HiGlass: Web-based visual comparison and exploration of genome interaction maps

Peter Kerpedjiev[1], Nezar Abdennur[2], Fritz Lekschas[3], Chuck McCallum[1], Kasper Dinkla[3], Hendrik Strobelt[3], Jacob M Luber[1,4], Scott Ouellette[1], Alaleh Azhir[1], Nikhil Kumar[1], Jeewon Hwang[1], Burak H Alver[1], Hanspeter Pfister[3], Leonid A Mirny[5,6], Peter J Park[1], Nils Gehlenborg[1,*]

[1] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
[2] Computational and Systems Biology Program, MIT, Cambridge, USA
[3] School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[4] Bioinformatics and Integrative Genomics Program, Harvard Medical School, Boston, MA, USA
[5] Department of Physics, MIT, Cambridge, USA
[6] Institute for Medical Engineering and Science, MIT, Cambridge, USA

** Correspondence should be addressed to nils@hms.harvard.edu.*

Abstract

We present HiGlass (http://higlass.io), a web-based viewer for genome interaction maps featuring synchronized navigation of multiple views as well as continuous zooming and panning for navigation across genomic loci and resolutions. We demonstrate how visual comparison of Hi-C and other genomic data from different experimental conditions can be used to efficiently identify salient outcomes of experimental perturbations, generate new hypotheses, and share the results with the community.

The development of chromosome capture assays measuring the proximity of two or more non-adjacent regions of the genome helped elucidate how the structure and dynamics of chromosome structure affect gene regulation and cellular function[1,2]. Techniques such as Hi-C have revealed features of genome organization such as A/B compartments, topologically associated domains (TADs) and loops[1,3–5]. They have helped implicate changes in genome organization in a variety of disorders, including acute lymphoblastic leukemia[6], colorectal cancer[7], and limb development disorders[8]. More fundamentally, they provide insights into the mechanisms of how genome conformation structures arise and are maintained[9–11]. Major efforts like the 4D Nucleome Network and the ENCODE project are generating such data at large scale across different cell lines and conditions with the aim of understanding the mechanisms that govern processes such as gene regulation and DNA replication as well as to cross-validate the results from different experimental assays[2,12132,12].

To obtain genome conformation capture maps, raw sequencing data are processed to identify captured proximity ligation events such as contacts or interactions between genomic loci, which are then binned to form contact matrices[14–16]; see Lajoie et al.[17] and Ay & Noble[18] for reviews of Hi-C data processing. The discovery of novel genomic structures and mechanisms, however, also requires sophisticated visual tools for exploring features relevant at scales ranging from tens to millions of base pairs[15,19,20]. Given the multiscale features of genome organization, it is crucial that such visualization tools support continuous zooming and panning, comparison across multiple scales and conditions as well as integration with additional genomic and epigenomic data. Existing tools support different ways of displaying contact frequencies, such as rectangular heatmaps, triangular heatmaps, arc plots, or circular plots, and different levels of interaction ranging from static plotting to interactive zooming and panning, as well as different levels of integration with other genomic data types[15,21–24]. None of the existing tools, however, provides flexible methods for fast, synchronized navigation across multiple conditions, resolutions and loci or the ability to share interactive visualizations with the broader community.

To address these shortcomings, we created HiGlass, an open source, web-based application with a customizable interface designed to support multiscale contact map visualization across multiple resolutions, loci, and conditions (Supplementary Methods). HiGlass introduces the concept of *composable linked views* for genomic data visualization (Supplementary Fig. 1). This approach enables users to compose the layout, content, and synchronization of locus, zoom-level, and other view properties across multiple views (Supplementary Fig. 2). These view compositions can range from the juxtaposition of two or more heatmaps (Fig. 1) to sophisticated collations of matrices, tracks and viewport projections (Fig. 2). View compositions serve to display data at multiple scales, to corroborate observations with other types of evidence and to facilitate comparisons between experiments. As a web-based tool, HiGlass also supports storing and sharing of view compositions with other investigators and the public via hyperlinks. The tool can be used to access selected public datasets at http://higlass.io, or it may be run locally and populated with private data using a provided Docker container.

To illustrate the utility of composable linked views, we used HiGlass to highlight key results of a recent study showing the effect of induced deletion of the cohesin loading factor *Nibpl* on chromosome organization in adult mouse hepatocytes[25]. We obtained Hi-C contact data and binned it at multiple resolutions starting at 1 kb for wild type (WT) and ΔNipbl primary hepatocytes (Supplementary Methods). We loaded both samples as separate views (Fig. 1, top) and linked them via location and zoom level. With the two linked views, we could navigate to regions clearly showing the disappearance of features in the ΔNipbl condition. For example, in the gene-poor region from chr14:80 Mb to chr14:100 Mb of mm9, we observe a robust "bleaching" of near-diagonal contact enrichment patterns. The patterns that disappeared are examples of TADs because they do not show the long range checkering characteristic of A/B compartments. In contrast, in the relatively gene-rich region upstream of chr14:80 Mb, we see an enhancement of the checkered pattern and the emergence of a finer division of A/B regions in the ΔNipbl condition. To explore this region more closely, we created two additional linked views for WT and ΔNipbl and navigated to the region between chr14:50 Mb and chr14:70 Mb

(Fig. 1, bottom). Adding H3K4me3 and H3K27ac ChIP-seq signal tracks revealed that these marks, while similar between conditions, correlate more strongly with the compartmentalization pattern in ΔNipbl. Finally, we used a viewport projection to mark the position of the bottom views relative to the top, resulting in the complete view composition shown in Fig. 1. This interactive visual recapitulation of key results from Schwarzer et al illustrates how synchronized navigation across loci and resolutions by linking views between multiple conditions facilitates the exploration of the complex effects of global perturbations on chromosome organization at multiple scales.

Using the same view composition we noticed the appearance of small dark patches away from the diagonal. To investigate them we created a new composition containing an overview and two zoom- and location-linked detail views (Fig. 2). By using the overview to find patches and comparing them using the detail views, we established that they are more enriched in the mutant condition than in the wild type, that they represent interaction between pairs of short active regions (type A compartment) and that they are closely aligned with gene annotations. Including RNA-seq and ChIP-seq tracks let us see that the genes which align with these patches are virtually always transcriptionally active. This further supports the conclusion of Schwarzer et al. that the absence of cohesin loading leads to strengthened A/B compartmentalization. It also provides avenues for further investigation and analysis of the structural and regulatory consequences of cohesin loss.

Using HiGlass to create the composable linked views shown in Figs. 1 and 2 enabled us to interactively explore the data generated by Schwarzer et al. across different conditions, zoom levels and loci. This gave us not only a clearer understanding of the results, but allowed us to see unexpected patterns and rapidly gather observations to be used in generating new hypotheses. We provide links to interactive online versions of the figures to disseminate the findings and encourage collaboration and discovery. Going forward, we aim to extend HiGlass to support not only additional data types but also new ways of combining and comparing tracks to offer new insights into genome organization.

Author Contributions

PK and NG conceived the research. PK, NA and NG wrote the manuscript with input from LAM, PJP, and BHA. PK, NA, FL, and CM wrote the software with help from KD, HS, JML, SO, AA, NK, and JH. BHA, HP, LAM, and PJP provided valuable input and advice for the project.

Acknowledgements

We thank Francois Spitz, Wibke Schwarzer, and Aleksandra Pekowska for providing the data presented in this paper. We also acknowledge important suggestions and feedback from members of the Park Lab at Harvard Medical School, the Mirny Lab at MIT, and the Dekker Lab at University of Massachusetts Medical School as well as members of 4D Nucleome Data

Coordination and Integration Center who provided input and feedback. This project was made possible by funding from the National Institutes of Health (U01 CA200059, R00 HG007583, and U54 HG007963).

References

1. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
2. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14,** 390–403 (2013).
3. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
4. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
5. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159,** 1665–1680 (2014).
6. Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351,** 1454–1458 (2016).
7. Seaman, L. *et al.* Nucleome Analysis Reveals Structure-function Relationships for Colon Cancer. *Mol. Cancer Res.* (2017). doi:10.1158/1541-7786.MCR-16-0374
8. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161,** 1012–1025 (2015).
9. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E6456–65 (2015).
10. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15,** 2038–2049 (2016).
11. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 996–1001 (2014).
12. Dekker, J. *et al.* The 4D Nucleome Project. *bioRxiv* 103499 (2017). doi:10.1101/103499
13. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
14. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9,** 999–1003 (2012).
15. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3,** 99–101 (2016).
16. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16,** 259 (2015).
17. Lajoie, B. R., Dekker, J. & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* **72,** 65–75 (2015).
18. Ay, F. & Noble, W. S. Analysis methods for studying the 3D architecture of the genome. *Genome Biol.* **16,** 183 (2015).

19. Gehlenborg, N. *et al.* Visualization of omics data for systems biology. *Nat. Methods* **7,** S56–68 (2010).
20. Nielsen, C. B., Cantor, M., Dubchak, I., Gordon, D. & Wang, T. Visualizing genomes: techniques and challenges. *Nat. Methods* **7,** S5–S15 (2010).
21. Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *bioRxiv* 112268 (2017). doi:10.1101/112268
22. Zhou, X. *et al.* Exploring long-range genome interactions using the WashU Epigenome Browser. *Nat. Methods* **10,** 375–376 (2013).
23. Akdemir, K. C. & Chin, L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* **16,** 198 (2015).
24. Yardımcı, G. G. & Noble, W. S. Software tools for visualizing Hi-C data. *Genome Biol.* **18,** 26 (2017).
25. Schwarzer, W. *et al.* Two independent modes of chromosome organization are revealed by cohesin removal. *bioRxiv* 094185 (2016). doi:10.1101/094185

Figures

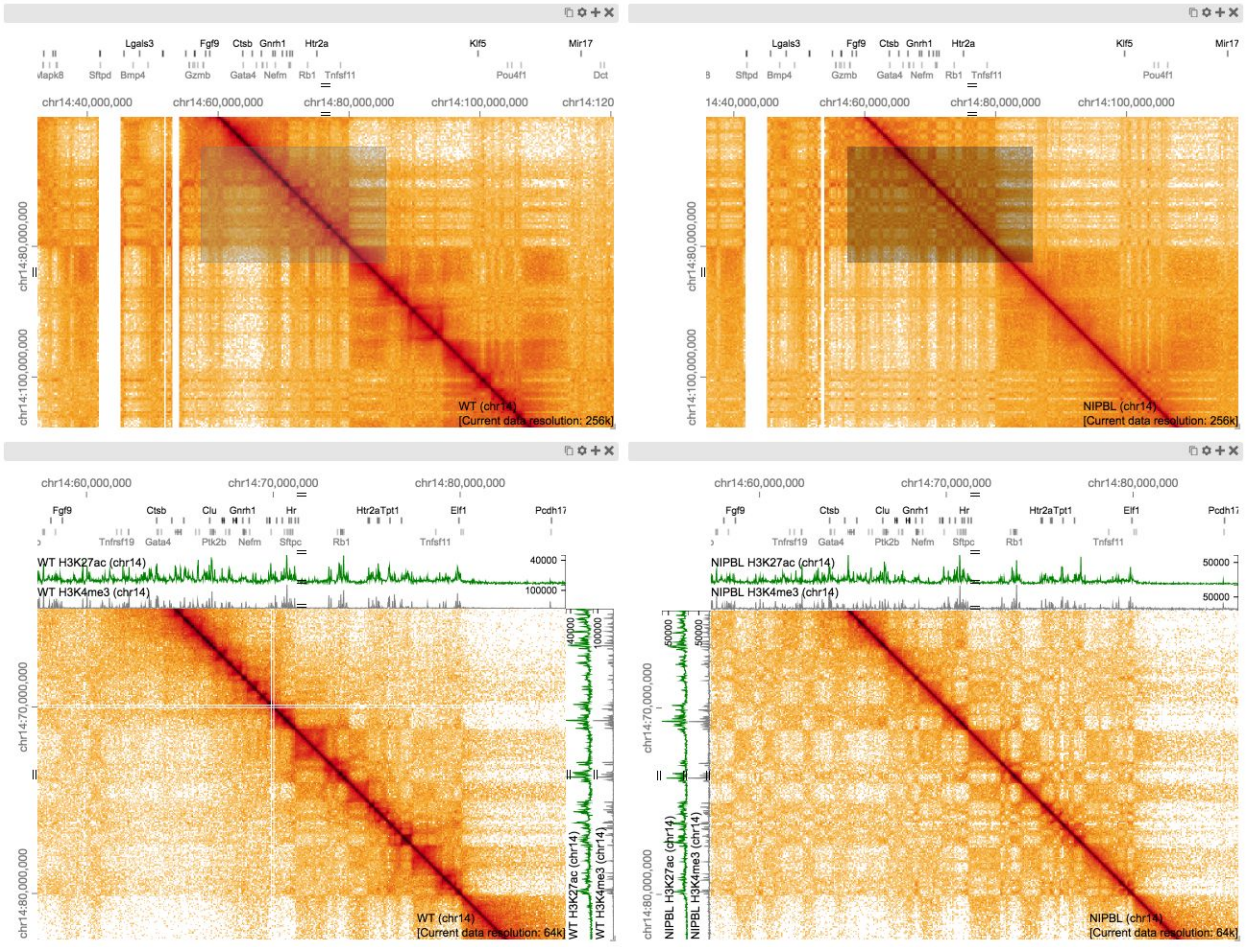

Figure 1 | A view composition highlighting the results from Schwarzer et al. with data from WT (left), and mutant (ΔNipbl, right) samples[25]. The top two views are linked to each other by zoom and location such that they always display the same region at the same resolution. Comparing the control (left) and mutant (right) condition at this zoom level reveals the bleaching of TADs in the gene-poor region in the lower right hand part of the maps. The bottom two views, also linked to each other by zoom and location, show a zoomed in perspective where a more fragmented compartmentalization of the ΔNipbl mutant (right) as compared to WT (left) can be seen. The black rectangles in the top views, which are referred to as viewport projections in HiGlass, show the positions and extent of the bottom views (Supplementary Fig. 3). The white lines in the bottom left panel are a result of bins filtered during matrix balancing. An interactive version of this figure is available at http://higlass.io/app/?config=Tf2-ublRTey9hiBKMlgzwg.

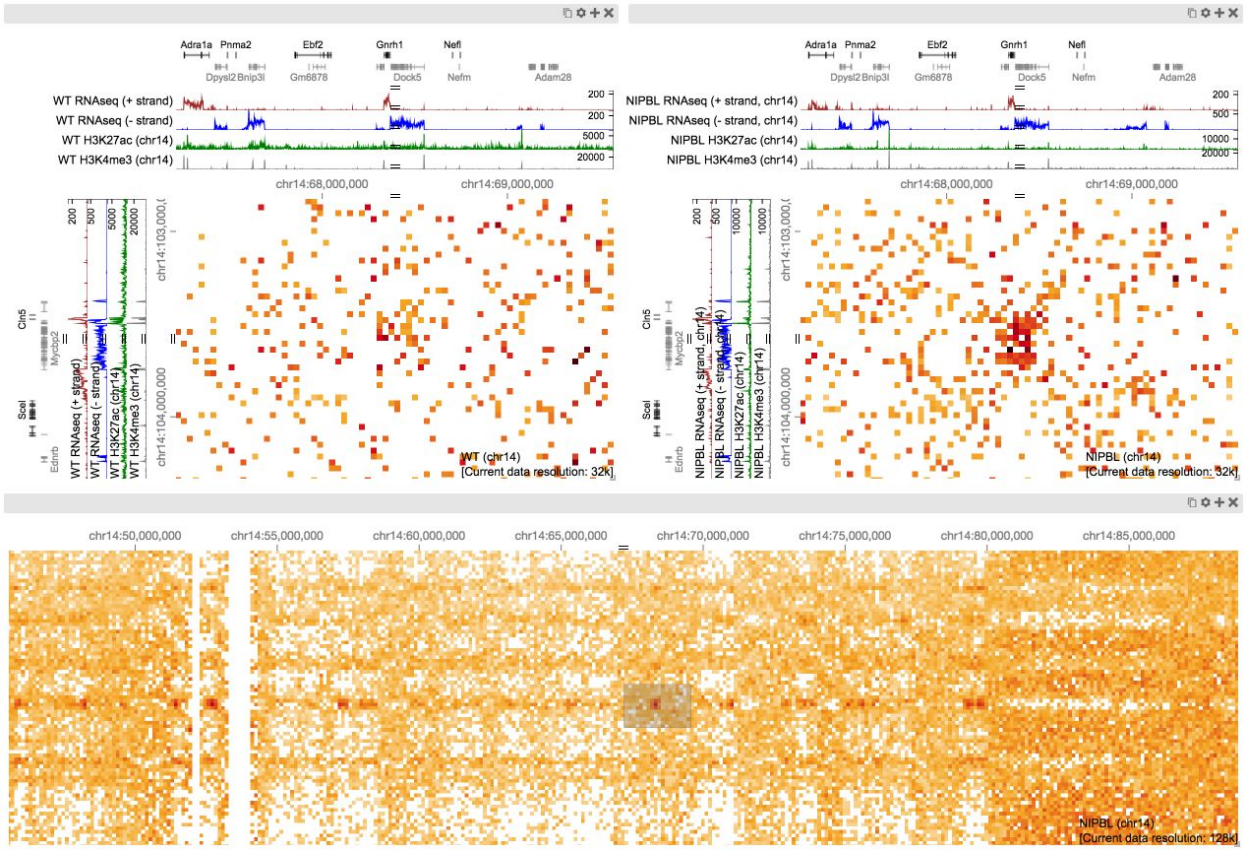

Figure 2 | A view composition containing two views linked by location and zoom (top) and an independent (unlinked) zoomed out overview (bottom) (Supplementary Fig. 4). The two views on top show data from chromosome 14 (mm9) in the wild type and ΔNipbl conditions, respectively. The bottom view shows data from the mutant condition as well as a projection of the viewport visible in the top views. The patch visible in the ΔNipbl condition (top left) is notably absent from the control (right). The gene annotations, RNAseq, H3K27me and H3K4me3 tracks show the presence and transcription of the *Dock5* and *Mycbp2* genes on the - strand as well as the presence and transcription of the *Gnrh1* and *Cln5* genes on the + strand. An interactive version of this figure is available at http://higlass.io/app/?config=Q5LdNchQRLSZ_0yKsTEoiw.