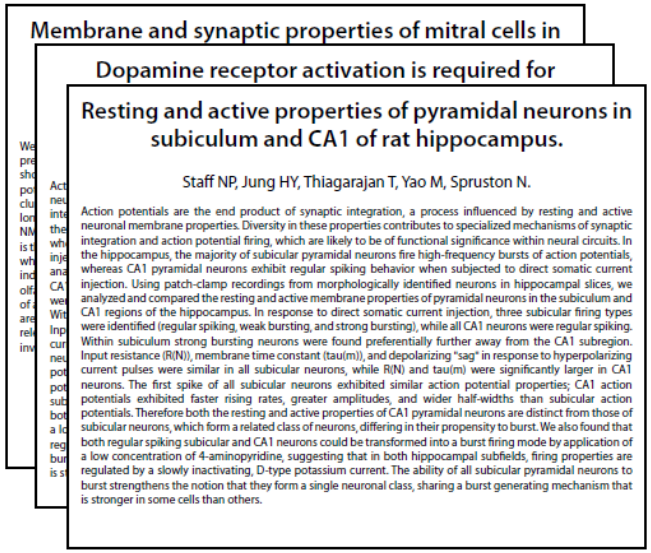
Insights on open data and metadata from a neuroinformaticist

Shreejoy Tripathy, Post-Doc
University of British Columbia
Email: stripathy@chibi.ubc.ca
Twitter: @neuronJoy

NeuroElectro.org: text-mining neuron properties from the existing literature

Electrophysiological studies



Membrane and synaptic properties of mitral cells in

Dopamine receptor activation is required for

Resting and active properties of pyramidal neurons in subiculum and CA1 of rat hippocampus.

Staff NP, Jung HY, Thiagarajan T, Yao M, Spruston N.

Action potentials are the end product of synaptic integration, a process influenced by resting and active neuronal membrane properties. Diversity in these properties contributes to specialized mechanisms of synaptic integration and action potential firing, which are likely to be of functional significance within neural circuits. In the hippocampus, the majority of subicular pyramidal neurons fire high-frequency bursts of action potentials, whereas CA1 pyramidal neurons exhibit regular spiking behavior when subjected to direct somatic current injection. Using patch-clamp recordings from morphologically identified neurons in hippocampal slices, we analyzed and compared the resting and active membrane properties of pyramidal neurons in the subiculum and CA1 regions of the hippocampus. In response to direct somatic current injection, three subicular firing types were identified (regular spiking, weak bursting, and strong bursting), while all CA1 neurons were regular spiking. Within subiculum strong bursting neurons were found preferentially further away from the CA1 subregion. Input resistance (Ri(N)), membrane time constant (tau(m)), and depolarizing "sag" in response to hyperpolarizing current pulses were similar in all subicular neurons, while Ri(N) and tau(m) were significantly larger in CA1 neurons. The first spike of all subicular neurons exhibited similar action potential properties; CA1 action potentials exhibited faster rising rates, greater amplitudes, and wider half-widths than subicular action potentials. Therefore both the resting and active properties of CA1 pyramidal neurons are distinct from those of subicular neurons, which form a related class of neurons, differing in their propensity to burst. We also found that both regular spiking subicular and CA1 neurons could be transformed into a burst firing mode by application of a low concentration of 4-aminopyridine, suggesting that in both hippocampal subfields, firing properties are regulated by a slowly inactivating, D-type potassium current. The ability of all subicular pyramidal neurons to burst strengthens the notion that they form a single neuronal class, sharing a burst generating mechanism that is stronger in some cells than others.

→

Semi-automated property extraction

Electrophysiological values

	CA1 pyramidal cell
V_{rest}	−66.2 ± 1.1 mV
IR	55.4 ± 3.7 MΩ
sag ratio	0.79 ± 0.01
$AP_{half\text{-}width}$	0.95 ± 0.03 ms

Experimental metadata

Strain	Wistar
Animal Age	14 - 63 days
Rec. Temp	35.0 ± 2.0°C
...	...

inspired by Aaron Swartz,
IBM's Watson, neurosynth.org

Tripathy et al, 2013. *Front Neuroinf*
Tripathy et al, 2015. *J Neurophys*

Large-scale data extraction from published texts

Table 1. Properties of Hippocampal Pyramidal Cells

	Regular Spiking
Electrophysiology	
R_N	55.0 ± 3.1 MΩ
Sag ratio	0.84 ± 0.02
Subthreshold dV/dt	2.61mV/ms ± 0.53mV/ms
ADP amplitude	17.9mV ± 1.3mV

- Simple algorithms that use simple text searching to identify:
  - Biophysical properties (in normotypic conditions)
  - Neuron types (from NeuroLex.org)
  - Biophysical data values
  - Methodological details
- Text-mined data is then checked by human curators

What does it mean to “share data”?

- Publishing a paper/book
- Making an Excel spreadsheet of “the data” available, either as supplemental data or on FigShare
- Making the raw data available (and the appropriate metadata!)

“Can I upload my raw data to your repository?”

Marianne Bezaire from UCI

Ephys data for NeuroElectro

casem to stripathy, rickg, nurban

Hi Shreejoy,

I am considering publishing a data article in the new nature journal "Scientific Data". It will contain some ephys data from experimentalists in my lab for various identified interneurons in the hippocampus, which I am going to analyze in a standard format to get the ephys properties from. I would like to then upload the data to NeuroElectro.

The journal lists recommended repositories, but they do not have NeuroElectro yet: http://www.nature.com/scientificdata/?page_id=777&preview=true#recommended-repositories, perhaps because they are focused on repositories that store raw data.

I would need to upload both my calculated ephys properties and the original AxoClamp traces to a community repository. The journal suggests generic repositories if I can't find a specific ephys one, but I was thinking it would be really nice to be able to upload AxoClamp traces directly to NeuroElectro. Also, since I will be tabulating all the internal and external recording solution components, it would be nice to store those tables in NeuroElectro as well, linked to the AxoClamp traces and possibly the calculated ephys properties.

Sharing data works…

- When everyone knows what exactly it means to share data in that subfield
- Repositories
- Standards
- File Formats
- Publisher/Funder mandates and using reviewers to check for compliance

“Sharing data” isn’t new in genetics

nature International weekly journal of science

Mandates for specific datasets

For the following types of data set, submission to a community-endorsed, public repository is mandatory. Accession numbers must be provided in the paper. Examples of appropriate public repositories are listed below.

Mandatory deposition	Suitable repositories
Protein sequences	Uniprot
DNA and RNA sequences	Genbank
	DNA DataBank of Japan (DDBJ)
	EMBL Nucleotide Sequence Database (ENA)
DNA and RNA sequencing data	NCBI Trace Archive
	NCBI Sequence Read Archive (SRA)
Genetic polymorphisms	dbSNP
	dbVar
	European Variation Archive (EVA)
Linked genotype and phenotype data	dbGAP
	The European Genome-phenome Archive (EGA)
Macromolecular structure	Worldwide Protein Data Bank (wwPDB)
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)
Microarray data (must be MIAME compliant)	Gene Expression Omnibus (GEO)
	ArrayExpress
Crystallographic data for small molecules	Cambridge Structural Database

NCBI

GEO Gene Expression Omnibus

GEO Publications | FAQ | MIAME | Email GEO

NCBI » GEO » Info » GEO and MIAME Login

GEO and MIAME (Minimum Information About a Microarray Experiment)

The MIAME guidelines outline the minimum information that should be included when describing a microarray experiment. Many journals and funding agencies require microarray data to comply with MIAME. GEO deposit procedures enable and encourage submitters to supply MIAME compliant data.

More information and background regarding GEO and MIAME are discussed in this **Nature Biotechnology correspondence**.

MIAME compliance is not related to the submission format or route, but rather to the content provided

The six most critical elements contributing towards MIAME are:

- The raw data for each hybridization (e.g., CEL or GPR files)
- The final processed (normalized) data for the set of hybridizations in the experiment (study) (e.g., the gene expression data matrix used to draw the conclusions from the study)
- The essential sample annotation including experimental factors and their values (e.g., compound and dose in a dose response experiment)
- The experimental design including sample data relationships (e.g., which raw data file relates to which sample, which hybridizations are technical, which are biological replicates)
- Sufficient annotation of the array (e.g., gene identifiers, genomic coordinates, probe oligonucleotide sequences or reference commercial array catalog number)
- The essential laboratory and data processing protocols (e.g., what normalization method has been used to obtain the final processed data)

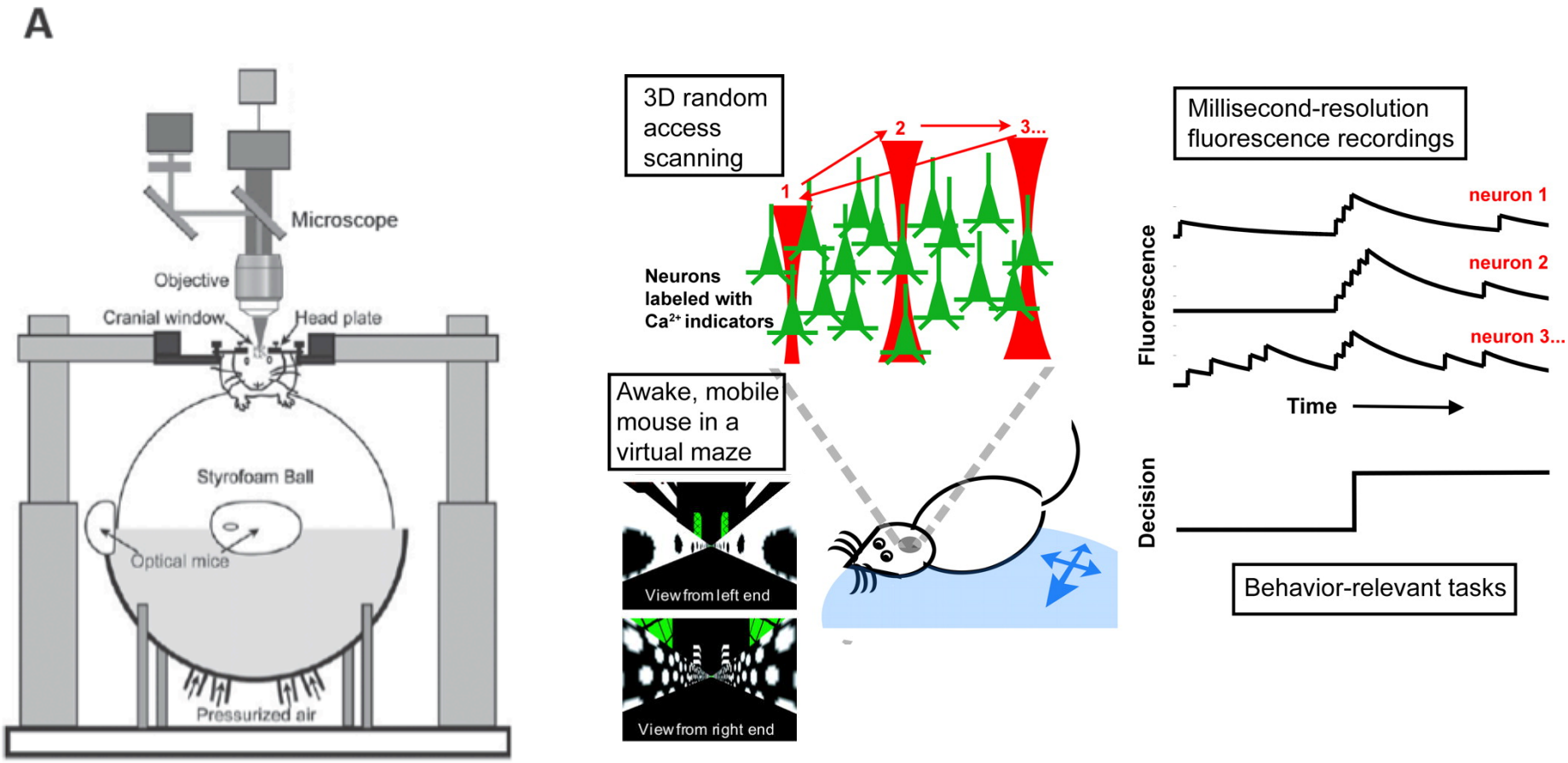
http://www.nature.com/authors/policies/availability.html

http://www.ncbi.nlm.nih.gov/geo/info/MIAME.html

Some questions on data sharing

- Ideally, every data creator would make their data publically available
- Is it fair to ask a researcher to make their data open? If doing so is (seriously) burdensome (as opposed to just a bit annoying)
- Is is helpful for researchers to make their data open, if it isn't all that well described or use standard metadata?

Data can be very complicated



Harvey et al, 2009; Kwan et al 2008

Neurodata-without-borders project

- A one year project to make a usable file format and data standards for neurophysiology data
- Produced during small working group meetings of data collectors and curators
- Designed around specific datasets which are typical of the subfield

https://crcns.org/NWB/Overview

Data sharing for neuroscientists

PLOS

Join the community
Sign up to get email alerts

Neuroscience Community

Submit your research

Subscribe to this page

Neuroinformaticists Decode the PLOS Data Policy at INCF 2014

Posted on September 3, 2014 — No Comments ↓

By Shreejoy Tripathy

The concept of data sharing is central to the annual Congress of Neuroinformatics, organized by the INCF (International Neuroinformatics Coordinating Facility).

> *We as a community recognize that many advances could be made in neuroscience if authors went beyond simply describing their data within a publication to sharing and uploading the entire dataset used to draw scientific conclusions.*

Monthly archives

Select Month

Recent posts

How common is "p-hacking"? Neuroskeptic interviews Megan Head April 9, 2015

Multisensory integration and causal inference in the brain April 3, 2015

This is your Brain on Wine: FMRI Signals of Alcohol Content March 25, 2015

Multimodal measures of brain

http://neuro.plos.org/2014/09/03/neuroinformaticists-decode-the-plos-data-policy-at-incf-2014/

Acknowledgements

- Pavlidis Lab @ UBC
- Urban Lab @ CMU
- Gerkin Lab @ ASU

Shreejoy Tripathy
Email: stripathy@chibi.ubc.ca
Twitter: @neuronJoy
URL: neuroelectro.org
Code: github.com/neuroelectro

ALLEN INSTITUTE for BRAIN SCIENCE
Fueling Discovery
NIF
incf International Neuroinformatics Coordinating Facility
Elsevier Research Data Services