PathRings User Guide

PathRings is a web-based application for visualization and analysis of biological pathways. Users can visualize and search pathways from Reactome, as well as perform ortholog and gene expression analysis by uploading orthologous and gene expression data. This project is part of the PathBubbles project: http://sites.google.com/a/umbc.edu/pathbbubles/. The recommended web browser to use is Chrome. It is known that some functions are not supported on Firefox browser.

The web-based PathRings is available for data analysis at http://raven.anr.udel.edu/~sunliang/PathBubbles/. All source code is available at Github here: https://github.com/ivcl/PathBubbles.

The PathRings manuscript is under review:

> Yongnan Zhu, Liang Sun, Alexander Garbarino, Carl Schmidt, Jinglong Fang, J. Chen. PathRings: an web-based tool for Exploration of Ortholog and Expression Data in Biological Pathways[J] (submitted to BMC Bioinfomatics)

1. Introduction

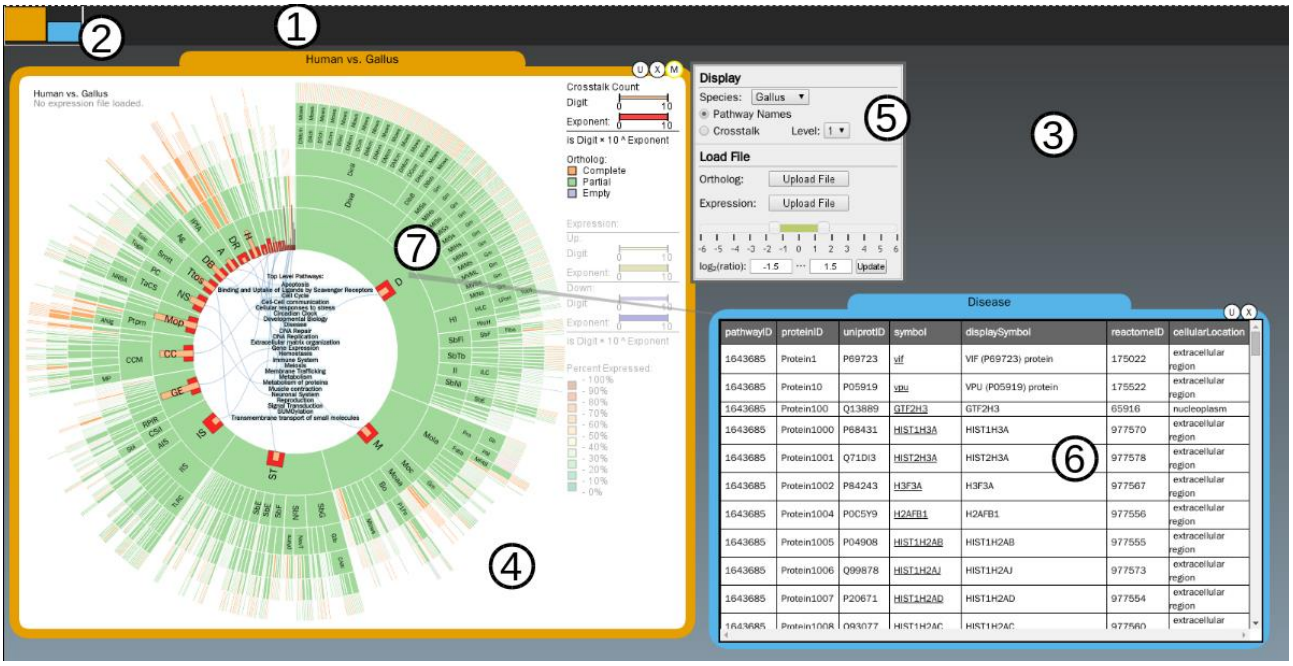


Figure 1: The PathRings interface. 1) The navigation bar, 2) Current view box, 3) main workspace, 4) A sunburst treering pathway visualization bubble, 5) The control menu. 6) Gene table bubble. 7) Edge link between two open bubbles to show the query sequence.

PathRings supports data exploration of four main types of pathway relationships: hierarchical gene pathways, cross-talk, ortholog, and gene expression. Each functional view can be queried and explored in a "bubble."

Figure 1 provides an overview of the PathRings interface. At the top is a navigation bar (1) which provides an overview of the current layout, and along with (2) which can be used to quickly scroll to different sections of the exploration canvas. The main workspace (3) is where the various bubbles are laid out. (4) and (6) are examples of two different bubble types. (5) is a menu for query configuration. (7) is a link from (4) to (6), showing the query sequence and relationship between the two.

The entire interface is a theoretically unlimited canvas. The **navigation bar** (Figure 1 (1)) spans the top of the PathRings interface and can be dragged to create a new query space. Each currently open bubble is displayed on the bar as a box. The current canvas view is indicated with the white box as shown in Figure 1 (2). Dragging that white box will change the current view.

Below the navigation bar is the main workspace (Figure 1 (3)), which is the main display area. Right-clicking on any empty spot in the main workspace will bring up a context menu with the following options:

- ***Open*** Entire Pathway: Creates a new, default pathway bubble which contains all pathways.
- ***Delete*** All: Will delete all bubbles.

- ***Open*** Help: Opens this document.
- ***Toggle*** Hints: Enable or disable the temporary hint bubbles.
- ***Toggle*** Links: Enable or disable the display of links between bubbles (Figure 1 (7)).

The entry point for this program is Figure 1 (4) where the entire human gene pathways are displayed in the sunburst and treering display.

2. General Bubble Mechanics

All bubbles share a few common traits, and are manipulated in a similar manner.

All bubbles can be moved, grouped or ungrouped, and deleted to facilitate data exploration.

While the mouse is over an outside edge of a bubble, the cursor will change. Left mouse clicking and dragging will allow the user to ***resize*** the bubble. Clicking and dragging anywhere else on the bubble's border, including the title, will allow the user to **move** the bubble.

Each bubble also has several buttons in the upper-right corner. All will have the 'X' button, for deleting the bubble, and the 'U' button, for ungrouping it. The 'M' button will ***open a menu*** for further data queries.

Bubbles may be grouped and moved together. **Grouping** is done by dragging and dropping one bubble with the cursor placed over another bubble, or done implicitly when a bubble is created from a parent bubble. Bubbles may then be ***ungrouped*** by clicking and grabbing the bubble, starting from the 'U' button in the top right corner.

3. PathRing Query Interfaces

3.1 The Sunburst Treering Pathway Bubble (Figure 1 (4))

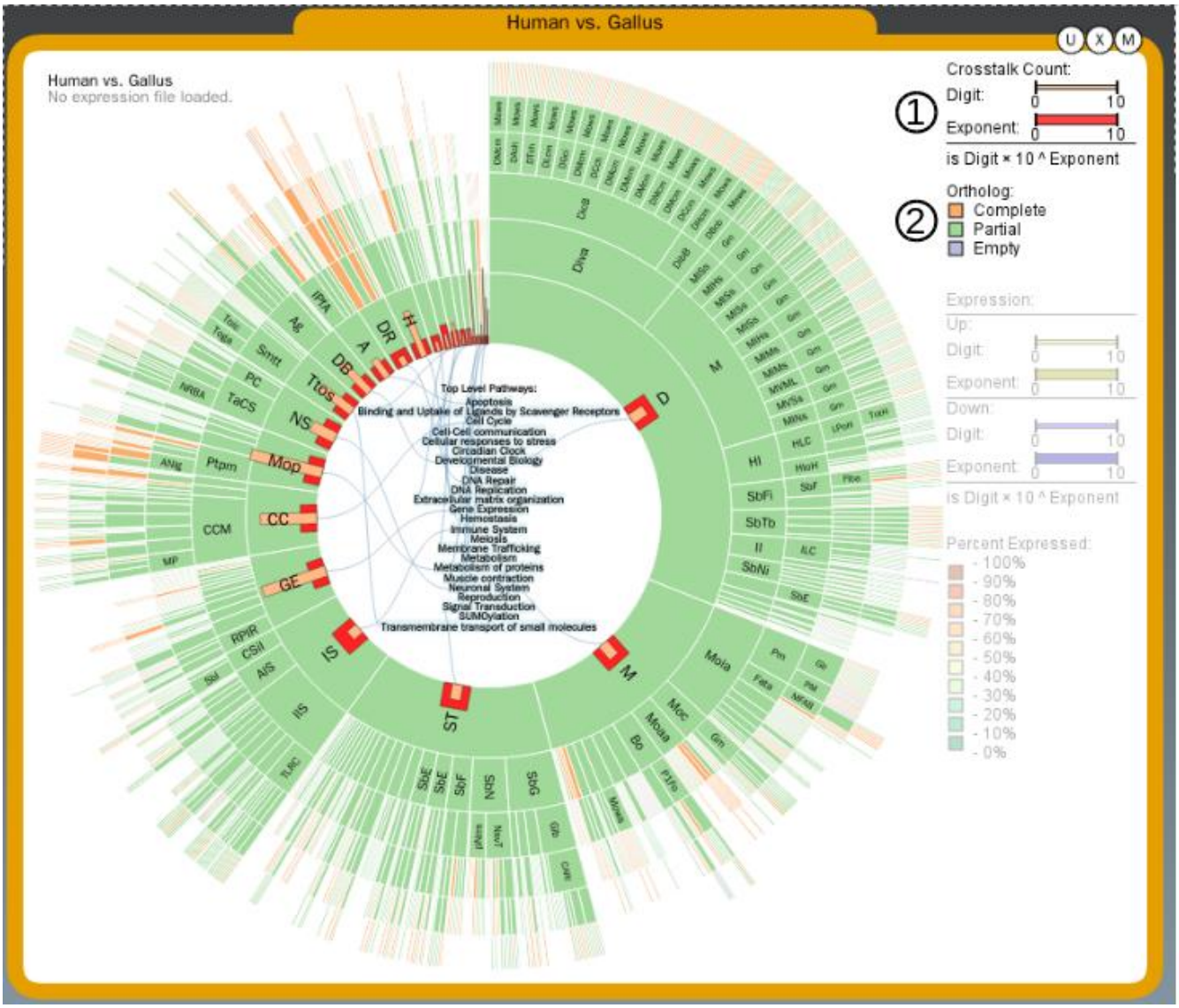


Figure 2: A Pathway Bubble showing crosstalk counts.

The Pathway Bubble visualizes up to the entire set of Reactome pathways. In the center of the bubble is a list of the top-level (level 1) pathway names being shown. There is a line drawn from each of these pathway names to its actual representation in the graph. We list these names because the font cannot be fit into the small ring circle. Hovering over the pathway name with the mouse will highlight this linking line, so that the actual pathway may easily be distinguished.

The majority of the bubble displays the pathway hierarchy. The inner-most ring (level 1) displays the top level pathways. Each successive ring (levels 2+) display the sub-pathways of the previous ring's pathway in a hierarchical fashion. Each pathways is colored according to how closely the selected species or uploaded ortholog data matches that of humans. (See Figure 3 (2)). It distinguishes between *Complete*, *Partial*, and *Empty* pathways. A Complete pathway has all of its genes also found in humans,

and an Empty one has no genes shared with humans.

The upper-left corner displays the currently selected species, as well as any loaded ortholog or expression file names.

On the right side of the bubble shows all the legend information. The parts of the legend not currently in use will be faded from view.

Query pathway and its sub-pathways: The actual diagram itself can be clicked on to bring up a new bubble. Left clicking on a pathway section will bring that pathway section and all its sub-pathways up in a new pathway bubble.

List all genes in a pathway: Right clicking on a pathway section in the ring circle will bring up a new table bubble **listing all of the components (e.g., genes)** of that pathway.

On level 1 of the diagram, **split bars** are used to indicate how many crosstalking genes there are in each pathway. These split bars are called order of magnitude markers (or OOMM). The thicker bar encodes the order of magnitude of the value, and the thinner one encodes the leading digit. (Figure 3 (1) is the portion of the legend for these split bars.) Splitting the numbers up in this way allows for finer comparison between different bars. Right clicking on a split-bar will bring up a new table bubble **listing all genes which contribute to that bar**, e.g. all crosstalking genes in that pathway.

3.2 The Control Menu (Figure 1 (5))

The control menu is opened with the circular 'M' button in the top right corner. It is used to load data files and otherwise manipulate the pathway bubble:

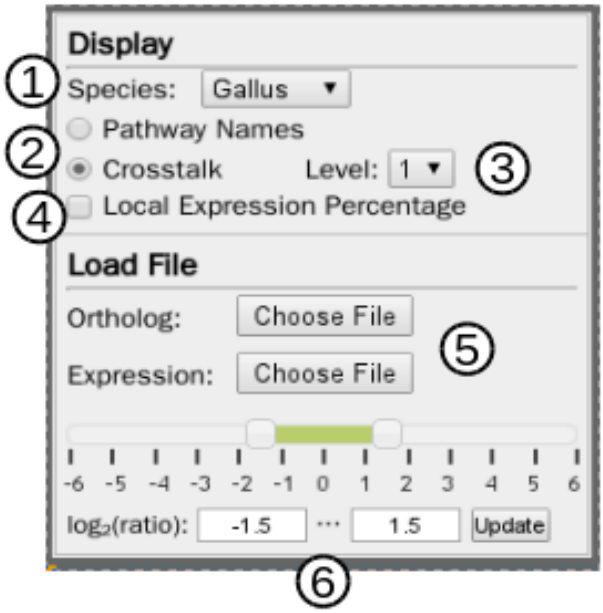


Figure 4: The Pathway Bubble's Menu.

1. This selects a species to *compare* vs. *humans*.
2. This will change the center of the sunburst treering so that instead of pathway names, it shows crosstalk links between pathways. (See figure 6). In addition, hovering over a split bar with the mouse will highlight any corresponding crosstalk links.
3. This selects at which ring level the split bars are displayed, as well as what level the crosstalk links are displayed. Level 1 is the innermost and highest ring displayed. Higher levels move towards the edge.
4. This checkbox is only visible once expression data has been loaded. While checked, this will shrink the "Percent Expressed" scale to have a smaller maximum (See Figure 5), based on the highest expressed pathway in that bubble. This impedes comparison between separate bubbles, as they will have different scales, but makes the existing bubble easier to examine on its own.

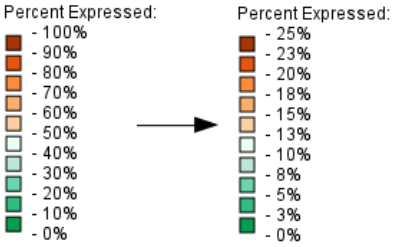


Figure 5: Local Expression Percentage

5. This lets you choose an orthology or expression file to load. See section 5 for the file formats. Section 3.3 explains the expression mode in greater detail.
6. The bottom portion lets you select exactly what constitutes up or down expressed. Using the $\log_2$ of the ratio, anything below the first value is considered down expressed, and anything above the second is considered up expressed. The values may be changed either with the sliders or by typing them in directly. A new expression file must be loaded, or the Update button pushed, before these changes will take effect.

In addition, any changes to species, ortholog, or expression data or display will be propagated to any linked bubbles as well.

3.3 Expression Display

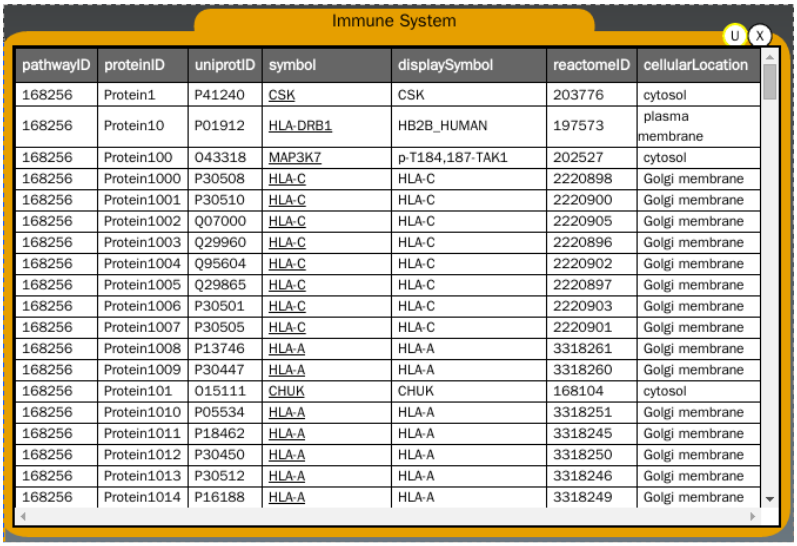
Figure 6: A Pathway Bubble showing gene expression.

Once an expression file is uploaded through the control menu (Figure 4 (5)), the pathway bubble will switch to its expression display form. This differs in two main ways:

Each pathway now has up to two sets of **split bars**. Instead of showing the number of crosstalking genes, they display the number of **up** and **down** expressed genes in the pathway. (See Figure 6 (1)). Right clicking on the bars still brings up a new table bubble **listing all genes which contribute to those bars**. This will list all up or down expressed genes in that pathway.

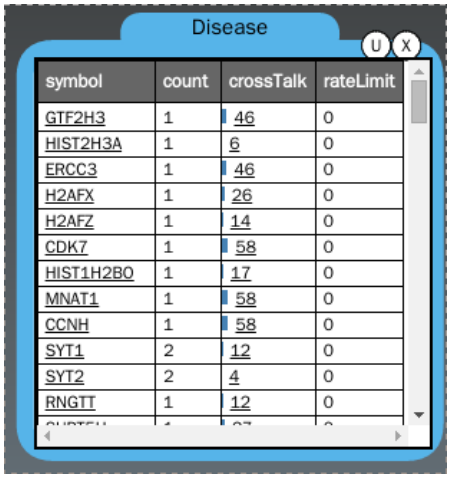
Each pathway is now color coded by the *number of expressed genes* instead of its similarity to humans. (See Figure 6 (2))

4. Table Bubble (Figure 1(6))



pathwayID	proteinID	uniprotID	symbol	displaySymbol	reactomeID	cellularLocation
168256	Protein1	P41240	CSK	CSK	203776	cytosol
168256	Protein10	P01912	HLA-DRB1	HB2B_HUMAN	197573	plasma membrane
168256	Protein100	043318	MAP3K7	p-T184,187-TAK1	202527	cytosol
168256	Protein1000	P30508	HLA-C	HLA-C	2220898	Golgi membrane
168256	Protein1001	P30510	HLA-C	HLA-C	2220900	Golgi membrane
168256	Protein1002	Q07000	HLA-C	HLA-C	2220905	Golgi membrane
168256	Protein1003	Q29960	HLA-C	HLA-C	2220896	Golgi membrane
168256	Protein1004	Q95604	HLA-C	HLA-C	2220902	Golgi membrane
168256	Protein1005	Q29865	HLA-C	HLA-C	2220897	Golgi membrane
168256	Protein1006	P30501	HLA-C	HLA-C	2220903	Golgi membrane
168256	Protein1007	P30505	HLA-C	HLA-C	2220901	Golgi membrane
168256	Protein1008	P13746	HLA-A	HLA-A	3318261	Golgi membrane
168256	Protein1009	P30447	HLA-A	HLA-A	3318260	Golgi membrane
168256	Protein101	O15111	CHUK	CHUK	168104	cytosol
168256	Protein1010	P05534	HLA-A	HLA-A	3318251	Golgi membrane
168256	Protein1011	P18462	HLA-A	HLA-A	3318245	Golgi membrane
168256	Protein1012	P30450	HLA-A	HLA-A	3318250	Golgi membrane
168256	Protein1013	P30512	HLA-A	HLA-A	3318246	Golgi membrane
168256	Protein1014	P16188	HLA-A	HLA-A	3318249	Golgi membrane

Figure 7: The pathway table.

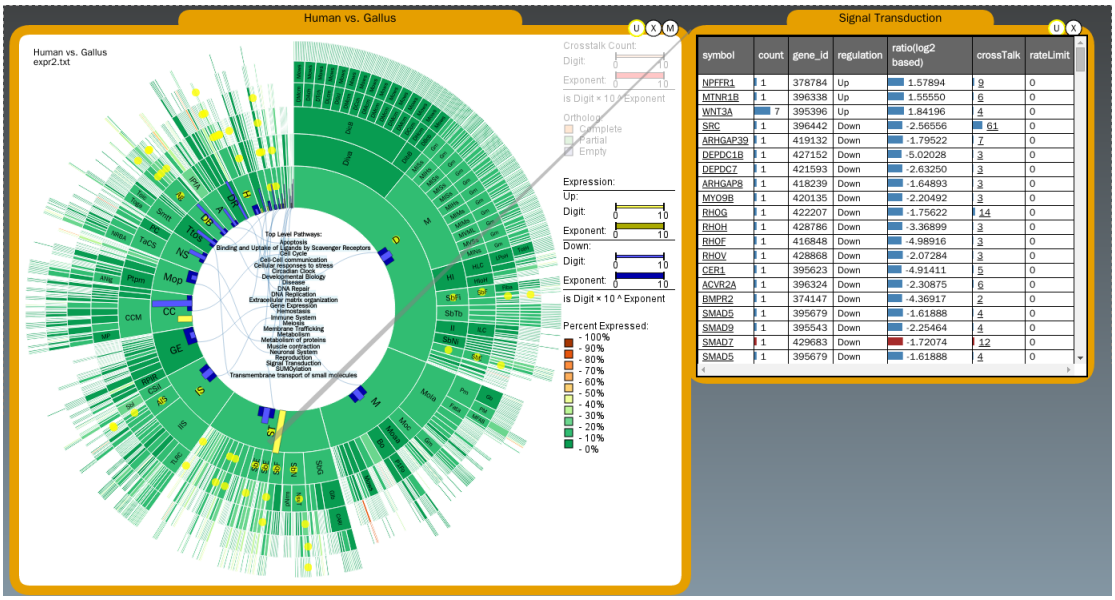


symbol	count	crossTalk	rateLimit
GTF2H3	1	46	0
HIST2H3A	1	6	0
ERCC3	1	46	0
H2AFX	1	26	0
H2AFZ	1	14	0
CDK7	1	58	0
HIST1H2BO	1	17	0
MNAT1	1	58	0
CCNH	1	58	0
SYT1	2	12	0
SYT2	2	4	0
RNGTT	1	12	0

Figure 8: The crosstalking table.

The table bubble shows information on individual genes in a pathway. There are three types of table bubbles: those that show *all* the genes in a pathway, and those that show genes that *contribute* to split bars, either from crosstalking or expression.

All of the table bubbles have the same common controls. The column headers may be left clicked to **sort** the table by that column. Clicking again will cycle through ascending and descending orders. A symbol column entry may be left clicked to **bring up an NCBI information pane** on them. Right clicking on a symbol element will bring up a new table bubble composed of just that symbol.



symbol	count	gene_id	regulation	ratio(log2 based)	crossTalk	rateLimit
NPFFR1	1	378784	Up	1.57894	9	0
MTNR1B	1	396338	Up	1.55550	6	0
WNT3A	7	395396	Up	1.84196	4	0
SRC	1	396442	Down	-2.56556	61	0
ARHGAP39	1	419132	Down	-1.79522	7	0
DEPDC1B	1	427152	Down	-5.02028	3	0
DEPDC7	1	421593	Down	-2.63250	3	0
ARHGAP8	1	418239	Down	-1.64893	3	0
MYO9B	1	420135	Down	-2.20492	3	0
RHOG	1	422207	Down	-1.75622	14	0
RHOH	1	428786	Down	-3.36899	3	0
RHOF	1	416848	Down	-4.98916	3	0
RHOV	1	428868	Down	-2.07284	3	0
CER1	1	395623	Down	-4.91411	5	0
ACVR2A	1	396324	Down	-2.30875	6	0
BMPR2	1	374147	Down	-4.36917	2	0
SMAD5	1	395679	Down	-1.61888	4	0
SMAD9	1	395543	Down	-2.25464	4	0
SMAD7	1	429683	Down	-1.72074	12	0
SMAD5	1	395679	Down	-1.61888	4	0

Figure 9: The expression table and crosstalk highlighting of pathways.

In addition, the crosstalking and expression table bubbles have a *crosstalk* column. Clicking an entry in

this column will **highlight each pathway** in the original pathway bubble which that gene is a member of. (See Figure 9)

5. Data Format

This section describes the file formats used in the uploading option for the Pathway Bubble.

5.1. Ortholog Gene Format

Your own ortholog gene data may be uploaded. It should be a tab-delimited, two column text file. The first column is the gene symbol, and the second is that gene's orthologous relationship between this species and human. If this gene cannot be found in any human pathway, \N should be used. Otherwise, it should be the NCBI Entrez gene ID.

There is an example ortholog file for use available at: http://raven.anr.udel.edu/~sunliang/PathBubbles/documents/Orthology.example1.txt.

Symbol	DbId
ADA	\N
CDH2	414745
AK13	421497
MED6	426282
NRE3	39528

Table 1: Ortholog Data Format Example

5.2. Expression Data Format

Your own gene expression data may be uploaded as well. It should be a tab-delimited, three column text file. The first column is the Entrez gene identification number, the second is the gene symbol, and the third is the expression ratio.

There is an example expression file for use available at: http://raven.anr.udel.edu/~sunliang/PathBubbles/documents/GeneExpression.TGF.txt

Gene Id	Symbol	Ratio
374096	SMAD6	0.6892992938620315
395132	SMAD3	7.128911138923654
395247	SMAD2	3.370212765957447
395543	SMAD9	Infinity
395679	SMAD5	3.013579576317219

Table 2: Expression Data Format Example