Published in final edited form as:

J Neurosci Methods. 2014 September 30; 235: 59–64. doi:10.1016/j.jneumeth.2014.06.002.

Validation of video motion-detection scoring of forced swim test in mice

Vance Gao^{a,*}, Martha Hotz Vitaterna^a, and Fred W. Turek^a

Martha Hotz Vitaterna: m-vitaterna@northwestern.edu; Fred W. Turek: fturek@northwestern.edu

^aCenter for Sleep and Circadian Biology, Department of Neurobiology, Northwestern University, 2205 Tech Drive, Hogan Hall 2-160 Evanston, Illinois 60208

1. Introduction

The forced swim test (FST), or Porsolt test, was created as an animal model to predict the effectiveness of potential antidepressant treatments. In this test, a rat (Porsolt and Pichon 1977) or a mouse (Porsolt et al. 1977) is placed into a container filled with water with no possibility of escape for several minutes. Usually, after an initial period of swimming and attempting to escape, the animal adopts a posture of immobility and floats passively in the water. In some protocols, usually with rats, the animal is tested again 24 hours later, during which its immobility is generally increased. The time spent immobile during the test is traditionally interpreted as a measure of “behavioral despair” or “learned helplessness”, qualities which are associated with depression. Many types of antidepressant compounds reduce the amount of time an animal spends immobile in the FST, which supports this idea (Porsolt et al. 1978). However, the despair interpretation has been criticized as overly anthropomorphic, and some researchers suggest that immobility in the FST is more due to adaptive learning rather than helplessness (De Pablo et al. 1989; West 1990). In addition, the FST’s psychophysiological similarity to human depressive disorder is tenuous; for instance, an acute dose of antidepressants changes FST behavior within an hour of administration, but treatment of depressive patients require weeks of chronic administration. Therefore it is safer to consider the FST as a test which does not directly model human depression symptoms, but nevertheless is very sensitive to monoaminergic alterations and other antidepressive treatments (Holmes 2003; Petit-Demouliere et al. 2005).

The FST is usually quantified visually by human scorers. Typically, scorers will view a video of a test trial and measure the amount of time the animal spends immobile, passively floating in the water. However, for experiments with large sample sizes, visual scoring can become a very time-consuming and tedious task. As large-scale genomics studies become more popular, sample sizes in the hundreds will become increasingly common. For example,

© 2014 Elsevier B.V. All rights reserved.

*Corresponding author: Vance Gao: Tel.: (847) 491-5687; v-gao@u.northwestern.edu; Address: Northwestern University, 2205 Tech Dr., Hogan 2-160, Evanston, Illinois 60208.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

486 F2 rats are phenotyped in Tomida et al. 2009 and 560 F2 mice are used in Yoshikawa et al. 2002. In addition to time considerations, there is considerable subjectivity in scoring the FST, making it difficult to split the task among multiple scorers while retaining consistency.

Because of this, several techniques for automated FST quantification have been created in the past. Shimazoe et al. 1987 measured the mechanical vibration on the tank walls caused by swimming rats, De Pablo et al. 1989 recorded disturbances of the electromagnetic field created by movement of water, and Tye et al. 2012 quantified hindpaw movement with magnetic induction. One commercially-available FST system uses infrared beam breaks to quantify movement (Kinder Scientific, CA, USA; Kurtuncu et al. 2005). Video analysis and motion tracking techniques have also been employed, first recording trials with cameras and then using computer programs to quantify the lateral distance moved (Hédou et al. 2001; Crowley et al. 2004), categorize movement as resting, climbing, or swimming (Nie et al. 2008; Biobserve, St. Augustin, Germany), measure amount of limb and tail motion (Gersner et al. 2009), or determine the amount of body submersion and speed of movement (Juszczak et al. 2008). However, all these methods require either specialized equipment or software.

Here, we have developed a simple method implemented in MATLAB (2013a, The MathWorks) to quantify the amount of motion in a digital video file and applied it to our FST data. Comparison of computer scoring and traditional visual scoring shows that results from the two techniques were well-aligned with each other. We then confirm that this method can detect group differences in activity between the first and second days of testing and between different genotypes, and the method replicates observed differences observed by visual scoring. Furthermore, we observe heterosis in the FST, in which hybrid F1 mice are more active in the test than either parental strain.

2. Methods

2.1. Animals

We obtained from the Jackson Laboratory male mice of four genotypes: the two parental strains C57BL/6J (B6) and A/J, plus (C57BL/6JxA/J) F1 and F2 hybrids. Mice were 4 weeks old when they arrived at the lab. They were singly-housed without enrichment items in 16×29 cm cages and maintained on a 12 hour light: 12 hour dark cycle (lights on at 7 am), at a room temperature of $23 \pm 2^{\circ}\text{C}$, with food and water *ad libitum*. The animals first experienced several stressors and behavioral tests, described below, before being tested in the FST. Animals were tested in the FST for two days in a row. However, for about 25% of the mice, only data from one of the two testing days is used in analysis; this is because two cameras recording in different conditions were used to record the FST trials, and video-analysis from only one of the cameras is presented here for consistency of results. Mice which were recorded with the other camera on one of the two days therefore only have data from one testing day. The sample sizes of each day are listed in Table 1. Protocols were approved by the Northwestern University Animal Care and Use Committee.

2.2. Previous Behavioral Tests and Stressors

The methods which will be described here are part of a much larger chronic-stress behavioral-testing protocol. However, presently we are only interested in validating this FST

scoring method, so we will restrict our attention to the FST and the stressors which occurred before the FST in the protocol. In the current analysis, these stressors are considered only as common previous experiences shared by all subjects, rather than as tests in their own right; therefore data from these additional stressors will not be presented here.

Mice experienced the open field arena (OFA) and elevated plus maze (EPM) at 7 weeks of age. In the OFA, the mouse is placed in a novel square arena (52.7 × 52.7 cm) with high walls. This is a test which is often used as a measure of anxiety in rodents (Prut and Belzung 2003); it is stressful because mice prefer enclosed areas rather than open fields, and because it is a novel environment. The mice's body weights were measured immediately after the OFA.

One day after the OFA, Stress-group mice experienced the EPM. The maze consists of a cross-shaped arena, with four arms (52.8 cm long × 5.7 cm wide) radiating at right angles from a center. The maze is elevated 30.5 cm off the ground. Two of the arms are enclosed by high walls (14 cm), and two of the arms consist only of platforms without walls. This is another frequently-used test of anxiety, which takes advantage of the conflict between a rodent's desire for exploration and desire for an enclosed environment (Walf and Frye 2007).

Mice in the Stress group experienced restraint stress at 8 weeks of age. Mice were restrained in hard plastic tubes for 15 minutes. Furthermore, tail bleeds were performed on these mice when the mouse was first placed in the tube, right before the mouse was taken out of the tube, and then 15 and 30 minutes after the mouse was taken out of the tube.

2.3. Forced Swim Test

Mice were tested in the FST at 9 weeks of age, and tested twice over two days. In each test, a mouse is placed in a glass beaker (25 cm tall, 15 cm diameter) filled with approximately 16 cm of water kept at 23.5 ± 0.5 °C. The mouse was left in the water for 6 minutes with no possibility of escape. The test was recorded using a digital video camera from the side at 30 frames per second, at a distance of approximately 30–40 cm. Ideally, the camera should be exactly at water level so that the amount of visible water movement is minimized, but our recordings varied somewhat in camera elevation. Because of the large number of mice tested in a short period of time, several animals were tested using the same water, and fecal boli were removed from the water in between tests. After testing, the mouse was placed back into its home cage, which was placed halfway onto a warming pad for 15 minutes, so that the mouse was able to thermoregulate by moving between the warmer and cooler ends of the cage. Mice were tested in the FST for a second time 24 hours later using the same procedure. Mice generally spend more time immobile during the second day of testing (MacQueen et al. 2001; Urani et al. 2001). The FST was performed between 6–11 hours after light onset (ZT 6–11).

The recorded video files were visually scored at one-second resolution. During each one-second interval, the mouse was counted as immobile if it was passively floating or gently paddling with one foot. The latency to immobility was the length of time from the beginning

of the test to when the mouse was first scored as immobile. The 134 videos which were visually scored were all scored by the same person.

2.4. Computer scoring

A method to quantify the amount of motion in a video file was written in MATLAB. For each frame k , each pixel's intensity value in each of the three 8-bit RGB channels is compared with that pixel's intensity value in frame $k + 1$. A difference is only counted if the difference is greater than or equal to a threshold value of 5. Adding the count of differences from the three 8-bit channels for all pixels will give the total difference between frame k and $k + 1$. This is then repeated for all frames in the video. Video files were recorded at 30 frames-per-second during the 6-minute long FST trial, so there were a total of 10800 frames per video; however, the first four seconds of video were not analyzed because of excessive camera jitter. A sample of one mouse's motion record is displayed in Fig. 1.

When videos of the forced swim tests were recorded, the distance between the camera and animals varied between trials. The mice and the beaker in the video's image appear larger when the camera is closer, causing the amount of motion detected in a video to be artificially increased. To correct for this, we measured the width of the beaker in each video and used this as a standard length to correct for distance. These corrections were applied after average motion values for each mouse had been calculated. The raw motion averages are divided by the square of the beaker length, since the two-dimensional size of the mouse's image in the video (and thus the motion detected) is proportional to distance squared. For future usage, this correction is not necessary if the recording camera's distance to the beaker is kept constant between trials.

$$Motion_{Dist.} = \frac{Raw\ motion}{(Beaker\ Width)^2}$$

An additional correction for the mouse's body weight was also applied, which we found slightly improved results. Assuming geometrical scaling, a mouse's body weight should be proportional to body length cubed, and the two-dimensional size of the mouse's image should be proportional to body length squared. Body weight data was collected 2 weeks prior to the FST, immediately following the OFA.

$$Motion_{BW} = \frac{Motion_{Dist.}}{(Body\ Weight)^{2/3}}$$

For calculating the immobility time and latency to immobility, a moving-average smoothing function was first applied to each mouse's movement data. For this, each frame's raw motion value was averaged with the motion values of the next 30 frames (30 frames = 1 second) to form the smoothed motion value for that frame. A threshold motion level was determined for each animal so that a mouse was counted as immobile when its motion was under the threshold. The threshold levels were based on a threshold constant, which was

then modified according to camera distance and body weight using the above equations. The threshold constant was set so that the resulting immobility values closely matched those from visual scoring. The latency to immobility was calculated by finding the first time the mouse's motion was below threshold for one full second.

2.5. Statistical Analysis

Differences between groups were tested with Student's *t*-test with equal variance. Analysis of variance (ANOVA) was used to determine if genotype had a significant effect on motion values.

3. Results

3.1. Validation of Computer Scoring

The Pearson correlation between computer-scored total motion and visual-scored immobility was $r = -0.80$ (Fig. 2a). The correlation between computer-scored immobility and visual-scored immobility was $r = 0.83$ (Fig. 2b). The two computer-based measures, total motion and immobility, were more correlated to each other ($r = -0.97$) than either was to visual-scored immobility. Computer-scored and visual-scored latency to immobility also were well-correlated with each other, $r = 0.79$ (Fig. 2c). Since the latency data is subject to outliers, caused by brief pauses in movement which were counted as immobility in one method but not the other, we also calculated the Spearman rank correlation, which was $\rho = 0.85$.

We wondered if the mouse's coat color would artificially increase or decrease motion detection values. A/J mice have the recessive *Tyrp1* allele (b/b) which causes brown fur, but is overridden by a tyrosinase recessive mutation (c/c) which makes them albino. B6 mice (B/B; C/C) and F1 mice (B/b; C/c) both have black coats, and our F2 mice showed the expected coat-color ratio of 4 albino: 9 black: 3 brown ($n = 17:44:12$). To see if coat color had a confounding effect on motion detected, we used the linear regression equation between computer-scored total motion and visual-scored immobility time to calculate predicted immobility time values from the total motion values, and then calculated the residuals between the predicted and actual immobility time values. We then tested if coat color had an effect on the residuals; for example, a coat color which artificially increases motion values should have predicted immobility times higher than those which were visually observed. There was no significant effect of coat color on these residuals ($p = 0.69$, $F(2,131) = 0.37$) (Table 2), suggesting that coat color does not have a large confounding effect on motion detection.

3.2. Group Differences of Day and Genotype

Next, we wanted to confirm that our motion detection method is able to reproduce observed differences between groups detected by traditional visual scoring. Using visual-scored immobility, computer-scored total motion, and computer-scored immobility measures, we performed T-tests to test for differences between Day 1 and Day 2 of testing, as well as for strain differences on each day of testing (Fig. 4). Results show that computer-scoring detected all significant differences which were detected by visual scoring, and also detected

some significant differences which visual-scoring did not (Table 3). When comparing within each genotype, only F2 mice showed a reduction of activity from Day 1 to Day 2. However, when grouping together A/J, B6, and F1 mice, this reduction in activity becomes apparent, but only when using the computer-scored data. Of the two parental strains, A/J mice were more active than B6. Interestingly, the hybrid F1 mice tend to have greater levels of activity than the parental A/J and B6 strains, rather than an intermediate amount, as might initially be expected.

4. Discussion

4.1. Automated Scoring

The automated scoring method that we developed agrees well with traditional visual scoring. There was high correlation between computer-scored total motion and visual-scored immobility, as well as between visual-scored and computer-scored immobility. No confounding effect of coat color on motion detection was observed. In group comparisons between days of testing and genotype, the computer-scored measures detected almost every significant difference that visual-scoring did; in addition, computer-scoring was able to detect differences which visual-scoring did not.

Between the two computer-scored measures, immobility was more correlated to visual scoring than total motion was, which may make this measure favored; on the other hand, measuring immobility requires an extra step of threshold calibration. Although we used all 134 of our data points to determine the threshold constant, we find that good correlations between computer-scoring and visual-scoring can be obtained with a relatively wide range of threshold constants. Indeed, when setting threshold constants using just two randomly-selected data points, the resulting correlations are almost always above $r = 0.81$, resulting in very similar between-group relationships, though the resulting immobility times can deviate from visual-scored immobility by quite a bit. If reduction of such deviations is desired, we recommend using at least 10 animals to calibrate the threshold constant.

When we originally performed these behavioral tests, we did not plan on using an automated scoring method, and thus we did not optimize the testing and recording setup for video processing. Our camera's distance and elevation was not constant, the lighting and background varied, and body weights were measured 2 weeks prior to the FST instead of immediately before the test. These deficiencies in recording setup could easily be corrected in future applications of this method. Standardizing these aspects of the test will reduce variability and presumably would have improved the correlation between computer-scoring and visual-scoring. In addition to scoring the FST, this method could probably also be used to quantify any other behavioral tests in which overall activity is a measure, such as a tail suspension test or circadian activity.

4.2. Strain Differences

A/J mice were more active than C57BL/6 mice in the FST. This result matches those previously reported by Lucki et al. (2001) which showed that A/J mice spent less time immobile, though it contradicts a different study showed that A/J mice were more immobile than C57BL/6 mice (Flaisher-Grinberg and Einat 2010). It is unclear what these different

results are due to. One possible source of discrepancy is that Lucki et al. visually scored their mice, whereas Flaisher-Grinberg & Einat used a commercial automated system (Biobserve), or perhaps subtle differences in housing conditions and testing procedures produced different behaviors in the test.

F1 mice were more active than both of the parental A/J and B6 strains in the FST. This is apparently an instance of heterosis, the phenomenon in which a hybrid organism shows a more extreme phenotype than either of its parental strains. (Whereas “heterosis” refers to the extreme phenotype of a hybrid cross between two strains or species, the related term “overdominance” refers to the extreme phenotype of a heterozygote. However, “overdominance” is also sometimes used in the first sense, especially when discussing F1 crosses of genetically identical inbred strains.) Heterosis was not observed in another FST study using (B6 x C3H/He) F1 mice (Yoshikawa et al. 2002) but was apparently present in a study with Wistar Kyoto x Fisher 344 rats, though this was not explicitly tested for (Solberg et al. 2004). In mice, researchers have demonstrated heterotic inheritance of traits such as wheel-running (Bruell 1964), alcohol preference (Blednov et al. 2010), fear conditioning (Owen et al. 1997), and sleep characteristics (Friedmann 1974). It may be the case that heterosis in FST is observed in crosses of some strains and not others, as is the case for the other traits. It would be interesting to test whether this effect is observed in crosses of other strains and whether such effects are actually indicative of alterations in psychiatric processes.

4.3. Other Testing Factors

Our mice were subjected to several stressors and tests before the FST, which has been demonstrated to increase immobility in the FST (Voikar et al. 2004). As an additional stressor, our mice were singly-housed, which could have affected their FST behavior. Vöikar et al. (2005) showed no effect of group vs. single-housing on B6 mice, but single-housing reduces immobility in experiments with C57BL/6JOLA^{Hsd} (Kuleskaya et al. 2011), DBA (Vöikar et al. 2005) and NIH-Swiss strains (Hilakivi et al. 1989; Karolewicz and Paul 2001); however, Swiss-Webster mice isolated for a shorter 24-hour period were more immobile in the FST rather than less (Yates et al. 1991). It is unknown how A/J mice react to isolation. Our mice also were not provided environmental enrichment; environmental enrichment has been shown to decrease immobility times in mice (Llorens-Martín et al. 2007) as well as rats (Magalhães et al. 2004). In addition, because of the large number of animals we tested (only a subset of which are presented here), we retested several mice using the same water. This has been shown to increase immobility in rats which were tested water previously used for other individuals (Abel 1991; Abel et al. 1992), and one can assume the same effect would be seen with mice. However, the order of mouse testing was relatively random, such that no particular genotype was consistently tested first or last, so any effects of water reuse in our data should be washed out as noise.

5. Conclusion

We successfully applied a video motion-quantification method to automatically score the FST, and we show that results from computer-scoring match results from visual-scoring well. We also use this method to demonstrate heterosis in the FST. Although other

automated FST methods already exist, they are commercial products with hardware and/or software which must be purchased. This method offers an effective and low-cost alternative.

Acknowledgments

This work was supported by The Defense Advanced Research Projects Agency (government contract/grant number W911NF1010066). The views, opinions, and/or findings contained in this article are those of the author and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense. Vance Gao is also grateful to NIH T32 HL007909 for support. Special thanks to Mark Dow of the Brain Development Lab at the University of Oregon for developing the motion detection method and providing the code online (http://lcnf.uoregon.edu/~dow/Geek_software/Video_motion/Video_motion_estimation.html#video_total_motion_estimation).

References

- Abel EL. Gradient of alarm substance in the forced swimming test. *Physiol Behav.* 1991; 49(2):321–3. [PubMed: 2062904]
- Abel EL, Altman HJ, Commissaris RL. Maudsley reactive and nonreactive rats in the forced swim test: Comparison in fresh water and solied water. *Physiol Behav.* 1992; 52(6):1117–9. [PubMed: 1484869]
- Blednov YA, Ozburn AR, Walker D, Ahmed S, Belknap JK, Harris RA. Hybrid Mice as Genetic Models of High Alcohol Consumption. *Behav Genet.* 2010; 40(1):93–110. [PubMed: 19798565]
- Bruell JH. Heterotic inheritance of wheelrunning in mice. *J Comp Physiol Psychol.* 1964; 58:159–63. [PubMed: 14215383]
- Crowley JJ, Jones MD, O’Leary OF, Lucki I. Automated tests for measuring the effects of antidepressants in mice. *Pharmacol Biochem Behav.* 2004; 78(2):269–74. [PubMed: 15219767]
- Flaisher-Grinberg S, Einat H. Strain-specific battery of tests for domains of mania: effects of valproate, lithium and imipramine. *Front Psychiatry.* 2010; 1:10. [PubMed: 21423422]
- Friedmann JK. A diallel analysis of the genetic underpinnings of mouse sleep. *Physiol Behav.* 1974; 12(2):169–75. [PubMed: 4361192]
- Gersner R, Gordon-Kiwkowitz M, Zangen A. Automated behavioral analysis of limbs’ activity in the forced swim test. *J Neurosci Methods.* 2009; 180(1):82–6. [PubMed: 19427533]
- Hédou G, Pryce C, Di Iorio L, Heidbreder CA, Feldon J. An automated analysis of rat behavior in the forced swim test. *Pharmacol Biochem Behav.* 2001; 70(1):65–76. [PubMed: 11566143]
- Hilakivi LA, Ota M, Lister R. Effect of isolation on brain monoamines and the behavior of mice in tests of exploration, locomotion, anxiety and behavioral “despair. *Pharmacol Biochem Behav.* 1989; 33(2):371–4. [PubMed: 2479036]
- Holmes PV. Rodent Models of Depression: Reexamining Validity Without Anthropomorphic Inference. *Crit Rev Neurobiol.* 2003; 15(2):143–74. [PubMed: 14977368]
- Juszczak GR, Lisowski P, Œliwa AT, Swiergiel AH. Computer assisted video analysis of swimming performance in a forced swim test: Simultaneous assessment of duration of immobility and swimming style in mice selected for high and low swim-stress induced analgesia. *Physiol Behav.* 2008; 95(3):400–7. [PubMed: 18656493]
- Karolewicz B, Paul IA. Group housing of mice increases immobility and antidepressant sensitivity in the forced swim and tail suspension tests. *Eur J Pharmacol.* 2001; 415(2–3):197–201. [PubMed: 11274999]
- Kuleskaya N, Rauvala H, Voikar V. Evaluation of Social and Physical Enrichment in Modulation of Behavioural Phenotype in C57BL/6J Female Mice. *PLoS One.* 2011; 6(9):e24755. [PubMed: 21931844]
- Kurtuncu M, Luka LJ, Dimitrijevic N, Uz T, Manev H. Reliability assessment of an automated forced swim test device using two mouse strains. *J Neurosci Methods.* 2005; 149(1):26–30. [PubMed: 15967510]
- Llorens-Martín MV, Rueda N, Martínez-Cué C, Torres-Alemán I, Flórez J, Trejo JL. Both increases in immature dentate neuron number and decreases of immobility time in the forced swim test

- occurred in parallel after environmental enrichment of mice. *Neuroscience*. 2007; 147(3):631–8. [PubMed: 17570601]
- Lucki I, Dalvi A, Mayorga A. Sensitivity to the effects of pharmacologically selective antidepressants in different strains of mice. *Psychopharmacology (Berl)*. 2001; 155(3):315–22. [PubMed: 11432695]
- MacQueen GM, Ramakrishnan K, Croll SD, Siuciak JA, Yu G, Young T, Fahnstock M. Performance of heterozygous brain-derived neurotrophic factor knockout mice on behavioral analogues of anxiety, nociception, and depression. *Behav Neurosci*. 2001; 115(5):1145–53. [PubMed: 11584927]
- Magalhães A, Summavielle T, Tavares MA, de Sousa L. Effects of Postnatal Cocaine Exposure and Environmental Enrichment on Rat Behavior in a Forced Swim Test. *Ann N Y Acad Sci*. 2004; 1025(1):619–29. [PubMed: 15542771]
- MATLAB Release. The Mathworks, Inc; Natick, Massachusetts, United States: 2013a.
- Nie, Y.; Ishii, I.; Yamamoto, K.; Takaki, T.; Orito, K.; Matsuda, H. High-speed video analysis of laboratory rats behaviors in forced swim test. 4th IEEE Conf Autom Sci Eng; 2008. p. 206–11.
- Owen EH, Logue SF, Rasmussen DL, Wehner JM. Assessment of learning by the Morris water 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*. 1997; 80(4):1087–99. [PubMed: 9284062]
- De Pablo JM, Parra A, Segovia S, Guillamón A. Learned immobility explains the behavior of rats in the forced swimming test. *Physiol Behav*. 1989; 46(2):229–37. [PubMed: 2602464]
- Petit-Demouliere B, Chenu F, Bourin M. Forced swimming test in mice: a review of antidepressant activity. *Psychopharmacology (Berl)*. 2005; 177(3):245–55. [PubMed: 15609067]
- Porsolt R, Bertin A, Jalfre M. Behavioral despair in mice: a primary screening test for antidepressants. *Arch Int Pharmacodyn therapie*. 1977; 229(2):327–36.
- Porsolt R, le Pichon M. Depression: a new animal model sensitive to antidepressant treatments. *Nature*. 1977; 266:730–2. [PubMed: 559941]
- Porsolt RDR, Anton G, Blavet N, Jalfre M. Behavioural despair in rats: a new model sensitive to antidepressant treatments. *Eur J Pharmacol*. 1978; 47(4):379–91. [PubMed: 204499]
- Prut L, Belzung C. The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review. *Eur J Pharmacol*. 2003; 463(1–3):3–33. [PubMed: 12600700]
- Shimazoe T, Shibata S, Ueki S. A new forced swimming test for the evaluation of antidepressants in rats by recording vibration of a water tank. *J Pharmacobiodyn*. 1987; 10(11):639–43. [PubMed: 3446771]
- Solberg LC, Baum AE, Ahmadiyeh N, Shimomura K, Li R, Turek FW, et al. Sex- and lineage-specific inheritance of depression-like behavior in the rat. *Mamm Genome*. 2004; 15(8):648–62. [PubMed: 15457344]
- Tomida S, Mamiya T, Sakamaki H, Miura M, Aosaki T, Masuda M, et al. Usp46 is a quantitative trait gene regulating mouse immobile behavior in the tail suspension and forced swimming tests. *Nat Genet*. 2009; 41(6):688–95. [PubMed: 19465912]
- Tye KM, Mirzabekov JJ, Warden MR, Ferenczi EA, Tsai H-C, Finkelstein J, et al. Dopamine neurons modulate neural encoding and expression of depression-related behaviour. *Nature*. 2012; 493(7433):537–41. [PubMed: 23235822]
- Urani A, Roman FJ, Phan VL, Su TP, Maurice T. The Antidepressant-Like Effect Induced by ζ 1-Receptor Agonists and Neuroactive Steroids in Mice Submitted to the Forced Swimming Test. *J Pharmacol Exp Ther*. 2001; 298(3):1269–79. [PubMed: 11504830]
- Võikar V, Polus a, Vasar E, Rauvala H. Long-term individual housing in C57BL/6J and DBA/2 mice: assessment of behavioral consequences. *Genes Brain Behav*. 2005; 4(4):240–52. [PubMed: 15924556]
- Voikar V, Vasar E, Rauvala H, Võikar V. Behavioral alterations induced by repeated testing in C57BL/6J and 129S2/Sv mice: implications for phenotyping screens. *Genes, Brain Behav* Munksgaard International Publishers. 2004; 3(1):27–38.
- Walf AA, Frye CA. The use of the elevated plus maze as an assay of anxiety-related behavior in rodents. *Nat Protoc*. 2007; 2(2):322–8. [PubMed: 17406592]

- West, aP. Neurobehavioral studies of forced swimming: the role of learning and memory in the forced swim test. *Prog Neuropsychopharmacol Biol Psychiatry*. 1990; 14(6):863–77. [PubMed: 2277851]
- Yates G, Panksepp J, Ikemoto S, Nelson E, Conner R. Social isolation effects on the “behavioral despair” forced swimming test: Effect of age and duration of testing. *Physiol Behav*. 1991; 49(2): 347–53. [PubMed: 2062907]
- Yoshikawa T, Watanabe A, Ishitsuka Y, Nakaya A, Nakatani N. Identification of Multiple Genetic Loci Linked to the Propensity for “Behavioral Despair” in Mice. *Genome Res*. 2002 Mar 1; 12(3): 357–66. [PubMed: 11875023]

Highlights

- We use a video motion-detection method in MATLAB to quantify the forced swim test.
- Data obtained by computer scoring correlate well with visual scoring.
- Computer scoring detected almost all group differences that visual scoring did, plus some which visual scoring did not.
- We observe heterosis, in which hybrid (A/J x C57BL/6) F1 mice are more active than the parental strains in the test.

Fig. 1.

Sample motion record. The motion record of a B6 mouse in the Acute Stress group is displayed here, after applying a smoothing function. 6 minutes of the FST is displayed. This mouse started with high activity while struggling to escape the beaker, ceased activity abruptly for several seconds before resuming swimming, then settled into alternating bouts of passive floating and swimming.

Fig. 2.

Computer scoring vs. visual scoring. (A) Comparison of computer-scored total motion and visual-scored immobility time. Units on the computer axis are arbitrary. Each point represents one mouse ($n = 134$). Pearson correlation $r = -0.80$. (B) Immobility time measured by computer-scoring and visual-scoring. Pearson $r = 0.83$. (C) Latency to immobility. Pearson $r = 0.79$, Spearman $\rho = 0.85$.

Fig. 3. Summed motion per minute of mice on (A) the first FST trial and (B) the second FST trial. Motion was computer-scored. All error bars are SE. One-way ANOVA testing for effect of genotype was computed for each minute; * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$.

Fig. 4.

Group comparisons. Day 1 vs. Day 2 differences for each strain are graphed using the following measures: (A) visual-scored mobility, (B) computer-scored total motion, and (C) computer-scored mobility. Time mobile is plotted in (A) and (C) instead of time immobile to make it visually easier to compare with the total-motion graph; time mobile is equal to 360 seconds minus time immobile. Error bars represent standard deviations.

Table 1**Sample sizes**

Number of mice of each genotype used in Day 1 and Day 2 of FST testing. All FST videos were both computer-scored and visual-scored. In total, n=134.

	Day 1	Day 2
A/J	11	11
B6	10	12
F1	10	7
F2	41	32

Table 2

Fur color residual analysis. Predicted immobility times were calculated based on each mouse's total motion data using the linear regression equation between computer-scored total motion and human-scored immobility time. Mean residuals (predicted minus actual immobility times) are displayed, as well as number of mice of each fur color. ANOVA revealed no significant effect of fur color on residuals ($p = 0.69$, $F(2,131) = 0.37$).

Fur Color	Number	Predicted – Actual (Mean $\pm$ SD)
Albino	40	3.0 ± 32.1 s
Black	82	-2.3 ± 40.4 s
Brown	12	5.6 ± 52.3 s

Table 3

To see if automatic motion detection could reliably detect differences between groups, T-tests were used to test for group differences using visual-scored and computer-scored measures. The corresponding graphs for these comparisons can be found in Fig. 4. “Inbred” refers to the A/J and B6 strains combined.

	Visual: Immobility Time	Computer: Total Motion	Computer: Immobility Time
<i>Day 1 vs. Day 2</i>			
A/J	-	-	-
B6	-	-	-
F1	-	-	-
F2	••	••	••
A/J, B6, F1	-	•••	•
<i>Strain Comparisons, Day 1</i>			
A/J - B6	•	•	••
A/J - F1	-	•	•
A/J - F2	-	-	-
B6 - F1	•••	•••	•••
B6 - F2	•••	••	•••
F1 - F2	-	-	-
Inbred - F1	••	•••	•••
<i>Strain Comparisons, Day 2</i>			
A/J - B6	•	•	•
A/J - F1	-	-	-
A/J - F2	-	-	-
B6 - F1	••	••	•
B6 - F2	•	••	•
F1 - F2	-	-	-
Inbred - F1	•	••	-

• denotes $p < .05$;

•• denotes $p < 0.01$, and

••• denotes $p < 0.001$.