

Methodology article

Open Access

VisANT: an online visualization and analysis tool for biological interaction data

Zhenjun Hu¹, Joseph Mellor¹, Jie Wu² and Charles DeLisi*^{1,2}

Address: ¹Bioinformatics Program, Boston University, Boston, MA 02215, USA and ²Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

Email: Zhenjun Hu - zjhu@bu.edu; Joseph Mellor - mellor@bu.edu; Jie Wu - jiewu@bu.edu; Charles DeLisi* - delisi@bu.edu

* Corresponding author

Published: 19 February 2004

Received: 11 September 2003

BMC Bioinformatics 2004, 5:17

Accepted: 19 February 2004

This article is available from: <http://www.biomedcentral.com/1471-2105/5/17>

© 2004 Hu et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: New techniques for determining relationships between biomolecules of all types – genes, proteins, noncoding DNA, metabolites and small molecules – are now making a substantial contribution to the widely discussed explosion of facts about the cell. The data generated by these techniques promote a picture of the cell as an interconnected information network, with molecular components linked with one another in topologies that can encode and represent many features of cellular function. This networked view of biology brings the potential for systematic understanding of living molecular systems.

Results: We present **VisANT**, an application for integrating biomolecular interaction data into a cohesive, graphical interface. This software features a multi-tiered architecture for data flexibility, separating back-end modules for data retrieval from a front-end visualization and analysis package. VisANT is a freely available, open-source tool for researchers, and offers an online interface for a large range of published data sets on biomolecular interactions, including those entered by users. This system is integrated with standard databases for organized annotation, including GenBank, KEGG and SwissProt. VisANT is a Java-based, platform-independent tool suitable for a wide range of biological applications, including studies of pathways, gene regulation and systems biology.

Conclusion: **VisANT** has been developed to provide interactive visual mining of biological interaction data sets. The new software provides a general tool for mining and visualizing such data in the context of sequence, pathway, structure, and associated annotations. Interaction and predicted association data can be combined, overlaid, manipulated and analyzed using a variety of built-in functions. **VisANT** is available at <http://visant.bu.edu>.

Background

The growing catalogue of biological data includes information discovered by methods that detect interactions between different biological molecules. Some of these techniques are direct and experimental (e.g. yeast two-hybrid, chromatin-immunoprecipitation (ChIP)) while others are indirect, predictive and computational (e.g., phylogenetic profiling [1], protein binding prediction[2],

and *cis*-element detection [3] and gene expression profiling) Instances of such interactions are the observed or predicted relationships between genes and proteins, and for the purpose of computational storage and analysis they can be represented as networks of functional association. Tools are needed which gather, display and facilitate analysis of these large data structures.

Interaction discovery techniques continue to emerge and evolve. Although they vary in accuracy, the confidence in any particular association is highest when made by a combination of measures [4]. What is true for the individual link in this case is also true for the network; that is, that the prediction of functional pathways and regulatory subunits in the cell is best accomplished by the combination of many measures of interaction, be they experimental or computational, between DNA, proteins, or any molecule in the cell. The results of Ideker and Thorsson, *et al*[5], Jansen, *et al*[6], and Yanai and DeLisi[7] suggest the potential value in combining multiple interaction types in analyzing global systems. In contrast to biological sequence databases, for which uses and applications are well established, the development of databases and associated tools for organizing, mining and analyzing molecular systems has begun relatively recently. To date, the focus in this rapidly evolving field has been mainly on tools for describing and visualizing experimental interaction networks [8-10], derived from gene expression and protein interaction data sources. Research on new methods, both computational and experimental, that describe associations among genes and proteins will continue to necessitate flexible data models that can grow to fit the needs of analysis and visualization. The broader problem of multiple data type integration will largely depend on the usefulness of these emerging data models.

Published databases such as BIND [11], KEGG [12], Predictome [13] and STRING[14] provide the conceptual platforms on which software for leveraging the full content of the interactome could operate. Early efforts in this area, such as the protein-protein binding databases DIP [15] and PathCalling [16], demonstrated usefulness in dynamic visualization of interaction networks, by allowing users to navigate among links in those particular data sets. Recent visualization and analysis tools such as Cytoscape [8], MintViewer [9] and Osprey [10] have expanded this concept. They include features for viewing and querying larger subsets of the interactome on a more global scale. The tools typically operate from the viewpoint of physical associations between proteins, or correlated gene expression, and include information that summarizes annotated functions, such as Gene Ontology (GO)[17] groupings, among subnetworks of linked genes or proteins. Missing from the current bioinformatics palette, though, is a generic interaction network tool capable of managing and analyzing the more abstract forms of interaction information that are available and regularly published. The VisANT tool, while not a complete solution in responding to this need, is nonetheless robust and useful for many different data types and analyses.

Important features that VisANT offers to the research community are (i) navigation of database-driven interaction

and association networks, (ii) visual comparison, manipulation and storage of known networks and uploaded user-defined data, (iii) the ability to uncover orthologous networks, and (iv) the ability to perform exploratory data mining and basic graph operations on arbitrary networks and sub-networks, including loop detection, degree distribution (the distribution of edges per node) and shortest path identification between various component genes or proteins.

Results

Design

One of the major design goals is flexibility – both with respect to the assimilation of new types of data, and the need for evolving a graphical interface that can fit new techniques for describing biological networks. For example, if new computational methods are published for identifying *cis*-regulatory elements upstream of yeast genes, we want putative interactions derived from these methods to be easily compared with other interactions, such as those determined by chromatin immuno-precipitation (ChIP) assays. For some problems, users might be interested only in those experimentally determined interactions, such as protein-protein or protein-DNA binding – the physical interactome. Where experimental data is limited, or biased towards genes with well-understood function, research of gene networks can benefit from use of systematically derived interactions produced *in silico*[2]. The increasingly data-driven state of research biology suggests that analysis of high-throughput data is necessarily more exploratory than hypothesis-driven. VisANT is designed to allow a wide range of exploratory questions. A researcher interested in interaction data for the uncharacterized gene *YLR089C* in *Saccharomyces cerevisiae*, which has uncharacterized function, could explore the network of known interactions in different species for conserved features. Investigations of biologically directed questions, such as finding transcription factors putatively linked downstream of known receptor proteins, could be used in generating hypotheses regarding new molecular functions or modes of gene regulation [8,18,19].

Regardless of the motivating problem, users should be able to identify potentially meaningful features of networks such as shortest paths, dense nodes (i.e. nodes with a large number of connected edges), highly-connected subgraphs, or network motifs such as directed loops or feedforward loops which appear to be biologically important [20-23].

Implementation

We applied the Model-View-Controller (MVC)[24] design pattern for the architecture of VisANT. The tiered system separates the data abstraction (the process by which a particular data type is represented and stored, e.g., a two-

dimensional adjacency matrix represents interactions between pairs of proteins) and retrieval layers from the presentation schema, which improves data integrity and increases flexibility. In particular, the tiered system allows us to put data control logic at the middle-tier to protect the data. Since the presentation of the stored data is separated from the data itself, users can modify the visual data (such as x, y coordinates, node size and labels) without modifying the data stored in the database, or make changes in any tier, without effecting the others, which makes the system easily maintainable and extensible.

The system is implemented using J2EE™ technology using a web service layer driven by the freely available Tomcat server. This data layer technology is both server- and platform-independent. It can be readily adapted to different computer systems and, with additional effort, other data sources. This enables other interaction databases to reuse VisANT as a visual analysis tool by implementing a web service layer with a database-specific Application Program Interface, and the VisANT data transportation format. Technical details of this implementation can be found in the source code of VisANT, as well as its user manual that is available at <http://visant.bu.edu/vmanual/>[25].

VisANT is accessed through its main web page, <http://visant.bu.edu>[26] using a compatible web browser and Java Runtime Engine (JRE). The visual tool has been tested on Netscape, Mozilla and Internet Explorer browsers, running Java (JRE 1.1 or greater) on Windows 2000/XP, Linux and Mac OSX. For the most reliable performance of the software, we recommend using a newer version of Java (JRE ≥ 1.4), freely available from Sun at <http://java.sun.com>[27]. The source code for VisANT is available on request.

Visual exploration of biological interaction network

The main interface of VisANT, the network visualization panel (Fig 1), displays a set of connected nodes or *vertices* corresponding to user selected gene IDs (the nodes) and the experimental methods that uncovered the connections (the connecting lines, or *edges*). Each vertex thus contains annotation information, and each edge stores the method used in assigning the link. Different experimental methods are captured on the screen by using edges of different color; consequently different edges can have different meanings. Some represent actual physical interactions between proteins (e.g. from yeast two-hybrid.); some connect a transcription factor to the protein encoded by the gene downstream of the regulatory sequence to which it binds (ChIP); others represent correlated functions (e.g. those determined by phylogenetic profiling[28].) Edges between transcription factors and the products of the genes they regulate are represented by

arrows, to indicate causal direction. All other edges are currently undirected.

A network is constructed by entering ORF IDs, GI numbers, or even KEGG pathway IDs for an arbitrary number of genes, and using data obtained by one or any combination of methods shown in the methods menu. Nodes corresponding to the selected genes will then appear on the screen, and by left clicking one or more times, they can be expanded into an increasingly complex set of interactions. Figure 1 is a screen shot of VisANT showing the connections in a segment of the MAPK regulatory network constructed by data from Lee et. al[29], and correlations in microarray experiments published by Hughes, *et al.* [30] VisANT algorithms find paths between receptors (STE2, SHO1, and MID2) and transcription factors (STE12, SWI4) in the MAPK network, revealing complex feedback relationships that possibly contribute to regulatory control in these pathways.

Additional functionality is supported by the Predictome database, which maintains look-up tables that store and associate synonyms and annotations for the same protein/gene, and which also facilitates the integrative analysis of the network with function, structure and sequence annotation. VisANT also provides functions to load user-defined interaction data with a single mouse-click, enabling easy comparison between different data sets. The number of viewable genes, proteins and interactions can range from few (as shown in Fig. 1) to thousands. To simplify and help filter the larger data sets, different layout algorithms combined with the built-in basic graph operations, such as closed loops, help to isolate network topology features that have potential biological implications. [20-22,31,32]

The relaxing layout algorithms implemented in VisANT are all based on a similar core heuristic algorithm [33] which models a two-dimensional network of physical objects with mechanical forces operating along the edges. The source code for these algorithms is based on modifications of a layout program distributed by Sun Corporation [27]. Although the algorithms have no biological meaning, they successfully separate the graph by the density of the connections between subgroups of nodes, providing a visual method of identifying relatively dense subgraphs within larger networks. Additional graph operations are generally provided through the various filters whose functions are detailed in the user manual on the VisANT website.

Data integration

Our public VisANT implementation currently draws information from the Predictome database, based on data from 66 fully sequenced microbial genomes. Higher

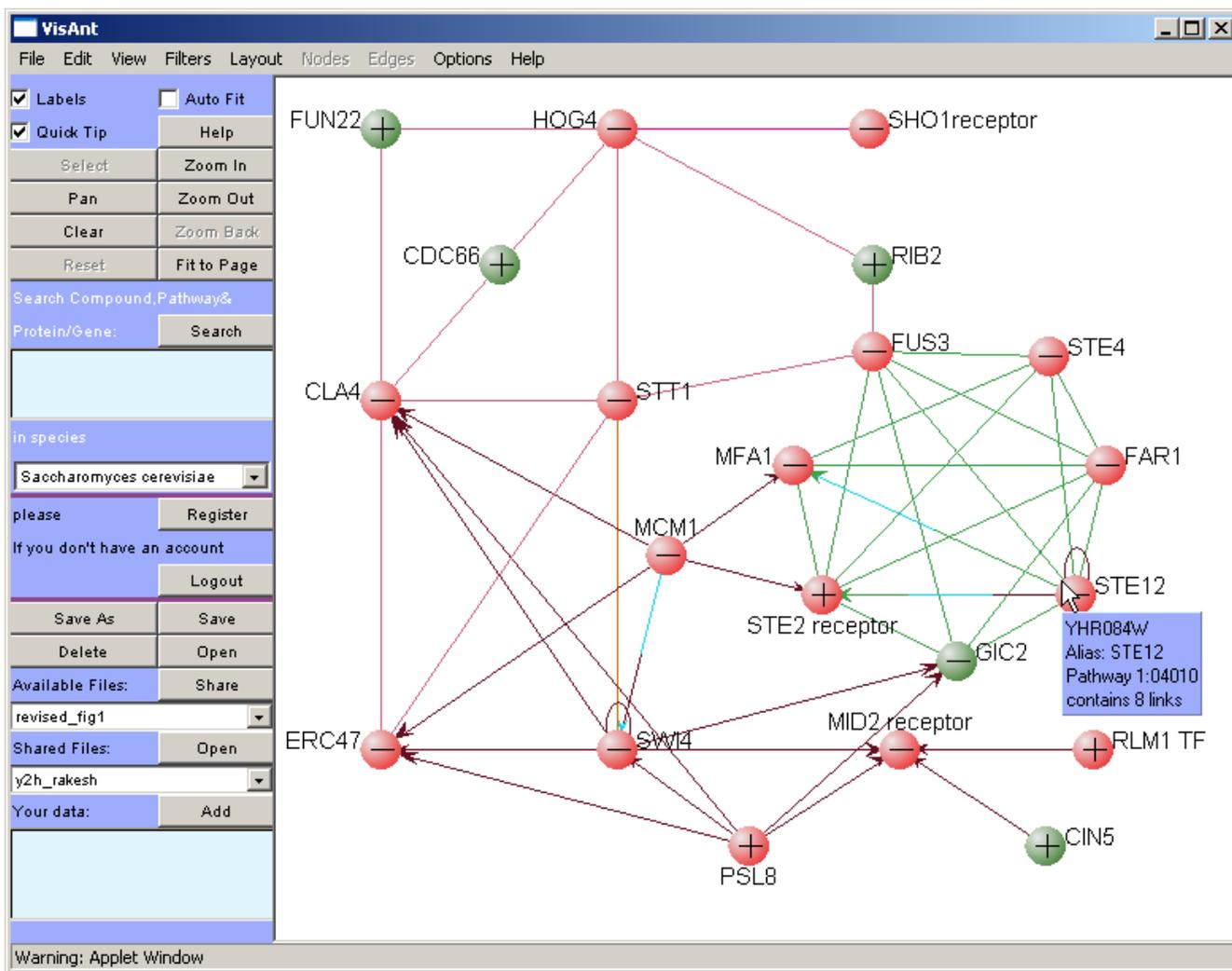


Figure 1
Sample view of a VisAnt application. Displayed are connections in a segment of the MAPK regulatory network constructed by data from Lee et. al.,[29] (Brown lines with arrows, indicating binding of protein to DNA) and correlations in microarray experiments published by Hughes, et al[30] (green lines), as well as links established by protein-protein binding etc. Genes for membrane-bound receptors, and related pathway proteins and transcription factors linked by physical interaction and gene expression relation are shown. Protein/DNA is represented as the nodes. Red nodes represent proteins that are annotated in at least one KEGG pathway (the quick-tip of node STE12 indicates that it maps to KEGG pathway 04010). A "-" indicates that the node is fully expanded (i.e. all connections are shown) while the "+" indicates that some links have not yet been displayed. Correlations between nodal proteins are indicated by connecting lines (edges), different colors corresponding to different experimental methods.

eukaryotes, including worm, human and mouse, are not yet supported, although we do include parsed versions of their genomes, so that networks orthologous to those in microbes can be mined. Computational methods in this database include phylogenetic profiling, gene fusion and gene proximity data. Experimental data drawn from publicly available data include protein-protein and protein-DNA interactions (*S. cerevisiae*), as well as gene expression

correlation and association data. VisANT provides a general platform for the integrative research on interaction networks in the context of pathway, sequence, structure and associated annotation. Pathway data is provided by the KEGG database based on the KEGG Markup Language (KGML)[35] which is currently available only for metabolic pathways. The COG[36] database was used to provide homology information for relationships between

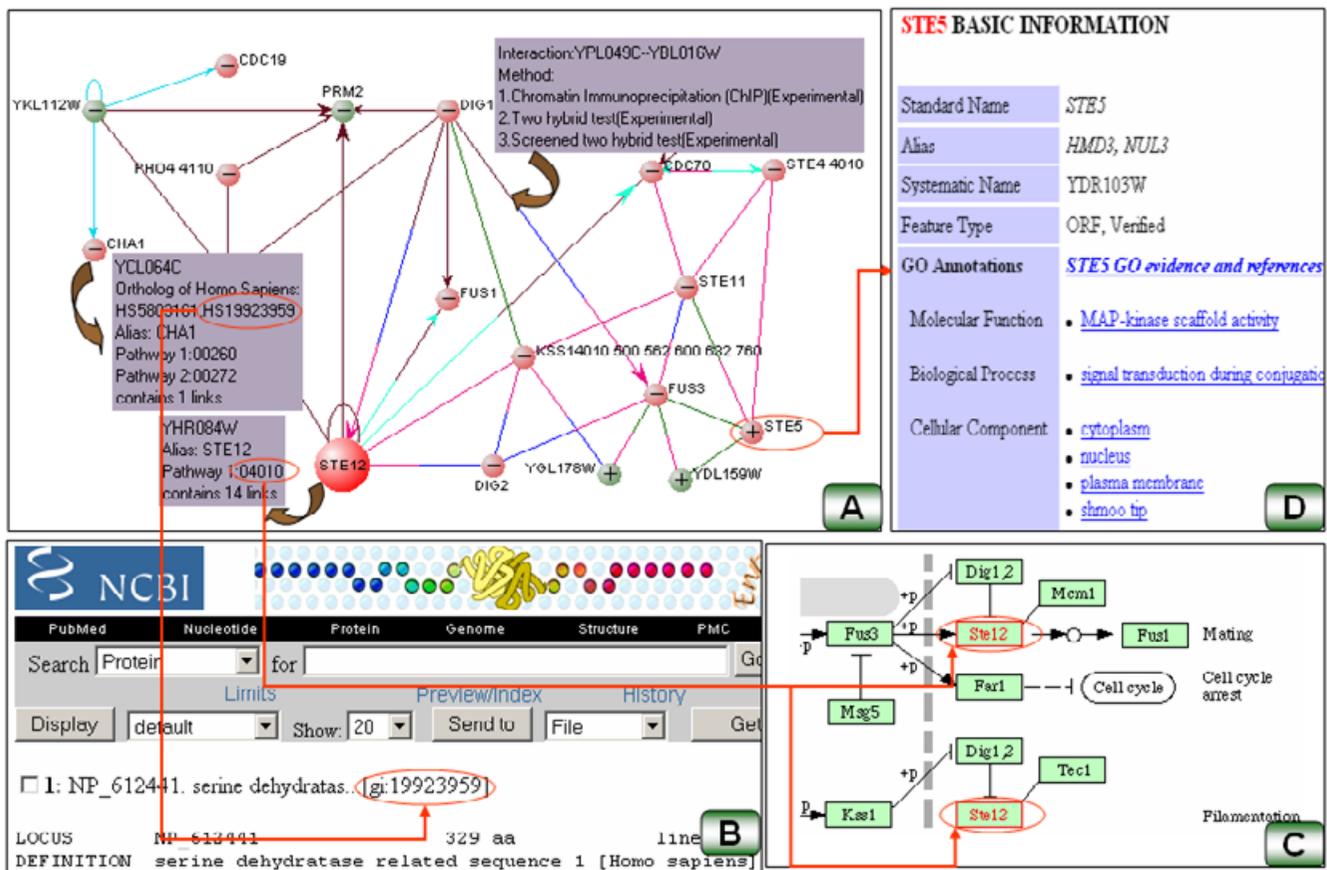


Figure 2
Illustration of data integration in VisANT. (A) The MAPK related network constructed from receptors and transcription factors in the pheromone-response pathway. Purple rectangles demonstrate the quick-tip obtained by mouse-overs of the edge between DIG1 and FUS3, and the nodes CHA3 and STE12 respectively. Most integration data are available only after the node has been queried against the databases, and are available under the "Available Links" submenu of the node. (B) GenBank[37] record of human homology protein for CHA1 based on COG database. The homology information is available after the corresponding filter has been processed. (C) STE12 is mapped to KEGG pathway 04010 (MAPK Singling Pathway) and the pathway has been loaded with corresponding nodes highlighted. (D) Functional annotation of STE5 is loaded through the cross-reference in SGD[49] database.

species. Annotation information is drawn from KEGG and the Gene Ontology, and cross-referencing of genes and proteins to GenBank [37] and SwissProt is provided.

When an interaction/association is discovered by more than one method, the corresponding edge will be segmented with different colors corresponding to the methods. These colors can be customized using the built-in method table. An example can be found in Figure 2A, which shows that the interaction between DIG1 and FUS3 is recovered by three different experimental methods. Red nodes in Figure 2A indicate that they have been mapped to KEGG pathways. For example, quick-tips show that STE12 has been mapped to KEGG pathway 04010, while

CHA1 is mapped to both KEGG pathways 00260 and 00272. KEGG pathways are directly referenced from within VisANT, with corresponding nodes highlighted as shown in Figure 2C. For *S. cerevisiae* proteins, annotation from SGD has also been referenced, as shown in Figure 2D. GenBank[37] sequence information has been referenced in similar fashion.

Interaction networks of arbitrary genes in microbes can be projected onto groups of orthologous human genes, providing hypothetical relationships between human genes. This projection is based on the COG ortholog database, coupled with filters provided by VisANT. Figure 2A shows that CHA1 has two ortholog proteins in human

(19923959 and 5803161, GI number) and Figure 2B displays the GenBank[37] record of human protein 19923959 directly referenced in VisANT.

Network storage and sharing

Visualization and comparison of different interaction networks (networks obtained with separate methods) is an important means of validation and understanding the relative contribution of different methods to functional understanding. VisANT allows users to enter customized data sets through the control panel as shown in Figure 1, and to overlay these data sets upon one another, or upon published datasets. Where multiple data sets based on similar methods have been published (e.g. yeast 2-hybrid screening in *S. cerevisiae*), the reference to each source is cited. The data format for user-specified data is simple tab-delimited, and can represent either directed or undirected associations. VisANT also provides password-protected saving of each customized graphical workspace to allow further analysis of a particular network at any time from anywhere on the internet. In addition, these individual workspaces can be securely shared, to promote collaboration within and among research groups.

Discussion

Although networks and pathways can be visualized and navigated using clickable images [15], the data mining process requires more than visualization. Visual data mining is mediated by a collection of interactive methods that support exploration of data sets by adjusting parameters to see how they affect the information being presented. The functionalities provided by VisANT reflect this approach, especially as it applies to biological networks.

Both genome-wide and conventional interaction data can be noisy and error prone [38]. The integration of interaction data from various data sources is critical for improving the accuracy of these data [38-42]. Data integration also requires the unification of heterogeneous data (such as expression data, sub-cellular localization information, and functional category etc.) into one general data model so that different analyses can be carried out easily. Clustering of gene expression, for example, may be guided by knowledge of protein localization, or participation of genes in the interaction network [43].

Other visual integration tools, such as Cytoscape [8], Osprey [10] and GenMAPP[44], are able to display varying aspects of physical interaction and expression data and relate this to functions and pathway annotation. VisANT differs conceptually from these tools in the notion that all such information – interaction, expression, function – can be represented and analyzed as a network. The dimensions of these networks can be very large, thus presenting a major and still incompletely met challenge for

visual integration and computation. Table 1 summarizes the differences between the three programs.

Future Directions

The goal of the VisANT project is to provide a general platform for visually mining process-level annotation[45]. This annotation, sometimes called functional, relates the genome to cellular processes: growth, apoptosis, differentiation and so on. Our first step focuses on protein/gene interaction mining and visualization. As the interaction network turns to functional modules/pathways and networks, corresponding functions will be implemented to support further analyses, including simulation of cellular activities.

Specifically, VisANT enhancements in the near future will include the following:

1. Visualization and graph manipulation. The data model will be further generalized to represent different types of bio-objects and the interactions between them. Visual representation of nodes and edges will be enhanced and standardized [46,47]. An immediate goal is to produce a data model and related functionalities that support abstract groupings and modularity based on function or experimental evidence, in order to facilitate the full integration with groupings such as KEGG pathways, GO annotations, and diverse objects such as protein complexes [23]. These groupings enable a more modular analysis of structure within interaction networks.
2. Inclusion of the full complement of KEGG pathways.
3. Support for higher eukaryotes including worm, human and mouse. Analysis and comparison of interactions across species will continue to be improved. Specifically, we are interested in the concept of cross-species mapping to facilitate direct comparison of the conservation of networks between different organisms.
4. The implementation of additional features for integrating data sources. For example, VisANT will be able to load microarray data either from standard databases (GEO [48] etc.), or from a user's local file. Third party open-source software, such as TM4 [34], will be integrated to enable direct analysis of expression data in context of mined networks.
5. VisANT's architecture will be further enhanced to enable pluggable parsers and filters, providing the flexible interfaces to facilitate the integration of heterogeneous data sources and third-party's analysis. Correspondingly, VisANT will be able to run as both a signed on-line java applet and standalone application.

Table 1: Comparison of VisANT against Cytoscape and Osprey

	VisANT	Cytoscape	Osprey
Graph Manipulation	Developed in house	Based on yWorks package	Developed in house
Data Generalization*	Node can either be Protein/Gene, chemical compound, KEGG pathway	Protein/Gene/Compound	Protein/Gene
Input	<ul style="list-style-type: none"> • Up-to-date Predictome database, both experimental/computation data • On-line network file saved per user • Tab-delimited user data through copy/paste 	<ul style="list-style-type: none"> • Local file with its own format for both network and expression data. 	<ul style="list-style-type: none"> • Up-to-date Grid database with experimental data only • local file with its own format
Output	<ul style="list-style-type: none"> • On line saving of work space per user • Export as tab-delimited file • Jpg image 	<ul style="list-style-type: none"> • Local file • Image printing 	<ul style="list-style-type: none"> • Local file • Image printing • Vector graph
Search	<ul style="list-style-type: none"> • On-line searching using either name, GI number, ORF ID, KEGG pathway ID etc. 	<ul style="list-style-type: none"> • Search node name on the graph 	<ul style="list-style-type: none"> • Search node name on the graph
Network Operation	<ul style="list-style-type: none"> • Various layouts • Network motif detection including feed-forward, feedback, shortest path, auto-regulation etc. 	<ul style="list-style-type: none"> • Various layouts 	<ul style="list-style-type: none"> • Various layouts
Filtering	<ul style="list-style-type: none"> • Filtering the data with the different methods • Compare different data sets • Select nodes with certain range of connections 	<ul style="list-style-type: none"> • Flexible filters with different attributes of node and edge 	<ul style="list-style-type: none"> • Filtering the data with the different methods/resources • Compare different data sets
Multi-species	<ul style="list-style-type: none"> • Support 66 species • Enable orthology searching against Human genome 	N/A	Support 4 species
Integration	<ul style="list-style-type: none"> • KEGG database • Genbank database • SGD database 	<ul style="list-style-type: none"> • Go database 	<ul style="list-style-type: none"> • Go database
Run Mode	<ul style="list-style-type: none"> • On-line java applet 	<ul style="list-style-type: none"> • Stand-alone application 	<ul style="list-style-type: none"> • On-line java applet • Stand-alone application
Java version	1.1 above	1.4 above	1.4 above
License	Open source	Open source	Commercial, free to academy

*All three programs can display any string as a node. The components listed in this row are those that have biological annotations

We expect that these and other directions of VisANT will also be augmented and assisted by feedback from the research community.

Authors' contributions

JM and ZH contributed to software concept. ZH implemented the system and performed major programming work. ZH, JM and JW contributed to the underlying Predictome database. This work was directed by CD. All authors have read and approved the final manuscript.

Acknowledgements

This work was funded by NIH grant IP20GM066401-01.

References

1. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO: **Assigning protein functions by comparative genome analysis: protein phylogenetic profiles.** *Proc Natl Acad Sci U S A* 1999, **96**:4285-4288.
2. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M: **A Bayesian networks approach for predicting protein-protein interactions from genomic data.** *Science* 2003, **302**:449-453.

3. Frith MC, Hansen U, Weng Z: **Detection of cis-element clusters in higher eukaryotic DNA.** *Bioinformatics* 2001, **17**:878-889.
4. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D: **Detecting protein function and protein-protein interactions from genome sequences.** *Science* 1999, **285**:751-753.
5. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L: **Integrated genomic and proteomic analyses of a systematically perturbed metabolic network.** *Science* 2001, **292**:929-934.
6. Jansen R, Greenbaum D, Gerstein M: **Relating whole-genome expression data with protein-protein interactions.** *Genome Res* 2002, **12**:37-46.
7. Yanai I, DeLisi C: **The society of genes: networks of functional links between genes from comparative genomics.** *Genome Biol* 2002, **3**:research0064.
8. Ideker T, Ozier O, Schwikowski B, Siegel AF: **Discovering regulatory and signalling circuits in molecular interaction networks.** *Bioinformatics* 2002, **18 Suppl 1**:S233-40.
9. Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G: **MINT: a Molecular INTERaction database.** *FEBS Lett* 2002, **513**:135-140.
10. Breitkreutz BJ, Stark C, Tyers M: **Osprey: a network visualization system.** *Genome Biol* 2003, **4**:R22.
11. Bader GD, Betel D, Hogue CW: **BIND: the Biomolecular Interaction Network Database.** *Nucleic Acids Res* 2003, **31**:248-250.
12. Kanehisa M, Goto S, Kawashima S, Nakaya A: **The KEGG databases at GenomeNet.** *Nucleic Acids Res* 2002, **30**:42-46.

13. Mellor JC, Yanai I, Clodfelter KH, Mintseris J, DeLisi C: **Predictome: a database of putative functional links between proteins.** *Nucleic Acids Res* 2002, **30**:306-309.
14. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B: **STRING: a database of predicted functional associations between proteins.** *Nucleic Acids Res* 2003, **31**:258-261.
15. Xenarios I, Fernandez E, Salwinski L, Duan XJ, Thompson MJ, Marcotte EM, Eisenberg D: **DIP: The Database of Interacting Proteins: 2001 update.** *Nucleic Acids Res* 2001, **29**:239-241.
16. **CuraGen Corporation.** <http://portal.curagen.com>.
17. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
18. Roberts CJ, Nelson B, Marton MJ, Stoughton R, Meyer MR, Bennett HA, He YD, Dai H, Walker WL, Hughes TR, Tyers M, Boone C, Friend SH: **Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles.** *Science* 2000, **287**:873-880.
19. Steffen M, Petti A, Aach J, D'Haeseleer P, Church G: **Automated modelling of signal transduction networks.** *BMC Bioinformatics* 2002, **3**:34.
20. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks.** *Science* 2002, **298**:824-827.
21. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of Escherichia coli.** *Nat Genet* 2002, **31**:64-68.
22. Ferrell J. E., Jr.: **Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability.** *Curr Opin Cell Biol* 2002, **14**:140-148.
23. Bader GD, Hogue CW: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC Bioinformatics* 2003, **4**:2.
24. Buschmann Frank: **Pattern-oriented software architecture : a system of patterns.** Chichester ; New York, Wiley; 1996:xvi, 467.
25. **VisANT Manual.** <http://visant.bu.edu/vmanual/>.
26. **VisANT.** <http://visant.bu.edu/>.
27. **Java.** <http://java.sun.com/>.
28. Wu J, Kasif S, DeLisi C: **Identification of functional links between genes using phylogenetic profiles.** *Bioinformatics* 2003, **19**:1524-1530.
29. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA: **Transcriptional regulatory networks in Saccharomyces cerevisiae.** *Science* 2002, **298**:799-804.
30. Hughes TR, Marton MJ, Jones AR, Roberts CJ, Stoughton R, Armour CD, Bennett HA, Coffey E, Dai H, He YD, Kidd MJ, King AM, Meyer MR, Slade D, Lum PY, Stepaniants SB, Shoemaker DD, Gachotte D, Chakraburty K, Simon J, Bard M, Friend SH: **Functional discovery via a compendium of expression profiles.** *Cell* 2000, **102**:109-126.
31. Bu D, Zhao Y, Cai L, Xue H, Zhu X, Lu H, Zhang J, Sun S, Ling L, Zhang N, Li G, Chen R: **Topological structure analysis of the protein-protein interaction network in budding yeast.** *Nucleic Acids Res* 2003, **31**:2443-2450.
32. Fox JJ, Hill CC: **From topology to dynamics in biochemical networks.** *Chaos* 2001, **11**:809-815.
33. Eades P: **A heuristic for graph drawing.** *Congressus Numerantium* 1984, **42**:142-160.
34. **TM4.** <http://www.tigr.org/software/tm4/index.html>.
35. **KGML.** <http://www.genome.ad.jp/kegg/xml/>.
36. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV: **The COG database: new developments in phylogenetic classification of proteins from complete genomes.** *Nucleic Acids Res* 2001, **29**:22-28.
37. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic Acids Res* 2003, **31**:23-27.
38. Edwards AM, Kus B, Jansen R, Greenbaum D, Greenblatt J, Gerstein M: **Bridging structural biology and genomics: assessing protein interaction data with known complexes.** *Trends Genet* 2002, **18**:529-536.
39. Bader GD, Hogue CW: **Analyzing yeast protein-protein interaction data obtained from different sources.** *Nat Biotechnol* 2002, **20**:991-997.
40. Gerstein M, Lan N, Jansen R: **Proteomics. Integrating interactomes.** *Science* 2002, **295**:284-287.
41. Sprinzak E, Sattath S, Margalit H: **How reliable are experimental protein-protein interaction data?** *J Mol Biol* 2003, **327**:919-923.
42. Tong AH, Drees B, Nardelli G, Bader GD, Brannetti B, Castagnoli L, Evangelista M, Ferracuti S, Nelson B, Paoluzi S, Quondam M, Zucconi A, Hogue CW, Fields S, Boone C, Cesareni G: **A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules.** *Science* 2002, **295**:321-324.
43. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M., Jr., Haussler D: **Knowledge-based analysis of microarray gene expression data by using support vector machines.** *Proc Natl Acad Sci U S A* 2000, **97**:262-267.
44. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR: **GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways.** *Nat Genet* 2002, **31**:19-20.
45. Stein L: **Genome annotation: from sequence to biology.** *Nat Rev Genet* 2001, **2**:493-503.
46. Kohn KW: **Molecular interaction map of the mammalian cell cycle control and DNA repair systems.** *Mol Biol Cell* 1999, **10**:2703-2734.
47. Pirson I, Fortemaison N, Jacobs C, Dremier S, Dumont JE, Maenhaut C: **The visual display of regulatory information and networks.** *Trends Cell Biol* 2000, **10**:404-408.
48. **GEO Database.** <http://www.ncbi.nlm.nih.gov/geo/>.
49. **SGD Database.** <http://www.yeastgenome.org/>.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

