A Hierarchical 3D-motion Learning Framework for Animal Spontaneous Behavior Mapping

Kang Huang[1,2,4], Yaning Han[1,2,4], Ke Chen[1,2], Hongli Pan[1], Gaoyang Zhao[1,2], Wenling Yi[1,2], Xiaoxi Li[1,2], Siyuan Liu[3], Liping Wang[1,2,5,*], Pengfei Wei[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. Pennsylvania State University, State College, Pennsylvania, 16802, USA.
4. Co-first authors
5. Lead Contact

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

Abstract:

Animal behavior usually has a hierarchical structure and dynamics. Therefore, to understand how the neural system coordinates with behaviors, neuroscientists need a quantitative description of the hierarchical dynamics of different behaviors. However, the recent end-to-end machine-learning-based methods for behavior analysis mostly focus on recognizing behavioral identities on a static timescale or based on limited observations. These approaches usually lose rich dynamic information on cross-scale behaviors. Inspired by the natural structure of animal behaviors, we addressed this challenge by proposing a novel parallel and multi-layered framework to learn the hierarchical dynamics and generate an objective metric to map the behavior into the feature space. In addition, we characterized the animal 3D kinematics with our low-cost and efficient multi-view 3D animal motion-capture system. Finally, we demonstrated that this framework could monitor spontaneous behavior and automatically identify the behavioral phenotypes of the transgenic animal disease model. The extensive experiment results suggest that our framework has a wide range of applications, including animal disease model phenotyping and the relationships modeling between the neural circuits and behavior.

Key Words: Behavioral structure-inspired; 3D motion capture; Behavioral dynamics; Computational ethology; Behavior phenotyping.

Introduction:

The structure of animal behavior follows a bottom-up hierarchy constructed by time-varying posture dynamics, which has been demonstrated to be classical in ethological theory [1,2] and recent animal studies [3–6]. Such behavioral organization is considered to coordinate with neural activities [7,8]. Previous studies [9–11] using large-scale neuronal recordings have provided preliminary evidence from the neural implementation perspective. As the central goal of modern neuroscience, fully decoding this cross-scale dynamic relationship requires comprehensive quantification of neural activity and behavior. Over the past few decades, scientists have been working on improving the accuracy and throughput of neural dynamics manipulation and capturing. Meanwhile, for behavior quantification, there has been a revolution from simple behavioral parameters extraction to machine-learning (ML)-based behavior sequence recognition [12,13]. However, most previous methods [14,15] often emphasized feature engineering and pattern recognition for mapping raw data to behavioral identities. These black-box approaches lack the interpretability of cross-scale behavioral dynamics. Thus, it is a challenging task, but with a strong demand, to develop a general-purpose framework for the dynamic decomposition of animal spontaneous behavior.

Previous researchers addressed this challenge mainly from two aspects. The first aspect is behavioral feature capturing. Conventional animal behavior experiments usually use a single camera top-view recording to capture the motion signal of behaving animals, leading to occlusions of the key body parts (e.g., paws), and these are very sensitive to viewpoint differences [16]. The recent emergence of ML toolboxes [17–19] has dramatically facilitated the animal pose estimation with multiple body parts. Thus, it enables us to study the animal kinematics more comprehensively and provides potential applications for capturing 3D animal movements. The second aspect is decomposing continuous time-series data into understandable behavioral modules. Previous studies on lower animals such as flies [10,20–22], zebrafishes [4,23–25] and C. elegans [26–28] utilized ML strategies and multivariate analysis to detect action sequences. However, mammalian behavior is highly complicated. Besides locomotion, animals demonstrate non-locomotor movement (NM) with their limbs (e.g., grooming, rearing, turning), and their organs have high-dimensional [29–31] and variable spatio-temporal characteristics. Even for similar behaviors, the duration and composition of postural sequences vary. To define the start and end boundaries to segment continuous data into behavioral sequences, many ML-based open-source toolboxes [21] and commercial software do excellent work in feature engineering. They usually 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 [14,15]. Although these approaches effectively identify specific behaviors, behavior recognition becomes problematic when the dynamics of particular behaviors cannot be represented by window features.

The present study proposes a hierarchical 3D-motion learning framework to address our contribution to these challenges. First, we acquired the 3D markerless animal skeleton with tens of body

parts by the developed flexible and low-cost system. Through the systematic validations, we proved that our system can solve the critical challenges of body occlusion and view disappearance in animal behavior experiments. Second, aiming at the parallel and hierarchical dynamic properties of spontaneous behavior, we were the first to propose a decomposition strategy preserving the behavior's natural structure. With this strategy, the high-dimensional, time-varying and continuous behavioral series can be represented as various quantifiable movement parameters and low-dimensional behavior map. Third, we obtained a large sample of the *Shank3B*$^{-/-}$ mouse disease model data resources with our efficient framework. The results showed that our framework could detect behavioral biomarkers that have been identified previously and discover potential new behavioral biomarkers. Finally, together with the further group analysis of the behavioral monitoring under different experimental apparatus, lighting conditions, ages, and sexes, we demonstrated our framework could contribute to the hierarchical behavior analysis, including postural kinematics characterization, movement phenotyping, and group level behavioral patterns profiling.

Results:

Framework of Hierarchical 3D-motion Learning

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, and they inform the natural characteristics of locomotion and NM. 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 degrees of freedom [32]. 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 was applied to the centralized animal skeleton postural data to obtain the NM space. Finally, together with the additional velocity-based locomotion dimension, unsupervised clustering was used to reveal the structure of the rodent's behavior.

Our framework has two main advantages. First, it addresses the multi-timescale of animal behavior [33]. Animal behavior is self-organized into a multi-scale hierarchical structure from the bottom up, including poses, movements, and ethograms [34,35]. The poses and movements are low- and intermediate-level elements [36], while higher-level ethograms are stereotyped patterns composed of movements that adhere to inherent transfer rules in certain semantic environments [37]. 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) [38] 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).

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

To efficiently and comprehensively characterize the kinematics of free-moving animals, we developed a 3D multi-view motion capture system (Fig. 2a, b) based on recent advances in techniques for pose estimation[17] and 3D skeletal reconstruction [39]. The most critical issues in 3D animal motion capture are efficient camera calibration, body occlusion, and viewpoint disappearance, which have not been optimized or verified [12]. To address these issues, we developed a multi-view video capture device (Supplementary Fig. 2a). This device integrates the behavioral apparatus, an auto-calibration module (Supplementary Fig. 2b, d), and synchronous acquisition of multi-view video streams (Supplementary Fig. 2c). While the conventional manual method requires half an hour to produce the required checkerboard for calibration, the auto-calibration module can be completed in one minute.

We collected the naturalistic behavioral data of free-moving mice in a featureless circular open-field (Supplementary Fig. 2a, and Supplementary 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. 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 timescale. We next investigated whether the 3D motion capture system could reliably track the animal in cases of body-part occlusion and viewpoint disappearance. We checked the DeepLabCut (DLC) tracking likelihood in the collated videos (0.9807 ± 0.1224, Supplementary Fig. 4a) and evaluated the error between the estimated 2D body parts of every training set frame and the ground truth (0.534 ± 0.005%, Supplementary Fig. 5b). These results indicated that in most cases, four cameras were available for 2D pose tracking. Since 3D reconstruction can be achieved as long as any two cameras obtain the 2D coordinates of the same point in 3D space from different views, the reconstruction failure rate caused by body occlusion and viewpoint disappearances is determined by the number of available cameras. Therefore, we evaluated the average proportion of available cameras in situations of body part occlusion and viewpoint disappearance. The validation results for body-part occlusion show an average reconstruction failure rate of only 0.042% due to body occlusion or inaccurate body-part estimation (Supplementary Fig. 5c). While for viewpoint disappearances, both tests (Supplementary Fig. 6, and Supplementary Video 4, 5) proved that our system has a high reconstruction rate for animal body parts. Moreover, the artifact detection and correction features can

recover the body parts that failed to be reconstructed. We calculated an overall reconstruction quality (0.9981 ± 0.0010, Fig. 2d) 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) [34,35], and research has focused on elucidating behavioral component sequences contained in stimuli-related ethograms [40]. 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 (Supplementary Fig. 7). Animal movement is continuous, and due to the high dimensionality of the mammalian skeleton, the behaviorally relevant posture variables are potentially infinite in number [12]. However, adjacent poses are usually highly correlated and redundant for behavior quantification and analysis [1], which is particularly evident in long-term recording. Therefore, for computational efficiency, we adopted a temporal reduction algorithm to merge 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 [41]: 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) [42]. 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. An appropriate distance metric is critical for modeling the temporal variability and optimizing the NM segmentation of a continuous postural time-varying series. Although dynamic time warping (DTW) has commonly been applied in aligning time-series data, it does not satisfy the triangle inequality [43]. Thus, we used the improved DTAK method 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 timescale 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 dimensionality reduction (DR).

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 (Supplementary 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. However, since the 936-D matrix cannot provide an informative visualization of behavioral structure, it is necessary to perform DR on this data. Various DR algorithms have been designed either to preserve the global representation of original data or to focus on local neighborhoods for recognition or clustering [44,45]. Thus, in animal behavior quantification, we face a trade-off between discretizing behavior to provide a more quantitative analysis and maintaining a global representation of behavior to characterize the potential manifolds of neural-behavioral relationships [46]. Therefore, we first evaluated the commonly used DR algorithms from the standpoints of preserving either the global or the local structure. The evaluation results show that UMAP can balance both aspects for our data (Supplementary Fig. 8) 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 investigate the behavior's spatio-temporal representation and identify the movement phenotypes. Most unsupervised clustering require a pre-specified number of clusters, and the number chosen can be data-driven or refer to the context of the practical biological problem [47]. In the single experimental data shown in Figure 4a, the data-driven Bayesian Information Criterion [48] in the R package *mclust* was adopted to determine that the optimal cluster number was 11 (Supplementary Fig. 10). 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 ± 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 (Supplementary Table 1). 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 [20,31,49]. 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 [13], 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 (e.g., running, rearing, sniffing, turning) identified with this framework were homogeneous. Manually inspecting the position, moving, bending, and other characteristics of the mouse limbs and trunk in the video clips of each phenotype group (Supplementary Video 3), we found reliable homogeneity for clips with high CQIs (CQI>0.75). To provide a kinematic validation of the identified behavioral phenotypes from the perspectives of visualization and quantification, we first visualized the average skeleton, which was averaged over all frames in 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. The detailed kinematic parameters, especially the velocity of each body part, could provide greater sensitive differences than the unclear visually-based assessments [50]. Therefore, we defined movement intensity (MI) as a metric for characterizing the kinematics of each body part in

each behavioral phenotype (see Supplementary Methods for further details). MI is related to velocity, and it contains both horizontal and vertical components. 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 details; we attribute their high overall vertical MI to the movement of the nose and front claws (Fig. 5a, c-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 dendrogram of the movements (Fig. 5a) 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 correlation and linear regression analysis 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 [51–53]. 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 20 mice (Fig. 6a, $n_{KO} = 10$, $n_{WT} = 10$) 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 20 animals and obtained 41 behavioral phenotypes (Fig. 6f, Supplementary Fig. 10). Compared with the single-session experiment, the group analysis revealed diverse behavioral types. We found that *Shank3B* Knock-out (KO, *Shank3B*$^{-/-}$) mice spent a significantly higher proportion of their time engaging in four of the movements (Fig. 6g, Supplementary Table 3). By manually reviewing the video clips of these four types, we annotated the 38th movement (M38 in Fig. 6g) as hunching; we also found that three of the movements were very similar (closely arranged on the behavioral dendrogram, Fig. 6g). Therefore, we grouped them and annotated them as self-grooming. In previous studies [54–56], self-grooming has been widely reported in *Shank3B*$^{-/-}$ mice. This is partly attributable to self-grooming being a long-lasting movement (4.48 ± 7.84 s, mean ± standard deviation [SD]) and thus easily recognized by human observation or software (Fig. 6i). Interestingly, although hunching has only previously been reported in a few related studies [57–59],

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 (1.29 ± 1.00 s, mean ± SD) as well as to the similarity between the kinematics of hunching and rearing (M31). 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. Besides the four phenotypes that KO mice preferred more than the WT mice did, the KO mice also showed four additional deficit behavioral phenotypes, namely stepping (M5), walking (M14), and two types of rising (M21 and M22). This result indicates that the locomotion intensity and vertical movement of KO mice were lower than those of WT mice. The locomotion result is consistent with the average velocity comparison shown in Fig. 6b.

Finally, we demonstrated that by modeling the time spent of multi-behavioral parameters, our framework could identify the animal types. We used UMAP to perform DR 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. 6h), even though there were large amounts of baseline movements with no significant difference. We defined these two types as "autistic-like behavior space." Recent reviews suggest that most previous methods [60,61], which usually only consider a few behavioral parameters and may lose critical insights, have been challenged in the animal phenotypes' identification. Hence, these findings indicate the potential advantages of our framework to automatically identify disease models.

Discussion

Inspired by the natural structure of animal behavior, 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 [14,62]. The multi-view motion capture system can avoid animal body occlusion and view-angle bias and estimate the pose optimally by flexibly selecting the view to use

according to the tracking reliabilities of the different views. We also confirmed the necessity of using multi-view cameras in complex experimental scenes, whereas in the simple experimental scenes, only three or even two cameras were needed (Supplementary Fig. 4). 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 DR 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, to study animal circadian rhythms, previous researchers have used behavioral recording approaches to characterize different brain states [63–65]. We used our framework to perform a continuous 24-hour behavioral recording, and the preliminary analysis proved that our framework could provide more comprehensive behavioral parameters and detailed quantification of behavior states (Supplementary Fig. 13). In addition, innate defensive behavior is considered to consist of three specific movement phases [37,66], 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 *Shank3B* mutant mice that show autistic-like behaviors. Our framework helped to reveal that *Shank3B*$^{-/-}$ engage in eight 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 [54,55] mentioned that the rearing behavior of *Shank3B* KO 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. Moreover, we further investigated the differences in the behavior patterns of *Shank3B* KO and WT mice at the group level. In addition to the data that had already been analyzed (collected under the condition: male mice, 5–6 weeks, white light, and circular open-field), we extended the group behavioral pattern analysis to include data collected under different conditions (i.e., different experimental apparatus, lighting, age, and sex; Supplementary Table 2). We calculated the cross-correlation coefficient matrix (CCCM) of all samples based on the movement fractions and used principal component analysis to extract the main variance factors of the CCCM (Supplementary Fig. 12 a, b). We found that when only a single condition was changed for male mice, there was no significant difference in population behavior patterns in mice with the same genotype (Supplementary Fig. 12 c). We also found that although some female KO mice had a weak tendency for autistic-like behavior, there was no significant difference between 5–6 week male and female KO mice at the group

level when tested under the white-light circular open field condition (Supplementary Fig. 12 c, d). Finally, we compared the behavior patterns when all conditions were the same except for the genotypes. The results showed that only the female group showed no significant difference between KO and WT genotypes, while significant differences in behavioral patterns were found between KO and WT male mice under all other conditions. These findings are consistent with previous reports that *Shank3B* KO male mice display more severe impairments than females do in motor coordination [67]. Accordingly, the behavior phenotyping on mouse disease model can be generalized to large animals such as non-human primates, dogs, and pigs which recently emerged as valuable models for studying neurological dysfunctions [52,53]. 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 encoding/decoding relationship rules of the neural activity generating behavior and behavior's neural representation, synchronization of large population activities and accurate measurement and identification of naturalistic, complex behavior are required. In the future, we will focus on combining our framework with free-moving two-photon microscopy and electrophysiological recording to link the neural activity patterns and functional brain connections with the cross-scale behavioral dynamics and timing patterns. Therefore, with further technical optimization and the open-source of a large sample, well-annotated disease model behavior database open source, our framework may contribute to resolving the relationships between complex neural circuitry and behavior, as well as to revealing the mechanisms of sensorimotor processing.

Lastly, we would like to discuss the limitations of our framework. When extending our framework to social behavior analysis, such as the analysis of mating, social hierarchy, predation, and defense behaviors, it is challenging to track multiple, visually indistinguishable (markerless) animals without identity-swapping errors (Supplementary Video 6, 7). Alternative methods mainly focus on tracking and identifying social behaviors at the population level, which only requires the identification of features unrelated to the animals' identities such as the positional differences between animals' body parts [68]. However, this approach is limited to specific behaviors and does not apply to interaction behaviors between social subjects of unequal status. Recent cutting-edge toolboxes such as DLC for multi-animal pose estimation [17], SLEAP [69], and AlphaTracker [70] have addressed the multi-animal tracking problem, but once animals with similar appearances are touching or even body-occluded, the inaccurate pose estimation of these toolboxes leads to off-tracking and identity-swapping errors. This is because when estimating multiple body parts of several animals in a single frame, the combination

of the poses of these animals is more complex and diverse, and identity-swapping in different views may happen at different times. Our 3D multi-view motion capture system promises to solve this problem by effectively reducing body-occlusion probability. As a next step, we are considering using computer vision technology (e.g., point cloud reconstruction) to fuse images from multiple views, then segment each animal's body, and estimate the body parts based on the reconstructed 3D animal. Solving these problems will extend the applicability of our framework to the benefit of the animal behavioral research community.

Acknowledgements

We thank Jianyuan Sun, Guoqiang Bi, Yu-Ting Tseng and Yunke Wang for comments on our manuscript; Feng Wang, Zhonghua Lu, and Yu Hu providing *Shank3B*$^{-/-}$ Mice.

Funding

This work was supported in part by Key Area R&D Program of Guangdong Province (2018B030338001 P.W., 2018B030331001 L.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., G.Z., K.C., and W.Y.; Manuscript Writing, K.H., Y.H., P.W., S.L., L.W., S.L., 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.

Data Availability

The data supporting the findings of this study are available within the article and its Supplementary information. Any other relevant data are available upon reasonable request. Source data are provided with this paper.

Code Availability

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

References

1. Hinde, R. A. & Bateson, P. P. G. *Growing Points Ethology*. (CUP Archive, 1976).
2. Tinbergen, N. *The study of instinct*. (Pygmalion Press, an imprint of Plunkett Lake Press, 2020).
3. Wiltschko, A. B. *et al.* Mapping Sub-Second Structure in Mouse Behavior. *Neuron* **88**, 1121–1135 (2015).
4. Marques, J. C., Lackner, S., Félix, R. & Orger, M. B. Structure of the Zebrafish Locomotor Repertoire Revealed with Unsupervised Behavioral Clustering. *Current Biology* **28**, 181-195.e5 (2018).
5. Berman, G. J., Bialek, W. & Shaevitz, J. W. Predictability and hierarchy in Drosophila behavior. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 11943–11948 (2016).
6. Johnson, R. E. *et al.* Probabilistic Models of Larval Zebrafish Behavior Reveal Structure on Many Scales. *Current Biology* **30**, 70-82.e4 (2020).
7. Hollon, N. G. & Jin, X. Neural Implementation of Behavioral Hierarchy. *Neuron* **105**, 402–404 (2020).
8. Kaplan, H. S. & Zimmer, M. Brain-wide representations of ongoing behavior: a universal principle? *Current Opinion in Neurobiology* **64**, 60–69 (2020).
9. Kaplan, H. S., Salazar Thula, O., Khoss, N. & Zimmer, M. Nested Neuronal Dynamics Orchestrate a Behavioral Hierarchy across Timescales. *Neuron* **105**, 562-576.e9 (2020).
10. Vogelstein, J. T. *et al.* Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science* **344**, 386–392 (2014).
11. Markowitz, J. E. *et al.* The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell* **174**, 44-58.e17 (2018).

12. von Ziegler, L., Sturman, O. & Bohacek, J. Big behavior: challenges and opportunities in a new era of deep behavior profiling. *Neuropsychopharmacology* 1–12 (2020) doi:10.1038/s41386-020-0751-7.
13. Datta, S. R., Anderson, D. J., Branson, K., Perona, P. & Leifer, A. Computational Neuroethology: A Call to Action. *Neuron* **104**, 11–24 (2019).
14. Hsu, A. I. & Yttri, E. A. B-SOiD: An Open Source Unsupervised Algorithm for Discovery of Spontaneous Behaviors. *bioRxiv* 770271 (2019) doi:10.1101/770271.
15. Ravbar, P., Branson, K. & Simpson, J. H. An automatic behavior recognition system classifies animal behaviors using movements and their temporal context. *Journal of Neuroscience Methods* **326**, 108352 (2019).
16. Günel, S. *et al.* Deepfly3D, a deep learning-based approach for 3D limb and appendage tracking in tethered, adult Drosophila. *eLife* **8**, 1–22 (2019).
17. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* **21**, 1281–1289 (2018).
18. Graving, J. M. *et al.* Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* **8**, 1–42 (2019).
19. Pereira, T. D. *et al.* Fast animal pose estimation using deep neural networks. *Nature Methods* **16**, 117–125 (2019).
20. Berman, G. J., Choi, D. M., Bialek, W. & Shaevitz, J. W. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of the Royal Society Interface* **11**, 20140672 (2014).
21. Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S. & Branson, K. JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nature Methods* **10**, 64–67 (2013).
22. Branson, K., Robie, A. A., Bender, J., Perona, P. & Dickinson, M. H. High-throughput ethomics in large groups of Drosophila. *Nature Methods* **6**, 451–457 (2009).
23. Sharma, A., Johnson, R. E., Engert, F. & Linderman, S. W. *Point process latent variable models of larval zebrafish behavior*. (2018).
24. Haesemeyer, M., Robson, D. N., Li, J. M., Schier, A. F. & Engert, F. The Structure and Timescales of Heat Perception in Larval Zebrafish. *Cell Systems* **1**, 338–348 (2015).
25. Bruni, G. *et al.* Zebrafish behavioral profiling identifies multitarget antipsychotic-like compounds. *Nature Chemical Biology* **12**, 559–566 (2016).
26. Brown, A. E. X., Yemini, E. I., Grundy, L. J., Jucikas, T. & Schafer, W. R. A dictionary of behavioral motifs reveals clusters of genes affecting Caenorhabditis elegans locomotion. *Proceedings of the National Academy of Sciences* **110**, 791–796 (2013).
27. Yemini, E., Jucikas, T., Grundy, L. J., Brown, A. E. X. & Schafer, W. R. A database of Caenorhabditis elegans behavioral phenotypes. *Nature Methods* **10**, 877–879 (2013).

28. Liu, M., Sharma, A. K., Shaevitz, J. W. & Leifer, A. M. Temporal processing and context dependency in caenorhabditis elegans response to mechanosensation. *eLife* **7**, (2018).
29. Storchi, R., Milosavljevic, N., Allen, A. E., Cootes, T. F. & Lucas, R. J. Beyond locomotion : in the mouse the mapping between sensations and behaviours unfolds in a higher dimensional space. *bioRxiv* 2020.02.24.961565 (2020) doi:10.1101/2020.02.24.961565.
30. Liu, D. *et al.* A common hub for sleep and motor control in the substantia nigra. *Science* **367**, 440–445 (2020).
31. Jhuang, H. *et al.* Automated home-cage behavioural phenotyping of mice. *Nature Communications* **3**, (2012).
32. Charles, J. P., Cappellari, O. & Hutchinson, J. R. A Dynamic Simulation of Musculoskeletal Function in the Mouse Hindlimb During Trotting Locomotion. *Frontiers in Bioengineering and Biotechnology* **6**, 61 (2018).
33. Berman, G. J. Measuring behavior across scales. *BMC Biology* **16**, 1–12 (2018).
34. Gris, K. V., Coutu, J. P. & Gris, D. Supervised and unsupervised learning technology in the study of rodent behavior. *Frontiers in Behavioral Neuroscience* **11**, 1–6 (2017).
35. Anderson, D. J. & Perona, P. Toward a science of computational ethology. *Neuron* **84**, 18–31 (2014).
36. Wang, L., Hu, W. & Tan, T. Recent developments in human motion analysis. *Pattern Recognition* **36**, 585–601 (2003).
37. Evans, D. A., Stempel, A. V., Vale, R. & Branco, T. Cognitive Control of Escape Behaviour. *Trends in Cognitive Sciences* **23**, 334–348 (2019).
38. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. (2018).
39. Sheshadri, S., Dann, B., Hueser, T. & Scherberger, H. 3D reconstruction toolbox for behavior tracked with multiple cameras. *Journal of Open Source Software* **5**, 1849 (2020).
40. Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M. & Mainen, Z. F. Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience* **17**, 1455–1462 (2014).
41. Zhou, F., De La Torre, F. & Hodgins, J. K. Aligned cluster analysis for temporal segmentation of human motion. *2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008* (2008) doi:10.1109/AFGR.2008.4813468.
42. Zhou, F., Torre, F. D. La & Hodgins, J. K. Hierarchical aligned cluster analysis for temporal clustering of human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 582–596 (2013).
43. Shimodaira, H., Nakai, M., Noma, K. & Sagayama, S. Dynamic Time-Alignment Kernel in Support Vector. *Nips* (2001).

522 44. Wang, Y., Huang, H., Rudin, C. & Shaposhnik, Y. Understanding How Dimension Reduction
523 Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMAP, and PaCMAP
524 for Data Visualization. 1–63 (2020).
525 45. Kobak, D. & Berens, P. The art of using t-SNE for single-cell transcriptomics. *Nature*
526 *Communications* **10**, (2019).
527 46. Gallego, J. A., Perich, M. G., Miller, L. E. & Solla, S. A. Neural Manifolds for the Control of
528 Movement. *Neuron* **94**, 978–984 (2017).
529 47. Valletta, J. J., Torney, C., Kings, M., Thornton, A. & Madden, J. Applications of machine
530 learning in animal behaviour studies. *Animal Behaviour* **124**, 203–220 (2017).
531 48. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. *mclust 5: Clustering, Classification and*
532 *Density Estimation Using Gaussian Finite Mixture Models*. http://cran-logs.rstudio.com.
533 49. Datta, S. R. Q&A: Understanding the composition of behavior. *BMC Biology* **17**, 1–7 (2019).
534 50. Sheets, A. L., Lai, P.-L., Fisher, L. C. & Basso, D. M. Quantitative Evaluation of 3D Mouse
535 Behaviors and Motor Function in the Open-Field after Spinal Cord Injury Using Markerless
536 Motion Tracking. *PLoS ONE* **8**, e74536 (2013).
537 51. Sur, I. & Taipale, J. Dissection of behavior and psychiatric disorders using the mouse as a
538 model. *Nat Reviews Cancer* **9**, 953–965 (2016).
539 52. Zhao, J., Lai, L., Ji, W. & Zhou, Q. Genome editing in large animals: Current status and future
540 prospects. *National Science Review* **6**, 402–420 (2019).
541 53. Sukoff Rizzo, S. J. & Crawley, J. N. Behavioral Phenotyping Assays for Genetic Mouse
542 Models of Neurodevelopmental, Neurodegenerative, and Psychiatric Disorders. *Annual*
543 *Review of Animal Biosciences* **5**, 371–389 (2017).
544 54. Peça, J. *et al.* Shank3 mutant mice display autistic-like behaviours and striatal dysfunction.
545 *Nature* **472**, 437–442 (2011).
546 55. Mei, Y. *et al.* Adult restoration of Shank3 expression rescues selective autistic-like
547 phenotypes. *Nature* **530**, 481–484 (2016).
548 56. Peixoto, R. T. *et al.* Abnormal Striatal Development Underlies the Early Onset of Behavioral
549 Deficits in *Shank3B*−/− Mice. *Cell Reports* **29**, 2016-2027.e4 (2019).
550 57. Rousseau, J. B. I., Van Lochem, P. B. A., Gispen, W. H. & Spruijt, B. M. Classification of rat
551 behavior with an image-processing method and a neural network. *Behavior Research Methods,*
552 *Instruments, and Computers* **32**, 63–71 (2000).
553 58. Kolevzon, A., Delaby, E., Berry-Kravis, E., Buxbaum, J. D. & Betancur, C. Neuropsychiatric
554 decompensation in adolescents and adults with Phelan-McDermid syndrome: A systematic
555 review of the literature. *Molecular Autism* vol. 10 50 (2019).
556 59. Orefice, L. L. *et al.* Targeting Peripheral Somatosensory Neurons to Improve Tactile-Related
557 Phenotypes in ASD Models. *Cell* **178**, 867-886.e24 (2019).

562 66. Wei, P. *et al.* Processing of visually evoked innate fear by a non-canonical thalamic pathway.
563 *Nature Communications* **6**, 1–13 (2015).
564 67. Wang, X. *et al.* Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms
565 of Shank3. *Human Molecular Genetics* **20**, 3093–3108 (2011).
566 68. Robie, A. A., Seagraves, K. M., Egnor, S. E. R. & Branson, K. Machine vision methods for
567 analyzing social interactions. *Journal of Experimental Biology* **220**, 25–34 (2017).
568 69. Pereira, T. D. *et al.* SLEAP: Multi-Animal Pose Tracking. *bioRxiv* 2020.08.31.276246 (2020)
569 doi:10.1101/2020.08.31.276246.
570 70. Chen, Z. *et al.* AlphaTracker: A Multi-Animal Tracking and Behavioral Analysis Tool.
571 *bioRxiv* 2020.12.04.405159 (2020) doi:10.1101/2020.12.04.405159.
572
573

Figures

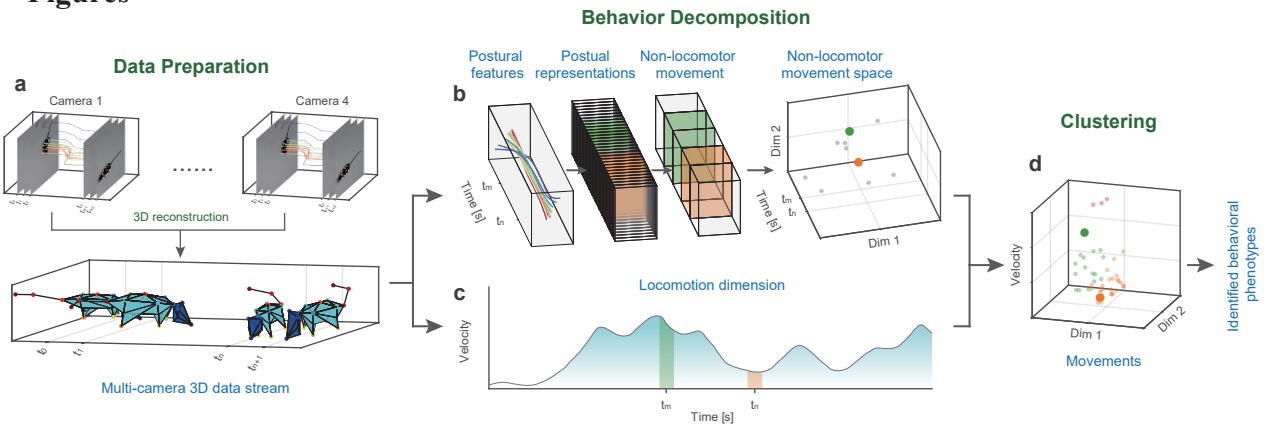

Fig. 1 | Hierarchical 3D-motion learning framework for animal behavior analysis. a Data preparation: 1) image streams captured from four cameras with different 2D views; 2) animal body parts are tracked to generates separate 2D skeletal trajectories (color-coded traces); 3) reconstructing 3D body skeleton 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.

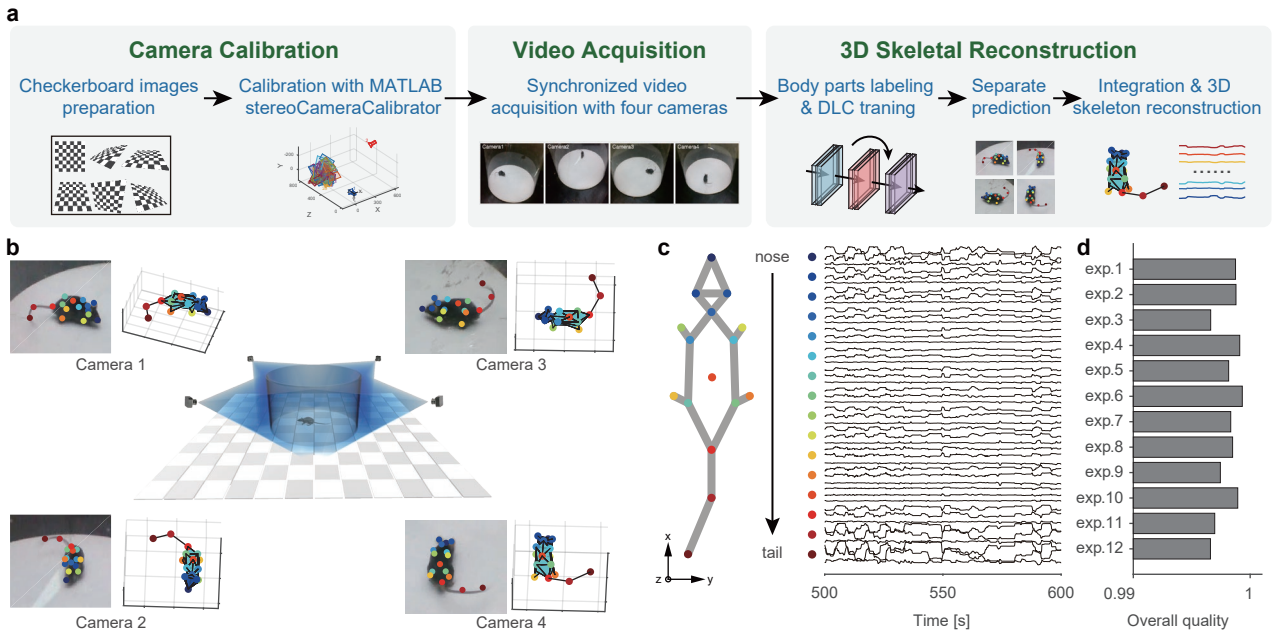

Fig. 2 | Collecting animal behavior trajectories via a 3D motion capture system. a Pipeline of 3D animal skeletal reconstruction. **b** Center, schematic diagram of recording animal behavior with four synchronized cameras; corners, frames captured by the cameras with the DLC labels (left) and the corresponding 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.

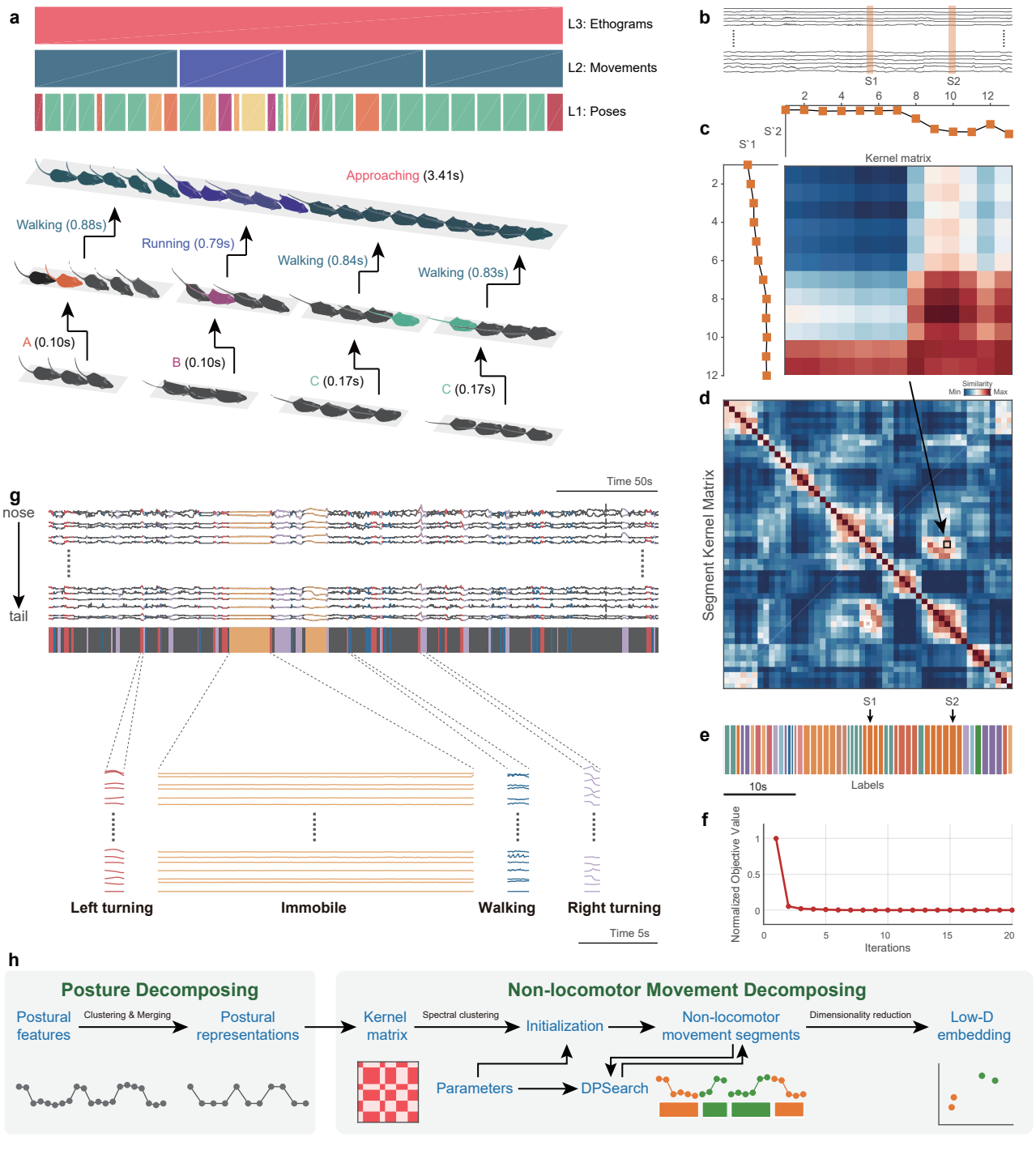

Fig. 3 | Dynamic temporal decomposition of multi-scale hierarchical behavior. a Illustration of the three-layer bottom-up architecture for behavior. Top: The color-coded bars indicate the types of behavior components in the corresponding time period at that layer; each upper layer component is composed 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 ith row and the jth 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 (maxi-mum 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.

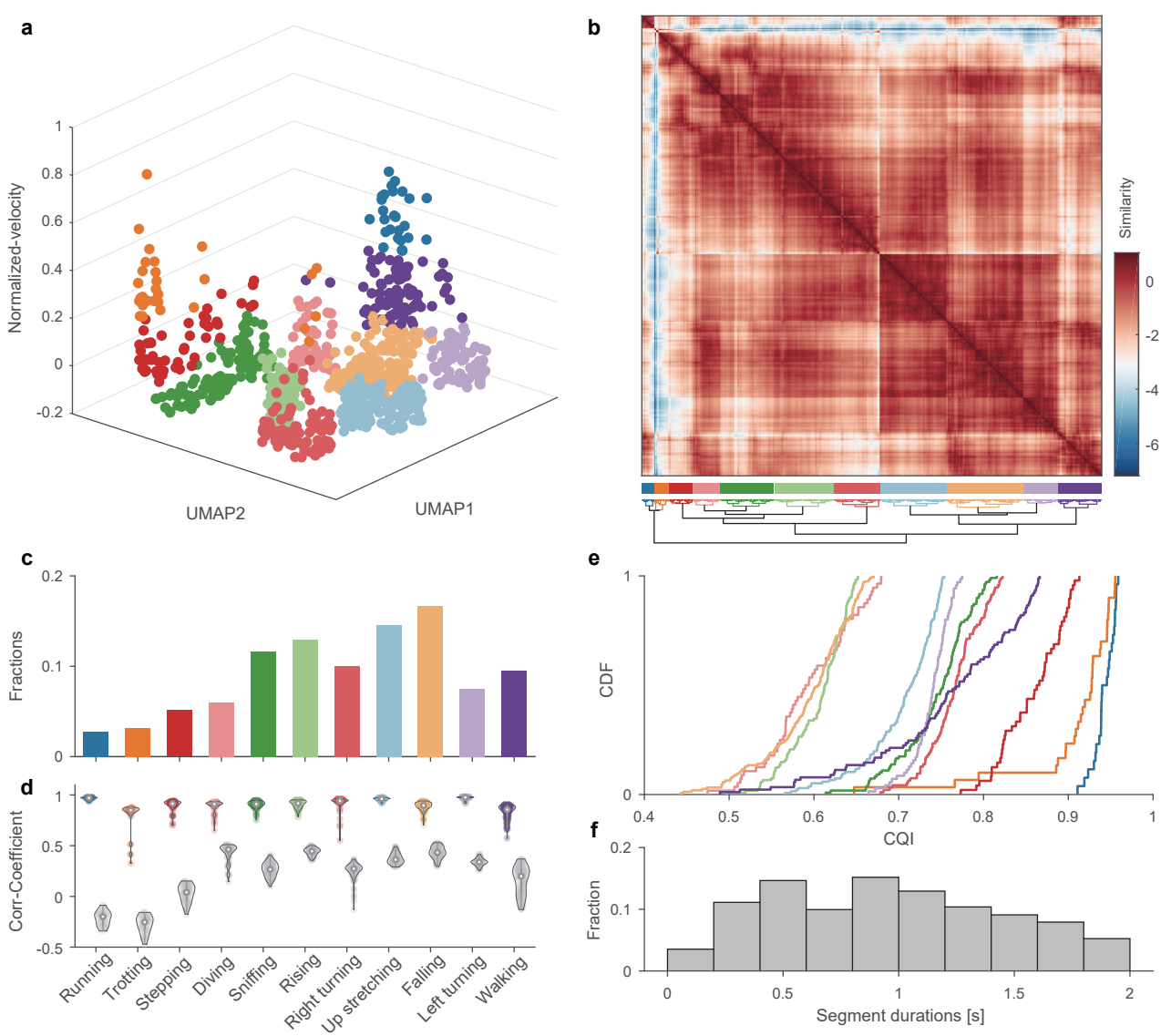

Fig. 4 | Identify movement phenotypes on single experimental data. a Spatio-temporal feature space of behavioral components. Each dot on the 3D scatter plot represents a movement bout (n = 935 bouts). The 11 different colors indicate the corresponding to 11 movement types shown in **d**. **b** Upper, 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 ith row and the jth column. The color-coded bars indicate the labels of clustered movement (middle). **c** Fractions of movement bouts number. For each subject, the behavior fractions are defined as the bouts number of each behavioral phenotype divide by the total number of behavior bouts the animal occurred during the experiment. **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 (CDF) 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).

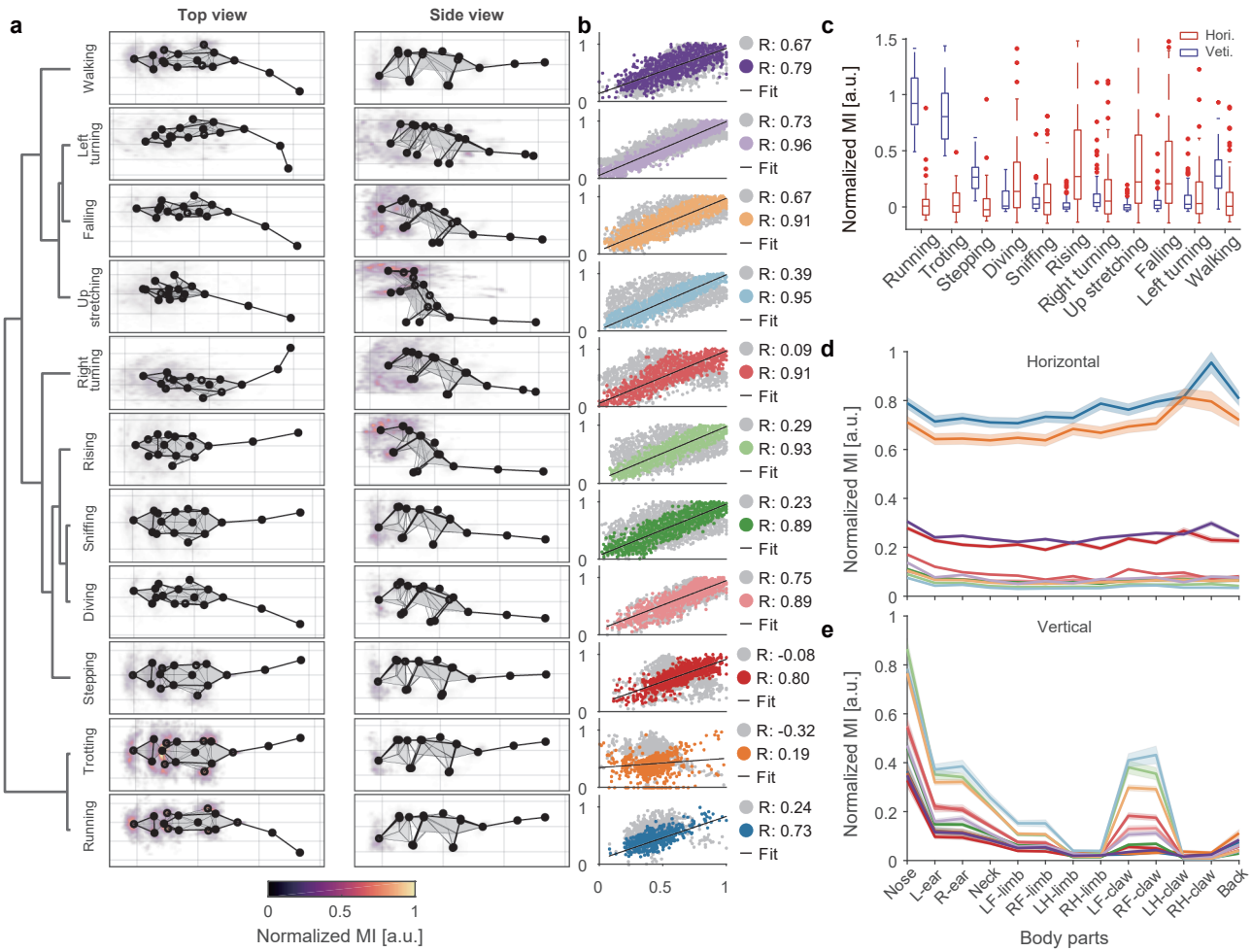

Fig. 5 | Visualization and quantification of behavioral kinematics. a Average skeleton of all frames within each movement phenotype. The skeletons are shown with solid lines and calculated by averaging poses of body parts across time. The heatmaps overlaid on the average skeleton are the distribution and movement intensity (MI; see Supplementary Methods for further details) corresponding to each movement phenotype. **b** Correlation and linear regression plot of movement phenotypes. The horizontal axis represents the target, and the vertical axis represents the reference (see Supplementary Methods for further details). The color-coded and gray dots correspond to the intra- and inter-cluster correlation coefficients, respectively. **c** The comparison of MI between different movement phenotypes. Each movement segment has two MI components (red boxes, horizontal; blue boxes, vertical). The boxes' values for each group contain the MIs of *n* behavioral modules (*n* is the number of behavioral modules of each group). **d, e** Horizontal and vertical MI of each body part in different movement phenotypes. The values on each line are the MIs of all behavior modules corresponding to the phenotype, shown by body parts separately and presented as mean ± standard deviation (SD).

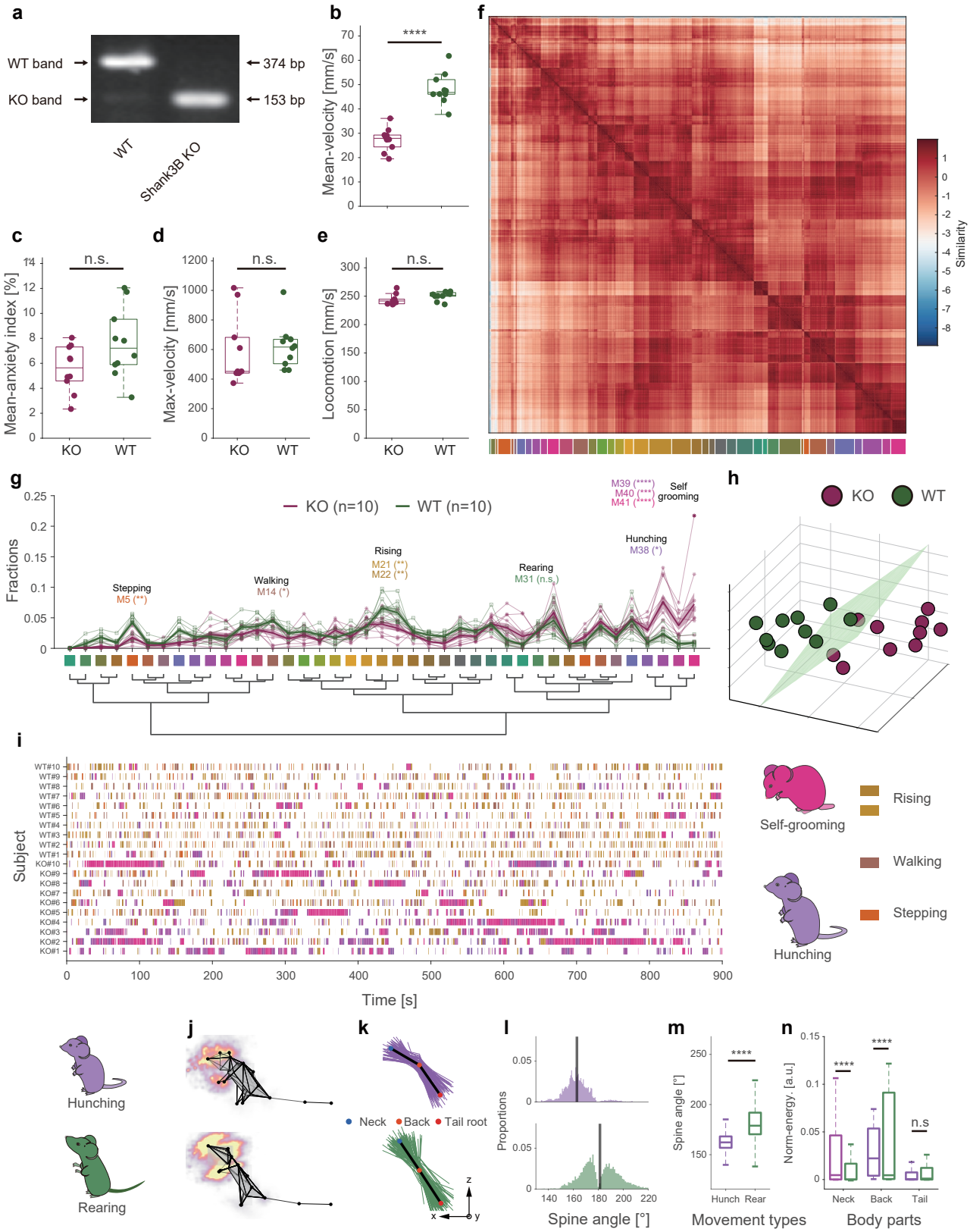

Fig. 6 | Spontaneous behavior analysis reveals autistic-like behaviors in *Shank3B* knock-out mice. **a** PCR genotyping for *Shank3B*$^{+/+}$(Wild Type, WT), *Shank3B*$^{-/-}$ (*Shank3B* Knock-out, KO) mice. **b-e** Box plot of mean velocity, mean anxiety index, maximum velocity, and locomotion of the two groups of animals (purple: KO, n = 10, green: WT, n = 10; Statistics: Mann-Whitney test for maximum velocity; Unpaired T-test for others, **** P<0.0001), values are represented as mean ± SD. **f** Top: recalculated

paired-wise similarity matrix. The movement bouts of all 20 mice involved were grouped (n = 16607) and rearranged in a dendrogram (**g**). Each pixel on the matrix represents the normalized similarity value of a pair of movement bouts at the ith row and the jth column. The color-coded bars (41 clusters) indicate the movements being clustered (bottom); **g** Comparison of the fraction 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 20 mice (purple, KO, n = 10; green, WT, n = 10). Middle color-coded labels and dendrogram indicate the movement types. Eight movements have significant differences between the two groups, and the fractions of the four movements that KO mice prefer are hunching (M38, KO = 3.00 ± 0.56%, WT = 0.94 ± 0.15%) and self-grooming groups (M39, K = 7.65 ± 1.21%, W = 2.34 ± 0.33%; M40, K = 3.73 ± 0.72%, W = 0.75 ± 0.19%; M41, K = 7.23 ± 1.88%, W = 0.90 ± 0.18%). ****P < 0001, **P < 0.01 by two-way ANOVA with Holm–Sidak post-hoc test. **h** Low-dimensional representation of the two animal groups (purple, KO, n = 10; green, WT, n = 10). The 20 dots in 3D space were dimensionally reduced from 41-dimensional movement fractions, and they are well separated. **i** Ethograms of the eight significant movements. **j-n** Kinematic comparison of rearing and hunching (upper row refers to hunching; lower row refers to rearing). **j** Average skeletons of all frames and normalized moving intensity (side view) of rearing and hunching. **k** Spine lines (the lines connecting the neck, back, and tail root) extracted from all frames (rearing, 16834 frames; hunching, 10037 frames) in movement types. For visualization purposes, only 1% of spine lines are shown in the figure (rearing, 168/16834; hunching, 100/10037). Black lines refer to the averaged spine line of the hunching and rearing phenotypes; **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 (181.34 ± 15.63°). By contrast, the spine angles of the rodents during hunching are consistently arcuate (162.88 ± 10.08°). **m, n** Box plot of spine angles of the two movement types. **n** Box plot of normalized MI of the three body parts involved. Statistics for **m, n**: Mann-Whitney test. **** P<0.0001.

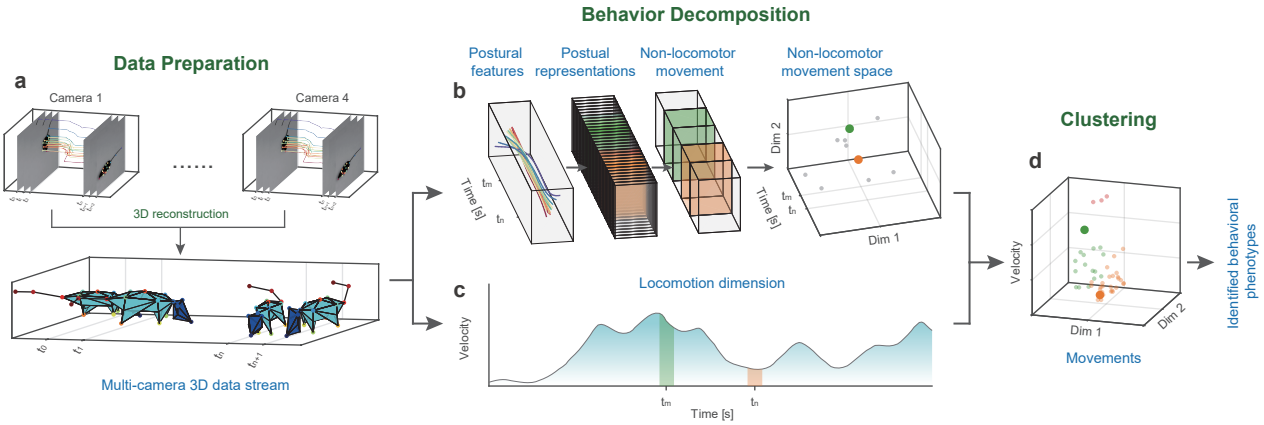

Fig. 1 | Hierarchical 3D-motion learning framework for animal behavior analysis. a Data preparation: 1) image streams captured from four cameras with different 2D views; 2) animal body parts are tracked to generates separate 2D skeletal trajectories (color-coded traces); 3) reconstructing 3D body skeleton 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.

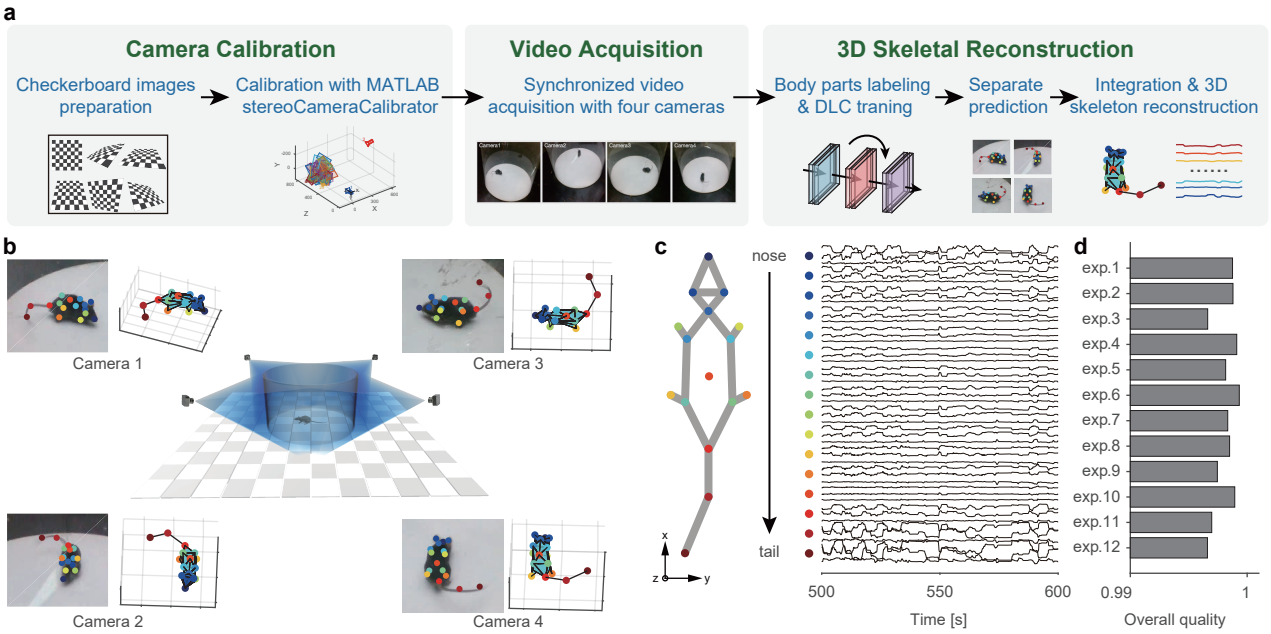

Fig. 2 | Collecting animal behavior trajectories via a 3D motion capture system. a Pipeline of 3D animal skeletal reconstruction. **b** Center, schematic diagram of recording animal behavior with four synchronized cameras; corners, frames captured by the cameras with the DLC labels (left) and the corresponding 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.

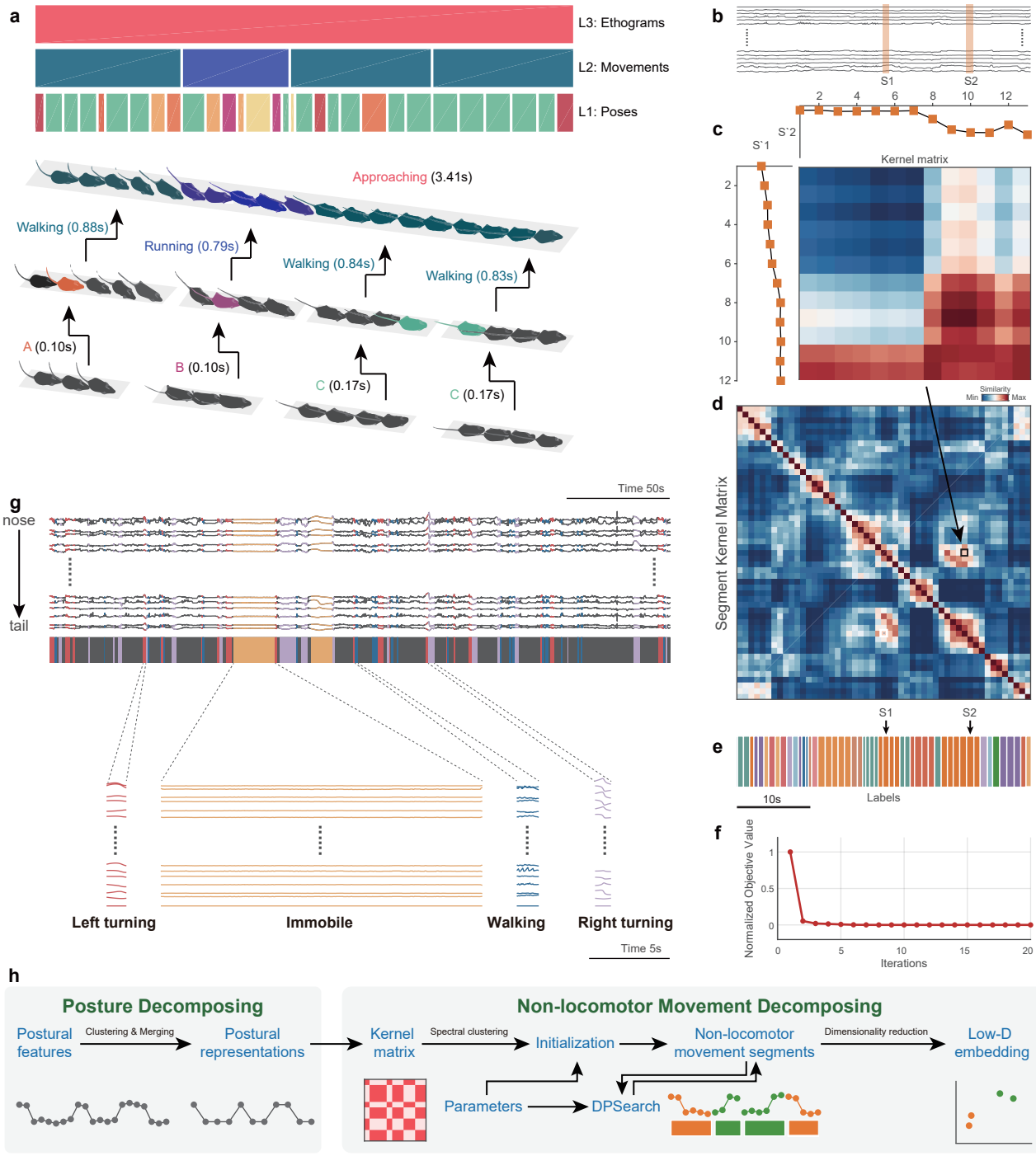

Fig. 3 | Dynamic temporal decomposition of multi-scale hierarchical behavior. a Illustration of the three-layer bottom-up architecture for behavior. Top: The color-coded bars indicate the types of behavior components in the corresponding time period at that layer; each upper layer component is composed 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 ith row and the jth 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.

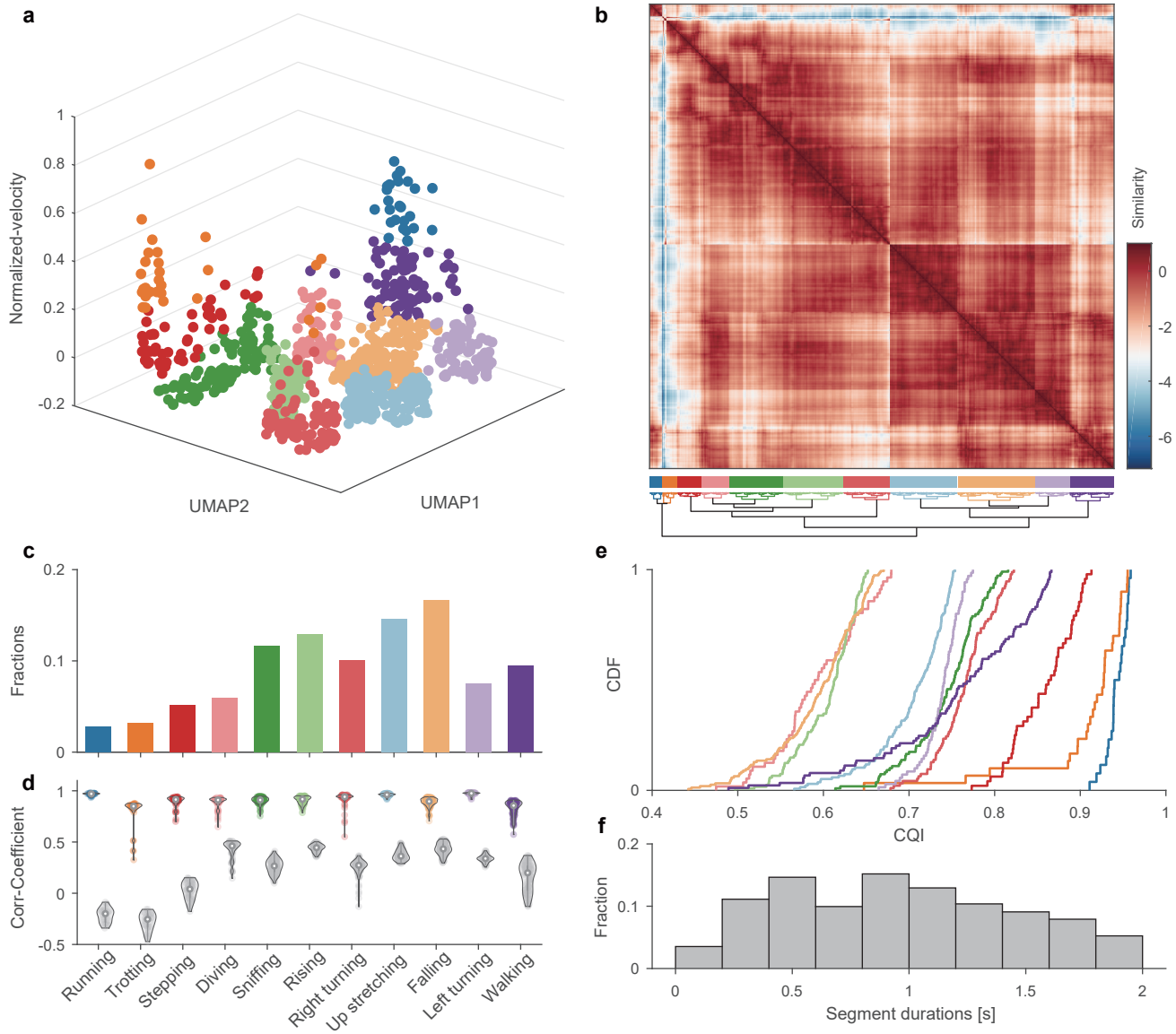

Fig. 4 | Identify movement phenotypes on single experimental data. a Spatio-temporal feature space of behavioral components. Each dot on the 3D scatter plot represents a movement bout (n = 935 bouts). The 11 different colors indicate the corresponding to 11 movement types shown in **d**. **b** Upper, 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 ith row and the jth column. The color-coded bars indicate the labels of clustered movement (middle). **c** Behavior fractions. For each subject, the behavior fractions are defined as the bout number of each behavioral phenotype divided by the total number of behavior bouts the animal produced during the experiment. **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 (CDF) 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).

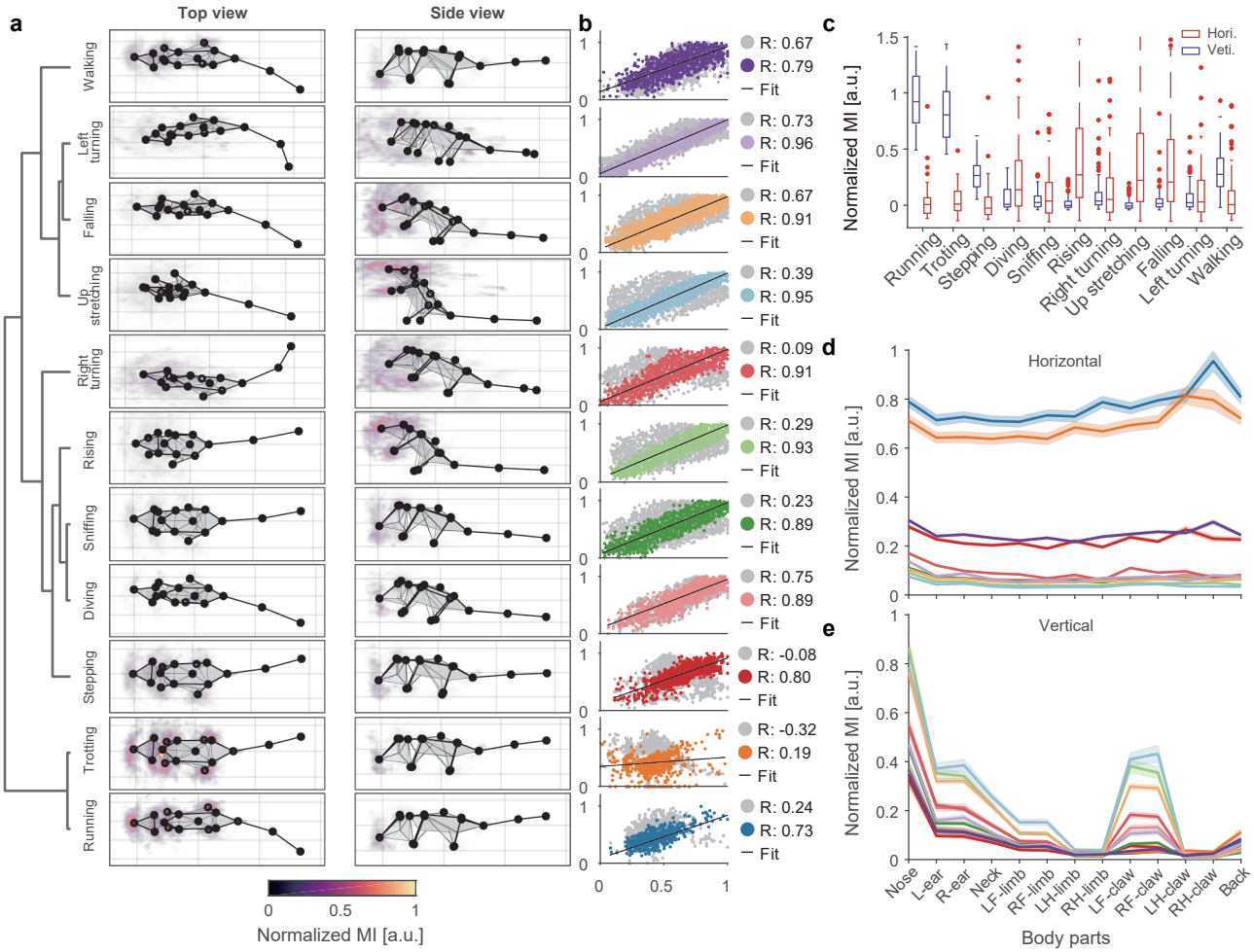

Fig. 5 | Visualization and quantification of behavioral kinematics. a Average skeleton of all frames within each movement phenotype. The skeletons are shown with solid lines and calculated by averaging poses of body parts across time. The heatmaps overlaid on the average skeleton are the distribution and movement intensity (MI; see Supplementary Methods for further details) corresponding to each movement phenotype. **b** Correlation and linear regression plot of movement phenotypes. The horizontal axis represents the target, and the vertical axis represents the reference (see Supplementary Methods for further details). The color-coded and gray dots correspond to the intra- and inter-cluster correlation coefficients, respectively. **c** The comparison of MI between different movement phenotypes. Each movement segment has two MI components (red boxes, horizontal; blue boxes, vertical). The boxes' values for each group contain the MIs of n behavioral modules (n is the number of behavioral modules of each group). **d, e** Horizontal and vertical MI of each body part in different movement phenotypes. The values on each line are the MIs of all behavior modules corresponding to the phenotype, shown by body parts separately and presented as mean ± standard deviation (SD).

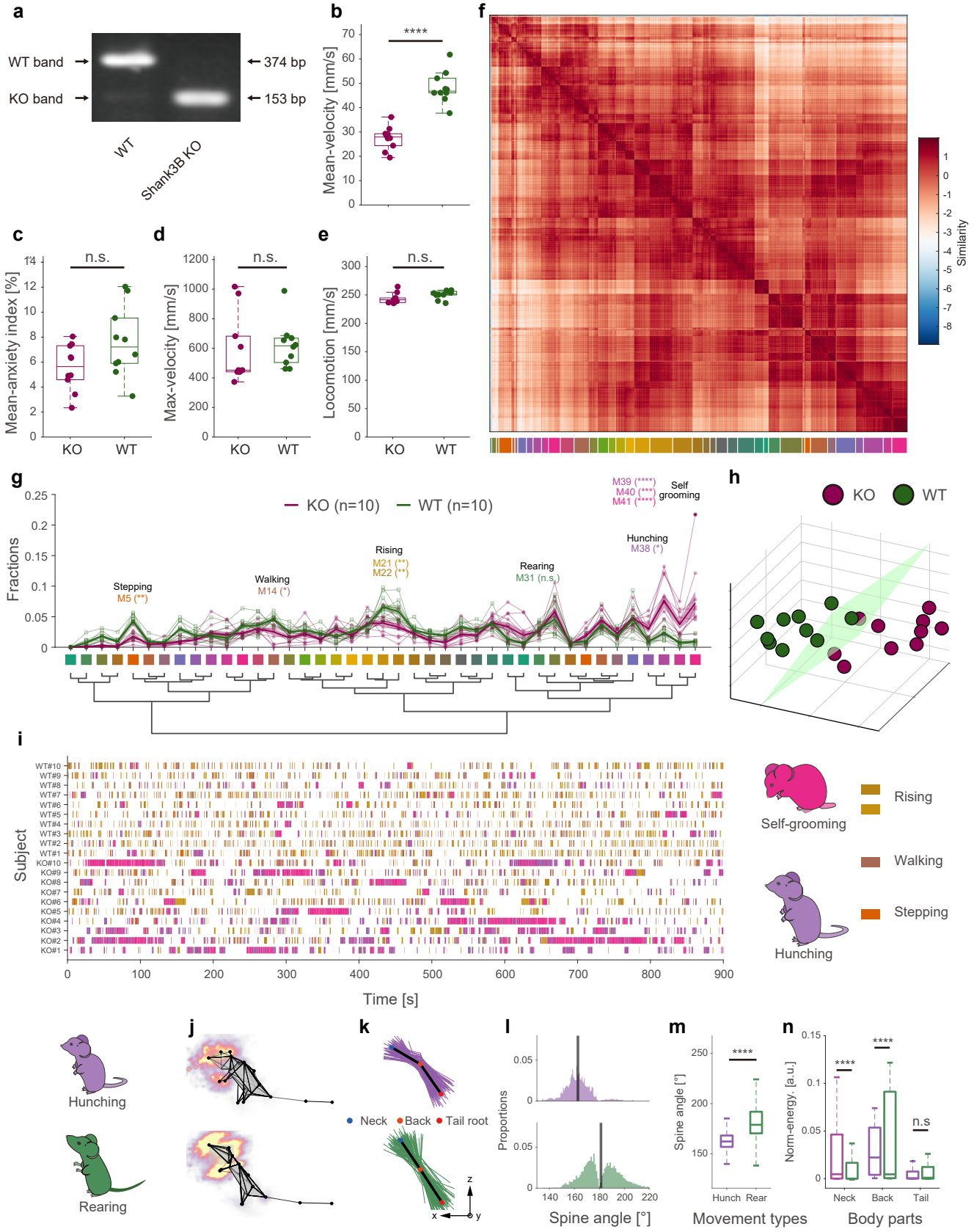

Fig. 6 | Spontaneous behavior analysis reveals autistic-like behaviors on shank3B knock-out mice. a PCR genotyping for Shank3B$^{+/+}$(Wild Type, WT), Shank3B$^{-/-}$ (Shank3B Knock-out, KO) mice. **b-e** Box plot of mean velocity, mean anxiety index, maximum velocity, and locomotion of the two groups of animals (purple: KO, n=10, green: WT, n=10; Statistics: Mann-Whitney test for maximum velocity; Unpaired T-test for others, **** P<0.0001), values are represented as mean±std. **f** Top: recalculated paired-wise similarity matrix. The movement bouts of all of the 20 involved mice were grouped (n = 16607) and were rearranged with dendrogram (**g**). Each pixel on the matrix represents the normalized similarity value of a pair of movement bouts at the ith row and the jth column. The color-coded bars (41 clusters) indicate the movements being

clustered (bottom); **g** Comparison of the fraction 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 20 mice (purple: KO, n=10, green: WT, n=10). Middle color-coded labels and dendrogram indicate the movement types. Eight movements have significant differences between the two groups, and the fractions of the four movements that KO mice prefer are hunching (M38: KO 3.00±0.56%, WT 0.94±0.15%) and self-grooming groups (M39: KO 7.65±1.21%, WT 2.34±0.33%; M40: KO 3.73±0.72%, WT 0.75±0.19%; M41: KO 7.23±1.88%, WT 0.90±0.18%;). ****P<0001, **P<0.01 by two-way ANOVA with Holm–Sidak post-hoc test. **h** Low-dimensional representation of the two animal groups (purple: KO, n=10, green: WT, n=10). The 20 dots in 3D space were dimensionally reduced from 41-dimensional movement fractions, and they are well separated. **i** Ethograms of the eight significant movements. **j-n** Kinematic comparison of rearing and hunching (upper row refers to hunching; lower row refers to rearing). **j** Average-skeletons of all frames and normalized moving intensity (side view) of rearing and hunching. **k** Spine lines (the lines connecting the neck, back, and tail root) extracted from all frames (rearing: 16834 frames, hunching: 10037 frames) in movement types. For visualization purposes, only 1% of spine lines are shown in the figure (rearing: 168/16834, hunching: 100/10037). 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 (181.34±15.63°). By contrast, the spine angles of the rodents during hunching are consistently arcuate (162.88±10.08°). **m, n** Box plot of spine angles of the two movement types. **n** Box plot of normalized MI of the three body parts involved. Statistics for **m, n**: Mann-Whitney test. **** P<0.0001.