# SVGMapping: an R package to map omic data sets onto pathways templates

Raphaël Champeimont, Christophe Leplat, Franck Chauvat, Jean-Christophe Aude

# SVGMapping: an R package to map *omic* data sets onto pathways templates

**R Champeimont**[a,b], **C Leplat**[a], **F Chauvat**[a] and **JC Aude**[a]

[a]Integrative Biology and Molecular Genetics Department
CEA, Institute of Biology and Technology Saclay
F-91191 Gif sur Yvette, France
christophe.leplat@cea.fr
franck.chauvat@cea.fr
jean-christophe.aude@cea.fr

[b]Génomique des Microorganismes
UMR 7238 CNRS-UPMC
15 Rue de l'Ecole de Médecine, 75006 Paris, France
raphael.champeimont@upmc.fr

**Mots clefs** : Vizualisation, Biology, Pathways, Microarrays, High throughput assays.

High-throughput *omic* technologies are now commonly used in large-scale experimental biology. The main characteristic of these *omic* approaches is that they usually produce large amounts of data. Results obtained through these analyses are mostly interpreted or assessed in terms of given hypotheses. In most cases, huge amount of results need to be transformed (*eg* using classification methods), integrated with other biological knowledge (*eg* pathways), and explored using mainstream or dedicated visualisation tools. Then, they can be meaningfully interpreted by biologists. Visualisation is crucial for an optimal understanding of the results emerging from the concerted analysis of shared material between experimental and computational researchers.

*Directed* visualisation methods [1] use *prior* knowledge in their process. In biology this knowledge is often depicted by networks. For example, Momin & *al.* [2] designed a method that combines a visualisation method and a prediction process to map transcriptomic data with predicted metabolite pools into pathways. Here we report `SVGMapping` [3], an R package to map *omic* experimental data onto custom-made templates which can be used to depict metabolic pathways, cellular structures or biological processes. `SVGMapping` allows the modification of color, opacity or shape of given graphical elements. It can be applied several times on the same template to combine various *omic* data types (*eg* protein and metabolite concentrations). This package has been designed to integrate the wealth of data generated by various strains (*eg* mutants *vs* wild-type), growth conditions (*eg* before *vs* after stress) or kinetic experiments.

**Templates:** In the `SVGMapping` framework, a template is an SVG file where shapes are specifically labeled. Labels are specified as *attributes* assigned to any kind of shape. These labels are used to pinpoint all attribute modifications (*ie* colors or opacity) to apply on the template. We have selected the SVG format for its versatility as a web-based and portable standard that can be rendered in many other graphics formats (*eg* PNG or PDF). Furthermore, javascript code can be embedded into SVG files to provide an interactive experience to the user when viewed within a compatible browser.

*Omic* **data mapping:** experimental data are provided as a numeric matrix $M$ with as many columns as conditions. Each row is labeled with a unique identifier (*eg* gene, protein or metabo-

lite). This identifier will be used to track the template shapes to modify.
For single condition experiments, $M$ values can be uniformly bound to a set of colors or a color gradient to modify the filling or stroke colors of shapes (see figure 1). For experiments with multiple conditions, one can use pie charts or colored stripes. In both cases, the filling color of each slice/stripe is set according to the $M$ value of the related shape identifiers (*ie* rows) and conditions (*ie* columns).
Besides this common usage, one can use $M$ values to alter the *opacity* or the *stroke width* of shapes. Another specific use is to simulate the filling of Erlenmeyers. In this case, $M$ values should be given as a proportion of the complete filling. This mode is of particular interest to simulate the relative concentrations of metabolites. Finally, since SVG is a web centric format, we have implemented a mechanism to add tooltips and hyperlinks (as URLs) to each shape.

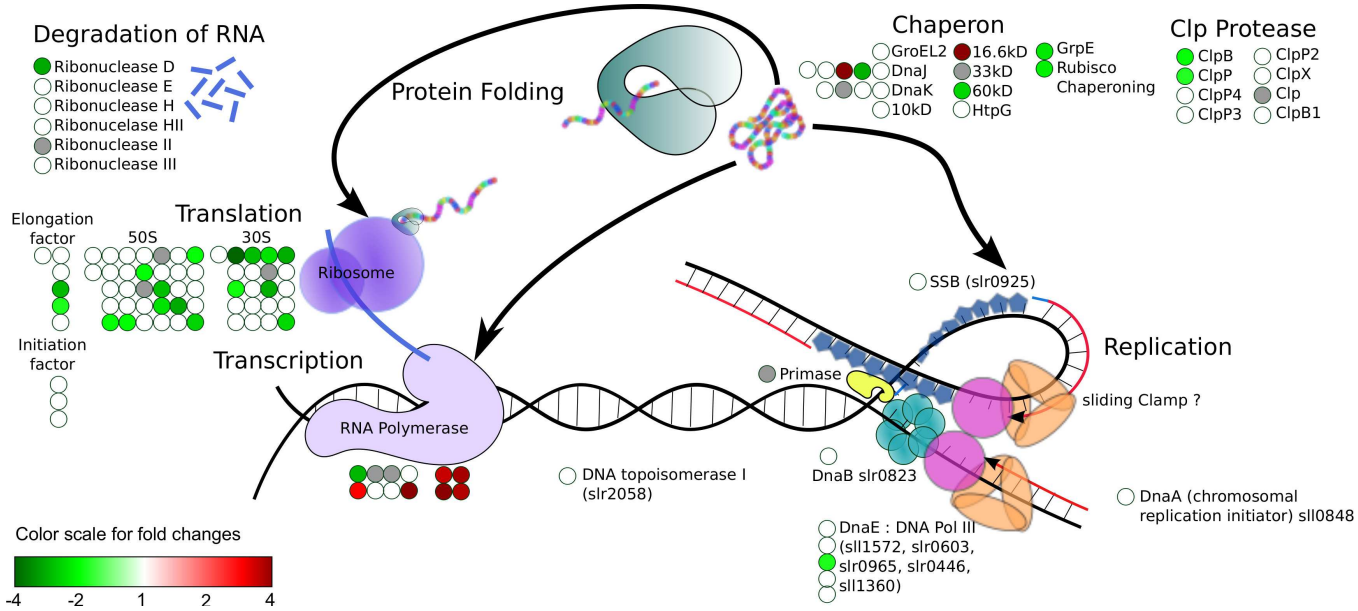


Figure 1: This figure represents DNA metabolic processes in *Synechocystis sp. PCC 6803* and illustrates a possible use of the `SVGMapping` package. Red (*resp.* green) circles depict induced (*resp.* repressed) genes involved in each biological mechanism corresponding to a microarray experiment in which cells where exposed to 3mM $H_2O_2$ for 30 minutes [4]. Matches between expression fold-changes and color levels can be obtained using the scale on the lower left of the figure. Circles filled in grey are related to probes that were not hybridized on the microarrays. Notice that many genes involved in protein translation were repressed while genes of RNA polymerase transcription were induced.

## References


[2] Keim, D. A. & *al.* (2006).Challenges in Visual Data Analysis. *Tenth International Conference on Information Visualisation (IV'06).*

[2] Momin, A. & *al.* (2011).A method for visualization of "omic" datasets for sphingolipid metabolism to predict potentially interesting differences. *The Journal of Lipid Research*, **52**(6), 1073-83

[3] http://svgmapping.r-forge.r-project.org/

[4] Houot, L. & *al.* (2007).Cadmium triggers an integrated reprogramming of the metabolism of Synechocystis PCC6803, under the control of the Slr1738 regulator. *BMC Genomics*, **8**, 350