National Center for Research Resources
NATIONAL INSTITUTES OF HEALTH

Navigating the Translational Researcher through a Complex of Animal and Biological Resources

Natcher Conference Center
Bethesda, MD

March 6-7, 2006

Final Workshop Report

September 2006

The views expressed in written conference materials or publications and by speakers and moderators at HHS-sponsored conferences do not necessarily reflect the official policies of the Department of Health and Human Services, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

I. Table of Contents

II. Executive Summary

The number and complexity of evolving disease models—naturally occurring, induced, and genetically engineered—is increasing much faster than our ability to effectively access and use the new information to speed life-saving therapies to the clinic. This current state of inefficiency defines a critical need for an innovative information service to provide clinical translational researchers with direct links to the full spectrum of clinically relevant model systems: animal models, cell culture, and molecular reagents.

This NCRR-sponsored workshop explored approaches for developing a disease-centric resource that would enable researchers to find and to use animal and other biological resources in ways not currently possible. The objectives were to identify the user community and its needs and to determine the nature and ideal characteristics of the resource to ensure that it will evolve in concert with the science and technology that it provides. The 50 invited workshop participants included animal researchers, clinical/translational science researchers, resource managers/developers, industry, and NIH intramural and extramural staff. (See list at end of report.)

The workshop began with presentations from NIH-supported researchers who provided an overview of current animal model databases and resources that are available and described some of their current limitations. Then, during breakout sessions, participants, guided by the following questions, discussed the potential new resource:

> *Who are the intended and potential users of the new resource? What tasks are they performing and what are their needs, particularly in translational science?*
>
> *What are the characteristics of this new resource that will best serve the users' needs?*
>
> *What framework for animal model systems is most appropriate to serve as the basis for this new resource that will also support later development? Possibilities might include one or more disease categories, body or organ systems or tissues, among others.*
>
> *How can currently available resources and distributed systems information technology projects be leveraged to develop the new shared infrastructure?*
>
> *What are the requirements of a cyberinfrastructure framework for animal model systems that are most critical in order to provide a foundation for this new resource that will also support rapid technological elaboration and expanded use?*

Discussion Highlights:

- *Vision*

  The problem at hand is not to structure information technology machinery but rather to identify tasks of the potential users and the challenges they face in achieving their research aims. The questions that users need to have answered must be articulated to determine what the most relevant technologies will be; that is, the vision should drive the technology, not the other way around.

  Previously, the standard approach was simply to start another database, while the new model is to join forces among existing databases and bring new value to an existing data infrastructure.

- *Users and their tasks*

The audience/user group for this proposed resource goes beyond clinician scientists, translational researchers, and basic scientists to those at the bedside—the clinician and patient should also be able to access information in a useable way.

New researchers need to look at promising models and determine how the model will help the patient, how it differs from the range of human disease, what its limitations are, and what an ideal model could do that other ones don't. However, there remain certain things that can only be addressed with a human model, and this situation needs to be clear when it occurs.

- *Challenges*

Data obtained from new methods must be linkable to other pieces of data. Realizing the full power of a model depends on facilitating seamless exchange of resources and making integrated genetic, genomic, and phenotypic information available to promote knowledge discovery.

Variation in the quality of information is an important consideration. The proposed model approximates the "what-if" question, that is, of being able to obtain information online without actually carrying out an experiment. While searching for the ideal, investigators must be aware that much of the information on the Web is inaccurate, incomplete, misleading, uncertain, and limited.

There is an undefined lag period for use of material by the scientific community, and it is a problem to keep information available so that it will be accessible when needed by an investigator.

This work poses significant sociological challenges—it is necessary not only to work in more collaborative ways, but also to help biomedical scientists learn to value computer scientists and see them as equal and valuable colleagues.

- *Use of existing tools and technologies*

The Biomedical Informatics Research Network (BIRN), a large consortium that is building an infrastructure of networked high-performance computers, data integration standards, and other emerging technologies, is one option proposed for navigating this system. The BIRN builds on evolving community standards for middleware and adds new capabilities required by projects. The components shaped by BIRN for the needs of the scientific community are authentication, authorization, auditing, workflows, visualization, and analysis.

A National Library of Medicine system that was previously used for human-to-human conceptual mappings has been successfully adapted for mouse-to-human mappings, illustrating that automatic mapping could find mapping not otherwise found by manual mapping.

Recommendations: follow-up activities:

1) Develop a catalog of current databases with a brief description of each
   a) Issue a Request for Information (RFI) to determine process – Should we initiate a database of animal systems, diseases, particular laboratories, etc.?
   b) Initial catalog development efforts will be partially focused to synergize and accommodate models needed for the effort described in 2).

2) Launch a prototypic model resource effort from a disease platform
   a) Start with a disease category or a specific disease
   b) Choose a prototypic effort which will be scalable
   c) Neurodegenerative diseases is a suggestion
   d) Focus on a single, cross-cutting disease with well defined models/databases and trans-NIH interest such as Alzheimer's.
   e) Establish an advisory committee to determine and guide process.

III. Purpose and Objectives of Workshop

NCRR plans to create a Web-based portal to integrate and coordinate use of all NIH-supported animal and related biological resources. This new resource will enhance access and retrieval of information from existing model databases and accommodate the addition of new ones. The use of translational science to address human health-related issues is becoming less of an option in today's biomedical research and is fast becoming a requirement needed to quickly and effectively move between discovery research and the clinic. As a result of this shifting research paradigm, there has never been a greater need for easily accessible and broadly informed disease modeling systems to guide the translational researcher to and through the best preclinical studies.

To plan the design of the resource, invited experts met on the NIH campus, March 6-7, 2006 for the workshop "Navigating the Translational Researcher Through a Complex of Animal and Biological Resources." Participants identified the user community and its needs and described the range of expertise and technology needed to staff and support the resource. Invited workshop members included animal researchers, clinical and translational science researchers, resource managers and developers, industry representatives, and NIH intramural and extramural staff.

IV. Summary of Presentations and Discussions

DAY 1 PRESENTATONS:

Animal Models and Resources – Biology

Monte Westerfield, Ph.D., Institute of Neuroscience, University of Oregon
Workshop cochair

The linkages among translational research, and animal and bio-materials resources were examined by looking at the problem with resources and limitations of current informatics support. Examples of solutions included terminologies to describe resources and ontologies to link resources with human disease.

The biological resources of NCRR's Division of Comparative Medicine encompass invertebrate, rodent, nonhuman primate, zebrafish, and biological materials, with multiple assets for most categories. But the difficulty is in how to use them, how to go from biological models to human diseases and back again to biological models, and where to find out what models are available.

This was illustrated with details of a search for model systems to study retinal degeneration. After an Internet search, many Web sites were visited including: the NCRR site; the Mouse Genomes Informatics (MGI) and Mutant Mouse Research Resource Center (MMRRC), the Zebrafish Information Network (ZFIN), the American Type Culture Collection (ATCC), the FlyBase, the *Saccharomyces* Genome Database (SGD) for yeast, and got no hits. Further searching for nonhuman primates revealed that most home pages do not have search engines.

There are a number of problems that impede finding animal models or biological materials as models for human disease. These include inconsistent search mechanisms and descriptive terms and lack of support for synonyms. There is also a lack of centralized one-stop shopping, with researchers finding heterogeneous information at multiple sites.

The next task is to determine how to overcome these problems. The newly established National Center for Biomedical Ontology, an initiative of the NIH Roadmap, has as one of its goals relating animal models to human disease, that is, mutant gene to mutant or missing protein to mutant phenotype (disease model). In the past, the only method of relating human and animal models has been by looking at gene sequence, but in many cases only the phenotypes, the clinical signs, and not the genes, are available. Researchers need to be able to make a connection at the phenotype as well as the genotype level.

A smooth functioning interface between our genotype and phenotype information collections is needed. The primary problem is lack of a centralized resource and this is the first issue that must be addressed. An individual researcher's search strategy must also be considered. A translational researcher might use a set of ontological keywords, while an animal model researcher would start from the animal model side and look for human diseases.

Primate Models and Resources

Stuart Zola, Ph.D., director of the Yerkes National Primate Research Center

Nonhuman primate (NHP) models have played an important role in studies of different aspects of cognition and cognitive decline. Early diagnosis of these conditions in humans provides an opportunity for more effective treatment, and it now appears that mild cognitive impairment might be a marker or risk factor for Alzheimer's.

Yerkes researchers are working with imaging of amyloid plaques to try to develop an early Alzheimer's diagnosis. A vaccine from the Elan Corporation (AN-1792) is designed to attack and clear out plaques. Mice immunized at a young age were protected from plaque deposits and they were halted and in some cases reversed in animals that already had deposits. However, clinical trials in human patients were halted when brain inflammation occurred in some of the subjects. The vaccine is currently being redesigned to be more selective. An important finding of the work so far is that deposits of amyloid in blood vessel walls appear to occur earlier than expected.

Science is on the brink of a new era with sequencing of the human and nonhuman primate (chimpanzee, rhesus monkey) genomes. Questions that need to be answered in Alzheimer's work involve where and when genes important for memory function are expressed, how cells expressing these genes are influenced by the environment and what individual differences are, and where to spot points of vulnerability that can cause impaired memory.

This will lead eventually to new era of predictive and preventive medicine. But many challenges remain in genomics, behavior, and imaging. All of the information for a single subject must be brought together and transformed into useable knowledge. This is where bioinformatics and programs such as BIRN will be critical. And basic research as well as translational research still has an important role to play in advancing diagnosis, interventions, and therapies. The real challenge is how these animal models can provide information that is useful.

Animal Models in Clinical and Translational Science

Keith Cheng, M.D., Ph.D., associate professor of pathology, biochemistry and molecular biology, and pharmacology at Hershey Medical Center, Pennsylvania State University

The audience/users of this proposed resource goes beyond clinician scientists, translational researchers, and basic scientists to those at the bedside—the clinician and patient should also be able to access information in a useable way. The material should also be available to the general public, to Ph.D. and M.D. trainees, college students, high school students, and legislators.

The diverse users need one-stop shopping, to engender thinking processes between genetics and reverse genetics, and to provide feedback by users, which is critical. New researchers need to look at promising models and determine how the model will help the patient, how it differs from the range of human disease, what its limitations are, and what an ideal model could do that current ones do not.

The most obvious animal models are the vertebrates—primates, mice, rats, and zebrafish. Others are the invertebrates (flies, worms), single cell organisms (phage, bacteria, yeast), and other models, including *Xenopus* (frogs), dogs, tetrahymena, other fish (fugu, Medaka, xiphophorus), and hydra. But there remain certain aspects that can only be answered by a human model—anatomy, histology, physiology, genetics, gene expression, disease correlations with animal models, environment, and psychosocial issues.

There are different ways of approaching the biological process and disease—through studies of aging, normal variation, neurophysiology, cancer, pigmentation, metabolism, or infectious disease, for example. There are also a number of ways of linking concepts: biochemical and signal transduction pathways, genes, gene expression, anatomy, histology, subcellular localization, lifespan, physiology, and disease. Often investigators who are focused on gene expression overlook the other areas. Tools for comparison are the core of translational research. Comparisons could be applied in many different ways: wild type versus mutant or morphological variant; older versus younger; treated versus untreated; one organism versus another; or multilevel among organs, tissues, cells, subcellular, or molecular materials.

A type of comparison was illustrated that is familiar to most researchers, comparing amino acids across species, specifically how human polymorphism rs1426654 affects a conserved amino acid. Comparisons can also be made among different parts of the genome or between different populations. Systems morphogenetics are an organizing tool through which biologists can work with engineers using the most up-to-date technology to answer questions about where, when, why (function), physiology, disease, and response to treatments including therapeutic effects, side effects, and rare adverse responses.

Questions about function, pathway, tissue specificity, subcellular localization, age specificity, and downstream effects need to be answered. But there is also a need to raise questions, to explore what is not known, to let that guide the research. It is important to direct questions, prioritize, identify information by degrees of uncertainty, identify conflicting opinions, and highlight important unknowns, clinical problems, and clinical goals.

The zebrafish serves as a model; studies of development, disease, drug development, and toxicology have the potential to serve as a bridge between invertebrates and mammals.

The models for the specific disease of cancer span a range from yeast to humans, and there are many diseases with many different pathophysiologies and complicated genetics. Treatments are primitive, and genomic instability creates a moving target. The National Cancer Institute's (NCI) new Cancer Genome Anatomy Project has been established to generate the information and technological tools needed to decipher the molecular anatomy of the cancer cell.

Another approach to translation is to study a general process such as pigmentation. Different genes are important in different settings, with the most common variations affecting color of skin, hair, and eyes. Two or three genes are the most important, but many modulate the effect, and they also have other potential effects including Parkinson's disease, age-related macular degeneration, and frostbite susceptibility. This as a "solvable" problem. Many aspects of pigmentation can be studied including variable anatomy and histology, transfer of pigment between cells, variable biochemistry, and variations in melanosome morphology. Proteomics are coming online, and there is extensive information about genetics and evolutionary conservation of function and normal and abnormal patterns in diseases.

Quality of information is an important consideration. While searching for the ideal, investigators must be aware that much of the information that is found on the Web is inaccurate, incomplete, misleading, uncertain, and limited. Even sites that seem reputable might cite erroneous information. Every researcher cannot possible keep up with every manuscript, and a means to extract accurate information is needed. This is highlighted by new findings involving glutaric acidemia type 1, or GA-1, a metabolic disease for which a new diet-induced mouse model allows mechanistic studies and drug development. This has led to key information about where the gene is expressed, where metabolites accumulate, why some parts of the brain are affected and others are not, and the cause of age-dependent resistance.

Information on the Web varies according to presentation, purpose and organization of site, names and file formats, relationships among important concepts, resolution, quality, and recency of information. The Semantic Web, which will allow for categorization and computation-dependent to extraction and organization of information, may provide one solution.

Translational researchers should plan now for a wide audience, including clinician and patient involvement. User feedback will be critical to discuss and respond to, and there is a need to integrate simpler models and humans. The human model will deal with critical issues that are not possible to study in other model organisms. Functional, developmental, and physiological data require greater prominence, and there is a need to highlight what is not known and link available information. It will be important to monitor limitations of models and data.

Mouse Models and Resources—Informatics

Mark Ellisman, Ph.D., director of Biomedical Informatics Research Network Coordinating Center, University of California, San Diego (UCSD)
Workshop cochair

The Biomedical Informatics Research Network (BIRN) project stands at the intersection of convergent revolutions in genomic medicine and information technology. The growth in the number, diversity, and utility of animal models parallels the rapid increase in the capabilities of distributed information systems to link data about these models. But the huge increase in diversity of models, the cost of models, and the need for linking information present challenges. The premise and motivation for this workshop is that biomedical researchers are not able to use animal models resources as effectively as they could if information about the models were more readily available and better integrated. The goal is to make it easier for researchers to find and use NIH-supported animal and other biological resources.

The BIRN consortium is comprised of the coordinating center, which develops and supports overall information technology (IT) infrastructure linking the testbeds, and three testbed projects that are conducting structural and functional studies of neurological disease. Morphometry BIRN is studying brain

structures related to unipolar depression, mild Alzheimer's disease, and mild cognitive impairment. Function BIRN is studying regional brain dysfunctions related to the progression and subtypes of schizophrenia. Mouse BIRN is studying animal models of multiple sclerosis, schizophrenia, Parkinson's disease, attention deficit/hyperactivity disorder, Alzheimer's, Tourette's disorder, and brain cancer.

Several groups are now forming. Researchers at Yerkes will be working on a nonhuman primate BIRN related to memory disorders. From Edinburgh, Scotland, the Medical Research Council and the Engineering and Physical Sciences Research Council of the United Kingdom will be working on gene expression maps in mouse development. An NIH intramural initiative, the National Database for Research on Autism, is a cooperative effort to collect human imaging and other data that involves the NIH Center for Information Technology, the National Institute of Mental Health (NIMH), the National Institute of Environmental Health Sciences (NIEHS), and the National Institute of Neurological Diseases and Stroke (NINDS). This group worked with the BIRN coordinating center and derived infrastructure and tools from human BIRN testbeds.

Data federation was illustrated using an example of multiscale, multimodal data from Mouse BIRN. The first step is to create databases at each site, then create conceptual links to a shared ontology, or hierarchy. The third step is to situate the data in a common spatial framework and then use an integration engine called a mediator to navigate and query across data sources.

Data models provide frameworks to integrate databases and integrated views are based on ontologies, which are a way to communicate a shared understanding of a field. Ontologies provide a number of frameworks including a representation of terminological knowledge, explicit specification of a conceptualization, concept hierarchy, and further semantic relationships between concepts ("is part of," "causes," etc.). Examples are Gene Ontology (GO); NeuroNames, a structured system of neuroanatomical terminology; a foundational model of anatomy; and the mouse anatomy coming out of Edinburgh.

Integrated views allow building of domain maps showing how all elements are linked and allowing the ability to navigate across different domains. In atlas-based spatial reference systems, multiscale and multimodal data are connected through ontologies. In storing data in a database, BIRN requires a unique identifier such as that provided by UMLS, the Unified Medical Language System.

The BIRN data integration environment bridges data models when users explore distributed data. For example, there are multiple ways to answer the request for images of medium spiny neurons, tract-traces, and histology of surrounding regions from the Parkinson's Disease (PD) alpha-synuclein (α-SYN) mouse model—it can be defined spatially on an atlas, it can be approached through ontologies or terminology or databases. Or it can be approached through transformations, other types of information that are related to a structure but not necessarily spatial in nature such as level of gene expression or activity in an area.

The BIRN SMART Atlas is a harbinger of things to come, an example of a data grid-based tool for spatial integration of multiscale distributed brain data. SMART is an acronym for spatial mark-up and rendering tool and it runs from a BIRN portal. It features a Java- and GIS-based brain navigation system and query interface, a database of spatial relationships between brain structures, a tool for registering and querying spatially registered data contained in distributed file systems or databases, and spatial and semantic integration of multiscale and multimodal brain data.

The Parkinson's disease mouse animal model is a BIRN testbed collaborative project that involves many sites with multidisciplinary contributions towards characterizing mouse models of Parkinson's. The National

Center for Microscopy and Imaging Research (NCMIR) is using multiscale imaging methods to characterize transgenic mouse models of PD produced by researchers at UCSD. NCMIR has also worked with them to acquire behavioral and molecular biology data on the same mouse models. At Duke University, the Center for *In Vivo* Microscopy is producing high resolution microscopic MRI data. Additional collaborators at Harvard University and University of Illinois are evaluating therapeutics, based on the multimodal data that is being collected.

PD researchers are looking at alpha-synuclein (α-SYN), a member of the synuclein family of synaptic proteins without a clearly defined role, with mutations that are associated with familial PD. When transgenic mice overexpress human alpha-synuclein, they exhibit motor deficits and α-SYN inclusions in neuronal cell bodies, neurites, and glial cells in cerebellum, hippocampus, and cortical regions. Multimodal studies of PD animal models include behavioral data (assessment of cognitive and motor function), multiscale imaging (correlation of large-scale mapping of immunolabeling and MRI studies and ultrastructural studies using electron microscopy), and chemistry and genetics (protein expression, Web quantitative trait loci [QTL] analyses, ligand biding studies) and can lead to pathological and neurological features.

PD research presents substantial challenges in making connections between *in vitro* preclinical models and human disease state. It is a spectrum of disorders with a complex etiology and no perfect model systems. The α-SYN mouse recreates pathological features of PD in several ways: it has widespread alpha-synuclein immunopositive aggregates, altered mGluR5 and dopamine receptor expressions, and a wide range of PD-like cognitive and motor neurological deficits. This provides opportunities for evaluating potential therapies, both pharmacologically and immunization. Investigators are almost ready to use these animals for preclinical screening of therapeutic approaches for PD.

In summary, BIRN is in its fifth year and growing, with almost 40 sites participating and more than 200 members at the last all-hands meeting in October 2005. Almost half of the group members are computer scientists. This highlights a sociological challenge of this work—it is necessary to not only work in more collaborative ways, but also to help biomedical scientists learn to value computer scientists and see them as equal and valuable colleagues. Much of the information and many of the tools in BIRN are publicly available. BIRN is helping communicate about the use collaborative research elements, and this is an opportunity to look at areas to utilize BIRN and open up new areas. It can be easy to see where a model fails as it is characterized, as one way to identify the limitations of models and point new translational researchers to alternative models they might want to use.

Mouse Models and Resources

Richard Woychik, Ph.D., director of The Jackson Laboratory

There are many challenges in the mouse community related to current resources and those anticipated coming online, as well as technologies being put together to solve them. Current advances may be illustrated by the fact that in 1958 all of the known genes of the mouse could be displayed on a single display board of mutant mice.

Today that has become much more complex with multiple mutant, inbred, and special inbred strains now available. From the first mutant strain created at Jackson there are now hundreds of inbred strains worldwide. In addition, special inbreds include wild-derived, recombinant, congenic, and consomic inbreds and recombinant congenics. The search for mutants has led to transgenic/targeted mutations, insertional mutations, agent-induced mutations, chromosome aberrations, and spontaneous mutations, with an explosion of different lines now available.

Jackson has more than 3,000 strains, and thousands of others are available throughout the world. The NIH Knockout Mouse Project, with the goal of engineering knockouts on all the genes of the mouse, further broadens the number of strains. Data resources derived from mice are also growing and include the NIH Transcriptome Project, cDNA and other libraries, and Jackson's Backcross mapping panels.

Realization of the full power of the mouse depends on facilitating seamless exchange of resources and making integrated genetic, genomic, and phenotypic information available to promote knowledge discovery.

The Mouse Genome Informatics (MGI) consortium is the recognized international community database resource for the mouse. It provides an integrated view of genomic and biological data, definitive reference datasets for mouse genomic and biological information, and a structured, integrated database for gathering, managing, analyzing, and accessing these data. The goal of MGI is to facilitate use of the mouse as model for human biology and disease and to use mouse mutants to help provide understandings of implications in humans. One example is achondroplasia; a mouse model has been developed with a short domed skull, short-limbed dwarfism, malocclusion, bulging abdomen as adults, respiratory problems, and shorted lifespan, the traits of this disorder in humans.

Data integration is a primary MGI activity. When data can be gathered from multiple sources, it is possible to factor out common objects, identify new objects, and assemble integrated objects, leading to integrated views that reveal data relationships not evident in the original source data. The MGI database covers sequences and maps, genes and gene products, gene expression, tumor biology, and phenotypes. It can be accessed at http://www.informatics.jax.org and users can search for and retrieve data, interactively explore details, generate map views, and download datasets. It also includes instructions for programmers. The Web site allows researchers to access nomenclature, mapping, orthologs, sequence, phenotype, gene ontology, expression, links, protein domain, etc.—a truly remarkable range of options.

MGI provides consensus/meta data views and data detail. It allows users to ask complex questions that simultaneously address areas of sequences, genes, mutations, phenotypes, expression, and function. For example, a mouse tumor frequency grid, which could respond with links to individual sequences and genes to a complex query such as: Provide a dataset of RefSeq RNA sequences for all genes in the mouse that have been annotated as kinases, are expressed in the heart, and have mutant alleles associated with heart or cardiovascular system phenotypes. Restrict to sequences from the C57BL/6J strain.

The Mouse Phenome Project (http://www.jax.org/phenome) is a database representing an international community effort to characterize genetic/phenotypic variability in inbred mouse strains. Phenotype data are assembled into a database for 40 different inbred strains, in 4 groups. This project can show phenotypic manifestations with the example of a search for serum cholesterol levels after 7 weeks on a high-fat diet in some strains, recapitulating the phenotypic diversity that exists in the human population. The Mouse Phenome Project features a large variety of phenotypes, high density single nucleotide polymorphism (SNP) discovery, an NIEHS-sponsored resequencing of 15 phenome strains, and phenotypic diversity as least as great as what exists in the human population.

A global catalog of mouse resources can be found at the International Mouse Strain Resources (IMSR) Web site (http://www.imsr.org). IMSR answers questions about where mouse resources can be found; in what state they are maintained; where more information is available; how to find information of mutations, alleles and phenotypes; and how the holder of the resource can be contacted. Searches can be done by

strain name, ID, state, and strain type; gene or allele and type of mutation carried; or repository or geographic region. IMSR currently has 20 repositories.

Another challenge is addressing the increasing number of mouse lines and coming up with better technology to maintain and distribute this information. The ideal solution would be a rapid, efficient, proven, reliable, and cost effective means of removal/recovery of a line of mice from the shelf. It would provide inexpensive and reliable long-term storage and easy shipment and importation. Embryo cryopreservation meets many of the requirements. This reliable long-term storage approach has been around for over 30 years. It is rapid, efficient, and cost effective and eliminates genetic drift. Shipment of embryos can be handled by most investigators, and most institutions have an embryo transfer laboratory. Speed cryopreservation can have hundreds of embryos ready for cryopreservation after overnight incubation.

Some of the projects described are separate databases and some are not. Some, such as Phenome, depend on the specific project supported. They are already integrated and steps are being taken to make them even more integrated.

Rat Models and Resources

Howard Jacobs, Ph.D., director, Human and Molecular Genetics Center, Children's Hospital, Medical College of Wisconsin

Different types of rat datasets were presented, including the Rat Genome Database (RGD), disease portals, comparative genomics, and building interactive physiological genomic databases.

The sequencing of the rat genome has changed the dynamics of what researchers need to talk about for data management. An extensive list of what is available includes the full genomic sequence with 10,000 genetic markers; four bacterial and two yeast artificial chromosome libraries; about 1.5 million SNPs, with maps under construction and chips being developed; microarray chips; more than 100,000 siRNA probes; 726 strains, knockouts and cloning possibilities; 1,029 QTL or biological phenotypes assigned to the rat genome; disease, phenotype, and gene ontologies; direct translation to human and mouse via comparative genomics; and more that 1.2 millions publications in the scientific literature.

The RGD (http://rgd.mcw.edu) was originally developed to support NIH rat genome project and initially consisted of just genetic maps. It now provides genomics/genetic and phenotyping data. For example, a recent search for the keyword "depression" in RGD turned up 68 hits—25 genes, 20 QTLs, 1 strain, 10 homologs, and 12 references. For those who know how to mine the information, RGD is a valuable source, but it can be difficult for a less experienced user. RGD offers a number of tools to search and analyze data; the G-viewer is one of the most valuable.

RGD provides genomics data and is being developed into a phenome database. But a problem is that most investigators who are interested in translational research will not know how to find the data they need. The solution is the development of disease portals, which allow researchers to enter a disease for the search term and come up with genes, QTLs, and strains in one search. The challenge is to make one of these for each specific disease of interest. Comparative genomics brings together investigators interested in same disease but who study different models or clinical samples. The rat, mouse, and human map to the same genetic location for many human diseases.

Currently, disease portals are strong in cardiovascular diseases because of funding from the National Heart, Lung, and Blood Institute (NHLBI) and local expertise. The genome browser can be used to show, for example, the region on the human genome where hypertension maps for specific populations and on the rat chromosome. Similar phenotypes are found to map to the same evolutionary location.

Systems biology and physiological genomics will provide new types of data, new ways to integrate it, and, hopefully, new ways to ask questions. Scientists have now learned that genome background changes and this has a great effect in managing data. Personalized medicine means developing this for each patient, and there is clearly a long way to go before this will be realized. With an interactive discovery tool, all physiological data are present, all phenotyping protocols are available, and users have the ability to ask questions and run statistical analyses on the dataset.

PhysGen (http://pga.mcw.edu/pga/index.html) is part of NHLBI's Programs for Genomics Applications. This Web site for physiological genomics focuses on understanding the genetic basis of fundamental mechanistic pathways of the heart, lung, kidney, blood, and vasculature through development of consomic rat panels and knockout models and physiological genomics, using environmental stressors. The home page provides an overview, data, components, and links, and all data are available for downloading. Consider the question: where in the genome of the SS rat does hypertension map? A Brown Norway rat that has been sequenced was benchmarked against 11 strains commonly used by industry to derive a chromosome sensitivity panel, changing one chromosome at a time. Then an ANOVA analysis on all strains determines which strains are statistically different from a specific type.

Another study used microarrays to identify differentially expressed genes in consomic rats in response to a high salt diet and identified a number of potentially important pathways. Researchers are trying to look at animal models, using ontologies to characterize phenotypes, quantitative traits, genes, and sequence.

In summary, rat genetics and genomics enable translation between different species but the extensive resources require integration. Rat bioinformatics continues to improve, with rat biology as a driver. There is a need to enable hypothesis testing from online data, with multiple contributors generating data, and a need to continue to capture data without extensive data curation, a major issue.

This model approaches the "what-if" question, of being able to obtain information online without actually carrying out an experiment. The mouse community is facing similar struggles and just because it is possible to generate many lines does not mean people are interested in them. Rat researchers are making targeted knockouts in a large number of genes and want to know the gene and phenotype, not just the phenotype. The problem is continuing to generate the line and not knowing what to do with it, and cryopreservation will be vital for this. There is an undefined lag period for use of material by the scientific community, and the problem is keeping resources available so they will be ready when an investigator wants them.

Ontological Resources for the Translational Researcher

Olivier Bodenreider, M.D., Ph.D., Lister Hill National Center for Biomedical Communications, National Library of Medicine

Integration is one of the key concepts and objectives of this workshop. The subdomains of primary interest are clinical repositories, genetic knowledge bases, biomedical literature, genome annotations, anatomy, and model organisms. Some are of greater interest to the bench scientist, some to the clinical researcher.

They can be integrated in several ways, and the use of ontology to do this is not far from being operational in some cases. The Unified Medical Language System (UMLS) can help with data integration.

Ontology is not the same as terminology. Terminology is a collection of terms (e.g., controlled vocabularies) and is useful for indexing and annotation as in the Gene Ontology (GO) and Medical Subject Headings (MeSH). Ontologies go farther and describe types of entities (e.g., substances, qualities, processes) and relations among them. This is useful for reasoning, as in the UMLS Semantic Network and SNOMED CT, the Systematized Nomenclature of Medicine Clinical Terms from the College of American Pathologists.

The first place to search when considering animals in ontological research is in a taxonomy browser, available through the National Center for Biotechnology Information (NCBI). NCBI tools not only give the whole lineage but also provide pointers to other NCBI resources, and from one given organism investigators can find information about sequences, structure, proteins, etc., from other databases. Animals in biomedical literature can be found through check tags in Medline citations; MeSH also includes a more complete hierarchy of animal models, although the level of detail will not be precise enough for annotating experiments. This was illustrated with examples from a PubMed search, first for animals, then for mice, and it can be integrated with mice categories in UMLS.

Other ontological resources are the Open Biological Ontologies (OBO; http://obo.sourceforge.net), the extended family of the GO, which includes collaborative development, and the National Center for Biomedical Ontology (http://bioontology.org). The OBO collection is extensive, as was illustrated with examples from gross anatomy, but broader integration is needed for comparative studies. It is important that phenotype ontologies be clear to characterize experiments.

Several issues need to be resolved. Ontology and formalism should not be confused, although differences might be subtle. Ontology languages are OWL and Protégé, which could be confused with markup languages; the markup languages (e.g., CellML, MAGE-ML) can be thought of as formats or syntax for exchanging data. But it is necessary to put content into syntax, and that is where ontology is needed. The danger in focusing on just syntax is that there might be more than one way of expressing a single concept, and there needs to be a way to reconcile different ways of expression. Focusing on formalism will not address this.

Another issue is related to ontologies and granularity. The information represented in most ontologies might not be fine-grained enough for some biological applications. Ontologies represent classes of entities while biological experiments refer to instances. Phenotype ontologies are emerging, but more progress is needed and phenotypes need to be fine grained and span across multiple organisms.

The goal of information integration can be achieved through ontology, which can stitch together the variety of resources investigators want to use. One example is a mapping system that was previously human-to-human, and was extended to mouse-to-human. The significance is that automatic mapping could find mapping not otherwise found by manual mapping. More of these studies should be done more systematically.

Charge to Breakout Groups and Discussion

Michael Katze, Ph.D., professor of microbiology, University of Washington, and associate director, Washington National Primate Research Center

To confront the challenge of integrating all relevant material, including clinical and animal information, the Katze lab developed its own bioinformatics group and secure database and purchased commercial products (Rosetta Biosoftware Resolver and Elucidator, GeneGo, Ingenuity) to build it. The laboratory's Expression Array Manager is accessible through a Web site. The lab is using an integrated approach to infectious disease with use of cell culture, mouse, and macaque animal models to study viral infection, combining traditional histopathological, virological, and biochemical approaches with functional genomics and proteomics to derive signatures of virulence and insights into mechanisms of host defense response, viral evasion, and pathogenesis.

As multiple NIH agencies become involved in translational research, nonhuman primates (NHPs) will experience an even greater demand, especially for infectious diseases such as influenza. The Department of Defense is investigating models in NHPs. Resources and tools now available make it possible to use NHPs in very sophisticated ways. One of the techniques is to get cells from bronchial brushings and then do sophisticated microarrays to look at pathways such as immune response. Studies have been done with low- pathogenicity viruses, and biocontainment is a major issue. Few centers have the facilities for high level biocontainment, and NIH and the Centers for Disease Control and Prevention (CDC) have not decided what level facility to use.

Jeffrey Taubenberger, M.D., Ph.D., former chief, Department of Cell Pathology, Armed Forces Institute of Pathology (AFIP), reconstructed the 1918 flu virus from genomic RNA sequences recovered from fixed paraffin-embedded samples in the AFIP National Tissue Repository from World War I soldiers and lungs from exhumed frozen corpses from Alaska. Work with the reconstructed virus has been done with macaques in a BSL-4 lab in Winnipeg and mouse experiments are being done at CDC. The 1918 virus is turning out to be very lethal in mice and macaques, although the results are not yet published.

A recent article about the future of virology described "finding function with functional genomics" for the 21st century. This will require biologists, data analysts, and computer specialists working closely together; hybrid bioinformatics software that can take validated low throughput wet lab results and make genome-wide functional predictions; and integration of emerging technologies including microarrays, proteomics, and microRNA. Long-term goals are far-reaching:

1. To define the complete host response at the RNA and protein level to every virus infection
2. To identify impacted pathways in common
3. To identify unique pathways
4. To cure all virus infections
5. To uncover links between virus infections and other diseases; e.g. malignancy

Discussion

Experimental biologists have not always taken advantage of the complexities of translational research. For example, rather than comparing mouse models that do not always correlate well and then going back to NHPs, would it be better to go back to mouse models and test, for example, 10 or 20 inbred strains? There is a need to become more sophisticated in how researchers think about an organism, and the field could gain significant mileage in looking at infections in different strains of mice. The range of possibilities is mind-boggling; one scientist suggested that if you take blood from 1,000 rhesus, you will get dozens of differential effects.

Researchers have not been working with the genome that long, while the infectious agents have been around a long time. It is important to be careful about self-confidence, to be less self-confident about what is known and what is not known. Understandings about what cells do are still primitive. Newly constructed

informatics resources must include a place for raw observational data so that data derived now can be used for comparisons later. Most of the systems being constructed now foreclose options for the future.

With complete datasets collected, biologists can do what they have always done with the information, but data become more useful if available to other disciplines. There are challenges in trying to store raw data, but researchers need to think about capturing as much mega-data about an experiment as possible and making it available through archive.

It is not always possible to link research to its clinical manifestation. Discoveries cannot always be orchestrated as planned, and it is essential to figure out ways that all pieces will be available that might be important later on. This is an important caveat about designing systems—do not lose the potential to bring in information. It is not always possible for an individual investigator to interpret the data he knows, especially when work is at a very early stage.

Group Discussion

Participants were asked to address five questions:

1. Who are the intended and potential users of the new resource? What tasks are they performing and what are their needs, particularly in translational science?
2. What are the characteristics of this new resource that will best serve the users' needs?
3. What framework for animal model systems is most appropriate to serve as the basis for this new resource that will also support later development? Possibilities might include one or more disease categories, body or organ systems, tissues, etc.
4. How can currently available resources and distributed systems information technology projects be leveraged to develop the new shared infrastructure?
5. What are the requirements of a cyberinfrastructure framework for animal model systems that are most critical in order to provide a foundation for this new resource that will also support rapid technological elaboration and expanded use?

When looking for an animal model, consider how the appropriate key words can be entered to return appropriate models. It is important to realize that these are models of diseases, not actual diseases, and this also applies to culture systems. Sophisticated ways of modeling are now available, but it is necessary to think about who would use them and what questions they would have about the data.

For example, is it possible to look at all the manipulations that led to a certain phenotype? How can a multifaceted condition like Parkinson's disease be handled? The problems are difficult and diffuse, and it is best to start with what can be done now and not what is desired 10 years down the road. It is difficult to anticipate and get started. Scientists creating ontologies are using different types of approaches, including that there is no single approach. Creating accession numbers would be helpful, so that investigators could register models just as they register gene sequences. There are many different issues that must be resolved, but he scope of the problem should be defined so that an answer can be achievable in stages.

Biologists of the future will be interested in developing datasets around organs or tissues or diseases. It would be good to establish ways for scientists to import data into an existing database rather than needing to start their own. The intent is not to create a new resource, but a resource of resources, a sort of Yellow Pages. In doing that, it is important to show new users how they can pick up the most useful data and transfer to their own data systems. biologists can come and go, but data stays integrated to a stable database, which is one of the big challenges of this task.

The proposed resource of resources needs to be more sophisticated than Yellow Pages. Investigators want to know how to find animals but they want more than that. Resources of lessons learned are also needed and available templates for people who want to come in and access the storehouses. The old model was to start another database. The new model is to join forces with existing databases and bring new value to an existing data infrastructure.

Answering the first question—who is this for?—is very important. Because this is such an ambitious undertaking, it might best begin with a very focused target. It could be overwhelming for most clinicians or patients to access this. Investigators coming from a resource-poor institution can't always use material such as infectious agents. It might be possible to manipulate the pieces that reside in a database without needing to use infectious agents in a particular lab. Users are a diverse community and resource designers need to look at potential customers in animal model communities who are begging for better ways of cross-connecting.

The first task is to determine the proper questions—not who is going to use the resource but what they will get out of it. Use cases usually develop from a vision of what is wanted for the future, what technology can enable today that it couldn't yesterday. Biologists will begin by looking at the range of resources available—this seems a simple task, but it requires hours of investigation. This is a starting point that would save people a great deal of time. One should begin with basic questions and build toward a vision.

The common questions are that a biologist will want answered need to be addressed, and if there is a need to standardize data so they are flexible, or if the current structure is adequate. The goal of Semantic Web is to draw data from different sources, but there are no standards. To create a resource, it is necessary to create a warehouse to store relevant information and make it accessible. Then investigators can selectively tap databases to answer the questions they are interested in; however, there is no panacea.

Computational tools are emerging and being used in early systems that have achieved some interesting integrations, but they are not as scalable as they need to be. The computer science community is struggling with important, big questions, and it needs to be tied to this group to understand the challenges. That can be done by scaffolding a system made of components: Start with the different types of data that are needed, get as much information as possible about each platform, and standardize wherever possible.

The participants broke into three breakout groups to work on use cases and a test case:

1. Infectious diseases
2. Developmental/genetics
3. Preclinical investigation/drug development

DAY 1 BREAKOUT REPORTS:

Development and Genetics Group:

The group felt that the focus must be on users, and it must be broad, to support the full range of biomedical research from basic to clinical. In the beginning, it might be smart to target specific groups, but it would be a mistake not to plan ahead to encompass a wide range.

The group came up with 10 core questions. The first four are experiment driven, and the rest are more practical and came from NIH representatives, who hear these questions.

1. Hypothesis testing of disease mechanisms: What model can I use to examine decreased activity of a biochemical pathway that I hypothesize is responsible for disease X?
2. Drug testing: What models are best for testing toxicity and efficacy of drug X?
3. Gene discovery: What human diseases have clinical signs most similar to the phenotype of the interesting new mutant I have discovered? What human genes are responsible for the disease and, hence, are potential orthologs, candidates that I've perturbed?
4. Gene discovery: I have mapped a human disease. What are the phenotypes of animal model mutant genes in this region?
5. Who are potential collaborators who can help me use this model?
6. What funding opportunities are available for using this model?
7. What technical support (i.e., husbandry, specialized equipment) is available?
8. Where are nearby cores (i.e., microarray or transgenic) that can help me?
9. What is needed to set up and maintain this model?
10. How can I compare anatomical structures, tissues, phenotypes, physiology, etc., between models? How can I get better support for cross-species comparisons?

Preclinical/Drug Development Group:

The group described four areas in defining animal models as surrogates for humans, and explored options, refinement and reconfiguration of models, how to facilitate collaboration in use, and documentation of results, both positive and negative.

A number of models or scenarios are possible. Type 2 diabetes, using the desert rat model, is one example of an animal that has a specific gene network that makes it susceptible to the disease. Hypertension, which is complex with many components need to be understood, presents a big informatics challenge. Other possibilities might include Parkinson's disease, Alzheimer's disease, cancer, neurodegeneration associated with memory disorders, heart disease and stroke, infectious disease, metabolic disease.

The group discussed the uses, tasks, and needs of the above models:

Uses discussed included: facilitation of preclinical investigations; provision of tools to help enable access to models at resources, to validate appropriateness of model resources, and to broker collaborations in use of models or their refinement; to document needs for and provide blueprints to build better animal models; to educate on lessons learned and pitfalls; to facilitate basic research into mechanisms; and to assist in deciding when it is appropriate to outsource.

Specific tasks to be considered include: identifying who will build, expand, integrate, and maintain ontologies; what are and who will provide the annotation tools; and what are the workflows?

This resource would need to be provided with: outcomes/responses, e.g. differential responses to drugs or other treatments; multidimensional, cross-referenced disease-to-model-to-disease ontologies; a series of rules for scientists to evaluate what they are doing, e.g. strain differences and genetic, environment, temporal impacts on models; regulatory requirements; translational opportunities such as funding, collaborations, and public/private partnerships; multidimensional model characterization (physiological, biochemical, pathological, behavioral); and a means to integrate with other resources (e.g., biochemical pathways).

Discussion

The disease focus would have to be narrowed to two or three. Any number could be listed as possible examples, but then narrowed for a demonstration activity; at least one disease should play across species. This can be approached in different ways—e.g., to enumerate diseases, enumerate questions, apply genetics questions. One approach would be to start with model organism databases to shed light on diseases.

Both diseases and questions need to be enumerated. It is important to drill down to a use case to bring interest to support what is needed. A thought-out summary of bringing animal models to IC work would be helpful. Needs shouldn't be stated with too many specifics: If questions are fairly generic across a large number of disease models, they are more likely to generate interest.

Infectious Diseases Group:

The purposes of looking at data are different for the basic and the applied research communities. The basic researcher is looking at the "ologies," or mechanisms (which ultimately leads to better models). The basic researcher wants to build a better model, to know more about how to approach a given problem. In contrast, the applied researcher is primarily interested in preventive and therapeutic issues. The applied researcher wants to use the best model that recapitulates human disease.

Types/examples of questions

1. What animal subject most closely represents the full spectrum or well-defined subset of human disease?
2. What are the parameters of the human condition?
3. Is there an animal model for that condition?
4. What aspects of human disease are reflected by/in animal?
5. Does animal pattern of response match human pattern of response?
6. What is disease phase at which model is relevant? [suggests need for model of disease]
7. What are the characteristics of this model (phenotype, clinical progression, response to different therapies)? User then decides whether/which are appropriate.
8. If trying to map mechanism, need to know phenotype; this is the clue.
9. Want to know who's doing what and what do they know.
10. What do we currently know about animal models that are involved in this in any way?
11. Is there an animal model that is relevant to this disease?
12. What strains have the given phenotype?
13. What is the current state of knowledge?

A good starting point would be a catalog of dababases with brief description of all models currently available to capture the known science for purpose of understanding fundamental mechanisms and disease pathways, detecting infection, and testing therapeutic and preventive interventions. Such a catalog would not be just a type of "Yellow Pages," but would present data that logically come together. The catalog could be create by drawing on existing databases as a start, using for example, five databases, five diseases, and five questions to ask.

Possible disease areas to focus on might include HIV/SIV, malaria, influenza, or hepatitis C. Databases that could form the core foundation could be drawn from existing databases of human, NHPs, the mouse, rats, drosophila, yeast, and worm. Specific questions to be asked would be determined on the basis of diseases and databases chosen, but might include those related to genes, pathways, phenotypes, disease frameworks (organs, systems, cells, etc.) at various levels, or pathology.

Important informatics issues include interfaces and differentia, as well as meta-data. Data need to be cataloged, indexed, described, and characterized. The GO could serve as a model. New ways to index biological knowledge are needed, that make it easier for naïve users to navigate existing knowledge and for both naïve and experienced users to mine new knowledge (phase 2).

Group Discussion

The group considered beginning with organs, systems, cells, and cell biology and then identify what cells are affected, rather than starting with a disease. Specific test questions are needed to determine what is happening at the gene level that affects the cells. It would be helpful to have something that assists investigators in laying out an organizational framework, tools they can draw on to put the genes that are found in the network, tools that will involve a molecular pathway, different ways that infectious agents get in, and potential places for intervention, because researchers need to be able to cascade through all of those tools.

Fairly extensive model organisms have been built and most have links to animal models, curated literature, and genes and gene pathways. The application should not be limited, but should set a starting point where one could imagine doing something in the near future to demonstrate capability. While there are data on many more species, little of it is well organized.

It would be valuable to be able to make the same sort of queries about humans as researchers can do in databases such as MGI: A human disease database is needed. Since many diseases have mitochondrial function as the focus of interest and much has been done to look at human mitochondria, it would be useful to map it to animal databases across a few diseases. Pathology is organized by principles (e.g., etiology, organ systems) and perhaps a database could be organized in that way. Mitochondria would cut across all systems. Looking at disease from different angles (e.g., cellular basis, immune basis, morphological basis) is critical to medicine.

DAY 2 BREAKOUT GROUPS:

Workshop participants reconfigured themselves into three breakout groups, to discuss disease models as the unifying themes and develop use cases. One group looked at the diversity of models for blood pressure issues, including interests related to type 2 diabetes. Another group discussed retinal degeneration, a simple system model. A third group examined neurodegenerative disease in humans and NHPs. An overarching theme should be how this is relevant to humans.

Hypertension Group:

Hypertension can be a model for a number of common complex diseases. There are several aspects of model use that should be considered in complex disease.

1. Human
    a. Clinical data/epidemiology
        i. Various forms including salt sensitive, low renin, high renin, neurogenic, vasculature, kidney, adrenals, environmental (diet)
        ii. Clusters within different ethnic groups, suggesting different backgrounds
        iii. Different susceptibilities to end-organ disease
    b. Genetic studies: association studies—HapMap and single gene. There will be an explosion of these studies and researchers must think about defining inclusion and exclusion criteria;

affected sibling pairs, QTLs available; Mendelian forms (14); end-organ sequelae (eye, renal, cardiac, vessels).
   c. Define the phenotypes clinically.

2. Animals
   a. Different species: human, rats, and mouse
   b. Types: surgical (e.g., putting a ligature around renal artery), genetic selection—strains in rats and mice, genetic, modified
   c. Ways to phenotype: tail cuff, direct with catheter, direct with telemetry
   d. Drug responses in different models
   e. Mechanisms of hypertension in different models (brain, kidney, vasculature, adrenal) and linking back to what can be inferred in human
   f. Genetic mapping: more than 200 QTLs (multiple species)

3. Ways to integrate
   a. Look at pathways implicated in hypertension, e.g. the renin angiotensin system (gene specific using approaches such as proteomics or microarray.
   b. Phenotype correlations with responses
   c. Use of known genes
   d. QTLs via comparative genomics
   e. Phenotypic response to environmental stressors, e.g., diet, hypoxia, toxins, infectious diseases
   f. Pathology correlations
   g. End-organ damage related to clinical sequelae
   h. Ontologies

4. Questions to ask
   a. Gene to phenotype
   b. Clinical phenotype to gene
   c. Mutant phenotype to clinical phenotype; e.g., insertional transgenic causes hypertension
   d. Phenotype correlations—biochemistry, pathology, pharmacological responses
   e. Clinical subtypes—need better information, can animal models help?
   f. Is end-organ damage dependent on a “flavor” of hypertension
   g. Longitudinal changes—correlation with disease, timing
   h. Gender responses, huge differences seen in timing and end-organ damage

5. Database resources
   a. Animal model databases
   b. Programs for Genomic Applications
   c. HapMap
   d. Online Mendelian Inheritance in Man database
   e. PubMed for humans
   f. Clinical records: study specific, clinical trials, FDA
   g. NCBI, European Molecular Biology Laboratory (sequencing information)
   h. UniProt (proteomics, Universal Protein Resource), e.g., protein-protein interaction and I.G. Bind
   i. Pathway databases, e.g., Reactome, Kyoto Encyclopedia of Genes and Genomes (more than 200 pathway databases)

j. Transcriptome databases: GEO, Gene Express, Novartis (GNF), Transfac, animal repositories.

This exercise points to the value of concentrating on a disease and the focus it provides. Researchers who work with nonhumans would like to see easy access to human samples.

Retinal degenerative diseases:

Many diseases have a silent early etiology, which points to the need for early diagnostics, before damage is done, and methods for early intervention. Most diseases are multigenic and each gene can be part of a pathway. This adds complexity, but also provides an opportunity because there are multiple sites for diagnosis and treatment.

The group proposed questions in two areas, preclinical and basic research:

Use cases—preclinical:

1. What models have phenotypes most similar to the clinical signs of this disease? This is important for gene discovery, therapeutics, etc. For example, the mouse does not have a macula, limiting its usefulness as a model for eye diseases.
2. What models have mutations in these pathways? This is important for etiology studies and drug testing.
3. What models have the most similar cellular and tissue structures?

Use cases—basic research:

1. What other models are there for this disease?
2. Who else is studying models of this disease?
3. Who is developing diagnostic methods for this disease? One possibility is CRISP (Computer Retrieval of Information on Scientific Projects), a searchable database of federally funded biomedical research projects.

This subject introduces a number of cross-cutting requirements, and it would be useful to expand farther into the nervous system and think, for example. about glaucoma and other loss of vision. The questions point to the ability to articulate advantages and disadvantages of different models with a disease-specific approach.

Neurodegenerative disease:

This group chose to focus on Alzheimer's disease. They envisioned a system that could be useful to a young investigator coming into the field knowing little or to a sophisticated researcher with a targeted point of view; it could address the needs of basic and clinical scientists. A portal would link to a federation of databases and entering "Alzheimer's disease" would produce an extensive index.

Phenotypes of the human condition follow a temporal axis from early stages with mild cognitive impairment to late stage, characterized by severe dementia. These could be linked to cognition, genetics, protein misfolding, or early detection.

Useful animal models would demonstrate items such as long-term memory impairment and overproduction of amyloid protein. These would link to information on publications, researchers or clinicians, reputable curated databases, and new less known or unknown models (e.g., CRISP), with alerts to controversial models. Users would also be able to interface with information about which animal models *have not*

worked; in such cases, small changes could make a big difference. Also emphasized would be specific information about what has been done with a particular model. Clinical cases and related information would be included to help communication between basic scientists and clinical researchers.

The group emphasized the importance of developing assessment tools to demonstrate when something is successful and the cost effectiveness of the investment dollars that are required.

The group also talked about whether it was necessary to have a forum to address controversies or conflicts in data. A Web site was suggested for alternative models and *the group discussed the fact that there are many different models and the possibility of linking to databases that have information about other models*. It would be useful to try to retain information on models that do not end up in the public domain.

V. Recommendations

1) Develop a catalog of current animal model databases with consistent meta-data and annotations. Priorities in catalog development will be partially driven to accommodate models needed for prototype in 2).

2) Launch the effort from a disease category or specific disease
    a) Choose a prototypic effort which will be scalable
    b) Narrow the initial focus to a single, cross-cutting disease with well-defined databases
    c) Possible prototypes include neurogenerative disease and metabolic syndrome.

VI. Conclusions

A Request for Information (RFI) was issued in July 2006 (NOT-RR-06-003 – http://grants.nih.gov/grants/guide/notice-files/NOT-RR-06-003.html) to determine the scope of the task from an informatics perspective. Responses are currently being analyzed.

A Steering Committee comprising both NIH and non-NIH members will be established to determine and guide the process of developing a prototype of the resource and portal.

VII. Contact information

For more information, please contact: Dr. Harold Watson (watsonh@mail.nih.gov) or Dr. Carol Bean (cabean@mail.nih.gov).

For more information about NCRR, please visit http://www.ncrr.nih.gov.

Appendices: Workshop Agenda
Participant List

Navigating the Translational Researcher Through A Complex of Animal and Biological Resources

Natcher Conference Center
National Institutes of Health
45 Center Drive
Bethesda, MD 20892

WORKSHOP AGENDA

March 6, 2006

7:30–8:30 a.m. **Registration and refreshments**

8:30–8:40 a.m. **Welcome and National Center for Research Resources Update**
Louise Ramm, Ph.D., Deputy Director, National Center for Research Resources

8:40–8:50 a.m. **Introduction**
Harold Watson, Ph.D., National Center for Research Resources, Division of Comparative Medicine
Carol Bean, Ph.D., National Center for Research Resources, Division of Biomedical Technology

8:50–9:20 a.m. **Presentation: Animal Models and Resources - Biology**
Co-Chair, Monte Westerfield, Ph.D., Institute of Neuroscience, University of Oregon

9:20–9:40 a.m. **Presentation: Primate Models and Resources**
Stuart Zola, Ph.D., Psychiatry and Behavioral Sciences, Yerkes National Primate Research Center

9:40–10:00 a.m. **Presentation: Animal Models in Clinical and Translational Science**
Keith Cheng, M.D., Ph.D., Pathology, Biochemistry and Molecular Biology, and Pharmacology, Pennsylvania State University, Hershey Medical Center

10:00–10:30 a.m. **Break**

10:30–11:00 a.m. **Presentation: Animal Models and Resources - Informatics**
Co-Chair, Mark Ellisman, Ph.D., Neurosciences and Bioengineering, University of California San Diego, School of Medicine

11:00–11:20 a.m. **Presentation: Rat Models and Resources**
Howard Jacob, Ph.D., Director of the Human and Molecular Genetics Center, Physiology, Medical College of Wisconsin

11:20–11:40 a.m. **Presentation: Mouse Models and Resources**
Richard Woychik, Ph.D., The Jackson Laboratory

11:40–12:00 p.m. **Presentation: Animal Models and Resources - Ontologies**
Olivier Bodenreider, M.D., Ph.D., National Institutes of Health, National Library of Medicine

12:00–1:00 p.m. **Lunch**

1:00–1:20 p.m. **Charge to Breakout Groups**
Michael Katze, Ph.D., Microbiology, Washington National Primate Research Center

1:20 –2:30 p.m. **Small Group Sessions - Discussions of Questions**

(Each Group will discuss all Questions)

Group A: Primary Focus – Questions 1 and 3
Leaders: *Maryanne Martone, Ph.D., Department of Neurosciences, University of California, San Diego*
Richard Yanagihara, M.D., Departments of Pediatrics, Public Health Sciences and Epidemiology, and Tropical Medicine and Medical Microbiology, John A. Burns School of Medicine, University of Hawaii at Manoa

Group B: Primary Focus – Questions 4 and 5
Leaders: *Janan Eppig, Ph.D., The Jackson Laboratory*
Edward Spack, Ph.D., Stanford Research Institute (SRI International)

Group C: Primary Focus – Question 2
Leaders: *Simon Twigger, Ph.D., Bioinformatics/Computational Biology Medical College of Wisconsin*
Dr. Michael Katze

2:30– 3:00 p.m. **Break**

3:00 –4:00 p.m. **Continuation of Small Group Sessions**

4:00–5:00 p.m. **Integration of Small Group Responses to Questions**

March 7, 2006

8:00–8:45 a.m. **Refreshments**

8:45–9:00 a.m. **Introduction**
Dr. Monte Westerfield
Dr. Mark Ellisman

9:00–9:30 a.m. **Report and Discussion – Question1**
Dr. Mary Martone

9:30–10:00 a.m. **Report and Discussion – Question 2**
Dr. Simon Twigger

10:00–10:30 a.m. **Report and Discussion – Question 3**
Dr. Richard Yanagihara

10:30–11:00 a.m. **Break**

11:00–11:30 a.m. **Report and Discussion – Question 4**
Dr. Janan Eppig

12:00–12:30 p.m. **Report and Discussion – Question 5**
Dr. Edward Spack

12:30–1:00 p.m. **Conclusions and General Discussion**
Dr. Monte Westerfield
Dr. Mark Ellisman

1:00–1:05 p.m. **Thank you and Goodbye**
Dr. Carol Bean
Dr. Harold Watson

March 2, 2006

National Center for Research Resources (NCRR),
Division of Comparative Medicine (DCM) and Division of Biomedical Technology (DBT)
Navigating the Translational Researcher Through
A Complex of Animal and Biological Resources
March 6–7, 2006

PARTICIPANTS LIST

Kristin M. Abraham, Ph.D.
Program Director
Division of Diabetes Endocrinology & Metabolic Diseases
National Institute of Diabetes and Digestive and Kidney Diseases
National Institutes of Health
6707 Democracy Boulevard
Room 795
Bethesda, MD 20892
Phone: 301-451-8048
Fax: 301-480-0475
E-mail: abrahamk@extra.niddk.nih.gov

Krishan K. Arora, Ph.D.
Health Scientist Administrator
Division of Research Infrastructure
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard
Room 938
Bethesda, MD 20892
Phone: 301-435-0763
Fax: 301-480-3770
E-mail: arorak@mail.nih.gov

Carol A. Bean, Ph.D.
Program Director
Division of Biomedical Technology
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard
Room 972
Bethesda, MD 20892
Phone: 301-435-0755
E-mail: cabean@mail.nih.gov

Andrea Beckel-Mitchener, Ph.D.
Chief
Functional Neurogenomics Program
National Institute of Mental Health
National Institutes of Health
6001 Executive Boulevard
Bethesda, MD 20892
Phone: 301-443-5288
E-mail: amitchen@mail.nih.gov

Olivier Bodenreider, M.D., Ph.D.
Staff Scientist
Lister Hill Center
National Library of Medicine
8600 Rockville Pike
38A/B1N28U
Bethesda, MD 20894
Phone: 301-435-3246
Fax: 301-480-3035
E-mail: olivier@nlm.nih.gov

Carol Bult, Ph.D.
Staff Scientist and Associate Professor
Mouse Genome Informatics
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609
Phone: 207-288-6324
E-mail: carol.bult@jax.org

Brian A. Canada
Graduate Student
Intergrative Biosciences
Pennsylvania State University
310 IST Building
University Park, PA 16802
Phone: 814-865-6168
E-mail: canada@psu.edu

Michael C. Chang, Ph.D.
Health Science Administrator
Division of Comparative Medicine
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard
Bethesda, MD 20892
Phone: 301-435-0750
E-mail: changmic@mail.nih.gov

Keith C. Cheng, M.D., Ph.D.
Associate Professor
Department of Pathology
Jake Gittlen Cancer Research Foundation
Pennslyvania State College of Medicine
500 University Drive
Room C7866A
Hershey, PA 17033
Phone: 717-531-5635
Fax: 717-531-5634
E-mail: kcheng76@gmail.com

Muriel T. Davisson, Ph.D.
Senior Staff Scientist
Director of Genetic Resources
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609
Phone: 207-288-6223
Fax: 207-288-6149
E-mail: muriel.davisson@jax.org

Nancy Desmond, Ph.D.
Program Officer
Division of Neuroscience
National Institute of Mental Health
National Institutes of Health
6001 Executive Boulevard
Room 7197, MSC 9645
Bethesda, MD 20892
Phone: 301-443-3563
E-mail: ndesmond@nih.gov

Linda C. Duffy, Ph.D.
Scientific Review Administrator
Office of Review
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard
Suite 1082
Bethesda, MD 20892
Phone: 301-435-0810
E-mail: duffyl@mail.nih.gov

Thomas B. Elliott, Ph.D.
Research Microbiologist
Department of Defense
Scientific Research Department
Armed Forces Radiobiology Research Institute
8901 Wisconsin Avenue
Building 42
Bethesda, MD 20889
Phone: 301-295-0898
Fax: 301-295-4078
E-mail: elliott@afrri.usuhs.mil

Mark Ellisman, Ph.D.
Professor and Director
Neurosciences and Bioengineering
Biomedical Informatics Research Network - Coordinating Center
Center for Research in Biological Systems
University of California, San Diego
9500 Gilman Drive
MC 0608, BSB 1000
La Jolla, CA 92093
Phone: 858-534-2251
Fax: 858-534-7497
E-mail: gosborne@ncmir.ucsd.edu

Janan T. Eppig, Ph.D.
Senior Staff Scientist
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609
Phone: 207-288-6422
Fax: 207-288-6132
E-mail: jte@jax.org

Donald Fine, Ph.D
Senior Director
Manufacturing, Testing, and Technical Services
DynPort Vaccine Company
64 Thomas Johnson Drive
Frederick, MD 21702
Phone: 301-607-5030
E-mail: dfine@csc.com

Colin F. Fletcher, Ph.D.
Program Director
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane
Bethesda, MD 20850
Phone: 301-451-1340
E-mail: fletcherc2@mail.nih.gov

Daniel Gallahan, Ph.D.
Associate Director
Division of Cancer Biology
National Cancer Institute
National Institutes of Health
6130 Executive Boulevard
Building EPN/Suite 5005A
Rockville, MD 20852
Phone: 301-435-5226
Fax: 301-48-2854
E-mail: dg13w@nih.gov

Mamta Gautam-Basak, Ph.D.
Scientific Review Administrator
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard Room 1066
Bethesda, MD 20892
Phone: 301-435-0965
E-mail: mg574v@nih.gov

Jonathan A. Glock
Health Specialist
Sexual Transmitted Infections Branch
Division of Microbiology and Infectious Diseases
National Institute of Allergy and Infectious Diseases
National Institutes of Health
6610 Rockledge Drive
Bethesda, MD 20817
Phone: 301-402-2140
Fax: 301-480-3617
E-mail: jglock@niaid.nih.gov

John Glowa, Ph.D.
Scientific Review Administrator
Office of Review
National Center for Research Resources
National Institutes of Health
5515 Charles Street
Bethesda, MD 20814
Phone: 301-435-0807
Fax: 301-480-3660
E-mail: glowaj@mail.nih.gov

Peter Good, Ph.D.
Program Director
Division of Extramural Research
National Human Genome Research Institute
National Institutes of Health
5635 Fishers Lane
Suite 4076
Bethesda, MD 20892
Fax: 301-480-2770
E-mail: goodp@mail.nih.gov

John D. Harding, Ph.D.
Director
Primate Resources
Division of Comparative Medicine
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard
Bethesda, MD 20892
Phone: 301-435-0776
Fax: 301-480-3819
E-mail: hardingj@mail.nih.gov

Jonathan Horsford, Ph.D.
Program Director
National Insitutes for Neurological Disorders and Stroke
National Institutes of Health
6001 Executive Boulevard
Bethesda, MD 20852
Phone: 301-496-5745
E-mail: horsforj@ninds.nih.gov

Howard J. Jacob, Ph.D.
Director and Associate Section Chief of Genetics
Human and Molecular Genetics Center
Children's Hospital of Wisconsin
Medical College of Wisconsin
HMGC - HRC 5th Floor
8701 Watertown Plank Road
Milwaukee, WI 53226
Phone: 414-456-4887
Fax: 414-456-6516
E-mail: jacob@mcw.edu

Michael G. Katze, Ph.D.
Professor
Department of Microbiology
University of Washington
Box 358070 Katze Lab
Seattle, WA 98195
Phone: 206-732-6135
Fax: 206-732-6056
E-mail: honey@u.washington.edu

Sonnie Kim, M.S.
Program Officer
Respiratory Diseases Branch
Division of Microbiology and Infectious Diseaseas
National Institutes of Health
6610 Rockledge Drive
Room 5034
Bethesda, MD 20852
Phone: 301-496-5305
E-mail: skim@niaid.nih.gov

Richard A. Knazek, M.D.
Contractor
Division for Clinical Research Resources
National Center for Research Resources
National Institutes of Health
One Democracy Plaza, 9th Floor
6701 Democracy Boulevard
Bethesda, MD 20892
Phone: 301-435-0792
Fax: 301-480-3661
E-mail: knazekr@mail.nih.gov

Francesca Macchiarini, Ph.D
Program Officer
Division of Allergy, Immunology, and Transplantation
National Institute of Allergy and Infectious Diseases
National Institutes of Health
6610 Rockledge Drive
Room 3070
Bethesda, MD 20892
Phone: 301-496-7551
E-mail: fmacchiarini@niaid.nih.gov

Maryann Martone, Ph.D.
Professor in Residence
Neuroscience
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093
Phone: 858-822-0745
Fax: 858-822-0828
E-mail: mmartone@ucsd.edu

Shelia McClure, Ph.D
Health Scientist Administrator
Division of Research Infrastructure
National Center for Research Resources
National Institutes of Health
One Democracy Plaza, 9th Floor
6701 Democracy Boulevard
Bethesda, MD 20892
Phone: 301-451-6536
E-mail: mcclursh@mail.nih.gov

Lori A. Mulligan, M.P.H
Director
Office of Science Policy and Public Liaison
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard
Room 984
Bethesda, MD 20852
Phone: 301-435-0897
E-mail: mulligal@mail.nih.gov

Nancy L. Nadon, Ph.D.
Director
Office of Biological Resources
National Institute on Aging
National Institutes of Health
7201 Wisconsin Avenue
GW 2C231
Bethesda, MD 20892
Phone: 301-402-7744
Fax: 301-402-5997
E-mail: nadonn@nia.nih.gov

Ray O'Neill, Ph.D.
Health Scientist Administrator
Division of Comparative Medicine
National Center for Research Resources
National Institutes of Health
One Democracy Plaza, 9th floor
6701 Democracy Boulevard
Bethesda, MD 20892
Phone: 301-435-0749
E-mail: oneillr@mail.nih.gov

John Quackenbush, Ph.D.
Professor
Biostatistics and Computational Biology
Department of Biostatistics
Dana-Farber Cancer Institute
44 Binney Street
Room M232
Boston, MA 02115
Phone: 617-582-8163
E-mail: johnq@jimmy.harvard.edu

William F. Rall, Ph.D.
Health Scientist Administrator
Division of Comparative Medicine
National Center for Research Resources
National Institutes Of Health
6701 Democracy Boulevard
Room 946
Bethesda, MD 20892
Phone: 301-434-0744
Fax: 301-480-3819
E-mail: rallw@mail.nih.gov

Louise E. Ramm, Ph.D.
Deputy Director
National Center for Research Resources
National Institutes of Health
9000 Rockville Pike
Bethesda, MD 20892
Phone: 301- 435-8079
E-mail: rammL@mail.nih.gov

Lorenzo M. Refolo, Ph.D.
Program Director
Neurodegeneration
National Institute of Neurological Disorders and Stroke
National Institutes of Health
6001 Executive Boulevard
NSC 2223
Rockville, MD 20852
Phone: 301-496-5446
Fax: 301-496-1080
E-mail: refolol@ninds.nih.gov

Daniel Rosenblum, M.D.
Medical Officer
Division for Clinical Research Resources
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard
Room 915
Bethesda, MD 20892
Phone: 301-435-4051
Fax: 301-480-3661
E-mail: rosenblumd@mail.nih.gov

Denise A. Russo, Ph.D.
Program Director
Division of Metabolism and Health Effect
National Institute on Alcohol Abuse and Alcoholism
National Institutes of Health
5635 Fishers Lane
Bethesda, MD 20892
Phone: 301-402-9403
Fax: 301-594-0673
E-mail: drusso@mail.nih.gov

Margaret Snyder, Ph.D.
Health Scientist Administrator
Office of Extramural Programs
Office of Extramural Research
Office of the Director
National Institutes of Health
6705 Rockledge Drive
Room 4184
Bethesda, MD 20892
Phone: 301-402-1058
Fax: 301-480-3530
E-mail: snyderm@mail.nih.gov

Edward G. Spack, Ph.D.
Senior Director, PharmaSTART
Biosciences
SRI International
333 Ravenswood Avenue
PN385
Menlo Park, CA 94025
Phone: 650-859-3064
Fax: 650-859-3041
E-mail: edward.spack@sri.com

James P. Stables,
Program Director
National Institutes of Neurological Disorders and Stroke
National Institutes of Health
6001 Executive Boulevard
Rockville, MD 20892
Phone: 301-496-1846
E-mail: stablesj@ninds.nih.gov

Danilo A. Tagle, Ph.D.
Program Director in Neurogenetics
Neurogenetics
Division of Extramural Research
National Institutes of Neurological Disorders and Stroke
National Institues of Health
Neuorscience Center, Room 2133
6001 Executive Boulevard
Bethesda, MD 20892
Phone: 301-496 5745
Fax: 301-402 1501
E-mail: tagled@ninds.nih.gov

Chris Taylor, Sc.D.
Program Officer
Respiratory Diseases Branch
National Institute of Allergy and Infectious Diseases
National Institutes of Health
6610 Rockledge Drive
Room 5045
Bethesda, MD 20892
Phone: 301-496-5305
Fax: 301-496-8030
E-mail: ct18m@nih.gov

Simon N. Twigger, Ph.D.
Assistant Professor
Physiology
Human and Molecular Genetics Center
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI 53226
Phone: 414-456-8802
Fax: 414-456-6595
E-mail: simont@mcw.edu

Harold L. Watson, Ph.D.
Health Scientist Administrator
Division of Comparative Medicine
National Center for Research Resources
National Institutes of Health
6701 Democracy Boulevard
Room 944
Bethesda, MD 20892
Phone: 301-435-0884
Fax: 301-480-3819
E-mail: watsonh@mail.nih.gov

Monte Westerfield, Ph.D.
Professor and Director
Institute of Neuroscience
University of Oregon
1254 University of Oregon
Eugene, OR 97403
Phone: 541-346-4607
Fax: 541-346-4548
E-mail: monte@uoneuro.uoregon.edu

Scott B. Winram, Ph.D.
Senior Project Officer
Office of Biodefense Research Affairs
Division of Microbiology and Infectious Diseases
National Institute of Allegy and Infectious Diseases
National Institutes of Health
6610 Rockledge Drive MSC 6604
Bethesda, MD 20892
Phone: 301-451-4806
Fax: 301-480-1263
E-mail: swinram@niaid.nih.gov

Richard P. Woychik, Ph.D.
Director
The Jackson Laboratory
600 Main Street
Bar Harbor, ME 04609
Phone: 207-288-6041
Fax: 207-288-6041
E-mail: rick.woychik@jax.org

Richard Yanagihara, M.D.
Professor of Pediatrics
John A. Burns School of Medicine
University of Hawaii at Manoa
651 Ilalo Street BSB 320L
Honolulu, HI 96813
Phone: 808-692-1610
Fax: 808-692-1976
E-mail: yanagiha@pbrc.hawaii.edu

Jane Ye, Ph.D
Health Scientist Administrator
National Heart, Lung, and Blood Institute
National Institutes of Health
6701 Rockledge Drive
Bethesda, MD 20892
Phone: 301-435-0513
E-mail: yej@nhlbi.nih.gov

Stuart M Zola, Ph.D.
Director
Yerkes National Primate Research Center
Emory University
954 Gatewood Road NE
Atlanta, GA 30329
Phone: 404-727-7707
Fax: 404-727-0623
E-mail: szola@rmy.emory.edu