

Meeting report

2001: a mouse genome odyssey

David R Beier and Bruce J Herron

Address: Genetics Division, Brigham and Women's Hospital and Harvard Medical School, 20 Shattuck St, Boston, MA 02115, USA.

Correspondence: David R Beier. E-mail: beier@rascal.med.harvard.edu

Published: 31 January 2002

Genome Biology 2002, **3**(2):reports4005.1-4005.4

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2002/3/2/reports/4005>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

A report on the 15th International Mouse Genome Conference, Edinburgh, UK, 21-24 October 2001.

The mouse genome sequence

In a year notable for the completion of a draft human genome sequence, it was appropriate that a major topic at the 15th International Mouse Genome Conference was a discussion of the status of the mouse genome project. Details of the public-domain effort were presented by John McPherson (Washington University, St Louis, USA), Kerstin Lindblad-Toh (Whitehead Institute, Cambridge, USA), Shaying Zhao (The Institute for Genomic Research, Rockville, USA), Anne-Marie Mallon (Medical Research Council Mouse Genome Centre and Mammalian Genetics Unit (MRC MGU/MGC), Harwell, UK), and Jim Thomas (National Human Genome Research Institute, National Institutes of Health (NIH), Bethesda, USA). The approach being taken combines sequencing of bacterial artificial chromosomes (BACs) - this was the main method for sequencing the human genome - and whole-genome shotgun sequencing, which is advocated by investigators at Celera Genomics. A fingerprinted physical map of 300,000 BAC clones has been generated and aligned to the mouse radiation-hybrid map using sequence-tagged sites (STSs) derived from microsatellite sequences and expressed sequence tags (ESTs). The precise position of the BACs on the map is being refined with the help of the sequences of the ends of many of the BACs. Nearly threefold (3x) coverage of the genome using shotgun sequence has been completed and deposited in public domain databases.

One aim of the public mouse genome-sequencing effort is to generate shotgun-sequence reads from different mouse strains, in order to identify single-nucleotide polymorphisms (SNPs). (As plans for this component are in development, we

propose that the validation and characterization of SNPs in a large number of commonly used inbred strains is as important as SNP discovery itself, and urge that appropriate resources be devoted to it). An important part of genomics is the informatics required for data management and analysis; programs for organizing and annotating mouse genome sequence in the Ensembl project [<http://mouse.ensembl.org/>] and at the National Center for Biotechnology Information (NCBI [http://www.ncbi.nlm.nih.gov/genome/guide/M_musculus.html]) were presented by Tim Hubbard (Sanger Centre, Hinxton, UK) and Deanna Church (NCBI, NIH, Bethesda, USA), respectively.

The other sequencing effort for the mouse genome is the proprietary sequence completed by Celera Genomics, described by Gene Myers (Celera Genomics, Rockville, USA). Myers and colleagues have assembled a reasonably complete mouse sequence using only 5.3x shotgun sequence coverage together with a bioinformatic assembly protocol that combines analysis of sequence identity between shotgun sequence reads with 'mate-pair' information, that is, information from the pairing of reads derived from opposite ends of each clone. Although some investigators have noted errors in the assembled sequence - and it is not clear what is being done to correct these - there can be little argument that Celera's work provides an extremely valuable resource to those who have access to it. It also validates the assertion made by Craig Venter and colleagues (Celera) that assembly of complex genome sequences using only shotgun sequence data is feasible; the Celera human genome assembly was not a conclusive validation because it used public-domain BAC sequences as well as shotgun sequence.

An emerging theme in genomic science is the power of comparative analysis for identification of genes and regulatory regions. This was highlighted in a plenary presentation by Jean Weissenbach (Genoscope, Paris, France) describing progress in sequencing the freshwater pufferfish

Tetraodon nigroviridis, which has a genome about an eighth the size of that of humans or mice. In addition to the 3x sequence coverage that has been deposited in the GenBank genome sequence survey database, Weissenbach and colleagues have developed a database of 'Ecores', sequences that are conserved between pufferfish and humans and that are therefore likely to be coding regions ([<http://www.genoscope.cns.fr/proxy/cgi-bin/exofish.cgi>]).

With the rapidly progressing sequencing of the mouse genome comes the possibility of performing true genome-wide investigation using techniques that permit massively parallel analysis, such as expression profiling, similar to the kinds of studies that have now been done in yeast and *Caenorhabditis elegans*. For example, a database of the tissue-specific expression levels of murine genes, containing 2.5×10^6 independent measurements, has been generated using hybridization to Affymetrix oligonucleotide arrays (Tim Wiltshire, Genomics Institute of the Novartis Foundation, San Diego, USA). Expression profiling has also been used to identify genes that are expressed in sexually dimorphic patterns during development; these were validated using whole-mount *in situ* analysis of gonads (Lee Smith, MRC MGC/MGU, Harwell). Among other approaches to genome-wide analysis that were also presented was a large set of full-length cDNAs generated by RIKEN, which is being further explored using a high-throughput two-hybrid screening approach to identify protein-protein interactions (Harukazu Suzuki, RIKEN, Yokohama, Japan). Over 500 interactions have been identified so far. Also, Bill Stanford (Samuel Lunenfeld Research Institute, Toronto, Canada) described one of several active large-scale programs for gene trapping, in which reporter constructs are randomly inserted into the genome and are expressed only if they insert within a gene. As well as being a resource with tremendous potential for annotation of gene function in its own right, the generation of large sets of STS-tagged gene-trapped embryonic cell lines by this project also provides a means to test candidate loci identified by phenotype-driven mutagenesis screens.

Mutagenesis

The mutagenesis session did not focus on large centers with broad aims but instead highlighted innovative approaches to screens focused on a particular region of the genome or on a specific phenotype. Although new approaches to the generation of mutations were briefly discussed, it is clear that ethylnitrosourea (ENU) remains the dominant method. Monica Justice (Baylor College of Medicine, Houston, USA) opened the session with an update on the Baylor ENU screen for mutations on mouse Chromosome 11. This screen uses a balancer-chromosome strain that provides a way to readily identify mice homozygous for an engineered inversion in recessive screens, making stock maintenance straightforward. A wide variety of mutant phenotypes were described, many of which are embryonic-lethal.

Specialized phenotype screens were a major focus of the session. Karen Steel (MRC Institute of Hearing Research, Nottingham) presented ongoing screens designed to identify and classify mice with inner-ear defects arising from the mutagenesis program at the MRC MGC/MGU. Using a combination of inner-ear pathology and genetic mapping, several complementation groups were defined; several of these mapped to previously known vestibular defects in mice, but new loci were also uncovered. One example of a new locus is *Pardon*, mutants in which have malformed auditory ossicles; *Pardon* has been identified as a mutation in *Emx2*, a homeobox-containing gene.

Discovery of new mutations from large mutagenesis centers will rely heavily on methods for screening large numbers of mice robustly and efficiently. Working in collaboration with the Munich ENU center (GSF Research Center for Environment and Health), Heinrich Flaswinkel (Technische Universität, Munich, Germany) presented a method to identify mice with immunological defects through fluorescence-activated cell sorting (FACS) and enzyme-linked immunosorbent assay (ELISA) analysis. This approach was reported to identify 54 heritable phenotypes from 8,857 lines screened. John Schimenti (Jackson Laboratory, Bar Harbor, USA) introduced a novel method for identifying mice with defects leading to genome instability. Mice are irradiated and bone marrow is isolated for flow cytometry; those with defects in DNA repair pathways, for example, will have a higher concentration of micronuclei and can be readily identified.

One benefit of point mutations, which are produced by ENU, is that they often produce hypomorphic alleles and could thus complement the functional analysis of a gene for which a null mutation has already been generated. Two talks presented approaches directed at identification of more mutations in known genes from populations of mutagenized mice. Jay Vivian (University of North Carolina, Chapel Hill, USA) discussed the use of denaturing high-performance liquid chromatography (DHPLC) to identify differences in *Smad* genes in mutagenized embryonic stem (ES) cells. Smads are intracellular proteins involved in signaling through the transforming growth factor β (TGF β) family of receptors. This approach identified approximately one coding-sequence mutation per 2 kilobases of the *Smad2* gene, and when the mice were generated from the mutant ES cells, one of the mutations resulted in a phenotype similar to that of *Smad2* knockout mice. Emma Coghill (MRC MGC/MGU) presented a sequence-based approach, screening archived sperm from the Harwell ENU mutagenesis program. Her analysis provided a new allele of connexin 26 that had an embryonic-lethal phenotype.

Gene discovery

As the whole of the mouse genome sequence becomes available, the positional cloning of many interesting mutations is

being accelerated, as was evident in the array of topics discussed in the Gene Discovery session of the meeting. The successes described below were not solely due to genome-sequencing efforts, however; other emerging technologies, such as cross-species mapping and gene trapping, were also used to identify mutations.

The session was opened by Adrian Bird (Wellcome Trust Center for Cell Biology, Edinburgh, UK) who presented recent data on a new mouse model for Rett syndrome, a progressive neurological disorder that primarily affects girls. The mutated gene in this syndrome in humans encodes the DNA-binding protein MeCP2, which is known to selectively bind CpG DNA motifs; the mechanism by which its absence results in the syndrome is unclear, however. Bird presented details of the mouse null mutation that faithfully recapitulates several aspects of the human disorder in heterozygotes, including ataxia and breathing dysrhythmia.

Mouse models of polycystic kidney disease have provided much information over the years about normal kidney development. We (D.B.) described our finding that the recessive polycystic kidney mutation *jck* is in a gene encoding a novel Nek-family kinase. We have taken a new approach to validate the function of the mammalian gene: using an antisense morpholino oligonucleotide in zebrafish to reduce expression of the kinase, the cystic phenotype was recapitulated, confirming the importance of this gene in normal kidney development. A gene that could be of potential therapeutic benefit was identified by analysis of heterozygote *Wlds*⁺ mice, which show resistance to motor-neuron axon degeneration after transection (Michael Coleman, University of Cologne, Germany). A novel fusion between the ubiquitination gene *E4B* and the gene for nicotinamide mononucleotide adenylyltransferase was identified in the *Wlds* strain. Expression of a transgenic fusion gene in mice reproduced the phenotype, implying that modulation of pyridine nucleotide metabolism may protect against axon degeneration.

Rescue of mutants by BAC transgenes is another powerful technique that is becoming common practice in positional-cloning projects. This approach was used by Philomena Mburu (MRC MGC/MGU) in combination with a detailed phenotypic analysis of inner-ear hair cells to narrow down the candidate genes to be tested for mutations in *whirler* mice, which have inner-ear defects. A novel gene was presented, encoding a PDZ-domain-containing protein, that is implicated in the development of stereocilia.

Genetic studies of coat-color defects in mice have provided much information on how the genes involved might interact. For example, the products of the *dilute*, *ashen*, and *leaden* genes are all thought to function in the same biochemical pathway, transporting melanin from melanosome to pigmented hair cells. Using a candidate gene approach, Lydia

Matesic (National Cancer Institute, Frederick, USA) has found a novel effector protein for the small GTPase Rab, called Mlph, that is mutated in *leaden*.

Cross-species comparative analysis is an approach that was used by several speakers to identify candidate genes for mouse mutations. This approach has facilitated the cloning of the genes defective in *myodystrophy* mutants (Pam Grewal, University of Nottingham, UK), in *flavivirus susceptibility* mutants (Tomoji Mashimo, Institut Pasteur, Paris, France) and in a rat mutant *queue courte* (Kazuhiro Kitada, Kyoto University, Japan). Along with other strategies presented in this session, the success of this technique shows that a major bottleneck in positional cloning - gene identification - is quickly becoming resolved.

Developmental genetics

Developmental genetics has been of increasing interest at successive international mouse genome meetings. This year's session was opened by two talks that focused on epigenetic programming in cloned embryos. Ian Wilmut (Roslin Institute, Roslin, UK) presented information about the success of embryo cloning in several species. The current low success rates are attributed to inappropriate gene expression that may be due in part to a failure to maintain developmentally appropriate imprinting in transferred somatic cell nuclei. This hypothesis was supported by a presentation from Kevin Eggan (Whitehead Institute, Cambridge, USA), who described differences in imprinted gene expression between cloned and normal mouse embryos.

The imprinting theme was continued in a talk from Jesse Mager (University of North Carolina, Chapel Hill, USA), who discussed X-chromosome inactivation. The *embryonic ectoderm development* gene was shown to be important for the maintenance of inactivation of the paternal X chromosome in female mice, as well as for development of trophoblast giant cells in female embryos. This gene has been proposed to interact with a histone deacetylase in the extraembryonic tissue, selectively deactivating the paternal X chromosome.

Skin and hair defects are a group of phenotypes that have been well studied by mouse geneticists, and three talks in this session were focused on a better understanding of these phenotypes. Karen Fitch (Stanford University, Palo Alto, USA) described the classification of mice with dark skin phenotypes from the GSF mutagenesis program (Munich, Germany). The mutants were further categorized using a melanocyte-specific reporter system, mapped, and assigned to ten linkage groups. Teresa Gunn (Cornell University, Ithaca, USA) described recent work characterizing alleles of *mahogany*, which was originally identified as a coat-color mutant but alleles of which have now been found to cause behavioral abnormalities caused by a degenerative neuropathy. Lastly, we (B.H.) described a mutation in a gene

encoding a p53-binding-protein homolog in *waved3* mice, which have abnormalities in skin and heart development.

A major benefit of mouse models is their utility in providing a better understanding of disease in humans. The presentation of Kenro Kusumi (University of Pennsylvania, Philadelphia, USA), describing the role of mutations in the *Dll3* gene, which encodes a Delta-like ligand for Notch proteins, in *pudgy* mice, which have vertebral and rib malformations, underscored the usefulness of this approach. Comparison of the mouse model to the human disorder spondylocostal dysostosis provides some insight into potential neurological defects in humans, as well as into the function of this gene in normal somitogenesis.

Finally, although mapping of the loci contributing to a complex trait is often readily done, identifying the causal gene(s) remains daunting, even in a well-defined genetic system such as the mouse. Sarah Lloyd (Imperial College of Medicine at St Mary's, London, UK) reported the mapping of quantitative trait loci affecting incubation times of prion disease, and Ayo Toyé (MRC MGC/MGU) reported mapping of diabetes intermediate phenotypes. Alain Balmain (University of California, San Francisco, USA) described progress in characterizing a locus that modifies susceptibility to skin cancer and reported that haplotype analysis of susceptible and resistant strains facilitated localization to a small genetic interval.

Next year's International Mouse Genome Conference will be held in San Antonio, USA. With the imminent completion of the public-domain mouse genome sequence, the productivity of the diverse mutagenesis programs, and the increasing sophistication of genome manipulation technologies such as gene trapping and BAC transgenesis, we anticipate that many exciting discoveries in mammalian biology will be reported.