who is coordinating the CGAP project. With just a click of the mouse, researchers should be able to determine how gene expression changes as a cancer progresses and ultimately begin to understand how tumors arise in the first place—all for no charge. "CGAP will be a national resource that all can tap into via the World Wide Web," says NCI director Richard Klausner. "It's a wonderful, farsighted thing to do," says Cavenee. "It will accelerate cancer research beyond anything [else] they could have done."

The project is the latest example of how a marriage between computer science and genetics—creating a brand-new field called "genomics"—is transforming whole areas of biology. It's also an example of fleet-footedness on the part of NCI. Klausner began talking about the project in mid-1996, even before some of the key technologies were in place. But they made their debut before the end of the year, and about the same time, Klausner got the go-ahead to fund CGAP (*Science*, 29 November 1996, p. 1456). NCI is putting $20 million into the effort this year, and it's also being supported by contributions from the National Library of Medicine's National Center for Biotechnology Information (NCBI), the Department of Energy (DOE), and several pharmaceutical companies.

And all that is just for starters. CGAP funds will also support the development of technologies for rapid analysis of gene activity patterns in the thousands of tumors likely to be seen in a hospital pathology lab. The hope is that, eventually, the tumor gene index and these new technologies will enable physicians to base a patient's diagnosis, prognosis, and treatment on the status of a particular tumor's genes and proteins rather than on its appearance under the microscope. "The notion that one can identify all the genes altered in a given cancer is very exciting," comments cancer geneticist Louise Strong at the M. D. Anderson Cancer Center in Houston. "It should [provide] really good information [for] classifying tumors and identifying targets for therapy."



CGAP

Cell removal. Laser-based dissection lifts small groups of cells *(right)* from a prostate tumor, leaving the rest of the tumor intact *(left)*.

Despite all this enthusiasm, however, this vision could take years to realize. Although various companies and academic researchers have begun developing methods for wholesale analysis of gene expression patterns, the methods have not yet been shown to work on anything near the large scale envisioned. And plans to expand the tumor gene index to include other cancers are still vague. A similar effort currently getting under way in Europe may, however, help out here (see sidebar). In time, NCI expects its Web site to be linked to the one the Europeans are planning.

Key techniques. The expectations are bold for a program that's still taking shape. But, as Klausner reported last month at the annual meeting of the American Association for Cancer Research, a team led by NCI pathologist Lance Liotta has already demonstrated the feasibility of constructing a gene index for one tumor: prostate cancer. The success of that work depended on two new techniques: one for pulling specific cells out of tumors, which are a heterogeneous mix of normal cells and cells in various stages of malignancy, and the other for analyzing the very tiny amounts of RNA present in the isolated groups of cells.

The melange of cell types in a typical tumor has long been a problem for researchers analyzing gene expression in cancer cells. They can't get an accurate picture by looking at RNAs extracted from a whole tumor, because it will be a mix of molecules from all the different cell types. And standard methods for dissecting out individual cells are not only tedious but also unreliable except in the most experienced hands, says Ramon Parson, a cancer geneticist at Columbia University College of Physicians and Surgeons in New York City. But last year, Liotta, Michael Emmert-Buck, also of NCI, and their colleagues developed a technique called laser capture microdissection that enables researchers to pick out just the cells they want to analyze in about one-tenth the time that standard microdissection requires (*Science*, 8 November 1996, p. 998).

The procedure starts with a thin slice of tumor tissue, placed on a glass microscope slide and covered with a transparent cap from a tiny vial, the underside of which is lined with a thin layer of plastic. A researcher simply scans the sample through a microscope to find a uniform group of cells and, when they are in the scope's cross hairs, squeezes a trigger, zapping that spot with a weak laser. The laser heats the plastic so it becomes sticky and adheres to the cells just underneath. When the cap is lifted from the sample, it pulls off the targeted cells with it. The rest of the sample remains intact, so by moving to a different section, a researcher can obtain cells in various stages of tumor development, all from one piece of tissue.

Emmert-Buck used this technique to collect four sets of cells—normal prostate tissue, cells just starting to transform, fully transformed cells, and invasive cancer cells—from a single frozen prostate tumor. Each set contained about 5000 cells.

The investigators' next challenge was to show that they could use these cells to generate libraries of complementary DNA—DNA from expressed genes—that accurately reflect the full complement of the cells' active genes. Making such libraries requires copying messenger RNA sequences into cDNAs that can then be cloned separately in bacteria. A big problem is that when there is too little RNA, the scarcest sequences can get lost. As a result, sequences of the more common messenger RNAs are overrepresented and others are underrepresented in the library.

Europe's Cancer Genome Anatomy Project

Not long after the Cancer Genome Anatomy Project (CGAP) goes online next month (see main text), the plan is to have it hook up with a similar project now taking shape in Europe. Eleven European academic and clinical laboratories have teamed up to create the Cancer Gene Expression Program, which is also intended to study gene expression in cancer "in a very comprehensive manner," says Charles Auffray, a cancer geneticist at the National Center for Scientific Research in Villejuif, France. Auffray and colleagues from Germany, Sweden, and the Netherlands will focus first on prostate, colorectal, breast, lung, skin, kidney, and pediatric brain cancers.

Once it raises the funding needed, the European team's first goal is to make libraries of full-length complementary DNAs (the DNA from active genes)—as opposed to shorter DNA fragments called expressed sequence tags, or ESTs—from these cancers. They will then select the subsets of sequences that appear to be specific to each of the tumor types. Each subset will be used to screen large numbers of that type of tumor to verify that indeed those genes are either aberrantly expressed (as in the case of oncogenes) or not expressed (as in tumor-suppressor genes) in that cancer, Auffray explains. He expects to make this information publicly available as quickly as possible.

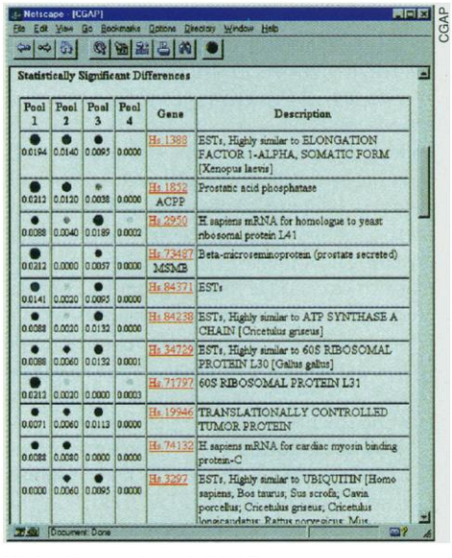
The project is still in the planning stages, but already Auffray has been talking with CGAP folks about collaborating and linking the European project's Web site to CGAP's. At first glance, the projects seem to be trying to accomplish the same goals. But both are needed, says Auffray, because "there are so many genes, so many different cancer types, and so many stages." He hopes even more teams will become involved in cataloging the cancer genes. –E.P.

Another method developed last year by Emmert-Buck, NCI molecular biologist David Krizman, and their colleagues seems to have solved that problem. It combines an efficient cloning technique with a few cycles of the polymerase chain reaction to increase the number of copies of each bit of sequence. When Krizman used this procedure on the cells Emmert-Buck obtained by laser capture microdissection, he found, he says, that "the diversity was really quite good [and] all the genes are kept in a good balance." Genes common in the prostate, such as the one encoding prostate-specific antigen, were well represented in the library, as would be expected. Moreover, there appeared to be some 360 new genes not previously described.

Creating a tumor index. Emmert-Buck, Krizman, and their colleagues will spend the next few months making similar cDNA libraries for normal, precancerous, and cancerous cells from lung, colon, breast, and ovarian tumors, as well as from more prostate cancers. Robert Waterston's team at Washington University in St. Louis will then spend 6 months sequencing several thousand bits of cDNA called expressed sequence tags, or ESTs—unique sequences useful for identifying active genes—from each of these 45 libraries. After that, the task will fall to a yet-to-be-named NCI grantee.

As fast as these EST sequences come in, NCBI researchers will post them in GenBank where they can be accessed from the CGAP Web site along with data about both the tumors and the libraries. This information is not currently available for ESTs in GenBank's database. "We want to present the ESTs in the context of the biology," says NCI's Strausberg.

A program developed at NCBI will enable any Web visitor to look at differences in the patterns of gene expression between any two libraries. Those differences could point researchers to genes important to the progression of cancer, or that could serve as markers of disease. "The hope is that what we'll find are patterns of gene expression that will define much better the subpopulations of cancers," says NCI's Kenneth Katz.



Web-site preview. A CGAP program computes the relative expression of genes from different stages of a tumor, displaying differences by dot size and shading.

That information might help clinicians identify diagnostic markers that can distinguish between normal and cancerous cells or aid in determining a patient's prognosis. Being able to predict a particular tumor's metastatic potential "would be of tremendous benefit," says molecular biologist Claire Fraser of The Institute for Genomic Research in Rockville, Maryland. An oncologist could then tailor the follow-up radiation or chemotherapeutic treatment according to how likely the tumor was to have spread.

In addition to building EST libraries, CGAP investigators will start constructing libraries with longer stretches of cDNA that contain more of an active gene's coding region. These will come from several dozen different types of cancers. Because larger amounts of starting material are needed to generate libraries with longer cDNAs, the RNA will be extracted from bulk rather than microdissected samples. A researcher finding an EST that appears to be unique to, say, precancerous tissue can then look for a match among the longer sequences in these libraries and, upon finding one, possibly isolate the gene faster.

Toward these ends, the NCI is putting $4 million this year into the tumor gene index. Several companies are kicking in additional support for the generation of ESTs. And DOE is allocating $1 million to help set up the bacterial clones in which the ESTs and longer clones are maintained. The IMAGE consortium at Lawrence Livermore National Laboratory in California will make these clones available to any researcher who wants them.

Another $6 million of NCI funds is slated to be distributed to grantees to develop new technologies for analyzing many genes at a time or to improve upon techniques for generating long, potentially full-length cDNAs. Several companies are already working on either DNA chips or microarrays, which arrange many thousands of bits of DNA in a small space. Once these companies have automated the process and sufficiently expanded the numbers of bits of DNA squeezed into an array or onto a chip, these devices can be used for rapid screening of large numbers of cDNA samples to detect expressed genes, Strausberg says. Finally, NCI will spend about $10 million on grants for researchers to come up with ways to put these basic data to use in the clinic.

All this should quickly produce a rich Web site. "The potential for data mining is just great," Emmert-Buck notes. And researchers not involved in CGAP couldn't agree more. "It will generate new directions for investigators," says cancer geneticist Kenneth Kinzler of Johns Hopkins University School of Medicine. All that potential has Cavenee's research team "salivating," he says. "We're checking the Internet daily to see if the [CGAP] home page has come up."

–Elizabeth Pennisi