

ANALYTICAL VALIDATION OF DIGITAL MOUSE DETECTION, IDENTIFICATION, AND ACTIVITY IN THE JAX ENVISION™ PLATFORM

| Summary

- The foundation of the JAX Envision™ platform is a set of machine learning algorithms that detect and identify individual mice within their home cage.
- Using an analytical validation framework, Envision demonstrates strong detection and identification efficiency and accuracy across a broad range of mouse strains and microenvironmental conditions.
- When mice are clearly visible, the JAX Envision platform detects and identifies mice with accuracy of 99.9% and 95.2%, respectively, against a human annotated validation dataset.
- The results of this validation highlight how the JAX Envision platform enables continuous experimental monitoring with highly accurate detection and tracking of individual mice in a cage for weeks to months at a time.

Rapid advances in sensor technologies and computational capabilities have provided a unique opportunity to enhance the value of animal studies. Complementing standard measures with continuous measures of behavior and physiology in the home cage environment provides a more dynamic, biologically, and clinically relevant characterization of disease progression and therapeutic effects.

Several approaches to addressing these issues have been identified, including home cage RFID tracking, photo beam interruption, and camera-based systems. Though laboratory mice spend most of their lifetime in the home cage, few available tools are capable of continuously monitoring mouse home cage behavior longitudinally.

Mice are social animals and both behavior and physiology are impacted by their social environment. However,

existing tools have not been able to reliably isolate and quantify the behavior of an individual animal in a group-housed environment. Additionally, most behavioral work occurs during the light cycle, despite mice being nocturnal animals. The dark cycle thus represents great unexploited potential to understand mouse behavior in context.

Through machine-learning-powered computer vision, cloud-based storage and analysis, and a readily accessible user interface, Envision™ by The Jackson Laboratory overcomes these barriers and empowers preclinical researchers with validated, translatable digital readouts of animal behavior in the home cage. Continuous monitoring provides unprecedented resolution of both discrete behavior and longitudinal variance – all aided by intuitive, flexible visualization tools. Moreover, by providing access to collaborative development environments, Envision paves the way for future advances in behavioral monitoring, improving the value of preclinical studies.

Envision facilitates automated machine learning behavioral analysis and supports longitudinal assessment of spontaneous behavior, enabling the detection of subtleties that may go unnoticed by gross cage-side observations. The platform continuously records and performs dynamic analysis of video for up to three uniquely ear-tagged mice housed in the same home cage. Envision enables continuous and seamless home cage analysis of individual mice within a group for weeks to months, in both the light and dark cycles, and with substantially greater capture of microenvironmental detail (Figure 1).

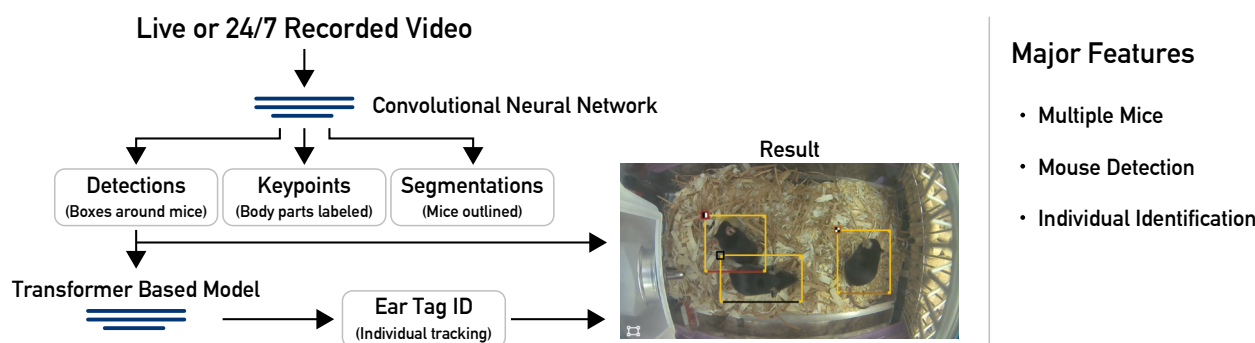


Figure 1

JAX Envision® platform utilizing (i) a custom convolutional neural network for mouse detection and (ii) transformer model for ear tag identification individually can track multiple mice within a home cage in both the light and dark cycles in a diversity of microenvironments.

| Verification & Validation of Mouse Detection, Mouse Identification, and Average Movement

The foundation of the JAX Envision platform is a set of ML-based computer vision algorithms that 1) detect mice, a process referred to as mouse detection; 2) identify individual mice within a cage, a process referred to as mouse identification; and 3) generate average activity measures at the cage and individual animal level¹. To ensure scientific rigor and provide evidence supporting the reliability and relevance of the detection, identification, and average activity algorithms, we analytically validated the models under several conditions in line with the concepts described in the V3 Framework originally published by the Digital Medicine Society² and adapted for preclinical use by the Digital In Vivo Alliance (DIVA)³, and adopted by DIVA and the 3Rs Collaborative^{4,5}.

Methods - Verification

The Envision platform features a highly optimized camera system, including a unique demosaicing scheme for night imagery, that maximizes video quality for off-line analysis and model inference performance. Various frame rates, resolutions, and codecs were experimented with to assess the system’s overall performance using widely adopted industry metrics such as Peak-Signal-To-Noise-Ratio and Visual Multimethod Assessment Function (VMAF). The net result is the JAX Envision platform produces extremely sharp high-definition video that enables the machine learning pipeline to track and identify mice in a variety of lightning conditions with high accuracy.

Analytical Validation - Detection and Identification

For validation of mouse detection and identification, one hundred video clips, each one minute in length of mice in the Allentown Discovery™ IVC cages, were selected. Cage dimensions were 6.14” x 5.04” x 12.64”, with 500 cm² floor space on the interior of the cage. These videos represented a diversity of environmental conditions, including contrasting mouse strains/coat colors (C3H/HeJ, C57BL/6J, and BALB/cJ), bedding (Alpha-dri®, Aspen Chips, Sani-chips), enrichment (wood blocks) and nesting materials (4-8 grams brown shredded paper and 1-2 cotton squares), levels of occlusion, and lighting conditions (both light and dark cycle). It is important to note that cage videos used for validation were not the same as those used for the training set. This separation of data by cage prevents the possibility of artificially inflated model performance due to information in the validation dataset being present in the training dataset (data leakage).

To compare the machine learning algorithms with known locations of mice, a ground truth dataset was generated by human observers. Annotators created bounding boxes outlining each visible mouse in the cage at short intervals between selected videos. Bounding box locations were interpolated linearly during these internals between human annotations. Bounding boxes were tagged with the ear tag of the mouse being annotated, so each box contained information about both mouse ID and location. Using this approach, the model was validated using 100 minutes of video data, comprised of approximately 180,000 frames of video and 498,099 mouse bounding boxes. The ground truth dataset represents both the light and dark cycle as well as agouti, black, and white mice. Several iterations of algorithms were validated against this dataset to examine the evolution of algorithm performance resulting from algorithm improvements.

Mice were identified with custom RapID® barcoded ear tags (**Figure 2**). One tag was placed in each ear of each mouse with one tag positioned anteriorly and the other positioned posteriorly.

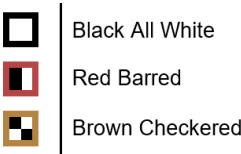


Figure 2
Patterns used on the custom RapID® ear tags for animal identification.

To assess the performance of the detection and identification models (2025 v3.0), the primary validation measures were efficiency and accuracy (**Figure 3**). For both detection and identification models, efficiency penalizes false negatives while accuracy penalizes false positives. *Detection efficiency* was determined by quantifying the number of detections made by the hydra (foundational) model for a video and then normalizing this quantity by the number of mice detected in the ground truth annotations for the same video. The number of mice in a frame of the ground truth dataset was 3 when there was no occlusion, but various sources of occlusion could reduce the number of mice in the ground truth, thus reducing the number of possible detections. Partially occluded mice were counted as valid opportunities for detection.

While detection efficiency simply compares the number of bounding boxes between the model inferences and the ground truth, *detection accuracy* quantifies the matches between model detections and ground truth annotations. A "match" was defined as any model detection where the predicted bounding box for the mouse overlapped a ground truth bounding box by at least 90% (based on Intersection over Union, or IoU). Detection accuracy was then calculated by dividing the number of these matches by the total number of model detections, resulting in the percentage of model detections that correspond to a true mouse.

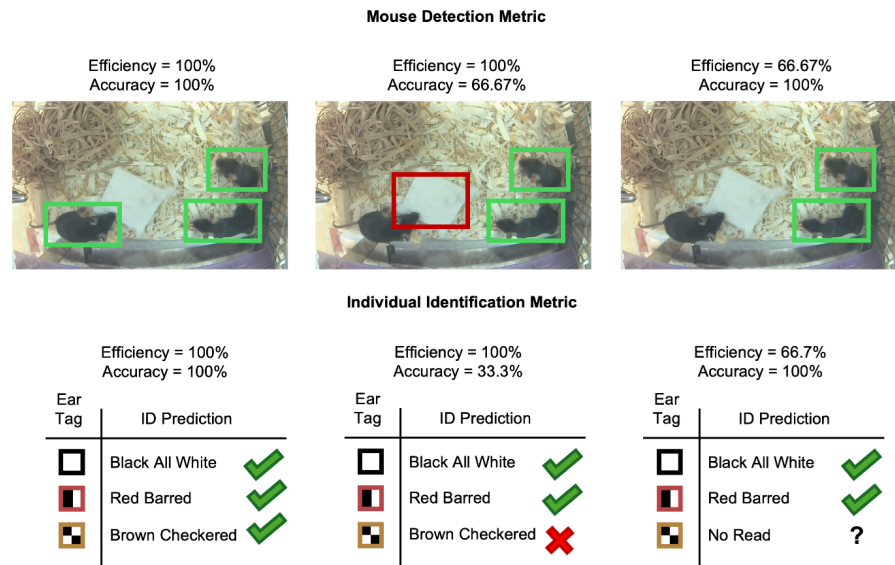


Figure 3
Examples of varying levels of *detection* efficiency and accuracy as well as *identification* efficiency and accuracy.

Identification efficiency was calculated as the number of times the model assigned a mouse identity over the total number of mouse detections in the ground truth dataset. Note that when an animal was partially occluded, as long as the mouse was detected by the model, this was still counted as an opportunity to identify an animal. *Identification accuracy* was calculated as the proportion of correct mouse identifications relative to the total number of identifications made by the model.

Videos were sorted by clear visibility, no occlusion, and partial occlusion. Clear visibility included 39,600 frames of video where mice were clearly visible and included no occlusion and no hanging or climbing for more than 5 seconds. No occlusion included 90,000 frames of video when mice were not occluded (but included climbing and hanging which influence visibility). Finally, partial occlusion included 90,000 frames of video when mice were partially occluded by nesting material, other mice in the cage, or cage furniture such as the water bottle. Performance metrics were then calculated for each condition.

Analytical Validation - Average Activity

For average activity in mice, the same set of human-annotated data was used for detection and ID validation. Euclidean distances between the centroids of interpolated bounding boxes were used for these human-annotated frames. The resulting dataset was used to evaluate the performance of the average activity metrics both at the cage level without ID and at the individual animal level with ID-dependent average activity measures. In addition, we utilized a dataset comparing mice that were orally dosed with either 16 mg/kg caffeine or vehicle within an hour of the lights coming on during the light cycle. We compared one hour of activity shortly after dosing to a circadian-matched hour from the day before dosing.

Results

Detection and Identification

Under conditions when mice were clearly visible, the JAX Envision platform detects and identifies mice with accuracy of 99.9% and 95.2%, respectively, against the validation dataset (**Figure 4**).

Detection		Identification	
Accuracy (%)	Efficiency (%)	Accuracy (%)	Efficiency (%)
99.9	93.4	95.2	99.9

Figure 4
Algorithm performance metrics for detection and identification efficiency and accuracy in conditions with clear visibility.

In unoccluded conditions, detection efficiency was generally high (94.4% in the light and 90.4% in the dark cycle conditions (**Figure 5**) and, when mice were detected, the accuracy of detection was very high (99.9% in the light and 100% in the dark cycle conditions (**Figure 5**). Individual mouse identification efficiency was also very high in unoccluded conditions (98.8-100%, **Figure 5**). Identification accuracy was also very high in most cases (above 92.1-95.3%, **Figure 5**).

Light Cycle	Detection		Identification	
	Accuracy (%)	Efficiency (%)	Accuracy (%)	Efficiency (%)
Light	99.9	94.4	95.3	98.8
Dark	100.0	90.4	92.1	100.0

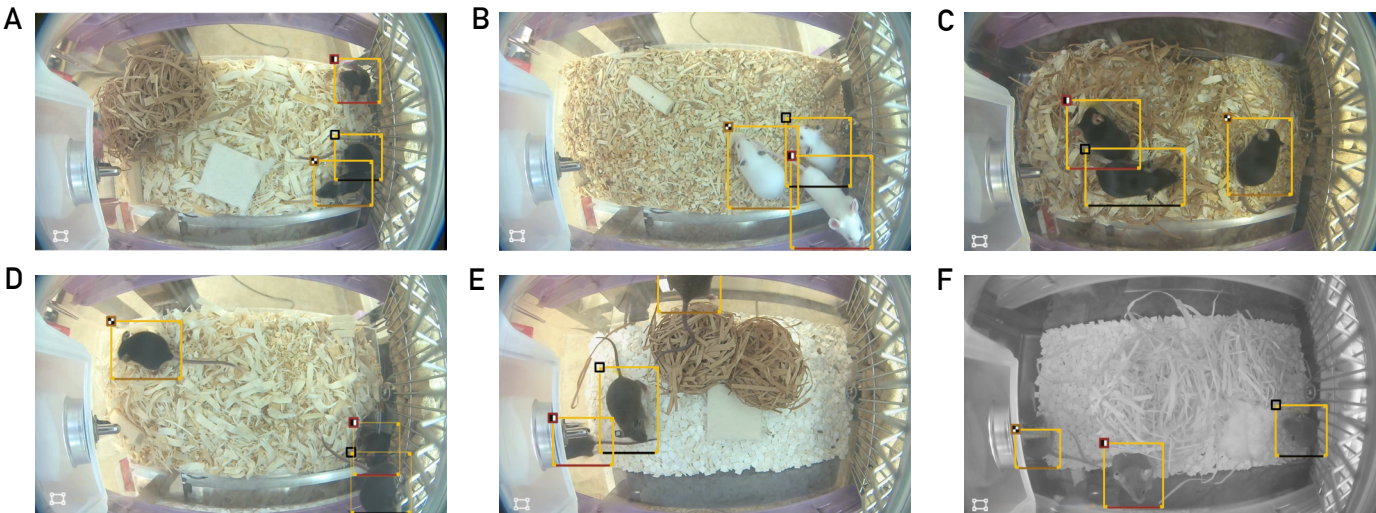


Figure 5
Algorithm performance metrics for detection and identification efficiency and accuracy in conditions without occlusion by light cycle. Example of a cage condition A) housing 3 C57BL/6J mice during the light cycle; and B) housing 3 BALB/cJ mice during the light cycle; and C) housing 3 C57BL/6J mice during the light cycle; and D) housing 3 C57BL/6J mice during the light cycle; and E) housing 3 C3H/HeJ mice during the light cycle; and F) housing 3 C3H/HeJ mice during the dark cycle.

In conditions of partial occlusion, there was a slight drop in detection efficiency (92.0% in the light and 84.5% in the dark cycle conditions, **Figure 6**), as occlusion reduced the opportunities for a mouse to be positively detected. Identification accuracy and efficiency was also high in partially occluded conditions (91.1-99.1% and 99.3-99.6%, respectively, **Figure 6**). Partial occlusion is often associated with less activity as mice are typically resting in their nests. However, identification accuracy was consistently high when mice were active, indicating that activity profiles are likely to be reliable indicators of individual activity.

Light Cycle	Detection		Identification	
	Accuracy (%)	Efficiency (%)	Accuracy (%)	Efficiency (%)
Light	97.3	92.0	91.1	99.3
Dark	99.9	84.5	91.0	99.6

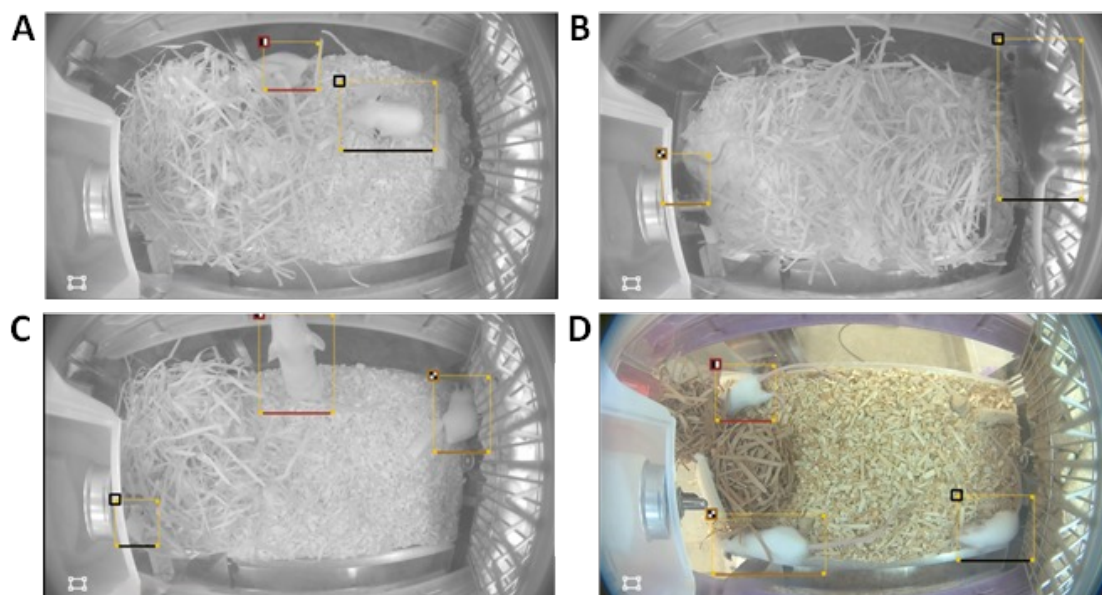


Figure 6

Algorithm performance metrics for detection and identification efficiency and accuracy in partially occluded conditions by light cycle. Example of a cage condition A) housing 3 BALB/cJ mice during the dark cycle; and B) during the dark cycle housing 3 C57BL/6J mice; and C) during the dark cycle housing 3 BALB/cJ mice; and D) during the light cycle housing 3 BALB/cJ mice.

Average Activity

Average activity is computed at the individual level with individual identification. Individual animal metrics are strongly correlated with human-annotated metrics (Pearson's $r = 0.875-0.954$, **Figure 7**). The raw data showed signs of heteroscedasticity; therefore, a Box-Cox power transform was utilized to identify a potential transformation to remediate this issue. Consequently, both the human and computer vision tracked datasets were square-root transformed for further analysis. The results are displayed in a square-root transformed coordinate space, but the X and Y axis labels are back-transformed. These results indicate that the output of mouse detection and identification can be used as substrates for derived metrics like locomotor activity.

To demonstrate that the system can produce biological insights on average activity, we utilized a dataset comparing mice that were orally dosed with either 16 mg/kg caffeine or vehicle within an hour of the lights coming on during the light cycle. We compared one hour of activity shortly after dosing to a circadian-matched hour from the day before dosing. In

a two-way ANOVA model looking at date of administration (pre-dose vs dose day), drug (caffeine vs. vehicle) and their interaction, we explained much of the variation in cage-level average activity as computed by average activity in an hour ($R^2 = 0.928$, **Figure 8**). This result demonstrates the utility of the system for identifying pharmacological effects on mouse behavior following drug administration.

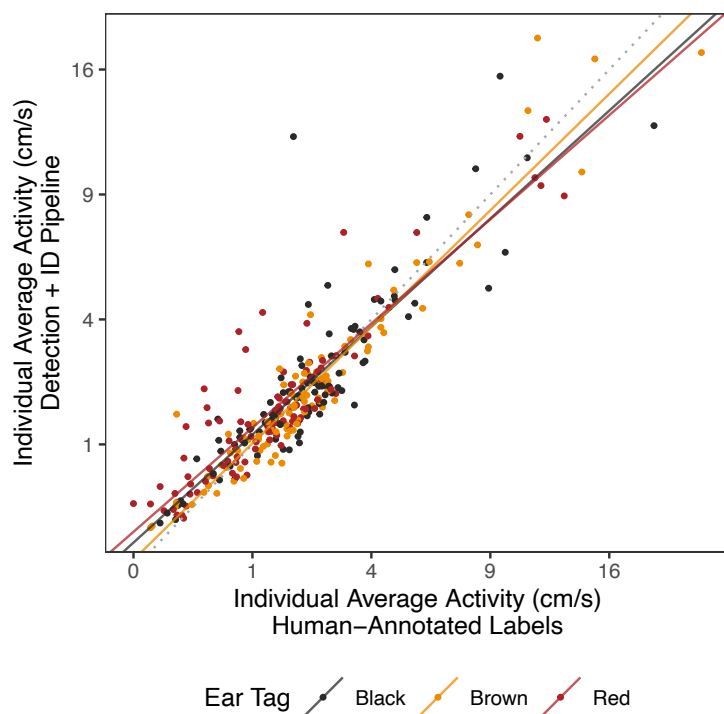


Figure 7

Average activity validation for 1-minute clips of video. Envision values for individual animal average activity metrics strongly correlate with human annotations of the same clips (Pearson's r ranges from 0.875 to 0.954).

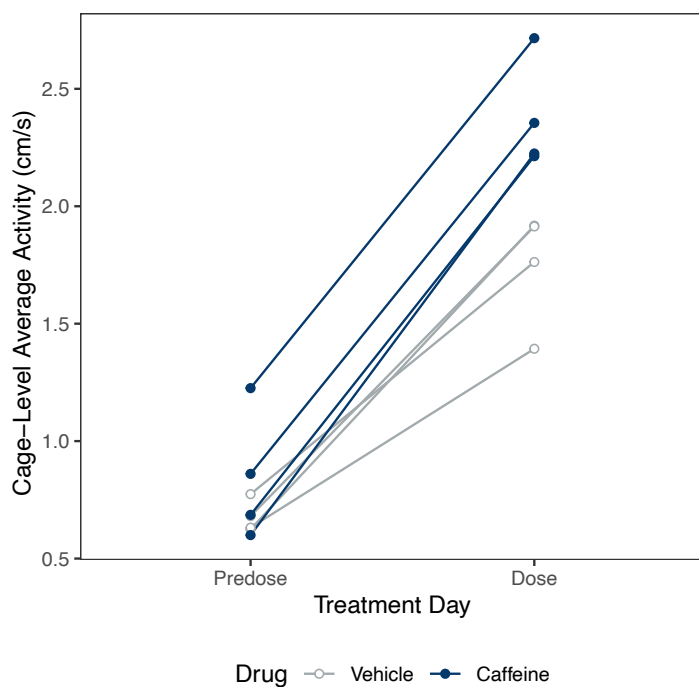


Figure 8

Summary of an experiment demonstrating pharmacologic effects using Envision. There was a substantial increase in average activity in mice following a high dose of caffeine (16 mg/kg, P.O.). Each set of points connected by a line is a single cage from the dataset. A two-way ANOVA of dose and timing explains 92.8% of the variance in the dataset.

Discussion

Under conditions when mice were clearly visible, the JAX Envision platform detects and identifies mice with 99.9% and 95.2% accuracy, respectively. These results indicate that the foundational mouse detection algorithm can detect mice with a high degree of accuracy when sufficient visual information is available. As is expected, detection accuracy drops slightly when mice are occluded. Therefore, occlusion by huddling or nesting material is thus a known limitation of detection, identification, and any derived metrics. As mice are typically asleep or resting while in the nest, it is often a fair assumption that mice are inactive in some way. However, we note that nesting and bedding occlusion can vary by cage, which should be considered in experimental design.

Individual-level average activity is strongly correlated with human-annotated datasets, and we capture a large amount of biological variation in a pharmacology experiment involving caffeine administration, demonstrating the potential application of this platform in pharmacological research. In conclusion, the JAX Envision platform demonstrates excellent detection and identification efficiency and accuracy across a range of mouse strains and microenvironmental conditions.

References

- [1] Robertson, T. L., Ellis, M., Bratcher Petersen, N. A., Ruidiaz, M. E., Harada, K., Toburen, D., Oberhauser, J. P., Grzenda, D., Peltier, N. E., Raza, M., Benway, J., Kiros, J., & Kumar, V. (2024, December 26). An integrated and scalable rodent cage system enabling continuous computer vision based behavioral analysis and AI enhanced digital biomarker development [Preprint]. bioRxiv. <https://doi.org/10.1101/2024.12.18.629281> v1
- [2] Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). *npj Digit Med*. 2020 Apr 14;3(1):1–15.
- [3] DIVA. DIVA. [cited 2024 Oct 15]. Home. Available from: <https://diva.bio/home/>
- [4] Baran SW, Bolin SE, Gaburro S, van Gaalen MM, LaFollette MR, Liu CN, Maguire S, Noldus LPJJ, Bratcher-Petersen N, Berridge BR. Validation framework for in vivo digital measures. *Front Toxicol*. 2025 Jan 8;6.
- [5] Baran, S. W., Bratcher, N., Dennis, J., Gaburro, S., Karlsson, J., Maguire, M., ... LaFollette, M. (2022). Emerging role of translational digital biomarkers within home cage monitoring technologies in preclinical drug discovery and development. *Frontiers in Behavioral Neuroscience*, 15, 758274. <https://doi.org/10.3389/fnbeh.2021.758274>



JAX® Mice, Clinical & Research Services

The Jackson Laboratory
Maine | Connecticut | California | China | Japan

jax.org
1-800-422-6423 (US, Canada & Puerto Rico)
1-207-288-5845 (from any location)

