Methods 170 (2020) 69–74

Contents lists available at ScienceDirect

Methods

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

Methods for comparative ChIA-PET and Hi-C data analysis

Dan Capurso[a], Zhonghui Tang[a], Yijun Ruan[a,b,*]

[a] The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA
[b] Department of Genetics and Genome Sciences, University of Connecticut Health Center, Farmington, CT, USA

ABSTRACT

The three-dimensional architecture of chromatin in the nucleus is important for genome regulation and function. Advanced high-throughput sequencing-based methods have been developed for capturing chromatin interactions (Hi-C, genome-wide chromosome conformation capture) or enriching for those involving a specific protein (ChIA-PET, chromatin interaction analysis with paired-end tag sequencing). There is widespread interest in utilizing and interpreting ChIA-PET and Hi-C. We review methods for comparative ChIA-PET and Hi-C data analysis and visualization. The topics reviewed include: downloading ChIA-PET and Hi-C data from the ENCODE and 4DN portals; processing ChIA-PET data using ChIA-PIPE; processing Hi-C data using Juicer or distiller and cooler; viewing 2D contact maps using Juicebox or Higlass; viewing peaks, loops, and domains using BASIC Browser; annotating convergent and tandem CTCF loops.

1. Introduction

The three-dimensional (3D) organization of chromatin in the nucleus is important for genome regulation and function. Advanced high-throughput sequencing-based methods have been developed for capturing chromatin interactions (Hi-C, genome-wide chromosome conformation capture) [21] or enriching for those involving a specific protein (ChIA-PET, chromatin interaction analysis with paired-end tag sequencing) [12,20]). ChIA-PET and Hi-C have been extensively applied and have uncovered topological chromatin domains with frequent interactions, and protein-mediated chromatin loops involved in genome architecture and gene regulation [19,24,28,25,29].

Hi-C experiments are performed as follows [21]. First, cells are subjected to single cross-linking (to stabilize protein-DNA interactions), cells are lysed, and chromatin is digested with a restriction enzyme. The DNA overhangs are filled with a biotinylated nucleotide, and proximity ligation is performed under dilute conditions to capture chromatin interactions. The junctions of chromatin interactions are extracted by reverse cross-linking, DNA shearing, pull down on streptavidin beads, and PCR amplification, and then the sample goes on to sequencing. The Hi-C protocol was improved to "*in situ* Hi-C", where proximity ligation is performed in intact nuclei to increase the efficiency of capturing chromatin interactions [24].

ChIA-PET experiments are performed as follows [28,20]. First, cells are subjected to dual cross-linking (to stabilize protein-DNA and protein-protein interactions), cells are lysed, and chromatin is fragmented by sonication. Immunoprecipitation enriches for chromatin complexes involving a specific protein, and DNA end repair and A-tailing are performed. Proximity ligation with a linker sequence captures chromatin interactions. The junctions of chromatin interactions are extracted by reverse cross-linking, tagmentation, immobilization on streptavidin beads, and PCR amplification, and then the sample goes on to sequencing. This description is of the second version of the ChIA-PET protocol, while the original ChIA-PET protocol used restriction digestion instead of sonication and a different read length (2×20 bp instead of 2×150 bp) [12].

Following the improvement in *in situ* Hi-C [24], ChIA-PET has been further improved to perform proximity ligation *in situ* in intact nuclei. The *in situ* ChIA-PET protocol now has a greatly increased efficiency for capturing intra-chromosomal chromatin interactions, with the tradeoff of a weaker binding enrichment of the protein. Similar protocols have been developed in other groups, notably HiChIP [23] and PLAC-seq [11].

There is now widespread interest in comparing *in situ* ChIA-PET and *in situ* Hi-C. We review methods for performing comparative analyses and visualizations. We note how to download public *in situ* ChIA-PET and *in situ* Hi-C data from web portals, and describe a collection of open-source computational tools to enable start-to-end data processing, analysis, and visualization. The Supplemental Material has example commands for installing and using these computational tools.

1.1. Downloading in situ ChIA-PET and in situ Hi-C data

Two major repositories for downloading *in situ* ChIA-PET and *in situ* Hi-C data are the ENCODE Portal [7,4] and the 4DN Portal [5] (see Data Availability). Both portals have intuitive user interfaces for browsing or searching for data by species, cell type, or experimental method. Within each dataset's webpage, there is an "Audits" section to indicate quality-

* Corresponding author at: The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.
E-mail address: yijun.ruan@jax.org (Y. Ruan).

https://doi.org/10.1016/j.ymeth.2019.09.019
Received 19 March 2019; Received in revised form 27 September 2019; Accepted 28 September 2019
Available online 16 October 2019

control flags.

Once a data set is identified that matches search criteria and has satisfactory quality control, the FASTQ files of raw sequencing reads can be downloaded. For *in situ* ChIA-PET and *in situ* Hi-C, these are paired-end FASTQ files (one file of R1 reads and one file of R2 reads). Once downloaded, the FASTQ files can be processed using open-source and fully automated pipelines (described in detail in later sections). Some datasets in the portals also have processed results files, such as the 2D contact matrix file. If using the results files directly, it is important to note the reference genome assembly to ensure consistency across the project.

To demonstrate results in the figures, we use data from the HFFc6 cell type downloaded from the 4DN portal (see Data Availability). Currently in the 4DN portal, the HFFc6 cell type and the H1 cell type have replicated *in situ* Hi-C datasets, CTCF *in situ* ChIA-PET datasets, and RNAPII *in situ* ChIA-PET datasets. These cell types are ideal starting points for comparative analyses.

1.2. Processing in situ ChIA-PET data using ChIA-PIPE

In situ ChIA-PET data can be processed using ChIA-PIPE [3], a fully automated pipeline for ChIA-PET data processing, quality assessment, analysis, and visualization. ChIA-PIPE runs from a single launch command on Linux high-performance computing clusters and is available for download (see Software Availability). ChIA-PIPE takes as input the names and directories of FASTQ files, and a configuration file with parameter settings and metadata.

Briefly, read pairs are parsed to identify the proximity-ligation linker sequence, and the genomic sequence "tags" are extracted from the reads. Read pairs that contained a linker sequence and two genomic sequence tags are referred to as paired-end tags (PETs). PETs are mapped to the reference genome, and uniquely mapped, non-redundant PETs are retained.

After the BAM file of final PETs is generated, multiple workflows of data processing proceed. First, input files are generated for viewing 2D chromatin contact maps. Juicer is used [8] to generate an input file for Juicebox [26], and cooler is used [2] to generate an input file for Higlass [15] (these tools are described in detail in a later section). Second, binding peaks are called using SPP [16] or MACS2 [30]. Third, loops are called and annotated with their binding-peak overlap, and these high-confidence loops are then used for calling chromatin contact domains. Fourth (optionally), allele-specific binding peaks and loops are resolved if phased SNP information is available for the relevant cell type (indicated via the configuration file).

As ChIA-PIPE runs, several key output files are automatically generated to facilitate interpretation. A summary statistics table is generated for quality assessment, and files of loops, peaks, and domains are generated for visualization in the WashU Epigenome Browser [31] and BASIC Browser [3] (this visualization tool is described in detail in a later section). Notably, ChIA-PIPE can also be readily applied for processing HiChIP or PLAC-seq data by modifying two parameters in the configuration file to indicate the experiment type and restriction-enzyme site.

1.3. Processing in situ Hi-C data using Juicer

In situ Hi-C data can be processed using Juicer [8], a fully automated pipeline for going from raw FASTQ files to final 2D chromatin contact maps. Juicer runs from a single launch command on Linux high-performance computing clusters or in the cloud with Amazon Web Services, and is available for download (see Software Availability). Briefly, read pairs are mapped to the reference genome, alignments are deduplicated, and read pairs with three or more alignments are filtered out [8]. The final BAM file of alignments is used to generate a 2D contact matrix (.hic file format) for heat map visualization in Juicebox [9,26]. Juicer enables generation of the 2D contact matrix with or without normalization by matrix balancing [17,14,8]. Juicer also automatically calls loops and domains from the 2D contact map (described in detail in a later section). Juicer reports extensive quality-assessment statistics, and generates a quality-assessment plot named the Aggregate Peak Analysis plot, [24,8], which depicts the average signal strength of aggregated loops compared to the background. Finally, the Juicer "dump" tool allows data to be extracted from the binary 2D contact matrix file and output to text files for analysis [8].

1.4. Processing in situ Hi-C data using distiller and cooler

In situ Hi-C data can also be processed using distiller [1], a fully automated pipeline that emphasizes reproducibility using the Nextflow workflow manager [6] with Docker containers [22]. Distiller maps read pairs to the reference genome, deduplicates alignments, and calls cooler [2] to generate a 2D contact matrix (.cool format) for visualization in HiGlass [15]. The .cool format is an efficient sparse matrix representation that allows random data access. Cooler provides command-line utilities and a Python library to create and access .cool files, and to normalize by matrix balancing [14]. Distiller and cooler are available for download (see Software Availability).

1.5. Viewing 2D contact maps using Juicebox and HiGlass

The 2D contact matrices generated for *in situ* ChIA-PET or *in situ* Hi-C can be viewed as interactive heat maps using Juicebox [9,26] and HiGlass [15]. The original Juicebox is a Desktop application [9], while the newer version, Juicebox.js, is a web browser [26]. HiGlass is a web browser [15]. The tools share many features: unlimited zoom of chromatin contact maps (full genome to a chromosome to a chromosomal region); customizable colors; ability to annotate the 2D contact map with loops and domains; and ability to display gene annotations, ChIP-seq tracks, and RNA-seq tracks above the 2D contact map. Juicebox.js and HiGlass allow customized visualization results to be shared as URLs.

For example, we used Juicebox.js to compare 2D contact maps of *in situ* ChIA-PET and *in situ* Hi-C data from the HFFc6 cell type (Fig. 1) (see Data Availability). 2D contact maps are shown for *in situ* Hi-C (in red), CTCF *in situ* ChIA-PET (in green), and RNAPII *in situ* ChIA-PET (in blue). 2D contact maps at multiple resolutions show that the *in situ* ChIA-PET and *in situ* Hi-C HFFc6 data are similar in terms of quality and broad topological features.

1.6. Calling loops and domains from in situ ChIA-PET and in situ Hi-C data

For CTCF *in situ* ChIA-PET data, loops and domains are automatically called during ChIA-PIPE data processing [3]. ChIA-PIPE calls loops using a modified version of the algorithm from its predecessor pipeline [18]. First, "self-ligation" PETs within close distance along the linear chromosome are filtered out, and the remaining "inter-ligation" intra-chromosomal PETs are used for loop calling. Second, each tag is extended in its 5' direction, and PETs that have both ends overlapping are merged into loops, where: "interaction frequency" is the number of PETs contributing to the loop; "anchors" are the contact points of the loop; and "span" is the linear chromosomal distance between the two anchors. The reason that tags are extended in the 5' direction is based on the nature of the ChIA-PET protocol: immunoprecipitation is performed at the protein-binding site of each fragmented chromatin complex, and then the loose DNA ends are ligated and captured for sequencing. Extension in the 5' direction of each tag moves toward the direction of the actual protein binding event. Third, loops are annotated with their binding-peak overlap (0, 1, or 2 anchors peak-supported). The final loops for analysis and visualization are those with interaction frequency of at least three and two anchors peak-supported.

Based on the loops, ChIA-PIPE calls domains using a modified version of the algorithm from [28]. First, loops are further refined by

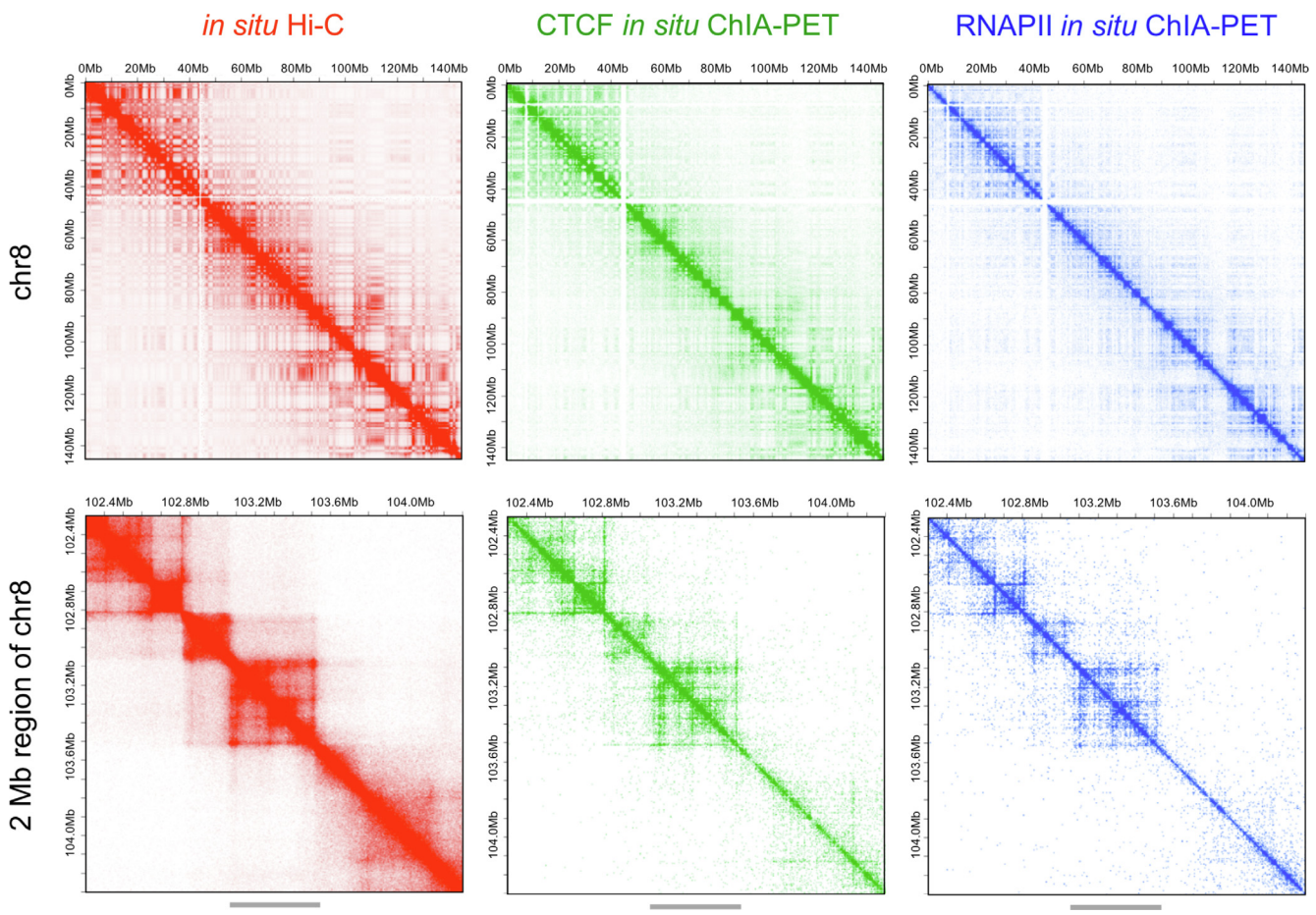

Fig. 1. 2D contact maps of *in situ* ChIA-PET and *in situ* Hi-C data and can be visualized using Juicebox.js. Chromatin interactions from the HFFc6 cell type are visualized using Juicebox.js [26]. The columns are: *in situ* Hi-C (red), CTCF *in situ* ChIA-PET (green), and RNAPII *in situ* ChIA-PET (blue) (see Data Availability). The first row displays a full chromosome (chr8 versus chr8). Topological domains can be seen along the diagonal. The second row displays a 2 Megabase region of chr8 (positions 102,303,987 to 104,303,987). Several topological domains can be seen. Within this region, *in situ* Hi-C and *in situ* ChIA-PET captured similar topological features. Heat maps have 250 kilobase resolution in the first row and 5 kilobase resolution in the second row. Heat maps are displayed with and without matrix balancing for *in situ* Hi-C and *in situ* ChIA-PET, respectively. The Juicebox.js saturation thresholds are: *in situ* Hi-C upper map: 738; *in situ* Hi-C lower map: 29; CTCF *in situ* ChIA-PET upper map: 27; CTCF *in situ* ChIA-PET lower map: 3; RNAPII *in situ* ChIA-PET upper map: 13; RNAPII *in situ* ChIA-PET lower map: 2. The gray bar beneath the second row denotes a topological domain that is magnified in Fig. 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

percentile filtering of interaction frequencies. Using the refined loops, domains are then defined as any genomic region that has continuous coverage by loops with no gaps for a minimum user-defined length (e.g., 10 or 25 kilobases).

For *in situ* Hi-C data sets, loops and domains are automatically called during Juicer data processing [8]. The HiCCUPS module in Juicer calls loops from the 2D contact matrix by examining each pixel and testing whether the number of contacts is significantly enriched compared to the contacts in neighboring regions [24,8]. To ensure specific loops, HiCCUPS defines the neighboring regions in multiple ways and then requires enrichment compared to all neighborhoods [24].

The Arrowhead module in Juicer calls domains from the 2D contact matrix [24,8]. The 2D contact matrix is converted to an “Arrowhead matrix”, where there is an arrow-shaped feature pointing to the left boundary of each domain from the 2D contact matrix. A “corner-score matrix” is then computed, which indicates each pixel’s likelihood of being located at an arrowhead [24]. Thus, the boundaries of domains can be inferred. Once loops and domains have been called from *in situ* ChIA-PET and *in situ* Hi-C, they can proceed to data visualization.

1.7. Viewing loops and domains using BASIC Browser

BASIC Browser (Browser for Applications in Sequencing and Integrated Comparisons) is a web browser for interactive, high-resolution visualization of *in situ* ChIA-PET loops, binding coverage, and domains (Fig. 2) [3]. BASIC Browser is available for download as a pre-compiled Docker image [22], which runs easily on a laptop or Linux server (see Software Availability). Once a dataset is uploaded to BASIC Browser, the data track remains stored in a local database. Thus, many tracks can be selected and viewed in BASIC Browser simultaneously.

BASIC Browser has many valuable features. First, there is an intuitive display of loops with the y-axis height indicating interaction frequency. Second, different data types can be uploaded and displayed to facilitate biological interpretation: gene annotations, *in situ* Hi-C loops, ChIP-seq, RNA-seq, ATAC-seq, chromHMM chromatin state annotations [10], and directional CTCF motifs. Third, BASIC Browser is highly customizable for colors, track heights, axis limits, and display features.

For example, we used BASIC Browser to compare loop and domain calls from *in situ* ChIA-PET and *in situ* Hi-C data in the HFFc6 cell type (Fig. 3) (see Data Availability). Results are shown in BASIC Browser for *in situ* Hi-C data (in red), CTCF *in situ* ChIA-PET data (in green), and RNAPII *in situ* ChIA-PET data (in blue). In the example genomic region, a similar domain is identified with *in situ* Hi-C and *in situ* ChIA-PET; the latter provides more fine-scale information on within-domain interactions.

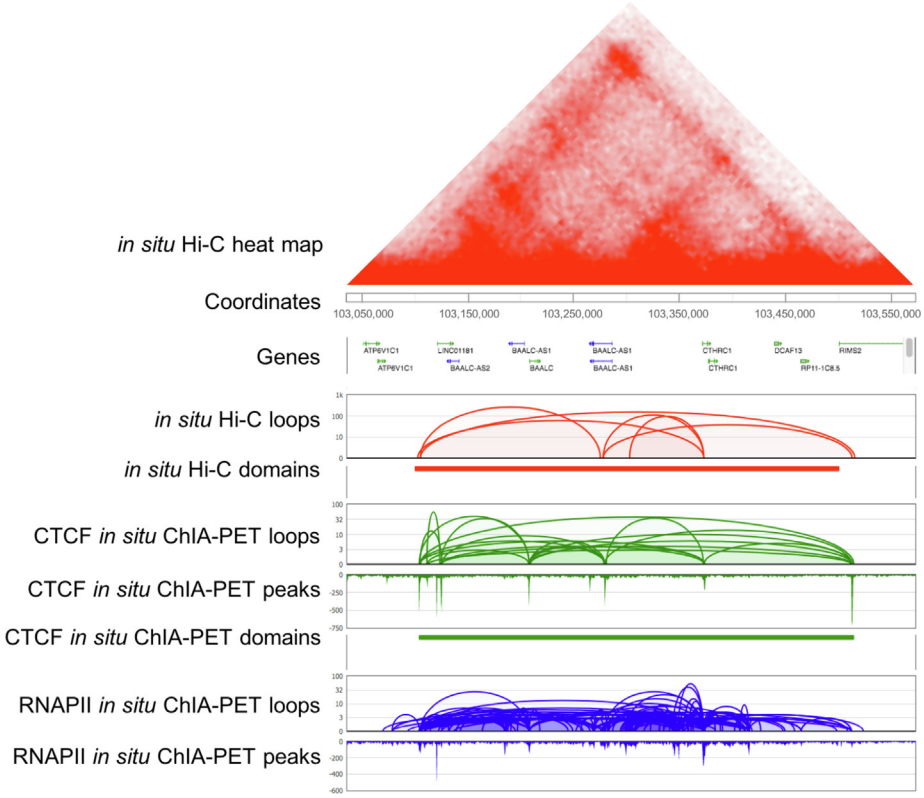

Fig. 2. Loops and domains from *in situ* ChIA-PET and *in situ* Hi-C data and can be visualized using BASIC Browser. Chromatin interactions from the HFFc6 cell type are visualized for a 500 kilobase region of chromosome 8 (positions 103,035,829 to 103,572,145), as denoted in Fig. 1 with a gray bar. The upper triangle is a Juicebox.js 2D contact map [26] of *in situ* Hi-C data, with visualization settings: 5 kilobase resolution; matrix balancing; saturation threshold of 29. A BASIC Browser visualization of the region is displayed next. First, there is a gene annotation track. Then, loops and domains are shown for *in situ* Hi-C (red), and loops, peaks, and domains are shown for CTCF *in situ* ChIA-PET (green) and RNAPII *in situ* ChIA-PET (blue). For *in situ* Hi-C, loops were called using HICCUPS in Juicer [8] and domains were called using Arrowhead in Juicer [8]. For *in situ* ChIA-PET, loops and domains were called using ChIA-PIPE [3]. In this region, a similar domain is identified from *in situ* Hi-C and CTCF *in situ* ChIA-PET. The CTCF *in situ* ChIA-PET data provides more fine-scale information on within-domain interactions, and also provides information on the CTCF binding profile. The RNAPII *in situ* ChIA-PET data provides information on the functional activity within the domain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1.8. Annotating tandem and convergent CTCF loops

After loops are called from *in situ* Hi-C or CTCF *in situ* ChIA-PET data, they can be annotated as tandem or convergent based on CTCF motif directionality in the two anchors. For CTCF *in situ* ChIA-PET data, the motif-directionality annotation is performed as follows. First, CTCF-motif coordinates and strands are extracted from Ensembl (version 95) using bio Mart [27]. Second, CTCF binding peaks are overlapped with CTCF motifs, and each peak is annotated based on motif direction(s). Peaks are retained for analysis that overlapped one CTCF motif, or that overlapped multiple CTCF motifs with the same direction. Third, loop anchors are overlapped with the refined peaks, and each anchor is annotated based on motif direction(s). Loops are retained for analysis that had at least one motif in each anchor and that had consistent motif directions within each anchor. Finally, the refined loops are annotated based on CTCF motif directionality: convergent, tandem (positive strand), tandem (negative strand), or divergent.

The analysis is performed similarly for *in situ* Hi-C data. However, as described above, it is beneficial to first use CTCF binding peaks for overlapping and retaining CTCF motifs. This is because the loop anchors may be broad, so overlapping them directly with CTCF motifs could result in many loops having ambiguous motif-directionality

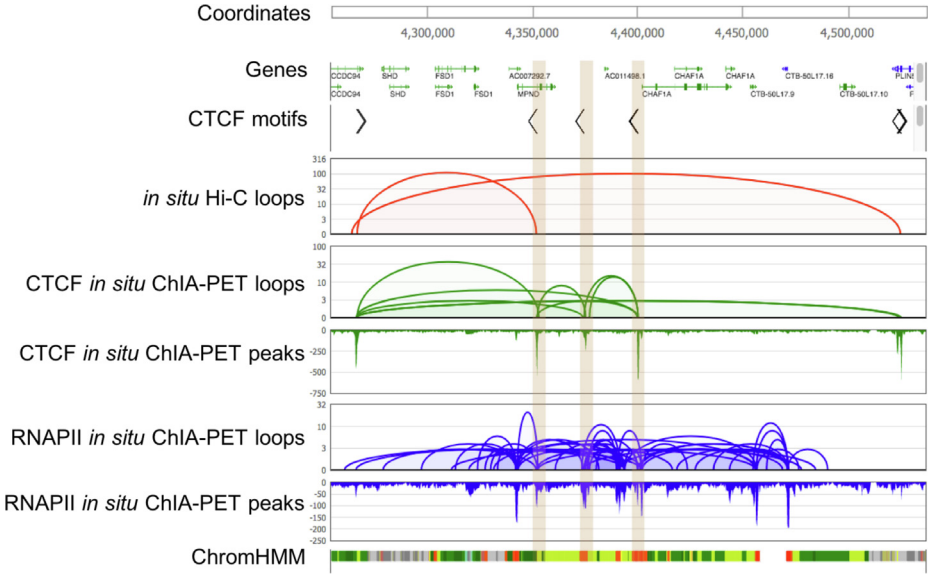

Fig. 3. The CTCF-motif directionality of loops from *in situ* ChIA-PET and *in situ* Hi-C data can be assessed and can be visualized using BASIC Browser. Chromatin interactions from the HFFc6 cell type are visualized using BASIC Browser for a 300 kilobase region of chromosome 19 (positions 4,255,132 to 4,536,942). First, there is a gene annotation track, and a track of CTCF motifs (black) and their directions (rightward arrows indicating positive-strand motifs; leftward arrows indicating negative-strand motifs). Next, loops are shown for *in situ* Hi-C (red), and loops and peaks are shown for CTCF *in situ* ChIA-PET (green) and RNAPII *in situ* ChIA-PET (blue). The *in situ* Hi-C loops were called using HICCUPS in Juicer [8], and the ChIA-PET loops were called using ChIA-PIPE [3]. The final track is chromHMM chromatin state annotations [10] (see Data Availability), where red indicate promoters. In this region, *in situ* Hi-C captured two convergent loops. *in situ* ChIA-PET captured the same convergent CTCF loops, but also captured tandem CTCF loops. The tandem CTCF loops in this region involve promoters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

annotations. Therefore, CTCF binding peaks can be downloaded for a ChIP-seq or *in situ* ChIA-PET data set from the same cell type and used when performing the motif-directionality annotation of *in situ* Hi-C loops.

For example, we performed the motif-directionality annotation of loops using CTCF *in situ* ChIA-PET and *in situ* Hi-C data from the HFFc6 cell type (Fig. 3) (see Data Availability). BASIC Browser to visualize the refined loops that passed filtering during the motif-directionality annotation, and to visualize the CTCF motifs and their directions (Fig. 3). In the example genomic region, *in situ* Hi-C identified two convergent CTCF loops. *in situ* ChIA-PET identified the same convergent CTCF loops, but also identified tandem CTCF loops. The tandem CTCF loops in this genomic region involved gene promoters (Fig. 3).

2. Conclusions

In situ ChIA-PET and *in situ* Hi-C are advanced high-throughput methods for capturing chromatin interactions. Now that high-quality replicated data sets are available of *in situ* Hi-C and *in situ* ChIA-PET (CTCF and RNAPII) in the same cell types, there is widespread interest in performing comparative analyses and visualizations. This review discussed a set of open-source tools for start-to-end analysis and visualization of *in situ* ChIA-PET and *in situ* Hi-C data – from downloading FASTQ files, to running processing pipelines, to visualizing the results, to performing downstream functional annotation. With this framework, many interesting questions could be addressed in the future by further comparative analyses. For example, a comprehensive categorization and functional analysis of the CTCF motif directionality of the loops would be of interest. Similarly, it would be of interest to determine if any loops were captured by *in situ* Hi-C but not by CTCF or RNAPII *in situ* ChIA-PET, which could suggest the involvement of other architectural proteins.

Software availability

This paper reviewed several tools for analyzing and visualizing *in situ* ChIA-PET and *in situ* Hi-C data. ChIA-PIPE [3] for processing *in situ* ChIA-PET data and calling loops and domains is available at https://github.com/TheJacksonLaboratory/chia_pipe. Juicer [8] for processing *in situ* Hi-C data and calling loops and domains is available at https://github.com/aidenlab/juicer/wiki. Juicebox.js [26] for visualizing 2D contact maps of *in situ* ChIA-PET and *in situ* Hi-C data is available at https://aidenlab.org/juicebox/. Distiller [1] and cooler [2] for processing *in situ* Hi-C data are available at https://github.com/mirnylab/distiller-nf and https://github.com/mirnylab/cooler. HiGlass [15] for visualizing 2D contact maps of *in situ* ChIA-PET and *in situ* Hi-C data is available at https://higlass.io/. BASIC Browser [3] for visualizing loops and domains from *in situ* ChIA-PET data and *in situ* Hi-C data is available at https://github.com/TheJacksonLaboratory/basic-browser.

Data availability

The ENCODE Data Portal is available at https://www.encodeproject.org/. The 4DN Data Portal is available at https://data.4dnucleome.org/. The HFFc6 *in situ* Hi-C data set used in this review is available for download from the 4DN Data Portal at https://data.4dnucleome.org/experiment-set-replicates/4DNES2R6PUEK/. The HFFc6 CTCF *in situ* ChIA-PET data set used in this review is available for download from the 4DN Data Portal at https://data.4dnucleome.org/experiment-set-replicates/4DNESCQ7ZD21/. The HFFc6 RNAPII *in situ* ChIA-PET data set used in this review is available for download from the 4DN Data Portal at https://data.4dnucleome.org/experiment-set-replicates/4DNESI1WZ5HT/. The HFF chromHMM annotation track used in this review is available for download from the ENCODE Data Portal at https://www.encodeproject.org/annotations/ENCSR666ARC/. The chromHMM track was lifted over [13] to the hg38 genome assembly for the example shown here.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgements

Y.R. was supported by ENCODE grant UM1 HG009409, 4DN grant (U54 DK107967), and JAX Director's Innovation Award 19000-18-02. We thank Yuliang Feng and Ping Wang for previously generating the HFFc6 *in situ* ChIA-PET data used as demo data in this review; thank Byoungkoo Lee for previously coordinating installation of a local Juicebox.js instance; and thank members of the Ruan Lab for helpful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymeth.2019.09.019.

References

[1] Abdennur, N., Schwarzer, W., Pekowska, A., Shaltiel, I.A., Huber, W., Haering, C.H., Mirny, L., Spitz, F. Condensin II inactivation in interphase does not affect chromatin folding or gene expression. bioRxiv.
[2] N. Abdennur, L. Mirny, Cooler: scalable storage for Hi-C data and other genomically-labeled arrays, Bioinformatics (2019) btz540.
[3] Capurso, D., Wang, J., Tian, S.Z., Cai, L., Namburi, S., Lee, B., Tjong, H., Tang, Z., Wang, P., Wei, C.L., Ruan, Y., Li, S. ChIA-PIPE: A fully automated pipeline for ChIA-PET data analysis and visualization. bioRxiv, doi: 10.1101/506683.
[4] C.A. Davis, B.C. Hitz, C.A. Sloan, E.T. Chan, J.M. Davidson, I. Gabdank, J.A. Hilton, K. Jain, U.K. Baymuradov, A.K. Narayanan, et al., The Encyclopedia of DNA elements (ENCODE): data portal update, Nucleic Acids Res. 46 (D1) (2018) D794–D801.
[5] J. Dekker, A.S. Belmont, M. Guttman, V.O. Leshyk, J.T. Lis, S. Lomvardas, L.A. Mirny, C.C. O'Shea, P.J. Park, B. Ren, et al., The 4D nucleome project, Nature 549 (7671) (2017) 219–226.
[6] P. Di Tommaso, M. Chatzou, E.W. Floden, P.P. Barja, E. Palumbo, C. Notredame, Nextflow enables reproducible computational workflows, Nat. Biotechnol. 35 (4) (2017) 316–319.
[7] I. Dunham, A. Kundaje, S.F. Aldred, P.J. Collins, C.A. Davis, F. Doyle, C.B. Epstein, S. Frietze, J. Harrow, R. Kaul, et al., An integrated encyclopedia of DNA elements in the human genome, Nature 489 (7414) (2012) 57–74.
[8] N.C. Durand, M.S. Shamim, I. Machol, S.S. Rao, M.H. Huntley, E.S. Lander, E.L. Aiden, Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments, Cell Syst. 3 (2016) 95–98.
[9] N.C. Durand, J.T. Robinson, M.S. Shamim, I. Machol, J.P. Mesirov, E.S. Lander, E.L. Aiden, Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom, Cell Syst. 3 (2016) 99–101.
[10] J. Ernst, M. Kellis, ChromHMM: automating chromatin-state discovery and characterization, Nat. Methods 9 (3) (2012) 215–216.
[11] R. Fang, M. Yu, G. Li, S. Chee, T. Liu, A.D. Schmitt, B. Ren, Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq, Cell Res. 26 (2016) 1345–1348.
[12] M.J. Fullwood, M.H. Liu, Y.F. Pan, J. Liu, H. Xu, Y.B. Mohamed, Y.L. Orlov, S. Velkov, A. Ho, P.H. Mei, et al., An oestrogen-receptor-alpha-bound human chromatin interactome, Nature 462 (2009) 58–64.
[13] Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res., 1;34(Database issue):D590-8.
[14] M. Imakaev, G. Fudenberg, R.P. McCord, N. Naumova, A. Goloborodko, B.R. Lajoie, J. Dekker, L.A. Mirny, Iterative correction of Hi-C data reveals hallmarks of chromosome organization, Nat. Methods 9 (10) (2012) 999–1003.
[15] P. Kerpedjiev, N. Abdennur, F. Lekschas, C. McCallum, K. Dinkla, H. Strobelt, J.M. Luber, S. Ouellette, A. Azhir, N. Kumar, et al., HiGlass: web-based visual exploration and analysis of genome interaction maps, Genome Biol. 19 (1) (2018) 125.
[16] P.V. Kharchenko, M.Y. Tolstorukov, P.J. Park, Design and analysis of ChIP-seq experiments for DNA-binding proteins, Nat. Biotechnol. 26 (2008) 1351–1359.
[17] P.A. Knight, D. Ruiz, A fast algorithm for matrix balancing, IMA J. Numer. Anal. 33 (2012) 1029–1047.
[18] G. Li, M.J. Fullwood, H. Xu, F.H. Mulawadi, S. Velkov, V. Vega, P.N. Ariyaratne, Y.B. Mohamed, H.S. Ooi, C. Tennakoon, et al., ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing, Genome Biol. 11 (2) (2010) R22.

[19] G. Li, X. Ruan, R.K. Auerbach, K.S. Sandhu, M. Zheng, P. Wang, H.M. Poh, Y. Goh, J. Lim, J. Zhang, et al., Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation, Cell 148 (1–2) (2012) 84–98.
[20] X. Li, O.J. Luo, P. Wang, M. Zheng, D. Wang, E. Piecuch, J.J. Zhu, S.Z. Tian, Z. Tang, G. Li, Y. Ruan, Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions, Nat. Protoc. 12 (2017) 899–915.
[21] E. Lieberman-Aiden, N.L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome, Science 326 (2009) 289–293.
[22] D. Merkel, Docker: lightweight Linux containers for consistent development and deployment, Linux J. 239 (2014) 2.
[23] M.R. Mumbach, A.J. Rubin, R.A. Flynn, C. Dai, P.A. Khavari, W.J. Greenleaf, H.Y. Chang, HiChIP: efficient and sensitive analysis of protein-directed genome architecture, Nat. Methods 13 (2016) 919–922.
[24] S.S. Rao, M.H. Huntley, N.C. Durand, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, A.L. Sanborn, I. Machol, A.D. Omer, E.S. Lander, E.L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping, Cell 159 (7) (2014) 1665–1680.
[25] S.S.P. Rao, S.C. Huang, B. Glenn St Hilaire, J.M. Engreitz, E.M. Perez, K.R. Kieffer-Kwon, A.L. Sanborn, S.E. Johnstone, G.D. Bascom, I.D. Bochkov, et al., Cohesin Loss Eliminates All Loop Domains, Cell 171 (2) (2017) 305-320.e24.
[26] J.T. Robinson, D. Turner, N.C. Durand, H. Thorvaldsdóttir, J.P. Mesirov, E.L. Aiden, Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data, Cell Syst. 6 (2018) 1–3.
[27] D. Smedley, S. Haider, S. Durinck, L. Pandini, P. Provero, J. Allen, O. Arnaiz, M.H. Awedh, R. Baldock, G. Barbiera, et al., The BioMart community portal: an innovative alternative to large, centralized data repositories, Nucleic Acids Res. 43 (W1) (2015) W589–W598.
[28] Z. Tang, O.J. Luo, X. Li, M. Zheng, J.J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, J. Wlodarczyk, B. Ruszczycki, et al., CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription, Cell 163 (2015) 1611–1627.
[29] A.S. Weintraub, C.H. Li, A.V. Zamudio, A.A. Sigova, N.M. Hannett, D.S. Day, B.J. Abraham, M.A. Cohen, B. Nabet, D.L. Buckley, Y.E. Guo, D. Hnisz, R. Jaenisch, J.E. Bradner, N.S. Gray, R.A. Young, YY1 Is a Structural Regulator of Enhancer-Promoter Loops, Cell. 171 (7) (2017) 1573–1588.e28.
[30] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, B.E. Bernstein, C. Nusbaum, R.M. Myers, M. Brown, W. Li, et al., Model-based analysis of ChIP-Seq (MACS), Genome Biol. 9 (2008) R137.
[31] X. Zhou, R.F. Lowdon, D. Li, H.A. Lawson, P.A. Madden, J.F. Costello, T. Wang, Exploring long-range genome interactions using the WashU Epigenome Browser, Nat. Methods 10 (2013) 375–376.