

BIF 701 LAB 10: Using OmicsNet to Explore Biological Networks

Christopher Eeles

Abstract—In this laboratory we will explore the recently published web-tool from McGill researchers Zhou, G. and Xia, J. (2018), aptly named OmicsNet. Our survey of this tool and its functionalities will provide insight into the cutting edge of biological network creation, analysis and visualization software. In doing so we will outline the types of analyses possible using OmicsNet and conduct a thorough investigation of one such methodology.

I. INTRODUCTION

Previously developed biological network tools mainly focused on protein-protein interaction (PPI) or metabolic networks and were often unable to handle high node volume without visualization becoming crowded and unreadable.³ In response to these limitations OmicsNet was developed to allow simple biological network creation from a diverse set of molecular interactions for analysis and visualization in 3D space.³ The tool is able to handle data sets from genes, proteins, microRNA, transcription factors and metabolites which can be combined and merged in a variety of ways.³

Given the increasing importance of systems biology in understanding how complex interactions create and affect the biology of cells, tissues, organisms and ecosystems, utilizing available network analysis tools and developing new ones is a necessity for bioinformaticians.² The possibilities in this field are vast, encompassing fields as diverse as biophysics and molecular physiology to epidemiology and ecology. Applications include development of precision medicines, mapping complex webs of drug-drug interactions, interpreting correlations between genetic variants and disease, and myriad other endeavors in basic as well as applied biological and medical sciences.⁴

II. METHODS⁵

A. OmicsNet

OmicsNet enables researchers to create networks to visualize relationships between genes, proteins, transcription factors and metabolites in 3D space. The web-tool contains a comprehensive, built-in knowledgebase allowing selection of multiple lists of molecules pertinent to a researcher's interests. Other data formats accepted include molecular lists—wherein the first column is read as the input type and the second column is read as the expression or quantitative measurement—and network files with .sif, .txt and .graphml extensions. The web-app is supported across multiple browsers including Chrome, Firefox and Safari with graphics generated using the WebGL Javascript API. Network generation starts from inputted molecular lists, which are then compared to the selected external interaction database to build the interaction network. The results are

displayed in a highly customizable Javascript application which enables exploration using different network models, annotated and coloured by the user.

B. Input Types

We began by selecting each of the gene/proteins, transcription factors (TF), miRNAs, and graph file options on the OmicsNet main page. Firstly, the gene/proteins option allowed selection of organism species as well as an extensive list of available database ID formats for upload to the server; in this exploration we will simply use the example set provided by OmicsNet. Secondly the transcription factors option displayed a similar input window species and database ID options specific to that tool. This pattern continued for the remaining options, allowing us to conclude that input types vary only in species for analysis and databases available. The exception was the graph file option, which instead prompted us for a file to upload or example files for small (.txt) and large (.graphml) networks.

C. Network Building

Examining the network building page we observed that concurrent input of gene, TF, miRNA and metabolite data was possible, but only three interactions could be selected at one time. Available interaction options include PPI, miRNA-gene interactions (MGI), metabolite-protein interaction (MPI) and TF-gene interactions (TGI) each of which had interaction database options. Network tools were also available to filter the input data to control network size. On submit, the network results summarize generated sub-networks, listing the number of nodes, edges and seeds for each.

D. Network Viewer

Proceeding to the network viewer, a Javascript interface displayed a visualization of our selected inputs and interactions. Input types were coloured with a legend available underneath which a node explorer, to select a subset of molecules of interest, allowed searching and deleting nodes based on the direction of your analysis. Edges appear to represent the relationships between nodes and can be customized for opacity to focus on each aspect respectively. Three layouts—spherical, force-directed and layered—allow different geometries to explore the networks relationships. To the right a number of explorer windows appear to allow more specific network analysis.

E. Detailed Analysis

For this lab we have select example data from Uniprot for protein and Kegg ID for metabolites, both for *H. sapiens*. Two comparisons were selected, PPI and MPI respectively, using the default interaction databases. On submission we were prompted to select an approach to expand the primary PPI network, from which 'target primary seed nodes' was chosen. No network tools were used to modify network size. Exploration of the network viewer will be recorded in the results section for subsequent analysis and discussion.

III. RESULTS⁵

The output labelled our proteins as genes, assumingly because our selected proteins are a result of expression these genes. Metabolites were highlighted in yellow. Edges were displayed in red and were bundled under the 'Edge' drop-down menu. Seed highlighting was also enabled to focus on the subset which was used to generate the network. Node labels were set to green for more easy visualization. Visualizations of the spherical, layered, and force-direction geometries are displayed in Fig. 1, Fig. 2 and Fig. 3, respectively.

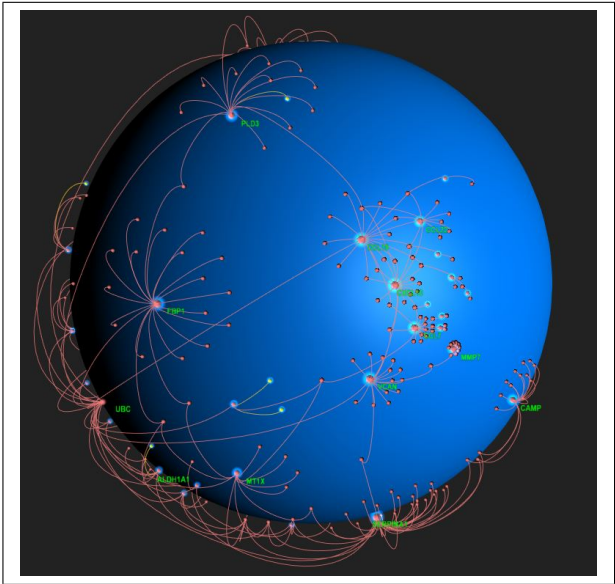


Fig. 1. Spherical Geometry

To explore the biological interpretation of these visualizations the HIST2HAC protein was selected due to its voluminous cluser of interactions. The HIST2H2AC protein and a subset of nodes—SMN1, TCF20 and ATXN7—were selected for indentification via the Uniprot database. The Fig. 4 displays the cluster selected, with the nodes of interest labelled and highlighted in blue. The batch selection function was used for this purpose.

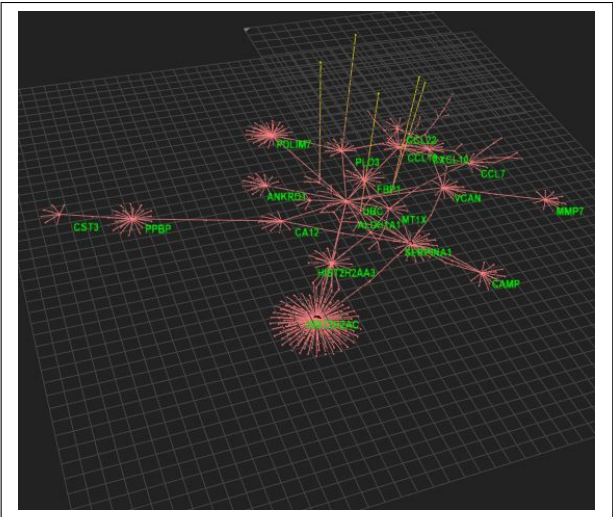


Fig. 2. Layered Geometry

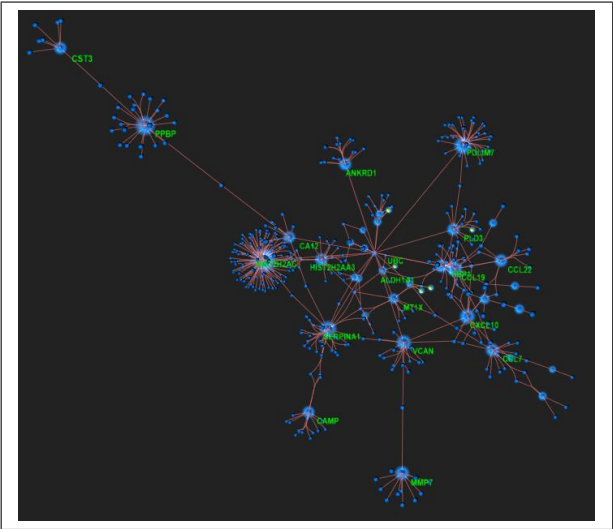


Fig. 3. Force-Direction Geometry

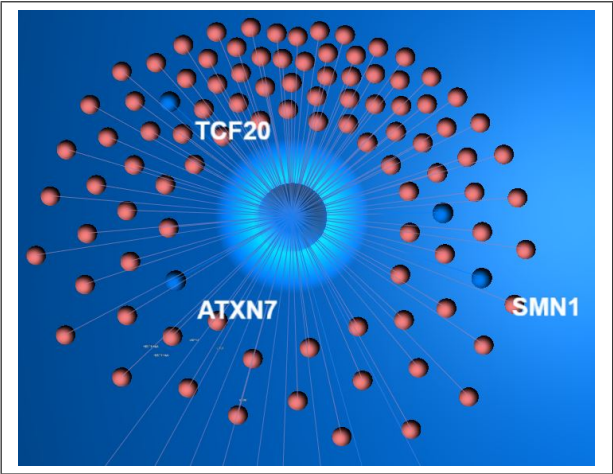


Fig. 4. HIST2HAC

IV. DISCUSSION AND CONCLUSION

Searching Uniprot for each protein identified in Fig. 4 and selecting the human variant we were able to identify each of these proteins, their function, and the biological processes in which they are involved. The seed node HIST2H2AC, highlighted in a light blue halo in Fig. 4, was identified as Histone H2A type 2-C which is a component of the nucleosome around which the DNA of chromosomes is wrapped for compaction and to limit accessibility to the machinery of transcription.⁶ Thus this histone plays an important role in transcription regulation as well as repair, replication and stabilization of chromosomes and the DNA within them.⁶ The major biological process in which HIST2H2AC is involved is chromatic organization.⁵ Based on this information we hypothesize that the selected interaction nodes within the HIST2H2AC cluster will play a role in organizing chromatin and the larger nucleus, as well as be involved in activation and execution of transcription.

Human SMN1 was identified as survival motor neuron protein.⁵ While not as clearly associated as hypothesized, this protein plays a role in catalyzing the assembly of small nuclear riboproteins (snRNPS), a component of the spliceosome needed to convert pre-mRNA into the mature, translation ready form.⁶ It is therefore involved in the transcription process and not contradictory to the hypothesis we formulated. The Uniprot results indicated that a key function of this protein is RNA binding within DNA-template transcription and termination as well as spliceosomal complex assembly processes.⁶ Interestingly, examining Fig. 4 we see two nodes identified as SMN1, potentially indicating multiple interactions with the histone seed protein.

Human TCF20 was identified as a transcriptional coactivator for the stromelysin gene.⁶ It therefore strongly supports the hypothesized function of this protein cluster. It was identified to have a role in a number of biological functions and processes including DNA-binding and transcription factor activity, RNA binding and regulation of DNA-templated transcription.⁶ The third node identified, ATX7, was identified as the human protein ataxin-7 which is a component of the STATA transcription coactivator-HAT complex.⁶ This complex is involved in histone post-translation modification of histones, thought to be essential to transcription activation and therefore regulation of gene expression. Specifically it plays a role in linking the nuclear cytoskeleton to activated histone complexes and therefore plays an essential role in movement of new transcripts out of the nucleus.⁶

The results of this analysis show that the identity of a seed node within an cluster can be used as an approximator of associated protein function. It is therefore reasonable to infer the reverse to be true, and placing newly identified proteins in such clusters provide invaluable clues about their identity and function. While our analysis focused on protein-protein interactions, the conclusions should be equally true for the other molecules accepted by OmicsNet. Given this information we can conclude that biological network analysis is an essential tool in mapping the various -omes of the species, aiding

scientists in placing novel molecules within the larger web of processes required for biological functioning. Therefore tools such as OmicsNet deserve considerable attention by bioinformatics researchers and dedicating resources to this endeavor will undoubtedly enable discovery and development of products which enhance human and animal well-being.

REFERENCES

- [1] School of Biological Sciences and Applied Chemistry. (2018). BIF701 Lab 10: Using OmicsNet to Explore Biological Networks. Seneca College: Toronto, ON.
- [2] School of Biological Sciences and Applied Chemistry. (2018). Topic 10: Systems Biology. Seneca College: Toronto, ON.
- [3] Zhou, G., Xia, J. (2018). OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic acids research*, 46(W1), W514-W522.
- [4] Ideker, T., Nussinov, R. (2017). Network approaches and applications in biology. *PLoS computational biology*, 13(10).
- [5] Xia Lab. (2018). OmicsNet: FAQs [Website]. McGill University: Montreal, QC.
- [6] National Institute of Health. (N.d.) UniProt [Website]. In *Elixir Core Data Resources*. Germany.