Whose Genome Is It, Anyway?

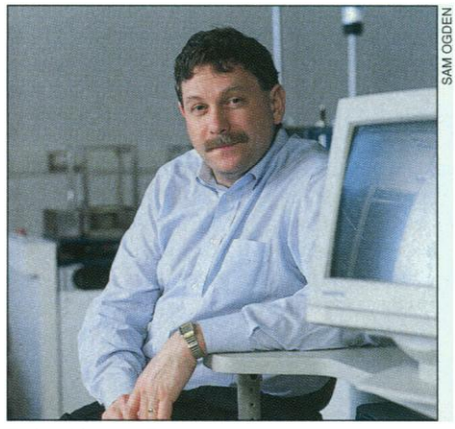
The genome program has issued guidelines to ensure that sequencing is done on DNA from diverse sources who have given informed consent and are anonymous. Most current sources don't meet those criteria

It may be the first question every nonexpert asks on learning about the Human Genome Project: Whose genome are we studying, anyway? It sounds naïve, says one government scientist—so naïve, in fact, that "we chuckle as we explain that we aren't sequencing anyone's genome in particular; we're sequencing a representative genome" made up of a mosaic of DNA from a variety of anonymous sources. And Bruce Birren, a clone-maker now at the Massachusetts Institute of Technology's (MIT's) Whitehead Center for Genome Research, says: "We spent many years pooh-poohing the question" of whose genome would be stored in the database. But now that labs have begun working on large stretches of human DNA—aiming to identify all 3 billion base pairs in the genetic code—the question no longer seems so laughable. To the distress of program managers in Bethesda, Maryland, the initial sources of DNA are not as diverse or as anonymous as they had assumed.

This past summer, six pilot projects in the United States plus the Sanger Center in Cambridge, England, geared up for large-scale sequencing (*Science*, 12 April, p. 188). As they did, the chiefs of the genome project and their scientific advisers began looking into DNA sources—and got a surprise. The DNA of major clone libraries now being fed into high-volume robotized labs appears to come primarily from a limited group of donors: three men and one woman. Worse, of these four subjects, two may not have been informed in advance that their genome would become part of a public database, and one of those two is dead, with no relatives to be found. And even worse, not all the donors appear to be anonymous: Sequencers in U.S. labs say they know the names of two of them, because they are colleagues. According to several leading researchers, two labs making clone libraries for sequencing—the Roswell Park Cancer Institute in Buffalo, New York, and the California Institute of Technology (Caltech) in Pasadena—used blood and sperm from staffers as a source of DNA.

The growing realization that these DNA sources may raise ethical and appearance problems created a headache for genome leaders this summer. And it forced the National Center for Human Genome Research (NCHGR) at the U.S. National Institutes of Health (NIH)—co-sponsor with the U.S. Department of Energy (DOE) of the U.S. genome program—to step in. NCHGR's chief, Francis Collins, and his advisory council conducted a confidential debate among ethicists and scientists in a closed session of NCHGR's council last May and through e-mail messages over the summer. And Collins met with the director of DOE's genome program, Aristides Patrinos, to iron out a new approach. The product of these talks, a statement of ethical guidelines dated 17 August, has been posted on a DOE Web site* and was released at NCHGR's advisory council meeting on 16 September.

The statement, signed by Collins and Patrinos, asks all U.S. genome sequencing centers to switch as soon as possible from existing DNA clone libraries to new ones now being made. These libraries will have better procedures for informed consent to cope with the ethical issues. To defuse some of the political concerns, they may also include more diverse sources of DNA, including more DNA from women. Last week, NCHGR handed out supplemental grants to Roswell Park and Caltech so that they can quickly increase the information in libraries and improve methods of selecting donors, enabling them to meet the new standards.



SAM OGDEN

Raised issue. Eric Lander asked at a genome meeting whether consent had been obtained.

In the meantime, says NCHGR's assistant director, Mark Guyer, NCHGR will permit sequencers to continue using the current libraries, provided the "people who made those libraries have gotten appropriate informed consent" from the donors. While it may not be possible to give former donors full anonymity, Collins and Patrinos have decided that it is more important to allow the research to go forward. That's a relief for some researchers, who were afraid that "all sequencing would be shut down in 1997," as David Smoller, founder of Genome Systems Inc. in St. Louis, puts it. But even with an interim waiver, Smoller says, he will have to scrap a $100,000 library he created and rebuild it to meet the new rules.

In Britain, the director of the largest sequencing project outside the United States, John Sulston of the Sanger Center, says, "We're taking a more relaxed view on this side of the Atlantic." His center has been using many of the same DNA sources as U.S. labs, but Sulston says that the sponsor of U.K. genome sequencing, the Wellcome Trust, feels that there is no reason to panic. His "bottom-line opinion, which is pretty much shared here, is: Let's not get so heated about it. We do not want to get in a legal mess ... but scientifically it doesn't matter at all." He anticipates using the new libraries as they become available.

Elitist DNA?

Researchers who spoke to *Science* about DNA sources for the genome project said they couldn't pinpoint when the DNA donor controversy erupted. But one scientist at the center of the discussions—Pieter De Jong, the principal investigator for a DNA clone library at Roswell Park—says the problem faced him squarely about 7 months ago. De Jong was attending a private meeting of genome researchers at a Bermuda resort last February, sponsored by the Wellcome Trust, to debate the ground rules for large-scale sequencing (*Science*, 26 April, p. 477). "I remember Eric Lander [director of the MIT genome center] saying, 'I hope it's all OK,' that everything has been done according to proper ethical guidelines. And I think Francis Collins pursued it" after that.

Guyer, who is implementing the new guidelines, says that he learned of the concerns after the six sequencing-center grants were announced this year. "We, along with everybody else, had been assuming that the DNA would come from multiple sources," Guyer says. But when the grant applications came in, NCHGR realized that "all of a sudden" people were focusing on just a couple of libraries: bacterial artificial chromosomes (BACs) created in Melvin Simon's laboratory at Caltech and a set of P1 artificial chromosomes (PACs) created by De Jong. Guyer says it became clear that "the assumptions [about diverse sources] we had been making all along were not going to be played out; Francis [Collins] got concerned."

* http://www.ornl.gov/TechResources/Human_Genome/archive/nchgrdoe.html

To ensure there will be no further confusion, the NIH-DOE statement lays out some standards that all new clone libraries must meet. It rules out the use of DNA donated by genome researchers themselves because this might appear "elitist." The statement also says it would be a mistake to take DNA from junior lab staffers because they might "feel pressure to donate," and it would be difficult to hide their identity. Furthermore, because "women have historically been underrepresented in research," the statement says, "concerns might arise if males (sperm DNA) were used exclusively as the source of DNA for large-scale sequencing." Sensitive to issues of research independence, the genome chiefs say that "the choice of donors will not be dictated to investigators." They concede that there is no scientific reason for insisting on diverse DNA sources or female donors (because DNA from a man provides all the requisite information, although in reduced quantity for the X chromosome). But "it is expected," the authors write, that clone libraries will be made from female as well as male sources.

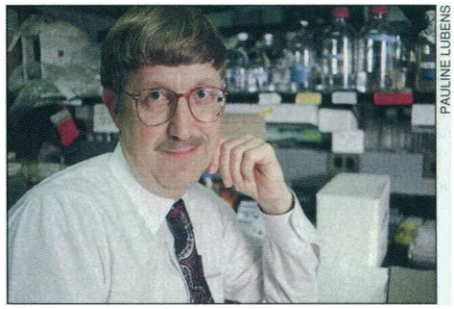
The new guidelines also require researchers to take special steps to ensure that donors remain anonymous and give informed consent to the use of their DNA. They ask that scientists create "as many disconnects between the identity of the donors and public available information" as possible. No demographic or personal information should be linked to the donor's DNA, the rules say, nor should donors have any way of identifying their own DNA in a database. The guidelines say that a diverse panel of clone libraries "must be made available" to research centers "to increase the likelihood that the first human DNA sequence will be an amalgam of regions sequenced from different sources." In addition, they require that plans for recruiting donors be approved by a local institutional review board "before work is initiated."

Principles and perceptions

Some researchers are convinced that the fuss over DNA for the genome project has more to do with appearances than substance. De Jong, for example, views it as a "perceptional issue." And Simon sees the debate over male versus female DNA as nonscientific, although he agrees that it will be necessary to obtain more rigorous informed consent agreements than in the past. He says he has now obtained them retroactively for his existing clone libraries. Gene sequencer Hamilton Smith of Johns Hopkins University thinks that much of the discussion about the need for a diverse pool of DNA donors is "a distraction ... a lot to do about nothing ... hokum."

Even Lander believes that the backstage debate, which he says has caused a slight delay in the program, may have more to do with public relations than ethics. He agrees, however, that "heightened scrutiny" and "a higher level of consent" from the donors are necessary for large-scale sequencing, because the donor's DNA will become part of a public, permanent reference library.

While Lander feels that the genome research community has essentially agreed to these guidelines, others take issue with some of the technical details. For example, some researchers—including Smith and Leroy Hood of the University of Washington—say it makes sense to use male donors. Unlike blood cells, sperm can be obtained in large quantity without resorting to laboratory production methods that risk significant changes in DNA—as may occur when blood cells are "immortalized" by inducing them to reproduce indefinitely. Sperm includes all 24 chromosomes, including the Y. Finally, as a germline cell, sperm does not include any developmental rearrangements of DNA, as occur in the immunoglobulin genes in somatic cells.


PAULINE LUBENS

Issued guidelines. Francis Collins wants DNA to come from diverse—and anonymous—sources.

Lander says that he also prefers to use germline DNA (sperm, because eggs are not donated for DNA studies). But germline DNA isn't essential. "If we sequence [germline DNA], we're going to have to sequence [somatic-cell DNA] anyway to see if it's different," Lander argues. He is prepared to begin with DNA from germline or somatic cells, male or female: "Either way is all right."

What does matter, however, according to Lander and other genome program leaders, is that each library of DNA clones be distinct, well characterized, and highly redundant. Although there have been some suggestions that the identity of donors be masked by pooling DNA donations from many individuals into a melange, most sequencers oppose this idea. For example, Robert Waterston, leader of the sequencing project at Washington University in St. Louis, says: "When working in one region or chromosome, it is important ... that clones come from one or a very few individuals." The reason: Natural variation produces genetic differences between individuals; the more of these mutations a sequencer must sort through, the harder it will be to align sequence data accurately. If there are a large number of mutations in the DNA being sequenced, it could become very difficult to determine whether any particular anomaly is a real mutation or just a sequencing error.

Arguments such as this seem to have won the day. Guyer says, "Our view is that the initial reference sequence" for the human genome project "should be a mosaic derived from a number of different individuals." But researchers need not pool the DNA, because "we're hearing that that's where you would run into trouble." Instead, says Guyer, "we're trying to make available a reasonable number of libraries, each of which is derived from a single individual," so that researchers will be working on several sources in parallel.

One big question remains unanswered, however: When will these new libraries be ready? NCHGR awarded a new grant on 20 September to Hiroaki Shizuya of Simon's lab—which has been supported by DOE to this point—to expand and improve the existing library of BAC clones. And about a week earlier, NCHGR gave a 2-year grant to De Jong to convert from PAC to BAC technology and produce yet another set of clones. By De Jong's estimate, the initial product of this venture, vetted by ethics reviewers, will be available for use by early summer next year. The completed library will be ready in 2 years—just in time, everyone hopes, for the anticipated ramp-up to full-speed sequencing of the human genome.

–Eliot Marshall

CLONE LIBRARIES FOR HUMAN GENOME SEQUENCING

Investigator	Institution	Source	Vector
Melvin Simon	Caltech	sperm	BAC
Pieter De Jong	Roswell Park	male blood	PAC
David Smoller	Genome Sys. Inc.	female blood	BAC
Giles Thomas	CEPH/Généthon	female blood	BAC

Developed at Caltech with Department of Energy funding, a vector called the bacterial artificial chromosome (BAC) is all the rage in labs planning for large-scale human genome sequencing. In contrast to yeast-based technology used for mapping, BACs are stable and predictable. They delete sequences less often and are far less likely to rearrange DNA in "chimeric" scrambles. And they suit the biology of robotic sequencing. De Jong's P1 artificial chromosomes (PACs) are a variation on the BAC theme. Thomas's library has not yet been widely sampled, according to sequencers. Caltech and Roswell Park are getting funded this fall to create newer, larger BAC libraries.