Quickstart guide

Cytosplore Transcriptomics

TUDelft   LUMC Leiden University Medical Center   NWO | Applied and Engineering Sciences

Table of Contents

Introduction

Cytosplore Transcriptomics is a stand-alonesoftware application for fast and interactive exploration of cell types from single cell sequencing data. It scale to millions of cells at still interactive speeds, and runs on PC (windows 10) and linux (Ubuntu). Cytosplore Transcriptomics was developed in close collaboration between the [Division of Image Processing](#) , the [Computational Biology Center](#) at [Leiden University Medical Center](#), and the [Computer Graphics and Visualization Group](#) at the [TU Delft](#).

Using Cytosplore Transcriptomics, you can:
- Perform interactive and very fast GPGPU-tSNE analyses for smaller dataset (< 300k cells)
- Perform interactive and very fast HSNE visual inspection of clusterings for large datasets (> 300k cells – 10 million cells)
- Perform mean-shift clustering on HSNE maps on the fly
- Enables manual correction of clustering results and labels for large amounts of cells with fast and easy user interactions.
- Explore transcriptome-wide gene expression in combination with metadata of individual cells.
- Perform differential analyses and statistics between manual selections of cells, or between pre-clusters throughout the cellular hierarchy.
- Visualize the results of differential analyses.
- Export the results of differential analyses for further analysis and visualization in external tools.

System requirements

The data size and computational demands of interactive GPGPU-tSNE and HSNE require sufficiently modern hardware to ensure the desired responsiveness in Cytosplore Transcriptomics. We recommend the following:
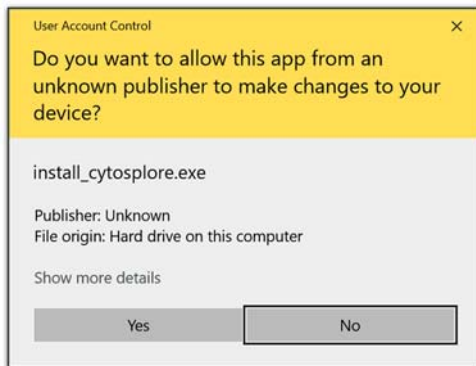- Windows 7 or higher (64bits) or linux ubuntu
- Intel Core i7 with 2 cores or better (preferably generation 6 or higher)
- 8GB of memory or more
- Dedicated Graphics Card (preferably nVidia GTX 960 or better) with at least 2GB of GPU memory

Configurations with lesser specifications may work, but will result in longer wait times for data loading and computing. We have tested Cytosplore on different configurations, including Windows 7, Windows 10 and linux Ubuntu, intel Core i7 and intel Xeon CPUs as well as different Graphics Cards (nvidia GTX 980, GTX 1080, GTX Titan, AMD Radeon r9 M290, and AMD Radeon Pro 460). It is highly recommended to update the drivers of the graphics card to the latest versions prior to installing Cytosplore Transcriptomics. Remote operation of Cytosplore Transcriptomics through Remote Desktop is not supported.
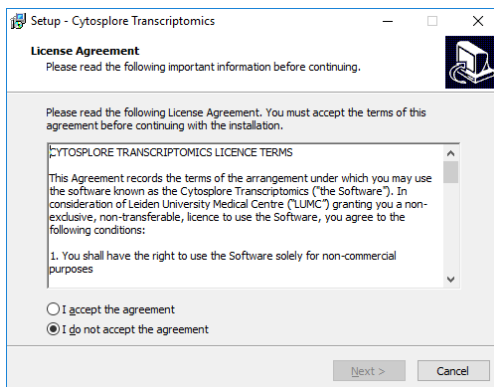
Installation and updates

The installer will ask for administrator rights on the computer. This is due to the fact that we include the Microsoft Visual C++ Redistributable Packages. Cytosplore Transcriptomics itself does not need or makes use of these rights. To continue the installation click *Yes*. You will be prompted to enter the administrator password.
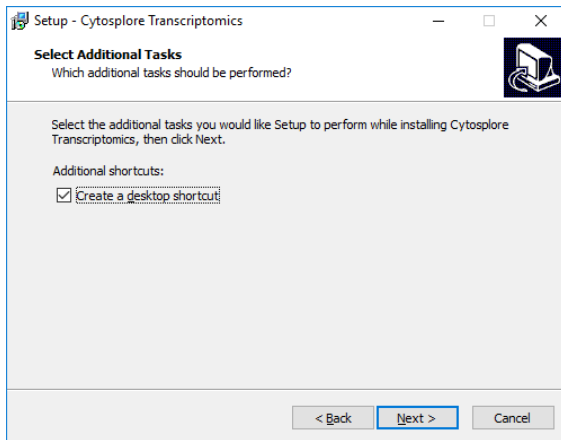
Click *Yes*, to continue the installation.

The installer will show the license agreement, that you have to accept. Most importantly it states that the software can be used free of charge.
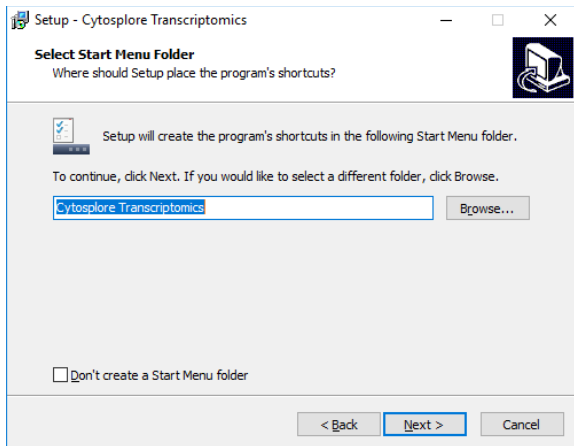
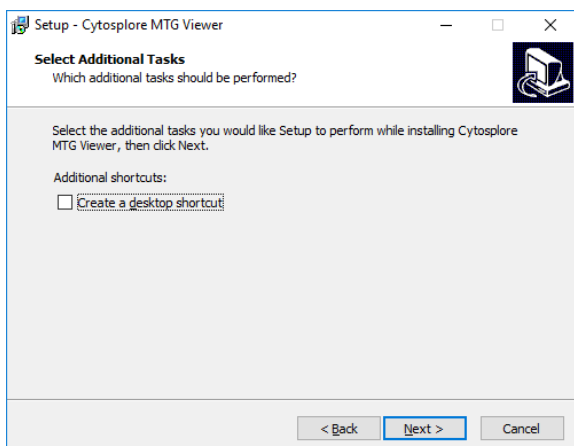Click *I accept the agreement* and then *Next >*.

Next you can define the location where Cytosplore Transcriptomics should be installed or simply click *Next >*.
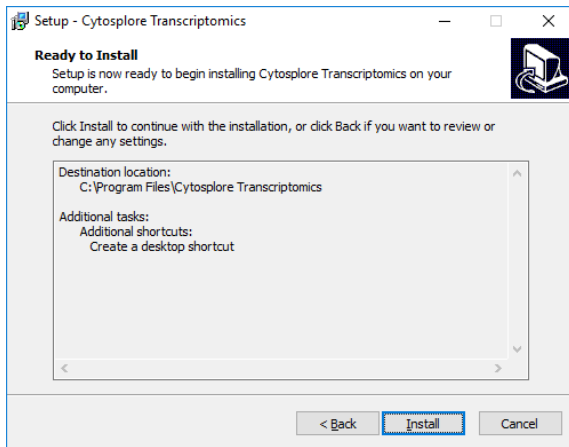
In the next dialog you can define the name of or decide not to create a Cytosplore Transcriptomics start menu folder. Click *Next >* to continue without changes.
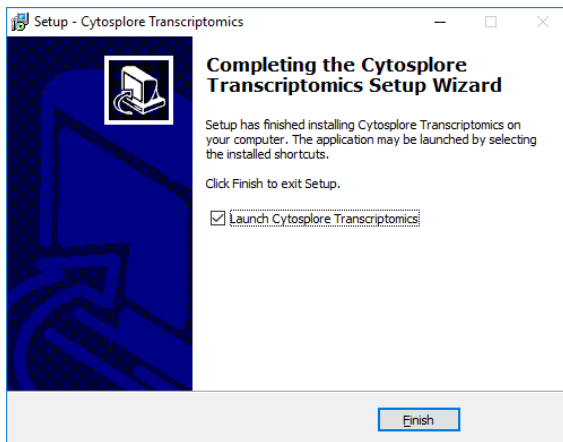

Now you will be asked if you want to create a desktop shortcut. If yes, just click *Next >* to continue.

The last dialog before putting the files in the appropriate locations shows a summary of the previous settings. Click *Install* to prepare the Cytosplore Transcriptomics installation.
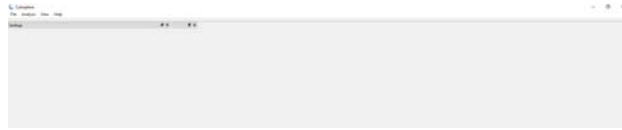
After a successful installation the following dialog will show. You can decide whether to launch Cytosplore Transcriptomics now, and click *Finish* to complete the process.
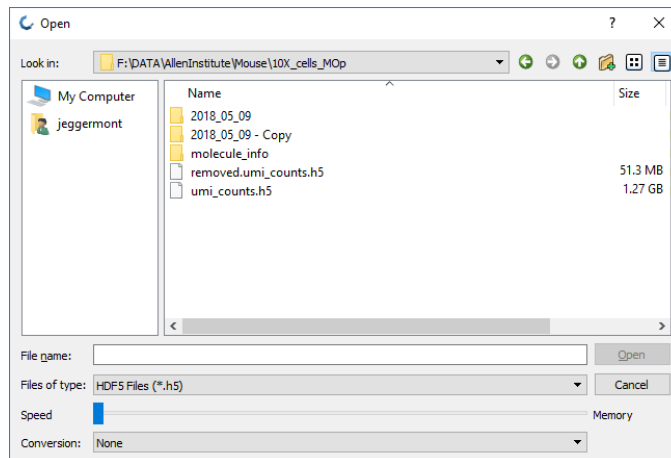
Getting Started

Up on starting Cytosplore Transcriptomics you see the following mostly empty window

To start an analysis go to File menu and select either of the options:
- Open CSV File: Open a comma-separated vales file
- Open HDF5 File: Open a 10x compliant HDF5 file.
- Open Metadata File: Open metadata per cell for the loaded dataset.
- Open H5AD data: Reads hdf5 files generated by Scanpy
- Open Matrix Market File
- Open TOME files: Reads hdf files generated by the Allen Institute
-

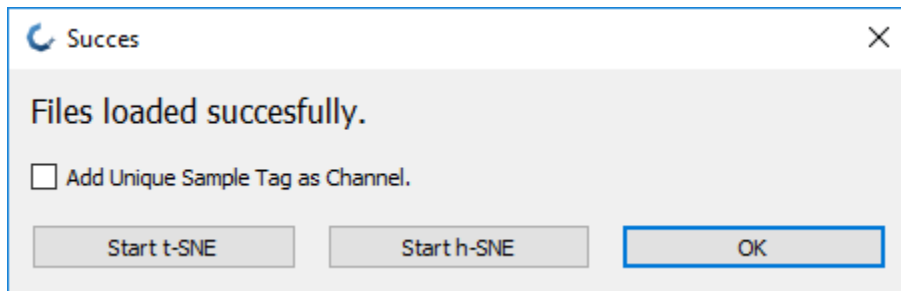For optimal data loading performance we recommend to use HDF5 Files for file-io.

When opening a TenX compliant HDF5 file there are 2 additional options in the File-dialog apart from selecting the file:
- Speed: This slider allows for balancing between speed and memory usage when loading the data which can be useful for very large datasets.
- Conversion: This allows you to select a standard conversion to be performed on the datavalues
    - None: No conversion
    - Log: All values are converted using log2(1+value)
    - Sqrt: All values are converted to their square-root value
    - Arcsin: All values are converted to arcsinh(value/5);

There is also a check box to normalize count matrices data per cell (into CPM)
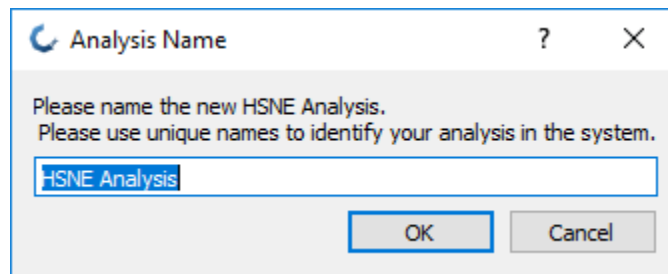
When the data is successfully loaded the following dialog will be shown to select the initial type of analysis:
- Start t-SNE : start GPGPU t-SNE analysis
- Start h-SNE: start an Hierarchical SNE analysis
- OK: Skips starting an analysis. Note that you can start analysis also from the "Analysis" menu.
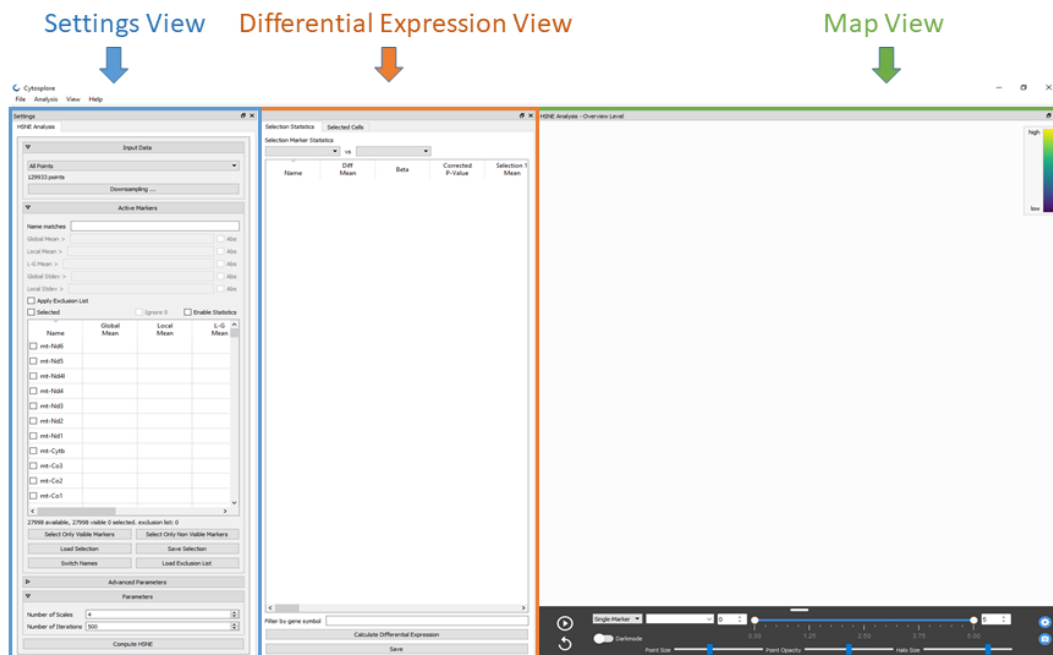

For large datasets we advise to start with an Hierarchical SNE analysis.

When starting an analysis it must be given a unique name within in the session. Typically the default name can be chosen by selecting OK.

Once the analysis has been named you will see 3 panels. From left to right these are the Settings View, Differential Expression View and the Map View (see below)
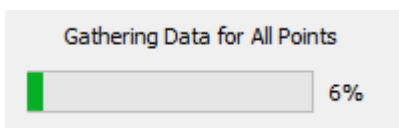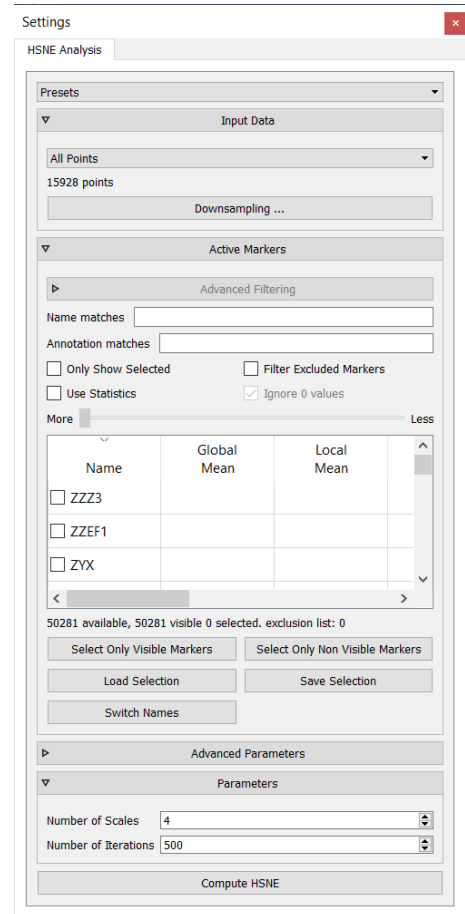
Settings View

The Data Selection panels for t-SNE and HSNE are very similar and consists of 5 parts:

- Input Data: this specifies which subset of the data will be used for SNE analysis. It also allows you to create a randomly downsampled subset of the data which pressing the Downsampling… button.
- Active Markers: This allows you to select (a subset of) the marker genes to use for the SNE analysis. See below for more detail.
- Advanced Parameters related to tSNE / HSNE settings
- Parameters: Standard parameters such as the number of iterations and for HSNE the number of scales to compute.
- Button to start either HSNE or t-SNE computation

Marker Selection

To allow for efficient marker gene selection for the SNE analysis there are several options.

1. Individual marker genes can be selected by checking the box in front of the gene symbol.
2. A selection of marker genes can be loaded from (and saved to) a .txt file which is a simple text file with 1 gene name per line.
3. An exclusion gene list can be loaded in the file menu for exclude for instance mitochondrial or X-and Y chromosome related genes; Checking the check box "Filter Excluded Markers" removes these markers from the statistics widget.
4. Genes can be selected based on statistical gene features. In order to do this check the "Enable Statistics" Option. The first time this is checked it will trigger computation of the statistics, which may take some time in large datasets.
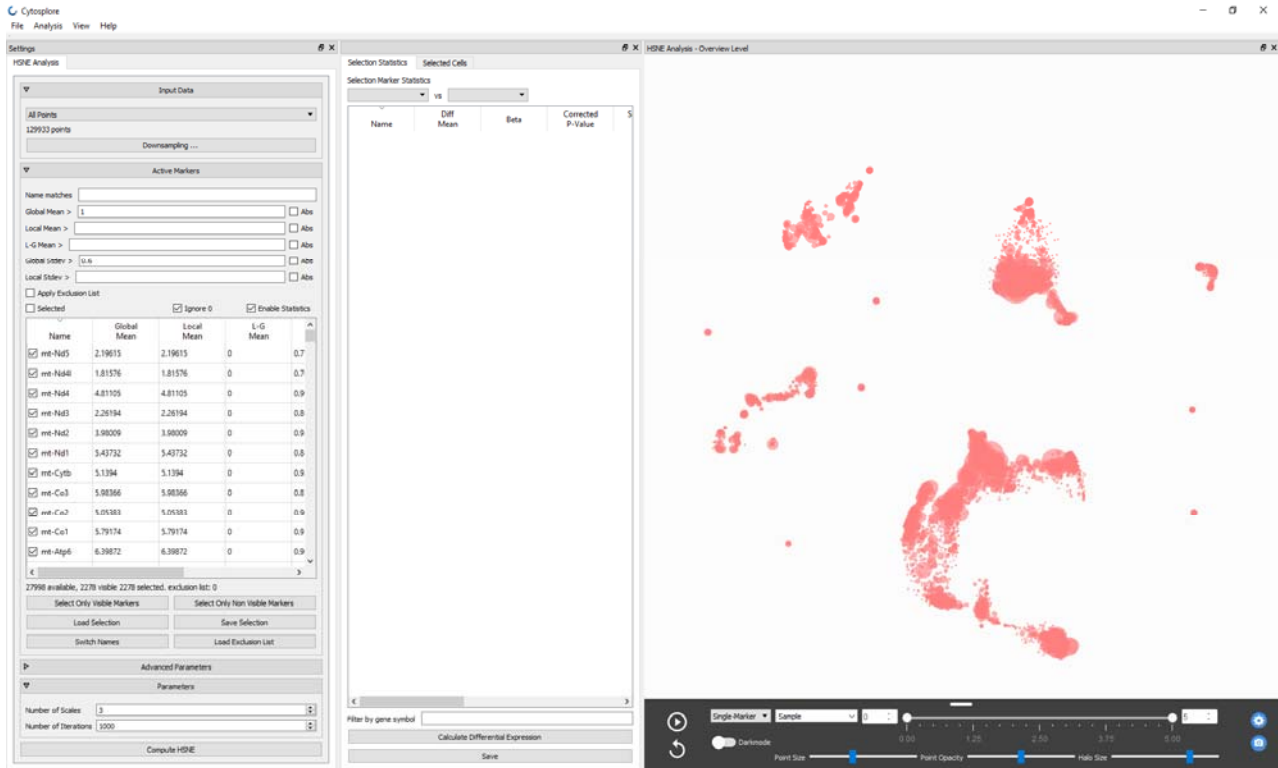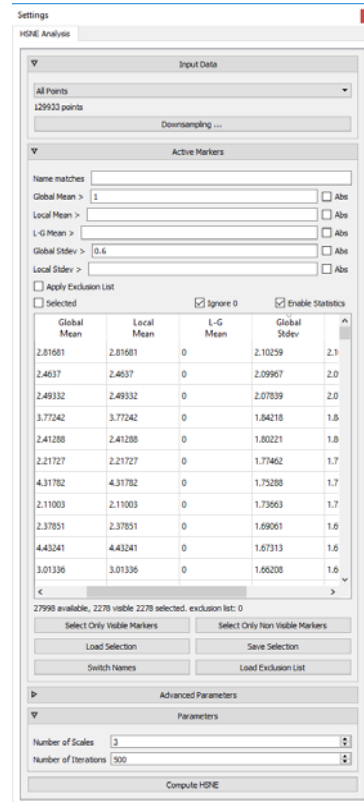
Once the statistics have been computed the table will be filled with the Global and Local statistics (mean and standard deviation). Global statistics are computed on "All Points" regardless of the selected subset while Local statistics are computed on the selected "Input Data". Note that by default "All Points" are selected as input data and therefore Global and Local statistics are the same. Since the mean and standard deviation can be influenced by a large number of 0-values in the dataset these 0-values can be ignored by checking the "Ignore 0" checkbox. With the statistics available the filter options above the statistics table can be used to limit the visible markers to for example values with a standard deviation above 0.6 and a mean value above 1.

Once you are satisfied with the remaining visible markers they can be selected by pressing the "Select Only Visible Markers" button. Alternatively the inverse selection can be made by pressing the "Select Only Non Visible Markers" button.

For convenience there is also the option to define an exclusion list of markers to exclude from the selection (e.g. mitochondrial or X and Y chromosome specific markers). This list can be loaded using the "Load Exclusion List" button. The format the same as for selections.

Once marker selection is completed, the number of scales and number of iterations per scale press "Compute HSNE" (or t-SNE) to start the SNE computation which will be displayed in the Map Viewer.

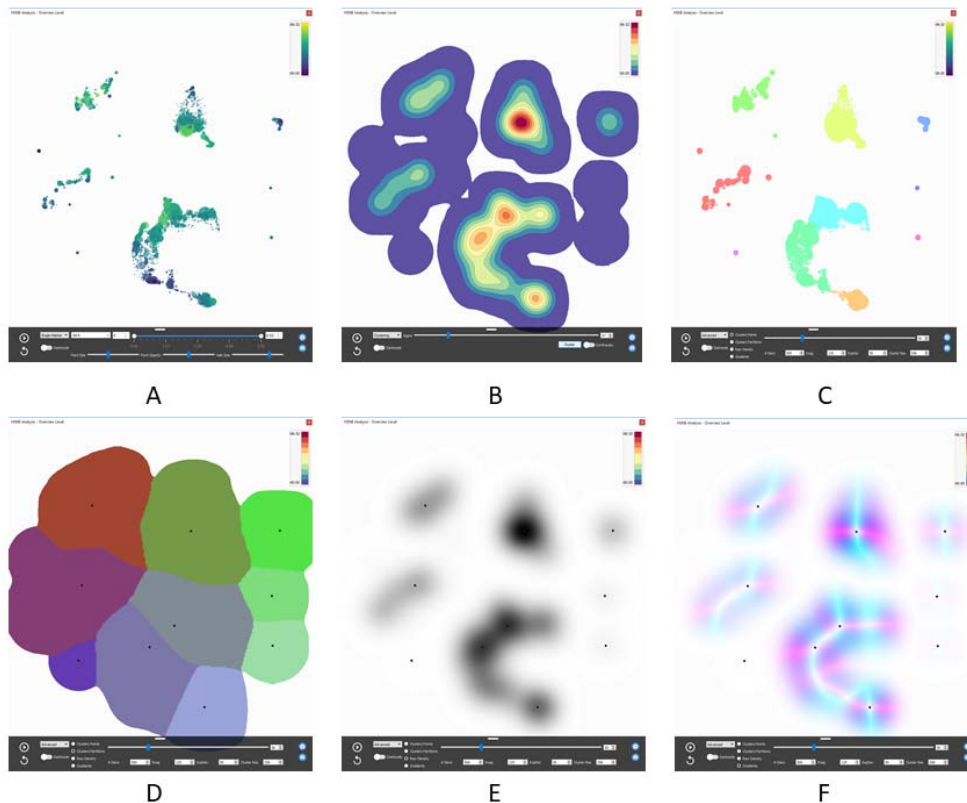

Note that for HSNE the map shown is the overview level embedding.

Map Viewer

The Map Viewer is designed for interactive exploration of tSNE and HSNE submaps at the single cell level, with the following features:

- SNE maps can be recomputed on the fly, and paused during computation for detail inspection and analysis. Note that linear structures during SNE computation with exponential force decay have been shown to suggest differentiation pathways and lineage relations [5].
- Single-Marker: (A) Expression of an arbitrary gene can be painted in the map by starting to type the gene name in the expression text box, and selecting the target gene from the drop down menu.
- Clustering: (B) Shows a density based visualization which can be used to interactively cluster the data points. The kernel size that is used for the density based estimation can be adjusted interactively by moving the slider. Pressing the Cluster button will cluster the embedding using the mean-shift algorithm using the density estimation as input.
- Advanced: Shows the clustering result either as points (C), Partitions (D) or shows the raw density map (E) or gradients (F).
- Metadata on each cell can be viewed by scrolling with the mouse wheel while hovering over the Metadata text box. This will bring up a legend box with the metadata items in that particular submap.
- Hovering over a cell brings up an information panel with the detail metadata on each nucleus.
- Manual annotations can be integrated by adding columns in the metadata. First scroll to Metadata viewing, and then click the "add category" button: the name of this category is added to the metadata. Then, a manual selection of cells can be annotated by pressing the "Label Selection" button, and a name for the selection can be chosen. Note that selecting a cluster in the higher HSNE levels propagates the label down in the hierarchy, enabling the annotation of large sets of cells in one annotation action.

A

B

C

D

E

F

Differential Expression Viewer

The Differential Expression Viewer is designed to mine the full feature set for differential expression between two selections of cells.
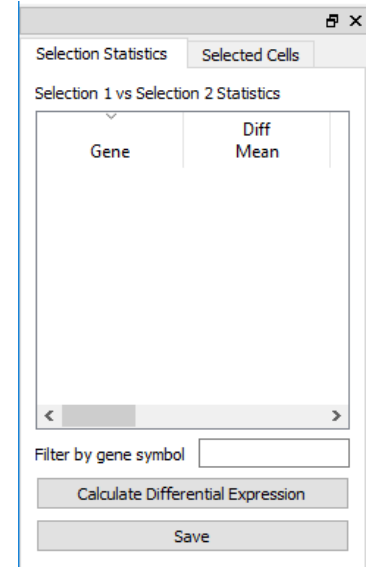
There are two methods to select cells / nuclei:
- Manual drawing of cell selections in the Map Viewer (Left-click and drag).
- Using the metadata legend in the map viewer: clicking on a metadata item in the legend selects the corresponding cells.

Selections can be emptied by left-clicking in an empty space in the map viewer. By default selections are stored in selection 1. Pressing the left-alt key while performing these actions will fill selection 2.

Once both selections are set, clicking the Calculate Differential Expression button will compute differential expression statistics over the full feature set and will fill the table.

The default behavior for the differential expression viewer is as follows:

- Differential expression and Beta are only computed if a gene is expressed in at least 5 cells in each selection, and otherwise set to 0.
- p-values for differential expression are computed if a gene is expressed in at least 10 cells in each selection using the Wilcoxon rank-sum test, or otherwise set to 1.
- p-values are corrected for multiple testing through Bonferroni correction.
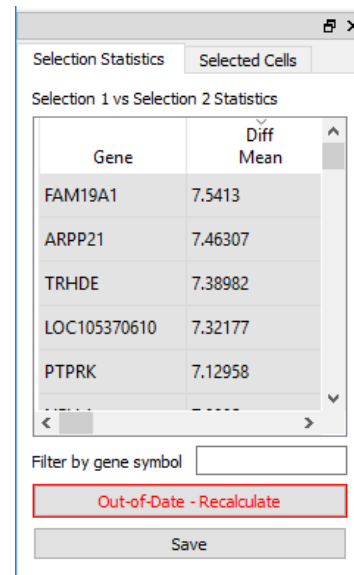- Compute times may increase with larger selections.

Once the differential expression computation has finished, the analysis results can be explored as follows:
- **Gene expression coloring in the map view** can be performed by clicking on a gene name in the left column. This will display the expression of the selected gene in the Map View
- **Gene sorting** can be performed by left-clicking on the sorting variable names at the top of each column. Left clicking again will sort the variable in reverse order. The columns can be reordered by clicking on the column name, and dragging it to a new position in the table.
- **Gene filtering** can be performed to reduce the number of listed genes by typing gene symbol characters in the "Filter by name" field. Markers from the pre-loaded gene exclusion list can be excluded by checking the "Filter excluded marker" list.
- **Cell selections** can be inspected, loaded from file or saved through a separate tab "Selected Cells".
- **Result export** can be achieved through the clipboard by right-clicking on the result table, or by pressing the "save" button.

If either of the selections is changed, the results in the table will no longer be up-to-date with the selections: this is indicated in the differential expression viewer by a darker background color for the statistics table and the button "Out-of-Date – Recalculate" will become visible. Note that the previously calculated differential expression results can still be interactively explored.

For an instructional video highlighting the user interface functionality, see: viewer.cytosplore.org

Context-sensitive user interactions in the Map and 2D Scatterplot viewers
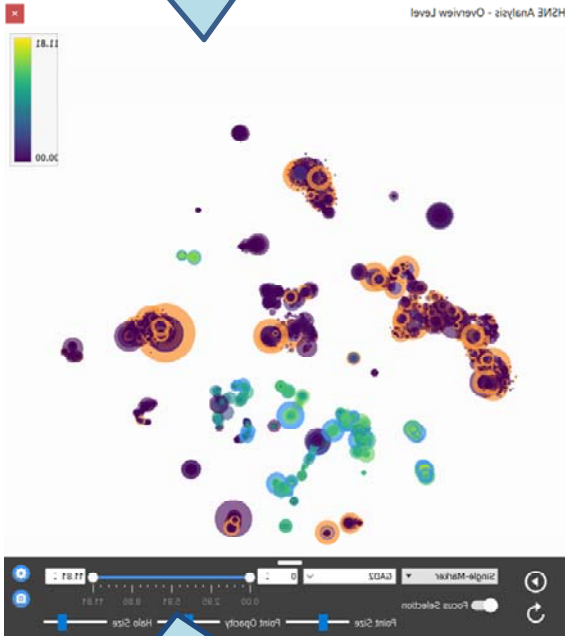
Map viewer
Viewing
- *Hover over a cell:* displays a widget with metadata on each nucleus under the arrow
- *Left-click on color bar:* enables selection of different expression color maps

Fill selection 1 (highlighted with blue halo)
- *Leftclick+drag*: **selects** nuclei within the dragged square (highlighted with blue halo)
- *Shift+leftclick and drag*: **adds new** nuclei to selection 1
- *Ctrl+A:* **selects all** nuclei In the current map view
- *Leftclick+drag* in an empty space: **deselects all** points
- *Ctrl+D:* **deselects all** points

Fill selection 2 (highlighted with orange halo)
- *Alt+leftclick and drag:* **selects** nuclei within the dragged square (highlighted with orange halo)
- *Alt+leftclick and drag in an empty space:* **deselects all** points
- *Alt+Shift+leftclick and drag:* **adds new** nuclei to selection 2
- *Altl+Ctrl+A:* **selects all** nuclei in the current map view
- *Alt+Ctrl+D:* **deselects all** points
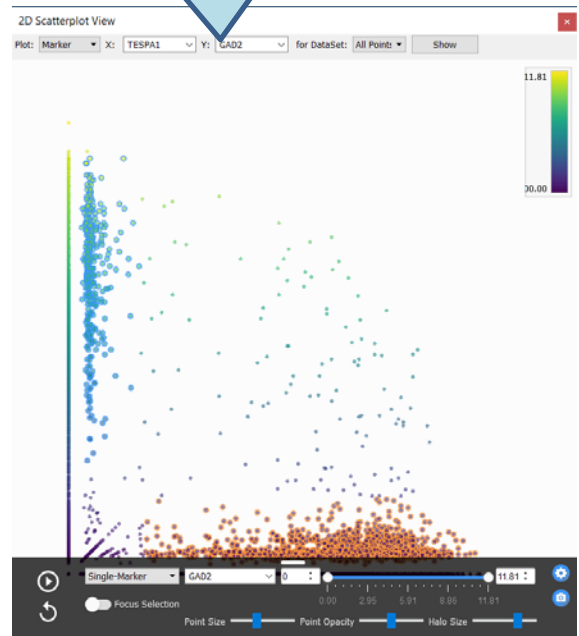
2D scatterplot viewer
Plot
- *Marker* enables plotting two of the markers in a bi-axial plot. Selections are linked with the Map viewer, enabling selection and annotation based on marker genes
- *Left-click on color bar:* enables selection of different expression color maps

Fill selection 1 (highlighted with blue halo)
- *Leftclick+drag*: **selects** nuclei within the dragged square (highlighted with blue halo)
- *Shift+leftclick and drag*: **adds new** nuclei to selection 1
- *Ctrl+A:* **selects all** nuclei In the current map view
- *Leftclick+drag* in an empty space: **deselects all** points
- *Ctrl+D:* **deselects all** points

Fill selection 2 (highlighted with orange halo)
- *Alt+leftclick and drag:* **selects** nuclei within the dragged square (highlighted with orange halo)
- *Alt+leftclick and drag in an empty space:* **deselects all** points
- *Alt+Shift+leftclick and drag:* **adds new** nuclei to selection 2
- *Alt+Ctrl+A:* **selects all** nuclei in the current map view
- *Alt+Ctrl+D:* **deselects all** points

Dark gray area
- GUI element dock that can be hidden with the small horizontal stripe at the top center.

Color Viewing switches
- Enable switching between viewing cluster colors, expression levels, and metadata. To select a gene in the expression mode: start typing the gene name, and select by mouse click the gene of interest that appears in autocomplete.
- *Middle mouse wheel scrolling* can be used to rapidly scroll through the drop-down items (genes or metadata)

Sliders
- Set different size and transparency properties

Play, Pause and Recompute buttons:
- Start, pause or recompute the A- tSNE computation. When pausing the map evolution, all exploration facilities are active.

Save a snapshot image of the current view
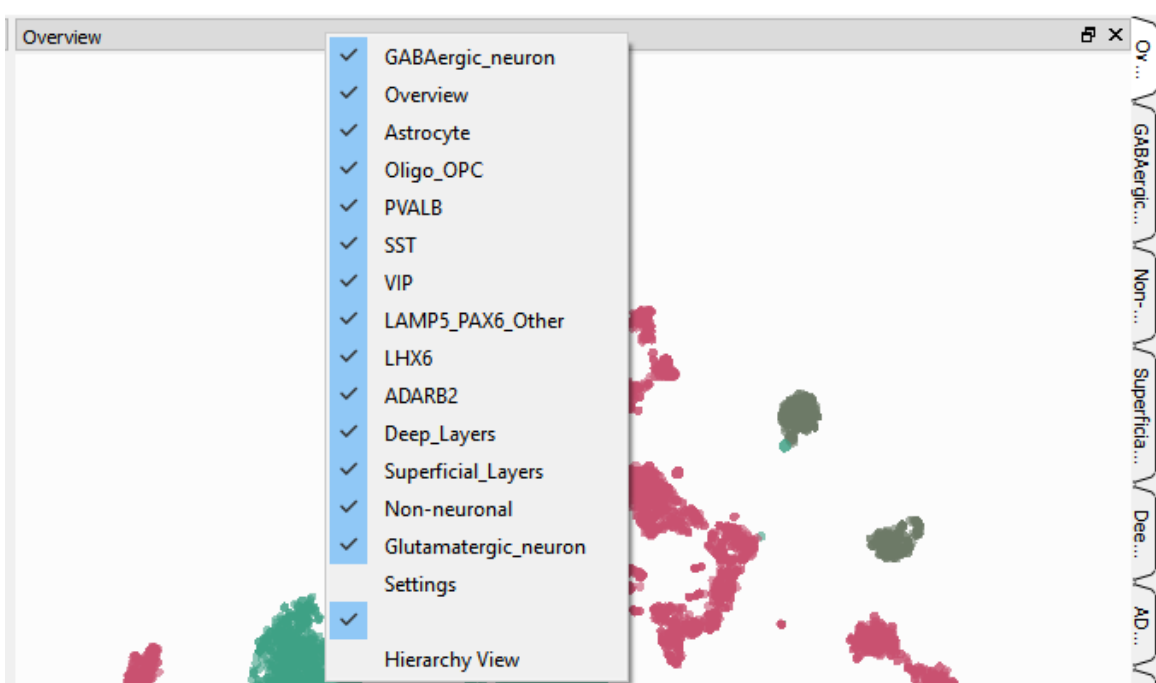- Press the camera icon in the right upper corner

Tab panels on the right border of the map view
- Enables switching between maps without the hierarchy view

General functions to organize views

The default organization of the different views can be customized by clicking and dragging on the top bar of each view. The organization of the windows can be cusomized as follows:

- **Move a view within the application window:** drag and drop within the application window will reorganize the other views. Dropping the view in the middle of another view will add it as a tab to the target view.
- **Resize a view:** each of the views can be resized by moving the mouse over the boundary of the view: arrows appear, and the border can be dragged to resize the view.
- **Dock a view outside of the application:** drag and drop to a location outside of the application window will float the view as a separate window, while maintaining the interactive links to all other views. This can be convenient for instance on a PC with two screens: views can be docked to a second screen.
- **Close a view:** a view can be closed by clicking the X in the right upper corner of each view.
- **Retrieve a closed view:** each view can be retrieved by right-clicking on the top bar of a view: this will bing up a list of views that can be toggled on or off. See below for an example.

Personal Settings

Personal default settings for t-SNE (such as the number of iterations) and visualization options (e.g. for the scatter plots) of Cytosplore Transcriptomics can be customized through the Cytosplore.ini file located in the program folder. Documention regarding the settings can be found in the ini file.

Acknowledgements

References

[1] T. Höllt, N. Pezzotti, V. van Unen, F. Koning, E. Eisemann, B.P.F. Lelieveldt, A. Vilanova, "Cytosplore: Interactive Immune Cell Phenotyping for Large Single-Cell Datasets", Computer Graphics Forum, vol. 35 (3), pp 171-180, doi 2016.

[2] V. van Unen, T, Höllt, N. Pezzotti, N. Li, M. Reinders, E. Eisemann, A. Vilanova, F. Koning, and B.P.F. Lelieveldt. "Visual Analysis of Mass Cytometry Data by Hierarchical Stochastic Neighbor Embedding Reveals Rare Cell Types". Nature Communications, 2017

[3] N. Pezzotti, J. Thijssen, A. Mordvintsev, T. Höllt, B. van Lew, B.P.F Lelieveldt, E. Eisemann, and A.Vilanova. "GPGPU linear complexity t-SNE optimization." IEEE transactions on visualization and computer graphics 26, no. 1 (2019): 1172-1181.

[4] T. Höllt, N. Pezzotti, Vincent van Unen, Frits Koning, Boudewijn P.F. Lelieveldt, and Anna Vilanova. CyteGuide: Visual Guidance for Hierarchical Single-Cell Analysis. IEEE Transactions on Visualization and Computer Graphics, vol 24(1), 2018.

[5] N.Li,* V. van Unen,* T. Höllt, A. Thompson, J. van Bergen, N. Pezzotti, E. Eisemann, A. Vilanova, S. M. Chuva de Sousa Lopes, B. P.F. Lelieveldt, F. Koning, "Multi-lineage Trajectory Analysis of Innate Lymphoid Cells Reveals an ILC3 Precursor in the Human Fetal Intestine", Journal of Experimental Medicine, DOI: 10.1084/jem.20171934, 2018.

[6] R. Hodge, T. Bakken et al, "Conserved cell types with divergent features between human and mouse cortex", biorXiv, 2018