Annotation Comparison Explorer (ACE)


ACE is a web (and associated R shiny) application for comparison of different annotations of the same dataset. ACE was designed (but is not limited to) use cases surrounding comparison of cell type definitions and other cell metadata. Examples include (i) cell type assignments (e.g., from different mapping/clustering algorithms or taxonomies), (ii) donor metadata (e.g., donor, sex, age), and (iii) cell metadata (e.g., anatomic location, QC metrics). Additionally, this tool can compare results across more than two taxonomies, for example for annotating a novel taxonomy with information from multiple existing taxonomies or for comparison of data from multiple studies of Alzheimer's disease.

Most folks will want to access ACE as a website at https://sea-ad.shinyapps.io/ACEapp/.

ACE can also be run in RStudio as a shiny app by following the instructions shown at the GitHub repository https://github.com/AllenInstitute/ACE. The codebase is open-source and we encourage contributions.

Note that ACE is under active development and may not match the content of the User Guide exactly. If you run into issues or more generally have any thoughts or opinions about the app, or you would like to contribute to this app, please reach out via the big button on the app or by submitting a GitHub issue.
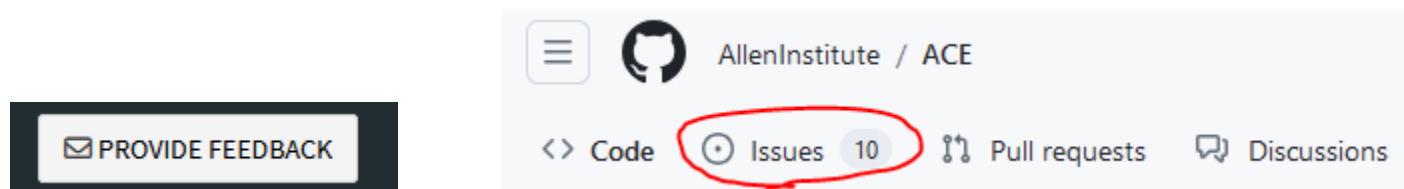
App developed by Jeremy Miller with support from Aaron Oster and Bosiljka Tasic, using some original code developed by Lucas Graybuck. Included annotation tables were created by Jeremy Miller, Kyle Travaglini, Tain Luquez, Rachel Hostetler, and Vilas Menon. Logo credit: Lauren Alfiler.

Use Cases

ACE was created to address a number of use cases not easily addressed using interactive tools.

- **Compare cell type results across published studies of Alzheimer's disease (AD)**: We have assigned cell types from the Seattle Alzheimer's Disease Brain Cell Atlas (SEA-AD) classification to cells from 10 studies of AD and have recorded changes in abundance with AD reported in each study. ACE allows users to interactively explore these relationships and uncover which cell types show common AD-associated changes across studies. (Check out Supertype Sst_25!)
- **Convert cell type names between different Allen Institute taxonomies**: Over the past decade the Allen Institute has created multiple cell type classifications in brain, including recent studies in whole mouse and whole human; however, understanding what a specific cell type defined in mouse visual cortex or human middle temporal gyrus in 2018 is now called is not straightforward. ACE includes a variety of translation tables providing easy conversion between these different taxonomies using matched data sets.
- **Compare mapping and clustering results for user-provided data**: MapMyCells provides an easy way for users to assign Allen Institute cell types to user-provided single-cell or spatial transcriptomics data. ACE allows researchers to see how results from their own clustering analysis relate to assigned types for a given set of cells. A separate tutorial for how these two tools can be used in parallel is under development.
- **Identifying which cell types are in which brain regions:** Recent publications of cell type classifications in whole mouse and whole human brain have found that many of the several thousand cell types show spatial localization. This extends earlier studies of multiple cortical areas, which found cell types specific to visual cortex and hippocampus, among other brain regions. ACE allows researchers who are focused on a particular brain region to easily filter cells by brain region (or other metadata) and determine which cell types are still present after filtering, or alternatively, to determine the proportion of cells of a given type that are found in different brain structures.
- **Assess data quality by viewing and comparing QC metrics per cluster:** Standard QC metrics, like # of genes detected, % mitochondrial reads, doublet scores, and lab-based pass/fail calls can be extremely useful in determining whether or not clusters defined in a cell type analysis represent "real" biological cell types/states. ACE allows you to plot any numeric values in the context of any categorical variable, so you can quickly scan whether certain clusters have higher or lower QC metrics than most others.
- **Visualize spatial location or morphoelectric features of cells of a given type**: A number of techniques allow a researcher to determine cell type of cells in other contexts. For example, MERFISH provides the specific spatial location of a cell with known type, while Patch-seq allows one to assign a transcriptomic type in a cell with measured morphology and electrophysiology. ACE includes tools for plotting any numeric values in the context of any categorical variable, so you can quickly scan whether certain clusters have higher or lower morphoelectric values. In addition, ACE can compare and color code any numeric values, allowing one to visualize where cells of a given type are located in physical space or in a plot of any two features of interest.
- **Comparing the abundance of cells in different situations (e.g., by sex, disease state, hemisphere, or brain region)**: Several glial cell states are known to appear in response to neurological insults, while

structural differences in left and right hemisphere may or may not translate to cell type differences. ACE provides a few tools for visualizing such abundance differences through confusion matrices and river plots, or direct inspection of individual metadata fields of cell types.

- **Visualizing metadata on a UMAP (or other 2D latent space)**: UMAPs are popular visualizations for exploring gene expression data. While other tools like CZI CellXGene are better suited for this task, if 2D coordinates are provided, ACE can color-code any subset of the cells by any provided metadata, allowing one to see where in the UMAP cells from a given cell type are located, for example.

- **Any other metadata comparison**: While ACE was created with cell type analysis in mind, it is ultimately just a tool for visualizing different properties of the same entities, meaning that it's uses are only limited by your data and your imagination. Other potential questions ACE could address include:
  - Do more students with red or blond hair also wear glasses?
  - What is the average age of workers in different professions?
  - Physically where in the USA are Republican vs. Democrat-leaning cities located?
  - What is the overlap between the Gyral and modified Brodmann anatomic atlases in human brain?

If you find ACE useful, **please let us know how you use this tool**! Also, if you are having trouble getting ACE working for your use case, please let us know that as well.

☑ PROVIDE FEEDBACK

How does ACE work?

ACE is an annotation comparison tool. Its input is a <u>single set of data</u> (typically, but not necessarily, cells) with <u>different annotations</u> for those cells (typically, but not necessarily, cell type assignments and/or other cell metadata). The application itself then visualizes these annotation relationships in a variety of ways.

As a simple example, let's assume that we have 11 adults participating in a medical study. They are first asked to measure and self-report their height ("Annotation 1" below), and all but the tallest two individuals comply. Researchers use these data to split these data into short vs. tall. Then a bit later, all 11 adults come into the office and get their heights measured ("Annotation 2"). With the remaining two individuals included, researchers now decide that there should actually be three categories of height: "Short", "Average" and "Tall".
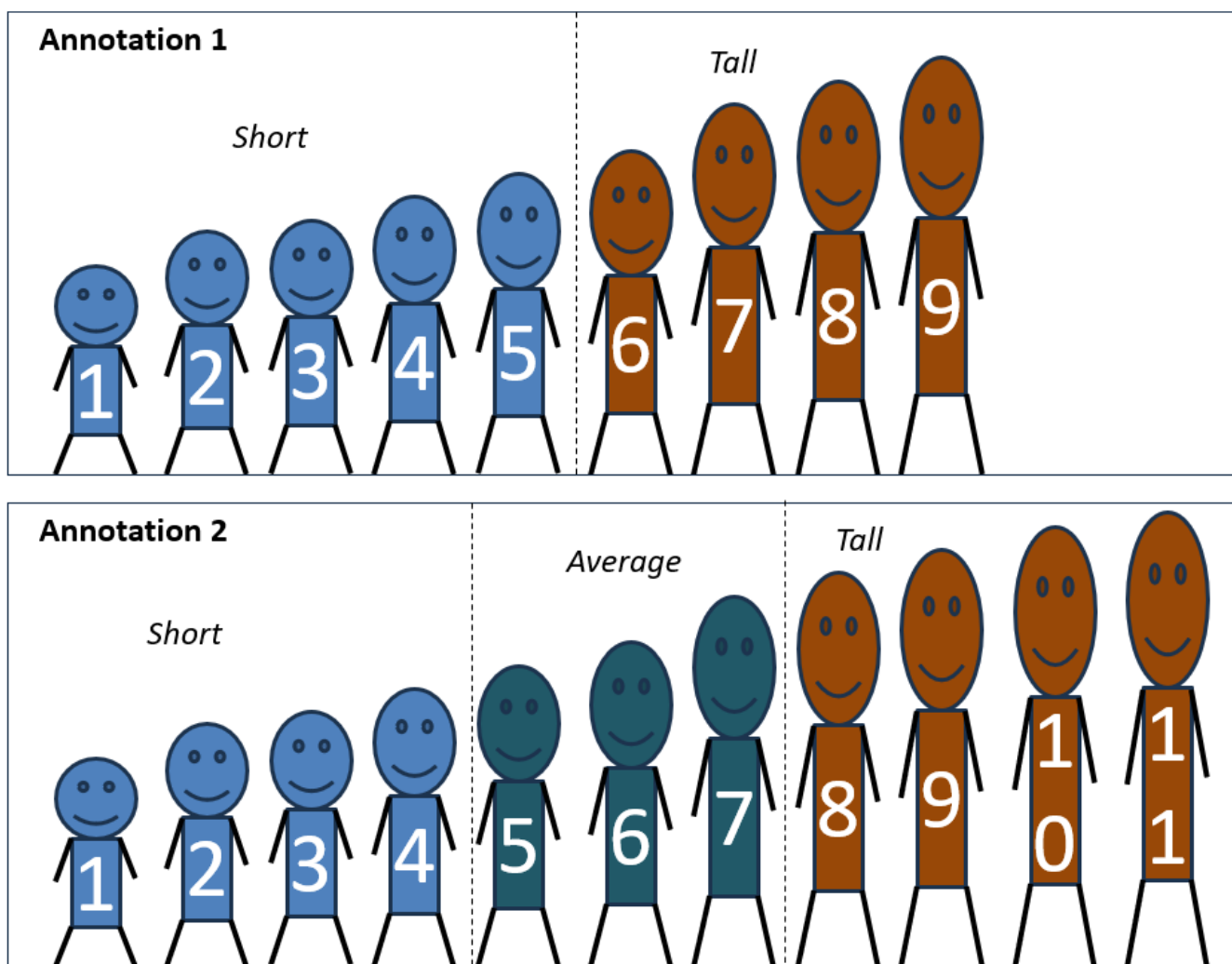
This example is meant to represent the case when we have a single set of cells that initially had one cell type annotation and later had another cell type annotation, potentially with the addition of more data. This is also a perfect example of data set that could be displayed in ACE. **But how?**

First, we need to prepare the input files. ACE takes as input two files. The first input is the cell (or in this case, person) annotation file:

Person #	Annotation 1	Annotation 2
1	Short.1	Short.2
2	Short.1	Short.2
3	Short.1	Short.2
4	Short.1	Short.2
5	Short.1	Average.2
6	Tall.1	Average.2
7	Tall.1	Average.2
8	Tall.1	Tall.2
9	Tall.1	Tall.2
10		Tall.2
11		Tall.2

Each row represents a single data point (person), and each column represents a different annotation. This is essentially a table version of the graphic from above. Because ACE relies on using the same data set, for any comparisons between Annotation 1 and Annotation 2, only the first nine individuals would be used (Person 10 and Person 11 did not measure their own height in Annotation 1 and therefore are excluded from comparisons).

The second (optional) input file is metadata information table, which provides additional information about each annotation. In this case there are five values used to describe each individual's heights across the two annotation columns:

cell_type	Alias	Height range	Annotation overview	Date
Short.1	Shorter adult	<1.85M tall	Self assessed height	7/23/2024
Tall.1	Taller adult	>=1.85M tall	Self assessed height	7/23/2024
Short.2	Short adult	<1.75M tall	Height at medical exam	7/29/2024
Average.2	Average adult	1.75M - 1.95M tall	Height at medical exam	7/29/2024
Tall.2	Tall adult	>1.95M tall	Height at medical exam	7/29/2024

These values can be whatever you want (although currently the column with the metadata values needs to be called "cell_type") and are only used in one of the visualizations (details below). Essentially, this allows you to easily review other knowledge of a single metadata value in context, which is helpful where there are thousands of values in total.

Once these data are input, ACE provides as number of ways to visualize the comparisons between two (or more) annotations. For example, ACE can make river plots (Sankey diagrams) showing how the annotations values of specific individuals change between two annotation fields. For our example above, it would look like this:
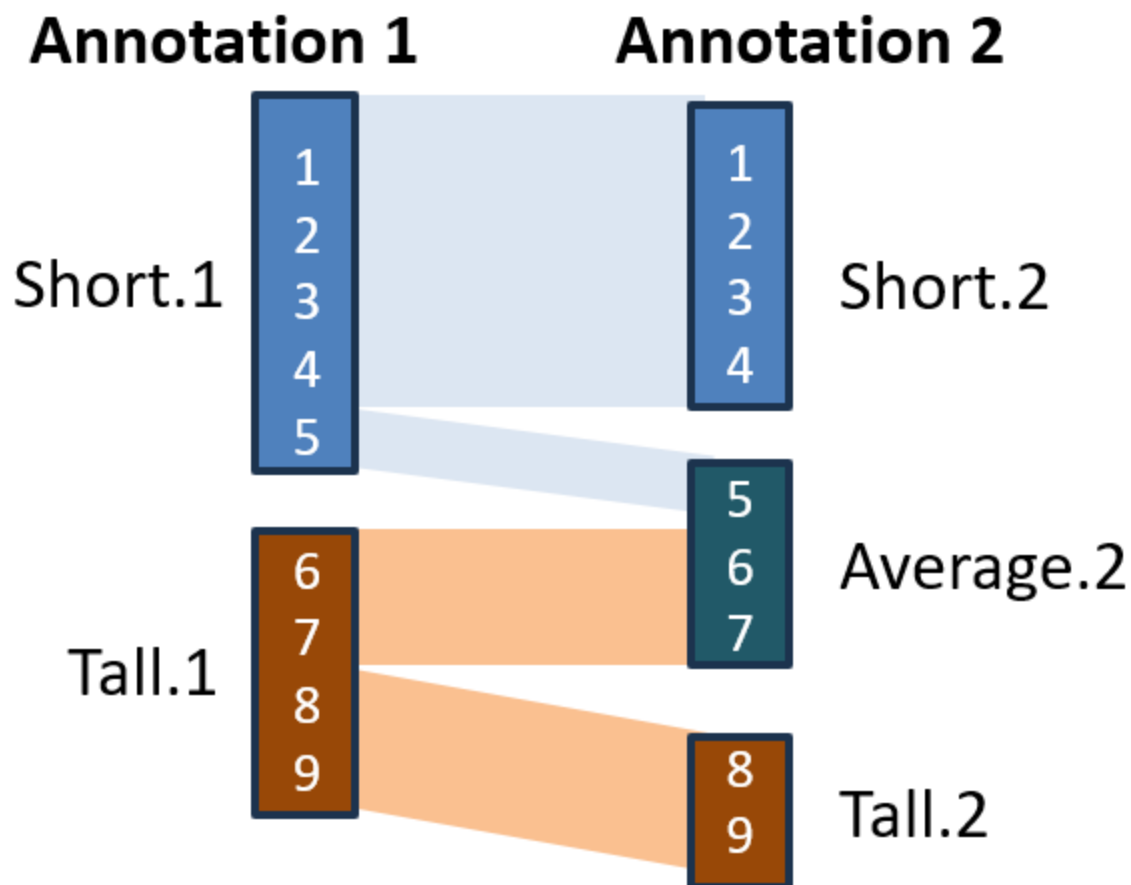
In this case, individuals 1-4 are always considered short, 8-9 are always considered tall, and 5-7 change their annotation in a way that can be visualized using the "rivers" between the two columns of colored rectangles. (Individuals 10-11 are considered tall in the second annotation but were not included in the first annotation and therefore are not in the plot.) If we look at the metadata table above, we can see that the reason these individuals change height categories is not because they change height, but rather because the definition of what we mean by short or tall has changed between the two annotations (e.g., 'short' changed from <1.85M to <1.75M). This is analogous to how in a cell type annotation, the underlying gene expression data for a given cell will remain the same, but the way in which cell types are defined can change. Unlike in this example, the reason that a cell changes from one annotation group to another is often complex, but also does not impact the visualizations.

ACE provides additional visualization options that are described in detail below, but hopefully this overview provides a general sense for what ACE is doing and how it works. The next section provides many additional use cases of ACE, and the remaining sections go into detail about how the tool itself is used.

Getting started

To use ACE, simply [go to the website](#) or launch the tool in R (to run locally, see the instructions [on GitHub](#)).

Once open, ACE is an interactive tool that consists of the four panes shown above:

1) **Informational pane**: The first pane contains general information about ACE, including what it is, links out to this guide and other resources to help you use the app, and acknowledgements of app developers. During use, this pane can be minimized by clicking on the three lines at the top of the screen.

2) **Select data set**: This pane allows you to choose a preset annotation table to compare or to point to your own table and is the starting point for all other ACE functionality. Preset annotations fall into a few categories:
   a. **Disease studies**: Currently just one table linking cell type assignments between SEA-AD and 10 published community studies of AD.
   b. **Human cell type taxonomies**: Multiple tables translating cell types between different cell type taxonomies focused on middle temporal gyrus (MTG). Additional tables connecting to whole human brain and focused on other brain structures to be developed soon.
   c. **Mouse cell type taxonomies**: Multiple tables translating cell types between different cell type taxonomies in visual cortex, additional cortical areas, and whole mouse brain, along with another table comparing brain cell types and anatomic structures.

   Please read the "Dataset description" that appears when selecting a comparison table to learn more.

3) **Filter cells in dataset:** After selecting a comparison table, this pane allows you to choose which rows (e.g., cells) in the comparison table to include in the visualizations. This step is critical when viewing tables with many rows. For example, one could choose to only look microglial cells, or to focus on cell types in hippocampus. There is also an 'invert' feature allowing you to select all cells except a given selection.

4) **Visualizations and statistics**: Once the annotation table and relevant rows are chosen, all of the functionality of ACE takes place here. The tabs in this pane provide different tools for comparing cell-level annotations and (optionally) providing additional information about each annotation. Most of this user guide is dedicated to these tools, but a brief overview of them can be found in the "Intro" tab, reproduced here:
   a. **Compare pairs of annotations**: This panel shows different plots depending on the data type. When comparing categorical variables (most cases), a confusion matrix is shown where points can be sized or colored in a variety of ways, including based on Jaccard distance (default). When comparing numeric vs. categorical values a box-and-whiskers + dot plot is shown. An example of this would be in viewing mapping probabilities for each cell type. Comparisons of pairs of numeric annotations is done in a separate panel.
   b. **Link 2+ annotations (river plots)**: This panel shows a river plot (also called a 'Sankey plot' or 'Sankey diagram') which chains together relationships between two or more categorical variables. The order variables are listed matters, as the overlap between each pair of adjacent annotations are shown.
   c. **Explore individual annotations**: This tab focuses on exploring individual annotations (e.g., the "subclass" called "SST"). It matched plots and graphs showing all values of different annotations for a given input annotation, as well as set of additional metadata for any annotation value selected from the table. This is the only place that the "Location of metadata (e.g., cluster) information" file is used.
   d. **Compare numeric annotations**: This tab focuses on comparing pairs of numeric annotations. Two common use cases are (1) visualizing 2D UMAP coordinates and (2) showing spatial positions of cells in a MERFISH tissue section. Cells are visualized using an interactive scatter plot with hover capabilities.

   Note that not all functionalities will be available for every annotation table. Depending on the types of data provided in the input tool, certain tabs may not work or may be omitted.

Input data for ACE

At its core, ACE is a method for comparing different cell-level annotations and (optionally) providing additional information about each annotation. Some comparison tables are provided, or users can provide their own.

The input to ACE one or two comparison tables:

(1) Cell data table

The primary input is a **required** table where rows correspond to individual items (typically cells or nuclei) and columns correspond to metadata such as cluster assignments, mapping results, different levels of the taxonomy hierarchy, donor information, or cell QC metrics:

	A	B	C	D	E	F	G	H
1	sample_name	SEAAD_supertype	subclass	class	GA_cluster	CA_cluster	sex	donor_name
2	AAACCCACAACTC	Pax6_1	Pax6	GABAergic	Pax6_1	Pax6_1	M	H18.30.002
3	AAACCCACACGGT	L5/6 NP_1	L5/6 NP	Glutamatergic	L5/6 NP_1	L5/6 NP_3	M	H18.30.002
4	AAACCCACACTCT	L5 IT_7	L5 IT	Glutamatergic	L5 IT_7	L5 IT_1	M	H18.30.002
5	AAACCCACATCAG	L6 CT_2	L6 CT	Glutamatergic	L6 CT_2	L6 CT_1	M	H18.30.002
6	AAACCCAGTGTCG	L4 IT_2	L4 IT	Glutamatergic	L4 IT_2	L4 IT_2	M	H18.30.002
7	AAACGAAAGCAC	Astro_1	Astrocyte	Glia	Astro_1	Astro_4	M	H18.30.002
8	AAACGAACACAA	L5 IT_2	L5 IT	Glutamatergic	L5 IT_2	L5 IT_1	M	H18.30.002
9	AAACGAACACGCT	L2/3 IT_5	L2/3 IT	Glutamatergic	L2/3 IT_5	L2/3 IT_3	M	H18.30.002
10	AAACGAAGTACCC	L5 IT_2	L5 IT	Glutamatergic	L5 IT_2	L5 IT_1	M	H18.30.002
11	AAACGAAGTTCCG	L5 IT_2	L5 IT	Glutamatergic	L5 IT_2	L5 IT_1	M	H18.30.002
12	AAACGAAGTTGCG	Vip_9	Vip	GABAergic	Vip_8	Vip_8	M	H18.30.002

User-provided files can either be located on the web (definitely on GitHub, and probably on other URLs but I haven't checked) or on the local machine. **Note that access to local machines is only possible using the R-Shiny version and NOT on the web version—an upload button to bypass this limitation is planned**.

Multiple file types are accepted as input:
- A **csv file**, like the one shown above.
- A **gzipped csv file** (e.g., XXXXX.csv.gz)
- An **h5ad file** in scrattch.taxonomy format (the obs field is read)—this would likely work for other h5ad files (e.g., input files for CZI CellXGene), but I have not tested it
- (For Allen Institute employees only: A directory containing an anno.feather file for visualization of taxonomies on molgen-shiny)

(2) Metadata table

The second file is optional, but if provided must be a csv file with information about each individual piece of metadata. Specifically, each row corresponds to an entry in one of the columns from the cell table (e.g., a specific cluster or subclass) and columns correspond to whatever information about the metadata that you want to share, and is only used in the "Explore individual annotations" tab. Here is one example that provides some information about cell sets from SEA-AD:

	cell_type	level	study	direction	direction_new	description	notes
1							
2	exc	class	SEA-AD	none	not_assessed	All excitatory neurons	
3	glia	class	SEA-AD	none	not_assessed	All non-neuronal cells (glial and non-neural types)	
4	inh	class	SEA-AD	none	not_assessed	All inhibitory neurons	
5	Astro	subclass	SEA-AD	up	up	Astrocytes	
6	Chandelier	subclass	SEA-AD	none	not_assessed	Chandelier MGE interneurons	
7	Endo	subclass	SEA-AD	none	not_assessed	Endothelial cells	
8	L2/3 IT	subclass	SEA-AD	down	down	Layer 2/3 intratelencephalic neurons	
9	L4 IT	subclass	SEA-AD	none	not_assessed	Layer 4 intratelencephalic neurons	
10	L5 ET	subclass	SEA-AD	none	not_assessed	Layer 5 extratelencephalic neurons	

User-provided files must be either csv or gzipped csv files and have the same restrictions for file locations as described above. There are minimal restrictions on what can be included in this table:

- **cell_type**: Currently the name of each metadata item **must** be shared in a column called "cell_type"; otherwise, this table is completely ignored. This constraint may be lifted soon.
- **direction** (optional): this field is currently the only way for coloring metadata on the "Explore individual annotations" tab (more details below). This tab looks for a column called "direction" which can be "down", "up", "none", or [anything else] and which determines color-coding. I'm open to a more flexible approach to color-coding cell metadata—if you have ideas, please provide feedback!
- **[any other columns]**: there are no restrictions on any other column names. Information about whatever other columns are included in this table will be reported

Efforts linking this tool with existing and in processes tools for cell type annotation are underway. This guide will be updated as integration progresses.

‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒‒
'Select data set' pane

This pane allows you to choose preset annotation tables to compare or to point to your own tables and is the starting point for all other ACE functionality.

This pane includes a few components:
1) **Select annotation category**: Choose from one of a few categories of pre-specified annotation categories (described above) or select "Enter your own location" to point to your own data and then skip to #4.
2) **Select comparison table**: Choose from one of the pre-specified annotation tables associated with this category. When you select a table a few things will happen immediately:
   a. File locations will populate in #3 and maybe #4. Do not change these file locations.
   b. The data will start to load. Depending on the table, this could take a couple seconds or up to a minute.
   c. Text will appear in the "Dataset description" section (#3). Read this text.
3) **Dataset description**: When a comparison table is selected, the text in this box will update to describe the content of the selected annotation table. Ideally this text would be sufficient to understand what the table represents, but if not, please provide feedback! Skip to #6.
4) **Input location of cell-level annotation information**: Enter the location of the primary table here. If an invalid file location is entered, you will see this error to the right of the bar: "ENTER VALID CELL ANNOTATION FILE."
5) **Location of metadata (e.g., cluster) information (optional; csv file):** Enter the location of the metadata table here. If you leave this blank or enter an invalid location, this table is ignored.
6) **Bookmark button:** <u>This is currently broken</u>! One you have been using the app and have things in a state that you would like to either share or return to at a later time, click the "Bookmark" button and a URL will appear below (see below). Copy this URL and when anyone pastes it into a web browser, it should return you to the current state of ACE. This applies to the web version of ACE only.
   Please reach out if you have any ideas about how to fix this Bookmark button! I've tried multiple approaches that all should work, but none of them do.
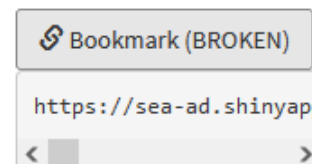
'Filter cells in a dataset' pane

Cell filtering is a critical component of ACE, as it allows you to define the context for all the visualizations and statistics. For example, since there are >5000 cell types in mouse whole brain, it is not practical to perform comparisons on all of them at once (and some visualizations may not work properly with that much data). What is practical, and much more useful, is to see how all the inhibitory cells collected from mouse visual cortex map between cell type taxonomy version, which dramatically decreases the size and scope of the data sets.

As another example (shown below), we can restrict our visualizations to only non-neuronal cells collected from a previous study of AD that successfully map to SEA-AD cell type. Here is how that filtering looks:

The "Choose Filter Set" box (#1) allows you to select one or more metadata columns on which to perform the filtering. This box (and all other boxes in the app where you can type) have autocomplete and will help you select possible options to select. For each chosen variable, the "Filter for" box (#2) allows you to choose how to filter the data. Categorical variables can be filtered by listing all of the values for that variable that you'd like to include in the visualizations for the app. For example, to retain all non-neuronal cells, you could either set the "Class" filtered to "Non-neuronal and Non-neural" (as done in this example) or you could set the "SEAAD_supertype" filter to include every non-neuronal cell type (or both). Categorical filters can also be "inverted", meaning that all cells except the ones filtered will be included. In this case, all of the cells that failed to map to SEA-AD (and therefore have a "Leng_2021_celltype" value of "noMappedCells_L21") are excluded. Numeric variables allow you to set a numeric range within which valid cells apply. For example, if there was a column corresponding to number of genes detected, you could set a minimum threshold of 500.

Finally, the bottom left corner of this box (#4) shows the filters that have been applied, whether cells from that filter and included or excluded, as well as the number of cells out of the total table that will be included for other app components.

Visualizations and statistics: Intro

Once the annotation table and relevant rows are chosen, all of the functionality of ACE takes place in the 'Visualizations and Statistics' pane. These tabs provide different tools for comparing cell-level annotations and (optionally) providing additional information about each annotation. The Intro pane provides a brief overview of the other visualization tabs, with more details described in the user guide sections below.

Each remaining tab within this pane corresponds to a different visualization that applies to the filtered data set defined above. **Note that not all functionalities will be available for every annotation table**. Depending on the types of data provided in the input tool, certain tabs may not work or may be omitted.

Finally, the Intro tab provides a calculation-free home page for this pane. Since some calculations are performed upon any state change in the app, we recommend adjusting the data set and filter set while this panel is showing, rather than while one of the main visualization tabs are shown.

Visualizations and statistics: Compare pairs of annotations

This tab allows comparison of any two pieces of categorical or numeric metadata. All of the tabs are laid out roughly the same way, with controls at the top, the main visualization in the middle, and download or other content on the bottom, details of the controls for this tab are shown a bit below.

COMPARING TWO CATEGORICAL VALUES

An example of the 'Compare pairs of annotations' tab is shown below, with default parameters for comparing a pair of categorical variables:
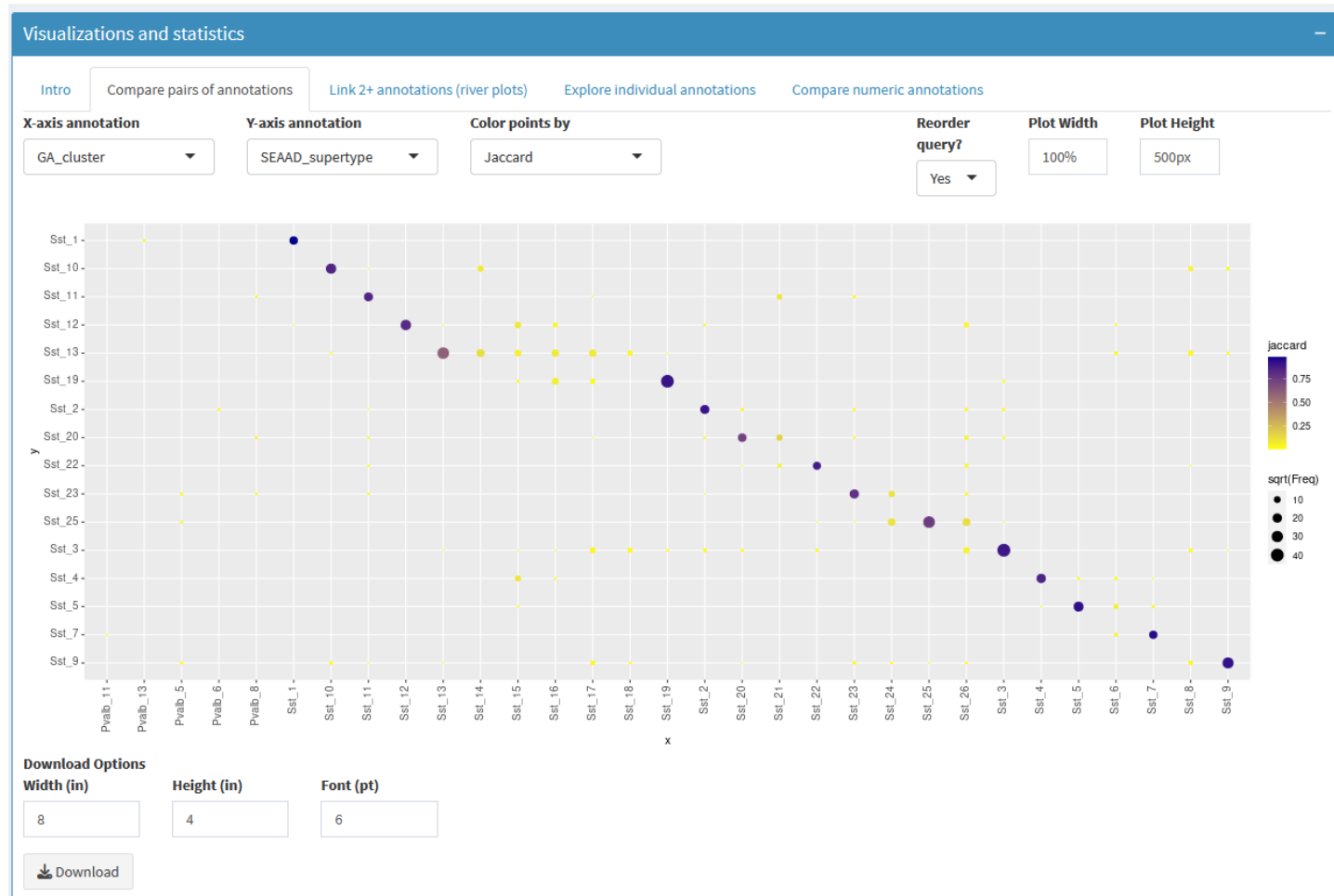
This example compares cell type assignments of neurons defined as SST in SEA-AD with their original cell type assignments in the great ape (GA) study. The confusion matrix shows a high correspondence between cell type definitions in these two studies, with most cells from SEA-AD supertypes (Y-axis) assigned to 1-2 clusters in the great ape study (X-axis). This result makes sense, since SEA-AD supertypes were defined using great ape assignments as a starting point. By default, Jaccard distance is shown, with blue values indicating a better correspondence and yellow values indicating a poorer correspondence. We can also change the plot to show the proportion of cells of each SEA-AD supertype assigned to each GA clusters (e.g., rows sum to 1):
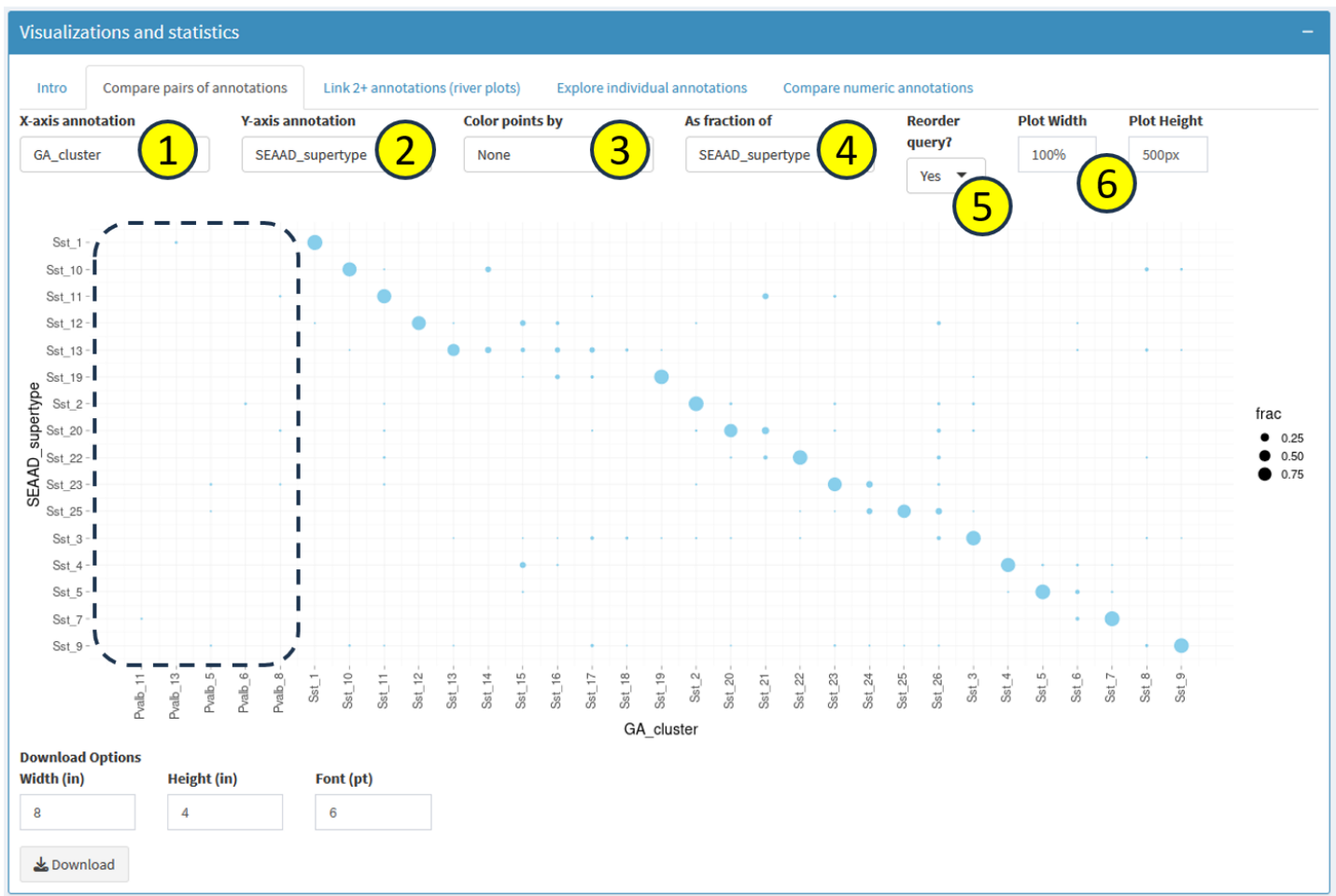
Again, we see that nearly all of the circles are either very big or tiny/missing, indicating good agreement in annotations. We also see a very small number of cells (tiny dots) that were defined as PVALB in GA but are not reassigned to SST types. Since the number of cells is small, and these subclasses are not perfectly distinct, such an overlap is nothing to be concerned with.

This panel includes the following controls, which can vary slightly depending on selections:

1) **X-axis annotation**: which annotation will be shown on the horizontal axis?
2) **Y- axis annotation**: which annotation will be shown on the vertical axis?
3) **Color points by**: Which metadata should points be colored by? For categorical visualizations the default value is Jaccard, which produces a plot like the one two figure above. In this case, color represents Jaccard similarity (bigger values and bluer colors represent higher similarity), while the size of the point scales by the square root of the number of cells in a given intersection. To change this scaling (using "As a fraction of" below), change the "Color points by" drop down to any of the other values.
4) **As a fraction of**: This drop-down is only shown if "Color points by" is not set to "Jaccard". Otherwise, there are three options: "None" (default), in which case the size of the points corresponds to the number of cells in that intersection. If changed to the metadata shown in y-axis or x-axis annotation, this converts the points to fractions where the values row- or column- sum to 100%.
5) **Reorder query?**: This parameter will attempt to sort the entries within the y-axis metadata to create as good looking of a diagonal as possible. This is extremely useful if the metadata are not sorted in the same way (or at all) and is set to 'Yes' by default. If set to 'No', the original order is retained.
6) **Plot Width / Plot Height**: Parameters that will change the width (in %) and height (in pixels) of the visualization on your screen. Note that you need to include the "%" sign in the Plot Width box.

Images can also be downloaded to your computer! See next page for more information.

COMPARING NUMERIC WITH CATEGORICAL VALUES

In addition to comparing categorical variables, ACE can compare categorical and numeric variables. In the case, a "swarm" plot is shown (#1 below), where each point represents a single cell, the width of the cells at a given vertical position showing the relative density, and an overlaid mean +/- interquartile range shown as red and black ticks, respectively.

In many cases these plots are useful for assessing data quality by reviewing the spread of a QC metric across cells of a given annotation (e.g., what fraction of SST cells have high mapping, do cells from a single donor have more mitochondrial reads that for others). In the plot below we use the example of mouse MERFISH data, showing the z-section from which all "01 IT-ET Glut" cells are located, divided by subclass and color-coded by supertype. We see that some subclasses tend to be more rostral (low values) than caudal (high values), while others have different supertypes spanning the brain (e.g., subclass 022 has teal-colored cells rostral and purple-color cells caudal).

Finally, the bottom of this tab (and several others) includes an option to download the current visualization (#2), along with a few parameters for the size of the images and included text. These defaults are not always reasonable, so it may be worth trying a few downloads to see which one looks best. The "Download" button defaults to using a generic file name and file location, so we recommend changing the names as appropriate.

Visualizations and statistics: Link 2+ annotations (river plot)

This panel creates an interactive river plot (also known as a "Sankey diagram") that allows you to compare two or more pieces of metadata (see below). In the Node Groups box (#1) you can enter any number of metadata fields in order (the order matters!), and then click "Go!" to generate the river plot. For each metadata field you get a stacked bar plot showing the number of cells that have each value within that metadata, with the value for each label shown (#2). In addition, for each adjacent pair of bar plots, you get "rivers" connecting each pair of metadata values together, where the thickness of each line represents the number of cells sharing the corresponding values from the two metadata fields (#3). Since this plot is interactive, you can hover over any bar or river and see what values and numbers correspond to what you are hovering over (#4). On the right side of the plot (#5) there are some extra controls that allows you to zoom, pan, and interact with the plot in various other ways. Finally, you can download the plot using the buttons on the bottom (not shown, but the same as describe above).
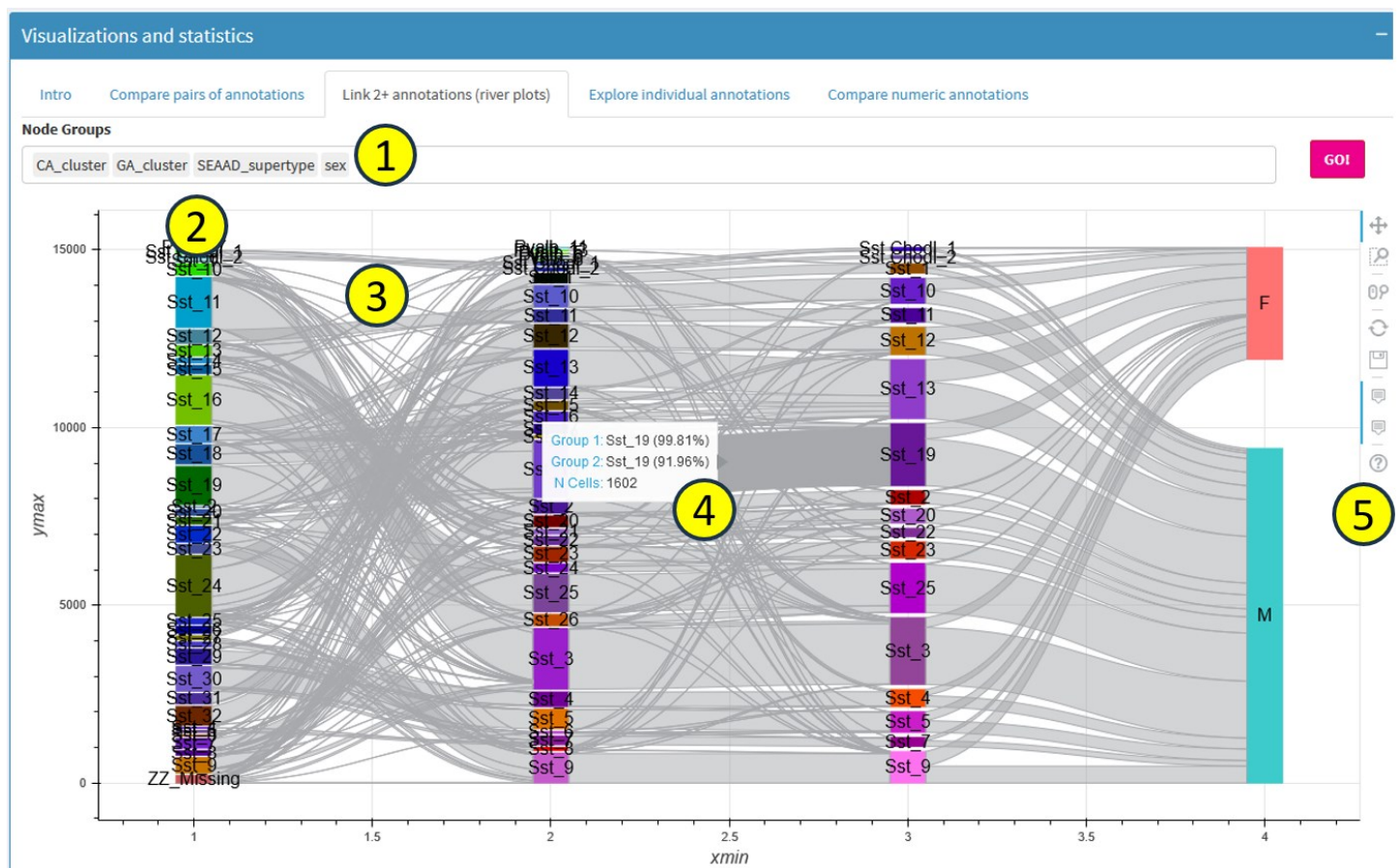
Currently there is no way to resize the box or to show the same metadata more than once in the Node Groups sequence, and this may or may not be updated later. Other planned improvements include adaptive reordering of clusters to match the "Compare pairs of annotations" tab.


Visualizations and statistics: Explore individual annotations

As the name suggests this tab is focused on exploring individual annotations (see below). *In essence, this tab shows basically the same information as river plots for a single metadata value, but without showing the rivers and with extra information about metadata fields optionally shown.* This can be useful for manual annotations of individual clusters (or supertypes, subclasses, etc.) or for understanding how cell types defined in one study compare with cell types defined in other studies. This is the only tab that uses the metadata information table. In this case we use the "SEA-AD: Alzheimer's cell type mapping" data set as an example.

Location of metadata (e.g., cluster) information (optional; csv file)

https://raw.githubusercontent.com/AllenInstitute/ACE/main/data/AD_study_cell_types_for_app.csv

Unlike the other tabs which compare all values for two or more pieces of metadata, this tab is anchored around a single "**Annotation value**" from a single "**Annotation group**" or metadata column (#1 below). Once this is chosen, the rest of the tab will be performed in comparison to that one annotation value. We next indicate which other "**Comparison groups**" (e.g., specific metadata columns) will be compared against the starting annotation value (#2 below). In this case, we choose mappings from the eight other AD studies that include neuronal types:
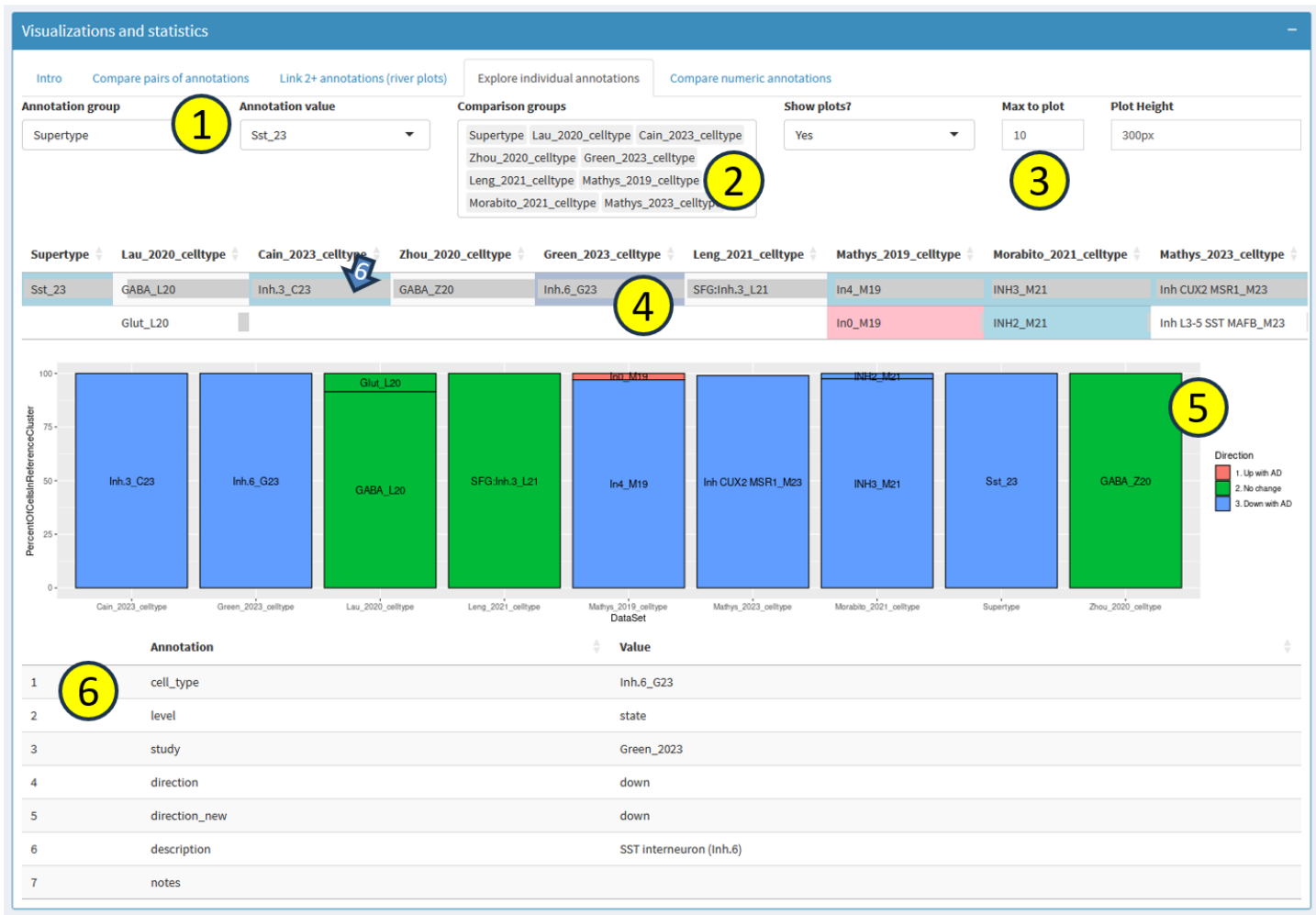
The other controls at the top describe whether/how comparison information will be displayed as a bar plot below the chart display (#3, more on that shortly). Specifically:

- **Show plots?:** A Yes/No call on whether the bar plot should be shown
- **Max to plot:** The maximum number of top overlaps to show in the bar plot (for example, if there are 20 SST clusters and "Annotation value" is set to SST subclass, then only the 10 must abundant SST clusters would be shown in the plot).
- **Plot Height**: Height of the plot in pixels.

The next section (#4) shows a chart of the selected annotation data along with the corresponding values from the comparison groups, sorted descending from highest to lowest overlap. Within each table entry, a grey bar indicates the fraction of cells from the selected annotation value that also have the comparison value listed. Boxes are also color-coded with whether they values are up (red), down (blue), or unchanged (white) with Alzheimer's disease (Note: I plan to make this a more generic coloring schema in the future). These calls rely on reading the "direction" column values in metadata information table. For both the chart and bar plot views, no color-coding is done if this column is omitted.

If shown, the bar plot (#5) displays the same information as the chart using a different graphic, with each bar indicating the fraction of cells from the selected annotation value that also have the comparison value listed, again colored by change in AD. Right now, there are some issues with coloring and ordering of bars that I'm still working out—in particular the color and order does not match between the chart and bar plot. If you have the time and knowledge to fix this, please reach out!

The final part of this tab is a table that displays additional provided information for a given metadata value (#6). If you click on any box in the chart (example above), then a table showing the provided metadata for that specific cluster will be shown below. I'm working on integrating this part with the information available in Taxonomy Development Tools, and more generally would appreciate any suggestions on how to make the information here (or the way to code it in) better.


Visualizations and statistics: Compare numeric annotations

This tab focuses on comparing pairs of numeric annotations. Cells are visualized using an interactive scatter plot with hover capabilities. Three common use cases are (1) visualizing 2D UMAP coordinates, (2) showing spatial positions of cells in a MERFISH tissue section, and (3) comparison of QC metrics or cellular features. It's worth noting that there are already a number of interactive tools dedicated to 2D scatter plot visualizations in the context of single cell data, including the Allen Brain Cell Atlas, CZI CellXGene, Cirrocumulus, and Cytosplore, which may be more appropriate than ACE for some use cases.

This tab is only activated for annotation tables with at least two pieces of numeric metadata. Otherwise, the tab either will not be present in the Visualization and statistics pane, or will return a warning in place of a plot here:
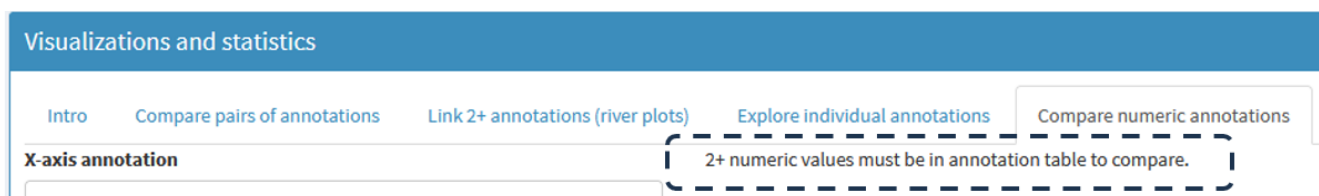
If two or more numeric metadata are available, a page like this will be displayed:
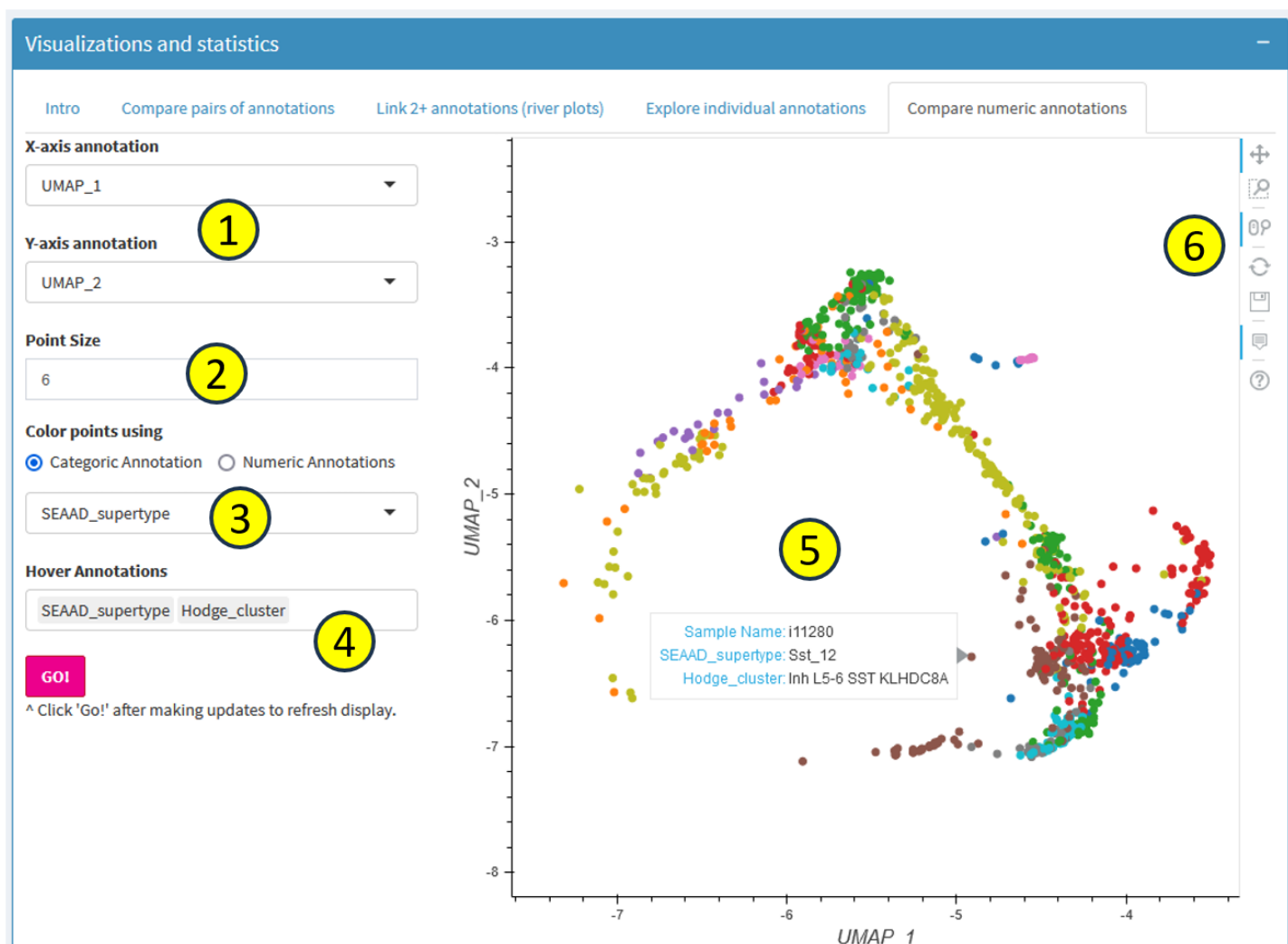
In the **X-axis annotation** and **Y-axis annotation** dropdowns (#1), you can select which annotation will be shown on the horizontal axis and vertical axes, respectively. For this pane only numeric annotations will be available to display. You can also select which **Point Size** to use (2), with the default (6) being reasonable in most cases.

The main use of this tool is for color-coding the points by metadata. You can color-code based on a single categoric annotation (#3 above) or based on up to three numeric annotations at once (alterative panel below).

Color points using

○ Categoric Annotation ⦿ Numeric Annotations

Numeric annotation 1 (Red)

| SEAAD_confidence | ▼ |

Numeric annotation 2 (Green)

| (none) | ▼ |

Numeric annotation 3 (Blue)

| (none) | ▼ |

Scaling

| Linear | ▼ |

Since this plot is interactive, you can hover over any point (#5) and see specified information about that cell (#4). On the right side of the plot (#6) there are some extra controls that allows you to zoom, pan, and interact with the plot in various other ways. Note that to save your image, you need to use the icon in control panel on the right (🖫). There is NOT a separate "Download" button.

Finally, note that the eventual goal of this tab is to compare two separate pieces of metadata for the same cells (e.g., color code cells that are only in hippocampus, that are only inhibitory neurons, that are neither, and that are both). Such functionality would be relatively unique from other 2D visualization options and would better align with ACE use cases.

Conclusions

Thank you for reading the user guide and for using ACE! Please reach out with any suggestions for how to improve the tool or associated documentation, or if you still have any questions or comments about the tool.

☰ ⬡ AllenInstitute / ACE

📧 PROVIDE FEEDBACK   〈〉 Code   ⊙ Issues 10   ↕ Pull requests   💬 Discussions