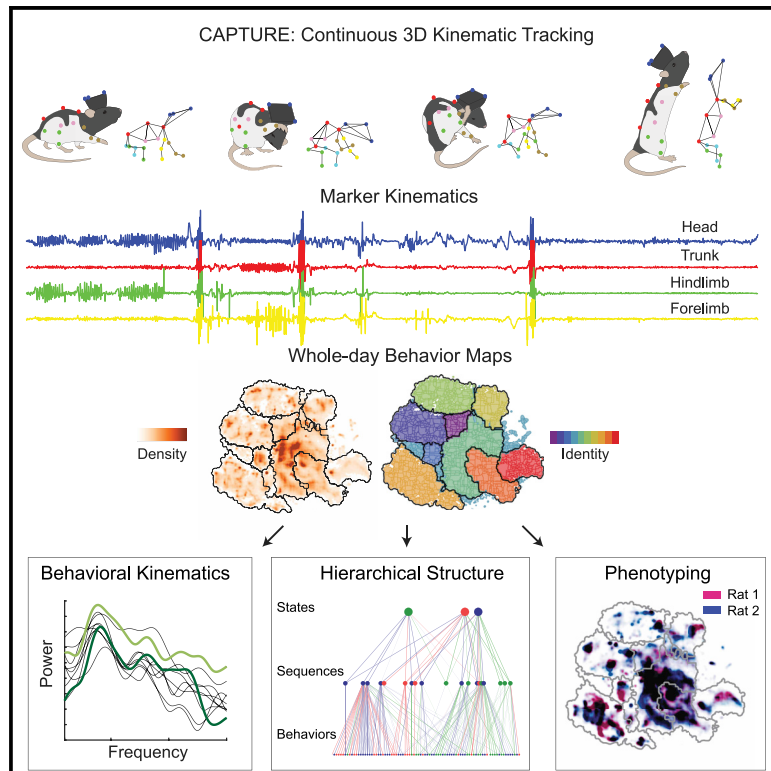


# Continuous Whole-Body 3D Kinematic Recordings across the Rodent Behavioral Repertoire

## Graphical Abstract



## Authors

Jesse D. Marshall, Diego E. Aldarondo, Timothy W. Dunn, William L. Wang, Gordon J. Berman, Bence P. Ölveczky

## Correspondence

jesse\_d\_marshall@fas.harvard.edu (J.D.M.),  
olveczky@fas.harvard.edu (B.P.Ö.)

## In Brief

Marshall et al. present CAPTURE, a new method for long-term continuous 3D motion tracking in freely behaving rats. Combining motion capture, body piercings, and deep learning, CAPTURE improves tracking precision many-fold over existing techniques. Comprehensive profiling of behavioral kinematics, usage, and hierarchical structure in normal and diseased animals is demonstrated.

## Highlights

- CAPTURE can track a rat's whole-body movements in 3D across days
- It is far superior in resolution and reliability to existing tracking approaches
- It allows comprehensive profiling of behavioral kinematics, usage, and hierarchy
- We demonstrate detailed phenotypic analysis after drugs and in *Fmr1*-KO rats



## NeuroResource

# Continuous Whole-Body 3D Kinematic Recordings across the Rodent Behavioral Repertoire

Jesse D. Marshall,<sup>1,\*</sup> Diego E. Aldarondo,<sup>2</sup> Timothy W. Dunn,<sup>3</sup> William L. Wang,<sup>1</sup> Gordon J. Berman,<sup>4</sup> and Bence P. Ölveczky<sup>1,5,\*</sup>

<sup>1</sup>Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

<sup>2</sup>Program in Neuroscience, Harvard University, Cambridge, MA 02138, USA

<sup>3</sup>Department of Statistical Science, Duke University, Durham, NC 27710, USA

<sup>4</sup>Department of Biology, Emory University, Atlanta, GA 30322, USA

<sup>5</sup>Lead Contact

\*Correspondence: [jesse\\_d\\_marshall@fas.harvard.edu](mailto:jesse_d_marshall@fas.harvard.edu) (J.D.M.), [olveczky@fas.harvard.edu](mailto:olveczky@fas.harvard.edu) (B.P.Ö.)

<https://doi.org/10.1016/j.neuron.2020.11.016>

## SUMMARY

In mammalian animal models, high-resolution kinematic tracking is restricted to brief sessions in constrained environments, limiting our ability to probe naturalistic behaviors and their neural underpinnings. To address this, we developed CAPTURE (Continuous Appendicular and Postural Tracking Using Retroreflector Embedding), a behavioral monitoring system that combines motion capture and deep learning to continuously track the 3D kinematics of a rat's head, trunk, and limbs for week-long timescales in freely behaving animals. CAPTURE realizes 10- to 100-fold gains in precision and robustness compared with existing convolutional network approaches to behavioral tracking. We demonstrate CAPTURE's ability to comprehensively profile the kinematics and sequential organization of natural rodent behavior, its variation across individuals, and its perturbation by drugs and disease, including identifying perseverative grooming states in a rat model of fragile X syndrome. CAPTURE significantly expands the range of behaviors and contexts that can be quantitatively investigated, opening the door to a new understanding of natural behavior and its neural basis.

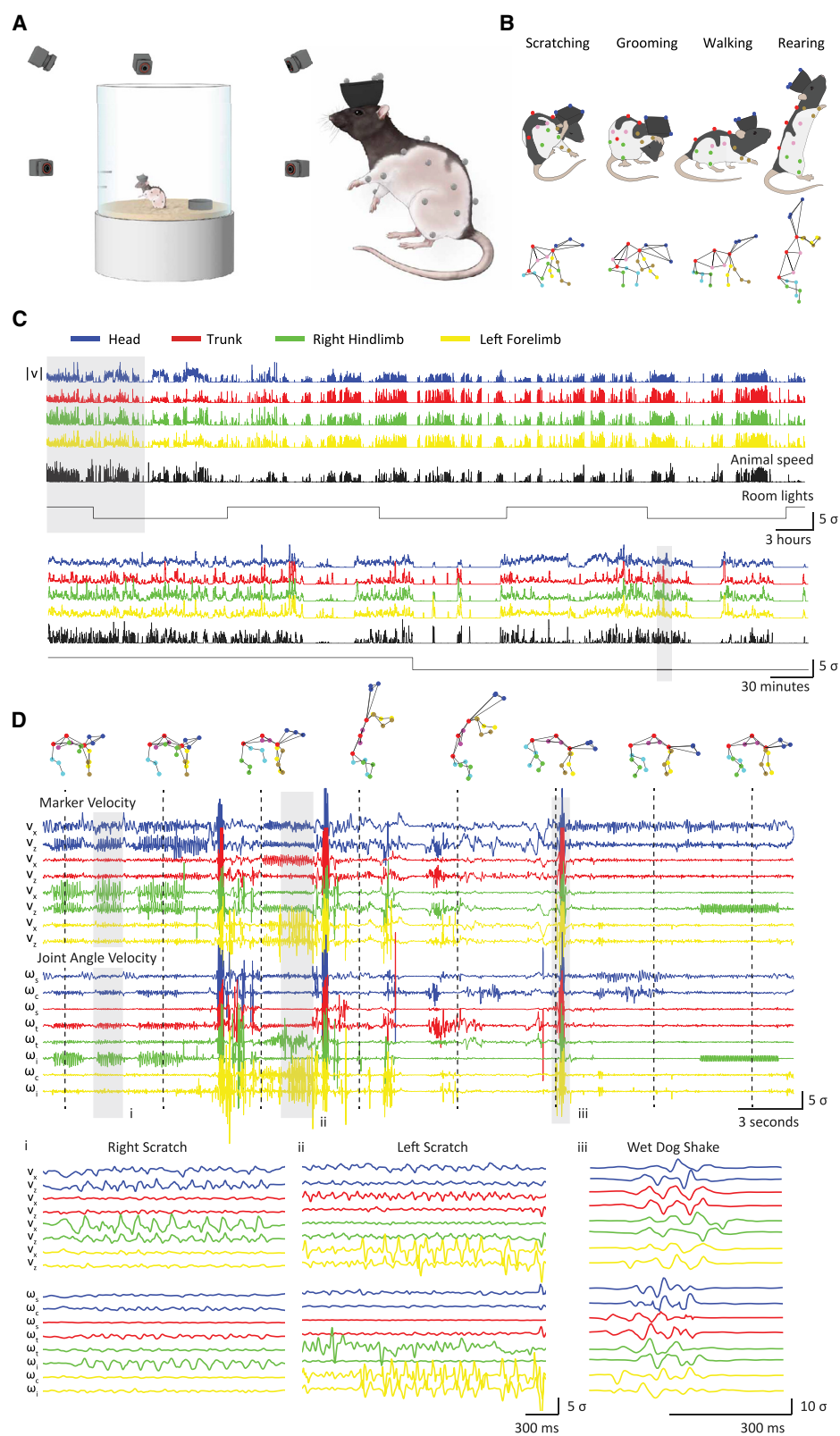
## INTRODUCTION

The overarching goal of neuroscience and psychology is to describe the neural principles and mechanisms that underlie natural behavior. A critical first step toward this goal is the development of tools and analysis frameworks capable of precisely measuring and describing the behavior of our experimental subjects (Anderson and Perona, 2014; Egnor and Branson, 2016; Krakauer et al., 2017). Advances have been made in quantifying movements and behavior in both humans and mammalian animal models during brief recording sessions in well-delineated environments (Machado et al., 2015; Mathis et al., 2018; Pereira et al., 2019). However, methods for precisely measuring behavior in naturalistic settings and over longer timescales have been lacking, permitting only coarse estimates of an animal's posture, movements, and behavioral state (Hong et al., 2015; Wiltischko et al., 2015). This technological limitation has impeded quantitative inquiries into the organization of naturalistic behavior and its neural underpinnings.

For instance, while movements, and the behaviors they produce, are thought to operate under organizational rules, much like the phonological and syntactical rules that govern language, only a handful of such rules have been described (Berridge and Fentress, 1987; Dawkins, 1976; Lashley, 1951; Tinbergen, 1950).

In contrast to most of biological research that deals with genes, cells, and species, there is no formal taxonomical structure for parsing and naming behaviors of laboratory animals, let alone their combinations into behavioral sequences or states. Precisely describing naturalistic behavior in terms of variables that are reproducible across experiments, such as body-part kinematics, could lead to the creation of rigorous and commonly accepted definitions of complex behaviors. Extending these measurements across temporal scales could lead to the establishment of a true behavioral *lingua franca*. Such standard metrics of behavior could greatly facilitate quantitative inquiries into the behavioral effects of cellular-, molecular-, and circuit-level pathologies. More generally, they could serve as biomarkers for various disease models, the utility of which is currently limited by the lack of behavioral reproducibility across laboratories (Brunner et al., 2015; Silverman et al., 2010).

Similarly, it has been hypothesized that the neural systems that control movement mirror the hierarchical structure of animal behavior, with different brain regions or spinal circuits controlling the movement of single limbs, the production of individual behaviors, and the selection of longer-timescale behavioral programs (Gallistel, 1982; Merel et al., 2019). However, while task-based studies have yielded detailed characterizations of how individual movements are produced (Dhawale et al., 2019;



(legend on next page)

Svoboda and Li, 2018), how more complex whole-body behaviors are controlled and generated remain poorly understood. Similar to how naturalistic visual stimuli were essential for illuminating the function and organization of the visual system (Simoncelli and Olshausen, 2001), effective tools for measuring naturalistic behavior should enable a new understanding of the principles by which the motor system stores and produces behavior (Merel et al., 2020).

The ideal tool for describing natural behavior, its neural underpinnings, and the effects of environmental and neural perturbations must deliver precise and continuous multiscale measurements of movement kinematics across an animal's natural behavioral repertoire. In mammals, achieving this goal requires tracking the position of points on the limbs, trunk, and head at high spatiotemporal resolution and in 3D, across diverse postures and locations. In rodents, this necessitates kinematic tracking with millimeter-scale precision and millisecond-timescale resolution, ideally continuously over hours and days to sample the full repertoire of rodent behavior and capture its long-timescale structure (Anderson and Perona, 2014; Egnor and Branson, 2016). While overhead video-based recordings or depth imaging can estimate an animal's pose in largely featureless environments, these methods fail to reliably track appendages (Hong et al., 2015; Jhuang et al., 2010; Wiltischko et al., 2015). Key-point tracking tools from computer vision and machine learning allow monitoring of visible landmarks on an animal's body (Branson et al., 2009; Machado et al., 2015; Mathis et al., 2018; Pereira et al., 2019) but to date have been limited due to an inability to robustly track occluded landmarks or across a broad range of animal postures.

Here, we present a technique capable of recording 3D movement kinematics across a rat's behavioral repertoire. By combining motion capture, deep learning, and body piercing, we achieve continuous long-term kinematic tracking of a rat's head, trunk, and appendages with substantially superior precision and robustness compared to convolutional networks. We use these continuous kinematic recordings to collect a definitive reference dataset of rat behavior, cataloging nearly every movement a rat makes over week-long timescales. To parse these recordings, we developed a machine learning analysis framework that allowed us to identify stereotyped organismal behaviors, behavioral sequences, and behavioral states. This framework allows us to describe organizing principles of natural behavior and comprehensively phenotype behavioral changes introduced by stimulants and in a rat model of fragile X syndrome.

## RESULTS

### CAPTURE: Continuous Appendicular and Postural Tracking Using Retroreflector Embedding

We sought to develop a system for continuous tracking of 3D whole-body kinematics during naturalistic rodent behavior. Because of its advantages in spatial precision, tracking speed, and data storage (STAR Methods), we used motion capture, in which a calibrated camera array tracks the position of retroreflective markers placed on a human or animal subject (Mischiati et al., 2015). While well established in humans, motion capture has seen limited use in animal models due to difficulties in stably attaching markers over long recording sessions (Courtine et al., 2008; Mimica et al., 2018; Takeoka et al., 2014). To overcome this limitation, we developed a method for chronically attaching retroreflective markers to animals using body piercings, a body modification approach also well established in humans (Seidler et al., 1992; Stim, 2003) (Methods S1). Our retroreflective markers consist of half-silvered high-index-of-refraction ball lenses ( $n = 2.0$ ) that serve as bright and durable motion-capture markers (Mischiati et al., 2015). We fuse these retroreflectors to biocompatible transdermal body piercings using high-strength epoxy. The vast majority of the piercings remain stably attached for the entirety of our experiments (Figure S1C), which can extend for weeks.

To track the rat's retroreflective body piercings, we constructed a rodent motion capture studio consisting of motion capture cameras positioned around a two-foot-diameter plexiglass arena (Figure S1A). Because markers must be seen by at least two cameras to be triangulated into 3D, we used 12 cameras to provide robustness to occlusions. To encourage behavioral diversity, we equipped it with bedding and objects and a lever for operant training (Kawai et al., 2015; Figures 1A and 1B; Videos S1 and S2).

We first assessed the ability of CAPTURE to record the kinematics of individual limbs across the rodent behavioral repertoire. We habituated our rats ( $n = 5$ ) to the arena and equipped them with the same set of 20 markers, which, once tracked, could be used to compute the position and orientation of the animal's head, trunk, forelimbs, and hindlimbs (Figure S1B). We tracked marker positions nearly continuously at 300 Hz for 1 week in each rat. Unlike depth-imaging approaches (Mallik et al., 2014), the use of bedding or objects does not interfere with our motion capture recordings, allowing the experimental arena to double as the animal's home cage (Figures 1C and 1D; Video S3). We found that our recordings showed

#### Figure 1. CAPTURE: Continuous Appendicular and Postural Tracking Using Retroreflector Embedding

(A) Schematic of the CAPTURE apparatus. Twelve motion capture cameras continuously track the position of 20 body piercings affixed to the animal's head, trunk, and limbs.  
(B) Top: schematic depictions of a rat with attached markers, colored by the major body segments tracked, engaging in different species-typical behaviors. Bottom: hypothetical wireframe representation of 3D marker positions tracked by motion capture for each of the depicted behaviors.  
(C) Speed of markers located on a single rat's head, trunk, and appendages across 72 h of near-continuous motion capture recordings (top) measured alongside the speed of the animal's center-of-mass and the state of the room lights (on or off). Investigation of kinematics on minute-long timescales (bottom) reveals alternating periods of movement and rest. Speed traces smoothed with a 30-s boxcar filter.  
(D) During a behavioral sequence of scratching, rearing, and walking (top), CAPTURE recordings reveal rhythmic modulation of body part kinematics (middle). Visualization of individual behaviors on millisecond timescales shows independent control of different appendages during scratching and wet dog shake behaviors (bottom). We defined joint angles with respect to sagittal (s), coronal (c), transverse (t), and inter-segment planes (i) (STAR Methods). Shaded regions in (C) and (D) denote expanded regions in lower panels. See also Figures S1–S3; Videos S1, S2, and S3; Table S1; and Methods S1 and S2.



sub-millimeter tracking precision ( $0.21 \pm 0.07$  mm; [STAR Methods](#)) and that segment lengths between markers remained stable over the recording session ([Figure S1D](#)), indicating that CAPTURE could reliably report limb kinematics over 1-week time spans. While tracking performance degraded slightly when using fewer cameras ([Figure S1E](#)), there is little incentive to reduce the number of cameras. This is because, unlike with traditional video camera-based systems, adding motion capture cameras does not meaningfully add to the experimental or computational effort ([Table S1](#); [Methods S1](#)).

To test whether body piercings altered the animal's behavioral repertoire, we attached a headcap fitted with retroreflectors to  $n = 3$  rats and tracked these for 2 days before piercings were attached ([Figure S2](#)). Animals with and without body piercings showed equal fractions of time spent moving in the arena and similar distributions and covariances of head velocities ([Figures S2A–S2C](#)). A classifier trained to identify the animal's behavior from the tracked movements of the headcap predicted equivalent behavioral usage before and after marker attachment ([Figures S2D–S2F](#)), altogether suggesting that piercings do not cause major behavioral changes in our animals.

Like all vision-based tracking approaches, unprocessed motion capture recordings were prone to dropouts of forelimb and hindlimb markers due to self- or environmental occlusions. The vast majority of these dropouts were brief ( $\sim 20$  ms in duration), allowing us to use standard interpolation methods based on the temporal history of marker position to faithfully reconstruct the position of dropped markers ([Liu and McMillan, 2006](#)) ([Figure S3](#)). However, as these methods do not incorporate constraints from neighboring markers or model long-timescale influences on marker position, they perform poorly for longer dropouts. To address this, we used our large collection of well-tracked motion capture data ( $\sim 25$  million frames per day) to train a more expressive deep learning architecture. In particular, we trained a temporal convolutional network ([Methods S2](#)) ([Oord et al., 2016a](#)) to predict a given marker's position using both temporal information about its past locations, as well as spatial information about the position of all other markers ([Figure S3](#)). Our imputation procedure resulted in a low ( $\sim 1$  mm) estimated median error during artificial dropout periods. Following imputation, all 20 markers were well tracked for  $\sim 99\%$  of frames when animals were active, resulting in sub-millimeter positional errors across markers ([Figures S1F–S1H](#)).

### Comparison with Key-Point Detection Using Convolutional Networks

Recently, 2D convolutional networks trained from scratch or on human-labeled pose databases have been applied to detect un-occluded visual landmarks, mostly in 2D, in individual behavioral tasks ([Mathis et al., 2018](#); [Pereira et al., 2019](#)). To test how our approach compares to these established networks in terms of 3D tracking across a larger range of behaviors, we quantified the tracking accuracy of DeepLabCut relative to CAPTURE. We first used the recommended approach of fine-tuning a pretrained network using a small number of hand-labeled frames (225) of data. We then used the fine-tuned network to detect key-points in frames from six synchronized, calibrated cameras, which we then triangulated across cameras to produce estimates of the

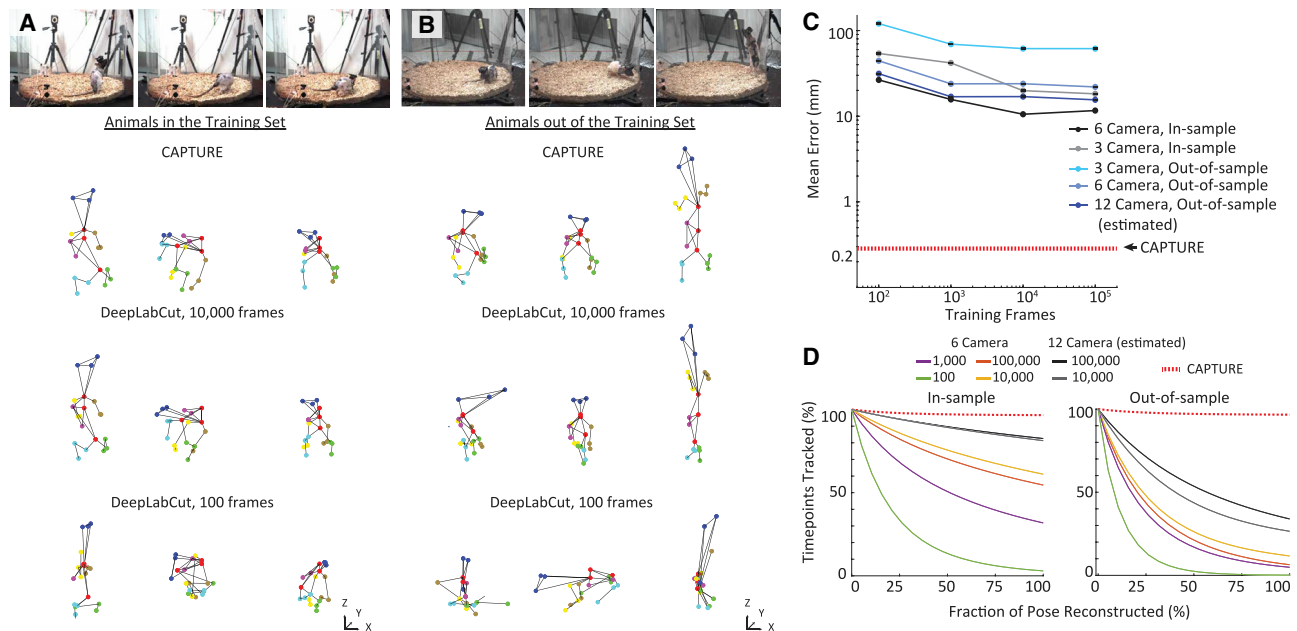
animal's 3D posture. Inspection revealed the key-point predictions to be poor, showing substantial deviation from human key-point labels on a held-out test dataset ([Figure S4](#)). These predictions were worse on the appendages or when using predictions from three cameras, indicating that networks especially struggled to track occluded markers.

It could be that these networks simply needed more training data to track naturalistic behavior in 3D. To test this, we trained DeepLabCut using 100–100,000 frames that were labeled by projecting ground truth marker positions (determined by motion capture) into video frames, a common practice in computer vision ([Video S4](#)) ([Ionescu et al., 2014](#)). Networks trained on large numbers of samples (100,000 and 10,000) showed accurate tracking, even on frequently occluded key-points on the forelimbs and hindlimbs ([Figure 2](#)). This performance degraded substantially for those trained on fewer numbers of frames (1,000 and 100), especially when using only three cameras.

However, even when trained on a large number of samples, these convolutional networks did not generalize to tracking out-of-sample rats bearing markers. Networks applied to out-of-sample animals showed 20–30 mm average tracking error, making it impossible to accurately reconstruct the animal's posture on the vast majority of frames ([Figure 2D](#)). Estimating performance with 6 additional cameras (12 total) did not rescue performance on these out-of-sample animals, consistent with past reports suggesting that dozens of cameras and tens to hundreds of thousands of domain- and view-specific hand labels are required for 3D tracking using 2D convolutional networks ([Bala et al., 2020](#); [Iskakov et al., 2019](#)). Lastly, training DeepLabCut using large numbers of labeled frames and then fine-tuning the network again on hand-labeled frames from an out-of-sample rat not bearing markers did not substantially improve tracking ([Figure S4](#)). Thus, while useful for tracking in constrained behavioral tasks, we find that 2D convolutional networks are not currently well suited to the more general problem of 3D tracking across multiple naturalistic behaviors in freely moving animals ([Figure S4E](#); [Table S2](#)).

### Comprehensive Profiling of Behavioral Kinematics

Having established CAPTURE's state-of-the-art ability to precisely track markers over long timescales, we next validated that these kinematic recordings can be used to identify the frequency and transition structure of known and novel rodent behaviors. To do so, we developed a framework for describing discrete behaviors from our kinematic recordings by combining ideas from two previous approaches: (1) supervised behavioral classification, which can be used to detect instances of hand-labeled example behaviors but are unable to identify novel behaviors ([Kabra et al., 2013](#)); and (2) unsupervised behavioral clustering, which can identify a much larger set of potentially novel behaviors but to date has not been shown to robustly detect known rodent behaviors such as grooming ([Berman et al., 2014](#); [Wiltchko et al., 2015](#)). To blend these approaches, we selected two sets of features that resembled those used in supervised and unsupervised classification approaches, respectively: 80 features that were informative about distinctions between a set of commonly recognized rodent



**Figure 2. 2D Convolutional Networks Are Ill-Suited for Whole-Body 3D Tracking across Multiple Behaviors**

(A and B) Example video images and associated wireframe representations of the tracked points using CAPTURE and DeepLabCut. We trained DeepLabCut using between 100 and 100,000 video frames with labels indicating the position of the 20 marker sites tracked using CAPTURE. DeepLabCut predictions were made using six cameras for animals in (A) or out of the training dataset (B).

(C) The average 3D distance between DeepLabCut predictions and the position of points tracked using CAPTURE as a function of training frames, camera number, and whether the rat was part of the training dataset or not (marked as in-sample or out-of-sample, respectively). The motion capture reprojection error is shown for comparison. Errors bars (mean  $\pm$  SEM) are within markers and computed over 306,356 and 249,241 frames for three in-sample and two out-of-sample animals, respectively.

(D) The 10- to 30-mm differences in precision between CAPTURE and DeepLabCut can produce dramatic changes in the ability to accurately reconstruct an animal's pose. We computed the fraction of frames in which the length of 19 different body segments, as reported by DeepLabCut, were within 18 mm of the true segment lengths tracked using CAPTURE, for animals in- and out-of-sample. The fraction of correct segment lengths for CAPTURE is shown for comparison. Shaded error bars (within lines) show standard deviation of 100 bootstrapped samples of frames.

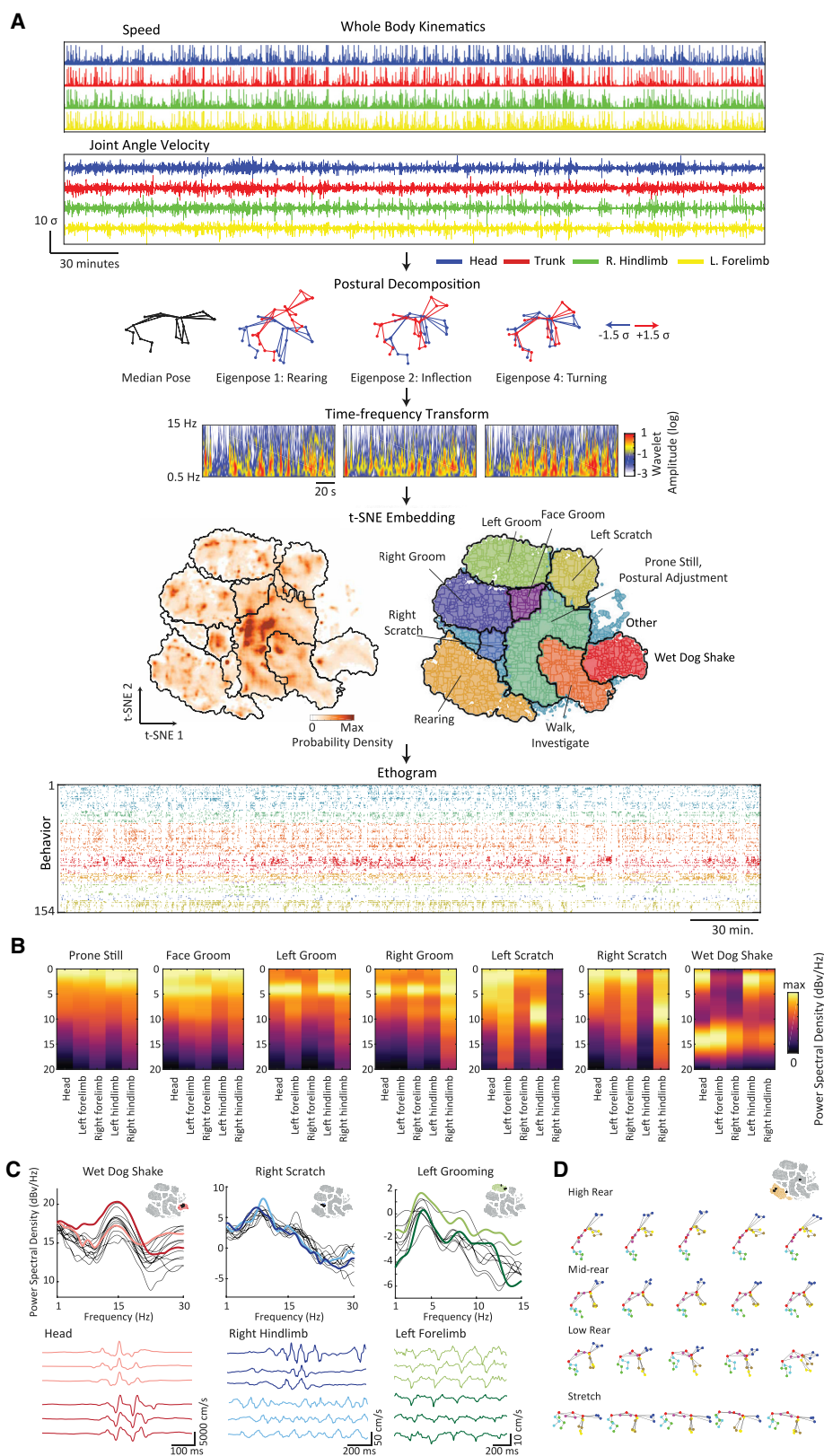
See also [Figure S4](#), [Video S4](#), and [Table S2](#).

behaviors ([Figure S5](#)) and 60 features that broadly encapsulated the animal's pose and kinematics. This yielded a 140 feature matrix that encapsulated the animal's behavior in an  $\sim$ 500-ms time window ([STAR Methods](#)).

To identify repeated instances of behaviors, we used a behavioral mapping approach in which this high-dimensional behavioral feature vector is embedded into two dimensions to create a map that facilitates clustering and exploratory data analysis ([Berman et al., 2014](#)). We collected time points across a representative set of CAPTURE recordings (16 rats,  $1.04 \times 10^9$  frames), subsampled this collection to yield a balanced representation of different behavioral categories, and embedded the subsampled frames into two dimensions using t-stochastic neighbor embedding (t-SNE) ([Figure 3A](#)) ([Maaten and Hinton, 2008](#)). The resulting embedding contained density peaks that corresponded to repeated instances of similar behaviors, which we clustered using a watershed transform ([Figure 3A](#); [Video S5](#)). Inspection revealed that behavioral clusters in different regions of the map corresponded to commonly recognized categories of rodent behaviors such as walking, rearing, grooming of the face and body using the tongue, and scratching of different body sites using the hindlimbs ([Video S6](#)). Individual behavioral

clusters within a category corresponded to postural and kinematic variants ([Video S7](#)).

CAPTURE thus offers the ability to comprehensively profile the kinematics of the rodent behavioral repertoire. As a demonstration of this, we examined the frequency spectrum of different body parts during rhythmic behaviors, specifically grooming, scratching, and wet dog shakes ([Figure 3B](#)). Grooming of the body consistently showed peaks in the frequency of the head and side-specific forelimb speed at 4 Hz and 7–9 Hz, consistent with past reports ([Berridge et al., 1987](#)). In contrast, scratching showed side-specific frequency peaks across a 7–12 Hz range, consistent with the 15–20 Hz frequency reported in mice when adjusted for body size ([Elliott et al., 2000](#)). Wet dog shakes showed a peak in trunk power at  $14 \pm 0.6$  Hz, consistent with past work using high-speed video analysis ([Dickerson et al., 2012](#)). Interestingly, while instances of wet dog shakes and grooming showed similar behavioral frequency but variable amplitude, scratching behaviors varied more broadly in both their frequency and amplitude, suggesting that they may be generated by more flexible control circuits ([Figure 3C](#)). Furthermore, CAPTURE also enables the detection of variability in non-rhythmic behaviors, for instance postural variability in static rearing behaviors ([Figure 3D](#)).



(legend on next page)

### New Insights into the Hierarchical Structure of Behavior

Animal behavior is thought to be hierarchically structured in time into repeated behavioral patterns (Dawkins, 1976; Tinbergen, 1950), yet, to date, systematic and quantitative means of identifying these structures have been lacking. To address this, we first probed the degree to which there is longer-timescale temporal structure in rodent behavior by examining the behavioral transition matrix at different timescales (Figure S6A). We found significantly more structure in the transition matrix at 10- to 100-s timescales than predicted by a first-order Markov chain (Berman et al., 2016), a time-invariant process that is commonly used to model behavioral dynamics (Berridge et al., 1987; Wiltschko et al., 2015). This enhancement in structure was far less pronounced at timescales of several minutes.

To elucidate the nature of these non-Markovian behavioral structures, we developed an algorithm that identifies temporal epochs with similar patterns of behavioral usage on a fixed timescale  $\tau$  (Figure 4A; STAR Methods). As examples of these patterns, we used 15 s and 2-min timescales, which identified distinct behavioral patterns (Figures S6B–S6D). On 15-s timescales, the algorithm identified sequentially ordered patterns in the behavior, such as “canonical” grooming sequences of the face followed by the body (Berridge et al., 1987) or the performance of stereotyped lever-pressing sequences acquired during training in our task (Kawai et al., 2015). We thus termed these detected patterns “sequences” (Figures 4B–4D; Video S8). On 2-min timescales, it identified epochs of varying arousal or task engagement, which often lacked stereotyped sequential ordering. We refer to these as “states” (Video S9). Consistent with this nomenclature, the transition matrices of patterns on 15-s timescales were significantly sparser than those on 2-min timescales, indicating they possessed a more stereotypic ordering between behaviors (Figure S6E).

We used these sequences and states to form a hierarchical representation of behavior (Figure 4D), which we, again, found to be significantly more structured than expected from Markovian behavior (Figures S6F–S6H). Rather than being organized as a strict, tree-based hierarchy (Berman et al., 2016; Dawkins, 1976), behaviors were shared across multiple behavioral sequences, which were then used differentially across behavioral states (Figure 4E). For example, grooming of the right fore-

limb was used both in persistent body grooming sequences and in shorter, more vigorous episodes of grooming, with these sequences being used to different extents in different behavioral states (Figure 4F).

Lastly, we used our pattern detection algorithm to identify usage patterns over a broad range of timescales, ranging from 7.5 to 480 s. We detected numerous distinct patterns between 15 and 120 s. In contrast, we found few patterns beyond 120 s timescales, and patterns below 15 s were similar across timescales (Figures S6B–S6D). Together with our analysis of non-Markovian structure in the transition matrix, these results indicate that behavioral usage is predominantly patterned over 10- to 100-s timescales.

### Multiscale Phenotyping of the Effects of Drugs and Disease

CAPTURE’s measurement precision and ability to record kinematics over long timescales enables an unprecedentedly comprehensive description of how drugs, neural circuit manipulations, and disease states affect behavior. As a first example of this, we performed recordings after acute administration of two stimulants, caffeine and amphetamine, at dosages known to similarly increase locomotor activity (Antoniu et al., 1998). Analysis of CAPTURE recordings recapitulated these findings, showing that both caffeine and amphetamine increase the amount of time animals spend moving compared to recordings at baseline or after injection of a saline vehicle (Figure 5A). However, we found that the changes elicited by these compounds differed substantially in the types of behaviors affected. While both compounds increased the amount of time spent in active behaviors at the expense of idling, caffeine, but not amphetamine, increased the proportion of time rats spent grooming (Figures 5A–5C). Amphetamine affected not only the types of behaviors animal’s expressed but also their patterning in time (Figure 5D). While the states and sequences expressed by animals after administration of caffeine were similar to naturally occurring states of arousal, amphetamine elicited entirely novel behavioral states that consisted of repetitive locomotor sequences such as circling (Figures 5D and 5E; Video S10). In the past, amphetamine was thought to elicit a biphasic behavioral response, stimulating locomotor activity at low doses and repetitive motor

### Figure 3. Comprehensive Kinematic Profiling of the Rat Behavioral Repertoire

(A) We developed a behavioral mapping procedure to identify behaviors in CAPTURE recordings. We first defined a set of 140 per-frame features describing the pose and kinematics of rats within a 500-ms local window. Many of these features were obtained by computing the eigenpostures of the rats from the measured kinematics (top), which consisted of commonly observed postural changes such as rearing or turning to the left and right. We then computed time-frequency decompositions of the eigenposture scores over time using a wavelet transform (middle). We subsampled features from all 16 rats recorded (1.047 billion frames) and co-embedded them in two dimensions using t-SNE to create a behavioral map that we clustered using a watershed transform (Figure S5) (STAR Methods). We annotated clusters in the behavioral map, which showed that behavioral categories segregated to different regions of the map. From these behavioral clusters, we then computed ethograms describing behavioral usage over time (bottom).

(B) The power spectrum of the speed of markers on different body segments during rhythmic behaviors. Because different body segments show large variance in overall power, power spectral densities are separately scaled within each marker group on the head, forelimbs, and hindlimbs. Power spectra computed from  $n = 4$  animals on 2 days.

(C) Top: power spectral densities of individual behavioral clusters belonging to rhythmic behaviors for one rat over 2 days. Power spectral densities were computed over the speed of one marker on the body segment listed. Colored lines correspond to examples below. Insets show position of the example behavioral clusters within the coarse behavioral category. Spectral densities are reported relative to  $1 \text{ (mm} \times \text{s}^{-1})^2 / \text{Hz}$ . Bottom: individual instances of marker kinematics randomly drawn from example clusters.

(D) Example poses selected from non-rhythmic behavior clusters associated with rearing and stretching. See also Figure S5 and Videos S5, S6, and S7.





**Figure 4. Rat Behavior Is Hierarchically Organized into Behavioral Sequences and States**

(A) We developed a temporal pattern matching algorithm to detect repeated behavioral patterns in our behavioral recordings. We smooth ethograms over 15-s or 2-min timescales and compute the pairwise correlation to yield a similarity matrix, which we threshold to extract high-value off-diagonal elements. These correspond to patterns of behavioral usage observed at two or more time points in the dataset. We cluster these patterns to identify repeated sequences and states, which can be organized to identify hierarchical structure. Ethograms and similarity matrices are shown from a full day of recording in a one animal smoothed with 4-s and 15-s boxcar filters, respectively. Dendrograms and clustered states are schematic examples. Behaviors colored following Figure 3A.

(B) Ethograms, smoothed with a 4-s filter for visualization, shown on minute- (top) and hour-long (bottom) timescales. Ethograms are shown for the subset of behaviors observed during each time period, sorted and shaded by their membership in different behavioral sequences (top) and states (bottom). States and sequences detected using the pattern matching algorithm are shown above and ordered in (B) and (C) by their membership in different behavioral categories.

(C) Heatmaps showing the composition of sequences (top) and states (bottom) in terms of different behavioral categories.

(D) Examples of behavioral patterning during different sequences and states. Sequences show repeated temporal patterns of behavioral usage, while states show average increases in behavioral frequency but without detailed temporal structure. For clarity, only a subset of frequently occurring behaviors are shown. Behaviors are colored by their membership in behavioral categories as in Figure 3A.

(E) To display the hierarchical organization of these sequences and states, we computed a stacked tree, whose lower links reflect the probability of observing different behaviors during each sequence and whose upper links reflect the probability of observing different sequences in each state. Behaviors and links are colored according to their membership in one of three categories: Grooming (grooming and scratching), Idling (prone still and postural adjustment), and Active

(legend continued on next page)



“stereotypies,” such as head swinging, at higher doses (Antoniou et al., 1998). Our results suggest that the increase in repetitive behavior triggered by amphetamine begins at much lower doses of administration and is at least partially dissociable from its effects on locomotion.

Having validated CAPTURE’s capacity to identify both known and novel effects of drugs on behavior, we next screened for behavioral changes in a rat model of fragile X syndrome, the most common monogenic form of autism spectrum disorders in humans (Miller et al., 2019). While behavioral assays using targeted behavioral tests have been performed in a variety of models of autism, including fragile X syndrome (Silverman et al., 2010), reports of behavioral changes often conflict across studies, especially those related to grooming behaviors (Brunner et al., 2015). We used CAPTURE to compare the behavioral repertoire of male *Fmr1*-KO rats and their wild-type cage mates over three continuous days of home cage behavior. While knockout rats spent comparable time moving, they spent significantly more time grooming (Figures 6A and 6B). This did not simply reflect an increase in the amount of time spent in normal grooming behaviors; *Fmr1*-KO also engaged in different types of grooming behaviors that tended to occur in longer bouts (Figure 6C). The grooming behaviors in knockouts were also assembled into different sequential patterns than controls, exhibiting altered usage of behavioral sequences and states (Figure 6D). This effect was driven in part by abnormal and idiosyncratic grooming sequences in knockout animals (Figure 6E). Taken together, this suggests that while *Fmr1*-KO rats largely overlap with wild-type animals in their locomotor behavior, they show idiosyncratic perseverative grooming sequences, a robust behavioral manifestation well suited for exploring the neural mechanisms underlying motor stereotypies present in autism spectrum disorders (Kalueff et al., 2016).

### Stability, Individuality, and Commonality of the Rodent Behavioral Repertoire

Rodents are known to exhibit individual differences in behavior as a result of interactions between genetic, environmental, developmental, and social factors (Lathe, 2004). Individuality is typically measured through coarse estimates of behavioral usage (Forkosh et al., 2019; Freund et al., 2013), leaving open the question of whether animals exhibit individuality along other dimensions, for instance in the kinematics and long-timescale patterning of behavior that can now be systematically explored using CAPTURE.

We therefore used CAPTURE to compare the behavior of five female LE rats across 3–5 days of baseline recordings. CAPTURE identified subtle kinematic differences in the vigor and fre-

quency of rhythmic behaviors across days and animals, but these kinematic differences were significantly smaller than those present across different behaviors, indicating that animals drew from a common behavioral repertoire (Figures 7A and 7B). However, the usage of these behaviors varied significantly across animals (Figures 7C–7E). This variation showed only moderate dependence on the coarseness of behavioral definitions, indicating that individuality in behavioral usage is not a trivial consequence of the slight differences in behavioral kinematics across animals (Figure 7F). Interestingly, dissimilarity across animals was most pronounced for grooming and rearing behaviors, suggesting they may be particularly sensitive to developmental or environmental differences (Figure 7E).

Beyond variation in overall behavioral usage, animals demonstrated individuality in the long-timescale patterning of behavior. Animals again drew from a common set of temporal patterns (Figure S7). Pattern usage in single animals across days showed ~80% similarity, but varied more significantly across animals, especially at long timescales (Figures 7G and 7H). Classifiers trained to recognize animal identities based on behavioral usage were significantly improved by adding additional information about long-timescale patterns (Figure 7I). This suggests that the patterning of behavior reflects a locus of individuality that is distinct from idiosyncrasies in behavioral usage.

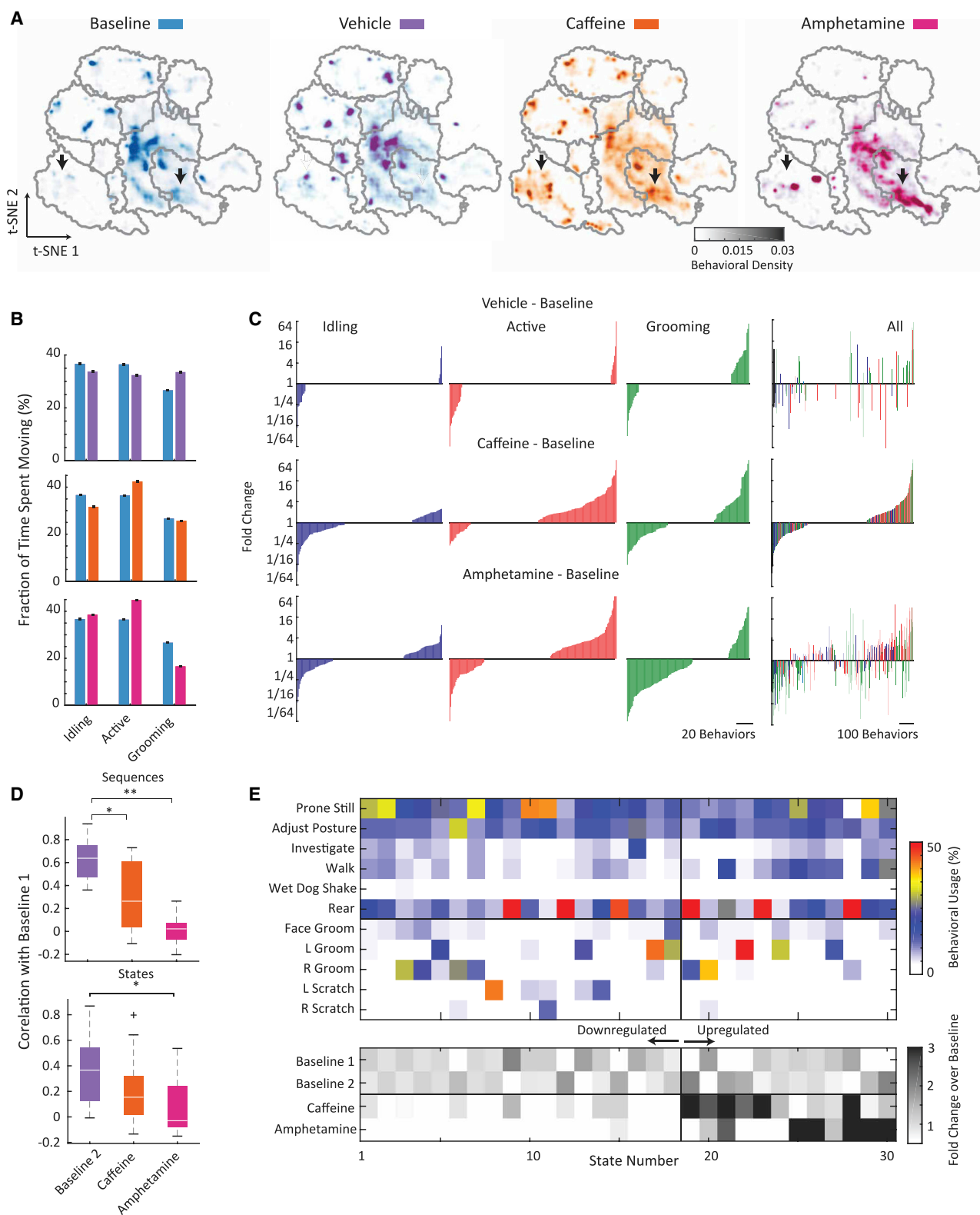
### DISCUSSION

We combined motion capture, deep learning, and body piercings to enable continuous kinematic tracking of the head, trunk, and appendages in freely behaving rats continuously over week-long timescales. In comparison to past approaches using depth cameras or convolutional neural networks, CAPTURE offers the ability to track limbs in 3D across the rat’s full behavioral repertoire (Figures 1 and 2). To parse these continuous 3D kinematic recordings, we developed an unsupervised computational analysis framework, which provides a synoptic picture of behavior encompassing discrete organismal behaviors and long-timescale behavioral patterns such as sequences and states (Figures 3 and 4). This analysis allowed us to both recapitulate the kinematic and sequential characteristics of known rodent behaviors and identify novel kinematic, sequential, and organizational properties of behavior, including a coupling between the rodent’s behavioral state and type of grooming observed (Figures 4F). Finally, we demonstrated how CAPTURE can be used to identify novel behavioral, sequential, and kinematic changes in response to drugs and disease states (Figures 5 and 6) and across individuals (Figure 7), in particular demonstrating the presence of perseverative grooming stereotypies in a rat model of autism spectrum disorder.

(investigate, walking, shaking, and rearing). For clarity, one in eight behaviors are shown, and we only visualize links showing probabilities of occurrence greater than 0.05 and 0.1 for behaviors and sequences, respectively. Shaded region corresponds to region highlighted in (F).

(F) Hierarchical arrangement of the right-side grooming sequences highlighted in (E) (left), as well as example ethograms of a subset of behaviors in each sequence (right). For clarity, in the hierarchy, we show only one in four behaviors and a subset of behavioral states used. Across animals, there was significant mutual information between behavioral state and the grooming sequence observed (0.62 versus  $10^{-3}$  rats,  $p = 0.004$ , sign test, over 9 days where grooming sequences were observed).

Example performance data in (B)–(F) shown for a single animal on one full day of recording. Example states and sequences numbered in parenthesis reference behavioral usages shown in (C). See also Figures S6 and S7 and Videos S8 and S9.



**Figure 5. Caffeine and Amphetamine Show Similar Effects on Arousal but Divergent Effects on Behavioral Organization**

(A) Behavioral density maps during baseline and after acute administration of a saline vehicle, caffeine (10 mg/kg), or amphetamine (0.5 mg/kg). The fraction of time animals spent moving compared to baseline ( $32\% \pm 1\%$ ) increased significantly after administration of caffeine ( $80\% \pm 1\%$ ) and amphetamine ( $84\% \pm 1\%$ ),

(legend continued on next page)

### Comparisons to the Emerging Behavioral Tracking Toolkit

A behavioral revolution is underway in neuroscience. Much like the diverse options available for monitoring electrical, calcium, hemodynamic, and neuromodulatory activity in the brain, the expanding behavioral toolkit contains a range of methods for behavioral measurements. These vary in the number of unique behaviors that can be discriminated, and in the ability to report the location of body key-points in 2D or 3D (Table S2). By combining large numbers of high-resolution camera views with high signal-to-noise markers, CAPTURE offers a robust and reliable framework for 3D kinematic measurements, continuously across the rat behavioral repertoire. In comparison, alternative approaches using depth cameras (Hong et al., 2015; Wiltchko et al., 2015) may offer reduced cost and increased throughput, but they struggle with infrared reflections (e.g., from bedding and other objects) and the ability to disassociate animals from objects in the arena, and they cannot track defined key-points, including the position of appendages. These limitations constrain the range of environments in which depth camera approaches can be applied, as well as their ability to test hypotheses relating neural activity with behavioral kinematics. Alternatively, we found that 2D convolutional networks, while useful for tracking non-occluded key-points in behavioral tasks, need extensive training datasets (100,000 fully labeled frames) to perform whole-body 3D tracking and, even then, fail to learn general visual features that readily transfer to new animals (Figure 2). Thus, these approaches are not suited to whole-body 3D tracking across an animal's behavioral repertoire.

CAPTURE, however, also faces challenges, as it is difficult to extend to species and regions of the body that cannot be equipped with markers. While we found that body piercings did not overtly modify animals' behavior (Figure S2), they may cause subtler behavioral changes or influence animals' social or environmental interactions. These potential adverse effects, as well as potential shifts in marker positions during longer experiments, should be carefully considered. Extending CAPTURE to a broader range of species and key-point locations could be helped by using it alongside synchronized video recordings to train deep learning algorithms for markerless 3D motion capture as is done in humans (Ionescu et al., 2014; Pavlakos et al., 2017). This could permit high-throughput markerless motion capture with as few as one camera.

### Behavioral Analysis across Spatial and Temporal Scales

Understanding the logic of biological systems can rarely be achieved by studies at a single scale. The "omics" era has demonstrated how complementary insights into the development and function of organisms can be derived across the single-cell, tissue, and organ-system scales, while neuroscience is increasingly driven forward by an integration of insights from molecular, systems, and computational approaches (Luo et al., 2018). Previous studies of behavior in rodents were restricted to high-resolution snapshots of individual behaviors or across a broader range of behaviors at a more qualitative level, stymieing efforts to integrate insights across scales (Berman, 2018). CAPTURE can precisely track kinematics over long timescales, allowing us to record both the occurrence of behaviors, their precise 3D kinematics, and their organization into sequences and states. This multiscale capability allowed us to quantitatively interrogate both the stereotypy and flexibility of rat behavior across timescales and individuals. We showed that behaviors are associated with different patterns of variability in frequency and amplitude (Figure 3C), that the same behaviors can be organized into different sequential patterns as a function of the animal's behavioral state (Figures 4E and 4F), and that different animals show individuality in behavioral kinematics, usage, and patterning (Figure 7).

Our results represent only a few select examples of the types of inquiry into behavioral organization possible with CAPTURE. Similar to recent comprehensive surveys of brain connectivity or cellular transcriptional properties, the high-resolution and comprehensive nature of these datasets can be used by the broader community to achieve a diverse set of goals. For example, they could serve to test new algorithms for behavioral analyses, as reference datasets for standardizing behavioral definitions, or to characterize the biomechanical properties of movement across the rodent behavioral repertoire.

### Comprehensive Phenotyping of Behavioral Perturbations

In comparison to existing approaches for motor and behavioral phenotyping that use task-based batteries of tests to probe specific kinematic or behavioral deficits (Jinnah and Hess, 2015; Silverman et al., 2010) or depth cameras to coarsely assess changes in usage across a broader range of natural behaviors (Wiltchko et al., 2015), CAPTURE can comprehensively identify

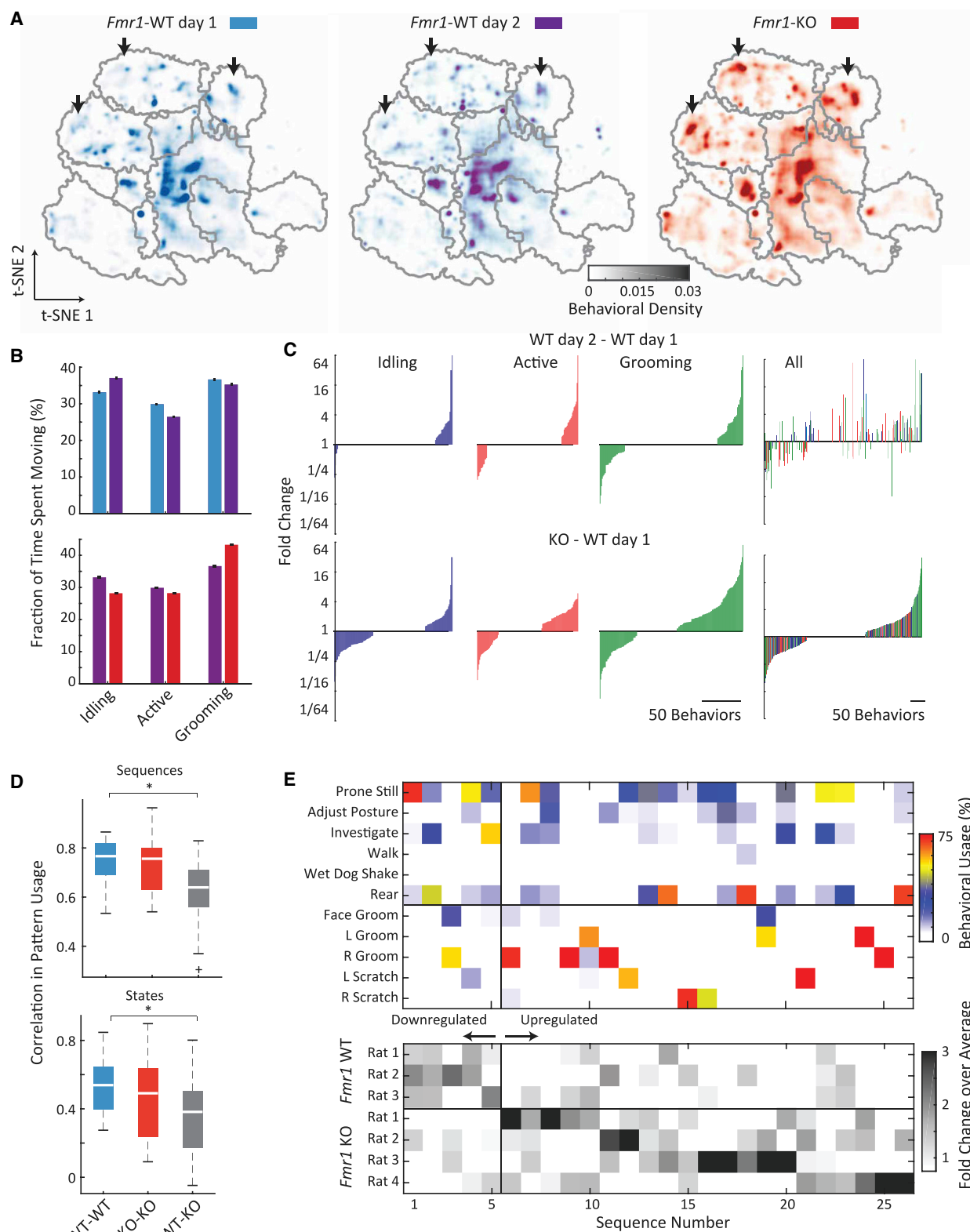
but not after a saline vehicle ( $17\% \pm 1\%$ ;  $p = 0.016$  for caffeine and amphetamine,  $p = 0.4$  for vehicle,  $n = 4$  rats for drug and vehicle conditions and  $n = 5$  rats for baseline condition, 4,635–7,391 s per condition). Both stimulant compounds altered the amount of time spent engaging in walking and rearing (arrows).

(B) Caffeine and amphetamine increased the fraction of time spent in active locomotor behaviors at the expense of idling behaviors but had divergent effects on grooming, with amphetamine alone showing suppression of grooming activity (all  $p < 10^{-5}$ , binomial test,  $n = 238$ –989 effective samples per behavioral category).

(C) Caffeine and amphetamine introduced new types of active and grooming behaviors that were rarely observed at baseline, such as high-velocity walks and more vigorous grooming. Colored bars denote fold change significantly modulated behaviors ( $p < 10^{-6}$ , Poisson probability of rates in the perturbed condition, Benjamini-Hochberg corrected). Sorting all behavioral changes by their modulation after administration of caffeine (right), reveals that caffeine and amphetamine both increase the frequency of many active behaviors and decrease the frequency of many grooming behaviors.

(D) Box-and-whisker plots showing the correlation coefficient of sequence (top) and state (bottom) probability vectors across baseline and drug conditions. Amphetamine and caffeine induced significant changes in the long-timescale organization of behavior ( $**p = 1.5 \times 10^{-6}$ ,  $*p < 0.005$ ; rank-sum test;  $n = 16$  pairs of days between conditions). Because of the relative lack of movement during vehicle and time-matched controls, baseline data here are taken from 2 days of recordings.

(E) Heatmaps showing the composition of states in terms of behavioral categories (top) and the fold change in the usage of these states compared to baseline (bottom), sorted to emphasize changes across conditions. Amphetamine introduced highly distinct states in animals, emphasized by black lines (Video S10). See also Video S10.



(legend on next page)

precise changes in both behavioral kinematics and usage. Additionally, the continuity of CAPTURE recordings, and compatibility with naturalistic arenas such as the animal's home cage, enables a more complete characterization of changes in long-timescale behavioral organization.

As an example of this capability, we demonstrated how caffeine and amphetamine elicit different changes in behavioral states and kinematics, with amphetamine producing a distinct, stereotypic, behavioral state with highly altered behaviors and behavioral kinematics (Figure 5). We also described idiosyncratic perseverative grooming sequences in a rat model of fragile X syndrome (Figure 6). The kinematic and behavioral resolution of CAPTURE should be especially useful for motor diseases such as Parkinson's, Huntington's, Tourette's, and dystonia, where key motor symptoms such as tremor, dyskinesia, stereotypies, and postural changes lack robust quantitative metrics in animal models (Gittis and Kreitzer, 2012; Pappas et al., 2014; Parker et al., 2018). The ability of CAPTURE to record natural behaviors over long timescales should be especially useful in models of behavioral disorders such as autism, anxiety, bipolar, and depression, in which symptoms are often associated with slower changes in behavioral patterns that have been difficult to reliably measure in animal models (Nestler and Hyman, 2010).

### Next-Generation Frameworks for Behavioral Analysis

High-dimensional data analysis, especially at the billion-time-point scale of CAPTURE recordings, is an emerging discipline. In the future, new frameworks for identifying behaviors and long-timescale structure will be aided by new approaches for dimensionality reduction (DeAngelis et al., 2019), clustering (Todd et al., 2017), feature set design (Du et al., 2015), and statistical modeling (Calhoun et al., 2019). However, such unsupervised analyses will only be an intermediate step toward the creation of standardized and rigorously defined behavioral taxonomies, which CAPTURE recordings should greatly facilitate (Anderson and Adolphs, 2014).

The ability of CAPTURE to readout postural kinematics also opens the possibility of creating analysis frameworks that go beyond behavioral identification to elucidate important axes of variation from the perspective of motor control (DeAngelis et al., 2019), behavioral exploration (Wu et al., 2014), and neural activity. Much like how task-optimized neural networks have

been valuable tools for elucidating the relevant axes of variation of poorly characterized sensory representations (Bao et al., 2020), the internal structure of deep neural networks trained to imitate CAPTURE recordings should be valuable tools for disentangling motor representations (Merel et al., 2018, 2020).

### Outlook

CAPTURE, in its current form, is a powerful tool for recording whole-body kinematics, but future improvements could provide an even more integrated picture of whole-organism behavior and physiology across species. Extension of CAPTURE's kinematic coverage can be made using increased numbers of smaller markers, active motion capture markers, or improved algorithms for imputation and body model fitting. To facilitate a broader understanding of behavior and physiological state, CAPTURE recordings could be paired with complementary measurements of musculoskeletal or physiological state, such as of skeletal or muscular dynamics (Nakamura et al., 2005) or eye, pupil, and whisker tracking (Meyer et al., 2018). CAPTURE can easily be extended to other mammalian model systems, such as marmosets or mice, or social settings by using distinguishable marker sets, although potential effects of body piercings should be evaluated in each new use case. Motion capture systems identify key-points in real time and so can be used with real-time behavioral recognition approaches to deliver closed-loop feedback, enabling precise and rapid reinforcement of natural behaviors and specific movement kinematics. Lastly, the comprehensive behavioral recordings enabled by CAPTURE, when combined with neural recordings, should be a powerful tool for disentangling the neural coding schemes for movement and influence of behavior on cognitive and sensory coding.

Overall, by providing a quantitative and comprehensive portrait of animal behavior, CAPTURE sets the stage for a broad range of new inquiries into animal behavior and its control by the nervous system.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

#### ● KEY RESOURCES TABLE

### Figure 6. *Fmr1*-KO Rats Show Idiosyncratic Perseverative Body Grooming Sequences

(A) Behavioral density maps of *Fmr1*-WT ( $n = 3$ ) and *Fmr1*-KO ( $n = 4$ ) age-matched cage mates for two full days of recording. Arrows highlight periods of increased grooming in *Fmr1*-KO rats. Both wild-type and knockout rats spent equal amounts of time moving ( $35.1\% \pm 2\%$  versus  $35.0\% \pm 0.7\%$ , rank-sum test,  $p = 0.85$ ;  $n = 6$  and 8 days of recording in wild-type (WT) and knockout (KO) rats respectively, 445,601–673,473 s per condition).

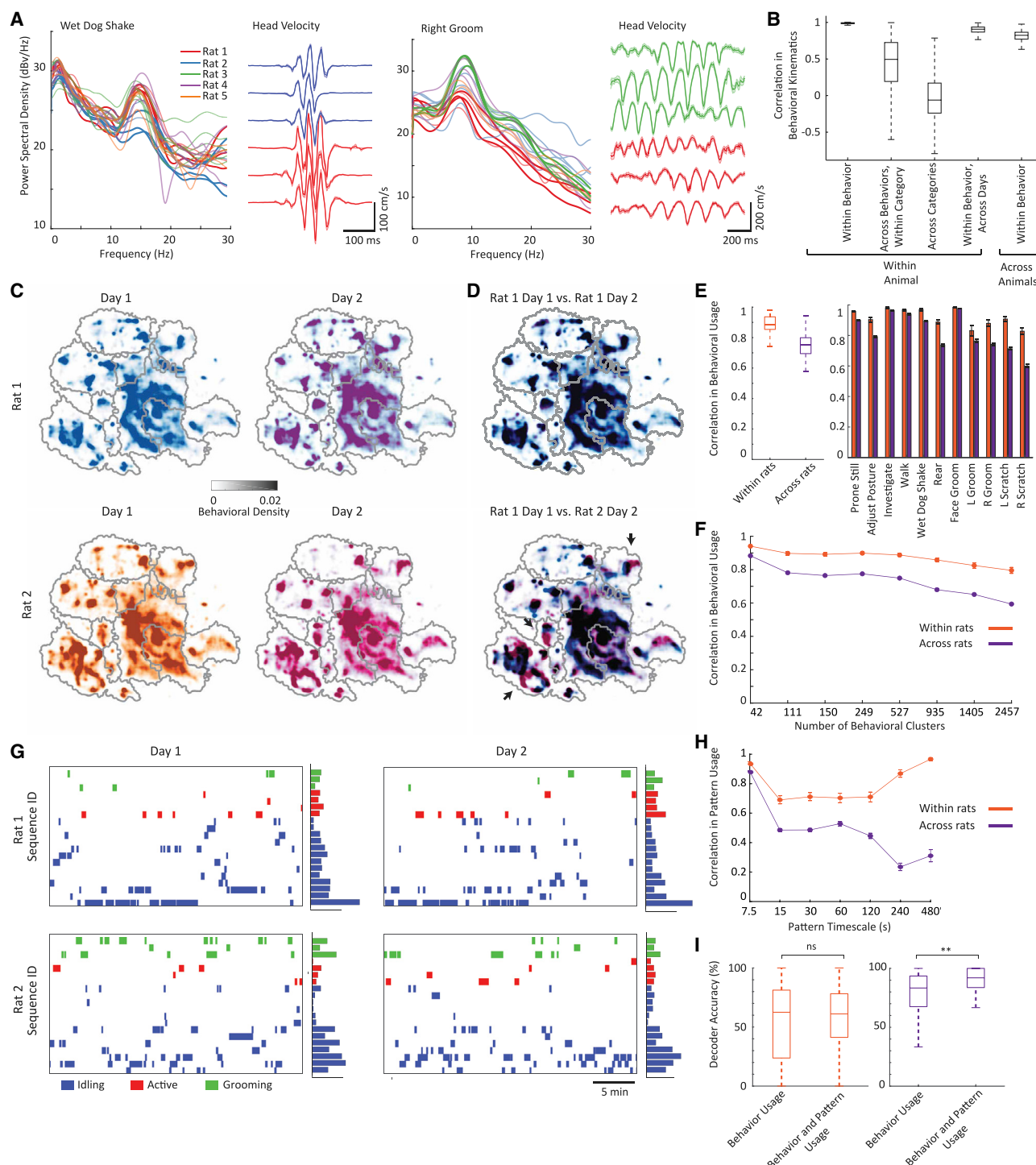
(B) *Fmr1*-KO, but not *Fmr1*-WT, animals spent increased time grooming, at the expense of idling behaviors (all  $p < 10^{-5}$ , binomial test,  $n_{\text{eff}} = 166$ –232 effective samples per behavioral category). This was accompanied by a significant increase in the dwell time of grooming behaviors alone ( $10\% \pm 2\%$  increase in dwell time between knockout and wild-type animals over all grooming behaviors,  $p = 0.03$  signed-rank test,  $n = 208$  grooming behaviors, compared to  $< 0.01\%$  changes in the average dwell time of active and idling behaviors).

(C) The composition of behavioral categories, especially grooming behaviors, were substantially modified in *Fmr1*-KO, but not *Fmr1*-WT, animals. Colored bars denote fold change of behaviors significantly modulated across the stated conditions ( $p < 10^{-6}$ , Poisson probability of altered rates, Benjamini-Hochberg corrected). Sorting all behavioral changes by their modulation in *Fmr1*-KO animals (right) shows little commonality between behaviors modulated in wild-type and knockout animals.

(D) Box-and-whisker plots showing the correlation of sequence and state usage probabilities for wild-type and *Fmr1*-KO rats. There was a significantly decreased correlation across genotypes ( $*p = 0.0006$ ,  $p = 0.01$  for sequences and states, respectively; rank-sum test,  $n = 48$  pairs of days between conditions).

(E) Heatmaps showing the composition of sequences in terms of behavioral categories (top) and fold change in the usage of these sequences compared to baseline (bottom) sorted to emphasize changes across conditions. Black lines highlight elevated levels of idiosyncratic grooming sequences in knockout rats.





**Figure 7. Long-Timescale Behavioral Structure Is a Locus of Behavioral Individuality**

(A) Left: power spectral densities of the velocity of the front head marker for two individual behavioral clusters for  $n = 5$  rats over 3–5 days each. Spectral densities are reported relative to  $1 \text{ (mm} \times \text{s}^{-1})^2 / \text{Hz}$ . Right: average Cartesian velocity of the head in each behavioral cluster shown for two rats on 3 days. Shaded bars represent SEM.

(B) Boxplots showing the Pearson correlation of the 140-dimensional behavioral feature vector for five separate comparisons: within-animal, within-day comparisons of (1) different instances of the same behavior, different behaviors in the same coarse category (2), or different behaviors in different coarse categories (3)

(legend continued on next page)

## ● RESOURCE AVAILABILITY

- Lead Contact
- Materials Availability
- Data and Code Availability

## ● EXPERIMENTAL MODEL AND SUBJECT DETAILS

### ● METHOD DETAILS

- Motion Capture Arena
- Retroreflective Body Piercings
- Drug Administration
- Surgery
- Motion Capture Recordings
- Motion Capture Postprocessing
- Performance Comparisons with 2D Convolutional Neural Networks
- Behavioral Feature Generation and Selection
- Behavioral Embedding, Clustering, and Systematics
- Temporal Pattern Matching Algorithm
- Behavioral Comparisons before and after Body Piercings

### ● QUANTIFICATION AND STATISTICAL ANALYSIS

- Motion Capture Precision
- Behavioral Power Spectra
- Transition Matrix Eigenvalue Analysis
- Similarity of Behavioral Patterns Across Timescales
- Cross-condition Behavioral Comparisons
- Individuality Analyses
- Data Analysis and Statistics

laboratory of Aaron Guerts and Simons Foundation Autism Research Initiative (SFARI) for generating the *Fmr1*-KO rat model. J.D.M. acknowledges support from a Helen Hay Whitney Foundation postdoctoral fellowship and NINDS K99/R00 Pathway to Independence Award K99NS112597 from the NIH. D.E.A. acknowledges support from National Science Foundation graduate fellowship DGE1745303. W.L.W. acknowledges support from a Harvard College Research Program fellowship. G.J.B. acknowledges grant support from the Research Corporation (25999), HFSP (RGY0076), and NIMH (R01MH115831-01). B.P.Ö. acknowledges support from SFARI (646706), the NIH (R01NS099323), and the Star Family Challenge Fund. Parts of this work were performed at the Aspen Center for Physics, which is supported by NSF grant PHY-1607611.

## AUTHOR CONTRIBUTIONS

The project was conceived by B.P.Ö. and J.D.M. J.D.M. developed the technique with consultation from B.P.Ö. J.D.M. performed all experiments with assistance from W.L.W. D.E.A. wrote the imputation package. J.D.M. performed all data analyses, with the aid of additional analysis design, software, and support from G.J.B. and T.W.D. J.D.M. wrote the manuscript in close consultation with the co-authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 27, 2020

Revised: October 1, 2020

Accepted: November 16, 2020

Published: December 18, 2020

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.neuron.2020.11.016>.

## ACKNOWLEDGMENTS

We thank all members of the Ölveczky lab for feedback and advice and acknowledge special experimental and computational assistance from Ashesh Dhawale, Steffen Wolff, Sasha Illeue, Alexander Barrett, and Mahmood Shah, as well as Harvard Center for Brain Science neuroengineers Ed Soucy, Joel Greenwood, Adam Bercou, and Brett Graham. We additionally thank the

## SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Agarwal and Triggs (2004); Alahi et al. (2016); Bai et al. (2018); Blake and Shiffrar (2007); Box and Jenkins (1976); Cappello et al. (1997); Chen et al. (2016); Collins et al. (2009); Deutscher et al. (2000); Dhawale et al. (2017); Gal and Ghahramani (2016); Gall et al. (2010); Graves et al. (2013); Heskes (1997); Kitagawa and Windsor (2008); Kobak and Berens (2019); Kobak and Linderman (2019); Lakshminarayanan et al. (2017); Lamb et al. (2016); Nix and Weigend (1994); Ohayon et al. (2013); Oord et al. (2016a), (2016b); Pearce et al. (2018); Seidler et al. (1992); Stim (2003); Taylor et al. (2007); Todd et al. (2017); Venkatraman et al. (2010) Yu and Koltun (2016).

and behavioral averages across different days within (4) and across (5) animals. All  $p < 10^{-10}$ , Kruskal-Wallis test with Bonferroni corrected post hoc testing,  $n = 1,731, 13,386, 135,094, 5,264$ , and  $22,934$  pairs of feature vectors for each comparison, respectively. Only behaviors with at least five instances per day were analyzed.

(C) Representative behavioral density maps showing consistency of behavior usage from day to day for two rats.

(D) Overlaid t-SNE density maps across days (top) and rats (bottom).

(E) Pearson correlation between the probability vectors of behavior usage over days, across all categories (left) and per behavioral category (right), over different days and rats. There was a significant decrease in correlation across rats, especially for rearing and grooming behaviors (signed-rank test,  $p < 10^{-7}$  for all comparisons;  $n = 36$  and  $348$  pairs of days within and across animals, respectively). There was a significant correlation between the behavioral usage change over days and the number of calendar days between recordings ( $p = 0.002$ , F-test compared to null model,  $R^2 = 0.23$ ,  $n = 36$ ), indicating that there was some drift in behavioral usage over time.

(F) The Pearson correlation between the probability vectors of behavior usage over days across 7 increasingly fine-grained clusterings of the behavioral space (STAR Methods).

(G) Example ethograms of sequence usage for two rats on two different days, as well as summary bar graphs showing the average sequence usage over the entire day. For clarity, we show only a subset of sequences, smoothed with a 10-s filter. Scale bar on bar plots corresponds to 10% of frames.

(H) Average Pearson correlation in the usage of temporal patterns across timescales. We found significantly greater similarity in usage across animals at short timescales compared to long timescales (signed-rank test,  $p < 10^{-10}$ , 174 pairs of days).

(I) We trained a random forest classifier to distinguish individual rats or individual days using behavioral usage statistics. Incorporating long-timescale pattern usage significantly improved identification of species, but not across days within species (signed-rank test,  $p = 0.01, 0.6$ , respectively,  $n = 21$  days).

See also Figure S7.

## REFERENCES

- Agarwal, A., and Triggs, B. (2004). Tracking articulated motion using a mixture of autoregressive models. In *Computer Vision: ECCV 2004*, T. Pajdla and J. Matas, eds. (Springer Berlin Heidelberg), pp. 54–65.
- Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., and Savarese, S. (2016). Social lstm: human trajectory prediction in crowded spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE)*, pp. 961–971.
- Anderson, D.J., and Adolphs, R. (2014). A framework for studying emotions across species. *Cell* 157, 187–200.
- Anderson, D.J., and Perona, P. (2014). Toward a science of computational ethology. *Neuron* 84, 18–31.
- Angel, E. (2009). *The Piercing Bible: the Definitive Guide to Safe Body Piercing (Celestial Arts)*.
- Antoniou, K., Kafetzopoulos, E., Papadopoulou-Daifoti, Z., Hyphantis, T., and Marselos, M. (1998). D-amphetamine, cocaine and caffeine: a comparative study of acute effects on locomotor activity and behavioural patterns in rats. *Neurosci. Biobehav. Rev.* 23, 189–196.
- Bai, S., Kolter, J.Z., and Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv*, 1803.01271 <https://arxiv.org/abs/1803.01271>.
- Bala, P.C., Eisenreich, B.R., Yoo, S.B.M., Hayden, B.Y., Park, H.S., and Zimmermann, J. (2020). Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nat. Commun.* 11, 4560.
- Bao, P., She, L., McGill, M., and Tsao, D.Y. (2020). A map of object space in primate inferotemporal cortex. *Nature* 583, 103–108.
- Berman, G.J. (2018). Measuring behavior across scales. *BMC Biol.* 16, 23.
- Berman, G.J., Choi, D.M., Bialek, W., and Shaevitz, J.W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. *J. R. Soc. Interface* 11, 20140672.
- Berman, G.J., Bialek, W., and Shaevitz, J.W. (2016). Predictability and hierarchy in *Drosophila* behavior. *Proc. Natl. Acad. Sci. USA* 113, 11943–11948.
- Berridge, K.C., and Fentress, J.C. (1987). Disruption of natural grooming chains after striatopallidal lesions. *Psychobiology (Austin Tex.)* 15, 336–342.
- Berridge, K.C., Fentress, J.C., and Parr, H. (1987). Natural syntax rules control action sequence of rats. *Behav. Brain Res.* 23, 59–68.
- Blake, R., and Shiffrar, M. (2007). Perception of human motion. *Annu. Rev. Psychol.* 58, 47–73.
- Box, G.E.P., and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*, Revised Edition (Holden-Day).
- Branson, K., Robie, A.A., Bender, J., Perona, P., and Dickinson, M.H. (2009). High-throughput ethomics in large groups of *Drosophila*. *Nat. Methods* 6, 451–457.
- Brown, A.E.X., Yemini, E.I., Grundy, L.J., Jucikas, T., and Schafer, W.R. (2013). A dictionary of behavioral motifs reveals clusters of genes affecting *Caenorhabditis elegans* locomotion. *Proc. Natl. Acad. Sci. USA* 110, 791–796.
- Brunner, D., Kabitzke, P., He, D., Cox, K., Thiede, L., Hanania, T., Sabath, E., Alexandrov, V., Saxe, M., Peles, E., et al. (2015). Comprehensive analysis of the 16p11.2 deletion and null Cntnap2 mouse models of autism spectrum disorder. *PLoS ONE* 10, e0134572.
- Calhoun, A.J., Pillow, J.W., and Murthy, M. (2019). Unsupervised identification of the internal states that shape natural behavior. *Nat. Neurosci.* 22, 2040–2049.
- Cappello, A., Cappozzo, A., La Palombara, P.F., Lucchetti, L., and Leardini, A. (1997). Multiple anatomical landmark calibration for optimal bone pose estimation. *Hum. Mov. Sci.* 16, 259–274.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A.L. (2016). Semantic image segmentation with deep convolutional nets and fully connected CRFs. *arXiv*, 1412.7062 <https://arxiv.org/abs/1412.7062>.
- Collins, T.D., Ghoussayni, S.N., Ewins, D.J., and Kent, J.A. (2009). A six degrees-of-freedom marker set for gait analysis: repeatability and comparison with a modified Helen Hayes set. *Gait Posture* 30, 173–180.
- Courtine, G., Song, B., Roy, R.R., Zhong, H., Herrmann, J.E., Ao, Y., Qi, J., Edgerton, V.R., and Sofroniew, M.V. (2008). Recovery of supraspinal control of stepping via indirect propriospinal relay connections after spinal cord injury. *Nat. Med.* 14, 69–74.
- Dawkins, R. (1976). Hierarchical organisation: a candidate principle for ethology. In *Growing Points in Ethology* (Cambridge University Press), pp. 7–54.
- DeAngelis, B.D., Zavatone-Veth, J.A., and Clark, D.A. (2019). The manifold structure of limb coordination in walking *Drosophila*. *eLife* 8, e46409.
- Deutscher, J., Blake, A., and Reid, I. (2000). Articulated body motion capture by annealed particle filtering. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (IEEE)*, pp. 126–133.
- Dhawale, A.K., Poddar, R., Wolff, S.B., Normand, V.A., Kopelowitz, E., and Ölveczky, B.P. (2017). Automated long-term recording and analysis of neural activity in behaving animals. *eLife* 6, e27702.
- Dhawale, A.K., Wolff, S.B.E., Ko, R., and Ölveczky, B.P. (2019). The basal ganglia can control learned motor sequences independently of motor cortex. *bioRxiv*. <https://doi.org/10.1101/827261>.
- Dickerson, A.K., Mills, Z.G., and Hu, D.L. (2012). Wet mammals shake at tuned frequencies to dry. *J. R. Soc. Interface* 9, 3208–3218.
- Du, Y., Wang, W., and Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE)*, pp. 1110–1118.
- Egnor, S.E.R., and Branson, K. (2016). Computational analysis of behavior. *Annu. Rev. Neurosci.* 39, 217–236.
- Elliott, G.R., Vanwersch, R.A.P., and Bruijnzeel, P.L.B. (2000). An automated method for registering and quantifying scratching activity in mice: use for drug evaluation. *J. Pharmacol. Toxicol. Methods* 44, 453–459.
- Forkosh, O., Karamihalev, S., Roeh, S., Alon, U., Anpilov, S., Touma, C., Nussbaumer, M., Flachskamm, C., Kaplick, P.M., Shemesh, Y., and Chen, A. (2019). Identity domains capture individual differences from across the behavioral repertoire. *Nat. Neurosci.* 22, 2023–2028.
- Freund, J., Brandmaier, A.M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., Sachser, N., Lindenberger, U., and Kempermann, G. (2013). Emergence of individuality in genetically identical mice. *Science* 340, 756–759.
- Gal, Y., and Ghahramani, Z. (2016). Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In *International Conference on Machine Learning (Omnipress)*, pp. 1050–1059.
- Gall, J., Rosenhahn, B., Brox, T., and Seidel, H.-P. (2010). Optimization and filtering for human motion capture. *Int. J. Comput. Vis.* 87, 75.
- Gallistel, C.R. (1982). *The Organization of Action: A New Synthesis* (Psychology Press).
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). *Bayesian Data Analysis* (CRC Press).
- Gittis, A.H., and Kreitzer, A.C. (2012). Striatal microcircuitry and movement disorders. *Trends Neurosci.* 35, 557–564.
- Graves, A., Mohamed, A., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE)*, pp. 6645–6649.
- Guo, J.Z., Graves, A.R., Guo, W.W., Zheng, J., Lee, A., Rodríguez-González, J., Li, N., Macklin, J.J., Phillips, J.W., Mensh, B.D., et al. (2015). Cortex commands the performance of skilled movement. *eLife* 4, e10774.
- Hartley, R., and Zisserman, A. (2003). *Multiple View Geometry in Computer Vision* (Cambridge University Press).
- Hebel, R., and Stromberg, M.W. (1976). *Anatomy of the Laboratory Rat* (Williams and Wilkins).
- Heskes, T. (1997). Practical confidence and prediction intervals. In *Advances in Neural Information Processing Systems 9*, M.C. Mozer, M.I. Jordan, and T. Petsche, eds. (MIT Press), pp. 176–182.

- Hong, W., Kennedy, A., Burgos-Artizzu, X.P., Zelinkowsky, M., Navonne, S.G., Perona, P., and Anderson, D.J. (2015). Automated measurement of mouse social behaviors using depth sensing, video tracking, and machine learning. *Proc. Natl. Acad. Sci. USA* 112, E5351–E5360.
- Hussein, M.A. (1973). Skin cleavage lines in the rat. *Eur. Surg. Res.* 5, 73–79.
- Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., and Schiele, B. (2016). Deeppicut: A deeper, stronger, and faster multi-person pose estimation mode (Springer: Proceedings of the European Conference on Computer Vision), pp. 34–50.
- Ionescu, C., Papava, D., Olaru, V., and Sminchisescu, C. (2014). Human3.6M: large scale datasets and predictive methods for 3D human sensing in natural environments. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 1325–1339.
- Isakov, K., Burkov, E., Lempitsky, V., and Malkov, Y. (2019). Learnable triangulation of human pose. In *International Conference on Computer Vision (IEEE)*, pp. 7718–7727.
- Jhuang, H., Garrote, E., Mutch, J., Yu, X., Khilnani, V., Poggio, T., Steele, A.D., and Serre, T. (2010). Automated home-cage behavioural phenotyping of mice. *Nat. Commun.* 1, 68.
- Jinnah, H.A., and Hess, E.J. (2015). Assessment of movement disorders in rodents. In *Movement Disorders (Second Edition)*, M.S. LeDoux, ed. (Academic Press), pp. 59–76.
- Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S., and Branson, K. (2013). JAABA: interactive machine learning for automatic annotation of animal behavior. *Nat. Methods* 10, 64–67.
- Kaluff, A.V., Stewart, A.M., Song, C., Berridge, K.C., Graybiel, A.M., and Fentress, J.C. (2016). Neurobiology of rodent self-grooming and its value for translational neuroscience. *Nat. Rev. Neurosci.* 17, 45–59.
- Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A.L., Dhawale, A.K., Kampff, A.R., and Ölveczky, B.P. (2015). Motor cortex is required for learning but not for executing a motor skill. *Neuron* 86, 800–812.
- Kitagawa, M., and Windsor, B. (2008). *MoCap for Artists: Workflow and Techniques for Motion Capture* (Focal Press).
- Kobak, D., and Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *bioRxiv*. <https://doi.org/10.1101/453449>.
- Kobak, D., and Linderman, G.C. (2019). UMAP does not preserve global structure any better than t-SNE when using the same initialization. *bioRxiv*, 2019.12.19.877522.
- Krakauer, J.W., Ghazanfar, A.A., Gomez-Marín, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron* 93, 480–490.
- Lakshminarayanan, B., Pritzel, A., and Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems* 30, I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds. (Curran Associates), pp. 6402–6413.
- Lamb, A.M., Goyal, A.G.A., Zhang, Y., Zhang, S., Courville, A.C., and Bengio, Y. (2016). Professor forcing: a new algorithm for training recurrent networks. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 4601–4609.
- Lashley, K.S. (1951). The problem of serial order in behavior. In *Cerebral Mechanisms in Behavior* (Wiley), pp. 112–131.
- Lathe, R. (2004). The individuality of mice. *Genes Brain Behav.* 3, 317–327.
- Liu, G., and McMillan, L. (2006). Estimation of missing markers in human motion capture. *Vis. Comput.* 22, 721–728.
- Luo, L., Callaway, E.M., and Svoboda, K. (2018). Genetic dissection of neural circuits: a decade of progress. *Neuron* 98, 256–281.
- Maaten, van der L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Machado, A.S., Darmohray, D.M., Fayad, J., Marques, H.G., and Carey, M.R. (2015). A quantitative framework for whole-body coordination reveals specific deficits in freely walking ataxic mice. *eLife* 4, e07892.
- Mallick, T., Das, P.P., and Majumdar, A.K. (2014). Characterizations of noise in kinect depth images: a review. *IEEE Sens. J.* 14, 1731–1740.
- Merel, J., Hasenclever, L., Galashov, A., Ahuja, A., Pham, V., Wayne, G., Teh, Y.W., and Heess, N. (2018). Neural probabilistic motor primitives for humanoid control. In *International Conference on Learning Representations*, <https://arxiv.org/abs/1911.09451>.
- Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* 21, 1281–1289.
- Merel, J., Botvinick, M., and Wayne, G. (2019). Hierarchical motor control in mammals and machines. *Nat. Commun.* 10, 5489.
- Merel, J., Aldarondo, D., Marshall, J., Tassa, Y., Wayne, G., and Ölveczky, B. (2020). Deep neuroethology of a virtual rodent. In *International Conference on Learning Representations*, <https://arxiv.org/abs/1811.11711>.
- Meyer, A.F., Poort, J., O’Keefe, J., Sahani, M., and Linden, J.F. (2018). A head-mounted camera system integrates detailed behavioral monitoring with multi-channel electrophysiology in freely moving mice. *Neuron* 100, 46–60.e7.
- Miller, E.A., Kastner, D.B., Grzybowski, M.N., Dwinell, M.R., Geurts, A.M., and Frank, L.M. (2019). Robust and replicable measurement for prepulse inhibition of the acoustic startle response. *bioRxiv*. <https://doi.org/10.1101/601500>.
- Mimica, B., Dunn, B.A., Tombaz, T., Bojja, V.P.T., and Whitlock, J.R. (2018). Efficient cortical coding of 3D posture in freely behaving rats. *Science* 362, 584–589.
- Mischiati, M., Lin, H.-T., Herold, P., Immler, E., Olberg, R., and Leonardo, A. (2015). Internal models direct dragonfly interception steering. *Nature* 517, 333–338.
- Nakamura, Y., Yamane, K., Fujita, Y., and Suzuki, I. (2005). Somatosensory computation for man-machine interface from motion-capture data and musculoskeletal human model. *IEEE Transactions on Robotics* 21, 58–66.
- Nestler, E.J., and Hyman, S.E. (2010). Animal models of neuropsychiatric disorders. *Nat. Neurosci.* 13, 1161–1169.
- Nix, D.A., and Weigend, A.S. (1994). Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN’94) (IEEE)*, pp. 55–60.
- Ohayon, S., Avni, O., Taylor, A.L., Perona, P., and Roian Egnor, S.E. (2013). Automated multi-day tracking of marked mice for the analysis of social behaviour. *J. Neurosci. Methods* 219, 10–19.
- Oord, A., van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016a). WaveNet: a generative model for raw audio. *arXiv*, 1609.03499 <https://arxiv.org/abs/1609.03499>.
- Oord, A., van den, Kalchbrenner, N., and Kavukcuoglu, K. (2016b). Pixel recurrent neural networks. *arXiv*, 1601.06759 <https://arxiv.org/abs/1601.06759>.
- Pappas, S.S., Leventhal, D.K., Albin, R.L., and Dauer, W.T. (2014). Mouse models of neurodevelopmental disease of the basal ganglia and associated circuits. *Curr. Top. Dev. Biol.* 109, 97–169.
- Parker, J.G., Marshall, J.D., Ahanonu, B., Wu, Y.-W., Kim, T.H., Grewe, B.F., Zhang, Y., Li, J.Z., Ding, J.B., Ehlers, M.D., and Schnitzer, M.J. (2018). Diametric neural ensemble dynamics in parkinsonian and dyskinetic states. *Nature* 557, 177–182.
- Pavlakos, G., Zhou, X., Derpanis, K.G., and Daniilidis, K. (2017). Harvesting multiple views for marker-less 3D human pose annotations. In *IEEE Conference on Computer Vision and Pattern Recognition (IEEE)*, pp. 1253–1262.
- Pearce, T., Zaki, M., Brintrup, A., and Neely, A. (2018). High-quality prediction intervals for deep learning: a distribution-free, ensembled approach. *arXiv*, 1802.07167 <https://arxiv.org/abs/1802.07167>.
- Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S.-H., Murthy, M., and Shaevitz, J.W. (2019). Fast animal pose estimation using deep neural networks. *Nat. Methods* 16, 117–125.
- Poddar, R., Kawai, R., and Ölveczky, B.P. (2013). A fully automated high-throughput training system for rodents. *PLoS ONE* 8, e83171.
- Seidler, H., Bernhard, W., Teschler-Nicola, M., Platzler, W., zur Nedden, D., Henn, R., Oberhauser, A., and Sjøvold, T. (1992). Some anthropological aspects of the prehistoric Tyrolean ice man. *Science* 258, 455–457.

- Silverman, J.L., Yang, M., Lord, C., and Crawley, J.N. (2010). Behavioural phenotyping assays for mouse models of autism. *Nat. Rev. Neurosci.* **11**, 490–502.
- Simoncelli, E.P., and Olshausen, B.A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216.
- Stephens, G.J., Johnson-Kerner, B., Bialek, W., and Ryu, W.S. (2008). Dimensionality and dynamics in the behavior of *C. elegans*. *PLoS Comput. Biol.* **4**, e1000028.
- Stirn, A. (2003). Body piercing: medical consequences and psychological motivations. *Lancet* **361**, 1205–1215.
- Svoboda, K., and Li, N. (2018). Neural mechanisms of movement planning: motor cortex and beyond. *Curr. Opin. Neurobiol.* **49**, 33–41.
- Takeoka, A., Vollenweider, I., Courtine, G., and Arber, S. (2014). Muscle spindle feedback directs locomotor recovery and circuit reorganization after spinal cord injury. *Cell* **159**, 1626–1639.
- Taylor, G.W., Hinton, G.E., and Roweis, S.T. (2007). Modeling Human Motion Using Binary Latent Variables. In *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J.C. Platt, and T. Hoffman, eds. (MIT Press), pp. 1345–1352.
- Tinbergen, N. (1950). The hierarchical organization of nervous mechanisms underlying instinctive behaviour. In *Symposium for the Society for Experimental Biology (Society for Experimental Biology)*, pp. 305–312.
- Todd, J.G., Kain, J.S., and de Bivort, B.L. (2017). Systematic exploration of unsupervised methods for mapping behavior. *Phys. Biol.* **14**, 015002.
- van Praag, H., Kempermann, G., and Gage, F.H. (2000). Neural consequences of environmental enrichment. *Nat. Rev. Neurosci.* **1**, 191–198.
- Venkatraman, S., Jin, X., Costa, R.M., and Carmena, J.M. (2010). Investigating neural correlates of behavior in freely behaving rodents using inertial sensors. *J. Neurophysiol.* **104**, 569–575.
- Whishaw, I.Q., and Kolb, B. (2005). *The Behavior of the Laboratory Rat: a Handbook with Tests* (Oxford University Press).
- Wiltschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abaira, V.E., Adams, R.P., and Datta, S.R. (2015). Mapping sub-second structure in mouse behavior. *Neuron* **88**, 1121–1135.
- Wu, H.G., Miyamoto, Y.R., Gonzalez Castro, L.N., Ölveczky, B.P., and Smith, M.A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat. Neurosci.* **17**, 312–321.
- Yu, F., and Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions. *arXiv*, 151107122 <https://arxiv.org/abs/1511.07122>.



## STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE                           | SOURCE                       | IDENTIFIER  |
|---|------------------------------|---|
| Chemicals, Peptides, and Recombinant Proteins |                              |   |
| Amphetamine                                   | Sigma-Aldrich                | A5880   |
| Caffeine                                      | Sigma-Aldrich                | C0750   |
| Experimental Models: Organisms/Strains        |                              |   |
| Long-Evans Rats                               | Charles River                | 006   |
| Fmr1-KO, WT Long-Evans Rats                   | Medical College of Wisconsin | RGD 11553874  |
| Software and Algorithms                       |                              |   |
| Motion Mapper                                 | Berman et al., 2014          | <a href="https://github.com/gordonberman/MotionMapper">https://github.com/gordonberman/MotionMapper</a>                       |
| TCN Imputation                                | This Work                    | <a href="https://github.com/diegoaldarondo/MarkerBasedImputation">https://github.com/diegoaldarondo/MarkerBasedImputation</a> |
| Behavioral Analysis Suite                     | This Work                    | <a href="https://github.com/jessedmarshall/CAPTURE_demo">https://github.com/jessedmarshall/CAPTURE_demo</a>                   |
| MATLAB  | Mathworks                    | <a href="https://www.mathworks.com/products/matlab.html">https://www.mathworks.com/products/matlab.html</a>                   |
| Other   |                              |   |
| Motion Capture System                         | Motion Analysis              | Kestrel 4200  |
| Video Cameras                                 | FLIR                         | Flea3 FL3-U3-13S2C  |
| 5 mm H-ZLAF ball lenses                       | Worldhawk Optoelectronics    | Custom, via <a href="https://www.alibaba.com/">https://www.alibaba.com/</a>   |
| Silvering Kit                                 | Angel Gilding                | Cat# A1204D   |
| Earrings                                      | Fire Mountain Gems           | Cat# H20-1585FN; H20-A5314FN  |
| High-strength epoxy                           | Loctite                      | EA0151  |
| Clamps  | Kent Scientific              | INS1005-5   |

## RESOURCE AVAILABILITY

## Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Bence P. Ölveczky ([olveczky@fas.harvard.edu](mailto:olveczky@fas.harvard.edu)).

## Materials Availability

This study did not generate new unique reagents.

## Data and Code Availability

Demonstration code and example datasets are available at [https://github.com/jessedmarshall/CAPTURE\\_demo](https://github.com/jessedmarshall/CAPTURE_demo). The core functions used for the behavioral embedding are available <https://github.com/gordonberman/MotionMapper>. The code used for imputation is available from the authors at <https://github.com/diegoaldarondo/MarkerBasedImputation>. The remainder of the code used for analysis were made using standard approaches in MATLAB 2017b and open source code extensions and are available from the corresponding author on request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

Experiments were performed using female and male Long-Evans rats aged 2-18 months. Female rats were obtained from Charles Rivers (Strain 006). Male *Fmr1*-KO rats (LE-*Fmr1*em4Mcowi) and cagemate controls were obtained from the Medical College of Wisconsin. Knockouts were generated by a CRISPR/SpCas9 knockout of *Fmr1* as described elsewhere (Miller et al., 2019). Loss of the *Fmr1* protein was confirmed by Western Blot. We recorded all *Fmr1*-KO and *Fmr1*-WT rats at 6-10 weeks of age. Genotype of individuals used in the study was confirmed by PCR.

## METHOD DETAILS

### Motion Capture Arena

We performed motion capture recordings using a commercial 12-camera motion capture system (Kestrel, Motion Analysis). To reduce background due to infrared reflections in motion capture recordings, we positioned cameras 5–10 feet from the center of the motion capture recording arena, a 2-foot diameter plexiglass cylinder. The arena was filled with pine chip bedding and a ceramic pedestal (5 inches diameter by 2 inches in height). We placed cameras at two different heights and angled them at 15° and 35° to the horizontal to track the half-silvered motion capture markers across the ~270° azimuthal extent of their retroreflectivity. We secured the plexiglass arena in a custom wooden platform sealed with urethane and painted with ultra-flat black paint (Krylon 1602). We fitted the arena with a custom lever and water spout for animal training (Kawai et al., 2015), and coated reflective elements outside the arena with dulling spray (Krylon 1301) to reduce reflectivity. We placed six RGB video cameras (Flea3 FL3-U3-13S2C, Point Grey) with varifocal lenses (Computar T10Z0513CS) on tripods (Slik Able 300 DX) above and on to the sides of the arena to provide video recordings. The arena was kept on a 12-hour light cycle, and red lights were used at night during RGB video recordings.

### Retroreflective Body Piercings

We fabricated motion capture markers from 5 mm diameter high-index of refraction ball lenses ( $n = 2.0$ , H-ZLAF, Worldhawk Optoelectronics) (Mischiati et al., 2015). To increase the reflectivity of the ball lenses, we cleaned and half-silvered them using a commercial silvering kit (A1204D, Angel Gilding) and custom rubber mold (Mischiati et al., 2015). The half-silvered markers were detectable in the motion capture arena across a 270° azimuthal range. To track the position and angular rotations of the head, three retroreflectors were attached to a custom 3D printed acrylic headcap (30x40 mm, 35 mm tall, 10 g) that was painted using flat black paint (Krylon 1602). We placed the markers in an isosceles triangle to improve marker assignment based on pairwise distances. We fabricated piercings for tracking the trunk and hips by soldering a 6 mm steel cup (H20-1585FN, Fire Mountain Gems) to a Monel clamp (INS1005-5, Kent Scientific). We used high-strength epoxy to affix retroreflectors to the steel cup (Loctite EA0151) that we cured at 175°C, resulting in light (0.6 g) and durable retroreflectors. We chose the angle of the cup to the clamp, and hence the angle of the retroreflectors, to maximize the observability of markers across the range of motion of the rat. We angled markers on the posterior of the animal toward the head to facilitate marker tracking when the animal reared. We angled lateral markers toward the midline to facilitate tracking during trunk twists. Because the skin on the lateral side of the animal was poorly suited to support the clamp piercings, we fabricated 0.4 g markers to be placed on the shoulders, forelimbs and hindlimbs by fusing retroreflectors to steel earstuds with 6mm wide cups using high-strength epoxy.

To track the position and orientation of the animal's head, trunk, and major appendages, as well as provide appropriate asymmetry to uniquely identify motion capture markers using a pairwise body model, we designed a custom 20 marker placement strategy. We placed markers according to skeletal landmarks identified under the skin (Hebel and Stromberg, 1976). We placed three markers along the animal's spine at the sixth thoracic vertebrae (Th6), the first lumbar vertebrae (L1) and the first sacral vertebrae (S1). To provide asymmetry for distinguishing the left and right sides of the animal, we also placed two markers on the animal's left trunk, midway along the anterior-posterior axis between each pair of adjacent spine markers and located 20 mm vertically beneath the spine. Markers over the animal's hips were placed along the femur above the trochanter minor. We used ten markers to track the position and configuration of the forelimbs and hindlimbs. We attached three markers to each forelimb: one over the scapula, 10 mm from the posterior endpoint, one over the olecranon (elbow), and one at the midpoint of the ulna. We attached two markers to each hindlimb: one on the animal's patella and one at the midpoint of the tibia.

### Drug Administration

We injected all drugs at 1 ml/kg. We dissolved Caffeine (Sigma C0750) in Phosphate Buffered Saline (PBS) to achieve a dilution of 10 mg/ml and injected at 10 mg/kg. We administered amphetamine (Sigma A5880) at 0.5 mg/kg in PBS. Vehicle injections were performed using PBS. To administer drugs, we briefly anesthetized animals and allowed them to recover for 10 minutes before we began recordings. Baseline recordings were matched to the same time of day as drug administration.

### Surgery

The care and experimental manipulation of all animals were reviewed and approved by the Harvard University Faculty of Arts and Sciences Institutional Animal Care and Use Committee. All surgical procedures were designed to limit pain and discomfort. Surgeries for attaching body piercings for motion tracking were performed under 1%–2% isoflurane anesthesia. Prior to surgery, we sterilized all tools. We sterilized body piercings by placing them in 70% ethanol for 30 minutes and rinsing with sterile water. We shaved the animal's head, trunk, and limbs using an electric razor. We removed remaining hair on the scalp using depilatory cream (Nair) and sterilized the scalp using betadine. To attach the custom acrylic headcap bearing head markers, we made a longitudinal incision over the animal's scalp and retracted the skin to expose the skull. We placed three skull screws over the cerebellum and temporal lobes and covered the skull with C&B Metabond (Parkell). We affixed the headcap using cyanoacrylate glue and dental cement (A-M Systems, 525000). Sites for the placement of body piercings on the skin were marked using a skin pen and then sterilized using alternating washes of betadine and 70% ethanol.

To attach markers to the spine, trunk and hips, we made two small incisions spaced by 1 cm, and drew a sterile, beveled, 18-gauge needle through the incisions to draw open the sites (Angel, 2009). We then inserted body piercings through the ends of the incision and secured them in place using pliers. Incisions and piercings were oriented perpendicularly to the skin lines of maximal tension (Hussein, 1973). For markers on the shoulders, forelimbs and hindlimbs, we similarly inserted a sterile, 18 gauge hollow needle through two points on the skin spaced 10 mm apart. We inserted the end of the piercing through the hollow end of the needle and retraced the needle from the skin. To then secure limb piercings, we attached earnuts (Fire Mountain Gems, H20-A5314FN) to the back of the piercings. We spaced the earnuts from the skin using damp wooden barriers and soldered them to the piercings using a soldering iron and solder flux. We applied antibiotic cream to marker sites and administered animals buprenorphine (0.05 mg/kg) and carprofen (5 mg/kg) subcutaneously following surgery. Following surgery, rats maintained a stable weight:  $R^2 = 0.06$  for a linear fit between normalized body weight and day after surgery,  $p$  value of the linear correlation coefficient after a linear fit to the data  $p = 0.25$ , 20 time points over 6 animals.

### Motion Capture Recordings

Prior to all recordings, we habituated animals to handling by the investigator for at least 5 days, and animals were handled daily during recording to provide enrichment (van Praag et al., 2000). We habituated animals to the arena for at least one day prior to marker attachment. We allowed animals to recover in the arena for at least one day before beginning recordings. To provide further enrichment we allowed animals to interact with a familiar, gender-matched, conspecific for 30–60 min.

We performed motion capture recordings using a commercial motion capture acquisition program (Cortex, Motion Analysis), on a custom acquisition computer (32 GB RAM, 3.6 GHz Intel i7). Motion capture was recorded continuously at 300 Hz with a 1/2500<sup>th</sup> second strobe and 750 nm illumination. We acquired data in 30-minute epochs with 5 s gaps in between and saved to a local server. We made simultaneous video recordings at 30–50 Hz. Points tracked by motion capture were assigned identities in real-time based on a model of pairwise distance relationships between markers that were fit offline.

Animals that were not part of Fragile X experiments (*Fmr1*-KO and *Fmr1*-WT) were previously trained in a motor skill task as previously described (Kawai et al., 2015; Poddar et al., 2013). Briefly, animals were water restricted to 85% of body weight and given three 1-hour training sessions per day in which they had to press a lever twice within a 700 ms interval to receive a water reward. Failed trials were penalized with a 1.2 s timeout. Animals were trained to asymptote in the automated recording boxes, with a mean inter-press interval trials within 10% of 700 ms and a coefficient of variation of the inter-press interval less than 0.25. We allowed animals to habituate to the motion capture arena until reliable task execution was observed.

### Motion Capture Postprocessing

We performed preprocessing and data analysis on a custom analysis workstation (128 GB RAM, 3.6 GHz Intel i7) using MATLAB (Mathworks, Natick MA). We first smoothed marker data using a 3-frame median filter. We then transformed marker positions into an egocentric reference frame centered on the middle of the animal's spine, which we placed at the origin. In this egocentric reference frame, we aligned all markers in the horizontal (x-y) plane so that the front of the animal's spine was oriented along the y axis. We defined joint angles as projections of the angles between adjacent body segments (Figure S1B) in the y-z (sagittal), x-z (coronal) and x-y (transverse) planes of the egocentric coordinate as well as the inter-segment plane defined by the three points of the adjacent segments.

We labeled frames as moving if the average speed of the spine in a 1 s moving window exceeded 1.5 mm/s. We marked markers as missing if they were not detected by the motion capture array or not fit by the motion capture body model. We additionally excluded measurements in which markers exceeded a velocity threshold of 25 mm/frame, or in which markers on the forelimb and hindlimb showed non-physical configuration relative to markers on the olecranon or patella, respectively.

To impute the position of missing markers we used a three-pronged approach. First, for gaps shorter than 5 frames (17 ms), we applied cubic interpolation separately across each gap for each Cartesian marker position. Second, for gaps longer than 5 frames, in frames in which a marker was missing but 5–8 adjacent markers were well tracked, markers were imputed using a random forest regression trained to predict the position of the marker from the well-tracked adjacent markers. We fit a separate model for each Cartesian component of each marker. Random forest models used 50 trees trained over each the 30-minute recording session, and had less than 1 mm median out-of-bag error for each marker group (Head: 0.53 mm Trunk: 0.34 mm: Forelimbs: 0.85 mm; Hindlimbs: 0.24 mm;  $n = 16$  animals,  $n = 73$  conditions). Lastly, to impute missing marker data over gaps > 17 ms in which a marker and one or more of its adjacency partners were missing, we used a temporal convolutional network that incorporates past temporal information about all markers positions to impute missing data (Oord et al., 2016a) (Figure S3; Methods S1). Our network uses 9 past time points sampled at 60 Hz (150 ms) in a 4-layer neural network that uses dilated causal convolutions with 512 filters per layer. Neurons in the network used a linear activation function. Networks using a rectified-linear activation performed similarly. We trained the network separately over all frames on each full day using a GPU cluster (4x Nvidia V100).

Following imputation, we re-applied the velocity and physical thresholds above to identify poorly imputed markers that we then marked as missing. We additionally applied two other quality thresholds: a threshold of 1 standard deviation on the summed jerk of all markers averaged across a 2.5 s moving window and an additional physical constraint on the sum of pairwise distance  $d_{ij}$  of head markers ( $50 \text{ mm} < \sum d_{ij} < 100 \text{ mm}$ ) for each frame to identify any remaining tracking errors.

To compare the number of markers detected by 6 or 12 cameras (Figure S1E), we recorded five animals using all 12 cameras, and stored the raw motion capture video files as well as the tracked marker positions. We used the motion capture video files to re-track the animals post hoc using Cortex. We tracked animals using the original 12 camera set, as well as four different subsets of 6 cameras: the upper and lower sets of cameras, as well as two different sets of alternating camera pairs. We assigned frames as moving where the velocity of the center of the spine, detected using all 12 cameras, was less than 3 mm/s, and compared detection ability over moving frames. For all points detected using 6 cameras, we assessed them as being on the animal if they were within 20 mm of a marker tracked using 12 cameras.

### Performance Comparisons with 2D Convolutional Neural Networks

Recently, 2D convolutional networks trained from scratch or on human-labeled pose databases have been applied to behavioral tasks to detect un-occluded visual landmarks in animals (Mathis et al., 2018; Pereira et al., 2019). These approaches, as well as keypoint prediction using classical computer vision approaches (Guo et al., 2015; Kabra et al., 2013), can be straightforwardly extended to 3D keypoint detection through triangulation of detected keypoints across multiple calibrated cameras. However, to date, these approaches have been restricted to 3D detection of un-occluded keypoints in individual behaviors. There are principled reasons for this. 3D tracking with these techniques requires that a keypoint be accurately identified in two different cameras before triangulation. As these 2D convolutional networks are pretrained and fine-tuned on human labeled datasets in which only un-occluded markers are labeled, these networks are typically only able to identify un-occluded keypoints in each view. Thus, detection and successful triangulation of keypoints on the often-occluded appendages across a broad behavioral repertoire requires large numbers of cameras and labeled examples. These problems are elegantly solved using motion capture, which uses high-sensitivity markers to improve labeling precision to sub-pixel resolution (10  $\mu$ m in our recordings), and on-board camera FPGAs to allow for integration of large numbers of high-resolution cameras with low data-bandwidth (10 GB/day compared to  $\sim$ 1000 GB/day for compressed video or depth camera recordings as computed from our data). Despite these limitations, these 2D approaches have been suggested as general purpose tools for 3D keypoint detection across behaviors.

To perform comparisons between CAPTURE and predictions from 2D convolutional networks, we synchronized the six video cameras (Motion Capture Recordings, above) with motion capture acquisition using a custom decimator circuit (Arduino Uno, Arduino) that downsampled 300 Hz motion capture trigger signals to 30 Hz. We acquired images at 1320  $\times$  1048 and compressed them in real-time using h264 compression. We encompassed the entire arena (600 mm diameter) in each camera's field of view, so that each camera's pixel size at the center of the arena was less than 1 mm. We calibrated cameras using custom scripts in MATLAB. We used a custom checkerboard with 17 mm spacing to compute camera intrinsics, and a motion capture L-frame (Motion Analysis) to compute camera extrinsics and align cameras to the motion capture world coordinate system. We manually inspected 2D projections of the 3D motion capture data, and in some experiments refined the camera calibrations by manually selecting points on the rat's headcap and using these to recalculate extrinsic camera parameters. All transformations between the 3D world coordinate system and the image frame of individual cameras accounted for radial and tangential distortions of images. All calibrations showed low  $\sim$ 0.3 mm reprojection errors across cameras.

To generate training datasets from motion capture data, we synchronized recordings from five rats bearing markers using two different sets of camera views ( $3.24 \cdot 10^5$  time points;  $1.94 \cdot 10^6$  images; 12 total views). We sampled 192,000 images from recordings from three of these rats, sampling across all views used. To ensure an adequate diversity of poses in this training dataset, we sampled these frames uniformly from 40 partitions of the motion capture recordings made by performing *k*-means clustering (*k* = 40) on the animal's pose estimated by motion capture. To make predictions across variable amounts of training data, we trained networks on randomly sampled subsets of 100, 1000, 10,000, or 100,000 frames from the initial 192,000 frame training dataset. We obtained keypoint predictions in poses for all 20 markers by projecting 3D keypoints detected by motion capture into 2D image frames from all views. We used the remaining two animals as a test dataset to evaluate the network's capacity to generalize to held out animals.

To generate training datasets in rats not bearing markers (Figure S4), we lightly shaved three animals and attached a headcap such that rats were subject to the same physical preparation as animals used for motion capture except for marker attachment. Two human labelers marked the position of each of the 20 keypoints corresponding the body position of motion capture markers that were visible in 225 frames from 3 views (25 unique time points per rat) to compute the inter-human variability in hand-labeling. We used the labels from one human as a training dataset, and a separate set of 225 frames (25 time points per rat) labeled by the same human as a test dataset. We drew samples from the same two sets of camera views used above.

We initialized DeeperCut (the algorithm used in DeepLabCut) using ResNet 101 and used default training configurations without the use of pairwise terms (Insafutdinov et al., 2016; Mathis et al., 2018). We fine-tuned networks on hand-labeled images or motion capture reprojections described above for  $1.03 \cdot 10^6$  steps. We also trained DeepLabCut on the full 192,000 frame set of labeled frames in animals bearing markers, and then additionally fine-tuned this network on the set of 225 hand-labeled images described above to attempt to improve generalization to rats not bearing markers. We triangulated the resulting marker predictions from 2D predictions across multiple views by taking the median vector across all individual pairwise triangulations, which we found was superior to multi-view singular value decomposition (Hartley and Zisserman, 2003). For 3-camera predictions, we repeated this procedure for all possible sets of 3 views and reported average statistics. To compare DeepLabCut with motion capture (Figure 2C), we compared DeepLabCut predictions to motion capture measurements on non-imputed recordings, ignoring frames in which motion capture markers were not tracked. To assess the ability of DeepLabCut or CAPTURE to reconstruct the animal's posture (Figure 2D) we

computed the fraction of body segments (Figure S1B) whose predicted lengths were within 18 mm of the segment lengths measured using motion capture. CAPTURE estimates of pose reconstruction efficacy are derived from the more extensive dataset used in Figure S1.

To compute expected enhancements in DeepLabCut performance when using 12 cameras, we used published measurements of markerless pose tracking performance using variable numbers of cameras (Bala et al., 2020; Iskakov et al., 2019). These reports demonstrate sub-linear increases in measurement precision when adding additional cameras, consistent with the precision of a noisy variable estimated using  $N$  independent measurements scaling as  $1/\sqrt{N}$ . We therefore estimated the error of DeepLabCut predictions made using 12 cameras as  $\sqrt{2}$  below those that use 6 cameras, for each frame, and used these reduced errors when estimating pose reconstruction efficacy (Figure 2D).

### Behavioral Feature Generation and Selection

To detect specific behaviors and kinematic variants of behaviors in our datasets, we generated a set of 140 per-frame features that describe the instantaneous speed and spatial configuration of the motion capture markers, as well as the spatio-temporal trajectories of the markers on a 500 ms timescale (Berman et al., 2014; Brown et al., 2013; Kabra et al., 2013; Stephens et al., 2008). We computed two sets of features for this analysis: (1) a set of 80 features specifically selected to provide information about 37 pairs of behavioral distinctions commonly recognized by rodent researchers, such as rearing, walking, and the subphases of facial and body grooming (Figure S5) (Whishaw and Kolb, 2005), and (2) a more general set of descriptors to enable discrimination between kinematic variants of these behaviors.

To select the features most informative about behavioral distinctions, we first generated a set of 985 features describing the pose kinematics of the rat in a 500 ms window. This 985-feature set consisted of both *per-marker* features describing the velocity of individual markers on different timescales, and *whole-organism* features, which conveyed information about the relative position and velocity of markers as a group. Per-marker features included the cartesian velocity components of each marker in the animal's egocentric reference frame, smoothed on 100, 300, and 1000 ms time intervals, as well as the moving standard deviation of each velocity components within each interval (Kabra et al., 2013). We additionally included features encapsulating the animal's overall speed: the average velocity and standard deviation of the animal in the world reference frame in each time interval. To compute whole-organism features, we combined information across the 10 markers on the top of the animal, which included the head, trunk, and hips. We reasoned that this set of posterior markers would be sufficient to classify the animal's behavior because of the success of depth cameras in the same task (Hong et al., 2015; Wiltchko et al., 2015). We computed the top 10 principal components of this marker set's Cartesian position, segment lengths and selected joint angles over time. To additionally compute features with frequency-specific information, we computed a wavelet transform of each of these pose features using 25 morlet wavelets spaced between 0.5 and 60 Hz. This yielded a set of 250 time-frequency coefficients that we compressed by computing the top 15 principal components of the wavelets. In all cases in which we computed the principal components of a set of pose or wavelet features, we used the top eigenvectors from one rat as a fixed basis set to compute the top principal components of each of these feature categories across all rats.

We then selected features that provided information about behavioral distinctions. Two observers used a custom graphical user interface to annotate a subset of  $1.56 \cdot 10^4$  motion capture frames with one of 37 commonly recognized rodent behaviors across  $n = 3$  animals. We computed the information each feature provided about pairs of behavioral distinctions (Figure S5). We selected 80 features, chosen by the presence of an approximate knee in the cumulative pairwise discriminability, that were most informative about behavioral distinctions: the top 10 and 6 principal components, respectively, of the Euclidean pose and segment lengths, the top 10, 15, and 15 principal components, respectively, of the wavelet transform of the joint angles of both the whole animal and, separately, the head and trunk, the relative speed of the head, hips, and trunk within a 100 and 300 ms window, the absolute speed of the trunk and its standard deviation in a 100 and 300 ms window, and the z-component of the trunk and head velocity averaged over a 100 and 300 ms window.

To provide greater ability to distinguish between new behaviors and kinematic variants of behavior not previously recognized by rodent researchers, we computed a set of more general features describing the configuration and kinematics of all 20 markers. We generated a tree of links between markers that approximated the major joint angles of the head, neck, trunk, forelimbs and hindlimbs (Figure S1), and computed the segment lengths, joint angles, and Cartesian pose of these links. We computed the top ten principal components of each of these feature categories. To provide a set of kinematic descriptors of each frame, we computed the wavelet transform of each of these ten principal components, using the same wavelet parameters as above. We computed all principal components using a common set of eigenvectors computed from one rat. Concatenating these postural and kinematic descriptors produced a 60-dimensional feature set that we combined with the features selected above to yield a 140-dimensional feature vector for each frame, which we then whitened.

To compare embeddings made using Cartesian or joint angle representations of posture (Figure S5E), we used the principal components and wavelet transforms of the Cartesian pose, or the joint angles and segment lengths, respectively.

This 140-dimensional feature set was sufficient for separating and commonly recognized behaviors and their kinematic variants (Figure 3). Some of these features, such as the principal components of the Cartesian and joint angle eigenpostures contain redundant information (Figure S5E) suggesting a more parsimonious feature space may be used. Conversely, some behaviors, such as sniffing, that are subtle or difficult to hand-annotate may not be emphasized in the feature selection approach above.



Other feature engineering approaches, guided by ground truth CAPTURE datasets, could be used to weight, select, or expand this to facilitate interpretable behavioral embedding and classification.

### Behavioral Embedding, Clustering, and Systematics

We created behavioral maps by embedding behavioral feature vectors in two-dimensions using t-SNE (Berman et al., 2014; Maaten and Hinton, 2008). To create a co-embedding across all rats, we concatenated the feature matrix of frames in which animals were moving for 16 rats across 73 different behavioral conditions ( $1.04 \cdot 10^9$  frames). We subsampled this feature matrix at 1 Hz to create a feature vector comprising  $\sim 10^6$  time points. Because t-SNE uses an adaptive similarity metric between points, when we created embeddings by uniformly sampling the data the embeddings were dominated by large regions when the animal was relatively still or adjusting its posture. We thus balanced the feature set (Berman et al., 2014) by performing k-means clustering on the full  $\sim 10^6$  frame feature matrix using 8 clusters. We drew 30,000 samples from each cluster to create a 240,000 frame feature matrix that we embedded using a multi-core implementation of t-SNE. We found that adding further samples resulted in overcrowding of the t-SNE space (Kobak and Berens, 2019). We performed all t-SNE embedding using the Barnes-Hut approximation with  $\theta = 0.5$  and the top 50 principal components of the feature matrix. For co-embeddings across rats, we used a perplexity of 200, which we found produced superior results for large feature sets. After creation of the embedding space, we re-embedded out-of-sample points in two steps. First, we found the 25 nearest neighbors to the out-of-sample point in the 140-dimensional feature space of whitened features. Next, we found the position of the first nearest neighbor in the embedding space. We took the position of the out-of-sample point as the median position of all the 25 nearest neighbors within a 3 unit radius of the closest nearest neighbor (Kobak and Berens, 2019).

To create behavioral clusters, we smoothed the behavioral map with a Gaussian kernel of width 0.125,  $\sim 2$  times the width of the spatial autocorrelation of the t-SNE map. We then clustered the data using a watershed transform (Berman et al., 2014). After clustering, two observers defined the kinematic criteria for assigning behavioral clusters into one of 12 coarse behavioral categories, such as walking, rearing, or grooming. The observers also established criteria for further assigning clusters to one of  $\sim 80$  fine behavioral categories, such as 'low rear', 'high rear', and 'right head scratch', that provided additional detail regarding the exact posture and kinematics of the animal. Each observer then watched 24 instances of each behavior selected at random from one animal and assigned each behavioral cluster into a coarse and fine behavioral category. Disagreements were resolved through discussion. Coarse behavioral boundaries drawn on the behavioral maps are hand drawn approximations to the occurrence of coarse behavioral labels in the dataset.

### Temporal Pattern Matching Algorithm

To identify patterns of repeated behaviors, we began with an ethogram encapsulating the occurrence of  $K$  behaviors over  $M$  frames. We smoothed the ethogram with a boxcar filter across the temporal dimension, using filter windows of  $\tau = 15, 120$  s. We normalized each frame, yielding a behavioral probability density matrix. In this matrix, individual frames reflect the probability of behavioral usage in a window of length  $\tau$ . We then computed the correlation coefficient between all frames of the density matrix, yielding a behavioral similarity matrix of dimensions  $M \times M$ .

Off-diagonal elements in this behavioral similarity matrix correspond to pairs of temporal epochs with similar behavioral usage. To identify repeated behavioral patterns, we thresholded the matrix, retaining regions with a correlation coefficient in their behavioral density vector greater than 0.3. We then performed a watershed transform of this thresholded similarity matrix, where each region extracted using the watershed transform corresponded to a pair of temporal epochs with similar behavioral usage in the dataset. We retained only regions of length  $> \sqrt{2} \tau$  seconds, and additionally excluded pairs of neighboring time points within  $2 \tau$  of one another, which may retain causal influence from the filtering process. For single day recordings, this procedure produced 50,000–100,000 epochs of similar behavioral usage to at least one other temporal epoch. To detect frequently occurring patterns, we computed the correlation distance between these epochs, and clustered the resulting distance matrix using hierarchical clustering, with a cutoff of 0.65 and tree depth of 3. The algorithm was linearly sensitive to the clustering cutoff used in a similar manner across animals. We chose a cutoff so that commonly accepted patterns such as the skilled tapping task were identified and not overly split across clusters.

The resulting clusters, which we refer to as 'sequences' and 'states' for short and long timescales, respectively, were marked as grooming if animals spent at least 40% of their time within the cluster grooming. Clusters were marked as active or inactive if they spent greater time during active behaviors (walking, investigating, rearing, and wet dog shake) compared to inactive behaviors (prone still, postural adjustment) or vice versa. Clusters were marked as task-related if they comprised at least 10% of the task-related time points, which were defined as those frames within 5 s or 30 s of a lever tap for sequences and states, respectively. Comparisons to Markovian behavior were made by simulating behavioral traces of identical length to the observed data using the average observed transition matrix and identifying sequences and states in the same manner as real data.

In cases when the algorithm was run over multiple days and animals, the number of epochs exceeded the memory capacity of hierarchical clustering algorithms, which rely on creating a full matrix of pairwise distances. To identify behavioral patterns in this case, we subsampled 50,000 epochs from the data in a balanced manner, by running k-means clustering on the dataset with  $k = 300$ , and then evenly sampling from the clusters without replacement. We performed hierarchical clustering using the same parameters as above on this subsampled dataset and performed nearest-neighbors re-embedding to assign clusters for the remainder of the data.

### Behavioral Comparisons before and after Body Piercings

To compare behavioral usage before and after attachment of body piercing, we created a 122-dimensional feature set describing the movements of the head in the arena. This feature set included the instantaneous position and velocity of the head in the arena's Cartesian coordinate frame, as well as in an egocentric coordinate frame centered on the head. We computed the roll and pitch angles of the head relative to the global Cartesian coordinate system. Because the yaw of the head can only be defined in a relative manner without measurements of the neck position, we only computed the absolute velocity of yaw deviations across neighboring frames. We smoothed all velocities with a 5-frame median filter, and to incorporate longer-timescale information about head movements, we convolved each Cartesian and egocentric velocity and position component with a boxcar filter on 200 ms and 1 s timescales. We additionally included 25 principal components of the wavelet transform of the Cartesian and egocentric velocity in, using the same wavelet parameters as described above.

To automatically identify the behavior of animals bearing head markers alone, we trained a random forest classifier with 50 trees to predict the animals' behavior from the 122-dimensional feature set describing the kinematics of the headcap. To train the classifier, we assigned behavioral labels to the days after marker attachment, and used half of the data as training set, reserving the second, interleaved, half as a test set. We assumed a uniform prior probability of observing each coarse behavioral class.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Motion Capture Precision

Motion capture uses FPGAs on cameras to identify marker positions in real time, obviating the need to stream high-definition video and permitting the use of large camera arrays to track points without prohibitive data storage and processing requirements. Because of the high signal-to-noise of markers and the large number of high-definition, high bit-depth cameras used, motion capture is extremely precise, and stationary markers in our arena were tracked with 10  $\mu\text{m}$  precision (Mischiati et al., 2015). Much of the error in motion capture measurements comes from changes in the subset of cameras used to record markers as animals move in the arena, which will vary in their estimate of a marker's position due to noise in the camera calibration. To estimate this error due to variation in the camera subset used, we computed the difference between the measured marker position in 3D space, averaged across all cameras, and the position of the marker projected into each camera. We then averaged this reprojection error across markers to arrive at an estimate of the overall precision taken over 5 rats:  $0.21 \pm 0.07$  mm for 12 cameras,  $n = 8 \cdot 10^7$  frames;  $0.22 \pm 0.08$  mm for 6 cameras,  $n = 2.7 \cdot 10^6$  frames; mean  $\pm$  s.d..

We then estimated the overall error of CAPTURE (Figure S1G) for each marker  $m$  as:  $E_m = f_r e_r + f_i \sum_{\tau} f_{\tau} e_{\tau}$ , where  $f_r$  is the fraction of time points when the markers was well-tracked,  $e_r$  is the motion capture reprojection error,  $f_i$  is the fraction of frames where the marker was imputed,  $f_{\tau}$  is the fraction of imputed frames with a gap length of  $\tau$  and  $e_{\tau}$  is the mean error across synthetic gaps with length  $\tau$  for the chosen marker.

### Behavioral Power Spectra

To compute the power spectra of marker velocities for each behavior, we extracted a  $[-1, +1]$  s windows of time around all instances of a behavior from a single day. We computed the instantaneous Cartesian velocity of each marker in each time window in the egocentric reference frame centered on the animal. We concatenated these velocities and computed the power spectrum of each component using Welch's method, with a 1 s bin size and 0.5 s overlap between bins. We averaged the power spectra across Cartesian components to yield the total power spectrum. To temporally align individual instances of behavior, we computed the cross-correlation between the z-velocity of a chosen marker across all behavioral instances, using a 500 ms lag. To further maximize the similarity of behaviors in each cluster, we divided instances into three clusters, and temporally shifted behaviors within a cluster to maximize the cross-correlation. Only clusters with at least 5 observed instances are shown.

### Transition Matrix Eigenvalue Analysis

To assess the existence and extent of non-Markovian structure in rat behavior, we computed the eigenvalue-decomposition of the behavioral transition matrix,  $T_{ij}(\tau)$ , where  $T_{ij}(\tau) = p(X_{t+\tau} = i | X_t = j)$  for two behaviors  $i$  and  $j$ . Since this transition matrix is right-stochastic in all cases, by the Perron-Frobenius Theorem, the leading eigenvalue should be exactly equal to one, and all other eigenvalues must have magnitudes less than one (Berman et al., 2016). The eigenvector corresponding to the largest eigenvalue is proportional to the overall probability of observing a given state, while other eigenvectors correspond roughly to linear combinations of behavior that exhibit the longest temporal correlations. Here, we look at the magnitudes of the largest eigenvalues ( $\lambda_1, \lambda_2, \dots$ ) of this matrix as function of  $\tau$  (Figure S6A). For comparison, we also computed the eigenvalues of the transition matrix  $T_{ij}(\tau)$  made from simulated data assuming that behavior is organized as a first-order Markov chain. We simulated a Markov chain separately for each day from the average observed transition matrix, to create a chain of identical length to the observed data.

### Similarity of Behavioral Patterns Across Timescales

To compute the similarity of temporal pattern usage on two different timescales  $s(t_1)$  and  $s(t_2)$  (Figure S6D), we computed the fraction of time the two time series of cluster labels were in agreement were in agreement after optimal permutation of the

cluster labels: for each temporal pattern  $j$  at short timescales  $t_1$  we found the temporal pattern  $i$  at the longer timescale  $t_2$  with the most similar distribution over time. We computed the fractional overlap of these two temporal patterns  $f(j) = |\{s(t_1) = i\} \cap \{s(t_2) = j\}| / |\{s(t_1) = i\} \cup \{s(t_2) = j\}|$  and computed the average of this fractional overlap over all patterns  $j$ , weighted by the frequency of the observed pattern, to yield the overall similarity.

To compute the average sparsity of the pattern transition matrix (Figure S6E), for each pattern  $p$  on timescale  $\tau$  we computed  $S_p(\tau) = \sum_i \rho_i \sum_j T_{i \rightarrow j}^2$ , where  $\rho_i$  is the frequency of observing behavior  $i$ , and  $T_{i \rightarrow j}$  is the probability of transitioning between behaviors  $i$  and  $j$ . If a behavior transitions to  $n$  other behaviors with equal probability, then  $\sum_j T_{i \rightarrow j}^2 = 1/n$ . Thus higher sparsity values  $S_p(\tau)$  indi-

cate that, on average, behaviors in a pattern transition to a smaller set of behaviors, and hence the pattern shows greater sequential organization. In Figure S6E we report the average sparsity over all patterns  $(1/M_p) \sum_p S_p(\tau)$ , where  $M_p$  is the number of patterns,

normalized by the average over all timescales. To ensure that our estimates of the transition matrix for each pattern are robust, we computed the sparsity for 25 different samples of behavior for each pattern by randomly selecting a consecutive subset of half the observed frames for each sample.

### Cross-condition Behavioral Comparisons

To compare the frequency of behaviors across perturbations induced by drugs and disease we computed the number of occurrences of individual behaviors or coarse behavioral categories for each rat and condition at 6 Hz sampling (drug experiments) or 1 Hz sampling (Fragile X experiments). To test for significant changes in the fraction of time animals spent moving or in coarse behavioral categories (Figures 5A, 5B, 6A and 6B), we first computed the effective sample size of temporal processes using  $n_{\text{eff}} = n / (1 + 2\tau)$ , where  $\tau$  is the temporal autocorrelation of the timeseries (Gelman et al., 2013). The autocorrelation of the binarized moving-not moving timeseries was  $\tau = 469 \pm 157$  s, while the autocorrelation of the coarse behavioral timeseries and behavioral timeseries were  $62 \pm 12$  s and  $0.65 \pm 0.1$  s, respectively (mean  $\pm$  s.d., computed over 10 days of recordings and 5 rats).

To examine the up- or downregulation of specific behaviors, we first restricted analyses to a set of frequent behaviors observed in at least 120 samples across all conditions (20 or 120 s for drug and Fragile X experiments, respectively). To determine whether these frequent behaviors were significantly modulated across conditions we computed the Poisson probability of the observed behavioral frequency in the perturbed condition, given the behavioral frequency in the baseline condition. We then adjusted all Poisson p values by the Benjamini-Hochberg procedure, and only showed behaviors whose adjusted p value was less than  $10^{-6}$ .

### Individuality Analyses

To compare behavior across days and animals, we selected days for analysis from 5 female Long-Evans rats where no drugs or vehicle were administered. We varied the number of clusters in Figure 7F by varying the kernel used to create a density map before watershed clustering. To compare usage of sequences and states, we ran our pattern detection algorithm over all animals and days considered. To maximize similarity in sequences and states detected across animals, we used a density of 0.2 for smoothing the behavioral map, which corresponds to 527 clusters in Figure 7F. We observed similar results using other densities.

To decode animal identity on each day given information about the animal's behavioral and pattern usage, we trained a random forest classifier with 50 trees. We divided periods of active behavior on each day into 10-minute segments and computed the average behavioral and pattern usage within each segment. We trained the algorithm on half of the segments of each day, and tested on the second, interleaved, half. We found that training the classifier using shorter input segment lengths showed poorer performance, especially when the classifier only had access to behavioral usage statistics.

### Data Analysis and Statistics

We performed all filtering and data analysis using custom software written in MATLAB (Mathworks). Unless otherwise stated, we used non-parametric statistical tests (two-tailed) to avoid assumptions that distributions were normal and had equal variance across groups. Boxes in all box-and-whisker plots corresponded to the interquartile range, with lines at the median. Whiskers spanned the smallest and largest data point within  $\pm 1.5$  times the interquartile range. No data was excluded in this study. Because of the proof-of-principle nature of this work, investigators were not blinded, nor were pre-determined sample sizes used to determine cohort sizes. Replication of these experiments was outside the scope of the current work.