

CytoHiC: a cytoscape plugin for visual comparison of Hi-C networks

Yoli Shavit* and Pietro Lio'

Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK

Associate Editor: Alfonso Valencia

ABSTRACT

Summary: With the introduction of the Hi-C method new and fundamental properties of the nuclear architecture are emerging. The ability to interpret data generated by this method, which aims to capture the physical proximity between and within chromosomes, is crucial for uncovering the three dimensional structure of the nucleus. Providing researchers with tools for interactive visualization of Hi-C data can help in gaining new and important insights. Specifically, visual comparison can pinpoint changes in spatial organization between Hi-C datasets, originating from different cell lines or different species, or normalized by different methods. Here, we present CytoHiC, a Cytoscape plugin, which allow users to view and compare spatial maps of genomic landmarks, based on normalized Hi-C datasets. CytoHiC was developed to support intuitive visual comparison of Hi-C data and integration of additional genomic annotations.

Availability: The CytoHiC plugin, source code, user manual, example files and documentation are available at: <http://apps.cytoscape.org/apps/cytohicplugin>

Contact: yolisha@gmail.com or ys388@cam.ac.uk

Received on January 22, 2013; revised on February 18, 2013; accepted on March 1, 2013

1 INTRODUCTION

The development of the Hi-C technique by Lieberman *et al.* (2009) introduced revolutionary concepts to the research of genome organization. A following analysis suggested a fractal globule organization, at 1 Mb resolution, where loci interact within (intra-) and between (inter-) chromosomes (Lieberman *et al.*, 2009). Since then, significant efforts were made to remove bias and normalize the data (Hu *et al.*, 2012; Imakaev *et al.*, 2012; Yaffe and Tanay, 2011), to identify governing features, such as CCCTC factor and binding sites (Botta *et al.*, 2010), and to suggest the role of nuclear architecture in regulation (Dong *et al.*, 2010) and translocation events (Engreitz *et al.*, 2012; Zhang *et al.*, 2012).

Hi-C data can be considered as an end result of disassembling a puzzle or shuttering a globe into small pieces and reporting the pieces that were found together. Normalization methods attempt to remove biases and, in turn, report a corrected contact value between the different pieces. Assembling back the pieces, or genomic bins, into the underlying spatial map is an important challenge. The ability to view and compare Hi-C maps as spatial networks with respect to genomic landmarks of interest can

provide important insights about the properties, dynamics and role of the nuclear organization.

Through visual comparison, researchers can instantly identify what has changed and what has remained conserved between different Hi-C datasets. One may be interested in comparing the interaction map of orthologs, or view the spatial changes through a progress of a disease or through developmental stages. Comparison of cells with different functions may reveal common and cell-specific spatial arrangements. Finally, comparing the same Hi-C dataset under different normalization methods, or before and after normalization has been performed, can aid in assessing differences in coverage and correction.

We have developed CytoHiC, a plugin for the Cytoscape platform (Shannon *et al.*, 2003), for generating and visually comparing spatial maps from normalized Hi-C data. CytoHiC provides a user-friendly interface for uploading Hi-C data and genomic landmarks and exploits the visualization power of Cytoscape to draw the corresponding spatial maps. Its implementation was designed to allow for easy identification of spatial changes and for visual integration of related properties such as methylation and gene expression. As a Cytoscape plugin, CytoHiC also makes it easier to combine and convey multi-omic information. To the best of our knowledge, CytoHiC is the first tool developed for the task of an interactive visual comparison of Hi-C data.

2 IMPLEMENTATION

After a simple installation (placing the CytoHiC plugin in the plugins directory of Cytoscape before launching Cytoscape), the CytoHiC plugin is available from the Plugins menu of Cytoscape. The user can then choose either to view or compare Hi-C networks.

Through simple dialogs, the user can upload a set (when viewing) or two sets (when comparing) of normalized Hi-C data and landmarks of interest. A Hi-C dataset consists of two files: a bin file, specifying the segmentation of chromosomes and the genomic location of each segment (bin), and a contact file, specifying the contact value (or expected and observed values) between the different bins. The file format should be specified by the user (two main formats are supported: a contact matrix or a list of observed and expected values). The landmarks file consists of the names and genomic locations of the landmarks of interest. The user may also specify a numeric property of interest (e.g. methylation level or gene expression level) that will be incorporated when visualizing the network (see Section 2.2 for details). While CytoHiC is mainly intended for normalized data, the user may

*To whom correspondence should be addressed.

upload data before normalization, if they are given in one of the supported formats. A user manual is incorporated in the Cytoscape's help content, providing details on general usage and supported file formats.

2.1 Network generation

Given a set of bins, a contact matrix and landmarks of interest, CytoHiC matches landmarks to bins and calculates the contact value between each pair of landmarks. When a landmark is spread over more than one bin, the contact value is calculated as a weighted sum, so that the weight is proportional to the overlap between the bin and the landmark. The spatial network is then generated, where landmarks correspond to nodes and (undirected) edges are weighted by the inverse of the contact value, for an approximation of physical proximity. A relationship of inverse proportion between contact frequencies and physical proximity was previously suggested by Fraser *et al.* (2009). When the contact value is not available or when the contact value is zero (owing to limited coverage of the Hi-C data or owing to normalization correction), there will be no edge between the landmarks.

2.2 Network visualization

The network is drawn using a force-directed layout algorithm (available from Cytoscape's layout algorithm collection), based on the edge weights. Users may also explore other layouts based on the edge weight attribute. The details of the nodes and edges are available by hovering and from the node and edge browsers of Cytoscape.

By default, nodes are colored by their chromosome affiliation (i.e. landmarks in the same chromosome will have the same color). However, when the user specifies an additional property of interest (for the landmarks), nodes are colored by a blue and red gradient with respect to the range of the property's values. For example, if a methylation level is specified, each node (landmark) will be colored accordingly. Highly methylated landmarks will appear in dark red and under-methylated landmarks will be colored in dark blue. Intermediate methylation levels will be assigned with intermediate colors. The edges are labeled by their weight (approximation of distance), with a line style matching the interaction type they represent. A solid line corresponds to an interaction within a chromosome (intra-interaction), and a dashed line corresponds to an interaction between chromosomes (inter-interaction).

2.3 Visual comparison of landmarks

The option to visually compare networks is available from the CytoHiC menu and from a right-click pop-up. The user may

compare two networks or compare a current network with another network. The nodes (landmarks) are labeled with indices so that matching landmarks can be easily compared (the indices are given according to the ordering of the landmarks file).

3 CONCLUSION

CytoHiC is a platform-independent easy to install plugin for Cytoscape. It offers visual exploration and comparison so users can instantly and easily observe trends and changes in the spatial environment of genomic landmarks. Users may explore spatial changes from evolutionary, functional, developmental and clinical point of views. As Hi-C data take its important place in the research of nuclear architecture, visual exploration applications such as CytoHiC become an important part of the required toolbox.

While the spatial networks generated by CytoHiC provide the user an approximation for physical proximity based on interaction frequencies, we hope to offer better methods for predicting the nuclear positions of landmarks and to fix coverage and correction errors in the future. Incorporation of such methods with a visualization tool can provide a more accurate representation of the nucleus topological map.

Funding: We thank FP7-Health-F5-2012, under grant agreement n° 305280 (MIMOmics).

Conflict of Interest: none declared.

REFERENCES

- Botta,M. *et al.* (2010) Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Mol. Syst. Biol.*, **6**, 426.
- Dong,X. *et al.* (2010) Human transcriptional interactome of chromatin contribute to gene co-expression. *BMC Genomics*, **11**, 704.
- Engreitz,J.M. *et al.* (2012) Three-dimensional genome architecture influences partner selection for chromosomal translocations in human disease. *PLoS One*, **7**, e44196.
- Fraser,J. *et al.* (2009) Chromatin conformation signatures of cellular differentiation. *Genome Biol.*, **10**, R37.
- Hu,M. *et al.* (2012) HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics*, 3–5.
- Imakaev,M. *et al.* (2012) Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods*, **9**, 999–1003.
- Lieberman-Aiden,E. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289–93.
- Shannon,P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Yaffe,E. and Tanay,A. (2011) Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.*, **43**, 1059–65.
- Zhang,Y. *et al.* (2012) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell*, **148**, 908–921.