

Single-Cell spatial Explorer Manual

Frédéric PONT

August 11, 2023

Contents

| | | |
|--|---|----|
| 0.1 | Single-Cell spatial Explorer presentation | 2 |
| 0.1.1 | Computer configuration | 3 |
| Minimal computer configuration | 3 | |
| Recommended computer configuration. | 3 | |
| Hard disk | 3 | |
| CPU | 3 | |
| Memory | 3 | |
| Display | 3 | |
| Graphic card driver | 4 | |
| 0.1.2 | Single-Cell Spatial Explorer features. | 4 |
| 0.2 | Software installation and use | 5 |
| 0.2.1 | Quick start | 5 |
| 0.2.2 | Preferences setting | 7 |
| 0.2.3 | Drawing gates on the microscopy image | 10 |
| 0.2.4 | Exporting gates | 11 |
| Importing gates | 11 | |
| 0.2.5 | Filtering data with gates | 11 |
| 0.2.6 | Plotting gates | 12 |
| 0.2.7 | Comparing gates | 12 |
| 0.2.8 | 2D interactive plots | 13 |
| 0.2.9 | Displaying cluster | 15 |
| 0.2.10 | Displaying expression | 15 |
| Slide Show | 17 | |
| 0.2.11 | Importing Cells | 18 |
| 0.2.12 | Saving images | 18 |
| 0.2.13 | Software compilation from source | 19 |
| 0.2.14 | Tips | 20 |
| How to save memory on a small computer ? | 20 | |
| How to calculate the scaling factor after resizing an image ? | 20 | |
| Working with large data sets such as Akoya Phenocycler | 20 | |
| 0.2.15 | Troubleshooting | 21 |
| The software crashes when I display clusters or expression. | 21 | |
| The software do not start on Linux with a Glibc error. | 24 | |
| There is a shift between the image and the cluster/expression map. | 24 | |

| | |
|---|----|
| My scatter plot is not displayed or partially displayed . . . | 25 |
| The software is slow | 25 |
| The preference file is corrupted or not compatible with an older version of the software | 26 |
| The legend mask the image | 27 |

0.1 Single-Cell spatial Explorer presentation

Single-Cell Spatial Explorer is a new tool, part of Single-Cell Signature Explorer software package [1] dedicated to spatial transcriptomics. It is able to display single cell data using color gradients on a microscopy image. It is possible to define and gate cell populations based on the cells or the single cell data. An unlimited number of data tables containing XY image coordinates can be filtered according to the gates. The content of the gates can be plotted with any XY coordinates or compared. Gates comparison is done by plotting an interactive volcano plot, in which it is possible to select points of interest and display the corresponding gene/pathway expression.

Files can be accessed at GitHub Single-Cell Spatial Explorer web page.

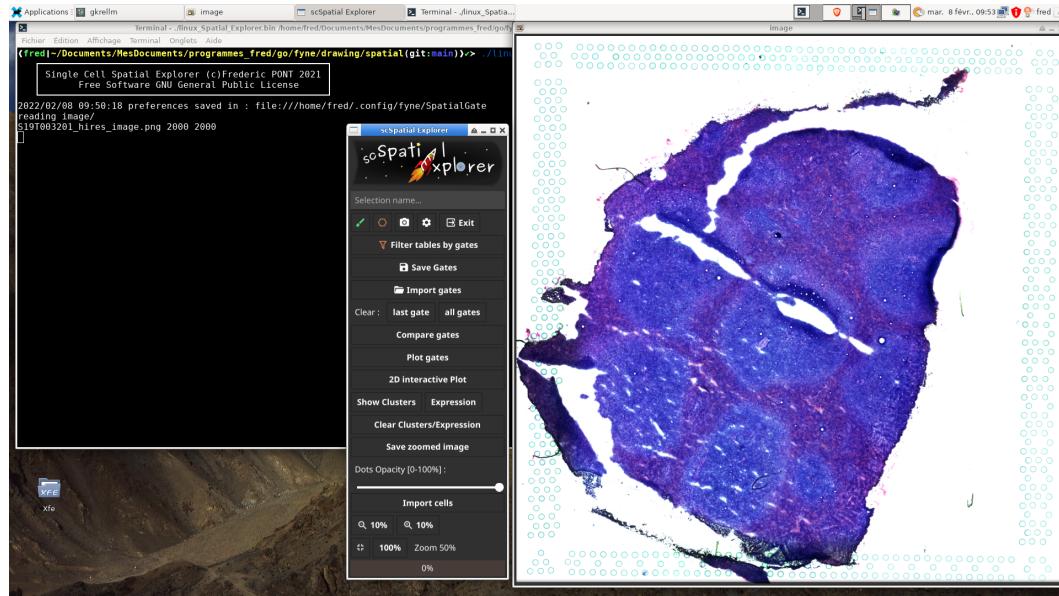


Figure 1: Single-Cell Spatial Explorer main windows at start. On the left, the terminal display log information. On the middle the main toolbox control the software. On the right the microscopy image re-scaled to fit the window.

0.1.1 Computer configuration

Minimal computer configuration

Single-Cell Spatial Explorer has been developed on linux Manjaro on a computer with an Intel ©Core™i5-3470 CPU @ 3.20GHz with 8 GB RAM, a 7200 rpm hard disk, a video card Nvidia GeForce GT630 and a HD display with a resolution 1920×1080 . This is a comfortable configuration for a small data set of about 3000 cells.

Although it could be possible to use Single-Cell Spatial Explorer on a laptop, we recommend to use a HD display.

Recommended computer configuration.

Hard disk

Single-Cell Spatial Explorer has been optimized to save RAM at the expense of hard disk access. As a result, Single-Cell Spatial Explorer will take a great advantage of fast disks such as SSD disks. Copy the software folder on your fastest disk or RAID.

CPU

Except tables filtering which can be massively parallel (1 table / CPU), most of the tasks uses 2-3 CPU. Single-Cell Spatial Explorer will be fast on CPU with a high single thread speed, such as the CPUs on top of this list.

Memory

The gates comparison is the most RAM demanding operation because the whole dataset have to fit in RAM if the user select all the columns in the comparison. For 3000 cells and 20 000 genes, 8GB RAM are enough (at least on a light Linux distribution), but for larger data sets, it could be necessary to increase RAM over 8GB.

Display

The display is very important for an optimal use of Single-Cell Spatial Explorer. We recommend at least one HD display, two if possible, or a high resolution display such as 4k displays. With a 4k display, it is possible to work on a 2000×2000 image at zoom 100% which is the more precise zoom since there is no pixel interpolation. Note that the expression/cluster layer is not updated by the zoom for performance reasons. The user has to click on the Show Cluster/Expression button when the final zoom value is obtained. This avoid a time consuming refresh of thousands/billions dots for each intermediate zoom values. **Very important : do not use display zooming, use the screen at the native resolution with a 100% zoom.**

Graphic card driver

The driver of the graphic card must support OpenGL or you will have an error : "The driver does not appear to support OpenGL". This is a requirement of the Fyne graphical toolkit.

On Linux, X11 is recommended, Wayland is not yet fully supported by Fyne ¹. On VirtualBox to enable OpenGL support, you must install the Guest Additions and enable 3D acceleration.

0.1.2 Single-Cell Spatial Explorer features.

1. Single-Cell Spatial Explorer is distributed with a detailed PDF manual and video tutorials
2. Single-Cell Spatial Explorer is ready to use in a pre-compiled binary, no installation required.
3. cross-platform (the graphical interface and the software are coded in pure Go)
4. low memory usage
5. compatible with any PNG image associated with any text file with tab separator containing XY coordinates of the image.
6. compatible with any numeric data : gene expression, pathway scores, antibody expression etc...
7. unlimited number of gates.
8. import/export gates in ImageJ/Fiji format.
9. extract cells and sub-tables delimited by the gates on an unlimited number of tables. Exportation is done in TAB separated files for great interoperability.
10. 2D plots of the cells inside the gates with any XY coordinates : t-SNE, UMAP, gene expression, pathway scores, antibody expression etc...
11. interactive 2D plot to show the selected cells on a t-SNE, UMAP or any other coordinates on the image and to filter the data tables into sub-tables.
12. cluster display with 3 color gradients, custom color palette, custom dot opacity and custom dot size. Shuffle color option change color positions on the map leading to almost 2 billions of possible images with 12 clusters.

¹Accordind to Fyne developpers, the application built for X11 will work on Wayland through XWayland. It's just applications built for Wayland that need some more work.

13. display any kind of cell expression (genes, pathways, antibodies...) with 7 preset gradients, custom legend color, dot opacity and custom dot size. The gradients are simple two colors maps and rainbow colors maps Turbo, Viridis and Inferno to optimize accuracy and details visualisation.
14. Min/Max intensity sliders to tune image contrast or remove artefacts due to outliers.
15. Expression opacity gradient with min/max threshold.
16. slide show to review many cell expression maps without need of repetitive click.
17. screenshot or native resolution image exportation.
18. import and display an unlimited number of cells list by repetitive click on the "import cells" button. The format is directly compatible with Single-Cell Virtual Cytometer [2].
19. compare two groups of gates together in the whole dataset .
20. compare one group of gates against all the remaining cells.
21. draw an interactive volcano plot after gate comparison.
22. plot cell expression of a selected dot in the volcano plot.
23. export volcano plot and the corresponding data table.
24. image zoom 10-200%

0.2 Software installation and use

Single-Cell Spatial Explorer has been compiled for GNU Linux, Mac OS and Microsoft®Windows™64 bits and require no complex installation procedure. A standalone executable file requiring no dependencies is provided.

0.2.1 Quick start

Windows and MacOS installation video.

1. unzip the software
2. **we recommend to test the software with the test files** (data table + image) first and to save a copy of the test files.
3. copy Single-Cell data, a text file with tab separator (extension does not matter and can be csv/tsv) to the “data” directory : the first row are gene or pathway (or anything else) names, the first column are cell barcodes and the next columns are the expression values, with separator TAB. **Avoid**

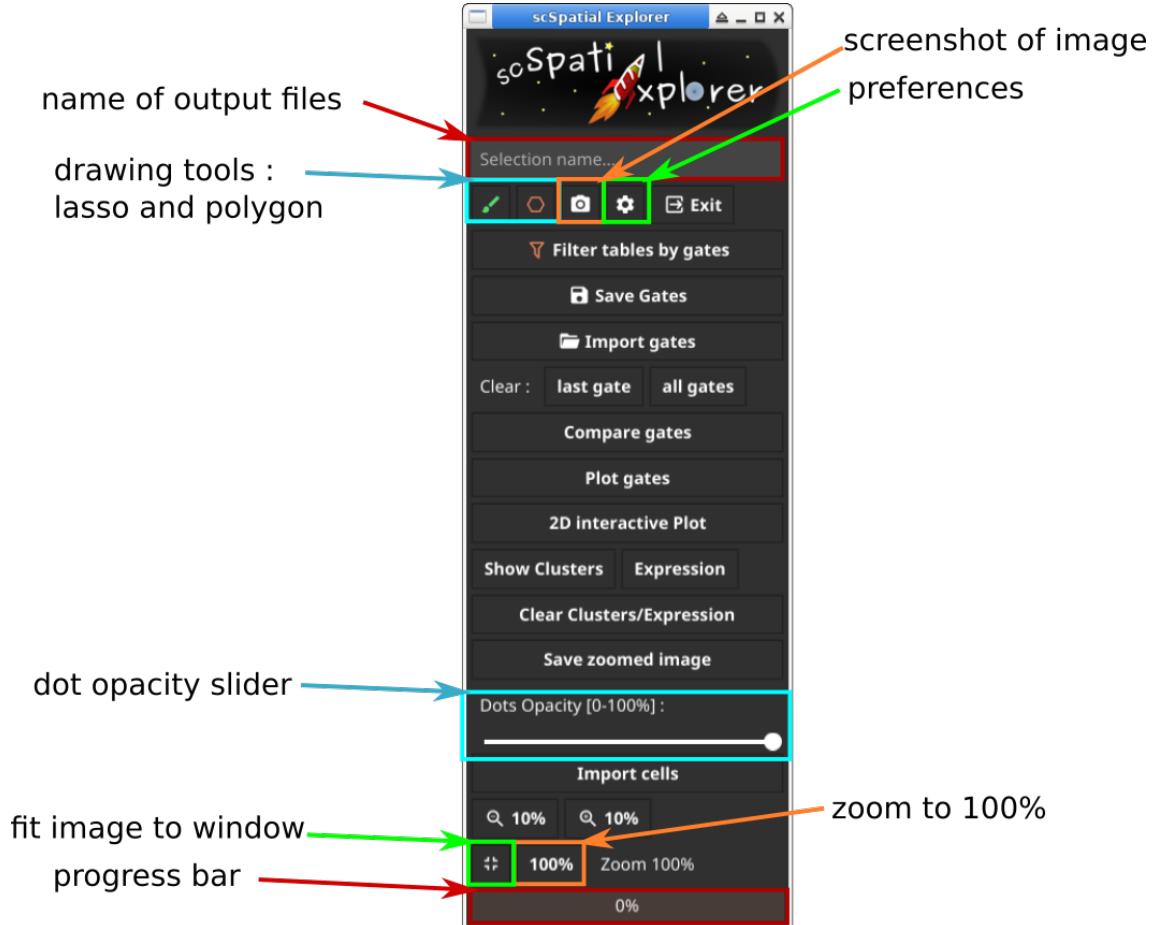


Figure 2: Single-Cell Spatial Explorer main toolbox.

spaces and special characters to name the files. The table MUST contain [X,Y] coordinates of the cells in the microscopy image. This is an example of the table :

| id | P53_PATHWAY | X_hires_image | Y_hires_image |
|--------------|-------------|---------------|---------------|
| cell_AAACTGC | 2.083 | 13378 | 4022 |
| cell_AAGACCT | 2.211 | 4340 | 12932 |
| cell_AACAGAG | 2.397 | 10102 | 2899 |

It is possible to copy an unlimited number of data tables in the “data” directory : the first table by alphabetical order will be used by the software, but all the tables can be filtered by gates.

4. copy the microscopy image in PNG format in the ”image” directory. Avoid spaces and special characters to name the files.

5. start the software using the binary file matching your operating system (see section 0.2.15).

On windows, the software can be started by a double click on the binary .exe file. Alternatively, it is possible to start the software from the command line.

On linux or Mac, we recommend to start the software from a terminal using the command ”./binaryfile” from the main directory of the software. The ”binaryfile.bin” must be executable. If necessary the permission to be executed can be changed using the command : chmod +x binaryfile.bin
On Mac, for security reasons, it is necessary to give execution rights to the binary to be able to start the software. This is illustrated in the Mac video.

Do not start multiple instances of the software. Multiple instance will try to use the same temporary files, such as the parameters preferences. This could give unexpected results or worse corrupt the preference file. After using any tool box, except the main panel of the software, is it highly recommended to close it, before opening another one.

Do not over-click, some tasks can be slow and the software needs time to read the data and process them. To be able to process large data sets on a standard computer the entire data set is not loaded in RAM and only the required data are read from the data file. This has a speed cost and repetitive click will not accelerate data processing, on the opposite it will start the task from zero again, slow the process and even crash the software.

6. Two windows should appear : the microscopy window and the main tool box (cf. fig. 1).
7. click on the ”Preferences” button and set all the preferences (cf. fig. 2). Close the ”Preferences” window and open it again to verify that all the settings are well saved. Note that it is necessary to restart the software to set some preferences, such as windows sizes. Do not forget to click the ”Rotate +90” radio button to test the software with test files.
8. Very important : when you display clusters/expression, do not forget to set the opacity slider between 50-100% opacity or you will see no results for clusters or expression.

0.2.2 Preferences setting

It is mandatory to set the preferences first before working with the software. Indeed the link between the microscopy image and the data table is made using column names and values that are in the preferences (cf. fig. 3). Preferences items :

- Scaling Factor : the scaling factor is the multiplicative number making the link between the [X,Y] coordinates in the data table and the image.

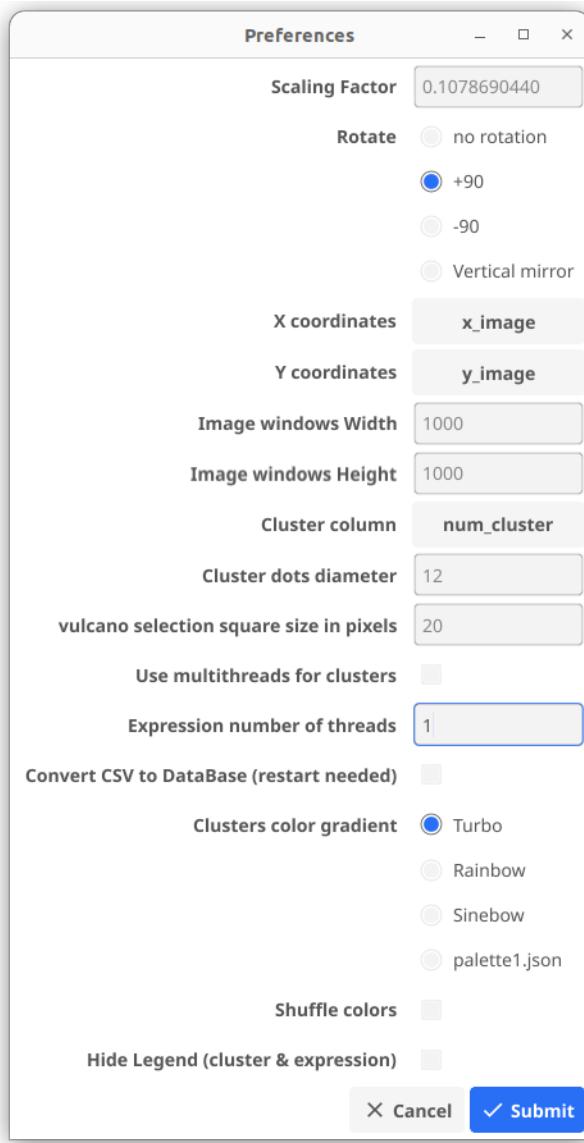


Figure 3: Single-Cell Spatial Explorer preferences window with the parameters for the demo files.

- Rotate : if the image is rotated (-90° , $+90^\circ$, Vertical mirror) compared to the coordinates in the table, check the corresponding option.
In our hands, for 10x Genomics data (image origin [0,0] on top left), select $+90$ and for Nanostring data (image origin [0,0] on bottom left), select "Vertical mirror".

- X,Y coordinates : the columns of the data table with the [X,Y] coordinates of the image.
- Image window width/height : the size of the image window. CAUTION the window size must fit the display. For example for a HD screen, the maximum resolution is 1920×1080 and the maximum recommended window size is 1000×1000 .
- Cluster column : the column of the data table with the cluster number. This column must contains **only integer numbers**. Each integer is used as a cluster number (see section 0.2.15).
- Cluster dots diameter : the point size in pixels to display clusters or expression.
- volcano selection square size in pixels : the volcano plot selection square size in pixels. This square is used to select dots in the volcano plot
- Clusters color gradient : the gradient use to display clusters. "palette1.json" is a custom palette with 26 colors. If the number of cluster is > 26 , then the Turbo gradient will be used instead.
For advanced users, it is possible to add custom palettes in json format in `src/palette/`, they will be active after a software restart. The json format must contain the RGB colors using this syntax :

```
{
  "pal": [
    {
      "R": "98",
      "G": "202",
      "B": "230"
    },
    {
      "R": "255",
      "G": "255",
      "B": "0"
    }
  ]
}
```

- "Use multithreads for clusters" : deactivated by default. this option is useful in two cases :
 1. When this option is activated, each cluster is rendered in a separate image in the directory `temp/clusters`. Just after displaying the clusters you can copy theses images to show each cluster separately.
 2. If you have millions of spots, multithread can increase speed

- "Expression number of threads" : 1 thread by default. Increasing the number of threads can increase speed when the number of spots is very high. Each thread will compute an image with $1/\text{nbThreads}$ spots, then the images are merged together. With 2.2 millions of spots the optimal number of threads is ~ 10 for a $3\times$ speed gain. When the number of spot is low, multithread will be slower than single threads, so we recommend to make a speed test to choose the best option.
- "Convert CSV to DataBase (restart needed)" : when the CSV data file has many rows/columns, a speed increase can be obtained by converting the CSV file to a pogreb database. The database will be located in "temp/pogreb/" and will have the name of the CSV file without extension. Avoid spaces and special characters to name the files. It is necessary to restart the software after database conversion. The databases in "temp/pogreb/" are not cleaned automatically to save the conversion time for further use. Clean the directory "temp/pogreb/" if some databases are no longer needed. With 2.2 millions of spots, the speed increase is $\sim 5\times$ for data reading.
- The "shuffle colors" checkbox, randomize the color attribution to clusters. This is a very powerful option to draw clusters. With 12 clusters, 3 gradients and one palette, the number of possible images is almost 2 billions. The color distribution is exported automatically in the "palettes" directory with RGB values and R code to use these colors in R plots.
- The "Hide Legend" checkbox, hide the legend in both cluster and expression views.

0.2.3 Drawing gates on the microscopy image

Single-Cell Spatial Explorer features can work with an unlimited number of gates.

- Select the brush or the polygon. The selected tool is green.
- To draw a gate with the brush, left click on the image with the mouse and drag the mouse. Then release the mouse button. The gate is automatically closed between the beginning and the end, it is not necessary that the beginning and the end are at the same place.
- To draw a gate with the polygon, click on the image, a dot appears, move the mouse and click again. To close the shape, right click. The gate is automatically closed between the beginning and the end, it is not necessary that the beginning and the end are at the same place.
- To clear all gates, click on the "clear all gates" button
- To clear the last gate, click on the "clear last gate" button

0.2.4 Exporting gates

Gates are exported in CSV format in the "gates" directory using the "Save gates" button. The name of the gate is the number of the gate, followed by the name in the text entry of the tool window. Gates are also automatically exported when the tables are filtered by gates. cf. 0.2.5

Importing gates

Single-Cell Spatial Explorer can import an unlimited number of gates. A gate file is a CSV file with a X,Y header followed by the edges coordinates of the polygon.

```
X,Y  
851,891  
1149,833  
1171,1073
```

This format is compatible with ImageJ/Fidji image analysis software.

The gates files have to be copied in the "import_gates" directory and then imported using the "Import Gates" button

A macro can be used to export the regions of interest : Macro ImageJ/FIJI to export Region Of Interest (ROI) coordinate from ROI manager:

```
for (i=0 ; i<roiManager("count"); i++) {  
    roiManager("select", i);  
    roiManager("Set Color", "yellow");  
    roiManager("Set Line Width", 0);  
    saveAs("Results", "MyFolder/XY_  
    Coordinates_"+i+".csv");  
}
```

0.2.5 Filtering data with gates

High throughput data extraction with Single-Cell Spatial Explorer video.

Single-Cell Spatial Explorer can filter an unlimited number of data tables in the "data" directory using the "Filter tables with gates" button. Please note that if more than one table are placed in the "data" directory, all the tables will be filtered, but only the first one by alphabetical order will be used to display clusters, expression, comparisons... The filtered tables and the cell names are in the "results" directory and the gates are automatically saved in the "gates" directory using the name in the text entry of the tool window. The names of the sub-tables can be set in the text entry on the top of the 2D plot tool window. Two types of files are produced : "cells" files contain the cell names and "filtered" files contain the sub-tables. The final file names are a string containing the following items : "cells/filtered_gate number_ text entry_orginal table name". In the sub-tables, a column labeled "GateNumber" is added at the end of the table.

0.2.6 Plotting gates

The cells in gate can be plotted with any 2D coordinates using the "Plot gates" button (cf. fig. 4). Do not forget to set the dots colors and the dots in background colors. If the dots colors is white, they will not appear in the scatter plot. An optimized set of colors is used first. If the number of gates exceed the number of preset colors (10), then the "Turbo" gradient is used to compute an unlimited number of colors through the rainbow. The plot are saved in the "plots" directory.

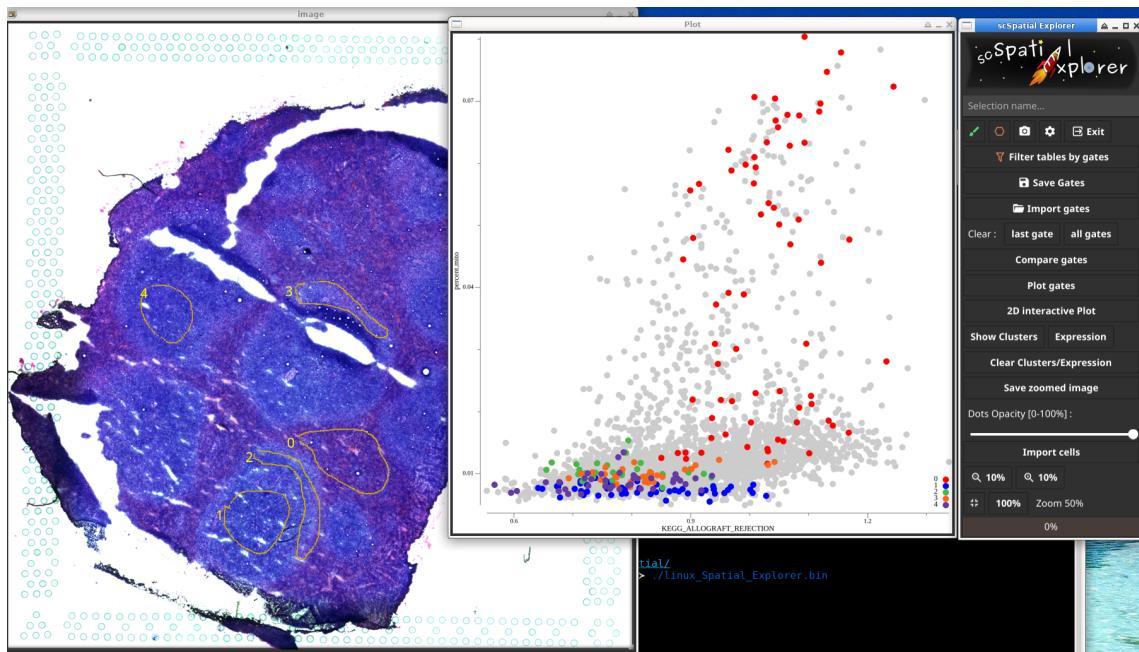


Figure 4: Single-Cell Spatial Explorer "Plot gates" window. Selected cells on the microscopy image can be plot with any system of XY coordinates, t-SNE, UMAP, gene-gene, pathways, antibodies ...

0.2.7 Comparing gates

Comparison of gates involves a Mann-Whitney U test corrected by Bonferroni. Gates and groups of gates can be compared using the "Compare gates" button. This button opens a "Compare" window (cf. fig. 5).

First enter a filename on the top of the window. It will be used to save the comparison files. If no name is entered, the default filename will be "comparison" followed by the time. The columns used for the comparison can be selected using check boxes. Cells inside two gates or two groups of gates can

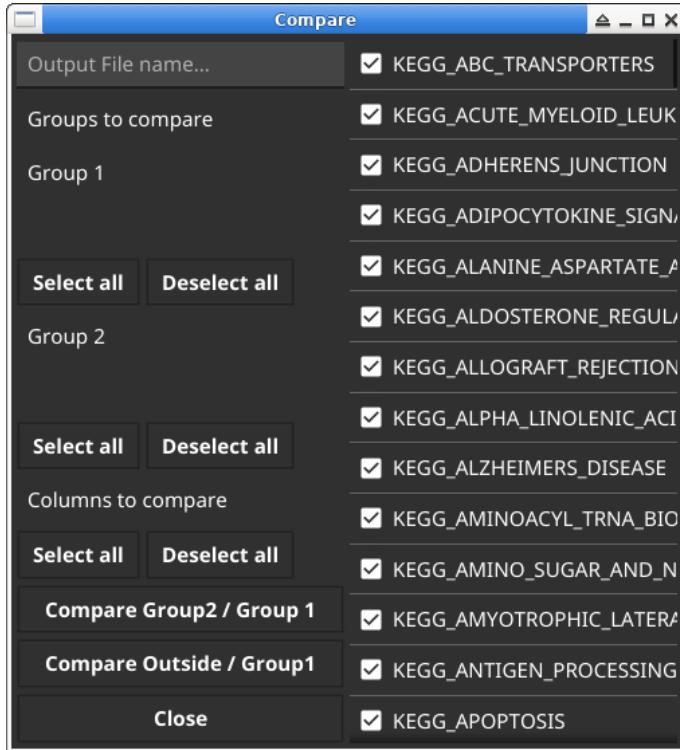


Figure 5: Single-Cell Spatial Explorer compare gates window.

be compared using the "Compare Group2 / Group1" button. Cells inside a gate or a group of gates can be compared with the cells outside the gates using the "Compare Outside / Group1" button. These buttons open an interactive volcano plot and a volcano tool window. To save the volcano plot, click on the "Save volcano plot" button. The volcano plot data are automatically saved in the "comparison" directory. A left click on the volcano plot display a selection square (the size is set in preferences). A right click list the dots names, fold change and pvalue in the volcano tool window. In the volcano tool window, it is possible to select a color gradient, a dot opacity (do not forget to set the dot opacity between 50-100% to see the dots) and a dot in the item table. The "Show expression" button will display the expression associated with the volcano dot on the image.

0.2.8 2D interactive plots

It is possible to display an interactive scatter plot by clicking on the "2D interactive plots" button (cf. fig. 6). Any 2D coordinates can be used, such as t-SNE, UMAP, gene expression, antibody signal etc... It is possible to draw gates on the 2D interactive plots. Left click on the image with the mouse and

drag the mouse. Then release the mouse button. The gate is automatically closed between the beginning and the end, it is not necessary that the beginning and the end are at the same place.

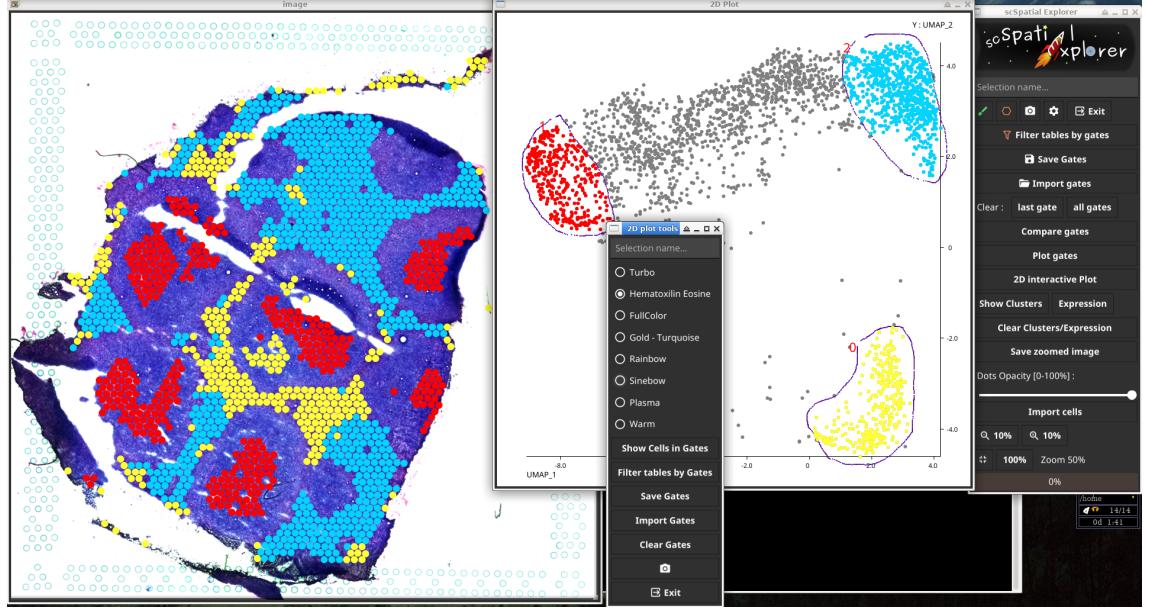


Figure 6: Single-Cell Spatial Explorer 2D interactive plot. The screenshot shows how cells can be selected on a UMAP plot, but any system of XY coordinates can be used.

1. Once the gates are drawn, it is possible to display the selected cells on the image by clicking on the "Show Cells in Gates" button.
2. Using the button "Filter tables by Gates", an unlimited number of data tables, in the "data" directory can be filtered to extract the sub-tables corresponding to the cells in the gates. The gates are automatically saved in the "gates" directory. The names of the sub-tables can be set in the text entry on the top of the 2D plot tool window. see 0.2.5.
3. To save gates (without filtering the tables), use the "Save Gates" button.
4. To import gates, use the "Import Gates" button. The gates must be copied in the "import_gates_2Dplot" directory. A gate file is a CSV file with a X,Y header followed by the edges coordinates of the polygon.

X,Y
851,891
1149,833
1171,1073

This format is compatible with ImageJ/Fidji image analysis software.

5. To clear gates, use the "Clear Gates" button.
6. To save a screenshot of the 2D plot press the camera button.

0.2.9 Displaying cluster

To be able to show clusters on the microscopy image, it is necessary to identify the clusters column in the preference setting. Then press the "Show Clusters" button. The colors gradient can be selected in the preference setting. When "Shuffle colors" is check, a new distribution of colors is obtained following repeated click on the "Show Clusters" button.

The color palettes are saved in the "palettes" directory, thus it is possible to use the colors in another plotting software. For very experienced users, an unlimited number of palettes can be placed in JSON format in src/palette (see Preferences). We do not recommend to modify the src/palette/palette1.json file. It is better to make a copy of this file and work on the copy. A software restart is required to make custom palettes appear in the preference form.

With 12 clusters, almost 2 billions of different images can be obtained with the 4 gradients and the shuffle colors.

If you select "Use multithreads for cluster" in the preference settings, a separate transparent image of each cluster will be available in the "temp /clusters/" directory just after cluster computation. Those images can be stacked in an image processing software such as Gimp to construct an image with only some clusters.

0.2.10 Displaying expression

Expression can be genes expression, antibody signal, gene sets expression, or any numerical data present in the data table.

The buttons "Previous slide" and "Next slide" are reserved to the slide show and cannot be used to navigate among expression items.

To display expression data on the microscopy image, press the "Expression" button. A window open (cf. fig. 7) :

1. on the top of the window, select a column of the data table in the drop down list and press "submit".
2. select a color gradient
3. if necessary change the legend color using the "Legend Text Color" button
4. press the "Plot Expression" button
5. the expression is displayed on the microscopy image. The dot opacity, and the image contrast can be adjusted with the sliders "Dots opacity", "Min", "Max". Press the "Apply" button to apply changes.

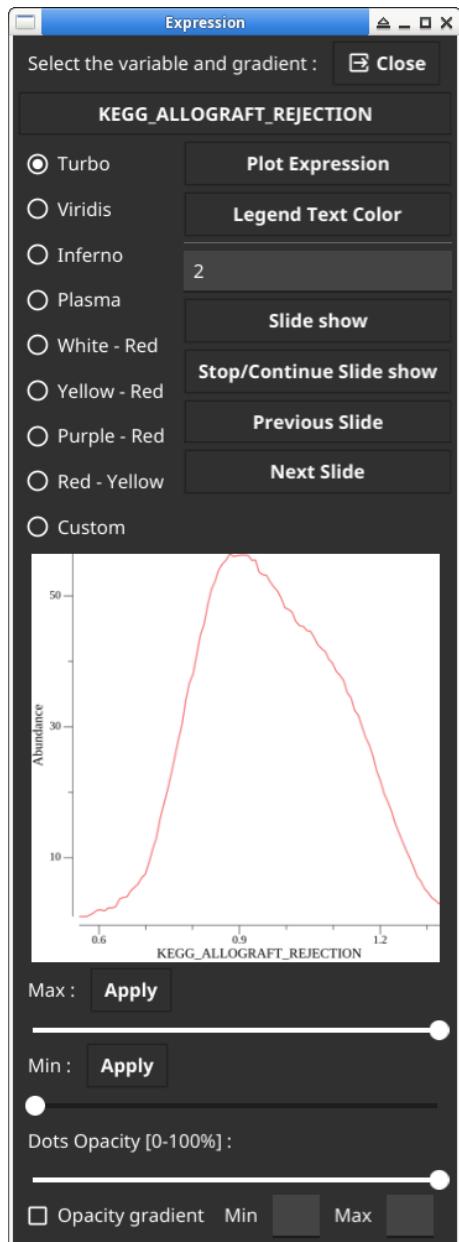


Figure 7: Single-Cell Spatial Explorer expression window.

6. when the check box "Opacity gradient" is selected, the opacity for each spot is computed from the expression value (cf. fig. 8). The min expression value correspond to a full transparency and the max a full opacity. The

min/max opacity threshold must be entered in the entry spaces. For example, if min opacity = 1.5, the spots with an expression ≥ 1.5 are fully transparent. If max opacity = 3.5, the spots with an expression ≤ 3.5 are fully opaque. When min opacity is closed to max opacity a kind of binarisation is obtained and only the spots with an expression between min and max are visible. min opacity must be superior to max opacity otherwise the software ignore the opacity computation.

7. In the expression window, a 2D plot show the distribution profile of the dots intensities. This plot is obtained by dividing the expression levels into 100 classes that are smoothed by a moving average on 20 points.

8. It is possible to use a custom gradient. Make a backup copy of the file /src/gradient/custom.csv, then edit it and enter hexadecimal colors. For example #FFFF33 is yellow.

The file must be formatted as follow. The filename must be "custom.csv", other filenames will be ignored but can be used to store many custom gradients. CSV file without header and with only one column, column separator (TAB), one hexadecimal colors per line. The color on the top of the list is the start of the gradient. Colors are interpolated between 2 consecutive colors. Example :

```
#6600CC
#FF6033
#FFFF33
```

Slide Show

The slide show is a way to review many expression maps without the need to select each item one by one². The Min/Max sliders and opacity gradient are reserved to expression and cannot be used in the slide show. To improve an image for a publication for example, use the procedure described previously in § 0.2.10.

To do a slide show :

1. Enter the pause between slides in the "Pause between slides" entry
2. Press the "Slide show" button
3. to stop/continue the slide show : press the "stop/continue slide show"
4. to show next/previous slide : press the "Next slide" / "Previous slide" buttons

²Please note that when the slide show is running, the title of the expression is not updated on the interface. Indeed, for technical reasons, Single-Cell spatial Explorer do not use drop down lists. Drop down lists have a limited capacity. They have been replaced by a button linked to a table, to have a quick and unlimited access. There is currently no way that the button open the table at a given position.

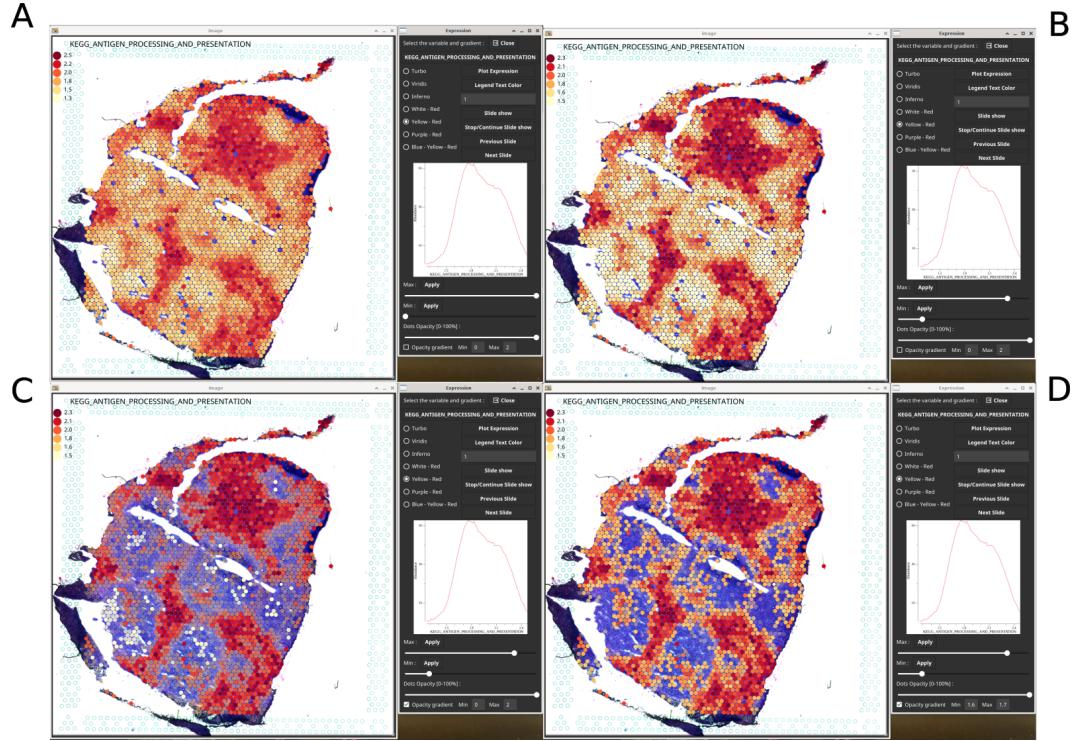


Figure 8: Single-Cell Spatial Explorer displaying a biological function with different settings of contrast and opacity. A) original image. B) Contrast was increased with the min/max slider. C) Same image as B) with a wide opacity gradient. D) Same image as C) with a narrow opacity gradient.

0.2.11 Importing Cells

An unlimited number of files containing cell names can be placed in the "import_cells". The files must be text file with tab separator (extension does not matter and can be csv/tsv) with one cell name per line without header. import and display an unlimited number of cells list by repetitive click on the "import cells" button. The format is directly compatible with Single-Cell Virtual Cytometer.

0.2.12 Saving images

There are two ways to save images : a screen shot or a high resolution image.

1. Enter a filename in the main toolbox (see fig. 2 - "name of output files")
2. adjust the zoom. For a screenshot, all the image must be visible in the image window. To export a high resolution image, it is recommended to

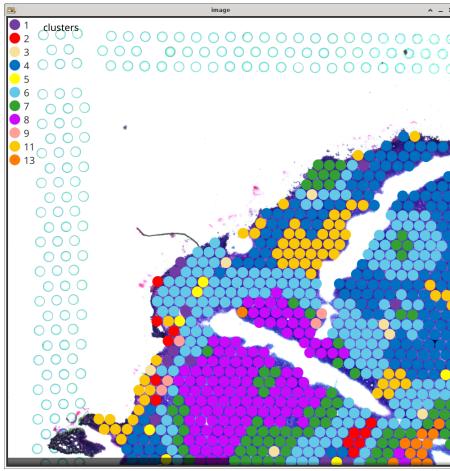


Figure 9: This screenshot show the image position required to export a high resolution image. Move the image in order to see the top left corner of the image. This is done by moving the XY scroll bars in the bottom and the right of the image window. The horizontal bar used to move the image laterally is visible on the bottom of the image.

apply a 100% zoom to obtain the best quality and to move the image in order to see the top left corner of the image. This is done by moving the XY scroll bars in the bottom and the right of the image window (see fig. 9).

3. draw your image (cluster, expression...)
4. click on the screenshot icon (camera) or on the "Save zoomed image" button. The progress bar monitors the exporting task.
5. the image is saved in the "plots" directory in png format (other formats are not supported by the graphical library).

0.2.13 Software compilation from source

To compile the software from source :

1. install Fyne dependencies for your operating system. See also the prerequisites of the Fyne graphical toolkit.
2. install Go programming language
3. rename the software folder "spatial-main" into "spatial"
4. all the needed packages for compilation are listed in go.mod and go.sum, so they are automatically downloaded during the compilation process. Open

a terminal in the software main directory (spatial/), type :
go mod tidy
go build .

5. a "spatial" binary file is build and can be executed.

0.2.14 Tips

How to save memory on a small computer ?

Each spot on the cluster/expression map is a small object using RAM. We recommend to save the cluster/expression picture as soon as possible and then to use the button "Clear Cluster/Expression" to release RAM. This is especially useful before doing gate comparisons. The gates comparison is the most RAM demanding computation because the whole dataset have to fit in RAM if the user select all the dataset columns in the comparison.

How to calculate the scaling factor after resizing an image ?

If an image is divided by a factor n, the scaling factor is $\frac{1}{n}$.

Example : the size of the initial image is 40000×30000 , a factor 10 reduction gives an image of 4000×3000 and the scaling factor is $SF = \frac{1}{10} = 0.1$.

Working with large data sets such as Akoya Phenocycler

Single-Cell spatial Explorer was designed to work with spatial transcriptomics data producing thousands of spots. Some technologies using microscopy and antibodies produce millions of spots.

Since version 2023-08, the rendering engine of Single-Cell spatial Explorer is able to display millions of spots, but the image have to be resized to a reasonable size, for example 4000×3000 .

Using Single-Cell spatial Explorer with antibodies based technology is possible (cf fig 13 15) but require some precautions.

When the data sets produce millions of spots we recommend the following strategy :

1. Do not over-click : displaying millions of spots takes time, you can watch image calculation on the progress bar of the main window. Over-clicking will not speed calculation, worse, it will start many processes in parallel and crash the software.
2. Use a computer with a fast video card and an optimized driver (we do not recommend to use the "nouveau" driver on linux with Nvidia cards, it is not fast/stable enough).
3. reduce the size of the image, for example 4000×3000 , or extract an interesting part of the image. Convert the image to PNG.

4. It is possible to use a black png image if your are not interested to interact with the microscopy image. A black background produce more contrasted images (see fig. 14)
5. If a resized microscopy image is used, the scaling factor has to be calculated and eventually to be adjusted using the cluser/expression view (see fig 12)
6. A resized microscopy image can be used without resampling. In the example fig 13, the image was reduced by a factor 10 and 2.2 millions of spots have been plotted with a diameter set to 1 in the software preferences.
7. If you have millions of spots and want a quick view of the data, reducing the number of spots by resampling is a good strategy. Sort the data set by XY (see fig 14) and then use the resampling function (see fig 10 11) to reduce the number of spots and to convert at least XY columns to integers. Indeed, Akoya at least produce coordinates in centroid, not in pixel. **Some software, such as Seurat, export some numbers enclosed by double quotes, remove them or the software will crash.** A copy of the original data file will be produced with the prefix "0_". The number of spots displayed will be reduced, but the data extraction by gates will be done on both files. Indeed, the data can be extracted on an unlimited number of files with (almost) unlimited number of cells, since extraction is done line by line. If you want to use a database instead of the CSV file after resampling, it is necessary to restart the software to convert the new CSV file to a database.
8. In the preference window, increase the number of threads for expression. For example, with 2.2 millions of spots the optimal number of threads is ~ 10 for a $3\times$ speed gain.

0.2.15 Troubleshooting

The software crashes when I display clusters or expression.

It is very important to test first the software with tests files and the preferences settings with the test files. If the software works with the test files and crashes with other data files, the possible causes are :

1. The cluster column is not found or does not contain Integers. Floats numbers are not allowed. (replace "2.0" by "2" for example)
2. preferences are not set for the data file, some columns cannot be found
3. the data table is not formated as required or contains special characters, blanks or strings instead of numbers
4. the datatable contains strings instead of numbers. This is a frequent problem when some cells coordinates are removed by quality controls. Check your table, it should not contains strings, except cells tags. Remove the rows containing "NA", "NaN", "missing" ...



Figure 10: Resampling button, on top of the window, for large data sets

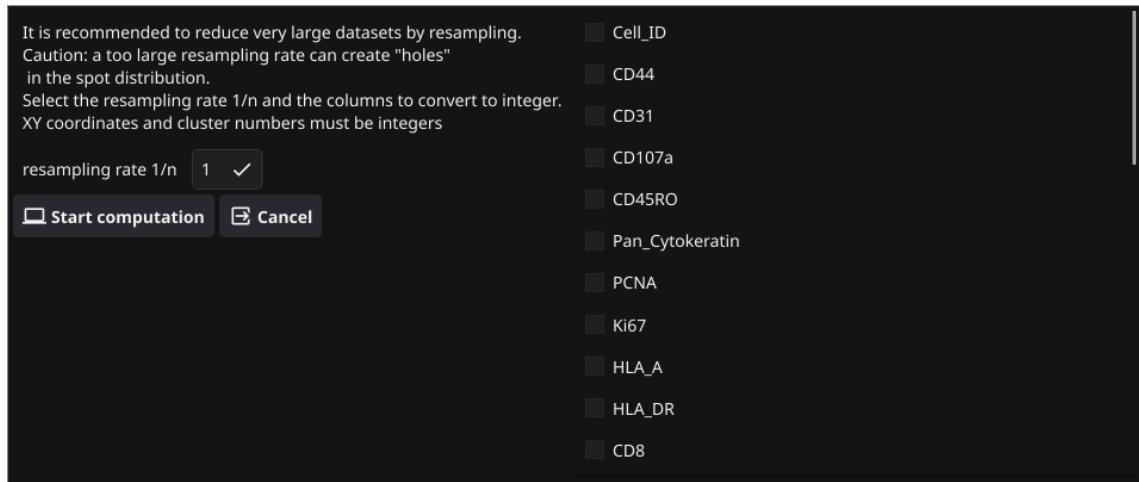


Figure 11: Resampling tool for large data sets. Some columns must be integers. For example XY coordinates and cluster number. Select the columns to convert to integers and the resampling rate. A resampling rate of 1 does not reduce the size of the output file, only selected columns are converted to integers. A sampling rate of 10 will take the line 1,10,20... of the original file.

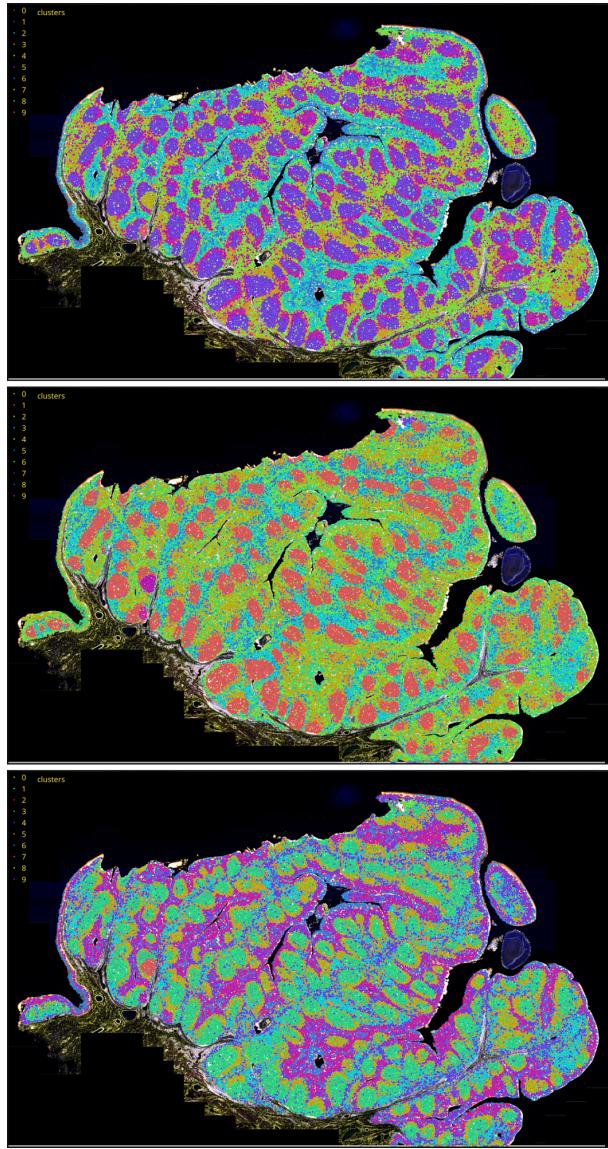


Figure 12: Cluster views obtain with Akoya large data set of > 2 millions of cells after resizing the background image and resampling ($factor = \frac{1}{10}$ ie $\sim 220.10^3$ cells.)

5. Do not open a data table with a spreadsheet while running the software. Spreadsheet (libreoffice, openoffice, Excel...) that create an invisible .lock file that cannot be processed in the data directory.

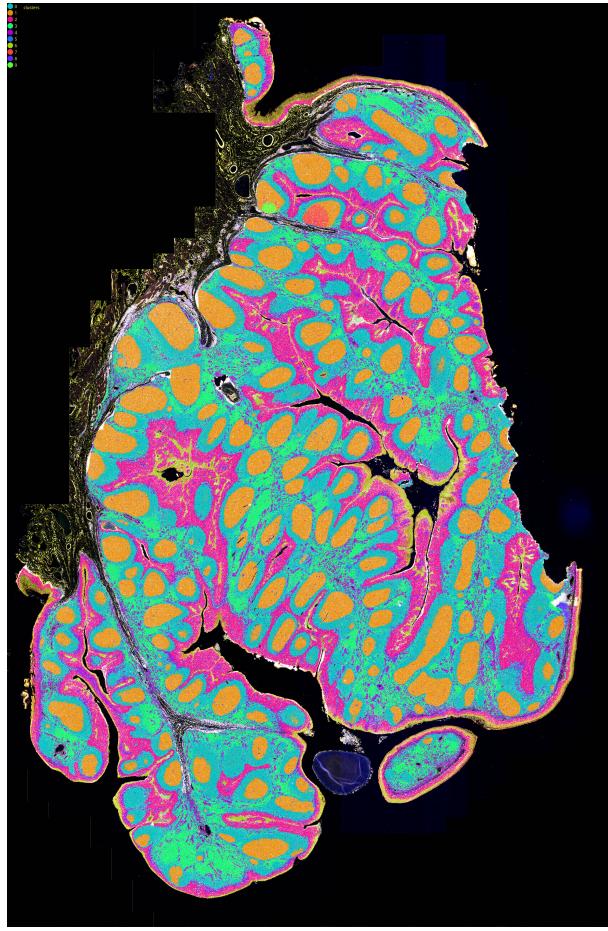


Figure 13: Cluster views obtain with Akoya large data set of > 2 millions of cells after resizing the background image (factor 1/10) without resampling.

The software do not start on Linux with a Glibc error.

The precompiled binaries for linux have been compiled on Manjaro Linux release 21.2.4 (up to 2022-11-16), Ubuntu 22.04.2 LTS (since 2023) and Xubuntu 20.04 LTS. The Manjaro/Ubuntu 22.04 binary is for recent linux distribution. The Xubuntu 20.04 LTS is for distributions of less than 2 years old. On an older Linux distribution, it is possible that Glibc is outdated for the binaries we provide. In that case a linux update is needed or the software can be recompiled from source (see § 0.2.13).

There is a shift between the image and the cluster/expression map.

1. the scaling factor is wrong

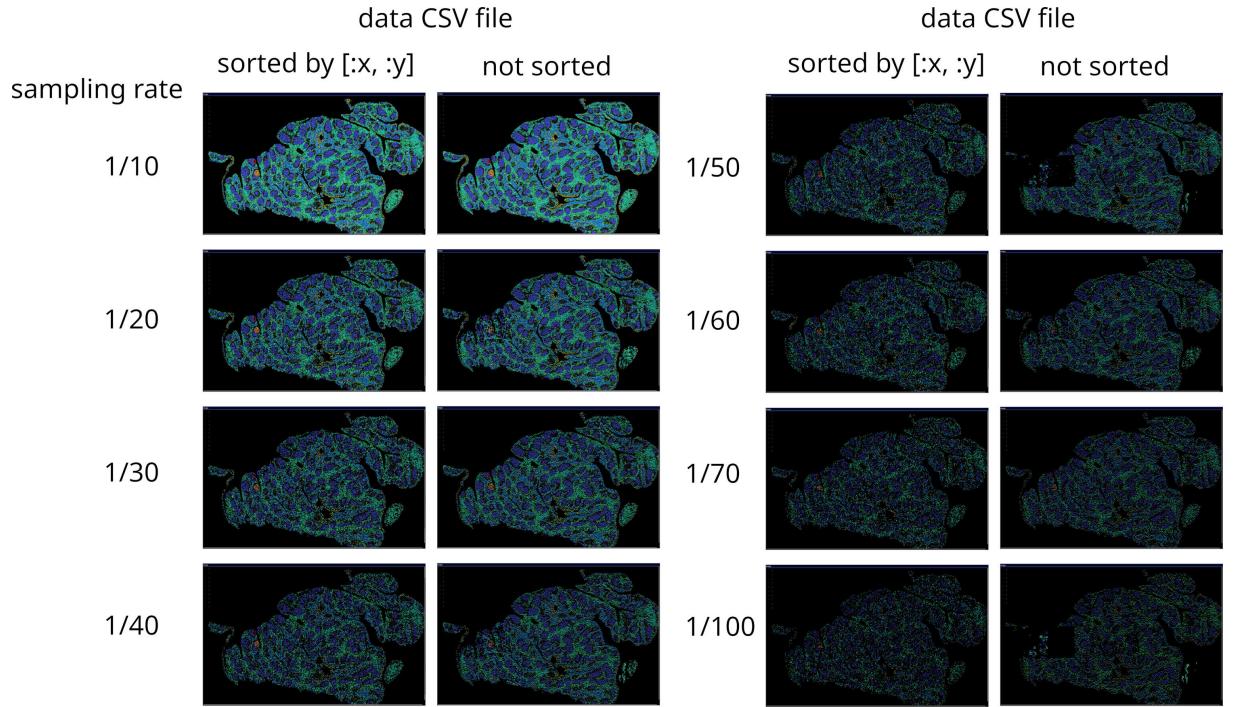


Figure 14: Benefits of sorting the data set by XY coordinates before resampling with Akoya large data set of > 2 millions of cells

2. the coordinates in the data file do not match the pixel of the image. Check the resolution of the image and the coordinates column in the data file.
3. A shift can appear if the title of the graph is too long. This issue is normally solved, the software trim the title when it's size is larger than the image.

My scatter plot is not displayed or partially displayed

1. Check the colors of the dots in gate and dots in background. A white spot is not visible over a white background.

The software is slow

The execution speed is related to the size of the data table. The number of rows but also the number of columns have a speed impact. To increase speed :

1. close other softwares, verify that the RAM of the computer is not saturated and that the disk is not used by another process (anti-virus scanning for example).

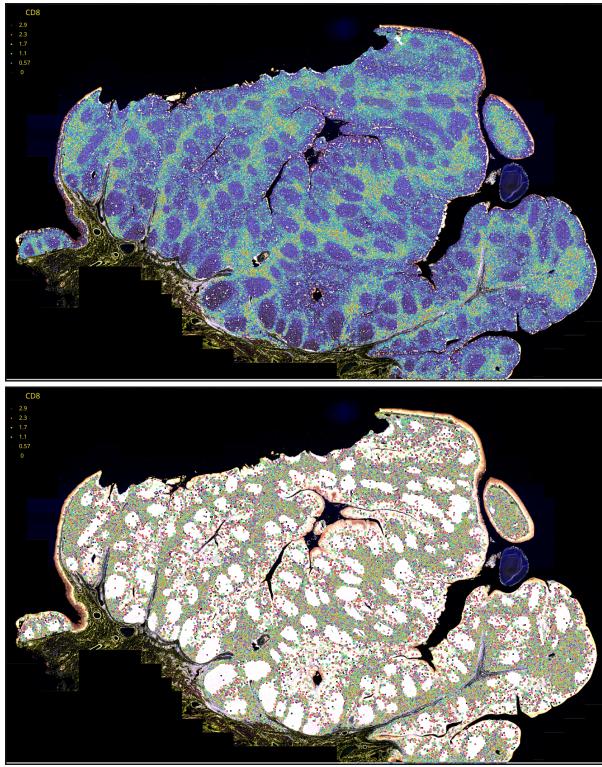


Figure 15: Expression opacity gradient on Akoya large data set

2. cut a large data table to decrease the number of columns
3. copy the software on the fastest disk, SSD if possible.
4. If the CSV file has many rows/columns, set preferences to convert the CSV to database
5. If the CSV file has hundred of thousands of rows, set preferences to use multithread (= using more than 1 processor).

The preference file is corrupted or not compatible with an older version of the software

As the software evolves, more preferences are stored to the preference file and may conflict with old ones. The location of this file is indicated just below the title in the Single Cell Spatial Explorer terminal. On Linux the preference file is generally located in : /home/user/.config/fyne/SpatialExplorer/preferences.json

1. Make a backup copy of the preference file if necessary to avoid loosing some parameter such as the scaling factor.

2. Remove the preference file.
3. Restart the software and set the new preferences.

The legend mask the image

The cluster/expression legend is plotted in the left side of the image. The optimal position and size of the legend depend on the margin size of the image and the number of clusters. Some image can have no margin. Since Single-Cell spatial Explorer is an interactive drawing software, the image cannot be resized to optimize the legend position such as non-interactive softwares. In the situations were the legend mask some interesting parts of the image a solution to make a high quality publication figure is :

1. In the preference settings, click on "Hide Legend"
2. Scale the image to fit the image window (screen resolution) or, for the best resolution set the zoom to 100%
3. Draw the cluster/expression without the legend
4. Save the image by clicking on the camera icon (screen resolution) or on the "save zoomed image" button for a native resolution
5. close the software and replace the microscopy image by a blank PNG image not wide, but high enough to contain the legend, in the "image" directory (for example 500×4000).
6. In the preference settings, deselect on "Hide Legend"
7. Draw the cluster/expression with the legend and save the legend. The legend is now in another image that can be cropped and scaled to be used in a publication.
8. If there are still collisions between the spots and the legend,
 - use another X or Y coordinate for the image spots, for example t-SNE, this will shift all the spots. The legend color is independent from spots positions, so the spots can be plot far from the legend.
 - the legend can also be drawn manually using a software such as Inkscape .

Bibliography

- [1] Pont, F., Tosolini, M., and Fournié, J. J. (2019) Single-Cell Signature Explorer for comprehensive visualization of single cell signatures across scRNA-seq datasets. *Nucleic acids research*, **47**(21), e133–e133.
- [2] Pont, F., Tosolini, M., Gao, Q., Perrier, M., Madrid-Mencía, M., Huang, T. S., Neuvial, P., Ayyoub, M., Nazor, K., and Fournié, J.-J. (2020) Single-Cell Virtual Cytometer allows user-friendly and versatile analysis and visualization of multimodal single cell RNAseq datasets. *NAR genomics and bioinformatics*, **2**(2), lqaa025.