



Matrix2png: a utility for visualizing matrix data

Paul Pavlidis^{1,*} and William Stafford Noble²

¹Columbia Genome Center, Columbia University, New York, USA and ²Department of Genome Sciences, University of Washington, Seattle WA, USA

Received on July 12, 2002; revised on August 27, 2002; accepted on September 2, 2002

ABSTRACT

We describe a simple software tool, 'matrix2png', for creating color images of matrix data. Originally designed with the display of microarray data sets in mind, it is a general tool that can be used to make simple visualizations of matrices for use in figures, web pages, slide presentations and the like. It can also be used to generate images 'on the fly' in web applications. Both continuous-valued and discrete-valued (categorical) data sets can be displayed. Many options are available to the user, including the colors used, the display of row and column labels, and scale bars. In this note we describe some of matrix2png's features and describe some places it has been useful in the authors' work.

Availability: A simple web interface is available, and Unix binaries are available from <http://microarray.cpmc.columbia.edu/matrix2png>. Source code is available on request.

Contact: pp175@columbia.edu

Many bioinformatics applications deal with the manipulation of matrices, and visualization of these matrices is often desirable. The application which motivated development of the tool we describe here are the data matrices handled in microarray studies, where typically rows in the matrices represent genes and columns individual samples. Rather than view a numerical 'spreadsheet', it is convenient to display microarray data in a visualization popularized in this context by Eisen *et al.* (1998), where varying expression levels are indicated in a grid of varying colors. The display of numerical values as color maps date to the 19th century have a very wide use in data visualization (see Feinberg, 1979 for review; see Tufte, 1990 and Ware, 2000 for practical issues relating to the use of color for visualization). Though originally most commonly used for display of demographic or geographic statistics on cartographic maps, color maps are generally useful for increasing the information content of multivariate data displays.

The 'Treeview' software developed by Michael Eisen for viewing expression data with hierarchical cluster-

ing dendrograms is a popular tool for the display of microarray data (Eisen *et al.*, 1998). Treeview allows interactive browsing of the data and rendering of clustering trees. However, Treeview runs only under Microsoft Windows, and outputs only Windows bitmap format files or postscript, both of which must be converted to other formats for display on web pages. In addition, Treeview permits limited control over the appearance of the image and does not provide scale bars. The limitations of Treeview and other applications we have encountered led us to develop our own image generating utility, matrix2png.

Matrix2png is implemented in the 'C' programming language and should run on a variety of platforms, though it has only been tested on Unix operating systems (Linux and Solaris). Matrix2png uses the 'gd' image production library developed by T. Boutell (<http://www.boutell.com/gd/>), and thus produces PNG (portable network graphics) format images (<http://www.libpng.org/pub/png/>). Like the better-known GIF format, PNG uses lossless compression to yield a very compact image file, without the licensing difficulties associated with the GIF format. The PNG format is widely supported by modern web browsers including Netscape and Internet Explorer, image processing programs such as Adobe Photoshop and Illustrator, and other major applications such as Microsoft Word, PowerPoint and Excel. Most users will therefore be able to use matrix2png output anywhere they might use GIF images, without any compatibility problems.

To make matrix2png available to a wider audience, we have implemented a simple web application that can be used to create images from matrices without the need to install the software locally (linked from <http://microarray.cpmc.columbia.edu/matrix2png/>). This also demonstrates one potential use for matrix2png, which is to generate images on the fly for the display in web applications. We encourage casual users to use the web interface, which gives access to the most important options. Additional options are currently accessible only from the command line, which requires installation of matrix2png on the user's computer. Matrix2png takes as input a single tab-delimited text file containing numerical values in a simple 'spreadsheet' format.

*To whom correspondence should be addressed at Russ Berrie Pavilion Room 121, 1150 St. Nicholas Ave, New York, NY 10025, USA.

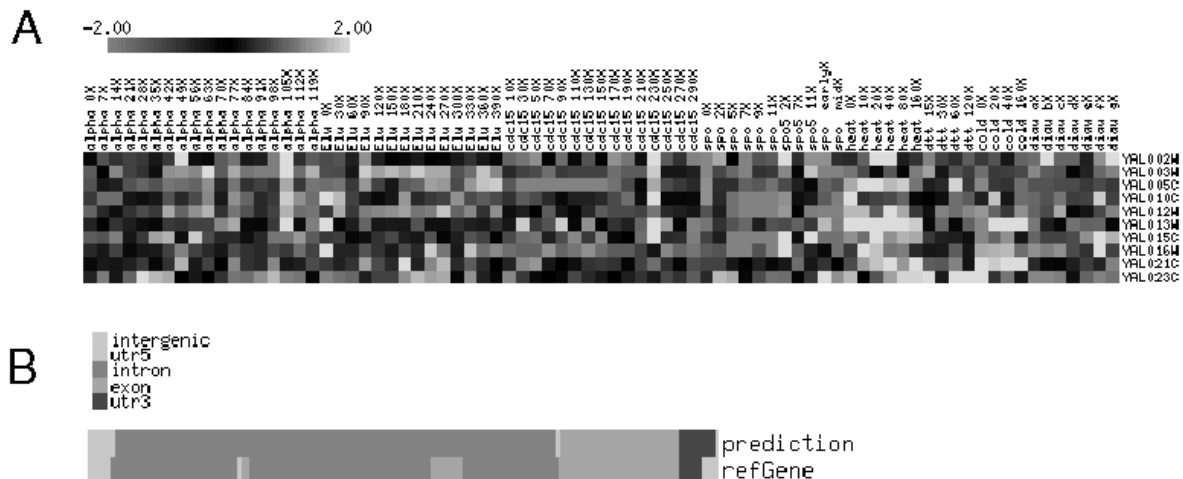


Fig. 1. (a) An image created with matrix2png from microarray data. Rows are genes, columns are samples. The values plotted are log ratios of gene expression. (b) Example of an image created using a mapping of categorical values to colors. In this case, the colors represent states in a model of a gene. This visualization helps assess the agreement of the prediction made by a gene-finding program (top line) with the true gene structure (bottom line).

Matrix2png affords considerable flexibility in image generation. Among the most important options are controls of value-to-color mapping. For continuous data (such as microarray data), the user can select from a range of preset color schemes, or choose their own color endpoints. Some of the hard-coded schemes are derived from the work of Brewer (2002), and include several which are suitable for color-blind viewers. For ‘discrete’ data (that is, where a small set of values present in the data are used to signify categorical or qualitative features, such as nucleotide bases), the user can rely on a default set of colors or define them in an auxiliary ‘discrete map’ file. The user can control image size, contrast, outlier trimming, the displayed data range, the display of text row and column labels, and, importantly, a scale bar.

We originally developed matrix2png with visualization of microarray data sets in mind (Example in Figure 1a). Examples in the existing literature of this use can be found in Pavlidis and Noble (2001) and Segal *et al.* (submitted). We soon discovered that matrix2png was useful for many other applications. For example, we have used it to visualize protein-protein interaction matrices, the outputs from hidden markov models of DNA sequences, and motif occurrences in multiple alignments of protein sequences. Figure 1b illustrates the use of matrix2png to visualize a comparison of a predicted gene structure with a reference structure. A gallery illustrating various uses of matrix2png is on our web site.

REFERENCES

Brewer,C. (2001) Color Brewer: Web tool for selecting map color

schemes, <http://www.personal.psu.edu/faculty/cia/cab38/>.
 Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **8:95**, 14863–14868.
 Feinberg,S.E. (1979) Graphical methods in statistics. *American Statistician*, **33**, 165–178.
 Pavlidis,P. and Noble,W.S. (2001) Analysis of strain and regional variation of gene expression in mouse brain. *Genome Biol.*, **2**, research0042.1–0042.15.
 Segal,N.H., Pavlidis,P., Antonesen,C.R., Maki,R.G., Noble,W.S., Woodruff,I.M., Lewis,J.I., Brennan,M.F., Houghton,A.N. and Cordon-Cardo,C. (submitted) Classification and subtype prediction of soft tissue sarcoma by functional genomics and support vector machine analysis.
 Tufte,E.R. (1990) *Envisioning Information*. Graphics Press, Cheshire CT.
 Ware,C. (2000) *Information Visualization: Perception for Design*. Morgan Kaufmann, New York.

To be balanced at final stage