Concurrent Dynamic Visualizations With Expressive Petri Net Representations to Enrich the Understanding of Biological and Pathological Processes: an Application to Signaling Pathways

F. Ramos[*1], C. Hallal[2], A. Nieto[3], D. García[4], J. Berúmen[5], D. Escárcega[*6]

[1, 4, 6] Tecnológico de Monterrey, Campus Cuernavaca.
Autopista del Sol Km 104+060. Col. Real del Puente, C.P. 62790
Xochitepec, Morelos, México.
*fernando.Ramos@itesm.mx; *daescarcega@gmail.com
[2, 3] Universidad Autónoma del Estado de Morelos. Facultad de Farmacia
Av. Universidad 1001. Col. Chamilpa. Cuernavaca, Morelos. C. P. 62209.
[5] Hospital General de México. Dr. Balmis 148, Colonia Doctores, delegación
Cuauhtémoc, C.P. 06726, México, D.F.

ABSTRACT

Dynamic visualizations and expressive representations are needed in systems biology to handle multiple interactions occurring during the biological processes of biopathway representations. Dynamic visualizations allow users an ease of interaction with pathway models. At the same time, representations of biopathways should express how interactions take place. In spite of the fact that diverse databases provide users with pathways, their information and representation are frequently different from each other and show restricted interactions because of their static visualization. An adopted solution is to merge diverse representations to obtain a richer one. However, due to different formats and the multiple links involved in the pathway representations, the merge results frequently in erroneous models and in a tangle web of relations very hard to be manipulated. Instead, this work introduces a concurrent dynamic visualization (CDV) of the same pathway, which is retrieved from different sites and then transformed into Petri net representations to facilitate the understanding of their biological processes by interacting with them. We applied this approach to the analysis of the Notch signaling pathway, associated with cervical cancer; we obtained it from different sources which we compared and manipulated simultaneously by interacting with the provided CDV until the user generated a personalized pathway.

Keywords: Petri nets, dynamic visualization, signaling pathways, pathway databases.

RESUMEN

En biología de sistemas la visualización dinámica y las representaciones expresivas son necesarias para representar interacciones múltiples que ocurren durante los procesos biológicos en bioredes. La visualización dinámica facilita a los usuarios interactuar con modelos de bioredes, mientras que las representaciones deben expresar como se llevan a cabo las interacciones dentro de éstas. A pesar de que diversas bases de datos proveen de redes a los usuarios, generalmente la información y representación contenidas en cada una son diferentes, y la interacción usuario-biored es restringida debido a la visualización estática. Una solución que se ha adoptado es hacer converger varias representaciones para obtener una más completa. Sin embargo, debido al uso de diferentes formatos incompatibles entre ellos y a las múltiples conexiones involucradas en las redes, la integración frecuentemente resulta en modelos erróneos y en una maraña de conexiones representadas en la red que son muy difíciles de analizar y manipular. En este trabajo introducimos la visualización dinámica concurrente (VDC) de una misma vía, la cual es recuperada de diferentes bases de datos y transformada a representaciones en redes de Petri para facilitar el entendimiento de los procesos biológicos y modificar las vías obtenidas interactuando con ellas. Hemos aplicado esta estrategia al análisis de la vía de señalización de Notch, asociada a cáncer cérvicouterino, obteniéndola de tres diferentes fuentes, comparándolas y manipulándolas simultáneamente interactuando con la VDC provista, hasta la generación de una vía personalizada.

1. Introduction

Systems biology models are focused on integrative approaches to understand the functioning of biopathways as a whole, instead of individual pieces. Hence, researchers aim to develop visualization of certain genes and/or proteins involved in pathways [1], as well as the discovery

of altered genes within regulatory mechanisms [2] and aids to comprehend the intricate biopathway crosstalk [3] [4]. Firstly, a system biology approach needs dynamic visualizations, which aim to allow users an ease of interaction with pathway models to manipulate, modify and update them. Secondly, representations of biopathways should be able to express how the multiple interactions occur within them. Thus, expressive models in this domain should be able to represent the interactions taking place within pathways, their dynamic nature as well as the capacity to represent several levels of abstraction. In addition, dynamic visualizations, along with tractable network representations on demand, are easily extendable and enabled for data integration as microarray high-throughput data (HTD) [5].

Pathway databases represent the major resource for computable data about biological processes and for capturing information about the functional interactions of molecular species [6]. Consequently, pathway databases should be done to facilitate a variety of analysis and simulation techniques that could enrich our understanding of biological processes. Conventionally, databases assemble the knowledge obtained within laboratories, computational techniques (such as text mining, information extraction, and retrieval applications for biology) and most of them through rigorous manual curation by experts [7].

On the one hand, a vast number of databases available can provide users with pathways; however, pathways derived from different databases are commonly represented in different incompatible formats each, and with different information concerning the pathway structure, representations and layouts. On the other hand, pathways are analyzed under static visualizations, restricting importantly the interaction between users and pathways [8]. In addition, most of them lack expressive models, making them difficult to understand even by experts. Expressive models in this domain should be able to represent the interactions taking place within pathways, their dynamic nature as well as the capacity to represent several levels of abstraction. Even more, data integration, such as HTD, requires visualizing genes and up-regulated and down-regulated proteins and within signaling pathways.

Since databases make their data available for analysis and reconstruction of pathways, different tools integrate their data as main input. This integration aims to obtain a richer pathway by combining several pathways through operations based on biological rules.

The following section analyzes relevant databases and their formats in order to expose important problems related with the difficulties of integrating pathways, which are derived from different sources (databases) of information [10].

1.1 Relevant databases and their formats

Databases such as Reactome [11], BioCyc collection of databases [12], Pathway Interaction Database (PID) [13], BioModels repository of computable models [14], WikiPathways database [15], Kyoto Encyclopedia of Genes and Genomes (KEGG) [16], and others listed by the Pathguide [17], store thousands of biopathways and molecular interactions of the Homo Sapiens and other organisms. Databases also characterize pathways using visual representations as diagrams or graphics close to 'cartoon' representations, for instance KEGG and Cell Signaling, respectively. Such representations avoid the interaction of the users with the biopathways and are inexpressive for the purpose of dealing with the interactions ocurring within pathways and with their dynamic nature. Indeed, a clear translation of abstract models such as eXtended Markup Language (XML) representations to tractable and executable models, e.g., Petri nets, is needed to improve the researchers' understanding of cellular processes [9].

Complementary, each database defines either standard formats for data exchange or proprietary file definitions. For both cases, key issues of databases involve data integration, model exchange and software interoperability [18]. The concept of integration is twofold in this work: 1) the fact of carrying out some operations between at least two pathways to obtain a new one that supposedly satisfies certain needs of the user; 2) the fact of incorporating high-throughput data into pathways to analyze its behavior as these data change. This section deals with the first option.

Moreover, only certain databases provide computer-readable formats needed for automatic retrieval and processing. Thanks to their positioning among the standard formats used in biological pathway modeling, we study in this work the following languages: Biological Pathway Exchange (BioPAX) [19], Systems Biology Markup Language (SBML) [20] and the KEGG Markup Language (KGML).

BioPAX and SBML represent the main standard formats of pathways supported by different databases (including Reactome and BioCyc) and KGML applies only to the KEGG database. BioPAX is a protocol for the specification and representation of cell signaling pathways, gene-regulatory networks, protein-protein interactions and other types of molecular interaction data, while SBML is an open XML-based language for describing models adequate to research in many areas of computational biology including cell signaling pathways, metabolic pathways and gene regulation. Finally, KGML is an exchange format of the KEGG pathway maps that are manually curated, drawn and updated only by KEGG specialists. KGML enables automatic drawing of KEGG pathways and provides facilities for computational analysis and modeling of protein networks and signaling networks, additionally other XML representations have been proposed [21].

1.2 The problem of integration

For integration purposes, databases and XML-based formats face different challenges. First, due to the fact that different XML formats are involved, there is a strong need for tools for the automatic translation of different pathways in a biological meaningful way, though this is not a trivial task. Several annotation problems and inconsistencies between formats exist in such a way that in many cases manual intervention is needed to achieve the integration. Thus, precise annotation of genes and chemical reactions taking place within signaling pathways are required to appropriately solve expressiveness problems. Moreover, databases could offer dissimilar representations of the same biological process, which complicates the interpretation and analysis of pathways. Finally, KGML defines a layout for pathway reconstruction using a position of biological elements in coordinates (x, y), while SBML

introduces a layout on level 3 and BioPAX, in nature, lacks a display layout.

As mentioned before, a systems biology approach requires handling the knowledge about genes and/or protein functions and interactions in the whole context of cellular machinery by tracing metabolic or signaling pathways, annotations of molecular functions, and a physical distribution of the components as a graphical layout. For this purpose, pathway databases [6] and dynamic visualization tools [9] provide researchers with qualitative tractable representations of cellular mechanisms [22].

This work introduces a methodology to enable the integration of BioPAX, SBML and KGML files to create enriched and biological meaningful signaling models or metabolic pathways. In the following sections, firstly, a review of related works is presented; then the methodology proposed is described and finally the application of this methodology is presented.

The main challenge of data integration is to identify differences between formats from diverse databases and integrate them into a meaningful model to facilitate the understanding of biological processes. This work deals with the description of signaling or metabolic pathways focusing on three formats: SBML, BioPAX and KGML. Although the three of them are used to describe the same biopathways, their purpose is semantically different: BioPAX and KGML are used for qualitative representations [19], while SBML is used for quantitative modeling [20]. Integration of XML representations of biopathways is an open issue in bioinformatics and systems biology. Authors focus on the integration of model representations to generate new enriched models; such is the case of the Systems Biology Pathway eXchange (SBPAX) alluding to SBML and BioPAX [23]. SBPAX creates one-to-one mappings. The objective of SBPAX is to provide a flexible expanded description around essential pathway data. In spite of the fact that SBPAX proposes to bridge between SBML and BioPAX, databases and tools require an adoption period before considering SBPAX a new standard.

Other tools focus on working with KGML data and generate XML files, for instance, the SBMLmerge

software [24] aids the user in combining models of biological subsystems with larger biological networks. The program associates all of the model elements with unique identifiers pointing to databases such as KEGG or Gene Ontology. Then, during merging, SBMLmerge detects and resolves various syntactic and semantic problems. At the end of the merging process it results in a new SBML model. Meanwhile, the Sig2BioPAX tool [25] was developed as an open-source command line tool that takes tabular data as input describing different types of molecular biochemical interactions. This tool converts such interactions into a BioPAX level 3 OWL format. KEGGconverter is a web application which uses KGML files as source in order to construct integrated pathway SBML models for simulation purposes. Such solutions partially address the problem of integration by focusing only on generating output files, leaving visualization, extension, manipulation and representation of models apart. Furthermore, the automatic integration of data from different repositories remains as key problem for new tools and databases. On this matter, Cytoscape [26] is capable of importing SBML, BioPAX [27] and KGML files using the appropriate plug-in, which leads the user to make a manual integration process. Finally, for manipulation and representation, other authors propose the direct conversion from KGML files into computational representations, for instance, KEGGgraph [28], which converts KEGG pathways (KGML data) into graph models and provides a collection of tools for graph manipulation, while MoVisPP [29] is an online resource that transforms KGML files into online Petri nets, though, as static visual representations. Several drawbacks, appearing when the integration mechanisms are applied, motivate the development of alternative approaches, which should take advantage of both dynamic visualizations and expressive representations. The main drawback concerns the resultant integrated pathway, whose final product is frequently a model represented by a tangle web of relations very hard to be manipulated [30]. Another relevant drawback is related with the lack of dynamic visualizations avoiding users the possibility of interacting with pathway models. Lastly, a considerable number of sources of

information provide no expressive representations making difficult the task of understanding how the interactions take place within pathways.

The work presented in this paper proposes a methodology dealing with the main following steps: 1) the automatic transformation of SBML, BioPax and KGML files into Petri net models; 2) a dynamic visualization of several pathways, the same pathway derived from different sources that are displayed concurrently to facilitate the user interaction aiming to obtain his own network that satisfies his/her needs; 3) the study of the pathways, both the sources of information and the user's pathway, under the integration of high-throughput data.

2. Methodology

In this section, the proposed methodology of concurrent dynamic visualization (CDV) is introduced; firstly, a brief introduction to Petri nets, then the corresponding mappings of SBML, BioPAX and KGML into Petri nets and finally the option of integrating microarray high-throughput data. See Figure 1 for a diagram of the proposed methodology.
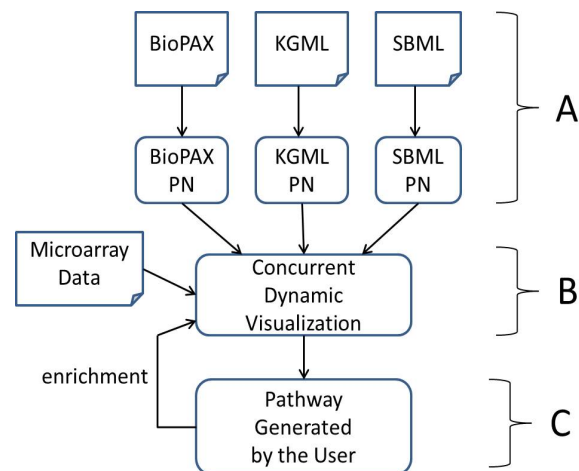

Figure 1. Diagram of the proposed methodology. (A) Transformation from XML representations into Petri net (PN) representations. (B) The CDV and integration of microarray data. (C) The user's pathway generated by interacting with the CDV.

2.1 Mapping SBML, KGML and BioPAX into a Petri net model

2.1.1 A brief introduction to Petri nets

PNs were developed by Carl Petri to model asynchronous distributed systems [31]. Ever since, PNs have been applied to different systems, among which the application of process modeling stands out. The basic PN is a directed bipartite graph consisting of two different nodes called *places* and *transitions*. Places model typically passive system elements such as conditions, states, or biological components, i.e., genes or proteins. Transitions stand generally for active system elements as events, or chemical reactions as activation, inactivation and post-tranductional modifications. Edges connect places to transition if the transitions depend on the state of the place. Transitions are connected to places if the transition has some effect on the place. In order to execute a process, each place can be marked by a number of tokens that represent resources; the tokens distributed throughout the places of the graph at the time *t*, called marking, represent the state of the system. PNs have recently become a promising tool among the various methods employed for the modeling and interpretation of biochemical networks [32] [33] [34] because their capacity to express state changes associated with their biological processes and the representations of pathways capable of expressing the interactions that take place within them, which is a key issue in systems biology modeling.

Tokens representing conditions or resources (molecules) that should be available within places to fire the transitions which can represent enzymatic activities such as phosphorylation or others chemical reactions related with the behavior of the biological processes associated with the dynamics of pathways. Thus, once the tokens are present in places, transitions linked to these places can be fired denoting in this way the execution of the network.

2.1.2 SBML, KGML and BioPAX

Each format requires an analysis of syntax and semantics to properly transform an XML format into a PN model. The SBML can be used to describe both quantitative and qualitative aspects of reactions that modify the entities. BioPAX is a specific XML format type for representing and sharing information concerning biomolecular networks, and does not deal with mathematical relations but provides many details concerning the molecular interactions within the cell. The KGML format is widely used in life science research and in some applications, information systems and developer libraries that support KGML, and also contains layout information.

2.1.2.1 Mapping SBML models into Petri nets

An SBML model consists of a set of chemical species that are linked by reactions. Species include simple molecules and large molecules such as RNAs or proteins. Each species is located in a compartment that describes the reaction environment. The reactions can transform one entity into another, transport entities between compartments, or describe the binding of the species. This is the general structure of an SBML model, though a number of more detailed features could be assigned to the objects of an SBML model. During translation of SBML files to PNs, each species is mapped into a place, while reactions are characterized as transitions. Compartments are not considered as part of the PN model, but they can be incorporated manually. The topology of the pathway is created automatically by using the positions of entities from the SBML file. However, this does not guarantee the identification of all locations, therefore the intervention of users may be necessary. Finally, parsing SBML models was achieved by using the JSBML library.

2.1.2.2 Mapping KGML model into Petri nets

The KGML files for metabolic pathway maps contain two types of graph objects; enzymes linked by relations and compounds linked by reactions; whereas KGML files for nonmetabolic pathway maps contain proteins, compounds, or pathways linked by relations. Relations correspond to different protein interactions such as activation, inhibition, expression, and so on. Also, the KGML files specify drawing information about the graphics objects. In the PN model proposed, places stand for enzymes, compounds, proteins and subpathways. Meanwhile, transitions represent relations such as interactions and reactions, among others. However, KGML files do not include all information of the pathway model, like relations with pathways. Figure 2 shows part of

```
1   <?xml version="1.0"?>
2   <!DOCTYPE pathway SYSTEM "http://www.genome.jp/kegg/xml/KGML_v0.7.1_.dtd">
3   <!-- Creation date: Mar 31, 2009 11:06:20 +0900 (JST) -->
4   <pathway name="path:hsa04330" org="hsa" number="04330"
5            title="Notch signaling pathway"
6            image="http://www.genome.jp/kegg/pathway/hsa/hsa04330.png"
7            link="http://www.genome.jp/kegg-bin/show_pathway?hsa04330">
8       <entry id="1" name="hsa:23385" type="gene"
9            link="http://www.kegg.jp/dbget-bin/www_bget?hsa:23385">
10           <graphics name="NCSTN, APH2, KIAA0253" fgcolor="#000000" bgcolor="#BFFFBF"
11               type="rectangle" x="263" y="290" width="46" height="17"/>
12       </entry>
13       <entry id="2" name="hsa:51107" type="gene"
14            link="http://www.kegg.jp/dbget-bin/www_bget?hsa:51107">
15           <graphics name="APH1A, 6530402N02Rik, APH-1, APH-1A, CGI-78" fgcolor="#000000" bgcolor="#BFFFBF"
16               type="rectangle" x="309" y="290" width="46" height="17"/>
17       </entry>
18       <entry id="3" name="hsa:5663 hsa:5664" type="gene"
19            link="http://www.kegg.jp/dbget-bin/www_bget?hsa:5663+hsa:5664">
20           <graphics name="PSEN1, AD3, FAD, PS1, S182..." fgcolor="#000000" bgcolor="#BFFFBF"
21               type="rectangle" x="309" y="273" width="46" height="17"/>
22       </entry>
23       <entry id="4" name="path:hsa04010" type="map"
24            link="http://www.kegg.jp/dbget-bin/www_bget?hsa04010">
25           <graphics name="MAPK signaling pathway" fgcolor="#000000" bgcolor="#FFFFFF"
26               type="roundrectangle" x="394" y="395" width="109" height="40"/>
27       </entry>
28       <entry id="5" name="hsa:3065 hsa:3066" type="gene"
29            link="http://www.kegg.jp/dbget-bin/www_bget?hsa:3065+hsa:3066">
30           <graphics name="HDAC1, DKFZp686H12203, GON-10, HD1, RPD3, RPD3L1..." fgcolor="#000000" bgcolor="#BFFFBF"
31               type="rectangle" x="663" y="325" width="46" height="17"/>
32       </entry>
33       <entry id="6" name="hsa:9541" type="gene"
34            link="http://www.kegg.jp/dbget-bin/www_bget?hsa:9541">
35           <graphics name="CIR1, CIR" fgcolor="#000000" bgcolor="#BFFFBF"
36               type="rectangle" x="707" y="302" width="46" height="17"/>
37       </entry>
38       <entry id="7" name="ko:K04497" type="ortholog"
```

Figure 2. A fragment of the KGML file used to obtain the network shown in Figure 7(a).

a KGML file that served to generate the model shown in Figure 5(a). In this particular case, the relation that links the MAPK signaling pathway with the rest of the pathway is missing. This lacking relations will appear as isolated places in the final PN. A parser for KGML data was developed using the dom4j library.

2.1.2.3 Mapping BioPAX model into Petri nets

BioPAX describes its objects using a hierarchy. The basic building blocks of a BioPAX model are the physical entity, interaction and pathway classes. Pathway represents a set of interactions that together form a network model. Each Interaction describes relationships between physical entities. Details concerning the interactions and entities are given by subclasses that describe various interactions and physical entity types. Physical entities are various objects that may interact, including molecular complexes, proteins, DNA, RNA, or small molecules. Each interaction type limits the possible roles and the number of potentially interacting entities. Similar to SBML, and KGML mappings, physical entities are mapped into

places and interactions into transitions. Nevertheless, the BioPAX format does not include layout data, thus the layout is obtained from KGML format. First, BioPAX entities, which match in both formats, inherit the KGML layout. Consequently, the drawbacks of KGML such as the lack of relations, will be inherited to BioPAX. Then, related entities are joined together and finally the rest of the components are displayed for the user's intervention. BioPAX files were parsed using the Paxtools utilities.

2.2 Translations of XML based models

Translation of pathways from different repositories involves operations with the pathway elements, entities and relations. Proposed solutions search for a transformation among models into a unified format [23], visual combination of models in single representation and inferring new elements within pathways. Nevertheless, the merge of different models into a single representation tends to complicate visualization and interpretation, for instance, see Figure 4. The work presented in this paper proposes to interact with graphical notations

that have been transformed from XML inputs into PN representations. The PN elements, places, transitions and relations minimize the interpretation of symbols. In addition, PN transitions provide users with a more expressive model. Afterwards, users interact with the pathways represented by PNs, which are displayed concurrently to facilitate their manipulation. In other words, pathways are not merged to build a single pathway model but the participant models, at least two and as much as the user can handle, are displayed at once in the same window. As the final result of the interaction, users generate their own model that aims to satisfy their needs or requirements. Layouts of each model are preserved and, through a color notation, common elements are emphasized.

In the following section, an example of the methodology is presented for the analysis and interpretation of a Notch signaling pathway. Models of the Notch pathway are obtained from databases Wikipathways (BioPAX), Reactome (SBML) and KEGG (KGML) and mapped into PNs. Lastly, it is introduced a concurrent visualization model of the Notch signaling pathway to aid researchers in the interpretation of the pathway by visualizing differences among models, extension of models and integration with high-throughput data. The methodology was implemented within a tool called BioPathXplore. This tool is available on request by addressing to daescarcega@gmail.com and fernando.ramos@itesm.mx.

3. Application of the CDV

This section introduces the application of the proposed methodology through the analysis of the Notch pathway. We aim firstly to show the disadvantages of handling different representations of the same pathway, provided by two or more different databases, as well as the inconvenient of working with merged models of pathways, which results in tangled web of interactions very difficult to be analyzed and interpreted. The application consists in generating a pathway that satisfies the need of a specific user consisting in linking a set of notch receptors with relevant molecules reported formally in the literature on Notch pathways associated with cervical cancer (CC).

3.1 The Notch signaling pathway

The Notch signaling pathway (NSP) is a highly conserved pathway for cell-cell communication. NSP is involved in the regulation of cell fate, cell proliferation and cell death in development. NSP is mediated by proteolysis and does not appear to involve any secondary messengers [35]. However, depending on the cellular context, the amplitude and timing of Notch activity can be regulated by posttranslational modifications to ligands and receptors and their trafficking. Therefore, successful targeting of NSP in cancer will require a thorough understanding of Notch regulation and the context-dependent interactions between Notch and other relevant pathways [36]. Understanding these interactions will increase the biologists' skills to design safer and more effective therapies for cancer and other human diseases [37] [38] [39]. The NSP, first studied in the Drosophila [35], consists of Notch receptors, ligands, negative and positive modifiers, and transcription factors [40]. The NSP importance for the understanding of cancer depends on a clear interpretation of the elements and relations that compile the NSP. Hence, different databases provide diagram representations of this and other pathways. Nonetheless, pathway representations, derived from different sources of information, differ in completeness, accuracy and description detail. In this example, three representations of the NSP were obtained from the KEGG, WikiPathways and Reactome databases, see Figure 3. The three representations describe the canonical NSP, though each database includes different details of the pathway and share data using dissimilar XML standards.

The KEGG database distributes pathway data using the KEGG Markup Language (KGML), Wikipathways provides data using different standards, including BioPax Level 3, and Reactome shares pathway data using the Systems Biology Markup Language (SBML). Online representations restrict users to freely manipulate the pathway or integrate new experimental data since online pathways are usually static models. Therefore, users require tractable models. Additionally, biologists have to choose an initial model and then extend it according to data or create a merged model from different sources. The latter requires additional work since merged models complicate the understanding and manipulation of pathways, an example is provided in the following section.
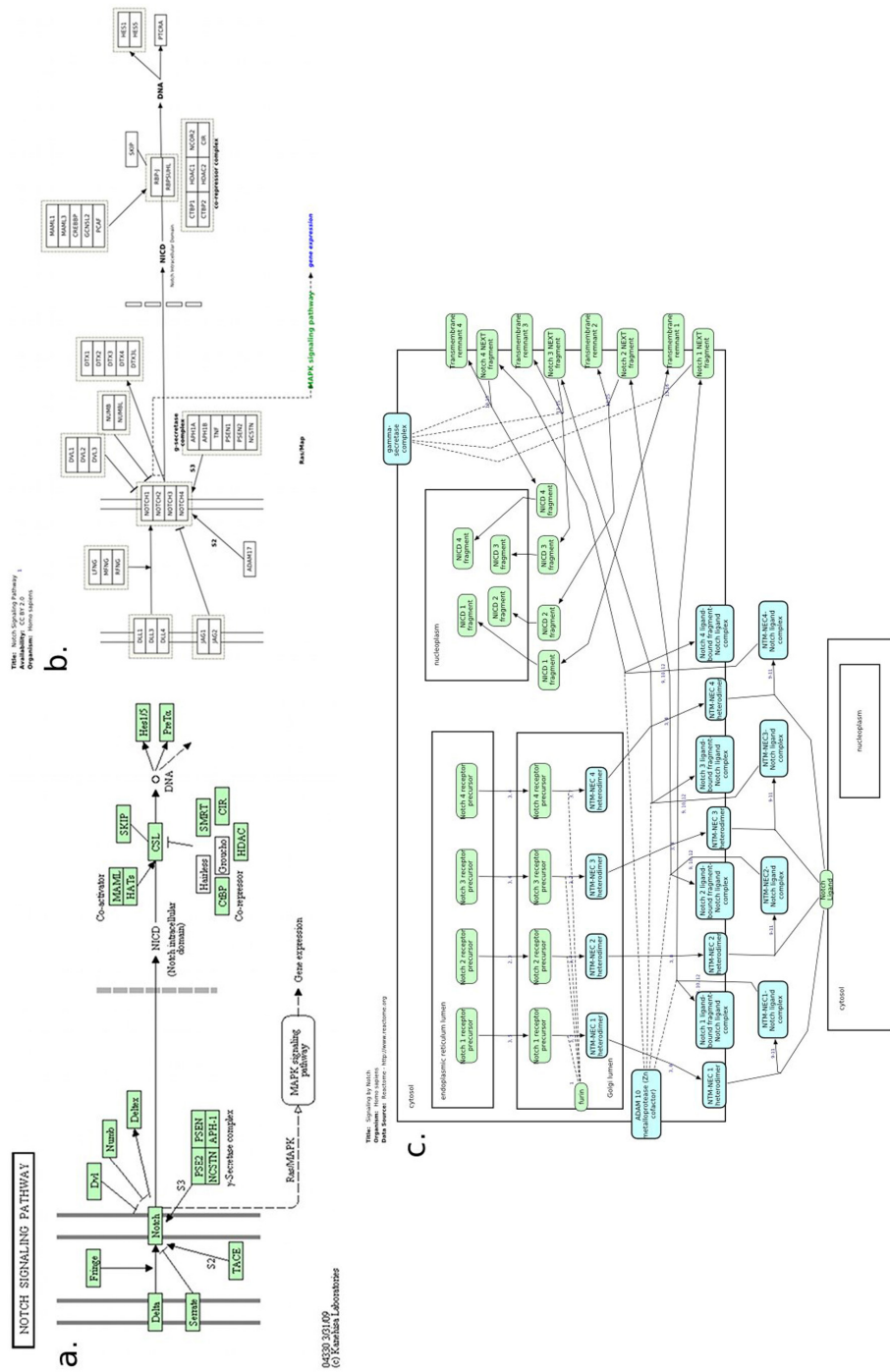
Figure 3. This figure shows (a) the representation of the NSP by KEGG, (b) the NSP modeled by WikiPathways and (c) the NSP updated by Reactome. Larger high-resolution images of each representation are available at their corresponding sites.

3.2 Merged models

Before dealing with the merged models, it is convenient to establish the difference between integration and merging when when it comes to talk about considering several pathways to be combined to obtain a new one. When an integration mechanism takes place between two or more pathways, a mapping process, component by component, is carried out. Then the new pathway is obtained retaining the common components. This mechanism results in several drawbacks: 1) certain components represent the same molecule in the pathways to be integrated, however they are not named identically, consequently they will not be considered in the resulting pathway; 2) given that the mapping between pathways does not consider the links between components certain relations disappear, obtaining graphs with an important number of isolated components; 3) the integration mechanism is automatic, without curation by the users. The resultant pathway is a very complex one to be analyzed and interpreted.

In our version of a merge mechanism, the role of the user is relevant, because he/she makes manually some curation operations. Nevertheless, users retrieve the pathways to be curated automatically in a Petri net representation.

In our merge mechanism, the user interacts with pathways through the CDV to curate manually the pathway to be obtained. The advantage of this method is that the user chooses the most convenient pathway and can perform the number of manipulation needed by acting directly on the pathways. For instance, the user can link components, add lacking components, exchange components between pathways, among the most important operations, until the requirements are satisfied.

Currently, automatic visualization, interactive exploration methods, and dynamic visualization are desired as the state-of-the-art for the systems biology approach [41]. Dynamic visualization refers to automatically transform databases information into tractable computational representations, either graphs or PNs, on the time that is required. Moreover, dynamic representations allow the integration of experimental data for visualization for qualitative and quantitative analyses. It is noteworthy that integration, at least in systems biology, could refer to add data to a computational model, as microarray expression levels, or to combining or merging pathways into a single representation.

Cytoscape is an open-source software platform very useful for visualizing complex networks and also provides the means to merge models. Nevertheless, by creating a merged model combining BioPAX, KGML and SBML data using a "plug-in", the generated graph loses the essence of the NSP, see Figure 4. Even for experts such tangled representations remain intractable for interpretation and extension. Hence, to avoid the limitations of static models and drawbacks of merged models, this work proposes a Concurrent Dynamic Visualization (CDV) with expressive representations of pathways based on PNs. In the CDV, a transformation from networks derived from different formats (SBML, KGML and BioPAX) into Petri nets is made automatically. Thus, the representation of the pathways in Petri nets allows users a curation of pathways interacting with them in a unique representation as well as taking advantage of the transitions that denote expressively semantic aspects of the dynamic nature of pathways. The CDV provides researchers with a tool to efficiently compare models, their individual layouts, shared elements, remarks the differences between models, and also creates a personalized network, enriched and updated. Furthermore, integration of high-throughput data in the CDV enables the analysis of genes altered in each model. Figure 5 shows concurrently the three participant pathways in this example, which are represented in PNs.

A biological pathway has an inherent complexity because of their important number of elements (genes or proteins) and another important number of relations between them. For instance, what could be a small pathway composed of about 200 elements, its number of potential relations between two elements belonging to the pathway would be $n(n-1)/2$, which represents a considerable number of relations (19,900). We have to point out that going from a specific element to another is one of the aspects that we are interested in when performing the analysis of a pathway. This process establishes a relation and, at the same time, composes a subgraph or a part of the pathway.

However, the complexity derived from the merging mechanism in Cytoscape involves the following problems: 1) Cytoscape performs an automatic mapping between the pathways to be merged. The mapping is carried out one-to-one. We have to recall that the mapping is performed between different formats, therefore each format can code, frequently, its elements differently. Thus, when the mapping is carried out, certain elements are not being mapped. 2) Discarding the relations in the mapping results in incoherent tangle of relations topologically incomprehensible to be analyzed and interpreted. In conclusion, the resulting complex pathway in Cytoscape is due to its inherent complexity plus the merging mechanism at the moment of performing the mapping.

As shown in Figures 5a and 5b, the automatic translation reveals that XML formats lack certain details that are not included in the release files, e.g., Figure 5a shows that the place called "Mapk signaling pathway" is not bound to a transition because such relation is not included in the KGML file. Another issue, shown in Figure 5c, is the integration of high-throughput data. Usually, genes

are identified by the standard "gene symbol" though not all databases or XML files, e.g., BioPAX, name each gene using the gene symbol. For such reason, none element in the 5c model was selected for modeling in the customized network. The concurrent visualization of several models enables the easy visualization of elements participating and the difference between models. Hence, users can interact with the most appropriate model or they can even create an enriched model from the model that best fits their needs complemented by elements of the other models; furthermore, they can update the model by incorporating more recent published data, as seen in Figure 6. Certainly, the concurrent display of more than three pathways could be not easy to be handled. In such case, the users will display the networks in small windows and they can choose the pathways of their interest and amplify them to facilitate their visualization. Another action that the user can do is to work with groups of three pathways.

In the next section, the analysis of NSP in the study of CC is used as an example of the application of CDV.
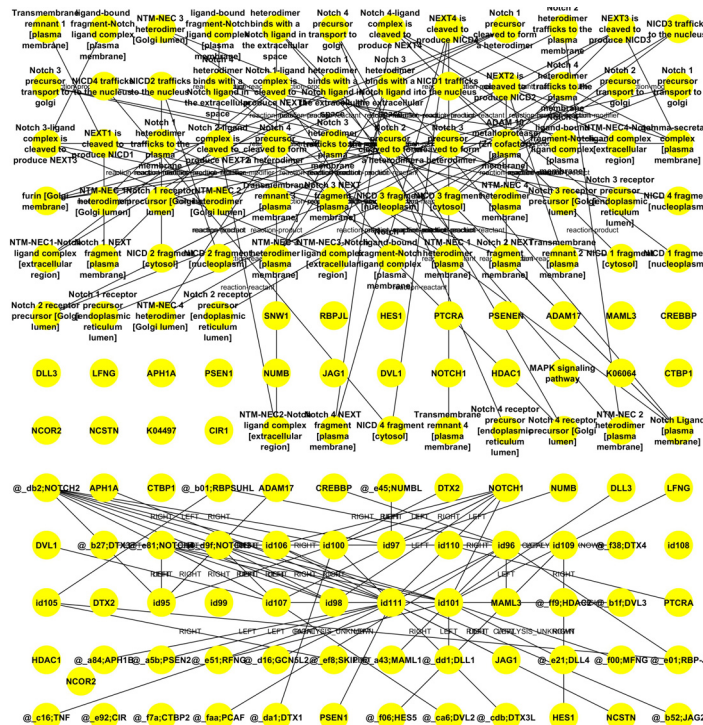

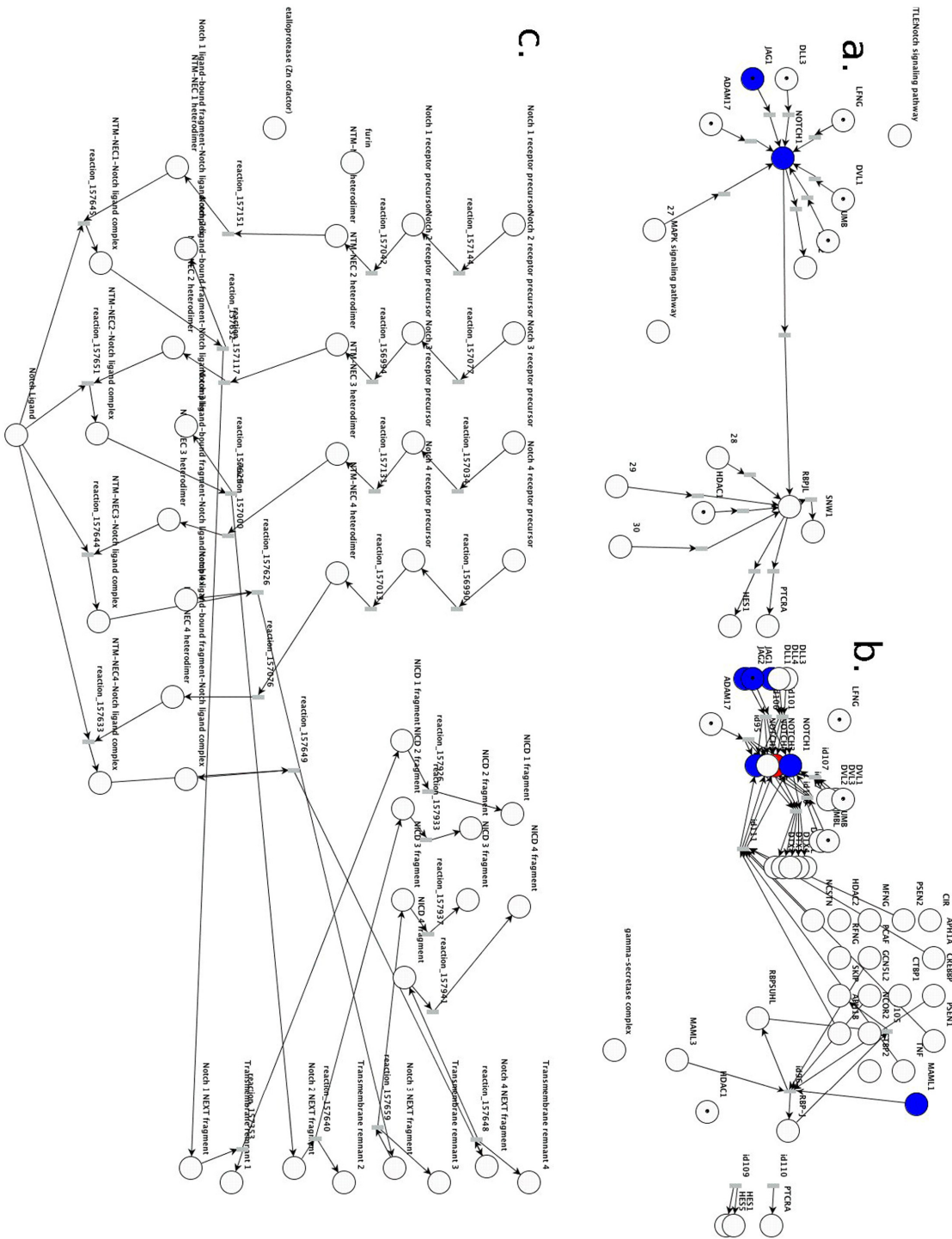Figure 4. NSP merged model by Cytoscape using the SBML, KGML and BioPAX formats.

Figure 5. Concurrent dynamic model of the Notch signaling pathway. Network (a) presents the KGML PN, (b) the mapped BioPAX PN, and (c) the SBML PN. Additionally, in this work, places in blue represent up-regulated genes, whereas places in red stand for down-regulated genes down.

3.3 Application of the CDV to the NSP associated with cervical cancer (CC)

Cervical cancer (CC) pathogenesis is closely related with the infection of the Human Papilloma Virus (HPV), and is sexually transmitted. 70% of infections clear within a year [42]. Though, persistent HPV infections lead to a sequel of various grades of cervical dysplasia that, ultimately, leads to cervical cancer (CC) [43]. Different proteins and pathways have been associated with CC, nevertheless, much remains uncovered. As an example of how CDV can aid to solve a biological need in a pathway related with CC, let us evoke a study of the NSP and its relation with the CC, where different isoforms of the Notch receptor, referred to as Notch 1, Notch 2, Notch 3 and Notch 4, have been formally reported. Notch family members play a role in a variety of developmental processes by controlling cell fate through a signaling pathway that regulates interactions between physically adjacent cells [44] [45] [46]. Additionally, up or down regulation of other molecules involved in NSP have been reported, e.g., Jagged1, Jagged2, Delta1 and MAML1 [47]. For this reason, the user's need is to generate a network linking the elements mentioned above. The generation of a model from a single source, for instance KEGG, supplies an incomplete representation (Figure 5a). Using concurrent visualization, other proteins involved are identified in the NSP (Figure 5b) and integrated into the KEGG model. Thus, using CDV, a manually curated pathway is generated; combining the KGML and BioPAX data into a single NSP, see Figure 6.
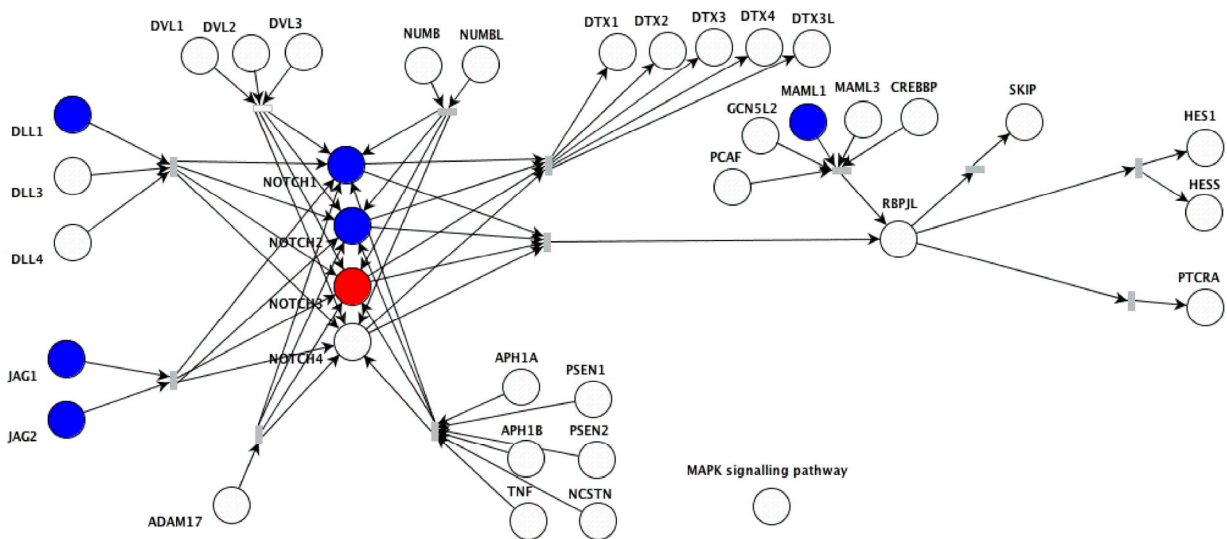

Figure 6. NSP model generated using CDV. Places in blue represent genes up regulated, whereas, places in red stand for genes down regulated. Additionally, transitions indicate the type of relation between places (genes or proteins). Transitions in blue represent activation or inhibition, transitions in dark green represent a binding/association relation, and transitions in light green indicate an indirect effect relation and transitions in red denote an expression relation.

Modeling with a PN allows representing models at different abstraction levels, depending on the users need. As a simpler model to be interpreted, Figure 7, derived from the model described in Figure 6, emphasizes the genes altered within the NSP. It is noteworthy, that models complement each other, providing richer models that cannot be achieved with a single source. Systems Biology requires tools that enable the construction and validation of models. Using holistic and comparative approaches, as the proposed CDV (Figure 6), and reduced models, as in Figure 7, complement the process to biologically evaluate candidate target genes and their relations within a biological process. The study of more specific networks may yield further insight into the cancer cascades [48].

The tool of dynamic visualization is focused on biological pathways in general. In this work a biological pathway is represented by a special graph known as Petri net, where nodes/places, which can be molecules such as genes or proteins, are linked by transitions. Therefore, biological pathways could be either signaling pathways, metabolic pathways or gene regulation pathways.

4. Discussion and Conclusions

The systems biology approach emphasizes the importance of disposing of appropriate pathway models to analyze and interpret them in order to understand their functioning and behavior. A key aim of systems biology is to systematically catalogue all molecules and their interactions within a living cell. Various types of interaction networks emerge from the sum of these interactions, including signaling, metabolic and transcription-regulatory networks [49]. CDV can be applied to the study of all these kind of networks.
Diverse formats are used in systems biology to describe signaling or metabolic pathways, either to assist users in modeling and simulation (SBML) or to collect biological pathways (BioPAX and KGML). Even though XML formats have elements in common, when they describe central data related to biopathways, pathways derived from the most important databases do not coincide, neither in the number of elements nor in the structure layout, which affects importantly the performance of the studies. The solution proposed by the community is to build methodologies to integrate pathways.
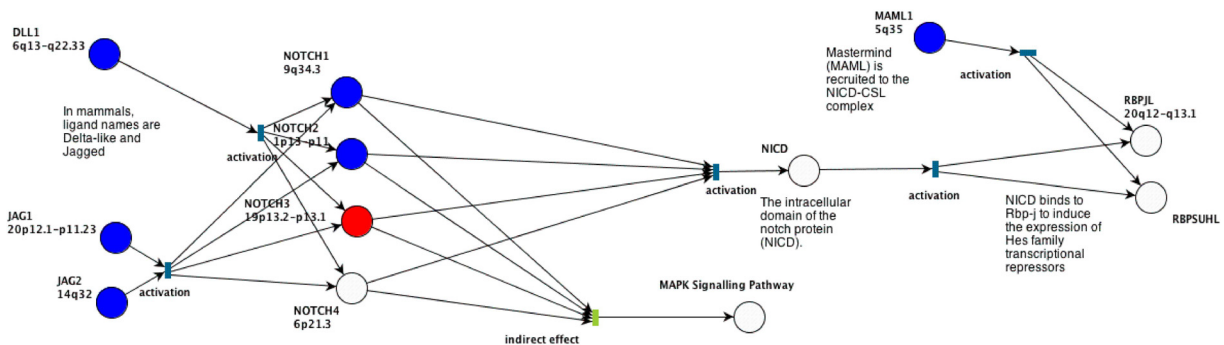

Figure 7. A reduced model of the NSP. A lower level of abstraction of the model of NSP shown in Figure 6, focusing on up- and down-regulated genes (blue and red places, respectively).

Nevertheless, the resulting pathway models are very complex, thus complicating importantly the visualization and interaction of pathways models with users.

In order to cope with this problem, instead of obtaining a merged pathway, we proposed a Concurrent dynamic visualization that displays the pathways participating in the study in the same window under the same expressive PN representation. Moreover, the proposed concurrent visualization demonstrates that a single representation of pathways is not adequate since complementary knowledge is provided by other databases and their corresponding formats and by data obtained from the literature. This alternative approach will allow users to interact with models in an easier way and compare the pathway representations for a better understanding of the biological processes involved. This paper provided an application of the proposed CDV for the understanding of the NSP associated with CC, which it would be worth applying in complex processes as those taking place in immune cell signaling, that have given rise to systems immunology [50]. Cancer cells express antigens that are not found on normal cells and these antigens appear foreign to the immune system, causing immune cells to attack the transformed cancer cells to identify and eliminate tumors. In cervical cancer, some of the antigens expressed are derived from humans, and when a cancer cell evades the immune system become carcinogenic. The possibility that cancers can be eradicated by specific immune responses has been the impetus for a large body of work in the field of tumor immunology [51]. Although high-throughput data derived from experimental technologies have been recently developed to generate large datasets to unravel the cellular complexity, cellular properties such as growth, proliferation, and immune response still remain largely elusive. To understand such complex and dynamic behavior of living systems, the development of systems biology approaches is required [52]. This work contributes to the study of the highly complex but coordinated modular networks of the immune system cell signaling and could address a better directed tumor immunotherapy.

As future work, three modules will work together within the same framework to enhance the tools for the analysis, interpretation and construction of personalized pathways. These modules are the following: the CVD described in this work; the integration of HTD into pathways to analyze their behavior in response to massive real data such as gene expression levels derived from information sources of complex diseases such as cervical cancer or others; and finally, the experiences that have been carried out to obtained a personalized pathway will be stored in a memory of experiences that will be retrieved in similar situations by using a case based reasoning module.

References

[1] Likić V.A., McConville M.J., Lithgow T. & Bacic A., Systems Biology: The Next Frontier for Bioinformatics, Advances in Bioinformatics, Vol. 2010, November 2010.

[2] Hornberg J. J., Bruggeman F.J., Westerhoff H., & Lankelma J., Cancer: A Systems Biology disease, BioSystems, Vol. 83, No. 2-3, February/March, 2006, pp.81–90.

[3] Wong, D.J. et al., Revealing targeted therapy for human cancer by gene module maps, Cancer Res., Vol. 68, No. 2, January, 2008, pp. 369-378.

[4] Pawson, T. et al., Oncogenic re-wiring of cellular signaling pathways, Oncogene, Vol. 26, No.9, February, 2007, pp.1268-1275.

[5] Itadani, H. et al., Can systems biology understand pathway activation? Gene expression signatures as surrogate markers for understanding the complexity of pathway activation, Curr. Genomics, Vol. 9, No. 5, August, 2008, pp. 349-360.

[6] OoiH. S., Schneider G., Lim T.T., Chan Y.L., Eisenhaber B. & Eisenhaber F., Biomolecular pathway databases, Methods Mol Biol, Vol 609, No. 1, 2010, pp. 129-144.

[7] Huang,Y.J. et al., Targeting the human cancer pathway protein interaction network by structural genomics. Mol. Cell. Proteomics, Vol. 10, No. 8, October, 2008, pp. 2048–2060.

[8] Suderman M. & Hallett M., Tools for visually exploring biological networks. Bioinformatics, Vol. 23, No. 20, October, 2007, pp.2651-2659.

[9] Sonnenschein C, Soto AM. Why Systems Biology and Cancer? Semin Cancer Biol.Semin Cancer Biol, Vol. 21, No. 3, June, 2011, pp.147- 149.

[10] Frolkis A. et. al. SMPDB: The Small Molecule Pathway Database. Nucleic Acids Res, Vol. 38 (Database issue), November, 2010, D480-D487.

[11] Vastrik I., D'Eustachio P., Schemidt E., et al., Reactome: a knowledge base of biologic pathways and processes, Genome Biol, Vol. 8, No. 3,March, 2007, R39.

[12] Karp P.D., Ouzounis C.A., Moore C., et al., 'Expansion of the BioCyc collection of pathway/genome databases to 160 genomes', Nucleic Acids Res, ,Vol. 33, No. 19, March, 2007, pp. 6083–6089.

[13] Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. PID: the Pathway Interaction Database. Nucleic Acids Res, Vol. 37(Database issue), October, 2008, D674-D679.

[14] Le Nove`re N., Borstein B., Broicher A., et al.: BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems, Nucleic Acids Res, Vol. 34 (Database issue), January, 2006, D689–D691.

[15] Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. PLoS Biol,Vol. 6, No. 7, July, 2008, xe184.

[16] Ogata H., Goto S., Sato K., et al., KEGG: Kyoto Encyclopedia of genes and genomes, Nucleic Acids Res, Vol. 27, No. 1, January, 1999, pp. 29–34. Database is accessible at: http://www.genome.jp/kegg/

[17] Bader GD, Cary MP, Sander C. Pathguide: a pathway resource list, Nucleic Acids Res, Vol. 1, No.34 (Database issue), January, 2006, D504-D506.

[18] Krallinger M, Valencia A, Hirschman L. Linking genes to literature: text mining, information extraction, and retrieval applications for biology, Genome Biol,Vol. 9, September, 2008, Suppl 2:S8.

[19] Demir E, Cary MP, Paley S, Fukuda K, Lemer C, Vastrik I, Wu G, D'EustachioP, Schaefer C, Luciano J, et al., The BioPAX community standard for pathway data sharing, Nat Biotechnol, Vol. 28, No. 9, September, 2010, pp. 935-942.

[20] Hucka M, Finney A, Sauro HM, et al., The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, Bioinformatics. Vol. 19, No. 4, March, 2003, pp. 524-531.

[21] Sauro HM, Bergmann FT., Standards and ontologies in computational systems biology, Essays Biochem., Vol. 45, 2008, pp. 211-222.

[22] Edelman L.B., Eddy J.A. & Price N.D. In silico models of cancer, Wiley Interdiscip Rev SystBiol Med, Vol. 2, No. 4, pp. July/August, 2010 pp. 4384-59.

[23] Ruebenacker O, Moraru II, Schaff JC, Blinov ML. Integrating BioPAX pathway knowledge with SBML models, IET SystBiol., Vol. 3, No. 5, September, 2009, pp. 317-328.

[24] Schulz M, Uhlendorf J, Klipp E, Liebermeister W. SBMLmerge, a system for combining biochemical network models,Genome Inform., Vol. 17, No. 1, 2006, pp. 62-71.

[25] Webb RL, Ma'ayan A. Sig2BioPAX: Java tool for converting flat files to BioPAX Level 3 format. Source Code BiolMed.,March, 2011, doi: 10.1186/1751-0473-6-5.

[26] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T.,Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res., Vol.13, No. 11, November, 2003, pp. 2498-2504.

[27] Zinovyev A, Viara E, Calzone L, Barillot E.,BiNoM: a Cytoscapeplugin for manipulating and analyzing biological networks, Bioinformatics, Vol. 24, No. 6, November, 2007, pp. 876-877.

[28] Zhang JD, Wiemann S. KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. Bioinformatics, Vol. 25, No. 11, June, 2009, pp. 1470-1471.

[29] Chen M, Hariharaputran S, Hofestädt R, Kormeier B and Spangardt S., Petri Net Models for the Semi-automatic Construction of Large Scale Biological Networks, Natural Computing, Vol. 10, No. 3, 2009, DOI 10.1007/s11047-009-9151-y

[30] Praneenararat T, Takagi T, Iwasaki W. Interactive, multiscale navigation of large and complicated biological networks, Bioinformatics, Vol. 27, No. 8, April, 2011, pp. 1121-1127.

[31] Petri C.A., Kommunikation mit Automaten, Ph.D. Thesis, Institut für instrumentelle Mathematik, Bonn, 1962.

[32] Pinney JW, Westhead DR, McConkey GA., Petri Net representations in systems biology. Biochem Soc Trans., Vol. 31, No. 6, December, 2003, pp. 1513-1515.

[33] Chaouiya C., Petri net modelling of biological networks, Brief Bioinform., Vol. 8, No. 4, July, 2007, pp. 210-219.

[34] Nagasaki M, Saito A, Jeong E, Li C, Kojima K, Ikeda E, Miyano S., Cell illustrator 4.0: a computational platform for systems biology. Stud Health Technol Inform., Vol. 162, 2011, pp. 160-181.

[35] Artavanis-Tsakonas S, Muskavitch MA., Notch: the past, the present, and the future. Curr Top Dev Biol., Vol. 92, September, 2010, pp. 1-29.

[36] Allenspach EJ, Maillard I, Aster JC, Pear WS., Notch signaling in cancer, Cancer Biol Ther., Vol. 1, No. 5, September/October, 2002, pp. 466-476.

[37] Stockhausen MT, Kristoffersen K, Poulsen HS., The functional role of Notch signaling in human gliomas, Neuro Oncol., Vol.12, No. 2, February, 2010, pp. 199-211.

[38] Westhoff B., ColalucaI.N.,D'Ario G., Donzelli M., Tosoni D., Volorio S., Pelosi G., Spaggiari L., Mazzarol G., Viale G., Pece S., Di Fiore P.P., Alterations of the Notch pathway in lung cancer,Vol. 6, No. 52, December, 2009, pp. 22293-22298.

[39] Cau E, Blader P., Notch activity in the nervous system: to switch or not switch?, Neural Dev., Vol. 2, No.4,October, 2009, pp. 36.

[40] Bray SJ., Notch signalling: a simple pathway becomes complex, Nat Rev Mol Cell Biol., Vol. 7, No. 9, September, 2006, pp. 678-689.

[41] Koch I., Reisig W., Schreiber F., Modeling in Systems Biology, The Petri Net Approach, 1stEd.,Springer, 2011, pp.

42] zurHausen H., Papillomaviruses in human cancers, Proc Assoc Am Physicians, Vol. 111, No. 6,November/December, 1999, pp. 581-587.

[43] Psyrri A, DiMaio D., Human papillomavirus in cervical and head-and-neck cancer, Nat Clin Pract Oncol., Vol. 5, No. 1, January, 2008, pp. 24-31.

[44] Talora C, Cialfi S, Segatto O, Morrone S, Kim Choi J, Frati L, et al., Constitutively active Notch1 induces growth arrest of HPV-positive cervical cancer cells via separate signaling pathways, Exp Cell Res., Vol. 305, No. 2,May,2005, pp. 343–354.

[45] Gray GE, Mann RS, Mitsiadis E, Henrique D, Carcangiu ML, Banks A, et al., Human ligands of the Notch receptor, Am J Pathol., Vol. 154, No. 3, March, 1999, pp. 785-794.

[46] Yeasmin S, Nakayama K, Rahman MT, Rahman M, Ishikawa M, Iida K, Otsuki Y, Kobayashi H, Nakayama S, Miyazaki K., Expression of nuclear Notch3 in cervical squamous cell carcinomas and its association with adverse clinical outcomes, Gynecol Oncol., Vol. 117, No. 3, June, 2010, pp. 409-416.

[47] Miele L, Golde T, Osborne B., Notch signaling in cancer, Curr Mol Med., Vol. 6, No. 8, December, 2006, pp. 905-918.

[48] L L Chen, N Blumm, N A Christakis, A-L Barabási, and T S Deisboeck: Cancer metastasis networks and the prediction of progression patterns British Journal of Cancer (2009) 101, 749–758.

[49] Albert-László Barabási & Zoltán N. Oltvai Network biology: understanding the cell's functional organization Nature Reviews Genetics 5, 101-113 (February 2004).

[50] Charlotte Schubert. Systems immunology: Complexity captured Nature 473, 113-114 (2011).

[51] S. V. Chiplunkar: The immune system and cancer. CURRENT SCIENCE, 81(5):542-48, 2001.

[52] Kumar Selvarajoo, Masaru Tomita and Masa Tsuchiya: Can Complex Cellular processes be governed by simple linear rules? Journal of Bioinformatics and Computational Biology 7(1):243–268, 2009.

Janus kinase/signal transducers and activators of transcription (JAK/STAT) pathway. Kyoto Encyclopedia of Genes and Genomes (KEGG). Available from: http://www.genome.jp/kegg/pathway/hsa/hsa04630.html

Jak/Stat Pathway. Cell Signaling Technology. Available from: http://www.cellsignal.com/pathways/jak-stat.jsp

kgmlreader. Cytoscape plugin for importing KEGG pathway files (KGML). Available from: http://code.google.com/p/kgmlreader/

JSBML. The Systems Biology Markup Language. Available from: http://sbml.org/Software/JSBML

Dom4j. yxymedia. Available from: http://www.dom4j.org/

Paxtools. Java library designed for accessing and manipulating data in BioPAX format. Computational Biology Center at MSKCC. Available from: http://www.biopax.org/paxtools.php