Revealing motor behavior patterns and functional improvement trajectories following spinal cord injury using unsupervised machine learning

ABSTRACT

Spinal cord injury (SCI) disrupts motor function, with recovery involving complex adaptive motor behaviors that cannot be fully assessed with traditional manual recovery assessments and their predefined recovery criteria. To move beyond these constraints, we employed motion sequencing (MoSeq), a machine learning-driven framework that automatically segments freely moving mouse behavior into discrete, repeated behavioral “syllables”, effectively capturing entire behavioral repertoires and dynamic behavioral sequences observed throughout SCI recovery. We observed that functional improvements involved the reemergence of preinjury behaviors, emergence of adaptive behaviors, and subtle individual differences between mice. Interestingly, the temporal organization of behavioral sequences evolved consistently across injury severities, suggesting a ceiling effect in sequence reorganization. MoSeq syllables also correlated with the Basso Mouse Scale (BMS) and outperformed BMS in correlation to several injury-related histological findings. We also introduced the “Recovery Score,” a new metric derived from aggregated MoSeq data that quantifies recovery based on syllable usage patterns. The Recovery Score identified clusters of syllables with high usages, which highlights the potential of using hallmark syllables as behavioral biomarkers for assessing recovery and informing therapies. Future research should incorporate more complex experimental designs to refine this aggregation approach and advance the pursuit of a new gold standard in recovery assessment.

INTRODUCTION

Spinal cord injury (SCI) induces profound changes to whole-body movement, with recovery involving complex adaptive motor behaviors[1–8]. Effective assessment of this recovery necessitates advanced techniques capable of capturing not only the intricacies of individual behaviors but also the sequences and patterns that define coordinated movement. Traditional rodent sensorimotor assays, such as the Basso, Beattie, Bresnahan (BBB) and Basso Mouse Scale (BMS) open field locomotor scales[9,10], horizontal ladder walking task[11–14], rotarod[15,16], grid walk[17], beam walking tests[18], force platform/arena methods[19], offer evaluations of locomotor function that correlate with gross neural damage and plasticity. However, these methods are limited by reliance on subjective human observation, which constrains their capacity to detect subtle or diverse motor patterns and often reduces behavior to isolated actions[9,10,20–22]. Advances in computer vision and machine learning, including tools like DeepLabCut[23,24], SLEAP[25], and OpenPose[26], have improved the precision of kinematic and gait analyses by automating movement tracking and reducing variability. Yet, these approaches still fall short in capturing the broader, integrated context of movement, particularly the continuous and coordinated nature of motor behavior orchestrated by the nervous system[27–29]. To overcome these limitations, a shift is needed toward frameworks that examine recovery by capturing comprehensive behavioral repertoires and the dynamic sequences underlying functional restoration.

One powerful approach for addressing these gaps is motion sequencing (MoSeq), a computer vision-based unsupervised machine-learning framework[28–33]. MoSeq analyzes freely moving mouse spontaneous behavior by segmenting it into discrete, repeated units called "syllables," such as grooming, locomotion, and rearing. These syllables provide an objective frame-by-frame description of behavior, enabling the precise capture of the spatiotemporal dynamics of behavior. MoSeq is grounded in principles of ethology[27,34,35], which posit that behavior is organized into modular components that can be flexibly assembled into meaningful sequences. This ethological perspective is particularly relevant in the context of SCI, as it provides a framework for examining how injury and recovery processes reorganize these behavioral modules into functional patterns over time.

Unlike traditional methods, MoSeq does not rely on predefined criteria[10,20], allowing for the unbiased identification and tracking of behavioral patterns. Its fine-grained temporal resolution further enables the linkage of observed behaviors to the responsible underlying neurobiology[29,31,32]. This is especially valuable in the context of SCI as central nervous system changes are key determinants of recovery extent[36–39]. Moreover, the technological

flexibility of MoSeq supports high-resolution, longitudinal analyses, which make it particularly well-suited for capturing the evolution of complex motor behaviors throughout the recovery process.

In this study, we demonstrate the utility of MoSeq in analyzing the evolution of motor behaviors following SCI. MoSeq revealed that, post-SCI, mice perform pre-injury behaviors within the physical constraints imposed by their injuries. This then allowed for the classification of behaviors into "normal," "injured," or "paralyzed" categories. While the usage of specific syllables was primarily influenced by injury severity and individual mouse identity, the temporal evolution of functional sequences was driven by timepoints post-injury. These results show that high-resolution behavior description from MoSeq can deconvolve various factors influencing functional recovery. Importantly, syllables aligned with established BMS[10] scores, but provided greater granularity and captured behaviors often not measurable by traditional scoring systems, such as grooming and rearing. Furthermore, lesion correlations demonstrated that syllables outperform BMS in several injury-related metrics. Quantitative analysis of syllable usages is a more precise scoring system as compared to BMS and revealed that distinctive behavior trajectories occur among different injury severities. Finally, syllables form clusters spanning from paralysis to normality, indicating different aspects of recovery and mobility promotes specific subsets of behaviors. From these syllable usage clusters, we observed that normal, but less neurally and physically demanding, behaviors are achievable for animals with injuries.

Using MoSeq, we uncovered motor behavior patterns that emerge post-SCI with a better precision and granularity than traditional scoring methods like BMS. The resulting high-resolution behavior descriptions from the MoSeq framework offer unprecedented insights into the interplay between injury severity, functional recovery, and the adaptive strategies animals employ to navigate their post-injury constraints.

RESULTS

Advanced computational approaches offer high-resolution investigation of behavior during recovery from spinal cord injury (SCI)

To capture and quantify behavioral changes following SCI and throughout recovery, we used a computational approach to analyze behavior patterns across a range of injury severities. We utilized four SCI models (**Fig. 1a**) — Transection ("Tx"), Moderate ("Mod", 50 kDyn) and Mild ("Mild", 30 kDyn) contusions, and Sham injuries (laminectomy only) — along with a Naive group for comparison, to capture the full spectrum of behaviors ranging from paralysis to normality over a 10-week recovery period. Behavioral recordings were conducted at key timepoints[10,40]: preinjury (before surgery), 2 days post-operation (DPO), 7 DPO, and weekly thereafter up to 70 DPO (10 weeks total). These SCI models were chosen to induce varying degrees of neural circuit damage and inflammatory responses[22,41,42], which would then allow us to examine a spectrum of behaviors that emerge post-SCI and to capture the temporal dynamics of recovery across different severities.

To identify phenotypes of functional recovery, we modified the existing MoSeq framework[33] to deconstruct complex behavior pre- and post-SCI. To evaluate the application of the MoSeq framework in SCI mice, we compared syllables to the gold standard for assessing locomotor deficits and recovery in SCI mice: the manually scored BMS[10]. MoSeq employs a downward-facing depth camera to capture spontaneous behavior in freely behaving mice in an open-field environment. The framework uses a machine vision method that analyzes 3D depth videos and segments the videos into distinctive, modulated, and repeated behavior motifs by the pose dynamics. These behavioral motifs are called “syllables.” Each syllable is represented and interpreted according to its frequency of usage, i.e. how often each particular behavior is performed. The existing MoSeq data extraction tools could not support SCI mice whose physical constraints of the injury, such as trunk and limb dragging, posed a challenge for the framework’s classical computer vision methods. Specifically, during the thresholding step, the lower trunk and limbs of injured mice were classified as noise and filtered out due to their significantly lower positioning compared to uninjured mice. To address this, we adapted the data processing step by leveraging the Detectron2 implementation of multi-task RCNN[43] (**Fig. 1b and Supplemental Fig. 1ab, see methods**), which then enabled accurate detection of the entire body of SCI mice without omitting any body parts. This adaptation allowed the MoSeq AR-HMM to better learn behavioral patterns, capturing the cessation of normal syllables and the emergence of neomorphic (injured and paralyzed) syllables from the more accurately extracted mouse data.

In both male and female models, we identified over 90 unique syllables (**Supplemental Tables 1-2**). Syllable similarities were assessed by calculating the pairwise distances between autoregressive matrices, which capture the dynamics of each syllable. Hierarchical clustering was then applied to group syllables based on these similarities (**Fig. 1c and Extended Data Fig. 1a**). Syllables were post-hoc labeled with names and semantic descriptions like "rear," "locomotion," and "groom," in normal and injured mice, by visually inspecting raw videos and the hierarchical clusters. Following injury, we observed that mice execute movements prior to the injury, albeit within the physical limitations imposed by their injury. Accordingly, we categorize syllables as "normal," "injured," or "paralyzed," based on the functional capacity of hindlimb movements. Here we define "normal syllables" as those observed preinjury, by the Naive group, and largely by the Sham group. As expected, numerous normal syllables reappear as an animal functionally recovers[44,45]. “Injured syllables" denote weightbearing movements that emerge during recovery and are absent in non-injured animals, such as instable locomotion[46–48]. These syllables may represent compensatory mechanisms developed to adapt to functional challenges posed by the SC. Lastly, "paralyzed syllables" refer to behavioral movements occurring while an animal drags its hindlimbs[49–52]. Paralyzed syllables are enriched in behavior expressed by the Transection group as well as shortly following injury in the Mild and Moderate groups. Semantic names in conjunction with functional capability were used to identify 17 syllable categories that spanned normal, injured, and paralyzed behaviors (**Supplemental Tables 1-2**).

MoSeq pipeline uses Principal Component Analysis (PCA)[53] to reduce the dimensionality of the data and the model takes the top 10 principal components as input (**Fig. 1d and Extended Data Fig. 1b**). PC1 embeds primarily the animal’s height and PC2 embeds body-wide dynamics. Instances with the same syllable label have similar trajectories in the 2D PC1 and PC2 histogram. The major differentiator between these normal, injured, and paralyzed versions of each syllable category is the height of the mouse off the ground (**Fig. 1e and Extended Data Fig. 1c**). Spinograms[54,55] further emphasize differences in animal height between normal, injured and paralyzed syllables execution (**Fig. 1f and Extended Data Fig. 1d**). Further validating our findings, functional recovery is denoted as the decrease in usages in a subset of paralyzed and injured syllables and the increase in usage in a subset of normal syllables (**Fig. 1g and Extended Data Fig. 1e**). In addition to syllable usages, the concatenation of syllables into behavioral sequences are different among injury types. UMAP analysis of syllable sequences reveals that normal, injured, and paralyzed syllables exhibit distinctive structural patterns, which highlights that these different types of syllables are embedded within fundamentally different behavioral sequences (**Fig. 1h, Extended Data Fig. 1f, Supplemental Fig. 1cd**).

In sum, MoSeq provides high-resolution descriptions of recovery following SCI. We discovered that different injury types lead to distinctive behavioral patterns, including dynamic changes in syllable usages and sequences. Moreover, behavior evolves significantly over the course of recovery, which emphasizes the complexity of adaptive motor responses after injury.

Behavioral variation following SCI is primarily influenced by reflects injury severity and mouse identity

We can analyze the recovery post-SCI by looking at the behavior syllable usages over time to characterize the behavior patterns and the variations among them. Specifically, we aimed to disentangle the relative contributions of injury severity, mouse individuality, and time post-injury in shaping mouse behavior throughout recovery. To achieve this, we first validated the robustness of our MoSeq approach, ensuring that observed differences were biologically meaningful and aligned with the context of injury and recovery. As expected[56], when examining traditional metrics such as total distance traveled and maximum velocity within a session, naive and sham animals exhibited similar mobility, while injured animals showed a marked decline immediately following injury. A subset of injured animals demonstrated gradual and partial recovery of motor functions over time (**Fig. 2a and Extended Data Fig. 2a**).

We then evaluated whether differences in metrics not typically assessed during recovery, but captured by MoSeq, reflected genuine biological phenomena (**Fig. 2b and Extended Data Fig. 2b**). The purpose of this analysis was to determine whether a broader spectrum of movements aligns with established knowledge of functional recovery following SCI, which is typically characterized by injured animals striving to return to their preinjury (normal) state while simultaneously deviating from normalcy due to the physical constraints of the injury[46–48]. Some of these individual movements might be important in understanding intrinsic animal states, yet they are often overlooked in standard assays due to the labor-intensive nature of manual scoring and its high susceptibility to error[9–18]. For example, we observe that grooming-like behaviors (referred to hereafter as

“grooming” syllables) are enriched in paralyzed animals, gradually diminish over time during recovery, and remain persistently elevated in animals with permanent paralysis (**Fig. 2c and Extended Data Fig. 2c**). This observation is consistent with prior studies on excitotoxic SCI and the onset of mechanical allodynia, which have shown that animals experiencing hypersensitivity following SCI display elevated grooming behaviors[57]. Importantly, discomfort, such as pain and disrupted sensorimotor functions, might play a significant role in functional recovery as it reflects underlying states that could influence an animal’s capacity for adaptation and engagement with its environment[58].

As recovery unfolds, patterns emerge not only within individual behaviors but also across the entire behavioral repertoire. UMAP analysis reveals distinct groupings of behaviors as they evolve during pre- and post-SCI with data points from similar timepoints and injury types aligning closely (**Fig. 2d and Extended Data Fig. 2d**). Normal, uninjured animals occupied one end of the UMAP distribution, while paralyzed animals were positioned at the opposite end. Between normal and paralyzed, lay a clear recovery gradient among different groups. These findings demonstrate a well-recognized yet not often evaluated phenomenon[59,60]: behaviors (like grooming) typically considered unrelated to functional recovery metrics (like locomotion) evolve alongside locomotor behaviors during recovery. Perhaps even more importantly, the clusters in UMAP shows that different groups may have very different behavior patterns beyond basic locomotion. Examining the entire behavioral repertoire could thus enrich the understanding of the recovery process.

The differences in behavior diversity among groups can be measured using entropy, a measure that quantifies the degree of randomness[61] (see methods). We observed that animals with severe hindlimb deficits exhibit rigid movement patterns, with this rigidity persisting in those with permanent paralysis (**Fig. 2e and Extended Data Fig. 2e**). More severely injured animals diverged significantly from their preinjury state and remained distinct over time. In contrast, those with mild or sham injuries became more similar to their preinjury state and maintained this stability at later time points (**Fig. 2f, Extended Data Fig. 2f, and Supplemental Fig. 2ab**). This pattern aligns with the expectation that the range of potential behaviors, or syllable usage entropy, decreases in animals with impaired hindlimb function and recovers as function is regained[62].

Linear classifier models demonstrated that animals were more easily distinguished at timepoints when motor deficits were most pronounced between groups (**Fig. 2g and Extended Data Fig. 2g**). Conversely, at timepoints when behaviors overlap among groups, the model's ability to differentiate between them decreased. This observation highlights the gradual convergence or divergence of behavior among recovering animals with different injury severities at different timepoints[41]. We next sought to disentangle the factors driving these behavioral differences. To explore this, we examined syllable usage heatmaps sorted by different variables. Sorting by injury severity and mouse identity produced distinct organized patterns, whereas sorting by time point failed to yield clear structure (**Fig. 2h and Extended Data Fig. 2h**). We applied generalized linear models to further investigate these factors, which showed that behavior patterns were more strongly influenced by injury severity and individual mouse identity than by time post-injury (**Fig. 2i and Extended Data Fig. 2i**). Quantification of variation confirmed that injury severity and mouse identity contributed significantly more to observed behavioral differences than timepoints (**Fig. 2j and Extended Data Fig. 2j**), with this effect being more pronounced in injured animals compared to uninjured ones (**Fig. 2k and Extended Data Fig. 2k**). These findings are consistent with prior studies showing that injury severity is the dominant factor influencing behavior, as it exerts the most substantial impact on overall movement[42]. However, individual mouse variability, often underexplored in efforts to simplify and standardize analyses[63], introduces subtle yet meaningful differences that reflect how each animal uniquely adapts to injury[30,64].

Previous studies demonstrate that behavioral fluctuations following SCI reflect a combination of functional recovery processes[65–68], adaptation to environment[69], and responses to discomfort or pain[70,71]. Building on this foundation, our findings reaffirm the critical role of injury severity in shaping behavioral outcomes, while also drawing attention to the underexplored impact of individual variability on recovery[30,63,64]. Thus, MoSeq helps to disentangle the multifaceted influences governing SCI recovery.

Degree of syllable transition changes shift temporally over the course of functional recovery

Behavior is not solely defined by individual syllables but also by the way these syllables are organized and expressed in sequences[28–33]. Rooted in the principles of ethology, motor behavior is inherently sequential, with individual syllables offering insight into an animal's behavioral repertoire and their dynamic sequencing revealing how successive motor actions are coordinated to produce fluid behavior[27,34,35]. Following SCI, transitions in

behavior have been extensively studied in the context of gait, with prior research demonstrating how injury-induced alterations disrupt gait patterns and transitions between gaits, as well as how these changes evolve throughout recovery[3,52,72–75]. Inspired by these findings, we sought to look beyond gait-specific transitions to encompass transitions across all behavioral categories and characterize the probabilistic structure of behavior as it changes throughout recovery. This broader perspective thus provides insights into the "future" actions of the animal[32], capturing the behavioral patterns it is likely to perform next. MoSeq's frame-by-frame behavior outputs offer a detailed analysis on how behavioral sequences unfold over an entire recording session. We examined averaged absolute changes in transition probabilities, where "transition probabilities" reflect the likelihood of the animal switching from one syllable to the next[33] (**Fig. 3a**). Similar to syllable usages (**Fig. 2 and Extended Data Fig. 2**), the entropy of transition probabilities – representing the diversity and complexity of transitions between behaviors – expectantly depended on the level of motor function, with animals exhibiting greater motor function displaying higher transition entropy (**Fig. 3b and Extended Data Fig. 3a**).

To investigate the factors driving transition probability changes, we leveraged our previously identified 17 syllable categories spanning normal, injured, and paralyzed behaviors (**Supplemental Tables 1-2**) to identify overarching patterns in transition probability dynamics. Initial analyses revealed a zero-sum nature of transitions, where increases in some transitions were offset by decreases in others. To better interpret these dynamics, we focused on the overall degree of transition changes rather than individual transitions, enabling us to quantify and compare the magnitude of change across timepoints and injury groups (**Fig. 3a**). Interestingly, our analyses showed that transition probability changes were mostly similar across groups within the same timepoint, suggesting that different severities of injury may exhibit a similar degree of changes in behavioral sequences over time. A notable exception was observed immediately post-operative, where the largest overall changes in motor behavior occurred across injured groups. This acute phase reflects the expected disruption caused by the injury itself[10,22,41], including changes in exhibited syllables used during this period. Shuffled data confirmed that this structure and similarity were not random but instead reflected real underlying dynamics (**Fig. 3c and Extended Data Fig. 3b**). Over time, the magnitude of these transition changes diminished, indicating the stabilization and habituation of behavior as time progressed[76–79] (**Fig. 3d and Extended Data Fig. 3c**).

UMAP analysis indicates more distinctive clustering based on timepoint, while structures based on mouse identity or injury types are less prominent (**Fig. 3e-g and Extended Data Fig. 3d-f**). This indicates that changes in transition probabilities are primarily driven by time-dependent adaptations rather than injury severity or individual variability. Our generalized linear models showed that changes in transition probabilities were more strongly influenced by timepoints post-SCI than by other factors (**Fig. 3h and Extended Data Fig. 3g**). A comparison to usage changes confirms results (**Supplemental Fig. 2c-h**). Coefficient of variation (CV) analyses further demonstrated that variation across timepoints, as well as within individual mice over time, was greater than variation across other groupings (**Fig. 3i and Extended Data Fig. 3g**), reinforcing the dominant role of temporal dynamics in shaping behavioral transitions. Further supporting the conclusion that temporal factors drive the observed results, no significant differences in variation of transition probabilities were detected when comparing injury and no-injury groups (**Fig. 3j and Extended Data Fig. 3i**). This finding suggests that the observed changes are not primarily influenced by the presence or absence of injury. Instead, these changes are more likely tied to time-dependent factors.

The study of behavior sequencing in SCI recovery has been largely overlooked, primarily due to the lack of tools capable of capturing this complexity. MoSeq enables us to examine transitions across all behaviors[33], which revealed the time-dependent changes in transition probabilities. Interestingly, the time-dependent changes appear to be conserved across groups with or without SCI. This finding demonstrates that there is a limitation to the amount of structural behavioral changes that can occur within a given temporal window, indicating a possible ceiling effect.

MoSeq provides enhanced granularity in describing behavior, promoting a higher resolution understanding of recovery than the gold standard BMS

Traditional scoring systems, such as BMS[10], assess motor recovery following SCI and provide a standardized framework for comparing outcomes across studies, labs, and SCI models. The motivation behind comparing MoSeq to BMS lies in evaluating the concurrent validity[80] of MoSeq syllables to ensure that this computational framework aligns with, and perhaps surpasses, the established standards of BMS in describing behavior and recovery. BMS relies on visual assessments by two trained researchers who observe each hindlimb's movement

as a mouse freely explores an open field environment (**Fig. 4a**). Key criteria such as Plantar Stepping, Coordination, and Trunk Stability are used to assess locomotor deficits post-SCI. BMS[10] is a 9-point scale ("BMS avg score") with an 11-point subscale ("BMS subscore") that provides a coarse evaluation of motor function, with higher scores reflecting better mobility. In our results, we observed expected recovery progression, with plateaus where recovery did not dramatically increase for the remainder of the study (**Fig. 4b and Extended Data Fig. 4a**). While BMS excels at capturing gross motor functions, it is inherently constrained to neglect the finer nuances of recovery[20].

By definition, animals can perform different actions to achieve the same BMS avg score/subscore combinations[10,81]. This limitation warrants the necessity for higher resolution behavioral descriptions, such as those that could be provided by MoSeq. To evaluate the concurrent validity[80] of MoSeq syllables,, i.e. measure the extent to which syllable usages and BMS agree, we performed Pearson's correlation analysis[82] between syllable usages and BMS criteria (**Fig. 4c and Extended Data Fig. 4b**). Expectantly, paralyzed syllables correlate strongly negatively with BMS criteria, while normal syllables correlate strongly positively. Moreover, injured syllables demonstrated a positive correlation with partial recovery markers like plantar stepping, but weaker, negligible, or even negative correlations with higher-level recovery markers such as coordinated stepping and trunk stability. These results suggest syllables follow established and well-documented patterns of behaviors as dictated by the gold standard BMS[10].

We next validated whether syllable usages could accurately encode BMS criteria. We observed that syllables could accurately predict BMS criteria, even those that were not directly visible by the downward-facing camera in the MoSeq framework (**Fig. 4de and Extended Data Fig. 4cd**). For example, paw position during locomotion can serve as a visible proxy[9,10,20] and computationally, these variations could reflect subtle differences in body dynamics. This then enables MoSeq to identify unique syllables. Moreover, MoSeq also identifies a multitude of additional syllables that extend beyond the scope of BMS which can provide a more holistic view of recovery.

The high dimensional syllable usages can be summarized into a 2D visualization using PCA. When PCA embedding plots are color-coded with BMS avg scores and subscores, the behavioral variability becomes evident through the spread of points across the plots (**Fig. 4f and Extended Data Fig. 4e**). Upon comparing the behaviors associated with specific BMS avg score/subscore to the syllables representing the same scores, we observed greater behavioral variability in the syllables (**Fig. 4gh and Extended Data Fig. 4fg**). This quantified variation confirmed that behavior varies differently at each BMS avg score/subscore (**Fig. 4ij and Extended Data Fig. 4hi**), which emphasizes not only the differences in how animals behave within a given BMS avg score/subscore combination, but also the inconsistencies between adjacent scores. These results thus underscore the broader behavioral repertoire captured by MoSeq reveal the inherent variability in recovery processes that BMS, constrained by its predefined categories, cannot fully encompass.

High resolution behavioral descriptions reveal bidirectional relationships between behavior and injury histological metrics

Motor recovery is intrinsically shaped by the neurobiological attributes of the injury, including the extent and spatial distribution of damaged tissue[83]. The relationships between behavior and lesion histology are well-recognized, with decades of SCI research illustrating how the neural characteristics of an injury manifest as observable behaviors[9,10,84–87]. This connection is exemplified by BMS and its strong correlation between avg scores and subscores and white matter sparing[10]. However, categorizing the extent of neurological damage to the spinal cord extends beyond white matter sparing alone. To gain a more complete understanding of the disrupted circuits involved in locomotion, sensory processing, and motor recovery[88–94] and their relationships to behavior, we quantified tissue damage by assessing total lesion volume and the spatial dimensions of the injury along the dorsoventral and rostrocaudal axes at the lesion epicenter (**Fig. 4k**).

Pearson's correlations revealed dynamic and bidirectional relationship between syllable usages and lesion metrics over time post-SCI. Specifically, some syllables showed positive correlations with lesion metrics, while others displayed negative correlations, with the strength of these relationships varying over time (**Fig. 4l left, Extended Data Fig. 4j left, and Supplemental Fig. 3a-d bottom**). This demonstrates that the complexity of

recovery involves both the increase and decrease of various unique behaviors. In contrast, BMS criteria exhibited only negative correlations, with minimal variation in the strength of these correlations over time (**Fig. 4l right, Extended Data Fig. 4j right, and Supplemental Fig. 3a-d top**). This homogeneity likely stems from BMS's reliance on human-defined "recovery" criteria, which may inadvertently constrain BMS to unidirectional changes where recovery is limited to the restoration of motor functions[20]. Moreover, these criteria are further constrained by their reliance on human visual observation, which forces assessments to be compressed into singular, locomotion-focused categories[10,20]. Ideally, a biomarker for recovery should capture both the increase of functionality and the decrease in a motor deficit.

Regression models demonstrated that MoSeq syllables achieved significantly better model fits across all timepoints for total injury volume and lesion metrics along the rostrocaudal axis, with comparable performance to BMS criteria along the dorsoventral axis (**Fig. 4m and Extended Data Fig. 4k**). The enhanced performance of syllable-based models may be attributed to their ability to capture diverse and complex behavioral patterns, which may more accurately reflect the intricate neural circuits affected by the longitudinal spread of the injury and its impact on sensorimotor function. In contrast, the comparable performance in lesion metrics along the dorsoventral axis aligns with the established strength of BMS in evaluating gross motor deficits, particularly those involving major motor pathways organized along this anatomical plane[10,20]. These findings could suggest that the higher granularity afforded by MoSeq may provide deeper insights into the broader neurobiological consequences of SCI, particularly in areas where traditional scoring methods may lack sensitivity. Additionally, the observed trends in correlations over the course of recovery suggest that the recovery process may involve multiple behaviors evolving together.

Aggregated numeric scoring of syllable usages strongly correlate with classical metrics associated with functional recovery and lesion severity

Previous SCI studies have identified behaviors ranging from paralysis to normality, with varying intermediate adaptive behaviors along this continuum[22,41,95,96]. Building on this foundation, we aimed to determine whether certain syllable usages covary as such patterns could suggest coordinated recovery processes or shared neurobiological mechanisms[87,97]. Our analysis of the relationships between tissue damage and syllable usage (**Fig. 4 and Extended Data Fig. 4**) revealed that groups of unique behaviors exhibit similar correlation trends to each other, further supporting the existence of shared behavioral patterns. As such, we aimed to quantitatively examine these trends and identify clusters of behaviors that evolve together during recovery. To better represent these patterns, we employed a PCA to reduce the dimensionality of syllable usage data. When we computed the mean trajectory of each group by timepoint, we observed distinctive temporal evolutions of behavior usages among the groups (**Fig. 5a and Extended Data Fig. 5a**). To explore potential quantification of these behavioral trajectories, we applied the Palantir approach[98], a computational method that uses diffusion components[99] to map developmental or evolutionary trajectories (**Fig. 5b and Extended Data Fig. 5b**). Originally developed to characterize cell fate and differentiation by analyzing variations in gene expression during cellular development, Palantir provides a conceptual parallel to the variation in syllable usage throughout recovery. Its ability to capture dynamic modulated shifts in syllable usages aligns well with the progressive nature of recovery following SCI[41,95,96]. Due to the high-dimensionality of syllable usages afforded by the MoSeq framework, Palantir serves as one example of how behavioral data can be quantitatively aggregated. Using Palantir, we developed the "Recovery Score," a quantitative metric that provides an aggregated description of behavior to represent an animal's overall motor functions. When we colored the PC embedding with Recovery Scores, the color gradient aligned as expected: lower scores corresponded to animals with motor deficits, while higher scores were associated with animals with more normal motor functions (**Fig. 5c and Extended Data Fig. 5c**). Similar to BMS, lower Recovery Scores correspond to more severe motor deficits, while higher scores indicate reduced deficits and/or normal motor function. Recovery Score curves (**Fig. 5d and Extended Data Fig. 5d**) show strong concordance with classical BMS avg score and subscore curves (**Fig. 4b and Extended Data Fig. 4a**), indicating that Recovery Scores are consistent with BMS in principle and share a similar behavioral semantic representation. Notably, however, Recovery Scores are derived from syllable usages, rather than predefined human-chosen criteria, and therefore provide a more enriched and unbiased assessment of motor function and recovery.

Regression models further validated the Recovery Score by demonstrating its relevance in the context of traditional open-field outcome measurements and a syllable-based behavior (**Fig. 5e and Extended Data Fig. 5e**). Consistent with open-field data (**Fig. 2a and Extended Data Fig. 2a**), Recovery Scores exhibited positive

correlations with the total distance traveled and maximum velocity. Additionally, behavior-level analysis confirmed that the Recovery Scores negatively correlate with grooming syllable usages (**Fig. 2c and Extended Data Fig. 2c**). A comparison of regression models between the first PC of the behavioral embedding and both BMS scores and Recovery Scores showed that Recovery Scores exhibited smaller and more evenly distributed residuals, indicating a better linear model fit across the continuum from paralysis to normal function (**Fig. 5f and Extended Data Fig. 5f**). Correlating BMS with Recovery Scores showed large variation in residuals, which suggests that BMS might not account for the broader range of behaviors that are represented across different Recovery Scores (**Fig. 5g, Extended Data Fig. 5g, and Supplemental Fig. 4ab**). Furthermore, Recovery Scores outperformed BMS avg scores and subscores in correlating with total injury volume and rostrocaudal axis distribution, but did not surpass BMS in correlating with dorsoventral axis injury metrics (**Fig. 5h and Extended Data Fig. 5h**). This suggests that while BMS maintains parity in capturing gross motor deficits linked to disruptions in major motor circuits along the dorsoventral axis[10,20], Recovery Scores may capture the more nuanced aspects of the lesion characteristics along the rostrocaudal axis and the impact on functional recovery. The results are similar to the correlations between lesion metrics and MoSeq syllable usages (**Fig. 4m and Extended Data Fig. 4k**), suggesting that aggregation and compression done to compute Recovery Scores still captures the necessary granularities in describing behavior and their relations to lesion metrics.

Syllable usage trends derived from Recovery Scores reveal clusters of syllable usages along the spectrum from paralysis to normality

Given the observed correlation trends among specific syllables during recovery, as indicated by their correlations with lesion volumes, we sought to determine whether distinct clusters of syllables emerge throughout the recovery process based on trends derived from Recovery Scores (**Fig. 5i, Extended Data Fig. 5i, and Supplemental Fig. 4cd**, see methods). To quantitatively characterize these clusters, we applied Leiden clustering[100] to syllable usage trends. This analysis revealed organized groups of syllables that exhibited similar usage patterns (**Fig. 5j, Extended Data Fig. 5j, and Supplemental Fig. 4ef**). When we color the PC embedding with the mean syllable usages from each cluster (**Fig. 5k and Extended Data Fig. 5k**), the colored points aligned well with trajectories associated with timepoint (**Fig. 5a and Extended Data Fig. 5a**) and the Recovery Score distribution (**Fig. 5c and Extended Data Fig. 5c**). Examining group-wise syllable usages, velocities, and heights across clusters validated our findings (**Supplemental Fig. 4gh**). We semantically named the identified syllable clusters, which represented behavior from most impaired to the most functional: Full Paralysis, Severe Paraparesis, Intermediate Paraparesis, "Easy" Normal, and "Challenging" Normal (**Supplemental Tables 45-46**). The first three clusters encapsulate traditionally-conceived injured behaviors[10,20]: Full Paralysis is marked by complete hindlimb dragging; Severe Paraparesis involved major deficits with some hindlimb movement; and Intermediate Paraparesis was characterized by hindlimb weight-bearing deficits and noticeable instability. Interestingly, we observed two “normal” syllable clusters: the Easy Normal cluster included behaviors that might be less physically demanding[101–103], such as wall-assisted rears and slower locomotion. These behaviors were achievable by animals with Mild SCI in later timepoints in the course recovery. The Challenging Normal cluster encompassed potentially more physically demanding behaviors that require significant hindlimb function[104,105], such as unassisted high rears and faster locomotion or darting. The Challenging Normal cluster was only achievable by animals without SCI. The emergence of these distinct normal clusters could suggest that recovery is influenced by the level of neural and motor demands required for specific behaviors.

Syllable usages provide a framework for quantifying behavioral recovery following SCI. The detailed behavior descriptions enable us to compute an aggregated Recovery Score and identify distinct behavioral clusters associated with different levels of motor functions. The emergence of clusters in which certain syllable usages decrease to allow others to increase indicates that recovery is not a straightforward linear restoration of function, but rather a dynamic reorganization of behaviors that are influenced by the interplay between injury severity and task demands.

DISCUSSION

SCI profoundly disrupts motor behavior, necessitating advanced frameworks to analyze recovery processes with precision and granularity[9,10,20,21]. This study demonstrates the potential of MoSeq[28–33] as a rigorous framework

for analyzing motor recovery following SCI, offering unprecedented insights through high-resolution behavior descriptions that surpass traditional assessments. By leveraging an unsupervised machine learning approach, MoSeq segments freely moving mouse behavior pre- and post-SCI into distinct and repeated behavioral syllables in normal, injured, and paralyzed conditions. Unlike traditional recovery assessments, which often rely on predefined metrics and/or human observation[9,10,20], MoSeq automatically delivers an unbiased, frame-by-frame granular analysis. The MoSeq framework not only matches BMS in evaluating motor recovery, but also reveals novel insights into the structure, sequencing, and adaptive strategies of behavior during recovery following SCI.

SCI induces system-wide changes that extend far beyond locomotor deficits, yet traditional recovery assessments have predominantly focused on locomotor outcomes as primary indicators of functional recovery[1,7,9–11,20,41,45,51]. This focus has been largely driven by the methodological limitations of manual recovery assessments, which inherently results in other potentially important behaviors[59,60] to go unnoticed. Reliance on manual scoring and human observation may result in non-locomotor-related behaviors getting overlooked or not accurately quantified by human observers. In contrast, MoSeq automatically identifies distinct syllables as they naturally emerge during behavior, capturing both locomotor and non-locomotor behaviors together. This granularity is necessary because the additional behaviors could be critical to understanding the responses to injury and recovery from it, as they may reflect underlying physiological states or compensatory mechanisms. For instance, grooming syllables, which have been associated with hypersensitivity and discomfort following SCI[57], change over the course of recovery. As such, this posits that behaviors traditionally considered unrelated to functional recovery may coevolve with locomotor behaviors. This coevolution may suggest that non-locomotor behaviors are not incidental, but integral to recovery. This challenges the conventional focus of SCI research and the prioritization of locomotor behaviors as the primary measure of functional recovery. This may also suggest that these seemingly unrelated behaviors are part of coordinated recovery processes or reflect shared neurobiological mechanisms.

By shifting the focus on the overall behavioral repertoire, MoSeq also revealed subtle differences in individual mice. While it is well-established that injury severity drives system-wide behavioral changes[50], this study presents individual differences following SCI as a factor in recovery. Indeed, intrinsic factors, including an animal's baseline preinjury behaviors and individual adaptive capacity, may play a role in shaping recovery[30,63,64]. This duality signifies there may be a non-negligible effect of individuality on behavior, such as on adaptive behaviors during recovery. Thus, individual mouse differences should be taken into consideration as warranted. MoSeq's focus on the full behavioral repertoire also represents a critical departure from traditional approaches that rely on gross predefined markers of recovery[9,10,20]. BMS[10], for example, is susceptible to overly emphasizing isolated motor events, such as the tail touching the ground during locomotion, which may result in disproportionate changes to an animal's score and potential misrepresentation of its overall recovery status. In contrast, MoSeq framework captures and aggregates all behaviors within the repertoire. This enables the observations of particular syllable usages that increase as recovery progresses and others that decrease as deficits resolve. As such, recovery milestones should reflect this bidirectional process. This bidirectionality could usher new multifaceted therapeutic approaches that reinforce adaptive behaviors to promote recovery while simultaneously targeting maladaptive behaviors to suppress deficits.

In addition to individual syllables, the organization and sequencing of these behaviors can offer us insights into motor recovery. In a healthy system, top-down mechanisms originating in the brain work synergistically with spinal-level processes to coordinate transitions between motor actions, whose regulation is reflected by changes in said transitions[88,106–108]. However, SCI can result in potential maladaptive reorganization within the circuits responsible for sequencing motor actions, disrupting this intricate interplay[36,37,39,109,110]. When analyzing the structure of behavior through syllable sequences, we observed an unexpected consistency in the degree of change within sequences across groups at a given timepoint, regardless of injury severity or even the presence of an injury. This temporal consistency could suggest the existence of a ceiling effect[102] that limits the extent of changes in behavioral sequences over time. This may indicate that the rate and scale of behavioral adaptation are constrained by the inherent limitations of the neuromuscular system, including its biomechanical properties and the fundamental architecture of neural circuits[102,103,111–113], rather than being exclusively driven by injury-induced neural plasticity. In the context of an injured system, this ceiling effect highlights a finite capacity for behavioral reorganization within a given temporal window[102], which might suggest that recovery processes operate within boundaries defined by the system's intrinsic properties. These findings raise intriguing questions about whether conserved mechanisms underlie this effect in both the injured and noninjured conditions. If such

mechanisms are indeed conserved, they could shed light on the principles governing motor sequence organization and the limits of adaptation within a given timeframe.

To compare behavior across mice following SCI, we compressed the high-dimensional behavior patterns into a unified numeric metric across mice. In this study, we generated the “Recovery Score” metric to address this need. Functionally similar to BMS[10], the Recovery Score offers a more detailed perspective on recovery because it is derived from automatically detected syllable usage data rather than predefined criteria. Unlike BMS, which simplifies behavior into course categories, the Recovery Score retains the necessary variation of syllable usages across a broad spectrum of behaviors. The Recovery Score encodes syllable usage trends, such that groups of syllables cluster across different levels of motor functions. The identification of distinct behavioral clusters from paralysis to normality agrees with findings from prior studies that show the selective restitution of motor functions during recovery. For instance, clinical studies suggest that improvements in bladder and sexual function often accompany gains in leg movement[114,115]. By extending this perspective to preclinical models, MoSeq provides a framework for integrating and understanding these multifaceted recovery processes. Examining these clusters revealed that achieving "normal" behaviors is indeed possible for animals with SCI, though these behaviors may be less physically demanding[101–105]. This observation may suggest that the nervous system has an upper limit to recovery during functional restoration. Future research could explore how the emergence of specific behaviors within clusters aligns with neural activity patterns and how these patterns differ between clusters. Furthermore, this clustering approach raises the possibility of using groups of hallmark syllables as behavioral biomarkers to assess recovery levels and inform therapeutic strategies.

The preliminary exploration of Recovery Score underscores the need for a more precise scoring system, as manually scored traditional recovery assessments alone cannot capture the complexity of SCI recovery. This study demonstrates that MoSeq offers a promising foundation for developing such a system by providing high-dimensional behavioral insights, yet there are several technical and experimental limitations warrant addressing. In terms of technical limitations, MoSeq requires specialized high-performance equipment and the current version does not accommodate varying mouse sizes[33]. The downward-facing camera setup may also limit the ability to capture certain aspects traditionally assessed in kinematics analyses, such as precise ankle positions. To address these technical limitations, ongoing efforts have focused on upgrading the framework to enhance its versatility and ensure long-term adaptability. For example, traditional kinematics readouts are achievable by adding additional customary cameras facing desired angles in the MoSeq recording setup. Recent advancements, such as Keypoint-MoSeq[116], also present opportunities to integrate both methods, combining the strengths of detailed kinematic metrics with behavioral sequencing capabilities. In terms of experimental limitations, this study was purposefully constrained to bilateral contusion SCI models whose behavior spanned paralysis to normality[10] to serve as a proof-of-concept for MoSeq’s utility in categorizing behavior post-SCI. The Recovery Score represents just one computational approach to aggregate high-dimensional behavioral data from MoSeq into a quantifiable recovery metric. Future studies should adopt more complex experimental designs, incorporating diverse mouse strains, injury models, treatment groups, etc. to further refine this approach and advance the pursuit of a new gold standard in recovery assessment.

MATERIALS AND METHODS

Subjects

Male and female 8- to 9-week-old C57BL/6 (Strain Code 027) mice (n = 71 male, n = 68 female) were procured from Charles River Laboratories (Wilmington, MA, USA). The mice were provided ad libitum access to food and water and were maintained in a 12-hour light and 12-hour dark cycle within the animal facility at the Division of Life Sciences, Nelson Biology Laboratories, Rutgers University. Spinal cord injury (SCI) surgeries were conducted on these mice at an age of 9-10 weeks. All housing, surgical procedures, behavioral experiments, and euthanasia were carried out in strict adherence to Rutgers University's Institutional Animal Care and Use Committee (IACUC) guidelines, with the experiments following the ARRIVE guidelines.

Surgical Procedures

The mouse thoracic SCI contusion model was conducted following the protocol described previously[10,117]. Briefly, 9- to 10-week-old mice were anesthetized with isoflurane (NDC 66794-017-25, Piramal Critical Care, Inc., Bethlehem, PA, USA), administered at 5% initially and reduced to 2% during surgery. A thermo-regulated heating

pad was utilized to monitor and sustain the body temperature throughout the surgery. Prior to surgical incision, the upper-back skin was shaved and sanitized using alternating preparations of Betadine scrub (NDC 67618-151-32, Purdue Products L.P., Stamford, CT, USA) and 70% ethanol. Preemptive and post-operative analgesia were provided by injecting 0.1 ml of 0.125% Bupivacaine (NDC 0409-1163-18, Hospira, Lake Forest, IL, USA) subcutaneously around the incision site. The slow-release Buprenorphine formulation, Ethiqa XR (Fidelis Pharmaceuticals, LLC, North Brunswick, NJ, USA), was subcutaneously administered at 3.25 mg/kg. To prevent corneal drying, lubricant ophthalmic ointment (Artificial Tears, NDC 59399-162-35, Akorn Animal Health, Lake Forest, IL, USA) was applied to the eyes post-anesthesia induction. A 3 cm skin incision was made along the median line on the animal's back, followed by laminectomy at the vertebrae level halfway between T9 and T10. The contusion SCI was induced using the IH Infinite Horizon Impactor (Precision Systems and Instrumentation, LLC; model IH-0400). The mouse was secured on the experimental table using spinal cord stabilizing forceps, and the exposed spinal cord was precisely positioned. The impactor tip was placed over the center of the exposed spinal cord. A force of 30 kDyn was used for mild injuries, and a force of 50 kDyn was employed for moderate injuries, both with a dwell time of 0 ms. The muscles were then sutured in two separate anatomical layers (Suture 4-0 Polyglactin 910 FS-2, Cat. No. 6549052, Henry Schein, Melville, NY, USA) and the incision closed with wound clips (Reflex-9, 9 mm, Cat. No. 201-1000, Cell Point Scientific, Inc.,Gaithersburg, MD, USA). Sham-injured animals underwent laminectomy but did not receive a contusion. Animals classified as "naive" received drugs during the approximate duration of the surgery but were not incised. Animals designated for complete transections underwent transections following laminectomy using aspiration and sectioning with microscissors as previously described[10,22] using great care to ensure complete severance of all axon fibers.

Postoperative monitoring and care

Following surgery, mice were individually housed in cages with white Alpha Dri bedding (W.F. Fisher & Son, Branchburg Township, NJ, USA) to monitor for potential bladder infections and bleeding. Subsequently, mice received subcutaneous injections of cefazolin (25 mg/kg) and 0.5 ml saline once daily for three days after injury. Additionally, cages were positioned partially on 37°C thermo-regulated heating pads for the duration of the study. Monitoring of food intake was performed using Labdiet Rodent Diet 5001 pellets (Cat. No. LD5001, W.F. Fisher & Son) placed within the cages, with DietGel (Clear H2O Dietgel 76A, Cat. No. CH2ODG762OZ, Fisher & Son) provided if necessary. Daily assessments for dehydration were conducted, and HydroGel (Clear H2O Hydrogel, Cat. No. CH2OHG2OZ) was administered as needed along with supplemental saline injections. Bladder care was administered twice daily from the day after injury until euthanization to prevent urine retention. Cases of hematuria were treated with Baytril (2.7 mg/kg) over a 10-day period. Mice were withdrawn from the study if they experienced weight loss exceeding 80% of their initial body weight or exhibited severe or persistent autophagia.

Behavior: Basso Mouse Scale (BMS)

At 2 days and 1-10 weeks post-spinal cord injury, locomotor recovery in mice was assessed using the BMS[10]. Mice were placed in the center of a small wading pool (open field) with a 44-inch diameter for 4 minutes, and hind limb movements were evaluated by two trained scorers on a scale ranging from 0 to 9. A score of 0 indicated no hind limb movement, while higher scores denoted improved hind-limb joint coordination, with 9 representing "normal" walking. In cases of discrepancy between scorers, the lower score was selected. If necessary, the sub-score table, considering parameters such as plantar stepping, coordination, paw position, trunk instability, and tail position, was utilized.

Behavior: MoSeq Framework

Here and below, we employed methodology as previously described[28–30,30–33] and customized it as required.

Data acquisition

On the day of acquisition, mice were transferred to the behavior room for a 30-minute habituation. Raw depth data was acquired using a Microsoft Kinect v2 camera, pointed downwards towards a circular arena, connected to a computer running the program kinect2-nidaq (v0.2.4alpha). Individual mice were gently placed in the arena, and their naturalistic behavior over a 20-minute period was recorded. Captured depth video frames were each 512 x 424 px, collected at 30 Hz, with each pixel encoded as a 16-bit unsigned integer indicating the distance of that pixel from the Kinect camera sensor in millimeters. Data was saved in binary format and transferred to another computer system for later processing.

Data preprocessing

Because injured mice often face paralysis, resulting in overall lower-to-the-ground body positions, we found it difficult to reliably segment mice from the noise and background using the classical computer vision approaches deployed in the conventional moseq2-extract pipeline. Therefore we developed a custom deep-learning-based extraction program, which outputs results that are compatible with the existing MoSeq ecosystem of tools. We adapted the Detectron2 framework released by Facebook AI Research[43], based on the state of the art mask-R-CNN architecture, as it could jointly propose segmentation masks and keypoint estimations. We use a ResNet 50 FPN backbone which feeds a Region Proposal Network (RPN), and both then feed a series of modular ROI Heads, namely Instance Bounding Box, Instance Segmentation Mask, and Instance Keypoint Regression. We generated a training set of 4,529 background-subtracted depth images, split into training (4,076 images) and test (453 images) subsets, which we annotated with masks and eight keypoints: nose, left/right ear, neck, left/right hip, tail base, and tail tip. During model training, we applied randomized augmentations including random rotation, random scaling, random brightness, random contrast, random gaussian white noise, and random particle noise (to simulate dust particles in the acquisition environment). We trained the model for 100,000 iterations, at which point performance plateaued with the following COCO performance metrics: bbox (AP: 64.812, AP50: 97.202, AP75: 73.423), keypoints (AP: 45.851, AP50: 93.768, AP75: 37.795), and segm (AP: 55.680, AP50: 98.673, AP75: 56.266).

For utilizing this model in extraction of real-world videos, we constructed a parallelized processing pipeline that would 1) read raw depth video frames from saved files, remove background and non-ROI pixels, 2) submit images for deep learning inference producing estimates of boxes, masks and keypoints, 3) post process these features and generate cropped and rotated mouse images and masks (in the same style as classical extraction), 4) generate a preview video visualization of the extraction results and save results in HDF5 format in a way that mimics the outputs produced by classical extraction. This whole process is able to extract a twenty minute depth video in about twelve minutes on a computer system equipped with a AMD Ryzen 7 3800X (8-Cores @ 3.89 GHz), 64 GB RAM, Nvidia RTX 2080 SUPER, and network attached storage (1Gbps link).

Behavioral Modeling and Analysis

The preprocessed data underwent dimensionality reduction and modeling through the MoSeq pipeline[33], resulting in frame-by-frame descriptions of the animal's dynamics (i.e., velocity, position) and syllable labels. Syllable usage fractions were computed across all sessions for analysis. Semantic labels were manually assigned to syllables by examining their associated dynamics and behaviors in the recorded movies. Noise syllables were excluded from further analysis, resulting in 92 syllables for male mice ("male model") and 93 syllables for female mice ("female model"). Behavioral diversity within a session was quantified using usage entropy, defined as $Entropy = -\sum_{x \epsilon X} p(x) \, log \, p(x)$, where x represents the usage fraction of each syllable within the session, and X is the usages of all syllables. Behavioral similarity between sessions was assessed using the Jensen–Shannon distance, calculated with the scipy.spatial.distance.jensenshannon function.

Transitions between syllables are characterized by outgoing transition probabilities, which represent the likelihood of one syllable transitioning into another. The diversity of these syllable transitions were quantified using outgoing entropy (same formula as above), where *x* represents the probabilities of one syllable transitioning into each other syllable in the session. To analyze changes in outgoing probabilities (Fig 3), syllables were aggregated into 17 syllable categories, so that outgoing probabilities are aggregated to represent one group of syllables transitioning to another group of syllables (i.e., from dart to pause, from locomotion to rear, etc.). The mean absolute differences in outgoing probabilities between two consecutive timepoints within each mouse were computed to capture transitions. As a control, outgoing syllable categories were shuffled within each mouse to generate a "shuffle control".

Classical locomotion metrics, such as total distance traveled and maximum velocity, were derived from MoSeq outputs. Distance traveled was calculated as the sum of the Euclidean distances between mouse centroid coordinates across two consecutive frames. Maximum velocity was determined as the 99 percentile of session velocities as a means of excluding extreme values from noise.

Behavioral Distance Analysis

MoSeq behavioral distance is a metric used to assess the similarity between two syllables. Behavioral distance was calculated as previously described[29] using the function "get_behavioral_distance()" from the moseq2-viz library using default parameters. Briefly, the function computes the correlation between the 3D pose trajectories associated with each syllable, which are encapsulated by the autoregressive process that models the two syllables. The behavioral distance is then derived by subtracting the correlation value from one. To analyze patterns in the behavioral distance matrix, hierarchical clustering was applied using a Euclidean distance metric and average linkage method. This clustering allowed for the grouping of syllables based on their behavioral similarity, which then facilitated downstream analysis of behavioral structure and patterns.

Predicting BMS Scores from Syllable Usages

Correlations between BMS features and syllable usages were computed using `scipy.stats.pearsonr()` for each pair of syllables and BMS metrics. Correlation R-values were plotted using seaborn.

Regression models to predict BMS metrics given syllable usages were constructed using scikit-learn. Values were first standardized using `StandardScalar()` and then fed to an `ElasticNet` model with the parameter `l1_ratio` set to 0.1. A 5-fold cross-validation strategy that balanced injury types and timepoints across CV splits was employed for model training and evaluation. Shuffle models were also generated, in which the syllable usage values were randomly shuffled prior to being fed to the model. Model performance, reported as the coefficient of determination (R^2) and measured on the held-out portion of data, across the distribution of CV splits was plotted as violins using seaborn. Model coefficients, averaged across the CV splits, were plotted as a heatmap using seaborn.

Syllable PC Trajectory

Syllable trajectories in the PC space were computed and visualized from the PC scores generated during the dimensionality reduction step in the MoSeq analysis pipeline. To smooth the trajectories, the PC scores were smoothed using a Gaussian kernel with a window length 5. The background of the visualization featured a binned 2D histogram to represent the density of points in PC space. To illustrate syllable dynamics, the syllable segments with the most frequent occurrence of same-length trajectories were selected to increase the count of representative examples. Trajectories originating from similar starting points in the PC space and exhibiting the closest shape in terms of trajectory dynamics were chosen to show the syllable dynamics in the PC space.

Syllable sequence Uniform Manifold Approximation and Projection (UMAP)

Syllables were labeled with semantic labels such as "locomotion," "rear," etc., and subsequently converted into one-hot encoding where a value of 1 indicated the animal is performing the syllable and 0 indicates it is not. Each row of the resulting dataframe corresponds to a single frame of behavior. To incorporate contextual information about temporally adjacent syllables, a Gaussian kernel was applied to the one-hot encoded dataframe, thereby allowing each row to reflect syllables performed in close temporal proximity. Each point on the UMAP represents a single frame with the UMAPs thus capturing structural patterns in syllable sequences over time.

UMAPs of syllable usages and degree of syllable category transition changes

UMAP was employed to reduce the dimensionality of the data for visualizing both syllable usages across groups and timepoints, as well as the degree of syllable category transition changes. For syllable usages, the syllable usage data from each session was z-scored before applying UMAP. The UMAP parameters for this analysis were set as: n_components=2, n_neighbors=20, metric='euclidean', and min_dist=0.1. For syllable category transition changes, the mean absolute changes across all syllable categories between two consecutive timepoints for each mouse were z-scored prior to UMAP analysis. The UMAP parameters for this analysis were set as: n_components=2, n_neighbors=20, metric='cosine', min_dist=0.3.

Linear classifiers

Linear classifiers were employed to predict group labels based on syllable usages across all sessions within each timepoint (**Fig. 2g and Ext. Data Fig. 2g**). Model accuracy was calculated from the results of 10

independent restarts of model runs. To ensure dataset balance, the input data was randomly resampled to equalize the number of sessions per group and standardized prior to modeling. A linear support vector classifier (SVC) was utilized, with a Stratified K-Fold (n_split=5) cross validation approach applied to maintain proportional representation of each group across folds. Accuracy was calculated as the ratio of true predictions to total predictions (True Labels/All Labels) and compared against results from shuffled data within each timepoint.

Within Cluster Sum of Squares (WCSS)

The WCSS was used to quantify the variability within each group, where lower values represent more similarity within the group, and higher values represent more dispersion within the group. To calculate WCSS for each type of sorting, the centroid of each group was computed by averaging the syllable usages across all sessions within the group for each syllable category. Then the sums of Euclidean distances between all sessions and their respective group centroids were computed. The sum of these distances within each group provided group-specific WCSS, which was then subsequently summed across all groups to produce the final WCSS metric. The calculated WCSS values were compared against those derived from data with shuffled group labels to evaluate their significance of the observed clustering.

General linear model (GLM)

GLMs were applied to analyze the relationship between factors such as group, mouse identity, and timepoint with syllable usage and transition probability changes. For syllable usage, GLMs were constructed using one-hot encoded factors for each session as predictors to model log-transformed, z-scored syllable usage. The models were: syllable usages ~ group + timepoint; usages ~ mouse + timepoint. For absolute outgoing transition probability changes, GLMs were similarly designed to predict log-transformed, z-scored average absolute outgoing transition probability changes based on one-hot encoded factors for each mouse at each timepoint. The models were: mean absolute change ~ group + timepoint; mean absolute change ~ Mouse + Timepoint. A linear regression model (ElasticNet with $\alpha=0.01$) was fit to model log-transformed, z-scored, session based, syllable usage for each.

Coefficient of Variation (CV)

The CV was calculated as the ratio between the standard deviation to the mean and is used to assess the variation of data within specific groupings. In Fig 2J, session CV for each syllable was computed across all sessions; for Inter-mouse, CV for each syllable is computed across mean syllable usages of each mouse; for intra-mouse, CV for each syllable is computed across mean syllable usages of each mouse in each timepoint; for injury, CV for each syllable is computed across mean syllable usages of each injury in each timepoint; for timepoint, CV for each syllable is computed across mean syllable usages of each timepoint in each mouse. The "No SCI vs. SCI" analysis followed a similar methodology, but the data was divided into two groups: no SCI (Naive, Sham) and SCI (Mild, Mod, Tx). In Fig 3I, the CV for each syllable category was computed across mean absolute outgoing probability changes instead of mean syllable usages and Fig 3J was split by no SCI and SCI. In Fig 4IJ, CV for each syllable was computed across mean syllable usages in each BMS average score and BMS average score+subscore combination.

Lesion metrics regressions

Linear regressions (ordinary least square) were utilized to predict lesion metrics (MBP volume, MBP along the dorsoventral axis, and MBP along the rostrocaudal axis) at each timepoint. For Fig 4m, the models were: lesion metric ~ BMS left + BMS right; lesion metric ~ BMS subscore items; lesion metric ~ MoSeq syllable categories. Similarly, for Fig 5h, the models were: lesion metric ~BMS avg score; lesion metric ~ bms subscore; lesion metric ~ recovery score. The R^2 for each model was compared across timepoints using t-tests.

Recovery scores and syllable usage clusters across different recovery scores

Recovery Scores and syllable usage clusters were derived using a cell differentiation algorithm, Palantir (REF), adapted to capture the progression of recovery from paralysis to normal function. Recovery Scores were calculated from pseudotime (as described in the Palantir algorithm) to represent what stage a session is at in the spectrum from full paralysis to normality. Briefly, the algorithm first applied PCA to reduce the dimensionality of syllable usages across all sessions. The dimensionality-reduced data are used to compute the diffusion maps[99] to capture developmental trends with parameters set to n_components=5 and knn=30. Sessions were

ordered along the diffusion components by calculating the shortest path between user-defined preinjury normal sessions (fully functional) and 2 DPO sessions (least functional) to assign pseudotime for each session that represent the relative distance to the predefined preinjury normal session. The user-defined sessions that were selected were sessions with the closest Euclidean distance to the respective group syllable means (i.e., the preinjury session with the closest distance to preinjury means, the 2 DPO session in Naive with the closest distance to 2 DPO Naive means, etc.). The Recovery Score was computed as $Recovery\ Score = (1 - Palantir\ Pseudotime) \times 10$ to scale values to a range between 0 and 10, analogous to theBMS scale. Lower Recovery Scores correspond to reduced motor function, while higher scores reflect more normal motor functions. For visualization, Recovery Scores were rounded to the nearest 0.5 (**Fig. 5c and Extended Data Fig. 5c**). Syllable usage trends along Recovery Scores were analyzed using Mellon Function Estimator fit to 500 evenly spaced numbers between 0 and 10, similar to the palantir.presults.compute_gene_trends function[118]. Nearest neighbor (n_neighbors=35) and Leidesn algorithm are used to cluster syllables with similar trends across the recovery scores. To visualize the cluster's mean usage hotspots on the PCA plots (**Fig. 5k and Extended Data Fig. 5k**), mean syllable usages are scaled with a MinMaxScalar so that the highest usage is 1 and lowest usage is 0.

Immunohistochemistry

Following the final behavioral tests at 10 weeks post-SCI, mice were deeply anesthetized with 5% isoflurane (NDC 66794-017-25, Piramal Critical Care, Inc., Bethlehem, PA, USA), followed by transcardial perfusion with saline containing 10 U/mL heparin (Cat No. H3393-100KU, Sigma-Aldrich, St. Louis, MO, USA), and 4% paraformaldehyde (Cat No. 15714-S, Electron Microscopy Sciences, Hatfield, PA, USA). The spinal cord columns were dissected, and the ventral side of the spinal cord was exposed. The tissue was postfixed in 4% paraformaldehyde overnight at 4°C, subsequently washed three times with phosphate-buffered saline (PBS), and stored in PBS containing 0.01% sodium azide for further processing.

For immunohistochemistry of the lesion site, 1 cm long spinal cords containing the lesion site at the center were cryoprotected by overnight immersion in 4°C PBS with increasing sucrose concentrations (10%, 15%, 20%), beginning with the lowest sucrose concentration overnight in each solution. Finally, spinal cords were incubated overnight with 50% Optimal Freezing Temperature (OCT) (Tissue-Tek, Cat No. 25608-930, VWR, Radnor, PA, USA) medium and 50% 20%-sucrose dissolved in PBS. Following these saturation steps, the tissue was frozen in OCT medium, and 20 µm thick serial sagittal cryosections were prepared, mounted on Diamond® White Glass Charged Microscope Slides (Globe Scientific Inc, Cat. No. 1358W), and stored at -20°C for immunostaining. The slides were thawed and dried for 1 hour at 20°C, washed three times with PBS, and then blocked with 3% bovine serum albumin (BSA) (Cat No. RLBSA50, IPO Supply Center, Piscataway, NJ, USA), 10% Normal Goat Serum (NGS) (Cat No. 01-6201, Fisher Scientific, Waltham, MA, USA), and 0.3% Triton-X 100 (X100-100ml, Sigma-Aldrich, St. Louis, MO, USA) for 1 hour at 20°C. Subsequently, the tissue sections were co-stained overnight at 4°C with a primary chicken antibody against the astrocyte marker Glial Fibrillary Acidic Protein (GFAP) (1:1000 in blocking solution; Cat No. 200-901-D60, VWR, Radnor, PA, USA) and a primary rat antibody against myelin basic protein (MBP) (1:50 in blocking solution; Cat. No. MAB386, Sigma-Aldrich, St. Louis, MO, USA). Following five 5-minute washes at 20°C with PBS, the tissue was incubated for 2 hours at 20°C with corresponding Alexa Fluor 488 (1:500, anti-Chicken coupled to Alexa Fluor 488, Cat No. A11039, Fisher Scientific, Waltham, MA, USA) and Alexa Fluor 565 (1:500, anti-rat coupled to Alexa Fluor 546, Cat No. A11081, Fisher Scientific, Waltham, MA, USA) secondary antibodies diluted in blocking solution. The tissue was washed five times for 5 minutes at 20°C with PBS, incubated for 10 minutes with DAPI solution diluted 1:5000 in PBS, and the sections were finally mounted under glass coverslips with Fluoromount-G (Cat No. 100241-874, VWR, Radnor, PA, USA).

Image acquisition and analysis

Quantitative analysis of the sagittal sections was performed using the automated microscope IN Cell Analyzer 6000 (GE Healthcare, Chicago, IL, USA). Four microscope slides were loaded into the system, and the image capturing protocol from the manufacturer was implemented. The tissue sections were precisely located via preview, and high-resolution images were acquired for subsequent analysis.

FIGURES

Fig 1. Leveraging the machine learning method Motion Sequencing (MoSeq) to uncover behavioral patterns in post-SCI recovery.

- **Panel a:** Individual mice (n = 71 male) were recorded in an open-field arena to assess their mobility in 12 sessions pre- and post-operation. Sessions included: preinjury, 2 days post operation (DPO), 7 DPO, and weekly thereafter for 10 weeks. Groups included: (no-SCI) Naive and Sham; and (SCI) Mild, Mod, and Tx. Two outcome measures were assessed: motion sequencing (MoSeq) and the Basso Mouse Scale (BMS). For MoSeq, animals were recorded using a depth camera in 20 min/session bouts. For BMS, two trained human scorers recorded an animal's mobility in 4min/session bouts.
- **Panel b:** Recorded videos were processed using custom segmentation algorithms and subsequently modeled with the established MoSeq framework[33]. The MoSeq pipeline applied dimensionality reduction via PCA (PCA), retaining the top 10 PCs as inputs. These components were then analyzed using MoSeq's autoregressive hidden Markov model (AR-HMM) to identify behaviors into distinct, modulated, and repeated sub-second behavioral motifs, referred to as "syllables" (see examples across different mobilities).
- **Panel c:** Syllables with similar dynamics are characterized by closer behavioral distance (see methods) when grouped by hierarchical clustering.
- Panel d: The top 10 PCs, used by the AR-HMM model to capture pose dynamics and segment syllables, illustrate distinct pixel weight distributions. For example, PC1 primarily reflects variations in head height.
- **Panel e:** Trajectories of example locomotion syllables in the PC embedding exhibit consistent patterns for unique syllables, with trajectories progressing from start (white) to end (blue) of emission (average trajectory in red). Trajectories of syllables across different mobilities are distinct in shape within the embedding.
- **Panel f:** Spinograms, a visualization plot that shows animal height and relative distance traveled throughout a syllable, illustrate how example locomotion syllables vary across different mobilities.
- **Panel g:** The fraction of usage for each syllable was calculated for each session. Usages of example locomotion syllables across different mobilities reveals how behavior patterns develop post operation. Data (n = 71) were analyzed by two-way analysis of variance (ANOVA) (Supplemental Table 3) and Dunnett's multiple comparison test (Supplemental Table 4).
- **Panel h:** Uniform Manifold Approximation and Projection (UMAP) plots of weighted syllable sequences, colored by locomotor syllables across different mobilities, demonstrate that structural patterns vary distinctly between syllables associated with different mobility states.

Extended Data Fig 1. Leveraging the machine learning method MoSeq to uncover behavioral patterns in post-SCI recovery (female model).

- **Panel a:** Syllables with similar dynamics are characterized by closer behavioral distance (see methods) when grouped by hierarchical clustering. Same as Fig 1c but for females (n = 68).
- **Panel b:** The top 10 PCs, used by the AR-HMM model. Same as Fig 1d but for females.
- **Panel c:** Trajectories of example locomotion syllables in the PC embedding exhibit consistent patterns for unique syllables, with trajectories progressing from start (white) to end (blue) of emission (average trajectory in red). Same as Fig 1e but for females.
- **Panel d:** Spinograms. Same as Fig 1f but for females.
- **Panel e:** The fraction of usage for each syllable was calculated for each session. Same as Fig 1g but for females. Data (n = 68) were analyzed by ANOVA (Supplemental Table 5) and Dunnett's multiple comparison test (Supplemental Table 6).
- **Panel f:** UMAP plots of weighted syllable sequences. Same as Fig 1h but for females.

Supplemental Fig 1. Custom segmentation and syllable analysis for SCI recovery in mice.

- **Panel a:** Illustrative example of the original MoSeq pipeline and our custom segmentation algorithms. The original MoSeq data extraction tools could not support SCI mice as lower trunk and limbs of injured mice were classified as noise and filtered out due to their significantly lower positioning compared to uninjured mice. We adapted the data processing step by leveraging the Detectron2 implementation of multi-task RCNN

- **Panel b:** Schematic of the extraction pipeline, including the addition of our custom segmentation algorithms.
- **Panel c:** UMAP plots of weighted syllable sequences, colored by all syllable categories (except for locomotion syllables) across different mobilities for the male model, demonstrate that structural patterns vary distinctly between syllables associated with different mobility states.
- **Panel d:** Same as c, but for the female model.

Fig 2. Distinct behavioral patterns after SCI are primarily influenced by injury severity and mouse identity, rather than timepoint.

- **Panel a:** Total distance traveled (left) and maximum velocity (right) by group over time Data (n = 71) were analyzed by ANOVA (Supplemental Table 7-8) and Dunnett's multiple comparison test (Supplemental Table 9-10) here and below unless indicated otherwise.
- **Panel b:** Behavioral summary heatmap with syllables ordered by type and mobility demonstrates distinct behavior patterns across groups over time. Example of grooming syllables (orange box) highlights these group and timepoint differences
- **Panel c:** Groom syllable usages (aggregating all groom syllables) by group over time showing that the usages of a non-locomotor behavior changes significantly different across different groups during recovery (statistical tests in Supplemental Tables 11-12).
- **Panel d:** UMAP plots of syllable usages, colored by groups (leftmost plot) and timepoints (right four plots), reveal that injury severities form distinct clusters, while timepoints lack clear structural separation with data points from different timepoints appearing intermixed.
- **Panel e:** Syllable usage entropy by group over time. Higher entropy indicates more diverse behavior patterns, while lower entropy indicates more predictable behavior pattern (statistical tests in Supplemental Tables 13-14).
- **Panel f:** Jensen Shannon distance to preinjury timepoint by group over time, denoting how similar timepoints post operation are to the preinjury timepoint (statistical tests in Supplemental Tables 15-16).
- **Panel g:** Linear classifier accuracy in distinguishing groups using syllable usages over time (top). Example confusion matrices at 3 timepoints (bottom) where the shading denotes the ratio of predicted labels and real labels. The darker shading along the diagonal indicates better performance, while the light gray shading indicates confusion (i.e., 56 DPO Naive and Sham).
- **Panel h:** Usage heatmaps of syllable categories (17) sorted randomly, by group, by mouse and by timepoint (top) and Within cluster sum of squares (bottom) to describe purity within structure where a smaller number denotes more structure (see methods).
- **Panel i:** Weights from generalized linear models with syllable usage as a function of two factor types: injury types and timepoints (left) and mouse ID and timepoints (right). A darker color denotes higher absolute value and therefore more sensitivity to the factor type.
- **Panel j:** The coefficient of variation (CV) of syllable usages was analyzed across various groupings: session (within all sessions), inter-mouse (across different mice), intra-mouse (within a single mouse), group (across different groups), and timepoints (across different timepoints). A higher CV within a specific grouping indicates greater variation in syllable usages across that grouping (unpaired t-test; Session vs Timepoint: $P = 1.33 \times 10^{-34}$, Intermouse vs Timepoint: $P = 6.13 \times 10^{-29}$, Intramouse vs Timepoint: $P = 8.11 \times 10^{-27}$, Injury vs Timepoint: $P = 1.04 \times 10^{-17}$).
- **Panel k:** Same as j, but split by injury (no SCI and SCI) revealing that SCI contributes more to the variation (unpaired t-test; Session, No SCI vs SCI: $P = 2.08 \times 10^{-11}$, Intermouse, No SCI vs SCI: $P = 2.69 \times 10^{-19}$, Intramouse, No SCI vs SCI: $P = 8.86 \times 10^{-9}$).

Extended Data Fig 2. Distinct behavioral patterns after SCI are primarily influenced by injury severity and mouse identity, rather than timepoint (female model).

- **Panel a:** Total distance traveled (left) and maximum velocity (right) by group over time. Same as Fig 1a but for females. Data (n = 68) were analyzed by ANOVA (Supplemental Table 17-18) and Dunnett's multiple comparison test (Supplemental Table 19-20) here and below unless indicated otherwise.
- **Panel b:** Behavioral summary heatmap with syllables ordered by type and mobility. Same as Fig 1b but for females.

- **Panel c:** Groom syllable usages (aggregating all groom syllables) by group over time. Same as Fig 1c but for females (statistical tests in Supplemental Tables 21-22).
- **Panel d:** UMAP plots of syllable usages, colored by groups (leftmost plot) and timepoints (right four plots). Same as Fig 1d but for females.
- **Panel e:** Syllable usage entropy by group over time. Same as Fig 1e but for females (statistical tests in Supplemental Tables 23-24).
- **Panel f:** Jensen Shannon distance to preinjury timepoint by group over time. Same as Fig 1f but for females (statistical tests in Supplemental Tables 25-26).
- **Panel g:** Linear classifier accuracy in distinguishing groups using syllable usages over time (top). Same as Fig 1g but for females.
- **Panel h:** Usage heatmaps of syllable categories (17) sorted randomly, by group, by mouse and by timepoint (top) and Within cluster sum of squares (bottom). Same as Fig 1h but for females.
- **Panel i:** Weights from generalized linear models with syllable usage as a function of two factor types: injury types and timepoints (left) and mouse ID and timepoints (right). Same as Fig 1i but for females.
- **Panel j:** The CV of syllable usages was analyzed across various groupings: session (within all sessions), inter-mouse (across different mice), intra-mouse (within a single mouse), group (across different groups), and timepoints (across different timepoints). Same as Fig 1j but for females (unpaired t-test; Session vs Timepoint: $P = 1.45 \times 10^{-27}$, Intermouse vs Timepoint: $P = 2.21 \times 10^{-15}$, Intramouse vs Timepoint: $P = 1.28 \times 10^{-15}$, Injury vs Timepoint: $P = 2.24 \times 10^{-8}$).
- **Panel k:** Same as j, but split by injury (no SCI and SCI) revealing that SCI contributes more to the variation. Same as Fig 1k but for females (unpaired t-test; Session, No SCI vs SCI: $P = 8.05 \times 10^{-13}$, Intermouse, No SCI vs SCI: $P = 4.21 \times 10^{-18}$, Intramouse, No SCI vs SCI: $P = 2.63 \times 10^{-13}$).

Supplemental Fig 2. Temporal dynamics and variability in syllable usage differences across groups and timepoints.

- **Panel a:** Heatmap plots for each group of the Jensen Shannon distance to preinjury timepoint over time for the male model, denoting how similar timepoints post operation are to the preinjury timepoint.
- **Panel b:** Same as a, but for the female model.
- **Panel c:** Average absolute syllable usage differences of 17 syllable categories heatmap over time using real (left) and shuffle (right) data.
- **Panel d:** Same as c, but for the female model.
- **Panel e:** UMAP of average absolute syllable usage differences of 17 syllable categories for the male model, colored by mouse ID (left), group (middle), and timepoint (right) (see methods).
- **Panel f:** Same as e, but for the female model.
- **Panel g:** Weights from generalized linear models for the male model with average absolute syllable usage differences as a function of two factor types: injury types and timepoints (left) and mouse ID and timepoints (right). A darker color denotes higher absolute value and therefore more sensitivity to the factor type.
- **Panel h:** Same as g, but for the female model.

Fig 3. The extent of syllable transition changes is influenced more by recovery timepoints than by group or mouse identity.

- **Panel a:** Illustrative example of syllable transitions evolving over time, with visualizations provided for 2 DPO (left), 7 DPO (middle), and the difference in transitions between 7 DPO and 2 DPO (right).
- **Panel b:** Average syllable outgoing entropy over time for different groups over time. Higher entropy reflects a more diverse selection of the next syllable, while lower entropy signifies more predictable transitions. Data (n = 71) were analyzed by ANOVA (Supplemental Table 27) and Dunnett's multiple comparison test (Supplemental Table 28) here and below unless otherwise indicated.
- **Panel c:** Average absolute outgoing transition differences heatmap over time using real (left) and shuffle (right) data.
- **Panel d:** Quantification of average absolute outgoing transition differences in c (statistical tests in Supplemental Tables 29-30).
- **Panel e:** UMAP of average absolute outgoing transition differences of 17 syllable categories, colored by mouse ID (see methods).

- **Panel f:** Same as e, but colored by group.
- **Panel g:** Same as e, but colored by timepoint.
- **Panel h:** Weights from generalized linear models with absolute outgoing transition changes as a function of injury types and timepoints (left) and mouse IDand timepoints (right).
- **Panel i:** The coefficient of variance (CV) of syllable usages was analyzed across various groupings: session (within all sessions), inter-mouse (across different mice), intra-mouse (within a single mouse), group (across different groups), and timepoints (across different timepoints) (unpaired t-test; Session vs Timepoint: $P = 0.03$, Intermouse vs Timepoint: $P = 8.31 \times 10^{-8}$, Intramouse vs Timepoint: $P = 0.06$, Injury vs Timepoint: $P = 9.89 \times 10^{-6}$).
- **Panel j:** Same as i, but split by injury (no SCI and SCI) did not reveal significant differences between the two groups (unpaired t-test; Session, No SCI vs SCI: $P = 0.7$, Intermouse, No SCI vs SCI: $P = 0.14$, Intramouse, No SCI vs SCI: $P = 0.51$, Timepoint, No SCI vs SCI: $P = 0.12$).

Extended Data Fig 3. The extent of syllable transition changes is influenced more by recovery timepoints than by group or mouse identity (female model).

- **Panel a:** Average syllable outgoing entropy over time for different groups over time. Same as Fig 1b but for females. Data (n = 68) were analyzed by ANOVA (Supplemental Table 31) and Dunnett's multiple comparison test (Supplemental Table 32) here and below unless otherwise indicated.
- **Panel b:** Average absolute outgoing transition differences heatmap over time using real (left) and shuffle (right) data. Same as Fig 1c but for females.
- **Panel c:** Quantification of average absolute outgoing transition differences in c. Same as Fig 1d but for females (statistical tests in Supplemental Tables 33-34).
- **Panel d:** UMAP of average absolute outgoing transition differences of 17 syllable categories, colored by mouse ID (see methods). Same as Fig 1e but for females.
- **Panel e:** Same as d, but colored by group. Same as Fig 1f but for females.
- **Panel f:** Same as d, but colored by timepoint. Same as Fig 1g but for females.
- **Panel g:** Weights from generalized linear models with absolute outgoing transition changes as a function of injury types and timepoints (left) and mouse IDand timepoints (right). Same as Fig 1h but for females.
- **Panel h:** The coefficient of variance (CV) of syllable usages was analyzed across various groupings: session (within all sessions), inter-mouse (across different mice), intra-mouse (within a single mouse), group (across different groups), and timepoints (across different timepoints). Same as Fig 1i but for females (unpaired t-test; Session vs Timepoint: $P = 0.03$, Intermouse vs Timepoint: $P = 2.43 \times 10^{-9}$, Intramouse vs Timepoint: $P = 0.06$, Injury vs Timepoint: $P = 3.68 \times 10^{-5}$).
- **Panel i:** Same as i, but split by injury (no SCI and SCI) did not reveal significant differences between the two groups. Same as Fig 1j but for females (unpaired t-test; Session, No SCI vs SCI: $P = 0.66$, Intermouse, No SCI vs SCI: $P = 0.01$, Intramouse, No SCI vs SCI: $P = 0.47$, Timepoint, No SCI vs SCI: $P = 0.03$).

Fig 4. MoSeq syllables provide greater resolution than BMS and outperform BMS in lesion estimation.

- **Panel a:** Illustrative schematic of manual BMS scoring conducted by two trained evaluators.
- **Panel b:** BMS Average Scores (left) and Subscores (right) by group over time (left). Data (n = 71) were analyzed by ANOVA (Supplemental Table 35-36) and Dunnett's multiple comparison test (Supplemental Table 37-38).
- **Panel c:** Pearson's correlation between syllable usages and BMS reveals a positive correlation between higher BMS scores and syllables associated with greater mobility and a negative correlation with syllables associated with reduced mobility.
- **Panel d:** Cross validations of regularized linear regression models of syllable usages predicting BMS of real and shuffled data.
- **Panel e:** Syllable weights in d, with darker colors indicating syllables that are important for predicting the respective BMS scores.
- **Panel f:** PCA plots computed from syllable usages colored by BMS Average Scores (left), BMS Subscores (middle), and groups with BMS Average Score of 5 and Subscore of 3 (right).
- **Panel g:** Heatmap of BMS criteria for sessions with a BMS Average Score of 5 and Subscore of 3 demonstrates that identical score-subscore combinations can arise from varying types of behavior.

- **Panel h:** Heatmap of syllable usages for sessions with a BMS Average Score of 5 and Subscore of 3 demonstrates the high resolution afforded by MoSeq syllables in describing behavior.
- **Panel i:** CV plots for each syllable usage within a given BMS Average Score and across all scores, which highlights the variation differences in behavior patterns within one score.
- **Panel j:** Same as i, but for all BMS Average Score and Subscore combinations.
- **Panel k:** Example histological images of sagittal sections of the spinal cord at the lesion site for each group. The abbreviations used: MBP, myelin basic protein (green); GFAP, glial fibrillary acidic protein (red).
- **Panel l:** Pearson's correlations for injury-induced demyelination volumes and syllable usages (left) and BMS criteria (right).
- **Panel m:** Adjusted R^2 values from correlation models of lesion metrics with BMS Scores, BMS Subscores, and MoSeq syllable category usages for all timepoints. Correlation models at each timepoint are: lesion~bms_left+bms_right, lesion~BMS_criteria, lesion~MoSeq_syllable_category_usages (see methods). Results indicate that MoSeq syllable categories yield better model fits for MBP volume and injury extent in the dorsoventral axis, which suggests that the high resolution of MoSeq better captures the extent of injury damage at the lesion site (unpaired t-test; MBP: syllables vs BMS avg score: $P = 1.67 \times 10^{-6}$, syllables vs BMS subscore: $P = 1.77 \times 10^{-5}$, MBP dorsoventral axis: BMS avg score vs BMS subscore: $P = 0.04$, MBP rostrocaudal axis: syllables vs BMS avg score: $P = 3.92 \times 10^{-5}$, syllables vs BMS subscore: $P = 0.01$, BMS avg score vs BMS subscore: $P = 0.004$).

Extended Data Fig 4. MoSeq syllables provide greater resolution than BMS and outperform BMS in lesion estimation (female model).

- **Panel a:** BMS Average Scores (left) and Subscores (right) by group over time (left). Data (n = 68) were analyzed by ANOVA (Supplemental Table 39-40) and Dunnett's multiple comparison test (Supplemental Table 41-42).
- **Panel b:** Pearson's correlation between syllable usages and BMS. Same as Fig 1c but for females.
- **Panel c:** Cross validations of regularized linear regression models of syllable usages predicting BMS of real and shuffled data. Same as Fig 1d but for females.
- **Panel d:** Syllable weights in d, with darker colors indicating syllables that are important for predicting the respective BMS scores. Same as Fig 1e but for females.
- **Panel e:** PCA plots computed from syllable usages colored by BMS Average Scores (left), BMS Subscores (middle), and groups with BMS Average Score of 5 and Subscore of 3 (right). Same as Fig 1f but for females.
- **Panel f:** Heatmap of BMS criteria for sessions with a BMS Average Score of 5 and Subscore of 3. Same as Fig 1g but for females.
- **Panel g:** Heatmap of syllable usages for sessions with a BMS Average Score of 5 and Subscore of 3. Same as Fig 1h but for females.
- **Panel h:** CV plots for each syllable usage within a given BMS Average Score and across all scores. Same as Fig 1i but for females.
- **Panel i:** Same as h, but for all BMS Average Score and Subscore combinations. Same as Fig 1j but for females.
- **Panel j:** Pearson's correlations for injury-induced demyelination volumes and syllable usages (left) and BMS criteria (right). Same as Fig 1i but for females.
- **Panel k:** Adjusted R^2 values from correlation models of lesion metrics with BMS Scores, BMS Subscores, and MoSeq syllable category usages for all timepoints. Same as Fig 1m but for females (unpaired t-test; MBP: syllables vs BMS avg score: $P = 0.00074$, syllables vs BMS subscore: $P = 0.00072$, MBP rostrocaudal axis: syllables vs BMS avg score: $P = 0.005$, BMS avg score vs BMS subscore: $P = 0.03$).

Supplemental Fig 3. Correlations with injury-induced demyelination across injury axes.

- **Panel a:** Pearson's correlations for injury-induced demyelination along the dorsoventral axis and BMS criteria (top) and syllable usages (bottom) for the male model.
- **Panel b:** Same as a, but for the female model.
- **Panel c:** Pearson's correlations for injury-induced demyelination along the rostrocaudal axis and BMS criteria (top) and syllable usages (bottom) for the male model.
- **Panel d:** Same as c, but for the female model.

Fig 5. The Recovery Score, computed from MoSeq syllables, demonstrates recovery trajectories across injury types and identifies syllable clusters during recovery.
Recovery trajectories based on syllables reveal temporally expressed clusters of behavior.

- **Panel a:** PCA plot colored by group, shaded by timepoint, and labeled with the progression of each group's averaged behavior over time.
- **Panel b:** Diffusion components heatmap computed from syllable usages. Diffusion components represent a low-dimensional embedding of syllable usage variations,with each component capturing different phenotypical features within the behavioral data.
- **Panel c:** PCA plot colored by Recovery Scores, a metric computed from syllable usage data to describe an animal's overall motor functions. Similar to BMS, animals with lower Recovery Scores exhibit motor deficits and those with higher scores exhibit more normal motor functions.
- **Panel d:** Recovery Scores by group over time. Data (n = 71) were analyzed by ANOVA (Supplemental Table 43) and Dunnett's multiple comparison test (Supplemental Table 44).
- **Panel e:** Correlation plots of Recovery Scores with total distance traveled (Ordinary least squares (OLS) linear regression here and below, $R^2 = 0.77$), maximum velocity ($R^2 = 0.85$), and groom usages ($R^2 = 0.80$) reveal consistent slopes but different intercepts between SCI and non-SCI groups.
- **Panel f:** Correlation plots of behavioral PC1 computed from MoSeq syllables with BMS Average Score (left, $R^2 = 0.83$) and Recovery Score (middle, $R^2 = 0.90$). Residual values plot (right) for both correlations demonstrate that Recovery Score has a better model fit.
- **Panel g:** Correlation plot of BMS Average Score and Recovery Score (left) and model residuals (right, $R^2 = 0.87$). While BMS and Recovery Score exhibit a correlation, the residuals underscore variations in how BMS aligns with the underlying behavioral data.
- **Panel h:** Adjusted R^2 values from correlation models of lesion metrics with BMS Average Scores, BMS Subscores, and Recovery Scores for all timepoints. Correlation models at each timepoint are: lesion~BMS_Avgerag_Score, lesion~BMS_criteria, lesion~Recovery_Score. Results indicate that Recovery Scores yield better model fits for MBP volume and injury extent in the dorsoventral axis (unpaired t-test; MBP: Recovery Score vs BMS avg score: $P = 0.001$, Recovery Score vs BMS subscore: $P = 0.008$, MBP rostrocaudal axis: Recovery Score vs BMS avg score: $P = 0.002$, Recovery Score vs BMS subscore: $P = 0.01$).
- **Panel i:** Heatmap of syllable usage trends calculated using the Recovery Score model (see methods). The Leiden algorithm clusters the trends into groups of syllables that evolve similarly with the Recovery Score metric.
- **Panel j:** Line plots of the normalized usage of syllables within each cluster across Recovery scores. Light lines denote individual syllables, solid lines are the average usage within the cluster, and shading represents the standard deviation.
- **Panel k:** PCA plots, colored by normalized mean syllable usages within each cluster, where darker shades represent higher usage and lighter shades indicate lower usage. These hotspots highlight the clusters of syllables within the PCA embedding that are most frequently utilized at specific scores. Clusters were semantically labeled based on the syllable identities associated with each cluster.

Extended Data Fig 5. The Recovery Score, computed from MoSeq syllables, demonstrates recovery trajectories across injury types and identifies syllable clusters during recovery (female model).

- **Panel a:** PCA plot colored by group, shaded by timepoint, and labeled with the progression of each group's averaged behavior through the PCA embedding over time. Same as Fig 1a but for females.
- **Panel b:** Diffusion components heatmap computed from syllable usages. Same as Fig 1b but for females.
- **Panel c:** PCA embedding plot colored by Recovery Scores. Same as Fig 1c but for females.
- **Panel d:** Recovery Scores by group over time. Same as Fig 1d but for females. Data (n = 68) were analyzed by ANOVA (Supplemental Table 47) and Dunnett's multiple comparison test (Supplemental Table 48).
- **Panel e:** Correlation plots of Recovery Scores with total distance traveled (OLS linear regression here and below, $R^2 = 0.71$), maximum velocity ($R^2 = 0.77$), and groom usages ($R^2 = 0.72$). Same as Fig 1e but for females.

- **Panel f:** Correlation plots of behavioral PC1 computed from MoSeq syllables with BMS Average Score (left, $R^2 = 0.81$) and Recovery Score (middle, $R^2 = 0.83$). Same as Fig 1f but for females.
- **Panel g:** Correlation plot of BMS Average Score and Recovery Score (left, $R^2 = 0.91$) and model residuals (right). Same as Fig 1g but for females.
- **Panel h:** Adjusted R^2 values from correlation models of lesion metrics with BMS Average Scores, BMS Subscores, and Recovery Scores for all timepoints. Same as Fig 1h but for females (unpaired t-test; MBP dorsoventral axis: Recovery Score vs BMS subscore: $P = 0.01$, MBP rostrocaudal axis: Recovery Score vs BMS avg score: $P = 0.02$).
- **Panel i:** Heatmap of syllable usage trends calculated using the Recovery Score model (see methods). Same as Fig 1i but for females.
- **Panel j:** Line plots of the normalized usage of syllables within each cluster across Recovery scores. Light lines denote individual syllables, solid lines are the average usage within the cluster, and shading represents the standard deviation. Same as Fig 1j but for females.
- **Panel k:** PCA plots, colored by normalized mean syllable usages within each cluster. Same as Fig 1k but for females.

Supplemental Fig 4. Trends of syllable usages across Recovery Scores and syllable clusters.

- **Panel a:** Residual values histogram for the correlation of BMS Average Score and Recovery Score for the male model.
- **Panel b:** Same as a, but for the female model.
- **Panel c:** Heatmap of raw syllable usage trends calculated using the Recovery Score model (see methods) for the male model.
- **Panel d:** Same as c, but for the female model.
- Panel e: Heatmap plots of the normalized usage of syllables within each cluster for the male model.
- **Panel f:** Same as e, but for the female model.
- **Panel g:** Line plots of usages (top), velocity (middle), and height (bottom) by group over time for each cluster for the male model. Data (n = 71) were analyzed by ANOVA (Supplemental Table 49) and Dunnett's multiple comparison test (Supplemental Table 50) here and below.
- **Panel h:** Same as g, but for the female model (n = 68, statistical tests in Supplemental Tables 51-52).

REFERENCES

1. Courtine G, Song B, Roy RR, et al. Recovery of supraspinal control of stepping via indirect propriospinal relay connections after spinal cord injury. *Nat Med*. 2008;14(1):69-74. doi:10.1038/nm1682
2. Frigon A, Akay T, Prilutsky BI. Control of mammalian locomotion by somatosensory feedback. *Compr Physiol*. 2021;12(1):2877-2947. doi:10.1002/cphy.c210020
3. Barrière G, Leblond H, Provencher J, Rossignol S. Prominent Role of the Spinal Central Pattern Generator in the Recovery of Locomotion after Partial Spinal Cord Injuries. *J Neurosci*. 2008;28(15):3976-3987. doi:10.1523/JNEUROSCI.5692-07.2008
4. Lavrov I, Courtine G, Dy CJ, et al. Facilitation of Stepping with Epidural Stimulation in Spinal Rats: Role of Sensory Input. *J Neurosci*. 2008;28(31):7774-7780. doi:10.1523/JNEUROSCI.1069-08.2008
5. Huie JR, Morioka K, Haefeli J, Ferguson AR. What Is Being Trained? How Divergent Forms of Plasticity Compete To Shape Locomotor Recovery after Spinal Cord Injury. *J Neurotrauma*. 2017;34(10):1831. doi:10.1089/neu.2016.4562
6. Eisdorfer JT, Smit RD, Keefe KM, Lemay MA, Smith GM, Spence AJ. Epidural Electrical Stimulation: A Review of Plasticity Mechanisms That Are Hypothesized to Underlie Enhanced Recovery From Spinal Cord Injury With Stimulation. *Front Mol Neurosci*. 2020;13:163. doi:10.3389/fnmol.2020.00163
7. Eisdorfer JT, Sobotka-Briner H, Schramfield S, et al. Chemogenetic modulation of sensory afferents induces locomotor changes and plasticity after spinal cord injury. *Front Mol Neurosci*. 2022;15. Accessed October 13, 2022. https://www.frontiersin.org/articles/10.3389/fnmol.2022.872634
8. Ballermann M, Fouad K. Spontaneous locomotor recovery in spinal cord injured rats is accompanied by anatomical plasticity of reticulospinal fibers. *Eur J Neurosci*. 2006;23(8):1988-1996. doi:10.1111/j.1460-9568.2006.04726.x
9. Basso DM, Beattie MS, Bresnahan JC. A sensitive and reliable locomotor rating scale for open field testing in rats. *J Neurotrauma*. 1995;12(1):1-21. doi:10.1089/neu.1995.12.1
10. Basso DM, Fisher LC, Anderson AJ, Jakeman LB, McTigue DM, Popovich PG. Basso Mouse Scale for locomotion detects differences in recovery after spinal cord injury in five common mouse strains. *J Neurotrauma*. 2006;23(5):635-659. doi:10.1089/neu.2006.23.635
11. Eisdorfer JT, Phelan MA, Keefe KM, et al. Addition of angled rungs to the horizontal ladder walking task for more sensitive probing of sensorimotor changes. *PLoS ONE*. 2021;16(2):e0246298. doi:10.1371/journal.pone.0246298
12. Metz GA, Whishaw IQ. The ladder rung walking task: a scoring system and its practical application. *J Vis Exp JoVE*. 2009;(28):1204. doi:10.3791/1204
13. Antonow-Schlorke I, Ehrhardt J, Knieling M. Modification of the ladder rung walking task-new options for analysis of skilled movements. *Stroke Res Treat*. 2013;2013:418627. doi:10.1155/2013/418627
14. Cummings BJ, Engesser-Cesar C, Cadena G, Anderson AJ. Adaptation of a ladder beam walking task to assess locomotor recovery in mice following spinal cord injury. *Behav Brain Res*. 2007;177(2):232-241. doi:10.1016/j.bbr.2006.11.042
15. Jones BJ, Roberts DJ. The quantiative measurement of motor inco-ordination in naive mice using an acelerating rotarod. *J Pharm Pharmacol*. 1968;20(4):302-304. doi:10.1111/j.2042-7158.1968.tb09743.x
16. Deacon RMJ. Measuring motor coordination in mice. *J Vis Exp JoVE*. 2013;(75):e2609. doi:10.3791/2609
17. Tillerson JL, Miller GW. Grid performance test to measure behavioral impairment in the MPTP-treated-mouse model of parkinsonism. *J Neurosci Methods*. 2003;123(2):189-200. doi:10.1016/s0165-0270(02)00360-6
18. Luong TN, Carlisle HJ, Southwell A, Patterson PH. Assessment of Motor Balance and Coordination in Mice using the Balance Beam. *J Vis Exp JoVE*. 2011;(49):2376. doi:10.3791/2376
19. Fowler SC, Miller BR, Gaither TW, Johnson MA, Rebec GV. Force-plate quantification of progressive behavioral deficits in the R6/2 mouse model of Huntington's disease. *Behav Brain Res*. 2009;202(1):130-137. doi:10.1016/j.bbr.2009.03.022
20. Sheets AL, Lai PL, Fisher LC, Basso DM. Quantitative Evaluation of 3D Mouse Behaviors and Motor Function in the Open-Field after Spinal Cord Injury Using Markerless Motion Tracking. *PLOS ONE*. 2013;8(9):e74536. doi:10.1371/journal.pone.0074536
21. Sato Y, Kondo T, Shinozaki M, et al. Markerless analysis of hindlimb kinematics in spinal cord-injured mice through deep learning. *Neurosci Res*. 2022;176:49-56. doi:10.1016/j.neures.2021.09.001
22. Ma M, Basso DM, Walters P, Stokes BT, Jakeman LB. Behavioral and histological outcomes following graded spinal cord contusion injury in the C57Bl/6 mouse. *Exp Neurol*. 2001;169(2):239-254.

doi:10.1006/exnr.2001.7679
23. Mathis A, Mamidanna P, Cury KM, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci*. 2018;21(9):1281-1289. doi:10.1038/s41593-018-0209-y
24. Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc*. 2019;14(7):2152-2176. doi:10.1038/s41596-019-0176-0
25. Pereira TD, Tabris N, Matsliah A, et al. SLEAP: A deep learning system for multi-animal pose tracking. *Nat Methods*. 2022;19(4):486-495. doi:10.1038/s41592-022-01426-1
26. Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. Published online May 30, 2019. doi:10.48550/arXiv.1812.08008
27. Dawkins R. Hierarchical organisation: A candidate principle for ethology. In: *Growing Points in Ethology*. Cambridge U Press; 1976.
28. Wiltschko AB, Tsukahara T, Zeine A, et al. Revealing the structure of pharmacobehavioral space through motion sequencing. *Nat Neurosci*. 2020;23(11):1433-1443. doi:10.1038/s41593-020-00706-3
29. Markowitz JE, Gillis WF, Beron CC, et al. The Striatum Organizes 3D Behavior via Moment-to-Moment Action Selection. *Cell*. 2018;174(1):44-58.e17. doi:10.1016/j.cell.2018.04.019
30. Levy DR, Hunter N, Lin S, et al. Mouse spontaneous behavior reflects individual variation rather than estrous state. *Curr Biol CB*. 2023;33(7). doi:10.1016/j.cub.2023.02.035
31. Gschwind T, Zeine A, Raikov I, et al. Hidden behavioral fingerprints in epilepsy. *Neuron*. 2023;111(9):1440-1452.e5. doi:10.1016/j.neuron.2023.02.003
32. Markowitz JE, Gillis WF, Jay M, et al. Spontaneous behaviour is structured by reinforcement without explicit reward. *Nature*. 2023;614(7946):108-117. doi:10.1038/s41586-022-05611-2
33. Lin S, Gillis WF, Weinreb C, et al. Characterizing the structure of mouse behavior using Motion Sequencing. *Nat Protoc*. 2024;19(11):3242-3291. doi:10.1038/s41596-024-01015-w
34. Tinbergen N, Tinbergen N. *The Study of Instinct: With a New Preface*. Oxford University Press; 1989.
35. Kaplan HS, Thula OS, Khoss N, Zimmer M. Nested Neuronal Dynamics Orchestrate a Behavioral Hierarchy across Timescales. *Neuron*. 2020;105(3):562. doi:10.1016/j.neuron.2019.10.037
36. Kadoya K, Lu P, Nguyen K, et al. Spinal cord reconstitution with homologous neural grafts enables robust corticospinal regeneration. *Nat Med*. 2016;22(5):479-487. doi:10.1038/nm.4066
37. Takase H, Regenhardt RW. Motor tract reorganization after acute central nervous system injury: a translational perspective. *Neural Regen Res*. 2021;16(6):1144. doi:10.4103/1673-5374.300330
38. Filous AR, Schwab JM. Determinants of Axon Growth, Plasticity, and Regeneration in the Context of Spinal Cord Injury. *Am J Pathol*. 2018;188(1):53-62. doi:10.1016/j.ajpath.2017.09.005
39. Mohammed H, Hollis ER. Cortical Reorganization of Sensorimotor Systems and the Role of Intracortical Circuits After Spinal Cord Injury. *Neurotherapeutics*. 2018;15(3):588-603. doi:10.1007/s13311-018-0638-z
40. Chen B, Li Y, Yu B, et al. Reactivation of Dormant Relay Pathways in Injured Spinal Cord by KCC2 Manipulations. *Cell*. 2018;174(3):521-535.e13. doi:10.1016/j.cell.2018.06.005
41. All AH, Al-Nashash H. Comparative analysis of functional assessment for contusion and transection models of spinal cord injury. *Spinal Cord*. 2021;59(11):1206-1209. doi:10.1038/s41393-021-00698-2
42. Reinhardt DR, Stehlik KE, Satkunendrarajah K, Kroner A. Bilateral cervical contusion spinal cord injury: A mouse model to evaluate sensorimotor function. *Exp Neurol*. 2020;331:113381. doi:10.1016/j.expneurol.2020.113381
43. Wu Y, Kirillov A, Massa F, Lo WY, Girshick R. Detectron2. Published online 2019. https://github.com/facebookresearch/detectron2
44. Dietz V, Fouad K. Restoration of sensorimotor functions after spinal cord injury. *Brain*. 2014;137(3):654-667. doi:10.1093/brain/awt262
45. Barbeau H, Rossignol S. Enhancement of locomotor recovery following spinal cord injury. *Curr Opin Neurol*. 1994;7(6):517.
46. Edgerton VR, Leon RD de, Harkema SJ, et al. Retraining the injured spinal cord. *J Physiol*. 2001;533(Pt 1):15. doi:10.1111/j.1469-7793.2001.0015b.x
47. Cai LL, Courtine G, Fong AJ, Burdick JW, Roy RR, Edgerton VR. Plasticity of functional connectivity in the adult spinal cord. *Philos Trans R Soc Lond B Biol Sci*. 2006;361(1473):1635-1646. doi:10.1098/rstb.2006.1884
48. Côté MP, Murray M, Lemay MA. Rehabilitation Strategies after Spinal Cord Injury: Inquiry into the Mechanisms of Success and Failure. *J Neurotrauma*. 2017;34(10):1841-1857. doi:10.1089/neu.2016.4577
49. Timoszyk WK, Nessler JA, Acosta C, et al. Hindlimb loading determines stepping quantity and quality

following spinal cord transection. *Brain Res*. 2005;1050(1-2):180-189. doi:10.1016/j.brainres.2005.05.041
50. Cha J, Heng C, Reinkensmeyer DJ, Roy RR, Edgerton VR, De Leon RD. Locomotor ability in spinal rats is dependent on the amount of activity imposed on the hindlimbs during treadmill training. *J Neurotrauma*. 2007;24(6):1000-1012. doi:10.1089/neu.2006.0233
51. Lovely RG, Gregor RJ, Roy RR, Edgerton VR. Effects of training on the recovery of full-weight-bearing stepping in the adult spinal cat. *Exp Neurol*. 1986;92(2):421-435. doi:10.1016/0014-4886(86)90094-4
52. de Leon RD, Hodgson JA, Roy RR, Edgerton VR. Locomotor capacity attributable to step training versus spontaneous recovery after spinalization in adult cats. *J Neurophysiol*. 1998;79(3):1329-1340. doi:10.1152/jn.1998.79.3.1329
53. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philos Transact A Math Phys Eng Sci*. 2016;374(2065):20150202. doi:10.1098/rsta.2015.0202
54. Gradwell MA, Ozeri-Engelhard N, Eisdorfer JT, et al. Multimodal sensory control of motor performance by glycinergic interneurons of the mouse spinal cord deep dorsal horn. *Neuron*. 2024;112(8):1302-1327.e13. doi:10.1016/j.neuron.2024.01.027
55. Bohic M, Pattison LA, Jhumka ZA, et al. Mapping the neuroethological signatures of pain, analgesia, and recovery in mice. *Neuron*. 2023;111(18):2811-2830.e8. doi:10.1016/j.neuron.2023.06.008
56. Shirota Y, Otani T, Wasada S, Ito S, Mieda T, Nakamura K. Inner and outer penetrating spinal cord injuries lead to distinct overground walking in mice. *IBRO Neurosci Rep*. 2024;16:345-352. doi:10.1016/j.ibneur.2024.02.005
57. Leem YJ, Joh JW, Joeng KW, Suh JH, Shin JW, Leem JG. Central Pain from Excitotoxic Spinal Cord Injury Induced by Intraspinal NMDA Injection: A Pilot Study. *Korean J Pain*. 2010;23(2):109. doi:10.3344/kjp.2010.23.2.109
58. Kloos AD, Fisher LC, Detloff MR, Hassenzahl DL, Basso DM. Stepwise motor and all-or-none sensory recovery is associated with nonlinear sparing after incremental spinal cord injury in rats. *Exp Neurol*. 2005;191(2):251-265. doi:10.1016/j.expneurol.2004.09.016
59. Ferguson AR, Irvine KA, Gensel JC, et al. Derivation of Multivariate Syndromic Outcome Metrics for Consistent Testing across Multiple Models of Cervical Spinal Cord Injury in Rats. *PLOS ONE*. 2013;8(3):e59712. doi:10.1371/journal.pone.0059712
60. Gensel JC, Tovar CA, Hamers FPT, Deibert RJ, Beattie MS, Bresnahan JC. Behavioral and Histological Characterization of Unilateral Cervical Spinal Cord Contusion Injury in Rats. *J Neurotrauma*. 2006;23(1):36-54. doi:10.1089/neu.2006.23.36
61. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948;27(3):379-423. doi:10.1002/j.1538-7305.1948.tb01338.x
62. Weber RZ, Mulders G, Kaiser J, Tackenberg C, Rust R. Deep learning-based behavioral profiling of rodent stroke recovery. *BMC Biol*. 2022;20(1):232. doi:10.1186/s12915-022-01434-9
63. Voelkl B, Altman NS, Forsman A, et al. Reproducibility of animal research in light of biological variation. *Nat Rev Neurosci*. 2020;21(7):384-393. doi:10.1038/s41583-020-0313-3
64. Goot MH van der, Kooij M, Stolte S, Baars A, Arndt SS, Lith HA van. Incorporating inter-individual variability in experimental design improves the quality of results of animal experiments. *PLoS ONE*. 2021;16(8):e0255521. doi:10.1371/journal.pone.0255521
65. Siegel CS, Fink KL, Strittmatter SM, Cafferty WBJ. Plasticity of intact rubral projections mediates spontaneous recovery of function after corticospinal tract injury. *J Neurosci Off J Soc Neurosci*. 2015;35(4):1443-1457. doi:10.1523/JNEUROSCI.3713-14.2015
66. Liu ZH, Yip PK, Adams L, et al. A Single Bolus of Docosahexaenoic Acid Promotes Neuroplastic Changes in the Innervation of Spinal Cord Interneurons and Motor Neurons and Improves Functional Recovery after Spinal Cord Injury. *J Neurosci Off J Soc Neurosci*. 2015;35(37):12733-12752. doi:10.1523/JNEUROSCI.0605-15.2015
67. Hou S, Duale H, Cameron AA, Abshire SM, Lyttle TS, Rabchevsky AG. Plasticity of lumbosacral propriospinal neurons is associated with the development of autonomic dysreflexia after thoracic spinal cord transection. *J Comp Neurol*. 2008;509(4):382-399. doi:10.1002/cne.21771
68. Fouad K, Torres-Espín A, Fenrich KK. Mechanisms of Behavioral Changes After Spinal Cord Injury. In: *Oxford Research Encyclopedia of Neuroscience*. Accessed January 29, 2025. https://oxfordre.com/neuroscience/display/10.1093/acrefore/9780190264086.001.0001/acrefore-9780190264086-e-245
69. Starkey ML, Bleul C, Kasper H, et al. High-Impact, Self-Motivated Training Within an Enriched Environment With Single Animal Tracking Dose-Dependently Promotes Motor Skill Acquisition and

Functional Recovery. *Neurorehabil Neural Repair*. 2014;28(6):594-605. doi:10.1177/1545968314520721
70. Hatch MN, Cushing TR, Carlson GD, Chang EY. Neuropathic pain and SCI: Identification and treatment strategies in the 21st century. *J Neurol Sci*. 2018;384:75-83. doi:10.1016/j.jns.2017.11.018
71. Gwak YS, Hulsebosch CE, Leem JW. Neuronal-Glial Interactions Maintain Chronic Neuropathic Pain after Spinal Cord Injury. *Neural Plast*. 2017;2017:2480689. doi:10.1155/2017/2480689
72. Danner SM, Shepard CT, Hainline C, Shevtsova NA, Rybak IA, Magnuson DS. Spinal control of locomotion before and after spinal cord injury. *bioRxiv*. Published online June 1, 2023:2023.03.22.533794. doi:10.1101/2023.03.22.533794
73. Battistuzzo CR, Rank MM, Flynn JR, et al. Gait recovery following spinal cord injury in mice: Limited effect of treadmill training. *J Spinal Cord Med*. 2016;39(3):335. doi:10.1080/10790268.2015.1133017
74. Pocratsky AM, Burke DA, Morehouse JR, et al. Reversible silencing of lumbar spinal interneurons unmasks a task-specific network for securing hindlimb alternation. *Nat Commun*. 2017;8:1963. doi:10.1038/s41467-017-02033-x
75. Shepard CT, Pocratsky AM, Brown BL, et al. Silencing long ascending propriospinal neurons after spinal cord injury improves hindlimb stepping in the adult rat. *eLife*. 10:e70058. doi:10.7554/eLife.70058
76. Seibenhener ML, Wooten MC. Use of the Open Field Maze to Measure Locomotor and Anxiety-like Behavior in Mice. *J Vis Exp JoVE*. 2015;(96):52434. doi:10.3791/52434
77. Bolivar VJ, Caldarone BJ, Reilly AA, Flaherty L. Habituation of Activity in an Open Field: A Survey of Inbred Strains and F1 Hybrids. *Behav Genet*. 2000;30(4):285-293. doi:10.1023/A:1026545316455
78. Brenes JC, Padilla M, Fornaguera J. A detailed analysis of open-field habituation and behavioral and neurochemical antidepressant-like effects in postweaning enriched rats. *Behav Brain Res*. 2009;197(1):125-137. doi:10.1016/j.bbr.2008.08.014
79. Ueno H, Takahashi Y, Suemitsu S, et al. Effects of repetitive gentle handling of male C57BL/6NCrl mice on comparative behavioural test results. *Sci Rep*. 2020;10(1):3509. doi:10.1038/s41598-020-60530-4
80. Fullerton JT. Evaluation of research studies: Part IV: Validity and reliability—Concepts and application. *J Nurse Midwifery*. 1993;38(2):121-125. doi:10.1016/0091-2182(93)90146-8
81. Li Y, Oskouian RJ, Day YJ, Kern JA, Linden J. Optimization of a mouse locomotor rating system to evaluate compression-induced spinal cord injury: correlation of locomotor and morphological injury indices. Published online February 1, 2006. doi:10.3171/spi.2006.4.2.165
82. Rodgers JL, Nicewander WA. Thirteen Ways to Look at the Correlation Coefficient. *Am Stat*. 1988;42(1):59-66. doi:10.2307/2685263
83. Chen K, Liu J, Assinck P, et al. Differential Histopathological and Behavioral Outcomes Eight Weeks after Rat Spinal Cord Injury by Contusion, Dislocation, and Distraction Mechanisms. *J Neurotrauma*. 2016;33(18):1667. doi:10.1089/neu.2015.4218
84. Courtine G, Sofroniew MV. Spinal cord repair: advances in biology and technology. *Nat Med*. 2019;25(6):898-908. doi:10.1038/s41591-019-0475-6
85. Rossignol S, Frigon A. Recovery of locomotion after spinal cord injury: some facts and mechanisms. *Annu Rev Neurosci*. 2011;34:413-440. doi:10.1146/annurev-neuro-061010-113746
86. Schucht P, Raineteau O, Schwab ME, Fouad K. Anatomical correlates of locomotor recovery following dorsal and ventral lesions of the rat spinal cord. *Exp Neurol*. 2002;176(1):143-153. doi:10.1006/exnr.2002.7909
87. Asboth L, Friedli L, Beauparlant J, et al. Cortico-reticulo-spinal circuit reorganization enables functional recovery after severe spinal cord contusion. *Nat Neurosci*. 2018;21(4):576-588. doi:10.1038/s41593-018-0093-5
88. Kiehn O. Decoding the organization of spinal circuits that control locomotion. *Nat Rev Neurosci*. 2016;17(4):224-238. doi:10.1038/nrn.2016.9
89. Grillner S, El Manira A. Current Principles of Motor Control, with Special Reference to Vertebrate Locomotion. *Physiol Rev*. 2020;100(1):271-320. doi:10.1152/physrev.00015.2019
90. Abraira VE, Ginty DD. The Sensory Neurons of Touch. *Neuron*. 2013;79(4). doi:10.1016/j.neuron.2013.07.051
91. Azim E, Jiang J, Alstermark B, Jessell TM. Skilled reaching relies on a V2a propriospinal internal copy circuit. *Nature*. 2014;508(7496):357. doi:10.1038/nature13021
92. Courtine G, Bloch J. Defining ecological strategies in neuroprosthetics. *Neuron*. 2015;86(1):29-33. doi:10.1016/j.neuron.2015.02.039
93. Harkema S, Gerasimenko Y, Hodes J, et al. Effect of epidural stimulation of the lumbosacral spinal cord on voluntary movement, standing, and assisted stepping after motor complete paraplegia: a case study.

Lancet Lond Engl. 2011;377(9781):1938-1947. doi:10.1016/S0140-6736(11)60547-3
94. Takeoka A, Vollenweider I, Courtine G, Arber S. Muscle spindle feedback directs locomotor recovery and circuit reorganization after spinal cord injury. *Cell*. 2014;159(7):1626-1639. doi:10.1016/j.cell.2014.11.019
95. Kuhn PL, Wrathall JR. A mouse model of graded contusive spinal cord injury. *J Neurotrauma*. 1998;15(2):125-140. doi:10.1089/neu.1998.15.125
96. Fouad K, Ng C, Basso DM. Behavioral testing in animal models of spinal cord injury. *Exp Neurol*. 2020;333:113410. doi:10.1016/j.expneurol.2020.113410
97. Grau JW, Huang YJ, Turtle JD, et al. When Pain Hurts: Nociceptive Stimulation Induces a State of Maladaptive Plasticity and Impairs Recovery after Spinal Cord Injury. *J Neurotrauma*. 2017;34(10):1873-1890. doi:10.1089/neu.2016.4626
98. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat Biotechnol*. 2019;37(4):451-460. doi:10.1038/s41587-019-0068-4
99. Coifman RR, Lafon S, Lee AB, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proc Natl Acad Sci*. 2005;102(21):7426-7431. doi:10.1073/pnas.0500334102
100. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep*. 2019;9(1):5233. doi:10.1038/s41598-019-41695-z
101. Sturman O, Germain PL, Bohacek J. Exploratory rearing: a context- and stress-sensitive behavior recorded in the open-field test. *Stress*. Published online September 3, 2018. Accessed January 29, 2025. https://www.tandfonline.com/doi/abs/10.1080/10253890.2018.1438405
102. J K, Eh B, A SS, et al. Task-specificity vs Ceiling Effect: Step-training in shallow water after spinal cord injury. *Exp Neurol*. 2010;224(1):178. doi:10.1016/j.expneurol.2010.03.008
103. Smith RR, Shum-Siu A, Baltzley R, et al. Effects of swimming on functional recovery after incomplete spinal cord injury in rats. *J Neurotrauma*. 2006;23(6):908-919. doi:10.1089/neu.2006.23.908
104. Neckel ND, Dai H, Bregman BS. Quantifying Changes Following Spinal Cord Injury with Velocity Dependent Locomotor Measures. *J Neurosci Methods*. 2013;214(1):27. doi:10.1016/j.jneumeth.2013.01.008
105. Hamers FPT, Koopmans GC, Joosten EAJ. CatWalk-assisted gait analysis in the assessment of spinal cord injury. *J Neurotrauma*. 2006;23(3-4):537-548. doi:10.1089/neu.2006.23.537
106. Jordan LM, Sławińska U. Chapter 17 - The Brain and Spinal Cord Networks Controlling Locomotion. In: Faingold CL, Blumenfeld H, eds. *Neuronal Networks in Brain Function, CNS Disorders, and Therapeutics*. Academic Press; 2014:215-233. doi:10.1016/B978-0-12-415804-7.00017-4
107. Guertin PA. The mammalian central pattern generator for locomotion. *Brain Res Rev*. 2009;62(1):45-56. doi:10.1016/j.brainresrev.2009.08.002
108. Drew T, Marigold DS. Taking the next step: cortical contributions to the control of locomotion. *Curr Opin Neurobiol*. 2015;33:25-33. doi:10.1016/j.conb.2015.01.011
109. Lavaud S, D'Andola M, Bichara C, Takeoka A. Electrophysiological signatures reveal spinal learning mechanisms for a lasting sensorimotor adaptation. Published online April 1, 2022:2022.03.30.486422. doi:10.1101/2022.03.30.486422
110. Bertels H, Vicente-Ortiz G, El Kanbi K, Takeoka A. Neurotransmitter phenotype switching by spinal excitatory interneurons regulates locomotor recovery after spinal cord injury. *Nat Neurosci*. 2022;25(5):617-629. doi:10.1038/s41593-022-01067-9
111. Magnuson DSK, Smith RR, Brown EH, et al. Swimming as a model of task-specific locomotor retraining after spinal cord injury in the rat. *Neurorehabil Neural Repair*. 2009;23(6):535-545. doi:10.1177/1545968308331147
112. Fouad K, Metz GAS, Merkler D, Dietz V, Schwab ME. Treadmill training in incomplete spinal cord injured rats. *Behav Brain Res*. 2000;115(1):107-113. doi:10.1016/S0166-4328(00)00244-8
113. Heng C, de Leon RD. Treadmill training enhances the recovery of normal stepping patterns in spinal cord contused rats. *Exp Neurol*. 2009;216(1):139-147. doi:10.1016/j.expneurol.2008.11.023
114. Benevento BT, Sipski ML. Neurogenic Bladder, Neurogenic Bowel, and Sexual Dysfunction in People With Spinal Cord Injury. *Phys Ther*. 2002;82(6):601-612. doi:10.1093/ptj/82.6.601
115. Lemos N, Fernandes GL, Ribeiro AM, et al. Rehabilitation of People With Chronic Spinal Cord Injury Using a Laparoscopically Implanted Neurostimulator: Impact on Mobility and Urinary, Anorectal, and Sexual Functions. *Neuromodulation*. 2023;26(1):233-245. doi:10.1016/j.neurom.2022.01.010
116. Weinreb C, Pearl JE, Lin S, et al. Keypoint-MoSeq: parsing behavior by linking point tracking to pose dynamics. *Nat Methods*. 2024;21(7):1329-1339. doi:10.1038/s41592-024-02318-2

117. Park JH, Kim JH, Oh SK, et al. Analysis of equivalent parameters of two spinal cord injury devices: the New York University impactor versus the Infinite Horizon impactor. *Spine J*. 2016;16(11):1392-1403. doi:10.1016/j.spinee.2016.06.018
118. Otto DJ, Jordan C, Dury B, Dien C, Setty M. Quantifying cell-state densities in single-cell phenotypic landscapes using Mellon. *Nat Methods*. 2024;21(7):1185-1195. doi:10.1038/s41592-024-02302-w

Figure 1

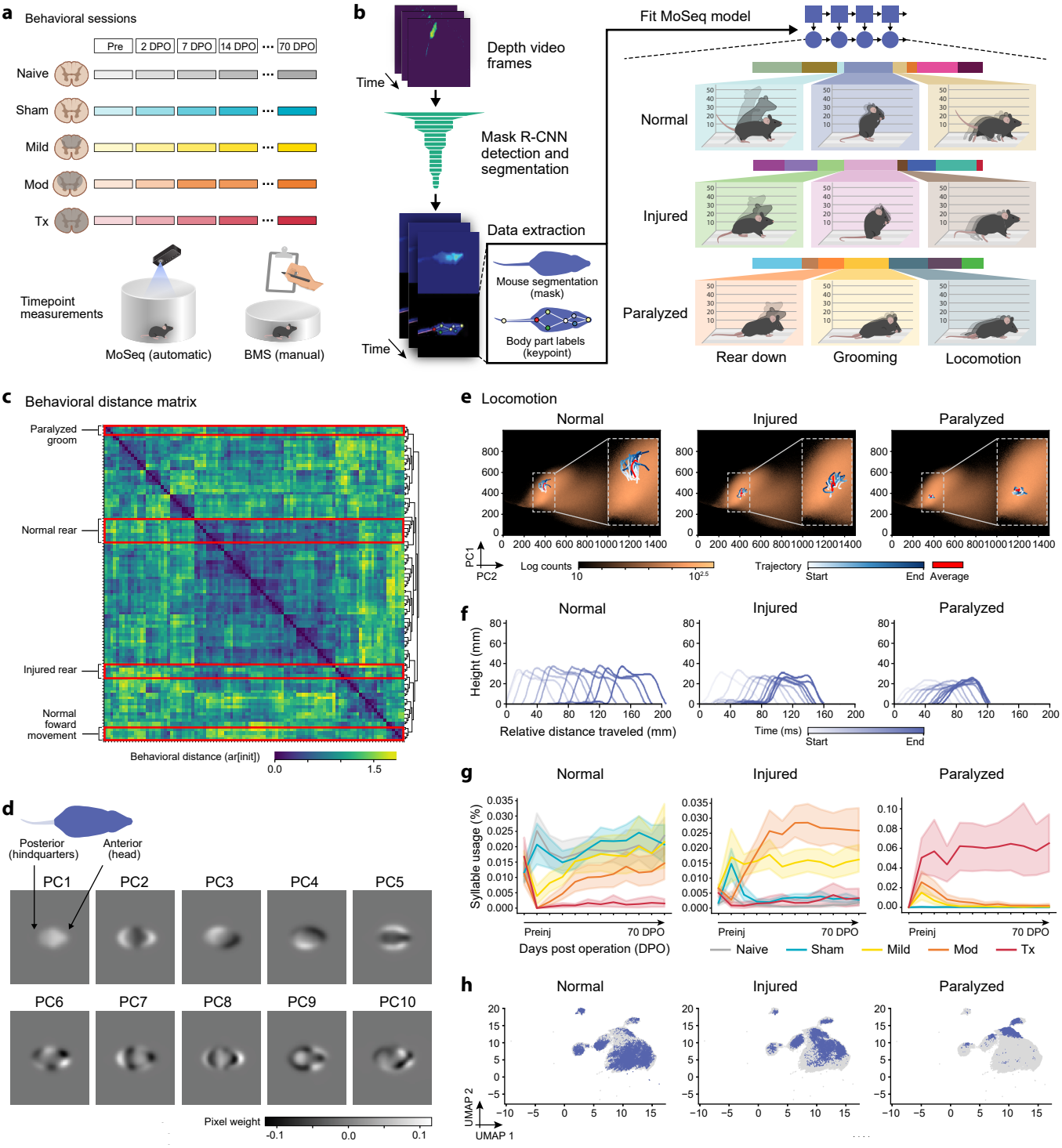

a
Behavioral sessions
Pre
2 DPO
7 DPO
14 DPO
70 DPO
Naive
Sham
Mild
Mod
Tx
Timepoint measurements
MoSeq (automatic)
BMS (manual)
b
Time
Depth video frames
Mask R-CNN detection and segmentation
Data extraction
Mouse segmentation (mask)
Body part labels (keypoint)
Fit MoSeq model
Normal
Injured
Paralyzed
Rear down
Grooming
Locomotion
c
Behavioral distance matrix
Paralyzed groom
Normal rear
Injured rear
Normal foward movement
Behavioral distance (ar[init])
0.0
1.5
d
Posterior (hindquarters)
Anterior (head)
PC1
PC2
PC3
PC4
PC5
PC6
PC7
PC8
PC9
PC10
Pixel weight
-0.1
0.0
0.1
e
Locomotion
Normal
Injured
Paralyzed
PC1
PC2
Log counts
10
$10^{2.5}$
Trajectory
Start
End
Average
f
Height (mm)
Relative distance traveled (mm)
Time (ms)
Start
End
g
Syllable usage (%)
Preinj
70 DPO
Days post operation (DPO)
Naive
Sham
Mild
Mod
Tx
h
UMAP 2
UMAP 1

Figure 2

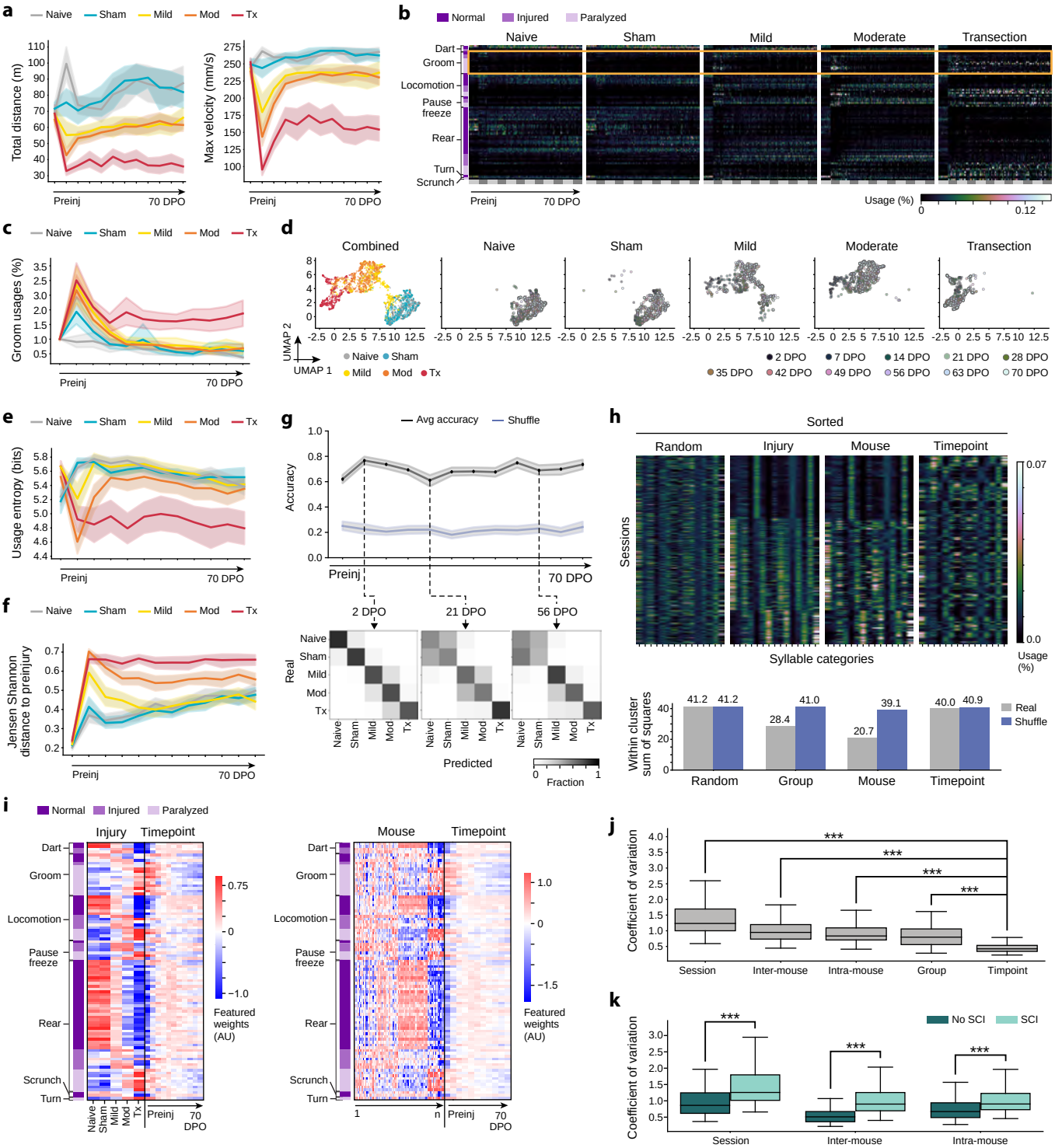

a
Naive Sham Mild Mod Tx
Total distance (m)
Max velocity (mm/s)
Preinj
70 DPO
b
Normal Injured Paralyzed
Naive Sham Mild Moderate Transection
Dart
Groom
Locomotion
Pause freeze
Rear
Turn
Scrunch
Usage (%)
0 0.12
c
Groom usages (%)
d
Combined Naive Sham Mild Moderate Transection
UMAP 2
UMAP 1
Naive Sham Mild Mod Tx
2 DPO 7 DPO 14 DPO 21 DPO 28 DPO 35 DPO 42 DPO 49 DPO 56 DPO 63 DPO 70 DPO
e
Usage entropy (bits)
f
Jensen Shannon distance to preinjury
g
Avg accuracy Shuffle
Accuracy
2 DPO 21 DPO 56 DPO
Real
Predicted
Fraction
h
Sorted
Random Injury Mouse Timepoint
Sessions
Syllable categories
Usage (%)
Within cluster sum of squares
41.2 41.2 28.4 41.0 20.7 39.1 40.0 40.9
Real Shuffle
Random Group Mouse Timepoint
i
Injury Timepoint
Mouse Timepoint
Featured weights (AU)
j
Coefficient of variation
Session Inter-mouse Intra-mouse Group Timpoint
k
No SCI SCI
Session Inter-mouse Intra-mouse

Figure 3

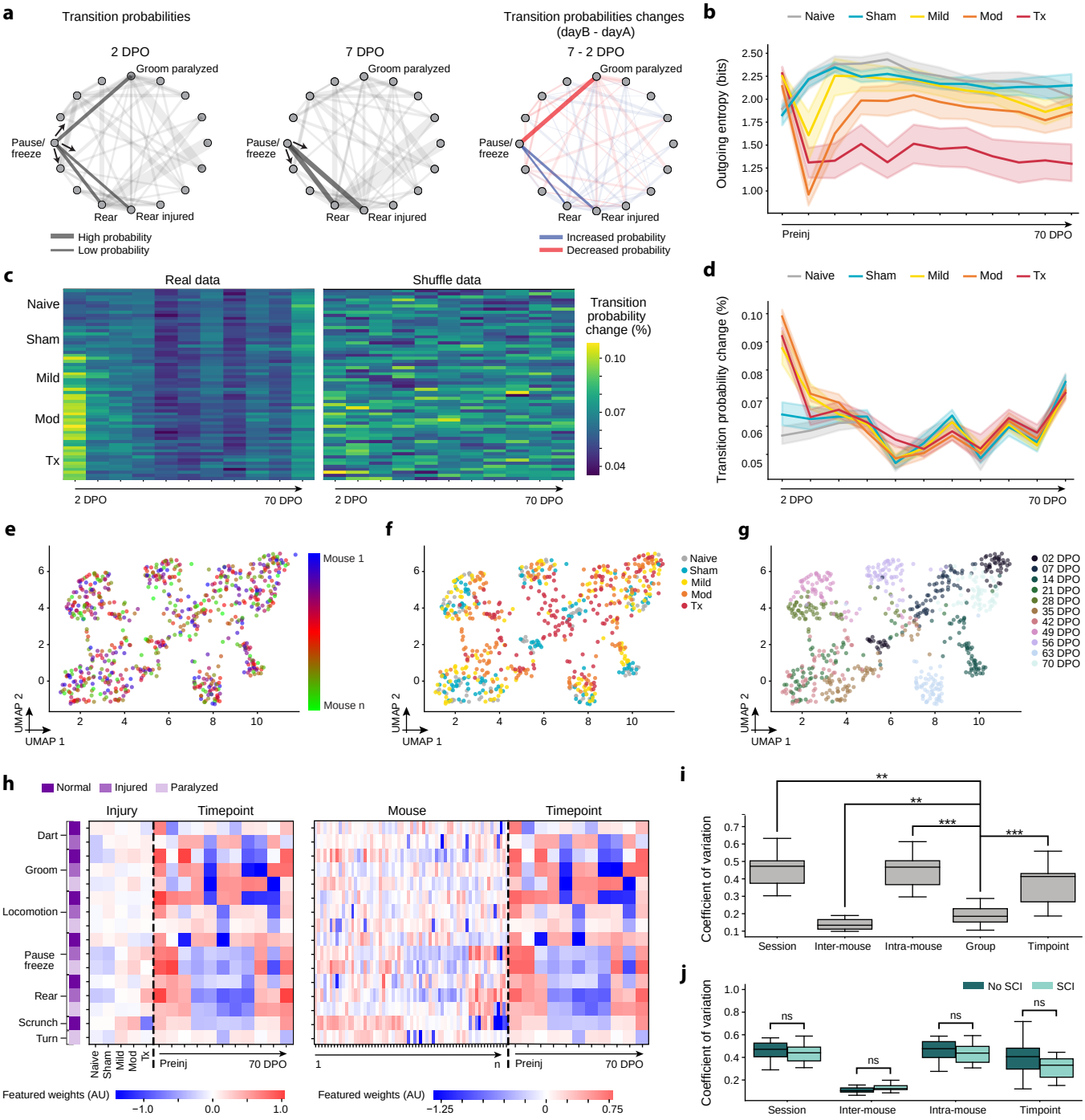

a
Transition probabilities
2 DPO
7 DPO
Transition probabilities changes (dayB - dayA)
7 - 2 DPO
Groom paralyzed
Pause/ freeze
Rear
Rear injured
High probability
Low probability
Increased probability
Decreased probability
b
Naive
Sham
Mild
Mod
Tx
Outgoing entropy (bits)
Preinj
70 DPO
c
Real data
Shuffle data
Naive
Sham
Mild
Mod
Tx
Transition probability change (%)
2 DPO
70 DPO
d
Transition probability change (%)
e
Mouse 1
Mouse n
UMAP 1
UMAP 2
f
g
02 DPO
07 DPO
14 DPO
21 DPO
28 DPO
35 DPO
42 DPO
49 DPO
56 DPO
63 DPO
70 DPO
h
Normal
Injured
Paralyzed
Injury
Timepoint
Mouse
Dart
Groom
Locomotion
Pause freeze
Rear
Scrunch
Turn
Featured weights (AU)
i
Coefficient of variation
Session
Inter-mouse
Intra-mouse
Group
Timpoint
j
No SCI
SCI
ns

Figure 4

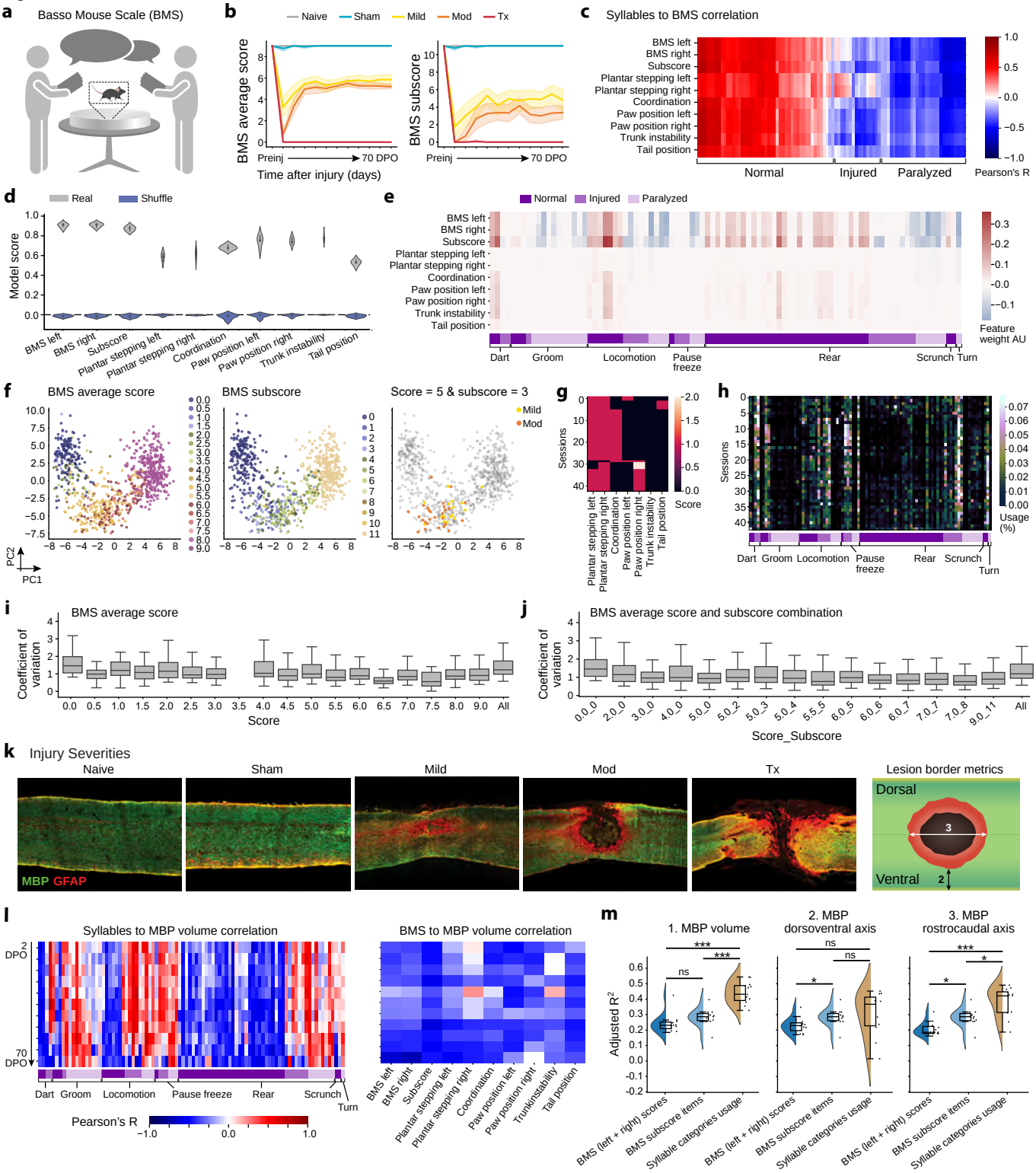
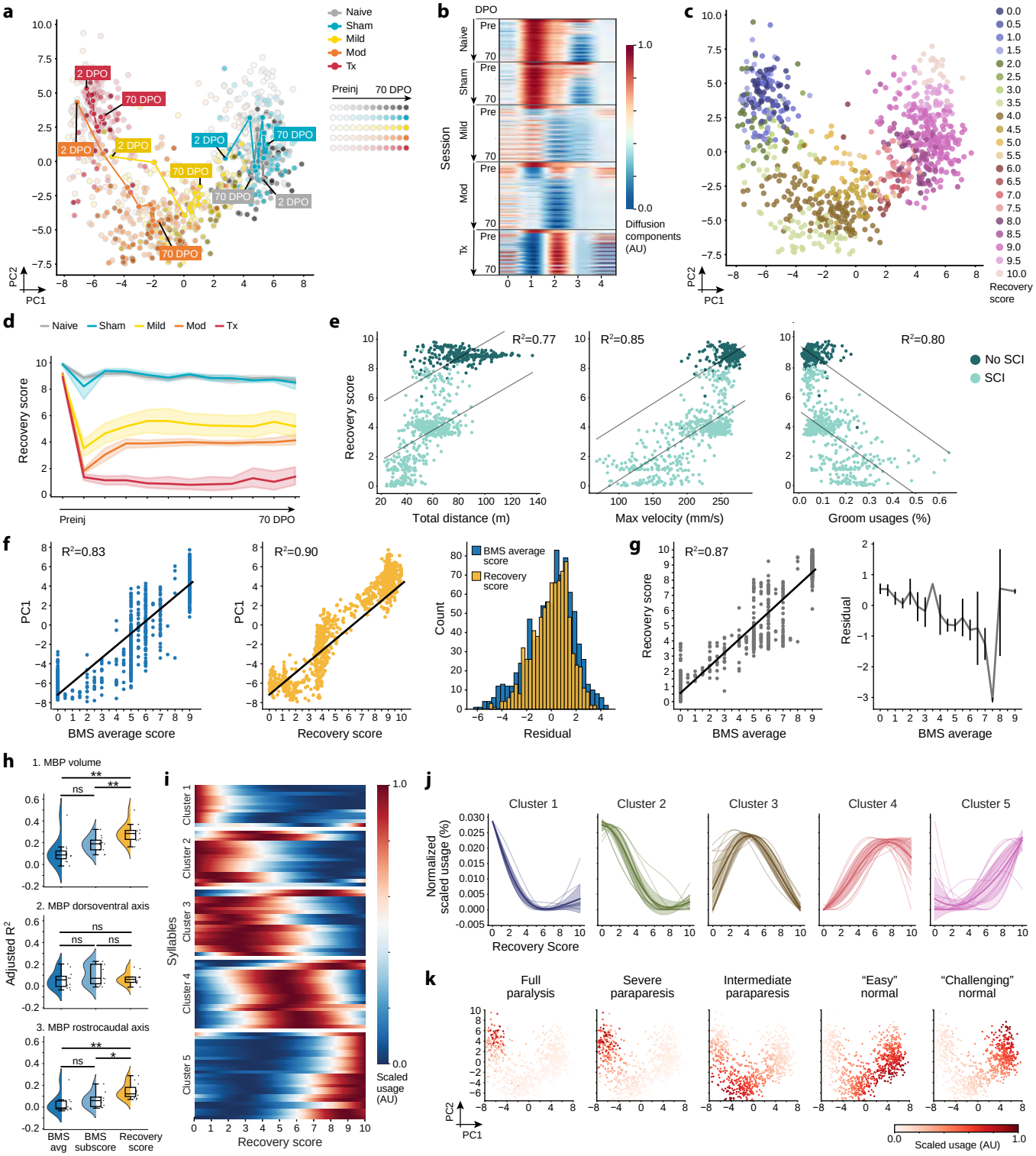
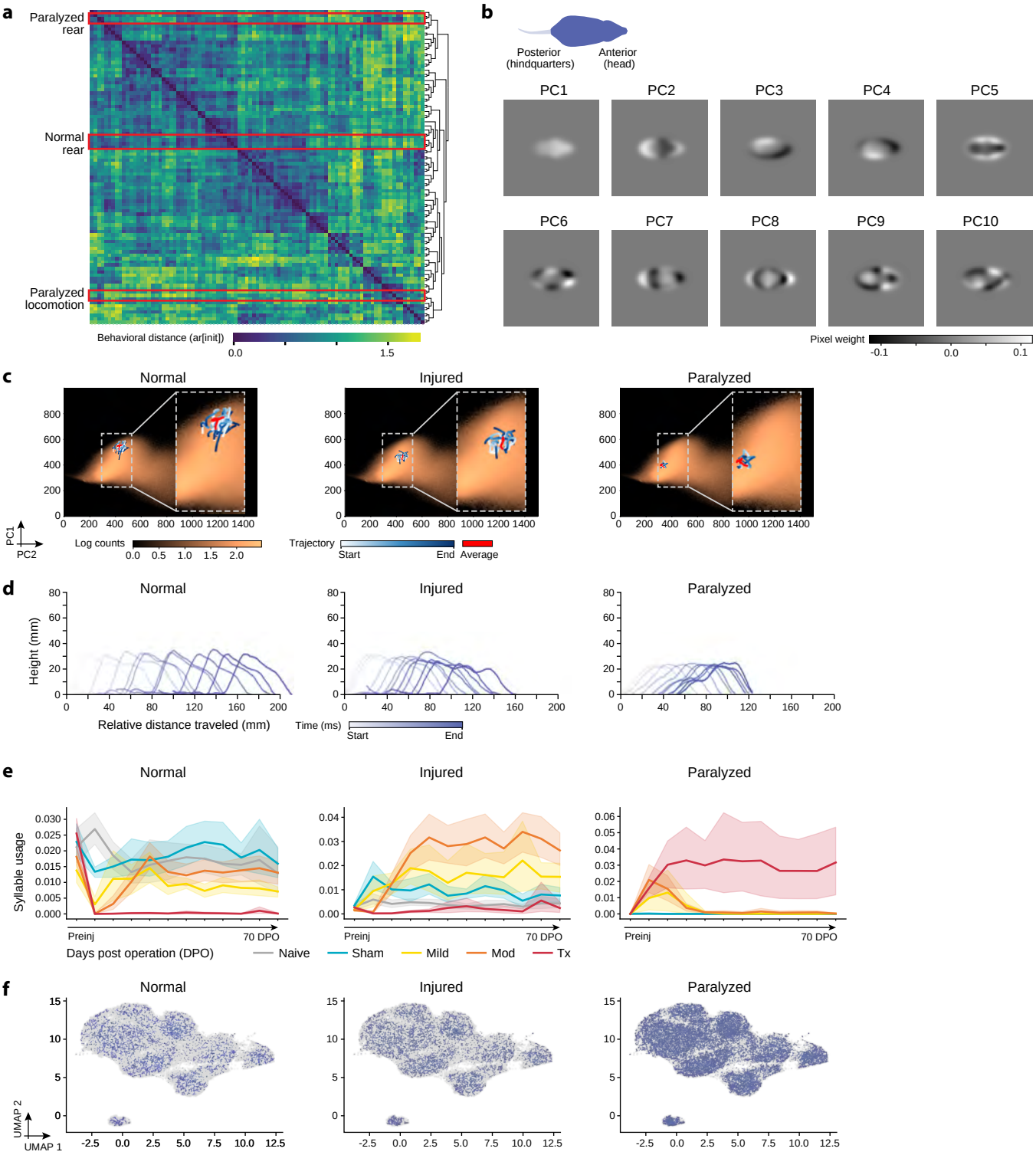

a
Basso Mouse Scale (BMS)
b
Naive
Sham
Mild
Mod
Tx
BMS average score
BMS subscore
Preinj
70 DPO
Time after injury (days)
c
Syllables to BMS correlation
BMS left
BMS right
Subscore
Plantar stepping left
Plantar stepping right
Coordination
Paw position left
Paw position right
Trunk instability
Tail position
Normal
Injured
Paralyzed
Pearson's R
d
Real
Shuffle
Model score
e
Feature weight AU
Dart
Groom
Locomotion
Pause freeze
Rear
Scrunch
Turn
f
BMS average score
BMS subscore
Score = 5 & subscore = 3
Mild
Mod
PC1
PC2
g
Sessions
Score
h
Usage (%)
i
Coefficient of variation
Score
j
BMS average score and subscore combination
Score_Subscore
k
Injury Severities
Naive
Sham
Mild
Mod
Tx
Lesion border metrics
Dorsal
Ventral
MBP GFAP
l
Syllables to MBP volume correlation
BMS to MBP volume correlation
2 DPO
70 DPO
Pearson's R
m
1. MBP volume
2. MBP dorsoventral axis
3. MBP rostrocaudal axis
Adjusted R²
BMS (left + right) scores
BMS subscore items
Syllable categories usage
ns
***
*

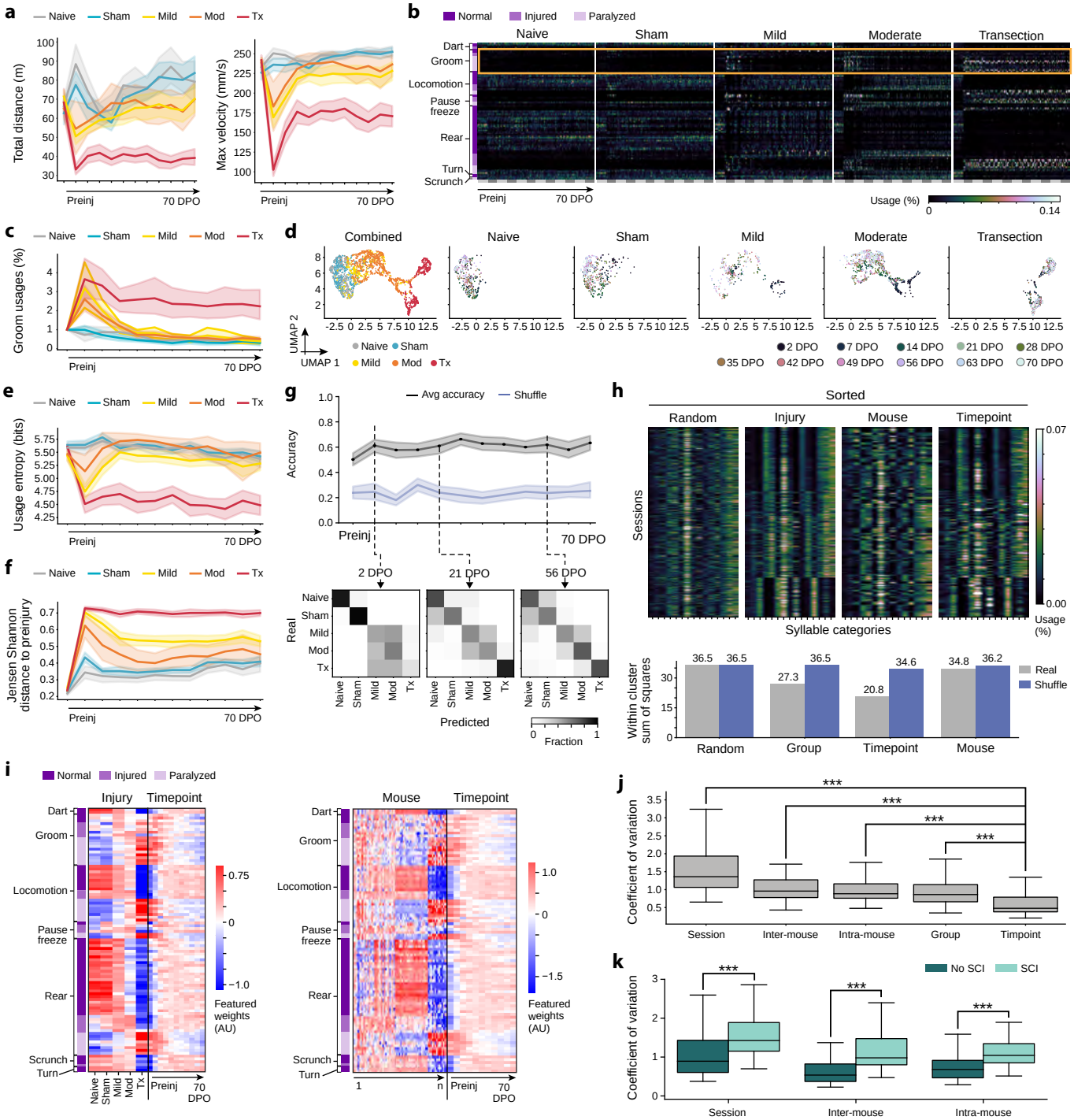
Figure 5


a
Naive
Sham
Mild
Mod
Tx
Preinj
70 DPO
2 DPO
70 DPO
PC2
PC1
b
DPO
Session
Pre
70
Diffusion components (AU)
c
Recovery score
d
Recovery score
Preinj
70 DPO
e
R²=0.77
R²=0.85
R²=0.80
No SCI
SCI
Total distance (m)
Max velocity (mm/s)
Groom usages (%)
f
R²=0.83
R²=0.90
PC1
BMS average score
Recovery score
Count
Residual
g
R²=0.87
Recovery score
BMS average
Residual
h
1. MBP volume
2. MBP dorsoventral axis
3. MBP rostrocaudal axis
Adjusted R²
ns
BMS avg
BMS subscore
Recovery score
i
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Syllables
Scaled usage (AU)
Recovery score
j
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Normalized scaled usage (%)
Recovery Score
k
Full paralysis
Severe paraparesis
Intermediate paraparesis
"Easy" normal
"Challenging" normal
PC2
PC1
Scaled usage (AU)

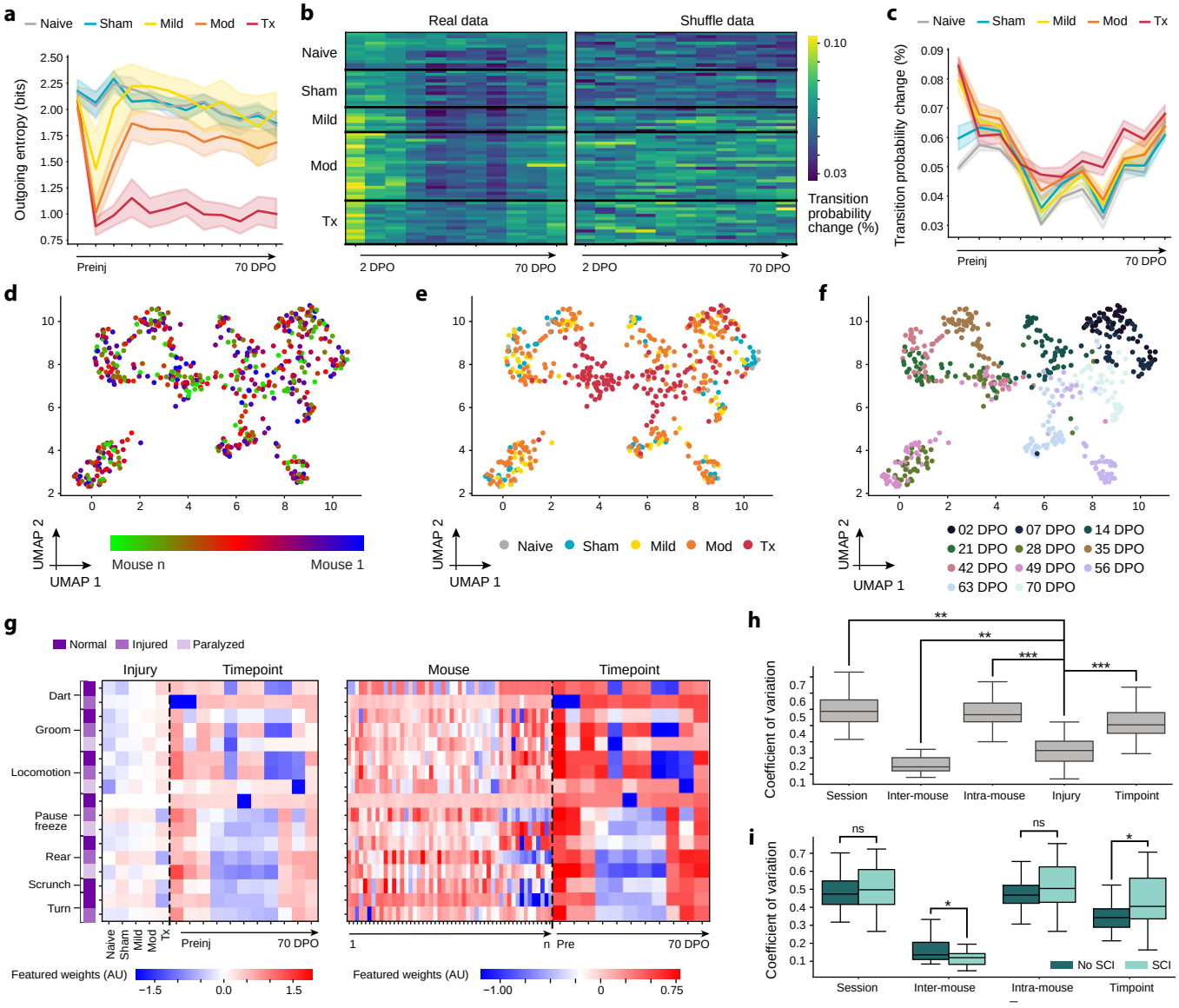
Extended Data Figure 1



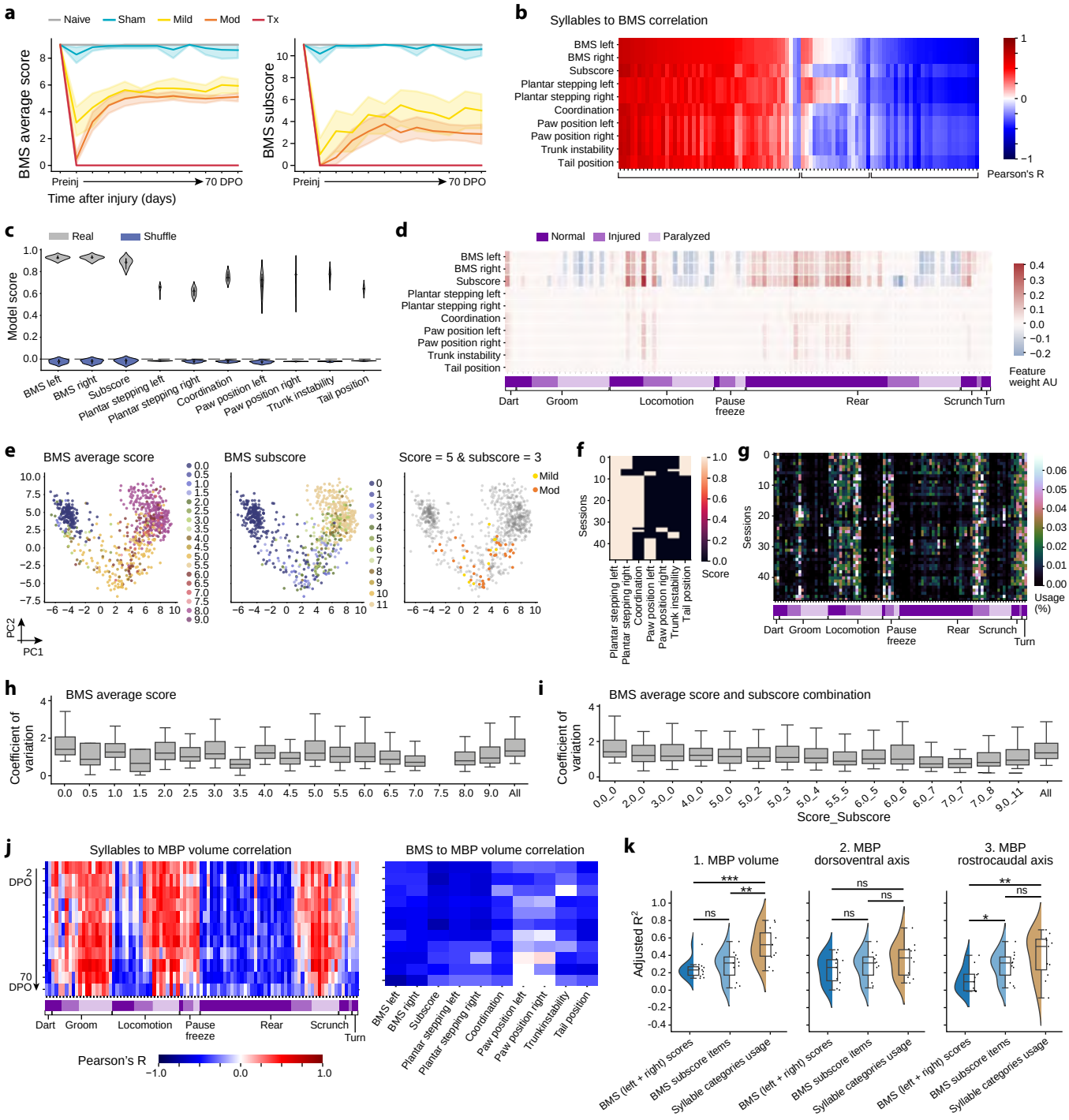
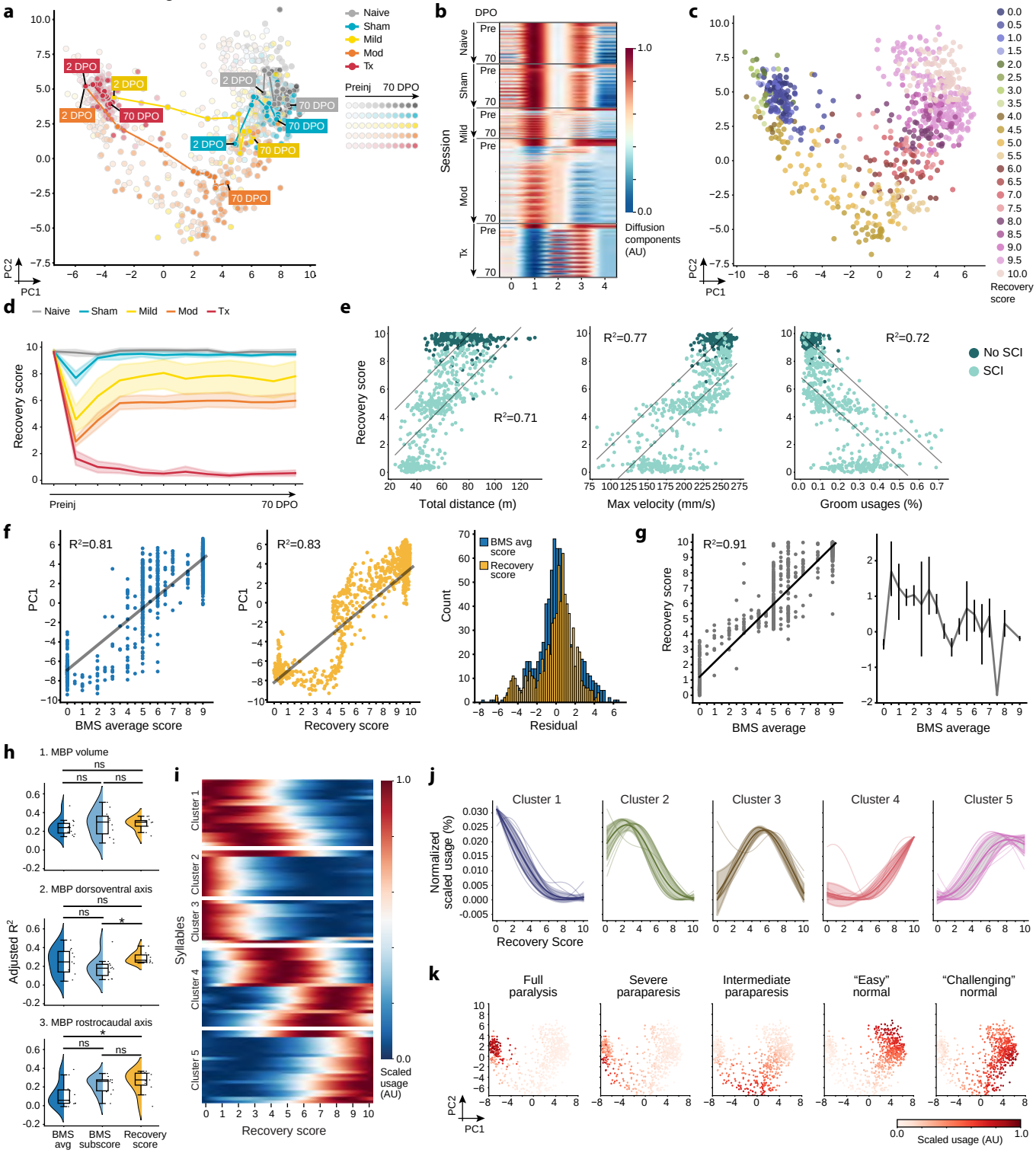
Extended Data Figure 2



Extended Data Figure 3



Extended Data Figure 4



Extended Data Figure 5