NIH BRAIN Circuit Program (U19 BCP)

BRAIN U19 BCP – Data Science Core Goals

1) Provide a service to the proposed U19 research components; and

2) Provide a prototype data science framework for facilitating the workflow for data aggregation and analysis between the proposed Research Components. The latter is considered a pilot effort in which high-risk-high impact methodologies are employed to optimize the framework for the U19 and produce generalizable approaches for the other BRAIN BCP efforts.

U19	Representative
A-Team	Guoqiang Yu (here)
Mouse V1	Liam Paninski (here)
Osmonauts	Rick Gerkin
TiM	Saskia de Vries (here) David Feng (here) Karel Svoboda (here)
Oxytocin	Peter Petersen (here) √
CausalityInMotion	Brooks M Musangu (here)
IBL-U19	Liam Paninski (here)
MoC3	Liam Paninski (here)
USArhythms	Gal Mishne (here)
Learning2Learn	Edgar Walker (here)
FlyLoops Team DOPE	Wei-Chung Lee (here) Mohammed Osman (here)

U19	**Representative**
Light-SPACE	Paul Nuyujukian ✓
MSCZ	Josh Vogelstein (Josh on paternity leave) Jeff Lichtman
DOPE	Mohammed Osman (Here)
Ripple	Ivan Raikov ✓
BRAIN CoGS	Manuel Schottdorf (here)
Sensation	Wolfgang Losert (here)
aABC	Yoav Freund
SCC	Tatiana Sharpee (here)

Tuesday July 5, 2022 - Planning meeting

1) Review current activities on Goal 2 from each U19 (using template in slide #3)
    a) IBL U19 pipelines
        i) Neuropixel, Ca Imaging
    b) Synergies (**a lot of overlap and redundancy!**)
        i) pre-processing of Ca Imaging data
        ii) Multiple ways to do similar things
2) Aggregating tools and methods → through **Common Interfaces**
    a) Share processing pipelines in Public Cloud; NIH to help with reducing costs (STRIDES, https://cloud.nih.gov/)
        i) Persistent storage on cloud is too expensive, need local solutions (or group-wide: Google Workspace Enterprise Basic Account); need to choose which long-term expense you want
        ii) Cost-effective - Spot Instance Kubernetes cloud processing, but need to copy the data all the time
    b) Common file formats and data conventions - are interoperable tools possible?
        i) Context dependent on downstream analysis needs
        ii) Standardize interfaces between tools (rather than committing to specific standards)
        iii) Terminology
    c) Protocols for common interfaces → need an API (NOT A FILE FORMAT)
        i) Users can load up their library and implement the API
        ii) Minimal development needed, already established interfaces
        iii) Thin wrappers around hardened API's

Tuesday July 5, 2022 - Planning meeting

1) Decide on Goals for next Data Science Consortium Meeting
    a) Best practices - high level commonalities across the Data Science Cores
    b) Barriers and challenges
    c) Making tools accessible (intuitive) at scale
    d) How circuit operate under perturbations (e.g.: stability, improving performance)
    e) What are the Common Interfaces needed?
2) Seek volunteers for co-chairs, moderators, speakers
    a) Common API development → white paper
    b) **Edgar, Manuel, Guoqiang**
3) Determine date(s), logistics
    a) Webinars - demo tech stacks, tutorials - each speaker sets up their own method for sharing
    b) Regular U19 subgroup meetings - Ca Imaging
    c) Modeling subgroup - common theories across U19's
    d) Behavioral subgroup - reboot on current state of the art

[U19 NAME] - Goal 2 activities

1. Title of a success story
2. Collaborations have you formed with other U19 BCPs?
3. Brief description of prototype framework (sandbox/pilot) you are creating (e.g.: workflows, analytical/computing resources)?
4. Domains are you reaching?
5. Outcomes?
6. Challenges/barriers?
7. **Basic technology of current API, tech stacks**
8. **Data API wishlist**

LIGHTspace - Goal 2 activities

1. Download-free dashboards to view/analyze data: HDF5 ros3 against DANDI archive
    a. View demo at lightspace.stanford.edu
2. No inter-U19 collaborations yet
3. Prototype: GitLab CI/CD end-to-end computational pipelines on Kubernetes (GCP GKE Autopilot, GitLab Kubernetes executor), with a common, intuitive Python API for data access, including uploading from acquisition systems for experimentalists (local GitLab shell executor)
4. Primarily targeting systems neuroscientists and computational modelers across rodent & NHP experimental sources
5. Outcomes: Research Data & Computation course offered (BIOE 301P), staff hires, formulation of training curriculum for students/postdocs
6. Challenges: Education, education, education…
7. Conduct training workshops on how to develop and deploy computational pipelines

Sensation - Goal 2 activities

1. NeuroWRAP and NeuroART
2. Collaborations have you formed with other U19 BCPs? Not yet
3. Brief description of prototype framework (sandbox/pilot) you are creating (e.g.: workflows, analytical/computing resources)?

    NeuroWRAP: Workflow integrator for Reproducible Analysis of multi-Photon NEUROimaging. It allows researchers to combine tools they already use in multiple languages, organizes analysis in a reproducible way, and includes validation through consensus analysis and metamorphic testing

    https://neurowrap.org/

    NeuroART: Real-time calcium image analysis to integrate data into theory-driven experimental protocols. Incorporating data analysis tools from multiple theory groups that are part of the U19.

1. Domains are you reaching? Aim to reach Calcium imaging community
2. Outcomes? Developing workflow with input from 5 experimental groups and 3 theory groups that are part of the U19. One conference proceeding. Two papers in preparation.
3. Challenges/barriers?
4. Next steps?

Braincogs (Circuits of cognitive systems) - Goal 2 activities

1. (i) Shared training & mouse management pipelines. (ii) Lightsheet + Histology pipelines. (iii) Preprocessing pipelines for ePhys & Calcium imaging.
2. We have profited substantially from other U19 / IBL (similar tools, dj etc.) & document/publish our tools. No direct collaborations.
3. (i) SQL DB backend, used by everyone, ~50GB. Access via simple GUIs and/or datajoint. Bulky raw data external (local university servers). (ii) Shared code base on github. All researchers have to commit, rigs auto-pull. Critical code with pull-requests.
4. (i) Systems neuroscientists, (ii) theory researchers, (iii) technicians running training through standardized GUIs (all behavior is virtual reality for head-fixed mice)
5. Outcomes: Automated training and pre-processing. Open pipelines and code on github (e.g. link). Staff hires.
6. Challenges: (i) Continuous integration of pipelines in highly dynamic projects, (ii) integration of university infrastructure. (iii) Getting researchers to contribute to shared infrastructure, and (iv) curate data beyond what is immediately necessary.
7. (i) Developing intuitive GUIs to further lower energy barrier (Done for training, at the moment for ePhys/imaging). (ii) Conduct training workshops on how to develop, improve and use pipelines.

TiM - Thalamus in the Middle - Goal 2 activities

1. Mouselight workstation in the cloud
2. None so far
3. Cloud data lake; evaluation of database, workflow, analysis tools (Datajoint, Django, Airflow, Vertex, etc)
4. Systems/computational neuroscientists; BICCN;
5. Outcomes: Selected cloud analytics platform (CodeOcean); defined metadata schema for multiple modalities; Hiring data & infrastructure engineers. Choose compression strategies for imaging and ephys.
6. Challenges: Cloud pricing; cost sharing for open science; bringing users to our data
7. Next steps: Implementation of platform and workflows

Oxytocin - Goal 2 activities

1. Success stories:
   a. CellExplorer - Framework for analyzing single cells: www.cellexplorer.org
   b. BrainSTEM - A collaborative electronic lab notebook for experimental neuroscience: www.brainstem.org
2. Collaborations: None so far
3. Prototype framework: BrainSTEM: A collaborative electronic lab notebook for experimental neuroscience.
4. Domains you are reaching? Electrophys, imaging, behavioral communities, systems neuroscience.
5. Outcomes? CellExplorer published, BrainSTEM public demo site
6. Challenges/barriers: finding pilot groups!
7. Next steps:
   a. Perform beta testing/recruit pilot groups
   b. API to interface with analysis tools and the rest of the FAIR ecosystem
   c. BioRxiv manuscript and documentation of BrainSTEM

IBL-U19, MouseV1, MoC3

1. Neuroscience Cloud Analysis as a Service (NeuroCAAS), Open Neurophysiology Environment (ONE), Alyx, pykilosort / SpikeInterface (IBL)
2. IBL-U19, MouseV1, MoC3
3. Currently developing cloud pipelines for behavioral video analysis and calcium / voltage compression / demixing
4. Systems and computational neuroscience
5. Several large raw and processed datasets released, eg https://www.internationalbrainlab.com/data
6. Robustness of current pipelines across diverse datasets
7. Hoping to share complete behavioral-vid and ca-imaging cloud pipelines by end of summer - feedback very welcome!

[Cervical Spinal Cord] - Goal 2 activities

1. *Interactive 3D Cervical Spinal Cord Atlas, Neural and biomechanical model for controlling flexor and extensor muscles during rhythmic motor behaviors*
2. Collaborations: *with ABC U19 to elucidate common principles in the organization of brainstem and spinal cord circuits controlling rhythmic and discrete motor behaviors*
3. Brief description of prototype framework (sandbox/pilot) you are creating (e.g.: workflows, analytical/computing resources)? *Software scripts and plug-ins for 3d atlas; Web-based interface to navigate anatomical, molecular, ephys and connectome data for cardinal classes and sub-classes of premotor neuron types in the spinal cord*
4. Domains are you reaching?: *Systems and computational neuroscientists working on sensorimotor circuits and movement, imaging and electrophysiological recording in the cord; Behavioral analysis, Molecular and Cellular Neuroscience*
5. Outcomes: *genetic signatures of neuronal classes in the cord (Ossewald et al Science 2021), 6 abstracts submitted this year to SfN; 1 paper submitted and 3 papers in preparation (a)3 papers in preparation (Neural and biomechanical model for rhythmic behavior, In vivo spinal recordings, 3D Atlas)*
6. Challenges/barriers?: .*Development of imaging data pipeline that accepts and integrates datasets from other groups into reference atlas. Understanding circuit stability/flexibility to changes in neural circuits and external perturbations in modeling studies; Integrating external datasets into our 3D Atlas reference atlas.*
7. Next steps:
   a) incorporating connectivity, ephys, scRNAseq and electrophysiology datasets into the 3D atlas;
   b)Refining the current neural and biomechanical models, with ephys, connectomics and perturbations data to simulate and analyze more complex rhythmic (e.g. rope pulling) and discrete (e.g. joystick) behaviors.
   c)Further our understanding of circuit stability to changes in neural circuits and external perturbations

Ripple - Goal 2 activities

1. NeuroH5: Parallel Large-Scale I/O Workflow for Computational Neuroscience Modeling
2. U19 BCP collaboration: not yet
3. Parallel large-scale computational data/simulation/analysis framework for biophysical neuroscience models
4. Domains: computational neuroscience, experimental ephys (tentative)
5. Outcomes: mature, flexible, highly efficient I/O infrastructure for parallel scientific computing environments
6. Challenges/barriers: currently difficult to bridge with NWB and other efforts related to experimental data formats, resources for software maintenance and user education
7. Next steps: possible collaborations with NSF Expeditions program

FlyLoops - Goal 2 activities

1. Title of a success story: ***brainmaps.io***
2. Collaborations have you formed with other U19 BCPs? ***Not that I know of***
3. Brief description of prototype framework (sandbox/pilot) you are creating (e.g.: workflows, analytical/computing resources)? ***Integrating connectomic data across datasets, tools, and models***
4. Domains are you reaching? ***Beyond U19 participant labs, endusers are mainly fly neuroscience researchers***
5. Outcomes? ***Web-based user interface for extracting fly connectomics data across datasets; data and code resources (https://projects.iq.harvard.edu/flyloops/resources)***
6. Challenges/barriers? ***Integrating new datasets as they come online, outputs into computational models***
7. Next steps? ***Integrating new datasets as they come online, outputs into computational models***

Team DOPE - Goal 2 activities

1. Systems Neuro Visualization and Analysis Tools: GLM Pipeline and Interactive Data Browser
2. N/A
3. Developing an open source pipeline for generalized linear model analysis of neural data as well as an interactive browser for neural and behavioral data
4. Systems/computational neuroscience
5. Tools developed and available on github (https://github.com/calebweinreb/SNUB)
6. Variation in data storage formats across labs in consortium
7. Address (6) by implementing NWB support

Learning 2 Learn - Goal 2 activities

1. AWS-based data sharing and common development environment
2. No outside collaborations yet
3. Data storage/sharing S3, JupyterHub hosted on Kubernetes on AWS, DataJoint for pre-processing data pipelines
4. Systems/computational neuroscience, machine learning
5. Outcomes: Centrally accessible human and NHP datasets and simple to use Python-based development environment with scaling
6. Challenges/barriers: 1. Integration with existing local infrastructure. 2. Wider adoption of the cloud-based tools, and 3.contribution into the common analysis code
7. Next steps: Integration of on-premise pre-processing of data (e.g. spike sorting) as part of the shared data processing pipeline through better integration of on-premise computation workflow