REVISED Manuscript

Click here to access/download;REVISED Manuscript;Lim et al_JNM submission_revisions_CL_GR_Clean.docx

Click here to view linked References

Comparison of automated video tracking systems in the open field test: ANY-Maze versus Ethovision XT.

Charmaine J.M. Lim[1], Bettina Platt[1], Sanna K. Janhunen[2], Gernot Riedel[1]*

[1]Institute of Medical Sciences, University of Aberdeen, Aberdeen, United Kingdom

[2]Forendo Pharma, Itäinen Pitkäkatu 4, 20520 Turku, Finland sannajanhunen@organon.com

Keywords: Video observation; animal tracking; mouse; light intensity

*Corresponding author:

Gernot Riedel

Institute of Medical Sciences

Foresterhill

University of Aberdeen

Aberdeen. AB25 2ZD

United Kingdom.

Email: g.riedel@abdn.ac.uk

Abstract

Background: ANY-Maze and EthoVision XT are two commonly used automated animal tracking systems employed to produce reliable and consistent results in behavioural paradigms. Data obtained with both tracking systems have presented differences, particularly when varying laboratory lighting conditions and contrasts of mice coat colour against the arena background in both water maze and tunnel maze.

Method: In this study, two fluorescent lighting conditions (58 and 295 lux), local to our laboratory, and different coat-coloured mouse lines (C57BL/6J - black; CD1 - agouti; C3H/HeN - white) were used to compare reproducibility in measures of tracking systems (ANY-Maze versus EthoVision) in the open field test.

Results: Differences between systems were reliant on the contrasts between coat and background colours. Surprisingly, black animals presented the greatest differences in read-outs between tracking systems, regardless of lighting conditions. Data from both video observation tools differed mainly in exploration-related parameters (distance travelled), but less in more static proxies (time in thigmotaxis zone). Overall, EthoVision XT return higher values for most parameters analysed relative to ANY-Maze. More inconsistencies in recording and analysis can be expected from other video recording systems.

Conclusion: Data analysis software provides an additional source of variation in need of consideration when reproducibility in behavioural neuroscience is required.

Insert graphical abstract here.

---

1. Introduction

Automated tracking of freely moving animals in behavioural paradigms offers numerous advantages to enhance efficiency, reliability, and consistency with minimal human interference (Noldus et al., 2001; Spruijt et al., 2014). This is particularly advantageous in quantifying complex behavioural responses (such as full or partial rotations of 360° completed by the animal's body; the angle between two consecutive vector movements of the animal; grooming; etc.) in addition to the typical parameters of distance travelled or time spent in specific zones. ANY-Maze and EthoVision XT, the most widely used automated video observation systems in academia and industry, track by contrasting the animal image against the background (Bailoo et al., 2010; Spink et al., 2001). Either a digital camera is used (as in this study), or an analogue image from the camera is first digitized by a frame grabber and examined as a series of frames, each comprising of a grid of pixels analysed frame-by-frame to differentiate tracked objects from the background. The contrast between animal and background can be influenced by the user depending on the study paradigm and subjects. For example, ANY-Maze allows the user to alter the sensitivity of animal detection against the arena background. EthoVision XT additionally provides tracking options such as grey-scaling or dynamic subtraction and detection sensitivity. Each video software, however, is based on its own algorithms, idiosyncratic to the manufacturer, for quantifications of the animal's behaviour (Noldus et al., 2001; Spink et al., 2001).

The threshold for contrast differences can be set manually by the user or automatically by the programme but is still extremely dependent on the lighting conditions and animal coat colour. Local lighting conditions are known to influence the contrast of coat colours with the arena background during tracking affecting the obtained results (Bailoo et al., 2010). In the case of low animal and background contrast, the intensity and type of illumination will impact the tracking systems. Uneven or dispersed illumination as well as reflections from surfaces of the apparatus generally cause tracking of shadows or noise in tracks, resulting in inaccurate path length measurements (Bailoo et al., 2010; Spink et al., 2001). The deviations in path lengths obtained by the tracking systems from the actual path taken have been obtained using inanimate and motor-driven such as discs (Bailoo et al., 2010; Lind et al., 2005). Under low contrast fluorescent lighting at 563 lux, the path lengths of motor-driven rotating inanimate disks were overestimated in ANY-Maze / underestimated in EthoVision XT for both water maze and open field tests, but otherwise presented relatively similar results for discs of other contrasts (Bailoo et al., 2010). Animals, however, often demonstrate unpredictable movements which cannot be represented accurately by inanimate objects. Lighting and contrast conditions may exacerbate the inaccuracy of recording unpredictable/non-systematic movements and sharp turns of displacement in animals.

The reliability of ANY-Maze and EthoVision XT will be assessed using mice in the open field test in this study. The open field test remains to be a popular sensorimotor paradigm which provides a simple and rapid method of assessing different motor activity levels, exploration habits, and anxiety traits in rodents (Denenberg, 1969; Robinson et al., 2018; Spruijt et al., 2014; Wahlsten et al., 2006). The simplicity of the apparatus in the open field test makes testing cost-effective and requires minimum to no experimenter expertise for the administration of the test, nor training of the test subject. As the open field test produces sufficiently reliable and repeatable measures on a range of independent variables, animal behaviour is generally observed first in the open field test before other behavioural assays. Clearly defined behaviours affected by genetic, physiological, and pharmacological manipulations are likely to be related to locomotion and/or motor activities which the open field test is sensitive to background of the animal (Robinson et al., 2018; Spruijt et al., 2014).

In this exploratory study, we used the open field at two different fluorescent light conditions (58 versus 295 lux) and fed the recorded videos into i) ANY-Maze 6.3 and ii) EthoVision XT 11.5 and analysed the animal's behaviour with a focus on activity and anxiety-related parameters. We compared three different mouse lines with black (C57BL/6J), agouti (C3H/HeN) and white (CD1) coat colour. Our data show that despite identical tracks from the two video systems, EthoVision XT appears to overestimate the distance parameters particularly at 58 lux relative to ANY-Maze (or that Any-Maze appears to underestimate relative to EthoVision XT). No differences in time in zone and ratios were observed.

2. Materials and Methods

2.1 Animals

Animals in this study were well handled and previously used in other non-invasive behavioural paradigms (not open field) and were approximately 15 weeks old at the start of testing in the open field. Ten C57BL/6J males (termed C57 from here on; registered in-house under the study plan 2019_004 in relation to a Home Office Project Licence), 10 Crl:CD1 (ICR), and 10 C3H/HeNCrl all females (registered in-house under the study plan R0165; and annotated as C3H and CD1 respectively from here on) from Charles River Laboratories (Margate, Kent, UK) were used in this experiment. They were chosen to determine the effect of different coat colour on video recording outcomes as they have black, agouti and white furs respectively. Comparisons between sexes and strains of animals and large sample sizes of animals per group were deemed of less importance for any comparisons between the analysis software, but animals in this study were of similar weight and sizes for outcomes to be comparable.

The animals were group-housed (approximately 6-8 per group) in stock cages (Techniplast 1292N), measuring 45 x 38 x 13 cm under a 12:12 hours light:dark cycle (lights on at 0700), with an average ambient temperature of 21 ± 2^{o}C and humidity of 50 ± 5%. Animals had free access to standard rodent food chow and water *ad-libitum* and were provided with clean bedding (corn cob and wood shavings) once per week. All animals were acclimatized to the facility environment for approximately 5 weeks prior to open field testing. Test subjects were handled by tail and scooping methods during the removal from their home cage and transfer to the open field arena. All housing and handling of animals were in accordance with international standards on animal welfare regulated by European Communities Council Directive 63/2010/EU and the UK Animals Scientific Procedures Act (1986). The experiments followed the study design, analysis and reporting methods recommended in the ARRIVE 2.0 guidelines and are detailed in the relevant segments below.

2.2 Apparatus, lighting conditions and testing procedures

Evaluation of general locomotor activity was conducted in a non-blinded manner due to the nature of the coat colours of the subjects and only one experimenter was involved throughout the tests and offline analysis. Animals of the same strain were allocated random identification numbers and the test sequence of subjects was randomized using the Williams Square Design (Wang et al., 2009). All animals within each strain were randomly assigned to an illumination group (5-6 within each genotype per illumination group).

The apparatus comprised of a square box (made up of white reflective Perspex material), measuring at 50 x 50 x 40 (height) cm. The arena floor was modified with black non-reflective material during testing with CD1 mice to contrast between the white coat colour of CD1 and the normally white floor of the arena. For comparison, bright (295 lux) and dim (58 lux) lighting conditions as frequently used in our laboratory were selected. Overhead fluorescent lights were used producing 58 lux in the arena; overhead fluorescent lights and 2 white wall LED lights facing upwards from the arena were used for brighter lighting conditions at 295 lux.

The open field tests were conducted in a dedicated sound-attenuated room, with the temperature and humidity maintained at 22 ± 2ºC and 50 ± 5% respectively. Animals were allowed to habituate to the room for approximately 30 minutes prior to testing. All animals were tested in the open field in a single day, within 8 hours from the start of the light phase. Each trial lasted 10 minutes. The mouse was placed in the centre of the arena to initiate the start of the test. Only one mouse was tested at any one time and the apparatus was thoroughly cleaned with odourless and alcohol-free wet wipes between each trial/animal.

2.3 Methods of automated tracking

An overhead camera (Imaging Source, DMK22AUCO3) was positioned 125 cm from the floor of the arena and all tests were recorded as MPEG4/h.264 (producing 30 frames per second) files on an adjacent computer. For calibration of arena alignments in both video recording programmes, images were taken from a test video and divided with zone boundaries (Figure 1) used prior to re-tracking for both tracking systems. All videos were tracked on EthoVision XT (Version 11.5, Noldus Information Technology, Wageningen, The Netherlands) and ANY-Maze (Version 6.3, Stoelting Co.) but were analysed offline (EthoVision XT (Version 14, ANY-Maze Version 6.3). Track-plots of the centre of gravity for each mouse line and illumination were obtained from both video software to reveal the exploration paths in the arena (Figure 2).

The automatic tracking option was defaulted in ANY-Maze, which is without means to manually adjust for animal size but for animal coat and apparatus background colours, allowing for the program to adjust tracking parameters between the environment and illumination. Detection settings for EthoVision XT were manually adjusted according to arena floor contrast, animal coat colour and animal size. Percentages of samples in which subjects were not found and the percentage of samples rejected, met an acceptable criterion of no more than 5%, according to the EthoVision XT manual. Sampling rates for re-tracking were maintained at 30 frames per second (fps) in both systems.

Initially, all parameters offered from each recording system were analysed and compared between study cohorts and converted into heatmaps. These heatmaps compared categories of parameters related to i) apparatus, ii) thigmotaxis and iii) centre point listing a total of 85 parameters for ANY-Maze and 161 for EthoVision XT. The most frequently reported proxies were then selected and used for (i) a comparison between mouse lines and (ii) a comparison between video analysis systems within each mouse line. Included in these proxies were: total distance travelled in the arena (cm); frequency of total rotations; thigmotaxic response (outer perimeter of 5cm width: time, distance, ratio), and average distance from the centre point (cm). The periphery measured the outer edge of the arena with 5 cm away from its walls.

Insert Figure 1 here

2.4 Statistical analysis

Exploratory data analysis using heat-maps was applied to reveal all differences between mouse line for all parameters that can be reasonably obtained in both ANY-Maze and EthoVision XT. Data presented in heat-maps denote p-values with a threshold of 0.05 (dark blue representing $p_{threshold}$ <0.01) obtained from Kolmogorov-Smirnov two-sample comparisons. Parameters in the heat-maps were first clustered for within-system (Figure 3) and between-system (Figure 3) comparisons for clarity (Gehlenborg & Wong, 2012; Timotius et al., 2019). Apparatus measures concerning total distance and other additional information regarding activity for the entirety of the test were categorized at the top of the map. Centre point and thigmotaxic zone measures which are generally used as standard parameters were clustered next. Heat-maps represent global read-outs over 10 minutes; time-dependent differences and scoring by segment of test were omitted. For simplicity, parameters from head and tail tracking were not considered.

For all comparisons between mouse lines and illumination recorded with different observation software, estimation analysis was conducted using R (Dabestr package, v. 0.3.0, Ho et al., 2018). Re-sampling at 10000 replacements and seed starting at 123456 was applied for estimation analysis. Data were visualized as Cummings estimation plots– the mean and standard deviation of each group is plotted as a gapped line next to the swarmplots. The mean of the null is the difference-axis origin (aligned with the mean of the test group), and unpaired mean differences were plotted with a shaded curve indicating the distribution of sampling error

for the difference between the means. Error bars on the difference axis depict the bootstrapped 95% confidence interval for differences between means. C57Bl6/J mice and ANY-Maze were selected in this study as the reference groups used to calculate mean differences in the within-system and between-system analyses, respectively.

For conventional statistical analysis, data were pooled and tested for normality which revealed the data to be skewed. This is not surprising given the small sample sizes (n=5-6/group – no power calculation performed a priori). Hence all differences between tracking sensitivity for each mouse line were averaged and contrasted using Kruskal-Wallis H tests for strain comparisons. These differences were further analysed with Bonferroni-corrected Mann-Whitney tests for comparisons against the reference group (C57Bl6/J in this study) where significant findings were observed in Kruskal-Wallis H tests. Mann-Whitney U tests were used in between-software comparisons for each strain. All conventional analyses were performed with statistical differences set at 95% confidence levels and are reported in detail in figures only. No outliers (with residuals more than two standard deviations away from the mean defined a priori) were detected. Correlations between tracking software, independent of coat colour and illumination, were performed with parametric Pearson's correlations (variables presented normality in Kolmogorov-Smirnov tests).

Estimation statistics, heatmaps and graphs were performed in R (v.1.2.5033, R Core Team (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria). All tests of normality and conventional statistics were performed in SPSS (IBM SPSS Statistics v.25.0. Armonk, NY: IBM Corp).

3. Results

3.1 Heatmap comparisons differ between Any-Maze and EthoVision despite identical track plots

Representative exploration paths for each mouse line with different coat colour (black, agouti, white) and light intensity (58 lux and 295 lux) were obtained via each tracking system and are presented in Figure 2. Differences in locomotor activity and spatial distribution occurred between mouse strains and light intensities despite feeding the same video input into both video tracking systems.

Differential comparison of proxies between CD1 (white) and C3H (agouti) against C57BL6/J (black) mice for each lighting condition was converted in a colour-coded heatmap. An initial total of 85 and 161 parameters were obtained for ANY-Maze and EthoVision XT respectively (Figure 3). A direct comparison of 28 parameters compatible in both recording programmes was performed and is depicted in Figure 3 and corresponding values are provided in Table 1. Regardless of lighting conditions, C57BL6/J mice consistently presented with the greatest number of differences (58 lux: 13 parameters and 295 lux: 12 parameters), followed by CD1 (58 lux: 9 parameters and 295 lux: 7 parameters) and C3H mice (58 lux: 5 parameters and 295 lux: 6 parameters). Amongst these parameters, differences ($p < 0.05$) common to all experimental groups were found in measures related to the thigmotaxic zone (Table 1; highlighted in blue): *absolute turn angle*, *thigmotaxic zone: maximum speed*, *thigmotaxic zone: minimum distance from*, and *thigmotaxic zone: absolute turn angle*. These differences in thigmotaxis are important parameters in the evaluation of the subject's emotionality and anxiety states.

Figures 2 & 4 and Table 1 here

3.2 Differences between mouse strains and lighting intensities affect tracking endpoints

A complete statistical summary of findings (visualised in Figure 5, 6) within each video analysis system can be found in Table 1. Mouse strain and lighting intensity affected tracking read-outs in both video analysis systems. Under dim light (58 lux), C3H mice presented with a lower *total distance travelled* relative to C57BL6/J mice in both tracking systems (Figure 5A); yet data were only significant for EthoVision XT recordings. The same profile was observed for

the distance moved in the thigmotaxic zone (Figure 5C). By contrast, ambulation was analysed differentially for CD1 mice with ANY-Maze returning higher values compared with C57BL6/J mice while EthoVision XT reported lower distances moved (Figure 5A). These differences appeared despite seemingly identical tracks extracted from the video files. In terms of overall activity, ANY-Maze reported a grading from highest to lowest of CD1 > C57BL6/J > C3H, whereas EthoVision XT reported C57BL6 > CD1 > C3H.

Average distance from the centre point (Figure 5B), *time in the thigmotaxic zone* (Figure 6A) and the derived *thigmotaxic ratios* (Figure 6B) were equal between C57BL6/J and C3H strains but lower for CD1 mice. This was similarly observed in both video-analysis tools. As for the total number of *rotations,* again there was no difference between C57BL6/J and C3H mice, but CD1 showed heightened number of rotations in both observation tools (Figure 6C; significantly different from C57BL6/J only in ANY-Maze). Overall, group-wise statistical comparison was more sensitive for data derived from ANY-Maze (Figure 5C, 6C).

Higher consistency was obtained when the illumination was increased to 295 lux. For both *total distance* and *distance in the thigmotaxis zone*, ANY-Maze and EthoVision XT reported greatest values for CD1 (Figure 5A, C). For the parameters *of distance from the centre*, both video tools returned significantly higher values for C3H mice relative to C57BL6/J but a smaller difference for the CD1 relative to C57Bl6/J were obtained (Figure 5B). When compared with C57BL6/J, similar outcomes were observed for *thigmotaxis ratio* (Figure 6B) with slightly elevated ratios in C3H mice, but little difference was reported for the CD1 strain. *Time spent in the thigmotaxis zone* was significantly lower in C57Bl6/J mice compared to C3H or CD1 mice in both tracking software, but patterns and outcomes of thigmotaxis time were similar for both software. Intriguingly, the *number of rotations* differed between video analysis tools, but demonstrate similar patterns. While ANY-Maze reported higher levels for C3H and CD1 strains relative to C57BL6/J, EthoVision XT reported slight elevations. All these variations did not reach significance.

Figures 5 & 6 and Table 2 here

3.3 Differences between video recording and analysis software for each mouse strain/coat colour

A reorganization of the data was undertaken to enable a direct comparison between video-observation systems for each mouse coat colour (depicted in Figures 7, 8 and Table 3). On

the same proxies, numerous software-related differences were identified, for which EthoVision XT typically provided higher readouts than ANY-Maze. At both light intensities, these differences were exclusive to measures of path length (*total distance moved*, Figure 7A, and *distance moved in thigmotaxis zone*, Figure 7C) and while observed for all mouse strain/coat colour, significances were observed for black coat colour in C57BL6/J mice only. Again, this overestimation of EthoVision XT / underestimation of ANY-Maze is surprising given that the tracks detected by both software systems were similar.

Identical values, however, were reported for the following parameters: *distance from centre* (Figure 7B); *time in thigmotaxis zone* (Figure 8A) and *thigmotaxis ratio* (Figure 8C). There was no effect of recording software on these data for any mouse strain/coat colour and were independent of light intensity. Finally, EthoVision XT presented a smaller *number of rotations* in all mouse strain/coat colour when recordings were conducted at 295 lux (Figure 8C), but similar number of rotations (except in CD1 mice) at 58 lux.

To address the issue of precision of tracking between the two video observation software packages measures of distance (*total distance travelled*, Figure 9A; *distance from centre point*, Figure 9B; *distance in thigmotaxis*, Figure 9C) were further analysed by Pearson correlations comparing the data for all subjects between both video systems at high and low intensity illumination. All correlations were positive and significant. Overall, correlations were close to R=1 for 295 lux, but R-coefficient was lower for 58 lux, particularly, in *total distance moved* and *distance moved in the thigmotaxis zone* (Figure 9A, C). For the latter, values for data from EthoVision XT were much higher than from ANY-Maze (see also Figure 7 and Table 3). Reasons for these light-intensity dependent differences remain elusive and data strongly suggest that work under brighter light intensities may achieve comparable values for both video software packages independent of coat colour.

Figures 7, 8, 9 and Table 3 here

---

4. Discussion

Here, two commercial video observation systems i.e., ANY-Maze and EthoVision XT, were used for tracking rodent movement patterns. They allow the elimination of subjective and labour-intensive manual scoring. The tracking systems require contrast between the subject, arena background and adequate lighting for comparable and reliable tracking of the subjects. ANY-Maze appeared to consistently present smaller read-outs (or EthoVision XT presented larger read-outs) for all measures of distance, and this difference between systems is exacerbated in low illuminations. Brighter light intensities of 295 lux, however, improved robustness of findings, presenting near identical values in both video recording systems, but it must be noted that higher illumination will affect certain behaviours, particularly anxiety, in animals. At low light intensities, differences in primary read-outs of activity (distance travelled) may be explained by increased dispersion of shadows observed, and the resultant low contrast between subjects and background. Tracking inaccuracies may be further exacerbated when this is coupled with sharper turns of angle displacement (particularly in the thigmotaxic zone) and non-systematic movements. The use of infrared backlight as a solution (such as in Bailoo and colleagues' work, 2010) would provide a high contrast and minimise the effects of light intensities and complexities of the arena on the tracking precision in future studies.

It is however, it was not in the remit of this work to determine which recording software presents the most precise measurements with the data presented here. The true path length could be determined by manually entering the position of the mouse in a frame-wise analysis and summing up the distances with subsequent calculations and independent statistics. Independent of this ground truth, we assume that testing of additional video systems including for instance VideoMot (TSE systems: https://tinateb.com/wp-content/uploads/2016/06/tse_videomot2.pdf), Smart video tracking (Panlab: http://www.panlab.com/en/products/smart-video-tracking-software-panlab), VideoTrack (Viewpoint, https://www.viewpoint.fr/en/p/software/videotrack), or open source video tracking systems (Zhang et al, 2020; Krynitskyet al., 2020) may increase variability and incoherence. If video tracking systems apply different algorithms for data input, one would expect deviations in raw values for all parameters under scrutiny and for all lighting intensities. Given that only some parameters differed significantly between EthoVison XT and ANY-Maze, and that brighter arena conditions increased similarity between data, makes differences at the front end unlikely. It rather suggests variation in algorithms for extraction and analysis of the different parameters, particularly those impinging on measurements of movement and exploration. These include *meandering* as an automatic endpoint in EthoVision XT, but it needs to be manually calculated for ANY-Maze (*absolute turn angle* against *total distance*

travelled) or *rotation frequency* and vertical activity as noticed in our heat-map analysis. It is consequently not surprising that differences between mouse strains occurred between tracking software, and this is not only due to the greater number of parameters analysed by EthoVision XT, or by the variation in activity between mouse lines.

With relevance to coat colours, black coat mice presented significant differences in distance moved between software comparisons, in contrast to the agouti and white coat colours. This could potentially be the result of light that may have reflected of the walls resulting in indirect and diffused lighting. This will in turn cause more prominent shadows which may not be distinguished from black coated mice, hence resulting in minute tracking errors. This would not be an issue with lighter fur, for example mice with white coat, as the shadow of a lighter coated mouse will not be confused with the mouse. Mice of the same strains may also have different coat colours: A^{vy}/a mice display variable expressivity ranging from yellow to agouti, with some mice having both yellow and agouti patches (Ounpraseuth et al., 2009). This provides an opportunity for future work to evaluate differences in detection settings between recording systems.

Strain and sex differences are known to also result in differences in animal size. Male mice for most genotypes are generally larger than female sizes, and in this study, the C3H female mice were noticeably the smallest and CD1 female mice were the largest. Hence animal strain and sex differences, and thereby animal size, could influence the tracking, for example, larger dispersion of shadows when tracking animals of a larger size. EthoVision can address this by allowing the user to manually define the maximum and minimum size of the animal. This option is, however, not available in ANY-Maze. The usability and type of detection settings therefore differ between tracking systems, and it is recommended that the user considers the genotype of mice and type of behavioural assay and apparatus prior to selecting the most optimal recording and tracking system. Within experiments, we do not recommend alternating between recording systems with different manufacturers and between different versions and particularly different detection methods within the same recording software. Under circumstances that this is inevitable, detection and recording methods should be factored in the analysis and reported in the study.

As a corollary, the reproducibility of experiments between laboratories may be low when different tracking software is applied. Towards this end, (Richter et al., 2011), used either ANY-Maze, Ethovision 3.0/3.1 or Ethovision XT software in an inter-laboratory comparison of 6 European laboratories for several tests of anxiety-like and exploratory behaviours, including the open field. Significant differences in distance travelled were obtained between laboratories. While the authors discuss environmental and experimental differences as

reasons for this variability, no mention is made to the differences in tracking software; they seemingly assume equivalence between tracking applications. While offering a great number of analytic features and versatility, details of the implemented algorithms for commercial video observation tools are not transparent to the user; this information, however, is readily available for freely distributed software applications for animal tracking. Twenty-eight of such freeware was investigated by Panadeira and coworkers (2021) for their features and strength and weaknesses. They report that only 3 programmes included calibration algorithms for the reduction of image distortion, which may substantially affect tracking accuracy and the analysis of activity-related parameters. Many other limitations (lack for export function of analysis metrics, lack for multiple animal recording, lack of updates and bug fixes in the last three years) were identified. Clearly, algorithms idiosyncratic to each system, were optimised for different video input/output types, type of animal/species being tracked, and calibration methods, making differences in animal phenotyping/profiling highly likely (Panadeiro et al., 2021). Taking this into consideration, a further comparison using more recording and tracking systems, as well as using a larger sample size to address the heteroscedasticity of the data in this study, could be performed to provide a more comprehensive understanding of the variation between tracking systems.

Fortunately, the heterogeneity in the data output between the video tracking and analysis systems utilised here has affected only few endpoints. Yet, these are the most frequently reported primary outcome measures on which decision about impairment/enhancement are typically based. It is of little conciliation that most parameters measured in this study returned similar and robust data independent of mouse strain (coat colour), light intensity or tracking software or hardware installation. Our unconventional approach of estimation statistics for data comparison provides direct visual information of the degree of differences/similarities between mouse strains on one hand, but also between the tracking applications.

5. Conclusion

Between-laboratory standardization and validation of read-outs is compromised by the use of different tracking software, calibration methods and lighting systems amongst others. Protocols for these factors are commonly kept idiosyncratic to each laboratory, to each experimenter and/or to each paradigm. This contributes to undeliberate systematic errors (Richter et al., 2009) and leads to seemingly irreproducible experimental results. Apart from generic factors of animal holding and maintenance, or specific experimental factors like water temperature or noise levels, we here provide compelling evidence that careful and detailed knowledge about the automated tracking software in use and the experimental environment is instrumental in ensuring that the same behaviour is indeed being probed. Finally, the differences in outcome may be intrinsic to the tracking application and given the great number of software tools that are available from vendors and as free ware, the reproducibility issue is difficult to resolve.

Acknowledgement:

This project included funding from the Innovative Medicines Initiative 2/EFPIA, European Quality in Preclinical Data (EQIPD) consortium under grant agreement number 777364. We would also like to acknowledge the staff of the Medical Research Facility for their support with animal care, handling and behavioural experiments.

Author Contributions

This research project was conceptualised by CL, SKJ and GR. The manuscript forms part of the PhD thesis of CL, who performed experiments and statistical analyses. CL and GR wrote the manuscript and all authors contributed to the final text and approved it for publication.

Data availability statement

All data are provided within the manuscript.

Conflict of interest:

The authors have no conflict of interest to report.

References

Bailoo, J. D., Bohlen, M. O., & Wahlsten, D. (2010). The precision of video and photocell tracking systems and the elimination of tracking errors with infrared backlighting. *Journal of Neuroscience Methods*, *188*(1), 45–52. https://doi.org/10.1016/j.jneumeth.2010.01.035

Denenberg, V. H. (1969). OPEN-FIELD BEHAVIOR IN THE RAT: WHAT DOES IT MEAN?*. *Annals of the New York Academy of Sciences*, *159*(3), 852–859. https://doi.org/https://doi.org/10.1111/j.1749-6632.1969.tb12983.x

Gehlenborg, N., & Wong, B. (2012). Heat maps. *Nature Methods*, *9*(3), 213. https://doi.org/10.1038/nmeth.1902

Ho, J., Tumkaya, T., Aryal, S., Choi, H., & Claridge-Chang, A. (2018). Moving beyond P values: Everyday data analysis with estimation plots. *BioRxiv*, 1–15. https://doi.org/10.1101/377978

Lind, N. M., Vinther, M., Hemmingsen, R. P., & Hansen, A. K. (2005). Validation of a digital video tracking system for recording pig locomotor behaviour. *Journal of Neuroscience Methods*, *143*(2), 123–132. https://doi.org/10.1016/J.JNEUMETH.2004.09.019

Noldus, L. P. J. J. J. J., Spink, A. J., & Tegelenbosch, R. A. J. J. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers*, *33*(3), 398–414. https://doi.org/10.3758/BF03195394

Ounpraseuth, S., Rafferty, T. M., McDonald-Phillips, R. E., Gammill, W. M., Siegel, E. R., Wheeler, K. L., Nilsson, E. A., & Cooney, C. A. (2009). A Method to Quantify Mouse Coat-Color Proportions. *PLOS ONE*, *4*(4), e5414-. https://doi.org/10.1371/journal.pone.0005414

Panadeiro, V., Rodriguez, A., Henry, J., Wlodkowic, D., & Andersson, M. (2021). A review of 28 free animal-tracking software applications: current features and limitations. *Lab Animal*, *50*(9), 246–254. https://doi.org/10.1038/s41684-021-00811-1

Richter, S. H., Garner, J. P., Zipser, B., Lewejohann, L., Sachser, N., Touma, C., Schindler, B., Chourbaji, S., Brandwein, C., Gass, P., van Stipdonk, N., van der Harst, J., Spruijt, B., Võikar, V., Wolfer, D. P., & Würbel, H. (2011). Effect of population heterogenization on the reproducibility of mouse behavior: a multi-laboratory study. *PloS One*, *6*(1), e16461–e16461. https://doi.org/10.1371/journal.pone.0016461

Robinson, L., Spruijt, B., & Riedel, G. (2018). Between and within laboratory reliability of mouse behaviour recorded in home-cage and open-field. *Journal of Neuroscience Methods*, *300*, 10–19. https://doi.org/10.1016/j.jneumeth.2017.11.019

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/. (n.d.).

Spink, A. J., Tegelenbosch, R. A. J., Buma, M. O. S., & Noldus, L. P. J. J. (2001). The EthoVision video tracking system—A tool for behavioral phenotyping of transgenic mice. *Physiology & Behavior*, *73*(5), 731–744. https://doi.org/10.1016/S0031-9384(01)00530-3

Spruijt, B. M., Peters, S. M., de Heer, R. C., Pothuizen, H. H. J., & van der Harst, J. E. (2014). Reproducibility and relevance of future behavioral sciences should benefit from a cross fertilization of past recommendations and today's technology: "Back to the future." *Journal of Neuroscience Methods*, *234*, 2–12. https://doi.org/10.1016/j.jneumeth.2014.03.001

469 Timotius, I. K., Canneva, F., Minakaki, G., Moceri, S., Plank, A.-C. C., Casadei, N., Riess, O., Winkler, J.,
470 Klucken, J., Eskofier, B., & von Hörsten, S. (2019). Systematic data analysis and data mining in
471 CatWalk gait analysis by heat mapping exemplified in rodent models for neurodegenerative
472 diseases. *Journal of Neuroscience Methods*, *326*, 108367.

473 Wahlsten, D., Bachmanov, A., Finn, D. A., & Crabbe, J. C. (2006). Stability of inbred mouse strain
474 differences in behavior and brain size between laboratories and across decades. *Proceedings of*
475 *the National Academy of Sciences of the United States of America*, *103*(44), 16364–16369.
476 https://doi.org/10.1073/pnas.0605342103

477

478

Table legends

Table 1. Statistical summary of differences between tracking systems for each mouse strain and light intensity in parameters which can be reasonably compared in both programs.

Parameters with statistical significance at $p<0.05$ revealed in all comparisons to C57BL6/J are highlighted in bold; parameters marked with an asterisk are of interest and further visualised in the figures. Differences were calculated using Kolmogorov-Smirnov tests using IBM SPSS Statistics 25 with statistical differences set at 95% confidence levels. *Ctrpt*, centre point; *Thig*, thigmotaxis; *ηp2*, partial eta-squared.

Table 1. Statistical summary of comparisons within tracking systems.

Data show estimated mean differences and confidence intervals, resampled 10000 times, with seed starting at 123456 using R (v.1.2.5033, RStudio, Inc.). Conventional statistics were performed using Kruskal Wallis H tests in IBM SPSS Statistics 25 and further analysed with Bonferroni-corrected Mann-Whitney tests (significant associations in bold) where significant findings were observed in Kruskal-Wallis H tests. Statistical differences were set at 95% confidence levels. *ηp2*, partial eta-squared (effect size).

Table 2. Statistical summary of comparisons between tracking systems.

Data show estimated mean differences and confidence intervals, resampled 10000 times, with seed starting at 123456 using R (v.1.2.5033, RStudio, Inc.). Conventional statistics were performed using Mann-Whitney U tests using IBM SPSS Statistics 25 with statistical differences set at 95% confidence levels. *ηp2*, partial eta-squared (effect size).

Figure legends

Figure 1. Calibrated images used for retracking in ANY-Maze and EthoVision XT.

Both images were taken at 58 lux light intensity with *a* white (left) *or* black (right) arena of identical dimensions (see Methods). The boundaries of the arena, thigmotaxic zone and the centre point were identically defined in both *ANY-Maze* and EthoVision XT.

Figure 2. Representative track-plots obtained for each experimental group in ANY-Maze and EthoVision XT at different illuminations.

Blue lines demarcate the borders of the thigmotaxic zone *(inner line)* and arena *(outer line)*; the center-point is represented by *a* '+' sign.

Figure 3. Heat-map demonstrating all differences observed in the open field between three mouse strains.

A total of 85 and 161 parameters were extracted from ANY-Maze (left) and EthoVision XT (right) respectively. C3H and CD1 were contrasted to C57 mice for both tracking systems and differences were analysed using p-values from Kolmogorov-Smirnov tests. Parameters were clustered according to measures pertaining to the apparatus (ANY-Maze, #1-20, and EthoVision XT #1-69), centre point (ANY-Maze, #21-36, and EthoVision XT #70-74) and thigmotaxis zone (ANY-Maze, #37-85, and EthoVision XT #75-161). Blue fields represent p-values with the threshold of 0.05 (dark blue = $p_{threshold} < 0.01$). The heat-map was visualized using R (v.1.2.5033, RStudio, Inc.).

Figure 3. Heat-map depicting differences in 28 parameters commonly reported by ANY-Maze and EthoVision XT.

The parameters were clustered according to measures related to apparatus (#1-8), centre point (#9-12) and thigmotaxic zone (#13-28). Mouse strains are presented individually, and blue fields represent p-values at a threshold of 0.05 (dark blue = $p_{threshold} < 0.01$) between tracking software. The heat-map was visualized using R (v.1.2.5033, RStudio, Inc.).

Figure 4. Comparison of ambulation of three mouse strains in the open field using two tracking systems and two light intensities

Parameters of ambulation are displayed for different mouse strains at 58 and 295 lux illumination for comparisons within ANY-Maze and EthoVision XT (EthoVision) The parameters of activity are as

follows: *total distance travelled* (A); *average distance from the centre point* (B); *thigmotaxic distance* (C). Single data points represent individuals for each strain. Cummings estimation plots represent means (gap) ± standard deviation (vertical lines) in the raw (top axis). The shaded curve and the error bar in the bottom axis show the distribution of sampling error and its respective 95% confidence interval for the difference between the means. Performance of C57BL6/J mice is represented by 0 line. Analysis and visualisation of estimated mean difference plots were performed using R (v.1.2.5033, RStudio, Inc.) with bootstrapping at 10000 samples, seed set at 123456, to estimate 95% confidence intervals. Post-hoc Bonferroni-corrected Mann-Whitney tests were performed for all pairwise comparisons between strains, where Kruskal Wallis H tests were significant, and are indicated as text in figure. Conventional statistics present significance values with alpha set to 5% ($p<0.05$). *Black* = C57 (N=5/6); *Green* = C3H (N=5); *Blue* = CD1 (N=5) mouse strains.

Figure 6. Comparison of thigmotaxis and stereotypic behaviour of three mouse strains in the open field using two tracking systems and two light intensities.

For three mouse strains at 58 and 295 lux illumination comparisons within ANY-Maze and EthoVision XT (EthoVision), we analysed parameters for thigmotaxis: time in thigmotaxis zone (A); thigmotaxis ratio (B) and number of rotations (C) for stereotypic behaviour. Single data points represent individuals for each strain. Cummings estimation plots represent means (gap**)** ± standard deviation (vertical lines) in the top axis. The shaded curve and the error bar in the bottom axis show the distribution of sampling error and its respective 95% confidence interval for the difference between the means**.** Performance of C57BL6/J mice is represented by 0 line. Analysis and visualisation of estimated mean difference plots were performed using R (v.1.2.5033, RStudio, Inc.). Post-hoc Bonferroni-corrected Mann-Whitney tests were performed for all pairwise comparisons between strains, where Kruskal Wallis H tests were significant, and are indicated as text in figure. Conventional statistics present significance values with alpha set to 5% ($p<0.05$). *Black* = C57 (N=5/6); *Green* = C3H (N=5); *Blue* = CD1 (N=5) mouse strains.

Figure 7. Direct comparison between tracking software for activity parameters extracted from behaviour in open field.

Three different mouse lines and two illumination intensities are presented. Parameters of activity (total distance travelled (A); mean distance from the centre point (B); thigmotaxic distance (C)), are compared between the 2 tracking systems ANY-Maze (ANY) and EthoVision (Etho). Single data points represent individuals for each strain. Cummings estimation plots represent means (gap) ± standard deviation (vertical lines) in the top axis. The shaded curve and the error bar in the bottom axis show the distribution of sampling error and its respective 95% confidence interval for the difference between the means. Data from ANY-Maze are represented by 0 line. Analysis and visualisation of estimated mean difference plots were performed using R (see legend Figure 5). Conventional statistics (Mann-Whitney U tests

between groups) present significance values with alpha set to 5% ($p<0.05$). *Black* = C57 (N=5/6); *Green* = C3H (N=5); *Blue* = CD1 (N=5) mouse strains.

Figure 8. Direct comparison between tracking software for thigmotaxis and stereotypy of mouse behaviour in open field.

For the different mouse strains and 58 and 295 lux illumination, parameters for thigmotaxis (time in thigmotaxis zone (D); thigmotaxis ratio (E)) and number of rotations (F) for stereotypic behaviour are compared between the 2 tracking systems ANY-Maze (ANY) and EthoVision (Etho). Single data points represent individuals for each strain. Cummings estimation plots represent means (gap) ± standard deviation (vertical lines) in the top axis. The shaded curve and the error bar in the bottom axis show the distribution of sampling error and its respective 95% confidence interval for the difference between the means. Data from ANY-Maze are represented by 0 line. Analysis and visualisation of estimated mean difference plots were performed using R (see Figure legend 5 for details). Conventional statistics (Mann-Whitney U tests between tracking systems) at alpha set to 5% ($p<0.05$) are indicated in the figure. *Black* = C57 (N=5/6); *Green* = C3H (N=5); *Blue* = CD1 (N=5) mouse strains.

Figure 9. Pearson correlations for measures of activity comparing between ANY-Maze and EthoVision XT.

Correlations of *total distance travelled* (A), *average distance from the centre point* (B) and *distance travelled in the thigmotaxic zone* (C) are represented for 58 (left) and 295 lux (right). Note that higher light intensities increase correlations (i.e. reproducibility) between the tracking applications. All graphs and correlation analyses (R- and p-values on figures) were performed and plotted using R (v.1.2.5033, RStudio, Inc.). *Black* = C57 (N=5/6); *Green* = C3H (N=5); *Blue* = CD1 (N=5) mouse strains.

Table (Editable version)

Click here to access/download;Table (Editable version);Lim et al_Tables submission_revision_CL.docx

Tables and figures

Table 1.

	ANY-Maze v. EthoVision XT	58 Lux									295 Lux								
		C57BL6/J			C3H			CD1			C57BL6/J			C3H			CD1		
	Parameters	Z-value	p-value	ηp2	Z-value	p-value	ηp2	Z-value	p-value	ηp2	Z-value	p-value	ηp2	Z-value	p-value	ηp2	Z-value	p-value	ηp2
1	*Distance	1.581	**0.013**	0.250	1.265	0.082	0.160	0.949	0.329	0.090	1.651	**0.009**	0.227	0.826	0.503	0.068	0.949	0.329	0.090
2	Mean speed	1.581	**0.013**	0.250	1.265	0.082	0.160	0.949	0.329	0.090	1.651	**0.009**	0.227	0.826	0.503	0.068	0.949	0.329	0.090
3	Max speed	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.265	0.082	0.160	1.651	**0.009**	0.227	1.651	**0.009**	0.273	1.581	**0.013**	0.250
4	Time mobile	1.265	0.082	0.160	0.632	0.819	0.040	0.632	0.819	0.040	1.321	0.061	0.145	0.716	0.685	0.051	0.632	0.819	0.040
5	Time immobile	1.265	0.082	0.160	0.632	0.819	0.040	0.632	0.819	0.040	1.321	0.061	0.145	0.716	0.685	0.051	0.632	0.819	0.040
6	*Clockwise rotations	0.316	1	0.010	0.632	0.819	0.040	0.632	0.819	0.040	1.651	**0.009**	0.227	1.101	0.177	0.121	0.949	0.329	0.090
7	*Anti-clockwise rotations	0.316	1	0.010	0.316	1	0.010	0.949	0.329	0.090	1.046	0.224	0.091	0.55	0.922	0.030	0.949	0.329	0.090
8	Absolute turn angle	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.651	**0.009**	0.227	1.651	**0.009**	0.273	1.581	**0.013**	0.250
9	*Ctrpt : mean distance from	0.632	0.819	0.040	0.316	1	0.010	0.632	0.819	0.040	1.046	0.224	0.091	0.55	0.922	0.030	0.316	1	0.010
10	Ctrpt : max distance from	0.632	0.819	0.040	0.632	0.819	0.040	1.581	**0.013**	0.250	1.321	0.061	0.145	0.771	0.593	0.059	0.949	0.329	0.090
11	Ctrpt : min distance from	0.632	0.819	0.040	0.632	0.819	0.040	0.316	1	0.010	0.44	0.99	0.016	0.44	0.99	0.019	0.632	0.819	0.040
12	Ctrpt : average absolute heading	1.581	**0.013**	0.250	0.949	0.329	0.090	0.949	0.329	0.090	1.651	**0.009**	0.227	1.321	0.061	0.175	0.949	0.329	0.090
13	Thig. : entries	1.581	**0.013**	0.250	0.632	0.819	0.040	0.949	0.329	0.090	0.716	0.685	0.043	0.44	0.99	0.019	0.632	0.819	0.040
14	*Thig. : time	0.316	1	0.010	0.316	1	0.010	0.316	1	0.010	0.44	0.99	0.016	0.55	0.922	0.030	0.316	1	0.010
15	*Thig. : distance	1.581	**0.013**	0.250	1.265	0.082	0.160	0.949	0.329	0.090	1.376	**0.045**	0.158	1.651	**0.009**	0.273	0.632	0.819	0.040
16	Thig. : latency to first entry	0.316	1	0.010	0.316	1	0.010	0.632	0.819	0.040	0.661	0.775	0.036	0.716	0.685	0.051	0.316	1	0.010
17	Thig. : latency to last entry	1.265	0.082	0.160	0.316	1	0.010	0.632	0.819	0.040	0.771	0.593	0.050	0.44	0.99	0.019	0.632	0.819	0.040
18	Thig. : average speed	1.581	**0.013**	0.250	1.265	0.082	0.160	1.581	**0.013**	0.250	1.651	**0.009**	0.227	1.101	0.177	0.121	0.949	0.329	0.090
19	Thig. : max speed	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.651	**0.009**	0.227	1.651	**0.009**	0.273	1.581	**0.013**	0.250
20	Thig. : max visit	0.316	1	0.010	0.316	1	0.010	0.632	0.819	0.040	0.385	0.998	0.012	0.44	0.99	0.019	0.632	0.819	0.040
21	Thig. : min visit	1.581	**0.013**	0.250	1.265	0.082	0.160	1.581	**0.013**	0.250	1.321	0.061	0.145	1.101	0.177	0.121	1.581	**0.013**	0.250
22	Thig. : mean visit	1.581	**0.013**	0.250	0.632	0.819	0.040	0.949	0.329	0.090	0.55	0.922	0.025	0.385	0.998	0.015	0.632	0.819	0.040
23	Thig. : time mobile	1.265	0.082	0.160	0.949	0.329	0.090	0.632	0.819	0.040	0.826	0.503	0.057	1.046	0.224	0.109	0.316	1	0.010
24	Thig. : time immobile	0.949	0.329	0.090	0.632	0.819	0.040	0.949	0.329	0.090	0.991	0.28	0.082	0.826	0.503	0.068	0.632	0.819	0.040
25	Thig. : max. distance from	0.949	0.329	0.090	0.949	0.329	0.090	1.581	**0.013**	0.250	1.376	**0.045**	0.158	1.321	0.061	0.175	1.581	**0.013**	0.250
26	Thig. : min distance from	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.651	**0.009**	0.227	1.651	**0.009**	0.273	1.581	**0.013**	0.250
27	Thig. : average absolute heading error	0.949	0.329	0.090	1.265	0.082	0.160	1.581	**0.013**	0.250	1.101	0.177	0.101	0.991	0.28	0.098	0.949	0.329	0.090
28	Thig. : absolute turn angle	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.581	**0.013**	0.250	1.651	**0.009**	0.227	1.651	**0.009**	0.273	1.581	**0.013**	0.250

Table 2.

Parameters	Light conditions	ANY-Maze				EthoVision XT			
		Statistics, p-value	ηp2	Mean difference [95% CI (lower; upper bounds)]		Statistics, p-value	ηp2	Mean difference [95% CI (lower; upper bounds)]	
				C3H minus C57BL6/J	CD1 minus C57BL6/J			C3H minus C57BL6/J	CD1 minus C57BL6/J
Total distance	58 lux	H(2)=6.32, p=0.034	0.451	-7.69 [-20.4; 1.22]	**17 [4.25; 28.8]**	H(2)=6.74, p=0.027	0.481	**-26.6 [-39.9; -17.4]**	-7.45 [-20.1; 4.58]
	295 lux	H(2)=3.78, p=0.159	0.268	-6.43 [-13.5; 2.75]	7.61 [-1.99; 16.1]	H(2)=3.47, p=0.184	0.247	-5.8 [-13.2; 4.08]	7.87 [-1.47; 16]
Average distance from centre-point	58 lux	H(2)=6.08, p=0.040	0.434	-0.001 [-0.021; 0.018]	**-0.030 [0.040; -0.020]**	H(2)=5.54, p=0.057	0.400	-0.002 [-0.024; 0.018]	-0.030 [-0.040; -0.021]
	295 lux	H(2)=7.75, p=0.013	0.553	**0.025 [0.012; 0.037]**	0.008 [-0.001; 0.021]	H(2)=7.71, p=0.012	0.550	**0.025 [0.012; 0.036]**	0.008 [-0.001; 0.021]
Distance in thigmotaxis	58 lux	H(2)=4.82, p=0.084	0.344	-4.95 [-11.6; -0.495]	3.09 [-2.21; 7.21]	H(2)=8.34, p=0.007	0.600	**-15.6 [-22; -8.18]**	**-11.8 [-19; -5.42]**
	295 lux	H(2)=0.96, p=0.647	0.068	0.445 [-2.1; 3.14]	5.92 [-0.783; 12.6]	H(2)=1.22, p=0.566	0.087	2.29 [-0.201; 4.93]	6.87 [-0.316; 14]
Time in thigmotaxis	58 lux	H(2)=4.38, p=0.105	0.313	-3.94 [-103; 77.3]	-95.8 [-165; -38.4]	H(2)=4.38, p=0.111	0.313	-1.01 [-106; 83.7]	-94.8 [-164; -36.8]
	295 lux	H(2)=8.72, p=0.005	0.623	**134 [65.8; 195]**	63 [10.2; 130]	H(2)= 8.72, p=0.005	0.623	**135 [69; 196]**	62 [9.36; 130]
Thigmotaxic ratio	58 lux	H(2)=3.12, p=0.217	0.223	-0.002 [-0.198; 0.135]	-0.125 [-0.203; -0.062]	H(2)=4.02, p=0.133	0.287	0.022 [-0.17; 0.162]	-0.137 [-0.208; -0.0758]
	295 lux	H(2)=2.69, p=0.279	0.192	0.188 [0.023; 0.351]	0.052 [-0.027; 0.16]	H(2)=3.12, p=0.219	0.223	0.183 [0.035; 0.32]	0.055 [-0.017; 0.158]
Rotations	58 lux	H(2)=8.00, p=0.011	0.571	-2 [-9.6; 4]	**12.2 [5; 17.8]**	H(2)=2.63, p=0.278	0.188	0.8 [-4.2; 05.4]	4.6 [-2.4; 10.6]
	295 lux	H(2)=3.62, p=0.172	0.258	-8.7 [-16.9; 0.546]	-4.9 [-10.4; 0.7]	H(2)=3.06, p=0.226	0.218	5.1 [0.167; 9.8]	1.3 [-3.9; 6.15]

Table 3.

Parameters	Light conditions	C57BL6/J			C3H			CD1		
		Statistics, p-value	ηp2	Mean difference [95% CI (lower; upper bounds)] EthoVision XT minus ANY-Maze	Statistics, p-value	ηp2	Mean difference [95% CI (lower; upper bounds)] EthoVision XT minus ANY-Maze	Statistics, p-value	ηp2	Mean difference [95% CI (lower; upper bounds)] EthoVision XT minus ANY-Maze
Total distance	58 lux	U=0, p=0.007	0.870	34 [24.4; 43.2]	U=4, p=0.099	0.592	15.2 [3.14; 26.5]	U=6, p=0.225	0.453	9.55 [-6.14; 24.7]
	295 lux	U=0, p=0.003	0.869	9.05 [6.12; 11.9]	U=6, p=0.230	0.453	9.69 [-1.26; 21.5]	U=6, p=0.226	0.453	9.32 [-2.22; 22]
Average distance from centre-point	58 lux	U=9, p=0.549	0.244	0.002 [-0.007; 0.011]	U=10, p=0.700	0.174	0.001 [-0.025; 0.028]	U=9, p=0.556	0.244	0.002 [-0.01; 0.013]
	295 lux	U=11, p=0.294	0.339	0.002 [-0.005; 0.009]	U=10, p=0.691	0.174	0.002 [-0.015; 0.017]	U=10, p=0.687	0.174	0.002 [-0.012; 0.018]
Distance in thigmotaxis	58 lux	U=0, p=0.007	0.870	20.3 [12.9; 25.6]	U=2, p=0.032	0.731	9.71 [3.79; 15.4]	U=5, p=0.152	0.522	5.48 [-0.498; 10.7]
	295 lux	U=1, p=0.004	0.821	5.1 [2.95; 7.09]	U=0, p=0.008	0.870	6.94 [4.01; 10.3]	U=8, p=0.419	0.313	6.05 [-3.31; 16.3]
Time in thigmotaxis	58 lux	U=11, p=0.840	0.104	-5.08 [-59.1; 45.7]	U=11, p=0.851	0.104	-2.15 [-122; 116]	U=10, p=0.699	0.174	-4.09 [-76.6; 66.7]
	295 lux	U=14, p=0.594	0.193	-3.44 [-27.5; 18.1]	U=12, p>0.999	0.0.35	-2.45 [-93.7; 83.7]	U=10, p=0.687	0.174	-4.52 [-84; 77]
Thigmotaxic ratio	58 lux	U=8, p=0.419	0.313	0.039 [-0.055; 0.126]	U=8, p=0.427	0.313	0.063 [-0.141; 0.272]	U=6, p=0.225	0.453	0.027 [-0.010; 0.058]
	295 lux	U=12, p=0.406	0.290	0.027 [-0.017; 0.071]	U=10, p=0.691	0.174	0.022 [-0.211; 0.225]	U=8, p=0.419	0.313	0.031 [-0.095; 0.153]
Rotations	58 lux	U=12, p=0.833	0.070	-1.6 [-7.6; 3.4]	U=11, p=0.735	0.141	1.2 [-5.6; 6.6]	U=4, p=0.084	0.594	-9.2 [-16.6; -2.2]
	295 lux	U=0, p=0.002	0.872	-17 [-22.5; -12.2]	U=10, p=0.691	0.174	-3.2 [-12.2; 4.6]	U=0.5, p=0.016	0.838	-10.8 [-16.2; -5.6]

Figure 1

Click here to access/download;Figure;Figure 1.tif

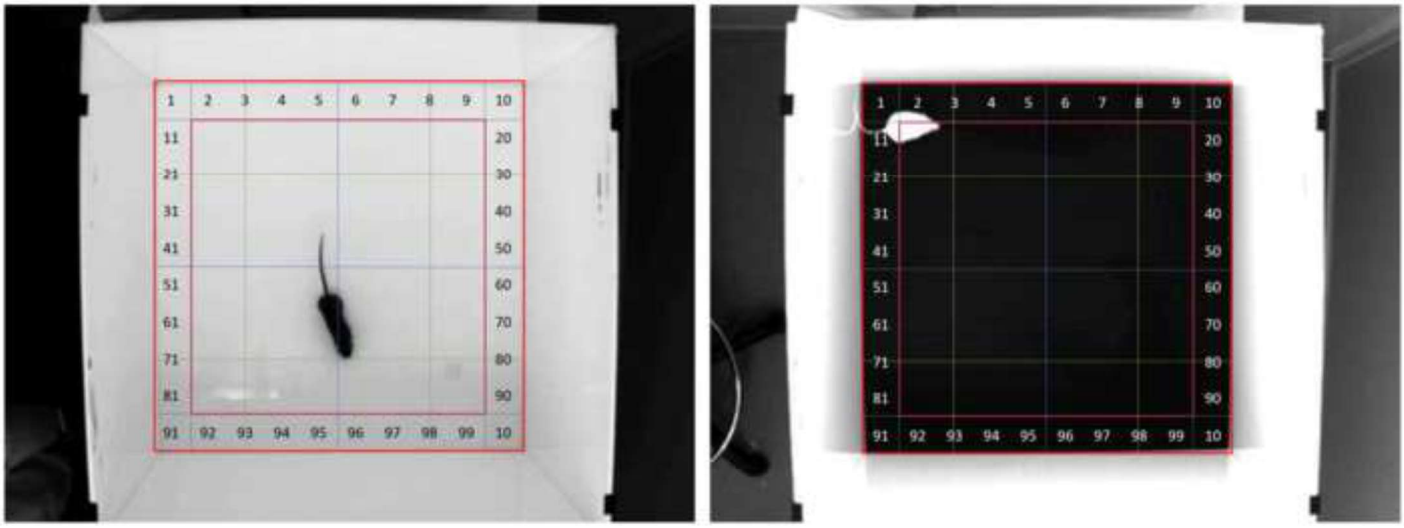


Figure 2

Click here to access/download;Figure;Figure 2.tif

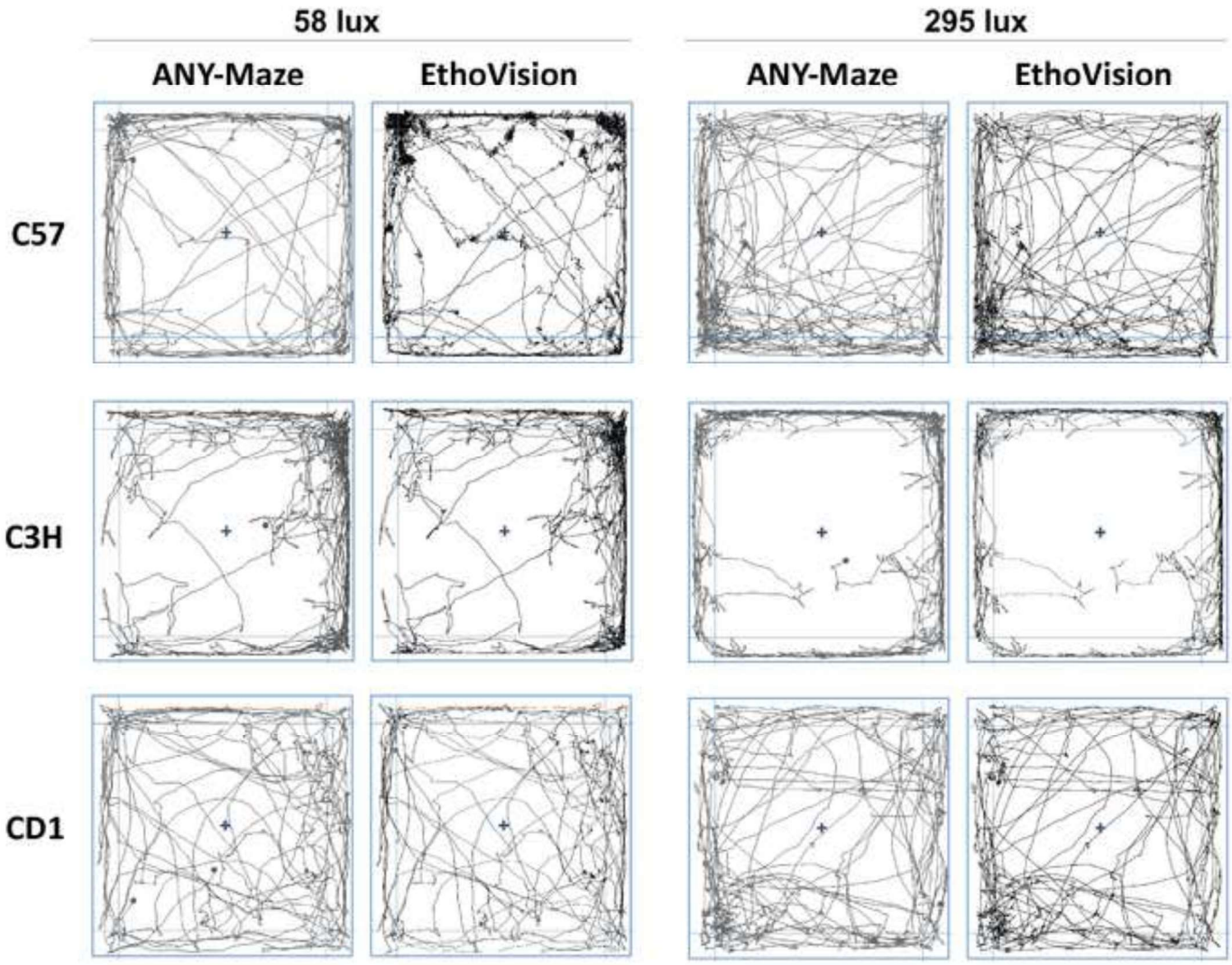


Figure 3

Click here to access/download;Figure;Figure 3.tif

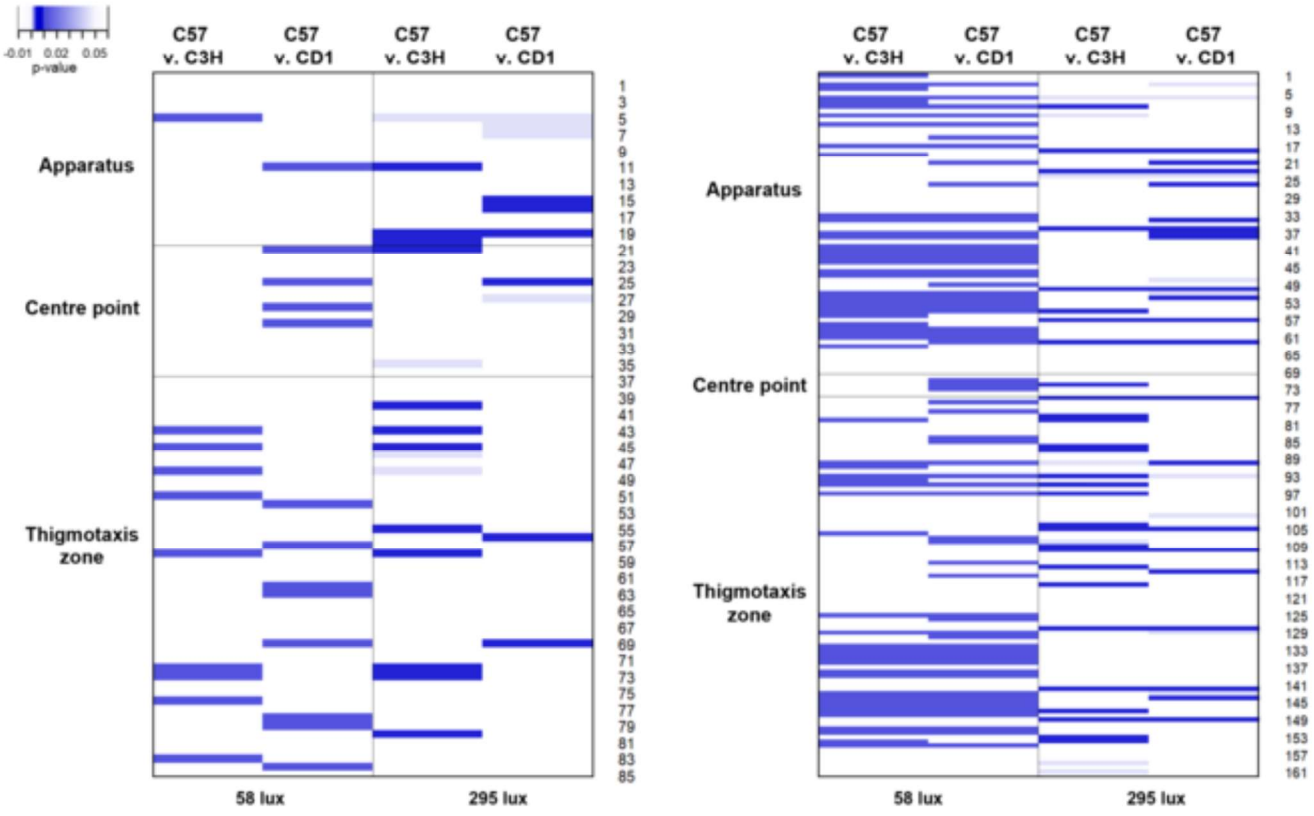


Figure 4

Click here to access/download;Figure;Figure 4.tif

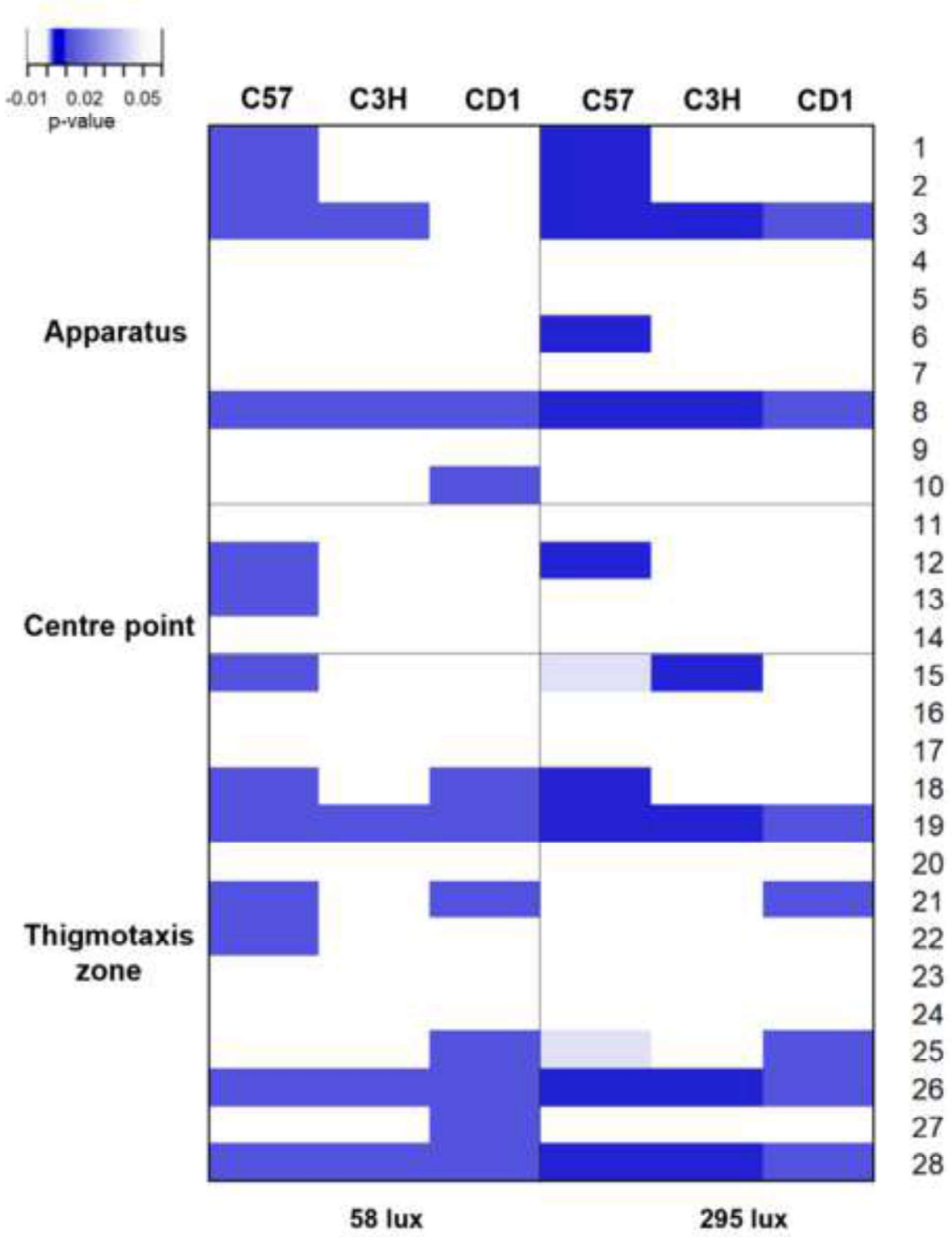


Figure 5

Click here to access/download;Figure;Figure 5.tif

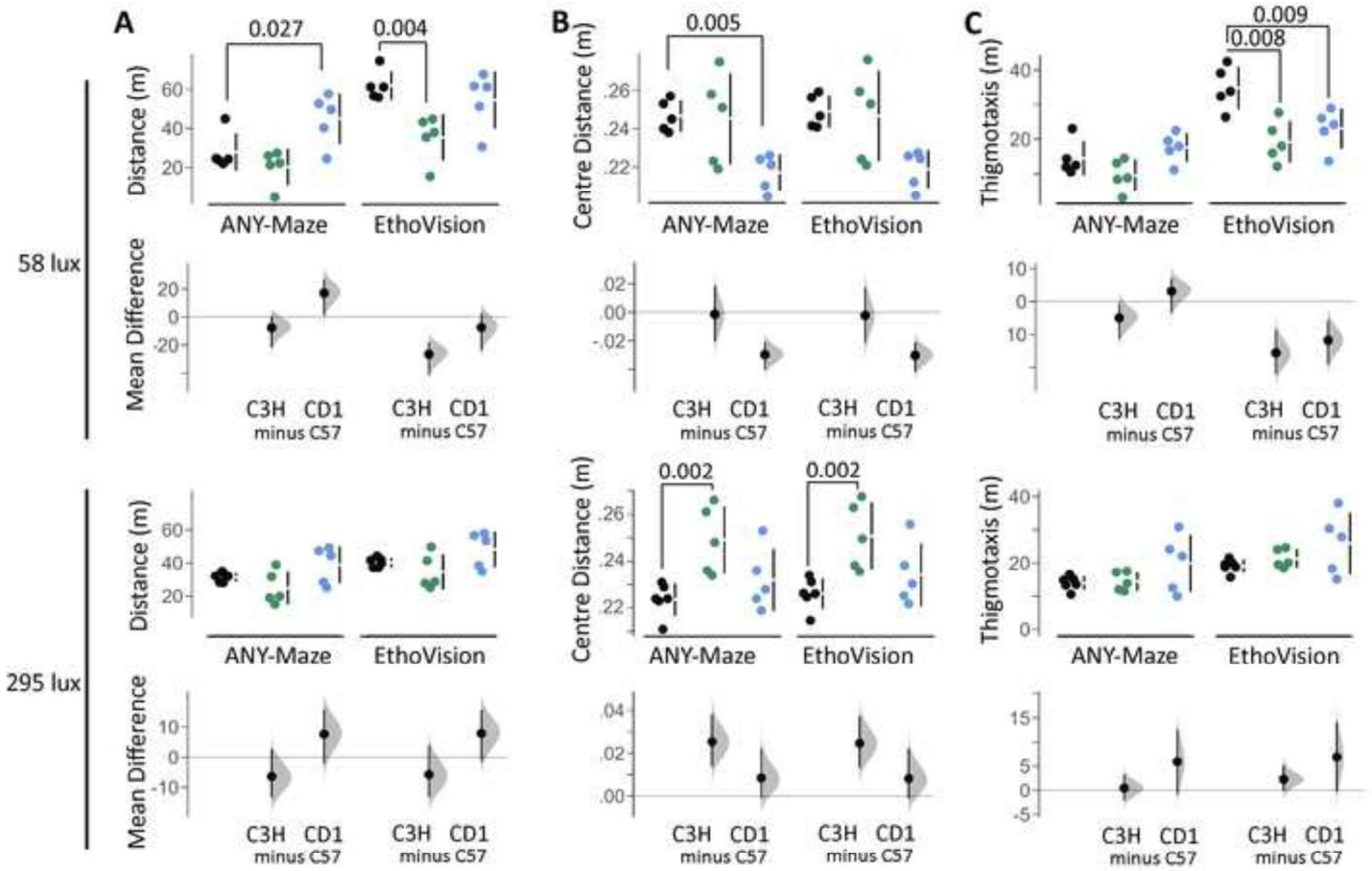


Figure 6

Click here to access/download;Figure;Figure 6.tif

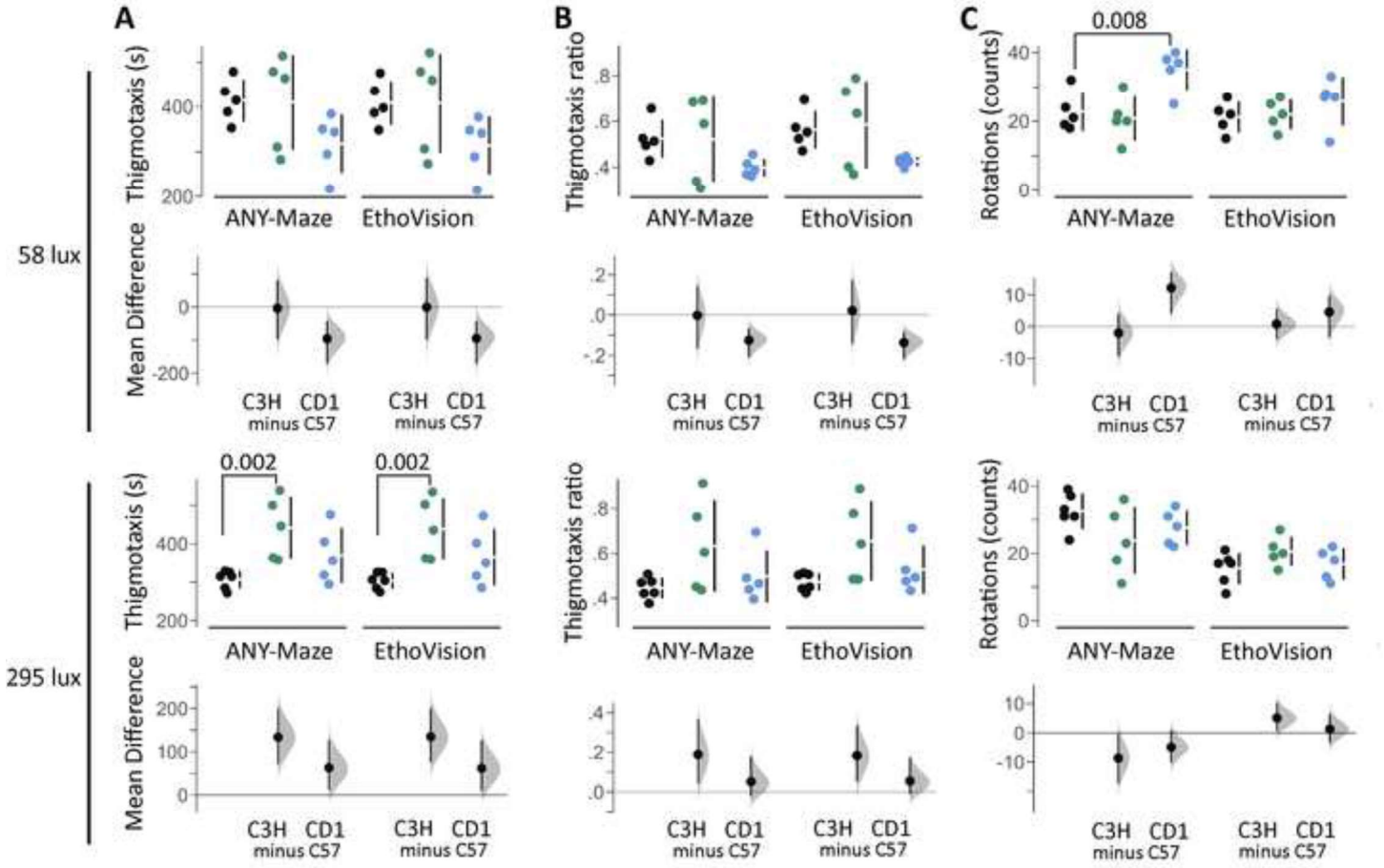


Figure 7

Click here to access/download;Figure;Figure 7.tif

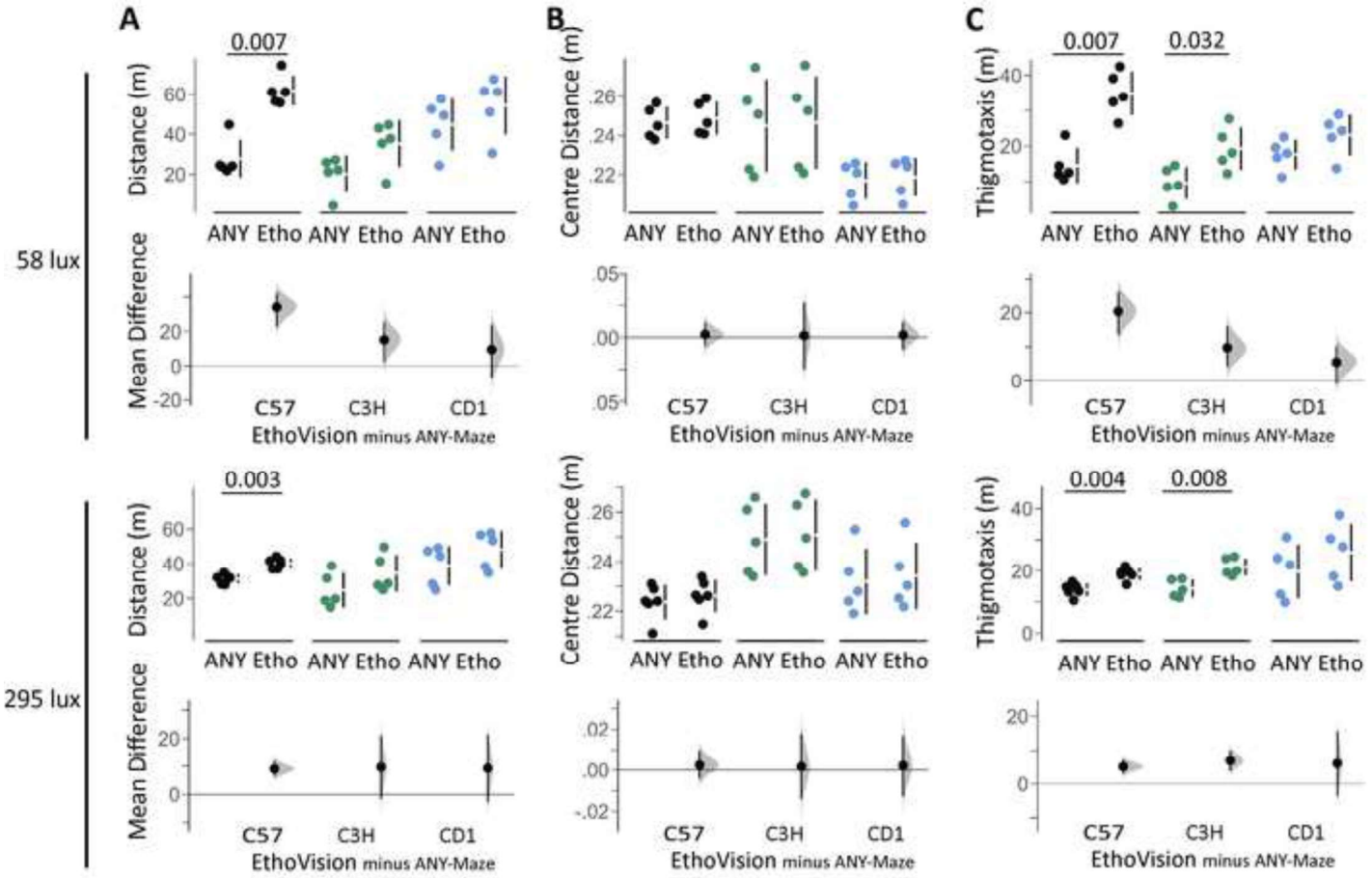


Figure 8

Click here to access/download;Figure;Figure 8.tif

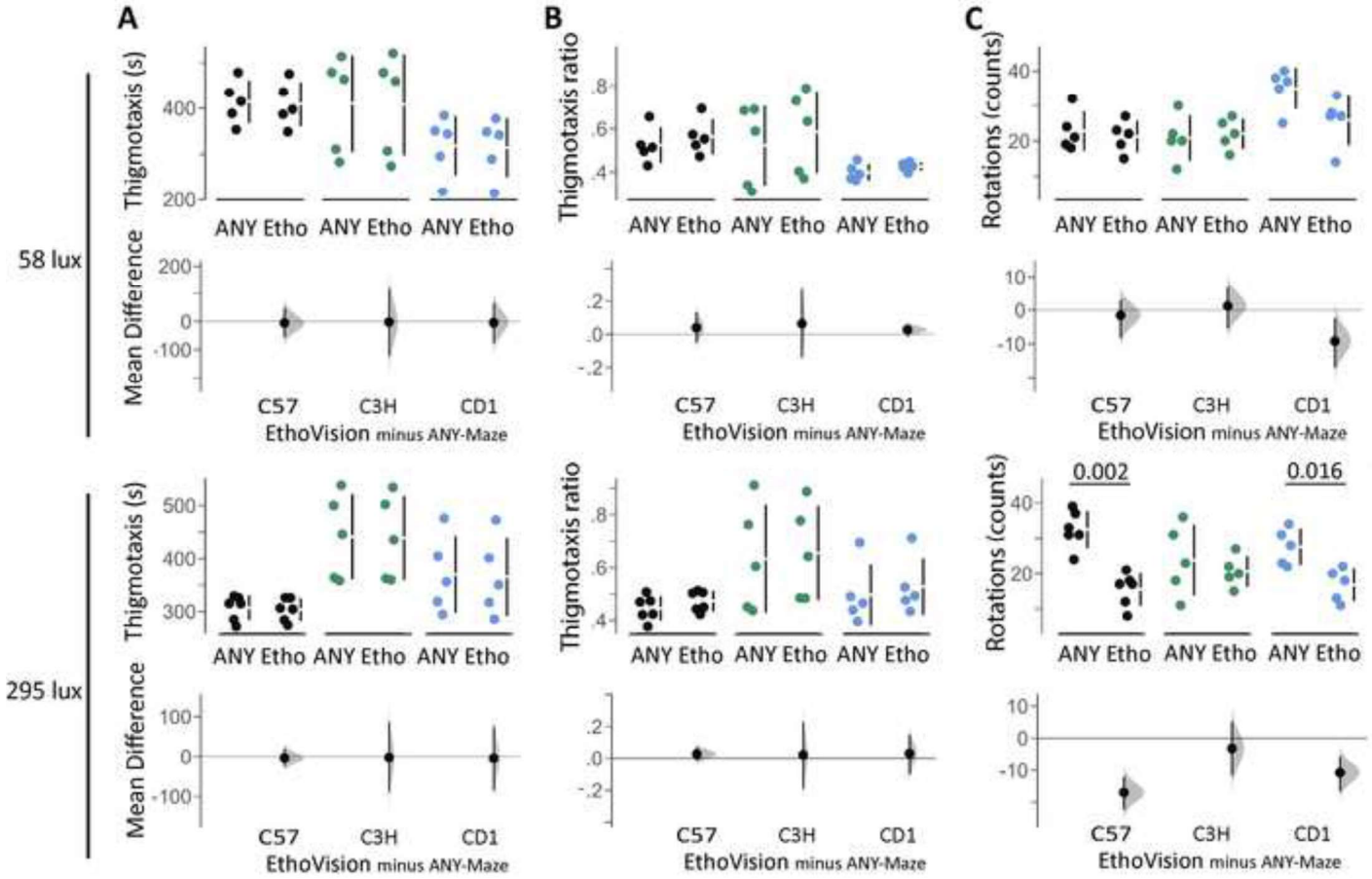
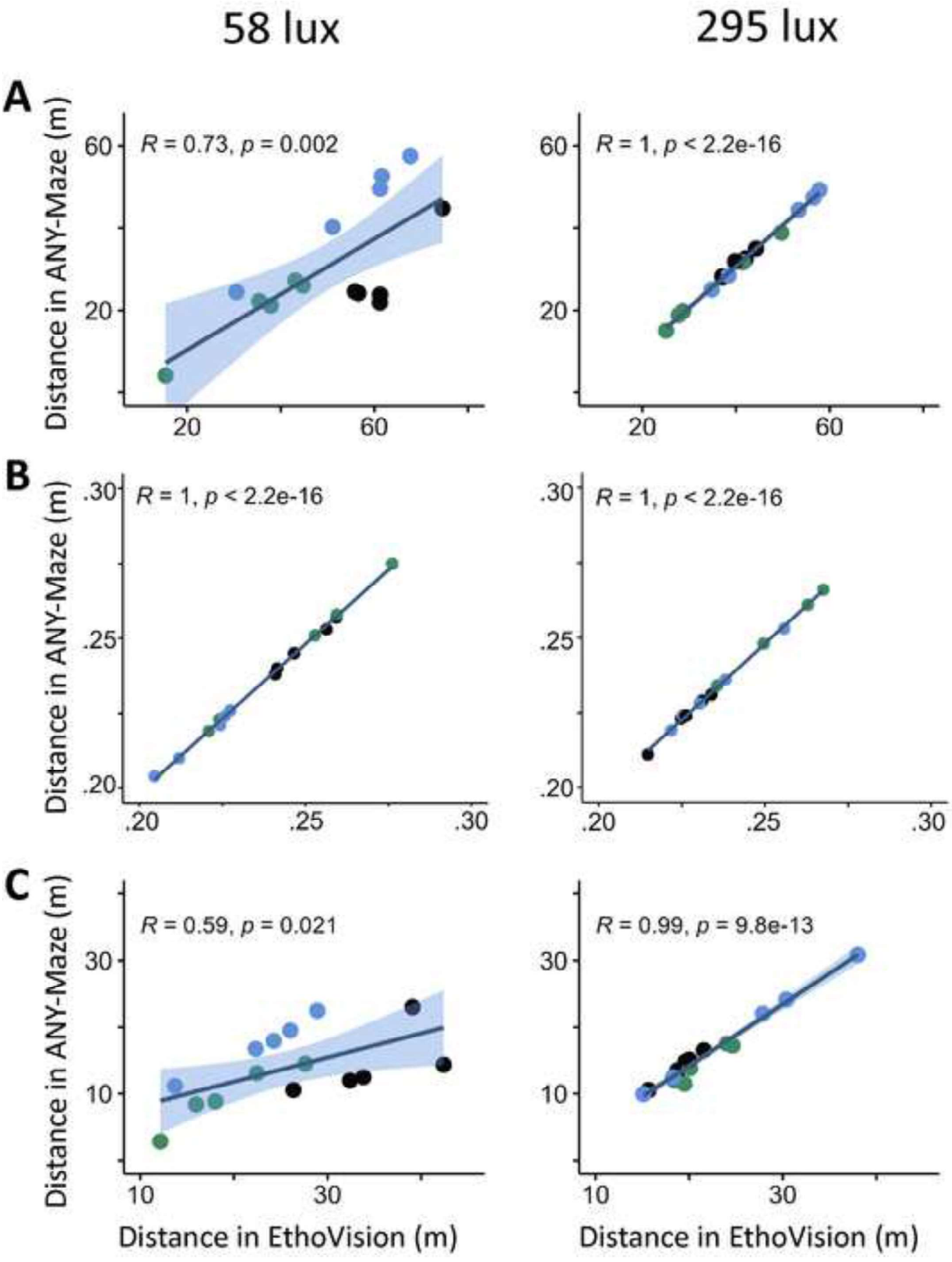


Figure 9

Click here to access/download;Figure;Figure 9.tif



Acknowledgement:

This project included funding from the Innovative Medicines Initiative 2/EFPIA, European Quality in Preclinical Data (EQIPD) consortium under grant agreement number 777364. We would also like to acknowledge the staff of the Medical Research Facility for their support with animal care, handling and behavioural experiments.

Author Contributions

This research project was conceptualised by CL, SKJ and GR. The manuscript forms part of the PhD thesis of CL, who performed experiments and statistical analyses. CL and GR wrote the manuscript and all authors contributed to the final text and approved it for publication.

Data availability statement

All data are provided within the manuscript.

Conflict of interest:
The authors have no conflict of interest to report.