2012

MetViz: an online visualization tool for regulons, genes and gene ontology

Achyuthan Vasanth
Iowa State University

MetViz: an online visualization tool for regulons, genes and gene ontology

by

Achyuthan Vasanth

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Co-majors: Computer Science; Human Computer Interaction

Program of Study Committee:

Eve Wurtele, Co-major Professor

David Fernandez-Baca, Co-major Professor

Vasant Honavar

Iowa State University

Ames, Iowa

2012

TABLE OF CONTENTS

LIST OF FIGURES

versions of the same were obtained by running simulations for Deuteranope, Protanope and Tritanope. The resulting images were evaluated for contrast. The colors seen in the Normal Image portion of the figure are the ones that were chosen to be used for the tool as they were able to produce sufficient amount of contrast in each of the simulations.

The colors that were chosen to be the ones producing sufficient contrast under different simulations are seen applied to the tool.

The figure displays the result of a search for gene HSPD1 in the Homo Sapien species. A dialog box is seen indicating that the gene was present in a regulon with ID 71. The regulon is also seen highlighted in the background.

The figure displays a window that contains ring summary information. It indicates the sorting type, the value range, the number of regulons in the range and the actual regulons contained.

The figure displays the regulon-gene content relationship chart. It can be inferred from the figure that with the increase in regulon ID number there tends to be a decrease in the number of genes that are present in the regulon.

LIST OF TABLES

CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

MetViz is an interactive web-based tool that uses novel visualization techniques to represent regulons, genes and the gene ontology hierarchy. The tool provides easier accessibility by making it available as a website [1] requiring no download. MetViz allows users to search for Gene Ontology terms, Regulons or Genes of interest and obtain information regarding them. It helps in understanding the relationship between Gene Ontology Terms and Regulons, Genes and Regulons & Gene Ontology terms and Genes. It provides vital statistics like the genes present in a particular regulon, their count, evidence data, the pearson correlation matrix between the probes associated with genes, shortest path between regulons to show how closely they are related etcetera. It also helps classify regulons based on different quantitative properties like intra-regulon density, degree of regulons and the number of genes present in them. MetViz also enables integration with other softwares such as MetaOmGraph [2]. It currently works for three species – Arabidopsis Thaliana, Homo-sapiens and Saccharomyces cerevisiae.

1.2 Thesis Organization

The thesis has been organized into a number of chapters.

Chapter 2 contains a manuscript that would be submitted to the BMC Bioinformatics journal. The format of this chapter is based on the specifications given by the journal. Achyuthan Vasanth (AV) and Eve Syrkin Wurtele (ESW) are the authors. All authors worked together on problem identification. AV developed the tool and proposed the visualization techniques. ESW

was responsible for determining new and evaluating existing functionalities of MetViz. All authors read and approved the final manuscript.

. Chapter 3 is a general conclusion for the research work. It also discusses future possibilities.

Appendix A contains the list of modifications made.

CHAPTER 2. METVIZ: AN ONLINE VISUALIZATION TOOL FOR REGULONS, GENES AND GENE ONTOLOGY

Modified from a paper to be submitted to
BMC Bioinformatics

Achyuthan Vasanth[1, 3] and Eve Wurtele[2]

2.1 Abstract

Background

Interpretation of large volumes of data that has information about genes, regulons, gene ontology and probes are important for the identification of the functionality of individual genes and their role with respect to the organism. Many software tools are available today, but they become difficult to interpret when visualizing large volumes of data and representing relationships between them.

Results

MetViz is an interactive web-based tool that uses novel visualization techniques to represent regulons, genes and the gene ontology hierarchy. The tool provides easier accessibility by making it available as a website [1] requiring no download. Instead of displaying all data and the relationship with each other at once, as is the case with many software tools like Cytoscape, MetViz incrementally displays and associates them based on the user's interaction with the tool. One can radially visualize and organize data based on different

---

[1]Department of Computer Science, Iowa State University, Ames, IA, USA

[2]Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA, USA

[3] Virtual Reality Applications Center, Iowa State University, Ames, IA, USA

quantitative properties such as Intra regulon density, gene count in a regulon and degree of regulons. It minimizes navigation between different windows and pages by maximizing screen utilization. It is also aimed at reducing the number of clicks performed to view and obtain information by providing easily accessible buttons to different functionalities, multiple ways to interact with data and perform the same function, and also making use of unique techniques such as a modified version of the icicle graph to display gene ontology information.

Conclusions

MetViz allows users to search for Gene Ontology terms, Regulons or Genes of interest and obtain information regarding them. It helps in understanding the relationship between Gene Ontology Terms and Regulons, Genes and Regulons & Gene Ontology terms and Genes. It provides vital statistics like the genes present in a particular regulon, their count, evidence data, the pearson correlation matrix between the probes associated with genes, shortest path between regulons to show how closely they are related etcetera. MetViz also enables integration with other softwares such as MetaOmGraph [2].

2.2 Keywords

Online, web-based, Visualization tool, genes, regulons, gene ontology, Pearson correlation, analysis

2.3 Background

One of the important goals within the fields of molecular biology and genetics is the identification of the functionality of individual genes and their role with respect to the organism. Research in these fields typically involves working with large volumes of data that has information about genes, regulons and metabolic pathways to name a few. In order to interpret this data and identify interesting patterns in them, many tools are being used.

AmiGO [3] is an open source web application maintained by the GO Consortium that allows users to query, browse and visualize ontologies and related gene product annotation data. The application primarily returns detailed text results in pages when viewing gene symbol data and provides limited visualization and interactivity when viewing the GO term hierarchy.

Cytoscape [4] is a widely used tool to visually represent biological pathways, gene expression profiles etcetera. An online network visualization library called Cytoscape Web was developed to support some of the basic features available in the Cytoscape project. One of the problems encountered in both Cytoscape and Cytoscape web is the prevalence of crisscrossing lines [Figure 2.1] when the data set is huge. Cytoscape also requires a download that might be inconvenient for some users.

GOFFA (Gene Ontology For Functional Analysis) [5] is another offline tool that displays a simple hierarchical structure which allows users to browse through the most significant gene

ontology terms and paths. One main disadvantage of this tool is the significant load time for different actions on a GO term.

agriGO [6], previously named easyGO, is a web-based tool and a database for GO analysis. It specifically provides support to the agricultural community by enabling analysis of 45 agricultural species. Results are visualized as HTML tables, tabulated text files, hierarchical tree graphs, and flash bar graphs.

REViGO [7] is a web server that takes in Gene Ontology terms and visualizes them in the form of scatterplots, interactive graphs and tag clouds. It depends on the user or other softwares like agriGO and GOrilla for its input. The tool as such does not contain quantitative data about genes or regulons. MetViz could possibly interact with this tool, although it requires user intervention in the form of extracting the GO terms from the exported XML file and then giving them as input to this tool.

g:Profiler [8] is a web-server used for analysis of gene lists. It provides a list of tools – g:GOSt, g:Convert, g:Orth , g:Sorter, g:Cocoa to the user. g:GOSt helps in retrieving GO Terms, pathways and also provides a Windows Explorer form of visual representation [Figure 2.2] of the GO graph. It provides limited interactivity.

QuickGO [9], another web-based tool, acts primarily as a browser for Gene Ontology terms. It provides mostly textual information about the GO terms. It also helps in the visualization

of the Gene Ontology hierarchy by providing an ancestor chart that displays the chosen GO terms' ancestors.

Many other visualization and analysis tools can be found at [10]

The visualization tool, MetViz, is aimed at combining the best features of the currently available tools. It avoids crisscrossing interconnections that are prevalent in visualization tools that deal with a large volume of related data. It adequately combines visual content with textual content and provides useful statistics. It acts as a browser for GO Terms, genes and regulons and also enables visual mapping with each other. Additionally, the tool is designed to be available online, to enable easier access without the trouble of having to download anything. The data is made available by assimilation from a number of sources into a central repository called MetNetDB[11].

2.4 Methods

The server side scripting was done in PHP. The client side script was written in JavaScript and it was based on HTML5 (canvas element).The client side script also included some library files that used jQuery. The database tables used were modified versions of the ones present in the MetNetDB database [11], developed to suit the needs of the MetViz tool.

The MetViz tool enables the identification of interesting and valuable genomic information. It provides statistical details about the genes, regulons and Gene Ontology terms that are

being viewed and also visually represents their relationship with each other. A snapshot of the tool is found in [Figure 2.3].

2.4.1 Regulon Visualization

The MetViz makes use of the hierarchical association between genes and regulons in that each gene is contained by a regulon, or, if the gene's functionality is not identified, it is not associated with any regulon (Hence they are visually hidden in the system.). The top pane [Figure 2.4] of the tool helps in the visualization of regulons and their relationship with each other. It consists of a number of circles arranged in concentric rings. The circles represent regulons. Each concentric circle represents a range of values of a particular property of the regulons. The concentric rings can be selected and summary data such as regulons present and range represented can be viewed. The regulons that have a property value that falls within this range is placed on that particular concentric circle. For example, a regulon having an intra-regulon density of 0.342 will be placed in the concentric circle with the range 0.3 to 0.4.

Two regulons are related if there exist probes (representing genes) in each of these regulons that are related to each other with a pearson correlation value great than 0.7. This relationship is represented by lines connecting the regulons. The connections are displayed only when the user clicks on a particular regulon, thus making it "incremental linking". Incremental linking avoids the problem of crisscrossing lines that is predominant in many other visualization tools that deal with large volumes of data.

2.4.2 Regulon Grouping

The properties based on which the regulons are grouped into concentric rings are

1. The intra-regulon density of the regulons
2. The number of genes present in them
3. The degree of each of the regulons.

These properties are viewed separately and hence are referred to as viewing modes. The tool gives the user the ability to change between the different viewing modes in run-time.

2.4.3 Modified Icicle Graph representation of Gene Ontology terms

The bottom pane [Figure 2.5] helps in the visualization of the Gene Ontology hierarchy. It is based on the icicle graph representation. It is based on a fixed height, fixed width (for a given zoom level) style in the sense that the width and the height of the pane does not depend on the number of GO Terms present at any particular level of the hierarchy. This is done to maximize usage of screen space.

The Gene Ontology terms are represented by rectangles. The width of a rectangle corresponds to the number of children that that particular GO Term has relative to the number of children that other GO Terms in the same level of the hierarchy have.
The representation allows a particular GO term to have more than one parent. Some GO Terms are also listed in a transparent color to indicate that they are children to another GO Term at a lesser depth from the root of the graph.

The GO terms are drawn based on a breadth first search methodology. That is, a GO Term that has multiple parents would be represented visually in the lowest depth in the hierarchy in which it appears (lower depths are closer to the root of the graph).

2.4.4 Accessibility of MetViz

The MetViz tool was designed taking into account people who have visually impairments. The different colors used for the tool are web safe. The web safe color palette [Figure 2.6] contains 6 shades of each of the major color components: red, green and blue. Not using web safe colors could lead to color approximation which in turn might lead to a decrease in the intended level of contrast amongst colors used thus affecting people with vision impairments. The chart in [Figure 2.7] presents the way in which different colors are perceived by people with normal vision and people with vision abnormalities.

Contrasting web safe colors were required for the background, unselected regulons and GO terms, the selected regulons and their reflection in the GO terms view panel, the related nodes to the selected regulons in the Regulons view panel, the selected GO terms and their reflection in the Regulons view panel, and, the related nodes to the selected regulons in the GO terms view panel. The ImageJ software was used with the Vischeck plugin to finalize on a set of chosen web safe colors (chosen based on guidelines mentioned in [12] and [13]) that appeared in contrast with each other with the normal eye and also in all three simulations - Protanopia, Deuteranopia and Tritanoptia. The final colors used and their three simulations can be found in [Figure 2.7]. A snapshot of the tool with the applied colors is present in [Figure 2.8].

2.5 Results and Discussion

With the MetViz tool, users interact with the regulons and Gene ontology terms by left clicking, right clicking, Click and drag select, and alt clicking (to select multiple items). Clicking of regulons highlights the associated Gene Ontology term(s) and vice-versa. Selecting a regulon gives the user a number of options that are available on the side pane and also on his right-click menu. Some of the functionalities are: Editing Concentric Rings, Magnifying tools, Gene Search, Regulon Search, Gene Ontology Search, View inter-regulon density, View relationship (between regulons), View Genes, View Common regulons, View Ring Summary. The users can view correlation matrices between the selected probes, load probes from MetaOmGraph, save probes and regulon information to the local hard drive (which can be loaded in to MetaOmGraph), view a chart that represents the relationship between regulons and the genes present in them, view relationship between Gene Ontology terms in terms of the common regulons, etcetera. Figures 2.9-2.13 illustrate some of these different functions.

2.5.1 Performance improvements

An online tool that works with huge volumes of data faces a lot of challenges. The amount of data that is being dealt with is in the order of a few tens of megabytes the largest being Arabidopsis Thaliana (around 100MB). Since it is a GUI based online tool, the response times should be low. It is not possible to interact with the server every single time a user interacts with the system as this would badly increase the response time and hence affect user experience. Since the graphical representation of data requires loading of the entire database

to analyze and draw/connect the different components, asynchronous downloading is also not possible. Hence the data had to be loaded all at once before the tool starts to be operational. Thus effective performance improvements had to be done.

One way to improve performance is caching. Caching refers to storing of information in the local system, so that the time required for querying the server and then transferring data from it is saved. The machine is made to obtain the local data instead. This was thought of as a way to improve performance in MetViz. The following have been implemented in order to achieve better load times.

1. Caching on the server side – The server, as soon as it runs the server side script for the page once, stores a local copy of the page. The next time the script is run, it checks for the existence of the file and runs it if present.
2. Caching on the client side – The browser is responsible for caching on the client side.

To get a better idea about how caching affects the performance of MetViz, the loading time was measured with cached data and without cached data. The tables 2.1- 2.3 and charts [Figure 2.14 – 2.16] detail the effects.

2.5.2 Pilot User Study

A pilot user study was conducted with 2 participants. The objective of the study was to identify problems and inadequacies in the design of the user study itself apart from identifying user-interface flaws. Users who had basic background knowledge in molecular biology were chosen for this study as they more accurately represent the target audience.

The participants worked in a relaxed setting. They were given a sample untimed task to get them acquainted to MetViz. Once they had completed it, they were given an additional set of ten tasks to perform using the tool. The users were timed when they performed these tasks. They were observed carefully to identify user interface issues that might exist in the system. The time to complete each was also noted.

After each task the participants were asked to rate on two metrics about the same - how easy it was to perform and how successful they think they were in accomplishing it. The first metric is a direct score on the usability of the tool. The users were asked to give a rating between 1 and 5 on the Likert scale [14] - 1 being the most difficult and 5 being the easiest to perform. The second metric was collected to ensure that the tool was not giving wrong information to the user and convincing them that it is the correct. This metric was again a value on the Likert Scale – 1 being the least confident and 5 being the most confident. It is important to note that, so long as the user's response to the question appropriately matches his confidence level, the tool is functioning appropriately. The following scenarios are acceptable

1. The user gives the correct answer and is confident about it.
2. The user gives the wrong answer and is not confident about it.

The following scenarios are not acceptable

1. The user gives the correct answer and is not confident about it.
2. The user gives the wrong answer and is confident about it.

The list of tasks are present in Table 2.4.The results of the pilot study are available in Table 2.5.

Feedback was obtained from the participants at the end of the study. They were asked to answer the following questions.

1. What did they learn from the tool
2. What did they like most about the usability of the tool
3. What suggestions do they have to improve the usability of the tool
4. How would they rate the ease of use of the tool
5. They were also asked for any comments in general.

Table 2.6 details the feedback obtained, problems identified and the modifications that were made to MetViz at the end of the pilot study.

2.5.3 User Study

Once the problems identified as a result of the pilot user study were addressed, the study was conducted on 7 individuals. Each participant was a graduate in a major in biology (again, to more accurately reflect the target population) and was made sure they had the relevant knowledge in the problem domain. The setting was very similar to the one in the pilot studies. A video recording of the sessions were made this time to identify common behaviors and thereby flaws in the system. The results obtained are documented in Tables 2.7 – 2.10. The changes made after the pilot testing and the user testing was done can be found in Appendix C.

The following were some of the feedback obtained from user testing regarding the usability and possible uses of the tool.

1. Integration of information to provide a comprehensive picture was interesting.
2. It could be used in evolutionary analysis, functional annotations and cross-species study.

3. Ability to access data by a variety of search terms.
4. Tool was useful in finding information at different levels from genes to classification based on Gene Ontology terms.
5. Clean interface.

2.5.4 User Study Results Analysis and Modifications to MetViz

From Tables 2.7 to 2.10, it is seen that some of the users found Task 5 to be a little difficult and they were also unsure about their solutions although their solutions were correct. This is a part of the primary learning curve that the user would face and hence should not be regarded as a user interface issue.

It is clear that there exists a problem with the functionality associated with Task 6. The issue was addressed by placing a separate tab for the summary data in the side pane.

It is also seen that the functionality associated with task 8 also had a problem. The problem in this case was that the button was hidden from the user's view and the user had to use the scrollbar to click on the functionality that produced the required result. This was dealt with as the entire window was moved to the side pane.

One of the users also found Task 9 to be cumbersome. The labeling of the section on the right pane was changed to better reflect its purpose. Also, help buttons were added for each section.

Task 10 was a cause of concern for one user as he was expecting for a section in the right pane or in the Settings icon to have a separate option for loading and saving user data. This has been dealt with by creating a new section in the side pane.

2.5.5 Future work

The MetViz tool, owing to the generic design of its underlying database schemas and database creation modules, can be extended in many possible ways. Currently, database tables have been developed for Homo Sapiens, Saccharomyces cerevisiae and Arabidopsis Thaliana. This can be extended to other species. Since the tool is based on JavaScript and PHP, one can also think of the possibility of integration with Cytoscape web which also makes use of the JavaScript platform. Although parallel comparisons of genes and regulons from different species and their relationship with the Gene Ontology can be done through multiple browsing windows, a closer integration could be made possible by introducing the ability to view them within the same browser tab.

2.6 Conclusions

Analyzing interactions between regulons and understanding gene pair up or down regulations could be helpful for hypothesis building from the wealth of microarray data. Unrelated biological processes can link up on analysis suggesting novel signaling events we might have missed or not even acknowledged. Although there are many softwares available today [3-10], they tend to be cumbersome to the user when it comes to visualizing and interpreting large volumes of data.

MetViz is a web-based tool that is simple to use and has the ability to display large volumes of data and information about them by an appropriate combination of visual elements and textual content. It acts as a browser for GO Terms, genes and regulons and enables visual mapping with each other. It is a tool that could be used along with tools such as the

MetaOmGraph to help identify interesting relationships and functionalities previously unknown.

2.7 Availability and Requirements

Project name: MetViz

Project home page: metnetdb.org/php/RWT_Viz/startPage2.html

Operating systems: Fully operational on the Google Chrome browser for Windows and MAC. Layout is fully operational on the Safari browser for Windows and MAC. Functional limitations exist. Not functional on Mozilla Firefox or Internet Explorer.

Programming language: JavaScript, PHP, HTML 5.0

Other requirements: N/A

License: Freely available under GNU GPL license.

Restrictions to use by non-academics: None

2.8 List of Abbreviations used

MAC – Macintosh operating system, PHP – PHP: Hypertext Preprocessor, GO – Gene Ontology, HTML – Hypertext Markup Language

2.9 Authors' contribution

All authors worked together on problem identification. AV developed the tool and proposed the visualization techniques. ESW was responsible for determining new and evaluating existing functionalities of MetViz. All authors read and approved the final manuscript.

2.10 Authors' information

AV did his Bachelor's degree in the field of Computer Science and Engineering at Madras Institute of Technology, India. He is doing his Masters' degree majoring in Computer Science and co-majoring in Human Computer Interaction. He works as the research assistant for ESW.

ESW is a Professor of Genetics, Development, & Cell Biology. She is also a VRAC Faculty Affiliate and HCI Graduate Faculty.

2.11 Acknowledgements

This work is supported by a non-federal grant Incentive GDCB Wurtele 490-61-01 490-61-01. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors. We would like to thank Dr.David Fernandez-Baca and Dr.Vasant Honavar for their general support throughout the project and also helping with and approving the final manuscript. We would like to thank Dr.Yaping Feng for valuable biological input during the development of this tool. It was the authors' decision to submit the manuscript for publication.

2.12 References

[1] MetViz website [http://metnetdb.org/php/RWT_Viz/startPage2.html]

[2] MetaOmGraph website [http://metnet.vrac.iastate.edu/MetNet_MetaOmGraph.htm]

[3] AmiGO: online access to ontology and annotation data.

Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S; AmiGO Hub; Web Presence Working Group.

Bioinformatics. 2009 Jan 15;25(2):288-9. Epub 2008 Nov 25.

PMID:19033274[PubMed - indexed for MEDLINE]

[4] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks.

Genome research 2003, 13(11):2498-2504.

[5] GOFFA: Gene Ontology For Functional Analysis – A FDA Gene Ontology Tool for Analysis of Genomic and Proteomic Data

Hongmei Sun, Hong Fang, Tao Chen, Roger Perkins and Weida Tong

Journal Name: BMC Bioinformatics,Cover Date: 2006-09-01,Publisher: BioMed Central,Issn: 1471-2105

[6] Zhou Du, Xin Zhou, Yi Ling, Zhenhai Zhang, and Zhen Su agriGO: a GO analysis toolkit for the agricultural community Nucleic Acids Research Advance Access published on July 1, 2010, DOI 10.1093/nar/gkq310. Nucl. Acids Res. 38: W64-W70.

[7] Supek F et al. "REVIGO summarizes and visualizes long lists of Gene Ontology terms" PLoS ONE 2011. doi:10.1371/journal.pone.0021800

[8] J. Reimand, T. Arak, J. Vilo: g:Profiler -- a web server for functional interpretation of gene lists (2011 update) Nucleic Acids Research 2011; doi: 10.1093/nar/gkr378

[9] QuickGO Website [http://www.ebi.ac.uk/QuickGO/]

[10] List of Visualization tools [http://geneontology.org/GO.tools_by_type.visualization.shtml]

[11] Wurtele E, Li L, Berleant D, Cook D, Dickerson J, Ding J, Hofmann H, Lawrence M, Lee E, Li J, Mentzen W, Miller L, Nikolau B, Ransom N, Wang Y: MetNet: Systems Biology Software for Arabidopsis. In Concepts in Plant Metabolomics. Springer Verlag; 2007:145-158.

[12] Designing for colour-blindness by using safe web colors [http://safecolours.rigdenage.com/colours1.html]

[13] Making Websites Accessible: Color Scheme Planning, Part I [http://www.disabilitytraining.com/wpblog/making-websites-accessible-color-scheme-planning-part-i/]

[14] Likert, Rensis (1932). "A Technique for the Measurement of Attitudes". Archives of Psychology 140: 1–55.

List of figures

Figure 2.1 Snapshot of demo using Cytoscape Web

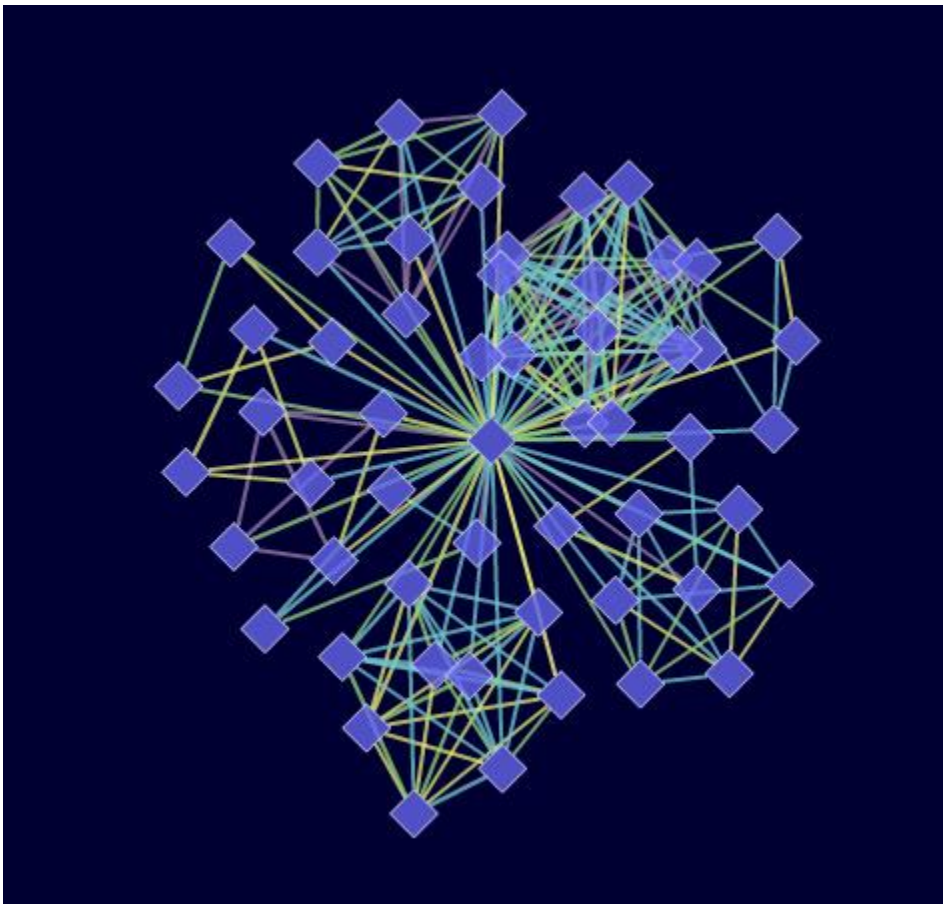


Figure 2.2 g:GOSt's Windows Explorer form of visual representation of the GO hierarchy

DAD1	P-value	T	Q	Q&T	Q&T/Q	Q&T/T	term ID	term domain and name
S	1.00e+00	15628	1	1	1.000	0.000	GO:0008150	BP biological_process (1)
e	1.00e+00	5536	1	1	1.000	0.000	GO:0032501	BP multicellular organismal process (2)
e	1.00e+00	4371	1	1	1.000	0.000	GO:0032502	BP developmental process (2)
e	1.00e+00	3759	1	1	1.000	0.000	GO:0048856	BP anatomical structure development (3)
e	1.00e+00	3924	1	1	1.000	0.000	GO:0007275	BP multicellular organismal development (3)
e	1.00e+00	855	1	1	1.000	0.001	GO:0009790	BP embryo development (3)
e	1.00e+00	457	1	1	1.000	0.002	GO:0009792	BP embryo development ending in birth or egg hatching (4)
e	1.00e+00	450	1	1	1.000	0.002	GO:0043009	BP chordate embryonic development (5)
e	1.00e+00	275	1	1	1.000	0.004	GO:0001701	BP in utero embryonic development (6)
e	1.00e+00	61	1	1	1.000	0.016	GO:0001824	BP blastocyst development (7)

Figure 2.3 **Snapshot of the tool**

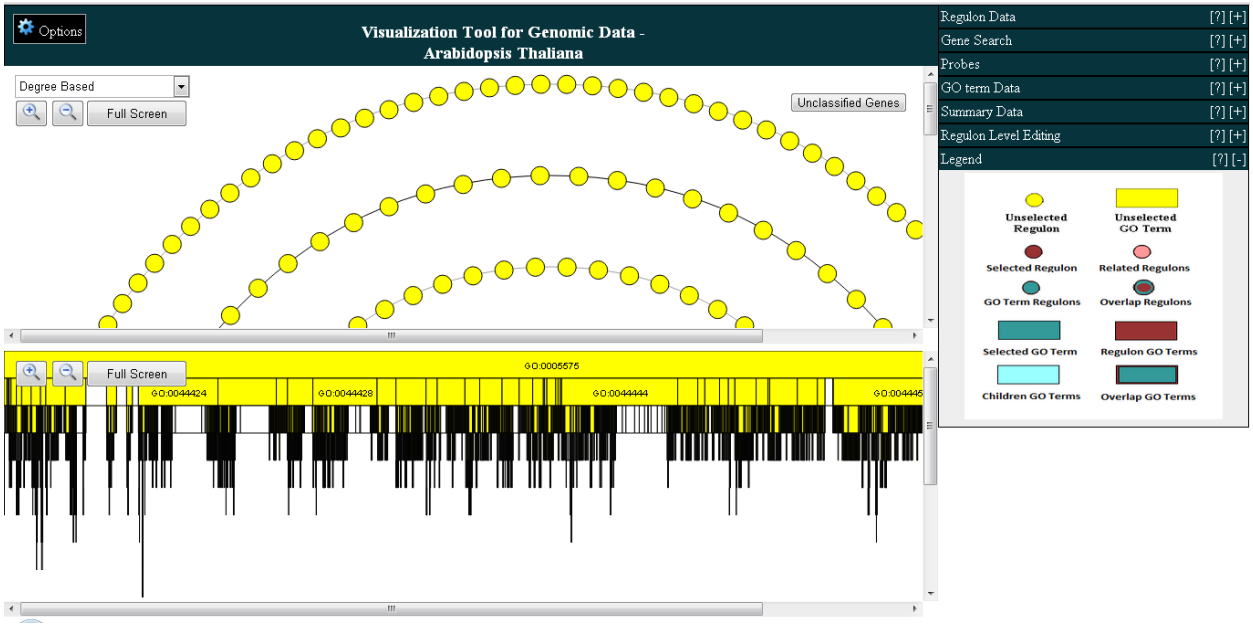


Figure 2.4 Top pane of the tool

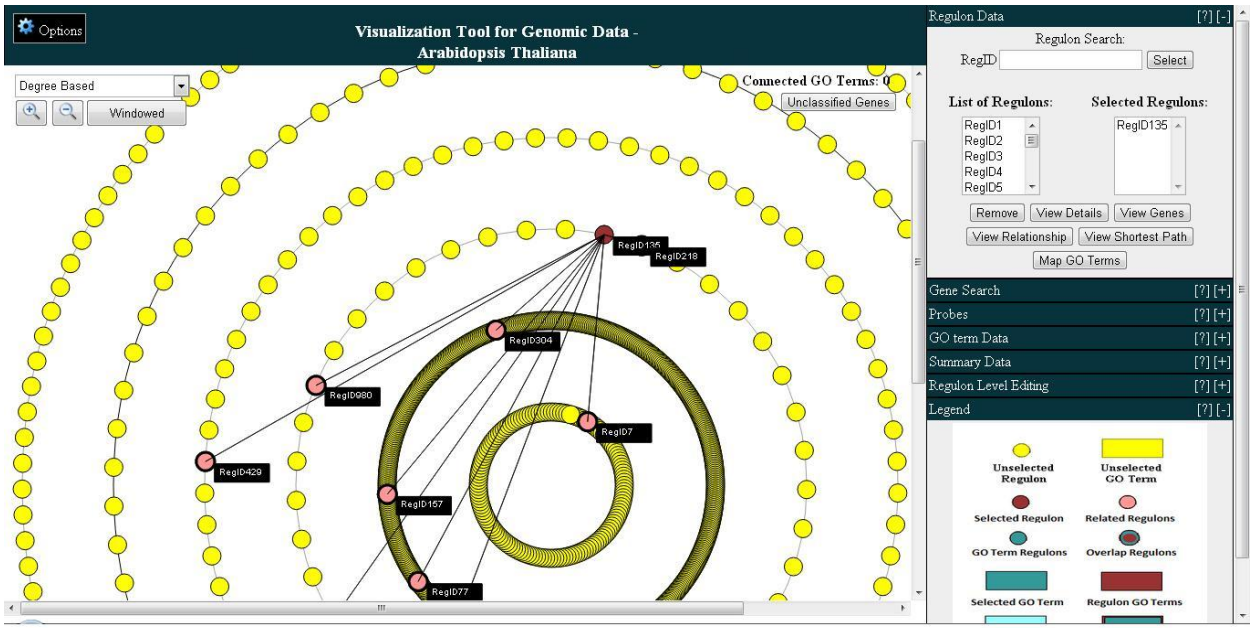


Figure 2.5 Bottom pane of the tool

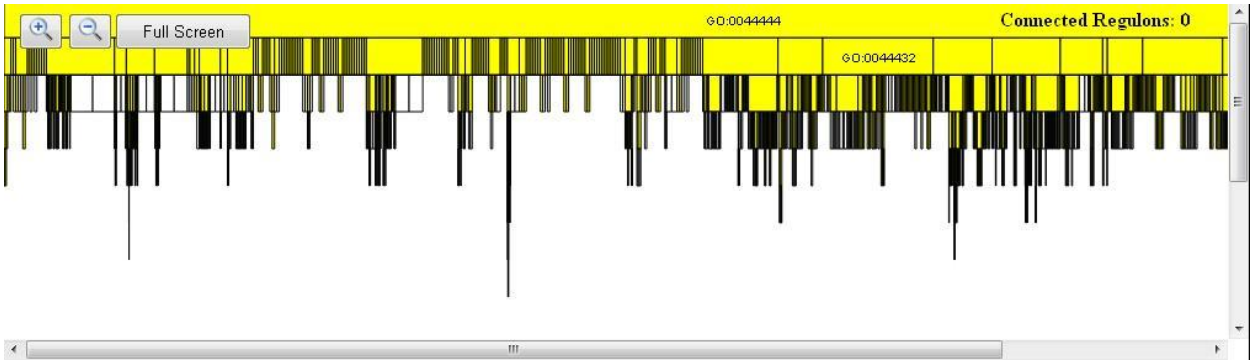
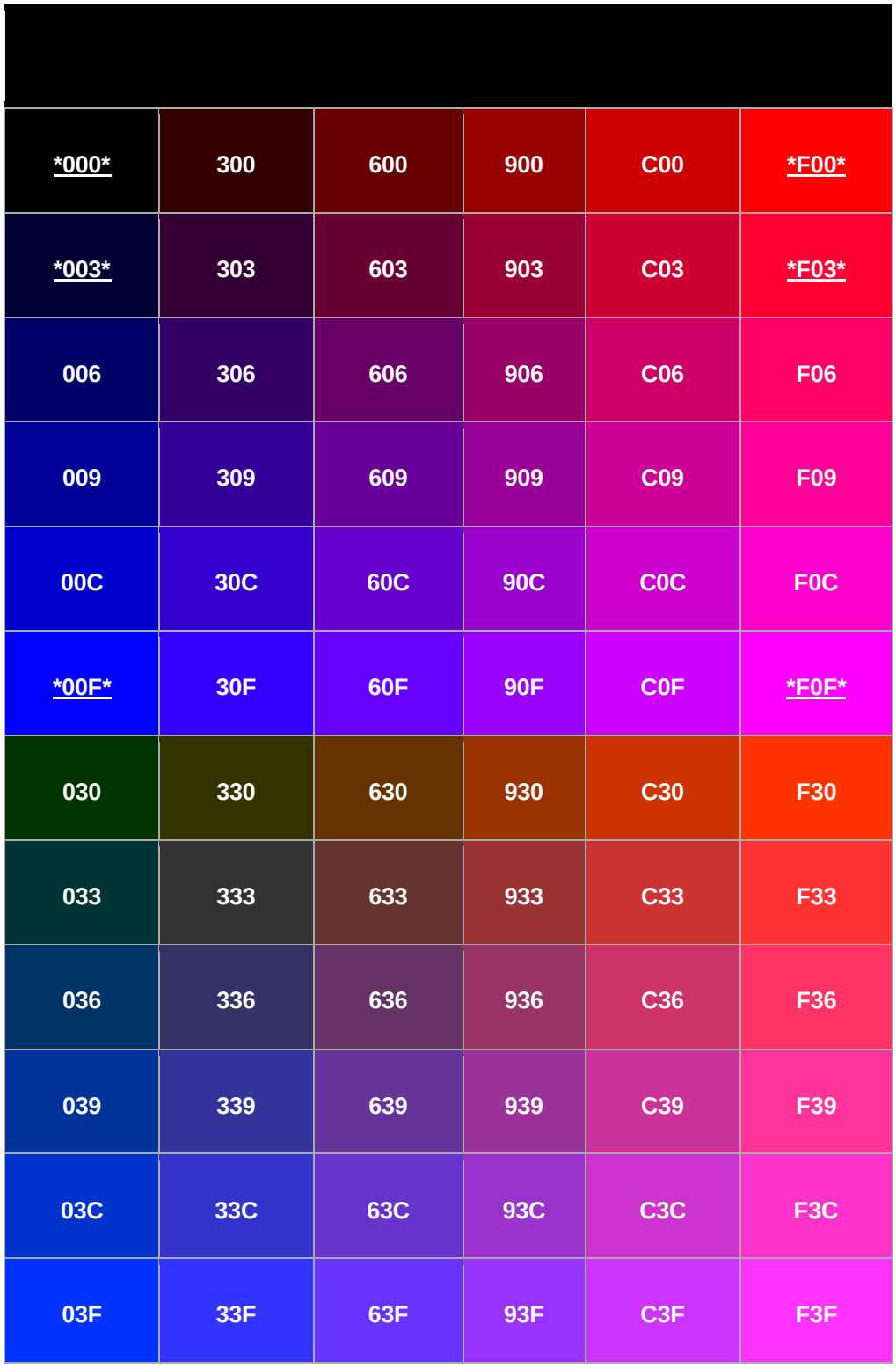


Figure 2.6 Web safe colors



060	360	660	960	C60	F60
063	363	663	963	C63	F63
066	366	666	966	C66	F66
069	369	669	969	C69	F69
06C	36C	66C	96C	C6C	F6C
06F	36F	66F	96F	C6F	F6F
090	390	690	990	C90	F90
093	393	693	993	C93	F93
096	396	696	996	C96	F96
099	399	699	999	C99	F99
09C	39C	69C	99C	C9C	F9C
09F	39F	69F	99F	C9F	F9F
0C0	3C0	6C0	9C0	CC0	FC0
0C3	3C3	6C3	9C3	CC3	FC3

0C6	3C6	6C6	9C6	CC6	FC6
0C9	3C9	6C9	9C9	CC9	FC9
0CC	3CC	6CC	9CC	CCC	FCC
0CF	3CF	6CF	9CF	CCF	FCF
0F0	3F0	*6F0*	9F0	CF0	*FF0*
0F3	*3F3*	*6F3*	9F3	CF3	*FF3*
0F6	*3F6*	6F6	9F6	*CF6*	*FF6*
0F9	3F9	6F9	9F9	CF9	FF9
0FC	*3FC*	6FC	9FC	CFC	FFC
0FF	*3FF*	*6FF*	9FF	CFF	*FFF*

Figure 2.7 Simulations of the chosen colors

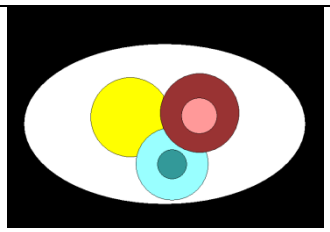
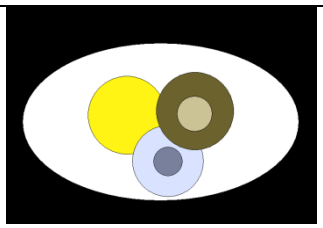
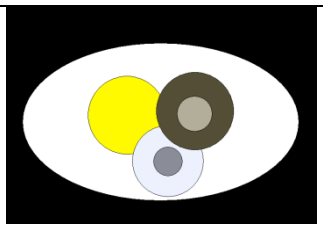
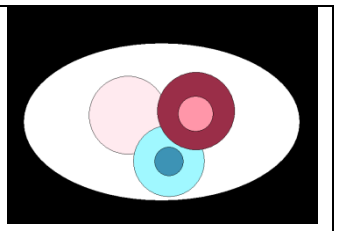
			
Normal Image	**Deuteranope Simulation**	**Protanope Simulation**	**Tritanope Simulation**

Figure 2.8 Snapshot of the tool with chosen set of colors

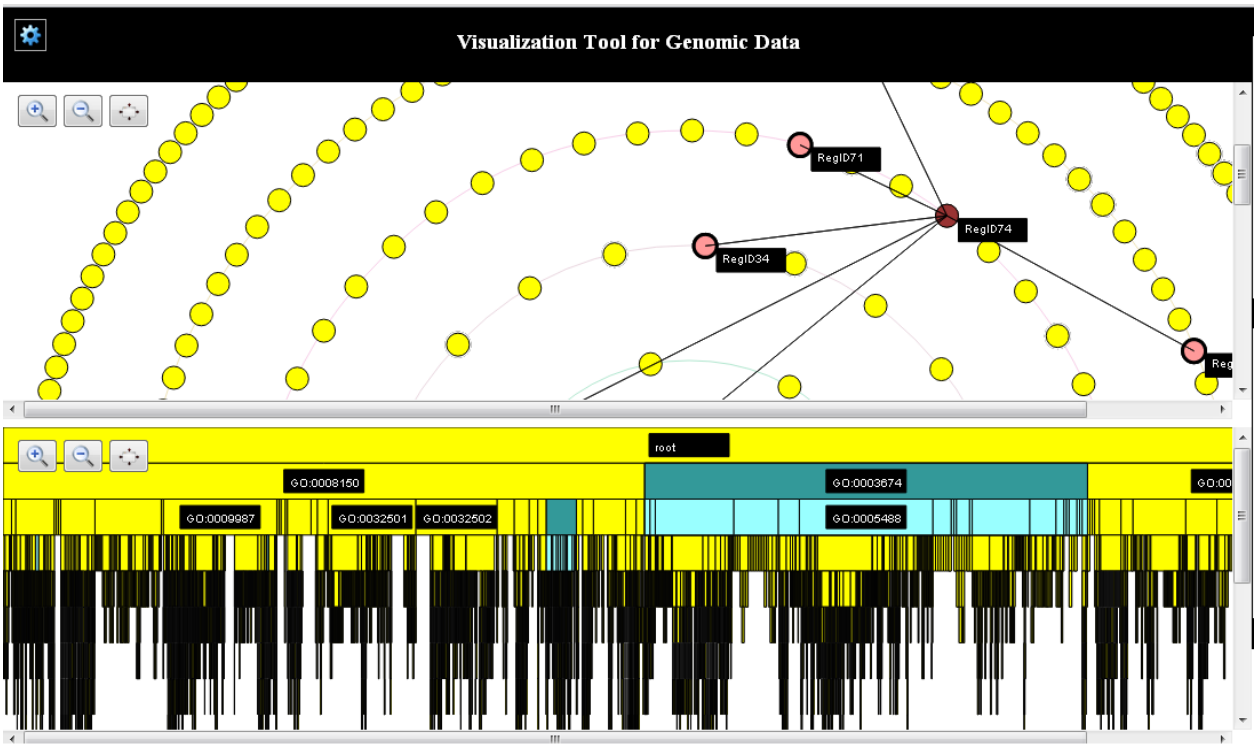


Figure 2.9 Gene search

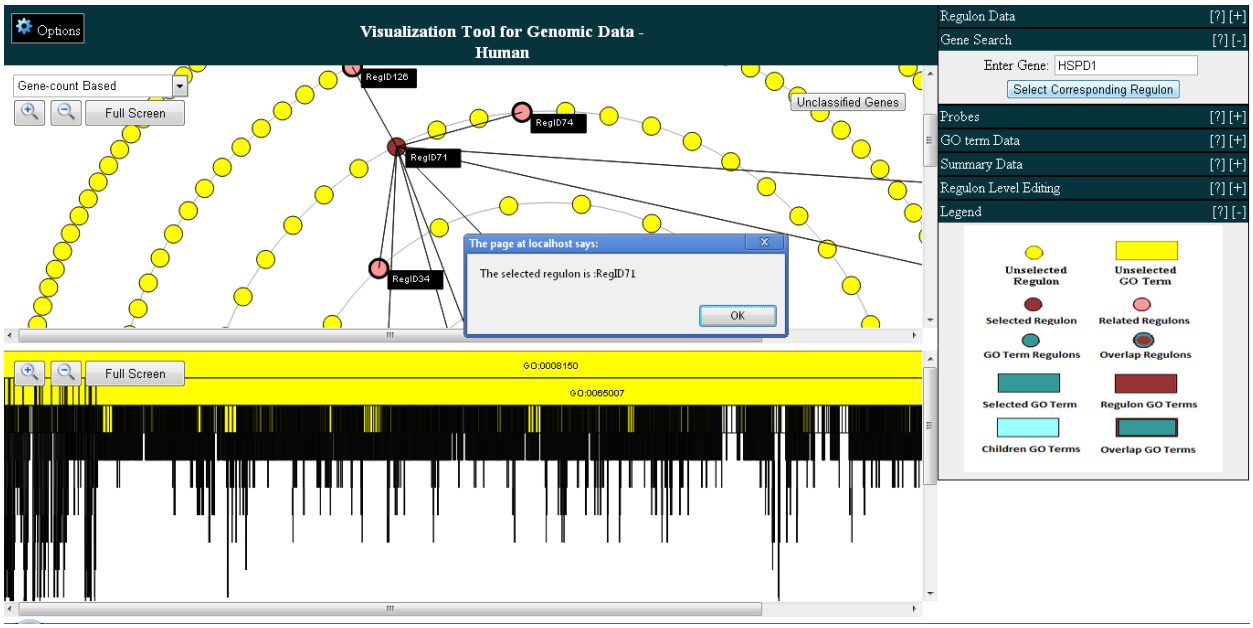


Figure 2.10 Ring summary

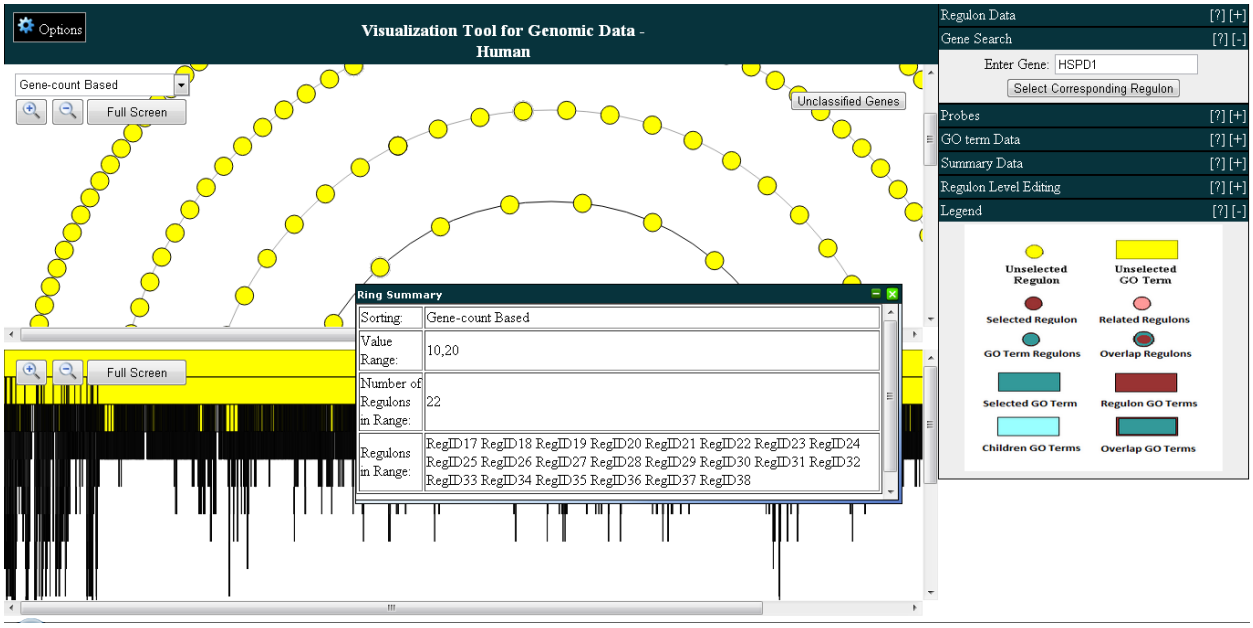


Figure 2.11 Summary data

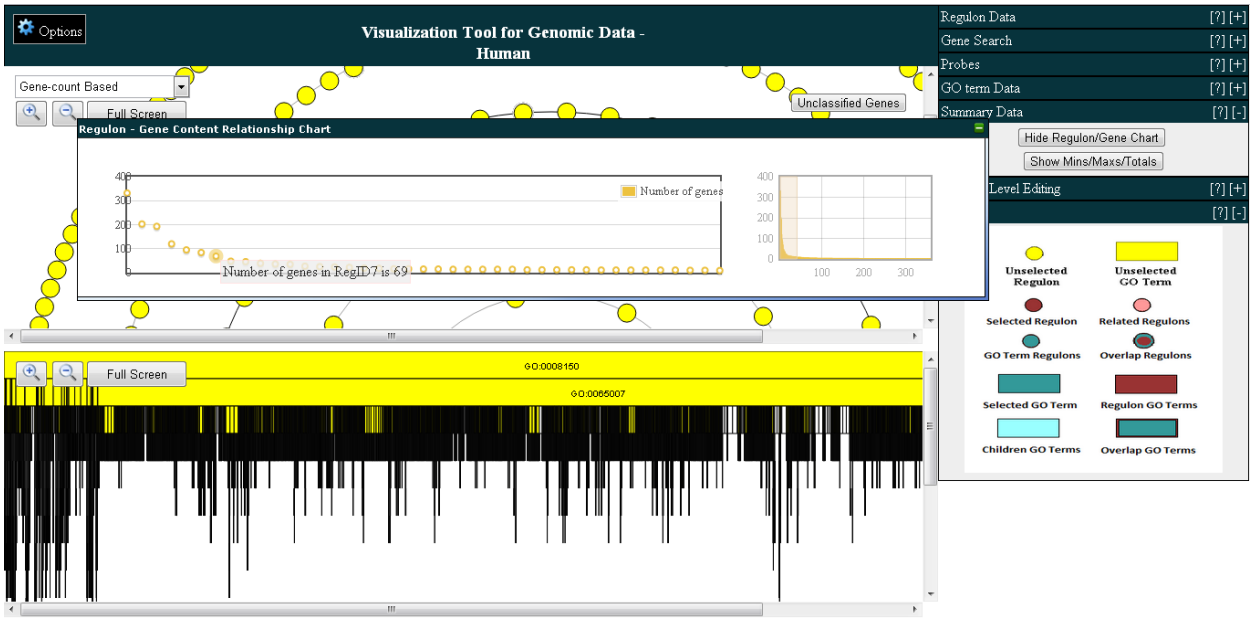


Figure 2.12 Gene Ontology – Gene mapping

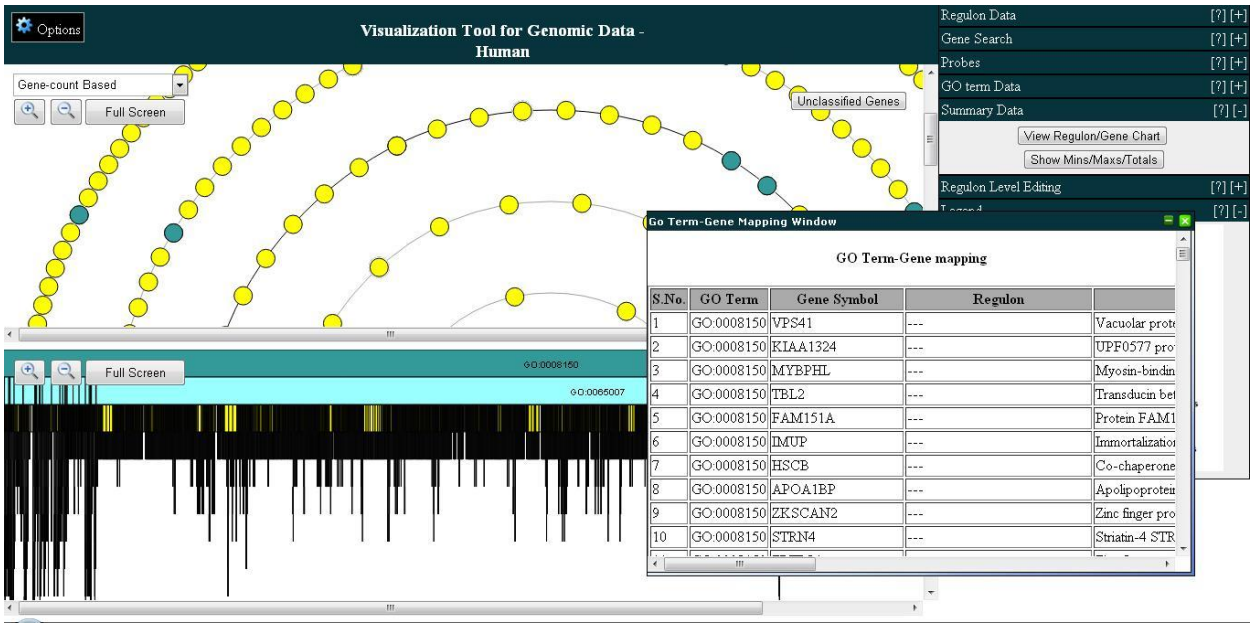
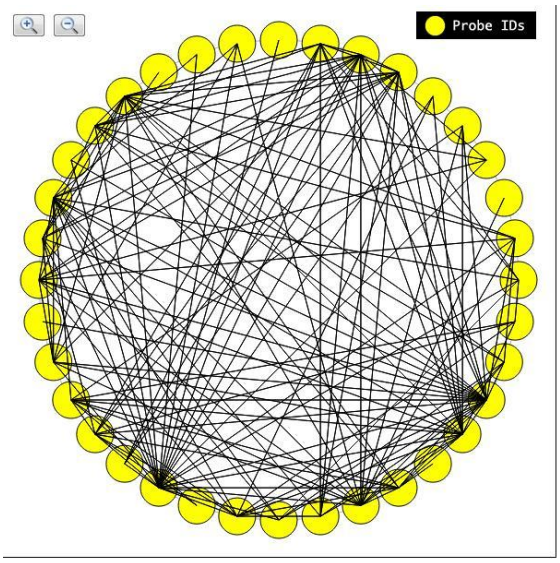


Figure 2.13 Correlation matrix and heat map



CORRELATION MATRIX

Probe IDs	200043_at	200093_s_at	200966_x_at	201119_s_at	201134_x_at	201226_at	201322_at	201
200043_at	1	0.82	---	---	---	---	---	
200093_s_at	0.82	1	---	---	0.75	---	---	
200966_x_at	---	---	1	0.75	---	---	0.72	
201119_s_at	---	---	0.75	1	0.73	---	0.77	
201134_x_at	---	0.75	---	0.73	1	---	0.71	
201226_at	---	---	---	---	---	1	0.71	
201322_at	---	---	0.72	0.77	0.71	0.71	1	
201441_at	---	---	0.74	0.82	0.72	---	0.7	
201754_at	---	---	---	0.72	0.81	---	---	
201840_at	0.71	---	0.7	0.7	---	---	---	
202233_s_at	---	0.74	---	0.71	0.77	---	0.73	
202941_at	---	---	---	---	---	---	---	
202961_s_at	0.73	---	---	0.79	0.7	---	0.7	
203478_at	---	---	---	---	---	---	---	
207507_s_at	---	0.7	---	0.75	---	---	0.76	
208909_at	---	0.7	---	0.75	0.71	---	0.83	
209224_s_at	---	---	---	0.72	0.7	---	---	
209329_x_at	---	---	---	0.72	---	---	---	
209492_x_at	---	---	---	0.75	0.79	---	---	
210149_s_at	---	---	---	0.72	---	---	---	

Figure 2.14 Loading time of MetViz for species - Arabidopsis Thaliana

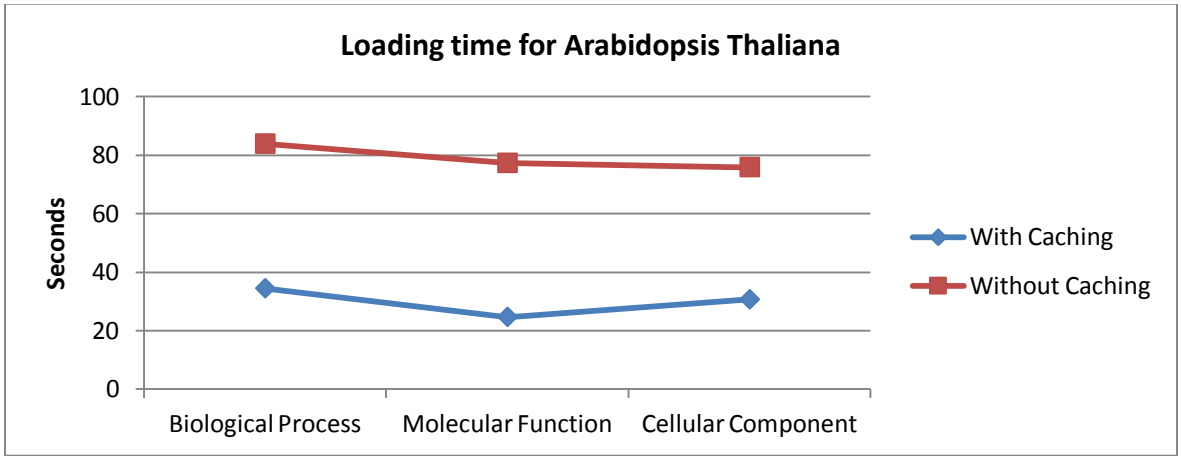


Figure 2.15 Loading time of MetViz for species - Homo Sapien

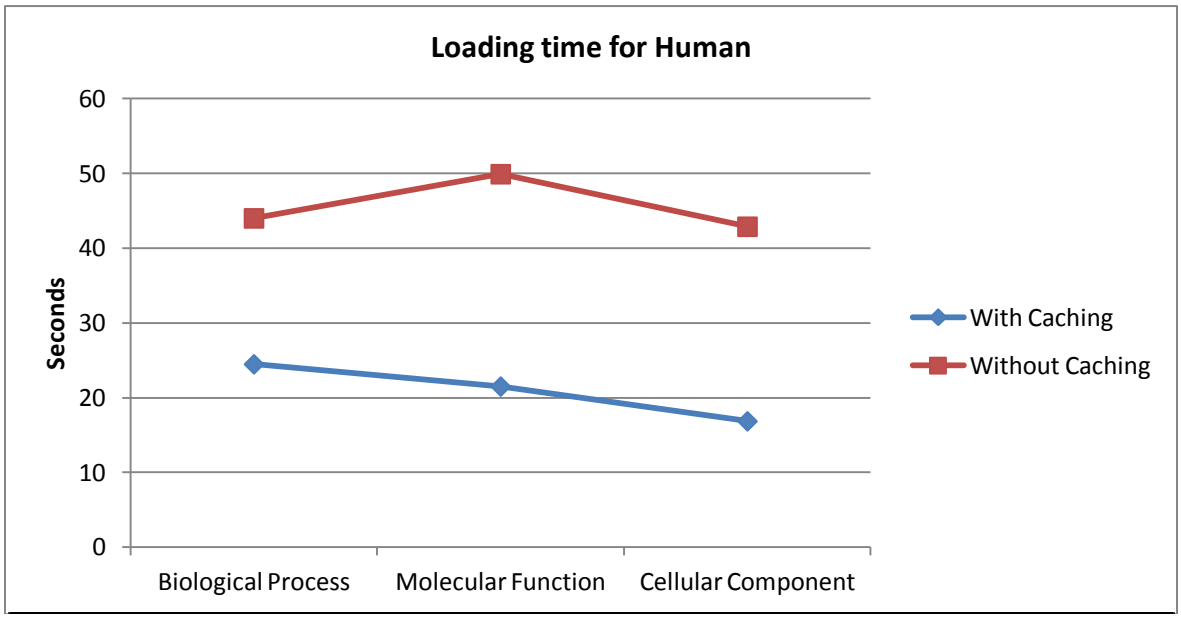
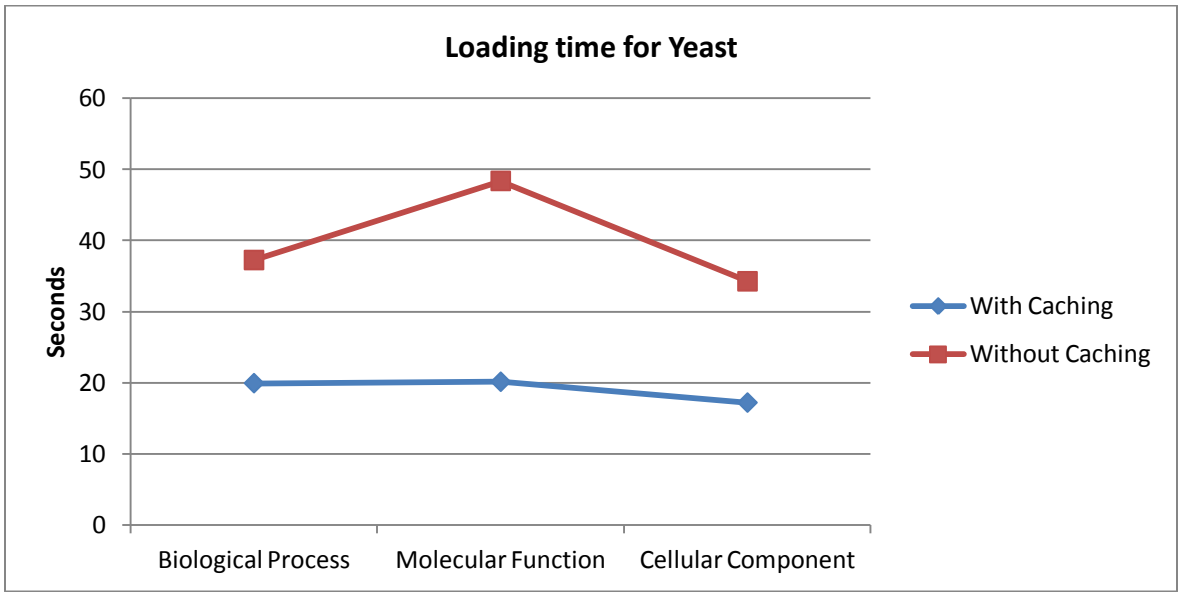


Figure 2.16 Loading time of MetViz for species - Saccharomyces cerevisiae



List of Tables

Table 2.1 Response times without caching

Gene Ontology Root Term	Species	Load time 1 (in seconds)	Load Time 2(in seconds)	Load Time 3(in seconds)	Average Load Time(in seconds)
Biological Process	Arabidopsis Thaliana	91.19	74.05	86.34	83.86
	Homo Sapien	48.72	42.10	41.15	43.99
	Saccharomyces cerevisiae	35.19	42.20	34.29	37.23
Molecular Function	Arabidopsis Thaliana	90.74	71.95	69.15	77.28
	Homo Sapien	41.63	52.23	55.85	49.90
	Saccharomyces cerevisiae	45.40	30.73	34.47	48.36
Cellular Component	Arabidopsis Thaliana	85.32	72.52	69.69	75.84
	Homo Sapien	43.52	42.87	42.22	42.87
	Saccharomyces cerevisiae	35.70	32.48	34.53	34.24

Table 2.2 Response times with caching

Gene Ontology Root Term	Species	Load time 1 (in seconds)	Load Time 2(in seconds)	Load Time 3(in seconds)	Average Load Time(in seconds)	Cached file size (in KB)
Biological Process	Arabidopsis Thaliana	42.50	26.32	25.94	34.41	86271
	Homo Sapien	23.21	25.73	22.48	24.47	37393
	Saccharomyces cerevisiae	20.65	19.17	27.22	19.91	27399
Molecular Function	Arabidopsis Thaliana	24.42	24.84	24.39	24.63	89199
	Homo Sapien	21.48	21.49	19.52	21.49	45458
	Saccharomyces cerevisiae	16.35	23.86	24.69	20.11	27103
Cellular Component	Arabidopsis Thaliana	25.17	36.08	23.62	30.63	84783
	Homo Sapien	19.06	14.61	12.21	16.84	45566
	Saccharomyces cerevisiae	19.89	14.46	13.83	17.18	27743

Table 2.3 Response times comparison

	Biological Process			Molecular Function			Cellular Component		
	Arabidopsis Thaliana	Homo Sapien	Saccharomyces cerevisiae	Arabidopsis Thaliana	Homo Sapien	Saccharomyces cerevisiae	Arabidopsis Thaliana	Homo Sapien	Saccharomyces cerevisiae
With Caching	34.41	24.47	19.91	24.63	21.49	20.11	30.63	16.84	17.18
Without Caching	83.86	43.99	37.23	77.28	49.90	48.36	75.84	42.87	34.24

Table 2.4 List of Tasks

Sample Task	Identify any 2 go terms that are mapped to the gene HSPD1
Task 1	Identify any 2 genes that correspond to the go term "biological process"(GO:0008150)
Task 2	Identify the number of regulons in the path between RegID52 and RegID145
Task 3	Change the view of the graph in the upper pane to sort it based on Intra-Regulon Density.
Task 4	In terms of the number of connecting gene pairs, could you tell how closely related are the Regulons "RegID4" and "RegID7"
Task 5	Identify a couple of regulons that are disconnected.
Task 6	Identify the regulon that contains the greatest number of genes?
Task 7	Identify the children for the GO term "adaptation of signaling pathway"(GO:0023058)
Task 8	Identify any 2 Probe ID pairs from RegID12 that have a pearson correlation value>0.7
Task 9	Delete level 501,12000 in the "Gene-count based" View of the top pane
Task 10	Load the file - "MyGenes.xml",and generate a correlation matrix for the genes present in the file.

Table 2.5 Result of pilot study

Tasks	Time to solve (in seconds)	How easy was this task to perform overall? (5- Easiest , 1 – Very Difficult)	How successful were you in accomplishing what you were asked to do? (5- Very Confident , 1 – Very doubtful)	Correct solution?
Task 1	29.3	5	5	Yes
Task 2	166.8	4	5	Yes
Task 3	12.3	5	5	Yes
Task 4	46.2	5	5	Yes
Task 5	30.8	5	5	Yes
Task 6	149.8	4	5	No
Task 7	79.7	5	5	Yes
Task 8	71	4	5	Yes
Task 9	22.5	5	5	Yes
Task 10	Could not solve	1	1	No

Table 2.6 Problems identified, corrections made and feedback obtained

Problems identified	Feedback obtained	Corrections made
GO term added multiple times Should define what 1 and 5 on the Likert scale mean Shouldn't include the time it takes for the user to read through the tasks. Not showing error when invalid regulon selected	Labelling was good It was not obvious that the go terms were actually selected when selecting through the right pane Move the select button to the right Set default button (clicking on enter should choose the button) The input text boxes should not be case sensitive.	GO term added multiple times – solved Move the select button to the right – Most functionalities were moved to the right pane It was not obvious that the go terms were actually selected when selecting through the right pane – The GO Terms were highlighted and zoomed into as they were selected from the side pane. set default button (clicking on enter should choose the button) – done The input text boxes should not be case sensitive.- done

		Should define what 1 and 5 on the Likert scale mean – done Shouldn't include the time it takes for the user to read through the tasks. – time not included in the actual study Not showing error when invalid regulon selected – a dialog box now appears

Table 2.7 Results of user study – How easy was it to accomplish the task

	Participants						
	1	2	3	4	5	6	7
Task 1	5	4	4	4	5	4	5
Task 2	5	3	2	5	3	3	5
Task 3	5	5	5	5	5	5	5
Task 4	5	4	5	5	5	3	5
Task 5	3	5	2	5	5	4	5
Task 6	1	4	1	3	3	1	2
Task 7	3	4	4	5	4	3	5
Task 8	2	5	4	4	4	3	4
Task 9	5	5	5	5	5	5	5
Task 10	5	5	3	5	5	4	5

Table 2.8 Results of user study – How confident are you about the answer

	Participants						
	1	2	3	4	5	6	7
Task 1	5	5	5	5	5	5	4
Task 2	5	5	4	5	5	5	5
Task 3	5	5	5	5	5	5	5
Task 4	5	3	5	5	5	5	5
Task 5	2	5	2	5	5	4	5
Task 6	1	5	1	5	5	1	2
Task 7	5	5	3	5	5	5	5
Task 8	2	5	5	5	5	4	5
Task 9	5	5	5	5	5	5	3
Task 10	5	5	1	5	5	5	5

Table 2.9 Results of user study – Were the users correct in their solutions

	Participants						
	1	2	3	4	5	6	7
Task 1	yes	yes	Yes	yes	yes	yes	Yes
Task 2	yes	yes	Yes	yes	yes	yes	yes
Task 3	yes	yes	Yes	yes	yes	yes	yes
Task 4	yes	yes	Yes	yes	yes	yes	Yes
Task 5	yes	yes	Yes	yes	yes	yes	Yes
Task 6	No	No	No	No	No	No	No
Task 7	yes	yes	Yes	yes	yes	yes	Yes
Task 8	No	yes	yes	yes	yes	yes	Yes
Task 9	yes	Yes	Yes	yes	yes	yes	Yes
Task 10	yes	yes	No	yes	yes	yes	Yes

Table 2.10 Results of user study – Time required to complete tasks

	Participants						
	1	2	3	4	5	6	7
Task 1	2:19	1:01	2:35	1:38	0:54	1:27	1:23
Task 2	0:49	1:35	3:47	1:47	5:20	2:53	1:00
Task 3	0:21	0:14	0:18	0:10	0:12	0:14	0:22
Task 4	1:01	2:24	0:38	0:59	0:51	2:20	1:00
Task 5	6:12	1:54	0:45	2:13	0:48	1:03	0:35
Task 6	5:26	5:27	2:38	3:40	3:23	5:11	4:41
Task 7	3:00	3:02	1:42	1:07	3:27	10:16	1:27
Task 8	4:25	1:13	2:29	2:28	1:41	3:40	3:28
Task 9	0:32	0:14	0:18	0:25	0:26	0:51	0:23
Task 10	0:24	1:17	1:13	2:11	1:20	2:39	1:10

CHAPTER 3. GENERAL CONCLUSION

Analyzing interactions between regulons and understanding gene pair up or down regulations could be helpful for hypothesis building from the wealth of microarray data. Unrelated biological processes can link up on analysis suggesting novel signaling events we might have missed or not even acknowledged. Although there are many softwares available today [3-10], they tend to be cumbersome to the user when it comes to visualizing and interpreting large volumes of data.

MetViz is a web-based tool that is simple to use and has the ability to display large volumes of data and information about them by an appropriate combination of visual elements and textual content. It acts as a browser for GO Terms, genes and regulons and enables visual mapping with each other. It is a tool that could be used along with tools such as the MetaOmGraph to help identify interesting relationships and functionalities previously unknown.

APPENDIX A - List of Modifications

List of modifications done to MetViz after pilot study

1. Change the name of the settings icon to "Options".
2. Put all the functionality as buttons on the right.
3. Minimize all panels other than the legend on the right - would present a simple a screen
4. Deselecting a go term redraws the text.
5. Moved the select button to the right.
6. Change the name of the full screen icon to "Full Screen"
7. Mouse hover on the list boxes would show entire name.
8. Error shows up when an invalid regulon is typed
9. Default button was set
10. On deselecting a GO term, deselect the related regulons and child go terms.
11. The text boxes were made to not be case sensitive.
12. Select all button was added to the Gene Window.
13. Indicate genes from the unclassified regulon.
14. GO Terms were zoomed in by default when selected on the right pane.

List of modifications done to MetViz after user study

1. 'Level Editing' renamed to 'Regulon Level Editing'
2. Moved legend down in the side pane
3. No hidden buttons were present
4. Correlation window was displayed on the side pane
5. Redesigning of the correlation window
6. Functionality removed from under the Options button
7. '?' help button placed to explain what different tabs mean
8. Handled the cases
    a. when "GO:" not given
    b. when "RegID:" given in the search boxes
9. Separate window was created for load and save data
10. Searching for a GO term would automatically zoom to that GO term.
11. Regulon summary button in the regulon data window - separate section for summary data.
12. Fullscreen->Windowed label switch
13. Changed label 'genes' -> 'probes' in view details
14. Increased size of side pane
15. Reset button in popup menu in GO term pane
16. Ring summary - user didn't initially understand what the modes meant. Ring summary gives him an idea.