

NaviCom

Python package and web interface to create interactive molecular network portraits using multi-level omics data

Tutorial

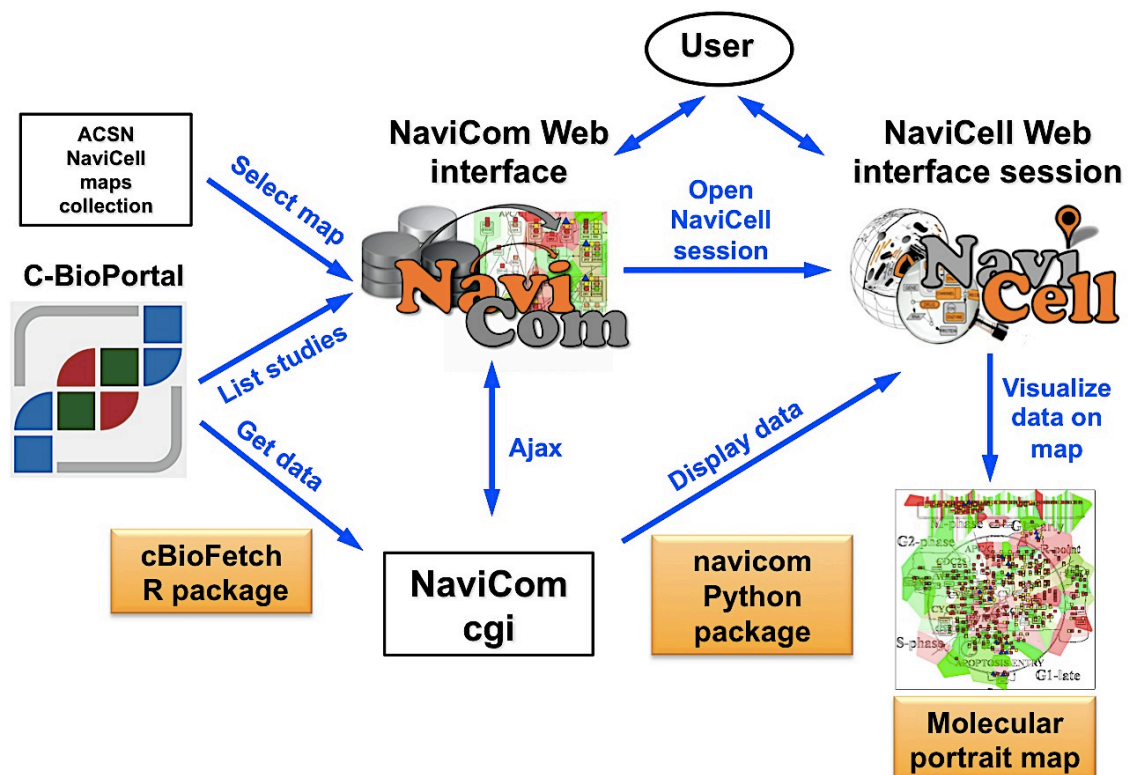


Table of Contents

1. Introduction	3
2. Instructions for data integration and visualisation	3
2.1 NaviCom web interface.....	3
2.2 Data selection	4
2.3 Map selection	4
2.3 Display modes	4
2.4 Display configuration.....	6
2.5 Processing.....	6
2.6 Errors messages.....	6
3. Using NaviCom from the command line	7
4. Application examples	8
4.1 Molecular portrait of Adrenocortical Carcinoma visualized on Cell cycle map	8
4.2 Molecular portrait of Glioblastoma visualized on Alzheimer’s signalling map	9
4.3 Molecular portrait of Sarcoma visualized on Ewing’s sarcoma map	10
4.4 Molecular portrait of Breast Invasive Carcinoma visualized on Cell cycle map	11

1. Introduction

[NaviCom](#) connects between [cBioPortal database](#) and [NaviCell web service](#) and allows to visualize various high-throughput data types simultaneously on the network maps in one click. NaviCom is empowered by a [cBioFetchR](#) R package to retrieve data from cBioPortal and the [python module naviacom](#) to display those data on NaviCell signalling network maps. The interactive maps with displayed data can be browsed and analyzed in [NaviCell](#) environment.

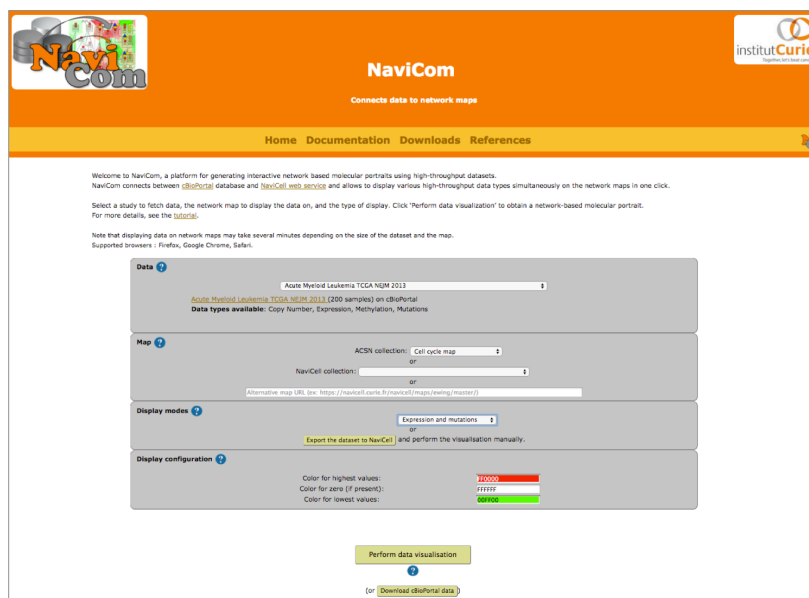
The studies available in the cBioPortal database contain large-scale cancer data sets including expression data for mRNA, microRNA, proteins expression; mutation, gene copy number and methylation profiles and beyond.

Depending on the nature of data, different types of visualisation modes can be required to achieve the informative picture. NaviCom provides a default visualisation setting for simultaneous integration of the data into the big comprehensive maps of molecular interactions to visualize a complex molecular portrait. Once chosen, these settings are applied automatically, significantly reducing the time required to perform the visualisation comparing to manual mode. It also allows launching the visualisation of several datasets on different maps in parallel.

The resulting maps with visualized data on top of them are interactive, they can be browsed using NaviCell Google-based navigation features that allow to visualize the data at different zoom levels, starting from the top level view, where patterns of integrated data can be grasped, up to the most detailed view at the level of individual molecules. In addition, since the whole dataset is already imported to NaviCell in a form of data tables, the user also may apply different types of analyses provided by NaviCell environment.

2. Instructions for data integration and visualisation

2.1 NaviCom web interface

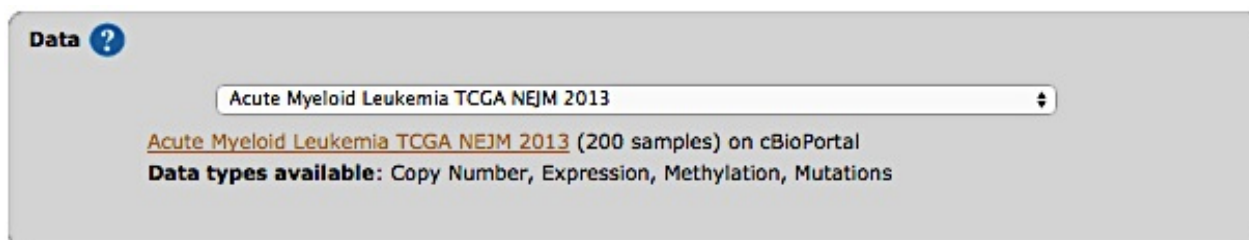


The screenshot displays the NaviCom web interface. At the top, there is a logo for NaviCom and Institut Curie. Below the header, a navigation bar includes links for Home, Documentation, Downloads, and References. The main content area contains a welcome message and instructions for using the platform. A central panel allows users to select a study, map, and display mode. The 'Data' section shows 'Acute Myeloid Leukemia TCGA NEM 2013' with 200 samples. The 'Map' section offers 'ACSIS collection' and 'Cell cycle map'. The 'Display modes' section includes 'Expression and mutations'. The 'Display configuration' section shows color scales for highest and lowest values. A 'Perform data visualisation' button is at the bottom, with a link to 'Download cBioPortal data'.

NaviCom platform is accessible via a web page that provides interactive selection panel to define study dataset, map and data display mode, to execute the visualisation and full downloadable [documentation](#) on the tool.

2.2 Data selection

The list of studies is obtained through the cBioPortal API, and thus contains all studies available from cBioPortal at the moment. Once the study is selected, a short summary including types of data and number of samples available, is provided.



Data ?

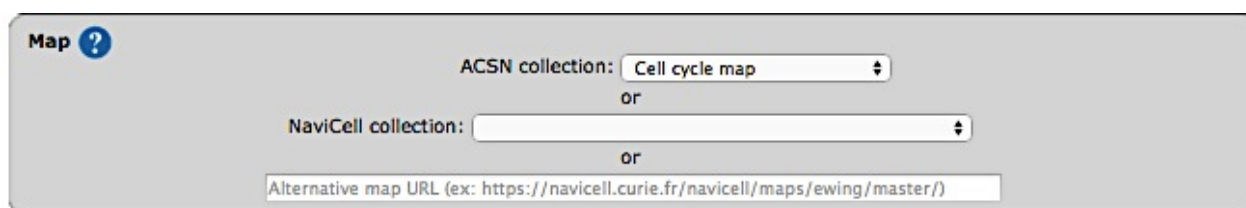
Acute Myeloid Leukemia TCGA NEJM 2013

Acute Myeloid Leukemia TCGA NEJM 2013 (200 samples) on cBioPortal
Data types available: Copy Number, Expression, Methylation, Mutations

NOTE: TCGA provisional studies have not been published yet, and can be subjected to restriction concerning their use in publication.

2.3 Map selection

Select a map from the set of curated maps of the [Atlas of Cancer Signaling Network collection](#), from the [NaviCell collection](#), or provide an URL to another NaviCell map. The user can import and visualize data via NaviCom on any type of signaling networks.



Map ?

ACSN collection: Cell cycle map

or

NaviCell collection:

or

Alternative map URL (ex: <https://navicell.curie.fr/navicell/maps/ewing/master/>)

ACSN maps collection

- ACSN global map
- Apoptosis map
- Cell survival map
- EMT and cell mobility map
- Cell cycle map
- DNA repair map







NaviCell maps collection


- Ewing Sarcoma signaling network
- Crosstalk between Notch and p53 signaling pathways
- Signaling pathways of Alzheimer's disease
- Mast Cell Activation Network

NOTE: The NaviCell collection map will only be used if no alternative URL is provided, and the ACSN collection map will only be used if no alternative URL is provided and no NaviCell collection map is selected.

2.3 Display modes

In NaviCell data can be graphically represented in a form of barplots, heat maps, glyphs or projected as a map staining, unique display mode coloring background areas around each entity according to the value of data value associated to this entity (Bonnet et al., 2015). NaviCom uses this graphics to overlay multiple data types. The settings are standardized and data display modes are pre-defined in NaviCom, projecting high-throughput data on NaviCell maps in automatized manner. Optimized visualization mode is chosen for each data type to achieve meaningful integrated view on complex molecular profile in the context of signalling network map.

Data display setting for visualization using NaviCom			
Data type	Visualization mode	Data display	Units
mRNA expression	 Up Down	Map staining	Level
Gene copy number		Heat map	Count
Mutation data		Glyph 1	Frequency
Methylation data		Glyph 2	Intensity
miRNA expression		Glyph 3	Level
Protein expression		Glyph 4	Level

NOTE: Click  icon on the upper panel of NaviCom homepage to open the display setting legend

Map staining:

The principle of map staining is in using the background of the map for visualizing the values mapped to individual molecular entities or group of entities (modules). The resulting colorful background of the network map, provides a possibility to grasp differences in the patterns of data distribution between samples or between groups of samples. To define a territory in the background occupied by a single molecular entity, the whole map territory is divided accordingly to the Voronoi cells computed from the positions of the nodes in the map. The sizes of the Voronoi cells are limited in order to avoid very large cells in the regions of the map empty from proteins or genes. Each Voronoi cell is then colored with semi-transparent color corresponding to the value mapped to the biological entity located in the center of the cell.

Heat map:

The data is displayed as color squares. Each square represents a sample in the data table, whereas the color corresponds to the individual values represented in the data table.


Glyph:


Glyph has three characteristics: shape, color and size, each one of those characteristics can be configured according to a different feature in the data. In NaviCom, the shape and the colour are associated with the data type and the size reflect the value in the data table.

NaviCom display modes:

Complete display	all data available for a given dataset
Mutations ad genomic data	expression-map staining/ copy number-heat map/ mutations-glyph 1
Expression and mutations	expression-maps staining/ mutations-glyph 1
Expression and proteomics	expression-maps staining/ proteome-glyph 4
Expression and miRNA	expression-maps staining/ miRNA-glyph 3
Expression and methylation	expression-maps staining/ pmethylation-glyph 2
Expression	expression-map staining

Select data display mode according to the data available in the chosen dataset and the scientific question to answer.

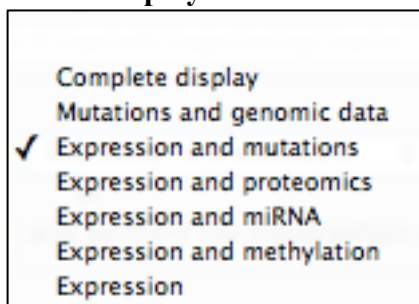
Display modes 

Expression and mutations 

or

[Export the dataset to NaviCell](#) and perform the visualisation manually.

Display modes

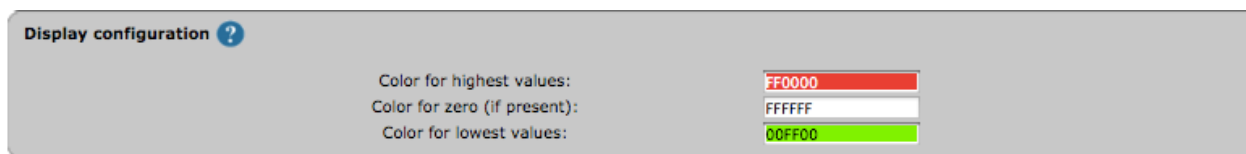


NOTE: The NaviCom platform provides the possibility to export the dataset from cBioPortal directly to NaviCell in order to perform some custom visualisations.

These pre-defined settings are flexible; after the data integration into the maps, the user can continue adjusting the visualization modes according to the scientific question.

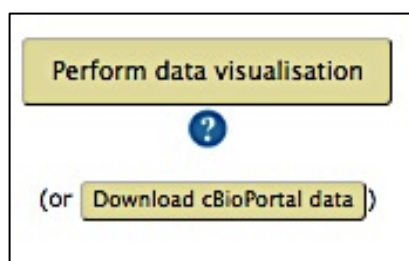
2.4 Display configuration

The color gradient for map staining and heatmaps can be adjusted. For colour adjustment, use the color selector or directly type a color hexadecimal code



2.5 Processing

After validation of the visualisation, a loading symbol will appear on the NaviCom page, which will terminate once the cBioPortal data have been downloaded by the navicom server.



NOTE: The NaviCom platform provides the possibility to download the dataset from cBioPortal directly to the user's computer.

The map with displayed data according to the chosen configurations will be opened in a separate NaviCell browser section. The resulting maps with visualized data on top of them are interactive, they can be browsed using NaviCell Google-based navigation features.

2.6 Errors messages

"Gateway TimeOut": wait a few minutes and try again

Meaning: the requested data were not cached yet, and were too big to be downloaded in less than

5 minutes.

"Invalid map": update HUGO names of entities on the map

Meaning: map entities are not annotated correctly, NaviCom cannot recover the HUGO names.

"Not a NaviCell map": update the url or update the correct map

Meaning: the URL or the map is not correct.

3. Using NaviCom from the command line

The activity of NaviCom requires the R package [cBioFetchR](#) and the [Python module navicom](#). With these packages, it is possible to use NaviCom with the command line, granting more flexibility in various ways, such as the configuration of the display in NaviCell or the list of genes downloaded from cBioPortal.

Download and save data from cBioportal using cBioFetchR

```
# Load the cBioFetchR package and connect to cBioPortal
library ( cBioFetchR )
conn = cBioConnect ( " http : // www . c b i o p o r t a l . o r g / " )
# Select the study id
listStudies ( conn , " ovarian " )
# Get the data from cBioPortal and prepare the data for
NaviCell export
datas = cBioNCviz ( " ov_t c g a _ p u b " , genes_list = " http : //
acsn . c u r i e . f r / f i l e s / a c s n _ v 1 . 1 . g m t " , name = "" ,
method = " p r o f i l e s " )
# Export in a format readable by the python package
saveData ( datas )
```

Import data in python

```
from navicom import _
# Prepare the connection
nc = NaviCom ( map_url = ' http s : // n a v i c e l l . c u r i e . f r /
n a v i c e l l / m a p s / c e l l c y c l e / m a s t e r / i n d e x . p h p ' , browser_
command = " google □ chrome %s " , display_config =
DisplayConfig ( 5 , na_color = " f f f f f f " ) )
# Load data for the visualisation
nc . loadData ( fname = " Ovarian _ Serous _ Cystadenocarcinoma _
TCGA _ Nature _ 2011 . txt " )
```

Select and display data on the NaviCell map

```
nc . listData ( ) # Get the data available
nc . listAnnotations ( ) # Get the groups available
# Display the data on the map
nc . completeDisplay ( )
# Display display for a specific group
nc . completeDisplay ( sample = " GRADE : DiseaseFree " )
```


4. Application examples

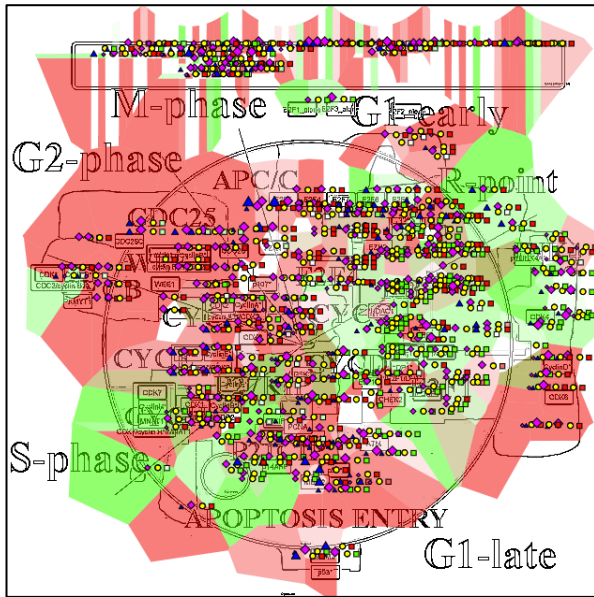
4.1 Molecular portrait of Adrenocortical Carcinoma visualized on Cell cycle map

Dataset: Adrenocortical Carcinoma TCGA Provisional

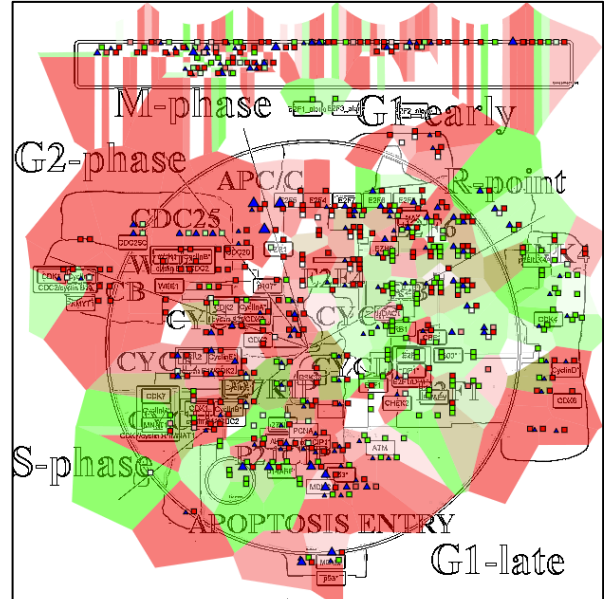
Data types available: Copy Number, Expression, Methylation, Mutations, Proteomics

Samples: 92

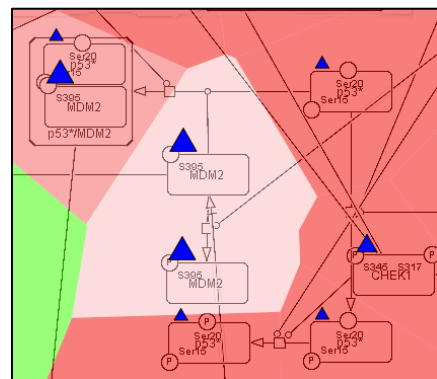
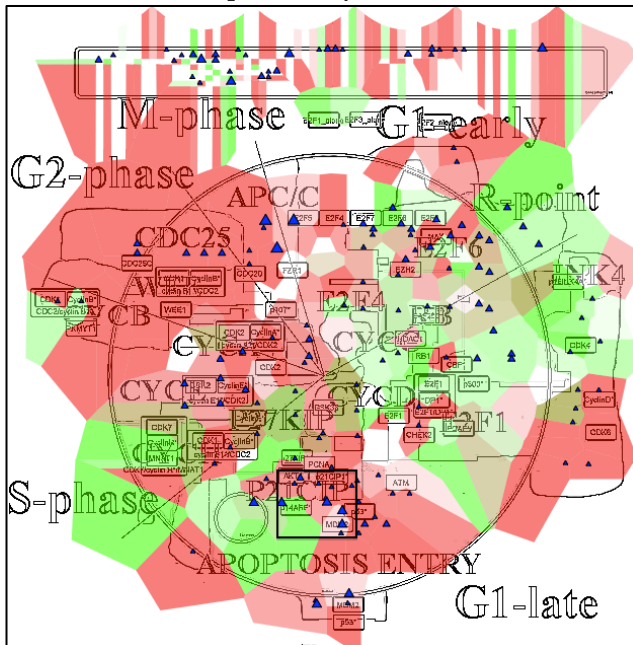
Expression/copy number/mutations/
methylation/proteome



Expression/copy number/mutation



Expression/mutations



Visualisation settings:

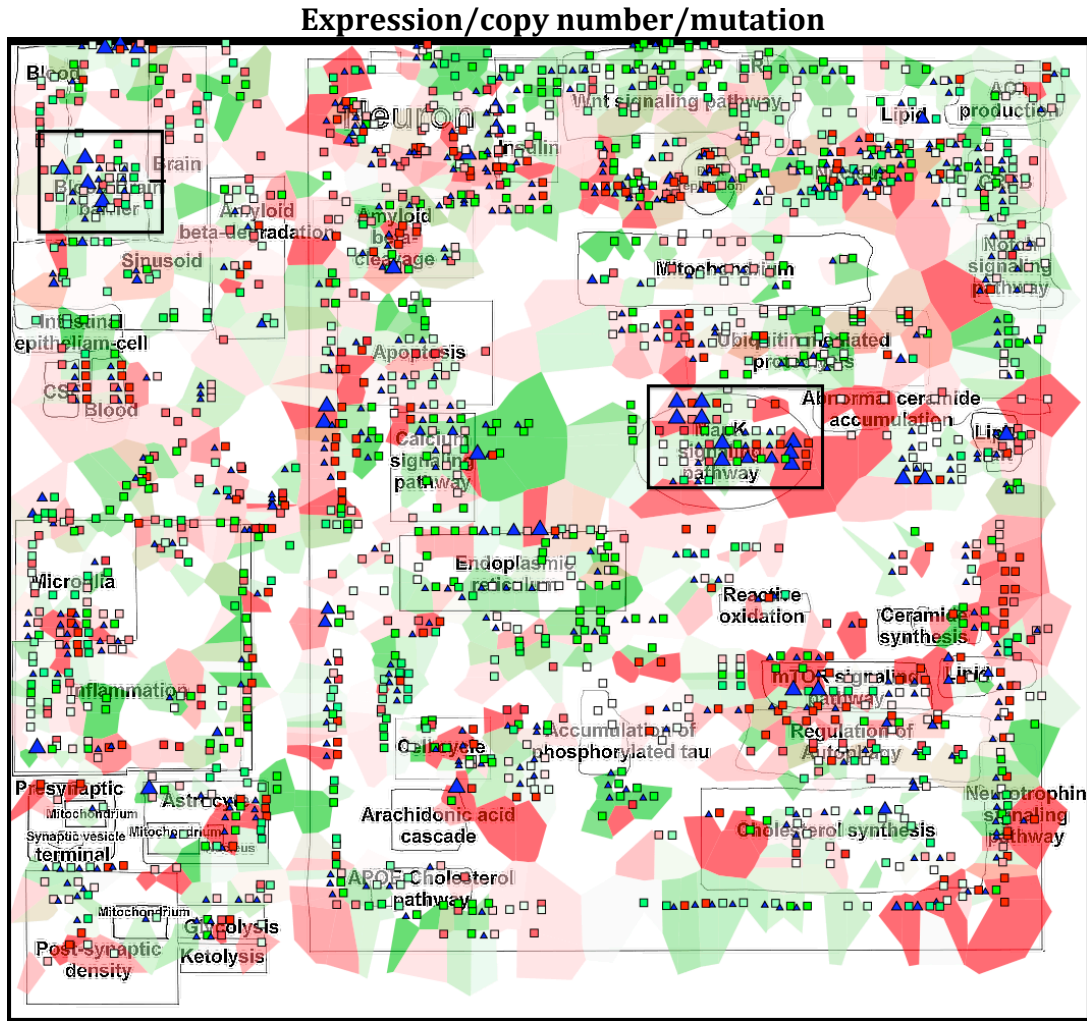
expression-map staining/ copy number-heat map/ mutations-blue triangle/ methylation-pink diamond

4.2 Molecular portrait of Glioblastoma visualized on Alzheimer's signalling map

Dataset: Glioblastoma TCGA Cell 2013

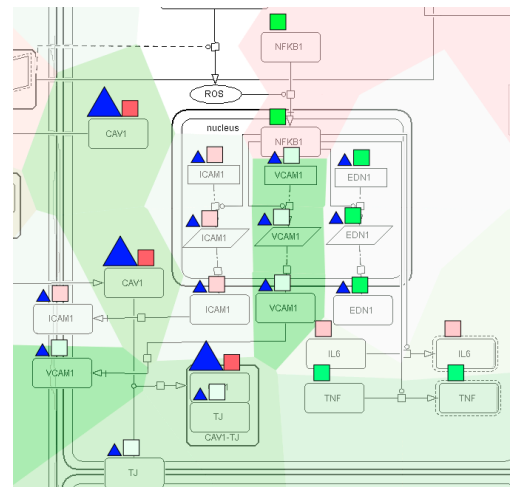
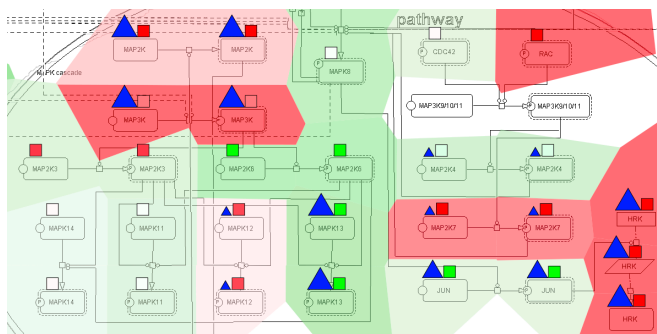
Data types available: Copy Number, Expression, Mutations, Proteomics

Samples: 580



MAPK pathway

Blood-Brain barrier signalling



Visualisation settings:

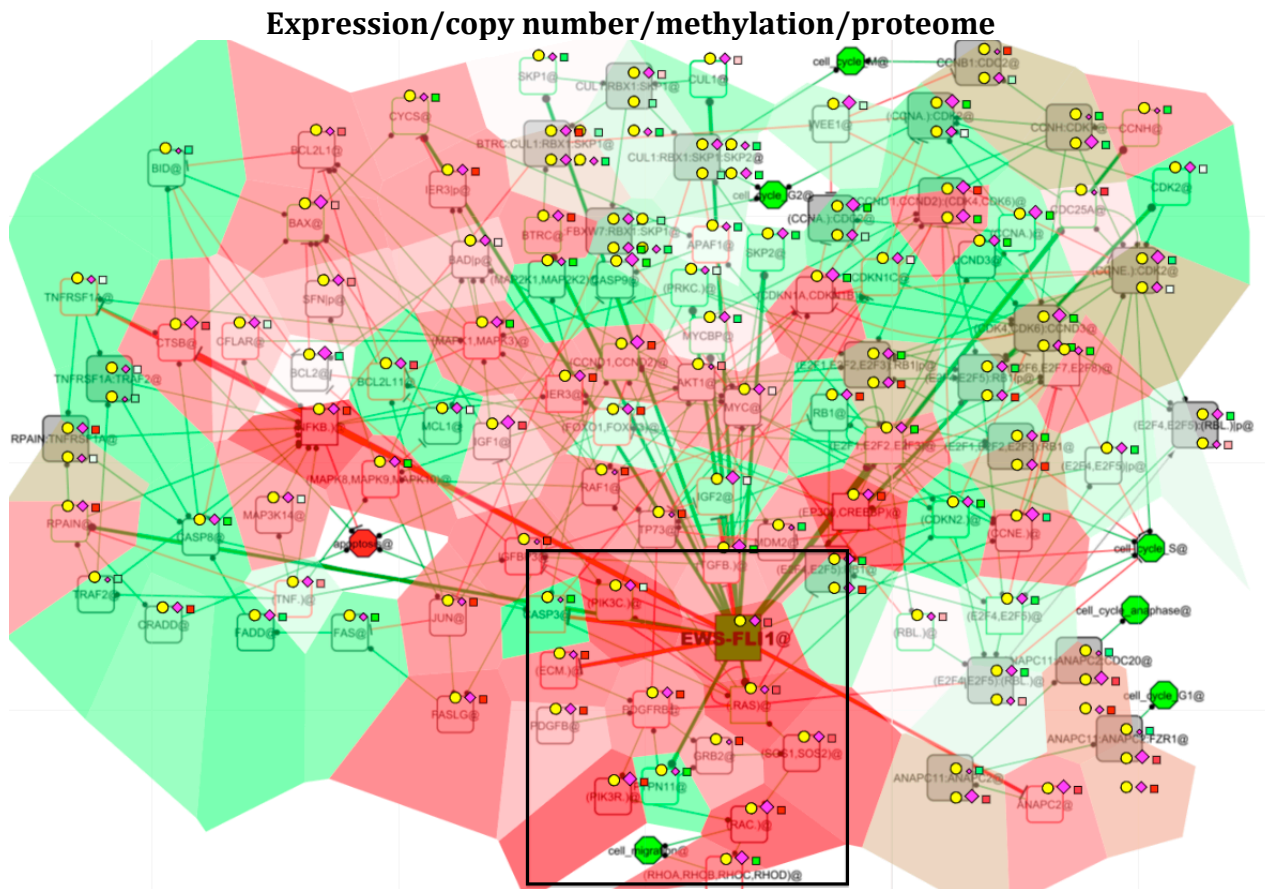
expression-map staining/ copy number-heat map/ mutations-blue triangle

4.3 Molecular portrait of Sarcoma visualized on Ewing's sarcoma map

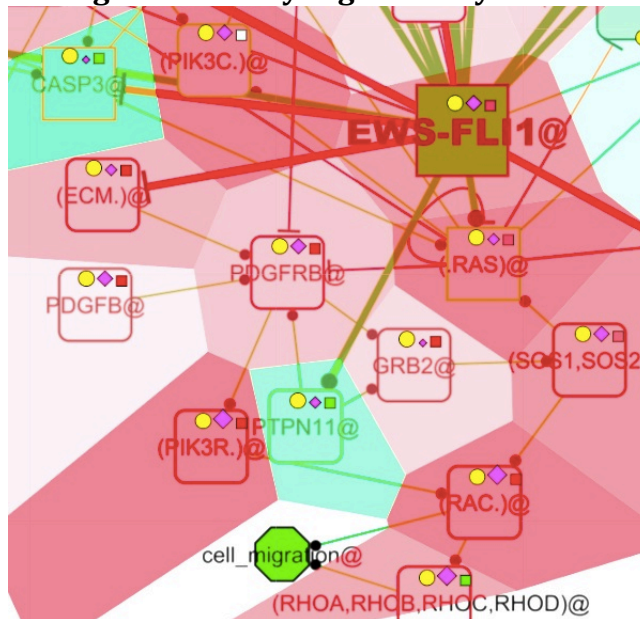
Dataset: Sarcoma TCGA Provisional_264 samples_

Data types available: Copy Number, Expression, Methylation, Proteomics

Samples: 264



Oncogenes directly regulated by EWS-FLI1



Visualisation settings:

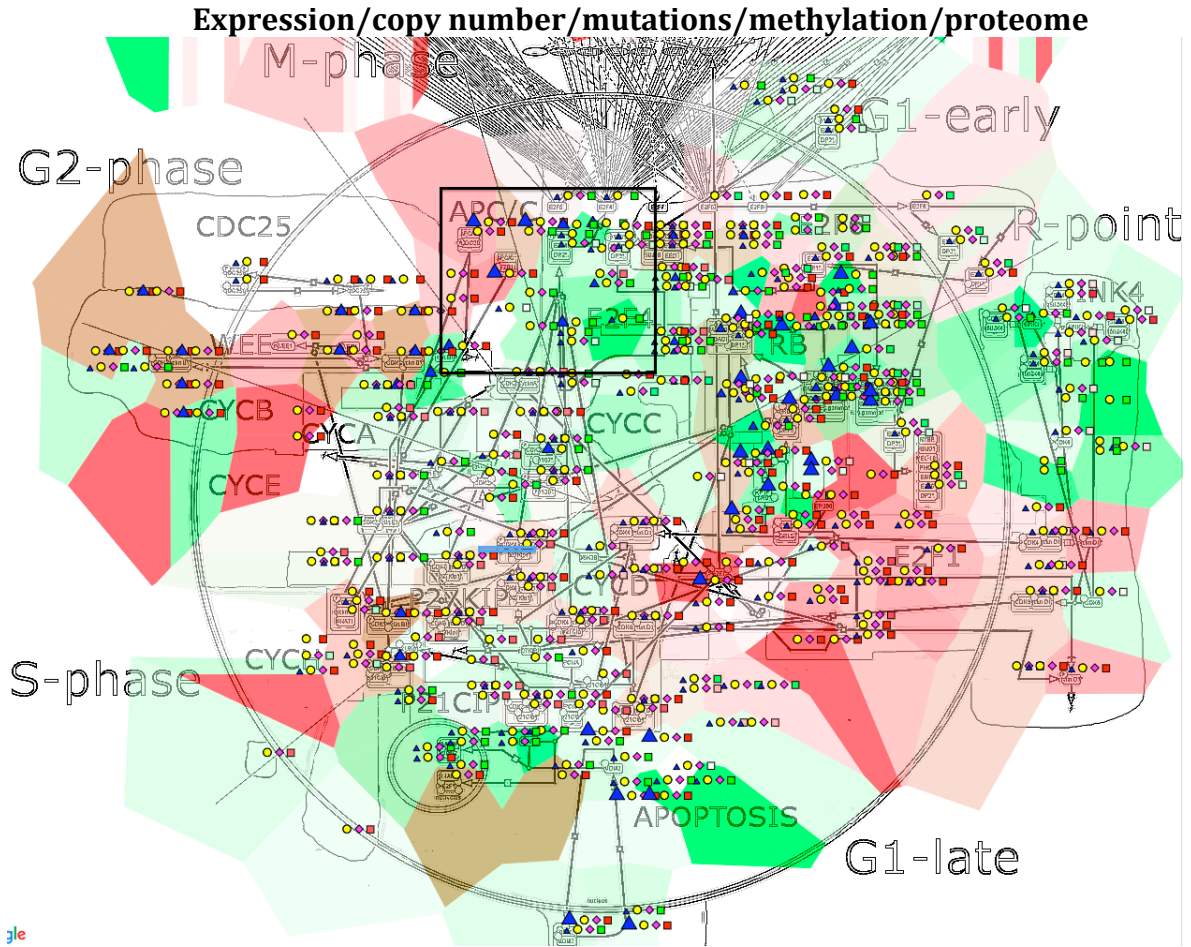
expression-map staining/ copy number-heat map/ methylation-pink diamond: proteomics-yellow circle

4.4 Molecular portrait of Breast Invasive Carcinoma visualized on Cell cycle map

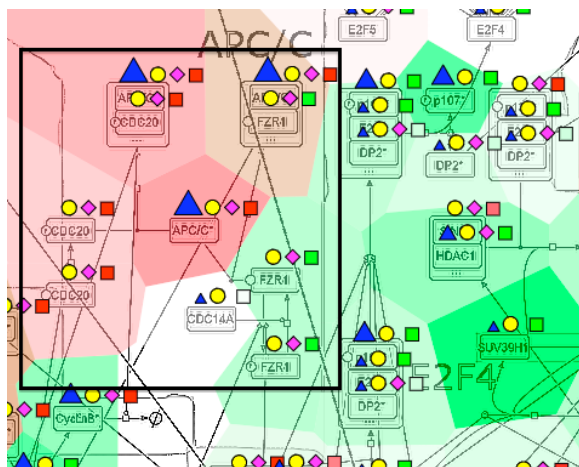
Breast Invasive Carcinoma TCGA Nature 2012

Data types available: Copy Number, Expression, Methylation, Mutations, Proteomics

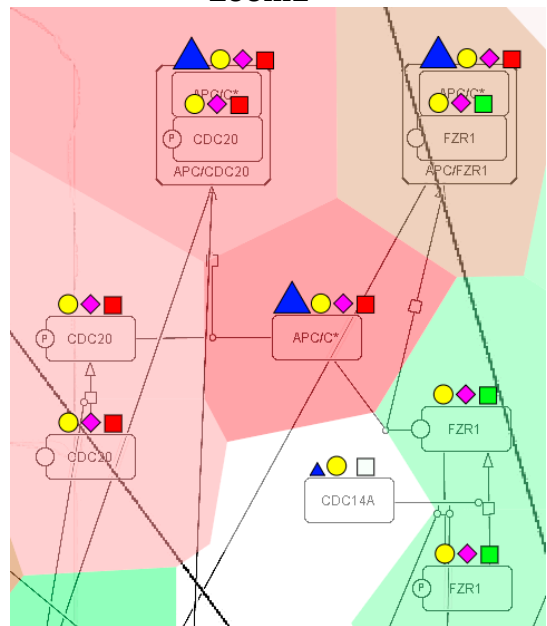
Samples: 825



Zoom1



Zoom2



Visualisation settings:

expression-map staining/ copy number-heat map/ mutations-blue triangle/ methylation-pink diamond/proteomics-yellow circle