Published in final edited form as:

Nat Methods. 2013 July ; 10(7): 597–598. doi:10.1038/nmeth.2517.

NetGestalt: integrating multidimensional omics data over biological networks

Zhiao Shi^{1,2,*}, Jing Wang^{3,*}, and Bing Zhang^{3,4,#}

¹Advanced Computing Center for Research & Education, Vanderbilt University, Nashville, TN 37235, USA

²Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235, USA

³Department of Biomedical Informatics, Vanderbilt University, Nashville, TN 37235, USA

⁴Department of Cancer Biology, Vanderbilt University, Nashville, TN 37235, USA

To the editor: Network visualization is typically based on node-link diagrams, which quickly become inadequate as network size and data complexity increase¹. To address this challenge, we developed NetGestalt (<http://www.netgestalt.org>), a web application for integrating multidimensional omics data over biological networks. NetGestalt exploits the inherent hierarchical modular architecture of biological networks² to achieve high scalability. Instead of using a two-dimensional node-link diagram (Fig. 1a), NetGestalt places the nodes of a network along the horizontal dimension of a webpage on the basis of the hierarchical architecture of the network (Fig. 1b) and displays the network module annotation at different scales below the ordered nodes (Fig. 1c). Visualization in the horizontal dimension conveys the functional relationship between different nodes (e.g. genes) as encoded in the network. Notably, this transformation makes it possible to render node-related information from different data sources and existing knowledge as tracks along the vertical dimension of the webpage for visual comparison and integration.

NetGestalt relies on a computational algorithm, available as an R package NetSAM (Network Seriation And Modularization), for revealing hierarchical organization of biological networks (Supplementary Methods, Supplementary Fig. 1). Using protein-protein interaction networks (PPINs) from human and mouse as examples, we showed that NetSAM could identify biologically coherent modules at different scales and create functionally meaningful one-dimensional ordering of the genes (Supplementary Note 1, Supplementary Table 1, Supplementary Fig. 2). Here we applied NetGestalt to a hierarchically and modularly organized human PPIN based on protein-protein interaction data from the Human Protein Reference Database (HPRD, <http://www.hprd.org>). We used two examples to demonstrate the features of NetGestalt and its potential in revealing important novel biological insights (Supplementary Note 1, Supplementary Fig. 3, Supplementary Fig. 4).

NetGestalt supports four track types corresponding to different data types (Supplementary Table 2). A single binary track (SBT) visualizes a single set of binary data, such as functional annotation data for genes or significant calls from a statistical analysis

[#]Correspondence should be addressed to B.Z. (bing.zhang@vanderbilt.edu).

^{*}Both authors contributed equally to this work

Author Contributions B.Z. led the project and oversaw the software development and data analysis. Z.S. designed the software architecture and implemented the software. J.W. prepared the data, developed the NetSAM package and performed the bioinformatics analysis. B.Z., J.W., and Z.S. wrote the manuscript.

(Supplementary Fig. 3j-k). A composite binary track (CBT) uses a heat map with two colors to simultaneously visualize a few related binary datasets, such as gene mutation status for a cohort of tumors (Supplementary Fig. 4l). A single continuous track (SCT) uses a bar chart to visualize a single set of continuous data, such as fold changes (Supplementary Fig. 3d). A composite continuous track (CCT) uses a heat map with colors ranging from blue to red to simultaneously visualize a few related continuous datasets, such as gene expression data for multiple samples (Supplementary Fig. 3c). Following the file formats described in the user manual (Supplementary Note 2), users can upload their own networks and data tracks into NetGestalt for visualization and analysis.

NetGestalt employs the Ajax (Asynchronous JavaScript and XML) technology and efficient software architecture to enable fast rendering and smooth navigation at scales ranging from individual genes to specific regions of interest to the whole network. Users can search for existing tracks in the database, add tracks for visualization, zoom in and out, pan left and right, reorder tracks, delete tracks, and generate node-link diagrams when appropriate. Venn diagram-based track comparison and enrichment analysis tools are also available to aid effective data analysis and hypothesis generation. All features implemented in NetGestalt are summarized in Supplementary Table 3.

Our first example combined heat map visualization, statistical analysis, pathway annotation, and cross datasets comparison to investigate gene expression in the advancement scheme of colorectal cancer known as the normal-adenoma-carcinoma sequence (Supplementary Note 1, Supplementary Fig. 3). The analysis highlighted an important role of chemokine signaling in colorectal cancer progression, and revealed novel dynamic gene expression patterns that may be useful for future therapeutic interventions based on anti-inflammatory strategies. Our second example integrated somatic mutation and copy number variation results from the TCGA glioblastoma multiforme (GBM) study³. The analysis identified known networks involved in GBM and discovered a novel FN1-centered network that may provide insights into both tumor progression mechanisms and clinical management of GBM (Supplementary Note 1, Supplementary Fig. 4). Moreover, by visualizing somatic mutation data from individual samples together with network module information, the analysis revealed a mutually exclusive mutation pattern for genes in the PIK3R1 centered network (Supplementary Note 1, Supplementary Fig. 4l).

Compared with typical network visualization tools, NetGestalt allows multi-scale representation and navigation of the data, and adjusts graphical presentations to the level of detail appropriate to a particular scale. It also allows simultaneous visualization of different types of data within the same framework to facilitate data integration. Moreover, NetGestalt complements the genome browsers⁴ in visualizing the functional relationship of genes located on different chromosome locations. This facilitates, for example, the discovery of important pathways and networks that link important genomic alterations (Supplementary Fig. 4p). A more detailed comparison between NetGestalt and other related tools is provided in Supplementary Table 4. Datasets used in this study (Supplementary Table 5) can be explored through the NetGestalt website, and a user manual is available in Supplementary Note 2.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health grants R01 GM088822, U24 CA159988, and P50 MH096972. This work was conducted in part using the resources of the Advanced Computing Center for Research

and Education at Vanderbilt University, Nashville, TN. The results published here are in part based upon data generated by The Cancer Genome Atlas pilot project established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at <http://cancergenome.nih.gov/>.

References

1. Gehlenborg N, et al. Visualization of omics data for systems biology. *Nat Methods*. 2010; 7:S56–S68. [PubMed: 20195258]
2. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet*. 2004; 5:101–113. [PubMed: 14735121]
3. TCGA-Research-Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008; 455:1061–1068. [PubMed: 18772890]
4. Kent WJ, et al. The human genome browser at UCSC. *Genome Res*. 2002; 12:996–1006. [PubMed: 12045153]

Figure 1. One-dimensional layout of a network preserves the hierarchical modular organization information

(a) Node-link diagram of a theoretical network. Colors of nodes and circles represent modules at different scales. (b) Placement of network nodes in one-dimension based on a computationally derived hierarchical tree. (c) Network modules at different hierarchical levels visualized as horizontal bars.