D7.2

Design and integrate pathway visualization

Project number:	668858
Project acronym:	PrECISE
Project title:	PrECISE: Personalized Engine for Cancer Integrative Study and Evaluation
Start date of the project:	1st January, 2016
Duration:	36 months
Programme:	H2020-PHC-02-2015

Deliverable type:	Other
Deliverable reference number:	PHC-668858 / D7.2 / v1.0
Work package contributing to the deliverable:	WP7
Due date:	December 2017 – M24
Actual submission date:	18th December 2017

Responsible organisation:	CI
Editor:	Laurence Calzone
Dissemination level:	PU
Revision:	v1.0

Abstract:	The visualization of genomic, transcriptomics and proteomics data onto maps describing molecular processes proves to be informative. At a first glance, by mapping these data onto maps of processes known to be altered in cancer, some signalling pathways or processes show a higher expression in specific patients or groups of patients and suggest which deregulations are associated with which (groups) of patients. In this deliverable, we demonstrate how visualization of expression data or of pathway activity can provide a first understanding of differences between patients with different clinical information.
Keywords:	data visualization, ACSN, NaviCell, proteomics and RNAseq visualization, pathway analysis

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 668858.

This work was supported (in part) by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 15.0324-2. The opinions expressed and arguments employed therein do not necessarily reflect the official views of the Swiss Government.

Editor

Laurence Calzone (CI)

Contributors (ordered according to beneficiary numbers)

Matteo Manica (IBM)

Nicolas Sompairac (CI)

Laszlo Puskas (ABT)

Disclaimer

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The users thereof use the information at their sole risk and liability.

Executive Summary

In the report, we show how it is possible to visualize genomic and expression data onto networks of cancer signalling pathways. Using a database of pathways that has been developed by CI partner, we give a step by step procedure on how to choose a map of a particular biological process and proceed to the mapping of available data onto these maps.

Mutation data, transcriptomics and proteomics data can be visualized simultaneously for a patient or a group of patient. We can then compare and interpret the visual portraits obtained for different cases.

Visual profiles of tumour stages, patient classification, inferred clones and tumour clonal composition, identified diagnostic, prognostic and therapeutic biomarkers, dysregulated pathways, suggested drug combinations and their mode of action can all be generated using the pathway database ACSN (atlas of cancer signalling network) using NaviCell software. Navicell is based on google-map technologies and allows the navigation through large maps as well as the mapping of data onto these maps.

For this report, we concentrated on the visualization of patient profiles onto networks centred on metabolic, cell signalling and immune signalling: RECON2, ACSN, and Innate immunity response in cancer meta-map. These networks are available online on the NaviCell platform (https://navicell.curie.fr/index.html). RNA expression, proteomics and mutation data for prostate cancer were used and observed on top of these three networks:

The mutation and expression data were visualized for individual patients. We also performed some analyses using ROMA (Representation of Module Activity). This software is designed to quantify the activity of a certain gene set or pathway. Using ROMA, we can show that certain signalling pathways behave differently between samples, their location or the Gleason group.

Contents

List of Figures

Chapter 1 Introduction

1.1 The visualization of data onto networks - Motivation

Biological knowledge is disseminated in many articles, in very different experiments with different conditions on different cell lines or animal models.

Cancer is known to be a network disease. A pathway that bears a mutation may affect many downstream events because of the numerous cross-tallks between the signalling pathways. To summarize all these interactions and cross-talks, we can visualize the pathways in the form of a graph where nodes are biological entities (proteins, genes, complexes, modified proteins, etc.) and edges can represent biochemical interactions (synthesis, degradation, phosphorylation, complexation, etc.). All information concerning an entity or a reaction can be annotated. These networks allow to

- Identify what is known and not known about biochemical interactions
- Find references for interactions
- Recapitulate knowledge about molecular mechanisms
- Integrate biochemical interactions from different publications into a comprehensive network

This type of representations can be used to understand the data in terms of signalling pathways. When mapping data on top of these networks, some insight can be gained about what is over-activated or under-activated in different types of samples: normal versus tumoural, for instance, or cell lines treated versus non-treated.

When we refer to the mapping of data, we mean that some nodes or areas around the nodes can be coloured according to their expression or their activity.

Visualizing data on these maps allow to

- Explore the data using a pathway-based approach
- Identify the pathways that are most affected in some cancers / patients

1.2 Visualization using SmartBioBank for PrECISE members

The ABT team collaborated with BCM and developed a secure biobank system named SmartBioBank (https://smartbiobank.astridbio.com) under Task 7.1. The ABT team upgraded the databank dashboard with a direct link to maps available in ACSN and in NaviCell repository, where the data can be visualized by the partners.

The pathway visualization tool, therefore, was integrated into the SmartBioBank system, where data management can be done in a secure way. The following snapshot shows how a partner can log in either to access a specific SmartBioBank project or to manage data visualizations for the ACSN/NaviCell maps.

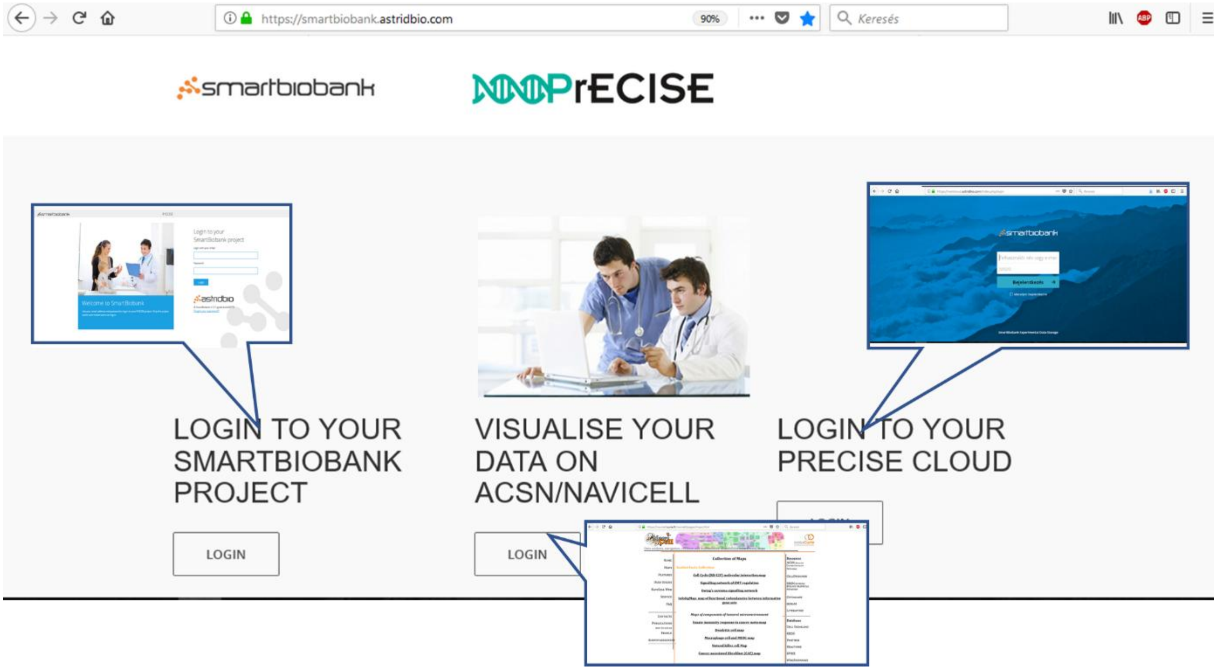

https://smartbiobank.astridbio.com
90%
Keresés
smartbiobank
PrECISE
LOGIN TO YOUR SMARTBIOBANK PROJECT
LOGIN
VISUALISE YOUR DATA ON ACSN/NAVICELL
LOGIN
LOGIN TO YOUR PRECISE CLOUD

Figure 1: SmartBioBank dashboard

1.3 Visualization of data stored in SmartBioBank

The ABT has also integrated a third link to the dashboard. The link is prepared to serve the users with a secure raw data storage. In the first project period, the ABT provided accessibility to a cloud service named NextCloud where the PrECISE members can directly store big volume of raw data from diverse types of omics experiments for further analyses. The mass stored data can be restructured and clinical relevant data will also be uploaded to the SmartBioBank.

Users from the PrECISE project will need to search data only on the https://smartbiobank.astridbio.com and under this dashboard, they can store their raw data into a secure NextCloud database and upload their clinical and omics data to the SmartBioBank. Moreover, they can visualize genomic and proteomic data on molecular maps provided by CI using ACSN/NaviCell.

Chapter 2 Methods for visualization of data

In this study, three maps have been used to visualize data:

- **ACSN** (Atlas of Cancer Signalling Network) [1]: is an interactive and comprehensive map encompassing multiple interconnected cancer-related signalling network maps containing a large number of reactions, proteins and other chemical species such as RNAs, ions or drugs This map is available online as a NaviCell resource at https://acsn.curie.fr/.
- **RECON2** [2]: is a comprehensive collection of metabolic pathways that contains also a large quantity of reactions (7.500) and metabolites (5.000) structured in 99 subsystems and compartmentalised into 8 cell's regions. This map is available online as a MINERVA resource at https://vmh.uni.lu/#reconmap and as a NaviCell resource at https://navicell.curie.fr/pages/maps_recon2.html.
- **Innate immunity response in cancer meta-map:** covers major signalling mechanisms occurring within and between different players of the innate immune component in TME. It represents both direct and indirect molecular interactions in innate immune cells, between immune cells and tumour cells. This map is available online at https://navicell.curie.fr/pages/maps_innateimmune.html.

To explore and visualize data on these maps, a web tool has been used: **NaviCell** [3]. This web tool for exploring large maps of molecular interactions was created using CellDesigner (http://celldesigner.org). The tool is characterized by a unique combination of three essential features: efficient map navigation based on Google maps engine, semantic zooming for viewing different levels of details on the map and an integrated blog for collecting community curation feedbacks.

To quantify the activity of our gene sets in groups and individual samples, we used **ROMA** [4], an algorithm allowing a fast and robust computation of the simplest linear model of gene regulation based on computing the first principal component of the expression data matrix and estimating the statistical significance of such approximation. We applied this method on RNAseq and proteomics data.

The data considered in this deliverable are made available from the consortium: PC39 data with mutation, RNAseq and proteomics data of 39 prostate cancer patients. For some of these 39 patients, punches from three different areas of the prostate were extracted.

Chapter 3 Data visualization

Data visualisation and analysis were performed by organising the samples by Gleason score groups, their punched location in the prostate and by comparing individual samples.

It was noted that when plotting data directly on top of our maps, there was no difference between groups or between individual samples (Figure 2). The expression data between two samples seem to be very similar at the level of individual genes. Thus, it was decided to explore the activity of pathways rather than the expression of individual genes to search for differences between samples.

We applied ROMA on our data to actually see the differences between our samples (see Chapter 2 for presentation of the tool).

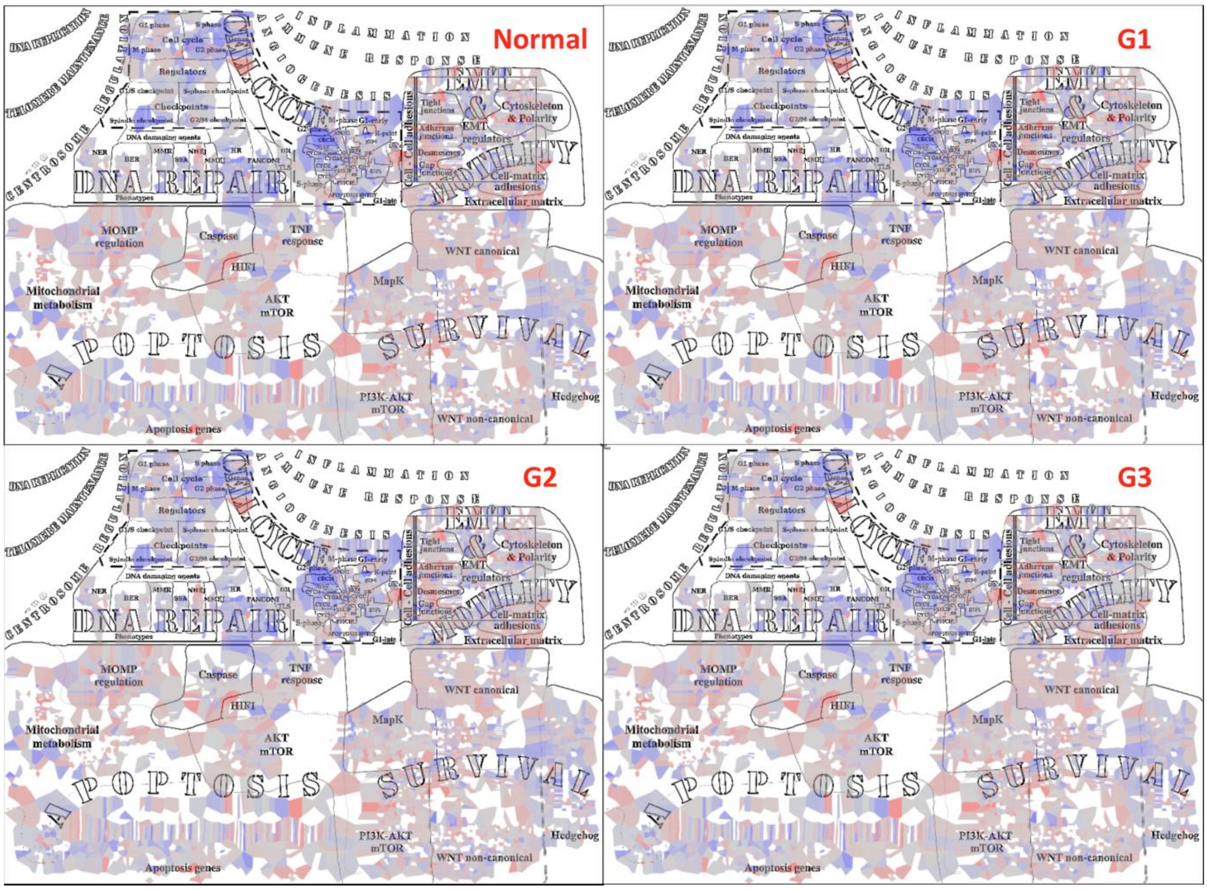


Figure 2: RNAseq data of Gleason groups (average value of samples belonging to the same Gleason group) visualized on ACSN through NaviCell. The colour gradient is chosen based on expression data with blue for lower expression level, red for higher expression level and grey for intermediate expression level.

In this chapter, we first present a tutorial on how to map data onto any network available in NaviCell map repository. We then show some results when considering clinical data: samples separated by Gleason score or samples divided into tumour areas. Finally, we show some examples of individual patients and discuss on heterogeneity of patients with the same clinical information.

Note that, in all sub-sections, the ROMA scores will be visualized on the maps, not the expression data for reasons we mentioned previously: there are no obvious differences at the level of gene or protein expression between samples.

3.1 Tutorial on how to visualize data with NaviCell

The atlas of cancer signalling networks (ACSN) is a pathway database recapitulating published facts about biochemical reactions involved in cancer. The pathways that are included in ACSN currently describe cell cycle, DNA repair, cell death (with autophagy, necroptosis, mitochondria metabolism, apoptosis, etc.), cell survival, EMT and cell motility. The navigation through ACSN uses NaviCell technology. NaviCell is a tool that allows the exploration of large maps using Google MapsTM engine and to map data onto these maps.

3.1.1 Step 1: Access ACSN website

Go to http://acsn.curie.fr

If you wish to navigate through the comprehensive map, click on “Atlas of Cancer Signalling Networks global map”, otherwise choose the individual maps.

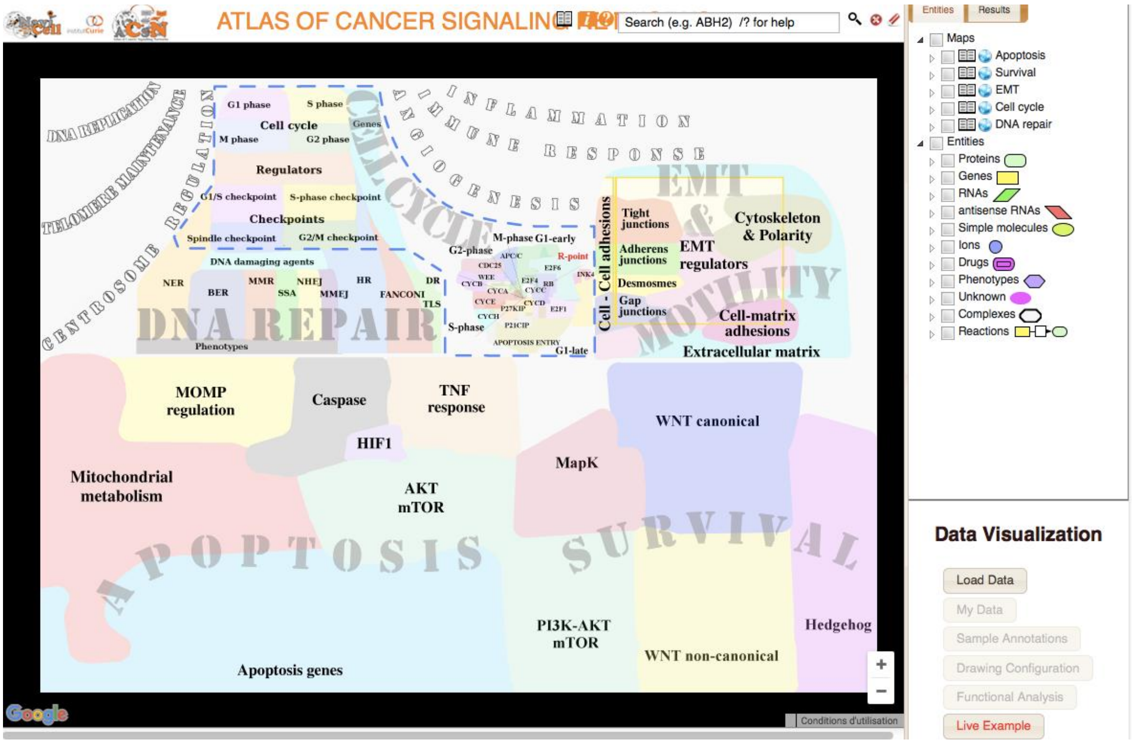


Figure 3:ACSN website. The atlas can be explored in the main window. The content of the map can be found in the upper right panel and the features for data visualization can be accessed through the lower right panel

For a step-by-step example, you can click on the Live Example in the Data Visualization panel.

3.1.2 Step 2: Load the data

To visualize the data onto the map, the genes/proteins need to be matched to their HUGO names. In the case of several “protein to gene” correspondence, we tend to choose the one with the most variation over the whole dataset. Moreover, for better visualization, we also compute the log10 values of the initial dataset.

The data to be visualized need to be prepared before.

1. If you want to visualize all samples one by one: upload the matrix of expressions with HUGO names.
2. If you want to visualize the mean expression per group (Gleason/areas of punches): for each protein, compute the average of all samples per group G1, G2, G3, TA1, TA2, Normal
3. If you want to visualize ROMA scores, you need to compute the scores using ROMA tool available at: https://github.com/Albluca/rRoma

In the Data Visualization panel, see Figure 4, choose “Load Data”:

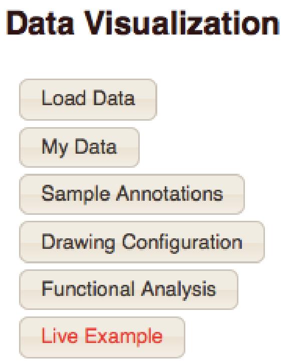
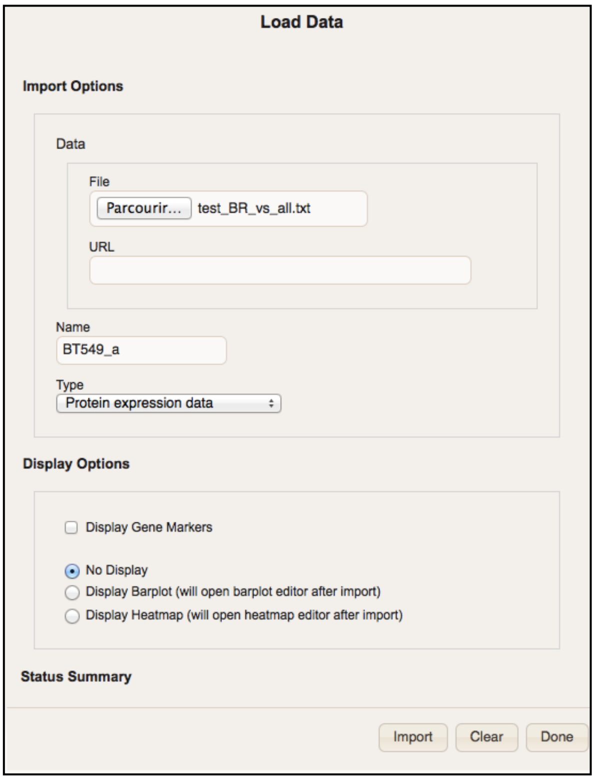


Figure 4: Data Visualization Panel

A window pops up, see Figure 5.

1. Upload your data: **RNAseq.txt.** Text files are used for datasets.
2. Name your matrix: **RNAseq**
3. Specify what type of data you are importing.
4. Here we select: **mRNA expression data**
5. If you do not wish to see which genes are found on the map, uncheck the box: **Display Gene Markers**. If you decide to check the gene markers, you will be able to erase them by clicking on this icon at the top right of the map panel: .
6. Click on **Import** then on **Done**

Do the same for Mutation data from the **Mutations.txt** matrix and selecting **Mutation data** in types. The mutation data have to be binary (0: no mutation, 1: mutation).



Load Data

Import Options

Data

File

Parcourir... test_BR_vs_all.txt

URL

Name

BT549_a

Type

Protein expression data

Display Options

□ Display Gene Markers

● No Display

○ Display Barplot (will open barplot editor after import)

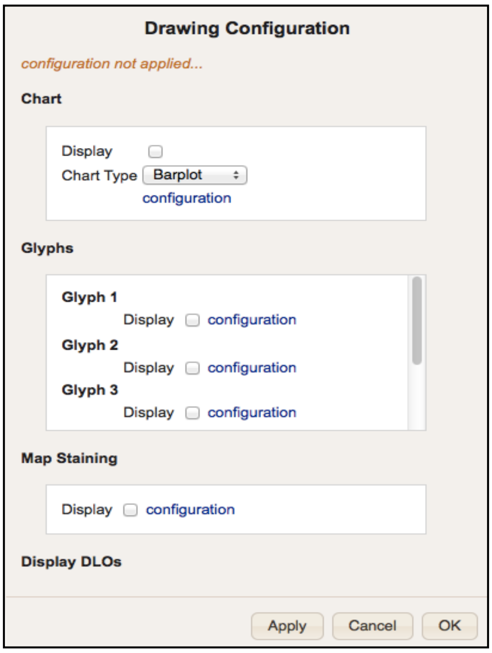
○ Display Heatmap (will open heatmap editor after import)

Status Summary

Figure 5: Data Upload

3.1.3 Step 3: Visualize the expression of H10N sample.

In the Data Visualization panel (lower right panel, see Figure 4), click on “Drawing Configuration”.



Drawing Configuration

configuration not applied...

Chart

Display □

Chart Type Barplot

configuration

Glyphs

Glyph 1

Display □ configuration

Glyph 2

Display □ configuration

Glyph 3

Display □ configuration

Map Staining

Display □ configuration

Display DLOs

Figure 6: Drawing Configuration

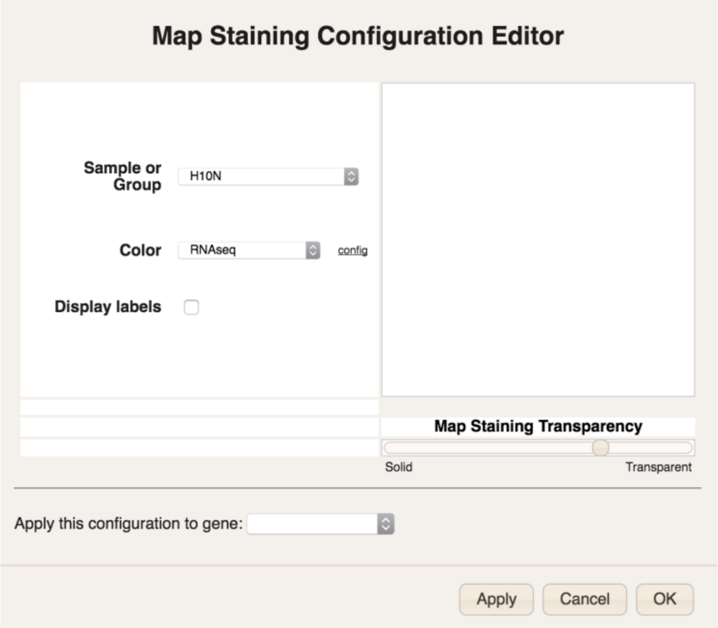
3.1.3.1 CASE 1 - Visualize individual samples

A window pops up. You can choose through several visualization options:

- Charts: “Heat maps” or “Bar plots”
- Glyphs with different shapes, forms and colours to visualize different types of information on the map
- Map staining to colour the territories around proteins according to their expression levels.

For our example, we choose to visualize the data with *Map Staining* style for expression data and *Glyphs plots* for mutation data.

Click on “Configuration” in the frame **“Map Staining”.**



Map Staining Configuration Editor

Sample or Group: H10N

Color: RNAseq

Display labels □

Apply this configuration to gene:

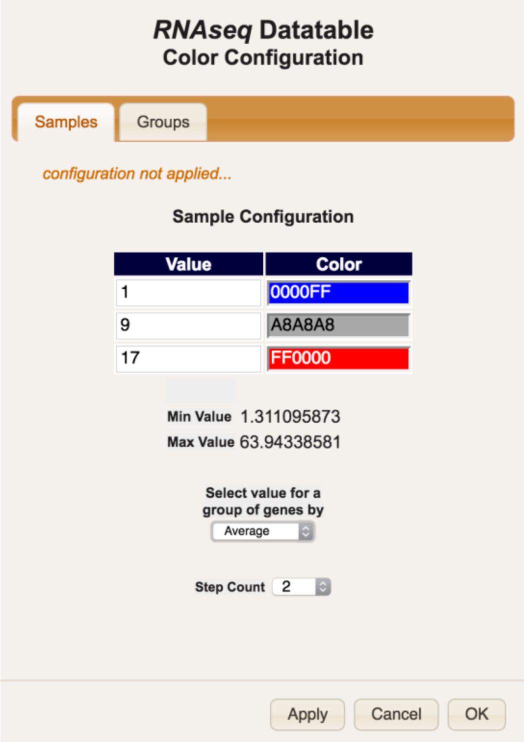
Figure 7: Map Staining Configuration Editor

In the Map Staining configuration editor, see Figure 7, we need to:

- Select a sample: **H10N** (which corresponds to the name found in the imported table)
- Select the Data table from which we select the values for the staining: **RNAseq** (corresponds to the name of the imported table we gave)
- Click on **config** to set parameters for the staining.
-

In Map Staining configuration, we can choose values and colours for the gradient, see Figure 8.

- Click on **Value** and put the needed limit
- Click on **Colour** to choose a colour from a box
- Click on “**OK**” when done to close the configurations
- Click on “**OK**” to apply the Map Staining and close the map Staining Configuration Editor window



***RNAseq* Datatable**
Color Configuration

configuration not applied...

Sample Configuration

Value	Color
1	0000FF
9	A8A8A8
17	FF0000

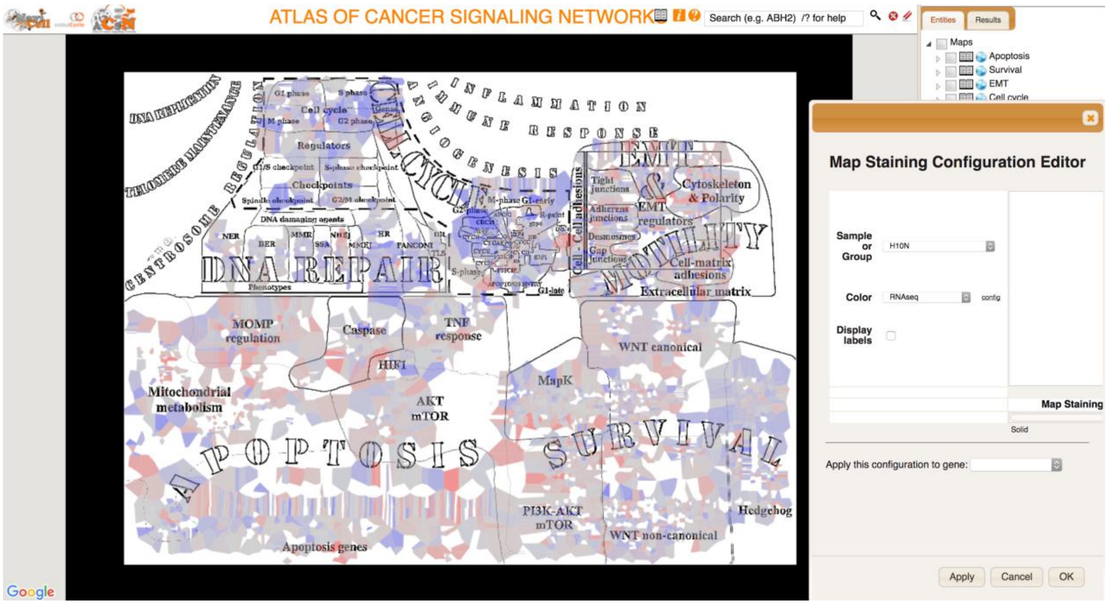
Min Value 1.311095873
Max Value 63.94338581

Select value for a group of genes by Average

Step Count 2

Figure 8: Colour Configuration

The colours on the map correspond to the gene expression for a particular patient.



Map Staining Configuration Editor

Sample or Group H10N

Color RNAseq config

Display labels □

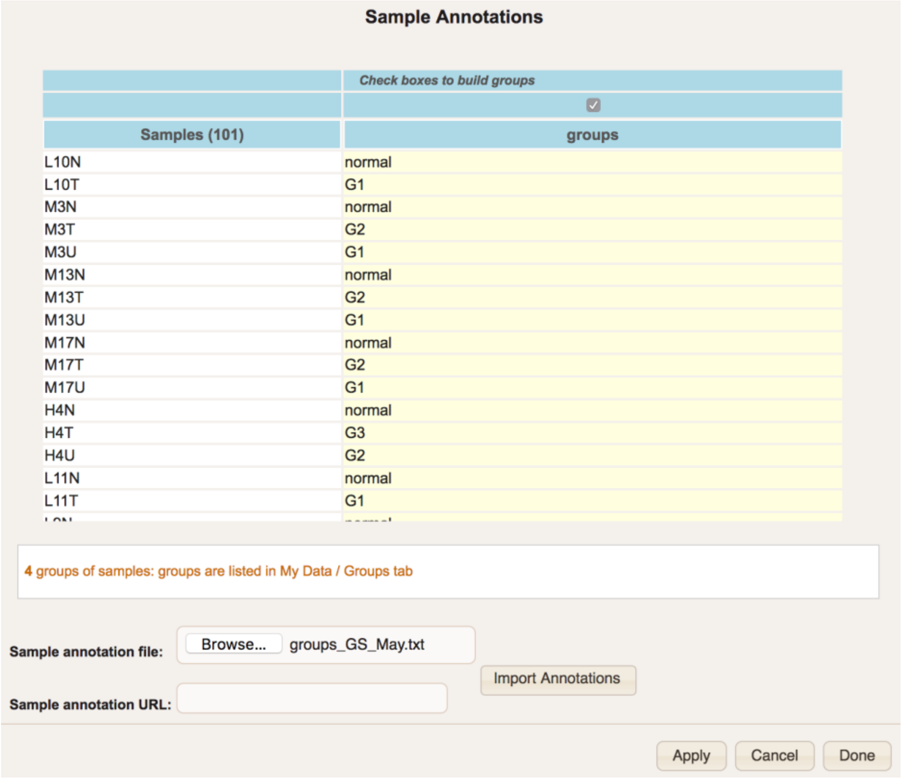
Map Staining

Solid

Apply this configuration to gene:

Figure 9: Visualization of RNAseq data on ACSN for the sample H10N through Map Staining

3.1.3.2 CASE 2 - Visualize groups of samples



Sample Annotations

	Check boxes to build groups
	☑
Samples (101)	groups
L10N	normal
L10T	G1
M3N	normal
M3T	G2
M3U	G1
M13N	normal
M13T	G2
M13U	G1
M17N	normal
M17T	G2
M17U	G1
H4N	normal
H4T	G3
H4U	G2
L11N	normal
L11T	G1

4 groups of samples: groups are listed in My Data / Groups tab

Sample annotation file: Browse... groups_GS_May.txt

Sample annotation URL:

Figure 10: Sample Annotations

On the ACSN main window, select “**Sample Annotations**”. A window will open, see Figure 10:

- “Browse” the annotation file **groups.txt** (contains samples in first column and the group in the second column or any clinical data available per sample)
- Click on **Import Annotations** to load the file
- Choose the groups you want to annotate your samples with by checking the boxes on top of columns
- Click on “**Apply**” and “**Done**” to confirm the annotation procedure

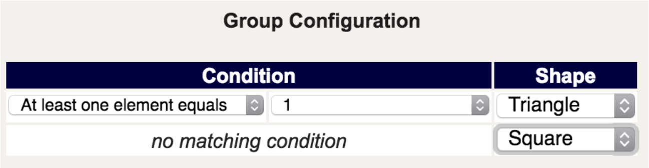
You can then proceed with the same steps to apply the Map Staining:

- Click on “**Drawing Configuration**”
- Click on “**configuration**” in Map Staining
- Select your group: **G2**.
- Configure your Map Staining values and colours in the Group tab
- Apply the Map staining for the group

To add glyphs for adding mutations on the map:

- Click on “**Drawing Configuration**”
- Click on “**configuration**” in the **Glyph 1** section
- Select your group: **G2**
- Select the data table you are taking for your glyph’s Shape, Colour and Size: **Mut** (name given when importing the matrix)
- Change the **Size on Map** to Large

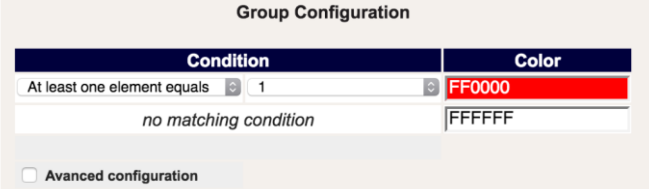
- In **Shape config**, set the condition “at least one element equals 1” as Triangle.



Group Configuration

Condition		Shape
At least one element equals	1	Triangle
no matching condition		Square

- In **Colour config**, set the condition “at least one element equals 1” as RED

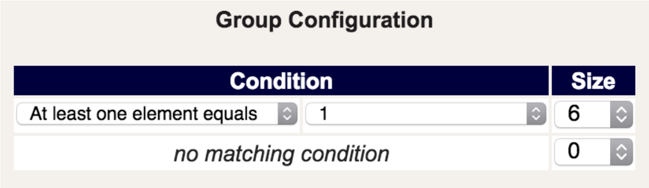


Group Configuration

Condition		Color
At least one element equals	1	FF0000
no matching condition		FFFFFF

□ Avanced configuration

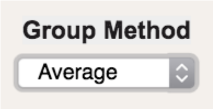
- In **Size config**, set the condition “at least one element equals 1” as 6 and the “no matching condition” as 0



Group Configuration

Condition		Size
At least one element equals	1	6
no matching condition		0

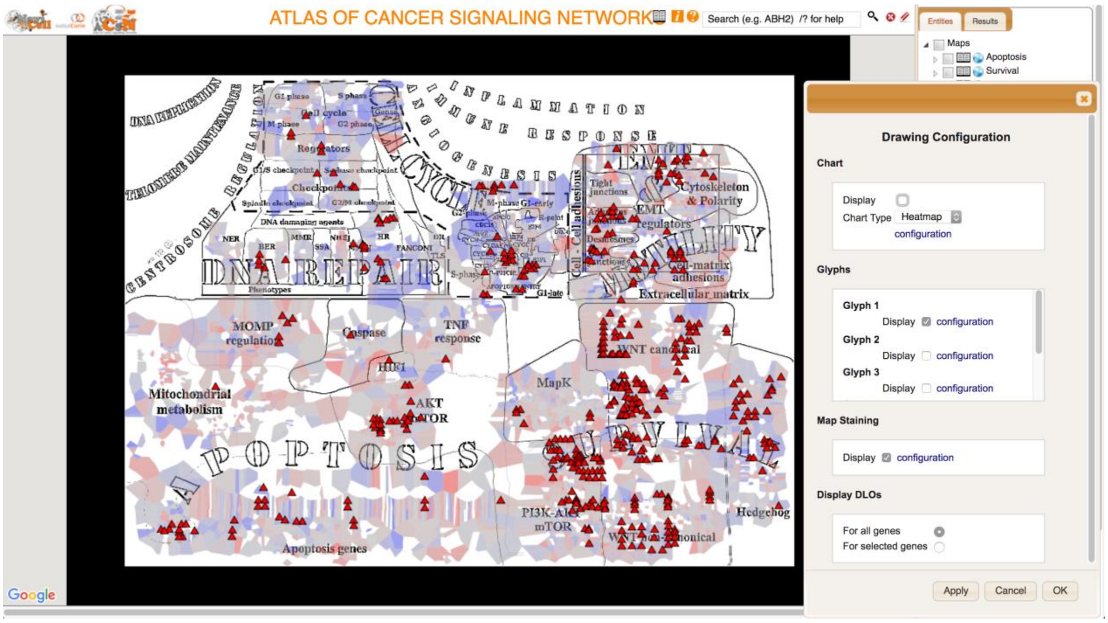
- You can change the “**Group Method**” to choose how the score will be calculated for a group.



Group Method

Average

- Click on “OK” to apply your settings



Drawing Configuration

Chart

Display □

Chart Type Heatmap

configuration

Glyphs

Glyph 1

Display ☑ configuration

Glyph 2

Display □ configuration

Glyph 3

Display □ configuration

Map Staining

Display ☑ configuration

Display DLOs

For all genes ●

For selected genes ○

Apply Cancel OK

Figure 11: Visualization of RNAseq data on ACSN for the group G2 through Map Staining. Mutation data was added in the form of Glyphs (Red Triangle).

3.2 Visualization of clinical data: Gleason scores

Some samples are considered normal (part of the biopsy with histopathological features close to the one observed in healthy tissues). For cancerous ones, three groups were set as follows:

- G1: corresponds to a Gleason score of 3+3
- G2: corresponds to a Gleason score of 3+4 and 4+3
- G3: corresponds to a Gleason score above 8

All data presented in this section correspond to the results of the ROMA analysis.

Note than when mapping results of ROMA, the scores can be visible through a Map Staining option using a colour gradient (blue being less active, red being more active and grey with intermediate activity). Through ROMA scores, we can assess the activity of certain signalling pathways (modules) based on the weighted activity of the genes that compose the gene sets.

On top of RNA and protein data, glyphs, in the form of a red triangle, were added to show the mutated genes.

3.2.1 *ACSN*

We use the whole atlas which integrates several signalling pathways known to be deregulated in cancer. It integrates several processes such as DNA repair, cell cycle, apoptosis, motility, survival, etc. Each process is composed of several modules (identified with different colours).

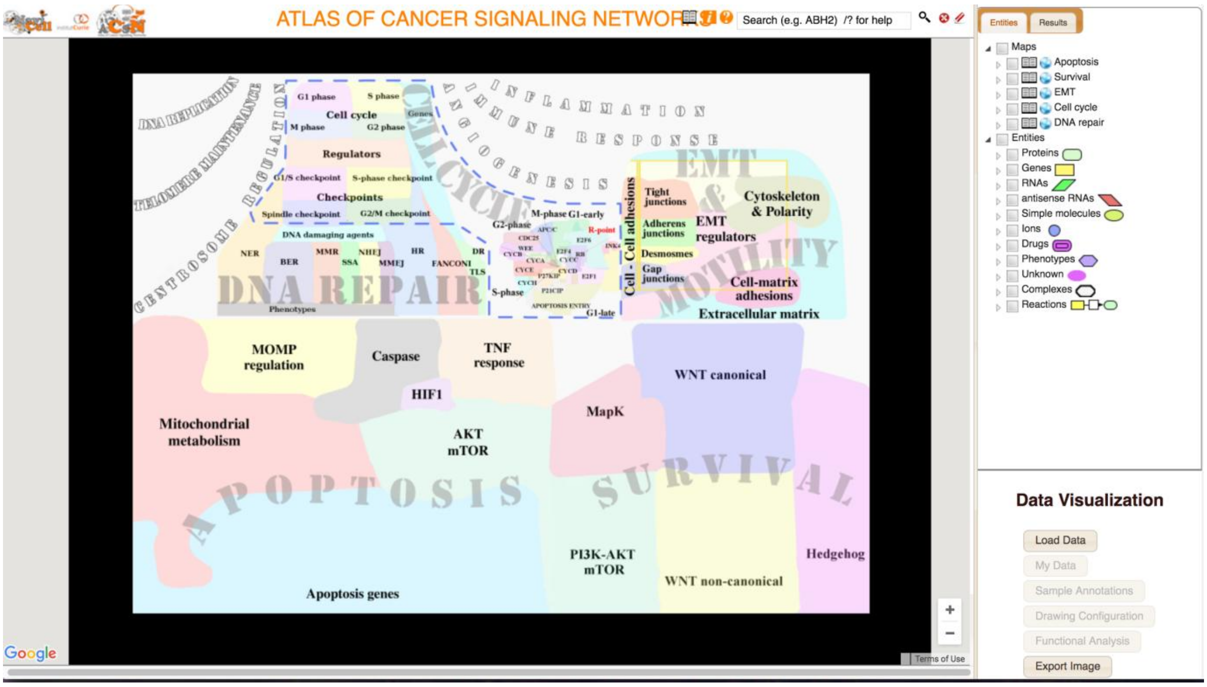


Figure 12: ACSN website. The scores of ROMA are computed based on the modules provided by the atlas of cancer signalling pathway. The territories are marked in different colours and correspond to the modules

We computed ROMA scores for both normal and tumour samples (Figure 13).

In normal cells, certain modules are less active than others (Figure 13, left panel). For instance, *DNA repair* and *Cell cycle* modules are both less active, while *Apoptosis*, *Survival* and *EMT&Motility* are all activated (could be understood as the fact that genes are transcribed and ready to be translated).

On the other hand, in tumour cells (G1, G2 or G3) (Figure 13, right panel), the behaviour is the same. *Apoptosis*, *Survival* and *EMT&Motility* are less active and *DNA repair* and *Cell cycle* pathways are active. This observation is indeed expected in cancer cells.

When looking at the mutations present in these tumour cells, most of them are located in the *Survival* and *EMT&Motility* modules, especially the *PI3K-AKT-mTOR*, *WNT canonical* and *WNT non-canonical* pathways are regrouping most of the mutations in tumour samples.

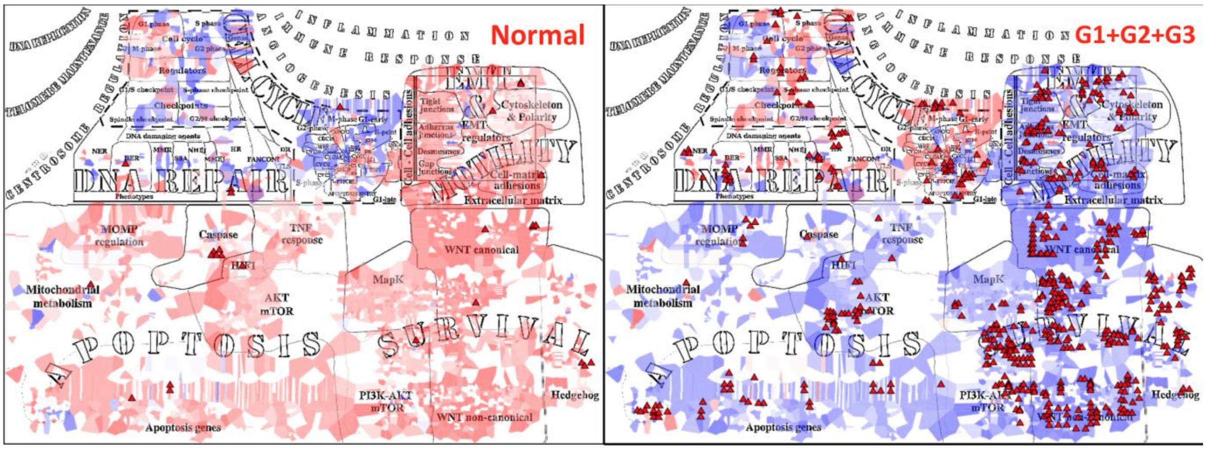


Figure 13: ROMA results of RNAseq data visualised on ACSN for Normal and Tumour samples (mean activity of each group). The colour gradient is chosen based on ROMA scores with blue as less active, red as more active and grey for values close to 0.

For proteomic data, Gleason groups show slight differences as well. However, when comparing to normal samples, we note something different from the RNA data. In normal samples, only *EMT&Motility* is more active. On the contrary, in tumour samples, *DNA Repair*, *Cell cycle*, *Apoptosis* and *Survival* are all active and *EMT&Motility* and some pathways of the *Survival* module (MapK and WNT canonical) are less active.

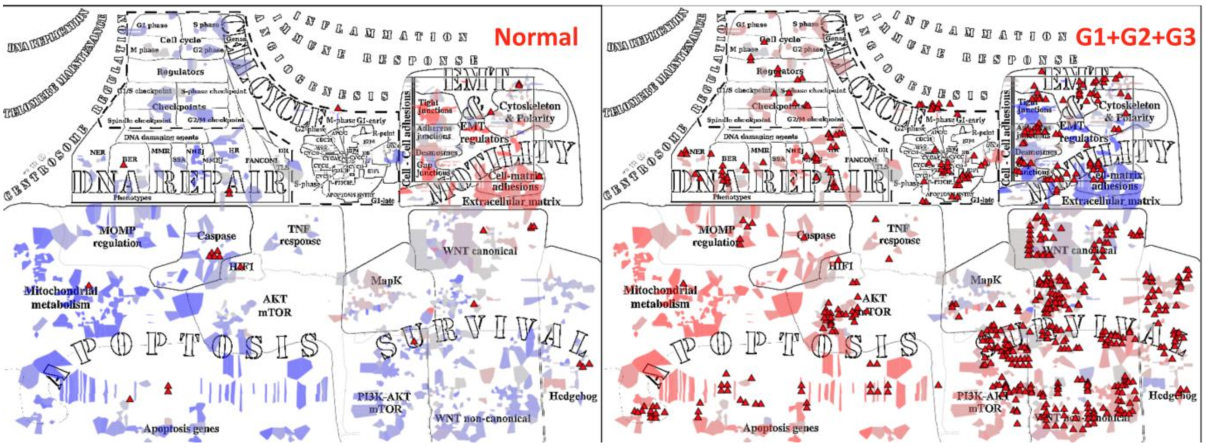


Figure 14: ROMA results of Proteomic data visualised on ACSN for Normal and Tumour samples (mean activity of each group). The colour gradient is chosen based on ROMA scores with blue as less active, red as more active and grey for values close to 0.

3.2.2 RECON2 map

On RECON2, we can observe what metabolic pathways are less or more active between our groups.

For RNAseq data (Figure 15), in normal cells, pathways that are less active are *Purine, Fructose* and *Manose Phosphatidylinositol* and *N-Glycan metabolism*.

What is more interesting is to observe modifications in tumour cells: *Purin and Pyrimidine metabolism* become much more active, as well as *Fructose and Manose metabolism*. The *Phosphatidylinositol, Folate* and some paths of the *Fatty Acid metabolism* are also more active while all other pathways are shown as less active. It is particularly evident that *Amino-acids transport* pathways are mostly active in cancer cells.

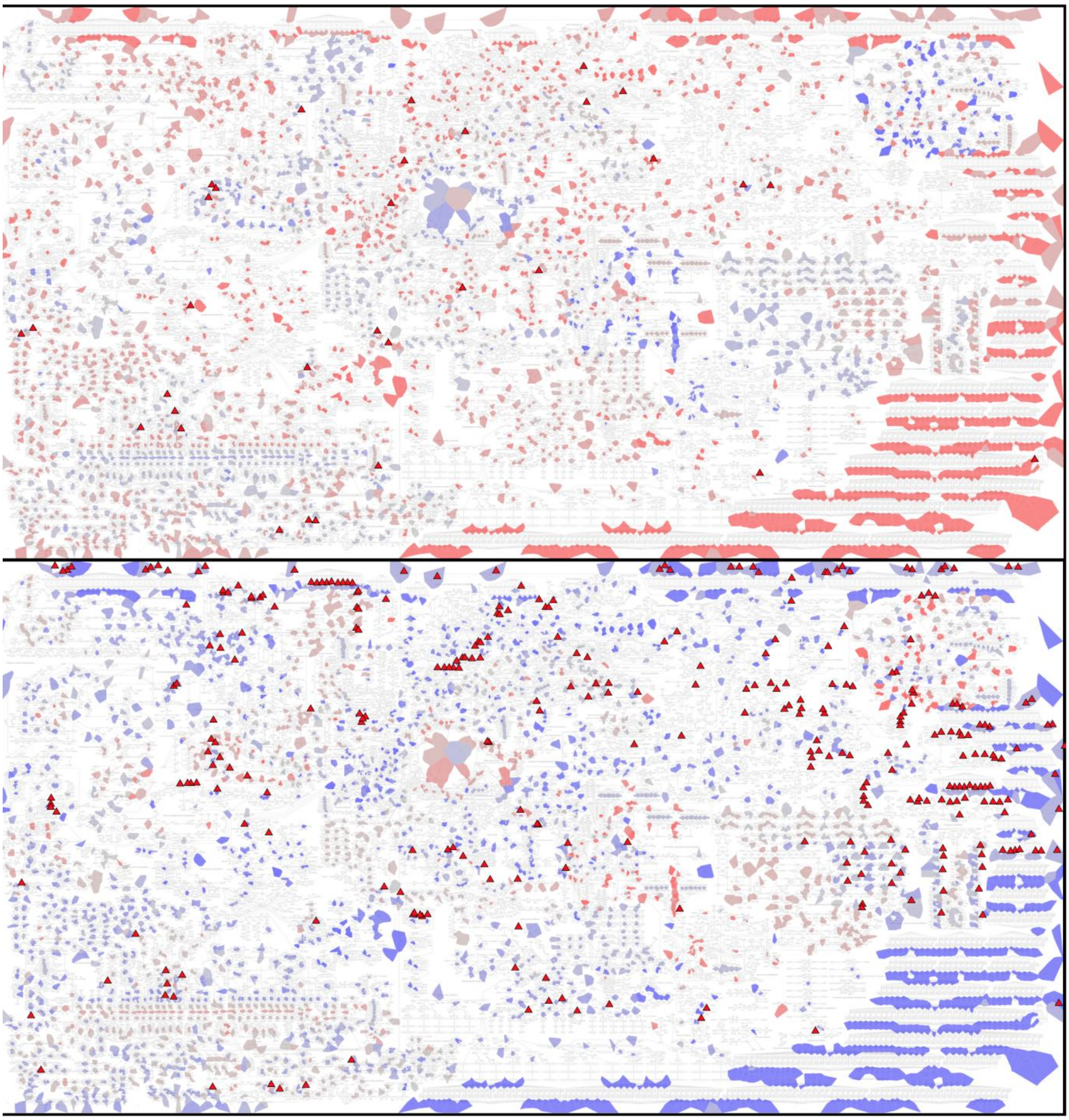


Figure 15: ROMA results of RNAseq data visualised on RECON2 for Normal (top) and Tumour (bottom) samples (mean activity of each group). The colour gradient is chosen based on ROMA scores with Blue as less active, Red as more active and Grey for values close to 0.

Looking at the proteomics data (Figure 16), it seems that all pathways are more active in cancer cells and less active in normal cells. Only the *Glycolysis* and *gluconeogenesis* pathways are less active in cancer cells in this case.

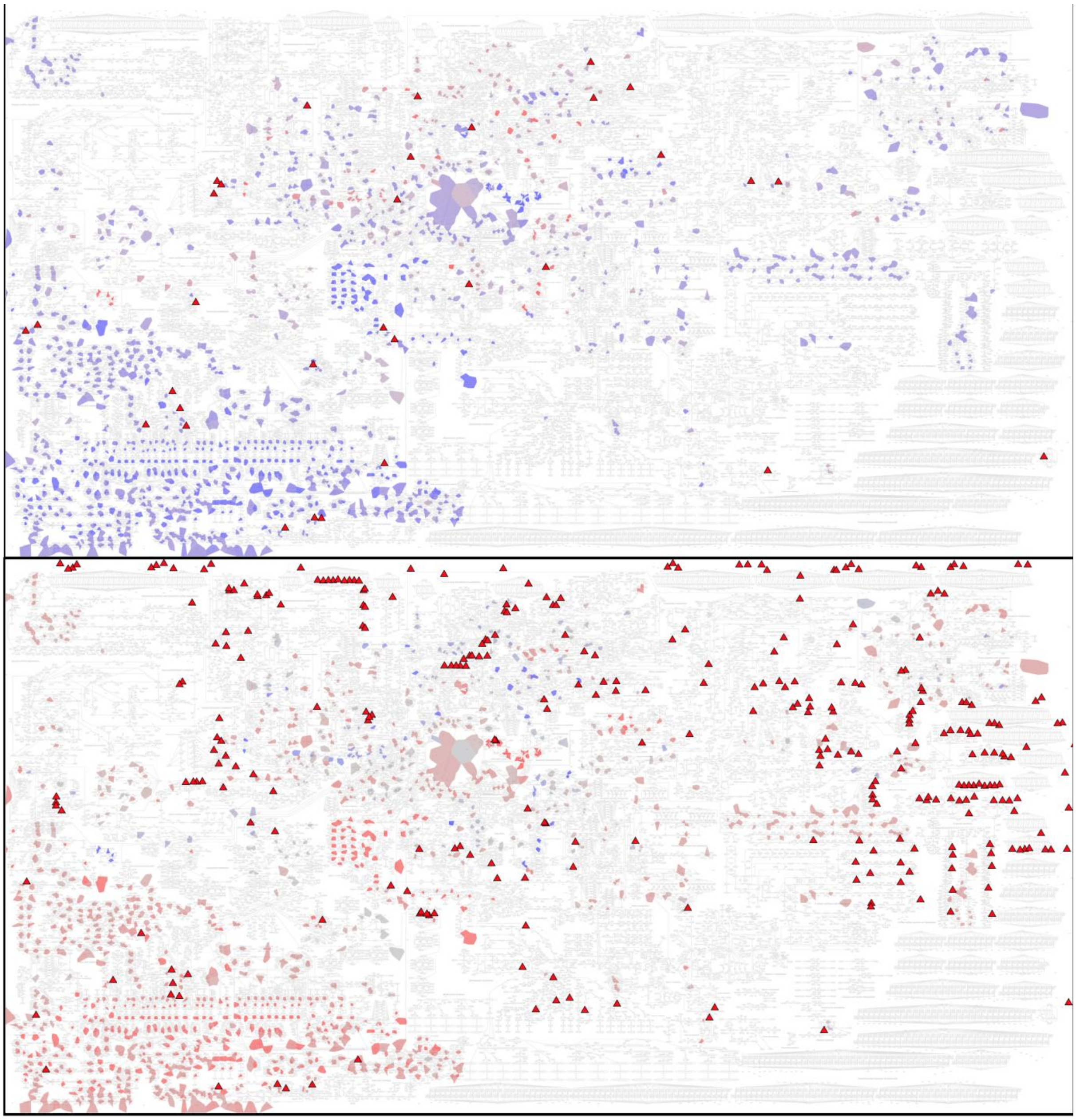


Figure 16: ROMA results of Proteomics data visualised on RECON2 for Normal (top) and Tumour (bottom) samples (mean activity of each group). The colour gradient is chosen based on ROMA scores with Blue as less active, Red as more active and Grey for values close to 0.

When mapping mutation data, we can see that it is in accordance with the RNAseq data: most of mutations are located on genes participating in *Amino acid transport* but also in *Tyrosine, Phenylalanine, Chondroitin, Starch, Sucrose* and *Bile Acid metabolisms*.

3.2.3 Innate Immunity map

In the case of the *Innate Immunity* map, most of the pathways are active for normal samples and less active in tumour samples. This shows that there is less interaction between prostate tumour cells and the tumour microenvironment (TME).

Most of the mutations in tumour cells are located in genes related to *Cytokine activities* but also with *NK cells*.

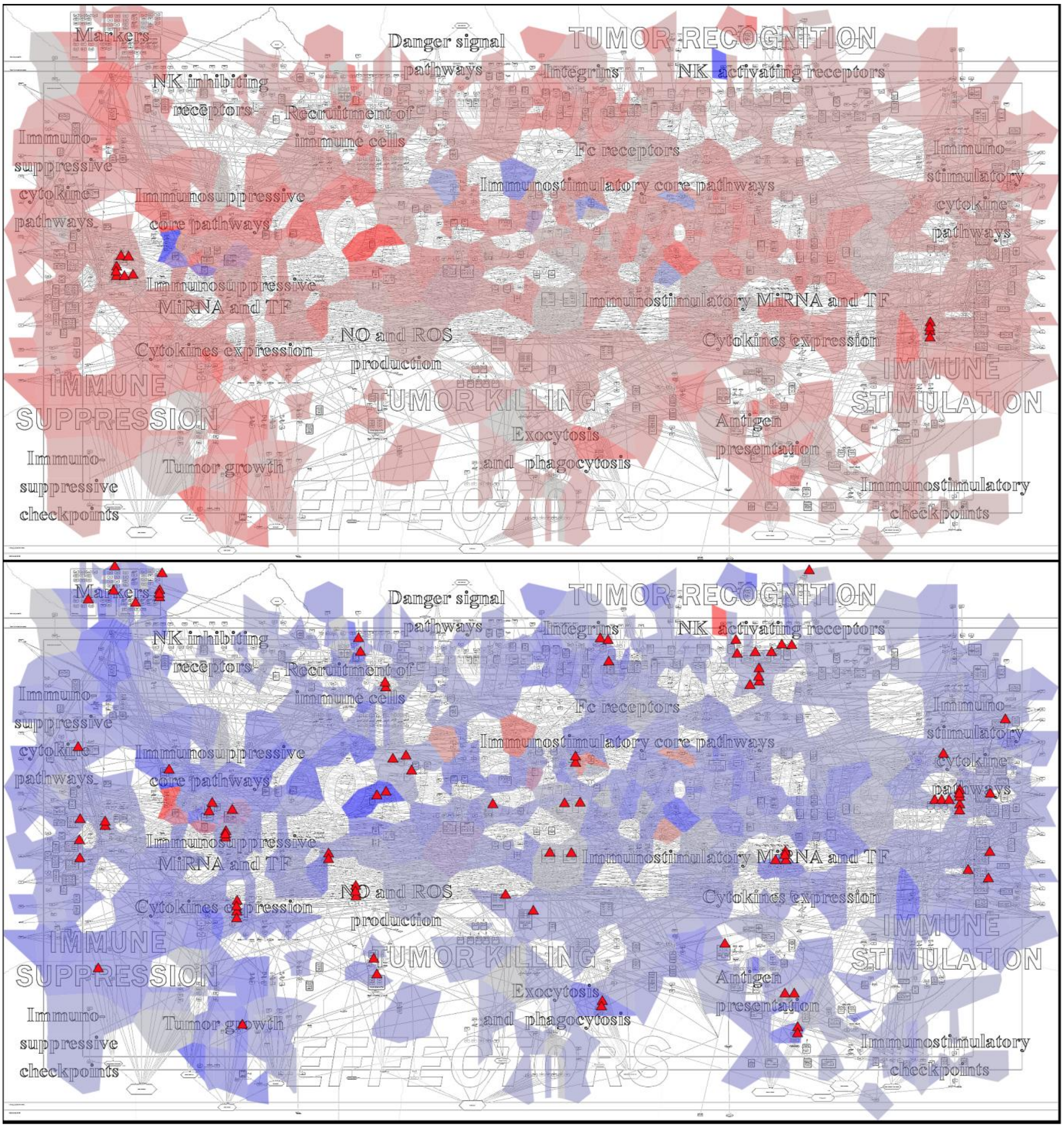

Markers
Danger signal pathways
TUMOR RECOGNITION
Integrins
NK activating receptors
NK inhibiting receptors
Recruitment of immune cells
Fc receptors
Immuno-suppressive cytokine pathways
Immuno-stimulatory cytokine pathways
Immunosuppressive core pathways
Immunostimulatory core pathways
Immunosuppressive MiRNA and TF
Immunostimulatory MiRNA and TF
NO and ROS production
Cytokines expression
Cytokines expression
IMMUNE SUPPRESSION
TUMOR KILLING
IMMUNE STIMULATION
Exocytosis and phagocytosis
Antigen presentation
Immuno-suppressive checkpoints
Tumor growth
Immunostimulatory checkpoints
EFFECTORS

Figure 17: ROMA results of RNAseq data visualised on the Innate Immunity map for Normal (top) and Tumour (bottom) samples (mean activity of each group). The colour gradient is chosen based on ROMA scores with blue as less active, red as more active and grey for values close to 0.

Unfortunately, too few proteins were overlapping between proteomic data and the map, thus not allowing a clear analysis of our results.

3.3 Visualization of clinical data: aggressiveness

Samples were classified in three categories based on their location to the tumour:

- Normal: sample coming from normal tissue
- TA1: most aggressive area of the tumour
- TA2: least aggressive area of the tumour

All data presented in this section correspond to the results of the ROMA analysis.

3.3.1 ACSN

Apart from the mutation data, there are not many differences between the origins of the samples (for both RNAseq data or proteomics). TA1 and TA2 groups show the same relative level of activity between them and Gleason groups (Figure 18).

Looking at the mutation data, disparities can be seen:

- Samples coming from TA1 present more mutations
- Mutations in TA1 are located more in the *Cell cycle*, *WNT non-canonical* and *PI3K-AKT mTOR* modules.

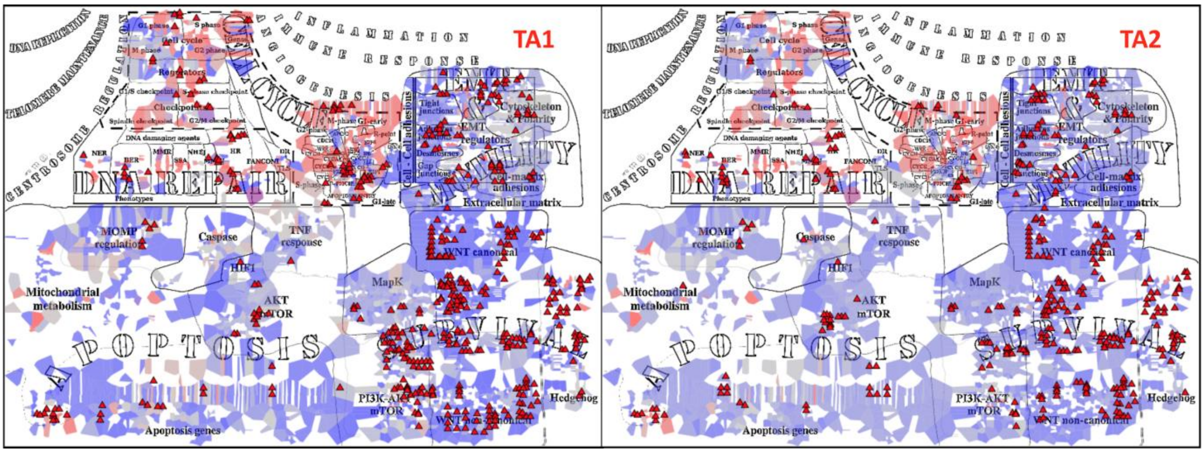


Figure 18: ROMA results of RNAseq data visualised on ACSN for differently located samples (mean activity of each group). The colour gradient is chosen based on ROMA scores with Blue as less active, Red as more active and Grey for values close to 0.

3.3.2 RECON2 map

For metabolic pathways, location of the samples seems to have little impact on the average activity (Figure 19). The same observations are made for both location groups as with Gleason groups.

Some differences could be noted for mutations:

- Mutations in TA1 are present mainly in genes related to Amino acid transport, Tyrosine, Phenylalanine metabolism and Bile acid synthesis.
- Mutations in TA2 are present mainly in gene related to Purine and Keratan sulfate metabolism.

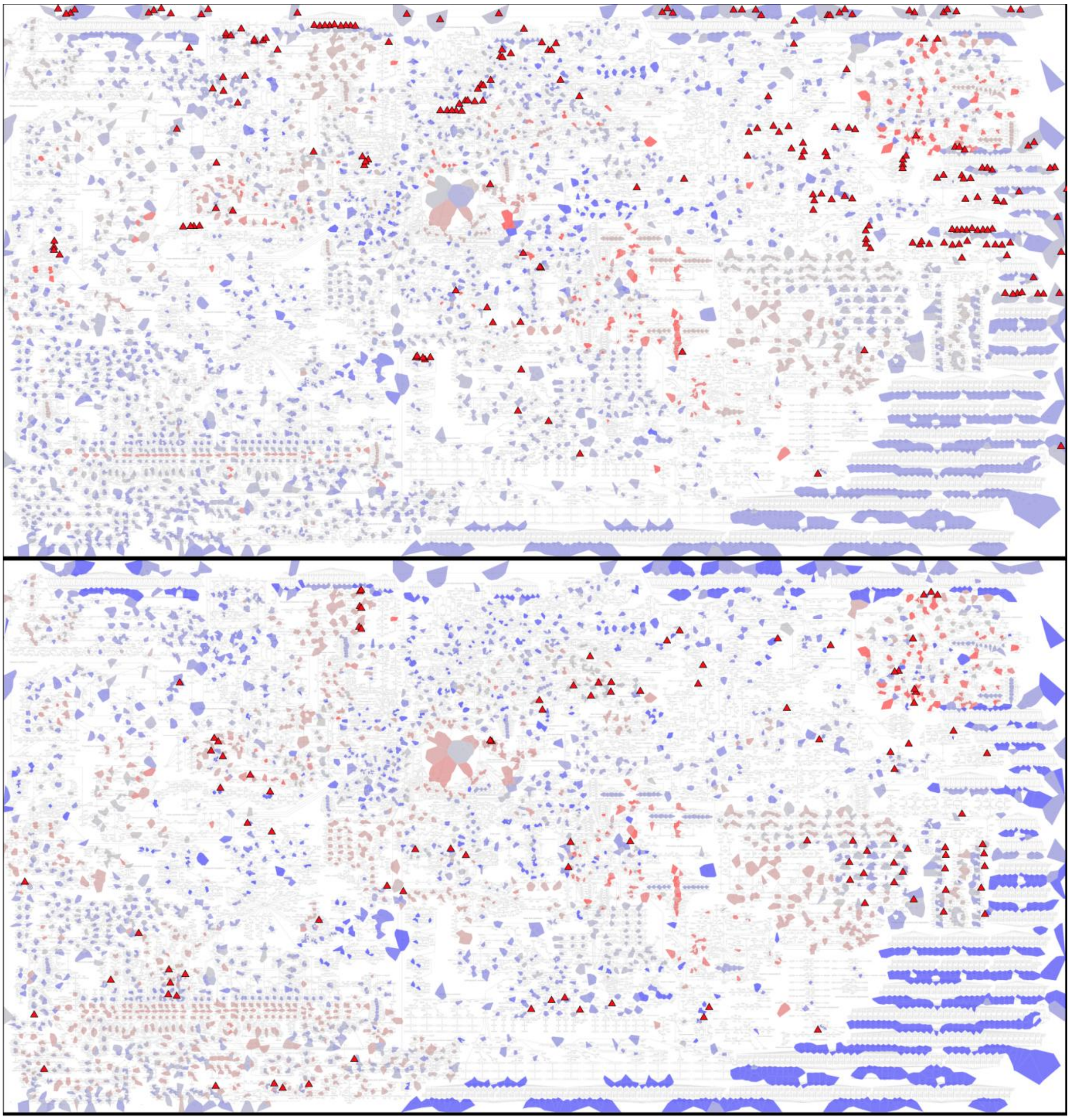


Figure 19: ROMA results of RNAseq data visualised on RECON2 for TA1 (top) and TA2 (bottom) samples (mean activity of each group). The colour gradient is chosen based on ROMA scores with Blue as less active, Red as more active and Grey for values close to 0.

3.3.3 Innate Immunity map

From the ROMA analysis of RNAseq data, we find differences between the TA1 and TA2 groups.

Samples coming from the TA1 region are more heterogeneous and show that some pathways are more active than others, like NK *activating receptors, Danger signal pathway* and *Immuno-stimulatory cytokine pathway.* However, in the TA2 regions, all pathways have a low level of activity.

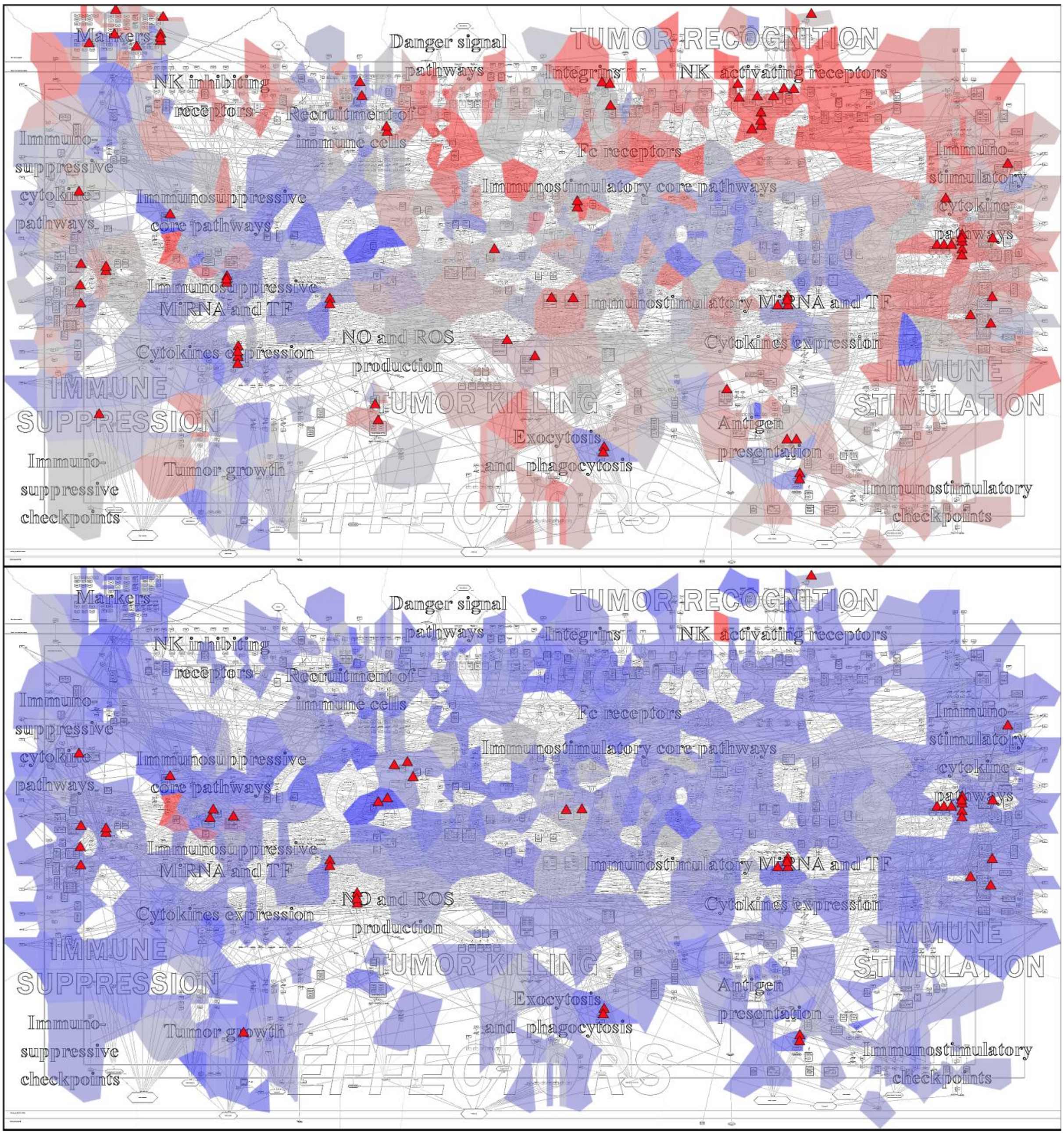

Markers
Danger signal pathways
TUMOR RECOGNITION
NK inhibiting receptors
Integrins
NK activating receptors
Recruitment of immune cells
Immuno-suppressive cytokine pathways
Fc receptors
Immuno-stimulatory cytokine pathways
Immunosuppressive core pathways
Immunostimulatory core pathways
Immunosuppressive MiRNA and TF
Immunostimulatory MiRNA and TF
Cytokines expression
NO and ROS production
Cytokines expression
IMMUNE SUPPRESSION
TUMOR KILLING
IMMUNE STIMULATION
Antigen presentation
Exocytosis and phagocytosis
Immuno-suppressive checkpoints
Tumor growth
Immunostimulatory checkpoints
EFFECTORS

Figure 20: ROMA results of RNAseq data visualised on the Innate Immunity map for TA1 (top) and TA2 (bottom) samples (mean activity of each group). The colour gradient is chosen based on ROMA scores with Blue as less active, Red as more active and Grey for values close to 0.

The most TA1 active pathways are part of 2 modules: Tumour Recognition and Immunity Stimulation.

3.4 Comparison between individual samples

Multiple samples have been analysed separately. We focused on some samples that exhibit very distinct number of mutations:

- Mutations: M13, M3, H2, L10
- No Mutations: M12, M10, H10, L3

All data presented in this section correspond to the results of the ROMA analysis.

3.4.1 ACSN

Using RNAseq data, it is hard to find common elements between samples from the same group (normal or tumour samples, for instance), as each sample exhibits a specific behaviour. No real difference has been noted in results between samples containing many mutations and those not containing mutations.

When looking at punches from several patients (L3, M12, M3, H2) we notice overall a lower activity of Apoptosis pathways in normal samples (Figure 21) whereas this module is found active when considering the whole normal group (Figure 13, left panel) (note that for the pathway score of the group, we compute the mean pathway score of patients composing the group). These differences are a sign of patient heterogeneity and this heterogeneity seems to be smoothed when taking the mean value of all normal samples together.

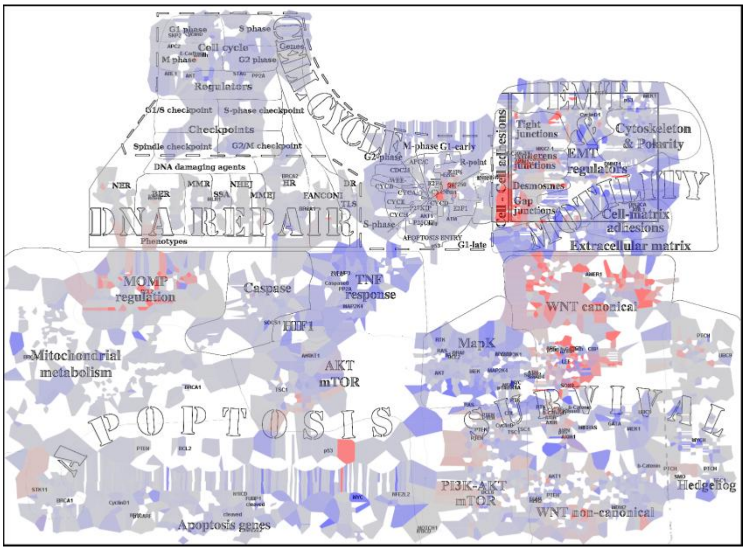


Figure 21: ROMA Scores of RNAseq data for M3 normal sample.

In all individual samples, tumour samples tend to have a diminished activity in modules such as *Survival* and *EMT&Motility* when compared to normal samples. However, tumour samples from H2, L10 and M12 patients remain active.

Similar observations can be made on proteomics data. When mapping proteomics data of individual patients, one can observe high heterogeneity: some tumour samples have pathways active in some patients such as *DNA repair*, *Apoptosis* and *Survival*, while these pathways are inactive in other tumour samples of patients (not shown).

3.4.2 RECON2 map

For metabolic pathways, samples show results closer to previously discussed observations. The most altered pathways are *Amino acids transport*, *Fatty acid synthesis* and *oxidation, Pyruvate and Biotin metabolisms*.

Some samples have peculiar results not concordant with previous observations:

- H10, L3, L10 and M12: Amino acid transport and Fatty acid synthesis pathways stay active in tumour samples
- L10 and M12: Fatty acid synthesis pathways stay active in tumour samples
- H10, M12: Pyruvate and Fatty acid metabolism pathways are significantly less active in normal samples

Proteomic data of each sample are, in this case, concordant with observations: with the majority of pathways being less active in normal cells and more active in tumour cells (not shown). Only two samples (H10, M12) have all pathways less active in both normal and tumour cells. Both samples are part of the most mutated groups across the samples.

3.4.3 Innate Immunity map

Unfortunately, taking samples individually yields to too tessellated results to be easily discernible and analysed.

3.5 Another representation of pathway activity scores

ROMA allows some statistical analyses of pathways/modules.

As previously mentioned, with ROMA, we try to identify which pathways have the highest variance across all the samples. For that we use a method based on PCA. Each pathway is a list a protein. For each pathway, we compute a score that corresponds to a weighted sum of the protein expression where weights are based on PCA [4].

These analyses were done only on ACSN and Metabolic pathways. A clustering algorithm was applied on the pathway scores for an easier visualisation of the general results.

3.5.1 Signalling pathways

We visualize the results of ROMA analyses as heatmaps. The pathways used as gene sets correspond to the modules defined in ACSN and RECON2. With this type of visualization, normal and tumoral samples separate well, but it is harder to see differences between the Gleason groups G1 and G3. We can also see that G2 samples are very heterogeneous (Figure 22).

We can observe that *Cell-cycle* pathways are less active in normal samples and *EMT&Motility* pathways are less active in cancer samples. It is less clear about *Innate Immunity* pathways due to the heterogeneity of samples, especially in G2, but the general tendency is that they are less active in tumour samples.

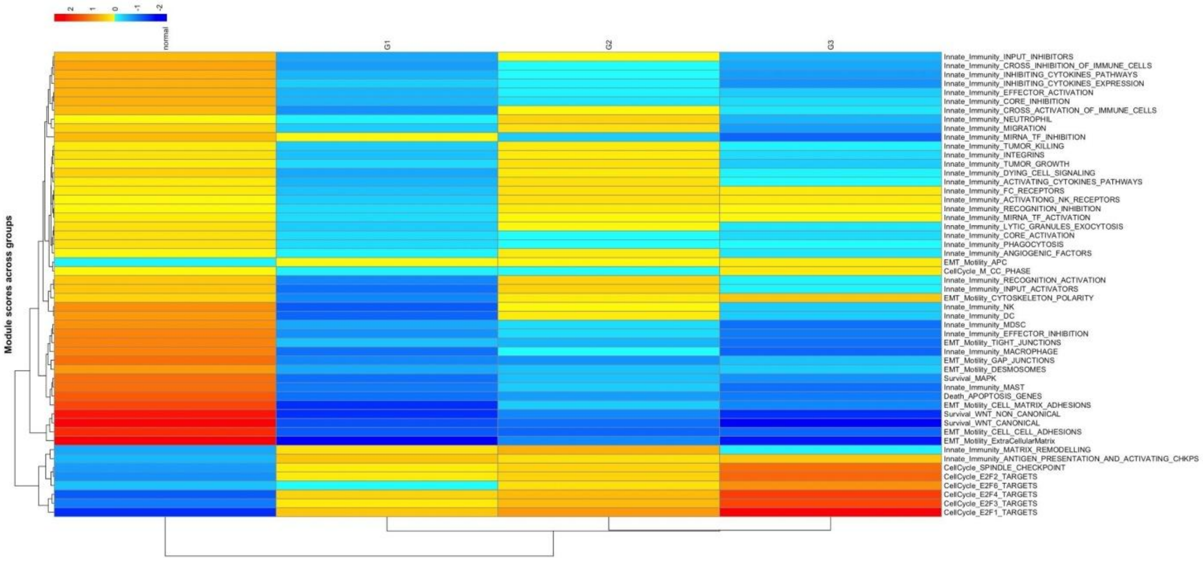


Figure 22: Heatmap of average Signalling Module activity (ROMA scores) of different Gleason Groups from RNAseq data.

Interestingly, TA1 samples present module scores with intermediate values between normal and TA2 samples. Differences of pathways such as *EMT&Motility* and *Innate Immunity* are more evident between normal and TA2 than between normal and TA1 (Figure 23). These observations are counter-intuitive and would require more in-depth analyses with the help of clinicians.

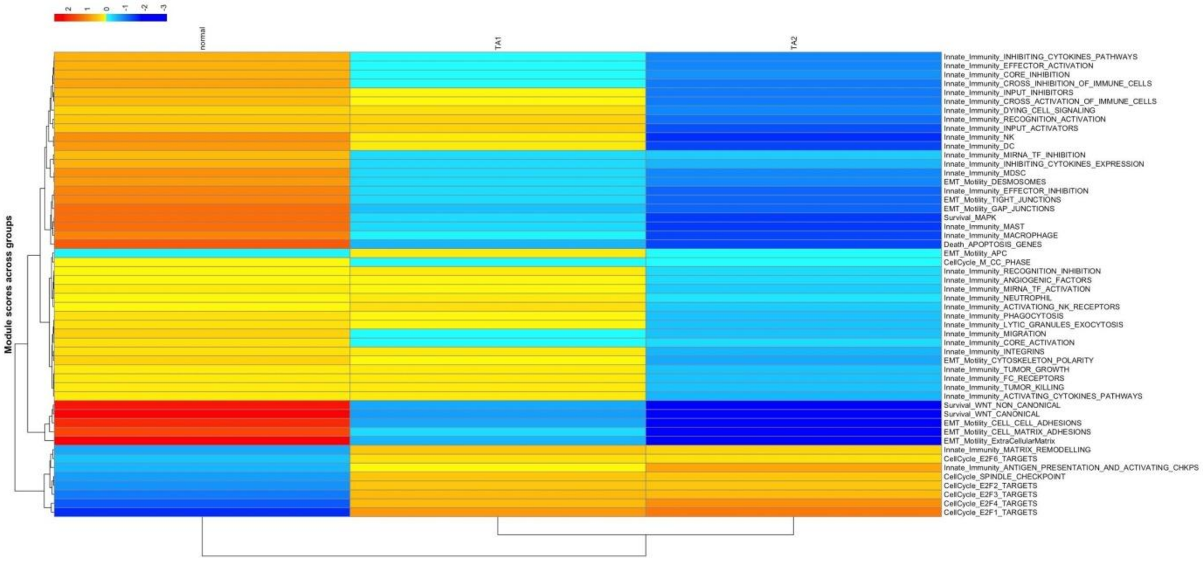


Figure 23: Heatmap of average Signalling Module activity of different biopsies location groups from RNAseq data.

When applying a clustering algorithm on samples, 3/4 clusters appear clearly but their constitution is less clear. All these clusters contain a mix of normal and tumour samples. The separation between normal and tumoural is not obvious when looking at individual samples but becomes more evident when considering groups of patients. For this case, we average the score across patients of the same group (Figure 24).

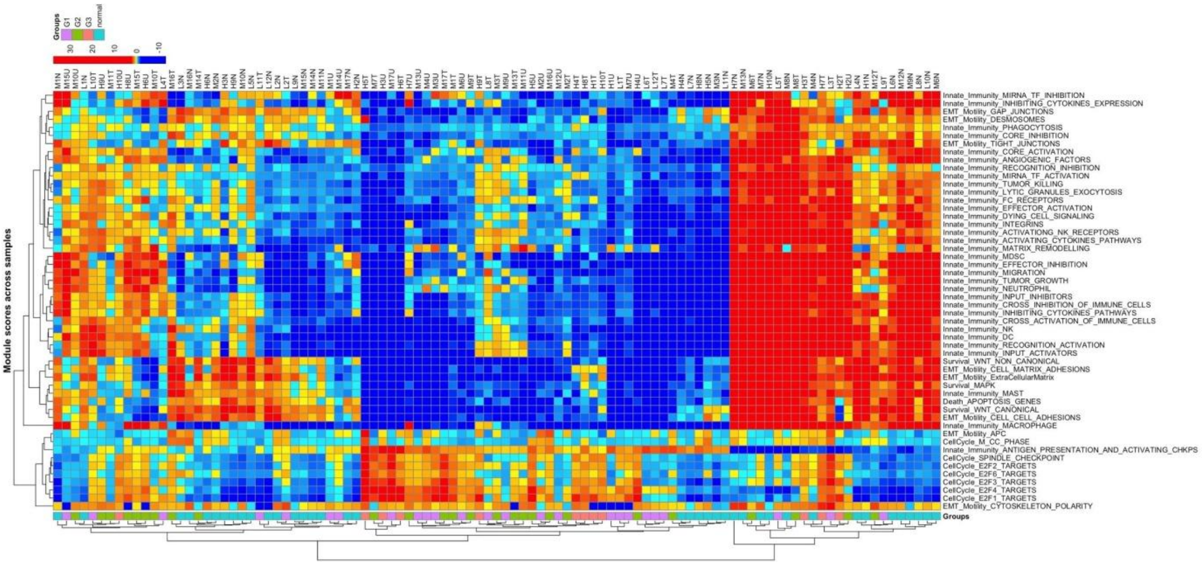


Figure 24: Heatmap of average Signalling Module activity of different samples from RNAseq data.

Looking at the correlations between module scores, two clusters clearly stand out and are anti-correlated: *Cell-cycle* pathways, belonging to the first cluster and *Innate Immunity* plus *EMT&Motility* belonging of the second cluster (Figure 25). *Innate Immunity* plus *EMT&Motility* seem to be well correlated.

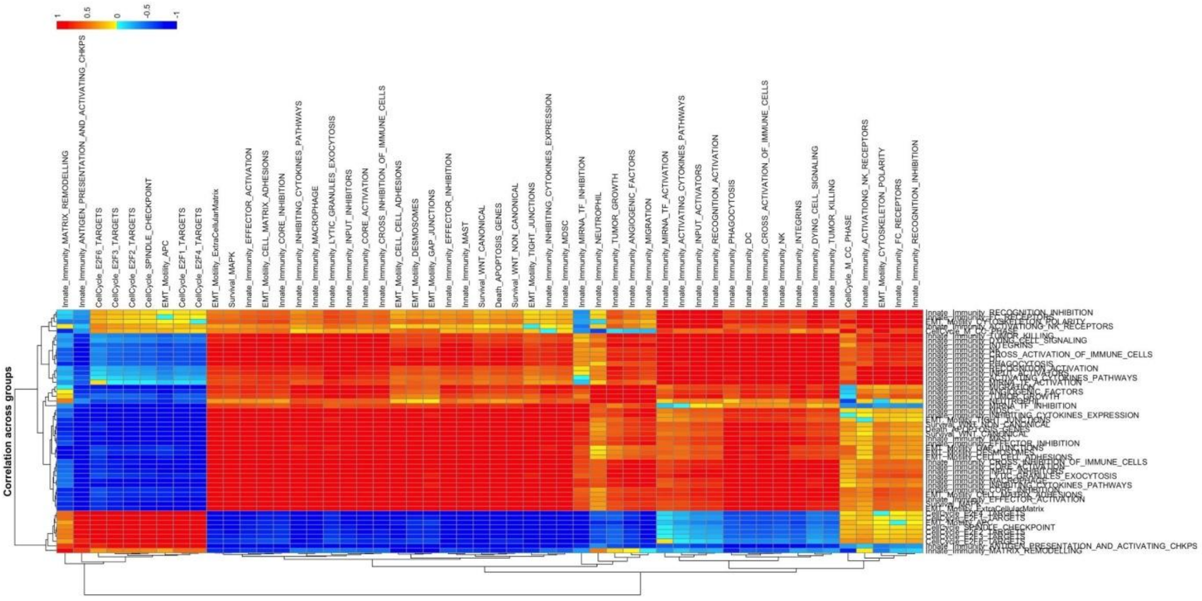


Figure 25: Correlations of average Signalling Module activity from RNAseq data.

3.5.2 Metabolic pathways

For the ROMA analyses of Metabolic pathways, RECON2, normal and tumoural samples separate well.

The correlation between pathways highlight three different clusters (Figure 26): the first one containing Exchange-demand reactions, Oxidative phosphorylation, Citric Acid cycle and Vitamin A metabolism; the second one containing the Fatty acid oxidation, Glycine-Serine-Alanine-Threonine-Valine-Leucine-Isoleucine and Tetrahydrobiopterin metabolisms; and the third one containing the rest of the pathways.

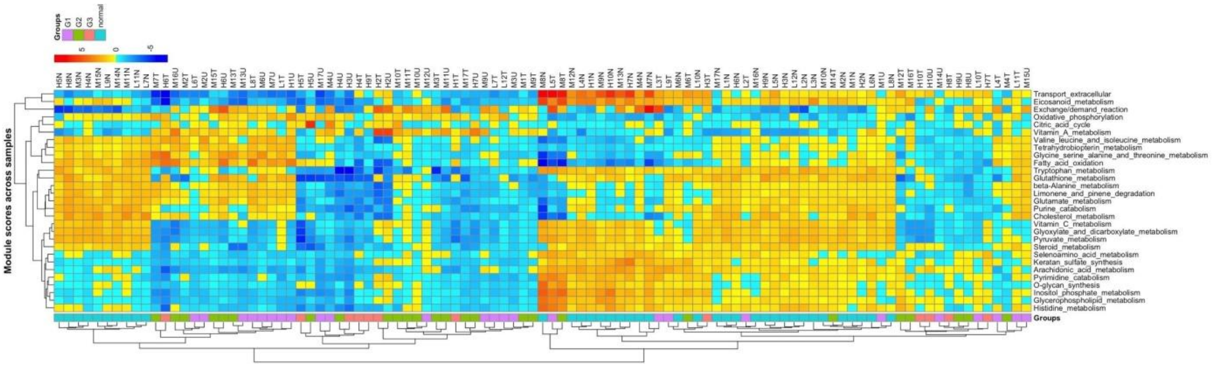


Figure 26: Heatmap of average Metabolic Modules activity (ROMA scores) of different samples from RNAseq data.

Chapter 4 Summary and Conclusion

Prostate cancer data (RNAseq, Proteomics, Mutation) from 39 patients have been visualized on top of three different networks. This allowed us to explore three processes such as Signalling, Metabolic and Innate Immunity pathways.

Mapping the expression data of individual patients did not permit to observe significant differences between samples or understand easily the possible mechanisms implicated. This is why the ROMA tool was used to extract a score indicating the level of relative activity of each of the pathways considered.

With success, several pathways have been noted as altered in prostate cancer. By simply visualizing the data, we found the level of activity of certain pathways different in cancer cells, such as the WTN and MAPK pathways, Amino Acids transports or Purin and Pyrimidine metabolism.

By grouping samples by Gleason score and by the area from which samples were extracted, we could see pathways changing gradually from normal to tumoural. For instance, the higher the Gleason score for a sample is, the higher the level of activity of Cell-Cycle pathway is. By clustering pathways based on their level of activity, we noted that pathways were grouped and their change in behaviour was correlated. Finally, thanks to the Metabolic network, some pathways can help differentiate mechanisms between normal and tumour cases such as the Inositol-phosphate and Glycerophospholipid metabolisms.

In conclusion, the visualisation of data on top of comprehensive networks of interconnected pathways allow an intuitive and quick analysis of the state of a sample and guide the observer to investigate into certain directions based on altered pathways and genes. Thanks to the available maps containing a vast number of pathways, this type of analyses can be easily performed and help to create hypotheses that could reorient research directions.

Chapter 5 List of Abbreviations

ACSN	Atlas of Cancer Signalling Network
TME	Tumour Micro-Environment

Chapter 6 Bibliography

[1] Kuperstein I, Bonnet E, Nguyen HA, Cohen D, Viara E, Grieco L, Fourquet S, Calzone L, Russo C, Kondratova M, Dutreix M, Barillot E, Zinovyev A. Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. Oncogenesis. 2015 Jul 20;4:e160. doi: 10.1038/oncsis.2015.19. Pubmed ID: 26192618.

[2] Thiele et al., A community-driven global reconstruction of human metabolism, Nat Biotech, 2013.

[3] Kuperstein I, Cohen DPA, Pooks S, Viara E, Calzone L, Barillot E and Zinovyev A. NaviCell: a web-based environment for navigation, curation and maintenance of large molecular interaction maps. BMC Syst Biol 2013 7(1):100

doi: 10.1186/1752-0509-7-100

[4] Martignetti, L., Calzone, L., Bonnet, E., Barillot, E., & Zinovyev, A. (2016). ROMA: Representation and Quantification of Module Activity from Target Expression Data. Frontiers in Genetics, 7, 18. http://doi.org/10.3389/fgene.2016.00018