

ANALYSIS

Genome II: The next frontier

NIH (Bethesda, MD) director Francis Collins has announced a new five-year plan for the next phase of the Human Genome Project (HGP), which some call Genome II. As well as focusing on functional genomics, the plan accelerates current HGP sequencing efforts to match industry's 2001 target date for a working or shotgun human genome sequence, with completion of its full sequence by 2003. Revised goals have been prompted in part by newer, more efficient sequencing that has allowed the HGP to do more with its money each year, and competition from Celera (Rockville, MD) and Incyte Pharmaceuticals (Palo Alto, CA). Collins claims the finished product will be more complete, accurate, and authoritative than that of industry rivals.

"Our new goals are ambitious, risky, and audacious," admits Collins, who says that Craig Venter's May 1998 leap into the sequencing race (*Nat. Biotechnol.* 16:497, 1998) had "stirred the pot," and helped spur on Collins' international team. "Not to do so would mean missing an opportunity to set the bar very high," Collins notes. No doubt Incyte's mid-August declaration that it, too, was entering the race (*Nat. Biotechnol.* 16:895, 1998) only added to their motivation.

Genome II was begun some time ago with the yeast genome, and will extend to all other genomes being studied in the HGP as genome sequences reach completion—*C. elegans* this year, *Drosophila* in 2002, and the mouse by 2008. As well as understanding human genetic variation, gene function, protein-protein interactions, and resequencing for variation, evolutionary and population biology, and ethical, legal, and social implications will be studied, says Collins.

One of the goals of Genome II is the creation of a genome-wide map of 100,000 single-nucleotide polymorphisms (SNPs). This puts the public genome program in competition with another leading genomics company, Genset (Paris), which plans to produce a 60,000 "bi-allelic" SNP map. The company has already completed a third to a half of the task. "SNPs will prove to be the next battleground in genomics, says Rochelle Long, chief of pharmacological and physiological sciences at the National Institute of General Medical Sciences (Bethesda, MD).

According to the original plan, the NIH and its US collaborators are responsible for 60% of the genome's sequence, the Department of Energy for 10%, and the Sanger Centre (Hinxton, UK) and the Wellcome Trust (London) for 30%.

At the start of the project in 1990, the NIH estimated that the US part of the HGP 15-



Collins: the HGP will create a genome-wide map of SNPs.

year program would cost \$3 billion dollars, or \$200 million per year. With yearly allocations from Congress steadily rising from \$59.5 million in 1990 to \$188.9 million in 1997 and an estimated \$218.4 in 1998, about \$1.8 billion has been spent to date. However, while spending has risen, costs have decreased—most notably sequencing costs have plummeted as technology and efficiency have improved—allowing the project to do more each year with its money. It is expected that the yearly increases in federal funding will continue, but that no special increase will be required to achieve the new goals of Genome II.

UK allocations are made in lump sums for a number of years: to date, the Wellcome Trust has spent \$120 million on Genome Campus (Hinxton) and will spend an additional \$350 million funding the Sanger Centre's sequencing efforts. Although Sanger has produced more human DNA sequence than any of the NIH's extramural academic collaborators, it recently acquired Amersham Pharmacia Biotech/Molecular Dynamics' MegaBASE system to further increase its sequencing capacity.

As well as completing the entire sequence by 2003, the HGP plans to have a working draft (30%) by 2001—the same target date Celera has set for completion of its own sequencing program (*Nat. Biotechnol.* 16:610, 1998).

The whole genome shotgun approach adopted by Celera involves randomly sequencing fragments from a genome that has been broken into short stretches. Although it is quick, when the data are assembled into a sequence it contains gaps.

In contrast, the HGP has a map-based approach and systematically sequences BAC-ordered libraries. However, the new goal of a working draft by 2001 will be attained using a shotgun style approach. The gaps will be filled during the more labor-intensive sec-

ond phase. Already partly done, the working draft sequence is accumulating in public databases at about twice the rate of the finished sequence.

For Incyte and Celera, which will use the new Perkin Elmer 3700 sequencer when it is released early next year, speed is key to quickly achieving commercialization. Celera plans to patent SNPs, and Incyte will continue providing sequence databases to the life science industry. "We are not interested in crossing all the t's," says Incyte president and chief scientific officer, Randy Scott, referring to the HGP's sequencing, which he and others in industry view as an academic exercise.

For immediate commercial purposes, a shotgun sequence may be adequate, says Michael Morgan, chief executive of the Wellcome Trust Genome Campus, but not for the future. "If we are interested in establishing genomics as the basis of the biological sciences for the next millennium, it will be necessary to understand why we carry around so much DNA—so-called junk DNA—for which we can find no purpose now," he says.

For Collins and Morgan, speed of the sequencing effort is necessary but not sufficient to produce a high-quality sequence of the human genome. "Two years ago, we considered doing a shotgun approach like Venter is now doing, but rejected it, because we felt there was no guarantee it would result in a complete, continuous, and accurate product," Morgan says.

Bickering between the two groups continues. Morgan notes that Venter's Institute for Genome Research's (TIGR; Rockville, MD) sequence of an isolate of tuberculosis using the shotgun method did not approach Sanger's in detail and accuracy. At the TIGR 10th International Genome Sequencing and Analysis Conference (September 17–20, 1998, Miami Beach, FL), Venter claimed TIGR's sequence of *C. elegans*—the first animal genome to be sequenced, due for completion by Celera in the next couple of years—will prove that shotgun sequencing is sufficiently accurate. However, Morgan points out that Sanger's sequencing of *C. elegans* will be done sooner—by the end of this year. In addition, Celera's Gene Myers and TIGR's Mark Adams justified Celera's shotgun approach with a series of complex mathematical algorithms.

Although Celera and Incyte argue against the thoroughness of HGP's approach, they are benefiting from it—HGP sequences are posted on a public database daily. By contrast, there is significant lag-time before industry discoveries are posted, in an effort by industry to achieve a commercial advantage.

Vicki Brower

1998 Nature America Inc. • <http://biotech.nature.com>