Published in final edited form as:

Methods Mol Biol. 2020 ; 2117: 65–78. doi:10.1007/978-1-0716-0301-7_3.

Evaluation of 3D Chromatin Interactions Using Hi-C

Gangqing Hu¹

¹Department of Microbiology, Immunology, and Cell Biology, West Virginia University School of Medicine, Morgantown, USA

Abstract

The invention of Hi-C has greatly facilitated 3D genome research through an unbiased probing of 3D chromatin interactions. It produces enormous amount of sequencing data that capture multiscale chromatin conformation structures. In the last decade, numerous computational methods have been developed to analyze Hi-C data and predict A/B compartments, topologically associating domains (TADs), and significant chromatin contacts. This chapter introduced the iHiC package that provides several utilities to facilitate Hi-C data analysis with public software and demonstrated its application to a Hi-C dataset generated for mouse embryonic stem (ES) cells.

Keywords

Hi-C; Bioinformatics; A/B compartments; TAD; Chromatin contacts

1 Introduction

DNA sequence within a nucleus is packed into a hierarchy of chromatin structures with histone proteins [1]. Chromatin conformation can be captured through chromosome conformation capture technique based on proximity ligation, such as Hi-C and its derivatives [2, 3], and ligation-free methods such as GAM [4], SPRITE [5], and Trac-looping [6]. Hi-C captures genome-wide chromatin interactions through ligation of chromatin DNA sequences located in space-proximity and cut by restriction enzymes, and applies high-throughput sequencing to measure interaction frequencies [2]. A typical Hi-C experiment would yield hundreds of millions of or even billions of pair-end tags (PETs) [7, 8], which impose computational challenges for data analysis.

Analysis of Hi-C contact matrix, which records the number of PETs linking two genomic regions, uncovers a hierarchy structure of chromatin organization. At a range up to millions of base pairs (bps), the chromatin is segregated into alternating A and B compartments, corresponding to transcription active and transcription repressive regions, respectively [2]. One compartment region may consist of one or more topologically associating domains (TADs); chromatin regions from one TAD preferentially interact with regions from the same TAD [9-11]. Within a TAD, interacting chromatin region may form structural loops anchored by convergent CTCT motifs [8] or regulatory loops for promoter enhancer interaction [6]. The structural loops could be readily identified by in situ Hi-C. The

prediction of regulatory loops requires a much higher sequencing depth but could be addressed by alternative methods that target chromatin regions with specific features [6].

Bioinformatic tools developed for Hi-C data analysis include (but not limited to) data preprocessing such as PET alignment and data normalization, identification of A/B compartments, TADs, and significant contacts, and data visualization [12, 13]. Multiple or even dozens of tools have been developed for each of the topics [14, 15]. This chapter will introduce the iHiC package that provides a handful of utilities to facilitate Hi-C data analysis with public software, including CscoreTool for compartments [16], TopDom for TADs [17], Fit-Hi-C for significant contacts [18], and WashU Epigenome Browser for data visualization [19]. It includes a step-by-step guidance on the application of the package to a Hi-C dataset generated for mouse ES cells E14 [20].

2 Materials

2.1 Public Hi-C Data

In-situ Hi-C sequencing data for mouse ES cell line E14 [20] is downloaded from SRA under sample accession number SRX1338543 (<https://www.ncbi.nlm.nih.gov/sra?term=SRX1338543>).

2.2 Public Software

Sratoolkit: <https://github.com/ncbi/sra-tools>.

Bowtie2 [21]: <https://github.com/BenLangmead/bowtie2>.

Samtools [22]: <https://github.com/samtools/samtools>.

Bedtools [23]: <https://github.com/arq5x/bedtools2>.

CscoreTool [16]: <https://github.com/scoutzxb/CscoreTool>.

TopDom [17]: <http://zhoulab.usc.edu/TopDom/> (see Note 1).

Fit-Hi-C [18]: <https://github.com/ay-lab/fithic>.

HTSlib for “bgzip” and “tabix”: <http://www.htslib.org/download>.

R: <https://www.r-project.org/>.

2.3 Online Web Servers

WashU Epigenome Browser: <https://epigenomegateway.wustl.edu>.

UCSC Genome Browser: <https://genome.ucsc.edu/>.

Cyverse: <https://www.cyverse.org>.

¹In case that the website is not available, <https://github.com/HenrikBengtsson/TopDom> maintains a copy of the original TopDom R script (“TopDom_v0.0.2.R”) and its documentation.

2.4 The iHiC Package

The iHiC package includes a set of utilities (mostly written in C++)

to facilitate Hi-C data analysis: it calculates statistics of PETs alignments, generates input files for several downstream software, and postprocesses their outputs for data visualization.

To install the package:

1. “`git clone https://github.com/hugangqing/iHiC.git`”: Download the package from github.
2. “`sh iHiC_install`”: compile the C++ codes with g++. Executive files will be saved to a subfolder named “bin”.
3. Add the “bin” folder to system PATH or copy the executive files to “/usr/bin/” or “/usr/local/bin”.

The package includes the following utilities:

- a. iHiC_BEDPE_Statistics: Calculate statistics of PET alignments from a simplified version of BEDPE file (*see* Note 2).
- b. iHiC_PETOrientations_Stat: Calculate % of PETs in different combinations of strand orientations as a function of genomic distance.
- c. iHiC_ContactFreq_Distance: Calculate contact probability as a function genomic distance.
- d. iHiC_BEDPE2III: Generate an intrachromosomal interaction index (III) file that records the number of PETs for interacting genomic bins from the same chromosomes.
- e. iHiC_III2WashU: Convert an III file into a “longrange” format (<http://wiki.wubrowse.org/Long-range>) accepted by the WashU browser.
- f. iHiC_BEDPE2HiCSummary: Convert a BEDPE file into a “HiCSummary” format (<http://homer.ucsd.edu/homer/interactions/HiCtagDirectory.html>) accepted by CscoreTool.
- g. iHiC_Cscore_Adj: Flip the sign of “C-score” chromosome-by-chromosome to ensure genomic regions with positive scores corresponding to A compartment status.
- h. iHiC_III2MTX4TopDom: Convert an III file into interaction matrices accepted by TopDom.
- i. iHiC_TADBoundary_TopDom: Extract TAD boundaries from TopDom predictions.

²The BEDPE format is described in <https://bedtools.readthedocs.io/en/latest/content/general-usage.html>. The simplified version keeps columns corresponding to “chrom1,” “start1,” “end1,” “chrom2,” “start2,” “end2,” “score,” “strand1,” and “strand 2” to save space storage. The simplified version is called as BEDPE throughout the chapter unless stated otherwise.

- j. iHiC_III2FitHiCInputs: Convert an III file into interaction matrices accepted by Fit-Hi-C.
- k. iHiC_FitHiC2WashU: Convert the significant contacts pre-dicted by Fit-Hi-C into a “longrange” format accepted by the WashU Epigenome Browser.
- l. TopDom_wrapper.R: an Rscript wrapper to run TopDom.

3 Methods

3.1 Sequencing Data Extraction

The “fastq-dump” utility from the sratoolkit is a convenient way to locate and download sequencing data from NCBI SRA simply based on accession numbers; the SRA explorer (<https://ewels.github.io/sra-explorer/>) is also an useful tool to explore data in SRA. The accession numbers for the SRA sample SRX1338543 are SRR2658622, SRR2658623, SRR2658624, SRR2658625, and SRR2658626.

1. “*mkdir your_folder*”: Create a folder for the Hi-C analysis demonstrated in this chapter. All commands listed in the chapter are executed in this folder under Linux/Unix environment (ubuntu LTS) unless stated otherwise.
2. Download sratoolkit binaries from <https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/> and copy them to /user/local/ bin/ or /user/bin/.
3. “*fastq-dump --split-files SRR2658622*”: Extract fastq files from SRR2658622 and repeat for other accession numbers (see Note 3).

3.2 Sequence Alignment

1. Download Bowtie index from <http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>.
2. “*bowtie2 -p 24 --local -x path-to-bowtie-index/index-prefix -l SRR2658622_1.fastq -2 SRR2658622_2.fastq | samtools view -b>SRR2658622.bam*”: Execute PETs alignment by “bowtie2” and save the alignment results in bam file (see Note 4).
3. “*samtools view -bf 1 -F 12 SRR2658622.bam | bamToBed -i stdin -bedpe | cut -f 1-6,8,9,10>SRR2658622.bedpe*”: Extract PETs where both ends map from the BAM file and convert the output to a simplified BEDPE format (see Note 2).
4. Repeat **steps 2 and 3** for all other accession numbers.

³The option “--split-files” denotes pair-end sequencing data; “fastq-dump” will generate two fastq files, one ending with “_1.fastq” for read 1 and the other ending with “_2.fastq” for read 2.

⁴The Hi-C PETs are generally chimeric with one end originated from one genomic region and the other end from another region, which may be separated by tens of thousands of base-pairs (bps) or even from a different chromosome. The sequence aligner Bowtie2 is flexible to capture PETs for both situations. The output SAM format demands huge space for storage. The “samtools view -b” command converts the SAM format into a binary BAM format. The “--local” option allows for an automatic trimming at read ends to maximize alignment score for “Bowtie2”; it helps to rescue PETs that are mixed with adapter sequences. The vertical bar (|) is a pipe symbol that redirects the output of one program (here “bowtie2”) to the input of another (here “samtools”) without creating any temporary file. The “>” symbol saves the output of “cut” to a local file.

5. “*cat SRX265862*.bedpe > SRX1338543.bedpe*”: Concatenate the BEDPE file for each accession to generate a pooled BEDPE file.

3.3 Quality Inspection

1. “*iHiC_BEDPE_Statistics SRX1338543.bedpe 10*”, where 10 is the minimal mapping quality (MAPQ) to extract uniquely mapped PETs: Calculate statistics for mapped PETs of a Hi-C library and present the output in a hierarchic order (Fig. 1a) (*see Note 5*).
2. “*iHiC_PETOrientations_Stat SRX1338543.bedpe 10*”, where 10 is the minimal MAPQ for uniquely mapped PETs: Calculate the percentages of PETs in different combinations of strand orientations and the standard deviation of the percentages as a function of genomic distance (Fig. 1b, c) (*see Note 6*).
3. “*iHiC_ContactFreq_Distance SRX1338543.bedpe 10 > SRX1338543_contactprob.txt*”, where 10 is the minimal MAPQ for uniquely mapped PETs: Calculate contact probability as a function of genomic distance interval (Fig. 1d) (*see Note 7*).

3.4 Visualization of Interaction Matrix

Tools and packages available to visualize interaction matrix of Hi-C data include the WashU Epigenome Browser [19], The 3D Genome Browser [24], Juicebox [25], HiCPlotter [26], and HiGlass [27]. The chapter uses the WashU Epigenome Browser as an example:

1. “*iHiC_BEDPE2III SRX1338543.bedpe 20000 20000 2000000 10 SRX1338543.iii*”, where the second, third, fourth, and fifth parameters stand for bin size, minimal distance, maximal distance, and minimal MAPQ for uniquely mapped PETs, respectively: Calculate numbers of PETs for genomic bin-pairs from the same chromosome and generate an intrachromosomal interaction index (III) file (*see Note 8*).
2. “*iHiC_III2WashU SRX1338543.iii 20000 SRX1338543.long-range*”, where 20,000 is bin size: Convert the III file into a “longrange” format, which should be

⁵Often, only PETs uniquely mapped are included for Hi-C data analysis. MAPQ measures the uniqueness of the mapping: a higher value indicates more unique, and a score higher than 10 is frequently used to exclude PETs mapped to multiple positions. PETs mapping to the same positions are likely caused by PCR amplification and therefore are redundant. The redundancy may depend on both library complexity and sequencing depth and therefore a high redundancy does not necessarily indicate a library failure, especially in situations where cell number is limited. For this reason, I recommended to base the quality inspection on the uniquely mapped PETs after removing redundancy. The “*Cis/Trans* ratio” in Fig. 1a is defined by the number of long-range *cis* PETs (20K bps) divided by the numbers of interchromosomal PETs and long-range *cis* PETs. According to the 4D Nucleome Consortium (4DN), a *Cis/Trans* ratio higher than 40% is required [13]. The percentage of “long-range *cis* reads” in Fig. 1a is defined by the number of long-range *cis* PETs over the number of uniquely mapped and nonredundant PETs. A number of 40% or higher indicates a good library.

⁶A PET originated from the same DNA fragment should have one end map to the “+” strand and the other end to the “-” strand. For SRX1338543, the PETs less than 1000 bps are dominated by the “+-” combination, interpreted as a result of self-ligations (Fig. 1b). The standard deviation of the percentages are smaller than 0.05 after 1K bps, and <0.002 after 20K bps (Fig. 1c); according to 4DN, the convergence is “good” if <0.05 and “very good” if <0.002.

⁷The curve follows a power-law relationship in the form of $P(s) \sim s^\lambda$ for genomic interval corresponding to TAD (10K–300K), where s and λ refer to distance and slope respectively. The slope λ is around -0.76 for mitotic samples [30]; it is around -0.77 for SRX1338543 (Fig. 1d).

⁸The intrachromosomal interaction index file (*.iii) is in the following format: chromosome, genomic position of bin1, genomic position of bin2, number of PETs.

sorted, compressed, and indexed (next three steps) before supply to the WashU Epigenome Browser.

3. `"sort -k1,1 -k2,2n SRX1338543.longrange > SRX1338543.longrange.sort"`: Sort the "longrange" file.
4. `"bgzip -f SRX1338543.longrange.sort"`: Compress the sorted file.
5. `"tabix -p bed -f SRX1338543.longrange.sort.gz"`: Index the compressed file. The output filename is ending with ".tbi".
6. Follow the instruction described in http://wiki.wubrowse.org/Track_hosting. to upload the compressed file and its index to Cyverse (<https://www.cyverse.org/>), and obtain an URL to be supplied to the WashU Browser: briefly, (a) register an account with Cyverse and login; (b) Choose "MY SERVICES" under the user's home and launch the "Discovery Environment" app; (c) Choose the "Data" app, and click the "upload" menu to upload files. (d) Click the file to be submitted to the WashU Browser and edit the "Info-Type" into "bedgz." (e) Click the "Send to: Genome Browser" to obtain the URL (see Note 9).
7. Open the WashU Epigenome Browser (<http://epigenomegateway.wustl.edu>), select "mouse/mm10", go to menu "Tracks/Custom Tracks", choose "longrange" as "Track type," paste the URL obtained from **step 6**, and click "submit." A screenshot of the WashU Browser for the interaction matrix for a genomic region in chromosome 6 is shown in Fig. 2.

3.5 A/B Compartments Generation

Conventional A/B compartment prediction is based on principal component analysis (PCA) applied to a correlation matrix that records the correlation of contact profiles between genomic bins [2]. The PC1 value defines compartment status with positive and negative values corresponding to A and B compartments, respectively. Examples of A/B compartment calculation based on PCA are described elsewhere [28]. An alternative method is based on maximum-likelihood estimation by calculating the chance (or "C-score") of a genomic region being in an A compartment region from long range Hi-C interacting PETs [16].

1. Download the CscoreTool package from <https://github.com/scoutzxb/CscoreTool>.
2. `"chmod +x CscoreTool1.1"`: Make the file CscoreTool1.1 executable.
3. `"g++ generateEqualLengthBed.cpp -O3 -o generateEqual-LengthBed"`: Generate an executive file used to partition the genome into bins of equal size. **Steps 3** and **4** are done in the CscoreTool folder.

⁹The tracks submitted to the WashU Browser need to be hosted under an HTTP server, which should be accessible to the Browser. While institute's Firewall will normally block the access, services provided by CyVerse can be used to host the tracks. Be advised that both the compressed file and its index file should be uploaded.

4. `./generateEqualLengthBed mm10_chrsizes.txt mm10_20k.bed 20000`, where 20,000 is the bin size: Partition the genome into bins of equal size and save in “mm10_20k.bed”. The “mm10_chrsizes.txt” file records the size of each chromosome, available from CscoreTool package or the “resource” folder from the iHiC package.
5. Copy the executive file CscoreTool1.1 and “mm10_20k.bed” to the Hi-C data analysis folder.
6. `“iHiC_BEDPE2HiCSummary SRX1338543.bedpe SRX1338543.hicsum”`: Convert the BEDPE file into a “HiCsummary” file (<http://homer.ucsd.edu/homer/interactions/HiCtagDirectory.html>) [29].
7. `“./CscoreTool1.1 mm10_20k.bed SRX1338543.hicsum SRX1338543 24 1000000”`, where “SRX1338543” is the prefix for output files, 24 is the number of threads, and 1,000,000 is the minimal distance of PET-ends used to estimate compartment status: Calculate compartment scores (namely, the “C-scores”), which are linearly transformed into values in range –1 and 1. The resultant file “SRX1338543_cscore.bedgraph” stores the “C-scores” for all chromosomal bins (see Note 10).
8. `“iHiC_Cscore_Adj SRX1338543_cscore.bedgraph mm10_CpG_ucsc.txt”`: to ensure that a positive value in “C-score” corresponds to A compartment status. The program calculates the average “C-score” for CpG islands of a chromosome and flips the sign of “C-score” if the average is negative for the chromosome (see Note 11); the adjusted “C-score” is saved in a file with name ending with “bedgraph.adj”. File “mm10_CpG_ucsc.txt” is for CpG annotation, downloaded from UCSC and also available from the “resource” folder of the iHiC package.
9. Follow the steps in Subheading 3.4 to sort (**step 3**), compress (**step 4**), make index (**step 5**) for the adjusted “C-score” file, upload to Cyverse, and obtain a URL for the WashU Browser (**step 6**).
10. Follow the instruction in Subheading 3.4 (**step 7**) to supply the URL to the WashU Browser (choose “bedGraph” as “Track type”). A screenshot of the WashU Browser for the “C-scores” of compartments is shown in Fig. 2 for a genomic region in chromosome 6.

3.6 TAD Prediction

TAD prediction is an active research subject with dozens of tools published so far. A comprehensive evaluation of the methods was recently done by considering five criteria:

¹⁰Compared to PCA-based methods, CscoreTool requires less time and memory. In addition, unlike the PC1 values, the “C-score” is directly comparable between different samples [16]. CscoreTools could efficiently estimate A/B compartments at high resolution (e.g., 10K bps) with only 35 million PETs (human genome) and reproduce the results estimated from 3.5 billion PETs. The “bedgraph” file is in the following format: chromosome, start, end, and score.

¹¹Genomic bins from the same chromosome share compartment status if they exhibit same sign in “C-score.” Whether positive values in “C-score” denote A compartment or not depends on the data initialization during the likelihood estimation. As the initialization is a randomized process, genomic bins with negative values may turn out to come from A compartment regions. Because CpG islands are enriched in A compartments, the average “C-scores” at CpG islands can be used to tell whether or not a positive score denotes A compartment status.

robustness to window size and normalization method, cost efficiency in terms of number of reads for concordant prediction, prediction reproducibility by other methods, biology relevance to prior knowledge on TAD features, and computing time [15]. TopDom is the only tool that satisfies all the five criteria among others [15].

1. *“iHiC_III2MTX4TopDom SRX1338543.iii 20000 mm10_chrsizes.txt SRX1338543”*, where 20,000 is the bin size and “SRX1338543” is a prefix for output files: Convert the III file into an interaction matrix file accepted by TopDom for each chromosome specified in “mm10_chrsizes.txt”. The filenames of outputs end with “.mtx4topdom”.
2. Download “TopDom_v0.0.2.R” from <https://github.com/HenrikBengtsson/TopDom/>.
3. *“Rscript TopDom_wrapper.R Path_for_TopDom_v0.0.2.R SRX1338543.chr6.mtx4topdom”*: Run TopDom through an Rscript wrapper (shown here “chr6” as an example). TopDom predicts TADs one chromosome at a time. The outputs include a “binSignal” file, which is used to extract TAD boundaries in the next step.
4. *“iHiC_TADBoundary_TopDom SRX1338543.chr6.mtx4top-dom.binSignal 0.000001 SRX1338543.chr6.tadboundary.bed-graph”*: Extract and merge genomic bins that are labeled as “-1” in the “local.ext” column and is less than 0.000001 in the “p-value” column from the “binSignal” file for TAD boundaries.
5. Follow the instruction in Subheading 3.4 (**steps 3–7**) to upload the “bedgraph” file (sorted and indexed) in the previous step to Cyverse, and obtain a URL to be supplied to the WashU Browser (choosing “bedGraph” as “Track type”).
6. Follow the instruction in Subheading 3.4 (**step 7**) to supply the URL to the WashU Browser (choose “bedGraph” as “Track type”). The TAD boundaries for a genomic region of 6.6 million bps in chromosome 6 are shown in Fig. 2.

3.7 Identification of Chromatin Interaction

A Hi-C interaction matrix captures both short-range and long-range significant contacts as well as those reflecting random polymer looping events. The last few years have witnessed the development of numerous tools for the identification of significant chromatin interactions. The chapter demonstrates the use of the Fit-Hi-C method, which calculates the statistical significance of chromatin interactions from a background model considering both random polymer effects and other biases in Hi-C data [18].

1. Download and install Fit-Hi-C from <https://github.com/ay-lab/fithic>. Take note on the path of “fithic.py”, located in the “fithic” folder of the package.
2. *“iHiC_III2FitHiCInputs SRX1338543.iii SRX1338543 mm10_chrsizes.txt 20000 20000 2000000”*, where the last three parameters are bin size, minimal distance, and maximal distance: Generate “fragments file” (ending with “frag.gz”) and “interactions file” (ending with cnts.gz”) for the required “-f” and

“-i” arguments of Fit-Hi-C, respectively. The two files will be generated for each chromosome as specified in “mm10_chrsizes.txt”.

3. *“python path_to_fit-hic-c.py -f SRX1338543.chr6.frag.gz -i SRX1338543.chr6.cnts.gz -r 20000 -L 20000 -U 2000000 -o fithic_output -l SRX1338543.chr6”*: Run Fit-Hi-C to calculate the significances of chromatin interactions. Fit-Hi-C makes the prediction one chromosome a time (shown here “chr6” as an example). The output file ending with “significances.txt.gz” records the significance of the interaction for bin-pairs.
4. *“cd fithic_output”*: Locate the “fithic_output” folder.
5. *“gunzip SRX1338543.chr6.spline_pass1.res20000.significances.txt.gz”*: Decompress the file.
6. *“iHiC_FitHiC2WashU SRX1338543.chr6.spline_pass1.res20000.significances.txt 20000 0.05”*; where 20,000 is the bin size, and 0.05 is a threshold on q-value: Extract bin-pairs with q-value <0.05 from the Fit-Hi-C predictions and convert to a “longrange” format accepted by the WashU Browser for visualization.
7. Follow the instruction in Subheading 3.4 (**steps 3–7**) to obtain a URL for the “longrange” file generated from the previous step to be supplied to the WashU Epigenome Browser.
8. Follow the instruction in Subheading 3.4 (**step 7**) to supply the URL to the WashU Browser (choose “longrange” as “Track type”). The significant contacts from Fit-Hi-C for a genomic region in chromosome 6 in shown in Fig. 2.

Acknowledgments

The author thanks Dr. Xiaobin Zheng from Carnegie Institution for Science and Mr. Aniello Infante in the lab for helpful comments. The work was partially supported by National Institute of General Medical Sciences Grant 5U54GM104942-04.

References

1. Dekker J, Belmont AS, Guttman M, Leshyk VO, Lis JT, Lomvardas S, Mirny LA, O’Shea CC, Park PJ, Ren B, Politz JCR, Shendure J, Zhong S, Network DN (2017) The 4D nucleome project. *Nature* 549 (7671):219–226. 10.1038/nature23884 [PubMed: 28905911]
2. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326 (5950):289–293. 10.1126/science.1181369 [PubMed: 19815776]
3. Denker A, de Laat W (2016) The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev* 30 (12):1357–1382. 10.1101/gad.281964.116 [PubMed: 27340173]
4. Beagrie RA, Scialdone A, Schueler M, Kraemer DC, Chotalia M, Xie SQ, Barbieri M, de Santiago I, Lavitas LM, Branco MR, Fraser J, Dostie J, Game L, Dillon N, Edwards PA, Nicodemi M, Pombo A (2017) Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543 (7646):519–524. 10.1038/nature21411 [PubMed: 28273065]
5. Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai MM, Shishkin AA, Bhat P, Takei Y, Trinh V, Aznauryan E, Russell P, Cheng C, Jovanovic M, Chow A, Cai L, McDonel P,

- Garber M, Guttman M (2018) Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* 174(3):744–757. e724. 10.1016/j.cell.2018.05.024 [PubMed: 29887377]
6. Lai B, Tang Q, Jin W, Hu G, Wangsa D, Cui K, Stanton BZ, Ren G, Ding Y, Zhao M, Liu S, Song J, Ried T, Zhao K (2018) Trac-looping measures genome structure and chromatin accessibility. *Nat Methods* 15(9):741–747. 10.1038/s41592-018-0107-y [PubMed: 30150754]
 7. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G (2017) Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* 171 (3):557–572. e524. 10.1016/j.cell.2017.09.043 [PubMed: 29053968]
 8. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7):1665–1680. 10.1016/j.cell.2014.11.021 [PubMed: 25497547]
 9. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398):376–380. 10.1038/nature11082 [PubMed: 22495300]
 10. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Bluthgen N, Dekker J, Heard E (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485 (7398):381–385. 10.1038/nature11049 [PubMed: 22495304]
 11. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148 (3):458–472. 10.1016/j.cell.2012.01.010 [PubMed: 22265598]
 12. Ay F, Noble WS (2015) Analysis methods for studying the 3D architecture of the genome. *Genome Biol* 16:183 10.1186/s13059-015-0745-7 [PubMed: 26328929]
 13. Lajoie BR, Dekker J, Kaplan N (2015) The Hitchhiker’s guide to Hi-C analysis: practical guidelines. *Methods* 72:65–75. 10.1016/j.ymeth.2014.10.031 [PubMed: 25448293]
 14. Forcato M, Nicoletti C, Pal K, Livi CM, Ferrari F, Bicciato S (2017) Comparison of computational methods for Hi-C data analysis. *Nat Methods* 14(7):679–685. 10.1038/nmeth.4325 [PubMed: 28604721]
 15. Zufferey M, Tavernari D, Oricchio E, Ciriello G (2018) Comparison of computational methods for the identification of topologically associating domains. *Genome Biol* 19(1):217 10.1186/s13059-018-1596-9 [PubMed: 30526631]
 16. Zheng X, Zheng Y (2018) CscoreTool: fast Hi-C compartment analysis at high resolution. *Bioinformatics* 34(9):1568–1570. 10.1093/bioinformatics/btx802 [PubMed: 29244056]
 17. Shin H, Shi Y, Dai C, Tjong H, Gong K, Alber F, Zhou XJ (2016) TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res* 44(7):e70 10.1093/nar/gkv1505 [PubMed: 26704975]
 18. Ay F, Bailey TL, Noble WS (2014) Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res* 24(6):999–1011. 10.1101/gr.160374.113 [PubMed: 24501021]
 19. Li D, Hsu S, Purushotham D, Sears RL, Wang T (2019) WashU Epigenome Browser update 2019. *Nucleic Acids Res* 47(W1): W158–W165. 10.1093/nar/gkz348 [PubMed: 31165883]
 20. Yan J, Chen SA, Local A, Liu T, Qiu Y, Dorighi KM, Preissl S, Rivera CM, Wang C, Ye Z, Ge K, Hu M, Wysocka J, Ren B (2018) Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res* 28(3):387 10.1038/cr.2018.18 [PubMed: 29497152]
 21. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9 (4):357–359. 10.1038/nmeth.1923 [PubMed: 22388286]
 22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16):2078–2079. 10.1093/bioinformatics/btp352 [PubMed: 19505943]
 23. Quinlan AR (2014) BEDTools: The Swiss-Army tool for genome feature analysis. *Curr Protoc Bioinformatics* 47:11.12.1–11.1234. 10.1002/0471250953.bi1112s47

24. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, Li D, Choudhary MNK, Li Y, Hu M, Hardison R, Wang T, Yue F (2018) The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 19(1):151 10.1186/s13059-018-1519-9 [PubMed: 30286773]
25. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL (2016) Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst* 3(1):99–101. 10.1016/j.cels.2015.07.012 [PubMed: 27467250]
26. Akdemir KC, Chin L (2015) HiCPlotter integrates genomic data with interaction matrices. *Genome Biol* 16:198 10.1186/s13059-015-0767-1 [PubMed: 26392354]
27. Kerpedjiev P, Abdennur N, Lekschas F, McCallum C, Dinkla K, Strobelt H, Lubner JM, Ouellette SB, Azhir A, Kumar N, Hwang J, Lee S, Alver BH, Pfister H, Mirny LA, Park PJ, Gehlenborg N (2018) HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol* 19 (1):125 10.1186/s13059-018-1486-1 [PubMed: 30143029]
28. Miura H, Poonperm R, Takahashi S, Hiratani I (2018) Practical analysis of Hi-C data: generating A/B compartment profiles. *Methods Mol Biol* 1861:221–245. 10.1007/978-1-4939-8766-5_16 [PubMed: 30218370]
29. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38 (4):576–589. 10.1016/j.molcel.2010.05.004 [PubMed: 20513432]
30. Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* 112 (47):E6456–E6465. 10.1073/pnas.1518552112 [PubMed: 26499245]

Fig. 1.

Quality inspection on a Hi-C library of mouse ES E14 (SRX1338543). **(a)** Statistics on the PET alignments from Bowtie2. **(b)** Proportion of different combinations of strand orientations as a function of genomic distance between PET ends. **(c)** Standard deviation of the proportions in **(b)** as a function of genomic distance. The convergence occurs after 1 K bp (dashed blue line). **(d)** Contact probability as a function of genomic distance intervals. Genomics ranges correspond to TAD indicated by blue dash lines

Fig. 2. A screenshot of the WashU Epigenome Browser for the interaction matrix (resolution = 20 K bps), “C-score” by CScoreTool for A/B compartments (blue for A and yellow for B), TAD boundaries by TopDom, and significant contacts by Fit-Hi-C for a genomic region of 6.6 million bps in chromosome 6 of mouse ES E14