



ALLEN INSTITUTE *for*  
BRAIN SCIENCE

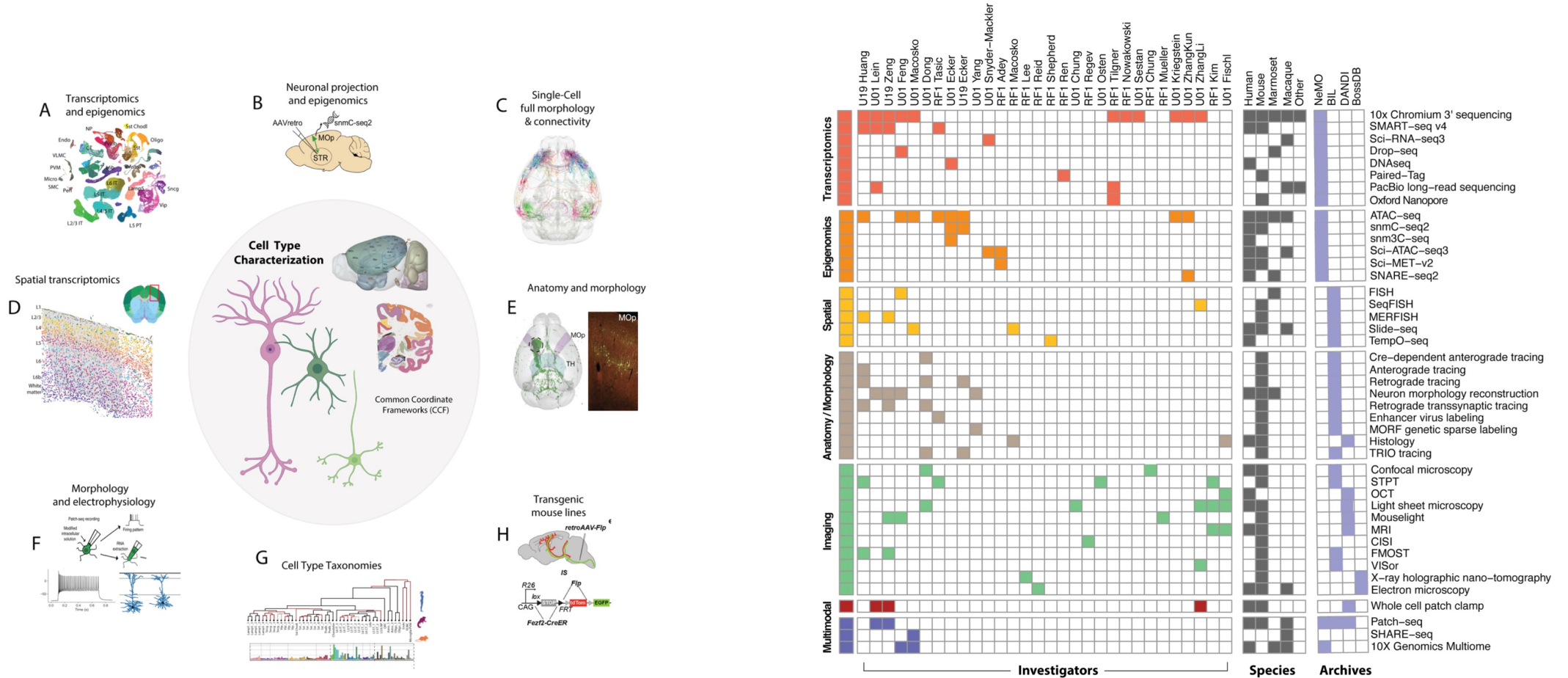
# ***The BRAIN Initiative Cell Census Data Ecosystem***

***Developing an International Data Governance Framework to  
Accelerate Global Neuroscience Research***

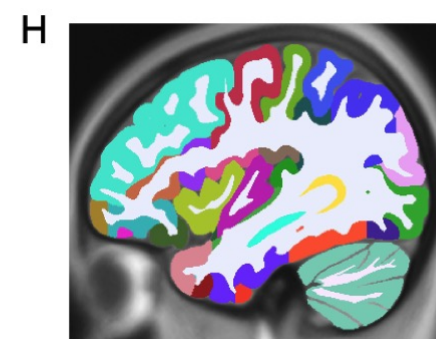
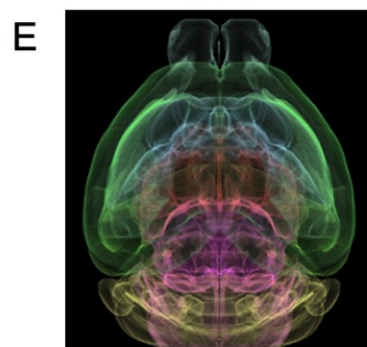
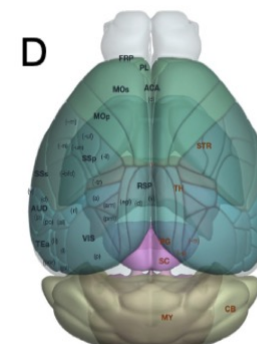
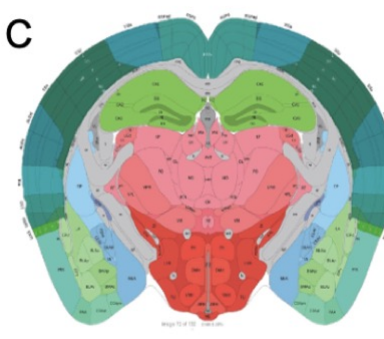
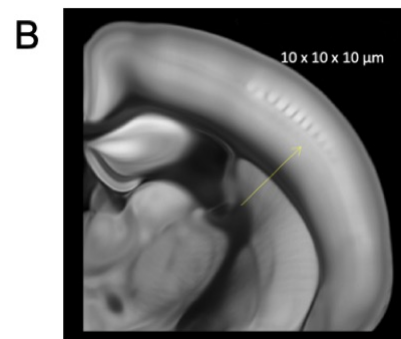
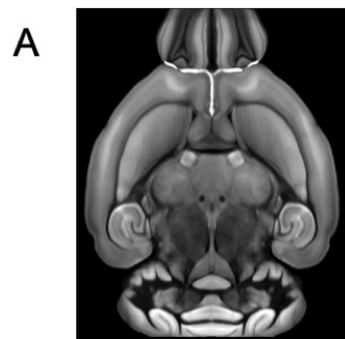
November 14, 2022

Michael Hawrylycz, *Allen Institute for Brain Science*

# BICCN Profiling Cell Type of Brain

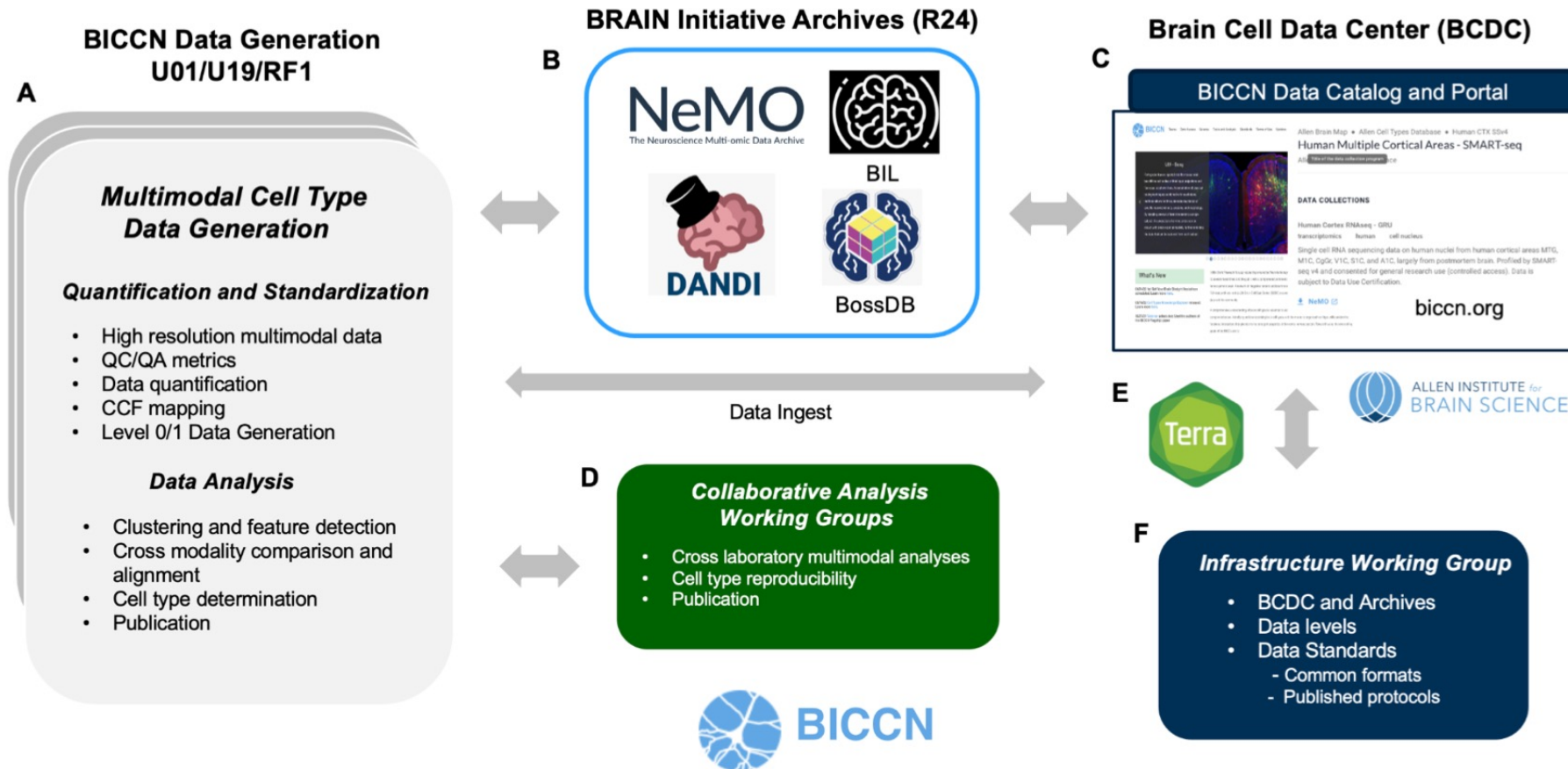


# Common Coordinate Frameworks





# BICCN Data Ecosystem



# BICCN Data Portal: [biccn.org](https://biccn.org)

### A Information Portal

 [Teams](#) [Data Access](#) [Science](#) [Tools and Analysis](#) [Standards](#) [Terms of Use](#) [Updates](#)



**BICCN Data Inventory**  
Learn about and access data from the network of BICCN centers and laboratories. Data spans multiple species, modalities, and techniques.

#### What's New

- 04/04/22 1st Get Your Brain Hackathon scheduled. [Learn more here.](#)
- 03/16/22 Cell Types Knowledge released. [Learn more here.](#)
- 10/27/21 Webinar scheduled: authors of the BICCN Flagship

### B Metadata and File Manifests

View by: Details | Sort by: Default

Filter

| Specimen ID           | Specimen Type | Species | Subject ID | Age... | Sex    | Genotype                                               | Subspecimen T  |
|-----------------------|---------------|---------|------------|--------|--------|--------------------------------------------------------|----------------|
| mouseID_497523_196469 | whole brain   | mouse   | 497523     | 11     | female | Calb2-CreERT2/wt;Ai166(TIT2L-MORF-ICL-1TA2)/wt         | reconstruction |
| mouseID_481967_194077 | whole brain   | mouse   | 481967     | 11     | male   | Plxnd1-CreER/wt;Ai166(TIT2L-MORF-ICL-1TA2)/wt          | reconstruction |
| mouseID_491994_196472 | whole brain   | mouse   | 491994     | 16     | female | Pdyn-T2A-CreERT2/wt;Ai166(TIT2L-MORF-ICL-1TA2)/wt      | reconstruction |
| mouseID_443055_191805 | whole brain   | mouse   | 443055     | 9      | male   | Dbh-Cre_KH212/wt;Ai166(TIT2L-MORF-ICL-1TA2)/wt         | reconstruction |
| mouseID_512894_201605 | whole brain   | mouse   | 512894     | 10     | male   | Pdyn-T2A-CreERT2/wt;Ai166(TIT2L-MORF-ICL-1TA2)/wt      | reconstruction |
| mouseID_339952_17782  | whole brain   | mouse   | 339952     | 14     | female | Gnb4-IRES2-CreERT2/wt;Ai82(TITL-GFP)/Ai140(TIT2L-G...  | reconstruction |
| mouseID_426812_191820 | whole brain   | mouse   | 426812     | 8      | female | Tac1-IRES2-Cre/wt;Sst-IRES-FlpO/wt;Ai188(TA2T21-EGE... | reconstruction |

### C Searchable Data Catalog

79 Projects

BICCN • U01 Lein • Cross-species M1  
**Cell type diversity in primary motor cortex across mammalian evolution**  
Ed Lein transcriptomics  
chimpanzee, common tree shrew, pig-tailed macaque, domestic cat, rhesus macaque, Norway rat, western gorilla, Arctic ground squirrel, green monkey, Mats night monkey, gray short-tailed opossum, nine-banded armadillo, rabbit, domestic ferret, pig, small-eared galago  
library, cell nucleus 10x Chromium T v3 sequencing  
10x v3 scRNA-seq data from primary motor cortex of multiple mammalian species

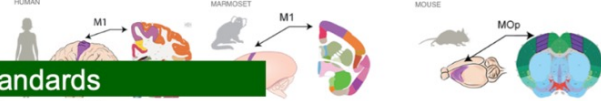
BICCN • U01 Sestan • Human, NHP and pig cell types  
**Developmental cell census of human, non-human primate and pig brain**  
Natal Sestan transcriptomics marmoset, chimpanzee, rhesus macaque, human, pig brain region  
10x Chromium T v3 sequencing  
Single cell/nucleus transcriptomics data to define cell types and states in multiple brain regions of human, rhesus macaque, chimpanzee, common marmoset and pig.

Integrative functional genomic analysis of human brain development and neuropsychiatric risks

### D Scientific Summary

Explore the BICCN Data Inventory  
This includes quick links to BICCN Data Catalog with filtered searches.

By Species



### E BICCN Standards

#### BICCN Standards, Best Practices, and Recommendations

As per FAIR, the BICCN seeks to utilize existing standards where they are available, while developing BICCN standards as necessary to fill in the gaps. The archives through the Infrastructure Working Group to ensure that standards are harmonized across the consortium where possible. Below are lists of standards BICCN.

#### Metadata and File Formats

**BCDC Metadata Standard for Collections and Samples**  
High level descriptive metadata to describe BICCN data collections. Instructions and templates for providing information about projects and data collections, and their descriptive metadata which must be submitted for each data collection milestone. **In use by:** [BCDC for the BICCN Data Catalog](#).

**Basic descriptive metadata for samples and specimens profiled in BICCN.** Instructions and templates for data submission are provided. An inventory of samples and associated cells or brains profiled for each sample for each dataset, reported on a quarterly basis. This is suitable for identifying the transgenic line and anatomic region sampled in a manageable spreadsheet, but is not intended to include more detailed provenance. **In use by:** [BCDC for the BICCN Data Catalog](#).

**Essential Metadata for 3D Microscopy**  
Metadata standard for a 3D microscopy dataset to support reuse by scientists who did not generate the data. 3D microscopy data submitted to BIL must adhere to this metadata standard. **In use by:** [BIL](#), [Data Archive](#).

**NeuroData Without Borders (NWB2.0)**  
Data standard for neurophysiology, providing neuroscientists with a common standard to share, archive, use, and build common analysis tools for neurophysiology data. **In use by:** [DANDI Archive](#). **Endorsed by:** [INCF](#).

**Project Instructions**  
[Data Collection Inventory](#)  
[BICCN Data Catalog](#)

**Sample Instructions**  
[Sample Inventory](#)  
[BICCN Data Catalog](#)

**Brain Image Library**  
[Data model](#)

**NWB** [help pages](#)  
[DANDI archive](#)  
[documentation](#)

[ELECTROPHYSIOLOGY](#)[SPATIAL TRANSCRIPTOMICS](#)[OPTOMICS AND EPIGENOMICS](#)[POPULATION IMAGING](#)

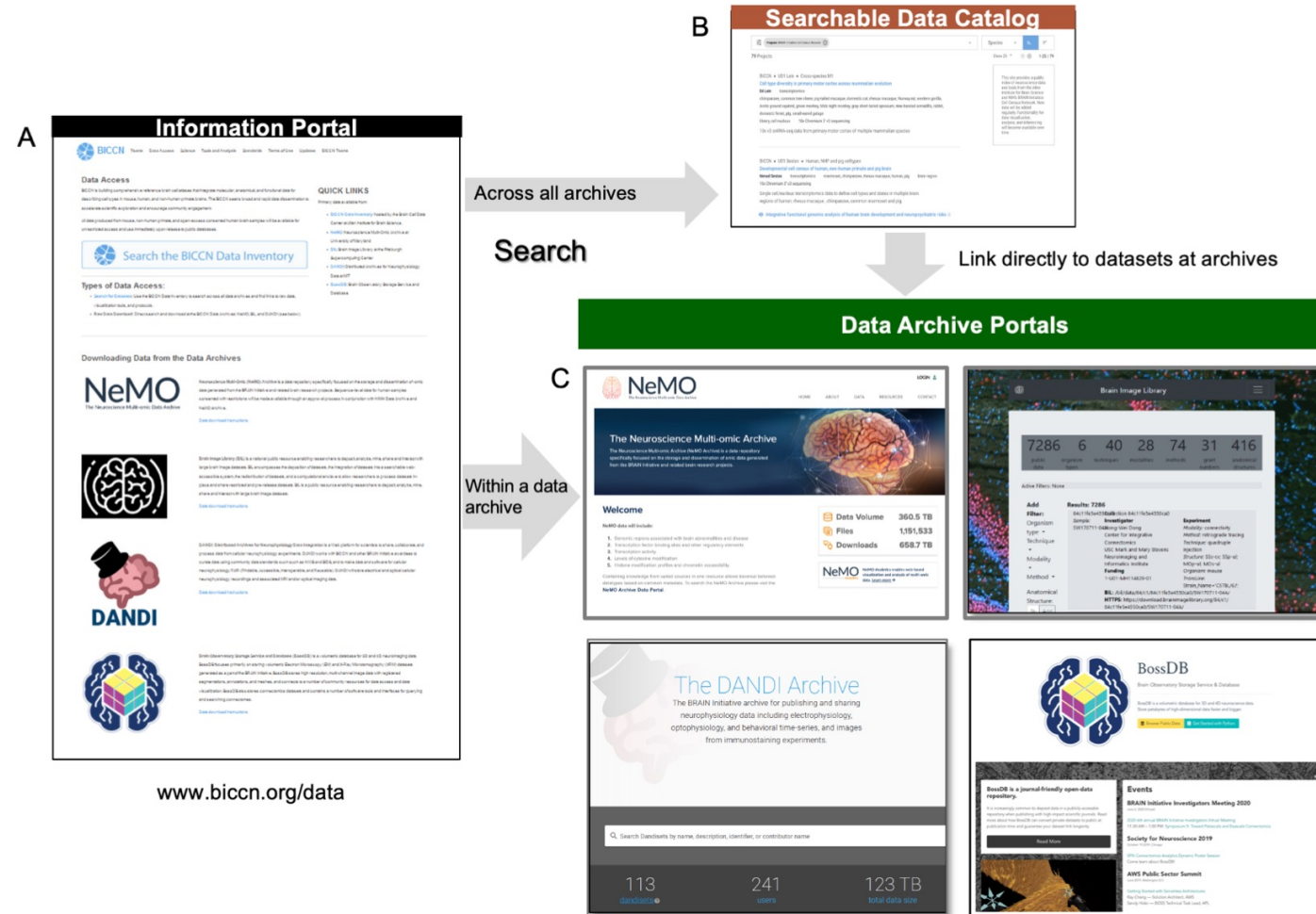
#### Primary Motor Cortex

Cell Census of the Primary Motor Cortex.

And article DOIs and to visualization tools that support access to data highlighted in those papers.

Carol Thompson

# Brain Initiative Data Archives

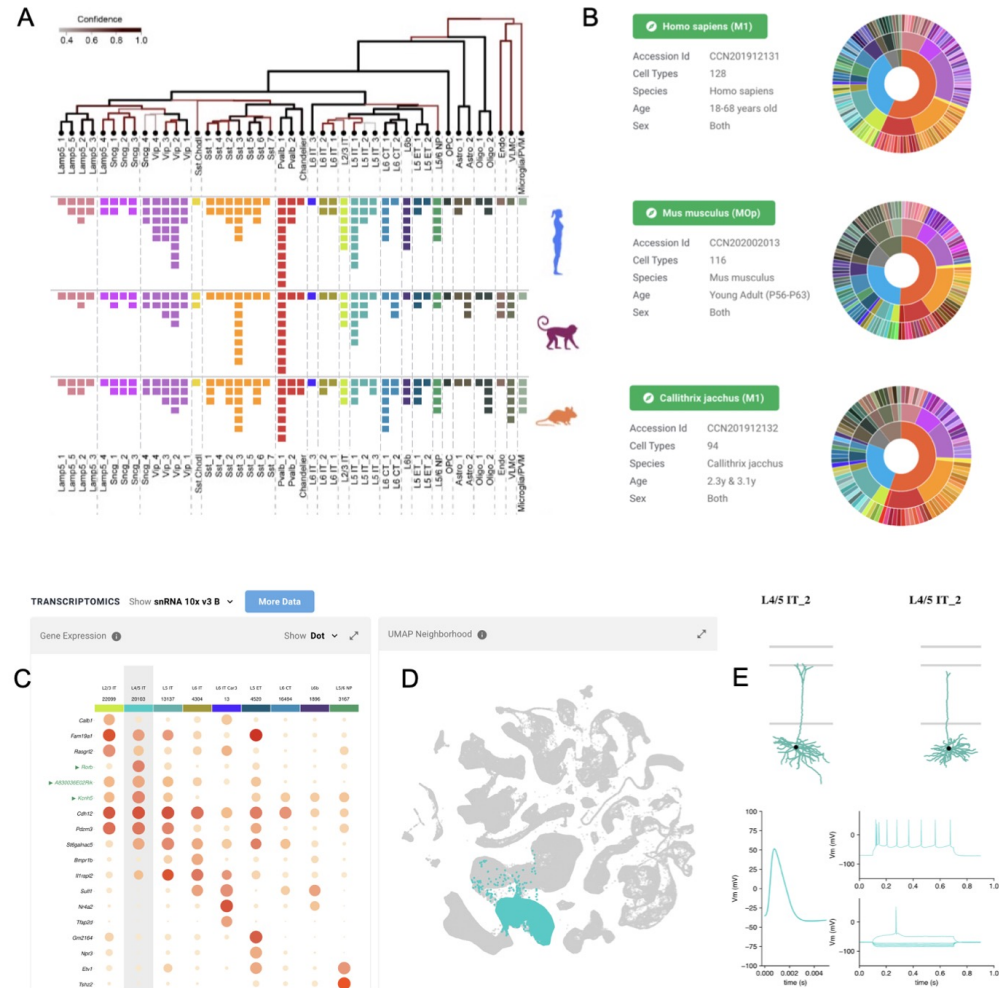


# Standardized Molecular Pipelines

| Pipeline                                                    | WARP WDL Code                          | Input Data                                                    | Overview                                        | Terra Workspace                        |
|-------------------------------------------------------------|----------------------------------------|---------------------------------------------------------------|-------------------------------------------------|----------------------------------------|
| Smart-seq2 Single Nucleus Multi-Sample<br>(RRID:SCR_021312) | Smart-seq2 Single Nucleus Multi-Sample | Single-cell data generated with Smart-seq2 assays             | Smart-seq2 Single Nucleus Multi-Sample Overview | Smart-seq2 Single Nucleus Multi-Sample |
| Optimus<br>(RRID:SCR_018908)                                | Optimus                                | 10x Genomics v2 and v3 3' single-cell and single-nucleus data | Optimus Overview                                | Optimus                                |
| Single-Cell ATAC (scATAC; RRID:SCR_018919)                  | scATAC                                 | Single-cell ATAC-seq data from nuclear isolates               | scATAC Overview                                 | scATAC                                 |
| (CEMBA; RRID:SCR_021219)                                    | CEMBA                                  | Multiplexed single-nucleus bisulfite sequencing data          | CEMBA Overview                                  | CEMBA                                  |

Terra, Broad Institute

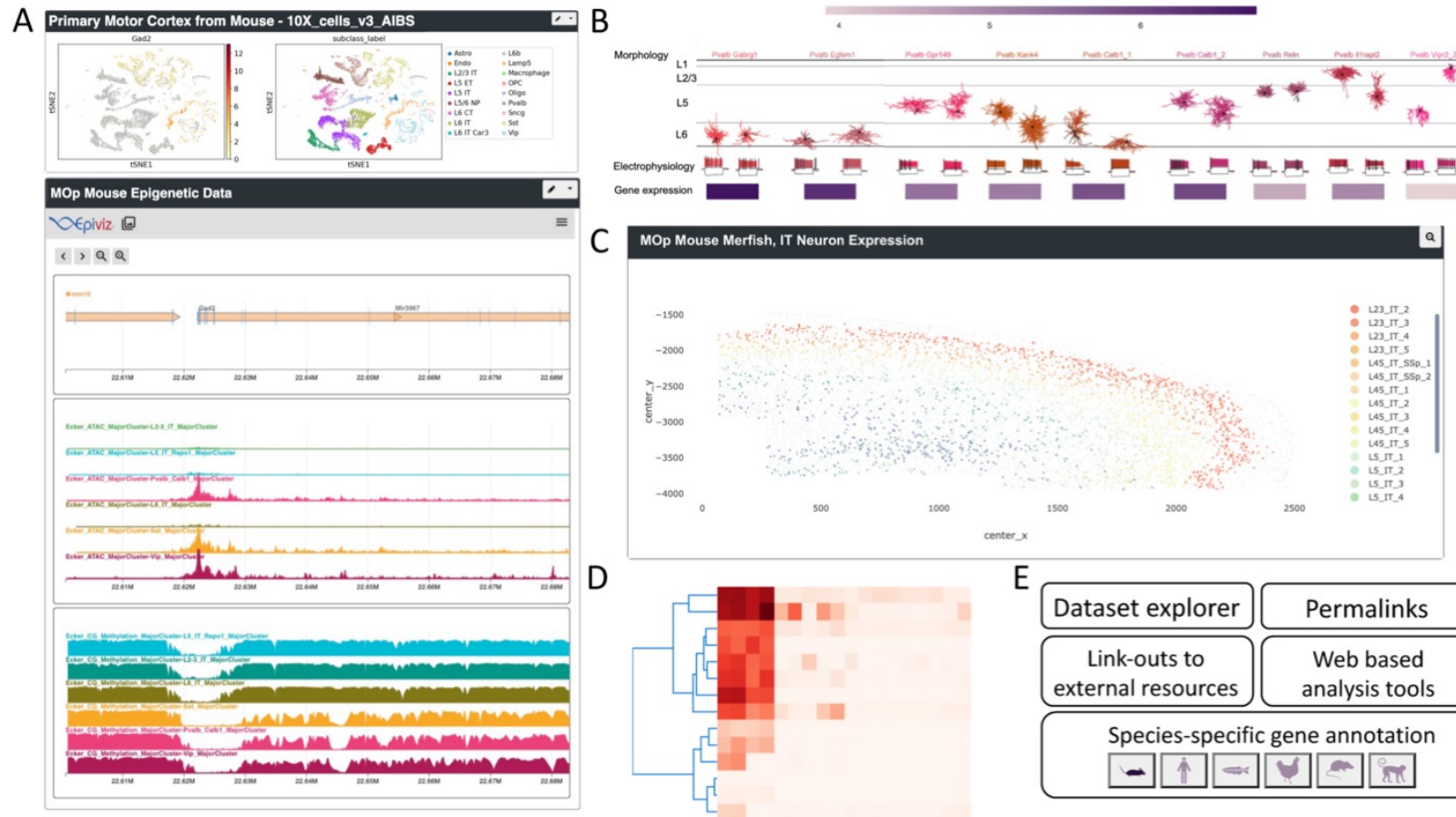
# Cell Type Knowledge Explorer



Raymond Sanchez

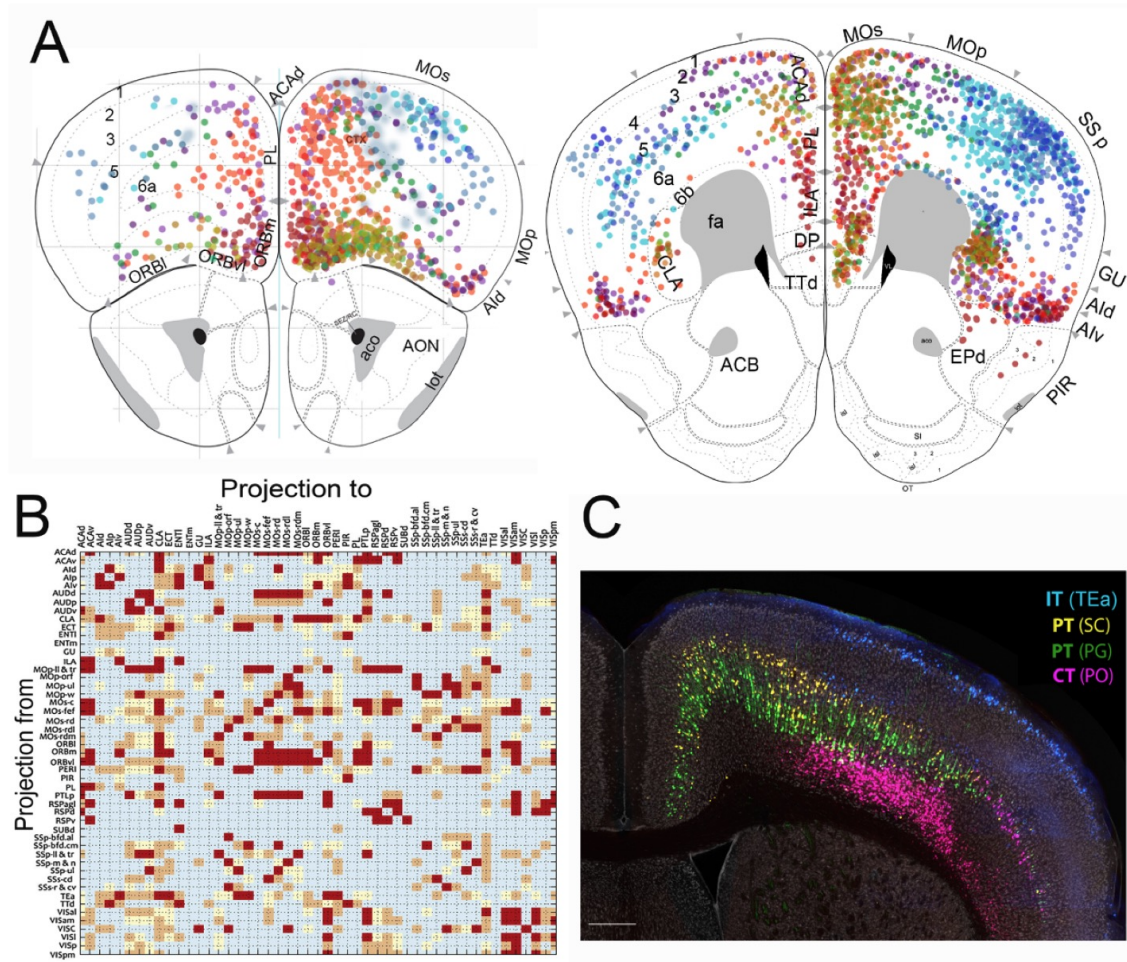


# NeMO Analytics



Ronna Hertzano

# Mouse Connectome Project



Hongwei Dong

# BICCN and FAIR Neuroscience

## BICCN and FAIR Neuroscience

### 1. Use of persistent identifiers and rich metadata to describe all data sets:

- BICCN Datasets receive a DOI or an equivalent persistent identifier from the archives,
- The BCDC and archives coordinate on standard metadata to accompany all datasets
- Archives implement dataset landing pages for machine-readable rich metadata about the datasets and access.

### 2. Providing detailed provenance

- BICCN datasets are versioned,
- Full citation metadata is supplied to support data citation,
- Investigators encouraged to link datasets to detailed experimental protocols deposited in at [BICCN group at Protocols.io](#)

### 3. Adherence to and definition of data standards

- Archives are implementing community data standards, including those developed through the US BRAIN Initiative.
- Archives have implemented common file format and metadata requirements for specific data types.
- Standards in use in BICCN are documented at [biccn.org](#).
- Several archives make use of standard identifier schemes for entities linked to the data such as [ORCIDS](#) for authors and [RRIDs](#) for organisms, antibodies, cell lines and tools.

### 4. Use of FAIR vocabulary

- BICCN has developed ontologies and controlled vocabularies to annotate data and map metadata such as the Brain Data Standards Ontology.
- Vocabularies are all maintained in GitHub repositories as described on the BICCN standards page

### 5. Providing a plurality of data attributes to aid in reuse

- Checklists for standard metadata for experimental types such as Patch-seq and for describing specimens
- Contact person identified to answer questions about the data, and code that can be used with the data.

### 6. The use of clear licenses and data use agreements

- All data that does not involve protected health information is made available under a CC-BY 4.0 attribution license.
- BICCN requires that those using the data follow formal citation principles for citing the data.
- Archives are making citations available per dataset to assist in proper citation.

Maryann Martone