

# KEGG-Based Pathway Visualization Tool for Complex Omics Data

**Nobuaki Kono**<sup>1,2</sup>  
t04331nk@sfc.keio.ac.jp

**Kazuharu Arakawa**<sup>1,3</sup>  
gaou@sfc.keio.ac.jp

**Yohei Yamada**<sup>1,2</sup>  
skipper@g-language.org

**Hirota da Mori**<sup>1,4</sup>  
hmori@sfc.keio.ac.jp

**Masaru Tomita**<sup>1,2</sup>  
mt@sfc.keio.ac.jp

<sup>1</sup> Institute for Advanced Biosciences, Keio University, Endo 5322, Fujisawa, Kanagawa 252-8520, Japan

<sup>2</sup> Department of Environmental Information

<sup>3</sup> Bioinformatics Program, Graduate School of Media and Governance

<sup>4</sup> Nara Institute of Science and Technology (NAIST), 8916-5 Takayama, Ikoma, Nara 630-0101, Japan

**Keywords:** pathway map, scalable vector graphics, visualization, systems biology

## 1 Introduction

To understand the physiological network of a whole cell, integrative systems biology approach is necessary to view the life as a complex system. However, this will result in huge masses of data to be handled systematically, and scientific visualization is a key technology to enhance our understanding of the data. For this purpose, we have developed a software system that visualizes complex omics data onto KEGG (Kyoto Encyclopedia of Genes and Genomes) [1] pathway map in vector graphics. This system can simultaneously map objects in different biological layers, such as the genes, enzymes, and metabolites, each in distinct color spectrums that represent their values corresponding to the concentrations or intensities. A database of visualized microarray data as well as a web-service of this visualization tool is available at: <http://www.g-language.org/data/marray/> [2].

## 2 Method and Results

The software system is developed upon the generic bioinformatics workbench, G-language Genome Analysis Environment [3]. At first, this system draws the pathway map in vector graphics based on the coordinates of nodes and interactions acquired from KGML (KEGG Markup Language) files, and subsequently maps genes and metabolism in color codes. Several pathway visualization tools already exist, but to the best of our knowledge, our tool is the only one that can simultaneously visualize complex omics data encompassing different biological layers in vector format, utilizing the familiar KEGG pathway data. The vector images are web-ready, and therefore platform independent. Required input for this tool is a list of name-value pairs, where the name is a KEGG compound ID for metabolites, EC number for enzymes, and canonical or general gene names for proteins and mRNA, and the value is an integer in the range of 1 to 100. The system interprets the type of an entity from the specified name, and maps the corresponding value on the pathway with color. Metabolites are represented in a color spectrum ranging from red (1) to blue (100), and gene products are likewise represented in color from yellow (1) to green (100). The first version of our software, which is currently available as a web-service, displays the mapped image in FLASH vector image. We are

further enhancing the tool to support another vector graphics format, SVG (Scalable Vector Graphics), which is based on XML and therefore easily editable by hand, with computer programs, or with commercial vector drawing software such as Adobe Illustrator. Moreover, the new version supports integrated pathway maps as opposed to the subdivided KEGG pathway maps, in order to view the entire metabolic pathway at a glance to capture the cell-wide activity, and use the merits of vector graphics to enlarge and see the specific subparts of the pathway when interested. Figure 1 is the integrated view of the entire carbohydrate metabolism pathways. Similarly, integrated maps are available for energy metabolism, lipid metabolism and so on.

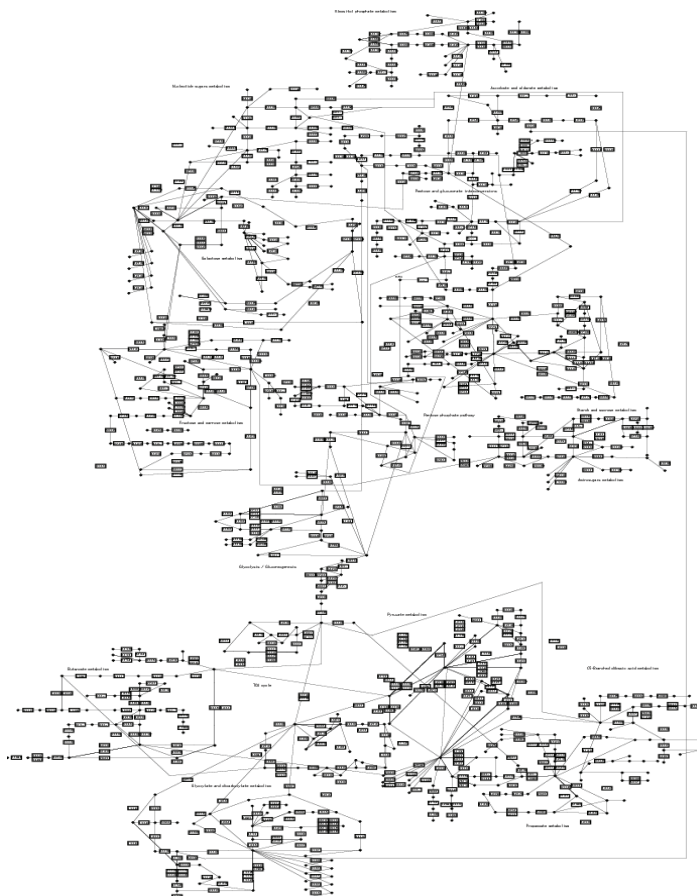


Figure 1: Integrated Map of Carbohydrate Metabolism

### 3 Conclusion

Pathway visualization is an important method to aid our understanding of complex nature of cellular systems. Our software, which is available as a web-service, is a powerful, yet easily accessible tool for this purpose.

### References

- [1] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M., KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, 27(1):29-34, 1999.
- [2] Arakawa, K., Kono, N., Yohei, Y., Hirotsada, M., and Tomita, M., KEGG-based pathway visualization tool for complex omics data, *In Silico Biol.*, 5:0039, 2005.
- [3] Arakawa, K., Mori, K., Ikeda, K., Matsuzaki, T., Kobayashi, Y., and Tomita, M., G-language genome analysis environment: a workbench for nucleotide sequence data mining, *Bioinformatics*, 19(2):305-306, 2003.