

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

Kang Huang<sup>1,2,4</sup>, Yaning Han<sup>1,2,4</sup>, Ke Chen<sup>1,2</sup>, Hongli Pan<sup>1</sup>, Gaoyang Zhao<sup>1,2</sup>, Wenling Yi<sup>1,2</sup>, Xiaoxi Li<sup>1,2</sup>, Siyuan Liu<sup>3</sup>, Liping Wang<sup>1,2,5,\*</sup>, Pengfei Wei<sup>1,2,5,\*</sup>

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.

34 **Introduction:**

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

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

68 The present study proposes a hierarchical 3D-motion learning framework to address our  
69 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<sup>-/-</sup>* 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<sup>32</sup>. 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<sup>33</sup>. Animal behavior is self-organized into a multi-scale hierarchical structure from the bottom up, including poses, movements, and ethograms<sup>34,35</sup>. The poses and movements are low- and intermediate-level elements<sup>36</sup>, while higher-level ethograms are stereotyped patterns composed of movements that adhere to inherent transfer rules in certain semantic environments<sup>37</sup>. 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

105 NM segments and thus equip the model to automatically search repeatable NM sequences. We then  
106 apply the uniform manifold approximation and projection (UMAP)<sup>38</sup> algorithm to visualize high-  
107 dimensional NM representations. After combining the locomotion dimension with NM space (Fig. 1c),  
108 we adopted hierarchical clustering to re-cluster the components and map the behavior's spatial structure  
109 (Fig. 1d).

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

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

120 We collected the naturalistic behavioral data of free-moving mice in a featureless circular open-  
121 field (Supplementary Fig. 2a, and Supplementary Video 1). We analyzed the mouse skeleton as 16 parts  
122 (Fig. 2c) to capture the movements of the rodent's head, torso, paws, and tail. The following motion  
123 quantification did not involve the motion features of two parts of the tail. The data obtained from  
124 tracking representative mouse poses tracking (Fig. 1d) includes the 3D coordinates (x, y, and z) of the  
125 body parts, which reveal that the high-dimensional trajectory series exhibits periodic patterns within a  
126 specific timescale. We next investigated whether the 3D motion capture system could reliably track the  
127 animal in cases of body-part occlusion and viewpoint disappearance. We checked the DeepLabCut  
128 (DLC) tracking likelihood in the collated videos ( $0.9807 \pm 0.1224$ , Supplementary Fig. 4a) and  
129 evaluated the error between the estimated 2D body parts of every training set frame and the ground  
130 truth ( $0.534 \pm 0.005\%$ , Supplementary Fig. 5b). These results indicated that in most cases, four cameras  
131 were available for 2D pose tracking. Since 3D reconstruction can be achieved as long as any two  
132 cameras obtain the 2D coordinates of the same point in 3D space from different views, the  
133 reconstruction failure rate caused by body occlusion and viewpoint disappearances is determined by the  
134 number of available cameras. Therefore, we evaluated the average proportion of available cameras in  
135 situations of body part occlusion and viewpoint disappearance. The validation results for body-part  
136 occlusion show an average reconstruction failure rate of only 0.042% due to body occlusion or  
137 inaccurate body-part estimation (Supplementary Fig. 5c). While for viewpoint disappearances, both  
138 tests (Supplementary Fig. 6, and Supplementary Video 4, 5) proved that our system has a high  
139 reconstruction rate for animal body parts. Moreover, the artifact detection and correction features can

140 recover the body parts that failed to be reconstructed. We calculated an overall reconstruction quality  
141 ( $0.9981 \pm 0.0010$ , Fig. 2d) to ensure that the data were qualified for downstream analysis.

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

143 Conceptually, behavior adheres to a bottom-up hierarchical architecture (Fig. 3a)<sup>34,35</sup>, and  
144 research has focused on elucidating behavioral component sequences contained in stimuli-related  
145 ethograms<sup>40</sup>. The purpose of the two-stage NM decomposition is to bridge the low-level vision features  
146 (postural time-series) to high-level behavioral features (ethograms). The first stage of the decomposition  
147 involves extracting postural representations from postural feature data. Since the definition of NM does  
148 not involve the animal's location or orientation, we pre-processed these data through center alignment  
149 and rotation transformation (Supplementary Fig. 7). Animal movement is continuous, and due to the  
150 high dimensionality of the mammalian skeleton, the behaviorally relevant posture variables are  
151 potentially infinite in number<sup>12</sup>. However, adjacent poses are usually highly correlated and redundant  
152 for behavior quantification and analysis<sup>1</sup>, which is particularly evident in long-term recording.  
153 Therefore, for computational efficiency, we adopted a temporal reduction algorithm to merge adjacent,  
154 similar poses as postural representations in a local time range.

155 In the second stage, NM modules are detected from temporal reduced postural representations.  
156 Unlike the static property of poses, mammalian movements have high dimensionality and large  
157 temporal variability<sup>41</sup>: e.g., the contents, phases, and durations of the three pose sequences were not  
158 the same (Fig. 3a). Hence, we adopted a model-free approach to dynamically perform temporal aligning  
159 and cluster the temporally reduced postural representation data (Fig. 3b)<sup>42</sup>. This problem is equivalent  
160 to providing a d-dimensional time-series  $X \in \mathbb{R}^{d \times n}$  of animal postural representations with  $n$  frames.  
161 Our task decomposes  $X$  into  $m$  NM segments, each of which belongs to one of the corresponding  $k$   
162 behavioral clusters. This method detects the change point by minimizing the error across segments;  
163 therefore, dynamic temporal segmentation becomes a problem of energy minimization. An appropriate  
164 distance metric is critical for modeling the temporal variability and optimizing the NM segmentation of  
165 a continuous postural time-varying series. Although dynamic time warping (DTW) has commonly been  
166 applied in aligning time-series data, it does not satisfy the triangle inequality<sup>43</sup>. Thus, we used the  
167 improved DTAK method to measure the similarity between time sequences and construct an energy  
168 equation (objective function) for optimization. The relationship between each pair of segments was  
169 calculated with the kernel similarity matrix  $K$  (Fig. 3c). DTAK was the used to compute the normalized  
170 similarity value of  $K$  and generate the paired-wise segment kernel matrix  $T$  (Fig. 3d).

171 Because dynamic temporal segmentation is a non-convex optimization problem whose solution  
172 is very sensitive to initial conditions, this approach begins with a coarse segmentation process based on  
173 the spectral clustering method, which combines the kernel k-means clustering algorithms. To define the  
174 timescale of segmentation, the algorithm sets the maximum and minimum lengths [ $w_{min}, w_{max}$ ] to

175 constrain the length of the behavioral component. For the optimization process, a dynamic  
176 programming (DP)-based algorithm is employed to perform coordinate descent and minimize energy.  
177 For each iteration, the algorithm updates the segmentation boundary and segment kernel matrix until  
178 the decomposition reaches the optimal value (Fig. 3e, f). The final segment kernel matrix represents the  
179 optimal spatial relationship between these NM segments, which can be further mapped into its feature  
180 space in tandem with dimensionality reduction (DR).

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

190 **Mapping Mouse Movements with Low-Dimensional Embeddings and Unsupervised  
191 Clustering**

192 We validated our framework in a single-session experiment with free-moving mouse behavioral  
193 data collected with the 3D motion capture system. First, the two-stage behavioral decomposition  
194 strategy decomposed the 15-minute experimental data into 936 NM bouts (Supplementary Video 2). A  
195  $936 \times 936$  segment kernel matrix was then constructed using the DTAK metric. This segment kernel  
196 matrix could flexibly represent the relationship and provide insight into the relationships between each  
197 behavioral component sequence in their feature space. However, since the 936-D matrix cannot provide  
198 an informative visualization of behavioral structure, it is necessary to perform DR on this data. Various  
199 DR algorithms have been designed either to preserve the global representation of original data or to  
200 focus on local neighborhoods for recognition or clustering<sup>44,45</sup>. Thus, in animal behavior quantification,  
201 we face a trade-off between discretizing behavior to provide a more quantitative analysis and  
202 maintaining a global representation of behavior to characterize the potential manifolds of neural-  
203 behavioral relationships<sup>46</sup>. Therefore, we first evaluated the commonly used DR algorithms from the  
204 standpoints of preserving either the global or the local structure. The evaluation results show that UMAP  
205 can balance both aspects for our data (Supplementary Fig. 8) and provide 2D embeddings of these NM  
206 segments. In addition, in our parallel feature fusion framework, the factor of an animal's interaction  
207 with the environment – i.e., velocity – is considered an independent dimension. Together with 2D NM  
208 embedding, they construct a spatio-temporal representation of movements (Fig. 4a).

209 We used an unsupervised clustering algorithm to investigate the behavior's spatio-temporal  
210 representation and identify the movement phenotypes. Most unsupervised clustering require a pre-  
211 specified number of clusters, and the number chosen can be data-driven or refer to the context of the  
212 practical biological problem <sup>47</sup>. In the single experimental data shown in Figure 4a, the data-driven  
213 Bayesian Information Criterion <sup>48</sup> in the R package *mclust* was adopted to determine that the optimal  
214 cluster number was 11 (Supplementary Fig. 10). We then recalculated the similarity matrices in the new  
215 feature space (Fig. 4b) and aggregated them using a hierarchical clustering method. Finally, we cut the  
216 original video into clips of  $0.963 \pm 0.497$  s and manually labeled them according to the behavior of the  
217 rodents in the clip: running, trotting, stepping, diving, sniffing, rising, right turning, up stretching,  
218 falling, left turning, and walking (Supplementary Table 1). The locomotion types of running, trotting,  
219 stepping, and walking accounted for 20.6% of the total activities, indicating that animals spent most of  
220 the time in the NM stage (Fig. 4c).

221 Although we phenotyped all the clips of the entire video, it was difficult to label the behaviors  
222 of the rodents with only 11 definitions. Further, there are various heterogeneous transition stages  
223 between bouts of stereotyped movements <sup>20,31,49</sup>. Therefore, we evaluated them by calculating the intra-  
224 cluster and inter-cluster correlation coefficients (intra-CC and inter-CC, respectively; Fig. 4d, Fig. 5b).  
225 Our results showed that running, up stretching and left turning have higher intra-CC and lower inter-  
226 CC, while walking and sniffing have both higher intra-CC and higher inter-CC. This is because walking  
227 and sniffing co-occur with other movements <sup>13</sup>, such as diving and turning, respectively. Finally, to  
228 evaluate the overall clustering quality, we integrated these two parameters and defined the Clustering  
229 Quality Index (CQI, Fig. 4e), which helped to determine the stereotyped/non-stereotyped movements.

## 230 Kinematic Validation of Mouse Behavioral Phenotypes

231 DTAK is an abstract extraction of animal motions that aims to simplify the complex temporal  
232 dynamics of behavior. Hence, we further elucidated whether the spatial kinematics of the original  
233 postural time-series of the behavioral phenotypes (e.g., running, rearing, sniffing, turning) identified  
234 with this framework were homogeneous. Manually inspecting the position, moving, bending, and other  
235 characteristics of the mouse limbs and trunk in the video clips of each phenotype group (Supplementary  
236 Video 3), we found reliable homogeneity for clips with high CQIs ( $CQI > 0.75$ ). To provide a kinematic  
237 validation of the identified behavioral phenotypes from the perspectives of visualization and  
238 quantification, we first visualized the average skeleton, which was averaged over all frames in each  
239 movement cluster (Fig. 5a). While some movements could be clearly recognized (e.g., left and right  
240 turning, and up stretching), the differences between movements with similar postures (running, trotting,  
241 walking, etc.) were not. The detailed kinematic parameters, especially the velocity of each body part,  
242 could provide greater sensitive differences than the unclear visually-based assessments <sup>50</sup>. Therefore,  
243 we defined movement intensity (MI) as a metric for characterizing the kinematics of each body part in

244 each behavioral phenotype (see Supplementary Methods for further details). MI is related to velocity,  
245 and it contains both horizontal and vertical components. The data show that the horizontal MI  
246 components of running and trotting are the highest, followed by stepping and walking. Vertical MI  
247 components (e.g., up stretching, rising, and falling) feature richer details; we attribute their high overall  
248 vertical MI to the movement of the nose and front claws (Fig. 5a, c-e). This approach of creating  
249 portraits for each type of movement provides further support for the efficacy of our framework in the  
250 decomposition of animal behavior. The dendrogram of the movements (Fig. 5a) revealed that similar  
251 movements were arranged were closely, such as running and trotting. Interestingly, falling and left  
252 turning were on close clades. Review of the video clips of these two groups demonstrated that 37.18%  
253 of the movements in this group occurred simultaneously with left turning (28.85% for right turning). A  
254 similar phenomenon occurred in the clades of diving and sniffing due to the co-occurrence of these  
255 behaviors. The correlation and linear regression analysis of these two pairs of clades showed that both  
256 intra-CC and inter-CC were relatively high (Fig. 5b), suggesting several concomitant descriptions of  
257 animal behavior. These clustering results occurred because these movements show more characteristics  
258 of the current class.

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

260 Animal disease models play an increasingly critical role in expanding understanding of the  
261 mechanisms of human diseases and novel therapeutic development <sup>51-53</sup>. Behavioral phenotyping  
262 provides a noninvasive approach to the assessment of neuropsychiatric disorders in animal models. By  
263 only evaluating spontaneous behavior without any induced conditions, we demonstrate the usability  
264 and unbiased character of our framework for animal phenotyping. We collected data from 20 mice (Fig.  
265 6a, n<sub>KO</sub> = 10, n<sub>WT</sub> = 10) with our 3D motion capture system and subjected them to routine velocity and  
266 anxiety index analyses (Fig. 6b-e). In agreement with prior research, we found a significant difference  
267 between the average velocities of the two groups.

268 We clustered the behavioral components of the 20 animals and obtained 41 behavioral  
269 phenotypes (Fig. 6f, Supplementary Fig. 10). Compared with the single-session experiment, the group  
270 analysis revealed diverse behavioral types. We found that *Shank3B* Knock-out (KO, *Shank3B*<sup>-/-</sup>) mice  
271 spent a significantly higher proportion of their time engaging in four of the movements (Fig. 6g,  
272 Supplementary Table 3). By manually reviewing the video clips of these four types, we annotated the  
273 38<sup>th</sup> movement (M38 in Fig. 6g) as hunching; we also found that three of the movements were very  
274 similar (closely arranged on the behavioral dendrogram, Fig. 6g). Therefore, we grouped them and  
275 annotated them as self-grooming. In previous studies <sup>54-56</sup>, self-grooming has been widely reported in  
276 *Shank3B*<sup>-/-</sup> mice. This is partly attributable to self-grooming being a long-lasting movement ( $4.48 \pm 7.84$   
277 s, mean  $\pm$  standard deviation [SD]) and thus easily recognized by human observation or software (Fig.  
278 6i). Interestingly, although hunching has only previously been reported in a few related studies <sup>57-59</sup>,

279 our framework frequently detected hunching movements in KO mice. This novel finding can be  
280 attributed to the duration of a single continuous hunching movement being too short to be noticed ( $1.29 \pm 1.00$  s, mean  $\pm$  SD) as well as to the similarity between the kinematics of hunching and rearing (M31).  
281 We proved that these two types of movements belong to distinct behavioral phenotypes. Specifically,  
282 during hunching, mice maintain an arcuate spine angle, while rearing is characterized by a stronger,  
283 wider range of necks and head motions (Fig. j-n). This ability to identify short-term and fine behavioral  
284 modules is one of the advantages of our framework. Besides the four phenotypes that KO mice preferred  
285 more than the WT mice did, the KO mice also showed four additional deficit behavioral phenotypes,  
286 namely stepping (M5), walking (M14), and two types of rising (M21 and M22). This result indicates  
287 that the locomotion intensity and vertical movement of KO mice were lower than those of WT mice.  
288 The locomotion result is consistent with the average velocity comparison shown in Fig. 6b.

290 Finally, we demonstrated that by modeling the time spent of multi-behavioral parameters, our  
291 framework could identify the animal types. We used UMAP to perform DR of the 41-dimensional  
292 behavioral proportion data of all movement types. As expected, the two genotypes of animals were well  
293 separated in the low-dimensional space (Fig. 6h), even though there were large amounts of baseline  
294 movements with no significant difference. We defined these two types as “autistic-like behavior space.”  
295 Recent reviews suggest that most previous methods<sup>60,61</sup>, which usually only consider a few behavioral  
296 parameters and may lose critical insights, have been challenged in the animal phenotypes’ identification.  
297 Hence, these findings indicate the potential advantages of our framework to automatically identify  
298 disease models.

## 299 Discussion

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

310 Our framework has two main advantages. First, our approach of tracking multiple body parts  
311 and acquiring 3D reconstruction data achieves better performance than similar recently reported rodent  
312 behavioral recognition frameworks<sup>14,62</sup>. The multi-view motion capture system can avoid animal body  
313 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<sup>63–65</sup>. 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<sup>37,66</sup>, 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*<sup>-/-</sup> 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<sup>54,55</sup> 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

351 level when tested under the white-light circular open field condition (Supplementary Fig. 12 c, d).  
352 Finally, we compared the behavior patterns when all conditions were the same except for the genotypes.  
353 The results showed that only the female group showed no significant difference between KO and WT  
354 genotypes, while significant differences in behavioral patterns were found between KO and WT male  
355 mice under all other conditions. These findings are consistent with previous reports that *Shank3B* KO  
356 male mice display more severe impairments than females do in motor coordination<sup>67</sup>. Accordingly, the  
357 behavior phenotyping on mouse disease model can be generalized to large animals such as non-human  
358 primates, dogs, and pigs which recently emerged as valuable models for studying neurological  
359 dysfunctions<sup>52,53</sup>. Our general-purpose framework further benefits from the significant advantage of  
360 being able to capture and analyze large animal movements, which have more complex 3D  
361 characteristics and temporal dynamics.

362 The dynamic, high-dimensional, and multi-scale characteristics of behavior can be attributed  
363 to similar properties of the nervous system produces it. While the most advanced large-scale  
364 neuroimaging and high spatiotemporal resolution electrophysiological techniques allow researchers to  
365 elucidate the details of the firing timing of all neurons and neurofunctional connections at all scales,  
366 they cannot inform the mapping of the neural-behavioral relationship without quantifying behavior at  
367 the corresponding level. In other words, to understand the encoding/decoding relationship rules of the  
368 neural activity generating behavior and behavior's neural representation, synchronization of large  
369 population activities and accurate measurement and identification of naturalistic, complex behavior are  
370 required. In the future, we will focus on combining our framework with free-moving two-photon  
371 microscopy and electrophysiological recording to link the neural activity patterns and functional brain  
372 connections with the cross-scale behavioral dynamics and timing patterns. Therefore, with further  
373 technical optimization and the open-source of a large sample, well-annotated disease model behavior  
374 database open source, our framework may contribute to resolving the relationships between complex  
375 neural circuitry and behavior, as well as to revealing the mechanisms of sensorimotor processing.

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

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

395 **Acknowledgements**

396 We thank Jianyuan Sun, Guoqiang Bi, Yu-Ting Tseng and Yunke Wang for comments on our  
397 manuscript; Feng Wang, Zhonghua Lu, and Yu Hu providing *Shank3B<sup>-/-</sup>* Mice.

398 **Funding**

399 This work was supported in part by Key Area R&D Program of Guangdong Province  
400 (2018B030338001 P.W., 2018B030331001 L.W.), National Key R&D Program of China  
401 (2018YFA0701400 P.W.), National Natural Science Foundation of China (NSFC 31500861 P.W.,  
402 NSFC 31630031 L.W., NSFC 91732304 L.W., NSFC 31930047 L.W.), the International Big Science  
403 Program Cultivating Project of CAS (172644KYS820170004 L.W.), the Strategic Priority Research  
404 Program of Chinese Academy of Science (XDB32030100, L.W.), the Youth Innovation Promotion  
405 Association of the Chinese Academy of Sciences (2017413 P.W.), Shenzhen Government Basic  
406 Research Grants (JCYJ20170411140807570 P.W., JCYJ20170413164535041 L.W.), Science,  
407 Technology and Innovation Commission of Shenzhen Municipality (JCYJ20160429185235132 K.H.),  
408 Helmholtz-CAS joint research grant (GJHZ1508 L.W.), Guangdong Provincial Key Laboratory of  
409 Brain Connectome and Behavior (2017B030301017 L.W.), the Ten Thousand Talent Program (L.W.),  
410 the Guangdong Special Support Program (L.W.), Key Laboratory of SIAT (2019DP173024 L.W.),  
411 Shenzhen Key Science and Technology Infrastructure Planning Project (ZDKJ20190204002 L.W.).

412

413 **Author Contributions**

414 Conceptualization, K.H., Y.H., P.W., and L.W.; Methodology, K.H., Y.H., K.C., P.W., L.W.,  
415 and S.L.; Data Analysis, K.H., Y.H., and K.C.; Animal Experiments and Data Collection, Y.H., H.P.,  
416 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  
417 Acquisition, P.W., and L.W.; and Supervision, L.W. and P.W.

418 **Declaration of Interests**

419 The authors declare no competing interests.

420 **Data Availability**

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

424 **Code Availability**

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

426 **References**

- 427 1. Hinde, R. A. & Bateson, P. P. G. *Growing Points Ethology*. (CUP Archive, 1976).
- 428 2. Tinbergen, N. *The study of instinct*. (Pygmalion Press, an imprint of Plunkett Lake Press,  
429 2020).
- 430 3. Wiltschko, A. B. *et al.* Mapping Sub-Second Structure in Mouse Behavior. *Neuron* **88**, 1121–  
431 1135 (2015).
- 432 4. Marques, J. C., Lackner, S., Félix, R. & Orger, M. B. Structure of the Zebrafish Locomotor  
433 Repertoire Revealed with Unsupervised Behavioral Clustering. *Current Biology* **28**, 181–  
434 195.e5 (2018).
- 435 5. Berman, G. J., Bialek, W. & Shaevitz, J. W. Predictability and hierarchy in Drosophila  
436 behavior. *Proceedings of the National Academy of Sciences of the United States of America*  
437 **113**, 11943–11948 (2016).
- 438 6. Johnson, R. E. *et al.* Probabilistic Models of Larval Zebrafish Behavior Reveal Structure on  
439 Many Scales. *Current Biology* **30**, 70-82.e4 (2020).
- 440 7. Hollon, N. G. & Jin, X. Neural Implementation of Behavioral Hierarchy. *Neuron* **105**, 402–  
441 404 (2020).
- 442 8. Kaplan, H. S. & Zimmer, M. Brain-wide representations of ongoing behavior: a universal  
443 principle? *Current Opinion in Neurobiology* **64**, 60–69 (2020).
- 444 9. Kaplan, H. S., Salazar Thula, O., Khoss, N. & Zimmer, M. Nested Neuronal Dynamics  
445 Orchestrate a Behavioral Hierarchy across Timescales. *Neuron* **105**, 562-576.e9 (2020).
- 446 10. Vogelstein, J. T. *et al.* Discovery of brainwide neural-behavioral maps via multiscale  
447 unsupervised structure learning. *Science* **344**, 386–392 (2014).
- 448 11. Markowitz, J. E. *et al.* The Striatum Organizes 3D Behavior via Moment-to-Moment Action  
449 Selection. *Cell* **174**, 44-58.e17 (2018).

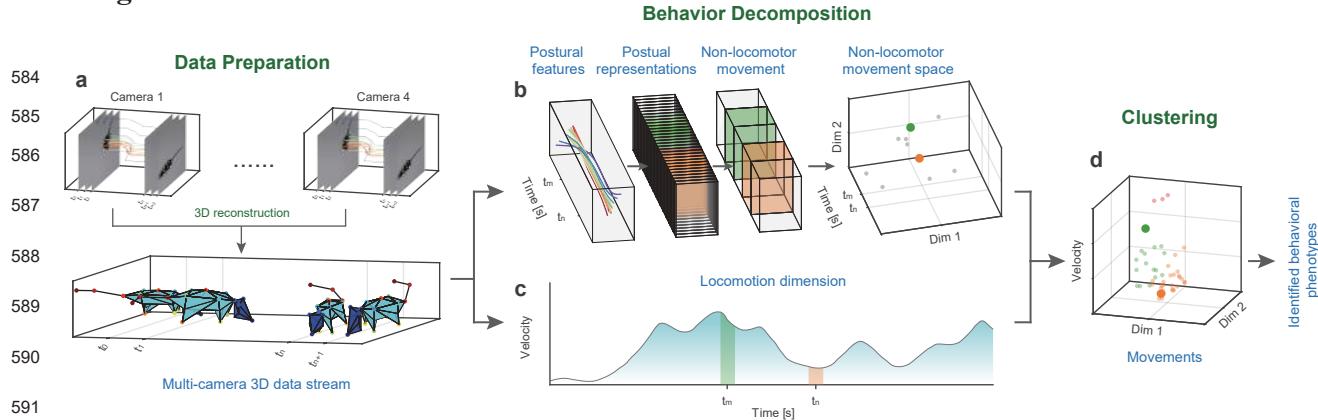
- 450 12. von Ziegler, L., Sturman, O. & Bohacek, J. Big behavior: challenges and opportunities in a  
451 new era of deep behavior profiling. *Neuropsychopharmacology* 1–12 (2020)  
452 doi:10.1038/s41386-020-0751-7.
- 453 13. Datta, S. R., Anderson, D. J., Branson, K., Perona, P. & Leifer, A. Computational  
454 Neuroethology: A Call to Action. *Neuron* **104**, 11–24 (2019).
- 455 14. Hsu, A. I. & Yttri, E. A. B-SOiD: An Open Source Unsupervised Algorithm for Discovery of  
456 Spontaneous Behaviors. *bioRxiv* 770271 (2019) doi:10.1101/770271.
- 457 15. Ravbar, P., Branson, K. & Simpson, J. H. An automatic behavior recognition system classifies  
458 animal behaviors using movements and their temporal context. *Journal of Neuroscience*  
459 *Methods* **326**, 108352 (2019).
- 460 16. Günel, S. *et al.* Deepfly3D, a deep learning-based approach for 3D limb and appendage  
461 tracking in tethered, adult *Drosophila*. *eLife* **8**, 1–22 (2019).
- 462 17. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with  
463 deep learning. *Nature Neuroscience* **21**, 1281–1289 (2018).
- 464 18. Graving, J. M. *et al.* Deepposekit, a software toolkit for fast and robust animal pose estimation  
465 using deep learning. *eLife* **8**, 1–42 (2019).
- 466 19. Pereira, T. D. *et al.* Fast animal pose estimation using deep neural networks. *Nature Methods*  
467 **16**, 117–125 (2019).
- 468 20. Berman, G. J., Choi, D. M., Bialek, W. & Shaevitz, J. W. Mapping the stereotyped behaviour  
469 of freely moving fruit flies. *Journal of the Royal Society Interface* **11**, 20140672 (2014).
- 470 21. Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S. & Branson, K. JAABA: Interactive  
471 machine learning for automatic annotation of animal behavior. *Nature Methods* **10**, 64–67  
472 (2013).
- 473 22. Branson, K., Robie, A. A., Bender, J., Perona, P. & Dickinson, M. H. High-throughput  
474 ethomics in large groups of *Drosophila*. *Nature Methods* **6**, 451–457 (2009).
- 475 23. Sharma, A., Johnson, R. E., Engert, F. & Linderman, S. W. *Point process latent variable*  
476 *models of larval zebrafish behavior*. (2018).
- 477 24. Haesemeyer, M., Robson, D. N., Li, J. M., Schier, A. F. & Engert, F. The Structure and  
478 Timescales of Heat Perception in Larval Zebrafish. *Cell Systems* **1**, 338–348 (2015).
- 479 25. Bruni, G. *et al.* Zebrafish behavioral profiling identifies multitarget antipsychotic-like  
480 compounds. *Nature Chemical Biology* **12**, 559–566 (2016).
- 481 26. Brown, A. E. X., Yemini, E. I., Grundy, L. J., Jucikas, T. & Schafer, W. R. A dictionary of  
482 behavioral motifs reveals clusters of genes affecting *Caenorhabditis elegans* locomotion.  
483 *Proceedings of the National Academy of Sciences* **110**, 791–796 (2013).
- 484 27. Yemini, E., Jucikas, T., Grundy, L. J., Brown, A. E. X. & Schafer, W. R. A database of  
485 *Caenorhabditis elegans* behavioral phenotypes. *Nature Methods* **10**, 877–879 (2013).

- 486 28. Liu, M., Sharma, A. K., Shaevitz, J. W. & Leifer, A. M. Temporal processing and context  
487 dependency in *caenorhabditis elegans* response to mechanosensation. *eLife* **7**, (2018).
- 488 29. Storchi, R., Milosavljevic, N., Allen, A. E., Cootes, T. F. & Lucas, R. J. Beyond locomotion :  
489 in the mouse the mapping between sensations and behaviours unfolds in a higher dimensional  
490 space. *bioRxiv* 2020.02.24.961565 (2020) doi:10.1101/2020.02.24.961565.
- 491 30. Liu, D. *et al.* A common hub for sleep and motor control in the substantia nigra. *Science* **367**,  
492 440–445 (2020).
- 493 31. Jhuang, H. *et al.* Automated home-cage behavioural phenotyping of mice. *Nature  
494 Communications* **3**, (2012).
- 495 32. Charles, J. P., Cappellari, O. & Hutchinson, J. R. A Dynamic Simulation of Musculoskeletal  
496 Function in the Mouse Hindlimb During Trotting Locomotion. *Frontiers in Bioengineering  
497 and Biotechnology* **6**, 61 (2018).
- 498 33. Berman, G. J. Measuring behavior across scales. *BMC Biology* **16**, 1–12 (2018).
- 499 34. Gris, K. V., Couto, J. P. & Gris, D. Supervised and unsupervised learning technology in the  
500 study of rodent behavior. *Frontiers in Behavioral Neuroscience* **11**, 1–6 (2017).
- 501 35. Anderson, D. J. & Perona, P. Toward a science of computational ethology. *Neuron* **84**, 18–31  
502 (2014).
- 503 36. Wang, L., Hu, W. & Tan, T. Recent developments in human motion analysis. *Pattern  
504 Recognition* **36**, 585–601 (2003).
- 505 37. Evans, D. A., Stempel, A. V., Vale, R. & Branco, T. Cognitive Control of Escape Behaviour.  
506 *Trends in Cognitive Sciences* **23**, 334–348 (2019).
- 507 38. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and  
508 Projection for Dimension Reduction. (2018).
- 509 39. Sheshadri, S., Dann, B., Hueser, T. & Scherberger, H. 3D reconstruction toolbox for behavior  
510 tracked with multiple cameras. *Journal of Open Source Software* **5**, 1849 (2020).
- 511 40. Gomez-Marin, A., Paton, J. J., Kampff, A. R., Costa, R. M. & Mainen, Z. F. Big behavioral  
512 data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience* **17**,  
513 1455–1462 (2014).
- 514 41. Zhou, F., De La Torre, F. & Hodgins, J. K. Aligned cluster analysis for temporal segmentation  
515 of human motion. *2008 8th IEEE International Conference on Automatic Face and Gesture  
516 Recognition, FG 2008* (2008) doi:10.1109/AFGR.2008.4813468.
- 517 42. Zhou, F., Torre, F. D. La & Hodgins, J. K. Hierarchical aligned cluster analysis for temporal  
518 clustering of human motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*  
519 **35**, 582–596 (2013).
- 520 43. Shimodaira, H., Nakai, M., Noma, K. & Sagayama, S. Dynamic Time-Alignment Kernel in  
521 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 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).

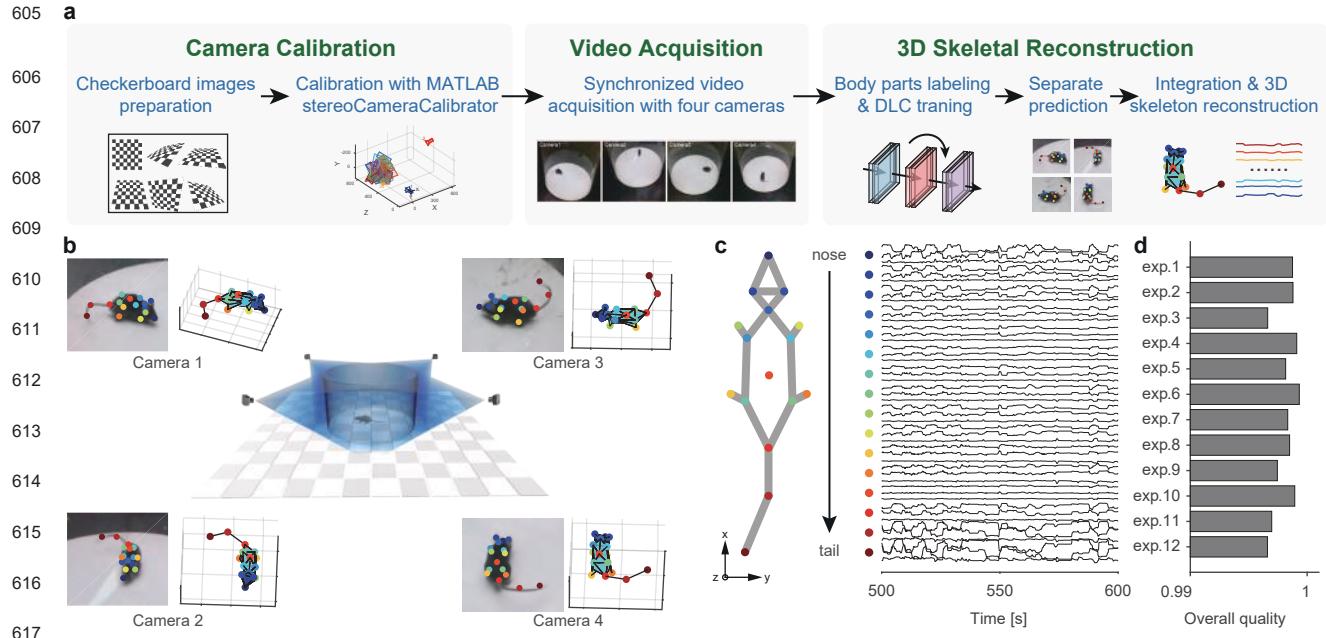
- 562 66. Wei, P. *et al.* Processing of visually evoked innate fear by a non-canonical thalamic pathway.  
563 *Nature Communications* **6**, 1–13 (2015).
- 564 67. Wang, X. *et al.* Synaptic dysfunction and abnormal behaviors in mice lacking major isoforms  
565 of Shank3. *Human Molecular Genetics* **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

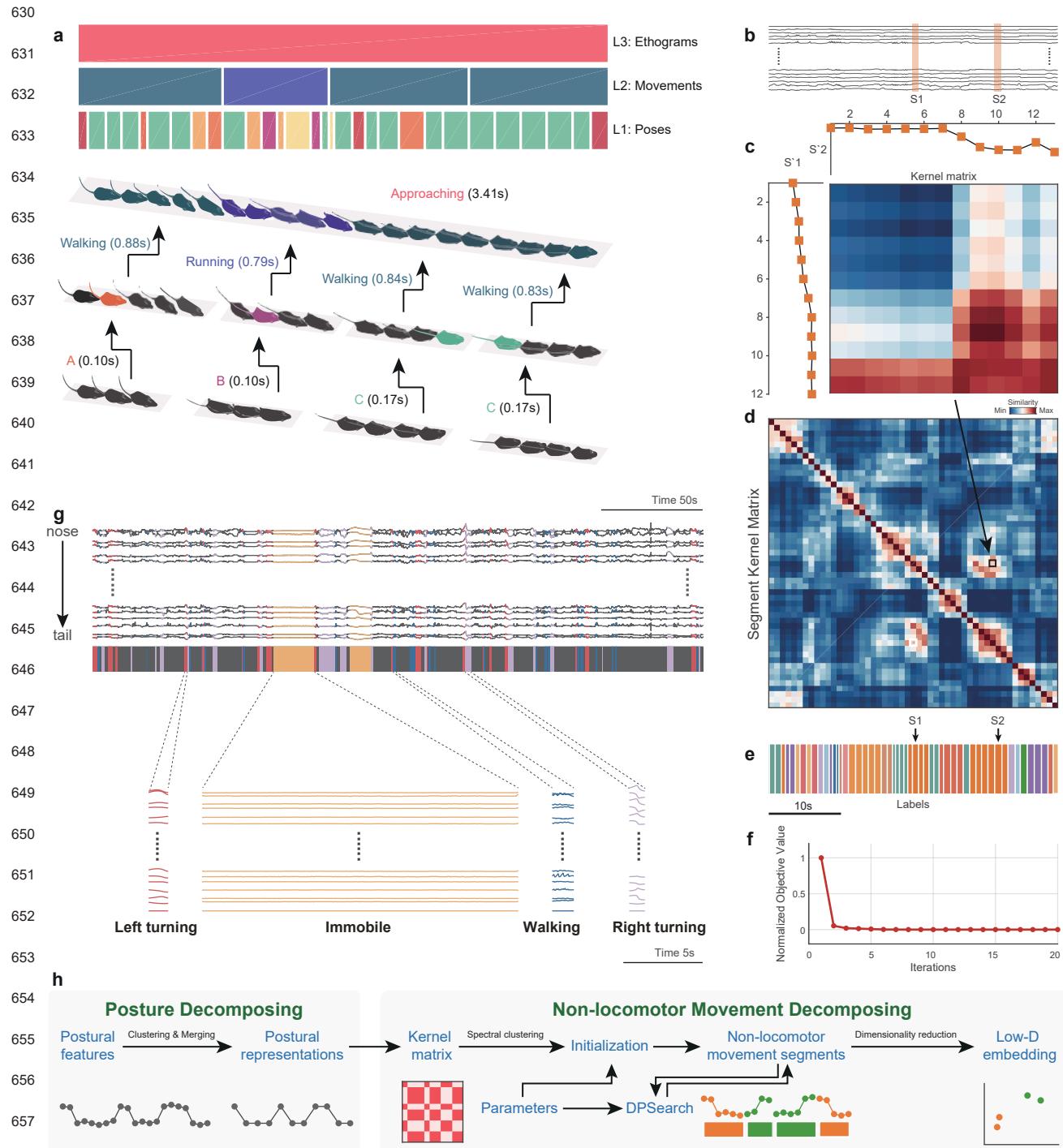
583 **Figures**



592 **Fig. 1 | Hierarchical 3D-motion learning framework for animal behavior analysis.** **a** Data  
593 preparation: 1) image streams captured from four cameras with different 2D views; 2) animal body  
594 parts are tracked to generates separate 2D skeletal trajectories (color-coded traces); 3) reconstructing  
595 3D body skeleton by integrating these four data streams. **b** Two-stage NM decomposition to generate  
596 the feature space: 1) pose decomposition groups continuous skeleton postural data into discrete postural  
597 sequences; 2) NM decomposition, two high-lighted (green and orange) blocks represent two NMs  
598 decomposed from the postural sequences; 3) NM sequences mapped to their 2D features space (right),  
599 where each dot on the 3D axis corresponds to the NM block on the left. **c** Calculation of locomotion  
600 dimension. The continuous velocity of the behaving animal is first calculated, then average the velocity  
601 of each segment obtained in the NM decomposition step. **d** 3D scatter plot represents the combined NM  
602 and locomotion feature space. All the movements are clustered into three types (red, green, and orange  
603 dots) with the unsupervised approach.

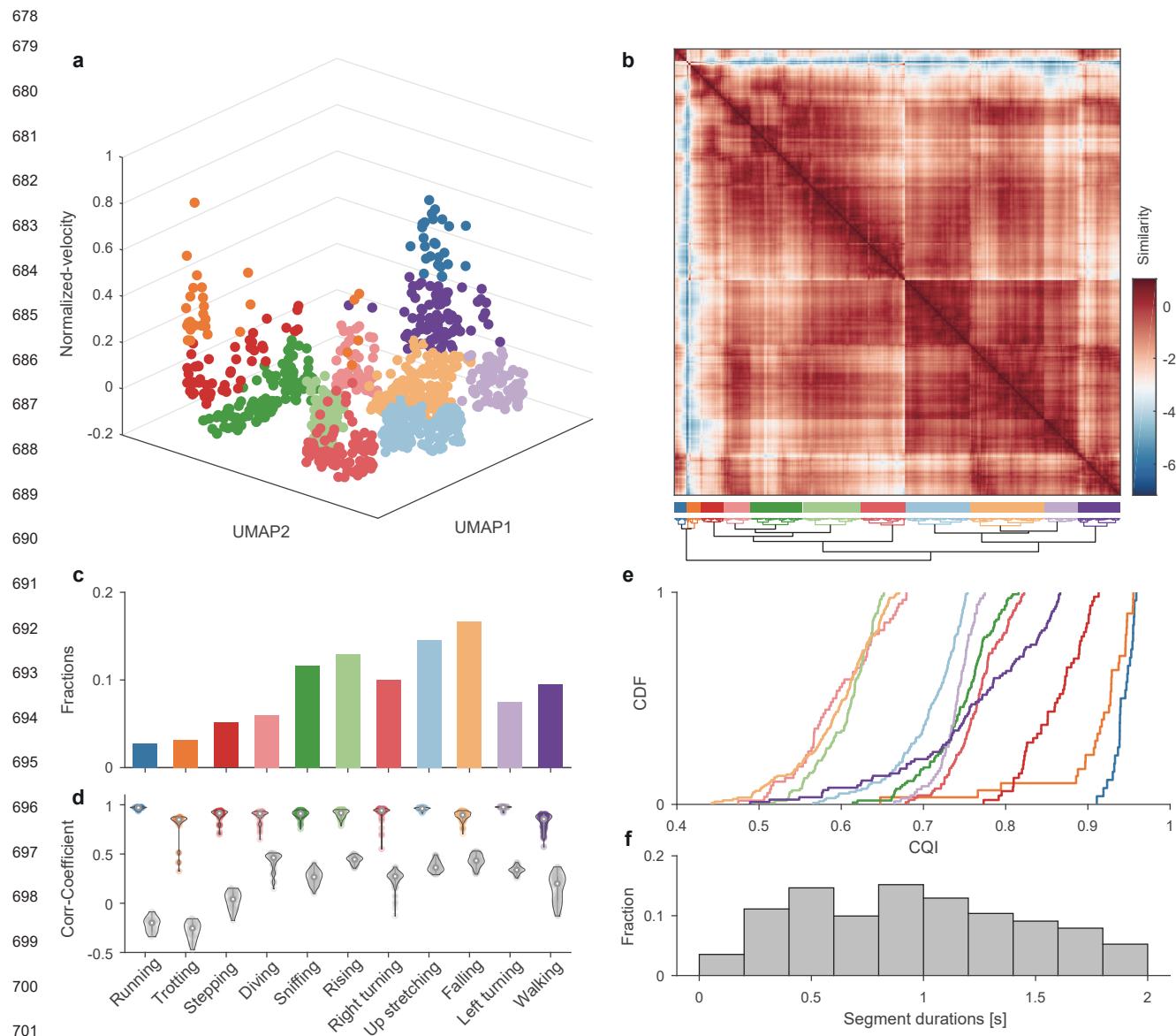
604





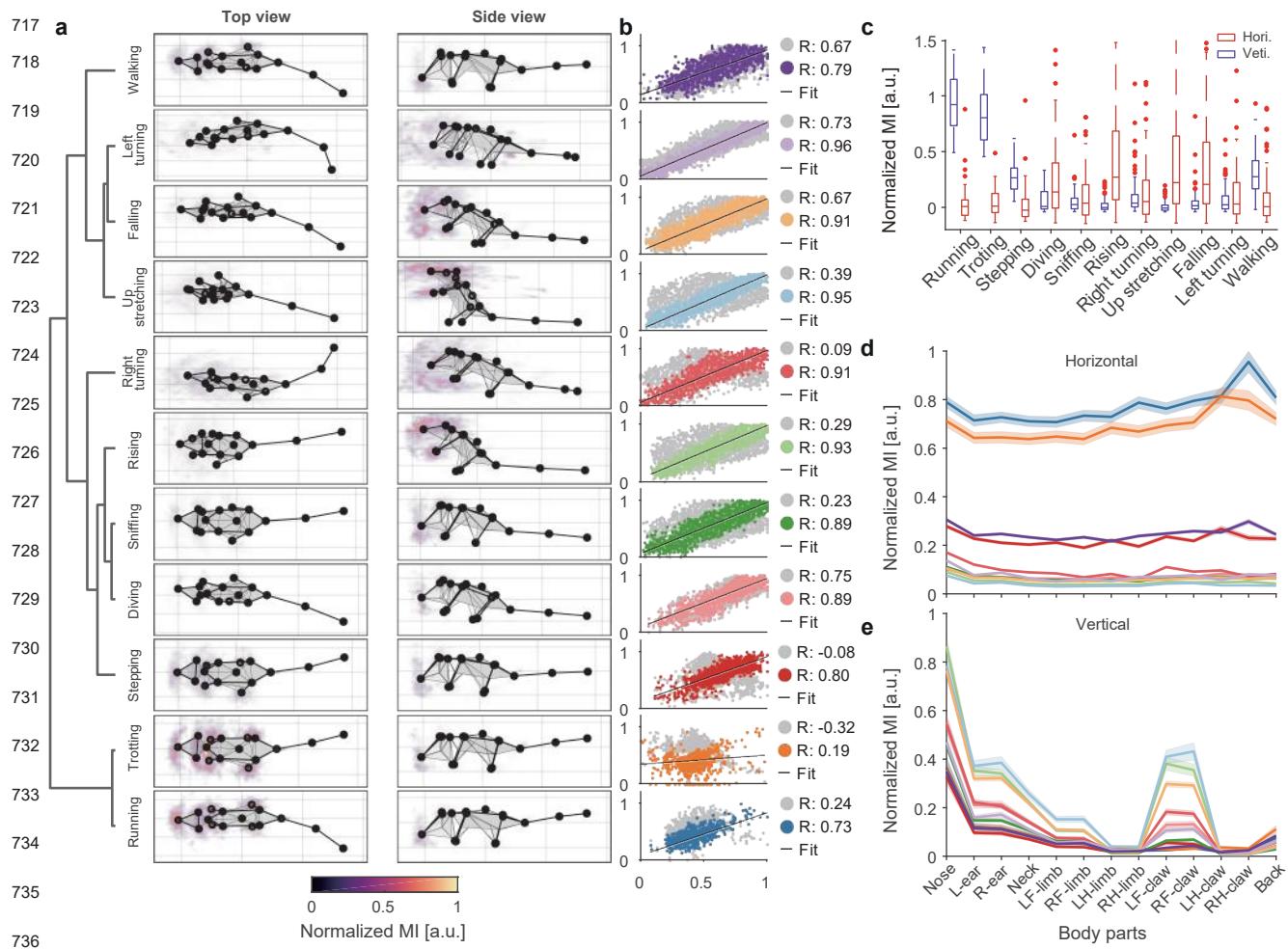
**Fig. 3 | Dynamic temporal decomposition of multi-scale hierarchical behavior.** **a** Illustration of the three-layer bottom-up architecture for behavior. Top: The color-coded bars indicate the types of behavior components in the corresponding time period at that layer; each upper layer component is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

668 normalized similarity value of the  $K$  for a pair of segments at the  $i^{th}$  row and the  $j^{th}$  column (e.g., the  
669 pixel in the black box indicates the final similarity of  $S1$  and  $S2$ ). **e** NM segments decomposed from the  
670 postural trajectories shown in **b** and their color-coded labels. Segments with the same color indicate  
671 that they belong to the same types due to their higher similarity. **f** Optimization process of dynamic  
672 temporal decomposition. Objective Value (OV) error decreases with each iteration until the termination  
673 condition is reached (maximum number of iterations or OV converges). **g** Top, representative 300-s  
674 skeletal traces, where the trace slices highlighted in colors corresponding to the four types of typical  
675 NMs (left turn, immobile, walk, right turn). Bottom, magnification of representative traces of these four  
676 movement types. **h** Workflow of the two-stage behavioral decomposition.  
677



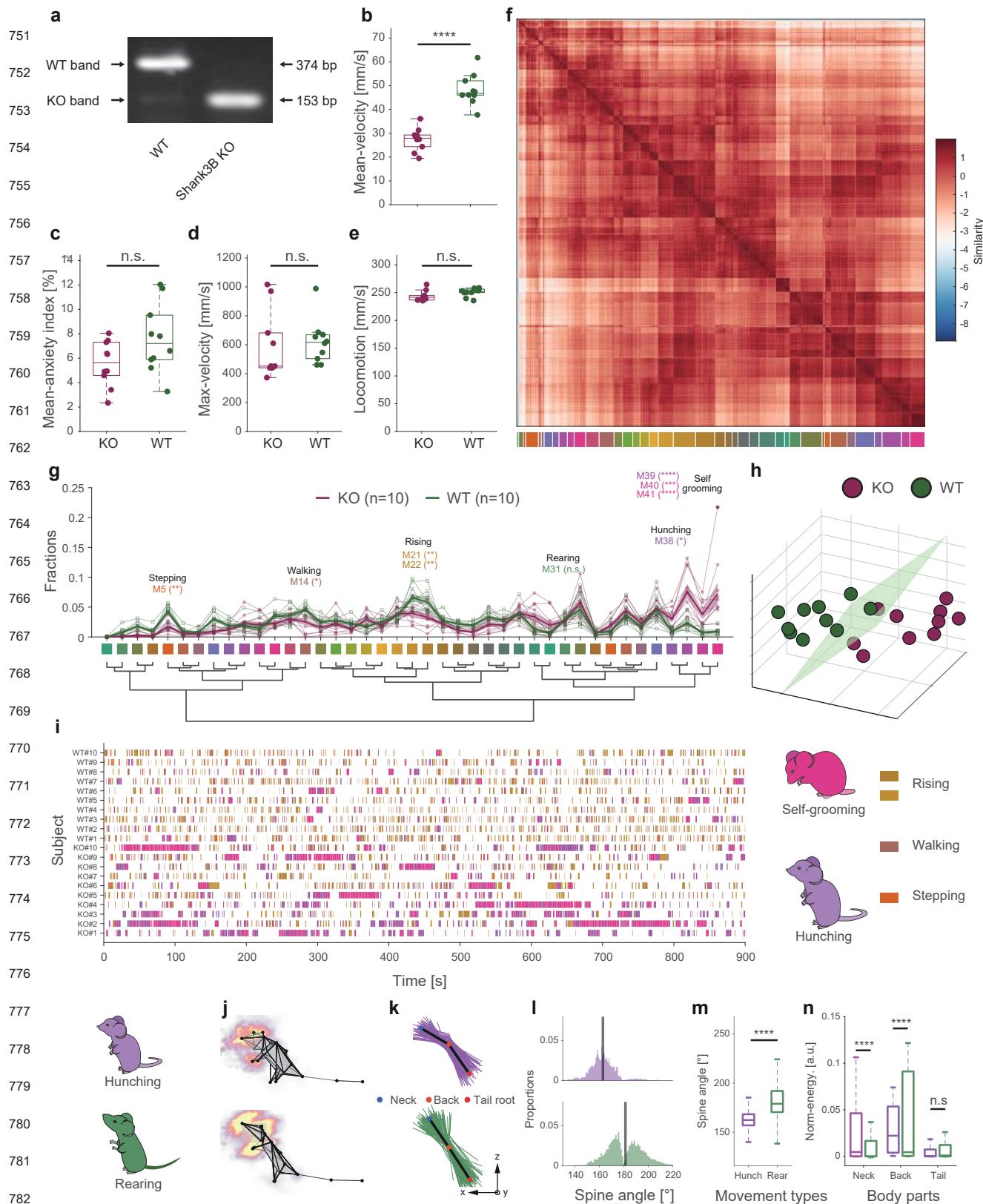
**Fig. 4 | Identify movement phenotypes on single experimental data.** **a** Spatio-temporal feature space of behavioral components. Each dot on the 3D scatter plot represents a movement bout ( $n = 935$  bouts). The 11 different colors indicate the corresponding to 11 movement types shown in **d**. **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  $i^{th}$  row and the  $j^{th}$  column. The color-coded bars indicate the labels of clustered movement (middle). **c** Fractions of movement bouts number. 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:  $0.91 \pm 0.07$ ; Inter-CC:  $0.29 \pm 0.19$ ). **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 ( $0.963 \pm 0.497$ s).

716



**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  $\pm$  standard deviation (SD).

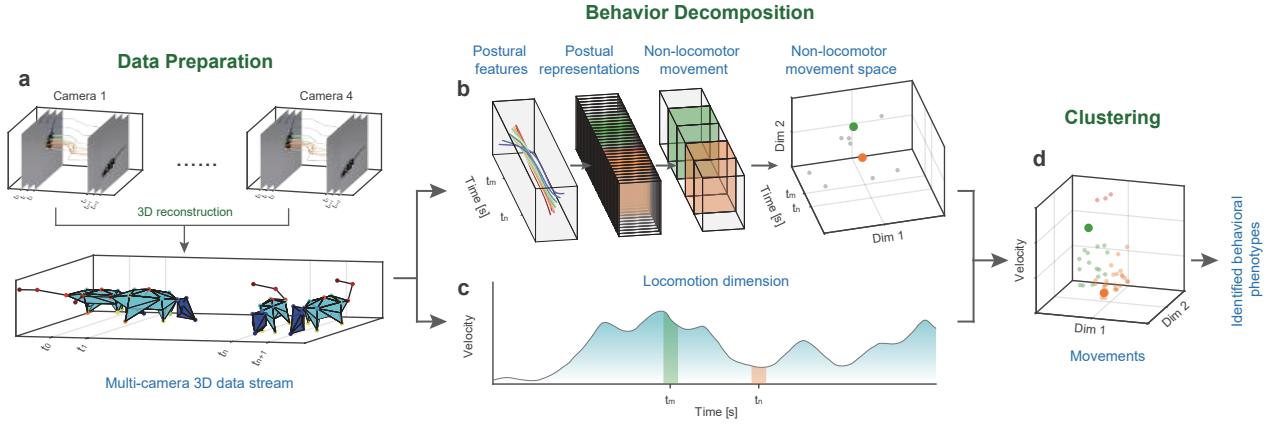
750



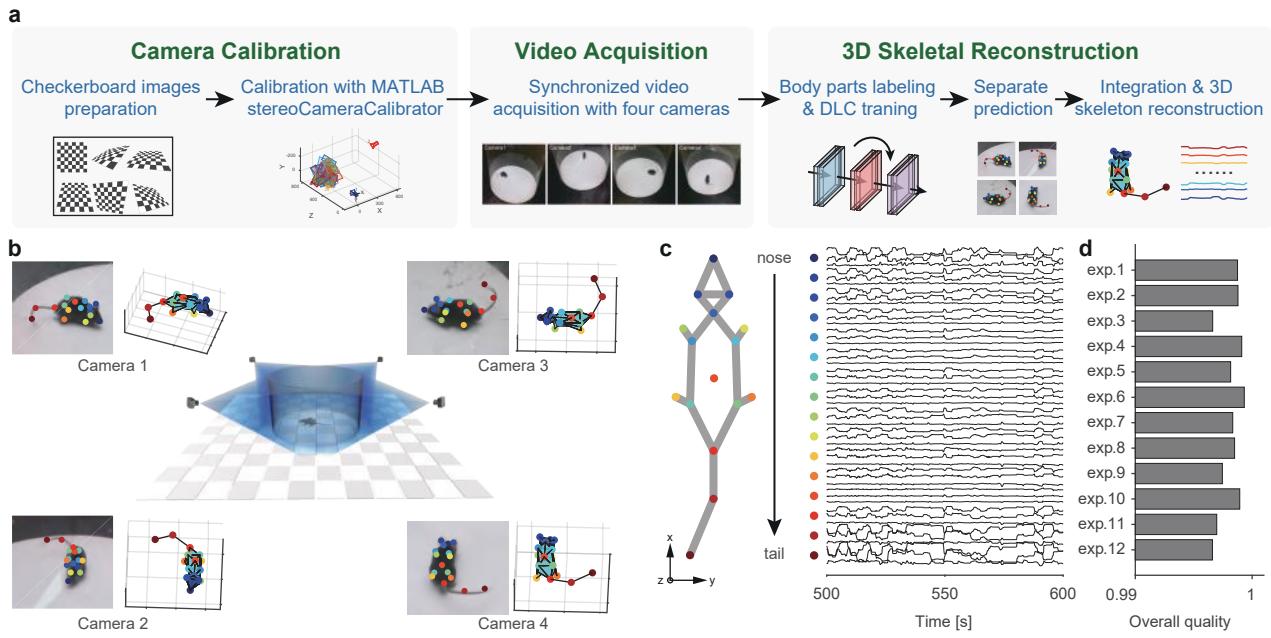
**Fig. 6 | Spontaneous behavior analysis reveals autistic-like behaviors in *Shank3B* knock-out mice.**

**a** PCR genotyping for *Shank3B*<sup>+/+</sup> (Wild Type, WT), *Shank3B*<sup>-/-</sup> (*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

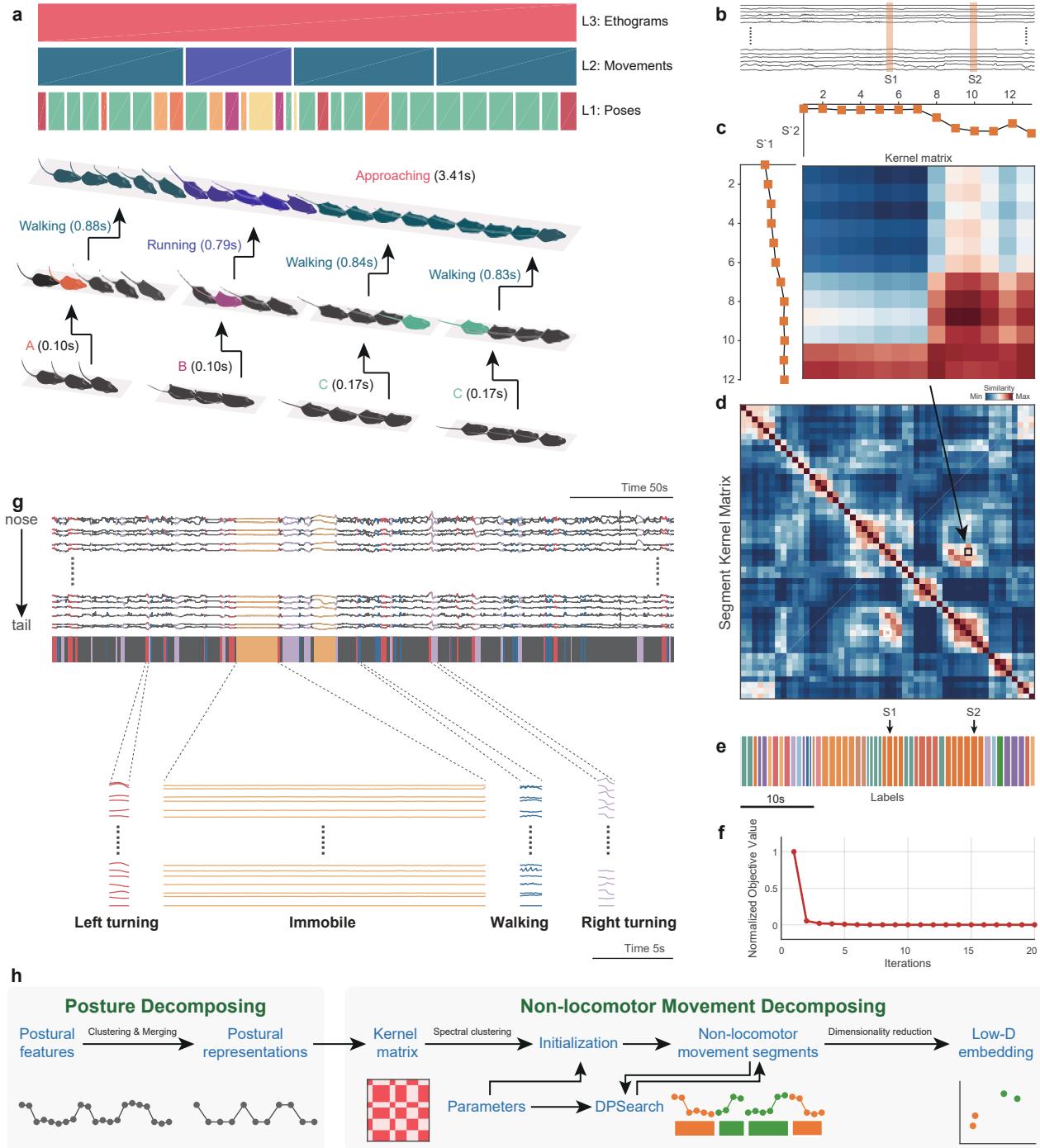
788 paired-wise similarity matrix. The movement bouts of all 20 mice involved were grouped ( $n = 16607$ )  
789 and rearranged in a dendrogram (g). Each pixel on the matrix represents the normalized similarity value  
790 of a pair of movement bouts at the  $i^{th}$  row and the  $j^{th}$  column. The color-coded bars (41 clusters) indicate  
791 the movements being clustered (bottom); **g** Comparison of the fraction of movement types between KO  
792 mice and WT mice. The bold traces and shadows indicate the mean  $\pm$  sem. Fractions of each group and  
793 light color traces are the fractions of all 20 mice (purple, KO,  $n = 10$ ; green, WT,  $n = 10$ ). Middle color-  
794 coded labels and dendrogram indicate the movement types. Eight movements have significant  
795 differences between the two groups, and the fractions of the four movements that KO mice prefer are  
796 hunching (M38, KO =  $3.00 \pm 0.56\%$ , WT =  $0.94 \pm 0.15\%$ ) and self-grooming groups (M39, K =  $7.65$   
797  $\pm 1.21\%$ , W =  $2.34 \pm 0.33\%$ ; M40, K =  $3.73 \pm 0.72\%$ , W =  $0.75 \pm 0.19\%$ ; M41, K =  $7.23 \pm 1.88\%$ , W  
798 =  $0.90 \pm 0.18\%$ ). \*\*\*P < 0.001, \*\*P < 0.01 by two-way ANOVA with Holm–Sidak post-hoc test. **h**  
799 Low-dimensional representation of the two animal groups (purple, KO,  $n = 10$ ; green, WT,  $n = 10$ ). The  
800 20 dots in 3D space were dimensionally reduced from 41-dimensional movement fractions, and they  
801 are well separated. **i** Ethograms of the eight significant movements. **j-n** Kinematic comparison of  
802 rearing and hunching (upper row refers to hunching; lower row refers to rearing). **j** Average skeletons  
803 of all frames and normalized moving intensity (side view) of rearing and hunching. **k** Spine lines (the  
804 lines connecting the neck, back, and tail root) extracted from all frames (rearing, 16834 frames;  
805 hunching, 10037 frames) in movement types. For visualization purposes, only 1% of spine lines are  
806 shown in the figure (rearing, 168/16834; hunching, 100/10037). Black lines refer to the averaged spine  
807 line of the hunching and rearing phenotypes; **l** Histograms of the spine angles (angle between three  
808 body parts). During rearing, the spine angles of the animals swing, and the average spine angle is  
809 straight ( $181.34 \pm 15.63^\circ$ ). By contrast, the spine angles of the rodents during hunching are consistently  
810 arcuate ( $162.88 \pm 10.08^\circ$ ). **m, n** Box plot of spine angles of the two movement types. **n** Box plot of  
811 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.

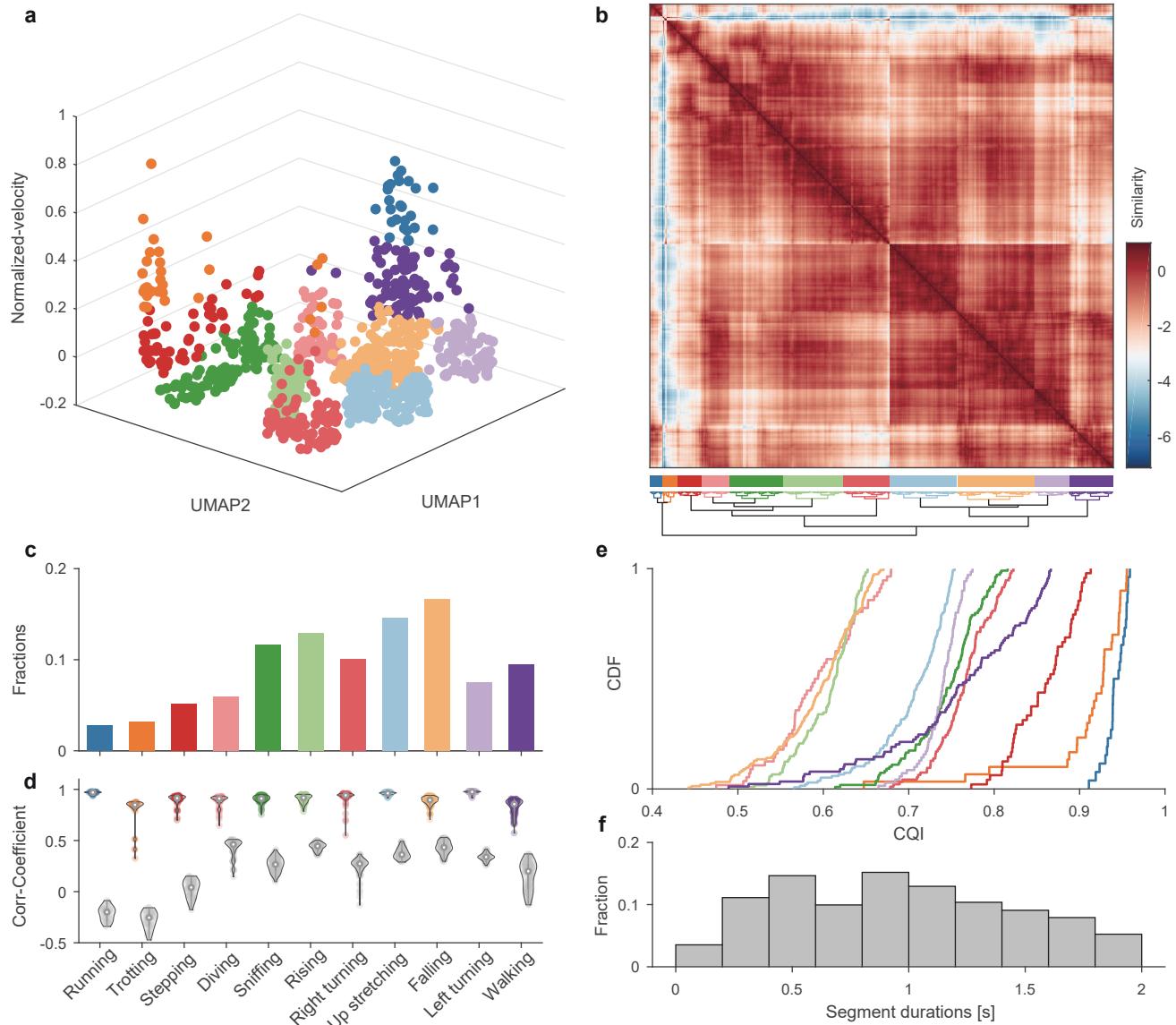


**Fig. 2 | Collecting animal behavior trajectories via a 3D motion capture system. a** Pipeline of 3D animal skeletal reconstruction. **b** Center, schematic diagram of recording animal behavior with four synchronized cameras; corners, frames captured by the cameras with the DLC labels (left) and the 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  $0.9981 \pm 0.001$ .

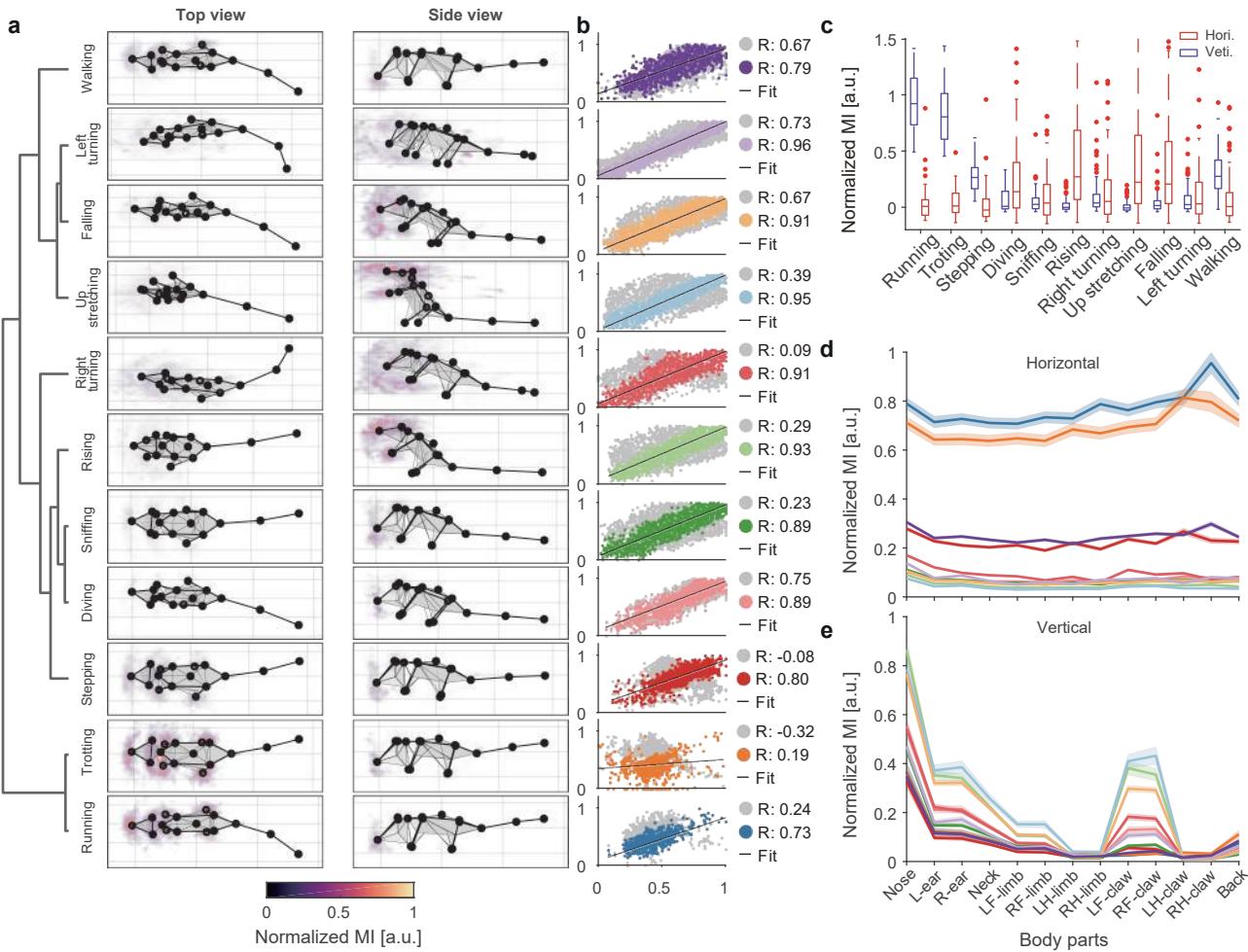


**Fig. 3 | Dynamic temporal decomposition of multi-scale hierarchical behavior.** **a** Illustration of the three-layer bottom-up architecture for behavior. Top: The color-coded bars indicate the types of behavior components in the corresponding time period at that layer; each upper layer component is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

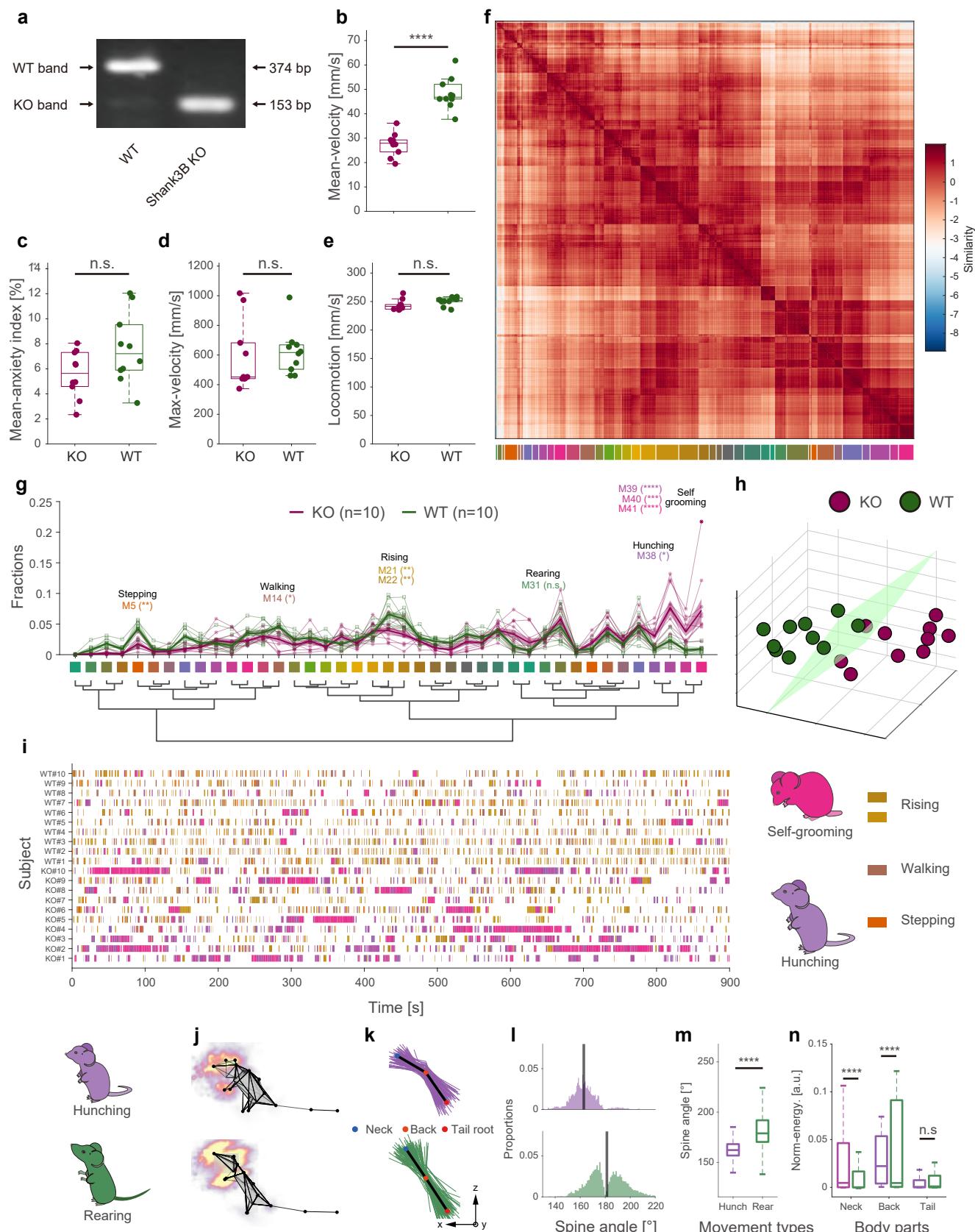
matrix represents the normalized similarity value of the  $K$  for a pair of segments at the  $i^{th}$  row and the  $j^{th}$  column (e.g., the pixel in the black box indicates the final similarity of  $S1$  and  $S2$ ). **e** NM segments decomposed from the postural trajectories shown in **b** and 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.



**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  $i^{th}$  row and the  $j^{th}$  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:  $0.91 \pm 0.07$ ; Inter-CC:  $0.29 \pm 0.19$ ). **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 ( $0.963 \pm 0.497$ s).



**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  $\pm$  standard deviation (SD).



**Fig. 6 | Spontaneous behavior analysis reveals autistic-like behaviors on shank3B knock-out mice.** **a** PCR genotyping for *Shank3B*<sup>+/+</sup>(Wild Type, WT), *Shank3B*<sup>-/-</sup>(*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 *i*<sup>th</sup> row and the *j*<sup>th</sup> 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 $\pm$ 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 $\pm$ 0.56%, WT 0.94 $\pm$ 0.15%) and self-grooming groups (M39: KO 7.65 $\pm$ 1.21%, WT 2.34 $\pm$ 0.33%; M40: KO 3.73 $\pm$ 0.72%, WT 0.75 $\pm$ 0.19%; M41: KO 7.23 $\pm$ 1.88%, WT 0.90 $\pm$ 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 $\pm$ 15.63°). By contrast, the spine angles of the rodents during hunching are consistently arcuate (162.88 $\pm$ 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.