



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

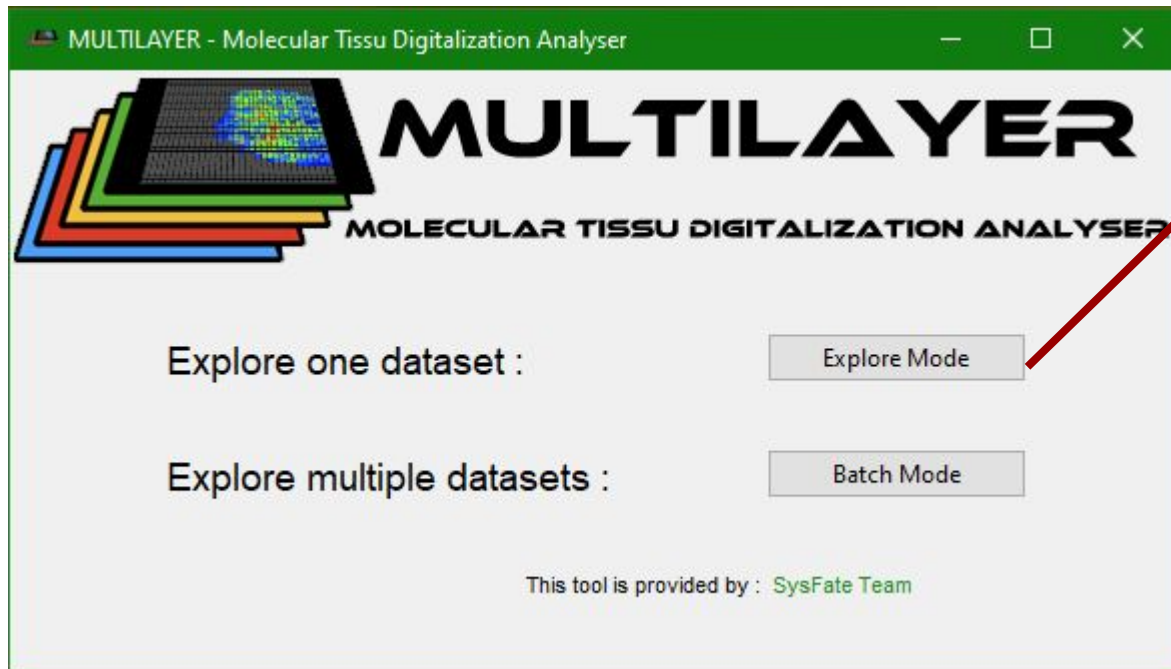
- | | |
|----------------|----------------------------|
| - Numpy | pip install numpy |
| - Matplotlib | pip install matplotlib |
| - Pandas | pip install pandas |
| - Scipy | pip install scipy |
| - Scikit-learn | pip install scikit-learn |
| - Seaborn | pip install seaborn |
| - Networkx | pip install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

2 - Run

```
python3 Multilayer.py
```

II - Explore molecular tissu with MULTILAYER

II - 1 - Mode



Explore mode : perform analysis over one sample (.tsv file).

II - 2 - Arguments

MULTILAYER - Molecular Tissue Digitalization Analyser

Select dataset(s) :

Raw Matrix Normalized Differential Expressed

Reset Selection

Reset your matrix selected.

Matrix Informations :

☐ : Transpose matrix

☐ : Round matrix

Size X of matrix : 32

Size Y of matrix : 32

Save matrix (norm, diff) : ☐

Back Next

Provide your dataset (.tsv file). The correct format is a matrix with genes name as rows, coordinates (coordinates should be in this format : 'XxY') as columns. You have to provide **at least a 'raw matrix'**. If you provide your own normalized matrix and / or differential expressed matrix Multilayer will use them.

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 maximum of Y. Default value is 32.

If you provide only raw matrix, you can save normalized and differential expressed matrix generated by Multilayer.

MULTILAYER - Molecular Tissu Digitalization Analyser

Arguments for differential gene expression :

Threshold up regulated :

Threshold down regulated :

Patterns detection :

Minimum n° of contiguous gexels :

Similarity Methods :

☒ Tanimoto

☐ Dice

Multiprocessing :

Number of thread : max (16).

Back Next

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

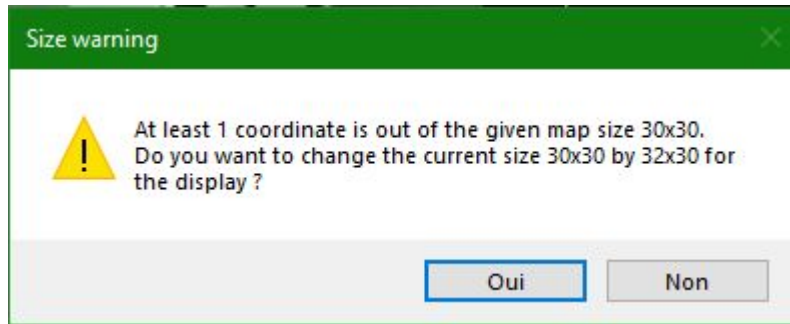
Pattern detection : The tool will use agglomerative clustering on gexels (<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html>) for determine a contiguous pattern. The tool will consider a pattern if the number of gexels is upper or equal than this value. Default value is 10.

Similarity methods will calculate similarity between patterns.

With Tanimoto, the similarity will be calculated with $A \cup B / A \cap B$.

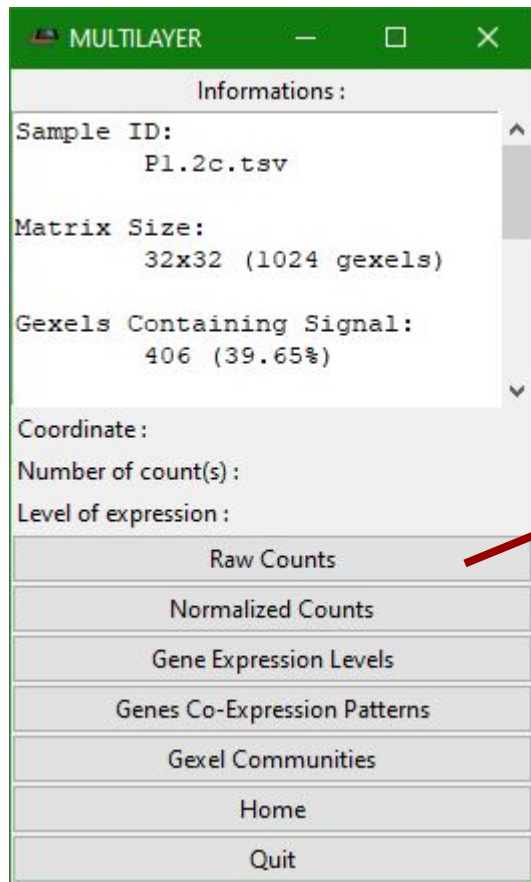
With Dice the similarity will be calculated with $2 * (A \cap B) / (A \cup B)$.

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

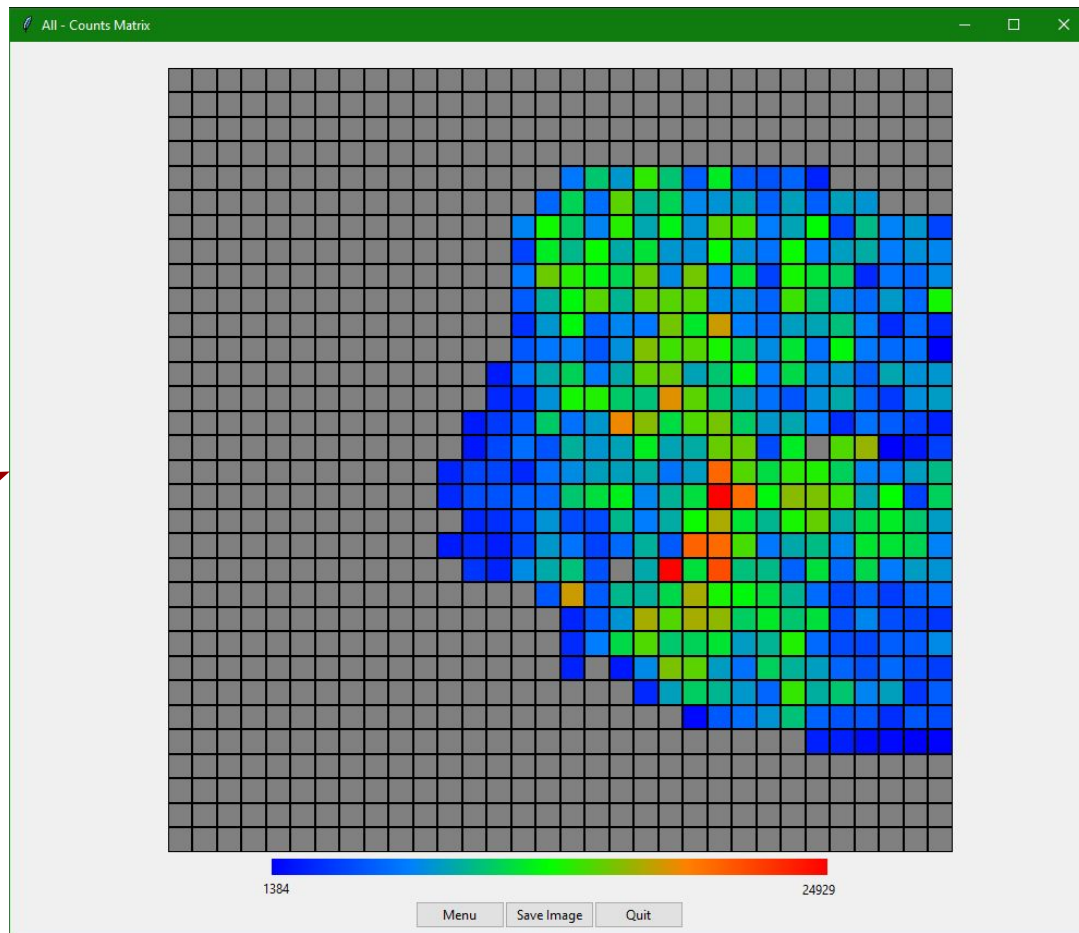


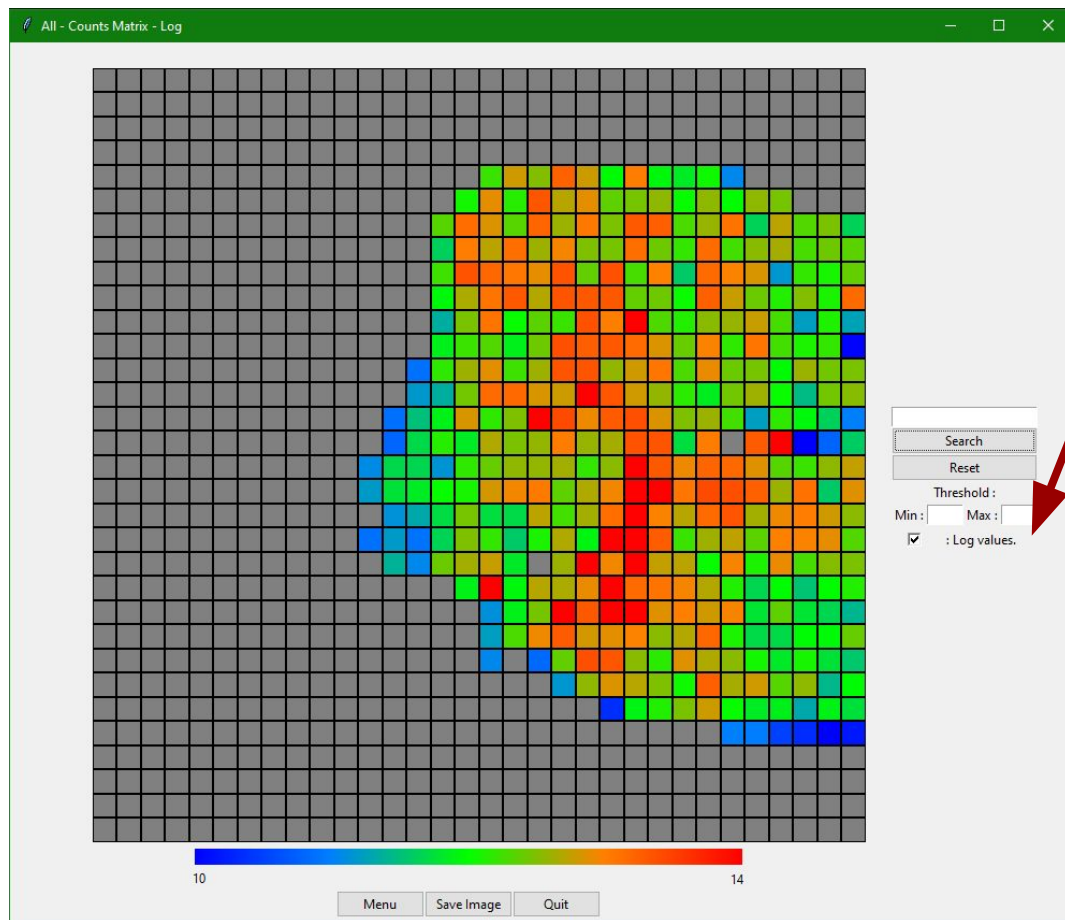
If Multilayer finds a coordinate out of the given map size. You can keep your size or take the size proposed by Multilayer.

II - 3 - Raw dataset



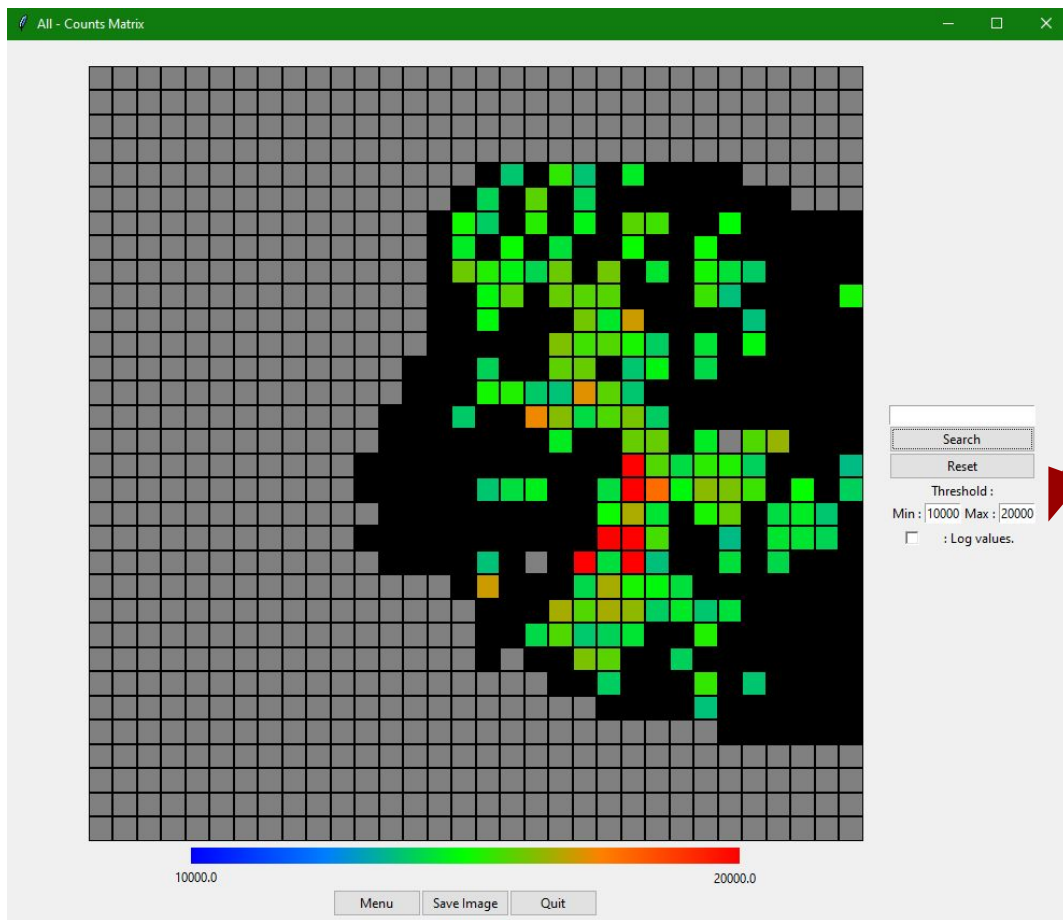
'Raw counts'
shows the input
dataset.



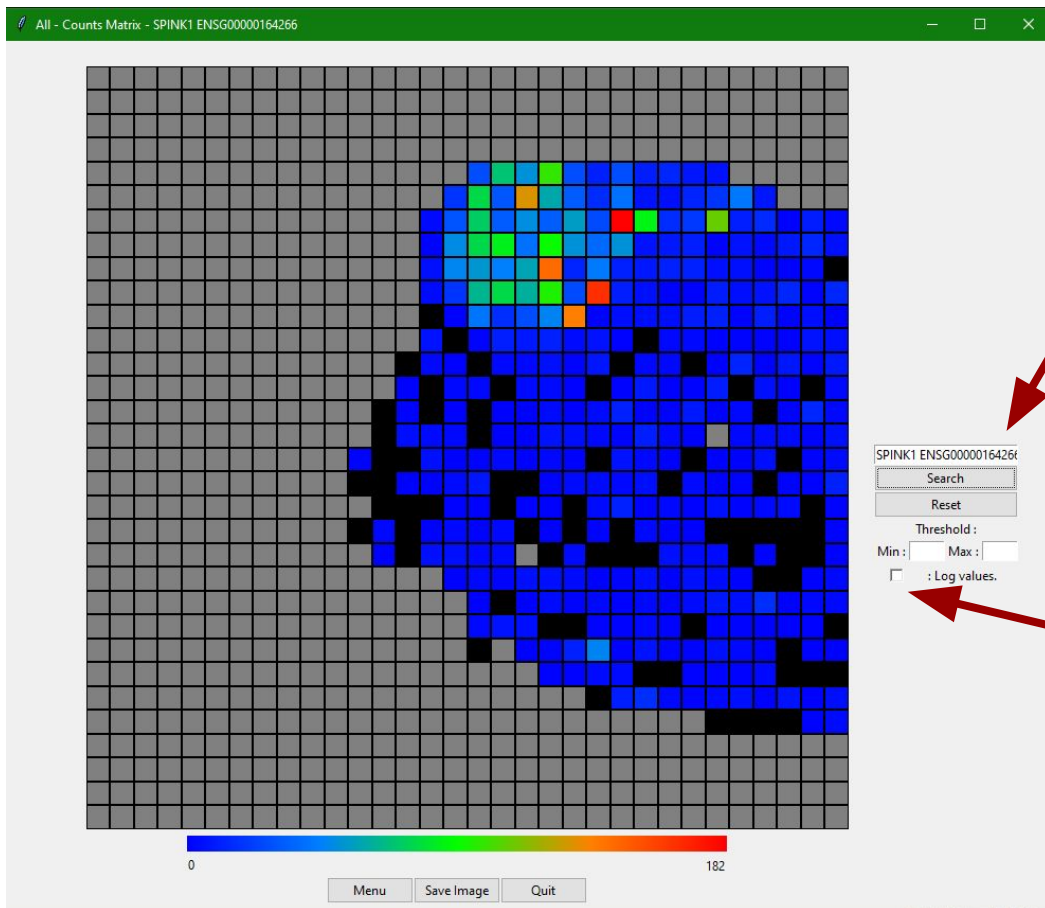


The right panel appears when you hit the menu button.

You can see the matrix with log values



You can adjust thresholds



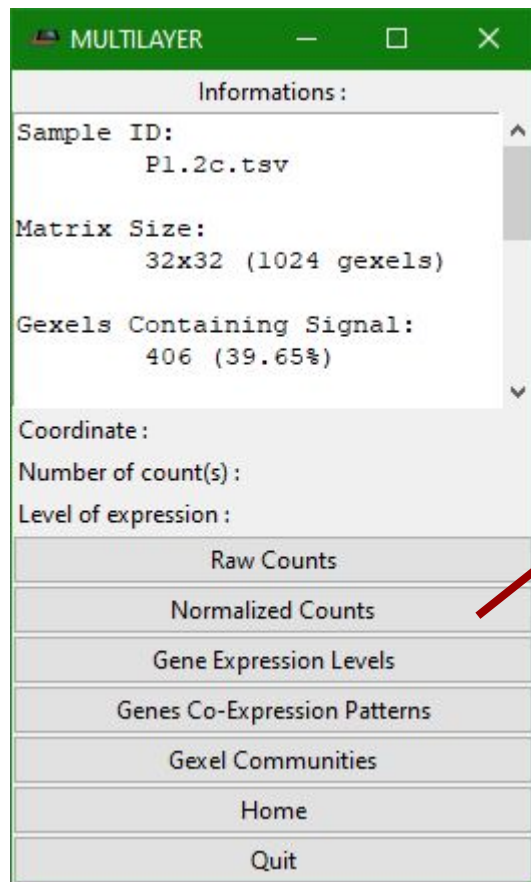
The right panel appears when you hit the menu button.

You can query a gene.

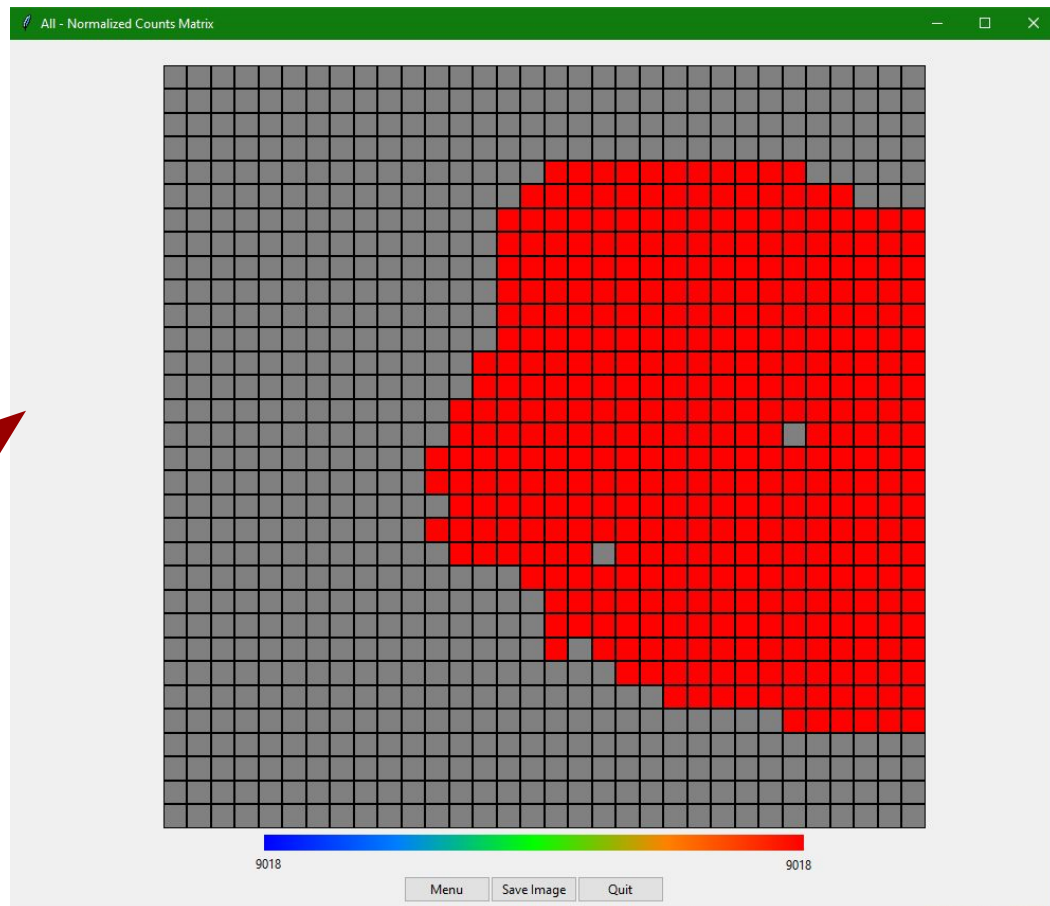
Display gene with log values

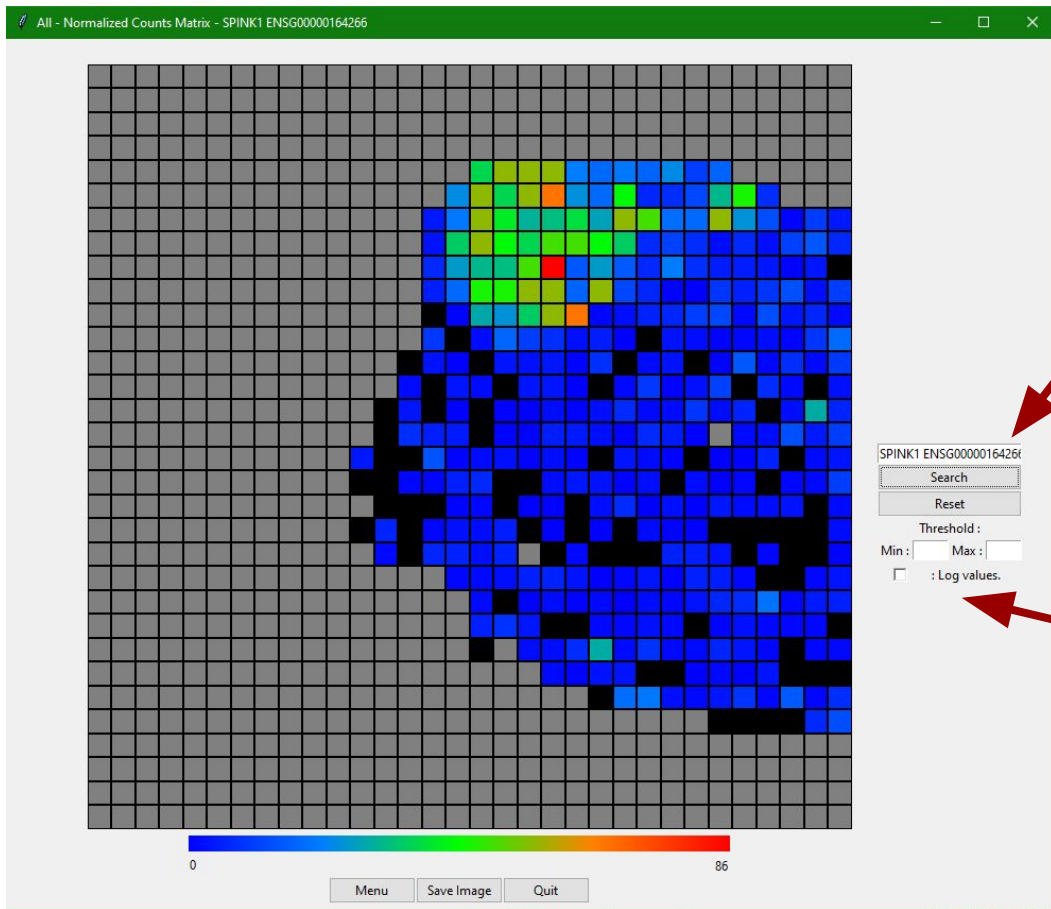
II - 3 - Normalized dataset

You can see the matrix with log values and adjust thresholds.



'Normalized counts' shows the normalized dataset.





The right panel appears when you hit the menu button.

You can query a gene.

SPINK1 ENSG00000164266

Search

Reset

Threshold :

Min : Max :

☐ : Log values.

Display gene with log values

II - 3 - Gene expressed levels dataset

MULTILAYER — □ ×

Informations:

Sample ID:
P1.2c.tsv

Matrix Size:
32x32 (1024 gexels)

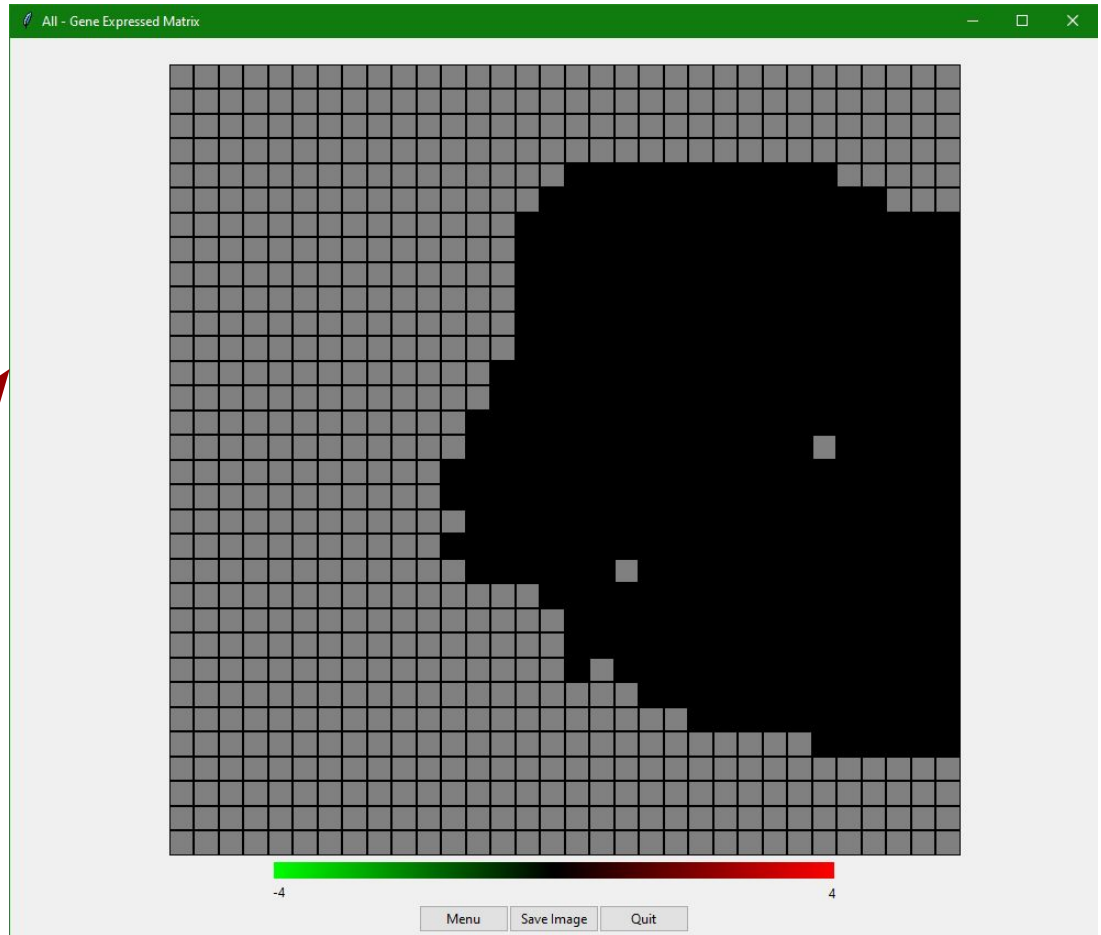
Gexels Containing Signal:
406 (39.65%)

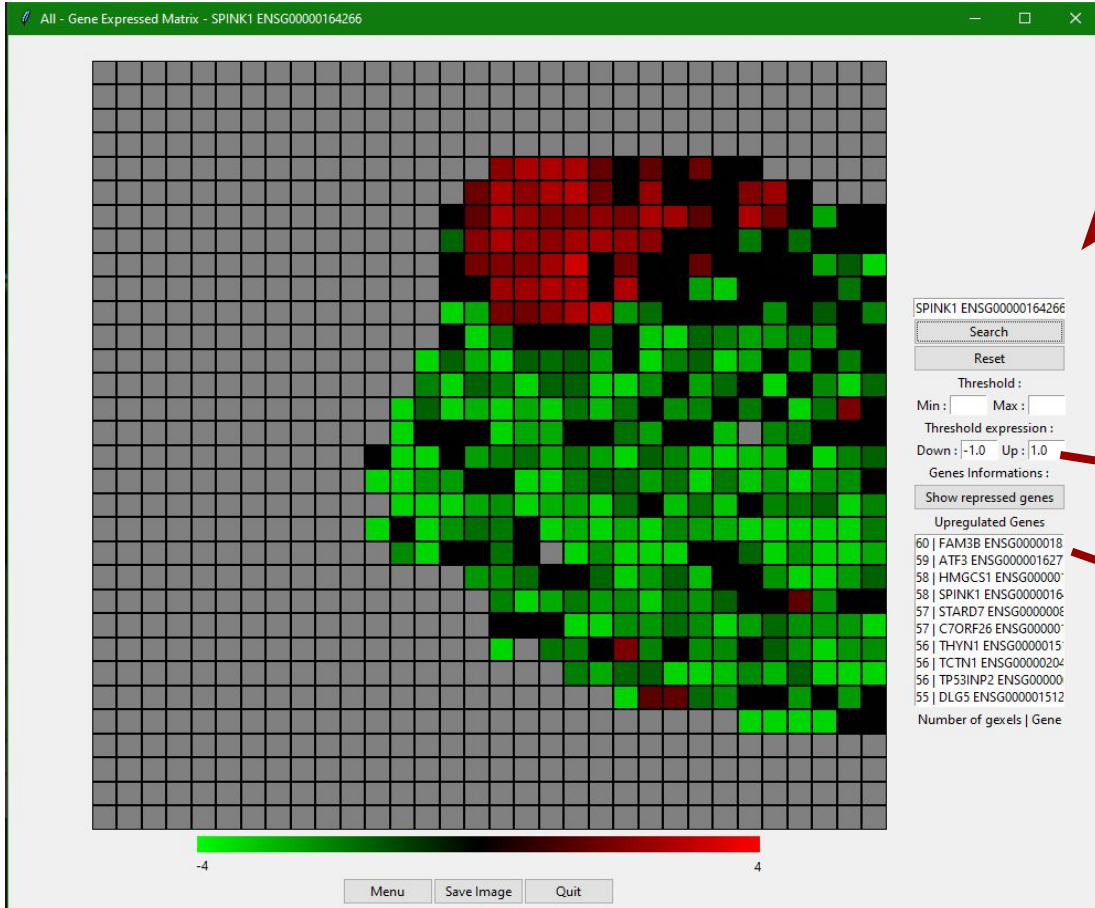
Coordinate:

Number of count(s):

Level of expression:

- Raw Counts
- Normalized Counts
- Gene Expression Levels
- Genes Co-Expression Patterns
- Gxel Communities
- Home
- Quit



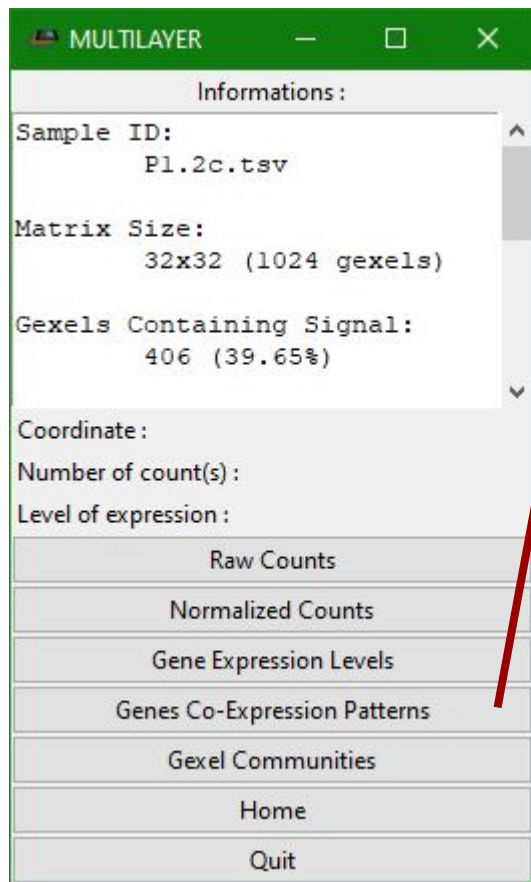


You can query a gene.

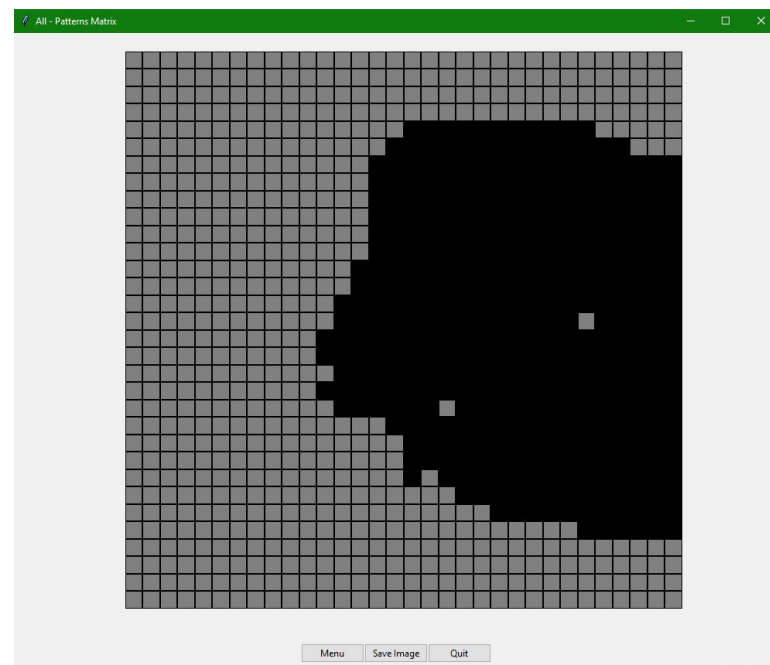
The button 'show repressed genes' show the top ranking down-regulated genes based on the number of gexels.

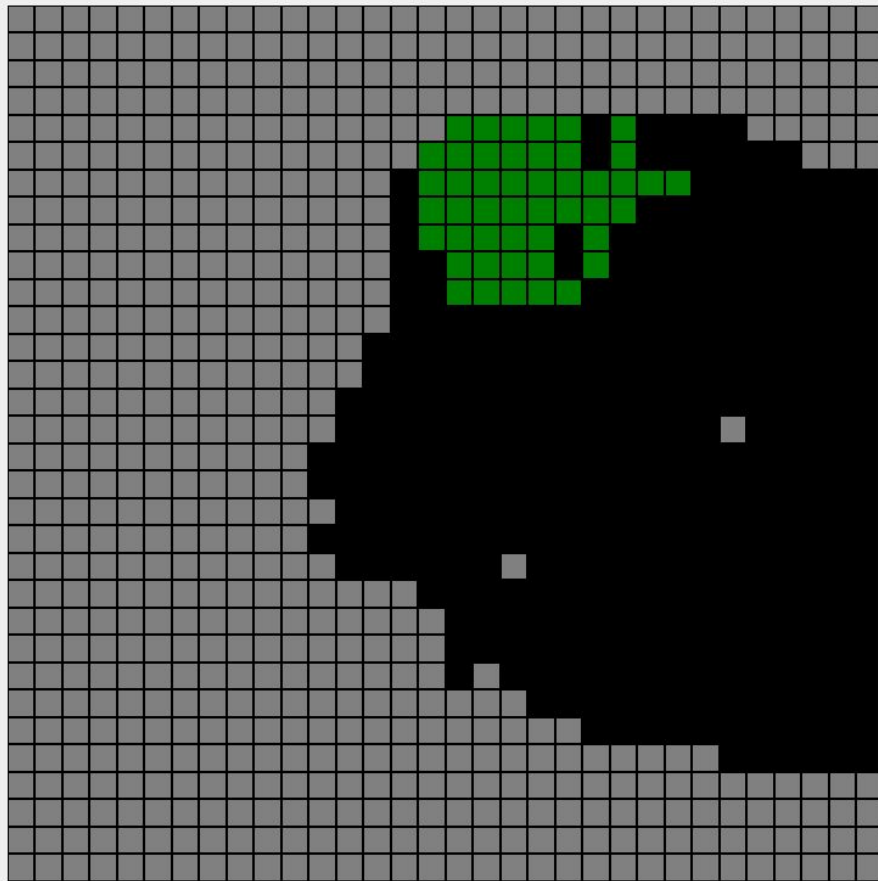
Top ranking of up-regulated genes based on the number of gexels.

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).





You can query a gene.

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

SPINK1 ENSG00000164266

Search

Pattern selected: 1

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	ENSG0000019102

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

Gene Ontology

Gene ontology analysis.

Menu

Save Image

Quit

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

Pattern selected : 1

Similarity > : %

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 ENSG0000019102

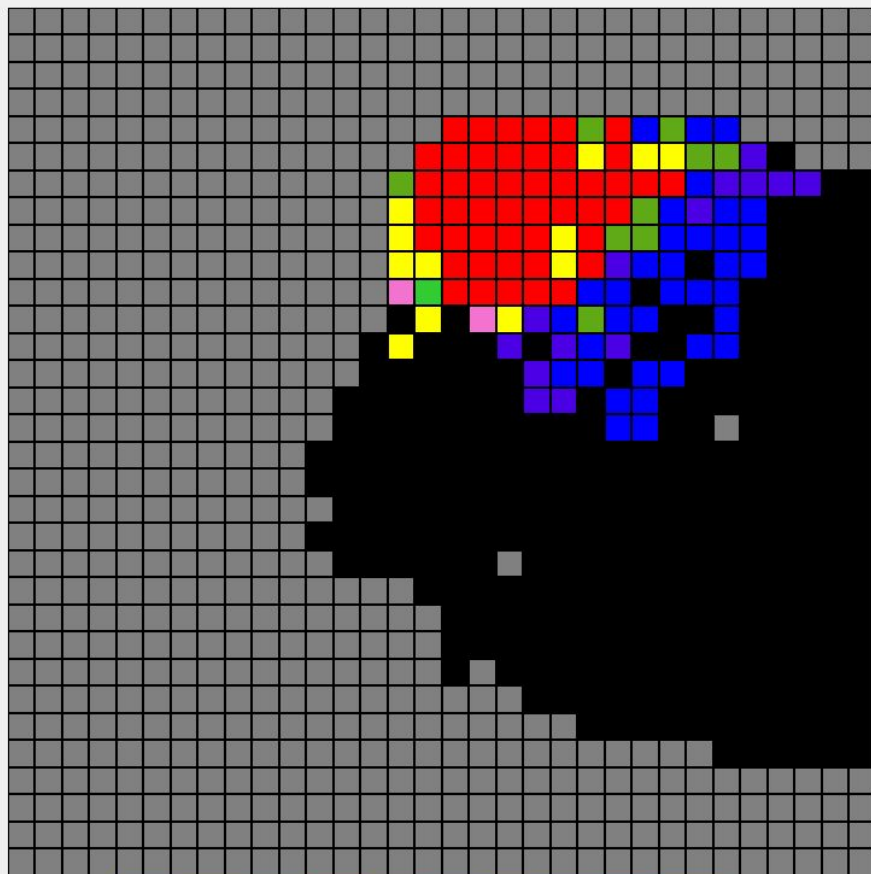
First step:

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'.



< 0%

100%

Menu

Save Image

Quit

SPINK1 ENSG00000164266

Search

Pattern selected : 1

Similarity > : 0 %

Similarity

Reset

Patterns Informations :
Number of gexel | Gene

47 | SPINK1 ENSG00000164266
 43 | FMO5 ENSG00000122176
 40 | AGR2 ENSG00000106541
 38 | PGC ENSG00000096088
 38 | ACSL5 ENSG00000197142
 37 | NPY ENSG00000122585
 34 | FAM3B ENSG00000183844
 32 | HPGD ENSG00000164120
 30 | TMEFF2 ENSG00000144339
 29 | VSIG2 ENSG0000019102

Gene Ontology

40	30	20	10	0
CLDN3 ENSG00000164266	LIPF ENSG00000164266	SPON2 ENSG00000164266	CLN3 ENSG00000164266	C3 ENSG00000164266
ELF3 ENSG00000164266	ORM1 ENSG00000164266	KEAP1 ENSG00000164266	TNS1 ENSG00000164266	COL1A1 ENSG00000164266
THBS4 ENSG00000164266	HMGCS2 ENSG00000164266	TRPM4 ENSG00000164266	LCN2 ENSG00000164266	DHR57 ENSG00000164266
PSCA ENSG00000164266	RP11-314O13.1 ENSG00000164266	RAP1GAP ENSG00000164266	THYN1 ENSG00000164266	EFNA1 ENSG00000164266
PDLIM5 ENSG00000164266	KCNN4 ENSG00000164266	TMSB10 ENSG00000164266	CST1 ENSG00000164266	C1S ENSG00000164266
NCAPD3 ENSG00000164266	ADAMTSL1 ENSG00000164266	PTGDS ENSG00000164266	AHI1 ENSG00000164266	CFD ENSG00000164266
HPGD ENSG00000164266		MFAP4 ENSG00000164266	PIGR ENSG00000164266	THBS4 ENSG00000164266
CRACR2B ENSG00000164266		ORM2 ENSG00000164266	COTL1 ENSG00000164266	IGF2 ENSG00000164266
		CAB39L ENSG00000164266		GSTP1 ENSG00000164266
		APOD ENSG00000164266		MGP ENSG00000164266
		STARD10 ENSG00000164266		NOMO2 ENSG00000164266
		ATF5 ENSG00000164266		
		WNK2 ENSG00000164266		
		PROC ENSG00000164266		
		APOE ENSG00000164266		

Save

Quit

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.

II - 4 - B - Pattern of co-expressed genes - Gene ontology

SPINK1 ENSG00000164266

Search

Pattern selected : 1

Similarity > : %

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	ENSG0000019102

Gene Ontology

Gene Ontology Analysis - All - Patterns Matrix - SPINK1 ENSG00000164266

GO terms	P-Values	Genes
----------	----------	-------

Select a GO terms database and press 'Run'.
Genes can be copied to clipboard by double-click on the table.

ARCHS4_Cell-lines

Run Save

Barplot Heatmap Quit

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).

Gene Ontology Analysis - All - Patterns Matrix - SPINK1 ENSG00000164266

GO terms	P-Values	Genes
Adenoma	4.606729889305634e-10	AGR2 APOE EFNA1 KLK4 C3 HGS API
Adenocarcinoma	3.5534874027347053e-09	AGR2 TMSB10 APOE ANPEP EFNA1
Prostatic Neoplasms	2.0138140432939682e-08	AGR2 STARD10 APOE SPON2 KLK4 K
Malignant neoplasm of prostate	3.3494919909225766e-08	AGR2 STARD10 APOE SPON2 ANPEP
Precancerous Conditions	2.9563078179287126e-07	AGR2 IGF2 PTGDS GSTP1 MGP SPINK
Malignant tumor of colon	3.29519650115297e-07	APOE SPON2 ANPEP CLN3 KLK4 C3
Colon Carcinoma	7.075737610983052e-07	APOE SPON2 ANPEP CLN3 EFNA1 KI
Colorectal Carcinoma	4.55792525570373e-06	AGR2 APOE SPON2 EFNA1 KLK4 KEA
Liver carcinoma	6.1097097777728139e-06	ATF5 AGR2 PROC APOE SPON2 ANP
High-Grade Prostatic Intraepithelial	6.599754624149162e-06	AGR2 GSTP1 APOD SPINK1 PSCA
Vascular inflammations	9.066017026042678e-06	KEAP1 THBS4 APOE ELF3 GJB1 LCN2
Hepatitis	1.214695051421898e-05	KEAP1 C1S GSTP1 APOE CYP27A1 PE
Neoplasm Metastasis	1.321232645743605e-05	AGR2 TMSB10 APOE ANPEP DHRS7
Stomach Neoplasms	2.213155065216265e-05	IGF2 REG4 GSTP1 APOE CLDN3 CST1
Prostate carcinoma	2.296127210839665e-05	AGR2 APOE ANPEP DHRS7 EFNA1 KL

Select a GO terms database and press 'Run'.
Genes can be copied to clipboard by double-click on the table.

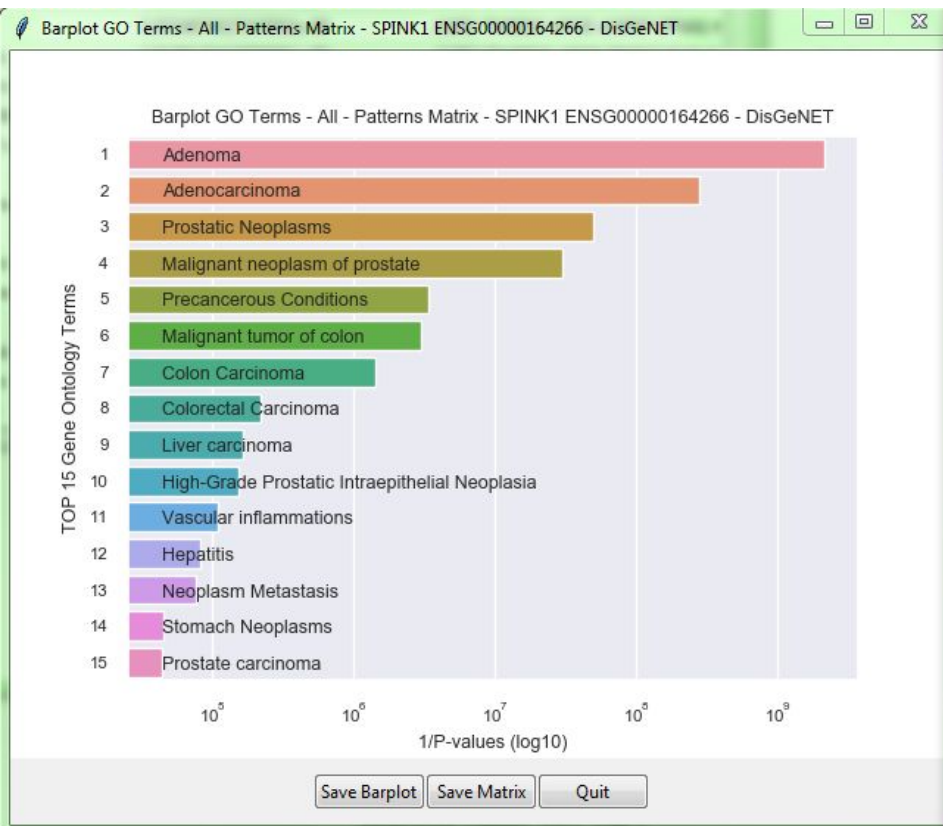
DisGeNET

Run Save

Barplot Heatmap Quit

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

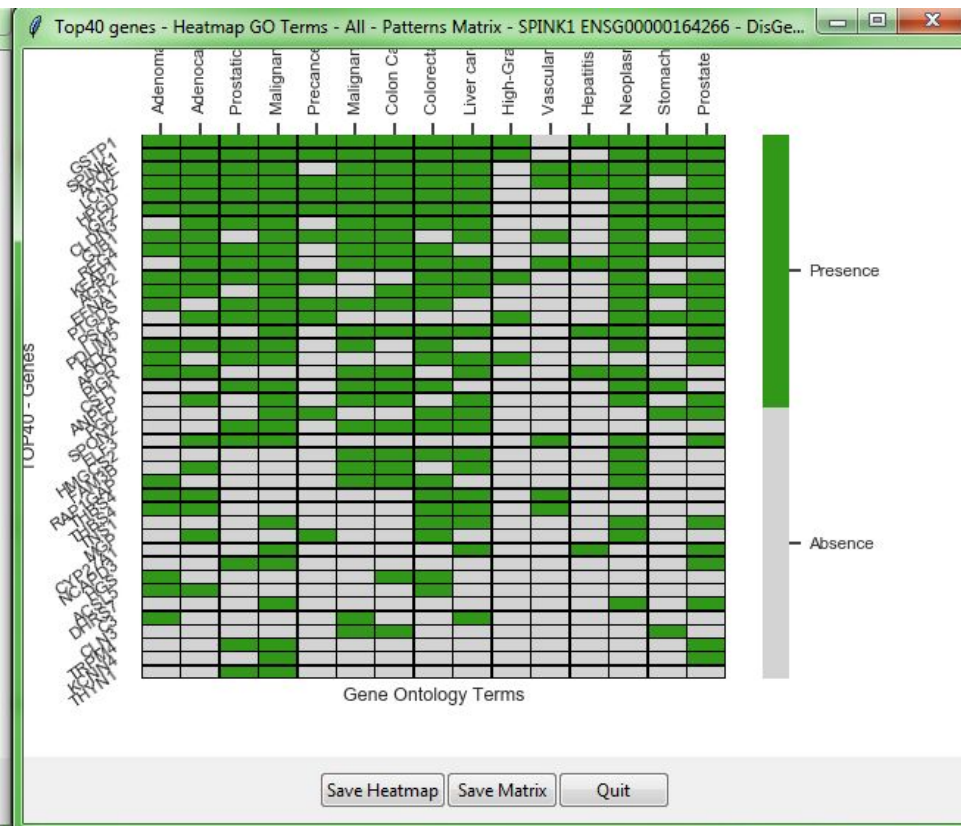
Barplot and heatmap will be introduced on the next page.



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'.

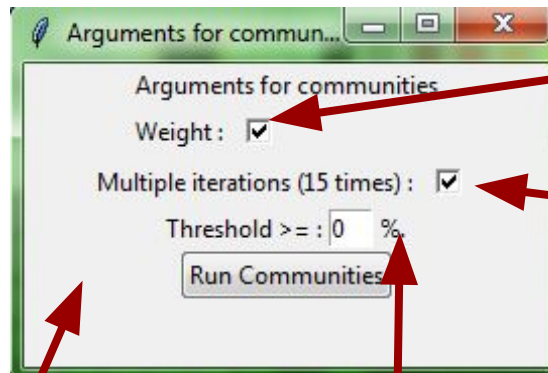
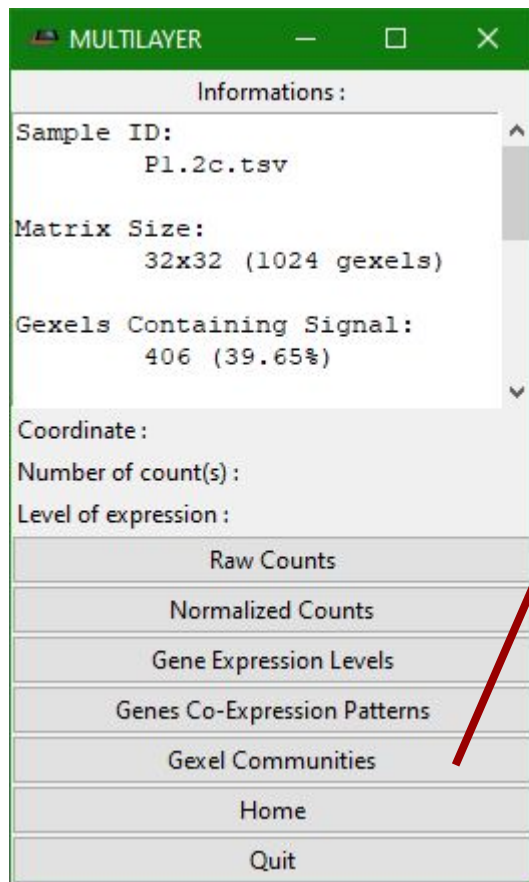


Heatmap shows TOP 15 best GO terms and the TOP 40 genes.

Save plot by clicking on 'save heatmap'.

Save the full matrix by clicking on 'save matrix'.

II - 5 - Gexel communities

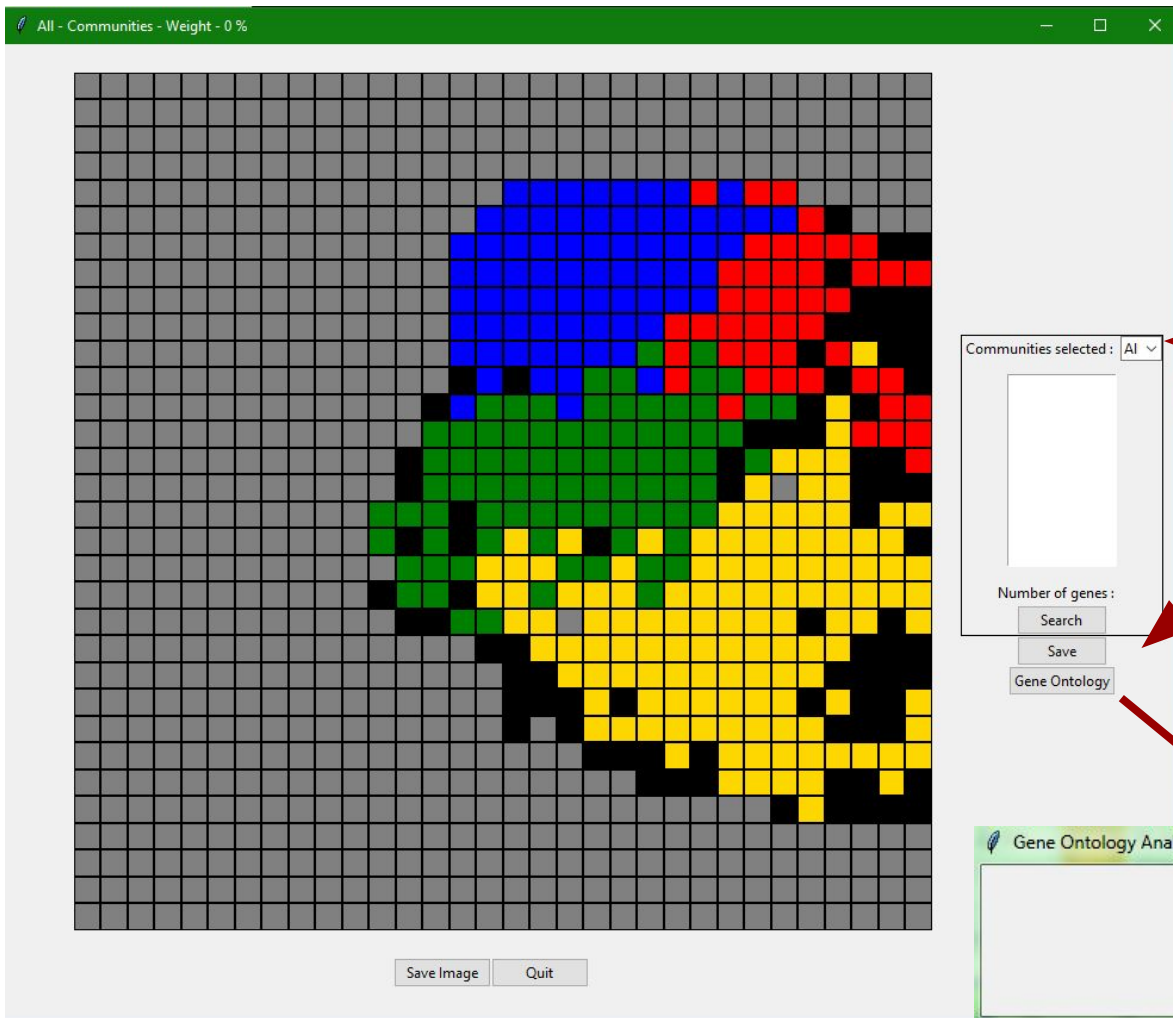


Consider the similarity weight as parameter in louvain algorithm.

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.

```
#####  
# LOG #  
#####  
1 [*] 80.0%  
2 [ 1] 13.333333333333334%  
3 [ 1] 6.666666666666667%
```

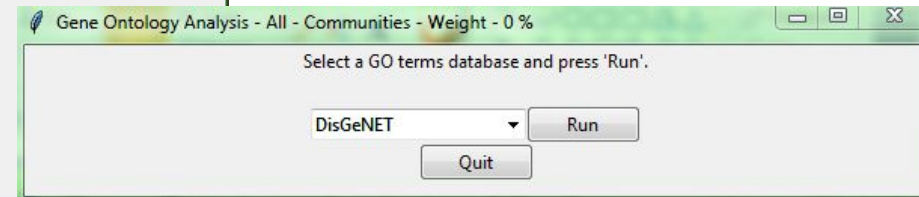



All communities are displayed. If there are overlaps between communities, the tool will display the gexel with community with the highest similarity.

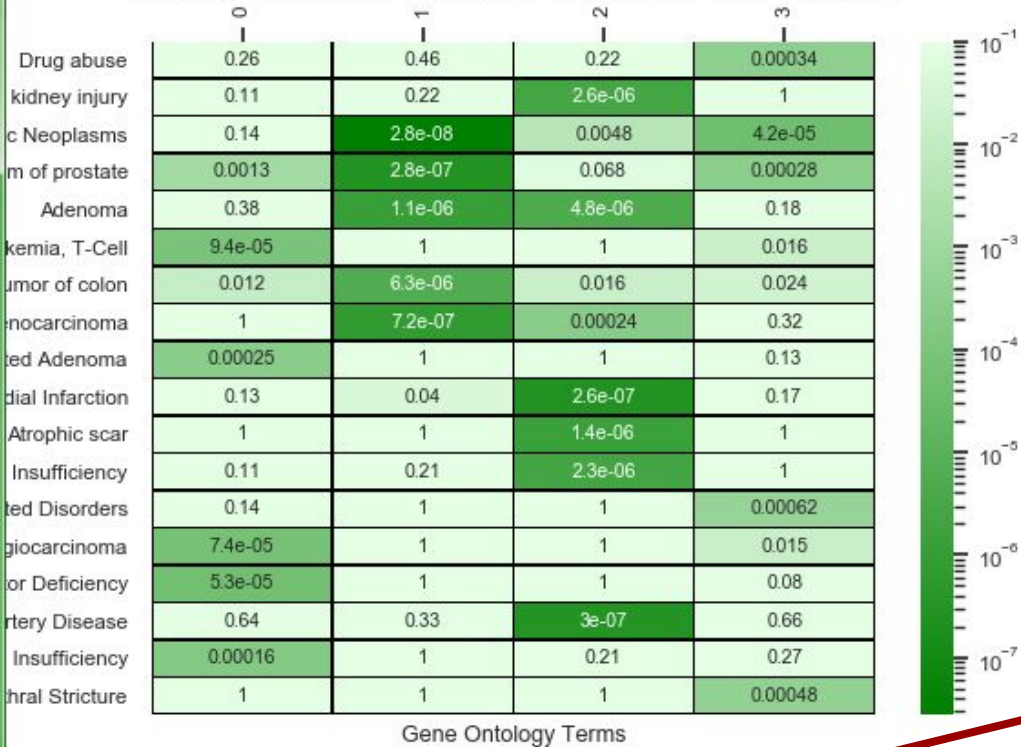
Display all communities or a specific one.

Save a tsv file (result of louvain)

Select database and 'run'.



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



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

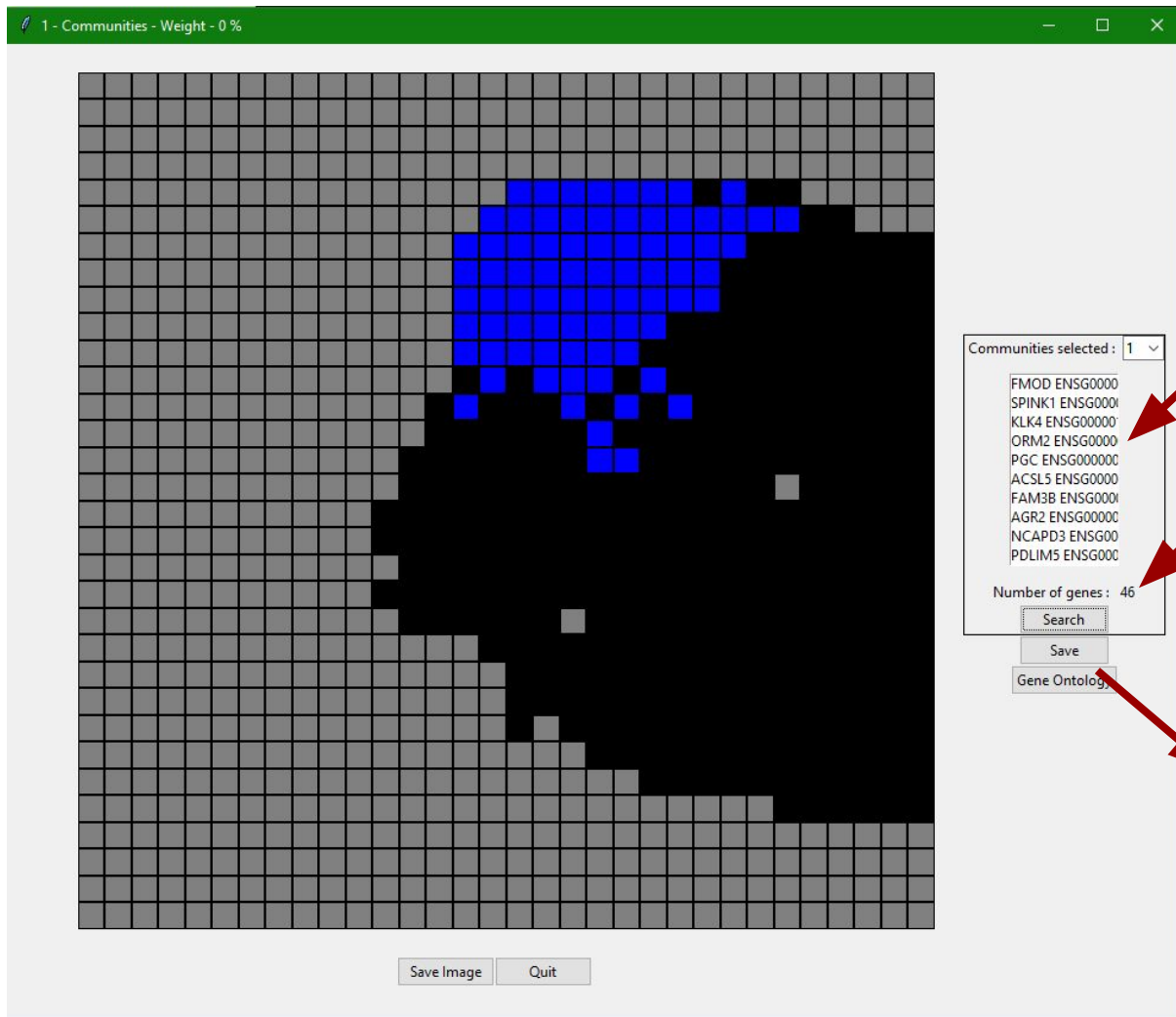
Save plot by clicking on 'save heatmap'

Save matrix by clicking on 'save matrix'

Save Heatmap

Save Matrix

Quit



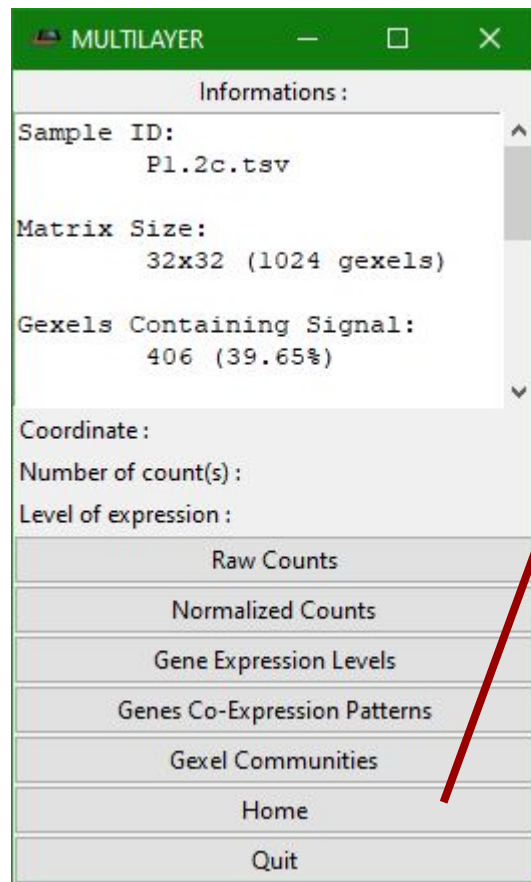
All communities are displayed. If there are overlaps between communities, the tool will display the gexel with community with the highest similarity.

If a specific community is selected, genes which compose the given community will be display.

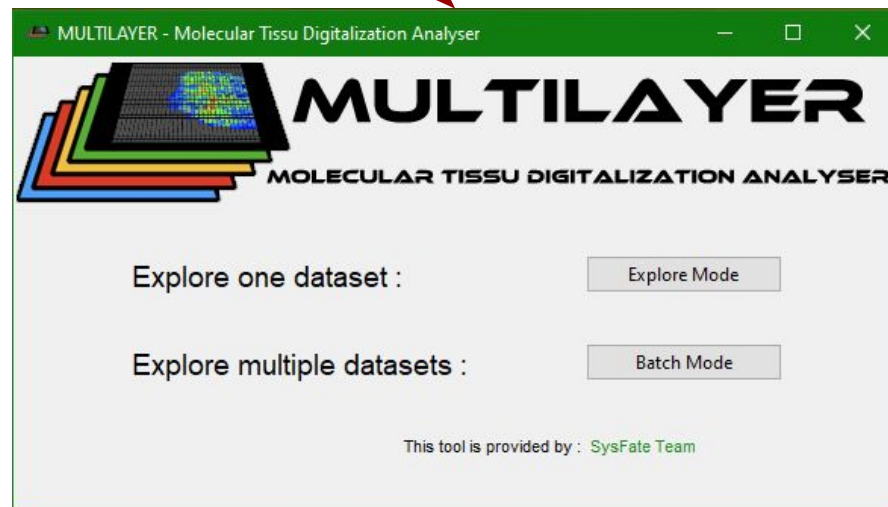
Toltal number of genes in the community

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

II - 6 - Home Button

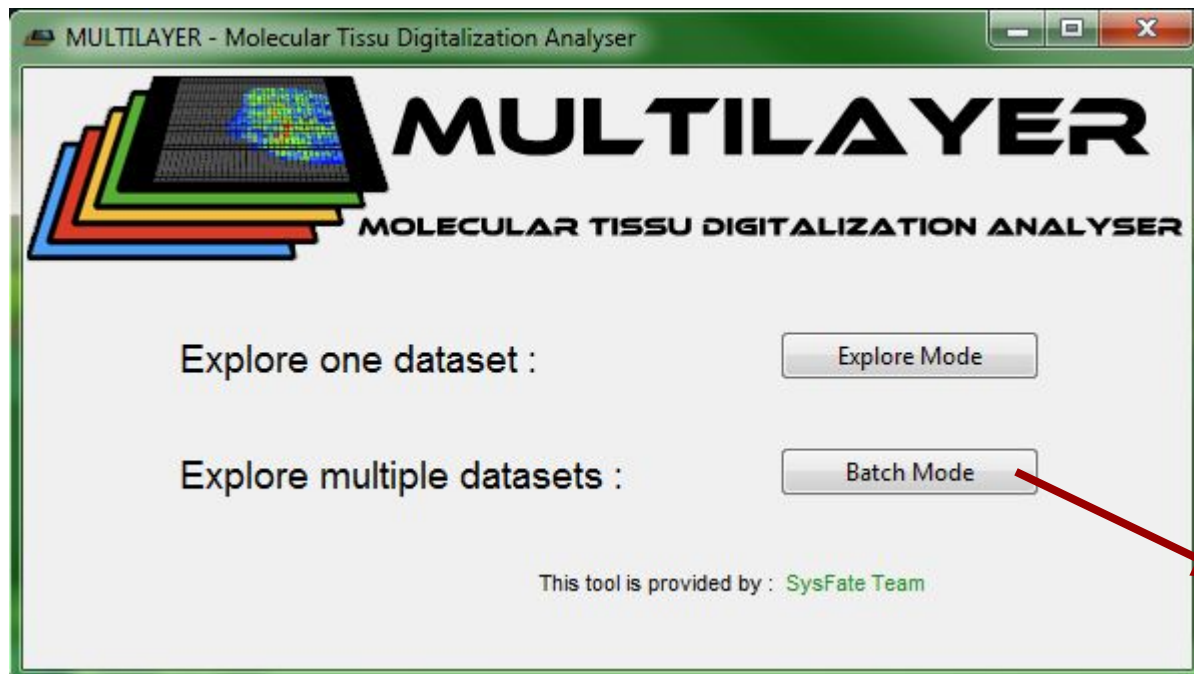


Back to home, all data loaded on this session will be lost.



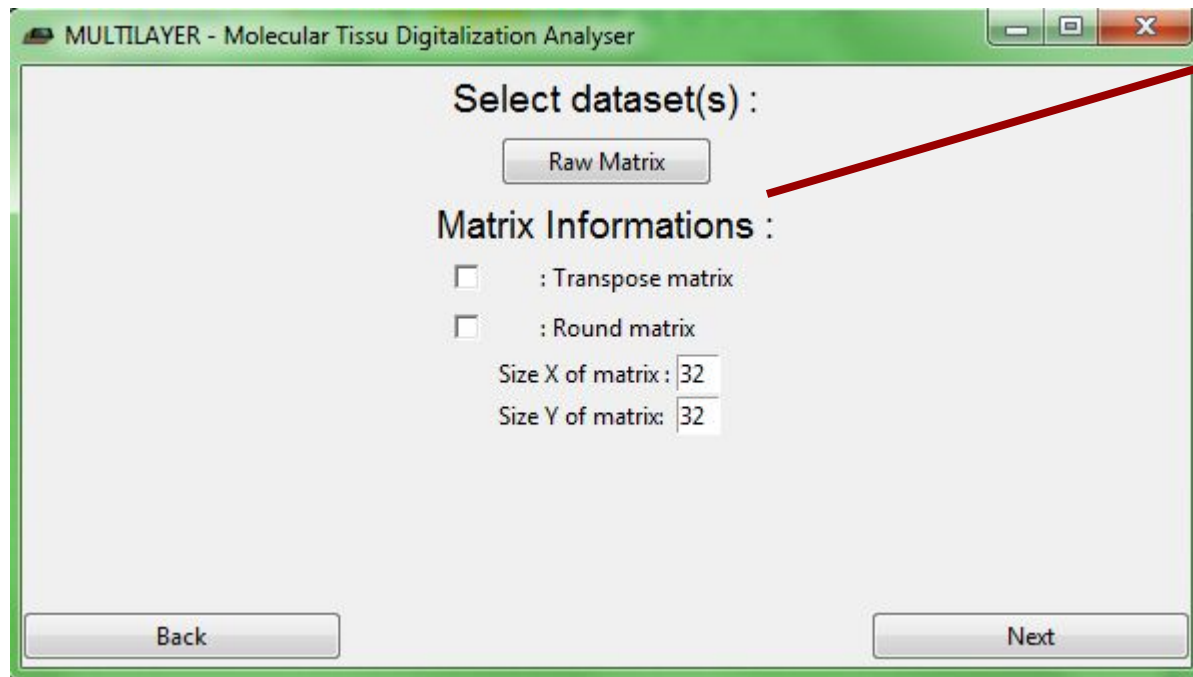
III - Batch mode with MULTILAYER (multiple samples)

III - 1 - Mode



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

III - 2 - Arguments



MULTILAYER - Molecular Tissu Digitalization Analyser

Select dataset(s) :

Raw Matrix

Matrix Informations :

☐ : Transpose matrix

☐ : Round matrix

Size X of matrix : 32

Size Y of matrix: 32

Back

Next

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.**

MULTILAYER

Select a sample :

P1.1c.tsv
P1.2c.tsv
P2.1c.tsv
P1.3c.tsv
P2.3c.tsv
P2.4c.tsv
P3.1c.tsv
P3.2c.tsv
P3.3c.tsv

Sample :

Coordinate :

Number of count(s) :

Level of expression :

Raw Counts

Normalized Counts

Gene Expression Matrix

Genes Co-Expression Patterns

Gexel Communities

Comparison datasets

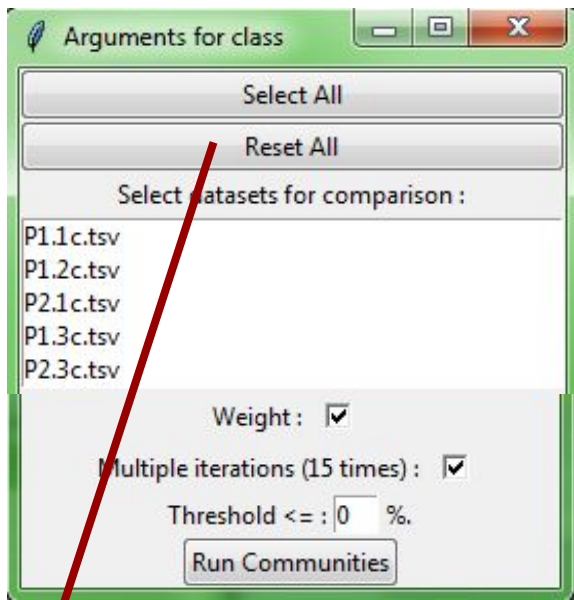
Home

Quit

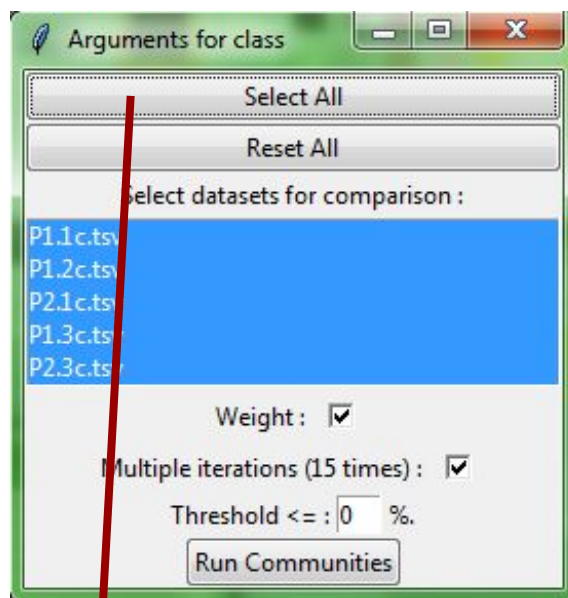
Select a specific sample for perform all previous analysis from pages 6 to 21.

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.

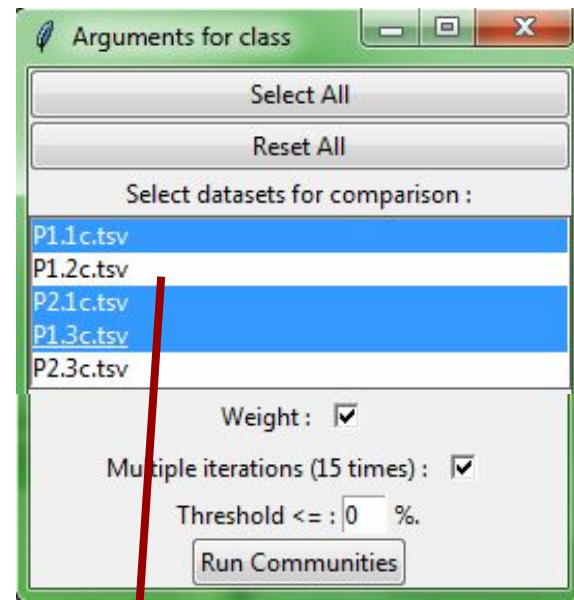
Comparision of all communities from all datasets.



Reset all selection

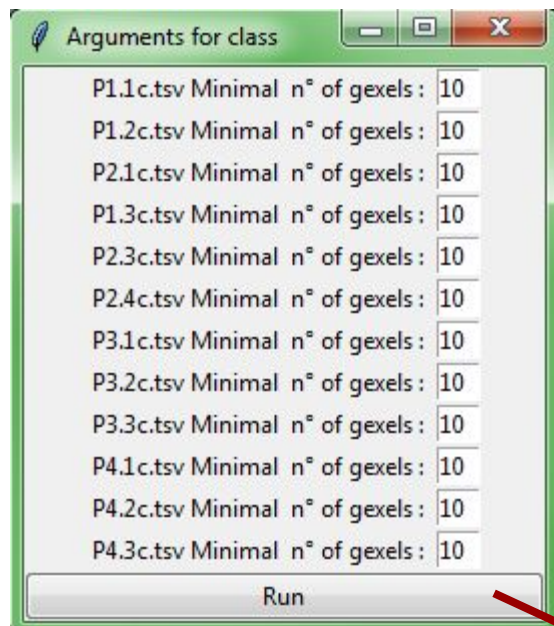


Select all datasets



Select several specifics datasets

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

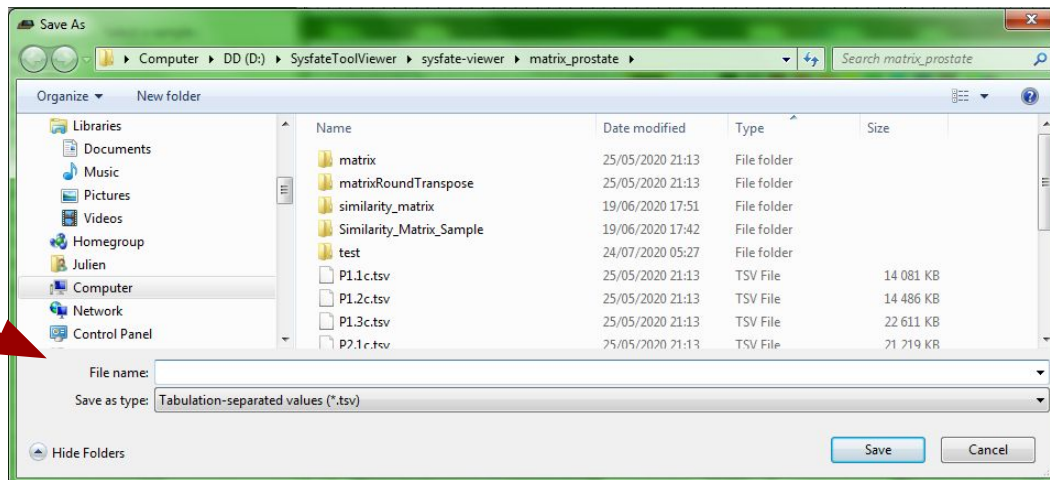


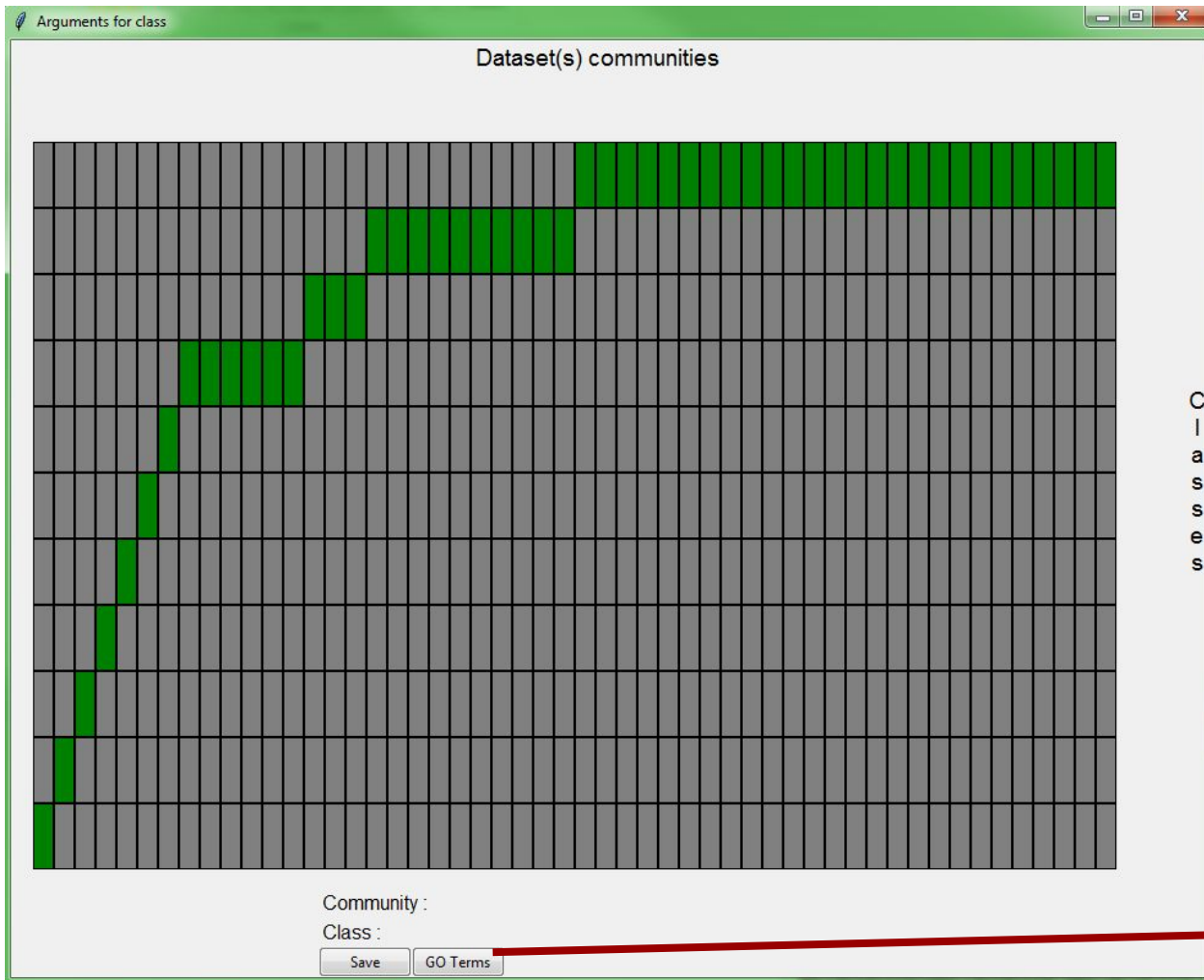
The displayed value is the default value.

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)





Interactive heatmap :

In X axis all datasets communities, in Y axis all classes.

A class is group of 1 or several communities.

C
l
a
s
s
e
s

GO terms

Gene Ontology Analysis - Class

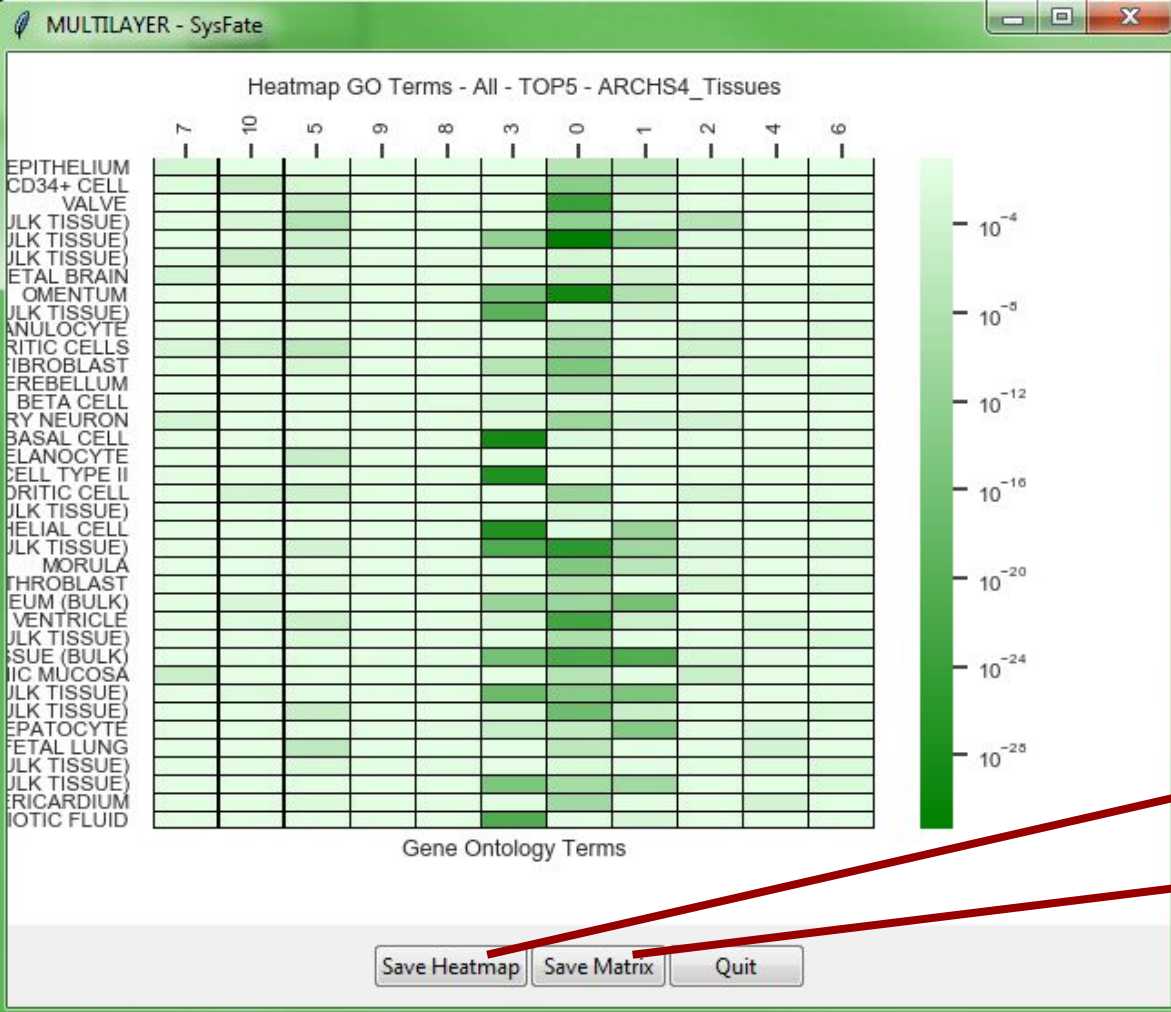
Select a GO terms database and press 'Run'.

Top GO terms : 5 ▼

DisGeNET ▼ Run

Quit

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



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.