# Foreword to the Report of the NIH Model Organism Database Workshop

On behalf of the National Institutes of Health (NIH), the National Heart, Lung, and Blood Institute (NHLBI) and the National Human Genome Research Institute (NHGRI) convened a Model Organism Database Workshop. This workshop was designed to provide the NIH with advice and concrete recommendations on reviewing, implementing, and organizing genomic, genetic, and phenotypic databases for model organisms.

The rate at which genomic and genetic information is emerging is increasing dramatically. One consequence of this information explosion is that practicing scientists are finding it difficult to keep up with the ever-expanding literature in the areas germane to their research programs. This difficulty is exacerbated by the realization that most biological phenomena are ancient and highly conserved over long evolutionary distances and, therefore, that an understanding of these phenomena requires a perspective on related information across large taxonomic territories.

Practicing scientists are becoming increasingly dependent on biological databases as key resources in identifying information of importance to their research efforts. Numerous databases are necessary to support the various needs of the scientific community. These databases fall into several classes.

One key class that is receiving a great deal of attention is the model organism databases. Model organisms are an important tool for understanding and dissecting human disease and biological process. An increasing amount of genetic, genomic, and phenotypic information is being generated on a variety of model organism systems. Databases have been established to disseminate this information to the scientific community. Currently, there are a number of different ways to create databases, maintain them, and establish community input.

The NIH is looking for guidance in initiating a coherent plan to establish, maintain, interconnect, and evaluate these public resources. The goals of this workshop were to contribute to the development of this plan by assessment of the kinds of data that the databases should incorporate, evaluation of data acquisition strategies, identification of means of community input, establishment of review criteria for new and existing database projects, and consideration of mechanisms to develop rich sets of interconnections among these databases at a variety of biological levels.

The report from this workshop, held December 7–8, 1998, at the Lansdowne Conference Center in Lansdowne, Virginia, can be found on the NHLBI Web site at **http://www.nhlbi.nih.gov/nhlbi/sciinf/sciinf2.htm** and on the NHGRI Web site at **http://www.nhgri.nih.gov/About_NHGRI/Der/modelorg_data_workshop.htm.**

# Executive Summary of the NIH Model Organism Database Workshop

Lansdowne Conference Center
Lansdowne, Virginia

December 7–8, 1998

---

Research on model organisms advances the understanding of biological processes and disease states. The NIH Model Organism Database Workshop was convened to provide recommendations on reviewing and implementing genomic, genetic, and phenotypic database projects that support and provide access to research on model organisms. Several themes emerged:

- **Biology-driven**: The driving force of a model organism database (MOD) is the biology, not the technology. The database should allow the user to ask critical biological questions. Thus, data content, organization, currency, and accuracy are paramount. The MOD database platform should be a solid, robust product that can support the data, not a product still in an experimental stage.

- **Breadth of information**: MODs must contain or link to information ranging from molecular structure and function to phenotype, with both high-throughput and community-generated information. Much genetic and phenotypic information is less structured and, thus, more labor-intensive and expensive to collect and annotate than genomic data; however, it is essential that it be included.

- **Curation:** MODs are value-added databases; their primary function is to integrate data from disparate sources, connecting related data and enhancing its contextual information. Expert domain knowledge is essential for high-quality data capture, and this will typically require that curators have Ph.D.-level research backgrounds.

- **User access:** MODs must recognize the hardware, software, and network capabilities of the community they serve. In addition, it is essential that end users receive the funding to keep up with the rapidly advancing technology needed to access MODs.

- **Communities:** MODs must have deep ties to their organismal communities, supporting their multiple roles as data consumers, data curators, and data providers. MODs must also be able to support inquiries from broader scientific communities.

- **Advisory groups:** External advisory groups should provide regular advice to the MOD, the research communities, and the funding agencies. Advisory groups should include MOD users, software specialists, and representatives from other MODs.

- **Priorities:** MODs must set clear priorities for data content, data capture, organization, curation, annotation, navigation, and presentation.

- **Database connections:** MODs must establish and maintain effective cross-links among themselves as well as with other types of databases. These links enable synthetic data analysis and permit databases to share information without redundant data management.

- **Leveraging existing projects:** New MODs should consider affiliating with existing MODs. Existing software should be reused or adapted when possible.

- **Research:** Bioinformatics research is important to support the infrastructural role of the MODs. MOD projects should perform informatics research and training.

# Report of the
# NIH Model Organism Database Workshop

Lansdowne Conference Center
Lansdowne, Virginia

December 7–8, 1998

## Introduction

Databases are vital for research in biology and medicine. Databases serve many roles, including the capture and organization of key information, integration of data from disparate sources, and facilitation of the formulation of new hypotheses and new perspectives. Research communities are facing many challenges, including a flood of new data, the rapid growth in data diversity, and the complexity of data produced by cross-disciplinary investigation. Robust and highly interconnected databases are essential to address these scientific challenges and to capitalize on new research opportunities. Insufficient support or ineffective implementation of model organism databases (MODs) will slow the pace and increase the cost of biological discovery.

### *Databases From the Perspective of Model Organism Research*

In the last 100 years, research on a handful of organisms has played a profound role in advancing our understanding of the biological and biomedical sciences. The need to capture, organize, and access data from these model organisms has driven the creation of organism-specific databases. These model organism databases have allowed researchers to sift through masses of data, to gain access to information or materials they might have missed, and to go in new research directions. Comparative analysis has proven to be valuable in increasing our understanding of biological processes, including those in humans. Because these MODs are of immense value, offer tremendous opportunities, and represent a significant fiscal investment, it is timely to examine issues pertaining to the establishment, maintenance, evaluation, and future directions of model organism databases. Thus, the NIH convened the Model Organism Database Workshop, which brought together an international group representing developers and users of established databases, investigators interested in developing new databases, and funding agencies. The goals were to assess the range of data that MODs capture, evaluate data acquisition strategies, identify means of community input and support, establish review criteria for new and existing database projects, and consider mechanisms to support coordinated efforts. It is mutually beneficial to all MODs that each of them is successful.

### *Recommendations of the Workshop*

This report addresses MODs as research resources. We outline the salient features toward which the MODs, individually and in concert, should strive. We do not have the knowledge to preordain a “one size fits all” database project plan. We can nevertheless state the general goals, exemplify some ways that MODs might achieve these goals, and consider how these goals should be

translated into review criteria for scientific and administrative evaluation of MODs. The report also addresses additional initiatives needed to support the MODs and ensure that the broad U.S. biomedical research community has access to them.

### *The MODs as Research Resources*

MODs deal with two sets of research communities, with different needs and expectations:

- **Model organism community**: This community provides the data to a MOD, adds value by contributing to the curation of the data, and comprises a major set of users who need access to a great deal of specialized information, such as strain collections.

- **General research community:** This community uses but does not directly contribute information to MODs. Unlike the model organism community, the general community does not usually understand the specialized jargon and nomenclature for a model organism. The MODs should provide accessible summaries of genomic, functional, and phenotypic information in addition to full access to the underlying datasets.

## The Model Organism Databases: A Life Cycle Perspective

Database projects have different needs and goals at different points in their life cycles. The overarching goal should always be to meet the needs of the research communities.

### *Some Features Common Throughout the Life Cycle*

- Tools to facilitate data submission should be developed or imported. Both human interface and automated machine-readable submission tools are needed.

- Where appropriate, raw data should be captured so that they can be reanalyzed. Because of the expense of capturing raw data, there must be a balance between taking in raw and summarized data.

- Curation is demanding and requires a high level of domain expertise. Ph.D.-level curators are needed, typically with research experience in the particular experimental system.

- Continual development of tools that support queries and graphical summaries of large data sets is important. Query tools and graphical viewers should address the needs of the general and the expert user communities, balancing ease of use with depth of information.

- Controlled vocabularies and standardized nomenclatures should be developed and implemented to support database organization and querying. The levels of controlled vocabularies and free text should be established and periodically reevaluated.

- Timely and effective user support is essential in maintaining good relations with the community.

- The MOD data presentation represents only one view of the biological world. Hence, the MODs should provide third parties with readily ported access to their entire data sets so that the information may be viewed in other ways.

- Database objects such as genes must have unique permanent identifier numbers in order to provide stable links, track changes in the names of the objects, and maintain synonym lists.

- Each MOD should establish extensive cross-links to other MODs and other types of relevant databases through the exchange of linked lists of objects and their identifiers.

- Each MOD should collaborate with other MODs and relevant databases to develop and share improved technologies, methods, and controlled vocabularies.

- The MODs should provide gene lists with Medline identifiers so that Medline curators can build the links to model organism genes reported in publications. This facilitates Medline-MOD links for users and aids the identification and parsing of the model organism literature.

- The MODs should encourage journals to develop mechanisms to promote MOD user submissions and to incorporate MOD object identifier numbers as well as valid names.

***Guiding Principles During the Establishment Phase***

When does an organism warrant its own database? Although it is difficult to come up with definitive answers, important criteria include the following:

- The experimental system really is a *model* system, which means that it is important for studying some biological processes or human health issues.

- The information should be rich enough to be the object of higher levels of analysis or of analysis not available in the primary literature.

- The community has an accepted system for nomenclature and a gene registry.

- The value-added data of the MOD should be of interest to both the organismal community and the general research community.

Once the need for a MOD is established, some priorities during the establishment phase are as follows:

- Particularly in the early phases of the project, there may be much to be gained by piggybacking on the software and technical expertise of existing MODs. Expanding an existing MOD or affiliating with one should be considered first. Highly portable database software could be considered next. Alternatively, shared data structures, schemas, and tools would enable software engineers to build rapidly on other database platforms. This would permit the new MOD to focus on issues of data curation while gaining a better, "field tested" view of the needs of its community and would promote the cost-effectiveness of the project.

- As considerable expertise, both technical and strategic, is available within the existing MODs, they should play a mentoring role in facilitating the establishment of new MOD projects. Hence, ways should be sought to provide interactions among existing and embryonic MODs. The planners of new MODs may wish to contact NIH staff early in the planning stage of a new database project; these staff members are knowledgeable about the existing projects and can facilitate the necessary contacts. Travel funds should be provided to the existing MODs for visitors' programs. Participation of individuals from the new MODs at periodic meetings of the MOD groups would also facilitate interaction. The

availability of a comprehensive WWW site describing MOD sites would be of considerable help for new MOD groups.

- The most essential needs of the model organism community should be addressed first. This is crucial to get those researchers who will be both providers and major users of the data to identify themselves as the "partners" in the MOD. The community needs should be assessed through a combination of advisors and directed surveys. Advisory committees should be established so that they are independent critics and intermediaries to the research community.

- Establishing a new database enterprise is a long-term and complex commitment and should be implemented in steps. This allows the MOD to address the initial organizational and logistical issues within the context of a reasonable set of production goals. MODs and funding agencies need to plan for the stepwise ramp-up in responsibilities and funding.

- Priority should be given to genomic and genetic data, and then more complex phenotypic data classes can be addressed. Phenotypic and expression pattern data should be treated as attributes of genomic/genetic data objects when possible. It should be recognized that much genetic and phenotypic data may be more expensive to collect than genomic data, but they are still essential for the scientific community.

### *Guiding Principles During the Maintenance Phase*

Many of the features inherent in the establishment phase continue to be important as the MOD matures, and there are additional responsibilities:

- With regular input from the advisory groups and others in the communities, the MOD should reevaluate its priorities, policies, and procedures with an eye towards maintaining a modern and effective resource that supports the rapidly advancing science in the model organism.

- Bioinformatics is a rapidly evolving field. Each MOD needs a budget for developing innovative solutions to problems or to migrate to new platforms while maintaining the daily operation of the database project.

- The MOD must address the needs of the general research community as well as the specialized organism community. Doing so requires outreach to the broader community and is also likely to involve alternative data views without jargon. Such outreach might include demonstrations at a range of scientific meetings and live on-line classes for scientists at their home institutions.

### *On the Reproductive, Senescence, and Death Phases of the MOD Life Cycle*

MODs are not static or immortal. Over time even a successful MOD may find it efficient to transfer some types of information to a central database, or it may become so large and cumbersome that it proves necessary to divide it into smaller projects. MODs have to be able to change as needed.

MODs are complicated projects and can run into difficulties for many reasons. The Human Genome Database (GDB) example shows that early recognition of such difficulties and early intervention is preferable to allowing the MOD to undergo

a lingering death. The workshop did not come to any explicit conclusions on how to achieve early detection and therapy, but part of the answer is to encourage critical and constructive review by the external advisory committees. Another part may be mechanisms that encourage interaction among the existing databases, such as periodic MOD workshops or visitor programs. A workshop would allow the database providers to talk freely about problems as well as solutions, which is essential for improving MOD projects through cooperative efforts.

Computer experts are in great demand. This makes MODs vulnerable to premature demise through the loss of key bioinformatics people. Affiliating new MODs with existing groups permits the technical groups to grow in size and therefore be less sensitive to staff loss.

**Guiding Principles for the Review and Funding of MODs**

- Each MOD is a critical research resource, which has important implications for evaluation.

- From the early stage of project development, MOD applicants should work with their advisory groups, other MOD projects, and NIH program staff to prevent as many pitfalls as possible in developing credible database applications.

- The initial review group must be put together carefully and must receive considerable education about the individual MOD. Representation of the specific and general research communities is essential, possibly including some external advisors and some officers of governing bodies that exist for some model organism communities. Other reviewers need to have the necessary computational or database project management expertise. The goal of the review committee education process is to ensure that, regardless of the funding mechanism, these grant applications are reviewed as research resources.

- Review criteria should be well established and understood consistently by both reviewers and applicants. As with any grant application, a complex mixture of positives and negatives must be distilled into a priority score and budget recommendations. Although the features of new and ongoing MODs were listed above as suggestions to provide flexibility and encourage innovation, the applicant must demonstrate that the goals of the review criteria have been met.

***Specific Review Criteria***

Documentation should be provided to demonstrate the following:

- The MOD is addressing critical needs of the model organism and general communities.

- The value added to the data for the primary community.

- The results of community surveys.

- The effective composition and use of external advisory committees.

- The effectiveness of user support.

- The outreach and education efforts to the user communities.

- Data on WWW database hits, by a method that NIH staff and MODs should

establish. Although these data have some problems, trends of hit frequency over time are informative.

- Interactions with other MODs and database groups.

- How the MOD has achieved cost-effectiveness and evaluated technology and software. Choices for the more expensive of alternative approaches must be carefully justified.

- The effectiveness of the curation models.

- How the appropriate data object relationships are represented in the MOD, and the types of queries that the database supports.

- Database performance, ease and transparency of use, interface design, documentation, and data access.

- How the MOD supports the advancement of science relating to the data it contains, and how it will respond to scientific advances.

**Funding Considerations**

- The workshop considered that the MODs and other database projects are substantially underfunded, using conservative figures of industry funding distributions (10 to 15 percent of research budget in informatics) and considering the amount of support for hypothesis-driven research in a model system. Budget increases for effective database support are essential for maintaining an outstanding publicly funded research enterprise.

- In general, established databases with strong track records should be on 5-year funding cycles, whereas those in flux require more frequent review. In both cases, periodic (typically annual) administrative review would be valuable, such as program officer visits to the MOD sites or attendance at external advisory committee meetings.

**Additional Recommendations**

- Many aspects of database development and implementation are still experimental. Funding independent research projects addressing these issues is important to support the MODs. These projects might focus on important areas, such as the development of functional ontologies or the production of reusable and readily portable software modules for data acquisition, maintenance, analysis, or display.

- There is serious concern that the capabilities offered by the MODs will outstrip the ability of users to take advantage of them. The difficulty of obtaining NIH research grant funding for computer hardware is completely at odds with the need for effective informatics infrastructure and should be resolved. The other potential bottleneck in delivering informatics support is network speed. Universal high-speed networks will be essential for transporting data sets and display tools across the WWW.

- The need for increased training in bioinformatics at all levels is well recognized, and the workshop encourages efforts to support such training. The MODs are important training sites in such programs, and affiliation of MODs with such programs should be fostered.

## Concluding Remarks

A great deal of important information exchange and consensus occurred during this effective workshop. The discussions were consistently frank and constructive. Nonetheless, there were many topics that this workshop could not do justice to in the constrained time, such as how various MODs should interact and coordinate, which data types should be provided to users from the nonorganismal community, and how curation should be done. Future workshops bringing together database providers, users, and NIH staff should be strongly encouraged. Other mechanisms for encouraging scientific interaction and collaboration among the database providers should also be considered.

# Agenda of the
# NIH Model Organism Database Workshop

Lansdowne Conference Center
Lansdowne, Virginia

December 7–8, 1998

Co-Chairs
William M. Gelbart, Harvard University
Joe Nadeau, Case Western Reserve University School of Medicine

---

**Monday, December 7**

8:30 a.m. Registration

***Plenary Session***

9:00 a.m. **Welcome and Charge**
Lisa Brooks and Stephen Mockrin

**Introduction**
Joe Nadeau

9:15 a.m. **Database Developers**

Presenters will review what worked, what did not, how diverse needs were balanced, and what they would do differently.

| | | |
|---|---|---|
| 9:15 a.m. | Yeast | David Botstein |
| 9:30 a.m. | Arabidopsis | Michael Cherry |
| 9:45 a.m. | Fly | William Gelbart |
| 10:00 a.m. | Worm | Richard Durbin |

10:15 a.m. **Break**

10:45 a.m. **Database Developers (continued)**

| | | |
|---|---|---|
| 10:45 a.m. | Mouse | Janan Eppig |
| 11:00 a.m. | GDB | Stan Letovsky |
| 11:15 a.m. | NCBI | James Ostell |
| 11:30 a.m. | Zebrafish | Monte Westerfield |
| 11:45 a.m. | Rat | Peter Tonellato |

**Monday, December 7 (continued)**

12 p.m. **Lunch**

1:00 p.m. **Group Discussion of the Questions**

Discussion leaders: Peter Cartwright and Howard Jacob

Group 1: Communities
Group 2: Information

3:00 p.m. **Break**

3:30 p.m. **Resume Group Discussion of the Questions**

Discussion leaders: Janan Eppig and Monte Westerfield

Group 3: Access
Group 4: Startup

5:50 p.m. **Recess**

6:30 p.m. **Dinner**

7:30 p.m. **Breakout Groups Convene to Produce One-Page Bullet Text**

Discussion leaders will facilitate dialogue and recommendations.

| | |
|---|---|
| Group 1—Communities | Peter Cartwright |
| Group 2—Information | Howard Jacob |
| Group 3—Access | Janan Eppig |
| Group 4—Startup | Monte Westerfield |

9:00 p.m. **Breakout Group Leaders Produce One-Page Document**

**Tuesday, December 8, 1998**

9:00 a.m. **Breakout Group Reports**

| | | |
|---|---|---|
| 9:00 a.m. | Group 1—Communities | Peter Cartwright |
| 9:45 a.m. | Group 2—Information | Howard Jacob |

10:15 a.m. **Break**

**Tuesday, December 8, 1998 (continued)**

| | | |
|---|---|---|
| 10:30 a.m. | **Breakout Group Reports (continued)** | |
| 10:30 a.m. | Group 3—Access | Janan Eppig |
| 11:15 a.m. | Group 4—Startup | Monte Westerfield |
| 12:00 p.m. | **Break** | |
| 12:15 p.m. | **General Discussion/Conclusion** | |
| | Discussion leaders: William Gelbart and Joe Nadeau | |
| 1:00 p.m. | **Lunch** | |
| 2:00 p.m. | **Summary Report Writing Session** | |
| | Workshop Chairs and Breakout Group leaders only | |

## Roster of the
## NIH Model Organism Database Workshop


Michael Ashburner, Ph.D.
Professor
European Molecular Biology Laboratory
European Bioinformatics Institute
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 15D
UNITED KINGDOM
44 1223 494412
44 1223 494470 FAX
ashburner@ebi.ac.uk

Ivan C. Baines, Ph.D.
Health Scientist Administrator
National Heart, Lung, and Blood Institute
National Institutes of Health
10218 Oldfield Drive
Kensington, MD 20895
(301) 897-0638
(301) 480-3541 FAX
ib9g@nih.gov

David Botstein, Ph.D.
Professor and Chairman
Department of Genetics
Stanford University
Lane Building, Room L329
300 Pasteur Drive
Stanford, CA 94305-5120
(650) 723-3488
(650) 723-7016 FAX
botstein@genome.stanford.edu

Josephine P. Briggs, M.D.
Director
Division of Kidney, Urologic and
Hematologic Diseases
National Institute of Diabetes and Digestive
and Kidney Diseases
National Institutes of Health
Building 31, Room 9A-17
31 Center Drive, Mailstop 2560
Bethesda, MD 20892
(301) 496-6325
(301) 402-4874 FAX
briggsj@hq.niddk.nih.gov

Lisa Brooks, Ph.D.
Program Officer
Genetic Variation Program
Genome Informatics Program
National Human Genome Research Institute
National Institutes of Health
Building 38, Room 614
38 Library Drive, Mailstop 6050
Bethesda, MD 20892-6050
(301) 496-7531
(301) 480-2770 FAX
lisa_brooks@nih.gov

Jill Carrington, Ph.D.
Health Scientist Administrator
National Center for Research Resources
National Institutes of Health
One Rockledge Center
6705 Rockledge Drive
Bethesda, MD 20892-7965
(301) 435-0744
(301) 480-3819 FAX
jillc@ncrr.nih.gov

Samuel Cartinhour, Ph.D.
Department of Plant Breeding
Cornell University
U.S. Department of Agriculture
Bradfield Hall
Ithaca, NY 14853
(607) 255-8091
(607) 255-6683 FAX
scartinh@greengenes.cit.cornell.edu

Peter Cartwright
President and Chief Executive Officer
Cimarron Software, Inc.
175 Southwest Temple, Suite 530
Salt Lake City, UT 84101
(801) 521-3210
(801) 521-3111 FAX
pc@cimsoft.com

Marvin Cassman, Ph.D.
Director
National Institute of General Medical Sciences
National Institutes of Health
Natcher Building, Room 2An.12
45 Center Drive, Mailstop 6200
Bethesda, MD 20892
(301) 594-2172
cassmanm@nih.gov

J. Michael Cherry, Ph.D.
Head
Genetics Database Group
Department of Genetics
Stanford University
Alway Building, Room M341
Stanford, CA 94305-5120
(650) 723-7541
(650) 723-7016 FAX
cherry@genome.stanford.edu

Hemin Chin, Ph.D.
Chief
Genetic Basis of Neural Function Program
Division of Basic and Clinical Neuroscience Research
National Institute of Mental Health
National Institutes of Health
5600 Fishers Lane, Room 10C-26
Rockville, MD 20857
(301) 443-1706
(301) 443-9890 FAX
hemin@nih.gov

Peter D'Eustachio, Ph.D.
Associate Professor
Department of Biochemistry
School of Medicine
New York University
550 First Avenue
New York, NY 10016
(212) 263-5779
(212) 263-8166 FAX
deustp01@mcrcr0.med.nyu.edu

Leslie K. Derr, Ph.D.
Office of Science Policy
National Cancer Institute
National Institutes of Health
Building 31, Room 11A03
31 Center Drive, Mailstop 2590
Bethesda, MD 20892-2590
(301) 496-1458
(301) 496-7807 FAX
derrl@osp.nci.nih.gov

Machi F. Dilworth
Division Director
Division of Biological Infrastructure
National Science Foundation
4201 Wilson Boulevard, Room 615 N
Arlington, VA 22230
(703) 306-1470
(703) 306-0356 FAX
mdilwort@nsf.gov

Daniel W. Drell, Ph.D.
Biologist
Life Sciences Division
Office of Biological and Environmental Research
U.S. Department of Energy
Building GTN, Room G-147
19901 Germantown Road
Germantown, MD 20974
(301) 903-4742
(301) 903-8521 FAX
daniel.drell@science.doe.gov

Richard Durbin, Ph.D.
Head
Informatics Division
The Sanger Centre
Hinxton Hall
Wellcome Trust Genome Campus
Hinxton, Cambridge CB10 15A
UNITED KINGDOM
44 01223 834244
44 01223 494919 FAX
rd@sanger.ac.uk

Geoffrey M. Duyk, M.D., Ph.D.
Chief Scientific Officer
Exelixis Pharmaceuticals, Inc.
260 Littlefield Avenue
South San Francisco, CA 94080
(650) 825-2218
(650) 825-2205 FAX
duyk@exelixis.com

Joe Ecker, Ph.D.
Professor
Department of Biology
University of Pennsylvania
415 South University Avenue
Philadelphia, PA 19104-6018
(215) 898-9384
(215) 898-8780 FAX
jecker@atgenome.bio.upenn.edu

Janan T. Eppig, Ph.D.
Staff Scientist
Mouse Genomics Laboratory
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609-1500
(207) 288-6422
(207) 288-6433 FAX
jte@jax.org

Jonathan Epstein, M.S.
Computer Scientist
Office of Scientific Research
National Institute of Child Health and Development
National Institutes of Health
Building 31, Room 2A31
31 Center Drive
Bethesda, MD 20894
(301) 402-4563
(301) 480-0655 FAX
jepstein@helix.nih.gov

Ken Fasman, Ph.D.
Vice President, R&D Informatics
AstraZeneca R&D Boston
35 Gatehouse Drive
Waltham, MA 02451
(781) 839-4518
(781) 839-4500 FAX
ken.fasman@astrazeneca.com

Elise Feingold, Ph.D.
Program Director
Genome Analysis
National Human Genome Research Institute
National Institutes of Health
Building 38A, Room 617
38 Library Drive, Mailstop 605
Bethesda, MD 20892-6050
(301) 496-7531
(301) 480-2770 FAX
elise_feingold@nih.gov

Mark Fishman, M.D.
Chief of Cardiology
Director of Cardiovascular Research Center
Massachusetts General Hospital
Bigelow 840
55 Fruit Street
Boston, MA 02114
(617) 726-3738
(617) 724-9564 FAX
fishman@cure.mgh.harvard.edu

Rainer Fuchs, Ph.D.
Vice President
ARIAD Pharmaceuticals, Inc.
26 Landsdowne Street
Cambridge, MA 02139-4234
(617) 494-0400
(617) 252-0851 FAX
rainer.fuchs@ariad.com

Jim Garrels, Ph.D.
President and Chief Executive Officer
Proteome, Inc.
100 Cummings Center, Suite 435M
Beverly, MA 01915
(978) 922-1643
(978) 922-3971 FAX
jg@proteome.com

William M. Gelbart, Ph.D.
Professor
Department of Molecular and Cellular Biology
The Biological Laboratories
Harvard University
16 Divinity Avenue
Cambridge, MA 02138-2020
(617) 495-2906
(617) 496-1354 FAX
gelbart@morgan.harvard.edu

Thomas J. Gill, III, M.D.
Menten Professor of Experimental Pathology
University of Pittsburgh School of Medicine
64 Pond Road
Duxbury, MA 02332
(781) 934-7348
(781) 934-7130 FAX
gilliii@massmed.org

Paul Gilna, Ph.D.
Program Director
Database and Computational Biology Activities
Division of Biological Infrastructure
National Science Foundation
4201 Wilson Boulevard, Room 615
Arlington, VA 22230
(703) 306-1470 EXT 6410
(703) 306-0356 FAX
pgilna@nsf.gov

Judith Greenberg, Ph.D.
Director
Division of Genetics and Developmental Biology
National Institute of General Medical Sciences
National Institutes of Health
45 Center Drive, Mailstop 6200
Bethesda, MD 20892
(301) 594-0943
(301) 480-2228 FAX
greenbej@nigms.nih.gov

Pascal Haffter, Ph.D.
Max-Planck-Institut für Entwicklungsbiologie
Spemannstrasse 35/III
Tuebingen 72076
GERMANY
49 7071 601 443
49 7071 601 384 FAX
pascal.haffter@tuebingen.mpg.de

Marnie E. Halpern, Ph.D.
Staff Scientist
Department of Embryology
Carnegie Institution of Washington
115 West University Parkway
Baltimore, MD 21210
(410) 554-1218
(410) 243-6311 FAX
halpern@mail1.ciwemb.edu

Jonathan Hodgkin, Ph.D.
Permanent Scientific Research Staff
MRC Laboratory of Molecular Biology
Hills Road
Cambridge CB2 2QH
UNITED KINGDOM
44 1223 402321
44 1223 412142 FAX
jah@mrc-lmb.cam.ac.uk

Philip M. Iannaccone, M.D., Ph.D.
George M. Eisenberg Professor
Department of Pediatrics
Children's Memorial Institute for
Education Research
Northwestern University Medical School
2300 Children's Plaza, Mailstop 204
Chicago, IL 60614
(773) 880-8236
(773) 880-8266 FAX
pmi@nwu.edu

Ian J. Jackson, M.A., Ph.D.
Senior Scientist
MRC Human Genetics Unit
Western General Hospital
Crewe Road
Edinburgh EH4 2XU
SCOTLAND
44 0131 467 8409
44 0131 343 2620 FAX
ian.jackson@hgu.mrc.ac.uk

Howard J. Jacob, Ph.D.
Associate Professor
Laboratory of Genetics Research
Department of Physiology
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI 53213
(414) 456-4887
(414) 456-6516 FAX
jacob@mcw.edu

Elke Jordan, Ph.D.
Deputy Director
National Human Genome Research Institute
National Institutes of Health
Building 31, Room 4B09
Bethesda, MD 20892
(301) 496-0844
elkej@mail.nih.gov

Gary H. Karpen, Ph.D.
Associate Professor
The Salk Institute
10010 North Torrey Pines Road
La Jolla, CA 92037
(619) 453-4100, ext. 1473
(619) 622-0417 FAX
karpen@salk.edu

Martin E. Kreitman, Ph.D.
Professor
Department of Ecology and Evolution
University of Chicago
1101 East 57th Street
Chicago, IL 60637
(773) 702-1222
(773) 702-9740 FAX
mkre@midway.uchicago.edu

Stanley I. Letovsky, M.S., Ph.D.
Director of Bioinformatics
Cereon Genomics, LLC.
270 Albany Street
Cambridge, MA 02139
(617) 551-1731
(617) 551-1814 FAX
Stanley.I.Letovsky@cereon.com

Goran Levan, Ph.D.
Professor of Genetics
Lundberg Laboratory
Department of Cell and Molecular Biology-Genetics
Göteborg University
Medicinareg.9C, Box 462
Göthenburg SE 405 30
SWEDEN
46-31-773-3290
46-31-826-286 FAX
Goran.Levan@gen.gu.se

David Lipman, M.D.
Director
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
Building 38A, Room 8N805
8600 Rockville Pike
Bethesda, MD 20894
(301) 496-2475
(301) 480-9241 FAX
lipman@ncbi.nlm.nih.gov

Cheryl L. Marks, Ph.D.
Program Director
Cancer Genetics Branch
Division of Cancer Biology
National Cancer Institute
National Institutes of Health
Executive Plaza North, Room 501
6130 Executive Boulevard
Bethesda, MD 20892-7381
(301) 435-5226
(301) 496-8656 FAX
cm74v@nih.gov

Victoria McGovern, Ph.D.
Program Officer
Burroughs Wellcome Fund
4709 Creekstone Drive
Durham, NC 27703
(919) 991-5112
(919) 941-5884 FAX
vmcgovern@bwfund.org

Jean McKay, M.L.S.
Office of Science Policy
National Cancer Institute
National Institutes of Health
Building 31, Room 501
31 Center Drive
Bethesda, MD 20892
(301) 496-1458
(301) 496-7807 FAX
jm279v@nih.gov

Victor McKusick, M.D.
University Professor of Medical Genetics
Center for Medical Genetics
Johns Hopkins University School of Medicine
Blalok 1007
600 North Wolfe Street
Baltimore, MD 21287-4922
(410) 955-6641
(410) 955-4999 FAX
mckusick@peas.welch.jhu.edu

Stephen C. Mockrin, Ph.D.
Deputy Director
Division of Heart and Vascular Diseases
National Heart, Lung, and Blood Institute
National Institutes of Health
Rockledge II, Room 9170
6701 Rockledge Drive, Mailstop 7940
Bethesda, MD 20892-7940
(301) 435-0477
(301) 480-1336 FAX
mockrins@gwgate.nhlbi.nih.gov

Joe Nadeau, Ph.D.
Professor
Genetics Department BRB-630
Case Western Reserve University School of Medicine
10900 Euclid Avenue
Cleveland, OH 44106
(216) 368-0581
(216) 368-3432 FAX
jhn4@po.cwru.edu

Kenji Nakamura, Ph.D.
Scientific Review Administrator
Office of Scientific Review
National Human Genome Research Institute
National Institutes of Health
Building 38A, Room 609
38 Library Drive, Mailstop 6050
Bethesda, MD 20892-6050
(301) 402-0838
(301) 480-2770 FAX
nakamurk@exchange.nih.gov

Robert Nussbaum, M.D.
Chief
Genetic Disease Research Branch
National Human Genome Research Institute
National Institutes of Health
Building 49, Room 4A68
49 Convent Drive, Mailstop 4472
Bethesda, MD 20892-4472
(301) 402-2039
(301) 402-2170 FAX
rlnuss@nhgri.nih.gov

Susan E. Old, Ph.D.
Health Scientist Administrator
Division of Heart and Vascular Diseases
National Heart, Lung, and Blood Institute
National Institutes of Health
Rockledge II, Room 9150
6701 Rockledge Drive, Mailstop 7940
Bethesda, MD 20892-7940
(301) 435-0477
(301) 480-1336 FAX
olds@gwgate.nhlbi.nih.gov

James Ostell, Ph.D.
Chief, Information Engineering Branch
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
Building 38A, Room 8N813
8600 Rockville Pike, Mailstop 53
Bethesda, MD 20894
(301) 435-5978
(301) 480-9241 FAX
ostell@ncbi.nlm.nih.gov

Jane Peterson, Ph.D.
Program Director
National Human Genome Research Institute
National Institutes of Health
Building 38A, Room 614
38 Library Drive, Mailstop 605
Bethesda, MD 20892
(301) 496-7531
(301) 480-2770 FAX
jane_peterson@nih.gov

Rodney Rothstein, Ph.D.
Professor
College of Physicians and Surgeons
Department of Genetics and Development
Columbia University
701 West 168th Street
New York, NY 10032-2704
(212) 305-1733
(212) 923-2090 FAX
rothstein@cuccfa.ccc.columbia.edu

Edward Rubin, M.D., Ph.D.
Department Head
Department of Genome Sciences
Lawrence Berkeley National Laboratory
University of California at Berkeley
One Cyclotron Road, Mailstop 84-171
Berkeley, CA 94720
(510) 486-5072
(510) 486-4229 FAX
emrubin@lbl.gov

Gerald M. Rubin, Ph.D.
Professor
Department of Molecular and Cellular Biology-HHMI
University of California at Berkeley
Room 3200, Box 539
Life Sciences Annex Building
Berkeley, CA 94720-3200
(510) 643-9945
(510) 643-9947 FAX
gerry@fruitfly.berkeley.edu

Alan F. Scott, Ph.D.
OMIM Scientific Director for Genes
Johns Hopkins University School of Medicine
600 North Wolfe Street
Baltimore, MD 21287-4922
(410) 955-2553
(410) 614-9752 FAX
afscott@welchlink.welch.jhu.edu

Grace L. Shen, Ph.D.
Program Director
Cancer Genetics Branch
Division of Cancer Biology
National Cancer Institute
National Institutes of Health
Executive Plaza North, Room 501
6130 Executive Boulevard
Rockville, MD 20892-7381
(301) 435-5226
(301) 496-8656 FAX
gs35r@nih.gov

Rochelle Small, Ph.D.
Health Scientist Administrator
Division of Human Communications
National Institute on Deafness and Other Communication Disorders
National Institutes of Health
Executive Plaza South, Room 400C
6120 Executive Boulevard, Mailstop 7180
Bethesda, MD 20892-7180
(301) 402-3464
(301) 402-6251 FAX
rochelle_small@nih.gov

Randall F. Smith, Ph.D.
Director
Bioinformatics Department
Smith Kline Beecham Pharmaceuticals
P.O. Box 1539 UW2230
709 Swedeland Road
King of Prussia, PA 19406
(610) 270-5774
(610) 270-5580 FAX
randall_f_smith@sbphrd.com

Chris Somerville, Ph.D.
Director/Professor
Carnegie Institution of Washington
Stanford University
260 Panama Street
Stanford, CA 94305
(650) 325-1521, ext. 203
(650) 325-6857 FAX
crs@andrew.stanford.edu

Marjorie Tingle, Ph.D.
Health Scientist Administrator
National Center for Research Resources
National Institutes of Health
Room 6154
6705 Rockledge Drive, Mailstop 7965
Bethesda, MD 20892-7965
(301) 435-0772
(301) 480-3659 FAX
macjoriet@ep.ncrr.nih.gov

Peter Tonellato, Ph.D.
Director
Informatics Research Center
Medical College of Wisconsin
8701 West Watertown Plank Road
Milwaukee, WI 53226
(414) 456-8570
(414) 456-6546 FAX
tone@mcw.edu

David Valle, M.D.
Professor of Pediatrics
Professor of Molecular Biology and Genetics
Johns Hopkins University School of Medicine
Room 802, PCTB
725 North Wolfe Street
Baltimore, MD 21205
(410) 955-4260
(410) 955-7397 FAX
dvalle@jhmi.edu

Susan J. Waldrop
Assistant Director for Program Coordination
Office of Science Policy
National Cancer Institute
National Institutes of Health
Building 31, Room 11A03
31 Center Drive
Bethesda, MD 20892
(301) 496-1458
(301) 496-7807 FAX
sw40j@nih.gov

Neal B. West, Ph.D.
Program Director
Division of Comparative Medicine
National Center for Research Resources
National Institutes of Health
6705 Rockledge Drive, Room 6166
Bethesda, MD 20892-7965
(301) 435-0749
(301) 480-3819 FAX
nealw@ncrr.nih.gov

Monte Westerfield, Ph.D.
Professor
Institute of Neuroscience
University of Oregon
1254 University of Oregon
Eugene, OR 97403
(541) 346-4607
(541) 346-4548 FAX
monte@uoneuro.uoregon.edu

## Planning Committee of the NIH Model Organism Database Workshop


Peter Cartwright
President and Chief Executive Officer
Cimarron Software, Inc.
175 Southwest Temple, Suite 530
Salt Lake City, UT 84101
(801) 521-3210
(801) 521-3111 FAX
pc@cimsoft.com

J. Michael Cherry, Ph.D.
Head
Genetics Database Group
Department of Genetics
Stanford University
Always Building, Room M341
Stanford, CA 94305-5120
(650) 723-7541
(650) 723-7016 FAX
cherry@genome.standford.edu

Janan T. Eppig, Ph.D.
Staff Scientist
Mouse Genomics Laboratory
The Jackson Laboratory
600 Main Street
Bar Harbour, ME 04609-1500
(207) 288-6422
(207) 288-6433 FAX
jte@jax.org

William M. Gelbart, Ph.D.
Professor
Department of Molecular and Cellular Biology
The Biological Laboratories
Harvard University
16 Divinity Avenue
Cambridge, MA 02138-202
(617) 495-2906
(617) 496-1354 FAX
gelbart@morgan.harvard.edu

Howard J. Jacob, Ph.D.
Associate Professor
Laboratory of Genetics Research
Department of Physiology
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI 53213
(414) 456-4887
(414) 456-6516 FAX
jacob@mcw.edu

Joe Nadeau, Ph.D.
Professor
Genetics Department BRB-630
Case Western Reserve University
School of Medicine
10900 Euclid Avenue
Cleveland, OH 44106
(216) 368-0581
(216) 368-3432 FAX
jhn4@po.cwru.edu

Monte Westerfield, Ph.D.
Professor
Institute of Neuroscience
University of Oregon
1254 University of Oregon
Eugene, OR 97403
(541) 346-4607
(541) 346-4548
monte@uoneuro.uoregon.edu

**On Behalf of the NIH**

Lisa Brooks, Ph.D.
Program Officer
Genetic Variation Program
Genome Informatics Program
National Human Genome Research Institute
National Institutes of Health
Building 38, Room 614
38 Library Drive, Mailstop 6050
Bethesda, MD 20892-6050
(301) 496-7531
(301) 480-2770 FAX
lisa_brooks@nih.gov

Stephen C. Mockrin, Ph.D.
Deputy Director
Division of Heart and Vascular Diseases
National Heart, Lung, and Blood Institute
National Institutes of Health
Rockledge II, Room 9170
6701 Rockledge Drive, Mailstop 7940
Bethesda, MD 20892-7940
(301) 435-0477
(301) 480-1336 FAX
mockrins@gwgate.nhlbi.nih.gov

Susan E. Old, Ph.D.
Health Scientist Adminstrator
Division of Heart and Vascular Diseases
National Heart, Lung, and Blood Institute
National Institutes of Health
Rockledge II, Room 9150
6701 Rockledge Drive, Mailstop 7940
Bethesda, MD 20892-7940
(301) 435-0477
(301) 480-1336 FAX
olds@gwgate.nhlbi.nih.gov