A Web-based Visual Analytics Application for Biological Networks

M. Krone[1,2], A. Dräger[2,3,4], E. Cobanoglu[1,3], M. O. Harke[1,3], M. Hoene[5], C. Weigert[5,6,7], and R. Lehmann[5,6,7]

[1] Big Data Visual Analytics in Life Sciences, Department of Computer Science, University of Tübingen, Germany
[2] Institute for Bioinformatics and Medical Informatics (IBMI), University of Tübingen, Germany
[3] Computational Systems Biology of Infection and Antimicrobial-Resistant Pathogens, Department of Computer Science, University of Tübingen, Germany
[4] German Center for Infection Research (DZIF), partner site Tübingen, Germany
[5] Institute for Clinical Chemistry and Pathobiochemistry, University of Tübingen, Germany
[6] Institute for Diabetes Research and Metabolic Diseases of the Helmholtz Zentrum München at the University of Tübingen, Germany
[7] German Center for Diabetes Research (DZD), Oberschleissheim, Germany

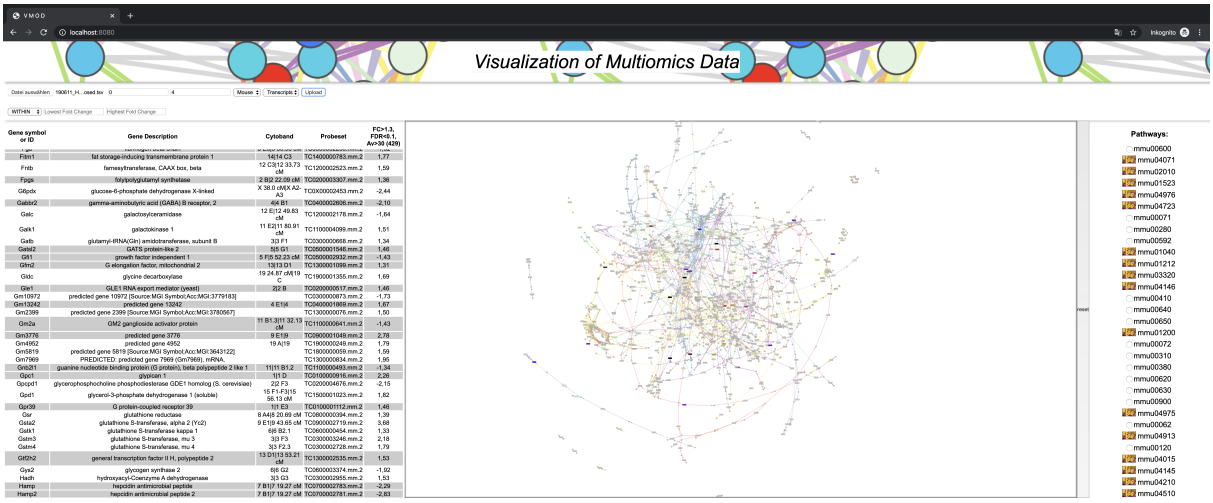

Figure 1: *Overview of our web-based visual analytics application. The table to the left lists information about all genes in the currently loaded data set (ID, description, cytoband, fold change). The list to the right shows all pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [KG00] database containing at least one of these genes. The central graph view shows an overview of the whole network of pathways. Our application, which is still work in progress, supports different ways to explore the data by visualizing details on demand.*

Abstract

Modern high-throughput methods enable rapidly obtaining transcriptomics data, which includes information about the expression rate of genes. The expression rates are usually given as fold change, which describes the over- or under-expression of each gene. Each gene can be part of one or more biological pathways. A pathway models the interactions between molecules in an organism that lead to a particular chemical change. Consequently, many applications in medical research need to analyze the impact of gene expression changes on the biological pathways of an organism. It allows concluding diseases or other conditions of the organism. We present a web-based visual analytics application that facilitates exploring the network of biological pathways corresponding to a given set of genes. The network is constructed from pathways derived from an external database. Users can interactively zoom and filter the network and get details on demand. Our application is currently work in progress and is developed in close collaboration with medical researchers. In subsequent steps, we strive to add more features, such as the ability to compare data from different individuals or to visualize time series data. Furthermore, we want to extend our application to visualize not just transcriptomics but multi-omics data.

CCS Concepts

• ***Human-centered computing*** → *Visualization systems and tools;* • ***Applied computing*** → *Systems biology; Transcriptomics;*

DOI: 10.2312/eurp.20201124

1. Introduction

Research in a clinical environment today is confronted with large amounts of multi-level high-throughput "omics" data, which are often obtained in a widely automated fashion. Linking the diverse data back to the context of molecular interactions and cellular processes remains challenging, despite considerable efforts taken over the past decades [JYS*16, HJ18, BYB*20]. Data protection and access restriction to computing devices constitute further problems in clinics where the execution of third-party software is not permitted.

We primarily focus on transcriptomics data, i.e., information about the expression rate of genes. The expression rates are usually given as fold change, which describes the over- or under-expression of each gene and, consequently, the protein that is encoded by this gene. Each gene can be part of one or more biological pathways. A pathway models the interactions between molecules in an organism that lead to a particular chemical change. KEGG [KG00, KSF*19, Kan19] is a database of biological pathways.

We present a web-based visual analytics application that facilitates exploring the network of biological pathways, corresponding to a given set of genes directly, in any web-browser. The analysis of biological pathways is inherent to many biomedical applications, e.g., in personalized medicine, where it can lead to valuable insights about deregulations or malfunctions on a microbiological as well as systems biological level for a specific patient or a group of patients that exhibit the same symptoms. To this end, many visualization tools and approaches for biological networks have been presented [GOB*10, MPB*19]. Various desktop applications are build on Cytoscape [SMO*03], an open-source software platform for visualizing networks (e.g., MODAM [EJP11]). More recent applications are often web-based [FLH*16, GDTC19] since it lowers the barrier of entry for the users [RCM*20].

2. User Requirements & Application Design

Our visual analysis application is developed in close collaboration with medical experts. Together with our project partners from the university hospital, we compiled an initial list of requirements (R1-4) for the application:

R1 Visualize the pathways that are affected by the regulated genes.
R2 Preserve the known KEGG layout of the pathway if possible
R3 Support the comparison of different measurements (either from the same individual or from different ones).
R4 The tool should be browser-based (no installation required).

The resulting application, which is still work in progress, is shown in Fig. 1. It takes a list of genes and their fold changes as input. Users can upload their own data as a CSV file that contains for all genes the respective gene ID, the fold change, and optionally an arbitrary number of columns with metadata (e.g., gene description, cytoband, and probeset; see Fig. 1). Our application allows the user to define the columns with the gene name and the fold change. The only other required information is the organism from which the samples were taken (selected via a dropdown menu). For each of the genes, it queries KEGG for all pathways containing this gene. These requests are executed by the a server, which was implemented in JavaScript using the popular *node.js* library. The server also merges all individual pathways into a vast network of pathways. This network is then sent back to the client. Our client app runs in the browser (R4) and uses D3 [BOH11] to generate the website and to visualize the network as a node-link diagram using a force-directed layout (Fig. 1). The gene and compound identifiers label the nodes. Rectangular nodes represent genes, and circular nodes denote compounds. The edges (representing interactions) are colored by the pathway to which they belong. This allows the user to spot interactions that also occur in another pathway quickly.

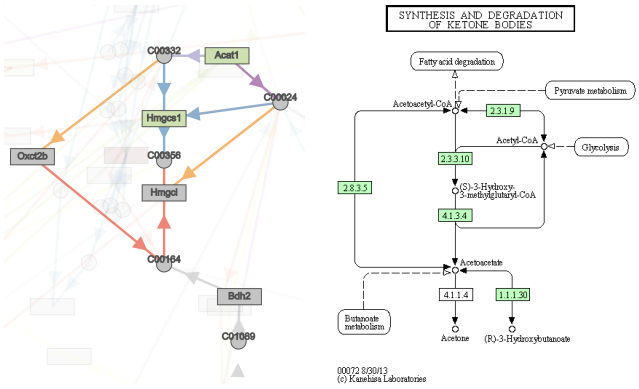

Figure 2: *Left: Detailed view of a selected pathway (synthesis and degradation of ketone bodies of the house mouse, KEGG ID: mmu00072) in our application. The focus is on the selected pathway, while all other pathways are de-emphasized using transparency. The selected pathway has been newly laid out to match the layout provided by KEGG [KG00] (shown to the right). Note that the white nodes in the KEGG reference image are genes that are not present in the selected organism (i.e., the mouse), therefore, we remove them in our visualization.*

Users can explore the pathway network by zooming and panning. If a user selects a gene in the table of input genes or clicks on a node in the network view, all pathways containing this gene are highlighted (R1). Individual pathways can be selected for closer inspection in the *Pathway* list. As shown in Fig. 2, the selected pathway is then newly laid out to resemble the layout provided by KEGG (R2). The rest of the network is still using the force-directed layout and acts as context. Genes can be colored according to the fold change using a diverging cool-to-warm color map: [Bre03, HB03]. Using this coloring and always the same layout for a pathway facilitates a comparison of different data sets (R3).

3. Discussion & Outlook

Our project partners are delighted with the current capabilities of our application and think that it can become a replacement for the previously used InCroMAP [WEZ12, WEBZ13], a desktop Java™ application that is no longer maintained. We aim to add more features, such as the ability to compare data from different individuals or to visualize time series data. Furthermore, we will extend our application to visualize not just transcriptomics but multi-omics data by integrating visualizations for proteomics and lipidomics data.

References

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec. 2011), 2301–2309. doi:10.1109/TVCG.2011.185. 2

[Bre03] BREWER C. A.: A Transition in Improving Maps: The ColorBrewer Example. *Cartography and Geographic Information Science 30*, 2 (2003), 159–162. doi:10.1559/152304003100011126. 2

[BYB*20] BUCHWEITZ L. F., YURKOVICH J. T., BLESSING C., KOHLER V., SCHWARZKOPF F., KING Z. A., YANG L., JÓHANNSSON F., SIGURJÓNSSON O. E., ROLFSSON O., HEINRICH J., DRÄGER A.: Visualizing metabolic network dynamics through time-series metabolomic data. *BMC Bioinformatics 21*, 1 (Apr. 2020), 130. URL: https://doi.org/10.1186/s12859-020-3415-z, doi:10.1186/s12859-020-3415-z. 2

[EJP11] ENJALBERT B., JOURDAN F., PORTAIS J.-C.: Intuitive Visualization and Analysis of Multi-Omics Data and Application to Escherichia coli Carbon Metabolism. *PLoS ONE 6*, 6 (2011). doi:10.1371/journal.pone.0021318. 2

[FLH*16] FRANZ M., LOPES C. T., HUCK G., DONG Y., SUMER O., BADER G. D.: Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics (Oxford, England) 32* (Jan. 2016), 309–311. doi:10.1093/bioinformatics/btv557. 2

[GDTC19] GHOSH S., DATTA A., TAN K., CHOI H.: SLIDE – a web-based tool for interactive visualization of large-scale – omics data. *Bioinformatics 35*, 2 (2019), 346–348. Publisher: Oxford Academic. doi:10.1093/bioinformatics/bty534. 2

[GOB*10] GEHLENBORG N., O'DONOGHUE S. I., BALIGA N. S., GOESMANN A., HIBBS M. A., KITANO H., KOHLBACHER O., NEUWEGER H., SCHNEIDER R., TENENBAUM D., GAVIN A.-C.: Visualization of omics data for systems biology. *Nature Methods 7*, 3 (Mar. 2010), S56–S68. doi:10.1038/nmeth.1436. 2

[HB03] HARROWER M., BREWER C. A.: ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal 40*, 1 (2003), 27–37. doi:10.1179/000870403235002042. 2

[HJ18] HARTMANN A., JOZEFOWICZ A. M.: VANTED: A Tool for Integrative Visualization and Analysis of -Omics Data. In *Plant Membrane Proteomics: Methods and Protocols*, Mock H.-P., Matros A., Witzel K., (Eds.), Methods in Molecular Biology. Springer, New York, NY, 2018, pp. 261–278. doi:10.1007/978-1-4939-7411-5_18. 2

[JYS*16] JANG Y., YU N., SEO J., KIM S., LEE S.: MONGKIE: an integrated tool for network analysis and visualization for multi-omics data. *Biology Direct 11*, 1 (2016), 10. doi:10.1186/s13062-016-0112-y. 2

[Kan19] KANEHISA M.: Toward understanding the origin and evolution of cellular organisms. *Protein Science 28*, 11 (2019), 1947–1951. doi:10.1002/pro.3715. 2

[KG00] KANEHISA M., GOTO S.: KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research 28*, 1 (2000), 27–30. doi:10.1093/nar/28.1.27. 1, 2

[KSF*19] KANEHISA M., SATO Y., FURUMICHI M., MORISHIMA K., TANABE M.: New approach for understanding genome variations in KEGG. *Nucleic Acids Research 47*, D1 (2019), D590–D595. doi:10.1093/nar/gky962. 2

[MPB*19] MARAI G. E., PINAUD B., BÜHLER K., LEX A., MORRIS J. H.: Ten simple rules to create biological network figures for communication. *PLOS Computational Biology 15*, 9 (2019), e1007244. doi:10.1371/journal.pcbi.1007244. 2

[RCM*20] REINA G., CHILDS H., MATKOVIC K., BÜHLER K., WALDNER M., PUGMIRE D., KOZLIKOVA B., ROPINSKI T., LJUNG P., ITOH T., GRÖLLER E., KRONE M.: The moving target of visualization software for an increasingly complex world. *Computers & Graphics* (2020). doi:10.1016/j.cag.2020.01.005. 2

[SMO*03] SHANNON P., MARKIEL A., OZIER O., BALIGA N. S., WANG J. T., RAMAGE D., AMIN N., SCHWIKOWSKI B., IDEKER T.: Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research 13*, 11 (2003), 2498–2504. doi:10.1101/gr.1239303. 2

[WEBZ13] WRZODEK C., EICHNER J., BÜCHEL F., ZELL A.: InCroMAP: integrated analysis of cross-platform microarray and pathway data. *Bioinformatics 29*, 4 (2013), 506–508. doi:10.1093/bioinformatics/bts709. 2

[WEZ12] WRZODEK C., EICHNER J., ZELL A.: Pathway-based visualization of cross-platform microarray datasets. *Bioinformatics 28*, 23 (2012), 3021–3026. doi:10.1093/bioinformatics/bts583. 2