Removing the cloak of invisibility: phenotyping the mouse

Monica J. Justice

If you study a human disease, it is likely that you have tried to generate a mouse model. Sometimes, these models are excellent; others are disappointing. Or, so we think. How often does our mouse mutant not model the human disease because of limitations in how we may look at it? In any living organism, many factors work together to produce the phenotype. Here, new phenotyping paradigms for assessing mouse biology and physiology are described and proposed. Advances in mouse phenotype assessments have paralleled human clinical diagnostics. The future brings a multitude of mouse strains that might be exposed to a variety of conditions. To assess health will require the ability to perform a broad-based phenotype assessment of every animal until we can understand how the perturbation of one system affects others.

The mouse in disease research

The mouse is the leading organism for disease research because of the ability to make any type of mutation in any gene (Oliver et al., 2007). In no vertebrate other than the mouse, may one target a gene to eliminate its function in the whole organism or in one tissue, carry out forward genetic screens, or introduce transgenes as cDNAs or bacterial artificial chromosomes. However, 12 years ago, the rat was the leading organism for mammalian physiology research, primarily because of the availability of assays. Scientists realized that the primary reason for the lack of physiological assessment in the mouse was that appropriate tools had not yet been developed for this tiny mammal. A trans-NIH initiative was formed that would promote mouse genomics, while it promoted the development of phenotyping tools, adding value to the powerful genetic systems that were already available (www.nih.gov/science/models/mouse) (Battey et al., 1999).

The diagnosis of a human health condition has advantages over the mouse. First, humans have the ability to speak. If our head aches, or our stomach is upset, we can convey that information. Mice cannot. Second, medical technology has provided a multitude of tests for the human that can be used to diagnose a multitude of complaints, determine acuity of eyesight or hearing, measure fitness, or detect toxins. Morphological, biochemical and genetic deviations can even be assessed in the fetus. To model human disease, many of the same clinical diagnostic assays that would be carried out on the human should be carried out on mice. A desire for a broad phenotyping platform, which would be used to assess biology, toxicology, pharmacology and quantitative systems relevant to clinical disease, was the guiding rule for the development of phenotype assays in the mouse (Table 1). Many of the new phenotyping assays that were initially developed and validated in the mouse were based on tools that were already available in the rat, although many of these, including telemetry chips, blood assays and imaging, required miniaturization. During the same time, however, large-scale mutagenesis screens were carried out in the mouse, which required phenotype screening assays that were rapid, robust and cost-effective. Such screens allowed for the assessment of a wide variety of phenotypes in many mouse strains and mouse mutants (Hrabe de Angelis and Balling, 1998; Kasarskis et al., 1998; Schimenti and Bucan, 1998; Hardisty et al., 1999; Nolan et al., 2000; Alessandrini et al., 2001; Nelms and Goodnow, 2001; Herron et al., 2002; Peters et al., 2002; Thaug et al., 2002; Beutler et al., 2003; Kile et al., 2003; Yu et al., 2004; Crawley, 2008; Fleischmannova et al., 2008; Gaultier and Gallego, 2008). Numerous assays were attempted, and some have led to the establishment of new, high-throughput robust tests for phenotyping, whereas others have been abandoned. Nevertheless, these efforts have brought the field into a new era.

You get what you look for

Why should we use these new assays? In our current scientific climate, a graduate student or postdoctoral fellow in a disease-focused lab would tackle his/her mouse mutant by asking if it has the phenotype that the laboratory expected. If not, it is often disappointing. If it does, it is a time for celebration. But how often are important phenotypes overlooked? A homozygous knockout of the anti-apoptotic protein Bcl-x causes early embryonic death at mid-gestation as a result of massive cell death of hematopoietic cells and neurons. According to the associated report, ‘heterozygous mutant mice (bcl-x+/-) were healthy and normal in size’ (Motoyama et al., 1995). A subsequent manuscript showed that heterozygous mice had low platelet counts, and this new information revealed that the Bcl-XL isoform is part of a molecular clock for timing platelet lifespan (Mason et al., 2007). With improved phenotyping tools, many phenotypes in heterozygous mutants can be found (Yu et al., 2004). A broad phenotype assessment of the Bcl-x mutant mouse would have detected the platelet abnormalities, but how many laboratories have easy, cost-efficient access to the equipment required to perform routine complete blood counts on mice?

Prior to the evolution of new phenotyping assays, a trait might be detected as a change in coat color, a movement disorder or a morphological difference. In fact, many traits carried by strains of mice could be described as ‘invisible’ (Stevens et al., 2007). Introducing assays for blood cells, blood chemistry, pain response, behavior, infection and cardiac function, to name a few, uncovers many more phenotypes, but where does one stop? The inbred strain

1Baylor College of Medicine, One Baylor Plaza R804, Houston, TX 77030, USA (e-mail: mjustice@bcm.edu)
C3H/HeJ harbors a mutation in the Toll-like receptor 4, which causes an abnormal response to bacterial antigens such as lipopolysaccharide (Sultzter et al., 1993). The identification of this abnormal response has led to a field dedicated to understanding the body’s response to infection, primarily through phenotype-driven assays that can assess one or many potential outcomes of infection (Beutler, 2005). Of note, the Toll receptors have no role unless microbes invade the body. Many genes probably function during times of stress, regardless of how that stress was induced. A future challenge will be to determine which stresses, similar to those relating to pathways and systems. But how will we prioritize the huge number of mouse mutations that will be available to determine what is important for biology? The mammalian genome is estimated to contain nearly 25,000 genes, and each scientist would ideally generate null, hypomorphic and conditional alleles of each one. Moreover, users of mouse models are expanding to include pharmaceutical companies and the medical community. Translational research will ask how a drug treatment affects the disease outcome and, ultimately, the health of the animal. A drug treatment may ameliorate only a part of the phenotype or cause unexpected outcomes that can only be assessed in the whole animal. Therefore, a future evaluation of mouse models may include many different types of mutations under the influence of different conditions, including drug treatments, and/or humanized backgrounds. It is prudent to consider how to manage and standardize this amount of phenotyping in the future.

Standard protocols for phenotyping mice have been developed in Europe through a program called EUMORPHIA (European Union Mouse Research for Public Health and Industrial Applications, www.eumorphia.org). In addition to providing detailed phenotype information for a number of mouse strains, groups within EUMORPHIA compared assays across laboratories for their replicability and robustness, to develop a database of over 100 standard operating procedures (SOPs) that can be used for phenotyping every major organ system, performing pathology, and assessing gene expression. Two databases, one in Europe, the Europhenome Database (www.europhenome.eu), and one in the USA, the Mouse Phenome Database (MPD, http://phenome.jax.org/pub-cgi/phenome/mpdcgi), house information gathered from a variety of inbred strains (Mallon et al., 2008; Bogue et al., 2007). A phenotype assessment paradigm called EMPReSSslim (European Mouse Phenotyping Resource for Standardised Screens, http://empress.har.mrc.ac.uk, http://empress.har.mrc.ac.uk/browser/index.html?pipe=one) tested by EUMORPHIA is being used to assess over 500 mouse strains, developed from the KOMP and EUCOMM projects, within a variety of laboratories that participate in the European Mouse Disease Clinic (EUMODIC, www.eumodic.org). Such assays must be performed on a large number of mutant mice and are designed to detect robust phenotypes in most body systems; therefore, throughput and cost are important considerations when choosing the tests (Brown et al., 2005).

**Where are we now?**

Soon, a mouse mutant or a targeted embryonic stem (ES) cell will be available for every gene in the mouse genome. Collaborative efforts between the USA and Europe through the Knockout Mouse Project (KOMP, www.komp.org) and the European Conditional Mouse Mutagenesis project (EUCOMM, www.eucomm.org) are generating a public resource of a null and/or conditional mutation in every gene in the mouse genome (Austin et al., 2004; Friedel et al., 2007). Such a resource will allow scientists to focus on a scientific hypothesis to tackle problems such as those relating to pathways and systems. But how will we prioritize the huge number of mouse mutations that will be available to determine what is important for biology? The mammalian genome is estimated to contain nearly 25,000 genes, and each scientist would ideally generate null, hypomorphic and conditional alleles of each one. Moreover, users of mouse models are expanding to include pharmaceutical companies and the medical community. Translational research will ask how a drug treatment affects the disease outcome and, ultimately, the health of the animal. A drug treatment may ameliorate only a part of the phenotype or cause unexpected outcomes that can only be assessed in the whole animal. Therefore, a future evaluation of mouse models may include many different types of mutations under the influence of different conditions, including drug treatments, and/or humanized backgrounds. It is prudent to consider how to manage and standardize this amount of phenotyping in the future.

Standard protocols for phenotyping mice have been developed in Europe through a program called EUMORPHIA (European Union Mouse Research for Public Health and Industrial Applications, www.eumorphia.org). In addition to providing detailed phenotype information for a number of mouse strains, groups within EUMORPHIA compared assays across laboratories for their replicability and robustness, to develop a database of over 100 standard operating procedures (SOPs) that can be used for phenotyping every major organ system, performing pathology, and assessing gene expression. Two databases, one in Europe, the Europhenome Database (www.europhenome.eu), and one in the USA, the Mouse Phenome Database (MPD, http://phenome.jax.org/pub-cgi/phenome/mpdcgi), house information gathered from a variety of inbred strains (Mallon et al., 2008; Bogue et al., 2007).

A phenotype assessment paradigm called EMPReSSslim (European Mouse Phenotyping Resource for Standardised Screens, http://empress.har.mrc.ac.uk, http://empress.har.mrc.ac.uk/browser/index.html?pipe=one) tested by EUMORPHIA is being used to assess over 500 mouse strains, developed from the KOMP and EUCOMM projects, within a variety of laboratories that participate in the European Mouse Disease Clinic (EUMODIC, www.eumodic.org). Such assays must be performed on a large number of mutant mice and are designed to detect robust phenotypes in most body systems; therefore, throughput and cost are important considerations when choosing the tests (Brown et al., 2005).

**What is the future?**

Today, many phenotyping tools are available to assess mice. However, the cost of assessing every mutant and drug treatment...
within each individual laboratory will become overwhelming. Further, to be most effective, each test requires experienced, trained personnel. The challenge will be to use assessment tools efficiently to uncover a large range of phenotypes, and then to direct information and mutants to disease-focused groups (Brown et al., 2006a). An International Mouse Phenotyping Consortium has been formed to discuss how scientists might prepare for this onslaught of mouse strains and conditions. A basic tenet learned from the EUMORPHA and EUMODIC projects is that it is too costly and time consuming to carry out all phenotype assays on all mice (even the broad-based assays). Therefore, different tiers of phenotyping are proposed (Fig. 1). A primary panel of phenotype assays should reveal a significant difference in mutant animals, even though they may not define the underlying basis for that phenotype, or its cause. These ‘first tier’ phenotyping assays would be broad based, hypothesis generating, and designed to economize mice and resources. For example, a fasting blood glucose assay is an inexpensive simple assay that may uncover a problem with glucose metabolism, but many other studies would be needed to determine the cause. A mouse strain with a high fasting blood glucose level could be referred to a specialized or ‘secondary’ phenotyping laboratory. ‘Second tier’ phenotyping is hypothesis driven, requires some biological expertise and would provide an in-depth analysis. ‘Third tier’ phenotyping should also be standardized, but may inquire about more detailed aspects of the phenotype, for example, whether diabetic neuropathy is an outcome (Sullivan et al., 2008). In the area of diabetes, such laboratories are already operating in the USA. The Mouse Metabolic Phenotyping Consortium (MMPC, www.mmopc.org) consists of a group of laboratories housed at outstanding academic institutions that perform a variety of metabolic assays for glucose metabolism, diabetes, obesity and diabetic complications. The goal of these laboratories is to broaden the scope of metabolic phenotyping tests for mice available to outside investigators, by providing tests that may require special equipment, or expertise, to carry out or interpret. The advantage of such an approach is that experts can standardize key methods, research can be expedited, costs can be reduced, and a database of information can be compiled.

The EUMODIC model was made possible through an unusual European-wide collaboration, which was funded through the European Research Council. It is not clear whether such a model would work in the USA, or other countries, where primary funding comes from disease-focused institutes. However, primary tier assays could mimic clinical diagnostics in humans, with reference laboratories providing the baseline tests, quality control and standards, while reporting values in a standard database format. In-depth second and third tier phenotyping could then be carried out by more specialized disease-focused laboratories.

Getting there
To carry out phenotyping, no matter what the model, guidelines must be established worldwide. We can prepare by standardizing test parameters, establishing quality control measures, improving technologies and connecting laboratories. For example, in the case of blood assays, the equipment, method of blood collection, use of anticoagulants and husbandry conditions must be reported as a part of the experiment. Further, journals should require that quality control measures and such parameters be reported; although, quality control measures are still lacking for many assays. For example, veterinary hematology analyzers assess the numbers of blood cells in the mouse. Such analyzers are commonly evaluated using human blood cell controls; however, mouse blood cells are smaller than human blood cells, requiring the apertures on the machine to be reset prior to performing assays on mice. Yet no standards exist for mouse blood. Large laboratories, such as the MRC-Harwell Mammalian Genetics Laboratory in the UK and The Wellcome Trust Sanger Institute in Hinxton, UK, can perform mouse strain controls, and determine whether they fall within a range of values. However, smaller laboratories that cannot afford to maintain additional control stocks of mice suffer. A larger service-based laboratory could afford such controls, and would be required to carry out primary phenotyping assays.

Additional infrastructures would be needed. Either of the models proposed above would require the animal facilities to be outfitted with phenotyping equipment, experienced personnel and databases. This would present a challenge to most animal facilities because of rising costs, since providing such resources would come at a price, even if cost recovery were the final outcome. Mouse phenotype ontologies must be developed (Smith et al., 2005). As phenotypes are described for mouse mutants under a variety of conditions, descriptive terms must be used that can be tracked in databases. Using the same terms that are used for humans will add value to the mouse phenotype descriptions. Finally, the databases should be coordinated and linked, which is not a simple task, but efforts have already started in Europe to do so through a project called CASIMIR (Coordination and Sustainability of International Mouse Informatics Resources, www.casimir.org.uk).

Over the coming years, it is likely that technologies for phenotyping mice will become even better. New validated assays may replace slower, less informative tests. In addition, advances in mouse phenotyping will probably continue to mimic advances in human disease detection and diagnosis. For example, pediatric biochemistries performed on mass spectrometers are very cost-effective and require very little blood, but can assess amino acids, cholesterol and fatty acids, among other blood components. At the German Mouse Clinic (www.mouseclinic.de), luminex bead assays for a variety of hormones are being developed and validated for use with mice. Currently, any assays for endocrine activity

---

**Fig. 1. A model for carrying out phenotyping in the mouse.** Primary phenotyping assays would be performed in tier 1, which would probably also house informatics and distribution. More specialized secondary (2°) and tertiary (3°) phenotyping (tiers 2 and 3) would be carried out at associated laboratories with disease specialties.
require serum to be pooled from multiple animals in order to carry out the test. The luminex immunoassay will allow for a sensitive accurate reading of a variety of hormones and endocrine parameters from individual mice. As technologies are improved, the numbers and efficacy of tests will increase, costs will decrease, and our ability to phenotype large numbers of mice will improve.

Challenges remain. Most of the tools described here assess postnatal or adult mice. It is difficult to examine mouse embryos in utero, even though microscopic ultrasound technologies have been developed for doing so (Brown et al., 2006b). These technologies require highly trained and specialized individuals, and handbooks have been devoted to methods for examining an embryonic phenotype (Papaioannou and Behringer, 2005). Further, different genetic backgrounds can complicate the analysis of any mutant line. Ultimately, our knowledge of mouse mutants may exceed our knowledge of human disease conditions and, in such cases, a mouse mutant may be described as a collection of disease phenotypes, rather than a disease model.

REFERENCES


