# Knowledge Formalization and High-Throughput Data Visualization Using Signaling Network Maps

Maria Kondratova[1,2,3,4], Emmanuel Barillot[1,2,3,4], Andrei Zinovyev[1,2,3,4] and Inna Kuperstein[1,2,3,4*]

[1]Institut Curie, France, [2]INSERM, U900, France, [3]Mines Paris Tech, France, [4]PSL Research University

*Corresponding author: Kuperstein I, Institut Curie, 26 rue d'Ulm, F-75248 Paris, France
Tel: +33 (0) 1 56 24 69 87
Email: inna.kuperstein@curie.fr

## ABSTRACT

Graphical representation of molecular biology knowledge in the form of interactive diagrams became widely used in molecular and computational biology. It enables the scientific community to exchange and discuss information on cellular processes described in numerous scientific publications and to interpret high-throughput data. Constructing a signaling network map is a laborious process, therefore application of consistent procedures for representation of molecular processes and accurately organized annotation is essential for generation of a high- quality signaling network map that can be used by various computational tools. We summarize here the major aims and challenges of assembling information in a form of comprehensive maps of molecular interactions and suggest an optimized workflow. We share our experience gained while creating a biological network resource Atlas of Cancer Signaling Network (ACSN) that was successfully applied in several studies. We explain the map construction process. Then we address the problem of user interaction with large signaling maps and suggest to facilitate navigation by hierarchical organization of map structure and by application of semantic zooming principles. In addition, we describe a computational technology using Google Maps API to explore signaling networks in the manner similar to global geographical maps and provide the outline for preparing a biological network for this type of navigation. Nowadays the most demanded application of signaling maps is integration and functional interpretation of high-throughput data. We demonstrate several examples of cancer data visualization in the context of comprehensive signaling network maps.

Keywords: Knowledge formalization; Network map construction: Data model; Network map curation; Network map navigation; Data visualization

## INTRODUCTION

Geographic maps, diagrams and flowcharts are examples of graphics that contain lots of information that is intuitive and relativity easy to grasp. Similarly, graphical representation of biological knowledge may allow to show complex processes of living cell in a visual and insightful way. Applying the principle of knowledge representation in the form of a diagram can help for systematic representation and formalization of molecular information distributed in thousands of papers. An additional advantage of representing the biological processes in a graphical form is catching collectively multiple cross-talks between components of different cell signaling processes. This allows understanding the global picture and connectivity between processes that is very difficult to keep in mind just from reading multiple scientific papers. Once the processes are depicted together as diagrams, the relationship between molecular circuits in cells can be appreciated, which makes signaling network maps also didactic tools [1].

Representation of biological processes as comprehensive signaling maps has three objectives: (i) generating a resource containing formalized summary of biochemical mechanisms as elucidated by many research groups, (ii) providing a platform for sharing information and discussing the mechanisms of biological processes, (iii) creating an analytical tool for high-throughput data integration and analysis. To achieve these goals, signaling map construction should become an accessible procedure that can be completed in a reasonable time. During last decade, the molecular biology community came up with several solutions of biological knowledge formalization that we shortly describe in this manuscript. We contribute to this global aim by describing the main principles of our established workflow for manual map construction. In addition, we demonstrate the ways of biological network map navigation facilitated by Google Maps technology and suggest our tools for performing data analysis and visualization on top of the signaling network maps.

There are four main modes of cell processes representation. Each mode uses a different logic to depict molecular information: (i) interaction diagrams showing simple binary relations between molecular entities;(ii) activity-flow or regulatory networks representing the flow of information or influences of one entity on another; (iii) entity relationship diagrams depicting relations in
which a given entity participates regardless of temporal aspects and (iv) process description diagrams (known in chemical kinetics as bi-partite reaction network graphs) where sequential order of biochemical interactions is explicitly represented [2]. Using the aforementioned modes of cell processes representation, several pathway databases have emerged [3]. They serve as biological knowledge information resource and as computational analytical tools for systems- based interpretation of data. Significant number of pathway databases have been developed in the private domain, but the majority of them are free open sources (Table 1).

The current status of representing cellular mechanisms in such databases as KEGG [4], Reactome [5], Panther [6], Spike [7], Consensus Path DB [8] and others, mainly remains at the level of drawing individual signal transduction pathways, that precludes clear representation of cross-regulations between pathways. The alternative solution is creating the seamless map of biological mechanisms covering multiple cell processes at one canvas as it is done in Atlas of Cancer Signaling Network (ACSN) [9] and KEGG metabolic pathway [10]. This “geography- inspired” approach to biological knowledge representation is a very attractive goal. However, achieving this goal is connected with a number of challenges related to creation, maintenance and navigation of the large signaling network

maps. This manuscript is partially devoted to discussing these challenges and suggesting solutions from our practical long-term experience.

Variety of uncompatible ways are used to represent biological maps in different pathway databases. This creates difficulties in combining complementary maps from multiple resources. With the aim to join the efforts in the field and create a collection of mergeable and exchangeable signaling maps, common rules of map drawing and standard graphical syntax should be developed and consistently applied. The current solution suggested in the field is Systems Biology Graphical Notation (SBGN) syntax, which is compatible with many pathway drawing and analytical tools, allowing to represent not only biochemical processes, but also cell compartments and phenotypes [11]. In such databases as Reactome, Panther pathways diagrams are represented in the SBGN graphical format. In addition, to enable cross-compatibility, several common pathway exchange formats were suggested such as BioPAX, SBML, PSI-MI etc [2].

For creating signaling maps, there exist several free and commercial tools for signaling map diagram construction. These tools use different syntax and also vary in their accessibility for the end users. Examples are SBGN-ED [12], visANT [13], CellDesigner [14] and others (Table 1).

Visualization and exploring biological network diagrams became an important issue, because size and complexity of molecular networks approach the modern geographic maps. Therefore, several tools such as CellPublisher [15], Pathway Projector [16] and NaviCell [17,18] have adopted the logic of navigation from Google Maps technology. They allow uploading networks diagrams, exploring big networks in a user-friendly manner using such Google Maps features as scrolling, zooming, markers and callouts (Table 1).

We provide a workflow where we briefly describe our methodology to meet challenges of map construction, navigation and data integration. We suggest a methodology that is neither unique nor universal, but provides practical solutions for comprehensive maps generation and manipulation. This methodology served for creating the maps for ACSN resource [9]. The approach was also successfully applied in several studies [19,20]. Each step of the workflow starts from the problem statement and description of the principles, followed by a solution suggestion demonstrated on a typical example.

**We discuss the following topics:**

- Defining the aim and the coverage of knowledge on signaling map
- Literature selection and signaling map drawing in CellDesigner tool using SBGN-like syntax[14]
- Preparation of CellDesigner maps in NaviCell format and generation of a NaviCell web-based pages
- Navigation modes using Google Maps-based NaviCell tool [17]
- High-throughput data visualization on top of the signaling maps using NaviCell Web Service module [18]

The entry point to the detailed description of the procedures is provided in the section 'Documentation', and available at https://navicell.curie.fr/pages/guide.html.

The suggested workflow for map generation and exploration is schematically depicted in Figure 1.

## GENERAL PRINCIPLES AND WORKFLOW FOR MAP CONSTRUCTION

### Work Organization

Signaling map construction requires an overview of broad scientific literature and a very minute work for correct representation of molecular processes in great details. Given large work content, map construction can be approached as a collective effort. To achieve efficient and coordinated teamwork, several important decisions should be made prior to the map construction by answering to the following questions: (i) What is the purpose of map construction? (ii) What map type is suitable for proper representation of the knowledge? (iii) What is the appropriate tool to build the map? (iv) What processes to include into the map? (iv) How the map will look like? Once the answers to those questions are found and agreement on the approach is achieved, the way of signaling diagram construction should be strictly followed to ensure generation of a homogeneous and accurate map. An additional important step before constructing a map is consulting similar efforts in the field and clear understanding of added value of a new map.

### Map Purpose and Type

As an example, we use the DNA repair map from ACSN resource available at https://acsn.curie.fr/navicell/maps/dnarepair/master/index.html. With purpose to understand how different types of DNA damage are repaired in the cell and how the coordination between various DNA repair mechanisms and the cell cycle takes place, we have decided to construct a comprehensive map of DNA repair and cell cycle signaling. We aim to use this map for detecting the modes of DNA repair machinery rewired in different pathological situations such as cancer or under genotoxic stress. The mechanisms of DNA repair are well studied and information on involved molecules and regulation circuits is available. Therefore, to preserve and depict accurately the processes in whole complexity, we have chosen the process description diagram type were the biochemical reactions can be explicitly depicted.

### Map construction tool, graphical standard and data model

Signaling processes are represented as biochemical reactions in CellDesigner diagram editor. CellDesigner uses standard Systems Biology Graphical Notation (SBGN) syntax [11] and is based on Systems Biology Markup Language (SBML) for further computational modeling of the map [14]. The data model, that is applied for our example is schematically depicted in Figure 2. This data model includes such molecular entities as proteins, genes, RNAs, antisense RNAs, simple molecules, ions, drugs, phenotypes, complexes. Biochemical reactions connect reactants to products and various types of reaction regulators are also depicted. Edges on the map represent biochemical reactions or reaction regulations including post-translational modifications, translation, transcription, complex formation or dissociation, transport, degradation, etc. Reaction regulations are catalysis, inhibition, modulation, trigger, and physical stimulation. It is also possible to depict cell compartments such as cytoplasm, nucleus, mitochondria, etc.
See http://celldesigner.org/documents.html for CellDesigner tool guide and http://www.sbgn.org/, Main_Page for SBGN syntax explanation.

## MAP BOUNDARIES, LAYOUT AND STRUCTURE

### Map Boundaries and Content

Given limitations of graphical tools and difficulties in manipulating large maps, it is recommended to define the boundaries of signaling maps. The most natural way to set map boundaries is to dedicate each map to one biological function (e. g. cell death, DNA replication, immune response) that is a difficult task per se due to 'fuzziness' of borders between processes and overlaps between players and pathways across cell signaling. Therefore, those function- driven maps should be assumed as components of the global atlas where the merging via common players or overlapping parts should be possible due to common standards applied and common identifiers for entities universally used through all maps. The decision about map boundaries highly depends on the opinion of the map creator, thus community-based curation of maps is crucial for making more objective decisions.

In our example, the boundaries of DNA repair map were defined in coordination with several specialists in the corresponding fields and based on the commonly accepted vision of pathways as they are presented in seminal reviews and in well-known databases. Accordingly to the current definition of DNA repair, it is possible to distinguish 10 different modes of repair depending on the type of damage and mechanisms of repair. The DNA damage types are clearly depicted on the map as input initiating DNA repair mechanisms. The ten DNA repair mechanisms with multiple crosstalk [21]; four cell cycle phases and check points including regulatory circuits between cell cycle to DNA repair mechanisms via checkpoints are included into the map [22] (Figure 3).

### Map Layout and Hierarchical Modular Structure

The aim of signaling map construction is not only to summarize molecular mechanisms, but also to allocate processes in a meaningful and biologically relevant way. Careful design of signaling map layout helps for intuitive understanding of 'what is where' and 'what is close and what is distant'. In addition, a bird's eye view on the map can give a general impression about map complexity based on interaction density between players in the network.

The structure of signaling map and its layout are messages by themselves. There are at least three ways of map layout choice: (i) Representing spatial localization of processes in the context of the global cell architecture. Most of the signal transduction maps mimic accepted view of cell organization, they include cellular compartments and signaling pathways are placed in the corresponding compartments on the map (see examples at https://acsn.curie.fr). (ii) Depicting process propagation in time, for instance, demonstrating propagation of the signaling through four phases of cell cycle (Figure 3A, cell cycle). (iii) Placing processes together according to involvement in a particular biological function. The most representative example is DNA repair pathways, where each pathway is depicting one biological function, these pathways are allocated next to each other, creating together a DNA repair machinery layer (Figure 3A, DNA repair). The combination of layout types at the same map is also possible as in the case of DNA repair map combining the three aforementioned layout types (Figure 3A and https://acsn.curie.fr/navicell/maps/dnarepair/master/index.html).

Type of chosen layout also can guide in separating big maps into sub-maps (modules) and help generating hierarchical modular structure of map. Each such a module can exist as a part of the global map and as an independent map. Exploring these module maps together with the global map can be

supported by Google Maps-based map navigation that facilitates understanding of depicted processes (discussed below). Map dimensions and layout should be defined prior to initiating the map drawing. This step is especially crucial in the case of collective map reconstruction approach where final global map is assembled from a number of sub-maps (or module maps).

In our example, the global layout of DNA repair map has been designed to emphasize the cross-regulation between three major blocks of the map: the knowledge of DNA repair machinery and its connection to the cell cycle and to the checkpoints is represented as layers. The upper layer depicts cell cycle, the middle layer represents cell cycle checkpoints coordinating the crosstalk between the cell cycle and the DNA repair machinery which is represented in the lower layer. The DNA map has modular structure composed of 18 functional modules corresponding to ten DNA repair mechanisms, four phases of cell cycle and four checkpoints, all interconnected, with multiple regulatory circuits (Figure 3A).

Typically, in the case of large maps, close up view on individual functional modules within the context of global map is difficult due to the distant location of some players and high density of edges crossing the map in multiple directions. To overcome these constraints, individual module layout can be designed for each modular map that can differ from the original one on the global map and aims to better represent the detailed biochemical reaction flow. An example of a module map with optimized layout for Base Excision Repair (BER) pathway is shown in Figure 3B. For modular map generation instructions see in the Map preparation in NaviCell format procedure (https://navicell.curie.fr/doc/NaviCellMapsPreparationProcedure.pdf).

## DATA EXTRACTION AND REPRESENTATION

### Manual Data Mining

Clear understanding of biological processes and experimental methodologies is essential for constructing a correct and usable signaling map. Different types of molecular interactions can be demonstrated by various experimental methods. Retrieving the correspondence between a method to evidence is needed for correct interpretation of the results described in the papers, followed by suitable graphical representation of the statements. The most common and reliable experimental methods confirming different types of molecular interactions in the cell are summarized in Table 2.

Similarly, it is possible to generate automatic layouts for module maps or even for maps of individual entity life-cycles with their related edges and reactions. This can be performed in Cytoscape using BiNoM plugin with help of modularization and automatic layout functions [23]. To facilitate integration of separate signaling diagrams, there are at least two methods for map merging: (i) Merge Model plugin in CellDesigner [14] and (ii) BiNoM plugin of Cytoscape which allows to reorganize, dissect and merge disconnected CellDesigner pathway diagrams [23].
See https://binom.curie.fr/docs/BiNoM_Manual_v2.pdf for map merging procedure in BiNoM.

One on the major aims of signaling network maps is to represent biological processes with great precision including post-translational modifications, transport, complex association, degradation etc. Phenotype nodes on the signaling maps normally serve for indication of signaling readouts or cell statuses or biological processes in general. This type of nodes can also serve for schematic representation of observation-type statements when the exact molecular mechanism is still unknown.

Some details on cell signaling might be skipped or in contrary, represented rigorously, depending on the purpose of the map drawing and opinion of the map creator. Persisting homogeneity in the presentation of information on the map will ensure correct stepwise appearance of details on different map views (discussed below).

For efficient map construction, we suggest applying systematic literature revision. Hierarchical organization of the map discussed in the previous paragraphs, reflects the principle of literature curation for map generation in our example. To define the map boundaries and content, seminal review papers in the field are used which provide a list of original references. It is also recommended to consult the major pathway databases (Table 1). The canonical pathways retrieved from major reviews and databases reflect the consensus view of the field and serve as a basis for drawing the core processes on the map. Thus, the details can be added to the map to depict information extracted from the recent literature with the requirement that the interactions and processes included into the map are supported by at least two independent investigations. Figure 4 represents an example from the DNA repair map on how the scientific text is translated into the process description diagram. For consistency of text to diagram conversion, we have developed several major rules for standard statements interpretation, summarized in:
Map creator guide
https://navicell.curie.fr/doc/NaviCellMapperAdminGuide.pdf
Map preparation in NaviCell format procedure
https://navicell.curie.fr/doc/NaviCellMapsPreparationProcedure.pdf

**Processing Formal Statements on Biochemical Reactions**

An additional useful method for knowledge to diagram conversion is using a formalized text- based intermediate language. In this method the sentences formulated in the "human" language are first converted into a set of formal statements describing reactions. These statements can be automatically converted into the graphical diagrams, using BiNoM Cytoscape plugin. For example, the set of statements corresponding to the diagram in Figure 4 is:

*BRCA1@cytoplasm -/> BRCA1@nucleus*
*BRCA1@nucleus+BARD1@nucleus -:> BARD1:BRCA1@nucleus*
*BRCA1-CHEK2|pho@nucleus->BRCA1|pho|active@nucleus*
*rBRCA1@cytoplasm-.> BRCA1@cytoplasm*
*gBRCA1@nucleus-|p53*@nucleus -BRIT1*:E2F1@nucleus -..>rBRCA1@cytoplasm*

These statements can be prepared in any simple text-based editor and imported to Cytoscape environment through BiNoM and then converted directly into the CellDesigner diagram presented in Figure 4. The syntax of this BiNoM Reaction Format (BRF) language have been depicted in [19,23] and in the BiNoM manual (https://binom.curie.fr/docs/BiNoM_Manual_v2.pdf).

**Map Entities Annotation**

Common entity annotation format and consistent integration of stable identifiers for map entities are essential for compatibility of maps with other tools. It also facilitates integration of data into maps and other manipulations as cross-curation of maps by specialists and incorporation of their corrections into the map. We have developed the NaviCell annotation format for each entity that is applied during map construction in CellDesigner (Figure 5). The annotation panel includes sections 'Identifiers',

‘Maps_Modules’, ‘References’ and ‘Confidence’. ‘Identifiers’ section provides standard identifiers and links to the corresponding entity descriptions in HGNC, UniProt, Entrez, SBO, Gene Cards and cross-references in REACTOME [5], KEGG [4], Wiki Pathways [25] and other databases. Metabolites and small compounds are annotated by corresponding identifiers and linked to ChEBI [26], PubChem Compound [27] and KEGG Compound [28] databases.
‘Maps_Modules’ section includes links to modules of the DNA map were the entity participates. In addition, since our example, DNA repair map, is part of Atlas of Cancer Signaling Network (ACSN) resource [1], the links to maps and modules of ACSN where the entity participates, are also provided. ‘References’ section includes notes added by the map manager and links to relevant publications (Figure 5A). Each entity annotation is represented as a post in the web- blog generated when the NaviCell map is generated from CellDesigner file (described below). The web-blog is providing a possibility of communication between map users and map managers. Comments can be submitted at the blog post page in text form together with files of any type (Figure 5A). The extended description of annotation formats for each type of entities on the map is provided in https://navicell.curie.fr/doc/NaviCellMapsPreparationProcedure.pdf.

**Confidence Scores**

We introduced two simple confidence scores for proteins complexes and reactions that are provided in “Confidence” in a form “five-star” diagram and calculated automatically while conversion of CellDesigner map to NaviCell format. Both scores represent integer numbers varying from 0 (undefined confidence) to 5 (high confidence). The reference score, marked by ‘REF’ indicates both the number and the ‘weight’ associated to publications found in the annotation of a given reaction, with weight equal 1 point for an original publication and 3 points for a review article. The functional proximity score, marked by ‘FUNC’ is computed based on the external network of protein-protein interactions (PPI), Human Protein Reference Database (HPRD) [29]. The score reflects an average distance in the PPI graph between all proteins participating in the reaction as reactants, products or regulators. The functional proximity is computed using BiNoM Cytoscape plugin [23] (Figure 5B).

## GENERATION OF NAVICELL MAP USING NAVICELL FACTORY AND EXCHANGE FORMATS

CellDesigner maps annotated in the NaviCell format or un-annotated, can be converted into a NaviCell web-based front-end, which represents a set of html pages with embedded JavaScript code that can be launched in a web browser locally or put on a web-server for further online use. Use of identifiers in the annotation of proteins (“HUGO:XXX” tag) will allow using NaviCell data visualization functionality. The NaviCell factory is embedded in the BiNoM Cytoscape plugin and also available as a stand-alone command line package (https://github.com/sysbio-curie/NaviCell). The detailed guide of using the NaviCell factory is provided at https://navicell.curie.fr/doc/NaviCellMapperAdminGuide.pdf.

The maps generated in CellDesigner and exposed in NaviCell format can also be provided in common exchange formats to ensure compatibility of maps with other computational tools. Currently, maps can be generated in BioPAX and PNG formats. In addition, the module composition of maps can be provided in a form of GMT files. The description of map preparation in various formats using the

BiNoM Cytoscape plugin is available in the BiNoM manual https://binom.curie.fr/docs/BiNoM_Manual_v2.pdf.

**MAP NAVIGATION**

Comprehensive signaling maps, as DNA repair map, contain large number of nodes and edges that makes navigation through the map difficult. To solve this problem, maps can be represented as clickable web pages with a clear graphic user interface. We have developed user-friendly NaviCell web-based environment empowered by Google Maps engine for visualization and navigation of the comprehensive maps [17]. These features are demonstrated in our example: scrolling, zooming, markers, callout windows and zoom bar are adopted from the Google Maps interface (Figure 3C). The map in NaviCell is interactive and all components of the map are 'clickable'. For finding the entity of interest, querying for single or multiple molecules using the search window is possible. Alternatively, the entity can be found in the selection panel or directly on the map.

We describe several solutions to optimize navigation through the maps. First, 'horizontal' navigation, that is facilitated by the hierarchical modular structure of DNA repair map, as it is described in previous paragraphs. This modular structure allows shuttling between the global map to the functional module maps and observe different processes on separate modular maps, containing only limited number of entities, as it is demonstrated in the DNA repair map example in Figure 3B.

Second, 'vertical' navigation, is facilitated by the semantic zooming feature of maps, especially prepared in NaviCell format. Semantic zooming simplifies navigation through the large maps of molecular interactions, showing readable amount of details at each zoom level. Gradual appearance of details allows exploration of the map content from the top-level toward detailed view. To prepare maps for this type of navigation, pruning of maps is performed in order to eliminate non-essential information for each zoom level. We recommend to prepare four zoom levels, although number of zooms is unlimited in NaviCell.

In our example of DNA repair map, the first, top-level view, shows modules of the map depicted as colored background shapes (Figure 3A). At the next level, a more detailed level of zooming shows canonical cell signaling pathways. These pathways are defined by intersecting the content of the map with the corresponding pathways in other databases (Figure 6A). Next zoom hides names of complexes, entities and reaction (Figure 6B) and the last zoom is the most detailed view where all map elements are present (Figure 6C). The module background coloring appears as a context layer in the background of all levels of zooming. For detailed instructions on map zoom levels creation see in:
Map creator guide
https://navicell.curie.fr/doc/NaviCellMapperAdminGuide.pdf
Map preparation in NaviCell format procedure
https://navicell.curie.fr/doc/NaviCellMapsPreparationProcedure.pdf

A third way to facilitate exploration of a map is focusing on individual entities of the map by highlighting them. We have developed a NaviCell function of selecting and highlighting species of interest or neighbors of a molecular species of interest. This function allows step-wise enlarging of the neighborhood coverage to understand propagation of signaling on the map as shown for p53* molecule on the DNA repair map in Figure 6D.

## MAP MAINTENANCE AND CURATION

Given the fact that biological knowledge about the majority of signaling pathways is not yet solid and continuously grows, one of the major problems of signaling maps is their fast obsolescence. To overcome the problem, permanent maintenance and updating of maps is essential. The community of users is the most reliable and trustable contributor to map maintenance, because specialists could support and update maps from the area of their own research that would ensure highest quality of maps update. To enable such a community-based effort, efficient curation tools should be created. To our knowledge, there is only one community curation tool for comprehensive maps, the Payoa plugin of CellDesigner (30).

We recommend to carry out map curation in the context of NaviCell environment. The process of map curation and maintenance in NaviCell involves map managers that regularly examine the comments posted in the blog of the maps (Figure 5A), check the latest scientific literature and update the maps and the annotations accordingly. An automated procedure supports the map updating and archives older versions of posts including comments, thus providing traceability of all changes on the maps and all discussions in the blog [17].

## VISUALIZATION OF OMICS DATA IN THE CONTEXT OF SIGNALING NETWORK MAPS

To make data visualization a straightforward and easy task, we have developed a built-in toolbox for visualization and analysis of high-throughput data in the context of - comprehensive signaling networks. The integrated NaviCell web-based toolbox allows importing and visualizing heterogeneous omics data on top of the maps and performing simple functional data analysis. It is also suitable for computing aggregated values for sample groups and protein families and mapping this data onto the maps. The tool contains standard heatmaps, barplots and glyphs as well as the novel map staining technique for displaying large-scale trends in the numerical values along the map. The combination of these flexible features provides an opportunity to adjust the modes of visualization to the data type and achieve the most meaningful picture [18]. An extended documentation, tutorial, live example and guide for data integration using NaviCell is provided at https://navicell.curie.fr/pages/nav_web_service.html.

To illustrate data visualization, the DNA map was used for analysis of omics data from breast cancer patients. To grasp the difference in the data distribution on top of the map and rewiring of signaling processes across different stages of breast cancers, the module activity was calculated based on transcriptomic data. The colors represent the average contribution of all module components (Figure 7A). There is a clear difference in pattern between non-invasive stage 1 group of breast cancer patients to stages groups 4 invasive group, indicating a major shift in signaling involvement while the cancer cells transformation from non-invasive to invasive. In addition, closer look at the cell cycle checkpoints shows that spindle checkpoint is activated in the stage 4, whereas all other cell cycle checkpoints are down regulated. The observation is consistent with the 'checkpoint addiction' phenomena, when the tumors develop dependence on spindle checkpoint to allow chromosomes separation despite accumulated genomic instability and ensures cell proliferation regardless of the status of DNA damage [31].
In another example, in order to illustrate the coverage of gene mutation frequencies over the DNA repair map, the information about mutations in breast cancer was obtained from the Catalogue

Of Somatic Mutations In Cancer (COSMIC) database [32] and mapped the most frequently mutated oncogenes and tumor suppressor genes (TSG) as glyphs using NaviCell data visualization toolbox (Figure 7B). TSG mutations are more frequently found in modules of DNA repair map than oncogenes. This might indicate that TSGs in those processes normally contribute to the restriction of uncontrolled divisions and unrepaired DNA, but are inactivated by mutations in cancer.

An additional way to visualize the omics data on top of the maps that we also developed, but do not detail here, is using Cytoscape plugin BiNoM were the map staining displaying module activity values and coloring individual protein nodes with corresponding transcriptome data is possible [23].

## CONCLUSIONS AND PERSPECTIVES

Representation of relationships between cell molecules in the form of a biochemical process diagram depicts our current understanding of how cell activity is coordinated at the molecular level. The advantage of drawing biological processes in the form of interconnected network is not only to bring together components that participate in the described process, but also to allow capturing non-trivial interactions and regulatory circuits between those components. As the load of knowledge about biological mechanisms increasingly grows, organization, structuring and systematized representation of this data is essential for creating the global picture. Standardized representation of biological processes, intuitive maps navigation tools, community contribution to revising and updating the networks diagrams simplify construction of new networks and facilitate maintenance of existing signaling diagrams collections. These comprehensive signaling maps serve as a basis for modeling of signaling networks and efficient analysis and interpretation of high-throughput data [70].

Here we have described the methodology developed in the group following our long- standing experience with comprehensive maps generation and manipulation. Using this approach we have created and currently maintain a pathway resource of Atlas of Cancer Signaling Network (ACSN) [9] and a collection of maps created in CellDesigner available at https://navicell.curie.fr/pages/maps.html

We have suggested a workflow for construction and annotation of signaling maps in CellDesigner, preparing the hierarchical modular structure of maps and also generation of different levels of the maps view, to allow semantic zooming-based exploration of maps in NaviCell. We have introduced NaviCell that is an environment for navigating large-scale maps of molecular interactions created in CellDesigner. NaviCell allows showing the content of the map in a convenient way, at several scales of complexity or abstraction; it provides an opportunity to comment on map content, facilitating curation and maintenance of the map. Finally we have shown how complex data can be visualized and interpreted in the context of the map.

Among many future challenges of the signaling network community are integration of similar efforts as improvement of network exchange formats and development of common network dynamic layouts. In addition, generation of comprehensive platforms for tools, data, and knowledge sharing in systems biology and biomedical research, similar to GARUDA initiative (http://www.garuda-alliance.org), will facilitate tools and resources compatibility improvement.

## DOCUMENTATION

CellDesigner introduction and tutorial
http://celldesigner.org/documents.html

SBGN
http://www.sbgn.org/Main_Page

BiNoM manual
https://binom.curie.fr/docs/BiNoM_Manual_v2.pdf

Map creator guide
https://navicell.curie.fr/doc/NaviCellMapperAdminGuide.pdf

Map preparation in NaviCell format procedure
https://navicell.curie.fr/doc/NaviCellMapsPreparationProcedure.pdf

NaviCell Web Service guide
https://navicell.curie.fr/doc/ws/NaviCellWebServiceGuide.pdf

NaviCell Web Service introduction, tutorial and case studies
https://navicell.curie.fr/pages/nav_web_service.html

Interactive demo on data visualization using NaviCell
https://navicell.curie.fr/navicell/maps/cellcycle/master/index.php?demo=on

ACSN introduction, tutorial and case studies
https://acsn.curie.fr/documentation.html

DNA repair map from ACSN resource
https://acsn.curie.fr/navicell/maps/dnarepair/master/index.html

## ACKNOLEDGMENTS

This work has been supported by the Agilent Thought Leader Award #3273; ASSET European Union Framework Program 7 project under grant agreement FP7-HEALTH-2010-259348; grant 'Projet Incitatif et Collaboratif Computational Systems Biology Approach for Cancer' from Institut Curie. This work has received support under the program «Investissements d'Avenir» launched by the French Government and implemented by ANR with the references ANR-11-BINF-0001 (Project ABS4NGS) and APLIGOOGLE program provided by CNRS. Funding for open access charge: Institutional budget of institut Curie - INSERM U900 department.

## REFERENCES


1. Kuperstein I, Grieco L, Cohen DPA, Thieffry D, Zinovyev A, et al. The shortest path is not the one you know: application of biological network resources in precision oncology research. Mutagenesis. 2015; 30: 191–204.
2. Cohen D, Kuperstein I, Barillot E, Zinovyev A, Calzone L. From a biological hypothesis to the construction of a mathematical model. Methods Mol Biol. 2013; 1021: 107–125.
3. Chowdhury S, Sarkar RR. Comparison of human cell signaling pathway databases--evolution, drawbacks and challenges. Database. 2015; 126.
4. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012; 40: 109–114.
5. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, et al. The Reactome pathway knowledgebase. Nucleic Acids Res. 2014; 42: 472–477.
6. Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, et al. The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res. 2005; 33: 284–288.
7. Paz A, Brownstein Z, Ber Y, Bialik S, David E, et al. SPIKE: a database of highly curated human signaling pathways. Nucleic Acids Res. 2011; 39: 793–799.
8. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 2013; 41: 793–800.
9. Kuperstein I, Bonnet E, Nguyen H-A, Cohen D, Viara E, et al. Atlas of Cancer Signalling Network: a systems biology resource for integrative analysis of cancer data with Google Maps. Oncogenesis. 2015; 4: e160.
10. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, et al. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res. 2014; 42: 199–205.
11. Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F. The Systems Biology Graphical Notation. Nat Biotechnol. 2009; 27: 735-741.
12. Czauderna T, Klukas C, Schreiber F. Editing, validating and translating of SBGN maps. Bioinformatics. 2010; 26: 2340-2341.
13. Hu Z, Mellor J, Wu J, DeLisi C. VisANT: an online visualization and analysis tool for biological interaction data. BMC Bioinformatics. 2004; 5: 17.
14. Kitano H, Funahashi A, Matsuoka Y, Oda K. Using process diagrams for the graphical representation of biological networks. Nat Biotechnol. 2005; 23: 961-966.
15. Flórez LA, Lammers CR, Michna R, Stülke J. CellPublisher: a web platform for the intuitive visualization and sharing of metabolic, signalling and regulatory pathways. Bioinformatics. 2010; 26: 2997-2999.
16. Kono N, Arakawa K, Ogawa R, Kido N, Oshita K. Pathway projector: web-based zoomable pathway browser using KEGG atlas and Google Maps API. PLoS One. 2009; 4: e7710.
17. Kuperstein I, Cohen DP, Pook S, Viara E, Calzone L. NaviCell: a web-based environment for navigation, curation and maintenance of large molecular interaction maps. BMC Syst Biol. 2013; 7: 100.
18. Bonnet E, Viara E, Kuperstein I, Calzone L, Cohen DP1. NaviCell Web Service for network-based data visualization. Nucleic Acids Res. 2015; 43: W560-565.
19. Calzone L, Gelay A, Zinovyev A, Radvanyi F, Barillot E. A comprehensive modular map of molecular interactions in RB/E2F pathway. Mol Syst Biol. 2008; 4: 173.
20. Chanrion M, Kuperstein I, Barrière C, El Marjou F, Cohen D. Concomitant Notch activation and p53 deletion trigger epithelial-to- mesenchymal transition and metastasis in mouse gut. Nat Commun. 2014; 5: 5005.]
21. Tian H, Gao Z, Li H, Zhang B, Wang G1. DNA damage response--a double-edged sword in cancer prevention and cancer therapy. Cancer Lett. 2015; 358: 8-16.
22. Wang H, Zhang X, Teng L, Legerski RJ. DNA damage checkpoint recovery and cancer development. Exp Cell Res. 2015; 334: 350-3588.
23. Bonnet E, Calzone L, Rovera D, Stoll G, Barillot E, et al. BiNoM 2.0, a Cytoscape plugin for accessing and analyzing pathways using standard systems biology formats. BMC Syst Biol. 2013; 7: 18.
25. Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR. WikiPathways: building research communities on biological pathways. Nucleic Acids Res. 2012; 40: D1301-1307.
26. Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 2008; 36: D344-350.
27. Li Q, Cheng T, Wang Y, Bryant SH. PubChem as a public resource for drug discovery. Drug Discov Today. 2010; 15: 1052-1057.
28. Kanehisa M. Molecular network analysis of diseases and drugs in KEGG. Methods Mol Biol. 2013; 939: 263-275.
29. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S. Human Protein Reference Database--2009 update. Nucleic Acids Res. 2009; 37: D767-772.
30. Matsuoka Y, Ghosh S, Kikuchi N, Kitano H. Payao: a community platform for SBML pathway model curation. Bioinformatics. 2010; 26: 1381-1383.
31. Bartek J, Lukas J. DNA damage checkpoints: from initiation to recovery or adaptation. Curr Opin Cell Biol. 2007; 19: 238-245.
32. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015; 43: D805-811.
33. MacLachlan TK, Dash BC, Dicker DT, El-Deiry WS. Repression of BRCA1 through a feedback loop involving p53. J Biol Chem. 2000; 275: 31869-31875.
34. Yang SZ, Lin FT, Lin WC. MCPH1/BRIT1 cooperates with E2F1 in the activation of checkpoint, DNA repair and apoptosis. EMBO Rep. 2008; 9: 907-915.

35. Chen CF, Li S, Chen Y, Chen PL, Sharp ZD, et al. The nuclear localization sequences of the BRCA1 protein interact with the importin-alpha subunit of the nuclear transport signal receptor. J Biol Chem. 1996; 271: 32863-32868.
36. Zhang J, Willers H, Feng Z, Ghosh JC, Kim S. Chk2 phosphorylation of BRCA1 regulates DNA double-strand break repair. Mol Cell Biol. 2004; 24: 708-718.
37. Xia Y, Pao GM, Chen HW, Verma IM, Hunter T. Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. J Biol Chem. 2003; 278: 5255-5263.
38. Ghosh S, Matsuoka Y, Kitano H. Connecting the dots: role of standardization and technology sharing in biological simulation. Drug Discov Today. 2010; 15: 1024-1031.
39. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013; 41: D808-815.
40. Chatr-Aryamontri A, Breitkreutz BJ2, Oughtred R3, Boucher L2, Heinicke S3. The BioGRID interaction database: 2015 update. Nuleic Acids Res. 2015; 43: D470-478.
41. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M. MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. 2012; 40: D857-861.
42. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur O. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res. 2011; 39: D685-690.
43. Schacherer F, Choi C, Götze U, Krull M, Pistor S. The TRANSPATH signal transduction database: a knowledge base on signal transduction networks. Bioinformatics. 2001; 17: 1053-1057.
44. Zinovyev A, Viara E, Calzone L, Barillot E. BiNoM: a Cytoscape plugin for manipulating and analyzing biological networks.Bioinformatics. 2008; 24: 876-877.
45. Wu G, Dawson E, Duong A, Haw R, Stein L. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis.Version 2. 1000.
46. Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P. iPath2.0: interactive pathway explorer. Nucleic Acids Res. 2011; 39: W412-415.
47. Pavlopoulos GA, Hooper SD, Sifrim A, Schneider R, Aerts J. Medusa: A tool for exploring and clustering biological networks. BMC Res Notes. 2011; 4: 384.
48. Day RN, Davidson MW. Fluorescent proteins for FRET microscopy: monitoring protein interactions in living cells. Bioessays. 2012; 34: 341-350.
49. Piehler J. New methodologies for measuring protein interactions in vivo and in vitro. Curr Opin Struct Biol. 2005; 15: 4-14.
50. Sklar LA, Edwards BS, Graves SW, Nolan JP, Prossnitz ER. Flow cytometric analysis of ligand-receptor interactions and molecular assemblies. Annu Rev Biophys Biomol Struct. 2002; 31: 97-119.
51. Berggård T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. Proteomics. 2007; 7: 2833-2842.
52. Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. Microbiol Rev. 1995; 59: 94-123.
53. Jubb H, Higueruelo AP, Winter A, Blundell TL. Structural biology and drug discovery for protein-protein interactions. Trends Pharmacol Sci. 2012; 33: 241-248.
54. Sambrook J, Russell DW. Detection of Protein-Protein Interactions Using the GST Fusion Protein Pulldown Technique. CSH Protoc. 2006.
55. Wu Y, Li Q, Chen XZ. Detecting protein-protein interactions by Far western blotting. Nat Protoc. 2007; 2: 3278-3284.
56. Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature. 2003; 422: 198-207.
57. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol. 1998; 280: 1-9.
58. Dey B, Thukral S, Krishnan S, Chakrobarty M, Gupta S. DNA-protein interactions: methods for detection and analysis. Mol Cell Biochem. 2012; 365: 279-299.
59. Stormo GD, Zhao Y. Determining the specificity of protein-DNA interactions. Nat Rev Genet. 2010; 11: 751-760.
60. Tompa M, Li N, Bailey TL, Church GM, De Moor B. Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol. 2005; 23: 137-144.
61. Kuhn DE, Martin MM, Feldman DS, Terry AV Jr, Nuovo GJ. Experimental validation of miRNA targets. Methods. 2008; 44: 47-54.
62. VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques.
2008; 44: 619-626.
63. Alam J, Cook JL. Reporter genes: application to the study of mammalian gene transcription. Anal Biochem. 1990; 188: 245-254.
64. Prediger EA. Detection and quantitation of mRNAs using ribonuclease protection assays. Methods Mol Biol. 2001; 160: 495-505.
65. Kurien BT, Scofield RH. Western blotting: an introduction. Methods Mol Biol. 2015; 1312: 17-30.
66. Bonner WA, Hulett HR, Sweet RG, Herzenberg LA. Fluorescence activated cell sorting. Rev Sci Instrum. 1972; 43: 404-409.
67. Manis JP. Knock out, knock in, knock down--genetically manipulated mice and the Nobel Prize. N Engl J Med. 2007; 357: 2426-2429.
68. Tiscornia G, Singer O, Ikawa M, Verma IM. A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA. Proc Natl Acad Sci U S A. 2003; 100: 1844-1848.
69. Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol. 2007; 3: 0337-0344.
70. Barillot E, Calzone L, Hupe P, Vert J-P, Zinovyev A. Computational Systems Biology of Cancer. CRC Press. 2012; 461-432.

Table 1: Pathways databases and network resources; navigation tools and high-throughput data visualization support

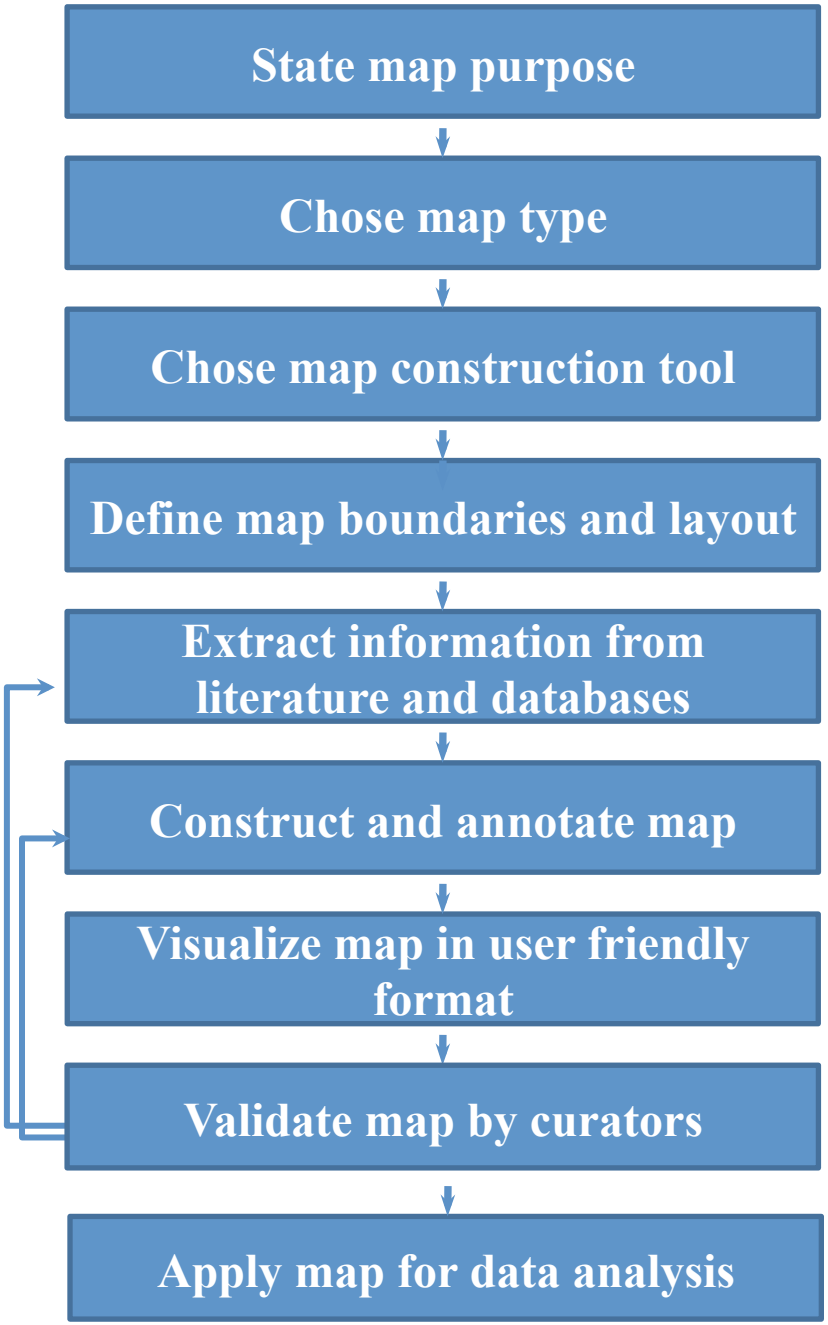
| Signaling pathways and networks resources | | | |
|---|---|---|---|
| *Name* | *Website* | *Description* | *Reference* |
| STRING | http://string-db.org | Integrated protein-protein interaction daatabase | (39) |
| BioGRID | http://thebiogrid.org | Integrated protein-protein and genetic interaction daatabase | (40) |
| MINT | http://mint.bio.uniroma2.it/mint/Welcome.do | Carefully curated PPI resource | (41) |
| PathwayCommons | http://www.pathwaycommons.org | Biological pathways resource collected from public pathway databases | (42) |
| TRANSPATH | http://www.biobase-international.com | Database of mammalian signal transduction and metabolic pathways | (43) |
| ConsensusPathDB | http://consensuspathdb.org | Integrated resource of interaction networks and pathways | (8) |
| Panther | http://pantherdb.org | Collection of biological pathways and data visualization and analysis tools | (6) |
| Spike | http://www.cs.tau.ac.il/~spike | Collection of curated, peer reviewed pathways and data visualization tools | (7) |
| WikiPathways | http://www.wikipathways.org | Collection of community curated signalling pathways | (25) |
| PID-NCI | http://pid.nci.nih.gov | Curated collection of information about biomolecular interactions and signalling pathways | |
| KEGG Pathway | http://www.genome.jp/kegg/pathway | Collection of manually drawn pathway maps visualization tool | (4) |
| Reactome | http://www.reactome.org | Collection of curated, peer reviewed pathways and data visualization/analysis tools | (5) |
| ACSN | http://acsn.curie.fr | Collection of curated, peer reviewed, interconnected cancer-related signaling networks and data visualization/analysis tools | (9) |
| Tools for network construction, visualization, navigation, and commenting | | | |
| *Name* | *Website* | *Description* | *Reference* |
| CellDesigner | http://www.celldesigner.org | Structured diagram editor for drawing gene-regulatory and biochemical networks | (14) |
| SBGN-ED | http://vanted.ipk-gatersleben.de/addons/sbgn-ed | VANTED add-on for create and edit three types of SBGN maps | (12) |
| CellPublisher | http://cellpublisher.gobics.de | KEGG database-associated tool for data visualization and analysis in the context of pathway maps | (15) |
| Cytoscape / BiNoM | http://www.cytoscape.org<br>http://binom.curie.fr | Software platform for maniputation of biological networks represented in standard systems biology formats | (44)<br>(23) |
| Payao | http://payao.oist.jp:8080/payaologue/index.html | Network curation tool for simultaneous map commenting using tag system | (30) |
| NaviCell | http://navicell.curie.fr | Web-based tool for heterogeneous data visualization and analysis in the context of signaling networks | (17)(18) |
| yEd graph editor | http://www.yworks.com/en/products/yfiles/yed/ | Application for generate high-quality diagrams construction | http://link.springer.com/chapter/10.1007/978-3-642-18638-7_8 |
| VisANT | http://visant.bu.edu/ | Tool for visual analyses of metabolic networks in cells and ecosystems | (13) |
| Pathway Map Creator | http://lifesciences.thomsonreuters.com/m/pdf/PathwayMapCreator-cfs-en.pdf | Tool for editing and analysis of canonical pathways maps | |

| Tools for visualisation of high-throughput data in the context of signalling networks | | | |
|---|---|---|---|
| *Name* | *Website* | *Description* | *Reference* |
| ReactomeFiViz | http://wiki.reactome.org/index.php/Reactome_FI_Cytoscape_Plugin | Cytoscape plugin for data integration into signaling networks | (45) |
| iPAth | http://pathways.embl.de | Web-based tool for data visualization in the context of pathway maps | (46) |
| Medusa | http://coot.embl.de/medusa | Tool for data visualization in the context of signaling network and network clustering | (47) |
| NaviCell | http://navicell.curie.fr | Web-based tool for heterogeneous data visualization and analysis in the context of signaling networks | (17)(18) |
| KEGG Mapper | http://www.kegg.jp/kegg/mapper | KEGG database-associated tool for data visualization and analysis in the context of pathway maps | (4) |

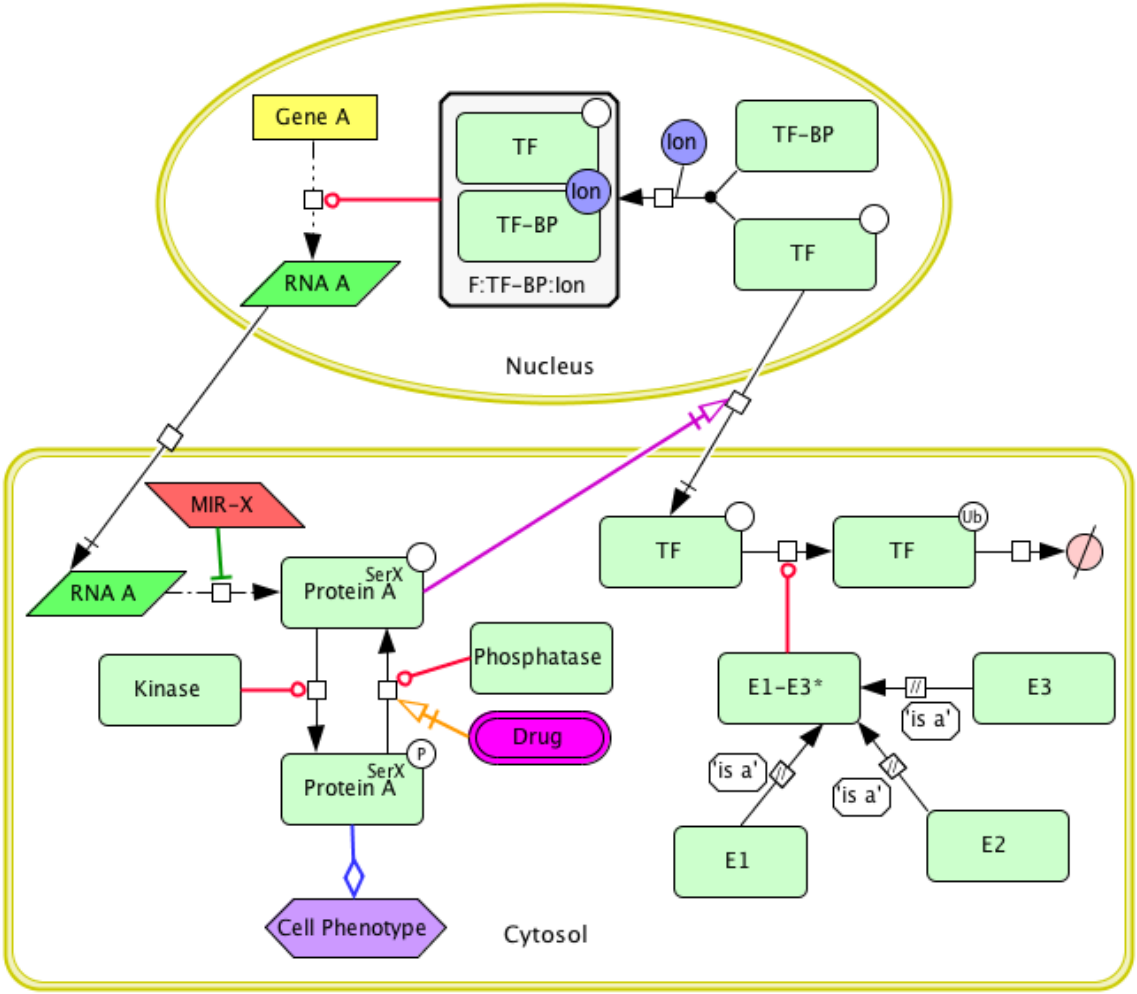
Table 2: Method for studying molecular and genetic interactions

| *Interaction* | *Method* | *Reference* |
|---|---|---|
| Ligand-receptor interactions | Resonance energy transfer (FRET and BRET) | (48)(49) |
| | Flow Cytometric Analysis | (50) |
| Direct protein-protein interactions | Co-immunoprecipitation (CoIP), | (51)(52) |
| | NMR, X-ray crystallography, | (53) |
| | GST-pull down assay | (54) |
| | Tandem affinity purification | (51)(49) |
| | Far Western blotting | (55) |
| | Phage display | (52) |
| | Mass-spectrometry | (56) |
| | Two hybrid assays (yeast and mammalia) | (51) |
| | Functional mutational analysis | (57) |
| Direct protein-DNA interaction (transcription regulation and co-regulation) | Chromatin immunoprecipitation (ChIP) | (58)(59) |
| | DNA footprinting, | (58) |
| | Electrophoretic mobility shift end supershift assays (EMSA) | (58) |
| | Computational prediction of transcription factors binding sites | (60) |
| MicroRNA binding | Direct miRNA binding assay, | (61) |
| | 3 'UTR reporter assay, | (61) |
| | Computational miRNA target prediction | (61) |
| Regulation of expression (mRNA and protein level) | Reverse transcription polymerase chain reaction (RT-PCR) | (62) |
| | Reporter assays | (63) |
| | RNase protection assay | (64) |
| | Nothern blot | (62) |
| | Western blot | (65) |
| | Fluorescence-activated cell sorting (FACS) | (66) |
| Genetic interactions | Genetic knock-out, knock-down, knock in or overexpression of effector molecules | (67)(68) |
| | Synthetic interaction detection assays | (52)(69) |

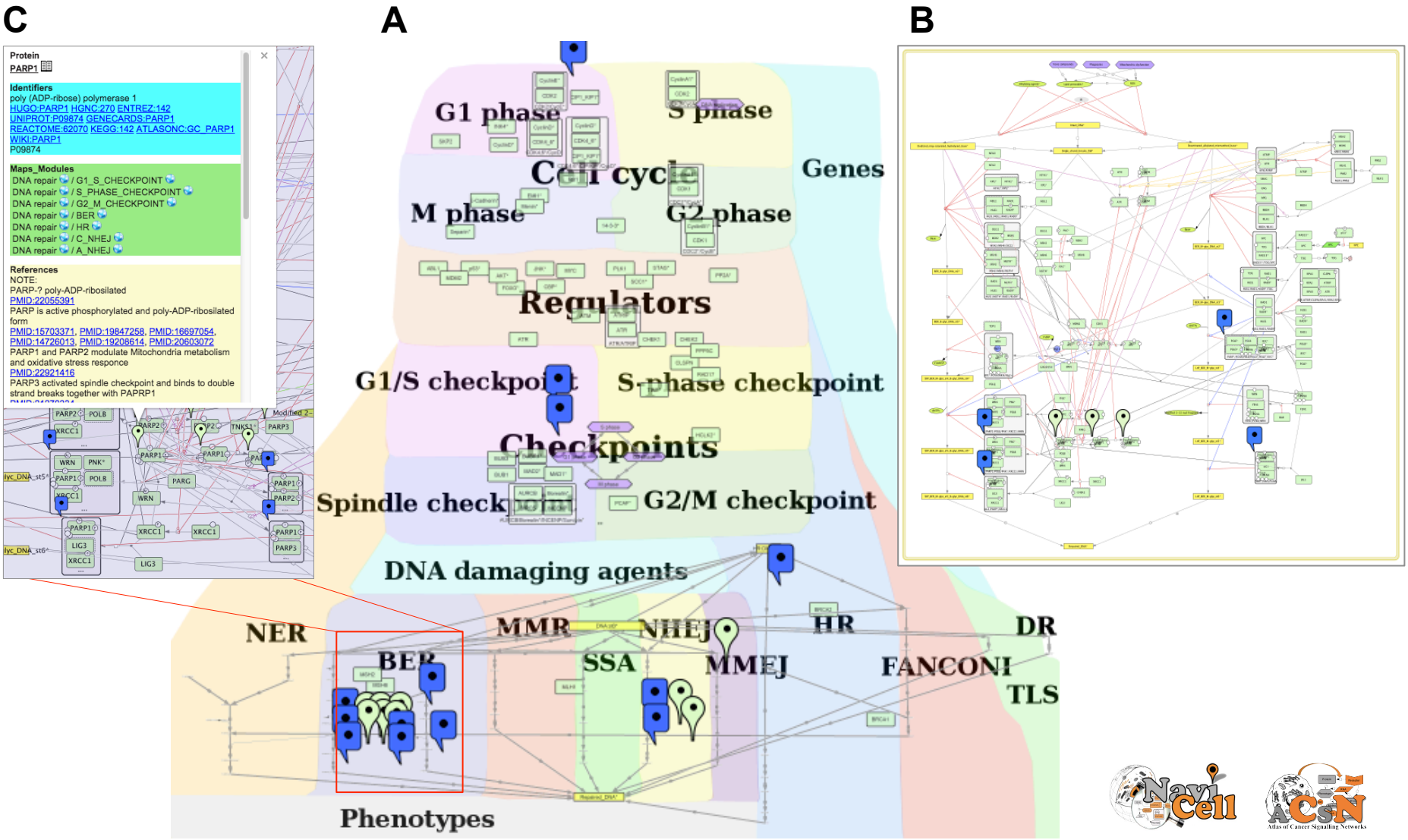
Figure 1


State map purpose
Chose map type
Chose map construction tool
Define map boundaries and layout
Extract information from literature and databases
Construct and annotate map
Visualize map in user friendly format
Validate map by curators
Apply map for data analysis

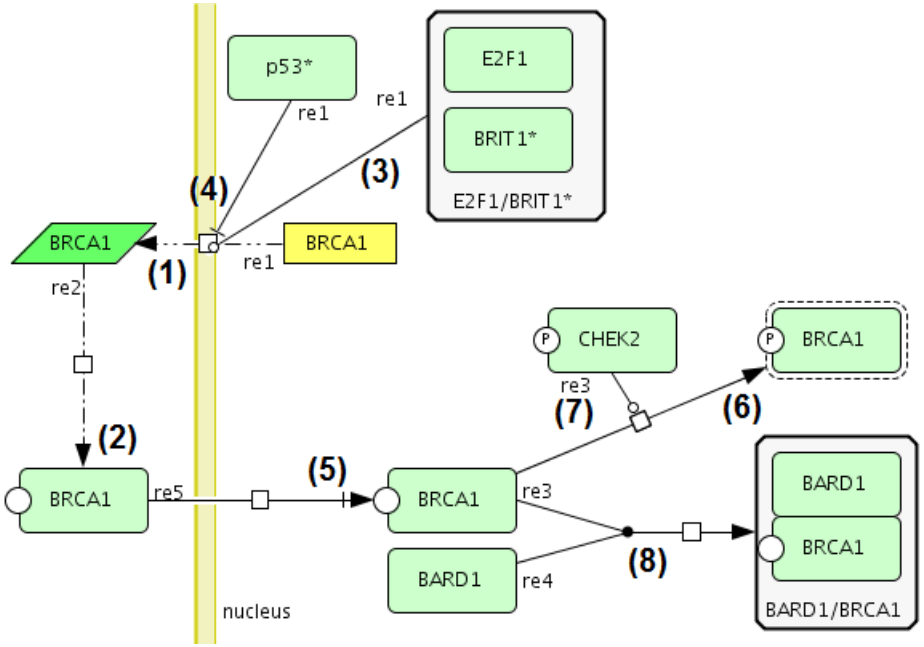
# Figure 2


Gene A
TF
Ion
TF-BP
Ion
TF-BP
TF
RNA A
F:TF-BP:Ion
Nucleus
MIR-X
RNA A
SerX
Protein A
TF
TF
Ub
Phosphatase
Kinase
E1-E3*
E3
'is a'
Drug
P
SerX
Protein A
'is a'
'is a'
E1
E2
Cell Phenotype
Cytosol

# Figure 3


C
A
B
Protein
PARP1
Identifiers
poly (ADP-ribose) polymerase 1
HUGO:PARP1 HGNC:270 ENTREZ:142
UNIPROT:P09874 GENECARDS:PARP1
REACTOME:62070 KEGG:142 ATLASONC:GC_PARP1
WIKI:PARP1
P09874
Maps_Modules
DNA repair / G1_S_CHECKPOINT
DNA repair / S_PHASE_CHECKPOINT
DNA repair / G2_M_CHECKPOINT
DNA repair / BER
DNA repair / HR
DNA repair / C_NHEJ
DNA repair / A_NHEJ
References
NOTE:
PARP-? poly-ADP-ribosilated
PMID:22055391
PARP is active phosphorylated and poly-ADP-ribosilated form
PMID:15703371, PMID:19847258, PMID:16697054, PMID:14726013, PMID:19208614, PMID:20603072
PARP1 and PARP2 modulate Mitochondria metabolism and oxidative stress responce
PMID:22921416
PARP3 activated spindle checkpoint and binds to double strand breaks together with PAPRP1
G1 phase
S phase
Cell cycle
Genes
M phase
G2 phase
Regulators
G1/S checkpoint
S-phase checkpoint
Checkpoints
Spindle checkpoint
G2/M checkpoint
DNA damaging agents
NER
MMR
NHEJ
HR
DR
BER
SSA
MMEJ
FANCONI
TLS
Phenotypes
NaviCell
ACSN
Atlas of Cancer Signalling Networks

# Figure 4


p53*
re1
re1
E2F1
BRIT1*
E2F1/BRIT1*
(4)
(3)
BRCA1
(1)
re1
BRCA1
re2
(2)
CHEK2
re3
(7)
(6)
BRCA1
BRCA1
re5
(5)
BRCA1
re3
BARD1
re4
(8)
BARD1
BRCA1
BARD1/BRCA1
nucleus

# Figure 5

## A

### p53*

Leave a reply

**Protein p53***

**Identifiers**
tumor protein p53
HUGO:TP53 HGNC:11998 ENTREZ:7157 UNIPROT:P04637
GENECARDS:TP53 REACTOME:69487 KEGG:7157
ATLASONC:P53ID88 WIKI:TP53

**Maps_Modules**
EMT / EMT_REGULATORS
Apoptosis / APOPTOSIS_GENES
Apoptosis / MOMP_REGULATION
DNA repair / G1_S_CHECKPOINT
DNA repair / G2_M_CHECKPOINT
Cell cycle / APOPTOSIS_ENTRY
Survival / MAPK
Survival / PI3K_AKT_MTOR
Survival / WNT_CANONICAL
Survival / WNT_NON_CANONICAL

**References**
PMID:21518799
p53 activates MIR200C
PMID:21483453
p53 activates microRNAs
PMID:21336307
p53 activates MIR34
PMID:17823410
PMID:6396087, PMID:21340684, PMID:20457558, PMID:20182602, PMID:20066118
P53 is a nuclear P38 target.
PMID:20506250
PMID:10721693
PMID:19459846

### Participates in reactions:

**As Reactant or Product:**

1. p53*|S15_pholactive@Nucleus → p53*@Nucleus
2. p53*@Nucleus → p53*|S15_pholactive@Nucleus
3. BCL2|unk@Mitochondrial outer membrane + p53*|S15_unk@Cytoplasm → BCL2:p53*@Mitochondrial outer membrane
4. BCL2-XL*@Mitochondrial outer membrane + p53*|S15_unk@Cytoplasm → BCL2-XL*:p53*@Mitochondrial outer membrane
5. p53*@Cytoplasm → p53*@Nucleus
6. rp53*@Nucleus → p53*@Cytoplasm
7. p53*@Nucleus → p53*|Ser15_pho@Nucleus
8. p53*|Ser15_pho@Nucleus → p53*|Ser15_phol Ser20_pho@Nucleus
9. MDM2@Nucleus + p53*@Nucleus → c_s731
10. p53*@Cytoplasm → degraded

## B

### Complex composition:

1. p53*
2. PARP1

PARP1:p53*@default

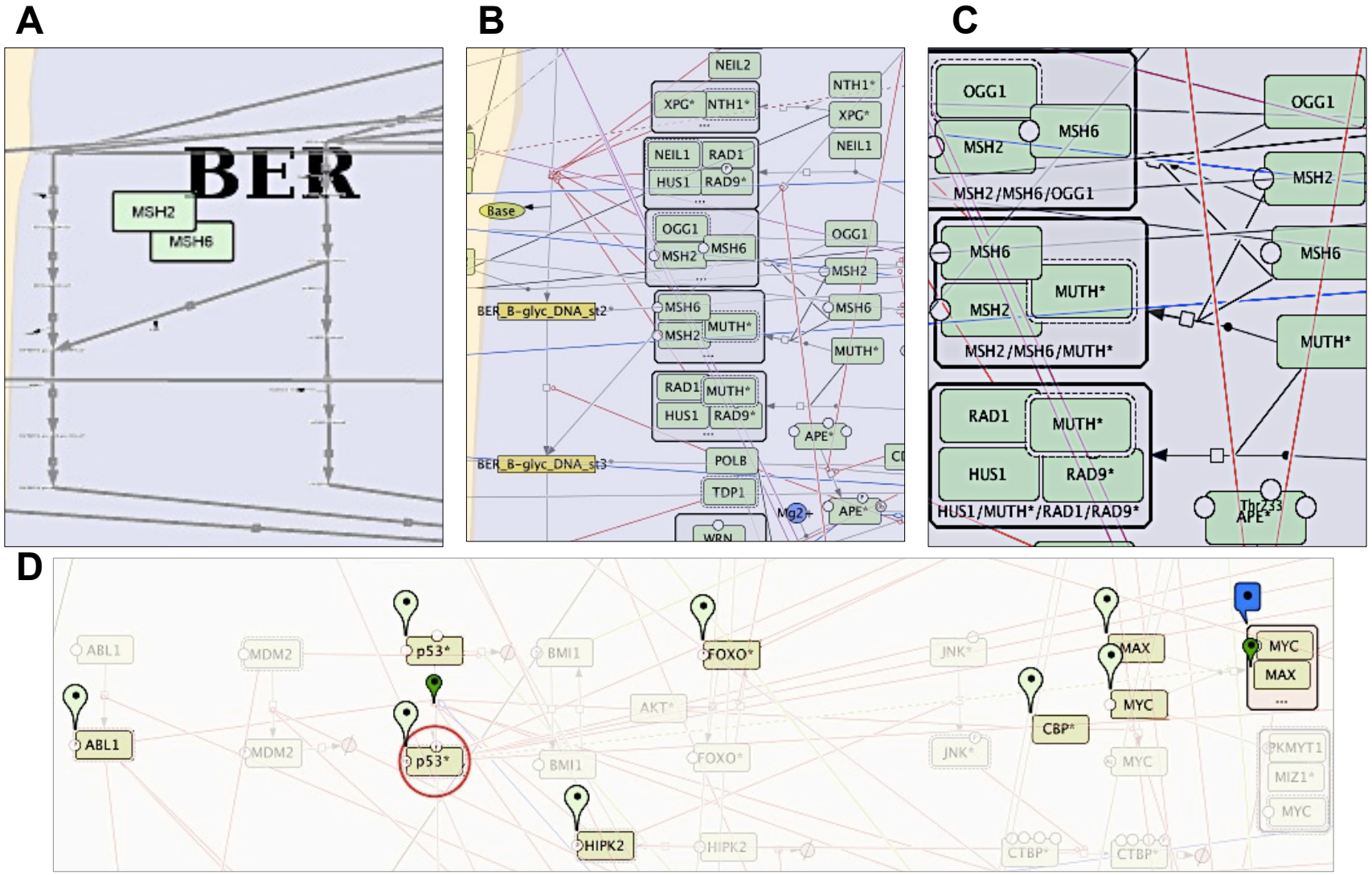
**Identifiers**
NAME:PARP1:p53*

**Maps_Modules**
DNA repair / G1_S_CHECKPOINT

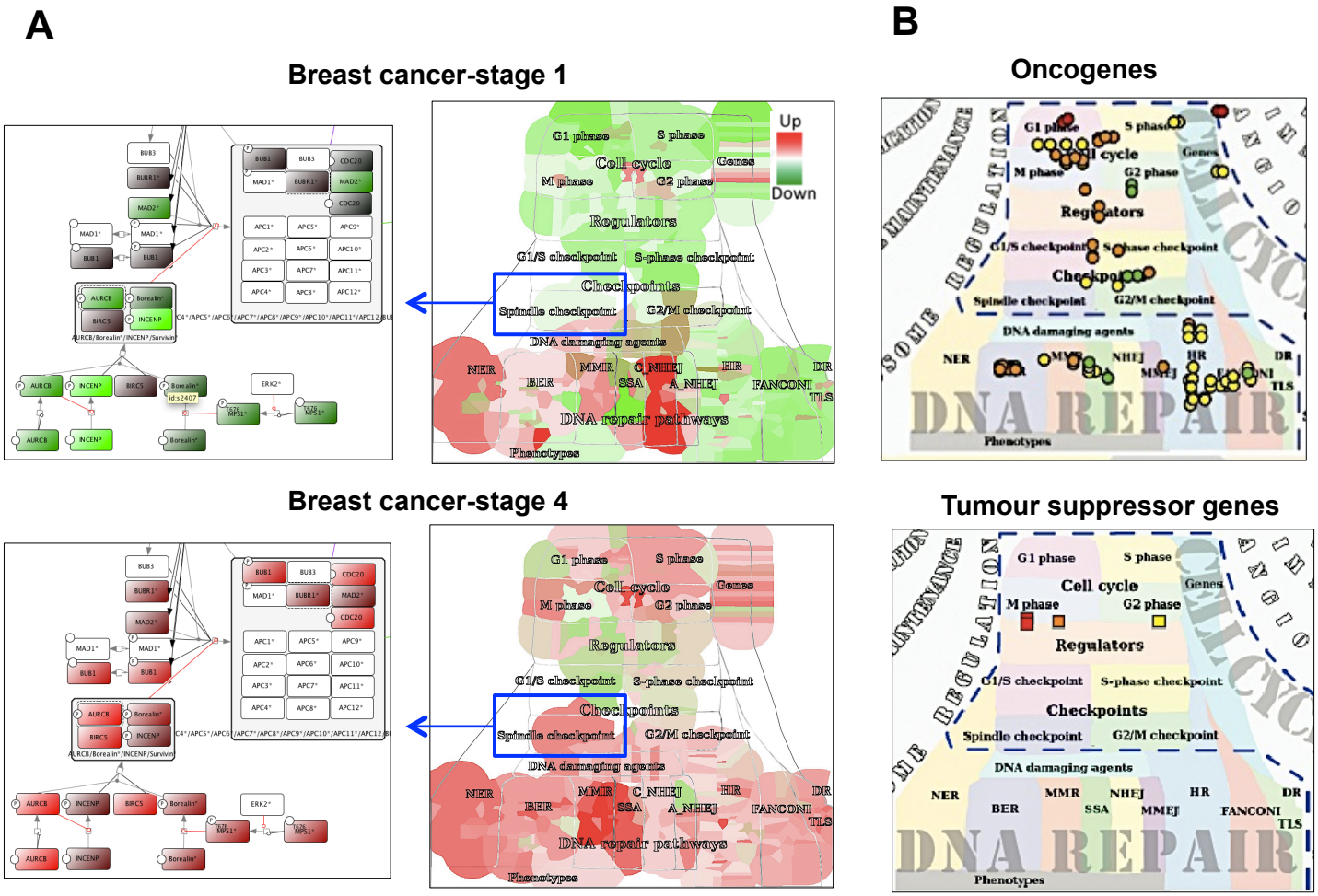
**References**
d_re115( DNA repair ):
PMID:16581787,
For inhibition of S-checkpoint by ATM:
PMID:15175241
Inhibition of Sphase by MRE complex:
PMID:17713585
By CycE*/CDK2: G1/S transition
By CycD*/CDK2/Clp/KIP: checkpoint G1/S

**Confidence**
★★☆☆☆ REF=2 FUNC=5

# Figure 6


A
BER
MSH2
MSH6
B
NEIL2
XPG*
NTH1*
NEIL1
RAD1
HUS1
RAD9*
Base
OGG1
MSH2
MSH6
BER_B-glyc_DNA_st2*
MUTH*
BER_B-glyc_DNA_st3*
POLB
TDP1
APE*
C
OGG1
MSH6
MSH2
MSH2/MSH6/OGG1
MUTH*
MSH2/MSH6/MUTH*
RAD1
HUS1
RAD9*
HUS1/MUTH*/RAD1/RAD9*
Thr233
APE*
D
ABL1
MDM2
p53*
BMI1
FOXO*
AKT*
JNK*
MAX
MYC
CBP*
HIPK2
CTBP*
PKMYT1
MIZ1*

# Figure 7


A
B
Breast cancer-stage 1
Oncogenes
Breast cancer-stage 4
Tumour suppressor genes
Up
Down
G1 phase
S phase
Cell cycle
Genes
M phase
G2 phase
Regulators
G1/S checkpoint
S-phase checkpoint
Checkpoints
Spindle checkpoint
G2/M checkpoint
DNA damaging agents
NER
BER
MMR
C_NHEJ
SSA
A_NHEJ
HR
FANCONI
DR
TLS
DNA repair pathways
Phenotypes
DNA REPAIR