



**HAL**  
open science

## Automatic monitoring of neural activity with single-cell resolution in behaving Hydra

Alison Hanson, Raphael Reme, Noah Telerman, Wataru Yamamoto,  
Jean-Christophe Olivo-Marin, Thibault Lagache, Rafael Yuste

► **To cite this version:**

Alison Hanson, Raphael Reme, Noah Telerman, Wataru Yamamoto, Jean-Christophe Olivo-Marin, et al.. Automatic monitoring of neural activity with single-cell resolution in behaving Hydra. Scientific Reports, 2024, 14 (1), pp.5083. 10.1038/s41598-024-55608-2 . pasteur-04626604v2

**HAL Id: pasteur-04626604**

**<https://pasteur.hal.science/pasteur-04626604v2>**

Submitted on 26 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN

## Automatic monitoring of neural activity with single-cell resolution in behaving *Hydra*

Alison Hanson<sup>1,2,4</sup>✉, Raphael Reme<sup>3,4</sup>, Noah Telerman<sup>1,4</sup>, Wataru Yamamoto<sup>1</sup>, Jean-Christophe Olivo-Marin<sup>3</sup>, Thibault Lagache<sup>3,5</sup> & Rafael Yuste<sup>1,5</sup>

The ability to record every spike from every neuron in a behaving animal is one of the holy grails of neuroscience. Here, we report coming one step closer towards this goal with the development of an end-to-end pipeline that automatically tracks and extracts calcium signals from individual neurons in the cnidarian *Hydra vulgaris*. We imaged dually labeled (nuclear tdTomato and cytoplasmic GCaMP7s) transgenic *Hydra* and developed an open-source Python platform (TraSE-IN) for the Tracking and Spike Estimation of Individual Neurons in the animal during behavior. The TraSE-IN platform comprises a series of modules that segments and tracks each nucleus over time and extracts the corresponding calcium activity in the GCaMP channel. Another series of signal processing modules allows robust prediction of individual spikes from each neuron's calcium signal. This complete pipeline will facilitate the automatic generation and analysis of large-scale datasets of single-cell resolution neural activity in *Hydra*, and potentially other model organisms, paving the way towards deciphering the neural code of an entire animal.

A major goal in neuroscience is to record the activity of all neurons in a behaving animal with single-neuron resolution to capture the emergent properties of neural circuits<sup>1</sup>. This remains a significant technological challenge in complex systems such as rodents<sup>2</sup>, monkeys<sup>3</sup>, and humans<sup>4</sup> in which whole brains can currently only be recorded with low spatiotemporal resolution. To overcome these challenges, neuroscientists have used simpler systems, including zebrafish<sup>5–8</sup>, *Drosophila*<sup>9,10</sup>, and *C. elegans*<sup>11,12</sup>, which contain far fewer neurons but exhibit more limited behavioral repertoires. However, even these simpler systems are still quite complex and present their own unique imaging challenges. Here, we have chosen to focus on a small freshwater cnidarian, *Hydra vulgaris*, with one of the simplest known nervous systems<sup>13–16</sup>. *Hydra* has key advantages, including a transparent body, small size, a limited number of neurons belonging to only a dozen cell types<sup>17</sup>, distributed nerve net lacking a true brain or ganglia, remarkable regenerative capacity<sup>18</sup>, and a behavioral repertoire that can be automatically quantified using machine learning<sup>19</sup>. Unfortunately, *Hydra* is also a highly deformable organism and is therefore one of the most difficult test cases for identifying and tracking single neurons in a moving animal.

Despite much progress, monitoring of single-cell resolution whole-body neural activity in a behaving animal remains a significant challenge, even among simpler systems. Indeed, even if high-resolution imaging of whole-body neural activity can be achieved, identifying and tracking individual neurons in a moving animal over time remains a major hurdle. Moreover, the robust prediction of individual spikes from each neuron's calcium signal also requires advanced processing of tracked calcium signals. Several approaches have been employed to facilitate the tracking of single neurons, including fixing the head of the animal to the microscope<sup>8–10,20</sup>, moving the microscope to keep the brain of the freely behaving animal in the center of the objective<sup>7,11</sup>, or immobilizing the entire animal<sup>5,21,22</sup>. In the first two cases, neural imaging is limited to the head only and does not allow organism-wide neural recording while behaving. In the last case, neural imaging is limited solely to immobilized, non-behaving animals.

New approaches in transgenic animals, hardware, and software have been developed to overcome these challenges. For example, a new transgenic *C. elegans* line was generated that expresses two fluorophores in the nuclei of neurons: a calcium-insensitive fluorophore (red fluorescent protein) and a calcium-sensitive fluorophore (GCaMP), which allows continuous visualization of the location of the neurons even when they are inactive<sup>23</sup>.

<sup>1</sup>Department of Biological Sciences, Neurotechnology Center, Columbia University, New York, NY, USA. <sup>2</sup>Department of Psychiatry, New York State Psychiatric Institute, Columbia University, New York, NY, USA. <sup>3</sup>UMR3691, Biologie Analytique Unit, Institut Pasteur, Université Paris Cité, CNRS, Paris, France. <sup>4</sup>These authors contributed equally: Alison Hanson, Raphael Reme and Noah Telerman. <sup>5</sup>These authors jointly supervised this work: Thibault Lagache and Rafael Yuste. ✉email: alison.hanson@nyspi.columbia.edu

**Figure 1.** Overview of the pipeline. (I) The pipeline includes generation of a novel dually labeled (nuclear tdTomato and cytoplasmic GCaMP7s) transgenic *Hydra* imaged on a custom built simultaneous two-color spinning disc confocal microscope, allowing high spatiotemporal resolution imaging of behaving animals. (II) Each nucleus is segmented using the StarDist deep learning algorithm. (III) Each segmented nucleus is then tracked over each frame in the movie by first using probabilistic temporal linking to form short tracklets, which are then stitched together using a cost-based method to determine the final full-length tracks. (IV) A region of interest (ROI) is then drawn around the center position of each nucleus and a sub-ROI method is used to detect the most likely corresponding neuronal cell body from which to extract GCaMP7s signal. (V) Both the extracted nuclear tdTomato and cytoplasmic GCaMP7s signals are used for signal processing to account for motion artifacts. Independent component analysis, followed by detrending and smoothing is run to remove imaging artifacts from the GCaMP7s signal. Neurons' spikes are predicted from the pre-processed calcium signal using FOOPSI. The final output is a raster plot with the spike predictions from individual neurons in a behaving animal. See text for further details.

These dually labeled animals make it significantly easier to detect and track individual neurons over time while the worm behaves. Novel imaging modalities include light sheet microscopy approaches, such as swept confocally aligned planar excitation (SCAPE)<sup>24</sup>, which has allowed fast whole-organism imaging with higher signal-to-noise and less photobleaching than widefield microscopy, and less photobleaching but poorer signal-to-noise than confocal microscopy approaches. On the software side, the advent of deep-learning algorithms has drastically improved the accuracy of single cell segmentation, a fundamental prerequisite for the tracking of neuronal activity with calcium fluorescent indicators. However, the monitoring of neuronal activity requires their robust tracking over time, which consists of pairing the neurons segmented at different time points into coherent tracks<sup>25</sup>. Tracking is a difficult task due to animal motion, occlusions, and intermittent activity of calcium probes. It is still performed manually or semi-automatically in many systems. Many advanced algorithms have also been developed for the accurate estimation of neuronal activity (spikes typically) from recorded calcium activity<sup>26–30</sup>. However, all these methods usually require signal pre-processing to attenuate imaging noise and remove non-linear fluorescence baseline that could bias the estimation of neurons' spikes from fluorescent calcium traces. Therefore, there is a need for a robust and versatile tool that can track individual neurons in deforming bodies, accurately estimate the neuronal activity from calcium fluorescence, and could also be applied to more simple cases where animals are fixed or minimally moving.

Here, we developed TraSE-IN, an end-to-end pipeline (Fig. 1) that allows the automatic tracking and signal extraction of individual neurons in behaving *Hydra*. This novel pipeline entails both hardware and software improvements over existing tools to create a versatile, robust, and accurate way to automatically detect and track individual neurons in a highly deformable organism while behaving. These improvements include the generation of a two-color transgenic *Hydra* and imaging method that allows the simultaneous recording of both nuclei and neural activity, and the development of an open-source Python platform (<https://github.com/raphaelreme/trase-in>). This platform consists of a tracking module to monitor the calcium activity of individual neurons in the behaving animal, and a series of signal processing modules for the robust extraction of spikes from calcium traces, paving the way for the analysis of the functional correlates underpinning animal behavior.

Altogether, hardware and software developed in the pipeline will significantly enhance our ability to correlate individual neural activity with behavior under different conditions, analyze large data sets over longer time scales, and bring us closer towards the goal of cracking the neural code of an entire animal.

## Results

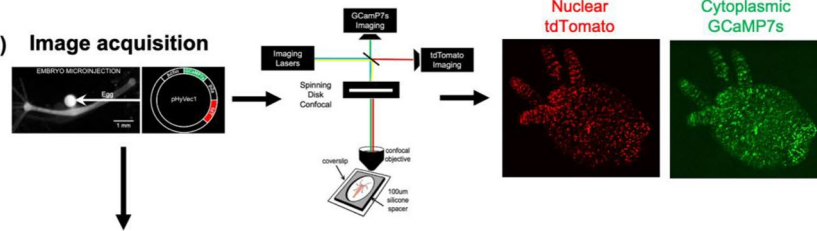
### Imaging dually labeled transgenic *Hydra*

A major hurdle in recording neural activity in a behaving *Hydra* is the faithful tracking of individual neurons over time as the animal moves. The initial transgenic *Hydra* created in our lab expressed GCaMP6s in all neurons, which allowed wide-field calcium imaging of whole animal neural activity while behaving<sup>16</sup>. However, each neuron had to be painstakingly tracked and analyzed manually. Despite the development of advanced tracking software (EMC<sup>2</sup>) that iteratively estimated the local deformation of the animal using spatial key points, and approximated the position of neurons when they were not firing and did not emit fluorescent signals<sup>25</sup>, tracking errors accumulate over time, which hinders the analysis of neural activity over long time-lapse sequences. To overcome this bottleneck, we developed a novel transgenic *Hydra* line that expresses two fluorescent indicators in its neurons: cytoplasmic GCaMP7s and nuclear tdTomato (Fig. 2, see Methods section for more details). GCaMP7s is a genetically encoded calcium indicator (GECI) that allows visualization of the calcium dynamics of each neuron<sup>31</sup> and tdTomato is a calcium insensitive molecule<sup>32</sup> used for single cell labelling. tdTomato is a high brightness indicator with emission in the red end of the spectrum (581 nm) ideal for use with the green emission spectrum of GCaMP7s (511 nm) with minimal cross talk between the fluorophores. Our transgenic strain allows for the spatial tracking of the continuous signal from tdTomato, facilitating the tracking of individual neurons, and the extraction of GCaMP7s signal whenever neurons fire.

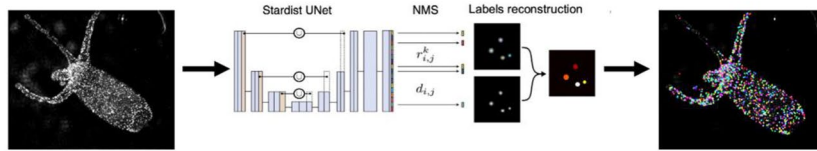
To confirm all neurons are indeed labeled in the double transgenic line, fixed animals were co-stained with anti-*Hydra* cadherin antibody that labels *Hydra* neurons<sup>33</sup> and anti-tdTomato antibody (Figure S1). In these animals, 96.53% ± 2.49% (std) cadherin positive cells also expressed nuclear tdTomato. Because some neurons might not have stained with anti-cadherin or anti-tdTomato during the immunostaining protocol, we conclude that essentially all neurons in the transgenic animal are expressing nuclear tdTomato. To confirm that each neuron that expresses nuclear tdTomato also expresses cytoplasmic GCaMP7s in the transgenic animals, calcium imaging movies were scored for the presence of nuclear tdTomato in active GCaMP7s neurons. This analysis showed that

**PIPELINE OVERVIEW**

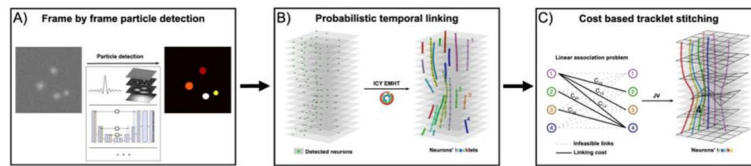
**I) Image acquisition**



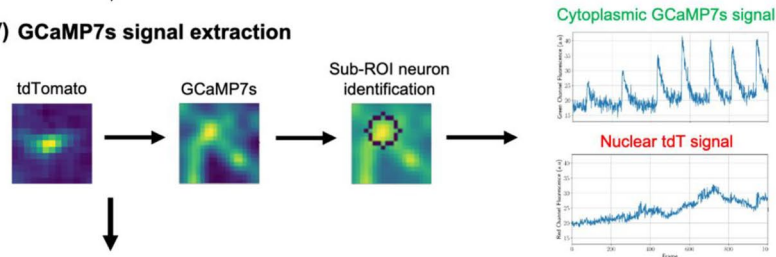
**II) tdTomato nuclei segmentation**



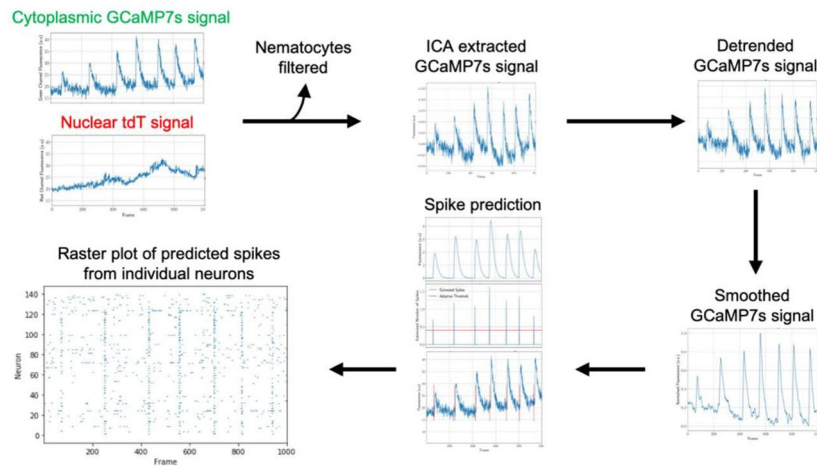
**III) tdTomato nuclei tracking**

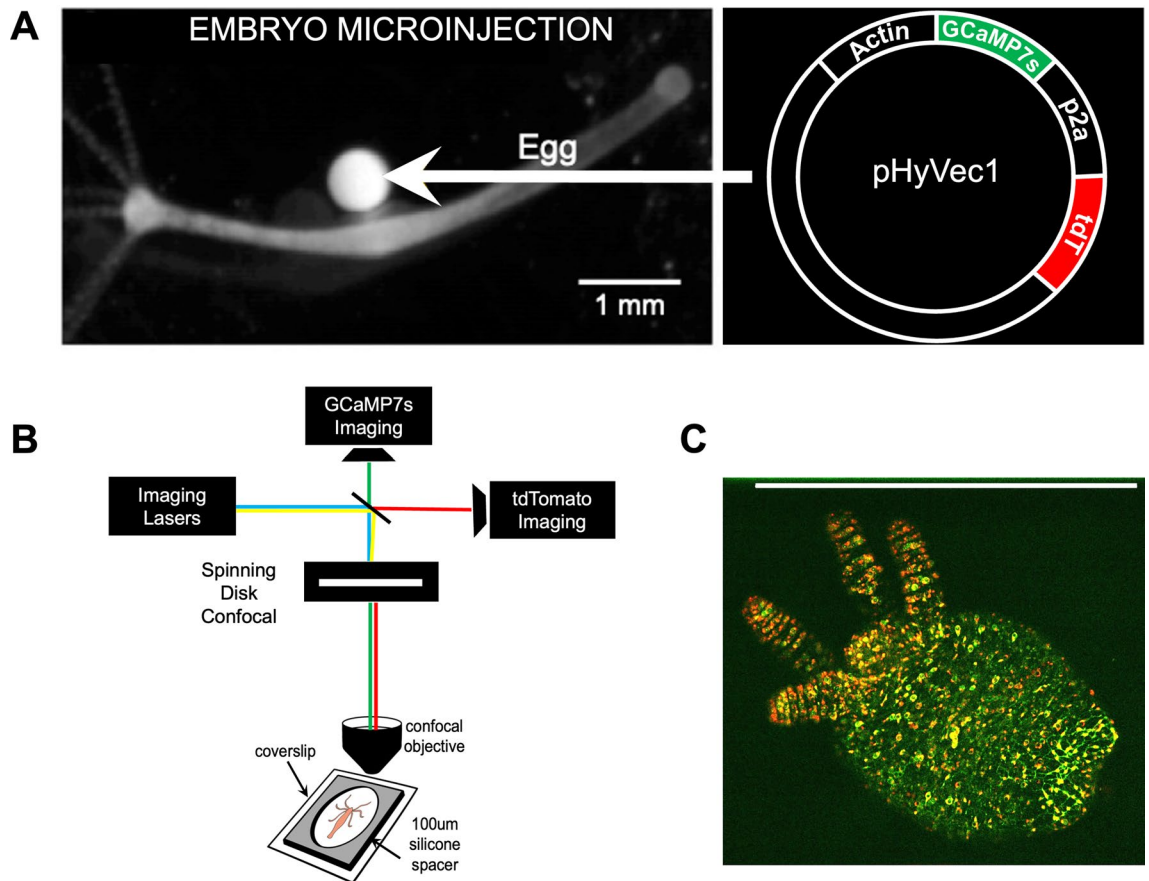


**IV) GCaMP7s signal extraction**



**V) GCaMP7s signal processing and spike prediction**





**Figure 2.** Generation of transgenic animals and imaging setup. (A) Transgenic animals were created by injecting the plasmid (right) into a fertilized *Hydra vulgaris* egg (left). Image on left modified with permission from<sup>16</sup>. (B) Dually labeled transgenic animals were placed in *Hydra* media in the center of a 100  $\mu\text{m}$  silicon spacer sandwiched between two coverslips and imaged with a custom-built simultaneous two-color (488 nm and 571 nm) spinning disc confocal microscope. (C) Representative image of a dually labeled (nuclear tdTomato and cytoplasmic GCaMP7s) *Hydra* imaged with the simultaneous two-color spinning disc confocal microscope setup shown in (B). Scale bar: 1 mm.

100% of GCaMP7s positive neurons also express nuclear tdTomato (Figure S1). Given that nearly all neurons co-stain with anti-cadherin and anti-tdTomato, and that all GCaMP7s neurons express nuclear tdTomato, we conclude that essentially all neurons in the transgenic animal expressed the dual cytoplasmic GCaMP7s-nuclear tdTomato plasmid. In addition to neurons, the interstitial cell lineage of *Hydra* produces germ cells, gland cells, and nematocytes<sup>34–37</sup>. Indeed, these cell types could also express the dual reporter plasmid in the transgenic animals, although in our experiments the only cell type clearly expressing the transgenes were nematocytes (Figure S1). The signal collected from these non-neuronal cells can be removed computationally in the analysis pipeline, if desired (see section “Measuring the activity of individual neurons”).

The novel dually labeled transgenic *Hydra* was prepared for imaging as previously described<sup>16</sup>. In this mounted preparation, *Hydra* is placed in media between two coverslips separated by a 100  $\mu\text{m}$  spacer, creating a single-plane imaging object while allowing the animal to exhibit a variety of behaviors<sup>38</sup>. The animal was then imaged on a custom-built spinning disc confocal microscope that allows simultaneous two-color imaging of both GCaMP7s and tdTomato activity (see Methods). Spinning disc confocal microscopy was adopted due to its superior performance over widefield, light sheet, and two-photon microscopy approaches in terms of its ability to perform simultaneous two-color recordings with sufficient spatiotemporal resolution and optimal signal-to-noise ratio. The fastest behaviors in *Hydra* occur at 5 Hz. Thus, to minimize motion blur in a rapidly contracting *Hydra*, images were acquired at 10 Hz. Recorded images were processed and adjusted for brightness and contrast using ImageJ analysis software.

### Segmenting and tracking individual neurons

Despite the calcium insensitiveness of nuclear tdTomato, the motion and deformation of the animal, the high density of neurons in some parts of the animal (peduncle and hypostome typically), and the potential occlusion or disappearance of neurons from the microscope focal plane due to axial motion of tissues, called for the development of a state-of-the-art pipeline for segmenting and tracking neurons’ nuclei.

With the exception of *C. elegans*, where neurons’ positions along time can be mapped and compared to an initial known atlas of positions after having corrected the deformation of the worm<sup>39</sup>, monitoring the calcium

activity of individual neurons in other model systems requires the development of standard single-particle-tracking (SPT) that consists of two distinct steps: first, the fluorescent spots of individual particles (neurons) are automatically segmented, and the positions of detected neurons are extracted in each time frame of the movie. The second step of SPT algorithms is the robust linking of extracted positions into coherent tracks representing the trajectories of individual particles.

The first module of the TraSE-IN platform consists of a tracking pipeline, that we call ByoTrack (<https://github.com/raphaelreme/byotrack>) (Fig. 3), which is a series of functions enabling the automatic detection of neurons' nuclei and their tracking along the time-lapse sequence.

### Segmentation of nuclei

In fluorescence microscopy, the objects of interest (neurons here) appear as spots brighter than the noisy background. When the signal-to-noise ratio (SNR) is low, objects' spots can be hardly distinguishable. Therefore, elaborated detection algorithms have been developed over the years<sup>40</sup>. Most of the analytical detection methods consist of image denoising (e.g., Gaussian smoothing, wavelet denoising), signal enhancement (e.g., background subtraction with top-hats or h-dome), and signal thresholding. Analytical methods have proven to be robust and fast for moderate to high signal-to-noise ratios. However, they usually require tweaking a few user-defined parameters, and, more importantly, they usually fail at separating closely apposed spots.

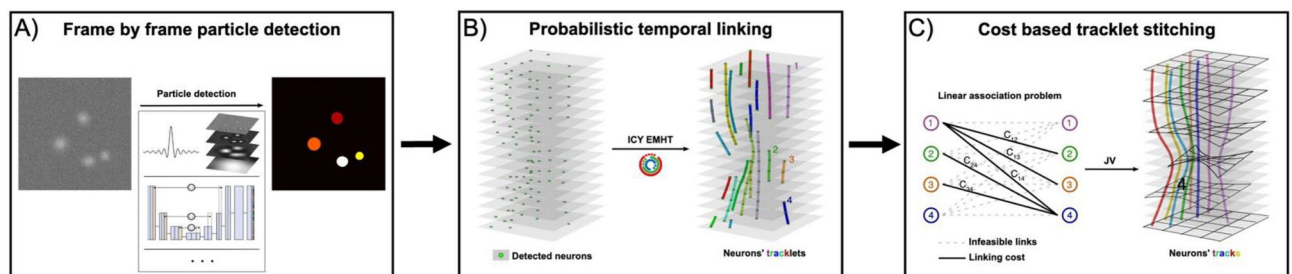
Since their recent introduction, deep-learning (DL) approaches have demonstrated a much-increased accuracy and robustness for detecting cells in noisy and cluttered images. Among DL approaches, StarDist is particularly well suited for segmenting and separating roundish objects such as cell nuclei as it uses a star-convex representation of objects' contours and a non-maximum-suppression (NMS) post-processing for separating touching cells<sup>41</sup>. However, DL approaches usually require manual annotation of data for model training.

In the first part of ByoTrack, we implemented different algorithms for neurons' segmentation and detection (Fig. 3A). In particular, we used an algorithm that executes the DL method StarDist, the training of the model being performed using the open-source Python code of StarDist<sup>41</sup>. We also implemented an analytical method based on undecimated wavelet transform and statistical thresholding of the image<sup>42</sup>. The wavelet thresholding approach has demonstrated a good performance on images with medium to low SNR<sup>43</sup>.

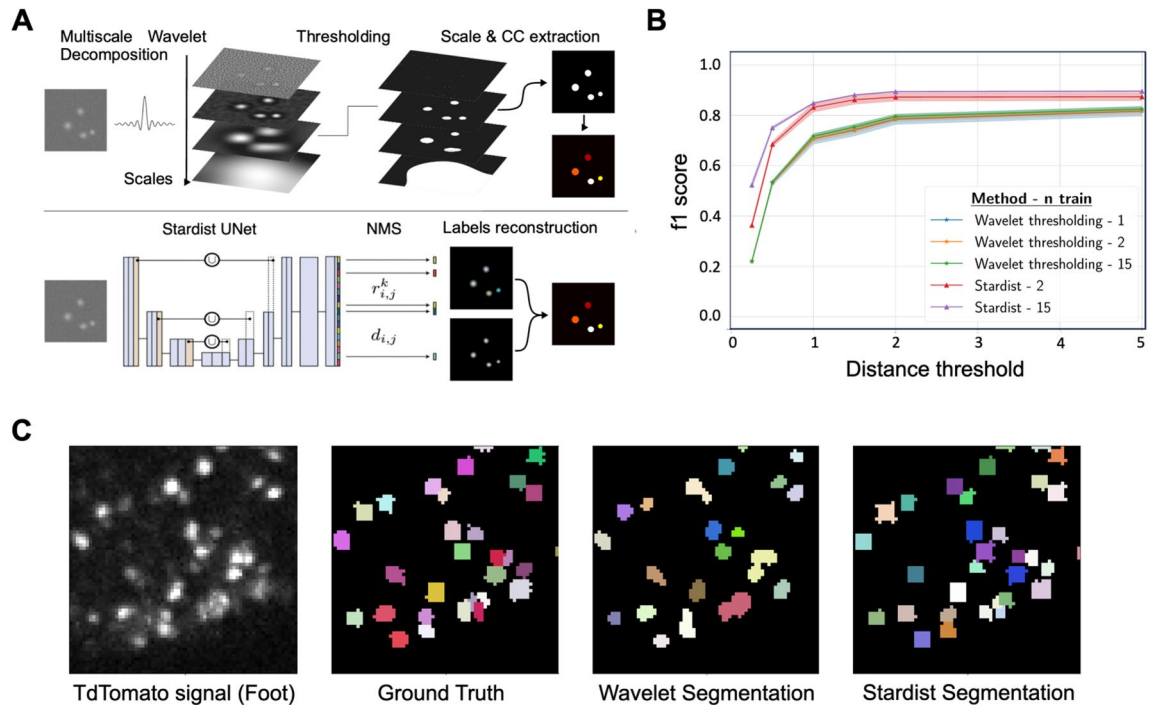
Using tdTomato images extracted from two-color time-lapse sequences, we trained a StarDist model and compared its performance with the analytical wavelet method (Fig. 4A). As manual annotation of cell nuclei can be tedious and time-consuming, we trained a StarDist model for an increasing number of tdTomato images (from 2 to 15, corresponding to a maximum of 9422 neurons' nuclei, see Methods) and measured how model accuracy increased with the size of the training dataset (Fig. 4B). In this manner, we could determine how to optimally balance model accuracy and training.

Wavelet thresholding is specifically designed for the detection and localization of fluorescent spots, rather than for their accurate segmentation. Thus, the comparison metrics (f1-score) between detected spots was computed using a distance threshold  $d$  between the center-of-mass of detected objects, rather than the commonly used intersection-over-union (IOU) between segmented spots. In other words, we considered that a predicted spot's detection was valid if its center-of-mass was at a distance less than  $d$  from the center-of-mass of a ground truth detection. Moreover, using the position of detected nuclei rather than their shape to assess the accuracy of the detection method is more meaningful for tracking, as we only use the position of each detected instance.

Using a distance threshold  $d = 1$  pixel we obtained, for trained StarDist, a f1-score of 84.9% (number of predicted spots  $n_{pred} = 2946$  with  $t_p = 2656$  true positives and  $f_p = 290$  false positives). The number of true instances (cell nuclei) was  $n_{true} = 3283$  over 5 testing images, the number of false negatives being equal to  $f_n = 627$  (Fig. 4B). The hyperparameters of the analytical wavelet thresholding method (scale and threshold) were tuned (grid search) by maximizing the method accuracy over 1 to 15 annotated images. With the same distance threshold  $d = 1$  pixel, we obtained a decreased f1-score of 72.0% ( $n_{pred} = 2747$  with  $t_p = 2171$  and  $f_p = 576$ ).



**Figure 3.** Tracking pipeline. The implemented tracking pipeline can be divided into three main steps: (A) the automatic detection of nuclei spots in tdTomato sequences using analytical methods such as wavelet transform and thresholding or deep learning methods such as StarDist detection, followed by (B) the robust linking of segmented nuclei in coherent tracklets with a state-of-the-art SPT method. Here, the probabilistic method enhanced Multiple-Hypothesis-Tracking (eMHT) method is called from the Icy software. (C) Finally, tracklets corresponding to the same neuron are stitched using a cost-based method. The cost here corresponds to the minimal distance between the forward- and backward-propagated positions of undetected nuclei. The optimal association between tracklets that minimizes the global cost of association is a linear association problem that we solved with the Jonker-Volgenant (JV) algorithm.



**Figure 4.** Comparison of a deep-learning approach (StarDist) versus an analytical approach (wavelet transform) for segmenting nuclei in tdTomato images. **(A)** The two methods compared for detection. On top, the analytical wavelet transform approach consists of decomposing the image with multiscale *a-trou* wavelets, before applying a statistical threshold on wavelet coefficients to denoise the image and extract the significant fluorescent signals (connected components CC). Wavelet decomposition favors the extraction of spotty signals, which makes it well-suited for the segmentation of fluorescent roundish objects such as cell nuclei. On the bottom, StarDist deep learning (DL) segmentation consists of applying a trained U-net convolutional network, followed by a non-maximum suppression (NMS) to segment individual nuclei in the image. **(B)** We compared the accuracy (f1-score, mean  $\pm$  standard error) of wavelet and DL segmentation algorithms applied to test images ( $n = 5$  images) for different distance thresholds and number of images used for StarDist training (2 (red) or 15 (purple)), or for the calibration wavelet method's parameters (scale and threshold: 1 (blue), 2 (orange) or 15 (green)). **(C)** A representative image of tdTomato nuclei in the foot of the animal, ground truth segmentation, wavelet segmentation, and StarDist segmentation of those same nuclei are shown.

Number of false negatives was equal to  $f_n = 1112$ . Therefore, the false negatives and false positives were almost two-fold higher using an analytical wavelet-thresholding method compared to the trained DL method. A representative image showing superior segmentation of StarDist over wavelet thresholding of nuclei in the foot of the animal is shown in Fig. 4C. In particular, StarDist was able to separate close neurons whereas the wavelet thresholding cannot.

We highlight that StarDist performances were nearly as good when trained with only 2 annotated images (one image used for training, and one used for validation) compared to 15 annotated images. Therefore, we conclude that StarDist can be easily trained for other imaging set-ups or conditions and showed a much-increased performance compared to more standard, analytical detection algorithms.

### Probabilistic tracking and tracklet stitching

To monitor the calcium activity of individual neurons, it is then necessary to link detected nuclei in the different time frames into coherent tracks. However, tracking is not trivial because of the errors made during the detection of particles (neurons) in each frame, the potential occlusion or undetectability of particles over several frames, and the difficult disentanglement of individual trajectories when particles are densely packed. Many elaborated algorithms have been developed in the last two decades<sup>44</sup>. These algorithms are either based on the global distance minimization (GDM) between the set of particles' positions in successive time frames or use a probabilistic framework to compute the most likely trajectories linking the different positions given a model for particles' motion. While requiring a higher computational load, the latter probabilistic framework has demonstrated an increased accuracy and robustness than GDM methods, especially in cluttered environments<sup>45</sup>. The second step of our tracking pipeline consists of linking neurons' detections with a state-of-the-art probabilistic algorithm (eMHT, Fig. 3B). This tracking algorithm is implemented in Icy, an open-source image analysis platform (<https://icy.bioimageanalysis.org/>)<sup>46</sup>. ByoTrack directly executes this step in Icy and gets the resulting tracks for post-processing and signal analysis.

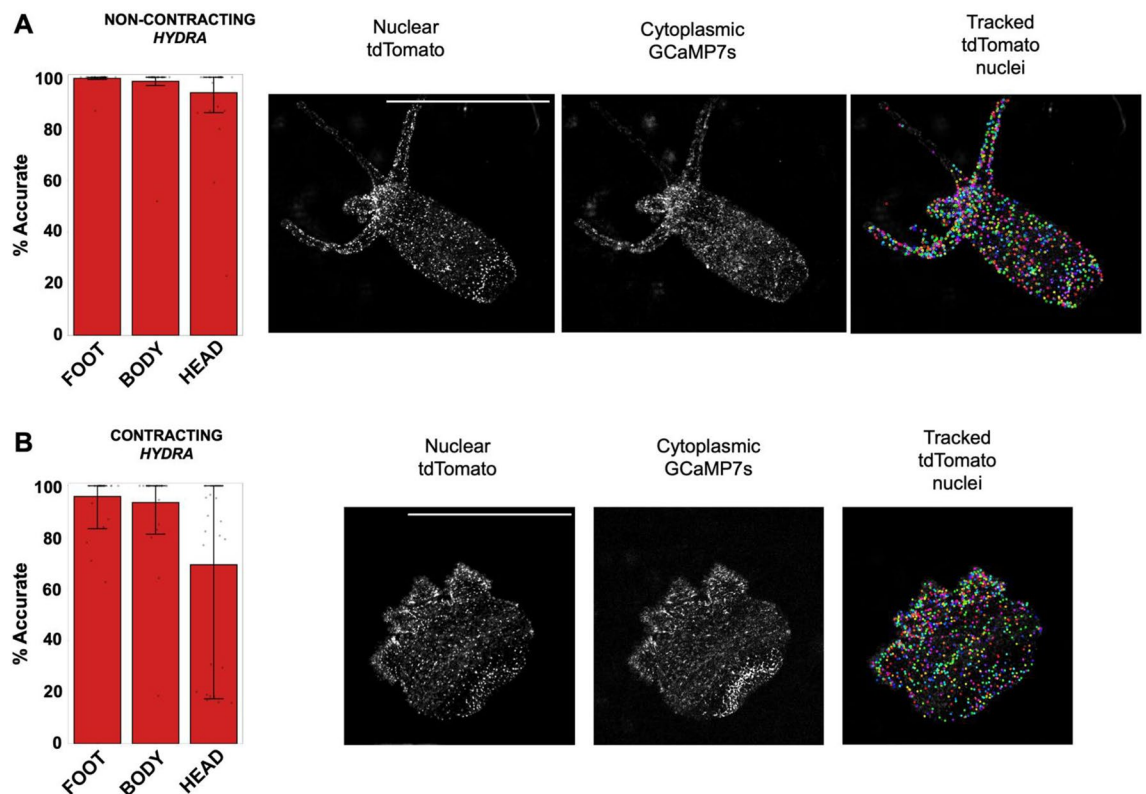
Due to particles' occlusion, moving out of focus, or intermittent detectability, standard tracking methods, such as eMHT, tend to produce multiple small tracks ("tracklets") for the same particle, hindering the long-term

identification of particles and, in the case of neurons, the monitoring of individual spiking activities. Indeed, when a particle is not detected over a few frames (more than  $\sim 5$  typically), SPT algorithms prematurely end tracks, and start a new track when the particle is re-detected.

To stitch the tracklets that represent the same neuron, we implemented the cost-based algorithm (EMC<sup>2</sup>)<sup>25</sup> that infers the positions of undetected neurons from neighboring tracklets (Fig. 3C). Briefly, when a neuron remains undetected over several frames, the forward position after the premature termination of the neuron's tracklet is estimated using a thin-plate-spline interpolation between neighboring tracklets. On the other hand, when a new track begins, the backward position of the tracked neuron is also estimated. Then, to stitch tracklets and obtain entire neurons' tracks, a cost matrix is computed between all the tracklets using the minimal distances between the forward and backward-estimated positions, and the corresponding linear association problem is solved with the Jonker-Volgenant algorithm<sup>47</sup>.

To measure the accuracy of our tracking pipeline on an experimental dataset, we tracked neurons' nuclei using ByoTrack in  $N = 3$  different animals (Fig. 5). The rhythmic longitudinal contractions of *Hydra* can drastically hinder the robust tracking of individual neurons. Indeed, the speed of contraction motion can surpass the acquisition rate of the movie (10 Hz typically), blurring the images and making the detection and tracking of nuclei spots difficult. Thus, for each animal, we determined two time-windows of  $T = 1,000$  frames where the animal contracts or not. For each of these six extracted movies (2 movies per animal, [https://datadryad.org/stash/share/vWz5obHbrC\\_yWiThtJG8V7mWTcjfkAdiIN1UVOURbrA](https://datadryad.org/stash/share/vWz5obHbrC_yWiThtJG8V7mWTcjfkAdiIN1UVOURbrA)), two independent operators chose randomly  $n = 15$  individual neurons throughout the *Hydra* body (5 neurons in the animal's peduncle (foot), central body, and hypostome (head)) and measured, for each neuron, the percentage of frames where it is correctly tracked. Obtained tracking accuracies are reported in Fig. 5 (mean  $\pm$  standard error),  $n = 30$  neurons per localization (foot, body, and head) and animal state (contracting or not). We observe a very good accuracy ( $> 90\%$ ) in all parts of the animal, except for the head of contracting animals where the accuracy drops to 70%. This local drop is due to the fast tentacle retraction and motion of the head of the animal during contraction cycles.

Altogether, these results indicate that the ByoTrack pipeline is sufficiently versatile and robust to track individual neurons in highly deforming *Hydra*. In particular, the use of a DL segmentation algorithm should be preferred to more standard approaches because it does not require extensive training, while showing much improved accuracy. Except for the head of the animal that moves faster than the microscope's acquisition rate during contraction phases, the combination of a state-of-the-art probabilistic tracking algorithm (eMHT) and



**Figure 5.** Accuracy of nuclei tracking in non-contracting and contracting *Hydra*. **(A)** Accuracy of tdTomato nuclei tracking in different parts of a non-contracting *Hydra* (left), with representative images of nuclear tdTomato, cytoplasmic GCaMP7s, and tracked tdTomato nuclei (right). **(B)** Accuracy of tdTomato nuclei tracking in different parts of a contracting *Hydra* (left), with representative images of nuclear tdTomato, cytoplasmic GCaMP7s, and tracked tdTomato nuclei (right) (Each bar is the mean over  $n = 30$  tracks, error bars are the 10% and 90% quantiles). Scale bars: 1 mm.



cost-based tracklet stitching applied to DL-segmented nuclei, showed very good performance, paving the way to the monitoring of the calcium activity of individual neurons in behaving animals.

## Measuring activity of individual neurons

### Linking calcium activity to individually tracked nuclei

The robust, long-term tracking of each neuron is a prerequisite for the monitoring of calcium activity, but additional signal processing is required for the robust extraction of neurons' spikes. First, even with perfect tracking of nuclei in the red channel, extracting the GCaMP7s signal is not as simple as mapping these tracks onto the video from the green channel. Indeed, optical and biological limitations cause misalignment between the two channels. Minor differences and imperfections in the optical path along with differing camera orientations cause each wavelength of light to be projected differently onto their respective cameras. Additionally, the indicators themselves are located in different parts of the cell, with tdTomato expressed in the cell's nucleus and GCaMP7s being expressed in the cell's cytoplasm. While some of this misalignment can be corrected with image processing software, the non-uniformity across the field of view is more complicated to correct. Moreover, multiple neurons often occur in close proximity, meaning the simple solution of taking the average fluorescence over a large region of interest (ROI) encompassing the nucleus will often include multiple neurons (Fig. 6A). This can create confounding results where the signals from two neurons are superimposed (Fig. 6B).

To handle this misalignment and extract robust single-neuron fluorescence information, a custom sub-ROI tracking method was developed (Fig. 6) to identify and track the green-channel fluorescence in the area surrounding the red-tracked point. This algorithm starts from the pixel locations of tracked nuclei and extracts a small ROI around these points in the green channel. The algorithm leverages the fact that cells, even when not fluorescing, are brighter than the background. Within the small ROI, the total brightness, proximity to the center, and proximity to the previously tracked neuron are weighted and summed to give a prediction of the location of the neuron in the green channel (see Methods). This algorithm, implemented as a module on the TraSE-IN platform, provides reliable single neuron fluorescence data that can be further processed for spike extraction.

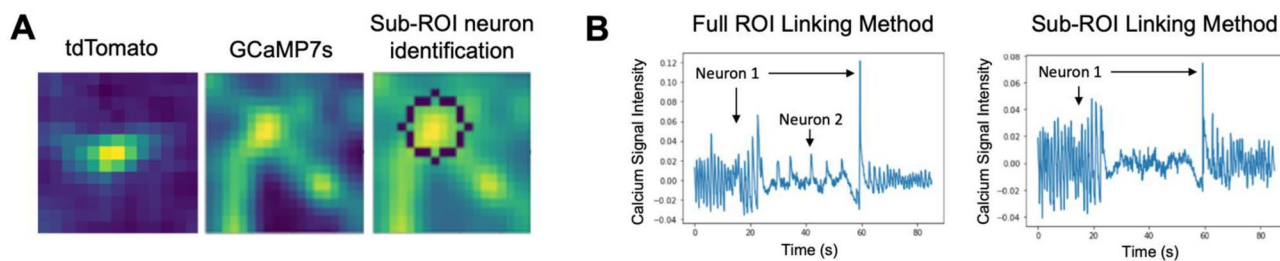
A second major hurdle to monitoring neuronal activity is the extraction of spikes from calcium fluorescence. Indeed, the fluorescent calcium traces are usually corrupted by background signal and noise, and the dynamics of conformational change of calcium indicators is much slower than actual voltage variations. For all these reasons, elaborate deconvolution methods have been developed over the years to robustly extract individual spikes<sup>26–28,48</sup>. These algorithms show good performance on synthetic and experimental datasets but require an input fluorescent signal with good signal-to-noise ratio and no, or moderate, fluctuations of fluorescence baseline. To robustly extract neurons' spikes from fluorescence intensity with these algorithms, we have therefore developed a series of pre-processing steps described hereafter (Fig. 7).

## Signal processing and spike prediction

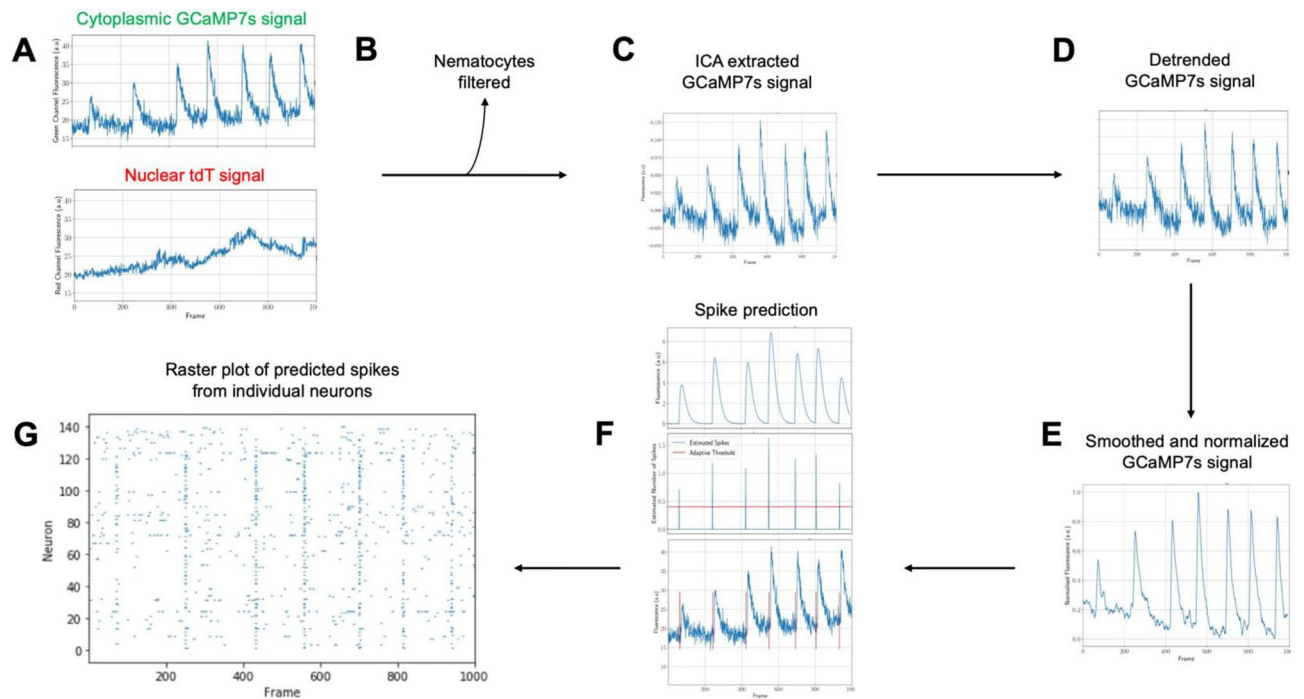
### Removing non-neuronal cells

Due to imperfect transgenic animals, where labeling occurs across i-cell derived lines, non-neuronal cells express tdTomato and GCaMP7s. Most of these cells do not display calcium activity so are simple to remove or ignore. Nematocytes, the stinging cells unique to cnidarians, are an exception. These cells produce very bright calcium signals and will be incorrectly analyzed by spike extraction methods if not removed.

Although little is known about the spiking activity of nematocytes, initial studies<sup>38</sup> indicate they do not spike like most neurons. Once they are activated, their calcium signal remains elevated for long periods of time (30 s to minutes). Thus, once activated, the variations of their fluorescence intensity are mostly due to imaging noise. Leveraging this fact allows filtering of nematocytes, if desired (Fig. 7B), based on their correlation with Gaussian noise (see Methods). It has also been shown that some neurons exhibit similar plateau maintaining calcium signals<sup>49</sup>, which might also be highly correlated with Gaussian noise. Accordingly, an interactive graphical interface in the TraSE-IN pipeline allows the user to visualize each cell that is highly correlated with Gaussian



**Figure 6.** Linking tracked tdTomato nuclei to GCaMP7s neuronal cell bodies. (A) A small region of interest (ROI) is drawn around the center position of the tracked tdTomato nucleus (left). The brightest pixel nearest the center in the corresponding GCaMP7s image (middle) is used to detect the most likely neuron in the full ROI, around which a sub-ROI is drawn (right) and GCaMP7s signal extracted (see text for further details). (B) Example GCaMP7s activity when using the full ROI versus the sub-ROI linking method. In the full ROI method, the signal from two neurons in the same ROI is erroneously reported, whereas in the sub-ROI method only the signal from one neuron is extracted.



**Figure 7.** Schematic of signal processing and spike prediction steps. (A) Raw cytoplasmic GCaMP7s and nuclear tdTomato signals are extracted from individual neurons in each frame. (B) Nematocytes can be filtered based on their correlation with Gaussian noise. (C) Independent component analysis (ICA) is used to extract the motion corrected GCaMP7s signal. (D) The motion corrected GCaMP7s signal is detrended with frequency filtering. (E) The detrended GCaMP7s signals from individual neurons are smoothed and (F) spikes are predicted using FOOPSI with a custom adaptive clustering and thresholding algorithm. (G) The final output is a raster plot of the predicted spikes from each individual neuron.

noise in the timelapse movie and decide whether or not to keep that cell for further downstream analysis (see Methods for further details).

#### Motion artefact removal

Motion artifacts occur when *Hydra* moves axially during imaging. Due to the narrow focal plane of confocal microscopy, and occlusion of light from outside this focal plane, any motion in Z will result in a significant change in intensity, and as *Hydra* is especially deformable, this type of motion occurs somewhat often in behaving animals. Additionally, due to the non-uniformity of *Hydra* motion, these artifacts do not apply to the entire frame at once, making some general motion correction algorithms useless.

Motion artifacts are notoriously challenging to correct algorithmically as they are highly sporadic, and the intensity and duration of their effect varies at random. However, we can leverage our simultaneous two-color imaging to provide some measure of correction. Since motion occurs in both the red and green channels identically, any large signal occurring in both channels is likely to be a motion artifact and we can use a variety of methods to remove it.

The standard method for removing motion artifacts in a two-color experiment is ratiometric correction. However, this method can lead to erroneous spikes being recorded: when both signals approach low values, the ratio between them can be very large even if the absolute difference is small<sup>50</sup>. Therefore, ratiometrically corrected signals can show a large increase in intensity when both signals are actually decreasing in intensity.

Thus, we used independent component analysis (ICA) (Fig. 7C) to separate the motion artifacts from changes in fluorescence due to calcium dynamics, as has previously been shown to perform better than ratiometric correction in *C. elegans* neurons expressing both GCaMP and RFP<sup>50</sup>. This method allows source separation by directly looking for common signals between the channels. ICA assumes that the only signals present in both the red and green channels are motion artifacts and calcium signals and seeks to find two component signals that can be recombined in differing proportions to reconstruct the original red and green channels. In practice, this method produces two signals: one mostly composed of motion signal, and the other composed of calcium signal. Noise is distributed equally between the outputs. As ICA is a stochastic algorithm, our implementation runs multiple times and selects the best resulting signals (see Methods). The calcium signal is identified from the outputs via its higher correlation with the original green channel signal.

#### Detrending and smoothing

Although ICA can reduce artifacts that occur in both channels identically, it cannot completely attenuate them, nor can it process artifacts that occur differently in each channel such as photobleaching, which occurs at different

rates in tdTomato and GCaMP7s. Additionally, spike identification algorithms are fragile. Baseline drift or abundant noise can drastically hinder their performance and cause them to miss or to incorrectly predict spikes, making downstream analysis far less reliable.

To improve the performance of our spike extraction, two additional preprocessing steps were performed: detrending to remove baseline drift (Fig. 7D) and smoothing (Fig. 7E) to clean the signals by reducing noise. Detrending was applied via frequency filtering. We deleted from each signal the frequencies smaller than a threshold tuned with user feedback. Smoothing was applied using a rolling average filter and was again tuned with user feedback to allow for optimal reduction of noise without removal of signal (see Methods).

#### Calcium signal and spike-time extraction

After filtering and preprocessing the extracted fluorescence signals, we utilized the FOOPSI spike extraction method (Fig. 7F) implemented in the CaImAn Python package<sup>30</sup>. This package is intended for use with mouse two photon calcium imaging data. However, since our data has low background noise and uses the same indicator used in mouse studies, we found this method was reliably able to extract spikes from our processed data with only minor tuning.

FOOPSI fits an autoregressive model to the provided signal and, using an underlying model of the expected calcium signal dynamics from GCaMP indicators, can predict the underlying fluorescence of GCaMP7s while ignoring the accompanying noise. In addition to extracting fluorescence, the FOOPSI algorithm provides spike times predicting the underlying electrical activity of the neuron from its calcium activity.

Spike time information from FOOPSI is provided as a probability of a spike having occurred at any given point in time. To extract robust spike times, a custom clustering and thresholding algorithm was developed (see Methods). The final output of the signal processing module is a raster plot (Fig. 7G) with predicted spikes for all tracked individual neurons.

## Methods

### Generation of transgenic *Hydra*

Following the method previously described<sup>51–53</sup>, a transgenic *Hydra vulgaris* (strain AEP) line was established to express GCaMP7s in the cytosol and tdTomato in the nucleus of cells derived from the interstitial stem cell lineage. The plasmid (Addgene catalog no. 102558)<sup>16</sup> was modified by replacing the actin promoter with the EF1-alpha promoter and inserting DsRed downstream of the GCaMP6s sequence. To ensure the nuclear localization of DsRed, a P2A self-cleaving peptide sequence containing a nuclear localizing signal (cccaagaagaagagggaaggtg) was inserted between GCaMP6s and DsRed. The nuclear localization of DsRed was confirmed by electroplating the plasmid into *Hydra*. To enhance fluorescence intensity, GCaMP6s was replaced with jGCaMP7s and DsRed was replaced with tdTomato. Finally, for microinjections, the EF1-alpha promoter was replaced with the Actin promoter. All fluorescent reporter gene sequences were codon optimized specifically for *Hydra*. Standard embryo microinjection of this plasmid was performed and transgenic hatchlings expressing cytoplasmic GCaMP7s and nuclear tdTomato in the interstitial cell lineage were isolated. Transgenic animals were bred until all neurons were expressing both cytoplasmic GCaMP7s and nuclear tdTomato. Transgenic *Hydra* were cultured using standard methods<sup>54</sup> in *Hydra* medium (1 mM calcium chloride dehydrate, 0.33 mM magnesium sulfate anhydrous, 0.5 mM sodium bicarbonate, 0.03 mM potassium chloride) at 18 °C on a 12 h light/12 h dark cycle. They were fed freshly hatched *Artemia* nauplii twice per week.

### Imaging

Dual-labeled transgenic *Hydra* were prepared for imaging as described<sup>16</sup>. Imaging was performed using a custom dual-channel spinning disc confocal microscope (Solamere Yokogawa CSU-X1) with a sCMOS camera for each channel (Teledyne-Photometrics Prime-BSI). Samples were simultaneously illuminated with both 488 nm and 561 nm lasers (Coherent OBIS). Emission light was split with a dichroic mirror sending green light to one camera and red light to the other. Cameras were aligned with pinholes and images were registered during processing. Images were captured with a frame rate of 10 frames per second using a 6X objective (Navitar HR Plan Apo 6X/0.3) and Micro-Manager software.

### Immunostaining

Immunostaining was performed using the previously established protocol optimized for double-labeling experiments in *Hydra* with the anti-*Hydra* cadherin antibody<sup>33</sup>, using the following antibodies: Primary antibodies: *Hydra* cadherin antibody (rabbit) 1:1000 (Thomas Holstein, Heidelberg), anti-tdTomato (goat) 1:200 (Origene, AB8181-200); Secondary antibodies: Alexa 488 donkey anti-rabbit 1:1000 (ThermoFisher A-21206), Rhodamine Red-X donkey anti-goat 1:250 (Jackson 705–295-147). Fixed and stained animals were imaged on the same spinning disc confocal setup described above. Image stacks were taken with a 40X objective (Olympus LUMPlanFl/IR 40X/0.8 W) every 0.5 µm and processed with ImageJ. Images shown are maximum intensity projections of short stacks (5–20 µm).

### Segmentation and tracking of nuclei (ByoTrack)

#### Segmentation of neuronal nuclei

To learn a segmentation model and validate our experiments, we sampled 20 tdTomato images of *Hydra* on various positions of the animal (e.g., contracted, elongated). We manually annotated these images using the ImageJ draw tool resulting in 12,705 segmented neurons (~600/image).

To measure the segmentation performance, 5 random images were kept as a testing dataset. To account for performance variability, we repeated this process 5 times with different seeds, resulting in different training and testing images. We measured the performance with each seed and report the results as (mean  $\pm$  std).

We used standard instance segmentation metrics (precision, recall, f1-score). But we did not base these metrics on IOU association, where a predicted instance is considered to be a true positive if it overlaps with a ground truth on a sufficiently large area. We rather based these metrics on distance association, where a predicted instance (spot) is considered valid if its mass-center is sufficiently close to the mass-center of a ground truth instance. This has two benefits: first, it is more meaningful for tracking as we only use the position of each detected instance. Second, it allows fair comparison with detection methods that do not try to produce a faithful segmentation or bounding box of the instances like the wavelet thresholding method.

The remaining 15 images were used to train and tune all the parameters of the segmentation model. We also measured the impact of the training data quantity by training the model with fewer images, from 1 to 15.

We compared two detection methods: an analytical method based on wavelet decomposition and thresholding, and a deep learning approach, StarDist.

#### *Calibrating wavelet thresholding method*

The method has only two hyperparameters: the scale of the spots and the noise threshold. Both can be easily tuned by hand on a single image. Nonetheless, to be fully autonomous, the system performs a grid search on the training images and chooses the best performing ones. No further training is required. Moreover, to reduce the number of false positives, the neurons with less than 5 pixels area were filtered out.

#### *Training DL method (StarDist)*

The deep neural network must be trained and the hyperparameters tuned. We therefore split the training images into training and validation as most standard deep learning approaches do. 20% of the training images were used as validation images. With less than 5 training images, one image was kept for validation. Therefore, the procedure that we show here requires at least 2 training images, one to train the weights of the network, the other to validate the hyper parameters.

The official implementation of StarDist (<https://github.com/stardist/stardist>) was used to train, validate, and evaluate the performances with our dataset. For tracking, we provide a wrapper to perform the detection process using a trained StarDist model (<https://partage.imt.fr/index.php/s/npwHJHZebxqGMPi>).

#### *Generating tracklets with probabilistic eMHT*

The eMHT algorithm is implemented on the open-source imaging platform Icy (plugin Spot Tracking), which is coded in Java language. Instead of translating all the code into Python, we decided to directly call the Spot Tracking plugin in Icy in a headless mode from the TraSE-IN platform. To ensure compatibility with Icy, we implemented different functions for inputs (detections) and outputs (tracks) wrapping.

We believe that calling implemented tracking solutions in major imaging platforms such as Icy and ImageJ is an efficient and sustainable solution because the main imaging platforms are open-source, the tracking plugins (e.g., Spot Tracking in Icy, Trackmate in ImageJ) are regularly updated by developers, wrapping input and output variables in TraSE-IN is much less tedious and prone to implementation errors than translating all the Java code of tracking methods in Python, and TraSE-IN will easily integrate new tracking solutions developed on other platforms, if needed.

#### *Stitching tracklets*

To stitch tracklets obtained with eMHT tracking, we implemented a cost-based algorithm. Briefly, a cost between all the tracklets is computed, and stitched tracklets correspond to the minimal global cost of all tracklets' linking. Computing the minimum global cost corresponds to a linear assignment problem that we solved using the Jonker-Volgenant algorithm<sup>47</sup>. The cost between tracklets was computed following the EMC<sup>2</sup> algorithm<sup>25</sup>. Input parameters of the stitching function (byotrack.implementation.refiner.stitching.emc2) are the smoothness  $\alpha$  of the Thin Plate Spline interpolation and the non-linking cost  $\eta$  (paid for unlinked tracklets). In our tracking scenario (Fig. 4C), we set  $\alpha = 10$  to provide enough regularization to be robust to eMHT linking errors and set non-linking cost  $\eta = 5$  pixels.

### **Linking calcium activity to tracked nuclei**

To solve the misalignment between nuclei and calcium signal, we implemented a sub-ROI tracking method. For each nucleus, we extracted a  $25 \times 25$  pixels ROI centered on each tracked nucleus. Then, to find neuronal candidates in the GCaMP7s channel, we iteratively computed  $n$  local maxima within each ROI ( $n = 5$  typically) and identified the most probable calcium signal. For this, we first applied a gaussian smoothing ( $\sigma = 1$  pixel) and first identified the global maximum in the ROI. We then deleted the pixels in the neighborhoods of this maximum and reiterated the process  $n$  times. To bias extracted maxima towards the nucleus position (and avoid selecting an outlier GCaMP7s signal) we weighted the ROI intensities with Gaussian prior centered on the nucleus position ( $\sigma = 5$  pixels).

In the first frame, the selected position of the calcium spot is the global maximum of the ROI (after having applied a Gaussian prior to bias intensities towards the nucleus position). For the following frames, we selected the closest maximum to the previous calcium position and updated the calcium position with an exponential moving average to attenuate the potential consequence of a wrong association. Finally, we extracted the calcium intensity for each neuron as the mean intensity within the 5 pixels radius circle centered on the tracked calcium position.

## Signal processing

### *Removing non-neuronal cells*

We compared extracted calcium signals with a normal Gaussian distribution using the `scipy.stats.normaltest` method. Any signal whose normality can be rejected with p-value lower than a user-defined threshold was kept. The other signals that might correspond to non-neuronal cells are displayed in an interactive feedback step to allow users to decide which cells they want to keep or discard from the analysis. This feedback step allows the user to visualize each cell corresponding to a Gaussian fluorescent signal. Then, by observing the cell's location, size, shape, and activity over the timelapse movie, users can decide if they would like to discard the cell or keep it for further downstream analysis.

### *Motion artefact removal*

We used the tdTomato intensity (extracted as the mean intensity in a 5 pixels radius circle centered on the nucleus position in the red channel) as a control signal. ICA between the control signal and the raw calcium signal was used to remove motion artefacts<sup>50</sup>. ICA relies on scikit-learn FastICA method<sup>55</sup> (we used a tolerance of 0.5). As the ICA algorithm is stochastic, we ran it several times and selected the best outcome as the one with the minimum correlation with the control signal.

### *Detrending and smoothing*

To remove the remaining baseline in the calcium signal, we applied frequency filtering using a 5<sup>th</sup> order Butterworth filter and a critical period of 100 frames. In addition, we smoothed the signal using a rolling average of size 5. Both the average size and critical period can be adapted with user feedback.

### *Calcium signal and spike-time extraction*

We used the constrained FOOPSI implementation in the CaImAn Python library (OASIS method of order 2)<sup>30</sup>. In *Hydra*, calcium spikes correspond to single action potentials<sup>16</sup>. Therefore, we added a clustering step to aggregate the action potentials (electrical spikes) predicted with the previous deconvolution FOOPSI algorithm (less than 5 frames away). The clustering consists of first temporal Gaussian blurring ( $\sigma = 5$  frames) of the estimated electrical spikes, followed by local maxima extraction. Finally, we filtered out electrical spikes with low probability/intensity using a user-defined threshold to reduce the number of erroneous predictions. Since these computed probabilities/intensities vary wildly between different signals, using a fixed threshold would not work across different signals. To rectify this, an adaptive threshold defined as the maximum probability minus 2 standard deviations was used. This method provided robust results across all cells, allowing spike extraction to be applied simultaneously across an entire dataset with minimal user input.

## Discussion

Here, we present TraSE-IN, an open access end-to-end protocol for single-neuron resolution calcium imaging in behaving animals. To our knowledge, this pipeline is the first to allow automatic detection, tracking, signal extraction, and spike prediction from individual neurons in a highly deformable organism such as *Hydra*.

Hardware improvements implemented in this pipeline include a novel dually labeled transgenic *Hydra* that is imaged with simultaneous two-color spinning disc confocal microscopy to allow high spatiotemporal resolution imaging of the entire animal in the field of view while behaving. The novel transgenic *Hydra* presented here is labeled with a calcium-insensitive fluorophore (tdTomato) in the nucleus and calcium-sensitive fluorophore (GCaMP7s) in the cytoplasm of each neuron. The presence of the calcium-insensitive fluorophore in the nucleus of each neuron allows visualization of each neuron even while it is inactive and significantly increases accuracy of tracking individual neurons over time. The high-resolution spinning disc confocal imaging also increases the accuracy of tracking individual neurons longitudinally because they are easier to detect in each frame due to the high spatial resolution, and easier to track frame-to-frame due to the high temporal resolution. Software improvements include more accurate segmentation of neurons in each frame using a machine learning algorithm (StarDist<sup>41</sup>) specifically trained to detect *Hydra* nuclei, an accurate single neuron tracking protocol using a state-of-the-art probabilistic algorithm (eMHT<sup>45</sup>) that is directly executed on the native open-source platform Icy<sup>46</sup>, tracklet stitching<sup>25</sup>, and, finally, a user-friendly Python pipeline for signal extraction, signal processing, and spike prediction for individual neurons in a behaving animal.

While the development of this pipeline is a major step forward towards the goal of automatic detection, tracking, signal extraction, and spike prediction from single neurons in behaving animals over longer timescales, multiple areas for future improvements remain. First, although the novel dually labeled transgenic *Hydra* line allows visualization of neurons even while inactive, transgenic expression is not specific to neurons. The fluorescent proteins are also expressed in the interstitial cell lineage of the animal, which is a stem cell lineage that makes four cell types: neurons, nematocytes, gland cells, and germ cells<sup>34–37</sup>. Thus, all these cell types could express the label in the transgenic animals. In the current pipeline, the non-neural cell types can be filtered out in the signal processing step of the pipeline. Instead of filtering out these different cell types during signal processing, it would be ideal to generate an animal with neuron-specific labeling using neuron-specific promoters<sup>17</sup>. In addition to more specific labeling of neurons in animals, it would be ideal to use different optical approaches that allow volumetric imaging of the entire animal during behavior. Here, *Hydra* is only imaged in a thick two-dimensional plane, which captures most of its neurons, but this method results in some neurons coming in and out of the imaging plane, so some neurons are lost over time. This is especially problematic for the tentacles, which move more than the body column of the animal. At present, current volumetric approaches are too slow to allow high spatiotemporal resolution imaging of an entire rapidly moving organism<sup>24</sup>. However, as fluorescent

labels and optics improve, shorter exposure times and higher frame rates may be possible to allow whole animal volumetric imaging while *Hydra* is behaving.

In addition to hardware improvements, multiple improvements can also be implemented on the software side. In terms of segmentation, even if StarDist provides accurate detection with relatively few manual annotations (~ 2 image patches for a total of ~ 1,000 segmented nuclei), transfer learning from pre-annotated public datasets would be a very valuable resource for the segmentation of cells in various imaging set-ups. In terms of tracking, we designed TraSE-IN so it does not code for a specific method, but rather executes state-of-the-art tracking algorithms directly from their native platform. In the present version, TraSE-IN is executing the eMHT probabilistic algorithm in Icy (<https://icy.bioimageanalysis.org/>), but future versions of the platform could implement other well-established tracking methods such as TrackMate in ImageJ<sup>56</sup>. For the tracklet stitching, the cost between tracklets corresponds to the minimal distance between the predicted forward- and backward-propagated positions of tracklets' ending- and starting-points, respectively<sup>25</sup>. This cost could also include visual features of tracklets, which should further improve the accuracy and robustness of tracklets' stitching, as recently shown<sup>57</sup>. In terms of signal extraction, many putative improvements reside in spike extraction. Indeed, the robust estimation of spiking times from noisy fluorescence traces is an active field of research, and recent developments in Bayesian analysis<sup>26,48</sup> might improve the accuracy of spike estimates compared with previous constrained deconvolution methods<sup>27,29,58</sup>.

In sum, we have created a robust, adaptative end-to-end pipeline that allows single-neuron resolution imaging of behaving *Hydra* followed by automatic detection, tracking, signal extraction, signal processing, and spike prediction of individual neurons. While room for future improvement remains, this new tool takes us one step closer towards the goal of automatically extracting every spike from every neuron in a behaving animal. Such data will allow visualization and analysis of the emergent properties of an entire nervous system under different conditions, making it possible to eventually crack the neural code of a whole animal.

## Data availability

Source code and documentation for the plugin are available at <https://github.com/raphaelreme/trase-in>. Movies are available on the Dryad platform at <https://doi.org/10.5061/dryad.h9w0vt4q3>. Our trained StarDist model for nuclei segmentation is available at <https://partage.imt.fr/index.php/s/npwHJHZebxqGMPi>.

Received: 25 September 2023; Accepted: 26 February 2024

Published online: 01 March 2024

## References

- Alivisatos, A. P. *et al.* The brain activity map. *Science* **339**, 1284–1285 (2013).
- Markicevic, M., Savvateev, I., Grimm, C. & Zerbi, V. Emerging imaging methods to study whole-brain function in rodent models. *Transl Psychiatry* **11**, 1–14 (2021).
- Milham, M. *et al.* Accelerating the evolution of nonhuman primate neuroimaging. *Neuron* **105**, 600–603 (2020).
- Lüsebrink, F. *et al.* Comprehensive ultrahigh resolution whole brain in vivo MRI dataset as a human phantom. *Sci. Data* **8**, 138 (2021).
- Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. & Keller, P. J. Whole-brain imaging at cellular resolution using light-sheet microscopy. *Nat. Methods* **10**, 413–420 (2013).
- Vladimirov, N. *et al.* Brain-wide circuit interrogation at the cellular level guided by online analysis of neuronal function. *Nat. Methods* **15**, 1117–1125 (2018).
- Zhang, Z. *et al.* Imaging volumetric dynamics at high speed in mouse and zebrafish brain with confocal light field microscopy. *Nat. Biotechnol.* **39**, 74–83 (2021).
- Fotowat, H. & Engert, F. Neural circuits underlying habituation of visually evoked escape behaviors in larval zebrafish. *eLife* **12**, e82916 (2023).
- Mann, K., Gallen, C. L. & Clandinin, T. R. Whole-brain calcium imaging reveals an intrinsic functional network in drosophila. *Curr. Biol.* **27**, 2389–2396.e4 (2017).
- Aimon, S. *et al.* Fast near-whole-brain imaging in adult *Drosophila* during responses to stimuli and behavior. *PLOS Biol.* **17**, e2006732 (2019).
- Nguyen, J. P. *et al.* Whole-brain calcium imaging with cellular resolution in freely behaving *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci.* **113**, E1074–E1081 (2016).
- Randi, F. & Leifer, A. M. Measuring and modeling whole-brain neural dynamics in *Caenorhabditis elegans*. *Curr. Opin. Neurobiol.* **65**, 167–175 (2020).
- Hadzi, J. Über das Nervensystem von Hydra. *Arb. Zool. Inst. Univ. Wien* **17**, 225–268 (1909).
- Burnett, A. L. & Diehl, N. A. The nervous system of hydra. I. types, distribution and origin of nerve elements. *J. Exp. Zool.* **157**, 217–226 (1964).
- Lentz, T. L. & Barrnett, R. J. Fine structure of the nervous system of hydra. *Am. Zool.* **5**, 341–356 (1965).
- Dupre, C. & Yuste, R. Non-overlapping neural networks in hydra vulgaris. *Curr. Biol.* **27**, 1085–1097 (2017).
- Siebert, S. *et al.* Stem cell differentiation trajectories in Hydra resolved at single-cell resolution. *Science* **365**, eaav9314 (2019).
- Vogg, M. C., Galliot, B. & Tsiaris, C. D. Model systems for regeneration: Hydra. *Development* **146**, dev177212 (2019).
- Han, S., Taralova, E., Dupre, C. & Yuste, R. Comprehensive machine learning analysis of Hydra behavior reveals a stable basal behavioral repertoire. *eLife* **7**, e32605 (2018).
- Woller, A., Bandow, P., Aimon, S. & Kadow, I. C. G. Preparing adult drosophila melanogaster for whole brain imaging during behavior and stimuli responses. *JoVE (Journal of Visualized Experiments)*. <https://doi.org/10.3791/61876> (2021).
- Schrödel, T., Prevedel, R., Aumayr, K., Zimmer, M. & Vaziri, A. Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light. *Nat. Methods* **10**, 1013–1020 (2013).
- Prevedel, R. *et al.* Simultaneous whole-animal 3D imaging of neuronal activity using light-field microscopy. *Nat. Methods* **11**, 727–730 (2014).
- Hallinen, K. M. *et al.* Decoding locomotion from population neural activity in moving *C. elegans*. *eLife* **10**, e66135 (2021).
- Hillman, E. M. C., Voleti, V., Li, W. & Yu, H. Light-sheet microscopy in neuroscience. *Annu. Rev. Neurosci.* **42**, 295–313 (2019).
- Lagache, T., Hanson, A., Pérez-Ortega, J. E., Fairhall, A. & Yuste, R. EMC2: A versatile algorithm for robust tracking of calcium dynamics from individual neurons in behaving animals. (2021) doi:<https://doi.org/10.1101/2020.06.22.165696>.

26. Deneux, T. *et al.* Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nat. Commun.* **7**, 12190 (2016).
27. Pnevmatikakis, E. A. *et al.* Simultaneous denoising, deconvolution, and demixing of calcium imaging data. *Neuron* **89**, 285–299 (2016).
28. Theis, L. *et al.* Benchmarking spike rate inference in population calcium imaging. *Neuron* **90**, 471–482 (2016).
29. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol.* **13**, e1005423 (2017).
30. Giovannucci, A. *et al.* CaImAn an open source tool for scalable calcium imaging data analysis. *eLife* **8**, e38173 (2019).
31. Dana, H. *et al.* High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657 (2019).
32. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol.* **22**, 1567–1572 (2004).
33. Keramidioti, A. *et al.* A new look at the architecture and dynamics of the Hydra nerve net. *eLife* **12**, (2023).
34. David, C. N. & Gierer, A. Cell cycle kinetics and development of hydra attenuata: III. Nerve and nematocyte differentiation. *J. Cell Sci.* **16**, 359–375 (1974).
35. David, C. N. & Murphy, S. Characterization of interstitial stem cells in hydra by cloning. *Dev. Biol.* **58**, 372–383 (1977).
36. Campbell, R. D. & Bode, H. R. Terminology for morphology and cell types. In *Hydra: Research methods* (ed. Lenhoff, H. M.) 5–14 (Springer, New York, 1983). [https://doi.org/10.1007/978-1-4757-0596-6\\_2](https://doi.org/10.1007/978-1-4757-0596-6_2).
37. Bode, H. R. The interstitial cell lineage of hydra: A stem cell system that arose early in evolution. *J. Cell Sci.* **109**, 1155–1164 (1996).
38. Yamamoto, W. & Yuste, R. Peptide-driven control of somersaulting in *Hydra vulgaris*. *Curr. Biol.* **33**, 1893–1905.e4 (2023).
39. Nguyen, J. P., Linder, A. N., Plummer, G. S., Shaevitz, J. W. & Leifer, A. M. Automatically tracking neurons in a moving and deforming brain. *PLoS Comput. Biol.* **13**, e1005517 (2017).
40. Smal, I., Loog, M., Niessen, W. & Meijering, E. Quantitative comparison of spot detection methods in fluorescence microscopy. *IEEE Trans. Med. Imaging* **29**, 282–301 (2010).
41. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell detection with star-convex polygons. [arXiv:1806.03535](https://arxiv.org/abs/1806.03535) 265–273 (2018) doi:[https://doi.org/10.1007/978-3-030-00934-2\\_30](https://doi.org/10.1007/978-3-030-00934-2_30).
42. Olivo-Marin, J.-C. Extraction of spots in biological images using multiscale products. *Pattern Recogn.* **35**, 1989–1996 (2002).
43. Izeddin, I. *et al.* Wavelet analysis for single molecule localization microscopy. *Opt Express* **20**, 2081–2095 (2012).
44. Chenouard, N. *et al.* Objective comparison of particle tracking methods. *Nat. Methods* **11**, 281–289 (2014).
45. Chenouard, N., Bloch, I. & Olivo-Marin, J.-C. Multiple hypothesis tracking for cluttered biological image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 2736–2750 (2013).
46. de Chaumont, F. *et al.* Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* **9**, 690–696 (2012).
47. Jonker, R. & Volgenant, A. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing* **38**, 325–340 (1987).
48. Diana, G., Sermet, B. S. & DiGregorio, D. A. High frequency spike inference with particle Gibbs sampling. doi:<https://doi.org/10.1101/2022.04.05.487201> (2022).
49. Giez, C. *et al.* Microbes as part of ancestral neuronal circuits: Bacterial produced signals affect neurons controlling eating behavior in Hydra. doi:<https://doi.org/10.1101/2023.04.28.538719> (2023).
50. Scholz, M. *et al.* Predicting natural behavior from whole-brain neural dynamics. 445643 (2018) doi:<https://doi.org/10.1101/445643>.
51. Wittlieb, J., Khalturin, K., Lohmann, J. U., Anton-Erxleben, F. & Bosch, T. C. G. Transgenic Hydra allow in vivo tracking of individual stem cells during morphogenesis. *Proc. Natl. Acad. Sci. U S A* **103**, 6208–6211 (2006).
52. Juliano, C. E., Lin, H. & Steele, R. E. Generation of transgenic Hydra by embryo microinjection. *J. Vis. Exp.* <https://doi.org/10.3791/51888> (2014).
53. Klimovich, A., Wittlieb, J. & Bosch, T. C. G. Transgenesis in Hydra to characterize gene function and visualize cell behavior. *Nat. Protoc.* **14**, 2069–2090 (2019).
54. Lenhoff, H. M. & Brown, R. D. Mass culture of hydra: an improved method and its application to other aquatic invertebrates. *Lab. Anim.* **4**, 139–154 (1970).
55. Pedregosa, F. *et al.* Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
56. Tinevez, J.-Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* **115**, 80–90 (2017).
57. Reme, R. *et al.* Tracking intermittent particles with self-learned visual features. In *2023 IEEE 20th international symposium on biomedical imaging (ISBI)* 1–5 (2023). doi:<https://doi.org/10.1109/ISBI53787.2023.10230664>.
58. Vogelstein, J. T. *et al.* Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* **104**, 3691–3704 (2010).

## Acknowledgements

We would like to thank John Szymanski and Jonathan Lovas for assistance with development of the dual reporter plasmid, Andrea Ogilvie Clark-Sevilla with assistance training the StarDist model, Thomas Holstein for generously providing the anti-*Hydra* cadherin antibody, and Charles David and Rob Steele for helpful discussions. This research was supported by the Leon Levy Fellowship in Neuroscience and K99 NS127851 (A.H.), French ANR-PRCI “Rebirth” (T.L. and R.R.) and the NSF (2203119; WY, NT, RY) Vannevar Bush Faculty Award (ONR N000142012828; WY, NT, RY).

## Author contributions

A.H. injected and screened the transgenic *Hydra* line, developed the custom imaging method and acquired all images, performed immunostaining experiments and analysis, trained and optimized the StarDist model, optimized tracking parameters and performed tracking accuracy analysis, and guided development of the pipeline. R.R. developed the platform TraSE-IN, implemented the functions for the detection (Wavelet and call for StarDist) and tracking (call for Icy *spot tracking* plugin, and Python implementation of EMC<sup>2</sup> algorithm), polished and debugged the calcium imaging analysis code, and performed the detection accuracy analysis. N.T. developed the calcium imaging analysis code to align images, extract fluorescence traces, perform signal processing and predict spike times, and provided tools for the analysis and visualisation of this data. W.Y. initiated a dual-color tracking project by constructing and validating the reporter plasmid using electroporation and injections. T.L. performed tracking accuracy analysis. All authors contributed to project conception, design, and manuscript writing. J.C.O.-M., T.L., and R.Y. assembled the team, directed the project, and secured equipment, resources, and funding.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-55608-2>.

**Correspondence** and requests for materials should be addressed to A.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



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

© The Author(s) 2024