More SNPs on the Way

Late last year, the National Cancer Institute (NCI) launched a project to find genome markers called single-nucleotide polymorphisms, or SNPs, to use in tracking down the hundreds of genes thought to affect cancer risk. NCI has already put about $1 million into the project, called the Genetic Annotation Initiative (GAI), which began generating SNPs in the spring. Researchers running the initiative are hoping that their approach will avoid many of the problems in using SNPs discussed at a recent conference in Skokloster, Sweden (see main text).

NCI is taking what NCI geneticist Ken Buetow, who oversees the GAI project, calls a "gene-based" approach. Instead of creating a genomewide map of anonymous SNPs, Buetow says, NCI will look for SNPs in the coding regions, and in the sequences at both ends, of several thousand genes suspected of contributing to cancer susceptibility or resistance. Besides the 100-plus known cancer-promoting oncogenes and the three dozen or so tumor suppressor genes, the pool will include DNA repair genes, genes that drive the cell division cycle, and genes involved in drug metabolism, immune responses, embryonic development, and cell migration and metastasis. Genes from the NCI's huge Cancer Genome Anatomy Project, which aims for a complete genetic profile of cancer cells (*Science*, 16 May 1997, p. 1023), will also be included as they're identified.

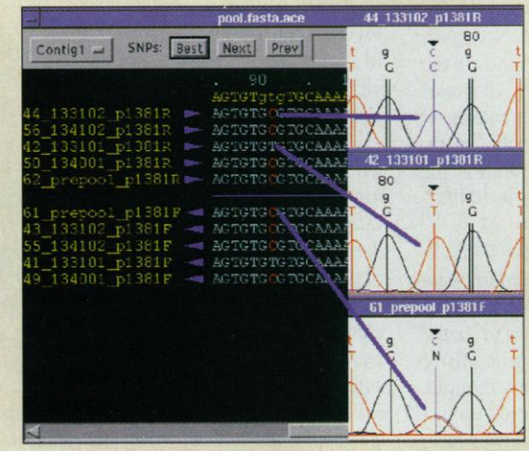
Buetow expects the average gene to yield three to five SNPs, a marker density that makes it much more likely that at least one will be close enough to any cancer mutation to be inherited with it as a block—a phenomenon called linkage disequilibrium. That doesn't ensure researchers won't miss the mutation when screening cancer patients—one of the researchers describing SNP problems at Skokloster had just such an experience with the sickle cell gene—but it should help. "We are less dependent on linkage disequilibrium relationships existing over long distances," says Buetow. "We're going to be right inside the genes." The data generated by GAI will also help determine how common the problems reported at the meeting are.

Once identified, the SNPs will be posted on a database of the National Center for Biotechnology Information, where researchers can access them and design and conduct "association studies" to see if the SNP patterns of cancer patients are different from those of controls. The hoped-for result: hundreds of new cancer genes. Cancer researchers welcome the new initiative. "Given the present technology, it seems to be the obvious next step," says Sofia Merajver, a breast cancer researcher at the University of Michigan, Ann Arbor.

Several issues are still up in the air, however. NCI hasn't decided which populations to screen for SNPs. Right now it's using the DNA of four people from the largely Caucasian families collected at the Centre d'Études du Polymorphisme Humain in France. The GAI wants more diversity, but no one agrees on what that means. "There is concern about stigmatization of populations and concern about what is a representative population," says Buetow. "There are going to be dramatic differences [in SNP frequency] based on geography."

Also under debate is the question of how deep to dig for cancer SNPs. Some would be satisfied with the common ones, in which case screening as few as eight individuals should yield the vast majority. But others argue that the newer, rarer SNPs are also needed, because they're more often in linkage disequilibrium with cancer mutations and thus more likely to show up in cancer association studies.

But the biggest question mark is what technology will be used to discover SNPs and then to detect or "score" them in cancer patients. "Not only has this not been done on a mass scale, but new technologies are being developed so fast, it's hard to know what to do," says the NCI's Mike Dean. To begin, Dean is using a high-performance liquid chromatography mutation-detection method developed by Peter Oefner of Stanford University. Buetow is doing conventional gel-based sequencing, which would be tedious and expensive for large-scale studies.



Gene guide? SNPs, such as the cytosine (C) to thymine (T) change shown here, may point to cancer genes.

One technology now in high demand is the DNA "chip," which can quickly identify SNPs across long stretches of DNA. Affymetrix, a Santa Clara, California, biotech company, has developed such chips, which researchers at the Whitehead Institute for Biomedical Research at the Massachusetts Institute of Technology and Affymetrix are using to do SNP prospecting (*Science*, 15 May, p. 1077). The National Institutes of Health is now negotiating with Affymetrix for a license, and both parties are optimistic. "We would be very happy to collaborate with the NIH in the area of SNP discovery," says Robert Lipshutz, Affymetrix's vice president of corporate development.

Whatever the outcome, Buetow is optimistic about finding methods that will make all kinds of cancer gene discovery projects easy. "We hope to push the technology to enable investigators to do any kind of study they want to do," he says. **–KEN GARBER**

Ken Garber is a science and health writer in Ann Arbor, Michigan.

veal the mutant gene. By analyzing DNA samples from 500 people randomly selected from around the world, Harding and her colleagues found that the β-globin gene has dozens of SNPs located in and around its coding sequences.

One of these SNPs turned out to be the sickle cell mutation itself. But when Harding looked at the frequencies of individual SNPs in the 500 samples, and also at inherited SNP patterns called haplotypes, searching for some sign that a particular SNP or haplotype was different, she found nothing that pointed to the sickle cell mutation. With SNP data alone, Harding concluded, "there will be nowhere near enough information to find something unusual and say 'there's a disease gene.'"

She predicts that, in addition to relying on SNPs, researchers will need to know about the patterns of disease and the history of the people being studied. "There has been this naïve idea that once you've gotten to the gene, you'll be able to decide which is the [pertinent] mutation," she adds. "But this is going to be very hard." Others concur. "You can't have just SNPs on their own," says Nigel Spurr, a geneticist with SmithKline Beecham in Harlow, United Kingdom. "You must have [other information and technology] to go with it."

For statistical geneticist Joseph Terwilliger of Columbia University in New York City, Harding's and Clark's experiences with SNPs are indicative of the underappreciated complexity of the genome and of the pitfalls of thinking SNPs will easily