Longitudinal tracking of neuronal activity from the same cells in the developing brain using Track2p

Jure Majnik[1*], Manon Mantez[1], Sofia Zangila[1], Stéphane Bugeon[1], Leo Guignard[2], Jean-Claude Platel[1,3*] and Rosa Cossart[1,3,4*]

1. Aix Marseille Univ, Inserm, INMED, Turing Center for Living Systems, Marseille, France
2. Aix-Marseille Univ, CNRS, IBDM, UMR7288, Turing Centre for Living Systems, Marseille, France
3. These authors contributed equally
4. Lead contact

*Correspondence:
jure.majnik@inserm.fr, jean-claude.platel@inserm.fr, rosa.cossart@inserm.fr

Summary

Understanding cortical circuit development requires tracking neuronal activity across days in the growing brain. While *in vivo* calcium imaging now enables such longitudinal studies, automated tools for reliably tracking large populations of neurons across sessions remain limited. Here, we present a novel cell-tracking method based on sequential image registration, validated on calcium imaging data from the barrel cortex of mouse pups over one postnatal week. Our approach enables robust long-term analysis of several hundreds of individual neurons, allowing quantification of neuronal dynamics and representational stability over time. Using this method, we identified a key developmental transition in neuronal activity statistics, marking the emergence of arousal state modulation. Beyond this key finding, our method provides an essential tool for tracking developmental trajectories of individual neurons, which could help identify potential deviations associated with neurodevelopmental disorders.

Introduction

Early postnatal development in rodents is a period of intense circuit wiring and remodelling at several scales through various major processes that include neuronal growth, synaptogenesis, apoptosis, migration, rise of intracortical connectivity, functional maturation of inhibitory synapses and the disappearance of transient connectivity schemes [1–4]. Critical periods for various sensory systems open and close during this time, further highlighting the profound reshaping of cortical networks [4]. This developmental period of remodelling is crucial for establishing the functional architecture of the mature cortex. Most importantly, all of these developmental processes are activity-dependent, with collective dynamics playing a critical role in the proper integration of neurons into functional networks [5]. These network dynamics sequentially unfold while relying on different mechanisms and circuits for their generation [1,5]. The emergence of functional brain circuits during development is therefore a precisely timed choreography, the timing of which is inherently tied to the age of the organism under study.

However, developmental age lacks precision due to significant variations in physical characteristics and growth patterns, even among offspring from the same genetic lineage (Fig. 1). Hence, longitudinal imaging from the same animal across days is the optimal solution to better capture the evolution of circuit dynamics during development. Additionally, developmental variability extends beyond the organism level, becoming even more pronounced at the single neuron level (Fig. 1). While population-level descriptions of cortical circuit development are crucial, individual neurons exhibit unique developmental trajectories rooted in their specific origin, birth timing and cellular identity [1]. In addition, the singular dynamics of sparse individual neurons can matter, as demonstrated for the rare hub cells [6,7]. Thus, in order to fully understand the circuit basis of cortical development in health and disease, it is crucial to track neuronal activity at both population and individual cell levels in the growing brain of the same animal. This hurdle spans both experimental methodology and data analysis.

Significant progress has been made over the recent years to develop innovative solutions to meet this challenge on the experimental side, using two-photon calcium imaging. Technical improvements included modified head plates and specialized surgical and care protocols [6,8–12]. Still, in many cases, different neurons from the same mouse were recorded at different ages, or the tracking of individual cells could only be achieved through visual inspection and manual annotation of relatively sparse, often genetically-subscribed neuronal populations expressing a calcium fluorescent reporter. The gold standard for developmental circuit analysis would be automated tracking of densely labeled neuronal populations, enabling efficient longitudinal monitoring while eliminating the burden of manual cell tracking. An algorithm capable of accurate automatic tracking of the same neurons across multiple days of brain development is thus essential.

Methods for tracking the activity of large populations of cells across days have been successfully developed and deployed in the adult brain [13–16]. In adult circuits, cell tracking is indeed relatively straightforward due to their structural stability, with minimal tissue growth, negligible morphological changes, and constant numbers of neurons. In contrast, the first weeks of mouse neocortical development are characterized by rapid and critical developmental changes spanning several scales from synapses, to single neurons and networks. These include extensive brain growth [17,18], substantial morphological changes [19] as

well as changes to cell numbers due to programmed cell death [10]. This makes tracking the same cells across sessions substantially more difficult during development compared to the adult brain.

To overcome the technical limitations detailed above, we developed an experimental protocol using chronic calcium imaging in mice and a novel cell tracking algorithm (Track2p), specifically tailored to development. This allowed us to track the activity of large, densely labelled, populations of neurons during early postnatal development, which has not been possible before. The algorithm overcomes the challenges of cell tracking during brain growth by applying sequential registration and cell matching steps, using the spatial overlap of cells on adjacent recordings as a matching criterion. Track2p is freely available as an open-source package with an interactive GUI, enabling researchers to analyze longitudinal calcium imaging data from both developing and mature circuits.

Applying the algorithm to a dataset recorded during the second postnatal week in mouse barrel cortex, a critical period for the formation of topographic maps in that region, yielded hundreds of identified neurons tracked across all days. Assessing the quality of the algorithm by benchmarking it on a newly generated ground truth dataset showed high tracking performance [20]. Leveraging ground truth benchmarking, we demonstrate that explicitly accounting for developmental processes, such as brain growth, are critical for accurately tracking cells during postnatal development. Our work thus shows that chronic calcium imaging and cell tracking using Track2p can be used to monitor the changing physiological properties of large populations of matched neurons during early cortical development. We demonstrate that the statistics of activity patterns in the tracked population display two periods of stability, with a critical transition point around postnatal day 11 (P11), marking the emergence of a stable behavioral state representation.

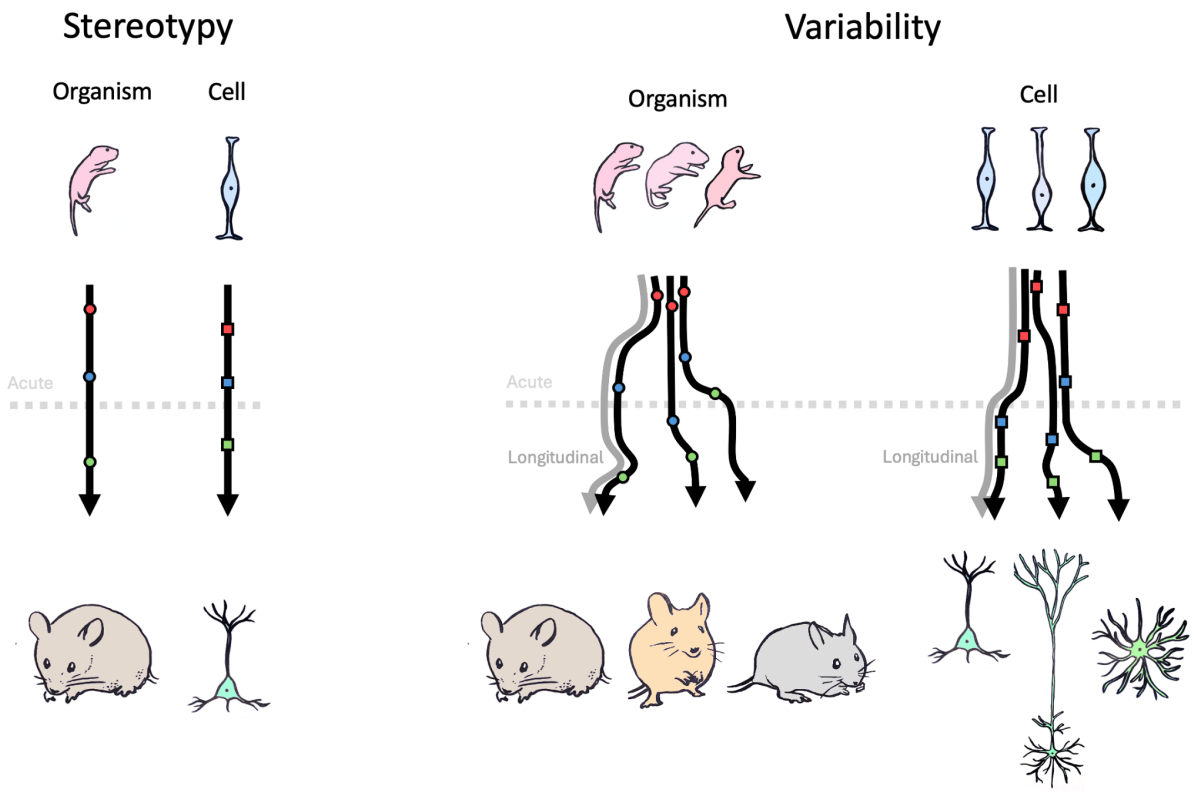


Figure 1: Tracking developmental trajectories at the organismal and cellular level. Postnatal development in mammals is not a strictly stereotyped process (left) but rather shows variability across individual organisms as well as across individual cells of the same organism (right). Performing an acute experiment (dashed grey line) only provides a single snapshot of the developmental trajectory of an individual organism (or cell). Alternatively, a longitudinal experiment (grey arrow) allows to track the properties of the same individual (or cell) throughout development, which is especially important in the case of variability in developmental trajectories (right).

Results

Cell tracking using image registration and overlap-based matching

Tracking neuronal activity across multiple days presents unique challenges due to the dynamic nature of brain development. We developed a novel tracking algorithm, called Track2p. As in other tracking algorithms for calcium imaging data [13,16,21], the final goal of Track2p is to follow individual cells across sessions allowing the user to compare their functional properties in downstream analyses. To achieve this goal, the algorithm takes as input a set of preprocessed recordings, each consisting of a set of regions of interest (ROIs i.e. putative neurons detected based on activity, see Methods) and their respective calcium fluorescence traces, as well as a mean image of the field of view (FOV). Briefly, the algorithm aims to match ROIs in any given pair of consecutive sessions based on their spatial overlap, assuming that the more the two overlap in anatomical space, the more likely they correspond to the same neuron. Due to developmental processes such as brain growth and other experimental factors, it is necessary to account for day-to-day changes that occur between the two recordings before computing spatial overlaps. This is achieved by performing affine image registration on the mean FOV images between consecutive days.

We apply the registration and spatial matching iteratively, starting with the first pair of sessions (s_0 and s_1) as follows (Fig. 2A):

Firstly, we estimate the spatial transformation between s_0 and s_1 using affine image registration (Ntatsis 2023) (i.e. allowing shifting, rotation, scaling and shearing, see Fig. 2B, the transformation is denoted as T). We employ affine transformation, since it can account both for rigid transformations (rotations and translations arising from minor mismatches in FOV alignment across experiments) as well as scaling and shearing (mostly due to brain growth). The changes across the two consecutive recordings are approximated as the transformation registering the mean FOV from session s_1 (green in Fig. 2B) to the mean FOV of s_0 serving as reference (red in Fig. 2B).

Secondly, the computed transformation is applied to the ROIs from session s_1 (green in Fig. 2C) to align them to ROIs from session s_0 (red in Fig. 2C). The amount of spatial overlap after registration (yellow in Fig. 2C, bottom), can indicate the accuracy of the estimated transformation between the two days. Assuming that the transformation is accurately estimated, ROIs corresponding to the same cells display substantial spatial overlap, with some ROIs from one recording also potentially overlapping poorly, if not at all, with any ROI from another.

Thirdly, once the ROIs are aligned, the algorithm proceeds with the matching (Fig. 2D). This is done by computing a spatial similarity metric (intersection over union - IoU) between each ROI from session s_0 and each transformed ROI from session s_1. Matches are then assigned in a globally optimal way by maximising the sum of IoU values across all matches using a linear sum assignment algorithm (Crouse 2016).

Finally, since two consecutive sessions contain different sets of detected cells (see * in Fig. 2D, bottom) and since ROIs can overlap even if the signal does not come from the same cells (see + in Fig. 2D, bottom) we perform an additional filtering step on the assigned matches.

Assuming that the IoU values for putative true and putative false matches come from different distributions, we would expect a bimodal distribution of IoU values across all assigned matches (see histogram in Fig. 2E). To reject the putative false positives, we compute a threshold based on automatic thresholding methods (Otsu's method [24]). This ensures that assigned matches with low spatial similarity are rejected (see + in Fig. 2D and E) while the matches with high similarity are accepted (see x in Fig. 2D and E). This yields the final matching for the first pair of consecutive imaging sessions (s_0 and s_1). In the case of more than 2 recordings, this tracking procedure is then iteratively applied for all consecutive pairs of sessions (s_0 to s_1, s_1 to s_2 … s_{N-1} to s_N , and so on), with tracks being extended sequentially from s_0 to s_N based on the identified matches, and terminated if a match was not identified at any particular session (Fig. 2F).

We provide an open-source implementation of the algorithm, combined with a user-friendly graphical user interface (GUI), allowing non-specialist users to run the algorithm and interact with its outputs (see Fig. 2A top, Fig. S1 and Supp. Info.: Software). Both the algorithm and the GUI come with a simple installation procedure and extensive documentation to facilitate the ease of use and accessibility of Track2p.

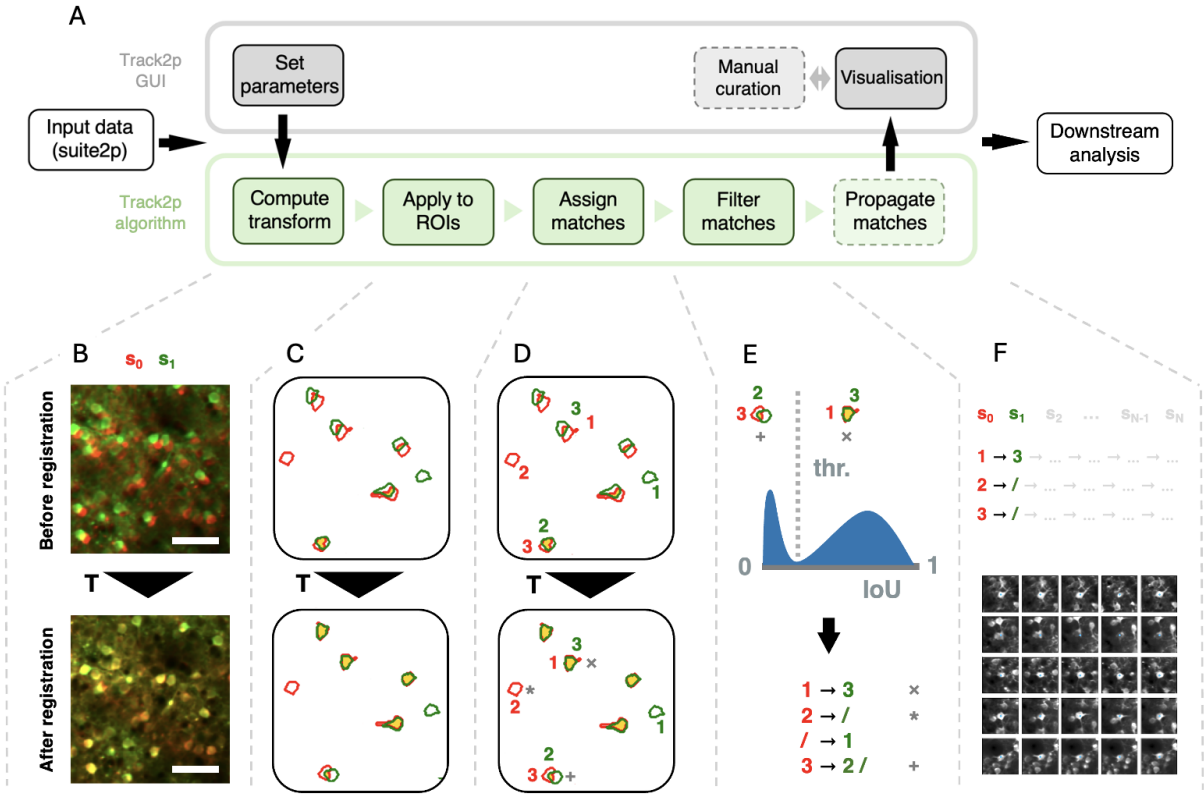


Figure 2: **The Track2p pipeline for tracking cells across recordings. A**: Schematic overview of the Track2p algorithm and the GUI capabilities. Dashed squares represent optional steps in the pipeline. **B**: The transformation (T) between two consecutive days is computed by registering the mean FOV images. Top: An overlay of the reference (red, s_0) and the moving (green, s_1) images before registration. Bottom: Same two images after registration. Scale bar: 50 µm. **C**: Applying the transform to ROIs from Suite2p segmentation. The ROI color code is the same as in B, the intersection of the ROIs is shown in yellow. Top: Overlap before T. Bottom: Overlap after T. Note: Only a few example ROIs are displayed for explanation purposes. **D**: Cell matching using linear sum assignment. ROIs are the same as C with indexes for the two recordings added for illustrative purposes. Cells from one recording are matched to cells from another maximising the sum of the intersection over union (IoU) across all matches. **E**: Filtering putative false and true matches by thresholding the IoU distribution. Top: The distribution of IoU values for matched ROI pairs shows a bimodal distribution, which is used to reject putative false positive matches. Bottom: Final result of the cell linking for a pair of recordings. **F**: Top: In the case of tracking across more than 2 imaging sessions, the steps from B to D are repeated sequentially to link the cells across all days. Bottom: Example matches for five cells (rows) successfully tracked across five consecutive days (columns). Note: The figures shown are for illustrative purposes only, see Fig. 3 for application to real data.

Tracking neurons across days in the early postnatal neocortex

In order to validate our method on experimental data, we next applied Track2p to a longitudinal dataset consisting of daily recordings of the same 720x720 µm FOV throughout the second postnatal week of mouse barrel cortex development (postnatal days 8 (P8) to P14, n=7 imaging sessions from one mouse; for more details see Methods.) In addition to the calcium indicator (GCaMP8m), we virally expressed a sparse anatomical marker (tdTomato) targeting GABAergic neurons using conditional expression in GAD67-Cre mice [22]. This dual-labelling strategy facilitated reliable FOV identification across sessions and provided an anatomical reference for Track2p's initial cross-session registration and tracking (Fig. 3, for comparison with tracking based on GCaMP8m registration see the 'Benchmarking on manually tracked cells' section). Before running Track2p, each recording was preprocessed using Suite2p [23] for active cell detection (segmentation) and calcium fluorescence trace extraction (see Methods). The outputs of Suite2p (ROIs, traces and FOV images) were then used as inputs to the Track2p algorithm.

The Track2p algorithm appeared to successfully register the mean FOV for each successive pair of imaging sessions (based on visual inspection of Track2p outputs; for the first and last pairs see Fig. 3A and B bottom left; for the equivalent of Fig. 3 across all days see Fig. S2A). Registering the ROIs for each pair of consecutive recordings showed a great degree of overlap for the majority of ROIs (yellow area in Fig. 3A and B bottom). We did however observe several ROIs that were only present in one session from the pair, likely due to differences in cell detection and developmental factors such as growth, silencing or cell death. When matching ROIs across sessions using spatial overlap (IoU), we would expect these ROIs to show significantly lower IoU values compared to the cells that were present in both sessions. Indeed, the IoU distribution of assigned matches revealed a bimodal distribution for each pair of imaging sessions, allowing us to use classical histogram thresholding methods [24], clearly separating the distributions of putative true and false matches (Fig. 3A and B bottom right, Fig. S2B). Propagating the putative matches yielded a total of 728 ROIs that were tracked across all days in this example mouse (out of 1988 ROIs detected in the first recording, Fig. 3E). We used the subset of cells successfully tracked across all days for all our future analyses.

Plotting ROIs overlaid on top of the mean image of the GCaMP8m imaging channel (Fig. 3D for contours on whole FOV, C for magnified example matches) and visually inspecting the matches using the Track2p GUI indicated excellent tracking with Track2p. These visualisations also revealed the substantial growth of the neocortex during the course of the experiment (FOV area covered by ROIs at P8 compared to P14 in Fig. 3D), with some cells growing out of the FOV (see top and left in Fig. 3D) and the pairwise distances between matched ROIs increasing by approximately 15% (Fig. 3F).

Applying Track2p on an example longitudinal imaging dataset thus demonstrated that it can successfully track activity from a large number of putatively matched neurons throughout early postnatal development in mice, as confirmed by visual inspection. To assess the tracking performance in a more quantitative and objective way, we next benchmarked Track2p's performance on manually generated ground truth for our specific experimental setting.

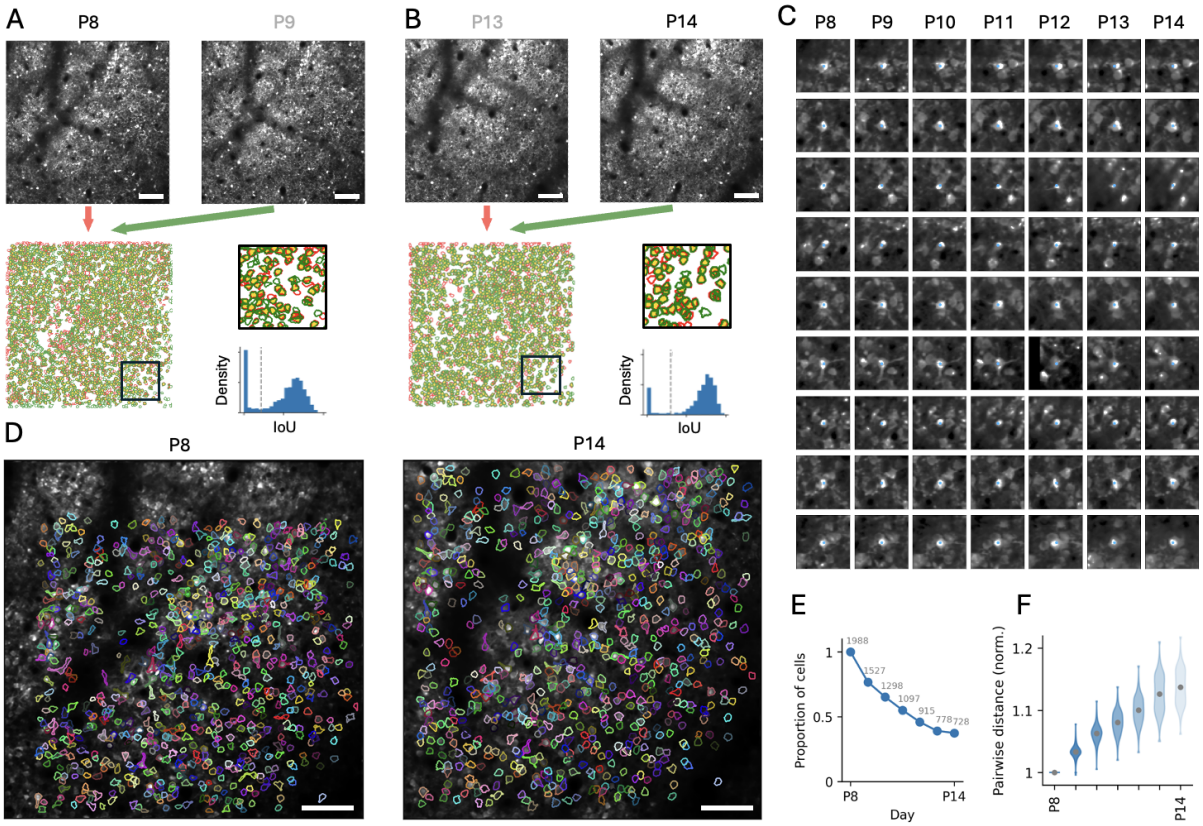


Figure 3: **Track2p tracks hundreds of cells throughout the second postnatal week of development in the mouse neocortex despite substantial brain growth. A**: Top left and right: Mean images of the 'anatomical' channel (tdTomato in GABAergic cells) for the first two imaging days (P8 and P9). Scale bar: 100 µm. Bottom left: Overlay of the two images after registration (pseudocolored as red and green respectively). Bottom middle: Overlay of the ROIs after registration (same colour code for P8 and P9). Bottom right: Distribution of IoU values for matched pairs showing the automatic threshold (grey dashed line). **B**: Same as A but for the last two imaging days (P13 and P14). **C**: Nine representative examples of matches visualised in the mean 'functional' channel (signal from pan-neuronal GCaMP8m expression) on the first and last days of recording (P8 and P14). The blue dot indicates the centroid of the ROI. **D**: Overlay of all ROIs successfully tracked across all days (N=728) on the mean 'functional' channel image of the first (P8, left) and last (P14, right) imaging days. Each tracked ROI is shown in the same color across the plots. Note the expansion of the FOV at P14 compared to P8. Scale bar: 100 µm. **E**: Graph plotting the proportion (y-axis) and absolute number (grey text) of cells successfully tracked from the first day of imaging onwards. **F**: Brain growth quantified as the relative increase in pairwise distances between tracked cells normalised to the first day. Grey dots represent the mean for each recording. Note: For visualisations in panels A, B, C and D across all days see Fig. S2.

Benchmarking on manually tracked cells

In order to benchmark Track2p, expert annotators manually tracked cells across sessions (similarly as in the Cell Tracking Challenge [20,25]). Benchmarking was performed based on a dataset from 3 mouse pups imaged under the same experimental conditions as described in the previous section (including the dataset shown in Fig. 3). For each experiment, we chose 64 homogeneously distributed ROIs detected on the first day and tracked them based on visual inspection across consecutive sessions (for more details see Methods and Fig. S3). This left us with on average 20 neurons per experiment that we were able to manually track across all days ('GT' in Fig 4.). We then proceeded to compare these to cell tracks identified by Track2p. For the purposes of evaluation we used a fully automatic tracking procedure (without manual curation and with default Track2p parameters).

Different metrics exist for evaluating cell tracking. Since we are aiming to track cells across all days, a robust cell tracking metric should reward perfect track matches with ground truth while imposing penalties for missed or incorrectly identified tracks. For this reason we used a biologically inspired 'complete tracks' quality metric, which corresponds to the F1 score for completely reconstructed full tracks (we refer to this value as 'CT' according to [20,25]). A CT score of 1 would correspond to perfect tracking, while a CT score of 0 would correspond to no matches or a large proportion of wrong matches (see Fig. 4A and Methods for more details). Additionally, to assess tracking performance over time, we quantified the proportion of reconstructed ground truth tracks over progressively longer time intervals (first two days, first three days etc. 'Prop. correct' in Fig. 4C-F, see Methods). This allowed us to understand how tracking accuracy depends on the number of successive sessions, as well as at which time points the algorithm might fail to successfully track cells.

Track2p was evaluated under three different scenarios, and compared to the performance of a widely used algorithm developed for longitudinal tracking in the adult brain (CellReg [21], see Methods). The initial evaluation was done using day-to-day registration based on a sparse anatomical marker (as in Fig. 3). Calculating the CT score in this condition showed remarkably high tracking performance for all datasets (see 'Anatomical' in Fig. 4B, mean CT = 0.93). To test whether a sparse marker was strictly necessary for successful tracking, we next ran the algorithm using the mean image of the GCaMP8m channel as a comparison. Interestingly, this evaluation showed a similar performance (see 'Functional' in Fig. 4B, mean CT = 0.91), indicating that dense calcium indicator labelling alone can provide comparable tracking performance, eliminating the requirement for sparse anatomical labeling.

To show the importance of a method tailored to the growing brain, we next compared the baseline tracking performance with two alternative conditions: firstly, by using Track2p without explicitly accounting for day-to-day growth, performing a rigid instead of affine registration of consecutive recordings; and secondly, by comparing Track2p to CellReg tracking [21]. In both scenarios, tracking performance significantly deteriorated compared to our baseline method (see 'Rigid' and 'CellReg' in Fig. 4B, mean CT = 0.22 and CT = 0 respectively). Interestingly however, these methods still managed to correctly reconstruct a portion of tracks across shorter age spans in certain cases (Fig. 4E and F), with the performance dropping significantly for longer tracks in comparison to the baseline condition (Fig. 4C and D).

These comparisons demonstrate Track2p's robust cell tracking performance in the developing brain, maintaining high accuracy over extended periods of development. However, tracking performance certainly depends on the type of experimental data (age, model system, brain area, imaging parameters, FOV alignment etc.), hence we also provide additional resources and documentation (https://track2p.github.io/home.html), allowing users to benchmark Track2p tracking for their particular use case.

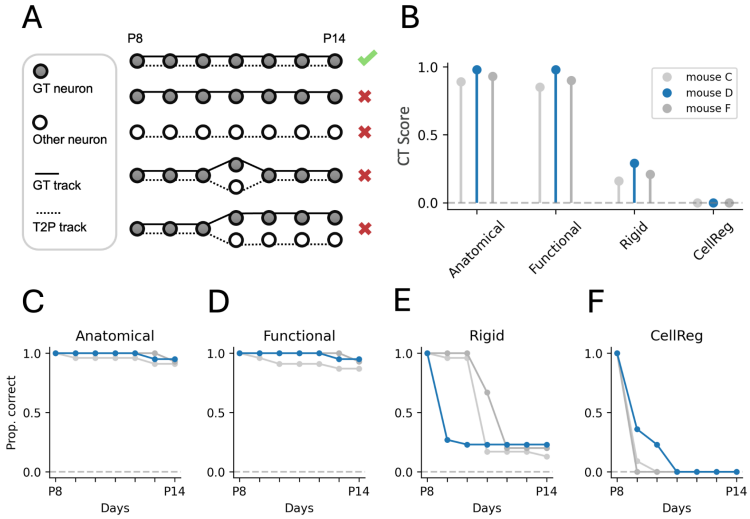


Figure 4: Evaluation of tracking performance on a manually tracked ground truth dataset during the second postnatal week of development in the mouse neocortex. A: A schematic representation of possible cases when comparing ground truth tracks (GT, solid lines) with those reconstructed by Track2p (T2P, dashed lines). The CT metric favours perfect matches (top row) and penalises all types of mismatches (bottom four rows). **B**: Graph showing the CT score for all three GT datasets across evaluated conditions. Mean CT scores of 0.93, 0.91, 0.22 and 0.00 for 'Anatomical', 'Functional', 'Rigid' and 'CellReg' respectively. Blue denotes the example mouse illustrated in Fig. 3. **C**: Proportion of fully correctly reconstructed GT traces for increasing time spans starting from P8 for the baseline ('Anatomical') condition. **D**, **E** and **F**: Same as C but for the 'Functional', 'Rigid' and 'CellReg' conditions respectively.

Development of firing statistics from hundreds of tracked neurons across postnatal development

Having validated the performance of cell tracking, we next analyzed the development of functional properties of the tracked population of neurons across the second postnatal week of mouse barrel cortex development. For this, we used a full dataset of 6 mice imaged daily for a minimum of 6 consecutive days within the second postnatal week (P7 to P14, see Fig. 5B for summary, we used this dataset for all subsequent analyses). On average 526 (± 190 std) neurons per mouse were successfully tracked across all days using Track2p, corresponding to 33 % (± 11 % std) of the neurons detected on the first day for each mouse. Based on the high tracking performance on ground truth (see Fig 4) we conclude that this drop is due to a failure in detection rather than tracking. During the course of the experiment, the weight of a mouse increased on average by 55% (± 18% std), with the pairwise distance between neurons increasing on average by 15% (± 7% std) (Fig. 5C and Fig. S5F). Weight was found to vary significantly across mice, as early as the first experimental day, and to correlate with brain growth (r=0.92, Fig. S5E). Our robust tracking capability provides a comprehensive view of the evolution of neuronal dynamics within single mice, while also allowing us to take into account the heterogeneity of the developmental timelines across mice.

The early postnatal period of cortical development studied here is characterized by the transition from synchronous calcium events recruiting many neurons to progressively more decorrelated population dynamics [26,27]. In order to visualise these global changes in our dataset of longitudinally imaged mice, we first plotted rasterplots of calcium fluorescence traces as a function of time for all tracked neurons on each imaging day in our example mouse (Fig. 5A for P8 and P14, all days in Fig. S4A). Visual inspection of these rasterplots clearly indicates the disappearance of recurring periods of highly synchronous activity from around P11 onwards as well as a global increase in single-neuron activity rates. Accordingly, quantifying calcium event rates as a function of age indicated a gradual increase in the mean and a widening of the distribution with age (Fig. 5D and Fig. S5A and G, all statistical comparisons were done between an ‘early’ (≤ P11) and ‘late’ (> P11) set of recordings). This evolution also signalled the transition towards long-tailed firing rate distributions, which are ubiquitous in adult brain circuits [28].

Next, we analyzed the distributions of pairwise correlations between activity traces from all longitudinally tracked neurons as a function of mouse age (Fig. 5D and Fig. S5B and H). Consistent with the disappearance of highly synchronous network events observed in the rasterplots, we found a significant decrease in mean pairwise correlations, indicating a progressive decorrelation of neuronal activity (Fig. 5E). The spatial distribution of pairwise correlations also evolved across days from highly correlated locally to more broadly distributed (Fig. 5F), suggesting the gradual breakdown of spatially clustered cell assemblies [5].

We finally turned to characterising the dominant population patterns of neural activity using principal component analysis (PCA). Interestingly, we observed an increase in the number of components required to explain a fixed amount of variance in the neural data across days, suggesting a developmental increase in dimensionality (Fig. 5G), consistent with the statistics of spontaneous activity described in the adult brain [29]. Notably, the plots of summary statistics for our entire dataset, indicated a clear outlier, consistent across all quantifications (Fig. 5D, E, F and G). This outlier mouse displayed a similar but delayed developmental

trajectory, compared to the other mice. Interestingly, it also showed the lowest initial weight and a less pronounced growth in weight and cortical size (Fig. 5C), likely indicating a lower maturation stage at the onset of the experiments, although the contribution of other experimental factors cannot be excluded given the invasiveness of imaging surgery.

To fully leverage our longitudinal approach, we next turned to comparing the functional properties of the same neurons across development, which is only possible when tracking cells across days.

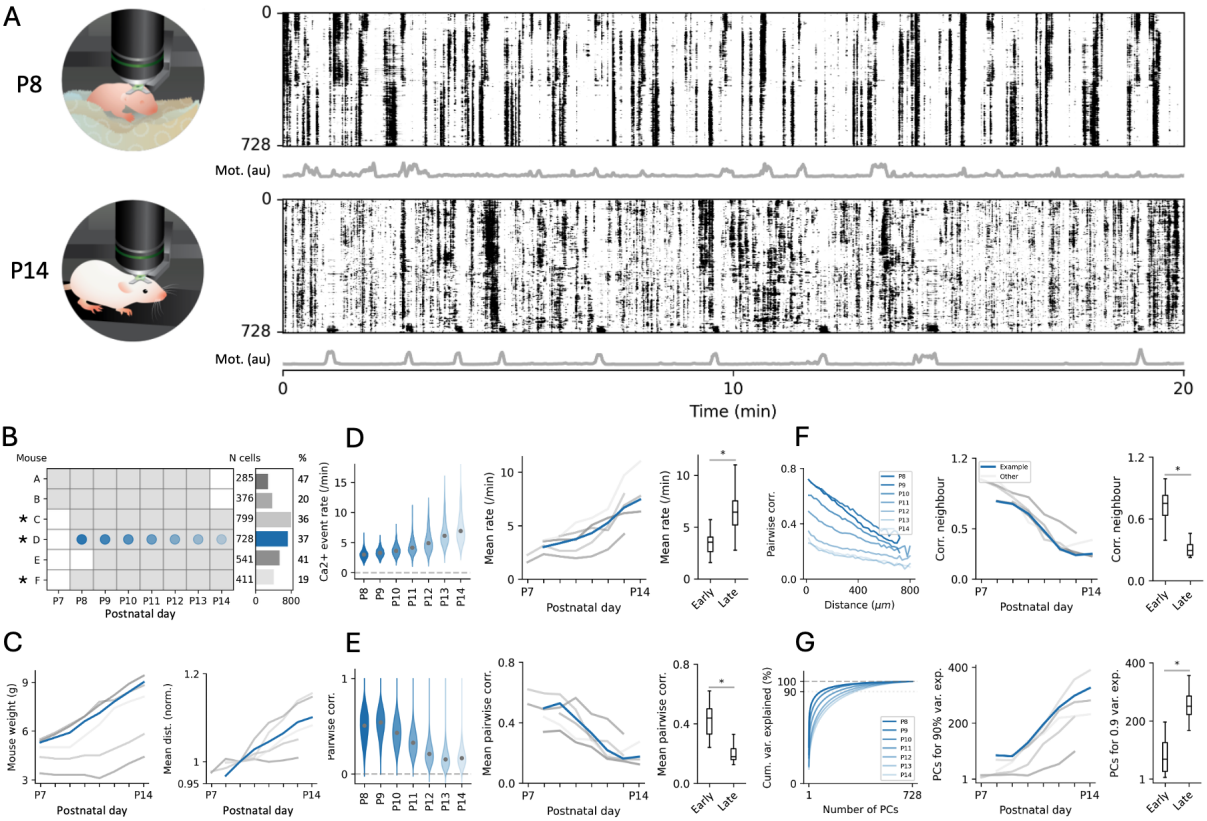


Figure 5: Evolution of neuronal activity statistics from hundreds of tracked neurons during the second postnatal week of mouse development. A: Raster plots showing the activity of all 728 tracked neurons as a function of time for the example mouse at P8 (top) and P14 (bottom). Each row in the raster corresponds to the trace of a single cell with the sorting determined by their Rastermap embedding computed at P14. Grey traces underneath the raster show a metric of global motion of the mouse computed from videography (see Methods). **B**: Overview of the dataset, indicating the imaging days for each mouse (left), the total number of cells successfully tracked across all recording days (N cells, right); * indicates mice used in the evaluation of the algorithm (Fig 3); blue denotes the example mouse. % cells: N cells as a proportion of all cells detected on the first day. **C**: Graphs plotting mouse weight (left) and the mean pairwise distance between tracked neurons (right; normalized to P9 corresponding to the earliest common day across all mice) as a function of imaging days t. **D**: Graphs plotting the distribution of calcium fluorescence event rates in all tracked neurons from the example mouse as a function of age (left), the mean value across days for all mice (middle) and a statistical comparison between the early (≤ P11) and late (> P11) epochs (right). Example mouse is shown in blue; same in E, F and G. *: Mann–Whitney U test, $p = 7.6 \times 10^{-6}$. For standard deviation see Fig. S5G. **E**: Same as D but for pairwise correlations. *: Mann–Whitney U test, $p = 1.8 \times 10^{-6}$. For standard deviation see Fig. S5H. **F**: Graphs plotting pairwise correlations as a function of anatomical distance for all pairs of tracked neurons across all ages in the example mouse (left), the estimated pairwise correlation of neighbouring neurons as a function of age for all mice (middle) and a statistical comparison across the two epochs (right). *: Mann–Whitney U test, $p = 2.7 \times 10^{-7}$. **G**: Cumulative distribution plot of the explained variance as a function of the number of principal components (PCs) for the example mouse across ages (left), number of PCs accounting for 90% of the variance as a function of age for all mice (right) and a statistical comparison across the two epochs (right). *: Mann–Whitney U test, $p = 4.0 \times 10^{-6}$.

A marked reorganisation of the functional network structure during the second postnatal week

We first examined the stability of the correlation structure, which we will refer to as 'Functional connectivity' (FC). Two alternative possibilities could be envisaged (see Fig. 6A,: (i) FC could be conserved across days, meaning that a highly connected pair on a given day would remain highly connected on the next (Fig. 6A, left); (ii) FC could reorganise, losing the structure from the previous day (Fig. 6A, right). In the first case, the FC values for all pairs would remain similar across the two days (Fig. 6A, bottom left) whereas in the second case they would be different (Fig. 6A, bottom right). To discriminate between the two cases, we defined a 'FC similarity' score for any pair of imaging days with matched neurons for a given mouse (Pearson correlation (r) across all neuron pairs, see Fig. 6A, bottom). Such FC similarity metric, encapsulates a variety of possible sources of network changes, from local or long-range connectivity, intrinsic excitability to neuromodulation.

Quantifying FC similarity across two early sessions for an example animal (P8 to P9, Fig. 6B top left) indicated remarkable stability, almost identical to the FC similarity within a given imaging day session which we took as a reference (computed as FC similarity between the first and second half of a same day recording, Fig 6B, bottom left). Conversely when comparing across more distant developmental ages, we noticed that a large part of the correlation structure was lost, resulting in lower FC similarity (P8 to P14, Fig. 6B, top middle). Interestingly, FC similarity between a pair of later developmental sessions, was again comparable to within-session similarity (P13 to P14 and P14 within, Fig. 6B, right), indicating that the functional network structure was stable, but different from that of earlier ages. Of note, within session stability was consistently lower later than earlier. Quantifying calcium event rate similarity across days indicated a similar pattern (see Fig. S6F-J).

To investigate this further we proceeded to compute the FC similarity for all possible combinations of sessions in all mice. Interestingly we observed a sigmoid-like decay of FC similarity when taking P8 as a reference (Fig. 6C). Plotting the full FC similarity matrix for all possible combinations of sessions also indicated two stable FC regimes with seemingly two blocks along the diagonal (Fig. 6D left for example, 6D right for average and 6E for statistical comparison). Interestingly, the sharpness of the transition in FC varied across mice (Fig. S6C) and could in part be explained by the weight of the mouse at the onset of the experiment (Fig. 6C, linear part of sigmoid fit). Such variation across mice could reflect either inherent individual differences in developmental processes or experimental factors, where mice with a lower initial weight may have experienced delayed development due to higher sensitivity to the imaging procedure.

Altogether, our analyses indicate that the second postnatal week marks a transition between two stable functional network structures in the barrel cortex of developing mice. Additionally, this shows how tracking cells during development opens new analysis possibilities using self-referencing between neurons, providing new insights into how the developmental choreography unfolds in the same population of neurons across days.

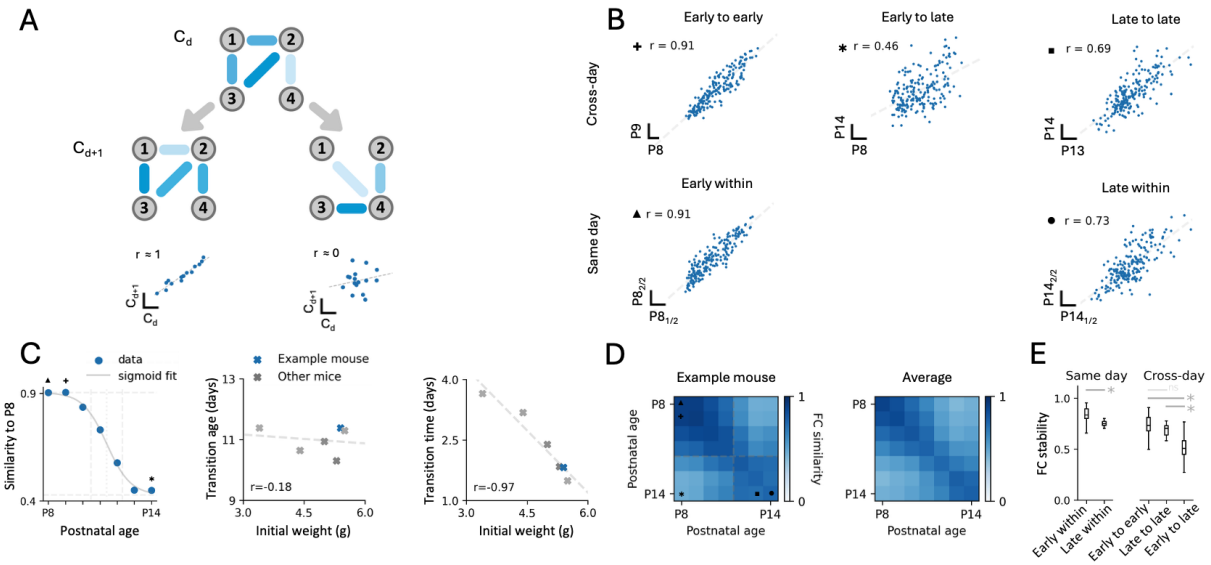


Figure 6: Transition between two stable functional network structures during the second postnatal week of barrel cortex development. A: Schematic explanation of the framework to compare functional connectivity across days; grey nodes represent neurons, strength of a functional connection is denoted by the opacity of the blue edge between two nodes. For a network of 4 neurons and a given connectivity C_d (16 connections, top) we can imagine that the connectivity on the next day (C_{d+1}) could be 'conserved' (middle left) or it could be 'reorganised' (middle right). A scatter plot comparing all 16 functional connections between C_d and C_{d+1} would indicate a high correlation in the 'conserved' case (bottom left) and low correlation in the reorganised case (bottom right). We refer to this correlation as 'FC similarity'. **B**: Scatter plots of FC for three different pairs of recording days: a pair of early sessions (P8 to P9, top left), an early and a late session (P8 to P14, top middle) and a pair of late sessions (P13 to P14, top right). Within-session FC similarity (bottom scatter plots) i.e. comparing the first and second half of an imaging session at early ($P8_{1/2}$ to $P8_{2/2}$, bottom left) and late ages ($P14_{1/2}$ to $P14_{2/2}$, bottom right). For visualisation purposes a random subset of 200 pairs is displayed; Linear fit and r (pearson correlation or 'FC similarity') are computed on all pairs; Symbols next to r values indicate the same values in panels C and D. **C**: FC similarity between P8 and all other days for the example mouse (left). Sigmoid fit: solid grey line; inflection point ('transition age', see Methods): dotted grey vertical line; linear portion of the sigmoid ('transition time', see Methods: two dashed grey vertical lines). Transition age (middle) and transition time (right) as a function of initial weight at P7 are plotted for all mice. **D**: FC similarity matrix (left) for all pairs of recording days (off diagonal), with within-session FC similarity on the diagonal for the example mouse (left) and average FC similarity across all mice for the same period (right). **E**: Box plots comparing within day (left) or across-day (right) FC similarity for early (≤ P11) and late (> P11) developmental epochs pooled across all mice. *: statistically significant; ns: not statistically significant (Kruskall-Wallis test: $p = 3.6 \times 10^{-22}$; post-hoc Mann–Whitney U tests with Bonferroni correction, early within - late within: $p = 3.3 \times 10^{-3}$, early to early - late to late: p = n. s., early to early - early to late: $p = 1.7 \times 10^{-12}$, late to late - early to late: $p = 4.2 \times 10^{-5}$, for all possible comparisons see Fig. S6K).

Developmental emergence of stable behavioral state modulation

As a first attempt to investigate the mechanisms underlying this network transition, we examined neural activity in relation to behavioral state. The second postnatal week marks the emergence of active sensation, suggesting that changes in arousal state modulation might drive this network reorganization. In the adult cortex, arousal strongly shapes both global firing rates and neural correlations [30–33], making it a promising candidate for orchestrating the developmental shift we observed.

To this aim, we sorted the neurons based on the similarity of their activity patterns, using Rastermap [34], and examined the relationship between neuronal activity and behavioral states. We assessed behavioural state indirectly using the videos capturing spontaneous mouse movement and quantifying them using a 'motion energy' metric (as in [30], for more details see Methods). Interestingly, this sorting revealed a subpopulation of neurons that were highly correlated with the animal's motion at later, but not earlier developmental stages (Fig. 5A, see Fig. S4A for all days). We first quantified this relationship by computing the temporal correlation between the first principal component of neuronal activity with the animal's motion. Interestingly, this correlation showed a steep increase after P11 for most mice (Fig. 7D). This suggested that global fluctuations in neural activity are more strongly modulated by active movement at later than earlier developmental stages.

Finally, we asked whether behavioral state could be decoded from population activity dynamics using regression analysis. Despite substantial variability, the same-day decoding performance increased with development (Fig. 7A-C). Since we tracked the same neurons across days, we could also probe the stability of this representation across days, by fitting a model on one day and testing it on all other days (cross-day decoding, Fig. 7E-H, Fig. S7B and C[15,35]). Interestingly, this analysis showed that, once developed, this representation was indeed stable, allowing for accurate cross-day decoding, with the same neurons showing either consistently positive or negative modulation across days (see Fig. S7D for the activity traces of an example cell with a high weight for decoding).

Hence, by combining Track2p longitudinal cell tracking with decoding approaches, one can systematically map the emergence and stability of cortical representations in the developing brain.

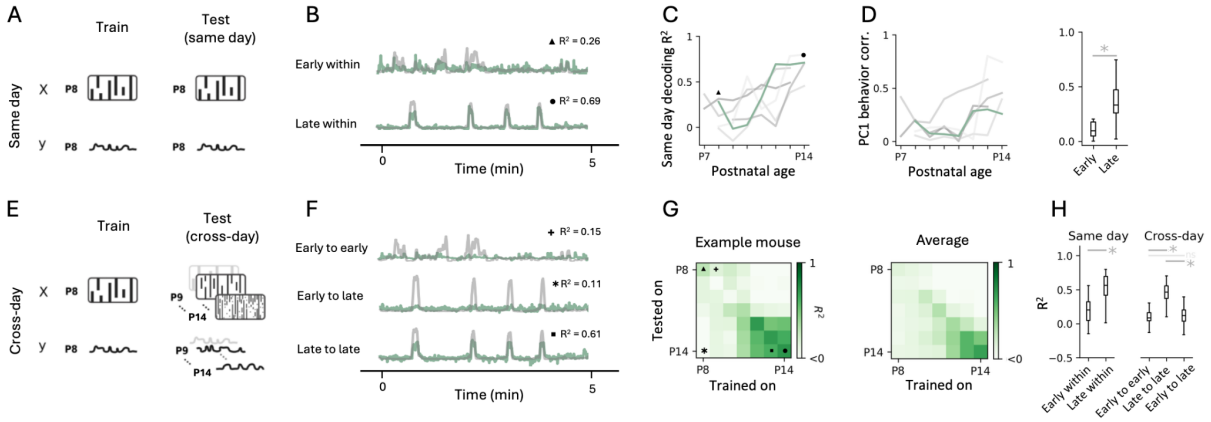


Figure 7: Regression analysis to study the development and stability of neural representations. A: Left: A decoding model is fitted on each day to predict a behavioural variable (y, mouse motion) from the simultaneously recorded calcium imaging data (X, activity raster). Right: The model is then tested on the same day (using cross-validation) to assess the encoding of the variable on that given day. **B**: Overlay of animal's motion (grey) and the predicted signal from neural activity (green) fit on the same day for example early (P8, top) and late recordings (P14, bottom). Symbols indicate the corresponding R^2 values in panels C and G. Only the first 5 min of the recording are shown for visualisation purposes, for full traces of all recordings see Fig. S7. **C**: Graphs indicating R^2 values for same-day cross-validated decoding performance as a function of mouse age. Green indicates the example mouse (also in D). **D**: Graph plotting the correlation between the animal's motion and the first principal component signal across days (left) and a statistical comparison of the early (≤ P11) and late (> P11) epochs (right). *: Mann–Whitney U test, $p = 1.2 \times 10^{-3}$. **E**: Left: Same as in A an individual model is fitted on each day. Right: To assess the stability of the representation, each model is tested across different days, which is only possible when tracking neurons across sessions. **F**: Same as B but for examples of cross-day decoding early to early (P9 to P8), late to late (P13 to P14) and early to late (P8 to P14). Symbols indicate the corresponding R^2 values in panel G. **G**: Full prediction performance matrix showing R^2 values for all combinations of fit and test datasets (rows and columns respectively) for the example mouse (left) and average across all mice for the same period (right). Diagonal entries correspond to same-day decoding, off-diagonal entries to cross-day recording. **H**: Box plots comparing R^2 values for same day (left) or across-day (right) decoding for early (≤ P11) and late (> P11) developmental epochs pooled across all mice.*: statistically significant; ns: not statistically significant (Kruskall-Wallis test: $p = 4.5 \times 10^{-10}$; post-hoc Mann–Whitney U tests with Bonferroni correction, early within - late within: $p = 1.6 \times 10^{-3}$, early to early - late to late: $p = 6.5 \times 10^{-6}$, early to early - early to late: p = n. s., late to late - early to late: $p = 2.4 \times 10^{-5}$, for all possible comparisons see Fig. S7I).

Discussion

Here we described Track2P, a cell tracking algorithm that can be used to follow the changing dynamics of hundreds of matched neurons from daily calcium imaging recordings in the growing brain of living mouse pups. At the core of the method lies sequential affine registration followed by cell matching for each subsequent pair of recordings, which lead to a better tracking performance during development when using Track2p compared to other methods. Using this benchmarked approach, we observed a sharp developmental transition from highly synchronized activity to multidimensional, behavior-state dependent neural dynamics. Beyond this key finding, our method opens a much needed tool for investigating the developmental functional trajectories of individual neurons during early postnatal brain development, and their deviations caused by genetic mutations or environmental perturbations.

Developing and benchmarking Track2p showed that it is able to track large populations of neurons despite substantial brain growth and other developmental changes through extended periods of time during development. We suggest this is mostly due to affine transform being a good approximation of the growth processes that occur between two consecutive days of imaging, with the cell matching step allowing to account for slight non-linearities in tissue growth day-by-day. Currently available methods typically register all imaging FOVs to a single reference, which we believe result in the accumulation of non-linearities making registration more difficult and resulting in poorer tracking performance.

From a practical perspective, Track2p reduces manual cell tracking time by transforming a tedious process of manual annotation into a fast and automated procedure. Indeed, considering the time needed for manually tracking all cells in our dataset, we estimate that it would require approximately 5 person-days (120 hours) of work per subject under our experimental conditions, highlighting the necessity of automatic tracking in large scale calcium imaging recordings. To aid future research, we provide Track2p fully open-source, with substantial documentation and an accompanying GUI facilitating the use of the algorithm by users without previous coding knowledge.

Besides developing the algorithm, we also showcase analysis techniques that can be used to gain unique insights from longitudinal calcium recordings, drawing inspiration from previous research studying plasticity, learning and representational drift in the adult brain [15,35,36]. We specifically highlight: 1) quantifications of functional statistics of the tracked population across days (Fig 4); 2) correlation analyses to compare these across days (Fig 5); 3) regression analyses to study the emergence and stability of neural representations across recording days (Fig 6).

Last, we would like to emphasize that this analytical advancement is grounded on optimized experimental procedures to image daily cortical dynamics through a glass window mounted on the developing pup head. Previous studies had described adaptations of the surgical procedure or head fixation to developing pups [8], however, by measuring the increase in pairwise distance between tracked neurons as a function of age per animal, we now provide a quantitative metric to estimate the invasiveness of the procedure. Interestingly, we observed a growth rate that closely matched *ex vivo* quantifications [18], indicating a minimal impact of our experimental preparation on anatomical brain growth, also confirmed by mouse weight

measurements. While we observed a slightly slower weight gain in imaged pups compared to their littermates, further investigation is needed to definitively assess whether repetitive daily imaging might influence cognitive developmental trajectories, for example through maternal separation.

Applying these techniques in our dataset during early postnatal development, we observed a significant and fast change in the nature of neuronal dynamics and state modulation in the barrel cortex, centered on P11, that manifests in several ways: (1) an increase in activity rates; (2) a decrease in pairwise correlations; (3) a shift from locally clustered to widely distributed correlations; (4) an increase in the dimensionality of spontaneous activity; (5) a remapping of the functional pairwise correlation structure; (6) the emergence of a stable representation of spontaneous active movement.

If the first three changes had already been described previously in different cortical areas [5,26,27,29,37–40], the last two could not be observed without the dense longitudinal tracking of the same neurons across several days in the growing brain permitted by Track2P. This multifaceted transition likely reflects several concurrent developmental processes, including changes in sleep architecture and neuromodulatory tone [38,41–43], maturation and rewiring of local inhibitory circuits [6,12,37,40,44–47] or disengagement from peripheral sensory inputs [38,48]. This postnatal spurt is correlated with sparsification, disappearance of the spindle-burst oscillations and increase in dimensionality of the representations [26,27,29,37,38,49–52]. Such drastic and fast changes are accompanied by significant behavioral changes, maybe the most salient ones being the change in the nature and duration of sleep as well as the start of active exploration [53]. The start of behavioural state-modulation and dimensionality increase could reflect the same phenomenon by which spontaneous motor behavior, including facial movements, drives multidimensional brainwide activity in the adult visual cortex [30]. This representation is present and stable from P11 onwards, as revealed using our decoding approach. The early absence of representation of spontaneous motor behavior is in agreement with the previously reported lack of reafferent brain activity in response to self-generated wake movement until P11 [54–56]. Thus, although we did not explicitly distinguish here between sleep and wake-generated movements, it is likely that the ‘motion energy’ metric combined with the decoding mostly took into account wake behavior, given the long periods of motor activity observed. Hence, we cannot exclude that twitching activity occurring during active sleep could contribute to an earlier representation of spontaneous motion, as reported previously [57]. More refined behavioral analysis, particularly of sleep-wake transitions and twitches, could provide additional insights.

Therefore, the mid-second postnatal week marks a transition between two stable functional connectivity structures, indicating that globally, early and late functional connectivities differ. However, singular developmental trajectories cannot be excluded. Future studies could examine whether different cell types transition at different times or different paces, for example depending on their time of birth [1], or whether unique neurons, such as hub cells, could maintain exceptional and stable functional connectivity [6]. One population of particular interest for tracking are the neurons that contribute the most to behavioral prediction. Indeed, these could form a distinct population of movement-correlated neurons embedded in specific wiring schemes, as recently demonstrated [58].

This developmental transition centered around P11, which corresponds roughly to birth in human brain development, spans a period between one and four days, depending on the initial weight of the imaged pup. This transition may not yet be the last step before a mature adult-like network but instead represent a transient state [59] preceding the emergence of mature activity patterns. Its timing, correlated with animal weight, suggests it may represent a conserved developmental milestone. Hence, comparative studies across brain regions and species, including potential parallels in human development, would be valuable.

We believe that combining Track2p tracking and the analysis methods described here provides a template for future investigations of more complex developmental phenomena, such as the emergence of sensory representations and cognitive functions. Also, longitudinal imaging uniquely enables investigation of activity-dependent development, including how early activity patterns predict later functional properties and assembly formation of the same cells. Indeed, there is nothing precluding further studies to continue tracking the activity of the same cells until adulthood. Additionally, this approach opens possibilities for targeted manipulation studies to examine how early perturbations affect subsequent circuit development. The ability to track the same cells throughout early postnatal development should thus open doors to entirely new classes of experiments not possible before, providing deeper mechanistic insights into developmental principles and pathologies. This is even more important considering that alterations of developmental trajectories at early postnatal time points are starting to be pointed out as the roots for many developmental disorders [5,60].

Resource availability

All code is available at the Track2p repository: https://github.com/juremaj/track2p, with more extensive documentation and demos available at: https://track2p.github.io/home.html. The data including pre-processed neural data for the tracked cells, behavioural data and ground truth data for cell tracking are available at: https://zenodo.org/records/17091226.

Acknowledgements

We thank all the members of the Cossart lab for helpful discussions and constructive feedback. We thank Drs. Lorenzo Fontolan and Michel Picardo for critical feedback on the manuscript. We thank INMED's animal facility and PBMC technological platform for excellent technical support. This work was supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement no. 951330), by the Fondation Bettencourt Schueller and the Fondation Roger de Spoelberch; R.C. is supported by CNRS. J-C. P. is supported by INSERM. The project leading to this publication has received funding from the French government under the "France 2030" investment plan managed by the French National Research Agency (reference : ANR-16-CONV000X / ANR-17-EURE-0029) and from Excellence Initiative of Aix-Marseille University - A*MIDEX (AMX-19-IET-004).

Supplemental information

Document S1: Figures S1-S7, Video S1 description, notes on using the Track2p GUI.
Video S1: Example of videography and 2-photon calcium imaging data acquired throughout the second postnatal week.

References

1. Cossart, R. & Garel, S. Step by step: cells with multiple functions in cortical circuit assembly. *Nat Rev Neurosci* **23**, 395–410 (2022).
2. Molnár, Z., Luhmann, H. J. & Kanold, P. O. Transient cortical circuits match spontaneous and sensory-driven activity during development. *Science* **370**, eabb2153 (2020).
3. Blankenship, A. G. & Feller, M. B. Mechanisms underlying spontaneous patterned activity in developing neural circuits. *Nature Reviews Neuroscience* **11**, 18–29 (2009).
4. Reh, R. K. *et al.* Critical period regulation across multiple timescales. *Proc National Acad Sci* **117**, 23242–23251 (2020).
5. Wu, M. W., Kourdougli, N. & Portera-Cailliau, C. Network state transitions during cortical development. *Nat. Rev. Neurosci.* **25**, 535–552 (2024).
6. Bollmann, Y. et al. Prominent in vivo influence of single interneurons in the developing barrel cortex. *Nat. Neurosci.* **26**, 1555–1565 (2023).
7. Wang, D. C., Santos-Valencia, F., Song, J. H., Franks, K. M. & Luo, L. Embryonically active piriform cortex neurons promote intracortical recurrent connectivity during development. *Neuron* (2024) doi:10.1016/j.neuron.2024.06.007.
8. Che, A. & García, N. V. D. M. An in vivo Calcium Imaging Approach for the Identification of *Cell*-Type Specific Patterns in the Developing Cortex. *Front. Neural Circuits* **15**, 747724 (2021).
9. Duan, Z. R. S. et al. GABAergic Restriction of Network Dynamics Regulates Interneuron Survival in the Developing Cortex. Neuron **105**, 75-92.e5 (2020).
10. Wong, F. K. et al. Pyramidal cell regulation of interneuron survival sculpts cortical networks. *Nature* **557**, 668–**67**3 (2018).
11. He, C. X., Arroyo, E. D., Cantu, D. A., Goel, A. & Portera-Cailliau, C. A Versatile Method for Viral Transfection of Calcium Indicators in the Neonatal Mouse Brain. Front. Neural Circuits **12**, 56 (2018).
12. Mòdol, L. et al. Assemblies of Perisomatic GABAergic Neurons in the Developing Barrel Cortex. Neuron 1–18 (2019) doi:10.1016/j.neuron.2019.10.007.
13. Beest, E. H. van et al. Tracking neurons across days with high-density probes. *Nat. Methods* 1–10 (**20**24) doi:10.1038/s41592-024-02440-1.
14. Lee, J. S., Briguglio, J. J., Cohen, J. D., Romani, S. & Lee, A. K. The Statistical Structure of the Hippocampal Code for Space as a Function of Time, Context, and Value. Cell **183**, 620-635.e22 (2020).
15. Ziv, Y. et al. Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience* 1–5 (2013) doi:10.1038/nn.3329.
16. Johnston, K. G. et al. Tracking longitudinal population dynamics of single neuronal calcium signal using SCOUT. *Cell Rep. Methods* **2**, 100207 (2022).
17. Lyck, L., Krøigård, T. & Finsen, B. Unbiased cell quantification reveals a continued increase in the number of neocortical neurons during early post-natal development in mice. *Eur. J. Neurosci.* **26**, 1749–1764 (2007).
18. Thompson, C. L. et al. A High-Resolution Spatiotemporal Atlas of Gene Expression of the Developing Mouse Brain. Neuron 83, 309–323 (2014).
19. Nakazawa, S., Mizuno, H. & Iwasato, T. Differential dynamics of cortical neuron dendritic trees revealed by long-term in vivo imaging in neonates. *Nat. Commun.* **9**, 3106 (2018).
20. Maška, M. et al. The Cell Tracking Challenge: 10 years of objective benchmarking. Nat. Methods 20, 1010–1020 (2023).

21. Sheintuch, L. et al. Tracking the Same Neurons across Multiple Days in Ca2+ Imaging Data. Cell Rep. 21, **110**2–1115 (2017).
22. Tolu, S. et al. A versatile system for the neuronal subtype specific expression of lentiviral vectors. *FASEB J.* **24**, 723–730 (2010).
23. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *Biorxiv* 061507 (2017) doi:10.1101/061507.
24. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst., Man, Cybern.* **9**, 62–66 (1979).
25. Ulman, V. et al. An objective comparison of cell-tracking algorithms. Nat. Methods 14, 1141–1152 (2017).
26. Rochefort, N. L. et al. Sparsification of neuronal activity in the visual cortex at eye-opening. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 15049–15054 (2009).
27. Golshani, P. et al. Internally Mediated Developmental Desynchronization of Neocortical Network Activity. *Journal of Neuroscience* **29**, 10**89**0–10899 (2009).
28. Mizuseki, K. & Buzsáki, G. Preconfigured, Skewed Distribution of Firing Rates in the *Hippocampus* and Entorhinal Cortex. *CellReports* **4**, 1010–1021 (2013).
29. Avitan, L. & Stringer, C. Not so spontaneous: Multi-dimensional representations of behaviors and context in sensory areas. Neuron (2022) doi:10.1016/j.neuron.2022.06.019.
30. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. Nature 1–21 (2019) doi:10.1038/s41586-019-1**34**6-5.
31. Benisty, H. et al. Rapid fluctuations in functional connectivity of cortical networks encode spontaneous behavior. Nat. Neurosci. 27, 148–158 (2024).
32. Niell, C. M. & Stryker, M. P. Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. Neuron **65**, 472–479 (2010).
33. Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F. & Petersen, C. C. H. Membrane Potential Dynamics of GABAergic Neurons in the Barrel Cortex of Behaving Mice. Neuron 65, 422–435 (2010).
34. Stringer, C. et al. Rastermap: a discovery method for neural population recordings. Nat. Neurosci. 28, 201–212 (2025).
35. Lütcke, H., Margolis, D. J. & Helmchen, F. Steady or changing? Long-term monitoring of neuronal population activity. *Trends Neurosci.* **36**, 375–384 (2013).
36. Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N. & Harvey, C. D. Dynamic Reorganization of Neuronal Activity Patterns in Parietal Cortex. Cell **170**, 986-999.e16 (2017).
37. Colonnese, M. T. Rapid Developmental Emergence of Stable Depolarization during Wakefulness by Inhibitory Balancing of Cortical Network Excitability. J. Neurosci. 34, 5477–5485 (2014).
38. Colonnese, M. T. et al. A conserved switch in sensory processing prepares developing neocortex for vision. Neuron 67, 480–498 (2010).
39. Dard, R. F. et al. The rapid developmental rise of somatic inhibition disengages hippocampal dynamics from self-motion. *Elife* **11**, e**7**8116 (2022).
40. Mòdol, L., Moissidis, M., Selten, M., Oozeer, F. & Marín, O. Somatostatin interneurons control the timing of developmental desynchronization in cortical networks. Neuron **112**, 201**5**-2030.e5 (2024).
41. Rio-Bermudez, C. D. & Blumberg, M. S. Sleep as a window on the sensorimotor foundations of the developing hippocampus. Hippocampus (2021) doi:**10**.1002/hipo.23**334**.

42. Herlenius, E. & Lagercrantz, H. Neurotransmitters and neuromodulators during early human development. *Early Hum. Dev.* **65**, 21–37 (2001).
43. Ahmad, M., Kim, J., Dwyer, B., Sokoloff, G. & Blumberg, M. S. Coincident development and synchronization of sleep-dependent delta in the cortex and medulla. *Curr. Biol.* **34**, 2570-2579.e5 (2024).
44. Marques-Smith, A. et al. A Transient Translaminar GABAergic Interneuron Circuit Connects Thalamocortical Recipient Layers in Neonatal Somatosensory Cortex. Neuron 89, 536–549 (2016).
45. Anastasiades, P. G. et al. GABAergic interneurons form transient layer-specific circuits in early postnatal neocortex. *Nature Communications* **7**, 10584 (2016).
46. Tuncdemir, S. N. et al. Early Somatostatin Interneuron Connectivity Mediates the Maturation of Deep Layer Cortical Circuits. Neuron 89, 521–535 (2016).
47. Chini, M., Pfeffer, T. & Hanganu-Opatz, I. An increase of inhibition drives the developmental decorrelation of neural activity. *eLife* **11**, e78811 (2022).
48. Gribizis, A. et al. Visual Cortex Gains Independence from Peripheral Drive before Eye Opening. Neuron **104**, 711-723.e3 (2019).
49. Minlebaev, M., Colonnese, M., Tsintsadze, T., Sirota, A. & Khazipov, R. Early Gamma Oscillations Synchronize Developing Thalamus and Cortex. Science 334, 226–229 (2011).
50. Shen, J. & Colonnese, M. T. Development of Activity in the Mouse Visual Cortex. J. Neurosci. 36, 12259–12275 (2016).
51. Micheva, K. D. & Beaulieu, C. Quantitative aspects of synaptogenesis in the rat barrel field cortex with special reference to GABA circuitry. *The Journal of Comparative Neurology* **373**, 340–354 (1996).
52. Luhmann, H. J. & Khazipov, R. Neuronal Activity Patterns in the Developing Barrel Cortex. Neuroscience **368**, 256–267 (2017).
53. Blumberg, M. S., Dooley, J. C. & Tiriac, A. Sleep, plasticity, and sensory neurodevelopment. Neuron 110, 3230–3242 (2022).
54. Tiriac, A., Del Rio-Bermudez, C. & Blumberg, M. S. Self-Generated Movements with “Unexpected” Sensory Consequences. *Curr Biol* **24**, 2136–2141 (2014).
55. Tiriac, A. & Blumberg, M. S. Gating of reafference in the external cuneate nucleus during self-generated movements in wake but not sleep. Elife 5, e18749 (2016).
56. Dooley, J. C. & Blumberg, M. S. Developmental “awakening” of primary motor cortex to the sensory consequences of movement. Elife 7, e41841 (2018).
57. Khazipov, R. et al. Early motor activity drives spindle bursts in the developing somatosensory cortex. Nature **432**, 758–761 (2004).
58. Inácio, A. R. et al. Distinct brain-wide presynaptic networks underlie the functional identity of individual cortical neurons. *bioRxiv* 2023.05.25.542329 (2024) doi:10.1101/2023.05.25.542329.
59. Domínguez, S. et al. A transient postnatal quiescent period precedes emergence of mature cortical dynamics. Elife 10, e69011 (2021).
60. Tasnim, A. et al. The developmental timing of spinal touch processing alterations predicts behavioral changes in genetic mouse models of autism spectrum disorders. Nat. Neurosci. 27, 484–496 (2024).
61. Ratsifandrihamanana, M. R., Dard, R. F., Denis, J., Cossart, R. & Picardo, M. A. Protocol to image and analyze hippocampal network dynamics in non-anesthetized mouse pups. *STAR Protoc.* **4**, 102760 (2023).
62. Stringer, C. et al. Spontaneous behaviors drive multidimensional, brainwide activity. Science **364**, 255 (2019).

63. Ntatsis, K. et al. itk-elastix: Medical image registration in Python. *Proc. 22nd Python Sci. Conf.* 101–105 (2023) doi:10.25080/gerudo-f2bc6f59-00d.
64. Crouse, D. F. On implementing 2D rectangular assignment algorithms. *IEEE Trans. Aerosp. Electron. Syst.* **52**, 1679–1696 (2016).
65. Varoquaux, G. et al. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *NeuroImage* **145**, 166–179 (2017).
66. Glaser, J. I. et al. Machine Learning for Neural Decoding. *eNeuro* **7**, ENEURO.0506-19.2020 (2020).

Methods

Data acquisition

Animals

All experimental procedures were approved by the French ethics committee (Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation (MESRI); Comité d'éthique CEEA-014; APAFiS # 30716-2021032215171216 v8) and conducted in agreement with the European Council Directive 86/609/EEC. GAD67-Cre mice were kindly donated by Dr. Hannah Monyer (Heidelberg University). Mice were bred and stored in an animal facility with room temperature (RT) and relative humidity maintained at 22 ± 1° C and 50 ± 20%, respectively. Mice were provided *ad libitum* access to water and food. A total of 6 mice were used in the study, all heterozygous GAD67-Cre transgenics.

Virus injections

We performed viral injections at P0 as previously described [61]. Briefly, we prepared a solution of AAV-hSyn-GCaMP8m and AAV-FLEX-tdTomato (2:1 volumetric ratio, 10^{12} genome copies per millilitre; Addgene) with a small volume (10:1) of 0.05% trypan blue (T8154 Sigma) to verify the success of the injection. We then briefly anaesthetised the mouse on ice and injected 2µL of the solution in the right lateral ventricle using a glass micropipette. Pups were left to recover on a heating pad at 37°C before being returned to the dam.

Cranial window surgery

All procedures were performed as in [6]. Surgeries were performed at P7 for all mice. Briefly, betadine and lidocaine cream were applied topically, covering the area of the intended incision. Isoflurane was used for induction of global anaesthesia and maintained via a nose cone throughout the procedure. A heating pad was used to maintain body temperature. After skin removal a head plate (4 mm inner diameter, Luigs and Neumann) was fixed to the part of the skull covering the barrel cortex using glue (SuperGlue3) and Super Bond (DSM Dentaire). A craniotomy was performed within the head plate opening before finally applying a thin layer of Kwik-Sil (WPI) to the surface of the dura and covering it with a glass cover-slip (3 mm, Warner Instruments). The cover-slip was last fixed to the headplate and the skull again using glue and Super Bond. Mice were left to recover on a heating pad at 37° C for at least 1 hour before returning them to the home cage.

Chronic 2-photon calcium imaging

Longitudinal two-photon calcium imaging was performed for each mouse for at least 6 consecutive days (see Fig. 5 for details). Imaging was performed using a Bruker (Ultima 2P) microscope with a Coherent Mai-Tai excitation laser (950 nm excitation light) and a 16x Nikon objective (NA 0.8). Before detection, emitted light was split into two optical paths each associated with a filter (red and green, 580-620 and 500-550 nm respectively), allowing us to simultaneously record the GCaMP8m and the tdTomato signal. The acquisition was performed using the Prairie view software. All recordings were performed in layer 2/3 (depth between 100-200 µm from the pial surface) with a 720 x 720 µm field of view and 512 x 512 pixel resolution. Imaging rate was 30 Hz (resonant scanner) and each session lasted 20 minutes. All experiments were performed in the dark, under sensory-minimised conditions, with mice being free to spontaneously run on a non-motorised treadmill (Luigs and Neumann). To

870 facilitate alignment and cell tracking, we kept the alignment of the head mount with respect to
871 the microscope fixed across all sessions for the same mouse. To record from the same region
872 across days we manually aligned the imaging plane in x, y and z to best match the reference
873 images of the red channel (tdTomato) from all previous recording days. During the course of
874 each recording, pups were kept warm by a heating element mounted to the imaging setup.
875 After each imaging session, the pups were returned to their home cage with the dam and their
876 littermates.
877
878 **Videography**
879 All videography was performed using a Basler camera (Basler ACE2 1920), with an infrared
880 LED light source (ThorLabs 850 nm) pointed at the mouse. Videos were recorded at 30 Hz,
881 with the microscope acquisition acting as a trigger for camera frame acquisition, also allowing
882 for simple synchronisation across the two modalities. All acquisition was done using custom
883 python scripts using the PyPylon wrapper for the Basler camera software (Pylon Camera
884 Software Suite, https://github.com/basler/pypylon).
885
886 **Processing of calcium imaging data**
887 Calcium imaging data was preprocessed using the Suite2p pipeline, sequentially performing
888 motion correction, ROI detection, signal extraction and spike deconvolution [23] for each
889 recording separately. Suite2p additionally provides a cell classification feature, providing a
890 probability of a classifier categorising an ROI as being a true cell. We considered all ROIs
891 above the default threshold of 0.5 as true cells. We used baseline corrected fluorescence
892 traces as our dF/F (using the default Suite2p parameters) for all subsequent analyses. To
893 facilitate the use of Track2p with other preprocessing pipelines, we implemented a data loader
894 that takes as input three simple NumPy arrays corresponding to ROIs, field of view and traces
895 (for more information see: https://track2p.github.io/run_inputs_and_parameters).
896
897 **Preprocessing videography**
898 We used the global movements of the mouse as a proxy of its arousal state. Similarly as in [62]
899 we quantified these by looking at the pixel-wise difference of consecutive frames in the
900 videography recording. Namely, we first took each two consecutive frames, computed their
901 pixelwise difference. We then squared all individual pixel-wise values and summed across
902 pixels. This yielded a scalar value quantifying the motion of the mouse at each time point,
903 which was used for all subsequent analyses.

904 # Cell tracking

905 **Image and ROI registration**
906 As explained in the main text, Track2p aligns the ROIs based on mean FOV image registration.
907 The implementation allows the user to choose which channel to use for computing the
908 transformation ('anatomical' or 'functional' channel), as well as which types of transformation
909 to estimate (rigid or affine). Unless otherwise stated, we used the 'anatomical' (tdTomato in
910 GAD67-Cre expressing cells) channel with affine transformation (referred to as the 'baseline'
911 condition). For the purposes of evaluation we additionally ran tracking using the functional
912 channel with affine registration ('functional' in Fig.4) and using the anatomical channel with
913 rigid registration ('rigid' in Fig.4). Once registering the images, we applied the same transform

914 to all ROIs from the subsequent session that were considered as cells (see main text). All
915 Track2p image registration is done using the itk-elastix toolbox [63].

Cell matching

Cell matching was done in a way to maximise the sum of the intersection over union (IoU) across all matches of ROIs between two sessions. For this reason we defined a cost matrix M with entries corresponding to:

$$M_{i,j} = 1 - \mathrm{IoU}(r_{s_k,i}, r'_{s_{k+1},j})$$

Where $r_{s_k,\,i}$ is the i-th ROI in session s_k and $r'_{s_(k+1),\,j}$ is the j-th ROI in the subsequent session s_{k+1} after registering it to the coordinate system of s_k. Assigning matches between the two sets of ROIs in this way corresponds to a linear sum assignment problem, with the goal to find a matrix X that yields the optimal assignment cost:

$$\min_X \sum_{i=1}^{m_s} \sum_{j=1}^{m_{s+1}} M_{i,j} X_{i,j}$$

Where $X_{i,j}$ equals 1 if and only if ROI $r_{s,i}$ is assigned to $r_{s+1,i}$ and 0 otherwise. Additionally, since the number of ROIs is not necessarily the same across two sessions, M and X are not necessarily square. In this case, if there are more rows than columns, then not every row needs to be assigned to a column, and vice versa. We solve this problem by using the algorithm described in [64] and implemented in SciPy.

Additionally, some of the matches can be putative false positives. To avoid this issue, we use an approach similar to the one described in [21], using a statistical threshold to filter matches based on their IoU distribution. To do this we use the Skimage implementation of Otsu's method applied to the IoU histogram [24].

Generating a ground truth dataset

Since manually tracking all cells would require prohibitive amounts of time (see Discussion), we decided to generate sparse manual annotations, only tracking a subset of all cells from the first recording day onwards. To do this, we took the first recording (s_0), and we defined a grid of 8 x 8 (64) equidistant points over the FOV and, for each point, identified the closest ROI in terms of euclidean distance from the median pixel of the ROI (see Fig. S3A). We then manually tracked these 64 ROIs across subsequent days. The manual tracking was done using the Suite2p GUI, by visualising the FOV and masks from two recordings side by side and choosing the matching ROI from the subsequent recording or terminating the track if we could not find a match (see Fig. S3B). Only neurons that were detected and tracked across all sessions were taken into account and referred to as our ground truth dataset ('GT' in Fig. 4). When comparing the GT to Track2p tracks, we only considered Track2p tracks that originated from one of the 64 ROIs chosen for evaluation initially. All manual tracking was performed blind to the Track2p outputs.

Tracking evaluation metrics

For the evaluation of Track2p under different conditions and comparison to CellReg, we used the 'Complete tracks' metric ([20,25], Fig. 3B), defined as:

$$CT = \frac{2 \cdot T_{rc}}{T_c + T_{gt}}$$

Where T_{rc} is the number of perfectly reconstructed tracks, T_c is the number of total tracks identified by Track2p for the 64 s_0 ROIs chosen for manual tracking and T_{gt} is the number of total tracks in the ground truth dataset (from the same 64 s_0 ROIs). In the case of perfect tracking, CT will be equal to 1 (where all computed tracks are equivalent to the ground truth tracks ($T_{rc}=T_c=T_{gt}$)). Conversely, in the case of failed tracking, the value of CT will be close to 0 (for example when there is a poor match between ground truth and reconstructed tracks ($T_{rc}\approx 0$) or when there are many falsely reconstructed tracks not present in the ground truth ($T_c >> T_{gt}$)). The CT metric is mathematically equivalent to an F1 score where true positives are defined as perfectly reconstructed tracks (Fig. 4A row 1), false negatives as tracks from GT without a match in Track2p (Fig. 4A row 2) and false positives as tracks from Track2p without a match in GT (Fig. 4A rows 3, 4 and 5).

In the final step of evaluation we looked at when the algorithms from the original evaluation failed in their tracking of the ground truth neurons. For this we used a metric of the proportion of correctly reconstructed tracks for an increasingly longer number of sessions (s) (referred to as 'Prop. correct' in Fig 4C-F):

$$T_{rc_{prop}}(s) = \frac{T_{rc}(s)}{T_{gt}}$$

Where T_{gt} is equivalent as above and $T_{rc}(s)$ is equivalent as T_{rc} but for full tracks up to session s. Since in this case we kept the GT consistent with the initial evaluation, including only the neurons identified across all days, this metric is agnostic to possible false positives at shorter time epochs.

To facilitate evaluation by individual users under different experimental conditions, we provide a helper script (Jupyter notebook) aiding the whole evaluation process by both defining a grid of cells to manually track (see previous section), as well as to compute the tracking quality metrics once the ground truth dataset is completed.

CellReg Tracking

To compare Track2p with CellReg, we ran the MATLAB implementation of CellReg tracking (https://github.com/zivlab/CellReg) according to the provided user manual. We first attempted tracking using the default rigid registration, which yielded an error, terminating the algorithm and indicating that subsequent sessions do not resemble the reference session and suggesting to use non-rigid registration. We then re-launched the algorithm with non-rigid registration and used those results to evaluate the tracking the same way as for Track2p (see Results). Notably, even when running CellReg using non-rigid registration, we noticed that the

algorithm did not find any tracks spanning all days (T_{rc} = 0), which explains the CT score of 0 for all day evaluation.

Functional properties of tracked neurons

Calcium event rates

To quantify the calcium event rate statistics we used SciPy's peak detection algorithm (Python). For this we first denoised the traces slightly by averaging using a bin size of 10 frames and then proceeded to detect any peaks with a height and prominence of at least one standard deviation. We then calculated the rate as the number of peaks per minute within the recording. When quantifying the stability of the rates across days we computed the Pearson correlation coefficient across all neurons for each possible combination of sessions recorded from the same mouse.

Pairwise correlations and PCA

We quantified pairwise correlations by computing the Pearson correlation between the full traces of all pairs of simultaneously recorded neurons within each session. To quantify the spatial properties of pairwise correlations, we additionally calculated the euclidean distance between ROI centroids for each corresponding pair of neurons. When plotting the pairwise correlations as a function of pairwise distance we averaged in bins of 30 µm. The correlation of neighbouring neurons was estimated as the intercept of an exponential fit to the full data.

When comparing the stability of the correlation structure across days, we refer to this as 'FC stability', which we calculate as the Pearson correlation between all pairs of original correlation matrices for a given mouse. When fitting a sigmoid using the first day as the reference, we fix the upper limit of the sigmoid to within-day FC stability for that day and leave the other parameters free. We calculate the inflection point ('transition age') and the linear portion ('transition time') of the sigmoid using the extrema of the first and fourth derivatives respectively.

For principal component analysis (PCA) we used the scikit-learn implementation (Pedregosa 2012, Python).

Decoding

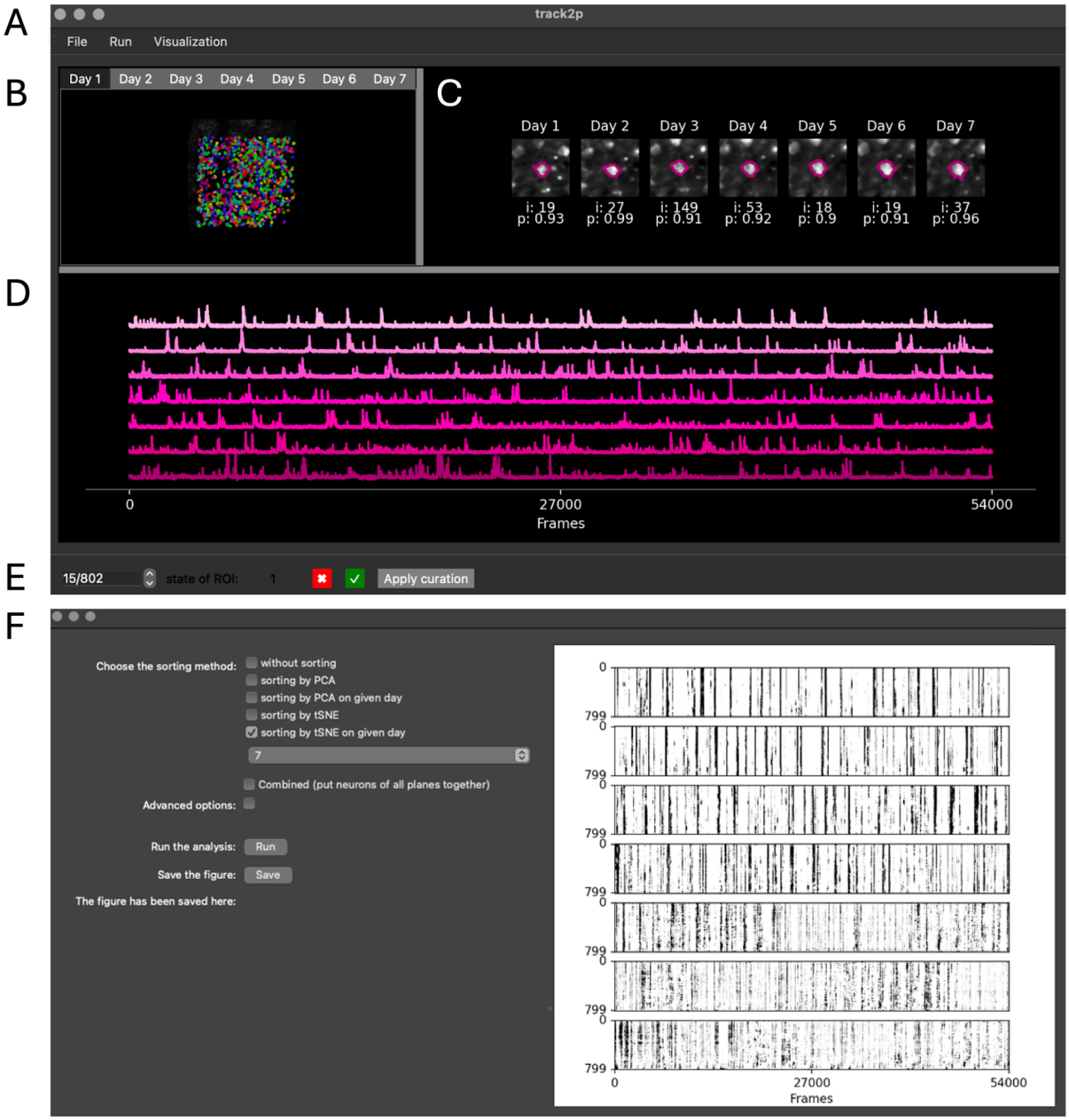
All decoding was done using linear regression with ridge regularisation (ridge regression) to avoid overfitting given the large number of neurons. Ridge regression optimises the weights (β) that minimise the following loss function:

$$\mathcal{L}(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2$$

Where in our case y is behavioural data, X is neural data and λ is a regularisation parameter. To choose the optimal λ and to reliably estimate the model performance on same-day decoding we used nested cross-validation (see for example [65,66]), ensuring data efficiency and a splitting between training, validation and test data. Briefly, a grid search for the optimal λ is performed by using the train set to fit models with different λ values and choosing the value corresponding to the model with the best performance on the held out validation set (model selection). This model is then evaluated on the test data that was used neither in model fitting

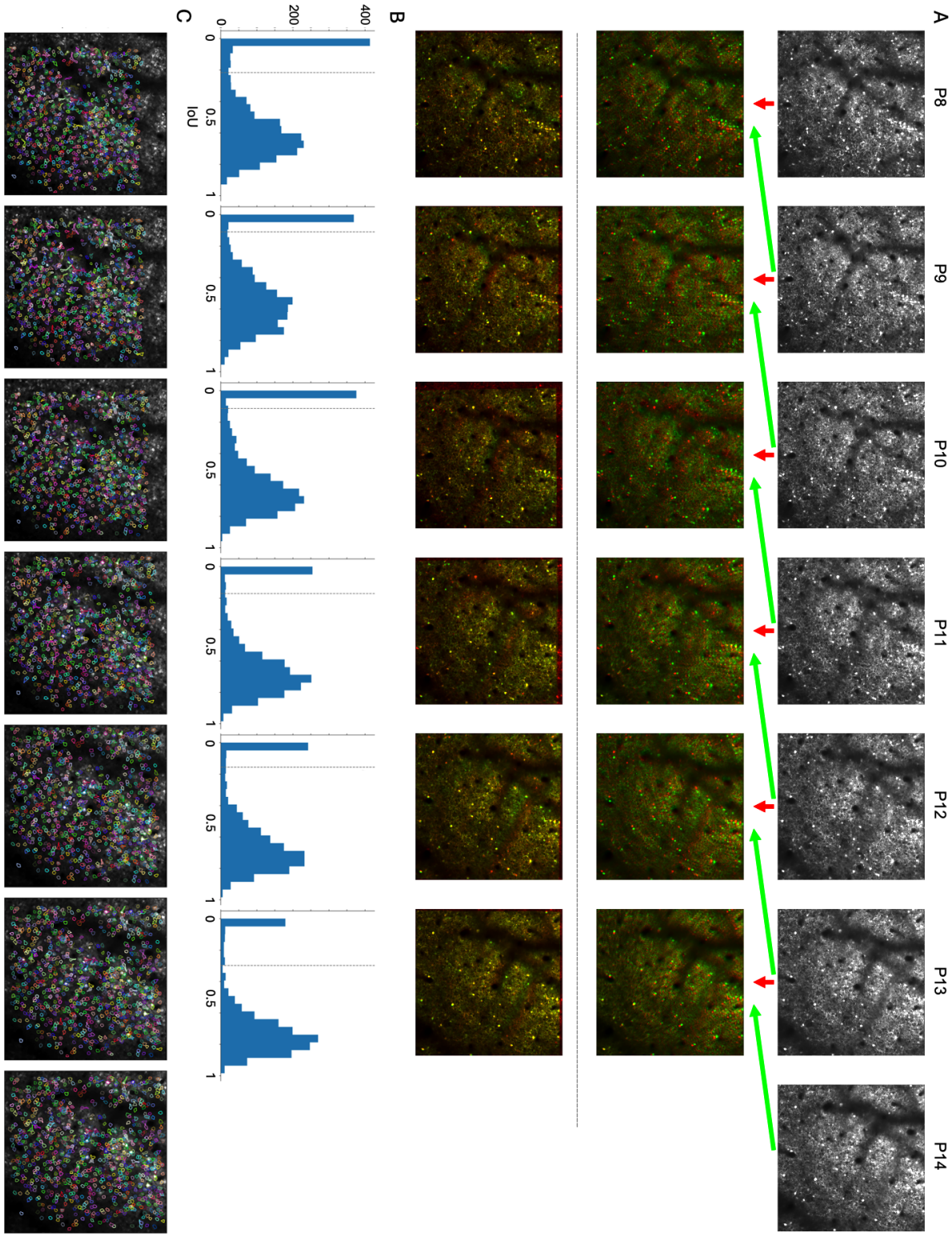
nor in the hyperparameter optimization (model evaluation). To ensure data efficiency, nested cross-validation repeats this procedure for all possible splits of the data, with an outer cross-validation loop used for model evaluation and an inner loop for model selection. We used 5 fold splits for both the inner and outer loops, splits were done on consecutive 2 minute blocks of the recording. For cross-day decoding we fit a new model with the optimal lambda for that given day and evaluated it on all other days. For all decoding analysis we slightly denoised the dF/F as well as the behaviour traces by averaging in bins of 10 consecutive timestamps. All models were implemented and fitted using PyTorch (Python).

Supplementary Information: Figures

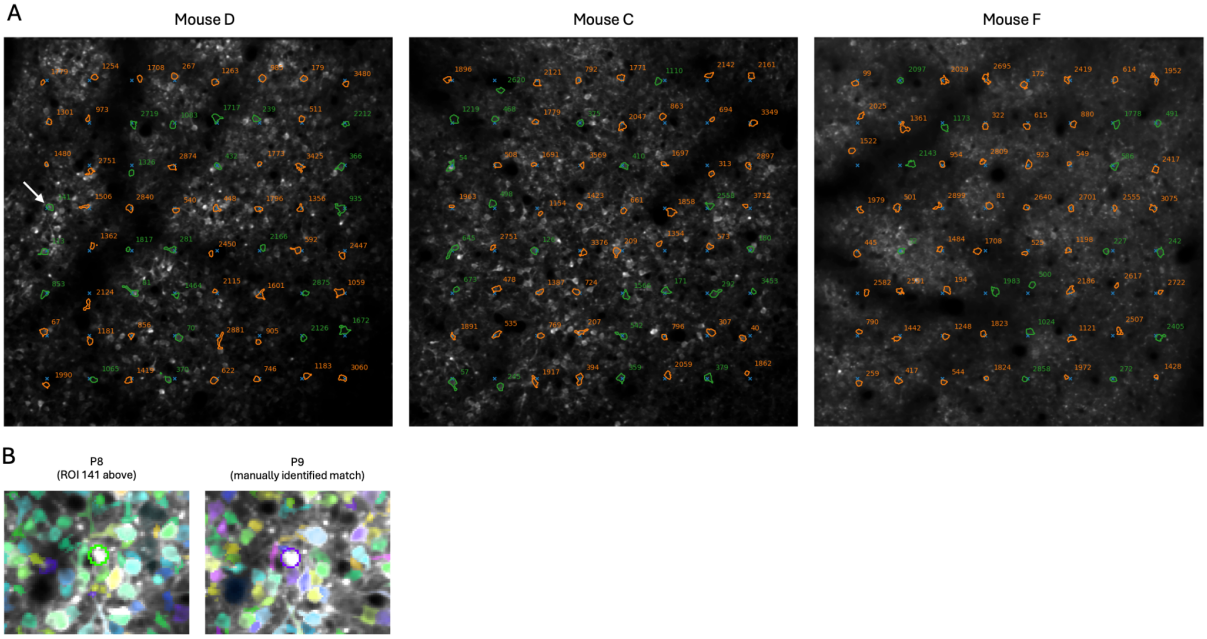


Supplementary Figure 1. Overview of the graphical user interface for interactive visualisation and curation of tracked cells. **A**. The menu bar allows the users to run the Track2p algorithm (under 'Run'), import the results of a previous run (under 'File') or to visualise population activity (under 'Visualisation'; opens the raster window in F) **B.** The interface displays the mean image for each day of the recording (the channel can be chosen by the user), with ROIs of the cells tracked across all days overlaid. This plot allows the user to select which cell to visualise across days. This visualisation corresponds to Fig. 3D. **C.** Once a cell is selected this section provides a high magnification image of the FOV for all days. Underneath each high magnification image the Suite2p index on that day (i) and the Suite2p cell probability (p) are displayed. This visualisation corresponds to Fig. 2E. **D.** The activity traces are presented in order from the first (top) to the last (bottom) session of the tracking. (The type of traces can be chosen by the user). **E.** Curation bar on the bottom of the window allows browsing all tracked cells to perform the manual curation. If a track is considered faulty during the manual

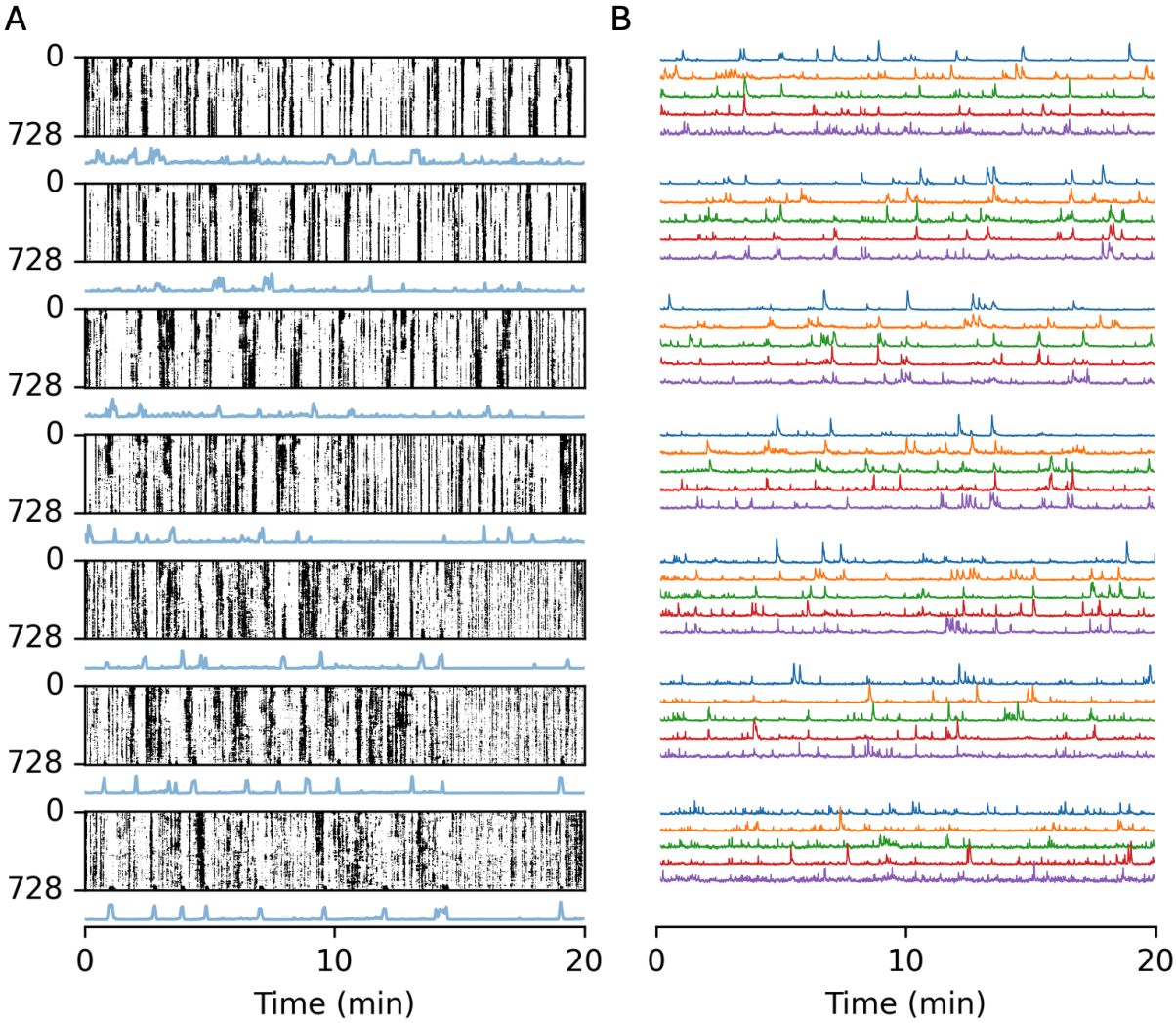
inspection, its state can be set to zero (by clicking on the cross). (The ‘Apply curation’ is used to update the FOV visualisations by coloring all contours of tracks assigned as bad in white.) **F**. Raster visualisation window allows the users to visualise the time series of the whole population and sorting the rasters to show population level features and their stability across days.



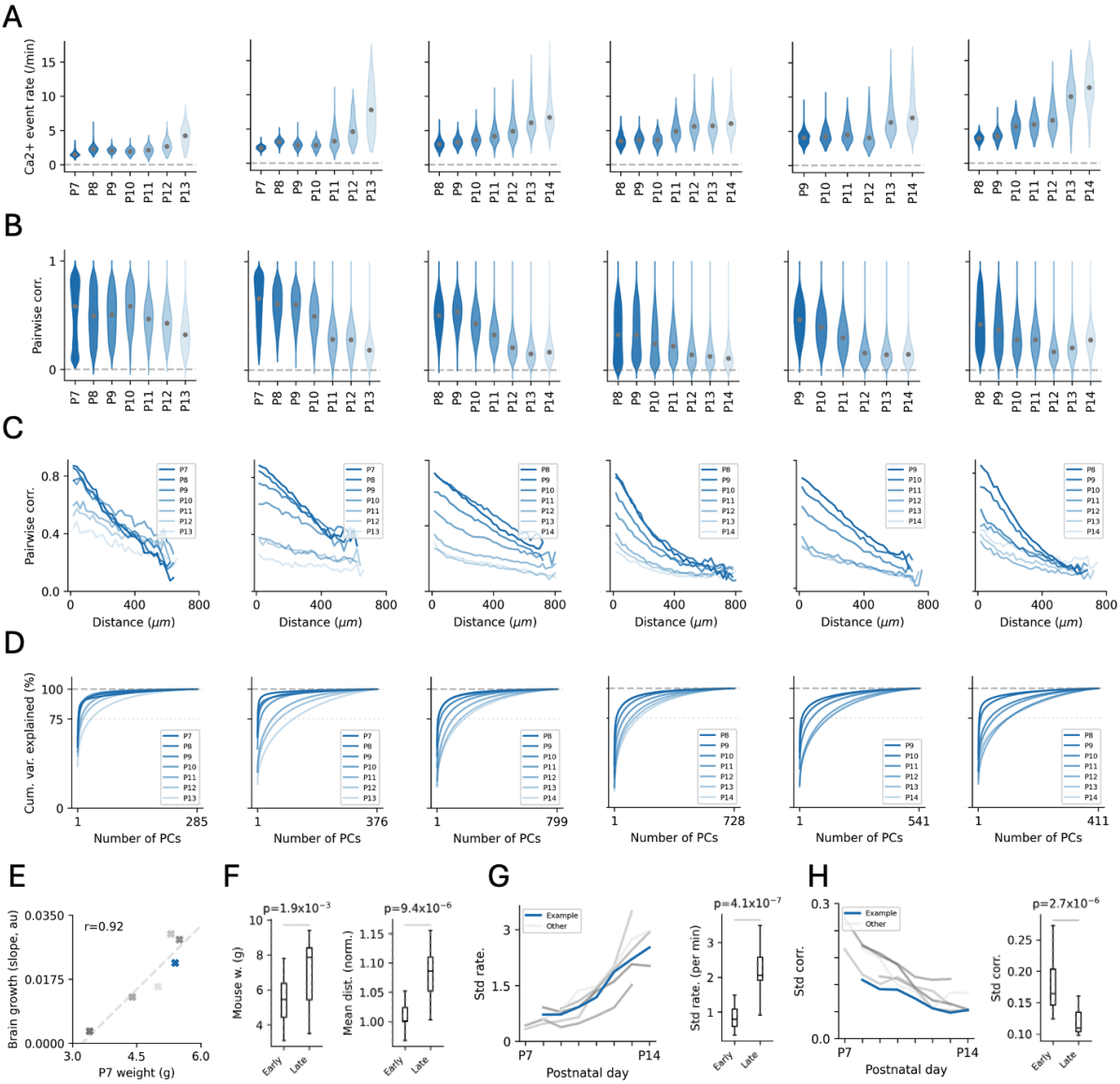
Supplementary Figure 2: Track2p outputs across all days for the example mouse. A. Top: Mean FOV of the sparse anatomical marker (tdTomato) expressed in GABAergic neurons for all days. Middle: Overlay of two neighbouring days before registration. Red: reference image, green: moving image. Bottom: Overlay after registration. **B**. Histogram of IoU values for all assigned ROI matches in each pair of subsequent recordings. Statistical threshold for filtering matches is shown with a dashed line. **C**. Overlay of all ROIs successfully tracked across all days on the mean FOV image of the GCaMP8m channel for all imaging days. Each tracked ROI is shown with the same contour color across all days.



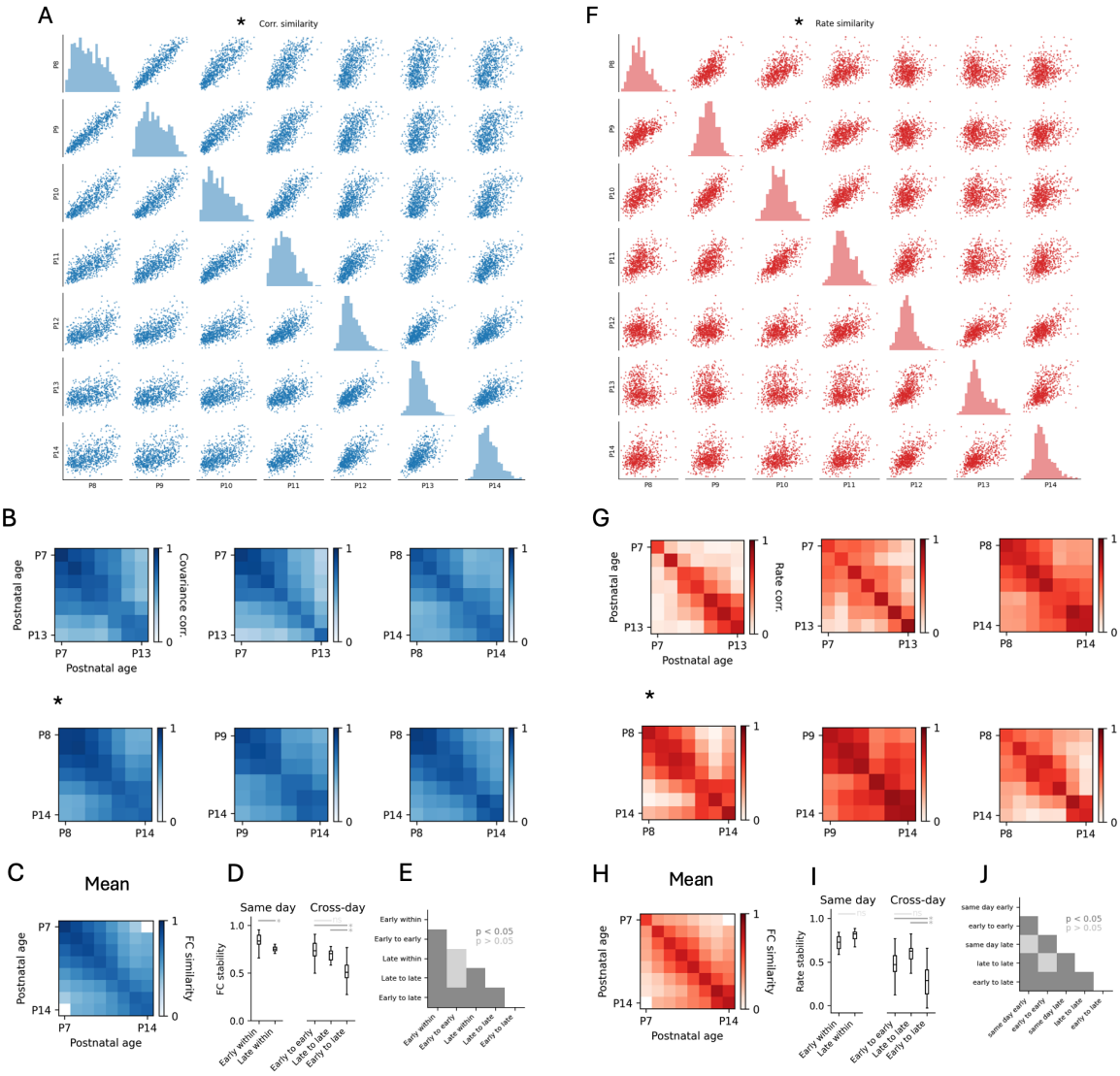
Supplementary Figure 3. Generating a ground truth dataset for Track2p evaluation. A: Overlay of ROIs selected for manual tracking on top of the mean image of the GCaMP8m channel on the first day of the recording for all three datasets used in the evaluation. Blue crosses indicate an 8x8 grid of equidistant points used to select ROIs for manual tracking. Contours are denoting the closest ROI to each of the points from the grid, with the number denoting the index of this ROI in the first dataset. Green ROIs denote ROIs that were manually tracked across all days, orange shows those that could not be manually tracked in at least one of the sessions. Note that green cells show a relatively homogenous distribution across the FOV. **B**. Example showing the manual tracking procedure. Left: A ROI chosen for manual tracking (index 141, green circle) visualised in the Suite2p GUI on the first day of recording (P8). Right: The same ROI identified manually on the next day (purple circle). Surrounding contours denote other surrounding ROIs detected on that imaging day.



Supplementary Figure 4. **Activity of tracked cells across all days for the example mouse. A**. Raster plots and simultaneously recorded mouse movements (blue traces) for all tracked neurons on each subsequent day from P8 (top) to P14 (bottom) for the example mouse (related to Fig.4A). **B**. Representative example dF/F traces of individual tracked neurons, with the trace of each neuron displayed in the same color across days. Neurons were chosen in equal increments along the rows of the corresponding raster plots in A.



Supplementary Figure 5. **Development of neuronal activity statistics in the tracked population of neurons for the full dataset. A**. Distributions of calcium event rates (related to Fig. 5D). **B**. Distributions of pairwise correlations (related to Fig. 5E). **C**. Pairwise correlations as a function of distance (related to Fig. 5F). **D**. Cumulative proportion of total variance explained by PCA (related to Fig. 5G). **E**: Relationship between the initial weight of each mouse and its brain growth estimated as the slope of the line of pairwise cell distances (Fig. 5C). **F**: Statistical comparison of mouse weight (left) and normalised mean distance between cells (right) between the two developmental epochs (related to Fig. 5C). **G**: Standard deviation of the calcium fluorescence event rates population of tracked neurons across days for all mice (left) and a statistical comparison between the early and late epochs (right) (related to Fig. 5D). **H**: Same as G but for pairwise correlations (related to Fig. 5E).



K Kruskal-Wallis p-value: $3.6 * 10^{-22}$

Pairwise Mann-Whitney U test with Bonferoni correction:

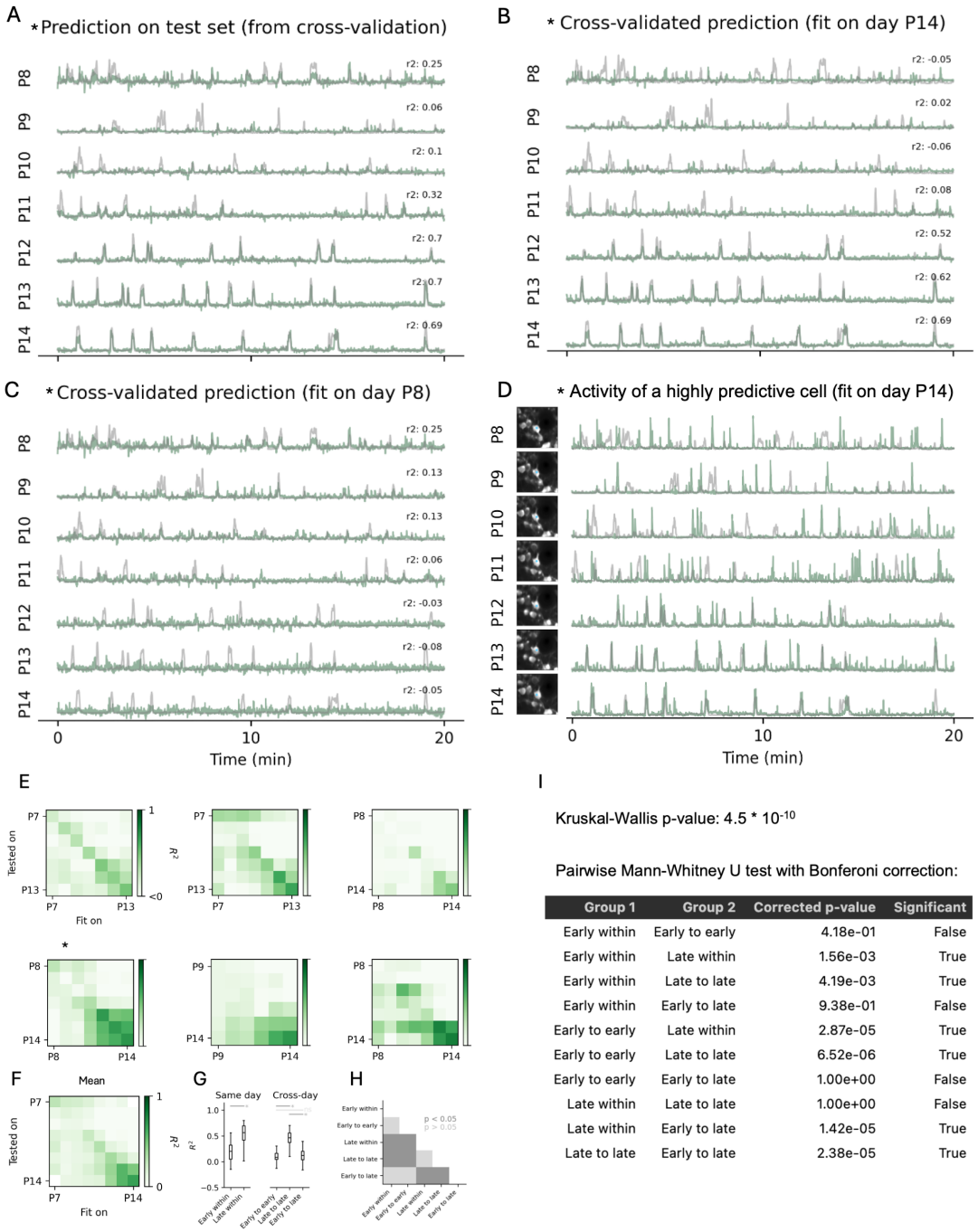
Group 1	Group 2	Corrected p-value	Significant
Early within	Early to early	4.42e-04	True
Early within	Late within	3.25e-03	True
Early within	Late to late	9.19e-05	True
Early within	Early to late	4.70e-12	True
Early to early	Late within	1.00e+00	False
Early to early	Late to late	1.00e+00	False
Early to early	Early to late	1.71e-12	True
Late within	Late to late	2.25e-02	True
Late within	Early to late	2.90e-08	True
Late to late	Early to late	4.17e-05	True

L Kruskal-Wallis p-value: $1.3 * 10^{-19}$

Pairwise Mann-Whitney U test with Bonferoni correction:

Group 1	Group 2	Corrected p-value	Significant
same day early	early to early	2.83e-07	True
same day early	same day late	8.48e-02	False
same day early	late to late	2.96e-02	True
same day early	early to late	1.18e-11	True
early to early	same day late	7.05e-07	True
early to early	late to late	4.15e-01	False
early to early	early to late	9.16e-04	True
same day late	late to late	1.43e-03	True
same day late	early to late	8.96e-09	True
late to late	early to late	1.96e-04	True

Supplementary Figure 6. **Stability of neural activity statistics for all mice in the dataset.** **A**: Scatter plots of FC values for all combinations of recordings for the example mouse (related to Fig. 5B). **B**. FC stability matrices for all mice in the dataset. Asterisk denotes the example mouse (related to Fig. 5D). **C**. Mean FC stability matrix averaged across all mice in the dataset. **D**. Box plots of the distribution of individual FC stability values in different conditions pooled across mice (same as Fig. 5E). **E**. Statistical significance of pairwise comparisons of data from D. **F-J**: Same as A-E but for stability of calcium event rates. **K**: P values for all statistical comparisons between FC stability of different developmental epochs (related D and E in this figure and Fig 6E). **L**: Same as K but for calcium event rate stability.



Group 1	Group 2	Corrected p-value	Significant
Early within	Early to early	4.18e-01	False
Early within	Late within	1.56e-03	True
Early within	Late to late	4.19e-03	True
Early within	Early to late	9.38e-01	False
Early to early	Late within	2.87e-05	True
Early to early	Late to late	6.52e-06	True
Early to early	Early to late	1.00e+00	False
Late within	Late to late	1.00e+00	False
Late within	Early to late	1.42e-05	True
Late to late	Early to late	2.38e-05	True

Supplementary Figure 7. **Relationship between spontaneous behaviour and neural activity for all mice in the dataset. A**. Cross-validated same day prediction for all recording days in the example mouse (related to Fig. 6B and C). **B**: Prediction of a decoder fit on P14 across all days (related to Fig. 6F and G). **C**. Prediction of a decoder fit on P8 across all days (related to Fig. 6F and G). **D**. Example dF/F trace of a highly predictive neuron for a model fit at P14. Notice that the activity closely corresponds to the movement trace also on two earlier days (P12 and P13). **E**. Full matrices of R^2 values for all combinations of train and test recordings in each mouse. Example mouse (*). **F**. Mean R^2 matrix averaged across all mice in the dataset. **G**. Box plots of the distribution of individual R^2 values in different conditions pooled across mice (same as Fig. 5E). **H**. Statistical significance of pairwise comparisons of data from G. **I**: P values for all statistical comparisons of decoder performance for all comparisons (related D and E in this figure and Fig 7. H).

Supplementary Information: Videos

Supplementary Video 1: Example of videography and 2-photon calcium imaging data acquired throughout the second postnatal week. Note the substantial growth of the mouse, the development of the whisker pad and opening of the eyes; and the loss of synchrony and spatial organisation in the calcium imaging data. For visualisation purposes, the videos were manually aligned using Fiji.

Supplementary Information: Software

Interactive visualisation and curation using the Track2p GUI

To facilitate the ease of as well as to visually evaluate and curate the quality of Track2p tracking, we've designed an intuitive graphical interface (GUI) that makes it easy for users to launch the Track2p algorithm and interact with the outputs. Track2p GUI allows the users to perform cell tracking as well as generate some preliminary visualisations of cell activity without requiring any programming skills on behalf of the user (Fig. S1A). Additionally, we've made the algorithm and the GUI simply installable via pip, and provide plenty of documentation on the project's website: https://track2p.github.io/.

Track2p provides several data visualisation functionalities. Firstly, while running the algorithm, a set of figures is generated that can be used to monitor the progress at each consecutive step, as well as to provide a visual overview of tracking quality, for example based on a visual assessment of registration quality, IoU histograms or zoomed-in images of example tracked cells. The GUI itself provides an interactive view of the FOV overlaid with ROIs tracked across all days (Fig. S1B, similar to Fig. 3D). The user can then select a cell from this window that is then displayed in a zoomed in view across all days (Fig. S1B, similar to Fig. 3C). The visualised cell can be selected interactively by selecting it from contours displayed over the whole FOV. Based on these visualisations, users can also assess the tracking performance and, if necessary, manually curate the results by manually identifying cases of incorrect tracking (Fig. S1E). Under conditions shown in this paper manual curation should not be strictly necessary (see Fig. 4 for evaluation without manual curation), however this will depend greatly on the particularities of the dataset used.

Additionally, the GUI also provides activity visualisation, plotting the activity time series of the selected ROI for each recording (Fig. S1C, similar to Fig. S4B), while also offering the visualisation of the activity of the whole population across all days by generating raster plots of neuronal activity (Fig. S1F, similar to Fig. 5A and Fig. S4A). These visualisations can allow the user to visually inspect changes to activity statistics at the single cell level as well as changes to population level phenomena such as synchronisation (assemblies) and sequences at the level of all tracked neurons. Raster sorting capabilities can also give a visual insight into the stability of these phenomena across recordings. Currently the user can choose between PCA (Principal Component Analysis) or t-SNE (t-distributed Stochastic Neighbor Embedding) techniques to re-sort the rows of the raster based on dimensionality reduction. These methods aim to sort cells with similar fluorescence traces in adjacent rows of the raster plot, visualizing co-varying activity patterns at the population level. Additionally, the user can choose between individual sorting and sorting that is preserved across days. Individual sorting better shows these population phenomena for each specific day, however in this case the correspondence between the cells across days is lost due to independent sorting. On the other hand, across-day sorting allows the user to focus on one day to visualize the stability of these population phenomena and see if they are preserved across other days, since the sorting of the neurons remains consistent across sessions.