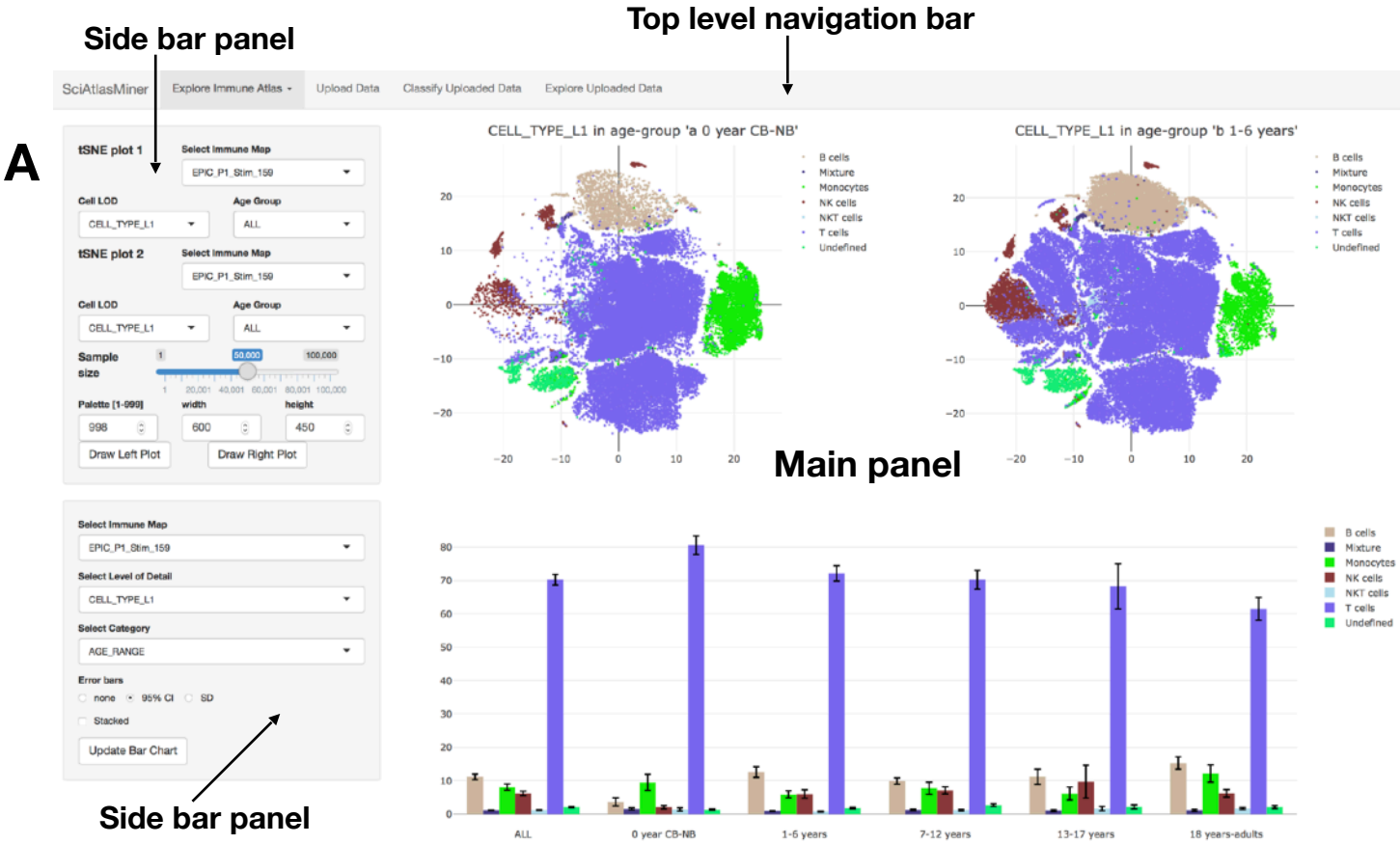
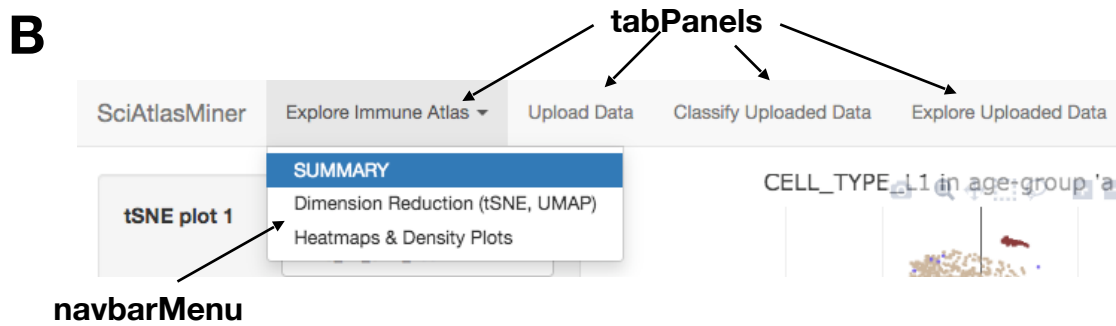
User Manual

SciAtlasMiner

1.1 *SciAtlasMiner* Layout & Navigation

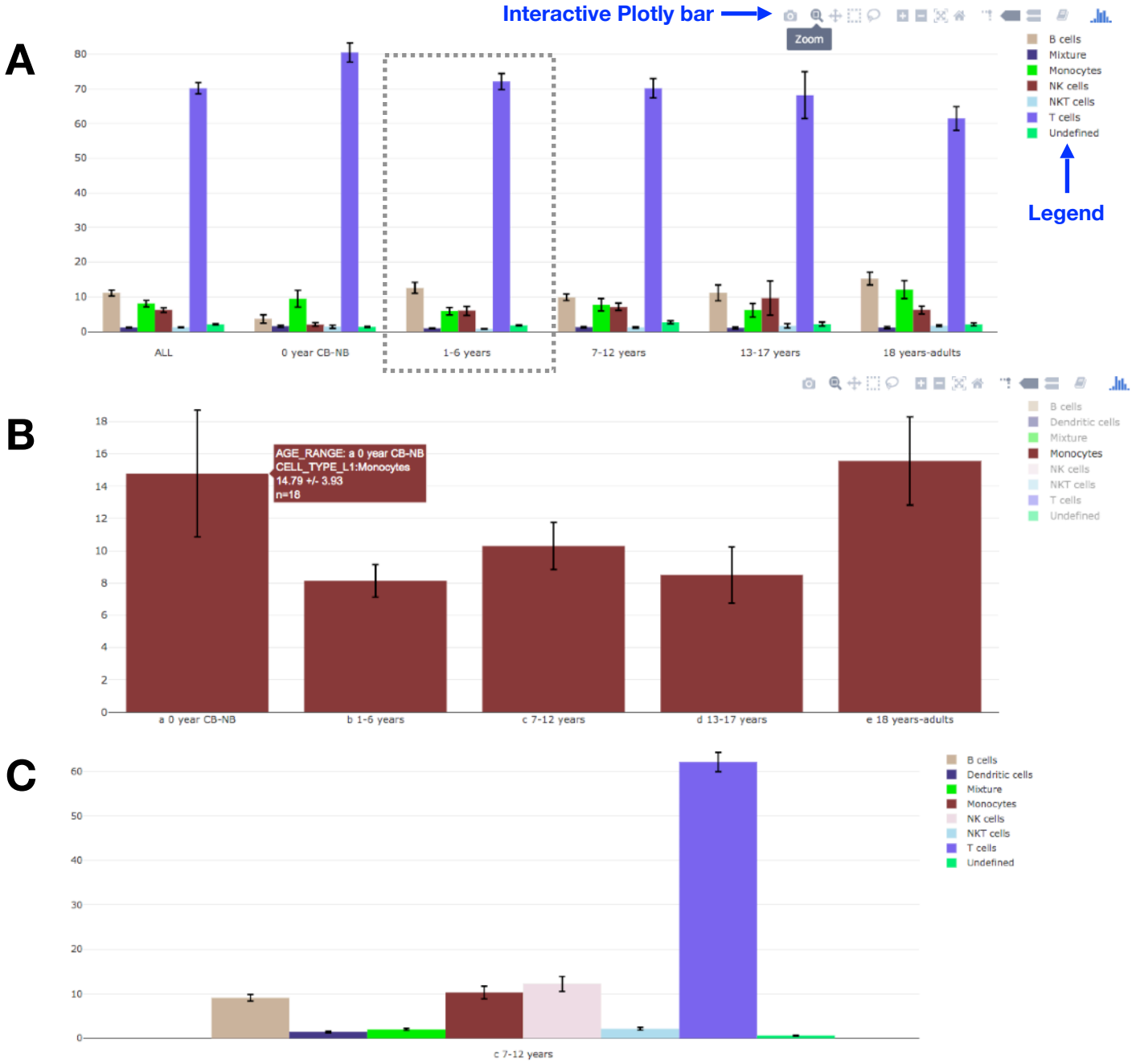




(A) Layout. The basic layout consists of top-level navigation bar that helps to switch between different panels, one or more side bar panels that contain the widgets of the user interface and a main panel that contains the interactive plots.

(B) Navigation. Click the tabPanels to navigate to the different functional panels. Embedded in the first tab 'Explore Immune Atlas' is a navBarMenu pull-down menu that controls navigation to multiple sub-panels.

1.2 *SciAtlasMiner* Interactive Plots



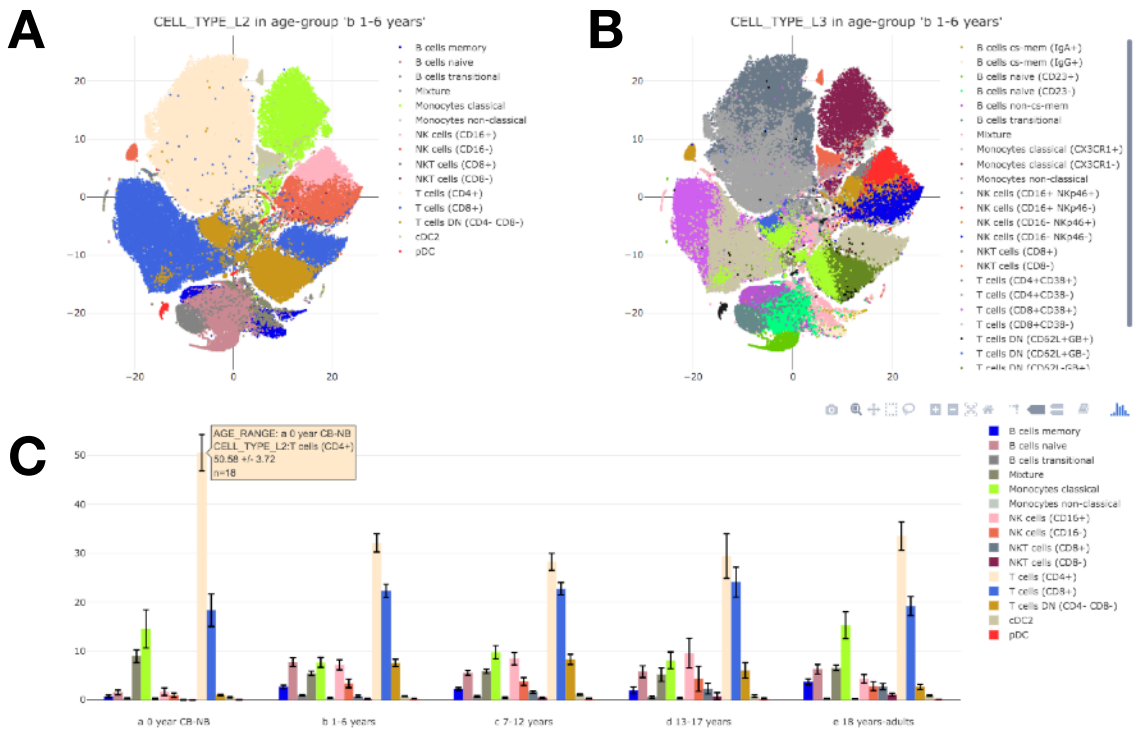
(A) The interactive Plotly Bar has multiple options, including (from left) download plot as PNG, zoom, pan, auto scale and reset axis.
Legend. Single mouse-click on a legend item toggles the visibility of a single trace, while **double mouse-click** toggles the visibility of all traces.

(B) Show data subset. A double mouse-click on 'Monocytes' in the legend isolates the cell subset in all age groups.

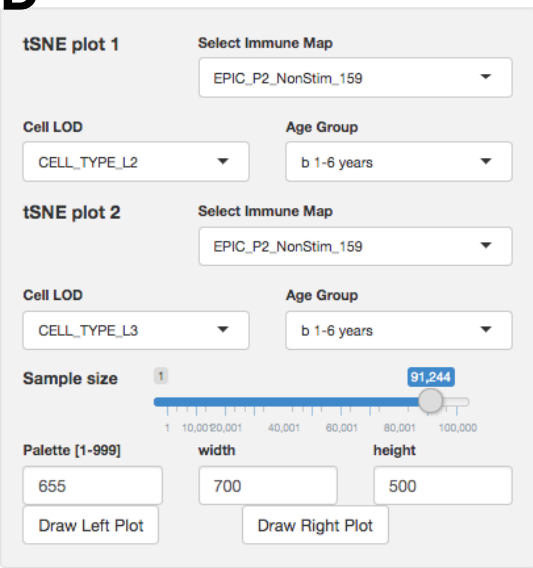
(C) Zooming. Zooming the 1-6 years olds in A.

2.1 Explore Immune Atlas

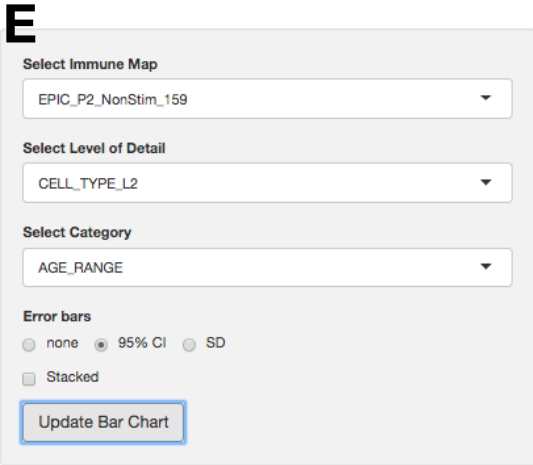
Summary: Development of the Immunome



D



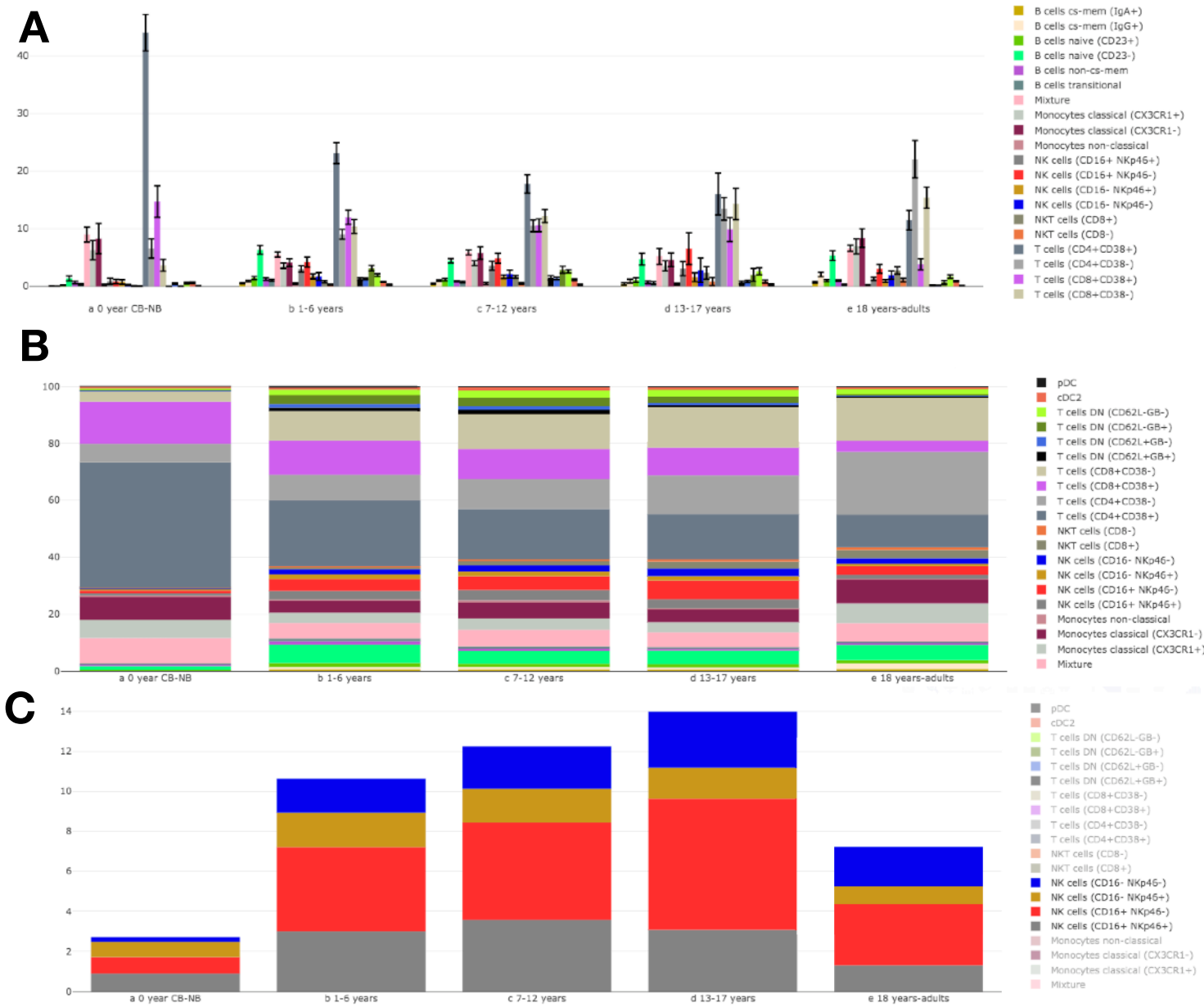
E



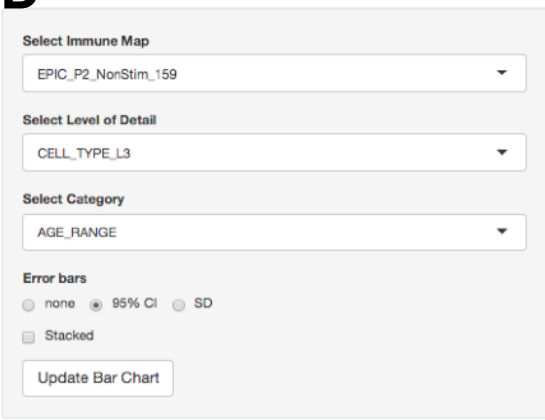
- **(A, B) tSNE** embedding plots compare the immunome of the same age group (1-6 years) at 2 different levels of detail (LOD). Colors represent cell populations.
- **(C) The bar chart plot** compares cell types of one LOD in 5 different age groups.
- **(D) Specify parameters** of scatter plots A & B.
  - Select immune map
  - Cell LOD: cellular level of detail
  - Age Groups
  - Sample size (default 50000)
  - Palette: color palette value from 1-99 sets the seed for random colour selection.
  - Width: panel width in pixels
  - Height: plot height in pixels
  - Press 'Draw Left/Right Plot' buttons to update plots
- **(E) Specify bar chart parameters.**
  - Immune map
  - Select Level of detail
  - Select category for grouping cells (default is AGE_RANGE)
  - Error bars: none, 95% confidence interval (95% CI) or standard deviation (SD)
  - Stacked: if checked, bars will be displayed in a stacked style (see next page)
  - Press 'Update Bar Chart' button

2.1 Explore Immune Atlas

Summary Bar chart: Development of the Immunome

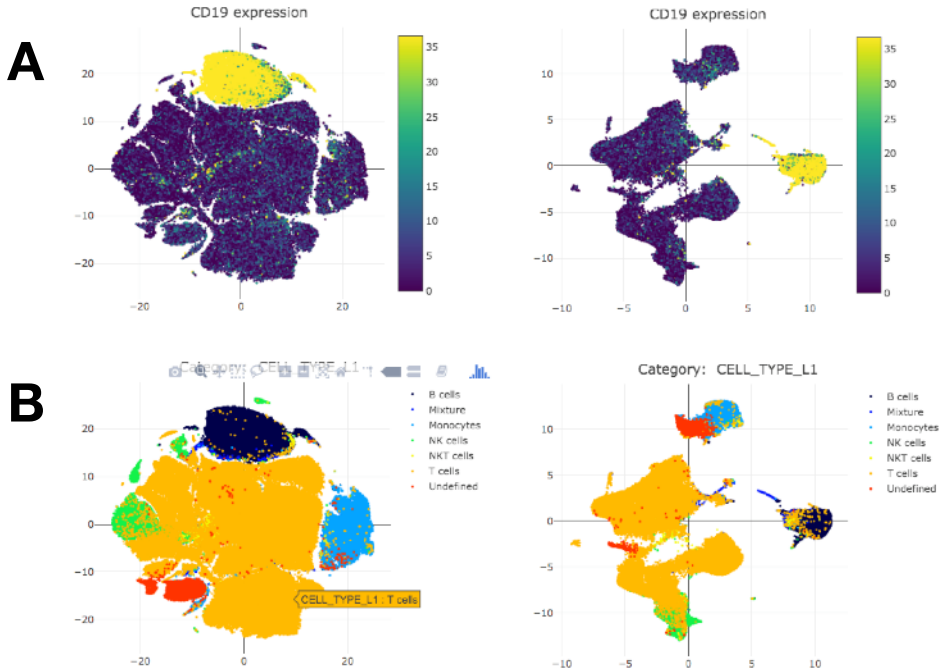


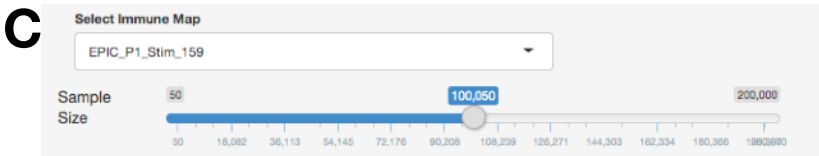
D

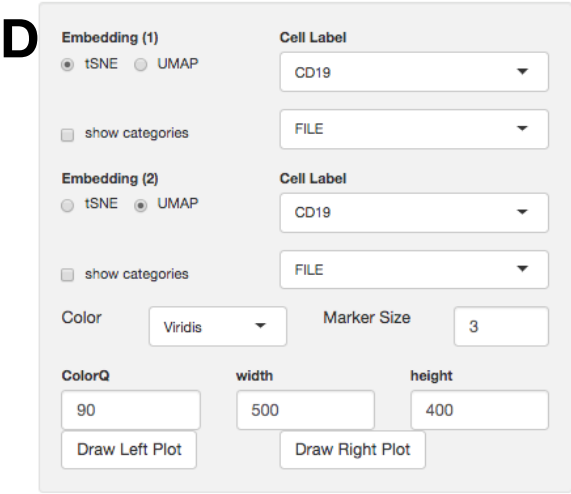


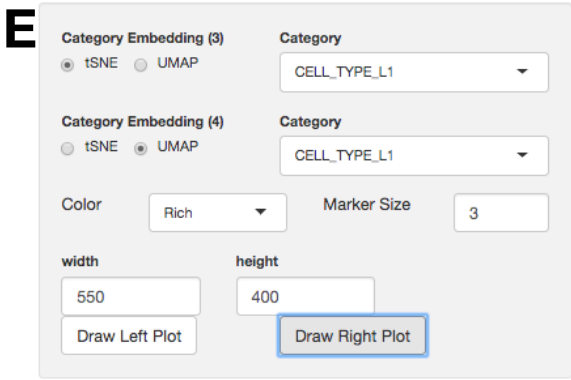
(A) Default vertical style bar chart style with error bars. The plot shows 25 cell types in 5 age groups.
(B) Check the Stacked CheckBox to switch to the stacked bar chart style of the same data
(C) Stacked bar chart showing 4 NK cell populations
(D) Control panel

2.2 Explore Immune Atlas Dimension Reduction Plots





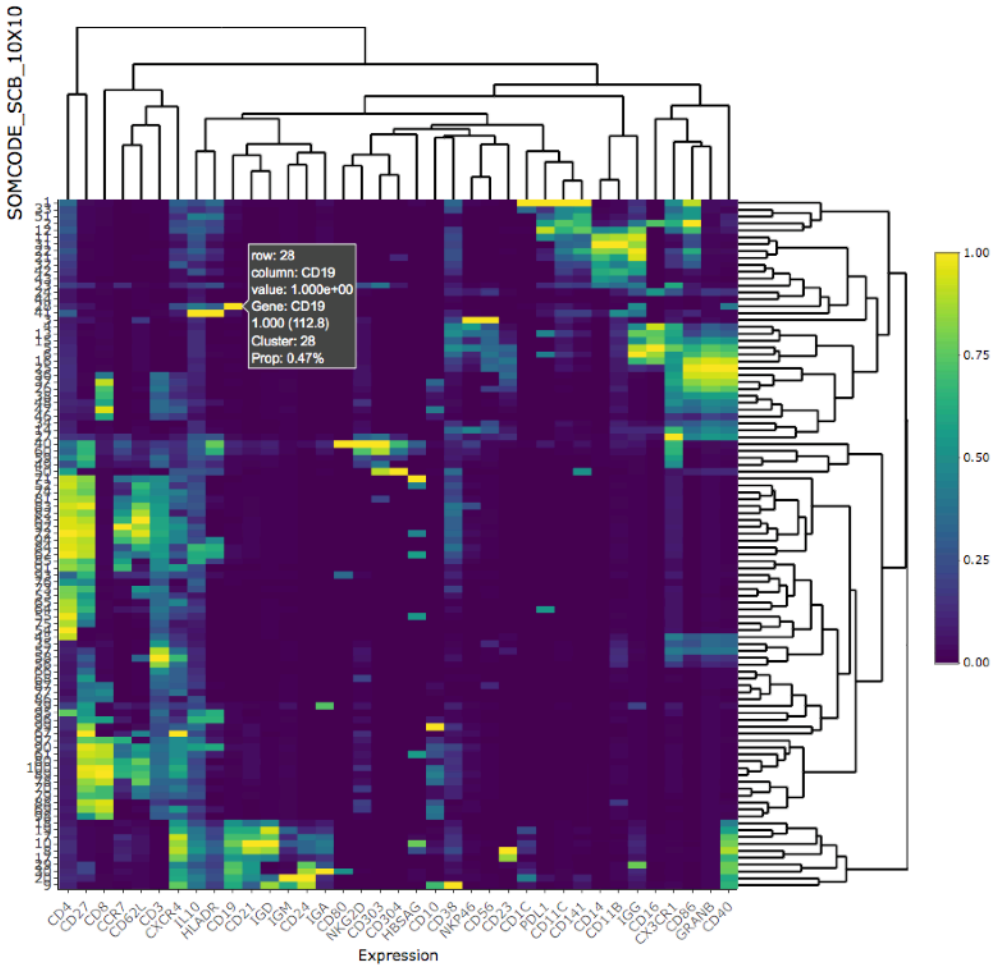




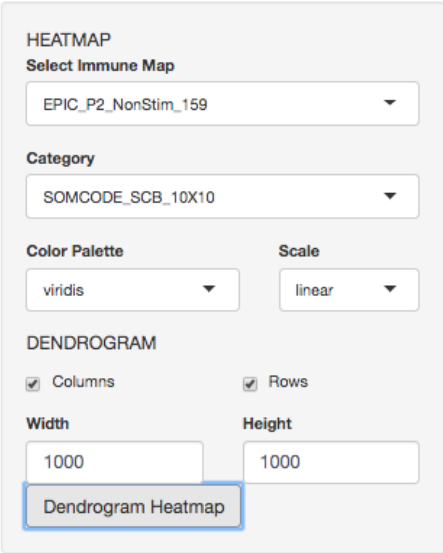
- **In the top level navigation bar, select *'Explore Immune Atlas' -> 'Dimension Reduction (tSNE, UMAP)'***
- **(C)** Select Immune Map and adjust slider to set sample size
- **(D)** Set parameters to display one or two embedding charts (panel A) that overlay **protein median expression** using a continuous color scale
  - Choose between tSNE or UMAP
  - Select cell label/antibody
  - 'show categories' groups data by category subsets that can be toggled on or off using a legend. Select the category in the input on the right of the checkbox.
  - Color: select the continuous color palette
  - Marker size
  - ColorQ: color quantile (default 90%). The quantile of expression values, above which all values are assigned the maximum colour value
  - Width, height: plot dimensions in pixels
  - Draw left/right plot buttons update the plots
- **(E)** set parameters of categorical non-linear embedding plots (panel B)
  - tSNE or UMAP embedding
  - Category: e.g. cellular LOD, cluster label
  - Color palette and marker size
  - width, height: plot dimensions in pixels
  - Draw left/right plot buttons update the plots

2.3.1 Explore Immune Atlas Dendrogram Heatmap

A

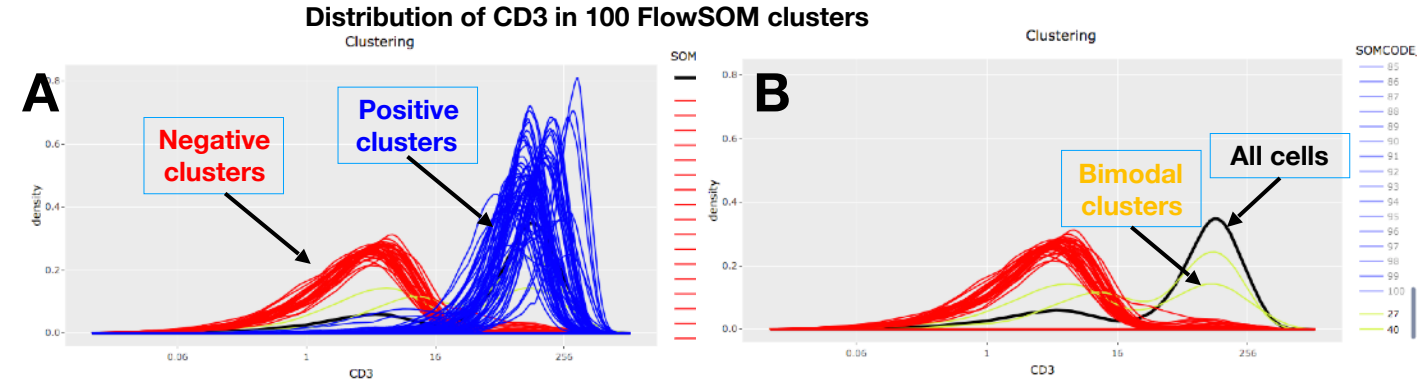


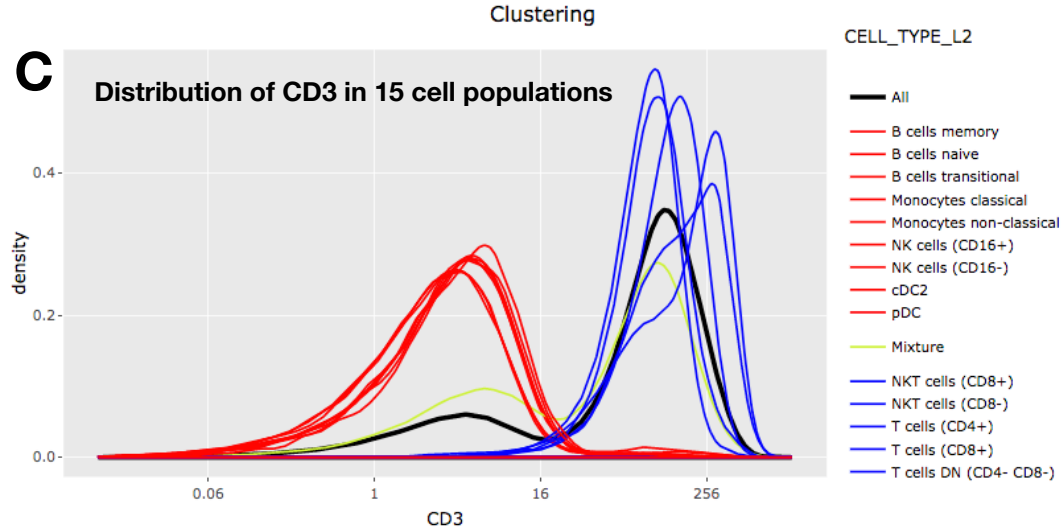
B



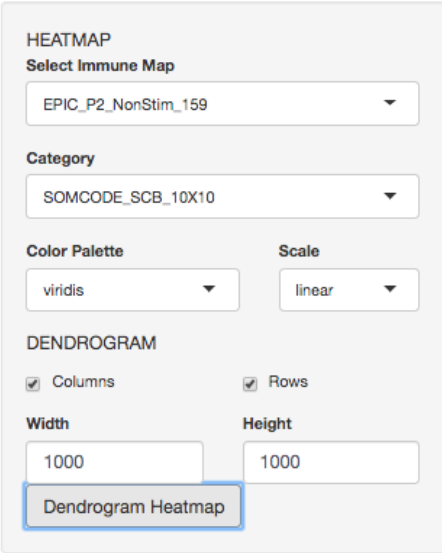
- **In the top level navigation bar, select *'Explore Immune Atlas' -> 'Heat maps & Density Plots'***
- **Set parameters of dendrogram heatmap**
  - Select Immune Map
  - Select Category, e.g. cluster labels or cell population
  - Color palette
  - Scale: no (linear) or arcsin scaling of expression values
  - Columns and Rows checkboxes indicate if markers and categories are subjected to hierarchical clustering (default checked)
  - Width and height adjust the plot dimensions in pixels
  - Click the dendrogram heat map button to update the heatmap

2.3.2 Density Plots

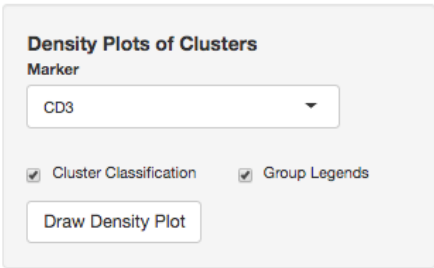




D

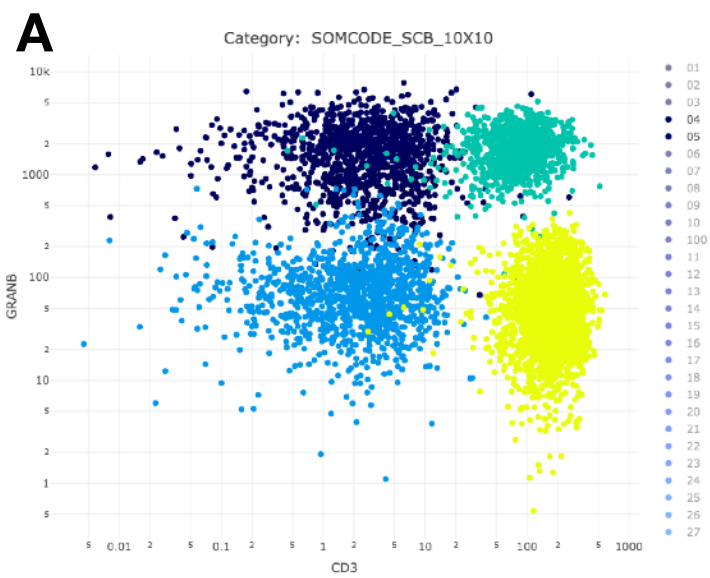


E

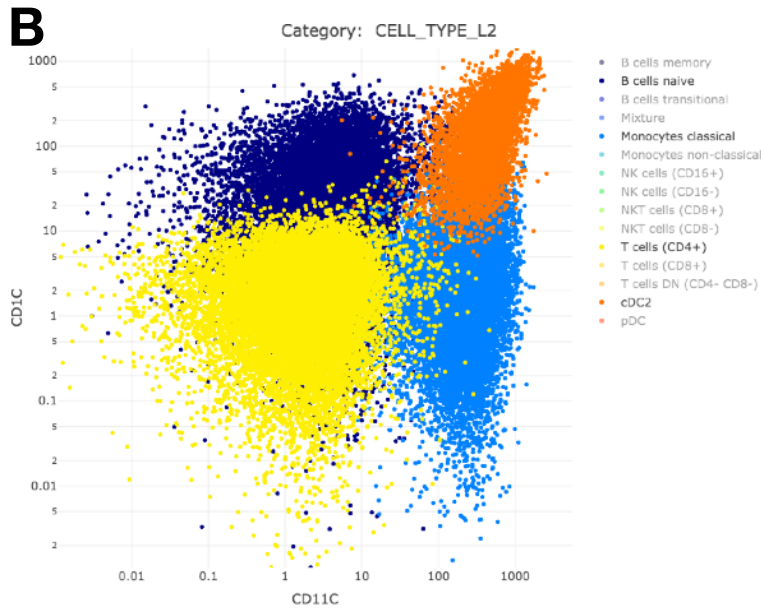


- **In the top level navigation bar, select *'Explore Immune Atlas' -> 'Heat maps & Density Plots'***
- **Set parameters of dendrogram heatmap**
  - Select Immune Map (D)
  - Select Category (D), e.g. cluster labels (A,B) (SOMCODE_SCB_10x10) or cell populations (C).
  - Select marker (E)
  - Tick "cluster classification" Checkbox to colorize KDE curves (smoothed histogram). Black=all cells, red=negative expression, blue=positive expression, green=bimodal distribution indicating mixed populations.
  - Tick "group legend" checkbox to collectively toggle the visibility of curves
  - Click the "Draw Density Plot" button to update the density plot

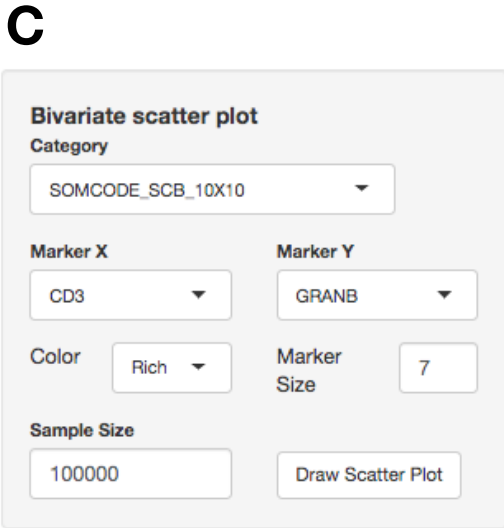
2.3.3 Bivariate scatter plots



View 4 out of 100 FlowSOM clusters (CD3 vs. GranB)

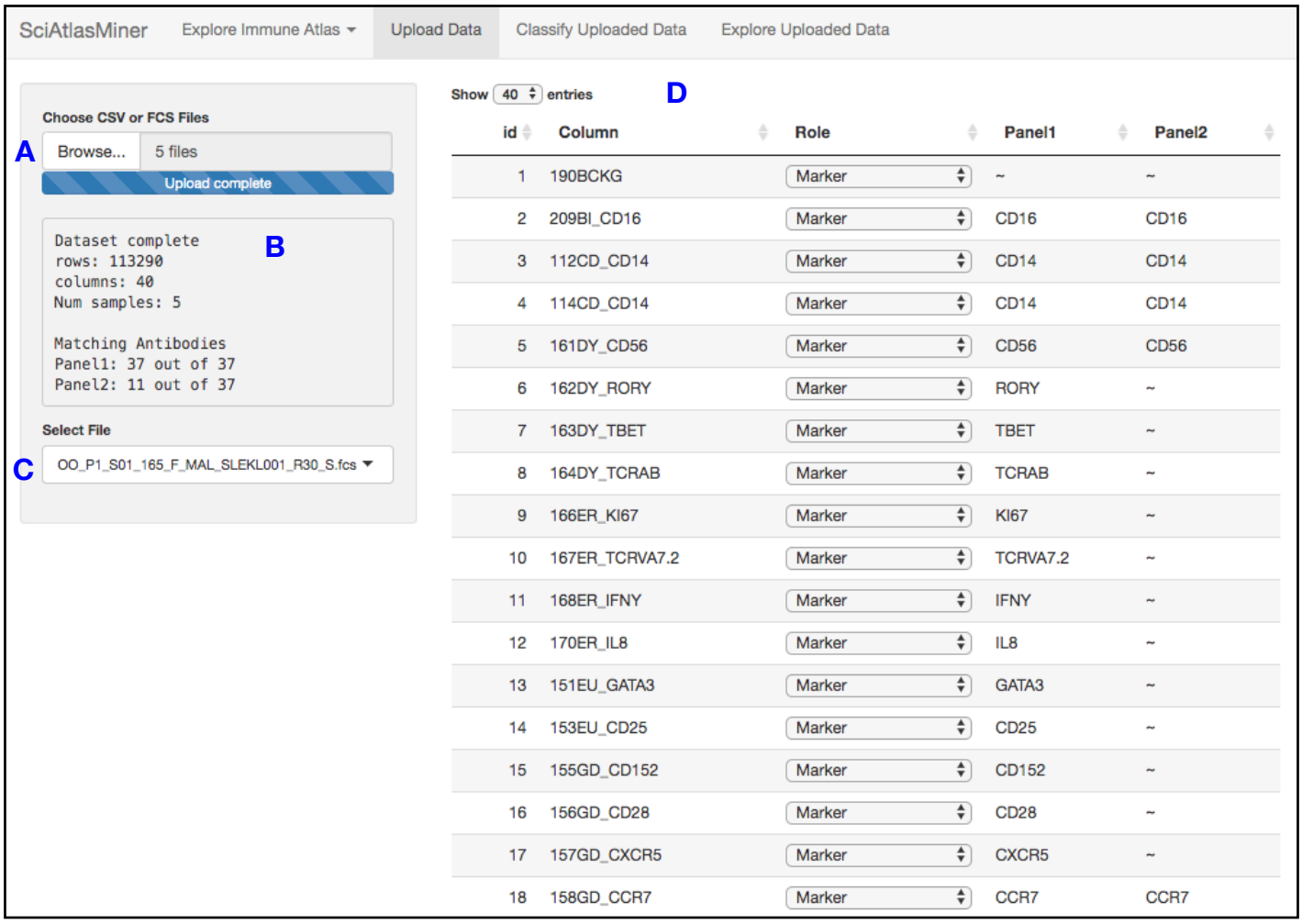


View 4 out of 15 cell population (CD11C vs. CD1C)



- **In the top level navigation bar, select *'Explore Immune Atlas' -> 'Heat maps & Density Plots'***
- **Scroll to the bottom menu 'Bivariate scatter plot' (C)**
- **Set parameters of scatter plot (C)**
  - Select Immune Map (page 7, panel E)
  - Select Category (C). It is possible to view all or a subset of cluster labels (A) or annotated cell populations (B).
  - Select X-axis marker
  - Select Y-axis marker
  - Select color palette and marker size
  - Select sample size to
  - Click the "Draw Scatter Plot" button to update the scatter plot

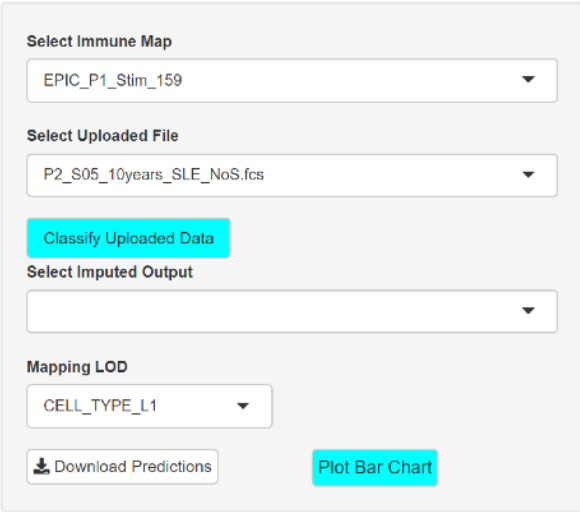
3. Upload Data



- **In the top level navigation bar, select *'Upload Data'***
- Click the Upload button titled "Choose CSV or FCS files" (A)
- In the file browser dialog, select one or multiple FCS or CSV files
- The program automatically concatenates files that share the same antibody/marker panel
- File details are shown in the left panel (B) and uploaded file names are appended to a dropdown list (C)
    - Number of rows, columns and files of newly created data-frame
    - Number of matching antibodies between uploaded file(s) and antibody panels of the atlas
- The table in the right panel (D) lists column details of the uploaded file(s) such as inferred matches to the antibodies of the database and inferred roles (marker, embedding, category, ignore)

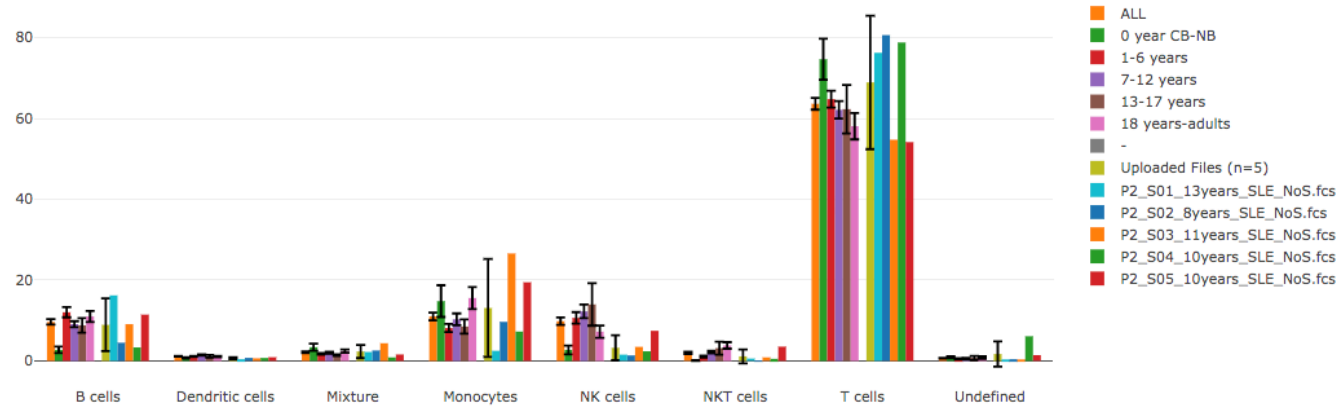
4.1 Classify Uploaded Data (Bar charts)

A

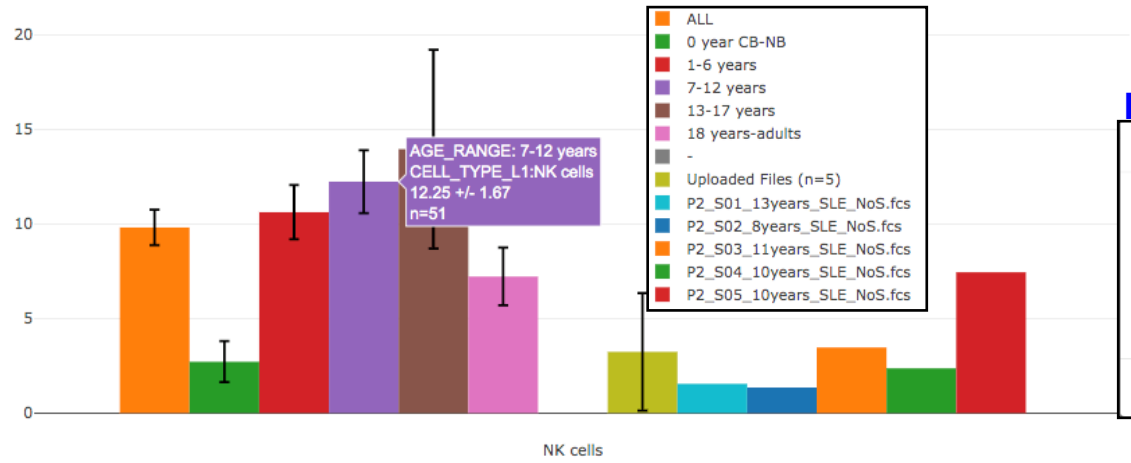


- **In the top navigation bar, select *'Classify Uploaded Data'***
- (A) Select Classification parameters
  - Immune Map
  - Uploaded File(s)
- Click "Classify Uploaded Data" button to map uploaded data to SOM of immune map
- (B) After mapping clusters (SOM codes) to annotated cell types, cell frequencies (% of PBMCs) for the specified level of detail (LOD) are shown.
- (C) Data for each cell type are shown in the following format
  - Mean frequencies of all cells and age groups of the immne map with error bars (95% CI)
  - Mean frequencies of all uploaded files (95% CI)
  - Frequencies of individual samples
- (D) If antibodies of immune map and uploaded data do not match completely, a warning message is displayed

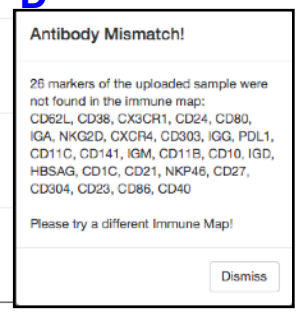
B



C

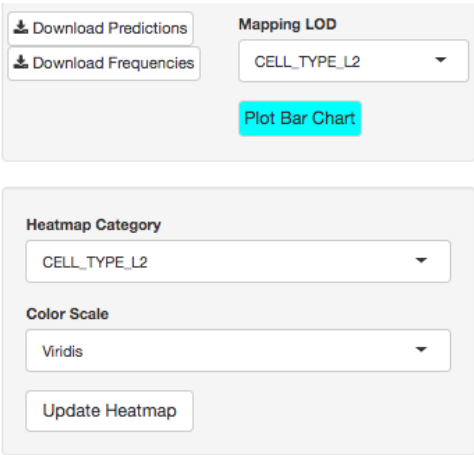


D



4.2 Classify Uploaded Data (Validation)

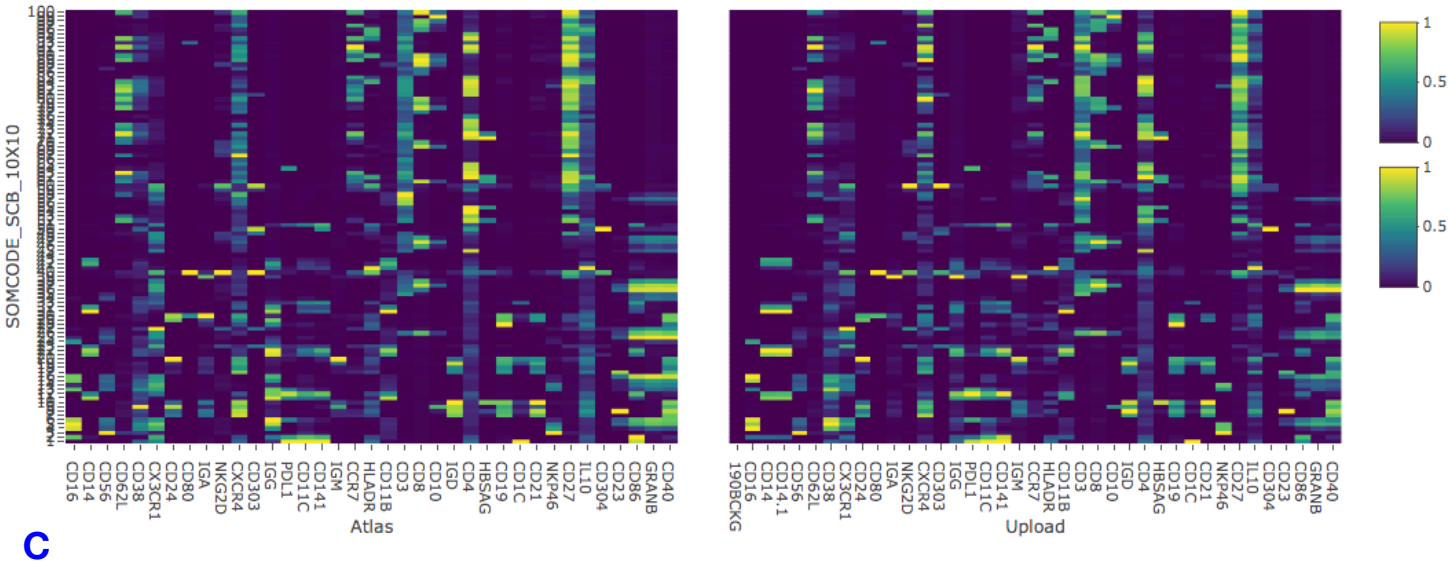
A



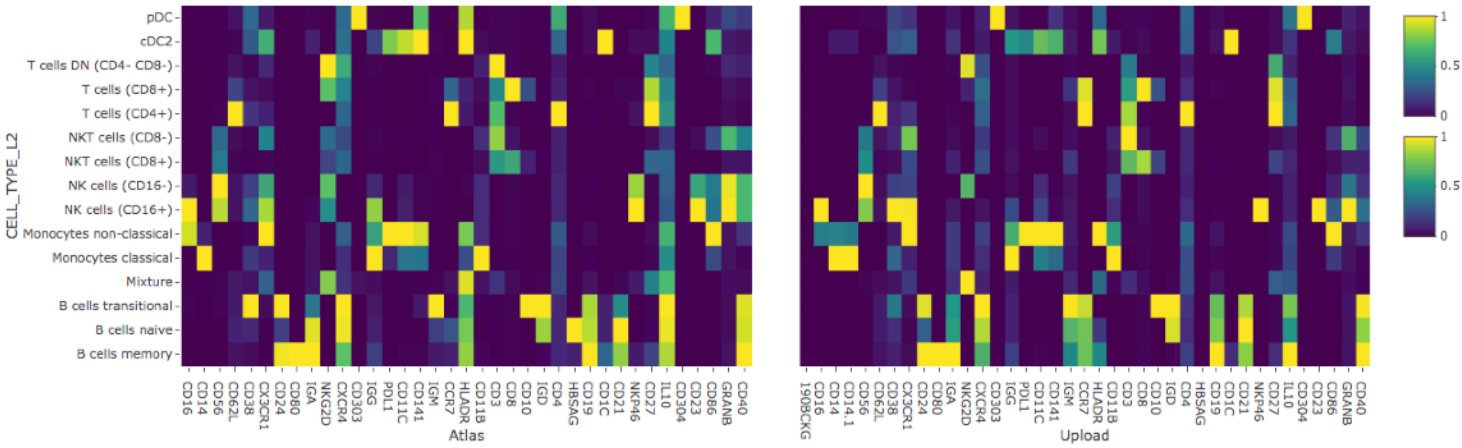
- **Heat maps** compare clustered protein expression patterns of atlas data and imputed clusters of uploaded data. Cells can be grouped according to cluster (B) or classified cell type (C).
- (A) Select heat map category, e.g. cluster method, cell annotation layers, etc.
- Choose color scale
- Click “Update Heatmap” button

- Uploaded data can be exported with classification outputs (cluster labels, cell types) by clicking the “**Download Predictions**” button. Downloaded CSV files are saved to “Download” folders according to OS. Downloaded results can be further analysed
    - Add ground truth labels obtained using manual gating and determine classification metrics such as recall and precision.
    - Re-import into SciAtlasMiner and explore using the “Explore Uploaded Data” module (see next page)

B

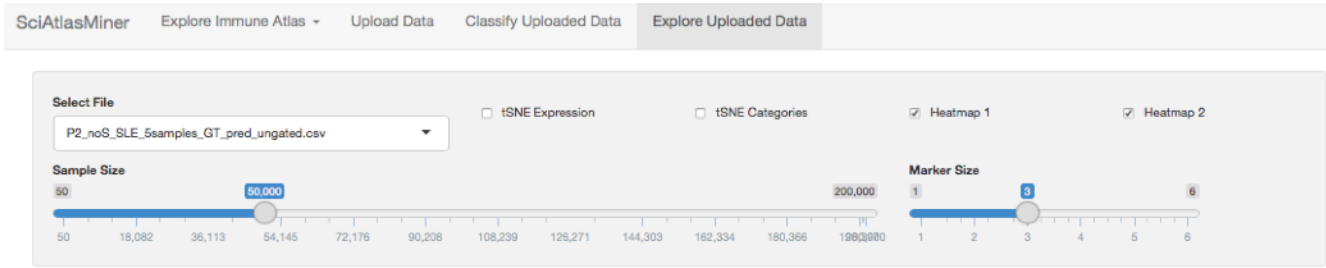


C



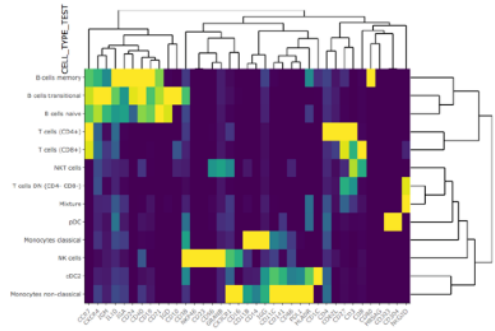
5. Explore Uploaded Data

A

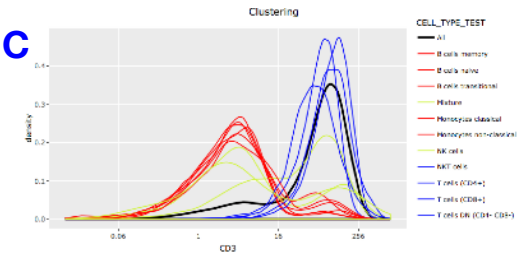


- This module is suitable for CSV files containing Cytometry expression data along with processed data (e.g. t-SNE) and/or categorical columns. FCS files do not contain sufficient information
- Upload CSV file as outlined in section 3 (page 10)
- In the top navigation bar, select *'Explore Uploaded Data'* and choose one of five modules
    - Modules 1 & 2: t-SNE embedding overlayed with marker expression or categories (not shown)
    - Modules 3 & 4: Dendrogram heatmaps (B) and density plots (C)
    - Module 5: Two-category heatmaps visualize the correspondence between manually created ground truth and automated clustering (D) or annotation (E), where the color scales represent frequency.

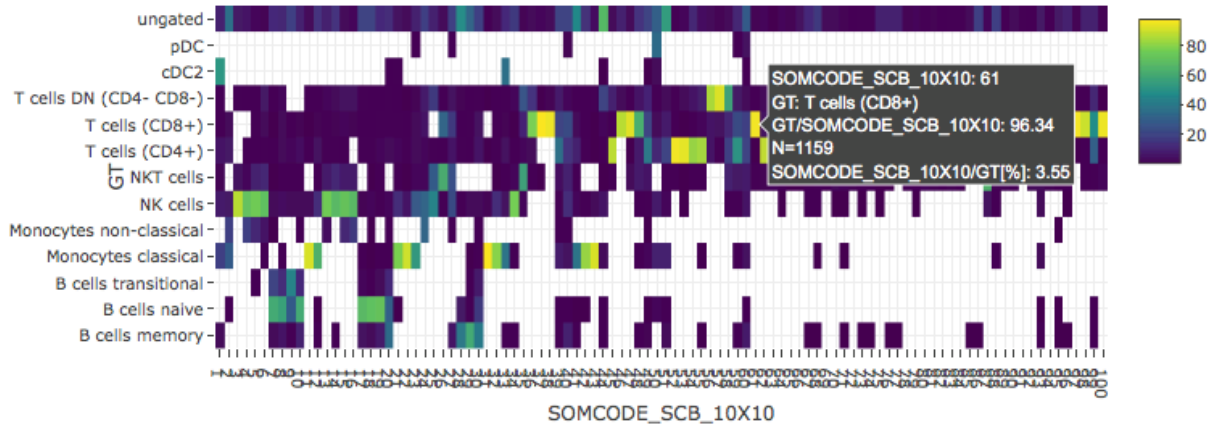
B



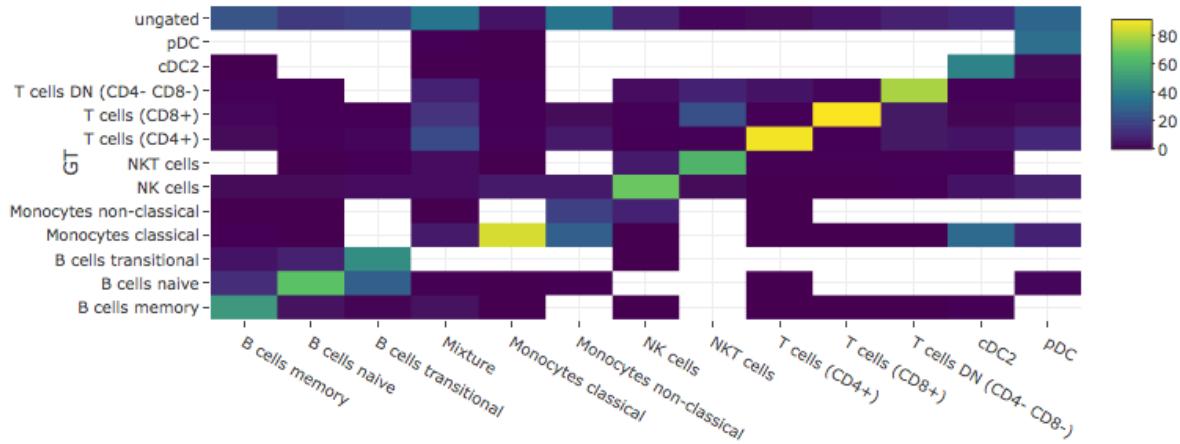
C



D



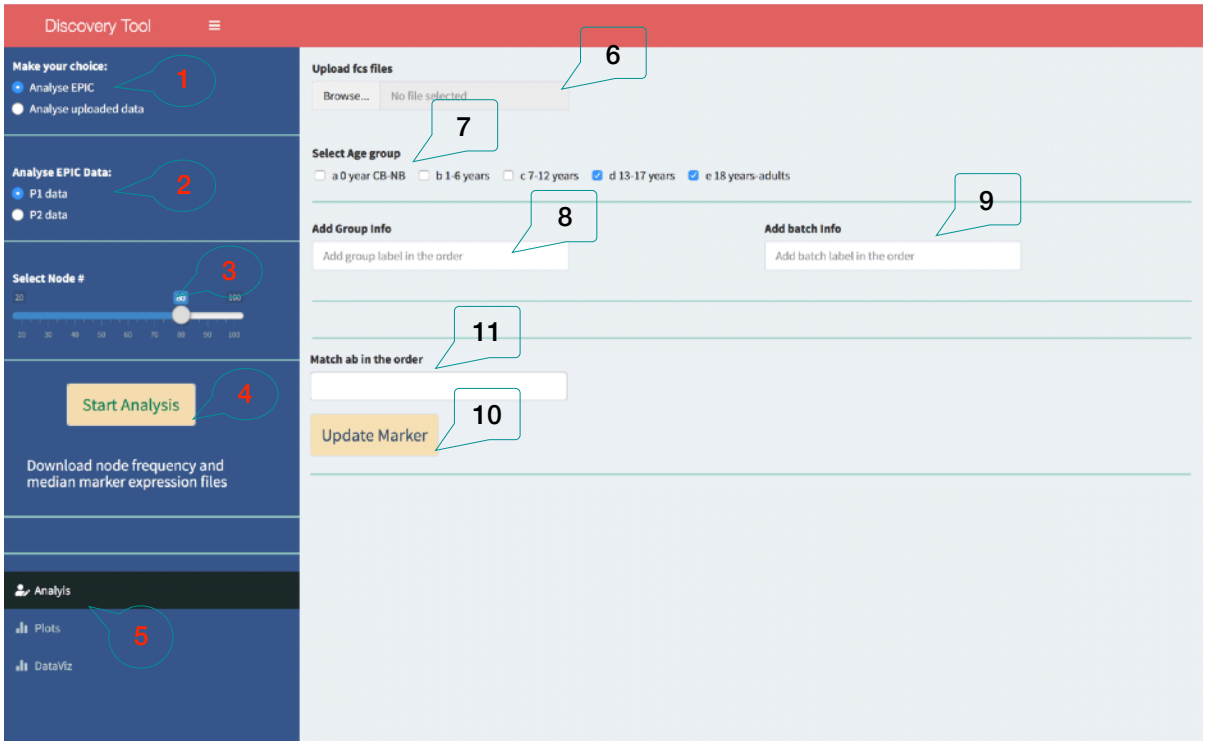
E



User Manual

Discovery Tool

Discovery Tool : User Interface (UI) layout



Sidebar panel **Main panel**

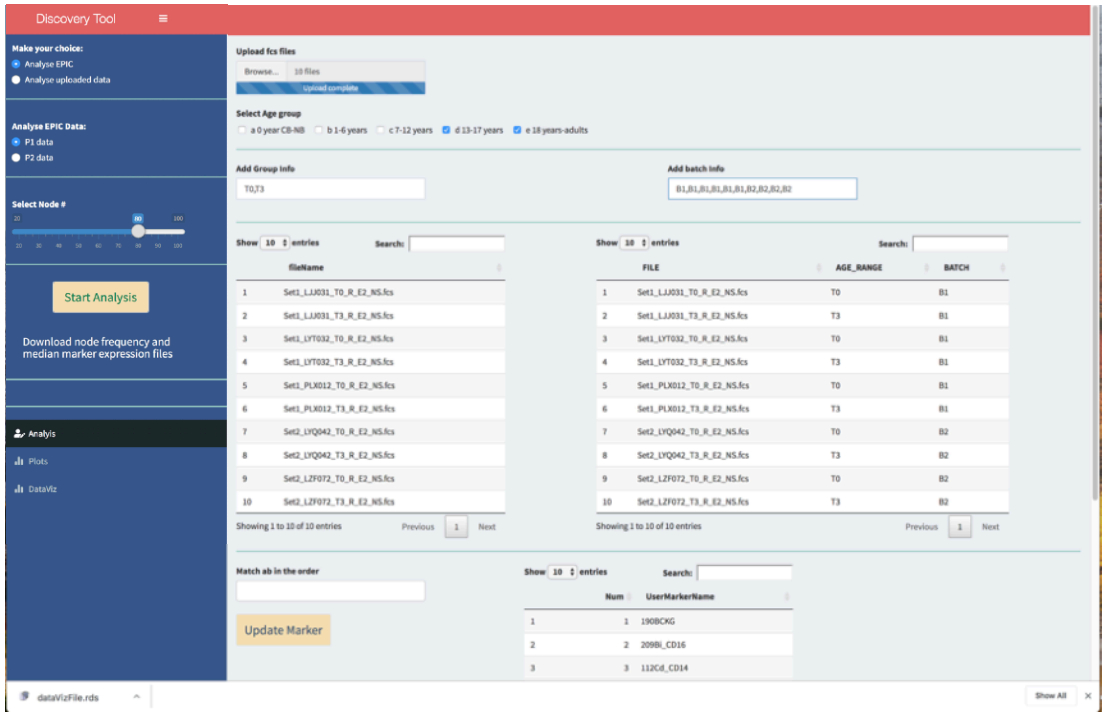
The application UI is diveded into sidebar panel (blue color) and main panel(gray color)

The sidebar UI:
1) Chooser UI (**Make your choice**) to select the analysis type: By default Analyze EPIC is selected
2) A Chooser UI conditioned on selection from **Make your choice** (1) If Analyse EPIC is selected **Analyse EPIC Data** UI will appear. (2) If Analyse uploaded data is selected **Analyse uploaded Data** will appear.
3) **Slider Input** UI to choose number of node for clustering the data: By default it's set on 80
4) **Start Analysis** Clickable Button to start the analysis
5) Links(**Analysis**, **Plots**, **DataViz**) for main panel

By default Analysis tab on the sidebar (5) is select to display the main panel. Selected panel is highlighted as black bar in the sidebar panel

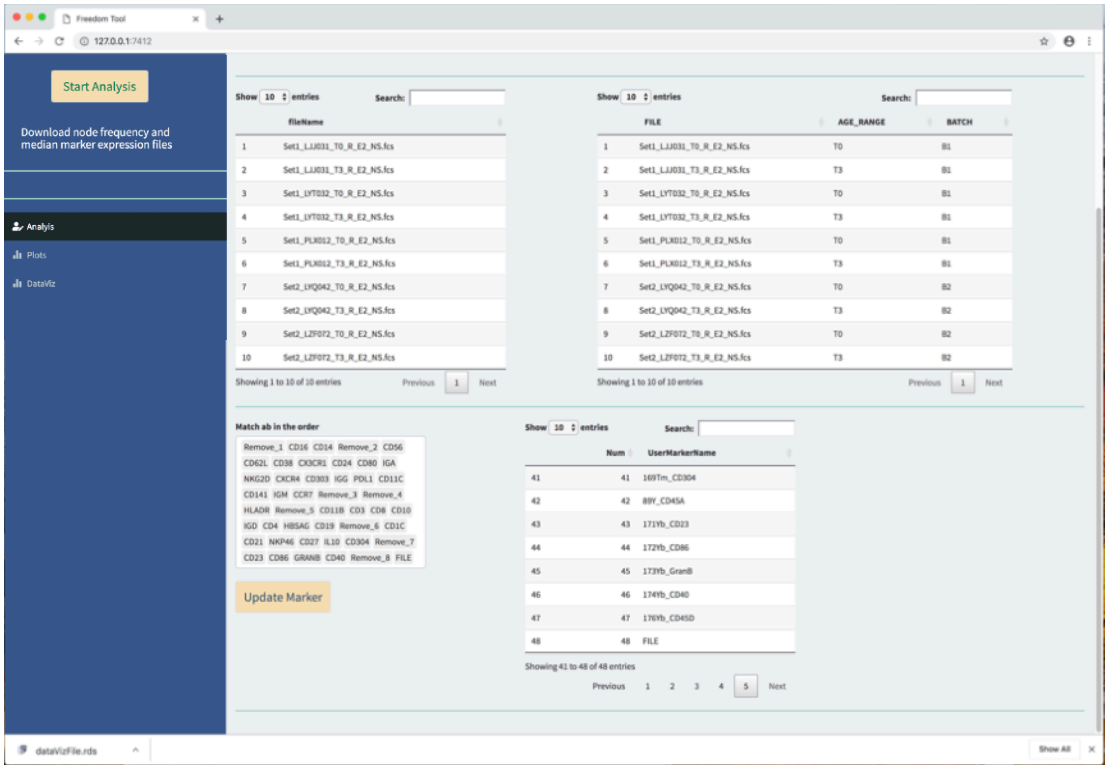
The main Analysis panel shows:
6) Upload file browser UI (**Upload fcs files**)
7) Checkbox UI (**Select Age group**)for selecting the age category to include in the analysis
8) Text input field (**Add Group Info**)to key in group information for each uploaded fcs file
9) Text input field (**Add batch Info**)to key in batch information for each uploaded fcs file
10) Update Marker button
11) If Analyse uploaded data is selected then an additional dropdown Input field (**Match ab in the order**) will appear.

Sidebar Panel tab :Analysis

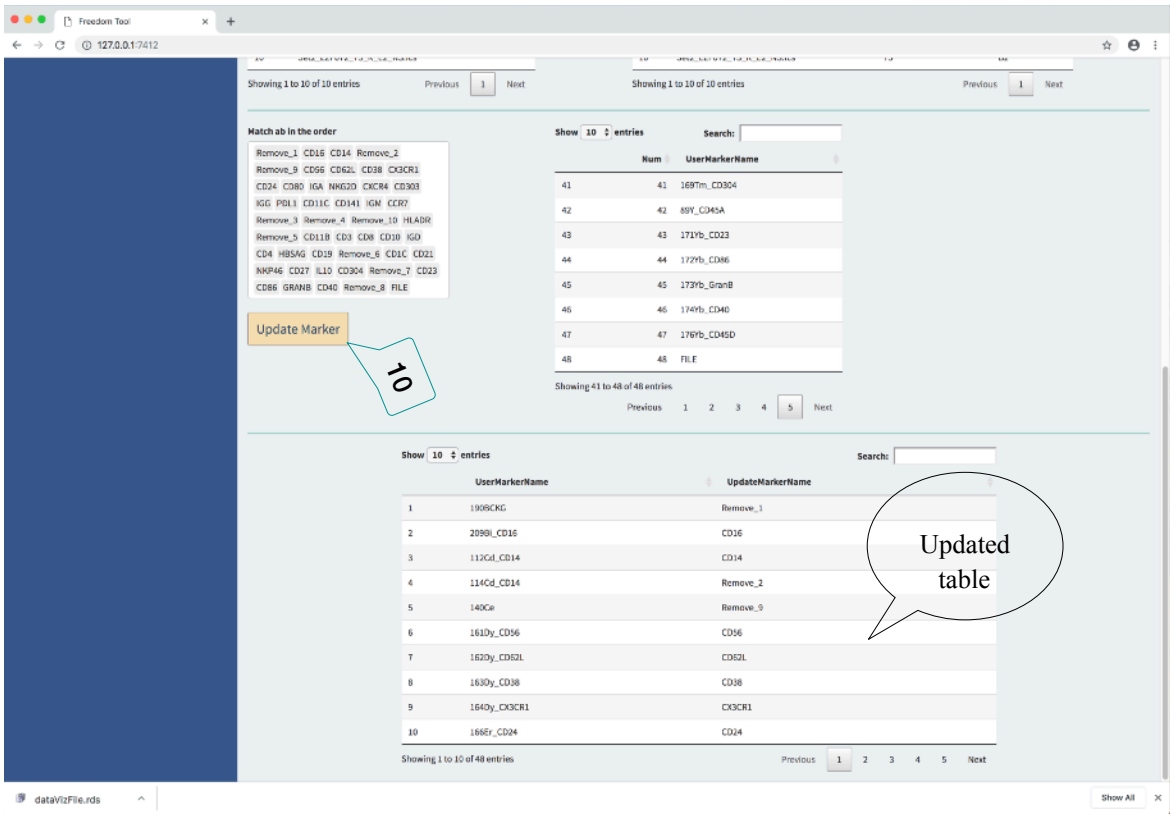


Once the user uploads the data using **Upload fcs files** UI (6) in the Analysis main panel, filename will be displayed below the **Add group info** tab (8). User must input the group and batch information for each file in the order the filename displayed in the **Add Group Info** input field (8) and in the **Add batch info** input field (9). Group and batch information for each files must be comma separated without any space between the text. If all the sample belong to one batch user can write only single label for all the file.

Once the user key in the group and batch info new table with filename and corresponding group and batch info will be displayed for verification.



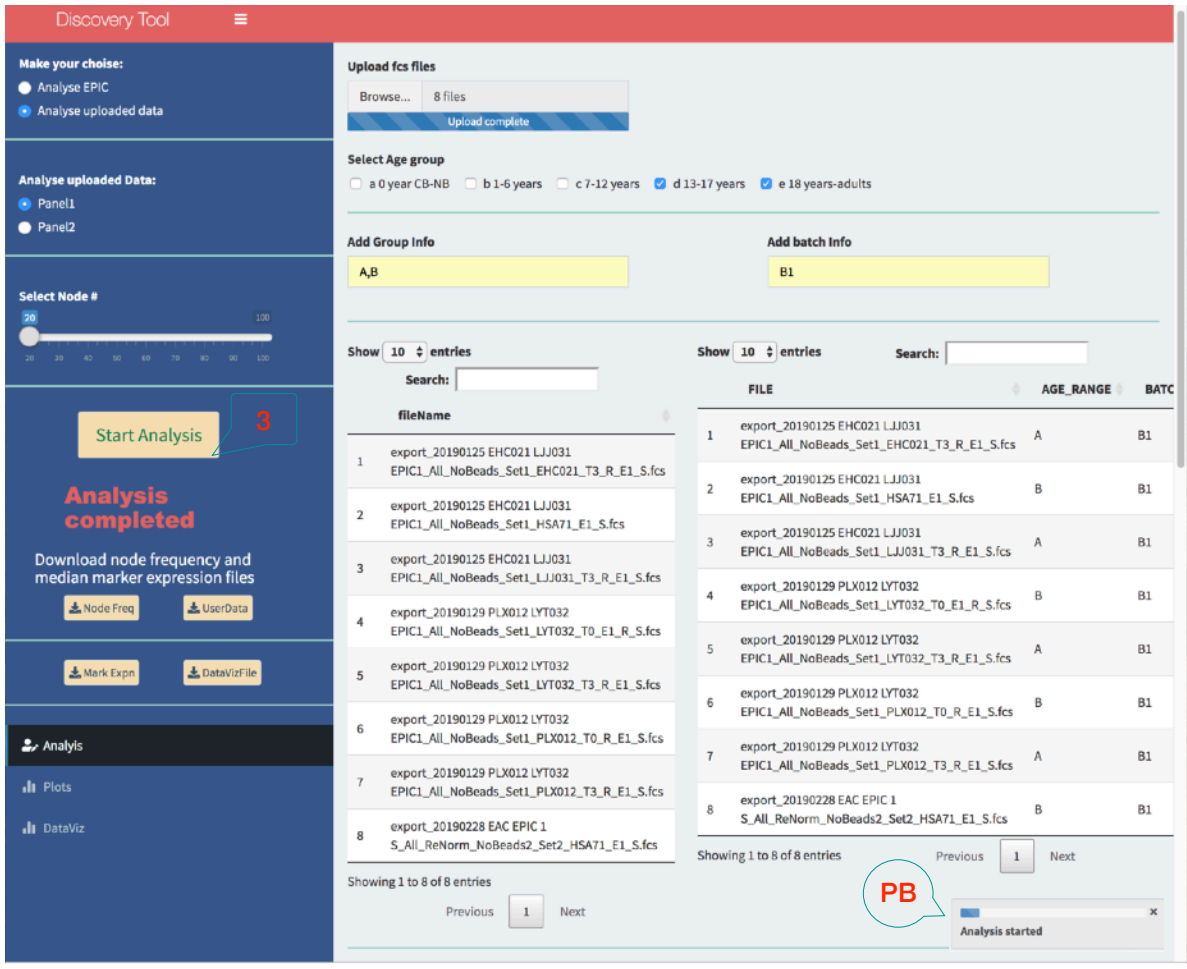
Next, user must choose the markers from dropdown chooser panel "**Match the ab in the order**" field (11). If markers in user files are not in the dropdown list user must choose one of the Remove_n tag to remove it from clustering and analysis. The Remove_n flag can also be used if user did not want to include a particular antibody to be used for analysis. The last marker must be FILE. User should check if the order of antibody in the "Match ab in the order" field matches with the table displayed next to it.



After inputing matching antibody user should click **Update Marker** button (10), after clicking this button a new table will appear below showing updated marker label.

Once user updated group and batch information key-in matching antibody and verified the updated antibody list, chosen the number of desired clusters and checked the boxes to include the age category to include the healthy samples for analysis, user can start analysis by clicking the start analysis button in the side bar panel (4).

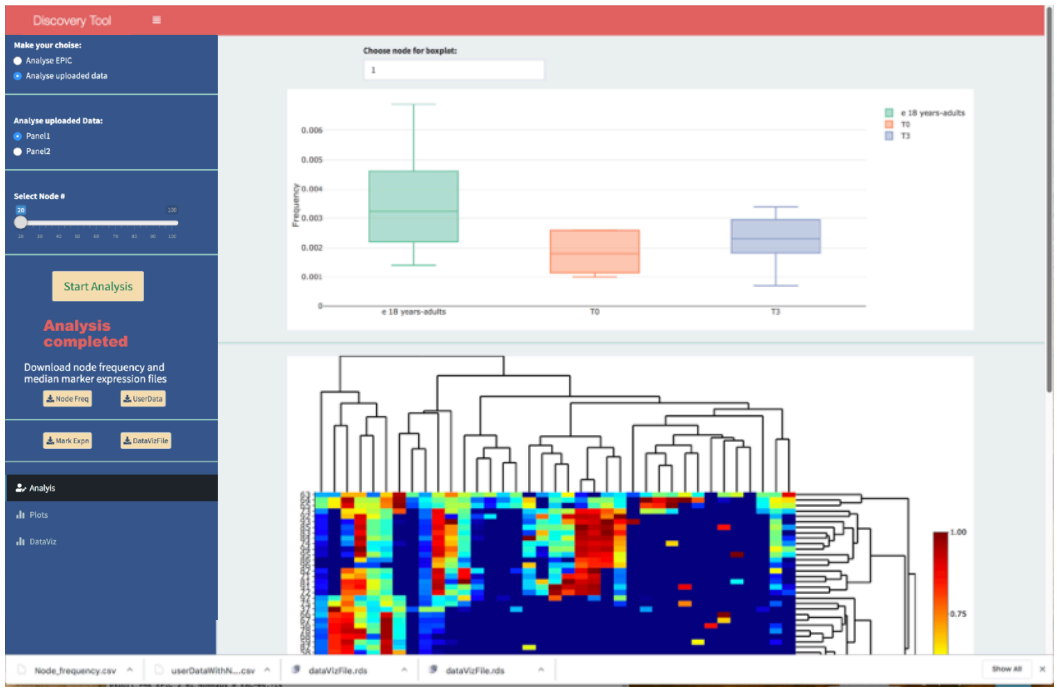
Once the analysis start a progress bar in the bottom right of the main panel will appear (PB). Wait for the analysis to completed. Once the analysis is complete progress bar will disappear and on the sidebar below the start analysis button a massage "**Analysis Complete**" will be displayed. Along with the massage 4 new download button will appear. Each download button allows to download data for further analysis and visualization.



Download buttons:

1) Node Frequency: Clicking of this button will download the frequency of each node for each of the uploaded files and the samples chosen from the EPIC database. File will be in csv text formate and it will contain column with file names , group and batch info, and node_n names.
2) UserData: This button will download the user uploaded single cell data with file name and corresponding Node number for each cell.
3) Mark Expn: This button will download csv file with median marker intensity for each node.
4) DataVizFile: Upon clicking this button an .rds file will be downloaded. This .rds file can be used latter for visualization using the DataViz tab in sidebar panel.

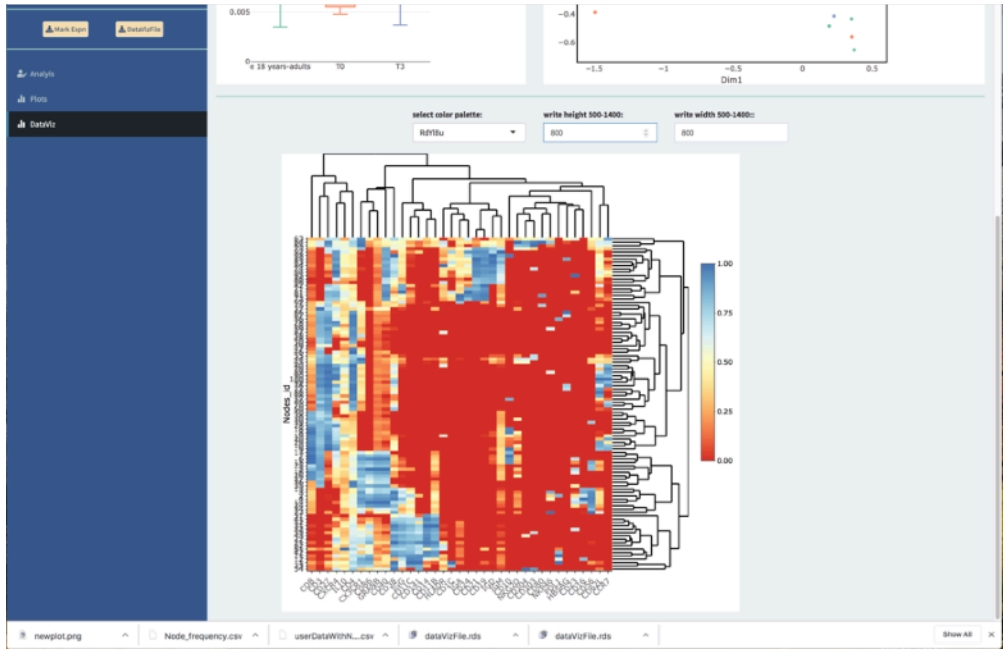
Sidebar Panel tab :Plot



Once the analysis is completed, user can click the Plot tab to switch from Analysis tab.
After analysis is completed Plot main panel will be populated with box-whiskers plot at the top and heatmap at the bottom. User can change the node number in the panel "**Choose node for boxplot**" to update the plot for respective selected node.

Boxplot will show the frequency of nodes across the group and heatmap can be used to find the phenotype of the nodes.

Sidebar Panel tab :DataViz



The DataViz.rds file downloaded after analysis using the Discovery tool can be uploaded to visualize the data. On clicking the DataViz tab, main panel will show UI for Upload rds file. Once user uploads the rds file, the panel will be populated with 3 kinds of data chart: (1) Box and whiskers plot at the upper left (2) Component analysis plot at the upper right (3) Heatmap at the bottom

Component analysis visualizes the frequencies of nodes for each sample on two principal components shown as Dim1 and Dim2. Each sample in the plot will be color coded for the group information.

Heatmap can be customized for height, width and color palette. A total of 7 different color palette option is available.