Tutorial

Julien Moehlin, Bastien Mollet, Bruno Colombo, and Marco Antonio Mendoza-Parra
UMR 8030 Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, University of Evry-val-d'Essonne, University Paris-Saclay, 91057 Évry, France.

I - Installation & run

Multilayer is runnable on every os (windows, linux mac osx) with python 3.
Python version 3.8 is recommended.

1 - Dependencies

- Numpy   pip install numpy
- Matplotlib   pip install matplotlib
- Pandas   pip install pandas
- Scipy   pip install scipy
- Scikit-learn   pip install scikit-learn
- Seaborn   pip install seaborn
- Networkx   pip install networkx
- Louvain   pip install python-louvain

2 - Run

python3 Multilayer.py

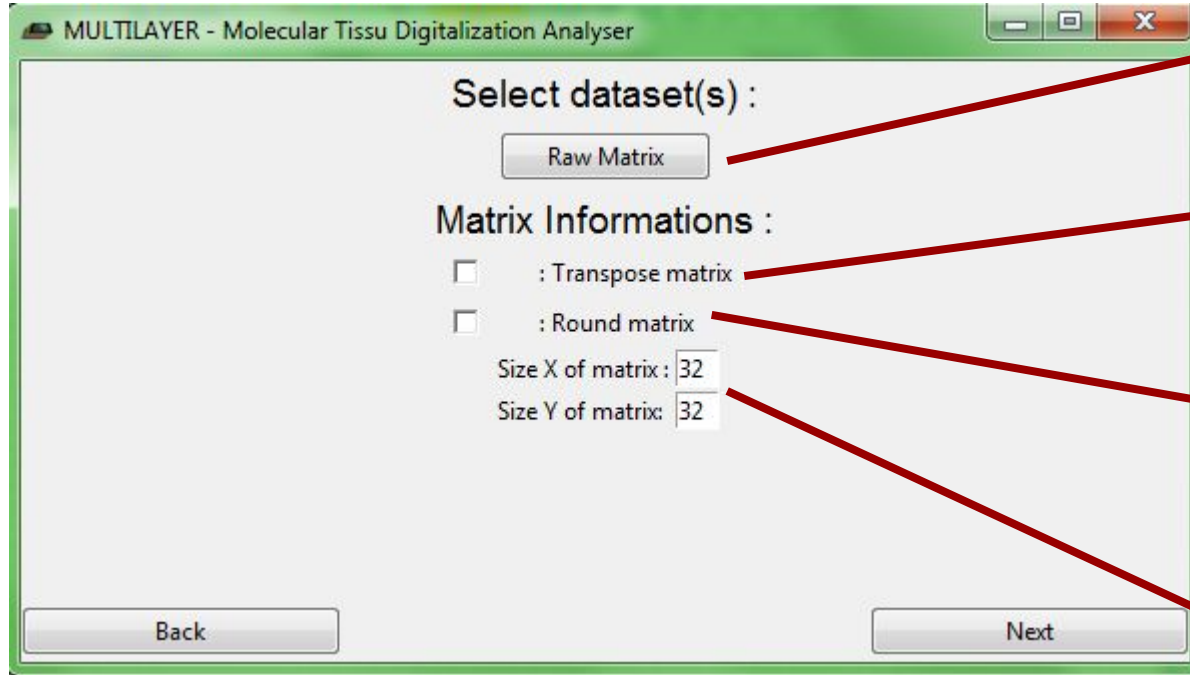# II - Explore molecular tissu with MULTILAYER

II - 1 - Mode
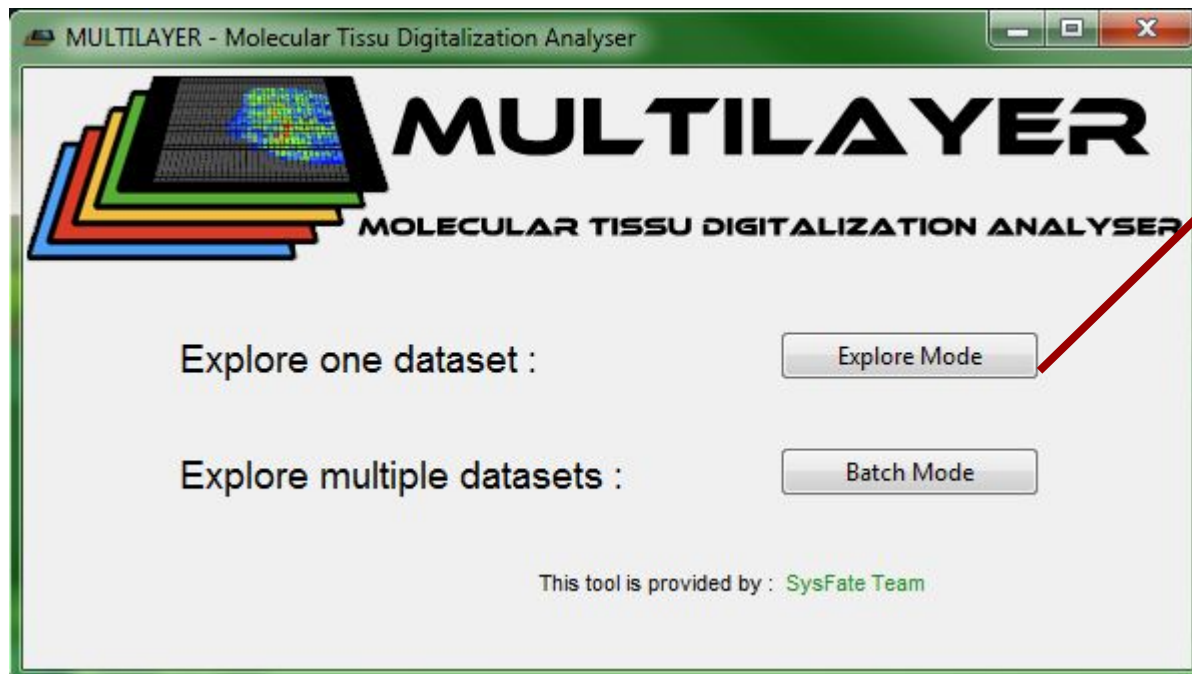
Explore mode : perform analysis over one sample (.tsv file).

Provide your dataset (.tsv file). The correct format is a matrix with genes name as rows, coordianates (coordinates should be in this format : 'XxY') as columns.

Transpose the dataset is an option. You can use this option if you have a matrix with genes name as columns and coordinate as rows.

Round matrix is an option. The tool needs integer as X & Y (coordinates). This option will round coordinate. For example : a given coordinate '3.36x13.94' will be converted to like this : '3x13'.

Indicate the **size** of dataset. The maximum of X and the maximun of Y. Default value is 32.

The tool will perform a differential gene expression analysis. Those thresholds are for define up-regulated genes and down-regulated genes.

Pattern detection : The tool will use agglomerative clustering on gexels (https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html) for determine a contiguous pattern. The tool will consider a pattern if the number of gexels is upper or equal than this value. Default value is 10.
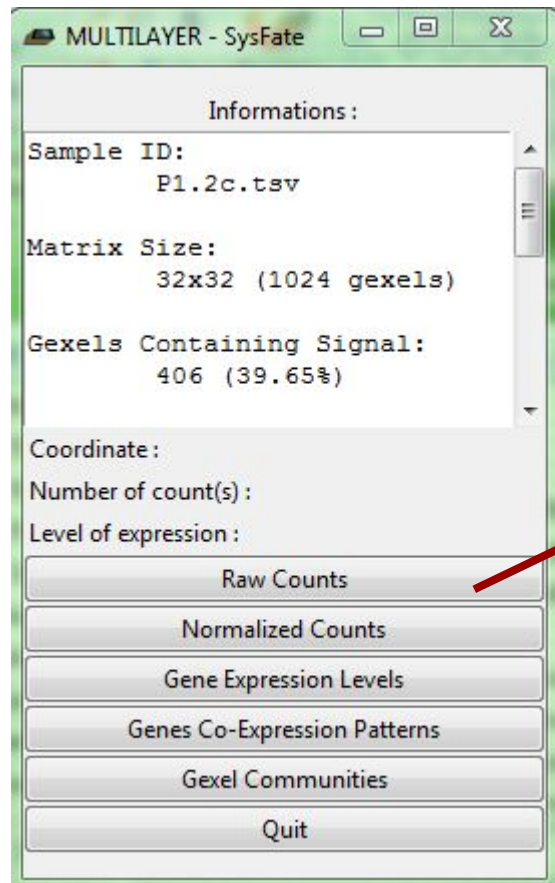
Similarity methods will calculate similarity between patterns.
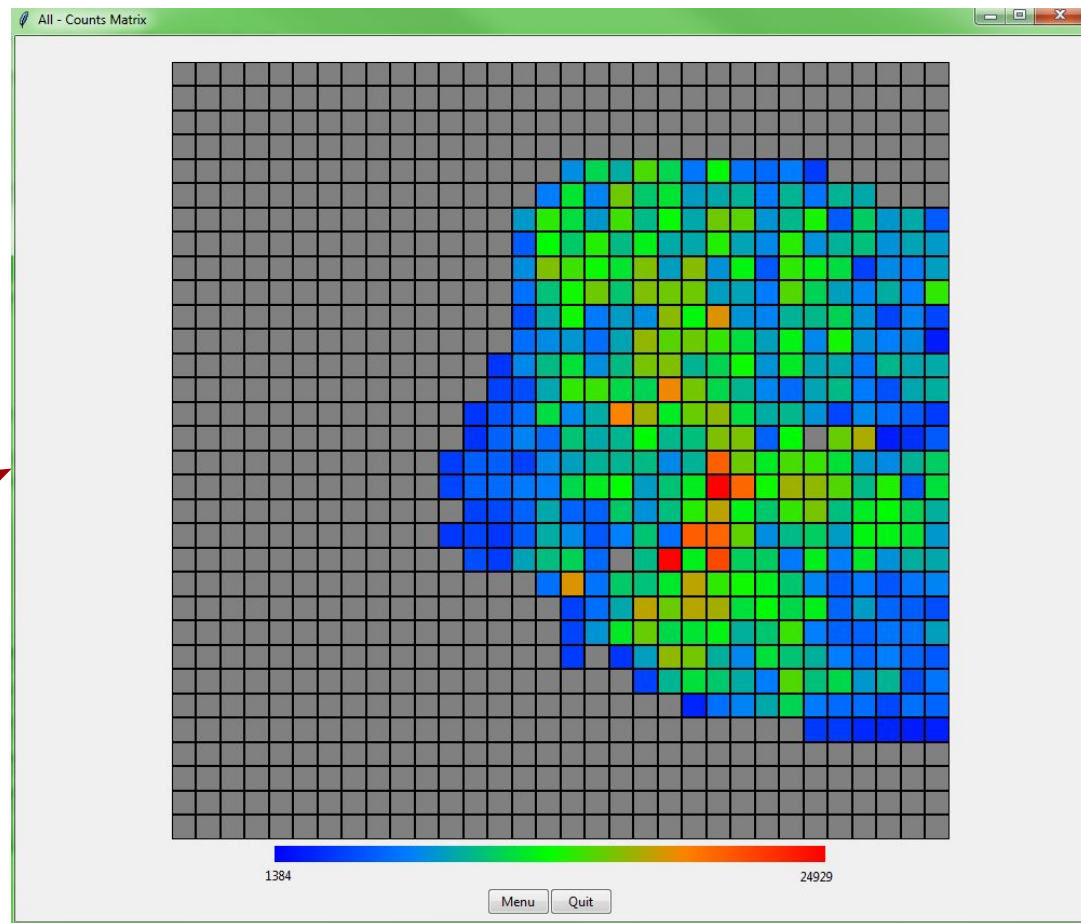With Tanimoto, the similarity will be calculated with A Union B / A Intersection B.
With Dice the similarity will be calculated with 2*(A Union B) / A Intersection B.

Multiprocessing : provide the number of threads for the analysis (always available for the Batch mode)

MULTILAYER - Molecular Tissu Digitalization Analyser

Arguments for differential gene expression :

Threshold up regulated : 1

Threshold down regulated : -1

Patterns detection :

Minimum n° of contiguous gexels : 10

Similarity Methods :

◉ Tanimoto

○ Dice

Multiprocessing :

Number of thread : 1 max (4).

Back

Next

II - 3 - Raw dataset

MULTILAYER - SysFate

Informations :

Sample ID:
    P1.2c.tsv

Matrix Size:
    32x32 (1024 gexels)

Gexels Containing Signal:
    406 (39.65%)

Coordinate :

Number of count(s) :

Level of expression :

[Raw Counts]
[Normalized Counts]
[Gene Expression Levels]
[Genes Co-Expression Patterns]
[Gexel Communities]
[Quit]

'Raw counts' shows the input dataset.
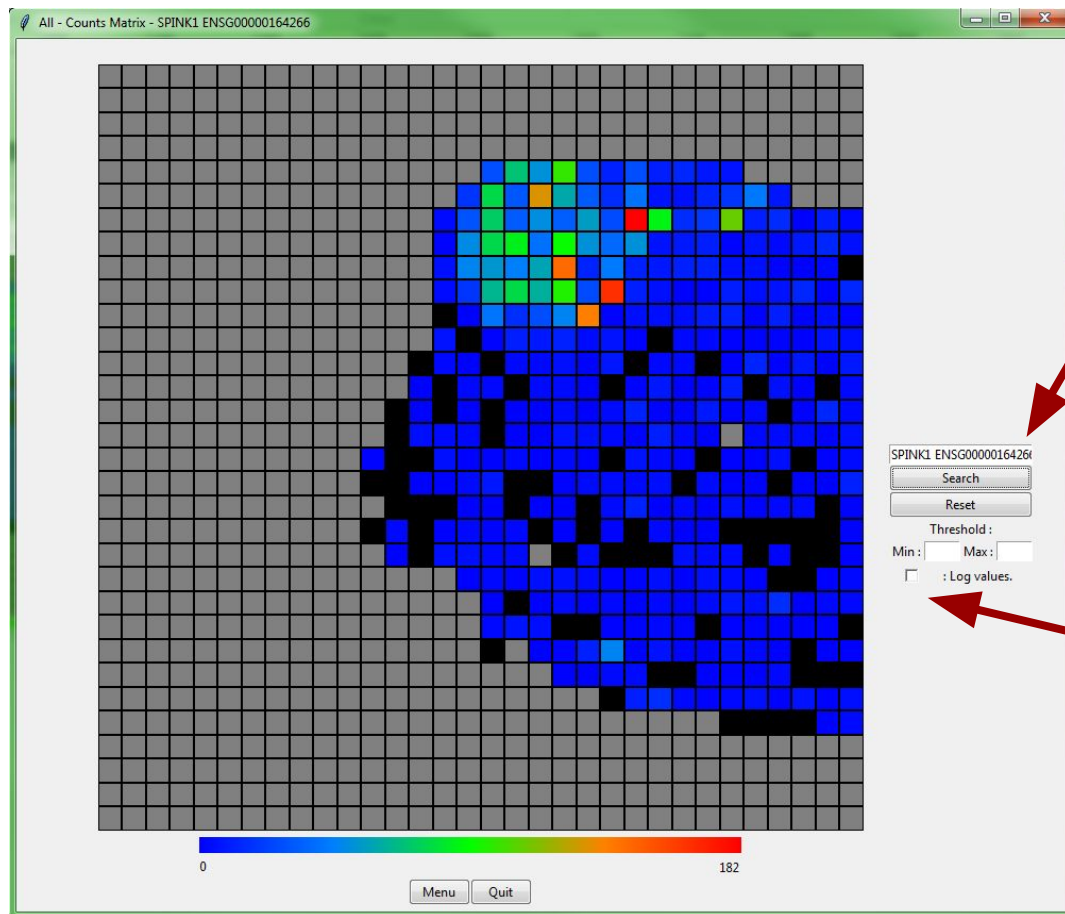
All - Counts Matrix

1384   24929

[Menu] [Quit]

The right panel appears when you hit the menu button.

You can query a gene.

Display gene with log values

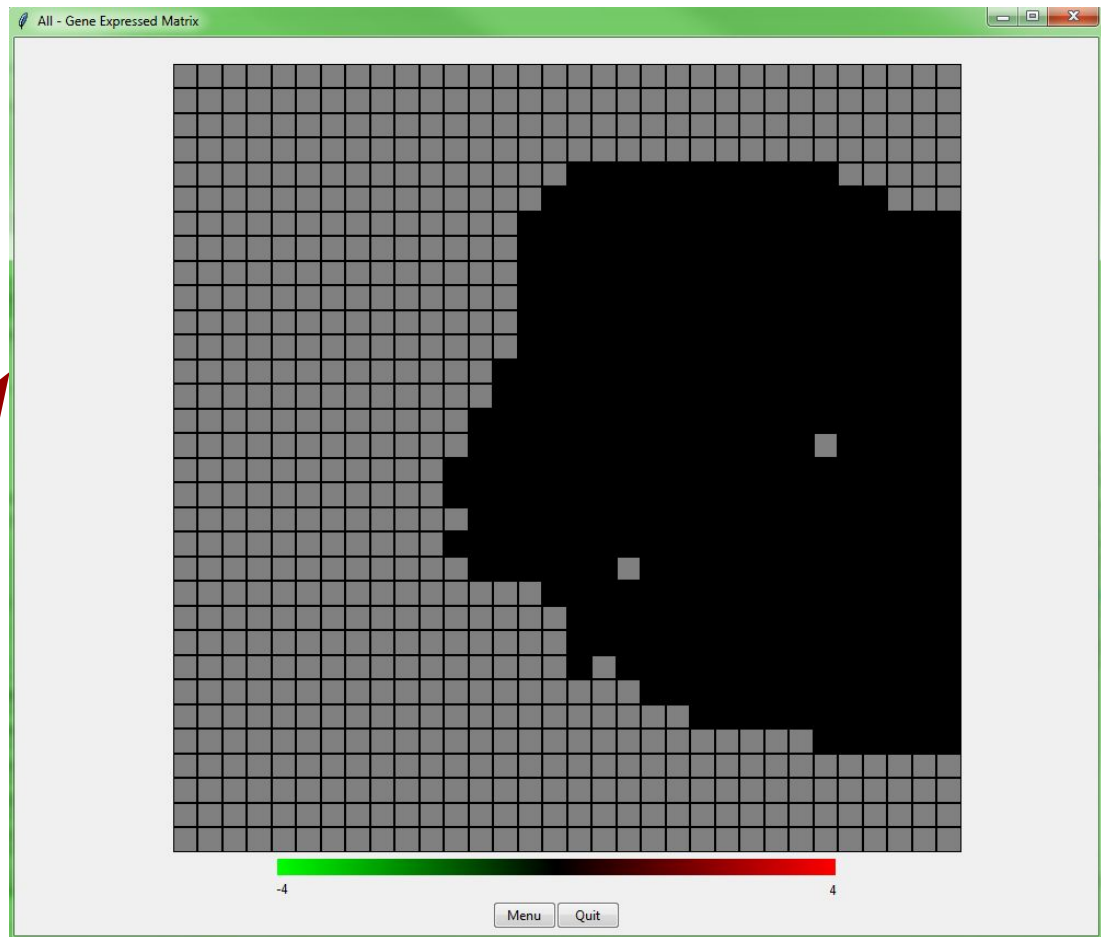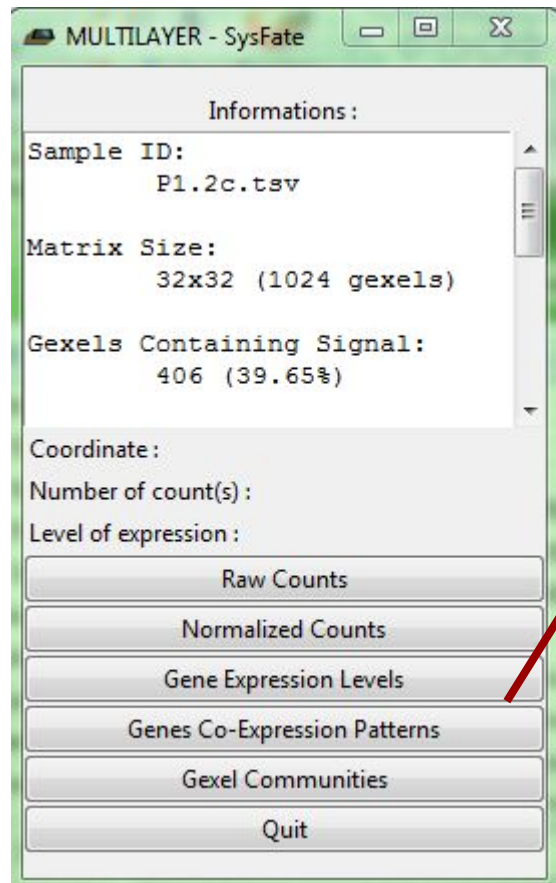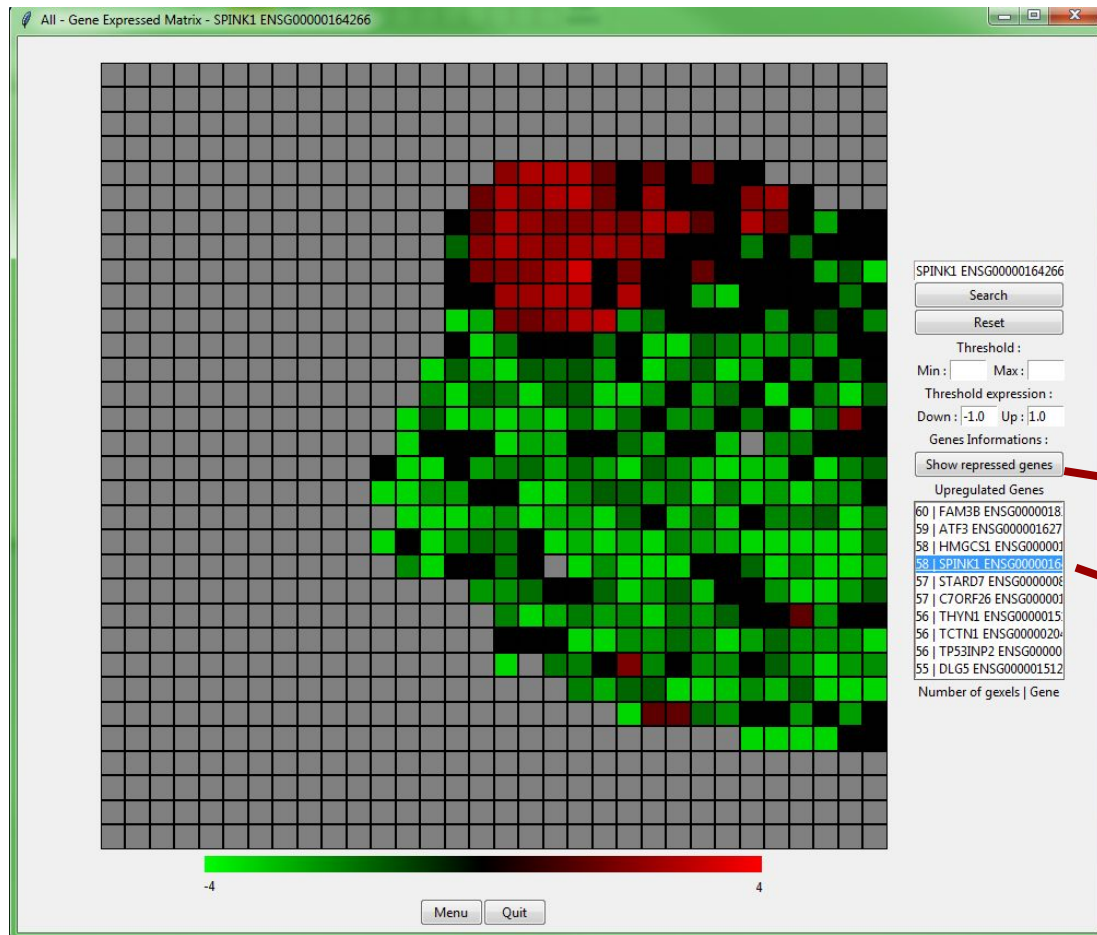'Normalized counts' shows the normalized dataset.
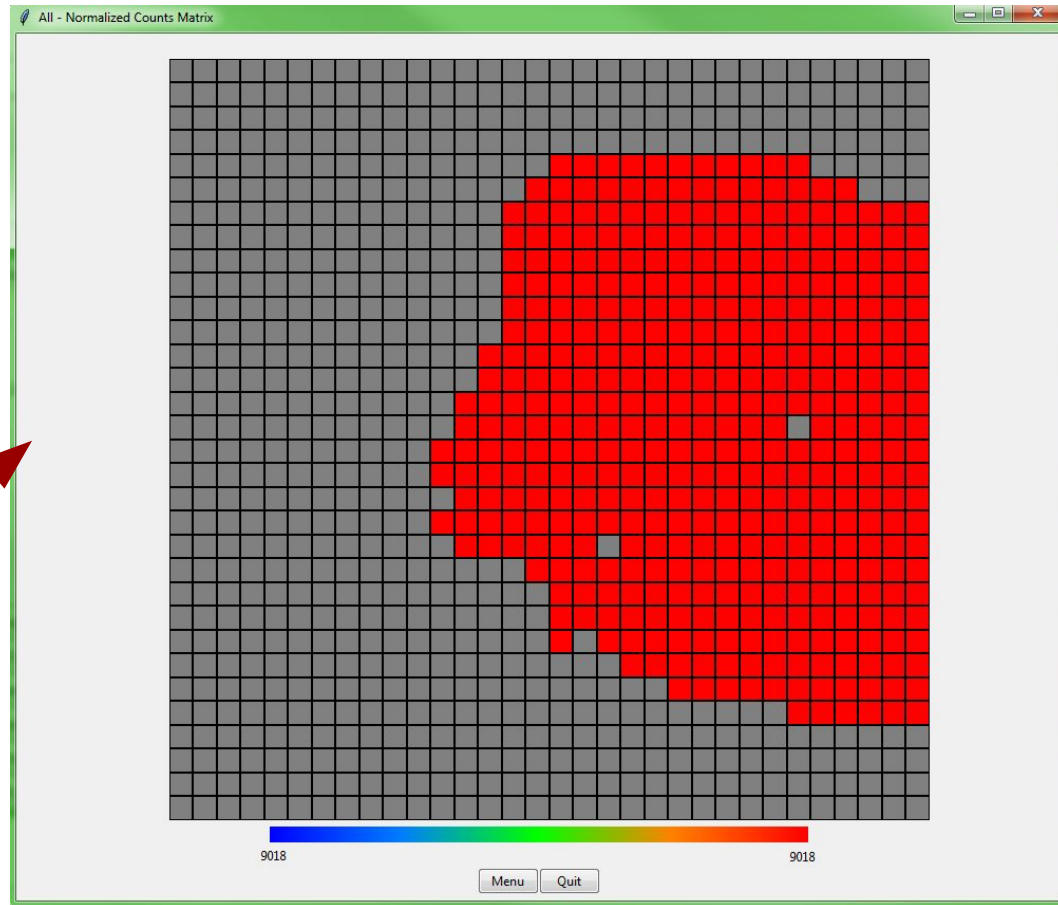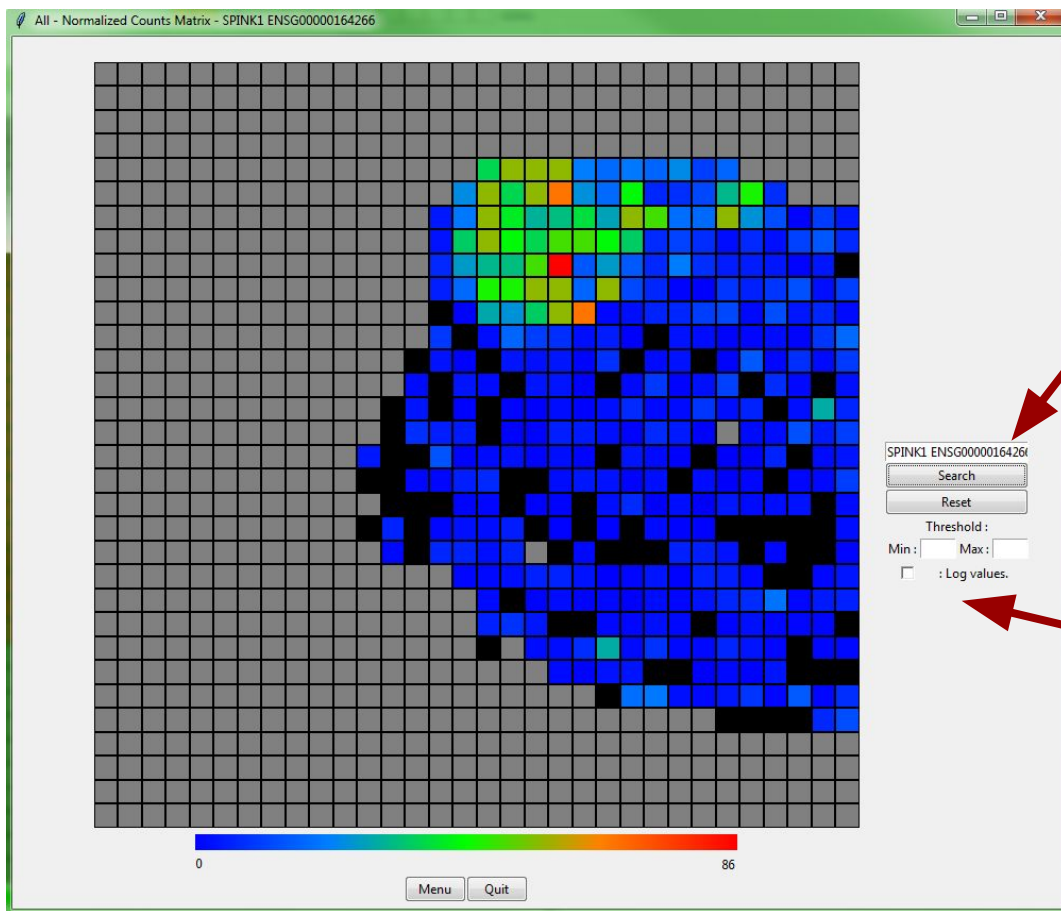
The right panel appears when you hit the menu button.

You can query a gene.

Display gene with log values
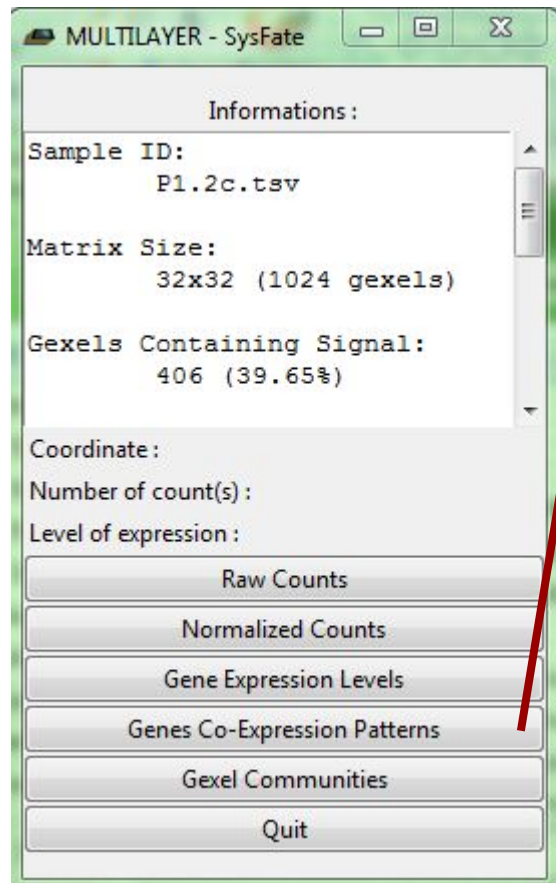
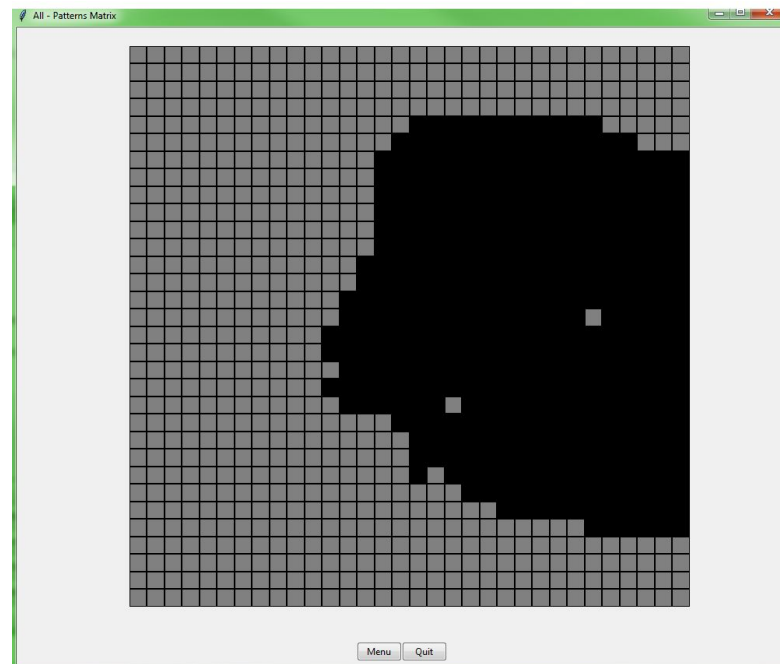## II - 3 - Gene expressed levels dataset

You can query a gene.

The button 'show repressed genes' show the top ranking down-regulated genes based on the number of gexels.
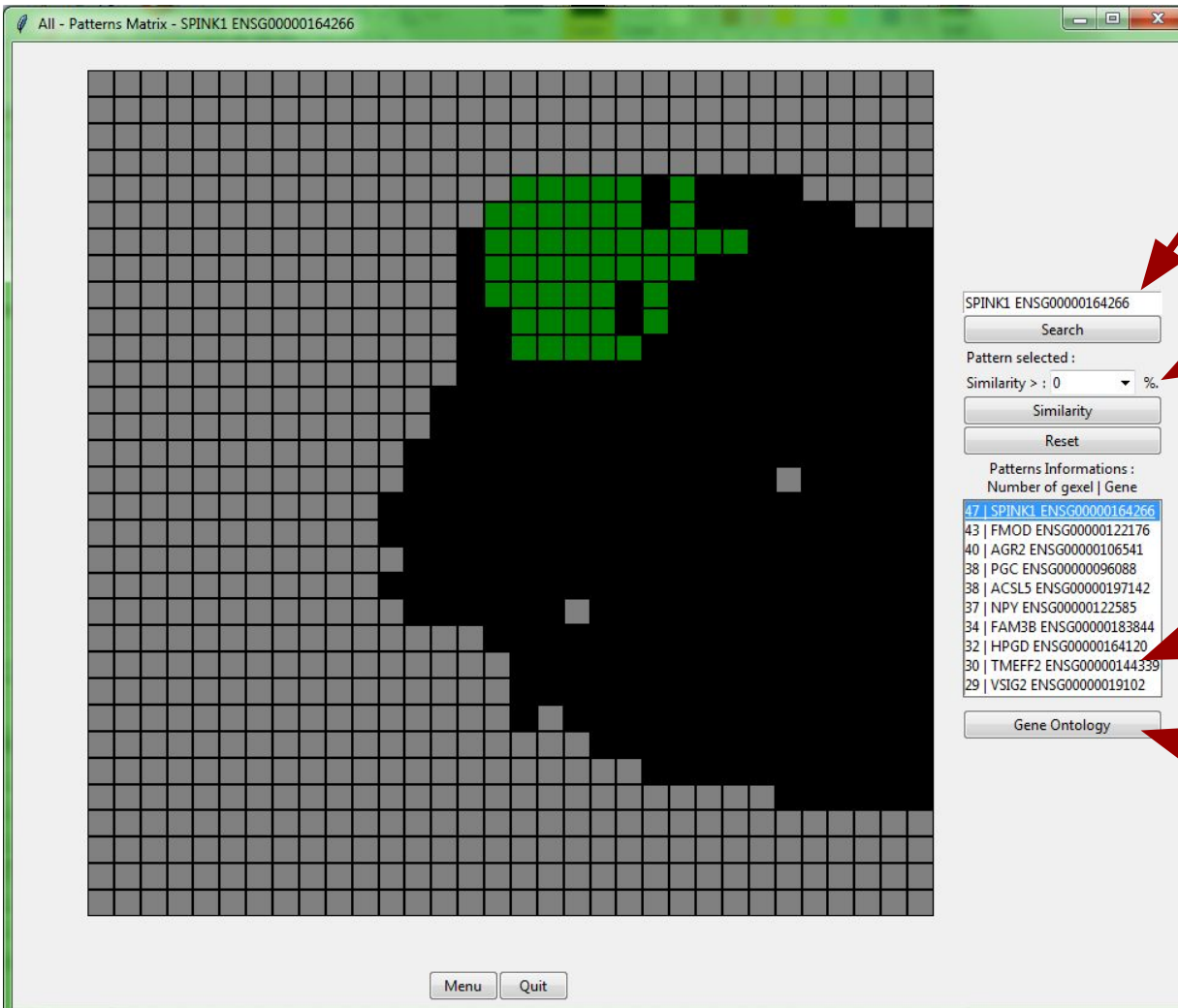
Top ranking of up-regulated genes based on the number of gexels.

II - 4 - Co-expression Patterns

The value is displayed here is the default value with whom the analysis was made. If you change this value, the tool will perform another agglomerative clustering for determine new clusters (this step can take a little time).

All - Patterns Matrix - SPINK1 ENSG00000164266

SPINK1 ENSG00000164266
Search

Pattern selected :
Similarity > : 0   %.
Similarity
Reset

Patterns Informations :
Number of gexel | Gene
47 | SPINK1 ENSG00000164266
43 | FMOD ENSG00000122176
40 | AGR2 ENSG00000106541
38 | PGC ENSG00000096088
38 | ACSL5 ENSG00000197142
37 | NPY ENSG00000122585
34 | FAM3B ENSG00000183844
32 | HPGD ENSG00000164120
30 | TMEFF2 ENSG00000144339
29 | VSIG2 ENSG00000019102

Gene Ontology

Menu   Quit

You can query a gene.

Pattern of co-expressed genes analysis based on their similarity.

Top ranking genes based on the number of gexels per pattern.

Gene ontology analysis.

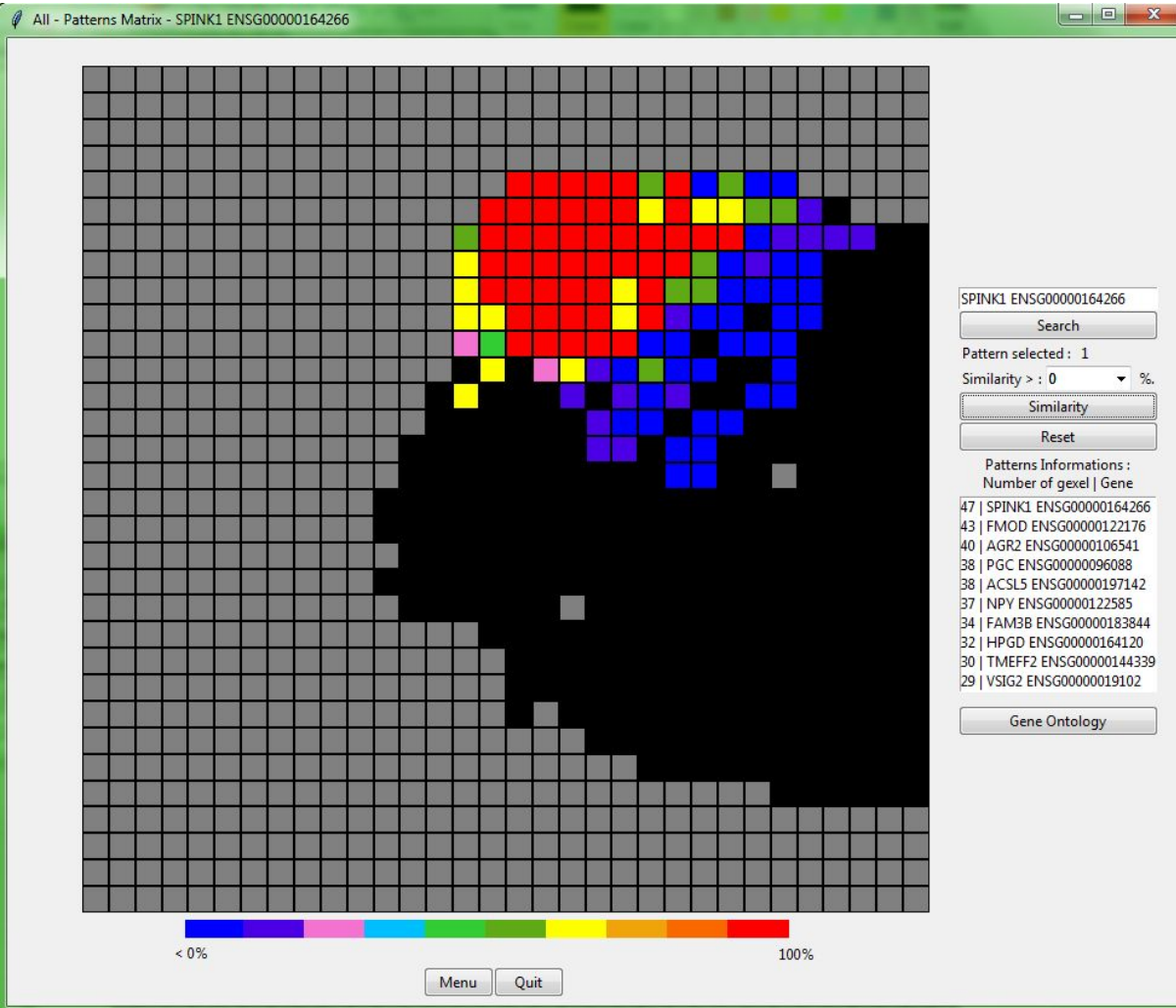# II - 4 - A - Pattern of co-expressed genes analysis

First step:
Select a pattern of genes by clicking on one gexel.

The selected pattern will appear here.

Second step:
Select a threshold of similarity and press 'Similarity button'.

The result of the co-expression pattern analysis

List of genes in each intervals of similarity.

Save the liste in tsv file by clicking on 'save' button.

Select a GO database. The user can provide his own database (paste the file with the correct format in the directory called 'GO_DB').

Once a database selected, click on 'run' button (this step can take few seconds).

Save the result (the table in 3 colums) in a tsv file by clicking on 'save' button.

Barplot and heatmap will be introduce on the next page.
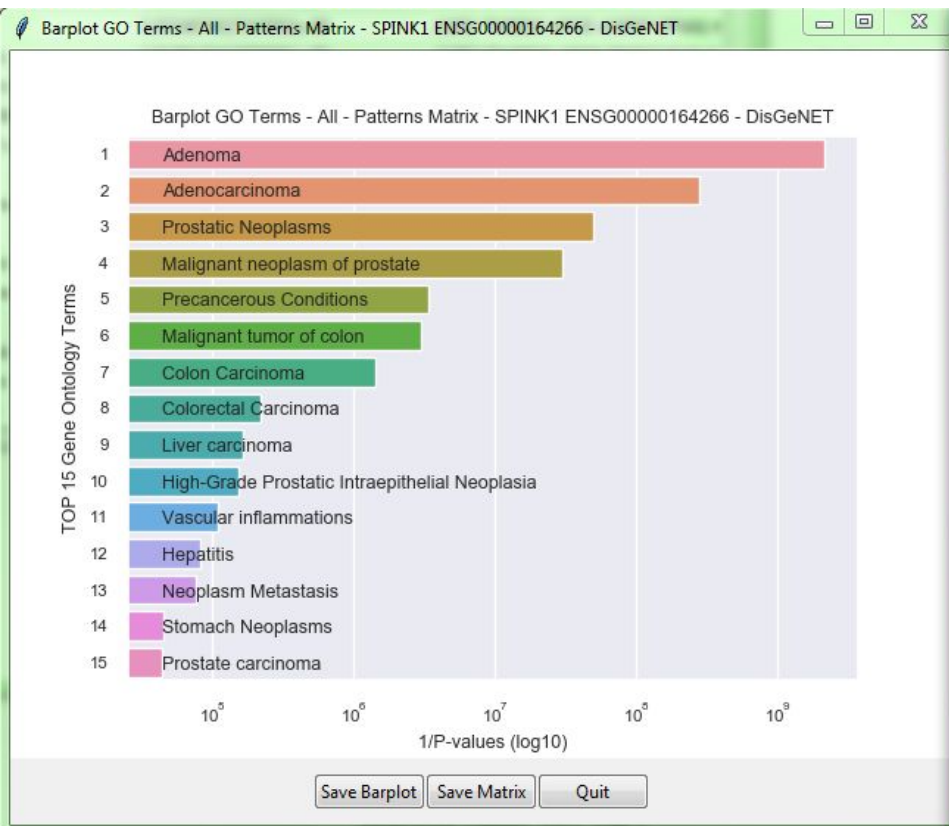
Barplot shows TOP 15 best GO terms based on p-values.

Save plot by clicking on 'save barplot'
Save the full matrix by clicking on 'save matrix'.
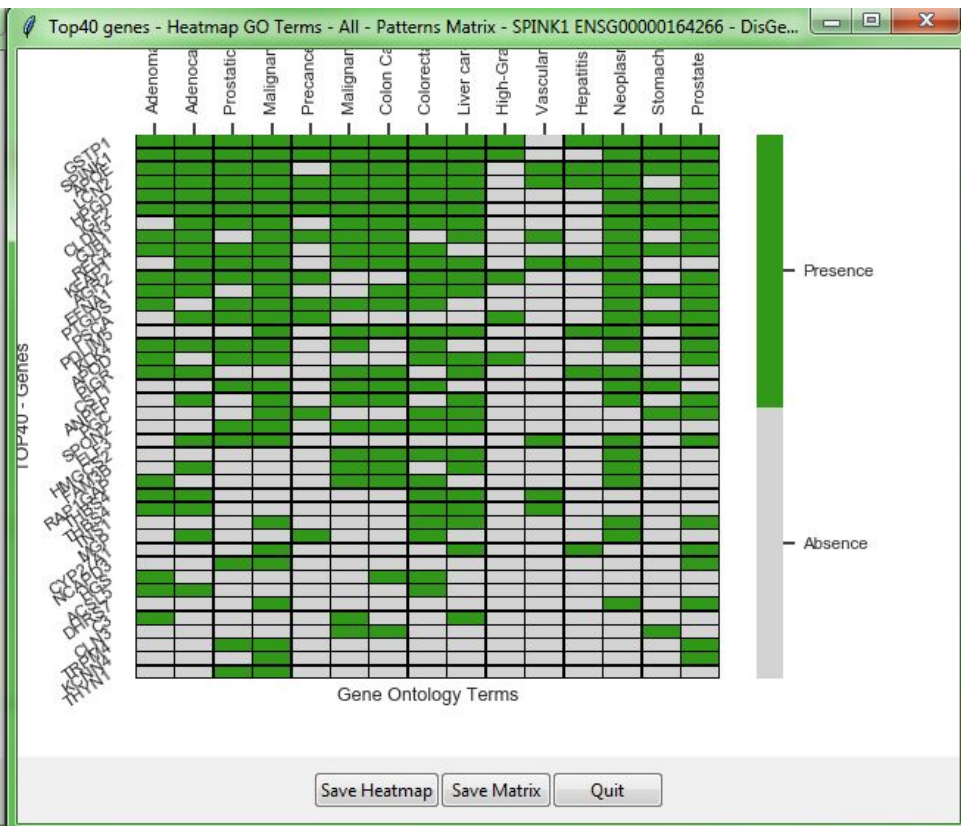
Heatmap shows TOP 15 best GO terms and the TOP 40 genes.

Save plot by clicking on 'save heatmap'.
Save the full matrix by clicking on 'save matrix'.
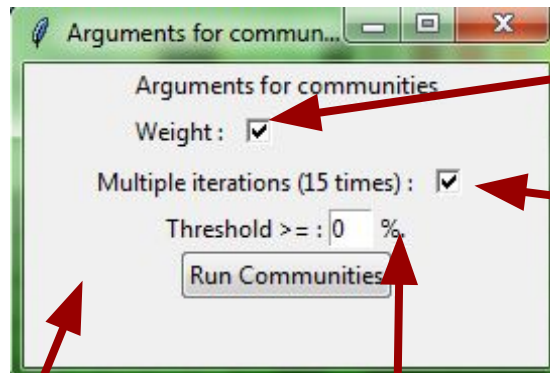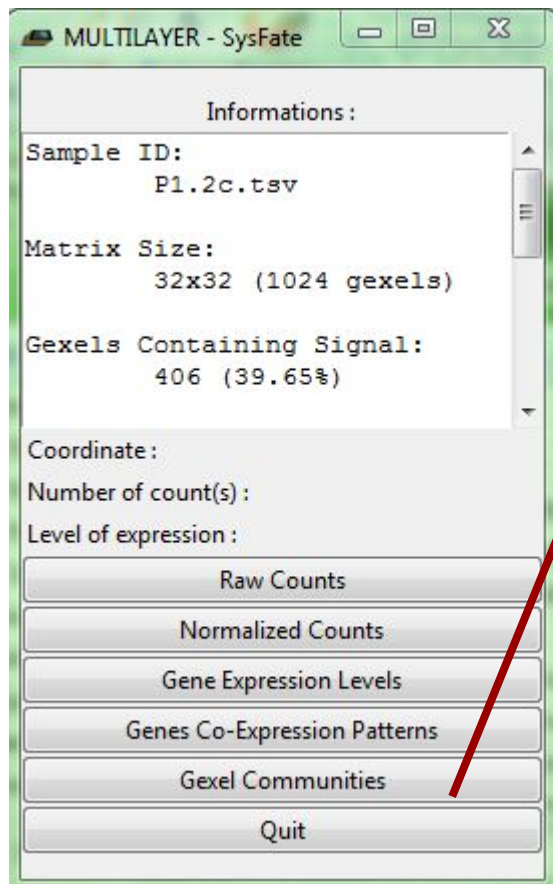
Consider the similarity weight as parameter in louvain algorithm.

Multiple iterations, perform 15 times louvain algorithm and keep the more frequent. Result will be show in the console.

Filter genes by similarity.
Default value is 0, all genes will be consider.
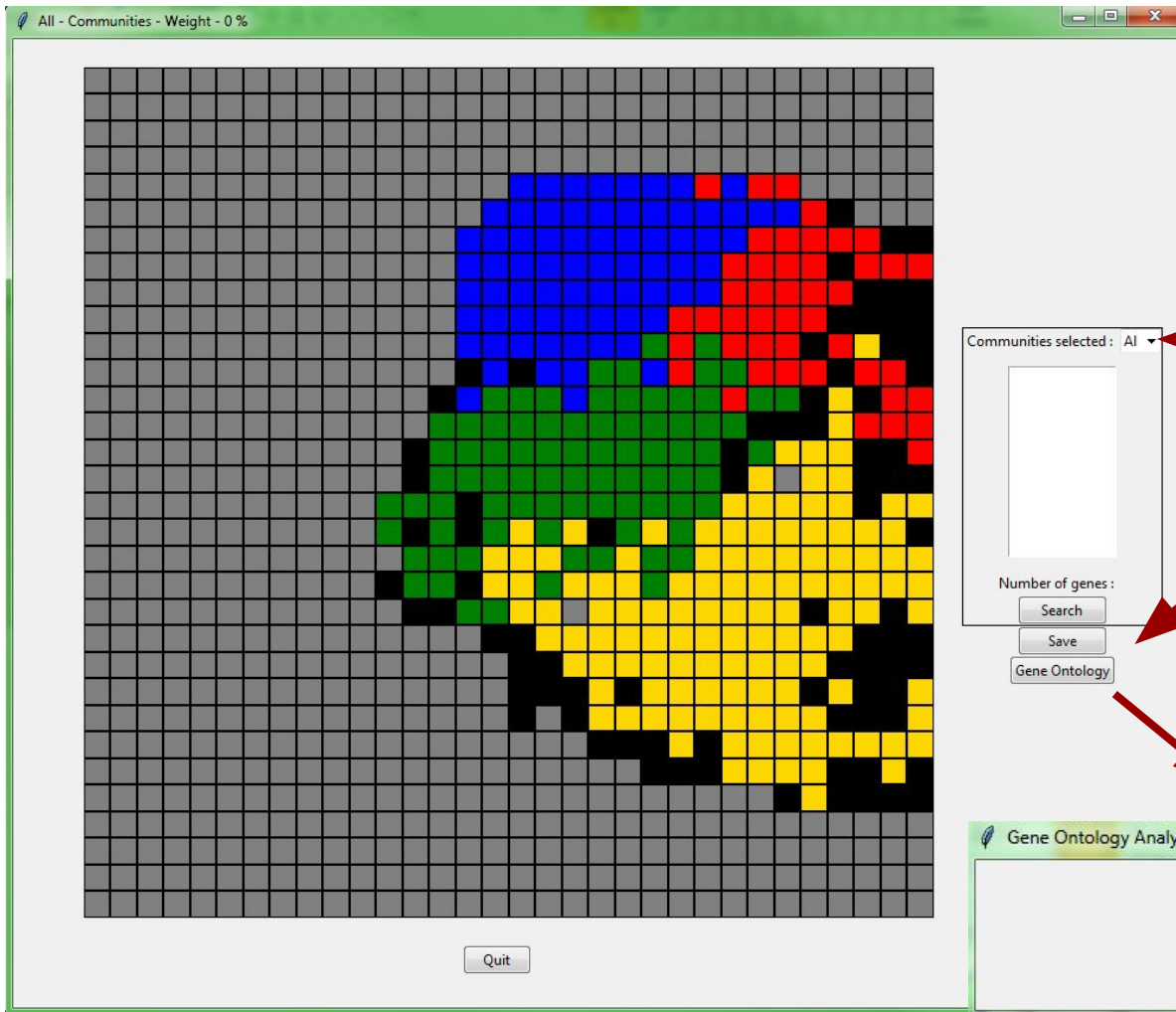
All communities are displayed. If there are overlaps between communities, the tool will display the gexel with community with the highest similarity.

Display all communities or a specific one.

Save a tsv file (result of louvain)

Select database and 'run'.

For each community, get the top 5 best GO terms.

Heatmap GO Terms - All - Communities - Weight - 0 % - TOP5 - DisGeNET

Gene Ontology Terms	0	1	2	3
Drug abuse	0.26	0.46	0.22	0.00034
kidney injury	0.11	0.22	2.6e-06	1
c Neoplasms	0.14	2.8e-08	0.0048	4.2e-05
m of prostate	0.0013	2.8e-07	0.068	0.00028
Adenoma	0.38	1.1e-06	4.8e-06	0.18
kemia, T-Cell	9.4e-05	1	1	0.016
umor of colon	0.012	6.3e-06	0.016	0.024
nocarcinoma	1	7.2e-07	0.00024	0.32
ted Adenoma	0.00025	1	1	0.13
dial Infarction	0.13	0.04	2.6e-07	0.17
Atrophic scar	1	1	1.4e-06	1
Insufficiency	0.11	0.21	2.3e-06	1
ted Disorders	0.14	1	1	0.00062
giocarcinoma	7.4e-05	1	1	0.015
or Deficiency	5.3e-05	1	1	0.08
rtery Disease	0.64	0.33	3e-07	0.66
Insufficiency	0.00016	1	0.21	0.27
hral Stricture	1	1	1	0.00048

Save plot by clicking on 'save heatmap'

Save matrix by clicking on 'save matrix'

All communities are displayed. If there are overlaps between communities, the tool will display the gexel with community with the highest similarity.
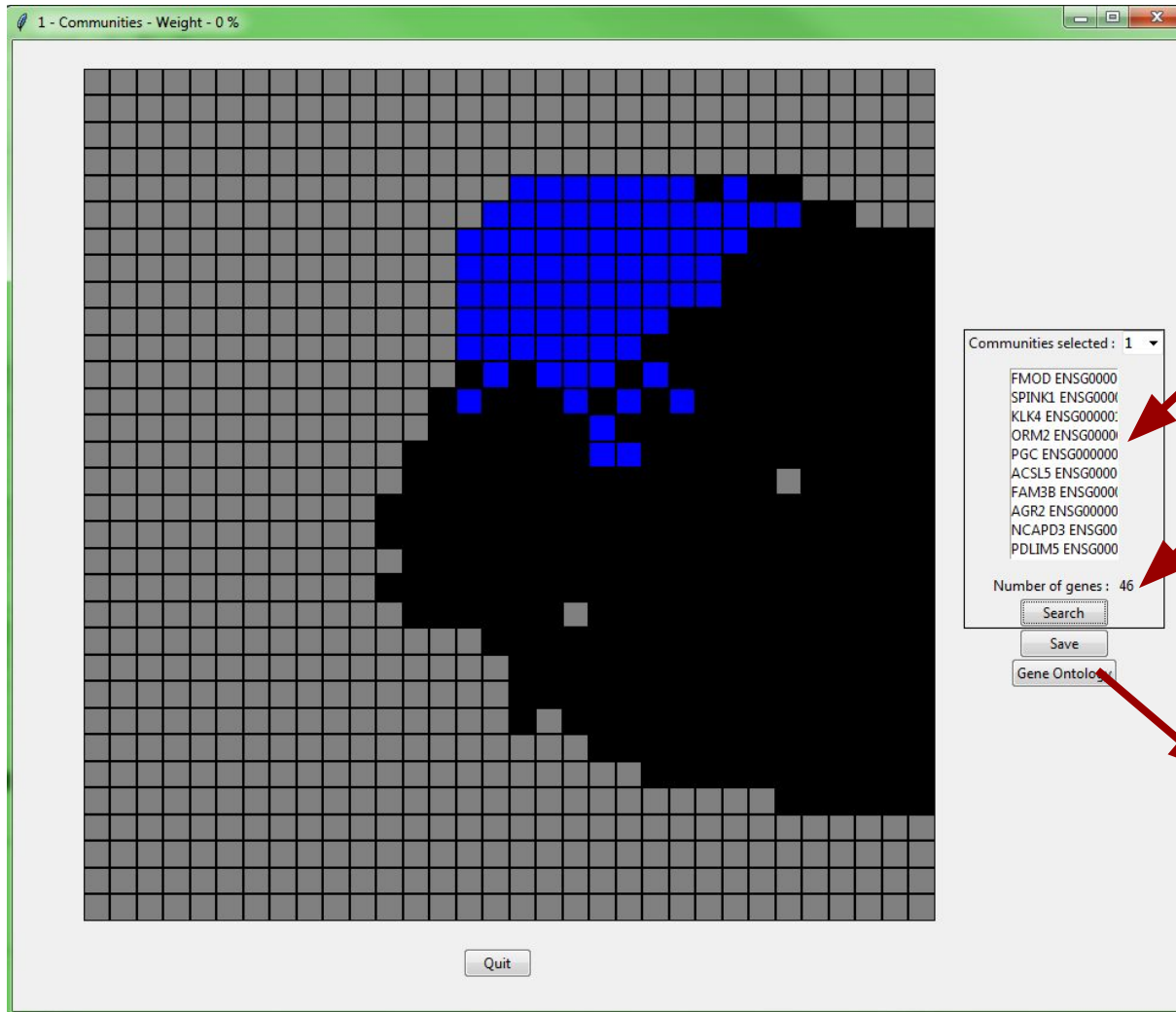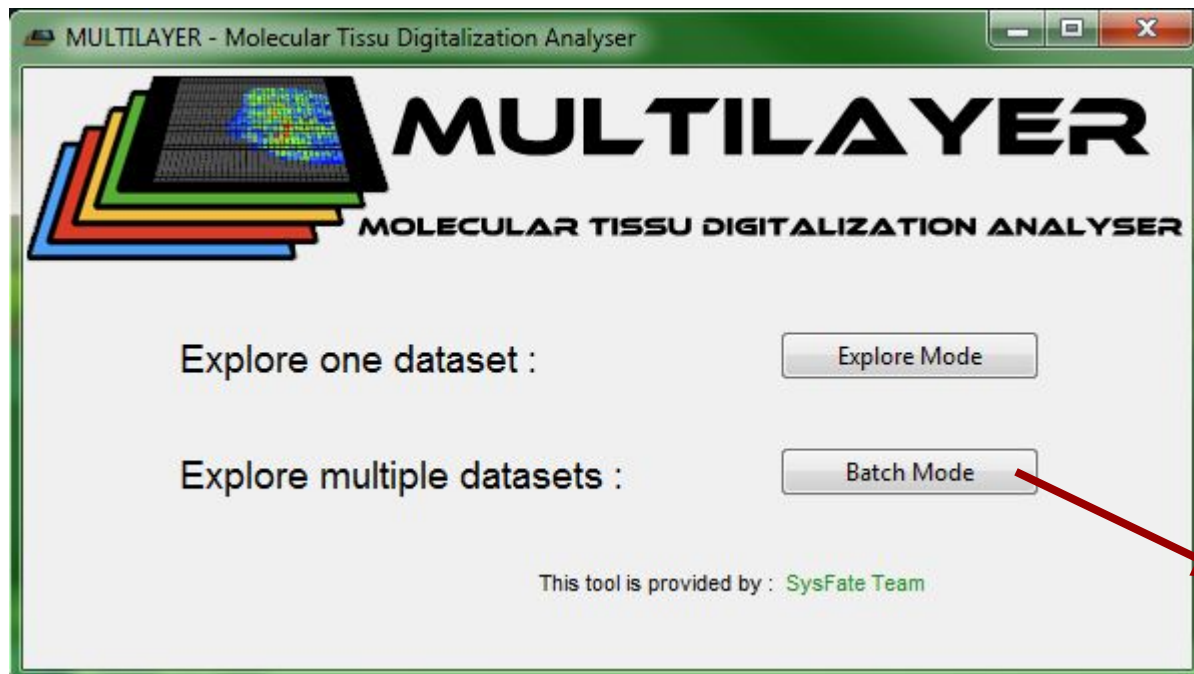
If a specific community is selected, genes which compose the given community will be display.

Toltal number of genes in the community

Same as parttern of co-expressed genes. Go back to pages 16 & 17.

Window title: 1 - Communities - Weight - 0 %

Panel:
Communities selected : 1

FMOD ENSG0000
SPINK1 ENSG0000
KLK4 ENSG000002
ORM2 ENSG0000
PGC ENSG000000
ACSL5 ENSG0000
FAM3B ENSG0000
AGR2 ENSG00000
NCAPD3 ENSG00
PDLIM5 ENSG000

Number of genes : 46

Search

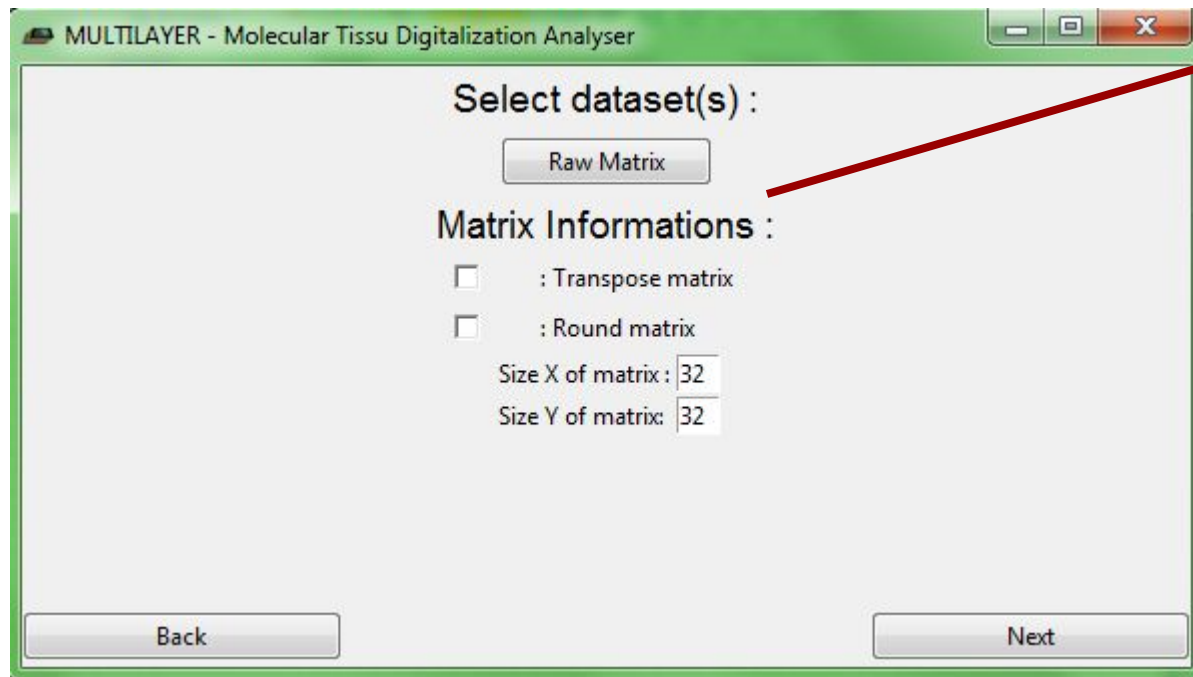Save

Gene Ontology

Quit

III - Batch mode with MULTILAYER (multiple samples)

III - 1 - Mode

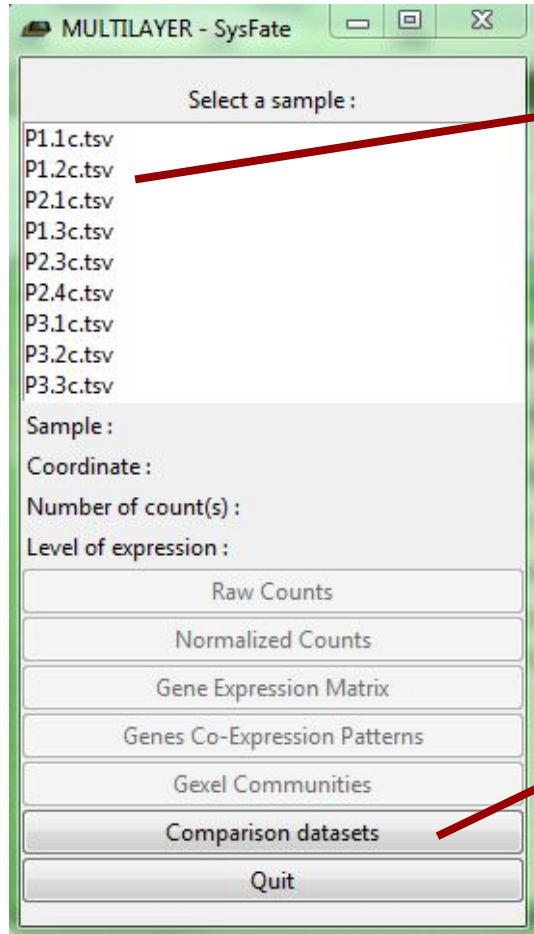Explore mode : perform analysis over several samples (.tsv files)

Same as Explore mode (page 4)

The single difference :
By clicking on raw matrix, you will choose a directory. All tsv files in the selected directory will be consider as a input Matrix.
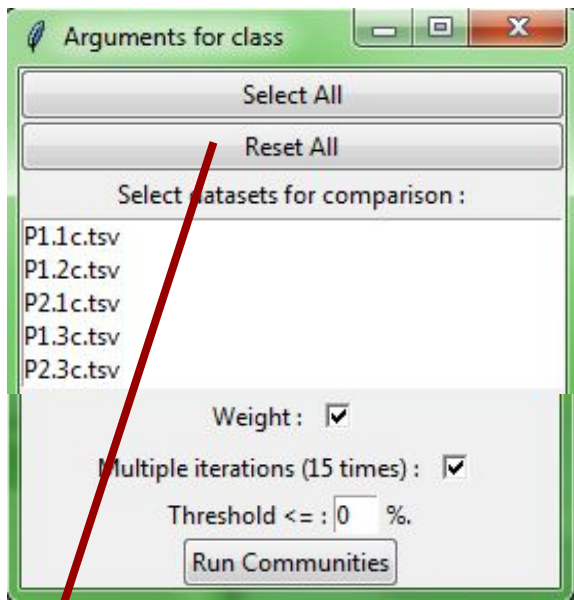
Select a specific sample for perform all previous analysis from pages 6 to 21.

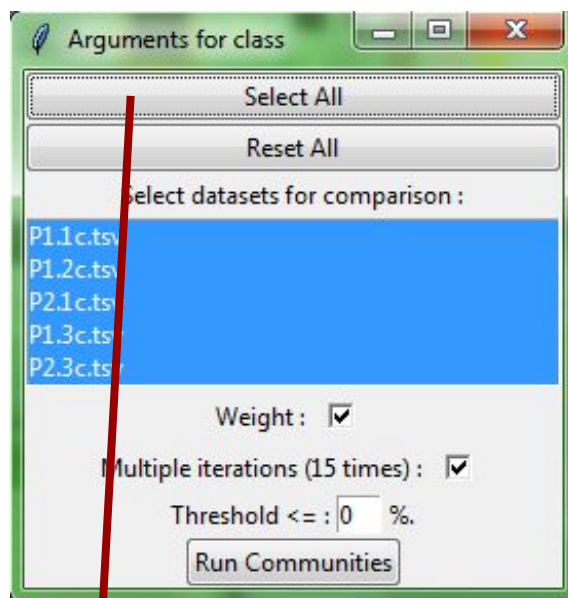**Important Note :** Raw counts and normalized counts are not stock in memory, they will be calculate every each click on 'raw counts' and 'normalized counts' buttons.

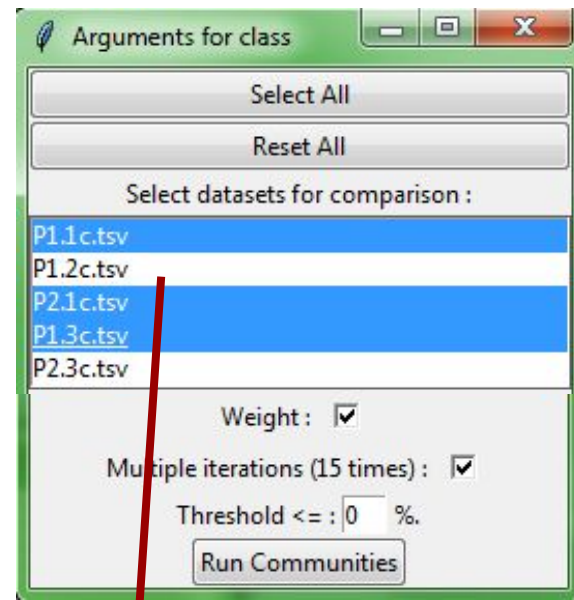Comparision of all communities from all datasets.
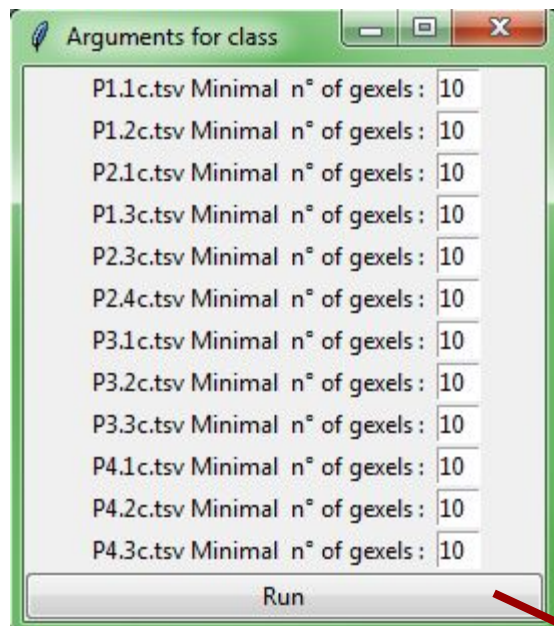
Reset all selection

Select all datasets

Select several specifics datasets

**Important note :** Arguments Weight, Multiple iterations and Filter are commented page 18.

The displayed value is the default value.

For the analysis, you can change minimal number of gexels for define patterns for each datasets separately.

If you change the value, the tool will compute with new value (this step can take a time).

You have to provide a name for saving files. The tool will save 2 files (ex: save.tsv, save_filter.tsv)

Arguments for class

Dataset(s) communities

C
l
a
s
s
e
s

Community :
Class :

Save | GO Terms
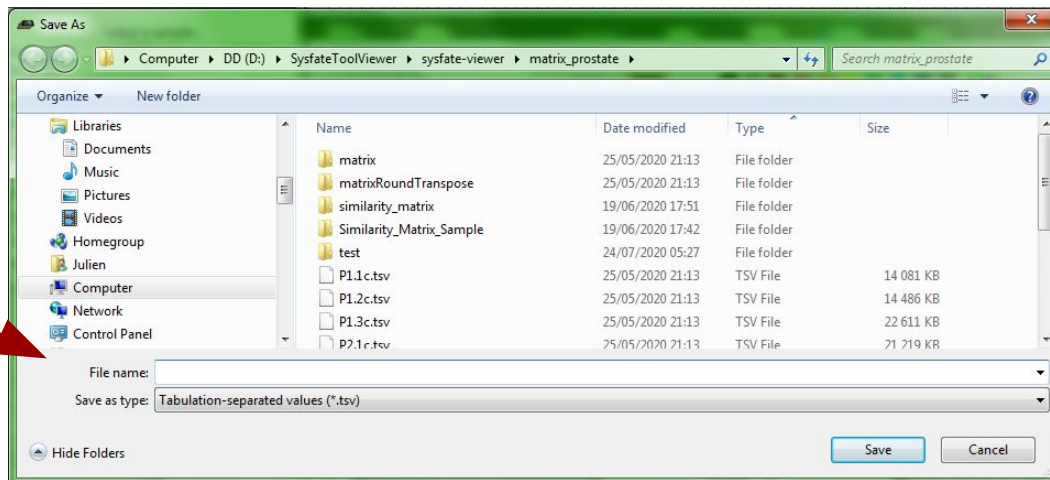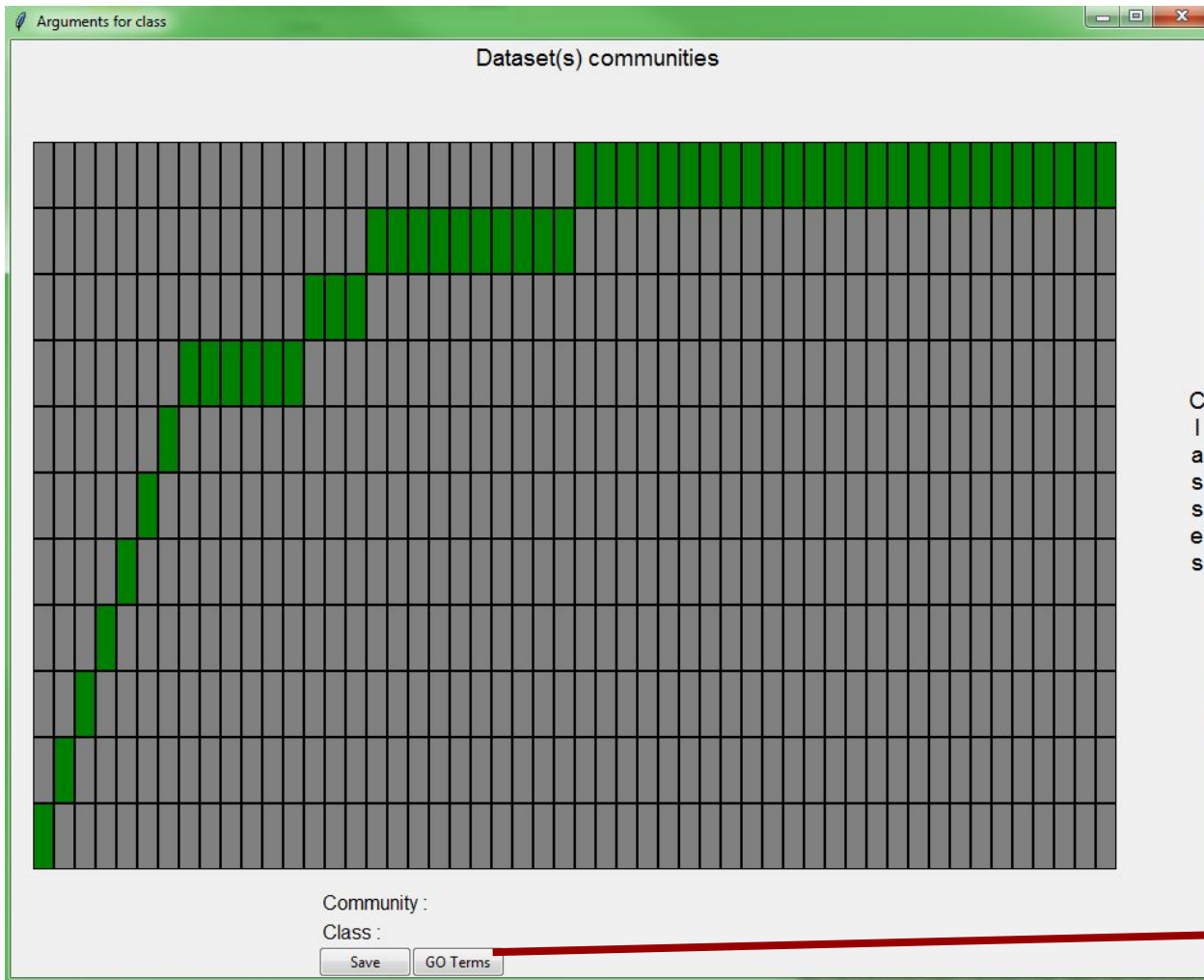
Interactive heatmap :
In X axis all datasets communities, in Y axis all classes.

A class is group of 1 or several communities.

GO terms

Select the number of GO terms selected.
Select the database.

Heatmap GO Terms - All - TOP5 - ARCHS4_Tissues

Save plot by clicking on 'save heatmap'

Save matrix by clicking on 'save matrix'
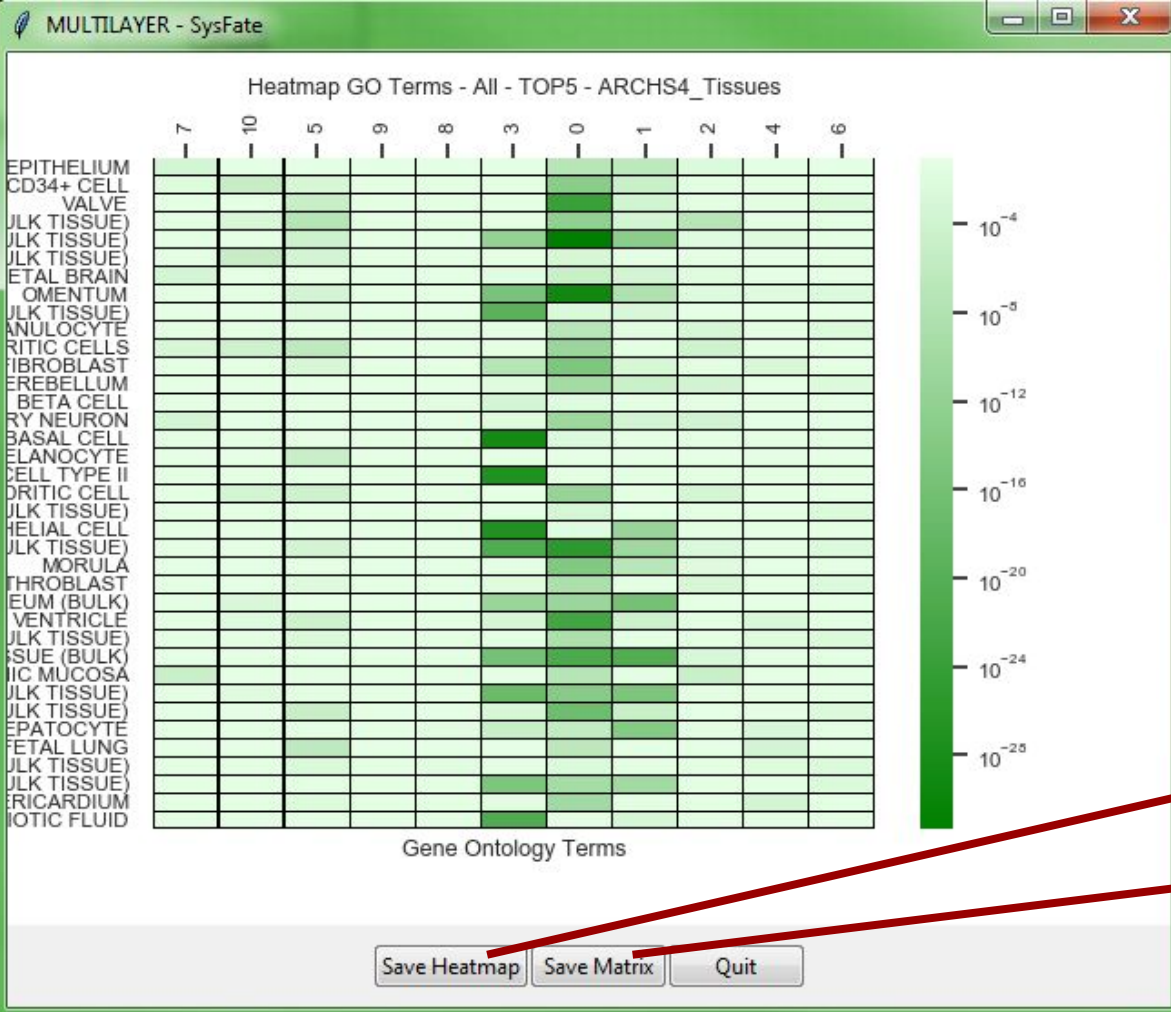
IV - MULTILAYER compressor

For high resolution dataset, we developed module for compress data. Multilayer compressor is able to merge several gexels in one big gexel. The format has to be in 3 columns like : barcodes (XxY) with header 'bc' - Genes with header 'gene' - Genes counts with header 'count'

Multilayer_Compressor.py has several arguments :
-i : input matrix.
-o : output matrix.
-cx : compressor factor.
-cy : (optional) compressor factor for Y. If -cy is not define -cy will be equal to -cx.

python3 Multilayer_Compressor.py -i input.tsv **-o** output.tsv **-cx** 100 **-cy** 100

In this given example : 10 000 gexels will be merge in 1 gexel.