

# Behavioral Phenotyping of Transgenic and Knockout Mice: Practical Concerns and Potential Pitfalls

*Kathleen R. Bailey, Nathan R. Rustay, and Jacqueline N. Crawley*

## Abstract

New technologies in molecular genetics have dramatically increased the number of targeted gene mutations available to the biomedical research community. Many mutant mouse lines have been generated to provide animal models for human genetic disorders, offering insights into anatomical, neurochemical, and behavioral effects of aberrant gene expression. A variety of assays have been developed to identify and characterize phenotypic changes. In the behavioral domain, our phenotyping strategy involves a comprehensive standardized methodological approach that assesses general health, reflexes, sensory abilities, and motor functions. This assessment is followed by a series of complementary tasks in the specific behavioral domain(s) hypothesized to reveal the function(s) of the gene. Our multitiered approach minimizes intersubject variability by standardizing the experimental history for all animals, improves interlaboratory reliability by providing a clearly defined experimental protocol, and minimizes artifactual interpretations of behavioral data by careful preliminary assessments of basic behaviors, followed by multiple tests within the behavioral domain of interest. Despite meticulous attention to experimental protocol, attention to environmental factors is essential. Differences in noise, light, home cage environment, handling, and diet can dramatically alter behavior. Baseline differences in the behaviors of inbred strains used to generate targeted mutant mouse lines can directly influence the behavioral phenotype of the mutant line. Strategies aimed at minimizing environmental variability and contributions of background genes will enhance the robustness of mouse behavioral phenotyping assays.

**Key Words:** background strain; behavior; behavioral test suites; colony housing; general health; learning and memory; mouse; flanking genes

## Behavioral Phenotyping

Advances in genetic technologies used to generate targeted gene mutations in mice highlight the need for rigorous, replicable, and dissociable behavioral test-

ing of mutant mice. Findings from our laboratory and others' suggest that many factors can influence the outcome of behavioral tests, indicating the need for a clearly defined approach to the behavioral phenotyping process (Crawley 1999, 2000, 2003; Paylor and Crawley 1997; Gold 1999; van der Staay and Steckler 2001, 2002; Wahlsten 2001). Our laboratory has developed a multitiered approach (Figure 1) that begins by assessing basic measures of general health and progresses in a systematic fashion to complex tests of behavioral domains targeted to address the hypothesized function of the gene of interest (Crawley 1999). This review focuses on the systematic procedures that have been developed for identifying behavioral phenotypes in genetically modified mice with a special focus on complicating factors (e.g., strain, handling, and housing and test order) that may dramatically affect performance and complicate the interpretation of behavioral results.

Behavioral phenotyping of a new mutant line begins with an assessment of overall general health. This assessment is accomplished by examining the fur condition, body weight, muscle and skeletal development, general activity level, and home cage social interactions of the mice. Many of these factors (e.g., fur condition, dowel biting, home cage activity) are scored by trained observers on a three-point scale (1 = below average/poor, 2 = average/good, and 3 = above average/superior). Other factors (e.g., pinnae reflex, positional passivity, forepaw reaching) are indicated as present or absent in each mouse. These data are quantified as the percentage of mice in the group ( $[\text{number of mice presenting}/\text{total number of mice in group}] \times 100$ ) that display the measure of interest. Poorly groomed, lethargic, isolated, weak, or malformed mice may indicate a postnatally lethal mutation or, at the least, a concern for the interpretability of subsequent behavioral test results. Deficits in subsequent behavioral tests may simply be an artifact of testing a sick mouse unable to perform the procedures of the task adequately.

The second and third stages of testing examine neurological, sensory, and motor functions of the mice. Simple tests of reflexes, including the righting, corneal, pinnae, and vibrissae reflexes, provide additional evidence that the genetic mutation has not compromised gross neurological function. Tests of motor coordination and sensory abilities address specific questions about the role of a targeted gene in locomotion, olfaction, vision, audition, and analgesia. We have included a subset of tests used by our laboratory to assess sensory abilities and motor functions. Additional tests can be incorporated for more detailed evaluation of specific sensory modalities and motor abilities. The auto-

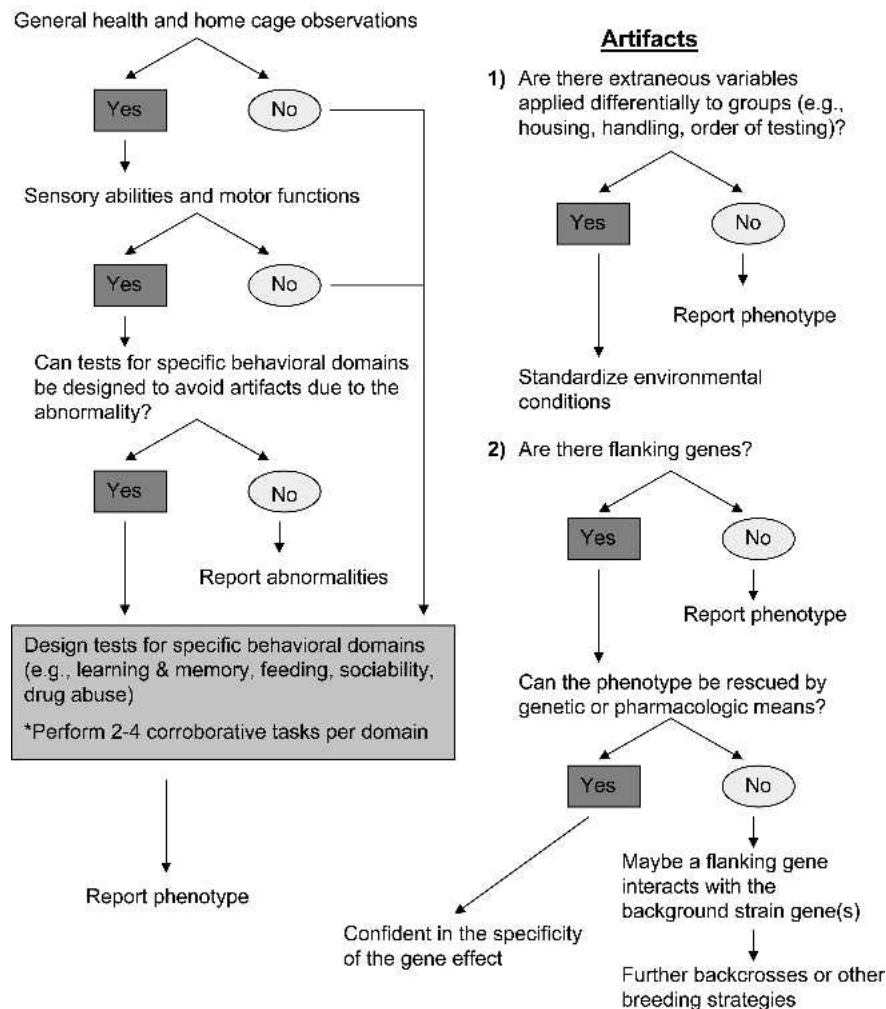
---

Kathleen R. Bailey, Ph.D., Nathan R. Rustay, Ph.D., and Jacqueline N. Crawley, Ph.D., are Postdoctoral Fellow, Postdoctoral Fellow, and Laboratory Chief, respectively, in the Laboratory of Behavioral Neuroscience, National Institute of Mental Health, Bethesda, Maryland.

Address correspondence and reprint requests to Dr. Bailey, Laboratory of Behavioral Neuroscience, NIMH, Bldg. 35, Room 1C909 Mailstop 3730, Bethesda, MD 20892-3730; Email: baileyka@mail.nih.gov; Tel: 301-451-9386; Fax: 301-480-4630. The inclusion of this information, which *ILAR Journal* does not typically make available, is provided at the authors' request.

## **Behavioral Phenotyping Strategy**

**Is your new mutant mouse significantly different from its wildtype littermate control?**



**Figure 1** Systematic behavioral phenotyping strategy.

mated open field test provides an index of locomotor activity. Two of the distinct questions addressed with this test include measuring an animal's activity level in a novel environment (5-min test) and assessing habituation as a reduction in an animal's activity level across time as the environment becomes familiar (30-min or 1-hr test). The accelerating rotorod test can identify mutant lines with deficits in balance and motor coordination. Pain sensitivity is assessed with the tail flick and hot plate tests. The acoustic startle test identifies mice with auditory deficits. Vision can be evaluated in a simple test like forepaw reaching, in which the mouse is suspended by the tail and lowered slowly toward a solid surface.

These three stages of testing may provide the necessary evidence to confirm or reject the hypothesized contribution of the gene product. However, if the gene of interest is

suspected of involvement in more complex behaviors, then these initial levels of testing provide critical controls to minimize the likelihood of false-positive interpretations in complex tasks. Investigators can then move forward to select tasks designed to assess behavior in the domain of interest (e.g., cognitive, attention, or anxiety-like behaviors). When sensory or motor deficits are detected in the preliminary sensory and neurological screens, it is often possible to choose complex tasks that do not require that particular sensory or motor ability. For instance, null mutant mice that are deaf would not be suitable test subjects for tone-shock pairing in auditory cued fear conditioning. Instead, the investigator could choose to pair the foot shock with a light cue, thereby avoiding the auditory sensory deficit.

The fourth level of this multitiered behavioral pheno-

typing strategy (Figure 1) addresses the investigator's specific research question within a behavioral domain. The behavioral neuroscience literature provides numerous examples of relevant tests appropriate for assessing performance in each of the behavioral domains, including tests for anxiety-like conflicts, drug abuse, motor function, social interactions, learning and memory, and nociception (Belknap et al. 1990; Bolivar et al. 2000; Crabbe et al. 1996; Crawley 1999, 2000; Holmes et al. 2002, 2003b; Logue et al. 1997, 1998; McFadyen et al. 2003; Owen et al. 1997; Paylor and Crawley 1997; Steiner et al. 2001). Determining the phenotype of a mutant line can best be accomplished by selecting several complementary tests within a behavioral domain. For example, anxiety-like behavior can be assessed in the elevated plus-maze, light↔dark exploration, and as a component of time spent in the center of an open field test. Each of these tasks is an approach/avoidance conflict, measuring the tendency of mice to explore a novel environment versus avoid brightly lighted open spaces. Yet each test may create varying levels of stress manifested as different types of anxiety-like behavior. The investigator will avoid false-negative results by incorporating several of these tests.

When assessing animals on several complementary behavioral tests in one or more behavioral domains, investigators may wish to consider test experience and test order. In an elegant study examining the effect of testing experience and test order, McIlwain and colleagues (2001) demonstrated that mice completing a behavioral phenotyping test battery performed significantly differently on certain tests (e.g., rotorod, open field) from naive mice completing only one selected test. However, performance on other tests (e.g., acoustic startle habituation, prepulse inhibition) appeared insensitive to prior test experience. In addition, performance on specific anxiety-like behavior tasks (e.g., elevated plus-maze, light↔dark exploration test) appeared sensitive to the order of the test within a test battery or to repeated testing (Holmes and Rodgers 1998; McIlwain et al. 2001; Vöikar et al. 2004). A multitiered approach to behavioral phenotyping should be designed to standardize the test experience for each mouse (Figure 1), thereby increasing the probability that results will be replicated within a laboratory and across different laboratories (Crabbe et al. 1999; van der Staay and Steckler 2001, 2002; Wahlsten et al. 2003b). In addition, researchers should be especially diligent when reporting their results, providing specific reference to methodological factors (e.g., housing, handling, test order, and genetic background) that may contribute to the experimental outcome.

These stages in the multitiered behavioral phenotyping strategy provide a comprehensive approach to evaluating the contributions that a mutant mouse model can make toward the understanding of human genetic disorders. Once a phenotypic anomaly has been identified, the mouse model can be applied to evaluate appropriate treatment interven-

tions for the disease (Cryan et al. 2004; Holmes et al. 2003b; Steiner et al. 2001).

## Colony Space

Evaluating the behavioral phenotype of a new transgenic or knockout (KO<sup>1</sup>) line usually requires a minimum of 10 mice per genotype. Generally investigators are interested in comparing the behavior of the homozygous (−/−) null mutants to their wild-type (WT,<sup>1</sup> +/+) littermates. If a “gene-dose” effect is suspected, it is advisable to include the heterozygous (+/−) littermates. Observed sex effects may also influence sample size for each of the genotypes. If preliminary behavioral data indicate significant differences between males and females within a genotype, it is advisable to test a minimum of 10 males and 10 females per genotype.

In addition to the general guidelines mentioned above to ensure an adequate sample size for statistical analysis, investigators may also wish to consider several other factors in choosing the initial sample size. Complex learning and memory tasks require significant training for mice to achieve performance criterion. Some mice fail to reach minimum performance levels and must be excluded from the experiment. Experiments with extensive training are also more likely to be characterized by attrition within the experimental groups due to health issues. Experiments involving surgical or pharmacological treatments may require additional animals in case of surgical error, cannula misplacement, or subsequent cannula loss during testing.

Colony space requirements are also affected by the genetic approaches favored by principal investigators. Large numbers of cages are required for high throughput behavioral screening in mutagenesis approaches (Crabbe and Morris 2004; Tecott and Nestler 2004). Inbred strain distributions and recombinant inbred strain comparisons require extensive colony space. Investigators working with newly developed transgenic mice may also wish to consider allocating colony space for testing mice of the background strain used to generate and breed the mutation. Several studies have examined differences in behavioral performance of inbred strains typically used to generate transgenic and knockout lines (Bolivar et al. 2000, 2001; Bouwknecht and Paylor 2002; Clapcote and Roder 2004; Cook et al. 2002; Crawley et al. 1997; Guillot et al. 1994; Logue et al. 1997, 1998; McFadyen et al. 2003; Owen et al. 1997; Vöikar et al. 2001; Wahlsten et al. 2003a,b). In addition to evaluating the behavior of the WT littermates as controls for the null (−/−) mutants in a specific knockout line, it would be useful to establish that performance of the WT littermates does not differ significantly from performance of mice from the relevant background strain. Importing background strain controls from an outside vendor and comparing behavioral performance with the WT of a mutant strain bred in-house is often useful.

<sup>1</sup>Abbreviations used in this article: ES, embryonic stem cell; KO, knockout; WT, wildtype; 5-HTT, 5-hydroxy tryptamine transporter.

## Other Colony Considerations

Veterinarians and animal care technicians follow specific guidelines regarding bedding changes, dietary and home cage enrichment, number of mice per cage, and health checks for mice under their care. To enable investigators to minimize outside factors that may contribute to variability in behavioral test data, housing conditions and the colony environment should be as similar as possible for all of the mice within an experiment. Light and humidity levels of many colony rooms are automated so that investigators can be relatively assured that these factors will be similar in different rooms. However, it is wise for investigators to consider how periodic adjustments in the light/dark cycle (e.g., daylight savings time) may affect ongoing experiments. In addition, when several investigators share vivaria space, investigators needing specific light/dark cycles (e.g., a reverse light/dark cycle) may affect space requirements for housing.

Another important recommendation for mutant mouse lines is to house littermates together whenever possible. Although this practice may increase the number of cages required, it maintains the social groups established among littermates. In contrast, home cage assignment based on genotype could increase the occurrence of fighting or stereotypies (i.e., barbering, excessive running, corner jumping) and introduce an environmental variable that is not equally experienced by all genotypes. For this reason, we house littermates together by sex but not by genotype. It should be noted that in instances in which a single male or female is born into a litter, combining that pup with others weaned at the same time is preferable to single housing the mouse because isolation is a stressor in mice (Garattini and Valzelli 1981; Valzelli 1973; Vöikar et al. 2005).

There is considerable evidence that differences in animal housing environments and animal care procedures at facilities can greatly influence affect, learning and memory performance, brain enzyme activity, and basal measures of neuroendocrine function in rodents (Benaroya-Milshtein et al. 2004; Jankowsky et al. 2005; Lambert et al. 2005; Lazarov et al. 2005; Olsson and Dahlborn 2002; van Praag et al. 2000; Wolfer et al. 2004). Even when extreme care is taken to equate the experimental design, testing apparatus, laboratory environment, and strain and age of mice tested across laboratories, investigators have found that results from different laboratories are not always identical (Crabbe et al. 1999; Wahlsten et al. 2003b,c).

Maintaining standardized animal husbandry procedures and home cage environments is a primary concern for the purpose of identifying phenotypic differences in transgenic and null mutant mice. Ideally, all mice within an experiment or series of experiments will be housed and handled by animal care staff in similar amounts and under similar conditions. For example, cage changes should occur at the same time on the same day, and inanimate objects used for home cage enrichment should be identical and equally available in all cages. Ideally, lighting in the colony rooms and noise

from mechanical equipment and animal husbandry duties (e.g., moving equipment cages and racks for cleaning, or staff conversation) are consistent between rooms and at a level that is not distressful to the mice. It is important for animal care technicians to be well informed about the effect of fluctuating noise, temperature, and lighting levels on behavioral testing.

Unanticipated factors can also affect the experimental history and home cage experience for a subset of mice within a behavioral study. If one mouse in a cage develops health problems, all of the mice in that cage, including healthy cagemates, may receive food supplements (e.g., fruit or gelatin) or medicated water as part of the treatment protocol for the ill mouse. In addition to dietary changes, mice flagged for health observations will be handled for examination more frequently, which may alter the behavioral performance of healthy cagemates as well. Investigators should discuss with animal care technicians and veterinarians how health concerns will be handled. Options include delaying treatment until a short behavioral experiment is completed, or moving the affected mouse to another cage for treatment.

The cage changing process within a colony is generally performed on a specified schedule. Investigators may wish to consider the impact of this process when determining the behavioral testing schedule for their experiments. Generally mice exhibit increased activity levels after being moved to clean cages. Testing for anxiety-like behaviors immediately after cage changing may yield results that are very different from results in mice that have not recently had their cages changed. In addition, learning and memory tasks, like the water maze, require consolidation of the learning experience to improve subsequent performance. This consolidation process occurs within a few hours after training. Although the effect that periodic cage changing or health checks have on this consolidation process remains to be explored systematically, investigators may wish to advise their animal care technicians to avoid cage changing and health checks during sensitive behavioral testing.

A recent development in the housing of laboratory animals used for biomedical research is a move from standard housing in unadorned cages to cages that offer a variety of forms of environmental enrichment or supplementation. Ethical and ethological concerns for the well-being of laboratory animals have prompted research into the benefits or potential problems of home cage enrichment, as well as how this may affect behavioral and physiological measures (Benaroya-Milshtein et al. 2004; Benefiel et al. 2005; Garner 2005; Hutchinson et al. 2005; Lambert et al. 2005; Lazarov et al. 2005; Moons et al. 2004; van Praag et al. 2000; Wolfer et al. 2004; Wurbel 2001). Researchers have divergent views on the effect that environmental enrichment will have on behavioral measures. One viewpoint suggests that without a thorough evaluation of the effect of various environmental enrichments, the variability in experimental results between different laboratories that use different enrichment protocols will increase (Benefiel et al. 2005).

An alternative perspective suggested by Garner (2005) is that enriched environments may reduce maladaptive and dysfunctional behaviors (e.g., infanticide, stereotypies, barbering) by increasing opportunities for mice to engage in species-relevant behaviors (e.g., nest building, social interactions). The opportunity to engage in ethologically relevant behaviors may in turn minimize the occurrence of idiosyncratic behaviors triggered by variability in other environmental factors (Wurbel 2001).

The effect of differing levels of enrichment or supplementation has yet to be evaluated systematically. Do minor changes in the environment (e.g., nesting material or cardboard tubing) alter physiology and subsequent behavior to the same extent as more elaborate enhancements (e.g., running wheels, climbing structures, and diet treats)? When environmental enrichment has been implemented as a standard protocol in an investigator's animal housing facility, it is useful to identify the supplementation protocol clearly when publishing results, to assist in equating conditions and to minimize interlaboratory variability in replication experiments.

Finally, it remains unclear whether stereotyped behaviors that are sometimes demonstrated by rodents in standard cages reflect an extreme end of normal brain function that manifests in captivity or reflect abnormal brain function (Garner 2005). For the purpose of screening phenotypic behaviors of mutant mice, this dichotomy is of concern. The variability associated with normal behavior is expected and inherent in behavioral studies, and is incorporated into appropriate statistical analyses. Rodents and other captive animals sometimes develop unusual behaviors due to constraints in the home cage environment that limit the expression or performance of species-relevant behaviors. These dysfunctional behaviors can indicate abnormal brain development or neurochemistry. In this situation, environmentally induced abnormal brain development or brain chemistry may produce a unique behavioral profile that is unrelated to the targeted gene mutation.

## Behavioral Testing Rooms

Behavioral testing rooms should be located in an area removed from noise and other distractions that may interfere with behavioral studies. Ideally, rooms can be assigned for continuous uses, such as to house large pieces of equipment like the Morris water maze or operant chambers, which tend to be cumbersome to move. Furthermore, by keeping these tests in the same place, there is less likelihood for variability in test rooms across studies, which can help to reduce the variability in mouse behavior. Separate rooms should be designated for tissue collection or surgery, to prevent the sounds and smells associated with these activities from interfering with the behavioral tasks. Some rooms can be used as multipurpose rooms into which small portable equipment can be moved during the day or week of testing.

Behavioral test rooms are most functional if they pos-

sess a desk with space for a computer and additional writing or note-taking space. Because many behavioral tasks result in equipment becoming dirty from mouse urine and feces, easy access to a large sink is required. In rooms used for the Morris water maze or the Porsolt swim task, a sink, faucet, and floor drain are usually included. Storage space in the form of shelves or cabinets can keep other equipment stored out of the way when not being used. Another very important consideration is to have control over light and temperature levels so that they can be adjusted to meet the demands of certain tasks. For example, low light levels on a separate electrical switch are needed for anxiety tests. A suggested design for the layout of a behavioral testing suite including test rooms and data analysis stations was proposed by Crawley (2003).

Locating the behavioral test rooms near a staging area serves many purposes. It allows mice to be taken from the colony area to a room that is closer to the test rooms, reducing transportation time and stress once the experiment is under way. It also allows the actual testing rooms to be smaller because minimal space is required for holding extra cages while mice are being tested. Lastly, it provides a consistent acclimation environment, which can reduce the stress that can be present in mice that are awaiting testing.

## Interpretational Issues in Behavioral Genetics Experiments

One major concern in the interpretation of knockout or transgenic mouse studies is the effect that the background strain may have on the behavioral expression of the genetic manipulation (Bucan and Abel 2002; Gerlai 1996, 2001; Phillips et al. 1999; Wolfer et al. 2002). Mutant founder mice possess genes from two or more background strains, including cells from both the embryonic stem (ES<sup>1</sup>) cell strain and the blastula donor strain. The founder mice are bred to WT mice of a chosen background strain to generate the knockout or transgenic mice for behavioral testing. This standard strategy can cause interpretational problems when the ES cells, blastula, and/or WT strain are of different genetic origin, which is often the case. For practical reasons, the ES cell is frequently harvested from a 129 strain and the blastula is from the C57BL/6J (B6) strain. In this example, even if the mutant mouse is backcrossed to the B6 background, there will be genes from the 129 strain still linked to the targeted gene (flanking genes) present in subsequent generations. Because we do not often know which genes are linked to the targeted gene, we cannot be sure whether phenotypic effects seen in the knockout or transgenic mice are due to the actual mutation, to the flanking genes, or to some interaction between B6 and 129 background genes.

Background gene issues are especially relevant if the different inbred strains used to generate the mutant mouse show different patterns of behavior in the tests of interest. Inbred strains differ widely in their performance on tests of learning and memory (Holmes et al. 2002; Owen et al.

1997), pain sensitivity (Belknap et al. 1990; Chesler et al. 2002), locomotor activity (Bolivar et al. 2000; Logue et al. 1997), motor coordination and balance (Rustay et al. 2003), and anxiety (Crabbe et al. 1999; Crawley et al. 1997). B6 and many of the 129 substrains show significantly different responses in many of these tests (Cook et al. 2002; Simpson et al. 1997). Creating a mutant mouse on a mixed genetic background can therefore introduce major variability that may mask the effect of the mutation on the behavioral phenotype. Furthermore, the mutation may have little to no physiological effect when generated on one background (and hence no behavioral effect), whereas it may have large effects when bred onto another background.

There are numerous reports that demonstrate the striking effect of genetic background on behavioral outcomes with knockout mice. For example, Holmes and colleagues (2003b) investigated the role of the serotonin transporter (5-hydroxytryptamine transporter, or 5-HTT<sup>1</sup>) on anxiety-related behavior using 5-HTT KO mice on both a B6 and 129S6 background. Each of these lines had been backcrossed to the respective background for at least 12 generations. Results showed that 5-HTT KO mice on the B6 background exhibited greater anxiety-like behavior than the WT mice on several tests; however, the mutation had no effect when it was bred onto the 129S6 background. The authors note that one potential explanation for this result was the fact that 129S6 mice show high baseline levels of anxiety that may have prohibited the detection of increased anxiety in the 5-HTT KO mice on that background.

In a study of the role of the *Fmr1* gene (responsible for fragile X syndrome in humans) in learning (Dobkin et al. 2000), the authors bred two lines of *Fmr1* KO mice: one on a mixed FVB/129 (approximately 75% FVB) background, and one on a B6/129 (97-98% B6B6) background. In a spatial position learning task, *Fmr1* KO mice on the B6/129 background did not show significant learning deficits, whereas those on the FVB/129 background did. Control mice of both lines showed similar learning curves. These data suggest that baseline performance differences could not account for the differences in background effects, implicating interactions between the *Fmr1* gene and unknown background gene(s).

Crabbe and coworkers (1996) reported that 5-HT 1B receptor KO mice drank significantly more ethanol than WT control mice. These mice were originally produced on a 129/Sv-ter background; however, over time, the knockout colonies ultimately included genetic material from different 129 substrains (Phillips et al. 1999). Subsequent attempts to replicate the original effect of the 5-HT 1B mutation on ethanol intake were unsuccessful, demonstrating that even small differences between substrains can influence the behavioral outcomes in KO studies.

There are some strategies for solving the problem of background strain effects in mutant lines of mice. Repeated backcrossing to the preferred background strain followed by genotyping for the presence or absence of the mutation can reduce the number of flanking genes carried along with the

targeted gene. However, this procedure is time consuming, and even with 12 generations of backcrossing, the number of flanking genes can still exceed 300 (Crusio 2004; Holmes et al. 2003a; Wolfer et al. 2002). Hence, backcrossing cannot completely eliminate the potentially confounding effect of flanking genes.

Another strategy to provide evidence that any behavioral effect is truly due to the genetic mutation is to perform a rescue experiment. Demonstrating that replacing the “lost” protein reverses the KO phenotype provides compelling evidence that the targeted gene mediated the phenotypic effect. For example, Schmitt and colleagues (2005) demonstrated that mice lacking a subunit of the AMPA glutamate receptor GluR1 show deficits in spatial working memory. When the subunit was reinserted in the forebrain of GluR1-deficient mice, the memory deficit was significantly reversed. Rescue strategies support the interpretation that the original deficit in the KO mice was due to the gene mutation and not to some other nonspecific effect.

Perhaps the most straightforward way to deal with the confounding variable of background strain effects on behavior is to eliminate the different strains altogether. Technology is emerging to enable the use of a single strain for the ES cell, blastula, and breeding strain for the generation of targeted gene mutations in mice (Schoonjans et al. 2003). Using this strategy, the only difference between the KO and WT littermates would be the single gene mutation. If available for numerous inbred strains, researchers will be able to select the “best” strain for their particular research question with which to generate their KO or transgenic mouse, thereby reducing the ambiguity in interpretations of behavioral results.

In summary, evaluating the behavioral phenotype of a new line of transgenic or knockout mice involves a multi-tiered approach that begins with the systematic assessment of general health, home cage behaviors, neurological reflexes, sensory abilities, and motor functions. Intact performance on these initial measures of procedural abilities allows investigators to proceed with specific complex tests in the behavioral domain of interest (e.g., anxiety-like tests, feeding, learning and memory tasks). If preliminary analyses indicate deficits in procedural abilities, then investigators may be able to select or modify complex behavioral tests to avoid false-positive results and artifactual interpretations of genotype differences as diagrammed in Figure 1. In addition, careful attention to the baseline behaviors of the background strain(s), and to the possibility that flanking genes and environmental factors may affect phenotypic effects, will maximize the correct attribution of interesting phenotypic anomalies to the targeted gene mutation.

## References

- Belknap JK, Lame M, Danielson PW. 1990. Inbred strain differences in morphine-induced analgesia with the hot plate assay: A reassessment. *Behav Genet* 20:333-338.

- Benaroya-Milshtein N, Hollander N, Apter A, Kukulansky T, Raz N, Wilf A, Yaniv I, Pick CG. 2004. Environmental enrichment in mice decreases anxiety, attenuates stress responses and enhances natural killer cell activity. *Eur J Neurosci* 20:1341-1347.
- Benefiel AC, Dong WK, Greenough WT. 2005. Mandatory "enriched" housing of laboratory animals: The need for evidence-based evaluation. *ILAR J* 46:95-105.
- Bolivar VJ, Calderone BJ, Reilly AA, Flaherty L. 2000. Habituation of activity in an open field: A survey of inbred strains and F<sub>1</sub> hybrids. *Behavior Genet* 30:285-293.
- Bolivar VJ, Pooler O, Flaherty L. 2001. Inbred strain variation in contextual and cued fear conditioning behavior. *Mamm Genome* 12:651-656.
- Bouwknicht JA, Paylor R. 2002. Behavioral and physiological mouse assays for anxiety: A survey in nine mouse strains. *Behav Brain Res* 136:489-501.
- Bucan M, Abel T. 2002. The mouse: Genetics meets behavior. *Nat Rev Genet* 3:114-123.
- Chesler EJ, Wilson SG, Lariviere WR, Rodriguez-Zas SL, Mogil JS. 2002. Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neurosci Biobehav Rev* 26:907-923.
- Clapcote SJ, Roder JC. 2004. Survey of embryonic stem cell line source strains in the water maze reveals superior reversal learning of 129S6/SvEvTac mice. *Behav Brain Res* 152:35-48.
- Cook MN, Bolivar VJ, McFayden MP, Flaherty L. 2002. Behavioral differences among 129 substrains: Implications for knockout and transgenic mice. *Behav Neurosci* 116:600-611.
- Crabbe JC, Phillips TJ, Feller DJ, Hen R, Wenger CD, Lessov CN, Schafer GL. 1996. Elevated alcohol consumption in null mutant mice lacking 5-HT<sub>1B</sub> serotonin receptors. *Nat Genet* 14:98-101.
- Crabbe JC, Morris RGM. 2004. *Festina lente*: Late-night thoughts on high-throughput screening of mouse behavior. *Nat Neurosci* 7:1175-1179.
- Crabbe JC, Wahlsten D, Dudek BC. 1999. Genetics of mouse behavior: Interactions with laboratory environment. *Science* 284:1670-1673.
- Crawley JN. 1999. Behavioral phenotyping of transgenic and knockout mice: Experimental design and evaluation of general health, sensory functions, motor abilities, and specific behavioral tests. *Brain Res* 835:18-26.
- Crawley JN. 2000. What's Wrong with My Mouse? Behavioral Phenotyping of Transgenic and Knockout Mice. New York: Wiley-Liss.
- Crawley JN. 2003. Behavioral phenotyping of rodents. *Comp Med* 53:140-146.
- Crawley JN, Belknap JK, Collins A, Crabbe JC, Frankel W, Henderson N, Hitzemann RJ, Maxson SC, Miner LL, Silva AJ, Wehner JM, Wynshaw-Boris A, Paylor R. 1997. Behavioral phenotypes of inbred mouse strains: Implications and recommendations for molecular studies. *Psychopharmacology* 132:107-24.
- Crusio WE. 2004. Flanking gene and genetic background problems in genetically manipulated mice. *Biol Psychiatry* 56:381-385.
- Cryan JF, O'Leary OF, Jin S, Friedland JC, Ouyang M, Hirsh BR, Page ME, Dalvi A, Thomas SA, Lucki I. 2004. Norepinephrine-deficient mice lack responses to antidepressant drugs including selective serotonin reuptake inhibitors. *Proc Natl Acad Sci U S A* 101:8186-8191.
- Dobkin C, Rabe A, Dumas R, El Idrissi A, Haubenstock H, Brown WT. 2000. Fmrl knockout mouse has a distinctive strain-specific learning impairment. *J Neurosci* 100:423-429.
- Garattini S, Valzelli L. 1981. Is the isolated animal a possible model for phobia and anxiety? *Prog Neuropsychopharmacol* 5:159-65.
- Garner JP. 2005. Stereotypies and other abnormal repetitive behaviors: Potential impact on validity, reliability, and replicability of scientific outcomes. *ILAR J* 46:106-117.
- Gerlai R. 1996. Gene-targeting studies of mammalian behavior: Is it the genotype or the background genotype? *TINS* 19:177-181.
- Gerlai R. 2001. Gene-targeting: Technical confounds and potential solutions in behavioral brain research. *Behav Brain Res* 125:13-21.
- Gold LH. 1999. Hierarchical strategy for phenotypic analysis in mice. *Psychopharm* 147:2-4.
- Guillot PV, Roubertoux PL, Crusio WE. 1994. Hippocampal mossy fiber distributions and intermale aggression in seven inbred mouse strains. *Brain Res* 660:167-169.
- Holmes A, Lit Q, Murphy DL, Gold E, Crawley JN. 2003a. Abnormal anxiety-related behavior in serotonin transporter null mutant mice: The influence of genetic background. *Genes Brain Behav* 1:55-69.
- Holmes A, Rodgers RJ. 1998. Responses of Swiss-Webster mice to repeated plus-maze experience: Further evidence for a qualitative shift in emotional state. *Pharm Biochem Behav* 60:473-488.
- Holmes A, Wrenn CC, Harris AP, Thayer KE, Crawley JN. 2002. Behavioral profiles of inbred strains on novel olfactory, spatial and emotional tests for reference memory in mice. *Genes Brain Behav* 1:55-69.
- Holmes A, Yang RJ, Lesch K-P, Crawley JN, Murphy DL. 2003b. Mice lacking the serotonin transporter exhibit 5-HT<sub>1A</sub> receptor-mediated abnormalities in tests for anxiety-like behavior. *Neuropsychopharmacology* 28:2077-2088.
- Hutchinson E, Avery A, VandeWoude S. 2005. Environmental enrichment for laboratory rodents. *ILAR J* 46:148-161.
- Jankowsky JL, Melnikova T, Fadale DJ, Xu GM, Slunt HH, Gonzales V, Younkin LH, Younkin SG, Borchelt DR, Savonenko AV. 2005. Environmental enrichment mitigates cognitive deficits in a mouse model of Alzheimer's disease. *J Neurosci* 25:5217-5224.
- Lambert TJ, Fernandez SM, Frick KM. 2005. Different types of environmental enrichment have discrepant effects on spatial memory and synaptophysin levels in female mice. *Neurobiol Learn Mem* 83:206-216.
- Lazarov O, Robinson J, Tang Y, Hairston LS, Korade-Mirnic Z, Lee VM-Y, Hersh LB, Sopalsky RM, Mirnic K, Sisodia SS. 2005. Environmental enrichment reduces Aβ levels and amyloid deposition in transgenic mice. *Cell* 120:701-713.
- Logue SF, Owen EH, Rasmussen DL, Wehner JM. 1997. Assessment of locomotor activity, acoustic and tactile startle, and prepulse inhibition of startle in inbred strains and F1 hybrids: Implications of genetic background for single gene and quantitative trait loci analyses. *Neuroscience* 80:1075-1086.
- Logue SF, Swartz RJ, Wehner JM. 1998. Genetic correlation between performance on an appetitive-signaled nosepoke task and voluntary ethanol consumption. *Alcohol Clin Exp Res* 22:1912-1920.
- McFadyen MP, Kusek G, Bolivar VJ, Flaherty L. 2003. Differences among eight inbred strains of mice in motor ability and motor learning on a rotorod. *Genes Brain Behav* 2:214-219.
- McIlwain KL, Merriweather MY, Yuva-Paylor LA, Paylor R. 2001. The use of behavioral test batteries: Effects of training history. *Physiol Behav* 73:705-717.
- Moons CPH, Van Wiele P, Ödberg FO. 2004. To enrich or not to enrich: Providing shelter does not complicate handling of laboratory mice. *Contemp Top Lab Anim Sci* 43:18-21.
- Olsson IAS, Dahlborn K. 2002. Improving housing conditions for laboratory mice: A review of "environmental enrichment." *Lab Anim* 36:243-270.
- Owen EH, Logue SF, Rasmussen DL, Wehner JM. 1997. Assessment of learning by the Morris water maze task and fear conditioning in inbred mouse strains and F1 hybrids: Implications of genetic background for single gene mutations and quantitative trait loci analyses. *Neuroscience* 80:1087-1099.
- Paylor R, Crawley JN. 1997. A proposed test battery and constellations of specific behavioral paradigms to investigate the behavioral phenotypes of transgenic and knockout mice. *Hormones Behav* 31:197-211.
- Phillips TJ, Hen R, Crabbe JC. 1999. Complications associated with genetic background effects in research using knockout mice. *Psychopharmacology* 147:5-7.
- Rustay NR, Wahlsten D, Crabbe JC. 2003. Assessment of genetic susceptibility to ethanol intoxication in mice. *Proc Natl Acad Sci U S A* 100:2917-2922.
- Schmitt WB, Sprengel R, Mack V, Draft RW, Seeburg PH, Deacon RM, Rawlins JN, Bannerman DM. 2005. Restoration of spatial working

- memory by genetic rescue of GluR-A-deficient mice. *Nat Neurosci* 8:270-272.
- Schoonjans L, Kreemers V, Danloy S, Moreadith RW, Laroche Y, Collen, D. 2003. Improved generation of germline-competent embryonic stem cell lines from inbred mouse strains. *Stem Cells* 21:90-97.
- Simpson EM, Linder CC, Sargent EE, Davisson MT, Mobraaten LE, Sharp JJ. 1997. Genetic variation among 129 substrain and its importance for targeted mutagenesis in mice. *Nat Genet* 16:19-27.
- Steiner RA, Hohmann JG, Holmes A, Wrenn CC, Cadd G, Juréus A, Clifton DK, Luo M, Gutshall M, Ma SY, Mufson, EJ, Crawley JN. 2001. Galanin transgenic mice display cognitive and neurochemical deficits characteristic of Alzheimer's disease. *Proc Natl Acad Sci U S A* 98:4184-4189.
- Tecott LH, Nestler EJ. 2004. Neurobehavioral assessment in the information age. *Nat Neurosci* 7:462-466.
- Valzelli L. 1973. The "isolation syndrome" in mice. *Psychopharmacologia* 31:305-320.
- van der Staay FJ, Steckler T. 2001. Behavioral phenotyping of mouse mutants. *Behav Brain Res* 125:3-12.
- van der Staay FJ, Steckler T. 2002. The fallacy of behavioral phenotyping without standardization. *Genes Brain Behav* 1:9-13.
- van Praag H, Kempermann G, Gage FH. 2000. Neural consequences of environmental enrichment. *Nat Rev Neurosci* 1:191-198.
- Võikar V, Kõks S, Vasar E, Rauvala H. 2001. Strain and gender differences in the behavior of mouse lines commonly used in transgenic studies. *Physiol Behav* 72:271-281.
- Võikar V, Polus A, Vasar E, Rauvala H. 2005. Long-term individual housing in C57Bl/6J and DBA/2 mice: Assessment of behavioral consequences. *Genes Brain Behav* 4:240-252.
- Võikar V, Vasar E, Rauvala H. 2004. Behavioral alterations induced by repeated testing in C57Bl/6J and 129S2/Sv mice: Implications for phenotyping screens. *Genes Brain Behav* 3:27-38.
- Wahlsten D. 2001. Standardizing tests of mouse behavior: Reasons, recommendations, and reality. *Physiol Behav* 73:695-704.
- Wahlsten D, Metten P, Crabbe JC. 2003a. Survey of 21 inbred mouse strains in two laboratories reveals that BTBR T/+ *tff/f* has severely reduced hippocampal commissure and absent corpus callosum. *Brain Res* 971:47-54.
- Wahlsten D, Metten P, Crabbe JC. 2003b. A rating scale for wildness and ease of handling laboratory mice: Results for 21 inbred strains tested in two laboratories. *Genes Brain Behav* 2:71-79.
- Wahlsten D, Metten P, Phillips TJ, Boehm II SL, Burkhart-Kasch S, Dorow J, Doerksen S, Downing C, Fogarty J, Rodd-Henricks K, Hen R, McKinnon CS, Merrill CM, Nolte C, Schalomon M, Sibert JR, Wenger CD, Dudek BC, Crabbe JC. 2003c. Different data from different labs: Lessons from studies of gene-environment interaction. *J Neurobio* 54:283-311.
- Wolfer DP, Crusio WE, Lipp HP. 2002. Knockout mice: Simple solutions to the problems of genetic background and flanking genes. *TINS* 25:336-340.
- Wolfer DP, Litvin O, Morf S, Nitsch RM, Lipp HP Würbel H. 2004. Cage enrichment and mouse behavior. *Nature* 432:821-822.
- Würbel H. 2001. Ideal homes? Housing effects on rodent brain and behaviour. *TINS* 24:207-211.