Erschienen in: Journal of Integrative Bioinformatics ; 7 (2010), 1. - 148
https://dx.doi.org/10.1515/jib-2010-148

Visualization and Analysis of a Cardio Vascular Disease- and MUPP1-related Biological Network combining Text Mining and Data Warehouse Approaches

Björn Sommer[1*], Evgeny S. Tiys[2], Benjamin Kormeier[1], Klaus Hippe[1], Sebastian J. Janowski[1], Timofey V. Ivanisenko[2], Anatoly O. Bragin[2], Patrizio Arrigo[3], Pavel S. Demenkov[4], Alexey V. Kochetov[2], Vladimir A. Ivanisenko[2], Nikolay A. Kolchanov[2], Ralf Hofestädt[1]

[1]Bio-/Medical Informatics Department, Bielefeld University, Universitätsstraße 25,
33615 Bielefeld, Germany
[2] Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences,
Lavrentyeva 10, 630090 Novosibirsk, Russia
[3] CNR ISMAC, Via De Marini 6, Genoa, Italy
[4] Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences,
4 Acad. Koptyug avenue, 630090 Novosibirsk, Russia

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Summary

Detailed investigation of socially important diseases with modern experimental methods has resulted in the generation of large volume of valuable data. However, analysis and interpretation of this data needs application of efficient computational techniques and systems biology approaches. In particular, the techniques allowing the reconstruction of associative networks of various biological objects and events can be useful. In this publication, the combination of different techniques to create such a network associated with an abstract cell environment is discussed in order to gain insights into the functional as well as spatial interrelationships. It is shown that experimentally gained knowledge enriched with data warehouse content and text mining data can be used for the reconstruction and localization of a cardiovascular disease developing network beginning with MUPP1/MPDZ (multi-PDZ domain protein).

1 Introduction

More than 4000 human diseases are known and defined [1]. Regarding the medical characteristics or main features one can see that any disease is defined or specified by particular symptoms and/or laboratory parameters. In practice the diagnosis problem is based on the fact that a lot of symptoms, such as fever, are related to many diseases. Therefore, the diagnostic procedure will always be a differential process which will produce a set of possible diseases. Furthermore, the so-called personalized medicine makes the problem of finding the patient-relevant diagnosis and recommendation for its treatment much more difficult. Based on the data of molecular biology, the development of new and more efficient tools for medical diagnosis and therapy process is becoming possible.

Today, more and more diseases can be reduced to simple metabolic processes, which more or less are based on mutations in related genes. OMIM [2] examplifies of well-known information systems which exactly represent this kind of knowledge. Overall, there are more than 1000 molecular database and information systems which represent various molecular and phenotypic data. These information resources were designed on the basis of either automatic data extraction or manual

[*] To whom correspondence should be addressed. E-mail: bjoern@CELLmicrocosmos.org

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-2tro85talstf3

annotation and curation. Behind these information systems there is one more specific and powerful information system which will present molecular and medical disease knowledge. The MEDLINE information system represents all relevant publications (abstracts and in the near future a complete listing of papers) which are relevant for molecular medicine or biomedicine. Overall one have access to more than 1000 powerful database and information systems which will help identify molecular knowledge about any disease. Furthermore, this data can be supported, enriched or fused by the extension of text and data mining techniques which allow the automatic extraction of medical and molecular knowledge from the PubMed system, which includes all relevant scientific results. Therefore, it is possible to construct or predict the metabolic network for any disease. This kind of work is relatively new and during the last years different database integration and data mining systems have been implemented. However, the problem of all these systems is, that data integration and mining tools will produce networks, which are too complex. Therefore, the development of special filter systems or visualization tools is a necessary step in understanding and analyzing these complex metabolic disease-related networks. In this paper it will be demonstrated how the data integration and data mining tools can be used to gather the molecular knowledge on diseases.

The focus of this application is on Cardiovascular diseases (CVDs), and more precisely the dilated cardiomyopathy, which is the leading cause of death in developed countries. Based on the experimental identification of a CVD relevant protein, two protein-protein interaction networks were constructed by using the network visualization and analysis tool VANESA [3] and the text mining tool ANDVisio [4], which is also able to identify the localization of network components. This localization information was extended, combined with the created networks and finally visualized in 3D by the CELLmicrocosmos 4.2 PathwayIntegration (CmPI) [5].

2 Basics

2.1 Metabolic disease networks

Much attention has been recently focused on the metabolic aspects of Cardiovascular diseases (CVDs). The discovery of new CVDs specific molecular targets promoted the investigation of proteins functional roles in their specific pathways. It is quite complex to evaluate the weight of each trigger factor (metabolism, hormones, exogenous factors, etc.) on CVDs emergence. Epidemiological studies constitute the starting point for molecular medicine screening. The advent of high throughput analytical techniques (DNA chip, protein arrays, molecular imaging) has improved the capability to screen new candidate target proteins (genes). The relations of metabolic pathways of a sample coming from patients affected with dilated cardiomyopathy (DCM) was the basis of study for this publication. The proteome analysis is based on experimental data on which integrative bioinformatics approaches have been applied to characterize a specific functional pathway deregulated in the pathological sample. In this study, the combination of data warehouse with text mining approaches is demonstrated by using different software applications.

2.2 Data integration

Since industrial research of molecular biology questions starting with the Human Genome Project, one of the main challenges in bioinformatics is the integration of molecular data. Today high throughput analysis delivers data of complete genomes, for instance short sequences of all genes in an organism or expression patterns of thousands of a cell in shortest time. Analysis of these high throughput data by manual investigation using publications or relevant databases is no longer

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

possible. Consequently, biologist has to be supported by tools and methods that can accumulate experimental data with complementary data sources, estimate the data and compare or classify these data. This challenge leads us to the problem of database integration.

Typically, data of genomes, genes, proteins, enzymes, chemical compounds, diseases, etcetera is stored in databases with worldwide availability. A good overview of important databases is provided by the annual special issue of Nucleic Acids Research [6]. The number of molecular databases is continuously increasing in the last decade. Molecular biological data has a high semantic heterogeneity that is caused by (experimental) data extracted from a series of experiments. Molecular biology deals with complex problems, hence enormous and versatile data is produced. The total number of databases, as well the data itself, is continuously increasing, as is the distribution and heterogeneity of the data. Particularly, data heterogeneity causes big problems in molecular biological data integration. Technical heterogeneity is caused by a high number of different formats and interfaces of the different data sources. Furthermore, the data is usually not available in a standard format which causes structural heterogeneity. Moreover, there is a level of semantic heterogeneity, because of missing standards and consensus for basic biological terms. In addition to the problems of molecular biological databases there are some more in data integration. Usually, data sources of an integrated system are distributed. That means, each and every source is located on separate systems and different locations. The distribution of several data sources leads automatically to the problem of autonomy. Regarding data integration, autonomy means independence of the data source that refers to access, configuration, development and administration.

The major problem of data integration is heterogeneity that is caused by autonomy. Moreover, distribution can also cause heterogeneity, but not generally. The development of an integrated database system is a complex task. Particularly, if a large number of heterogeneous databases have to be integrated. Data warehouses (DWH) are one of the widely used structures for database integration. For that purpose a software infrastructure for building life science data warehouses using different common relational database management systems is introduced. The BioDWH [7] system is realized as a Java-based open source application that is supported on different platforms with an installed Java Runtime Environment (JRE). BioDWH is a flexible DWH infrastructure for bioinformatics; it is independent from the underlying RDBMS. Furthermore, the data warehouse approach provides an easy-to-use graphical user interface for administration and configuration. The main feature of the BioDWH tool-kit is the automatic storage and visualization of data content and information from different public databases into a homogeneous and consistent data warehouse. It provides integrated data from different widely-used life science databases, such as BRENDA [8], EMBL [9], ENZYME [10], GO [11], HPRD [12], KEGG [13], OMIM [2], Reactome [14], SCOP [15], Transfac [16], Transpath [17] and UniProt [18] and microarray data. Additionally, configuration of the infrastructure and its tools is also possible via XML, because it is human readable, well-formatted, easy accessible and standardized. A logging mechanism observes the integration process and begins a simple recovery process to guarantee a consistent state of the data warehouse. The data warehouse BioDWH addresses the aforementioned aspects of data integration.

Based on the data from the warehouse infrastructure, the CardioVINEdb [19], a data warehouse approach, was developed to browse and explore life science data. Furthermore, a DWH system to search integrated life science data and simple navigation called DAWIS-M.D. was implemented based on the life science data from the BioDWH toolkit. In addition, the network editor VANESA uses the data from DAWIS-M.D to generate biological networks and enrich them with additional information.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

2.3 Text mining

Work with scientific literature and factual databases is required for research in every knowledge area. The size of this information pool is immense and expands exponentially. The PubMed database alone contains over 19 million abstracts [20] and their number increases annually by 1 million. Thus, the development of computer algorithms for automated text analysis (text-mining) becomes a timely task.

In microbiology and biomedicine the most important type of interactions is molecular-genetic. Extraction of facts concerning such interactions from literature and providing access to them can be divided in two directions: manual analysis of literature and automated data analysis with text-mining techniques. Manual curation is most accurate but also is a highly time-consuming task [21]. The automated data analysis methods are not so accurate but allow the processing of larger amounts of data in less time and usually are used in three main tasks: extraction of data on molecular-genetic interactions between biological objects, discovery of new associations between different sources of biological information and biological data classification.

For the automated extraction of molecular-genetic interactions between biological objects and new associations from the literature various methods exist:

- methods based on the co-occurrence of objects,
- natural language methods based on the deep syntactic analysis of single sentences (full parsing),
- natural language methods based on rules and templates analysis (shallow parsing).

Co-occurrence is based on the statistically important values of the joint frequency of names of biological objects in texts. The main advantage of this method is that it is easy to implement and achieves good results in regards to search completeness, but it is not so accurate. Moreover such approach does not allow detection of different parameters of interconnections between objects, such as type of interaction and its direction. PUBGENE [22] and FACTA [23] are based on this approach. In BioGene [24] it was implemented for prioritization of genes.

“Full-parsing” is based on the definition of the language with formal grammar. There are many various types of grammars as well as descriptions of the complete sentence structure. The main limitation of this approach is its low time efficiency so it cannot be used for all tasks. MedScan from the PathwayStudio [25], GeneScene [26] are examples of a full parsing-based system, also it has been used by Fayruzov [27] for protein relation extraction. A text analysis algorithm based on formal grammar implemented in this system shows high accuracy but is also very time-consuming.

“Shallow parsing” (deep parsing) is based on the extraction of information from sentences by using the partial connection between words in the sentence with the help of specified rules and templates. A SUISEKI system is based on it [28]. In the Chilibot system [29], the deep parsing method was implemented for the classification of extracted proteins (genes) from PubMed abstracts. The relations between two proteins may result from the existence or non-existence of an interaction or co-location.

The biological data classification task is based on the idea of classification of various sources of data by user-specified features. For the solution of this task, different algorithms are applicable, such as hidden Markov models [30] or Bayesian networks. The BioBayesNet server [31], based on Bayesian networks can be used for such classification.

Most of the modern text-mining systems are combining various methods. The ANDVisio, ALI BABA [32] and PolySearch [33] systems are based on the combination of co-ocurrence and shallow-parsing methods. The co-occurrence method is used for mapping of the biological objects

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

in texts with dictionaries, and then the deep parsing method should be used for the identification of interactions between mapped objects.

Another important implementation of text-mining in micro biology and biomedicine field is automated building of thematic thesauruses of names of microbiology entities. Such dictionaries are crucial for the co-occurrence method in general, as well as for various text-mining systems that are based on the mapping of microbiological objects for the identification of interaction between them. This task can be partially solved by using data from semantic databases (database-mining) such as: Uniprot, Ensembl [34], PharmGKB [35], DrugBank [36] etc. Information contained within such databases has a high degree of confidence and is well-structured, but its rate of replenishment is inferior to the growth rate of the total number of publications. Another significant disadvantage for using dictionaries based exclusively on the information generated by database-mining for the identification of interactions between biological objects in literature is the lack of synonyms for biological objects. This problem is caused by authors often altering the canonical names of objects by adding back various special characters (dashes, colons, etc.), replacing the letters of the Greek alphabet to their transcriptions and vice versa, etc. in their publications. This is why the generation of a thesaurus with technologies combining text-mining and database-mining approaches achieves the best results. Tyne Liang and colleges used statistical approaches verified with real corpora in the thesaurus construction module of their bacterial Textual Processing and Retrieval System with thesauruses created by the analysis of databases and it showed good results [37].

2.4 3D-visualization

The scope of the 3D visualization introduced here is defined by two main areas: the pathway visualization in 3D and cell visualization and simulation.

An established approach lies in the 2.5D Visualization of metabolic networks [38], which offers comparison methods for two different biological networks: On the first 2D layer a metabolic pathway is presented, on layer two a protein interaction network and on the third layer, located in the middle of the 3D space, the overlapping nodes are shown. In other 2.5D visualization approaches the layer concept is used for the inter-organismic [39] or inter-domain large-scale [40] comparison of related metabolic networks. Another analogy with those 2.5D approaches is the use of KEGG [13] as the metabolic data source.

MetNetVR introduced the possibility of visualizing complex large-scale, hierarchical networks interactively by implementing different 3D layout algorithms [41]. Virtual Reality techniques are used to extend displays into the third dimension. In addition, the network layouts of MetNetVR may follow the cellular compartmentation, but only on a very abstract level, refusing cell component internal mapping.

BioCichlid is another tool which visualizes and animates time-dependent gene expression data, correlated with protein interaction, signalling and regulatory networks in 3D [42].

Different cell simulation environments have been extended from 2D to 3D during the last few years, but the included cell models of the mentioned approaches are based on a very high grade of simplification:

For example CompuCell3D is a software framework to simulate the development of multicellular organisms with stochastic rules and differential equations [43]. E-Cell3D is implementing meta-algorithms also based on differential equations to simulate nonlinear interactions between functional modules [44]. The Virtual Cell simulation environment (VCell) allows the formulation and simulation of cell biological models in 3D [45].

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

3 Applied Software Tools

3.1 VANESA

In the last decades, many different methods of modeling and simulation of biological networks have been introduced. In this paper a software application called VANESA (Visualization and Analysis of Networks in System Biology Applications) (http://vanesa.sf.net) is presented. VANESA creates a large-scale biological network based on the DAWIS-M.D. data warehouse information system to examine gene-controlled processes. The BioDWH data warehouse infrastructure was used to integrate life science data from multiple data sources for DAWIS-M.D. and VANESA. Using VANESA, different fields of studies are combined such as life-science, database consulting, modeling, visualization and simulation for a semi-automatic reconstruction of complex biological systems. The main function of VANESA is to trim down data to a manageable yet relevant size and to analyze and identify new as well as altered versions of interaction patterns in dynamic interaction networks.

The idea of VANESA is to extend any molecular data based network by new targets and interacting elements. The software solution is a new editor-controlled information system for the representation of research data in the form of biomedical network representations. Information is visualized in a clear and understandable manner to meet the purposes of underlying research activities. The user is enabled to record research results and thoughts in the form of a digital network model. The user is not limited to any kind of biological model; moreover it is possible to create an individual system that meets the requirements of each research activity.

As a case study of VANESA and the data warehouse BioDWH and DAWIS-M.D. information system, the modeling and exploration of biological systems in cardiovascular diseases from an EU project is presented here[46]. The case study is based on a cardiovascular-disease related to gene-regulated biological networks. Based on the project experimental data, literature and the integrated databases it was begun by exploring and reconstructing specific pathways derived from misleading proteins in cardiovascular diseases in VANESA.

In addition to experimental data, external databases and literature had to be examined for meaningful information to map out the important biomedical networks and systems. It is essential for scientists to access and analyze information from multiple heterogeneous data sources to meet their objectives.

The communication between VANESA and the biomedical data sources is realized by a web service. Spanning multiple databases containing biochemical and metabolic information from databases such as KEGG and HPRD enabled the modeling and visualization of the most important pathways based on the proteins in the discussed microarray sample of our case study. The data from the BioDWH system was analyzed on a large scale and visualized in a biological meaningful way. Multi-dimensional data annotations was considered in a way suitable for the knowledge discovery process.

As a result of the predicted gene-controlled processes and protein-protein interaction pathways scientists were provided with new opportunities for the discovery of novel biomarkers, and unknown therapeutic targets. In addition, the use of VANESA in combination with the data warehouse infrastructure BioDWH and the DAWIS-M.D. information system can allow investigation of the biological functionality of a gene or of a protein in its specific genetic or functional pathway.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

3.2 ANDVisio

The computer system ANDVisio-ANDCell was developed for automated extraction of knowledge from PubMed abstracts and databases concerning molecular-genetic interactions, gene regulations, catalytic processes, polymorphism gene – disease associations and other associations between facts and their representation as semantic association networks [4]. The vertices of such networks are molecular-genetic objects, diseases and processes while the edges between the vertices represent types of associations. Considered are the following objects: genes, proteins, microRNAs, metabolites, molecular processes, pathways and cellular components. The system has the following types of interaction between objects: direct interaction, catalytic reaction, proteolysis, treatment, co expression, expression regulation, activity/function regulation, stability regulation and transport regulation. For molecular interactions and associations, data on cell types and organisms are represented. Knowledge extracted from different types of publications was stored into the base of knowledge: ANDCell. A graphical user interface is realized in the ANDVisio program. ANDVisio allows the graphical visualization and analysis of the associative networks, reconstructed by using queries sent to the ANDCell knowledge base.

The knowledge base contains about 5 millions facts. For development of the base of knowledge ANDCell, data from the PubMed abstracts was analyzed, as well as different databases such as IntAct [47], MINT [48], NCBI GENE [49], TRRD [50], KEGG, PIMRider®, InterPro [51]. The system has been provided with a user-friendly interface and implemented links to molecular-genetic databases. Also, articles for additional information were extracted. The developed system may be useful for resolving a wide range of tasks in biology and biomedicine, such as expansion and complementation of the genetic networks reconstructed by the experts, identification of associations of genetic networks with diseases, search for the existing molecular mechanisms of associations between pathologies, identification of gene-candidates for genotyping, mutation that reduce disease, interpretation data of microchip analysis of gene expression etc.

ANDVisio system, particularly, was used for the analysis of potential molecular mechanisms of interconnection between myopia and glaucoma. In the course of this work, a list of potential gene-candidates for the genotyping of myopia and open-angle glaucoma [52] was detected. Also the ANDVisio system was used for the analysis of data from high-performance proteomic experiments in researching of Helicobacter pylori and their connection with progressive gastritis and gastric tumors [53].

3.3 CELLmicrocosmos PathwayIntegration

CELLmicrocosmos 4.2 PathwayIntegration (CmPI) is an approach to visualize and analyze inter-cellular and intra-compartmental relationships by correlating pathways with an abstract cell environment in 3D space. By using data coming from DAWIS-M.D., metabolic pathways from KEGG can be parsed. The pathway structure, consisting of enzymes, their substrates and products with the connecting reactions, can be shown in 2D as original KGML layout and directly compared to the 3D layout in the cell. The cell can be modelled by using a variety of different eucaryotic cell component models, which are mainly abstractions of Electron and Light Microscopic Images. In addition, first approaches of 3D microscopic-based cell component models exist, based on electron tomographic data. The composition of the cell may vary according to the needs of the visualization or mapping information.

For the enzymatic localization, terms from the databases BRENDA and UniProt (UniProt 2008) are used. Usually information exists on the subcellular level – but also mapping information about the intra-compartmental mapping may be derived . Focusing, for example on mitochondria, UniProt contains more than 50, BRENDA more than 20 different localization definitions. The quality of the

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

data varies: BRENDA contains only localization information reviewed by an curator, but UniProt provides additionally unreviewed information. These results may be compared directly to the PubMed abstract, if the corresponding database provides the link. Different terms may belong to the same localization: “Mitochondrial inter-membrane space” and “mitochondrial lumen” both need to be mapped onto 3rd mitochondrial layer. Often different mapping information are found and stored in an interactive localization table. The user may choose which of these options should be used to place the enzymatic spatial positioning or refuse the propositions from the databases and predict the localization. Sometimes the localization information from the database contains comments specifying more precisely the whereabouts of a protein then the regular cell component information. In this case, CmPI uses the comment for mapping.

The localization of different pathways is comparable by using the same position for each enzyme of the same type located in 3D space. An Inverted Self-Organizing Graphs (ISOM) layout is used for the distribution of nodes [54] onto unit hypersphere: Connected nodes are placed in proximity to each other. A six-degrees-of-freedom (6DoF) navigation offers different possibilities to navigate through the cell environment. Following the Focus+Context paradigm [55], also every single node of the pathway can be spatially focused and examined according the information acquired from the different databases. In addition, Stereoscopy [56] is implemented, compatible to e.g. nVidia® Quadro® FX cards, to take full advantage of the 3D perspective.

The Webstart application is located at http://Cm4.CELLmicrocosmos.org.

4 Application

4.1 Experimental Data

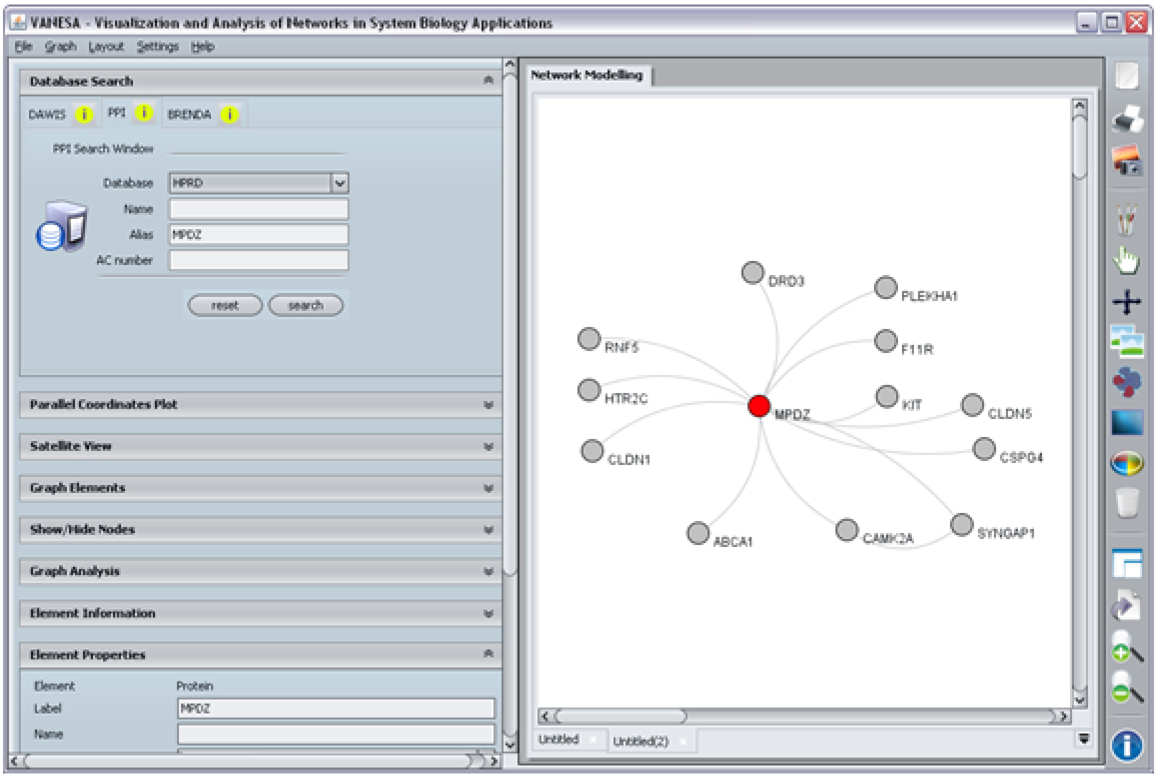
First, the point of interest has been defined as a Cardiovascular Disease related pathway. The data used here represents a dilated cardiomyopathy (DCM) coming from a female DCM patient with renal insufficiency aged 52 years. The analysis has been carried out in an extracted cytoplasmic sample of cultured aortic smooth muscle cells (S12 fraction). In order to highlight its associated disregulated pathways, the proteomic profile of the sample had to be investigated. The proteome analysis has been carried out by a Clontech Ab Microarray^{TM} 500 (Lot no. 7030444, Clontech, CA, USA). From the set of identified proteins, the Multiple PDZ Domain protein (MUPP1/MPDZ) has been chosen for former analysis. MUPP1 is a 13 PDZ domain holding protein, showing a large diversity of interacting proteins [57] and viruses [58, 59].

4.2 Network Reconstruction

Using VANESA, the environment of MPDZ has been investigated. 12 interacting proteins have been identified by using the BioDWH integration of HPRD: ABCA1, CAMK2A, CLDN1, CLDN5, CSPG4, DRD3, F11R, HTR2C, KIT, PLEKHA1, RNF5 and SYNGAP1. Moreover, CAMK2A and SYNGAP are also interacting with each other (Fig. 1).

Further investigations according the MPDZ protein were carried out using the KEGG implementation of VANESA. MPDZ was identified as being part of the human Tight junction signaling pathway (hsa04530) (Fig. 2). The tight junction is the closely associated area of the plasma membranes of two different cells. They form an impermeable surface area to ions and molecules. The corresponding pathway regulates the passage of substances through the protein complexes. The tight junction localization is in agreement with various publications [58, 59].

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).



VANESA - Visualization and Analysis of Networks in System Biology Applications

PPI Search Window

Database: HPRD

Name:

Alias: MPDZ

AC number:

Element Properties

Element: Protein

Label: MPDZ

Name:

Figure 1: The relevant sub-network (from Fig. 2) of direct protein interactions with MPDZ, computed by the BioDWH integration of the HPRD database in VANESA.

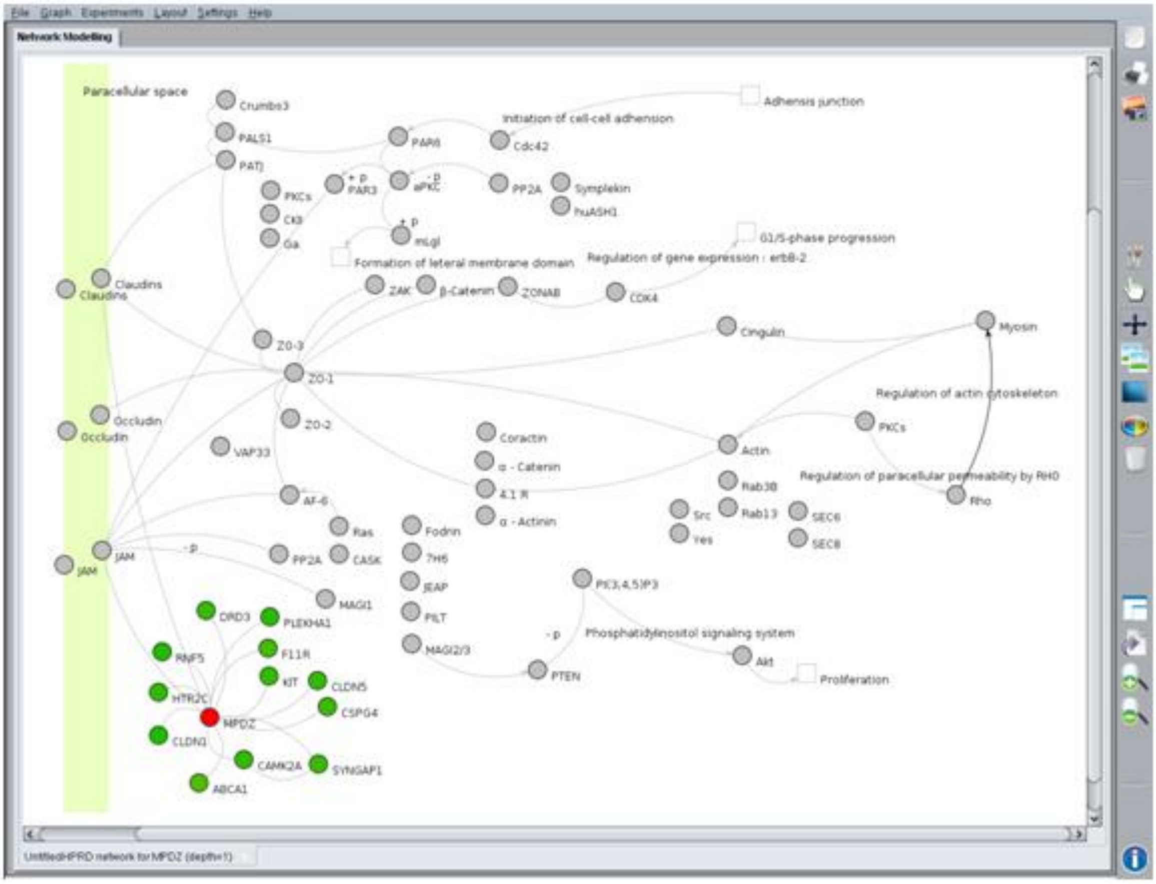


Figure 2: Visualization of the Tight junction signaling pathway (hsa04530) from KEGG by VANESA. The place marked red is the relevant protein MPDZ on the microarray sample.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

The ANDVisio system also has been used for the determination of the MUPP1 surrounding proteins by searching for its synonym MUPP1. Eight interacting proteins has been revealed and five of them are new: AMOT, CLD8, CLIC6, GABR2, PKHA2. After the curation of ANDVisio, all newly found associations were confirmed. Although the protein RNF5 was found by Vanesa, the curation of the ANDVisio text mining results could not verify the link between MUPP1 and RNF5. The synonym NG2 is pointing here to the protein CSPG4, although NG2 is often used as a synonym for RNF5. Sentences introducing these links are found in Table 1.

ANDVisio IDs	Uniprot KB IDs	Association Confirmation	PubMed ID	Sentence from the abstract which confirm the association
AMOT	Q4VCS5	confirmed	17397395	Using yeast two-hybrid screening, we found here that MUPP1 interacts with angiomotin (Amot), JEAP/Amot-like 1 and MASCOT/Amot-like 2, which we refer to as Amot/JEAP family proteins.
CLD5	O00501	confirmed	12403818	MUPP1 and claudin-5 colocalized in the incisures, and the COOH-terminal region of claudin-5 interacts with MUPP1 in a PSD-95/Disc Large/zona occludens (ZO)-1 (PDZ)-dependent manner.
CLD8	P56748	confirmed	12839333	The interaction of claudin-8 and MUPP1 in vivo was confirmed by co-immunolocalization and co-immunoprecipitation in MDCK cells.
CLIC6	Q96NY7	confirmed	14499480	In two-hybrid system, CLIC6 also interacted with MUPP1 and radixin but not GIPC, suggesting it could take part in a complex with D(2)-like receptors, not only by direct interaction with their C-termini, but also through interactions with scaffolding proteins.
CSPG4	Q6UVK1	confirmed	10967549	The fusion proteins fail to bind NG2 missing the C-terminal half of the cytoplasmic domain, emphasizing the role of the NG2 C-terminus in the interaction with MUPP1.
GABR2	O75899	confirmed	17145756	Biochemical analysis confirmed that full-length Mupp1 and PAPIN interact with GABA(B)R2 in cells.
PKHA2	Q9HB19	confirmed	11802782	We show that TAPP1 and TAPP2 interact with the 10th and 13th PDZ domain of MUPP1 through their C-terminal amino acids.
RNF5	Q99942	*not confirmed*	10967549	The fusion proteins fail to bind NG2 missing the C-terminal half of the cytoplasmic domain, emphasizing the role of the NG2 C-terminus in the interaction with MUPP1. *Explanation: The term NG2 is in the context of the referenced publication no synonym for RNF5 (see CSPG4 instead).*

Table 1. Sentences from PubMed proving associations to MUPP1/MPDZ extracted by ANDVisio.

4.3 Localization

The reconstructed network of MPDZ should be investigated according the localization of the different interacting proteins. Because the original sample discussed above (see 6.1) is taken from cytoplasm, the previous experimentally achieved knowledge was merely that the proteins are localized within the cell and outside the nucleus. The correlation to the tight junction pathway using KEGG is pointing towards the cell junction. Therefore it is mainly searched for this cellular part "tight junction" during the localization process of MPDZ and the 12 interacting proteins using the CmPI. Table 2 is showing the localization accuracy classes in this context. By using the BioDWH

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

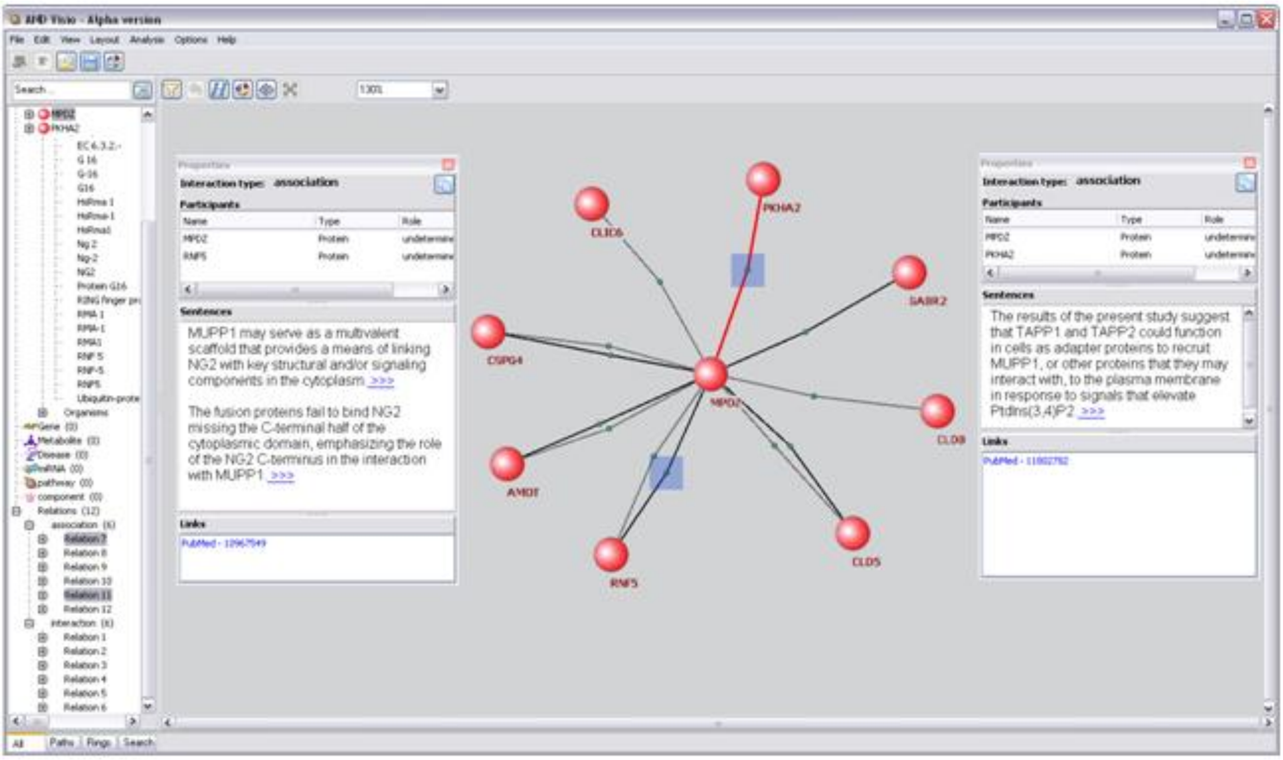
integrating BRENDA, UniProt, GO and Reactome, all of the 13 proteins are localized. The most precise results are achieved for five proteins, including MPDZ, by pointing to the tight/cell junction. For five proteins, the cell membrane, and for another two proteins, a membrane fraction has been identified as a possible localization (Fig. 4). For SYNGAP1 only the term "intracellular" inferred from electronic annotation by the InterPro database has been found, which is not accurate enough.

Localization Accuracy Class	Terms
High	cell/tight junction (organisation)
Middle	cell/plasma membrane/projection/surface integrin cell surface interaction 'transmembrane proteins with a single transmembrane pass, a cytoplasmic domain, and an extracellular domain'
Low	membrane/membrane fraction
No	actin filament chromosome collagen/collagen type VI cytoplasm, intracellular cytosol death inducing signaling complex endocytic vesicle endoplasmatic reticulum extracellular matrix filamentous actin golgi mitochondrium nucleus/nucleoplasm/pronucleus, ribosome sarcoplasmic reticulum vimentin X chromosome

Table 2: Localization Accuracy Classes of the terms found by CmPI and ANDVisio in comparison to the reference term "cell junction"

Now the question should be investigated, if the text mining data from PubMed abstracts created by ANDVisio can verify and/or improve the accuracy of the results. For this purpose, the proteins of the MPDZ interacting protein network are successfully identified. In a second step, the corresponding localizations of every protein are searched by focusing only on PubMed results. ANDVisio uses a dictionary which connects different synonyms and spellings to one localization term. "Peripheral plasma membrane protein", "juxta-membrane" and "juxtamembrane" are for example connected to the term "extrinsic to plasma membrane". 34 different cell components are found in ANDVisio (Fig. 6) and 10 of 13 proteins are localized by using the mapping table of CmPI.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).



Interaction type: association

Participants

Name	Type	Role
MPDZ	Protein	undetermine
RNF5	Protein	undetermine

Sentences

MUPP1 may serve as a multivalent scaffold that provides a means of linking NG2 with key structural and/or signaling components in the cytoplasm >>>

The fusion proteins fail to bind NG2 missing the C-terminal half of the cytoplasmic domain, emphasizing the role of the NG2 C-terminus in the interaction with MUPP1 >>>

Links

PubMed - 10967549

Interaction type: association

Participants

Name	Type	Role
MPDZ	Protein	undetermine
PKHA2	Protein	undetermine

Sentences

The results of the present study suggest that TAPP1 and TAPP2 could function in cells as adapter proteins to recruit MUPP1, or other proteins that they may interact with, to the plasma membrane in response to signals that elevate PtdIns(3,4)P2 >>>

Links

PubMed - 11802782

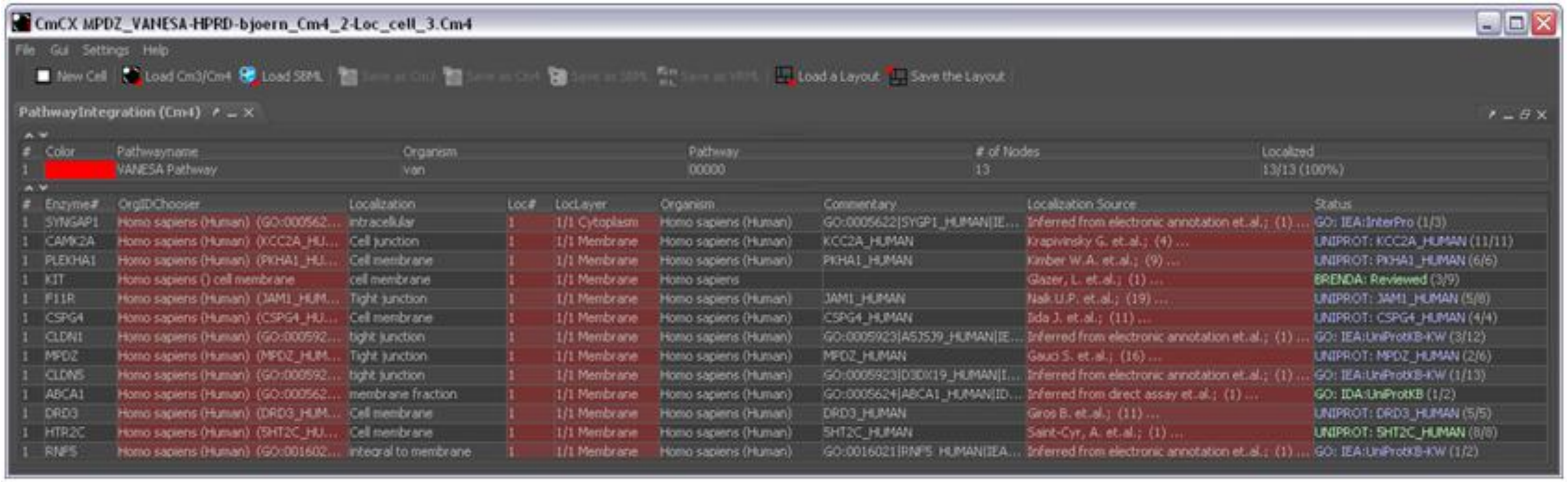
Figure 3. The MPDZ protein-protein interaction network based on PubMed abstracts in ANDVisio.

Importing the results to CmPI, four proteins could be localized to the tight/cell junction and four proteins to the cell membrane. For one protein the Nucleus and for another protein only a membrane faction could be found as results (Fig. 5).

The results from CmPI are combined with those from ANDVisio, showing that the results pointing to the cell junction are not improved. But four results proposing the cell junction are now double-proofed (Fig. 7). In addition, ANDVisio improved the result on ABCA1: CmPI could localize this protein only to a membrane (fraction), ANDVisio found results pointing to the cell membrane. The complete results can be found in Table 3.

After localizing the protein-protein interaction network created with VANESA, the network created with ANDVisio needed to be localized. CmPI can localize four of eight proteins to the cell junction and the remaining proteins to the cell membrane. ANDVisio can localize five proteins. Two of these proteins, which were not identified by VANESA, namely AMOT and GABR2, are found at the cell membrane according to PubMed abstracts. This is in affirmation of the CmPI results for these two proteins (Fig. 7).

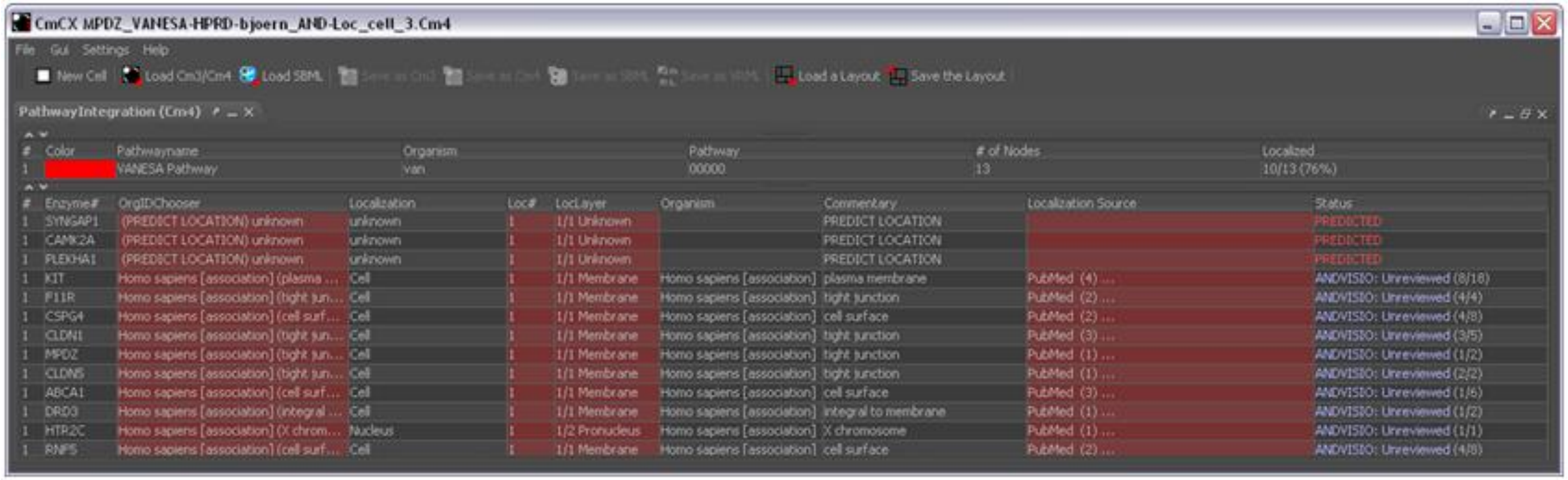
Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).



#	Color	Pathwayname	Organism	Pathway	# of Nodes	Localized
1	🟥	VANESA Pathway	van	00000	13	13/13 (100%)

#	Enzyme#	OrgIDChooser	Localization	Loc#	LocLayer	Organism	Commentary	Localization Source	Status
1	SYNGAP1	Homo sapiens (Human) (GO:000562...	intracellular	1	1/1 Cytoplasm	Homo sapiens (Human)	GO:0005622|SYGP1_HUMAN|IE...	Inferred from electronic annotation et.al.; (1) ...	GO: IEA:InterPro (1/3)
1	CAMK2A	Homo sapiens (Human) (KCC2A_HU...	Cell junction	1	1/1 Membrane	Homo sapiens (Human)	KCC2A_HUMAN	Krapivinsky G. et.al.; (4) ...	UNIPROT: KCC2A_HUMAN (11/11)
1	PLEKHA1	Homo sapiens (Human) (PKHA1_HU...	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	PKHA1_HUMAN	Kimber W.A. et.al.; (9) ...	UNIPROT: PKHA1_HUMAN (6/6)
1	KIT	Homo sapiens () cell membrane	cell membrane	1	1/1 Membrane	Homo sapiens		Glazer, L. et.al.; (1) ...	BRENDA: Reviewed (3/9)
1	F11R	Homo sapiens (Human) (JAM1_HUM...	Tight junction	1	1/1 Membrane	Homo sapiens (Human)	JAM1_HUMAN	Naik U.P. et.al.; (19) ...	UNIPROT: JAM1_HUMAN (5/8)
1	CSPG4	Homo sapiens (Human) (CSPG4_HU...	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	CSPG4_HUMAN	Iida J. et.al.; (11) ...	UNIPROT: CSPG4_HUMAN (4/4)
1	CLDN1	Homo sapiens (Human) (GO:000592...	tight junction	1	1/1 Membrane	Homo sapiens (Human)	GO:0005923|A5J5J9_HUMAN|IE...	Inferred from electronic annotation et.al.; (1) ...	GO: IEA:UniProtKB-KW (3/12)
1	MPDZ	Homo sapiens (Human) (MPDZ_HUM...	Tight junction	1	1/1 Membrane	Homo sapiens (Human)	MPDZ_HUMAN	Gauci S. et.al.; (16) ...	UNIPROT: MPDZ_HUMAN (2/6)
1	CLDN5	Homo sapiens (Human) (GO:000592...	tight junction	1	1/1 Membrane	Homo sapiens (Human)	GO:0005923|D3DX19_HUMAN|I...	Inferred from electronic annotation et.al.; (1) ...	GO: IEA:UniProtKB-KW (1/13)
1	ABCA1	Homo sapiens (Human) (GO:000562...	membrane fraction	1	1/1 Membrane	Homo sapiens (Human)	GO:0005624|ABCA1_HUMAN|ID...	Inferred from direct assay et.al.; (1) ...	GO: IDA:UniProtKB (1/2)
1	DRD3	Homo sapiens (Human) (DRD3_HUM...	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	DRD3_HUMAN	Gros B. et.al.; (11) ...	UNIPROT: DRD3_HUMAN (5/5)
1	HTR2C	Homo sapiens (Human) (5HT2C_HU...	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	5HT2C_HUMAN	Saint-Cyr, A. et.al.; (1) ...	UNIPROT: 5HT2C_HUMAN (8/8)
1	RNF5	Homo sapiens (Human) (GO:001602...	integral to membrane	1	1/1 Membrane	Homo sapiens (Human)	GO:0016021|RNF5_HUMAN|IEA...	Inferred from electronic annotation et.al.; (1) ...	GO: IEA:UniProtKB-KW (1/2)

Figure 4: The localization of the MPDZ interacting protein network (Fig. 1) using CmPI: 5 proteins could be localized to the tight/cell junction (including MPDZ), 5 proteins to the cell membrane, 2 protein to membrane and 1 protein to the cytoplasm (SYNGAP1).



#	Color	Pathwayname	Organism	Pathway	# of Nodes	Localized
1	🟥	VANESA Pathway	van	00000	13	10/13 (76%)

#	Enzyme#	OrgIDChooser	Localization	Loc#	LocLayer	Organism	Commentary	Localization Source	Status
1	SYNGAP1	(PREDICT LOCATION) unknown	unknown	1	1/1 Unknown		PREDICT LOCATION		PREDICTED
1	CAMK2A	(PREDICT LOCATION) unknown	unknown	1	1/1 Unknown		PREDICT LOCATION		PREDICTED
1	PLEKHA1	(PREDICT LOCATION) unknown	unknown	1	1/1 Unknown		PREDICT LOCATION		PREDICTED
1	KIT	Homo sapiens [association] (plasma ...	Cell	1	1/1 Membrane	Homo sapiens [association]	plasma membrane	PubMed (4) ...	ANDVISIO: Unreviewed (8/18)
1	F11R	Homo sapiens [association] (tight jun...	Cell	1	1/1 Membrane	Homo sapiens [association]	tight junction	PubMed (2) ...	ANDVISIO: Unreviewed (4/4)
1	CSPG4	Homo sapiens [association] (cell surf...	Cell	1	1/1 Membrane	Homo sapiens [association]	cell surface	PubMed (2) ...	ANDVISIO: Unreviewed (4/8)
1	CLDN1	Homo sapiens [association] (tight jun...	Cell	1	1/1 Membrane	Homo sapiens [association]	tight junction	PubMed (3) ...	ANDVISIO: Unreviewed (3/5)
1	MPDZ	Homo sapiens [association] (tight jun...	Cell	1	1/1 Membrane	Homo sapiens [association]	tight junction	PubMed (1) ...	ANDVISIO: Unreviewed (1/2)
1	CLDN5	Homo sapiens [association] (tight jun...	Cell	1	1/1 Membrane	Homo sapiens [association]	tight junction	PubMed (1) ...	ANDVISIO: Unreviewed (2/2)
1	ABCA1	Homo sapiens [association] (cell surf...	Cell	1	1/1 Membrane	Homo sapiens [association]	cell surface	PubMed (3) ...	ANDVISIO: Unreviewed (1/6)
1	DRD3	Homo sapiens [association] (integral ...	Cell	1	1/1 Membrane	Homo sapiens [association]	integral to membrane	PubMed (1) ...	ANDVISIO: Unreviewed (1/2)
1	HTR2C	Homo sapiens [association] (X chrom...	Nucleus	1	1/2 Pronucleus	Homo sapiens [association]	X chromosome	PubMed (1) ...	ANDVISIO: Unreviewed (1/1)
1	RNF5	Homo sapiens [association] (cell surf...	Cell	1	1/1 Membrane	Homo sapiens [association]	cell surface	PubMed (2) ...	ANDVISIO: Unreviewed (4/8)

Figure 5: The localization of the MPDZ interacting protein network (Fig. 1) using CmPI and exclusively results from ANDVisio (Fig. 6): 4 proteins could be localized to the tight/cell junction, 4 proteins to the cell membrane, 2 proteins to the Nucleus and 3 proteins not at all.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

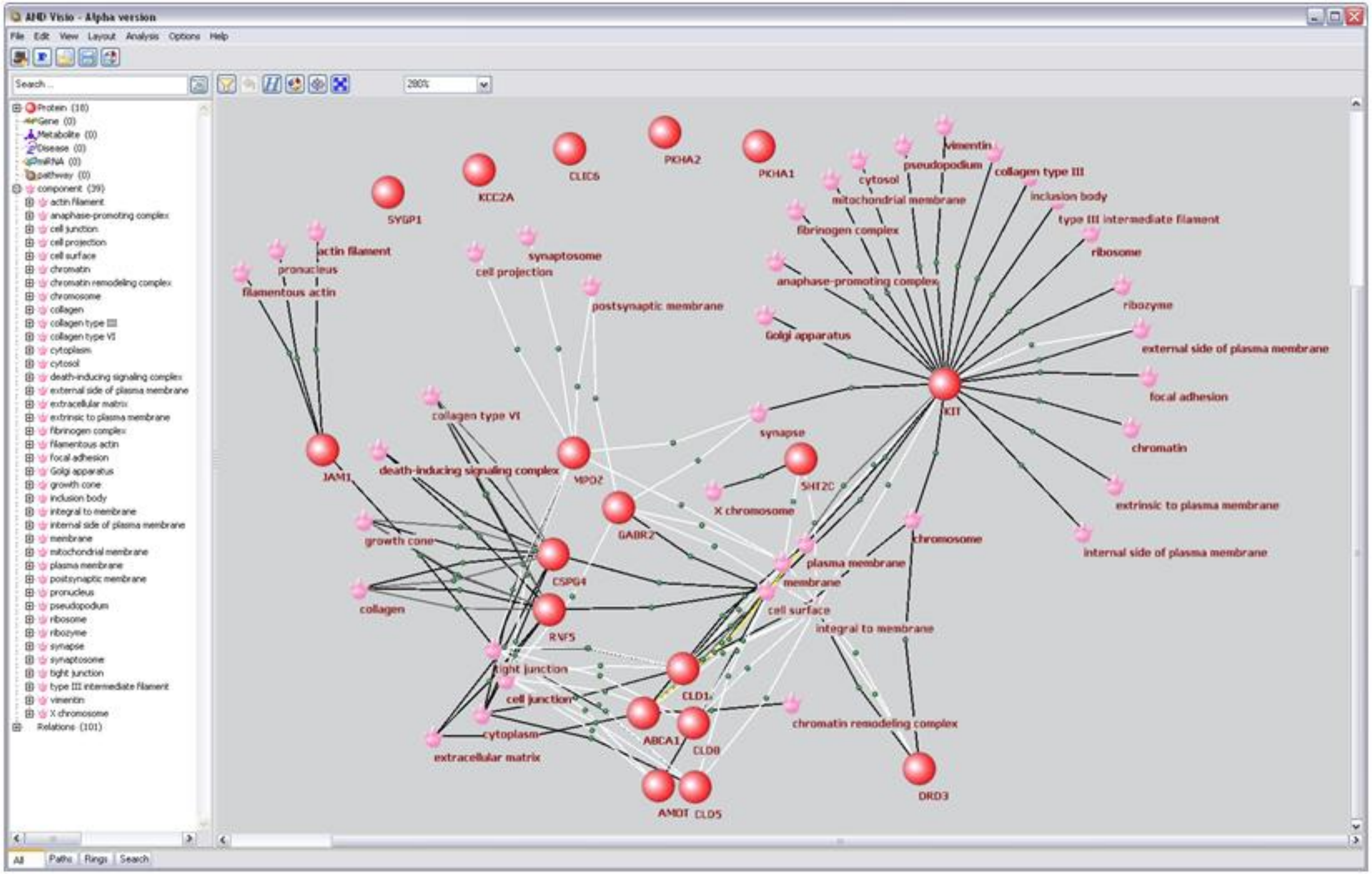
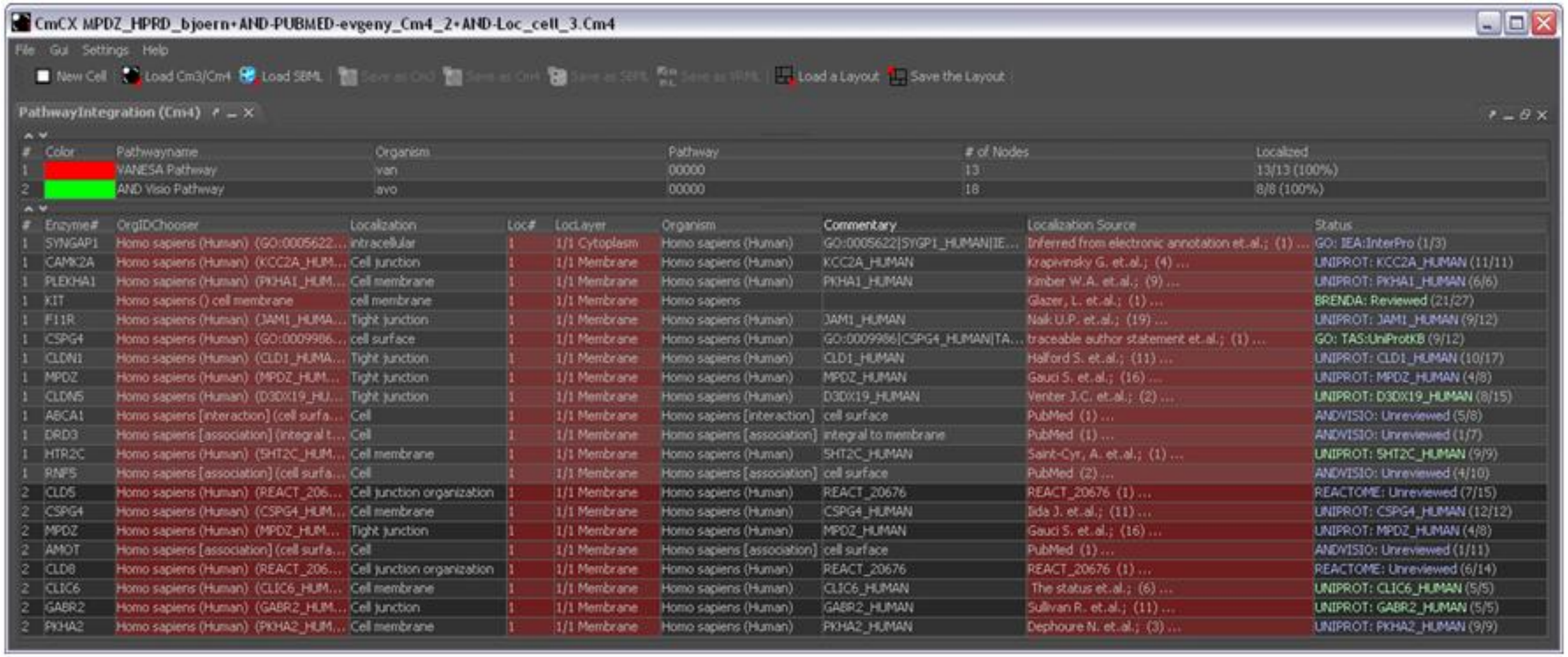


Figure 6: The sub-network shown in Fig. 2 supplemented by using ANDVisio and its Localization results extracted from PubMed entries. The protein identifiers are here synonyms for the identifiers used in VANESA. Six proteins are not localized: PKHA1 (PLEKHA1), PKHA2 (PLEKHA2), KCC2A (CAMK2A), CLIC6, CLD8 and SYGP1 (SYNGAP1). Localization descriptions obtained by ANDVisio have different detail levels. For example the KIT protein is connected to the following descriptions related to the cell membrane: cell surface, plasma membrane, pseudopodium, external/internal side of plasma membrane and extrinsic to plasma membrane. Grey lines show "interactions", black lines show "association" and the yellow line shows an "activity regulation".

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).



#	Color	Pathwayname	Organism	Pathway	# of Nodes	Localized
1	🟥	VANESA Pathway	van	00000	13	13/13 (100%)
2	🟩	AND Visio Pathway	avo	00000	18	8/8 (100%)

#	Enzyme#	OrgIDChooser	Localization	Loc#	LocLayer	Organism	Commentary	Localization Source	Status
1	SYNGAP1	Homo sapiens (Human) (GO:0005622…	intracellular	1	1/1 Cytoplasm	Homo sapiens (Human)	GO:0005622|SYGP1_HUMAN|IE…	Inferred from electronic annotation et.al.; (1) …	GO: IEA:InterPro (1/3)
1	CAMK2A	Homo sapiens (Human) (KCC2A_HUM…	Cell junction	1	1/1 Membrane	Homo sapiens (Human)	KCC2A_HUMAN	Krapivinsky G. et.al.; (4) …	UNIPROT: KCC2A_HUMAN (11/11)
1	PLEKHA1	Homo sapiens (Human) (PKHA1_HUM…	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	PKHA1_HUMAN	Kimber W.A. et.al.; (9) …	UNIPROT: PKHA1_HUMAN (6/6)
1	KIT	Homo sapiens () cell membrane	cell membrane	1	1/1 Membrane	Homo sapiens		Glazer, L. et.al.; (1) …	BRENDA: Reviewed (21/27)
1	F11R	Homo sapiens (Human) (JAM1_HUMA…	Tight junction	1	1/1 Membrane	Homo sapiens (Human)	JAM1_HUMAN	Naik U.P. et.al.; (19) …	UNIPROT: JAM1_HUMAN (9/12)
1	CSPG4	Homo sapiens (Human) (GO:0009986…	cell surface	1	1/1 Membrane	Homo sapiens (Human)	GO:0009986|CSPG4_HUMAN|TA…	traceable author statement et.al.; (1) …	GO: TAS:UniProtKB (9/12)
1	CLDN1	Homo sapiens (Human) (CLD1_HUMA…	Tight junction	1	1/1 Membrane	Homo sapiens (Human)	CLD1_HUMAN	Halford S. et.al.; (11) …	UNIPROT: CLD1_HUMAN (10/17)
1	MPDZ	Homo sapiens (Human) (MPDZ_HUM…	Tight junction	1	1/1 Membrane	Homo sapiens (Human)	MPDZ_HUMAN	Gauci S. et.al.; (16) …	UNIPROT: MPDZ_HUMAN (4/8)
1	CLDN5	Homo sapiens (Human) (D3DX19_HU…	Tight junction	1	1/1 Membrane	Homo sapiens (Human)	D3DX19_HUMAN	Venter J.C. et.al.; (2) …	UNIPROT: D3DX19_HUMAN (8/15)
1	ABCA1	Homo sapiens [interaction] (cell surfa…	Cell	1	1/1 Membrane	Homo sapiens [interaction]	cell surface	PubMed (1) …	ANDVISIO: Unreviewed (5/8)
1	DRD3	Homo sapiens [association] (integral t…	Cell	1	1/1 Membrane	Homo sapiens [association]	integral to membrane	PubMed (1) …	ANDVISIO: Unreviewed (1/7)
1	HTR2C	Homo sapiens (Human) (5HT2C_HUM…	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	5HT2C_HUMAN	Saint-Cyr, A. et.al.; (1) …	UNIPROT: 5HT2C_HUMAN (9/9)
1	RNF5	Homo sapiens [association] (cell surfa…	Cell	1	1/1 Membrane	Homo sapiens [association]	cell surface	PubMed (2) …	ANDVISIO: Unreviewed (4/10)
2	CLD5	Homo sapiens (Human) (REACT_206…	Cell junction organization	1	1/1 Membrane	Homo sapiens (Human)	REACT_20676	REACT_20676 (1) …	REACTOME: Unreviewed (7/15)
2	CSPG4	Homo sapiens (Human) (CSPG4_HUM…	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	CSPG4_HUMAN	Iida J. et.al.; (11) …	UNIPROT: CSPG4_HUMAN (12/12)
2	MPDZ	Homo sapiens (Human) (MPDZ_HUM…	Tight junction	1	1/1 Membrane	Homo sapiens (Human)	MPDZ_HUMAN	Gauci S. et.al.; (16) …	UNIPROT: MPDZ_HUMAN (4/8)
2	AMOT	Homo sapiens [association] (cell surfa…	Cell	1	1/1 Membrane	Homo sapiens [association]	cell surface	PubMed (1) …	ANDVISIO: Unreviewed (1/11)
2	CLD8	Homo sapiens (Human) (REACT_206…	Cell junction organization	1	1/1 Membrane	Homo sapiens (Human)	REACT_20676	REACT_20676 (1) …	REACTOME: Unreviewed (6/14)
2	CLIC6	Homo sapiens (Human) (CLIC6_HUM…	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	CLIC6_HUMAN	The status et.al.; (6) …	UNIPROT: CLIC6_HUMAN (5/5)
2	GABR2	Homo sapiens (Human) (GABR2_HUM…	Cell junction	1	1/1 Membrane	Homo sapiens (Human)	GABR2_HUMAN	Sullivan R. et.al.; (11) …	UNIPROT: GABR2_HUMAN (5/5)
2	PKHA2	Homo sapiens (Human) (PKHA2_HUM…	Cell membrane	1	1/1 Membrane	Homo sapiens (Human)	PKHA2_HUMAN	Dephoure N. et.al.; (3) …	UNIPROT: PKHA2_HUMAN (9/9)

Figure 7: The combined localization results for the VANESA (#1) and the ANDVisio (#2) pathway. The VANESA Pathway includes the localization of the MPDZ interacting protein network (Fig. 1) using CmPI including results from ANDVisio: Five proteins could be localized to the tight/cell junction (including MPDZ), seven proteins to the cell membrane, one protein only to the cytoplasm (SYGP1).

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Protein	CmPI Localization				ANDVisio Localization
	BRENDA	GO*	Reactome	UniProt	PubMed Abstracts
	VANESA Pathway Reconstruction				
ABCA1		m(U)		m	cm:4, em, nu
CAMK2A	c, cs, m, nu, sa	cj(U), cs(R), nu(R)	cs, nu	cj, cm, m	
CLDN1		cj(U):2	cj	cj:2, cm:2, m:2	cj, cm:2, c, nu
CLDN5		cj(U):2	cj	cj:2, cm:2, m:2	cj, cs
CSPG4		cm(U)		m, cm	cm, cs:2, em:3
DRD3				cm, m	m, nu
F11R		cj(U)	cj, cm	cj, cm, m	cj, cs:2, nu
HTR2C		cm(U)		cm, m	nu
KIT	cm:3	m(I), em(U)**		m:2	cs:3, cm:4, em:2, gg, mi, nu:3, ri
MPDZ		cj(U)		cj, cm, m	cj, c
PLEKHA1		c(U), nu(H)		cm, c, m, nu	
RNF5				m	cm, c, cs, em:4
SYNGAP		c(I):3			
	ANDVisio Pathway Reconstruction* **				
AMOT		cj(U), cm(M,U):2, cp(U):2, vs(M)		cj	cm
CLD8		cj(U):2, er(U)	cj	cj:2, cm:2, m:2	
CLIC6		c(U), cm(U)		c, cm, m	
GABR2		cj(U)		cj, cm, m	cm
PKHA2		c(U):2, cm(U), nu(U):2		c, cm, m, nu	

Table 3: Localization Results in CmPI and ANDVisio
Unique Connections to the following Cell Components:
c: cytoplasm, intracellular; cj: cell/tight junction (organisation); cm: cell/plasma membrane/projection/surface, integrin cell surface interaction, "transmembrane proteins with a single transmembrane pass, a cytoplasmic domain, and an extracellular domain"; cs: cytosol, death inducing signaling complex, filamentous actin, actin filament, vimentin; em: extracellular matrix, collagen (type VI); er: endoplasmatic reticulum; gg: golgi; m: membrane (fraction); mi: mitochondrium; nu: nucleus/nucleoplasm/pronucleus, (X) chromosome; ri: ribosome; sr: sarcoplasmic reticulum; vs: endocytic vesicle; numbers behind the colon show multiple results for one localization
Comments:
*** the letter in brackets indicates if the GO result is achieved data from InterPro (I), Human Protein Atlas (HPA), MGI (M), Reactome (R) or UniProt (U)**
**** in UniProt, this term is only found in the GO keywords, not as a cell component definition, therefore it is not found by the direct UniProt search in CmPI in this case**
***** italic-faced proteins are also part of the ANDVisio network**

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

5 Discussion

Analysing the Localization Results in Table 3, different important observations are made:

CmPI already achieves good results by combining four different databases:

- BRENDA: With two localized enzymes BRENDA does not perform well. The reason is that BRENDA is working with EC numbers. For many proteins discussed here there is no EC entry. Therefore BRENDA is unable to find localizations for those enzymes.
- GO: The Gene Ontology provides links to different external databases. The links to UniProtKB have been specially relevant for this sample. But there have been also links to Reactome, InterPro, Human Protein Atlas (HPA) [60] and a human annotation from the Mouse genome informatics (MGI) [61]. 16 proteins have been localized by GO, indicating that it is a very good localization resource. In particular the protein SYNGAP1 could only be localized by using GO.
- Reactome: For a few proteins, five in number, results are coming from Reactome. Most of them are pointing to the cell junction, which is the most relevant localization.
- UniProt: All proteins except SYNGAP1 has been localized by UniProt. This indicates, that it is the best resource discussed here for localizing proteins by their gene names.

Focusing now on the results from ANDVisio, the advantages of the text mining approach emerges. It should be mentioned, that ANDVisio is also supporting other data sources, but the results discussed here were restricted to PubMed Abstract data. In spite of this, the approach performs well for our purpose. Four proteins could be localized to the tight/cell junction and six proteins to the cell membrane. One hit is pointing to a nucleus or a membrane, and one hit only to the nucleus. Another six proteins could not be localized at all. If it would make sense to rank the localization sources according the hits for this sample, it would be UniProt, GO, ANDVisio PubMed Abstracts, Reactome, BRENDA.

Naturally, it is not the intention of the text mining approach discussed here to compete with an established database. Particularly the problem with the ambiguous synonym NG2 pointing to CSPG4 as well as RNF5 show that it is important to evaluate crucial data from ANDVisio. But the final question reads as follows: Is ANDVisio able to increase the precision of the localization? It was mentioned before that three localizations from CmPI were very imprecise. The results from ANDVisio are fixing two of these problems by adding plasma membrane similar terms. Comparing all localizations in Table 3 in detail it shows, that ANDVisio finds new localizations for ten proteins, which are not included in the databases accessed by CmPI. This fact shows the high importance of tools like ANDVisio: They can be used to verify, improve and extend the localization results. Tools like this should be used by database curators to search for new results expanding the knowledge of their databases.

Finally one question remains: How to solve the problem with the imprecisely localized protein SYNGAP1? Analyzing the networks, the interacting proteins should be taken into account. MPDZ and CAMK2A are interacting with SYNGAP1, as the protein interaction network indicates. In addition, ANDVisio can be used to search the PubMed abstracts for interacting proteins with SYGP1. One of three interaction nodes is the protein KCC2G, which belongs to the same enzyme complex (EC 2.7.11.17) as KCC2A, the synonym for CAMK2A. Therefore, this connection is double-proofed. The localizations shown by ANDVisio, “cytoplasm”, “chromosome” and “growth cone” are not satisfactory in this case, because the search is focusing on the tight junction complex (Fig. 8). In CmPI, MPDZ as well as CAMK2A are localized at the tight junction. The logical consequence is, that SYNGAP1 can also be found with a very high probability at the tight junction complex.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

With the accumulated knowledge, a virtual cell environment is created based on the localization information from CmPI. With this visualization, it is possible to compare the 2D network with the localized 3D visualization, the localization table (Tab. 3) and in addition, both networks created with VANESA and ANDVisio (Fig. 9-12). Of course the CmPI Visualization is more useful for showing inter- and intra-organelle relationships instead of processes restricted only to one region like the tight junction discussed here. The visualizations of the alternative localizations in Fig. 10 and 12 give an idea of this ability.

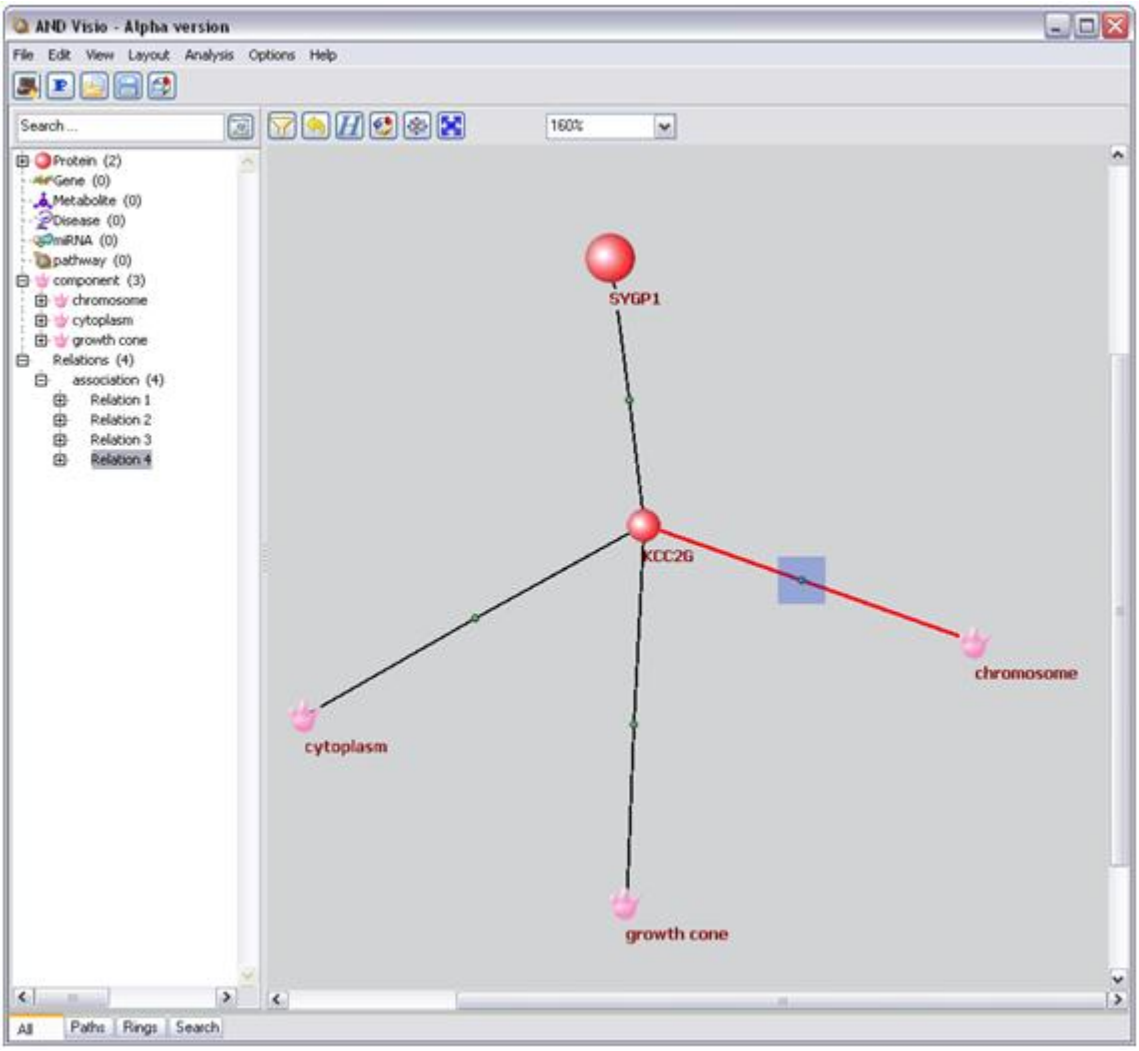


Figure 8: The protein interaction network in ANDVisio shows also the direct connection between SYGP (SYNGAP1) and KCC2G, which is part of the same protein complex (2.7.11.17) as KCC2A (CAMK2A). The localization result of KCC2G could be assigned as well to SYGP1: chromosome, cytoplasm and growth cone.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

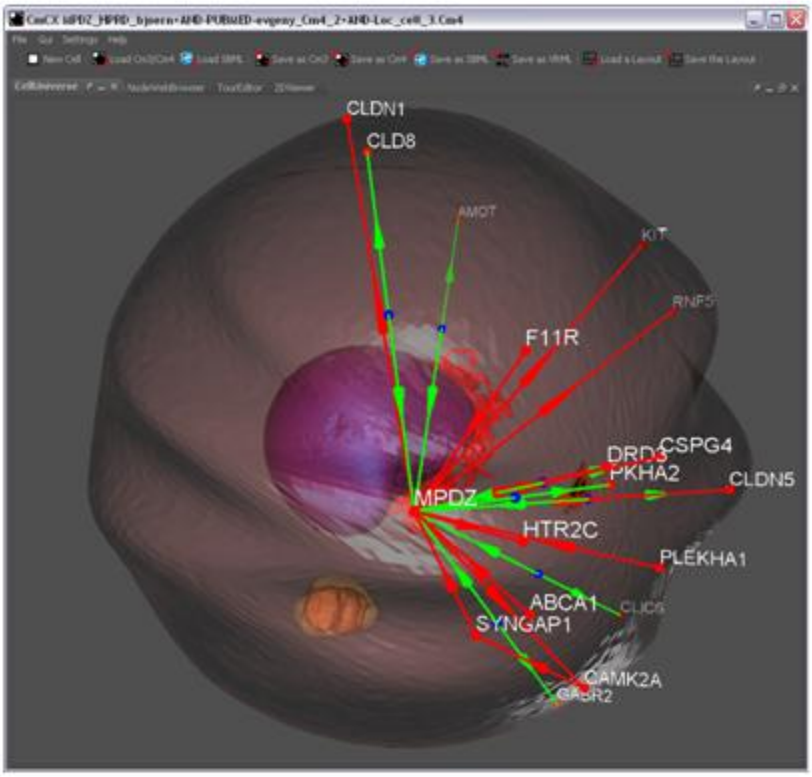


Figure 9: This CmPI visualization shows the localized network as shown in Fig. 7. The red network comes from VANESA, the green from ANDVisio. All enzymes are localized at the cell membrane. SYNGAP1, originally localized at the cytosol, can also be mapped directly onto the cell membrane by combining the newly gained knowledge.

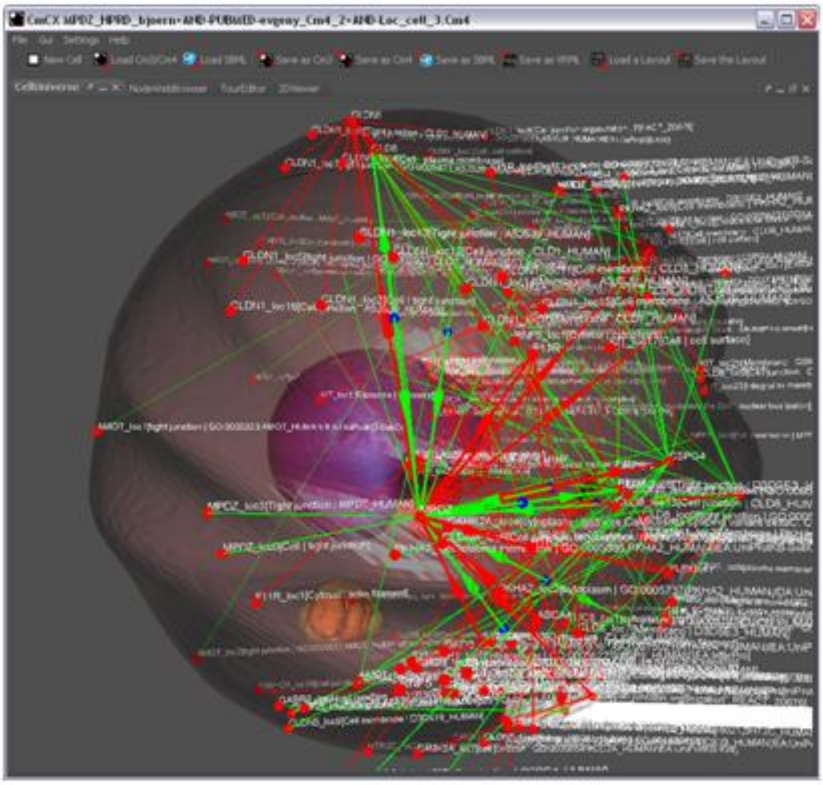
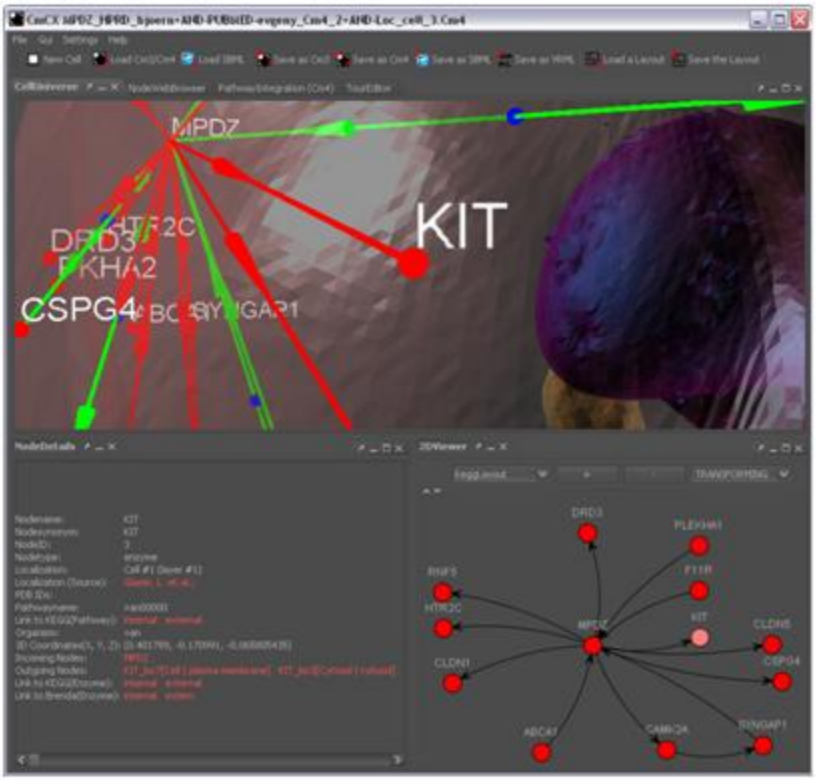


Figure 10: The CmPI visualization from Fig. 7 showing additionally all alternative localization of the enzymes. The information overflow can be limited by using the Focus+Context paradigm.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).



Nodename: KIT
Nodesynonym: KIT
NodeID: 3
Nodetype: enzyme
Localization: Cell #1 (layer #1)
Localization (Source): [illegible]
PDB IDs:
Pathwayname: [illegible]
Link to KEGG(Pathway): internal external
Organism: hsa
3D Coordinates(X, Y, Z): [illegible]
Incoming Nodes: MPDZ
Outgoing Nodes: [illegible]
Link to KEGG(Enzyme): internal external
Link to Brenda(Enzyme): internal external

Figure 11: The comparison of the 2D and 3D visualization. In this case CmPI uses the original 2D layout from VANESA and maps it into the 3D cell environment. The NodeDetails window provides information about the actual state of the node. Every window shown here can be used for the navigation.

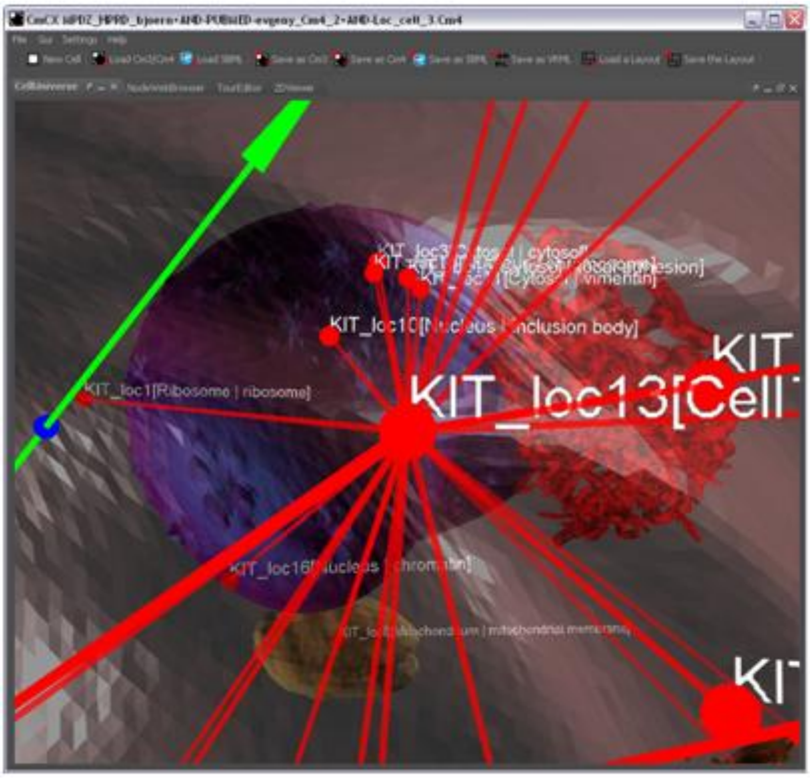

CellUniverse
KIT_loc3[Cytosol | cytosol]
KIT_loc1[Cytosol | vimentin]
KIT_loc10[Nucleus | inclusion body]
KIT_loc1[Ribosome | ribosome]
KIT_loc13[Cell
KIT
KIT_loc16[Nucleus | chromatin]
KIT_loc[Mitochondrium | mitochondrial membrane]
KI

Figure 12: Focusing KIT, the incoming reaction from MPDZ (thick red line) and connections to the alternative locations (thin red lines) are shown, like the Ribosome, Nucleus and Mitochondria.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

6 Conclusion

This case study showed a way to combine experimental data, data warehouse and text mining approaches in order to create protein-protein interaction networks by using VANESA and ANDVisio. All in all, 17 proteins were identified to interact with MPDZ. Then the logical assumption was examined, that the interacting proteins with MPDZ, which is a part of the tight junction pathway, can be localized at the tight junction or at least at the cell membrane by combining results from CELLmicrocosmos 4.2 PathwayIntegration and ANDVisio PubMed abstract text mining. With this methods, eight proteins, including MPDZ, could be localized at the tight junction complex, nine proteins at the cell membrane and one protein, SYNGAP1, could be imprecisely localized at the inner cell. By combining the localization results with the pathway structure it was shown, that SYNGAP1 could also be indirectly localized at the cell junction. Moreover, it was verified, that ANDVisio is an important tool which can be used to search for localization alternatives to extend the content of curated databases by identifying information gaps. With all acquired knowledge it was finally possible to create a curated cell visualization showing the intracellular relationships of the network discussed here (Fig. 13).

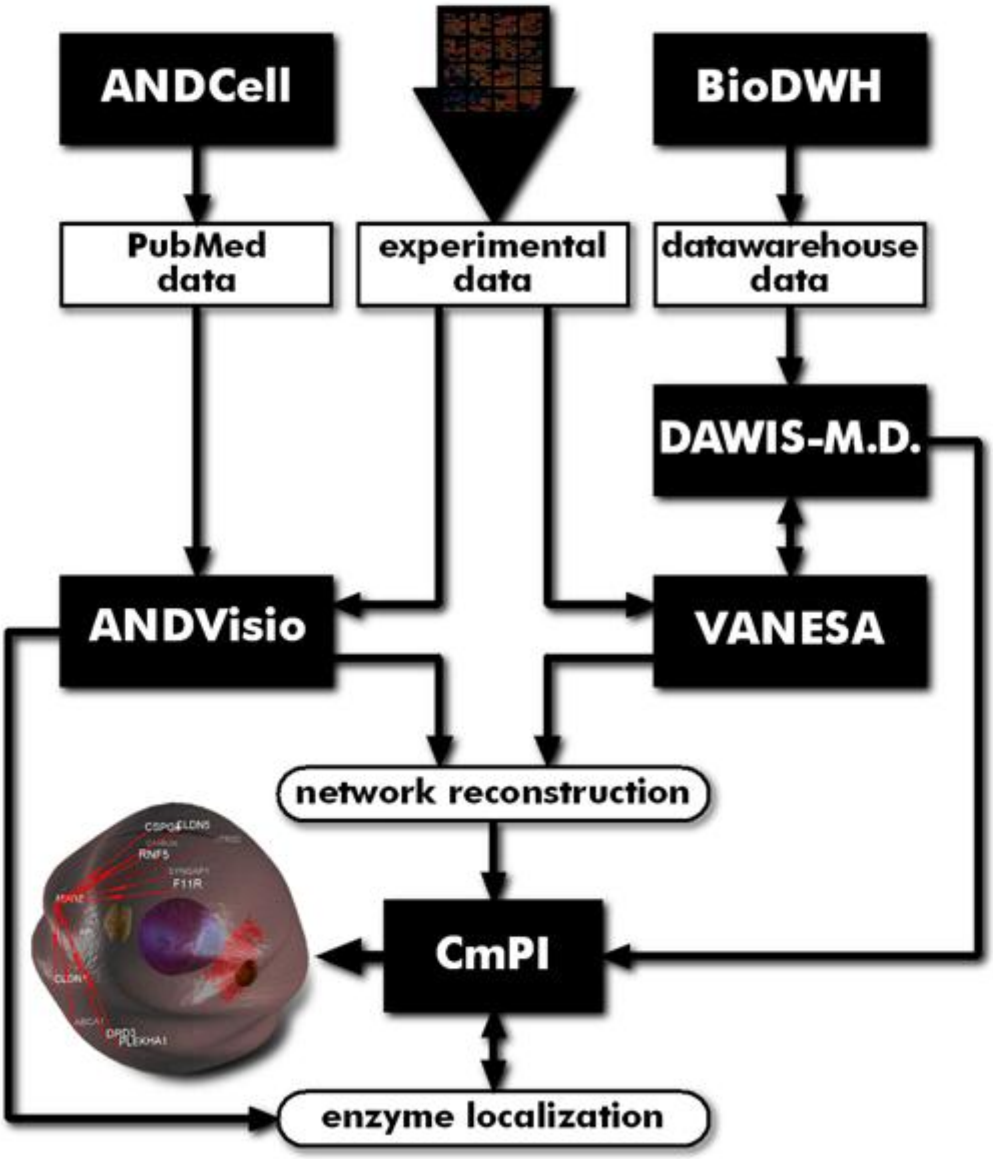


Figure 13: Knowledge generation by the combination of data from seven different projects.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Supplementary Data

The Supplementary Data is located at: http://jib2010.Cm4.CELLmicrocosmos.org

Acknowledgements

This work was supported in part by:

7th Framework Programme (FP7) project "New Algorithms for Host Pathogen Systems Biology",

BMBF Internationale Zusammenarbeit in Bildung und Forschung mit Russland, Projekt RUS 08/005 (BS, RH),

DAAD Leonard Euler Program grant Nr. 50024820 (AB, ET, TI),

EU-FP6, CARDIOWORKBENCH project, to FA,

Graduate College for Bioinformatics (GK635) (which is sponsored by the Deutsche Forschungs Gesellschaft (DFG)) (BS),

Interdisciplinary integrative Project of SB RAS N111 "Intelligent computer analysis of scientific texts for search, retrieval and integration of knowledge: the application to catalysis in chemistry and biology",

Russian Ministry of Science & Education (02.740.11.0705; P721) (AK),

State contract with MES RF 02.514.11.4123,

State contract with MES RF 02.740.11.0882.

References

[1] K. Wiemann. Das MSD Manual der Diagnostik und Therapie, 7. Auflage. Elsevier GmbH, München, 2007. ISBN 978-3-437-21761-6. Online Journal: http://www.msd.de/msdmanual/

[2] A. Hamosh, A. F. Scott, J. S. Amberger, C. A. Bocchini and V. A. McKusick. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res., 33:D514–D517, 2005.

[3] S. Janowski, B. Kormeier, T. Töpel, K. Hippe, R. Hofestädt, N. Willassen, R. Friesen, S. Rubert, D. Borck, P. Haugen and M. Chen. Modeling of cell-to-cell communication processes with Petri nets using the example of quorum sensing. Silico Biologdemenkovy, 10(0003), 2010. Online Journal: http://www.bioinfo.de/isb/2010/10/0003/

[4] P. S. Demenkov, E. Aman and V. A. Ivanisenko. Associative network discovery (AND) - the computer system for automated reconstruction networks of associative knowledge about molecular-genetic interactions. Computational technologies, 13(2):15-9, 2008.

[5] B. Sommer, J. Künsemöller, N. Sand, A. Husemann, M. Rumming and B. Kormeier. CELLmicrocosmos 4.1: An interactive approach to Integrating Spatially Localized Metabolic Networks into a Virtual 3D Cell Environment. Proceedings of Bioinformatics 2010, 90-5, 2010.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

[6] G. R. Cochrane and M. Y. Galperin. The 2010 Nucleic Acids Research Database Issue and online Database Collection: a community of data resources. Nucleic Acids Res., 38:D1–D4, 2010.

[7] T. Töpel, B. Kormeier, A. Klassen and R. Hofestädt. BioDWH: A Data Warehouse Kit for Life Science Data Integration. Journal of Integrative Bioinformatics, 5(2):93, 2008.

[8] A. Chang, M. Scheer, A. Grote, I. Schomburg and D. Schomburg. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res., 37:D588-D592, 2009.

[9] T. Kulikova, P. Aldebert, N. Althorpe, W. Baker, K. Bates, P. Browne, A. van den Broek, G. Cochrane, K. Duggan, R. Eberhardt, N. Faruque, M. Garcia-Pastor, N. Harte, C. Kanz, R. Leinonen, Q. Lin, V. Lombard, R. Lopez, R. Mancuso, M. McHale, F. Nardone, V. Silventoinen, P. Stoehr, G. Stoesser, M. N. Tuli, K. Tzouvara, R. Vaughan, D. Wu, W. Zhu and R. Apweiler. EMBL Nucleotide Sequence Database in 2006. Nucleic Acids Res., 35:D1–D5, 2006.

[10] A. Bairoch: The ENZYME database in 2000. Nucleic Acids Res., 28:304–5, 2000.

[11] Gene Ontology Consortium: The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res., 32:D258–61, 2004.

[12] T. S. K. Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. H. Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, A. Pandey: Human Protein Reference Database - 2009 update. Nucleic Acids Res., 37:D767–D772, 2009.

[13] M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu and Y. Yamanishi: KEGG for linking genomes to life and the environment. Nucleic Acids Res., 36:D480-D484, 2008.

[14] L. Matthews, G. Gopinath, M. Gillespie, M. Caudy, D. Croft, B. de Bono, P. Garapati, J. Hemish, H. Hermjakob, B. Jassal, A. Kanapin, S. Lewis, S. Mahajan, B. May, E. Schmidt E., I. Vastrik, G. Wu, E. Birney, L. Stein and P. D'Eustachio P.: Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res., 37:D619-22, 2009.

[15] A. G. Murzin, S. E. Brenner, T. Hubbard and C. Chothia. SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. J. Mol. Biol., 247:536–540, 1995.

[16] E. Wingender, X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. P. I. Reuter and F. Schacherer. TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res., 28:316–9, 2000.

[17] M. Krull, S. Pistor, N. Voss, A. Kel, I. Reuter, D. Kronenberg, H. Michael, K. Schwarzer, A. Potapov, C. Choi, O. Kel-Margoulis and E. Wingender: TRANSPATH: an information resource for storing and visualizing signaling pathways and their pathological aberrations. Nucleic Acids Res., 34:D546–D551, 2006.

[18] R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi and L. S. Yeh. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res., 32:D115–D119, 2004.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

[19] B. Kormeier, K. Hippe, T. Töpel and R. Hofestädt. CardioVINEdb: a data warehouse approach for integration of life science data in cardiovascular diseases. Journal of Integrative Bioinformatics, 7(1):142, 2010.

[20] M. E. Falagas, E. I. Pitsuoni, G. A. Malietzis, G. Pappas. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. The FASEB Journal, 22(2):338-42, 2008.

[21] H. Jose, T. Vadivukarasi, J. Devakumar. Extraction of protein interaction data: a comparative analysis of methods in use. EURASIP J. Bioinform. Syst. Biol., 53096, 2007.

[22] T. K. Jenssen, A. Laegreid, J. Komorowski and E. Hovig. A literature network of human genes for high-throughput analysis of gene expression. Nat. Genet., 28(1):21-8, 2001.

[23] Y. Tsuruoka, J. Tsujii and S. Ananiadou. FACTA: a text search engine for finding associated biomedical concepts. Oxford Journal, 1460-2059, 2008.

[24] A. Rolfs, Y. Hu, L. Ebert, D. Hoffmann, D. Zuo, N. Ramachandran, J. Raphael, F. Kelley, S. McCarron, D. A. Jepson, B. Shen, M. M. A. Baqui, J. Pearlberg, E. Taycher, C. DeLoughery, A. Hoerlein, B. Korn and J. LaBaer. A Biomedically Enriched Collection of 7000 Human ORF Clones. PLoS ONE 3(1), e1528, 2008.

[25] A. Nikitin, S. Egorov, N. Daraselia and I. Mazo. Pathway studio - the analysis and navigation of molecular networks. Bioinformatics, 19(16):2155-7, 2003.

[26] G. Leroy and H. Chen. Filling preposition-based templates to capture information from medical abstracts. Pac. Symp. Biocomput., 350-61, 2002.

[27] T. Fayruzov, M. De Cock, C. Cornelis and V. Hoste. Linguistic feature analysis for protein interaction extraction. BMC Bioinformatics, 10:374, 2009.

[28] C. Blaschke and A. Valencia: The potential use of SUISEKI as a protein interaction discovery tool. Genome Inform, 12:123-134, 2001.

[29] H. Chen and B. M. Sharp. Content-rich biological network constructed by mining PubMed abstracts. BMC Bioinformatics, 5:147, 2004.

[30] S. R. Eddy. Profile hidden Markov models. Bioinformatics, 4(9):755-63, 1998.

[31] S. Nikolajewa, R. Pudimat, M. Hiller, M. Platzer and R. Backofen. BioBayesNet: a web server for feature extraction and Bayesian network modeling of biological sequence data. Nucleic Acids Res., 35:W688-W693, 2007.

[32] P. Palaga, L. Nguyen, U. Leser and J. Hakenberg. High-Performance Information Extraction with AliBaba. Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, St. Petersburg, Russia, 360, 2009.

[33] D. Cheng, C. Knox, N. Young, P. Stothard, S. Damaraju and D. S. Wishart. PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. Nucleic Acids Res., 36:399-405, 2008.

[34] T. J. P. Hubbard, B. L. Aken, S. Ayling, B. Ballester, K. Beal, E. Bragin, S. Brent, Y. Chen, P. Clapham, L. Clarke, G. Coates, S. Fairley, S. Fitzgerald, J. Fernandez-Banet, L. Gordon, S. Graf, S. Haider, M. Hammond, R. Holland, K. Howe, A. Jenkinson, N. Johnson, A. Kahari, D. Keefe, S. Keenan, R. Kinsella, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, K. Megy, P. Meidl, B. Overduin, A. Parker, B. Pritchard, D. Rios, M. Schuster, G. Slater, D. Smedley, W. Spooner, G. Spudich, S. Trevanion, A. Vilella,

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

J. Vogel, S. White, S. Wilder, A. Zadissa, E. Birney, F. Cunningham, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, J. Herrero, A. Kasprzyk, G. Proctor, J. Smith, S. Searle and P. Flicek. Ensembl 2009, Nucleic Acids Res., 37:D690-D697, 2009.

[35] M. Hewett, D. E. Oliver, D. L. Rubin, K. L. Easton, J. M. Stuart, R. B. Altman and T. E. Klein. PharmGKB: the Pharmacogenetics Knowledge Base. Nucleic Acids Res., 30(1):163-5, 2002.

[36] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang and J. Woolsey. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res., 34:D668-D672, 2006.

[37] T. Liang and J.-S. Chen: Multi-class Named Entities Extraction from Biomedical Literature. J. Inf. Sci. Eng., 22(6):1339-53, 2006.

[38] D. C. Y. Fung, S. H. Hong, D. Koschützki, F. Schreiber and K. Xu, 2008: 2.5D Visualisation of Overlapping Biological Networks. Journal of Integrative Bioinformatics, 5(1):1-17, 2008.

[39] U. Brandes, T. Dwyer and F. Schreiber. Visual Understanding of Metabolic Pathways Across Orga-nisms using Layout in Two and a Half Dimensions. Journal of Integrative Bioinformatics, 1(1), 2004.

[40] G. A. Pavlopoulos, S. I. O'Donoghue, V. P. Satagopam, T. G. Soldatos, E. Pafilis and R. Schneider. Arena3D: visualization of biological networks in 3D. BMC Syst. Biol., 2:104, 2008.

[41] Y. Yang, E. S. Wurtele, C. Cruz-Neira and J. A. Dickerson. Hierarchical Visualization of Metabolic Networks Using Virtual Reality. Proc. ACM Intl. Conf. on Virtual Reality Continuum and Its Applications - VRCIA '06, 377-81, 2006.

[42] R. R. Ishiwata, M. S. Morioka, S. Ogishima and H. Tanaka. BioCichlid: central dogma-based 3D visualization system of time-course microarray data on a hierarchical biological network. Bioinformatics, 25(4):543-4, 2009.

[43] R. M. H. Merks and J. A. Glazier. A Cell-Centered Approach to Developmental Biology. Physica A, 352:113-30, 2005.

[44] M. Sugimoto, K. Takahashi, T. Kitayama, D. Ito, M. Tomita. Distributed Cell Biology Simulations with E-Cell System. Lecture Notes in Computer Science, 20-31, 2005.

[45] L. M. Loew and J. C. Schaff. The Virtual Cell: A Software Environment for computational Cell Biology. Trends Biotech., 19(10):401-6, 2001.

[46] A. Camargoa and F. Azuaje. Identification of dilated cardiomyopathy signature genes through gene expression and network data integration . Genomics, 92(6):404-13, 2008.

[47] B. Aranda, P. Achuthan, Y. Alam-Faruque, I. Armean, A. Bridge, C. Derow, M. Feuermann, A. T. Ghanbarian, S. Kerrien, J. Khadake, J. Kerssemakers, C. Leroy, M. Menden, M. Michaut, L. Montecchi-Palazzi, S., N. Neuhauser, S. Orchard, V. Perreau, B. Roechert, K. van Eijk, and H. Hermjakob. The IntAct molecular interaction database in 2010. Nucleic Acids Res., 38:D525-D531, 2009.

[48] A. Zanzoni, L. Montecchi-Palazzi, M. Quondam, G. Ausiello, M. Helmer-Citterich and G. Cesareni. MINT: a Molecular INTeraction database. FEBS Lett., 513(1):135-40, 2002.

[49] D. Maglott, J. Ostell, K. D. Pruitt and T. Tatusova. Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res., 33:D54–D58, 2005.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).

[50] N.A. Kolchanov, E.V. Ignatieva, E.A. Ananko, O.A. Podkolodnaya, I.L. Stepanenko, T.I. Merkulova, M.A. Pozdnyakov, N.L. Podkolodny, A.N. Naumochkin and A.G. Romashchenko. Transcription Regulatory Regions Database (TRRD): its status in 2002. Nucleic Acids Res., 30(1):312-7, 2002.

[51] S. Hunter, R. Apweiler, T. K. Attwood, A. Bairoch, A. Bateman, D. Binns, P. Bork, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, A. Laugraud, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, J. Mistry, A. Mitchell, N. Mulder, D. Natale, C. Orengo, A. F. Quinn, J. D. Selengut, C. J. Sigrist, M. Thimma, P. D. Thomas, F. Valentin, D. Wilson, C. H. Wu and C.Yeats. InterPro: the integrative protein signature database. Nucleic Acids Res., 37:D224-8, 2009.

[52] O. P. Podkolodnaya, E. E. Yarkova, P. D. Demenkov, O. S. Konovalova, V. A. Ivanisenko and N. A. Kolchanov. Application of the ANDCell Computer System to Reconstruction and Analysis of associative Networks describing potential Relationships between Myopia and Glaucoma. Vestnik VOGiS, 14(1):106-15, 2010.

[53] K. T. Momynaliev, S. V. Kashin, V. V. Chelysheva, O. V. Selezneva, I. A. Demina, M. V. Serebryakova, D. Alexeev, V. A. Ivanisenko, E. Aman and V. M. Govorun. Functional divergence of Helicobacter pylori related to early gastric cancer. J. Proteome Res., 9(1):254-67, 2010.

[54] B. Meyer. Self-Organizing Graphs - A Neural Network Perspective of Graph Layout. Lecture Notes in Computer Science, 1547:246-62, 1998.

[55] A. J. Robinson and T. P. Flores. Novel Techniques for Visualizing Biological Information. Proceedings of the 5th International Conference on Intelligent Systems for Molecular Biology, 241-249, 1997.

[56] M. H. P. H. van Beurdena, G. van Hoeyb, H. Hatzakisc, W. A. Ijsselsteijna. Stereoscopic Displays in Medical Domains: A Review of Perception and Performance Effects. Human Vision and Electronic Imaging XIV, Proceedings of the SPIE, 7240:72400A-72400A-15, 2009.

[57] C. Ullmer, K. Schmuck, A. Figge, H. Lübbert. Cloning and characterization of MUPP1, a novel PDZ domain protein. FEBS Lett., 424(1-2):63-8, 1998.

[58] C.B. Coyne, T. Voelker, S. L. Pichla, J. M. Bergelson. The coxsackievirus and adenovirus receptor interacts with the multi-PDZ domain protein-1 (MUPP-1) within the tight junction. J. Biol. Chem., 279(46):48079-84, 2004.

[59] I. J. Latorre, M. H. Roh, K. K. Frese, R. S. Weiss, B. Margolis, R. T. Javier. Viral oncoprotein-induced mislocalization of select PDZ proteins disrupts tight junctions and causes polarity defects in epithelial cells. J Cell Sci., 118(Pt 18):4283-93, 2005.

[60] L. Berglund, E. Björling, P. Oksvold, L. Fagerberg, A. Asplund, C. A. Szigyarto, A. Persson, J. Ottosson, H. Wernérus, P. Nilsson, E. Lundberg, A. Sivertsson, S. Navani, K. Wester, C. Kampf, S. Hober, F. Pontén and M. Uhlén. A gene-centric human protein atlas for expression profiles based on antibodies. Mol Cell Proteomics, (10):2019-27, 2008.

[61] J. T. Eppig, J. A. Blake, C. J. Bul, J. E. Richardson, J. A. Kadin, M. Ringwald, MGI staff. Mouse genome informatics (MGI) resources for pathology and toxicology. Toxicol. Pathol., 35(3):456-7, 2007.

Copyright 2010 The Author(s). Published by Journal of Integrative Bioinformatics.
This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/3.0/).