

RESEARCH ARTICLE

HiC-3DViewer: a new tool to visualize Hi-C data in 3D space

Mohamed Nadhir Djekidel¹, Mengjie Wang², Michael Q. Zhang^{3,1,*} and Juntao Gao^{1,*}

¹ MOE Key Laboratory of Bioinformatics; Bioinformatics Division and Center for Synthetic & Systems Biology, TNLIST; Department of Automation, Tsinghua University, Beijing 100084, China

² School of Pharmaceutical Sciences, Tsinghua University, Beijing 100084, China

³ Department of Biological Sciences, Center for Systems Biology, The University of Texas at Dallas, Richardson, TX 75080-3021, USA

* Correspondence: michaelzhang@tsinghua.edu.cn, jtgao@biomed.tsinghua.edu.cn

Received July 9, 2016; Accepted August 26, 2016

Background: Although significant progress has been made to map chromatin structure at unprecedented resolution and scales, we are short of tools that enable the intuitive visualization and navigation along the three-dimensional (3D) structure of chromatin. The available tools people have so far are generally script-based or present basic features that do not easily enable the integration of genomic data along with 3D chromatin structure, hence, many scientists find themselves in the obligation to hack tools designed for other purposes such as tools for protein structure study.

Methods: We present HiC-3DViewer, a new browser-based interactive tool designed to provide an intuitive environment for investigators to facilitate the 3D exploratory analysis of Hi-C data along with many useful annotation functionalities. Among the key features of HiC-3DViewer relevant to chromatin conformation studies, the most important one is the 1D-to-2D-to-3D mapping, to highlight genomic regions of interest interactively. This feature enables investigators to explore their data at different levels/angles. Additionally, investigators can superpose different genomic signals (such as ChIP-Seq, SNP) on the top of the 3D structure.

Results: As a proof of principle we applied HiC-3DViewer to investigate the quality of Hi-C data and to show the spatial binding of GATA1 and GATA2 along the genome.

Conclusions: As a user-friendly tool, HiC-3DViewer enables the visualization of inter/intra-chromatin interactions and gives users the flexibility to customize the look-and-feel of the 3D structure with a simple click. HiC-3DViewer is implemented in Javascript and Python, and is freely available at: <http://bioinfo.au.tsinghua.edu.cn/member/nadhir/HiC3DViewer/>. Supplementary information (User Manual, demo data) is also available at this website.

Keywords: Hi-C; 3D genome visualization; chromatin structure prediction

INTRODUCTION

The breakthrough in chromatin conformation study made in the last decade enabled us to understand many aspects of gene transcription mechanisms and revealed the important role of chromatin crosstalk during this process [1–4]. One of the mainly used chromatin conformation study techniques is Hi-C [1], which enables the mapping of genome-wide interactions to identify structural features and interactions among all chromosomes, by generating a matrix of interaction frequencies that can be represented

as two-dimensional heat maps with genomic positions along the two axes. This information intrigued investigator to develop 3D genome prediction methods to infer the underlying 3D chromatin structure in order to understand the transcription mechanisms behind chromatin organization [5–7].

A key limitation of the currently existing chromatin conformation visualization and modeling tools is their extensive reliability on scripting knowledge such as the different Hi-C 2D-based analysis packages available in Bioconductor [8,9]. Even the 3D visualization tools that

provide user friendly interfaces are very limited in terms of functionalities and manipulation flexibility (Table 1). Generally, the existing chromatin 3D structure study tools fall into one of two categories: 3D structure prediction or visualization. The 3D structure prediction tools [5,7] generally produce outputs compatible with the Protein Data Bank (PDB) format designed to be used with protein visualization tools, such as PyMol [10]. However, the PDB format is designed to represent atoms and amino acid residues and the usage of this format for chromatin 3D structure representation leads to the loss of genomic coordinate information and makes it challenging to overlay genomic data (such as genes positions, ChIP-Seq signal) on the top of the 3D structure. On the other hand, chromatin 3D visualization tools are generally available as databases where users can just visualize pre-calculated 3D models and do not have the ability to upload their own data [13,14].

To fill this gap, we developed HiC-3DViewer, a new interactive chromatin visualization tool designed to provide an intuitive environment for investigators that facilitates the 3D exploratory analysis of Hi-C data and enables 3D chromatin structure prediction. Beside visualization, HiC-3DViewer enables investigators to interactively highlight any specific genomic regions through the usage of an interactive 2D Hi-C contact map or by uploading a BED file of given regions. Additionally, HiC-3DViewer allows users to visualize inter- and intra-chromatin interactions and to change the look-and-feel of the 3D chromatin model for more comfortable visualization.

RESULTS

Design and implementation

As a browser-based interactive 3D chromatin visualization tool, HiC-3DViewer is designed based on a client-server architecture (Figure 1), from which the client exploits the features of HTML5 and WebGL to achieve data visualization. All 3D displays are based on the BufferGeometry class provided by the Three.js javascript library [15] which enables the display of large genomes with reasonable memory and CPU consumption. The

server-side, on the other hand, is implemented in Python and based on the Flask micro-framework [16], that enables the establishment of a very lightweight web-server without any additional configuration. The Flask micro-framework can either run as standalone application on a single user computer or be configured to work with a webserver such as APACHE without the need to do any code changes.

The server side is mainly responsible for data storage and 3D prediction. Two types of data are available on the server: (i) pre-built 3D models for yeast [17], human [1] and *Drosophila* [18] genomes and (ii) the user's own uploaded data. The 3D prediction is built on the top of the software package Pastis [5] that implements four 3D structure prediction algorithms—two based on multi-dimensional scaling algorithms and the other two optimizing the likelihood of a Poisson-based model. Additionally, the server implements some modules to do file format conversion and to manipulate genomic ranges.

The browser-based client is responsible for the 3D model and 2D Hi-C contact map visualization. Different menus are available to customize the display of chromosomes in order to help investigators focus their studies on the chromosomes of interest (Supplementary Figure S1).

HiC-3DViewer navigation environment

The viewer provides a single-page environment that embeds different features in three panels according to their roles (Figure 2A). Features handling data input and output are all on the top-left panel. In this panel, investigators are able to select some already pre-calculated 3D models (from species such as Human, Yeast and *Drosophila*) for display and visualization. If user has uploaded her/his own model, she/he will also be able to select it from the model selection section. Basic information such as name and resolution is available for each model. The time required for loading a model depends on the size of the Hi-C data. We also compiled a set of cytobands for different species that can be loaded by the user along with the 3D model. Cytoband information for some species such as Yeast is not available so it will be displayed as a white band in the viewer.

The latest technological advancement in chromatin

Table 1. Comparison of some widely used tools for 3D chromatin visualization.

Tool	Specific to 3D genome display	3D model display	Cis-interaction display	Trans-interaction display	Custom annotation	Need scripting
PyMol [10]	×	✓	×	×	✓	✓
GMol [11]	✓	✓	×	×	×	×
Genome3D [12]	✓	✓	×	×	✓	×
Tadkit [6]	✓	✓	×	×	✓	×
HiC-3DViewer	✓	✓	✓	✓	✓	×

Comparison between the features of HiC-3DViewer and some chromatin 3D visualization tools. ✓ indicates yes, and × means no.

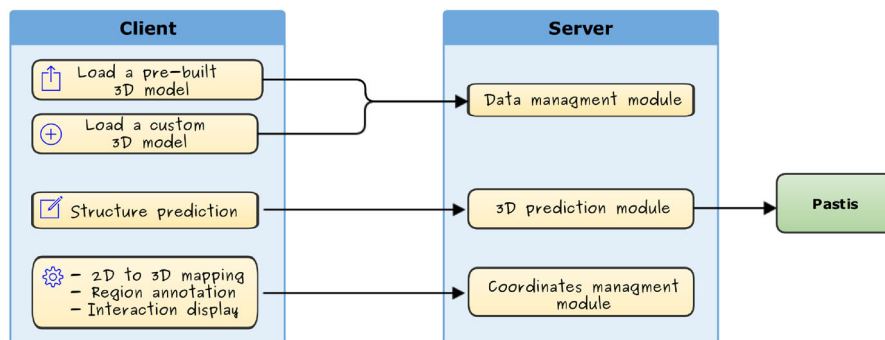


Figure 1. Client-server architecture of HiC-3DViewer. The client uses HTML 5 WebGL features for the 3D visualization. The server (implemented in python) is responsible for data storage and data processing.

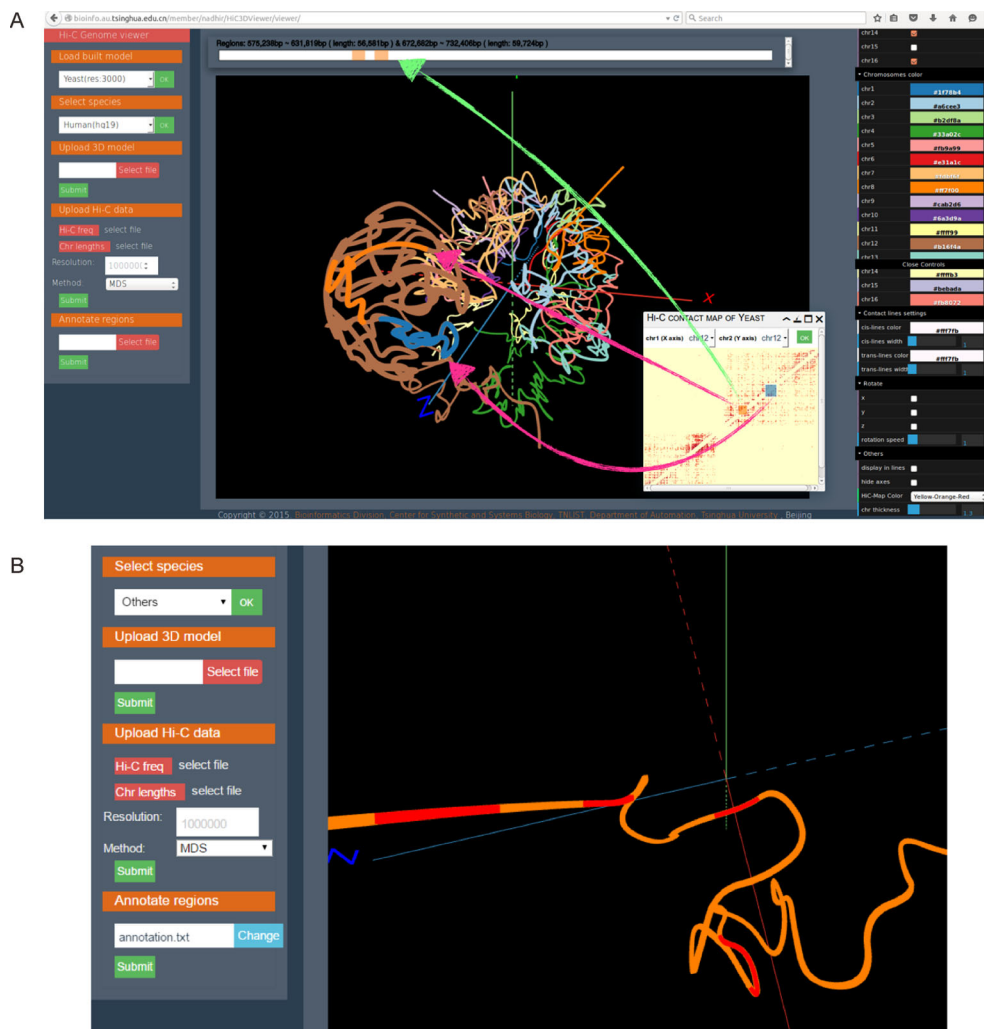


Figure 2. Display of some of HiC-3DViewer functionalities. (A) A screenshot showing the main interface of HiC-3DViewer with the Yeast genome displayed. Pink arrowheads indicate the 2D to 3D mapping of the highlighted region in Hi-C heat map on the 3D model. The green arrowhead indicates the 2D to 1D mapping of the selected region in 1D yeast genome (shown in orange for the recently selected region). Each selected region will be mapped to an individual 1D genome. The 1D genome of the highlighted regions can be browsed using the scrollbar on the right of the 1D genome window. If no cytoband information is available, the genome will be shown as a white bar. (B) An example showing the customized annotation feature. Here a file (bottom-left file input box) is uploaded that contains the genomic coordinates of some regions to be highlighted (shown in red). This feature is another alternative to the interactive Hi-C matrix which can be used for bulk annotation of genomic regions such as ChIP-Seq peaks, genes.

conformation study enabled investigators to mine chromatin fiber cross-talk at an unrepresented scale [19] and intrigued investigators to develop 3D prediction methods to crack the mechanisms underlying chromatin interaction establishment. Thus, in order not to create “yet another 3D genome prediction algorithm” we give investigators the ability to upload and visualize their own predicted 3D model using their preferred 3D genome prediction algorithm, with the only prerequisite that the model should be provided in the form of a tabulated text file (see User Manual).

Alternatively, HiC-3DViewer enables users to predict the 3D structure of their own Hi-C data. The 3D structure prediction is based on the 3D model prediction algorithm developed by N. Varoquaux *et al.* [5] available in the Pastis package which implements four 3D model prediction algorithms: two based on the multidimensional scaling (MDS) algorithm and the other two statistical models assuming that the counts between two loci follow a Poisson distribution with decreasing intensity between distant loci. The prediction speed depends on the data size, the specified resolution and the algorithm selected, with the MDS algorithm as the fastest one and the Poisson distribution ones the slowest. To add more flexibility, different formats for Hi-C data are supported for upload, ranging from the matrix format to the 3-columned format.

Once the model is loaded, a couple of panels are displayed on the right side of the screen (Figure 1A), with a control panel displayed on the top-right (Supplementary Figure S1 and text of the User Manual of HiC-3DViewer). The control panel enables investigators to customize the display of every chromosome, such as the color, thickness, visibility, 3D rotation, the display of inter- and intra-chromatin interactions and the customization of the color map used to plot the Hi-C 2D heat map.

On the top-down side, a floating interactive panel with Hi-C 2D contact map is displayed which enables investigators to use the mouse to select regions on the Hi-C heat map and to highlight the corresponding 3D positions (Figure 2A, Supplementary Figure S2). Users can also highlight regions by uploading a 4-columned BED file on the “region annotation” section.

1D to 2D to 3D mapping

As depicted in the previous sections, one of the key features relevant to chromatin conformation studies is the interactive 1D to 2D to 3D mapping between 1D genomic sequence, 2D Hi-C contact map and the 3D genome model (Figure 2A, Supplementary Figure S2). Though simple, this feature is lacking in all existing Hi-C exploratory tools. Furthermore, this feature is very practical, as in many cases investigators would like to make a correspondence between the Hi-C contact matrix

and the 3D model, for example to check the spatial proximity between two genomic regions. If a cytoband is selected, then one dimensional mapping on the cytoband will be performed for each annotated region (see the cytobands at the top of Supplementary Figure S2). For the case that the cytoband for the selected species is not available, a white band will be displayed (Figure 2A).

Another point of interest for investigators is to study the spatial clustering of genomic signals or a group of genes or mutations. Investigators can upload a custom BED file of the genomic regions to highlight signals such as ChIP-Seq peaks, genes, SNPs (Figure 2B, Supplementary Figure S3). If the user associates a score with each region, a gradient color “Yellow-Orange-Red” will be used to highlight the regions, with the lowest scores colored in yellow and larger scores in red.

Gene annotation and distance estimation

Another option of interest for investigators is the highlight of genes located in a region of interest and measuring the distance between different loci. Figure 3 shows an example in which the genes located in the neighborhood of a selected region are displayed along with a small panel that displays information about them. The gene info panel gives the names of the genes with a link to UCSC, their position and link to papers (if any) in which the name of the gene and the fluorescent *in situ* hybridization (FISH) experiment are mentioned.

Exploration of inter- and intra-chromatin interactions

Chromatin conformation studies revealed that chromatin interactions present non-random organizational structures in which intra-chromatin interactions tend to show clustered regions known as Topologically Associated Domains (TADs) [20], while existing technologies can not tell us about the true signal-to-noise ratio presented in the inter-chromatin interactions. Nonetheless, this information is exploited by 3D prediction algorithms to position chromosomes in the 3D space.

In HiC-3DViewer we enable investigators to visualize *cis*- and *trans*-interactions so they can assess the accuracy of the 3D predictions by checking if highly connected elements are positioned in close spatial positions. It also enables users to get some hints about how chromatin topological structures, such as domains or higher hierarchies, are organized, depending on the resolution. Investigators can customize the color and the thickness of the displayed interactions to get a better view or better screenshots (Figure 4).

In Figure 4A, we show the intra-chromatin interactions in chromosome 2 of Yeast genome. Here, we set

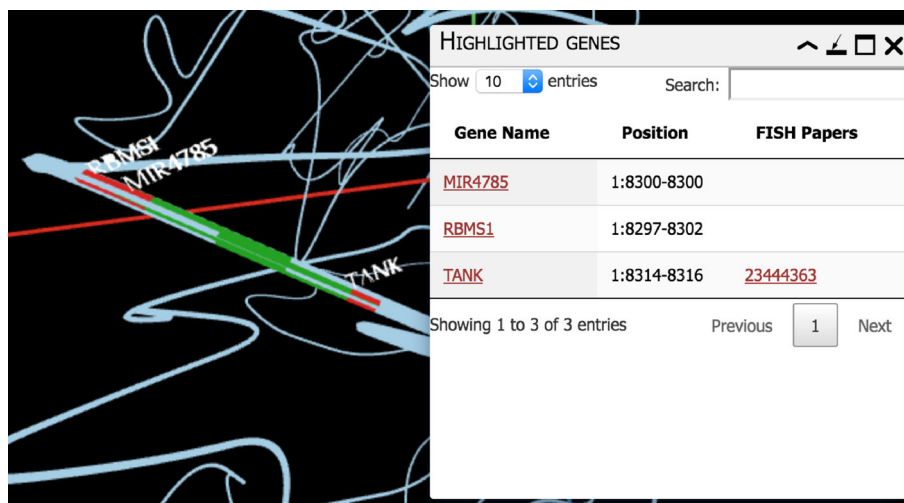


Figure 3. Gene annotation functionality. Investigators have the potential to select genomic region (shown here in green) by a mouse click. The genes located in the proximity of the selected region will be highlighted in red. The name, the position and the potential FISH experiments related papers of the neighboring genes will be displayed in a floating panel. The gene names are linked to the UCSC browser and the papers are linked to pubmed.

chromatin interactions to be thicker and with yellow color. The transparency of each line indicates the interaction frequency strength. We can see that genomic regions located in a short genomic distance have brighter colors than the ones located far away. This observation can also be made for the case of *trans*-interactions. In Figure 4B, we display chromatin interactions between chromosomes 2 and 5. We also notice that regions with shorter spatial proximity have brighter interaction lines than the further ones.

The examination of inter- and intra-chromatin interactions can be useful to check the accuracy of the 3D prediction model. Even though it is quantitative, it gives many hints on the behavior of 3D model prediction algorithm.

3D visualization as a quality control step for 3C-based experiments

The use of genome 3D models in elaborating biological hypothesis is not a new practice. For example, Ay *et al.* [21] exploited the 3D model inferred from Hi-C data to understand the genome structure in initial stages of malaria parasite infections. Other groups such as Bau *et al.* [6] used FISH data to build a 3D model of the alpha-globin locus and investigated the chromosomal changes in different cell lineages.

We also argue that genome 3D model can give many hints on the quality of 3C-based experiments. To demonstrate this, we used HiC-3DViewer to evaluate the Hi-C experiment quality with two different samples (Figures 2A and 5A). Figure 2A shows the 3D model of

the yeast genome predicted from good quality data generated from 3C-based technique developed by Duan *et al.* [17]. The clustering of the centromeres of chromosomes at one pole of the nucleus can be clearly noticed. However, in the noisy yeast 3D conformation data (Figure 5A), obtained from a failed Hi-C experiment data, we can not see such organization.

Annotation of GATA1 and GATA2 spatial binding

In the work published in 2012, Lan *et al.* [22] combined Hi-C interaction data with the ChIP-Seq signal of 45 transcription factors and 9 histone modifications. Their results showed that the GATA-factors (GATA1 and GATA2) tend to co-bind to a specific subset of loci, indicating the aggregation of the interacting loci into different interaction hubs.

Therefore, we overlaid GATA1 and GATA2 signals onto the top of the 3D structure of human genome, as one potential visualization example for HiC-3DViewer application. Figure 5B shows the binding of GATA1 (in yellow) and GATA2 (in red) on chromosome 11. The clustering of the binding signals of the two factors can be clearly visualized. By close examination we notice that these factors tend to cluster into the chromatin curvatures that show a short spatial proximity.

DISCUSSION

Structure determines function. The three-dimensional chromatin organization in cell nucleus has an essential regulatory function on the activity of gene expression,

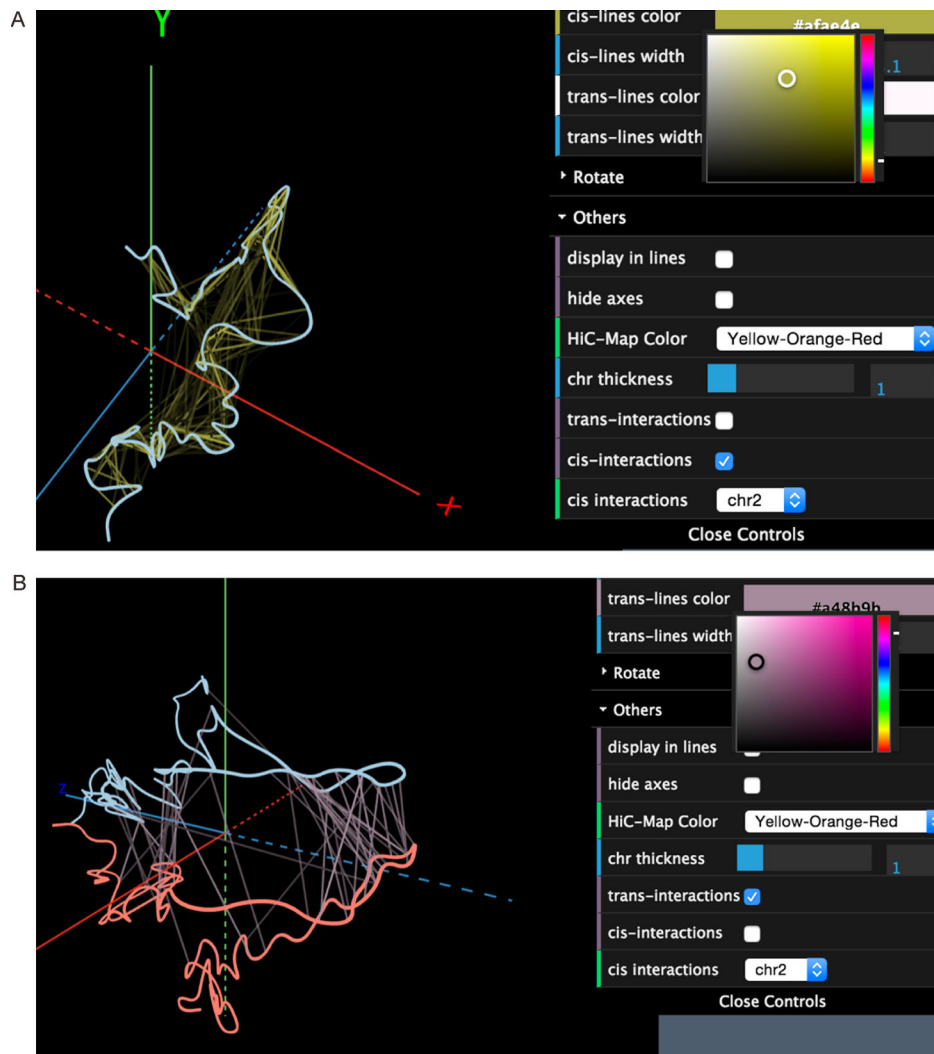


Figure 4. Cis- and trans-chromatin interactions and their control panel. Intra- and inter-chromatin interactions are displayed in the form of white straight lines. The transparency of the lines indicates the strength (or frequency) of the interactions. (A) Users can customize the color and the width of *cis*-chromatin interactions to get a better view. The intra-chromatin interactions of the selected chromosome are displayed. (B) *Trans*-chromatin interactions can also be customized. In this panel, only *trans*-interactions between visible chromosomes, but not intra-chromatin interactions, are displayed.

transcription, replication, gene repair and other biological processes. The study of the 3D structure (referred to as 3D genomics) and function of the whole genome helps us understand the mechanism about how genes are guided, get expressed and transcribed in a particular space and time.

HiC-3DViewer provides such an environment for us to visualize and evaluate how chromatins are positioned in 3D space, with the corresponding 1D genomic sequence information and 2D Hi-C contact matrix. As a novel interactive tool that enables the intuitive exploration of chromatin structure, HiC-3DViewer is a highly customizable tool that requires no expert knowledge in 3D genome

prediction algorithms on the user side. Users have the ability to select and utilize one of the four available algorithms or they can upload their own 3D model for visualization and annotation. We also showed that HiC-3DViewer can be used to examine the quality of chromatin structure experiments. Additionally, the presented tool can be used to explore the spatial position of different genomic regions such as genes, SNPs and ChIP-Seq signals. We also integrated some features that enable users to annotate a large number of regions by uploading a BED file. The consideration, integration and visualization of the 1D genome sequence, 2D interaction matrix, and 3D gene structure and regulatory elements at the same

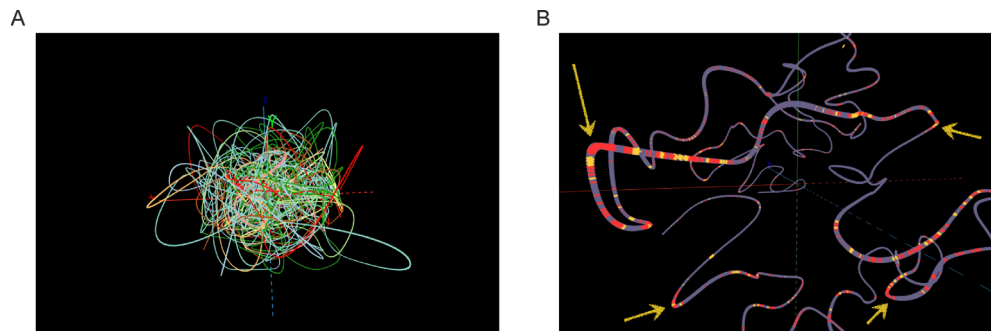


Figure 5. The potential applications of HiC-3DViewer. (A) HiC-3DViewer can be used for the quality control of 3C-based experiments. Here the 3D model of a noisy yeast Hi-C heat map was uploaded. The non-structured shape can be seen very easily, compared to the 3D model in Figure 1A, which has good quality. (B) Annotation of the binding sites of GATA1 and GATA2. The yellow arrowheads indicate some regions that show the clustering of GATA1 (in yellow) and GATA2 (in red) signals at short spatial proximity, confirming previous findings [22].

time will provide novel insights into the genome regulatory functions.

METHODS

3D genome model construction for different species

The yeast 3D model was downloaded from the original paper by Duan *et al.* [17]. For the human genome, we at first normalized the Hi-C contact map published by Lieberman *et al.* [1] using the HiCNorm method [23], then did 3D prediction using Pastis [5]. For *Drosophila* genome, the chromatin interaction map [18] was downloaded at first, then HiCNorm and Pastis were used for data normalization and 3D model prediction, respectively.

GATA1 and GATA2 ChIP-Seq signals

ChIP-Seq data provided by ENCODE consortium for GATA1 (GSM1003608) and GATA2 (GSM935373) was used in this study.

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at DOI 10.1007/s40484-017-0091-8.

ACKNOWLEDGMENTS

This work is supported by State Key Research Development Program of China (No. 2016YFC1200303), and the National Natural Science Foundation of China (Nos. 31361163004 and 31671383). We thank Yanjian Li (Tsinghua University) for sharing his Hi-C data. MQZ was partially supported by UTD funds.

COMPLIANCE WITH ETHICS GUIDELINES

The authors Mohamed Nadhir Djekidel, Mengjie Wang, Michael Q. Zhang,

and Juntao Gao declare that they have no conflict of interests. All the data sets the authors used are from public repositories.

REFERENCES

- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragozy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326, 289–293
- Li, G., Ruan, X., Auerbach, R. K., Sandhu, K. S., Zheng, M., Wang, P., Poh, H. M., Goh, Y., Lim, J., Zhang, J., *et al.* (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell*, 148, 84–98
- Sanyal, A., Lajoie, B. R., Jain, G. and Dekker, J. (2012) The long-range interaction landscape of gene promoters. *Nature*, 489, 109–113
- Göndör, A. and Ohlsson, R. (2009) Chromosome crosstalk in three dimensions. *Nature*, 461, 212–217
- Varoquaux, N., Ay, F., Noble, W. S. and Vert, J.-P. (2014) A statistical approach for inferring the 3D structure of the genome. *Bioinformatics*, 30, i26–i33
- Baù, D., Sanyal, A., Lajoie, B. R., Capriotti, E., Byron, M., Lawrence, J. B., Dekker, J. and Marti-Renom, M. A. (2011) The three-dimensional folding of the α -globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.*, 18, 107–114
- Wang, S., Xu, J. and Zeng, J. (2015) Inferential modeling of 3D chromatin structure. *Nucleic Acids Res.*, 43, e54
- Thongjuea, S., Stadhouders, R., Grosveld, F. G., Soler, E. and Lenhard, B. (2013) r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. *Nucleic Acids Res.*, 41, e132
- Phanstiel, D. H., Boyle, A. P., Araya, C. L. and Snyder, M. P. (2014) Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics*, 30, 2808–2810
- Schrödinger, LLC (2010) The PyMOL Molecular Graphics System, Version 1.3r1. Available: <http://www.pymol.org>
- Nowotny, J., Wells, A., Xu, L., Cao, R., Trieu, T., He, C., Cheng, J.

- (2016) GMOL: an interactive tool for 3D genome structure visualization. *Sci. Rep.* 6, 20802
12. Asbury, T. M., Mitman, M., Tang, J. and Zheng, W. J. (2010) Genome3D: a viewer-model framework for integrating and visualizing multi-scale epigenomic information within a three-dimensional genome. *BMC Bioinformatics*, 11, 444
 13. Peng, C., Fu, L.-Y., Dong, P.-F., Deng, Z.-L., Li, J.-X., Wang, X. T. and Zhang, H. Y. (2013) The sequencing bias relaxed characteristics of Hi-C derived data and implications for chromatin 3D modeling. *Nucleic Acids Res.*, 41, e183
 14. Teng, L., He, B., Wang, J. and Tan, K. (2015) 4DGenome: a comprehensive database of chromatin interactions. *Bioinformatics*, 31, 2560–2564
 15. Dirksen, J. (2013) *Learning Three.js: The JavaScript 3D Library for WebGL*. Birmingham: Packt Publishing
 16. Grinberg, M. (2014) *Flask Web Development*. Sebastopol: O'Reilly Media
 17. Duan, Z., Andronescu, M., Schutz, K., McIlwain, S., Kim, Y. J., Lee, C., Shendure, J., Fields, S., Blau, C. A. and Noble, W. S. (2010) A three-dimensional model of the yeast genome. *Nature*, 465, 363–367
 18. Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A. and Cavalli, G. (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell*, 148, 458–472
 19. Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159, 1665–1680
 20. Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485, 376–380
 21. Ay, F., Bunnik, E. M., Varoquaux, N., Bol, S. M., Prudhomme, J., Vert, J. P., Noble, W. S. and Le Roch, K. G. (2014) Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res.*, 24, 974–988
 22. Lan, X., Witt, H., Katsumura, K., Ye, Z., Wang, Q., Bresnick, E. H., Farnham, P. J. and Jin, V. X. (2012) Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic Acids Res.*, 40, 7690–7704
 23. Hu, M., Deng, K., Selvaraj, S., Qin, Z., Ren, B. and Liu, J. S. (2012) HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics*, 28, 3131–3133