eNeuro

Research Article: Open Source Tools and Methods | Novel Tools and Methods

DBscorer: an open source software for automated accurate analysis of rodent behavior in forced swim test and tail suspension test

https://doi.org/10.1523/ENEURO.0305-21.2021

Cite as: eNeuro 2021; 10.1523/ENEURO.0305-21.2021

Received: 17 July 2021
Revised: 9 September 2021
Accepted: 3 October 2021

Open Source Tools and Methods

Title: DBscorer: an open source software for automated accurate analysis of rodent behavior in forced swim test and tail suspension test

Abbreviated title: Tool for automated analysis of FST and TST

Author list: Arnab Nandi[1], Garima Virmani[1], Aatmika Barve[1], Swananda Marathe[1,*]

Affiliations:

[1] Centre for Neuroscience, Indian Institute of Science, Bangalore, India 560012

*Correspondence may be addressed to: Dr. Swananda Marathe, Centre for Neuroscience, Indian Institute of Science, Old TIFR building, Gulmohar Marg, Bangalore, India 560012

email: swanand@iisc.ac.in or swanandlab@gmail.com

Number of:

1. Figures: 4
2. Tables: 1
3. Multimedia: Video:1, .zip:1

Number of words:

1. Abstract: 215
2. Introduction: 700
3. Discussion: 684

Acknowledgments:

Authors would like to thank Mr M. Manjunath and the staff at the central animal facility at IISc for technical help.

Conflict of interest statement:

The authors declare no competing or financial interests.

Funding Sources:

This work was supported by the INSPIRE faculty grant from the Department of Science and Technology, Ministry of Science and Technology, India (DST) to S.M. (DST/INSPIRE/04-I/2016/000002) and Early Career Research Award from Science and Engineering Research Board (SERB) to S.M. (ECR/2017/003240). A.N. and G.V. were supported by the Council of Scientific and Industrial Research, India (CSIR-NET) Junior and Senior Research fellowships.

Abstract:

Forced swim test (FST) and tail suspension test (TST) are commonly used behavioral tests for screening antidepressant drugs with a high predictive validity. These tests have also proved useful to assess the non-motor symptoms in the animal models of movement disorders such as Parkinson's disease and Huntington's disease. Manual analysis of FST and TST is a time-consuming exercise, and has large observer-to-observer variability. Automation of behavioral analysis alleviates these concerns, but there are no easy-to-use open source tools for such analysis. Here, we describe the development of DBscorer, an open source program installable on Windows®, with an intuitive graphical user interface, that helps in accurate quantification of immobility behavior in FST and TST from video analysis. Several calibration options allow customization of various parameters to suit the experimental requirements. Apart from the readout of time spent immobile, DBscorer also provides additional data and graphics of immobility/mobility states across time revealing the evolution of behavioral despair over the duration of the test and allows the analysis of additional parameters. Such comprehensive analysis allows a more nuanced understanding of the expression of behavioral despair in FST and TST. We believe that DBscorer would make analysis of behavior in FST and TST unbiased, automated and rapid, and hence prove to be helpful to the wider neuroscience community.

Significance Statement:

FST and TST are commonly used to rapidly screen novel antidepressant compounds. They are also used to assess the non-motor behaviors in models of Parkinson's and Huntinton's disease. However, the manual analysis of behavior is time-consuming and subject to observer-to-observer variability. The tools available for automation are either difficult to use, are expensive, require special apparatus, or not comprehensively validated. DBscorer, described here, is an open source software for Windows® with a user-friendly GUI, which we have extensively validated against the performance of highly trained human scorers. We believe that the ease of installation and use, as well as the high accuracy would lead to a widespread adoption of DBscorer by the neuroscience community.

Introduction:

The serendipitous discovery of antidepressant drugs in the 1950s (Kuhn, 1958; Loomer et al., 1957) led to a quest to understand their mechanisms of action. This necessitated the development of suitable rodent models to study the action of antidepressant drugs. In the 1970s, Porsolt and colleagues described a new test to model behavioral despair in rodents (Porsolt et al., 1978, 1977). The test, called the Forced Swim Test (FST), involves introducing the animal (mouse or rat) into a narrow cylindrical container containing water. After the initial period of vigorous swimming activity, the emergence of behavioral despair is evident from the time spent immobile. It was shown that a single injection with various classes of antidepressant drugs was sufficient to decrease the time spent immobile in the FST (Detke et al., 1997; Porsolt et al., 1978, 1977). Further, a dry version of FST, called the Tail suspension test (TST), was proposed, where a mouse is suspended by its tail and the time spent immobile is scored as a measure of despair (Steru et al., 1985).

Both FST and TST have become classical tests for assessing depressive-like behavior in rodents. Since the initial study by Porsolt and colleagues, several other studies also reported that a single injection with antidepressant drugs significantly reduces the immobility in both FST and TST, allowing the test to be used to rapidly screen novel compounds for antidepressant-like activity (Castagné et al., 2009). Since these tests require the rodents to perform vigorous and coordinated motor activity under a stressful environment, these tests also find use in the fields of movement disorders (Bonito-Oliva et al., 2014; Campos et al., 2013; Du et al., 2015; Pouladi et al., 2009; Soylu-Kucharz et al., 2016).

However, these tests often suffer from observer-to-observer variability; and reproducibility is often a concern (Kara et al., 2017; Smalheiser et al., 2021; Trunnell and Carvalho, 2021). Moreover, the manual analysis of behavior is extremely time-consuming, thus hampering the utility of these tests for high throughput screening of candidate compounds.

Automation of the behavioral analysis in FST and TST could help in objective, quick and highly reproducible analysis of behavior. TST lends itself to easy automation through the use of a strain gauge to which a mouse is suspended by its tail and the changes in the force on the strain gauge is measured, which can then be used to distinguish mobility from immobility. This approach was first described in 1987 and multiple manufacturers now have strain gauge-based TST devices available for purchase (Cryan et al., 2005; Mombereau et al., 2004; Steru et al., 1987). Apart from the costs of acquisition and maintenance, a big disadvantage with this method is the number of mice that can be studied in parallel, which is dictated by the configuration of the device. On the other hand, the automation of FST is principally done through video analysis (Crowley et al., 2004; Gersner et al., 2009). Analysis of TST can also be done through video analysis, thus bypassing the need to use a specialized apparatus. Such analyses are often performed using commercially available software by tracking the frame-to-frame variations, changes in the position of the centroid etc (Juszczak et al., 2006). However, to our knowledge, there are no easy-to-use free open-source softwares available for FST and TST. Furthermore, an easy and intuitive graphical user interface (GUI) is only possible if the tool is specifically designed for FST and TST, since the general purpose tools would not allow easy calibration of parameters that are relevant to FST and TST. Here, we describe the development of DBscorer (Depression Behavior Scorer), a MATLAB®-based tool, designed for rapid and automated phenotyping of behavioral analysis in FST and TST. We validated the performance of the software against that of trained scorers and compared the behavior in an experiment involving a Chronic Mild Unpredictable Stress (CMUS) in mice. As long as the background provides sufficient contrast, DBscorer can be used on black as well as white animals. We believe that ease of installation and an intuitive GUI of DBscorer would help researchers with no programming knowledge to perform automated analysis of behavior in FST and TST and would thus help in a standardized, unbiased and objective analysis of behavior.

Materials and methods:

Animals: 6 months old C57BL6/J male mice were used for experiments. Animals were bred and housed in the [Author University] and the experiments were performed in accordance with the protocols approved by the institutional animal ethics committee of [Author University]. Mice were housed in groups of 3-5 animals per cage and were maintained on a 12:12 hour light:dark cycle with access to food and water *ad libitum*. Usage of animals was reduced as much as possible in accordance with the principle of 3Rs.

Forced Swim Test: To quantify the behavioral despair in FST, the mice were allowed to swim in a 2L glass beaker filled upto 75% of its capacity with water. Before each test, the beaker was thoroughly cleaned. Water temperature was maintained between 21°C and 25°C. Mouse was placed in the middle of the container by gently holding its tail. Their behaviour was videotaped for 5 minutes and the behavior over the entire 5 minute period was analyzed. The FST protocol was adapted from previous studies (Castagné et al., 2011; Jaggar et al., 2017). 6 minutes long FST protocols have also been used, with the first 2 minutes removed from analysis (Can et al., 2012a; Kara et al., 2017). The choice of the method in this study was based on a protocol standardized earlier in a similar chronic stress paradigm (Virmani et al., 2021). Every effort was made to minimise the effects of reflected light from the water surface. Lights were mounted on top of the behavioral setup at a height of 10 feet, with a light intensity of 100-120 lx at the level of the beakers. Non-reflective matte-finished white surface was placed behind the beakers to minimize the glare. The camera was set horizontally at water level and at a distance of 3 feet from the beakers. The position of the camera and beakers remained fixed over all the video recordings across multiple days.

After the test, mice were dried and put into a warm cage (30°C-33°C) for 20 minutes before returning to their home cage. For manual analysis, the mouse was considered immobile when floating passively and making only movements that were required to keep its head above water. Time spent immobile was reported as a % of total time.

Tail Suspension Test: For the tail suspension test, mice were hung by their tails for a period of 5 minutes. Since C57BL6/J mice show extensive tail climbing behaviour as reported earlier (Can et al., 2012b), we passed the tail through a lightweight plastic tube (0.5g) to prevent tail climbing as described previously (Can et al., 2012b). The lighting conditions and camera placement were similar to FST. After the tests were done, mice were returned temporarily to a holding cage until all the animals from their home cage were tested.

CMUS: Chronic mild unpredictable stress was performed as described earlier (Virmani et al., 2021). In brief, mice were subjected to 1-2 stressors per day for 21 consecutive days. Stressors were picked randomly from a set of stressors described earlier (Virmani et al., 2021) and were administered at a random time during the day. FST was performed on day 1 and day 19. TST was performed on days 2 and 17. These tests not only served as stressors, but helped in the assessment of the emergence of behavioral despair as a result of CMUS.

Video Acquisition: The videos were acquired using a Nokia 6.1 mobile camera mounted horizontally at the level of the animal. The videos were converted to 15 fps mp4 format using the ffmpeg tool (http://ffmpeg.org/). The lighting conditions as well as the white background were selected so as to provide the best possible contrast, while minimizing the glare.

Development of DBscorer: Code for the DBscorer software was written in MATLAB® 2020 (www.mathworks.com). Code can either be run on MATLAB® or as a standalone program on Windows® requiring matlab runtime 9.9 (www.mathworks.com). The user interface was built using the MATLAB® App Designer. The interface is very intuitive and easy to use. The program analyzes video files in multiple video formats like .avi, .mp4, .mov which are supported by MATLAB®. It takes various user inputs like start and endpoints, area within the video to be analyzed, blurring and threshold values. Then it converts each frame to binary images according to user inputs. Threshold values were extensively tested and the recommended thresholds are given in Table 1.

Difference between the area of the binarized image as a percentage of area of the previous frame is then calculated and averaged for each second. When this number is below a given threshold, the animal is considered to be immobile. The data is analyzed in blocks of user-defined timelengths, and immobility time is then calculated for each of these blocks.

For optional calibration or manual scoring of behaviour, videos can be analysed manually by the user using the software interface. It is recommended to use a video with approximately equal time of two states (mobility and immobility) to avoid any bias during calibration. The user has to press a toggle button when the animal becomes immobile to start quantification of immobility time, and then to press it every time when the animal transitions between the mobility and immobility states. For each video, an output file is created containing the timestamp for every second. Score of ‘1’ is assigned to immobility (magenta toggle) and ‘0’ to mobility (green toggle) for each second.

We have used 60% of all videos from each set to get the optimum threshold. The change in percentage area is then analyzed automatically by the software using various area threshold values and results for each second are systematically compared with the manual score to generate a ROC curve. We observed that humans typically take between 1-3 seconds to respond to the change of behavioral state with a key press, perhaps as a result of a combination of indecisiveness, inherent ambiguity in assessing an animal's behavior and a lag in motor response. On the other hand, the behavioral state transitions are instantaneous in the case of automated analysis. Hence, the match between the 2 methods was poor around the boundaries between immobility and mobility states, but the match was stronger away from the boundaries. Since the time lags for a human scorer are not constant throughout, we could not artificially align the data by shifting the time-series. Hence, we removed 3 seconds around the boundaries while obtaining the optimum thresholds. The optimal threshold is determined from the ROC curve and reported in the user interface and in the output file. The formula is given below, where c is the cut-point.

Optimum Cut-Point (c)= Sensitivity(c) $\times$ Specificity(c) (Unal, 2017)

We take the threshold with the maximum value of the optimal cut-point.

Code Accessibility: The source code is available as supplemental data and in the online repository at http://github.com/swanandlab/DBscorer under a GPLv3 license. Periodic updates to this code, if any, will be made available on the link provided above.

Statistical Analysis: All analyses were done in MATLAB® (2020b) or GraphPad Prism. Pearson's correlation and corresponding p-values were calculated and the scatter plots with linear regression were made using GraphPad Prism. The Bland-Altman plot was plotted and the bias was calculated using GraphPad Prism. The two group comparisons for the CMUS experiments were made using paired 2-tailed student's *t*-test and the data were plotted as bar graphs depicting mean ± SEM with overlaid scatter plot using GraphPad Prism. All other plots were prepared in MATLAB®.

Results:

Development of DBscorer and the GUI, and the estimation of optimum threshold

We developed DBscorer in MATLAB® as described in the materials and methods. The GUI (Figure 1A) was designed in MATLAB App Designer and the script was compiled as an executable file using MATLAB Compiler™. This allows direct installation on Windows® as a standalone program. A typical workflow (Figure 1A, Movie 1) involves loading the video, selecting time boundaries of the part of the video that needs to be analyzed and marking the minimum area such that the animal remains within the selected area throughout the test. The area is marked using a multi-point selection, after which a rectangular area from the video is cropped out for analysis. This is followed by automated thresholding, though it is possible to input a user-defined binary threshold value. By pressing the “Background Fill” button and by marking the outer boundaries of the animal, we can estimate the background behind the animal. Following this, the analysis can be done completely automatically by pressing “Automatic Analysis”. DBscorer can also be used to perform the analysis manually by pressing the “Manual Analysis” button. For manual analysis, video playback can be controlled using the Play/Pause toggle switch. Another toggle switch called “State” is used to alternate between mobility and immobility during manual analysis.

For automated analysis, we tested 3 different parameters for their correlation with human scorers. These parameters included changes in object length, object area and frame-by-frame variation. While change in object length was marginally better in the case of TST, change in area was more versatile and suited for both FST as well as TST. Hence, change in the area of the object was used for further validation (Figure 1B). Δ Area Threshold as a % of the area of the previous second is used to classify the data into mobility or immobility (Figure 1B). Furthermore, manual analysis can be done to automatically calibrate the Δ Area Threshold (%). Table 1 summarizes the Δ Area Thresholds from our calibrations, which we recommend to the users.

We used the ROC curves to obtain an optimum threshold for each test as described in the materials and methods (Figure 1C and 1D; and Table 1). Using these thresholds, we sought to compare the performance of DBscorer and human scorers in FST and TST.

Comparison of automated analysis with DBscorer against manual analysis in TST

To validate the performance of the DBscorer, comparisons with the manual analysis were made for both TST as well as FST. For TST, we used 20 videos from C57BL6/J mice subjected to TST. We computed 3 parameters for each mouse, namely percent time spent immobile, Latency to the first bout of immobility, and the longest bout of mobility. Although the traditional method of manual analysis using a stopwatch does not allow the measurements of parameters other than the time spent immobile, manual analysis on DBscorer allows these measurements. We used the Bland-Altman plots to assess the agreement between the two methods for all 3 parameters. We found a good agreement between DBscorer and manual scoring on all 3 parameters for TST (Figure 2 A, 2B and 2C). The bias for the immobility % was 5.4 and 95% limits of agreement from -7.40 to 18.20 (Figure 2A). For latency to immobility, the bias was -1.35 the 95% limits of agreement were -18.05 to 15.35 (Figure 2B). In the case of the longest bout, the bias was 4.25 and the 95% limits of agreement were from -9.13 to 17.63.

We next assessed the correlation between the results obtained using the 2 methods. In the percent time spent immobile, we found a strong correlation between DBscorer and manual analysis (R^2 (20) = 0.87, $p < 0.0001$) (Figure 2D). The correlation coefficients and p values for latency to immobility were R^2 (20) = 0.83 and $p < 0.0001$ respectively (Figure 2E), while those for the longest bout were R^2 (20) = 0.52 and $p < 0.0003$ respectively (Figure 2F).

Comparison of automated analysis with DBscorer against manual analysis in FST

For FST, we used 20 videos from C57BL6/J mice subjected to FST. We found a good agreement between DBscorer and manual scoring on all 3 parameters for FST (Figure 3A, 3B and 3C). The bias for % immobility was 2.3 and 95% limits of agreement were from -8.58 to 13.18 (Figure 3A). For latency to immobility, the bias was -1.10 the 95% limits of agreement were -29.55 to 27.35 (Figure 3B). In the case of the longest bout, the bias was -5.95 and the 95% limits of agreement were from -37.57 to 25.67 (Figure 3C).

We next assessed the correlation between the results obtained using the 2 methods. In the percent time spent immobile, we found a strong correlation between DBscorer and manual analysis (R^2 (20) = 0.90, $p < 0.0001$) (Figure 3D). The correlation coefficients and p values for latency to immobility were R^2 (20) = 0.87 and $p < 0.0001$ respectively (Figure 3E), while those for the longest bout were R^2 (20) = 0.73 and $p < 0.0003$ respectively (Figure 3F).

Analysis of TST and FST behavior in the CMUS paradigm

To test DBscorer in a real-world experiment, we used a 21 days CMUS paradigm. It was shown that such a paradigm leads to a significant increase in immobility and other related parameters (Virmani et al., 2021). TST was a part of the CMUS paradigm as mentioned in the materials and methods section and was performed on day 2 and day 17 of the paradigm. In agreement with previously reported data, there was a significant increase in immobility at day 17 as compared to day 2 as seen from the raster plots of immobility (Figure 4A and 4B). Comparisons using paired t-tests also showed a statistically significant increase in the % time spent immobile (Day 2: 34.92 ± 3.25 %, Day 17: 63.08 ± 3.95 %; p = 0.0015) (Figure 4C), a decrease in the latency to immobility (Day 2: 50.88 ± 3.98 seconds, Day 17: 19.63 ± 4.44 seconds; p = 0.0002) (Figure 4D) and an increase in the length of the longest bout of immobility (Day 2: 19.00 ± 2.74 seconds, Day 17: 29.50 ± 2.89 seconds; p = 0.0221) (Figure 4E).

Time spent immobile in FST was analyzed on day 1 and day 19 of the CMUS paradigm, which showed an increase at day 19 as compared to day 1 as can be seen from the raster plots for FST (Figure 4F, 4G). Paired t-test analyses also showed a significant increase in time spent immobile (Day 1: 44.00 ± 4.34 %, Day 19: 73.17 ± 2.76 %; p = 0.0013) (Figure 4H), a decrease in the latency to immobility (Day 1: 67.63 ± 9.37 seconds, Day 19: 6.12 ± 1.63 seconds; p = 0.0003) (Figure 4I) and a significant increase in the duration of the longest bout of immobility (Day 1: 29.63 ± 4.44 s, Day 19: 44.50 ± 5.27 seconds; $p = 0.0232$) (Figure 4J).

Taken together, the automated analysis on DBscorer revealed the emergence of the depressive-like behavior in a CMUS paradigm not just with the conventional parameter of the % of time spent immobile, but also with additional parameters and a raster plot. We believe that the DBscorer would prove to be a faster and more objective analysis method that will accelerate the screening for novel antidepressant compounds.

Discussion:

Shortcomings of currently prescribed antidepressants have necessitated the search for novel antidepressant drugs. This search requires screening methods that are easy, efficient, and objective. FST and TST provide a quick behavioral test for screening novel compounds for their antidepressant-like activity. The analysis of behavioral despair in these tests typically involve the quantitation of the time that the rodents spend immobile. While manual analysis by a trained scorer is a norm, such manual analyses are often tedious, inefficient and subject to high inter-scorer variability. Fortunately, these methods do lend themselves to automation based on video analysis. Automated detection of the rodents in the video frame followed by measurements of the changes in various features of the object between frames has been tried and published before (Cryns et al., 2006; Gao et al., 2014; Hayashi et al., 2011; Hédou et al., 2001; Juszczak et al., 2008; Kulikov et al., 2010; Kurtuncu et al., 2005; Pennington et al., 2019; Rocha et al., 2005; Sturman et al., 2020; Yuman et al., 2008). TST has also been automated using a strain-gauge apparatus (Alexandre et al., 2006; Crowley et al., 2004; Cryan et al., 2005; Mombereau et al., 2004; Steru et al., 1987; Strekalova et al., 2004). But video analysis is preferred since it does not require any special apparatus, and the number of animals that can be simultaneously tested is not dictated by the configuration of the apparatus (Can et al., 2012b).

To our knowledge, there is a dearth of video analysis tools that are user friendly with a GUI, free-to-use and extensively validated. Here we describe the development of DBscorer, an open-source software written in MATLAB®, and has an intuitive GUI for ease of use. We tested the performance of DBscorer against that of experienced scorers. We found a significant correlation between manually scored data and the data obtained from DBscorer (Figure 2, 3). Furthermore, we also used DBscorer to analyse a real-world experiment to monitor behavioral changes as a result of a 21-day CMUS paradigm. In addition to the parameters that can be calculated manually, DBscorer returned additional parameters as well as a raster plot depicting the evolution of behavioral despair as a function of time (Figure 4). These parameters and graphics can also be generated using event recording tools such as ETHOM (Shih and Mok, 2000), but this needs a completely manual analysis of behavior. We believe that the detailed behavioral analysis by DBscorer can provide additional insights into the effects of experimental interventions.

When an animal is first introduced into the water during FST, or suspended by its tail for TST, they typically exhibit erratic mobility behavior for some initial duration that can obscure the real effects of the treatment being studied. Hence, typically the initial period is removed from the analysis. Every lab has its standard operating procedure, and anywhere from 0-2 initial minutes are removed from the analysis. We believe that this would not only depend on the species and the strain of animals being tested but may also vary depending on the specific experimental intervention being studied. The raster plot provided by DBscorer would help make a more informed and objective decision on the specific period to analyse from the total length of the test. In addition, other parameters such as latency to immobility and the duration of the longest bout could also provide additional insights into the animals' responses to the experimental interventions. Furthermore, the source code for DBscorer has been made open source so that the wider community can collaboratively improve upon the analysis and add additional features that may further our understanding of behavioral despair in rodents.

We believe that the availability on the Windows® platform and an intuitive GUI would help DBscorer to be easily adopted by users with no knowledge of computer programming. On the other hand, more advanced users can modify and adapt the software as per their requirements. Moreover, continued collaborative development of DBscorer would help further improve the software. In summary, we believe that DBscorer would prove to be incredibly useful for the scientific community working on depression and antidepressant treatments in rodent models.

Author Contributions:

Conceptualization: A.N., S.M.; Methodology: A.N., G.V., A.B.; Software: A.N.; Formal analysis: A.N., G.V., A.B.; Investigation: A.N., G.V., A.B.; Resources: S.M.; Writing - original draft: S.M.; Writing - review & editing: A.N., S.M.; Supervision: S.M.; Project administration: S.M.; Funding acquisition: S.M.

References:

Alexandre C, Popa D, Fabre V, Bouali S, Venault P, Lesch K-P, Hamon M, Adrien J (2006) Early Life Blockade of 5-Hydroxytryptamine 1A Receptors Normalizes Sleep and Depression-Like Behavior in Adult Knock-Out Mice Lacking the Serotonin Transporter. J Neurosci 26:5554.

Bonito-Oliva A, Masini D, Fisone G (2014) A mouse model of non-motor symptoms in Parkinson's disease: focus on pharmacological interventions targeting affective dysfunctions. Front Behav Neurosci 8.

Campos FL, Carvalho MM, Cristovão AC, Je G, Baltazar G, Salgado AJ, Kim Y-S, Sousa N (2013) Rodent models of Parkinson's disease: beyond the motor symptomatology. Front Behav Neurosci 0.

Can A, Dao DT, Arad M, Terrillion CE, Piantadosi SC, Gould TD (2012a) The Mouse Forced Swim Test. J Vis Exp.

Can A, Dao DT, Terrillion CE, Piantadosi SC, Bhat S, Gould TD (2012b) The Tail Suspension Test. J Vis Exp.

Castagné V, Moser P, Porsolt RD (2009) Behavioral Assessment of Antidepressant Activity in Rodents In: Methods of Behavior Analysis in Neuroscience. 2nd Edition , CRC Press/Taylor & Francis.

Castagné V, Moser P, Roux S, Porsolt RD (2011) Rodent models of depression: forced swim and tail suspension behavioral despair tests in rats and mice. Curr Protoc Neurosci Chapter 8.

Crowley JJ, Jones, O'Leary OF, Lucki I (2004) Automated tests for measuring the effects of antidepressants in mice. Pharmacol Biochem Behav 78.

Cryan JF, Mombereau C, Vassout A (2005) The tail suspension test as a model for assessing antidepressant activity: review of pharmacological and genetic studies in mice. Neurosci Biobehav Rev 29.

Cryns K, Shamir A, Shapiro J, Daneels G, Goris I, Van Craenendonck H, Straetemans R, Belmaker RH, Agam G, Moechars D, Steckler T (2006) Lack of Lithium-Like Behavioral and Molecular Effects in IMPA2 Knockout Mice. Neuropsychopharmacology 32:881–891.

Detke MJ, Johnson J, Lucki I (1997) Acute and chronic antidepressant drug treatment in the rat forced swimming test model of depression. Experimental and Clinical Psychopharmacology.

Du X, Pang TY, Mo C, Renoir T, Wright DJ, Hannan AJ (2015) The influence of the HPG axis on stress response and depressive-like behaviour in a transgenic mouse model of Huntington's disease. Exp Neurol 263.

Gao V, Hotz Vitaterna M, Turek F (2014) Validation of video motion-detection scoring of forced swim test in mice. J Neurosci Methods 235:59–64.

Gersner R, Gordon-Kiwkowitz M, Zangen A (2009) Automated behavioral analysis of limbs' activity in the forced swim test. J Neurosci Methods 180:82–86.

Hayashi E, Shimamura M, Kuratani K, Kinoshita M, Hara H (2011) Automated experimental system capturing three behavioral components during murine forced swim test. Life Sci 88:411–417.

Hédou G, Pryce C, Di Iorio L, Heidbreder CA, Feldon J (2001) An automated analysis of rat behavior in the forced swim test. Pharmacol Biochem Behav 70:65–76.

Jaggar M, Weisstaub N, Gingrich JA, Vaidya VA (2017) 5-HT receptor deficiency alters the metabolic and transcriptional, but not the behavioral, consequences of chronic unpredictable stress. Neurobiol Stress 7:89–102.

Juszczak GR, Lisowski P, Sliwa AT, Swiergiel AH (2008) 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 95:400–407.

Juszczak GR, Sliwa AT, Wolak P, Tymosiak-Zielinska A, Lisowski P, Swiergiel AH (2006) The usage of video analysis system for detection of immobility in the tail suspension test in mice. Pharmacol Biochem Behav 85.

Kara NZ, Stukalin Y, Einat H (2017) Revisiting the validity of the mouse forced swim test: Systematic review and meta-analysis of the effects of prototypic antidepressants. Neurosci Biobehav Rev 84:1–11.

Kuhn R (1958) THE TREATMENT OF DEPRESSIVE STATES WITH G 22355 (IMIPRAMINE HYDROCHLORIDE). American Journal of Psychiatry.

Kulikov AV, Morozova MV, Kulikov VA, Kirichuk VS, Popova NK (2010) Automated analysis of antidepressants’ effect in the forced swim test. J Neurosci Methods 191:26–31.

Kurtuncu M, Luka LJ, Dimitrijevic N, Uz T MH (2005) Reliability assessment of an automated forced swim test device using two mouse strains. J Neurosci Methods 149:26–30.

Loomer HP, Saunders JC, Kline NS (1957) A clinical and pharmacodynamic evaluation of iproniazid as a psychic energizer. Psychiatr Res Rep Am Psychiatr Assoc 8:129–141.

Mocking RJT, Naviaux JC, Li K, Wang L, Monk JM, Taylor Bright A, Figueroa CA, Schene AH, Ruhé HG, Assies J, Naviaux RK (2021) Metabolic features of recurrent major depressive disorder in remission, and the risk of future recurrence. Transl Psychiatry 11.

Mombereau C, Kaupmann K, Froestl W, Sansig G, van der Putten H, Cryan JF (2004) Genetic and Pharmacological Evidence of a Role for GABA B Receptors in the Modulation of Anxiety- and Antidepressant-Like Behavior. Neuropsychopharmacology 29:1050–1062.

Nobis A, Zalewski D, Waszkiewicz N (2020) Peripheral Markers of Depression. J Clin Med Res 9.

Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, Mohr DC, Schatzberg AF (2016) Major depressive disorder. Nature reviews Disease primers 2.

Overstreet DH (2012) Modeling depression in animal models. Methods Mol Biol 829.

Pennington ZT, Dong Z, Feng Y, Vetere LM, Page-Harley L, Shuman T, Cai DJ (2019) ezTrack: An open-source video analysis pipeline for the investigation of animal behavior. Sci Rep 9:19979.

Porsolt RD, Anton G, Blavet N, Jalfre M (1978) Behavioural despair in rats: a new model sensitive to antidepressant treatments. Eur J Pharmacol 47.

Porsolt RD, Bertin A, Jalfre M (1977) Behavioral despair in mice: a primary screening test for antidepressants. Arch Int Pharmacodyn Ther 229.

Pouladi MA, Graham RK, Karasinska JM, Xie Y, Santos RD, Petersén A, Hayden MR

(2009) Prevention of depressive behaviour in the YAC128 mouse model of Huntington disease by mutation at residue 586 of huntingtin. Brain 132.

Rocha BA, Fleischer R, Schaeffer JM, Rohrer SP, Hickey GJ (2005) 17β-Estradiol-induced antidepressant-like effect in the Forced Swim Test is absent in estrogen receptor-β knockout (BERKO) mice. Psychopharmacology 179:637–643.

Shih H-T, Mok H (2000) ETHOM: Event-Recording Computer Software for the Study of Animal Behavior. Acta Zoologica Taiwanica.

Smalheiser NR, Graetz EE, Yu Z, Wang J (2021) Effect size, sample size and power of forced swim test assays in mice: Guidelines for investigators to optimize reproducibility. PLoS One 16:e0243668.

Soylu-Kucharz R, Baldo B, Petersén Å (2016) Metabolic and behavioral effects of mutant huntingtin deletion in Sim1 neurons in the BACHD mouse model of Huntington's disease. Sci Rep 6:1–10.

Steru L, Chermat R, Thierry B, Mico J-A, Lenegre A, Steru M, Simon P, Porsolt RD (1987) The automated tail suspension test: A computerized device which differentiates psychotropic drugs. Progress in Neuro-Psychopharmacology and Biological Psychiatry.

Steru L, Chermat R, Thierry B, Simon P (1985) The tail suspension test: A new method for screening antidepressants in mice. Psychopharmacology.

Strekalova T, Spanagel R, Bartsch D, Henn FA, Gass P (2004) Stress-Induced Anhedonia in Mice is Associated with Deficits in Forced Swimming and Exploration. Neuropsychopharmacology 29:2007–2017.

Sturman O, von Ziegler L, Schläppi C, Akyol F, Privitera M, Slominski D, Grimm C, Thieren L, Zerbi V, Grewe B, Bohacek J (2020) Deep learning-based behavioral analysis reaches human accuracy and is capable of outperforming commercial solutions. Neuropsychopharmacology 45:1942–1952.

Trunnell ER, Carvalho C (2021) The forced swim test has poor accuracy for identifying novel antidepressants. Drug Discov Today.

Unal I (2017) Defining an Optimal Cut-Point Value in ROC Analysis: An Alternative Approach. Comput Math Methods Med 2017.

Virmani G, D'almeida P, Nandi A, Marathe S (2021) Subfield-specific effects of chronic mild unpredictable stress on hippocampal astrocytes. Eur J Neurosci.

Walker WH, II, Walton JC, Courtney DeVries A, Nelson RJ (2020) Circadian rhythm disruption and mental health. Transl Psychiatry 10.

Yuman N, Idaku I, Kenkichi Y, Takeshi T, Kensuke O, Hiroshi M (2008) High-speed video analysis of laboratory rats behaviors in forced swim test In: 2008 IEEE International Conference on Automation Science and Engineering , Presented at the 2008 IEEE International Conference on Automation Science and Engineering (CASE 2008) IEEE.

Legends:

Figure 1:

Development of DBscorer and the GUI, and the estimation of optimum threshold

(A) shows a screenshot of the DBscorer user interface showing a tail suspension test (TST) video loaded for analysis. **(B)** The representative line plot shows the % change in the area as a function of time. Epochs of immobility are shown in magenta and epochs of mobility are depicted in green based on the optimized threshold. Receiver operating characteristic (ROC) curves for **(C)** TST and for **(D)** Forced Swim Test (FST) with 0 (red), 1 (blue), 2 (magenta) or 3 (green) seconds removed from the behavioral state transitions. Sensitivity values at the optimum thresholds are shown by black dots for each ROC curve.

Figure 2:

Comparison of automated analysis with DBscorer against manual analysis in the tail suspension test (TST)

Bland-Altman plots for TST are plotted using the data obtained with DBscorer and manual analysis for **(A)** % time spent immobile, **(B)** Latency to the first bout of immobility (seconds) and **(C)** the duration of the longest bout of immobility (seconds). Black solid line at zero denotes complete agreement between the two methods. Dotted black line denotes the mean. Dotted magenta lines depict 1.96 standard deviations (95% limits of agreement) away from the mean. Scatter plots with an overlaid linear fit showing the correlation between the manually obtained data with the data obtained from DBscorer for **(D)** % time spent immobile, **(E)** Latency to the first bout of immobility (seconds) and **(F)** the duration of the longest bout of immobility (seconds). Pearson's correlation coefficient and *p*-value were used to determine the correlation. n=20 videos.

Figure 3:

Comparison of automated analysis with DBscorer against manual analysis in the forced swim test (FST)

Bland-Altman plots for FST are plotted using the data obtained from DBscorer and with manual analysis for **(A)** % time spent immobile, **(B)** Latency to the first bout of immobility (seconds) and **(C)** the duration of the longest bout of immobility (seconds). Black solid line at zero depicts complete agreement between the two methods. Dotted black line denotes the mean. Dotted magenta lines depict 1.96 standard deviations (95% limits of agreement) away from the mean. Scatter plots with an overlaid linear fit showing the correlation between the manually obtained data with the data obtained from DBscorer for **(D)** % time spent immobile, **(E)** Latency to the first bout of immobility (seconds) and **(F)** the duration of the longest bout of immobility (seconds). Pearson's correlation coefficient and *p*-value were used to determine the correlation. n=20 videos.

Figure 4:

Analysis of tail suspension test (TST) and forced swim test (FST) behavior in the chronic mild unpredictable stress (CMUS) paradigm

Raster plots show the bouts of immobility (black) against time in the TST at **(A)** day 2 as compared to **(B)** day 17 of the 21 day long CMUS paradigm. Bar graphs comparing the behavior in TST at day 2 and day 17 show **(C)** a significant increase in the % of time spent immobile, **(D)** a significant decrease in the latency to the first bout of immobility (seconds) and **(F)** a significant increase in the duration of the longest bout of immobility (seconds). Raster plots reveal the bouts of immobility (shown in black) as a function of time in the FST at **(F)** day 1 and **(G)** day 19 of the 21 day long CMUS paradigm. Bar graphs show **(H)** a significant increase in the % of time spent immobile, **(I)** a significant decrease in the latency to the first bout of immobility (seconds) and **(J)** a significant increase in the duration of the longest bout of immobility (seconds) in the FST paradigm shows. * denotes $p < 0.05$, 2-tailed paired *t*-test. n= 8 mice per group.

Table 1:

Receiver-Operating Characteristic (ROC) Analysis and optimum threshold values

To find the optimum Δ Area thresholds for each test, ROC curves were plotted with 0-3 seconds removed around the behavioral state transitions. The table shows the area under the ROC curves and corresponding optimum thresholds. ** denotes the recommended Δ Area Threshold (%) for each test.

Movie 1:

Demonstration of installation and use of DBscorer

The video shows how to install the appropriate version of the MATLAB® runtime and DBscorer and how to perform the analysis automatically as well as manually.

Extended Data 1 Code zip file:

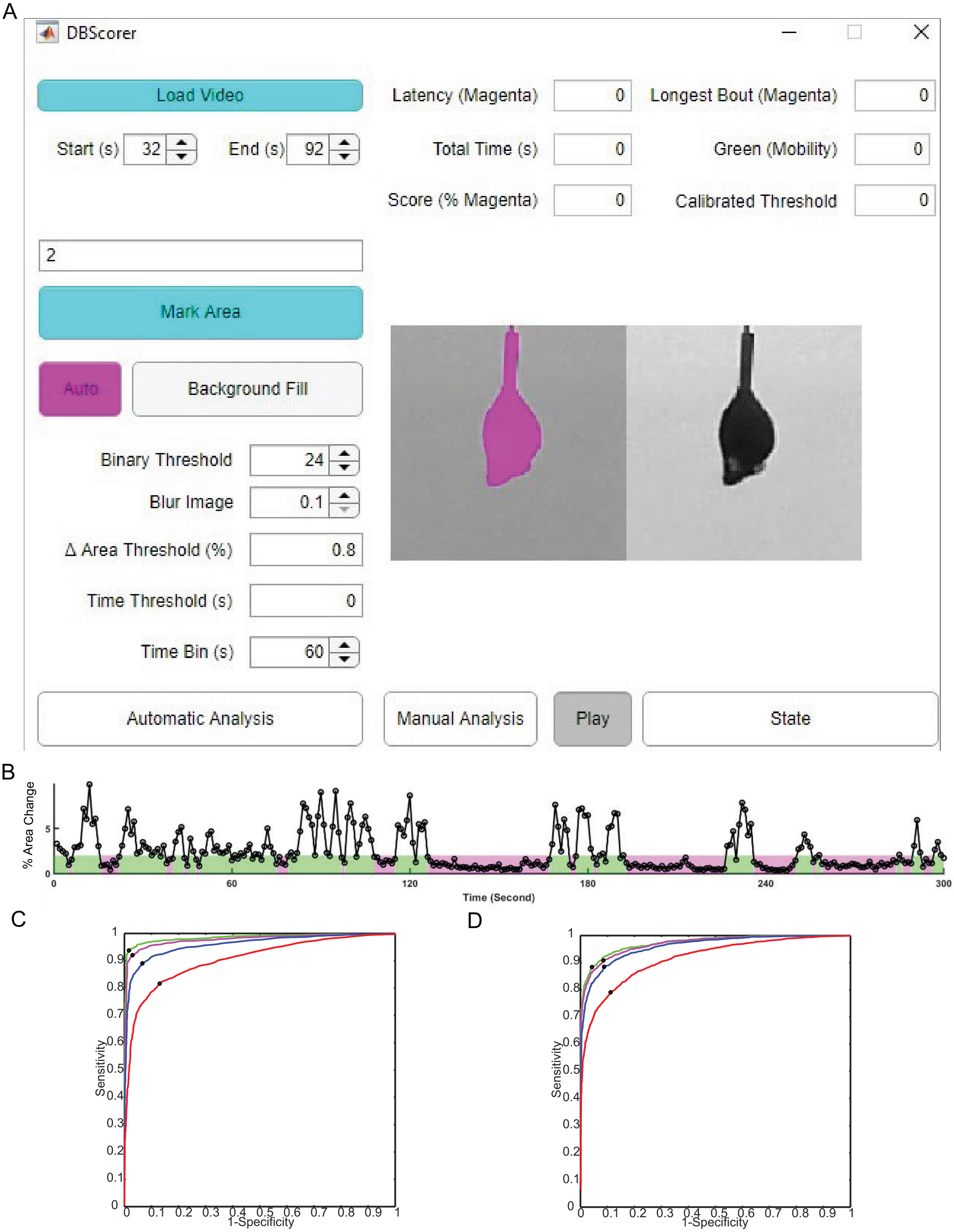
Zip file containing the source code of DBscorer

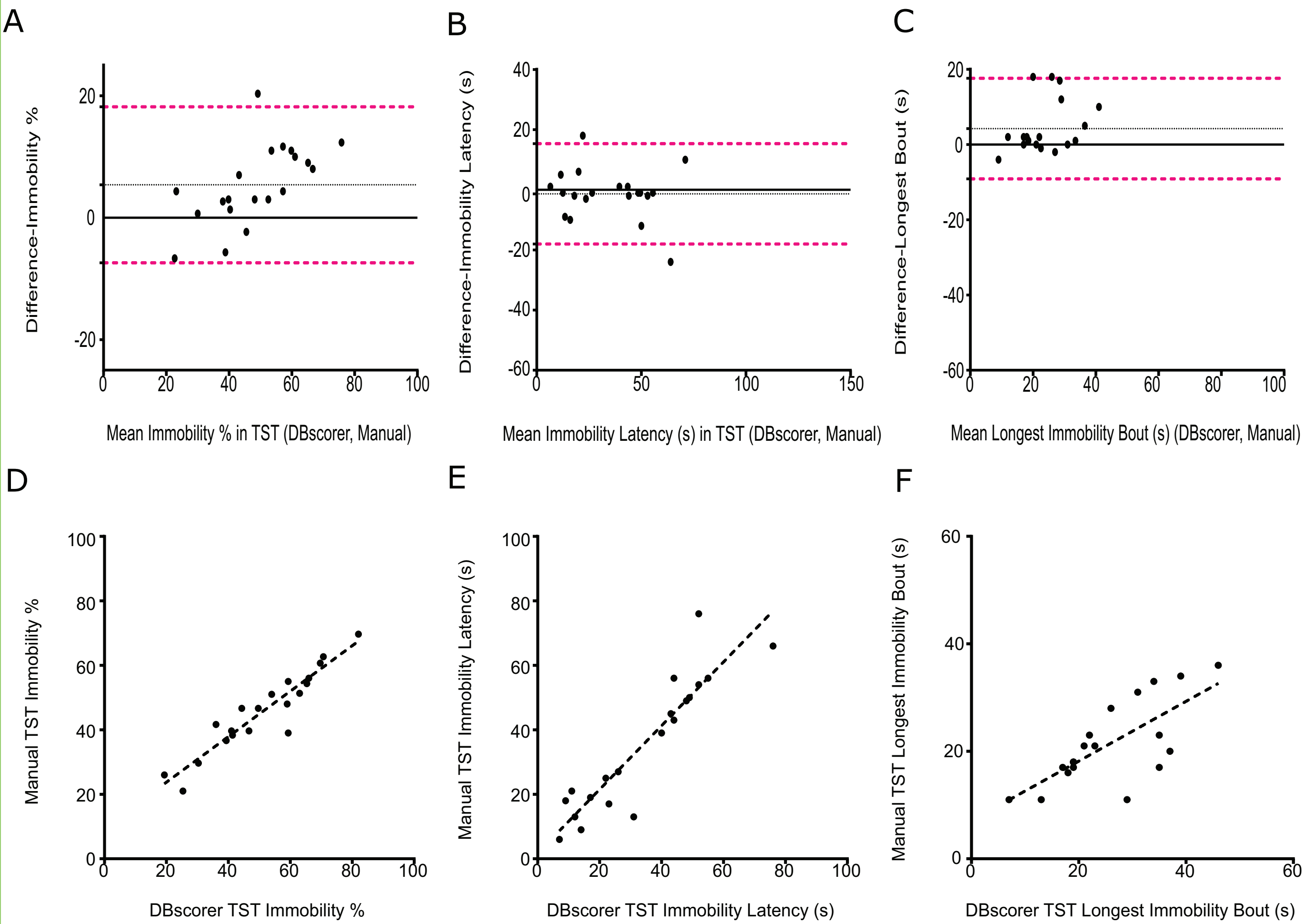
Zip file contains the original source-code of DBscorer written in MATLAB® as well as the installation files. The source code is open source under a GPLv3 license and the updated versions can be accessed in future on http://github.com/swanandlab/DBscorer.

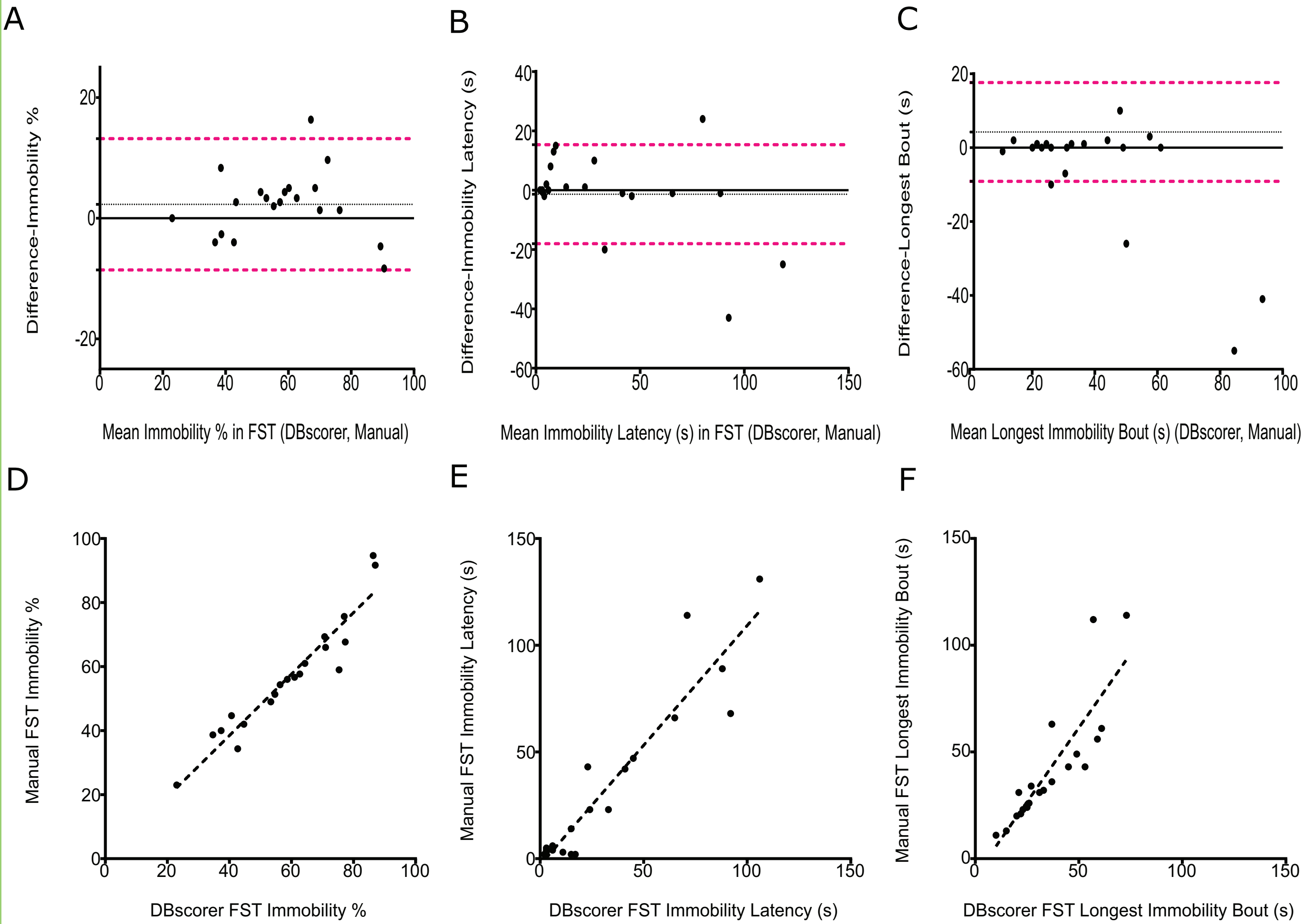
Table 1:

TST Time (Sec) removed	Area Under the Curve	Optimum Threshold
0	0.9047	0.5806
1	0.9598	0.6506
2	0.9775	0.7407
3	0.9847	0.7808**
FST Time (Sec) removed	Area Under the Curve	Optimum Threshold
0	0.9154	2.2207
1	0.9600	2.2306
2	0.9694	2.2306
3	0.9735	2.5861**

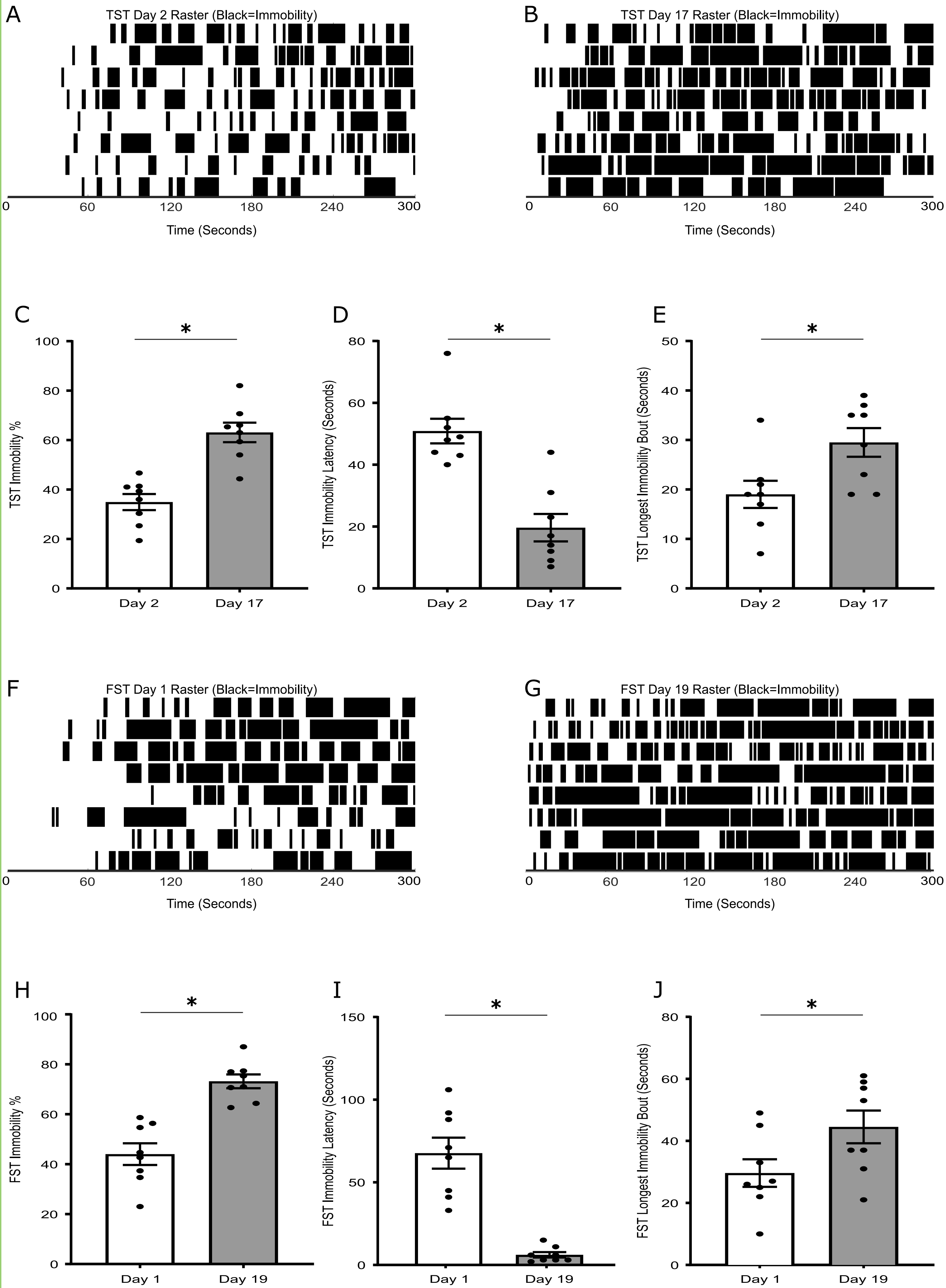
**Recommended thresholds for each test.


A
DBScorer
Load Video
Start (s) 32
End (s) 92
Latency (Magenta) 0
Longest Bout (Magenta) 0
Total Time (s) 0
Green (Mobility) 0
Score (% Magenta) 0
Calibrated Threshold 0
2
Mark Area
Auto
Background Fill
Binary Threshold 24
Blur Image 0.1
Δ Area Threshold (%) 0.8
Time Threshold (s) 0
Time Bin (s) 60
Automatic Analysis
Manual Analysis
Play
State
B
% Area Change
Time (Second)
C
Sensitivity
1-Specificity
D
Sensitivity
1-Specificity


A
Difference-Immobility %
Mean Immobility % in TST (DBscorer, Manual)
B
Difference-Immobility Latency (s)
Mean Immobility Latency (s) in TST (DBscorer, Manual)
C
Difference-Longest Bout (s)
Mean Longest Immobility Bout (s) (DBscorer, Manual)
D
Manual TST Immobility %
DBscorer TST Immobility %
E
Manual TST Immobility Latency (s)
DBscorer TST Immobility Latency (s)
F
Manual TST Longest Immobility Bout (s)
DBscorer TST Longest Immobility Bout (s)



A
Difference-Immobility %
Mean Immobility % in FST (DBscorer, Manual)
B
Difference-Immobility Latency (s)
Mean Immobility Latency (s) in FST (DBscorer, Manual)
C
Difference-Longest Bout (s)
Mean Longest Immobility Bout (s) (DBscorer, Manual)
D
Manual FST Immobility %
DBscorer FST Immobility %
E
Manual FST Immobility Latency (s)
DBscorer FST Immobility Latency (s)
F
Manual FST Longest Immobility Bout (s)
DBscorer FST Longest Immobility Bout (s)



A
TST Day 2 Raster (Black=Immobility)
0 60 120 180 240 300
Time (Seconds)
B
TST Day 17 Raster (Black=Immobility)
0 60 120 180 240 300
Time (Seconds)
C
TST Immobility %
Day 2
Day 17
D
TST Immobility Latency (Seconds)
Day 2
Day 17
E
TST Longest Immobility Bout (Seconds)
Day 2
Day 17
F
FST Day 1 Raster (Black=Immobility)
0 60 120 180 240 300
Time (Seconds)
G
FST Day 19 Raster (Black=Immobility)
0 60 120 180 240 300
Time (Seconds)
H
FST Immobility %
Day 1
Day 19
I
FST Immobility Latency (Seconds)
Day 1
Day 19
J
FST Longest Immobility Bout (Seconds)
Day 1
Day 19