

# VisANT 3.0: new modules for pathway visualization, editing, prediction and construction

Zhenjun Hu<sup>1</sup>, David M. Ng<sup>3</sup>, Takuji Yamada<sup>2</sup>, Chunnuan Chen<sup>3</sup>, Shuichi Kawashima<sup>4</sup>, Joe Mellor, Bolan Linghu<sup>1</sup>, Minoru Kanehisa<sup>2</sup>, Joshua M. Stuart<sup>3</sup> and Charles DeLisi<sup>1,\*</sup>

<sup>1</sup>Center for Advanced Genomic Technology, Boston University, Boston, MA 02215, USA, <sup>2</sup>Bioinformatics Center, Institute for Chemical Research, Kyoto University, Japan, <sup>3</sup>Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA and <sup>4</sup>Human Genome Center, Institute of Medical Science, University of Tokyo, Japan

Received January 31, 2007; Revised April 9, 2007; Accepted April 12, 2007

## ABSTRACT

With the integration of the KEGG and Predictome databases as well as two search engines for coexpressed genes/proteins using data sets obtained from the Stanford Microarray Database (SMD) and Gene Expression Omnibus (GEO) database, VisANT 3.0 supports exploratory pathway analysis, which includes multi-scale visualization of multiple pathways, editing and annotating pathways using a KEGG compatible visual notation and visualization of expression data in the context of pathways. Expression levels are represented either by color intensity or by nodes with an embedded expression profile. Multiple experiments can be navigated or animated. Known KEGG pathways can be enriched by querying either coexpressed components of known pathway members or proteins with known physical interactions. Predicted pathways for genes/proteins with unknown functions can be inferred from coexpression or physical interaction data. Pathways produced in VisANT can be saved as computer-readable XML format (VisML), graphic images or high-resolution Scalable Vector Graphics (SVG). Pathways in the format of VisML can be securely shared within an interested group or published online using a simple Web link. VisANT is freely available at <http://visant.bu.edu>.

## INTRODUCTION

Biological pathways are often represented as pixel images (JPEG, GIF, etc.) or vector graphics (Scalable Vector Graphics (SVG) or PostScript). Typical examples of such static representations include those presented in databases

such as KEGG (1), Reactome (2), BioCarta (<http://www.biocarta.com>) and EcoCyc (3). Although a static representation is intuitive and informative and has been widely used in textbooks and illustrations, it is difficult to edit, or to reuse for analysis, modeling and simulation. As a result, important resources such as the KEGG database cannot be fully exploited. Notable steps toward meeting the challenge of computable representations include the development of BioPAX (Biological Pathways Exchange, <http://www.biopax.org/>) and KGML (KEGG Markup Language, <http://www.genome.jp/kegg/xml/>) with BioPAX focusing on detailed ontology while KGML includes layout information.

A number of software tools (4–9) have been developed to visually build computable models of pathways. These tools are usually based on graphical models in which nodes represent genes, proteins or chemical compounds, and edges represent various types of interactions or associations. To date, few tools support the conditional dependencies of molecular and genetic entities and their associations. Thus, pathways encoded with existing tools may lack key information needed for interpreting the pathway's functioning.

In order to combine multiple pathways in a manner that is useful for modeling cellular behavior, two main challenges must be addressed. First, models must allow a hierarchical visual representation (6,10–12). Second, data representation is complicated when several complexes share some of their proteins, because the role of a common protein generally depends on context (13,14). Methods such as semantic zooming or hierarchical decomposition (10,12,15–20) are needed to aggregate and abstract entire pathways or pathway portions into small units that can be displayed within larger pathway systems. Hierarchical structures are also very common in the computable representation of biological knowledge in BioPAX and KGML formats. A protein complex must

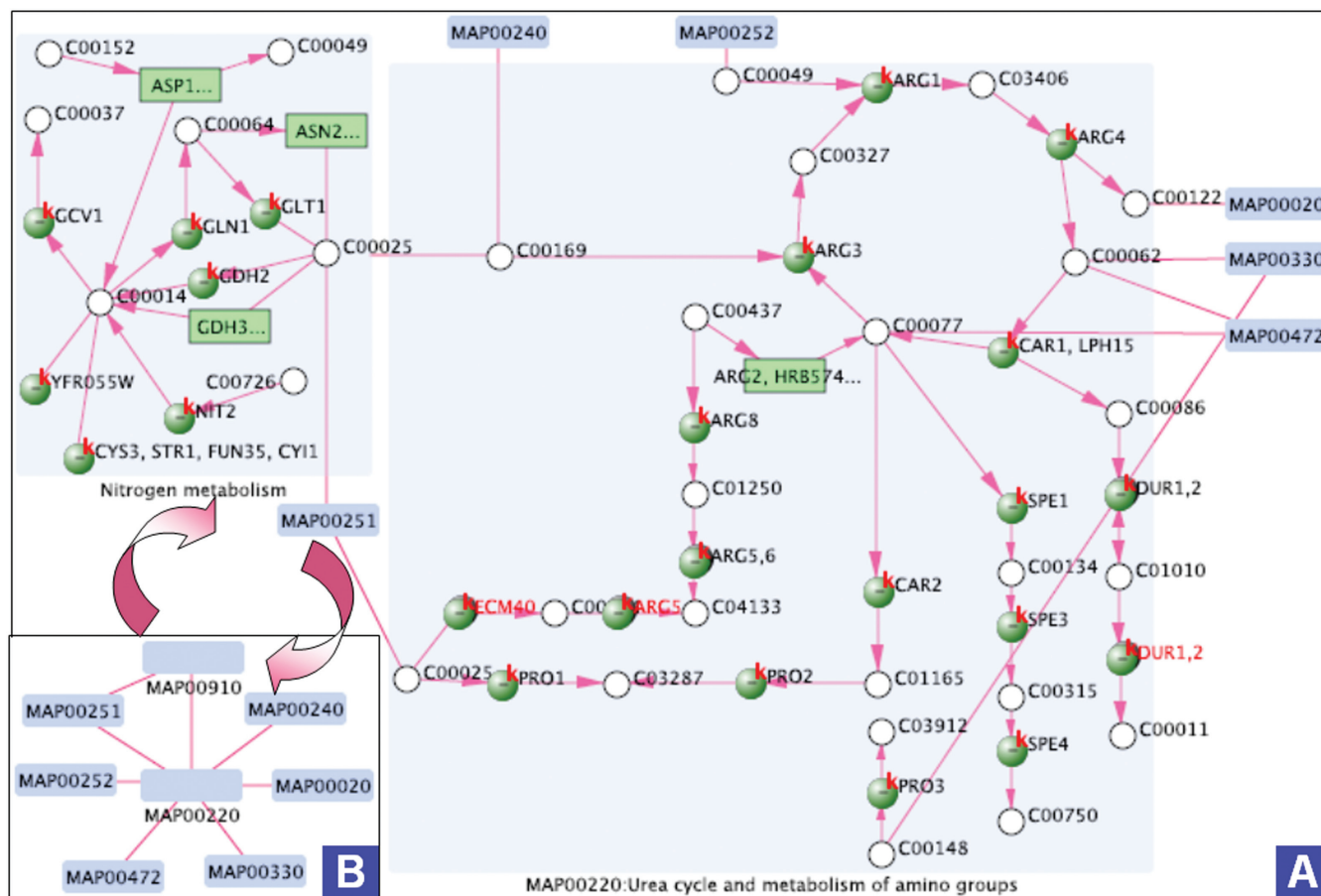
\*To whom correspondence should be addressed. Tel: +617 353 1122; Fax: +617 353 3333; Email: [delisi@bu.edu](mailto:delisi@bu.edu)

Present address:

Joe Mellor, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Illustration of multi-scale visualization of pathways using VisANT's metagraph capability. Green boxes are contracted metanodes representing a group of enzymes that catalyze a common reaction; filled green circles represent enzymes and open circles represent compounds. (A) KEGG Yeast pathways shown in VisANT by first loading MAP00220 and then expanding MAP00910. The connection between MAP00910 and MAP00251 is shown only after MAP00910 is expanded. (B) The eight pathways are all represented as contracted linked metanodes. A and B can be alternately displayed by toggling the state of the metanodes.

often be represented as a node containing a set of nodes, one for each subunit. Each subunit in turn may itself contain a set of nodes representing conserved domains identified in the subunit's 3D structure or primary sequence. Representing a protein complex as a simple, non-hierarchical, node often obscures properties of the proteins because attributes of the simple node are aggregated across multiple proteins, each of which may have different attributes with respect to one another. An obvious workaround for this issue is to model protein complexes as 'compound nodes' (10,11,15) or 'metanodes', which are nodes with recursive internal structure (Figure 1) (20).

While biological systems contain an appreciable amount of hierarchical organization, molecular components are reused across subsystems, making it impossible to perfectly capture all of the information into a nested set of relations. Strict hierarchical representations can capture biological substructure but cannot model overlap between protein complexes. Related to this idea, nodes that represent only a single protein may not have a unique state but may instead behave in a condition-dependent manner. It is common practice (21) to use multiple nodes

to represent different states of the same protein to maintain clarity of control logic and conditional dependency in pathways. However, this can lead to an explosively growing chain of nodes. It also breaks data integrity and introduces data redundancy, as the same protein is represented by multiple nodes. More importantly, the exact conditional-dependent state of a given protein can be unclear or unknown in many pathways. A typical example can be found for protein STE20 in the MAK1 signaling pathway for yeast ([http://www.genome.ad.jp/dbget-bin/show\\_pathway?sce04010+YHL007C](http://www.genome.ad.jp/dbget-bin/show_pathway?sce04010+YHL007C)), which most likely has different activities under different conditions, but the exact nature of the state differences is currently unknown. How can such conditional dependencies be represented and modified when corresponding biological information becomes available?

Protein-protein interaction data sets obtained from either large-scale experiments or computational predictions, as well as coexpressed genes predicted from large-scale expression data, can be used to help fill gaps in incomplete pathways (22–25). Although many tools provide facilities to visualize expression data in the context of pathways (4,5,7,9,26), facilities to enrich pathways in

a computationally based visualization system, using both interaction and expression profiles, are missing.

Here we report new developments in VisANT 3.0, a Web-based platform with new modules supporting exploratory pathway analysis using metagraphs (20) to address multi-scale visualization of multiple pathways; editing and annotating pathways using a KEGG compatible visual notation; visualization of expression data in the context of pathways; enriching pathways using either coexpressed components of known pathway members predicted from expression data in the SMD (27) and GEO (28) databases or proteins with known physical interactions and assigning genes/proteins of unknown function to known pathways. The new version of VisANT will help users take full advantage of the large number of available resources in the KEGG pathway database when building new pathways.

## METHODS

### Metagraphs

A metagraph is a data structure for representing nodes, edges and subnetworks in a nested structure. One significant difference between a compound graph and a metagraph is that metagraphs allow one node to have multiple instances and these instances are automatically tracked. This capability allows a metanode in a metagraph to share nodes: each metanode has its own instance of the same node. Metanodes have two semantic states: an expanded state that reveals the associated subgraph inside, and a contracted state that hides the internal structure, rendering the metanode as a simple node. Edges between the nodes in an expanded metanode have the usual meaning (associations based on experimental data or computationally inferred correlations); edges between metanodes either reflect a correlation between standard (hidden) nodes or indicate that the same gene/protein occurs in both metanodes (20).

### KGML and pathway integration tools

KGML is an exchange format for KEGG graph objects, particularly KEGG pathways, which are manually drawn and updated. The KGML files for KEGG metabolic pathways specify how enzymes (boxes) are linked by a relation and how compounds (circles) are linked by a reaction. In contrast, the KGML files for KEGG regulatory pathways contain only the former. KGML files for all supported species in VisANT have been preprocessed to map genes to their KEGG pathways, and a VisANT user can identify pathways for a specified gene either by searching for its interactions or resolving (normalizing) its names or IDs as explained subsequently.

Two pathway recommendation web services for identifying functionally related genes from transcriptional profiles are integrated in VisANT through its plugin architecture (20). Given a set of query genes, typically the known genes of a pathway, these services recommend additional genes in the same pathway as the query set. Both search engines support five species: *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*

and *Saccharomyces cerevisiae*. When VisANT is run as an online applet, connections to the services are mediated by the VisANT server.

GeneRecommender (29) discovers new genes with similar function to a given list of genes (the query) already known to have closely related function. It ranks genes according to how strongly they correlate with a set of query genes in those experiments for which the query genes are most strongly coregulated.

ClueGene (30) uses the pattern of how genes cluster together in sets of experiments to recommend new genes in a pathway. ClueGene bases its recommendations on the query set and on a cluster compendium. Each set of experiments is clustered independently. The collection of clusters constitutes the cluster compendium. Each gene in the genome is given a co-clustering score. Higher scoring genes are more highly recommended and tend to be found in small clusters in the cluster compendium along with query genes.

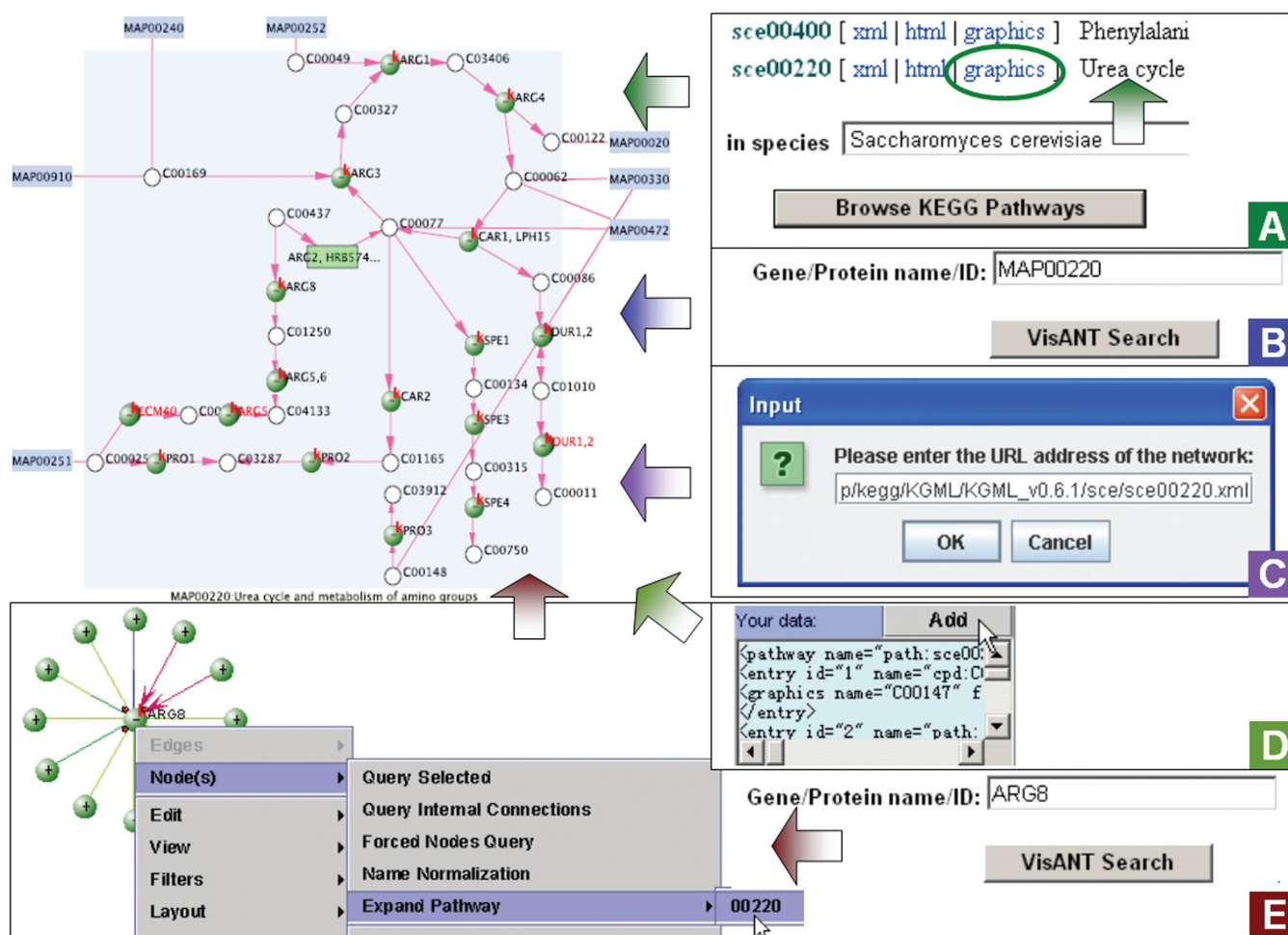
The use of VisANT (20,31,32) to mine, integrate and display biological interactions based on KEGG pathways and expression data is facilitated by a name-normalization service which resolves IDs used by different databases. In addition, customized ID mappings, as well as corresponding Web links, can be easily added to the network through a simple tab-delimited format. VisANT is developed using Java technology. In addition to the Web browser applet interface, VisANT can also be run as a stand-alone application which implements an auto-upgrading detection system to keep it up-to-date. Detailed information on VisANT's three-tier structure (31) and plugin framework(20) can be found at <http://visant.bu.edu>. In addition, a new error-reporting system has been implemented to enhance the integration reliability of distributed systems: users will have the option to report critical errors to the plugin authors and VisANT team.

## INPUT AND OUTPUT

### Input

VisANT automatically recognizes the format of an input file based on its content. Only those formats that related with the new functions will be discussed here. The full list of supported files can be found in VisANT's web site.

Pathways can be loaded into VisANT using several different input methods as detailed in Figure 2. In particular, double-clicking on a contracted pathway node (e.g. the blue boxes in Figure 1) will also load the pathway if the corresponding KGML file is available from the KEGG. Expression data is input from a common tab-delimited file. The first column can be an Entrez Gene ID, an Access ID/GI number, a gene name or an ID from an organism-specific database. The file can have a header line to indicate the names of the different experiments; otherwise, VisANT will use a sequential number to identify different experiments. If the expression data is to be overlaid on an existing pathway, the name normalization service should be utilized first so that genes in the network and in the expression data can be matched to each other.



**Figure 2.** Different methods of loading pathways into VisANT. (A) KEGG pathways for a given species can be directly browsed and pathways of interest can be loaded by clicking the 'graphics' link. A pathway can also be located and loaded by using its (B) name or ID, (C) a URL or (D) KGML contents copied/pasted into the 'Add' box. (E) Pathway IDs will be shown in a node's tooltip if the protein/gene is involved in one or more pathways. In such a case, a pathway can be directly loaded using a set of drop down menus as shown above.

## Output

All data shown in VisANT can be saved in an XML format using the VisANT Markup Language (VisML). Registered users can save the network on the VisANT server so that it can be accessed wherever the internet is available. VisML uses a version number to facilitate compatibility and extensibility. A description of VisML can be found at <http://visant.bu.edu>. In addition, pathways can be exported without visual information, as tab-delimited edge and node lists. Pathways can also be saved as pixel images, or as high-quality SVG for publication and illustration. An SVG file can be further polished with an SVG editor.

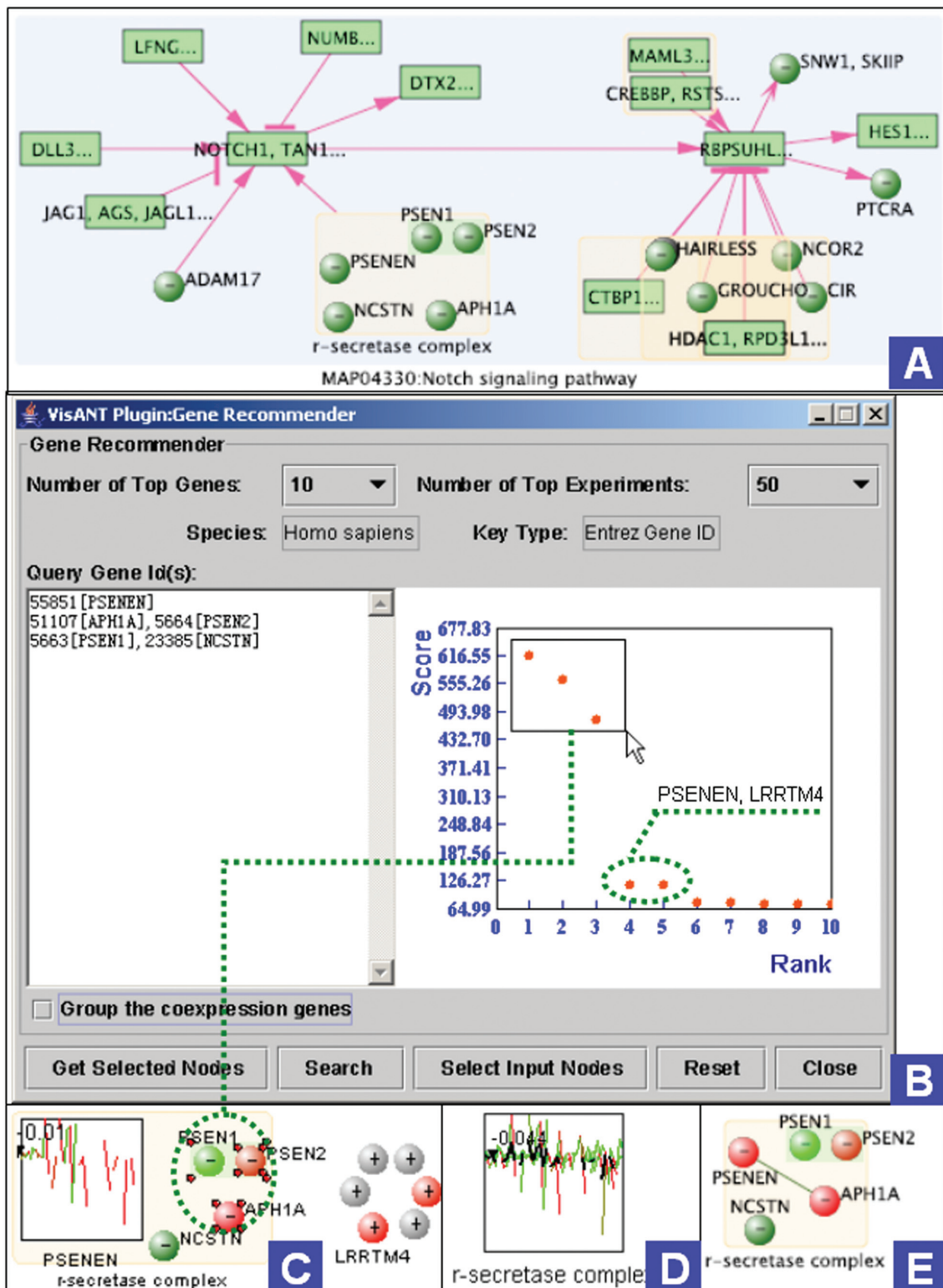
## PROGRAM DESCRIPTION

### Pathway visualization, navigation and editing

Each pathway is represented as a metanode which may be nested within other metanodes (Figure 1). If links to other pathways are available in KGML, these pathways are represented as contracted metanodes.

VisANT adopts the KEGG notation for graphics annotation so that users will have consistent views of the KEGG pathways. However, a few changes were necessary. In particular, a single protein/gene is represented as a filled green circle, and a metanode displayed as a green box is used to represent multiple proteins/genes. Additionally, the number of proteins/genes contained in a metanode can be revealed by double-clicking the box. Use of a metanode for a protein complex is also introduced (Figure 3). Multiple instances of the same node can exist even in the same pathway (ARG5,6 in Figure 1). These instances can be tracked by pressing the right mouse button over the corresponding node. Dashed lines will connect all instances of the node. The lines vanish once the mouse button is released.

Pathways can be easily edited in VisANT. Nodes and edges can be modified, added or deleted. Additional components can be added to pathways by a simple drag and drop. Pathways can be easily ungrouped or regrouped as one large pathway, depending on the user's needs.



**Figure 3.** Illustration of pathway enrichment. (A) *H. sapiens* notch signaling pathway, the  $\gamma$ -secretase complex is represented as nested metanodes. Five components are shown in the complex. (B) VisANT plugin for the GeneRecommender service searches for the top 10 coexpressed genes with a given set of query genes shown in the left panel. Score plot of the top 10 coexpressed genes is shown in the right panel. (C) Expression data of the top 50 experiments, as well as additional genes excluding the query set, are returned and integrated with the existing network. New genes are shown using the circle layout. Expression of PSENEN is shown as an embedded plotter. Color is used to represent expression for the rest nine other genes with red for low expression values and green for high values. Gray is used to indicate that the expression value of the gene for the current experiment is missing. (D) The metanode of the  $\gamma$ -secretase complex is contracted and the expression plotter is turned on. All expression profiles for the member components are plotted together; the average profile is shown in black. The discontinuity is due to missing values for some components of corresponding experiments. (E) Querying against the Predictome database indicates that APH1A and PSENEN are functionally related (36).

### Multi-scale visualization, pathway overview and crosstalk

As with the extension of interactions for a given protein/gene, pathways can be extended by double-clicking on a pathway node. Using this method, a network of pathways can be quickly constructed. Figure 1A shows the network of pathways by first loading pathway MAP00220 and then expanding the pathway MAP00910. It is worth noting that crosstalk between MAP00251 and MAP00910 mediated by the compound C00025 is only visible after MAP00910 is expanded.

Because the state of a metanode can be toggled by mouse-clicking, an overview of the pathway shown in Figure 1B can be easily achieved by contracting the two pathway nodes MAP00910 and MAP00220. Thus, VisANT is capable of easily exploring pathways at different scales: a pathway overview enables users to observe the topology of large sets of pathways, while the detailed internal structure of any particular pathway or set of pathways is easily revealed by mouse-clicking.

### Overlaying expression data

VisANT provides two methods to visualize expression data over pathways: either the node color is used to represent the expression value in a particular experiment, or a plot of the expression profile is embedded in the node, as shown in Figure 3. The two methods can be toggled either for individual nodes or for the whole network. Different experiments can be navigated using a sliding bar and the navigation process can be animated. When the expression profile is shown, the corresponding experiment and expression value is indicated by a cursor.

In VisANT it is convenient to determine whether genes in the same pathway are coexpressed, as all the expression profiles of the nodes contained in a metanode (pathway), as well as the average profile, are drawn together as one plot with average profiles in black. Figure 3D shows such an example for a node representing a protein complex.

### Pathway prediction

Sets of genes in the same pathway are often activated together and may have very similar expression profiles; their protein products may also interact, either physically or functionally, to achieve a specific task. VisANT provides functions to assign genes/proteins with unknown function to the known KEGG pathways based on these observations. Predictome (33) can easily be queried for sets of proteins that interact either functionally or physically with a specified protein. VisANT also has editing capabilities that allow any such set to be augmented with a user's own data set.

Genes with similar expression profiles can be identified using the ClueGene and GeneRecommender plugins and the genes so identified can be associated with one or another KEGG pathway in accordance with user specified criteria based on either functional or physical links (Figure 2E) (25,34,35). Query genes can be placed in identified pathways by a simple drag and drop.

We suggest that users test the coexpression of query genes with known genes in the potential pathways and

compare scores using either ClueGene or GeneRecommender. In addition, expression profiles can be compared if query genes are searched using GeneRecommender.

### Pathway construction, enrichment and update

New pathways can be created from scratch or from relevant KEGG pathways, the latter of course being substantially more convenient because of KEGG documentation. In collaboration with the KEGG, the VisANT web site lists all pathways for which KGML is available, allowing easy access and loading into VisANT (Figure 2A). These reference KEGG pathways can also be updated when necessary. When loaded into VisANT, they can be enriched either by querying functionally associated components from experimental and computational results accessible from the VisANT-Predictome system, or by searching for coexpressed genes as indicated above.

We next describe a use-case scenario to illustrate some of the new features of VisANT. Suppose a user is interested in the  $\gamma$ -secretase complex which acts in the *H. sapiens* notch signaling pathway (Figure 3A), and wishes to get more knowledge about related genes or the internal structure of the  $\gamma$ -secretase complex. First, the GeneRecommender plugin can be used to search for potential genes coexpressed with the five component members of the complex: APH1A, NCSTN, PSEN1, PSEN2 and PSENEN. GeneRecommender returns the top 10 coexpressed genes scored in the top 50 experiments. As can be seen from Figure 3B, the scores of the coexpressed genes can be separated into three groups. The top group, APH1A, PSEN1 and PSEN2, has much higher scores than the second group, PSENEN and LRRTM4. The plotter is linked to the network and selecting a spot in the plotter will select the corresponding node in the network (Figure 3B and C). Note that query gene NCSTN is not included in the top 10 coexpressed genes, indicating that NCSTN is not positively correlated with other members of the complex. Anti-correlations are very common in signaling pathways (Figure 3A); future implementations of the search engines will support identification of anticorrelated genes. Users may select different combination of query genes to achieve the best results. In addition, the degree of coexpression between members of a given metanode can be viewed by contracting the metanode and turning on the expression plotter option, as shown in Figure 3D. To further test the correlation of the 11 genes shown in Figure 3C, interactions between pairs of genes are queried against the Predictome database, which reveals the interaction between PSENEN and APH1A identified by coimmunoprecipitation (36), as shown in Figure 3E.

In addition, pathways can be updated against the KEGG database so that the latest pathway information can be easily incorporated into existing pathways customized by the users.

### FUTURE DEVELOPMENT

Among our goals for further development of VisANT is supporting pathways from other databases, including

Reactome (2), BioCarta (<http://www.biocarta.com>), EcoCyc (3) and INOH (<http://www.inoh.org/>). Since computable representations of pathways from these databases are available in BioPAX format, one way to proceed would be to increase VisANT's support of BioPAX. This will require developing an automatic layout algorithm since BioPAX, unlike KGML, does not contain layout information. More importantly, a standard visual notation for different types of nodes and edges will also need to be developed for different types of biological components, and for the relations between them. Second, unlike KGML in which each pathway is usually stored in its own file, pathways in BioPAX format are usually represented in one large file which can exceed 100 MB, making it impractical to load them all at once and also preventing exploratory navigation of pathways. New efforts, such as the latest developments in CPath (<http://cbio.mskcc.org/cpath/home.do>) have made significant progress to overcome this problem by providing corresponding Application Programming Interfaces (APIs) that can retrieve pathways one by one in the format of BioPAX. We expect obstacles discussed above will be removed in the near future and pathways from these databases will be ready for use in VisANT.

## AVAILABILITY

VisANT along with the full user manual and tutorials are available on the VisANT web site, <http://visant.bu.edu>

*Conflict of interest statement.* None declared.

## REFERENCES

- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M. and Hirakawa, M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
- Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R. *et al.* (2005) Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
- Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M. and Karp, P.D. (2005) EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.*, **33**, D334–D337.
- Chung, H.J., Park, C.H., Han, M.R., Lee, S., Ohn, J.H., Kim, J., Kim, J. and Kim, J.H. (2005) ArrayXPath II: mapping and visualizing microarray gene-expression data with biomedical ontologies and integrated biological pathway resources using Scalable Vector Graphics. *Nucleic Acids Res.*, **33**, W621–W626.
- Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C. and Conklin, B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.*, **31**, 19–20.
- Demir, E., Babur, O., Dogrusoz, U., Gursoy, A., Nisanci, G., Cetin-Atalay, R. and Ozturk, M. (2002) PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, **18**, 996–1003.
- Mlecnik, B., Scheideler, M., Hackl, H., Hartler, J., Sanchez-Cabo, F. and Trajanoski, Z. (2005) PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways. *Nucleic Acids Res.*, **33**, W633–W637.
- Ng, A., Bursteinas, B., Gao, Q., Mollison, E. and Zvelebil, M. (2006) pSTING: a 'systems' approach towards integrating signalling pathways, interaction and transcriptional regulatory networks in inflammation and cancer. *Nucleic Acids Res.*, **34**, D527–D534.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Fukuda, K. and Takagi, T. (2001) Knowledge representation of signal transduction pathways. *Bioinformatics*, **17**, 829–837.
- Demir, E., Babur, O., Dogrusoz, U., Gursoy, A., Ayaz, A., Gulesir, G., Nisanci, G. and Cetin-Atalay, R. (2004) An ontology for collaborative construction and analysis of cellular pathways. *Bioinformatics*, **20**, 349–356.
- Herman, I., Melancon, G. and Scott Marshall, M. (2000) Graph visualization and navigation in information visualization: a Survey. *IEEE Trans. Vis. Comput. Graph.*, **6**, 24–43.
- Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141–147.
- Spirin, V. and Mirny, L.A. (2003) Protein complexes and functional modules in molecular networks. *Proc. Natl Acad. Sci. USA*, **100**, 12123–12128.
- Sugiyama, K. and Misue, K. (1991) Visualization of structure information: automatic drawing of compound digraphs. *IEEE Trans. Syst. Man Cybern.*, **21**, 876–892.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabasi, A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, **297**, 1551–1555.
- Gagneur, J., Krause, R., Bouwmeester, T. and Casari, G. (2004) Modular decomposition of protein-protein interaction networks. *Genome Biol.*, **5**, R57.
- Klukas, C. and Schreiber, F. (2007) Dynamic exploration and editing of KEGG pathway diagrams. *Bioinformatics*, **23**, 344–350.
- Saraiya, P., North, C. and Duca, K. (2005) Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Information Visualization*, **4**, 191–205.
- Hu, Z., Mellor, J., Wu, J., Yamada, T., Holloway, D. and Delisi, C. (2005) VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res.*, **33**, W352–W357.
- Kitano, H., Funahashi, A., Matsuoka, Y. and Oda, K. (2005) Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.*, **23**, 961–966.
- Stuart, J.M., Segal, E., Koller, D. and Kim, S.K. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–255.
- DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680–686.
- Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. and Davis, R.W. (1997) Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl Acad. Sci. USA*, **94**, 13057–13062.
- Segal, E., Wang, H. and Koller, D. (2003) Discovering molecular pathways from protein interaction and gene expression data. *Bioinformatics*, **19**(Suppl. 1), i264–i271.
- Junker, B.H., Klukas, C. and Schreiber, F. (2006) VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, **7**, 109.
- Ball, C.A., Awad, I.A., Demeter, J., Gollub, J., Hebert, J.M., Hernandez-Boussard, T., Jin, H., Matese, J.C., Nitzberg, M. *et al.* (2005) The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Res.*, **33**, D580–D582.
- Barrett, T., Troup, D.B., Wilhite, S.E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I.F., Soboleva, A., Tomashevsky, M. *et al.* (2007) NCBI GEO: mining tens of millions of expression profiles – database and tools update. *Nucleic Acids Res.*, **35**, D760–D765.
- Owen, A.B., Stuart, J., Mach, K., Villeneuve, A.M. and Kim, S. (2003) A gene recommender algorithm to identify coexpressed genes in *C. elegans*. *Genome Res.*, **13**, 1828–1837.

30. Ng,D.M., Woehrmann,M.H. and Stuart,J.M. (2007) Recommending pathway genes using a compendium of clustering solutions. *Pac. Symp. Biocomput.*, **12**, 379–390.
31. Hu,Z., Mellor,J., Wu,J. and DeLisi,C. (2004) VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics*, **5**, 17.
32. Hu,Z., Mellor,J. and DeLisi,C. (2004) In Baxevanis,A., Davison,D., Page,R., Petsko,G., Stein,L. and Stormo,G. (eds), *Current Protocols in Bioinformatics.*, John Wiley & Sons, Inc., New Jersey, USA.
33. Mellor,J.C., Yanai,I., Clodfelter,K.H., Mintseris,J. and DeLisi,C. (2002) Predictome: a database of putative functional links between proteins. *Nucleic Acids Res.*, **30**, 306–309.
34. Ge,H., Liu,Z., Church,G.M. and Vidal,M. (2001) Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. *Nat. Genet.*, **29**, 482–486.
35. Jansen,R., Greenbaum,D. and Gerstein,M. (2002) Relating whole-genome expression data with protein-protein interactions. *Genome Res.*, **12**, 37–46.
36. Hasegawa,H., Sanjo,N., Chen,F., Gu,Y.J., Shier,C., Petit,A., Kawarai,T., Katayama,T., Schmidt,S.D. *et al.* (2004) Both the sequence and length of the C terminus of PEN-2 are critical for intermolecular interactions and function of presenilin complexes. *J. Biol. Chem.*, **279**, 46455–46463.