



Next-Generation Clustered Heat Maps (NG-CHMs) for Interactive Exploration of Patterns in Omic data

Bradley M. Broom¹, Michael C. Ryan⁴, Chris Wakefield¹, Robert Brown⁴, Futa Ikeda⁴, Mark Stucky⁴, David W. Kane³, Rehan Akbani¹, James Melott¹, and John N. Weinstein^{1,2}.
University of Texas MD Anderson Cancer Center, Depts of Bioinformatics and Computational Biology¹ and Systems Biology², Houston, TX; ³Santeon, Inc., Fairfax, VA; ⁴In Silico Solutions, Falls Church, VA.

See <http://bioinformatics.mdanderson.org/TCGA/NGCHMPortal/>

THE UNIVERSITY OF TEXAS
MD Anderson
Cancer Center

MD Anderson Cancer Center
Row Centered
Z Normalized
RNA expression - Illumina (HiSeq) for 447 cancer interesting genes in 469 top KIRC samples, using the TCGA Pan-Cancer freezeV4 data.

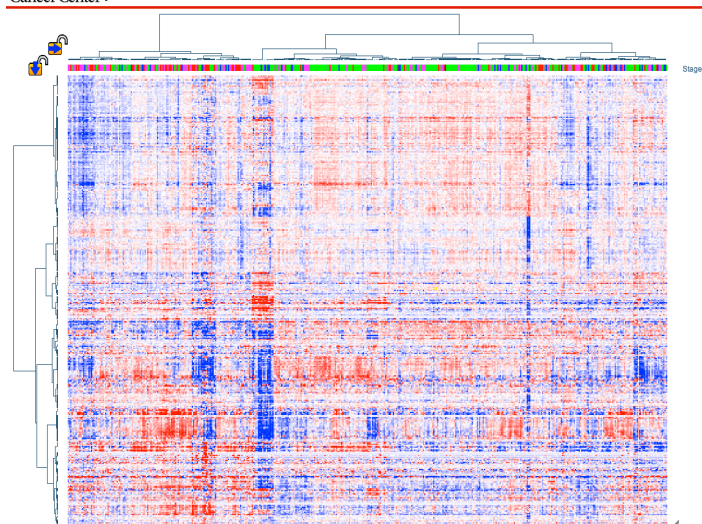


Fig. 1. A high-level data view and accompanying dendrograms. (This dataset is gene by sample mRNA expression in TCGA kidney renal clear cell carcinoma.)

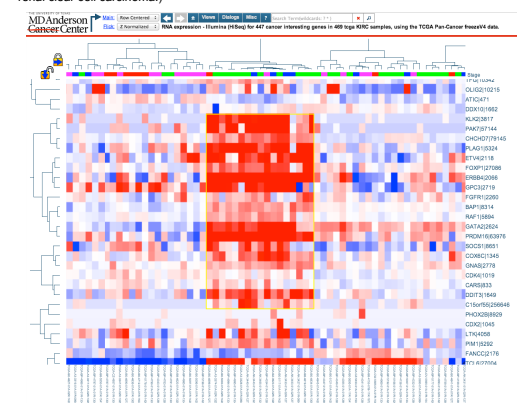


Fig. 2. The view after zooming in on a small region of highly expressed genes in a subset of samples shown in Fig. 1. Individual data values and row and column labels are now visible.

Approach

- Web-based map viewer implemented using HTML, JavaScript and CSS. No browser plugins are required.
- Four NG-CHM builder levels for different levels of expertise
 - Simple GUI-based web-generator
 - Full-featured GUI-based generator
 - R package for better integration with statistical processing
 - Command line tool

Introduction

We introduced Clustered Heat Maps (CHMs) in the early 1990's for visualizing pharmacogenomic data (Weinstein, Stern Cells, 1994) and later for integrated visualization of genomic, transcriptomic, proteomic, and functional data (Weinstein, Science, 1997). As the ubiquitous first-order way of visualizing omic data, CHMs have appeared in thousands of publications and have been used extensively to represent TCGA data. Their limitations have been summarized, along with cautions about improper use (Weinstein, Science, 2008).

One such limitation is that currently popular approaches to the generation of CHMs don't scale well. They produce fundamentally static images. It's difficult to use a single representation for both the entirety of a large data set and details of individual data points. What we really wanted was a visualization tool suited to dynamic exploration of large omic data sets as a whole and also in detail. We therefore initiated the Next-Generation CHM project, focusing it on TCGA data. The new web-based visualization tool brings new levels of scalability and interactivity to this popular visualization paradigm.

<http://bioinformatics.mdanderson.org/main/NG-CHM-Overview>

Goals

- Interactive
- Customizable
- Production-quality graphics to meet journal publication requirements
- Scalable

Other Features

- Maintains and reports detailed metadata to aid reproducibility
- Fluidly navigates from summary views of large data sets down to a detailed level
- Allows toggling between different data layers
- Decoupled from specific clustering algorithms and implementations

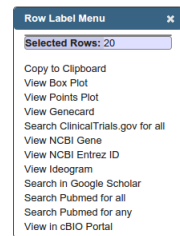


Fig. 3. The navigator shows your overall context as you zoom in and navigate around the data.

Fig. 4. Context-specific menus link to content-specific external resources

Context-specific menus

- Configurable menus let map makers link the data in the map with external resources and provide other operations
- Based on an easily extensible, type-based database:
 - Map builder specifies types of axes
 - Appropriate link-outs automatically added to resulting NG-CHM

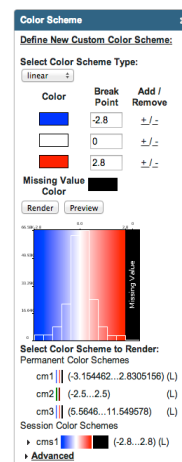


Fig. 5. Customize the color scheme with multiple breakpoints

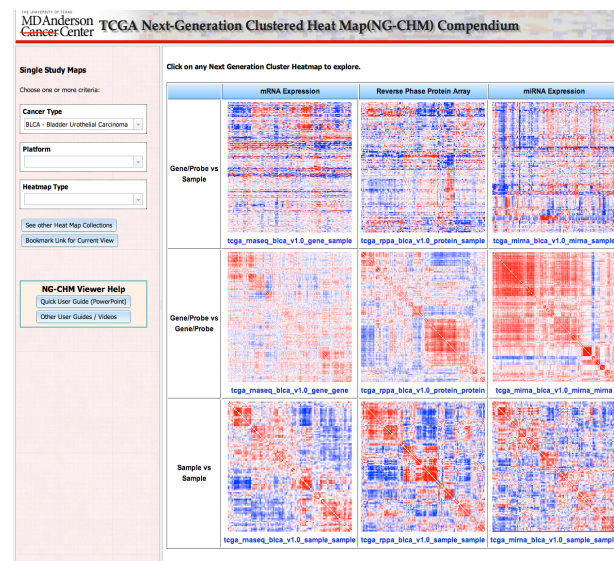


Fig. 6. We created a compendium of 667 NG-CHMs showing mRNA expression, protein abundance, miRNA expression, and methylation status for 447 cancer interesting genes for tumor types included in the TCGA PanCancer analysis. Selecting one or more criteria in the left panel will display a matrix of heat map images in the right panel. Clicking on any will open the corresponding 'Next-Generation' Clustered Heat Map. The compendium is publicly available at <http://bioinformatics.mdanderson.org/TCGA/NGCHMPortal/>.

Try It

We provide a test server running the simple GUI-based web-generator. You can use example data or upload your own data at <http://bioinformatics.mdanderson.org/testchm/>

Availability

You can freely download the NG-CHM system for use either on your own server or in the cloud. We provide it as Docker container images: <http://bioinformatics.mdanderson.org/main/NG-CHM-Docker>

Conclusions

Our CHM tool offers new levels of scalability and interactivity necessary for exploring large data sets. The JavaScript-based approach works well to balance the competing requirements of high-level and detailed views of very large heat maps, and users can fluidly navigate between the two. There are suitable interfaces to incorporate a wide range of interactions and annotations. CHM is customizable to support many different data types.

Support

We are grateful for the support of ITCR under grant U24CA199461 and cloud pilot supplement. Also supported by TCGA under grant U24CA143883. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. Also supported in part by a gift from the H.A. & Mary K. Chapman Foundation and a grant from the Michael & Susan Dell Foundation.

References

- Weinstein, J.N. ... Paul, K.D. Stem Cells 12; 13, 1994.
- Weinstein, J.N. ... Paul, K.D. Science 275;343, 1997.
- Myers, T.G. ... Weinstein, J.N. Electrophoresis 18; 467, 1997.
- Eisen, M.B. ... Botstein, D. Proc. Natl. Acad. Sci. U.S.A. 14863, 1998.
- Golub, T.R. ... Lander, E.S. Science 286; 531, 1999.
- Ross, D.T. ... Brown, P.A. Nature Genetics 24; 227, 2000.
- Scherf, U. ... Weinstein, J.N. Nature Genetics 24; 236, 2000.
- Zeeberg, B.R. ... Weinstein, J.N. BMC Bioinformatics 6; 168, 2005.
- Weinstein, J.N. Science 319; 1772, 2008.