BIOINFORMATICS APPLICATIONS NOTE Vol. 20 no. 7 2004, pages 1191–1192
DOI: 10.1093/bioinformatics/bth052

ChromoViz: multimodal visualization of gene expression data onto chromosomes using scalable vector graphics

Jihoon Kim[1], Hee-Joon Chung[1], Chan Hee Park[1], Woong-Yang Park[2] and Ju Han Kim[1,3,*]

[1]Seoul National University Biomedical Informatics (SNUBI), [2]Department of Biochemistry and Molecular Biology and [3]Human Genome Research Institute, Seoul National University College of Medicine, Seoul 110-799, Republic of Korea

Received on September 9, 2003; revised and accepted on November 24, 2003
Advance Access publication February 5, 2004

ABSTRACT

Summary: ChromoViz is an R package for the visualization of microarray gene expression data, cross-species and cross-platform comparisons, as well as non-expression genomic data obtained from public databases onto chromosomes. Chromosomal visualization format is proposed for the clear decoupling of the data layer from the procedure layer and the combined visualization of genomic data from heterogeneous data sources. Visualization with Javascript-enabled scalable vector graphics enables interactive visualization and navigation of data objects on the Web.

Availability: http://www.snubi.org/software/ChromoViz/

Contact: juhan@snu.ac.kr

INTRODUCTION

Positional clusters of co-expressed genes on consecutive chromosomal locations are known in prokaryotic operons and described in several eukaryotes (Roy *et al*., 2002), including human (Lercher *et al*., 2002). Some development-related genes are physically clustered (Ramalho-Santos *et al*., 2002). Genomic DNA copy number alterations play an important role in the development and progression of cancer (Pollack *et al*., 2002) and may suggest the oncogenic potential of the involved genes (Mu *et al*., 2003). Therefore, mapping gene expression data in chromosomal order with associated annotations is a powerful tool for understanding the higher-level chromatin structures coordinating transcriptional regulation.

Recently developed tools (Sturn *et al*., 2002; Gentleman 2003, http://www.bioconductor.org/), however, have some limitations. They are monolithic, i.e. the data layer and the procedure layer are not clearly separated (Pressman, 1997). Sometimes it is not easy to directly compare the multiple features of microarray data such as multi-channel intensities and derived values for different transformations and normalizations. More importantly, it is not easy to add tracks for informative non-expression data for chromosomal analysis such as CpG islands, gene density, variations, repeats and genomic DNA copy number alterations from array CGH. Applying comparative genomic information may uncover patterns of gene expression conserved between human and mouse. Cross-platform visualization of gene expression data from different technologies (i.e. cDNA and oligonucleotide arrays) and the combined analysis of gene expression profiles from different experiments may also uncover new biological knowledge.

PROGRAM OVERVIEW

ChromoViz is a Web-enabled multi-track visualization tool written in the R-statistical language (Ihaka and Gentleman, 1996) with its output in Scalable Vector Graphics (SVG; http://www.w3.org/TR/SVG). To clearly separate the data layer from its procedure layer following the principle of Gene Finding Format (GFF; http://www.sanger.ac.uk/Software/formats/GFF/), ChromoViz uses a data format called Chromosomal Visualization Format (CVF) to store the position, value and other data associated with tracks (i.e. tab-delimited with the nine attributes, track/taxonomy/chromosome/strand/begin/end/id/value/title). A Web site, http://www.snubi.org/software/ChromoViz/, has been provided to generate CVF files with human–mouse homology tracks from accession numbers and expression levels.

For the purpose of illustration, microarray datasets for human and mouse stem cell differentiation are mapped onto human chromosomes with cytoband, mouse–human homology and cross-platform-comparison information (Fig. 1). If one uploads a list of <*id*>–<*value*> pairs (i.e. accession numbers and expression levels) from microarray experiments and inputs the <*track*> and <*taxonomy*> labels at the ChromoViz Web site, the associated position and strand information (i.e. *chromosome/strand/begin/end*) to

*To whom correspondence should be addressed.

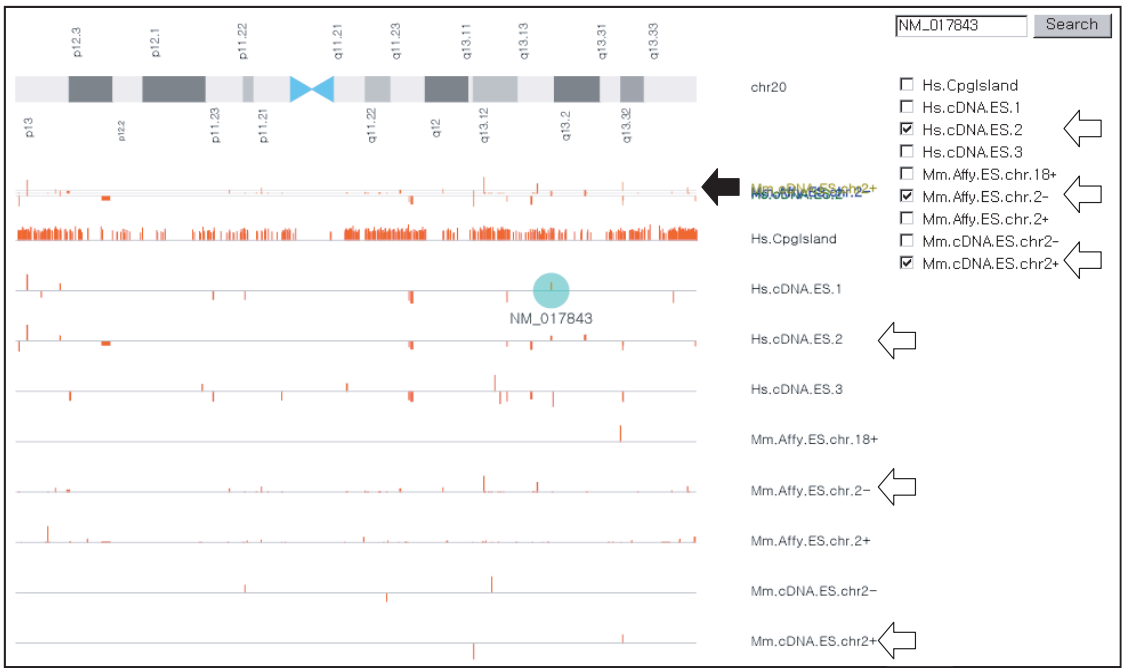

Fig. 1. ChromoViz screen shot. Microarray gene expression profiles from human and mouse stem cells are mapped onto human chromosome 20 with the parallel tracks of expression/non-expression and cDNA/oligonucleotide data. Three tracks (open arrows) selected by a user are overlaid together (closed arrow) for direct comparison. The green circle highlights a probe searched by the user.

create complete CVF output are automatically resolved by matching the accession to the pre-computed tables of GenBank, UniGene, LocusLink and UCSC Golden Path: refGene, knownGene, all_mRNA and all_est. Homologous pairs for the probes are resolved using NCBI's Homologene. NetAffyx is used for Affymetrix oligonucleotide arrays (http://www.affymetrix.com/analysis/index.affx). The track for a cytogenetic band is created by parsing NCBI's report for chromosome and cytogenetic band (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens/ maps/mapview/).

From its input, a CVF file, ChromoViz outputs Javascript-enabled SVG files that can be visualized on the client side with a Web browser using a plug-in for SVG. Javascript-enabled SVG provides ChromoViz with interactive functionalities. Users can navigate through chromosomes by zooming in and out for detailed analysis. Each object can be searched and highlighted in response to a user query. Object animation features like overlaying a track onto another for direct comparison can be added. Integrated visualization and clearly defined input/output formats of ChromoViz will benefit biological knowledge discovery.

ACKNOWLEDGEMENTS

The authors would like to thank Robert Gentleman and Peter J.Park for helpful comments. This study was supported by a grant from Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea (03-PJ1-PG3-21000-0009) and in part by a grant from Stem Cell Research Center of Frontier R&D Program (SC11021).

REFERENCES

Ihaka,R. and Gentleman,R. (1996) R: a language of data analysis and graphics. *J. Comput. Graphic Stat.*, **5**, 299–314.

Lercher,M.J., Urrutia,A. and Hurst,L.D. (2002) Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.*, **31**, 180–183.

Mu,D., Chen,L., Zhang,X., See, L.H., Koch,C.M., Yen,C., Tong,J.J., Spiegel,L., Nguyen,K.C., Servoss,A. *et al.* (2003) Genomic amplification and oncogenic properties of the KCNK9 potassium channel gene. *Cancer Cell*, **3**, 297–302.

Pollack,J.R., Sørlie,T., Perou,C.M., Rees,C.A., Jeffrey,S.S., Lonning,P.E., Tibshirani,R., Botstein,D., Borresen-Dale,A.L. and Brown,P.O. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl Acad. Sci., USA*, **99**, 12963–12968.

Pressman,R.S. (1997) *Software Engineering: A Practitioner's Approach*, 4th edn. McGraw-Hill, New York, NY, USA, pp. 341–361.

Ramalho-Santos,M., Yoon,S., Matsuzaki,Y., Mulligan,R.C. and Mettoh,D.A. (2002) "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science*, **298**, 597–600.

Roy,P.J., Stuart,J.M., Lurd,J. and Kim,S.K. (2002) Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature*, **418**, 975–979.

Sturn,A., Quackenbush,J. and Trajanoski,Z. (2002) Genesis: cluster analysis of microarray. *Bioinformatics*, **18**, 207–208.