

SOFTWARE

Open Access

PathRings: a web-based tool for exploration of ortholog and expression data in biological pathways

Yongnan Zhu^{1,2†}, Liang Sun³, Alexander Garbarino¹, Carl Schmidt^{3*†}, Jinglong Fang^{2*} and Jian Chen^{1*†}

Abstract

Background: High-throughput methods are generating biological data on a vast scale. In many instances, genomic, transcriptomic, and proteomic data must be interpreted in the context of signaling and metabolic pathways to yield testable hypotheses. Since humans can interpret visual information rapidly, a means for interactive visual exploration that lets biologists interpret such data in a comprehensive and exploratory manner would be invaluable. However, humans have limited memory capacity. Current visualization tools have limited viewing and manipulation capabilities to address complex data analysis problems, and visual exploratory tools are needed to reduce the high mental workload imposed on biologists.

Results: We present PathRings, a new interactive web-based, scalable biological pathway visualization tool for biologists to explore and interpret biological pathways. PathRings integrates metabolic and signaling pathways from Reactome in a single compound graph visualization, and uses color to highlight genes and pathways affected by input data. Pathways are available for multiple species and analysis of user-defined species or input is also possible. PathRings permits an overview of the impact of gene expression data on all pathways to facilitate visual pattern finding. Detailed pathways information can be opened in new visualizations while maintaining the overview, that form a visual exploration provenance. A dynamic multi-view bubbles interface is designed to support biologists' analytical tasks by letting users construct incremental views that further reflect biologists' analytical process. This approach decomposes complex tasks into simpler ones and automates multi-view management.

Conclusions: PathRings has been designed to accommodate interactive visual analysis of experimental data in the context of pathways defined by Reactome. Our new approach to interface design can effectively support comparative tasks over substantially larger collection than existing tools. The dynamic interaction among multi-view dataset visualization improves the data exploration. PathRings is available free at <http://raven.anr.udel.edu/~sunliang/PathRings> and the source code is hosted on Github: <https://github.com/ivcl/PathRings>.

Background

Biology has entered an era when our ability to collect data has outstripped our ability to turn that data into knowledge. In particular, high-throughput sequencing is providing enormous amounts of information about gene-expression patterns in a large number of species.

In many cases, the experimental objective is to compare two or more distinct biological states, such as disease and control, in order to understand the ramifications of changes in gene expression. Typically these high-throughput data are interpreted in the context of signaling and metabolic pathways, and several resources are available that warehouse and provide visualization tools for pathway analysis (e.g., [1-3]). Visualization is a valuable way to assist in pathway exploration by rendering large amounts of data, thereby aiding investigators' ability to generate testable hypotheses and knowledge.

* Correspondence: schmidt@udel.edu; fjl@hdu.edu.cn; jichen@umbc.edu

†Equal contributors

³Department of Animal & Food Sciences, University of Delaware, Newark, DE, USA

²Department of Computer Science, Hangzhou Dianzi University, Hangzhou, Zhejiang Province, P.R. China

¹Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD, USA

Some existing tools either list overlaps between datasets, such as Lists2Networks [4] and Kerfuffle [5], or overlay information onto the individual pathway graph network, such as BioVenn [6], VisANT [7], CHIBE2 [8], and MIMO [9]. These tools are useful in identifying the interconnections between multiple datasets, but they lack a general overview of all pathways. Such an overview would be valuable in rapidly interpreting the global implications of changes in gene expression patterns and to developing a general view of differences among multiple biological states.

Other tools provide an overview for pathway relationships by overlaying information onto pathways, e.g., iPath gives an overview of regulatory pathways [10]. However, manual intervention is still required to identify interesting pathways. Cellular overview provides organism-specific metabolic map diagrams, but additional information such as expression data is still displayed in a single pathway diagram [11]. Many pathway databases such as Reactome use a tree view to list the hierarchical pathways, but no high-level overview is available to place experimental results within the context of a large biological network.

Most existing tools are restricted to presenting visualization results in a single view at a time, limiting data comparison [12] and forcing humans to store information in their working memory during analysis process. Additionally, displaying data is limited by the physical size of the screen [13]. Some tools, such as Reactome [1]

and Cell Overview [11], use a zoomable user interface that lets users navigate in a fixed view; however, this approach does not support multiple and complex dataset comparison.

Here we present PathRings, a web-based framework for interactive exploration of biological pathway networks (Figure 1). Its features include: (1) **Bubbles views:** applying the ideas of VisBubbles [13], PathRings extends traditional static multiple views to a bubble-based interface. Bubbles do not overlap, but can be grouped and ungrouped, which helps users perform dynamic analysis and switch views for comparative studies. (2) **Virtual working space:** PathRings applies the idea of “virtual screen space extension” [13] by providing a panning navigation bar at the top of the window and extending the current view space to a much larger continuous working space. (3) **Sunburst compound graph visualization:** PathRings displays a hierarchical view of all human pathways from Reactome [1] in a single sunburst visualization, and each arc of a sunburst view is a unit of query and can be dragged out to form a new sunburst compound graph in a new view for sub-pathway analysis. Gene expression, crosstalk and ortholog data can be overlaid. Our exploration workflow supports overview to detail exploration: while sunburst visualization can give an overview of all the pathways, it also displays detailed sub-pathway information in a greater scale. The extendable interface can create new detailed views and support a comparison of multiple experimental data.

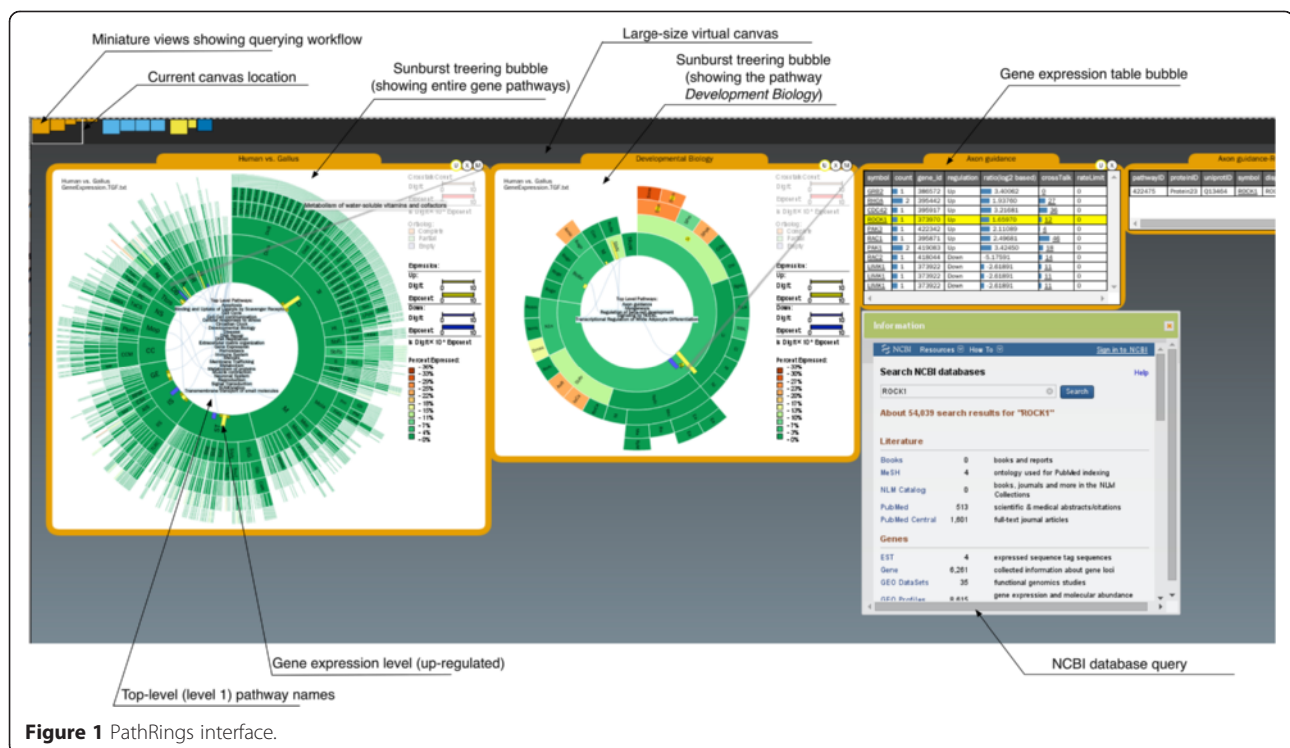


Figure 1 PathRings interface.

PathRings attempts to solve several challenges in pathway visualization that vary depending upon users' exploration goals. At one level, comparing pathways between species allows insight into evolution. Such comparison requires identification of orthologous sequences between a reference and a target genome. PathRings uses Human Reactome as the reference to predict pathways in other species. Interspecies comparisons are then displayed that depict pathway differences between the two species. A second usage scenario is to evaluate gene or protein expression data from a single species in the context pathways. One complexity that current visualizations rarely handle is that individual gene products can affect multiple pathways. Changes in the expression of such cross-talking gene products may affect multiple pathways and have significantly more impact on biology than gene products that function in a single pathway. In addition, few current visualization schemes indicate rate-limiting gene products in the context of pathways. PathRings addresses these issues by identifying both cross-talking and rate-limiting gene products in the context of pathways.

Implementation

PathRings is written in JavaScript Language and HTML5. The main bubble-based interface framework was developed in HTML5 two-dimensional (2D) Canvas API, while the Sunburst visualization is implemented in D3 [14]. It uses Ajax to get data from the server along with MySQL databases to manage the biological pathway data. PHP is used for extracting data from the server and MySQL databases. PathRings is designed to be easily extensible and allows developers to create new features for interactive data analysis.

Results

Pathway generation

We have extended the Reactome human pathways [1] to support cross-species analysis. PathRings supports the analysis of human, mouse, chicken, alligator, and turtle gene expression data. Here the mouse, chicken, alligator, and turtle pathways are predicted based on orthologous relationships between the human and target genomes, thus our pathways are more complete than Reactome. Investigators can use PathRings for other species when an orthology mapping is available between the targeted species and those of human gene products. For chicken Reactome, pathways have been augmented by including orthologous genes identified by RNAseq analysis (Schmidt, unpublished). Hence, we refer to the chicken Reactome as Gallus Reactome Plus.

PathRings: an overview

PathRings' sunburst visualization can depict the impact of the expression data on all Reactome pathways. The

user can select interesting pathways for further analysis by clicking a certain arc of the sunburst to create a new sunburst bubble visualization to examine sub-pathways or to create a gene table bubble visualization listing all the affected genes (Figure 1). The user can select a symbol name in the table to obtain gene information from NCBI [15].

Comparison of multiple pathways can be made by grouping concurrent views and editing one view will affect the other view. No analysis will need to be deleted and moving the current canvas location on the panning bar to another view (or analytical process), while keeping the previous exploratory analysis in context (Figure 1).

PathRings supports pathway exploration for the four types of relationships between pathway members: hierarchical relationships, cross-talking relationships, orthologous relationships, and gene expression relationships.

Hierarchical relationships between pathways

Reactome provides a hierarchical depiction of pathways by dividing large pathways into multiple sub-pathways in a tree structure, depicted in a Sunburst view (Figure 2). Sunburst represents the hierarchical structure in a radial layout to confine information in a space, useful for visualizing mid-sized trees [16] while keeping an intuitive hierarchical structure [17]. In PathRings, every arc of the sunburst shows a parent-children pathway relationship, and overlaying experimental data on the pathways facilitates concurrent exploration.

Here the visualized gene expression level is represented using the order of magnitude markers approach (OOMM) to fit the large-dynamic range expression data [18]. An expression level is represented using the scientific notation of $A \times 10^B$, where A is the digit (a real) and B is the exponent (an integer). The integer exponent B uses a wider bar and the real digit A is shown using a narrow bar. Both A and B are on a linear scale from 0 to 10. For example, an expression level of 99 will be rewritten as 9.9×10^1 , and the wider bar showing 1 whilst the narrow bar showing 9.9. In this way, both large and small expression values can be precisely perceived.

Cross-talking relationships between pathways

In biological pathways, biological crosstalk refers to instances in which one or more components of one pathway affects another; and such crosstalk is often described using cross-talking gene products, as an interconnection relationship between two or more pathways [19]. PathRings shows a cross-talking relationship between pathways by edge links (Figure 3). Details about cross-talking gene products are listed in a table when needed showing the experimental data analysis process. Users can still click the cross-talking gene

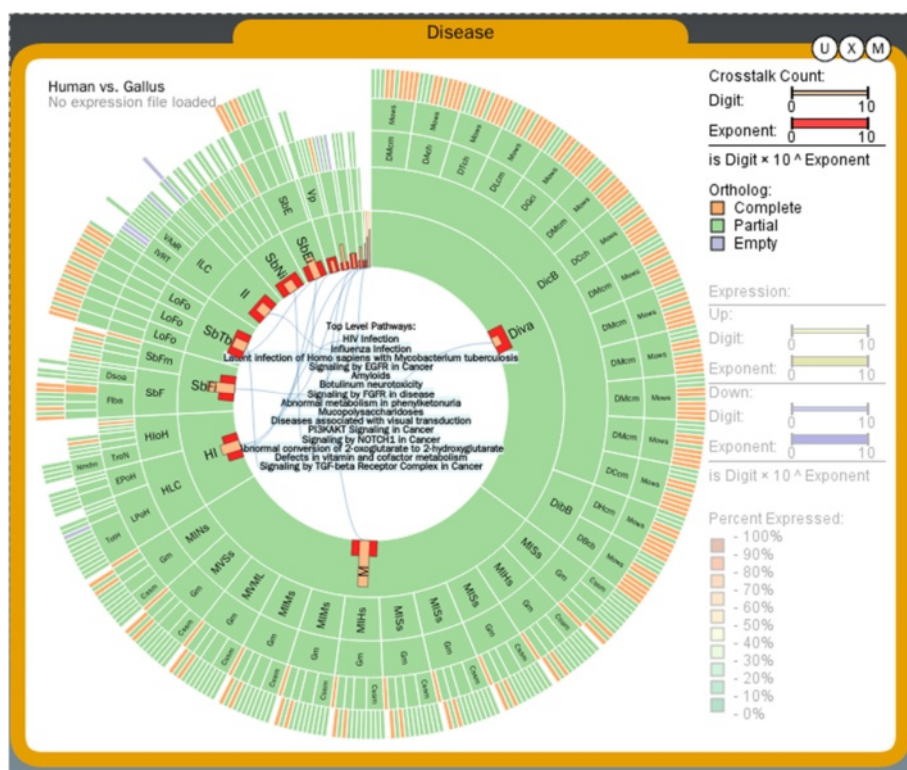


Figure 2 Pathway hierarchical relationships.

from a crosstalk table to highlight the cross-talking pathways marked in yellow dots. The number of cross-talk genes is also represented using the OOMM approach.

Orthologous relationship between Species

Pathway prediction based on orthology between a reference genome (human for Reactome) and a target (such as chicken) will yield three types of orthologous pathway relationships between the reference and target species: complete, partial, and empty defined as the following (Figure 4).

- Complete: All genes in a pathway can be identified in a target species based on orthology to the reference.
- Partial: Not all genes in a pathway are present in the target species.
- Empty: No orthology is present in the target species.

In PathRings, all three relationships of a pathway are encoded by color overlaying the arc of the sunburst (green, yellow, and purple for complete, partial, and empty accordingly). The default view of the sunburst shows the orthologous relationships between human and chicken. Other species of mouse, alligator, and turtle can be loaded for further analysis.

We provide an overview of orthologous genes (shared proteins) of a pathway between two species by embedding bar charts on each arc of sunburst. The height of the bar corresponds to the number of orthologous gene, so that the user can easily find the interesting pathways. Orthologous genes can be queried and listed in a table that includes gene symbol, the number of gene products in the pathway, the number of cross-talking pathways influenced by this gene and the gene product's rate-limiting status all shown in bar charts in different columns (Figure 4). The user can reorder the table by clicking the header of each column to search for data of interests, get detailed gene product information, and see the cross-talking relationship. Views are linked such that clicking on the cross-talking gene (highlighted in yellow in the table) will highlight the cross-talking pathways that contains these genes in the sunburst view in large yellow dots. Finally, the user can also open two or more sunburst bubbles and load their ortholog data for comparative analysis.

Visualization of gene expression information

PathRings supports the display of gene expression data from experiments or in controlled conditions (Figure 5). Gene expression levels, calculated from existing orthology mappings [20] color pathways in the sunburst. Users can analyze their gene expression data from

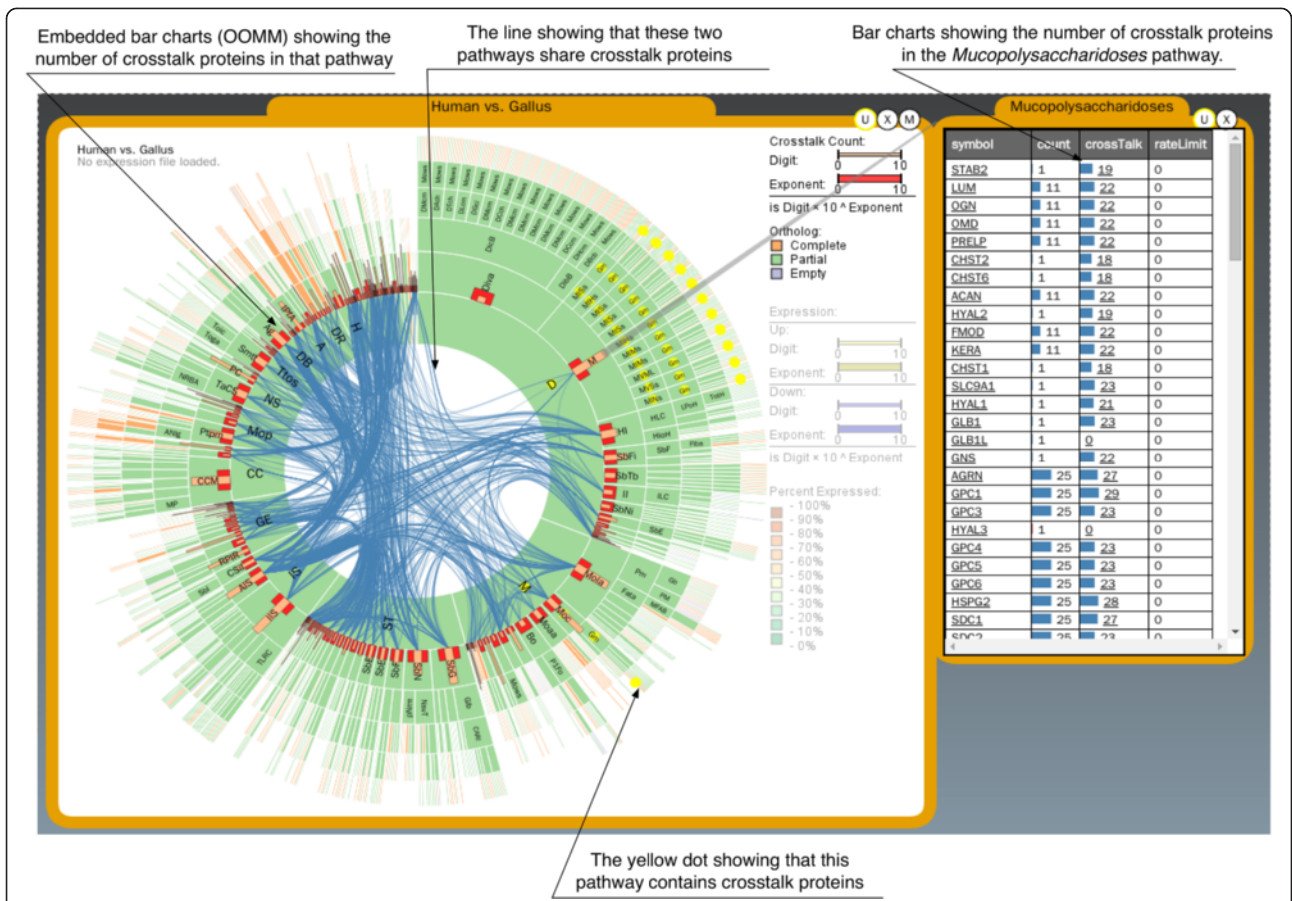


Figure 3 Cross-talking relationships between pathways. This example shows cross-talking at level two in the pathway hierarchy. The OOMMs represent the number of shared proteins of two species (here human and gallus).

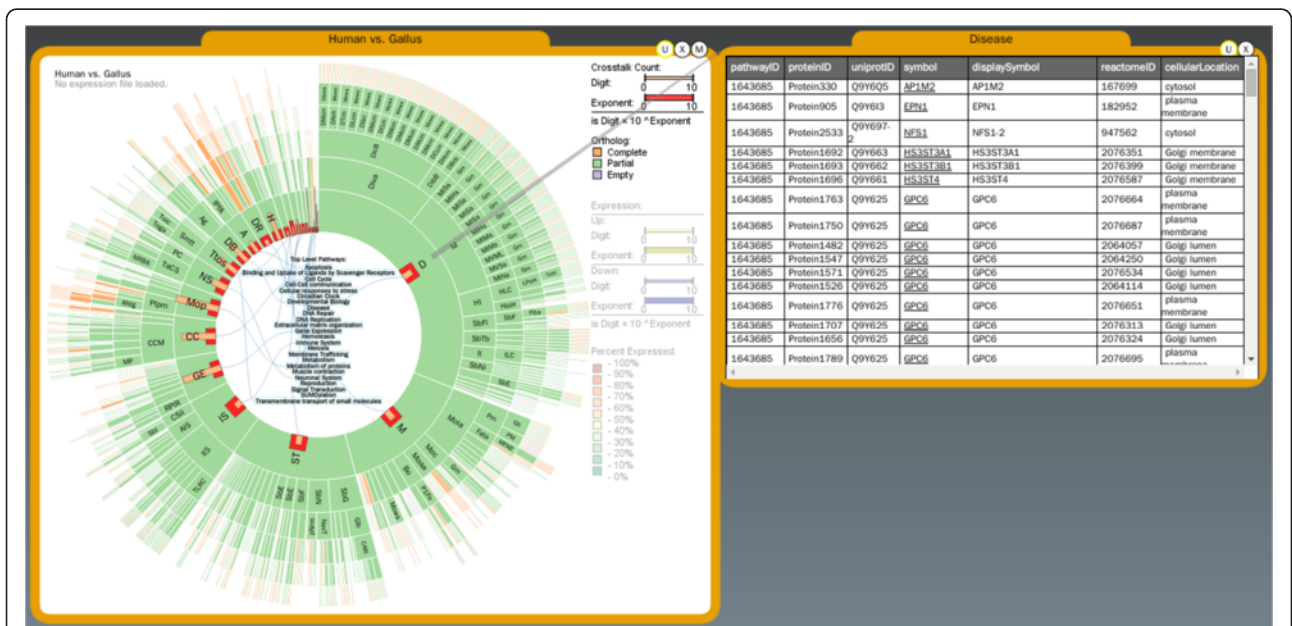


Figure 4 Orthologous relationships between species. Ortholog is complete or partial or empty. Users can query the information of the ortholog genes by clicking the OOMM bars and a table bubble will display the ortholog table (here human and gallus).

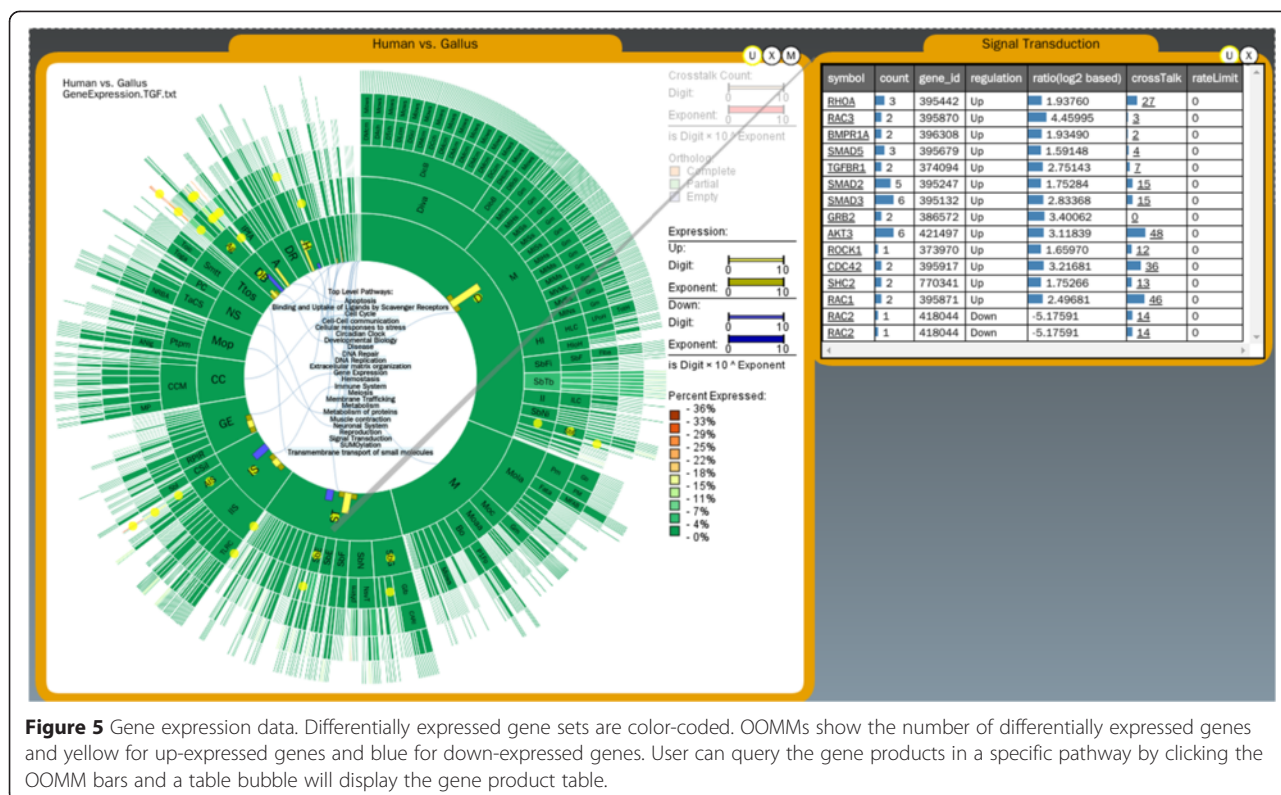


Figure 5 Gene expression data. Differentially expressed gene sets are color-coded. OOMMs show the number of differentially expressed genes and yellow for up-expressed genes and blue for down-expressed genes. User can query the gene products in a specific pathway by clicking the OOMM bars and a table bubble will display the gene product table.

different biological states according to the Entrez gene identification number, the gene symbol, and the ratio of the expression level. The user may input a cutoff threshold value for analysis or just use the default value. Color is used to encode the ratio between differentially expressed gene sets (sum of up-expressed genes and down-expressed genes) and the total number of orthologous proteins of each pathway, providing an overview of the gene expression relationships in all pathways. OOMMs are embedded on the sunburst to show the number of differentially expressed gene sets. To identify the genes affecting a pathway, the user can click the pathway to retrieve a table that includes the Entrez gene identification number, gene symbol, number of gene products in the pathway, log-2-based ratio of each symbol, regulation status (up and down), number of cross-talking pathways influenced by this gene, and the gene product's rate-limiting status.

Discussion

Species consideration

Our current implementation only supports a few species. Our long-term vision is to integrate this exploratory analysis to iPlant infrastructure [21] thereby making several thousand species accessible to the community.

Exploration vs. automatic statistical approach

Automated methods play a central role in biological pathway analysis. For example, we could potentially use Gene Set Enrichment Analysis [22] for interpreting gene expression data by assigning a score to a pathway through a statistical analysis to measure similarities in pathways. However, many other features might be difficult to identify through automated methods alone. Visualization can help exploration especially when performing unstructured browsing and locating tasks or pre-filtering a set of entities. Broadly speaking, no visualization tools exist for comparing large sampling data that scale beyond small stretches of several pathways. Even current tools, such as Reactome could be scaled to several pathways, we find that the design often did not allow for large scale comparisons and focused explorations thus is lack of visual scalability.

Conclusions

PathRings lets biologists explore pathway datasets in a dynamic fashion that pathway hierarchies, ortholog, crosstalk, and NCBI are integrated to answer biological questions. It provides an overview of all the pathways for analyzing experimental data and supports uploading experimental data. Novel visual interface design supports rapid visual retrieval and comparative exploration for efficient data inspection.

Availability and requirements

- **Project Name:** ABI Development: PathBubbles for Dynamic Visualization and Integration of Biological Information
- **Project home page:** <https://sites.google.com/a/umbc.edu/pathbubbles/>
- **Operating system(s):** Platform independent
- **Programming language:** HTML, Javascript, PHP
- **Other requirements:** Web browser
- **License:** BSD license
- **Any restrictions to use by non-academics:** no restriction

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YZ and AG programmed PathRings, JC, CS, YZ, AG, LS, and JF designed PathRings. YZ, JC, and CS drafted the manuscript. All authors read and approved the final manuscript. The work is conducted when Yongnan Zhu is visiting the Interactive Visual Computing Lab at the University of Maryland, Baltimore County, MD, USA.

Acknowledgements

This work is supported in part by National Science Foundation under grant numbers NSF DBI-1260795, DBI-1147029, and IIS-1302755 and by Agriculture and Food Research Initiative Competitive Grant No. 2011-67003-30228 from the USDA National Institute of Food and Agriculture. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Jian Chen, Carl Schmidt, and Jinglong Fang are the corresponding authors.

Received: 16 December 2014 Accepted: 22 April 2015

Published online: 19 May 2015

References

- Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 2011;39:D691–7.
- Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
- Lachmann A, Ma'ayan A. Lists2Networks: integrated analysis of gene/protein lists. *BMC Bioinformatics.* 2010;11:87.
- Aboukhalil R, Fendler B, Atwal GS. Kerfuffle: a web tool for multi-species gene colocalization analysis. *BMC Bioinformatics.* 2013;14:22.
- Hulsen T, de Vlieg J, Alkema W. BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics.* 2008;9:488.
- Hu Z, Mellor J, Wu J, DeLisi C. VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics.* 2004;5:17.
- Babur Ö, Dogrusoz U, Çakır M, Aksoy BA, Schultz N, Sander C, et al. Integrating biological pathways and genomic profiles with ChiBE 2. *BMC Genomics.* 2014;15:642.
- Di Lena P, Wu G, Martelli PL, Casadio R, Nardini C. MIMO: an efficient tool for molecular interaction maps overlap. *BMC Bioinformatics.* 2013;14:159.
- Yamada T, Letunic I, Okuda S. iPath2.0: interactive pathway explorer. *Nucleic Acids Res.* 2011;39:W412–5. Web Server issue.
- Latendresse M, Karp PD. Web-based metabolic network visualization with a zooming user interface. *BMC Bioinformatics.* 2011;12:176.
- Lechat P, Souche E, Moszer I. SynTView - an interactive multi-view genome browser for next-generation comparative microorganism genomics. *BMC Bioinformatics.* 2013;14:277.
- Li G, Bragdon AC, Pan Z, Zhang M, Swartz SM, Laidlaw DH, Zhang C, Liu H, Chen J. VisBubbles: a workflow-driven framework for scientific data analysis of time-varying biological datasets. In: SIGGRAPH Asia Posters 2011. USA: ACM Press.
- Bostock M, Ogievetsky V, Heer J. D3: Data-Driven Documents. *IEEE Trans Vis Comput Graph.* 2011;17(12):2301–9.
- NCBI website [<http://www.ncbi.nlm.nih.gov/ncbisearch/>]; Access date: April 2015.
- Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 2011;39:W475–8.
- Stasko J, Zhang E. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: IEEE Symposium on Information Visualization. USA: IEEE; 2000; 57–65.
- Borgo R, Dearden J, Jones MW. Order of magnitude markers: an empirical study on large magnitude number detection. *IEEE Trans Vis Comput Graph.* 2014;20(12):2261–70.
- Donaldson R, Calder M. Modular modelling of signalling pathways and their cross-talk. *Theor Comput Sci.* 2012;456:30–50.
- Decker K, Anday P, Sun L, Schmidt C. Using expression data to help pathway curation. In: IEEE International Conference on Bioinformatics and Biomedicine Workshops 2012; 535–539.
- Goff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D, et al. The iPlant collaborative: cyberinfrastructure for plant biology. *Front Plant Sci.* 2011;2:34.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005;102(43):15545–50.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

