Independent Student Research Guide: Cell Types + Cell Type Knowledge Explorer

Introduction

Goals

- This is a basic guide to conducting your own cell types research project using the Cell Type Knowledge Explorer.
- This guide expands on the online lab activities found in the "Cell Types Lesson" to support students doing research independently.
- You will read scientific papers and supporting resources to understand better the foundational knowledge surrounding cell types. This will help you develop the background knowledge you need to interpret data in this online tool and design a research question of your own.
- You will learn different approaches to scientific research (explore, compare, investigate) and mimic a chosen approach like a scientist does.
- You will be given sample research questions to consider to help you get started with your own research project.
- *This guide assumes you have gone through the "Cell Types Lecture" via your class or your own self-study. The "Cell Types Lecture" can be found here.*

Cell Types Background

As shown in "Cell Types Lesson," using online databases allows scientists to **explore**, **compare**, and **investigate** data, which may be either the main focus of their research or help inform other projects they will do later. Scientists constantly analyze new data, and in turn they will never stop learning to help understand that data. Having open scientific tools & data allows anyone to ask their own research questions and be a scientist.

Background literature and resources

- Cell Types Support: Comprehensive list of supporting material to help better understand the science behind cell types and how to use Allen Institute's scientific tools. The "What is" section contains recommended introductory resources such as scientific papers. Recommended to read the review paper *"What is cell type and how to define it?"*. In addition, there is "How to use Cell Type Knowledge Explorer" section that will help you navigate the Cell Type Knowledge Explorer.
- Scientific American: A Cell Atlas Reveals the Biodiversity Inside Our Head

- Cell Types YouTube Playlist: Playlist of all YouTube videos that Allen Institute has created over the years with topics regarding cell types.
- News Article: Unraveling the complexity of the mammalian brain

Selecting a cell type

Pick a subclass that interests you, either one listed below or one found at the subclass* level on Cell Type Knowledge Explore. Choose to either **explore, compare and/or investigate** by following the guide below**.** If there is no subclass that particularly interests
**we recommend subclass level because as of 2024, it will be difficult to find other studies that define cell types at the single-cell transcriptomic level as it is a newer technology that not many other labs have adopted*

Here are some example subclasses and interesting findings that would make scientists interested in learning more about that subclass.

- **Parvalbumin (Pvalb):** Parvalbumin neurons are the largest subclass of cortical GABAergic neurons, at estimates from 30-50% (Apicella 2022). Studies indicate that schizophrenia patients have a lower density of Pvalb neurons (Kaar, S. J., et al. (2019) Pre-frontal parvalbumin interneurons in schizophrenia: a meta-analysis of post-mortem studies).
- **Somatostatin (Sst):** Somatostatin neurons are the second largest subclass of cortical GABAergic neurons, with estimates Sst is 30% of total cortical GABAergic neurons (Apicella 20220). Studies indicate that in Alzheimer's Disease, Sst neurons show one of the most rapid declines in early progression ([Pre-print] Gabitto, M. I., Travaglini, K. J., et. al. (2023). Integrated multimodal cell atlas of Alzheimer's disease).
- **Vasoactive intestinal polypeptide (Vip):** Vip neurons historically were a subgroup of ionotropic serotonin receptor (5HT3aR) neurons, but transcriptomic studies have separated 5HT3aR to two groups, Vip and Sncg (Tasic 2018). Studies indicate that Vip neurons target other inhibitory neurons, such as Sst, to help regulate their action **(**Millman, D. J., et al. (2020). VIP interneurons in mouse primary visual cortex selectively enhance responses to weak but specific stimuli).

Explore

What is explore?

When you are exploring, you don't have a specific research question in mind yet, you're just looking for something intriguing or further expanding your knowledge.

Why explore?

- You don't know what you don't know - a large part of science research is exploring literature and other resources to have a better understanding of the field.
- Online data visualization tools allow exploratory research that helps scientists investigate – with a goal in mind, or just to explore.

How to explore

Pick 1-4 cell types within your chosen subclass.

Explore the highest gene expression markers (largest and darkest circle) and/or gene markers (indicated in green with triangle next to it) connect back to NIH Gene Database and learn more about those genes.

Exploration questions to consider:

- Does that gene's function relate to that cell type's function?
- What are some genes that are consistent across cell types within the same subclass, and what genes are different?
- What are some functional differences between the genes that are listed as markers versus genes that are highly expressed?
- How do the gene markers and/or highly expressed genes change or stay the same in different species (mouse to human to marmoset)?

After you made your observations, how would you transform your exploration questions into a deeper research question? Example: You discovered gene x stayed the same across species. Was gene x consistent across species because of convergent evolution or parallel evolution?

Comparison

What is compare?

Examine how your results match or differentiate from other published results.

Why compare?

- Cell Type Knowledge Explorer and many other Allen Institute tools focus on creating healthy, control data. This allows scientists who study areas like disease, focus on their models of interest and not spend much effort on creating standard, healthy control.
- In addition, if scientists are also studying healthy models, by accessing Allen Institute's vast healthy dataset, they can compare their data with Allen Institute's data to ensure it is aligning as expected.
- Single-cell transcriptomics is an expensive and newer technique, and not every lab can define their morphology or electrophysiology data at the transcriptomic

level. Having free multimodal data allows scientists to compare their data from one modality and predict what the other corresponding modalities might be.

How to compare

Pick either one of the research focus & accompanying papers listed below that match your chosen cell type, or use resources like PubMed or Google Scholar to find other papers. See how the paper's data compares to the data listed in the Cell Type Knowledge Explorer. Be sure to note any variable differences such as species, brain region, if the animal is healthy or diseased, etc.

Research focus: Morphology

Subclass: Pvalb

"Distinct Physiological Maturation of Parvalbumin-Positive Neuron Subtypes in Mouse Prefrontal Cortex" Miyamae (2017). DOI: 10.1523/JNEUROSCI.3325-16.2017

Subclass: Sst

"Distinct Subtypes of Somatostatin-Containing Neocortical Interneurons Revealed in Transgenic Mice" Ma (2006). DOI: 10.1523/JNEUROSCI.0661-06.2006

"Diversity and Connectivity of Layer 5 Somatostatin-Expressing Interneurons in the Mouse Barrel Cortex" Nigro (2018). DOI: 10.1523/JNEUROSCI.2415-17.2017

Subclass: Vip

"Characterizing VIP Neurons in the Barrel Cortex of VIPcre/tdTomato Mice Reveals Layer-Specific Differences" Prönneke (2015). DOI: 10.1093/cercor/bhv202

"Morphological and Functional Characterization of Non-fast-Spiking GABAergic Interneurons in Layer 4 Microcircuitry of Rat Barrel Cortex" Emmenegger (2018). DOI: 10.1093/cercor/bhx352

Research focus: Electrophysiology

Subclass: Pvalb

"Effects of optogenetic stimulation of basal forebrain parvalbumin neurons on Alzheimer's disease pathology." Wilson (2020). DOI: 10.1038/s41598-020-72421-9

Subclass: Sst

"Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex." Gentet (2012). DOI: 10.1038/nn.3051

Subclass: Vip

"Electrophysiological evidence that the vasoactive intestinal peptide receptor antagonist VIP6–28 reduces nociception in an animal model of osteoarthritis" Schuelert (2006). DOI: 10.1016/j.joca.2006.04.016.

Research questions to consider:

- How do the morphology or electrophysiology features in your paper compare to the ones listed in Cell Type Knowledge Explorer?
- What could account for the differences?
- If your paper is studying a disease model, what do the differences say about how the disease affects electrophysiology or morphology?
- If no transcriptomic data is listed in the paper, are you able to guess the cell type at the transcriptomic level by finding the closest matching features?

Investigate

What is investigate?

Scientists can examine and review raw data to interpret and reuse as they see fit.

Why investigate?

- Having accessible, open data allows scientists to further investigate data
- Public tools that contain open data allow scientists to examine the data themselves so scientists can reuse that data and further investigate the raw data themselves

How to investigate

Below is quote from research paper stating that Sst Chodl has been identified as a separate subclass from the main Sst subclass. This was due to Sst Chodl having distinct transcriptomic mapping profile, and multiple research studies finding different electrophysiology features. You can use the Cell Type Knowledge Explorer to investigate if you see the same electrophysical differences following the instructions below.

"These 'Sst Chodl' neurons are rare and, based on expression of specific marker genes, correspond to the only known cortical interneurons with long-range projections. Recent studies using the multimodal cell phenotyping method Patch-seq confirmed that 'Sst Chodl' cell sets characterized based on morphology and electrophysiology match those defined by transcriptomic profiles." (Miller 2020) https://doi.org/10.7554/eLife.59928

1. Use the metadata file to find the file name of the data of interest
    a. Go to Github link of metadata file and click on the download icon to download the file (.csv). *CSV file is similar to Excel file* https://github.com/AllenInstitute/CTKE_viz/blob/main/ephys/Tolias_m1_patchseq_meta_data.csv
    b. Find your cell type of interest (in this example, Sst vs Sst Chodl) by sorting RNA type in column T.
    c. Write down the mouse ID (column F) and sample number (column E). You will need to these identification names to know what raw data to look for.

2. Open the raw data on DANDI*
    a. Go back to Cell Type Knowledge Explorer and your cell type of interest. By “Electrophysiology”, go to “Download Data” – this will open a separate page called DANDI where the data is stored
    b. Click on “Files” on the right side to view all the electrophysiology data. Refer to your mouse name (column F) written down
    c. Click on the mouse ID. For the sample number (column E) written down, go to “Open With” and select “MetaCell/NWBExplorer”. *This may take a while to load*
    d. Click on the eyeball to view different recordings

*DANDI: Distributed Archives for Neurophysiology Data Integration is a data repository supported by the BRAIN Initiative to upload neurophysiology data, such as electrophysiology data used in Cell Type Knowledge Explorer. To learn more about DANDI visit: https://www.dandiarchive.org/

Research questions specific to the example above:

- What type of firing patterns are shown in Sst vs Sst Chodl (adapting, burst, irregular, fast spiking)?
- What marker genes found in Sst Chodl correspond to only cortical interneurons with long-range projections? What is that expression in Sst?
- Which Sst cell type most closely resembles Sst Chodl?
- Do you agree with the claim made in Miller 2020 that Sst Chodl is a distinct subclass from Sst?

Try to find another research paper about your cell type of interest, and see if you can find the raw data that supports their findings.

Questions to think about regarding open data:

- Where was their raw data hosted?
- How accessible was their raw data?
- Were you able to find and understand the metadata?
- Do you agree with their findings once examining the raw data yourself?

To learn more about open research and best practices, we recommend starting by reading the FAIR principles.

Cell Type Knowledge Explorer Citations

Patch-Seq (Mouse): Scala, F., et al. (2021). Phenotypic variation of transcriptomic cell types in mouse motor cortex. Nature. DOI: 0.1038/s41586-020-2907-3
Cross-Species: Bakken, T. E., Jorstad, N. L., Hu, Q., et al. (2021). Comparative cellular analysis of motor cortex in human, marmoset and mouse. DOI 10.1038/s41586-021-03465-8
Transcriptomics & Epigenomics: Yao, Z. et al. (2021) A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. Nature, 598(7879):103-110. DOI 0.1038/s41586-021-03500-8