Event Report

Benjamin Dichter and Anna Szonja Weigl
CatalystNeuro

Ryan Ly and Oliver Rübel
Lawrence Berkeley National Laboratory

Contents

1

NEURODATA WITHOUT BORDERS

2

1 Executive Summary

Overview: The DANDI Archive (by the time of the event) contained 150+ neurophysiology datasets in the Neurodata Without Borders (NWB) format spanning many species, brain areas, task types, and imaging modalities. These included high-value datasets, e.g. from The Allen Institute, the MICrONS project, and the International Brain Laboratory, as well as diverse contributions from neuroscience labs around the world.

NeuroDataReHack 2023 (see Section 3) was held at the Hotel Andalucia Center in Granada, Spain as a satellite of the International Brain Research Organization (IBRO) World Congress 2023, with the goal to make the event accessible to diverse participants who might not otherwise have the opportunity to participate in the workshop. 31 participants from 25 unique institutions and 15 different countries attended the event (see Section 3.1).

Prior to the workshop, we hosted the virtual Open Neurodata Showcase on the Gather platform (see Section 2) where attendees were able to meet the contributors behind the neurophysiology datasets and explore virtual posters. A diverse group of 73 participants from all over the world registered for the Open Neurodata Showcase (see Section 2.2).

Projects: During NeuroDataReHack 2023 attendees learned about the open neurophysiology datasets available on the DANDI Archive and how to maximally utilize the archive and the NWB standard to incorporate existing data into their scientific workflows. Participants applied their skills to reanalyze data from the DANDI archive in a diverse range of self-defined projects with assistance by the event instructors. Attendee projects spanned different neural recording modalities, species, and scientific questions. Impressively, by the end of the 4-day workshop, many of the attendees had performed in-depth data analysis and generated figures to demonstrate preliminary results. Several attendees also incorporated the analysis tools introduced by the guest speakers in their projects. The projects were summarized by the attendees in a shared project reporting document (see Section 3.3)

Neurodata Discovery Award: Following the event, participants were invited to apply for the Kavli Foundation Neurodata Discovery Award, which awards $50,000 (USD) of funding to continue data reanalysis projects that were initiated at the NeuroDataReHack event (see Section 3.4).

Conclusion: The event was very well received by the participants (see Section 3.5) and was successful in training a cohort of scientists in the reuse of publicly available, standardized neurophysiology data and in demonstrating that such reanalysis can be done with the current ecosystem of software and available data.

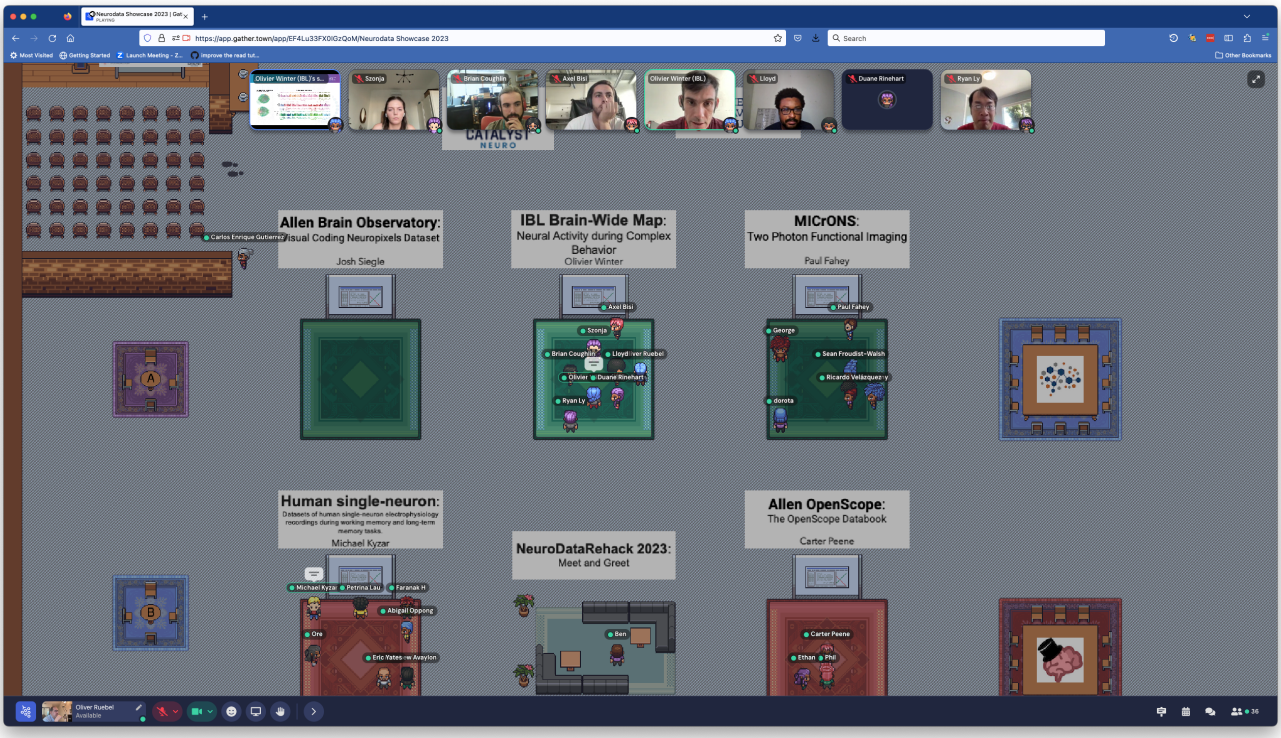
Organizing Committee:

- Benjamin Dichter, CatalystNeuro
- Oliver Rübel, Lawrence Berkeley National Laboratory
- Ryan Ly, Lawrence Berkeley National Laboratory
- Stephanie Albin, The Kavli Foundation

Event Website:

- https://neurodatawithoutborders.github.io/nwb_hackathons/HCK16_2023_Granada_RH

2 Open Neurodata Showcase



We organized Open Neurodata Showcase, a virtual event where open neurophysiology data contributors presented their datasets to participants interested in analysis of open datasets. The event was hosted on the Gather platform. This year, we had 8 registered presenters and 73 registered participants.

2.1 Program

Session	Speakers	Time: Pacific
Gather space opens to participants		8:55am
Introduction to the virtual data showcase	Benjamin Dichter (CatalystNeuro)	9:00-9:20am
Virtual poster session	Data contributors (see below)	9:20-10:30am

Poster	Speakers
Allen Brain Observatory: Visual Coding Neuropixels Dataset	Josh Siegle, Allen Institute
A Brain-Wide Map of Neural Activity during Complex Behaviour	Olivier Winter, International Brain Lab (IBL)
MICrONS Two Photon Functional Imaging	Paul Fahey, Baylor College of Medicine
The OpenScope Databook	Carter Peene, Allen Institute
Three datasets of human single-neuron electrophysiology recordings during working memory and long-term memory tasks	Michael Kyzar, Rutishauser Lab
Challenges and neuroethical considerations in sharing human neural data	Angelique C Paulk, Massachusetts General Hospital
Data on auditory cortex plasticity and oxytocin neuron responses in co-parental mice	Robert Froemke, NYU

2.2 Participants

A diverse group of 73 participants from all over the world registered for the event. During the event approximately 30 to 40 participants where typically in attendance at any given time.

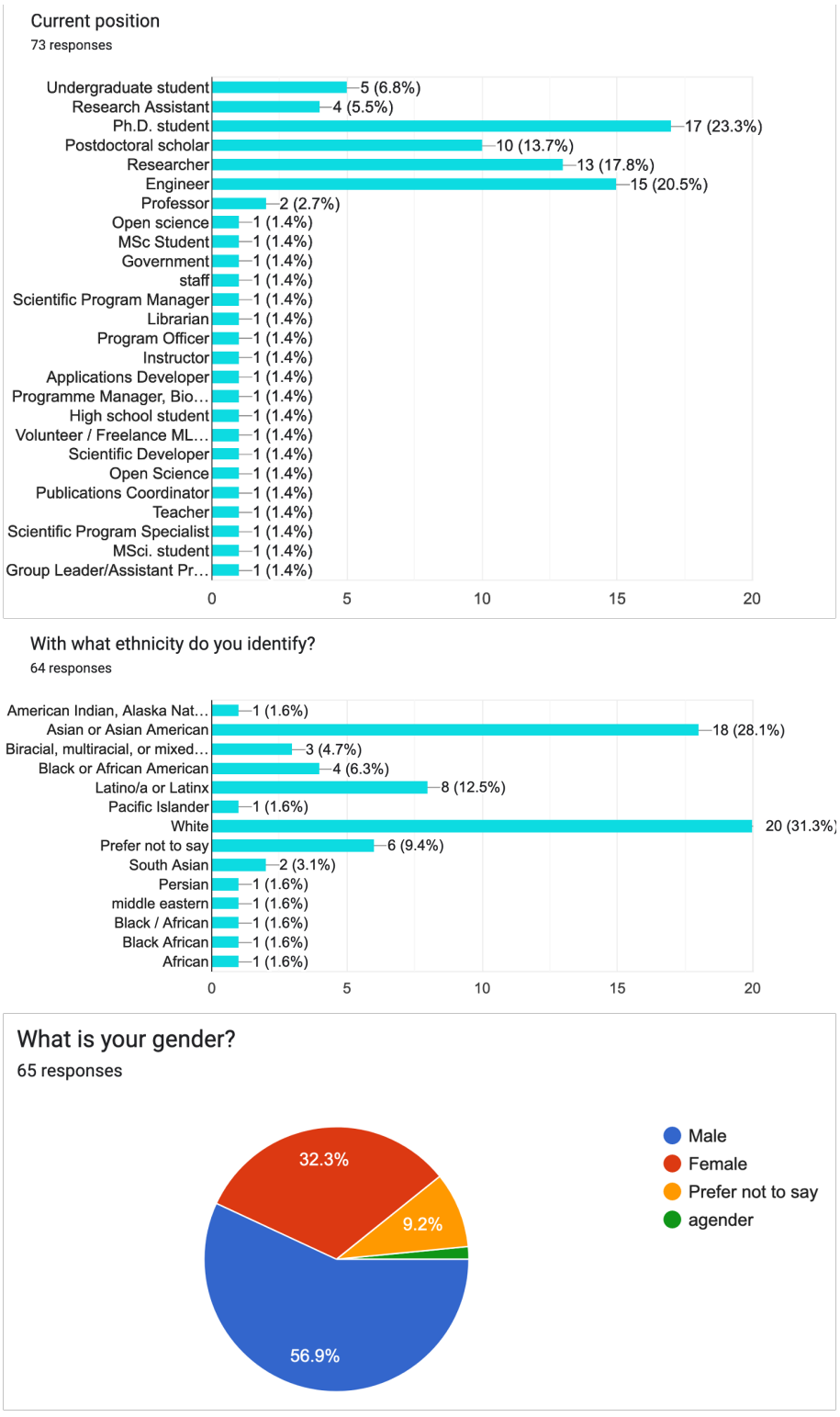


Figure 1: Current position, ethnicity, and gender reported by registered participants of the Open Neurodata Showcase 2023.

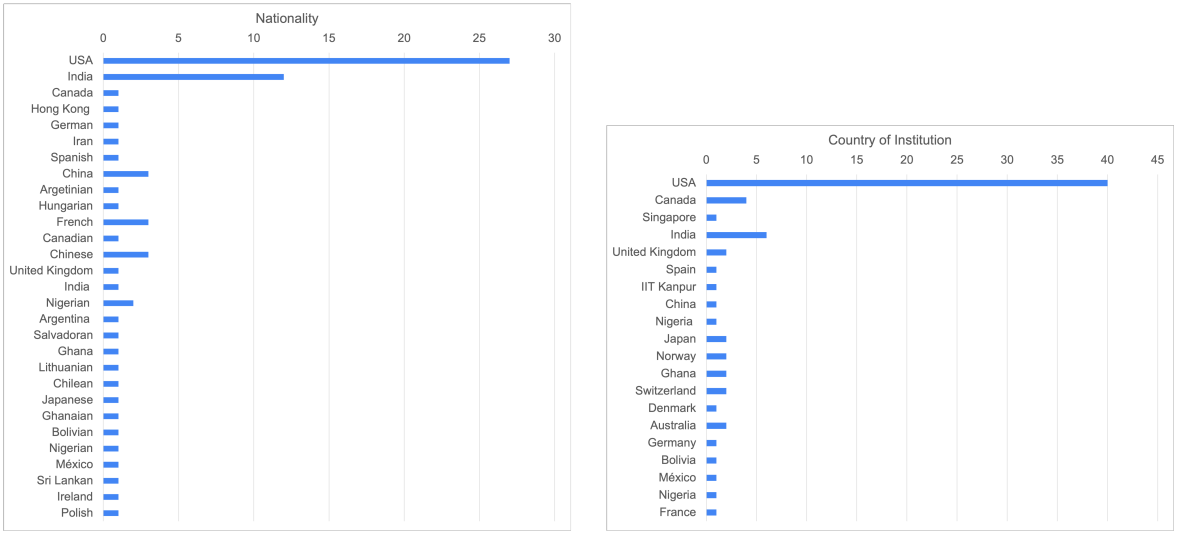


Figure 2: Nationality and the country of the institution reported by registered participants of the Open Neurodata Showcase 2023.

2.3 Exit Survey

9 attendees responded to the exit survey. Based on the collected feedback, attendees had an overall good experience, and found the virtual event useful for engaging with open data contributors and other scientists. We received comments about how to improve next year's event; such as creating a separate booth that introduces the posters such that attendees have a sense about the talks they are about to attend.

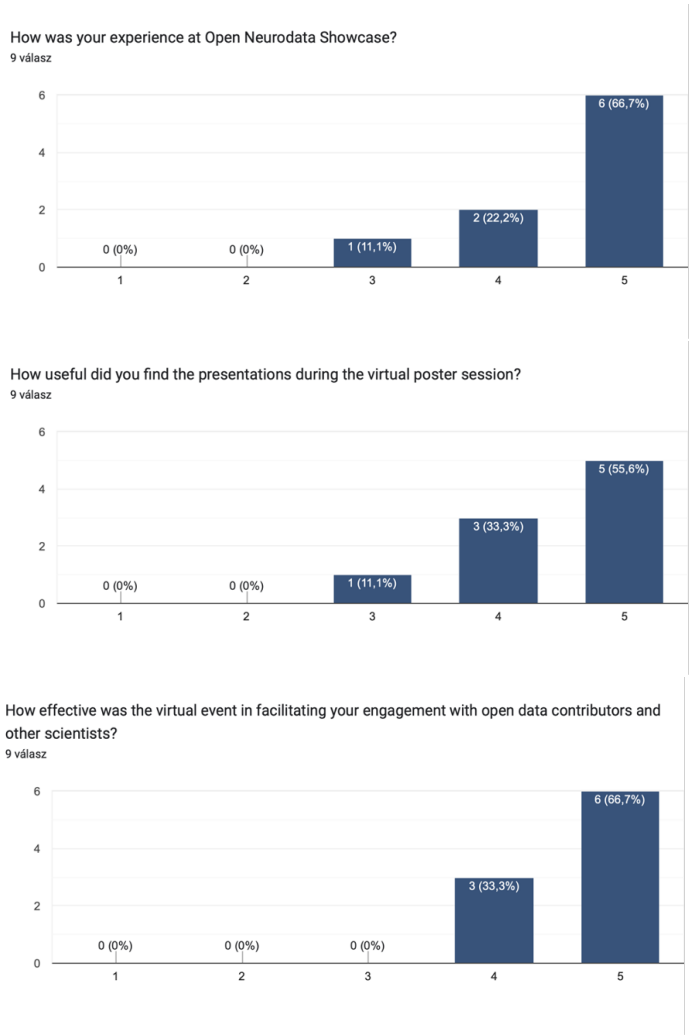


Figure 3: Participants of Neurodata Rehack 2023. **Left:** Rooftop social on day 4 (Sept. 8). **Right:** Flamenco performance at *Restaurante-Flamenco Quera La Rocio* on day 2 (Sept. 6).

3 NeuroDataReHack

3.1 Participants

Country	Institution	Count
United States	Johns Hopkins Applied Physics Lab	3
	Lawrence Berkeley National Laboratory	2
	Massachusetts Institute of Technology	2
	MBF Bioscience	1
	CatalystNeuro	1
	Princeton University	1
	Columbia University	1
United Kingdom	University College London	3
Spain	Pompeu Fabra University, Spain	1
	Universitat Pompeu Fabra	1
	IDIBPAS	1
Germany	Ernst Struengmann Institute for Neuroscience	1
	Max Planck Institute for Neurobiology of Behavior	1
	University Tübingen	1
India	Indian Institute of Technology Kanpur	1
Denmark	University of Copenhagen	1
Nigeria	University of Port Harcourt	1
Canada	Krembil Brain Institute	1
China	Casibrain	1
France	INS, INSERM, Aix-Marseille University	1
Mexico	UNAM	1
Australia	Monash University	1
Italy	”La Sapienza” University of Rome	1
Portugal	Champalimaud Center for the Unknown	1
Switzerland	EPFL	1

Table 1: NeuroDataReHack 2023 participants and instructors. There were 31 participants from 25 unique institutions.

3.2 Program

Color Legend				
Talks	Food	Hacking	Group discussions	Break

Day 0 (Monday, September 4)			
Start Time	End Time	Duration	Topic
7:00 PM	8:30 PM	1:30:00	Dinner

Day 1 (Tuesday, September 5)				
Start Time	End Time	Duration	Topic	Speaker
8:00 AM	9:00 AM	1:00:00	Breakfast	
9:00 AM	9:30 AM	0:30:00	Welcome to NeuroDataReHack 2023	Ben Dichter
9:30 AM	10:10 AM	0:40:00	Introduction to NWB	Oliver Ruebel
10:10 AM	10:40 AM	0:30:00	How to read NWB files	Ryan Ly
10:40 AM	11:10 AM	0:30:00	Introduction to DANDI	Nima Dehghani
11:10 AM	11:40 AM	0:30:00	Searching DANDI and advanced queries	Ben Dichter
11:40 AM	12:30 PM	0:50:00	Hacking on projects	
12:30 PM	1:30 PM	1:00:00	Lunch	
1:30 PM	1:55 PM	0:25:00	A tour of NWB analysis software	Ryan Ly
1:55 PM	2:35 PM	0:40:00	Invited lecture: Open Data from International Brain Lab	Olivier Winter
2:35 PM	3:15 PM	0:40:00	Invited lecture: Learnable latent embeddings for joint behavioural and neural analysis with CEBRA	Steffen Schneider
3:15 PM	3:30 PM	0:15:00	Refreshments	
3:30 PM	4:30 PM	1:00:00	Project roundtable	Everyone
4:30 PM	5:10 PM	0:40:00	Hacking on projects	
5:10 PM	5:30 PM	0:20:00	The Kavli Discovery Award	Stephanie Albin
5:30 PM	6:30 PM	1:00:00	Speed Networking	Everyone
6:30 PM	8:00 PM	1:30:00	Dinner	

Day 2 (Wednesday, September 6)				
Start Time	End Time	Duration	Topic	Speaker
8:00 AM	9:00 AM	1:00:00	Breakfast	
9:00 AM	9:15 AM	0:15:00	Introduction to Day 2	Ben Dichter
9:15 AM	9:45 AM	0:30:00	Remote invited lecture: SpikeInterface	Alessio Buccino
9:45 AM	12:30 PM	2:45:00	Hacking on projects	
12:30 PM	1:30 PM	1:00:00	Lunch	
1:30 PM	3:45 PM	2:15:00	Hacking on projects	
3:45 PM	4:00 PM	0:15:00	Refreshments break	
4:00 PM	4:30 PM	0:30:00	Remote invited lecture: Neural Data Analysis with Pynapple	Guillaume Viejo
4:30 PM	6:30 PM	2:00:00	Hacking on projects	
6:30 PM	8:00 PM	1:30:00	Dinner	
8:00 PM	8:30 PM	0:30:00	Break	
8:30 PM	10:30 PM	2:00:00	Dancing performance in caves	

Color Legend				
Talks	Food	Hacking	Group discussions	Break

Day 3 (Thursday, September 7)				
Start Time	End Time	Duration	Topic	Speaker
8:00 AM	9:00 AM	1:00:00	Breakfast	
9:00 AM	9:15 AM	0:15:00	Introduction to Day 3	Ben Dichter
9:15 AM	9:45 AM	0:30:00	Project check-ins	Everyone
9:45 AM	12:30 PM	2:45:00	Hacking on projects	
12:30 PM	1:30 PM	1:00:00	Lunch	
1:30 PM	3:30 PM	2:00:00	Hacking on projects	
3:30 PM	3:45 PM	0:15:00	Refreshments break	
3:45 PM	6:30 PM	2:45:00	Hacking on projects	
6:30 PM	8:00 PM	1:30:00	Dinner	

Day 4 (Friday, September 8)				
Start Time	End Time	Duration	Topic	Speaker
8:00 AM	9:00 AM	1:00:00	Breakfast	
9:00 AM	9:15 AM	0:15:00	Introduction to Day 4	Ben Dichter
9:15 AM	12:30 PM	3:15:00	Hacking on projects	
12:30 PM	1:30 PM	1:00:00	Lunch	
1:30 PM	4:30 PM	3:00:00	Project presentations	Everyone
4:30 PM	4:45 PM	0:15:00	Refreshments break	
4:45 PM	6:00 PM	1:15:00	Discussion and feedback	Everyone
6:00 PM	8:30 PM	2:30:00	Break	
8:30 PM	10:00 PM	1:30:00	Rooftop dinner	
10:00 PM	12:00 AM	2:00:00	Rooftop cocktail pool party	

3.3 Projects

Participants applied the skills they learned to implement their own projects and analyze existing neurophysiology data to generate new insights. Participants reported on 19 unique projects via a shared GoogleDoc, included below.

Projects: 2023 NeuroDataReHack

Neuronal ensembles in different brain regions

Key Investigators

- Ricardo Velázquez Contreras

Project Description

Neuronal ensembles, as originally proposed by Donald Hebb, represent groups of neurons that display coordinated and alternating activity patterns. These ensembles and their dynamics play a crucial role in governing various behaviors and encoding sensory stimuli within the brain. While these ensembles have predominantly been observed in the primary visual cortex of rodents, where they contribute to the encoding of visual stimuli such as drifting gratings, our understanding of this interpretation of neuronal populations and its applicability to comprehending stimulus encoding or behavioral development in other brain regions remains somewhat limited.

In pursuit of this project, my objective is to identify neuronal ensembles across different types of calcium imaging datasets. I will be examining records of neural activity in three distinct brain regions: the primary visual cortex (DANDI dataset ID:000039), the piriform cortex (DANDI dataset ID:000167), and the CA1 region of the hippocampus (DANDI dataset ID:000219). My primary method for achieving dimensional reduction and ensemble identification will be Singular Value Decomposition (SVD).

In this project, I hypothesize that it is possible to discern various patterns of neural activity and encoding within these ensembles. These patterns may arise from neuronal coactivity and/or differences in the sequences of activation. This research aims to shed light on the diverse roles played by neuronal ensembles across different brain regions and their implications for understanding neural coding and behavior.

Objectives

- Read the NWB datasets from the DANDI archive, extracting neuronal activity and the temporal presence of each stimulus or behavior.
- Identify neuronal ensembles using the Singular Value Decomposition method.
- Synchronize the activity of each identified ensemble and compare it with the presence of each stimulus or behavior.

Approach and Plan

- Download multiple datasets from the DANDI Archive, encompassing various brain regions and types of stimuli using the DANDI API.
- Extract the fluorescence activity of each individual neuron using PyNWB and conduct basic preprocessing on the data to generate a binary representation of neuronal activity by computing the first derivative of the calcium signal and establishing an activation threshold.

- Create a similarity matrix based on the most active population vectors across each recording session, using the cosine of the angle formed by each pair of population vectors.
- Apply Singular Value Decomposition to the similarity matrix in order to extract the number and timing of activations for each ensemble.
- Compare the activation timing of each ensemble with the presence of each stimulus.

Progress and Next Steps

- I have acquired the DANDI datasets containing activity data and have begun signal preprocessing.
- In the process of generating the similarity matrix for each dataset and conducting some code sanity checks.
- Working on coding a straightforward approach for performing Singular Value Decomposition on the data.
- Extracting the timing of stimuli and comprehending how it is represented in each dataset.
- Synchronizing the timing of stimuli with the activity of ensembles.
- For my next steps, I plan to:
    - Conduct sequence analysis using the activity of each ensemble and compare these sequences with the presence of each stimulus or behavior.
    - Refine the algorithm to identify ensembles with greater precision, taking into account the unique properties of each dataset.
    - Explore additional datasets and compare the encoding of information, considering both the perspective of neuronal coactivation within the ensembles and the sequencing of activations.

Materials

- Allen Institute – Contrast tuning in mouse visual cortex with calcium imaging - https://dandiarchive.org/dandiset/000039
- Two photon calcium imaging of mice piriform cortex under passive odor presentation - https://dandiarchive.org/dandiset/000167
- Two photon calcium imaging in the CA1 region of the hippocampus in neonatal mice. - https://dandiarchive.org/dandiset/000219

Background and References

- Hebb, D. O. (1949). *The organization of behavior*. Wiley.
- Carrillo-Reid L, Miller JE, Hamm JP, Jackson J, Yuste R. Endogenous sequential cortical activity evoked by visual stimuli. J Neurosci. 2015 Jun 10;35(23):8813-28. doi: 10.1523/JNEUROSCI.5214-14.2015. PMID: 26063915; PMCID: PMC4461687.

Identifying latent states in decision-making tasks to uncover choice strategies

Key Investigators

- Zeinab Mohammadi

Project Description

Identifying latent states in decision-making tasks is an interesting modeling framework. Rodent decision-making relies on distinct latent states which switch on a timescale of tens to hundreds of trials. Different brain regions' contributions to state-dependent strategies remain an important open problem. We can address this challenge using model-based analysis of datasets on the DANDI Archive. We can also analyze the correlation between the neural and behavioral data to find the neural pattern corresponding to different states. To achieve this, we will consider fitting Generalized Linear Models (GLMs), Hidden Markov Models (HMMs), and a combinational GLM-HMMs framework to analyze these datasets.

Objectives

- In this project, we will leverage Multi-state GLM-HMM to uncover the intricate interplay among various choice strategies across trials. These findings will provide essential insights for identifying and analyzing states within behavioral data, further advancing our understanding of sensory decision-making processes in computational neuroscience and machine learning. If possible, we will also attempt to identify the correlation between the behavioral and neural data.

Approach and Plan

1. Preprocessing the data to select the appropriate sessions or animals.
2. Creating the design matrix for the model, incorporating suitable regressors.
3. Fitting the GLM, 1-state GLM-HMM, to the data.
4. Using step 3 as an initialization for the GLM-HMM.
5. Fitting the model to the DANDI data using varying numbers of states.
6. Utilizing cross-validation to compare different models and select the best one.
7. Assessing the roles of various covariates in the model and generating coresponding psychometric curves.

Progress and Next Steps

1- Writing a code to
   - Read the different sections of the data
   - Select the good trials
   - Convert regressors/output to an appropriate format for the algorithm (binary format): Stimulus rate, past stimulus, animal choice, etc

2- Working toward a GLM-HMM for this data using different covariates and animal choice

Materials

- The DANDI dataset: https://dandiarchive.org/dandiset/000016/draft/files?location=
- This is the related paper: https://pubmed.ncbi.nlm.nih.gov/31753580/

Background and References

- Z. Ashwood, et al. "Mice alternate between discrete strategies during perceptual decision-making." *Nature Neuroscience*, 2022
- S.S. Bolkan & I.R. Stone et al. Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. *Nature Neuroscience*. 2022.

Neural trajectories in the anterior cingulate cortex during reward

Key Investigators

- Olivier Winter
- Petrina Lau

Project Description

Compute neural trajectories from raw AP data in the region ACA (anterior cingulate cortex) of the mouse around the reward event in the IBL task.
We expect unsupervised clustering algorithm such as Umap or Cebra to distinguish between rewarded and unrewarded trials, and between left and right trials.

Objectives

- Create a pipeline to select a brain region and a behavior, read in the raw data and relate to the task events in the dandiset 000409.

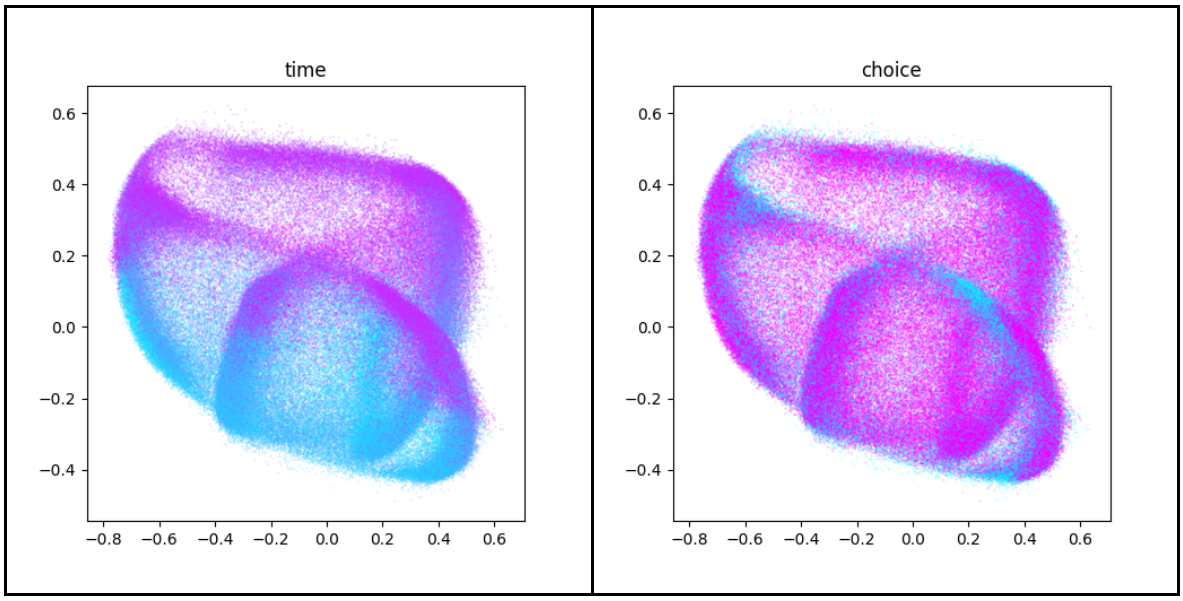
Approach and Plan

- Query the BWM dataset for recordings in ACA
- Extract raw data snippets around the feedback event of the IBL task
- Pre-process the raw data to extract a PSTH-like signal
- Input the data into an unsupervised algorithm CEBRA to extract structure
- Label the manifold with corresponding labels, we expect to see behavior related structure

Progress and Next Steps

- Here is a couple of Manifold with colormap corresponding to labels.

Next step would be to compare with a standard PSTH to see if there is some information gain in using the raw AP band signals or not.



Bridging the gap between function and structure

Key Investigators

- Hannah Gooden
- Victoria Rose
- Daniel Xenes

Project Description

We propose continued secondary analysis of the MICrONS dataset, including the dataset of two-photon calcium imaging hosted on the DANDI archive and the co-registered electron microscopy data hosted in the BossDB archive. In particular, the comparison and generation of functional connectivity networks and structural connectivity networks have been under-explored in this dataset and require joint analysis across archives. We propose to generate functional connectivity estimates, using a rolling time window, using the two-photon calcium imaging data in the MICrONS dataset to create a sequence of graphs over time and across stimuli conditions. We will utilize an existing table of co-registered neuron IDs, which connect the structural graph derived from the electron microscopy data, to align the functional networks to the structural networks. We will create visualizations of the graphs and characterize common graph metrics in the structural and functional graphs.

If successful, this work will generate novel insight into the stimulus-dependent and time-dependent activation of functional subnetworks. This will also allow direct investigation of causal connectivity estimation using functional data. Follow-up work could investigate improvements to functional connectivity estimation methods.

Even more exciting, however, is the potential to generate a wide range of hypotheses relating to the structure and function of neural networks within the mammalian cortex.

Objectives

- Provide a single notebook that connects the functional time-series data from DANDI, the structural EM and reconstruction data from BossDB, and the annotation data from CAVE to lower the barrier of entry for multi-modal analysis on MICrONS.
- Create a visualization widget that leverages *Neuroglancer* to show functional and structural data in a shared coordinate frame.
- Investigate tuning curves of functional fluorescence traces stratified on morphological cell type using *Pynapple*.

Progress and Next Steps

- 9/4/23 - Determined DANDI dataset and identified external data sources from BossDB and CAVE
- 9/5/23 - Developed initial data loading functions and prototype neuroglancer script
- 9/6/23 - Ran analysis with session 7 scan 4, field 4 as an example
- 9/7/23 - Cleaned up code and integrated NWB file loader to PyNapple analysis notebook
- 9/8/23 - Created presentation and developed figures

Background and References

- https://doi.org/10.1101/2021.07.28.454025
- https://github.com/dxenes1/neurodatarehack2023

Predicting cell type identity based on functional connectivity

Key Investigators

- Margaret Conde Paredes

Project Description

The goal is to study task-specific neural trajectories and covariance in primate motor cortex electrophysiology data to gain insights into the neural mechanisms underlying motor control and movement generation. To understand how the motor cortex represents and processes motor information, we ask how different neurons or neural populations contribute to motor tasks, and how motor commands are translated into specific movements. These analyses contribute to our understanding of motor function and can have implications for areas such as neuroprosthetics or rehabilitation strategies.

Objectives

- Determine the connectivity between cells by analyzing the correlation across time series.

Approach and Plan

- Get calcium traces adjusted for baseline fluorescence (df/f0).
- Label each cell based on IHC markers
- Apply rank-1 NMF to mitigate background activity due to motion-induced blood flow.
- Split cells into testing and training groups
- For each training cell, compute its correlation with other training cells. Extract features: For each cell, calculate its highest correlation with each cell type (avoid self-correlation).
- Train a **Random Forest classifier** on these features to predict cell types.
- Test the classifier on the testing set.
- For validation: Shuffle cell labels, retrain, and evaluate to understand the classifier's robustness.

Background and References

- Schneider, Aidan et al. “Transcriptomic cell type structures in vivo neuronal activity across multiple timescales.” *Cell reports*, vol. 42,4 112318. 29 Mar. 2023, doi:10.1016/j.celrep.2023.112318
- Geiller, Tristan et al. “Large-Scale 3D Two-Photon Imaging of Molecularly Identified CA1 Interneuron Dynamics in Behaving Mice.” *Neuron* vol. 108,5 (2020): 968-983.e9. doi:10.1016/j.neuron.2020.09.013

Analyzing spatiotemporal patterns from ECoG recordings

Key Investigators

- Tim Näher
- Lisa Bastian

Project Description

My proposed project aims to analyze spatiotemporal patterns from ECoG recordings to identify patterns of neural activity in the sensorimotor cortex that correspond to different speech sounds and articulatory movements.

Building on the research presented in the original study, I will broaden the scope of the analysis to consider spatiotemporal patterns like travelling waves, which have been shown to be influenced by a variety of variables including visual inputs, working memory load, and motor commands. Due to the topological organization of the somatosensory cortex, travelling waves may spread throughout the cortical sheet to produce dynamics in both space and time that could provide be crucial for the creation of sequences of motor events, which is the building block for speech production.

I will use these spatiotemporal patterns to train a deep learning model to differentiate between the different produced sounds based on the neuronal responses. The classification ability for different frequency band dynamics could ultimately be compared to infer which frequency pattern carries the most information about the produced sound. By integrating spatiotemporal patterns into the analysis, one could gain a deeper

understanding of the time-space dependency of functional organization in the sensorimotor cortex during speech production.

Objectives

- We would like to find neural spatiotemporal correlates of speech production using the LFP in sensori motor areas. To achieve this, we will transform the ECoG LFP data into vector fields and analyze the dynamics of the resulting time series.

Approach and Plan

- After appropriate preprocessing (z-score, interpolating outlier electrodes, trial parsing), we identify spectral peaks in the LFP. We achieve this by fitting the 1/f background spectrum using the FOOOF tool box. We then model spectral peaks as gaussian deviations from this 1/f fit. Subsequently, we filter the LFP in the identified frequency bands. This allows us to use only frequency bands which are likely to have meaningful temporal structure. After that, we compute vector fields with an optical flow algorithm. Finally. We extract features from these fields such as direction, magnitude, consistency, divergence and curl and try to classify the spoken consonant vowel pairs.

Integrating Rule-Based Modulation & Context-Dependent Sensorimotor Adaptation in Cortical Areas

Key Investigators

- Tenghai Long

Project Description

Integrating Rule-Based Modulation & Context-Dependent Sensorimotor Adaptation in Cortical Areas

Objectives

- Explore the integration of rule-based modulation and context-dependent sensorimotor adaptation across cortical areas by analyzing electrophysiological data from DANDI:000239 and DANDI:000232.

Approach and Plan

- Data acquisition: Download and preprocess data from both dandisets in NWB format.
- Neural representation analysis: Investigate encoding of rule-based and context-dependent variables in cortical areas and compare neural representations.
- Integration analysis: Analyze neural activity patterns associated with integration across cortical areas and identify involved neural circuits.
- Sequence learning and rule-guided behavior analysis: Explore interactions between sequence learning and rule-guided behavior processes.

- Computational modeling: Develop models to describe neural mechanisms underlying integration, and test predictions using electrophysiological data.
- Expected Outcomes: Generate new insights into neural mechanisms underlying the integration of rule-based modulation and context-dependent sensorimotor adaptation. Findings may contribute to developing effective strategies for motor rehabilitation in patients with sensorimotor deficits and provide valuable insights into the neural basis of sensorimotor flexibility and rule-guided behavior.

IBL Brain-wide map: Adaptation and history biases across the brain hierarchy

Key Investigators

- Melanie Tschiersch

Project Description

I propose to analyze a dataset from the DANDI Archive to gain a deeper understanding of the nature of adaptation (repulsion from the previous stimulus) versus serial dependence (attraction to the previous stimulus) in visual processing versus higher-order cognitive areas. The dataset IBL dataset 000049 “Brain-wide map” offers the opportunity to do so through viewing a visual decision-making task in mice while performing recordings from many different areas across the entire brain. To analyze the data, I plan to first view traces of adaptation in behavior, for which I will propose a new method of measuring response accuracy in single trials in this task. I will then view if responses are serially dependent through analyzing the strength of the single trial error on the difference between the previous and current trial’s luminosity. I will then analyze if neurons are indeed repelled from the previous item through viewing neural manifolds in CEBRA for different brain regions. Since recent evidence supports the possibility that serial dependence is generated in higher-order areas (e.g. [1]), I expect to see less adaptation in the higher-order cognitive areas (RSP), than in visual areas (VIS). Analyzing this dataset can therefore help in clarifying the opposing roles and possible differences in the neural mechanisms underlying adaptation and serial dependence, which are of recent interest in the field.

Background and References

- [1] Sheehan & Serences, 2022, PLOS Biology

Normative Models of Cognitive Representations

Key Investigators

- Will Dorrell

Project Description

‘Why is that neuron firing like that?’ is perhaps the most fundamental neuroscience question. A large body of work in theoretical neuroscience attempts to develop frameworks with which to understand this problem. One of the most attractive of these is the normative framework, that answers this question with a set of goals: i.e. these neurons are firing like that because they want to achieve goals A, B, and C. A classic example is sparse coding in the visual cortex (Olshausen & Field 1996): this argues that V1 neurons want to represent visual scenes (goal one) using a set of positive sparse firing rates (constraint one - firing rates must be positive, constraint two - and also sparse, i.e. only a few neurons fire at once). Optimising this objective leads to Gabor-like tuning curves as seen in the visual cortex - ‘explaining’ the observed findings, extracting generalisable principles, and making experimentally testable predictions; a very attractive triad of achievements.

My PhD work has been developing a novel normative representational objective, which we call actionability. This augments existing approaches that I broadly classify as either functional or biological: functional are those that ask the representation to *be a representation* i.e. to actually encode information about the object it is representing. Biological constraints are things like sparse coding that argue from things like energy efficiency to place additional constraints on the code. Our work has tried to formalise a third idea: representations must not simply encode information about the world, they must be the locus of all the computations we use to understand the world. For example, in many situations we don’t just want to represent a variable, such as our position in an unknown city. We also want to have the power to consider hypotheticals: what will happen if I take this sequence of streets? Will I reach the bakery? When we measure neural representations, we are not just reading the brain sending information to itself, we are measuring the fumes of its ongoing computations.

In this project I attempted to use the tools we have developed to make this idea tractable to understand a dataset available on Dandi. I chose dandiset 000239. In this experiment mice were asked (among other things) to lick a port that appeared in one of five locations. The port appeared in a predictable sequence, zigzagging up and down the five locations. The mouse was rewarded every 12 correct licks, and eventually became very good at this behaviour. The experimenters then recorded firing rates in cortical area ALM that appeared to encode the abstract latent sequence variable.

I analysed the representation looking for some surprising representational properties that we could try to understand using our normative frameworks. This led to some non-trivial representational properties. We started with a ‘null model’ that let us work out what was surprising. I asked a representation to encode the upcoming position the mouse had to lick and whether it would get rewarded - this led to neural representations that were categorically not found in the representation, suggesting it is a poor model of the area. I then developed this further by asking it to embed the sequence nature of the task. We can think of this as saying that the animal is able to predict what location it will have to lick next from its current representation at all times. When this constraint was added, making the representation an internal model of the task, the optimal representation changed dramatically and displayed features that were qualitatively similar to those measured.

Further work could try to solidify the representational analysis (what proportion of neurons had each of these tuning properties? Are we extracting justifiable time periods?). In addition, it would be good to come up with some more inventive way of comparing theoretical and experimental results, rather than just looking. All in all it was a promising start for these approaches though, and I would be excited to try on more dandisets!

Investigating the relationship between neuronal activity patterns and behavior in humans during decision-making tasks

Key Investigators

- Kshitij Kumar

Project Description

Analyzing DANDI:000397 to investigate the relationship between neuronal activity patterns and behavior in humans during decision-making tasks. This would include recordings from high-density silicon arrays during neurosurgical procedures, as well as corresponding behavioral data. The project would start by preprocessing the data, such as filtering, spike sorting, and aligning the neuronal activity with the behavioral data. Then, the project would identify neuronal activity patterns associated with specific aspects of the decision-making task, such as perceptual processing, motor planning, or reward anticipation. This could involve applying unsupervised learning techniques such as clustering or dimensionality reduction to identify meaningful patterns in the neuronal activity data. Once the neuronal activity patterns are identified, the project would investigate their relationship with behavioral measures such as response time, accuracy, or subjective confidence. This could involve applying regression or classification models to predict behavior based on the neuronal activity patterns. The project would aim to provide new insights into the neural basis of human decision-making and could have implications for the development of more effective interventions for neurological and psychiatric disorders. It could also contribute to the development of new analysis techniques and tools for high-density neuronal recordings in humans.

Activity of single neurons during a declarative memory task

Key Investigators

- Fatmaalzahraa Aboalasaad

Project Description

In this data set with iD 000004 we can see how does the activity of single neurons during a declarative memory task correlate with the encoding, retrieval, and consolidation processes of memory formation. In the beginning: we extract relevant features from the neuronal activity data. This could include spike timing, firing rates, spike amplitude, or other relevant measures that capture the neuronal responses during the memory task. Then follow the following steps 1: Encoding Analysis Investigate the neuronal activity patterns during the encoding phase of the memory task by looking for specific neural signatures that indicate the encoding of new memories, such as increased firing rates or coordinated activity among specific neuron populations. 2: examining the neuronal activity during the retrieval phase of the memory task. 3: Analyze the patterns of neural activity associated with successful retrieval and compare them to unsuccessful retrieval attempts. Look for consistent firing patterns or reactivation of specific neuron populations related to memory retrieval. Correlation Analysis: doing correlations between neuronal activity patterns and memory performance measures.

Determine whether specific neuronal responses during the memory task are predictive of successful memory encoding, retrieval, or consolidation.

Modulation of spontaneous neural activity by arousal

Key Investigators

- Martynas Dervinis

Project Description

Neural activity can be broadly divided into fast and slow types. Slow activity consists of neuronal membrane potential fluctuations occurring on the timescale of seconds, including slow and infra-slow subtypes centred around 0.3 and 0.03 Hz, respectively. During wakefulness, it is typically thought to underlie attentional processes and brain state transitions. Spontaneous neural activity dynamics on this timescale in the mouse brain (hippocampus, cortex, thalamus) were the focus of data I collected using Neuropixels probes in my recent research project. In addition, I recorded pupil area size as a measure of the animal's level of arousal. Preliminary findings indicate that spontaneous neural activity on this timescale is governed by the arousal modulation gradient. This gradient is defined by the ratio of positively and negatively arousal-modulated neurons in cell populations across the brain.

With the increasing number of publicly available neurophysiology datasets, it is possible to test whether my preliminary findings hold in other datasets. Relevant datasets would need to contain both extracellular electrophysiology recordings of spontaneous neural activity in the mammalian brain combined with video recordings of pupil size. I identified two such repositories: Allen Institute's Visual Coding – Neuropixels "Functional Connectivity" dataset and IBL's "Eight-probe Neuropixels recordings during spontaneous behaviors" dataset. I intend to focus on the former dataset.

Disparities in parental response to visual and auditory cues from pups

Key Investigators

- Eberechi Wogu

Project Description

Early maternal care is essential for the development and survival of offspring. In mammals, multiple sensory modalities such as visual, olfactory and auditory cues from the pups have been shown to trigger parental responses. A recent study by Carcea et al.,2021 on mice has shown that oxytocin expressed in the Paraventricular Nucleus (PVN) plays a vital role in initiating parenting behavior towards pups by the biological mother-dams. More interesting is that even non-biological virgin female mice develop alloparenting behaviours towards pups after periods of cohabitation with experienced mother dams and litters. This social learning

depended on multiple sensory inputs such as pup distress calls (auditory cues) and pup moving away from nest (visuals cues) in brain for learned maternal care. But it is unclear which pop cues (visual or auditory) elicit a more dominant parental response from the mother dams and the virgin female mice.
This study aims to investigate the disparity between parental response to the visual pop cues and the auditory pop cues by the mother dams and the virgin female mice after some period of cohabitation using open source neurophysiological analytic tools. Data source: Dandi archive (dandiest ID DANDI:000114) containing multi electrode extracellular electrophysiological recording and behavior datasets obtained from 11 house mice.

Identifying homologies and diversifications of intrinsic physiological properties in rodents

Key Investigators

- Felipe Yáñez

Project Description

I'm currently working on the generation of a large representative dataset of single-cell properties of GABAergic interneurons (INs) in the rat barrel cortex (S1). In my prospective project, I'd like to identify homologies and diversifications of intrinsic physiological properties in rodents. Specifically, I'd systematically assess the degree and character of the variability of intrinsic physiological properties across the entire cortical depth of rat S1 (currently converting 300+ recordings to NWB), as well as mouse primary visual (V1) and primary motor (M1) cortices. I'd first carefully curate a sample of 494 V1 and 370 M1 INs and quantitatively test their representativeness against the absolute number of neurons per molecular sub-class (i.e., PV, SST, VIP, etc.) in each cortical area. A representative sample is paramount in order to avoid overrepresentation and/or underrepresentation of e-types. For each neuron, I would compute a comprehensive list of descriptive features based on its spiking patterns in response to somatic current injections. Then, I would assign neurons into e-types using standardized clustering methods. Finally, I'd evaluate e-type predictability per area and across areas in order to identify area-specific versus canonical e-types. This project would provide quantitative insight into the relationships between the main attributes that are currently used to define e-types as a function of cortical area and species.

The effect of different analysis techniques on the detection of fNIRS responses

Key Investigators

- Ishara Paranawithana

Project Description

Functional Near-Infrared Spectroscopy (fNIRS) is an emerging non-invasive optical neuroimaging technique that can be used in a wide range of neuroscience applications. However, as there is no standard data preprocessing pipeline and/or analysis procedure, which is accepted unanimously by the fNIRS community, the differences in analysis procedures could complicate the comparison and interpretation of data and eventually lead us to draw different conclusions across studies. I plan to use the dataset of “human fNIRS recordings of motor cortex during finger-tapping task” (DANDI ID: 000122) to investigate the impact of different analysis procedures on the morphology, detection, and lateralization of fNIRS responses in the motor cortex as a result of a finger-tapping task. Specifically, I intend to determine whether block-averaging or generalized linear model (GLM)-based analysis generates the same or different conclusions in block-design experimental studies. Further, I would like to investigate the use of short channels (short channel data are available in the dandiest) in reducing non-neuronal systemic hemodynamic responses and assess whether this approach improves the detection of responses at individual-subject level. I believe generating new insights through a secondary analysis of this dataset is particularly important for the fNIRS community due to the diverse range of analysis procedures and increasing adoption across a wide range of clinical applications.

Transcriptomic profiles and action potentials in GABAergic interneurons in the mouse visual cortex

Key Investigators

- Raymond L Wang

Project Description

I'd like to propose a project that focuses on understanding the relationship between transcriptomic profiles and action potentials in GABAergic interneurons in the mouse visual cortex. Utilizing the Patch-seq dataset collected by the Allen Institute, I aim to investigate the complex relationship between transcriptomic profiles, action potentials, and neuronal cell types.

My approach involves developing a generative AI model, such as a Variational Autoencoder, Generative Adversarial Network, or Diffusion Model, to simulate neurons based on their transcriptomic profiles by generating action potentials. This will allow us to gain deeper insights into the influence of genetic attributes on electrophysiological properties.

Additionally, I plan to build a machine learning-based classifier that can accurately identify transcriptomic types using electrophysiological data. This tool could serve as a valuable resource for cell type classification in neuroscience research and potentially enable in-vivo classification based on cell patching, eliminating the need to destroy the cell and facilitating further study.

By combining these two components, I hope to contribute to the ongoing efforts to establish a systematic framework for understanding and classifying neuronal cell types in the mouse visual cortex, shedding light on the complex interplay between genetic, morphological, and electrophysiological properties of neurons.

Investigating neural representations of visual stimuli during encoding and retrieval

Key Investigators

- Sofia Raglio

Project Description

As a project, I would like to investigate how different the neural representation of visual stimuli is during encoding and retrieval. I'd like to organize the work in three steps: first, decoding each trial stimulus from single units' activations, then comparing each stimulus representation in medial temporal lobe (MTL) the first time that it is presented (encoding) to the second time (recognition), and finally look for inter-trials and inter-subjects differences in latent-space representations of the stimuli and correlate them to patients' behavior. The idea is to investigate how different is the role and the activation of MTL in sensory-driven vs memory-driven representations. Comparing the neural activation corresponding to the same stimulus in the two different contexts could give some insights of how recognition memory of an object changes its internal representation. Moreover, considering a targeted dimensionality reduction of the neural activity, it could be possible to cluster the stimuli depending on their category or their novelty in a latent space. Computing the distances between the clusters of novel vs recognized items in different subjects or comparing the relative positions of two stimuli in the same category, checking if the stimulus position shifts when considering encoding vs recognition, could give some information about inter-trials and inter-subjects different behavioral performances, correlating with errors and confidence levels declared by the patients.

Contribution of physiological rhythms (respiration) to neural population geometries

Key Investigators

- Araceli Ramirez Cardenas

Project Title - Akanksha Gupta

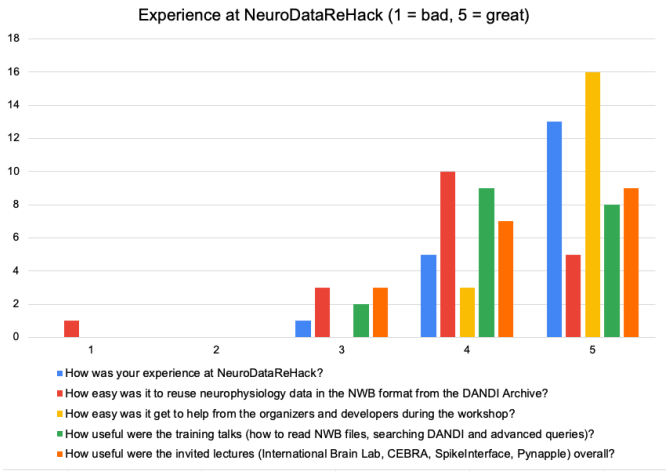
3.4 The Kavli NeuroData Discover Award

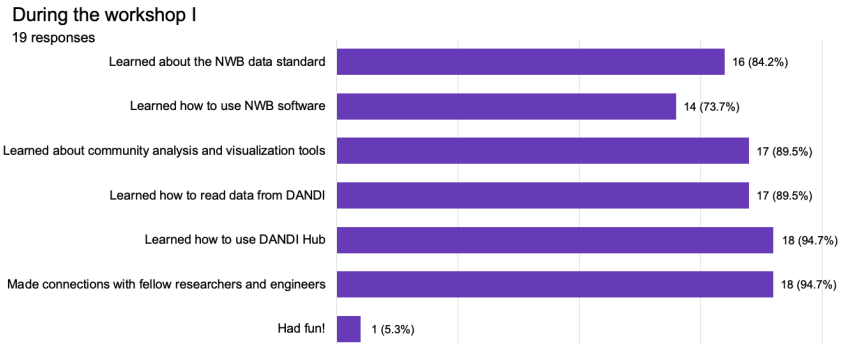
The Kavli Foundation issued a new award to generate new discoveries from reanalysis of data sets in the NWB format that is open only to attendees of the NeuroDataReHack workshop. The goals of this award are to:

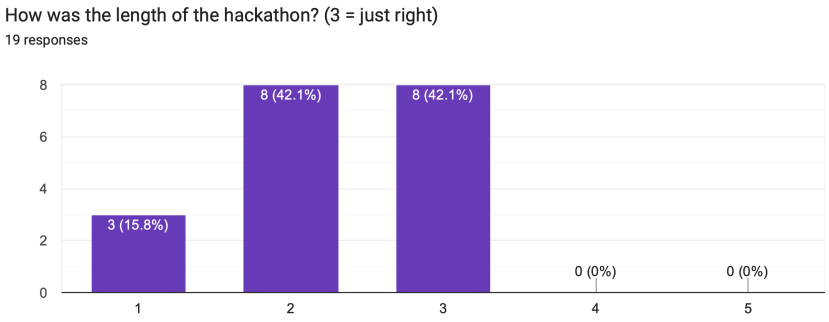
1. Support innovative analyses of existing NWB-formatted datasets through conventional or novel analytic methods,
2. Promote studies and new approaches to analysis that will drive discoveries and accelerate the pace of fundamental research in neuroscience, and
3. Demonstrate that the secondary analysis of data can be used to examine questions beyond the scope of the original data.

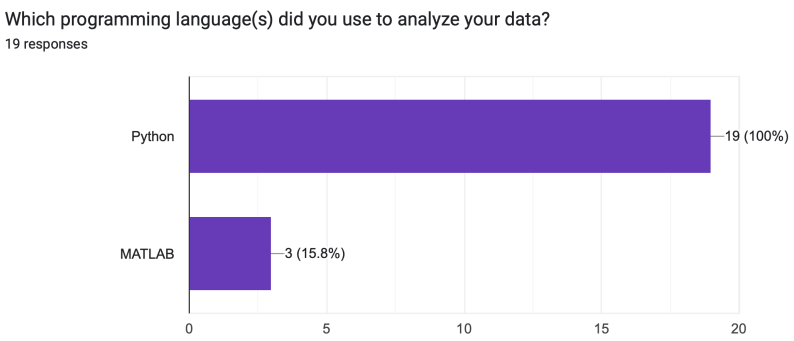
The award offers 50,000 USD for one year with a deadline of October 17th, shortly after the end of the NeuroDataReHack event. This funding opportunity received 14 applications, and they are currently under review.

3.5 Exit Survey



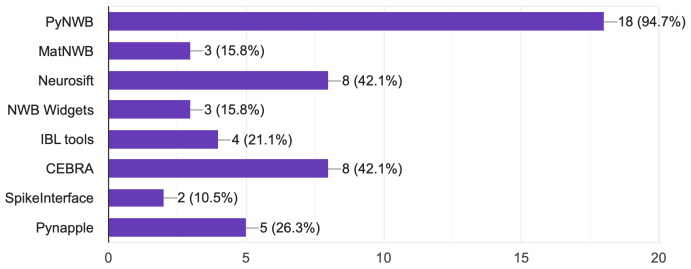






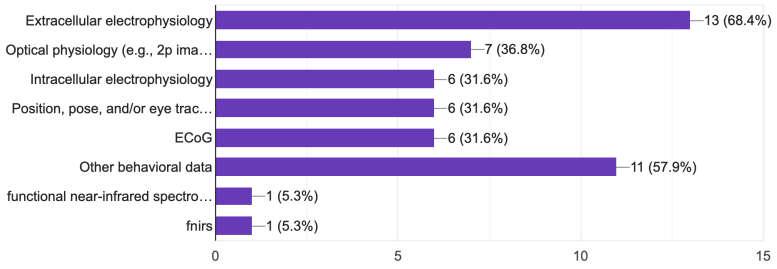
Which software tools are you using to interact with your data?

19 responses



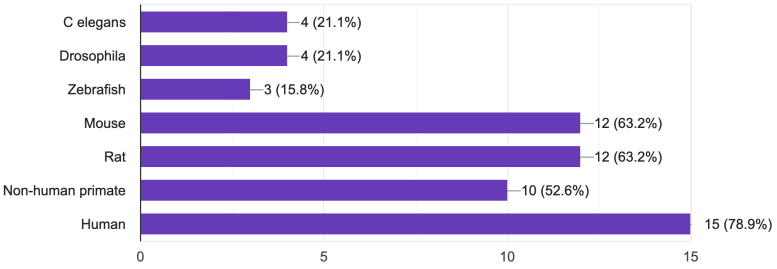
What types of data are you interested in reusing from DANDI?

19 responses



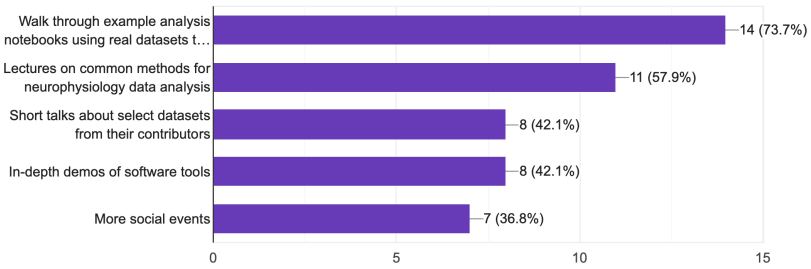
From what animal models are you interested in reusing data on DANDI?

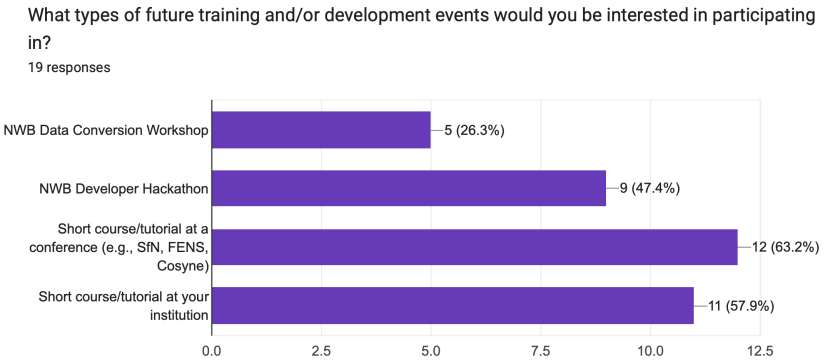
19 responses

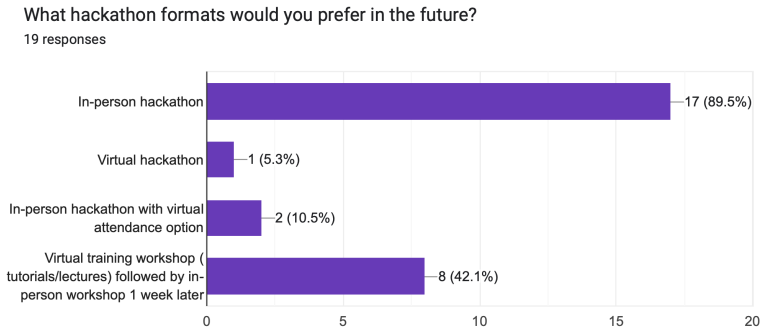


What types of sessions would you like to see in future NeuroDataReHack events?

19 responses







What were the most helpful parts of the workshop? (14 responses): **1)** Talking with organizers and other people about my project **2)** Hacking **3)** Learning about how to interact with DANDI, different analyses tools and aspects from a wider electrophysiology community was really good! **4)** Talks on NWB and loading data, invited talks **5)** Open tools for neurophysiological data visualization and analysis **6)** The experts being present at the workshop **7)** The experts and organisers being present to answer questions. **8)** Meeting tool developers, working with both experimenters and analysts. Dedicated hacking time and on-call support. New (isolated) environment to completely immerse oneself. **9)** The networking sessions at the beginning were excellent for icebreaking. The initial workshops teaching us about nwb and pynwb were very useful and I appreciated that the slides were posted for us to refer to **10)** The very high level, learning about the existence of DANDI etc. was great **11)** Reading data snippets, cebra, the possibility yo easily ask questions **12)** The most useful things in the workshop were the organizers' great willingness to provide support with doubts and problems, and the constant support and feedback from the other participants **13)** The hands on tutorials and interaction with organizers/other students **14)** getting started with NWB DANDI and CEBRA

What could have been improved at the hackathon? (11 responses): **1)** The event was really well organized, many thanks! **2)** I would have done the project pitch earlier, maybe in between a few lectures the first morning. It would give people more time to think about joining other project initiatives. **3)** It would be good to have a pre-section for those who lacks programming skills, most participates work on things they are used to or use methods they have used to. **4)** Sharing some relevant and useful materials before the workshop so that participants can get prepared a bit better/earlier **5)** Extended duration (days) 14 days would be awesome. **6)** Maybe a day longer **7)** I would've liked to have one more day to work on the dataset, a more stable internet connection and more breaks during the final presentations on the last day. **8)** I think the venue was just ok, was getting a little repetitive and definitely would've preferred have a nice dinner out once a week than 4 days of ok food. **9)** Wifi, more sunlight **10)** I believe it would be good to have at least a couple more days to develop the projects. It would also be beneficial to have a session for sharing ideas and possibly forming teams, preferably after the introductory talks. **11)** Longer project time, a little more preparation before

What kinds of training/hacking events on open neurophysiology data/software would you like to see us organize in the future? Where and how should we hold these events? (9 responses): **1)** There is a lot of requisites for people to start using datasets on DANDI, which is a bit disappointing. **2)** More regional/local smaller events outside of the US/Europe would have been great! **3)** It would be nice if you could spread out more with regards to hosting destination of the academy to other continents sometimes, if that's possible. **4)** I think in person is important, because the social exchange is crucial. I also really like the idea of local small hackathons. These could even follow a common topic like NHP specific data or rat behavior, etc **5)** Data analysis training for master students and this type of hackathons. **6)** I would like to see hackathons for best practice for data generators to create and modify NWB files. **7)** This one was awesome. The diversity of the attendees and the accessibility to you guys made the experience very enjoyable and productive. **8)** I think it would be beneficial to have workshops where the main figures from the paper associated with the dataset are reproduced. This would be appealing because it demonstrates a highly practical use of the dataset and also serves as a peer review exercise, which is necessary in scientific research. These workshops can be conducted both virtually and in person. In the case of in-person workshops, there is the opportunity to go deeper into the analysis or conduct individual exercises for peer review of another dataset. **9)** I think satellite events at conferences is working well, also invited hackathons in different institutes will be great

In your opinion, what would make it easier for others to reuse neurophysiology data? What would incentivize others to reuse neurophysiology data? (12 responses): **1)** Highlighting successful cases where data was reused **2)** A centralised modular data transformation would be a great help for starting working on a specific datasets **3)** Having a low barrier to access/load new data and more sample codes/tutorials to get a headstart **4)** Make notebooks for datasets available, if possible **5)** To make getting started as easy as possible. The code examples are great and you can easily copy and past code snippets together. I still think that having a "getting started" default notebook would be fantastic for many (unexperienced and potentially new to python) users **6)** Get a concise introduction on how to load and interact with the data present on dandi in the NWB format in form of a video tutorial **7)** Full metadata and confidence that all the data is "there". Good search engine and low-cost tools to determine if the secondary analysis is worth it. **8)** More awareness events like these. Make sure people know about these datasets **9)** Have a big how to on the dandi main website with link to dandihub example code **10)** Something that could facilitate the use of datasets would be to create step-by-step notebooks where some of the figures from a research paper are reproduced, demonstrating in a straightforward manner how you can quickly go from data to publishable figures. At the same time, it highlights that having access to data from scientific research promotes open science, peer review, and the decentralization of knowledge. **11)** An educational part that can encourage undergrad and master students to use data in thesis projects, advertising in computational labs that need to apply some theoretical modeling **12)** 1.One way to make neurophysiology data more accessible and user-friendly is by providing example Jupyter Notebooks that walk researchers through the steps of how to analyze the data. This not only serves as a tutorial but also provides a framework that can be customized for other related research. 2.Neurophysiology data is rich and often multidimensional. One dataset can potentially answer multiple questions, not just the question it was initially collected to address. If researchers provide an extensive metadata description and pose additional research questions that could be explored using the same dataset, it would incentivize others to dig into the data for their purposes. 3. Adding the element of competition, coupled with tangible rewards and industry sponsorship, you not only incentivize the reuse of neurophysiology data but also expand its impact beyond academia. Hosting a competition focused on decoding or analyzing neurophysiology data can substantially boost interest and engagement in data reuse. A tangible reward, such as a cash prize or resources for research, can serve as a strong incentive for participation. It could attract not only neuroscientists but also data scientists, machine learning engineers, and researchers from other disciplines, enriching the methods and approaches used in the analysis. Having the event sponsored by an industry player in brain-machine interfaces can offer several advantages. Firstly, it lends credibility and attracts more visibility to the competition. Secondly, it may provide additional resources, such as specialized hardware or software, that could be used in the competition or offered as part of the prize. Lastly, it bridges the gap between academic research and industry applications, showing how the dataset can be used in real-world applications.

Any additional questions, comments, or suggestions? (9 responses): **1)** Thanks for organising, great event! **2)** Optional training course beforehand would be really useful, they can be recording and notebooks shared virtually! **3)** Run events periodically around the world (places outside US and Europe, i.e., Asia and Australia) to increase awareness about NWB and DANDI, A bit longer hackathon would be great so that participants have more time for hacking **4)** Thanks for an awesome experience at NeuroDataReHack academy **5)** great workshop! thanks for organizing :) **6)** Thanks so much for hosting! **7)** Thanks for everything, it's been a true pleasure **8)** I think it would be great to see the DANDI Archive at open-source software development events, like Hacktoberfest by DigitalOcean and GitHub. I'm also very interested in becoming an "NWB format/DANDI expert" and promoting the use of the format at my university. If you

develop any strategies for this, I would love to participate. In the meantime, I will be looking for opportunities to promote it through my channels at my university. **9)** Sponsored by a Brain-Machine Interface Company

Testimonials

"DANDI is an excellent resources to exploratory work. DANDIhub is an awesome start point for novice without worrying about environment issues." – Petrina Lau, UCL.

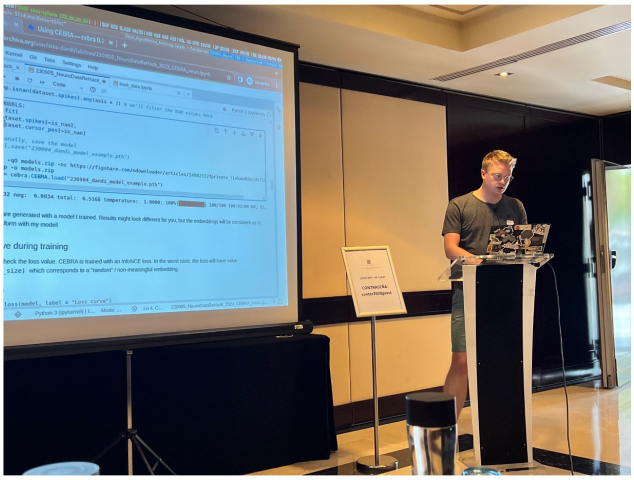
"The workshop has been a great opportunity to connect with other colleagues and, at the same time, learn how to use the NWB format as well as the DANDI Archive. Moreover, being in a group with great diversity, I was able to learn new techniques, perspectives, and ideas that will be very useful for my research. The organizers and all the participants were highly willing to support each other and learn in a fun and motivating environment. At the same time, there is an encouragement of the culture of sharing and facilitating access to science. Without a doubt, NeuroDataReHack 2023 has been a great experience. ." – Ricardo Velázquez, Institute of Neurobiology in Mexico.

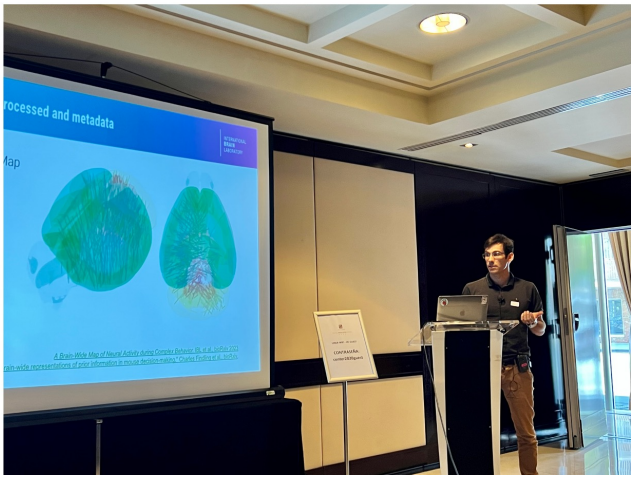
"I would encourage any computational and systems neuroscientists to try NWB and DANDI. They are great open-source tools for beginners as well as experts to dive into really interesting neuroscience problems." – Ishara Paranawithana, Monash University, Bionics Institute, Australia.

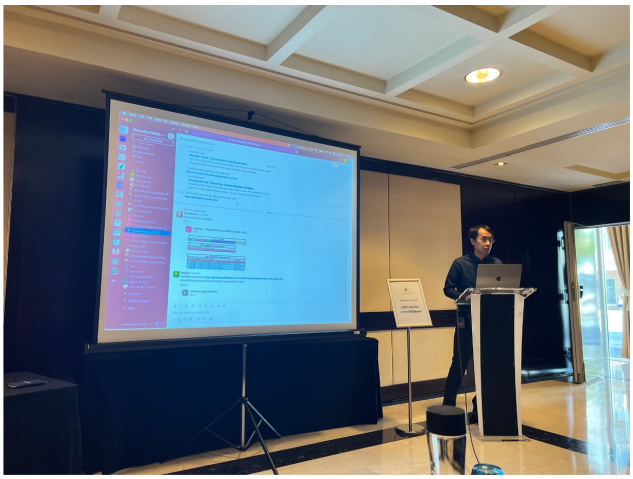
"NeurodataReHack is a great way to get together with scientists, infrastructure managers and dataset providers!" – Anonymous.

3.6 Photos







Bridging the gap between
function and structure


Neuronal activity in maternal response to pup
visual clues



Acknowledgements

Financial support for the event was provided by The Kavli Foundation. We would like to thank Stephanie Albin (The Kavli Foundation) for tireless organizational support. We also want to thank the Kavli Foundation for the Discovery Award RFP.

We would like to thank the presenters and hosts of the breakout sessions for their contributions to the workshop (see Sec. 3.2). We would like to thank all participants for the great enthusiasm and for making the event a great success!

Organization and operations of the event and NWB research activities were supported by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under Award Number U24NS120057 to O. Rübel (LBNL) and B. Dichter (CatalystNeuro).

Disclaimer