Development of Integrated Visualization Tool for Differentially Expressed Genes and Gene Ontology Analysis

Somyung Oh[1], Junghyeon Ha[2], Kyungwon Lee and Sejong Oh

Dept. of Digital Media, Ajou University
206 Worldcup-ro, Yeongtong-gu, Suwon 16499, Korea
Dept. of Software Science, Dankook University
152, Jukjeon-ro, Suji-gu, Yongin 16890, Korea
[1]*somyungoh@gmail.com,* [2]*hajh0104@gmail.com,* [3]*swissrain@gmail.com,*
sejongoh@dankook.ac.kr

Abstract

To study biological significance, biologists deal with differentially expressed genes and corresponding gene ontology data. For more intuitive research, biologists visualize them with heatmap and gene ontology graph, respectively. We propose a tool that integrates these two visualization and enables them to have interaction. By integrating two independent visualization, we expect to get more insight in biological analysis trough the proposed tool.

Keywords: microarray, differentially expressed gene, gene ontology, heatmap.

1. Introduction

Biologists compare the expression levels of genes between organisms to find the biological meaning of life. If the expression level exceeds the threshold value, it is determined to be a significant gene and classified as differentially expressed genes (DEG). Expression levels of genes are generally measured by microarray [1]. Data of each gene is extracted by point and color through microarray as shown in Figure 1, and numerical expression value is calculated according to color. Heatmap visualization is commonly used to easily compare and recognize the expression values of genes between each organism (see Figure 1).
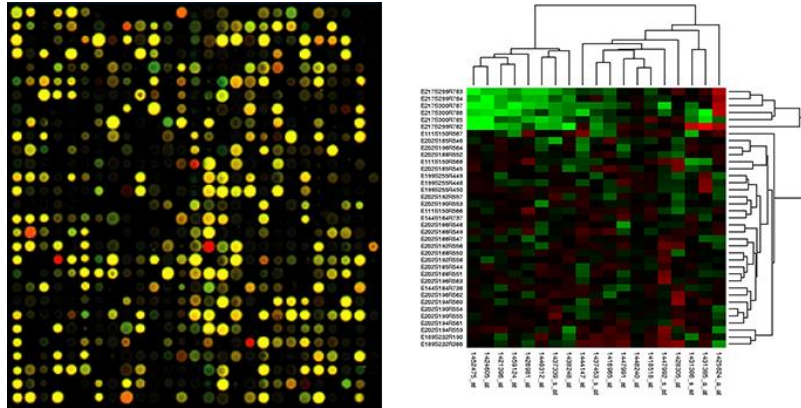
Figure 1. An example of microarray

DEG has biological meanings when several are assembled together. The conceptualization of the biological meanings of genes is called gene ontology [2]. The gene ontology database is structured in hierarchical directed acyclic graph. To increase the chances of finding new biological meanings a good gene ontology graph visualization is needed.

In this paper, we propose an integrated visualization tool an analysis tool for the heatmap and the gene ontology graph (GO graph). This tool is an advanced version of the tool already introduced [3]. The proposed tool provides the function of extracting DEG from microarray input data and finds corresponding gene ontology from previous found DEG set. Based on the extracted data, the proposed tool visualizes heatmap and GO graph. The proposed tool has overcome the difficulties of users in analyzing process due to existence of heatmap and GO graph independently, by integrating two visualizations interact with each other in real-time. The proposed tool also focused on the aesthetics of allowing users to easily and intuitively understand and analyze complex data [4].
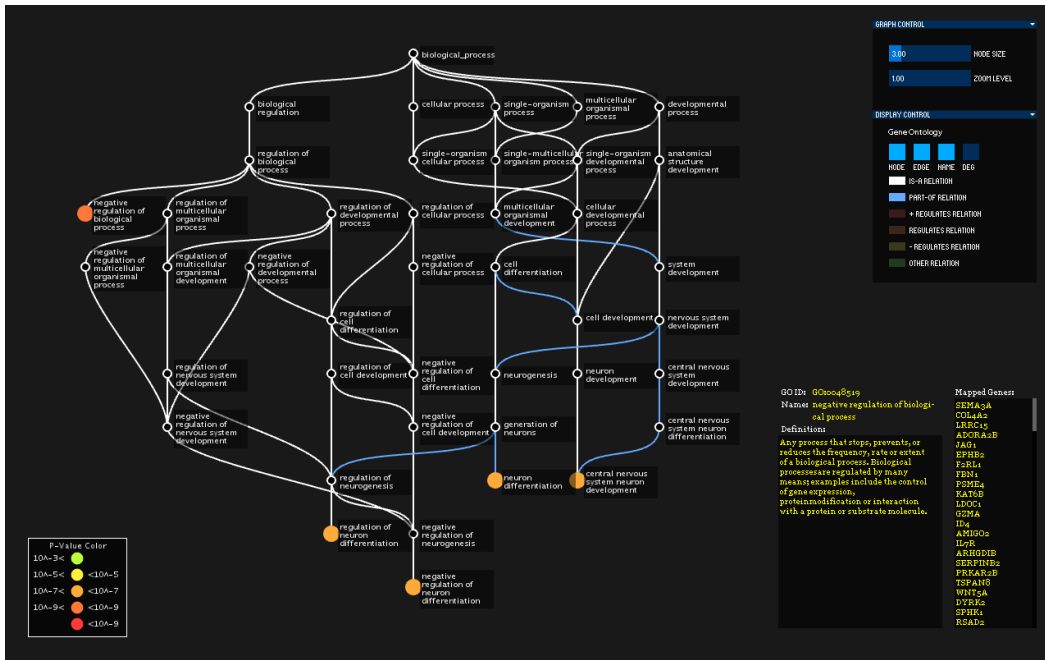
1.1. Related works

Various programs are available for DEG and gene ontology analysis. Typically, there are GOrilla [5] and BiNGO [6] developed using software graphViz [7] and Cytoscope [8]. The heatmap and GO graph provided by them have static limits that the user cannot modify. Other tools that are available also have limitations that can be analyzed independently of either the heatmap or the gene ontology. However, considering that there is a deep correlation between DEG and gene ontology in the biological research process, it is necessary that each analysis process is integrated into one and making available for heatmap and the GO graph to interact each other.
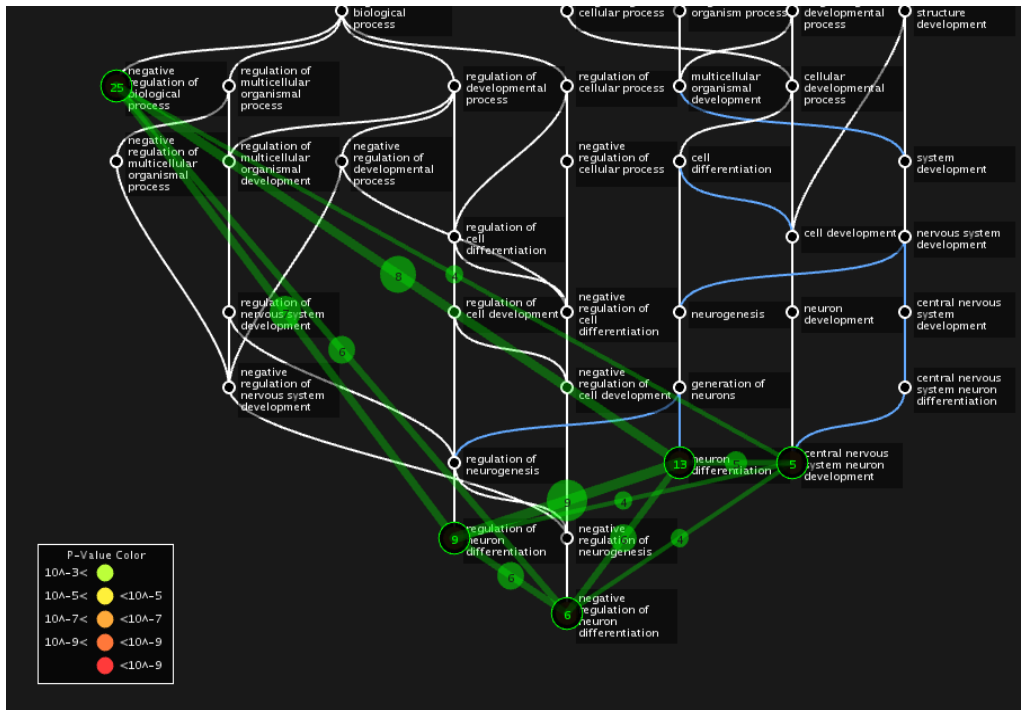
2. Methods

2.1. DEG heatmap and gene ontology graph visualization

Two visualization - DEG heatmap and GO graph provided by the proposed tool have been developed to allow users to interact with the user to view and observe vast amounts of information. First, the heatmap shows specific information about the gene, such as gene ID and organism name, when the user hovered over a particular gene, as shown in Figure 2. Secondly, GO graph enables various display settings such as ontology name and relationship.

Also, like the heatmap, when the user hovers over a particular ontology, specific information about the ontology appears on the right side of the screen, as shown in Figure 3(a).

2.2. Interactivity between DEG heatmap and gene ontology graph

The main novelty of the proposed tool is that heatmap and GO graph interact with each other. Since DEG and corresponding ontology has relationship, the proposed tool connects two visualization with real-time interactivity. For example, if the user selects the DEG set in the heatmap, the corresponding ontologies respond in the GO graph. Conversely, if you select a specific ontology in the GO graph, the genes associated with it will react in the heatmap (see Figure 4). This minimizes the inconvenience of existing tools as they provide independent visualization, and helps to find new information that was not seen before.

Figure 2. Visualization of DEG heatmap

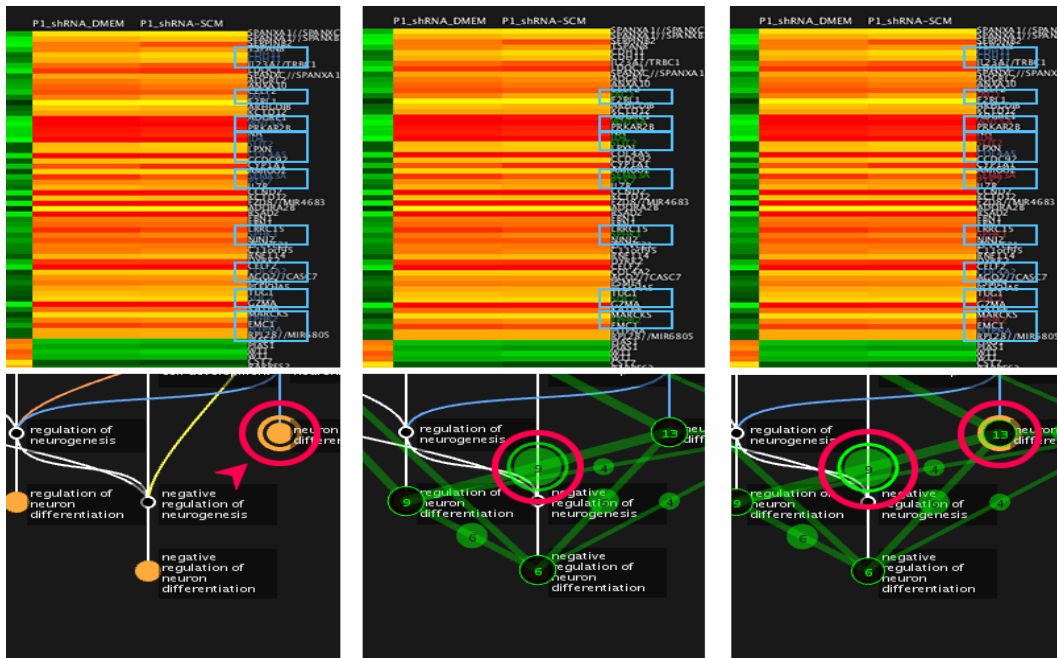**(a) Overall gene ontology graph**

(b) Visualization of GO Terms sharing genes

Figure 3. Visualization of gene ontology graph
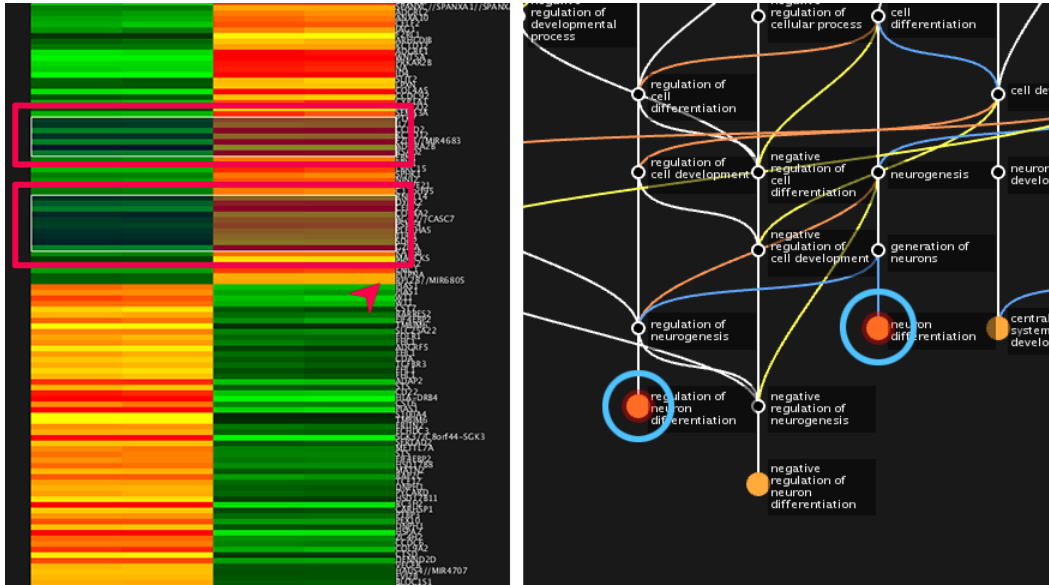
(a) Response (highlighted cyan) from GO Graph when DEG set is selected (highlighted magenta)

(b) Response(highlighted cyan) from heatmap in three ways of GO term selection(highlighted magenta)

Figure 4. Interaction between heatmap and gene ontology graph

3. Conclusion

The task of analyzing genes by aggregation is an important methodology in biological research. Gene ontology defines the meanings when bound to a gene set. The proposed tool provides visualization functions that allow more intuitive analysis of gene data and GO. We expect the proposed tool to contribute to the genetic analysis process and helps biologists to discover new biological meanings.

Acknowledgement

References

[1]   N. Altman, "Replication, variation and normalisation in microarray experiments", Applied bioinformatics, Vol. 4, No. 1, pp. 33-44, **(2005).**

[2]   Gene Ontology Consortium, "The Gene Ontology (GO) database and informatics resource". Nucleic acids research, 32(supply 1), D258-D261, **(2004)**.

[3]   S. Oh, J. Ha, K. Lee, and S. Oh, Editors. "Integrated Visualization Tool for Differentially Expressed Genes and Gene Ontology Analysis". In Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, October 2-5; **(2016),** Seattle, United States.

[4]   N. Cawthon and A.V. Moere, Editors. "The effect of aesthetic on the usability of data visualization". In Information Visualization, 2007. IV'07. 11th International Conference, July 4-6; Zurich, **(2007),** Switzerland.

[5]   E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini. "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists". BMC bioinformatics, Vol. 10, No. 1, pp. 48, **(2009).**

[6]   S. Maere, K. Heymans, and M. Kuiper. "BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks". Bioinformatics, 21(16), pp. 3448-3449 **(2005)**

[7]   J. Ellson, E. Gansner, L. Koutsofios, S.C. North and G. Woodhull, Editors, "Graphviz—open source graph drawing tools". In International Symposium on Graph Drawing, September 23-26; **(2001),** Vienna, Austria.

[8]   P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks", Genome Research, Vol. 13, No. 11, pp. 2498-2504, **(2003).**