

## Building a better mouse test

As more mouse models are produced, researchers studying neuropsychiatric diseases will need better ways to evaluate them and more realistic assessment of the results.

The toolbox for mouse genetics has never been fuller. The International Knockout Mouse Consortium (IKMC) has generated thousands of mouse embryonic stem cell lines. The plan is to turn all of these into mouse strains, each with a different gene knocked out. Other projects are also quickly yielding many new strains of laboratory mice and rats.

Some of these rodents will almost certainly be useful resources for studying neuropsychiatric diseases. Several rodents with single-gene knockouts have interesting correlates to human disorders such as Rett's syndrome and obsessive-compulsive disorders. So do those carrying disease-linked human mutations. Environmentally induced models and inbred strains are also routinely used to study conditions such as depression and autism.

Having more animals to study can only go so far toward biological insight. Researchers also need robust experimental protocols. In humans, most neuropsychiatric diseases are assessed not by molecular markers but by behavior, and this is the desirable strategy in animal models as well. But these diseases are heterogeneous and hard to diagnose in humans; detecting relevant phenotypes in mice is harder still.

A multicountry team of scientists organized as the International Mouse Phenotyping Consortium is giving itself a decade to run IMKC strains through a standard battery of tests. The team will provide the data in a centralized database as they are ready. The testing pipeline was still being finalized as this issue went to press, but so far neurological and behavioral assessments include gross observations, assessments of grip strength and balance, the open field test (often used to assess exploratory and anxiety-like behavior), prepulse startle inhibition (a quick test with intriguing correlations to schizophrenia) and simple assessments of vision, hearing, pain sensation and smell. More complex tests, such as those examining cognition and sociability, are beyond the IMPC's ability to implement. Other scientists will need to fill in the gaps.

Behavioral testing is expensive, labor-intensive and time consuming, and requires unerring attention to details. Researchers must be aware of endless sources of artifacts. In one study measuring factors affecting response to painful stimuli, the handler made more of a difference than the genotype (Chesler, E.J. *et al. Nat. Neurosci.* 5, 1101–1102; 2002). Tests performed days or weeks earlier can affect performance in a new assay. So can subtle changes to diet or noise from adjoining labs.

The availability of so many more animal strains will increase the research invested in them. Thus, it is increasingly important that the assays implemented be up to the task. In fact, one of IMPC's contributions to the study of neuropsychiatric diseases will be in the enormous amount of data about the operation of its basic assessments and potential environmental impacts on test results. The IMPC database will facilitate recording metadata on behavioral experiments such as food type, cage conditions and even the opaqueness of the field test walls, and will incorporate acquired wisdom of which conditions affect which tests. Such information will make researchers more aware of potential artifacts and confounding variables, and help them assess other researchers' data and home in on important variables for their own research.

But the accumulation of data is not enough. Appropriate attitudes are also required. Researchers too often simply assert or assume the validity of their models and assays. It is essential to discuss the strengths and weaknesses of a model, such as how well an environmental insult mirrors a human neurodegenerative disease. Indeed, the question of whether or not a model is valid is too simplistic. Instead of validity being considered in terms of 'yes' and 'no', it should be evaluated as a continuum, with some models having more evidence of validity and others less, both in terms of their replication of disease symptoms and their modeling of underlying causes.

Although no animal will be a perfect model for any human disease, good mouse models can be very useful. But the validity of a particular model depends on the goal: animals used to screen for new antidepressants based on reference compounds may not be the best choice to investigate the etiologies or mechanisms of disease. A recent analysis by Eric Nestler and Steven Hyman emphasized not a need for more data but for more frank assessments: "Most important in developing, examining and reporting on animal models of disease is to be clear about the goals of the model and, in that context, to critically judge construct, face and predictive validity" (Nestler, E.J. and Hyman, S.E. *Nat. Neurosci.* 13, 1161–1169; 2010).

Tangible steps can be taken to realize this ideal. An appropriate database for recording experimental conditions could help. So will publishing checklists that prompt explicit statements on hypotheses, aspects of the illness to be modeled and assessments of validity in a specific experiment. Although expanding options for mouse models are certainly a good thing, these options will be wasted without a concomitant commitment to developing better assays.