TITLE

Using behavioral biomarkers to redefine epochs of spontaneous recovery following spinal cord injury

AUTHORS

Jaclyn T Eisdorfer*[1,2], Josh Thackray*[1], Thomas Theis*[2], Ana Vivinetto[3], Matthew T Ricci[4], Sherry Lin[5], Olisemeka Oputa[6], Alana M Martinez[1], Hannah D Nacht[1], Monica Tschang[7], Malaika Mahmood[8], Ashley Tucker[9], Manon Bohic[1], Shailee Pusuloori[1], Lance Zmoyro[1], Abraira Lab Computational Group[1], Phillip Popovich[10], Adam R. Ferguson[11], Dana McTigue[10], Vicki M Tysseling[6], Jennifer Dulin[9], Edmund Hollis II[3], Sandeep Robert Datta[5], Victoria E Abraira[1]

*Asterisks denote co-first author.

1. Rutgers, The State University of New Jersey;
2. AstraZeneca;
3. Burke Neurological Institute/Weill Cornell Medicine;
4. Brown University;
5. Harvard Medical School;
6. Northwestern University;
7. University of Washington;
8. University of Pennsylvania;
9. Texas A M;
10. Ohio State;
11. University of California San Francisco (UCSF)

ABSTRACT

Uncovering the intricacies of the recovery trajectory following spinal cord injury (SCI) has remained a critical exploration for researchers and clinicians, fostering the need of innovative approaches to offer insight into the underlying dynamics of this complex phenomenon. Existing methods, such as the Basso Mouse Scale (BMS) and kinematic analyses, have provided valuable insights, yet limitations in their ability to comprehensively capture behavioral nuances call for more sophisticated approaches. This study addresses segregating the intricate trajectory of recovery following SCI into discrete epochs through the use of behavioral biomarkers. Leveraging a machine learning-driven video analysis technique known as Motion Sequencing (MoSeq), we identified distinct behavioral modules, elucidating shared patterns across diverse injury severities. Our analysis highlights the correlation between these behavioral biomarkers and established recovery metrics, such as BMS criteria and histological markers. Importantly, behavioral biomarkers enabled for deeper understanding of mouse behavior, capturing nuanced features often overlooked by traditional measures. These findings underscore the potential of behavioral biomarkers in characterizing discrete recovery epochs and signatures at the transition from one phase to the next.

INTRODUCTION

Unraveling the nuanced trajectory of recovery following spinal cord injury (SCI) has remained a pivotal quest, fostering the need of innovative approaches to offer insight into the underlying dynamics of this complex phenomenon. Pioneering efforts in the rodent model, such as the Basso Mouse Scale (BMS)[1] and kinematic analyses[2], have yielded invaluable insights into the intricate neural mechanisms and observable behavioral patterns indicative of recovery. However, these methods predominantly focus on isolated aspects of behavior, such as stepping patterns and joint actions, thereby limiting the comprehensive understanding of each recovery epoch in the rodent. Consequently, there remains a conspicuous gap in fully elucidating complex motor behavior, leading to an incomplete grasp of the constituent behaviors characterizing individual epochs and the transitional behaviors signaling shifts between these distinct phases.

Behavior serves as a tangible reflection of underlying neural mechanisms[3–7]. Therefore, leveraging behavioral patterns not only aids in predicting underlying neural changes but also serves as a valuable therapeutic strategy to enhance and optimize the recovery process. Nevertheless enhancing our comprehension of the recovery process is crucial, as it can help alleviate misinterpretations and temper overinflated expectations during the translation of research findings[8]. Previous literature has underscored the importance of accurately delineating recovery epochs, highlighting their pivotal role in tailoring patient rehabilitation programs and therapeutic interventions. Leveraging behavioral indicators indicative of the transitional phases between these epochs has the potential to refine the targeting of therapeutic strategies, thereby potentially optimizing the overall recovery trajectory. Even incremental advancements resulting from such interventions can yield substantial improvements in a patient's overall quality of life[9], illustrating the significance of this research direction in the field of SCI rehabilitation.

This study utilizes behavioral biomarkers as a tool to redefine the delineation of recovery epochs and reveal shared behavioral patterns across varying degrees of injury severity. We employed a machine learning-driven, artificial intelligence, three-dimensional video analysis technique known as Motion Sequencing[10–15]. MoSeq integrates depth imaging with unsupervised computational algorithms, enabling the segmentation of mouse behavior into succinct (~350 ms) behavioral "modules," including actions like “locomotion,” “rearing,” and “grooming” that occur in the preinjury/normal condition and adaptive counterparts that arise following SCI. Modules are characterized by their stereotyped nature, reusability, and concatenation into more extensive complex behaviors. This method facilitates the phenotyping of mice, allowing for the identification of deleterious or novel behaviors that may emerge due to dysfunction within spinal cord circuits[11,14], particularly in the aftermath of SCI. Importantly, MoSeq captures an array of behavioral features, including those that might evade detection when focusing solely on traditional outcome measures such as individual joint movement. Our data-driven analysis reveals that behavioral biomarkers effectively partition the recovery process into discrete epochs, indicating common movement patterns across various injury severities. Additionally, we find strong correlations between these behavioral biomarkers and established metrics of recovery assessment, including individual components of the Basso Mouse Scale (BMS), as well as histological markers quantifying injury severity. Notably, the insights provided by behavioral biomarkers offer a more nuanced understanding of movement dynamics within each specific recovery epoch.

RESULTS

Behavioral biomarkers segmentally organize recovery following SCI.

To gain insights into the dynamics of recovery, we deconstructed complex motor behavior pre- and post-SCI using unsupervised machine learning approaches[10–15]. We employed four distinct models of SCI — Transection ("Tx") injuries, Moderate ("Mod," 50 kDyn) contusion injuries, Mild (30 kDyn) contusion injuries, and Sham injuries (laminectomy only) — as well as included a fifth group of Naive animals for comparative analysis (**Supplemental Figure 1A**). Naive animals did not have surgery, but they were exposed to operative anesthesia for the approximate duration of a SCI surgery as well as postoperative analgesics. Our experimental timeline included recording naturalistic behavior in 20 min bouts before surgery, 2 days post surgery, 7 days post surgery, and every week thereafter or 10 weeks total (**Supplemental Figure 1B**).

Figure 1A illustrates our motion sequence (MoSeq) modeling process, leveraging a downward-facing depth camera to capture spontaneous behavior in a freely behaving mouse. Each "mouse" pixel was characterized by three-dimensional coordinates (x, y, z — where z represents the mouse's height above the ground of the arena). Subsequently, a hyperparameter enabled the identification of prevalent behavioral shifts at the sub-second to second timescale, demarcating distinctive behavioral modules such as "rear," "locomotion," and "groom," among others, observed across both normal and injured mice. We confirmed the precision of our models in distinguishing between various groups and time points (Supplemental Figure 1C-E).

In **Figure 1B**, the heatmap rows depict the distribution of module similarities with module numbers clustered into a dendrogram using a hierarchical clustering approach with Euclidean distances. The columns are ordered by module usages exhibited by each group and time point using the same clustering approach. Employing a data-driven strategy, modules naturally coalesce into distinct epochs regardless of the severity of the injury. Prominent clusters or “usage hotspots” emerge, particularly: in the heatmap's top left corner, housing modules signifying paralysis, encompassing the Tx group and early time points of the Mod and Mild groups; the middle of the heatmap's lower section, embodying modules reflective of recovery or adaptive behaviors, representing mid to late time points of the "Mod" and "Mild" groups; and lastly, the middle of the heatmap's right side accommodates clusters of animals exhibiting normal behavior, including the Sham, Naive, and the preinjury time point.

The characterization of recovery, represented at the bottom of the heatmap's central region, is more finely delineated through an analysis of the *F* univariate statistical test[10] (**Figure 1C**). This analytical examination more clearly ascribes the most relevant modules for discerning differences between recovery epochs. The rationale underscores the significance of modules whose frequency of use is statistically interdependent on a particular condition, thus serving as valuable markers for linear classification and indicative of that condition. These findings reveal various facets of recovery, each distinguished by its distinctive module usage patterns and marked transitions towards modules considered "normal." Furthermore, the results illustrate that behavioral patterns observed two days after Sham injuries align with later stages of recovery from SCI, implying that laminectomies may not serve as true "sham" controls.

Upon reducing the data's dimensionality to a 2-dimensional linear discriminant analysis (LDA) space, a structured representation of recovery emerges (**Figure 1D and Supplemental Figure 1F**). Within this spatial arrangement, modules linked with paralysis are positioned in the upper left quadrant of the LDA space, while modules associated with normal behavior occupy the upper right quadrant. Serving as the intermediary, a recovery trajectory emerges between these two quadrants, demonstrating an overlap between the moderate and mild injury groups. The LDA space mirrors the “usage hotspots” in Figure 1B’s heatmap. This insight underscores the intrinsic nature of recovery patterns, which, when projected into the LDA space, reveals distinct clusters of behaviors that persist regardless of the severity of the initial injury.

The discovery of these enduring behavioral clusters illuminates a fresh perspective on the intricacies of the recovery process. Behavioral biomarkers can thus unveil distinct recovery epochs, forming cohesive modules that transcend traditional injury severity categorizations.

Behavioral biomarkers correlate to traditional recovery outcome measures.

The Basso Mouse Scale (BMS) stands as the established metric for quantifying and comparing recovery among groups of SCI mice[1]. This 9-point scale provides a coarse evaluation of locomotor capability, utilizing indicators like Plantar Stepping, Coordination, and Trunk Stability, alongside other locomotion features. Notably, our experimental groups adhered to conventional recovery patterns (**Supplemental Figure 2A**). To confirm the applicability of behavioral biomarkers in gauging recovery, we conducted Pearson's correlation analysis linking BMS criteria with modules (**Figure 2A**). Our findings revealed a robust negative (blue) correlation between modules denoting paralysis (top left usage hotspot on heatmap) and BMS criteria indicative of normal locomotion or advanced recovery stages, including Plantar Stepping and Coordination. Similarly, we observed a significant positive (red) correlation between modules reflecting uninjured movements (bottom right usage hotspot on heatmap) and BMS criteria signifying regular locomotion or advanced recovery phases, such as "up" Tail Position and Trunk Stability. Specifically focusing on the usage hotspot in the bottom center of the heatmap, in columns representing the Mod (orange) or Mild (green) groups, we excitingly found a substantial positive correlation between modules illustrating adaptive behavior during recovery and Plantar Stepping (Right, Left) and BMS scores (Left, Right, Avg). However, this correlation was comparatively less positive, slightly negative, or absent in the case of BMS criteria Coordination, Tail Position, Paw Position, and Trunk Stability, aligning with expectations. Indeed, during the intermediate phase of recovery, animals exhibit plantar stepping but do not typically manifest coordinated stepping, elevated tail positions, parallel paw positions, or stable trunks.

With Pearson's correlation analysis, we explored the association of behavioral biomarkers with prevalent histological indicators linked to injury (**Figure 2AB**). We included Myelin Basic Protein (MBP, green)[16], a marker of damaged central nervous system (CNS) myelinated axons resulting from injury, and Glial Fibrillary Acidic Protein (GFAP, red)[17], a CNS astrocytic marker and a constituent of astrocytes that aggregate around the lesion post-injury, in the analysis. Additionally, we examined the inflammation-to-lesion volume ratio (MBP/GFAP ratio). We observed a notably strong positive correlation between larger MBP and GFAP volumes and behavioral biomarkers suggestive of paralysis, matching anticipated outcomes. Moreover, irrespective of the lesion volume (GFAP), increased inflammation (MBP volume) demonstrated a positive association with modules reflecting both paralysis and recovery. However, larger lesion volumes showed notably strong positive correlations primarily with paralysis modules.

We conducted an analysis of entropy, representing the capacity for diverse behavioral manifestations, in relation to BMS criteria (**Figure 1C**) and lesion severity (**Figure 1DE**). In the context of SCI, greater entropy reflects a broader range of behavioral patterns, while lower entropy signifies limited behavioral variations. Post-paralysis, the reduction in behavioral diversity is observed due to the limitation in available behavioral biomarkers. Notably, we observed a significant positive correlation between higher entropy and BMS criteria linked to normal locomotion or recovery (Figure 1C). Additionally, we identified a strong association between higher entropy and less severe injuries (Figure 1D, GFAP: slope = -0.08719611; y-intercept = 0.62797916), as well as reduced inflammation (Figure 1E, MBP: slope = -0.09328262; y-intercept = 0.84207945).

These findings suggest a robust correlation between behavioral biomarkers and the established gold standard metrics used to assess recovery in SCI cases. Moreover, the utilization of

behavioral biomarkers offers an enriched dataset, capturing nuanced movement patterns that are often not accounted for in conventional metrics.

DISCUSSION

Our investigation sought to redefine the course of spontaneous recovery post-spinal cord injury (SCI) through the utilization of computationally identified behavioral biomarkers. Our findings unveiled persistent behavioral clusters, independent of the severity of the injury. These behavioral signatures not only exhibited correlations with conventional measures assessing functional impairments and injury severity but also provided a nuanced portrayal of the recovery process, emphasizing the considerable potential inherent in behavioral biomarkers.

Delving deeper into the granularity of behavioral modules can offer insights into specific signatures crucial for recovery, even when adaptive behavior appears to stabilize. This nuanced approach challenges the conventional notion of a recovery plateau and underscores the potential for continued progress as identified through behavioral biomarkers. These modules may be positioned at the intersection of epochs, and targeting them with therapies may facilitate the transition from one phase to the next. Furthermore, establishing correlations between behavioral signatures and underlying activities, such as brain and muscle activity, alongside histological markers, presents a promising avenue for comprehending the intricate interplay between behavior and underlying neural circuit changes.

MATERIALS AND METHODS

Subjects

Male and female 8- to 9-week-old C57BL/6 (Strain Code 027) mice 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[1]. 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 kDyne was used for mild injuries, and a force of 50 kDyne was employed for moderate injuries. 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[1,18] 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

At 2 days and 1-10 weeks post-spinal cord injury, locomotor recovery in mice was assessed using the Basso Mouse Scale (BMS)[1]. 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: 3D pose analysis

Data acquisition

Here and below, we employed methodology as previously described[10–13] and customized it as required.

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[19], 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 scaleign, 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.

	AP	**AP50**	**AP75**
bbox	64.812	97.202	73.423
keypoints	45.851	93.768	37.795
segm	55.680	98.673	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).

Dimensionality Reduction and Behavioral Segmentation

We utilized moseq2-pca (v1.1.3) to reduc the dimensionality of the extracted depth data through principal component analysis (PCA). This involved projecting the extracted depth data onto the first 10 learned PCs, resulting in a 10-dimensional time series that characterizes the 3D pose dynamics of the mouse. Moreover, we conducted a model-free changepoint analysis, which identifies abrupt discontinuities in mouse pose dynamics, offering an empirical estimation of timescale behaviors without any model constraints. The quality of the results was evaluated by visually examining the pixel weights assigned by each PC and the cumulative distribution of the percent variance explained by each PC. Subsequently, we trained an autoregressive hidden Markov model (AR-HMM, also known as the “MoSeq model”) using moseq2-model (v1.1.2). This model consisted of a vector autoregressive process that captured the evolution of the first 10 PCs over time for each state (i.e., behavioral module) and a hidden Markov model that described the switching dynamics between states. A maximum of 100 behavioral states was permitted.

Identifying behavioral module usage and transition probabilities

Module usage (percentage) was determined by tallying the number of emissions for each module and dividing it by the total sum of module emissions in each recording session. Transition matrices were computed by tallying the total number of occurrences of each module transitioning to any other module and then dividing by the sum of the matrix (bigram normalization). Statistical analysis followed previously described methods[11]. In brief, group comparisons were executed for each module by generating 1,000 bootstrap samples of the module’s frequency for each group, followed by a z-test conducted on these two distributions. The Benjamini-Hochberg procedure was implemented to correct for multiple hypothesis testing and regulate the false discovery rate.

We conducted linear discriminant analysis (LDA) using the scikit-learn[20] implementation with the eigen solver and two components. The data was divided into train and test subsets at a 70:30 ratio, and LDA models were trained on the train subset. The LDA model received group labels, along with normalized module usages or bigram transition probabilities as inputs, and its performance was evaluated on the test subset. Additionally, we performed a permutation test (sklearn.model_selection.permutation_test_score). In this procedure, a collection of models was trained using 100 sets of randomly permuted labels. The resulting distribution of model scores from the shuffled data was compared to the final model score to derive a p-value. Plots were generated using seaborn and matplotlib.

Identifying important features

We employed an *F* univariate statistical test as previously described[10–13] to discern the most discriminatory modules among groups (Tx, Mod, Mild, Sham, and Naive). The rationale was to identify modules whose usage frequency strongly correlated with specific conditions, making them valuable for linear classification and characteristic of the corresponding state.

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.

DATA AND CODE AVAILABILITY

LINK(s):

- https://github.com/tischfieldlab/moseq2-detectron-extract
- https://github.com/tischfieldlab/moseq-reports

REFERENCES

1. 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

2. 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

3. 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

4. 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

5. 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

6. Takeoka A, Arber S. Functional Local Proprioceptive Feedback Circuits Initiate and Maintain Locomotor Recovery after Spinal Cord Injury. *Cell Rep*. 2019;27(1):71-85.e3. doi:10.1016/j.celrep.2019.03.010

7. 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

8. Reier PJ, Lane MA, Hall ED, Teng YD, Howland DR. Translational spinal cord injury research: preclinical guidelines and challenges. *Handb Clin Neurol*. 2012;109:411-433. doi:10.1016/B978-0-444-52137-8.00026-7

9. Kelly-Hedrik M, Abd-El-Barr MM, Aarabi B, et al. Importance of Prospective Registries and Clinical Research Networks in the Evolution of Spinal Cord Injury Care. *J Neurotrauma*. 2023;40(17-18):1834-1848. doi:10.1089/neu.2022.0450

10. 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

11. Wiltschko AB, Johnson MJ, Iurilli G, et al. Mapping Sub-Second Structure in Mouse Behavior. *Neuron*. 2015;88(6):1121-1135. doi:10.1016/j.neuron.2015.11.031

12. 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

13. 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

14. Gradwell MA, Ozeri-Engelhard N, Eisdorfer JT, et al. Multimodal sensory control of motor performance by glycinergic interneurons of the spinal cord deep dorsal horn. Published online August 1, 2023:2022.05.21.492933. doi:10.1101/2022.05.21.492933

15. 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

16. Marinelli S, Vacca V, De Angelis F, et al. Innovative mouse model mimicking human-like features of spinal cord injury: efficacy of Docosahexaenoic acid on acute and chronic phases. *Sci Rep*. 2019;9. doi:10.1038/s41598-019-45037-x

17. Vivinetto AL, Kim ID, Goldberg DC, et al. Zeb2 Is a Regulator of Astrogliosis and Functional Recovery after CNS Injury. *Cell Rep*. 2020;31(13):107834. doi:10.1016/j.celrep.2020.107834

18. 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

19. Wu Y, Kirillov A, Massa F, Lo WY, Girshick R. Detectron2. Published online 2019. https://github.com/facebookresearch/detectron2

20. Buitinck L, Louppe G, Blondel M, et al. API design for machine learning software: experiences from the scikit-learn project. Published online September 1, 2013. doi:10.48550/arXiv.1309.0238

FIGURE CAPTIONS:

Figure 1. Behavioral biomarkers segmentally organize recovery following SCI. (A) Illustrative schematic of machine learning-guided motion sequence (MoSeq) pipeline, including recording of 20 min bouts of naturalistic behavior using a depth camera, data extraction, dimensionality reduction using principal component analysis (PCA), and model fitting. Point clouds denote normal movements (top) and injured counterparts (bottom). **(B)** Heatmap denoting module usages across all groups and time points. Inset, left: The five groups include Transection (Tx, pink), Moderate contusion (Mod, orange), Mild contusion (green), Sham injury (purple), and Naive (blue). The following 12 time points were recorded: preinjury, 2 days post-injury, 7 days post-injury, and every week thereafter for 10 weeks total. Heatmap, right: The heatmap rows depict the distribution of module similarities with module numbers clustered into a dendrogram using a hierarchical clustering approach with Euclidean distances. The columns are ordered by module usages exhibited by each group and time point using the same clustering approach. Prominent clusters or “usage hotspots” are observed in the heatmap's top left corner, housing modules signifying paralysis, encompassing the Tx group and early time points of the Mod and Mild groups; the middle of the heatmap's lower section, embodying modules reflective of recovery or adaptive behaviors, representing mid to late time points of the "Mod" and "Mild" groups; and lastly, the middle of the heatmap's right side accommodates clusters of animals exhibiting normal behavior, including the Sham, Naive, and the preinjury time point. **(C)** *F* univariate statistical test denoting relevant modules for discerning differences between recovery epochs. Each epoch is distinguished by its distinctive module usage patterns and marked transitions towards modules considered "normal." Results illustrate that behavioral patterns observed two days after Sham injuries align with later stages of recovery from SCI, implying that laminectomies may not serve as true "sham" controls. **(D)** Linear discriminant analysis (LDA) of behavioral biomarkers. Modules linked with paralysis are positioned in the upper left quadrant of the LDA space and modules associated with normal behavior occupy the upper right quadrant. Serving as the intermediary, the recovery trajectory between these two quadrants includes an overlap of behavior exhibited by the moderate and mild injury groups. The LDA space mirrors the “usage hotspots” in the heatmap in B.

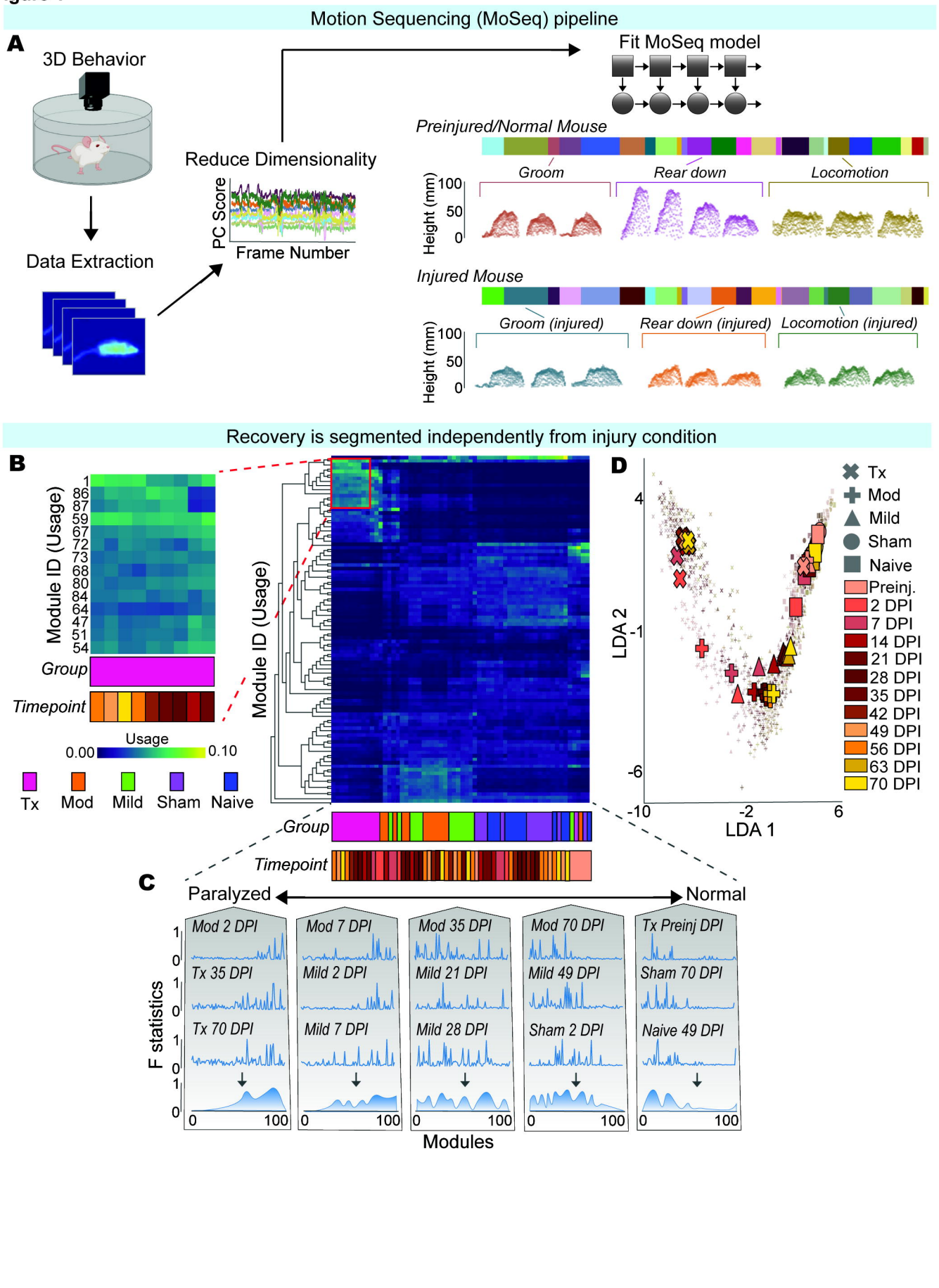
Supplemental Figure 1. Study design and model validation. (A) There are five groups in the study: Transection (Tx), Moderate contusion (Mod), Mild contusion, Sham injury, and Naive animals. **(B)** Experimental Timeline. Recordings occurred on the following time points: preinjury, 2 days post-surgery, 7 days post-surgery, and every week thereafter for a total of 10 weeks. Asterisks indicate recording days for MoSeq and BMS. **(C-E)** Linear Discriminant Analysis (LDA) Classification Confusion Matrix (C), Validation Curve (D), and Permutation Score (E). The model was able to appropriately discriminate between groups and time points. **(F)** LDA dimensionality reduction. Subplots indicate different groups within the LDA space.

Figure 2. Behavioral biomarkers correlate to traditional recovery outcome measures. (A) Pearson’s R Correlations of BMS criteria and lesion severity (left heatmap) vs. module usages (right heatmap). BMS: We observed strongly negative (blue) correlations between BMS criteria indicative of normal locomotion or advanced recovery stages, including Plantar Stepping and Coordination and modules denoting paralysis (top left usage hotspot on heatmap). We further observed a significant positive (red) correlation between BMS criteria signifying regular locomotion or advanced recovery phases and modules reflecting uninjured movements (bottom right usage hotspot on right heatmap). In the center of the bottom of the left heatmap, in columns representing the Mod (orange) or Mild (green) groups, we observed a positive correlation between Plantar Stepping (Right, Left) and BMS scores (Left, Right, Avg) and

modules illustrating adaptive behavior during recovery. This correlation was comparatively less positive, slightly negative, or absent in the case of BMS criteria Coordination, Tail Position, Paw Position, and Trunk Stability. Lesion: We observed a strong positive correlation between larger MBP (demyelination) and GFAP (astrocytic inflammation) volumes and paralysis modules. Increased inflammation (MBP volume) had a positive correlation to modules that are associated with paralysis and recovery. Larger lesions (GFAP volume) demonstrated a strong positive correlation to paralysis modules. Groups include Transection (Tx, pink), Moderate contusion (Mod, orange), Mild contusion (green), Sham injury (purple), and Naive animals (blue). time points are denoted by colors. **(B)** Injury severity in five groups with demyelination (MBP) volume in green and lesion volume (GFAP) in red. **(C)** Pearson's R Correlations of BMS criteria vs. entropy. Entropy denotes the capacity to demonstrate diverse behaviors. We observed a positive correlation between BMS criteria linked to normal locomotion or recovery and higher entropy. **(DE)** Correlations of lesion severity vs. entropy. We observed a strong correlation between less severe injuries (D, GFAP: slope = -0.08719611; y-intercept = 0.62797916), as well as reduced inflammation (E, MBP: slope = -0.09328262; y-intercept = 0.84207945) and higher entropy.

Supplemental Figure 2. BMS Scores. (A) All animals follow a typical recovery curve: Transections (Tx, purple), Moderate contusions (Mod, red), Mild contusions (green), Sham injuries (orange), and Naïve animals (blue).

Figure 1


Motion Sequencing (MoSeq) pipeline
A
3D Behavior
Data Extraction
Reduce Dimensionality
PC Score
Frame Number
Fit MoSeq model
Preinjured/Normal Mouse
Groom
Rear down
Locomotion
Height (mm)
Injured Mouse
Groom (injured)
Rear down (injured)
Locomotion (injured)
Recovery is segmented independently from injury condition
B
Module ID (Usage)
Group
Timepoint
Usage
0.00
0.10
Tx
Mod
Mild
Sham
Naive
D
LDA 1
LDA 2
Preinj.
2 DPI
7 DPI
14 DPI
21 DPI
28 DPI
35 DPI
42 DPI
49 DPI
56 DPI
63 DPI
70 DPI
C
Paralyzed
Normal
Mod 2 DPI
Tx 35 DPI
Tx 70 DPI
Mod 7 DPI
Mild 2 DPI
Mild 7 DPI
Mod 35 DPI
Mild 21 DPI
Mild 28 DPI
Mod 70 DPI
Mild 49 DPI
Sham 2 DPI
Tx Preinj DPI
Sham 70 DPI
Naive 49 DPI
F statistics
Modules

Supp. Figure 1

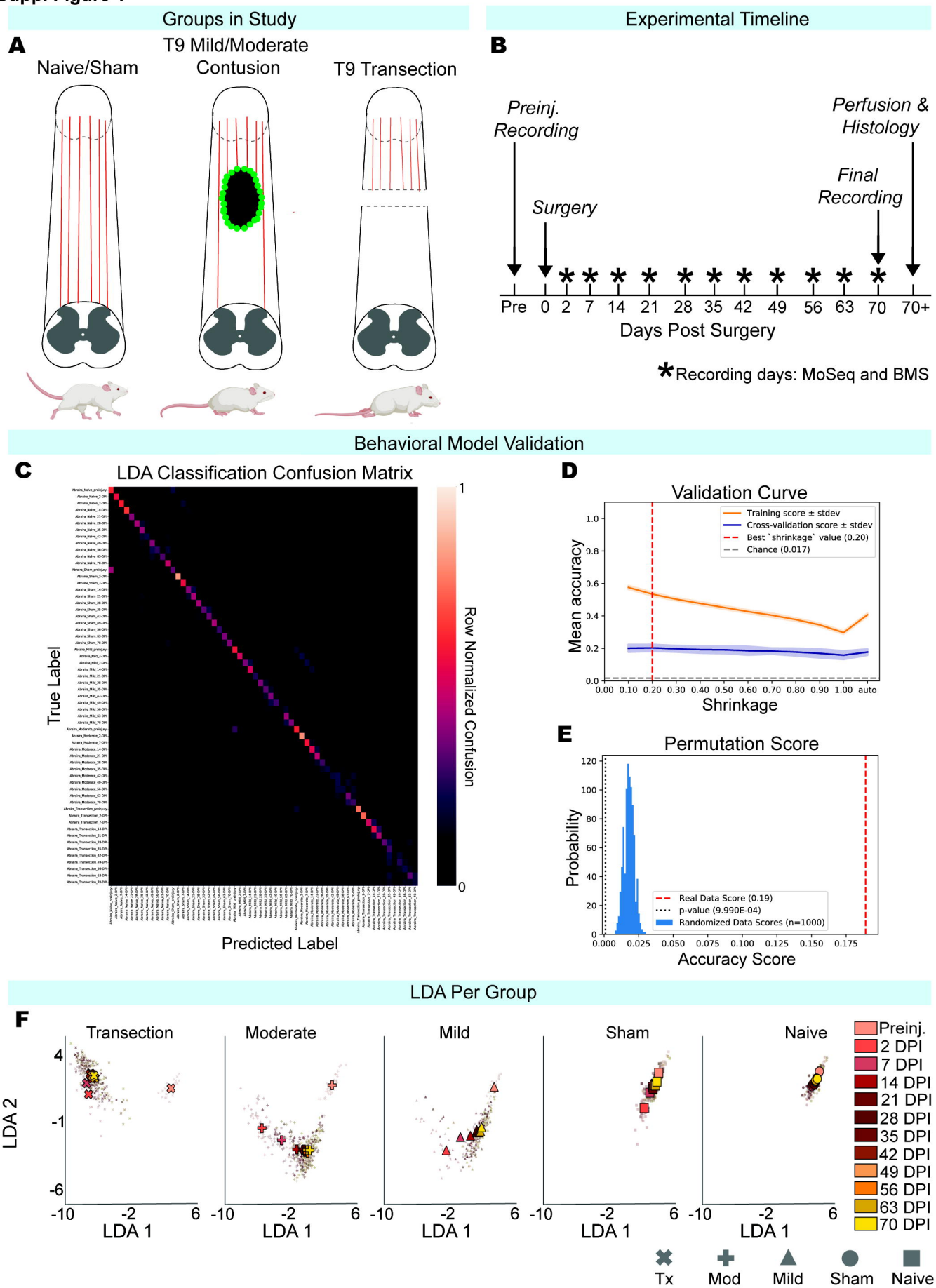
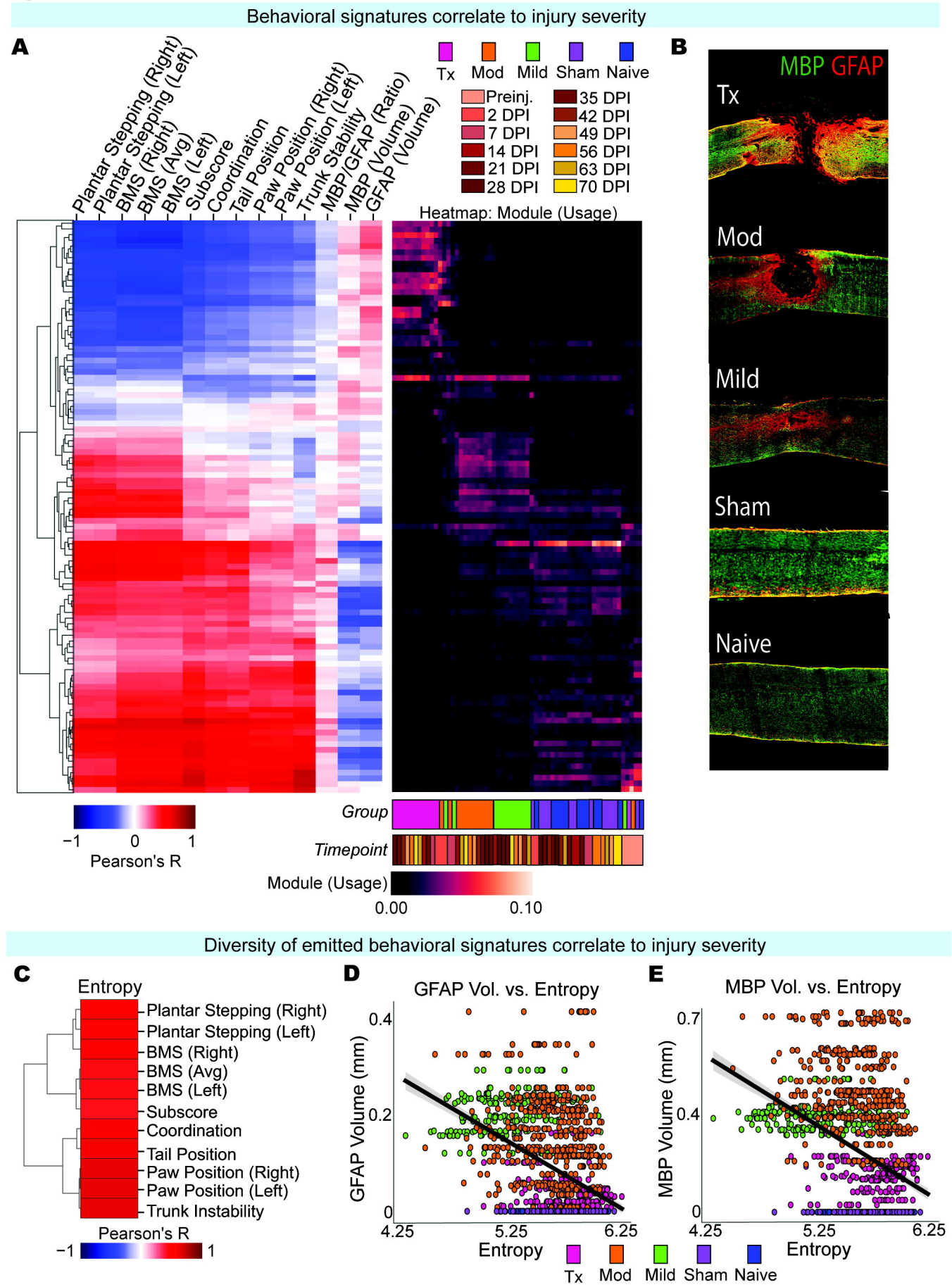


Figure 2


Behavioral signatures correlate to injury severity
A
B
Plantar Stepping (Right)
Plantar Stepping (Left)
BMS (Right)
BMS (Avg)
BMS (Left)
Subscore
Coordination
Tail Position
Paw Position (Right)
Paw Position (Left)
Trunk Stability
MBP/GFAP (Ratio)
MBP (Volume)
GFAP (Volume)
Tx
Mod
Mild
Sham
Naive
Preinj.
2 DPI
7 DPI
14 DPI
21 DPI
28 DPI
35 DPI
42 DPI
49 DPI
56 DPI
63 DPI
70 DPI
Heatmap: Module (Usage)
MBP GFAP
Tx
Mod
Mild
Sham
Naive
−1
0
1
Pearson's R
Group
Timepoint
Module (Usage)
0.00
0.10
Diversity of emitted behavioral signatures correlate to injury severity
C
D
E
Entropy
Plantar Stepping (Right)
Plantar Stepping (Left)
BMS (Right)
BMS (Avg)
BMS (Left)
Subscore
Coordination
Tail Position
Paw Position (Right)
Paw Position (Left)
Trunk Instability
Pearson's R
−1
1
GFAP Vol. vs. Entropy
GFAP Volume (mm)
0.4
0.2
0
4.25
5.25
6.25
Entropy
MBP Vol. vs. Entropy
MBP Volume (mm)
0.7
0.4
0
4.25
5.25
6.25
Entropy
Tx
Mod
Mild
Sham
Naive

Supp. Figure 2