





A deep learning-based toolbox for Automated Limb Motion Analysis (ALMA) in murine models of neurological disorders

Almir Aljovic^{1,2,3,5}, Shuqing Zhao ^{1,2,3,5}, Maryam Chahin^{1,2,3}, Clara de la Rosa^{1,2,3},
Valerie Van Steenbergen ^{1,2}, Martin Kerschensteiner ^{1,2,4} & Florence M. Bareyre ^{1,2,4}✉

In neuroscience research, the refined analysis of rodent locomotion is complex and cumbersome, and access to the technique is limited because of the necessity for expensive equipment. In this study, we implemented a new deep learning-based open-source toolbox for Automated Limb Motion Analysis (ALMA) that requires only basic behavioral equipment and an inexpensive camera. The ALMA toolbox enables the consistent and comprehensive analyses of locomotor kinematics and paw placement and can be applied to neurological conditions affecting the brain and spinal cord. We demonstrated that the ALMA toolbox can (1) robustly track the evolution of locomotor deficits after spinal cord injury, (2) sensitively detect locomotor abnormalities after traumatic brain injury, and (3) correctly predict disease onset in a multiple sclerosis model. We, therefore, established a broadly applicable automated and standardized approach that requires minimal financial and time commitments to facilitate the comprehensive analysis of locomotion in rodent disease models.

¹Institute of Clinical Neuroimmunology, University Hospital, LMU Munich, 81377 Munich, Germany. ²Biomedical Center Munich (BMC), Faculty of Medicine, LMU Munich, 82152 Planegg-Martinsried, Germany. ³Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universitaet Munich, 82152 Planegg-Martinsried, Germany. ⁴Munich Cluster of Systems Neurology (SyNergy), 81377 Munich, Germany. ⁵These authors contributed equally: Almir Aljovic, Shuqing Zhao. ✉email: florence.bareyre@med.uni-muenchen.de

Research into neurological conditions often attempts to uncover how the structural and functional deficits of individual neurons and circuits relate to behavioral outcomes¹. Over the years, a number of elaborate tools based on chemogenetics, optogenetics, and connectomics have been designed to manipulate and record the structure and function of individual circuits and circuit elements^{2–6}; nevertheless, a full understanding of the consequences of such interventions can only be achieved using refined behavioral analysis.

To this extent, a wide range of behavioral tests have been developed that can detect specific aspects of behavior in a range of neurological conditions^{7–9}. Although such tests have substantially improved our ability to selectively monitor deficits, for example in motor function, the evaluation tools are still hampered by a number of limitations. These include a lack of automatization, which makes such tests both highly time-consuming or susceptible to observer bias. For example, the use of the open field¹⁰ to assess locomotor behavior in rodents requires two independent investigators trained to distinguish between the subtle paw and tail positions that could be prone to interpretation. Another limitation is often the high costs of recording and analysis equipment¹¹, which often restricts access to a few specialized labs. Such limitations also affect the evaluation of locomotion that plays a central role in disease-related neuroscientific research, as many common neurological conditions caused by trauma, ischemia, or inflammation prominently affect walking abilities.

Unraveling the complexity of locomotion is best approached via the generation of gait parameters based on the precise position of the limbs. Since Hildebrand's early description of the notion of gait¹², several tests have been designed to evaluate limb motion and paw placement. Early work relied on collecting footprints^{13–15} from an animal after inking the paws, but the data that could be retrieved were often incomplete due to rapid drying of the ink, and analysis was cumbersome to perform. Later systems, including the use of the commercial automated Catwalk[®] system, which uses foot placement to derive a range of gait parameters that reflect locomotion on a static ramp, offered important improvements^{16,17}. The Catwalk[®] system has proven quite useful in the analysis of gait parameters in rodents subjected to genetic manipulation or injuries^{18,19}. However, this system also has some limitations, as the tracking of light-furred rodents can be difficult, and potential mistakes in the tracking of footprints need to be manually corrected by the experimenter. In addition, the system is static and does not allow the locomotion speed to be controlled. The gait can also be recorded at variable speeds while the animal is running on a treadmill using three-dimensional video recordings coupled with a kinematic tracing system¹⁵ or using the well-established motion-capture system VICON with eight infrared cameras^{20–22}. Both require reflective markers to be manually and bilaterally attached at key joints (for example, iliac crest, lateral malleolus, the tip of the toe, etc.), and use of the latter system is further limited by the cost of the acquisition²³. Therefore, recent new advances have been made to improve and facilitate locomotion tracking. A key step has been the development of the DeepLabCut (DLC) method that provides a markerless approach to labeling and tracking moving joints (DeepLabCut²⁴) and, thereby, facilitates the kinematic analysis of gait as well as arm movements in animals and humans^{25–28}. However, even this approach requires substantial specialized expertise and processing times to translate the limb coordinates into kinematic profiles and parameters that quantify the distinct aspects of locomotion.

Here, we describe how we overcame this final challenge by developing an open-source computational “toolbox”, Automated Limb Motion Analysis (ALMA), which provides a fully automated and comprehensive analysis of locomotion and fine paw placement in mice with minimal costs, time requirements, and previous expertise. This toolbox which is based upon pose

estimation obtained from deep lab cut (DLC) and includes a graphical user interface (GUI) with functionalities for automated kinematic parameter computation and automated footfall detection, kinematic data analysis with random forest classification and principal component analysis, and visualization of the gait kinematics. To make it amenable to as many users as possible, we used only inexpensive equipment (custom-made ladders and a commercial treadmill) and a single standard high-speed action camera. We applied this toolbox to mouse models of common traumatic and inflammatory diseases of the brain and spinal cord to demonstrate its capability to robustly and sensitively monitor the evolution of locomotor deficits in a broad range of neurological conditions. Notably, our results show that such an automated comprehensive analysis can delineate the specific parameters of the locomotor function that are best suited to track injuries of the brain or spinal cord or are sensitive enough to predict disease onset during the prodromal phase of a multiple sclerosis model.

Results

Pipeline for automated limb motion analysis: DLC markerless labeling, model training, and automated kinematic analysis and footfall detection in mice. In this paper, we present a new open-source toolbox for the analysis of gait kinematic parameters of locomotion and fine paw placement in mice. To do so, we focused on two behavioral tasks that were applied to mice: walking on a treadmill (Fig. 1a) to determine gait kinematic parameters and walking along ladders (Fig. 1b) with regularly or irregularly spaced rungs to analyze fine paw placement. All tasks were filmed with a GoPro 8 camera positioned parallel to and at a fixed distance and angle from the treadmill and ladder. We used DLC to perform markerless limb tracking of six hindlimb joints (toe, metatarsophalangeal joint (MTP), ankle, knee, hip, and iliac crest) for treadmill kinematic analysis, and of limb extremities for the ladder rung paradigm. (Fig. 1a, b). For both analyses, we focused only on the hindlimb trajectories.

We then trained the neural network model for body part coordinate extraction. For each behavioral paradigm, 2D body part coordinates were obtained using a deep residual network model (ResNet-50), which was trained and refined using DLC (Fig. 1a, b). By choosing the most representative frames through k-means clustering, we were able to obtain a finetuned deep residual network model that can reliably detect body part coordinates using only up to 450 manually labeled frames, as described in the methods. We then performed automated kinematic analysis on the coordinate data obtained from the treadmill task and extracted a selection of 44 kinematic parameters that best reflected the locomotion of the hindlimbs according to the literature^{20,29}, which could also easily be expanded depending on the study goal. We grouped those parameters into five main categories: joint angles, spatial variability, the temporal feature of gait, limb endpoint trajectories, and dragging (Fig. 1a and Supplementary Table 1). These parameters could then be classified and compared using random forest and principal component analyses (PCA) to extract the most important parameters (Fig. 1a). Our toolbox also allows via a scree plot the visualization of the percentage of explained variance for all components of the PCA. For the ladder rung analysis, we performed automated detection of hindlimb footfalls (Fig. 1b), which were manually validated. Three parameters were then extracted, the number, depth, and duration of footfalls (Fig. 1b).

Tracking locomotion after spinal cord injury with the ALMA toolbox. We first tested the applicability of our toolbox for analyzing overground locomotion following spinal cord injury (SCI) (Fig. 2a). To do so, mice walking on a treadmill were

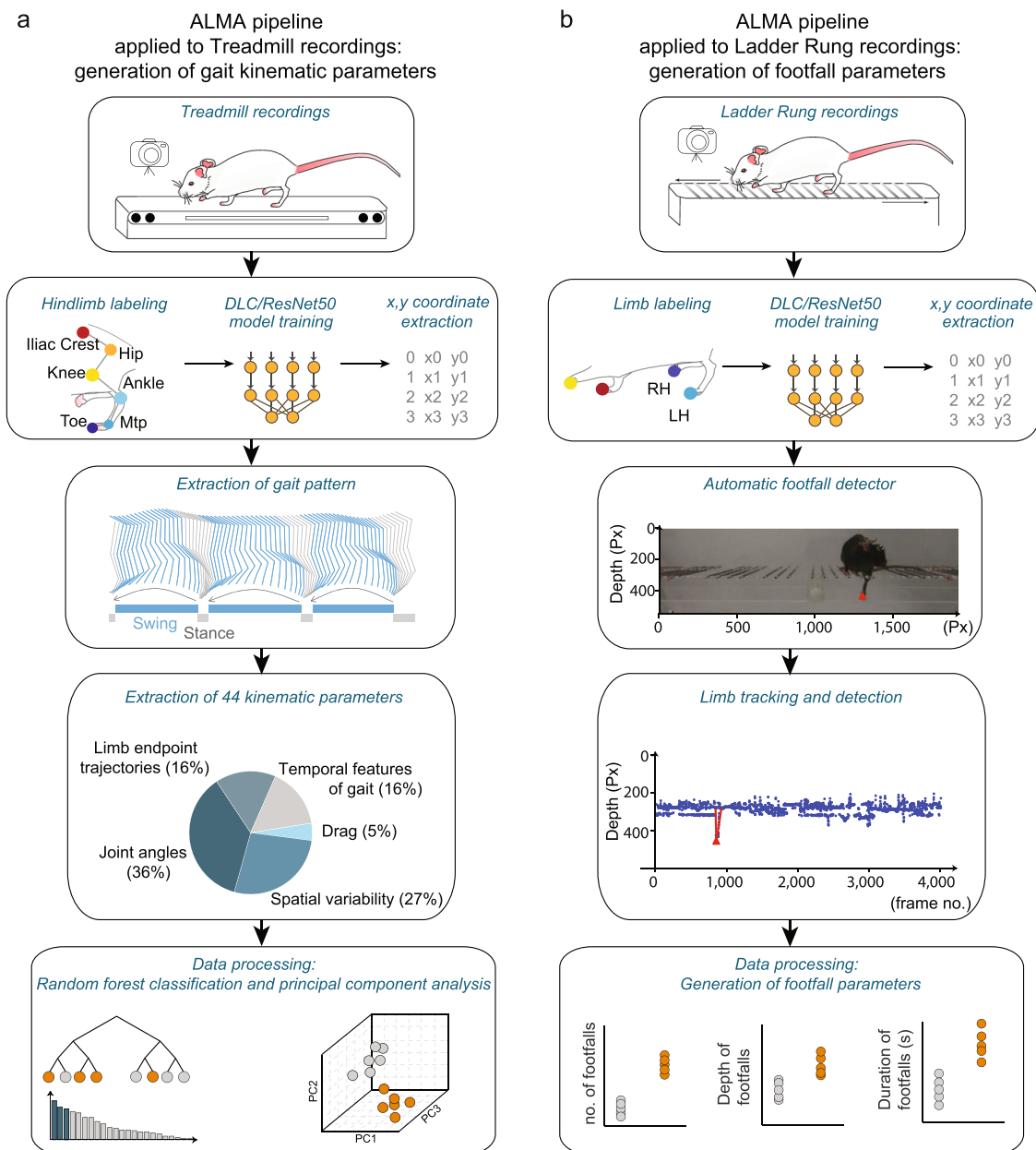
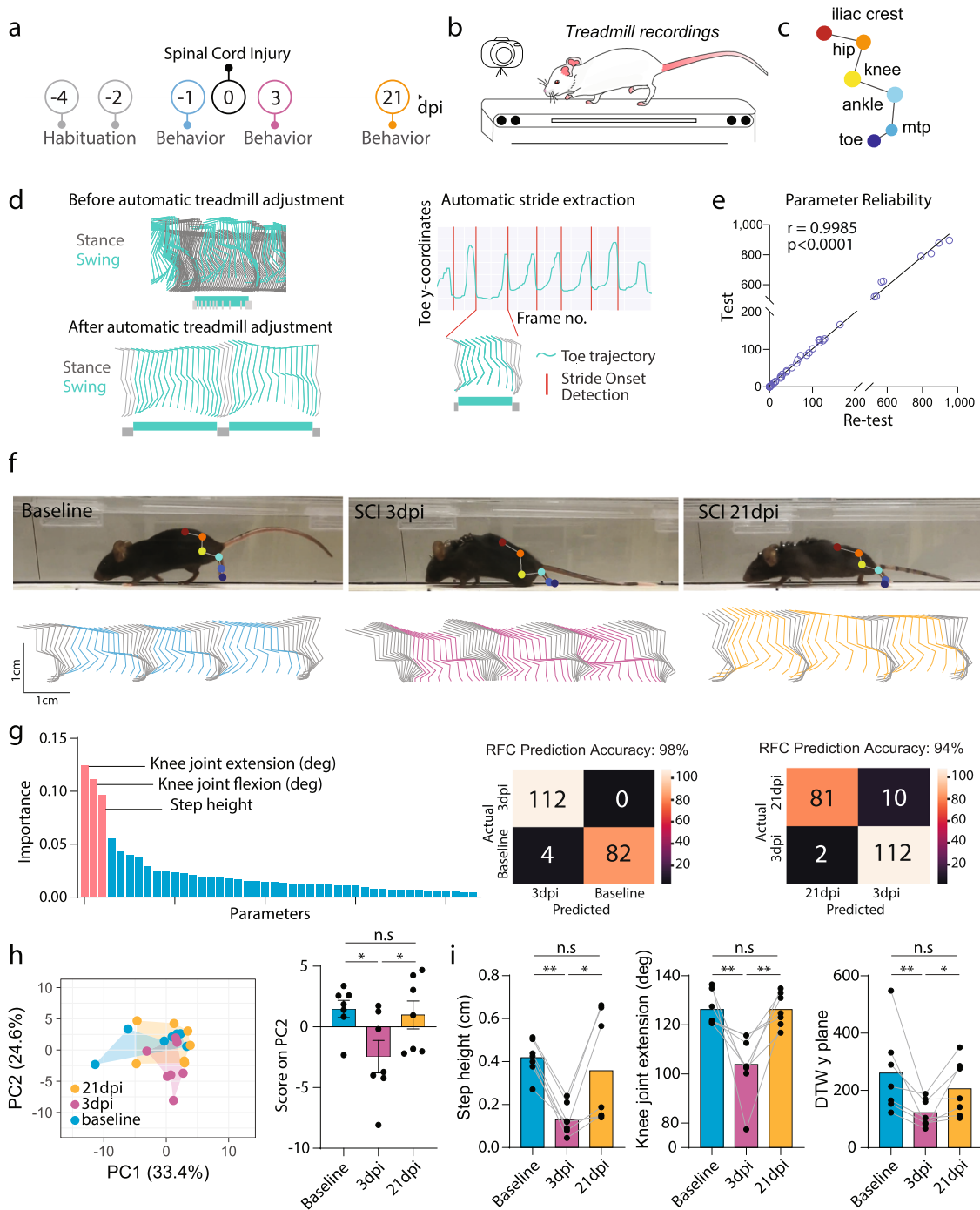


Fig. 1 Automated behavioral analysis using ALMA toolbox. **a** Schematic of ALMA toolbox application to treadmill recordings for the generation of gait kinematic parameters. First, treadmill videos were recorded with a GoPro 8 camera placed parallel to the treadmill. Then, markerless limb labeling and modeling was trained and refined using ResNet-50 and DeepLabCut (DLC) and coordinates were extracted. Coordinates were tracked for six hindlimb joints (toe, metatarsophalangeal (MTP) joint, ankle, knee, hip, and iliac crest) in the treadmill kinematic paradigm, and the coordinates were processed using the toolbox to generate hindlimb trajectories. This allowed the generation of 44 kinematic parameters that represented joint angles, spatial variability, limb endpoint trajectories, temporal features of gait, and dragging. Data processing was conducted in ALMA to obtain principal component analysis and random forest classification of the parameters. **b** Schematic of ALMA toolbox application to the ladder rung recordings for the generation of footfall parameters. First, ladder rung videos were recorded with a GoPro 8 camera placed parallel to the ladder. Then, markerless limb labeling and modeling was trained and refined using ResNet-50 and DLC, and coordinates were extracted. Coordinates could be tracked for all four paws. The toolbox used the automated footfall detector to extract limb tracing and footfall detection. The on- and off-set of locomotor errors (footfalls) were estimated using signal processing methods and subjected to manual validation. Three parameters, the number, depth, and duration of the footfalls, were extracted. Px pixels.

recorded with a GoPro 8 camera (Fig. 2b), and we used DLC to perform markerless labeling of the six hindlimb joints as described above (Fig. 2c and Supplementary Video 1). The ALMA toolbox then automatically detected the treadmill speed and adjusted the body part coordinates accordingly. The adjusted body part coordinate data were then used to extract stride onsets and offsets, as well as to distinguish the stance and swing phases (Fig. 2d and Supplementary Fig. 1). This procedure enabled

automated kinematic parameter extraction for each stride (Fig. 2d). We validated the reliability of the extracted data with baseline recordings to check the reproducibility of the 44 parameters measured. To do so, we tested and re-tested mice on consecutive days on the treadmill task at baseline (e.g., before injury), and we found that reliability was excellent for all parameters measured (Fig. 2e; $r = 0.9985$, $P < 0.0001$, Pearson's correlation) between the initial test and the re-test conditions.



For the SCI experiment, we used mice that were first tested in the Basso Mouse Scale (BMS)¹⁰ and showed at least one point of recovery over time (BMS score baseline 9 ± 0 ; BMS score 3 days post-injury (dpi) 3.86 ± 1.08 and BMS score 21 dpi 6.29 ± 0.81). We then analyzed the treadmill recordings collected before spinal cord injury and at 3 and 21 days following SCI in these mice and generated the adjusted hindlimb trajectories (Fig. 2f). Those trajectories were profoundly altered in SCI mice at 3 days post-injury (dpi) but had partially recovered by 21 dpi (Fig. 2f). We applied random forest classification to the data for the animals based on the 44 extracted parameters, and we could predict the injury status to a 98% accuracy when we compared baseline to animals at 3 dpi, and the recovery status to a 94% accuracy when we compared 3 dpi to 21 dpi animals. The step height, knee joint flexion, and knee joint extension showed the highest Gini

impurity-based feature importance (Fig. 2g). We then compared the gait of the mice before and after spinal cord injury using principal component analysis of our 44 kinematic parameters and determined that the data for injured mice at 3 dpi clustered separately from those of the mice at baseline, while mice at 21 dpi were clustered in between, again indicating incomplete recovery of locomotor function (Fig. 2h). PC1 and PC2 together represented 58% of the variance, with PC2 better reflecting the between-group differences (Fig. 2h). After calculating the factor loadings, we identified three parameters that best-represented PC2 and were, thus, most likely to track SCI-related differences (Fig. 2h). Step height, knee joint extension, and dynamic time warping (DTW) demonstrated significant alterations at 3 dpi, and recovered over time until they were no longer significantly different from baseline (Fig. 2h and Supplementary Table 2).

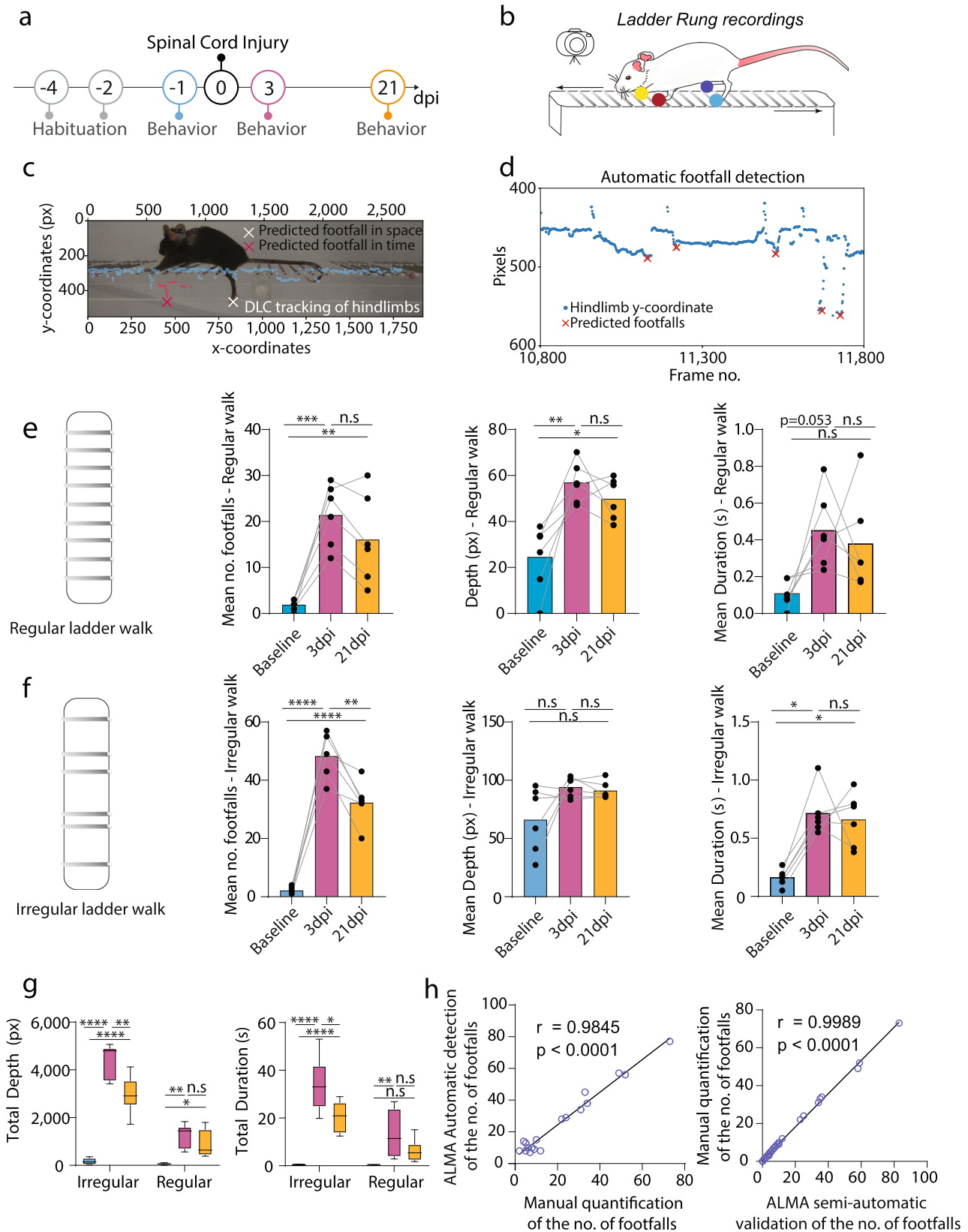
Fig. 2 ALMA analysis of gait changes in spinal cord injured mice tested on the treadmill. **a** Timeline of the traumatic spinal cord injury (SCI) experiment. **b** Schematic of the treadmill system used to record the behavior of mice during the SCI experiment. **c** Schematic of DeepLabCut (DLC) markerless joint labeling. Six joints were labeled: iliac crest, hip, knee, ankle, metatarsophalangeal joint (MTP), and toe. **d** Representation of hindlimb trajectories (left panel) before the adjustment in ALMA (top; green, swing; gray, stance) and after adjustment with ALMA (bottom; green, swing; gray, stance). Note, that after adjustment, swing and stance were efficiently separated. Representation of automatic stride extraction from the toe coordinates and frame number (right panel). **e** Quantification of parameter reliability; baseline data were tested and re-tested and demonstrated a high correlation coefficient ($r = 0.9985$, $P < 0.0001$; Pearson's correlation coefficient). **f** Photographic images (top) of mice running on the treadmill showing the markerless labeling of hindlimb joints using DLC at baseline, 3 dpi, and 21 dpi, and hindlimb trajectories for baseline (cyan), 3 dpi (fuchsia), and 21 dpi (orange). **g** Random forest classification (RFC) of 44 parameters extracted from the ALMA toolbox for the analysis of gait following spinal cord injury and accuracy injury status prediction based on the 44 parameters using confusion matrices for 3 dpi vs. 21 dpi (Gini impurity-based feature importance for RFC: knee joint extension, 0.125; knee joint flexion, 0.112; step height, 0.097. RFC prediction accuracy: baseline vs. 3 dpi 98% and 3 dpi vs. 21 dpi 94%, tested in $n = 84$ –92 step cycles). **h** Principal component analysis of data obtained on the treadmill and processed with the ALMA toolbox for spinal cord injury, and plot of scores of PC2 that represent 22.3% of the variability (principal component analysis, PC1 36.1%, PC2 22.3%, repeated-measures one-way ANOVA followed by Tukey's test; baseline vs. 3 dpi [$P = 0.022$]; baseline vs. 21 dpi [$P = 0.920$], 3 dpi vs. 21 dpi [$P = 0.044$]; $n = 7$). **i** Quantitative evaluation of parameters associated with PC2, such as step height, knee joint extension, or dynamic time warping (DTW) y plane, at baseline, 3 dpi, and 21 dpi. Repeated-measures one-way ANOVA followed by Tukey's test was used to analyze knee joint extension (baseline vs. 3 dpi, $P = 0.005$; baseline vs. 21 dpi, $P > 0.999$; 3 dpi vs. 21 dpi, $P = 0.005$; $n = 7$), Friedman and Dunn tests were used for DTW y plane (baseline vs. 3 dpi, $P = 0.004$; baseline vs. 21 dpi, $P > 0.999$; 3 dpi vs. 21 dpi, $P = 0.049$; $n = 7$), Repeated-measures one-way ANOVA followed by Tukey's test was used to analyze step height (baseline vs. 3 dpi, $P = 0.063$; baseline vs. 21 dpi, $P = 0.012$; 3 dpi vs. 21 dpi, $P > 0.999$; $n = 6$). In all panels, data are presented as mean \pm SEM; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Px pixels.

Our analysis thus suggests that these kinematic parameters are best suited to monitor locomotor deficits resulting from an incomplete spinal cord injury. In addition, we validated a subset of kinematic parameters by manually labeling the different joints and the onset of step cycle detection, and found no significant differences with the parameters obtained through the automatic method (Supplementary Fig. 2). Importantly, those changes are specific of SCI as our toolbox could not pick up any gait kinematic changes following sham injury (Supplementary Fig. 3).

Tracking skilled paw placement after spinal cord injury with the ALMA toolbox. While overground locomotion strongly depends on the function of intraspinal circuits, skilled paw placement requires supraspinal input and is, thus, commonly used to assess the regeneration and remodeling of descending motor tracks, including corticospinal projections. We, therefore, assessed whether the ALMA toolbox could also be used to monitor skilled paw placement following spinal cord injury (Fig. 3a). To this end, we recorded mice walking on a horizontal ladder with regularly or irregularly spaced rungs. Similar to the overground locomotion experiment, all videos were recorded using a single GoPro 8 camera, and we used DLC to perform markerless labeling of the hind paws (Fig. 3b and Supplementary Video 2). We then applied ALMA to determine footfall characteristics with a peak detection algorithm (Fig. 3c, d and Supplementary Fig. 4). The parameters extracted from ALMA showed that spinal cord injury caused an increase in the mean number and depth of footfalls at 3 dpi, while the mean duration of the footfalls was only slightly increased, for the ladder with regularly spaced rungs (Fig. 3e). Those parameters remained altered over a prolonged period so that they remained different from baseline at 21 dpi (Fig. 3e). Similarly, in the ladder with irregular rungs experiment, the mean number of footfalls and duration of footfalls were found to increase at 3 dpi and remained elevated till 21 dpi. However, the mean number of footfalls significantly recovered from 3 to 21 dpi. The mean depth of footfalls was, in this case, unchanged throughout the study period (Fig. 3f). We also calculated the total duration and total depth of the footfalls on the ladders with regularly and irregularly spaced rungs and observed a significant increase in the total depth and total duration at 3 dpi and 21 dpi (Fig. 3g). Finally, we compared the ALMA automated detection of the number of footfalls using a deviation algorithm with (i) the semi-automatic detection (that includes a manual correction provided by the GUI) and (ii) a fully manual count of the

mistakes. In both cases, we found a highly significant correlation (Fig. 3h) confirming that the ALMA toolbox can provide precise quantification of skilled paw placement. While there is a slight overestimation through the fully automated detection (see “Methods”), the fully manual quantification and the semi-automated quantifications with ALMA are similar (Fig. 3h).

The ALMA toolbox can reveal behavioral consequences of traumatic brain injury. Next, we wanted to determine whether our toolbox, which was initially designed to reveal motor abnormalities related to spinal cord pathology, was also capable of detecting the locomotor changes induced by neurological conditions primarily affecting the brain. This is of particular importance for traumatic brain injuries, as it has often been challenging to detect and reliably quantify motor deficits arising from such insults, in particular in the acute injury phase^{30,31}. We, therefore, induced moderate brain injury in mice using a traumatic brain injury (TBI) impactor and recorded the locomotion and paw placements of mice walking on the treadmill and ladders, respectively, before and 1 day and 10 days after traumatic brain injury (Fig. 4a). We applied the ALMA toolbox as described above and initially analyzed the treadmill recordings (Fig. 4b, c). The findings showed that the limb trajectories were strongly affected at 1 dpi and had partially recovered by 10 dpi (Fig. 4d). We applied random forest classification to all animals at baseline and 1 dpi based on the 44 extracted parameters, and we were able to demonstrate the parameters with the highest Gini impurity-based feature importance (step height, hip joint, hip joint amplitude, and hip joint flexion) and predict the injury status with 83% accuracy. This indicates that our toolbox makes it possible to robustly detect even the subtle locomotor changes induced by a moderate TBI (Fig. 4e). We then used principal component analysis to reduce the dimensionality of our data and determine the individual movement parameters that best identified differences between the groups. We found that data for mice at baseline clustered closer to the data for the mice at 10 dpi than to those at 1 dpi, indicating that the injury and recovery effects can be ascertained based on the kinematic parameters (Fig. 4f). As PC1 represented almost 40% of the variance, we plotted the principal component analysis scores and demonstrated the above-mentioned effects at 1 dpi (Fig. 4f). Based on PC1, we identified the top three parameters that contributed to the differences: step height, stride length, and DTW distance. These parameters showed significant



changes at 1 dpi and later recovery at 10 dpi, indicating that our toolbox can also detect changes in locomotion following brain lesions (Fig. 4g and Supplementary Table 3).

We then analyzed the ladder task recordings using the ALMA toolbox (Fig. 4h). We extracted the three footfall parameters described previously and demonstrated that mice with brain

injuries displayed an increased number of footfalls in the regular-rung ladder task, which returned to baseline at 10 dpi (Fig. 4i). In contrast, the irregular-rung ladder performance was unaffected by moderate traumatic brain injury. Taken together, these results indicate that the ALMA toolbox can be applied to assess neurological conditions of the brain and is capable of sensitively

Fig. 3 ALMA analysis of fine paw placement in spinal cord injured mice in the ladder rung test. **a** Timeline of the traumatic spinal cord injury experiment. **b** Schematic of the ladder rung system used to record the fine paw placement of mice during the spinal cord injury experiment indicating the DeepLabCut (DLC) markerless paw labeling (yellow, red, dark blue and light blue dots). **c** Photographic image of a mouse running on the treadmill showing the automatic detection of footfall, as predicted in time and space by the toolbox. **d** Automated detection algorithm used to predict footfalls in space and time. **e** Quantitative parameters extracted from ALMA for the regular walk showing the mean number, mean depth, and mean duration of footfalls for all time points (cyan, baseline; purple, 3 dpi; and orange, 21 dpi). Repeated one-way ANOVA followed by Tukey's test was used to analyze the regular ladder rung results (mean no. footfalls, baseline vs. 3 dpi [$P = 0.0002$], baseline vs. 21 dpi [$P = 0.021$], 3 dpi vs. 21 dpi [$P = 0.223$]; mean depth, baseline vs. 3 dpi [$P = 0.003$], baseline vs. 21 dpi [$P = 0.013$], 3 dpi vs. 21 dpi [$P = 0.612$]; and mean duration, baseline vs. 3 dpi [$p = 0.053$], baseline vs. 21 dpi [$P = 0.131$], 3 dpi vs. 21 dpi [$P = 0.838$]; $n = 6$). **f** Quantitative parameters extracted from ALMA for the regular walk showing the mean number, mean depth, and mean duration of footfalls for all time points (cyan, baseline; purple, 3 dpi; and orange, 21 dpi). Repeated one-way ANOVA followed by Tukey's test was used to analyze the irregular ladder rung results (mean no. footfalls, baseline vs. 3 dpi: [$P < 0.0001$], baseline vs. 21 dpi [$P < 0.0001$], 3 dpi vs. 21 dpi [$P = 0.002$]; mean depth, baseline vs. 3 dpi [$P = 0.745$], baseline vs. 21 dpi [$P > 0.999$], 3 dpi vs. 21 dpi [$P > 0.999$]; and mean duration, baseline vs. 3 dpi, $P = [0.028]$, baseline vs. 21 dpi [$P = 0.028$], 3 dpi vs. 21 dpi [$P > 0.999$]; $n = 6$). **g** Quantitative evaluation of the total depth and total duration of footfalls for all time points (cyan, baseline; purple, 3 dpi; and orange, 21 dpi). Repeated one-way ANOVA followed by Tukey's test was used to analyze total footfall depth on regular ladder rungs (baseline vs. 3 dpi, $P = 0.0047$; baseline vs. 21 dpi, $P = 0.043$; and 3 dpi vs. 21 dpi, $P = 0.175$; $n = 6$), total depth on irregular ladder rungs (baseline vs. 3 dpi, $P < 0.0001$; baseline vs. 21 dpi, $P < 0.0001$; 3 dpi vs. 21 dpi, $P = 0.0039$; $n = 6$), total duration on regular ladder rungs (baseline vs. 3 dpi, $P = 0.0042$; baseline vs. 21 dpi, $P = 0.173$; 3 dpi vs. 21 dpi, $P = 0.101$; $n = 6$), and total duration on irregular ladder rungs (baseline vs. 3 dpi, $P < 0.0001$; baseline vs. 21 dpi, $P = 0.004$; 3 dpi vs. 21 dpi, $P = 0.041$; $n = 6$). **h** Correlation between the ALMA automatic detection of the number of footfalls using the deviation algorithm and a fully manual detection (left panel; $r = 0.9845$, $P < 0.0001$; Pearson's correlation coefficient) and between the ALMA semi-automatic detection of the number of footfalls using the deviation algorithm and a fully manual detection (right panel; $r = 0.9989$, $P < 0.0001$; Pearson's correlation coefficient). In all panels, data are presented as mean \pm SEM; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Px pixels, dpi days post-injury.

tracking both alterations in locomotor kinematics and foot placement after TBI.

The ALMA toolbox can monitor disease symptoms and predict disease onset in a multiple sclerosis model. Finally, we sought to investigate the applicability of our toolbox for assessing neurological conditions with a less predictable disease course, such as those caused by CNS inflammation. Therefore, we induced a commonly used mouse model of multiple sclerosis, experimental autoimmune encephalomyelitis (EAE), by immunizing mice with the myelin oligodendrocyte glycoprotein (MOG). We then recorded the locomotion of the mice on the treadmill at baseline, disease onset (defined as the first day with clinical symptoms), disease peak (3 or 4 days after onset), and recovery (10 or 11 days after onset; Fig. 5a, b). As expected, as the disease progressed, the mice developed ascending paresis and paralysis that strongly altered their hindlimb trajectories at the peak of the disease (and precluded the use of the ladder tests). While the hindlimb trajectories recovered slightly over the next few days, they remained clearly distinct from the baseline pattern (Fig. 5c). The principal component analysis confirmed that the data formed distinct clusters at the onset, peak, and recovery of disease compared to baseline (Fig. 5d). As PC1 represented 76.1% of the variability, we plotted the scores in PC1 to recapitulate the evolution of motor systems (Fig. 5e) and extracted the key parameters clustering with PC1. Those parameters (stride length, stride height, toe-crest distance, and dragging percentage) all showed significant changes at the peak of the disease and tended towards a later recovery, indicating that our toolbox is well suited to monitoring the locomotor alterations resulting from the formation and resolution of inflammatory lesions (Fig. 5f and Supplementary Table 4).

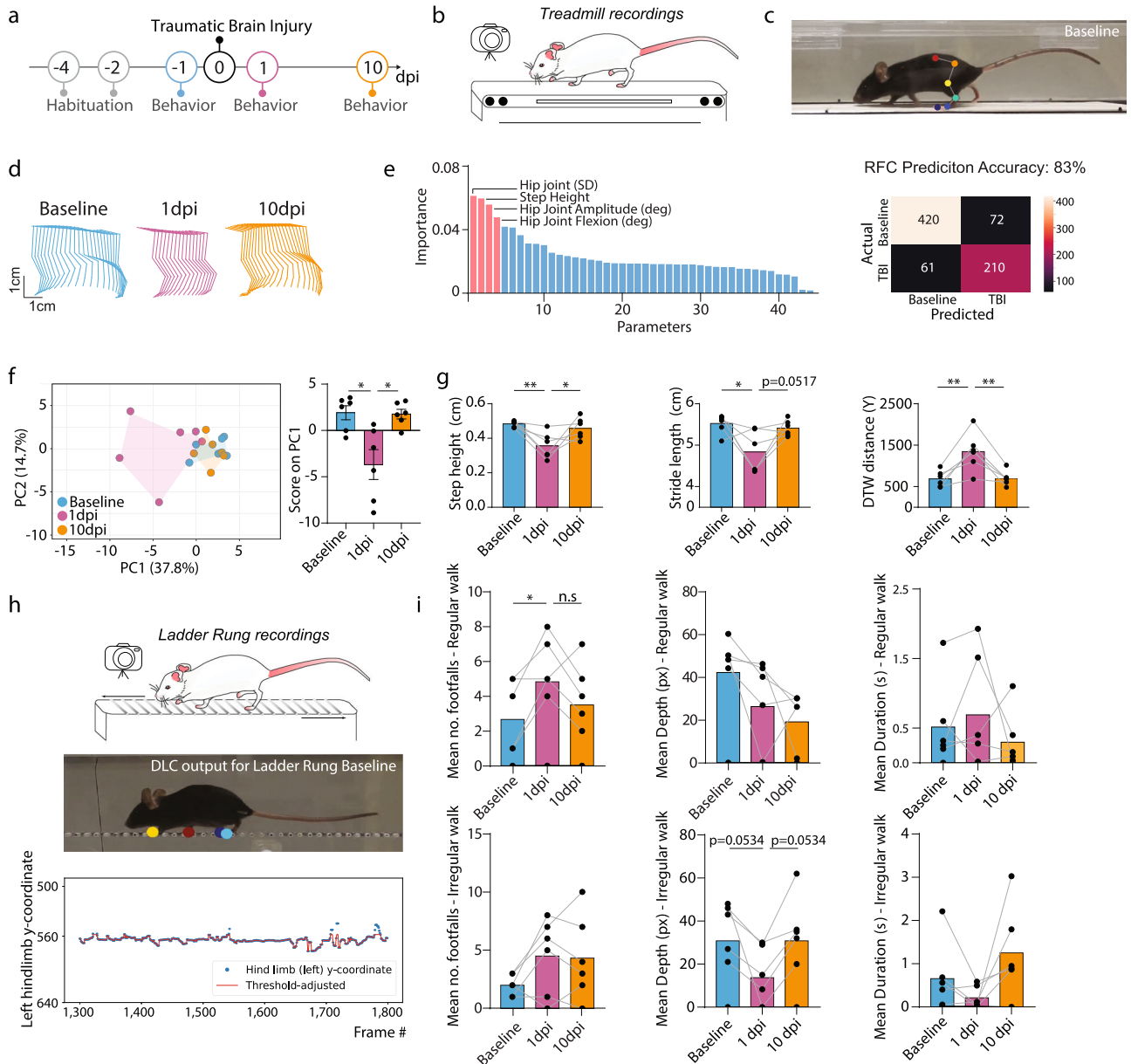
We next examined whether this refined analysis of locomotion kinematics is sufficiently sensitive to pick up the prodromal stages of the disease, which are not apparent in classical EAE scoring (based on simple observations of mouse mobility). In this context, we were particularly interested in whether we could detect the more subtle locomotor alterations associated with the initial formation of the lesions, which can predict the subsequent onset of overt disease symptoms (defined as the first day of detectable symptoms in the EAE score; often seen as tail paralysis). This would be useful, as it is often important to precisely initiate interventions, e.g., pharmacological treatment, when CNS lesions

first start to form, as, once initiated, the CNS inflammatory process can self-perpetuate. Therefore, we used the ALMA toolbox to analyze the treadmill recordings starting 3 days prior to disease onset (defined by the EAE score). Using random forest classification, we found we could predict whether and when the mice would subsequently show EAE symptoms with 75% accuracy 3 days before onset, with 78% accuracy 2 days before onset, and with 86% accuracy 1 day before onset (Fig. 5g). This illustrated the ability of our toolbox to accurately predict whether an immunized mouse will develop overt disease symptoms in the prodromal disease phase and even predict the day of onset. This renders the initial formation phase of CNS lesions amenable to study.

Discussion

The quantitative assessment of locomotor performance is critical in the analysis of physiological gait patterns and their perturbations in neurological diseases. Conventional evaluation strategies often rely on resource-intensive and time-demanding observation and analysis setups. Together with the specialized expertise required to operate these setups, these demands currently prevent the broad application of refined locomotor analyses in basic and clinical neuroscience. In this study, we used deep-learning strategies to implement ALMA, an open-source toolbox that facilitates refined analysis of overground locomotion and skilled paw placement at a fraction of the cost of specialized behavioral setups. ALMA is fully automated, saving time and preventing observer bias, and it can be used without previous expertise through its user-friendly interface. Therefore, ALMA makes a comprehensive analysis of locomotion accessible to every research group interested in revealing the behavioral consequences of nervous system dysfunction and disease.

The use of machine-learning approaches to the study of behavior in both rodents and humans has dramatically increased in the last 5 years^{24–27,32}. An important step in this process was the development, by Mathis and colleagues, of a markerless method (DLC²⁴) to label the joints and follow their motion reliably over time. Here, we made use of this method to mark important joints in mice and import the DLC coordinates to track locomotion and fine paw placements. Alternatively, our toolbox can also be efficiently used with coordinates generated through other techniques, such as VICON, the newly published DANNCE²⁸ methods or any other



excellent pose estimation platforms such as SLEAP (sleap.ai)³³, MARS³⁴, and DeepPoseKit³⁵.

To translate these limb and joint coordinates into locomotor patterns on a treadmill or horizontal ladder and to subsequently extract the parameters of gait and footfall, we used model training. In healthy mice, this allowed us to define a list of 44 hindlimb parameters— validated with almost perfect re-test reliability ($r = 0.9985$)—that best represent gait features but can also be increased or made more comprehensive to meet one’s needs. It is remarkable that this highly robust kinematic analysis can not only be achieved with a fully automated analysis pipeline but also that it can be achieved with the use of an affordable single-action camera. Particularly as, to date, scientists have had to rely on the use of high-end cameras and software modules to manually extract gait cycles and parameters in the analysis of kinematic behavior. Previous techniques came with several pitfalls, including high software costs, limited troubleshooting support, and cumbersome and time-consuming post-experimental processing. In contrast, the ALMA toolbox simply requires the import of DLC output using a GUI and allows the extraction of the final gait kinematic parameters. Another important aspect of our

approach is that the ALMA toolbox allows the tracking of locomotor patterns based on side-view camera recordings, which, while harder to model, provide much more gait parameters compared with the bottom-view recordings used in the catwalk approach or in recent automated approaches^{19,27}. As demonstrated here, 2D side-view tracking and analysis provides important information on each joint angle, step height, and body support, which is missing in top- or bottom-view recordings of gait analysis.

In addition to kinematic analysis, the ALMA toolbox offers the first automated analysis of the number, duration, and depth of footfalls on regularly and irregularly spaced ladder rungs. This is an important improvement for those investigating skilled paw placement, as it replaces a cumbersome analysis protocol that had to be performed frame by frame^{36,37}. The analysis of skilled paw placement is complementary to the obtention of kinematic parameters, as it provides a readout of the damage and recovery of supraspinal circuits^{18,38}. Finally, it should be noted that while we focused our analysis on both the gait and footfall aspects of mouse hindlimb movements, the ALMA toolbox should (with minor modifications) be equally suitable for the analysis of

Fig. 4 ALMA monitoring of gait changes and differences in fine paw placement in brain-injured mice. **a** Timeline of the traumatic brain injury experiment. **b** Schematic of the treadmill system used to record the behavior of the mice during the traumatic brain injury experiment. **c** Photographic images of the mice running on the treadmill showing markerless labeling of hindlimb joints using DeepLabCut (DLC) at baseline. **d** Hindlimb trajectories for baseline (top, cyan), 1 dpi (middle, purple), and 10 dpi (bottom, orange). **e** Random forest classification (left) of the 44 parameters extracted from the ALMA toolbox for the analysis of gait following traumatic brain injury, and confusion matrix (right) for determining prediction accuracy of the injury status based on the 44 parameters (Gini impurity-based feature importance for RFC: hip joint, 0.061; step height, 0.059; hip joint amplitude, 0.056; hip joint flexion, 0.047; RFC prediction accuracy: baseline vs. 1 dpi 83%; tested in $n = 282$ –481 step cycles). **f** Principal component analysis of data obtained from the treadmill task and processed with the ALMA toolbox, and plot of PC1 scores that represent 37.8% of the variability and associated factor loadings (principal component analysis, PC1 37.8%, PC2 14.7%; repeated one-way ANOVA followed by Tukey's test, baseline vs. 1 dpi [$P = 0.012$], baseline vs. 10 dpi [$P = 0.665$], 1 dpi vs. 10 dpi [$P = 0.014$]; $n = 6$). **g** Quantitative evaluation of factors associated with PC1, i.e., step height, stride length, and dynamic time warping (DTW) distance, at baseline, 1 dpi, and 10 dpi. Repeated one-way ANOVA followed by Tukey's test was used to analyze step height (baseline vs. 1 dpi, $P = 0.0095$; baseline vs. 10 dpi, $P = 0.730$; 1 dpi vs. 10 dpi, $P = 0.033$; $n = 6$), stride length (baseline vs. 1 dpi, $P = 0.0214$; baseline vs. 10 dpi, $P = 0.855$; 1 dpi vs. 10 dpi, $P = 0.0517$; $n = 6$), and DTW distance (baseline vs. 1 dpi, $P = 0.001$; baseline vs. 10 dpi, $P = 0.9943$; 1 dpi vs. 10 dpi, $P = 0.0012$; $n = 6$). **h** Schematic of the ladder rung system used to record the behavior of mice during the traumatic brain injury experiment, and photographic images of a mouse running on the treadmill showing markerless labeling of hindlimb paws using DLC at baseline and showing the algorithm detection of footfall. **i** Quantitative evaluation of three parameters extracted from ALMA for footfalls at baseline, 1 dpi, and 10 dpi. Friedman followed by Dunn's test was used to analyze the regular ladder rung mean no. footfalls (baseline vs. 1 dpi, $P = 0.0315$, baseline vs. 10 dpi, $P = 0.2557$; 1 dpi vs. 10 dpi, $P = 0.1583$), mean depth (baseline vs. 1 dpi, $P = 0.1299$; baseline vs. 10 dpi, $P = 0.0628$; 1 dpi vs. 10 dpi, $P > 9999$), and mean duration (baseline vs. 1 dpi, $P > 0.9999$; baseline vs. 10 dpi, $P > 0.9999$; 1 dpi vs. 10 dpi, $P > 9999$; $n = 6$) and the irregular ladder rung mean no. footfalls (baseline vs. 1 dpi, $P = 0.3371$; baseline vs. 10 dpi, $P = 0.9370$; 1 dpi vs. 10 dpi, $P > 0.9999$; $n = 6$), mean depth (baseline vs. 1 dpi, $P = 0.0534$, baseline vs. 10 dpi, $P > 0.9999$; 1 dpi vs. 10 dpi, $P = 0.0534$; $n = 6$) and mean duration (baseline vs. 1 dpi, $P > 0.9999$; baseline vs. 10 dpi, $P = 0.9370$; 1 dpi vs. 10 dpi, $P = 0.3371$; $n = 6$). In all panels, data are presented as mean \pm SEM; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Px pixels, dpi days post-injury.

forelimb function, forelimb–hindlimb coordination, or the tracking of locomotor function in larger rodents such as rats.

We further showed that the ALMA toolbox can be applied to a range of neurological conditions and enables the automated tracking of locomotor deficits in mice with traumatic and inflammatory injuries of the brain and spinal cord. Using spinal cord injury models, we confirmed that tracking of locomotor deficits is robust, as demonstrated by the high reproducibility and the correct prediction of injury status. As spinal cord injuries lead to pronounced locomotor deficits, a large number of studies have used behavioral testing strategies to assess the functional outcomes of genetic or therapeutic modulations^{20,21,39–45}. Importantly, using ALMA, we find profound disturbances in gait and fine paw placement at acute time points following incomplete spinal cord injury and demonstrate a recovery at more chronic time points. This is similar to previous reports that use similar spinal cord injury models^{18,20,29,38,46,47}. The application of the ALMA toolbox in such models should both improve the reliability of the analysis (as observer and selection biases are removed) and provide a more refined view of locomotor disturbance and recovery, as it allows the operator to precisely determine which component or components of locomotor function are regulated by a specific circuit modulation or therapeutic intervention.

The locomotor deficits observed after spinal cord injury are pronounced compared with the less obvious deficits resulting from mild-to-moderate brain injuries. It has often been challenging to detect and reliably monitor locomotor deficits following such brain injuries in mice, particularly in the early phase when the effects are subtle^{30,31}. Our findings showed that the kinematic tracking of gait parameters, in particular, step height, stride length, and DTW (which measures the similarity between two step sequences), allowed us to sensitively reveal both the emergence of locomotor deficits (1 day after injury) and their recovery (by 10 days after injury). Interestingly, while pronounced abnormalities of locomotion were observed at the initial recording after injury, the animals later recovered a movement pattern almost identical to the pattern observed at baseline. This is clearly distinct from the comparably incomplete recovery process observed after spinal cord injury and may indicate that the persistence of intraspinal circuits is critical for re-establishing the

“original” physiological movement pattern, while the reorganization of intraspinal (but not supraspinal) circuits results in a compensatory adaption of the movement pattern.

In comparison to studies of the injured brain and spinal cord, only a few attempts have been made to apply refined behavioral testing to models of inflammatory CNS damage, such as the EAE model^{48–50}. One limitation has been, that, due to the disseminated nature of inflammatory infiltration, which results in variable neuronal tracking systems being affected in different mice, inter-individual differences in the pattern of locomotor changes are to be expected. For this purpose, targeted EAE models have been developed that allow inflammatory lesions to be directed to a predetermined anatomical location in the brain or spinal cord^{51–53}. However, the stereotactic injection procedure required for the induction of such models limits their application, e.g., for studying disease initiation. Here we were able to show that the comprehensive kinematic analysis generated by the ALMA toolbox allows for the refined monitoring of locomotor deficits, even in classical disseminated EAE models. While such locomotor symptoms can also be tracked by the classical EAE scoring scale, which is based on visual inspection of the walking abilities of mice, kinematic analysis provides a number of advantages. First, the ALMA toolbox allows for fully automated and standardized analyses, removing observer bias and, thereby, making behavioral assessments more comparable between different observers and labs. Second, it stands to reason that the quantitative assessment of 44 distinct gait parameters should be more sensitive to differences in locomotor function compared with visual inspection. This is particularly important when evaluating the behavioral consequences of therapeutic manipulations targeting neuronal protection and repair⁵⁴ that are expected to lead to more subtle changes in locomotor function. Third, and related to the latter, our current analysis showed that the ALMA toolbox is able to accurately predict the onset of overt motor symptoms up to 3 days before the onset of disease can be detected by conventional EAE scoring. This indicates that even subtle motor symptoms that arise during this stage⁵⁵, which are presumably related to the formation of the first CNS lesions, can be detected by the ALMA toolbox. Our toolbox thus facilitates studies of the prodromal stage of the disease and its targeted modulation by therapeutic interventions.

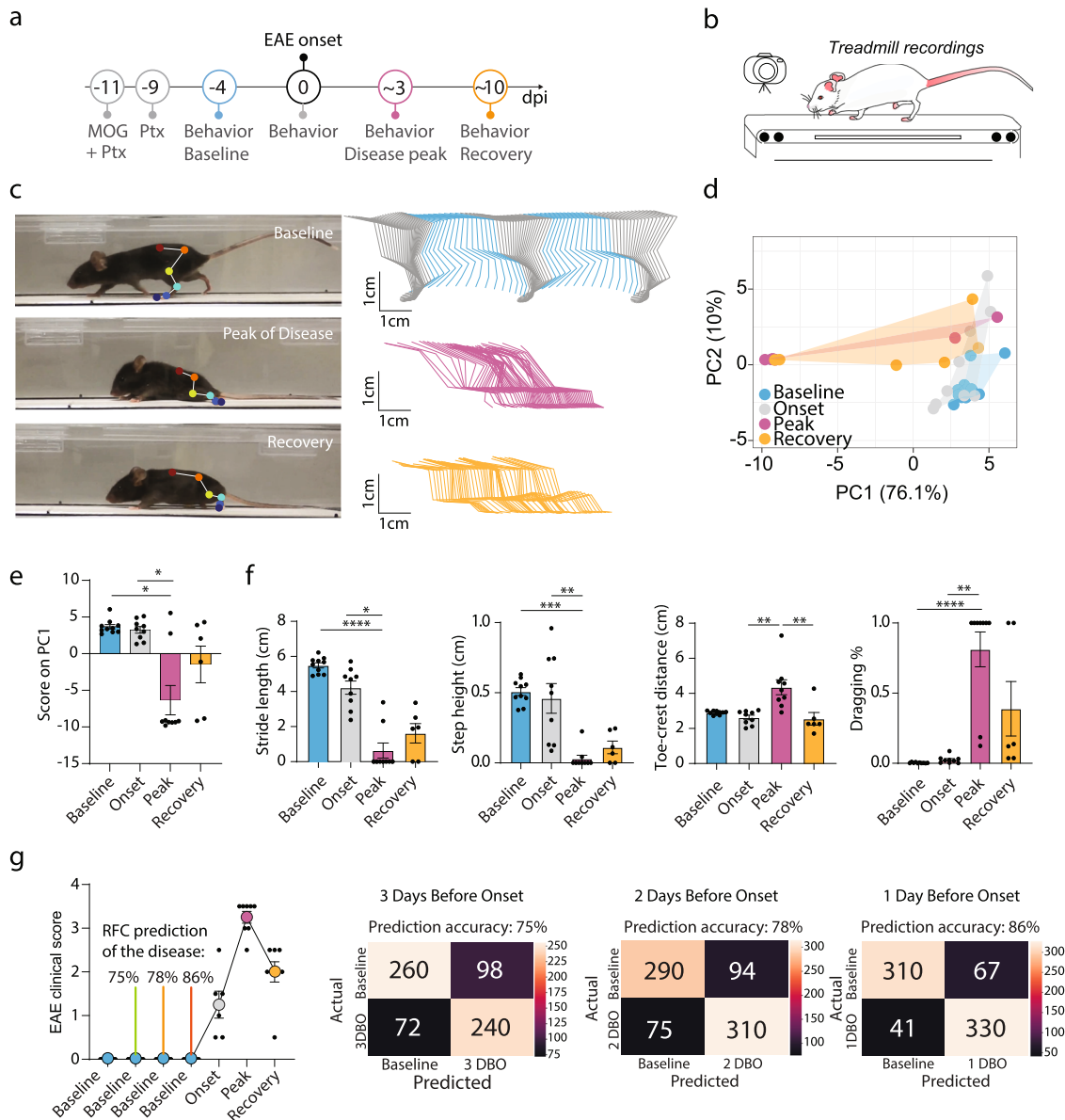


Fig. 5 ALMA monitoring of locomotor changes in mice that developed experimental autoimmune encephalomyelitis and accurate prediction of disease development during the prodromal phase. **a** Timeline of the EAE experiment. **b** Schematic of treadmill system used to record the behavior of the mice during the EAE experiment. **c** Photographic images of mice running on the treadmill (left) showing markerless labeling of hindlimb joints using DeepLabCut (DLC) at baseline (top), onset of disease (middle), and disease recovery (bottom), and hindlimb trajectories (right) for baseline (top), onset of disease (middle), and disease recovery (bottom). **d** Principal component analysis of data obtained on the treadmill and processed with the ALMA toolbox. **e** Plot of the PC1 scores that represent 76.1% of the variability and associated factor loadings (Kruskal-Wallis followed by Dunn’s test; baseline vs. peak, $P = 0.0104$; onset vs. peak, $P = 0.0418$; baseline vs. recovery, $P = 0.9806$; peak vs. recovery, $P > 0.9999$). **f** Quantitative evaluation of factors associated with PC1, i.e., stride length, step height, toe-crest distance, and dragging, at baseline and different stages of EAE (Kruskal-Wallis followed by Dunn’s test; stride length, baseline vs. peak [$P < 0.0001$], onset vs. peak [$P = 0.0231$], baseline vs. recovery [$P = 0.0019$], peak vs. recovery [$P > 0.9999$]; $n = 6$; step height, baseline vs. peak [$P = 0.0002$], onset vs. peak [$P = 0.0020$], baseline vs. recovery, [$P = 0.0382$], peak vs. recovery [$P > 0.9999$]; toe-crest distance, baseline vs. peak [$P = 0.0737$], onset vs. peak [$P = 0.0036$], baseline vs. recovery [$P = 0.0382$], peak vs. recovery [$P = 0.7374$]; and dragging (%), baseline vs. peak [$P < 0.0001$], onset vs. peak [$P = 0.0082$], baseline vs. recovery [$P = 0.0039$], peak vs. recovery [$P > 0.9999$]; $n = 6$). **g** EAE clinical score and prediction of the disease onset based on random forest classification in the prodromal phase (3, 2, and 1 days before onset) using ALMA. In all panels, data are presented as mean \pm SEM; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Px pixels, dpi days post-injury.

While we have outlined the real advantages of using ALMA to analyze gait and fine paw placements in the context of CNS damage, ALMA also comes with inherent limitations. For example, gait analysis was, in this paper, performed only on hindlimbs. This can be easily circumvented by further using pose estimation to label the forelimb joints which would allow the generation of gait parameters for the forelimb and the generation

of coupling data between forelimbs and hindlimbs. Likewise, ALMA currently provides the investigators with the two main components of the PCA following analysis. Using a scree plot (also provided in the toolbox), more components can be visualized and studied. For the determination of mistakes in the ladder rung test, we used a “deviation peak detection” algorithm that slightly overestimates the raw number of mistakes. This can be

circumvented, as we describe in the paper, by manual validation of the number of mistakes within the GUI in our semi-automatic detection. Alternatively, other algorithms (also implemented in the toolbox) could be used such as the “threshold” or “baseline correction” algorithms but pose additional constraints on recording set-ups.

Taken together, we provide a user-friendly open-source toolbox that requires minimal time and resource commitments; is applicable to a wide range of neurological conditions affecting the brain and spinal cord; and provides an unbiased, robust, and comprehensive assessment of locomotion.

Methods

ALMA toolbox. This research aimed to provide a toolbox for the analysis of gait and footfall in mouse models of neurological disorders. This toolbox includes a graphical user interface (GUI) with functionalities for (i) automated kinematic parameter computation, (ii) automated footfall detection, (iii) data analysis of the computed kinematic parameters with random forest classification and principal component analysis, and (iv) visualization of gait kinematics. The ALMA toolbox is an open-source Python repository for the automatic processing of DLC coordinates, gait cycle detection, and kinematic parameter extraction. Parameters extracted from ALMA include joint angles, limb endpoint trajectories, drag, temporal features of gait, and spatial variability. This last category represents a range of pairwise dynamic time warping (DTW) parameters that measure similarity between limb endpoint trajectories despite different duration and speed³². Following kinematic parameter extraction, the ALMA toolbox enables the use of machine-learning algorithms, such as random forest classification and principal component analysis, to reduce dimensionality and identify the most relevant kinematic parameters in the scope of health and disease. In addition to automated kinematic analysis, the toolbox GUI provides a code for the accurate detection of footfalls during fine motor tasks (e.g., traversing ladder rungs) for manual validation of each footfall. Details and the open-source toolbox can be found at <https://github.com/sollan/alma>.

Feature labeling and model training. To train the DLC model, we used ~450 image frames from different disease models (spinal cord injury, traumatic brain injury, and EAE) and time points. To predict hindlimb kinematic positions, we manually labeled six different body parts (toe, MTP joint, ankle, knee, hip, and iliac crest) in all ~450 image frames. The model was trained for up to 650,000 iterations using a deep residual network structure (ResNet-50), based on the pretrained model weights from DLC. To detect footfalls in the ladder rung test, we manually labeled all four paws on ~200 image frames for different mouse models (see above) and time points. The ResNet-50 model was trained for 400,000 iterations, based on the pretrained model weights from DLC. To train our models, we used a computer with 64 GB RAM, AMD Ryzen 9 3900 × 12-Core Processor × 24, and GeForce RTX 2080 Ti 11 GB graphics card. Our trained models for kinematic and footfall analyses are publicly available.

To perform random forest (RF) classification following gait analysis, we used scikit-learn⁵⁶. To build the classifiers, ALMA provides several built-in functions such as outlier removal, assignment of group labels to the single gait cycle kinematic data, and concatenation of the kinematic data from different groups into a single data frame. For the random forest hyperparameters, we followed the default from the random forest classifier function in sklearn, which was sufficient for learning the mapping⁵⁷. Specifically, we used 100 decision trees in each RF classifier, with Gini impurity as the splitting criterion. We did not set a maximum depth for the trees and allowed an unlimited number of leaf nodes. The individual trees were built with samples from the entire training dataset. Importantly, we created new classifiers for each assay (i.e., one RF classifier for healthy vs. 3 dpi, another for healthy vs. 21 dpi, and so on). We used a 75%/25% split of training and testing data. As there was roughly an equal amount of data from the healthy/disease conditions, we based the splitting on random selection. The accuracy results presented in this paper were based on test data not included in the training set to demonstrate the generalizability of the models.

Animals. We used C57bl6 female mice of 2–4 months of age. The mice were kept in our animal facility under a regular day/night cycle (12 h/12 h). Animals had constant access to food and water. All animal experiments were carried out in accordance with the German animal welfare guidelines and previously authorized by the local regulatory committees (Regierung von Oberbayern).

Spinal cord injury. Mice were anesthetized with MMF (medetomidin 0.5 mg/kg, Orion Pharma; midazolam 5.0 mg/kg, Ratiopharm; fentanyl 0.05 mg/kg, B. Braun). Once the mice presented no reflex reaction from paw pinching, their backs were shaved and a laminectomy was performed at T8 levels. The dura was exposed, and a dorsal hemisection was performed using irridectomy scissors^{38,58} (Bradley et al.⁵⁸; Loy et al.⁵⁹). After the hemisection, the wound was closed, and the skin was

sutured. An antagonist mix was given (atipamezole 2.5 mg/kg, flumazenil 0.5 mg/kg, and naloxon 1.2 mg/kg), and mice were kept on a heating pad until completely awake. Mice received meloxicam (Metacam, 1.5 mg/ml oral suspension) at 12, 24, and 48 h following the injury.

EAE. Active EAE was induced in female mice by subcutaneous injection of 400 µg of purified recombinant MOG (N1-125) in complete Freund’s adjuvant (freshly made by adding 10 mg/ml *Mycobacterium tuberculosis* H37 Ra, Sigma-Aldrich). Then, 400 ng of pertussis toxin (Sigma-Aldrich) were administered intraperitoneally (i.p.) on days 0 and 2 after immunization⁶⁰. The mice were weighed daily and scored for neurological deficits according to the following EAE scores: 0, no clinical signs; 0.5, partial tail weakness; 1, tail paralysis; 1.5, gait instability or impaired righting ability; 2, hindlimb paresis; 2.5, hindlimb paresis with dragging of one foot; 3, total hindlimb paralysis; 3.5, hindlimb paralysis and forelimb paresis; 4, hindlimb and forelimb paralysis; 5, death.

Stages of the disease were defined as follows: disease onset was defined as the first day with clinical symptoms in the EAE score, disease peak occurred 3 or 4 days after onset, and recovery occurred 10 or 11 days after onset).

Traumatic brain injury. Mice were anesthetized by intraperitoneal MMF injection. Once the mice presented no reflex reaction from paw pinching, they were put on a stereotaxic frame (Precision Systems & Instrumentation, LLC). A skin incision was made followed by the drilling of a window on the right skull hemisphere, which was positioned rostrocaudal, between the bregma and lambda and under the sagittal suture. Mice were then placed on the TBI-0310 impactor (Precision Systems & Instrumentation, LLC) to undergo traumatic brain injury (TBI). The tip of a 3-mm diameter steel rod was used to induce injury to the somatosensory cortex according to the following settings: 6 m/s, 150 ms dwell time, 0.5 mm depth⁶¹. Mice were removed from the impactor, the skull window was repositioned and sealed with Vetbond glue (3 M Vetbond, 3 M United States), and the skin was stitched. Mice were placed on a heating pad and injected with the antagonist mix before receiving a subcutaneous glucose injection (glucose 5% B. Braun Infusionslösung).

Behavior setup for kinematics. Mice were recorded using a GoPro 8 camera at 120 frames per second while running on a treadmill (Harvard Apparatus; speed varying from 2 cm/s to 25 cm/s depending on the disease model; Supplementary Fig. 1). The distance between the treadmill and camera was 14.5 cm, and the camera was placed equidistant from the two ends of the treadmill. We chose this camera position to obtain comprehensive information on joint angles, spatial variability, and limb endpoint trajectories, which are missing from bottom recordings of gait analysis. Each mouse was recorded for at least 1 min to sample enough step cycles. Every incomplete step cycle was automatically excluded by the toolbox. To extract meaningful datasets, a minimum number of 1200 frames were captured, i.e., 10 s of recording. One prerequisite to accurate data extraction was the blind selection of frames with representative step cycles (no pause in animal locomotion, no grooming, no back turns etc.). When the animal’s limbs are completely dragging, our algorithm will only compute parameters that are independent to the step cycles. All preprocessing and processing steps undertaken by the ALMA toolbox can be found in Supplementary Fig. 1. We tested the validity of the automated kinematic analysis methods of ALMA by computing a selection of temporal and spatial kinematic parameters through manual labeling and quantification, based on a subset of the video data recorded at baseline (Supplementary Fig. 2). The manual validation was done using a custom-written Python script using OpenCV (openCV.org). An experienced observer labeled the body parts used for markerless pose estimation (toe, MTP, ankle, knee, hip, iliac crest) in more than 200 frames from a video recording at baseline (120 fps, treadmill speed 30 cm/s). We then calculated the Euclidean distance between the body part coordinates generated by manual labeling and markerless tracking from DLC. We then computed a number of kinematic parameters for up to seven-step cycles based on manually determined step onsets, using both manually labeled coordinates and automatically detected coordinates.

Behavior setup for a ladder rung. Mice were recorded using a GoPro 8 camera at 120 frames per second while making four consecutive runs on our custom-made ladders. In this test, the animals had to cross 1-m horizontal ladders, and footfalls were recorded. We evaluated the rhythmic locomotion in the regular walking task on a ladder with evenly spaced rungs and the animal’s fine coordination paw placement ability using irregularly spaced rungs (irregular walking task). The camera was placed 17.5 cm distant from the ladder equidistant from the two ends. Animals were habituated on a ladder with regularly spaced rungs before any experiment was performed (2/3 habituations each for a total max. time of 3 min). We used DLC to perform markerless labeling of the hind paws and then applied ALMA to determine footfall characteristics with a peak detection algorithm. In the toolbox, we included three preprocessing algorithms that accommodate a range of different recording conditions and can be chosen by the experimenter. Here we chose to use a preprocessing procedure based on the deviation algorithm that shows a great correlation with manual counting. In particular, we used the “deviation” algorithm in automated ladder rung analysis, based on peak detection applied to the raw *y*-coordinate signal output from markerless pose estimation

methods, to determine mistakes. In order to avoid missed footfalls, we set a low prominence (footfall depth) threshold in the peak detection step, leading to a slightly higher number of false positives due to the signal-noise ratio. To reduce the number of false positives while using the “deviation” algorithm, we then adjusted the prominence threshold in the peak detection function. The data generated by ALMA in this paper include a manual validation step using the graphic interface provided in ALMA. We also validated fully manually all datasets obtained using ALMA by manually counting all mistakes made by the animals. To do so, we only analyzed consecutive steps of the hindlimbs. Therefore, the last step before or after any interruptions were not scored. Placements were considered a mistake when mice either totally missed a rung or if they slipped from a rung (deep or slight slip). Placements were considered as correct when the mice correctly placed all the feet or only a portion of the foot on the rungs. Then the number of mistakes over a standard distance was calculated quantitatively. All data preprocessing and analysis steps undertaken by the ALMA toolbox can be found in Supplementary Fig. 3.

Basso mouse scale. All mice with SCI were evaluated preoperatively and at 7 and 21 dpi. The scores regarding the locomotor ability of mice were given according to the original paper¹⁰ by fully trained observers.

Statistics and reproducibility. All results are given as mean \pm standard error of the mean (SEM) in the figures. In the expanded tables, data are given as mean \pm standard deviation (SD). Statistical analysis and the construction of the graphs for data illustration were carried out on GraphPad Prism 8.4.3 for Windows (GraphPad Software). All datasets were tested for normality using the Shapiro–Wilk test. Parametric data from the SCI and TBI experiments were analyzed using repeated-measures ANOVA and Tukey’s post hoc test. Nonparametric datasets from the SCI and TBI experiments were analyzed using the Friedman test followed by the Dunn post hoc test. Datasets from the EAE experiments were analyzed with the Kruskal–Wallis test followed by the Dunn post hoc test, as they were distributed non-normally. To determine correlation, we used Pearson’s correlation coefficient. In order to classify animals we used random forest classification, based on individual step cycles. Feature importance in the Random Forest Classification was determined by Gini impurity-based feature importance which ranged between 0 and 1. Statistical significance levels are indicated as follows: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Reproducibility. All experiments in this study include at least five biological replicates. The number of replicates (ns) is mentioned in the text or figure legend.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data generated and analyzed in this study are included in this published article or are publicly available on Figshare repository⁶².

Code availability

We have provided the code for the ALMA toolbox at <https://github.com/sollan/alma> and Zenodo⁶³.

Received: 26 July 2021; Accepted: 21 January 2022;

Published online: 15 February 2022

References

- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., Maciver, M. A. & Poeppel, D. Perspective neuroscience needs behavior: correcting a reductionist bias. <https://doi.org/10.1016/j.neuron.2016.12.041> (2017).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
- Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potently activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104**, 5163–5168 (2007).
- Wickersham, I. R., Finke, S., Conzelmann, K. K. & Callaway, E. M. Retrograde neuronal tracing with a deletion-mutant rabies virus. *Nat. Methods* **4**, 47–49 (2007).
- Fenno, L., Yizhar, O. & Deisseroth, K. The development and application of optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
- Roth, B. L. DREADDs for neuroscientists. *Neuron* **89**, 683–694 (2016).
- Crawley, J. N. Behavioral phenotyping strategies for mutant mice. *Neuron* **57**, 809–818 (2008).
- Brooks, S. P. & Dunnett, S. B. Tests to assess motor phenotype in mice: a user’s guide. *Nat. Rev. Neurosci.* **10**, 519–529 (2009).
- Fonio, E., Golani, I. & Benjamini, Y. Measuring behavior of animal models: faults and remedies. *Nat. Methods* **9**, 1167–1170 (2012).
- Basso, D. M. et al. Basso mouse scale for locomotion detects differences in recovery after spinal cord injury in five common mouse strains. *J. Neurotrauma* **23**, 635–659 (2006).
- Merriault, P., Dupuis, Y., Boutteau, R., Vasseur, P. & Savatier, X. A study of vicon system positioning performance. *Sensors* **17**, 1591 (2017).
- Hildebrand, M. Vertebrate locomotion: an introduction: how does an animal’s body move itself along? *Bioscience* **39**, 764–765 (1989).
- Metz, G. A. S., Dietz, V., Schwab, M. E. & Van de Meent, H. The effects of unilateral pyramidal tract section on hindlimb motor performance in the rat. *Behav. Brain Res.* **96**, 37–46 (1998).
- Fouad, K., Metz, G. A. S., Merkler, D., Dietz, V. & Schwab, M. E. Treadmill training in incomplete spinal cord injured rats. *Behav. Brain Res.* **115**, 107–113 (2000).
- Ueno, M. & Yamashita, T. Kinematic analyses reveal impaired locomotion following injury of the motor cortex in mice. *Exp. Neurol.* **230**, 280–290 (2011).
- Hamers, F. P. T., Lankhorst, A. J., Van Laar, T. J., Veldhuis, W. B. & Gispen, W. H. Automated quantitative gait analysis during overground locomotion in the rat: Its application to spinal cord contusion and transection injuries. *J. Neurotrauma* **18**, 187–201 (2001).
- Bhimani, A. D. et al. Functional gait analysis in a spinal contusion rat model. *Neurosci. Biobehav. Rev.* **83**, 540–546 (2017).
- Jacobi, A. et al. FGF 22 signaling regulates synapse formation during post-injury remodeling of the spinal cord. *EMBO J.* **34**, 1231–1243 (2015).
- Hamers, F. P. T., Koopmans, G. C. & Joosten, E. A. J. CatWalk-assisted gait analysis in the assessment of spinal cord injury. *J. Neurotrauma* **23**, 537–548 (2006).
- Takeoka, A., Vollenweider, I., Courtine, G. & Arber, S. Muscle spindle feedback directs locomotor recovery and circuit reorganization after spinal cord injury. *Cell* **159**, 1626–1639 (2014).
- Courtine, G. et al. Recovery of supraspinal control of stepping via indirect propriospinal relay connections after spinal cord injury. *Nat. Med.* **14**, 69–74 (2008).
- Huang, R. et al. Machine learning classifies predictive kinematic features in a mouse model of neurodegeneration. *Sci. Rep.* **11**, 1–16 (2021).
- Pfister, A., West, A. M., Bronner, S. & Noah, J. A. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *J. Med. Eng. Technol.* **38**, 274–280 (2014).
- Mathis, A. et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
- Arac, A., Zhao, P., Dobkin, B. H., Carmichael, S. T. & Golshani, P. Deepbehavior: a deep learning toolbox for automated analysis of animal and human behavior imaging data. *Front. Syst. Neurosci.* **13**, 20 (2019).
- Zimmermann, C., Schneider, A., Alyahyay, M., Brox, T. & Diester, I. FreiPose: a deep learning framework for precise animal motion capture in 3D spaces. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.02.27.967620> (2020).
- Fiker, R., Kim, L. H., Molina, L. A., Chomiak, T. & Whelan, P. J. Visual Gait Lab: a user-friendly approach to gait analysis. *J. Neurosci. Methods* **341**, 108775 (2020).
- Dunn, T. W. et al. Geometric deep learning enables 3D kinematic profiling across species and environments. *Nat. Methods* <https://doi.org/10.1038/s41592-021-01106-6> (2021).
- Wenger, N. et al. Spatiotemporal neuromodulation therapies engaging muscle synergies improve motor control after spinal cord injury. *Nat. Med.* **22**, 138–145 (2016).
- Liu, N. K. et al. A semicircular controlled cortical impact produces long-term motor and cognitive dysfunction that correlates well with damage to both the sensorimotor cortex and hippocampus. *Brain Res.* **1576**, 18–26 (2014).
- Bondi, C. O. et al. Found in translation: Understanding the biology and behavior of experimental traumatic brain injury. *Neurosci. Biobehav. Rev.* **58**, 123–146 (2015).
- Nica, I., Deprez, M., Nuttin, B. & Aerts, J. M. Automated assessment of endpoint and kinematic features of skilled reaching in rats. *Front. Behav. Neurosci.* **11**, 255 (2018).
- Pereira, T. D. et al. SLEAP: Multi-animal pose tracking. Preprint at <https://doi.org/10.1101/2020.08.31.276246> (2020).
- Segalin, C. et al. The Mouse Action Recognition System (MARS) software pipeline for automated analysis of social behaviors in mice. *Elife* **10**, e63720 (2021).
- Graving, J. M. et al. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* **8**, e47994 (2019).
- Metz, G. A. & Whishaw, I. Q. Cortical and subcortical lesions impair skilled walking in the ladder rung walking test: a new task to evaluate fore- and hindlimb stepping, placing, and co-ordination. *J. Neurosci. Methods* **115**, 169–179 (2002).

37. Metz, G. A. & Whishaw, I. Q. The ladder rung walking task: a scoring system and its practical application. *J. Vis. Exp.* <https://doi.org/10.3791/1204> (2009).
38. Loy, K. et al. Semaphorin 7A restricts serotonergic innervation and ensures recovery after spinal cord injury. *Cell. Mol. Life Sci.* **78**, 2911–2927 (2021).
39. Silver, J. & Miller, J. H. Regeneration beyond the glial scar. *Nat. Rev. Neurosci.* **5**, 146–156 (2004).
40. Bareyre, F. M. et al. The injured spinal cord spontaneously forms a new intraspinal circuit in adult rats. *Nat. Neurosci.* **7**, 269–277 (2004).
41. Ueno, M., Hayano, Y., Nakagawa, H. & Yamashita, T. Intraspinal rewiring of the corticospinal tract requires target-derived brain-derived neurotrophic factor and compensates lost function after brain injury. *Brain* **135**, 1253–1267 (2012).
42. Lai, S. et al. Quantitative kinematic characterization of reaching impairments in mice after a stroke. *Neurorehabil. Neural Repair* **29**, 382–392 (2015).
43. Maier, I. C. et al. Differential effects of anti-Nogo-A antibody treatment and treadmill training in rats with incomplete spinal cord injury. *Brain* **132**, 1426–1440 (2009).
44. Chen, K. et al. Sequential therapy of anti-Nogo-A antibody treatment and treadmill training leads to cumulative improvements after spinal cord injury in rats. *Exp. Neurol.* **292**, 135–144 (2017).
45. Hanlon, L. A., Huh, J. W. & Raghupathi, R. Minocycline transiently reduces microglia/macrophage activation but exacerbates cognitive deficits following repetitive traumatic brain injury in the neonatal rat. *J. Neuropathol. Exp. Neurol.* **75**, 214–226 (2016).
46. Brown, A. R. & Martinez, M. Chronic inactivation of the contralesional hindlimb motor cortex after thoracic spinal cord hemisection impedes locomotor recovery in the rat. *Exp. Neurol.* **343**, 113775 (2021).
47. Jacobs, B. Y. et al. The Open Source GAITOR suite for rodent gait analysis. *Sci. Rep.* **8**, 1–14 (2018).
48. Peruga, I. et al. Inflammation modulates anxiety in an animal model of multiple sclerosis. *Behav. Brain Res.* **220**, 20–29 (2011).
49. de Bruin, N. M. W. J. et al. Multiple rodent models and behavioral measures reveal unexpected responses to FTY720 and DMF in experimental autoimmune encephalomyelitis. *Behav. Brain Res.* **300**, 160–174 (2016).
50. Herold, S. et al. CatWalk gait analysis in a rat model of multiple sclerosis. *BMC Neurosci.* **17**, 1–13 (2016).
51. Kerschensteiner, M. et al. Remodeling of axonal connections contributes to recovery in an animal model of multiple sclerosis. *J. Exp. Med.* **200**, 1027–1038 (2004).
52. Merkler, D., Ernsting, T., Kerschensteiner, M., Brück, W. & Stadelmann, C. A new focal EAE model of cortical demyelination: multiple sclerosis-like lesions with rapid resolution of inflammation and extensive remyelination. *Brain* **129**, 1972–1983 (2006).
53. Jafari, M. et al. Phagocyte-mediated synapse removal in cortical neuroinflammation is promoted by local calcium accumulation. *Nat. Neurosci.* **24**, 355–367 (2021).
54. Faissner, S., Plemel, J. R., Gold, R. & Yong, V. W. Progressive multiple sclerosis: from pathophysiology to therapeutic strategies. *Nat. Rev. Drug Discov.* **18**, 905–922 (2019).
55. Zhan, J. et al. High speed ventral plane videography as a convenient tool to quantify motor deficits during pre-clinical experimental autoimmune encephalomyelitis. *Cells* **8**, 1439 (2019).
56. Pedregosa Fabianpedregosa, F. et al. Scikit-learn: machine learning in Python. Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. *Mathieu Perro. J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
57. Hsu, A. I. & Yttri, E. A. B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nat. Commun.* **12**, 1–13 (2021).
58. Bradley, P. M. et al. Corticospinal circuit remodeling after central nervous system injury is dependent on neuronal activity. *J. Exp. Med.* **216**, 2503–2514 (2019).
59. Loy, K. et al. Enhanced voluntary exercise improves functional recovery following spinal cord injury by impacting the local neuroglial injury response and supporting the rewiring of supraspinal circuits. *J. Neurotrauma.* **35**, 2904–2915 (2018).
60. Locatelli, G. et al. Mononuclear phagocytes locally specify and adapt their phenotype in a multiple sclerosis model. *Nat. Neurosci.* **21**, 1196–1208 (2018).
61. Littlejohn, E. L., Scott, D. & Saatman, K. E. Insulin-like growth factor-1 overexpression increases long-term survival of posttrauma-born hippocampal neurons while inhibiting ectopic migration following traumatic brain injury. *Acta Neuropathol. Commun.* **8**, 1–15 (2020).
62. Aljovic, A. et al. Data used in this publication. <https://doi.org/10.6084/m9.figshare.17212856.v1> (2021).
63. Zhao S. & Aljovic A. sollar/alma: first major release. <https://doi.org/10.5281/zenodo.5767422> (2021).

Acknowledgements

The authors would like to thank Luca Fabbio for animal handling as well as Dana Matzek and Bianca Stahr for animal husbandry. Work in F.M.B.'s lab is supported by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB 870 Project ID 118803580, TRR274 Project ID 408885537), by the Munich Center for Neurosciences (MCN) and the Institute for Research on Paraplegia (IRP). F.M.B. is also supported by the Munich Center for Systems Neurology (DFG, SyNergy; EXC 2145/ID 390857198). V.V.S. is supported by a post-doctoral fellowship from the Humboldt foundation.

Author contributions

F.B., A.A., and S.Z. designed the experiments. A.A., C.R., and M.C. performed all surgical procedures. A.A., S.Z., and V.V. collected and analyzed the data. F.B., A.A., S.Z., and M.K. wrote the manuscript. All authors approved the final version of the paper.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information


Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-03077-6>.

Correspondence and requests for materials should be addressed to Florence M. Bareyre.

Peer review information *Communications Biology* thanks Ronaldo Ichiyama and the other, anonymous, reviewers for their contribution to the peer review of this work. Primary Handling Editor: Luke R. Grinham. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022