The University of Michigan
Medical Center
Ann Arbor, Michigan 48109-0618

Department of Internal Medicine
Division of Medical Genetics
(313) 763-2532

Francis S. Collins, M.D., Ph.D.
Division Chief
[redacted]

Thomas D. Gelehrter, M.D.
[redacted]

Andrew P. Feinberg, M.D., M.P.H.
[redacted]

July 18, 1989

Dr. Anthony V. Carrano
Director Human Genome Project
Genetics Section Leader
Lawrence Livermore National Laboratory
University of California
P.O. Box 5507
Livermore, CA 94550

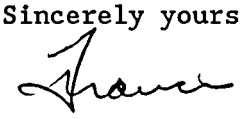
Dear Tony,

Thank you for your letter of June 25, 1989. I apologize for a slightly late response but I have just recently returned from Nigeria and a great deal of paperwork has piled up. I have reviewed your draft document on "Sharing Genome Information and Byproducts", and find it to be an excellent set of guidelines regarding responsibilities of those who generate materials useful for the genome project. I am in agreement with essentially all of the conditions which you have included in this draft.

As your letter points out, what is not presented in the document is the responsibilities of recipients, which to my mind is equally important if the information is to be collected in an optimal fashion. I do feel that users should be obligated to feed back information using arrayed libraries to the distribution center of those libraries, so that the data can be assembled. Clearly the mapping effort will lag months or years behind if this is not done in a foolproof fashion. Appropriate carrots and sticks would need to be included in such a policy. For example, compliance with the policy would ensure that an investigator would be able to obtain other libraries, whereas failure to do so might be noted by a study section in a future grant review and taken as a strong negative for future support. In general, I think the distribution center needs to maintain control over these arrayed libraries in order for chaos not to ensue. Therefore, I would think that rearraying of a library should in general not be done by users except by special arrangement with the distribution center. Furthermore, distribution of clones ought most appropriately to be done by the center, which is another way of guaranteeing that the information is reaching them. A strong distribution center would also answer point three of your letter, since it would be immediately apparent which investigators were working with a particular arrayed library, facilitating formal or informal interchanges of information between users on a horizontal level.

I hope these brief comments will be useful, and I commend you on the important effort which you have already made. This excellent start makes the job much easier for the rest of us. Best regards.

Sincerely yours,



Francis S. Collins, M.D., Ph.D.
Associate Investigator
Howard Hughes Medical Institute
Chief, Division of Medical Genetics

FSC/bs(1-carr)

Biomedical & Environmental Sciences Divisions
Lawrence Livermore National Laboratory

June 25, 1989

Dr. Francis Collins
University of Michigan Medical School
Medical Genetics
4708 Medical Sciences II
Box 0618
Ann Arbor, MI 48109

Dear Francis:

As you are aware, the HUGO Committee on Physical Mapping has appointed our subcommittee to look into some of the aspects of dealing with the distribution of arrayed libraries including possible user obligations and mechanisms for enforcing standards of scientific collaboration. I would like to extend our committee's charge to include not just arrayed libraries but all physical materials which are likely to be derived from a comprehensive genome effort. This includes cell lines, libraries, probes, and information. It seems that the same principles should be applied to all materials.

I have delayed in writing to you to begin this task in order to await discussion of some of these principles by both the NIH and DOE advisory committees to the genome effort. The two groups have met and both public and private discussions of these and related topics took place. Recommendations have been made by these committees and I will try to summarize the three salient points.

1. A source of library material for the human genome effort will likely be the cosmid and lambda chromosome-specific libraries derived from flow-sorted chromosomes. In fact, Lawrence Livermore and Los Alamos National Laboratories have constructed such libraries for about ten of the human chromosomes. They are currently being characterized and likely to be available for distribution in 2-4 months. Other laboratories (e.g. H. Lehrach) also have chromosome-specific libraries to contribute.

2. It has been suggested and anticipated that once the cosmid libraries are characterized and are acceptable, they will be distributed either as primary arrays or as amplified libraries to qualified scientific laboratories. Filters of the arrays or clones from the arrays will also be available. The extent of the distribution of the libraries will be limited by the resources available to the distributing laboratory.

3. It was also felt that some guidelines should be established for both the distributing laboratory and the recipient laboratory. These guidelines would serve two purposes. First, they would help protect the originating laboratories research interests for the amount of time and effort they used to construct and characterize the library. Second, guidelines might be helpful to insure that the recipient laboratory provides feedback information on the characteristics of the library and its clones to the rest of the scientific community. The general principle of guidelines has stimulated some debate with a few scientists of the opinion that such guidelines are not needed.

As a member of the Department of Energy Steering Committee, I was asked to draft a document that might be used to establish minimal guidelines for sharing materials derived from the human genome effort. A copy of that draft document is attached. I would solicit your opinions on this document and the suggestions contained therein. It perhaps might serve as source material for initial discussion and possible elaboration. It lacks in many specifics for which I would also solicit your opinion. When I receive your suggestions and we reach concensus, they will be added. Please consider the following issues and any others you might think appropriate.

1. Does the document provide a minimal set of acceptable guidelines for sharing materials? If not, please offer suggested changes or recommendations.

2. With regard to reference arrayed libraries, should there be any requirements imposed upon the recipients of these arrays? For example: Should users be obligated to feedback information to the distribution center for the libraries? Should the users be given freedom to re-array a library or should this be done by the distribution center? Should the users be required to disseminate clones to other users or should these come from the distribution center? Are there any additional obligations for users of such reference library materials?

3. For the reference array libraries, should a formal or informal mechanism be established to encourage cooperation or to enforce guidelines? If so, what is this mechanism? Should enforcement, if any, be the responsibility of the funding agencies or a formal scientific body? Should peer-pressure drive cooperation?

These are not easy issues and they have already stirred some emotion. However, the issues are timely since some of the chromosome-specific libraries have been arrayed and will shortly be distributed. I believe we can make a contribution to this effort, if we suggest to the scientific community a uniform policy that is also fair. This policy statement will be a recommendation from our subcommittee to HUGO.

Since HUGO does not have any funds to bring our subcommittee together, we will have to act either via the mails, FAX, or phone. I am attaching the names and addresses of our committee members. I would be grateful if you could respond to me with your FAX number when you reply to this letter. Rather than let this issue sit too long, please try to have your response to me by 15 July. This will give us time to iterate the process over the summer.

Thank you for your cooperation in this important matter and do not hesitate to call me or others should you desire to discuss these issues.

Sincerely,

Tony Carrano

Anthony V. Carrano, Ph.D.
Director, Human Genome Project
Genetics Section Leader

HUGO COMMITTEE ON PHYSICAL MAPPING
SUBCOMMITTEE ON USER OBLIGATIONS

Anthony V. Carrano
Biomedical Sciences Division, L-452
Lawrence Livermore National Laboratory
7000 East Avenue
Livermore, CA 94550
[redacted]

Francis Collins
University of Michigan Medical School
Medical Genetics
4708 Medical Sciences II
Box 0618
Ann Arbor, MI 48109
[redacted]

Jean-Louis Mandel
LGME/CNRS - U 184 INSERM
Institut de Chimie Biologique
11 rue Humann
[redacted]
France

John Sulston
MRC Laboratory of Molecular Biology
Hills Road
Cambridge CB2 2QH
England
[redacted]

SHARING GENOME INFORMATION AND BY-PRODUCTS

The human genome project will generate an extraordinary amount of data and by-products in the course of mapping and sequencing the genome of man. The products of this research will take the form of information such as restriction maps, contig maps, and DNA sequences as well as material objects such as cell lines, recombinant libraries, vectors, clones, and probes. A uniform policy should be established to protect the researcher and, at the same time, to insure rapid and reliable dissemination of information and products to qualified users. Because the shared resources have a unique personality, they will be discussed separately.

A. Cell Lines

Primary or transformed cell lines serve as source material for most mapping programs. Often a cell line from a single donor will form the basis for mapping programs in several laboratories. Cell lines representing the entire genomes of individuals take the form of fibroblast cultures or lymphoblastoid cell lines. Hybrid cell lines have been exceptionally useful for mapping panels or as starting material from which to purify individual chromosomes. Many laboratories have major programs to create appropriate cell lines for mapping. They invest considerable time and manpower in obtaining suitable donor material, selecting appropriate clones, creating useful hybrids, and providing extensive characterization at the cellular and cytogenetic level. Thus the investment made by the laboratory in creating the cells must be protected. There are a number of ways to insure this.

Recommendations

1. All cell lines should be made freely available to the general scientific community at the time a description of the cell line is published. To insure that this will occur and to demonstrate good faith on the part of the originator, it is suggested that the cell line be placed in a repository at the time of initial characterization and a hold be placed on distribution until publication at which time issuance becomes automatic. In the event no publication is forthcoming, a reasonable period for withholding the cell line is one year from the date of its establishment. To insure quality control, it is recommended that cell lines be distributed by the originating laboratory or qualified repository.
2. Investigators are strongly encouraged to release cell lines prior to publication either by a "no-strings" donation or by active collaboration with other groups. It is appropriate, in some cases, that the cell lines remain in the originating laboratory. A typical example would be hybrid mapping panels. Investigators who wish to have a probe mapped could send the probe to the mapping laboratory in return for the map data. If resources are not available for the holders of the cell lines to perform such functions, they should act as a host laboratory and invite collaborators to send a person(s) and/or supplies to perform the needed tasks.

B. Libraries

A *primary library* consists of the clones derived from the initial growth of the ligated material. An *amplified library* results from the growth of the primary library as pooled clones. Amplified libraries are likely to be less representative of the genome than the primary libraries, however, the stock of the amplified libraries will be more plentiful. To

minimize the loss of representation, an *arrayed library* is often made. To do this, individual clones are selected from the primary library and propagated as such. They are most useful for making filters for hybridization of probes to the array. An *ordered library* consists of subset(s) of one or more libraries containing clones for which relative map information is established. The ordered library will span a region of the genome and will likely have genetic, fingerprint, or restriction map information associated with it. These libraries are extremely valuable for further mapping studies and for targeting specific regions of the genome for sequencing. A *special purpose library* will consist of a unique set of clones. Examples are chromosome-specific, cDNA, linking, and jumping libraries. They usually require considerably more technical skill and effort to construct then genomic libraries.

The value of the library to the scientific community depends upon many factors such as the source of DNA, the ease of construction, the vector/cloning system, the amount of characterization available, and the representation. Creation of general purpose lambda and cosmid libraries from genomic DNA is now a fairly routine task. The use of yeast and P1 phage cloning systems for library construction is not yet routine, even for genomic DNA. Construction of chromosome-specific libraries is also not a routine procedure as it requires unique expertise in chromosome isolation and separation procedures and manipulation of small quantities of DNA. The use of flow sorting to purify chromosomes is one example of such a procedure requiring a commitment to a multidisciplinary staff and sophisticated expensive instrumentation.

Once libraries are constructed, some characterization indicating the quality of the library should be determined prior to release. Depending on the nature of the library, this could be a considerable effort. Often the originating laboratory will send aliquots of the library to qualified investigators to assist in the characterization on a collaborative

basis prior to publication and/or general release. This is to be encouraged. Often, however, casual dissemination of partially or uncharacterized libraries can have untoward consequences. If the recipient laboratory is unfamiliar with the host/vector system used or with library characterization methods, the originating laboratory may be in a position of doing more extensive work to resolve inconsistencies arising from the inexperience of the recipient laboratory. Thus the dissemination of libraries to unqualified investigators or to "collectors" is not encouraged. Criteria for defining a qualified user should include: 1) experience in handling recombinant libraries and 2) experience in some aspect of mapping, e.g. chromosomal localization using hybrid panels, *in situ* hybridization, or genetic linkage analysis.

Recommendations

1. *Primary libraries*, or aliquots thereof, may be distributed at the discretion of the originating laboratory. General distribution of primary libraries is not encouraged as these libraries represent the original unamplified source material which is generally limited in quantity. The originating laboratory is encouraged to enter into collaborative arrangements, if sufficient material exists.
2. *Amplified libraries* that are general purpose should be made available to qualified investigators when the manuscript describing the library is published. Dissemination prior to publication is encouraged and some protection can be guaranteed to the originator by requesting collaborative arrangements as a condition of release. Unpublished libraries should be made available within one year of characterization. It is desirable that the originating laboratory or its designated representative (e.g. a qualified repository) serve as the sole source for the amplified library. The amplified library should not, in general, be further distributed from the recipient's laboratory.

3. *Arrayed libraries* may be distributed at the discretion of the originating laboratory. Distribution of filters from the array is encouraged to allow others to identify clones of interest. If the originating laboratory does not have the resources to distribute the filters or perform hybridizations with probes sent to them, they should allow collaborating laboratories to provide the manpower and supplies to perform the necessary work.
4. *Ordered libraries* should be maintained in the originating or a single laboratory. Distribution of the library prior to completion is not encouraged. Because each clone will generate a wealth of map and sequence information, these libraries and the needs of the scientific community are best served if they are conservatively maintained. Individual clones from the library should, however, be made available to qualified investigators for the purposes of further characterization. Recipients of the clones, in turn, should be required to provide map or sequence information to the donor laboratory. This information is to be rapidly incorporated into a public physical map database, published, or otherwise made available to the research community. It is incumbent upon the originating laboratory to periodically disseminate the current status of the library and individual clones. Release of the minimal ordered library, or segments thereof, to qualified investigators at the time of completion is at the discretion of the investigator. The recipient laboratory is expected to: 1) have experience in library handling and clone characterization; 2) agree not to further distribute the collection; and 3) provide information to the originating laboratory on library characterization, map locations, new clones or probes, etc.
5. *Special purpose libraries* can be quite esoteric in nature and the degree to which they are distributed may be limited by the number of interested groups.

Broad dissemination is encouraged consistent with protecting the interests of the originating laboratory. They should be subject to the same principles stated above for primary and amplified libraries.

C. Vectors, Clones, and Probes

Vectors for library construction are continually improved and become more specialized. Research in vector construction moves rapidly and often the vector is outdated before the information on it is published. Vectors may be subject to patent protection and they can generate considerable commercial interest. Thus, there may be special circumstances which will tend to limit their distribution. Clones and probes derived from clones may also have some commercial value. Often clones and probes are modified by the user laboratory to make them suitable for their particular application. Wide dissemination of these materials is highly encouraged.

Recommendations

1. Vectors that are not being commercialized should be made freely available to the scientific community no later than the date of the publication describing the vector. Prior to that time, distribution by collaboration is desirable. If no publication will be forthcoming, the vector should be made available by the originator or through transfer to a central repository for distribution (e.g. ATCC) no later than one year after its characterization. If the original vector is modified in the user laboratory, it should be made available to the donor laboratory on a collaborative basis when characterized. In keeping with current practice, once the vector becomes freely available, transfer among recipient laboratories is appropriate. The original vector stock should be maintained in the originating laboratory or a repository.

2. Clones should be made available no later than the time that they are described in a publication or, in the absence of publication, within one year of characterization. In the interim, collaborations should be established.
3. Probes should be also made available no later than the time that they are described in a publication or, in the absence of publication, within one year of characterization. In the interim, collaborations should be established.

D. Map and Sequence Data

If the human genome project is to succeed, it must have the cooperation of all investigators in releasing and sharing data. Knowledge of the latest results should help minimize redundancy and accelerate research on important regions of the genome. Moreover, discrepancies in map and sequence data will undoubtedly occur and these will have to be resolved. Only by interchange and peer-examination of data can this cross-checking occur. The timing of data release and the nature of the released data will influence progress on the genome.

Recommendations

1. Both the complete processed and raw data should be made available to the scientific community no later than the time of publication or when the data are entered into a public domain database. Selected release of information prior to this time is strongly encouraged. When data are released prior to the publication, it is appropriate for the originating investigator to request assurance from the privileged recipient that he/she will not publish a derivative study prior to the publication of the originating investigator. Failure of the recipient investigator to adhere to this assurance should be subject to review and possible censure.

2. Should the originating investigator fail to publish map or sequence information within a year of collection, publication of similar and derivative data obtained in the laboratory of a privileged recipient is warranted.

Sharing information and the products of research is best governed by a sense of fairness and spirit of cooperation. Yet, scientists recognize that future funding is contingent upon progress and, in the course of intense research, may tend to overlook the ethics of mutual interaction. Unless funding agencies enforce policy, the scientific community must act as its own regulator. There are and will continue to be exceptions to the policies and it is not possible to address each one herein. Resolution of the exceptions is dependent upon the continued good intentions of the scientific community.

Suggested Procedures for Sharing Products of the Human Genome Effort

Product	Release Required	Release Encouraged
Cell lines	At the time of publication or one year after establishment	Anytime by collaboration
Primary libraries	At the discretion of the investigator	Anytime by collaboration
Amplified libraries	At time of publication, transmittal to ATCC, or within one year of characterization	Anytime by collaboration
Arrayed libraries	At the discretion of the investigator	Distribution of filters with DNA from array is recommended
Ordered libraries	Release after completion at the discretion of the investigator	Clones should be made available
Special purpose libraries	At time of publication or within one year of characterization	Anytime by collaboration
Vectors, clones, and probes	At time of publication or within one year of characterization	Anytime by collaboration
Data, raw and processed	At time of publication or entry into a public domain database	Anytime by agreement

College of Physicians & Surgeons of Columbia University | *New York, N.Y. 10032*

DEPARTMENT OF GENETICS AND DEVELOPMENT 701 West 168th Street

TELEPHONE [redacted] TELEFAX [redacted]

May 16, 1989

Dr. Francis Collins
Division of Medical Genetics
Dept. of Medicine
University of Michigan Medical School
4708 Medical Sciences II, Box 0618
Ann Arbor, MI 48109

VIA FAX [redacted]

Dear Francis:

I enclose a draft of the minutes of our meeting at Cold Spring Harbor. Please contact me at once if you feel the general thought of it is at odds with our actual discussion and conclusions.

Sincerely yours,

Charles R. Cantor
Higgins Professor and Chairman

CRC/ah
Attachment (3 pp.)

Minutes of the First Meeting of the HUGO Committee on Physical Mapping
Cold Spring Harbor Laboratory
April 28, 1989

Attendees: Charles Cantor (chair), Walter Bodmer, Anthony Carrano, Daniel Cohen, Francis Collins, Hans Lehrach, Jean-Louis Mandel, Maynard Olson, Nobuyoshi Shimizu, James Watson (absent: Sydney Brenner, Tom Caskey)

Those present had a free-ranging discussion about the prospects for using shared common samples in human physical mapping and also heard a brief summary of a meeting held the previous day by R. Waterston, A. Carrano, M. Olson, C. Smith, E. Hildebrandt, P. de Jong and others which concluded that it was not certain that C. elegans was a useful model for organization of human mapping. Based on the discussions there was a general consensus on the following issues:

1. It would be highly desirable that at least some cell lines be commonly available for mapping to permit comparisons between different efforts but that this does not appear to pose any current problems.

2. Both single chromosome and genomic libraries need to be commonly available to participants in the human genome project.

3. Procedures need to be developed to assure quality control for libraries that will form the common sample pool.

4. A plan is needed to establish ordered arrays of these libraries.

5. Rules need to be established governing the obligations and responsibilities of users of these arrayed libraries, and mechanisms for enforcement of these rules need to be devised.

6. Over the next few years parallel physical mapping efforts on individual chromosomes need to be encouraged to coalesce into physical

mapping committees which could coordinate efforts, share data, and coordinate with appropriate gene mapping committees.

7. A plan is needed on how to create, structure and support a limited network of sample distribution centers to optimize use of common, arrayed libraries.

8. It is appropriate that fees for service defray much of the cost of handling arrayed libraries.

9. The role of HUGO should be to provide the horizontal communication in all of this. The cost of HUGO's role and the role of specific sample distribution centers and mapping committees should be shared internationally.

10. Subcommittees should be established immediately to draw up drafts of specific proposals and policies. These should communicate or meet as appropriate during the summer.

11. A fall meeting should be held of the committee as a whole to integrate the efforts of the separate subcommittees, perhaps in Berkeley with NIH and DOE asked to share the costs.

Initial members of the subcommittees are proposed as follows. Their membership can be expanded as needed with other Committee members or outsiders.

Sample Accession (principally items 3 and 4 above):
- H. Lehrach (chair)
- G. Evans
- P. de Jong
- M. Olson

User Obligations (principally item 5 above):
- A. Carrano (chair)
- F. Collins
- J.-L. Mandel
- J. Sulston

Distribution Network (principally items 7 and 8 above):

D. Cohen (chair)
W. Bodmer
D. Maglott
N. Shimizu

In addition all members of the Committee and subcommittees are encouraged to think further about what steps should be taken, and when such steps should be taken, to encourage the formation of chromosome committees.

CRC/ah