Report to the
Director,
National Institutes
of Health

AD HOC PROGRAM ADVISORY COMMITTEE ON COMPLEX GENOMES

WHITEHEAD INSTITUTE

Dr. David Baltimore
Director
617/258-5155

John Pratt
Associate Director
617/258-5104

June 14, 1988

Dr. James Wyngaarden
National Institutes of Health
Building 1, Room 124
9000 Rockville Pike
Bethesda, MD 20814

Dear Dr. Wyngaarden,

Enclosed is the final report of the NIH Ad Hoc Program Advisory Committee on Complex Genomes. This report, skillfully put together by an independent rapporteur and reviewed by the Committee members, reflects well and concisely the discussions at the single meeting of the Ad Hoc Committee that took place on 29 February and 1 March, 1988. Embodied in it are the major recommendations that emerged from the Discussions.

I trust that this document can serve as the basis for NIH to develop a vigorous program to exploit the opportunities in Genome Analysis that have arisen in the last few years. Such a program would complement NIH's other efforts in biomedical research and would lead our science into even-deeper understanding of how complex genomes are organized and the nature of the information they encode. By interdigitating NIH's efforts with those of other agencies and other countries, an international program of great power and importance should emerge.

Sincerely,

David Baltimore

NATIONAL INSTITUTES OF HEALTH

RECOMMENDATIONS AND PRIORITIES DEVELOPED

BY THE AD HOC PROGRAM ADVISORY COMMITTEE ON COMPLEX GENOMES

- Research and development for competitive grants exploring methods for more cost-effective cosmid acquisition, verification, and storage: $1 to $2 million per year, to begin immediately.

- About 4,000 RFLPs are needed to develop a map having a 1-centiMorgan resolution. The anticipated cost is $1 million per year for 5 years, for a total of $5 million.

Estimated Costs for Training Needs

- The program should be training 50 predoctoral students and recruiting 50 postdoctoral scientists--building over a 5-year period to 150 predoctoral and 150 postdoctoral researchers per year. Computing annual training costs at $20 thousand and $25 thousand per individual, respectively, the training program will grow from $2 million to $7 million per year in 5 years.

- Information handling will require a special component for training in biology and computer/mathematical analysis:

  -- Initially it should support three to five such programs with three to five students each, at an initial program cost of $700,000 per year, building to 50 to 75 students per year after 5 years with an annual program cost of about $7.5 million

  -- Training must extend to postdoctoral level scientists and senior visiting scientists, with the annual program initially set at $1 million for about 20 postdocs and five visiting scientists but growing after 5 years to 50 postdocs and 10 visiting scientists for a total annual cost of about $6 million.

  -- Training of general molecular biologists in computational methods of genomic analysis, including the development of curricula and educational materials, with the expectation that all molecular biology graduate students should eventually take one such course. Cost: $750,000 per year.

Construction and Infrastructure Costs

- For building and remodeling laboratory facilities, particularly as efforts are begun to conduct large-scale DNA sequencing at the megabase level and beyond, an overall budget of $30 million is recommended (alternatively, an appropriate estimate of this figure is 20 percent of the total budget), with outlays of about $3 million beginning in the second year and the remainder to be spent over the next several years.

- Provision must be made for administrative costs incurred because of the new program.

* U.S. GOVERNMENT PRINTING OFFICE: 1988-217-780

- Information research through a grants program: An additional ten 3-year grants (with an average size $300,000 to $500,000) should be funded the first year, building to a total of 30 such grants running per year after 5 years. The annual budget would grow from about $3 to $6 million to about $12 million.

- In addition to support for existing national data bases, 10 to 15 percent of total costs of mapping and sequencing projects should be added on for the purpose of data organization.

- Centers of excellence devoted to information analysis should be established either within institutions already conducting mapping and sequencing projects or at universities without such activities but with more general interests in related areas, such as protein structure analysis or genetics. Core facilities and central research support will require $1 million per year for each center. Thus initially, two to three will cost $3 million per year, growing after 5 years to about 10 with an aggregate total budget of $10 million.

Biological Materials Working Group

- The cost for developing and maintaining a full collection of overlapping cosmids of the human genome could amount to $100 million, based on current estimates of $10 million per year for 10 years. However, technological developments are expected to reduce this budget significantly, possibly by an order of magnitude.

- Before any undertaking to build a full collection begins, pilot projects to improve technology are needed, and they should be funded through competing grants totalling $1 to $2 million per year.

- The estimated expense for setting up a U.S. repository for familial cell lines to construct reference maps is $1 to $2 million per year for a 5-year period, or a total cost of $5 to $10 million.

- Some 20 to 25 regional human genetics centers, each employing one to two experts, should be established at an aggregate annual cost of $2 to $3 million for 10 years. In addition, the Camden, N.J. repository should be budgeted for $1 million per year for 10 years. Total Cost: $20 to $40 million.

- Costs could amount to $5 million (total) over 5 years for hybrid cell lines containing a full collection of human chromosomes with necessary replicates.

- As an approximation dependent on unrealized (not yet undertaken) improvements in technology, the working group calculates that introduction of 10,000 cosmids per year into a repository requires an annual budget of $10 million, for a projected cost of $50 million for 5 years.

GENERAL POLICY ISSUES

The prospect of mapping and sequencing complex genomes, particularly the human genome, looms as an important challenge for the biological community. Discussions about the pace and strategy for best proceeding have sometimes been intense. But there is also growing recognition within the biomedical research community that parts of an informal program already have been taking shape because of rapid progress in molecular biology and genetics.

To assess these developments, Dr. James B. Wyngaarden, the Director of the National Institutes of Health, recently convened a group of experts as an Ad Hoc Program Advisory Committee on Complex Genomes (CCG)* (Attachment I). At a meeting late in February, the Committee expressed confidence that, by building on the groundwork now in place, a well planned effort to intensively analyze the human genome could bring dramatic success in a few years--and substantial completion of the project within a 15-year period. However, to meet this ambitious goal, the CCG believes that a systematic, centrally-coordinated initiative is required.

Many benefits would accrue from such a complete analysis of the human as well as other complex genomes. Besides the wealth of biomedical information this concerted effort will provide, enormous scientific and technological advances can be expected, having both basic and commercial applications. Many challenges lie ahead if this effort is to succeed, but it is likely to enhance U.S. competitiveness, particularly in the growing international arena of biotechnology.

The Ad Hoc CCG has developed a set of guidelines and recommendations for establishing a formal program within NIH, tentatively named the Human Genome Research Program. The Committee is well aware that plans for such a program are likely to evolve rather rapidly, particularly as expected technological developments pave the way for improved efficiency in different phases of this undertaking. Indeed, the members of the Committee did not reach consensus on some issues of significance. Nevertheless, CCG members agreed that certain priorities must be set now, and some principles for shaping the overall NIH program must be unequivocally established from the outset. They include:

- Peer review and permanent advisory committee. Stringent peer review will be vital during all stages of this project, with the emphasis on selection of the best scientific proposals put forth by scientists drawn from a broad base. Because the program will highlight special technology needs and unusual targeted efforts compared to other research components within NIH, special study sections should be established as needed. Moreover, a permanent Program Advisory Committee made up of expert scientists from appropriate disciplines will play a crucial role in keeping efforts on target.

* The NIH Ad Hoc Program Advisory Committee on Complex Genomes met February 29-March 1, 1988, in Reston, Virginia.

- Establish an Office of Human Genome Research. The Ad Hoc Program Advisory Committee unanimously endorses a proposal to establish an Office of Human Genome Research, headed by a new Associate Director, within the Office of the Director of NIH. CCG believes that separate management of this new program not only will stress its unique identity, but will emphasize that it should not in any way disrupt other NIH research programs. The Associate Director and staff is expected to work closely with the permanent Program Advisory Committee, as needed, and to coordinate related efforts with representatives from all NIH components and elsewhere within the research community.

- Role of model systems in providing insights. Insights into human diseases will be rapidly forthcoming from this effort, and careful attention must be given to disseminating medically useful information to the wider community. Despite this obvious emphasis on human genetics, vital insights can and will be gained by continuing reliance on pertinent model systems--microbial, lower and higher animals, and even plants when appropriate scientific justifications are put forward. Considerable interplay is expected with other research fields, and free exchange of information and technologies must be fully encouraged at all levels.

- Rolling 5-year plan with annual review. Because the technologies for mapping, sequencing, and handling the biological materials and data to be generated are expected to change dramatically, an annual review and adjustment of program priorities are anticipated. Thus, although a tentative 5-year agenda should be set now, there must be ample flexibility to change tactics when new opportunities arise, particularly during the first few years of this enterprise.

- Training needs. Special attention must be given, particularly during the early phase of this program, to assessing and providing for the specialized training needs that such an interdisciplinary undertaking entails.

- National Academy of Sciences Report. The Ad Hoc Program Advisory Committee agrees with many additional recommendations outlined in the report, "Mapping and Sequencing the Human Genome" of the National Research Council of the National Academy of Sciences (February 1988) (Attachment II).

SPECIFIC PROGRAM NEEDS

The Ad Hoc Committee believes that the Human Genome Program represents a unique opportunity in modern biological research. The grand scale of the program makes the undertaking unique in biological research, and it also calls upon scientists with a uniquely wide variety of skills to work closely together. Moreover, the nature of the undertaking dictates that scientists representing different disciplines work cooperatively, with little or no hierarchical relationships. Thus, for instance, the insights of a computer scientist or a biochemical engineer working on some seemingly obscure aspect

COST APPENDIX DEVELOPED BY THE AD HOC PROGRAM ADVISORY COMMITTEE ON COMPLEX GENOMES

Mapping Working Group

- Construction of a genetic map with an average 1-centiMorgan resolution should proceed as rapidly as possible. It will cost $10 to $15 million per year for 3 to 5 years, for a total of $30 to $50 million. Work on nonhuman model systems should also be planned at an additional total cost of $4 million per year, or $20 million for 5 years.

- Physical mapping of chromosomes also should be undertaken with particular attention paid to improving technologies. Funding should support 10 large and varied projects, such as mapping the Drosophila genome or a small human chromosome, each costing $1 million per year for 3 to 5 years. Once completed, such advances set the stage for mapping human chromosomes during the next 5 to 10 years.

- Targeted technology development that is not incorporated into specific mapping projects during the first 3- to 5-year period will cost $5 million per year, for a total of $15 to $25 million.

DNA Sequencing Working Group

- The principal short-term goal is to establish 10 to 20 exploratory research programs of varied size, with a substantial component in technology development. At the outset of a 5-year period, spending will be $7 million per year, growing to $60 million per year.

- Technology changes could dramatically affect this component of the overall effort, so that in the second 5-year period, complete sequencing of small-sized human chromosomes will be undertaken, leading to an effort on the entire genome between years 10 and 15.

Information Working Group

- A high priority is to establish a national information coordinating center. Starting from an initial outlay of $3 million per year, the budget for the center should grow to $8 to $9 million per year after 3 years.

  [These figures are based on: 30 to 35 FTEs at the steady state (at a cost of $3 to $3.5 million); $2.5 million per year in computational resources; and $2.5 million per year in targeted extramural support.]

of a particular project could well prove vital to the entire program. Hence, proper attention must be given to streamlining communications among all participants and for encouraging a continuous exchange of ideas.

During the course of its deliberations, the Ad Hoc CCG scrutinized the prospective program by dividing it into four main activity areas: mapping, sequencing, information, and biological materials. Each of these groups developed goals and tentative plans for meeting them as well as projected costs for doing so (for projected costs, see Attachment III).

However, certain components of the program and the costs for meeting such needs are general, cutting across all the activity areas. The priorities identified by the CCG that transcend working group subject areas fall into two major categories--training needs and infrastructure and construction requirements. Key requirements include the following:

- The [Human] Genome initiative entails new training needs for scientists with interdisciplinary skills. To begin with, the program should be training 50 predoctoral students and recruiting 50 postdoctoral scientists--building over a 5-year period to 150 predoctorals and 150 postdoctorals.

- Provision must also be made for training master's and bachelor's level scientists and technicians, who will make valuable contributions to this overall program.

- Information handling will require a special component to encourage students to obtain coordinated training in both biology and computer/mathematical analysis.

  -- Initially three to five such programs with three to five students each should be supported, building to 50 to 75 students per year after 5 years.

  -- Training must extend to postdoctoral level scientists and senior visiting scientists, with the annual program initially set for about 20 postdocs and five visiting scientists but growing after 5 years to 50 postdocs and 10 visiting scientists.

- Some provision must be made for building and remodeling laboratory facilities, particularly as efforts are begun to conduct large-scale DNA sequencing at the megabase level and beyond.

- Provision must be made for new administrative costs that arise from the program, including the efforts associated with the proposed new Office of Associate Director and staff, a permanent Program Advisory Committee, and new study sections devoted to reviewing Human Genome Program proposals.

Mapping and Sequencing

The Ad Hoc CCG carefully examined the specific scientific and technological tasks that make up the new Human Genome Research Program. Although the overall goals of the program are well defined, many of the intermediate achievements will necessitate adjustments along the way. Because much of this anticipated development will lead to improved and more efficient methodologies, the Committee members stress the importance of building a diversified effort that is critically reviewed at every step--particularly during the early stages.

Similar thinking dictates that major DNA sequencing efforts be phased in gradually, with each increment awaiting expected improvements in technology and careful evaluation of their impact on costs before projects ascend to the next level in volume and complexity. Committee members feel that breakthroughs in technology must come before there is any concerted move to a massive "assembly-line" approach to DNA sequencing. Indeed, the need for a centralized effort in DNA sequencing conceivably may never arise if appropriate technological developments in automation enable a fully decentralized approach. However, some consolidation of efforts is expected as improvements in technology dictate. Despite the realization that better approaches to DNA sequencing are needed, however, there also is a need to develop them in the context of sequencing biologically relevant segments of DNA. Moreover, decisions to proceed will involve a constant interplay of costs and value considerations.

The current limitations on technology for mapping genes are different than those faced for sequencing DNA. Genetic mapping now can go ahead on a substantial scale. The Committee recognizes that the available technology is ready for construction of a 1-centiMorgan average resolution linkage map, and the pace of this effort will depend primarily on available funding--not on an awaited technology. Although significant improvements are to be actively sought, physical mapping can now be done on a substantial scale with a fairly centralized effort. Directed projects to develop such maps are expected to spur developments in technology.

Specific recommendations and needs for mapping and DNA sequencing include the following:

- Three types of maps should be developed in parallel, beginning immediately (high resolution genetic maps, macrorestriction maps, and linked libraries).

- Support is needed for the following smaller-scale efforts to improve mapping and DNA handling technologies:

  -- Physical means for separating intact human chromosomes or other large DNA fragments with high resolution;

  -- Isolating and maintaining human chromosome fragments within cultured cell lines;

  -- Cloning and purifying large DNA fragments;

-- Ordering adjacent DNA fragments in a DNA clone bank; and

-- Automating DNA mapping, including DNA purification and hybridization analysis, handling of DNA samples, and mathematical analysis to aid in map construction.

- Full-scale DNA sequencing should not begin until technology developments dictate greater efficiency of effort. To achieve that goal, a gradual build-up through practical sequencing undertakings will be helpful, particularly as they provide increased confidence in the scientific value of the data being obtained. Thus, contiguous megabase sequencing on model systems (such as the major histocompatibility locus) could begin in 1 to 2 years, building to sequencing of small human chromosomes within 5 years. The full human genome might be tackled during years 10 to 15.

- Great emphasis must be placed on technology development and its dependence on interdisciplinary efforts. Even modest improvements in slow ("rate-limiting") steps in sequencing can bring about significant gains in efficiency.

- Cooperation between industrial, academic, and Federal laboratories also must be expressly encouraged. The successful transfer of new technologies is one of the most important achievements expected from this program. For example, new instruments most certainly will be developed as an outgrowth of this effort.

- The Ad Hoc Program Advisory Committee carefully discussed the important role of genetic model systems in complementing efforts to better understand the human genome. In different phases of this project, various nonhuman model systems will provide valuable insights in establishing techniques and providing crucial comparative landmarks in this uncharted territory. Sound scientific proposals that focus on model systems--including Drosophila, the mouse, or even simple plant systems--may well provide unique insights into technical problems and could also lead to solving problems directly related to human disease. A continuing effort will be needed to coordinate efforts with other established biomedical research programs to avoid unnecessary overlap.

- Consideration should be given to supporting some projects to explore the feasibility of constructing cDNA maps. The Committee recognizes the many challenges involved in undertaking such projects, particularly in deciding what types of cells should be used for source materials and in obtaining low abundance molecules of messenger RNA. Also, it is not yet known to what extent eukaryotic cells rely on control at the transcriptional level. Nonetheless, some exploratory efforts could uncover highly valuable information about cellular functions.

Funding Projection

- Estimated cost: $200 million per year, to be reached during the third year of the project.
- Estimate is based on a projected total of 1,200 individuals @ $100,000 annually.
- "The committee's possible scenario divides the project into 3 five-year periods. During each period, mapping and sequencing efforts five times as complex as the next lower numerical designation would be undertaken at constant cost, reflecting five-fold increments in technological sophistication. This plan requires the following:
  - A major effort must be expended in technological development;
  - "New methods of DNA subcloning and processing will have to be developed (or present ones automated) to stay within the estimated costs;
  - DNA clones from an ordered DNA clone collection will be sequenced, thereby producing large contiguous stretches of DNA sequence that are immediately useful;
  - This effort will require the recruitment of scientists with extensive experience in mapping and sequencing. The multidisciplinary centers supported by the project will play a key role in training new independent scientists--a major benefit to the biological community.

Information and Biological Materials

The Human Genome Program poses major new challenges in the handling, storage, and analysis of biological materials and information. Special resources will be needed for dealing with materials, and new methods must be developed to improve the efficiency with which materials are exchanged among research groups. Similarly, the program will generate vast amounts of data, and they also must be stored, handled, and analyzed efficiently.

Special provisions must be made for collecting genetic materials from families in which genetic diseases segregate. The coordination of this effort also very quickly takes on information handling challenges. Moreover, it tends to involve existing research activities outside the new program. Such questions of overlap help prompt the need for a subcommittee to deal specifically with oversight of information resources. Key needs for dealing with information and biological materials include the following:

- Biological resource collections are vital for distributing valuable biological materials and preserving them in archives. For the program, central, federally supported repositories are crucial for two categories of materials--cells and cloned DNA segments and probes.
- The familial cell lines needed to construct a 1-centiMorgan average resolution map are already available as cell lines kept in a repository in Paris. However, a more conveniently accessible repository for such cell lines should perhaps also be established in the United States.
- The collection of genetic materials from families in which genetic diseases segregate is largely underfunded. Efforts to collect such materials badly need better coordination. However, the central collection effort may not itself be a part of the genome project, and other Institutes of NIH must be called upon to deal with collections needed for studying particular diseases, such as schizophrenia and heart disease.
- Cloned DNA segments are a vital but, currently, very costly item to collect, analyze, and store. Pilot projects must be undertaken to solve technical and cost problems.
- A national center will be required for coordination and reasonably uniform integration of the many data bases that the genome project will generate. The National Library of Medicine may be the right place for establishing such a national center. Conceivably, the current Biotechnology Information Center could grow into this role.
- Coordination of data bases is an absolute requirement. A community minimum standard should be established for all NIH-sponsored data bases that would allow format and database management system inter-convertibility and maximum ease of cross-indexing. The existing and needed data bases, not to be limited to those containing information on human genetics, should include:

  - The committee is not in favor of establishing a few large production centers for mapping and sequencing, given the current state of technology.

- Support for some centers and pilot projects may be amenable to the contract funding mechanism awarded on a competitive basis.

- Competition between centers should be encouraged.

- Include selected other organisms required for interpretation of the human genome map and sequence.

Management Strategies

- "It is imperative to design a management system that will provide oversight, coordination, review of progress, and forward planning."

- The committee reached a consensus for a lead agency, either NIH or DOE.

- "Although the lead agency would have the ultimate responsibility for funding and policy decisions, it would draw on the advice and expertise of a Scientific Advisory Board." Responsibilities of the Board would be to:

  - Facilitate coordination;

  - Assure accessibility of information and materials;

  - Monitor quality of research by helping to assure a uniform standard of peer review;

  - Suggest mechanisms for quality control on mapping and sequence data;

  - Promote international cooperation;

  - Make recommendations concerning the establishment of large sequencing endeavors; and

  - Publish periodic reports on progress, problems, and research recommendations.

- The Scientific Advisory Board would require funding, perhaps from private institutions as well as the lead agency."

- If funding is provided by several separate U.S. government agencies, as well as by private funds, an effective reviewing body will be needed to avoid excessive duplication of effort and to oversee cooperation between research groups.

- The committee believes it will become necessary to have some major organized mechanism for international cooperation. The objective would be to collate data and ensure rapid accessibility to it, and to distribute materials, such as cloned DNA fragments.

  -- DNA sequence, protein sequence;

  -- Genetic linkage maps (restriction fragment length polymorphisms);

  -- Physical maps of overlapping clones and of restriction sites;

  -- Cytogenetic maps;

  -- Tracking materials generated by the program;

  -- Data bases such as the Brookhaven Crystallographic Data Base;

  -- Bibliographic references relevant to data generated;

  -- Full-text data bases, such as Mendelian Inheritance in Man;

  -- Secondarily derived data bases of patterns, such as those consisting of patterns defining protein or DNA functional features; and

  -- Genetic markers and data bases of genes, such as the Yale-Howard Hughes data base.

- The proposed national genome information coordinating center will have as one of its charges taking basic research in informatics and applying it to existing data bases that represent national research resources. Having a national center in which all available data bases are maintained and made accessible makes possible other local centers that will be free to explore different approaches and uses.

- Distribution of and free access to the data bases must be fully encouraged. Thus, the data must be in the public domain, and the redistribution of the data should remain free of royalties. Moreover, the operation (collection and organization) of the human genome databases should not be linked to income from its distribution. All data coming from federally funded research related to the Human Genome Research Program should be expeditiously submitted and entered into national databases.

- Basic research needs in information include research to improve data representation, man-machine interfacing, investigations into new chips and new architecture, and developing new means for examining genomic variation and three-dimensional structure, finding similarities, recognizing sites and features, and analyzing mapping strategies. It should be done at the proposed center and extramurally, and it should feed back into the data bases.

Information Handling

- All human map data should be accessible from a single data base.
- "To derive the full benefit of the human genome sequence will require many new tools, including a comprehensive data base of DNA sequences from other organisms."
- A centralized facility is needed to collect, store, analyze, and distribute information.
  - An initial analysis of these data should be carried out in the facility;
  - All data must be provided to the center in electronic or magnetic form; and
  - The information center must be linked to data users via a computer network.
- "Decisions for a major push on bulk sequence data collection, as distinct from the envisioned pilot projects that push technology development, would depend on how fast the new sequencing technologies develop."
- Encourage the activities of those individuals who combine skills in computer programming and biology as they will be needed to generate the DNA sequence search routines of greatest utility to the biological community.

Implementation

- Funding ought not to be provided at the expense of currently funded biological research.
- Funding ought to be distributed by peer review, grants to be awarded for 3- to 7-year periods. The committee specifically recommends the form of peer review in place at NIH.
- "Establishment of a competitive grant program specifically focused on improving in 5- to 10-fold increments the scale or efficiency of mapping and sequencing the human genome. These grants would be designed to support work that is more technologically oriented than most ongoing biological research."
- "This project ought to include work by both small research laboratories and larger multidisciplinary centers formed by juxtaposing several small research groups having different expertise."
  - Multidisciplinary centers would comprise 3 to 10 research groups, each with an outstanding independent scientific director and a different but related focus. The center could share equipment and personnel as a core facility.

- Separating large DNA fragments with higher resolution;
- Ordering the adjacent DNA fragments in a DNA clone bank, including mathematical and statistical work that would aid in map construction;
- Automating DNA mapping, including DNA purification and hybridization analysis, and handling of many DNA samples simultaneously; and
- Data recording, storage, and analysis.

Sequencing

- "Initially, improvements in existing technology and the development of new technology directed toward the long-range goal of a complete human genome sequence should be vigorously encouraged. This effort would include applications of automation and robotics at all steps in cloning and sequencing." "The awarding of competitive grants to individuals and to larger groups organized into cooperative, multidisciplinary centers is viewed by the committee as the most effective way to achieve these goals."
- "The disparities between the capabilities of current technology and the magnitude of the work required to sequence the human genome suggests that fundamentally different technologies deserve serious exploration."
- "Human gene sequencing by individual research groups interested in specific genes should be strongly supported by standard research grants."
- Encourage development of technology to extend the length of contiguous sequence that can be determined on a single polyacrylamide gel.
- A pilot study to be "initiated immediately would define as its goal sequencing approximately 1 million nucleotides of continuous sequence."
- A mechanism of quality control must be developed to monitor the groups that are contributing extensive sequence DNA information. One might consider an external group that functions as the National Bureau of Standards does to provide independent quality control.

Biological Materiel

- Ordered DNA clone collections should be completed. "A facility for collecting and distributing material should be organized to handle the cloned DNA fragments generated and mapped in the many different laboratories involved." This facility would store the appropriate clones, index them according to plan, and then redistribute them upon request.
- Stability of stored DNA fragments is still a problem.
- There may be a need for more than one production center (in addition to CEPH) to grow cells and to produce and distribute DNA.

National Research Council
Commission on Life Sciences
Board on Basic Biology
Committee on Mapping and Sequencing the Human Genome*

RECOMMENDATIONS

Mapping

- Full-scale mapping (not gene by gene), both genetic linkage and physical, should begin immediately.

- "Because the technology needed for genetic linkge mapping with RFLPs is more advanced than that for physical mapping, an immediate emphasis should be placed on completing the genetic linkage map. A project with the goal of attaining of a fully connected map with an average resolution of 1 cM is strongly recommended."

- All types of maps (restriction-site maps, cDNA maps, ordered DNA clone maps, and genetic linkage maps) "need to be coordinated as part of a human genome project."

- Encourage researchers' natural tendency to construct detailed maps of chromosomal regions of particular interest.

- "The committee specifically recommends against a centrally imposed plan to proceed from lower to higher resolution as is implicit, for example, in proposals to complete the entire physical map before initiating pilot sequencing projects."

- "Most support should be to groups that are attempting to map large genomes, with support for different mapping methods proceeding in parallel."

- Development and refinement of techniques should be emphasized early in the mapping part of the project.

- Improved methods for the following would facilitate map construction and usefulness:

  - Physically separating intact human chromosomes;

  - Isolating and immortalizing identified fragments of human chromosomes in cultured cell lines;

  - Cloning cDNA from low abundance mRNA and obtaining "normalized" cDNA libraries;

  - Cloning and purifying large DNA fragments;

* Mapping and Sequencing the Human Genome©, 1988, National Academy of Sciences.

AD HOC PROGRAM ADVISORY COMMITTEE ON COMPLEX GENOMES

Sheraton International Conference Center
Reston, Virginia

February 29-March 1, 1988

List of Committee Members

Chairman: David Baltimore, Ph.D.
Director
Whitehead Institute
9 Cambridge Center
Cambridge, Massachusetts 02142
(617) 258-5155

Francisco Ayala, Ph.D.
Distinguished Professor
of Biological Sciences
Department of Ecology and Evolutionary

C. Thomas Caskey, M.D.
Investigator
Howard Hughes Medical Institute
Director and Professor

[redacted]

[redacted]

[redacted]

ATTACHMENT II

AD HOC PROGRAM ADVISORY COMMITTEE ON COMPLEX GENOMES

Sheraton International Conference Center
Reston, Virginia

February 29-March 1, 1988

NIH Working Group on Complex Genomes

Chairman: James B. Wyngaarden, M.D.
Director
National Institutes of Health
[redacted]

Vice Chairman:
Jay Moskowitz, Ph.D.
Associate Director for Science Policy
and Legislation
National Institutes of Health
[redacted]

Executive Secretary:
Rachel E. Levinson, M.A.
Program Planning Branch
Office of Science Policy
and Legislation
National Institutes of Health
[redacted]

Duane Alexander, M.D.
Director
National Institute of Child Health
[redacted]

Ruth Kirschstein, M.D.
Director
National Institute of General
Medical Sciences
[redacted]

Phillip Gorden, M.D.
Director
National Institute of Diabetes and
Digestive and Kidney Diseases
National Institutes of Health
[redacted]

Donald A.B. Lindberg, M.D.
Director
[redacted]icine
Health
[redacted]94

George E. Palade, M.D.
Senior Research Scientist
Department of Cell Biology
Yale University School of Medicine
[redacted]

Betty H. Pickett, Ph.D.
Director
Division of Research Resources
National Institutes of Health
[redacted]

Additional NIH Participants

Vida H. Beaven, Ph.D.
Assistant Director for Program
Coordination
National Institutes of Health
[redacted]

Katherine L. Bick, Ph.D.
Deputy Director for Extramural Research
National Institutes of Health
Shannon Building, Room 107
[redacted]

James Cassatt, Ph.D.
Deputy Director
Biophysics and Physiological
Sciences Program
National Institute of General
[redacted]

Cheryl Corsaro, Ph.D.
Executive Secretary
Special Review Section
Division of Research Grants
[redacted]

Judith H. Greenberg, Ph.D.
Acting Director
Genetics Program
National Institute of General
Medical Sciences
[redacted]

Delbert H. Dayton, M.D.
Chief
Genetics and Teratology Branch
Center for Research for Mothers
and Children
National Institute of Child Health
[redacted]

Irene A. Eckstrand, Ph.D.
Chief
Molecular and Medical Genetics
Section
Genetics Program
National Institute of General
[redacted]

Richard Farishian, Ph.D.
Chief
Program Analysis and Evaluation
Section
Office of Program Planning and
Evaluation
National Institute of Diabetes and
[redacted]

Donald H. Luecke, M.D.
Deputy Director
Division of Research Grants
[redacted]

David A. Smith, Ph.D.
Department of Energy Liaison
Acting Deputy Director
Health Effects Research Division
ER 72
Office of Health and Environmental
[redacted]

J. Craig Venter, Ph.D.
Chief
Section of Receptor Biochemistry
and Molecular Biology
National Institute of Neurological
[redacted]s and Stroke
[redacted]h

Walter Gilbert, Ph.D.
Chairman
Department of Cellular
and Developmental Biology
Carl M. Loeb University Professor
Biological Laboratories
[redacted]

Diane R. Hinton
Manager
Human Gene Mapping Project
[redacted]

Patricia Hoben, Ph.D.
Analyst
Biological Applications Program
[redacted]

Michael Hunkapiller, Ph.D.
Scientific Director
Applied Biosystems, Inc.
[redacted]
94404

David J. Lipman, M.D.
Research Scientist
Mathematical Research Branch
National Institute of Diabetes and
Digestive and Kidney Diseases
[redacted]

Jacob V. Maizel, Jr., Ph.D.
Laboratory of Mathematical Biology
Division of Cancer Biology and
Diagnosis
National Cancer Institute
National Institutes of Health
[redacted]

Daniel R. Masys, M.D., F.A.C.P.
Director
Lister Hill National Center
for Biomedical Communications
National Institutes of Health
[redacted]

Ira Pastan, M.D.
Chief
Laboratory of Molecular Biology
National Cancer Institute
National Institutes of Health
[redacted]

Lesley M. Russell, Ph.D.
Professional Staff
Energy and Commerce Committee
[redacted]

David Schlessinger, Ph.D.
Professor
Department of Microbiology
and Immunology
Washington University
School of Medicine
[redacted]

Mark S. Guyer, Ph.D.
Program Administrator
Genetics Program
National Institute of General
Medical Sciences
[redacted]

Robert Katz, Ph.D.
Director
Metabolic Diseases Research
National Institute of Diabetes
[redacted]

Jane Peterson, Ph.D.
Program Administrator
Genetics Program
National Institute of General
Medical Sciences
[redacted]

J. Edward Rall, M.D., Ph.D.
Deputy Director for
Intramural Research
[redacted]

James Willett, Ph.D.
Chief
Office of Program Planning and
Evaluation
Division of Research Resources
[redacted]

AD HOC PROGRAM ADVISORY COMMITTEE ON COMPLEX GENOMES

Sheraton International Conference Center
Reston, Virginia

February 29-March 1, 1988

List of Participants

Meeting Chairman: James B. Wyngaarden, M.D.
Director
National Institutes of Health
[redacted] 124
[redacted] 892

Christian Burks, Ph.D.
Staff Member
Coprincipal Investigator
GenBank®
Los Alamos National Laboratory
[redacted]

Igor B. Dawid, Ph.D.
Chief
Laboratory of Molecular Genetics
National Institute of Child Health
[redacted]

John E. Burris, Ph.D.
Director
Board on Basic Biology
[redacted] s
[redacted] .W.

Helen Donis-Keller, Ph.D.
Senior Research Director
Human Genetics Department
Collaborative Research, Inc.
[redacted] ts 01730
[redacted]

Jeffrey L. Fox, Ph.D.
Science Writer
[redacted]

George F. Cahill, Jr., Ph.D., M.D.
Vice President
Scientific Training and Development
Howard Hughes Medical Institute
[redacted]

Donald S. Fredrickson, M.D.
Guest Researcher
National Institutes of Health
[redacted]