SVGMap: configurable image browser for experimental data

Xavier Rafael-Palou[1,†], Michael P. Schroeder[1,†] and Nuria Lopez-Bigas[1,2,*]

[1]Department of Experimental and Health Sciences, Research Unit on Biomedical Informatics (GRIB), Universitat Pompeu Fabra–IMIM, Dr Aiguader 88 and [2]Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain

ABSTRACT

Summary: Spatial data visualization is very useful to represent biological data and quickly interpret the results. For instance, to show the expression pattern of a gene in different tissues of a fly, an intuitive approach is to draw the fly with the corresponding tissues and color the expression of the gene in each of them. However, the creation of these visual representations may be a burdensome task. Here we present SVGMap, a java application that automatizes the generation of high-quality graphics for singular data items (e.g. genes) and biological conditions. SVGMap contains a browser that allows the user to navigate the different images created and can be used as a web-based results publishing tool.

Availability: SVGMap is freely available as precompiled java package as well as source code at http://bg.upf.edu/svgmap. It requires Java 6 and any recent web browser with JavaScript enabled. The software can be run on Linux, Mac OS X and Windows systems.

Contact: nuria.lopez@upf.edu

1 INTRODUCTION

Modern biological experiments generally generate big datasets that need to be visualized effectively in order to interpret the results. Some available tools for genomic data analysis focus on representing huge datasets in one figure using visualization techniques like heat maps and circular plots (Krzywinski *et al.*, 2009; Perez-Llamas and Lopez-Bigas; Pritchard *et al.*, 2006).

This focus on big data should not leave out the development of tools to represent singular data, for example representing expression of a single gene. In such cases, the viewer can be provided with figures that are more straightforward to understand, like a textbook figure. Particularly, figures in SVG format, the W3C standard to design high-quality images and graphics, are well suited for the graphical representation of biological data.

With SVGMap, we present a tool that is well fitted for the visualization of biological data through the generation of individualized figures in a generic and interactive way. It is a standalone Java application and can be executed on any system with

---

**To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.*

Java 6 or higher installed. The software instantiates a web server that can be accessed through common web browsers. SVGMap can be executed on a personal computer or can be incorporated into an existing web server optionally using an external database.

2 BASIC USAGE AND IMPLEMENTATION

SVGMap requires two different kinds of inputs from the user: (i) an Scalable Vector Graphic (SVG) image file and (ii) a tab separated mapping file. Optionally, the user can also provide another tab separated file with annotations for each item (e.g. synonyms for the genes, descriptions, etc). We refer to an entity of these files as an experiment. The user is responsible to provide the SVG image for the application. It may have to be created from scratch or an existing image that was published in public domain may be reused.

Since SVG is a markup-language format, like the more popular HTML and XML, different objects (regions) in the SVG image can be annotated with region IDs. The regions are freely defined areas in the SVG image and their IDs will be used to map values of items (e.g. gene expression values) to the region in question using the mapping file. This relation can be used to dynamically map the values of a measurement to a certain region on the SVG image and color it accordingly. We use Batik SVG Toolkit to automatically generate SVG images (http://xmlgraphics.apache.org/batik).

The mapping file contains three or more columns. The first column indicates a region ID in the SVG image. The second column the item (e.g. gene name), and all further columns contain values of that item in the corresponding region for a certain condition.

The graphical user interface (GUI), the SVGMap browser, is a web server that consists of a restricted administration section and an unrestricted browser section. The administration section is designed for creating, deleting and configuring experiments. For each experiment, the user must upload an SVG image, the mapping file and optionally the annotation file. Some configurable options are available, such as selecting which type of color scale (*P*-value, linear or linear two-sided) will be used, which color range and which values of the mapping file will be selectable. More information about the experimental conditions can be represented in the title and description fields of the experiment. The unrestricted browser section lets the users browse the data in a sortable and searchable table where data to be visualized in the image can be selected. It also allows the user to adjust the scales to be used and the range and colors. After having generated a figure via the SVGMap browser, a ZIP folder is offered for download, which includes the selected data, the generated figures and the scale. The workflow is represented in Figure 1.
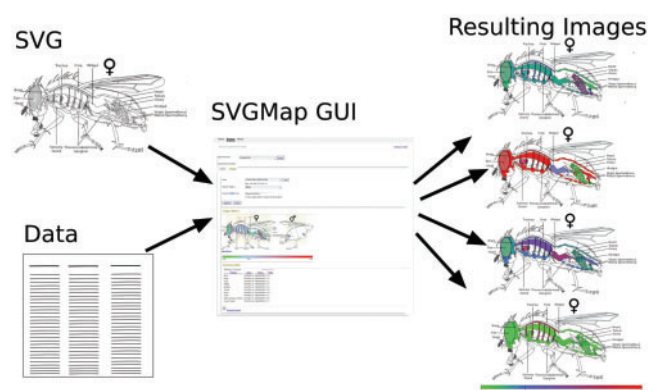
Fig. 1. From left to right: the SVG image and the data are loaded into the SVGMap via its GUI, the SVGMap browser. In the SVGMap browser, the user selects the desired data and adjusts for color and data range in order to generate the figure. Repeating this step for other selection results in more figures, each of which can be downloaded in SVG format, including the scale and the data file corresponding to the image.

Apart from the SVGMap browser, advanced users can also use the SVGMap package for batch generating data-annotated images with the command-line package, documented on the project's website.

3 DISCUSSION

The development of SVGMap and its SVGMap browser was motivated to cover the need we had in collaboration with an experimental group to support the study of gene expression in *Arabidopsis thaliana* roots, but has been programmed in a generic manner, such that it can be adapted to any other data types. Our tool helped them to study their genes of interest and gave insight in which tissue they are expressed and with which level of expression. Furthermore, it serves as well as a platform for the publication of their data and images. In this collaboration, SVGMap has been proved to be straightforward and very useful. In our opinion, SVGMap has a wide range of applications to represent biological data. To illustrate that we have prepared four diverse examples: (i) visualization of expression levels of yeast cell cycle genes in each part of the cell cycle (M,G1,S,G2) (Rustici *et al.*, 2004); (ii) visualization of copy number variation data for 10 different types of cancer from IntOGen data (Gundem *et al.*, 2010); (iii) visualization of gene expression in different *Drosophila* tissues from FlyAtlas (Chintapalli *et al.*, 2007) and (iv) visualization of cell compartments affected by transcriptomic alterations in 10 cancer types (Gundem *et al.*, 2010).

With the flexible SVG format as basis of our tool, it is important to stress that SVG images can be reused and adapted because the graphics are described by text. Any SVG that has been published in public domain, like the Wikimedia documents (http://commons.wikimedia.org), can be loaded in a SVG editor in order to add or remove objects and annotate them.

4 CONCLUSIONS

In this article, we have presented SVGMap, an intuitive tool, which aims to help researchers in the visualization and generation of high-quality images for singular data items.

The whole code of SVGMap including its examples has been released as open source and may be freely downloaded and reused. We also welcome and encourage contributors to ideas and code.

REFERENCES

Chintapalli,V.R. *et al.* (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat. Genet.*, **39**, 715–720.

Gundem,G. *et al.* (2010) IntOGen: integration and data mining of multidimensional oncogenomic data. *Nat. Methods*, **7**, 92–93.

Krzywinski,M. *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639–1645.

Perez-Llamas,C. and Lopez-Bigas,N. (2011) Gitools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS One*, **6**, e19541.

Pritchard,L. *et al.* (2006) GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics*, **22**, 616–617.

Rustici,G. *et al.* (2004) Periodic gene expression program of the fission yeast cell cycle. *Nat. Genet.*, **36**, 809–817.