Khomtchouk *et al.*

MicroScope: comprehensive genome analysis software suite for gene expression heatmaps

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

[*]Correspondence:
b.khomtchouk@med.miami.edu
[1]Center for Therapeutic Innovation and Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, 1120 NW 14th ST, Miami, FL, USA 33136
Full list of author information is available at the end of the article

Abstract

We propose a user-friendly, comprehensive genome software suite for the interactive visualization and analysis of gene expression heatmaps, including integrated features to support: statistical analysis, gene ontology analysis, and dynamic network visualization of differentially expressed genes directly from a heatmap.

MicroScope is hosted online as an R Shiny web application 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.

Introduction

Most currently existing heatmap software produce static heatmaps (Saeed et al. 2003, Reich et al. 2006, http://www.qlucore.com, GENE-E, Khomtchouk et al. 2014), without features that allow the user to dynamically interact with, explore, and analyze the landscape of a heatmap. Such features would allow the user to engage the heatmap data in a visual and analytical manner while in real-time, thereby allowing for a deeper, quicker, and more comprehensive data exploration experience. An interactive, non-reproducible heatmap tool was previously employed in the study of the transcriptome of the *Xenopus tropicalis* genome (Tan et al. 2013). Likewise, manual clustering of dot plots depicting RNA expression is an integral part of the Caleydo data exploration environment (Turkay et al., 2014). Chemoinformatic-driven clustering can also be toggled in the user interface of Molecular Property Explorer (Kibbey and Calvet, 2005). Also, prior software concentrated primarily in hierarchical clustering and searching gene texts for substrings, and serial analysis of genomic data, with no features to conduct gene ontology or network analysis (Saldanha 2004, Caraux and Pinloche 2005, Wu et al. 2010). No free, open-source heatmap software has yet been proposed to explore heatmaps at multiple levels of genomic analysis and interactive visualization capacity. Specifically, we propose a user-friendly genome software suite designed to handle dynamic, on-the-fly JavaScript visualizations of gene expression heatmaps as well as their respective statistical analysis, gene ontology analysis, and network analysis of differentially expressed genes.

MicroScope leverages the cumulative utility of R's d3heatmap (Cheng et al. 2015), shiny (Chang et al. 2015), htmlwidgets (Vaidyanathan et al. 2015), RColorBrewer (Neuwirth, 2014), dplyr (Wickham et al. 2015), data.table (Dowle et al. 2015), goseq (Young et al. 2010), GO.db (Carlson, 2015), and networkD3 (Gandrud et

al. 2015) libraries to create an integrative web browser-based software experience requiring absolutely no programming experience of the user, or even the need to download R on a local computer. MicroScope employs the Bioconductor package edgeR (Robinson et al. 2010) to create a one-click, built-in, user-friendly statistical analysis feature that provides differential expression analysis of gene expression data. This 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 with a high degree of statistical significance. This information, in turn, is used to investigate the top gene ontologies of differentially expressed genes, which can then be conveniently visualized as interactive network graphics.

Results

Figure 1 shows the MicroScope user interface (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 gene expression heatmaps in mind, where individual gene labels overlap and render the text unreadable. However, MicroScope allows the user to repeatedly zoom in to any sector of the heatmap to investigate a region, cluster, or even a single gene. MicroScope also allows the user to hover the mouse pointer over any specific gene to show gene name, expression level, and column ID.

One of the user-friendly features within MicroScope is that it is responsive to the demands asked of it by the user. For example, gene ontology analysis buttons are not provided in the UI until a user both generates a heatmap and runs statistical analysis on its contents, both of which constitute prerequisite steps required prior to conducting a successful gene ontology analysis. In other words, MicroScope is user-responsive in the sense that it automatically unlocks new features only as they become needed when the user progresses through successive stages in the software. Furthermore, MicroScope automatically provides short and convenient written guidelines directly in the UI to guide the user on the next steps in the usage of the software. As such, complex analytical operations can be performed by the user in a friendly, step-by-step fashion, each time facilitated by the help of the MicroScope software suite, which adjusts to the needs of the user and provides written guidelines on the next steps to pursue.

Some of MicroScope's initial login features include:

- User-specified file input widget
- log_2 data normalization widget
- Multiple heatmap color schemes widget
- Hierarchical clustering widget
- Row/column dendrogram branch coloring widget
- Row/column font size widget
- Heatmap download widget
- Variety of other download widgets to give users a brief look ahead

After a user inputs a file, a heatmap is automatically produced in the Heatmap panel and a Specify Control Samples widget automatically appears, in preparation for statistical analysis. Details about this widget and the ensuing statistical analysis

(Figure 2) are provided in-depth for users in the Instructions panel of the Micro-Scope software. Following the successful completion of the statistical analysis, a user is automatically supplied with five more UI widgets:

- Genome database widget
- Number of top gene ontologies widget
- Gene ontology stratification widget
- Gene ontology p-value cutoff widget
- Gene ontology FDR cutoff widget

Specifying values for these features and clicking the Do Gene Ontology Analysis button returns a list of the top gene ontologies according to the exact specifications set by the user (Figure 3). The successful completion of this step can be followed up by running a network analysis on the top gene ontologies, thereby generating network graphics corresponding to the number of top gene ontologies requested by the user (Figure 4). Nodes represent either gene names or gene ontology identifiers, and links represent direct associations between the two entities. In addition to serving as a visualization tool, this network analysis capability automatically identifies differentially expressed genes that are present within each top gene ontology, information not readily available by running gene ontology analysis alone. By immediately extracting the respective gene names from each top gene ontology category, MicroScope's network analysis features serve to aid the biologist in identifying the top differentially expressed genes in the top gene ontologies. Figure 5 compares interactive network visualizations of the top two gene ontologies, thereby demonstrating the immediate responsiveness of MicroScope's network graphics to user-specified settings.

Conclusion

We provide access to a user-friendly web application 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 providing a variety of complementary visualization and analysis features to investigate gene expression heatmaps. Coupled with a built-in analytics platform to pinpoint statistically significant differentially expressed genes, a built-in gene ontology platform to categorize the top gene ontologies, and a built-in network analysis platform to visualize gene ontology categories at the gene-specific level, MicroScope presents a significant advance in heatmap technology over currently available software.
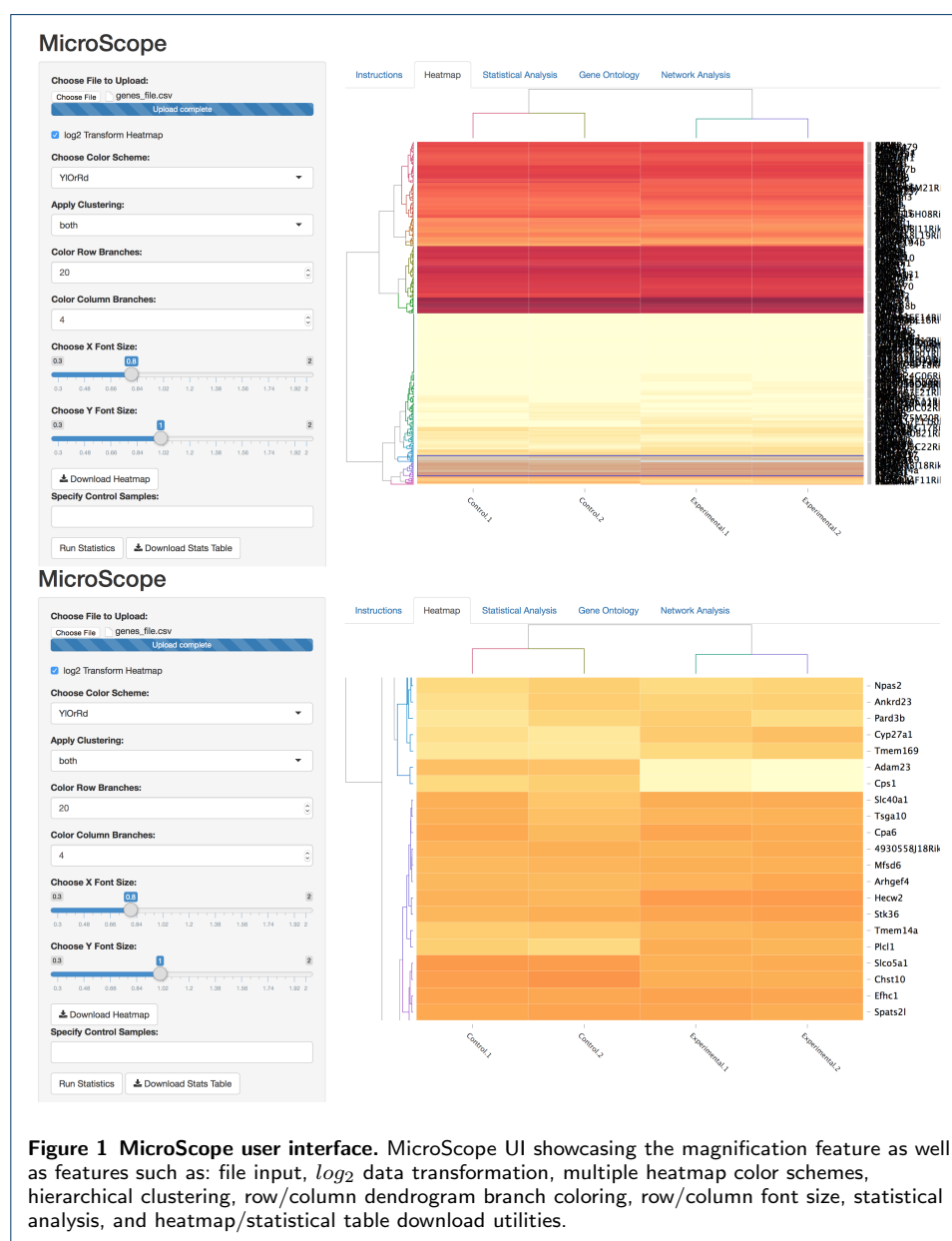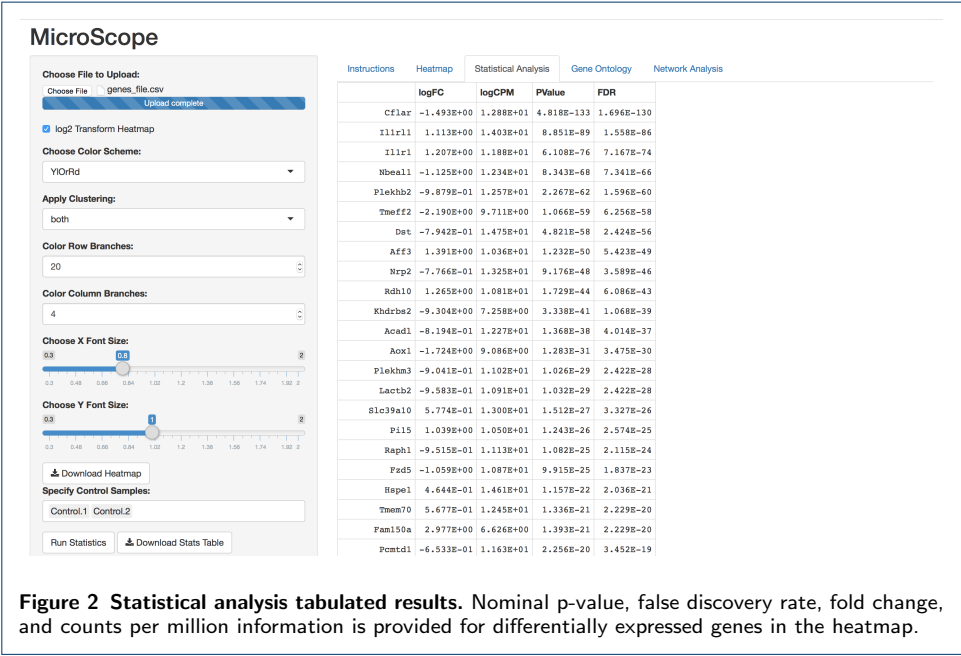
Author details
[1]Center for Therapeutic Innovation and Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, 1120 NW 14th ST, Miami, FL, USA 33136. [2]Department of Microbiology and Immunology, University of Miami Miller School of Medicine, 1600 NW 10th Ave, Miami, FL, USA 33136. [3]Department of Computer Science, University of Miami, 1365 Memorial Drive, Coral Gables, FL, USA 33146.

References

1. Caraux G, Pinloche S: *Permutmatrix: A Graphical Environment to Arrange Gene Expression Profiles in Optimal Linear Order.* Bioinformatics. 2005, 21: 1280-1281.
2. Carlson M: *GO.db: A set of annotation maps describing the entire Gene Ontology.* 2015. R package version 3.2.2.
3. 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.
4. 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.
5. Dowle M, Srinivasan A, Short T, Lianoglou S, Saporta R, Antonyan E: *data.table: Extension of Data.frame.* 2015. R package version 1.9.6.
6. Gandrud C, Allaire JJ, Russell K, Lewis BW, Kuo K, Sese C, Ellis P, Owen J, Rogers J: *networkD3: D3 JavaScript Network Graphs from R.* R package version 0.2.8.
7. Gould J: *GENE-E software hosted at the Broad Institute.*
8. 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.
9. Kibbey C, Calvet A: *Molecular Property eXplorer: a novel approach to visualizing SAR using tree-maps and heatmaps.* J Chem Inf Model. 2005, 45 (2): 523-532.
10. Neuwirth E: *RColorBrewer: ColorBrewer Palettes.* 2014. R package version 1.1-2.
11. Qlucore Omics Explorer: *The D.I.Y Bioinformatics Software.*
12. Robinson MD, McCarthy DJ, Smyth GK: *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics. 2010, 26: 139-140.
13. 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.
14. 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.
15. Saldanha AJ: *Java Treeview – extensive visualization of microarray data.* Bioinformatics. 2004, 20 (17): 3246-3248.
16. 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.
17. Turkay C, Lex A, Streit M, Pfister H, Hauser H: *Characterizing cancer subtypes using dual analysis in Caleydo StratomeX.* IEEE Comput Graph Appl. 2014, 34 (2): 38-47.
18. Vaidyanathan R, Xie Y, Allaire JJ, Cheng J, Russell K, RStudio: *htmlwidgets: HTML Widgets for R.* 2015. R package version 0.5.
19. Wickham H, Francois R, RStudio: *dplyr: A Grammar of Data Manipulation.* 2015. R package version 0.4.3.
20. Wu HM, Tien YJ, Chen CH: *GAP: A Graphical Environment for Matrix Visualization and Cluster Analysis.* Computational Statistics and Data Analysis. 2010, 54: 767-778.
21. Young MD, Wakefield MJ, Smyth GK, Oshlack A: *Gene ontology analysis for RNA-seq: accounting for selection bias.* Genome Biology. 2010, 11: pp. R14.

Figures

Figure 1 MicroScope user interface. MicroScope UI showcasing the magnification feature as well as features such as: file input, log_2 data transformation, multiple heatmap color schemes, hierarchical clustering, row/column dendrogram branch coloring, row/column font size, statistical analysis, and heatmap/statistical table download utilities.

Figure 2 Statistical analysis tabulated results. Nominal p-value, false discovery rate, fold change, and counts per million information is provided for differentially expressed genes in the heatmap.
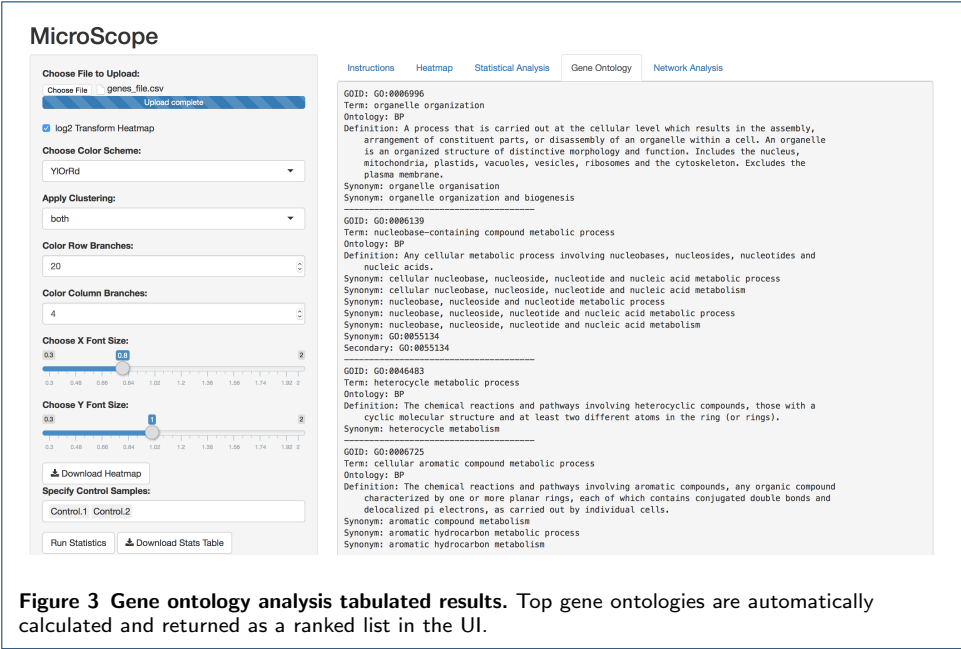
Figure 3 Gene ontology analysis tabulated results. Top gene ontologies are automatically calculated and returned as a ranked list in the UI.

Figure 4 Network graphics visualizations of top gene ontologies. Differentially expressed gene names are automatically computed during the network analysis of user-specified gene ontology settings.

Figure 5 Network visualizations of top ranked gene ontology vs. top two ranked gene ontologies. Comparison of dynamic and interactive network graphics at various user-specified gene ontology settings (e.g., 'Choose How Many Top Gene Ontologies to Display' panel in the UI).