Optimizing behavioral assays for mouse models of autism

BY **JACQUELINE CRAWLEY**

23 SEPTEMBER 2014

As the number of genetic mutations linked to autism **rises into the hundreds**, and the number of corresponding rodent models climbs concomitantly, we in the research community are well positioned to advance our knowledge about the genetic causes of autism. Because autism is still defined solely by its behavioral symptoms, **measuring autism-relevant behaviors** in rodent models is crucial to developing new therapies.

Now may be a good time to remind ourselves of best practices for conducting assays of mouse and rat behaviors. Our mission as autism researchers is to frame our interpretations of mouse behavior accuratelyin a way that is relevant to the diagnostic symptoms of the disorder.

Let's say you are pursuing a compelling hypothesis about a mutation implicated in autism, and have just generated a new mouse with the mutation. Your laboratory's expertise may be in genetics or neurobiology, but not in mouse behavior. How do you proceed?

Ideally, you identify an expert behavioral neuroscientist and initiate a productive collaboration. Or you might contract with a behavioral core facility to test your mice. You will want to look for a core with established competence, standard equipment and careful methods.

Behavioral assays are not technically difficult, but require a high level of attention to a large number of details that affect results. Look for a behavioral neuroscience collaborator or a core director with knowledge and published experience in conducting tests relevant to autism, including social and **repetitive behavior** assays. Many core testing facilities are staffed by investigators who were trained in only one specific behavioral domain, such as drugs of abuse, learning and memory or movement disorders.

Intellectual impairments, anxiety, seizures, sensory, motor and other associated abnormalities appear in a sizeable proportion of people with autism, and are useful to include when looking at

mouse models. But the core diagnostic symptoms, which include social and communication deficits and repetitive behaviors with restricted interests, should arguably be your primary focus.

Background effects:

Next is the challenge of breeding cohorts of mice for behavioral testing. One aspect to consider is **the background strain of mice**, both when placing the mutation in stem cells and when mating the mutant mice.Any given inbred strain of mice harbors some rare variants that could interact with your targeted mutation to enhance or attenuate phenotypes of interest.

Researchers commonly use the C57BL/6J (B6) inbred strain because of its generally average traits. However, even the B6 mouse genome contains unusual alleles that could influence autism-relevant behaviors. For example, B6 carries a gene that leads to age-related hearing loss[1]. FVB is another good choice because it shows high social behaviors and few baseline repetitive behaviors, and also produces large litters. Although the FVB/NJ substrain has poor vision, the recently developed FVB/AntJ substrain has normal sight[2].

Large litters are helpful. Behavioral testing usually requires about 15 mice of each genotype or treatment group, which should come from the same litters. This helps control for individual variability, which is present in all biological systems. Behavioral variability is particularly acute because of the many small and large influences of the environment on behavior, even among genetically identical mice living in a well-controlled vivarium. Parental care, dominance hierarchy, handling by investigators and vivarium noises, and many other environmental factors can affect individual mice differently and modify their behaviors during testing. Large numbers of mice, in the range of 12 to 20 per genotype, will average out these random differences.

Wildtype littermates without the mutation provide the best control for the home-cage environment. If you cannot generate enough littermates of each genotype all at once from multiple breeding cages, you can combine mice from a second round of breeding with the first group to reach the required numbers for the first cohort. But these combined subgroups are valid only if scores from the wildtypes do not differ significantly between subgroups.

A cohort is defined as the full number of mice in each experimental group. To confirm reproducibility of findings, researchers can breed a second cohort with the full numbers and test it identically to the first cohort. There is no substitute for dedicated cage space to breed sufficient numbers of littermate mice of each genotype for behavioral studies.

Assessing behavior:

Once you have enough mice, how do you select the ideal constellation of assays to address your hypothesis? The best path forward is to start with the wealth of well-validated tasks in the behavioral neuroscience literature. Methods, control procedures, standardized equipment and

required reagents are widely published. A reputable behavioral neuroscience collaborator or core facility will follow the gold standards.

Although it is always fun to invent new mouse behavioral tasks, the validation process is extensive. If a study detects no symptoms on standard tasks, and reports symptoms only with a new task that was just developed by the research team and is minimally described, this suggests that the mutation has weak effects.

Tests for mouse models of autism must include the primary diagnostic criteria of the syndrome: social abnormalities and repetitive behaviors. Preferably, two or more tests within each of these domains are employed to corroborate findings and show that abnormalities generalize across related tests[3].

Examples of assays for social behavior include social interactions between juveniles, interactions between adult males and females, and recordings of ultrasonic vocalizations. Ourthree-chambered social-approach test, for comparing time spent with a novel mouse versus a novel object, is a simple, automated social test that is widely used.

Examples in the repetitive behavior domain include high levels of self-grooming, digging in the litter and burying marbles, as well as motor stereotypies such as rapid circling and vertical jumping. Researchers have conceptualized insistence on sameness as an animal's impaired ability to switch from one action to another. For example, a mouse model of autism may learn a Morris water maze platform location with no problem, but have trouble reversing to a new platform location. This is interpreted as resistance to change.

Many excellent standardized assays in the behavioral neuroscience literature test cognitive abilities, anxiety-related behaviors, responses to sensory stimuli, hyperactivity and other traits that are analogous to associated symptoms of autism[4]. The order of testing does not appear to be critical for most behavioral assays, as long as a day or two intervenes between tests[5].Elevated plus maze, a conflict test, is usually conducted first, as scores on this anxiety-related task are sensitive to prior experience. Stressful tests such as water-maze learning, which requires vigorous swimming, are placed at the end of the sequence.

Be sure to include a battery of general health measures. A sick mouse will not behave normally in any behavioral task. A sedated or motor-impaired mouse cannot move around sufficiently for most behavioral assays.

When designing control measures, try to think like a mouse. Anthropomorphizing is a mistake; recognizing the mouse's point of view is essential. For example, people tend to be afraid of the dark, whereas mice avoid the light. In a new environment, people generally gravitate toward familiar objects and friends, whereas mice explore novel objects and new mice.

Test and repeat:

Publishing your research findings is an exercise in correctly interpreting the data. Statistical analyses of most behavioral data employ an overall Analysis of Variance (ANOVA), which takes into account the variability across all groups. Repeated Measures Analysis of Variance is often appropriate, to evaluate factors such as genotype, sex, time or drug treatment. A follow-up (*post hoc*) analysis to determine which groups differ from each other is justified when the ANOVA value is significant. Stringent *post hoc* tests such as Newman-Keuls, Dunnett's, Tukey's and Bonferroni-Dunn yield more biologically meaningful results than weak *post hoc* tests such as Fisher's Least Significant Difference.

Certain tasks may require specific statistical considerations. For example, our three-chambered social-approach test offers a simple yes-or-no measure of sociability[6]. The mice either spend more time in the chamber interacting with the novel mouse than in the chamber with the novel object — meeting the definition of sociability — or they do not. Sociability or absence of sociability usually replicates well across cohorts.

However, the actual amount of time a mouse spends with a novel mouse varies greatly across repeated testing of the same mice, and across cohorts of the same genotype. The test is therefore not a useful quantitative measure. Seconds spent with the novel mouse cannot be used to quantitatively compare genotypes, or to compare controls with drug treatments, although many investigators have made this mistake.

A more sensitive social task, such as reciprocal social interactions, provides quantitative scores for many useful observer-scored parameters that measure different aspects of the same behavioral trait. For example, mouse reciprocal social interactions include nose-to-nose sniffing, anogenital sniffing, following the other mouse, physical contact and emitting ultrasonic vocalizations. **Videotracking software** that produces only a single composite parameter or index will miss these various elements of social interactions. Although not every parameter may show significant genotype differences, a preponderance of parameters strengthens the conclusion that a mouse model displays social interaction deficits.

Similarly, interpretation of a major abnormality within an autism-relevant domain will benefit from corroborative tasks within the domain. Consistent deficits in two or more social tasks, or high scores on two or more repetitive behaviors, increase confidence that an abnormality generalizes within that domain.

Corroboration of findings in two or more cohorts of mice is essential to demonstrate replicability. What to do if an effect occurs in the first cohort but not in the second? A third cohort could be considered as a tiebreaker, but a more accurate interpretation is to describe the phenotype as minor or weak.

Failure to replicate across labs is a red flag in all fields of scientific research. Methodological issues are often the cause, and are best discussed between the labs and detailed in publications. Findings that do consistently replicate across many labs, such as the social deficits seen in the **BTBR inbred strain of mice**, raise confidence in the reliability of the mouse model and its usefulness as a translational research tool.

Drawing conclusions:

To ensure appropriate interpretations of your findings, remember to take time to talk to the clinical experts, read the clinical literature and, if possible, observe people with autism.

For example, too few vocalizations in mouse pups separated from their mother **may be relevant to human infant crying**, but is probably different from the kind of socially inappropriate communication seen in people with autism. Ultrasonic vocalizations during social interactions in juvenile and adult rodents could be more relevant to autism communication deficits. As another example, high-functioning adults with autism tend to be good at routine housekeeping tasks, suggesting that poor nest building in mice is not a good model for autism-relevant behavior. Further, a mouse model that displays only a 20 percent elevation in open field activity is not highly relevant to the clinical definition of hyperactivity.

It does the autism research field no good to overemphasize a statistically significant but small finding, or to claim that a gene plays a central role in autism when behaviors relevant only to associated symptoms, such as anxiety, differ between genotypes. We all want to publish our findings in high-profile journals. We all need to obtain funding based on our discoveries. However, we all recognize that our higher calling is to discover the truth about the causes of autism. The translational value of autism research discoveries ultimately rests on the rigor and honesty of every investigator. Our common goal is to frame the interpretations of our mouse behavioral data accurately.

Jacqueline Crawley is professor of psychiatry and behavioral sciences at the MIND Institute, University of California Davis School of Medicine in Sacramento, California.

References:

1: Johnson K.R. *et al. Genomics* **70**, 171-180 (2000) **PubMed**

2: Errijgers V. *et al. Genes Brain Behav.* **6**, 552-557 (2007) **PubMed**

3: Silverman J.L. *et al. Nat. Rev. Neurosci.* **11**, 490-502 (2010) **PubMed**

4: Crawley J.N. (2007). *What's wrong with my mouse? Behavioral phenotyping of transgenic and knockout mice.* Hoboken, NJ: John Wiley & Sons, Inc.

5: McIlwain K.L. *et al. Physiol. Behav.* **73**, 705-717 (2001) **PubMed**

6: Yang M. *et al. Curr. Protoc. Neurosci.* **Jul**, Chapter 8 (2011) **PubMed**