



The Human Genome Project Aims for 2003

Laurie Goodman

Genome Res. 1998 8: 997-999

Access the most recent version at doi:[10.1101/gr.8.10.997](https://doi.org/10.1101/gr.8.10.997)

References

This article cites 2 articles, 1 of which can be accessed free at:
<http://genome.cshlp.org/content/8/10/997.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

The Human Genome Project Aims for 2003

Laurie Goodman¹

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11743 USA

Bold is the operative word for the goals in the new five-year plan for the U.S. Human Genome Project as presented at the National Human Genome Research Institute (NHGRI) advisory council meeting on September 14, 1998. (See Box 1 for an overview of the 5-year plan. The complete plan is published in the October 23, 1998 issue of *Science*.) The plan itself includes such words as “ambitious” and even “audacious,” reflecting the committee’s own perception that the agenda is one that will require extreme diligence and effort to achieve. The most notable point is the call for the completion of the human genome sequence in the year 2003—a full 2 years earlier than originally planned and corresponding to the 50th anniversary of the discovery of the double helix structure of DNA (Watson and Crick 1953).

The proposal was approved by the NHGRI council and by the U.S. Department of Energy (DOE); human genome project research in the U.S. is sponsored jointly by the NHGRI at the National Institutes of Health (NIH) and the Office of Biological Environmental Research at the DOE. Approximately two-thirds of the human genome sequence is estimated to be completed by NIH- and DOE-funded projects and the remainder by international Wellcome Trust-funded projects.

In addition to the new time frame for the completion of the human genome sequence, the plan also covers new agendas for the (often overlooked) wide range of other areas embraced under the heading of the Human Genome Project. New plans are in effect for supporting and developing novel technology (from sequencing hardware to computer software); setting goals for model organisms; devising genomic resources, such

as cDNA libraries, for the broader biological community; educating and establishing new resources in manpower; and continuing programs to investigate the impact of this new genetic information on the ethical, legal, and social aspects of the population.

2003 or Bust

The goal of completing the human genome sequence by 2003 instead of 2005 is what is catching most of the attention from the press and public. Most wonder if this maneuver is simply a response to the claims by private sector companies that they will provide a sequence of the human genome by 2003—2 years ahead of the initial 2005 date set by the publicly funded genome initiative.

Francis Collins (Director, NHGRI) admits that the plan presented in May by the collaborative venture between Craig Venter and Perkin-Elmer (Celera Genomics Corp.), followed by that announced in August by Incyte Pharmaceuticals to sequence the entire human genome by the year 2003, certainly had a “crystallizing effect.” Collins adds, however, that “the ability to do this [in publicly funded genome centers] is clearly here, and if Celera decided not to do it [pursue sequencing the human genome by 2003] tomorrow, we wouldn’t suddenly decide against going ahead with this bolder plan.”

Richard Gibbs (Baylor College of Medicine Sequencing Center, Houston, TX) likewise pointed out that the publicly funded project has been poised to make such a move for some time. Previous publications and announcements have made it clear that the public project is already 2 years ahead of schedule (Collins 1995)—making 2003 as a completion date a logical outcome. Many in the private sector, however, remain convinced that the faster agenda

can be nothing other than a response to their plans.

To a great extent, whether or not the 2003 mark for the complete sequence is the publicly funded project’s response to private sector plans is rather unimportant when taken out of context with respect to what the scientific community will be getting. If the publicly funded projects were sacrificing quality and completeness for speed in a response to private sector plans, this would indeed be unacceptable. One of the clear aims of the publicly funded projects has been their target of a highly accurate (error rate = $1/10,000$), long-range contiguous sequence for the human genome. That this remains the final goal of the public project even with the earlier 2003 date is the most important factor for the scientific community.

There are two points in this 5-year agenda that are expressly new to the overall human genome sequencing plan. One point is that roughly one-third of the sequence, that is, approximately 1 Gb of sequence with several contigs of no less than 20 Mb in length, will be finished by 2001. This is achievable because of advances in technology over the last 5 years and expected advances in the upcoming years as well as the anticipated scale-up of the best methodologies for sequencing as determined over the last 5 years. A peer-review process will also be set in place to prioritize specific regions to be finished first. Such ranking of genome regions will reflect the needs of the international scientific community and should dovetail with the sequencing strategies at the time, such that large-scale sequencing disruptions are at a minimum. The addition of a selection process for sequencing specific areas of interest or gene-rich regions may reflect a desire to provide early public availability of regions that may otherwise end up solely in the hands of the private sector; con-

¹E-MAIL goodman@cshl.org; FAX (516) 367-8334.

Box 1. Human Genome Project Goals for 1998–2003

- Goal 1: The Human Genome Sequence
 - a. Complete the human genome sequence by end of 2003
 - b. Complete one-third of the human genome by end of 2001
 - c. 90% coverage of genome in working draft with mapped clones by end of 2001
 - d. Make entire sequence freely accessible
- Goal 2: Sequence Technology
 - a. Increase throughput and reduce cost of sequencing technology
 - b. Support novel technology research that can lead to significant improvement in sequencing technology
 - c. Development of methods for introduction of new technologies into sequencing process
- Goal 3: Sequence Variation
 - a. Develop technologies for rapid, large-scale identification of SNPs or other DNA sequence variants
 - b. Identify common variants in coding regions of the majority of identified genes
 - c. Create 100,000-marker SNP map of the human genome
 - d. Develop intellectual foundations for studies of sequence variation
 - e. Create public DNA sample and cell line resources that are totally anonymous and carry appropriate restrictions to maintain their anonymity (Establishment of separate, appropriately marked sample resources after taking into account issues examined in Goal 6)
- Goal 4: Functional Genomics Technology
 - a. Create cDNA resources of both human and model organisms
 - b. Develop improved technologies for analysis of noncoding sequences
 - c. Generate methods and technologies for comprehensive analysis of gene expression
 - d. Develop technologies for creating genome-wide mutations and analyzing the results
 - e. Develop experimental and computational means for global protein analysis
- Goal 5: Comparative Genomics
 - a. Complete the *C. elegans* genome sequence in 1998
 - b. Complete the *Drosophila* genome sequence by 2002 (or earlier because of anticipated work in the private sector)
 - c. Develop genomic resources for the mouse
 - i. Develop an integrated physical and genetic map
 - ii. Develop cDNA resources
 - iii. Complete the mouse genome sequence by 2008 (or earlier given lessons and new technologies from Goals 1 and 2)
 - d. Identify other model organisms of great use to understanding the human genome or that will significantly aid in other genomic studies
- Goal 6: Ethical, Legal, and Social Implications (ELSI)
 - a. Analyze implications resulting from extensive information about human sequence and its variation
 - b. Examine issues involved in integration of genetic technologies and information into health care and public health activities
 - c. Examine issues involved in integration of knowledge about genomics and gene-environment interactions into nonclinical and research settings
 - d. Explore interaction of genetic knowledge, philosophy, theology, and ethics
 - e. Explore ways in which racial, ethnic, and socioeconomic issues affect genetic research, services, and policy development
- Goal 7: Bioinformatics and Computation Biology
 - a. Improve database content and utility
 - b. Develop better tools for data generation, capture, and annotation
 - c. Develop and improve tools and databases for comprehensive functional studies
 - d. Develop and improve tools for representing and analyzing sequence similarity and variation
 - e. Develop and support effective approaches to produce robust, exportable software that can be widely shared
- Goal 8: Manpower and Training
 - a. Nurture training of scientists skilled in genomics research
 - b. Encourage establishment of academic career paths for genomic scientists
 - c. Increase number of scholars with knowledge in both genomic and genetic sciences and in ethics, law, or social sciences

*Information obtained from draft of the New Goals for the U.S. Human Genome Project 1998–2003, as provided at the September 14, 1998 public forum meeting of the NHGRI advisory council. Complete content available in the October 23, 1998 issue of *Science*.

cerns about what effect having large amounts of DNA information tied up in intellectual property rights will have on the public welfare are still unresolved.

The second new addition to the plan is the generation of what is being called a "working draft" of the human genome by 2001. This draft will be based on mapped clones that provide at least 90% coverage of the genome, will have a 99% accuracy, and should be quite useful for finding genes. The one issue stressed about this working draft is that it is a stage that must be passed through in the progress toward the final sequenced genome goal. This was an essential component of the plan, as no one wanted to shunt money into generating a quick, low-quality version of the genome sequence that would ultimately be discarded once the completed version could take its place. The concern here is twofold: (1) the perceived waste of money in generating something that would ultimately be discarded and (2) the potential that the lower-quality version could possibly be chosen as an ersatz final version were it deemed appropriate by individuals with narrower goals for the use of the sequence.

The Gold-Standard Version of the sequence proposed for 2003 does retain a few caveats, which the board was careful in trying to define. Several committee members expressed concern that 2003 would arrive and, without a more detailed indication of what a "finished" sequence was, there was always the potential to simply claim the human genome sequence complete. As indicated by a footnote in the proposal, the board expects that there will remain some regions (an indeterminate, though likely rare, number) that will prove difficult to clone and sequence. The other regions not likely to be finished by 2003 include highly repeated sequences in the centromeres and constitutive heterochromatic regions. Thus, the complete sequence for the human genome in 2003 is expected to retain some gaps that require closure. The completed sequence therefore refers to that portion of human DNA that can be stably cloned and sequenced by current technology.

Not Just Sequence

Publicly funded projects embrace a great deal beyond just determining the human genome sequence. Thus any com-

parison of money spent and completion of goals by privately and publicly funded projects must take into account the additional aspects provided by each.

That the 5-year plan for the human genome project covers more than just human genome sequencing per se is made immediately apparent by the fact that Goal 1 covers the completion of the human genome sequence, whereas Goals 2–8 cover all the additional components that are also under the umbrella of the publicly funded Human Genome Project (see Box 1). These areas have long-term impact not only on the simple achievement of the completion of the human genome sequence by the desired date, but on our ability to understand and use this information.

New technology, better methodologies, and more advanced bioinformatics analysis tools and resources are required for progress in sequencing capability. This is not only essential to meet the 2003 completion date, but also to establish efficient and inexpensive means to continue to utilize high-throughput sequencing technology for additional experimental purposes beyond this date. Genomic studies in model organisms, comparative studies, and further bioinformatics advances are absolutely required to aid in deciphering genome sequences (from gene identification to gene function) and in dissecting whole chromosome function, evolution, and activity. The biological community in general has made tremendous use of EST resources and will benefit more so from planned cDNA libraries, cell line, and DNA sample libraries. Obviously, an entire highly trained work force is required to carry out these plans, and training, funding, and encouraging establishment of such skilled individuals is therefore included in the goals defined for the next 5 years. Finally, all this work and planning could be worth very little if the public ultimately rejects the potential benefits from the genome projects because of misconceptions or fear of the potential misuse or abuse of this newfound knowledge. Thus, the importance of the ethical, legal, and social Implications (ELSI) programs can not be overstated. Scientists and the general public must come together in making decisions about how this information is to be used in both clinical and nonclinical settings.

Taking the entire new 5-year plan

in at a glance, "bold" may be too tame a word.

REFERENCES

- Collins, F.S. 1995. *Proc. Natl. Acad. Sci.* 92: 10821–10823.
- Watson, J.D. and F.H.C. Crick. 1953. *Nature* 171: 737–738.