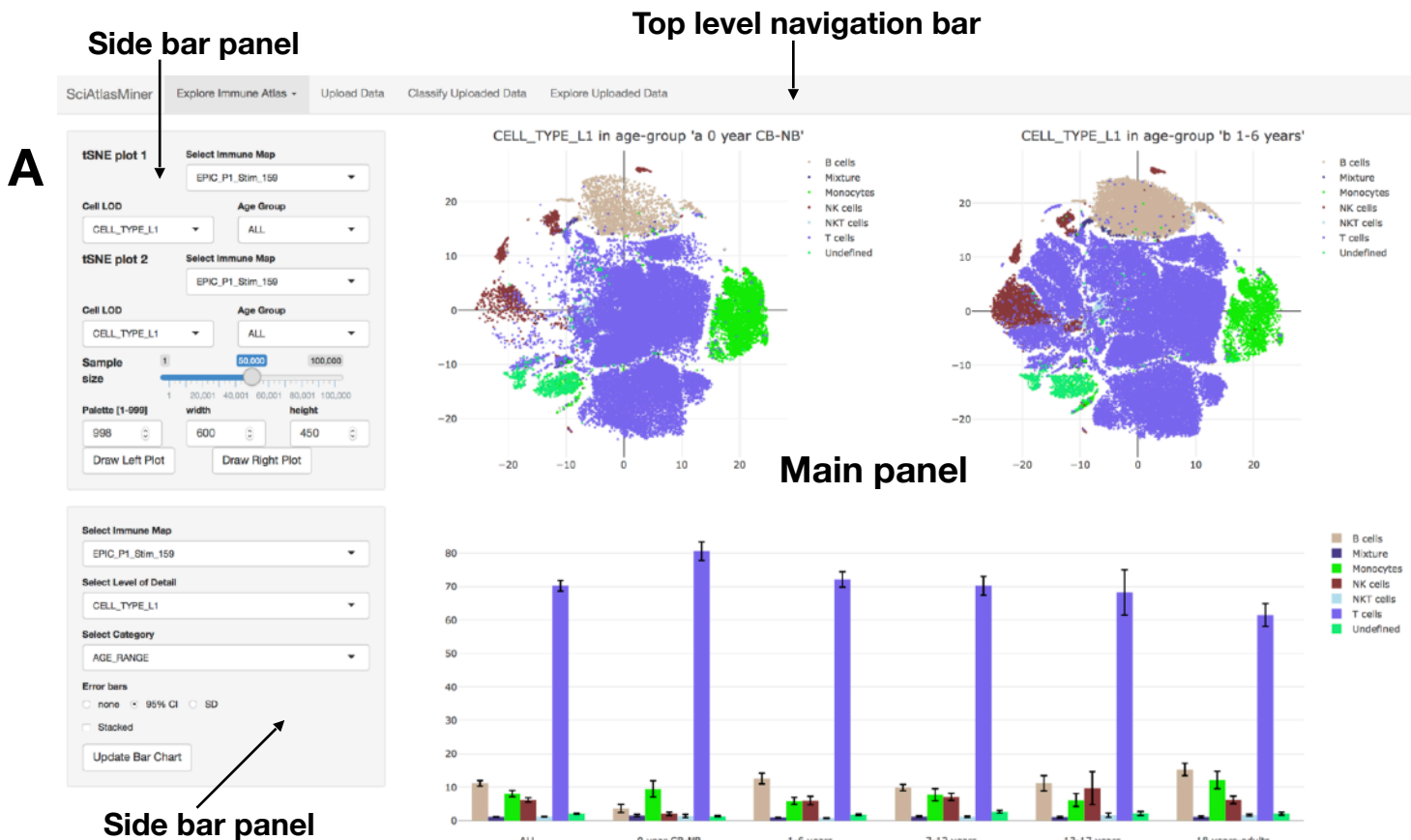


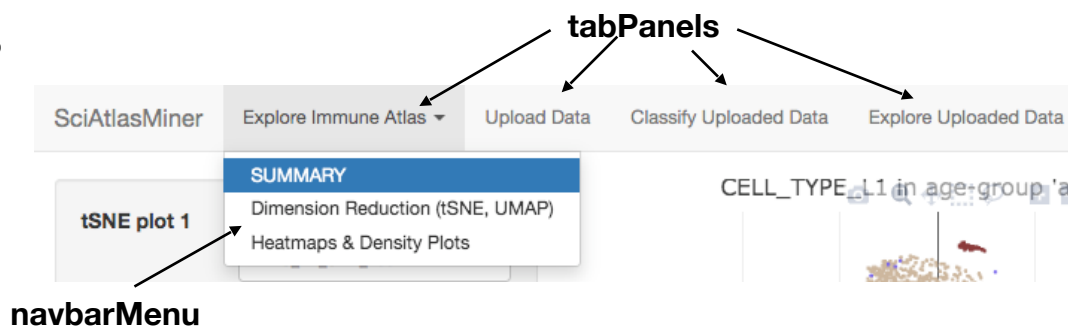
# **User Manual**

# **SciAtlasMiner**

# 1.1 SciAtlasMiner Layout & Navigation



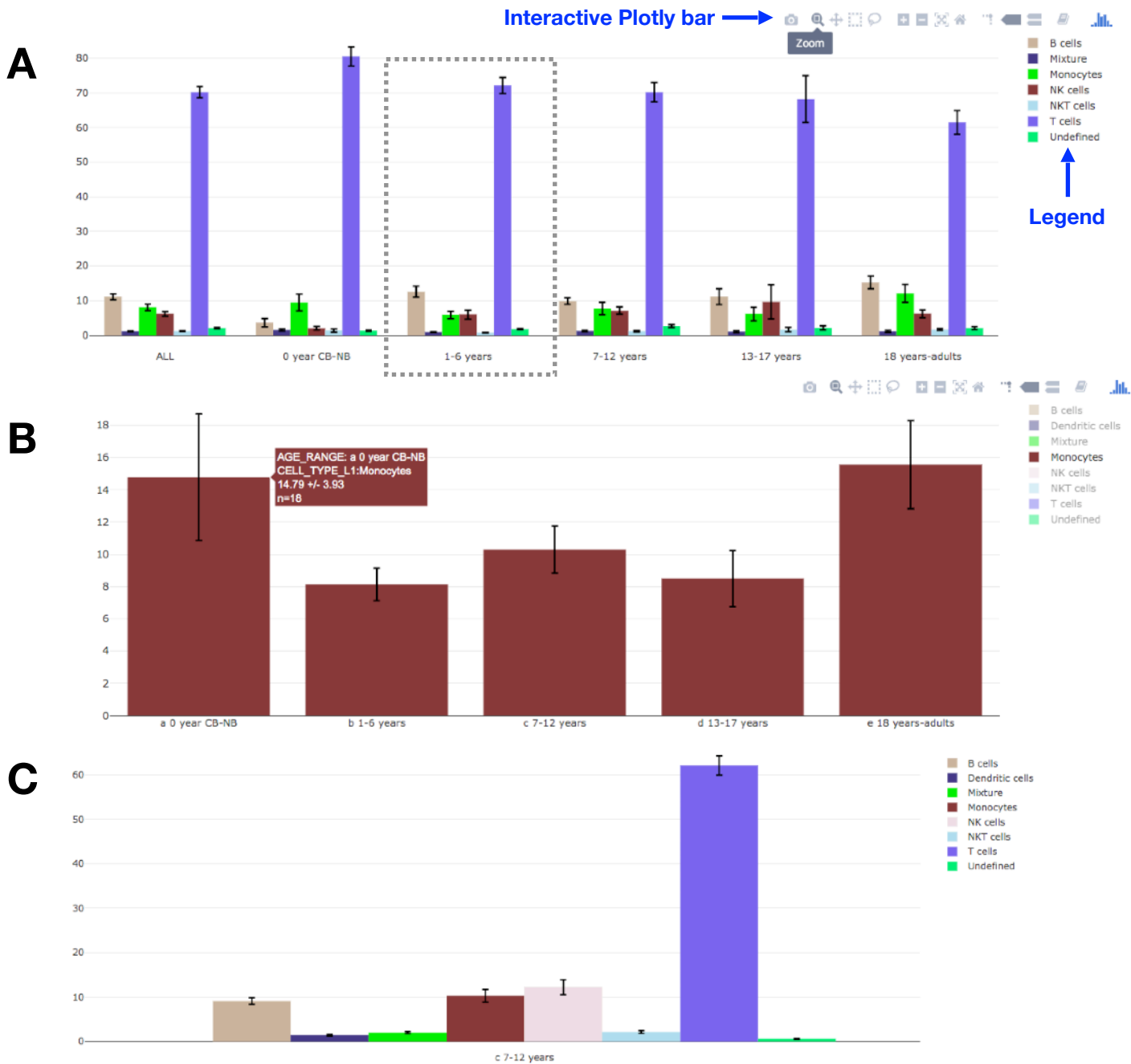
**B**



**(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



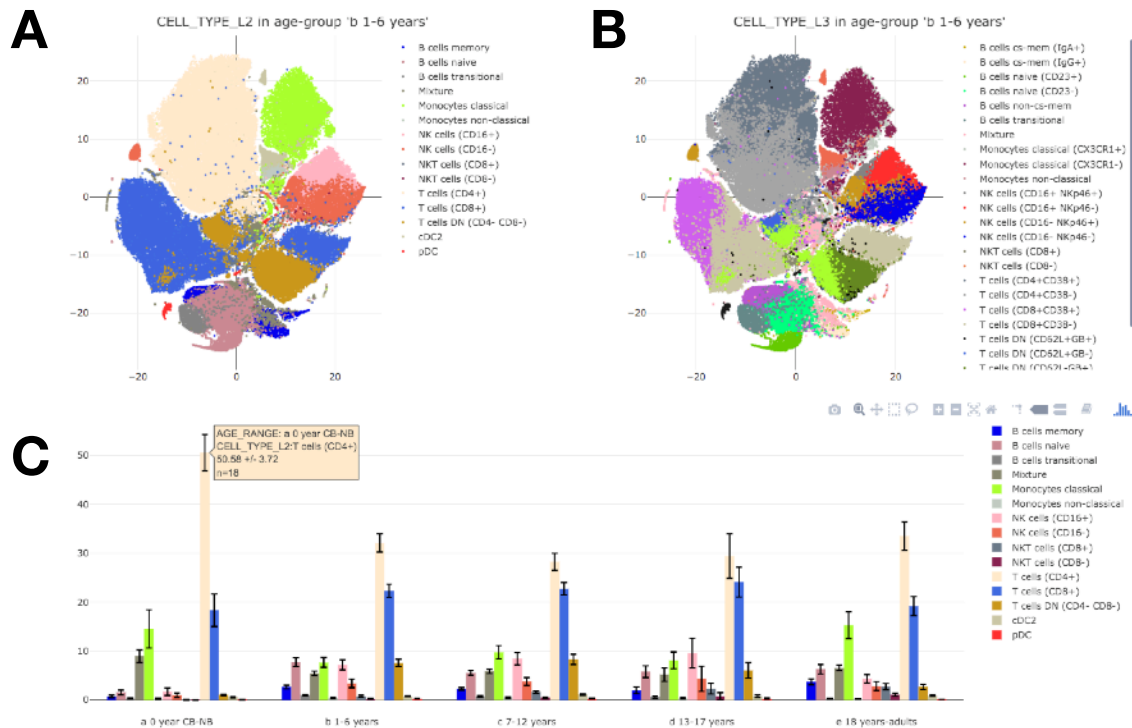
**(A) The interactive Plotly Bar** has multiple options, including (from left) download plot as PNG, zoom, pan, auto scale and reset axis.  
**Legend.** **Single mouse-click** on a legend item toggles the visibility of a single trace, while **double mouse-click** toggles the visibility of all traces.

**(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**

**tSNE plot 1** Select Immune Map  
EPIC\_P2\_NonStim\_159

Cell LOD Age Group  
CELL\_TYPE\_L2 b 1-6 years

**tSNE plot 2** Select Immune Map  
EPIC\_P2\_NonStim\_159

Cell LOD Age Group  
CELL\_TYPE\_L3 b 1-6 years

Sample size 1 91,244

Palette [1-999] width height  
655 700 500

Draw Left Plot Draw Right Plot

- **(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**

Select Immune Map  
EPIC\_P2\_NonStim\_159

Select Level of Detail  
CELL\_TYPE\_L2

Select Category  
AGE\_RANGE

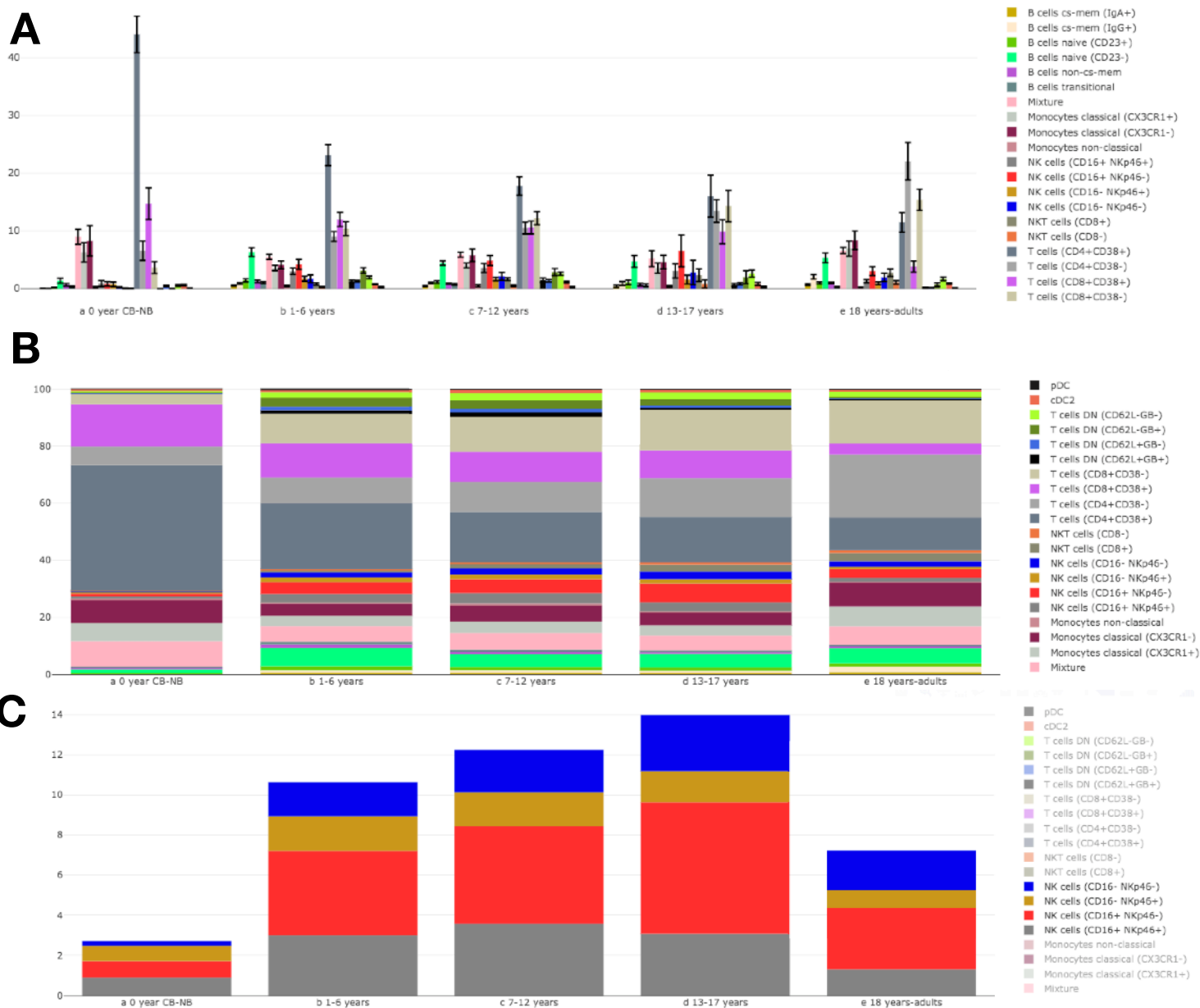
Error bars  
☐ none ☒ 95% CI ☐ SD  
☐ Stacked

Update Bar Chart

- **(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



## Summary Bar chart: Development of the Immunome



D

Select Immune Map

EPIC\_P2\_NonStim\_159

Select Level of Detail

CELL\_TYPE\_L3

Select Category

AGE\_RANGE

Error bars

☐ none

☒ 95% CI

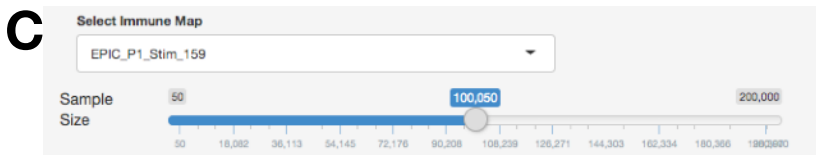
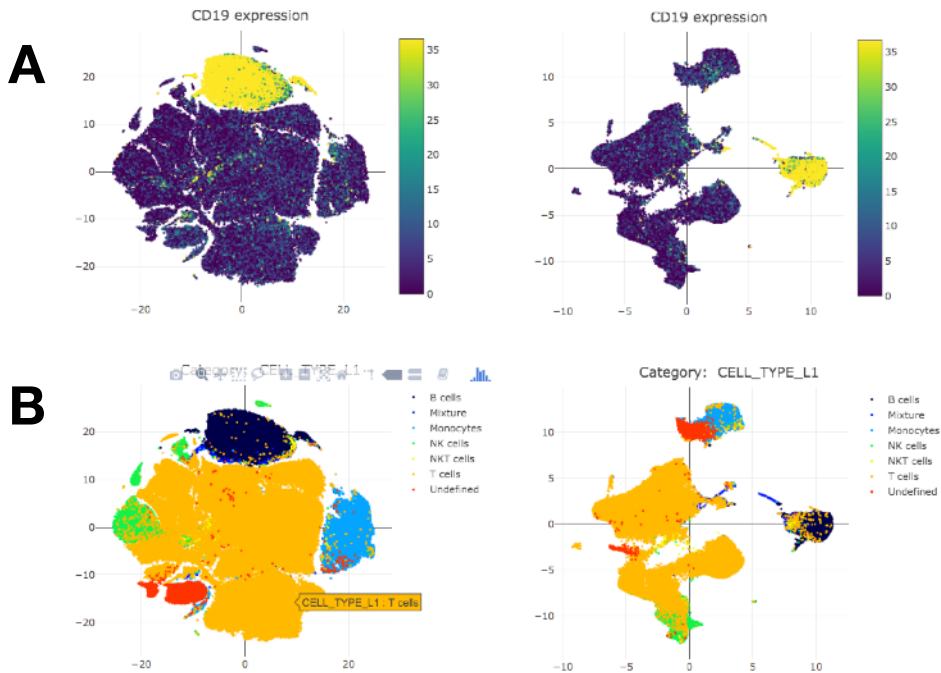
☐ SD

☐ Stacked

Update Bar Chart

- (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



**D**

**Embedding (1)**

☒ tSNE ☐ UMAP

☐ show categories

**Embedding (2)**

☐ tSNE ☒ UMAP

☐ show categories

**Color** Viridis **Marker Size** 3

**Cell Label**

CD19

FILE

**ColorQ** 90 **width** 500 **height** 400

Draw Left Plot Draw Right Plot

E

Category Embedding (3)

☐ tSNE ☐ UMAP

Category

CELL\_TYPE\_L1

Category Embedding (4)

☐ tSNE ☒ UMAP

Category

CELL\_TYPE\_L1

Color

Rich

Marker Size

3

width

550

height

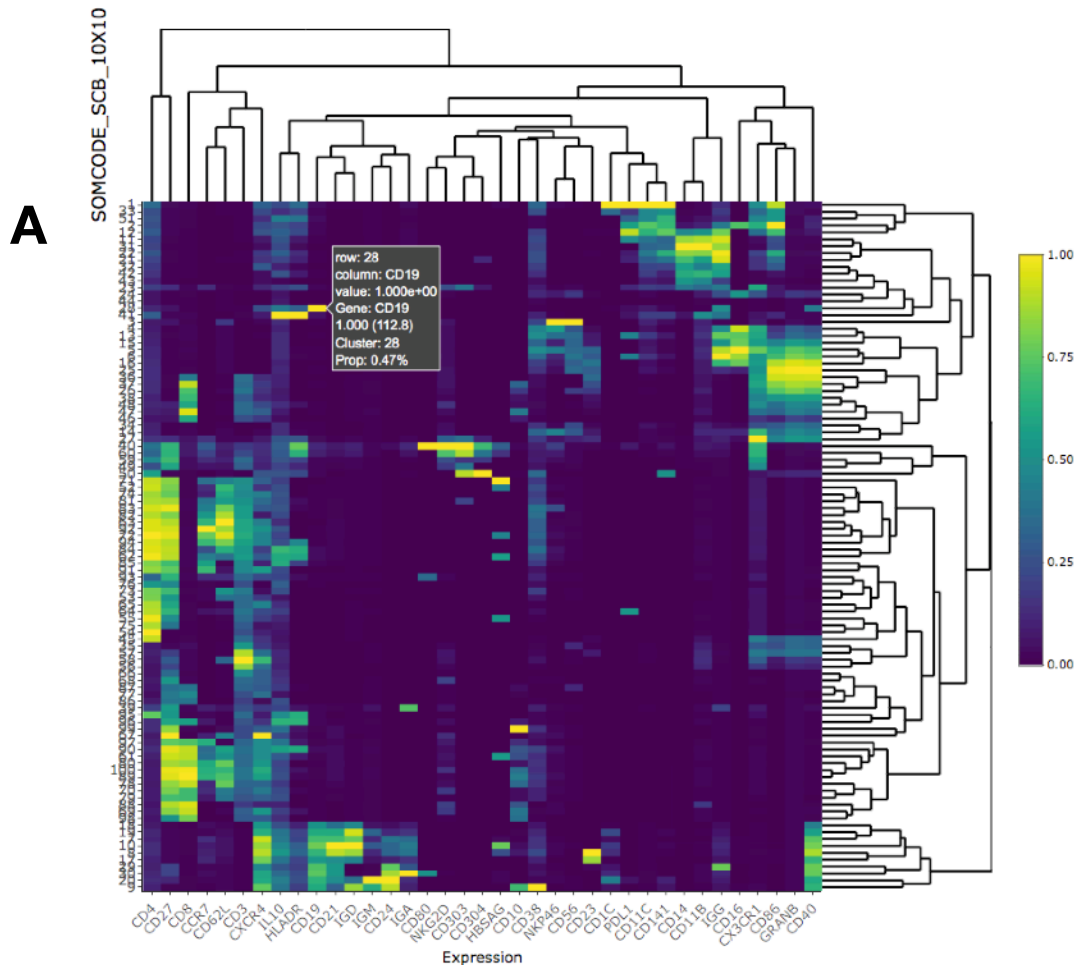
400

Draw Left Plot

Draw Right Plot

- 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



**B**

**HEATMAP**

Select Immune Map  
EPIC\_P2\_NonStim\_159

Category  
SOMCODE\_SCB\_10X10

Color Palette  
viridis

Scale  
linear

**DENDROGRAM**

☒ Columns ☒ Rows

Width  
1000

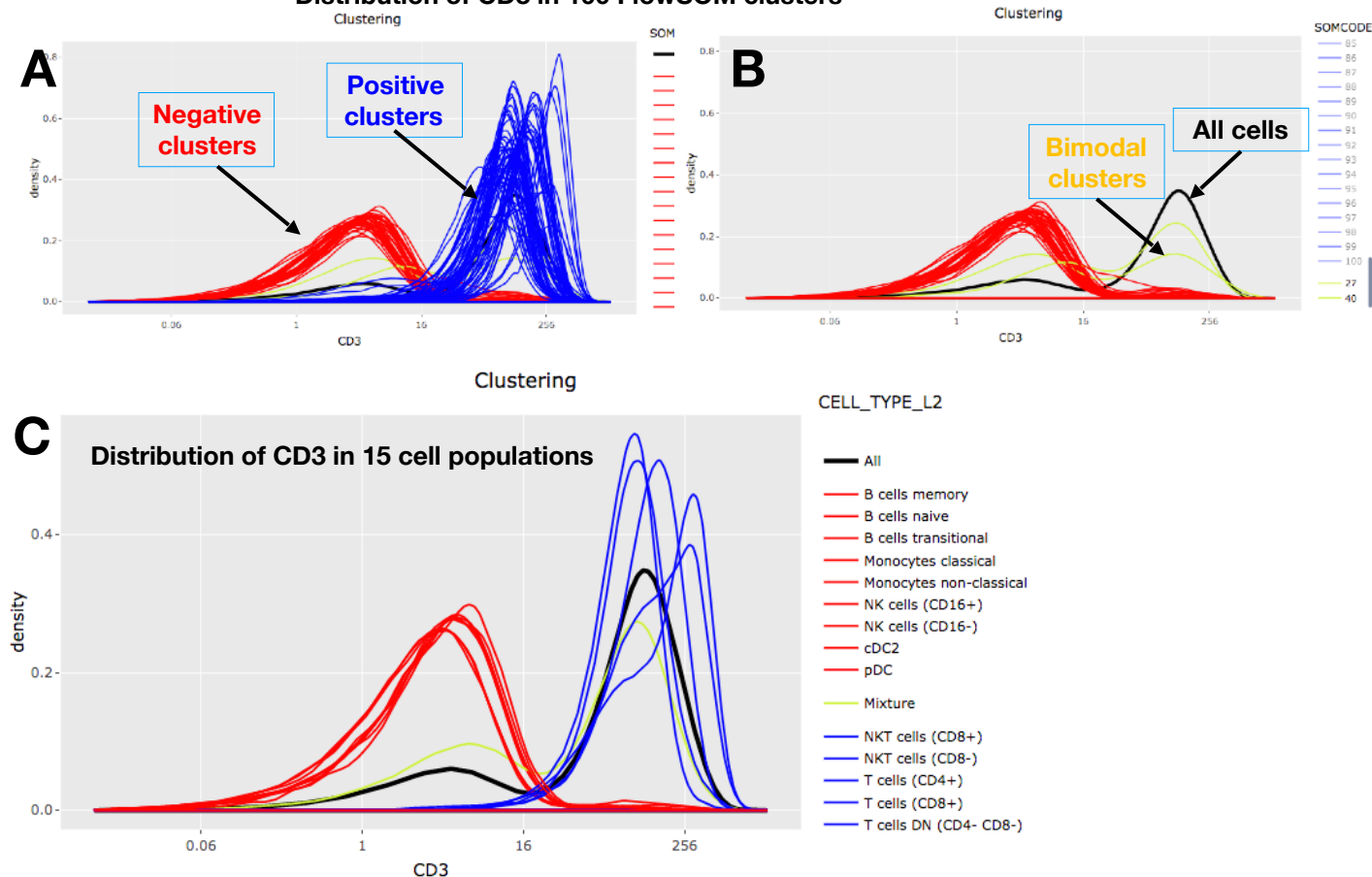
Height  
1000

Dendrogram Heatmap

- 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

Distribution of CD3 in 100 FlowSOM clusters



D

**HEATMAP**

Select Immune Map  
 EPIC\_P2\_NonStim\_159

Category  
 SOMCODE\_SCB\_10x10

Color Palette  
 viridis

Scale  
 linear

**DENDROGRAM**

☒ Columns ☒ Rows

Width  
 1000

Height  
 1000

**Dendrogram Heatmap**

- 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

E

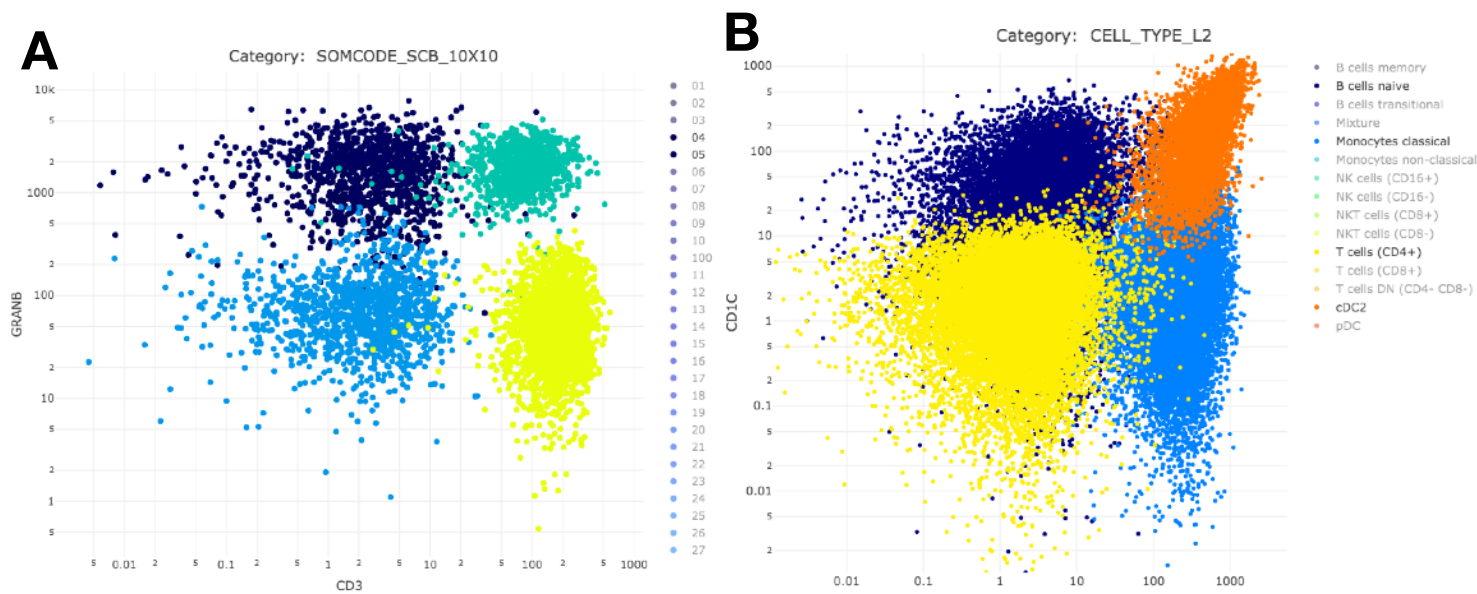
**Density Plots of Clusters**

Marker  
 CD3

☒ Cluster Classification ☒ Group Legends

**Draw 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)

**C**

**Bivariate scatter plot**

Category:

Marker X:

Marker Y:

Color:

Marker Size:

Sample Size:

- 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

SciAtlasMiner
Explore Immune Atlas
Upload Data
Classify Uploaded Data
Explore Uploaded Data

Choose CSV or FCS Files

A
Browse...
5 files
Upload complete

B
Dataset complete  
rows: 113290  
columns: 40  
Num samples: 5  
  
Matching Antibodies  
Panel1: 37 out of 37  
Panel2: 11 out of 37

C
Select File  
OO\_P1\_S01\_165\_F\_MAL\_SLEKL001\_R30\_S.fcs

D
Show 40 entries

id	Column	Role	Panel1	Panel2
1	190BCKG	Marker	~	~
2	209BI_CD16	Marker	CD16	CD16
3	112CD_CD14	Marker	CD14	CD14
4	114CD_CD14	Marker	CD14	CD14
5	161DY_CD56	Marker	CD56	CD56
6	162DY_RORY	Marker	RORY	~
7	163DY_TBET	Marker	TBET	~
8	164DY_TCRAB	Marker	TCRAB	~
9	166ER_KI67	Marker	KI67	~
10	167ER_TCRVA7.2	Marker	TCRVA7.2	~
11	168ER_IFNY	Marker	IFNY	~
12	170ER_IL8	Marker	IL8	~
13	151EU_GATA3	Marker	GATA3	~
14	153EU_CD25	Marker	CD25	~
15	155GD_CD152	Marker	CD152	~
16	156GD_CD28	Marker	CD28	~
17	157GD_CXCR5	Marker	CXCR5	~
18	158GD_CCR7	Marker	CCR7	CCR7

- 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

Select Immune Map  
EPIC\_P1\_Stim\_159

Select Uploaded File  
P2\_S05\_10years\_SLE\_NoS.fcs

Classify Uploaded Data

Select Imputed Output

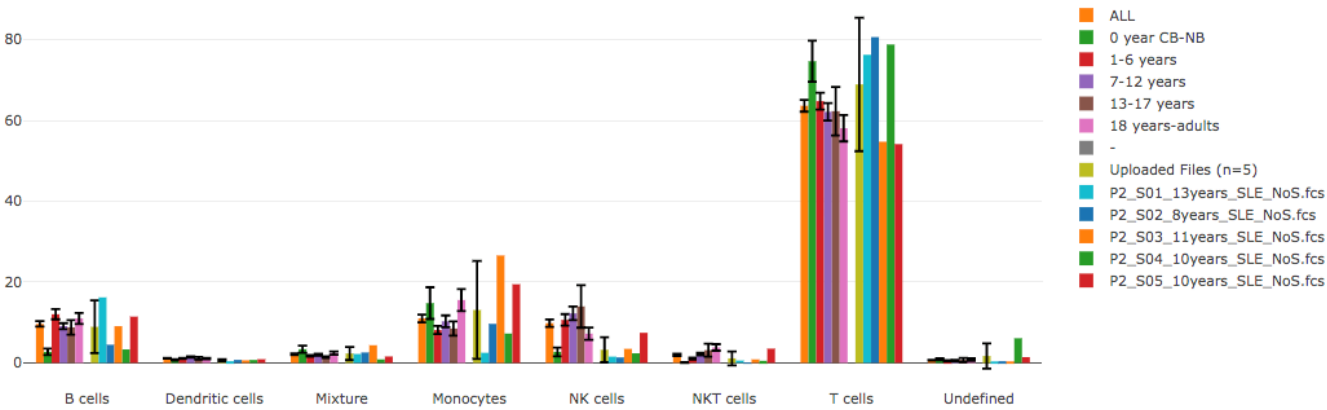
Mapping LOD  
CELL\_TYPE\_L1

Download Predictions

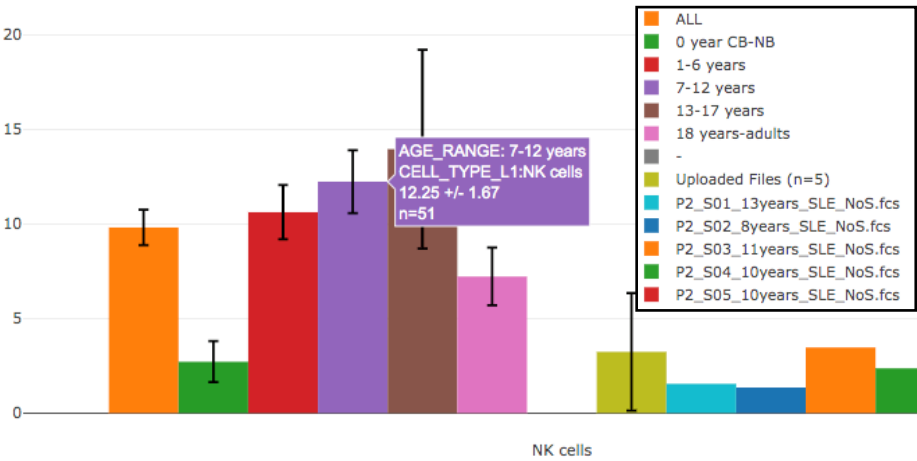
Plot Bar Chart

- 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 immune 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

Antibody Mismatch!

26 markers of the uploaded sample were not found in the immune map:  
CD62L, CD38, CX3CR1, CD24, CD80, IGA, NKX2D, CXCR4, CD303, IGG, PDL1, CD11C, CD141, IGM, CD11B, CD10, IGD, HBSAG, CD1C, CD21, NKP46, CD27, CD304, CD23, CD86, CD40

Please try a different Immune Map!

Dismiss

## 4.2 Classify Uploaded Data (Validation)

A

Download Predictions

Download Frequencies

Mapping LOD

CELL\_TYPE\_L2

Plot Bar Chart

Heatmap Category

CELL\_TYPE\_L2

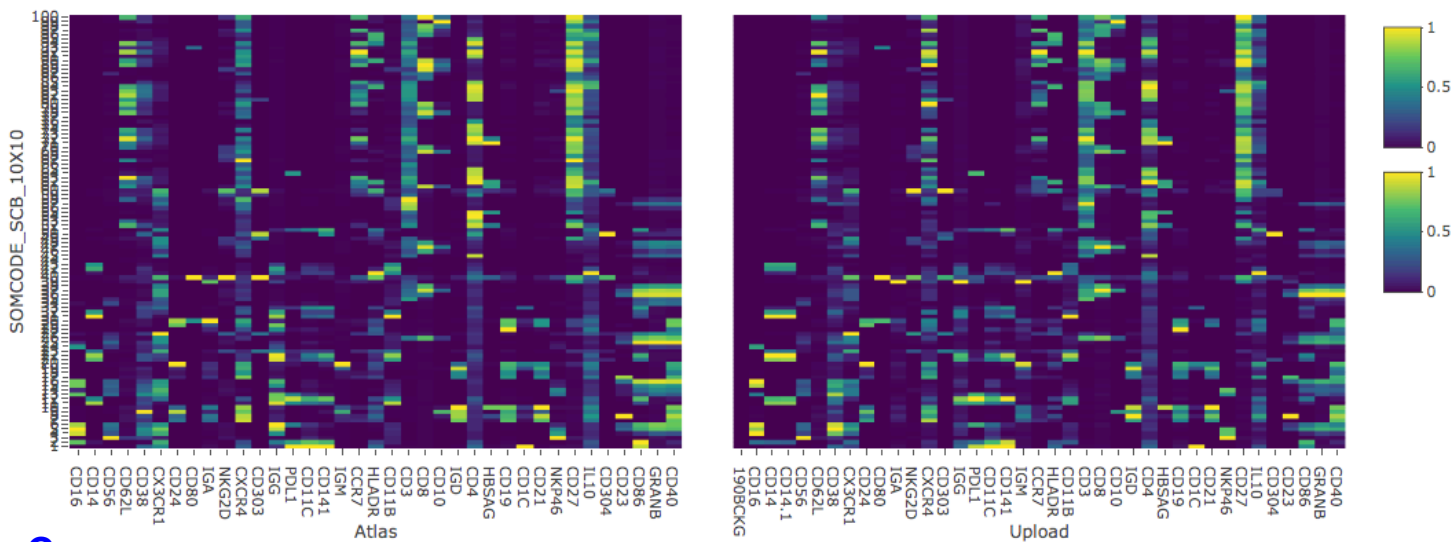
Color Scale

Viridis

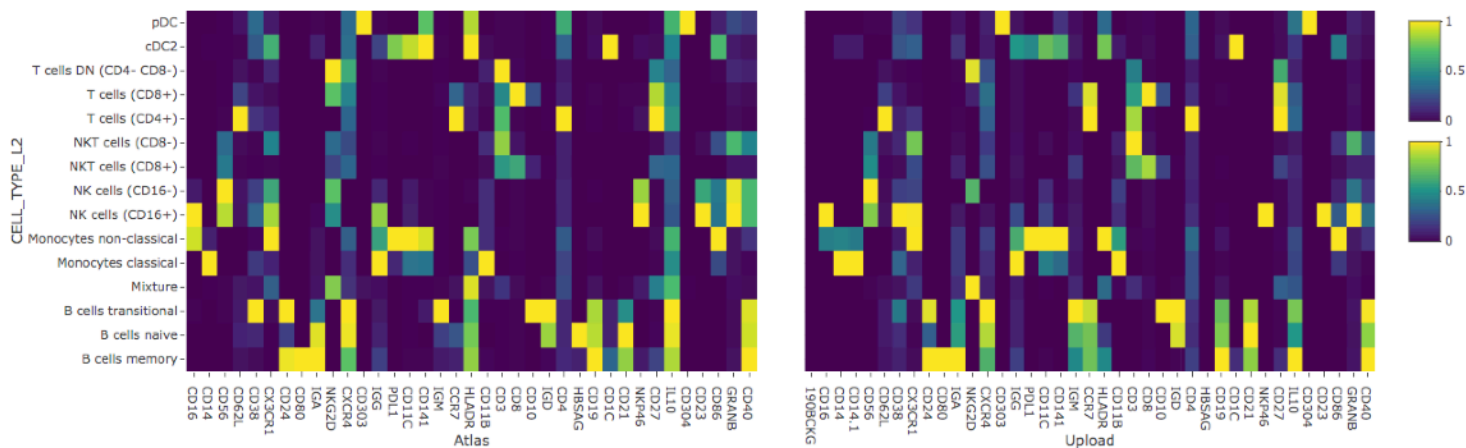
Update Heatmap

- **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**



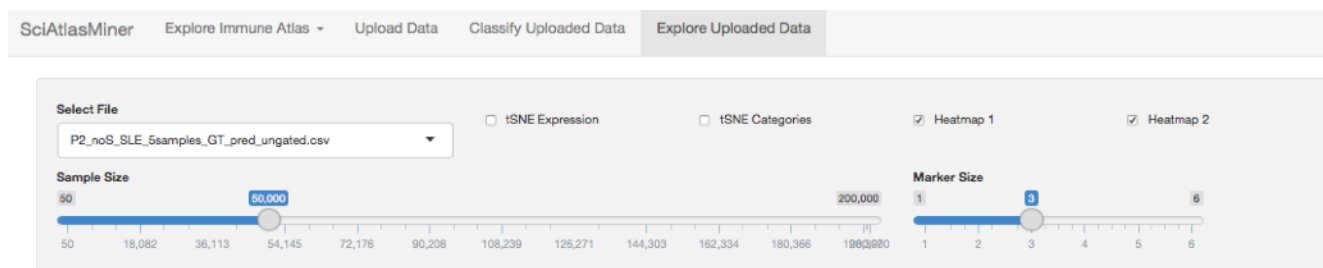
C



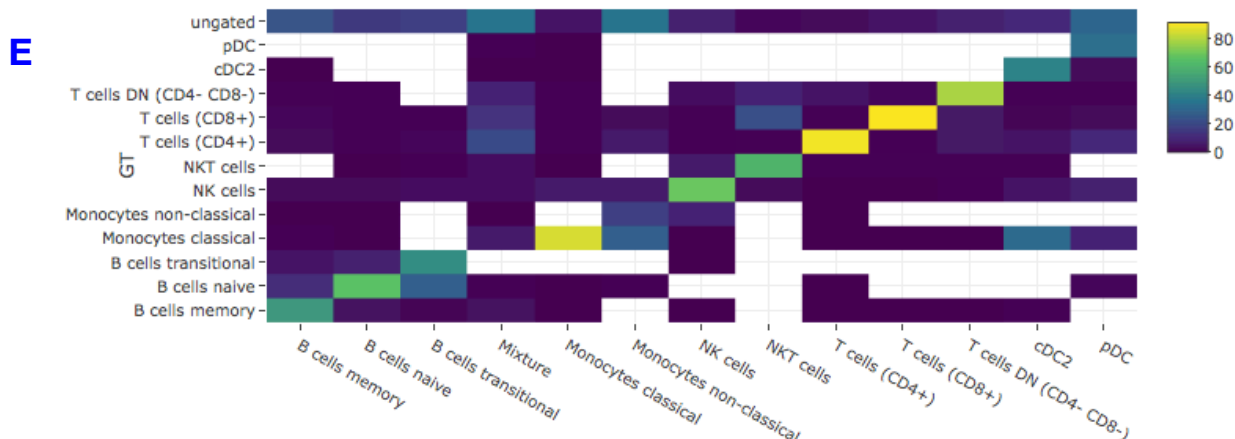
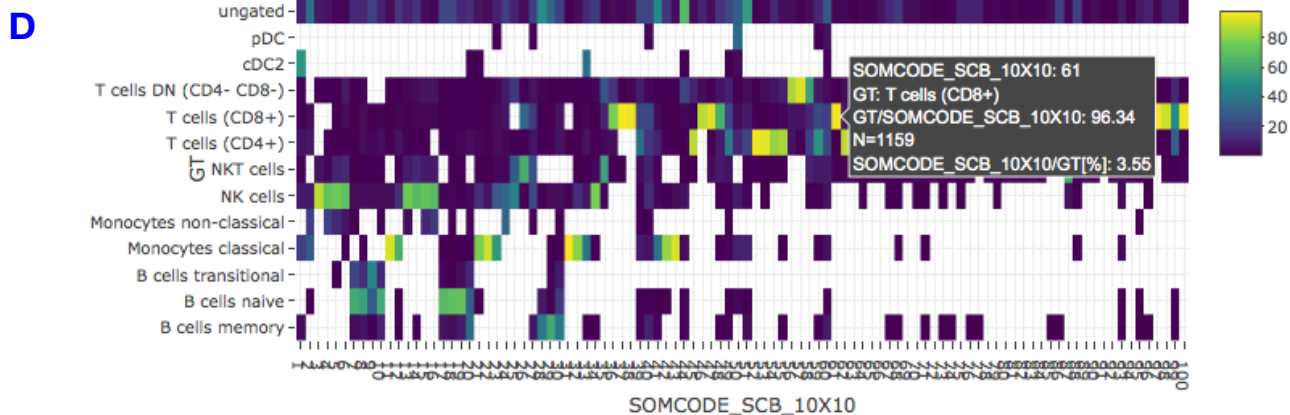
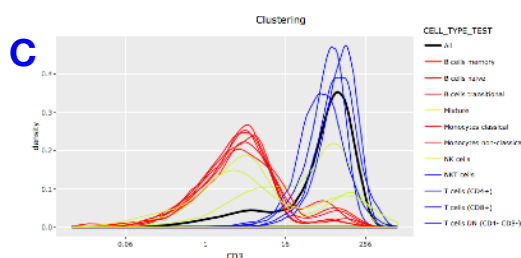
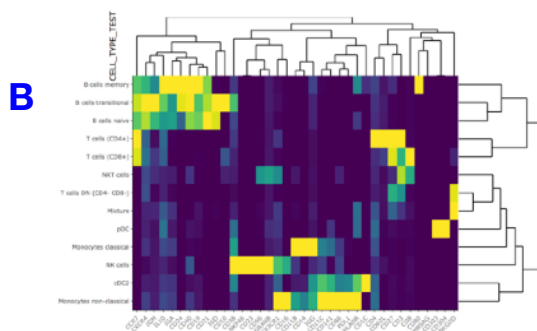


# 5. Explore Uploaded Data

A



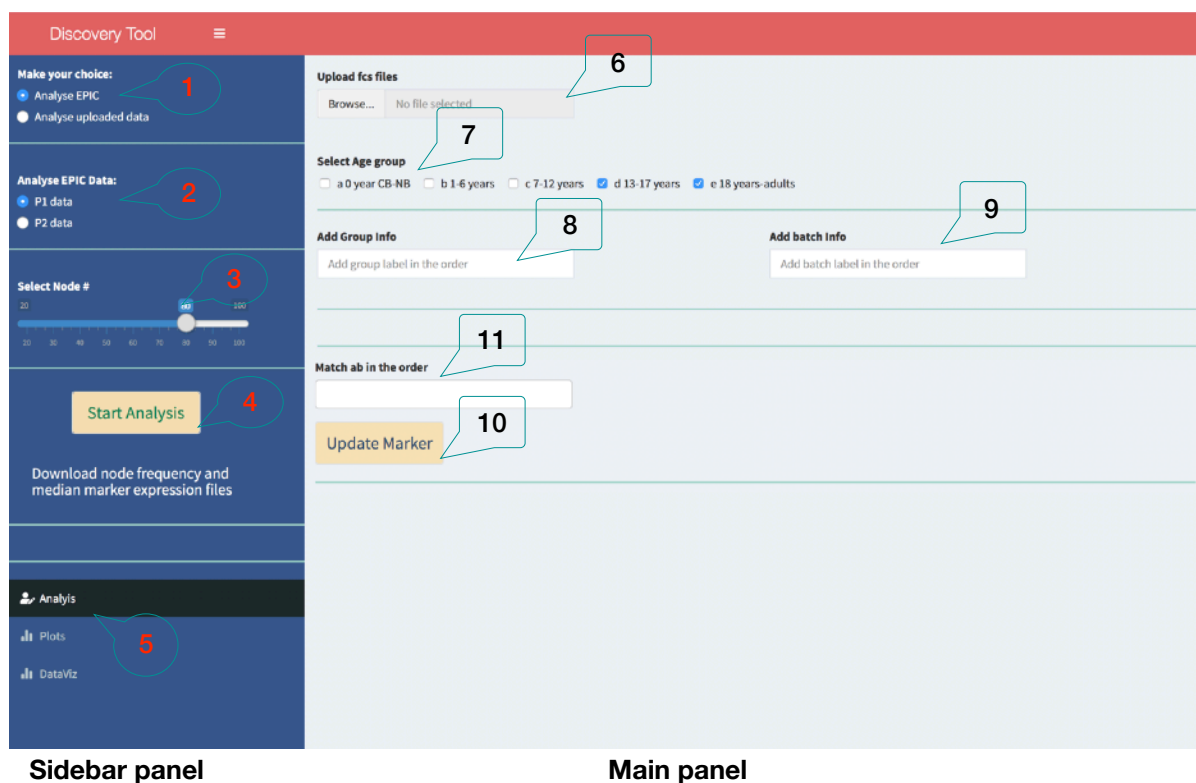
- This module is suitable for CSV files containing Cytometry expression data along with processed data (e.g. t-SNE) and/or categorical columns. FCS files do not contain sufficient information
- Upload CSV file as outlined in section 3 (page 10)
- In the top navigation bar, select 'Explore Uploaded Data' 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.



# **User Manual**

# **Discovery Tool**

## Discovery Tool : User Interface (UI) layout



The application UI is divided 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

Discovery Tool

Make your choice:

- Analysis EPIC
- Analyze uploaded data

Analyze EPIC Data:

- P1 data
- P2 data

Select Node #

Start Analysis

Download node frequency and median marker expression files

Analysis

Plots

DataViz

Upload fcs files

Browse... 10 files

Select Age group

a 0 year CB-NB b 1-6 years c 7-12 years d 13-17 years e 18 years-adults

Add Group info

T0,T3

Add batch info

B1,B1,B1,B1,B1,B1,B2,B2,B2

Show 10 entries

Filename
1 Set1_LJX031_T0_R_E2_NS.fcs
2 Set1_LJX031_T3_R_E2_NS.fcs
3 Set1_LYT032_T0_R_E2_NS.fcs
4 Set1_LYT032_T3_R_E2_NS.fcs
5 Set1_PLX012_T0_R_E2_NS.fcs
6 Set1_PLX012_T3_R_E2_NS.fcs
7 Set2_LYQ042_T0_R_E2_NS.fcs
8 Set2_LYQ042_T3_R_E2_NS.fcs
9 Set2_LZF072_T0_R_E2_NS.fcs
10 Set2_LZF072_T3_R_E2_NS.fcs

Show 10 entries

FILE	AGE_RANGE	BATCH
1 Set1_LJX031_T0_R_E2_NS.fcs	T0	B1
2 Set1_LJX031_T3_R_E2_NS.fcs	T3	B1
3 Set1_LYT032_T0_R_E2_NS.fcs	T0	B1
4 Set1_LYT032_T3_R_E2_NS.fcs	T3	B1
5 Set1_PLX012_T0_R_E2_NS.fcs	T0	B1
6 Set1_PLX012_T3_R_E2_NS.fcs	T3	B1
7 Set2_LYQ042_T0_R_E2_NS.fcs	T0	B2
8 Set2_LYQ042_T3_R_E2_NS.fcs	T3	B2
9 Set2_LZF072_T0_R_E2_NS.fcs	T0	B2
10 Set2_LZF072_T3_R_E2_NS.fcs	T3	B2

Show 1 to 10 of 10 entries

Match ab in the order

Update Marker

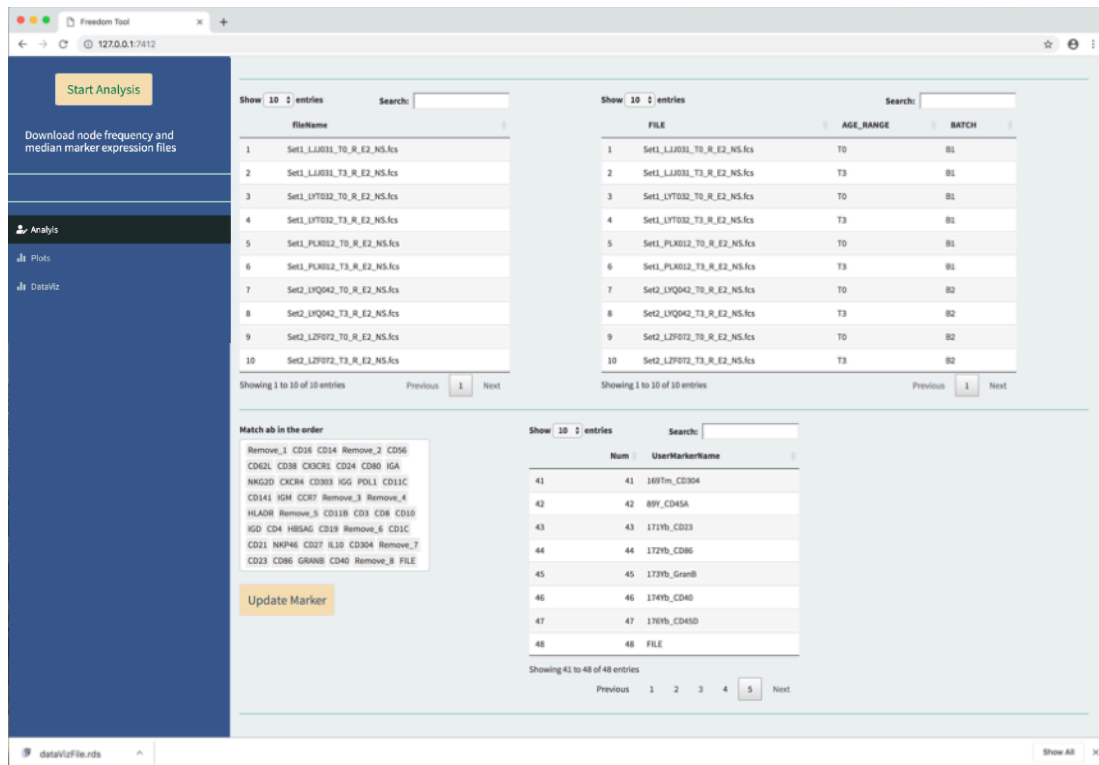
Show 10 entries

Num	UserMarkerName
1	190BCKG
2	209BL_CD16
3	112C0L_CD14

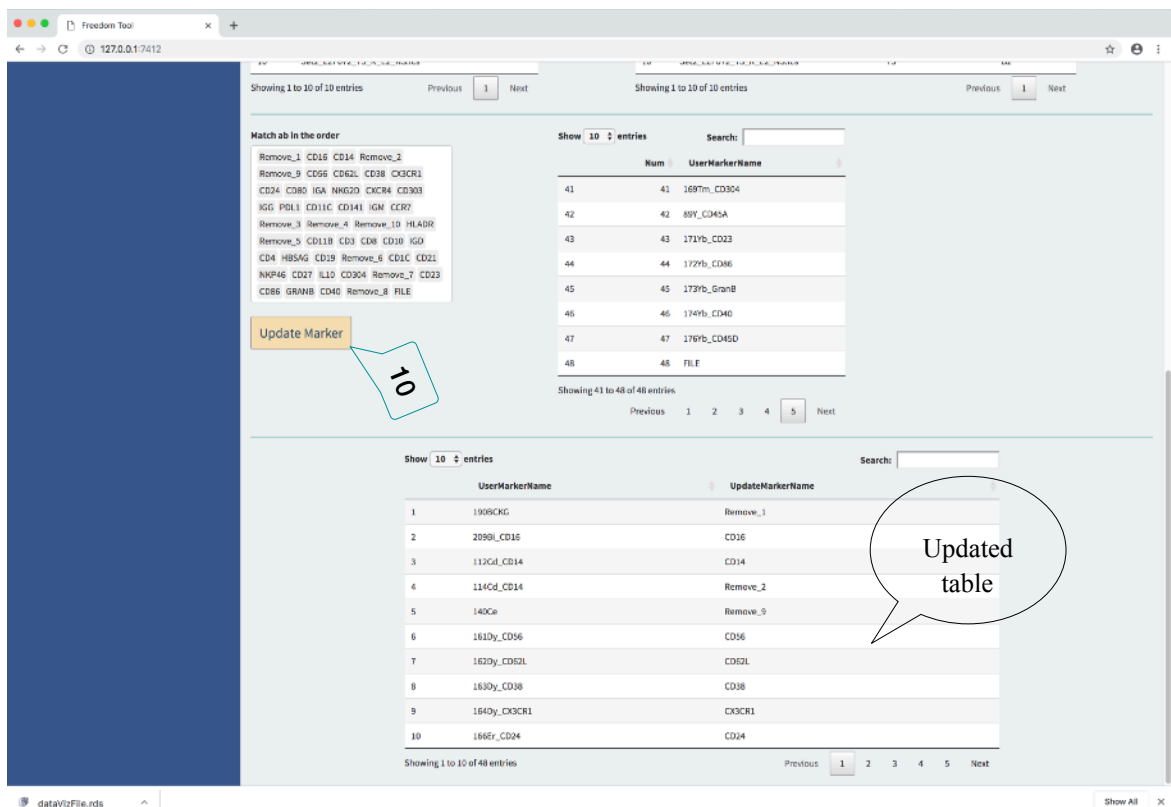
Show All

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 inputting 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).

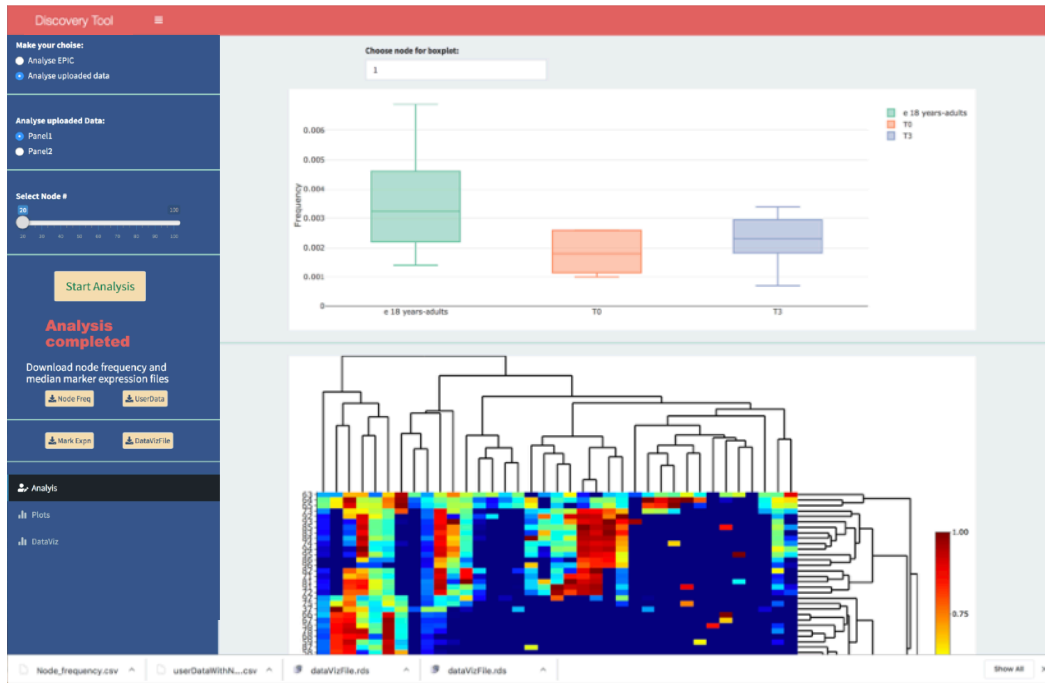
Once the analysis start a progress bar in the bottom right of the main panel will appear (PB). Wait for the analysis to completed. Once the analysis is complete progress bar will disappear and on the sidebar below the start analysis button a message “**Analysis Complete**” will be displayed. Along with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

The screenshot displays the Discovery Tool interface. The sidebar on the left includes a 'Make your choice' section with radio buttons for 'Analyse EPIC' and 'Analyse uploaded data'. Below this is 'Analyse uploaded Data' with 'Panel1' selected. A 'Select Node #' slider is set to 20. The 'Start Analysis' button is highlighted with a red box and labeled '3'. Below the sidebar, a message 'Analysis completed' is displayed, along with four download buttons: 'Node Freq', 'UserData', 'Mark Expn', and 'DataVizFile'. The main panel shows the 'Upload fcs files' section with a 'Browse...' button and '8 files' listed. The 'Select Age group' section has radio buttons for age ranges, with 'd 13-17 years' and 'e 18 years-adults' selected. The 'Add Group Info' section shows 'A,B' and the 'Add batch Info' section shows 'B1'. Below these are two tables of file entries. The first table lists file names, and the second table lists file names, age ranges, and batch information. A progress bar (PB) is visible in the bottom right corner, and a message 'Analysis started' is shown in a box at the bottom right.

### 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 format 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

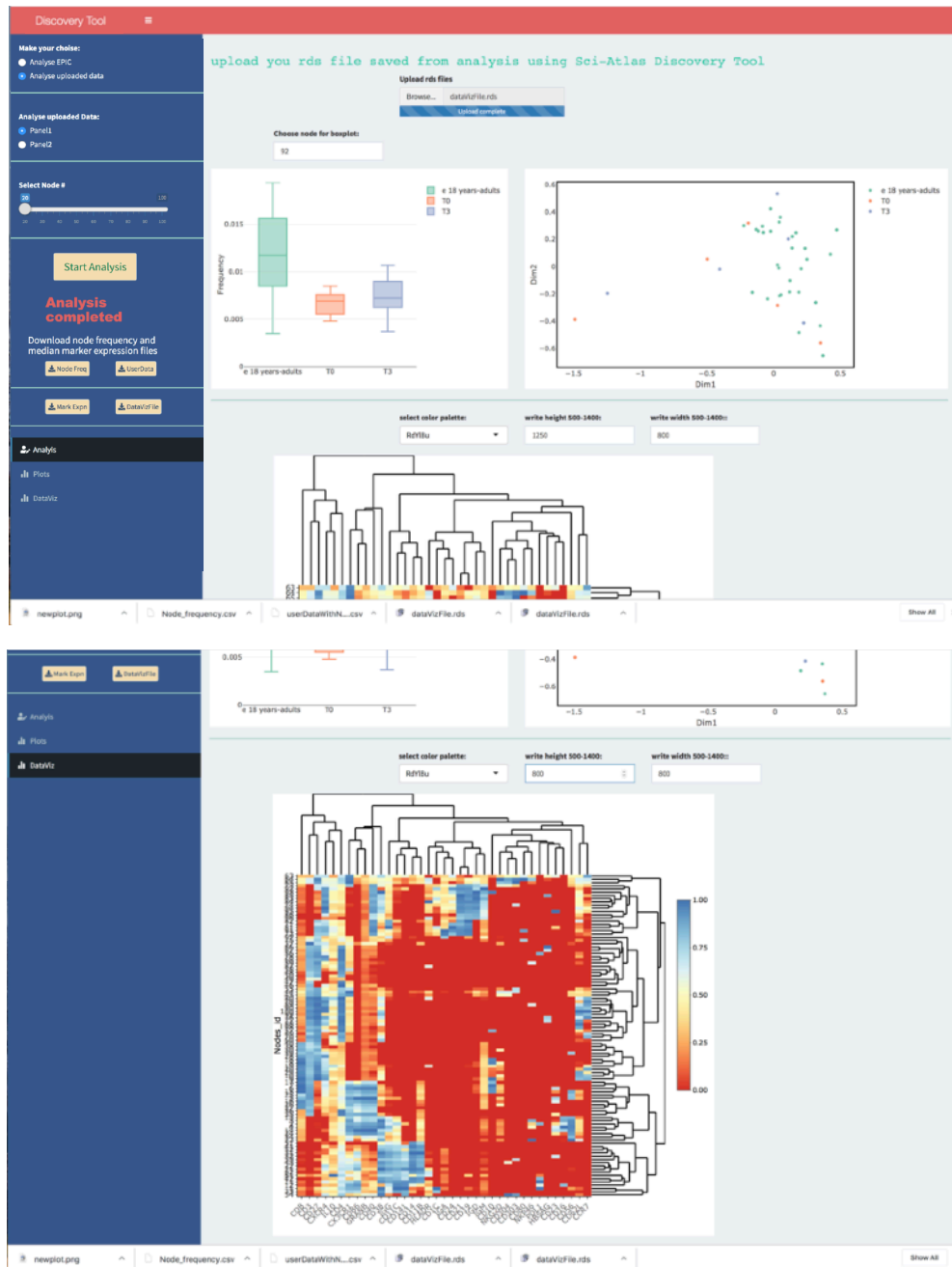


Once the analysis is completed, user can click the Plot tab to switch from Analysis tab.

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.