An action plan for mouse genomics

James Battey[1], Elke Jordan[2], David Cox[3] & William Dove[4]

The mouse has become the leading animal model for studying biological processes in mammals. Creation of additional genomic and genetic resources will make the mouse an even more useful model for the research community. On the basis of recommendations from the scientific community, the National Institutes of Health (NIH) plans to support grants to generate a 'working draft' sequence of the mouse genome by 2003, systematic mutagenesis and phenotyping centres, repositories for mouse strain maintenance, distribution and cryopreservation and training fellowships in mouse pathobiology.

Over the last two decades, major technological advances have enabled scientists to manipulate the mouse genome in a targeted and predictable fashion. There is now a rich collection of mouse strains harbouring naturally occurring, ENU-induced or targeted gene mutations. These resources, coupled with sophisticated genetics enabling gene identification and analysis of gene function and interactions, have made the mouse a key research tool for the study of normal and disordered biological processes in mammals. The value of the mouse as a model could be further enhanced if more complete resources were available, such as high-resolution maps and genomic sequence, an increased number and length of mouse DNA sequences, additional mouse mutants covering a wide scope of phenotypes, increased capacities of repositories for preserving useful mouse strains and more veterinarians, physicians and scientists trained to study mouse pathobiology, physiology and behaviour.

In March 1998, the Director of the NIH, Harold Varmus, invited a distinguished panel of international scientists, co-chaired by David Cox (Stanford University) and William Dove (University of Wisconsin), to make recommendations to the NIH regarding priorities for generating mouse genomics and genetics resources. About 60 scientists came to Bethesda, Maryland, USA, for three days to discuss these priorities. They recommended NIH support for resource development in four areas: structural analysis of the mouse genome, functional analysis of mouse biology, methods and facilities for storing and distributing interesting mouse strains and training in mouse pathobiology. The full report of this meeting can be obtained from the NIH web site (http://www.nih.gov/welcome/director/reports/mgenome.htm). In this article, the current plans for implementing these four key recommendations are described.

A Trans-NIH Mouse Genomics and Genetics Resources Coordinating Group, co-chaired by Elke Jordan (Deputy Director, National Human Genome Research Institute, NHGRI) and James Battey (Director, National Institute on Deafness and Other Communication Disorders), was created to plan the implementation of these recommendations. This Coordinating Group has representation from each of the Institutes and Centers of the NIH. After taking an inventory of ongoing activities, new initiatives were proposed which together addressed the four goals. On 5 October 1998, a subset of the extramural scientists who formulated the March 1998 report were invited back to Bethesda to review plans for implementation of the recommendations, and to provide advice on how well these plans would meet the needs of the research community. Representatives from the Department of Energy (DOE, USA) and the Medical Research Council (MRC, UK) were also present to coordinate plans for generating new mouse genomics and genetics resources. The NIH aims to coordinate its plans for mouse resource development with any related efforts supported by other funding agencies in the US and abroad.

Structural analysis of the mouse genome

The mouse genomic sequence is a critical resource for all areas of mouse research, and it will be of immediate benefit to have the sequence available as soon as possible. A consensus for the analysis of the mouse genome has emerged that will enable an integrated map of the genome to be developed in parallel with the assembly of a significant amount of sequence information. A strategy for rapidly generating a 'working draft' of the mouse genome sequence by 2003 was established. This strategy will serve the immediate needs of the research community in their efforts to identify mutations in genes and generate a useful intermediate on the way to a finished mouse genome sequence of quality comparable to the human genome sequence.

The first part of this strategy will involve characterization of a BAC library to be used for sequencing. The BAC library will be generated from the C57BL/6J strain, with an average insert size of approximately 200 kb and about 15-fold coverage of the mouse genome. There was considerable debate in the research community about which mouse strain should be used to construct the library and ultimately serve as the index mouse genome sequence. At the 5 October meeting, the extramural advisors unanimously decided that the best choice was the C57BL/6 strain. The rationale for selecting C57BL/6 rather than other strains (such as 129s, for example) includes confidence of strain derivation, widespread use among the research community and favourable breeding characteristics.

The library will be characterized by fingerprinting and BAC-end sequencing of all clones. The goal of this characterization is not to construct comprehensive sequence-ready maps, but rather

[1]*National Institute on Deafness and Other Communication Disorders, NIH Building 31, 9000 Rockville Pike, Bethesda, Maryland 20892, USA.* [2]*National Human Genome Research Institute, NIH Building 31, 9000 Rockville Pike, Bethesda, Maryland 20892, USA.* [3]*Departments of Genetics and Pediatrics, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305, USA.* [4]*McArdle Laboratory for Cancer Research, Madison Medical School, University of Wisconsin, 1400 University Avenue, Madison, Wisconsin 53706–1599, USA. Correspondence should be addressed to J.B. (e-mail: batteyj@nidcd.nih.gov).*

to use the fingerprints to construct initial contigs and to generate BAC-end sequences to perform subsequent walking. In parallel, a high-resolution radiation hybrid panel (with resolution of 100–300 kb) will be created and anchored to existing maps by cross-mapping with approximately 15,000 existing markers. This radiation hybrid mapping resource will be used for subsequent anchoring of BAC clones or contigs as they are sequenced.

The second part of the strategy will consist of genomic sequencing. This can be divided conceptually into three phases. The first is a seeding phase entailing shotgun sequencing of BAC contigs and clones, chosen both randomly and on the basis of biological interest. This phase will continue until a substantial portion of the genome is covered. The optimal amount of the genome to cover will depend on the size of the contigs and the remaining gaps, but it is expected that it will be greater than 50%. Second, a walking phase will be undertaken in which BACs will be selected using BAC-end sequences to extend contigs and close gaps. This phase will probably continue until at least 90–95% of the genome is covered. Finally, a targeted closure phase, possibly involving specialized strategies, will be carried out to close any remaining gaps.

It will not be necessary to assign specific regions of the genome to specific groups in the seeding phase. Instead, various groups will contribute to the overall effort provided that a central computer server is developed to maintain current information about library characterization and clones undergoing sequencing. The sequence produced is expected to consist of ordered and oriented sequence contigs with few gaps, serving as a useful intermediate step towards producing finished sequence at a later point. Mouse sequence data will be released to public databases within 24 hours of their generation, as is the case for public human genome sequencing efforts. This plan has the significant advantage of rapid generation of a well-characterized BAC resource as well as large quantities of genomic sequence, both of which will be valuable to investigators engaged in positional cloning and the cloning of new genes structurally related to known genes. In addition, much will be learned about gene regulatory sequences from comparative sequence analysis of the human and mouse genome. To take full advantage of this opportunity, powerful new tools for comparative sequence analysis need to be developed; many institutes are presently supporting research to improve these methods.

At present, NIH-supported human sequencing centres in the US are permitted to use up to 10% of their capacity for mouse sequencing in selected regions of the mouse genome. To generate the capacity needed to support the latest initiative, new sequencing centres focused primarily on mouse genome sequencing will be needed, as will expansion of the capacity of existing centres. The NHGRI expects to fund grants to initiate systematic mouse genomic sequencing in fiscal year 1999. The funds expected to be available will allow production of a 'working draft' sequence of each BAC with about fivefold coverage by 2003 for an estimated total cost of $126 million. More complete coverage will be possible if additional funding or increased sequencing efficiencies become available. Coordination with mouse genome sequencing initiatives supported by the DOE and MRC will enhance NIH-supported sequencing capacity and expedite completion of this initiative.

International efforts are currently underway to map about 33,000 ESTs on a low-resolution radiation hybrid panel. The mouse is a powerful model for obtaining a complete collection of cDNA libraries, because libraries can be constructed from many different tissues at varying stages of development, including the brain, as well as a spectrum of different tumour types. Several institutes have planned efforts to generate tens to hundreds of thousands of mouse ESTs from these libraries over the next few years.

A trans-NIH group has been established to coordinate the development of techniques needed to generate cDNA libraries with a higher proportion of full-length clones. Isolation of intact mRNA is an important prerequisite for generating cDNA libraries enriched in full-length clones; rapid dissection of mouse tissues will facilitate rapid mRNA extraction, leading to a high proportion of full-length templates for first-strand cDNA synthesis. A wealth of new mouse cDNAs, coupled with advances in cDNA library construction methods, will help to overcome long-standing barriers to gene discovery and characterization.

The funds expected to be available will allow production of a 'working draft' sequence of each BAC with about fivefold coverage by 2003 for an estimated total cost of $126 million.

NIH's Center for Inherited Disease Research is collaborating with Perkin-Elmer and intramural scientists to assemble an initial panel of 320 simple sequence length polymorphism (SSLP) markers evenly spaced throughout the mouse genome, which will be used to genotype 40 mouse strains commonly used for gene-mapping studies. This resource will facilitate mapping of genes of interest and quantitative trait loci in strains other than C57BL/6. It is anticipated that this panel of markers will be useful for low-resolution (approximately 10 cM) mapping of mutant genes, and will be expanded in the future to provide a more comprehensive panel of markers, including single nucleotide polymorphisms (SNPs), that will facilitate high-resolution mapping efforts. In addition to generating polymorphic markers, the NIH is supporting research to construct, maintain and distribute panels of consomic mouse strains, providing an additional resource to facilitate gene mapping.

Functional analysis of mouse biology

Mutant mouse strains provide new and critical insights into the molecular mechanisms governing normal and disrupted biological processes. Two centres outside the USA are currently aimed at increasing the number of mutant strains by mutagenesis using N-ethyl-N-nitrosourea followed by phenotyping: the MRC Mammalian Genetics Unit, Harwell, UK (http://www.mgc.har.mrc.ac.uk/mutabase/) and the ENU-Mouse Mutagenesis Screen Project, Neuherberg, Germany (http://www.gsf.de/isg/groups/enu/enu_cpt.html). New NIH-supported mutagenesis and phenotyping centres are being planned to further the number of mutant strains available to the research community. These centres will be supported by one or more NIH institutes that will issue a Request for Applications during the fiscal year 1999. Each centre will employ a distinct mutagenesis strategy, followed by a battery of tests to detect abnormalities in one or more organ systems or during development (for example, the central nervous system, the immune system and musculoskeletal, cardiovascular and pulmonary development). Additional targeted phenotype determination will be incorporated into each centre to the maximum extent possible to accommodate specialized goals (for example, identification of mouse models for visual or hearing impairment). The creativity and merit of centre grant applications will be evaluated by initial peer review to establish the details of a particular strategy for mutagenesis and phenotype determination for a given centre. Grant applicants will be invited to develop improved high-throughput screening assays ideally suited for the initial detection of mutants arising in the mutagenesis and phenotyping centres. Applicants will be required to describe plans

for developing coordinated databases to allow the research community access to information on strains and to outline plans for the provision of mutant strains to the research community in a timely fashion.

Facilities for storing and distributing mouse strains

As mutagenesis and phenotyping centres create hundreds to thousands of interesting new mouse mutant strains, the need for mouse strain repositories will grow substantially. A Request for Applications will be published in fiscal year 1999, soliciting applications to create several regional centres to expand resources for depositing and distributing mouse strains, as well as conducting research to improve the current techniques for cryopreserving mouse sperm, eggs and embryos. These centres will be viewed as a complement to the excellent service currently provided by the Jackson Laboratory. Some of the activities of the new mouse resource centres may overlap with the activities in mutagenesis and phenotyping centres; coordination of plans will be essential as these two types of centres are formed.

Training in mouse pathobiology

There is a need for scientists with expertise in mouse pathology, physiology and behaviour to collaborate with geneticists in characterizing new and existing mouse mutants. This need is predicted to expand as new mutagenesis and phenotyping centres generate an extensive number of new mutant strains. NIH Fellowship and Career Development awards will be offered so that entry and mid-career level scientists, physicians and veterinarians can obtain training in evaluating mouse pathology, physiology and behaviour. In all likelihood, some fellowships will be developed in conjunction with mutagenesis and phenotyping centres, although support of training in pathobiology outside the centres will also be provided. In addition to training opportunities in mouse pathobiology, the Jackson Laboratory currently offers a course in cryopreservation methodology.

A team effort

Overall, this plan of action emphasizes a balance between functional investigations and structural analysis of the mouse genome. Advice from distinguished scientists in the research community has played a major role in determining NIH priorities for developing mouse genomics and genetics resources. A productive dialogue with the research community should continue as the NIH implements this exciting series of new initiatives. A website (http://www.nih.gov/science/mouse) has been established on which information relevant to mouse genomics and genetics resources, including Requests for Applications relevant to implementing these recommendations, will be posted as the NIH develops them. Additional meetings will be held to allow members of the extramural research community to continue to provide advice on any needed course correction as implementation of initiatives addressing the recommendations moves forward.