

1 Data release - "2022_Q4_IBL_et_al_BWM" - Technical White Paper

2 International Brain Laboratory

3 Contents

4 Introduction	2
5 Recording strategy	2
6 Data set overview	4
7 Data access	6
8 Contact details	6
9 Methods	7
Overview	7
Animals	8
Headbar implant surgery	8
Materials and apparatus	8
Habituation, training, and experimental protocol	8
Electrophysiological recording using Neuropixels probes	9
Serial section two-photon imaging	9
Probe track tracing and alignment	9
Spike sorting	10
Coverage	10
20 Data quality	11
Sessions and insertions	11
Trials	11
Neurons and brain regions	11
Wheel data	11
Video data	11
Raw video QC	13
Pose estimation QC	13
Electrophysiology data	16
Raw ephys data QC	16
Spike sorting QC	16
31 Acknowledgments	28
32 Diversity Statement	28

33 **Introduction**

34 The International Brain Laboratory (IBL; www.internationalbrainlab.org) is a collaboration of 22 labs with the shared mission of
35 advancing our understanding of the neural systems and circuits that underlie behavior. The IBL approaches this challenge with a
36 distributed experiment model, in which scientists within each lab contribute to the same shared scientific experiment. Data
37 collection takes place across 12 labs and the data are preprocessed and stored centrally. The collaboration emphasizes an open
38 science ethic in which data and code are shared freely.

39 Here, we present our first release of data from the "Brain-wide Map" project, in which we have systematically recorded from
40 nearly all major brain areas with Neuropixels probes. Our recording strategy, described in detail below, used a grid system for
41 unbiased sampling and replicated each recording site in at least two laboratories. These data constitute a brain-wide map of
42 activity at single-spike cellular resolution during a decision-making task. In addition to the map, this data set contains other
43 information gathered during the task: sensory stimuli presented to the mouse; mouse decisions and response times; and mouse
44 behavioral information from analyzed video recordings.

45 This is the first release of a living dataset: additional recordings are ongoing and future data releases will increase the size of
46 the dataset.

47 **Recording strategy**

48 Our recording strategy aimed to achieve these goals:

- 49 1. Map the activity from all brain regions, spanning from rostral tip of the left forebrain to the right brainstem and cerebellum
50 caudally. However, we excluded the olfactory bulb region.
- 51 2. Obtain a uniform coverage across the brain volume
- 52 3. Assess the reproducibility of recordings across subjects and laboratories (via the "repeated site")
- 53 4. Assess bilateral, interhemispheric interactions (via the "bilateral sites")

54 To accomplish this, we designed a grid of recording sites with 500 μm spacing (Fig 1a-c). The grid locations were sampled with
55 one of two probes per session (typically, one coming from the left of the subject and one from the right). The grid is focused on
56 the left hemisphere to maximize the proportion of pairs of areas that share anatomical connections (for this reason, cerebellum
57 and medulla sampled from the right hemisphere instead of left, since connections from these areas to the rest of the brain cross
58 the midline). To assess the reproducibility of our physiological measurements across subjects and laboratories, we selected
59 one of these recording sites as the "repeated site" (Fig 1d) and targeted a recording to this site once in every mouse of the
60 dataset (*The International Brain Laboratory*, 2022a). Recordings from the "repeated site" sampled activity in the visual cortex,
61 hippocampus and thalamus. To test for hemisphere-specific activity, three locations were chosen to be recorded bilaterally
62 (i.e. left and right hemisphere of the same coordinate simultaneously) in a small number of subjects (Fig 1e). The first location
63 sampled the visual, retrosplenial cortices areas, midbrain and pons ; the second location was the "repeated site" ; the third
64 location sampled the motor, orbital, and olfactory cortical areas.

65 This grid served as an initial target; more insertions are being performed, in particular to reach areas that are currently
66 lacking coverage. Such additional datasets will be released in coming years once processing is completed.

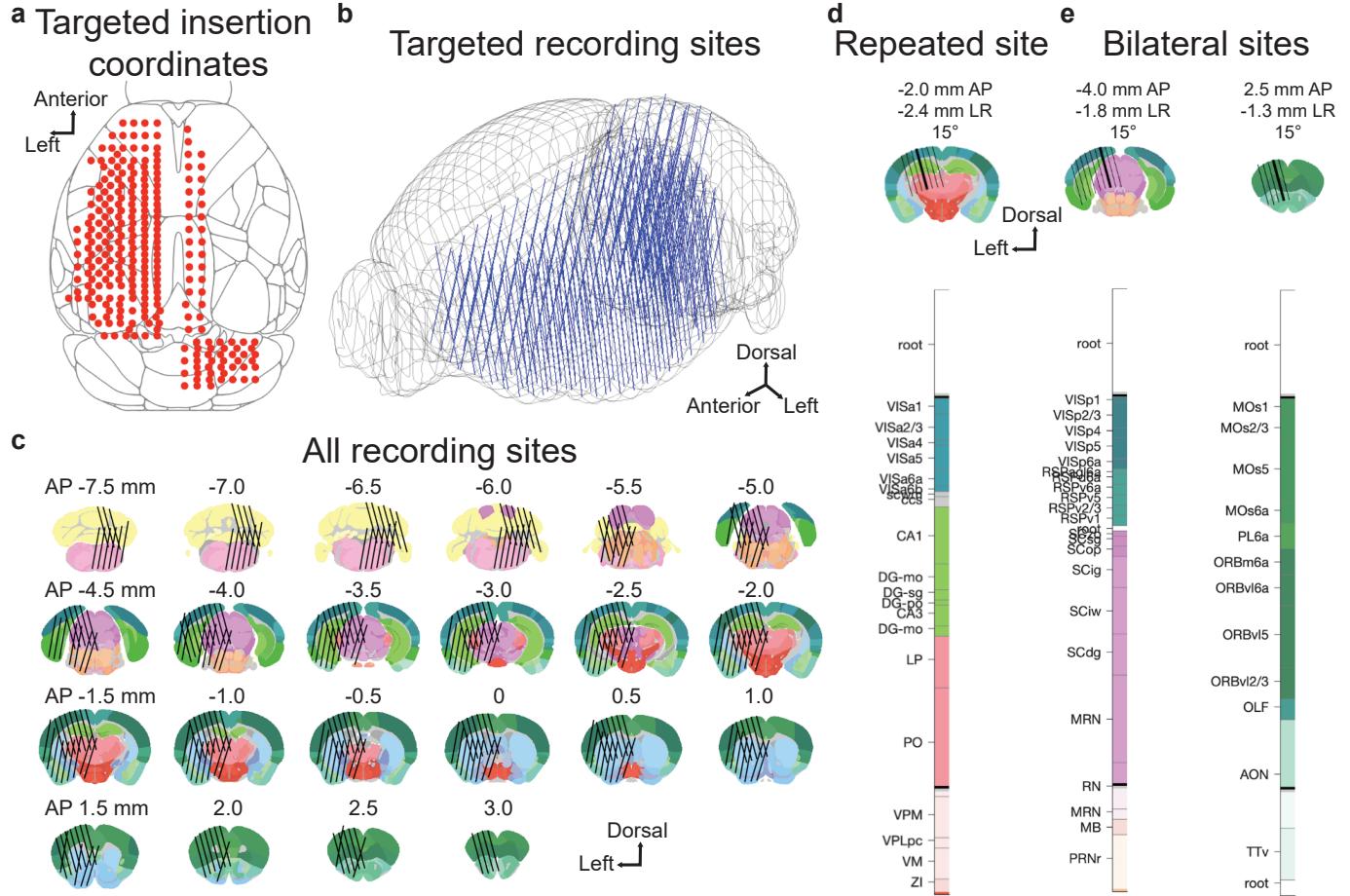


Figure 1. Recording strategy. **a.** Targeted insertion locations on the brain surface. **b.** 3D positions of all recording sites. Each line represents the 3.84 mm span of electrode locations recorded by a single probe. **c.** Exact locations of all targeted recording sites (black lines) shown on coronal slices of the Allen CCF atlas. **d.** The "Repeated Site" location. Top, A coronal slice of the brain at -2.0 mm ap with the depicted penetration highlighted in dark black. Bottom, The brain regions traversed along this trajectory, as well as those 1 mm above and below, for context. The recorded area is anticipated to be the bright area between the two solid black lines, i.e. from the top of VISa to near the bottom of PO, but not including VPM, etc. **e.** The "Bilateral Site" locations. Conventions as in d. In addition to these two sites, the Repeated Site was also recorded bilaterally.

67 Data set overview

68 The brain-wide-map dataset of IBL consists of 295501 units (see section *Data quality* for definition), recorded using Neuropixels
69 probes in 116 mice doing the IBL task, distributed across 12 labs in Europe and the US. The neural data is obtained from at most
70 two Neuropixels probes recorded simultaneously per behavioral session (N session = 354), with each mouse undergoing multiple
71 of such sessions, totalling to 547 insertions covering most of the left hemisphere (Fig2a).

72 Though recordings were targeted to an evenly spaced grid (Fig 1), differences between the subject's brain and the atlas space,
73 along with experimental variability, led to actual recording locations as reconstructed with histology (see section *Methods*) being
74 at slightly different locations. Accounting for the true location of each recording, our coverage in this dataset is 90% of the total
75 targeted brain volume.

76 After spike sorting and single unit quality control (*The International Brain Laboratory*, 2022b), the brain region of each recorded
77 neuron was identified. We consider a brain region to have been recorded in this dataset when at least two insertions, which each
78 contained at least 10 neurons, were completed. By this definition, we recorded from 192 distinct parcellated regions in the Allen
79 CCF (Wang *et al.*, 2020) (see online table for details). Fig 2b shows the distribution of recorded neurons (i.e. good units) across
80 brain regions, with horizontal black lines in the bars separating different recordings.

81 The decision making task (Fig2c,d) requires the mouse to move the visual stimulus 35° to the center of the screen, by moving
82 a wheel with its two front paws (see section *Methods* for details). The stimulus is presented on the left or right side of the screen
83 with a probability that varies from one block of trials to the next. Of interest, zero-contrast stimuli are presented within these
84 blocks, which when responded to, offer an insight on the mouse prior to the likely side of the stimulus.

85 In order to study neural correlates of the behavior, we recorded both behavioral and neural data simultaneously. The major
86 behavioral events of interest for analysis are the stimulus onset time, the animal's choice, the feedback delivery time, the
87 feedback type (reward vs error tone), the probability of the stimulus being on the left (block type), and the wheel motion onset
88 time. Of note, in this task the time between stimulus onset and wheel motion onset - the reaction time of the animals - is short,
89 with a median of 150 ms (inset in Fig2d).

90 Besides these scalars describing the events ongoing in the trials, there are continuous recordings of the mice' behavior using
91 the wheel, as well as features extracted from videos of the animals (*The International Brain Laboratory*, 2021a). Fig2e illustrates
92 three such continuous behavioral time series, namely: the wheel speed, the whisker pad motion energy, and licking (obtained
93 from tongue tracking on the video images), aligned to neural activity recorded in various brain regions simultaneously.

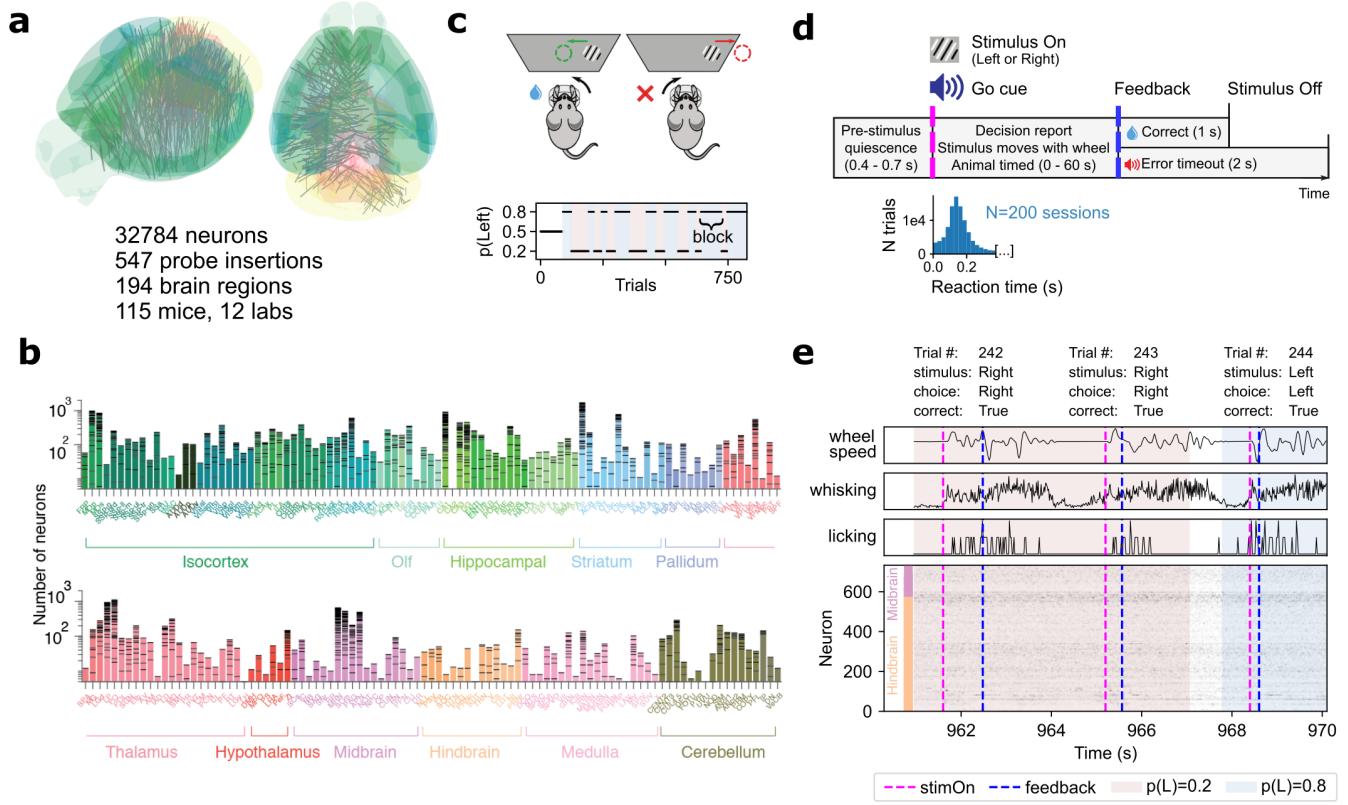


Figure 2. Task and data types. **a.** Neuropixel probe trajectories shown within a 3D brain schematic. A mesh of recordings covers most of the left hemisphere. **b.** Number of neurons (i.e. good units) recorded in each region, with horizontal black lines separating different recordings. **c.** The IBL task schematic and block structure of an example session. The probability of the stimulus being on the left side is varied in blocks of consecutive trials. **d.** Temporal details of the IBL task with histogram of reaction times taken from 200 sessions. **e.** Example time series and key trial information for 3 trials, resulting from the rotary encoder of the wheel, video analysis and electrophysiological recordings.

94 **Data access**

- 95 • The data collected can be viewed at <https://viz.internationalbrainlab.org/app>.
- 96 • To download the data, see the instructions at https://int-brain-lab.github.io/iblenv/notebooks_external/data_release_brainwidemap.html.
- 97 • To assess the quality of the data, see the general QC spreadsheet (see section **Data quality** for definitions)
- 98

99 **Contact details**

- 100 • Issues with the data? Post a question on one of these platforms:
- 101 - <https://github.com/int-brain-lab/iblenv/issues>
- 102 - <https://neurostars.org/>; make sure to apply the tag `ibl`
- 103 • General questions about the datasets or publications? Email info@internationalbrainlab.org

104 Methods

105 All procedures and experiments were carried out in accordance with the local laws and following approval by the relevant
106 institutions: the Animal Welfare Ethical Review Body of University College London; the Institutional Animal Care and Use
107 Committees of Cold Spring Harbor Laboratory, Princeton University, University of Washington, and University of California at
108 Berkeley; the University Animal Welfare Committee of New York University; and the Portuguese Veterinary General Board.

109 Overview

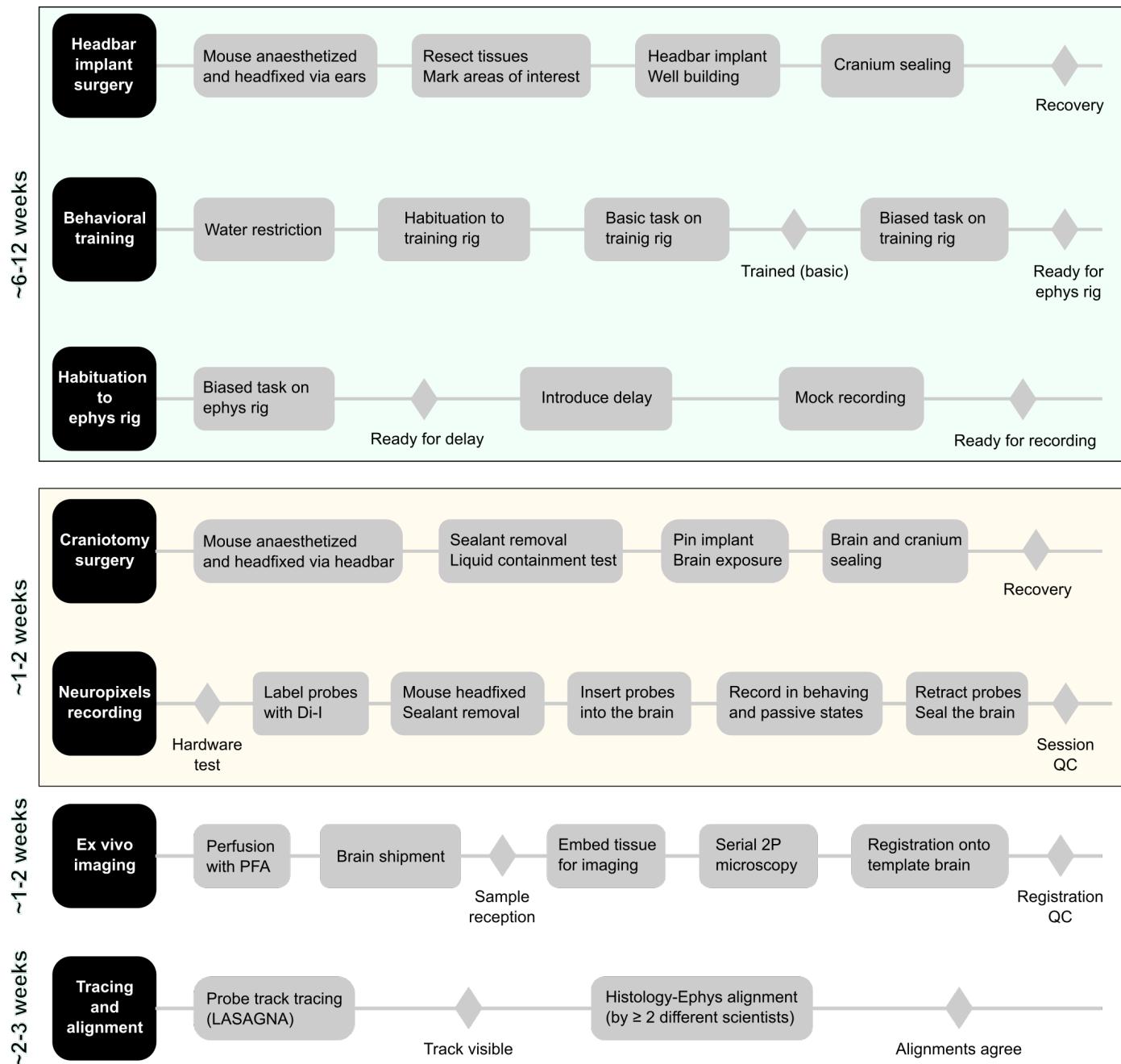


Figure 3. Overview of the experimental pipeline. The experiment follows the main steps indicated in the left-hand black squares in chronological order from top to bottom. Within each main step, actions are undertaken from left to right; diamond markers indicate points of control. Time estimate for each step of the pipeline is indicated on the left.

110 **Animals**

111 Mice were housed under a 12/12 h light/dark cycle (normal or inverted depending on the laboratory) with food and water
112 available ad libitum, except during behavioural training days. Electrophysiological recordings and behavioural training were
113 performed during either the dark or light phase of the cycle depending on the laboratory. N=115 adult mice (C57BL/6, male and
114 female, obtained from either Jackson Laboratory or Charles River) were used in this study. Mice were aged 13-93 weeks and
115 weighed 16.1-36.2 g on the day of electrophysiological recording.

116 **Headbar implant surgery**

117 A detailed account of the surgical methods for the headbar implant is in Appendix 1 of (*The International Brain Laboratory*, 2021c).
118 Briefly, mice were anesthetized with isoflurane and head-fixed in a stereotaxic frame. The hair was then removed from
119 their scalp, much of the scalp and underlying periosteum was removed and bregma and lambda were marked. The head was
120 positioned in stereotaxic coordinates. The head bar was then placed in one of three stereotactically defined locations and
121 cemented in place. The location of the future craniotomies were measured using a pipette referenced to Bregma, and marked on
122 the skull using either a surgical blade or pen. The exposed skull was then covered with cement (Vetbond) and clear UV curing
123 glue (Norland Optical Adhesives), ensuring that the remaining scalp was unable to retract from the implant.

124 **Materials and apparatus**

125 For detailed parts lists and installation instructions, see Appendix 1 of (*The International Brain Laboratory*, 2022a).

126 Briefly, each lab installed a standardized electrophysiological rig, which differed slightly from the apparatus used during
127 behavioral training (*The International Brain Laboratory*, 2021c). The general structure of the rig was constructed from Thorlabs
128 parts and was placed on an air table (Newport, M-VIS3036-SG2-325A) surrounded by a custom acoustical cabinet, which design
129 varied from lab to lab. A static head bar fixation clamp and a 3D-printed mouse holder were used to hold a mouse such that its
130 forepaws rested on the steering wheel (86652 and 32019, LEGO) (*The International Brain Laboratory*, 2021c). Silicone tubing
131 controlled by a pinch valve (225P011-21, NResearch) was used to deliver water rewards to the mouse. The display of the visual
132 stimuli occurred on a LCD screen (LP097Q × 1, LG). To measure the precise times of changes in the visual stimulus, a patch of
133 pixels on the LCD screen flipped between white and black at every stimulus change, and this flip was captured with a photodiode
134 (Bpod Frame2TTL, Sanworks). Ambient temperature, humidity, and barometric air pressure were measured with the Bpod
135 Ambient module (Sanworks), wheel position was monitored with a rotary encoder (05.2400.1122.1024, Kubler). Videos of the
136 mouse were recorded from 3 angles (left, right and body) with USB cameras (CM3-U3-13Y3M-CS, Point Grey) sampling at 60, 150,
137 30 Hz respectively (for details see Appendix 1 of (*The International Brain Laboratory*, 2022a)). A custom speaker (Hardware Team
138 of the Champalimaud Foundation for the Unknown, V1.1) was used to play task-related sounds, and an ultrasonic microphone
139 (Ultramic UM200K, Dodotronic) was used to record ambient noise from the rig. All task-related data was coordinated by a Bpod
140 State Machine (Sanworks). The task logic was programmed in Python and the visual stimulus presentation and video capture
141 was handled by Bonsai (Lopes et al., 2015) and the Bonsai package BonVision (Lopes et al., 2021).

142 All recordings were made using Neuropixels 1.0 probes (Imec (Jun et al., 2017); some probes were 3A and 3B prototype
143 models), advanced in the brain using a micromanipulator (Sensapex, uMp-4) tilted at a 15 degree angle from the vertical line.
144 Data were acquired via an FPGA (for 3A probes) or PXI (for 3B and 1.0 probes, National Instrument) system using SpikeGLX, and
145 stored on a PC.

146 **Habituation, training, and experimental protocol**

147 For a detailed protocol on animal training, see section *Methods* in (*The International Brain Laboratory*, 2021c, 2022a).

148 All the data analysed in this study were acquired when the mouse performed the task named "ephysChoiceWorld". Briefly, at
149 the beginning of each trial, the mouse was required to not move the wheel for a quiescence period of 400–700 ms. If the wheel
150 moved during this period, the timer was reset. After the quiescence period, a visual stimulus appeared on either the left or right
151 ($\pm 35^\circ$ azimuth) of the screen, with a contrast randomly selected from a predefined set. Simultaneously, an onset tone (5 kHz sine
152 wave) was played for 100 ms. When the stimulus appeared, the mouse had 60 s to move the wheel. A response was registered if

153 the center of the stimulus crossed the $\pm 35^\circ$ azimuth line from its original position. If the mouse correctly moved the stimulus 35°
154 to the center of the screen, it immediately received a $1.5 \mu\text{L}$ reward; if it incorrectly moved the stimulus 35° away from the center,
155 it received a timeout. If the mouse responded incorrectly or failed to reach either threshold within the 60 s window, a noise burst
156 was played for 500 ms and the inter-trial interval was set to 2 s. In trials where the visual stimulus contrast was set to 0, the mouse
157 had to respond as for any other trial, i.e. by turning the wheel in the correct direction in order to receive a reward - however, in
158 such a case the mouse was not able to perceive whether the stimulus was presented on the left or right side of the screen.

159 **Electrophysiological recording using Neuropixels probes**

160 For details on the craniotomy surgery, see Appendix 3 of (*The International Brain Laboratory*, 2022a).

161 Briefly, upon the first day of electrophysiological recording, the animal was anaesthetised using isoflurane and surgically
162 prepared. The cement and glue were removed, exposing the skull over both hemispheres. A test was made to check whether
163 the implant could hold liquid, and if successful a grounding pin was implanted. One or multiple craniotomies (approximately
164 $1 \times 1 \text{ mm}$) were made over the marked locations. The dura was left intact, and the brain was lubricated with ACSF. DuraGel
165 was applied over the dura as a moisturising sealant, and covered with a layer of Kwikcast. The mouse was administered with
166 analgesics subcutaneously, and left to recover in a heating chamber until locomotor and grooming activity were fully recovered.

167 After a further recovery period of at least 2 hours, subjects could be head-fixed in the apparatus for recording. Once a
168 craniotomy was made, up to 4 subsequent recording sessions were made in that same craniotomy. Once the first set of
169 craniotomy was fully recorded from, a subject could undergo another craniotomy surgery in accordance with the institutional
170 licence. Up to two probes were inserted in the brain on a given session.

171 **Serial section two-photon imaging**

172 Mice were given a terminal dose of pentobarbital and perfuse-fixed with PBS followed by 4% formaldehyde solution (Thermofisher
173 28908) in 0.1M PB pH 7.4. Whole mouse brain was dissected, and post-fixed in the same fixative for a minimum of 24 hours
174 at room temperature. Tissues were washed and stored for up to 2-3 weeks in PBS at 4C, prior to shipment to the Sainsbury
175 Wellcome Centre for image acquisition. For full details, see Appendix 5 of (*The International Brain Laboratory*, 2022a).

176 For imaging, brains were equilibrated with 50mM PB solution and embedded into 5% agarose gel blocks. The brains were
177 imaged using serial section two-photon microscopy (Ragan et al., 2012; Economo et al., 2016). The microscope was controlled with
178 ScanImage Basic (Vidrio Technologies, USA), and BakingTray, a custom software wrapper for setting up the imaging parameters
179 (Campbell, 2020). Image tiles were assembled into 2D planes using StitchIt (Campbell, 2021). Whole brain coronal image stacks
180 were acquired at a resolution of $4.4 \times 4.4 \times 25.0 \mu\text{m}$ in XYZ, with a two-photon laser wavelength of 920 nm, and approximately 150
181 mW at the sample. The microscope cut 50 μm sections but imaged two optical planes within each slice at depths of about 30 μm
182 and 55 μm from the tissue surface. Two channels of image data were acquired simultaneously using multialkali PMTs ('Green' at
183 525 nm ± 25 nm; 'Red' at 570 nm low pass).

184 Whole brain images were downsampled to 25 m XYZ pixels and registered to the adult mouse Allen common coordinate
185 framework (Wang et al., 2020) using BrainRegister (West, 2021), an elastix-based (Klein et al., 2010) registration pipeline with
186 optimised parameters for mouse brain registration. For full details, see Appendix 7 of (*The International Brain Laboratory*, 2022a).

187 **Probe track tracing and alignment**

188 Neuropixels probe tracks were manually traced to yield a probe trajectory using Lasagna (Campbell et al., 2020), a Python-
189 based image viewer equipped with a plugin tailored for this task (see Appendix 6 of (*The International Brain Laboratory*,
190 2022a)). Traced probe track data was uploaded to an Alyx server (Rossant et al., 2021); a database designed for experimental
191 neuroscience laboratories. Neuropixels channels were then manually aligned to anatomical features along the trajectory using
192 electrophysiological landmarks with a custom electrophysiology alignment tool developed by (Faulkner, 2020). Another example
193 use case of the alignment tool can be found in (Liu et al., 2021).

194 **Spike sorting**

195 The spike sorting pipeline used at IBL is described in detail in (*The International Brain Laboratory*, 2022b). Briefly, spike sorting
196 was performed using a modified version of the Kilosort 2.5 algorithm (Steinmetz et al., 2021). We found it necessary to improve
197 the original code in several aspects (scalability, reproducibility, and stability, as discussed in (*The International Brain Laboratory*,
198 2022a)), and developed an open-source Python port; the code repository is here: (*The International Brain Laboratory*, 2021b).

199 **Coverage**

200 The coverage has been evaluated by taking a cylinder of radius 354 μm around each probe insertion. Voxels that fall within this
201 cylinder are denoted as “recorded”. We iterated through all insertions in our Brainwide map recordings to evaluate how many
202 voxels have been targeted by one or more insertions. We recorded 90% of the targeted volume with at least 1 probe and 76.5 %
203 with at least 2 probes.

204 Data quality

205 Sessions and insertions

206 Electrophysiology sessions were included if the mice performed at least 400 trials, with a performance of at least 90% correct on
207 easy contrasts (across block types).

208 Also, sessions were included if they reached a set threshold on defined hardware tests (definitions can be found here).
209 Sessions for which less than 200 trials passed trial exclusion criteria (see below) are released but marked as 'FAIL' in the
210 *TrialEvents* column of this general QC spreadsheet.

211 Insertions were disregarded if the raw neural data presented major artifacts, or if the probe tract could not be recovered
212 during the histology procedure. Furthermore, insertions are included only if their alignments have been resolved (see Appendix
213 6 of (*The International Brain Laboratory*, 2022a) for definition).

214 Trials

215 The following trial exclusion criteria were considered: Any of the following trial events could not be detected, and hence is
216 set to NaN: choice, probabilityLeft, feedbackType, feedback times, stimOn times, firstMovement times. The difference between firstMovement times and stimOn times is not in the range of 0.08 and 2.0
217 seconds. Note that the data for trials not passing these criteria is still released. The exclusion criteria were solely used to assess
218 which sessions are marked as 'FAIL' (see above).

219 Detailed trials quality control outputs and the associated code can be found in the associated github repository.

221 Neurons and brain regions

222 Units output by the spike sorting pipeline were marked as bad units if they failed one of the three criteria described in (*The*
223 *International Brain Laboratory*, 2022b): amplitude > 50 uV ; noise cut-off < 20 ; refractory period violation. Units that pass all the
224 three criteria are considered good for analysis and are termed "neurons" in this document. Note that we computed the single
225 unit quality metrics for every single recording session, regardless of the raw ephys or spike sorting quality outcomes (described
226 in sections below). Out of the 295501 units collected, 32766 were considered neurons. We released all units as part of the data
227 release. Instructions on how to filter for good units upon data downloading can be found at this url.
228

Brain regions were included in the final count only if they had at least 2 included sessions and at least 10 neurons.

229 Wheel data

230 For the wheel position data derived from the rotary encoder, we checked whether a timestamp has been assigned to each wheel
231 movement, and whether any of the data sets contains only NaN values. The wheel data for all sessions included in this release
232 passed our quality control checks, as indicated in the *Wheel* column of the general QC spreadsheet.

233 Detailed wheel quality control outputs and the associated code can be found in the associated github repository.

234 Video data

235 Video data were recorded from three separate cameras per session and analyzed using an automated pipeline. We provide an
236 overview of the pipeline here, full details can be found in the IBL video hardware and software white paper (*The International*
237 *Brain Laboratory*, 2022c). The data viewer at <https://viz.internationalbrainlab.org/app> can be used to visually inspect raw videos
238 overlaid with pose estimation results from individual datasets.

239 The IBL ephys rig contains three cameras: a LEFT camera (1280x1024, 60 FPS) and a RIGHT camera (640x512, 150 FPS) that
240 record the mouse's face, paws, and upper body from the side, and a BODY camera (640x512, 30 FPS) that records the mouse's
241 body in the holder from above. We use DeepLabCut (DLC) (Mathis et al., 2018) to perform pose estimation on the three videos
242 from each session (Fig 4). We also compute the motion energy of the whisker pads and the body, the times of lick events and the
243 pupil diameter as additional behavioral features. Below we describe various automated quality control (QC) metrics computed
244 on the raw videos and the DLC outputs.

245 The overall video and DLC QC outcomes for each camera can be found in the respective columns in the general QC spreadsheet.
246 Note that for 44 sessions the data for one or more cameras was not recorded or not released due to severe issues, and is indicated
247 as 'MISSING'. Otherwise, the values 'PASS', 'WARNING' or 'FAIL' indicate the worst of all QC checks for a given camera and data
248 type. The full list of QC outcomes for each camera can be found in the associated github repository.

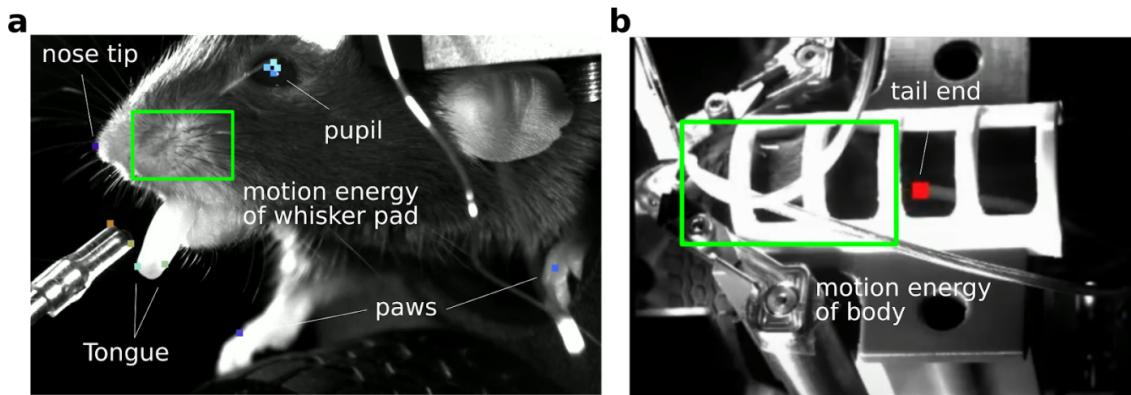


Figure 4. Video analysis pipeline outputs: **a.** Frame of left camera video with location of the points that are routinely tracked for LEFT and RIGHT videos. Green rectangular region indicates the whisker pad region used to compute motion energy. **b.** Frame of BODY camera video with the location of the tail end which is tracked in this view. Green rectangle indicates the region for which the motion energy of the animal's body is calculated.

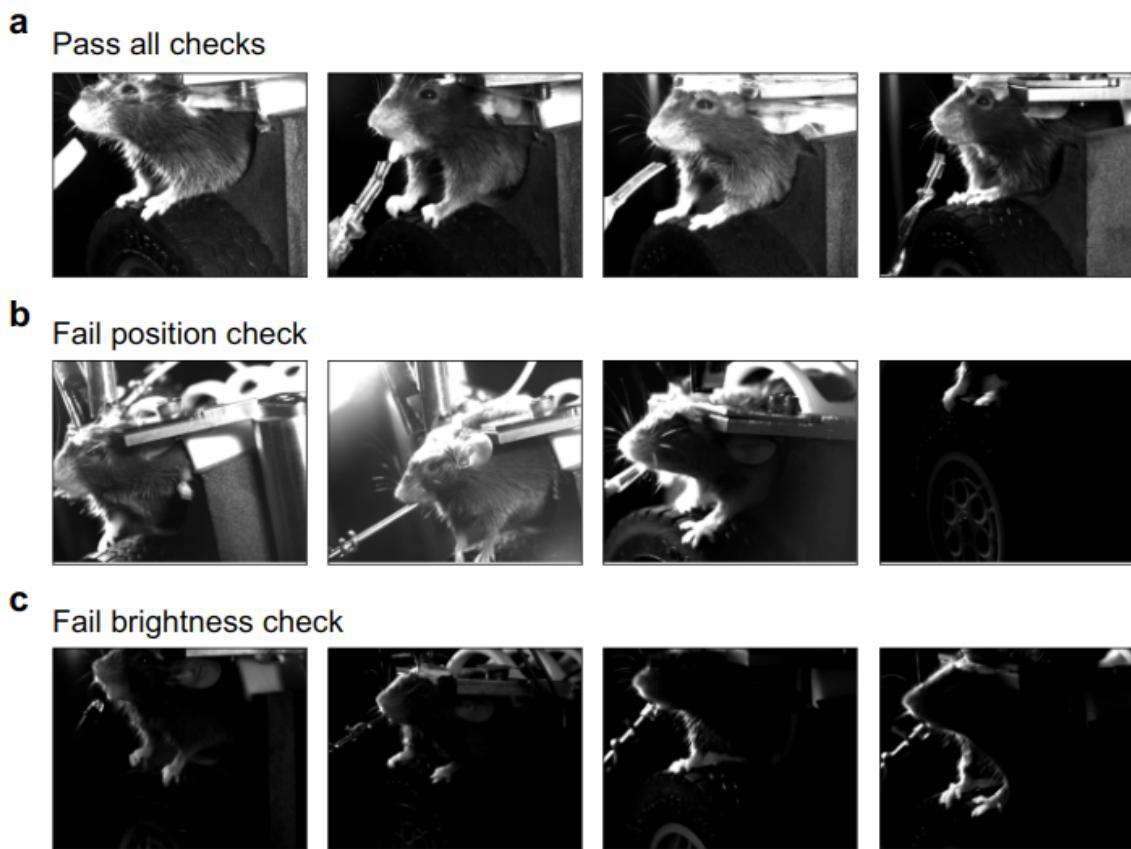


Figure 5. Raw video QC examples from left camera: **a.** Example frame from each of four different sessions that pass all raw video QC checks. **b.** Examples that fail the position check. **c.** Examples that fail the brightness check.

249 Raw video QC

250 We report a standardized set of QC metrics for each video, such as whether a timestamp can be associated with each frame
251 of the video; whether the mouse is positioned correctly in the camera frame; and whether the camera brightness is adequate.
252 Figure 5 displays example frames of LEFT videos from several sessions that either pass all checks or fail a specific check.

253 All raw video QC metrics are defined here and further described in the aforementioned IBL video white paper (*The International*
254 *Brain Laboratory*, 2022c). We also report an overall outcome for each video based on these metrics. Below are the number of
255 videos from each of 354 sessions that have an overall QC outcome of PASS/WARNING/FAIL/MISSING:

- 256 • RIGHT: 33/263/18/40
- 257 • LEFT: 43/236/43/32
- 258 • BODY: 69/235/9/41

259 Many of the WARNING/FAIL checks result for videos that are not positioned ideally or are too dark; these videos can
260 nevertheless be run through the pose estimation pipeline (described below) and often lead to reasonable behavioral features
261 (see for example top panels of Fig 6), which have their own set of QC metrics. Some of the WARNING/FAIL checks stem from
262 issues in the alignment between the video data and the behavioral data. While these videos can also be run through pose
263 estimation, aligning the resulting behavioral features to trial-based events or neural activity should be performed with care.
264 Details on the outcomes for each QC check can be found in the associated github repository.

265 Pose estimation QC

266 We report a standardized set of QC metrics for the DLC traces, such as a check that the length of the DLC traces matches the
267 length of the video. All DLC QC metrics are defined here and their outcomes can be found in the associated github repository.

268 **Paws.** Paw movements feature prominently in several IBL analyses, and as such we provide a specific QC metric for paw tracking.
269 For a given paw in one video, we compute the fraction of time points where paw likelihood falls below a given threshold in a
270 window of data aligned to stimulus onset for each trial (Fig 6). This fraction is then used to assign PASS/FAIL/NOT_COMPUTED
271 to the tracking of each paw in each video. This procedure is performed for both paws separately.

- 272 • RIGHT close paw: 282/8/64
- 273 • RIGHT far paw: 231/59/64
- 274 • LEFT close paw: 287/7/60
- 275 • LEFT far paw: 247/47/60

276 Overall tracking is better for the paw closest to the camera (“close paw” above), as the far paw is more often occluded by the body
277 or experimental equipment. Many of the FAIL cases reported here are due to underlying raw video issues (position, brightness,
278 etc.).

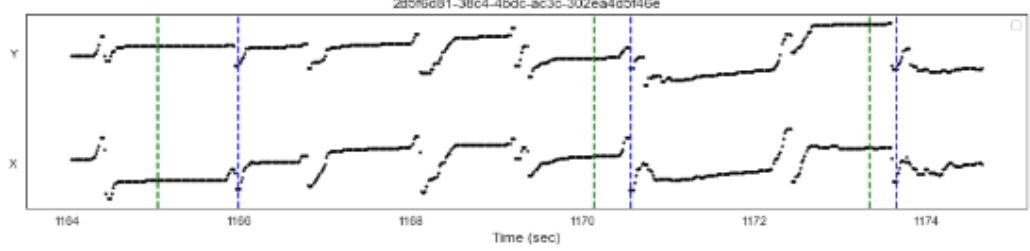
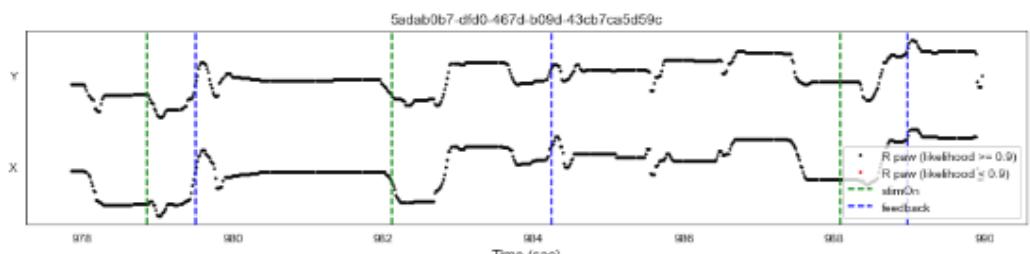
279 **Pupil.** We also provide an estimate of the pupil diameter from the LEFT and RIGHT cameras. Pupil diameter is estimated
280 using the top, bottom, right and left pupil markers after dropping low-likelihood markers. The resulting 1-dimensional trace
281 y is noisy and often contains missing data. We compute a smoothed version of this trace \hat{y} using a Savitzky-Golay filter that
282 interpolates through missing data. Both y and \hat{y} are made available. We estimate the signal-to-noise ratio (SNR) by computing
283 $SNR = \text{var}(\hat{y})/\text{var}(\hat{y} - y)$ (code can be found here). An SNR threshold is used to assign PASS/FAIL/NOT_COMPUTED to each
284 pupil diameter estimate.

- 285 • RIGHT: 48/286/20
- 286 • LEFT: 76/271/7

287 The pupil diameter is not well tracked in our current pipeline, often due to the quality of the raw videos and/or the presence of
288 moving whiskers and infrared light reflections on the eye (Fig 7). Pupil diameter estimates are somewhat more reliable from
289 the LEFT camera, which has higher spatial resolution than the RIGHT camera. We are working to improve pupil diameter
290 estimation from the video data.

a

Pass paw QC check

**b**

Fail paw QC check

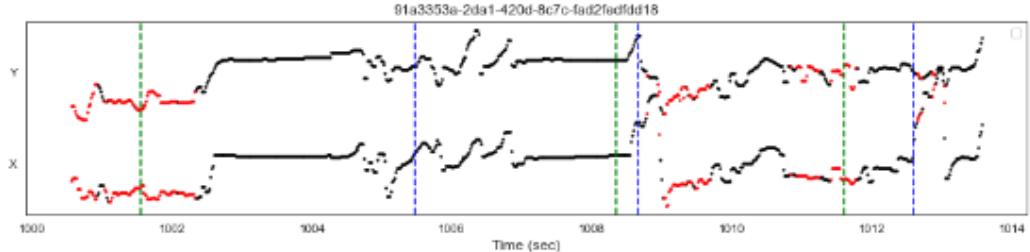
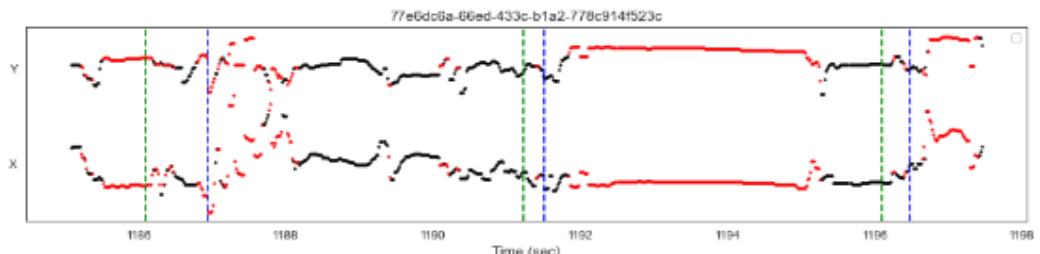


Figure 6. Paw QC examples from left camera: **a.** Example frame and traces from each of two different sessions that pass paw QC check. Top line is paw y-coordinate, bottom line is paw x-coordinate; high likelihood points (likelihood ≥ 0.9) are plotted in black, low likelihood points (<0.9) are plotted in red. Green dashed vertical lines indicate stimulus onset, blue dashed vertical lines indicate feedback onset. **b.** Examples that fail the paw QC check.

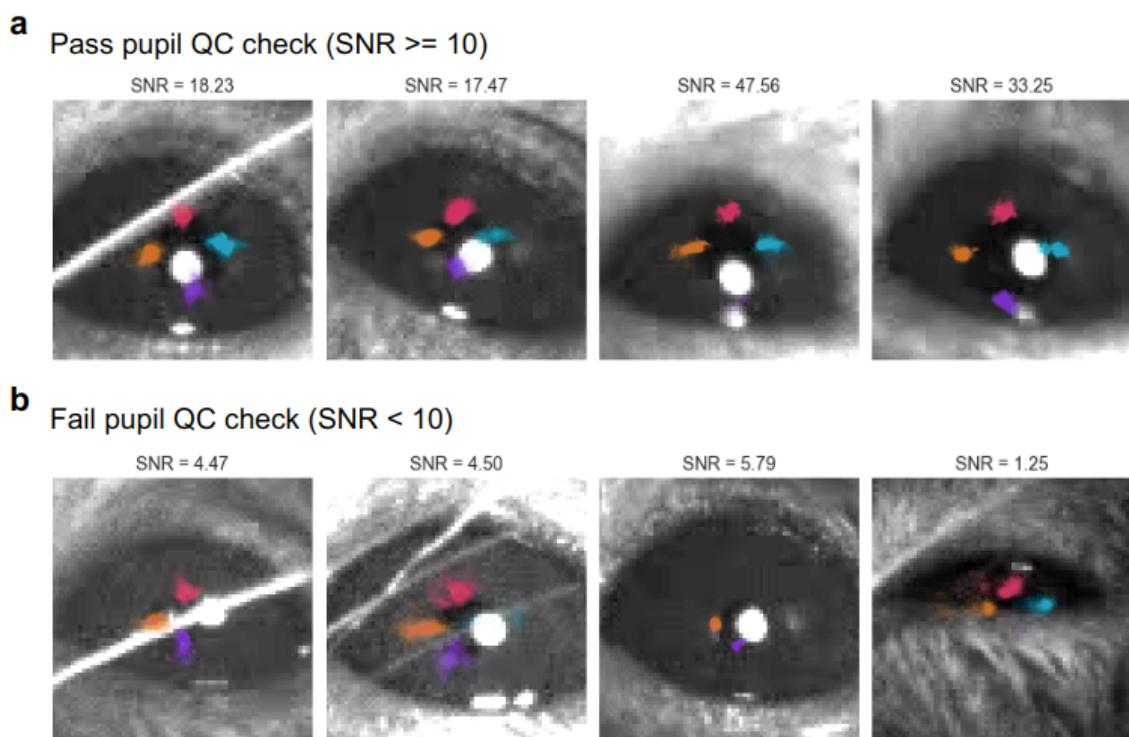


Figure 7. Pupil QC examples from left camera: **a.** Example frame from each of four different sessions that pass pupil QC check, overlaid with tracked points across the session. **b.** Examples that fail the pupil QC check. Frames have been brightened to increase pupil visibility.

291 **Electrophysiology data**

292 We visually inspected the raw data as well as the spike sorting output, and labelled them as PASS, WARNING or FAIL based
293 on criteria defined below. Overall, 92 % of the raw ephys data was marked as PASS, 6 % as WARNING and the remaining 1 %
294 judged as FAIL was not released. For the spike sorting output however, we decided to release the severe cases as the raw data
295 appears of reasonable quality. 62 % of the spike sorting output was marked as PASS, 17 % as WARNING, and 20 % as FAIL.

296 Detailed output table (CSV) can be found at https://github.com/int-brain-lab/paper-brain-wide-map/tree/main/data_checks,
297 and in the general QC spreadsheet.

298 **Raw ephys data QC**

299 Specifically, we inspected 40ms snippets of raw data taken at three different time points in the recording : T=600, 1800 and 3000
300 seconds. The raw data was plotted using a gray-scale along the probe depth (example Fig 8); dark and white colors indicate a
301 shift from baseline. Spikes detected were overlaid (good units in green, and multi-units in red). Brain regions were displayed
302 along the probe depth next to the gray image, to determine whether noisy-presenting channels were indeed inside the brain and
303 should be considered as such.

304 When inspecting the raw data, we highlighted these issues:

- 305 • Recordings with noisy channels (example Fig 9, where the top channels are noisy)
- 306 • Recordings with brief regular line noise artifacts (example Fig 10, see the vertical stripy bands appearing regularly).
- 307 • Recordings with brief regular artefacts appearing on part of the channels (example Fig 11, channels 140-160)
- 308 • Recordings with brief artifacts spanning the whole probe (example Fig 12 at T=523s)

309 Recordings presenting such defaults are all marked as WARNING in the associated QC spreadsheet (column "Quality (Raw data)").

310 More severe cases were removed from the data release.

311 **Spike sorting QC**

312 When checking the spike sorting output, we looked both at the matching between the raw data and the spikes detected, and
313 at the spikes raster overall (see Fig 13 for an example raster). We highlighted these issues, of which we show severe cases as
314 examples:

- 315 • Recordings with a sharp discontinuity in the raster ; labelled as Discontinuity
 - 316 1. Recordings that appear stable, but have a sudden loss of spikes (example Fig 14)
 - 317 2. Recordings that appear stable, but have sudden appearance of spikes (example Fig 15)
318 Note: In some cases the raster presents major discontinuities, but the raw data appears stable throughout the
319 recording (example Fig 20 and Fig 22)
 - 320 3. Recordings that appear stable, but have sudden horizontal artifacts (example Fig 16) ; also labelled as Horizontal
321 bands
- 322 • Recordings with "smooth" loss / gain of spikes over time (example Fig 17) ; labelled as Loss of spikes
- 323 • Recordings with jittery-looking raster (example Fig 18) ; labelled as Jitter
324 Note: Often these recordings have stable raw data throughout the recording, but the spike detection does not seem to
325 capture the underlying activity seen in the raw data (example Fig 18 and Fig 19)
- 326 • Recordings presenting an abnormal band of spikes, not obviously linked to the activity present in the raw data (example
327 Fig 23-25) ; labelled as Horizontal bands
328 Note: This issue is likely due to the whitening process of the spike sorting algorithm.

329 Insertions presenting such issues are marked as WARNING or FAIL in the associated QC spreadsheet (column "Quality (Spike
330 sorting)"), depending on their severity. Insertions presenting such issues in a very mild manner, or not at all, were marked as
331 PASS otherwise.

332 We also have noticed issues with the spike detection, that we did not label systematically:

- 333 • Cases of putative single spikes being detected several times (example Fig 26)
334 • Cases where the raw data presents activity, but the spikes detected are nearly null (example Fig 27)

335 In short, it appears that the current spike sorting pipeline has considerable room for improvement, and users of this data
336 should keep these issues in mind and use the data viewer at <https://viz.internationalbrainlab.org/app> to check individual
337 datasets as needed. We hope to correct these issues in future data releases.

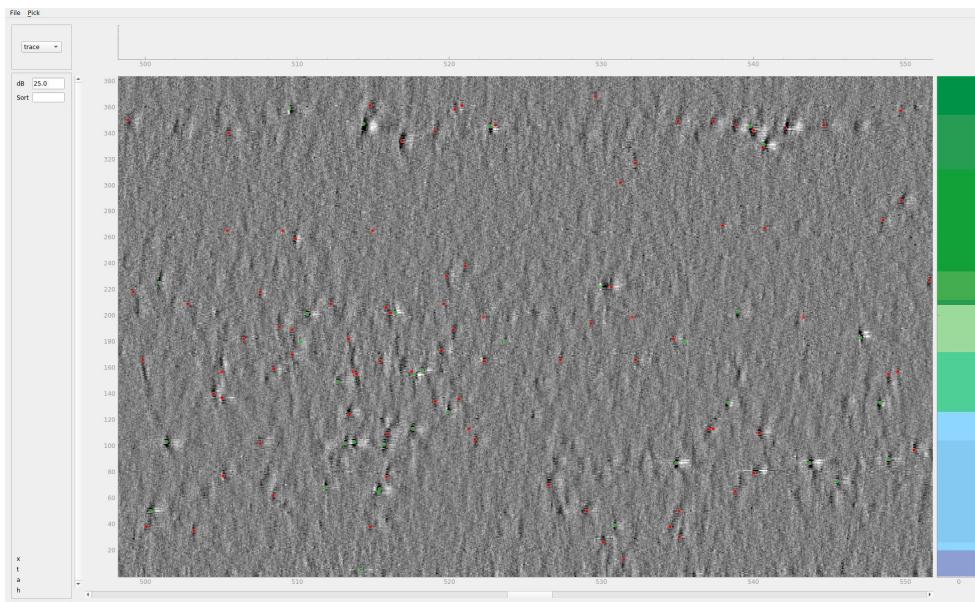


Figure 8. Example of good raw data. We present over time (x-axis) 40ms of raw data, displayed in a gray-scale for each channels on the probe (y-axis). The colors on the right indicate the brain regions crossed by the channels (using the Allen nomenclature). PID: 1bf5c05b-3654-4d5c-bac9-2b96edd12adf. Data displayed at T=1800s in recording.

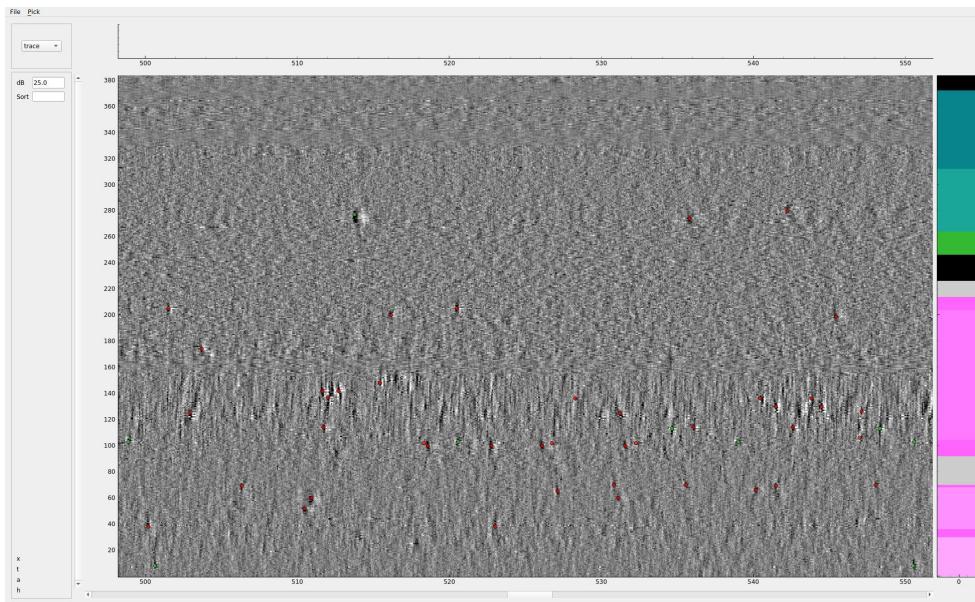


Figure 9. Example of noisy raw data, here where channels at the top (in cortical layers) appear as "blurred". PID: 578ca001-8cf5-4d30-b58f-1d574eaf808a. Data displayed at T=600s in recording.

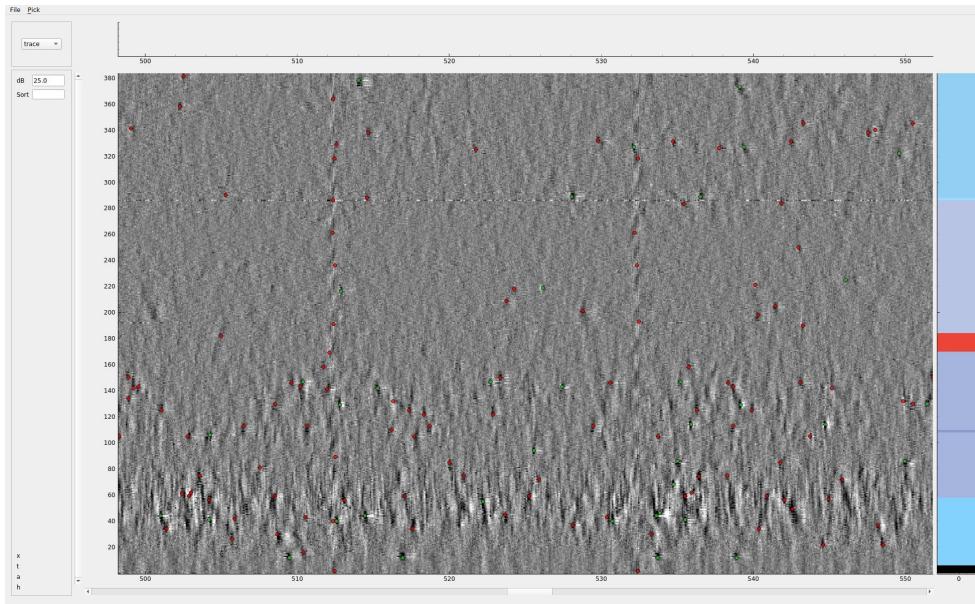


Figure 10. Example of noisy raw data, here displaying line noise. See the vertical stripy bands appearing regularly at a frequency nearing 50Hz. PID: a5f2ec22-0ff3-4249-bd2f-6247c3990e53. Data displayed at T=600s in recording.

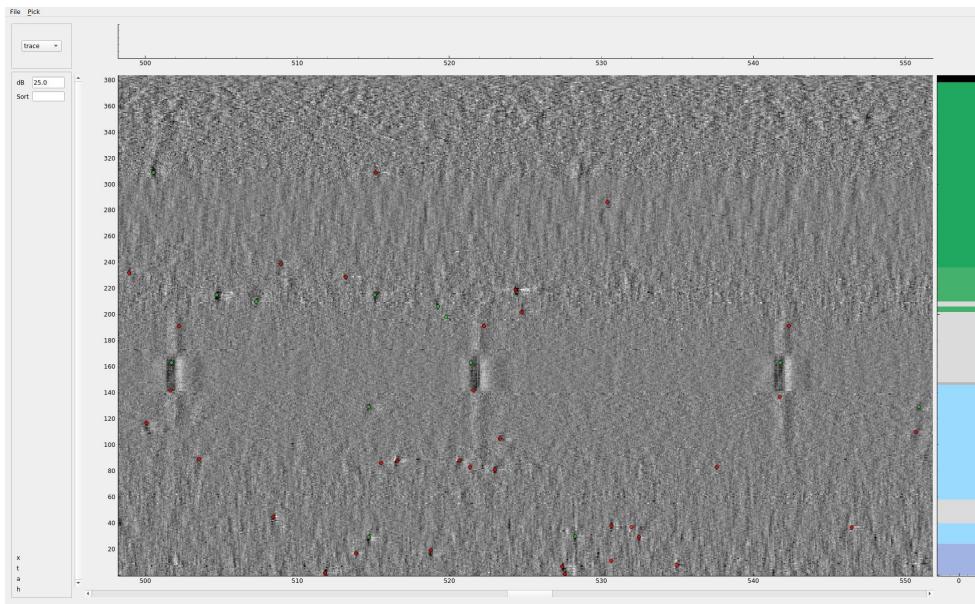


Figure 11. Example of noisy raw data, here displaying artifacts present only on some channels. See the "square" artifact visible on channels 140-160, appearing at regular (line noise) intervals. PID: af2a0072-e17e-4368-b80b-1359bf6d4647. Data displayed at T=600s in recording.

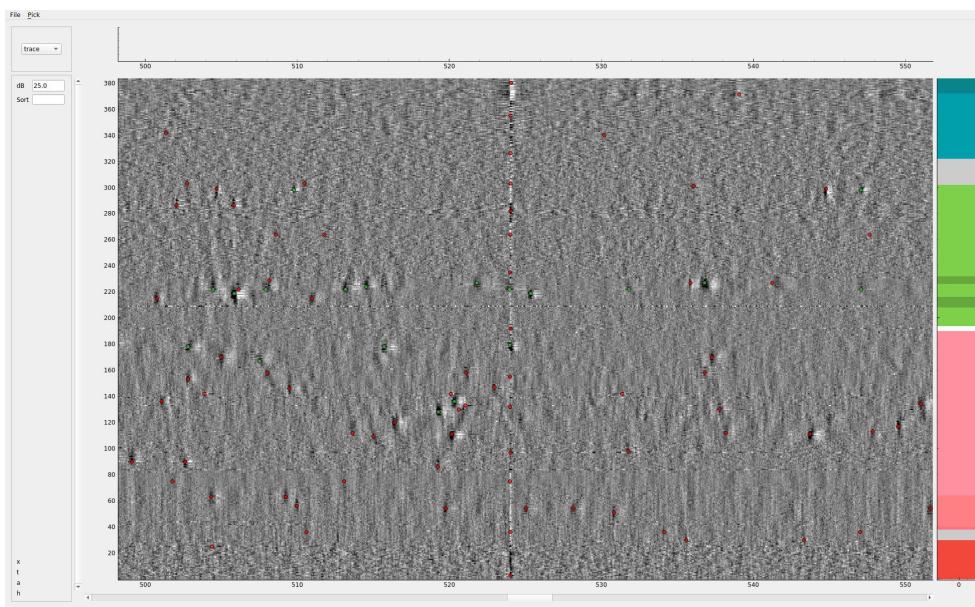


Figure 12. Example of noisy raw data, here presenting a brief artifact spanning the whole probe. See the vertical line at x=523. PID: c07d13ed-e387-4457-8e33-1d16aed3fd92. Data displayed at T=600s in recording.

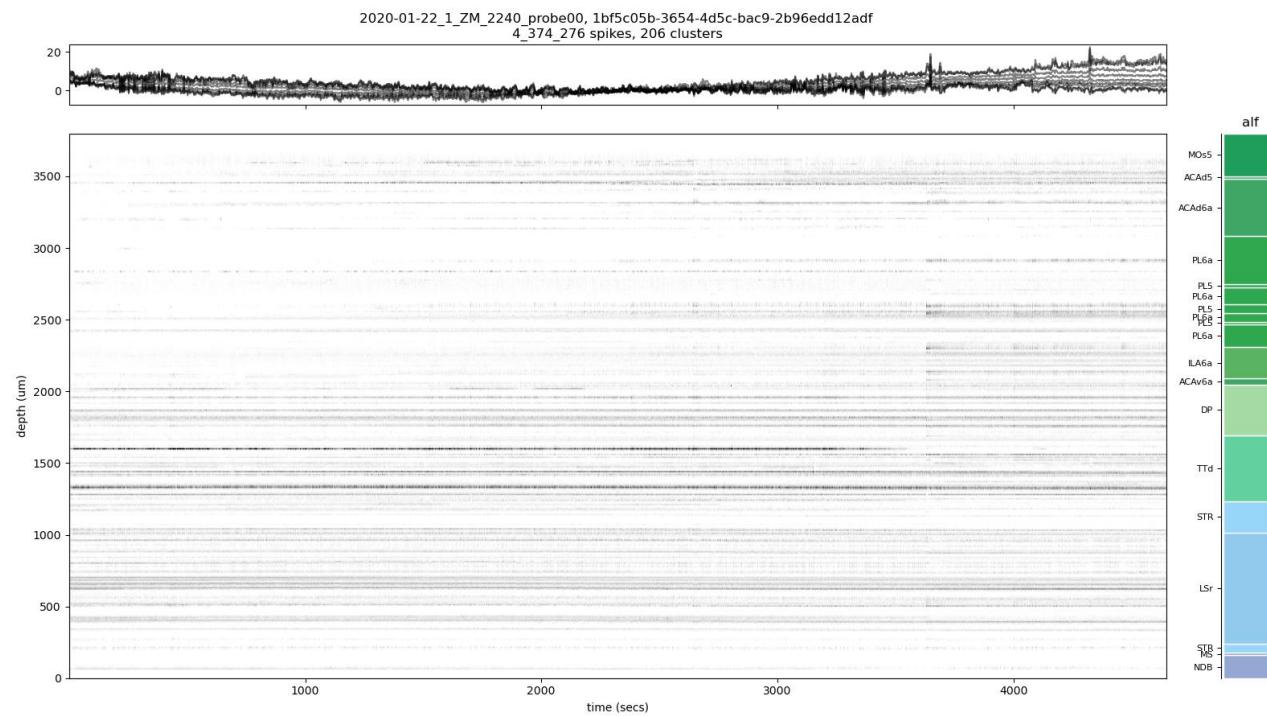


Figure 13. Example of good raster (good units). On top of the raster, we show the drift correction applied as part of the spike sorting processing (non-rigid drift correction applied here; each trace corresponds to the drift correction applied to a spatial subset of the probe). On the right, the brain regions crossed at labelled (as per the Allen nomenclature). PID: 1bf5c05b-3654-4d5c-bac9-2b96edd12adf

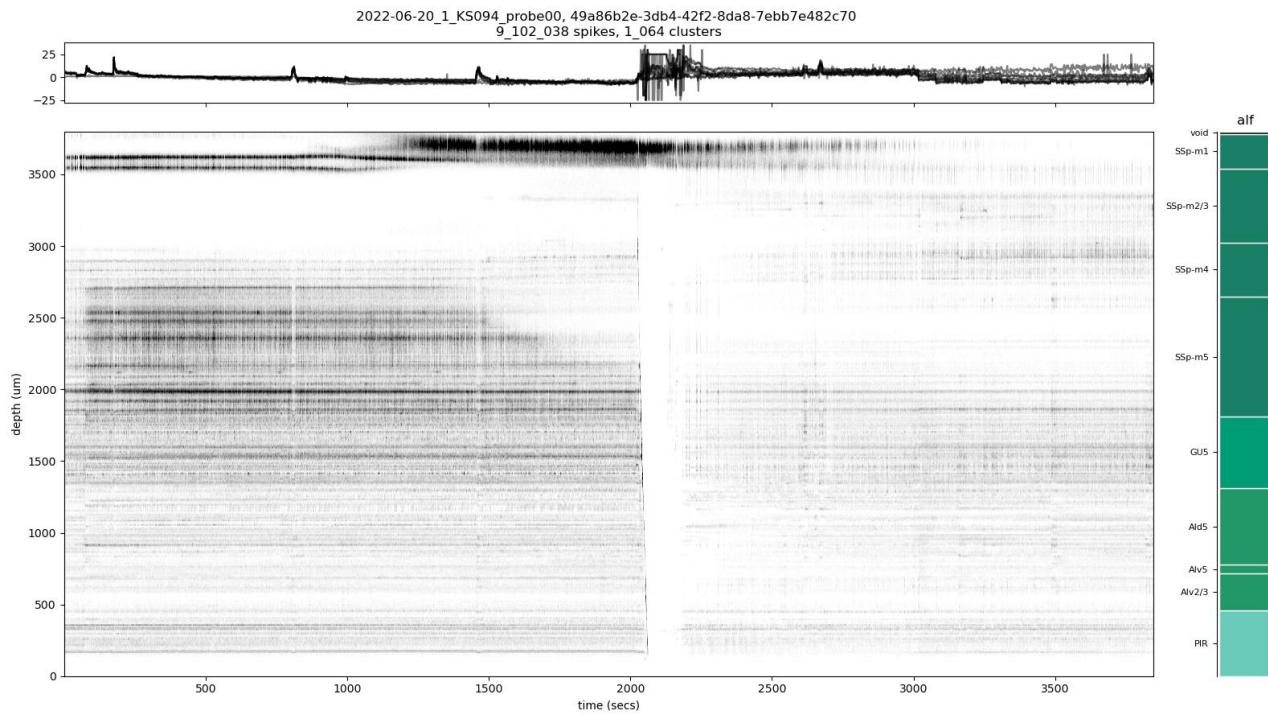


Figure 14. Example of a raster presenting a discontinuity, here a sudden loss of spikes (all units). See the gap in spikes after T=2000s. PID: 49a86b2e-3db4-42f2-8da8-7ebb7e482c70

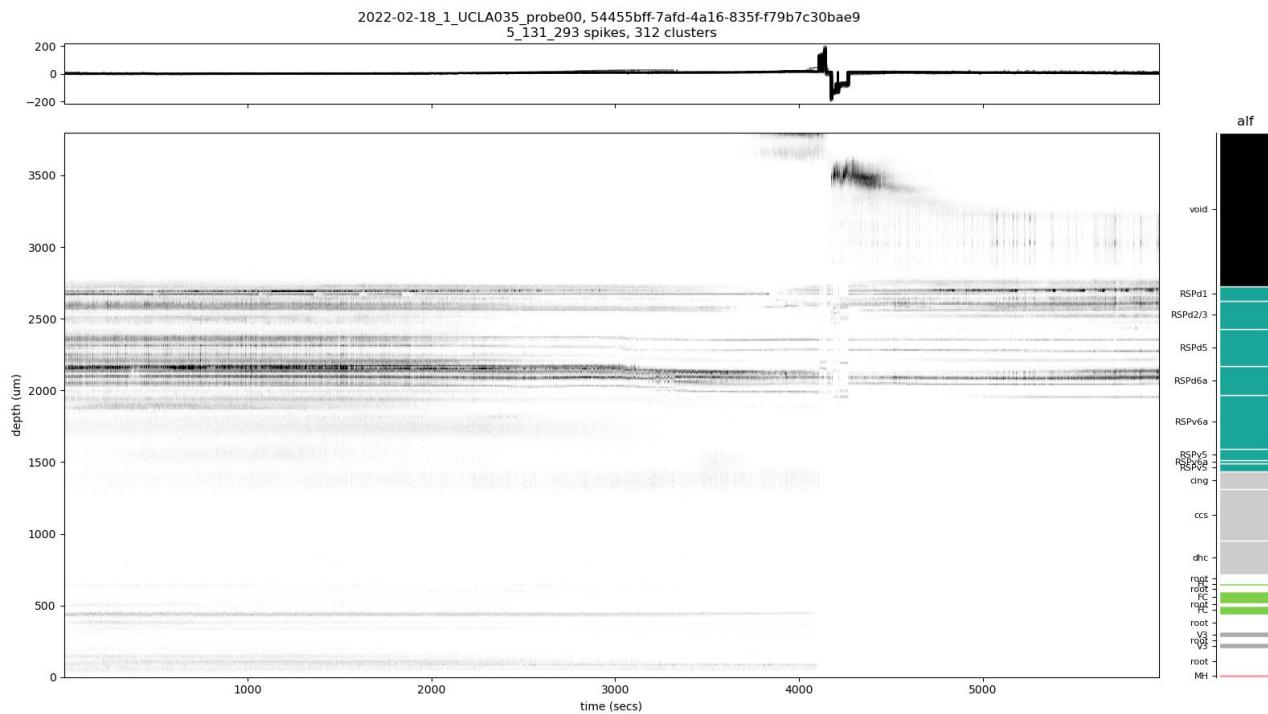


Figure 15. Example of a raster presenting a discontinuity, here a sudden gain of spikes (all units). See the loss and rebound in spikes after T=4000s, especially near depth=3500 μm. PID: 54455bff-7af8-4a16-835f-f79b7c30bae9

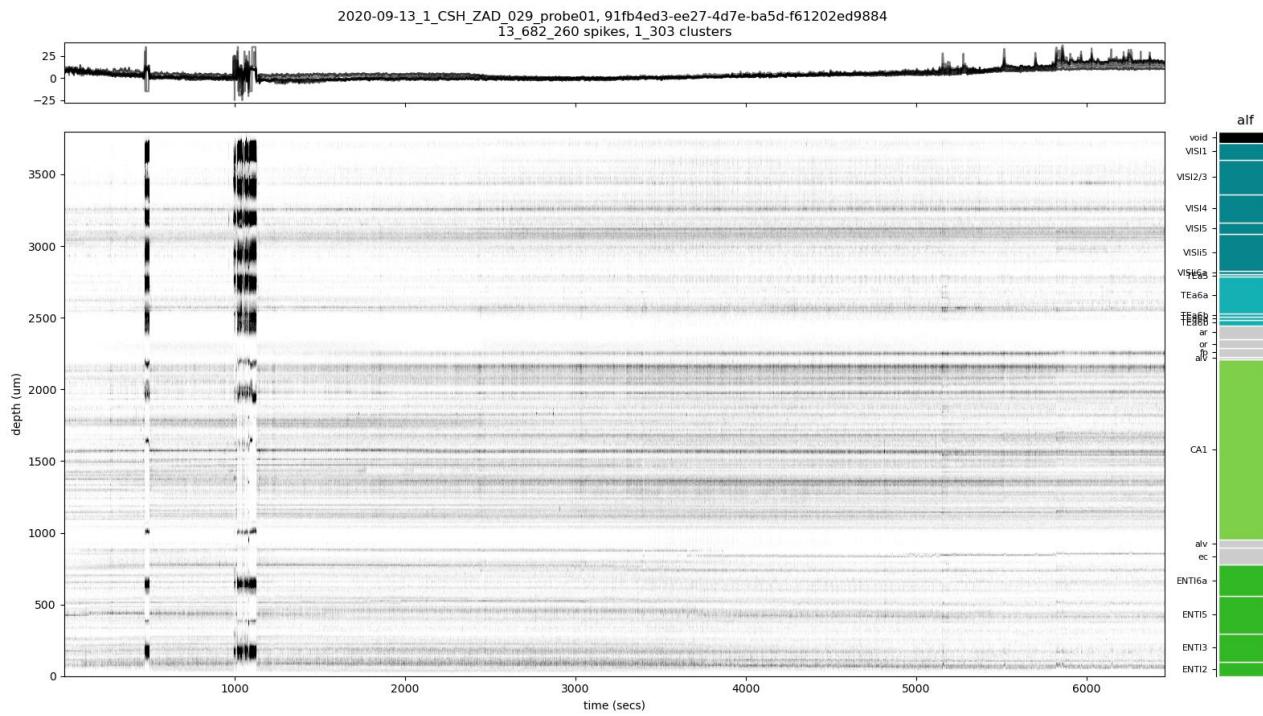


Figure 16. Example of a raster presenting a discontinuity, here sudden horizontal bands (all units). PID: 91fb4ed3-ee27-4d7e-ba5d-f61202ed9884

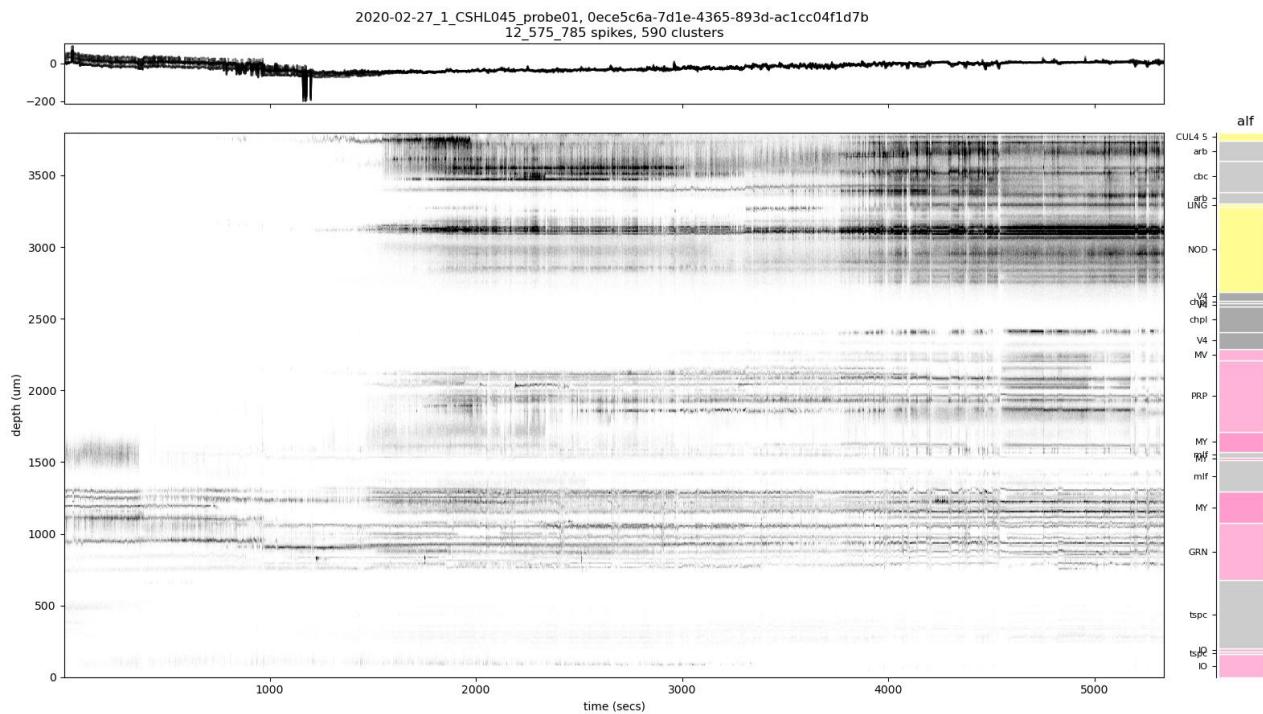


Figure 17. Example of a raster presenting a discontinuity, here a smooth gain in spiking activity over time, visible particularly near depth=2000-3500 μm (all units). PID: 0ece5c6a-7d1e-4365-893d-ac1cc04f1d7b

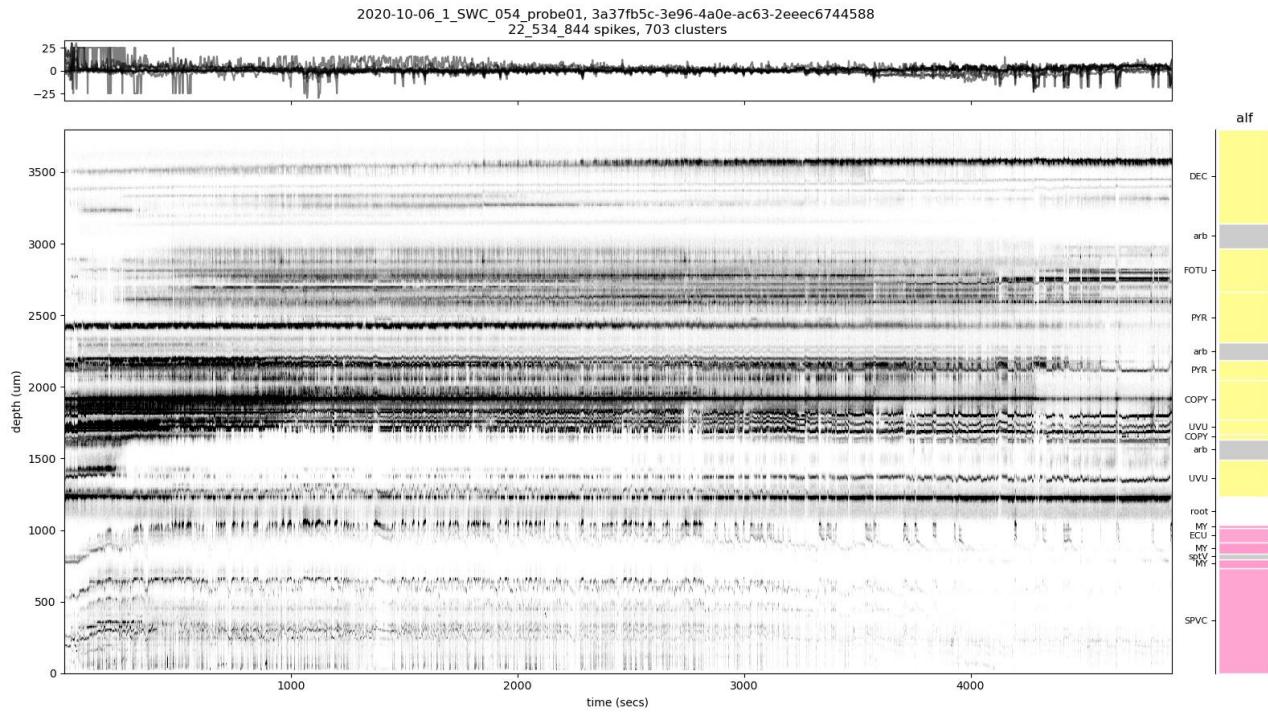


Figure 18. Example of a raster presenting a discontinuity, here a severe case of jitter (all units). See for example near depth=1000 μm.
 PID: 3a37fb5c-3e96-4a0e-ac63-2eeec6744588

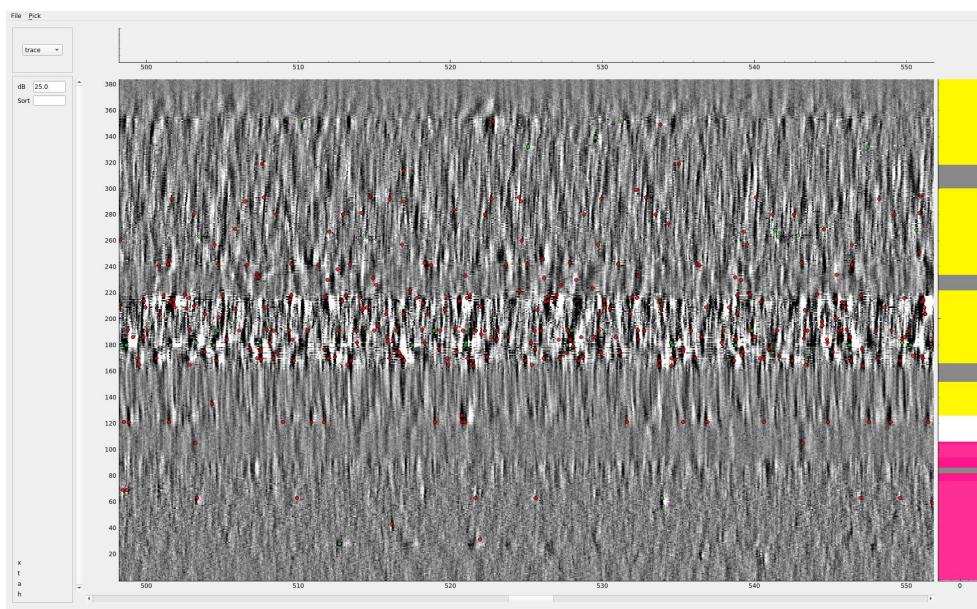


Figure 19. Example of raw data in the case of a jittery-looking raster. The raw data presents highly-active patterns, some of which is not caught by the spike detection. PID: 3a37fb5c-3e96-4a0e-ac63-2eeec6744588, same as in Fig fig:3a37fb5c_, aster. Data displayed at T = 600s in recording.

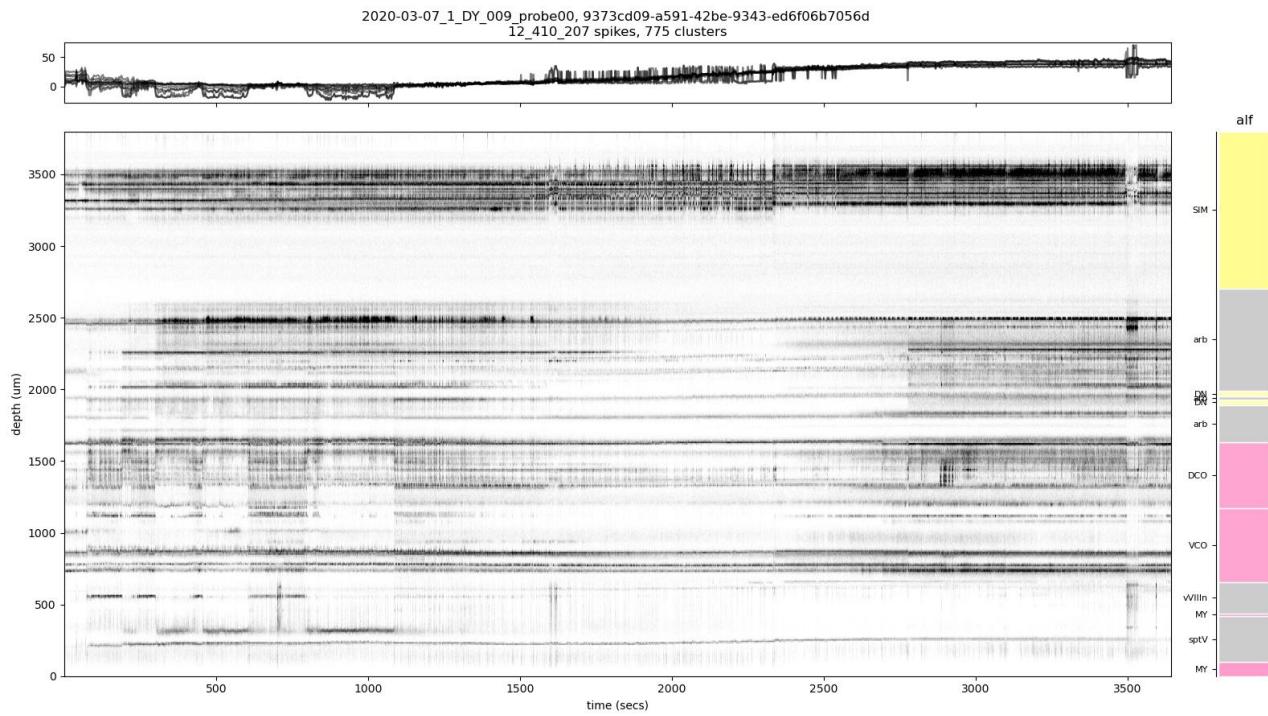


Figure 20. Example of a raster presenting discontinuities, but the raw data appears stable (all units). For example, at time T=600s, there is a clear diminution of activity compared to e.g. at time T=1250s or T=3000s. These sharp changes are not observed in the raw data. See the corresponding figures below, showing raw data at these two extreme time points (600s ; 3000s). PID: 9373cd09-a591-42be-9343-ed6f06b7056d

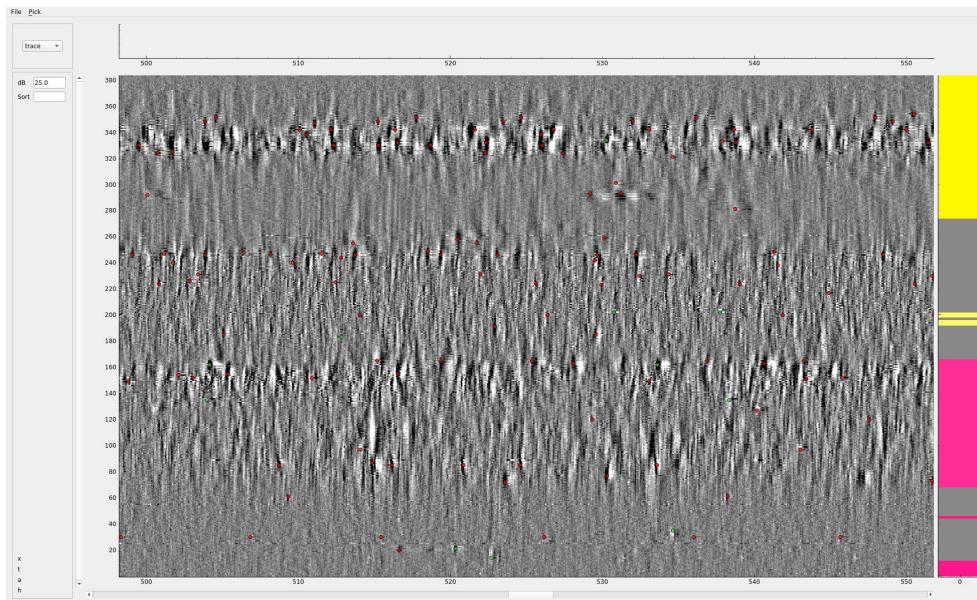


Figure 21. Example snippet of raw data at a time where we see a clear "dip" in the raster of Fig 20. PID: 9373cd09-a591-42be-9343-ed6f06b7056d. Data displayed at T=600s in recording.

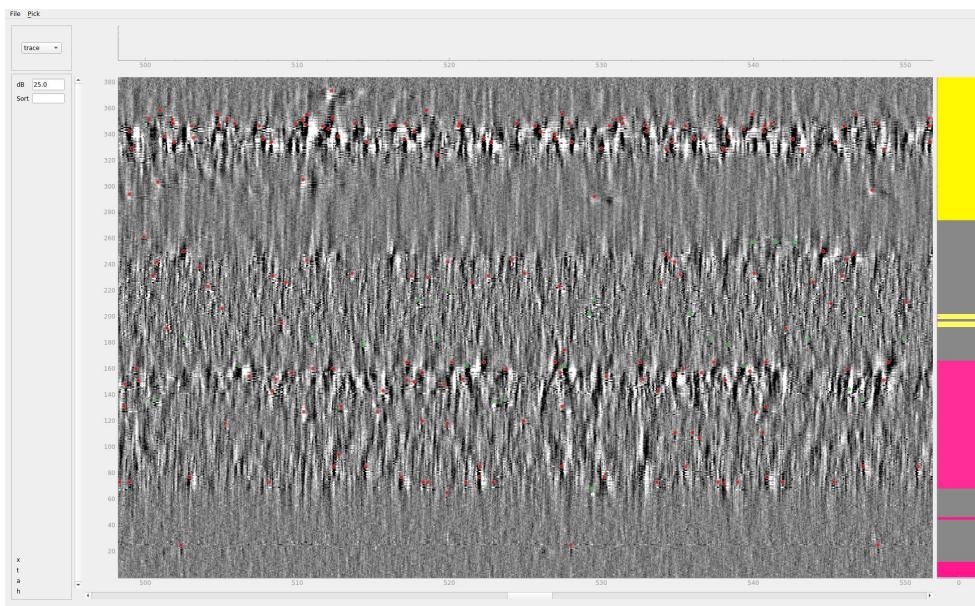


Figure 22. Example snippet of raw data at a later time in the raster of Fig 20. The activity is comparable to that seen in Fig 22. PID: 9373cd09-a591-42be-9343-ed6f06b7056d. Data displayed at T=3000s in recording.

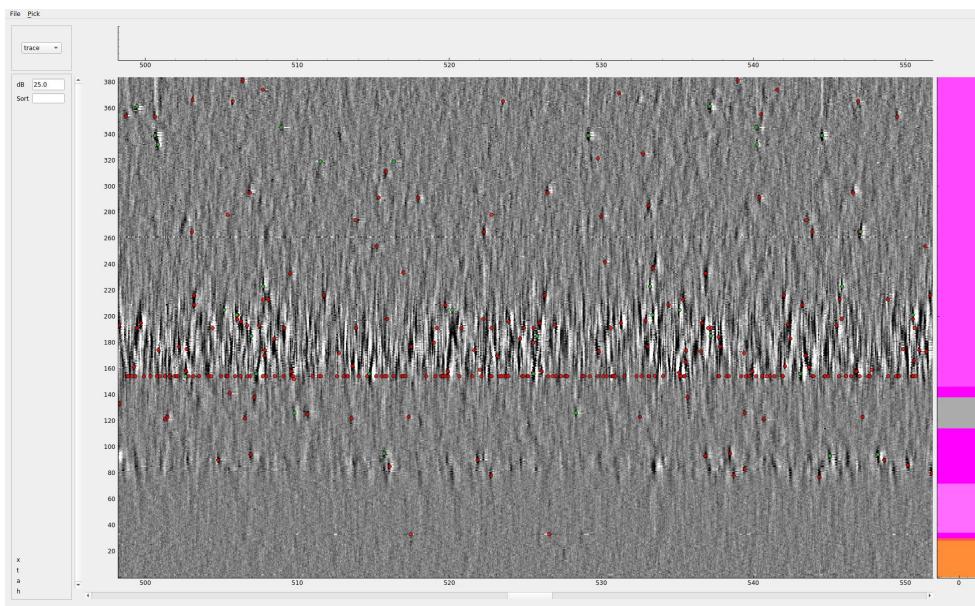


Figure 23. Example of raw data with abnormal spike detection. See the band of red (multi-units) spikes detected near channel 155. These spikes are not correlated with any activity seen on the raw data. PID: 3c283107-7012-48fc-a6c2-ed096b23974f. Data displayed at T=600s in recording.

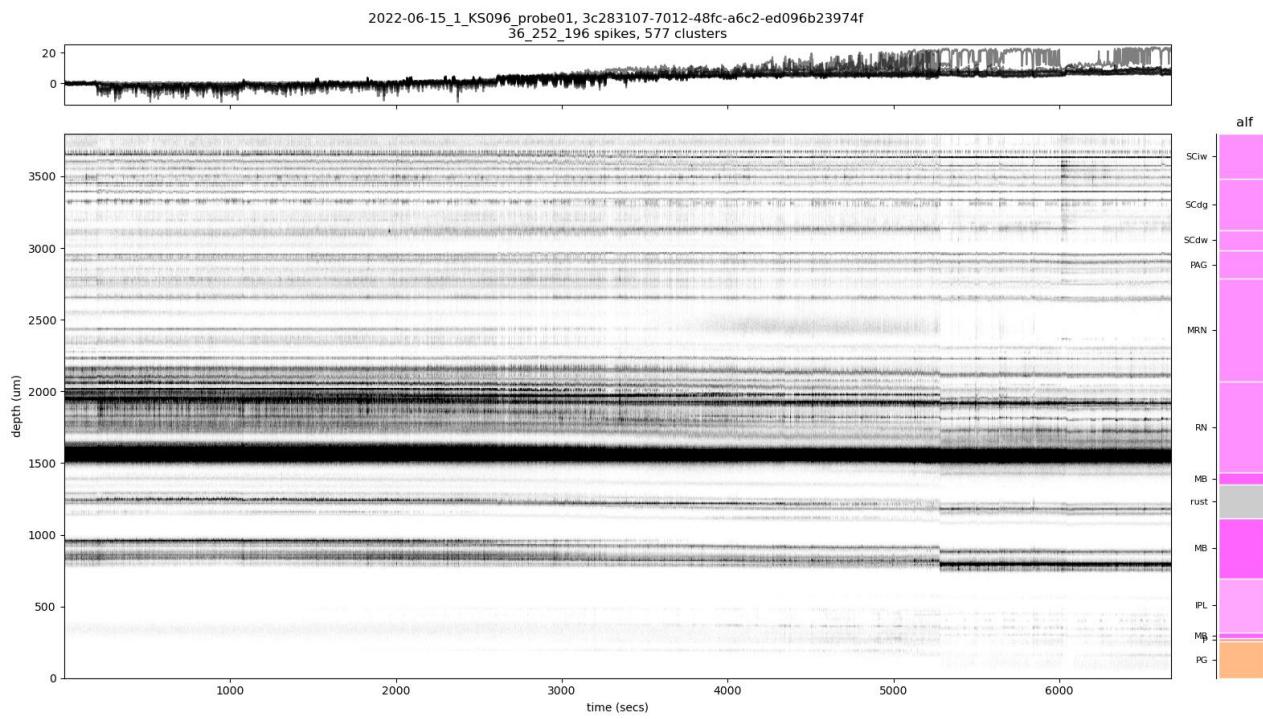


Figure 24. Example of raster with abnormal spike detection (all units). See the dark band near depth $1500\mu\text{m}$, emerging from the poor spike detection visible from the raw data display in Fig 23. PID: 3c283107-7012-48fc-a6c2-ed096b23974f

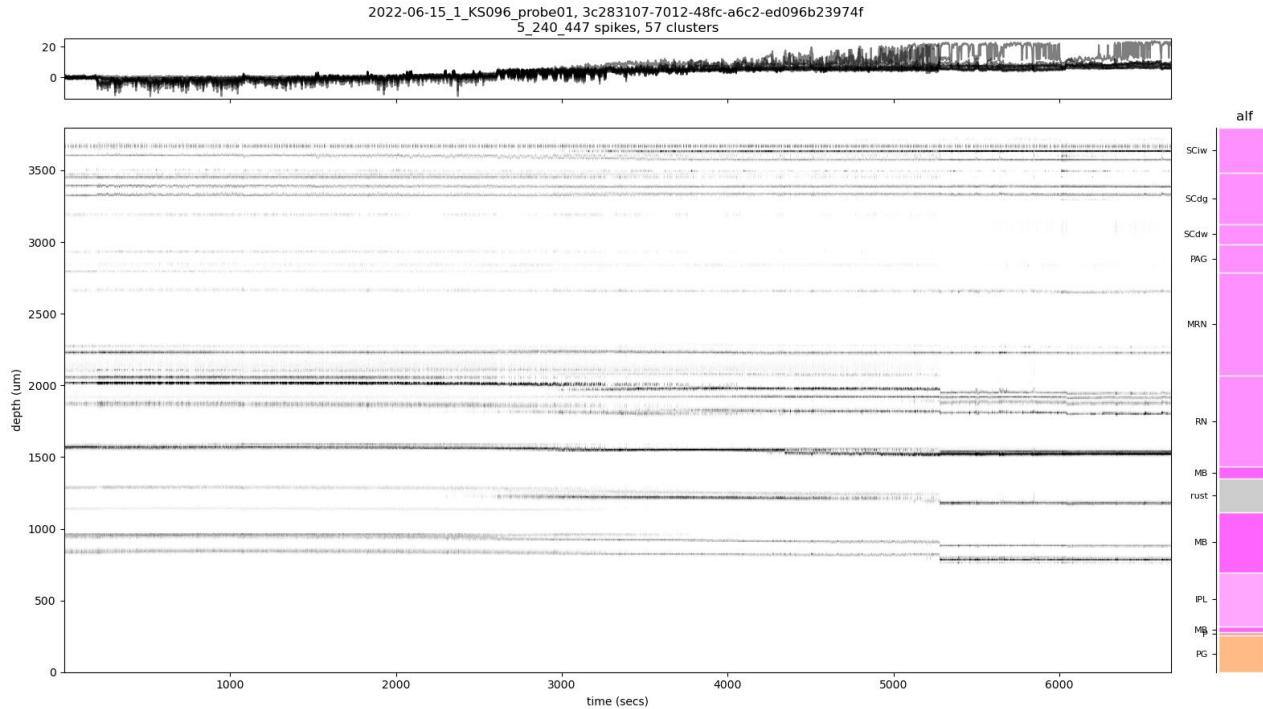


Figure 25. Example of raster with abnormal spike detection (good units). The dark band near depth $1500\mu\text{m}$ that appeared in Fig 24 is removed by the filtering for good units. PID: 3c283107-7012-48fc-a6c2-ed096b23974f

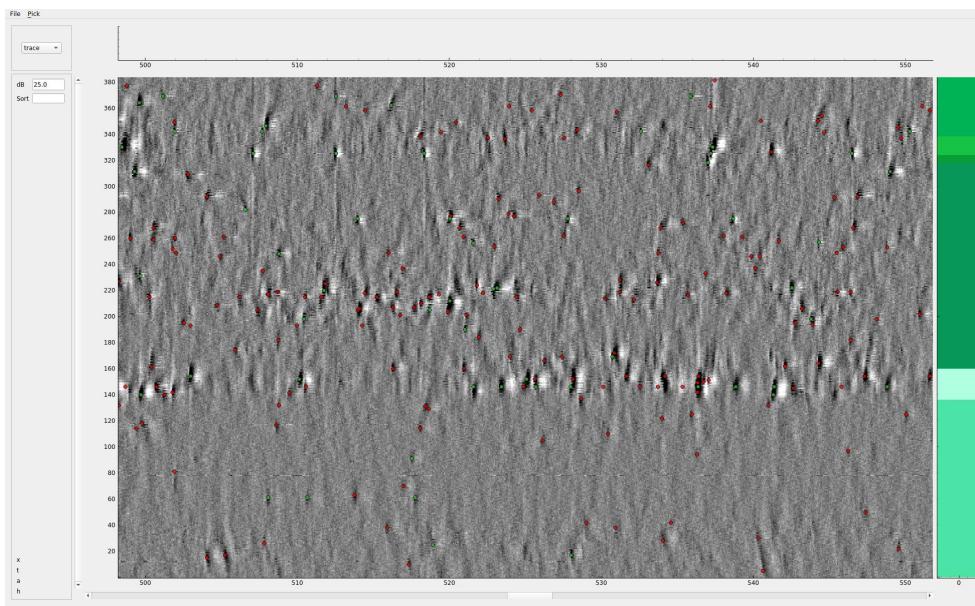


Figure 26. Example of raw data, with putatively single spikes appearing as being detected multiple times. See an example spike at x=532, y=140, with two green dots overlaid (these are spikes associated with good units). PID: df350b09-a420-42bb-9952-17862cbb9f1c Data displayed at T=600s in recording.

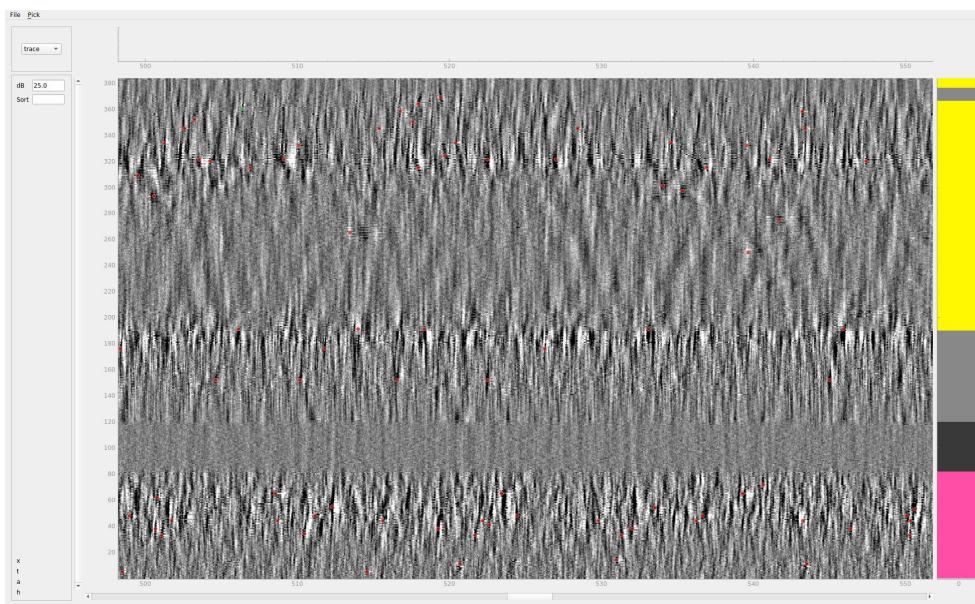


Figure 27. Example of raw data, with low spike detection compared to the activity present in the raw data. PID: 577e4741-4b15-4e91-b81b-61304a09bfb5. Data displayed at T=1800s in recording.

338 Acknowledgments

339 This work was supported by grants from the Wellcome Trust (209558 and 216324), National Institutes of Health (1U19NS123716,
340 including a Diversity Supplement and an Administrative Supplement to Support Enhancement of Software Tools for Open
341 Science) and the Simons Foundation.

342 Diversity Statement

343 We support inclusive, diverse and equitable conduct of research. One or more of the IBL members self-identifies as a member of
344 an underrepresented ethnic minority in science. One or more of the IBL members self-identifies as a member of the LGBTQIA+
345 community.

346 References

- 347 **Campbell R**, BakingTray; 2020. <https://github.com/SainsburyWellcomeCentre/BakingTray>, doi: <https://doi.org/10.5281/zenodo.3631609>.
- 348 **Campbell R**, StitchIt; 2021. <https://github.com/SainsburyWellcomeCentre/StitchIt>, doi: <https://zenodo.org/badge/latestdoi/57851444>.
- 349 **Campbell R**, Blot A, Rousseau C, Winter O, Lasagna; 2020. <https://github.com/SainsburyWellcomeCentre/lasagna>, doi: [10.5281/zenodo.3941894](https://doi.org/10.5281/zenodo.3941894).
- 351 **Econo MN**, Clack NG, Lavis LD, Gerfen CR, Svoboda K, Myers EW, Chandrashekhar J. A platform for brain-wide imaging and reconstruction of
352 individual neurons. *eLife*. 2016; 5(e10566). doi: [10.7554/eLife.10566](https://doi.org/10.7554/eLife.10566).
- 353 **Faulkner M**, Ephys Atlas GUI; 2020. <https://github.com/int-brain-lab/iblapps/tree/master/atlacelectrophysiology>.
- 354 **Jun JJ**, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, Lee AK, Anastassiou CA, Andrei A, Aydin Ç, et al. Fully integrated silicon
355 probes for high-density recording of neural activity. *Nature*. 2017; 551(7679):232–236.
- 356 **Klein S**, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity based medical image registration. *IEEE Transactions on
357 Medical Imaging*. 2010; 29(1):196–205. doi: [10.1109/TMI.2009.2035616](https://doi.org/10.1109/TMI.2009.2035616).
- 358 **Liu LD**, Chen S, Hou H, West SJ, Faulkner M, Economo MN, Li N, Svoboda K, the International Brain Laboratory. Accurate localization of linear
359 probe electrode arrays across multiple brains. *eNeuro*. 2021; 8(6). doi: [10.1523/ENEURO.0241-21.2021](https://doi.org/10.1523/ENEURO.0241-21.2021).
- 360 **Lopes G**, Bonacchi N, Frazão J, Neto JP, Atallah BV, Soares S, Moreira L, Matias S, Itsikov PM, Correia PA, et al. Bonsai: an event-based framework
361 for processing and controlling data streams. *Frontiers in neuroinformatics*. 2015; 9:7.
- 362 **Lopes G**, Farrell K, Horrocks EA, Lee CY, Morimoto MM, Muzzu T, Papanikolaou A, Rodrigues FR, Wheatcroft T, Zucca S, et al. Creating and
363 controlling visual environments using BonVision. *Elife*. 2021; 10:e65541.
- 364 **Mathis A**, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M. DeepLabCut: markerless pose estimation of user-defined body
365 parts with deep learning. *Nature neuroscience*. 2018; 21(9):1281–1289.
- 366 **Ragan T**, Kadiri LR, Venkataraju KU, Bahlmann K, Sutin J, Taranda J, Arganda-Carreras I, Kim Y, Seung HS, Osten P. Serial two-photon
367 tomography for automated ex vivo mouse brain imaging. *Nat Methods*. 2012; 9(3):255–8. doi: [10.1038/nmeth.1854](https://doi.org/10.1038/nmeth.1854).
- 368 **Rossant C**, Winter O, Hunter M, Huntenburg J, Faulkner M, Wells M, Steinmetz N, Harris K, Bonacchi N, Alyx; 2021. <https://github.com/cortex-lab/alyx>.
- 370 **Steinmetz NA**, Aydin C, Lebedeva A, Okun M, Pachitariu M, Bauza M, Beau M, Bhagat J, Böhm C, Broux M, Chen S, Colonell J, Gardner RJ, Karsh
371 B, Kloosterman F, Kostadinov D, Mora-Lopez C, O'Callaghan J, Park J, Putzeys J, et al. Neuropixels 2.0: A miniaturized high-density probe for
372 stable, long-term brain recordings. *Science*. 2021; 372(6539):eabf4588.
- 373 **The International Brain Laboratory**, iblvideo; 2021. <https://github.com/int-brain-lab/iblvideo>.
- 374 **The International Brain Laboratory**, pykilosort; 2021. <https://github.com/int-brain-lab/pykilosort>.

- 375 **The International Brain Laboratory.** Standardized and reproducible measurement of decision-making in mice. *eLife*. 2021; 10:e63711. doi:
376 10.7554/eLife.63711.
- 377 **The International Brain Laboratory.** Reproducibility of in-vivo electrophysiological measurements in mice. *bioRxiv*. 2022; .
- 378 **The International Brain Laboratory.** Spike sorting pipeline for the International Brain Laboratory. *figshare*. 2022; doi:
379 10.6084/m9.figshare.19705522.
- 380 **The International Brain Laboratory.** Video hardware and software for the International Brain Laboratory. *figshare*. 2022; doi:
381 10.6084/m9.figshare.19694452.
- 382 **Wang Q**, Ding SL, Li Y, Royall J, Feng D, Lesnar P, Graddis N, Naeemi M, Facer B, Ho A, Dolbeare T, Blanchard B, Dee N, Wakeman W, Hirokawa KE,
383 Szafer A, Sunkin SM, Oh SW, Bernard A, Phillips JW, et al. The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell*.
384 2020; 181(4):936–953.e20. doi: 10.1016/j.cell.2020.04.007.
- 385 **West SJ**, BrainRegister; 2021. <https://github.com/stevenjwest/brainregister>.