Computational and Structural Biotechnology Journal 23 (2024) 3467–3471

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Software/Web server Article

ppHiC: Interactive exploration of Hi-C results on the ProteinPaint web portal

Akanksha Rajput [1], Colleen Reilly [1], Airen Zaldivar Peraza [1], Jian Wang, Edgar Sioson, Gavriel Matt, Robin Paul, Congyu Lu, Aleksandar Acic, Karishma Gangwani, Xin Zhou [*]

Department of Computational Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, United States

ARTICLE INFO

Keywords:
Hi-C
ProteinPaint
Contact matrix
Visualization
Web server
Genomic rearrangement

ABSTRACT

The ProteinPaint Hi-C tool (ppHiC) facilitates web-based visualization and collaborative exploration of Hi-C data, a vital resource for understanding three-dimensional genomic structures. ppHiC allows researchers to easily analyze large Hi-C datasets on a web browser without requiring the computational expertise that has heretofore limited access to this complex genomic data. The platform is compatible with multiple Hi-C data versions and boasts a highly customizable interface, including a configuration panel for the precise adjustment of key visualization parameters. The tool's interactive features offer a broad range of views, from whole-genome landscapes to detailed interactions between pairs of loci, that are accessible within a single, integrated environment. Here, we demonstrate how using ppHiC to visualize an altered chromatin conformational landscape in neuroblastoma can inform understanding of the genomic rearrangements in this cancer.

1. Introduction

Hi-C, a high-throughput genomic assay for mapping the spatial organization of chromatin within the nucleus, has been an invaluable tool for interrogating the dynamic relationship between DNA three-dimensional structure and biological function [1]. Hi-C improves upon previous chromosome conformation capture methods like 3 C, 4 C, and 5 C by increasing the scope to include all the interactions between all fragment pairs. While this comprehensive approach has the potential to significantly advance understanding of genomic regulation and chromosomal architecture, it produces complex data that can be challenging to analyze and visualize.

The primary output of Hi-C experiments is a 'contact matrix', a two-dimensional array in which each cell quantifies the interaction frequency between a pair of genomic loci [2,3]. The patterns revealed by the contact matrix allow for decipherment of the chromatin's folding or looping behaviors, information that is integral to understanding the functional implications of chromosome organization. These matrices are often extremely large and dense, making them computationally intensive to process and visualize. Moreover, Hi-C data is subject to various experimental biases and noise (e.g., sequencing depth, GC content, and fragment length), necessitating proper normalization techniques for accurate interpretation. Visualization and analysis is further complicated by the fact that chromatin interactions occur at a range of genomic scales, from individual genes to entire chromosomes. Though informative on its own, Hi-C data often needs to be integrated with other genomic and epigenomic data types to produce comprehensive insights, a process that requires sophisticated computational methods. Finally, the large size of Hi-C datasets has made sharing and collaborative analysis difficult, potentially hindering scientific progress.

Current tools widely used to visualize Hi-C data, such as Juicebox [4], HiGlass [5], and the 3D Genome Browser [6], are beset by various limitations. Some lack support for all Hi-C file versions. Desktop tools are less useful than web-based tools for accessing large data files and broadly sharing Hi-C data [7]. Many tools have limited visualization features and may not support global, genome-wide views, detailed views between pairs of chromosomes or loci, or genome browser-based views. Many tools also lack genomic annotation datasets for interpreting chromatin interactions.

To address these shortcomings and to improve the overall accessibility and interpretability of Hi-C data, we developed ppHiC, an advanced visualization tool that is part of the ProteinPaint platform [8,

Abbreviations: ppHiC, ProteinPaint Hi-C; SVG, Scalable vector graphics; CSS, Cascading style sheets; NBL, neuroblastoma; SV, structural variation.

* Corresponding author.
E-mail address: xin.zhou@stjude.org (X. Zhou).

[1] Contributed equally.

https://doi.org/10.1016/j.csbj.2024.09.020
Received 10 July 2024; Received in revised form 26 September 2024; Accepted 26 September 2024
Available online 2 October 2024
2001-0370/© 2024 St. Jude Children's Research Hospital. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

9]. ppHiC delivers greater analytical depth and an enhanced user experience through a number of features, including: (i) compatibility with the three most recent Hi-C file versions (v9, v8, and v7); (ii) the ability to integrate custom annotation tracks; (iii) the ability to select from a variety of normalization methods to adjust for biases in data; (iv) the availability of multiple matrix types to suit different analytical needs; and (v) multiple viewing modes and the ability to transition between them seamlessly, allowing users to view intra- and inter-chromosomal interactions from the level of nucleotides to the whole genome.

Our Hi-C visualization tool makes it easier to turn raw Hi-C data into biological insights. ppHiC allows researchers to visualize and interpret complex Hi-C datasets with greater ease and precision, facilitating investigations of the long-range interactions between genomic elements, regulatory patterns, and three-dimensional genome structure. Furthermore, by making it possible to share Hi-C data visualizations over the web, ppHiC will advance scientific collaboration in genomics and related fields. Altogether, this tool will contribute to understanding of the spatial arrangement of the genome and its consequences for cellular function and disease pathogenesis.

2. Methods

2.1. Web server design

The ppHiC module is part of the ProteinPaint codebase, with features that allow effective management of complex Hi-C data. The data is stored on a remote server that runs on Node.js, a JavaScript runtime environment suitable for high performance web services, ensuring reliable and continuous user access. On the server backend, ppHiC calls the Juicebox tool [4] for efficient retrieval of Hi-C data from both local and remote sources.

ppHiC's user interface is built with the D3.js library, allowing for the dynamic manipulation of SVG, HTML5, and CSS to produce detailed images of Hi-C data, including heatmaps and arc diagrams that show how often different parts of the DNA come into contact in three-dimensional space. The Hi-C visualization module provides a user-friendly interface for rapid, in-depth genomic analysis.

The ppHiC web portal is a transformative tool that revolutionizes the way researchers work with Hi-C data. It requires no advanced technical knowledge, allowing researchers to make detailed genomic studies with ease. The platform simplifies data adjustments and corrections to ensure accurate results. Its interactive features provide a clear view of the genome's 3D structure, and its annotations add valuable context to the data. Unlike other tools, such as Galaxy HiCExplorer 3 [10] (which requires fastq files), HiCeekR [11] (which uses.bam,.bai, or.txt files), and the 3D Genome Browser [6] (which accepts.BUTLR files), ppHiC accommodates user input in multiple formats, including Juicebox Hi-C files, compressed BEDs, and text files, enhancing its adaptability (Table S1).

2.2. Multiple viewpoints of ppHiC

The ppHiC web server provides four different views through which researchers can explore intra- and inter-chromosomal interactions. These views are described below, as are methods to transition from a general view to a detailed view.

A. *Genome View*: The Genome View provides a global picture of chromatin organization across the genome, showing how various parts of DNA interact across all chromosomes. This bird's eye view allows for quick identification of outstanding interactions between any pair of loci and provides a general portrait of genome organization that serves as a foundation for subsequent, more detailed analyses. In the Genome View, clicking on a rectangle representing interactions between a pair of chromosomes will bring users to the "Chromosome Pair View".
B. *Chromosome Pair View*: With the Chromosome Pair View, researchers can focus on a pair of specific chromosomes to observe their interactions. This view provides better resolution than Genome View and can be applied to both intra- or inter-chromosomal pairs. Clicking on any point from this view will launch the "Detailed View".
C. *Detailed View*: The Detailed View provides an up-close look at the contact between two regions of interest identified from the Chromosome Pair View. Researchers can explore the intricate folding patterns and local interactions at a high resolution, making it possible to investigate the roles of genes and regulatory elements in a targeted genomic area. Users can pan and zoom the two genomic regions in the Detailed View or launch the "Genome Browser View" by clicking the "Horizontal View" button.
D. *Genome Browser View*: In the Genome Browser View, two chromosomal regions from the Detailed View are displayed side-by-side in a genome browser format. This view allows visualization of each region's intra-region contact as well as contacts between regions, making it particularly useful for examining intra- and inter-chromosomal long-range interactions. Users can pan and zoom on each region independently, providing additional flexibility.

2.3. Visualization algorithm

The ppHiC visualization algorithm efficiently renders large-scale Hi-C data in real-time within a web browser environment. When a user requests chromatin interaction data for a pair of chromosomal regions, ppHiC uses the Juicebox tool [4] to retrieve the relevant portion of Hi-C data from the server. This approach circumvents the need to load the entire dataset into memory, allowing large datasets to be handled more efficiently. Users can also select from multiple normalization methods to account for experimental biases; these include Knight-Ruiz (KR), Iterative Correction, and Eigenvector decomposition (ICE), and Vanilla Coverage (VC) [12,13].

ppHiC dynamically adjusts the resolution of the displayed data according to the level of zoom. At a genome-wide view, data is binned at a lower resolution (e.g., 1 Mb), while zoomed-in views use higher resolution bins (e.g., 5 kb) or fragment resolutions. Interaction frequencies are mapped to a color scale for visualization. ppHiC uses a logarithmic color scale by default to better represent the wide range of interaction frequencies, but users can choose alternative color mappings.

The D3.js library is used to render the Hi-C data as an interactive heatmap. Each pixel in the heatmap represents a bin of interactions, with the color intensity indicating the frequency of interaction. The algorithm supports real-time panning and zooming. As a user interacts with the visualization, the algorithm efficiently recalculates which data needs to be displayed and then updates the view accordingly. Genomic annotations are rendered alongside the Hi-C heatmap. The algorithm ensures that annotations are correctly aligned with the corresponding genomic positions in the Hi-C data.

2.4. Collaboration and data sharing

The web-based framework of the ppHiC module is designed with collaboration in mind. It facilitates data and visualization sharing between research teams and across the scientific community. Users can export their findings for further analysis or inclusion in scientific publications. The Hi-C visualization module in ProteinPaint offers a range of interactive features that enhance collaborative research. For instance, researchers can create and share custom views of chromatin interactions, highlighting specific regions of interest or patterns observed in the data.

In summary, the ppHiC module in ProteinPaint is a powerful tool with a simple interface. It provides researchers with a rich set of features to explore chromatin conformation and its regulatory roles, improving

understanding of gene regulation and genomic organization. The platform's collaborative capabilities and comprehensive data integration make it a useful resource for advancing genomic research and scientific discovery.

3. Results and discussion

3.1. Workflow

The ppHiC workflow for Hi-C data visualization is both efficient and user-friendly. Users begin by preparing their Hi-C experiment data in Juicebox Hi-C format files. They then open the ProteinPaint portal on a web browser and bring up a ppHiC view of the Hi-C files. On the portal, users can choose the reference genome version and bring up annotation tracks to provide genomic context for the Hi-C data, either selecting from predefined annotation tracks (e.g., RefGene, GENCODE, or RepeatMasker) or adding their own custom tracks. The ppHiC platform facilitates data visualization at various scales and formats, including heatmaps for intensive Hi-C experiment results and arc graphs, which can be defined in compressed BED files, for sparse interactions. For data analysis, users can zoom in on specific regions to explore chromatin interactions and the broader organization of the genome. The interface allows for seamless navigation and detailed examination of genomic regions.

3.2. Comparison with existing tools

Though many of ppHiC's features are shared by existing tools, ppHiC is distinguished for its comprehensive feature set and integration with the ProteinPaint framework (Table S1). Unlike existing tools, ppHiC allows users to transition seamlessly between multiple view modes, including genome-wide, chromosome-pair, and genome browser views. This unique capability is essential for understanding Hi-C data across different scales of genomic organization, from whole-genome patterns to specific locus interactions.

ppHiC supports all recent Hi-C file versions (v7, v8, and v9) and offers both heatmap and arc plot visualizations, features shared by only a few other tools. Moreover, ppHiC's web-based interface and ability to accept user input data from local computers enhance its accessibility and flexibility. Whereas most current tools are limited to the visualization of Hi-C data, ppHiC's integration within a broader data portal framework allows users to analyze Hi-C data in the context of other genomic and

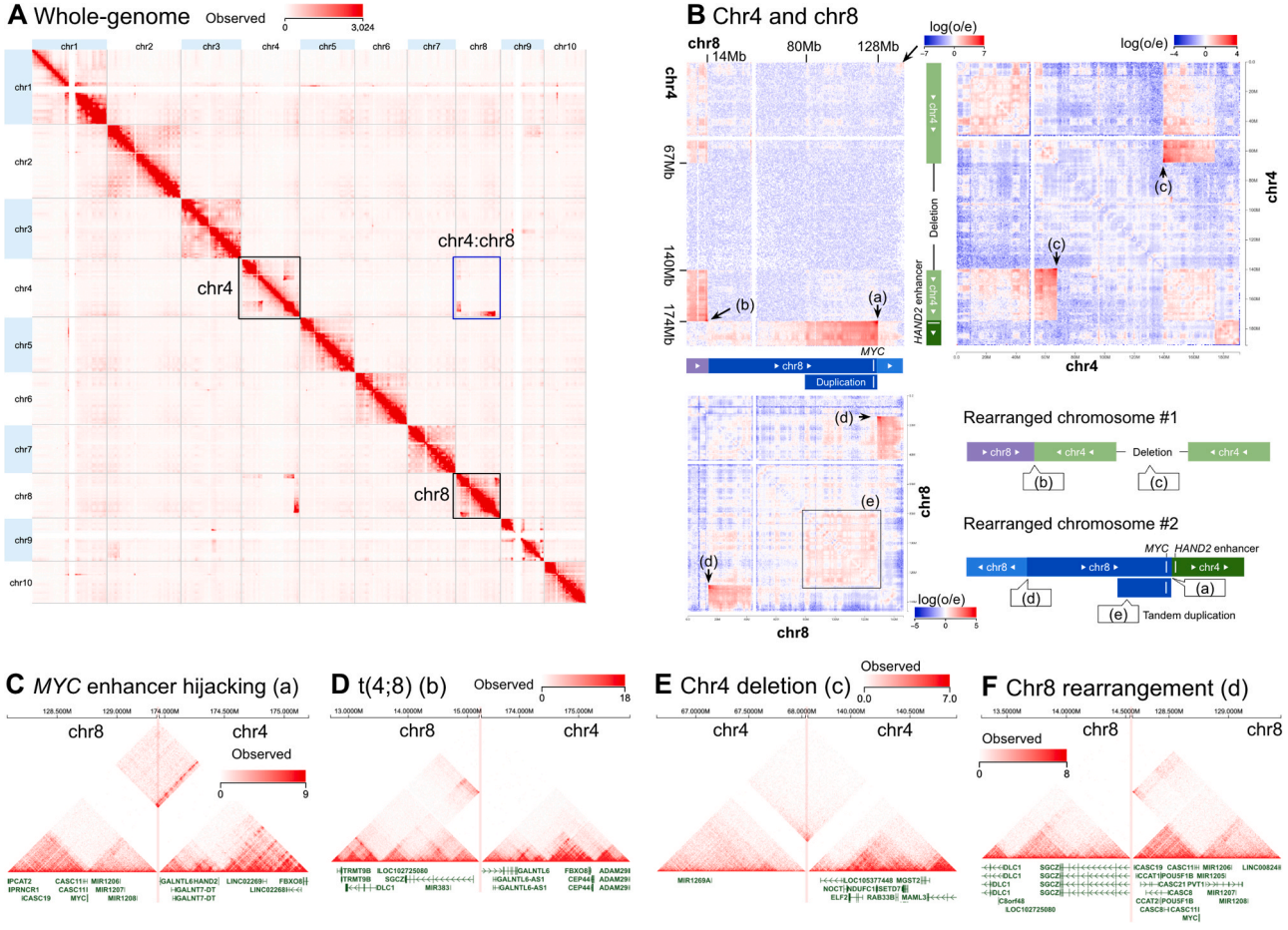

Fig. 1. ppHiC visualization of Hi-C data illustrates genomic rearrangements in NBL cell line NB69. (A) Genome-wide view of intra- and inter-chromosomal observed interactions, from chr1 up to chr10. See Fig. S1A for whole-genome view. Chr4-chr8 (blue box), intra-chr4 and intra-chr8 (black boxes) interactions are highlighted. (B) Intra- and inter-chromosomal interactions of chr4 and chr8 and a model of rearranged chromosomes. Hi-C maps (upper left, upper right, bottom left) show log (observed/expected) values. Top left: inter-chr4-chr8 interactions. Top right: intra-chr4 interactions. Bottom left: intra-chr8 interactions. Genomic events are marked on Hi-C maps with coordinates indicated on inter-chr4-chr8 Hi-C map: (a) translocation resulting in *MYC* enhancer hijacking, (b) novel t(4;8) translocation, (c) intra-chr4 deletion, (d) intra-chr8 rearrangement, (e) tandem duplication in chr8. Between Hi-C maps: chr4 and chr8 diagrams, segments produced by chromosomal breakage are color-coded and shown to-scale. Arrows indicate forward strand. White bands indicate *MYC* gene and *HAND2* enhancer. Bottom right: models of two rearranged chromosomes. Callouts indicate genomic events. Left/right arrows indicate forward/reverse strands. Segment lengths are not to scale. (C) Genome browser Hi-C track view of event (a). (D) Genome browser Hi-C track view of event (b). (E) Genome browser Hi-C track view of event (c). (F) Genome browser Hi-C track view of event (d).

clinical data, making the ability to share Hi-C data particularly useful. Combined, these features make ppHiC a powerful tool for researchers seeking to understand the complex relationships between chromatin structure and biological function.

3.3. Use case

Dysregulation of the *MYC* oncogene is a hallmark of tumorigenesis in pediatric neuroblastoma (NBL) and is frequently associated with aggressive disease progression. Aberrant *MYC* expression affects cell proliferation, differentiation, and apoptosis, making it a critical determinant of NBL molecular pathology [14]. Tumor-derived cell lines are excellent models of the complex biological and genetic drivers of cancer progression. A recent study reported hijacking of the *MYC* enhancer through a t(4;8) translocation in primary NBL tumors as well as the NBL cell line NB69 [14]. Here, we use ppHiC to investigate the in-situ Hi-C results for the NB69 cell line, uncovering an intricate global landscape of altered chromatin interactions from which we were able to identify an extensive genomic rearrangement.

In the ppHiC Genome View (Fig. 1A), we observed various altered chromatin contacts, most notably the inter-chromosomal contacts between chr4 and chr8, as well as long-range intra-chromosomal contacts for both chr4 and chr8. We used the Chromosome Pair View and Genome Browser View to explore these events in detail.

In the Chromosome Pair View, the inter-chr4-chr8 interaction comprises two translocation events. The first event (Fig. 1B, point (a)) involves a connection between the centromeric region of chr8:128 Mb (a position on chromosome 8 that is approximately 128 megabases from the start of the chromosome) where *MYC* resides (Fig. 1B, dark blue box), and the telomeric region of chr4:174 Mb (a position on chromosome 4 that is approximately 174 megabases from the start of the chromosome), where the *HAND2* enhancer resides (Fig. 1B, dark green box). Examination of this event in the Genome Browser View reveals extensive contact between *MYC* and the *HAND2* enhancer (Fig. 1C) that results in the aberrant *MYC* expression driving tumorigenesis [14].

In the second t(4;8) translocation event (Fig. 1B, point (b)), one breakpoint collocates with the *MYC* activation event at chr4:174 Mb. Through this event, the telomeric region of chr8:14 Mb (a position on chromosome 8 that is approximately 14 megabases from the start of the chromosome) (Fig. 1B, purple box) is joined with the centromeric region of chr4:174 Mb (Fig. 1B, light green box). Details of this event are further illustrated in the Genome Browser View provided in Fig. 1D.

The intra-chromosomal contact maps of both chr4 and chr8 show unexpected long-range contacts. In chr4, a contact forms over a 73 Mb distance (Fig. 1B, point (c); 67 Mb to 140 Mb of light green box, Fig. 1E). The reduced proximal contact inside chr4:67Mb-140Mb indicates a one-copy genomic deletion. This is supported by matching structural variation (SV) breakpoints and copy number loss calls previously discovered in the whole-genome sequencing data of NB69 cells (Fig. S2A) [15]. By assessing the chr4 deletion with the novel t(4;8) translocation, the chr4 deletion region shows an absence of contact with the telomeric region of chr8:14 Mb (Fig. S2B and S2C), indicating that the intra-chr4 deletion is *in-cis* with the novel t(4;8) translocation and the intact chr4 is not *in-cis* with chr8.

In chr8, a contact forms over a 114 Mb distance (Fig. 1B, point (d), 14 Mb to 128 Mb of dark blue box, Fig. 1F). The lack of copy number gain calls over chr8:14Mb-128Mb (Fig. S3A) and the absence of proximal contact around the 14 Mb position in chr8 (Fig. S3B) indicates a joining of the centromeric region of chr8:14 Mb (Fig. 1B, dark blue box) with the centromeric region of chr8:128 Mb (Fig. 1B, light blue box) rather than a tandem duplication of the entire 114 Mb segment. Within that segment, an increase in copy number indicates a 48 Mb tandem duplication that includes the *MYC* gene (Fig. S3A), which is also supported by increased proximal contact (Fig. 1B, box (e)). Further, the translocation underlying *MYC* activation displayed increased chr4-chr8 contact over the chr8 duplicated region compared to the non-duplicated region (Fig. S3C), suggesting that the chr8 tandem duplication is *in-cis* with the *MYC* activating translocation, where the duplicated copy of *MYC* is proximal to the *HAND2* enhancer. Further, the telomeric region of chr4:174 Mb (Fig. 1B, dark green box) and the telomeric region of chr8:128 Mb (Fig. 1B, light blue box) show weak contact (Fig. S3C), suggesting that the two segments are *in-cis* but distal from each other. Taken together, this evidence allows us to construct two rearranged chromosomes using chr4 and chr8 segments delineated by genomic breakpoints (Fig. 1B), explaining the abnormal chromatin contact patterns observed in the ppHiC visualizations.

In addition to the complex rearrangements of chr4 and chr8, ppHiC allowed us to characterize translocation events between chromosomes 2 and 11 (Fig. S1B and C) and between chromosomes 1 and 17 (Fig. S1D and E), but no oncogene activation was observed in those cases.

4. Conclusion

ppHiC overcomes common challenges of visualizing intricate chromatin conformation data, providing a platform that is accessible to researchers who may lack the computational resources and expertise previously required to analyze Hi-C datasets. As a web-based application, it allows researchers to easily access and share data that would otherwise be cumbersome to handle and disseminate. The tool's comprehensive features and user-friendly design streamline the analysis process while providing a high level of customization.

ppHiC is a valuable addition to the genomics research toolkit that will accelerate discoveries in the field of chromatin conformation and enhance understanding of the regulatory mechanisms governing the genome. We are committed to the continuous improvement of ppHiC and plan to regularly update and enhance its software features. This ongoing development will ensure ppHiC's place at the forefront of Hi-C data visualization and analysis tools. We encourage the scientific community to not only adopt ppHiC but also contribute to its development by adding new features, improving existing functionalities, and integrating additional datasets. Together we will advance our shared understanding of the genome's architecture and activity.

CRediT authorship contribution statement

Xin Zhou: Writing – review & editing, Writing – original draft, Funding acquisition. **Aleksandar Acic:** Software. **Karishma Gangwani:** Software. **Robin Paul:** Software. **Congyu Lu:** Software. **Edgar Sioson:** Software. **Gavriel Matt:** Software. **Airen Zaldivar Peraza:** Software. **Jian Wang:** Software. **Akanksha Rajput:** Writing – review & editing, Writing – original draft, Software. **Colleen Reilly:** Software.

Declaration of Competing Interest

Authors declare no competing interest.

Availability

ppHiC live examples, and user tutorial are available at https://proteinpaint.stjude.org/hic/. Source code is available at https://github.com/stjude/proteinpaint.

Acknowledgement

This work was supported by American Lebanese Syrian Associated Charities (ALSAC), the Pediatric Cancer Dependencies Accelerator of the Broad Institute, Dana-Farber Cancer Institute, and St. Jude Children's Research Hospital: peddep.org. We thank Sarah Frank for proofing the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.09.020.

References

[1] Yardımcı GG, Noble WS. Software tools for visualizing Hi-C data. Genome Biol 2017;18:26.
[2] Lévy-Leduc C, Delattre M, Mary-Huard T, Robin S. Two-dimensional segmentation for analyzing Hi-C data. Bioinformatics 2014;30:i386–92.
[3] Zhang S, Chasman D, Knaack S, Roy S. In silico prediction of high-resolution Hi-C interaction matrices. Nat Commun 2019;10:5449.
[4] Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst 2016;3:99–101.
[5] Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. Genome Biol 2018;19:125.
[6] Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D Genome browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol 2018;19:151.
[7] Pal K, Ferrari F. Visualizing and annotating Hi-C Data. Methods Mol Biol 2022; 2301:97–132.
[8] Zhou X, Edmonson MN, Wilkinson MR, Patel A, Wu G, Liu Y, et al. Exploring genomic alteration in pediatric cancer using ProteinPaint. Nat Genet 2016;48:4–6.
[9] Matt GY, Sioson E, Shelton K, Wang J, Lu C, Zaldivar Peraza A, et al. St. Jude Survivorship Portal: sharing and analyzing large clinical and genomic datasets from pediatric cancer survivors. Cancer Discov 2024. https://doi.org/10.1158/2159-8290.CD-23-1441.
[10] Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, et al. Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res 2020;48: W177–84.
[11] Di Filippo L, Righelli D, Gagliardi M, Matarazzo MR, Angelini C. HiCeekR: a novel Shiny App for Hi-C data analysis. Front Genet 2019;10:1079.
[12] Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods 2012;9:999–1003.
[13] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 2009;326:289–93.
[14] Zimmerman MW, Liu Y, He S, Durbin AD, Abraham BJ, Easton J, et al. Drives a subset of high-risk pediatric neuroblastomas and is activated through mechanisms including enhancer hijacking and focal enhancer amplification. Cancer Discov 2018;8:320–35.
[15] Zhou X, Wang J, Patel J, Valentine M, Shao Y, Newman S, et al. Exploration of coding and non-coding variants in cancer using GenomePaint. Cancer Cell 2021; 39:83–95. e4.