

Poster J-27

KEGG-based pathway visualization tool for complex omics data



Authors:

Nobuaki Kono (*Institute for Advanced Biosciences, Keio University*)

Kazuharu Arakawa (*Japan Society for the Promotion of Science*)

Yohei Yamada (*Institute for Advanced Biosciences, Keio University*)

Hirotsada Mori (*Nara Institute of Science and Technology*)

Masaru Tomita (*Institute for Advanced Biosciences, Keio University*)

Short Abstract: We have developed a pathway mapping tool for the visualization of multiple omics data, including transcriptome, proteome, and metabolome. The software maps given data onto an integrated pathway map based on the KEGG pathway diagrams in SVG format. The software is available as a web application at: <http://www.g-language.org/data/marray/>.

Long Abstract:

The introduction of high throughput measurement technologies resulted in the rapid accumulation of comprehensive data, and in order to understand the omics data in the context of physiological networks of a whole cell, integrative systems biology approach is necessary to view the life as a complex system. Scientific visualization is a key technology to enhance our understanding of the huge masses of data. For this purpose, we have developed a software system that visualizes complex omics data onto KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway map in vector graphics. This system can simultaneously display the values of multiple layers of omics, including transcriptome, proteome, and metabolome. 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/>.

Several pathway mapping and visualization tools already exist, but most of them are focused on a specific layer of data such as the transcriptome data from microarray experiments, and therefore are unable to map data from multiple layers. This is especially important for the understanding of systems biology analysis using computer simulation of the cellular pathways, which produces quantitative data for enzyme and metabolite concentrations. Moreover, most software systems produce graphics in bitmap, so that they are not scalable and cannot be easily manipulated afterwards. Our visualization tool enables simultaneous mapping of gene, mRNA, protein, and metabolite data on a single pathway map, in vector formats in FLASH (SWF) and Scalable Vector Graphics (SVG) formats. The latter format is a subset of an eXtensible Markup Language (XML) and therefore it can be easily handled with computer programs or text editors. SVG files can also be loaded and edited by common image editing software such as GIMP and Adobe Illustrator, so that the users can add on new components and graphics onto the mapped pathway image, and customize the colors and values manually. Both vector image formats are web-ready, and therefore platform independent. The software system is developed upon the generic bioinformatics workbench, G-language Genome Analysis Environment (Arakawa, K. et al. 2003) with Perl programming language.

Given a list of canonical/common gene names, enzyme classification numbers, or KEGG compound IDs with their values for mapping, and the choice of pathway in a choice of species, 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 metabolites in color codes. The input values should be an integer in the range of 1 to 100, and genes/mRNAs/proteins are colored with red to green spectrum, and metabolites are colored with blue to yellow spectrum. For example, a metabolite with the value of 1 will appear in bright blue, and a protein with the value of 100 will be in bright green. Heteropolymeric enzymes composed of multiple genes/proteins are correctly mapped, by subdividing the box representing the enzyme with component genes.

The latest version of our mapping tool can produce integrated pathway maps combining the pathway specifically subdivided KEGG pathway maps to show, for example, the entire carbohydrate metabolism map, in order to view the entire metabolic pathway at a glance to capture the cell-wide activity. Using the merits of vector graphics to be scalable, users can interactively view the specific subparts of the pathway of interest by zooming in on certain parts without losing high image resolutions. Entities mapped in this integrated pathway are also clickable, leading to external databases for further information. Pathway visualization is an important method to aid our understanding of complex nature of cellular systems, and our software, which is available as a web-service, is a powerful, yet easily accessible tool for this purpose.

Arakawa K, Mori K, Ikeda K, Matsuzaki T, Kobayashi Y, Tomita M, "G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining", *Bioinformatics*, 2003, 19(2):305-306

Arakawa K, Kono N, Yamada Y, Mori H, and Tomita M, "KEGG-based pathway visualization tool for complex omics data", *In Silico Biology*, 2005, 5:0039