Information Visualization (2005), 1–15
© 2005 Palgrave Macmillan Ltd. All rights reserved 1473-8716 $30.00
www.palgrave-journals.com/ivs

Visualizing biological pathways: requirements analysis, systems evaluation and research agenda

Purvi Saraiya[1,2]
Chris North[1,2]
Karen Duca[3]

[1]Center for Human–Computer Interaction, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; [2]Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA; [3]Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, USA.

Correspondence: Chris North, 660 McBryde Hall, Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0106.
Tel: +1 540 231 2458;
Fax: +1 540 231 6075;
E-mail: north@vt.edu

Abstract

Pathway diagrams are used by life scientists to represent complex interactions at the molecular level in living cells. The recent shift towards data-intensive bioinformatics and systems-level science has created a strong need for advanced pathway visualizations that support exploratory analysis. This paper presents a comprehensive list of requirements for pathway visualization systems, based on interviews conducted to understand life scientists' needs for pathway analysis. A variety of existing pathway visualization systems are examined, to analyze common approaches by which the contemporary systems address these requirements. A heuristic evaluation, by biology domain experts, of five popular pathway visualization systems is conducted to analyze the end-user perception of these systems. Based on these studies, a research agenda is presented concerning five critical requirements for pathway visualization systems. If addressed effectively, these requirements can prove to be most helpful in supporting exploratory pathway analysis. These include: (1) automated construction and updating of pathways by searching literature databases, (2) overlaying information on pathways in a biologically relevant format, (3) linking pathways to multi-dimensional data from high-throughput experiments such as microarrays, (4) overviewing multiple pathways simultaneously with interconnections between them, (5) scaling pathways to higher levels of abstraction to analyze effects of complex molecular interactions at higher levels of biological organization.

Information Visualization advance online publication, 23 June 2005; doi:10.1057/palgrave.ivs.9500102

Keywords: Information visualization; biological pathways; human–computer interaction (HCI); requirements analysis

Introduction

Biological pathways represent networks of complex reactions at the molecular level in living cells. They model how biological molecules interact to accomplish a biological function and to respond to environmental stimuli. Pathways capture the current knowledge of biological processes and are derived through scientific experimentation and data analysis. Life scientists use pathways to integrate results from literature, formulate hypotheses, capture empirical results, share current understanding, and even simulate processes. A common goal of research in the life sciences is to develop an ever-broadening library of pathway models for biological processes of many different organisms. Such pathways can have significant broad impacts, such as making products in biotech applications and drug discovery in the pharmaceutical industry.

Received: 25 January 2005
Revised: 18 April 2005
Accepted: 29 April 2005

Pathways also serve as a focal point to integrate other diverse related information, such as literature citations, research notes, and experimental data. In recent years, high-throughput data capture technology has vastly improved life scientists' ability to detect and quantify gene, protein, and metabolite expression. Such experiments can simultaneously provide data about thousands of entities.[1–4] All this data must be analyzed in the context of the pathway diagrams to enable biologists to make inferences about the underlying biological processes and to improve the current pathway models. Hence, the increasing complexity of pathway diagrams derives not only from their size and representations, but also from the large amount of important related information.

The increasing importance of exploratory pathway analysis corresponds to a major shift in emphasis in biological research; a shift beyond the reductionist scientific process, which rigorously examines individual interactions of biological molecules, towards *systems-level* science, which simultaneously explores entire systems of many biological molecules. Systems-level science highlights that the whole is greater than the sum of the parts. A challenging goal for pathways is to try to convey complex global functionality, interconnections with other pathways, and their dynamic behavior.

To facilitate the exploratory analysis of complex pathways, visual representations are necessary. Pathways are typically represented as network diagrams (see Figure 1 for examples). Some pathway diagrams are manually generated such as those found in textbooks[7] or KEGG,[8] whereas others are generated by interactive visualization software such as GenMAPP[9] and PathwayAssist.[10] However, although several pathway visualization systems have been developed recently, there is little guidance for the design of such tools (e.g.[11,12]). Though there have been a few studies on graph layout and aesthetics,[13,14] their utility and impact for pathway visualizations is yet unclear.

In discussions with life scientists, we found that many are skeptical about the biological value of current pathway visualizations. When considering cost *vs* benefit, the cost seems to outweigh the benefits. They are reluctant to invest time required to overcome the learning curve for many of these systems. A large amount of effort is required to gain biologically meaningful insight for specific projects from most of these systems. The tools lack many important data analysis capabilities that scientists need. Thus, to truly enable a shift towards systems-level science, more rigorous requirements analysis and evaluation of pathway visualization systems are needed.

This paper aims to apply human–computer interaction (HCI) methods to enable a more principled scientific approach to solve the difficult problem of pathway visualization. The first goal is to understand life scientists' usage of pathway diagrams through open-ended and informal interviews and questionnaires, and to generate a

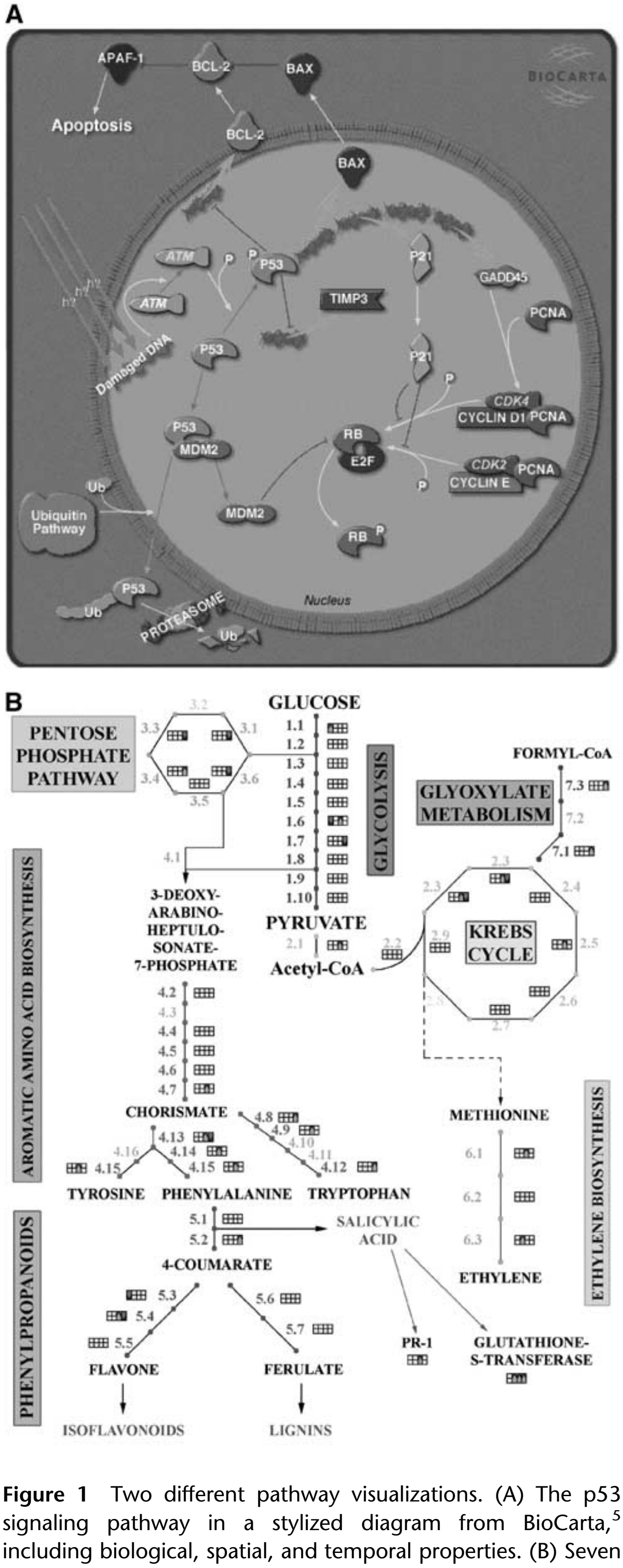

Figure 1 Two different pathway visualizations. (A) The p53 signaling pathway in a stylized diagram from BioCarta,[5] including biological, spatial, and temporal properties. (B) Seven inter-related metabolic pathways in *Arabidopsis thaliana*, including gene expression measurements on a time series at four time points indicated in the small colored boxes.[6]

comprehensive list of requirements for pathway visualization software. The second goal is to evaluate several existing pathway visualization systems, and to examine some common approaches used by the contemporary systems to address user requirements. We evaluated a few popular pathway visualization tools, including GenMAPP,[9] Cytoscape[15] and Pathway Assist,[10] with life scientists (the intended end users and domain experts) with respect to the requirements in order to examine end-user perception of these systems. Finally, based on these studies, we identify critical areas of visualization design that can prove to be most important with respect to user requirements, and a research agenda to seek out most effective solutions. Our hope is to provide guidance to bioinformatics software designers in the future development of pathway visualizations, and to focus HCI and visualization researchers on these critical needs.

Procedure

Generally in HCI, analysis of requirements starts with interviewing and observing current work practices of users. These observations can be contextual (users are observed as they carry out their tasks), or participatory (users are engaged in discussions). Results of these observations are scenarios and requirements that help developers understand how users will eventually use a system and its impacts.[16]

We focused on life scientists as the primary user class, and life science research as the primary usage scenario. To understand pathway usage, we interviewed four research professors and postdoctoral fellows having diverse research interests and several years of research experience, over a period of 6 months. We met with each researcher usually once or twice a week. The researchers were selected based on their availability and willingness to participate in the discussions.

We generally interviewed only one researcher at a particular time. Each interview session lasted for about 1–2 h. Most of these interviews were informal and participatory. We did not ask the researchers a specific predefined set of questions. The life scientists explained their research work to us and its biological significance. They also explained importance of biological pathways, different contexts in which pathways are used, different types of information needed from pathways and the current methods to obtain this. The life scientists also discussed their research work, experiments, data analysis tasks, and how pathway diagrams fit into their overall research goals. We also attended presentations and seminars conducted by these life scientists to understand their work in a broader context.

In addition to the interviews, we conducted two focus group meetings, with about 10 life scientists (two of these were researchers we interviewed extensively). In the group meetings, we discussed the requirements derived from earlier interviews. In addition, we attended the journal club meetings of a life science research group, where we discussed published research about high-throughput data experiments. Based on these studies and group meetings, we derived a final list of requirements for pathway analysis. To get feedback from additional life scientists, a short questionnaire was sent via email listservs. The scientists were requested to rate the degree to which they agree or disagree with the requirements.

To analyze the end-user perception of existing pathway visualization systems, we conducted a heuristic evaluation with six life scientists on five pathway analysis systems. Participation in the evaluation was voluntary. This heuristic evaluation was a form of user study in which biology domain experts reviewed systems to suggest advantages and disadvantages against the list of requirements.[17] This approach helps to further elucidate the requirements and how the systems meet biologists' needs. The results provide useful guidance for developing pathway visualization software.

Biological pathways

Pathway description

There is not yet a standardized language for pathway components, as it is highly dependent on the domain and the particular need that motivates the construction of any given pathway. In many cases, a 'pathway' is the user-defined network of the biological interactions under study in a particular research group. Pathways in life science research are extremely diverse. Some capture higher level abstractions, while others are very specific. Some are sketchy, while others are rigorous. Figure 1 shows two examples of different types of pathways. Overall, pathways provide an approximate model or explanation of the underlying biological process.

Typically pathways are represented as a graph, consisting of nodes and edges. A node in a pathway usually represents a biological molecule, but could also be used to summarize another entire pathway that interconnects with the one under study, or to represent any other relevant phenomena such as an environmental stimulus (e.g., heat or light). A node representing a biological molecule in a pathway diagram may be either a metabolite, nucleic acid, or protein. Nucleic acids can be DNA, mRNA, tRNA, and structural RNA, etc. Proteins can be enzymes, structural proteins, chemical effectors, etc. Enzymes are further divided into ligases, phosphatases, kinases, etc. Structural protein can be microtubules, actin filaments, etc. Chemical effectors can be hormones, cytokines, chemokines, growth factors, etc. An edge in a pathway usually represents a relationship or some form of interaction between the nodes. The interaction could be of many types: gene expression, inhibition, catalysis, chemical modification, etc.

Pathway graphs can be complex multi-modal or hypergraphs. While simple graphs can capture the very basic events represented in the pathway, complex biochemical dynamics do not lend themselves well to basic graph representations. An edge could connect three nodes or might connect a node to another edge. For example, an inhibitory interaction (edge) actually indicates a deeper process by which one molecule (node) might prevent some other interaction (edge) from occurring.

Based on the overall effect they have on the functioning of an organism, pathways may be divided into several different categories. Three example categories are: metabolic pathways, gene regulation/transcription pathways, and signal transduction pathways.

In this paper, we emphasize this fairly broad notion of pathways. We do not focus on one type of pathway or specific set of pathway elements because (a) the requirements to analyze different kinds of pathways are similar, and (b) it is a long-term goal to produce software that can integrate a broad variety of pathways to support the grand vision of combined systems-level analysis. Unless explicitly stated otherwise, a pathway in this discussion refers collectively to all types.

User classes

The primary users of pathway visualization tools are advanced academic, industrial and government researchers in the life sciences (i.e. biologists, biochemists, chemists, biomedical researchers, etc.). Their goals are to construct pathway diagrams that model biological phenomena as closely as possible, based on literature and experimental results. This is somewhat analogous to a computer scientist attempting to reverse engineer an algorithm by running the compiled code on a variety of inputs and examining the outputs. Each researcher is generally focused on contributing to a small set of pathways representing their area of interest and expertise. They are very knowledgeable about the details of these pathways. However, they must make use of other pathways for which they may have only general knowledge or know little about.

The life scientists interviewed in this study work in small teams of about 5–10 people. A team includes undergraduate and graduate students, lab technicians, postdocs and senior researchers. Data to construct pathways is generally provided by more senior investigators. Multiple research scientists in the same or different research institutes may collaborate on identical problems. At the highest levels, there are internationally renowned scientists who curate newly made discoveries and resolve discrepancies in research findings, for example, The Alliance for Cellular Signaling (AfCS).[18]

Pathway research process

Pathway research is strongly iterative and evolving. A critical component of the research process that enables biologists to continue the experimental feedback loop[19] is inference. Inference enables them to turn experimental data results into refined hypotheses. Some common pathway inference tasks that biologists perform include: (1) recognition of changes between experiment and control or between time points; (2) detection of changes in relationship between components of a pathway or between entire pathways; (3) identification of global patterns across a pathway; and (4) mapping pathway state to phenotype (observable effects at the physical level in living organisms) or other biological information.[20] Sometimes, the new discoveries fail to support past assumptions, leading to further experimentation and research, culminating in modified pathways. Pathway modification is a continuous, evolutionary process.

Some hypotheses and research questions are relatively simple, and can be answered through scientific reduction methods. However, with the advent of systems-level analysis, it is becoming more common to examine hypotheses that are significantly more complex. Researchers are typically interested in pathways that contain approximately 50–500 nodes. However, when inputs to these nodes from other pathways (that in turn may be affected by several other pathways) need to be taken into account, things quickly get more complicated. Inferences that must be made in these cases are equally complex, requiring the recognition of subtle effects at various levels of scale involving multiple pathway networks. These inferences are well beyond the capabilities of current pathway visualization techniques.

Requirements analysis

Based on the interviews and focus group meetings with life scientists, a list of requirements for pathway visualizations were developed as shown in Table 1. The requirements are grouped into three main categories: pathway assembly, information overlay, and pathway analysis. These categories are described in the following subsections.

Accomplishing these requirements will require interactive dynamic visualizations. Static, textbook-like pathway representations will not be adequate in the long term. While these functional requirements provide guidance, they do not directly dictate visualization design. It might not be possible to adequately satisfy all requirements with a single design, and tradeoffs will likely need to be carefully balanced.

Category: pathway assembly

These requirements support the assembly and maintenance process for pathways.

R1. Construct and update: A complete pathway is generally not available from a single source. Life scientists often must combine different parts of a pathway from various sources, including reference archives such as KEGG,[8] research articles, etc. It is also important to continually capture updates of source information in order to keep a pathway in sync with the latest knowledge.

R2. Context: A pathway may be clear to the author because of deeper understanding of the components (nodes and edges) involved. However, the same diagram may be difficult to understand by someone not familiar with the underlying biological process. It is therefore advisable to include information such as pathway

Table 1 Summary of requirements for pathway visualization systems

Categories	*Requirements*	*Tasks*
Pathway assembly	1. Construct and Update	Collect and link pathways from multiple resources
	2. Context	Provide information about pathways
	3. Uncertainty	Maintain alternate hypotheses and information reliability
	4. Collaboration	Enable group work
Information overlay	5. Node and edge representation	Details about network entities and interactions
	6. Source	Details about source resources
	7. Spatial information	Physical locations of pathway entities in the cell
	8. Temporal information	Time-related properties
	9. High-throughput data	Expression data from high-throughput experiments
Pathway analysis	10. Overview	Comprehend large or multiple pathways
	11. Inter-connectivity	Intra- and inter-pathway effects of entities on each other
	12. Multi-scale	Relate networks at different levels of abstraction
	13. Notebook	Track accumulated research information

The requirements are grouped into three main categories: pathway assembly, information overlay, and pathway analysis.

significance, specific conditions for it to function, collective effects of the pathway components, history of updates, etc., in some form when creating a pathway. If a pathway from a community resource is modified, then the rationale for doing so should be stated explicitly.

R3. Uncertainty: Pathways are constantly evolving. Some relationships between pathway components may be uncertain, and may require more research to be accepted. Known facts should be distinguished from hypotheses. Representations for alternate, potentially conflicting, hypothesis should be supported.

R4. Collaboration: More than one life scientist can be working together on the same pathways. They need ways to communicate effectively with each other.

Category: information overlay

Pathways are tightly linked to many other types of biological information, and it is critical that pathway visualizations depict this richness of information in order to be biologically relevant. Pathway visualizations that look like simple ball-and-stick graph drawings are likely to be considered information-poor, and not biologically meaningful.

R5. Node and edge representation: Pathway nodes and edges have information attributes that visualizations should reveal through their visual representations. Quick interactive access to further details should also be provided. Pathway nodes can represent many different types of entities (e.g., genes, enzymes, etc.), which may have different chemical properties that visualizations should depict. Nodes labels for the entity names must be clearly visible. Life scientists need to attach notes to pathway nodes for future reference, and be able to link them to databases such as GenBank and Gene Ontology for up-to-date information. An edge between two nodes usually implies a certain type of relationship (e.g., expression, catalysis, etc.), perhaps with properties such as rates, that visualizations should depict.

R6. Source: To evaluate a pathway, it is important to have access to the source information for its components, such as literature citations, experimental data, etc.

R7. Spatial information: Visualizations should represent the physical, spatial attributes of the biology of the pathway, such as location within the cell, relative distance, containment, nodes bound to each other, etc. Sometimes the entity represented by the node can be present in different parts of the cell in different states.

R8. Temporal information: Pathways often have time lag information associated with edges. Events can occur strictly in a particular sequence, simultaneously, cyclic, or mutually exclusive. Many pathways have a primary linear structure, with supporting secondary branches.

R9. High-throughput data: A crucial requirement is to examine changes in pathway components based in high-throughput data experiments such as microarrays. Microarrays allow life scientists to measure expression of several thousand genes simultaneously.[1,2] The raw data set needs to be preprocessed before it can be used for analysis.[3,4] Typically, for each experiment, data can be captured for each gene over multiple time points as well as multiple conditions. Hence, pathway nodes contain multi-dimensional quantitative data. This data could also be generated through simulation.

Category: pathway analysis

Pathway visualizations must enable analysis of complex pathways and hypotheses, beyond simple small effects to very large systems-level interactions.

R10. Overview: Pathways can be large, containing hundreds or even thousands of nodes, with complex interactions throughout. Furthermore, since each pathway provides a specialized focused 'view' on a certain biological function within the larger biological system, pathways are neither independent nor isolated. Life scientists need to overview multiple pathways collectively, with layouts that reveal global patterns and effects

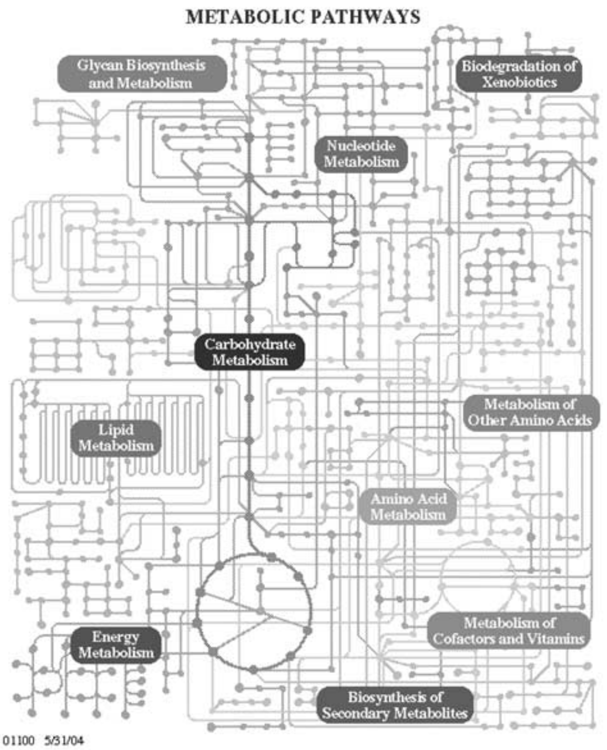
METABOLIC PATHWAYS
Glycan Biosynthesis and Metabolism
Biodegradation of Xenobiotics
Nucleotide Metabolism
Carbohydrate Metabolism
Metabolism of Other Amino Acids
Lipid Metabolism
Amino Acid Metabolism
Metabolism of Cofactors and Vitamins
Energy Metabolism
Biosynthesis of Secondary Metabolites
01100 5/31/04

Figure 2 Provides an overview diagram that shows interconnectivity between metabolic pathways, taken from KEGG.[8]

in context. Figure 2 from KEGG,[8] provides a comprehensive overview for metabolic processes.

R11. Interconnectivity: Pathways are highly interconnected. Components can affect each other directly or indirectly. A single node could be involved in multiple pathways. As complexity increases, it becomes more difficult to understand connections between distant components. Life scientists need to see both upstream and downstream effects from a local region of interest, including other pathways that might be affecting the focal pathway.

R12. Multi-scale: Higher level pathways can be composites of more basic pathways. In the extreme, a small change in a molecular interaction can have substantial effects at physiological levels. In such cases it is necessary to create multiple levels of abstractions to relate molecular components to higher level abstractions, and to be able to relate effects across these levels of scale.

R13. Notebook: A research group might work for several years on a set of pathways. During this time, they might obtain many results about the pathway entities. They need a logical way to keep track of collected information, along with textual notes.

Questionnaire

To validate and prioritize requirements and get feedback from more life scientists, we sent a questionnaire to about 100 life scientists using email listservs. We asked the scientists to rate each requirement according how much they agreed or disagreed with the requirement. A total of 10 scientists responded to the questionnaire. Requirements that are highly rated (strongly agree) by more scientists provides a basis for priority over lower rated (strongly disagree) requirements. Appendix A describes the questionnaire and the number of responses.

Most of the life scientists agreed with the requirements list we compiled. A few of the requirements received many high ratings. The need to assemble pathways from different resources, to link source information, to infer the change in pathway components over several different experiment treatments, and to analyze the influence of pathways on one another were considered very important requirements. Most life scientists commented that they were not satisfied with diagrams provided by current network visualization software. The visualizations should provide information about the biological properties and about the spatial and temporal relationships between the pathway components.

Survey of pathway visualization systems

A large number of systems are available for pathway visualization.[21–23] It would be very difficult to review all the pathway systems. Here, we focus on systems that were selected based on availability, popularity in the bioinformatics community, and visualization and data analysis capabilities. Although the list is not exhaustive, it provides a general overview of capabilities provided and approaches used by the current pathway visualization systems. Owing to the wide range of requirements, it would be difficult for any one system to address all. We group the systems based on the category of requirements they address and the approach that they use.

Category: pathway assembly

A large number of systems have been developed to facilitate pathway construction, using different approaches. Table 2 groups some of these systems based on the pathway assembly requirements they address and the approaches used by these systems to meet the requirements. Reference archives such as KEGG[8] provide a comprehensive list of pathways for different cellular processes. Life scientists frequently use these databases for accurate and up-to-date information on pathway components. A comprehensive list of such reference databases is provided by Pathway Databases.[24] The visualizations provided by these databases are typically static and textbook-like.

Editor tools, such as Pathway Editor[28] and Knowledge Editor,[29] allow users to create pathway visualizations manually. A large number of systems, such as PathwayAssist,[10] PathwayFinder,[31] and PubGene,[32] use Natural Language Processing (NLP) algorithms to generate pathways automatically from research articles retrieved from search engines. Systems such as GenePath[36] infer pathways from microarray data. Vector PathBlazer[34] can

Table 2 Groups systems by the pathway assembly requirements addressed and approaches used

Requirements	*Approaches*	*Systems*
R1: Construct & Update	Reference	KEGG,[8] BIND,[25] STKE,[26] BioCarta,[5] EcoCyc[27]
	Pathway editor tools	Pathway Editor,[28] Knowledge Editor,[29] Unipath[30]
	Construct pathways using NLP algorithms on literature databases	PathwayAssist,[10] PathwayFinder,[31] PubGene,[32] GENIES,[33] Vector PathBlazer,[34] Omniviz[35]
	Construct pathways from microarray data	GenePath,[36] GeneSys,[37] GENEW[38]
	NLP algorithms to update local database	PathwayAssist[10]
	Update database manually	Patika[39]
	Update pathways manually	GenMAPP,[9] Cytoscape[15]
R2: Context	Attach notes	GenMAPP, PathwayAssist, Cytoscape
R3: Uncertainty	Manipulate node and edge properties (e.g., shape, size and color)	GenMAPP, Cytoscape
R4: Collaboration	Facilitate sharing across group members	OmniViz,[35] Biological Story Editor[40]

Table 3 Groups systems by the information overlay requirements addressed and approaches used

Requirements	*Approaches*	*Systems*
R5: Node and edge representation	Manipulate node and edge visual properties (shape, size, color, etc.)	GenMAPP,[9] Cytoscape,[15] GScope[41]
	Provide shapes for different types of nodes	Unipath,[30] Patika,[39] PathwayAssist[10]
R6: Source	Attach source information on nodes and edges	GenMAPP,[9] Cytoscape,[15] PathwayAssist[10]
R7: Spatial information	Provide different shapes to show different cellular locations	GenMAPP
	Manipulate node properties or use fixed layout	Cytoscape, GenMAPP, STKE,[26] PathwayAssist
	Divide visualization into different areas	Patika
R8: Temporal information	Manipulate edge length, or layout pathway elements in the order in which they react	Cytoscape , GenMAPP, PathwayAssist, Vector PathBlazer[34]
	Animations	STKE
R9: High-throughput data	Overlay data on nodes (using color), one condition at a time	Cytoscape, Pathway Assist, GenMAPP
	Embedded views, for multiple conditions (data visualizations such as heatmaps or line charts embedded on or near nodes)	GScope[41]
	Multiple linked views, for multiple conditions (pathways linked to other data visualizations)	GeneSpring[42]
	Visualizations for a functional group	MapMan[43]
	Automatically infer relationships between entities from data	GenePath[36]
	Overlaying replicates	GenMAPP

create pathways by combining information from different reference databases such as KEGG[8] and BIND.[25]

Category: information overlay

Table 3 presents pathway systems grouped by the information overlay requirements they address and the approaches they use. Different systems provide different ways to visually represent biological properties of pathway elements. Biological properties of pathway elements are represented in Cytoscape[15] by manipulating visual node properties such as shape, size, and color. Systems such as Patika,[39] PathwayAssist,[10] and GenMAPP[9] provide predefined shapes to represent different types of pathway nodes. The Patika visualization is spatially divided into fixed areas to represent different cellular locations, such as nucleus or cytoplasm. Temporal information can be shown through animation, and is often partially revealed with top-to-bottom or left-to-right ordering of primary pathway flows. Since the amount of information to overlay on nodes is large, visualizations can easily become confusing if too many node properties are visually represented.

MapMan[43] enables users to analyze microarray data for genes grouped by their functional relationships. Users can zoom into pathways to focus on areas of interest. GenMAPP (Figure 3), Cytoscape (Figure 4), and PathwayAssist (Figure 6) allow users to overlay data from microarray experiments on pathways. Usually, the color of a node is used to encode its expression value in an experiment, using a standard color ramp from green

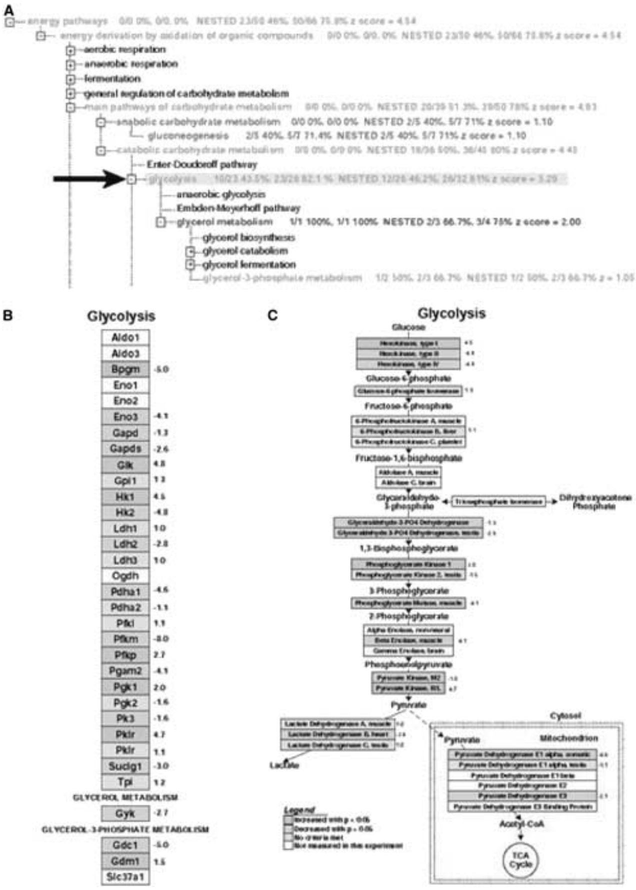

Figure 3 GenMapp[9] A visualization of glycolysis pathway in GenMapp linked to MAPPFinder.[44] MAPPfinder, along with GenMapp, lets users perform statistical analysis on pathways to identify the most changed for a treatment. Results are displayed using the GO hierarchy as shown in (A). Users can click a pathway of interest in the hierarchy (A) for more detailed information. Pathway nodes are listed in (B). The relationships between nodes are shown in (C). The nodes are color coded based on their expression in a microarray treatment (B, C).

(down-expressed) to yellow (no change) to red (up-expressed). Most tools limit users to overlay microarray data for one experiment condition at a time. Then, users can animate the colors to infer changes across conditions. GScope (Figure 5)[41] allows users to overlay expression data for several experiment conditions at once, by embedding small charts onto each node within the pathway visualization. GeneSpring[42] uses multiple views to display separate data visualizations (such as parallel-coordinate plots or heatmaps) of multiple experiment conditions, which are interactively linked to the pathway visualization. Users can then relate the information by interactively selecting nodes in the pathway to highlight the corresponding nodes' data in the data visualizations, and vice versa.

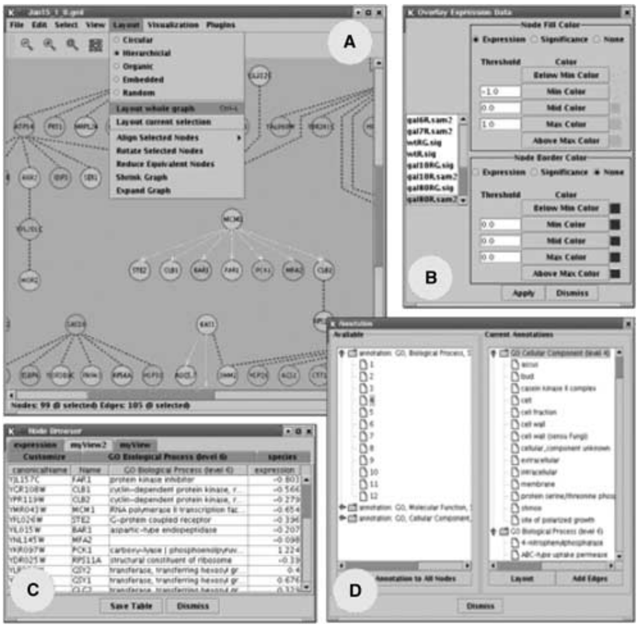

Figure 4 Cytoscape.[15] The color of nodes corresponds to expression data for a microarray experiment as shown in (A). Users are provided with various menus to manipulate node and edge properties (B). It is also possible to overlay annotation (C) and gene ontology information (D) on pathway nodes.

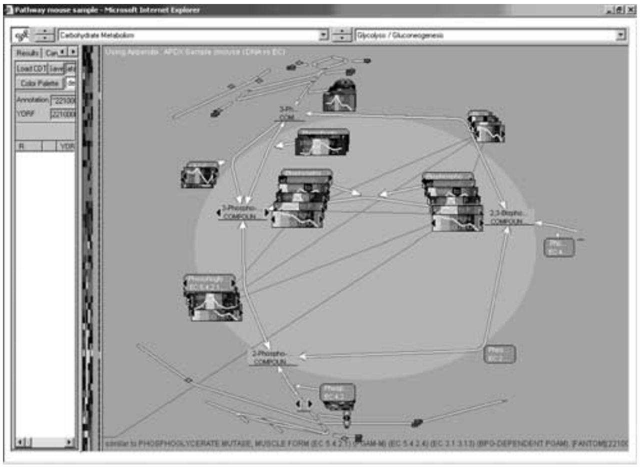

Figure 5 GScope.[41] Fish-eye view is used to reveal details within global context. Multiple treatments of microarray time-series data are overlaid on pathways, using colored heatmaps and line charts.

Category: pathway analysis

Table 4 groups systems by the analysis requirements they address and approaches used. As shown in Figure 2, KEGG[8] provides an overview representing all the interconnections between the metabolic pathways. GScope[41] uses fish-eye techniques to provide an overview for pathways, with a magnified focus region for details.

Table 4 Groups systems by the pathway analysis requirements addressed and approaches used

Requirements	*Approaches*	*Systems*
R10: Overview	Functional groups	KEGG,[8] MapMan[43]
	Zooming	Cytoscape[15]
	Fish-eye views	GScope[41]
R11: Inter-connectivity	Up-down cascades	GScope
	Query pathways	PathwayAssist, Patika
R12: Multi-scale	Chromosome location + pathways	GeneSpring[42]
R13: Notebook	Attach notes to nodes and edges	GenMAPP, Cytoscape
	Build stories about pathway elements	Biological Story Editor[40]

Gscope also allows users to dynamically simulate the effects of a change in a relationship between two nodes on all networks of interest. Patika and PathwayAssist let users query pathway interconnections, such as finding all nodes between two nodes of interest, or finding relationships between pathways of interest. As one form of multi-scale view, GeneSpring[42] links pathways to separate visualizations of gene locations on the chromosome. Biological Story Editor[40] uses a novel metaphor of story telling to organize and share research information and arguments about a pathway among collaborators.

Heuristic evaluation

Based on the systems survey (previous section), we selected six systems for evaluation against the requirements with users. These were selected based on their availability. Some users had favorable experiences with GenMAPP and PathwayAssist and requested their inclusion in our analysis. The systems were evaluated with six life scientists divided into two groups. Although most users were not familiar with all systems, their reviews are important as end-user perception, and valuable to visualization designers. The systems are listed in the order in which they were evaluated.

User reviews

GenMAPP: GenMAPP (Figure 3) provides drafting tools for biologists to create pathways. Although the scientists felt that the tool was easy to use, they said that they would be interested in using GenMAPP only if premade pathways for their interests were available. Creating pathways from scratch would be too time consuming.

GenMAPP does not allow users to link pathways and analyze interconnectivity between them. The life scientists felt that it would be difficult to show concurrent, dependent and mutually exclusive events. Unless arrows representing relationships were labeled it was not easy to tell their type (e.g., stimulatory or inhibitory). Ability to overlay information from microarray experiments was considered helpful. GenMAPP allows users to overlay information from one experimental treatment at a time. GenMAPP also recognizes and highlights replicates in a microarray experiment. The scientists were skeptical of the statistical algorithms used by MAPPFinder,[44] but said it can provide a good start to suggest pathways of interest from a long list.

Cytoscape: The life scientists commented it would be very difficult to understand maps created by someone else in Cytoscape (Figure 4). Some commented that the tool represents computer scientists' conceptions of pathways. In the overview mode, it was difficult to see the labels of genes and their properties. Without this information, a pathway is not helpful to them. They felt it would be difficult to include spatial and temporal information in Cytoscape. While information about connectivity of a node to other nodes in a pathway can be analyzed, it is difficult to comprehend overall pathway connectivity. As a result of these fundamental problems, they were not impressed by the zooming capability to overview pathways. Cytoscape is created for analyzing microarray data in pathway context and provides various analytical plug-ins. Our users were mainly focused on the visualization aspects.

GScope: For life scientists not familiar with them, fish-eye views were confusing. The distorted view and the re-orientation of the nodes when moving the fish-eye caused disorientation. Visualizations either showed too much information in the overview, or too few nodes in the case of the 'clipped view' option. It was difficult to see how a single node is related to the overall pathway. GScope (Figure 5) lets users simultaneously overlay gene expression data for multiple experimental treatments on the nodes. However, the pathway nodes are divided to show values for different conditions using heat map visualizations. The division of nodes, combined with fish-eye distortion, made it difficult to see overall changes in the pathway for different conditions. The scientists preferred animating the pathway node colors, showing one experiment condition at a time as done in GenMAPP, over the GScope approach.

There were mixed comments about the 'cascade' functionality that simulates the effect of a node manipulation on the overall network. One group said that this could be helpful when combined with a better means to overview the network. The other group, which was more familiar with pathway simulation tools that use differential equations (e.g., Copasi[45]), was skeptical of this implementation.

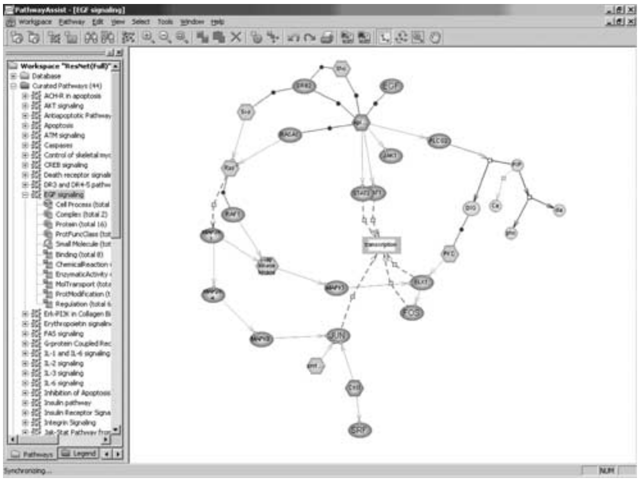

Figure 6 PathwayAssist.[10] EGF signaling pathway visualized in PathwayAssist. The pathway is constructed automatically using NLP algorithms, and needs to be curated by a researcher. The color and shape of the nodes denote different types of biological molecules. Also, the edges indicate if the relationship between two biological molecules is inhibitory or stimulatory. The research papers from which the information is obtained are linked to the edges.

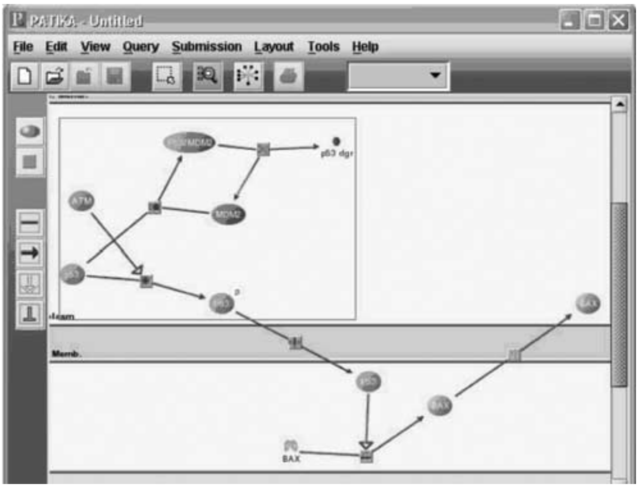

Figure 7 Patika.[39] The pathway diagram is divided into different regions to represent different cellular regions, such as nucleus, cytoplasm, etc. The visual properties of nodes indicate their biological properties.

PathwayAssist: All the scientists were impressed with PathwayAssist's (Figure 6) pathway assembly capabilities. Some wanted to analyze the software to check if the tool really fulfills its claims of creating pathways automatically by searching the literature. They liked the ability to create pathways directly from the ResNet database[8] and from PubMed using NLP algorithms. They were excited to learn that its database has information about more than 140,000 entities, and that more can be added as required. They said that the ability to automatically link scientific references with node interactions was very helpful. The visualization also depicts the interaction type. One of the scientists was concerned about the possibility for misuse and failure to appreciate the shortcomings of NLP. Proper indication of the reliability of NLP-derived information should be indicated.

Patika: Currently, Patika (Figure 7) is a niche product for use in cancer research. A serious limitation is that its database is limited, and has information for just 4,000 different entities. The scientists stated that visualizations provided by Patika were more informative than other tools, because it shows multiple states of a molecule in a pathway and shows the cell compartments where the reactions take place. If information is available from the database, they found it easy to create a pathway in Patika by formulating simple queries to search for connecting entities.

BioCarta: Although we had not originally planned to include it, several scientists commented during the analysis that pathway diagrams provided by BioCarta (Figure 1A) are among the best they have seen for providing biological context to pathways. Different types of pathway entities, the sequence of reactions between them, and the spatial relationships are all shown clearly. The symbols, shapes, and organization of the diagrams are familiar, and similar to those found in textbooks. Simply clicking on a node name reveals more information about a pathway entity. They said it is easy to comprehend the information-richness of biological pathways from these cartoon-like visualizations. They felt that none of the other pathway analysis tools provided as much information in such a helpful and biologically meaningful visual format. It should be noted that BioCarta, unlike the other tools discussed, is simply a repository of pathway diagrams. The diagrams are manually constructed. It does not provide features like the other tools to automate pathway analysis or overlay gene expression data, but can serve as a reference library for users to construct their pathways. Hence, it serves as an excellent educational resource.

Conclusions and research agenda

This work attempts to provide a comprehensive list of requirements for pathway visualizations. We also conducted a software survey and heuristic evaluation to analyze how existing pathway visualization tools address user needs. We found that most tools allow users to perform broader data analysis tasks. A serious shortcoming of these tools at present is that they do not provide adequate domain-specific biological context, and users must perform many tedious operations to search for and extract relevant information. Unless the tools provide users with rapid biologically relevant insight that relates the data to the underlying biological meanings (e.g., to phenotype), most life scientists will be reluctant to use them. The following sections discuss the most important unmet requirements, and a research agenda to address these shortcomings.

Pathway construction and update: Life scientists use many references to construct the pathways they need. Hence, creating pathways requires a significant time investment. Most life scientists pointed out that no matter how valuable the other visualization capabilities, they will not be interested in tools that require them to create large pathways (approximately greater than 100 nodes) from scratch; it is simply too large a time investment, and requires a huge amount of background work to make it meaningful. The tools must be able to construct pathways by retrieving and building on previous relevant pathways. All the life scientists in this study showed particular interest in PathwayAssist, because this tool allows users to automatically search for relevant pathway information and periodically update local databases. The life scientists felt that this capability could save them a significant amount of time and effort. At the same time, users were very wary about a completely automated pathway builder and wanted some degree of human curation.

Information overlay: Much information needs to be overlaid on pathway entities. Most tools let users impart various entity attributes by manipulating simple visual properties of nodes and edges. Different graph layouts can help reveal spatial and temporal relationships. Patika visualizations were appreciated by life scientists due to the representation of different states of molecules, along with their spatial cellular locations. BioCarta diagrams were considered most biologically meaningful, and were preferred by life scientists over ball-and-stick graphs. None of the visualizations capture the actual complexity of network dynamics. For example, STKE[26] provides some animated visualizations to explicitly show sequences of events in a signaling pathway, including movement of biological molecules within the cellular structure. One potential approach for more meaningful visualizations is to represent pathways based on central dogma. Pathway entities can be presented based on their categories such as genes, RNA message, proteins, metabolites, etc.

Defining consistent representations for pathways and entities is needed. Although a large number of pathway visualization systems exist, there is no standardized vocabulary. The green–yellow–red color encoding for gene expression data is one of the few standardized features among these tools (a side effect of microarray imaging technology). This is also true for reference databases and other reference sources. Scientists must constantly learn new representation styles for visualizations created in different systems. An important research area is to define a consistent language for pathways and their visual representations.

Overlay data from high-throughput experiments: The goal of high-throughput data analysis is to infer biological meaning. Life scientists must observe high-throughput data within the context of information-rich pathways. In a separate evaluation study of microarray data visualization tools, it was found that the lack of pathway context severely hampered scientists' ability to derive biologically meaningful insight from the microarray data.[20] Further work is needed to effectively combine pathway and microarray visualization tools.

Designing visualizations that relate pathway network diagrams to quantitative multi-dimensional microarray data, consisting of expression values for potentially multiple treatments and multiple time points, is difficult. In general, there are several possible design alternatives that must be comparatively evaluated to determine effectiveness:

- Nodes-as-glyphs: Most pathway tools will color nodes according to a single microarray treatment (usually the green/red color scale for down/up-regulated).
- Pathway animation: Cycling through several nodes-as-glyphs views over time enables the visualization of a time series. Sliders or other controls can be used to directly navigate the animation loop.
- Small multiples[46] of pathways: Layout several nodes-as-glyphs pathway views in miniature form, likely in a grid of treatments *vs* time series (Figure 8).
- Complex node glyphs, or data visualizations embedded within nodes: While nodes-as-glyphs supports only one value per node, embedding small visualizations of microarray data within each node enables the simultaneous display of values for multiple treatments or time points. For example, GScope embeds heatmaps and line charts. Cytoscape has explored the use of radial bars of different lengths around a node.[47]

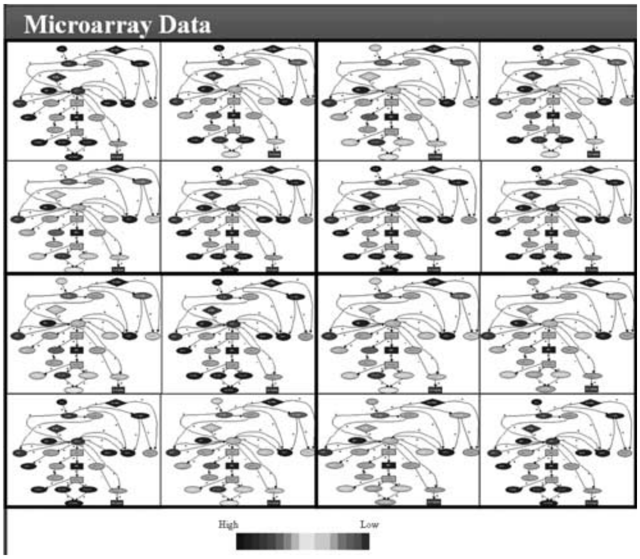

Figure 8 Small multiples visualization of 16 microarray treatments (4 conditions by 4 time points) overlaid on a pathway. Each treatment is overlaid on a separate miniaturized view of the pathway.

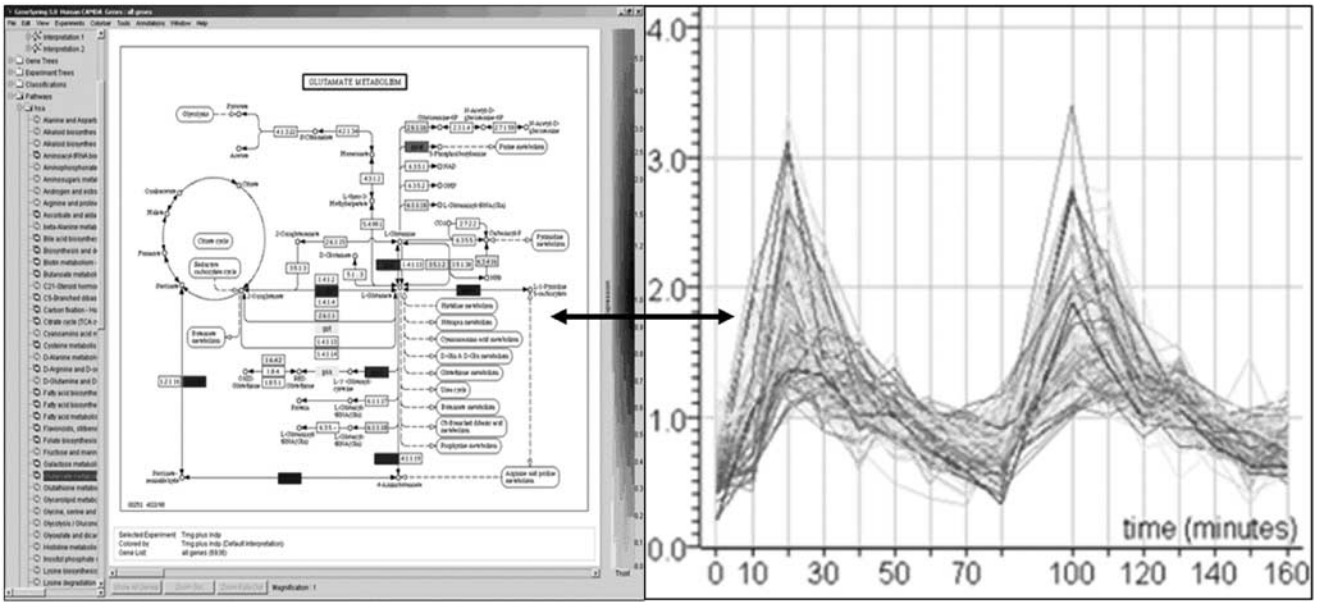

Figure 9 Pathway visualizations in GeneSpring[42] are linked to multi-dimensional visualizations such as timeseries charts. Brushing and linking between the views enables users to select nodes in the pathway to highlight corresponding microarray data in the timeseries, and vice versa.

A disadvantage is that these visualizations can become complex and difficult to read.

- Linked pathway and microarray visualizations: Pathway and microarray visualizations can be separated, enabling advanced microarray data visualization methods such as parallel coordinates and clustering (e.g. GeneSpring, as in Figure 9). The visualizations are interactively linked to enable users to relate nodes to their corresponding microarray data values.

Pathway overview and interconnectivity: Most systems list pathway names (as Windows Explorer lists directory names) to let users select a particular pathway of interest. Life scientists prefer visualizations that provide an overview of pathways displaying interconnections between them, as in Figure 2. Incoming and outgoing visual links could enable users to view how other pathways can potentially affect or be affected by the focus pathway at each node. In a densely populated pathway, it is important to be able to analyze connectivity between components. Simple interactive queries for pathway analysis, such as up-stream and down-stream components from a node at predefined depths or steps, are considered more useful than having to do this manually. This all suggests highly interactive pathway visualizations.[48]

Multi-scale pathways: As pathways become large and complex, methods such as semantic zooming[49] or hierarchical decomposition[50] are needed to aggregate and abstract entire pathways or pathway portions into small units that can be displayed within larger pathway systems. These aggregates should be simple visual representations that reveal enough information of its contents to enable analysis of the high-level effects. For most applications, pathway visualizations must provide sophisticated multi-scaling to view lower level molecular interactions in the context of higher level physiological changes.

Thus, though a large number of pathway tools have been developed, those that allow researchers to effectively explore large complex biological systems of many integrated pathways are still needed. We believe that pursuit of this research agenda to develop tools that address the requirements listed here will lead to significant improvements in life scientists' ability to utilize pathway representations, and facilitate the transition to systems-level science in bioinformatics.

References

1 Duggan D, Bittner B, Chen Y, Meltzer P, Trent J. Expression profiling using cDNA microarrays. *Nature Genetics* 1999; **21**: 11–19.

2 Shi L. DNA Microarray – Genome Chip. [WWW document], http://www.gene-chips.com/GeneChips.html#What (accessed 24 April 2005).

3 Churchill G. Fundamentals of experimental design for cDNA microarrays. *Nature Genetics* 2002; **32**: 490–495.

4 Quackenbush J. Microarray data normalization and transformation. *Nature Genetics* 2002; **32**: 496–501.

5 BIOCARTA. Charting pathways of life [WWW document]. www.biocarta.com (accessed 24 April 2005).

6 Scheideler M, Schlaich N, Fellenberg K, Beissbarth T, Hauser N, Vingron M, Slusarenko A, Hoheisel J. Monitoring the switch from housekeeping to pathogen defense metabolism in *Arabidopsis thaliana* using cDNA arrays. *Journal of Biological Chemistry* 2002; **277**: 10555–10561.

7 Lodish H, Berk A, and Zipursky S, Matsudaira P, Baltimore D, Darnell J. *Molecular Cell Biology*. W.H. Freeman, New York, 2000.

8 Kanehisha Laboratories. KEGG: Kyoto Encyclopedia of Genes and Genomes. [WWW document] http://www.genome.jp/kegg/ (accessed 24 April 2005).

9 Dahlquist K, Salomonis N, Vranizan K, Lawlor S, Conklin B. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genetics* 2002; **31**: 19–20.

10 PathwayAssist™. Ariadne Genomics. [WWW document]. http://www.ariadnegenomics.com/products/pathway.html (accessed 24 April 2005).

11 Stevens R, Goble C, Baker P, Brass A. A classification of tasks in bioinformatics. *Bioinformatics* 2001; **17**: 180–188.

12 Michal G. On representation of metabolic pathways. *BioSystems* 1998; **47**: 1–7.

13 Purchase HC. Metrics for graph drawing aesthetics. *Journal of Visual Languages and Computing* 2002; **13**: 501–516.

14 Ware C, Purchase H, Colpoys L, McGill M. Cognitive measurements of graph aesthetics. *Information Visualization* 2002; **1**: 103–110.

15 Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003; **13**: 2498–2504.

16 Rosson M, Carroll J. *Usability Engineering: Scenario-Based Development of Human Computer Interaction*. Morgan Kauffman, Los Altos, CA, 2001.

17 Nielsen J. Finding usability problems through heuristic evaluation. *In Proceedings of CHI 92*. ACM Press, New York 373–380.

18 The alliance for cellular signaling (AfCS). *Nature* 2002; **420**: 6916.

19 Heath L, Ramakrishnan N. The emerging landscape of bioinformatics software system. *IEEE Computing* 2002; **35**: 41–45.

20 Saraiya P, North C, Duca K. An insight-based methodology for evaluating bioinformatics visualization. *IEEE Transactions on Visualization and Computer Graphics* 2005; **11**: 443–456.

21 Leung Y. Network Pathway analysis software [WWW document]. http://ihome.cuhk.edu.hk/%7Eb400559/arraysoft_pathway.html (accessed 24 April 2005).

22 Bolshakova N. Microarray Software Catalogue [WWW document]. http://www.cs.tcd.ie/Nadia.Bolshakova/softwaretotal.html (accessed 24 April 2005).

23 Bioinformatics Links Directory. [WWW document], http://bioinformatics.ubc.ca/resources/links_directory/ (accessed 24 April 2005).

24 Pathway database. [WWW document], http://www.bioinf.mdc-berlin.de/~schober/AnnotationDTBs.htm (accessed 24 April 2005).

25 Baderr G, Donaldson I, Wolting C, Ouellete B, Pawson T, Hogue C. BIND – the Biomolecular Interaction Network Database. [WWW document], http://bind.ca/ (accessed 25 April2005).

26 STKE. Signal Transduction Knowledge Environment. [WWW document], http://stke.sciencemag.org/ (accessed 24 April 2005).

27 Karp P, Collado-Vides J, Ingraham J, Paulsen I, Saier M. Ecocyc: Encyclopedia of *Escherichia coli* K12 Genes and Metabolism. [WWW document] http://www.ecocyc.org/ (accessed 24 April 2005).

28 Krishnamurthy L, Nadeau J, Ozsoyoglu ZM, Ozsoyoglu G, Schaeffer G, Tasan M, Xu W. Pathways database system: an integrated system for biological pathways. *Bioinformatics* 2003; **19**: 930–937.

29 Toyoda T, Hirosawa K, Konagaya A. KnowledgeEditor: a new tool for interactive modeling and analyzing biological pathways based on microarray data. *Bioinformatics* 2003; **19**: 433–434.

30 Lee M, Hyun S, Park S. UniPath: a knowledge representation system for biological pathways. *Genome Informatics* 2003; **14**: 681–682.

31 Yao D, Qu K, Wang J, Lu Y, Noble N, Sun H, Zhu X, Lin N, Payan D, Li M. PathwayFinder: paving the way towards automatic pathway extraction. *Proceedings of the Second Conference on Asia-Pacific Bioinformatics*. (Dunedin, New Zealand) 2004; **2**: 53–62.

32 PubGene™ [WWW document], http://www.pubgene.com/ (accessed 24 April 2005).

33 Friedman C, Kra P, Yu H, Krauthammer M, Rzhetsky A. GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics* 2001; **17**: S74–82.

34 Vector PathBlazer™. Informax Inc. Solutions. [WWW document] http://register.informaxinc.com/solutions/pathblazer/ (accessed 24 April 2005).

35 OmniViz®. [WWW document], http://www.omniviz.com/applications/pathways.htm (accessed 24 April 2005).

36 Zupan B, Demsar J, Bratko I, Juvan P, Halter J, Kuspa A, Shaulsky G. GenePath: a system for automated construction of genetic networks from mutant data. *Bioinformatics* 2003; **19**: 383–389.

37 Glass Ä, Gierl I. *Proceedings of Second International Conference of Biosystems and Medical Technology, September 7–9*, 2000 (Rostock-Warnemunde, Germany) 2000; 52 pp.

38 Kim P, Lee K, Cho H, Park S, Shin M, Kang E. Comparative analysis workbench for genetic networks. *Genome Informatics* 2003; **14**: 380–381.

39 Demir E, Babur O, Dogrusoz U, Gursoy A, Nisanci G, Cetin-Atalay R, Ozturk M. PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics* 2002; **18**: 996–1003.

40 Kuchinsky A, Graham K, Moh D, Creech M. Biological storytelling: a software tool for biological information organization based upon narrative structure. *Proc ACM Advanced Visual Interfaces Conference* (Trento, Italy) 2002.

41 Toyoda T, Mochizuki Y, Konagaya A. GSCOPE: a clipped fisheye viewer effective for highly complicated biomolecular network graphs. *Bioinformatics* 2003; **19**: 437–438.

42 GeneSpring™. Silicon Genetics. www.silicongenetics.com [WWW document], (accessed 24 April 2005).

43 Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller L, Rhee S, Stitt M. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal* 2004; **37**: 914–939.

44 Doniger S, Salomonis N, Dahlquist K, Varnizan K, Lawlor S, Conklin B. MAPPFinder: using Gene Ontology and GenMAPP to create a global gene expression profile from microarray data. *Genome Biology* 2003; **4**: R7.

45 COPASI. Complex Pathway Simulator. [WWW document], http://www.copasi.org/tiki-index.php (accessed 24 April 2005).

46 Tufte E. *The Visual Display of Quantitative Information* (Graphic Press, Cheshire, CT) 1983.

47 Markiel A. Cytoscape: a network modeling environment with applications to biomolecular interaction networks. *The IEEE Symposium on Information Visualization.* (Seattle, Washington) 2003; Interactive Demos.

48 Herman I, Melancon G, Marshall M. Graph visualization and navigation in information visualization: a survey. *IEEE Transactions on Visualization and Computer Graphics* 2000; **6**: 24–43.

49 Bederson B, Hollan J. Pad++: a zooming graphical interface for exploring alternate interface physics. *In Proceedings of User Interface Software and Technology (UIST 94)* 1994. ACM Press, New York, 17–26.

50 Feiner S. Seeing the forest for the trees: Hierarchical display of hypertext structure. *Proceedings of the ACM Conference on Office Information Systems* (Palo Alto, CA, March) 1988; 205–222.

Appendix A

The questionnaire and the number of responses are given in Table A1. The life scientists were requested to rate each requirement according to how much they agree or disagree with it. The table shows the number of scientists (out of 10) that agree or disagree with each individual requirement. There were no 'strongly disagree' ratings.

Table A1 The questionnaire used to rate each individual requirement.

Pathway questions		*Strongly agree*	*Agree*	*Neutral*	*Disagree*
Category: Pathway assembly					
R1: Construct and update					
1	In my work, the entire pathway(s) is generally not available from a single source	4	6		
2	It would be valuable to have tools that allow pathway import from multiple sources	7	3		
3	Assembling the pathway manually is one of the most time consuming processes in the whole endeavor	4	4	1	
4	Tools that can partially build the pathway from literature or other sources would be of great value to me	6	2	2	
R2: Context					
5	For my work, even if the pathway is fairly well known, I need to be able to modify it if I got it from a published source	2	5	3	
R3: Uncertainty					
6	I want to represent hypothetical connections and/or nodes that have not yet been validated	2	4	4	
R4: Collaboration					
7	I collaborate with others and need my tool to allow them to enter changes from remote sites	1	4	4	1
Category: Information overlay					
R5: Node and edge representation					
8	I am satisfied if just the name of the bio-molecules is displayed on the network diagram		2	4	4
9	I need to have more information displayed on the network diagram than just names and connectivity		8	2	
10	If two molecules interact, a line drawn between them is adequate for my needs	1	1	5	3
11	I want the edge between the interacting components to have information about the nature of the interaction attached	3	6	1	
12	I need the edges to provide more information about the nature of the interaction	4	5	1	
13	I need the line to indicate in some manner how certain it is that the interaction actually exists.	3	5	2	
14	I want the lines to indicate in some manner alternate options/theories in network connectivity	1	6	3	
R6: Source					
15	I need to link the molecule to a database or other sources of additional information	6	3	1	
16	I need to have a lot of annotation and references for my diagram	2	7	1	
R7: Spatial information					
17	Representing the cellular compartment where the components are located is important for my work	3	3	4	
R8: Temporal information					
18	I need to view time series data and want to see how the networks change with time	2	4	3	
19	I need to view how components move between cell compartments over time	1	6	3	
R9: High-throughput data overlay					
20	Adding results from multiple experiments to the network diagram would be of value to me	2	7	1	
21	I need my pathway tool to link to statistical programs for further analysis	3	4	2	1
Category: Pathway analysis					
R10: Overview					
22	I need information about how the pathway I am viewing links to other pathways not displayed	7	3		

Table A1 **(*continued*)**

Pathway questions	*Strongly agree*	*Agree*	*Neutral*	*Disagree*
R11: Interconnectivity				
23 I need a large amount of interactivity with the network diagram	2	3	5	
R13: Notebook				
24 I need to have a history function to record all the changes I've made to the diagram with reasons for them	4	4	2	
25 I perform repetitive steps for pathway analysis session to session	1	5	4	