

## Short communication

## Automated scoring of fear-related behavior using EthoVision software

Jon Pham<sup>a</sup>, Sara M. Cabrera<sup>a,b</sup>, Carles Sanchis-Segura<sup>a,c</sup>, Marcelo A. Wood<sup>a,\*</sup><sup>a</sup> Department of Neurobiology and Behavior, Center for the Neurobiology of Learning and Memory, University of California, Irvine, United States<sup>b</sup> Department of Biological Sciences, Allergan, Irvine, CA, United States<sup>c</sup> Area de Psicobiologia, Universitat Jaume I, Castellon, Spain

## ARTICLE INFO

## Article history:

Received 26 September 2008

Received in revised form 16 December 2008

Accepted 18 December 2008

## Keywords:

Mobility

Fear conditioning

Behavior

Automated scoring

## ABSTRACT

Fear conditioning is a frequently used paradigm for assessing learning and memory in rodents. Traditionally researchers have relied upon scoring of fear-related behavior by human observation, which can be difficult and subjective and thus vary among investigators. The goal of this study was to evaluate the ability of EthoVision tracking software (Noldus Information Technology Inc.) to reliably and accurately score fear-related behavior in mice. Specifically, we were interested in its ability to accurately track mice and score immobility as a fear-related behavior during contextual and cued fear conditioning. Contextual and cued fear conditioning were performed in modified PhenoTyper chambers (Noldus Information Technology Inc.) fitted with grid floors to deliver a scrambled foot shock. Our results demonstrate that we have identified parameters in EthoVision that can accurately track mice and be used for automated scoring of immobility that is nearly identical to scoring by human observation. Together, EthoVision software and the modified PhenoTyper chambers provide an excellent system for the reliable and accurate measurement of fear-related behavior in a high-throughput manner.

© 2008 Elsevier B.V. All rights reserved.

## 1. Introduction

Contextual and Cued fear conditioning are Pavlovian or classical conditioning tasks in which an animal learns the relationship between an aversive event and external stimuli that predict the aversive event. Fear conditioning is widely used to investigate the neurobiology of learning and memory at the molecular, cellular, system, and behavioral level (reviewed in Maren and Quirk, 2004). Our goal was to examine the ability of EthoVision software to accurately track and measure immobility in mice subjected to contextual and cued fear conditioning.

Measuring fear-related behavior by a human observer can be difficult, especially when attempting to analyze the behavior of multiple cages in large experiments. Human observers can also have subtle yet significant differences in their subjective measurements of fear-related behavior. This variability can potentially mask important performance differences in fear conditioning. In addition, human observer scoring is highly time-consuming, making it unsuitable for scientific endeavors that require high-throughput analysis. Finally, although human scoring is performed by blind observers, there are cases in which different groups can be easily

identified (i.e. when comparing strains with different coat colors, drug treatments causing obvious changes in behavior, lesions, etc.), which compromise the objectivity of human scoring. Because of the disadvantages of using human observers to score behavior, the popularity and number of automated scoring software packages has increased rapidly. Thus, several studies have focused on utilizing automated scoring software to measure fear-related behavior in this paradigm. In this study, we examined whether a hardware and software package (PhenoTyper chambers and EthoVision software; Noldus Information Technology Inc.), that are commercially available and specifically designed to be operated together, could be used to measure fear-related behavior. We found EthoVision was capable, when using a proper configuration, of excellent tracking quality, matched human scoring of fear-related behavior, and that in combination with modified PhenoTyper chambers can be utilized for reliable and precise automated scoring of immobility in mice subject to contextual or cued fear conditioning. The exact values for each relevant identified parameter in measuring fear conditioning are reported here.

## 2. Materials and methods

## 2.1. Animals

Male C57BL/6J mice bred in our animal facility from mice originally obtained from Jackson Laboratories were used in most experiments. Mice were 8–16 weeks old and had free access to food

\* Corresponding author at: University of California, Irvine, Department of Neurobiology and Behavior, Center for the Neurobiology of Learning and Memory, 106 Bonney Research Lab, Irvine, CA 92697-3800, United States. Tel.: +1 949 824 2259; fax: +1 949 824 6633.

E-mail address: [mwood@uci.edu](mailto:mwood@uci.edu) (M.A. Wood).

and water in their homecages. Lights were maintained on a 12:12-h light/dark cycle, with all behavioral testing carried out during the light portion of the cycle. All experiments were conducted according to National Institutes of Health guidelines for animal care and use and were approved by the Institutional Animal Care and Use Committee of the University of California, Irvine.

## 2.2. Experimental apparatus and software

Fear conditioning experiments were performed using a set of four modified Noldus PhenoTyper (Model 3000) chambers (Leesburg, VA) with shock floors as previously described (Lattal et al., 2007). The PhenoTyper Model 3000 chamber has a 30 cm × 30 cm floor and is 40 cm in height. The PhenoTyper chamber is equipped with a top unit including a matrix of infrared LED lights and an infrared CCD camera, with a high-pass filter blocking visible light. The floor of the cages was modified to include a stainless steel grid (inter-bar separation: 0.9 cm) connected to an electric shock generator (Shock Scrambler ENV-414S; Med Associates, St. Albans; VT). Automated tracking and shock delivery control were performed using EthoVision 3.1 software (EthoVision 3.1; Noldus Information Technology, Leesburg, VA; Noldus et al., 2001; Spink et al., 2001; see <http://www.noldus.com/ethovision> for additional references and information on EthoVision software). EthoVision software works as a contrast engine, recognizing objects by their differences in shade. The program takes samples, or screenshots, of video data and calculates the area of the subject depending on the difference of contrast between the subject and the defined arena, as designated by each pixel's gray value. The detection method itself can be configured as a subtraction, or gray scaling method. A subtraction method uses a reference image of the experiment background, without the subject, to calculate where the subject is by subtracting real-time video data from the reference image. This method works well for experiments, in which the experimental background does not change often. The gray scaling method uses light and dark thresholds to determine the subject within the designated experiment arenas. For the purpose of our experiments, we chose to work with a gray scaling method, as it provided the most flexibility.

Every experiment is also recorded as MPEG-4 format video files using video capture software (Mediacruise, Canopus, San Jose, CA). This allowed us to run MPEG videos through EthoVision 3.1 software to optimize parameters used to score immobility. Parameters include: sample rate (described in detail in Section 3); processing and detection method (grey scaling); object intensity (object is always darker than background); immobility threshold (5%); strong mobile threshold (95%); mobility averaging interval (20); erosion and dilation filters (erosion by 1 pixel, then dilation by 1 pixel). The mobility averaging interval averages the current sample taken with the previous 19 (when set to 20) in order to adjust for false sample detection issues. Freezing behavior was also assessed by hand (Fanselow, 1980). Freezing was defined by the absence of detected movement and continuously monitored across the whole session length by an experienced single observer blind to the experimental conditions.

## 2.3. Fear conditioning protocol

Mice were handled for three consecutive days for 1 min each day. Both, contextual and cued fear conditioning protocols consisted of one session of conditioning followed, 24 h later, by a retention test (Vecsey et al., 2007; Wood et al., 2005, 2006). For contextual fear, during the conditioning session C57BL/6J mice were placed into the conditioning chamber and received three 2 s 0.75 mA scrambled footshocks 2.5 min, 3.5 min, and 4.5 min after placement into the chamber. Mice were removed from the chamber after 5 min. During the testing session, mice received one 5 min exposure to the

same conditioned context in the absence of shock. For cued fear conditioning, C57BL/6J mice handled in a similar fashion to those used in contextual fear conditioning were placed into a chamber and an audible cue (a built in buzzer housed in the PhenoTyper chamber) was presented three times during a 5-min conditioning trial. The buzzer was turned on for 30 s at 2 min, 3 min, and 4 min after placement into the chamber. The CS-only group received only those three audible cues. The CS/US group was subject to the same three audible cues, but they each co-terminated with a 2-s 0.75 mA footshock (thus three CS/US pairings). Mice were removed from the chamber 5 min after initial placement into the chamber. A 24-h retention test was performed the next day. Mice were introduced into a novel context (the same conditioning chamber modified to have a smooth flat floor, altered dimensions, and a novel odorant) for a 5-min test trial. Fear testing was performed for two separate durations during the test; the first 2 min without the audible cue (Pre-CS) and from 2 min to 5 min when the audible cue was present (CS). For both contextual and cued fear conditioning, the EthoVision software was used to control shock periods/amplitude, cue presentation (for cued fear conditioning), and experimental parameters such as time trial and testing zones via user-defined variables.

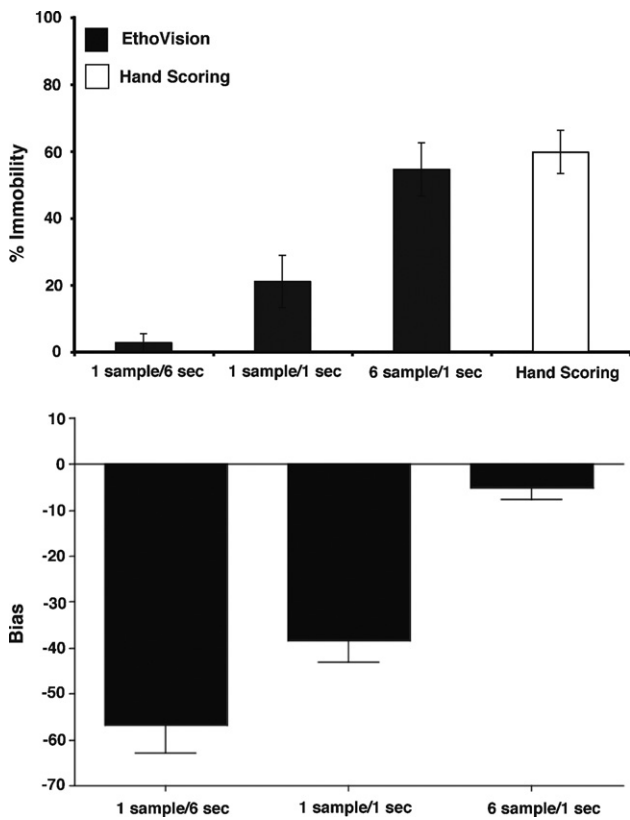
## 2.4. Data analysis/interpretation

Data were analyzed using SPSS (Version 13.0 for MAC OS X). Most of the results were analyzed by means of ANOVA, including a repeated measures factor when necessary. Follow up between groups multiple comparisons were conducted using Student–Newman–Keuls or, in those cases involving a set of comparisons against a control group, Dunnett's post hoc test (Seaman et al., 1991).

Covariance and agreement between automated and trained-observer scores of freezing/conditioned fear were assessed by means of Pearson's correlation index as well as the intraclass correlation (Shrout and Fleiss, 1979) and bias coefficients (as derived from Bland–Altman plots; Bland and Altman, 1986), respectively. The intraclass correlation coefficient is an omega-squared like statistic used as a measure of inter-judges reliability because it estimates the proportion of variance in the data that is due to group differences. This is because in contrast to Pearson's Correlation, where the variables of interest are modeled as two distinct traits, with the mean and variance of each being estimated separately (and, therefore, both variables do not share neither their metric or variance), in the intraclass correlation the trait's mean and variance are derived from pooled estimates across all members of all groups. Therefore, while the Pearson gives the proportion of shared variance between the two members of a pair without respect to group membership, the intraclass correlation gives the proportion of variance accountable to group differences, rather than differences in the judges, judge × subject interaction, or error (McGraw and Wong, 1996). On the other hand, the Bland–Altman plots is the standard statistical approach used in clinical research when comparing two different methods of measuring the same variable (McCluskey and Gaaliq, 2007) because it provides both an average discrepancy index between both methods (bias) as well as an individual treatment of the sample studied

## 3. Results

In the first experiment, the objective was to establish parameters within EthoVision to accurately score fear behavior in C57BL/6J mice subjected to contextual fear conditioning. To score fear-related behavior, we used the mobility detection function on EthoVision, which has a lower threshold to separate immobility from mobility, and an upper threshold to separate mobility from strong mobil-



**Fig. 1.** Establishing automated scoring of immobility using EthoVision. (A) Mice were subject to contextual fear conditioning and then received a 24-h retention test in the conditioned context. The retention test lasted for 5 min during which EthoVision was used to score immobility and compared to scoring by human observation. MPEG video was generated during the retention test in order to run multiple times through EthoVision to optimize parameters to match performance scored by human observation. Sample rate configurations included 1 sample/6 s, 1 sample/1 s, and 6 samples/s. (B) 6 samples/s most closely matched freezing scored by human observation (Hand Scoring) and reduced measurement bias as calculated by Bland–Altman plots. Values are mean  $\pm$  S.E.M.

ity. The lower threshold was set to 5.0% to score immobility, which means there can be no more than 5.0% change in the pixels of a detected object between current sample and previous sample (see section 12.3.3 of EthoVision 3.1 Reference Manual). Another important factor in adjusting the automated scoring to match our hand scoring was to set the averaging interval to 20.0. The averaging interval will smooth the mobility parameter, which means potential problems with tracking will be smoothed out.

Using these parameters (more fully described in the Section 2), we studied the effect of different sample rates on the score accuracy provided by EthoVision software as compared to those provided by an experienced human observer assessing the same group of mice (Fig. 1). Increasing sample rate resulted in higher freezing scores. This effect was confirmed by means of an ANOVA, which yielded a significant difference between groups ( $F_{(3,20)} = 17.0$ ,  $p < 0.0001$ ). As determined by Dunnett's post hoc comparisons, low to intermediate sample rates (1 sample/6 s or 1 sample/s) resulted in a poor identification of freezing, that is, significantly lower than the scores provided by a human observer ( $p < 0.01$ ). However, when sample rate was increased up to 6 samples/s, there were no differences between automated and hand scoring (Fig. 1A).

These results are in agreement with the increasing covariance observed between hand-scoring and EthoVision software provided scores as a function of sampling rate ( $r = 0.49$ ,  $p = .41$ ;  $r = 0.81$ ,  $p = 0.05$ ;  $r = 0.96$ ,  $p = 0.003$  for 1 sample/6 s, 1 sample/s, 6 samples/s, respectively). However, covariation does not necessarily imply

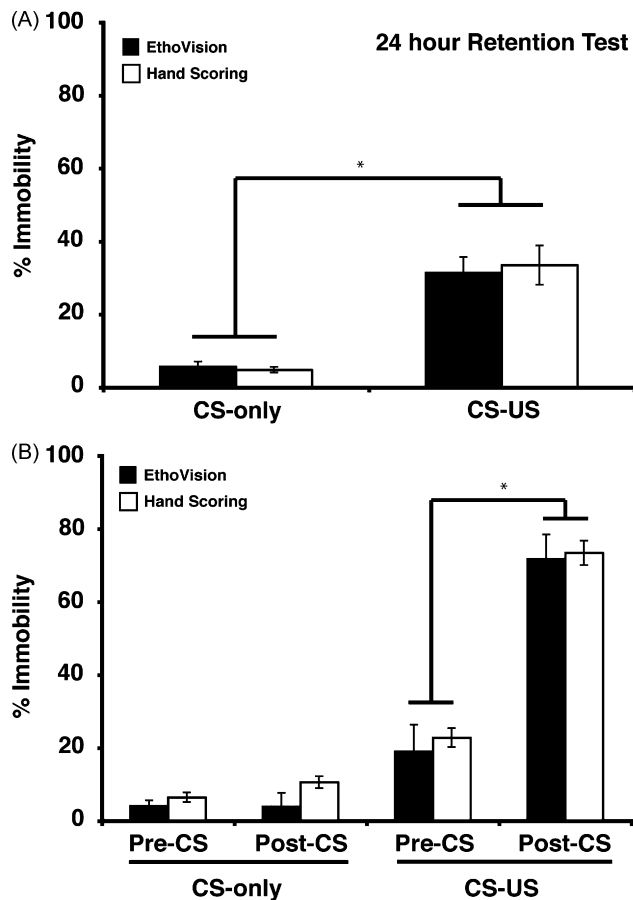
agreement and comparing two methods of measuring the same variable using only this kind of covariation indexes can be misleading (i.e. two methods can show a clear concordance in the relative allocation of the subjects of a sample but little agreement in the overall estimate of the measured variable). Therefore, we calculated the intraclass correlation between the freezing EthoVision software scores obtained at each sample rate and those obtained by a human observer. Once again, we observed that increasing the sample rate allowed EthoVision scores to reach human observation scores (intraclass correlation indexes were 0.024,  $p = 0.266$ ; 0.239,  $p = 0.01$ ; 0.909,  $p = 0.001$ , for 1 sample/6 s, 1 sample/s, 6 samples/s, respectively). Finally, we constructed different Bland–Altman plots and calculated the bias index provided by this method. More specifically, the bias (average discrepancy) was calculated by averaging the difference between the scores obtained for each individual by hand scoring and by the software at each sample rate. Fig. 1B depicts how bias is reduced as a function of the sample rate. This effect was confirmed by means of ANOVA [ $F_{(2,10)} = 105.656$ ,  $p < 0.0001$ ] and subsequent SNK post hoc multiple comparisons, which revealed that higher the sample rate, lower the bias (1 sample/6 s versus 1 sample/1 s,  $p < 0.05$ ; 1 sample/6 s versus 6 samples/1 s,  $p < 0.01$ ). Taken together, these data suggest that with the appropriate sample rate, freezing scores obtained by the EthoVision system are analogous to those obtained by well-trained human observers.

We next examined the ability of EthoVision to score fear-related behavior in contextual and cued fear conditioning experiments. For contextual fear conditioning, one group (CS) only received exposure to the context, but no shock. A different group received three CS–US pairings during conditioning. There was a significant difference between the CS and CS–US groups in a 24 h retention test ( $F_{(3,28)} = 23.0$ ,  $p < 0.0001$ ). We observed no differences between automated scoring of immobility and our hand scoring in either the CS group or the CS–US group (Fig. 2), using the sample rate of 6 samples/s. Importantly, a higher sample rate of 15 sample/s results in much higher false positive scoring (data not shown). For cued fear conditioning, one group (CS) only received exposure to the context and tone, but no shock. A different group received three CS–US pairings during conditioning, in which the CS is a 30-s tone that co-terminates with a foot shock. The 24-h retention test, performed in a novel context, consists of a pre-CS period (time interval before the tone comes on during the retention test) and a post-CS period (time interval during which the tone is on in the retention test). There was a significant difference between CS and CS–US groups in the post-CS period during the retention test ( $F_{(7,46)} = 52.0$ ,  $p < 0.0001$ ). There were no differences within the CS only group with regard to pre-CS and post-CS immobility. These results demonstrate that the modified PhenoTyper cages can be used for contextual fear conditioning and that EthoVision software parameters can be set to reliably and accurately match human scoring of fear-related behavior.

#### 4. Discussion

The main conclusion of the present study is that, under the proper software configuration, the use of the hardware–software tandem provided by PhenoTyper–EthoVision resulted in fully automated, reliable, and precise quantification of immobility in mice subject to contextual or cued fear conditioning. In this regard, the use of this system might provide a standardized high-throughput automated set-up to measure fear-related behavior across different laboratories. Such a standard would be clearly advantageous when trying to compare results from different sources in the current attempt to identify the molecular, cellular and physiological mechanisms underlying learning and memory processes.

Unlike many systems, EthoVision software interfaces with the PhenoTyper chambers directly to operate the chambers and it also scores the behavior. Further, MPEG videos can be recorded simul-



**Fig. 2.** Modified PhenoTyper chambers in combination with EthoVision software can be used for contextual and cued fear conditioning. (A) Mice were subject to contextual fear conditioning and then received a 24-h retention test in the conditioned context. One group was exposed to the context in the absence of a shock (CS-only group). A different group received three CS-US pairings (CS-US group). EthoVision scored immobility nearly identical to scoring by human observation. (B) Mice were subject to cued fear conditioning and then received a 24-h retention test in a novel context. The retention test lasted 5 min, which was comprised of a period of time in the novel context without a tone (Pre-CS) followed by the onset of tone (Post-CS). One group received the tone but no shock (CS-only). A different group received three CS-US pairings (CS-US group). Again, EthoVision scored immobility very similarly to scoring by human observation. Further, these results also demonstrate the use of PhenoTyper chambers for contextual and cued fear conditioning. Values are mean  $\pm$  S.E.M. \* $p < 0.05$ .

taneously such that the videos can be replayed to perform multiple rounds of additional scoring of other behaviors of interest. Thus, hardware operation and animal behavior scoring can be finely tuned through the same computer software system. EthoVision permits exquisite control over user defined parameters for automated scoring of behavior allowing the experimenter to adjust tracking sensitivity and behavioral scoring to achieve measures that are virtually identical of those reported by well trained human observers. Importantly, this close match between human observer and EthoVision scores remained regardless of the levels of freezing (e.g. the same degree of concordance was observed at low and high levels of freezing; see Fig. 2).

The PhenoTyper-EthoVision system provides several additional advantages over human observers. First, this system allows for the simultaneous monitoring of multiple chambers, which increases productivity and data acquisition in this time-consuming process. Second, because it can operate under infra-red conditions, we have observed that the PhenoTyper-EthoVision system can provide the same reliable scores under a wide variety of conditions, including total darkness (data not shown). This might be a crucial feature

for growing number of studies dealing with the involvement of circadian rhythms in learning and memory, and many other behavioral domains (Eckel-Mahan et al., 2008; Perreau-Lenz et al., 2007; Zueger et al., 2006). Third, because EthoVision software processes digital video, behavioral set-ups can be run in nearly any location as long as a digital video camera is used to capture video, which can then be processed by EthoVision.

Our results extend previous reports on the different possible applications of this equipment used to measure locomotor behavior in rodents (de Visser et al., 2005, 2006, 2007; Dalm et al., 2008). In summary, the present study shows that the combined use of PhenoTyper chambers with the EthoVision software package provides an integrated system able to measure fear-related behavior generated by fear conditioning in a fast, reliable and flexible manner.

## Acknowledgments

This work was supported by the Whitehall Foundation and NIMH (R01MH081004) (M.A.W.) and a Fundacion Alicia Koplowitz Fellowship (C.S.-S.).

## References

- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;307:310.
- Dalm S, de Visser L, Spruijt BM, Oitzl MS. Repeated rat exposure inhibits the circadian activity patterns of C57BL/6J mice in the home cage. *Behav. Brain Res* 2008 [published online August 3].
- de Visser L, van den Bos R, Kuurman WW, Kas MJ, Spruijt BM. Novel approach to the behavioural characterization of inbred mice: automated home cage observations. *Genes Brain Behav* 2006;5:458–66.
- de Visser L, van den Bos R, Spruijt BM. Automated home cage observations as a tool to measure the effects of wheel running on cage floor locomotion. *Behav Brain Res* 2005;160:382–8.
- de Visser L, van den Bos R, Stoker AK, Kas MJ, Spruijt BM. Effects of genetic background and environmental novelty on wheel running as a rewarding behaviour in mice. *Behav Brain Res* 2007;177:290–7.
- Eckel-Mahan KL, Phan T, Han S, Wang H, Chan GC, Scheiner ZS, Storm DR. Circadian oscillation of hippocampal MAPK activity and cAMP: implications for memory persistence. *Nat Neurosci* 2008 [Published online August 10].
- Fanselow MS. Conditioned and unconditional components of post-shock freezing. *Pavlov J Biol Sci* 1980;15:177–82.
- Lattal KM, Barrett RM, Wood MA. Systemic or intrahippocampal delivery of histone deacetylase inhibitors facilitates fear extinction. *Behav Neurosci* 2007;121:1125–31.
- Maren S, Quirk GJ. Neuronal signalling of fear memory. *Nat Rev Neurosci* 2004;5:844–52.
- McCluskey A, Gaaliq A. Statistics I: data and correlations. *Contin Educ Anaesth Crit Care Pain* 2007;7:95–9.
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psych Methods* 1996;1:30–46.
- Noldus LPJ, Spink AJ, Tegelenbosch RAJ. EthoVision: a versatile video tracking system for automation of behavioral experiments. *Behav Res Methods Instrum Comput* 2001;33:398–414.
- Perreau-Lenz S, Zghoul T, Spanagel R. Clock genes running amok: clock genes and their role in drug addiction and depression. *EMBO Rep* 2007;8:S20–3 [Spec No].
- Seaman MA, Levin JR, Serlin RC. New developments in pairwise multiple comparisons: some powerful and practicable procedures. *Psychol Bull* 1991;110:577–86.
- Shrout P, Fleiss JL. Intraclass correlation: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–42.
- Spink AJ, Tegelenbosch RAJ, Buma MOS, Noldus LPJ. The EthoVision video tracking system: a tool for behavioral phenotyping of transgenic mice. *Physiol Behav* 2001;73:731–44.
- Vecsey CG, Hawk JD, Lattal KM, Stein JM, Fabian SA, Attner MA, Cabrera SM, McDonough CB, Brindle PK, Abel T, Wood MA. Histone deacetylase inhibitors enhance memory and synaptic plasticity via CREB:CBP-dependent transcriptional activation. *J Neurosci* 2007;27:6128–40.
- Wood MA, Attner MA, Oliveira AM, Brindle PK, Abel T. A transcription factor-binding domain of the coactivator CBP is essential for long-term memory and the expression of specific target genes. *Learn Mem* 2006;13:609–17.
- Wood MA, Kaplan MP, Park A, Blanchard EJ, Oliveira AM, Lombardi TL, Abel T. Transgenic mice expressing a truncated form of CREB-binding protein (CBP) exhibit deficits in hippocampal synaptic plasticity and memory storage. *Learn Mem* 2005;12:111–9.
- Zueger M, Urani A, Chourbaji S, Zacher C, Lipp HP, Albrecht U, Spanagel R, Wolf DP, Gass P. mPer1 and mPer2 mutant mice show regular spatial and contextual learning in standardized tests for hippocampus-dependent learning. *J Neural Transm* 2006;113:347–56.