Mathematical methods for using biological networks in the analysis of Big Data in cancer research

Andrei Zinovyev

Computational Systems Biology of Cancer

U900 **Institut Curie**/INSERM/Ecole des Mines Paristech
Paris, France

Computational Systems Biology of Cancer

group at Institut Curie http://sysbio.curie.fr

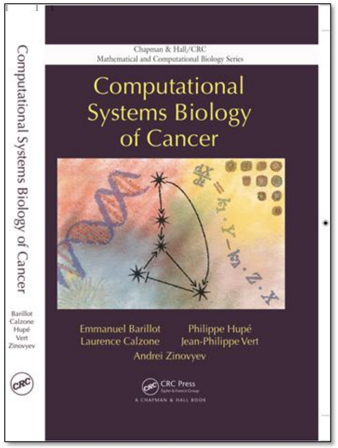
Directions:

1) Omics data analysis using **biological networks**

2) Statistical **analysis of multi-level omics data**

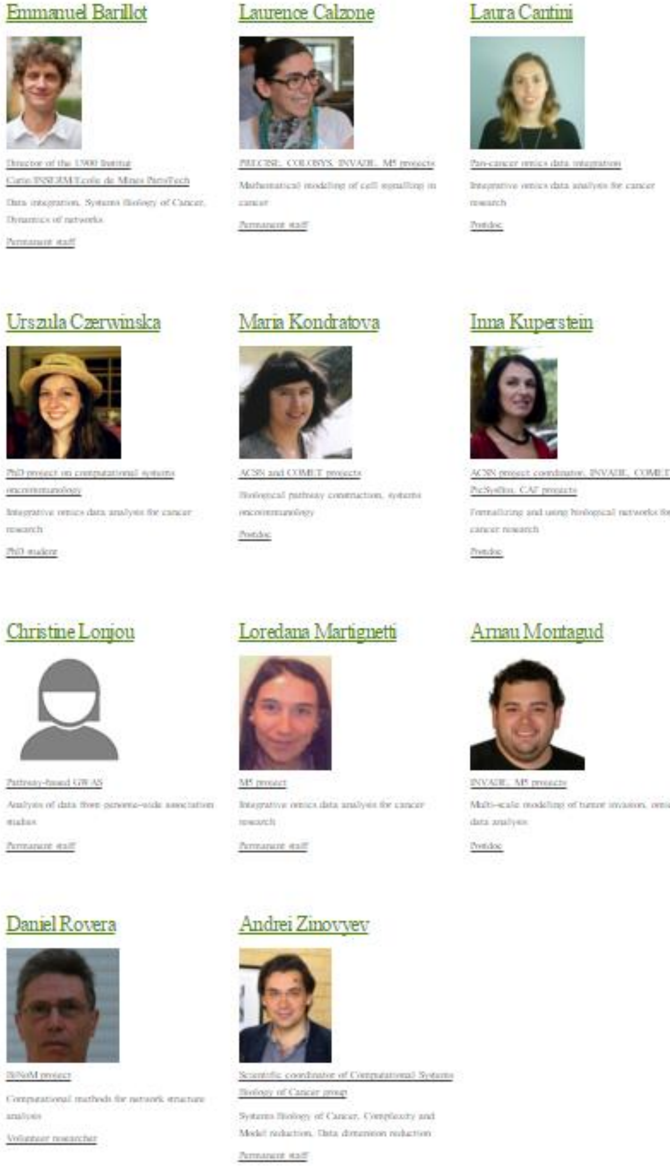
3) **Mathematical modeling** of networks involved in tumor growth, interaction with microenvironment, metastases

4) Methods and **software development** for systems biology


Chapman & Hall/CRC
Mathematical and Computational Biology Series
Computational Systems Biology of Cancer
Emmanuel Barillot Philippe Hupé
Laurence Calzone Jean-Philippe Vert
Andrei Zinovyev
CRC Press

Dynamic international and multidisciplinary environment
Numerous collaborative projects on concrete questions cancer biology and cancer treatment
Access to original large-scale data from application of latest technologies

We are HIRING postdocs!


Emmanuel Barillot
Director of the U900 Institut Curie/INSERM/Ecole de Mines ParisTech
Data integration, Systems Biology of Cancer, Dynamics of networks
Permanent staff

Laurence Calzone
PRECISE, COLOSYS, INVADE, M5 projects
Mathematical modeling of cell signalling in cancer
Permanent staff

Laura Cantini
Pan-cancer omics data integration
Integrative omics data analysis for cancer research
Postdoc

Urszula Czerwinska
PhD project on computational systems oncoimmunology
Integrative omics data analysis for cancer research
PhD student

Maria Kondratova
ACSN and COMET projects
Biological pathway construction, systems oncoimmunology
Postdoc

Inna Kuperstein
ACSN project coordinator, INVADE, COMET
PicSysBio, CAF projects
Formalizing and using biological networks fo
cancer research
Postdoc

Christine Lonjou
Pathway-based GWAS
Analysis of data from genome-wide association studies
Permanent staff

Loredana Martignetti
M5 project
Integrative omics data analysis for cancer research
Permanent staff

Arnau Montagud
INVADE, M5 projects
Multi-scale modeling of tumor invasion, omi
data analysis
Postdoc

Daniel Rovera
BiNoM project
Computational methods for network structure analysis
Volunteer researcher

Andrei Zinovyev
Scientific coordinator of Computational Systems Biology of Cancer group
Systems Biology of Cancer, Complexity and Model reduction, Data dimension reduction
Permanent staff

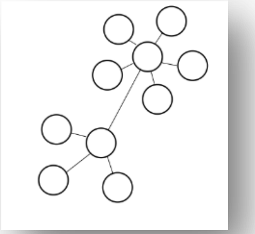
institut Curie

Inserm
Institut national de la santé et de la recherche médicale

Networks in Molecular Biology short history

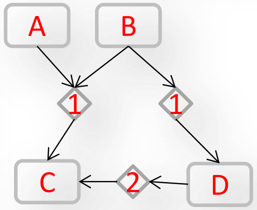
- Network of (bio-)chemical reactions governing cellular life (metabolism, signaling)
- Graph theory methods in mathematical chemistry (structure vs dynamics)
- Pattern of connections between molecular entities is more important than the properties of the entities
- Notion of emergence, non-reductionism
- Robust network functioning (scale-free properties)
- Idea of general network theory
- Nowadays: network biology, network medicine

Molecule-Molecule interaction network



(weighted undirected graph)

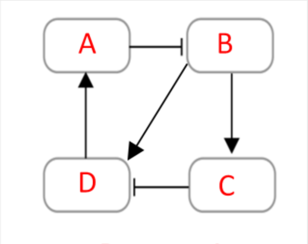
Reaction networks (chemical kinetics)



A + B -> C
B -> D
D -> 2C

(bipartite directed graph)

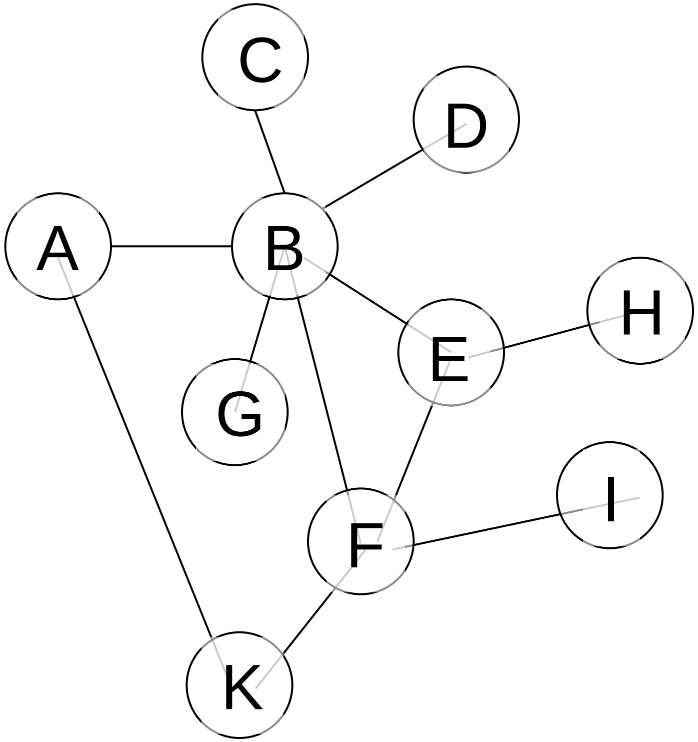
Regulatory networks

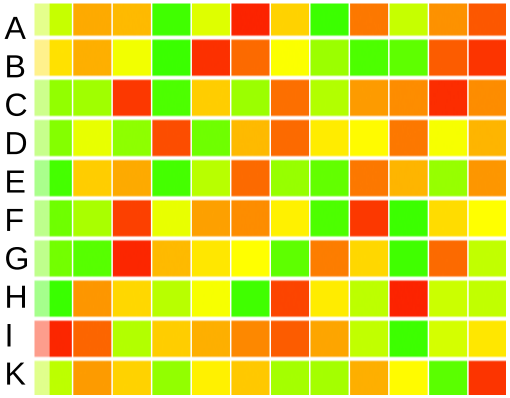


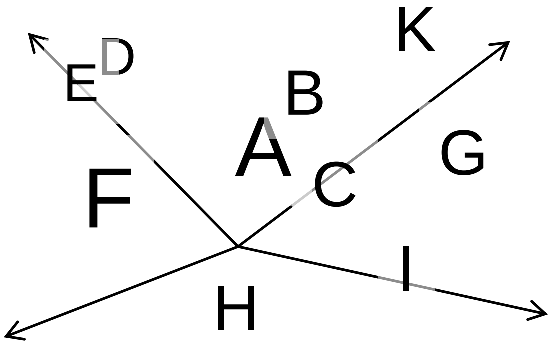
B <- not A
A <- D
C <- B
D <- B and not C

(signed directed graph)

Mathematical methods for using biological networks in data analysis







institut Curie

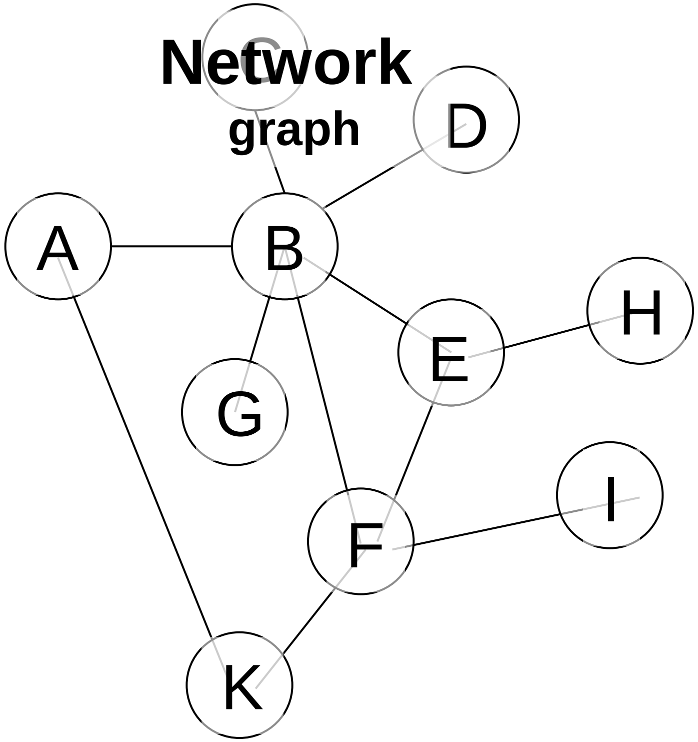
Inserm
Institut national
de la santé et de la recherche médicale

MINES PARIS

Mathematical methods for using biological networks in data analysis

Network

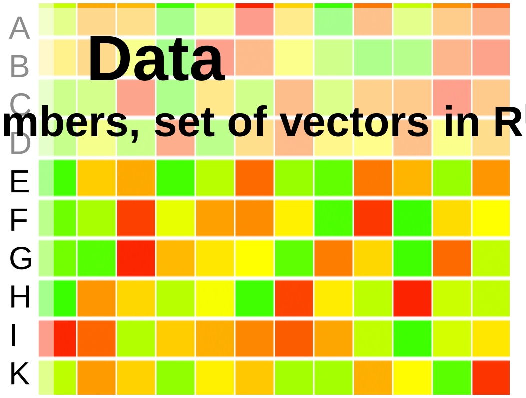
graph

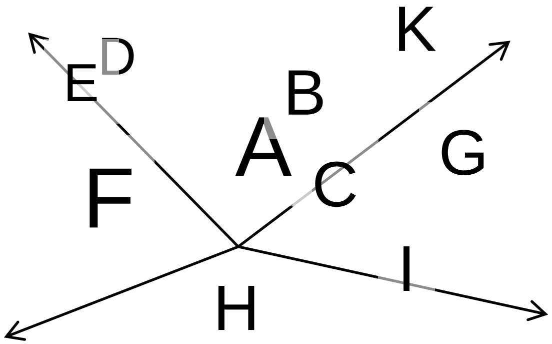


undirected graph, directed graph
directed signed graph, bipartite graph

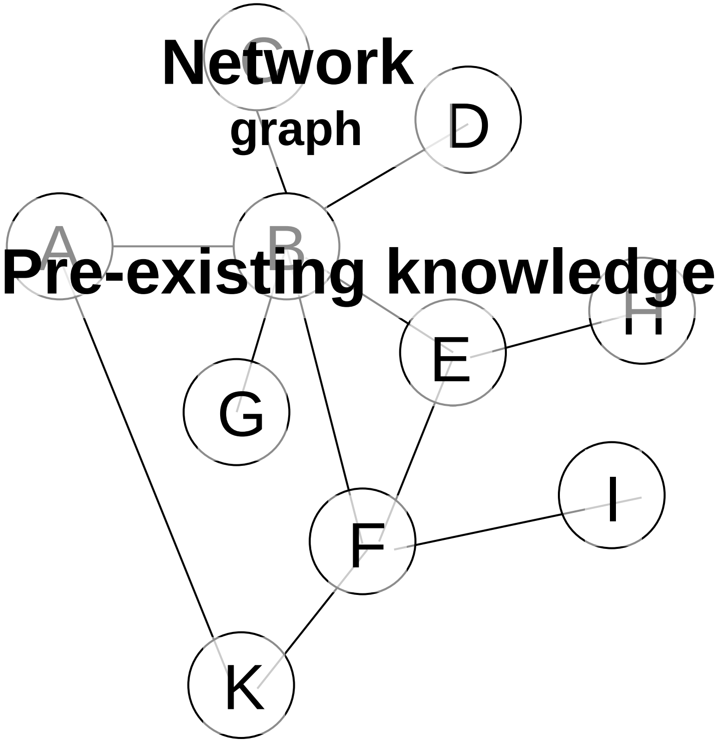
Data

(table of numbers, set of vectors in R^N)





Mathematical methods for using biological networks in data analysis

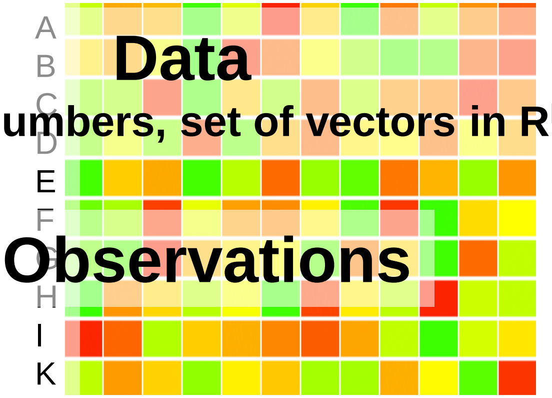


Network

graph

Pre-existing knowledge

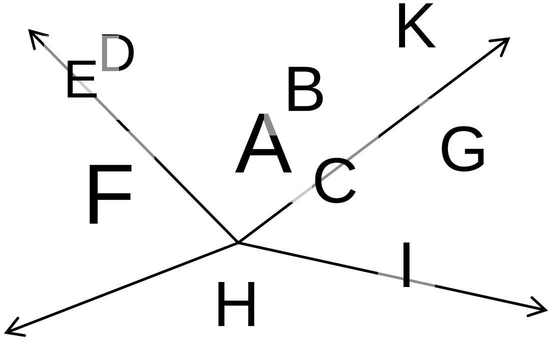
undirected graph, directed graph
directed signed graph, bipartite graph



Data

(table of numbers, set of vectors in R^N)

Observations

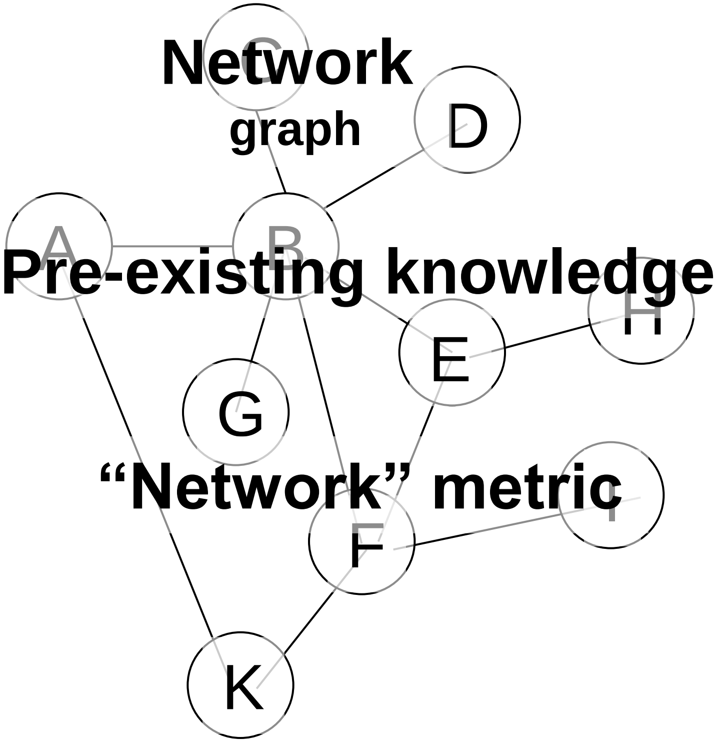


institutCurie

Inserm
Institut national
de la santé et de la recherche médicale

MINES PARIS

Mathematical methods for using biological networks in data analysis



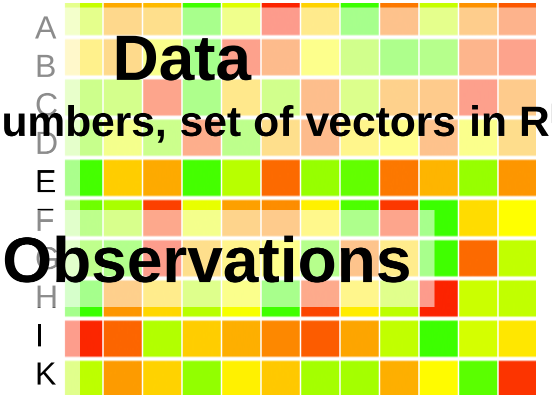
Network

graph

Pre-existing knowledge

“Network” metric

undirected graph, directed graph
directed signed graph, bipartite graph

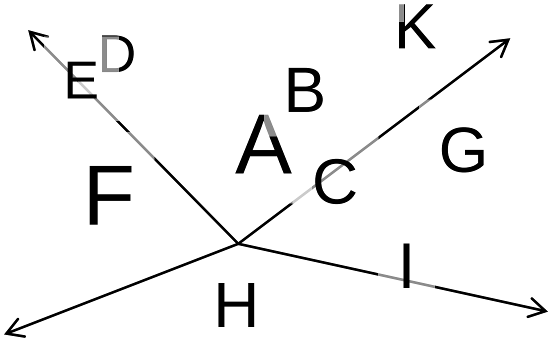


Data

(table of numbers, set of vectors in R^N)

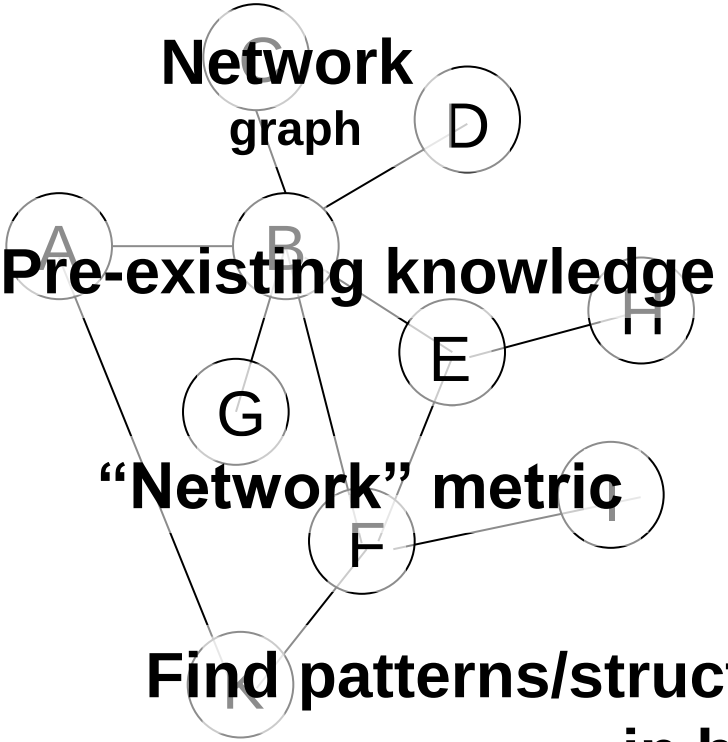
Observations

“Data” metric



Institut national de la santé et de la recherche médicale

Mathematical methods for using biological networks in data analysis



Network

graph

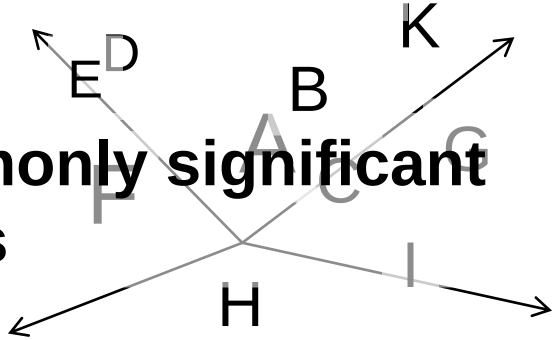
Data

(table of numbers, set of vectors in R^N)

Pre-existing knowledge

Observations

“Network” metric



“Data” metric

Find patterns/structures commonly significant in both metrics

undirected graph, directed graph
directed signed graph, bipartite graph

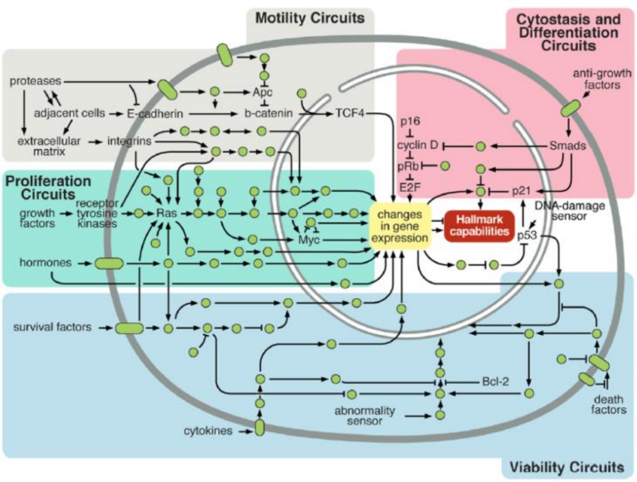
Plan of the talk

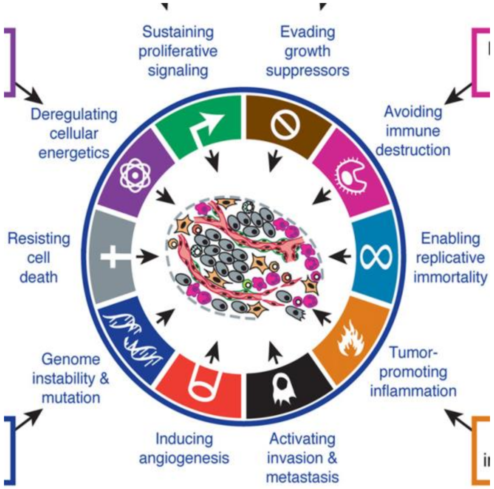
- Constructing biological networks from reading literature (knowledge formalization)
- Visualizing data onto network layouts
- Finding significant subnetworks
- Network propagation principle
- Minimal network cuts
- ~~*Mathematical modeling of network dynamics*~~

Atlas of cancer signaling network:

project inspired by the authors of hallmarks of cancer

Hanahan and Weinberg, 2001



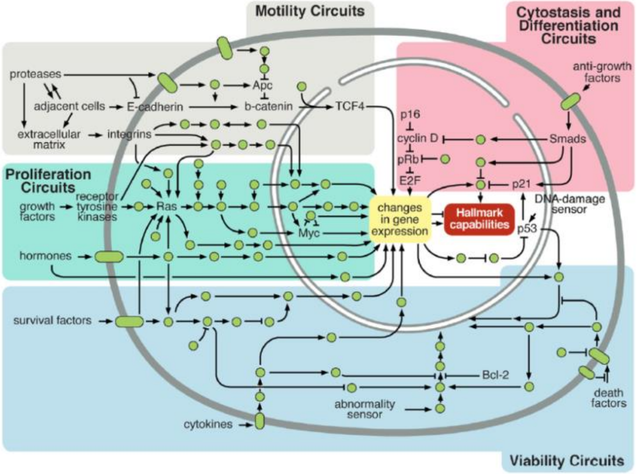


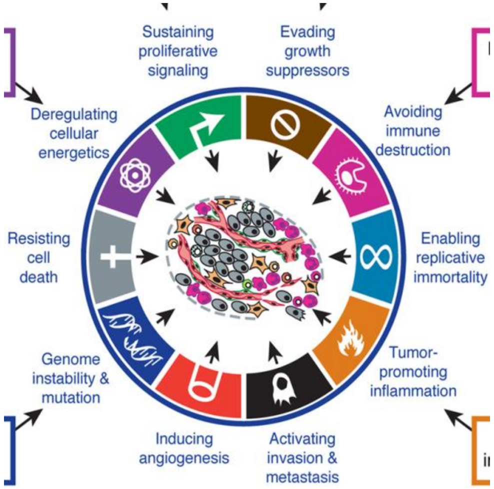
Hanahan and Weinberg, 2011, Cell

Atlas of cancer signaling network:

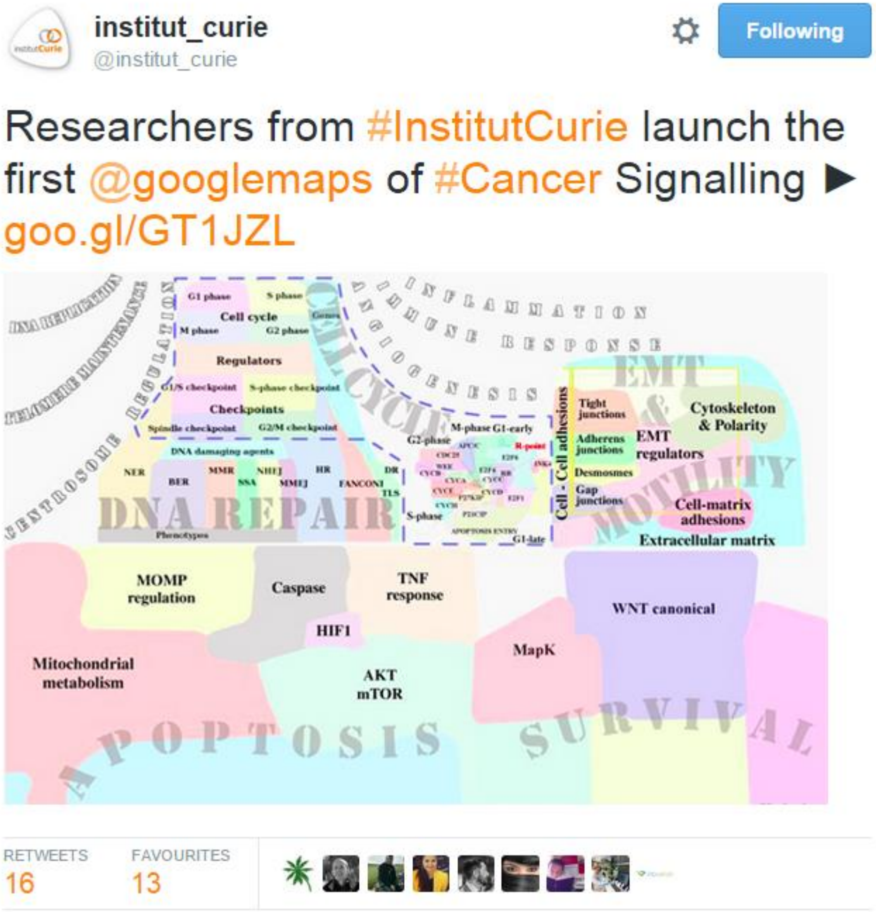
project inspired by the authors of hallmarks of cancer

Hanahan and Weinberg, 2001





Hanahan and Weinberg, 2011, Cell



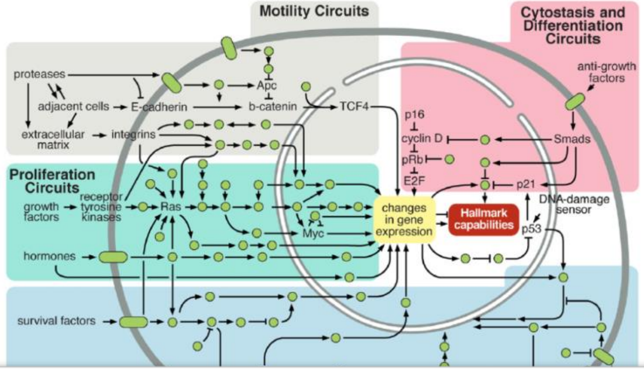
institut_curie
@institut_curie

Researchers from #InstitutCurie launch the first @googlemaps of #Cancer Signalling ▶ goo.gl/GT1JZL

RETWEETS 16
FAVOURITES 13

Atlas of cancer signaling network:

project inspired by the authors of hallmarks of cancer

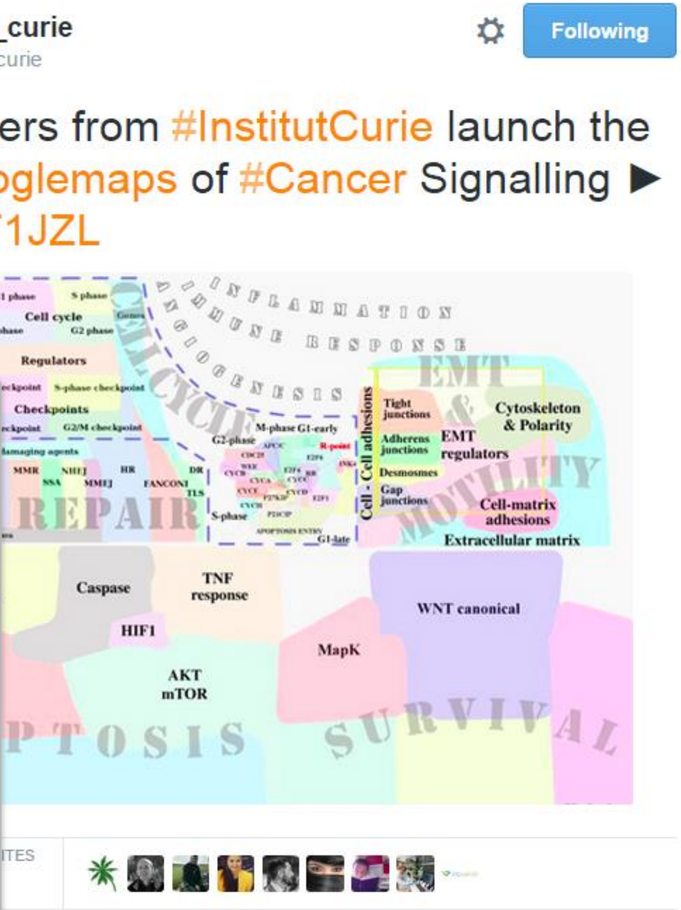


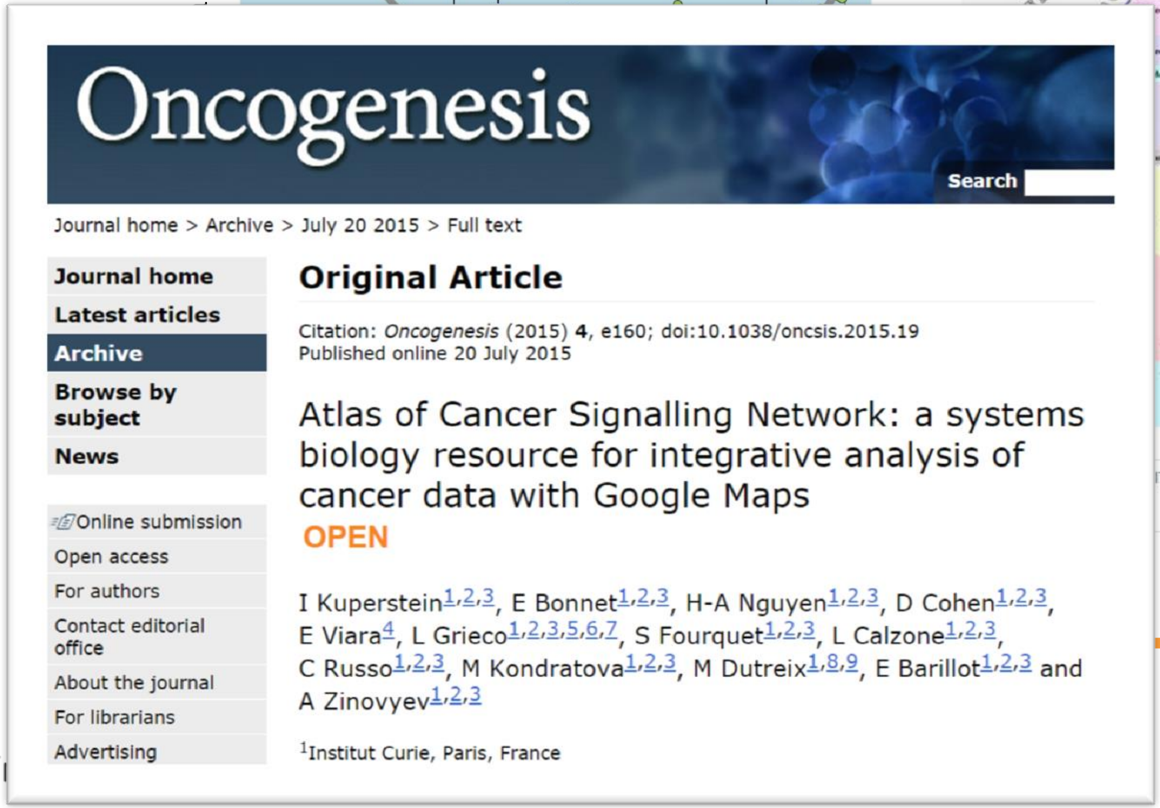
han and Weinberg, 2001

institut_curie
@institut_curie

Following

Researchers from #InstitutCurie launch the first @googlemaps of #Cancer Signalling ▶ goo.gl/GT1JZL





Oncogenesis

Search

Journal home > Archive > July 20 2015 > Full text

- **Journal home**
- **Latest articles**
- **Archive**
- **Browse by subject**
- **News**

- Online submission
- Open access
- For authors
- Contact editorial office
- About the journal
- For librarians
- Advertising

Original Article

Citation: *Oncogenesis* (2015) **4**, e160; doi:10.1038/oncsis.2015.19
Published online 20 July 2015

Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps

OPEN

I Kuperstein[1,2,3], E Bonnet[1,2,3], H-A Nguyen[1,2,3], D Cohen[1,2,3], E Viara[4], L Grieco[1,2,3,5,6,7], S Fourquet[1,2,3], L Calzone[1,2,3], C Russo[1,2,3], M Kondratova[1,2,3], M Dutreix[1,8,9], E Barillot[1,2,3] and A Zinovyev[1,2,3]

[1]Institut Curie, Paris, France

i

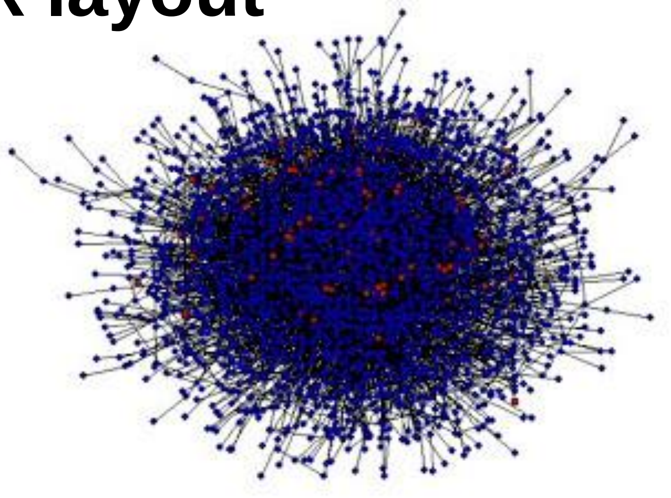
MINES PARIS

Biological network and Biological network map

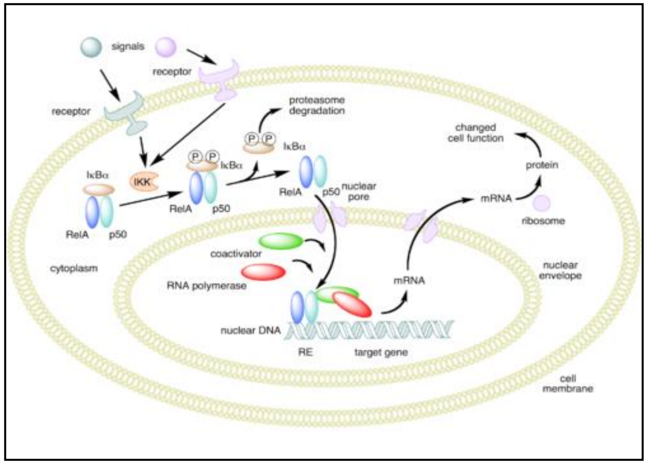
- **Network** – set of connections between biological entities

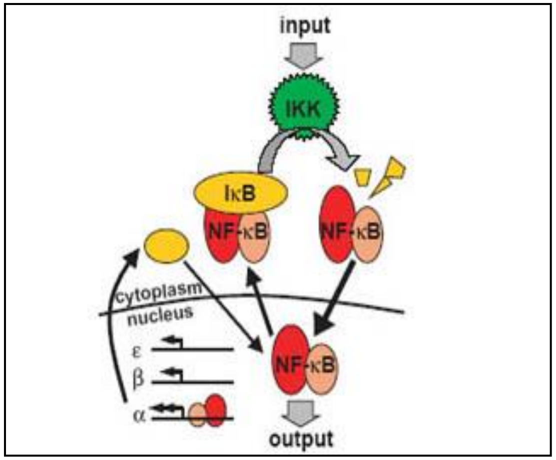
- **Network map** – graphical representation of the network, layout + decorations

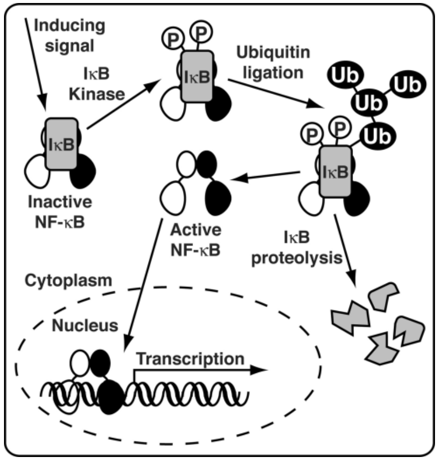
- Problem of meaningfull **network layout**

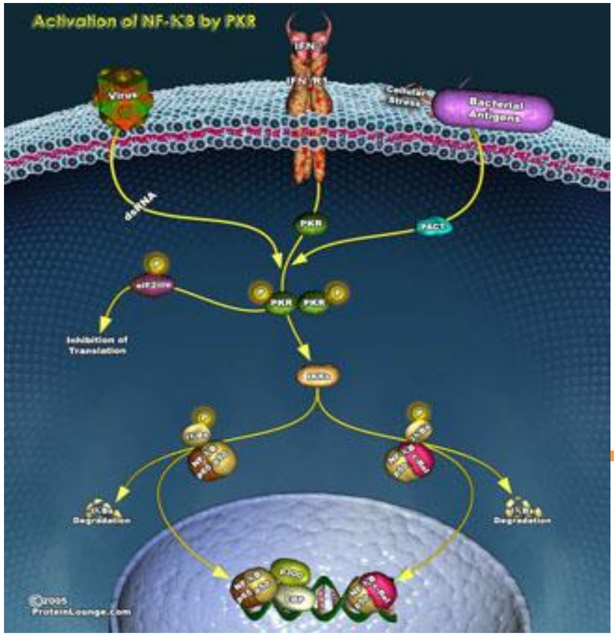


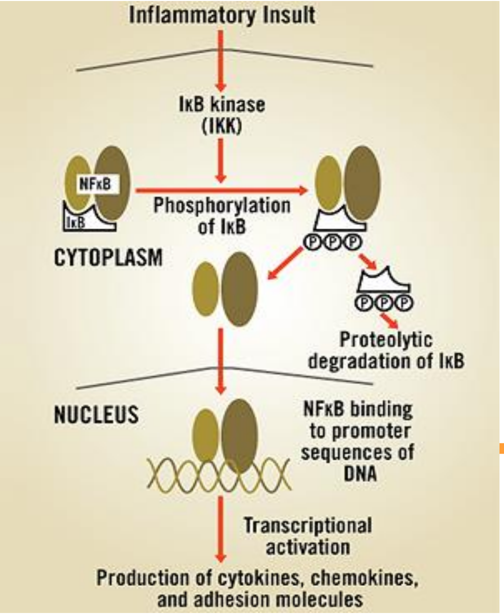
Diagrams+human language vs computer-readable formal language

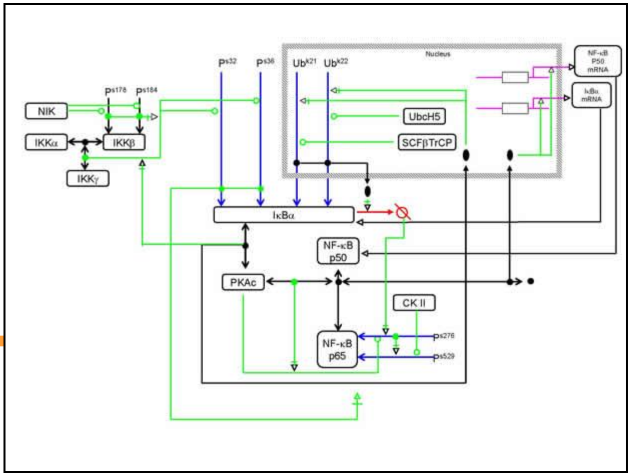




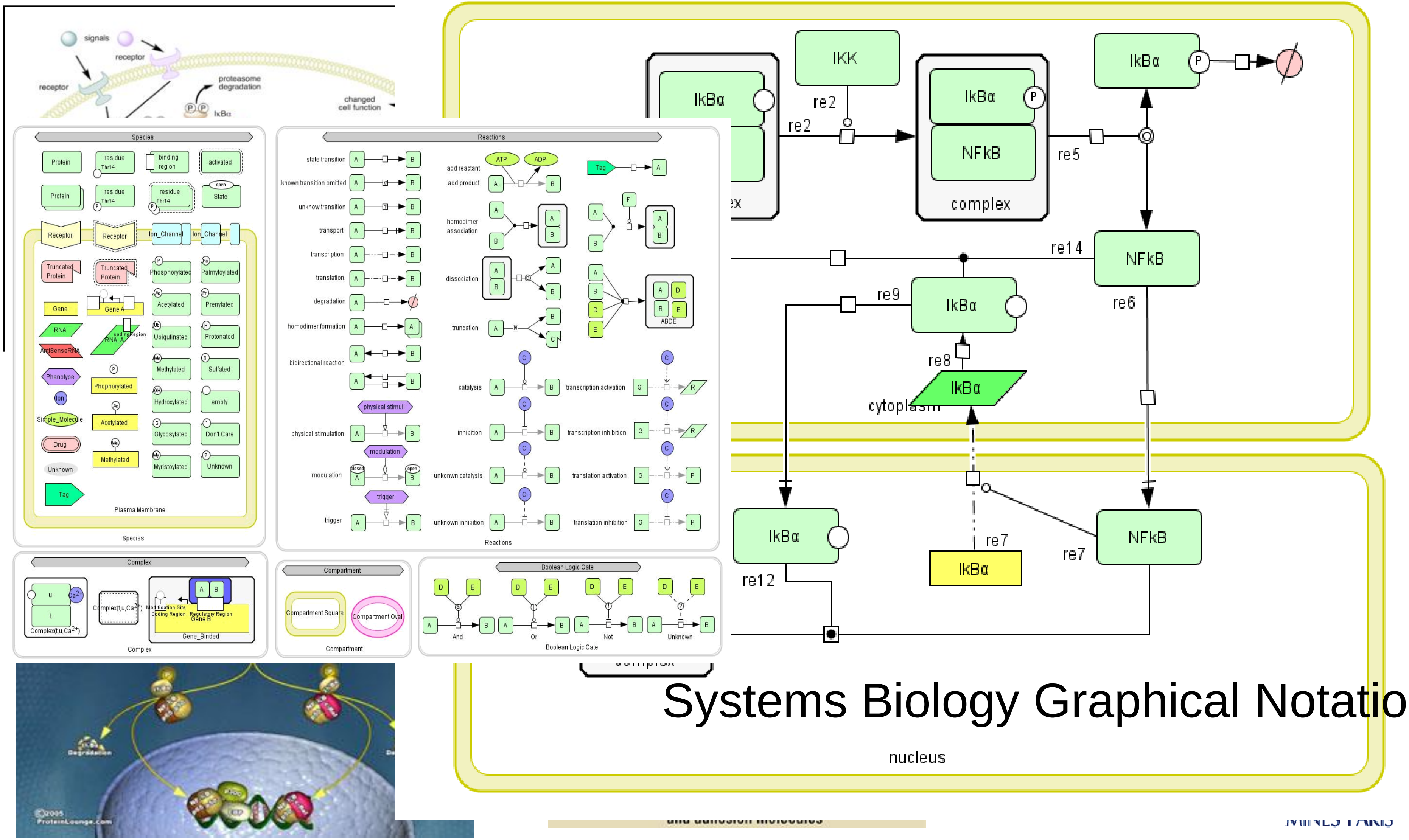








Diagrams+human language
vs computer-readable formal language



Systems Biology Graphical Notation

Atlas of Cancer Signaling Network and NaviCell

http://acsn.curie.fr http://navicell.curie.fr

- **ACSN** – “world map” of cancer biology

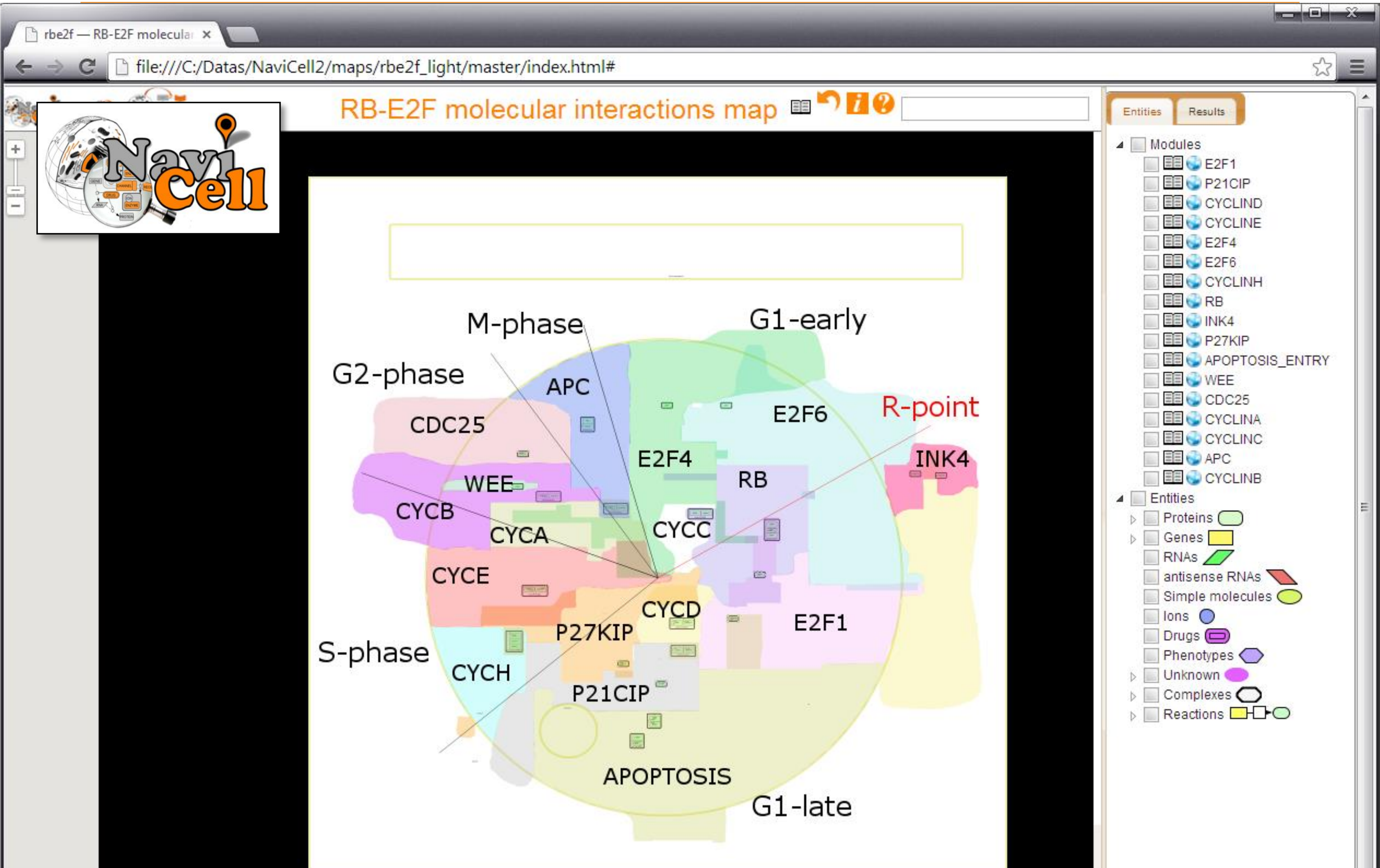
the largest cancer-specific reaction network map

(4826 reactions, 2371 proteins, 5975 chemical species)

- **NaviCell** – Google Maps-based engine to browse VERY LARGE biological network maps and visualizing data on top of them

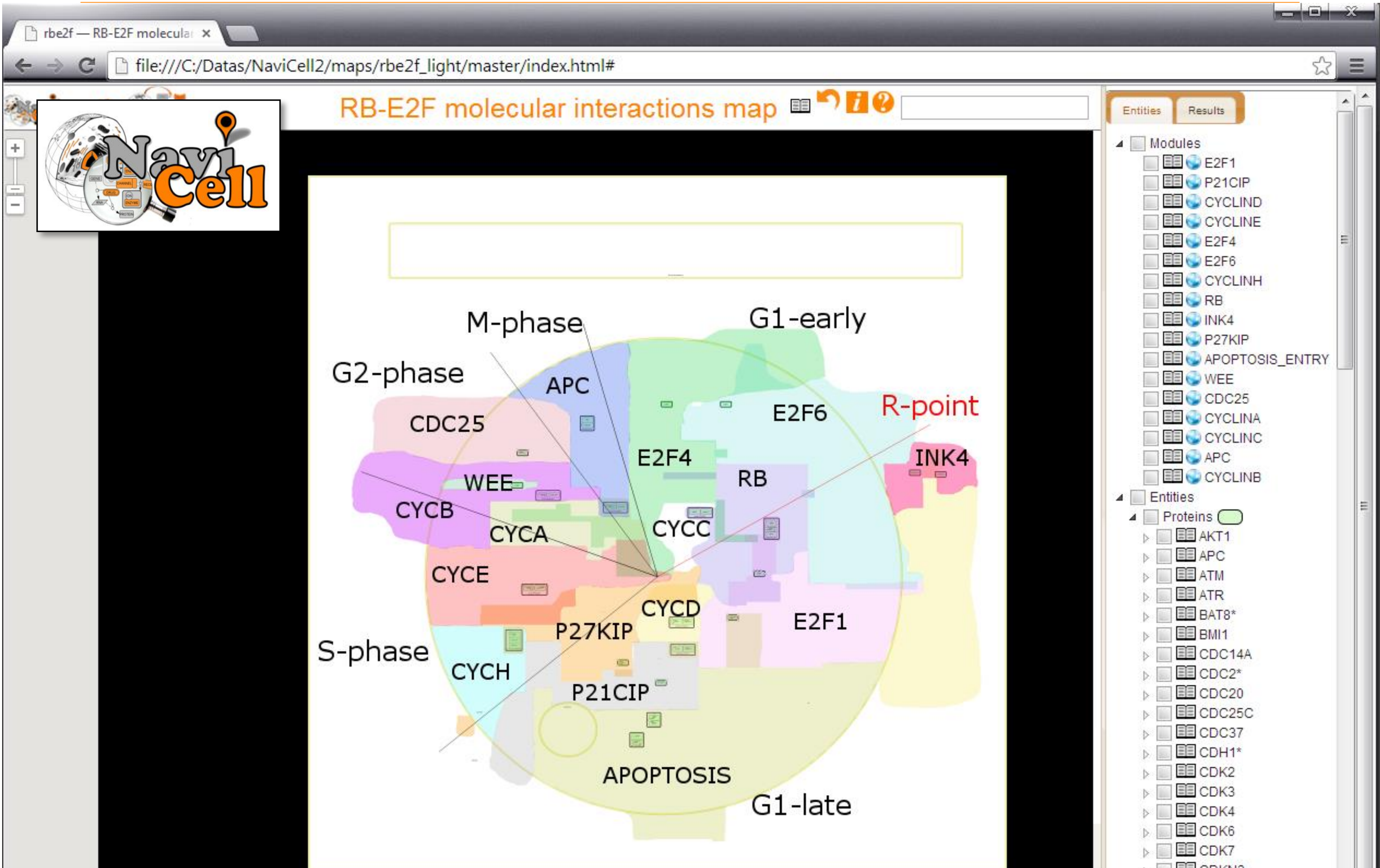
Example: Google maps of cell cycle

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013; Bonnet et al, NAR, 2015)



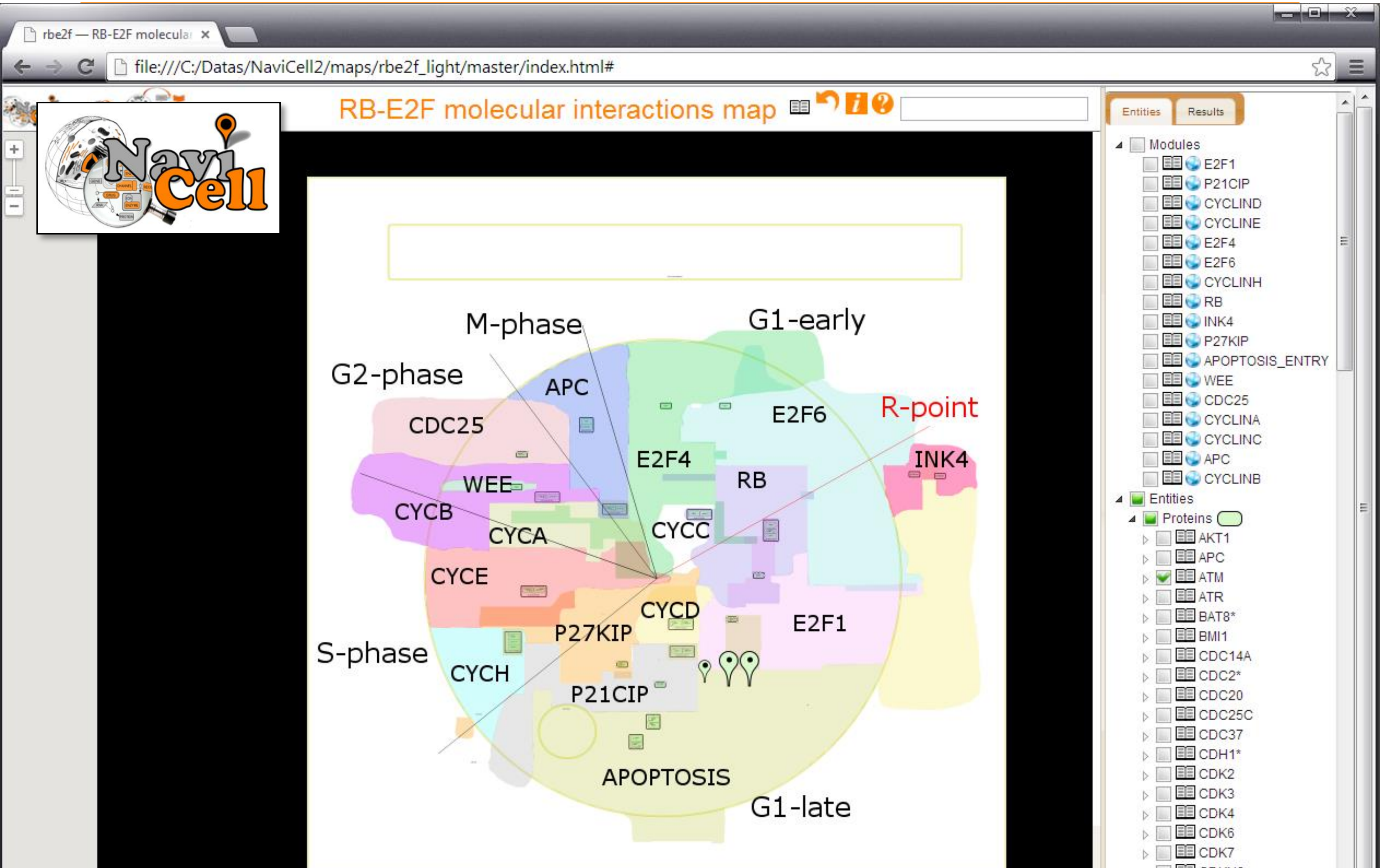
Example: Google maps of cell cycle

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013; Bonnet et al, NAR, 2015)



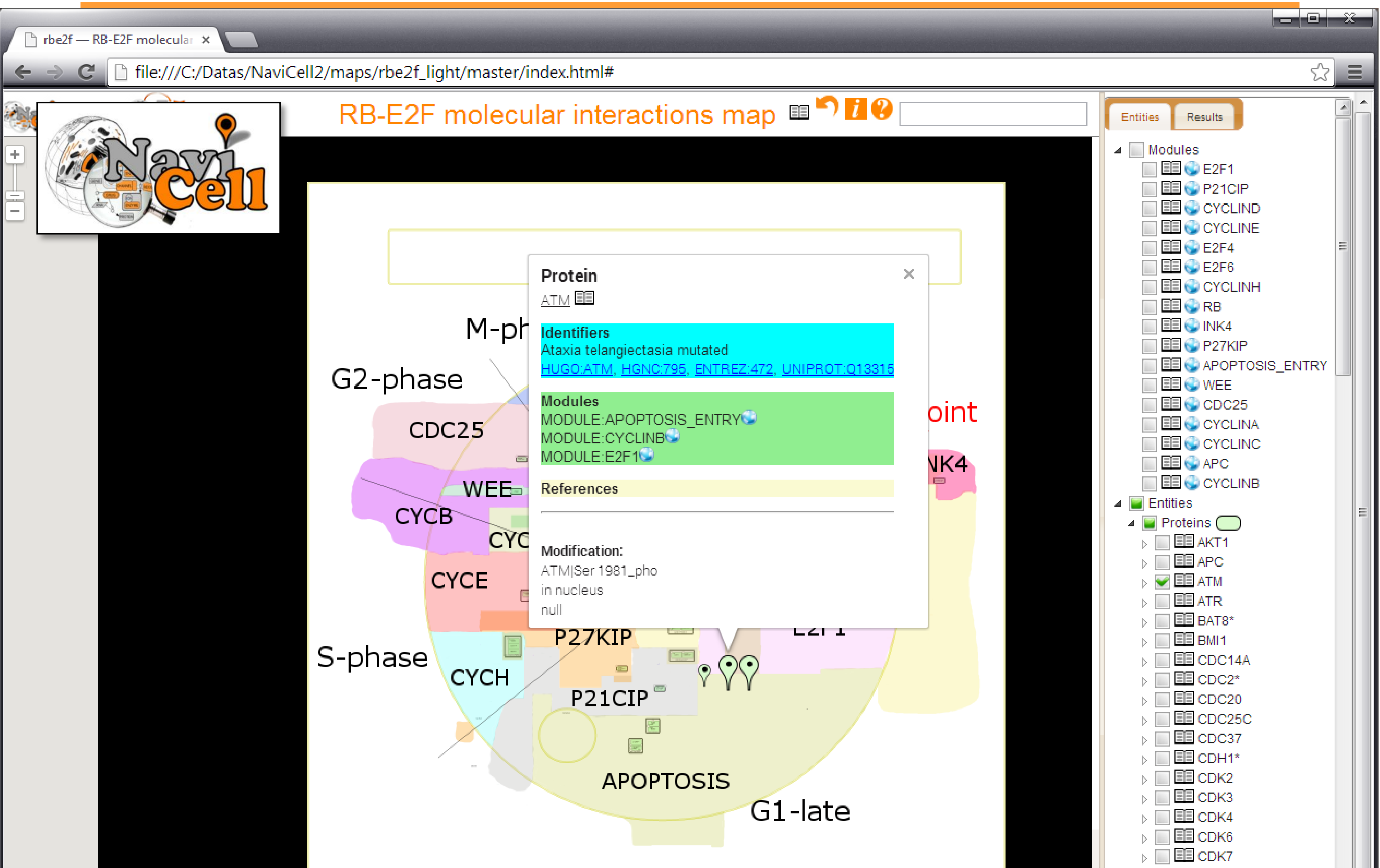
Example: Google maps of cell cycle

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013; Bonnet et al, NAR, 2015)



Semantic zoom in NaviCell

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013; Bonnet et al, NAR, 2015)



Protein

ATM

Identifiers

Ataxia telangiectasia mutated

HUGO:ATM, HGNC:795, ENTREZ:472, UNIPROT:Q13315

Modules

MODULE:APOPTOSIS_ENTRY

MODULE:CYCLINB

MODULE:E2F1

References

Modification:

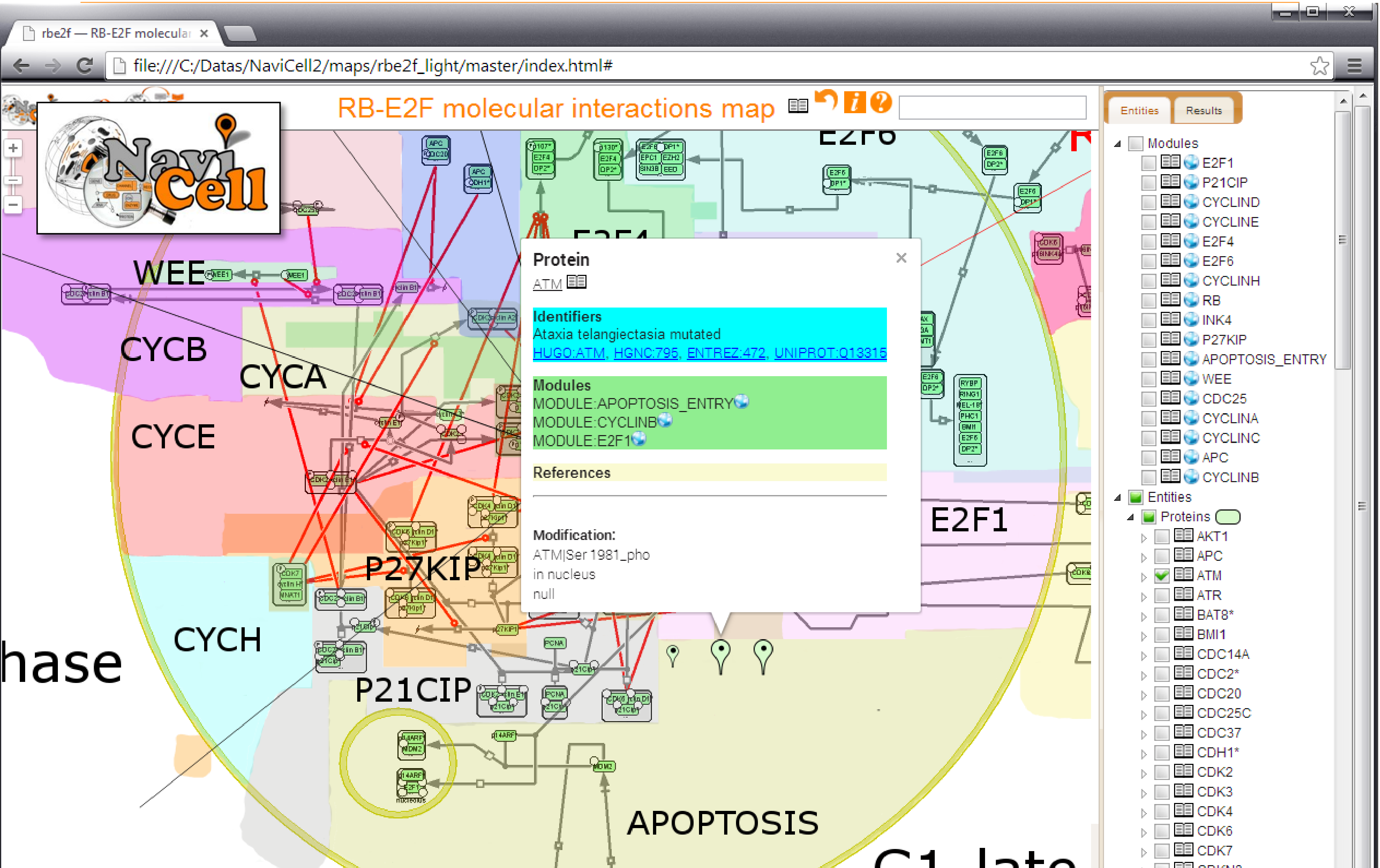
ATM|Ser 1981_pho

in nucleus

null

Semantic zoom in NaviCell

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013; Bonnet et al, NAR, 2015)



rbe2f — RB-E2F molecular

file:///C:/Datas/NaviCell2/maps/rbe2f_light/master/index.html#

RB-E2F molecular interactions map

Protein

ATM

Identifiers

Ataxia telangiectasia mutated

HUGO:ATM, HGNC:795, ENTREZ:472, UNIPROT:Q13315

Modules

MODULE:APOPTOSIS_ENTRY

MODULE:CYCLINB

MODULE:E2F1

References

Modification:

ATM|Ser 1981_pho

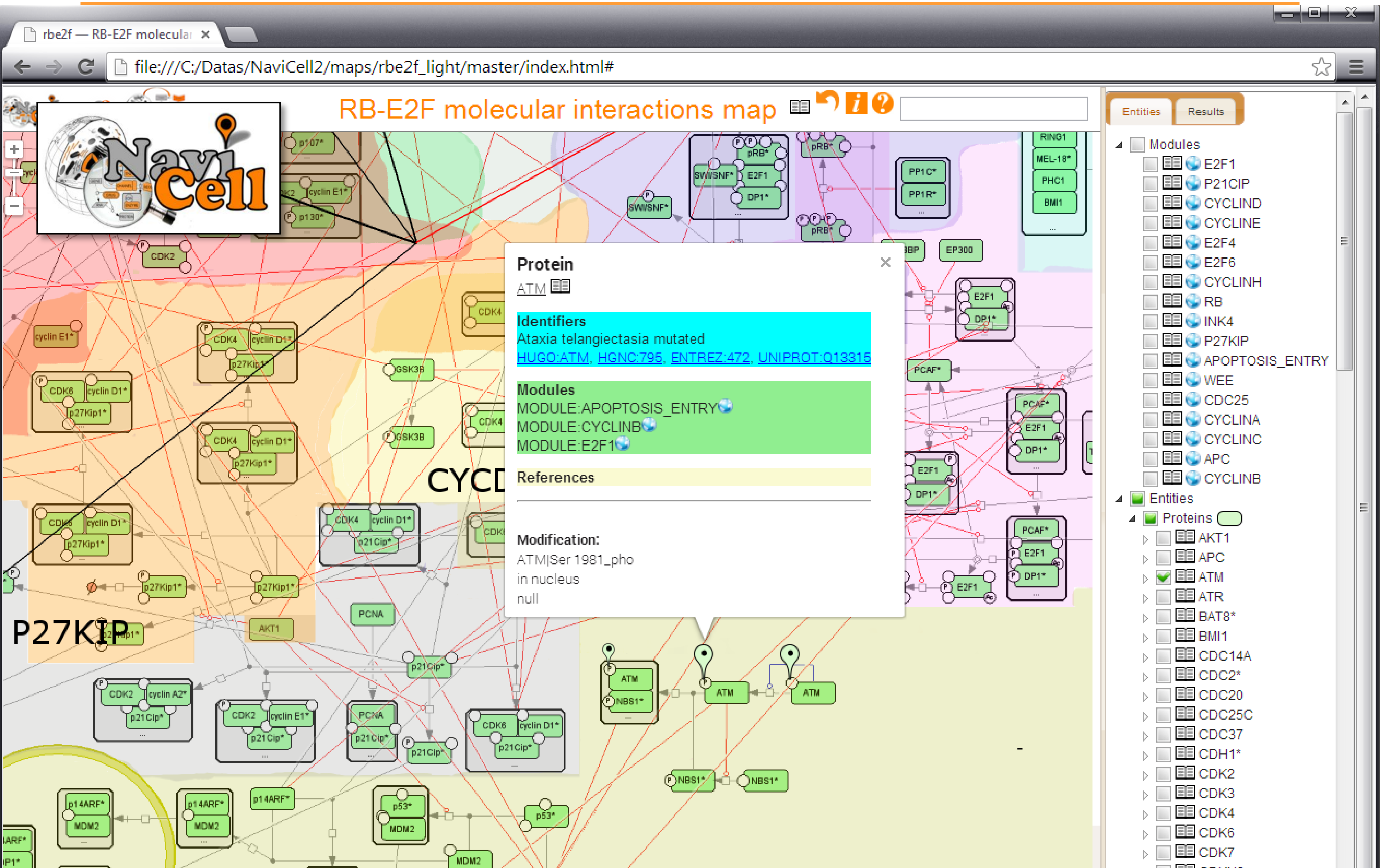
in nucleus

null

- ☐ Modules
  - ☐ E2F1
  - ☐ P21CIP
  - ☐ CYCLIND
  - ☐ CYCLINE
  - ☐ E2F4
  - ☐ E2F6
  - ☐ CYCLINH
  - ☐ RB
  - ☐ INK4
  - ☐ P27KIP
  - ☐ APOPTOSIS_ENTRY
  - ☐ WEE
  - ☐ CDC25
  - ☐ CYCLINA
  - ☐ CYCLINC
  - ☐ APC
  - ☐ CYCLINB
- ☑ Entities
  - ☑ Proteins
    - ☐ AKT1
    - ☐ APC
    - ☑ ATM
    - ☐ ATR
    - ☐ BAT8*
    - ☐ BMI1
    - ☐ CDC14A
    - ☐ CDC2*
    - ☐ CDC20
    - ☐ CDC25C
    - ☐ CDC37
    - ☐ CDH1*
    - ☐ CDK2
    - ☐ CDK3
    - ☐ CDK4
    - ☐ CDK6
    - ☐ CDK7

Semantic zoom in NaviCell

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013; Bonnet et al, NAR, 2015)



rbe2f — RB-E2F molecular

file:///C:/Datas/NaviCell2/maps/rbe2f_light/master/index.html#

RB-E2F molecular interactions map

Protein

ATM

Identifiers

Ataxia telangiectasia mutated

HUGO:ATM, HGNC:795, ENTREZ:472, UNIPROT:Q13315

Modules

MODULE:APOPTOSIS_ENTRY

MODULE:CYCLINB

MODULE:E2F1

References

Modification:

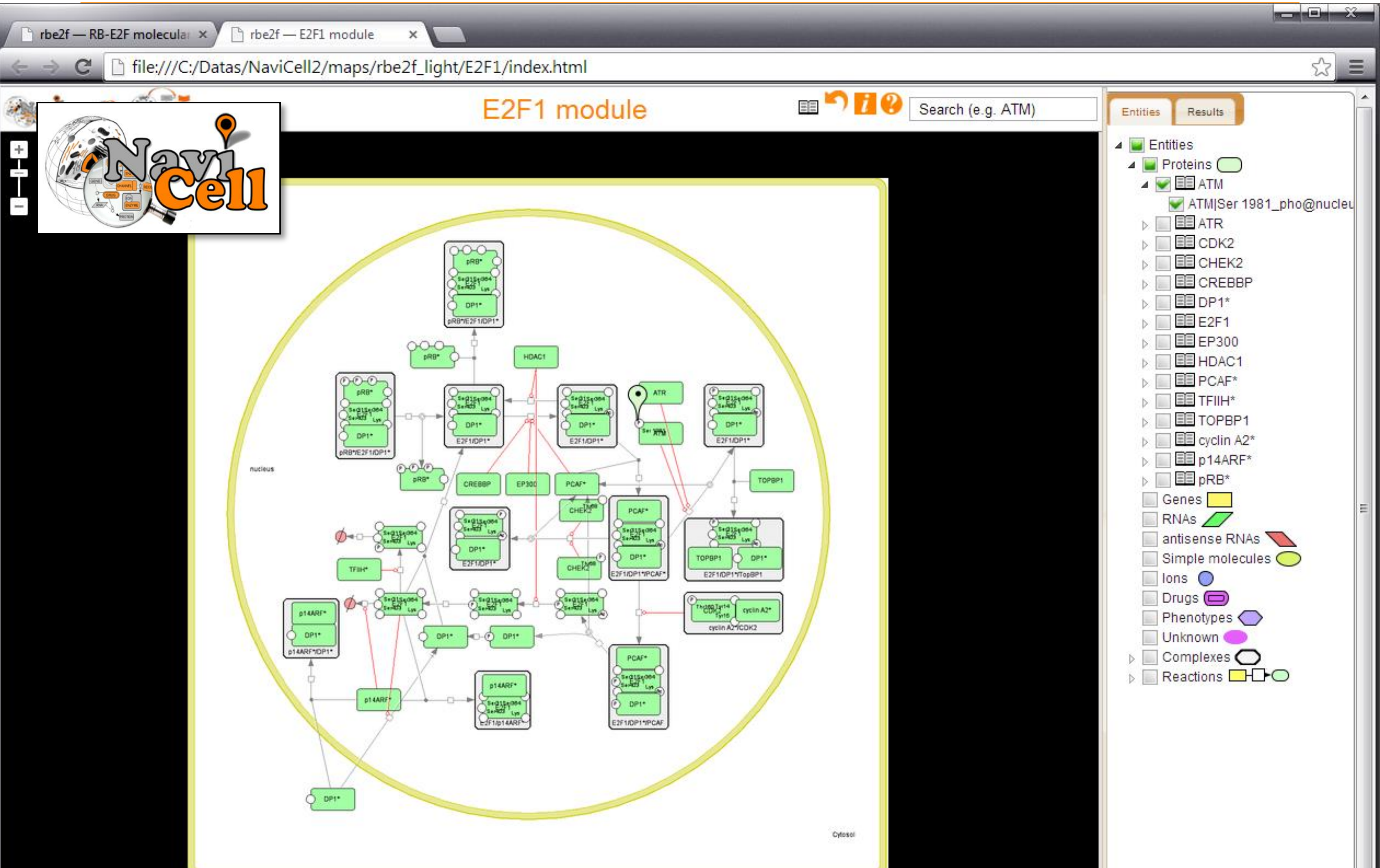
ATM|Ser 1981_pho

in nucleus

null

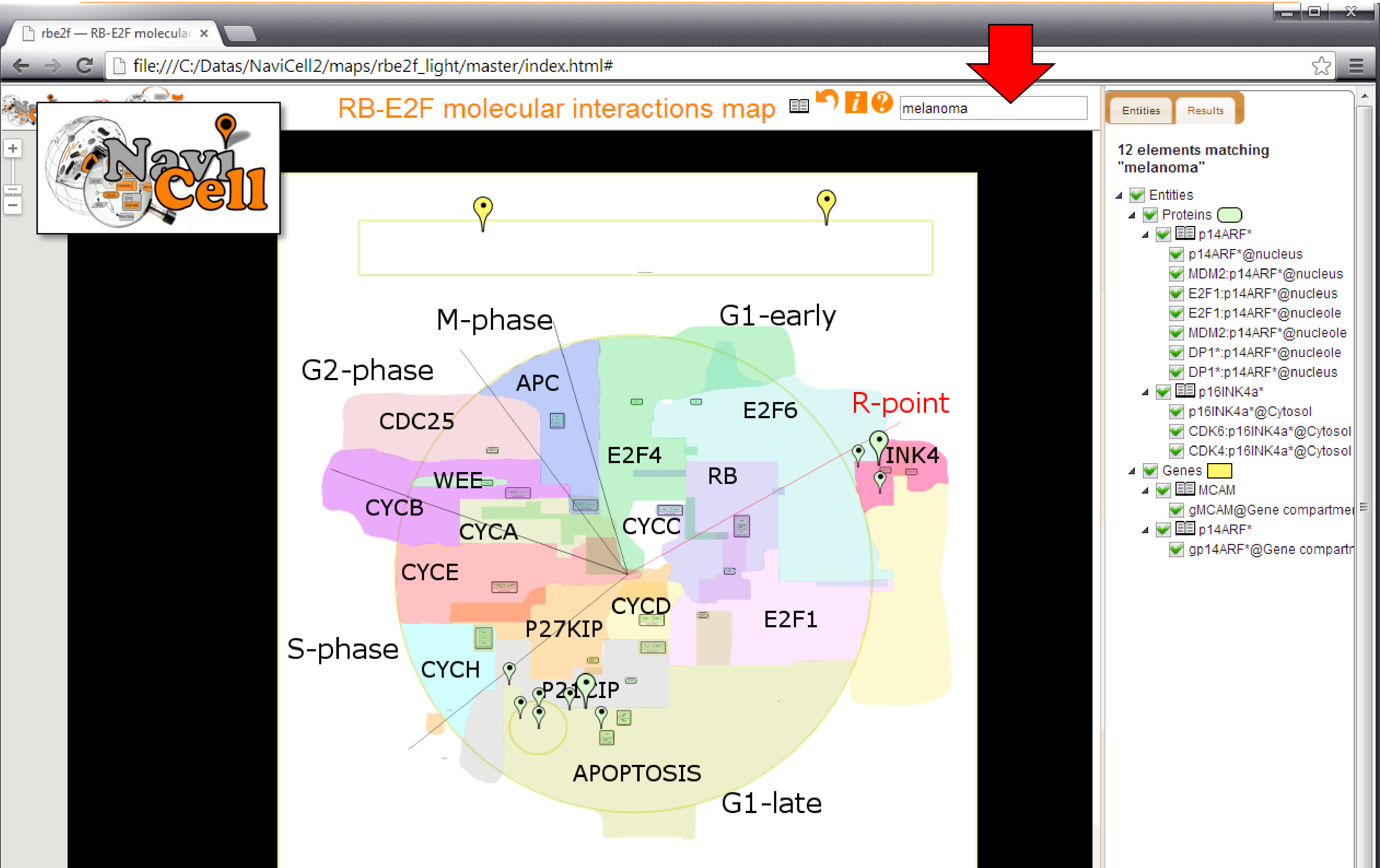
Module maps with simpler layout

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013; Bonnet et al, NAR, 2015)



Keyword search in NaviCell

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013)



rbe2f — RB-E2F molecular

file:///C:/Datas/NaviCell2/maps/rbe2f_light/master/index.html#

RB-E2F molecular interactions map

melanoma

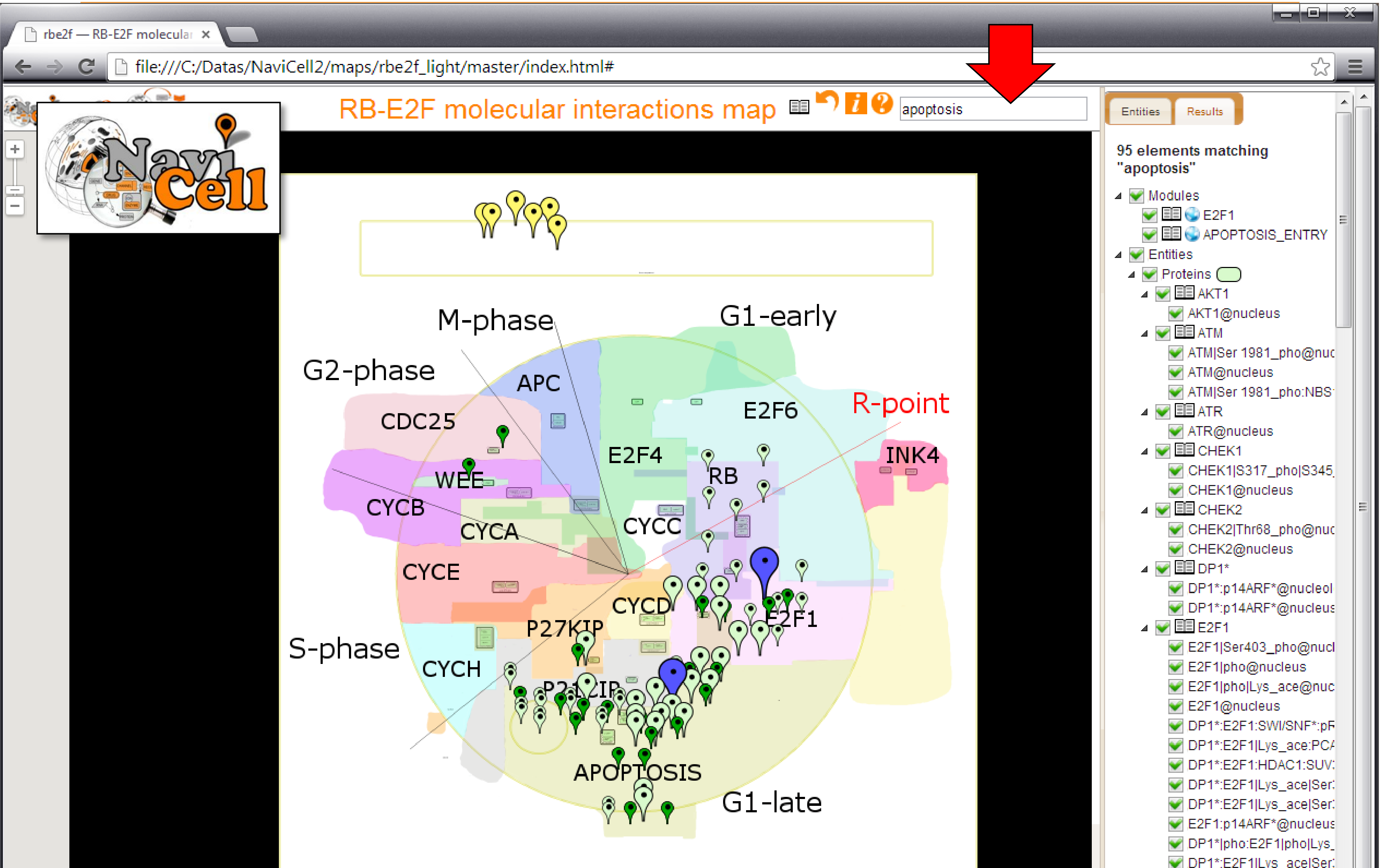
Entities Results

12 elements matching "melanoma"

- ☑ Entities
  - ☑ Proteins
    - ☑ p14ARF*
      - ☑ p14ARF*@nucleus
      - ☑ MDM2:p14ARF*@nucleus
      - ☑ E2F1:p14ARF*@nucleus
      - ☑ E2F1:p14ARF*@nucleole
      - ☑ MDM2:p14ARF*@nucleole
      - ☑ DP1*:p14ARF*@nucleole
      - ☑ DP1*:p14ARF*@nucleus
    - ☑ p16INK4a*
      - ☑ p16INK4a*@Cytosol
      - ☑ CDK6:p16INK4a*@Cytosol
      - ☑ CDK4:p16INK4a*@Cytosol
  - ☑ Genes
    - ☑ MCAM
      - ☑ gMCAM@Gene compartme
    - ☑ p14ARF*
      - ☑ gp14ARF*@Gene compartn

Keyword search in NaviCell

(Calzone et al, MSB, 2008; Kuperstein et al, BMC Sys Bio, 2013; Bonnet et al, NAR, 2015)



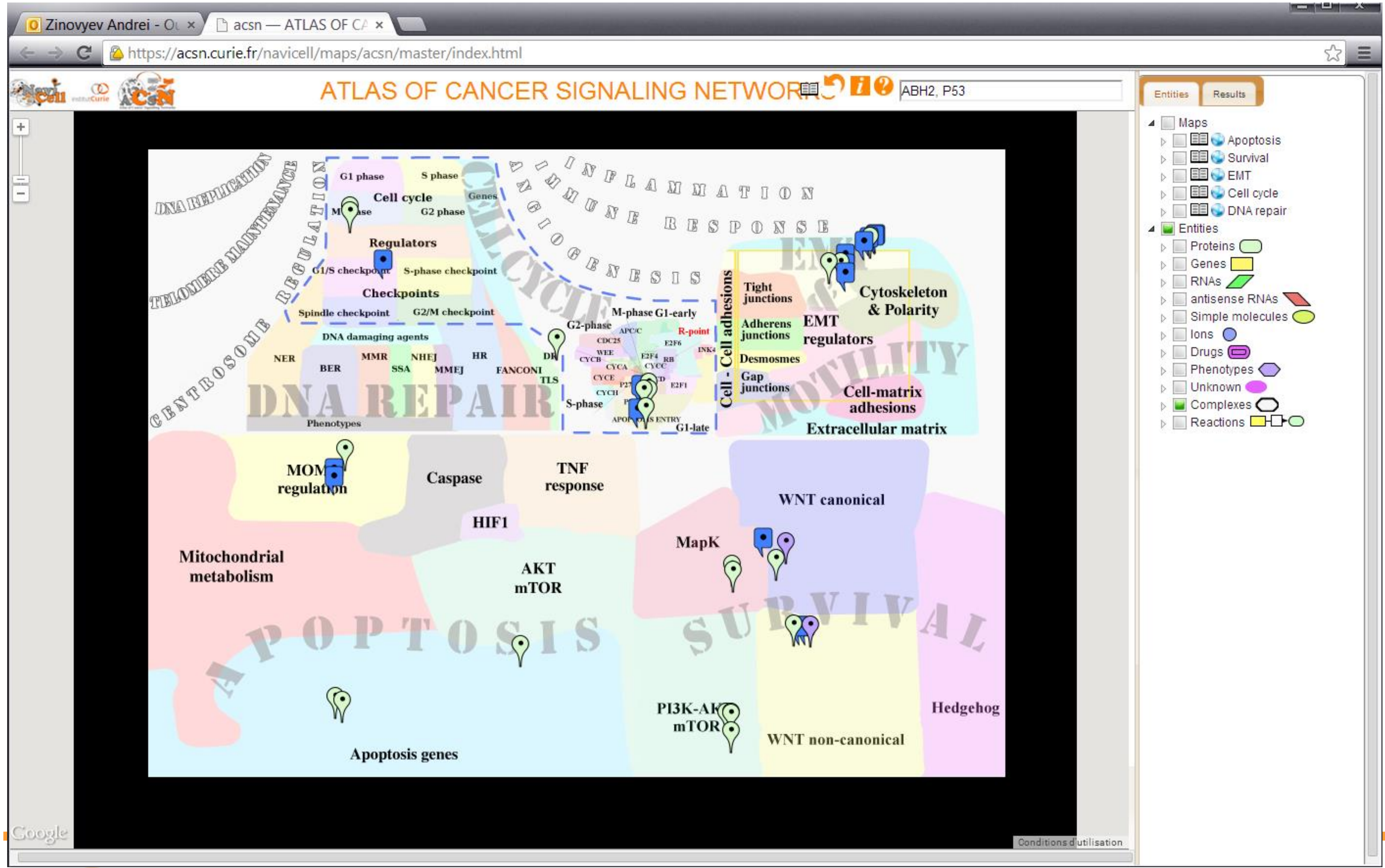
apoptosis

95 elements matching "apoptosis"

- ☑ Modules
  - ☑ E2F1
  - ☑ APOPTOSIS_ENTRY
- ☑ Entities
  - ☑ Proteins
    - ☑ AKT1
      - ☑ AKT1@nucleus
    - ☑ ATM
      - ☑ ATM|Ser 1981_pho@nuc
      - ☑ ATM@nucleus
      - ☑ ATM|Ser 1981_pho:NBS
    - ☑ ATR
      - ☑ ATR@nucleus
    - ☑ CHEK1
      - ☑ CHEK1|S317_pho|S345
      - ☑ CHEK1@nucleus
    - ☑ CHEK2
      - ☑ CHEK2|Thr68_pho@nuc
      - ☑ CHEK2@nucleus
    - ☑ DP1*
      - ☑ DP1*:p14ARF*@nucleol
      - ☑ DP1*:p14ARF*@nucleus
    - ☑ E2F1
      - ☑ E2F1|Ser403_pho@nucl
      - ☑ E2F1|pho@nucleus
      - ☑ E2F1|pho|Lys_ace@nuc
      - ☑ E2F1@nucleus
      - ☑ DP1*:E2F1:SWI/SNF*:pF
      - ☑ DP1*:E2F1|Lys_ace:PCA
      - ☑ DP1*:E2F1:HDAC1:SUV
      - ☑ DP1*:E2F1|Lys_ace|Ser
      - ☑ DP1*:E2F1|Lys_ace|Ser
      - ☑ E2F1:p14ARF*@nucleus
      - ☑ DP1*|pho:E2F1|pho|Lys_
      - ☑ DP1*:E2F1|Lys_ace|Ser

Atlas of Cancer Signaling Network

5 “continents” and 52 “countries”



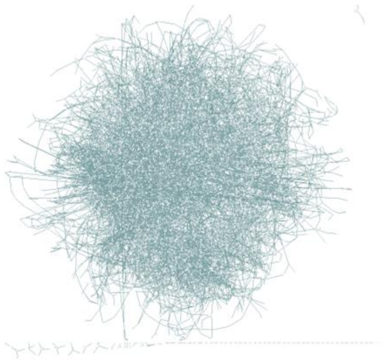
institutCurie

Inserm
Institut national de la santé et de la recherche médicale

MINES PARIS

ACSN

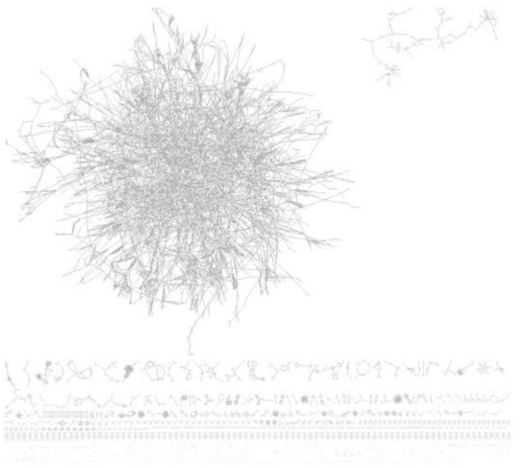
of distinct proteins = **1814**
# of protein complexes = **1306**
Reactions per protein = **2.77**
Complexes per protein = **0.72**



LCC size = **5875** species/**4305** reactions
LCC Charateristic path length, directed =**43.6**
LCC Charateristic path length, undirected =**16.3**

NCI PID

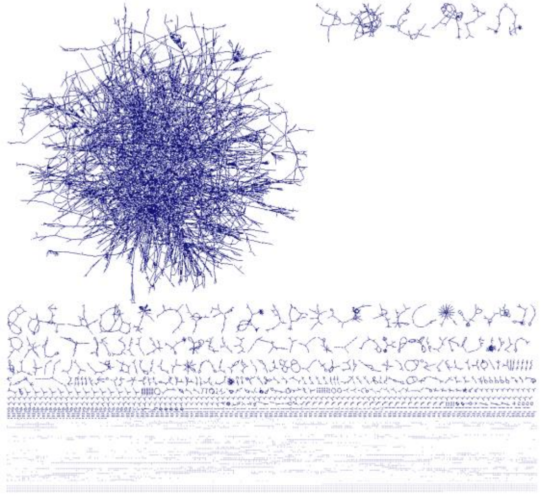
of distinct proteins = **3399**
# of protein complexes = **3904**
Reactions per protein = **2.67**
Complexes per protein = **1.15**



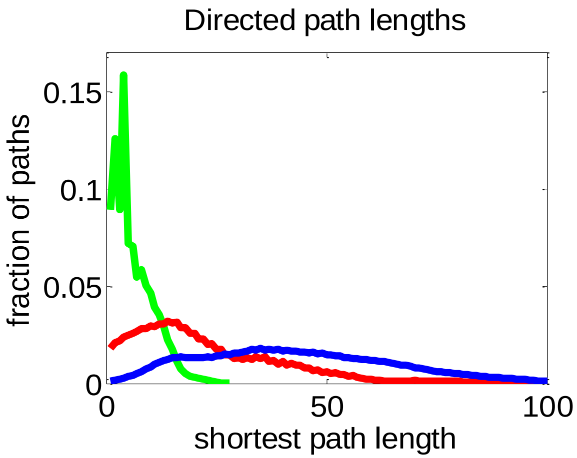
LCC size = **5038** species/**4501** reactions
LCC Characteristic path length, directed = **6.4**
LCC Characteristic path length, undirected = 1**6.8**

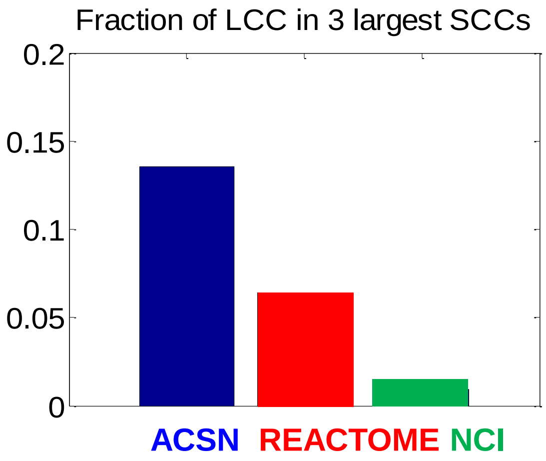
REACTOME

of distinct proteins = **8082**
# of protein complexes = **6552**
Reactions per protein = **1.06**
Complexes per protein = **0.85**



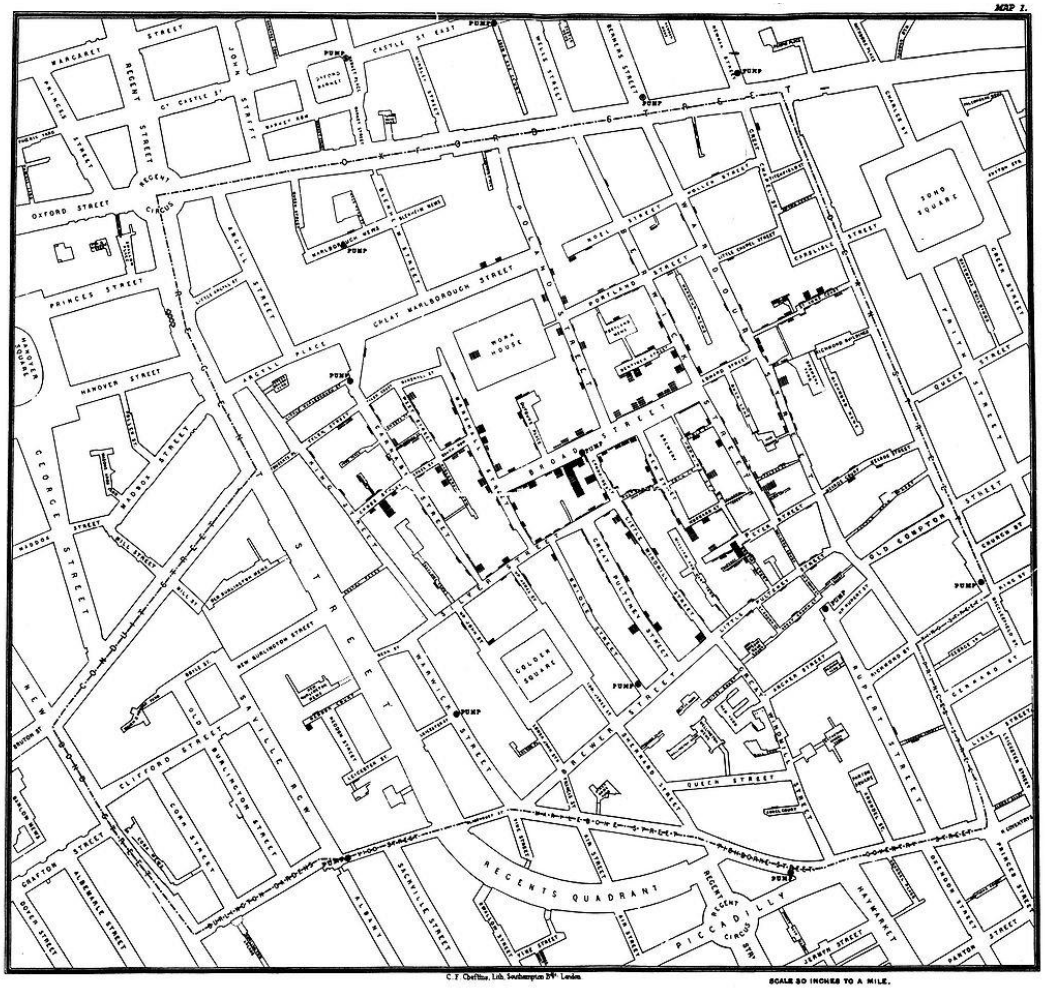
LCC size = **5181** species/**3392** reactions
LCC Characteristic path length, directed = **23.6**
LCC Characteristic path length, undirected = **24.6**





«Geographic» method

First application of geographic method

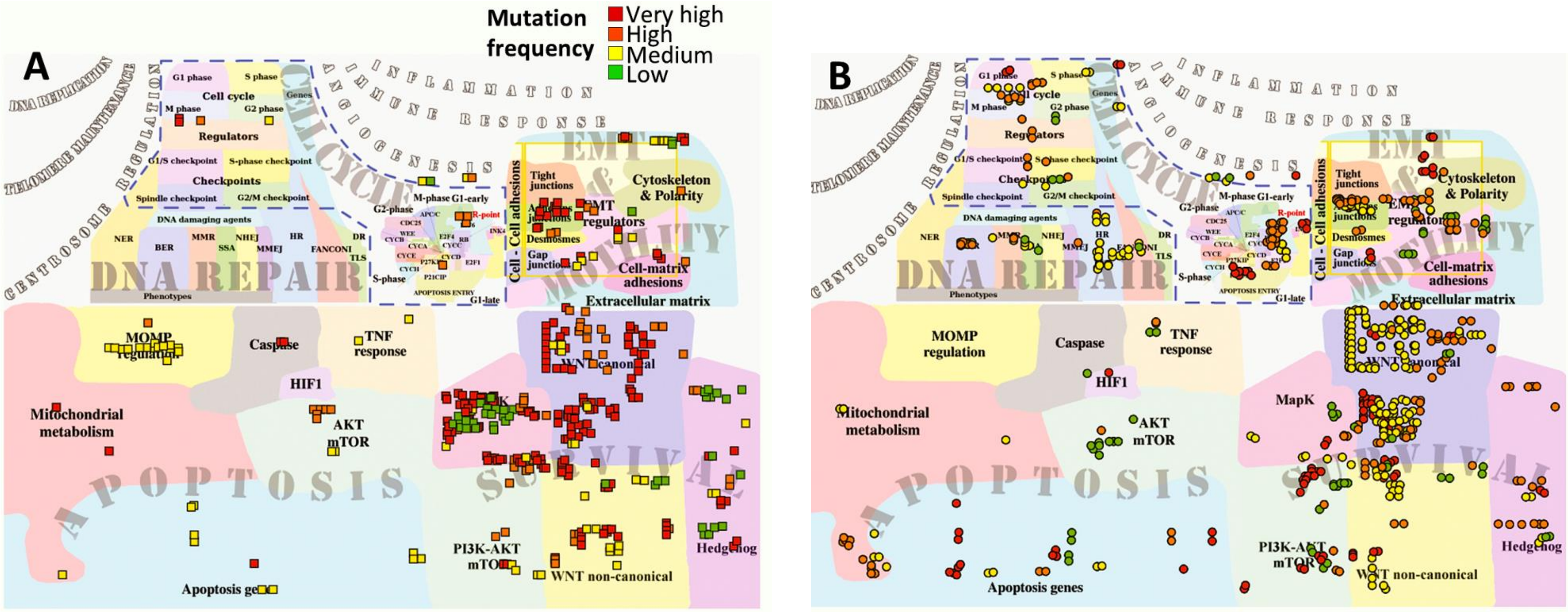


Original map by John Snow showing the clusters of cholera cases in the London epidemic of 1854. The pump is located at the intersection of Broad Street and Cambridge Street.

Data visualization with NaviCell 2.0

NaviCell Web Service (Bonnet et al, 2015, NAR)

Cancer driver genes in ACSN

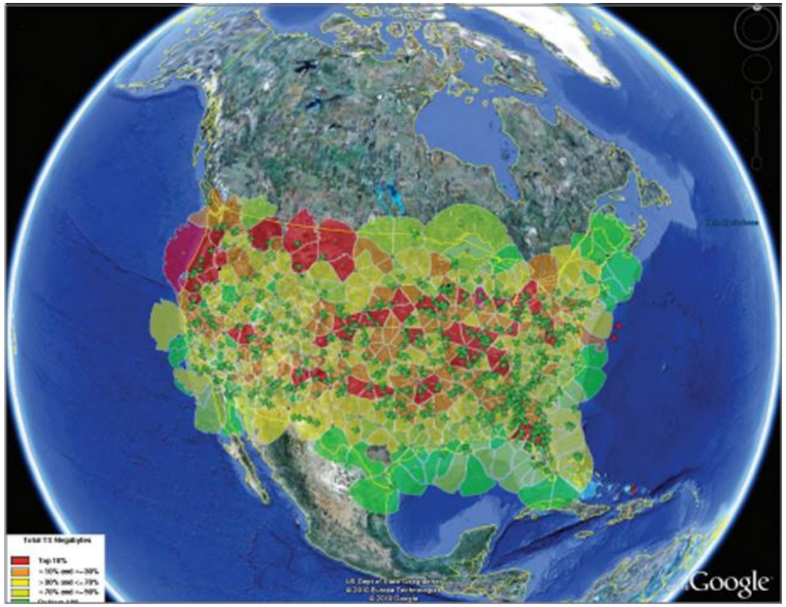


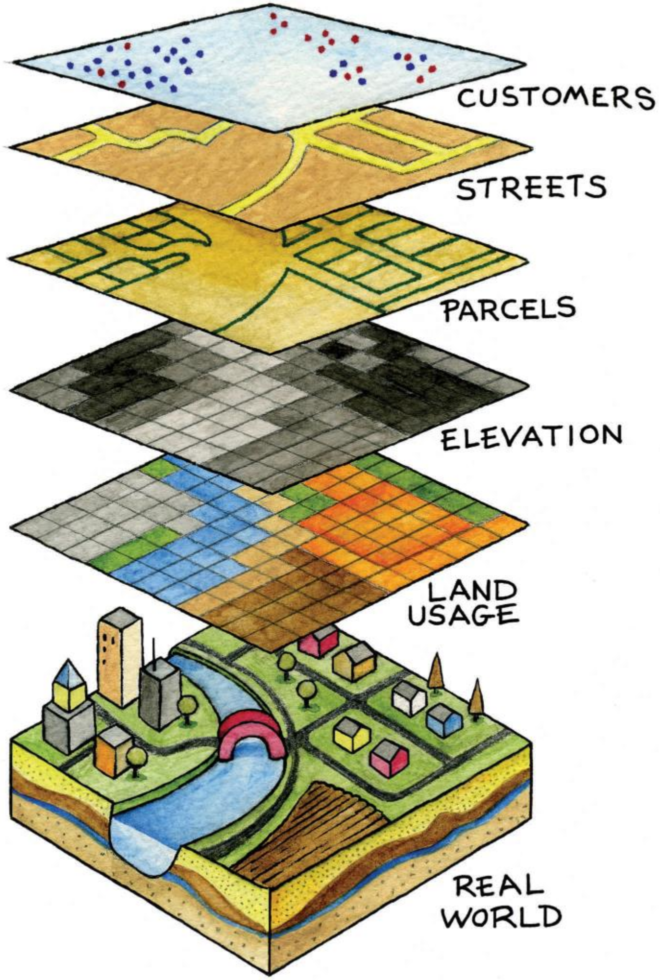
Oncogenes

Tumour suppressors

ACSN: Towards “Geographic” information system (GIS) for molecular biology

- Duality: an object exists in geographical space and in data space at the same time





From https://www.cachecounty.org/gis/

Principles of multi-layer omics data visualization in ACSN and NaviCell

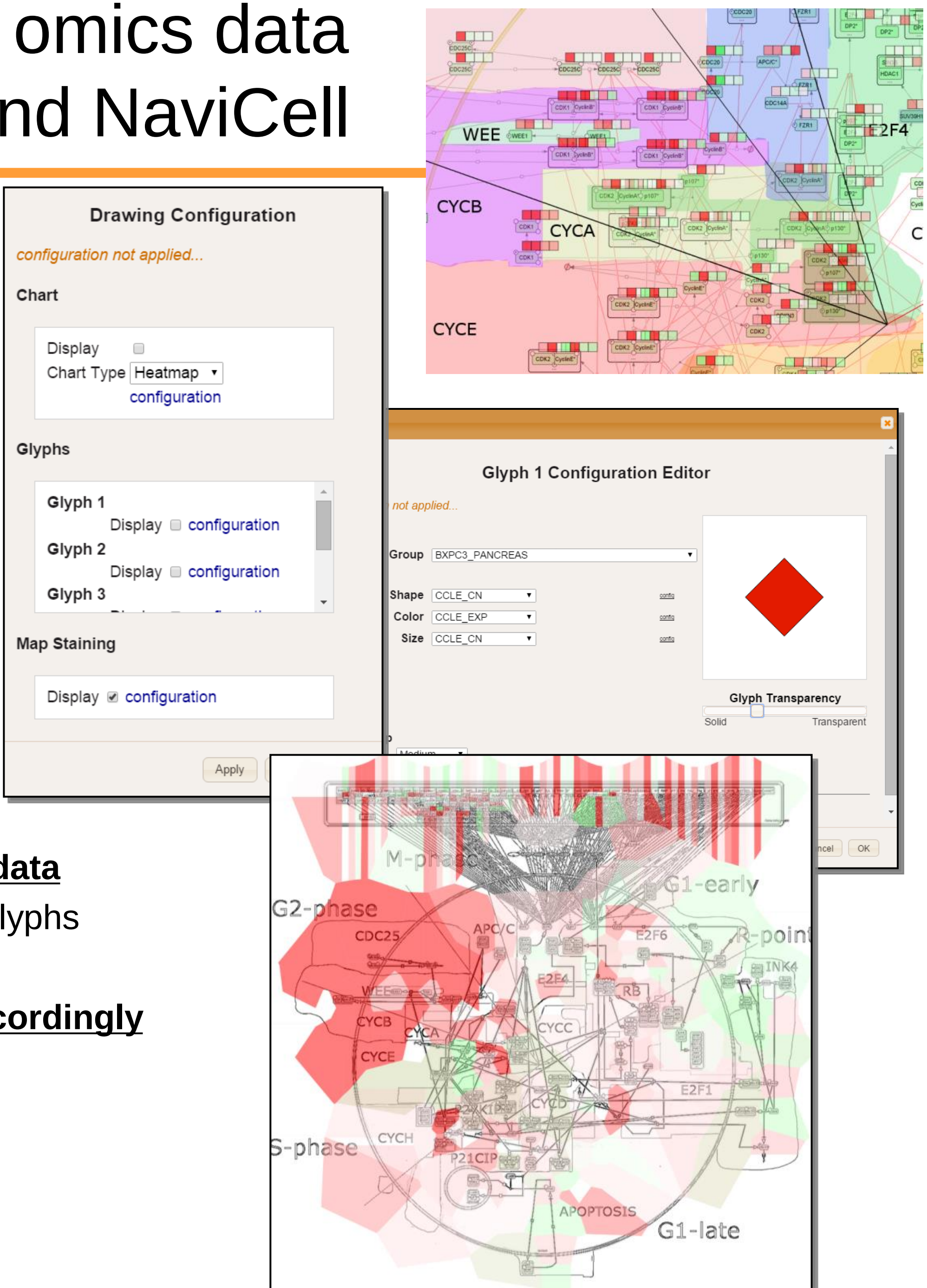
- **<u>Intrinsic data types</u>**
  - lists of objects
  - numerical object/sample tables (discrete and continuous)
  - categorical tables object/sample tables
  - sample annotation

- **<u>Data visualization channels</u>**
  - Charts: Barplots and heatmaps
  - Glyphs
  - Using map background (map staining)

- **<u>Assigning a data channel to specific omics data</u>**

  i.e. Expression in map staining, mutations in glyphs

- **<u>Aggregating measurements into groups accordingly to sample annotations (i.e. clinical data)</u>**



Drawing Configuration

configuration not applied...

Chart

Display ☐
Chart Type: Heatmap
configuration

Glyphs

Glyph 1 Display ☐ configuration
Glyph 2 Display ☐ configuration
Glyph 3

Map Staining

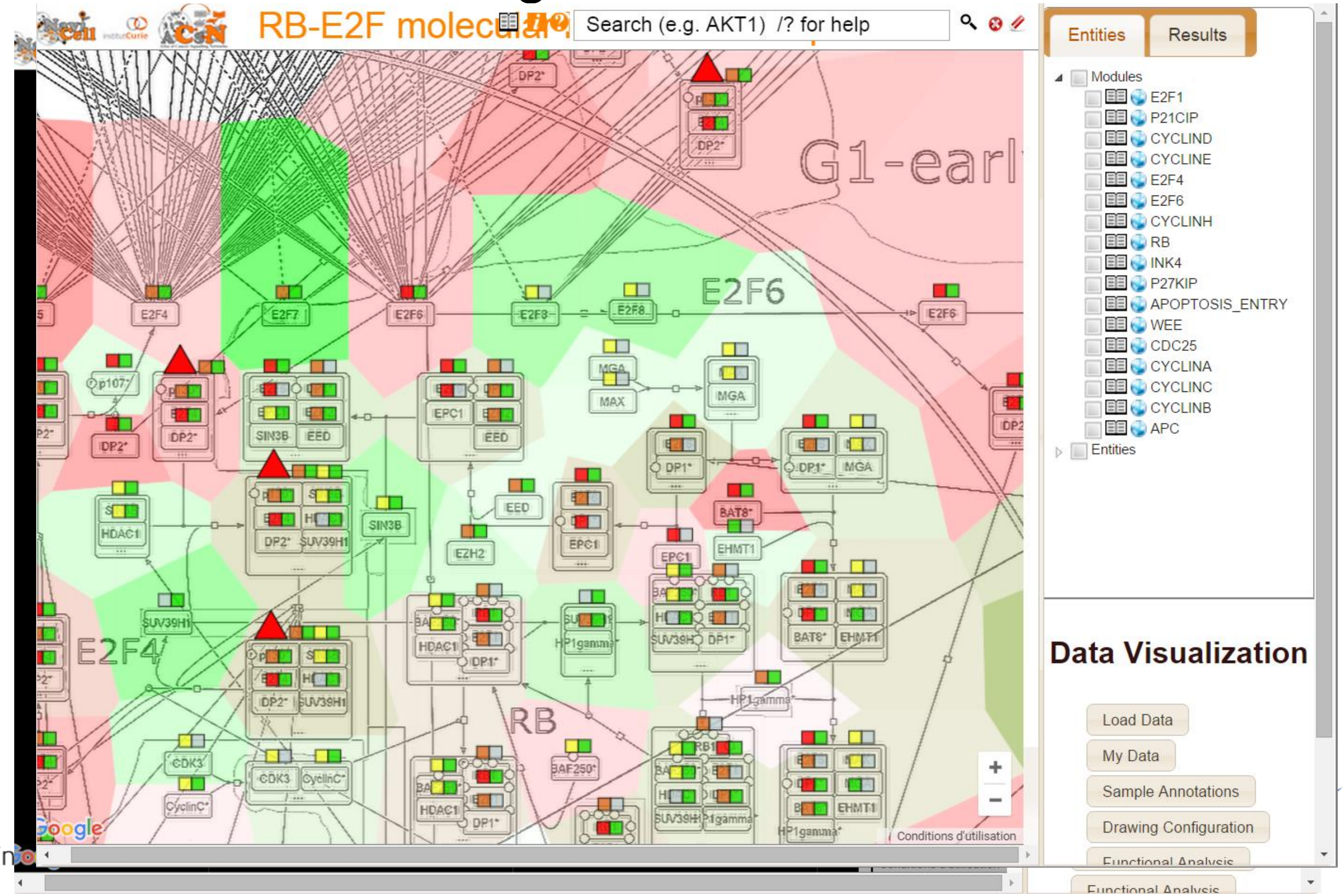
Display ☑ configuration

Glyph 1 Configuration Editor

not applied...

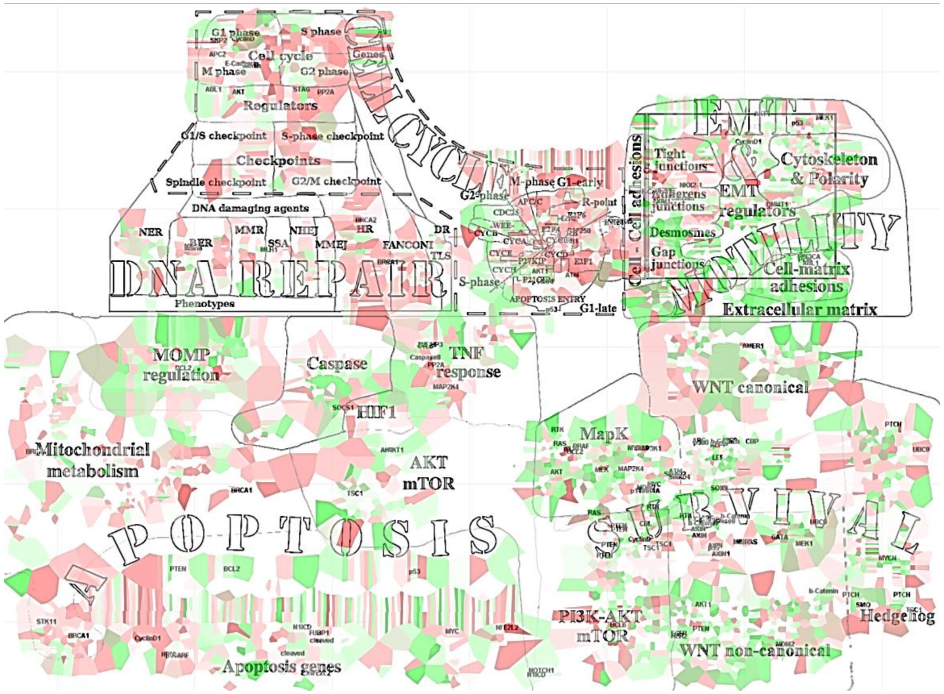
Group: BXPC3_PANCREAS
Shape: CCLE_CN
Color: CCLE_EXP
Size: CCLE_CN

Complex interactive data visualization scenes in NaviCell: combining data visualization channels

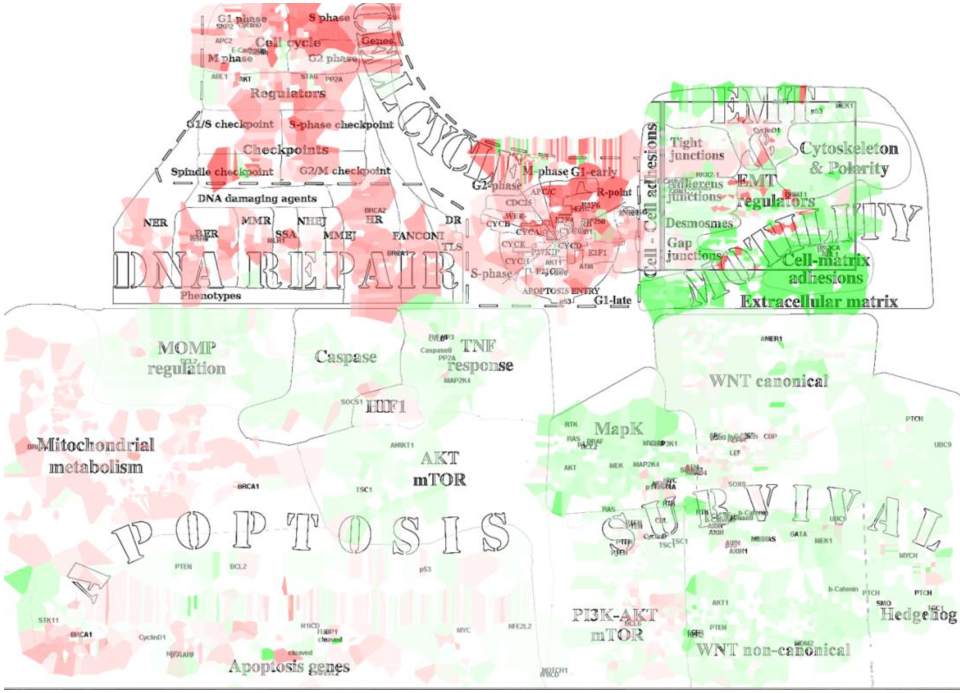


Pathway-based data abstraction in ACSN

“network smoothing” of the data



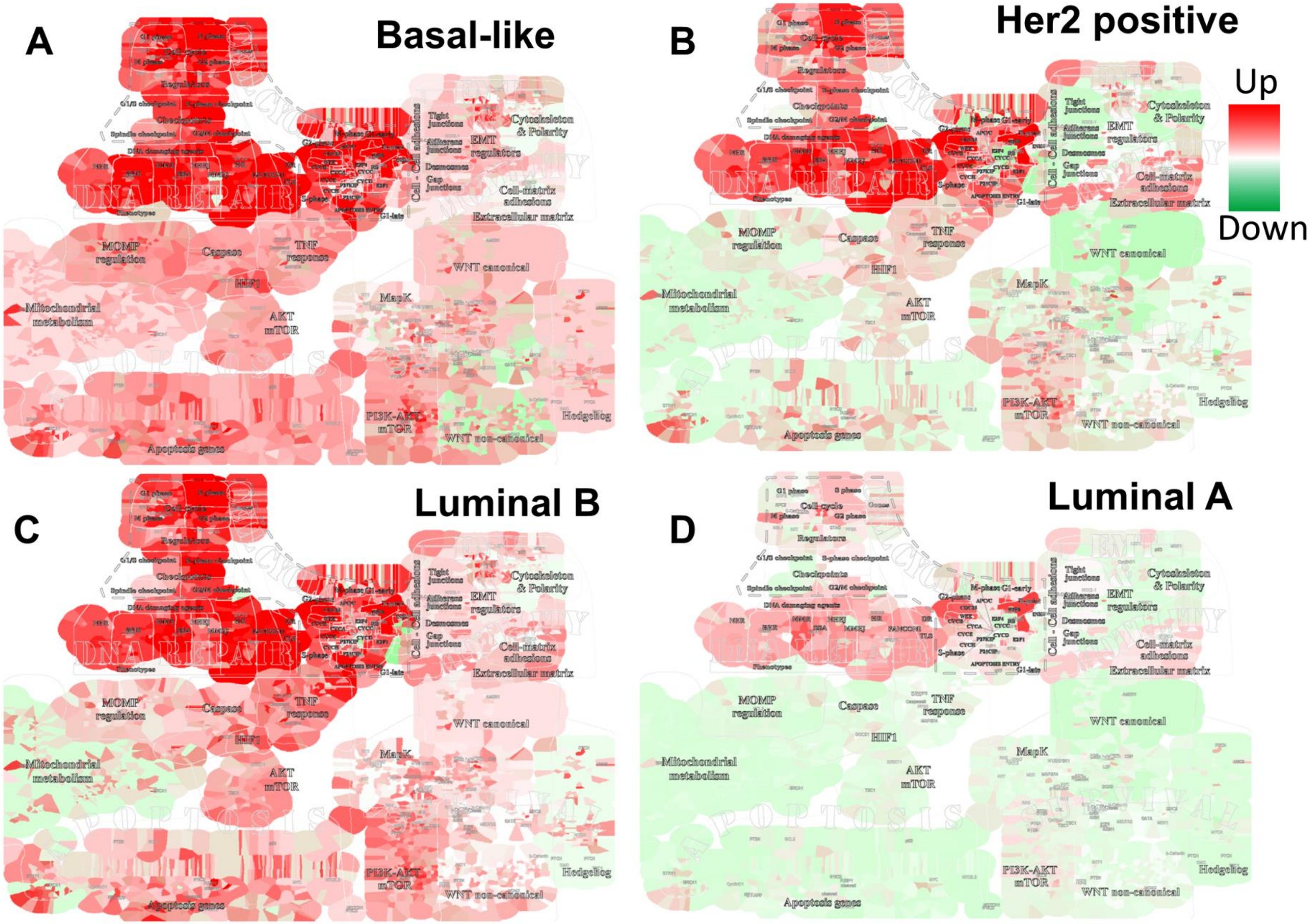
protein-based map staining



module-based map staining

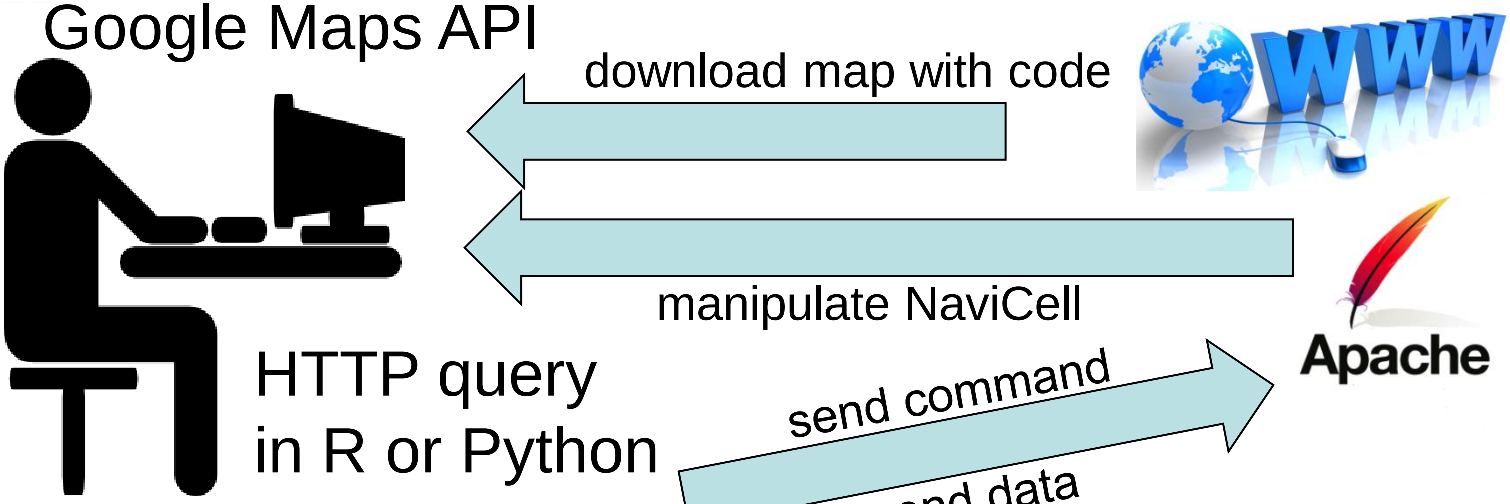
Data visualization with NaviCell and ACSN

map staining data visualization technique



NaviCell and ACSN as a web-server with API

NaviCell = Browser+JavaScript+ Google Maps API



download map with code

manipulate NaviCell

HTTP query in R or Python or Java

send command

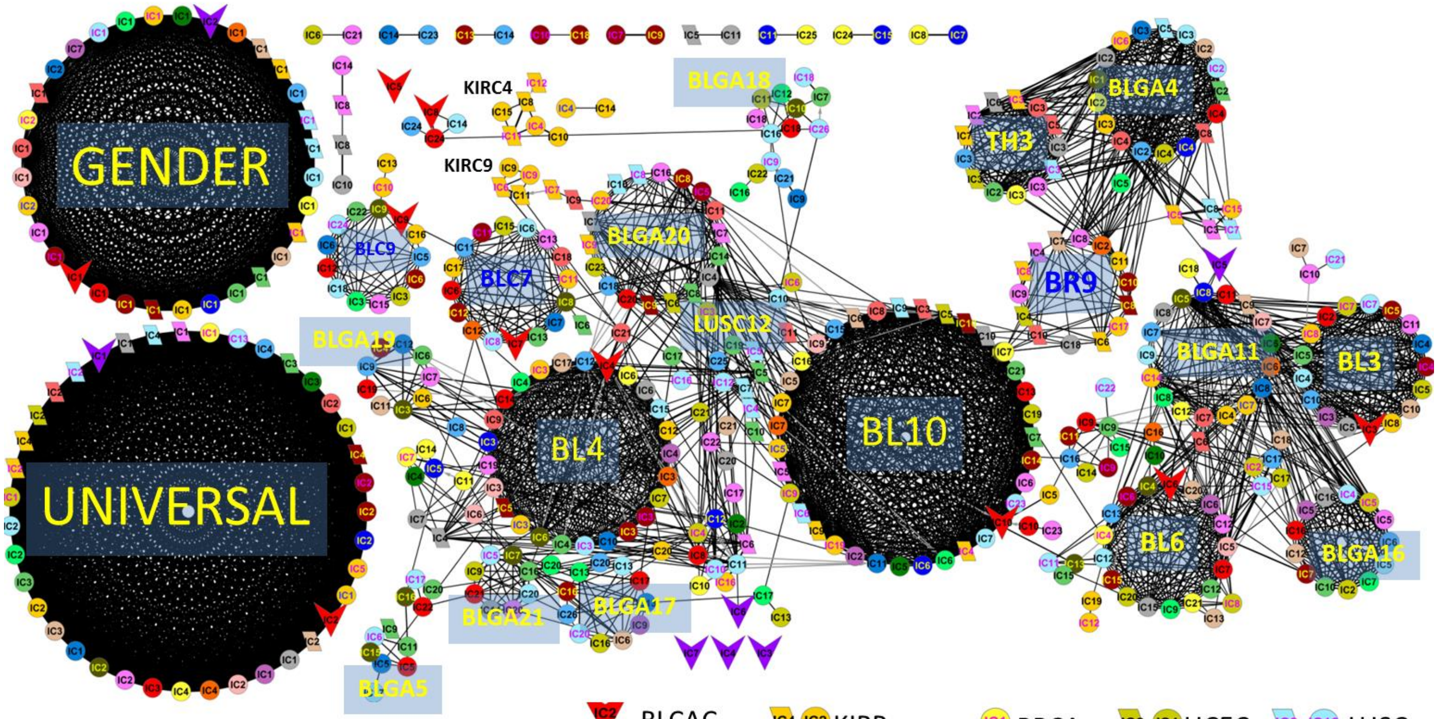
send data

NaviCell beyond ACSN

http://navicell.curie.fr

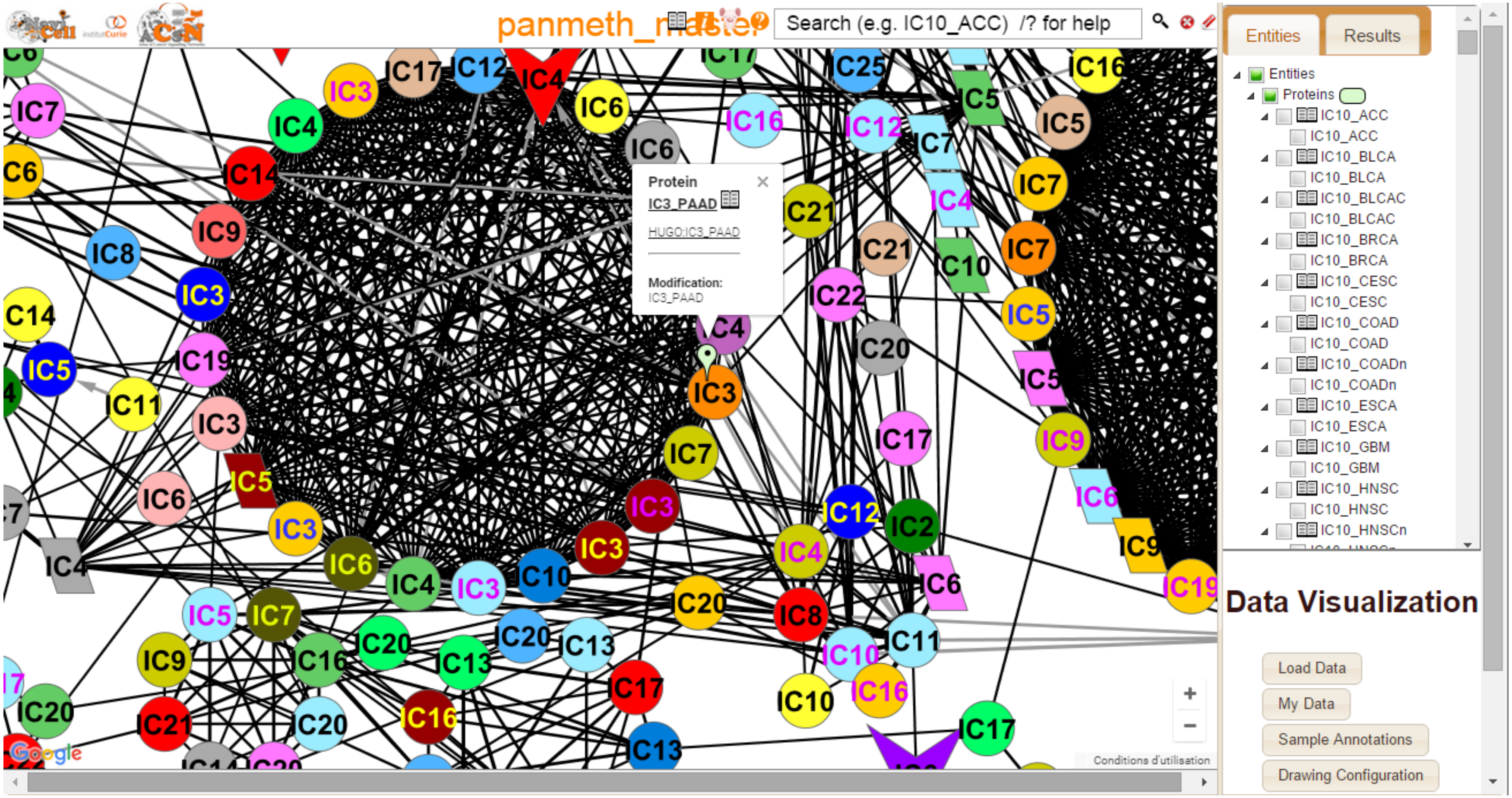
- NaviCell can be used to
  - Browse any CellDesigner map not from ACSN
  - Browse any non-CellDesigner network map (i.e. any biological network that can be imported to Cytoscape)
  - Browse any graph, not necessary biological network
- The benefits are
  - Web-based environment, clickable interactive content
  - All functionality of NaviCell (semantic zoom, search, highlighting, data visualization toolbox)

Pan-cancer methylome deconvolution by Independent Component Analysis



8181 tumor samples
745 normal tissues
27 cancer types

NaviCell for (1) **interactive** (2) **programmable** (3) **equipped with a discussion forum visualization** of any large graph

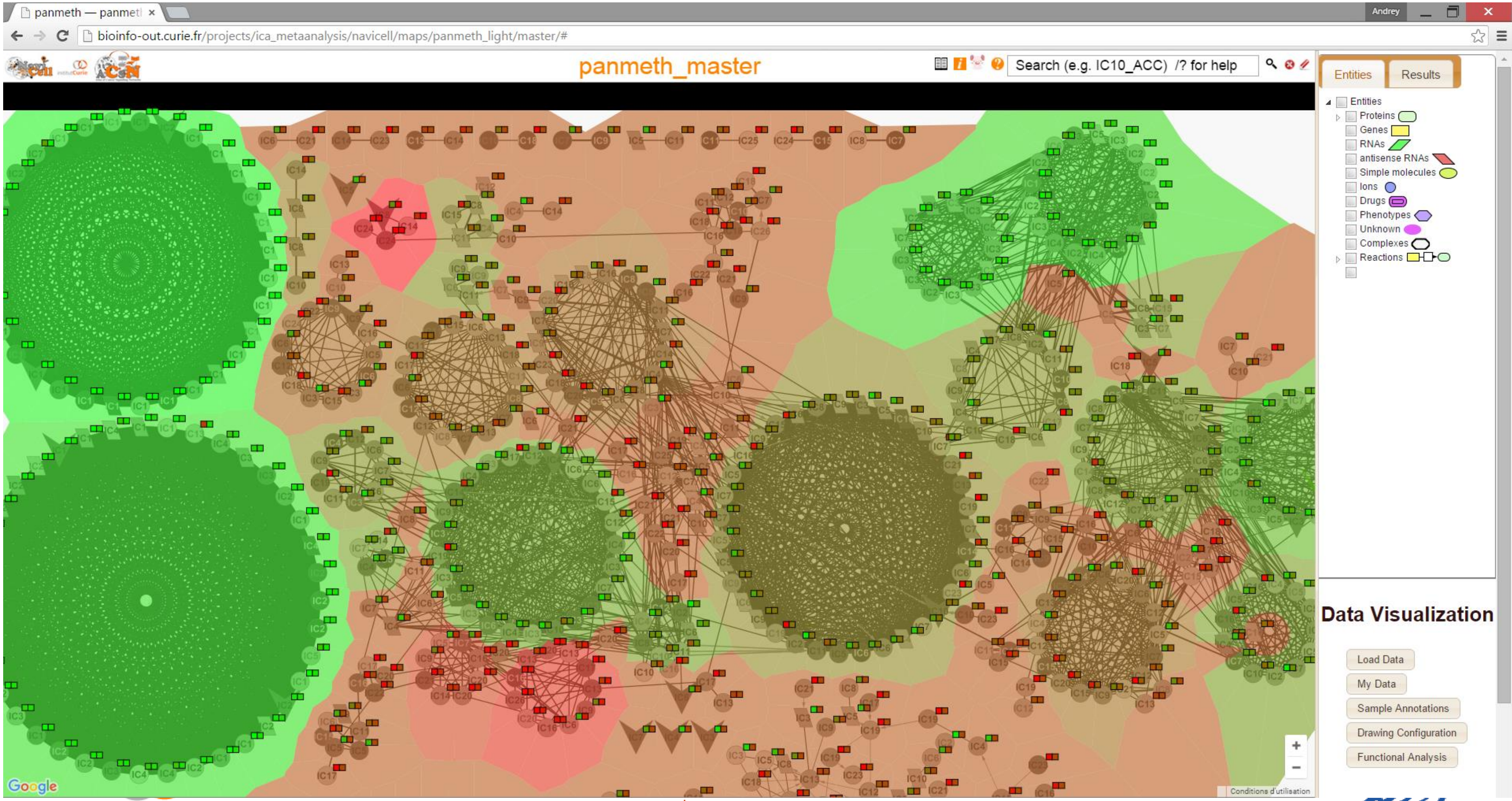


institutCurie

Inserm
Institut national
de la santé et de la recherche médicale

MINES PARIS

NaviCell for (1) **interactive** (2) **programmable** (3) **equipped with a discussion forum visualization** of any large graph



institutCurie

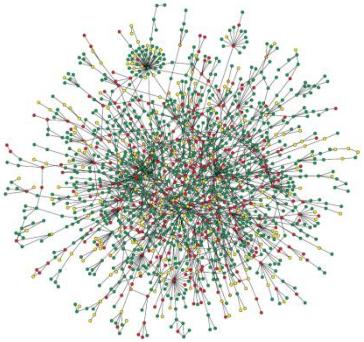
Institut national de la santé et de la recherche médicale

MINES PARIS

Network and data: two major problem formulations

- “Field of significance” defined from data

(defined on nodes or edges)



1) Detecting “significant” subnetworks (modules)

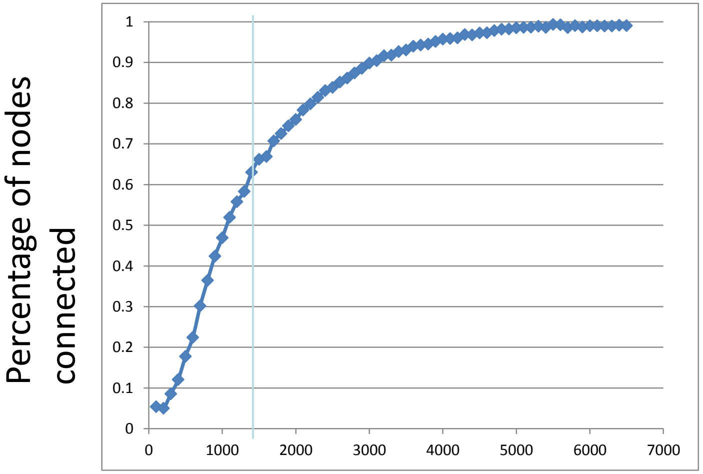
  - find a subgraph which nodes satisfy certain conditions resulting from data analysis (eg, maximum proportion of nodes having large changes between two conditions)

2) Network propagation

  - transform the data such that “significant” nodes would “spread” significance to more or less distant graph neighbours

Detecting “significant” subnetworks (modules)

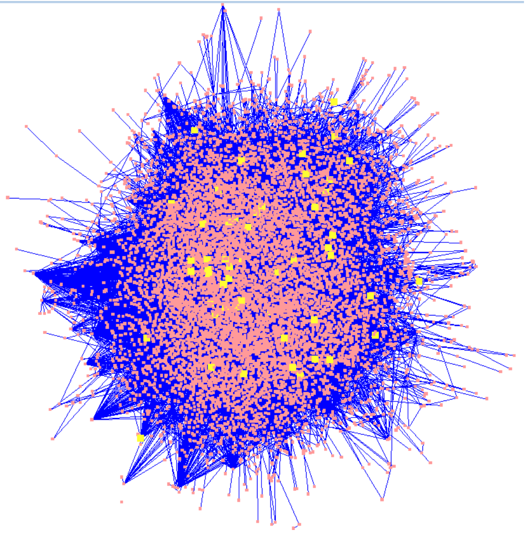
- Methods
  - combinatorial, optimisation, greedy
- Simple “percolation” model (Kairov et al, 2012)



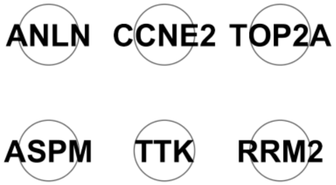
TOP2A	6.68
CCNE1	6.52
ASPM	6.5
ANLN	6.14
RRM2	6.1
MGC57827	6.04
CDC2	5.87
CCNE2	5.81
DLG7	5.71
NUSAP1	5.7
TTK	5.66
RRM2	5.65
KIF11	5.56
CDCA1	5.56
CDC2	5.51
TPX2	5.49
C10orf3	5.43
TOPK	5.35
FKSG14	5.35
UHRF1	5.29
RAMP	5.27
DEPDC1	5.27
MELK	5.26
ANLN	5.25
…	
SCT	0.000683
FLJ11236	0.000642
ITPR1	0.000588
KHDRBS3	0.000409
COL22A1	0.000128
DOCK2	0.000102

Ranking of the genes accordingly to an independent component score

20.000 genes

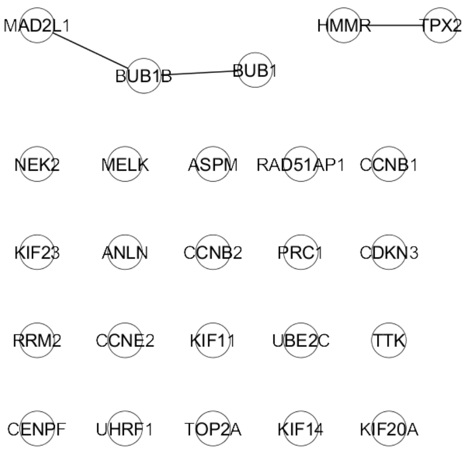


Take the top 10 genes



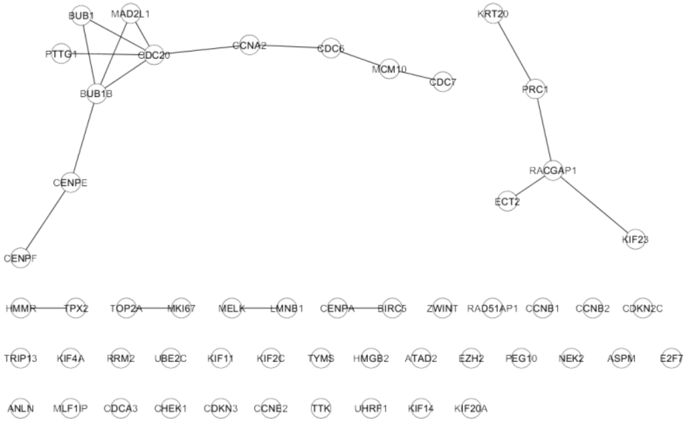
Size of connected component = 0

Take the top 50 genes



Size of connected component = 3

Take the top 100 genes

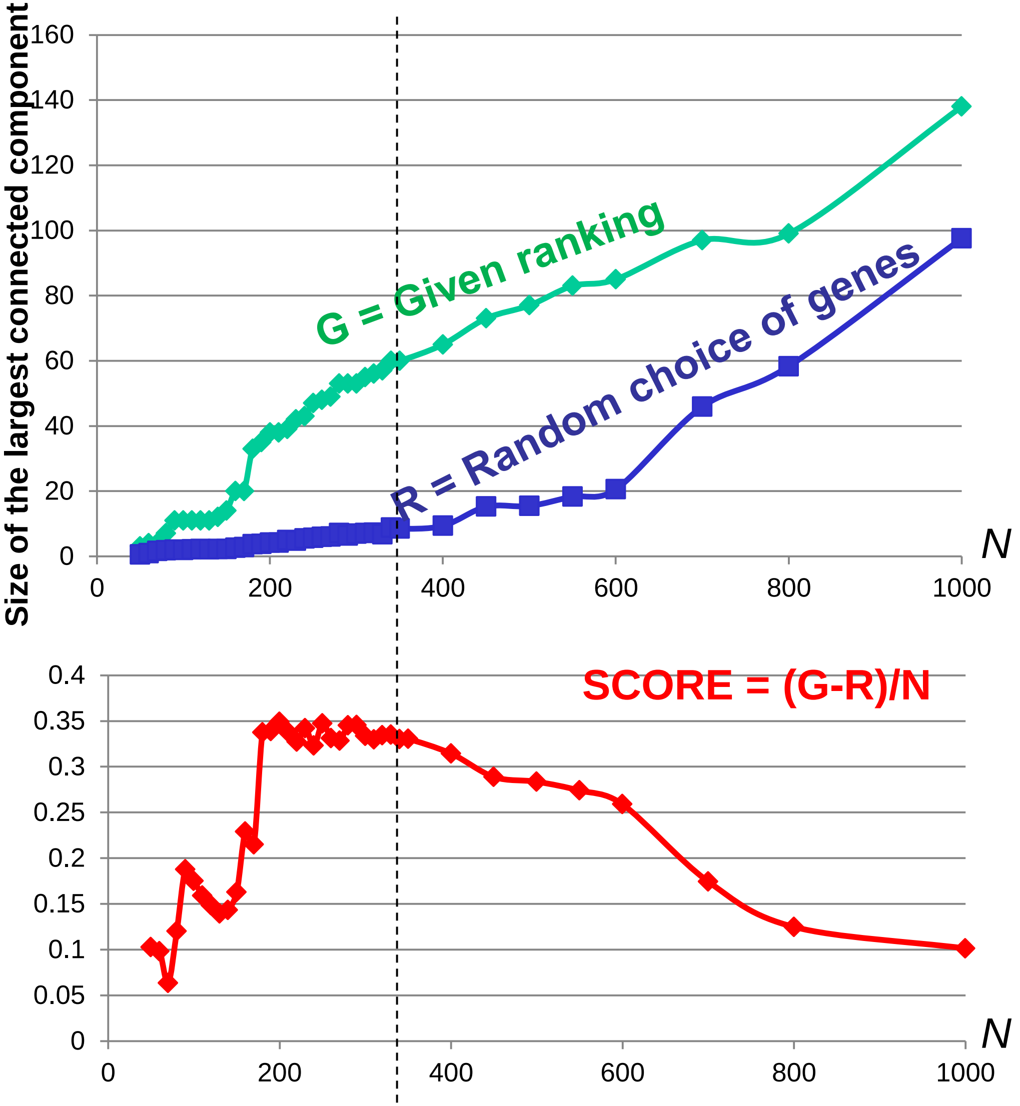


Size of connected component = 11

TOP2A	6.68
CCNE1	6.52
ASPM	6.5
ANLN	6.14
RRM2	6.1
MGC57827	6.04
CDC2	5.87
CCNE2	5.81
DLG7	5.71
NUSAP1	5.7
TTK	5.66
RRM2	5.65
KIF11	5.56
CDCA1	5.56
CDC2	5.51
TPX2	5.49
C10orf3	5.43
TOPK	5.35
FKSG14	5.35
UHRF1	5.29
RAMP	5.27
DEPDC1	5.27
MELK	5.26
ANLN	5.25
…	
SCT	0.000683
FLJ11236	0.000642
ITPR1	0.000588
KHDRBS3	0.000409
COL22A1	0.000128
DOCK2	0.000102

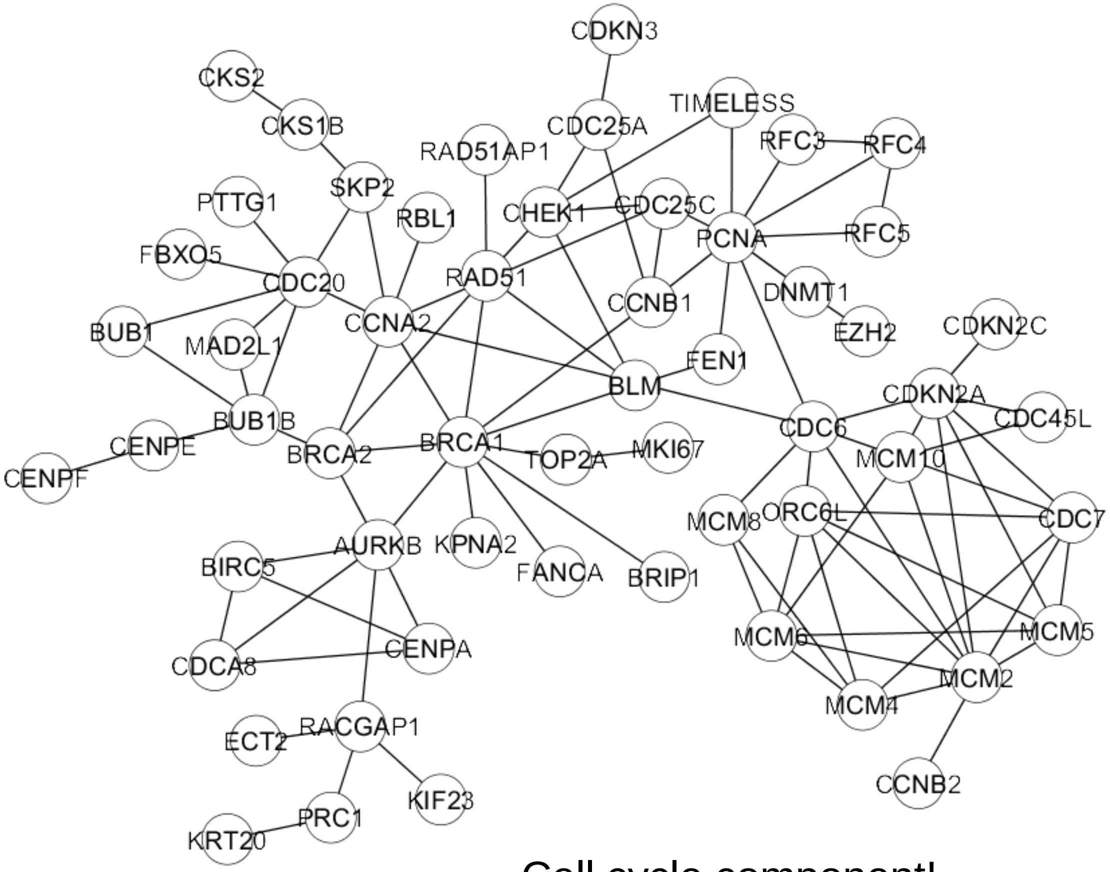
Ranking of the genes accordingly to a score

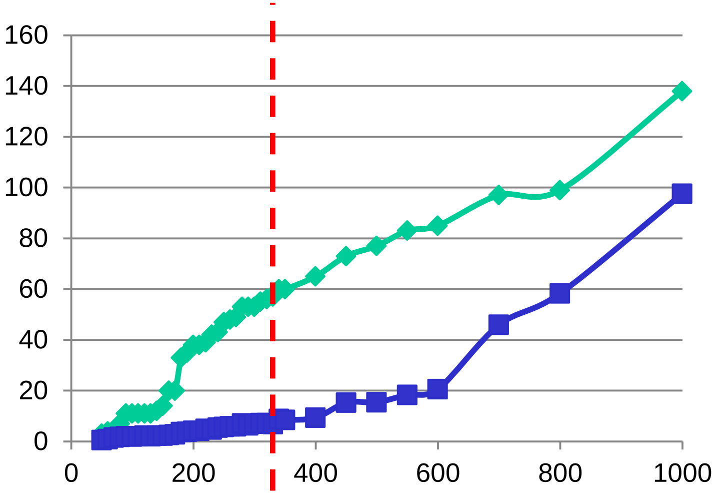
20.000 genes



330 genes

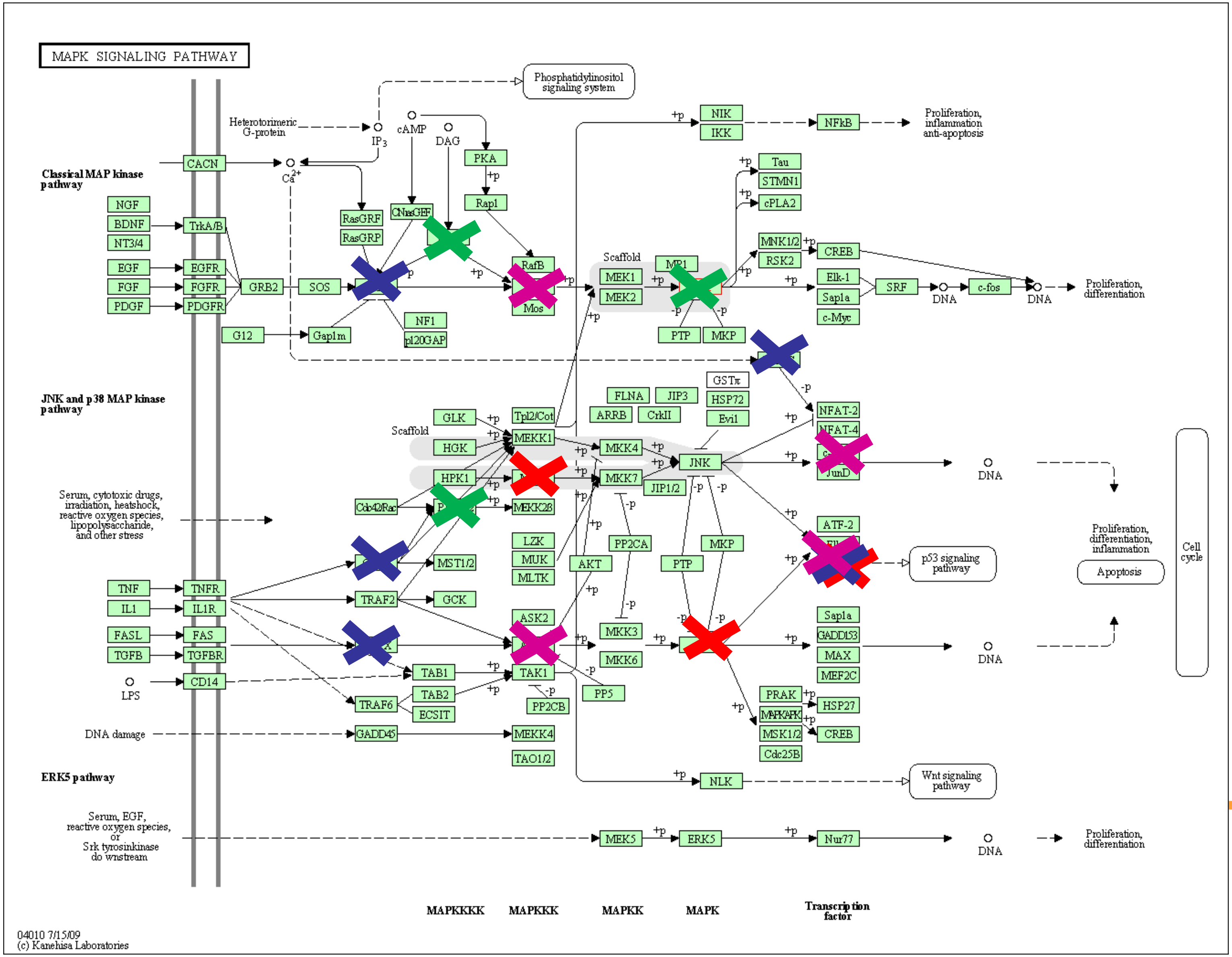
Optimally functionaly enriched network (OFTEN)





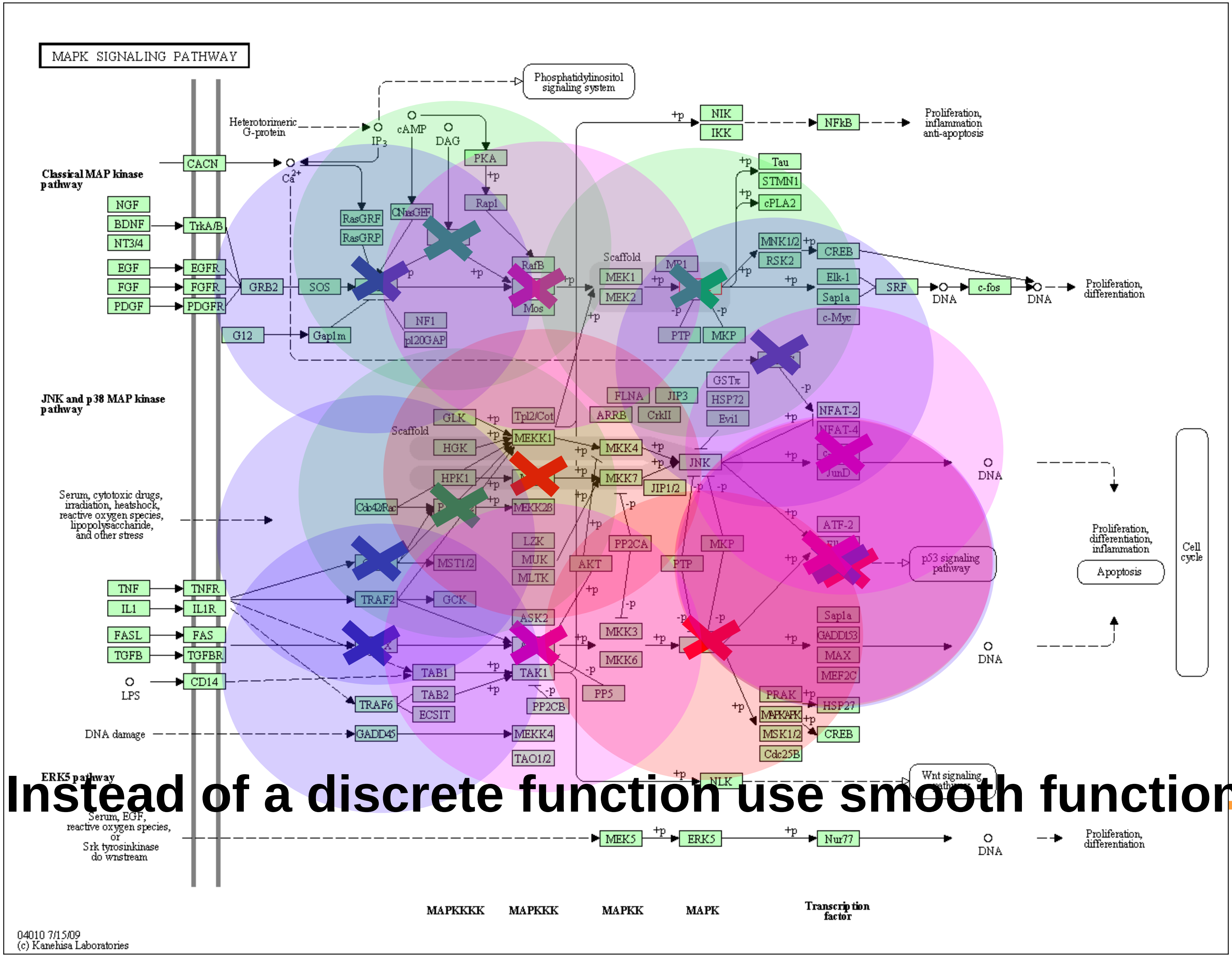
Cell cycle component!

Effect of mutation on a «pathway»: Anna Karenina’s principle



Patient 1
Patient 2
Patient 3
Patient 4

Network propagation of data



Patient 1

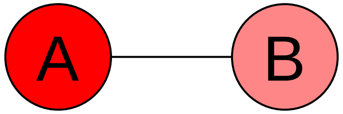
Patient 2

Patient 3

Patient 4

Instead of a discrete function use smooth function on a graph

Graph Laplacian and dosage balance in interaction networks

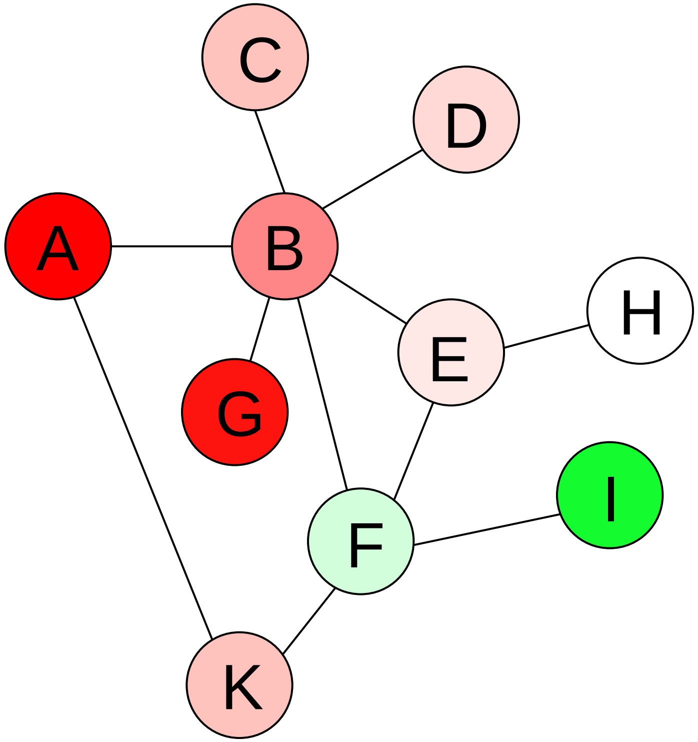


Assumption: Interacting molecular entities A and B should be balanced in their concentrations (local amounts in space and time)

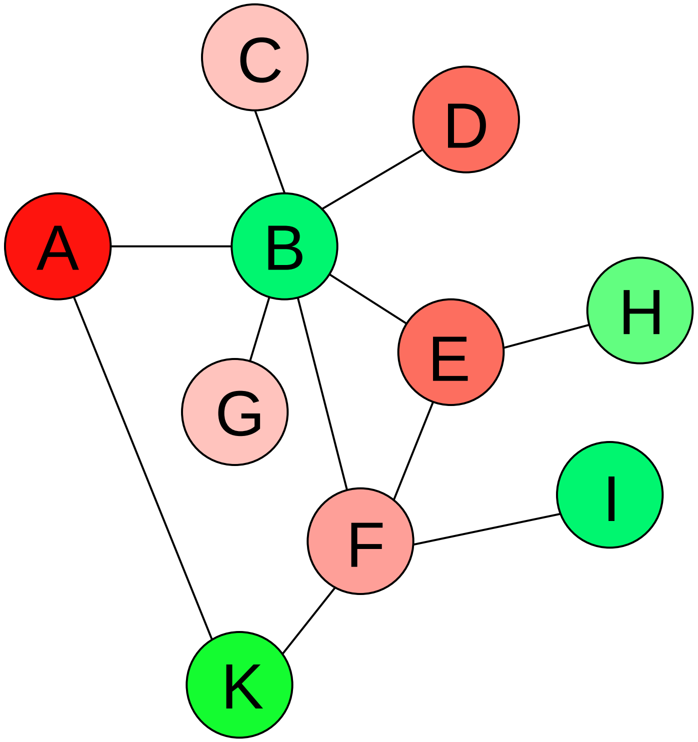
Examples from molecular biology:

1) A and B form complex, A or B along is toxic
2) A and B form functional complex, production of A or B is expensive
3) A is a scaffold for B and C, complex of A:B:C performs a function
4) A regulates B (catalyzes, titrates, ..)
5) A and B compete for a resource, this competition is decisive for a cell fate

Graph Laplacian and dosage balance in interaction networks



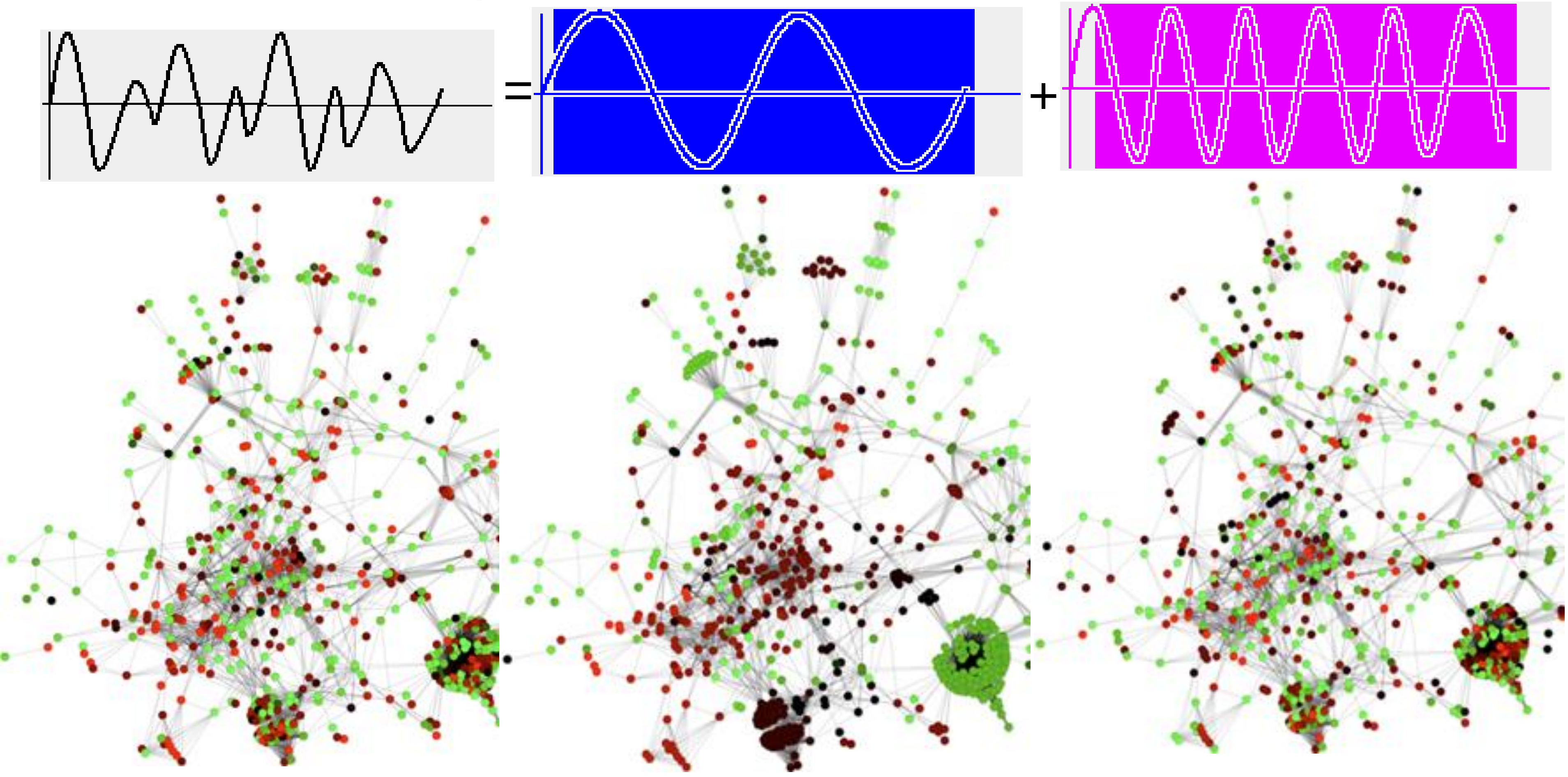
Smooth distribution,
balanced dosage



Non-smooth distribution,
unbalanced dosage

Spectral graph analysis (Fourier transformation on graphs)

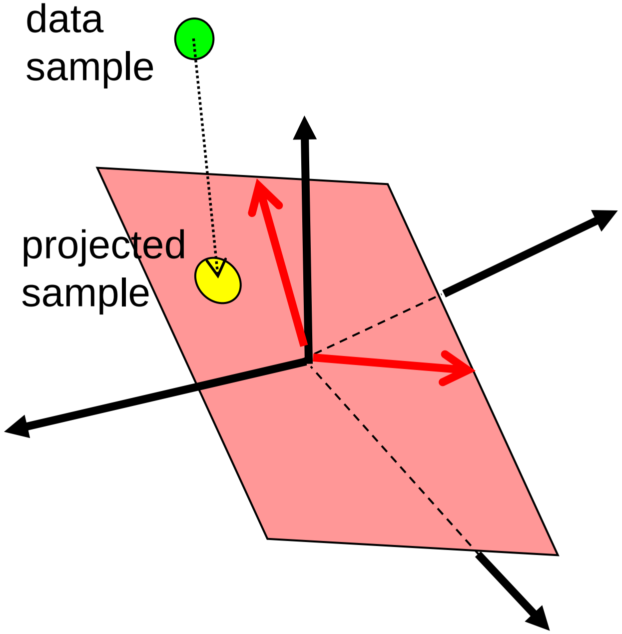
Function on graph



Slow, smooth component

Fast, high frequency component

The data are projected into the subspace of “distributions smooth on a biological graph”



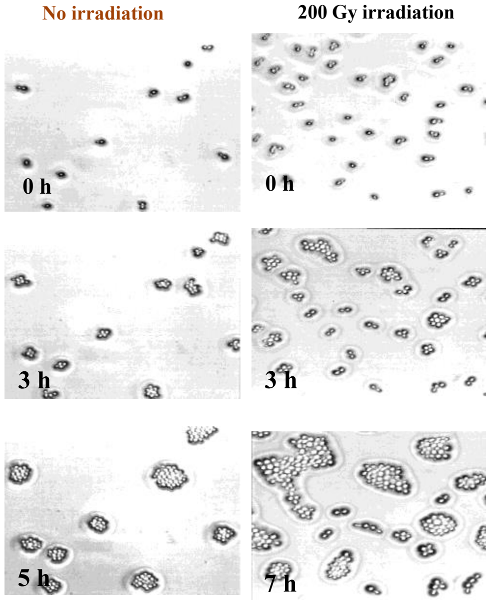
initial space R^{20000}

subspace of functions smooth on a gene network

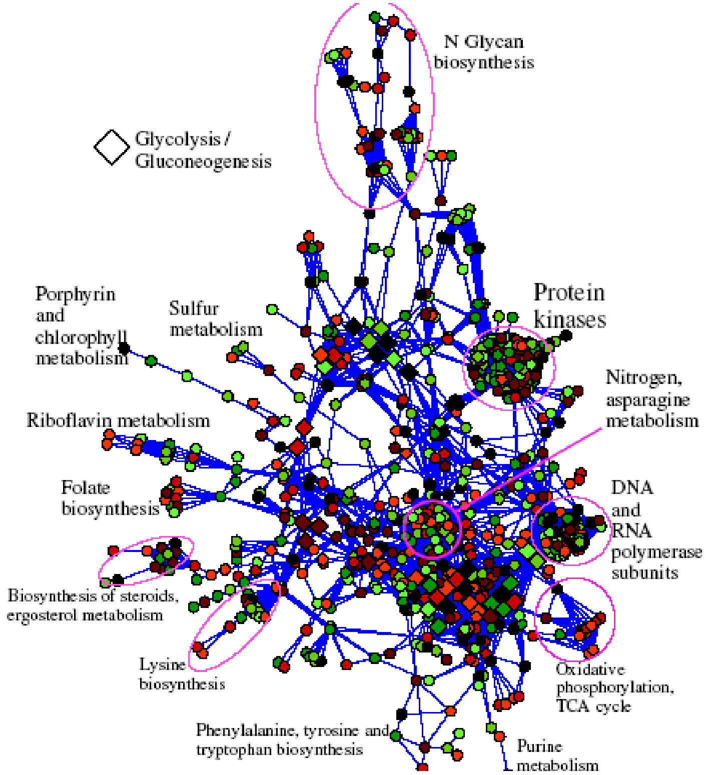
basis vectors are eigenvectors of the graph Laplacian

Classifier smooth on biological graph

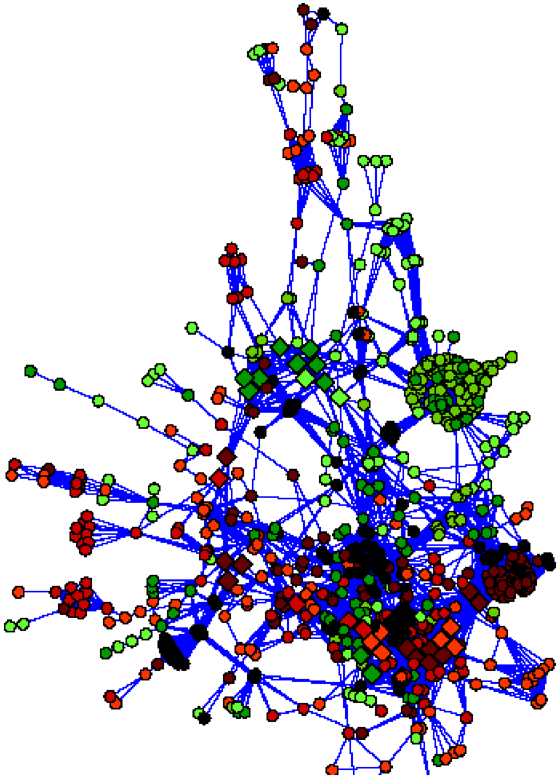
Rapaport et al., BMC Bioinformatics, 2007



Data from
Marie Dutreix
(Institut Curie)



“Classical” SVM



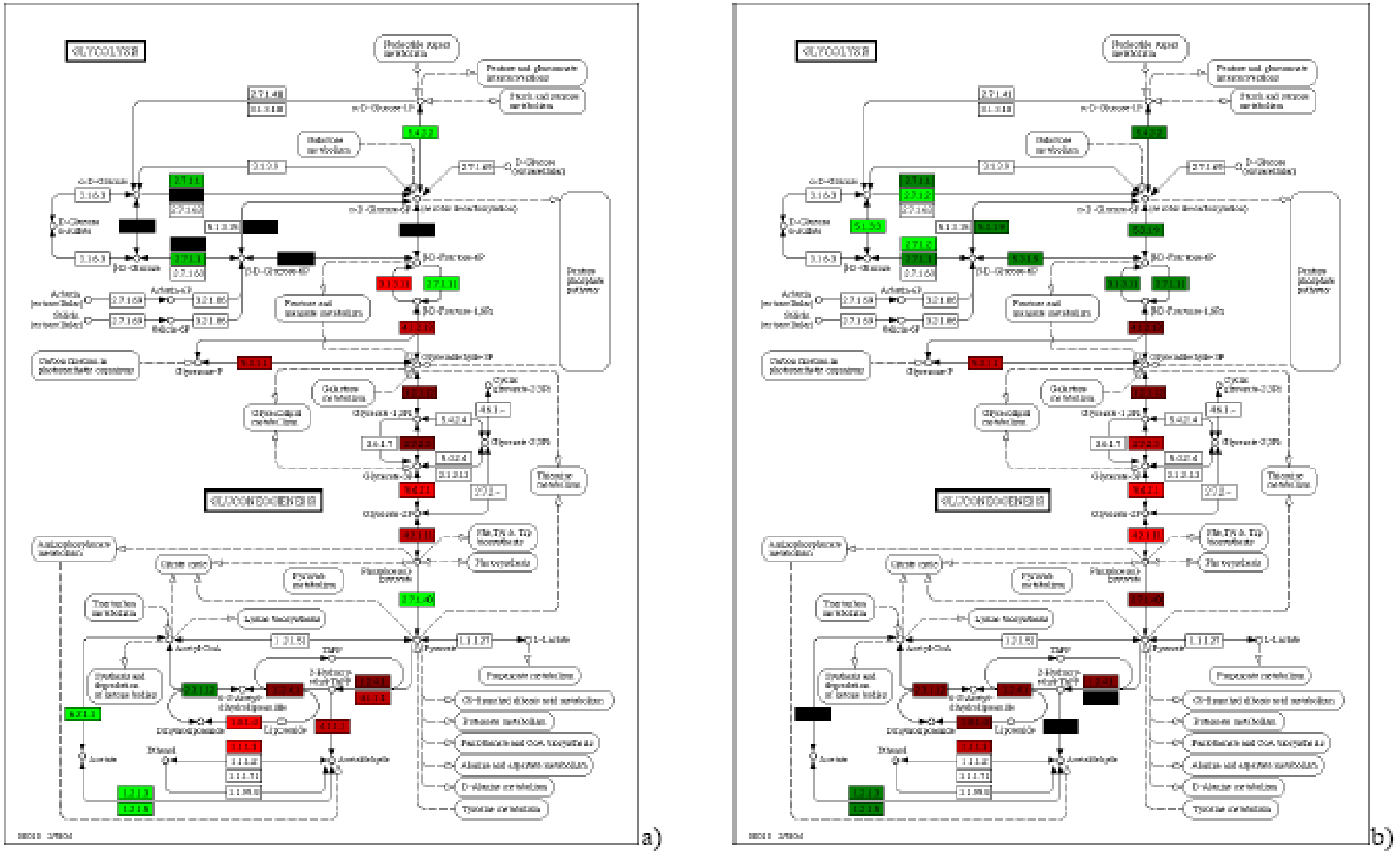
SVM done in the reduced subspace of smooth functions (first 20% of Laplacian eigenvalues)

institutCurie

Inserm
Institut national
de la santé et de la recherche médicale

MINES PARIS

Glycolysis/Gluconeogenesis switch following yeast irradiation



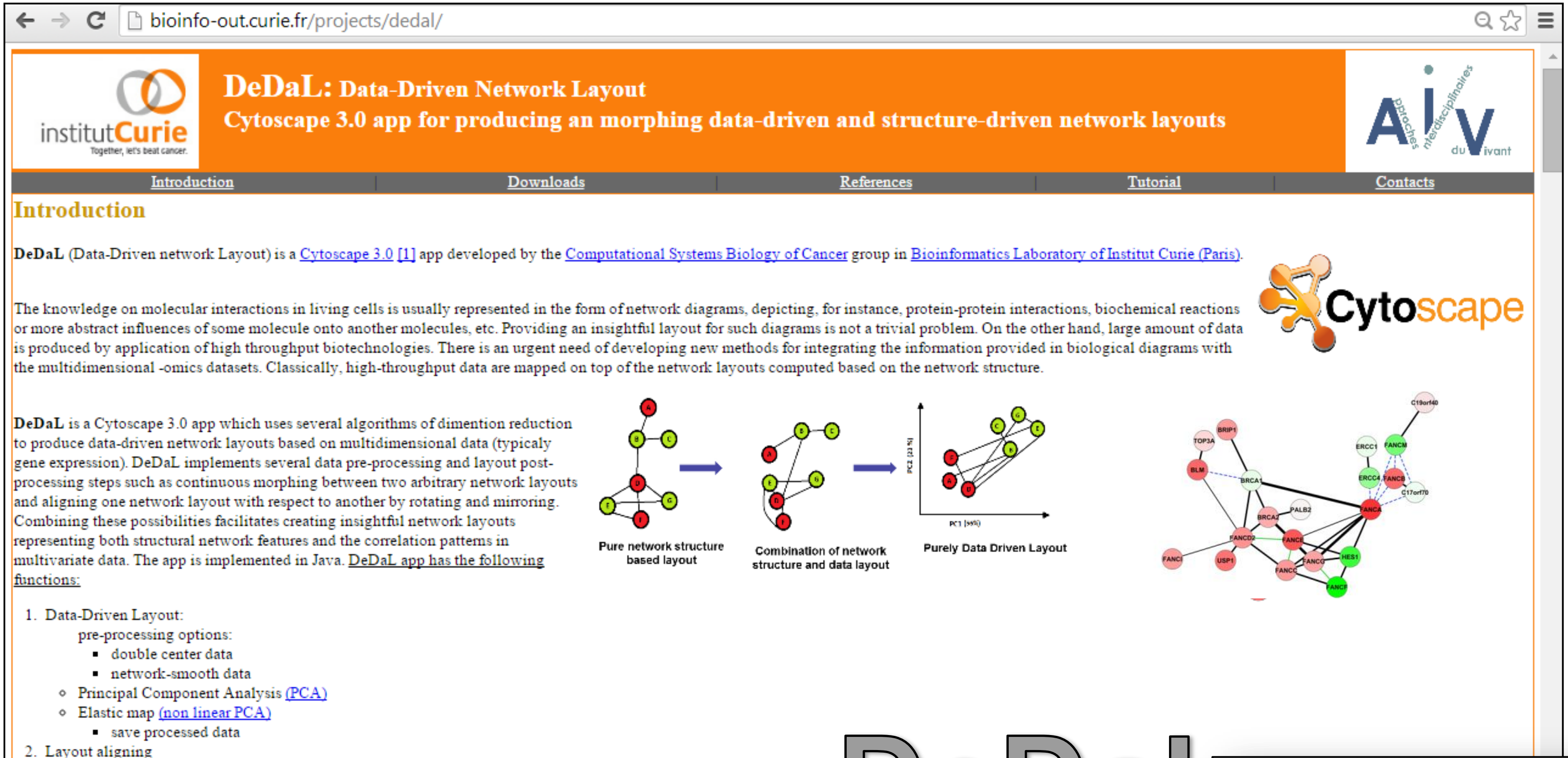
a)

b)

institutCurie

Inserm
Institut national
de la santé et de la recherche médicale

MINES PARIS



DeDaL: Data-Driven Network Layout

Cytoscape 3.0 app for producing an morphing data-driven and structure-driven network layouts

Introduction

DeDaL (Data-Driven network Layout) is a Cytoscape 3.0 [1] app developed by the Computational Systems Biology of Cancer group in Bioinformatics Laboratory of Institut Curie (Paris).

The knowledge on molecular interactions in living cells is usually represented in the form of network diagrams, depicting, for instance, protein-protein interactions, biochemical reactions or more abstract influences of some molecule onto another molecules, etc. Providing an insightful layout for such diagrams is not a trivial problem. On the other hand, large amount of data is produced by application of high throughput biotechnologies. There is an urgent need of developing new methods for integrating the information provided in biological diagrams with the multidimensional -omics datasets. Classically, high-throughput data are mapped on top of the network layouts computed based on the network structure.

DeDaL is a Cytoscape 3.0 app which uses several algorithms of dimention reduction to produce data-driven network layouts based on multidimensional data (typicaly gene expression). DeDaL implements several data pre-processing and layout post-processing steps such as continuous morphing between two arbitrary network layouts and aligning one network layout with respect to another by rotating and mirroring. Combining these possibilities facilitates creating insightful network layouts representing both structural network features and the correlation patterns in multivariate data. The app is implemented in Java. DeDaL app has the following functions:

1. Data-Driven Layout:
   pre-processing options:
     - double center data
     - network-smooth data
   - Principal Component Analysis (PCA)
   - Elastic map (non linear PCA)
     - save processed data
2. Layout aligning
3. Layout morphing

Downloads

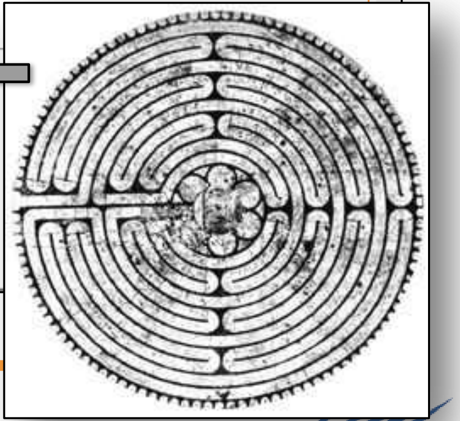
Install DeDaL test version

Download File DeDaL (7Mb) last release.

Using the Cytoscape app manager

1. Launch the Cytoscape app manager (menu "Apps -> App Manager-> Install from File . . . ").

DeDaL



institutCurie

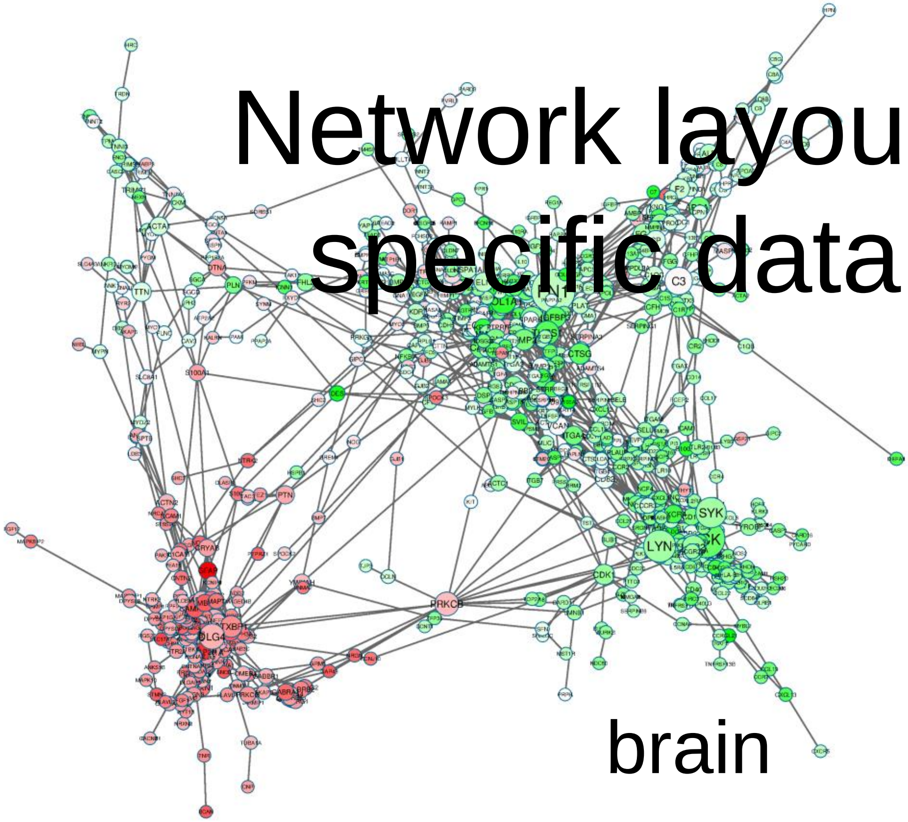
Inserm
Institut national
de la santé et de la recherche médicale

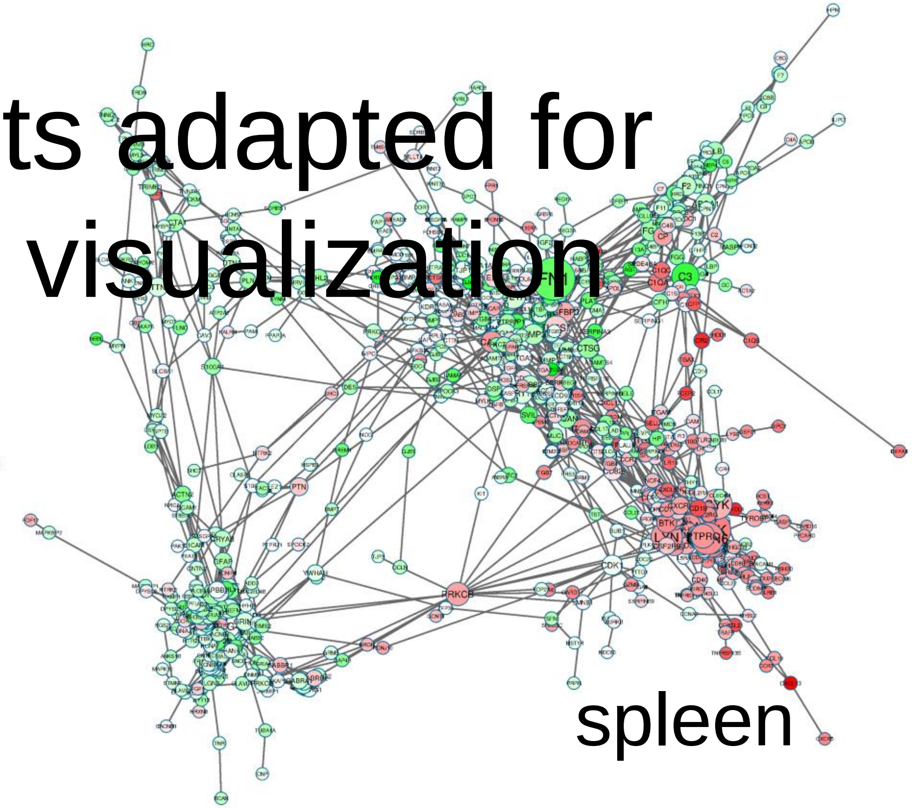
MINES PARIS

Healthy tissue-specific subnetwork of HPRD
(28 healthy tissues, RNASeq data)

application of network smoothing+non-linear PCA

Network layouts adapted for specific data visualization


brain


spleen

(from Czerwinska et al, BMC Sys Bio, 2015)

institut Curie

Inserm
Institut national
de la santé et de la recherche médicale

MINES PARIS

Disconnecting complex graphs

SECRET RM-1573 10-24-55 -37-

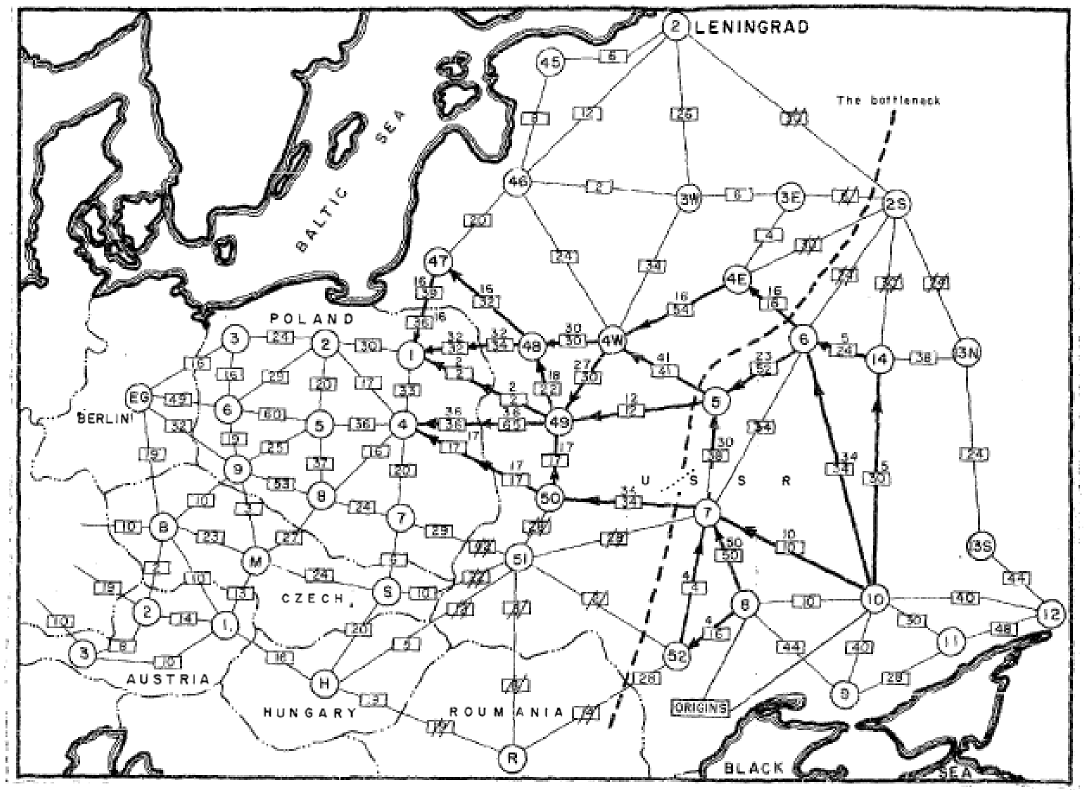


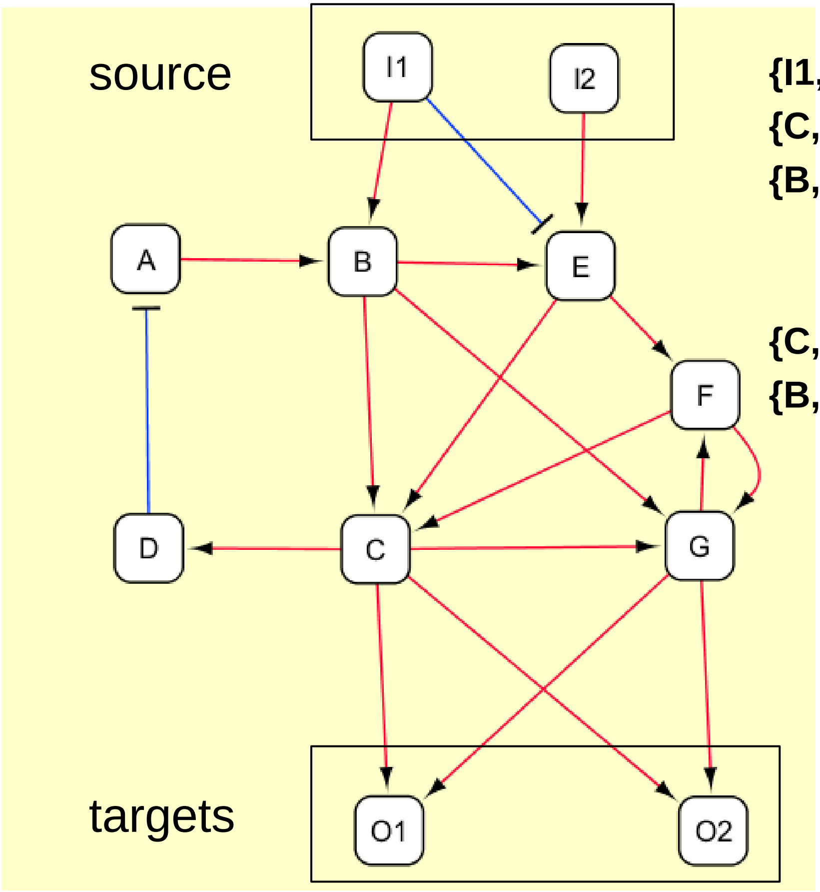
Fig. 9 — Traffic pattern: certain divisions inoperative

Transversal Hypergraph Generation

Minimal Hit Set Problem

Harris TE, Ross FS. (1955) Fundamentals of a Method for Evaluating Rail Net Capacities. In Research Memorandum RM-1573, The RAND Corporation

Minimal cutting set problem for biological pathways



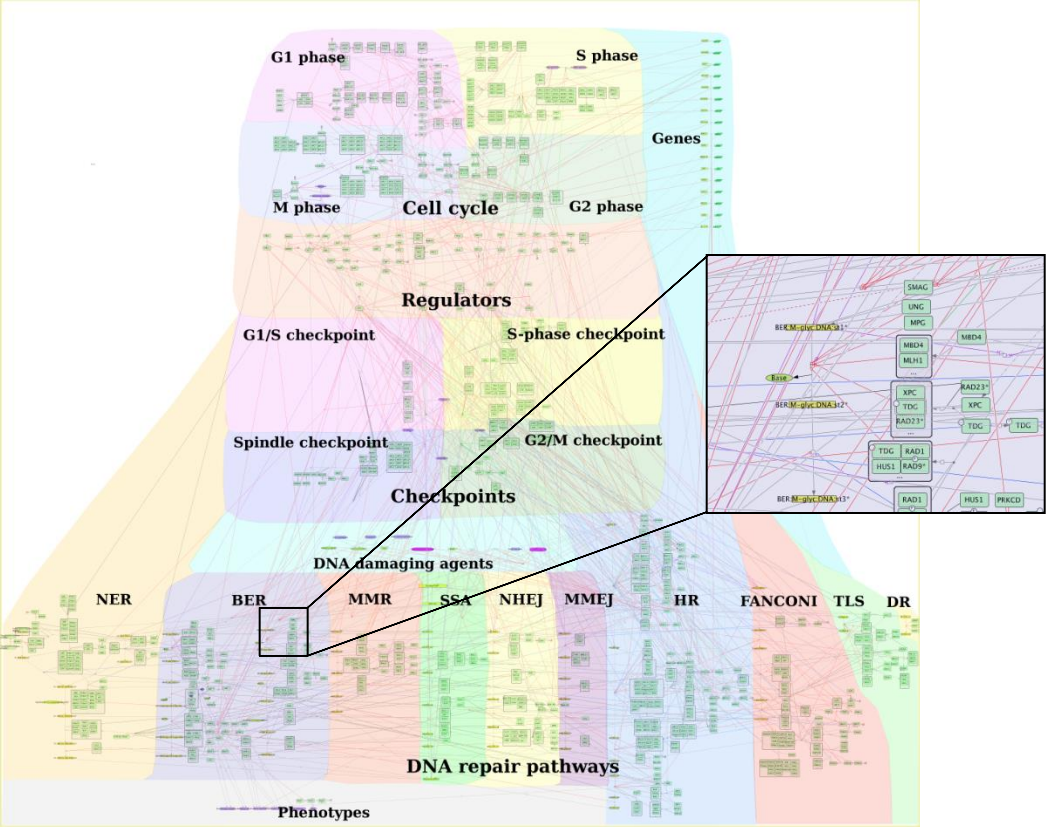
{I1, I2}
{C,G}
{B,E}

{C,F,I1}
{B,C,F}

Vera-Licona, P., et al. OCSANA: optimal combinations of interventions from network analysis. (2013). *Bioinformatics* **15**:29(12).

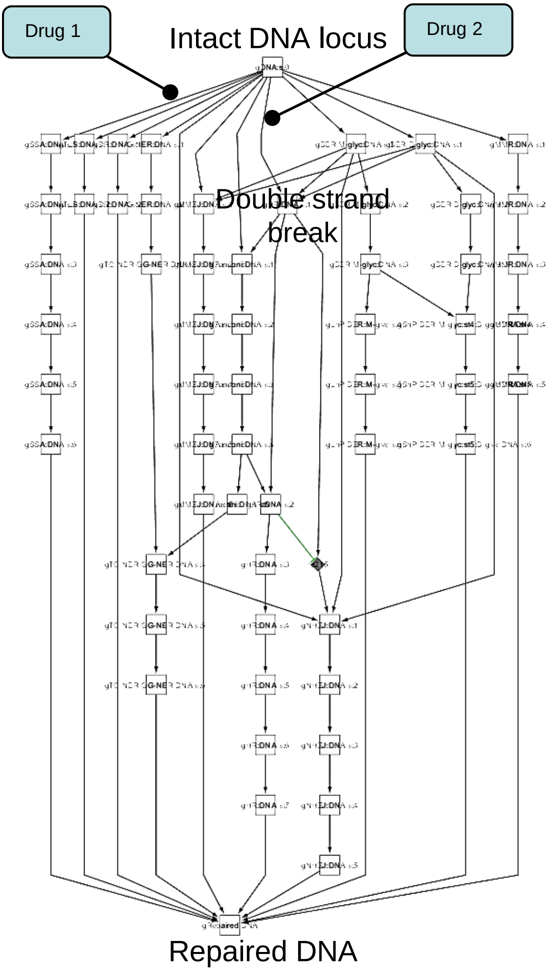
OCSANA: tool for structural analysis of LARGE biological networks

Predicting synthetic lethal interactions in DNA repair, dysregulated in various cancers (Inna Kuperstein)



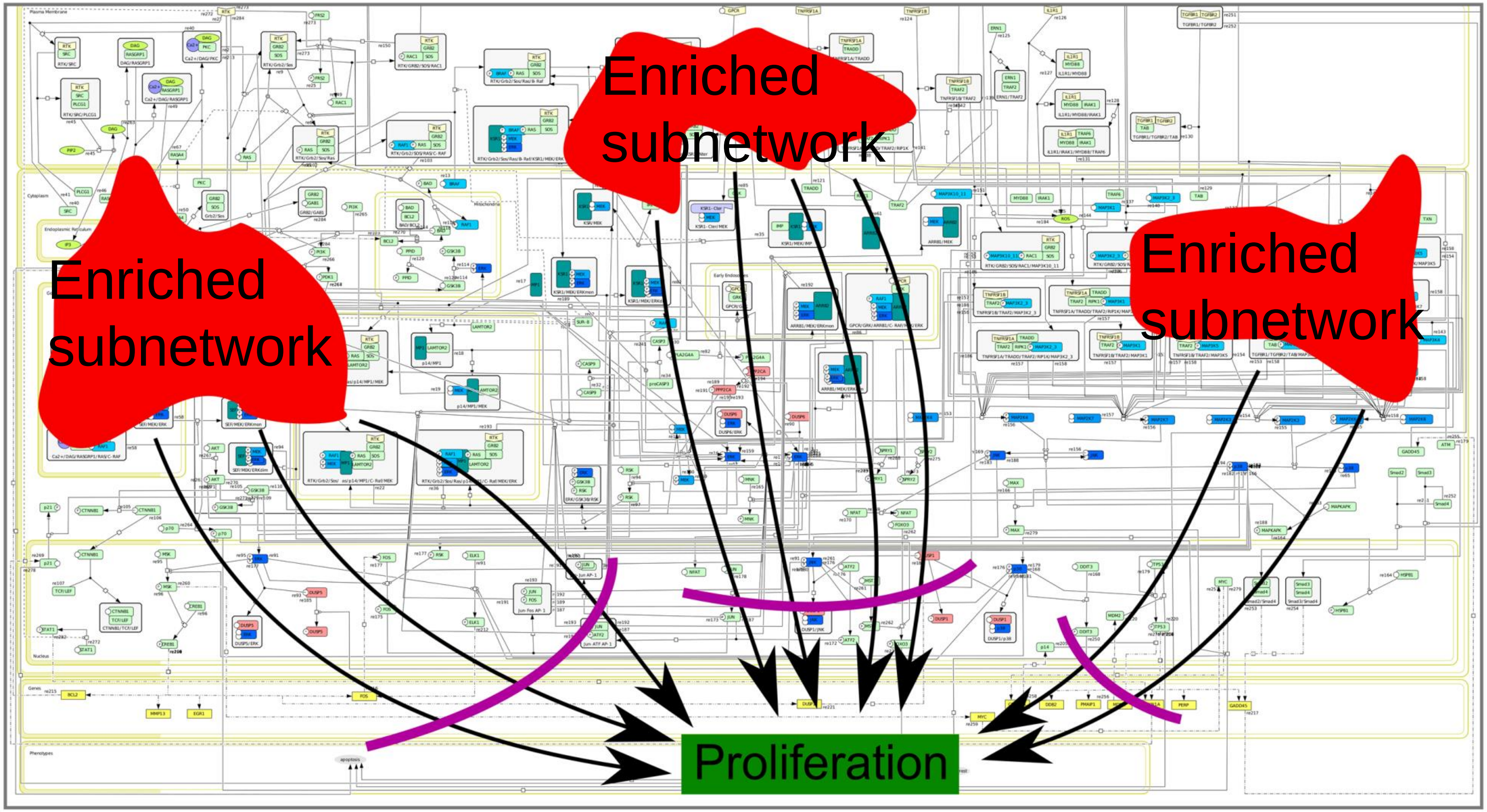
Comprehensive reconstruction of DNA repair network

DNA repair map “skeleton”: damaged DNA state transition diagram



MAPK pathway function in bladder cancer

(Kuperstein et al, Mutagenesis, 2015)



institutCurie

Inserm
Institut national de la santé et de la recherche médicale

MINES PARIS

Instead of conclusion

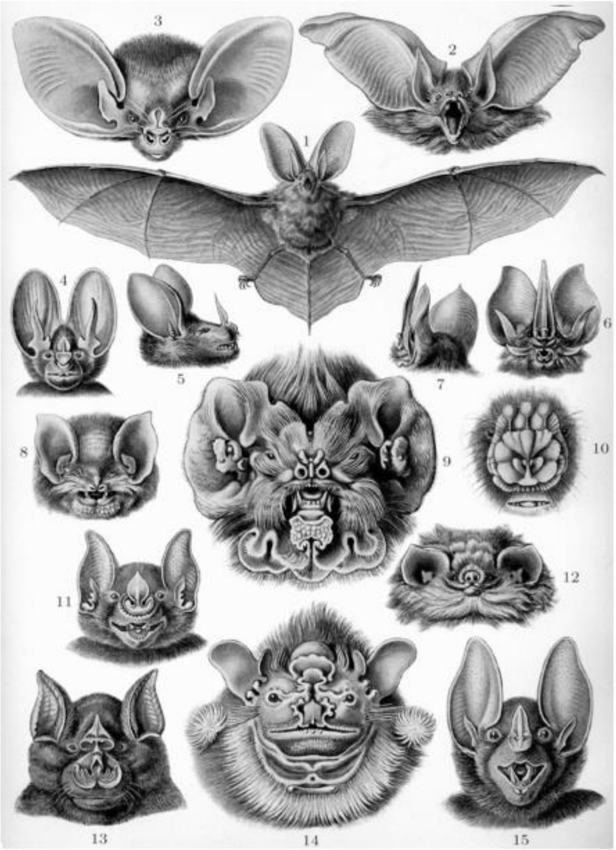


Fig. 1 Bats by Ernst Haeckel.
Lithograph from “Kunstformen der Natur” (1900)

“I think all you can do is to proceed like a child learning a language, a child mathematician, though you be biology-blind as a bat (Fig.1): you ask questions and listen to the echoes – the answers you get from biologists. You do not understand the “meaning” of the words, you are not supposed to “see” the objects they represent, but the familiar music of logic starts reverberating in your ears. If this “music” metamorphoses inside your head into a mathematically consistent arrangement of patterns, you might call it “life” or “the real world.” (c) Misha Gromov “Pattern formation in Morphogenesis”

Acknowledgements

Present and past members of Computational Systems Biology of Cancer group (Institut Curie)

Inna Kuperstein
Eric Bonnet
Emmanuel Barillot
Laurence Calzone
David Cohen
Mathurin Dorel

Luca Grieco
Christophe Russo
Maria Kondratova
Simon Fourquet
Hien-Ahn Nguyen
Urszula Czerwinska

All experts contributed to map curation
(see full list at the web-site)

Sysra company
Eric Viara and Stuart Pook

INSERM U1021
Marie Dutriex

Agilent company
Thought Leader Award-2013

More than 15 projects including European ones, supported ACSN map construction in 2008-2015

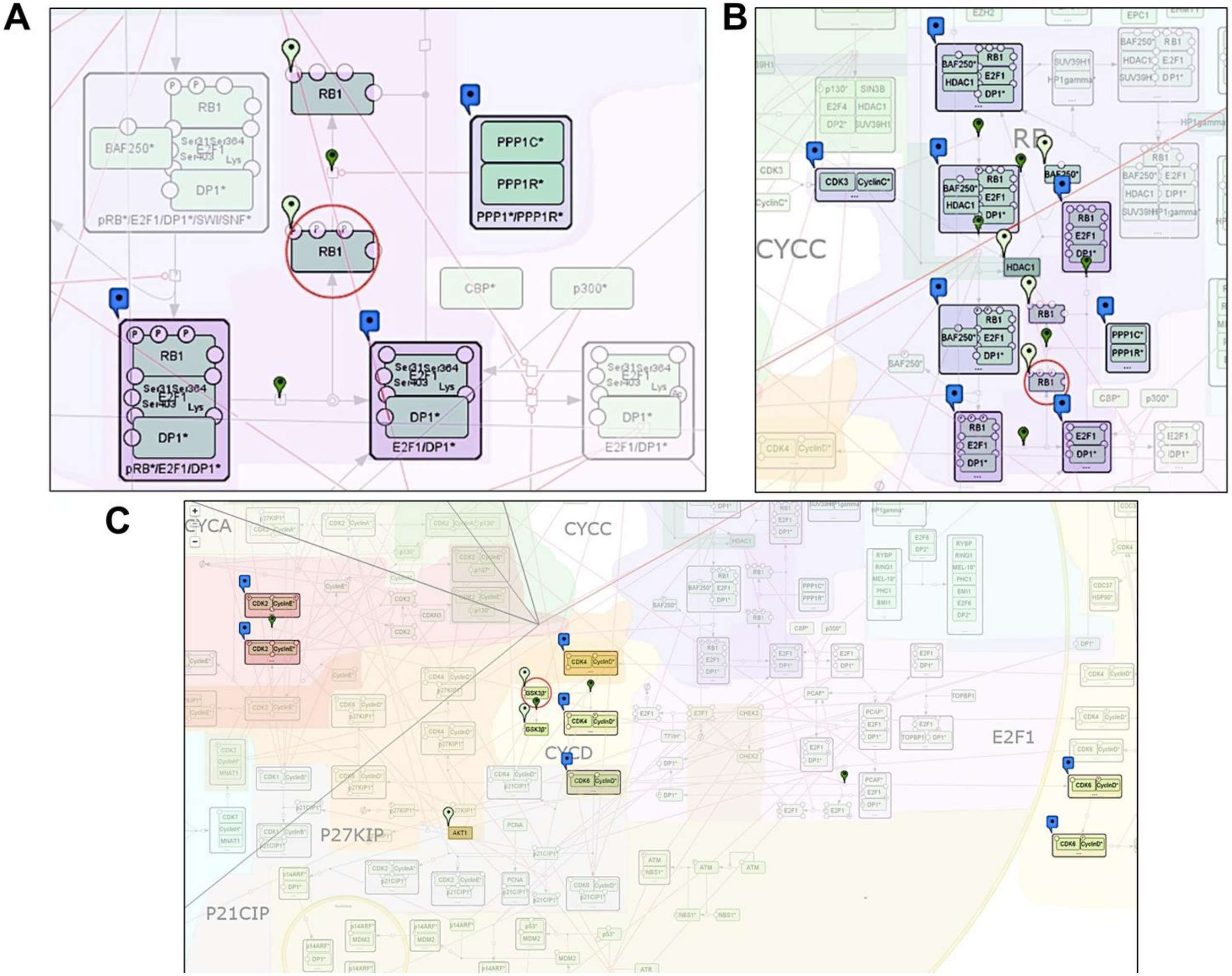
institutCurie

Inserm
Institut national
de la santé et de la recherche médicale

MINES PARIS

ACSN-NaviCell new functionality

- Highlighting part of the map
- Exploring protein/species network neighborhood



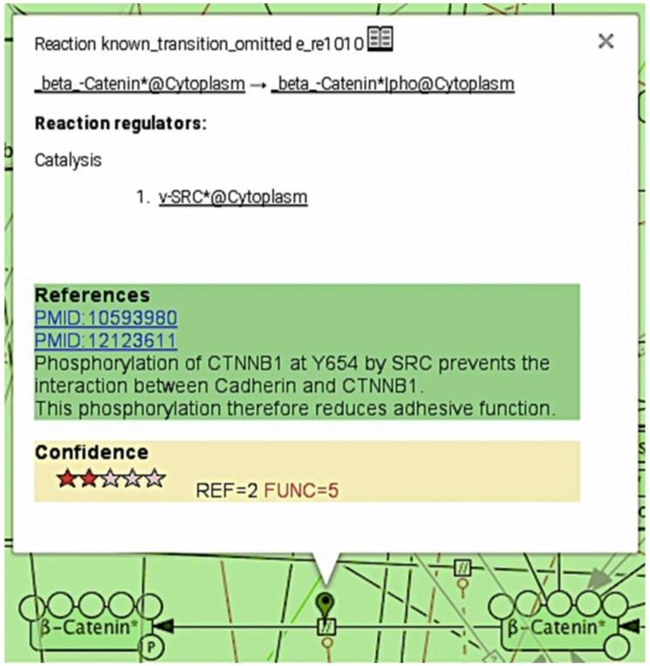
institutCurie

MINES PARIS

ACSN-NaviCell new functionality

- Confidence scores for reactions and complexes
  - **Reference score** (number of publications in the annotation)
  - **Functional proximity score** (average distance between products and reactants or protein complex components in a PPI graph)

A



Reaction known_transition_omitted e_re1010

beta-Catenin*@Cytoplasm → _beta_-Catenin*|pho@Cytoplasm

Reaction regulators:

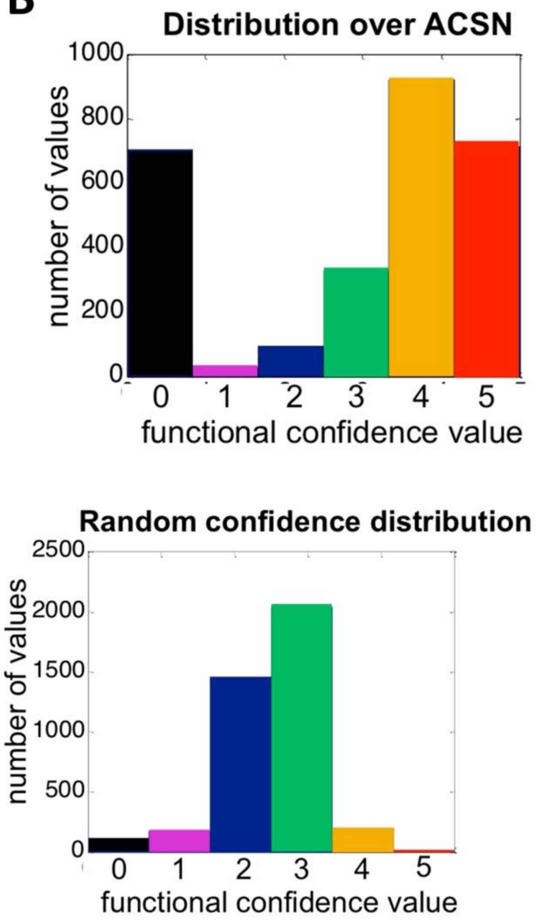
Catalysis

1. v-SRC*@Cytoplasm

References
PMID:10593980
PMID:12123611
Phosphorylation of CTNNB1 at Y654 by SRC prevents the interaction between Cadherin and CTNNB1.
This phosphorylation therefore reduces adhesive function.

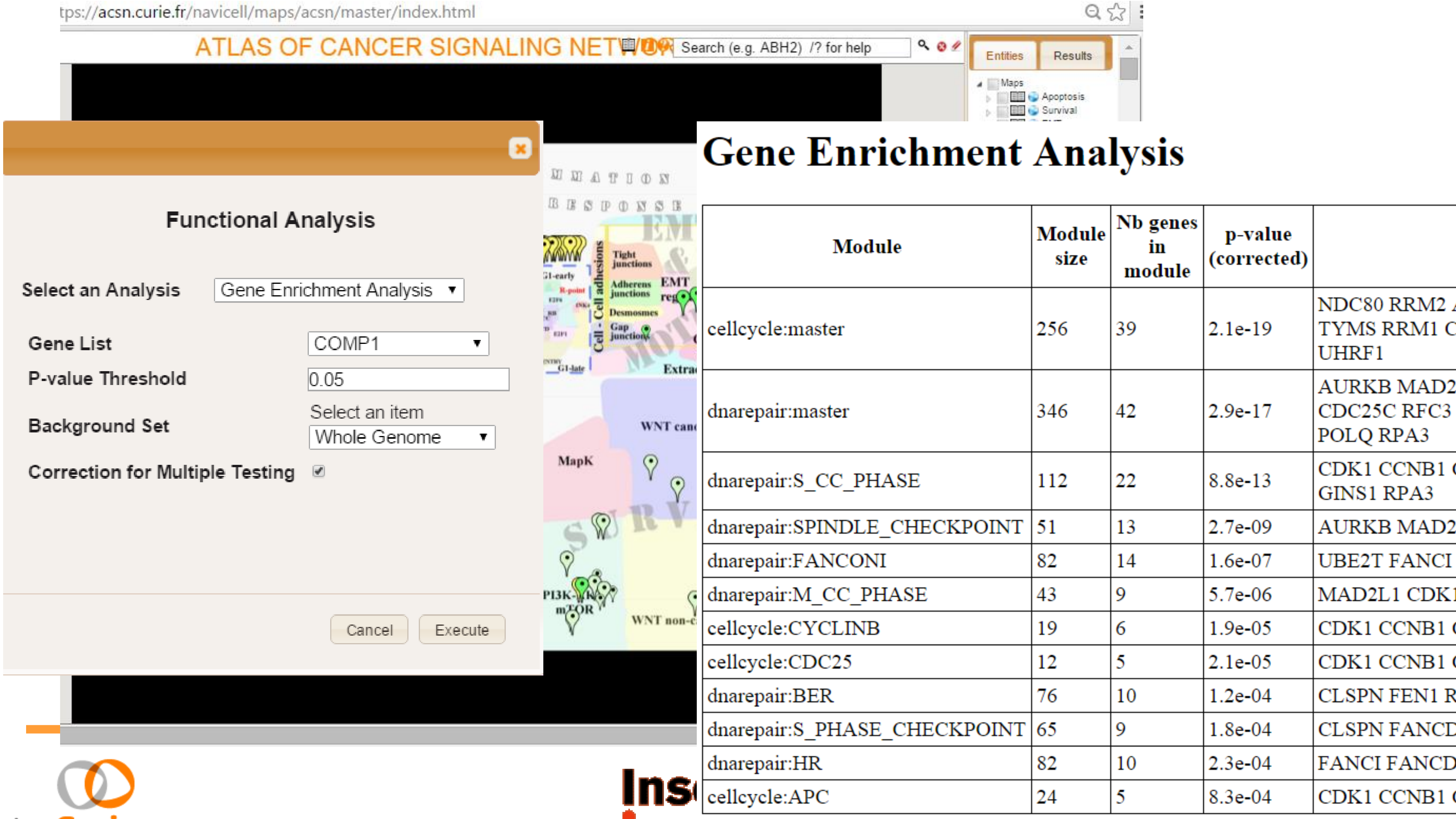
Confidence
★★☆☆☆ REF=2 FUNC=5

B



ACSN-NaviCell new functionality

- Built-in module enrichment analysis



Functional Analysis

Select an Analysis: Gene Enrichment Analysis

Gene List: COMP1

P-value Threshold: 0.05

Background Set: Select an item: Whole Genome

Correction for Multiple Testing ☑

Gene Enrichment Analysis

Module	Module size	Nb genes in module	p-value (corrected)	
cellcycle:master	256	39	2.1e-19	NDC80 RRM2 TYMS RRM1 UHRF1
dnarepair:master	346	42	2.9e-17	AURKB MAD2 CDC25C RFC3 POLQ RPA3
dnarepair:S_CC_PHASE	112	22	8.8e-13	CDK1 CCNB1 GINS1 RPA3
dnarepair:SPINDLE_CHECKPOINT	51	13	2.7e-09	AURKB MAD2
dnarepair:FANCONI	82	14	1.6e-07	UBE2T FANCI
dnarepair:M_CC_PHASE	43	9	5.7e-06	MAD2L1 CDK
cellcycle:CYCLINB	19	6	1.9e-05	CDK1 CCNB1
cellcycle:CDC25	12	5	2.1e-05	CDK1 CCNB1
dnarepair:BER	76	10	1.2e-04	CLSPN FEN1 R
dnarepair:S_PHASE_CHECKPOINT	65	9	1.8e-04	CLSPN FANCD
dnarepair:HR	82	10	2.3e-04	FANCI FANCD
cellcycle:APC	24	5	8.3e-04	CDK1 CCNB1