bioRxiv preprint doi: https://doi.org/10.1101/034694; this version posted January 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

BIOINFORMATICS

Vol. 00 no. 00 2016
Pages 1–2

MicroScope: real-time statistical analytics platform and visualization engine for gene expression heatmaps

Bohdan B. Khomtchouk [1,*], Vytas Dargis-Robinson [2], James R. Hennessy [3] and Claes Wahlestedt [1*]

[1]University of Miami Miller School of Medicine, Center for Therapeutic Innovation and Department of Psychiatry and Behavioral Sciences, 1120 NW 14th Street Suite 1463, Miami, FL 33136, USA
[2]University of Miami Miller School of Medicine, Department of Microbiology and Immunology, 1600 NW 10th Ave, Miami, FL 33136
[3]University of Miami Miller School of Medicine, Department of Computer Science, 1365 Memorial Drive, Coral Gables, FL 33146

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: Large gene expression heatmaps (≥ 200 genes) are difficult to read: while general expression level patterns may be observed via hierarchical clustering, the identities of specific gene names become entirely unreadable at large scale. More importantly, current state-of-the-art heatmaps are entirely static, require programming skills to produce, and do not include any statistical significance information indicating the identities of differentially expressed genes.

Results: We create a user-friendly web app to create heatmaps that are both dynamic and interactive. A few mouse clicks allows a user to generate a heatmap based on an input file and proceed to dynamically navigate to any sector of the heatmap via zooming in and out to a specific region, cluster, or even single gene. In addition to visual magnification, MicroScope also allows users to analytically perform real-time statistical analyses with simple button clicks. Furthermore, MicroScope allows users to hover the mouse pointer over any specific gene to show precise expression level details, as well as create row and column dendrograms with colored branches using simple one-click features. As such, MicroScope presents a significant advance in heatmap visualization technology over current standard protocols.

Availability and implementation: MicroScope is hosted online as an R Shiny web app based on the D3 JavaScript library: https://microscope.shinyapps.io/microscope. The methods are implemented in R, and are available as part of the MicroScope project at: https://github.com/Bohdan-Khomtchouk/Microscope.

Contact: b.khomtchouk@med.miami.edu

1 INTRODUCTION

Currently existing heatmap software produce static heatmaps [1-5], with no features available to dynamically interact with and explore the landscape of a heatmap. Such a feature would allow the user to interact with the data in a visual manner in real-time, thereby allowing for a deeper data exploration experience. An interactive, non-reproducible heatmap tool was previously employed in the study of the transcriptome of the *Xenopus tropicalis* genome [6]. However, no open-source, dynamic and interactive heatmap software has yet been proposed to explore an arbitrary user-specified input dataset. Likewise, currently existing heatmap software do not provide built-in, user-friendly statistical analysis features.

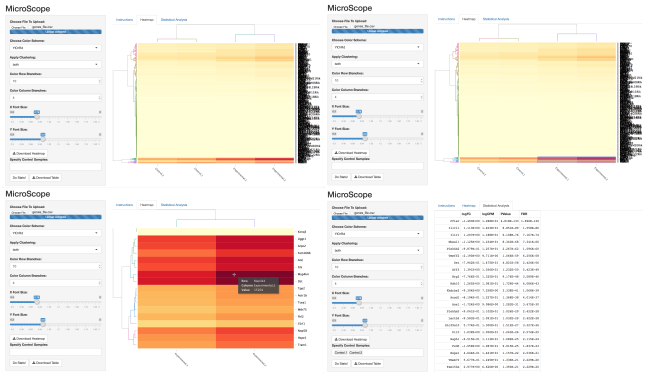

Fig. 1. MicroScope user interface showcasing the magnification feature as well as features such as: file input, multiple heatmap color schemes, hierarchical clustering, row/column dendrogram branch coloring, row/column font size, statistical analysis, and heatmap/statistical table download utilities.

2 APPROACH

In this paper, we leverage R's d3heatmap [7], shiny [8], htmlwidgets [9], and RColorBrewer [10] libraries to create a dynamic, interactive heatmap web app called MicroScope, in recognition of the software's magnification utility (described in Results section). We employ the Bioconductor package edgeR [11] to create a one-click, built-in, user-friendly statistical analysis feature that provides differential expression analysis of gene expression data. This

*to whom correspondence should be addressed

© Khomtchouk et al. 2016.

bioRxiv preprint doi: https://doi.org/10.1101/034694; this version posted January 5, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

supplies the user with rank-based information about nominal p-value, false discovery rate, fold change, and counts per million in order to establish which specific genes in the heatmap are differentially expressed to a high degree of statistical significance. All heatmaps and statistical analyses are generated within the R environment on the server-side, thereby completely obviating the need for any programming skills on the client-side.

3 RESULTS

MicroScope is an R Shiny and JavaScript (D3.js) software program designed to produce dynamic, interactive heatmaps in a web browser. Figure 1 shows the UI in action. MicroScope allows the user to magnify any portion of a heatmap by a simple click-and-drag feature to zoom in, and a click-once feature to zoom out. MicroScope is designed with large heatmaps in mind (e.g., gene expression heatmaps with thousands of genes), where individual entries quickly become unreadable as more and more add up. However, MicroScope allows you to repeatedly zoom in to any sector of the heatmap to investigate a region, cluster, or even a single gene. You can scroll up and down the page of your web browser to see more genes than automatically fit your window. MicroScope also allows you to hover the mouse pointer over any specific gene to show precise expression level details. Some of the user-friendly features of MicroScope include:

- User-specified file input widget
- Multiple heatmap color schemes widget
- Hierarchical clustering widget
- Row/column dendrogram branch coloring widget
- Row/column font size widget
- Statistical analysis widget
- Heatmap download widget
- Statistical analysis table download widget

4 CONCLUSION

We provide access to a user-friendly web app designed to produce dynamic and interactive heatmaps within the R programming environment, without any prerequisite programming skills required of the user. Our software tool aims to enrich the genomic data exploration experience by allowing for the ability to magnify any portion of a gene expression heatmap and thereby exploring specific gene regions along with their respective expression level details. Coupled with a built-in analytics platform to pinpoint statistically significant differentially expressed genes, MicroScope presents a significant advance in heatmap visualization technology over current standard protocols.

ACKNOWLEDGEMENTS

BBK wishes to acknowledge the financial support of the United States Department of Defense (DoD) through the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program: this research was conducted with Government support under and awarded by DoD, Army Research Office (ARO), National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a.

REFERENCES

[1]Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J: *TM4: a free, open-source system for microarray data management and analysis.* Biotechniques. 2003, 34 (2): 374-378.

[2]Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP: *GenePattern 2.0.* Nat Genet. 2006, 38 (5): 500-501. 10.1038/ng0506-500.

[3]Qlucore Omics Explorer. [http://www.qlucore.com]

[4]Gould J: GENE-E software hosted at the Broad Institute. [http://www.broadinstitute.org/cancer/software/GENE-E/]

[5]Khomtchouk BB, Van Booven DJ, Wahlestedt C: *HeatmapGenerator: high performance RNAseq and microarray visualization software suite to examine differential gene expression levels using an R and C++ hybrid computational pipeline.* Source Code for Biology and Medicine. 2014, 9 (1): 1-6.

[6]Tan MH, Au KF, Yablonovitch AL, Wills AE, Chuang J, Baker JC, Wong WH, Li JB: *RNA sequencing reveals a diverse and dynamic repertoire of the Xenopus tropicalis transcriptome over development.* Genome Research. 2013, 23: 201-216. [http://jason.chuang.ca/research/xenopus/ensembl.html]

[7]Cheng J, Galili T, RStudio Inc, Bostock M, Palmer J: *d3heatmap: Interactive Heat Maps Using 'htmlwidgets' and 'D3.js'*. 2015. R package version 0.6.1. [https://cran.r-project.org/web/packages/d3heatmap/index.html]

[8]Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J, RStudio, jQuery Foundation, jQuery contributors, jQuery UI contributors, Otto M, Thornton J, Bootstrap contributors, Twitter Inc, Farkas A, Jehl S, Petre S, Rowls A, Gandy D, Reavis B, Kowal KM, es5-shim contributors, Ineshin D, Samhuri S, SpryMedia Limited, Fraser J, Gruber J, Sagalaev I, R Core Team: *shiny: Web Application Framework for R*. 2015. R package version 0.12.2. [https://cran.r-project.org/web/packages/shiny/index.html]

[9]Vaidyanathan R, Xie Y, Allaire JJ, Cheng J, Russell K, RStudio: *htmlwidgets: HTML Widgets for R*. 2015. R package version 0.5. [https://cran.r-project.org/web/packages/htmlwidgets/index.html]

[10]Neuwirth E: *RColorBrewer: ColorBrewer Palettes*. 2014. R package version 1.1-2. [https://cran.r-project.org/web/packages/RColorBrewer/index.html]

[11]Robinson MD, McCarthy DJ, Smyth GK: *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics. 2010, 26: 139-140.