

SPATIAL TRANSCRIPTOMICS VIEWER v0.6.3 USER MANUAL

Jose Fernandez Navarro

September 16, 2016

Preface

The Spatial Transcriptomics viewer (ST viewer) is a desktop application that allows users to securely access and visualize spatially distributed gene expression profile data with their respective tissue image. At the same time it allows users to analyze the data directly, or export it for their own analyses. The application can obtain the data from a secured server (configuration files will need to be updated for this) and/or from local files.

For installation instructions, please see the README.md file. This file describes the install instructions on Windows, OSX and Linux.

Licensing

This is open source software under the MIT license. Read LICENSE for more information about the licensing terms.

Contact

Please report bugs, feedback, questions and errors to:

Jose Fernandez Navarro (jose.fernandez.navarro@scilifelab.se) or

Alexander Stuckey (alexander.stuckey@scilifelab.se)

Contents

1 General Overview	3
1.1 User Interface	3
1.2 Gene Display Modes	5
2 Using the Viewer	9
2.1 Selection Mode	9
2.2 Differential Expression Window	11
3 Known Issues and Future Plans	13
3.1 Known Issues	13

List of Figures

1.1	The spatial transcriptomics viewer default interface	3
1.2	Toolbar showing username after logging in.	3
1.3	Datasets view.	4
1.4	Default view after loading a dataset	4
1.5	Gene List	5
1.6	The toolbar when a dataset is loaded.	5
1.7	The visual canvas of the ST viewer in each map mode.	6
1.8	Gene display configuration	7
2.1	The visual canvas of the ST viewer showing different selection types.	9
2.2	The selection window.	10
2.3	The genes in the current selection.	10
2.4	The differential expression window.	11

Chapter 1

General Overview

1.1 User Interface

When running the spatial transcriptomics viewer for the first time, you will be presented with an interface as shown in figure 1.1. If the viewer has been configured to access datasets through the ST API then the login window will be shown, otherwise it will not be shown.

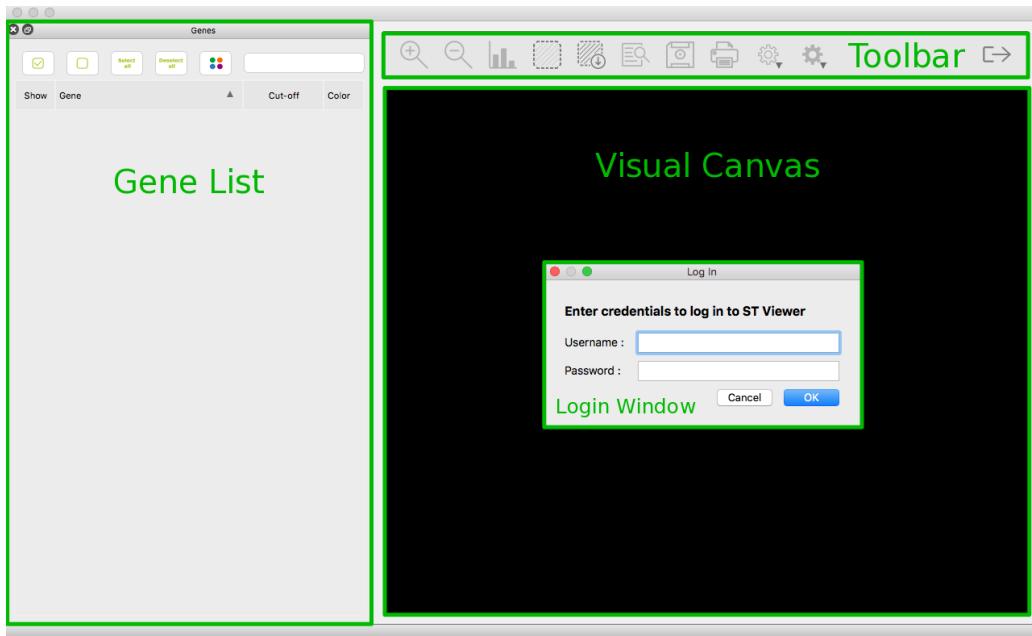


Figure 1.1: The default interface presented when the ST viewer is run for the first time. The elements present are the gene list window on the left, the toolbar in the top right, the visual canvas on the right, and optionally the log in window (if the viewer is configured to access datasets through the ST API).

Once you have logged in (if applicable), your user name will be displayed in the toolbar, as shown in figure 1.2.



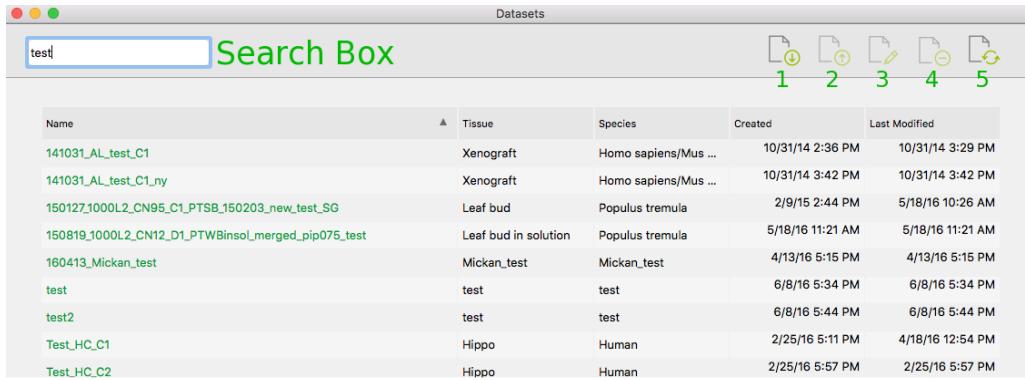
Figure 1.2: Toolbar showing username after logging in.

There are two other main windows in the ST viewer, datasets and selections. They can both be accessed from the views menu by clicking views → datasets and views → selections. The datasets window shows all the datasets that you have access to. There is a search box that can be used to find a keyword in the dataset names to narrow down the list of datasets (see figure 1.3). There are five buttons to the right of the dataset search box (labelled 1 — 5). They are:

1. Import dataset from file (local file on your computer)

2. Open selected dataset (for datasets stored in the database)
3. Edit selected dataset (change name, add comments)
4. Delete selected dataset
5. Refresh list of datasets

The datasets displayed in the dataset view can also be sorted by each column (e.g. name or species).



A screenshot of the 'Datasets' view. At the top, there is a search box containing the text 'test'. Below the search box is a table with columns: Name, Tissue, Species, Created, and Last Modified. The table lists several datasets, including '141031_AL_test_C1' and '141031_AL_test_C1_ny' which are highlighted in green, indicating they match the search criteria. Other datasets listed include '150127_1000L2_CN95_C1_PTSB_150203_new_test_SG', '150819_1000L2_CN12_D1_PTWBinsol_merged_pip075_test', '160413_Mickan_test', 'test', 'test2', 'Test_HC_C1', and 'Test_HC_C2'. The last two datasets are also highlighted in green.

Figure 1.3: Datasets view. The search box is being used to display only the datasets that contain the word "test" in their name.

The selection view will be discussed further in chapter 2.

Note: Each dataset is associated with a substantial amount of data, which will be accessed after a dataset has been selected and opened. As such the transition between the views might require a few moments. It is worth noting that the data is cached, which implies that once it is downloaded, the next time the same dataset will open much quicker. This does not apply if the dataset is updated in the cloud in the meantime.

When a dataset is loaded, the gene list will populate with the list of all genes that are associated with the current dataset, and the visual canvas will display the image of the tissue section (see figure 1.4).



Figure 1.4: The default view after loading a dataset. The gene list is populated by all the genes in the data, and an image of the tissue section is presented on the visual canvas.

The gene list is headed by five buttons and a text box. The buttons are, in order:

1. Show selected
2. Hide selected
3. Select all
4. Deselect all
5. Set colour
6. Search box

The text box enables you to search the list of genes for genes that are of interest. Shown genes are marked with a checkmark in the box to the left of the gene name / ensemble ID. Selected genes have their names highlighted in orange instead of green, and have a darker grey background. Note that shown genes and selected genes may differ (as seen in figure 1.5). The colour column allows you to choose custom colours for single genes or groups of genes, in order to increase visual clarity. The gene list can also be sorted by gene name and cut-off value.

Show	Gene	Cut-off	Color
<input checked="" type="checkbox"/>	ENSMUSG00000098178	112	Red
<input checked="" type="checkbox"/>	ENSMUSG00000026576	109	Red
<input type="checkbox"/>	ENSMUSG00000027273	98	Red
<input checked="" type="checkbox"/>	ENSMUSG00000070880	88	Red
<input type="checkbox"/>	ENSMUSG00000002985	88	Red
<input type="checkbox"/>	ENSMUSG00000015090	86	Red
<input type="checkbox"/>	ENSMUSG00000036564	86	Red

Figure 1.5: Gene list for the current dataset, sorted by cut-off value.

When a dataset is loaded into the viewer the buttons above the visual canvas become active (figure 1.6). The buttons on the toolbar are:

1. Zoom in
2. Zoom out
3. Feature distribution histograms
4. Selection mode toggle
5. Create selection object
6. Feature selection by regular expression
7. Save a picture of the visual canvas
8. Print a picture of the visual canvas
9. Gene display configuration
10. Visual canvas options
11. Logout button



Figure 1.6: The toolbar when a dataset is loaded.

The gene display configuration is shown in figure 1.8. From top to bottom, the options are:

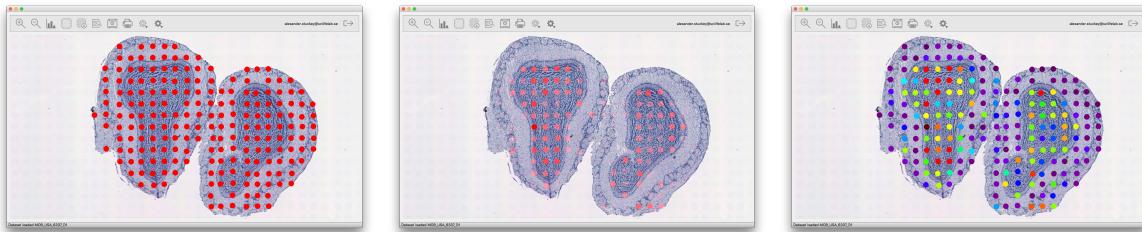
Show Grid toggles the display of the original array and its border simplified to 5×5 features per square.

Show Genes toggles the display of the currently selected genes.

Choose Colour Grid opens a menu to change the colour of the displayed array grid.

1.2 Gene Display Modes

In **Normal Mode** (figure 1.7a) the selected genes in each feature are treated independently. When multiple genes are selected, the final colour of the feature will be a mix of the different colours of the selected genes. In normal mode, the threshold setting treats the genes in each feature separately. Therefore, the threshold can hide some of the selected genes in a feature, or even all the genes in a feature. In all cases, the final colour of the feature will be adjusted accordingly.



(a) Visual canvas in normal mode. (b) Visual canvas in dynamic mode. (c) Visual canvas in heatmap mode

Figure 1.7: The visual canvas of the ST viewer in the normal, dynamic and heatmap map modes. All genes are selected in these images, with all colours set to the defaults.

In **Dynamic Range Mode** (figure 1.7b) the selected genes in a feature are all treated as one unit (as opposed to normal mode). This means that each feature is treated as a single unit where the level of expression is determined by the combination of all the genes selected in the feature. The final expression is normalised and the alpha channel (the brightness) of the feature is adjusted according to the normalised final expression of the feature. This way, users can get an idea of the total expression level of a feature compared to all other features. The colour of each feature will be computed in the same way as in the normal mode (it combines the colours of the genes in each feature). The threshold in dynamic mode affects features as a whole, instead of gene-by-gene as in normal mode. When the threshold is changed the intensity for all the features is re-calculated, as the normalisation factor might have changed.

Heatmap Mode (figure 1.7c) treats genes in a feature in the same manner as the dynamic range mode. Instead of adjusting the brightness of the colour, a new colour is computed for the feature representing the overall level of expression. The low frequencies are blue and represent a low hit count, while the high frequencies are red and represent a high hit count. The heatmap spectrum and threshold levels can be seen on the legend, when the legend is enabled. The threshold works the same way as in dynamic range mode, when the threshold is changed the colour range and colour of features will be recalculated.

The opacity, size and shape parameters change the display of the features on the visual map mode.

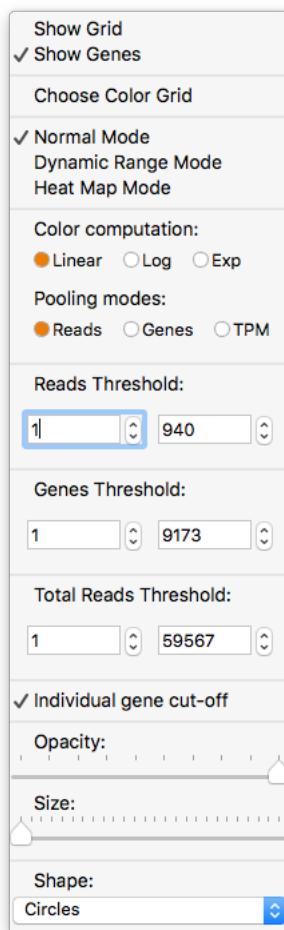


Figure 1.8: Gene display configuration

Chapter 2

Using the Viewer

2.1 Selection Mode

Selection mode is toggled using button 4 in 1.6. When selection mode is enabled the user is able to select features on the canvas by clicking and dragging a box on the visual canvas (figure 2.1a). The user can select disjointed areas by holding down shift while selecting an area (figure 2.1b). Areas can also be removed from the selection by either shift + control (Windows and Linux) or shift + command on Mac (figure 2.1c).

In addition to selecting via clicking and dragging on the visual canvas, users can also select genes via regular expressions (button 6 in figure 1.6). Regular expressions provide the user with a powerful means to select specific genes by matching the gene names against a specific string pattern. The regular expression syntax is modeled after the Perl regular expressions. In addition the selection dialog provides three options; to include ambiguous genes, to select non-visible genes, and to make the regular expression case sensitive. By default the regular expression selection excludes ambiguous genes and ignores case differences.

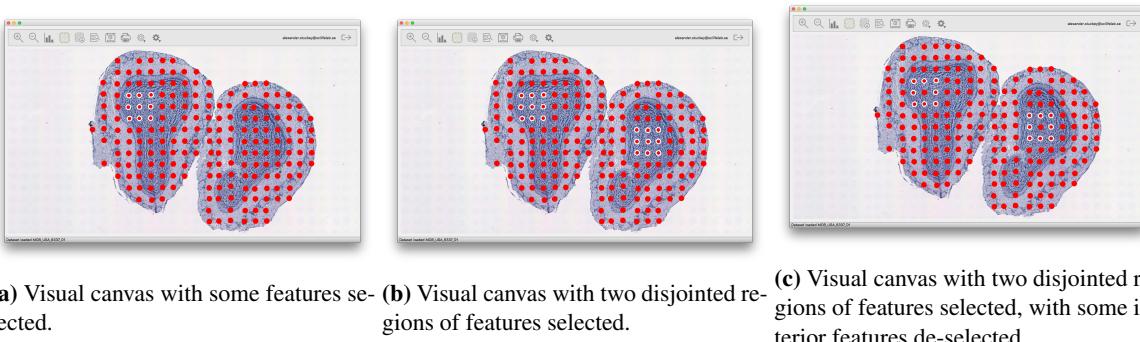


Figure 2.1: The visual canvas of the ST viewer showing the different ways of selecting features.

When the desired features are selected, the user can create a selection object by clicking button 5 in 1.6. The selections can then be viewed by clicking on views → selection. This brings up the selection window (figure 2.2).

The selection menu lists all the selections that have been made in the current sample. The textbox on the top left enables the user to narrow down the list of selections. The buttons on the right, in order, show an image of the visual canvas with the selection that was used, saves the current selection to the cloud, shows the list of genes in the selection (along with the count of genes, figure 2.3), performs a differential expression analysis on the selected selections (figure), saves the current selection to the local computer, edits the name and details of the current selection, and remove the current selection.

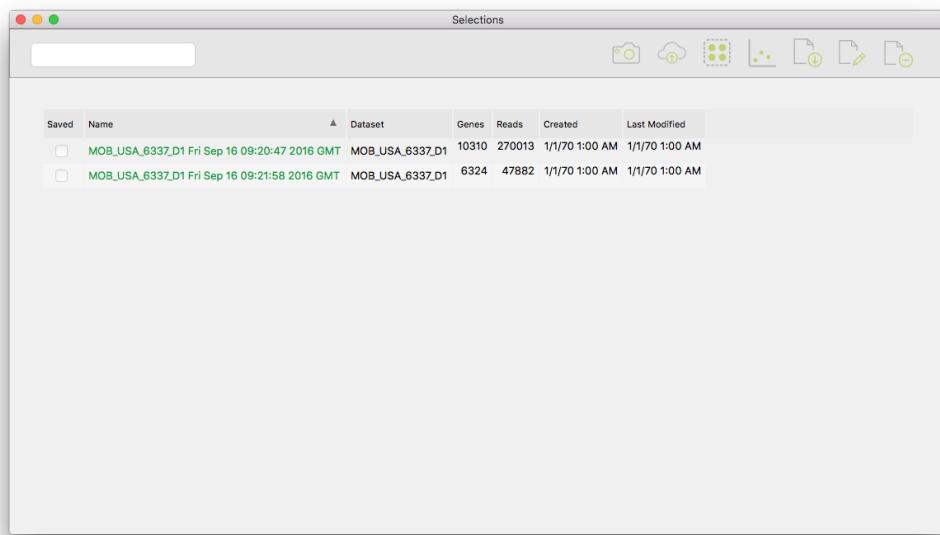


Figure 2.2: The selection window.

Gene	Total count
ENSMUSG00000000001	14
ENSMUSG00000000056	3
ENSMUSG00000000058	7
ENSMUSG00000000085	33
ENSMUSG00000000088	183
ENSMUSG00000000093	1
ENSMUSG00000000094	1
ENSMUSG000000000126	1
ENSMUSG000000000131	8
ENSMUSG000000000134	9

Figure 2.3: The genes in the current selection.

2.2 Differential Expression Window

The differential expression window shows a log-log graph of counts for each gene in both selections, a list of all the genes and their counts in both selections and some statistics for the differential expression analysis. The statistics include The number of genes in both selections, the number of genes present in both selections, and the number of genes unique to each selection. Also displayed is the Pearson correlation for the differential expression. It is also possible to change the threshold for the number of reads used in the differential expression analysis. Changing the threshold changes both what is displayed on the graph, and also recomputes the Pearson correlation. The save button will save the current graph to the local computer.

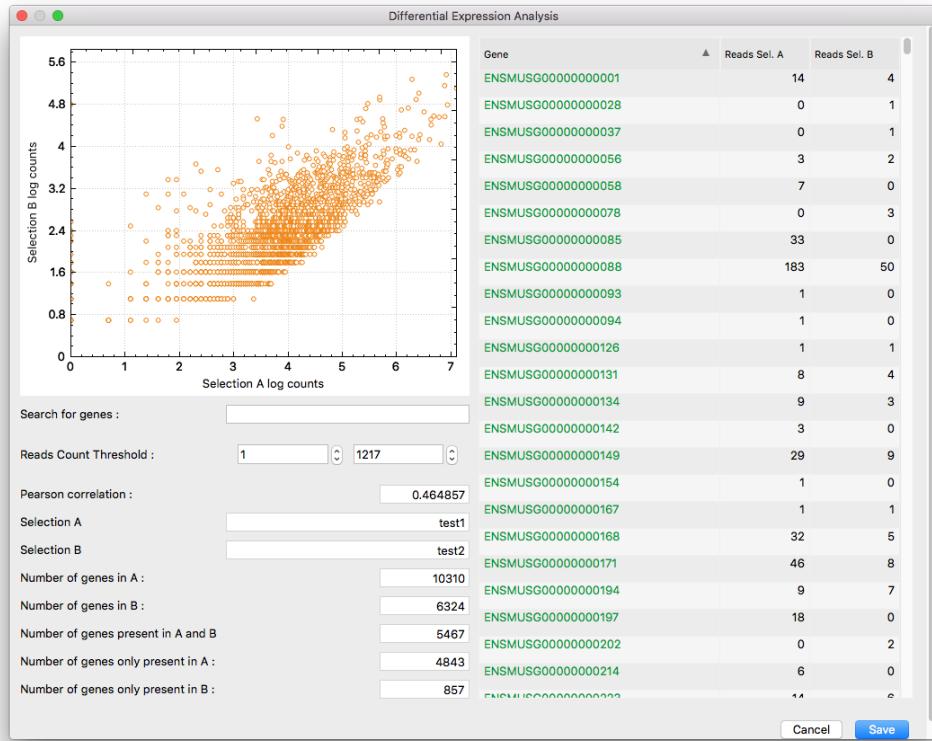


Figure 2.4: The differential expression window.

Chapter 3

Known Issues and Future Plans

3.1 Known Issues

As the application is still in the early alpha phase it may contain functional as well as visual discrepancies and issues. Some of these are known but were out of the scope of this release.

Below follows a list detailing each known issue, the afflicted platform as well as a brief description of its effect:

- General
 - Speed and memory issues
- Linux
 - Some GUI related artifacts
- Windows
 - Some GUI related artifacts