Perspective

The Human Genome and Other Initiatives

BERNARD D. DAVIS AND COLLEAGUES*

WHEN TECHNIQUES FOR SEQUENCING SEGMENTS OF DNA became available, it seemed clear that the most valuable materials to analyze were the regions that could be associated with a function or a disease. As more than 95% of the human genome does not code for the kinds of functions that we can recognize, it has been, temporarily, called "junk." A few years ago, however, it was proposed that systematic sequencing of the entire human genome was now feasible and a major challenge that we should welcome.

This human genome program (HGP) had an unusual origin. It was not initiated by a committee of molecular geneticists dealing with a pressing need or by the major biomedical funding agency, the National Institutes of Health (NIH). Instead, it was advanced by a politically astute administrator in the Department of Energy (DOE), convinced that the powerful tools of molecular biology made it appropriate to introduce centrally administered "big science" into biomedical research.

The idea quickly developed strong political appeal. Sequencing the entire human genome was as definite and highly visible as putting a man on the moon. It promised large benefits for human health and for our understanding of human biology, and it was claimed to be more efficient and cheaper than reaching the goal eventually by piecemeal sequencing. One Nobel laureate even asserted that knowing the whole sequence would tell us what human beings really are. Small wonder that the HGP became politically unstoppable. It has now acquired a distinguished leader and set of advisers and funding has grown rapidly.

Meanwhile several changes have led to growing concern in the biomedical research community, which stemmed from many scientists' uneasiness about starting big science in their area. First, recent cuts in the funding of grants in other areas have drawn attention to the problem of competition from the HGP. Second, within the HGP, the aim of sequencing the human genome has been replaced by a very different plan. Even though many of the revised goals involve kinds of work that would be going on even if there were no HGP, the program has inherited a high level of support and a centralized mechanism of funding from the initial plan. These changes suggest the need to reevaluate the HGP.

NIH study sections consider two kinds of investigator-initiated applications: new projects and "competing renewals," that is, applications for renewal of grants whose committed period has expired. Within the past few years the combined number of such grants funded annually has fallen from more than 6000 to 4600, and among the total grants approved by study sections, the fraction funded has fallen from more than 40% to less than 25%. Indeed, some study sections have dropped much lower this year. Study sections thus face the depressing prospect of virtually tossing a coin to choose among scientists with a continuing record of distinguished achievement, and promising beginners are even more at risk. The conclusion seems inescapable that far too many excellent scientists are not being funded. In addition, to spread the limited funds, the NIH is cutting the budgets of all awarded grants by 10 to 20% from the levels recommended by study sections.

The consequences are far-reaching. Among them, the continuity of the research enterprise, so essential for its effectiveness, is being disrupted. Moreover, bright students are being discouraged from seeking careers in science at a time when other factors have already seriously eroded the interest of our youth in science.

Part of the reason for the decline in the number of competing renewal grants is simply the greater average length of awards in recent years, which shifts the proportions in the noncompeting and competing categories. However, in addition, there is a real decline in the support of research, with appropriations failing to keep pace with increases in its cost and in the number of excellent applications.

In principle the HGP should not be part of the problem, for it was initially presented as an exciting opportunity for which Congress would be happy to provide additional money. However, even though NIH officials maintain that HGP is not competing, that cannot be satisfactorily demonstrated. Although the HGP is listed as a separate line item, the structure of the budget does not reflect the psychological elements in its formation. These inevitably force Congress, especially in times of stringency, to focus most sharply on the total amount for biomedical research rather than meeting each need separately.

It thus appears that the human genome initiative is competing with the initiatives of investigators in other areas. Furthermore, the scale of the competition is not negligible. One way of quantitating this is by comparing the total budget (NIH+DOE) for HGP to the total NIH budget for grants in biomedical research. The President's budget for next year recommends $4174 million for new and competing plus continuing grants through the NIH. Hence the ultimate annual total for the HGP, $200 million, would be about 5% of the amount allocated by NIH for untargeted research in the biomedical sciences. In fact, half of next year's estimated allocation of $154 million for HGP would fund 385 grants of $200,000—a substantial amount of famine relief for untargeted research.

The perception of competition on this scale is engendering bitterness in the research community. The concern also has a deeper cause, stemming from doubts about the scientific justification for the present status of the HGP. Many are not convinced that a crash program for analyzing the structure of genomes will advance either health or the life sciences, for many years to come, as much as studies of specific physiological and biochemical functions and their abnormalities. In addition, in contrast to some areas of physics, which require extremely expensive facilities, biology does not have an obvious need for "big science." Our country's spectacular success in this area has depended in large part on the wide support of independent, investigator-initiated, peer-reviewed research. Many biologists fear that the precedent of the HGP will weaken this tradition.

To judge these arguments we must consider the altered content of the HGP, and whether its different parts need a centralized approach.

*Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston MA 02115. This article is endorsed by the following members of the faculty of the author's department: Harold Amos, Jonathan R. Beckwith, R. John Collier, Ronald C. Desrosiers, Bernard N. Fields, Deirdre B. Furlong, Michael E. Greenberg, Ulla Hansen, Alice S. Huang, Paul L. Kaplan, Elliott Kieff, David M. Knipe, Roberto G. Kolter, Yen Li, E. C. C. Lin, John J. Mekalanos, Jean L. Patterson, Kathryn Ruoff, Ruth Sager, Priscilla Schaffer, Charles Stiles, and P. C. Tai. It is adapted from the Annual Mahlon Hoagland Lecture at the Worcester Foundation for Experimental Biology.

When a National Research Council committee appraised the initial HGP in 1986 it recognized that starting to sequence 3 billion nucleotides, at a cost of more than $5 each, would be quite impractical. It recommended a different program, which would postpone systematic sequencing for at least 5 years and would concentrate on several other activities. These included developing a map of identifiable short sequences, which would make it much easier to locate any new human disease genes or other genes; sequencing human DNA regions of interest; developing more economical procedures for sequencing; developing systems for effectively storing and accessing the massive data; and studying the genomes of other organisms, ranging from *Escherichia coli* to the mouse.

These additions to the HGP are all scientifically sound. In particular, going beyond humans to other systems makes a great deal of sense, because these organisms can be used as experimental models for human functions and disorders. We will need experimental approaches to solve the most challenging problems, such as understanding polygenic traits (which are involved in most human diseases); elucidating the mechanisms by which a fertilized egg cell develops into an organism; elucidating the molecular basis of information storage and processing in the brain; and acquiring insight into the functions of the enormous amount of so-called "junk" DNA in higher organisms.

Nevertheless, the extensive changes in the new HGP mean that it has moved far from the initial dream. Indeed, most of it, except for the centralized plan for mapping, differs only in mechanism of funding, and in scale, from what we would be doing if the concept of an HGP had never arisen. If these activities were to be proposed as a novel program today it is doubtful that they could generate the strong political appeal of the original proposal.

To evaluate the justification for the present program we must look at its major components separately. A map of signposts along the road will be valuable in locating new genes. Moreover, because finding the locator sites for the map is not itself very interesting, this part of the HGP evidently requires a centralized, quasi-industrial organization rather than conventional grants aimed at more creative discoveries. However, in making decisions about the scale of support, certain limits to its payoff should be recognized. Diseases caused by a malfunction in one gene tend to be rare, and the genes involved in other diseases will probably be studied mostly in experimental animals, which seem uneconomical to map extensively. For example, mapping and sequencing would be just as expensive for the mouse genome as for the human genome.

Another goal of the HGP, systematic sequencing of the entire human genome (now postponed), presents deeper problems. First, even if future automation should bring the cost of sequencing to less than $1 per nucleotide, the effective storage and retrieval of the 3 billion units of information will also be costly, and the total final estimate that $3 billion will yield the full sequence in 15 years is very uncertain.

A more fundamental question is whether identifying the last nucleotide in a human genome really has deep scientific value, apart from its public relations impact. Sequencing the much smaller genomes of a few viruses has been a gratifying milestone, but it has not obviously added a great deal to what had already been learned from the study of specific regions. It is even harder to see how a complete sequence could be useful for understanding the organization of the huge human genome: the magnification is wrong, like viewing a painting through a microscope.

It thus appears that the most meaningful studies of the human genome will identify units within the chromosomes that have functions—of unknown as well as of familiar classes—and will then sequence these units. Several approaches have been fruitful in finding functional regions. It is straightforward to proceed from a known protein to its gene. The reverse process is also effective: locating the gene responsible for a disease by its linkage to known markers in the genome, and then using the sequence of that gene to help identify the mechanism of the disease. An increasingly refined map of the genome is making this "reverse genetics" much easier. In a third approach, blind sequencing of the genome can also lead to the discovery of new genes through recognition of sequences with the characteristics that suggest coding for a protein; but this is not an efficient process. On average it would be necessary to plow through 1 to 2 million "junk" bases before encountering an interesting sequence; and then finding its unknown function would also be difficult.

To be sure, it may be necessary to start from sequences in approaching one of the largest challenges: identifying the unknown functions that almost certainly exist in much of the "junk" DNA. But the answers seem likely to come from detailed studies that would include experimental manipulation of small samples of the genome; it is hard to see how knowing the whole sequence would help.

For these several reasons it is perhaps no great loss that the goal of systematic sequencing has been postponed. When it is reevaluated, 5 years from now, it may seem quite unimportant. Among the other activities of the current HGP, the research on techniques clearly overlaps with the goals of mapping and sequencing. However, the HGP mechanism of funding may give this research an unwarranted advantage over other kinds. Finally, the other major activity, sequencing of interesting regions in humans or in lower organisms, does not have any obvious need for centralized organization, or obvious justification for insulation from competition with other kinds of research.

Although all the goals of the HGP, except for the complete sequencing of the human genome, are clearly worthwhile, there is concern over its competition with other research for funds at a time of financial stringency, and doubt that its scientific benefits justify its rapid expansion and its organization in the pattern of big science, have engendered widespread dissatisfaction. This is illustrated by the virtual unanimity of the departmental faculty that is endorsing this statement. The HGP may therefore need reevaluation.

In any such reevaluation there would be no difficulty in justifying a centralized organization for the mapping, and probably for the research on methods. However, it is not obvious that these activities justify support for the HGP at a level equivalent to over 20% of all other biomedical research. The other main activity, sequencing regions of interest in various organisms, does not appear to need centralized organization, and its grant applications might be judged more equitably if they were fitted into the general program of investigator-initiated research.

In reevaluation of the HGP, we should recognize that excessive funding competition may interfere with attainment of the project's own goal, because investigations in many areas of research can lead to DNA regions of interest. Our fundamental goal is to understand the human genome and its products, and not to sequence the genome because it is there.