Cortex-Wide Substrates for Body Schema and Action Awareness

Kyle S. Severson[1,*], Jinghao Lu[1], Wenxi Xiao[1], He Jiang[1], Ting Lou[1], Kian A. Caplan[1], Manuel S. Levy[1], Tianqing Li[2], Timothy W. Dunn[2] , Fan Wang[1,*]

[1]Department of Brain and Cognitive Sciences, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Biomedical Engineering, Duke University, Durham, NC, USA.
*Corresponding authors. e-mail: kyleseve@mit.edu; fan_wang@mit.edu

Abstract

The body schema, an internal model of the body's configuration in space, is essential for embodied perception and voluntary action. Cortical lesions can disrupt the body schema; however, it remains unclear how the cortex efficiently represents the vast repertoire of body postures during natural behavior. Here, we unveil the representational format underlying the body schema by combining large-scale electrophysiology with 3D full-body motion capture in freely moving mice. Across posterior parietal (PPC) and sensorimotor cortices (SMC), neurons encoded diverse postural features in distinct reference frames: PPC neurons emphasize midline-referenced features related to left-right turning, whereas SMC neurons encode body elevation and pitch relative to the ground, revealing a gravicentric reference frame. Strikingly, beyond static posture encoding, we identified neural ensembles selectively activated during stereotyped actions such as walking, rearing, and grooming. These action ensembles fire in phase-locked sequences that tile the entire action cycle, providing a substrate for action awareness. A network model recapitulated these dynamics, showing that population activity clustered into action-specific subspaces. Together, these findings demonstrate that the cortex organizes the body schema around actions, compressing high-dimensional kinematics into an efficient, gravity-anchored, and action-aware neural code.

Introduction

In the early 20th century, Head and Holmes (1911) studied patients with cortical lesions and discovered that the most common deficits involved an inability to localize body parts in space. Patients also frequently failed to perceive passive movements, for example, when an experimenter moved the patient's limb without their volition. These observations bore the concept of the "body schema": an internal model of the body's posture in three-dimensional space. Subsequent research showed that cortical lesions, especially to the posterior parietal cortex (PPC) in the right hemisphere, disrupted the body schema and led to numerous neurological symptoms, including impaired awareness of movement and delusional misattributions in which patients perceive their body parts as belonging to someone else (Bermúdez, 1995; Maravita et al., 2003; Haggard & Wolpert, 2004). Despite its central role in body perception, embodiment, and

action, the neural basis of the body schema, and specifically its representational format in the brain, remains poorly understood.

Understanding the neural basis of the body schema requires addressing three fundamental questions. The first concerns dimensionality. In vertebrates, postural changes occur at the joints, and mammalian bodies are composed of many joints, most allowing multiple degrees of freedom. This results in a vast, high-dimensional posture space, comprising all possible combinations of joint angles. Proprioceptive afferents innervating muscles, tendons, and joints encode this high-dimensional joint space and transmit detailed, local proprioceptive information to the cortex (Tsay et al., 2016). Does the brain maintain this high-dimensional representation to construct the body schema, or does it encode a more efficient, low-dimensional format?

The second question concerns the reference frame of posture representations. By definition, posture is the configuration of the body relative to gravity (Berthoz, 1997; Mittelstaedt, 1997). For example, standing upright and lying flat on the ground may involve nearly identical joint angles in the limbs and torso but are perceived and function as distinct postures due to their orientation with respect to gravity. Gravity thus provides a stable, external reference frame for the body schema. Sensory systems including the vestibular apparatus, proprioception, and vision together offer reliable cues for inferring the direction of gravity (Angelaki & Cullen, 2008; Chesler et al., 2016). How does the brain incorporate these multisensory signals to anchor the body schema in a gravity-centered reference frame?

The third closely related question concerns the format of the body schema in the context of dynamic actions. Unlike static postures, actions involve coordinated changes in the body configuration across the temporal dimension. Yet, we do not consciously perceive these changes on a millisecond timescale. Instead, actions are typically experienced as discrete modes, such as running or squatting. How does the body schema incorporate dynamic postural changes during actions? Answering these fundamental questions is critical for understanding how the brain supports motor control and action awareness.

To begin addressing these questions and investigating the neural basis of the body schema, we examined cortical representations of full-body posture in freely moving mice, which naturally exhibit a rich and diverse repertoire of body configurations during spontaneous behaviors. We combined large-scale extracellular recordings with 3D full-body posture tracking to relate cortical activity with naturalistic postural dynamics. Our primary focus was on the posterior parietal cortex (PPC) and the primary and secondary somatosensory cortices (S1 and S2), given their putative roles in body awareness and proprioceptive integration.

The PPC has been implicated in awareness of the body and voluntary actions. In humans, lesions can cause asomatognosia (loss of awareness of body parts), hemi-spatial neglect (failure to attend to one side of the body), and deficits in recognizing self-generated movements (MacDonald et al., 2003; Sirigu et al., 2004; Vallar & Ronchi, 2009; Vallar & Calzolari, 2018). Electrical stimulation studies further suggest a causal role, with PPC stimulation evoking subjective intentions to move and illusory movement without overt muscle activity at higher stimulation

intensities (Desmurget et al., 2009). In rodents, PPC encodes lateralized turning during navigation (Whitlock et al., 2012) and efficiently encodes head and trunk posture during natural behavior (Mimica et al., 2018), though it remains an open question whether and how PPC represents the entire body, including the tail and limb appendages.

Primary (S1) and secondary (S2) somatosensory cortices are also strong candidates for encoding body schema, particularly given their roles in processing proprioceptive input. The formation and maintenance of the body schema are thought to depend on multisensory integration, combining proprioceptive, tactile, vestibular, and visual inputs. Among these, proprioceptive feedback plays a central role in signaling joint angles and postural changes (Roll et al., 1991; Maravita et al., 2003; Tsay et al., 2016). In anesthetized rodents, muscle- and joint-responsive neurons have been identified preferentially in the agranular cortex (primary motor cortex, M1), in proprioceptive areas of S1, as well as in S2, where neurons respond to tendon afferents spanning the entire limb (Chapin & Lin, 1984; Lee & Kim, 2012; Marasco et al., 2017). In awake animals, however, proprioceptive responses are also distributed across S1 in both primates and rodents (Prud'homme & Kalaska, 1994; Alonso et al., 2023).

To determine the respective contributions of cortical regions to the body schema representation, we performed chronic in vivo recordings from PPC and from S2 and S1, including the perigranular (S1pgz), dysgranular (S1dz), and S1-M1 transition (S1tz) zones. For comparison, we also recorded from M1, primary visual cortex (V1), and hippocampus (HPC). Across 20 mice and 392 one-hour sessions, we examined how different cortical regions encode the body schema during natural behavior.

Results

Full-body posture tracking and joint angle computations

To investigate how the cortex encodes the body schema during natural behavior, we developed a high-resolution multi-camera recording system for tracking full-body posture in freely moving mice. Spontaneous behavior was recorded in a custom-built motion capture arena equipped with six synchronized, calibrated cameras, enabling high-precision, non-invasive 3D tracking (Fig. 1a). We defined a comprehensive skeletal model with 26 anatomically grounded keypoints (Extended Data Fig. 1a,b) which were used to train a DANNCE model (Dunn & Marshall et al., 2021) for markerless 3D pose estimation. The network trained on 775 human-annotated frames (Fig. 1b,d) achieved a mean Euclidean error of 2.38 mm on held-out test frames (Fig. 1c), and generalized robustly across millions of frames. All 3D keypoint predictions were calibrated to the center of the arena floor, representing continuous x-, y-, and z-coordinate time series for each keypoint, with z measuring absolute height from the floor in world coordinates.

To derive interpretable measures of head and body orientation, we grouped these keypoints into 17 anatomically defined body segments spanning the head, trunk, limbs, and tail (Extended Data

Fig. 1c,d). To compute the 3D positions and orientations of these body segments, we defined each segment as plane including of three adjacent keypoints (e.g., snout and ears for the head) and decomposed world-centered segment rotations into yaw, pitch, and roll Euler angles (Extended Data Fig. 1e). Yaw represents allocentric (world-centered) heading, while pitch and roll capture segment orientation relative to gravity.

Because the size and shape of body segments in adult mice remain essentially constant (Extended Data Fig. 1d), postural changes arise from joint rotations driven by muscle activity (Monsees et al., 2022). Rotations at the joints and head thus serve as biomechanical correlates of proprioceptive and vestibular input—key components from which the brain constructs the body schema. We constructed a joint-centered geometric model of the skeleton by defining 16 joints connecting adjacent body segments (Extended Data Fig. 1g,h). Each joint was assigned up to three rotational degrees of freedom to parameterize full-body posture (Fig. 1e). Ball-and-socket joints (e.g., shoulder, hip) allow flexion-extension (*fe*), abduction-adduction (*aa*), and internal-external rotation (*ie*), while hinge joints (e.g., elbow, knee) support only flexion-extension. This joint-centered geometric skeletal model provided the basis for the parameterization of full-body posture.

Using our geometric skeletal model, we computed 38 joint angles as Euler rotations of each joint's distal segment relative to its proximal segment (Extended Data Fig. 1f,h; See Methods). These joint angles, along with head pitch and roll, compress the posture parameter space into 40 interpretable dimensions that capture the body configuration at each time point. When enriched with derived postural dynamic features such as joint velocities, our framework forms the foundation for linking cortical activity to detailed full-body kinematics in freely behaving mice.

PPC preferentially encodes postural angles of midline body parts

To examine how cortical neurons encode body posture, we first performed extracellular recordings from the medial PPC, posterior to the S1 trunk region (Fig. 2a,b), using chronic tetrode microdrives (N=4 mice). In total, we isolated 996 single units sampled across all cortical depths, while simultaneously acquiring behavioral video for automated full-body tracking (61 sessions, totaling 3,660 minutes). To first examine how the activity of PPC neurons correlates with posture, we aligned spike rasters with height traces of snout, forepaws, and hindpaws, as well as traces from all 40 joint and head angles (Fig. 2d for reference on joint angle definitions). We observed that neural activity was modulated by changes in posture, including shifts in body height, head pitch, and joint angles (e.g., Fig 2c). To systematically quantify posture tuning, we computed tuning curves for every neuron to 64 posture features, including joint angles, body segment orientations, and height of different body parts (see Methods), and compared their relative tuning strength using a variance explained (R^2) as a metric. Among the 751 regular-spiking (RS) PPC neurons, 51.5% (387/751) were tuned to one or more posture features. A heatmap of normalized tuning curves across the posture-tuned population revealed that most neurons preferred extreme angles (Fig. 2e), confirming efficient coding (Barlow, 1961) of posture angles in PPC (Mimica et al., 2018).

To understand which posture features were most strongly represented in PPC, we examined the fraction of neurons best-tuned to each individual postural feature. Consistent with prior work (Mimica et al., 2018), we observed prevalent tuning to head, neck, and trunk angles (Fig. 2f-k). Furthermore, our full-body tracking revealed that while relatively few PPC neurons were tuned to limb-related features (58/461, 12.6%), surprisingly, a substantial fraction (139/461, 30.1%) were tuned to tail angles (Fig. 2g). In total, 87.4% (403/461) of posture-tuned PPC neurons encoded postural angles of midline body parts, such as the neck, head, spine, and tail (e.g., Fig 2l,m). Notably, across the PPC population we found an overrepresentation of posture features reflecting left-right axial trunk rotation, including tail *aa*, tail roll, spine *aa,* and spine *ie* angles, as well as yaw angular velocity of the head or thorax. These observations suggest that PPC preferentially represents postural deviations from the body midline, particularly those associated with turning.
To test this idea, we asked whether PPC tuning could be captured in lower-dimensional postural axes. We applied principal component analysis (PCA) to identify the "eigen-poses" that capture the dominant axes of the body configuration (see Methods). The first six Posture PCs explained 61% of the variance in full-body posture (Fig. 2o). We visualized the extreme bins of each Posture PCs dimension as averaged 3D poses, revealing that the top three Posture PCs represented orthogonal posture dimensions: vertical elevation (PC1), forward–backward leaning (PC2), and left–right turning (PC3, Fig. 2n). We then computed the tuning of PPC neurons to each posture PC. Remarkably, 51% of posture-tuned neurons were preferentially tuned to PC3 (Fig. 2p), confirming a strong population-level preference for left-right turning (Fig. 2f), consistent with our tuning analysis and prior work in rats (Whitlock et al., 2012). Fig. 2q,r illustrate an example PC3-tuned PPC neuron that selectively increased firing during leftward but not rightward turns.

Together, these results demonstrate that PPC efficiently encodes head, trunk, and tail posture angles, with a strong bias toward turning postures. Posture deviations from the midline facilitating left-right turning appear to serve as a key reference axis for the body schema representation in mouse PPC. This midline-centered encoding may reflect a broader role for PPC in stabilizing posture during turning, coordinating axial body movements, or integrating vestibular inputs to support balance control.

SMC represent posture in a gravicentric reference frame

Because body schema representations likely depend on proprioceptive inputs, we next expanded our recordings to sensorimotor cortices (SMC) and compared posture tuning across multiple cortical areas. Specifically, we targeted S1 (n=4,404 units from 6 mice), including the proprioceptive subregions S1pgz, S1dz, and S1tz (Fig. 3a). For comparison, we conducted additional recordings in S2 (n=971 units from 2 mice), M1 (n=611 units from 3 mice), V1 (n=691 units from 2 mice), and HPC (n=1,077 units from 5 mice).

Similar to our PPC findings, tuning analysis revealed robust posture selectivity across excitatory RS neurons in SMC, but the preferred features differed markedly by region. Tuning analysis revealed prominent posture tuning in RS neurons across SMC (M1: 197/453, 43.5%; S1: 1314/3327, 39.5%) and PPC (401/751, 53.4%). In contrast, posture tuning was relatively sparse in S2 (150/707 neurons, 21.2%) and V1 (90/557 neurons, 16.1%) and minimal in HPC (20/512

neurons, 3.9%). To determine how these different neural populations encoded posture, we extended the Posture PC analysis to all cortical regions. Unlike PPC, which overrepresented left-right turning (PC3; Fig. 3c, right), posture-tuned neurons in SMC were most strongly tuned to PC1 (M1: 88/197, 44.7%; S1: 648/1331, 48.7%; S2: 55/150 neurons, 36.7%; Fig 3b), corresponding to changes in body height relative to the floor (Fig. 3c, left). To directly compare encoding of global postural features (e.g., body segment height) versus local joint configurations (i.e., joint angles), we compared the R^2 of each neuron's best height feature to that of its best joint angle feature. In both S1 and M1, a greater proportion of neurons showed stronger tuning to best height (Fig. 3d,e), whereas PPC neurons exhibited the opposite pattern, with stronger tuning to joint angles (Fig. 3f). Population rasters from an example S1 session (60 simultaneously recorded neurons) illustrate widespread modulation by changes in snout, forepaw, and hindpaw heights, as well as posture angles (Fig.3g). Fig. 3h-k further illustrate example S1 neurons with strong firing rate modulation aligned to the height of the head or thorax. Together, these results indicate that SMC preferentially encodes posture in a gravicentric reference frame, emphasizing body height and pitch relative to gravity rather than fine-grained joint angles.

Fast-spiking neurons robustly encode global body motion

In addition to encoding static postures, many cortical neurons were modulated by dynamic posture changes. While analyzing tuning to kinematic features (i.e., postural features and their derivatives), we found that a substantial fraction of FS putative interneurons across M1, S1, S2, and PPC exhibited symmetric tuning to many joint angular velocities, regardless of direction and joint identity (Fig. 3l). Thus, rather than encoding velocity of specific joints, these neurons were more strongly driven by the average angular speed across all joints, which we refer to as "body motion" (Fig. 3m).

Motion-sensitive FS neurons typically increased their firing rates in proportion to body motion (e.g., Fig. 3o,p; Fig. 3n), although a small subset of motion-sensitive FS and RS neurons showed the opposite pattern, firing more when non-specific body motion was slower (e.g., Fig. 3q,r). Across all cortical regions, a substantial fraction of neurons was best tuned to either body motion or locomotion—collectively classified as "non-specific motion" (M1: 49%, S1: 36%, S2: 39%, PPC: 55%, V1: 46%, HPC: 49%; Fig. 3t). Notably, body motion is distinct from locomotion: while locomotion refers to a change in body location (i.e., X,Y position) across frames, body motion refers to the averaged joint speed. Locomotion and body motion are correlated during walking but become dissociated during stationary actions such as self-grooming. When comparing tuning to locomotion versus body motion, neurons in V1 and HPC were more tuned to locomotion, consistent with prior work (Neill and Stryker, 2010; Sun et al., 2018). In contrast, FS neurons in M1, S1, S2, and PPC showed stronger tuning to body motion (Fig. 3s, right). Together, these findings suggest that putative inhibitory interneurons in SMC and PPC are specialized for encoding body motion speed, complimenting the posture-selective tuning of RS excitatory neurons.

Posture is the principal encoded feature category in SMC

Having established that RS neurons in SMC preferentially encode posture while FS neurons encode body motion, we next asked how these feature preferences distribute across categories. We computed tuning strength for each neuron and compiled a histogram of preferred feature tuning (Fig. 3s). SMC contained large proportions of neurons best-tuned to body height features, especially head height. Specifically, tuning to forepaw and thorax height was more pronounced in S1 than in S2 or M1. Across SMC, head pitch was also strongly represented, especially in S2. Next, we assigned each neuron into six broader functional categories based on their preferred tuning feature: Body Posture, Head Angles, Allocentric, Paw Touch, Postural Dynamics, and Non-specific Motion. Histograms comparing the distribution of best-tuned features across regions revealed that Body Posture, comprised of joint angles, segment heights, and posture PCs, was the dominant tuning category in M1 (62.1%), S1 (61.8%), S2 (43.5%), and PPC (68.8%, Fig. 3t). Surprisingly, only a small fraction of RS neurons in SMC (6-10%) were best tuned to Postural Dynamics (i.e., derivatives of joint angles or height), suggesting that dynamic features were less emphasized than static postural configurations. Few neurons in SMC (5-8%) were categorized as encoding Paw Touch, derived from the proximity of the paws to the floor, walls, or head. In V1, Head Angle features, especially head pitch and head yaw velocity (head turning), were strongly preferred by the RS population (67.8%), consistent with integration of head orientation (Mimica et al., 2023). As expected, the majority of RS neurons in HPC were Allocentric (84.5%), preferring “place” above all other features (Fig. 3s, left).

Across all cortical regions and cell types, head- and trunk-related kinematic features were overrepresented (Fig. 3u). Tail features were more strongly represented in PPC than in M1, S1, or S2, while head-related kinematics were overrepresented in V1. Surprisingly, only a modest fraction of S1 neurons (23.0%) showed preferential tuning to limb-related features, despite the fact that we targeted S1 subregions known to receive primary proprioceptive input from the limbs (Chapin & Lin, 1984). These findings suggest that, rather than encoding high-dimensional, primitive proprioceptive inputs, SMC compresses posture into a low-dimensional representation as body height and body angle (pitch) relative to gravity. This gravicentric posture representation, referenced relative to the ground and aligned with the gravitational vector, provides a stable and ethologically relevant reference frame for encoding the body schema. Moreover, through regional and cell-type specialization, the cortex maintains distinct but complementary representations: RS neurons in SMC encode posture as height and pitch, RS neurons in PPC encode midline posture as roll and left–right turning, and FS interneurons encode non-specific motion. Together, these complementary representations support a moment-by-moment, low-dimensional body schema distributed across specialized cortical subnetworks.

Prominent action ensembles in the cortex

Continuous representations of the body configuration may be metabolically costly and computationally inefficient. Yet, animals do not navigate the world by consciously evaluating snapshots of every posture. Instead, animals compose spontaneous behaviors from a repertoire of species-typical actions (Marshall et al., 2021) (e.g., walking, grooming, rearing) that are stereotypical and repeated routinely throughout their lives. For example, during walking, we do

not track joint angles across each footstep; we simply lean forward and initiate a 'walking' sensorimotor program. We hypothesized that the brain similarly employs a lower-dimensional, action-level representation that abstracts away the finer details of postural configurations. To test this, we examined how neuronal activity was modulated by actions, define here as dynamic, stereotyped movement trajectories centered around distinctive full-body postures. In our behavioral arena, mice primarily performed nine common actions: Walk, Rear, Crouch, Sniff, Head Groom, Body Groom, Left or Right Scratch, and Rest (Fig. 4a). We annotated example periods with these categorical labels and trained a supervised action segmentation model, implemented as an encoder-decoder temporal convolutional network (AS-EDTCN; based on Lea et al., 2017), to classify actions from kinematic data (Extended Data Fig. 2a-e; see Methods). The model achieved high accuracy (81%) and aligned more closely with human annotations than an unsupervised model (Weinreb et al., 2024). Segmented actions coincided with widespread neuronal activations, revealing large clusters of neurons, sorted via Rastermap (Stringer et al., 2025), that were selectively active during specific actions, such as Sniff or Groom in S1 (e.g., Fig. 4b) and Rear or Scratch in PPC (e.g., Fig. 4c). These observations suggested that distinct groups of neurons are selectively engaged during specific actions, forming what we term "action ensembles".

Building on these observations, we next sought to identify neurons whose firing rates were significantly modulated by actions. To identify action ensembles, we used a simple threshold on the mean Z-scored modulation across actions (Fig. 4d; see Methods). This straightforward classification approach revealed substantial action ensembles across cortical areas, especially for Rear, Groom, and Scratch (Fig. 4d). Remarkably, SMC and PPC contained large proportions of strongly action-tuned neurons (M1: 70.2%, S1: 60.4%, S2: 55.4%, PPC: 70.3%), whereas V1 and HPC contained smaller fractions (V1: 27.2%, HPC: 26.7%) with weaker action tuning. Moreover, neurons within each action ensemble exhibited kinematic tuning profiles that aligned with the defining features of the corresponding action. For example, neurons in the Walk ensemble were tuned to locomotion speed, Rear ensemble neurons were selective for head pitch and head height, and Scratch ensemble neurons were tuned to hindpaw height (Extended Data Fig. 4), reflecting the dominant kinematic features associated with each action.

Having observed prominent neural ensembles representing each of these nine common actions, we next asked whether individual neurons were dedicated to a single action or contributed to multiple action ensembles. By our definition, neurons could participate in more than one ensemble. To examine this overlap, we computed a matrix of shared ensemble membership across all action pairs (Fig. 4e). Overlap was most pronounced among Head Groom, Body Groom, and Scratch ensembles, likely reflecting shared postures and movement trajectories. On average, each pair of ensembles overlapped by less than 14% in RS neurons (Fig. 4f). In contrast, FS neurons, were more promiscuous, with ~20% overlap across ensembles. Nevertheless, 26–50% of neurons within each action ensemble were exclusive to that action across cortical areas (Fig. 4g). These findings suggest that stereotyped actions are represented by unique, but partially overlapping, cortical subnetworks.

We next asked how this Specific Actions feature class compared to other kinematic feature classes in driving neuronal selectivity. Incorporating Specific Actions into the preferred feature analysis revealed that sizeable fractions of RS neurons in SMC and PPC (M1: 20.7%; S1: 23.5%; S2: 28.3%; PPC: 19.6%) now showed strongest tuning to discrete action categories compared to continuous kinematic features in other classes (Fig. 4h, left). Nonetheless, the largest fraction of RS neurons remained tuned to Body Posture (Fig. 4h, left). For FS neurons, although a notable proportion now preferred Specific Actions, the majority across cortical areas continued to be best tuned to Non-specific Motion (Fig. 4h, right). Together, these results demonstrate that cortical populations form distinct action ensembles, with RS neurons preferentially encoding specific actions and postures and FS neurons contributing broader, motion-related modulation across actions. These prominent action ensembles may serve as a neural substrate for action awareness.

Action ensembles encode phase as neural sequences

Each action unfolds as a characteristic sequence of postural transitions. For example, Rear is a ballistic, single-cycle action in which the animal lifts its forepaws, rapidly extends its hindlimbs to reach peak height, then descends back into a crouched or prone position (Fig. 5a). In contrast, Head Groom consists of repeated, stereotyped cycles in which the mouse extends its forepaws from behind the ears toward the snout, then retracts them before initiating the next cycle (Fig. 5b). Actions like Rear, Head Groom, and Walk are therefore inherently phasic, consisting of stereotypical sequences of postures, whereas Rest and Crouch involve relatively static postures. To determine whether these phasic action dynamics are reflected in cortical activity, we defined the relative position within each action cycle as action phase (see Methods). This allows for phase alignment across repetitions of the same action, even if executed with varying speeds or durations (e.g., fast versus slow Rears, or different Walking gaits). When neural activity was aligned to action phase, striking temporal structure emerged: neurons within a given action ensemble consistently fired at the same phase across repetitions of that action (e.g., Fig. 5c,d; see also Supplemental Videos).

To quantify this phase specific firing pattern, we computed peri-event time histograms (PETHs) aligned to action phase onset and offset. Across repetitions, individual neurons showed highly consistent modulation at specific action phase (Fig. 5e,f). Remarkably, when phase tuning curves from all neurons in an action ensemble were pooled across brain regions, they formed a smooth sequence across the population that tiled the full action phase cycle (Fig. 5g; see region-specific heatmaps in Extended Data Fig. 5). Similarly, neurons that were suppressed during an action exhibited phase-specific suppression sequences, suggesting coordinated inhibition (Extended Data Fig. 5).

To further validate these patterns, we quantified phase modulation by centering each neuron's activity on its preferred phase and computing the Z-scored average across neurons within each action ensemble. This analysis, performed separately for activated and suppressed populations,

revealed strong modulation around preferred phase across each action ensemble (Fig. 5h). In contrast, phase modulation was negligible in shuffled spike controls. Moreover, phase modulation was action-specific: neurons in one action ensemble showed markedly reduced modulation when aligned to the phases of other actions, across SMC, PPC, and even HPC (Fig. 5i).

Notably, the sequential firing patterns observed in action ensembles are unlikely to result solely from somatotopically-ordered cutaneous inputs. For example, both left and right scratch ensembles were recorded from the right hemisphere (Fig. 3a), yet each exhibited strong phase tuning despite receiving tactile input from different sides of the body. If these patterns were driven purely by touch, right-hemisphere neurons would not be reliably activated during right-sided scratching. Moreover, while tactile input may serve as a cue to anchor the neural sequence to the action (e.g., forepaw contact at the onset or offset of a rear, brief whisker-forepaw contact during head grooming), touch events are sparse and unevenly distributed across the action cycle (Extended Data Fig 5). If sequential patterns of tactile input were the primary driver of these sequences, we would expect tuning to certain phases (i.e., those associated with tactile events) to be overrepresented and others (i.e., lacking tactile input) to be underrepresented or missing. In contrast, the phase tuning we observed smoothly tiled the entire action cycle (Fig. 5e).

Given the prominent action phase tuning observed in cortical neurons, we once again re-evaluated all recorded neurons to assess which feature class best explained their activity, now including action phase as a feature in the Specific Action class. The results revealed that for a stunningly large proportion of both RS and FS neurons, action phase emerged as the top explanatory feature (Fig. 5j), highlighting its role as a key organizing principle of cortical dynamics during natural behavior. Thus, action sequence ensembles serve as a neural substrate that supports not only the awareness of which action is being performed, but also the awareness of the current state within the action cycle.

A kinematic encoder model reveals neural subspaces for distinct actions

To test whether actions are indeed a dominant organizing principle underlying cortical dynamics, we developed an encoder model to predict population activity from continuous kinematic features, without providing any explicit information about action labels or phase. To this end, we implemented a modular EDTCN architecture incorporating 1) a shared “Core” kinematic encoder, trained across all sessions to learn a common set of latent features from kinematic inputs, and 2) a session-specific “Readout” layer that maps these shared features onto the activity of simultaneously recorded neurons (Fig. 6a). Importantly, we excluded discrete action labels and action phase from the model input, ensuring that predictions relied solely on continuous kinematic variables—including joint angles, segment positions, and allocentric features such as head direction and spatial location—to uncover the principles governing cortical activity. Our “Core+Readout” model outperformed other encoder architectures in predicting neural activity (Extended Data Fig. 6; see Methods), with especially strong performance in SMC and PPC. On held-out test data, the model explained substantial variance in RS neuron activity across S1 (26.4%), S2 (15.8%), M1 (31.1%), and PPC (30.7%) (Fig. 6b). In contrast, prediction performance

was notably lower in V1 (10.3%) and HPC (13.1%), even though locomotion and allocentric features (i.e., “place” and “head direction”) were included among the inputs.

Importantly, beyond predicting single-neuron activity, the model recapitulated the complex population dynamics observed in the recorded data (e.g., Fig. 6c). When we aggregated the model outputs into a pseudo-population of 8,763 neurons (as if recorded simultaneously), sorting with Rastermap revealed distinct functional clusters that aligned to specific actions (Extended Data Fig. 6c). To visualize clustering at the population level, we applied t-SNE to the modeled neural activity from an example recording session (see Methods). Time points were colored by the corresponding action, revealing that the high-dimensional neural population vector segregated into discrete subspaces on a neural manifold, each corresponding to a distinct action (Fig. 6d). Moreover, when the modeled neural activity was sorted into action ensembles based on previously observed action phase tiling, the modeled population exhibited clear sequential activity aligned to action phase (Fig. 6e). Together, these results demonstrate that even when driven solely by high-dimensional, continuous kinematic input, cortical dynamics during natural behavior organize into action-specific subspaces and represent each action cycle through phase-aligned neural sequences. These findings demonstrate that cortical neurons compress complex kinematic information into an efficient body schema, ultimately to support action awareness.

Discussion

Here, we show that the cortex encodes high-dimensional body kinematics in compact, structured, and ethologically grounded neural representations of body schema and action. By combining full-body kinematic tracking with large-scale in vivo electrophysiology in freely moving mice, we demonstrate that the cortex encodes posture using distinct reference frames, integrates gravitational context, and organizes neural activity into action sequence ensembles. Together, these principles provide a foundation for how proprioceptive inputs are integrated into a coherent internal model of the body in space, with implications for restoring sensorimotor function in cases of injury or neurological disease.

Understanding the body schema requires identifying its representational format, particularly the reference frames in which it is encoded. Prior work (Wolpert & Ghahramani, 2000; Alonso et al., 2023) proposed egocentric reference frames for the posture representation. However, based on our findings during natural behavior, the body schema is anchored to gravity and the body’s midline. We found that PPC primarily encodes postural angles relating to left-right turning postures, while SMC predominantly encode vertical height and pitch of the body with respect to the ground. This compact, physically-grounded coordinate system for representing 3D body posture allows the cortex to decompose high-dimensional kinematic inputs into ethologically-relevant components distributed across anatomically distinct regions. In parallel, our results are consistent with prior findings that V1 encodes posture and visual space in a head-centered reference frame (Mimica et al., 2023), and that the hippocampus encodes location and locomotion in an allocentric, world-centered reference frame (Moser et al., 2017).

To navigate and interact effectively with the physical world, land-dwelling vertebrates must anchor their posture representation relative to the ground. Crucially, "ground" is not fixed—it can be the flat surface of an arena, a tree branch, or the floor of a tunnel—so the brain must maintain an internal reference rooted at the base and aligned with gravity to stabilize posture across diverse terrains. The “idiotropic vector”, a concept introduced by Mittelstaedt, refers to this centrally stored, gravity-aligned axis that anchors the perception of upright and supports spatial orientation across contexts (Mittelstaedt, 1999). Our findings provide evidence in support of this conceptual model, suggesting that the sensorimotor cortex integrates tactile input from ground contact with proprioceptive and potentially vestibular cues to construct body-part positions in a gravicentric coordinate system. This representation enables rapid readout of body elevation relative to the ground, supporting efficient postural control. Unlike joint-centric angles, gravity-referenced features, such as elevation and pitch, provide a global representation of posture within the gravitational field. Our findings support the view that SMC contribute to a grounded, gravity-referenced body schema, enabling the brain to monitor and adjust posture to support balance, facilitating stable motor control across diverse environments and behavioral demands. Future work is needed to determine how sensorimotor and parietal areas integrate multisensory cues to construct and maintain this gravity- and ground-referenced body schema.

From an evolutionary perspective, it is adaptive to represent midline-referenced body configurations separately from ground-referenced postures. Vertebrates first evolved in water, where the body’s dorsoventral axis naturally aligned with the gravitational vector. Consequently, the primary postural changes during swimming arose from alternating left-right flexion along the body midline, driven by spinal segment and tail shifts. In the absence of a ground plane, a midline-referenced body schema, aligned with vestibular signals from the inner ear, likely provided the most effective internal representation of posture.

The observed overrepresentation of PPC neurons tuned to tail and trunk angles, particularly those associated with left-right deviations from the midline, raises the possibility that PPC contributes not only to real-time encoding of posture, but also to anticipatory postural control. During movement, animals must generate preemptive adjustments to maintain balance, such as initiating compensatory tail movements before the onset of whole-body turns. In this context, encoding of tail abduction–adduction and roll angles in PPC may reflect predictive motor representations that help coordinate such adjustments. Supporting this idea, PPC neurons in rodents encode upcoming turns during maze navigation, consistent with a broader role in predictive coding of future actions (Harvey et al., 2012; Whitlock et al., 2012). Our finding that the PPC encodes posture relative to the midline is consistent with clinical and experimental observations of hemi-spatial neglect, in which PPC lesions impair awareness of the contralateral side of the body (Brain, 1941; King & Corwin, 1993; Karnath et al., 2002). Thus, PPC is an integral part of an internal forward model that predicts the physical consequences of turning actions, enabling smooth and coordinated voluntary movement.

During natural behavior, the brain must not only track posture but also anticipate the body’s biomechanics across different action phases. The varying gravitational demands across phases likely contribute to the formation of distinct action-specific neural ensembles. For example, during

an unsupported rear, the rear-up phase requires muscular effort to raise the body's center-of-mass against gravity. The hold phase demands postural stability under sustained gravitational load, as the animal balances upright with its full weight supported by the hind limbs. In the rear-down phase, gravity assists the descent, but controlled forelimb extension and muscle relaxation are required for a stable landing. Finally, the landing phase engages forelimb and trunk reflexes upon forepaw-ground contact to absorb impact and restore balance in a quadrupedal stance. Each phase reflects distinct biomechanical interactions with gravitational forces—rising against it, stabilizing under it, descending with it, and bracing for impact—likely recruiting distinct neural populations. Indeed, we observed neurons across the rear ensemble that represented rear up, rear hold, and rear down phases. More broadly, each distinct action imposes unique biomechanical demands across different phases, imposing a need for phase- and action-specific representations. Across all cortical areas examined, we found that a substantial fraction of neurons exhibited highly stable, phase-specific firing patterns, organized into coherent action-specific ensembles. Hence, rather than encoding static postures, these neurons preferentially represented transitions between dynamic posture changes, aligned to distinct action phases.

Our findings also suggest that the cortex compresses the high-dimensional space of dynamic postural configurations into more compact representations that reflect the stereotypical posture and movement patterns associated with individual actions. Supporting this idea, our modeling showed that population activity is organized into low- dimensional subspaces associated with distinct actions. Thus, the brain effectively addresses the challenge of representing the vast space of kinematic configurations in a freely moving, multi-jointed body. Natural behavior thus constrains this kinematic space based on natural statistics, which the brain efficiently encodes as action subspaces. These action subspaces parallel continuous attractor models in which continuous variables, such as head orientation or spatial location, are encoded as localized activity clusters on a low-dimensional manifold embedded within a high-dimensional neural state space (Kakaria & de Bivort, 2017; Khona & Fiete, 2022; Kutschireiter et al., 2023). While previous studies reported sequential neural dynamics during isolated, constrained behaviors such as treadmill locomotion (Foster et al., 2014; Xing et al., 2022; Borgognon et al., 2025) or a trained reaching task (Churchland et al., 2012), our findings unveiled the prevalence of distinct action ensembles whose sequential activity tiled the full phase cycle of multiple species-typical actions during freely moving behavior. Our observation that each action recruits its own distinct neural sequence revealed a previously underappreciated organizational principle of cortical dynamics during natural behavior (Miller et al., 2022).

Note that our discovery of action ensembles in PPC aligns with prior work showing that low-intensity stimulation of PPC evokes the subjective experience or urge to perform actions without triggering motor output (Desmurget et al., 2009). In our dataset, a large fraction of putative inhibitory neurons encoded non-specific body motion or locomotion, with a subset displaying action-specific, phasic activation. This suggests that inhibitory interneurons may contribute to network-level gating, suppressing competing ensembles and allowing specific action sequences to unfold. Electrical microstimulation, by simultaneously activating both excitatory and inhibitory elements, could bias the network toward a particular attractor by tipping the balance of excitation and inhibition. This raises the possibility that microstimulation evokes percepts by recruiting

action-specific ensemble dynamics, including their temporal progression, rather than activating isolated neurons. Such ensemble recruitment could underlie both the execution and conscious experience of actions—contributing to action awareness.

While our behavioral paradigm captured a broad range of spontaneous full-body actions, it did not encompass the full richness of the mouse behavioral repertoire. In particular, our setup lacked fine forelimb behaviors such as skilled reaching, manual manipulation, or climbing. As such, our findings may not generalize to cortical representations underlying dexterous limb control (Grier et al., 2025). Future work in enriched environments and social or task-motivated contexts may reveal additional action ensembles or finer-grained structure within action subspaces (Dunn & Marshall et al., 2021; Marshall et al., 2021; Klibaite et al., 2025). Nevertheless, our framework predicts that any species-typical action would recruit distinct cortical ensembles. Phase-aligned activation patterns within action ensembles reflected the stereotyped structure of natural behavior, where both posture and neural activity unfold in predictable sequences. Specific neural sequences may encode upcoming postural transitions, rather than current or past sensory states, thus enabling a predictive internal model of motor intention. In this view, sequential dynamics could also support transitions between actions, such as rearing to walking, providing a circuit-level substrate for the planning of voluntary actions and continuity of behavior.

Together, our findings uncover core organizing principles underlying neural dynamics during spontaneous behavior. By revealing how neural ensembles encode posture and action, we provide a foundation for understanding the neural basis underlying the body schema and action awareness. These insights could inform future brain-machine interfaces and bio-inspired robotics, where embedding ground- and gravity-referenced sensing may improve body-part localization, postural control, and autonomous movement planning. Finally, integrating physically realistic biomechanics and physics-based simulations (Aldarondo et al., 2024), together with action context into dynamical systems models may enable more adaptive and flexible motor control across diverse tasks and environments.

References

1. Aldarondo, D. *et al.* A virtual rodent predicts the structure of neural activity across behaviours. *Nature* **632**, 594–602 (2024).
2. Alonso, I. *et al*. Peripersonal encoding of forelimb proprioception in the mouse somatosensory cortex. *Nat. Commun.* **14**, 1866 (2023).
3. Angelaki, D. E. & Cullen, K. E. Vestibular system: the many facets of a multimodal sense. *Annu. Rev. Neurosci.* **31**, 125–150 (2008).
4. Barlow, H. B. Possible principles underlying the transformations of sensory messages. In *Sensory Communication* (ed. Rosenblith, W. A.) 217–234 (MIT Press, Cambridge, MA, 1961).
5. Bermúdez, J. L., Marcel, A. & Eilan, N. (eds) *The Body and the Self* (MIT Press, Cambridge, MA, 1995).
6. Berthoz, A. *Le sens du mouvement* (Odile Jacob, Paris, 1997).
7. Head, H. & Holmes, G. Sensory disturbances from cerebral lesions. *Brain* 34, 102–254 (1911).
8. Borgognon, S. *et al.* Regional specialization of movement encoding across the primate sensorimotor cortex. *Nat. Commun.* **16**, 5729 (2025).
9. Brain, W. R. Visual disorientation with special reference to lesions of the right cerebral hemisphere. *Brain* **64**, 244–272 (1941).
10. Chapin, J. K. & Lin, C. S. Mapping the body representation in the SI cortex of anesthetized and awake rats. *J. Comp. Neurol.* **229**, 199–213 (1984).
11. Chesler, A. T. *et al.* The role of PIEZO2 in human mechanosensation. *N. Engl. J. Med.* **375**, 1355–1364 (2016).
12. Churchland, M. M. *et al.* Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
13. Desmurget, M. *et al.* Movement intention after parietal cortex stimulation in humans. *Science* **324**, 811–813 (2009).
14. Dunn, T. W. *et al.* Geometric deep learning enables 3D kinematic profiling across species and environments. *Nat. Methods* **18**, 564–573 (2021).
15. Foster, J. D. *et al.* A freely-moving monkey treadmill model. *J. Neural Eng.* **11**, 046020 (2014).
16. Grier, H. A., Salimian, S. & Kaufman, M. T. Mouse sensorimotor cortex reflects complex kinematic details during reaching and grasping. *eLife* **14**, RP106270 (2025).
17. Haggard, P. & Wolpert, D. Disorders of body scheme. In *Higher-Order Motor Disorders* (eds Freund, H.-J., Jeannerod, M., Hallett, M. & Leiguarda, R.) 261–271 (Oxford Univ. Press, 2005).
18. Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
19. Kakaria, K. S. & de Bivort, B. L. Ring attractor dynamics emerge from a spiking model of the entire protocerebral bridge. *Front. Behav. Neurosci.* **11**, 8 (2017).
20. Karnath, H.-O., Milner, A. D. & Vallar, G. The cognitive and neural bases of spatial neglect. *Nat. Rev. Neurosci.* **3**, 554–564 (2002).
21. Khona, M. & Fiete, I. R. Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* **23**, 744–766 (2022).
22. Klibaite, U. *et al*. Mapping the landscape of social behavior. *Cell* **188**, 2249–2266.e23 (2025).

23. King, V. & Corwin, J. V. Comparisons of hemi-inattention produced by unilateral lesions of the posterior parietal cortex or medial agranular prefrontal cortex in rats: neglect, extinction, and the role of stimulus distance. *Behav. Brain Res.* **54**, 117–131 (1993).
24. Kutschireiter, A., Basnak, M. A., Wilson, R. I. & Drugowitsch, J. Bayesian inference in ring attractor networks. *Proc. Natl Acad. Sci. USA* **120**, e2210622120 (2023).
25. Lea, C., Flynn, M. D., Vidal, R., Reiter, A. & Hager, G. D. Temporal convolutional networks for action segmentation and detection. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)* 156–165 (2017).
26. Lee, T. & Kim, U. Descending projections from the dysgranular zone of rat primary somatosensory cortex processing deep somatic input. *J. Comp. Neurol.* **520**, 1021–1046 (2012).
27. MacDonald, P. A. & Paus, T. The role of parietal cortex in awareness of self-generated movements: a transcranial magnetic stimulation study. *Cereb. Cortex* **13**, 962–967 (2003).
28. Marasco, P. D., Bourbeau, D. J., Shell, C. E., Granja-Vazquez, R. & Ina, J. G. The neural response properties and cortical organization of a rapidly adapting muscle sensory group response that overlaps with the frequencies that elicit the kinesthetic illusion. *PLoS One* **12**, e0188559 (2017).
29. Maravita, A., Spence, C. & Driver, J. Multisensory integration and the body schema: close to hand and within reach. *Curr. Biol.* **13**, R531–R539 (2003).
30. Marshall, J. D. *et al*. Continuous whole-body 3D kinematic recordings across the rodent behavioral repertoire. *Neuron* **109**, 420–437.e8 (2021).
31. Miller, C. T. *et al.* Natural behavior is the language of the brain. *Curr. Biol.* **32**, R482–R493 (2022).
32. Mimica, B., Dunn, B. A., Tombaz, T., Bojja, V. & Whitlock, J. R. Efficient cortical coding of 3D posture in freely behaving rats. *Science* **362**, 584–589 (2018).
33. Mimica, B. *et al.* Behavioral decomposition reveals rich encoding structure employed across neocortex in rats. *Nat. Commun.* **14**, 3947 (2023).Mittelstaedt, H. Interaction of gravity and posture control in humans. *Adv. Space Biol. Med.* **6**, 135–157 (1997).
34. Mittelstaedt, H. The role of the otoliths in perception of the vertical and in path integration. *Ann. N. Y. Acad. Sci.* **871**, 334–344 (1999).
35. Monsees, A. *et al.* Estimation of skeletal kinematics in freely moving rodents. *Nat. Methods* **19**, 1500–1509 (2022).
36. Moser, E., Moser, M.-B. & McNaughton, B. Spatial representation in the hippocampal formation: a history. *Nat. Neurosci.* **20**, 1448–1464 (2017).
37. Niell, C. M. & Stryker, M. P. Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron* **65**, 472–479 (2010).
38. Prud'homme, M. J. & Kalaska, J. F. Proprioceptive activity in primate primary somatosensory cortex during active arm reaching movements. *J. Neurophysiol.* **72**, 2280–2301 (1994).
39. Roll, P., Roll, R. & Velay, J.-L. Proprioception as a link between body space and extra-personal space. In *Brain and Space* (ed. Paillard, J.) 112–132 (Oxford Univ. Press, New York, 1991).
40. Sirigu, A. *et al.* Altered awareness of voluntary action after damage to the parietal cortex. *Nat. Neurosci.* **7**, 80–84 (2004).
41. Stringer, C. *et al*. Rastermap: a discovery method for neural population recordings. *Nat. Neurosci.* **28**, 201–212 (2025).
42. Sun, C., Schinder, G. H. R. J. & Schmidt-Hieber, C. Characterizing speed cells in the rat hippocampus. *Cell Rep.* **25**, 1930–1940.e4 (2018).

43. Tsay, A. J., Giummarra, M. J., Allen, T. J. & Proske, U. The sensory origins of human position sense. *J. Physiol.* **594**, 1037–1049 (2016).
44. Vallar, G. & Calzolari, E. Unilateral spatial neglect after posterior parietal damage. In *Handbook of Clinical Neurology* Vol. 151, 287–312 (Elsevier, 2018).
45. Vallar, G. & Ronchi, R. Somatoparaphrenia: a body delusion. A review of the neuropsychological literature. *Exp. Brain Res.* **192**, 533–551 (2009).
46. Weinreb, C. *et al.* Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics. *Nat. Methods* **21**, 1329–1339 (2024).
47. Whitlock, J. R., Pfuhl, G., Dagslott, N., Moser, M.-B. & Moser, E. I. Functional split between parietal and entorhinal cortices in the rat. *Neuron* **73**, 789–802 (2012).
48. Wolpert, D. J. & Ghahramani, Z. Computational principles of movement neuroscience. *Nat. Neurosci.* **3** (Suppl.), 1212–1217 (2000).
49. Xing, D., Truccolo, W. & Borton, D. A. Emergence of distinct neural subspaces in motor cortical dynamics during volitional adjustments of ongoing locomotion. *J. Neurosci.* **42**, 9142–9157 (2022).

Methods

Animal Care

Adult C57BL/6J (JAX stock #000664) mice were obtained from the Jackson Laboratory. During electrophysiological and behavioral experiments, mice were single housed under a 12-hour normal light/dark cycle with ad libitum access to food and water. All experiments were conducted during the light phase. Details regarding the sex and age of all animals are provided in Supplemental Table S1. All procedures were performed in accordance with protocols approved by the Institutional Animal Care and Use Committees at the Massachusetts Institute of Technology.

Procedures

Mice were implanted with head-posts to fix the head during optical imaging, microdrive implantation, and headstage connection. Mice were anesthetized with isoflurane (3 % for induction, 1.0-1.5 % for maintenance) delivered in oxygen at 0.8 L/min and positioned in a stereotaxic frame (Model 963, David Kopf Instruments). A custom stainless steel headbar (Pololu; 0.060″ ± 0.004″, Grade 304) was affixed to the skull using cyanoacrylate adhesive (Krazy Glue, #KG585) and sealed with a protective layer of dental cement (Stoelting).

To guide tetrode microdrive targeting, intrinsic optical signal (IOS) imaging was used to localize cortical regions responsive to vibrotactile stimulation of different body parts. Following full recovery from headpost implantation, mice were lightly anesthetized with isoflurane (0.5–1%). Kwik-Cast and dental cement covering the skull were removed to provide optical access. Body temperature was maintained at 37°C with a heating pad (FHC), and respiration was continuously monitored. Illumination was delivered using high-power LEDs (Thorlabs M530L3, green for vascular reference; M625L4, red for hemodynamics) oriented at oblique angles onto the cortical surface. Reflected light was collected by a tandem lens macroscope (objective lens: Canon 50mm f1.2; imaging lens: Nikon 85mm f1.4) and imaged by a sCMOS camera (PCO Edge 4.2). Images were acquired at 20 Hz at 512×512-pixel resolution (with 4x average binning) using PCO acquisition software. Vibrotactile stimulation (Seeed, Part #316040001; 4s, 10s inter-stimulus interval) was applied to the skin and fur of the contralateral neck, back, tail, forepaw, hindpaw, or E-row whiskers. White noise was presented continuously to minimize auditory cortex activation. Reflectance changes ($\Delta R/R$) were calculated by normalizing pixel intensities to pre-stimulus baselines and averaging across trials (≥ 10 repetitions). Post hoc analysis aligned regions-of-interest showing stimulus-locked decreases in reflectance with vascular landmarks to guide subsequent craniotomies and tetrode microdrive implantation.

Prior to microdrive implantation surgery, mice were anesthetized with isoflurane (3 % for induction, 1.0-1.5 % for maintenance) delivered in oxygen at 0.8 L/min and secured in a stereotaxic frame (Model 963, David Kopf Instruments). Preoperative analgesia (meloxicam, 0.5 mg/kg, subcutaneous) was administered to reduce inflammation. A ground screw was inserted into the contralateral hemisphere and pinned to the EIB. Following a small craniotomy and

durotomy at the IOS-targeted area, the tetrode bundle was lowered just below the brain surface. Moveable components of the microdrive and the craniotomy were coated with sterile ophthalmic ointment (Puralube). The microdrive was affixed to the skull with dental cement. A shield, made of transparency film laminated with copper tape and connected to the EIB, was cemented to the skull and covered with black electrical tape to protect the implant during postoperative recovery. Postoperative analgesia (buprenorphine, 0.05 mg/kg, subcutaneous) was administered on the day of surgery. Animals were monitored and received continued analgesia (meloxicam, 0.5 mg/kg, subcutaneous) for at least two additional days and were allowed to recover for a minimum of seven days prior to recording.

Electrophysiology

For microdrive construction, tetrodes were prepared by clamping four insulated nichrome alloy microwires (Sandvik, #PX000003) into a tetrode spinner (NeuraLynx, version 2.0). In manual mode, the wires were rotated 100 turns clockwise, followed by 40 counter-clockwise turns to release residual torsion. The tetrode bundle was heat-treated with a heat gun (350–400 °C, low setting) held 2 cm from the wires for 5 seconds from three directions to fuse the bond coat while preserving the insulation. Individual tetrode wires were pinned by gold pins (Neuralynx, #31-0603-0102) to a custom 64-channel electronic interface board (EIB, designed in DipTrace and printed by OSH Park) equipped with two 32-channel Omnetics connectors. The fused end of each tetrode was routed through a polyimide tube (ID: 99 µm, OD: 168 µm, Polymicro Technologies) arranged in a 2×8 or 4×4 array and epoxied to the nano-Drive (Cambridge Neurotech). Once routed, tetrodes were glued to the polyimide array and trimmed to equal length using tungsten carbide scissors (FST, #14502-14). Prior to implantation, tetrodes were immersed in gold-plating solution (Neuralynx, #31-0611-0100) and electroplated using Nano-Z 1.2 (White Matter) to a target impedance of 200 kΩ. In some cases, tetrodes were coated with the fluorescent dye DiI (Invitrogen, #V22885) to assist in post hoc histological confirmation of recording sites.

For chronic electrophysiological recordings, mice were briefly head-fixed to minimize mechanical strain while connecting the 2x32 channel Omnetics connectors on the 64-channel headstage (Intan, Part #C3325). The brain was allowed to relax for at least 15 minutes before transferring the animal to the denoised recording arena, enclosed within a grounded Faraday cage. The headstage was tethered via a 6-ft ultra-thin SPI cable (Intan, Part #C3216) to a custom passive commutator, allowing unrestricted movement during recording sessions. Extracellular voltage signals were acquired at 30 kHz using an RHD2000 Acquisition Board (Intan, Part #C3100) and RHX Data Acquisition Software (Intan). Signals were hardware filtered between 0.1 Hz to 10 kHz to enable separation of fast- and regular-spiking neuronal populations based on waveform shape. At the end of each session, mice were briefly head-fixed to allow safe removal of the headstage and advancement of the tetrode array by 62.5 µm increments. Recording sites were verified by post hoc histology.

For spike sorting, initial automated single-unit isolation was performed using Kilosort 3.0 (https://github.com/MouseLand/Kilosort). Tetrodes were mapped onto 16 “shanks” of four channels each. Extracellular voltage signals were high-pass filtered at 300 Hz and whitened prior

to automated clustering in Kilosort3 using rigid drift correction (nblocks = 1). Clusters were subsequently manually curated in MClust 4.4 (http://redishlab.neuroscience.umn.edu/MClust). Isolation quality was assessed using: false alarm (FA) rate, defined as the fraction of spikes with inter-spike intervals (ISI) shorter than 2 ms (< 2% threshold); mean firing rate (FR > 0.25 sp/s threshold); and signal-to-noise ratio (SNR > 5 s.d. threshold, defined as mean waveform amplitude divided by the standard deviation of baseline noise for the peak channel). Units not meeting these criteria were excluded from further analysis. Units with a narrow waveform shape (trough-peak duration < 0.45 ms) were labeled as putative fast-spiking (FS), and units with a wide waveform shape (trough-peak duration > 0.45 ms) were labeled as putative regular-spiking (RS).

Histology

Mice were deeply anesthetized and transcardially perfused with phosphate-buffered saline (PBS; Gibco), followed by 4% paraformaldehyde (PFA) in PBS. Brains were post-fixed overnight and then cryoprotected in 30% sucrose in PBS at 4°C. Tissue was embedded in OCT compound (Sakura Finetek) and coronally sectioned at 80 µm using a cryostat (Leica Biosystems). Sections were stained with Neurotrace Blue (1:500; Thermo Fisher Scientific, #N21479) in 0.3% Triton X-100/PBS. The sections were then briefly rinsed with PBS and mounted on slides with a custom-made mounting medium (Takatoh & Prevosto et al., 2021). All histological images were acquired with a Keyence microscope (BZ-X800). Locations of recording tracks were manually registered to the Allen CCF.

Behavior Recordings

The recording arena was constructed on a 1 × 1 m anodized aluminum breadboard (Thorlabs, Part #MB30) mounted on rubber feet (Thorlabs, Part #RDF1) to minimize vibration. The enclosure was assembled using a custom T-slotted aluminum frame with grounded aluminum side panels to provide electrical shielding. The arena floor consisted of a multi-layered base (foam underlay, PVC baseboard, and green vinyl film) designed to provide traction and durability. A cylindrical glass enclosure (hurricane candle holder; inner diameter 189 cm, height 205 cm) surrounded the arena, allowing unobstructed visual access. Illumination was provided by three to four grounded 30 W LED arrays (IKAN), ensuring uniform lighting conditions. Between recording sessions, both the glass cylinder and arena floor were cleaned by wiping with 70% ethanol.

Six machine-vision color cameras (acA1920-150uc; Basler) were operated via Pylon Software (Basler) to record video at 1152 × 1024 resolution in RGB pixel format using on-chip PGI 5 × 5 debayering. Cameras acquired frames with a global shutter synchronized by external hardware frame triggers at 100 fps with 2 ms exposure time. Each camera was mounted with an 8 mm fixed-focal-length lens (M0824-MPW2, Computar) with 300 mm working distance. Polarizing filters (PR120-30.5; Midwest Optical) were mounted on the lenses to reduce glare and reflections from the glass cylinder and other reflective surfaces. Cameras were mounted using ½″ posts and accessories (Thorlabs) and arranged radially around the arena at ~60° intervals. To provide sufficient parallax for 3D triangulation, cameras were angled vertically at alternating inclinations of 0° (horizontal) and −15° (downward).

Intrinsic camera calibration (lens distortion and focal parameters) was performed using a 10 × 11 cm checkerboard. Extrinsic calibration (camera pose and orientation) was conducted using a custom 3D-printed L-frame affixed to a circular plexiglass base (8″ diameter, 0.25″ height). Camera parameters were computed using custom MATLAB scripts and achieved sub-pixel reprojection error for each view.

Multi-camera triggering and video recording were managed by Campy python software (https://github.com/ksseverson57/campy). Camera triggering was synchronized by a microcontroller (Teensy 3.2) running custom Arduino triggering software (Campy trigger module). Video frames were acquired using custom Python recording software (CamPy), with each stream compressed in real time using FFmpeg configured with H.264 or H.265 NVENC hardware encoding on NVIDIA GPUs. Multi-camera video files were acquired as single MP4s. Total video acquired were reduced in size from ~10 TB/hr to ~10 GB/hr when compressed at 4 Mb/s constant bit rate, making saving, storing, and loading video datasets manageable. To enable precise alignment of video frames with electrophysiological recordings, trigger TTL pulses were simultaneously recorded on analog input channels of the RHD2000 board at 30 kHz.

Pose Estimation and Kinematics Calculation

DANNCE pose estimation models were architected and trained similarly to procedures described in Dunn, Marshall et al., 2021. Briefly, a total of 835 frames from 15 sessions were manually annotated using the Label3D GUI to define 3D positions for a 44-keypoint mouse skeleton. 775 frames were used for training, with 60 frames held out for validation. A 2D convolutional neural network (CNN) was first trained to estimate the center of mass (COM), defined as the geometric mean of the labeled keypoint positions. The 3D CNN DANNCE model was finetuned from pre-trained weights originally trained on the Rat7M dataset (Dunn & Marshall et al., 2021). During DANNCE model finetuning, data augmentation included random jittering of the centroid position, rotation, and mirroring (with left-right keypoint re-indexing) of the 3D volumetric grid to improve generalization. Synchronized frames from the six cameras were projected into the augmented 80×80×80 volumetric grid (1.5 mm isometric voxel resolution) using projective geometry. Additional augmentations included random variation in view selection, brightness, and hue. To improve smoothness of keypoint predictions, a static “weighted average” layer on top of the “MAX” network outputted the weighted average position of the probability map. Keypoint prediction accuracy was evaluated on a held-out validation set (n=60 frames) by computing the 3D Euclidean distance between predicted and ground truth keypoint positions. Once validated, the trained COM and DANNCE models were applied to all behavioral recordings (392 sessions across 20 mice, totaling 141.12 million frames). For tuning analyses, keypoint positions were smoothed using an 11-frame moving average filter.

We developed a geometric anatomical framework to quantify full-body kinematics with high precision. This geometric anatomical model transforms 3D keypoint tracking data (x,y,z) into angles defined in world-, body-, and joint-centered reference frames (Extended Data Fig. 1). A total of 26 anatomical keypoints were tracked across the body to capture the position and rotation of each major segment and joint, with at least three keypoints per segment to resolve 3D plane orientation. Joint-centered axes were aligned to the anatomical planes of each segment: the

anteroposterior axis (X) followed the proximal-distal orientation, the mediolateral axis (Y) aligned to segment width, and the vertical axis (Z) was defined as the cross-product of X and Y, capturing long-axis rotation. Using this geometric body model, we defined rotations for 16 major skeletal joints (i.e., neck, thoracic-lumbar and lumbar-sacral spine, tail, shoulders, elbows, wrists, hips, knees, and ankles) encompassing 38 rotational degrees of freedom. This comprehensive framework provides a continuous, anatomically-defined description of the body configuration in freely moving animals, enabling detailed parameterization of body posture, motion, and postural dynamics.

For allocentric angle calculation, rotation matrices for each body segment were constructed from the X, Y, and Z axis vectors derived from 3D Cartesian keypoint positions. Allocentric rotation angles (yaw, pitch, roll) were then computed relative to the calibrated world coordinate frame using a "YXZ" Euler angle sequence. This sequence was selected over the more common "XYZ" convention to avoid gimbal locks, as pitch angles in our mouse behavior recordings occasionally exceed ±90° (e.g., during high rearing or grooming), while head roll rarely exceeds 90°. Subsequent transformations, including body-centered alignments, were computed using an "XYZ" Euler sequence applied to the allocentric angles.

For joint-centered angle calculation, joint rotation matrices were computed by multiplying the rotation matrix of the distal (target) segment by the transposed rotation matrix of the proximal (basis) segment. This "rotation-of-coordinates" convention preserves the hierarchical structure of the skeleton, allowing angles to propagate from proximal to distal joints and enable full reconstruction of 3D keypoint positions via forward kinematics. The resulting joint rotation matrix (R_j) captures the relative rotation of the distal segment with respect to the proximal segment at joint *j*. To maximize the dynamic range of each joint and avoid gimbal locks, we centered the rotation at the mean or "T-pose" orientation for each joint by multiplying the inverse of the singular value decomposition (SVD) of the average joint rotation matrix with the joint rotation matrix. With this SVD T-pose alignment procedure, Euler sequence choice was less impactful.

$$R_j = {R_{dist}}^{-1} * R_{prox}$$

$$U, V = SVD\left(\frac{1}{n}\sum R_j\right)$$

$$R_{Tpose} = U * V^{-1}$$

$$R_{j,center} = {R_{Tpose}}^{-1} * R_j$$

Joint-centered Euler angles were then extracted from these rotation matrices using axis sequences selected to yield interpretable components: flexion–extension (*fe*), abduction–adduction (*aa*), and internal–external rotation (*ie*). For each joint, we validated the selected Euler sequence by binning angles and visualizing average aligned keypoint positions across bins to confirm correspondence to expected biomechanical axes (Extended Data Fig. 1). To account for biomechanical constraints, special considerations were made: elbow and knee joints were modeled as single-axis hinge joints and restricted to *fe* only, computed via trigonometric angle between vectors; tail angles were restricted to *fe* and *aa*, as *ie* rotation could not be reliably

estimated with a single midline keypoint on the tail. In total, 38 joint angles were calculated. Angles were smoothed using either a mild moving average (21-frame sliding window) or Savitzky-Golay filter (fifth order, 21-frame sliding window) applied to unwrapped angles to reduce noise while preserving high-frequency postural dynamics. Parallelized computation using MATLAB's "pagemtimes" function enabled efficient, frame-independent Euler angle extraction.

We performed principal component analysis (PCA) on a combined set of joint angles and head and body segment orientation angles to identify low-dimensional representations of full-body posture. Angles were first centered in circular coordinates, and outliers were removed using percentile-based filtering. Wrist joints were excluded due to nosier tracking. Features from 37 full-length (one hour) recording sessions were concatenated and z-scored prior to PCA. PCA was performed on the pooled feature matrix to identify dominant orthogonal postural modes (PCs), and cumulative variance explained was used to assess the dimensionality of the posture space. PC loadings were saved and projected onto additional sessions; for example, posture PC scores were used to assess neural tuning by correlating firing rates with percentile-binned PCs. To visualize posture dimensions captured by individual PCs, timepoints were binned into quintiles along each PC axis, and the mean 3D keypoint positions were computed for the lowest and highest quintiles. Pose data were smoothed and aligned to a common reference frame (lumbar spine segment). The mean pose was rendered in 3D using a predefined keypoint skeleton. This procedure revealed interpretable low-dimensional posture motifs along PC axes, such as vertical elevation (Z), forward and backward leaning (X), and leftward or rightward turning (Y).
Non-specific body motion was quantified as the smoothed mean of absolute angular velocity across all joints, excluding wrists and ankles to reduce noise. Angular velocities were clipped to the 1st–99th percentile range to mitigate outliers, smoothed using a 21-frame moving average, and expressed in degrees per second. Locomotion speed was defined as the magnitude of frame-to-frame head displacement in the horizontal (XY) plane, also smoothed with a 21-frame moving average and expressed in millimeters per second.

To estimate proximity (touch) of body parts to either the ground or other body parts as proxies for touch, we computed two features: distance-to-floor (d2f) and distance-to-face (d2face). D2f features were computed as the minimum Euclidean distance between each paw segment and the floor or wall, whichever was closer, clipped at 5 mm. D2face features were computed as the Euclidean distance between each paw and the snout keypoint, as well as the distance between each paw and the ipsilateral ear, clipped at 5 mm. Derivatives of d2f and d2face were computed to estimate touch dynamics.

For raster plotting, Rastermap was used to sort unit order (Stringer et al., 2025). Briefly, spike rates were computed from spike counts using a 40-ms Gaussian convolution. Spike rates were Z-scored and inputted into Rastermap (https://github.com/MouseLand/rastermap) for single neuron sorting with the following settings: n_PCs = n_neurons/2, locality = 0.5, time_lag_window = 100, symmetric = False. For rasters showing activations, spike rates were Z-scored, sorted by the Rastermap neuron order, plotted as an image, and color limited to positive range to emphasize activations.

To quantify stimulus selectivity, we computed marginal (1D) tuning curves by estimating each neuron's conditional firing rate as a function of each continuous behavioral or kinematic variable. Stimulus variables were first filtered to exclude outliers using percentile-based thresholds. For most features, stimuli were binned into 25 equal-percentile bins. Categorical variables (e.g., action category, action phase) were handled separately using discrete levels, while uniformly spaced binning was applied to touch-related variables (e.g., distance to floor, distance to face), head direction, and place (in 5 × 5 grid). For each bin, the average spike rate and SEM were computed across timepoints assigned to that bin. Tuning curves were generated only for bins with at least 50 samples. To assess tuning reliability, we computed cross-validated R^2 values using a five-fold data split following the partitioning strategy described above. Mutual information (MI) between stimulus and response was calculated from the bin-wise firing rate conditional expectations and the stimulus distribution. As a control, shuffled tuning curves were generated by randomly permuting spike-stimulus pairings, and 95% bootstrap confidence intervals were computed across resampled datasets.

To perform feature selection using marginal tuning models to identify the most informative kinematic- or action-related feature for each neuron, we applied marginal tuning models across an extensive feature set (294 total features). For each neuron-feature pair, we computed the tuning curve and assessed its predictive power using cross-validated R^2. Stimuli were binned before computing the conditional mean firing rate in each bin, which served as a lookup table to predict responses on held-out data. Predictions were then compared to actual firing rates, and R^2 was calculated as the proportion of variance explained (see Coefficient of Determination). This process was repeated across all candidate features, and the feature yielding the highest cross-validated R^2 was selected as the neuron's best-tuned feature. To avoid noisy estimates from low sample counts, negative or out-of-bound R^2 values were clipped between 0 and 1. Units with no marginal tuning model R^2 above 0.05 were considered "untuned". This method provided a simple and interpretable estimate of feature importance for each unit in the recorded population.

Kinematic features were grouped into functional categories to aid in interpretability of tuning analyses. Posture features included static joint angles, body segment pitch and roll angles, and segment heights; Postural dynamics included first and second derivatives of posture features (head, body segment, and joint angular velocity and acceleration, lateral and vertical motion and acceleration); Head angles included head pitch and head roll angles; Touch-related features included proximity measures (e.g., distance-to-floor, distance-to-face); Allocentric (world-centered) features included place and head direction; Movement features included whole-body motion metrics (mean joint speed, locomotion speed). Action features included coarse action categories classified by the supervised action segmentation model (see below). The full list of features and their assigned categories is provided in Ext. Data Figure 2.

Action Segmentation

Ground truth action annotations were manually classified by human labelers using a modified version of the DeepEthogram GUI (Bohnslav et al., 2021), adapted to display multiple camera views simultaneously. Short video epochs were selected to represent each major action class observed in our mouse behavior recordings (see definitions below). Each frame was assigned a

single, mutually exclusive label corresponding to one of eight primary action categories: walk, rear, crouch, sniff, head groom, body groom, scratch, or rest. Frames containing infrequent behaviors, such as wet dog shake, forepaw shake, or stretch, were labeled as other. A total of 477,372 frames (79.6 minutes) were annotated from six mice (3 male, 3 female). Final labels were exported from the GUI as a one-hot logical matrix in CSV format, enabling framewise accuracy calculation for classifier evaluation on held-out test partitions (see below).

We used the following operational definitions of species-typical actions: *Walk*: Alternating forepaw and hindpaw movements for forward locomotion and/or turning. *Rear*: Both forepaws lifted off the ground and the torso height >40% of maximum. *Crouch*: Intermediate torso height (<40%) without other active behaviors. *Sniff*: Either (a) prone posture (resting) with snout lifted to sniff air, or (b) prone posture (walking) with head tilted downward to sniff ground or consume food. *Head Groom*: Use of forepaws to clean the head (ears, face, whisker pad, snout, mouth) or use of the tongue to lick the glabrous side of the forepaw between head grooming bouts. *Body Groom*: Crouched posture with head directed to groom the body (arms, abdomen, back, legs) with mouth, or grooming hindpaws using forepaws and mouth; also includes licking the dorsal surface of the forepaw. Scratch: Hindpaw lifted to rapidly claw (~10-20 Hz) at the head or body, or brought to the mouth to lick between scratching bouts. *Rest*: Sedentary periods in prone posture with low movement speed. *Other*: Rare behaviors (e.g., wet dog shake, forepaw shake, stretch) comprising ~1% of annotated actions.

To perform unsupervised action segmentation from 3D keypoint data, we implemented keypoint-MoSeq (Weinreb et al., 2024) with custom adaptations. Because the original pipeline was designed for DeepLabCut keypoints, we developed a data loader to import DANNCE-generated 3D keypoint coordinates, converting them into a compatible format. We specified anterior and posterior body parts for egocentric alignment and excluded tail keypoints to focus on core body and limb movements. The model was trained on six hours of data (2,160,000 frames) from six mice (2 male, 4 female). PCA was used for initialization, retaining components that captured 90% of variance. To achieve a median syllable duration of ~300 ms, we tuned the stickiness parameter (κ) over a range from 1×10^3 to 1×10^{15}, settling on $\kappa = 9\times10^9$. After 50 AR-HMM initialization steps, the full model was fit using 500 iterations of Gibbs sampling, with convergence tracked via log joint probability. The final model identified 95 unique syllables (e.g., “Syllable 0” to “Syllable 94”). After visual inspection of short behavior video snippets, we assigned semantic labels to each syllable and grouped similar syllables into 25 intermediate clusters (e.g., “rear up,” “rear down”), which were then mapped to nine coarse behavior categories (“walk”, “rear”, “crouch”, “sniff”, “head groom”, “body groom”, “left scratch”, “right scratch”, and “rest”). For syllables with overlapping features (e.g., “rear and sniff”), we assigned the category with higher priority. Model accuracy was evaluated on a separate set of human-annotated frames excluded from training. The unsupervised clustering provided by keypoint-MoSeq was instrumental in discovering the coarse action categories and neural correlations with actions.

To perform supervised action segmentation, we implemented a custom AS-EDTCN model in PyTorch, adapted from a previously described action segmentation architecture (Lea et al., 2017). This encoder-decoder network compresses temporal kinematic dynamics into a low-dimensional

representation via an encoder block, and outputs categorical action labels via a decoder block. The encoder consists of a 1D convolutional layer followed by layer normalization, ReLU activation, max pooling, and dropout for regularization. The decoder mirrors this structure, with upsampling, dropout, 1D convolution, layer normalization, and ReLU activation. The model processes an acausal bidirectional input sequence of 128 frames, downsampled 2x, resulting in a total temporal receptive field of ± 1280 ms. The temporal resolution was downsampled via max pooling across two encoder blocks and restored the original sequence length through two decoder blocks via upsampling. A static attention mechanism, implemented as a trainable einsum operation, reduced the sequence dimension to a single output frame. The predicted label was assigned as the class with the maximum value (argmax). Categorical cross-entropy loss with class weighting was used to prevent overfitting to more frequent action classes. Full training hyperparameters are listed in Supplemental Table S2, including: AdamW optimizer, exponential learning rate decay with warm-up ($\gamma = 0.87$), batch size = 64, pool size = 4, sequence length = 128, hidden sizes = [384, 384], kernel size = 8, learning rate = 1×10^{-6}, L2 regularization = 1×10^{-6}, elastic net ratio (L1:L2) = 0.01:0.99, and dropout = 0.25. Generalization performance was evaluated using leave-one-subject-out validation, ensuring that test data from the unseen subject were held out from training. Accuracy of the supervised (AS-EDTCN) and unsupervised (Keypoint-MoSeq) action segmentation models were assessed by averaging the per-class true positive rate (TPR) across all action classes in the held-out test data. Specifically, the supervised AS models were trained on labeled data from five out of six subjects and tested on the remaining subject to assess generalization. Unsupervised action labels predicted by the Keypoint-MoSeq model were validated on the same labeling dataset (N=6 subjects). The true positive rate (TPR), was defined for each class as the proportion of correctly classified positive examples relative to all actual positive examples:

$$TPR = \mathrm{TP} / (\mathrm{TP} + \mathrm{FN})$$

Here, TP (True Positive) refers to instances where the action was correctly predicted, FN (False Negative) refers to instances where the action occurred but was incorrectly classified as another action. In parallel, the false positive rate (FPR) was computed to quantify the proportion of negative cases (i.e., timepoints not belonging to the given action class) that were incorrectly classified as positive, where FP (False Positive) refers to timepoints wrongly predicted as belonging to the class, and TN (True Negative) refers to timepoints correctly predicted as not belonging to the class:

$$FPR = \mathrm{FP} / (\mathrm{FP} + \mathrm{TN})$$

Overall classification accuracy was calculated as the average TPR across all classes:

$$\text{Accuracy} = \frac{1}{n}\sum TPR$$

To account for the non-uniform distribution of action classes, the cross-entropy loss during supervised training was weighted by the inverse class frequency. This class-balancing approach prevents over-weighting common actions and under-weighting rare actions, thereby ensuring that accuracy metrics reflect performance across the full range of actions.

To identify frequency-domain signatures of specific behaviors, we computed the power spectral density (PSD) of kinematic features across stereotyped actions. Kinematics from a subset (N=21) recording sessions were aggregated, including keypoint positions and action label identified by our AS-EDTCN classifier (e.g., grooming, scratching, rearing). Kinematic signals were first preprocessed by subtracting a local moving average (midpoint filter) to emphasize relatively fast movements. The resulting signals were then bandpass filtered (0.1–20 Hz) using a third-order Butterworth filter. For each behavior and feature (e.g., head height, forepaw height, etc.), PSDs were estimated using Welch's method (256-point window, 128-point overlap, 512-point FFT) with a Hamming window. PSDs were averaged across behavior episodes and plotted as frequency-by-feature matrices to visualize unique kinematic "fingerprints" associated with each action class.

To estimate the continuous phase (i.e., polar position within the action cycle) of ongoing actions, we manually selected a representative kinematic feature for each action whose dynamics reflected its stereotyped, periodic trajectory. These features were chosen based on prior PSD analysis and neural correlations to capture the dominant motion and neural encoding associated with each action. Each selected feature was bandpass filtered using a third-order zero-phase Butterworth filter, with filter parameters tuned to the dominant frequency range of each action. The instantaneous phase was computed using the Hilbert transform, yielding continuous action phase ranging from $-\pi$ to π, defined only within bouts of the corresponding action:

Walk: difference in hindpaw anteroposterior (AP) position; 1-3 Hz
Rear/Crouch: head height; 0.1-2 Hz
Sniff: head pitch; 2-8 Hz
Head groom: mean forepaw AP position; 0.5-2 Hz
Body groom: inverted head height; 2-8 Hz
Left/Right scratch: left/right hindpaw height; 0.5-2 Hz
Rest: not assigned (not phasic)

To discretize these continuous phase signals, the phase cycle of each action was divided into three equal-width phase bins corresponding to the rising, peak, and falling phases of the stereotyped trajectory. This procedure produced $3\times N+1$ total bins for the $N = 8$ phasic actions, with an additional bin assigned to "rest", resulting in 25 discrete action phase bins.

To identify functional ensembles of neurons tuned to specific actions, we classified units based on their action-specific modulation strength using z-scored tuning values. Tuning to coarse action categories (N=9 categorical bins) was first computed (see *Tuning Curve Calculation*), and the motion-independent component of tuning (motion-subtracted residual action tuning) was used for thresholding. We computed action modulation as the total modulation of the Z-scored action tuning curve (max-min). This approach allowed us to categorize neurons into ensembles for each action. A unit was classified as "activated" by a particular action if its positive z-scored modulation was greater than a positive threshold (0.25 s.d.), "suppressed" if its negative modulation was below a negative threshold (-0.25 s.d.), or "modulated" if the total modulation exceeded the threshold (0.25 s.d.). This effect size threshold-based method has the advantage of non-exclusive ensemble membership, and thus a neuron could be part of multiple ensembles if it exhibited

strong modulation to more than one action. This contrasts with clustering methods, which typically enforce mutually exclusive assignment and may obscure shared functional networks across actions.

Phase Tuning Analysis

To analyze phase-aligned peri-event time histograms (PETHs), we aimed to characterize the temporal relationship between neural activity and ongoing cyclic actions. We computed phase-aligned PETHs by aligning neural activity and stimulus features to the onset and offset of actions. We then aligned spike rates and characteristic action features (e.g., head height, hindpaw height) to individual phase cycles of the action, enabling averaging across repetitions despite variable action durations. Continuous spike rates were estimated from binned spike counts using a Gaussian smoothing kernel with width equal to half the phase cycle duration. Spike rates were normalized by z-scoring the log-transformed firing rates (using the numerically stable `log1p` function) across the session. Action cycles were first segmented based on the action category and further separated into distinct action cycles based on action phase (see Action Phase Calculation). For each cycle, the representative stimulus feature (e.g., hindpaw AP position for walking, head height for rearing) was extracted and time-warped by interpolating the phase cycle duration to the target duration (100 samples). The corresponding spike rate trace was also interpolated across the same normalized phase axis using linear interpolation. Trials were thus aligned and binned into a fixed number of phase points per cycle (e.g., 100 phase bins ranging from $-\pi$ to π) to enable direct averaging across cycles. Phase-aligned PETHs were computed separately for each action class and unit, allowing fine-grained dissection of firing patterns across stereotyped action cycles. This analysis revealed structured, phase-locked firing dynamics across actions such as walking, grooming, and rearing.

To demonstrate tiling of action phase tuning, we sorted neuron order within each ensemble (e.g., activated, suppressed, modulated) according to their preferred phase. Phase tuning curves were extracted for each neuron using the z-scored firing rate across discrete action phase bins (25 total; see Action Phase Calculation), separately for each action. Tuning curves were upsampled by a factor of 10 and smoothed to reduce sampling noise and allow finer sorting resolution. Units were sorted based on their peak or trough response phase, depending on ensemble type (e.g., peak for activated, trough for suppressed). Interpolated tuning curves were z-scored per unit, normalized to the [0,1] range, and plotted as heatmaps sorted by preferred phase. This visualization revealed structured tiling of phase-tuned responses across actions. Unit IDs for each ensemble, sorted by preferred phase, were used to sort predicted population activity (see below).

To quantify action phase modulation, we computed the z-scored tuning curve for each neuron and extracted the amplitude of modulation across the action cycle. For each action and neuron, Z-scored tuning curves were aligned to the preferred phase using circular shifting. Modulation was quantified as the difference between the maximum and minimum response across the phase cycle. Modulation was also computed for shuffled data, in which spike rates were randomly permuted across phase bins, providing a baseline for comparison, as alignment to peak phase will always produce a sequence, albeit with minimal modulation. Finally, for each ensemble and

region, we computed an action–phase modulation matrix capturing how, on average, neurons in each action ensemble modulated activity across all other action phase cycles to evaluate phase-specific ensemble recruitment.

Encoding Models

We performed static data partitioning to avoid temporal auto-correlation. Each one-hour recording session was divided into ten continuous 6-minute segments ("deciles"). For five-fold cross-validation, each decile was split into 80% training and 20% test data. To prevent temporal overlap, a three-second buffer around each training–test boundary was excluded from both sets, following the approach in Syeda et al., 2024. Partition masks were fixed (non-random) and reused consistently across all encoding model training and evaluation procedures.

Regression model (neural encoder) performance was assessed by computing the coefficient of determination (R^2) on held-out test data, where y is the true response, $\hat{y}$ is the predicted response, and $\bar{y}$ is the mean response:

$$\textit{Residual sum of squares:} \quad SSR = \sum(y - \hat{y})^2$$

$$\textit{Total sum of squares:} \quad SST = \sum(y - \bar{y})^2$$

$$\textit{Coefficient of determination:} \quad R^2 = 1 - SSR \,/\, SST$$

We trained a core–readout encoder-decoder temporal convolutional network (EDTCN) to predict neural activity solely from kinematic features (with no action information). In benchmark testing, this architecture outperformed alternative models, including GLMs, gradient boosted decision trees (XGBoost), multilayer perceptrons (MLPs), Kolmogorov-Arnold networks (KAN), LSTMs, and other CNNs on a neural regression (encoder) task. The model consisted of a shared temporal convolutional encoder ("Core") that transformed sequences of kinematic inputs into a lower-dimensional latent representation, and a set of session-specific fully connected decoders ("Readouts") that mapped this latent activity onto population neural responses. The encoder comprised two layers of bidirectional 1D convolutions (kernel size = 8, ReLU activation, dropout = 0.25, hidden size = 384), each followed by max pooling (pool size = 2), resulting in a total temporal receptive field of ±640 ms. Input features included joint angles, head orientation, paw, head, and trunk positions, and their temporal derivatives. These kinematic features were z-scored per session using statistics from the training set. Neural firing rates were normalized between 0 and 1. The model was trained across 392 recording sessions from multiple brain regions (M1, PPC, S1, S2, V1, HPC) using mini-batch gradient descent (AdamW optimizer; batch size = 128; initial learning rate = 1×10^{-5}; linear warmup schedule). Training was performed in mixed precision (FP16) and regularized by elastic net (L1:L2 ratio = 0.1:0.99). Model loss was computed as the root mean squared log error (RMSLE) between predicted and observed firing rates. To prevent overfitting, we used five-fold cross-validation with early stopping. For each fold, preprocessed data were stored as memory-mapped arrays to optimize input/output performance. Final model accuracy was assessed using the coefficient of determination (R^2), computed independently for each neuron.

We used t-SNE embedding to perform dimensionality reduction on predicted population activity. To predict population neural activity from the kinematic inputs in an example session, we first applied the trained Core EDTCN module to predict latent representation, followed by a Readout layer concatenated across all recorded sessions, outputting a total of 8,763 RS units. The resulting population activity matrix (timepoints × units) was Z-scored across time and subjected to principal component analysis (PCA) for dimensionality reduction. The first ~140 principal components (PCs), explaining at least 90% of the variance, were retained. These components were then embedded into two dimensions using the openTSNE implementation of t-SNE (https://github.com/pavlin-policar/openTSNE), using perplexity = 500 and default early exaggeration settings optimized for visualization. Each point in the resulting 2D embedding represents a population activity vector at a single timepoint. To visualize clustering with respect to action category, each point was color-coded by action label, based on the AS-EDTCN-predicted discrete action identity at that timepoint. This procedure enabled visual inspection of action-specific structure in the predicted neural state space, capturing separable trajectories and clusters corresponding to different species-typical actions.

Additional References

50. Bohnslav, J. P. *et al.* DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. *eLife* **10**, e63377 (2021).
51. Syeda, A. *et al*. Facemap: a framework for modeling neural activity based on orofacial tracking. *Nat. Neurosci.* **27**, 187–195 (2024).
52. Takatoh, J. *et al.* The whisking oscillator circuit. *Nature* **609**, 560–568 (2022).

Acknowledgements

We thank members of the Wang and Dunn laboratories for valuable discussions and feedback. We thank Paul M. Thompson and Seonmi Choi for technical assistance. This work was supported by NIH grants DP1 NS137188 to F.W. and F32MH122995 to K.S.S.

Author contributions

F.W. and K.S.S. conceived the project. K.S.S. and F.W. designed experiments with input from T.W.D. K.S.S. performed all experiments with assistance from W.X., H.J., and M.S.L. K.S.S. analyzed the data with help from J.L., H.J., T. Lou, M.S.L., and W.X. T.W.D., T. Li, and K.S.S. developed the DANNCE models. J.L. and K.S.S. developed the geometric anatomical framework. H.J., T. Lou, K.A.C., and W.X. assisted K.S.S. with action segmentation models. T. Lou, H.J., and T. Li assisted K.S.S. with the neural encoder models. K.S.S. and F.W. wrote the manuscript with input from all authors. F.W. supervised the project.

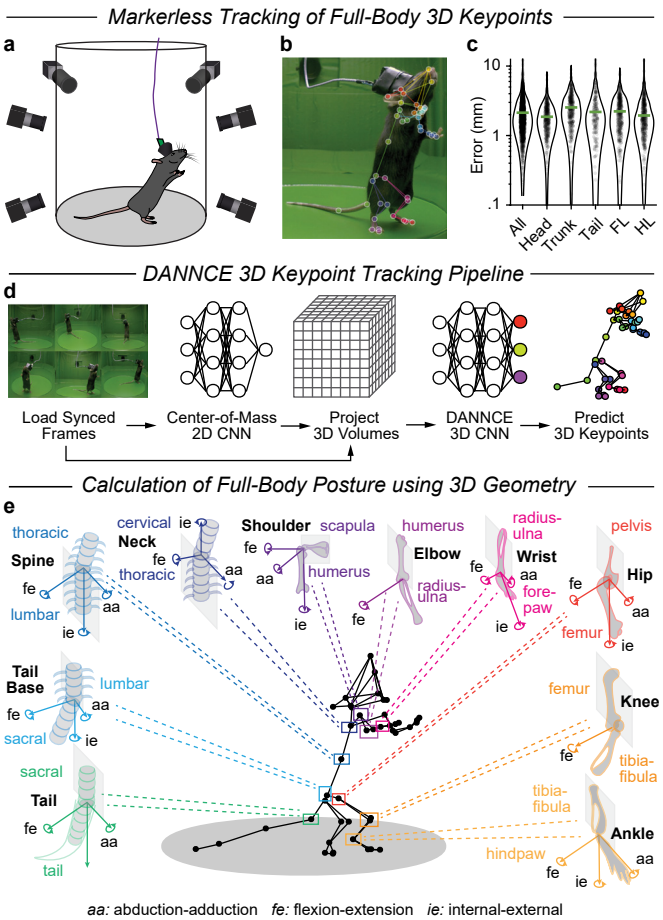

Figure 1. Tracking full-body kinematics with 3D geometric models.
a, Markerless motion capture setup with six synchronized cameras surrounding a freely-moving mouse in an enclosed arena. **b,** Single-camera frame with DANNCE keypoint predictions (colored circles) connected by links (colored lines). The mouse was implanted with a tethered 64-channel microdrive for simultaneous electrophysiology. **c,** Euclidean distance between human-labeled and DANNCE-predicted 3D positions across validation frames (n=60 frames, 12 mice); green lines, median error for each body part; mean error 2.38 mm across body parts. **d,** DANNCE markerless pose estimation applies 2D CNNs to multi-view frames to triangulate the 3D center-of-mass (COM), projects them into a COM-centered volumetric grid via projective geometry, and uses a 3D CNN to predict 3D keypoint positions. **e,** Joint-centered geometric skeletal model estimating joint angles via 3D trigonometry from the rotation of the distal body segment relative to each proximal segment.

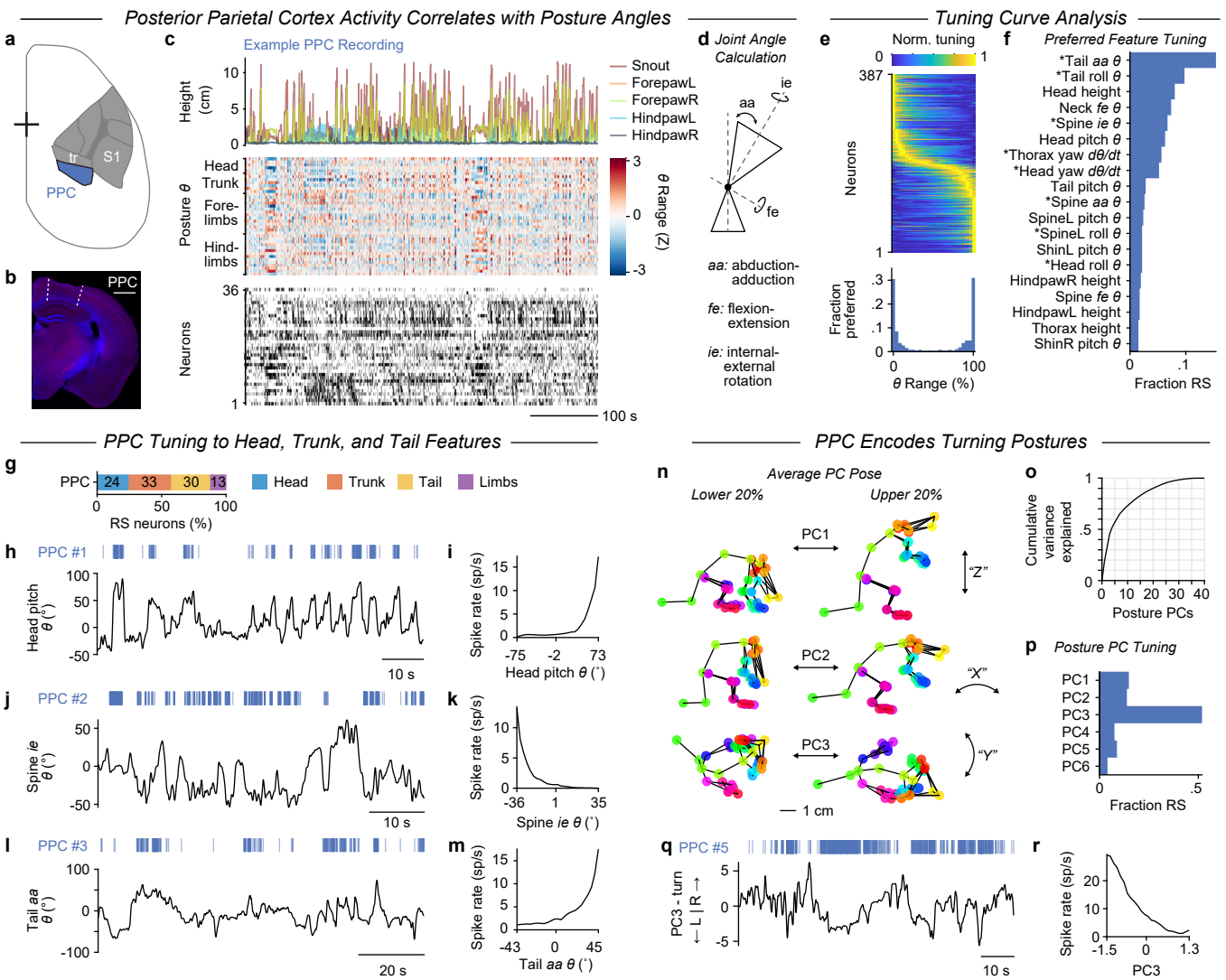

Figure 2. PPC efficiently encodes turning postures through preferential tuning to the angles of midline body parts.
a, Dorsal view of mouse right hemisphere with medial PPC outlined. Tetrode microdrives were chronically implanted in right PPC, posterior to the empirically mapped S1 trunk (tr) region identified via intrinsic optical signal imaging. **b,** Cortical section of the right hemisphere after tetrode removal, showing the targeted PPC region. Scale bar, 1 mm. **c,** Time-aligned example traces. Top, keypoint height (cm) of distal body parts. Middle, Z-scored time series of 40 posture angles. Bottom, spike raster of 36 simultaneously recorded PPC neurons, sorted by Rastermap; grayscale firing rates were Z-scored and saturated (range: 0–0.1) to highlight activations. **d,** Schematic of abduction-adduction (aa), flexion-extension (fe), and internal-external (ie) rotation angles measured at the joint (black circle). **e,** Top, normalized tuning curves for the best-tuned posture variable across all posture-tuned RS neurons in PPC, sorted by preferred angle bin. Bottom, histogram of preferred angle percentile bins. Preference toward the extremes suggests efficient coding. **f,** Histogram of best-tuned features for all kinematics-tuned RS neurons (n=461), sorted by frequency. *, turning-related features. **g,** Distribution of best-tuned kinematic features grouped by body part: head (blue), trunk (orange), tail (yellow), limbs (purple), for the same RS neurons in (f). **h–i,** Example spike raster (h) and tuning curve (i) for neuron #1 aligned to head pitch. **j–k,** Same for neuron #2 aligned to spine ie rotation. **l–m**, Same for neuron #3 aligned to tail aa joint angle. **n,** Average poses from bottom and top 20th-percentile bins of first three Posture PCs. **o,** Cumulative variance in posture angles explained by Posture PCs. **p,** Histogram of tuning preference to top six Posture PCs for all PC- or posture-tuned RS neurons (n=401); PC3 captured turning postures. **q–r,** Spike raster (q) and tuning curve (r) for example neuron #5 aligned to PC3 (turning axis; negative = left, positive = right).

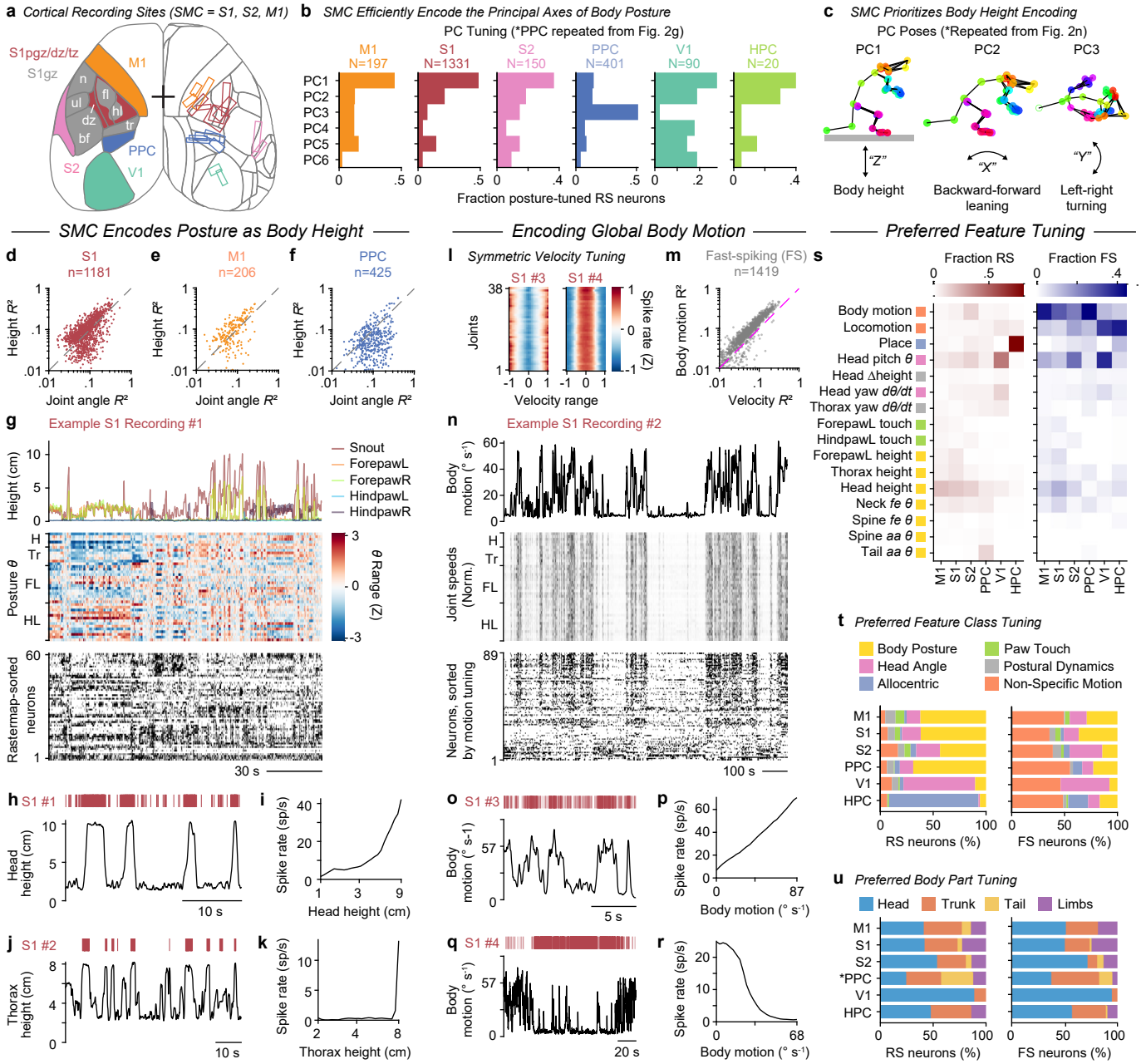

Figure 3. Cortex encodes global posture and motion features in a gravicentric reference frame.
a, Brain atlas based on the Allen CCF (gray outlines), with bregma marked by a black crosshair (total length = 1 mm). All recordings were performed in the right hemisphere, with targeted regions outlined: M1 (orange), S2 (pink), PPC (blue), V1 (green), and S1pgz/dz/tz (red). For reference, granular S1 (S1gz, dark gray) shows somatotopic zones for the snout (sn), barrel field (bf), trunk (tr), hindlimb (hl), and forelimb (fl). **b,** Tuning preference histograms for the top six Posture PCs across posture- or PC-tuned RS neurons in SMC. **c,** Averaged poses from first three Posture PCs, representing body height (Z), pitch (X), and turning (Y). **d–f,** Scatter plots comparing variance in firing rates explained by the best joint angle model versus best segment height model in S1 (d), M1 (e), and PPC (f). **g,** Time-aligned example traces. Top, keypoint height (cm) of distal body parts. Middle, Z-scored posture angles, grouped by body part: head (H), neck/trunk (Tr), forelimbs (FL), and hindlimbs (HL). Bottom, spike raster of 60 simultaneously recorded S1 neurons. **h–i,** Example spike raster (h) and tuning curve (i) for S1 neuron #1 aligned to head height. **j–k,** Same for S1 neuron #2 aligned to thorax height. **l,** Z-scored tuning curves for all joint angular velocities (dθ/dt) for example S1 neurons #3 and #4. **m,** Scatter plot comparing variance in tuned FS neuron firing explained by body motion versus joint angular velocity (see Methods). **n,** Top, example body motion trace. Bottom, spike raster of 89 simultaneously recorded neurons. **o–p,** Example spike raster (o) and tuning curve (p) for S1 neuron #3 activated by body motion. **q–r,** Same for S1 neuron #4, suppressed by body motion. **s,** Histograms quantifying fraction of RS (left) and FS (right) neurons preferring each of the top selected features. Colored squares next to the y-axis labels indicate feature class (see legend in t). **t,** Distribution of preferred feature classes for RS (left) and FS (right) neurons across cortical regions. **u,** Histograms of preferred tuning to head-, trunk-, tail-, or limb-related features for RS (left) and FS (right) neurons across regions. *, Data from PPC RS neurons in b, c, and u are repeated from Fig. 2 for comparison.

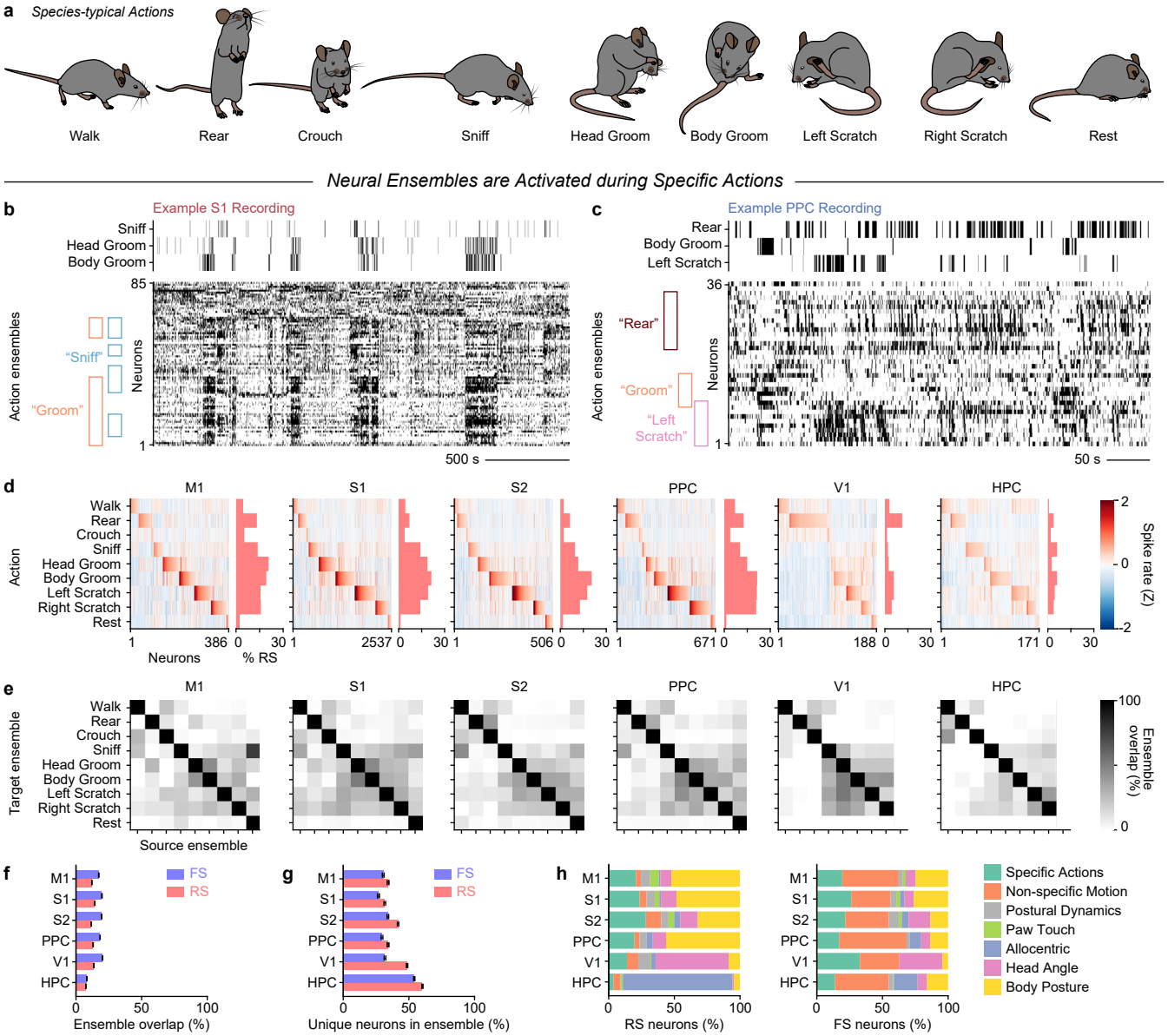

Figure 4. Prominent action ensembles in SMC and PPC.
a, Nine common species-typical actions observed during spontaneous behavior. **b,** Example snippet from behavioral and S1 recordings in a mouse. Top, ethogram showing occurrence of "sniff", "head groom", and "body groom." Bottom: spike raster aligned in time, sorted by Rastermap. For illustrative purposes, clusters of neurons aligned to repeated actions were qualitatively labeled as putative "action ensembles." **c,** Same as (b), but for PPC activity aligned to "rear," "body groom," and "left scratch." **d,** Heatmaps: mean Z-scored spike rate of action-activated RS neurons (columns) for each action (rows). Histograms: percentage of RS neurons in each region belonging to each (non-exclusive) action ensemble. **e,** Ensemble overlap grids quantifying the percentage of RS neurons in each source ensemble also present in target ensembles. **f,** Mean percentage pairwise ensemble overlap for FS (blue) and RS (red) neurons. Error bars: 95% CI. **g,** Mean percentage unique FS (blue) and RS (red) neurons per ensemble not shared with others. Error bars: 95% CI. **h,** Distribution of preferred feature classes for RS (left) and FS (right) neurons across cortical regions. Same format as Fig. 2t, but recalculated with the addition of the Specific Actions feature class.

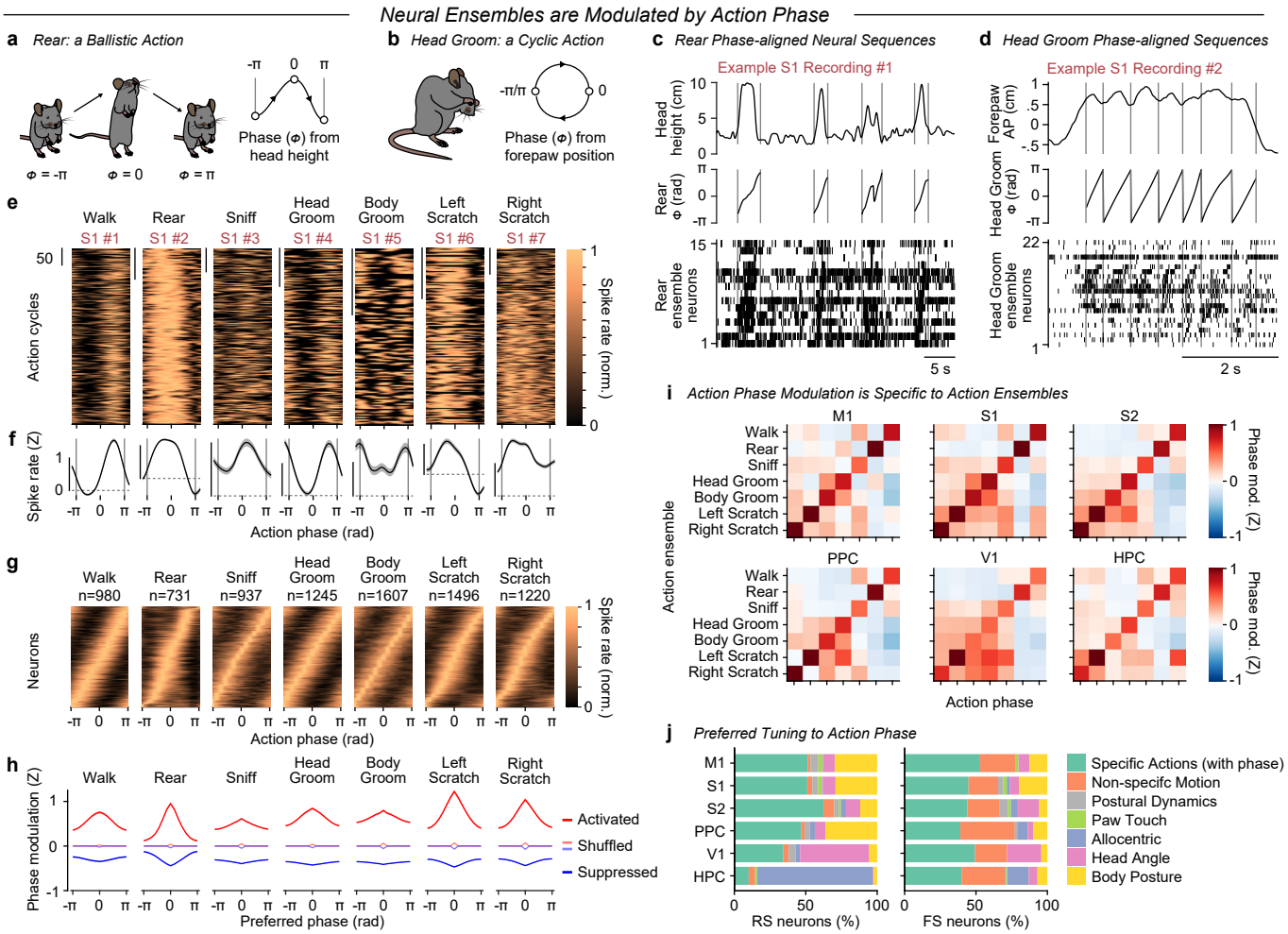

Figure 5. Action ensembles encode action phase as neural sequences.
a, Illustration of the ballistic Rear action, with phase defined by head height. **b,** Illustration of the cyclic Head Groom action, with phase defined by forepaw anterior-posterior (AP) position. Phase resets from π to –π at each cycle offset/onset. **c,** Example S1 recording containing multiple rearing events. Top, head height; middle, rear phase; bottom, raster from 15 simultaneously recorded neurons in the Rear ensemble. Vertical gray lines mark Rear onset and offset. **d,** Example S1 recording containing a single head grooming bout. Top, mean forepaw AP trace; middle, Head Groom phase; bottom, raster from 22 simultaneously recorded neurons in the Head Groom ensemble. Vertical gray lines mark Head Groom cycle onset and offset. **e-f,** Peri-event time histograms (PETHs) aligned to action phase cycles for seven example S1 neurons tuned to the seven phasic actions. Individual PETHs are normalized 0–1 within each cycle (e), and corresponding average firing rates are shown as Z-score normalized traces (f), with shading ±SEM. **g,** Action phase tuning curves for neurons in each activated action ensemble concatenated across cortical regions, normalized to range 0-1 for each neuron and sorted by preferred phase. **h,** Average phase modulation (Z-score normalized), centered on each neuron's preferred phase, for action-activated (red) and action-suppressed (blue) neurons. Shuffled controls in light red/blue. **i,** Phase modulation grids for all combinations of action ensemble (rows) and action phase (columns) for each cortical region. **j,** Distribution of preferred feature classes for RS (left) and FS (right) neurons across cortical regions. Same format as Fig. 4h, but recalculated after including action phase in the Specific Actions feature class (see Methods).

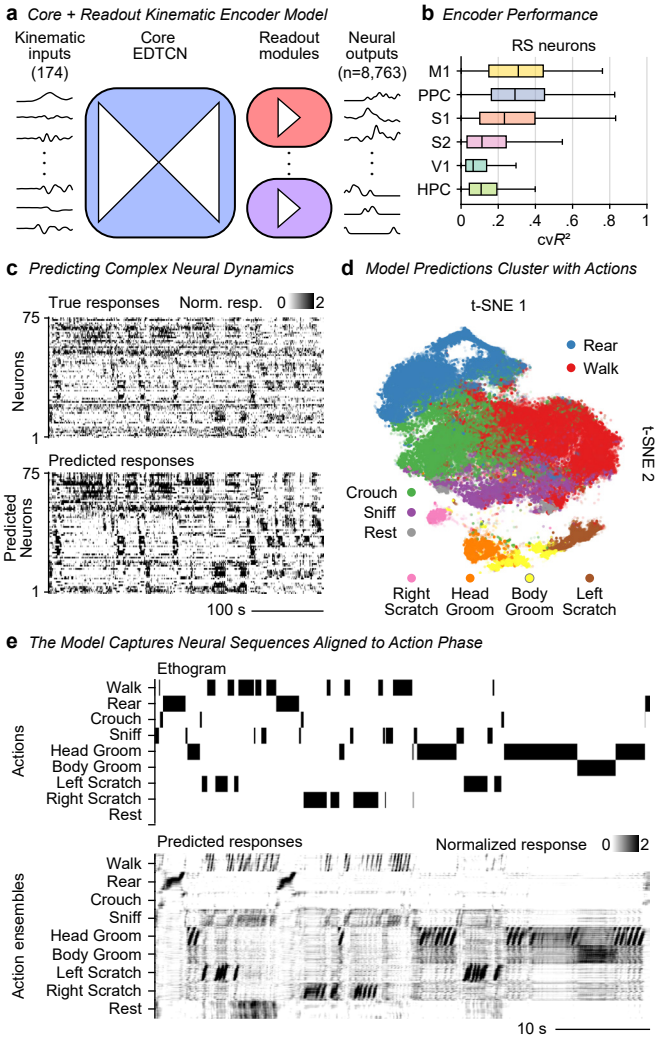

Figure 6. Action phase-aligned neural sequences recapitulated by a kinematic encoder model.
a, Architecture of the Core+Readout kinematic encoder model. A batch of kinematic features is inputted to a shared encoder–decoder temporal convolutional network (EDTCN "Core"), producing 384 latent features. These latent features are then mapped through session-specific Readout layers, each projecting into the neurons recorded in that session (total n=8,763 neurons across 392 sessions; number of neurons varies by session). During training, input features and Readout mappings were repeated for all sessions, allowing the Core to learn a shared kinematic representation, while each Readout module captured the neural translation. **b,** Box-and-whisker plot quantifying Core+Readout encoder model cross-validated fit (cvR²) for RS neurons recorded in each cortical area. Boxes represent interquartile ranges, lines mark median cvR², and whiskers span the full distribution. **c,** Example session showing true neural responses (top; Z-score normalized to 0–2 s.d.) and predicted responses (bottom; same normalization and sorting by Rastermap), capturing much of the dynamics in recorded data. **d,** t-SNE embedding of the predicted population activity. Dots represent time points, colored according to the nine actions classified by an independent action segmentation model. **e,** Example epoch showing the action ethogram (top) and predicted responses (bottom) for action ensembles (Walk: n=322; Rear: n=601; Sniff: n=917; Head Groom: n=1146; Body Groom: n=1436; Left Scratch: n=1243; Right Scratch: n=1016). Neurons are sorted by preferred phase tuning (as in Fig. 5g). Predicted responses are Z-score normalized to 0–2 s.d. to emphasize activations.