Detecting ataxia using an automated analysis of motor coordination and balance of mice on the balance beam

Lucas Wahl[1#], Fabian M.P. Kaiser[1,2#], Mieke Bentvelzen[1], Joshua White[1], Martijn Schonewille[1], Aleksandra Badura*[1]

[1] Department of Neuroscience, Erasmus MC, Rotterdam, 3015 CN, The Netherlands
[2] Department of Immunology, Erasmus MC, Rotterdam, 3015 CN, The Netherlands
# equal contribution

Correspondence:
Aleksandra Badura (a.badura@erasmusmc.nl)
Department of Neuroscience, Erasmus MC
Wytemaweg 80, 3015 CN Rotterdam
tel: 0031-(0)10 704 35 89

Keywords

Behavior, in vivo models, balance beam, automated classification, locomotion, cerebellum, machine learning, phenotyping, motor coordination

Abstract

Background

The balance beam assay is a well-known paradigm to assess motor coordination in mouse models of neurodegenerative diseases. Classically, these experiments have been quantified using manual annotation, which is time-consuming and prone to inter-experimenter variability.

Methods

We present an open-source analysis pipeline that allows for the automated quantification of motor function. Using an established ataxia model, we validated the pipeline by comparing its output to three independent observers.

Results

Pcp2-Ppp3r1 mutant animals showed a significant increase in the number of missteps and increased time to traverse the beam.

Conclusion

We show that our pipeline can reliably report crossing time, missteps, stops, and falls, thereby offering a high-throughput option with increased inter-experimenter reliability for the analysis of balance beam data.

Significance statement

The analysis of mouse behaviour and motor coordination experiments often lacks automation and standardisation and is thus time-consuming and prone to inter-experimenter variability and poor reproducibility. The balance beam assay is a well-established motor coordination paradigm and currently lacks such an automated, standardised analysis platform. We have developed an open-source pipeline that allows for the automated analysis of crossing time, missteps, stops, and falls. Together with a standardised, easy to assemble balance beam setup, our pipeline offers a cost-effective, and high-throughput option to generate and analyse balance beam data with increased inter-experimenter and intra-experimenter reliability.

Introduction

Mouse models of neurodegenerative disorders such as ataxia are indispensable for a better understanding of the mechanistic underpinnings of loss of motor coordination and balance. Until recently, the analysis of experimental data to quantitatively assess motor deficits has heavily relied on manual scoring by observers. Manual analysis of animal behaviour is time-consuming, arduous, labour-intensive, and subjective, often resulting in poor reproducibility across different observers due to inter-experimenter variability (1–5). Machine learning and deep neural networks have greatly accelerated the analysis of animal behaviour across disciplines and provide standardised and objective results that are not affected by inter-experimenter variability (6). Such analysis strategies rely on the training of automated classifiers by experienced investigators to annotate pre-defined behavioural or motor patterns until the software can reliably identify those without additional user input (3, 7–13).

The balance beam assay has emerged as a widely used and highly sensitive paradigm to study motor coordination and balance in mice (14). The primary measurements of the assay are the time needed to traverse a beam that connects two platforms and the amount of missteps or falls the

animals make during that period. The incentive to traverse the beam is created by either placing food or the animal's homecage on the other platform. The task can be adapted by using flat or round beams of decreasing diameter or by varying the angle of the beam between the two platforms. The balance beam has proven to be more sensitive than the rotarod - another classic paradigm for motor coordination and balance assessment (15).

Other than the automated analysis of gait rhythmicity (16), the analysis of balance beam data has thus far been limited to the manual scoring of missteps or falls. Here we report the development of an open-source analysis pipeline that we generated using the Janelia Automatic Animal Behavior Annotator (JAABA) (7), which reliably reports crossing time, missteps, stops, and falls during the balance beam experiment. Using male and female adult *Pcp2-Ppp3r1* mice with Purkinje cell-specific knock-out of calcium/calmodulin-activated protein-phosphatase-2B (PP2B) on the *C57BL/6J* background - an established ataxia mouse model (17) - we provide proof-of-principle that the pipeline can discern phenotypic differences between ataxic mice and their wildtype littermates. We show that the pipeline can reliably report crossing time, missteps, stops, and falls in a cost-effective, quick, and high-throughput manner with increased inter-experimenter reliability.

Material and Methods

Mice

Calcium/calmodulin-activated protein phosphatase 2B, also known as calcineurin, PP2B or protein phosphatase 3 (Ppp3r1), was selectively deleted from Purkinje cells with the Cre-loxP-system, using mice with loxP sites flanking the regulatory subunit,CNB1 (*Ppp3r1*^{tm1Stl}) (18) and mice carrying the L7-Cre transgene (*Tg(Pcp2-Cre)3555Jdhu) (19)* as described in (17), to generate *Pcp2-cre*$^{+/-}$*-Ppp3r1*$^{fl/fl}$ mice ('*Pcp2-Ppp3r1* mice', hereafter referred to as mutant animals) and *Pcp2-cre*$^{-/-}$*-Ppp3r1*$^{fl/fl}$ mice (hereafter referred to as wildtype animals).

All mice were maintained on a C57BL6/J background. Additionally, 28 adult wildtype C57BL6/J animals were used for classifier training. Experimental animals were between 13 and 19 weeks of age (114 ± 17 days) and housed with food and water available *ad libitum* in a 12:12 h light/dark cycle. All experiments were approved by local and national animal ethical committees (CCD approval: AVD1010020197846).

Balance Beam setup

A balance beam setup was constructed using commercially available supplies. The setup was constructed with two opposing metal platforms positioned 50 cm above an acrylic glass plate with a distance of 86 cm between platforms (Figure 1A). Round aluminium beams of varying diameters (12/10 mm) were clamped below each platform with red markings on the beam indicating the start and end position of a section that measured 70 cm in length, thereby leaving 8 cm distance from the start and end positions on the track to each platform (Figure 1B). The 8 cm gap was excluded from analysis because mice generally paused to prepare for entering or exiting the beam on this section. A moveable aluminium arm with two joints that allows for adjustment of the camera was

inserted at one side of the setup and the camera was placed at the same height as the beam. Recording was performed with a GoPro X HERO7 Black, capturing 29.97 frames per second at a resolution of 2704 x 1520 pixels in linear mode and MP4 format. To improve contrast, a white board was placed behind the setup (for white mice we recommend using a black board).

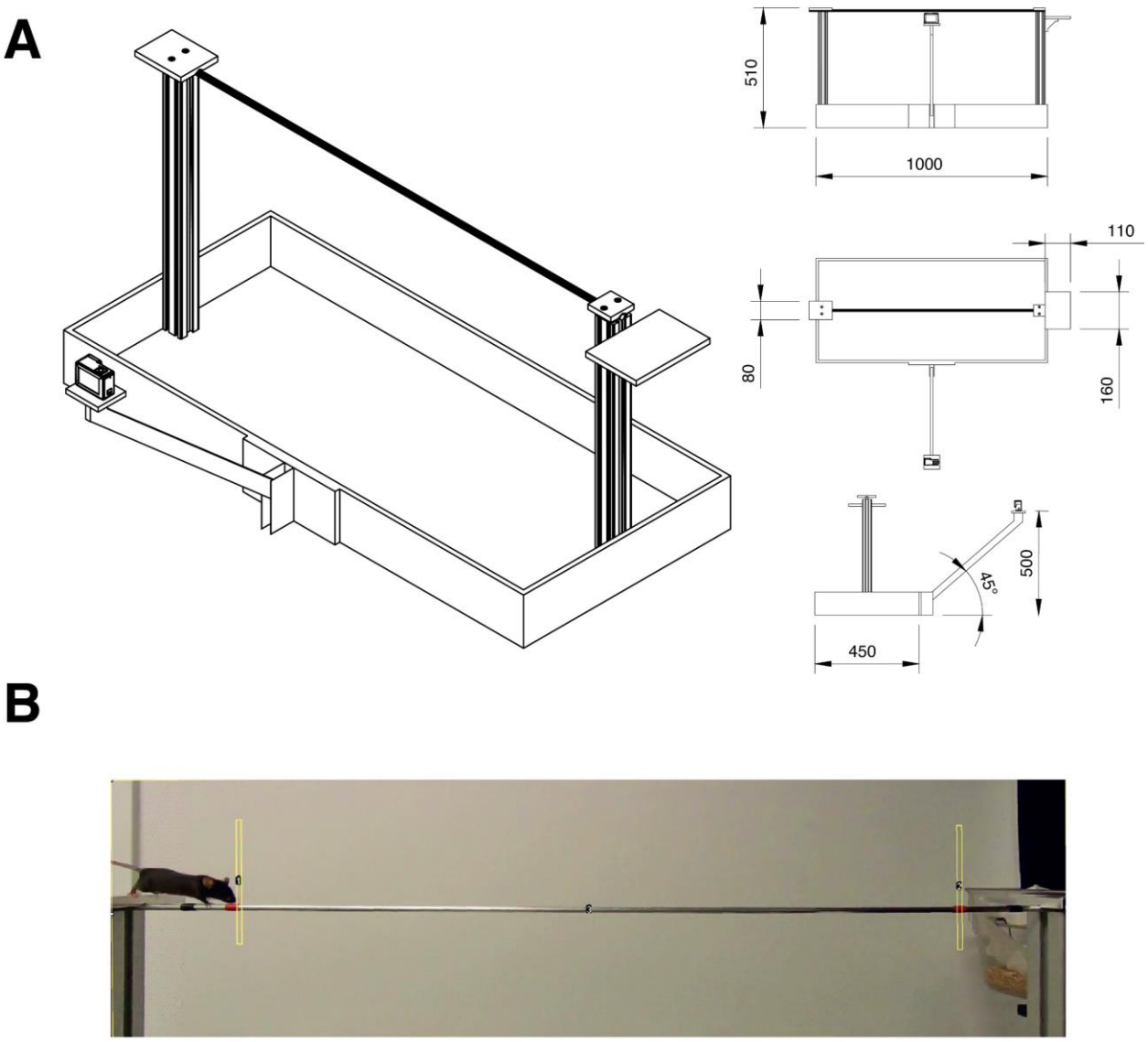

Figure 1. Setup standardisation.

(**A**) Schematic of the balance beam apparatus used to quantify locomotor function in mice. Shown are the isometric view (left) and orthographic projection with size indicators. Measurements are shown in millimetres. (**B**) Side view of the beam in the balance beam

test. Yellow rectangles indicate user-defined start and stop areas used in measuring behaviours and time to traverse the beam.

Experimental design

Mice were habituated to the experimenters - who remained blinded to the genotype - by handling for 1 h per day for three consecutive days. During the course of the experiment, mice were allowed to acclimate to the room for 1 h prior to the begin of the experiment. On day 1, mice were placed into the experimental cage with bedding of the home cage for 10 min prior to being placed on the platform on the opposite pole of the setup. On day 1 and day 2, mice were trained to traverse the balance beam on beams of decreasing diameter in three consecutive trials per beam. On day 1, beams with diameters of 12 mm and 10 mm were used. On day 2, mice were challenged to walk on beams with diameters of 10 mm, 8 mm, and 6 mm. Mutant mice were unable to traverse the 8 mm and 6 mm beams, therefore only the 12 and 10 mm data is included in the results. On day 3, mice were evaluated on all beams in the order of decreasing diameter. After three consecutive trials on the same beam, mice were placed into the home cage and the setup was cleaned with 70% ethanol before the subsequent mouse was tested. Each mouse was given 1 min of rest after each trial. If mice failed three times during consecutive trials, the experiment was terminated and mice were tested on additional beams. At the beginning of each trial, a piece of paper displaying information about the animal, the beam diameter, and the current trial was briefly recorded. Afterwards, the animal was placed on the starting platform and allowed to traverse the beam. The arms of the investigator did not enter the field of view of the camera. Once the animal had reached the platform on the opposite side of the setup, recording was stopped. If the animal fell during the course of a trial, the trial was repeated.

Pre-processing of videos

All subsequent analysis steps were performed on a computer equipped with an AMD Ryzen Threadripper 1950X 16-Core 3.4 GHz Processor with 16 cores and 64 GB RAM, using Windows 10 Pro 64-bit as the operating system. In the first step after data acquisition, videos were pre-processed in order to create files of the adequate format and length as well as to reduce the size, by integrating FFMPEG (https://ffmpeg.org/) and Fiji (20) into MATLAB using the MIJ package (https://imagej.net/plugins/miji). Using FFMPEG, videos were first cropped to cover the entire beam length and remove unnecessary areas above and below the beam. The video file was subsequently converted to the uncompressed AVI format to render it compatible with Fiji. Within Fiji, three regions of interest (ROI) were marked in the standardised video format to label the entire beam as well as the ROIs that designated the start and finish on the beam. The video was then trimmed based on the increase in signal in the start and finish ROIs as the mice traverse the beam in order to remove video segments during which mice were not on the beam. After this step, the files were ready to be used as input for the tracking software to determine the time needed to traverse the beam.

Tracking

In the subsequent step, animals were tracked using the open-source software Motr (2). Motr tracks elliptical trajectories of mice in a closed environment with minimal user input. During the training, the user is asked to correct ellipses that are placed on the animal in order to compute the optimal segmentation threshold. The software uses previous frames to generate multiple hypotheses for the position of the mouse in the current frame and expectation maximisation algorithms are used to estimate the finite most probable position of the mouse in that frame.

Training of the JAABA classifiers

In the final step, we used the open-source software JAABA (7) to train three independent classifiers to identify three common indicators of impaired motor coordination and balance: missteps, stops, and falls. Slips, stops or falls were defined as true positive events in the separate classifiers respectively and the sequences of the remaining video material were labelled as true negative events. JAABA uses supervised machine learning to create libraries with the designated motion patterns, processing information such as the position and angle of the animal as well as the trajectory data. We first implemented a batch-processing step in MATLAB for converting the trajectory data that was generated using Motr into the correct input format for JAABA. We next trained a classifier to identify missteps, which we defined as slipping of the hindpaws that results in a downward motion of the hip, using 483 videos and ~30 hours of training. In total, the classifier was trained on 26374 frames with 4611 frames labelled as missteps and 21763 frames without such events. The minimum bout length was set to three frames (90 ms) by post-processing of the data.

Subsequently, we trained a classifier to recognize moments of stopping as the animals advanced on the beam. We used 80 videos of 28 adult male and female C57BL6/J mice to train the classifier. We defined stops as a pause in the movement of the animals over several sequential frames during the trial and trained the classifier with such as true positive events, using the remaining sequences - most including walking, missteps, or falls - as true negative events. In sum, the classifier was trained on 13235 frames with 6525 frames showing stops and 6711 frames without such events, requiring a total of ~12 hours of training. We further set the amount of classifier training iterations

to 200 and the cross-validation folds to 6. The minimum bout length was set to two frames (60 ms).

Next, we trained a classifier to identify episodes during which the animals fell from the beam. We used 75 videos of 28 adult male and female C57BL6/J mice to train the classifier. Falls were defined as episodes during which the centre of mass of the animal dropped below the beam. In summary, the classifier was trained on 10118 single frames including 5076 frames with and 5043 frames without falls, requiring a total of ~3 hours of training. Iterations were changed to 150 and the cross-validation fold was increased to 7 for improved accuracy and sensitivity. The minimum bout length was set to 15 frames (500 ms).

Code accessibility

The code, classifiers and supporting files described in the article are freely available online at https://doi.gin.g-node.org/10.12751/g-node.el1sd7/.

Results

All experiments were performed on a standardised balance beam setup (Figure 1) built using readily available materials. We set up a standardised analysis pipeline (Figure 2) aimed at training a JAABA classifier to recognize three common indicators of impaired motor coordination and balance: missteps, stops, and falls, as well as automatically measure the time required to traverse the beam.

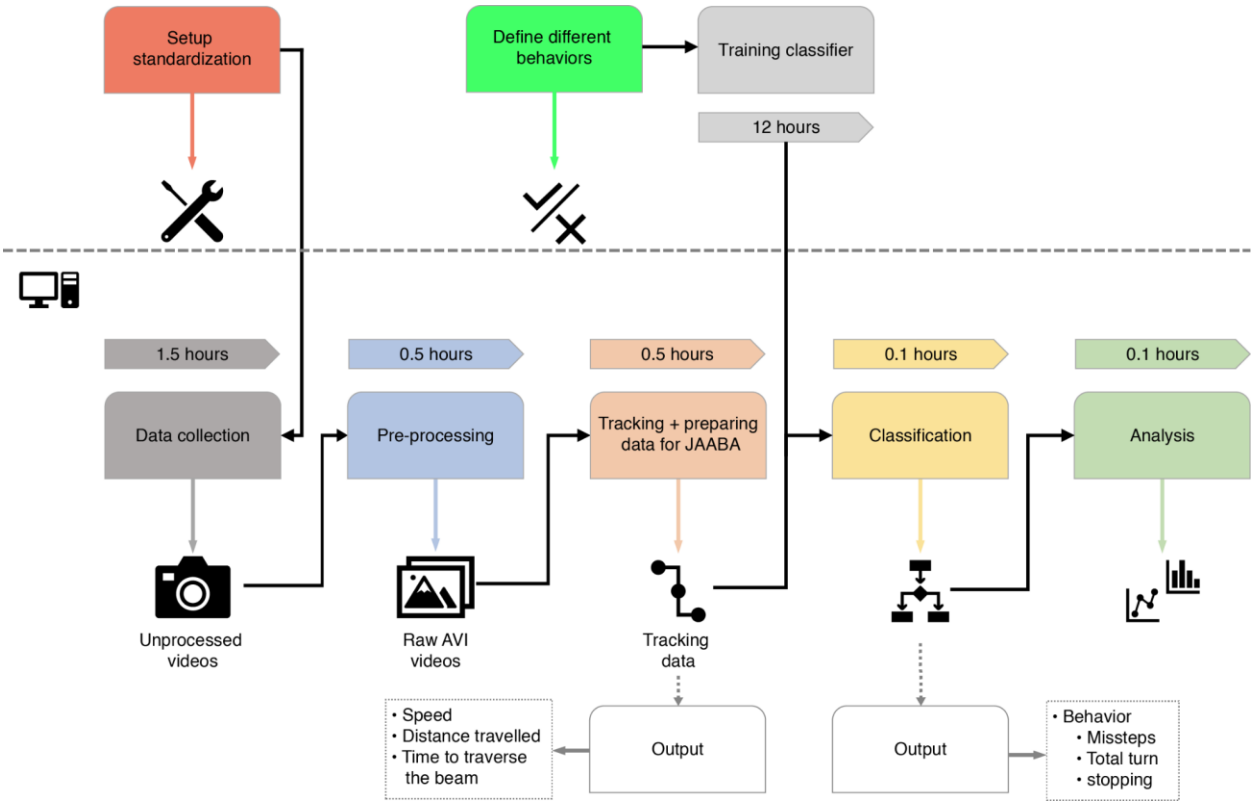

Figure 2. Pipeline for setting up standardised classification of the balance beam test for mice.

The first step in setting up the balance beam test consists of building or acquiring the hardware apparatus, after which initial behavioural data can be collected. These videos have to be pre-processed to convert them for the open-source tracking software (Motr). After the videos have been tracked, they can either be used to train a new behavioural classifier or they can be directly analysed with pre-existing classifiers. Behaviours are then classified using the open-source software JAABA.

Building an analysis pipeline to analyse missteps, stops, falls and speed on the balance beam

We trained three JAABA classifiers to recognize missteps (Figure 3A, left), stops (Figure 3A, middle), and falls (Figure 3A, right), respectively, evaluating their accuracy using cross-validation. Cross-validation quantitatively assesses the accuracy of the classifier based on 1/7 folding by withholding a fraction of the dataset that is used for training of the classifier in order to validate its accuracy (7).

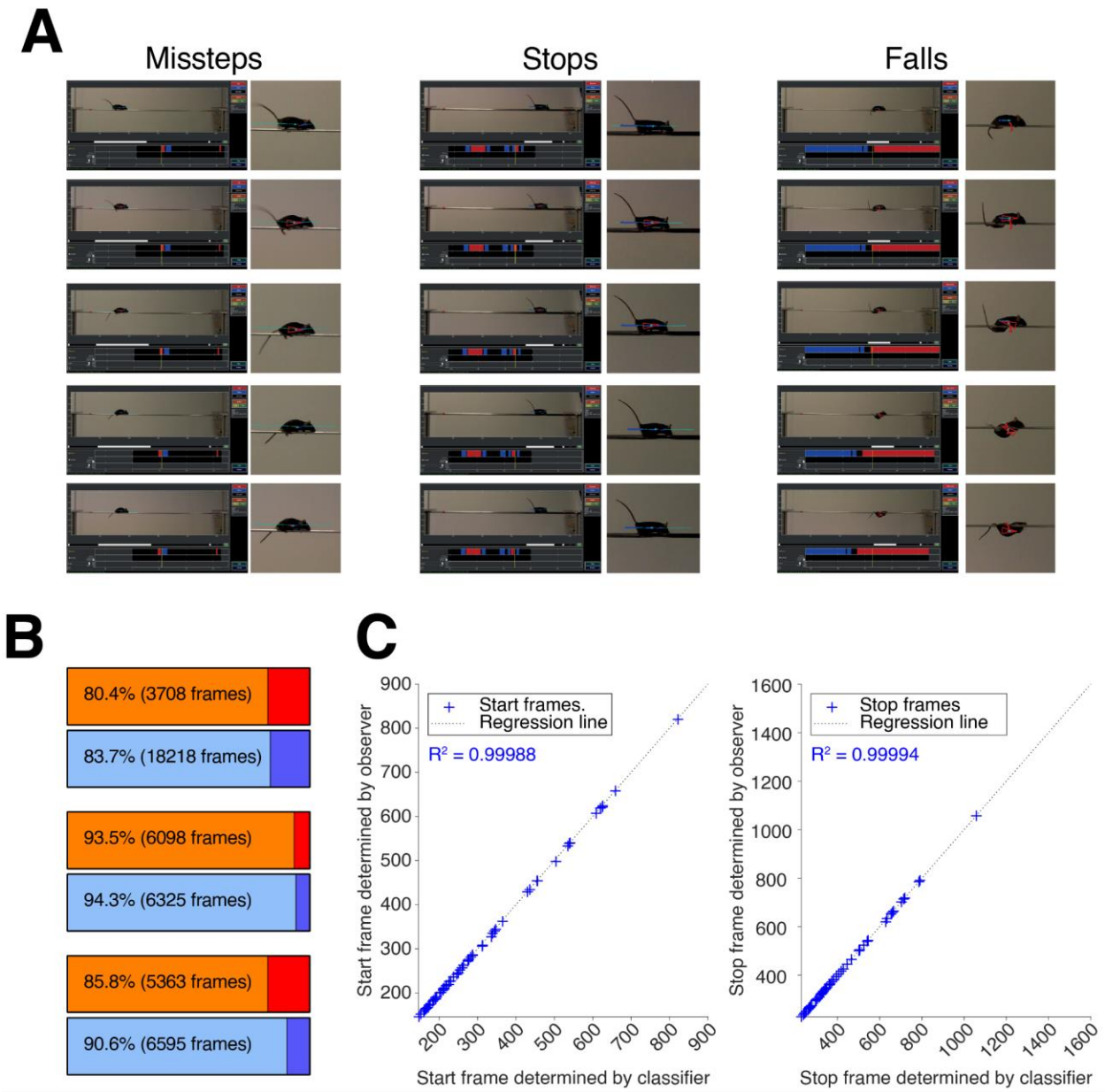

Figure 3. Classifier scoring and accuracy.

(**A**) Sequence of selected frames of the classifier trained for the quantification of missteps in the balance beam test. Shown are the missteps (left), stops (middle) and falls (right). (**B**) Cross-validation results of manually annotated frames for the missteps (top), stops (middle) and falls (lower) classifiers. The orange bar represents correct annotation of frames containing the relevant behaviour, while red represents frames incorrectly scored as the relevant behaviour by the classifier. Light-blue bar represents correct annotation of non-behavior frames, whereas dark-blue shows incorrectly annotated non-behavior frames. (**C**) Manual scoring of start and stop frames in a set of 60 videos revealed highly correlated time to traverse the beam with JAABA, as indicated by Pearson's R square value.

We defined missteps as slipping of the hindpaws that resulted in a downward motion of the hip (Figure 3A, left). Importantly, we noticed that missteps did not always result in a dropping of the paw below the beam but instead frequently presented as a sliding downward motion of the paw on the beam. In both cases, the hip of the animal dropped below its previous height, which was thus chosen as the defining criterion for missteps. The corresponding classifier reached a sensitivity of 80.4% to identify such and a specificity of 83.7% to identify segments without missteps with a minimum length of three frames (90 ms) required for the identification (Figure 3B, top). Stops were defined as pausing of the animal for at least 2 frames (60 ms) before continuing forward locomotion. With regard to stops, the classifier reached a sensitivity of 93.5% to accurately identify those and a specificity of 94.3% to identify segments without stops (Figure 3B, middle).

We defined falls as a downward drop of the entire animal below the beam; in most cases the animals were still able to cling onto the beam and did not immediately fall onto the bottom of the setup. Given that the classifier initially scored multiple falls if the animal dropped below the beam and continued forward locomotion by climbing along the beam, we adjusted the minimum number of frames that were required to score a fall to 15 (500 ms). As a result, the corresponding classifier reached a sensitivity of 85.8% and a specificity of 90.6% (Figure 3B, bottom).

Measuring the time to traverse the beam

In order to determine whether the automated analysis pipeline was able to accurately determine the time needed to traverse the beam, we manually analysed 60 videos of adult male and female C57BL6/J mice. We observed a high correlation between the manual and automated analysis in

determining the start (Figure 3C, left; R^2 = 0.99988) and stop frame (Figure 3C, right; R^2 = 0.99994), respectively.

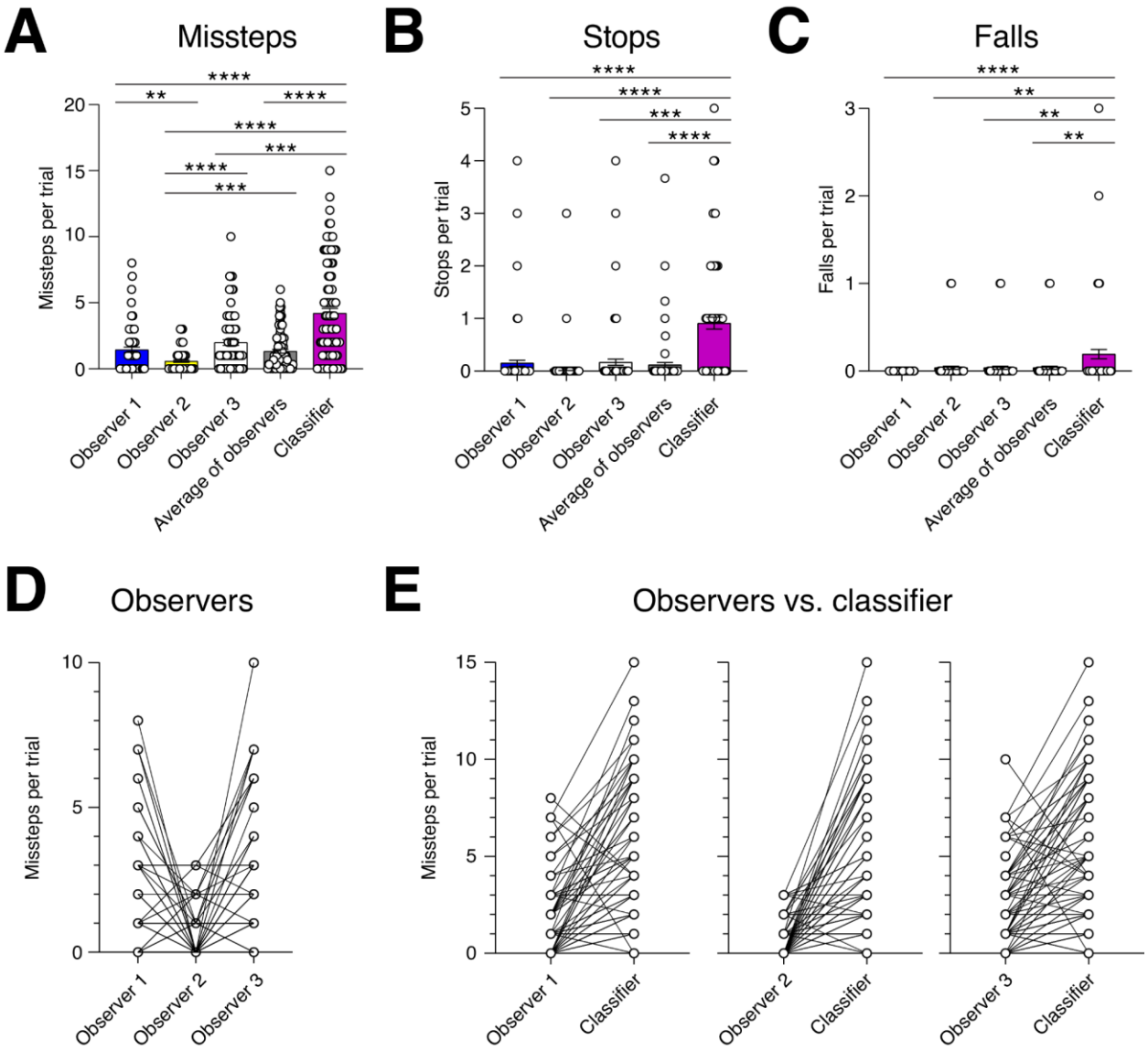

Figure 4. Classifier validation.

(**A**) Number of missteps in 102 videos, manually scored by four observers as well as a trained classifier. (**B**) Number of stops in 102 videos, manually scored by four observers as well as a trained classifier. (**C**) Number of falls in 102 videos, manually scored by four observers as well as a trained classifier. (**D**) Ranking correlation between scoring of the number of missteps per trial by observers. (**E**) Ranking correlation between scoring of the number of missteps per trial by observers and the classifier. Data presented as mean with

SEM. Missteps, stops and falls, Kruskal-Wallis test. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, n = 17 mice.

Comparison of classifiers to observers' scores

We subsequently compared the classifier performance to three observers that were blinded to the genotype of the animals and given an introduction to the underlying principle of the assay. We used 102 videos of male and female *Pcp2-Ppp3r1* adult mice with Purkinje cell (PC)-specific deletion of the PP2B protein, resulting in a lack of parallel fibre-to-PC long-term potentiation (17, 21) and their wildtype littermates for this task. With regard to missteps, we observed substantial variation in scoring by manual observers (Figure 4A). When comparing the identification of stops, the classifier identified more stop events than the three observers (Figure 4B). The number of falls was scored similarly by the observers and the corresponding classifier (Figure 4C). Observers showed large variability in the ranking of mice when scoring missteps per trial (Figure 4D), with overall consistent results with the classifier (Figure 4E).

Automated classification of behaviours and speed using an ataxia mouse model

In order to determine whether the classifiers were capable of reliably identifying differences in motor coordination and balance - as reflected by quantification of missteps, stops, and falls - between ataxic and wildtype mice, we evaluated the performance of the *Pcp2-Ppp3r1* mutant mice and wildtype littermates, with the use of automated classification. Given the severe motor coordination impairments of the mutant animals, we were only able to evaluate their performance on 12 mm and 10 mm beams.

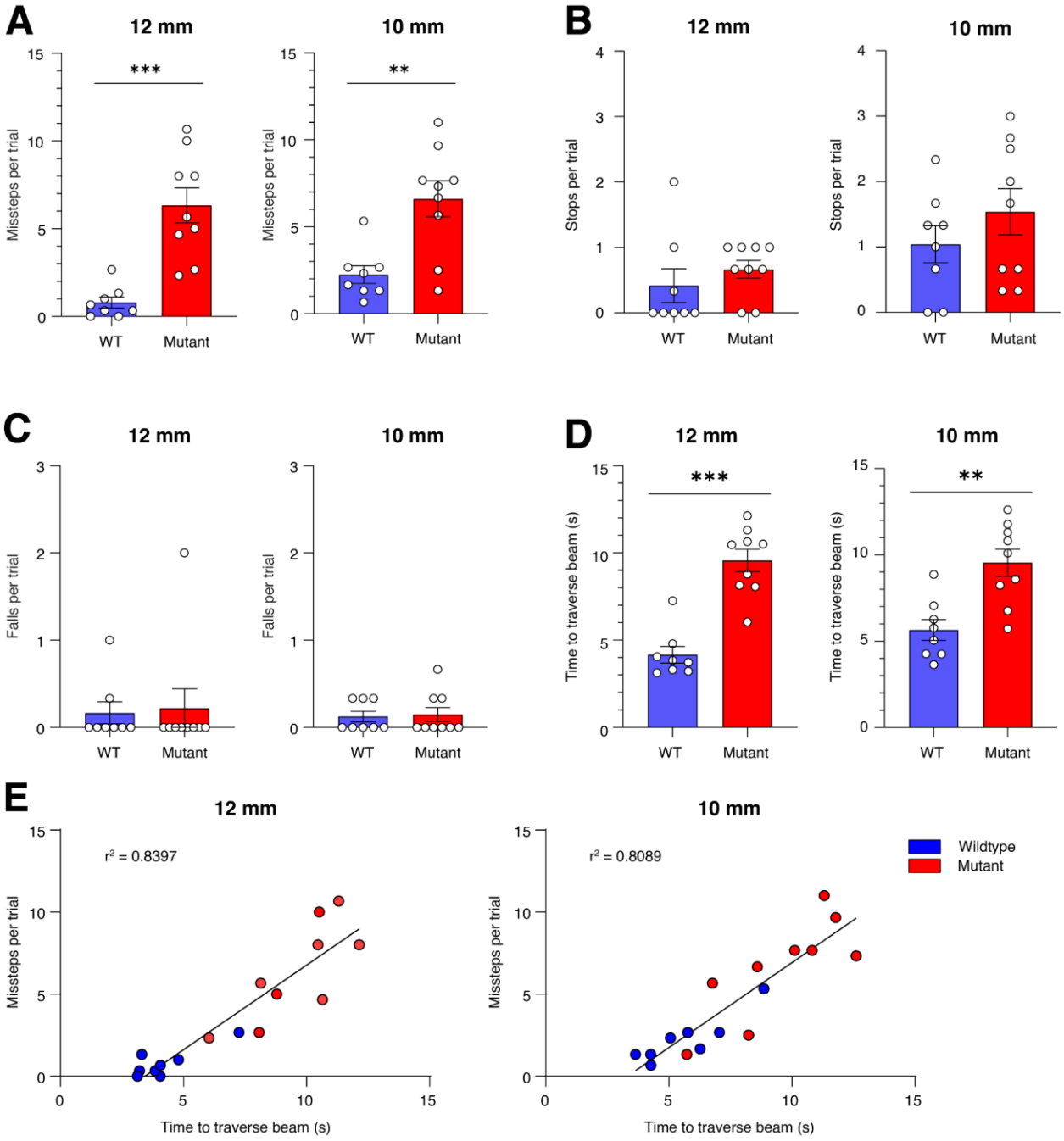

Figure 5. Automated behavioural classification of missteps, stops and falls in the balance beam test.

Automated classification of behaviours are shown for *Pcp2-Ppp3r1* mutant mice (Mutant, red) and wildtype littermates (WT, blue) during the balance beam test). (A) Number of missteps per trial. (B) Number of stops per trial. (C) Number of falls per trial. (D) Time to traverse the beam per trial. (E) Correlation between the missteps per trial and the time needed to traverse the beam. Data presented as mean with SEM, number of missteps (12

mm and 10 mm), two-tailed t test; number of stops (12 mm), Mann-Whitney test; number of stops (10 mm), two-tailed, unpaired t test; number of falls (12 mm and 10 mm), Mann-Whitney test; time to traverse the beam (12 mm), Mann-Whitney test; time to traverse the beam (10 mm), two-tailed t test. *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, n = 8 for WT mice and n = 9 for Mutant mice.

Mutant mice showed a significant increase in missteps per trial on both the 12 mm beam (Figure 5A, left; $p = 0.0002$) and the 10 mm beam (Figure 5A, right; $p = 0.0024$), confirming the ataxic phenotype of PP2B-mutant mice. No significant difference was found when analysing the number of stops per trial between wildtype and mutant animals (Figure 5B; 12 mm beam, $p = 0.3000$; 10 mm beam, $p = 0.2240$), nor the number of falls per trial (Figure 5C; 12 mm beam, $p = 0.7176$; 10 mm beam, $p > 0.9999$). In order to determine whether the animals exhibit altered locomotor patterns on the balance beam, we used our automated analysis pipeline to measure the speed at which animals traverse the beam. Mutant animals took significantly longer to traverse both the 12 mm beam (Figure 5D, left; $p = 0.0002$) and the 10 mm beam (Figure 5D, right; $p = 0.0016$), unveiling another metric that differs between wild-type and mutant animals. The number of missteps scored by the classifier and the time needed to traverse the beam were highly correlated for the 12 mm beam (Figure 5E, left) as well as the 10 mm beam (Figure 5E, right).

Discussion

The analysis of rodent behaviour and motor coordination experiments has been facilitated and accelerated by the introduction of machine learning algorithms and deep neural networks, generating quantitative, reliable, and high-throughput measurements without inter-experimenter variability. The balance beam assay - a well-established and sensitive paradigm for the analysis of motor coordination and balance deficits in rodents - currently lacks such an analysis strategy. Here we describe and validate an automated analysis pipeline that reports the time needed to cross the beam as well as the amount of missteps, stops, and falls.

A feasible and cost-efficient setup and analysis strategy

First, we provide a manual for setup assembly using commercially available material from local hardware stores and a GoPro X HERO7 Black camera (Figure 1A). We estimate the construction time to be approximately 6 h with estimated costs of ~500 euro. Using adult male and female C57BL6/J mice, we proceeded to train three classifiers using MATLAB and open-source software packages. In the first step after data acquisition, we processed videos using FFMPEG and Fiji in order to acquire the time required for crossing the beam and to prepare the videos for the subsequent generation of trajectory data (Figure 2). We compared OptiMouse (22) and motr (2) - two open-source software packages for tracking the animals. In contrast to motr, tracking by OptiMouse was complicated by the frequent misidentification of the nose and tail, which caused the angle of the animal to turn regularly. This was most commonly observed when the tail of the animal escaped identification when positioned at the height of or below the beam. We therefore proceeded to use motr for the generation of the trajectory input for JAABA. Motr provides reliable, high-quality tracing data, but demands substantial processing power.

Fast analyses by automated classifiers with increased inter-experimenter reliability

We trained a classifier to recognize missteps as slipping of the hindpaws with a subsequent downward motion of the hip (Figure 3A, left) and compared its performance to the manual analysis by the three observers. Whereas observer 1 and 3 showed consistent results across genotypes, observer 2 experienced great difficulty in identifying missteps (Figure 3D), which underlines the known problem of inter-experimenter variability in the manual analysis of behavioural and motor coordination experiments with rodents.

When testing the classifier on videos of an established cerebellar ataxia mouse model (17), automated analysis showed significant increases in missteps by mutant animals on both the 12 mm and 10 mm beams compared to wildtype animals (Figure 5A), confirming the ataxic phenotype of *Pcp2-Ppp3r1* mice on the balance beam.

We subsequently trained a second classifier to recognize stops on the beam (Figure 3A, middle). The classifier identified more stops in the trials of mice on the balance beam. We have set the minimum number of frames per stop to two, thereby also defining short moments of pausing as stops. These short bouts may have been difficult to discern for the observers given their short nature, particularly in animals that quickly advance on the beam and mutant animals that exhibit an ataxic and thus arrhythmic gait. No significant differences in the number of stops between wildtype and mutant animals were found (Figure 5B).

Furthermore, we trained a classifier to identify falls (Figure 3A, right). Both the observers and the classifier identified falls, showing a similar level of performance (Figure 3F). There were no discernable differences in the number of falls between wildtype and mutant animals (Figure 5C).

Last, our pipeline also measures the time needed for animals to traverse the beam. The pipeline provided fast and consistent output that showed no discernable difference in start and stop times

between three independent observers and the automated pipeline (Figure 3C). Mutant animals required significantly more time to traverse both the 12 mm and 10 mm beams (Figure 5D), further confirming the ataxic phenotype of the cerebellar *Pcp2-Ppp3r1* mouse model. Additionally, animals showed a strong positive correlation in the number of missteps per trial, scored by the classifier, and the time needed for animals to traverse the beams (Figure 5E), showing that the reduced speeds are due to the ataxia phenotype and not a coping mechanism to reduce errors on the beam.

In sum, both the classifier and the observers were able to discern substantial motor coordination and balance deficits in mice by scoring the number of missteps, stops, falls, and the time needed to traverse the beam. Machine learning and deep neural networks have greatly accelerated the analysis of behavioural and motor coordination experiments. Our pipeline with three automated classifiers is capable of precisely measuring the time required to traverse the beam and discern differences in the amount of missteps, stops, and falls between wildtype animals and a Purkinje cell-specific knock-out of PP2B - a known ataxic mouse model. It provides a fast and reliable alternative to the manual time-consuming, error-prone, and variable analysis by observers, relies on open-source software, and can be easily applied in other experimental setups.

Future perspective

This pipeline can reliably report motor function in mice, thereby offering a cost-effective and high-throughput alternative to manual scoring of behaviour. This will enable researchers to increase inter-experimenter and intra-experimenter reliability for the analysis of balance beam data.

Executive summary

Background: We developed a cost-effective and high-throughput alternative to manual scoring of motor function on the balance beam.

Methods: Using an established ataxia model, we validated the open-source analysis pipeline by comparing its output to three independent observers and quantifying the differences in motor function between wildtype and mutant animals on the balance beam.

Results:Pcp2-Ppp3r1-mutant animals showed a significant increase in the number of missteps and increased time needed to traverse the beam.

Conclusion: We show that the pipeline can reliably report crossing time, missteps, stops, and falls, thereby offering a cost-effective and high-throughput option with increased inter-experimenter and intra-experimenter reliability for the analysis of balance beam data.

Acknowledgements

The authors are grateful to Maaike Imthorn for assistance with animal experiments.

Conflict of Interest

Authors report no conflict of interest.

Funding sources

This work was supported by Netherlands Organization for Scientific Research (NWO) grants VIDI/917.18.380,2018/ZonMw (A.B.) and 016.Veni.192.270/NWO (J.J.W) and Erasmus MC Fellowship (J.J.W).

References

1. **Tecott, L.H., and E.J. Nestler**. 2004. Neurobehavioral assessment in the information age. Nat. Neurosci. *7*:462–466.

2. **Ohayon, S., O. Avni, A.L. Taylor, P. Perona, and S.E. Roian Egnor**. 2013. Automated multi-day tracking of marked mice for the analysis of social behaviour. J. Neurosci. Methods *219*:10–19.

3. **de Chaumont, F., R.D.-S. Coura, P. Serreau, A. Cressant, J. Chabout, S. Granon, and J.-C. Olivo-Marin**. 2012. Computerized video analysis of social interactions in mice. Nat. Methods *9*:410–417.

4. **Spencer, C.M., D.F. Graham, L.A. Yuva-Paylor, D.L. Nelson, and R. Paylor**. 2008. Social behavior in Fmr1 knockout mice carrying a human FMR1 transgene. Behav. Neurosci. *122*:710–715.

5. **Wahl, L., A.M. Punt, T. Arbab, I. Willuhn, Y. Elgersma, and A. Badura**. 2022. A Novel Automated Approach for Improving Standardization of the Marble Burying Test Enables Quantification of Burying Bouts and Activity Characteristics. eNeuro *9*.

6. **Bateson, M., and P. Martin**. 2021. Measuring Behaviour: An Introductory Guide (Cambridge University Press).

7. **Kabra, M., A.A. Robie, M. Rivera-Alba, S. Branson, and K. Branson**. 2013. JAABA: interactive machine learning for automatic annotation of animal behavior. Nat. Methods *10*:64–67.

8. **Nilsson, S.R.O., N.L. Goodwin, J.J. Choong, S. Hwang, H.R. Wright, Z.C. Norville, X. Tong, D. Lin, et al.** 2020. Simple Behavioral Analysis (SimBA) – an open source toolkit for computer classification of complex social behaviors in experimental animals.

9. **Jhuang, H., E. Garrote, J. Mutch, X. Yu, V. Khilnani, T. Poggio, A.D. Steele, and T. Serre**. 2010. Automated home-cage behavioural phenotyping of mice. Nat. Commun. *1*:68.

10. **Hong, W., A. Kennedy, X.P. Burgos-Artizzu, M. Zelikowsky, S.G. Navonne, P. Perona, and D.J. Anderson**. 2015. Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. Proc. Natl. Acad. Sci. U. S. A. *112*:E5351–E5360.

11. **Wiltschko, A.B., M.J. Johnson, G. Iurilli, R.E. Peterson, J.M. Katon, S.L. Pashkovski, V.E. Abraira, R.P. Adams, and S.R. Datta**. 2015. Mapping Sub-Second Structure in Mouse Behavior. Neuron *88*:1121–1135.

12. **Unger, J., M. Mansour, M. Kopaczka, N. Gronloh, M. Spehr, and D. Merhof**. 2017. An unsupervised learning approach for tracking mice in an enclosed area. BMC Bioinformatics *18*:272.

13. **Arakawa, T., A. Tanave, A. Takahashi, S. Kakihara, T. Koide, and T. Tsuchiya**. 2017. Automated Estimation of Mouse Social Behaviors Based on a Hidden Markov Model. Methods Mol. Biol. *1552*:185–197.

14. **Luong, T.N., H.J. Carlisle, A. Southwell, and P.H. Patterson**. 2011. Assessment of motor balance and coordination in mice using the balance beam. J. Vis. Exp.

15. **Stanley, J.L., R.J. Lincoln, T.A. Brown, L.M. McDonald, G.R. Dawson, and D.S. Reynolds**. 2005. The mouse beam walking assay offers improved sensitivity over the mouse rotarod in determining motor coordination deficits induced by benzodiazepines. J. Psychopharmacol. *19*:221–227.

16. **Lang, J., E. Haas, J. Hubener-Schmid, C.J. Anderson, S.M. Pulst, M.A. Giese, and W. Ilg**. 2020. Detecting and Quantifying Ataxia-Related Motor Impairments in Rodents Using Markerless Motion Tracking With Deep Neural Networks. Conf. Proc. IEEE Eng. Med. Biol. Soc. *2020*:3642–3648.

17. **Schonewille, M., A. Belmeguenai, S.K. Koekkoek, S.H. Houtman, H.J. Boele, B.J. van Beugen, Z. Gao, A. Badura, et al.** 2010. Purkinje cell-specific knockout of the protein phosphatase PP2B impairs potentiation and cerebellar motor learning. Neuron *67*:618–628.

18. **Zeng, H., S. Chattarji, M. Barbarosie, L. Rondi-Reig, B.D. Philpot, T. Miyakawa, M.F. Bear, and S. Tonegawa**. 2001. Forebrain-specific calcineurin knockout selectively impairs bidirectional synaptic plasticity and working/episodic-like memory. Cell *107*:617–629.

19. **Barski, J.J., K. Dethleffsen, and M. Meyer**. 2000. Cre recombinase expression in cerebellar Purkinje cells. Genesis *28*:93–98.

20. **Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, et al.** 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*:676–682.

21. **Romano, V., L. De Propris, L.W. Bosman, P. Warnaar, M.M. Ten Brinke, S. Lindeman, C. Ju, A. Velauthapillai, et al.** 2018. Potentiation of cerebellar Purkinje cells facilitates whisker reflex adaptation through increased simple spike activity. Elife *7*.

22. **Ben-Shaul, Y.** 2017. OptiMouse: a comprehensive open source program for reliable detection and analysis of mouse body and nose positions. BMC Biol. *15*:41.