

Published in final edited form as:

*J Neurosci Methods*. 2013 September 30; 219(1): 10–19. doi:10.1016/j.jneumeth.2013.05.013.

## Automated multi-day tracking of marked mice for the analysis of social behavior

Shay Ohayon<sup>1</sup>, Ofer Avni<sup>2</sup>, Adam L. Taylor<sup>2</sup>, Pietro Perona<sup>1,3</sup>, and S. E. Roian Egnor<sup>2</sup>

<sup>1</sup>Computation and Neural Systems, California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA

<sup>2</sup>Howard Hughes Medical Institute, Janelia Farm Research Campus, 19700 Helix Drive, Ashburn, VA 20147, USA

<sup>3</sup>Division of Engineering and Applied Science, California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA

### Abstract

Quantitative description of animal social behavior is informative for behavioral biologists and clinicians developing drugs to treat social disorders. Social interaction in a group of animals has been difficult to measure since behavior develops over long periods of time and requires tedious manual scoring which is subjective and often non-reproducible. Computer-vision systems that could measure complex social behavior automatically would have a transformative impact in biology. Here, we present a method for tracking group-housed mice individually as they freely interact over multiple days. Each mouse is bleach-marked with a unique fur pattern. The patterns are automatically learned by the tracking software and used to infer identities. Trajectories are analyzed to measure behavior, as it develops over days, beyond the range of acute experiments. We demonstrate how our system may be used to study the development of place preference, association and social relationships by tracking four mice continuously for five days. Our system enables accurate and reproducible characterization of wild-type mouse social behavior, and paves the way for high-throughput long-term observation of the effects of genetic, pharmacological and environmental manipulations.

### Introduction

Mouse models have been recently developed to study cognitive and social deficits observed in autism (Jamain et al., 2008; Penagarikano et al., 2011), schizophrenia (Hikida et al., 2007; Tremolizzo et al., 2002), down syndrome (Olson et al., 2004; Reeves et al., 1995) and fragile X syndrome (Kooy et al., 1996; Zang et al., 2009). Social relationships in mice develop and evolve over the course of many days (Hurst et al., 1993; Poole and Morgan, 1975). The ability to carry out thorough, quantitative, long-term observations would likely have transformative effects in understanding and measuring social behavior and its pathologies. Yet, widely used assays are often performed for short durations that can miss the persistent durable trait (Kabra et al., 2013). A key challenge in performing long assays is the ability to obtain reliable annotation. However, it is not practical to have it done by human experts because it is tedious, expensive and not easily reproducible (de Chaumont et al., 2012; Spencer et al., 2008). Computer vision systems that are able to analyze animal

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

behavior automatically hold much promise (Reiser, 2009). Despite recent progress, state-of-the-art computer vision systems are limited to the observation of two mice sharing an unfamiliar enclosure for a period of 10–20 minutes, often in partition cages, which limit social interaction (de Chaumont et al., 2012; Spencer et al., 2008). Significant progress has been reported recently on classification of actions, once animal trajectories are computed (Burgos-Artizzu et al., 2012; de Chaumont et al., 2012; Jhuang et al., 2010). However, reliable tracking and animal identification of multiple mice sharing the same enclosure for several days remains an open problem.

Keeping track automatically of the identity of multiple animals in a video sequence is difficult. Current approaches are either based on the assumption that the animals are always visible, do not overlap, and do not move too quickly or employ heuristics, such as size difference across animals (Dankert et al., 2009), constrained environments (Branson et al., 2009) or artificial colored markers (EthoVision, Noldus) to resolve animal identities. Attached colored markers are easily groomed out and are not discriminable in infrared lighting which is required for observation during dark cycles. All of the above approaches can fail and require human verification and correction of results (de Chaumont et al., 2012). Furthermore, mice have flexible bodies and are highly interactive (cuddling, chasing, jumping on top of each other, mounting, etc.) and live in fairly complex environments (e.g. involving nests and bedding into which mice burrow, which makes them invisible to the camera for periods of time), making tracking and identification challenging, particularly when prolonged observation of social behavior is desired.

We present a method capable of tracking individual mice interacting socially in a group over days without confusing identities; identities are maintained even when individuals hide and burrow in the bedding. The method consists of a single-camera computer vision system that learns automatically the appearance of each mouse and uses it to infer each animal's identity throughout the experiment. We developed a set of uniquely discriminable patterns for marking the back of each animal; the patterns are produced by applying harmless hair bleach to the fur, they cannot be groomed out and can be tracked under infrared illumination during both dark and light cycles. The trajectories computed by our system may be used to detect and quantify mouse social behavior (courtship, aggression, dominance, etc.), and study its evolution over days. The system is easily reproducible, inexpensive, user-friendly, and scalable allowing high-throughput. Using our system we characterize how social interaction develops in groups of four wild-type mice (two males and two females) over a five day period.

## Results

### Method Overview

Recognizing individual mice from overhead pictures is difficult for both human observers and machines. To overcome this limitation, we developed a method to apply a distinct pattern on the back of each mouse using hair bleach (see Fig. 1a, Methods). After patterning, each mouse is filmed alone for 5–10 minutes to collect diverse samples of its appearance during normal behavior (Fig. 1b, 1c). The samples are then used to train image classifiers (one per mouse). All mice are then placed together in the same enclosure for the actual study, where they are video-recorded continuously for five days under infrared lighting. A purpose-built computer vision system tracks the position of the mice and computes their trajectories (Fig. 1d). As a final step, the system computes mouse identities for each trajectory using Bayesian inference (Fig. 1e). On a single CPU processing of a video frame is ~300 ms (10x slower than real time). Processing can be done on a computer cluster to improve performance. Processing a five-day video (at 30 FPS) takes approximately 12 hours

on a cluster of one-hundred 2.66GHz four-core processors. Short sequences (1–2 hours) can be easily analyzed on a single computer overnight.

### Mouse patterns

Inspired by natural occurring patterns from the animal world (Gordon, 1985), as well as patterns used in error-correcting codes (Blahut, 2003), we designed and tested more than a dozen different patterns, ten of which are presented in Figure 1a. Patterns included large spots and thick stripes at different orientations and positions. Many more patterns can be generated using the same dyes. Our goal was to design patterns that are easy and quick to draw reproducibly on the back of mice, and highly discriminable from each other. Patterns on the fur slowly fade due to dark hair regrowth but remain visible for almost three weeks.

To train our computer vision system to identify mice, we filmed each patterned mouse alone for several minutes (5–10 min) as it was exploring the arena. Our tracking algorithm detected the position and orientation of the mouse in each frame and extracted a small image patch, centered and aligned on the mouse (Fig. 1a, Supplementary Fig. 1, see Supplementary Text). Dense Histogram of Gradients (Dalal and Triggs, 2005) (HOG) features were extracted from each image patch and were used to train a classifier to discriminate each mouse from all other mice patterns (1 vs. all, see Supplementary Fig. 2, Supplementary Text).

The performance of each pattern classifier was then evaluated in a cross validation procedure ( $k=4$ ), by testing it against patterns from all ten mice (10k samples per mouse) in order to discover which patterns are maximally discriminable.

We found that most patterns could be discriminated with high accuracy. The average true positive rate (TPR) was  $0.9\pm 0.04$  and the average false positive rate (FPR) was  $0.01\pm 0.06$  (confusion matrix in Fig. 1f). However, we found that some patterns more easily confused than others. For example, pattern five (two vertical stripes) was more likely to be confused with pattern eight (three vertical stripes). A manual inspection of misclassified samples revealed that errors occurred when patterns were heavily deformed (due to the flexible nature of a mouse body), partially obscured or completely occluded. This typically occurred when mice sat or reared.

To find the optimal set of four patterns, we tested all possible pattern quadruplets and computed for each quadruplet the error frequency (average false positive + false negative) (Supplementary Fig. 3a,b). We found that many quadruplets of patterns produce roughly the same performance level (the top ten combinations are given in Supplementary Fig. 3c), indicating that the method is relatively robust to the particular patterns used. For all our experiment we chose patterns 1–4 (Fig. 1a).

Small image patches obtained from videos showing only one mouse in the imaging setup (“solo samples”, Fig. 1g) contained less variability compared to samples obtained from a video with four mice in the imaging setup (“group samples”, Fig. 1i). Classifiers were trained on solo samples, requiring no human annotation. Classifiers performed well on solo samples (Fig. 1h, average TPR  $0.96\pm 0.01$ ); their performance dropped when tested on group samples (average TPR  $0.88\pm 0.13$ , Fig. 1j). Thus, frame-by-frame classification was not always reliable, due to occlusion and to large variations in appearance (Fig. 1i), suggesting integration of information from multiple frames is needed to accurately recover identities.

### Detection and tracking

The function of the tracker in our system is to detect and track the pose (position and orientation, modeled by an ellipse) of multiple mice, without concern for identity

(Supplementary Text, section 3). The tracker works incrementally from the beginning to the end of the video. For each new frame, the pose of mice from the previous frame is extrapolated and perturbed randomly to generate multiple hypotheses regarding mice position in the current frame. Multiple instances of the Expectation Maximization (Bishop, 2006) (EM) algorithm are initialized with these random hypotheses to estimate the most likely pose in the current frame. The best fitting hypothesis is then selected as the current pose, and it is associated to the corresponding pose in the previous frame. Tracking of a mouse stops when not enough pixels are available (e.g. when it burrows in bedding) and reinitializes when new unassigned pixels appear (e.g. when the mouse emerges from the bedding). Multiple mice disappearing and reappearing (e.g. due to burrowing) do not pose a problem since identities are resolved in a later step (see Supplementary Text, Supplementary Fig. 4). The process is repeated for all frames in the video in a single pass from the beginning to the, thus obtaining four trajectories. In order to speed up processing, the video is automatically split into shorter segments which are processed in parallel on different computers (see Supplementary Text, section 3.1,3.3, Supplementary Fig. 5).

Each trajectory obtained from the tracker may track different mice at different time instants; this is because when two mice meet and tumble together the identities may swap. These identity errors are resolved in the next step by using the patterns on each mouse.

### Propagating Identity Information

Once trajectories are obtained (previous step) the mouse identity classifiers are used to assign identities to the mice that are associated to each trajectory at each frame. A good identity assignment is one where at each frame the mouse's identity is consistent with its appearance, and each mouse's trajectory is smooth.

Our system uses a Hidden Markov Model (HMM) to associate the most likely mouse identities to each trajectory at each frame. The model is defined over all possible assignments of trackers to identities. For example, given a frame with four mice, there are 24 (4!) possible ways to assign identities to the four detected ellipses (two of such assignments are shown in Fig. 2a, each identity is color coded). The identity classifiers assign probabilities to each identity assignment. The probability of transitioning between one identity assignment to another is low when the mice are well separated in space and high when the mice are very close to each other (Fig. 2a, Supplementary Text, section 4.3). The probability of each identity assignment, purely based on frame-by-frame appearance-based identity classification, is shown in Figure 2b for a short (15 minute long) sequence. Each row corresponds to an identity assignment, and each column represents a frame. States with high identity probability are denoted in red.

Selecting the most probable identity (ID) assignment in each frame, purely based on mouse appearance, results in a jagged solution (see pink outline in Fig. 2c) because the most probable identity of the mouse in each trajectory changes frequently when visual classification is ambiguous. Comparison to ground truth identities shows that selecting the most likely assignment frame-by-frame has about 10% error rate. The HMM uses the additional constraint that cross-trajectory swaps are only likely when two trajectories come very close (i.e., see example in Fig. 2d), thus it computes a better assignment of identities yielding 100% correct ID (Fig. 2e).

### Validation

In order to evaluate our system's performance we classified each mouse as huddled, when it was in close contact with another mouse, and non-huddled otherwise (see methods and Fig. 3b). Huddled mice are typically sleeping clustered together and very difficult to tell apart,

posing a difficult problem both for correct segmentation and for identification for both human and automatic annotators. This has little effect on behavior analysis: since huddled mice are mostly sleeping, their behavior is easily classified even when identification is uncertain. By contrast, correct mouse identification in non-huddled events is crucial to studying individual and social behavior. Huddling events were abundant and account for 55% of the video frames. They are much more frequent during the light cycle (when mice are less active) than the dark cycle, and increase in number over the course of the five day experiment (Fig. 3a).

We quantified the performance of our system in estimating mouse pose and found it has comparable performance to human annotators, whether mice are huddling or not. To this end, we trained two human observers to draw tight ellipses around the bodies of mice in 470 frames randomly sampled from our video recordings. We found that the average discrepancy in determining position of a mouse between two human annotators was  $1.6 \pm 0.8$  mm while the discrepancy between a human annotator and machine was  $1.8 \pm 2.8$  mm (see Supplementary Fig. 6, Supplementary Text, sections 6–7).

We also measured the accuracy of our system in classifying mouse identities over long periods of time. A human annotator manually labeled mouse identities in hour-long sections during the dark and light cycles spanning five days (Fig. 3b). We compared the annotator-determined identities with those computed by our algorithm on one frame every 5 seconds in the annotated sections. Overall, 34416 mouse images were manually annotated, which amounts to 12 hours of annotated video (out of a total of 120).

Mice were correctly identified during non-huddling in 97.3% (19649/20193) of the images. Performance was approximately constant across the five days of the experiment. Identification errors (2.7%) were in part due to segmentation errors (Fig. 3c). Huddled events posed a much harder problem for our system; we found that 58% (8262/14223) of those frames contained correct segmentation and correct identities, while the remaining 28% of the mouse images were poorly segmented and 13% were properly segmented, but were assigned incorrect identities. Thus, our system was capable of maintaining correct identities during active behavior over days, spanning both dark and light cycle and errors were almost entirely limited to mice who were huddled together and motionless.

To further evaluate the performance and generalization of our system we recorded 12 continuous hours of video of six mice in the imaging setup during a dark cycle. We ground-truthed the video by manually annotating mouse identities every 30 seconds regardless of huddling condition. Out of 8400 annotated mouse images, 99.4% were properly segmented and correctly identified, 0.3% were assigned incorrect identities and 0.3% were segmentation errors (Fig. 3d).

Fighting behavior can often involve very rapid movements, as mice jump and tumble over each other. We identified several fighting bouts in one of our 5 day sequences by thresholding mouse velocity. Out of 10 randomly selected fighting bouts (four are shown in Supplementary Fig. 12), only 5% of the frames contained incorrect identities of the fighting mice. In all cases, identities were correct just before and just after the fight. Fights typically lasted 15–60 frames (0.5–2 sec).

### Development of social behavior in wild-type mice

We characterized the behavior of six sets of four C57BL/6J wild-type mice (two brothers and two sisters) over five days. Males and females had been housed separately prior to the experiment allowing us to observe how social hierarchy develops when mice are grouped together for the first time. At the beginning of recording the mice were added to a large (.6x.

6x.6 m) home cage equipped with food, water, and two tubes shelters (see Fig. 1a, Supplementary Fig. 7).

After capturing five-day videos (12,960,000 frames), we used our system to compute the trajectories of each individual over the entire period. We analyzed the trajectories by calculating statistics (places visited, velocity, and distance between mice) as well as by detecting actions. For the latter task we employed JAABA, a freeware software tool for detecting behaviors in animal trajectories (Kabra et al., 2013).

Figure 4a shows how much time mice in the first set spend at any given location in the enclosure. The four corners, the entrance to the tubes and inside the tubes are preferred locations (Fig. 4b). A similar pattern was observed across multiple experiments (Supplementary Fig. 8). Fig. 4c shows a histogram of times spent at these locations. We find that mice switch, as a group, between the two tubes during the light cycle (events marked by white arrows in Fig. 4c). We observed this phenomenon in all groups; it appeared to be spontaneous and not associated with human presence or disturbance. Also, over days they tend to spend more time at one of the corners (in this case, the bottom left, see Fig. 4c).

Mice spent overall less time at the corners compared to the tubes and tube entrances ( $p < 0.0001$ , U-test, Supplementary Fig. 9a). This is true for all mice in all experiments except one male in Experiment 5 (Supplementary Fig. 9a, fourth experiment column). Mice spent more time at the corners on the last day compared to previous days ( $p < 0.05$ , U-test, Supplementary Fig. 9b).

To quantify how groups are formed and which ones were more frequent we counted all possible mice group configurations. We considered two mice to be in the same group if the minimal distance between their ellipses was smaller than half their body width. Given four mice, 15 possible group configurations are possible ranging from all mice forming a single group (Fig. 5a, first row, group configuration #1) to every mouse in isolation (Fig. 5a, last row, group configuration #15). We found that mice spent the majority of their time during the first dark cycle on their own (Fig. 5b, top). However, this behavior gradually changed and mice spent less and less time in isolation over the next days. We found this trend to be significant ( $p < 0.01$  one-way ANOVA). Two-way ANOVA factoring the husbandry conditions for each experiment (standard or enriched) did not reveal any significant effect of rearing conditions on this behavior ( $p < 0.001$  for day,  $p > 0.5$  for husbandry). We also observed significant increase in the fraction of time mice spent all together, and, again, no difference between husbandry conditions (Fig. 5b,  $p < 0.001$  for day,  $p > 0.1$  for husbandry, two factor-ANOVA, Fig. 5b bottom). These changes in group composition suggest that the social relationships of the mice were developing continuously throughout the five days experiment.

Preferred location and preferred associates in a group are passive proxies of social preference. To investigate active behaviors we quantified social interaction by focusing on male following behavior (e.g. both male-following-male and male-following-female; see Supplementary Text for follow classifier details). An example of male following is shown in Figure 6a. In both standard and enriched conditions following behavior was strongly circadian, with the vast majority of follows occurring during the dark cycle (Figure 6b,  $p < 0.006$ ). In all cases the largest number of follow events occurred in the first dark cycle. In the enriched condition cages (Exp 4,5 and 6) intermediate levels of following were maintained over the five days, while in two of the three standard condition cages (Exp 1 and 2) follow rates dropped to very low levels after the first dark cycle, suggesting a reduction in social interaction in these cages. Follow duration and speed distributions were similar across experiments (see Supplementary Figure 10a,b).

It has been shown that male mice develop dominance relationships, in which one male is both successful in agonistic interactions and has more mating opportunities (Dewsbury, 1981) and higher reproductive success (D'amato, 1988; Hurst et al., 1993) than subordinates. We wondered whether following behavior would also display a similar asymmetry between males, with the prediction that one male would do the majority of following, both of the other male and of the females. To explore this possibility we developed two following indices: one based on male-male following behavior and one based on male-female following behavior (see Methods). The male-male index is based on the amount of time each male spent following the other male, such that a value of +1 indicates that all the male-male follows were performed by male 1 following male 2, while a value of -1 indicates that all the male-male follows were performed by male 2 following male 1. An example of the male-male index as a function of time is shown in figure 6c (open circles, data from Exp 1). Both males started by following each other equally (index close to zero), but as time progressed male 1 spent more time following male 2. The male-female index was computed similarly as the amount of time each male spent following the females (see Methods). We also observed gradual increase in the female follow index of male 1 over the first 12 hours (Fig. 6c, filled circles).

We then plotted the male and female follow indices against each other for every hour, to produce a follow index graph (see figure 6d). In order to simplify comparison across cages, we designated the male with the higher male-male index in the first twelve hours as male 1, and the other as male 2. If the male-male and male-female indices are correlated and stable, all values of male and female follow indices should be greater than 0 and should result in points in the upper right hand corner of the follow index graph (as in Fig 6d, first dark cycle of Exp 1). The follow index graph for all six cages is shown in figure 6e. In all the enriched cages (Exp 4–6), the male-male and male-female follow indices were greater than zero from the first block—indicating that a single male was responsible for the majority of both the male-male follows and the male-female follows, while all of the standard cages had values outside the upper right-hand corner in the first dark cycle, indicating that male-male behavior and male-female behavior were not completely correlated at first. By the end of the first dark cycle (12 hours), however, all six cages had resolved to have male and female follow indices in the upper right-hand corner.

The previous analysis focuses on following behavior, detected using the output of our tracker to train a behavioral classifier. It is important to note that many different behaviors could easily be quantified using this system. For example, the system can also be used to detect simple behaviors like walking (Kabra et al, 2013) or more complex behaviors like mating events (see Supplementary materials).

## Discussion

We developed a method for tracking multiple socially interacting, individually identified mice for multiple days without confusing their identities. Our system is fully automated and requires minimal human intervention. Our method integrates information over time and reliably computes the identity of each mouse even for video frames where instantaneous identity is difficult to discriminate due to pattern occlusion or deformation. We demonstrated the applicability of our system by tracking several groups of four mice over a five days period and observing how behavior evolves over hours and days. In order to verify the applicability of our method to different numbers of mice, we computed trajectories in a six-mouse cage with excellent identification performance.

We measured proxies of social behavior (preferred location, group setting, following) and found them to change across days. In addition, we found that there are no differences

between standard-reared and enriched-reared mice for simple social metrics like group association, but there were differences in more complex metrics, such as male and female following behavior. The lack of difference between standard and enriched cages for simple association may be due to the tendency of mice to associate with each other, even across dominance relationships (Uhrich, 1938). This observation underscores the importance of quantitative and detailed behavioral descriptions in untangling social deficits. Such behavior would be difficult to assess in a short-term. In addition, our method was able to demonstrate that animals that experienced enriched rearing environments more quickly adopted consistent social roles, an observation that has been previously made using labor-intensive manual scoring (Branchi et al., 2006).

Our method was designed with cost and reproducibility in mind. It is based on a single overhead camera to reduce the need to store and process multiple video feeds. Processing long videos (days) is fast on a large computer cluster but shorter experiments (spanning few hours) may be analyzed on a single CPU.

The ability to keep track of correct identities over long periods of time opens up a wide range of possibilities for developing new assays to study aggression and courtship. We expect that our system will prove to be a valuable tool in genetic screening by allowing examination of the effects of genetic, pharmacological and environmental manipulations on long term social behavior.

## Materials and methods

### Animals

Male and female C57Bl/6J mice (Jackson Labs), aged 6–17 weeks, were used. Prior to recording, two female mice (sisters) and two male mice (brothers) were housed in separate cages. Mice were raised in either standard or enriched conditions. Standard reared mice were acquired from Jackson Labs at 3 weeks of age and housed in same-sex pairs (siblings) in large mouse cages until the recording session. Enriched reared mice were born as the second of three litters into a large ( $0.61 \times 0.61 \times 0.61 \text{m}^3$ ) population cage with two adult males and two adult females. Enriched reared mice were removed from the population cage at 3 weeks of age and housed in same-sex pairs (siblings) in large mouse cages until the recording session.

We exposed the female mice in the study to bedding from the males to be used in the study, at least 7 days in advance of recording, to insure that the females were cycling regularly (Whitten, 1959). Vaginal smears from both females were then collected and used to determine estrus state. Recordings were started when both females were in proestrus. Mice always had *ad libitum* access to food and water.

### Fur patterns

Individually distinctive patterns were bleached into the fur of the mice. Mice were anesthetized with isoflurane (2%) in an induction chamber. Lab tape was used to mask out a chosen pattern on the back of each anesthetized mouse. Human hair bleach (Clairol niceneasy Born Blond Maxi) was mixed using the manufacturer's instructions. Bleach was applied to the top of the fur only to avoid irritating the skin. The tape was removed and the mice were maintained under anesthesia (1.5–2% isoflurane) for 20 minutes. The bleach was then rinsed thoroughly using warm water, the fur was dried and the mice were placed in a heated cage to recover from anesthesia.

**Mouse enclosure and recording equipment**—Mice were housed in a  $0.61 \times 0.61 \times 0.61 \text{m}^3$  polycarbonate population cage. Bedding was a 25%/75% mix of corn cob and alpha-dri (Shepherd). Shelters for the mice were custom-made square-section tunnels made of IR-transparent acrylic (cylindrical-section tunnels distorted the image of the mice within the tunnel, degrading tracking performance). Video was recorded using an overhead Basler A622f monochrome 1394 camera (16 mm Fixed Focal Length Lens with Manual Focus and Iris, C-mount, 2/3" format, F-stop: 1.4, Filter: 25.5mm, Pitch: 0.5, graftek.com; part # HF16HA-1B). The camera was placed centrally facing downwards, approximately 120cm above the cage floor (see Supplementary Fig. 7). Illumination was provided by four infrared LED light sources placed adjacent to the camera (IR-LT30, 850nm, 30 degree beam, Reytec Imaging). Because the mice were filmed continuously across multiple days and were on a 12 hour day/night cycle, an infrared-pass filter (Hoya RM72 Infrared filter, B&H Photo; OIR7252) was used to minimize the effect of changes in ambient illumination in the recordings as the room lights turned on and off. Video recording was monitored from an adjacent control room. Video (30 Hz, 1024×768 pixel image) was streamed continuously to an external hard drive using StreamPix 5 software (Norpix). Camera gain and black level were adjusted ahead of the experiments to obtain good contrast between the mice and the background, without saturating the mice.

We recorded the four mice for five days, and then recorded single-mouse video used to train the mouse classifiers. This was done so that all mice would be new to the enclosure at the beginning of the experiment.

### Huddled mice

We define an image of a mouse as “huddled” if the minimal distance between the mouse ellipse and the closest other ellipse is smaller than a pre-defined threshold set at 6mm and mouse velocity is smaller than 3 pix/frame (7.2 cm/s).

### Follow Index

We define the male and female follow indices as follows:

$$\begin{aligned} \text{male-male follow index} &= \frac{m1m2 - m2m1}{m1m2 + m2m1} \\ \text{male-female follow index} &= \frac{m1f - m2f}{m1f + m2f} \end{aligned}$$

Where  $m1m2$  is the amount of time male 1 spent following male 2,  $m2m1$  is the amount of time male 2 spent following male 1,  $m1f$  is the time male 1 spent following females and  $m2f$  is the time male 2 spent following females.

### Statistical methods

Duration and speed distributions for follow events were compared using paired Kolmogorov-Smirnov tests, with Bonferroni correction for multiple comparisons. Comparison of follow number were two-factor repeated measures ANOVAs.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

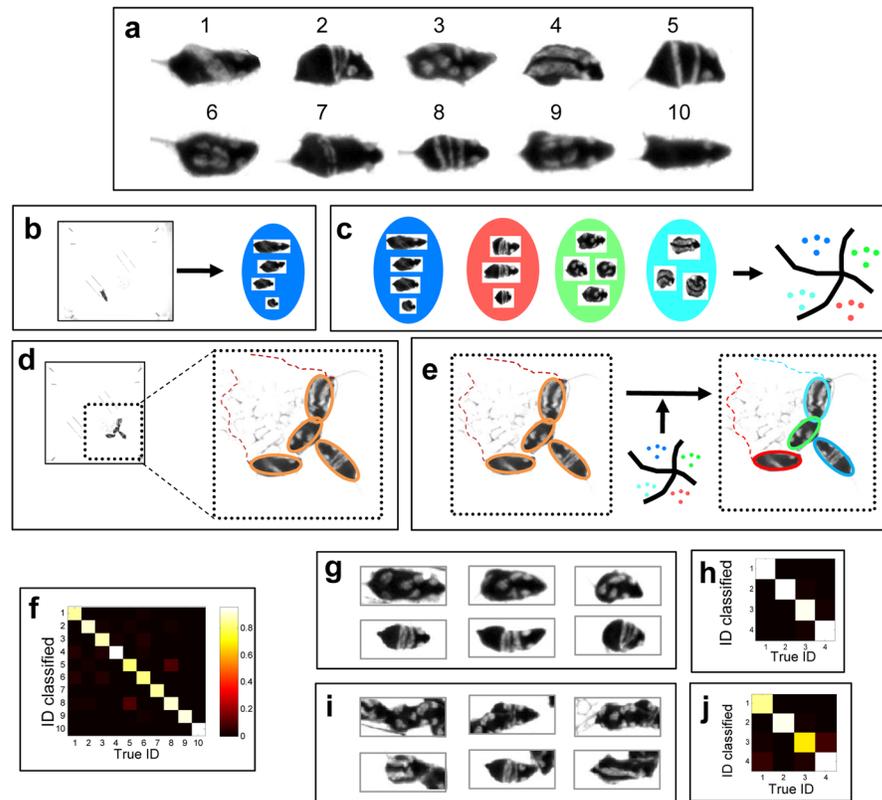
### References

- Bishop CM. Pattern Recognition and Machine Learning. 2006  
 Blahut, RE. Algebraic Codes for Data Transmission. Cambridge University Press; 2003.

- Branchi I, D'Andrea I, Fiore M, Di Fausto V, Aloe L, Alleva E. Early social enrichment shapes social behavior and nerve growth factor and brain-derived neurotrophic factor levels in the adult mouse brain. *Biol Psych*. 2006; 60:690–696.
- Branson K, Robie AA, Bender J, Perona P, Dickinson MH. High-throughput ethomics in large groups of *Drosophila*. *Nat Methods*. 2009; 6:451–457. [PubMed: 19412169]
- Burgos-Artizzu, XP.; Dollar, P.; Lin, D.; Anderson, DJ.; Perona, P. Social behavior recognition in continuous video. Paper presented at: IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2012.
- D'amato FR. Effects of male social status on reproductive success and on behavior in mice (*Mus musculus*). *J Comp Psychol*. 1988; 102:146–151. [PubMed: 3396312]
- Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. Paper presented at: "International Conference on Computer Vision & Pattern Recognition; San Diego: IEEE Computer Society; 2005.
- Dankert H, Wang L, Hoopfer ED, Anderson DJ, Perona P. Automated monitoring and analysis of social behavior in *Drosophila*. *Nat Methods*. 2009; 6:297–303. [PubMed: 19270697]
- de Chaumont F, Coura RD, Serreau P, Cressant A, Chabout J, Granon S, Olivo-Marin JC. Computerized video analysis of social interactions in mice. *Nat Methods*. 2012; 9:410–417. [PubMed: 22388289]
- Dewsbury D. Social Dominance, Copulatory Behavior, and Differential Reproduction in Deer Mice (*Peromyscus maniculatus*). *Journal of Comparative and Physiological Psychology*. 1981; 95:880–895.
- Gordon RD. The Coccinellidae (Coleoptera) of America north of Mexico. *Journal of the New York Entomological Society*. 1985; 93:1–912.
- Hikida T, Jaaro-Peled H, Seshadri S, Oishi K, Hookway C, Kong S, Wu D, Xue R, Andrade M, Tankou S, et al. Dominant-negative DISC1 transgenic mice display schizophrenia-associated phenotypes detected by measures translatable to humans. *Proc Natl Acad Sci U S A*. 2007; 104:14501–14506. [PubMed: 17675407]
- Hurst J, Fang J, Barnard C. The role of substrate odours in maintaining social tolerance between male house mice *Mus musculus domesticus*. *Animal Behaviour*. 1993; 45:997–1006.
- Jamain S, Radyushkin K, Hammerschmidt K, Granon S, Boretius S, Varoqueaux F, Ramanantsoa N, Gallego J, Ronnenberg A, Winter D, et al. Reduced social interaction and ultrasonic communication in a mouse model of monogenic heritable autism. *Proc Natl Acad Sci U S A*. 2008; 105:1710–1715. [PubMed: 18227507]
- Jhuang H, Garrote E, Mutch J, Yu X, Khilnani V, Poggio T, Steele AD, Serre T. Automated home-cage behavioural phenotyping of mice. *Nat Commun*. 2010; 1:68. [PubMed: 20842193]
- Kabra M, Robie AA, Rivera-Alba M, Branson S, Branson K. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat Methods*. 2013; 10:64–67. [PubMed: 23202433]
- Kooy RF, D'Hooge R, Reyniers E, Bakker CE, Nagels G, De Boule K, Storm K, Clincke G, De Deyn PP, Oostra BA, et al. Transgenic mouse model for the fragile X syndrome. *Am J Med Genet*. 1996; 64:241–245. [PubMed: 8844056]
- Olson LE, Roper RJ, Baxter LL, Carlson EJ, Epstein CJ, Reeves RH. Down syndrome mouse models Ts65Dn, Ts1Cje, and Ms1Cje/Ts65Dn exhibit variable severity of cerebellar phenotypes. *Dev Dyn*. 2004; 230:581–589. [PubMed: 15188443]
- Penagarikano O, Abrahams BS, Herman EI, Winden KD, Gdalyahu A, Dong H, Sonnenblick LI, Gruver R, Almajano J, Bragin A, et al. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell*. 2011; 147:235–246. [PubMed: 21962519]
- Poole T, Morgan D. Aggressive behaviour of male mice (*Mus musculus*) toward familiar and unfamiliar opponents. *Anim Behav*. 1975; 23:470–479. [PubMed: 1172394]
- Reeves RH, Irving NG, Moran TH, Wohn A, Kitt C, Sisodia SS, Schmidt C, Bronson RT, Davisson MT. A mouse model for Down syndrome exhibits learning and behaviour deficits. *Nat Genet*. 1995; 11:177–184. [PubMed: 7550346]
- Reiser M. The ethomics era? *Nat Methods*. 2009; 6:413–414. [PubMed: 19478800]

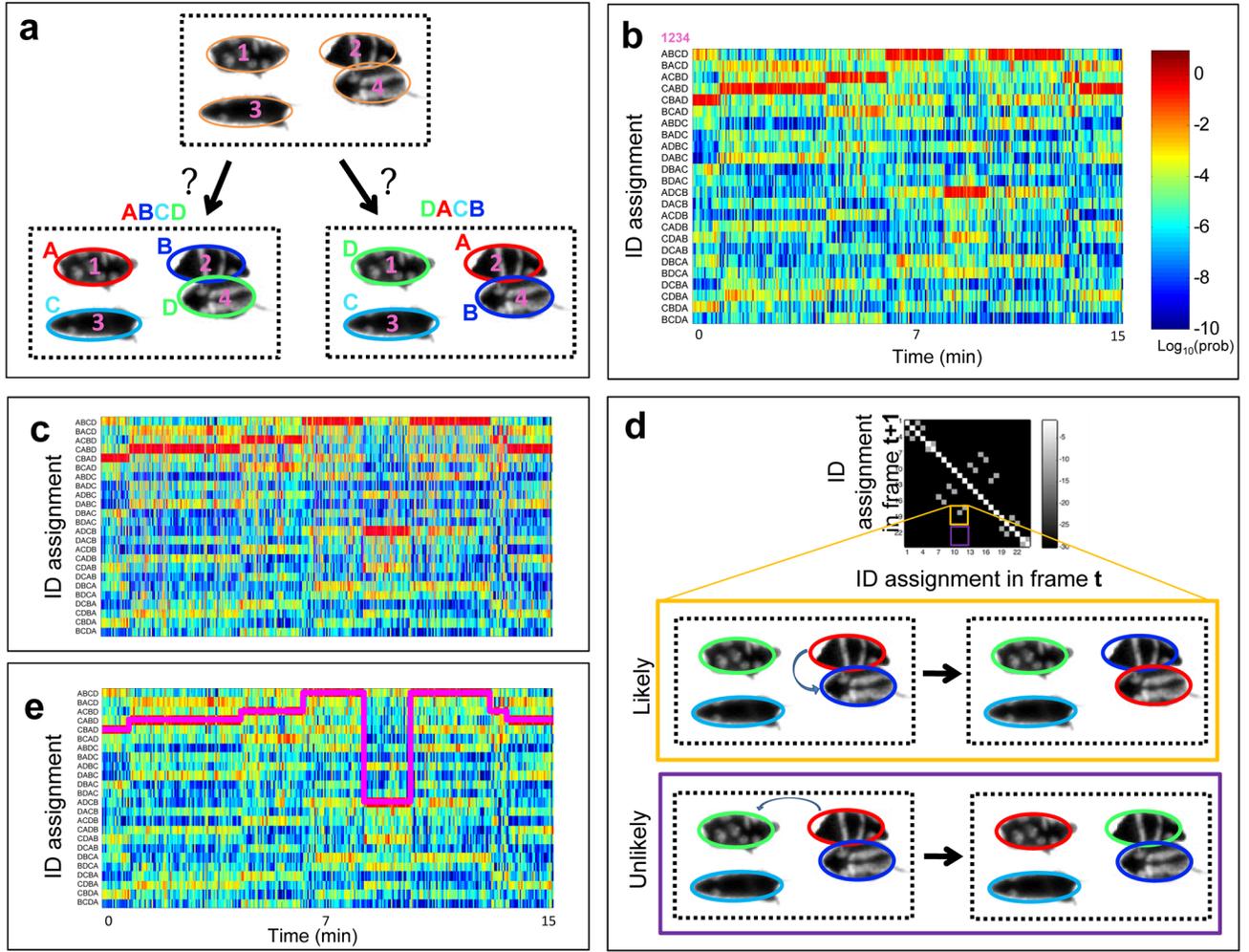
- Spencer CM, Graham DF, Yuva-Paylor LA, Nelson DL, Paylor R. Social behavior in Fmr1 knockout mice carrying a human FMR1 transgene. *Behav Neurosci*. 2008; 122:710–715. [PubMed: 18513141]
- Tremolizzo L, Carboni G, Ruzicka WB, Mitchell CP, Sugaya I, Tueting P, Sharma R, Grayson DR, Costa E, Guidotti A. An epigenetic mouse model for molecular and behavioral neuropathologies related to schizophrenia vulnerability. *Proc Natl Acad Sci U S A*. 2002; 99:17095–17100. [PubMed: 12481028]
- Uhrich J. The social hierarchy in albino mice. *J Comp Psychol*. 1938; 25:373–413.
- Whitten WK. Occurrence of anoestrus in mice caged in groups. *J Endocrinol*. 1959; 18:102–107. [PubMed: 13620841]
- Zang JB, Nosyreva ED, Spencer CM, Volk LJ, Musunuru K, Zhong R, Stone EF, Yuva-Paylor LA, Huber KM, Paylor R, et al. A mouse model of the human Fragile X syndrome I304N mutation. *PLoS Genet*. 2009; 5:e1000758. [PubMed: 20011099]

- A fully automated system to track multiple animals in a large arena without losing their identities is presented.
- The system learns unique bleach patterns on the mice fur and tracks them during both dark and light cycles.
- Identification of six mice in the experimental setup is 97% correct during non-sleep intervals.
- As a proof of principle, we track groups of four mice and report social trends that develop across hours and days.

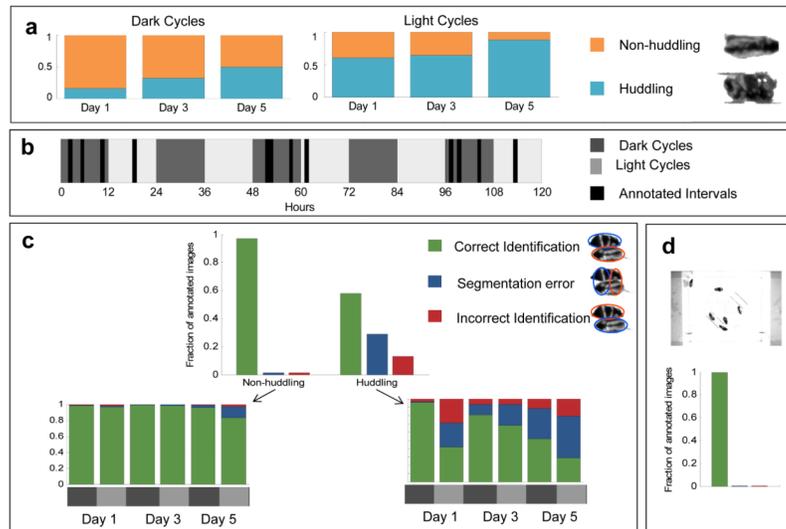


**Figure 1.**

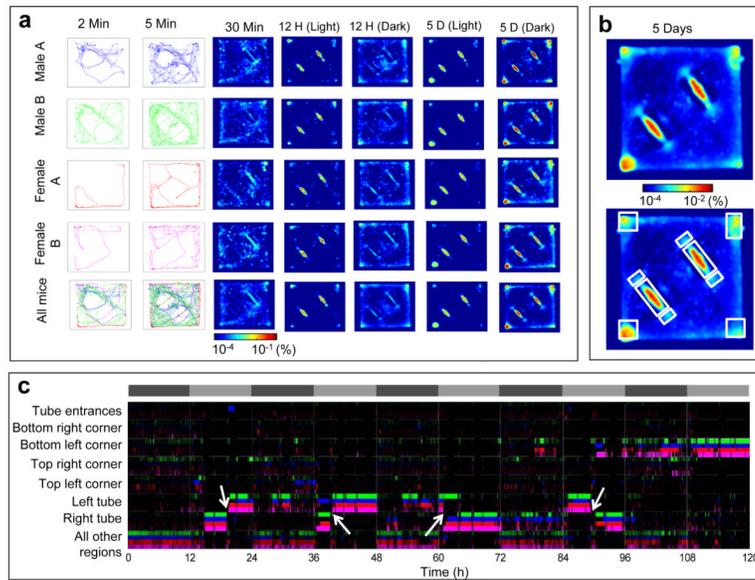
**System Framework.** (a) Ten patterns dyed on the back of mice. Each pattern is created by bleaching the fur for several minutes (see Online Methods). (b) A single mouse is placed in the imaging setup and filmed for 5–10 minutes. Multiple images of the mouse are collected. (c) The process is repeated and images are collected for each individual in the group. Using the four mice's images as a training set, a classifier is trained to distinguish individual mice. (d) The mice are tracked in the video of the experiment without using identity information, generating trajectories with possible identity swaps. (e) Information from trajectories and the learned classifiers is combined to generate correct identity-preserving trajectories. (f) Performance of mouse identity classifiers on images collected when each mouse is filmed alone. Each column represents the performance of a classifier trained to identify a single mouse. Entries on the main diagonal entries represent the true positive rate (correct identification of the mouse in a test set). Off-diagonal entries correspond to false alarm rate (incorrectly assigned mouse identity). (g) Examples of appearance variations found when mice are filmed alone. (h) Performance of four classifiers trained to identify patterns [1–4] from (a) and tested on images obtained from single mouse videos. (i) Examples of appearance variations in a 30 min long video with four mice. (j) Performance of the same classifiers in (d) on images obtained from a 30 min long video with four mice simultaneously present in the enclosure. Rather than using directly the classifier output, our method identifies mice by combining classifier's identity with trajectory information (see Fig. 2).



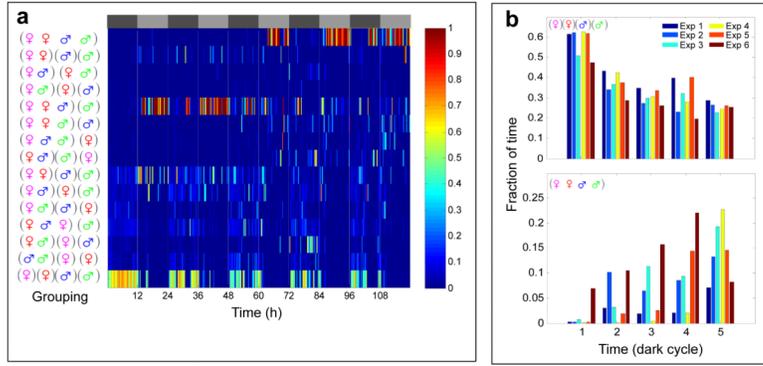
**Figure 2.** Propagating identity information. **(a)** In each frame, the identities (identified by letters and corresponding colors) of the four tracked mice (identified by numbers) are unknown. 24 possible identity assignments are possible (two are depicted). **(b)** Identity assignment probability matrix for a 15 minute video (red denotes high likelihood). Each row represents a fixed identity assignment to the four tracked mice. Each column corresponds to a video frame. **(c)** Identities selected according to the maximum likelihood found in each frame, i.e., by the classifiers shown in Fig. 1. Notice the jagged solution, suggesting that assignments are switched frequently (incorrectly) between different trajectories. **(d)** Identities can only change when mice are in close proximity. Some identity swaps are more likely than others, given the current identity assignment. For example, swapping the red and blue identities is more likely (due to their proximity) than swapping the red and green identities. **(e)** Identity likelihood computed by mouse classifiers is combined with mouse proximity by a hidden Markov model (HMM) producing correct identity assignment over the entire video sequence (piecewise-constant pink trace).



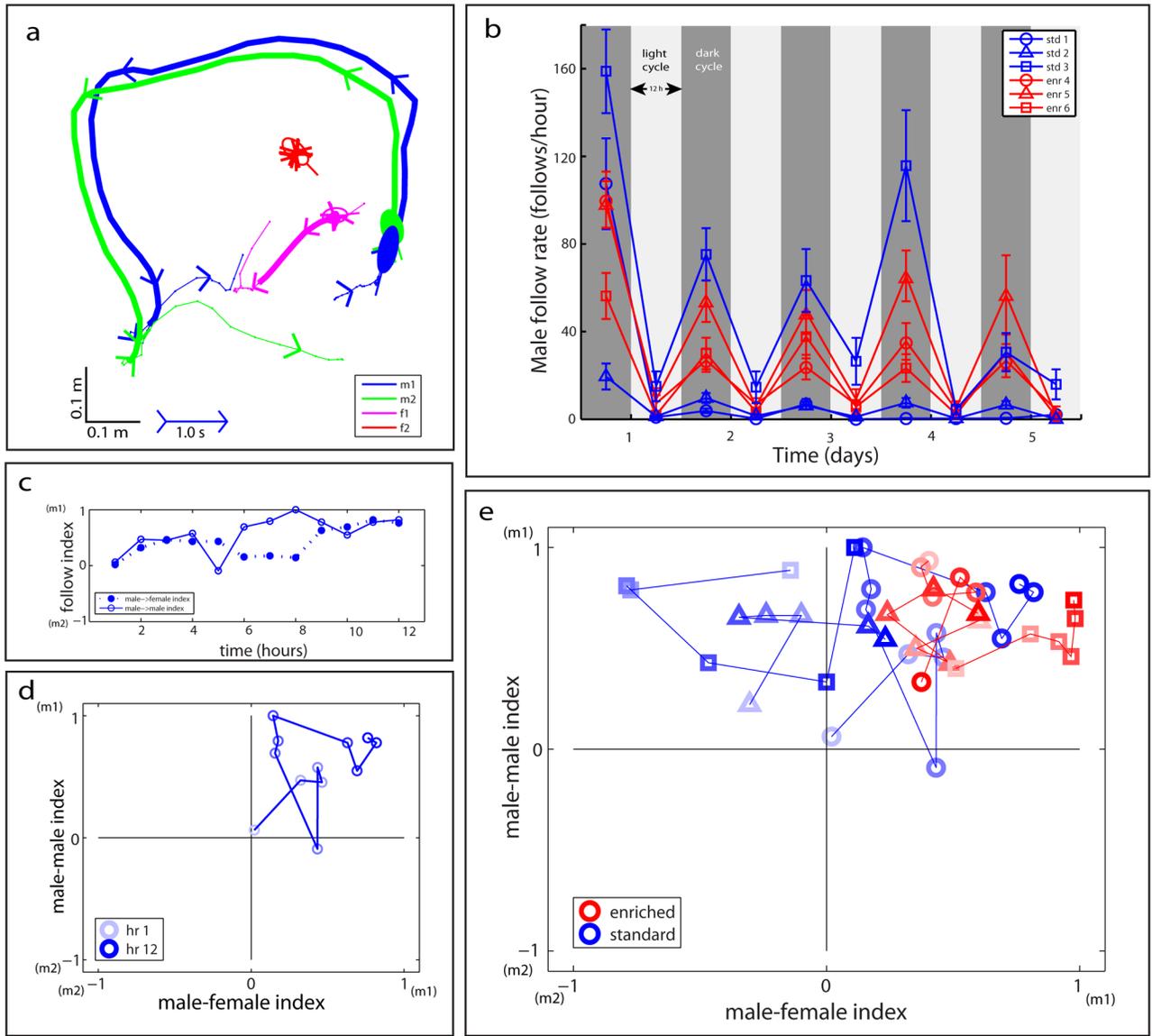
**Figure 3.** Validation of identity assignment. **(a)** Fraction of time spent on each day in a huddling configuration (orange) compared to non-huddled configuration (cyan). Left: during dark cycles, right: during light cycles. **(b)** Intervals in a 5 days test sequence for which mouse identity was established by a trained human observer in experiment 5 (black). Dark cycles are represented in dark gray and light cycles are represented in light gray. **(c)** Performance of the tracking system, measured in terms of correct identification (green), incorrect identification with correct segmentation (red) and incorrect segmentation (blue). The upper plot denotes the performance averaged across the entire five day experiment, broken down to huddling and non-huddling events. Bottom plot depicts performance as a function of days. **(d)** Performance of the system tracking six mice for 12 continuous hours during a dark cycle. Conventions same as **(c)**. A frame from the video is shown.



**Figure 4.** Mouse trajectories and dwelling places for experiment 5. **(a)** Example trajectories and position histograms for each individual mouse and for the entire group. Data presented for 2 min, 5 min, 30 min, 12 hours (first light cycle), 12 hours (first dark cycle), all light cycles (5 days), all dark cycles (5 days). Each colored histogram was constructed by computing the percentage of time spent in a given pixel. Data was smoothed and is presented in log scale for improved visualization. **(b)** 2D position histogram for all mice (top) and selected regions for monitoring (bottom, highlighted in white). **(c)** Ethogram summarizing the fraction of time each mouse spent in each of the monitored regions. Color codes denote mice identities, similar to (a). White arrows denote events in which mice changed their sleeping place from one tube to the other.



**Figure 5.** Group configuration analysis. **(a)** Ethogram denoting the percent time spent in one of 15 possible group configurations. Group is denoted by color-coded male and female symbols on the left. Dark and light cycles are denoted by gray bars on top. **(b)** Top: fraction of time spent during dark cycles in group configuration 15 (every mouse on its own). Each color denotes a different five day experiment. Bottom: fraction of time spent during dark cycles in group configuration 1 (all mice in a single group). **(c)** Difference in the fraction of time male 1 spent in a group with females (but not with male 2). Each color denotes a different experiment.

**Figure 6.**

**Male Following Behavior.** (a) Example of male 1 (—) following male 2 (—). The trajectory line is thick during the following event and becomes thin at the end of the event, time between arrows is 1s. Ellipses indicate the position of the four mice at the beginning of the follow event; sticks indicate the tails. The position and movement of the female mice is indicated by the pink and red symbols. (b) Following rate as a function of time for all six experiments. (c) Example of male-male follow index (open circles) and male-female follow index (filled circles) for the first dark cycle of Experiment 1. (d) Data from the first dark cycle in experiment 1 (standard rearing conditions). Each hour of observation is represented by an open circle. The male follow index is plotted as a function of female follow index; time is indicated by color saturation, with more saturated color later in time. (e) Data on following for all six experimental cages. Standard cages are in blue, enriched in red. All enriched cages are fully contained in the upper right-hand quadrant while each standard cage produces data points that spill into other quadrants, indicating a more complex evolution of male-female and male-male social interaction patterns.