DRAGON View: information visualization for annotated microarray data

Christopher M. L. S. Bouton [1,3] and Jonathan Pevsner [1, 2,]*

[1]Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD 21205, USA and [2]Department of Neurology, The Kennedy Krieger Institute, Baltimore, MD 21205, USA

ABSTRACT

Summary: The DRAGON View information visualization tools aid in the comprehensive analysis of large-scale gene expression data that has been annotated with biologically relevant information through the generation of three types of complementary graphical outputs.

Availability: The DRAGON View tools are freely available on the DRAGON web site at http://pevsnerlab. kennedykrieger.org/dragonview.htm.

Contact: pevsner@kennedykrieger.org; cbouton@jhmi.edu

INTRODUCTION

We recently reported the development of a World Wide Web (www) accessible database called 'Database Referencing of Array Genes ONline' or DRAGON (Bouton and Pevsner, 2000). The DRAGON Database can be used to annotate large lists of genes derived from gene expression analysis technologies such as cDNA microarrays. To complement the annotations provided by the DRAGON Database, we have developed the DRAGON View information visualization tools. There are three complementary tools available in DRAGON View: DRAGON Families, Order and Paths. Presently, these tools are designed for use with ratio expression data such as those derived from two-channel (e.g. Cy3 & Cy5) fluorescent microarrays. Although the DRAGON View tools have been designed in conjunction with the DRAGON Database they are distinct from the database and can be used with any ratio expression data set that has been annotated with classifying information.

Information provided by the DRAGON Database 'Annotate' tool classifies genes into groups based upon shared biological characteristics of the genes or their encoded proteins. Examples are the accession numbers from the Pfam database (URL 1, Bateman *et al.*, 2000), controlled vocabulary of keywords of the SWISS-PROT database (URL 2, Bairoch and Apweiler, 2000) and the pathway numbers from the Kyoto Encyclopedia of Genes and Genomes (KEGG; URL 3, Kanehisa and Goto, 2000).

The DRAGON Families tool uses these types of classifying information to provide a graphical display of the ratio expression values of all of the genes in a data set organized into rows based upon the classifying information provided (Figure 1a). This allows for the rapid identification of families that contain large numbers of similarly expressed genes.

DRAGON Order automatically sorts data sets by ratio expression values and then displays sets of lines, corresponding to genes, organized into rows (Figure 1b). Each row represents all of the genes in the submitted data set that are defined by a specific type of information. The genes at the far left and right ends of the rows represent the most up- and down-regulated genes in the data set and clusters of lines at one end or the other (or both) represent groups of related genes that are co-expressed.

DRAGON Paths maps the location of differentially expressed human genes onto any one of 83 KEGG cellular pathway diagrams (Figure 1c). The diagrams and text files (all files with the form *_gene.coord at URL 4) containing the x, y coordinates of every gene on each diagram are directly downloaded from KEGG. The resulting view of the data allows for the analysis of putative effects of differential regulation on protein–protein interactions in complex cellular networks.

SYSTEM, METHODS AND IMPLEMENTATION

A Dell PowerEdge 6300, dual Xeon 550 MHz processor-based server running Red Hat Linux 6.2 serves the DRAGON web site with Apache (URL 5). The DRAGON View tools are implemented as common gateway interface (.cgi) scripts written in Perl. Please refer to the DRAGON View web site (URL 6) for detailed instructions, relevant literature references, and web links.

---

*To whom correspondence should be addressed.
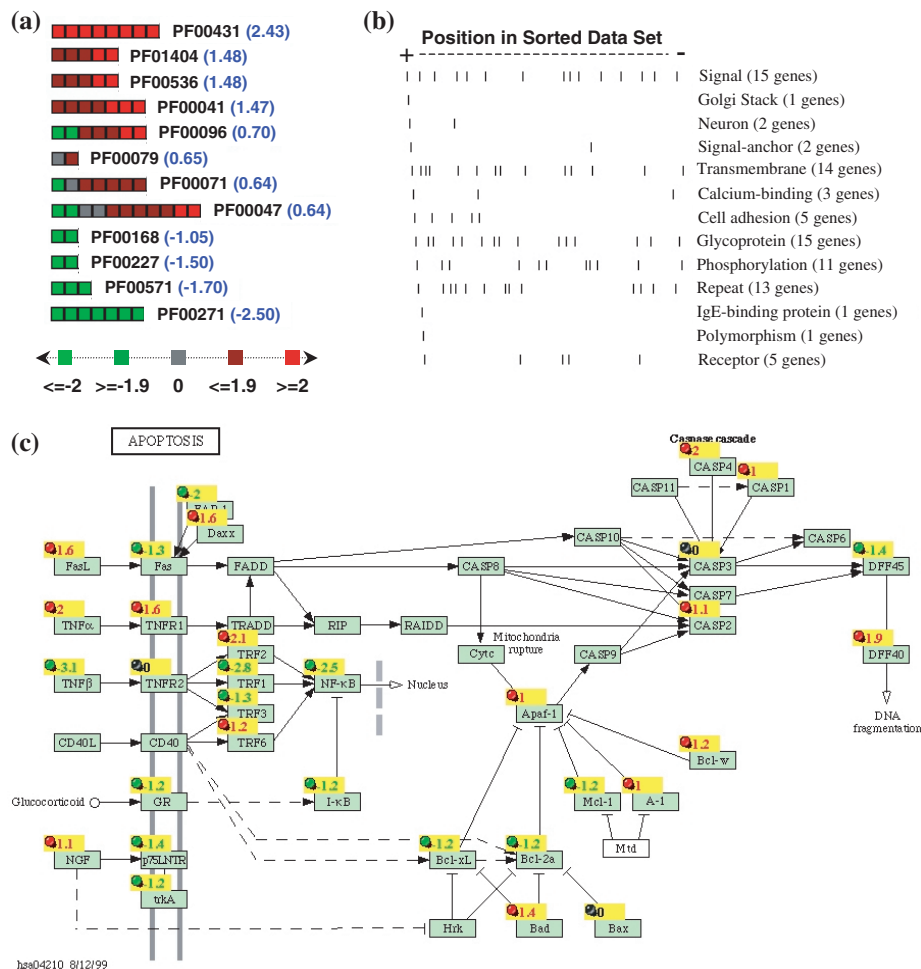[3]Present address: LION Bioscience Research Inc.

Fig. 1. Examples of the graphical outputs of the three types of DRAGON View tools. (a) DRAGON Families produces rows of green (down-regulated), red (up-regulated) and grey (unchanged) boxes (see scale for the range of ratio values represented by each color). Each box represents one gene and is hyperlinked to its corresponding UniGene entry. Each row has a type identifier to its right that is hyperlinked to its description. To the far right is the average ratio expression value for all of the genes in that family. All rows are sorted from the most up-regulated family to the most down-regulated family. (b) DRAGON Order produces rows of black lines. Each line represents one gene and its location in the row represents its position on a gene list sorted by ratio expression values. Lines at the far left represent the most up-regulated genes (+) and lines at the far right represent the most down-regulated (−). Each row's type (e.g. SWISS-PROT keywords) is listed to the right. (c) DRAGON Paths maps the location and ratio expression value of genes from the submitted gene list on to KEGG cellular pathway diagrams. A green (down-regulated), red (up-regulated) or grey (unchanged) circle followed by the ratio expression value is mapped to the upper left corner of each corresponding protein box. Each protein box is hyperlinked to its corresponding LocusLink entry.

URL REFERENCES

(1) http://www.sanger.ac.uk/Software/Pfam/

(2) http://www.expasy.ch/sprot/

(3) http://www.genome.ad.jp/kegg/

(4) ftp://kegg.genome.ad.jp/pathways/hsa/

(5) http://www.apache.org

(6) http://pevsnerlab.kennedykrieger.org/dragonview.htm

REFERENCES

Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.

Bateman,A., Birney,E., Durbin,R., Eddy,S.R., Howe,K.L. and Sonnhammer,E.L.L. (2000) The Pfam protein families database. *Nucleic Acids Res.*, **28**, 263–266.

Bouton,C.M. and Pevsner,J. (2000) DRAGON: Database Referencing of Array Genes Online. *Bioinformatics*, **16**, 1038–1039.

Kanehisa,M. and Goto,S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.