



Basic Neuroscience

Validation and implementation of a novel high-throughput behavioral phenotyping instrument for mice

Jesse Brodkin ^{a,*}, Dana Frank ^a, Ryan Grippo ^b, Michal Hausfater ^b, Maria Gulinello ^c, Nils Achterholt ^d, Christian Gutzen ^d

^a Behavioral Instruments, 5 Jill Court Unit 1, Hillsborough, NJ 08844, United States

^b Department of Psychology, Rutgers University, Piscataway, NJ 08854, United States

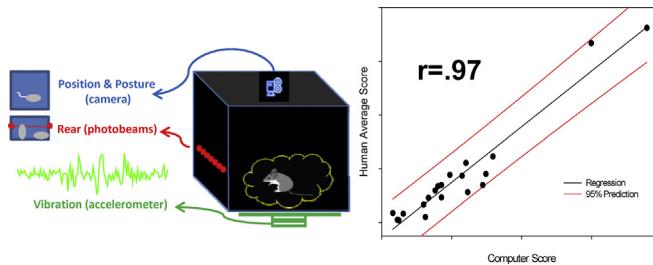
^c Behavioral Core Facility, Albert Einstein College of Medicine of Yeshiva University, Dominick P. Purpura Department of Neuroscience, Rose F. Kennedy Center, RM 925, 1410 Pelham Pkwy S., Bronx, NY 10461, United States

^d BIOBSERVE GmbH, Siegburger Str. 35, 53757 St. Augustin, Germany

HIGHLIGHTS

- We created a new instrument for measuring mouse behavior using video and vibration.
- The Behavioral Spectrometer scored mice similarly to expert human observers.
- It passed tests of face and construct validity.
- It readily identified models of autism, pain and stress.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 14 October 2013

Received in revised form

18 December 2013

Accepted 22 December 2013

Keywords:

Automated detection
Neurobehavioral assessment
Behavioral screening
Ethological analysis
Grooming
Video analysis

ABSTRACT

Background: Behavioral assessment of mutant mouse models and novel candidate drugs is a slow and labor intensive process. This limitation produces a significant impediment to CNS drug discovery.

New method: By combining video and vibration analysis we created an automated system that provides the most detailed description of mouse behavior available. Our system (The Behavioral Spectrometer) allowed for the rapid assessment of behavioral abnormalities in the BTBR model of Autism, the restraint model of stress and the irritant model of inflammatory pain.

Results: We found that each model produced a unique alteration of the spectrum of behavior emitted by the mice. BTBR mice engaged in more grooming and less rearing behaviors. Prior restraint stress produced dramatic increases in grooming activity at the expense of locomotor behavior. Pain produced profound decreases in emitted behavior that were reversible with analgesic treatment.

Comparison with existing method(s): We evaluated our system through a direct comparison on the same subjects with the current "gold standard" of human observation of video recordings. Using the same mice evaluated over the same range of behaviors, the Behavioral Spectrometer produced a quantitative categorization of behavior that was highly correlated with the scores produced by trained human observers ($r = 0.97$).

Conclusions: Our results show that this new system is a highly valid and sensitive method to characterize behavioral effects in mice. As a fully automated and easily scalable instrument the Behavioral Spectrometer represents a high-throughput behavioral tool that reduces the time and labor involved in behavioral research.

© 2013 Elsevier B.V. All rights reserved.

* Corresponding author at: Behavioral Instruments, 67 Maple Run, Basking Ridge, NJ 07920, United States. Tel.: +1 908 240 6901.

E-mail addresses: Brodkin@behavioralinstruments.com, jessebrodkin@yahoo.com (J. Brodkin), christian.gutzen@biobserve.com (C. Gutzen).

1. Introduction

While molecular biology, medicinal chemistry, and information technology have advanced considerably in the past 20 years, tools for behavioral analysis have lagged behind. Despite the considerable increase in mutant mouse models, success rates for CNS drugs are some of the lowest in the industry. The high rate of attrition for these drugs can be attributed to a lack of efficacy especially in the early stages of their development (Kola and Landis, 2004). Every candidate drug or mutant mouse model developed for a CNS indication requires extensive preclinical behavioral evaluation. Unfortunately, behavioral assays are slow and labor-intensive (for review see Tecott and Nestler, 2004) and thus result in a low rate of throughput. Several factors contribute to this low throughput including: (1) the need for extensive training of animals (e.g., radial mazes and operant procedures); (2) the need for human observation (e.g., pain assessment and reflex testing); and (3) the inability of an experimenter to simultaneously assess multiple animals (e.g., elevated plus maze and water mazes). Despite these drawbacks, behavioral data remain the main drivers of CNS drug discovery. Yet, the limitations of these procedures create a bottleneck in the success of CNS programs. Relieving this bottleneck could substantially accelerate drug discovery for CNS indications.

Several companies have attempted to develop automated solutions to behavioral evaluation like PhenoMaster (TSE-systems), PhenoTyper (Noldus, Inc.) and, most recently, PhenoCube (New-Behavior and Psychogenics, Inc.). These systems have generally relied on video analysis to recognize a small number of behaviors, and modular gadget add-ons to evaluate specific behavioral domains (e.g., lickometers, pellet dispensers and operant manipulanda). Here we have taken a different approach; instead of focusing on specific behavioral domains, we have endeavored to capture a highly detailed representation of the spontaneous behavior of a mouse in an open field. Several recent publications have shown that the open field behavior of rodents contains sufficient complexity and sensitivity to elucidate a wide range of drug effects (Kafkafi et al., 2014; Yucel et al., 2009; Kafkafi et al., 2009; Benjamini et al., 2010). These studies have shown through a careful analysis of video records of open field behavior it is possible to differentiate between therapeutic classes of drugs and mouse strains.

Our system is capable of automatically identifying 23 unique behaviors (see Table 1) and providing a complete, real-time profile of mouse behavior. We accomplish this by the integration of state-of-the-art video analysis (Viewer³, BiObserve, Inc.) with advanced vibration analysis (developed at Behavioral Instruments, Inc.) to create the Behavioral Spectrometer; a tool with the highest resolution of behavioral identification in the industry. This capability allows for rapid, reliable and robust quantitative profiling of mutant mice and candidate drugs, which may increase success rates for therapeutic interventions and identification of underlying mechanisms of CNS disorders.

This report summarizes the validation and implementation of our system. We assessed the face validity of our system by comparing it to the current “gold standard” of human observation. We assessed the construct validity of our measures by using pharmacological, genetic and physiological manipulations in mouse models that produce well-characterized behavioral alterations. We examined water misting, a classic intervention to stimulate mild grooming (Kyzar et al., 2011; Hartley and Montgomery, 2008); restraint stress, which has been used as a model for stress and dramatically increases grooming (Zhang et al., 2011; Dunn and Swiergiel, 1999); and the recently described model for autism, the BTBR mouse (Pobbe et al., 2010; Amodeo et al., 2012). To explore the wider implementation of our system, we also examined pain behavior. Pain has traditionally been assessed through evoked responses from an animal in pain, but there has been

Table 1
Ethogram.

Behavior type	Variable	Description
Rear	Rear_Still	Rearing and still
	Rear_Sniff	Rearing and sniffing with small short head rotations
	Rear_Bob	Rearing and moving whole head and/or body up and down
	Rear_Climb	Rearing and trying to climb using limbs
Groom	Paw	Grooming paws using very short paw and head movements
	Nose	Grooming nose using short paw movements
	Head	Grooming head using paws in long circular strokes
	Face	Grooming face and eyes using paws in medium strokes
	Cheek	Rubbing side of face with arm/shoulder
	Leg	Grooming leg with medium head strokes
	Back	Grooming back with long head strokes
	Genital	Grooming genitals with rapid short head strokes
	Tummy	Grooming abdomen with medium head strokes
	Shimmy	Rapid short rotations of the body like ‘hula dancer’ or wet dog
Orient	Scratch	Rapid strong movements of the leg
	Orient_Sniff	Head extension with small short head rotations
	Orient_Look	Head and shoulder extension with whole head movements
	Orient_Shuffle	Repositioning feet with little displacement of body
Ambulation	Orient_Creep	Moving 1 or 2 steps slowly
	Still	Stationary with no body movements
	Walk	Slow locomotor movement with flat posture
	Trot	Medium locomotor movement with arched posture
	Run	Fast locomotor movement with arched posture

shift in interest placing more focus on spontaneous pain (Mogil and Crager, 2004). Given that spontaneous pain is a major clinical symptom of pain, a full description of the spontaneous behavior, like the one produced by our instrument, is particularly relevant. Together, these experiments establish the validity of the measures produced by our instrument and show the potential application of our system to a wide variety of CNS disorders.

2. Materials and methods

2.1. Apparatus

The Behavioral Spectrometer (Behavioral Instruments, NJ and BiObserve, DE who market it under the name Behavior Sequencer) consists of a 40 cm by 40 cm square arena enclosed at a height of 45 cm. The removable floor is made of aluminum honeycomb sheet that rests on three vibration sensors. A miniature color CCD camera is mounted in the ceiling above the center of the arena. A row of 32 infrared transmitter and receiver pairs is embedded in the walls at a height of 6.5 cm. The Spectrometer is equipped with 4 dimmable

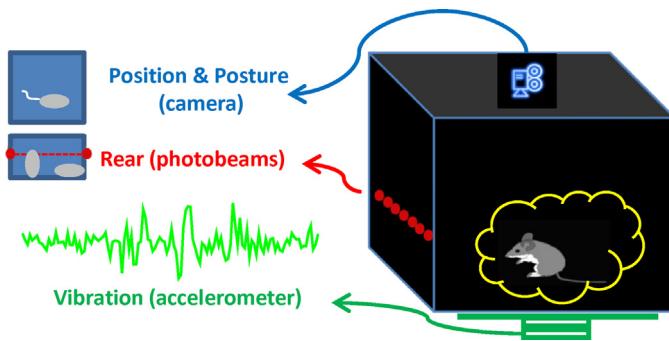


Fig. 1. Diagram of the Behavioral Spectrometer. The box contains a camera, an instrumented floor to measure vibrations and a row of photobeams to capture rearing behavior. One mouse is run at a time and behavior identification takes place automatically in real-time.

10-W halogen lights placed on the ceiling of the enclosure. One side of the enclosure is a door allowing access to the arena.

The mouse's behavior is captured by three types of sensors (Fig. 1): a camera captures video, from which position and posture are extracted (Viewer³, BiObserve); accelerometers embedded in the floor supports capture the mouse's vibrations; and the infrared beams detect when the animal rears. By using a combination of these sensors we were able to generate very detailed descriptions of mouse behavior. Each behavior produces a distinct pattern of sensor data. The software contains 23 different patterns of sensor readings against which to compare any current readings. Through use of a proprietary algorithm utilizing multidimensional clustering to the nearest neighbor in the sensor data space the computer determines the most likely behavior being emitted by the mouse at every second. A video record synchronized with the computer scored behavior is available for post-session inspection (for details see: <http://www.youtube.com/watch?v=8GUw0wR43dY>). The Spectrometer software interface allows for numerous experimental variables (e.g., experiment name, mouse number, treatment, etc.) to be attached to each run as well as the automation of the parameters of each session (e.g., start, delay, length, interval, etc.).

2.2. Animals

Seventy male CD1 mice between 6 and 10 weeks of age were obtained from Harlan Laboratory (Frederick, MD, USA) for the human video correlation, effect of water spray and pain studies. Forty two male C57 mice, 6–24 weeks of age, were obtained from Harlan Laboratory (Frederick, MD, USA) for use in the restraint stress study. The 12 BTBR T+tf/J (BTBR) and 12 C57Bl/6J (C57) controls were obtained from Jackson Laboratory (Bar Harbor, ME, USA). All animals were group housed under a 12:12 h light-dark schedule (lights on at 06:00 h) with constant access to standard rodent chow and room temperature tap water. Animals were housed for two weeks of facility habituation before testing commenced. All procedures were performed according to protocols approved by Behavioral Instruments and Albert Einstein College of Medicine Laboratory Animal Services in accordance with the National Institute of Health (NIH) Guide for the Care and Use of Laboratory Animals.

2.3. Drugs

Naproxen sodium and λ -Carrageenan were purchased from Sigma Chemical Company (St. Louis, MO). Naproxen sodium was dissolved in isotonic saline in a concentration of 3 mg/ml

and administered in a volume of 10 ml/kg intraperitoneally. λ -Carrageenan was dissolved in water with the aid of sonication and mild heat at a 1% (w/v) concentration and administered into the hind paws (i.pl.).

2.4. Procedure

2.4.1. Correlation with human scores

To validate the measures of our system we compared our measures to human scoring of video recordings which is currently considered to be the "gold standard" method for detailed analysis of unconditioned behavior. To maximize confidence in this effort we employed a concurrent validation design where we had our system and human scorers evaluate the same mouse sessions. Video records for human scoring were obtained by placing an additional camera outside the chamber for a lateral view of the mice while the system collected behavioral data. The door to the arena was left open to obtain an unobstructed view of the mice during the sessions. A 12 cm diameter clear plastic cylinder was placed in the middle of the arena floor to restrict the mice from engaging in locomotor behavior. Each mouse was sprayed with five sprays of room temperature tap water from a standard spray bottle and placed in the center of the cylinder to enhance grooming. The Behavioral Spectrometer analyzed and recorded behaviors for each 10 min session. Video from the side-view camera was analyzed separately by two experienced human scorers. Each human scorer was familiarized with the ethogram (Table 1) and instructed to attempt to score every second of the session. In practice, each human was only able to confidently score 56% of the seconds. For comparisons with the computer, which scored all seconds, only seconds scored by at least one human were included. This resulted in about 30% more scores for the computer than for the average of the human scores. Due to this asymmetry, it was decided that a correlation of overall raw session score values was the best comparison method.

2.4.2. Wet mice

In order to further validate our system, we employed the "Known-Groups" method of testing for construct validity (Hattie and Cooksey, 1984). This method involved measuring behavior in two groups of animals that are expected to be different based on prior knowledge. For this experiment we placed wet mice in the chambers with the expectation that when compared to dry mice these mice would display increased grooming (Kalouff and Tuohimaa, 2004). Wet mice were sprayed three times with room temperature tap water from a standard spray bottle and placed in the Spectrometer arena. Control mice were not sprayed. Behavioral data were collected by the Behavioral Spectrometer for 20 min.

2.4.3. Reproducibility of results

To investigate the stability of the behavioral measures extracted by our system, six control mice were run the next day according to the same procedure to produce the within-subject correlation data. To examine reproducibility of the measures data obtained from two different cohorts of mice were compared to produce the between-subject correlation data.

2.4.4. Restraint stressed

We sought to further support our construct validity by looking at the effects of restraint stress. After being removed from restraint, mice engage in high levels of self grooming (Zhang et al., 2011; Dunn and Swiergiel, 1999). Stress was produced by placing mice in a ventilated cylinder, 4 cm in diameter and 8 cm long, for 2 h prior to testing. Control animals were not restrained and remained in their home cages prior to testing. Mice were then placed in the arena and behavioral data was collected for 20 min.

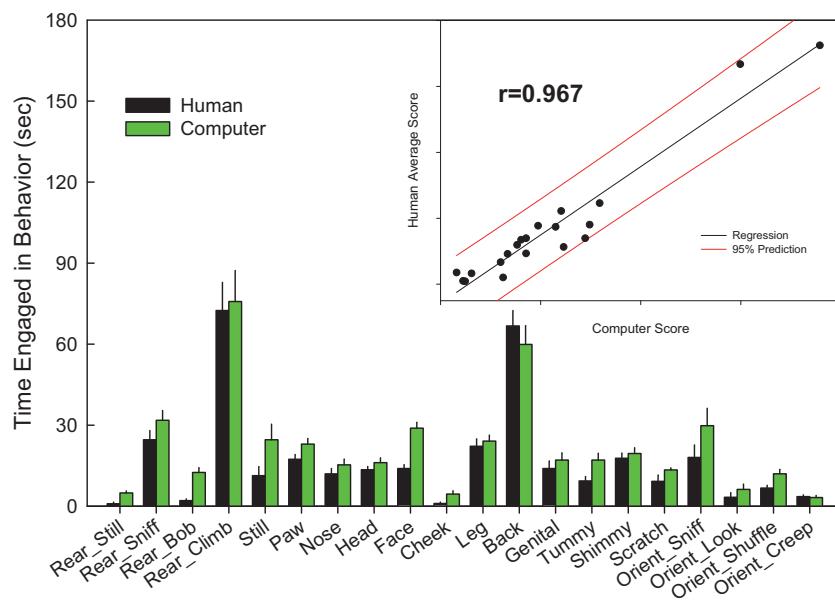


Fig. 2. Analysis of human versus computer scoring shows excellent correspondence. Human scores represent the average score of two observers for each mouse expressed as a mean (+S.E.M.) of all mice over 10 min. Computer score is the score for the same mice expressed as the mean (+S.E.M.). Inset graph shows the relationship between the human average score and the computer score across all behaviors measured. R value computed from Pearson's Product Moment Correlation test. $N=10$ CD1 mice for frequency graph, $N=20$ behaviors for inset graph.

2.4.5. BTBR

To further establish construct validity, as well as show the range of implementation of our system, we measured the behavior of BTBR mice. These mice have recently been characterized as a model for autism and have been reported to show increased levels of grooming (Silverman et al., 2010). We examined their behavior by placing them in a 20 cm by 15 cm clear plastic cage with 1 cm of woodchip bedding on the floor in the center of the arena. The clear plastic did not obstruct the infrared photobeams. The dimensions of the cage were chosen to limit the locomotion of the animal but not restrain it. Consistent with the parameters laid out in the earlier study of BTBR mouse grooming (Silverman et al., 2010) the BTBR or C57 control mice were placed in the cage and allowed to acclimate for 5 min. After acclimation, behavioral data was collected for 10 min.

2.4.6. Inflammatory pain

We explored the implementation of our system by measuring the effect of inflammatory pain. This experiment served to exemplify the potentially broad applications of our system. Mice were injected with 20 μ l of 1% λ -Carrageenan into the plantar surface of each rear footpad (i.pl.), as previously described (da Silva et al., 2012; Khan et al., 2013; Shin et al., 2010). Control mice were injected with saline and analgesic treated mice were injected with 30 mg/kg naproxen sodium in a volume of 10 ml/kg intraperitoneally (i.p.) immediately prior to λ -Carrageenan. Untreated mice were injected i.p. with saline. After injections, mice were placed back in their home cage for 3 h prior to testing. For testing, mice were placed, one at a time, in the arena and behavior was recorded for 20 min.

2.5. Data analysis

To correlate human and computer scored data, the two human scorers' average scores for each category was computed and compared to computer score values. The strength of the association between scores was quantified using the Pearson Product Moment Correlation test. Using the same test, the correlation between the same subjects run on two consecutive days and the correlation

between two separate cohorts of mice was computed. Analyses of all ethogram variables for the different treatments (i.e., wet, stressed and BTBR) were done using unpaired t -tests for each variable (Pearson et al., 2011). Pain data were analyzed using a one-way ANOVA with the post hoc Multiple Comparisons versus Control Group Holm–Sidak method using saline i.pl plus saline i.p. as the control group. A summary depiction of all the experiments (Fig. 7) was made by taking the difference between the treatment and the control for each ethogram variable and expressing the value as a color coded multiple of the standard deviation of the control for that measure.

3. Results

All sessions were successfully completed without any occurrence of technical malfunction or animal escapes, thereby ensuring the integrity of the system.

3.1. Validation using human scoring

The correlation between human and computer-scored mouse behavior was strong and highly significant ($r=0.967$, $p<0.001$) (Fig. 2). The Behavioral Spectrometer was able to accurately predict how a human scored mouse behavior.

3.2. Reproducibility of automated scoring

When the same mice were run twice, the system scored them similarly (Fig. 3A). Computer scores of the same mice on two consecutive days revealed a strong and highly significant correlation ($r=0.97$, $p<0.001$). Interestingly, the point that fell farthest from the line (89,129) was the measure Still. This discrepancy in values (i.e., value was larger the second day) can probably be explained by a habituation effect on the second day (for review see Leussis and Bolivar, 2006). When data obtained from separate cohorts of animals were compared (Fig. 3B) a strong and highly significant correlation ($r=0.98$, $p<0.001$) was observed.

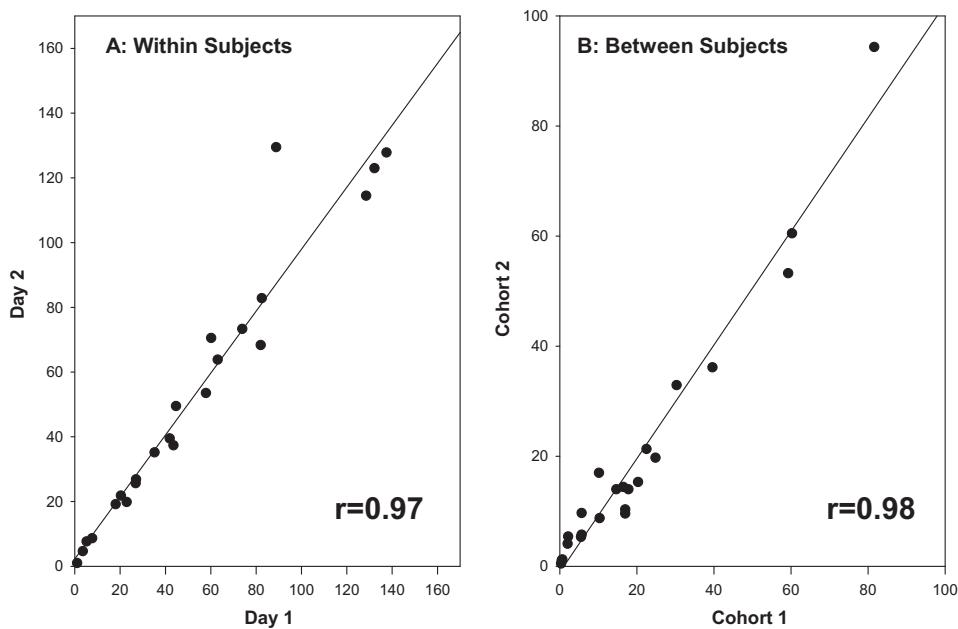


Fig. 3. Automated measures are stable within mice and consistent between groups. (A) Behavioral Spectrometer data of the 23 categories is plotted for the same mice run on two consecutive days for 20 min. Each point represents the mean score of a behavior for the group plotted against the score the following day. $N=6$ CD1 mice for each of 23 behaviors. (B) Data obtained from two different cohorts of mice are plotted against each other. $N=6$ (cohort 1) and 12 (cohort 2) for each of 23 behaviors over 8 min. R value computed from Pearson's Product Moment Correlation test.

3.3. Validation using wet mice

Wet mice displayed elevations in measures of grooming nose, head, face, leg, back and tummy as well as scratching (Fig. 4, $p < 0.05$). Our system measured significantly less walking as well as orienting in the sniffing and creeping categories ($p < 0.05$) in mice sprayed with water. As it has been previously reported that spraying mice with water leads to an increase in grooming behavior (Kalueff and Tuohimaa, 2004), this finding further validates our system's measurements of behavior. While the decreases in orienting and walking were not anticipated, they are not surprising since

the mice must decrease certain types of behavior to account for the increase in grooming time.

3.4. Validation using stressed mice

The Behavioral Spectrometer detected a dramatic effect of restraint stress on behavior (Fig. 5). There were large increases in grooming of all body parts (i.e., paw, face, head, cheek, leg, back, and genitals) ($p < 0.05$), accompanied by a moderate increase in scratching ($p < 0.05$). Conversely, stress produced dramatic decreases in locomotor behavior (i.e., walk, trot and run) and

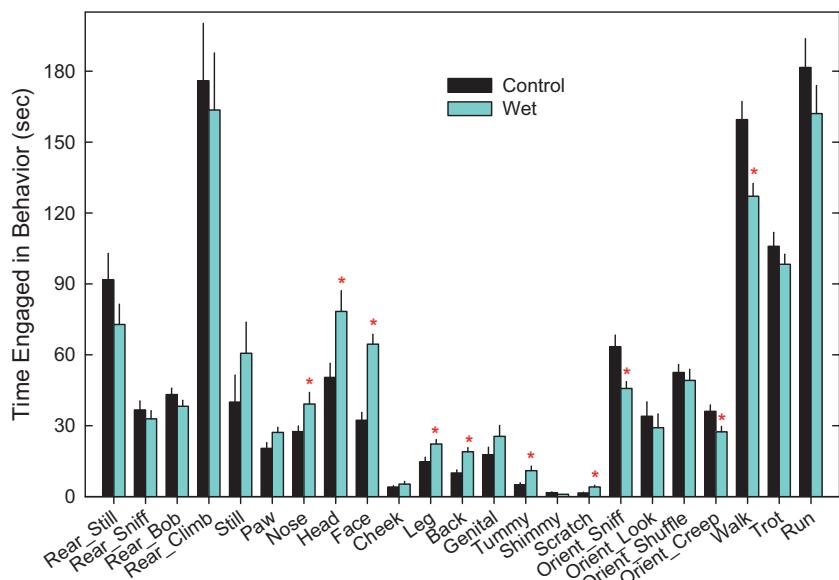


Fig. 4. Wet mice displayed more grooming. Data obtained from the Behavioral Spectrometer are shown as number of seconds of behavior scored by behavioral category for control (dry) and water sprayed (wet) mice (mean +S.E.M.) over 20 min. Wet mice showed more grooming of the nose, head, face, leg, back and tummy categories. Statistical comparisons were made between wet and dry mice for each category. $N=12$ CD1 mice per treatment, $*p < 0.05$.

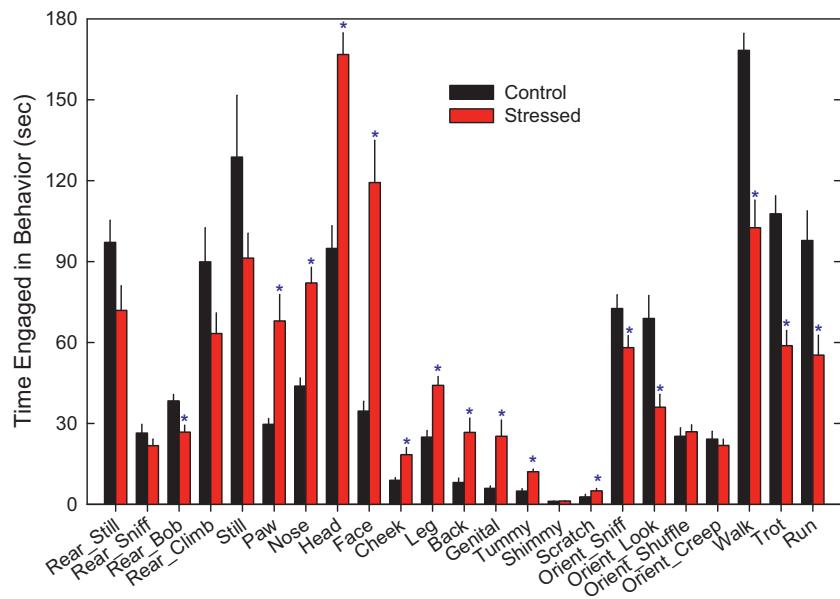


Fig. 5. Stressed mice displayed much more grooming and less locomotor behavior. Data obtained from the Behavioral Spectrometer are shown as number of seconds of behavior scored by behavioral category for control (unrestrained) and stressed (2 h of prior restraint) mice (mean +S.E.M.) over 20 min. Stressed mice showed more grooming in all categories and less locomotor activity. Small decreases were also evident in the Rear_Bob, Orient_Sniff and Orient_Look categories. Statistical comparisons were made between control and restrained mice for each category. $N = 12$ C57 mice per treatment, * $p < 0.05$.

a mild decrease in the orienting behaviors of sniff and look ($p < 0.05$). Again, this was not surprising considering the increase in grooming had to come at the expense of other behaviors. The observed increase in grooming was consistent with previous reports (Zhang et al., 2011; Dunn and Swiergiel, 1999) and served to further support the validity of the measurements made by our system.

3.5. BTBR mice

BTBR mice showed significant increases in grooming of the paw, nose, head, face, leg, back, genitals and tummy (Fig. 6, $p < 0.05$). There were also increases in scratching and orient looking ($p < 0.05$).

However, the most dramatic effects observed were decreases in all forms of rearing, including still, sniff, bob and climb ($p < 0.05$). The increase in grooming was consistent with several previous reports (Pearson et al., 2011; Silverman et al., 2012) and, once again, suggests that the Behavioral Spectrometer is making valid measures of behavior. While no previous reports have suggested that rearing behavior is impaired in these animals, there have been no attempts to find which behaviors are decreased to make way for the increased grooming. Additionally, these mice have been reported to exhibit less interest in the outside world (e.g., other mice) than in self-directed activity (e.g., self-grooming) (McFarlane et al., 2008); this finding of decreased exploratory rearing is consistent with that description.

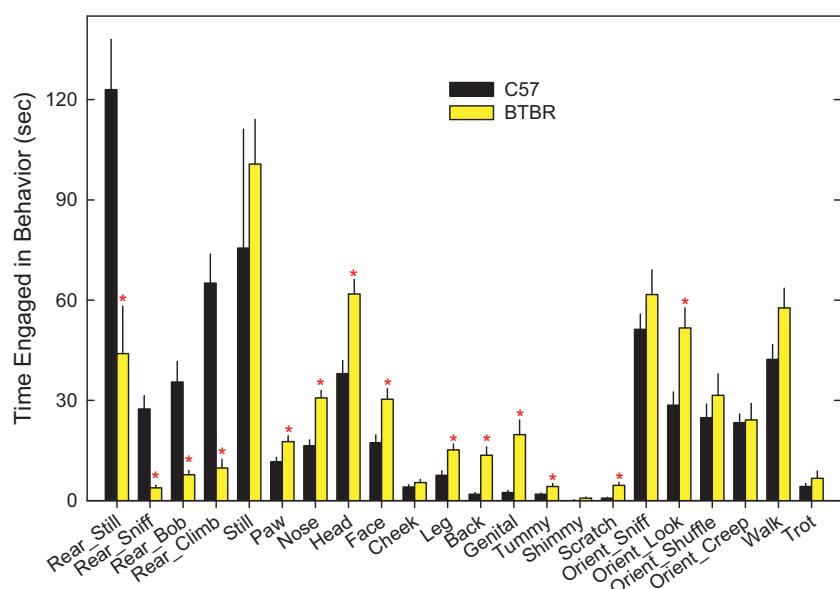


Fig. 6. BTBR mice displayed more grooming and less rearing. Data obtained from the Behavioral Spectrometer are shown as number of seconds of behavior scored by behavioral category for control (C57) and BTBR mice (mean +S.E.M.) over 10 min. BTBR mice showed more grooming in all categories except cheek. BTBR mice also displayed dramatically less rearing behavior. Statistical comparisons were made between control and BTBR mice for each category. $N = 12$ mice per treatment, * $p < 0.05$.

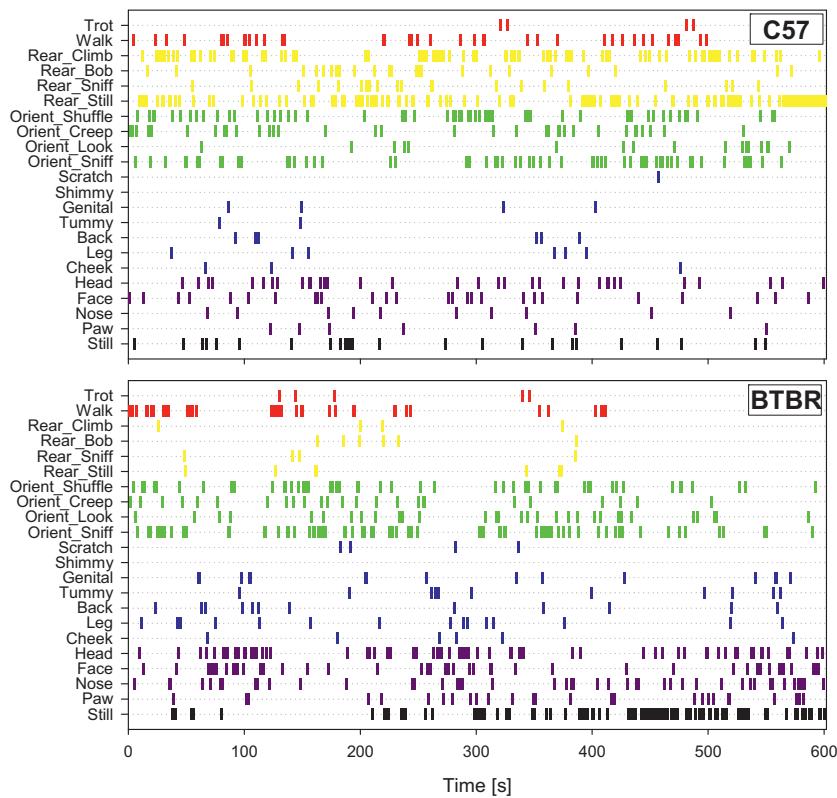


Fig. 7. Representative ethograms of the BTBR mouse. These two panels show the second-by-second categorization of behavior produced by the Behavioral Spectrometer in real time. Upper panel shows an ethogram of a control (C57) mouse and the lower panel shows a BTBR mouse. Behavioral (y-axis) events are marked by time (x-axis) in a color coded fashion. Locomotor activity is marked in red, rearing activity in yellow, orienting in green, vigorous grooming in blue, fine grooming in purple and still in black. Comparing ethograms shows less rearing and more grooming in the BTBR (lower panel) compared to the control mouse (upper panel). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7 shows the ethogram chart for representative samples from the BTBR and C57 groups. The behavioral differences are readily apparent through inspection of the color distribution on the chart. The control C57 mouse shows much more rearing (yellow) while the BTBR mouse shows much more grooming (purple and blue). This chart is created in real time for each mouse during the sessions and serves as a useful means to assess the overall behavior of the subject.

3.6. Effects of pain

Mice injected with λ -Carrageenan into their rear paws (i.e., in pain) showed a dramatic increase in the amount of time they spent stationary (still in **Fig. 8**, $p < 0.05$). This increase came at the expense of almost all other categories of behavior. Except for a few low frequency behaviors (cheek, shimmy and scratch) and orient look, all other behaviors were significantly reduced ($p < 0.05$). This large increase in immobile behavior is consistent with most animal pain studies (Zhu et al., 2012). Interestingly, when we compared the saline injected group to previous control groups of the same strain (**Fig. 4**) we observed a significant increase in still and head grooming scores. The increase in still behavior is consistent with an interpretation that saline injection produced a small degree of pain (i.e., still moved in the same direction that λ -Carrageenan injection produced) and the increase in head grooming is consistent with the injection procedure being stressful (i.e., head grooming moved in the same direction that restraint stress produced, **Fig. 5**). When λ -Carrageenan-injected animals were administered the powerful analgesic naproxen, the blockade of pain-induced behavior was absolute. Not only was there no increase in still behavior, but all other categories of behavior were evident at normal or slightly

above normal (walk and trot) levels. By examining the effect of an analgesic, we substantiated that our measurements were, in fact, indicative of pain. These findings suggest that our system is sensitive to behavioral changes induced by pain and sensitive to the effects of analgesics.

3.7. Behavioral signatures

Fig. 9 shows the behavioral data from all experiments expressed as a change from controls and normalized by the standard deviation of the control group. By graphing the results this way it was possible to compare the degree of the behavioral effect between treatments as well as effects on individual behaviors within treatments. We used this information to select head grooming in the BTBR group as a moderate effect on a behavior within a moderate treatment effect. Results of a Power analysis (Power = 0.80, Alpha = 0.05) on these data revealed that the minimum sample size for this degree of effect was 7 mice per group. This depiction also highlights the potential of the system to create a unique signature profile for various experimental manipulations. Here, five different behavioral patterns are shown for five different treatments. These patterns can serve as reference baselines to compare effects of treatments (as was done with pain in the current report). They can also be used to develop a library of “signatures” that can serve as references for profiles of novel treatments.

4. Discussion

The validation and implementation of the Behavioral Spectrometer, a novel instrument that measures mouse behavior, has been summarized in this report. This video- and vibration-based system

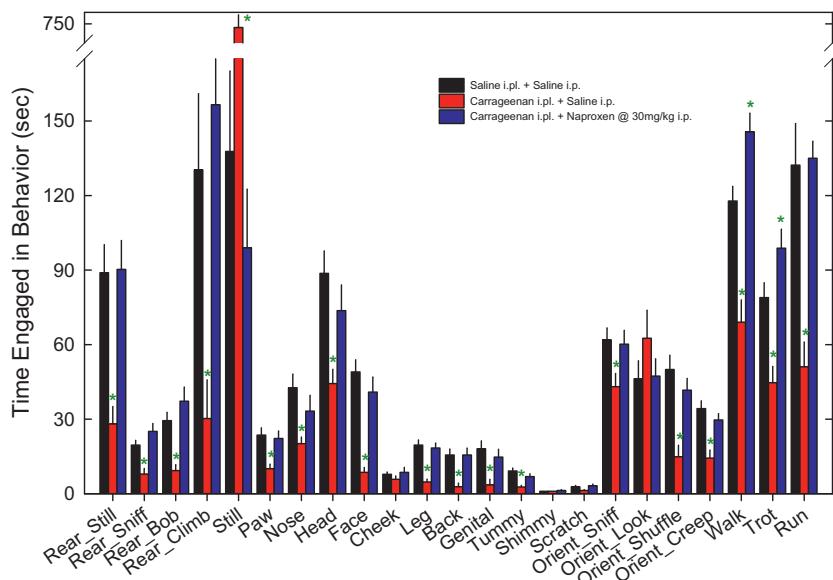


Fig. 8. Inflammatory pain produces a profound and reversible effect on behavior. Data obtained from the Behavioral Spectrometer are shown as number of seconds of behavior scored by behavioral category for control (saline i.pl. + saline i.p.), inflamed (Carrageenan i.p.l. + Saline i.p.) and treated (λ -Carrageenan i.p.l. + naproxen i.p.) mice (mean \pm S.E.M.) over 20 min. Mice in pain showed a dramatic inhibition of all behavior except sitting still. When naproxen was co-administered with the irritant, behavior was returned to normal levels (with small increases in walk and trot). Statistical comparisons were made between groups within each behavioral category with control mice serving as the control comparison. $N=12$ CD1 mice per treatment, $*p < 0.05$.

exhibited excellent agreement with trained human observers in the identification of 23 different behaviors emitted by mice in an open field. Furthermore, our system was able to identify numerous statistically significant behavioral effects in three mouse models of CNS disorders.

Validation of the Behavioral Spectrometer was accomplished through several rigorous tests. We employed the “gold standard” of human scoring to produce categorization scores of behavior and then compared those to our system’s automatic scores as a test of face validity. We then employed three tests of construct validity using a “Known Groups” method. These experiments consisted of testing different treatments of animals that had been reported in the literature to show increased grooming behavior: wet mice, stressed mice and BTBR mice. Our system found robust and significant differences between the groups in the three models. In all cases the changes observed were consistent with previous reports, suggesting that the categorizations produced were accurate and valid.

Interestingly, our human-scored correlation experiment showed that the humans, despite attempting to score every second, were unable to come up with scores for 44% of the seconds. This is one of the difficulties of human scoring and highlights a unique feature of our box: every second is scored. By producing a score for every second our system ensures that no behavior will be missed, whether it is anticipated or not. The algorithm assigns even unique behaviors to closely approximated categories. This “forced fit” method ensures that all behavioral abnormalities show up in the data. For example, we have observed (*data not shown*) that clonic seizures are commonly classified as “shimmy”. Once a signal is identified in the categories, it can be observed directly in the video record for further subjective description, if desired.

The BTBR mouse has been reported to have deficits in social behavior and show increased levels of grooming reminiscent of stereotypic self-stimulatory behaviors seen in children with autism (Pobbe et al., 2010). Due to the critical need for tools to facilitate autism research, we decided to test the BTBR mouse with our instrument. Consistent with previous reports, we found increased

levels of self-grooming. Not only were the grooming variables increased, but the amount of increase was similar to previous reports (Pearson et al., 2011). Additionally, we found that the BTBR mice show a profound rearing deficit. To our knowledge, ours is the first report of this behavioral deficit. A lack of exploratory rearing is entirely consistent with the autism-like phenotype of this mouse (i.e., a less outwardly focused animal). This finding highlights the utility of the full spectrum behavioral profiling provided by our instrument.

We wanted to explore the potential of our system to be used in a diversity of mouse models of CNS disorders. Since many models are based on interventions (pharmacological or surgical), we chose to look at one of the more popular models of inflammatory pain based on injection of a chemical irritant. Inflammatory pain caused the mice to profoundly limit their behavior to mostly sitting stationary. Typically, this type of response would be problematic for behavioral testing since there would be little to no spontaneous behavior to analyze. While this was true in our case, we had a plethora of affected behavioral measures which to compare against an analgesic response. When we administered the analgesic naproxen we saw a dramatic reversal of every behavior. Although somewhat surprising, the complete reversal was not inconsistent with reports in the literature of the high analgesic efficacy of NSAIDs in inflammatory pain (Jain et al., 2002; Berk et al., 2009; Sahin et al., 2001). These findings are especially exciting because the rich background of data gathered can be used as contrast for seemingly dull alterations in behavior. This allows for a very informative assessment of possible therapeutics. This feature in our system is ideal for exploring models whose return to normal behavior is the focus of the research. For example, in schizophrenia research, treatment with psychotomimetics agents like PCP is often used to model the human condition. However, PCP produces massive increases in locomotor behavior which makes behavioral assessment problematic. A lot of research has focused on merely measuring the degree to which an agent can reduce this locomotor activity; this too is problematic as any agent that sedates the animal will look therapeutic. With the detailed background of what a normal mouse looks like, it is

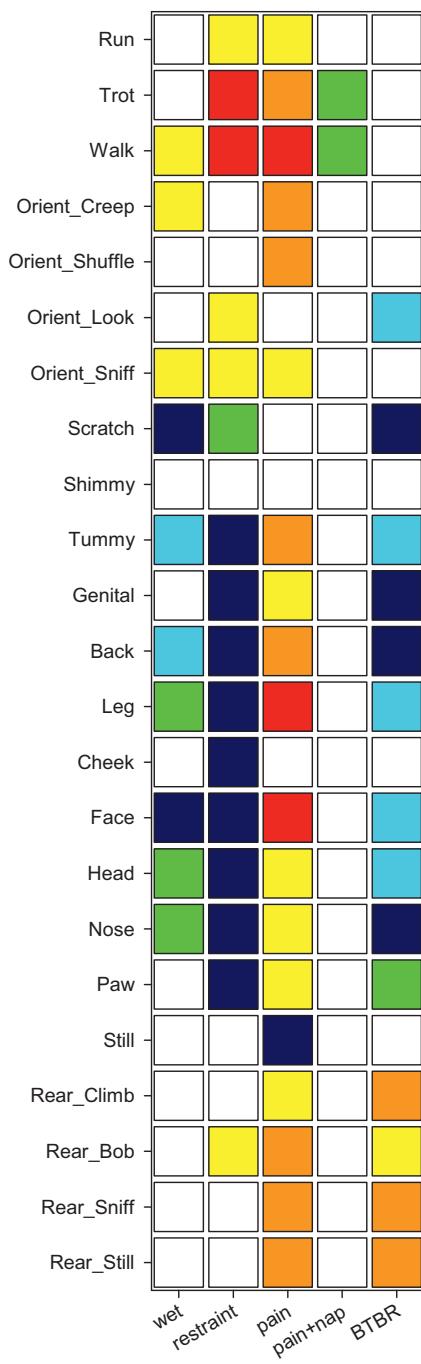


Fig. 9. Behavioral spectrograms. The effect of each treatment was computed as the change in behavior relative to control expressed as a multiple of the standard deviation (SD) of the mean of the control values. The depiction is color coded to correspond to degree of change in the occurrence of the behavior, with dark blue representing an increase of more than 2 SDs, light blue between 1.5 and 2 SDs, green between 1 and 1.5 SDs and red representing as decrease of more than 2 SDs, orange between 1.5 and 2 SDs and yellow between 1 and 1.5 SDs. Results that were not statistically significantly different from control values or where the difference was less than 1 SD are plotted in white. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

eminently possible to distinguish between mere sedation (or stimulation in our pain experiment) and true return to normal behavior, as our pain experiment nicely shows.

Because our system is highly standardized it allows for the benefit of comparisons across experiments to be made. With each experiment users would add to a library of patterns which describe

drug effects and/or diseases. This library would then provide a reference from which to compare the results of all past and future experiments. Once the library is sufficiently large, results from our system would produce a result that tells the user what class of drugs or disease models most closely resemble their treatment. And, in fact, early adopters of our system have begun to build this library through the publication of ethograms associated with a model of Parkinson's and a model of abnormal neural development (Paumier et al., 2013; DeBoer et al., under review, respectively). By testing for multiple indications in a single session through library reference comparisons, the long standing need for a truly high-throughput assay would be satisfied.

Overall, these results show that the Behavioral Spectrometer makes valid and detailed measurements of behavior which are sensitive to numerous experimental manipulations. Importantly, the assessments made by our system are truly high-throughput in that they are automated, scalable and obtained in a single short session. We assert that this instrument represents the most detailed, high-throughput behavioral recognition system commercially available. We hope that our instrument can be helpful at relieving the bottleneck which has heretofore plagued behavioral evaluation in CNS research.

Conflicts of interest

JB owns Behavioral Instruments.
CG owns BiObserve.

References

- Amodeo DA, Jones JH, Sweeney JA, Ragozzino ME. Differences in BTBR T+tf/J and C57BL/6J mice on probabilistic reversal learning and stereotyped behaviors. *Behav Brain Res* 2012;227(February (1)):64–72.
- Benjamini Y, Lipkind D, Horev G, Fonio E, Kafkafi N, Golani I. Ten ways to improve the quality of descriptions of whole-animal movement. *Neurosci Biobehav Rev* 2010;34(July (8)):1351–65.
- Berk B, Erol DD, Kupeli E, Yesilada E. Design and synthesis of some (S)-2-(6-methoxynaphthalen-2-yl)-N-substituted ethyl propanamide derivatives as potent non-ulcerogenic anti-inflammatory and analgesic agents. *Arzneimittelforschung* 2009;59(4):195–201.
- da Silva KA, Manjavachi MN, Paszczuk AF, Pivatto M, Viegas CJr, Bolzani VS, et al. Plant derived alkaloid (–)-cassine induces anti-inflammatory and anti-hyperalgesic effects in both acute and chronic inflammatory and neuropathic pain models. *Neuropharmacology* 2012;62(February (2)):967–77.
- DeBoer EM, Vega TA, Ricardo A, Brodkin J, Akamatsu W, Okano H, et al. Prenatal deletion of the RNA binding protein HuD disrupts postnatal cortical circuit maturation and behavior. *J Neurosci* 2014 [under review].
- Dunn AJ, Swiergiel AH. Behavioral responses to stress are intact in CRF-deficient mice. *Brain Res* 1999;845(October (1)):14–20.
- Hartley JE, Montgomery AM. 8-OH-DPAT inhibits both prandial and waterspray-induced grooming. *J Psychopharmacol* 2008;22(September (7)):746–52.
- Hattie J, Cooksey RW. Procedures for assessing the validities of tests using the 'Known Groups' method. *Appl Psychol Measure* 1984;8(3):295–305.
- Jain NK, Kulkarni SK, Singh A. Modulation of NSAID-induced antinociceptive and anti-inflammatory effects by alpha2-adrenoceptor agonists with gastrophrope effects. *Life Sci* 2002;70(May (24)):2857–69.
- Kafkafi N, Mayo CL, Elmer GL. Mining mouse behavior for patterns predicting psychiatric drug classification. *Psychopharmacology* 2014;231(January (1)):231–42.
- Kafkafi N, Yekutieli D, Elmer GL. A data mining approach to in vivo classification of psychopharmacological drugs. *Neuropsychopharmacology* 2009;34(February (3)):607–23.
- Kalueff AV, Tuohimaa P. Grooming analysis algorithm for neurobehavioural stress research. *Brain Res Brain Res Protoc* 2004;13(August (3)):151–8.
- Khan S, Shehzad O, Chun J, Kim YS. Mechanism underlying anti-hyperalgesic and anti-allodynic properties of anomalin in both acute and chronic inflammatory pain models in mice through inhibition of NF- κ B, MAPKs and CREB signaling cascades. *Eur J Pharmacol* 2013;718(October (1–3)):448–58.
- Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004;3(August (8)):711–5.
- Kyzar E, Gaikwad S, Roth A, Green J, Pham M, Stewart A, et al. Towards high-throughput phenotyping of complex patterned behaviors in rodents: focus on mouse self-grooming and its sequencing. *Behav Brain Res* 2011;225(December (2)):426–31.
- Leussis MP, Bolivar VJ. Habituation in rodents: a review of behavior, neurobiology, and genetics. *Neurosci Biobehav Rev* 2006;30(7):1045–64.

- McFarlane HG, Kusek GK, Yang M, Phoenix JL, Bolivar VJ, Crawley JN. **Autism-like behavioral phenotypes in BTBR T+tf/J mice.** *Genes Brain Behav* 2008;7(March (2)):152–63.
- Mogil JS, Crager SE. **What should we be measuring in behavioral studies of chronic pain in animals?** *Pain* 2004;112(November (1–2)):12–5 [review].
- Paumier KL, Sukoff Rizzo SJ, Berger Z, Chen Y, Gonzales C, Kaftan E, et al. **Behavioral characterization of A53T mice reveals early and late stage deficits related to Parkinson's disease.** *PLoS One* 2013;8(August (8)):e70274.
- Pearson BL, Pobbe RL, Defensor EB, Oasay L, Bolivar VJ, Blanchard DC, et al. **Motor and cognitive stereotypies in the BTBR T+tf/J mouse model of autism.** *Genes Brain Behav* 2011;10(March (2)):228–35.
- Pobbe RL, Pearson BL, Defensor EB, Bolivar VJ, Blanchard DC, Blanchard RJ. **Expression of social behaviors of C57BL/6J versus BTBR inbred mouse strains in the visible burrow system.** *Behav Brain Res* 2010;214(December (2)):443–9.
- Sahin G, Palaska E, Kelicen P, Demirdamar R, Altinok G. **Synthesis of some new 1-acylthiosemicarbazides, 1,3,4-oxadiazoles, 1,3,4-thiadiazoles and 1,2,4-triazole-3-thiones and their anti-inflammatory activities.** *Arzneimittelforschung* 2001;51(6):478–84.
- Shin EM, Zhou HY, Xu GH, Lee SH, Merfort I, Kim YS. **Anti-inflammatory activity of hispidol A 25-methyl ether, a triterpenoid isolated from Poncirus Immaturus Fructus.** *Eur J Pharmacol* 2010;627(February (1–3)):318–24.
- Silverman JL, Smith DG, Rizzo SJ, Karras MN, Turner SM, Tolu SS, et al. **Negative allosteric modulation of the mGluR5 receptor reduces repetitive behaviors and rescues social deficits in mouse models of autism.** *Sci Transl Med* 2012;4(April (131)):131ra51.
- Silverman JL, Tolu SS, Barkan CL, Crawley JN. **Repetitive self-grooming behavior in the BTBR mouse model of autism is blocked by the mGluR5 antagonist MPEP.** *Neuropsychopharmacology* 2010;35(March (4)):976–89.
- Tecott LH, Nestler EJ. **Neurobehavioral assessment in the information age.** *Nat Neurosci* 2004;7(May (5)):462–6. Review.
- Yucel Z, Sara Y, Duygulu P, Onur R, Esen E, Ozguler AB. **Automated discrimination of psychotropic drugs in mice via computer vision-based analysis.** *J Neurosci Methods* 2009;180(June (2)):234–42.
- Zhang J, Wu Z, Zhou L, Li H, Teng H, Dai W, et al. **Deficiency of antinociception and excessive grooming induced by acute immobilization stress in Per1 mutant mice.** *PLoS One* 2011;6(January (1)):e16212.
- Zhu CZ, Mills CD, Hsieh GC, Zhong C, Mikusa J, Lewis LG, et al. **Assessing carrageenan-induced locomotor activity impairment in rats: comparison with evoked endpoint of acute inflammatory pain.** *Eur J Pain* 2012;16(July (6)):816–26.