# Mapping Mouse Behavior with an Unsupervised Spatio-temporal Sequence Decomposition Framework

**Kang Huang[1,2,4], Yaning Han[1,2,4], Ke Chen[1,2], Hongli Pan[1], Wenling Yi[1,2], Xiaoxi Li[1,2], Siyuan Liu[3], Pengfei Wei[1,2,5,*], Liping Wang[1,2,5,*]**

1. Shenzhen Key Lab of Neuropsychiatric Modulation and Collaborative Innovation Center for Brain Science, Guangdong Provincial Key Laboratory of Brain Connectome and Behavior, CAS Center for Excellence in Brain Science and Intelligence Technology, Brain Cognition and Brain Disease Institute (BCBDI), Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Shenzhen 518055, China.
2. University of the Chinese Academy of Sciences, Beijing 100049, China.
3. Smeal College of Business Penn State University, 8082 University Park, Pennsylvania, US.
4. Co-first authors
5. Lead Contact

*Correspondence: pf.wei.ac.cn (P.W.), lp.wang@siat.ac.cn (L.W.)

## Abstract:

Objective quantification of animal behavior is crucial to understanding the relationship between brain activity and behavior. For rodents, this has remained a challenge due to the high-dimensionality and large temporal variability of their behavioral features. Inspired by the natural structure of animal behavior, the present study uses a parallel, multi-stage approach to decompose motion features and generate an objective metric for mapping rodent behavior into the animal's feature space. Incorporating a three-dimensional (3D) motion-capture system and unsupervised clustering into this approach, we developed a framework that can automatically identify animal behavioral phenotypes from experimental monitoring. We demonstrate the efficacy of our framework by generating an "autistic-like behavior space" that can robustly characterize a transgenic mouse disease model based on motor activity without human supervision. Our results suggest that our framework features a broad range of applications, including animal disease model phenotyping and the modeling of relationships between neural circuits and behavior.

**Key Words:** Behavioral mapping; 3D motion capture; Computational ethology; Unsupervised learning; Behavior phenotyping.

## Introduction:

A fundamental issue in modern neuroscience research is linking specific neural circuit activity to its corresponding behavior [1–3]. In rodent model-based neural circuit studies, cutting-edge

neural population activity recording and cell-specific neural circuit manipulating techniques provide us unprecedented opportunities to deeply gain insight into the mechanisms of neural-behavioral encoding and generation [4,5]. Meanwhile, automated and high-throughput quantification and description of animal behavior are becoming increasingly popular [6,7]. The recent emergence of automated animal pose estimation toolboxes has dramatically facilitated body parts tracking [8–10], specific well-defined behaviors (e.g., grooming, locomotion) can thus be detected based on body features with supervised approaches [11]. However, most naturalistic rodent behaviors are highly complex and variable; The labor-intensive, repetitive and biased manual labeling is insufficient to produce high-quality training sets [12,13]. Therefore, how to identify and categorize these un-predefined behavior remains a challenging task.

Existing studies on lower animals such as flies [14,15], zebrafishes [16–18] and C. elegans [19–22] have used unsupervised strategies and multivariate analysis to segment behavioral modules. These leveraged approaches have applied to behavior structure uncovering and modeling [19,23,24], neural circuits dissecting [25–27], and brain-wide neural-behavioral mapping [15,28]. However, the application of this strategy to the study of mammals has been obstructed by two problems. First, mammalian behavior is relatively more dynamic and high-dimensional. Dozens of degrees of freedom (DoF) [29] and complex 3D characteristics result in the high variability of mammalian behavior on a temporal scale. To define the start and end boundaries to segment continuous data into behavioral sequences, many machine learning-based open-source toolboxes [30] and commercial software did excellent works in feature engineering. They first compute per-frame features that refer to position, velocity, or appearance-based features. The sliding windows technology then converts them into window features to reflect the temporal context [11,31]. Although these approaches effectively identify specific behaviors, behavior recognition becomes problematic when the dynamics of particular behaviors cannot be represented by window features. Furthermore, mammalian behavior is highly variable. Even for similar behaviors, the duration and composition postural sequences vary. The apparently simple action of recognizing behavior actually corresponds to the extraction of shared high-level features from low-level vision data [32]. A recently developed toolbox called MoSeq [33] assumed that the behavioral modules can be modeled as an autoregressive hidden Markov model (AR-HMM). Thus, the variability of behavior can be reduced by inferring a specific number of hidden states.

The present study incorporates a 3D motion-capture system into a general-purpose framework to decompose animal behavior into metrizable behavioral modules. Informed by the natural structure of animal behavior reported in previous theoretical studies, our framework uses a two-stage (pose and movement) behavior decomposition to reduce variation. We characterize motion dynamics by applying the dynamic time alignment kernel (DTAK) method, which generates an objective metric to measure the similarity between movement sequences. Combined with locomotion information, this

metric constructs a feature space of naturalistic behavior. We then apply this framework to the discovery of behavior modules by implementing unsupervised clustering and demonstrate that this framework can assess the spontaneous behavior of a transgenic animal disease model. Hence, by mapping mouse behavior into a feature space without human supervision, we show that this approach can reveal the behavioral signature of animals with different genotypes.

## Results:

### Framework of the Unsupervised Animal Movement Analysis

Our framework first requires the preparation of the animal postural feature data (Fig. 1A). These data can be continuous body parts trajectories that comprehensively capture the motion of the animal's limbs and torso. Theoretically, the natural characteristics of animal movement involve locomotion and non-locomotor movement (NM) [34–36]. Locomotion can be represented by velocity-based parameters. NM is manifested by movement of the limbs or organs without movement of the torso and is controlled by dozens of DoF. Hence, we adopted a parallel motion decomposition strategy to extract features from these time-series data independently (Fig. 1B, C). A two-stage dynamic temporal decomposition algorithm is applied to the centralized animal skeleton postural data to obtain the NM space. Finally, together with the additional velocity-based locomotion dimension, unsupervised clustering is used to reveal the structure of the rodent's behavior.

Our framework has two main features. First, it addresses the multi-time scale of animal behavior[37]. Animal behavior is self-organized into a multi-scale hierarchical structure from the bottom up, including poses, movements, and ethograms. [38,39]. The poses and movements are low- and intermediate-level elements [32], while higher-level ethograms are stereotyped patterns composed of movements that adhere to inherent transfer rules in certain semantic environments [40]. Our two-stage pose and movement decomposition focuses on extracting the NM features of the first two layers. Second, our framework emphasizes the dynamic and temporal variability of behavior. The most critical aspect of unsupervised approaches is to define an appropriate metric for quantifying the relationship between samples. However, the duration and speed of NM segments of the same cluster may differ. To address this, we used a model-free approach called DTAK as a metric to measure the similarity between the NM segments and thus equip the model to automatically search repeatable NM sequences. We then apply the uniform manifold approximation and projection (UMAP) [41] algorithm to visualize high-dimensional NM representations. After combining the locomotion dimension with NM space [Fig. 1C], we adopted hierarchical clustering to re-cluster the components and map the behavior's spatial structure [Fig. 1D].

### Colleting Mouse Motion Data with a 3D Multi-view Motion Capture System

To fully characterize the kinematics of free-moving animals, we developed a 3D multi-view motion capture system (Fig. 2A, B). This system integrates the behavioral apparatus (Supp. Fig. S1), camera calibration (Supp. Fig. S3, 4), multi-view video stream acquisition, pose estimation (Supp. Fig. S5) [8], and 3D skeletal reconstruction [42]. We collected the naturalistic behavioral data of free-moving mice in a featureless circular open-field (Fig. S1, Video 1). We analyzed the mouse skeleton as 16 parts (Fig. 2C) to capture the movements of the rodent's head, torso, paws, and tail, and the following motion quantification did not involve the motion features of two parts of the tail (). The data obtained from tracking representative mouse poses tracking (Fig. 1D) includes the 3D coordinates (x, y, and z) of the body parts, which reveal that the high-dimensional trajectory series exhibits periodic patterns within a specific time scale. On the trajectory, there are pattern switches between these segments. To evaluate the quality of the 3D skeletal data, we checked the DeepLabCut (DLC) tracking likelihood in the 12×4 videos (0.9807±0.1224, Fig. 5). After artifact detection and correction, we calculated the overall reconstruction quality (0.9981±0.0010, Fig. 2D, Fig. S6A) to ensure that the data were qualified for downstream analysis.

## Decomposing Non-Locomotor Movements with Dynamic Time Alignment Kernel

Conceptually, behavior adheres to a bottom-up hierarchical architecture (Fig 3A) [38,39], and research has focused on elucidating behavioral component sequences contained in stimuli-related ethograms [43]. The purpose of the two-stage NM decomposition is to bridge the low-level vision features (postural time-series) to high-level behavioral features (ethograms). The first stage of the decomposition involves extracting postural representations from postural feature data. Since the definition of NM does not involve the animal's location or orientation, we pre-processed these data through center alignment and rotation transformation (Supp. Fig. 6). The continuous sampling of animal poses is usually subject to redundancy attributable to neighboring poses of varying similarity [44]. Therefore, for computational efficiency, we adopted a temporal reduction algorithm to merge the adjacent similar poses as postural representations in a local time range.

In the second stage, NM modules are detected from temporal reduced postural representations. Unlike the static property of poses, mammalian movements have high dimensionality and large temporal variability [45]: e.g., the contents, phases, and durations of the three pose sequences were not the same (Fig 3A). Hence, we adopted a model-free approach to dynamically perform temporal aligning and cluster the temporally reduced postural representation data (Fig. 3B) [46]. This problem is equivalent to providing a d-dimensional time-series $X \in \Re^{d \times n}$ of animal postural representations with $n$ frames. Our task decomposes $X$ into $m$ NM segments, each of which belongs to one of the corresponding $k$ behavioral clusters. This method detects the change point by minimizing the error across segments; therefore, dynamic temporal segmentation becomes a problem of energy minimization. To model the temporal variability and provide a suitable metric for evaluating the

differences between NM segments, we used the DTAK method, which extends from dynamic time warping (DTW), to measure the similarity between time sequences and construct an energy equation (objective function) for optimization. The relationship between each pair of segments was calculated with the kernel similarity matrix *K* (Fig. 3C). DTAK was the used to compute the normalized similarity value of *K* and generate the paired-wise segment kernel matrix *T* (Fig. 3D).

Because dynamic temporal segmentation is a non-convex optimization problem whose solution is very sensitive to initial conditions, this approach begins with a coarse segmentation process based on the spectral clustering method, which combines the kernel k-means clustering algorithms. To define the time scale of segmentation, the algorithm sets the maximum and minimum lengths $[w_{min}, w_{max}]$ to constrain the length of the behavioral component. For the optimization process, a dynamic programming (DP)-based algorithm is employed to perform coordinate descent and minimize energy. For each iteration, the algorithm updates the segmentation boundary and segment kernel matrix until the decomposition reaches the optimal value (Fig. 3E, F). The final segment kernel matrix represents the optimal spatial relationship between these NM segments, which can be further mapped into its feature space in tandem with dimension reduction.

We demonstrate the pipeline of this two-stage behavior decomposition (Fig. 3H) in a representative 300-s sample of mouse skeletal data. The raw skeletal traces were segmented into NM slices of an average duration of 0.89 ± 0.29 s. In these segments, a few long-lasting movements occurred continuously, while most others were intermittent (Fig. 3G). The trajectories of these movement slices can reflect the actual kinematics of the behaving animal. For instance, when the animal is immobile, all of its body parts are still; when the animal is walking, its limbs show rapid periodic oscillations. Consistent with our observations, the movements corresponding to the other two opposite NMs, left and right turning, tended to follow opposite trajectories. These preliminarily results demonstrated that DTAK can be used for the decomposition and mapping of NMs.

## Mapping Mouse Movements with Low-Dimensional Embeddings and Unsupervised Clustering

We validated our framework in a single-session experiment with free-moving mouse behavioral data collected with the 3D motion capture system. First, the two-stage behavioral decomposition strategy decomposed the 15-minute experimental data into 936 NM bouts (Video 2). A 936×936 segment kernel matrix was then constructed using the DTAK metric. This segment kernel matrix could flexibly represent the relationship and provide insight into the relationships between each behavioral component sequence in their feature space. For visualization purposes, we adopted an algorithm called UMAP that can preserve both the local and global structure of the dataset and provide 2D embeddings of these NM segments. In addition, in our parallel feature fusion framework, the factor of an animal's interaction with the environment – i.e., velocity – is considered an

independent dimension. Together with 2D NM embedding, they construct a spatio-temporal representation of movements (Fig. 4A).

We used an unsupervised clustering algorithm to verify the spatio-temporal representation of animal behavior and identify the movement phenotypes. First, we determined the number of behavioral phenotypes of this dataset. Using the Bayesian Information Criterion (BIC) [47] to model the structure of the data and combining the model with the practical situation of the sample distribution in 3D space, the optimal cluster number was determined to be 11 (Supp. Fig. 9). We then recalculated the similarity matrices in the new feature space (Fig. 4B) and aggregated them using a hierarchical clustering method. Finally, we cut the original video into clips of $0.963 \pm 0.497$ s and manually labeled them according to the behavior of the rodents in the clip: running, trotting, stepping, diving, sniffing, rising, right turning, up stretching, falling, left turning, and walking (Supp. Behavioral phenotypes definition). The locomotion types of running, trotting, stepping, and walking accounted for 20.6% of the total activities, indicating that animals spent most of the time in the NM stage (Fig. 4C).

Although we phenotyped all the clips of the entire video, it was difficult to label the behaviors of the rodents with only 11 definitions. Further, there are various heterogeneous transition stages between bouts of stereotyped movements [12,14,36]. Therefore, we evaluated them by calculating the intra-cluster and inter-cluster correlation coefficients (intra-CC and inter-CC, respectively; Fig. 4D, Fig. 5B). Our results showed that running, up stretching and left turning have higher intra-CC and lower inter-CC, while walking and sniffing have both higher intra-CC and higher inter-CC. This is because walking and sniffing co-occur with other movements [2], such as diving and turning, respectively. Finally, to evaluate the overall clustering quality, we integrated these two parameters and defined the Clustering Quality Index (CQI, Fig. 4E), which helped to determine the stereotyped/non-stereotyped movements.

## Kinematic Validation of Mouse Behavioral Phenotypes

DTAK is an abstract extraction of animal motions that aims to simplify the complex temporal dynamics of behavior. Hence, we further elucidated whether the spatial kinematics of the original postural time-series of the behavioral phenotypes identified with this framework were homogeneous. Manually inspecting the position, moving, bending, and other characteristics of the mouse limbs and trunk of the video clips of each phenotype group (Video 3), we found reliable homogeneity for clips with high CQIs (CQI>0.75). We examined the average skeleton of all frames for each movement cluster (Fig. 5A). While some movements could be clearly recognized (e.g., left and right turning, and up stretching), the differences between movements with similar postures (running, trotting, walking, etc.) were not. However, as such unclear differences should be reflected in the moving intensity (MI) of the body parts, we computed the MI of all body parts during movement.

The data show that the horizontal MI components of running and trotting are the highest, followed by stepping and walking. Vertical MI components (e.g., up stretching, rising, and falling) feature richer detail; we attribute their high overall vertical MI to the movement of the nose and front claws (Fig. 5A, C, D, E). This approach of creating portraits for each type of movement provides further support for the efficacy of our framework in the decomposition of animal behavior. The movement lineage analysis revealed that similar movements were arranged were closely, such as running and trotting. Interestingly, falling and left turning were on close clades. Review of the video clips of these two groups demonstrated that 37.18% of the movements in this group occurred simultaneously with left turning (28.85% for right turning). A similar phenomenon occurred in the clades of diving and sniffing due to the co-occurrence of these behaviors. The linear regression of these two pairs of clades showed that both intra-CC and inter-CC were relatively high (Fig. 5B), suggesting several concomitant descriptions of animal behavior. These clustering results occurred because these movements show more characteristics of the current class.

## Identification of the Behavioral Signatures of the Mouse Disease Model

Animal disease models play an increasingly critical role in expanding understanding of the mechanisms of human diseases and novel therapeutic development [48–50]. Behavioral phenotyping provides a noninvasive approach to the assessment of neuropsychiatric disorders in animal models. By only evaluating spontaneous behavior without any induced conditions, we demonstrate the usability and unbiased character of our framework for animal phenotyping. We collected data from 12 mice (Fig. 6A, $n_{KO}$=6, $n_{WT}$=6) with our 3D motion capture system and subjected them to routine velocity and anxiety index analyses (Fig. 6B-E). In agreement with prior research, we found a significant difference between the average velocities of the two groups.

We clustered the behavioral components of the 12 animals and obtained 41 behavioral phenotypes (Fig. 6F, Supp. Fig. 8,9). Compared with the single-session experiment, the group analysis revealed diverse behavioral types. We found that KO (Shank3B Knock-out, Shank3B$^{-/-}$) mice spent a significantly higher proportion of their time engaging in two of the movements (Fig. 6G). We manually reviewed the video clips of these two types and annotated them as self-grooming and hunching. In previous studies [51–53], self-grooming has been widely reported in Shank3B$^{-/-}$ mice. This is partly attributable to self-grooming being a long-lasting movement (7.13±7.81 s) and thus easily recognized by human observation or software (Fig. 6I). Interestingly, although hunching has only previously been reported in a few related studies [54–56], our framework frequently detected hunching movements in KO mice. This novel finding can be attributed to the duration of a single continuous hunching movement being too short to be noticed (1.47±0.31) as well as to the similarity between the kinematics of hunching and rearing. We proved that these two types of movements belong to distinct behavioral phenotypes. Specifically, during hunching, mice maintain an arcuate

spine angle, while rearing is characterized by a stronger, wider range of necks and head motions (Fig. J-N). This ability to identify short-term and fine behavioral modules is one of the advantages of our framework (Fig. 6H).

Finally, we determined whether the animal types could be identified simply by the proportions of time spent engaging in different behaviors. In fact, when assessing behavior, considering each behavioral component as a dimension may affect the outcome of the animal phenotype. Therefore, we used UMAP to perform dimensionality reduction of the 41-dimensional behavioral proportion data of all movement types. As expected, the two genotypes of animals were well separated in the low-dimensional space (Fig. 6E), even though there were large amounts of baseline movements with no significant difference. We defined these two types as "autistic-like behavior space." Hence, our findings indicate the potential use of our framework to identify disease models automatically.

## Discussion

The current study presents a framework for discovering quantifiable behavioral modules from high-dimensional postural time-series by combining dynamic temporal decomposition and unsupervised clustering. Behavior decomposition adopts a parallel, two-stage approach to extract animal motion features in accordance with the natural structure of animal behavior. We used DTAK to measure the similarity between behavioral modules and applied further low-dimensionality embedding to represent the behavior's underlying feature space. The unsupervised clustering identified behavioral phenotypes from the feature space and helped to automatically assess the behavioral experiment data. In addition, the clustering step could quickly generate large amounts of distinct unlabeled behavior groups. By manually assigning annotations to each group, our framework will potentially facilitate semi-supervised behavior recognition.

Our framework has two main advantages. First, our approach of tracking multiple body parts and acquiring 3D reconstruction data achieves better performance than similar recently reported rodent behavioral recognition frameworks [11,57]. The high signal-to-noise ratio of the data yielded by the present method avoids animal body occlusion and view-angle bias in single-camera top-view monitoring. More importantly, our behavior decomposition framework emphasizes the extraction of the temporal dynamics of movements. Without making model assumptions, similar movements with various time durations and temporal variability can be efficiently represented by the self-similarity matrix. We proved that this similarity matrix is a reliable objective metric by evaluating the consistency of clustered behavior phenotypes. We further performed dimension reduction to visualize the behavioral map, which facilitates exploring the evolution of movement sequences of higher-order behavior and behavioral state transition caused by neural activity. For example, innate defensive behavior is considered to consist of three specific movement phases [40,58], but data supporting this

idea is lacking. Hence, our future work will focus on modeling the transition patterns of innate behavior based on the behavioral map.

Comprehensive and unbiased behavioral phenotyping is becoming a powerful approach to the study of behavioral abnormalities in animal models of neuropsychiatric disorders. In this study, we demonstrate its application to the monitoring of Shank3 mutant mice that show autistic-like behaviors. Our framework helped to reveal that Shank3B$^{-/-}$ engage in two types of spontaneous behaviors significantly more often than WT mice; While grooming has been extensively observed in murine models of restricted, repetitive behavior, short-term hunching behavior has not. Previous studies [51,52] mentioned that the rearing behavior of Shank3 mice also differs from that of WT mice; however, because hunching is kinematically similar to rearing, it is difficult to distinguish these two types by human observation or algorithms. Our 3D and sub-second methods will help to identify new behavioral biomarkers and advance understanding of the neural circuit mechanisms underlying behavioral changes caused by genetic mutations. In addition, large animals such as non-human primates, dogs, and pigs have recently emerged as valuable models for studying neurological dysfunctions [49,50]. Our general-purpose framework further benefits from the significant advantage of being able to capture and analyze large animal movements, which have more complex 3D characteristics and temporal dynamics.

The dynamic, high-dimensional, and multi-scale characteristics of behavior can be attributed to similar properties of the nervous system produces it. While the most advanced large-scale neuroimaging and high spatiotemporal resolution electrophysiological techniques allow researchers to elucidate the details of the firing timing of all neurons and neurofunctional connections at all scales, they cannot inform the mapping of the neural-behavioral relationship without quantifying behavior at the corresponding level. In other words, to understand the precise temporal relationship between neural activity and behavior, neural-behavioral mapping requires the complete characterization of behavior dynamics and how neural activity and behavior sequences co-evolve over time. Further, hierarchically measuring behavior at multiple scales may potentially reveal the mechanisms of hierarchical sensorimotor processing. Our framework and further technical optimization may, therefore, contribute to resolving the relationships between complex neural circuitry and behavior.

## Acknowledgements

We thank Guoqing Bi and Jianyuan Sun for comments on our manuscript; Feng Wang, Zhonghua Lu, Gaoyang Zhao, and Yu Hu providing Shank3B$^{-/-}$ Mice.

## Funding

This work was supported in part by Key Area R&D Program of Guangdong Province (2018B030331001 L.W., 2018B030338001 P.W.), National Key R&D Program of China

(2018YFA0701400 P.W.), National Natural Science Foundation of China (NSFC 31500861 P.W., NSFC 31630031 L.W., NSFC 91732304 L.W., NSFC 31930047 L.W.), the International Big Science Program Cultivating Project of CAS (172644KYS820170004 L.W.), the Strategic Priority Research Program of Chinese Academy of Science (XDB32030100, L.W.), the Youth Innovation Promotion Association of the Chinese Academy of Sciences (2017413 P.W.), Shenzhen Government Basic Research Grants (JCYJ20170411140807570 P.W., JCYJ20170413164535041 L.W.), Science, Technology and Innovation Commission of Shenzhen Municipality (JCYJ20160429185235132 K.H.), Helmholtz-CAS joint research grant (GJHZ1508 L.W.), Guangdong Provincial Key Laboratory of Brain Connectome and Behavior (2017B030301017 L.W.), the Ten Thousand Talent Program (L.W.), the Guangdong Special Support Program (L.W.), Key Laboratory of SIAT (2019DP173024 L.W.), Shenzhen Key Science and Technology Infrastructure Planning Project (ZDKJ20190204002 L.W.).

## Author Contributions

Conceptualization, K.H., Y.H., P.W., and L.W.; Methodology, K.H., Y.H., K.C., P.W., L.W., and S.L.; Data Analysis, K.H., Y.H., and K.C.; Animal Experiments and Data Collection, Y.H., H.P., K.H., K.C., and W.Y.; Manuscript Writing, K.H., Y.H., P.W., L.W., and X.L.; Funding Acquisition, P.W., and L.W.; and Supervision, L.W. and P.W.

## Declaration of Interests

The authors declare no competing interests.

## Code Availability

The code of this framework can be accessed at https://behavioratlas.tech/

## References

1. Krakauer JW, Ghazanfar AA, Gomez-Marin A *et al.* Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron* 2017;**93**:480–90.

2. Datta SR, Anderson DJ, Branson K *et al.* Computational Neuroethology: A Call to Action. *Neuron* 2019;**104**:11–24.

3. Love BC. Cognitive Models as Bridge between Brain and Behavior. *Trends Cogn Sci* 2016;**20**:247–8.

4. Jennings JH, Stuber GD. Tools for Resolving Functional Activity and Connectivity within Intact Neural Circuits. *Curr Biol* 2014;**24**:R41–50.

5. Luo L, Callaway EM, Svoboda K. Genetic Dissection of Neural Circuits: A Decade of Progress. *Neuron* 2018;**98**:256–81.

6. Mathis MW, Mathis A. Deep learning tools for the measurement of animal behavior in neuroscience. *Curr Opin Neurobiol* 2020;**60**:1–11.

7. Datta SR, Anderson DJ, Branson K *et al.* Computational Neuroethology: A Call to Action. *Neuron* 2019;**104**:11–24.

8. Mathis A, Mamidanna P, Cury KM *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 2018;**21**:1281–9.

9. Graving JM, Chae D, Naik H *et al.* Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* 2019;**8**:1–42.

10. Pereira TD, Aldarondo DE, Willmore L *et al.* Fast animal pose estimation using deep neural networks. *Nat Methods* 2019;**16**:117–25.

11. Hsu AI, Yttri EA. B-SOiD: An Open Source Unsupervised Algorithm for Discovery of Spontaneous Behaviors. *bioRxiv* 2019:770271.

12. Datta SR. Q&A: Understanding the composition of behavior. *BMC Biol* 2019;**17**:1–7.

13. Altimus CM, Marlin BJ, Charalambakis NE *et al.* The Next 50 Years of Neuroscience. 2020;**40**:101–6.

14. Berman GJ, Choi DM, Bialek W *et al.* Mapping the stereotyped behaviour of freely moving fruit flies. *J R Soc Interface* 2014;**11**:20140672.

15. Vogelstein JT, Park Y, Ohyama T *et al.* Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science (80- )* 2014;**344**:386–92.

16. Bruni G, Rennekamp AJ, Velenich A *et al.* Zebrafish behavioral profiling identifies multitarget antipsychotic-like compounds. *Nat Chem Biol* 2016;**12**:559–66.

17. Marques JC, Lackner S, Félix R *et al.* Structure of the Zebrafish Locomotor Repertoire Revealed with Unsupervised Behavioral Clustering. *Curr Biol* 2018;**28**:181-195.e5.

18. Sharma A, Johnson RE, Engert F et al. Point Process Latent Variable Models of Larval Zebrafish Behavior., 2018.

19. Brown AEX, Yemini EI, Grundy LJ *et al.* A dictionary of behavioral motifs reveals clusters of genes affecting Caenorhabditis elegans locomotion. *Proc Natl Acad Sci U S A* 2013;**110**:791–6.

20. Yemini E, Jucikas T, Grundy LJ *et al.* A database of Caenorhabditis elegans behavioral phenotypes. *Nat Methods* 2013;**10**:877–9.

21. Liu M, Sharma AK, Shaevitz JW *et al.* Temporal processing and context dependency in caenorhabditis elegans response to mechanosensation. *Elife* 2018;**7**, DOI: 10.7554/eLife.36419.

22. Jordan D, Kuehn S, Katifori E *et al.* Behavioral diversity in microbes and low-dimensional phenotypic spaces. *Proc Natl Acad Sci U S A* 2013;**110**:14018–23.

23. Berman GJ, Bialek W, Shaevitz JW. Predictability and hierarchy in Drosophila behavior. *Proc Natl Acad Sci U S A* 2016;**113**:11943–8.

24. Johnson RE, Linderman S, Panier T *et al.* Probabilistic Models of Larval Zebrafish Behavior Reveal Structure on Many Scales. *Curr Biol* 2020;**30**:70-82.e4.

25. Hoopfer ED, Jung Y, Inagaki HK *et al.* P1 interneurons promote a persistent internal state that enhances inter-male aggression in Drosophila. *Elife* 2015;**4**:1–27.

26. Wang Q, Taliaferro JM, Klibaite U *et al.* The PSI-U1 snRNP interaction regulates male mating behavior in Drosophila. *Proc Natl Acad Sci U S A* 2016;**113**:5269–74.

27. Cande J, Namiki S, Qiu J *et al.* Optogenetic dissection of descending behavioral control in Drosophila. *Elife* 2018;**7**, DOI: 10.7554/eLife.34275.

28. Marquez-Legorreta E, Constantin L, Piber M *et al.* Brain-wide visual habituation networks in wild type and fmr1 zebrafish. *bioRxiv* 2019:722074.

29. Charles JP, Cappellari O, Hutchinson JR. A Dynamic Simulation of Musculoskeletal Function in the Mouse Hindlimb During Trotting Locomotion. *Front Bioeng Biotechnol* 2018;**6**:61.

30. Kabra M, Robie AA, Rivera-Alba M *et al.* JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nat Methods* 2013;**10**:64–7.

31. Ravbar P, Branson K, Simpson JH. An automatic behavior recognition system classifies animal behaviors using movements and their temporal context. *J Neurosci Methods* 2019;**326**:108352.

32. Wang L, Hu W, Tan T. Recent developments in human motion analysis. *Pattern Recognit* 2003;**36**:585–601.

33. Wiltschko AB, Johnson MJ, Iurilli G *et al.* Mapping Sub-Second Structure in Mouse Behavior. *Neuron* 2015;**88**:1121–35.

34. Storchi R, Milosavljevic N, Allen AE *et al.* Beyond locomotion□: in the mouse the mapping between sensations and behaviours unfolds in a higher dimensional space. *bioRxiv* 2020:2020.02.24.961565.

35. Liu D, Li W, Ma C *et al.* A common hub for sleep and motor control in the substantia nigra. *Science (80- )* 2020;**367**:440–5.

36. Jhuang H, Garrote E, Mutch J *et al.* Erratum: Automated home-cage behavioural phenotyping of mice (Nature Communications (2010) 1: 68 DOI:10.1038/ncomms1064). *Nat Commun* 2012;**3**, DOI: 10.1038/ncomms1399.

37. Berman GJ. Measuring behavior across scales. *BMC Biol* 2018;**16**:1–12.

38. Gris K V., Coutu JP, Gris D. Supervised and unsupervised learning technology in the study of rodent behavior. *Front Behav Neurosci* 2017;**11**:1–6.

39. Anderson DJ, Perona P. Toward a science of computational ethology. *Neuron* 2014;**84**:18–31.

40. Evans DA, Stempel AV, Vale R *et al.* Cognitive Control of Escape Behaviour. *Trends Cogn Sci* 2019;**23**:334–48.

41. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 2018.

42. Sheshadri S, Dann B, Hueser T *et al.* 3D reconstruction toolbox for behavior tracked with multiple cameras. *J Open Source Softw* 2020;**5**:1849.

43. Gomez-Marin A, Paton JJ, Kampff AR *et al.* Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nat Neurosci* 2014;**17**:1455–62.

44. Hinde RA, Bateson PPG. *Growing Points Ethology*. CUP Archive, 1976.

45. Zhou F, De La Torre F, Hodgins JK. Aligned cluster analysis for temporal segmentation of human motion. *2008 8th IEEE Int Conf Autom Face Gesture Recognition, FG 2008* 2008, DOI: 10.1109/AFGR.2008.4813468.

46. Zhou F, Torre FD La, Hodgins JK. Hierarchical aligned cluster analysis for temporal clustering of human motion. *IEEE Trans Pattern Anal Mach Intell* 2013;**35**:582–96.

47. Scrucca L, Fop M, Murphy TB et al. Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.

48. Sur I, Taipale J. Dissection of behavior and psychiatric disorders using the mouse as a model. *Nat Rev Cancer* 2016;**9**:953–65.

49. Zhao J, Lai L, Ji W *et al.* Genome editing in large animals: Current status and future prospects. *Natl Sci Rev* 2019;**6**:402–20.

50. Sukoff Rizzo SJ, Crawley JN. Behavioral Phenotyping Assays for Genetic Mouse Models of Neurodevelopmental, Neurodegenerative, and Psychiatric Disorders. *Annu Rev Anim Biosci* 2017;**5**:371–89.

51. Peça J, Feliciano C, Ting JT *et al.* Shank3 mutant mice display autistic-like behaviours and striatal dysfunction. *Nature* 2011;**472**:437–42.

52. Mei Y, Monteiro P, Zhou Y *et al.* Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. *Nature* 2016;**530**:481–4.

53. Peixoto RT, Chantranupong L, Hakim R *et al.* Abnormal Striatal Development Underlies the Early Onset of Behavioral Deficits in Shank3B−/− Mice. *Cell Rep* 2019;**29**:2016-2027.e4.

54. Rousseau JBI, Van Lochem PBA, Gispen WH *et al.* Classification of rat behavior with an image-processing method and a neural network. *Behav Res Methods, Instruments, Comput* 2000;**32**:63–71.

55. Kolevzon A, Delaby E, Berry-Kravis E *et al.* Neuropsychiatric decompensation in adolescents and adults with Phelan-McDermid syndrome: A systematic review of the literature. *Mol Autism* 2019;**10**:50.

56. Orefice LL, Mosko JR, Morency DT *et al.* Targeting Peripheral Somatosensory Neurons to Improve Tactile-Related Phenotypes in ASD Models. *Cell* 2019;**178**:867-886.e24.

57. Luxem K, Fuhrmann F, Kürsch J *et al.* Identifying Behavioral Structure from Deep Variational Embeddings of Animal Motion. *bioRxiv* 2020:2020.05.14.095430.

58. Wei P, Liu N, Zhang Z *et al.* Processing of visually evoked innate fear by a non-canonical thalamic pathway. *Nat Commun* 2015;**6**:1–13.

## Figure Legends

**Figure 1 | Spatio-temporal decomposition framework for animal behavior analysis. A.** Data preparation: 1) image streams captured from four cameras with different 2D views; 2) animal body parts are tracked to generates separate 2D skeletal trajectories (color-coded traces); 3) reconstructing 3D body skeleton by integrating these four data streams. **B**. Two-stage NM decomposition to generate the feature space: 1) pose decomposition groups continuous skeleton postural data into discrete postural sequences; 2) NM decomposition (two high-lighted [green and orange] blocks represent two NMs decomposed from the postural sequences; 3) NM sequences mapped to their 2D features space (right), where each dot on the 3D axis corresponds to the NM block on the left. **C**. Calculation of locomotion dimension. The continuous velocity of the behaving animal is first calculated, then average the velocity of each segment obtained in the NM decomposition step. **D**. 3D scatter plot represents the combined NM and locomotion feature space. All the movements are clustered into three types (red, green, and orange dots) with the unsupervised approach.

**Figure 2 | Collecting animal behavior trajectories via a 3D motion capture system. A**. Pipeline of 3D animal skeletal reconstruction. **B.** center, schematic diagram of recording animal behavior with four synchronized cameras; corners, frames captured by the cameras with the DLC labels (left) and the corresponding reconstructed skeletons (right). **C**. Left: 16 key body parts include the nose, left ear, right ear, neck, left front limb, right front limb, left hind limb, right hind limb, left front claw, right front claw, left hind claw, right hind claw, back, root tail, middle tail, and tip tail. Right: representative mouse body tracking trace data collected over 100 s showing 48 data vectors obtained by DLC for each body part (indicated with a color-coded dot) encoded by x, y, and z coordinates. For visualization purposes, mean normalization is applied to each trace. **D**. 3D reconstruction quality assessment: 1-best quality, 0-worst quality. The quality of the data obtained from the 12 mice averaged at 0.9981±0.001.

**Figure 3 | Dynamic temporal decomposition of multi-scale hierarchical behavior. A**. Illustration of the three-layer bottom-up architecture for behavior. Top: The color-coded bars indicate the types of

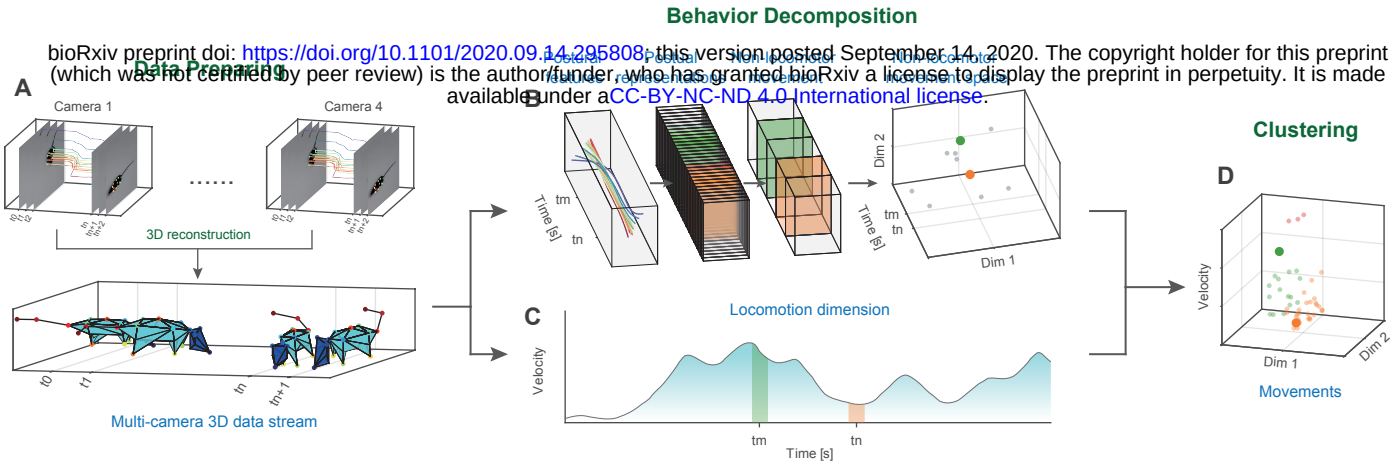
behavior components in the corresponding time period at that layer; each upper layer component is composed of the sequence of the lower layer. The instance of "approaching" is at the ethogram level which is composed of three movement level sequences, and each movement sequence includes a set of postural representations. **B**. Representative animal postural trajectories (black traces) with two selected similar NM segments *S1* and *S2* (orange bars masked). **C**. Discrete postural sequences *S'1* (12 points) and *S'2* (13 points) were decomposed from *S1* and *S2* and used to calculate their similarity kernel matrix *K*. **D.** Segment kernel matrix *T* calculated with DTAK. Each pixel on the matrix represents the normalized similarity value of the *K* for a pair of segments at the $i^{th}$ row and the $j^{th}$ column (e.g., the pixel in the black box indicates the final similarity of *S1* and *S2*). **E.** NM segments decomposed from the postural trajectories shown in B and their color-coded labels. Segments with the same color indicate that they belong to the same types due to their higher similarity. **F.** Optimization process of dynamic temporal decomposition. Objective Value (OV) error decreases with each iteration until the termination condition is reached (maximum number of iterations or OV converges). **G.** Top, representative 300-s skeletal traces, where the trace slices highlighted in colors corresponding to the four types of typical NMs (left turn, immobile, walk, right turn). Bottom, magnification of representative traces of these four movement types. **H.** Workflow of the two-stage behavioral decomposition.

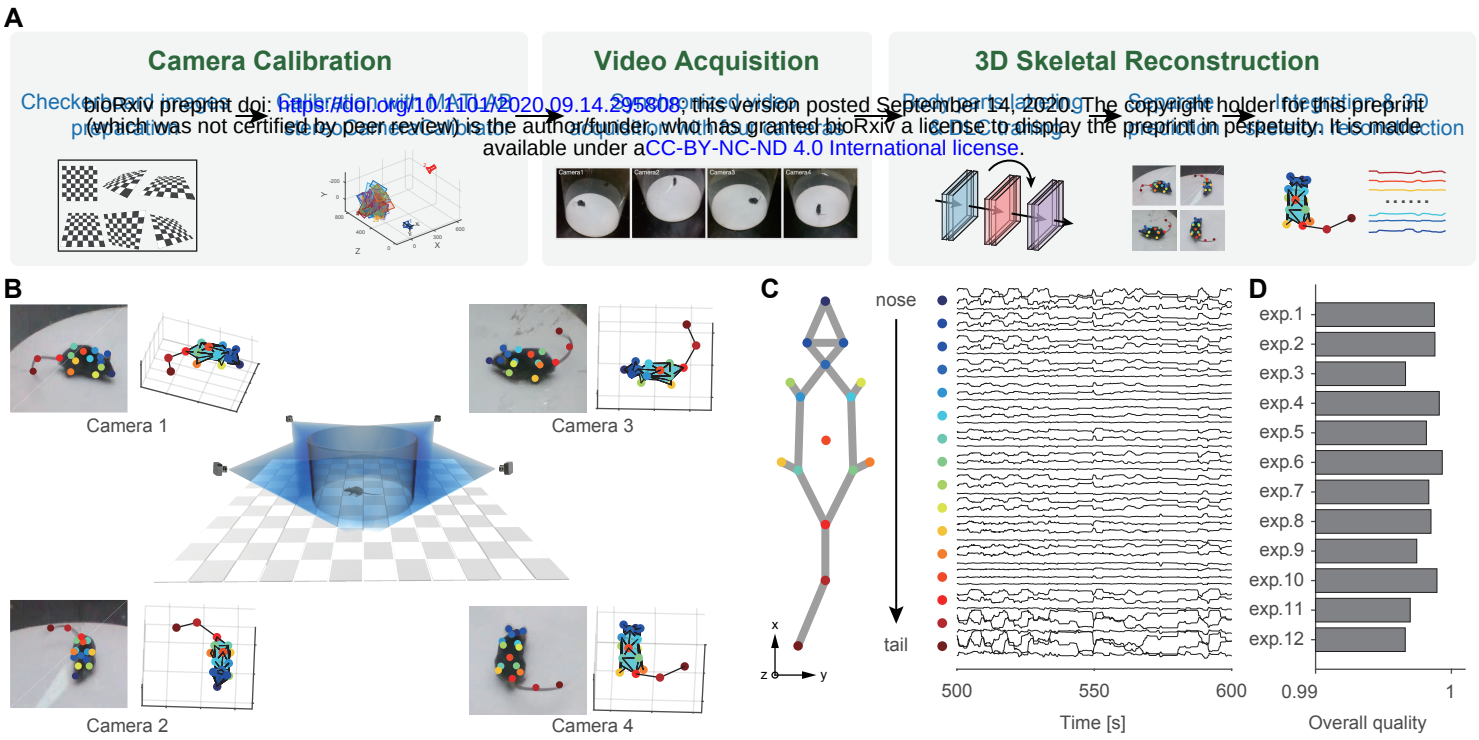
**Figure 4 | Identify movement phenotypes on single experimental data. A.** Spatio-temporal feature space of behavioral components. Each dot on the 3D scatter plot represents a movement bout (n = 935 bouts). The 11 different colors indicate the corresponding to 11 movement types. **B.** Upper, recalculated paired-wise similarity matrix, and they were rearranged with a dendrogram (lower). Each pixel on the matrix represents the normalized similarity value of a pair of movement bouts at the $i^{th}$ row and the $j^{th}$ column. The color-coded bars indicate the labels of clustered movement (middle). **C.** Fractions of movement bouts number. **D.** Intra-CC (color-coded) and inter-CC (grey dots) of each movement group. The dots on each violin plot represents their intra-CC or inter-CC, and dots number in a pair of violin plot in each group are the same (Intra-CC: 0.91±0.07; Inter-CC: 0.29±0.19). **E.** Cumulative Distribution Function of CQI of the movement clusters. The clusters represented by the curves on the right side have better clustering qualities, and their corresponding movements are more stereotyped. **F.** The histogram of the duration of all movements (0.963±0.497s).

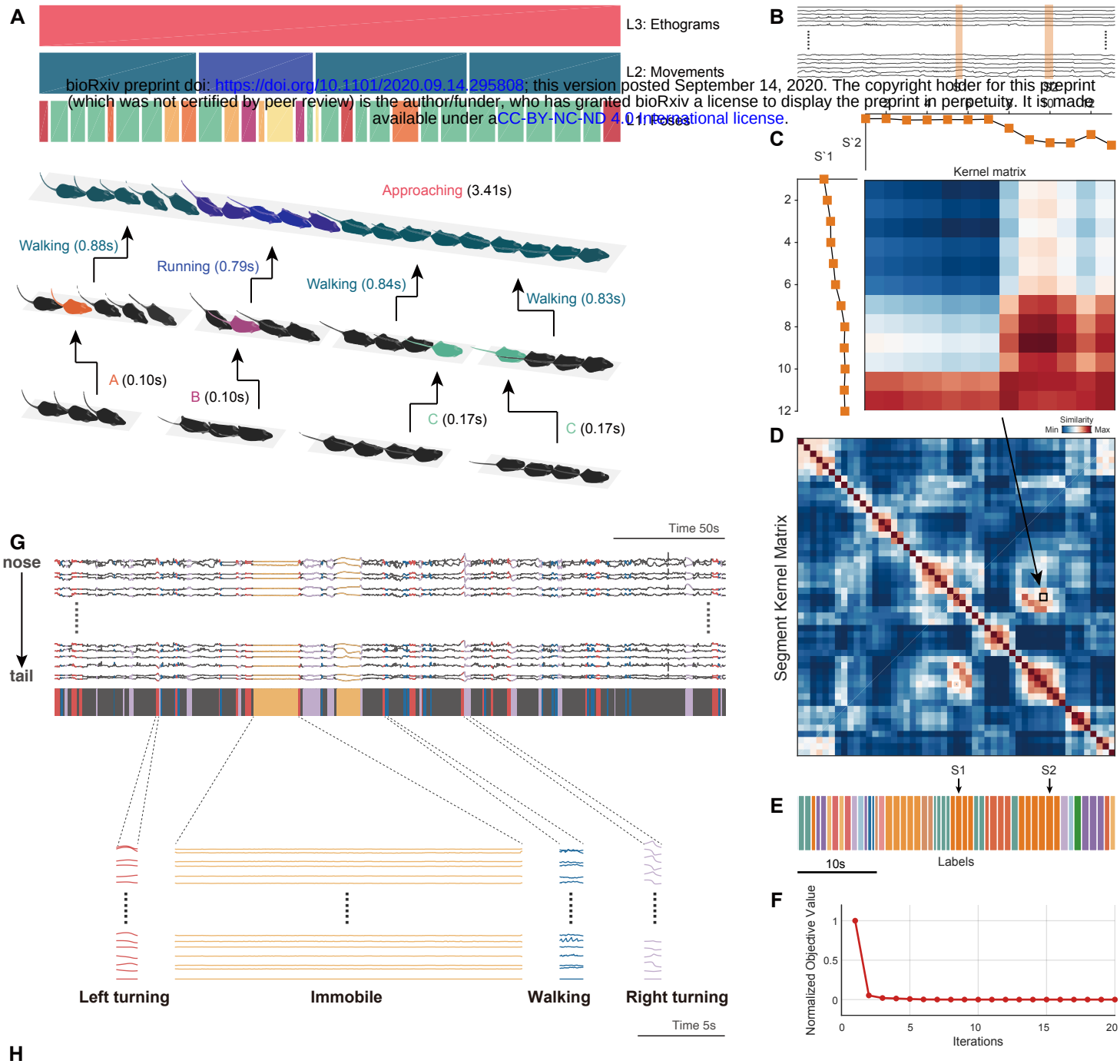
**Figure 5 | Visualization and quantification of behavioral kinematics. A.** Average-skeleton of all frames within each movement phenotype. **B.** Linear regression plot of movement phenotypes. The horizontal-axis represents the target, and the vertical-axis represents the reference. The color-coded and gray dots correspond to the intra- and inter-cluster correlation coefficients, respectively. **C.** Box

plot of normalized MI of movement groups. Red boxes: horizontal MI; Blue boxes: vertical MI. **D, E.** Horizontal and vertical MI of each body part.

**Figure 6 | Spontaneous behavior analysis reveals autistic-like behaviors on *shank3B* knock-out mice. A.** PCR genotyping for Shank3B$^{+/+}$ (Wild Type, WT), Shank3B$^{-/-}$ (Shank3B KO) mice. **B-E.** Box plot of mean velocity (upper left), mean anxiety index (upper right), maximum velocity (lower left), and locomotion (lower right) of the two groups of animals (red: KO, n=6, blue: WT, n=6; Statistics: unpaired T-test for mean velocity, Mann-Whitney test for mean anxiety index; ***P<0.001), values are represented as mean±std. **F.** Upper: recalculated paired-wise similarity matrix. The movement bouts of all of the 12 involved mice were grouped (n = 9495) and were rearranged with dendrogram (lower). Each pixel on the matrix represents the normalized similarity value of a pair of movement bouts at the $i^{th}$ row and the $j^{th}$ column. The color-coded bars (41 clusters) indicate the movements being clustered (middle); **G.** Comparison of the proportion of movement types between KO mice and WT mice. The bold traces and shadows indicate the mean±sem. Fractions of each group and light color traces are the fractions of all 12 mice (red: KO, n=6, blue: WT, n=6). Middle color-coded labels and dendrogram correspond to B. Two movements differed significantly between the two groups: hunching: KO 5.56±3.84%, WT 1.07±0.68%; self-grooming: KO 3.16±1.12%, WT 2.31±1.15%. ****P<0001, **P<0.01 by two-way ANOVA with Holm–Sidak post-hoc test. **H**. Low-dimensional representation of the two animal groups (red: KO, n=6, blue: WT, n=6). The 12 dots in 3D space were dimensionally reduced from 41-dimensional movement fractions, and they are well separated. **I.** Ethograms of the two significant movements (orange: self-grooming, green: hunching). **J-N.** Kinematic comparison of rearing and hunching (upper row refers rearing; lower row refers to hunching). **J.** Average-skeletons of all frames and normalized activity intensity (side view) of rearing and hunching. **K.** Spine lines (the lines connecting the neck, back, and tail root) extracted from all frames (rearing: 9854 frames, hunching: 15359 frames) in movement types. For visualization purposes, only 1% of spine lines are shown in the figure (rearing: 985/9854, hunching: 1535/15359). Black lines refer to the averaged spine line of the hunching and rearing; **L.** Histograms of the spine angles (angle between three body parts). During rearing, the spine angles of the animals swing, and the average spine angle is straight (181.49±15.48°). By contrast, the spine angles of the rodents during hunching are consistently arcuate (168.74±11.19°). **M.** Box plot of spine angles of the two movement types. **N.** Box plot of normalized MI of the three body parts involved.








H