SCIENCE CHINA PRESS

SCIENCE CHINA Life Sciences
life.scichina.com link.springer.com
https://doi.org/10.1007/s11427-023-2441-0

RESEARCH PAPER

# Hi-Tag: a simple and efficient method for identifying protein-mediated long-range chromatin interactions with low cell numbers

Xiaolong Qi[1†], Lu Zhang[1†], Qiulin Zhao[1], Peng Zhou[1], SaiXian Zhang[1], Jingjin Li[1], Zhuqing Zheng[1], Yue Xiang[1], Xueting Dai[1], Zhe Jin[1], Yaobang Jian[1], Xinyun Li[1,2,3*], Liangliang Fu[1,2*] & Shuhong Zhao[1,2,3*]

[1]Key Lab of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education and Key Laboratory of Swine Genetics and Breeding of Ministry of Agriculture, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China;
[2]The Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China;
[3]Hubei Hongshan Laboratory, Frontiers Science Center for Animal Breeding and Sustainable Production, Wuhan 430070, China

†Contributed equally to this work
*Corresponding authors (Xinyun Li, email: xyli@mail.hzau.edu.cn; Liangliang Fu, email: fuliangliang2017@mail.hzau.edu.cn; Shuhong Zhao, email: shzhao@mail.hzau.edu.cn)

Received 15 May 2023; Accepted 12 July 2023; Published online 23 January 2024

**Protein-mediated chromatin interactions can be revealed by coupling proximity-based ligation with chromatin immunoprecipitation. However, these techniques require complex experimental procedures and millions of cells per experiment, which limits their widespread application in life science research. Here, we develop a novel method, Hi-Tag, that identifies high-resolution, long-range chromatin interactions through transposase tagmentation and chromatin proximity ligation (with a phosphorothioate-modified linker). Hi-Tag can be implemented using as few as 100,000 cells, involving simple experimental procedures that can be completed within 1.5 days. Meanwhile, Hi-Tag is capable of using its own data to identify the binding sites of specific proteins, based on which, it can acquire accurate interaction information. Our results suggest that Hi-Tag has great potential for advancing chromatin interaction studies, particularly in the context of limited cell availability.**

**Hi-Tag | chromatin interactions | transposase tagmentation | chromatin proximity ligation | low cell numbers**

## INTRODUCTION

Over the past two decades, the developments of chromosome conformation capture (3C) (Dekker et al., 2002) and its derivatives, circular chromosome conformation capture (4C) (Simonis et al., 2006) and 3C-copy (5C) (Dostie and Dekker, 2007), as well as chromosome conformation capture coupled with sequencing (Hi-C) (Lieberman-Aiden et al., 2009), chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) (Li et al., 2017), chromosome conformation capture coupled with immunoprecipitation (HiChIP) (Mumbach et al., 2016), and proximity ligation-assisted ChIP-Seq (PLAC-Seq) (Yu et al., 2021) have greatly advanced our ability to explore the three-dimensional (3D) structure of chromatin and its role in gene expression regulation (Krijger and de Laat, 2016; Wang et al., 2015). However, these techniques still have some limitations, including the requirement for large numbers of cells, low capture efficiency, and difficulty in analyzing sparse cell samples. In this context, we develop Hi-Tag, a novel method that utilizes transposase tagmentation and chromatin proximity ligation to identify high-resolution, long-range, protein-mediated chromatin interactions using as few as 0.1 million cells. In this study, we demonstrate the potential of Hi-Tag to effectively characterize the 3D chromatin conformation in the nucleus, providing a simpler and more accessible approach for studying chromatin interactions.

## RESULTS

### Hi-Tag can efficiently profile chromatin interactions mediated by histone H3-lysine-27 acetylation

The Hi-Tag method is developed to address the limitations of existing techniques for studying protein-mediated chromatin interactions. In Hi-Tag, cells are crosslinked with formaldehyde; DNA is digested with a restriction enzyme (AluI) to generate blunt ends; and long-range genomic DNA contacts in the nuclei are established *in situ* using a phosphorothioate-modified linker marked with biotin at the junction (see MATERIALS AND METHODS). The cleavage under targets and tagmentation (CUT&Tag) (Kaya-Okur et al., 2019) strategy is employed to construct a library and directly capture long-range interactions associated with a target protein. After all the proteins are digested by proteinase K, the biotin-containing fragments are retained by streptavidin beads. The appropriate number of amplification cycles is determined using quantitative PCR to reduce GC content and amplicon size bias during library amplification to avoid saturation (Figure 1A).

We performed histone H3-lysine-27 acetylation (H3K27ac) Hi-Tag to specifically capture genome-wide interactions between enhancers and active promoters in K562 cells. Statistical analysis showed that the correlation coefficient between Hi-Tag biological replicates was above 0.8 (Figure S1A in Supporting Information),

indicating the excellent reproducibility of Hi-Tag. To evaluate the sensitivity of H3K27ac Hi-Tag, we downsampled reads from the raw data. We found a high correlation between low-sequencing and high-sequencing depths, demonstrating the extremely high efficiency of Hi-Tag (Figure S1B in Supporting Information). The identified paired-end tags (PETs) accounted for more than 30% of the total sequenced H3K27ac Hi-Tag reads in the K562 cell line (Table S1 in Supporting Information). A total of 117,053 significant H3K27ac-mediated chromatin loops were detected at a resolution of 5 kb, with interaction region size ranging from 20 kb to 2 Mb (Table S1 in Supporting Information). Additionally, two replicates yielded highly consistent results, and both of them revealed a high proportion of cis long-range interactions (Figure S1C and D in Supporting Information). The distance distributions of interactions identified by Hi-Tag were extremely similar under different sequencing depths (Figure S1E in Supporting Information). Hi-Tag captured more chromatin interactions at a higher resolution and a lower cell number than Hi-C and HiChIP, suggesting that Hi-Tag has a higher capture efficiency (Figure S1F and G in Supporting Information). Next, we compared chromatin interactions captured by Hi-Tag, HiChIP, and Hi-C at different resolutions (100, 25, 10 and 5 kb). At resolutions of 25 and 100 kb, the correlation coefficient between Hi-Tag and Hi-C data surpassed 0.7, while at 5 kb resolution, the correlation coefficient was reduced to 0.5 (Figure 1B). To reveal the differences in chromatin interactions captured by Hi-C and Hi-Tag, we generated normalized interaction maps at 5 kb resolution and a heatmap of signal differences (Hi-C signals subtracted by Hi-Tag signals) (Figure 1C). We found that there were significantly more Hi-Tag signals than Hi-C signals at many loci, implying that Hi-Tag could detect chromatin interaction signals that Hi-C was unable to detect at high resolution (Figure 1C and D). These findings suggest that the concordance between Hi-Tag and Hi-C data is influenced by genomic resolution. It is worth noting that the correlation coefficients between Hi-Tag data and HiChIP data at four resolutions (5, 10, 25 and 100 kb) all exceeded 0.8 (Figure 1B). A map of normalized interactions at different resolutions also

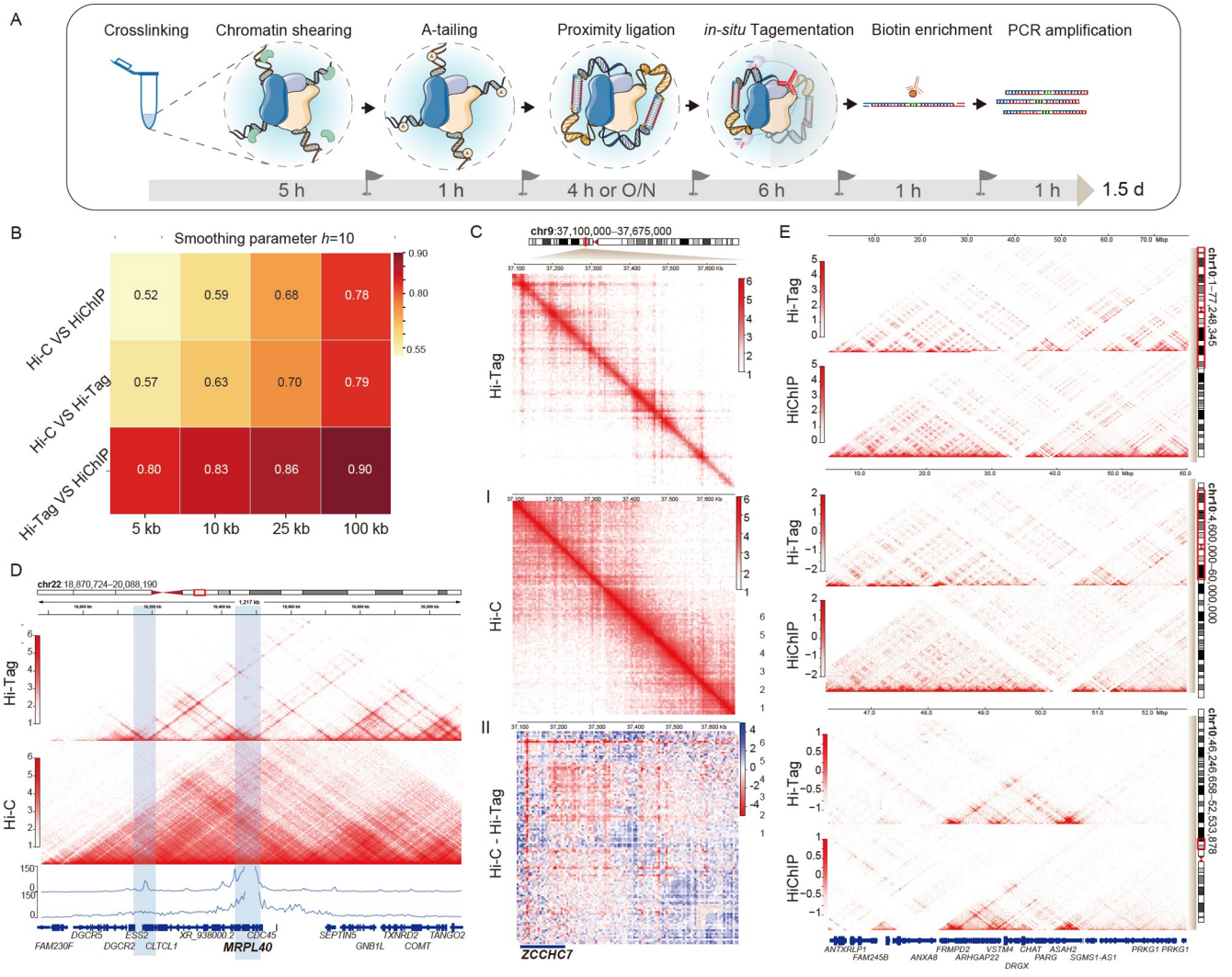


**Figure 1.** Hi-Tag identifies *in situ* chromatin loops mediated by H3K27ac. A, Schematic of the Hi-Tag method. B, Correlation analysis between Hi-Tag, HiChIP and Hi-C raw reads at resolutions of 100, 25, 10 and 5 kb. C, Normalized interaction maps generated using Hi-Tag (up) and Hi-C (middle) data at the *ZCCHC7* locus. The heatmap of interaction signals at the *ZCCHC7* locus (bottom) after the Hi-Tag matrix was subtracted from the Hi-C matrix. D, Virtual 4C plot and raw interaction maps generated using Hi-Tag and Hi-C data at *CDC45* locus. The anchors are indicated by shading. The virtual 4C signals are shown as paired-end tag counts in a 10-kb bin. E, K562 H3K27ac Hi-Tag and HiChIP raw interaction maps of a locus on chromosome 10 at the resolution of 100 kb (up), 25 kb (middle) and 5 kb (down).


https://doi.org/10.1007/s11427-023-2441-0

showed a high degree of similarity between Hi-Tag and HiChIP (Figure 1E). We also compared the CTCF Hi-Tag data with published ChIA-PET and HiChIP data. Hi-Tag significantly reduced the experimental timeline and the requirements for cell input and sequencing depth, compared with ChIA-PET and HiChIP (Figure S1I and J in Supporting Information). Hi-Tag technology simplifies experimental procedures and reduces operational complexity by utilizing an *in situ* transposition reaction to obtain specific chromatin complexes, meanwhile ensuring comparable results to traditional methods. Therefore, Hi-Tag enables more efficient detection of chromatin interactions mediated by specific proteins with greater ease.

## Hi-Tag can efficiently identify genome-wide protein binding sites and chromatin interactions

A Hi-Tag experiment produces two genome-wide data sets, namely protein-binding site data (the same as CUT&Tag data) and chromatin interaction data. Further, we explored whether the protein binding sites generated by Hi-Tag could replace those generated by CUT&Tag or ChIP-seq in the FitHiChIP model to identify chromatin interactions (Bhattacharyya et al., 2019). We separately utilized all reads and self-ligation reads from Hi-Tag data to identify significant peaks, which we named Hi-Tag one-dimensional (1D) peaks. We found that the peak set identified from CUT&Tag data was highly correlated with the Hi-Tag 1D peak set and that both peak sets exhibited high MACS2 confidence scores (Figure 2A–C; Figure S2 in Supporting Information). These results suggest a high degree of positional overlap between the two peak sets of Hi-Tag 1D and CUT&Tag, with each unique peak occurring in weak-signal regions. In addition, the signal intensity of the Hi-Tag 1D peak is higher than that of CUT&Tag in the overlapping region (Figure 2D and H). The genome-wide distribution of Hi-Tag 1D peaks was consistent with that of the CUT&Tag peaks, indicating that Hi-Tag has the potential to produce both a map of protein-specific binding sites in the entire genome and a map of the chromatin interactions mediated by these binding sites. Next, we compared the significant loops obtained from the FitHiChIP model based on the aforementioned Hi-Tag 1D peaks and CUT&Tag peaks. We defined the loops obtained based on the Hi-Tag 1D peaks as "1D peak-based loops", and similarly, the loops obtained based on the CUT&Tag peaks were named "CUT&Tag-based loops". We found that almost all of the anchors of 1D peak-based loops overlapped with those of CUT&Tag-based loops (Figure 2E). 1D peak-based loops and CUT&Tag-based loops showed similar distance distributions (Figure 2F). The overlapping loop anchors had a stronger enrichment of functionally active histones than non-overlapping loop anchors (1D peak-based loops or CUT&Tag-based loops) (Figure 2G and H). These findings further confirm that Hi-Tag can produce both a map of protein-specific binding sites across the entire genome and a map of the chromatin interactions mediated by these sites. This eliminates the need for additional input of protein-binding site information to identify chromatin interactions, thus simplifying the analysis workflow and reducing software requirements.

## Hi-Tag can handle sparse biological samples

To determine whether sample input affects the performance of Hi-Tag, we conducted Hi-Tag experiments on transcription factor CTCF within cell numbers ranging from 0.05 million to 1 million. We found that all experiments yielded high-quality chromatin profiles, and that Hi-Tag provided stable high-quality loop output at different cell numbers and sequencing depths (Figure 3A). Based on CUT&Tag peaks or Hi-Tag 1D peaks, Hi-Tag identified approximately 7,000 CTCF-mediated chromatin interactions at 1 million cells. Even only at 0.05 million cells, Hi-Tag identified 2,266 chromatin interactions. However, we found that when the number of cells exceeded 0.1 million, the number of identified chromatin interactions stably increased with the increasing number of input cells. Therefore, we recommend using more than 0.1 million cells as the sample size to optimize the results. Although fewer chromatin interactions are identified when using low cell numbers, the identified chromatin interactions with low cell numbers as input exhibit high consistency with those identified with 1 million cells as input, indicating the robustness of the Hi-Tag technique in the case of limited amounts of starting materials (Figure 3B and E). In CTCF-mediated chromatin loops, the CTCF motif in anchors typically converges in orientation. To better understand the high-confidence loops identified by FitHiChIP using Hi-Tag data, we characterized the binding of CTCF to both end anchors of the loops. Our results showed that as expected, up to 90% of the loops were anchored by CTCF (Figure 3C), and about 80% of the high-confidence loops identified by Hi-Tag exhibited strong CTCF convergent patterns (Figure 3D). In addition, to further test Hi-Tag's ability to capture chromatin interactions at low cell numbers, we used the enhancer/promoter-associated epigenetic mark H3K27ac (Creyghton et al., 2010; Zhu et al., 2022), as a candidate to specifically capture the interactions between enhancers and active promoters across the entire genome at 0.1 million GM12878 cells via Hi-Tag. We identified 16,289 and 17,383 chromatin loops in two replicates, respectively. Correlation analysis confirmed the high stability of the output data from Hi-Tag under low cell number conditions (Figure S3 in Supporting Information). Overall, these results suggest that Hi-Tag is a robust and reliable technique for studying chromatin interactions mediated by specific proteins, even in the case of limited starting materials.

## CONCLUSION

Hi-Tag is a fast, efficient, and technically simplified method for identifying specific protein-mediated chromatin conformations. Hi-Tag uses transposase tagmentation to achieve *in situ* shearing of specific protein-mediated chromatin complexes. Sequencing adapters are directly added to the target fragments, thus greatly reducing experimental steps, compared with conventional techniques. Hi-Tag experiments can be performed in a single reaction vessel, hence eliminating the need to use instruments with unstable test conditions (such as sonicators) and improving the consistency of valid data output. In this study, we produced DNA-binding profiles and H3K27ac- and CTCF-mediated chromatin interaction maps using the Hi-Tag technique. We found that Hi-Tag provides comparable data at a low cell number input, and thus Hi-Tag is less costly than other existing techniques.

The critical technical advancement of Hi-Tag lies in the incorporation of hyperactive PG/A-Tn5 transposase-based chromatin and the addition of a sequencing adapter into a one-step reaction to improve the efficiency of chromatin DNA library construction, which eventually greatly improves the efficiency of

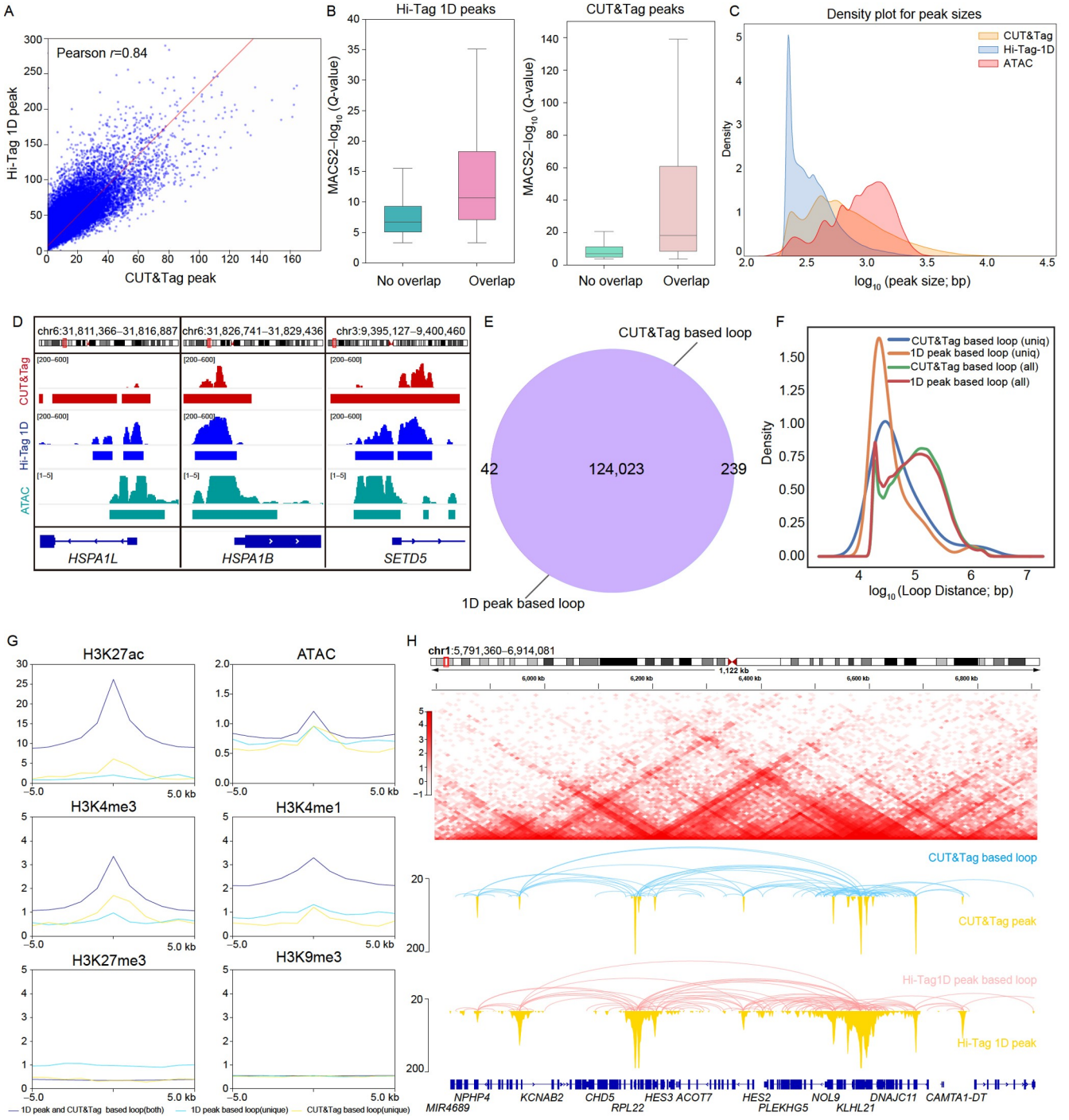


**Figure 2.** Two genome-wide datasets generated by H3K27ac Hi-Tag. A. Correlation between H3K27ac CUT&Tag peaks and H3K27ac Hi-Tag 1D peaks. B. MACS2 confidence scores for overlapping and non-overlapping peaks between CUT&Tag (right) and Hi-Tag 1D (left). C. Density plots of the peak sizes of H3K27ac CUT&Tag, H3K27ac Hi-Tag 1D, and ATAC. D. The peak widths of H3K27ac CUT&Tag, H3K27ac Hi-Tag 1D, and ATAC in the genomic regions of *HSPA1L*, *HSPA1B*, and *SETD5*. E. Venn diagram showing the overlap of loop anchors identified using Hi-Tag 1D peaks and CUT&Tag peaks. The Hi-Tag 1D peak-based loop anchors and the CUT&Tag-based loop anchors were both extended upstream and downstream by 10 kb. The loops with anchors overlapping at either end were defined as overlapping, and the loops with no overlapping anchor at either end were designated as specific. F. Density plot of interaction distance for Hi-Tag all 1D peak-based loops, all CUT&Tag-based loops, unique Hi-Tag 1D peak-based loops and unique CUT&Tag-based loops. G. Enrichment of H3K27ac, H3K4me3, H3K4me1, H3K27me3, H3K9me3, and ATAC signals in the anchor regions of all 1D peak-based loops, unique Hi-Tag 1D peak-based loops and unique CUT&Tag-based loops. H. Genomic screenshots of Hi-Tag 1D peak-based loops and CUT&Tag-based loops. Raw interaction heat map demonstrating chromatin interactions captured by H3K27ac Hi-Tag. The blue and pink lines show 1D peak-based loops and CUT&Tag-based loops, respectively. Yellow indicates the CUT&Tag and Hi-Tag 1D peak signal of H3K27ac.

chromatin interaction capture. Key experimental steps, including restriction digestion, proximity ligation, transposase fragmentation, and sequencing adapter addition, are performed *in situ*, thus maintaining the integrity of individual cells or nuclei. Therefore, the Hi-Tag method can be further applied to various single-cell platforms for developing single-cell Hi-Tags.


https://doi.org/10.1007/s11427-023-2441-0

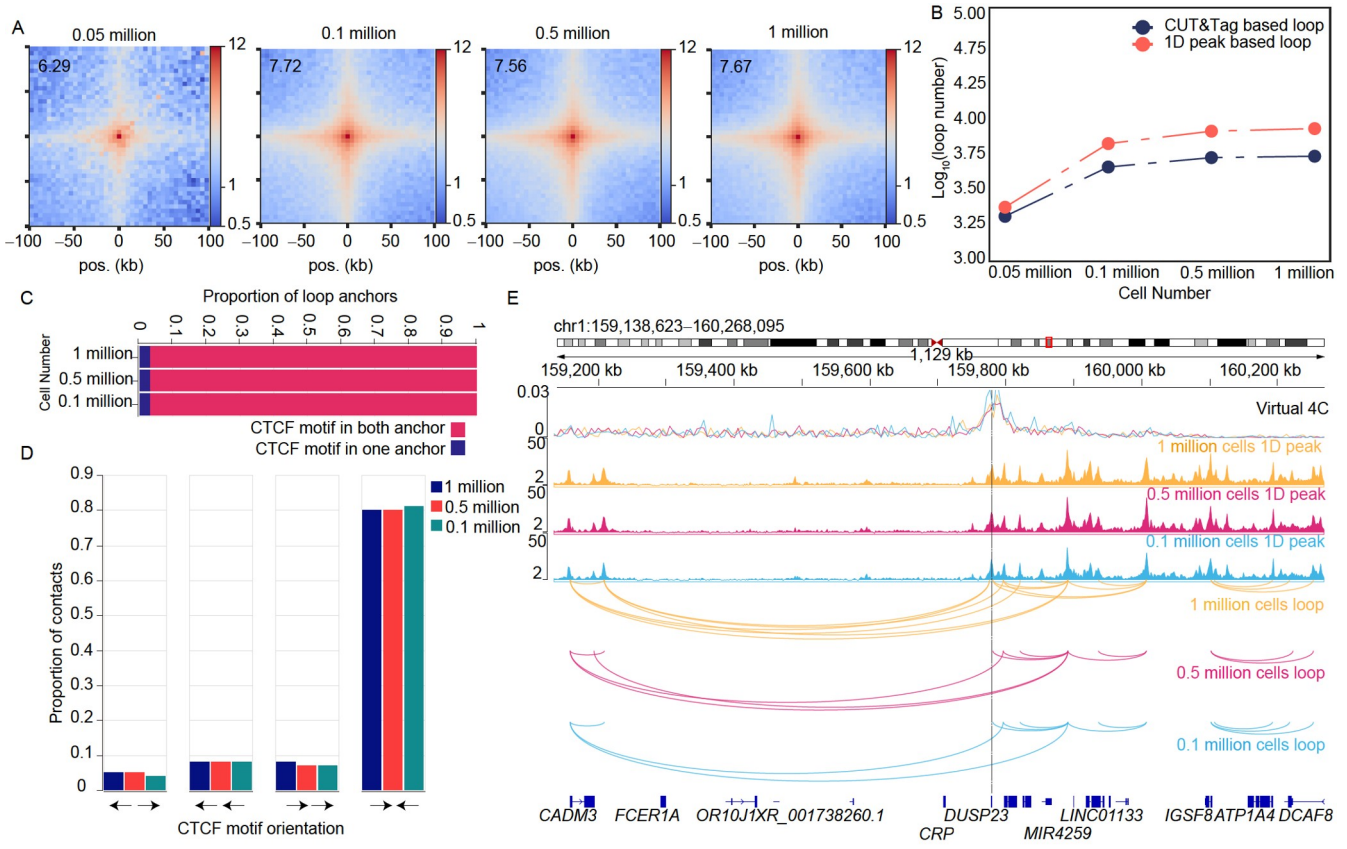


**Figure 3.** Stability of Hi-Tag data for CTCF-mediated loops obtained using different numbers of cells. A. APA plots of loops identified by Hi-Tag analysis of CTCF-mediated chromatin loops performed using different numbers of cells (1 million, 0.5 million, 0.1 million and 0.05 million). B. Line chart of the number of CUT&Tag-based loops and Hi-Tag 1D peak-based loops identified at different cell numbers. C. Bar plot demonstrates the presence of CTCF peaks at one or both anchor points of CTCF 1D peak-based loops. D. Bar chart of the proportions of contacts with different CTCF motif orientations identified using different cell numbers. E. Genomic screenshots showing CTCF-mediated chromatin loops identified by Hi-Tag in the *DUSP23* locus across different cell numbers.

Compared with existing chromatin interaction capture techniques such as HiChIP and ChIA-PET, Hi-Tag exhibits multiple advantages in simplifying experimental procedures and overcoming sample-size limitations. Hi-Tag is capable of identifying protein-mediated chromatin interactions using only a cell number of 0.05 million, which is much smaller than what is required by existing technologies. Similar to PLAC-Seq, Hi-Tag has a specific analysis pipeline designed for the characteristics of the library, but Hi-Tag is different from HiChIP, which commonly uses the general pipeline of Hi-C. Compared with HiCuT (Sati et al., 2022), Hi-Tag provides a more straightforward sample handling process, since Hi-Tag uses a single fixation method and the highly efficient Bridge linker to connect adjacent DNA ends. Hi-Tag allows specific enrichment of chromatin interactions through the use of biotin-labeled linkers, whereas HiCuT does not exhibit such an enrichment. Hi-Tag enhances the sensitivity and specificity of active chromatin loop detection, compared with HiCuT. Additionally, gel purification, rather than magnetic bead-based recovery, is employed to avoid small fragment bias during library fragment selection to ensure the recovery of authentic library fragment distribution. In Hi-Tag, the sequencing approach utilizes 150 bp paired-end reads, resulting in longer read lengths to ensure accurate alignment. Comparative analysis of sequencing data reveals that HiCuT yields an unusually low proportion of valid interaction pairs due to its lack of specific chromatin interaction enrichment. In contrast, Hi-Tag data exhibit a higher proportion of valid pairs in the final library, enabling the identification of more high-confidence loops using mainstream software (Figure S4 in Supporting Information). In conclusion, the Hi-Tag method represents an advancement in the study of chromatin interactions. By providing a simple and efficient alternative, Hi-Tag overcomes the limitations of sample availability and experimental complexity in existing techniques.

## MATERIALS AND METHODS

### Hi-Tag library construction

Cell culture and fixation. About 1 million cultured cells were resuspended in freshly made 1% formaldehyde (Sigma-Aldrich, USA) to a volume of 30 mL. The suspension was incubated in a rotator at room temperature for 5 min. To terminate the formaldehyde crosslinking reaction, 1.5 mL glycine (2.5 mol $L^{-1}$) was added to final concentration of 125 mmol $L^{-1}$, and the sample was incubated at room temperature for 5 min with rotation. The cell suspension was centrifuged at 2,500×*g* for 10 min at 4°C in a horizontal centrifuge. After centrifugation, cells were resuspended in cold PBS, flash-frozen in liquid nitrogen, and stored at −80°C or directly used for Hi-Tag experiments.

Cell lysis and *in situ* digestion. To obtain crosslinked nuclei, 500 μL of lysis buffer (containing 10 mmol $L^{-1}$ Tris-HCl (pH 8.0),

10 mmol $L^{-1}$ NaCl, 0.2% v/v Igepal CA630, and 1× protease inhibitors (Sigma-Aldrich)) was added, and the sample was incubated on ice for 15 min. The mixture was centrifuged at 2,000×*g* for 5 min at 4°C. Afterwards, the precipitate was collected, washed with phosphate-buffered saline plus 0.1% Tween-20 (PBST) and CutSmart wash buffer (1× CutSmart (NEB, USA) and 0.1% Tween-20 (Sigma-Aldrich)), resuspended in 0.3% sodium dodecyl sulfate (SDS) buffer (1× CutSmart and 0.3% SDS (Thermo Fisher Scientific, USA)), and incubated at 62°C for 2 min. To neutralize the SDS, the precipitate was resuspended with 200 μL Triton X-100 (10%) (Thermo Fisher Scientific) and incubated at room temperature for 10 min. The cell culture was centrifuged at 2500×*g* for 5 min at 4°C in a horizontal centrifuge to obtain nuclei, which were washed with PBST and CutSmart wash buffer. Finally, the nuclei were resuspended in a reaction system containing 1× CutSmart buffer, 1% Triton X-100, and 1× protease inhibitors. The AluI (Thermo Fisher Scientific) restriction enzyme was added to the reaction system to a final concentration of 0.6 U $\mu L^{-1}$, and the sample was incubated at 37°C for 4 h in a Thermomixer (1,200 r $min^{-1}$, 15 s/pause, 2 min).

*In situ* chromatin ligation with DNA bridge linker. The cell samples were centrifuged at 2,500×*g* for 5 min at 4°C, and the supernatant was removed to obtain the precipitate. The precipitate was washed with 1× NEB buffer 2 containing 0.1% Tween 20, resuspended in an A-tailing reaction system (containing 1× NEB buffer 2, 1% Triton X-100, 0.2 mmol $L^{-1}$ dATP, and 10 μL Klenow 3′-5′ exo- (NEB)), and incubated at 37°C for 1 h in a Thermomixer (900 r $min^{-1}$, 15 s/pause, 2 min). The mixture was centrifuged at 2500×*g* for 5 min at 4°C, and the supernatant was removed to obtain the precipitate. The precipitate was washed with PBST and 1× T4 DNA ligase buffer (NEB) containing 0.1% Tween 20. Finally, the nuclei were resuspended in 996 μL of ligation mix (1× T4 DNA ligase buffer, 1% Triton X-100, 0.1 mg $mL^{-1}$ bovine serum albumin (BSA, Sigma-Aldrich), and 6 μL T4 DNA ligase), and 4 μL (200 ng $\mu L^{-1}$) of the Phosphorothioate & Biotin Linker (forward primer: CG*CG*A-T*AT*C/iBiotin-dT/T*AT*CT*GA*CT; reverse primer: GT*CA*-GA*TA*AG*AT*AT*CG*C*GT) was added. Subsequently, the mixture was incubated at 25°C for 4 h.

*In situ* chromatin immunoprecipitation. The cell culture was resuspended in 1 mL of wash buffer (20 mmol $L^{-1}$ HEPES (pH 7.5), 150 mmol $L^{-1}$ NaCl, 0.5 mmol $L^{-1}$ spermidine (NEB), and 1× protease inhibitor), and temporarily stored on ice. ConA beads were pre-warmed to room temperature and washed twice with binding buffer (20 mmol $L^{-1}$ HEPES (pH 7.5), 10 mmol $L^{-1}$ KCl, 1 mmol $L^{-1}$ $CaCl_2$, and 1 mmol $L^{-1}$ $MnCl_2$ (Sigma-Aldrich)). Afterward, the cleaned ConA beads were resuspended in 50 μL of binding buffer. Then, the ConA beads were gently mixed with cells at room temperature for 10 min. The beads were collected by placing a tube on a magnet stand, and the liquid was removed. The beads were resuspended in 500 μL of Antibody buffer (20 mmol $L^{-1}$ HEPES (pH 7.5), 150 mmol $L^{-1}$ NaCl, 12.5 μL 0.5 mmol $L^{-1}$ spermidine, 0.05% digitonin, 2 mmol $L^{-1}$ EDTA, and 0.1% BSA), 2 μg of primary antibody (H3K27ac and CTCF) was added, and the beads were incubated for 2 h at room temperature on a rotator to allow the binding of the target protein. After the removal of the primary antibody binding solution, 500 μL of Dig-wash buffer (20 mmol $L^{-1}$ HEPES (pH 7.5), 150 mmol $L^{-1}$ NaCl, 12.5 μL 0.5 mmol $L^{-1}$ spermidine, 0.05% digitonin) containing 2 μg of secondary antibody (anti-IgG H&L) was added, and the sample was incubated for 1 h on a rotator at room temperature. The beads were collected and washed three times with 700 μL of Dig-wash buffer. The cleaned beads were resuspended in 500 μL of Dig-300 buffer (20 mmol $L^{-1}$ HEPES (pH 7.5), 300 mmol $L^{-1}$ NaCl, 12.5 μmol $L^{-1}$ spermidine, 0.01% digitonin (Sigma-Aldrich)) containing 4 μmol $L^{-1}$ of Hyperactive pG/A-Tn5 (Vazyme, Nanjing, China) and incubated for 1 h. Finally, the beads were washed more than three times with Dig-300 buffer to remove free PG/A-Tn5 as completely as possible to reduce non-specific cleavage. Subsequently, the beads were resuspended in 500 μL of tagmentation buffer (Dig-300 buffer with 10 mmol $L^{-1}$ $MgCl_2$ (Sigma-Aldrich)), and incubated at 37°C for 1 h in a shaking Thermomixer at 600 r $min^{-1}$ to activate the transposition reaction. To stop transposition and digest the protein, 16.7 μL of 0.5 mol $L^{-1}$ EDTA, 10 μL of 10% SDS, and 15 μL of 20 mg $mL^{-1}$ Proteinase K (NEB) were added to the bead suspension solution, and the sample was incubated at 55°C for 1 h.

DNA purification and library preparation. Phase-Lock tubes were used for the final step of the immunoprecipitation process. For DNA extraction, the tubes were spun at 12,000×*g* for 1 min, and then 500 μL of phenol:chloroform:isoamyl alcohol (25:24:1) was added into each tube. After a brief centrifugation, tubes were placed on a magnet stand, and the supernatant was transferred to a Phase-Lock tube. The tubes were vigorously shaken until the contents became milky white. The tubes were then centrifuged at maximum speed for 5 min, and the aqueous layer was carefully transferred to a new tube. To the new tube was added 50 μL of 3 mol $L^{-1}$ NaAc (Sigma-Aldrich) and 2 μL of 20 mg $mL^{-1}$ glycogen, and the mixture was gently vortexed. Then, 550 μL of isopropanol was added, and the mixture was incubated at −20°C for 30 min, followed by centrifugation at 12,000×*g* for 30 min at 4°C. The supernatant was removed, and the precipitate was washed with 75% ethanol. After a quick centrifugation, all the liquid was removed with a 10 μL pipette. The tube was air-dried at room temperature, and 100 μL of water was added to dissolve the DNA precipitate. Streptavidin C1 beads (Thermo Fisher Scientific) were used for biotin pull-down assays. The enriched DNA was then subjected to quality control and library preparation following the HiChIP protocol (Krueger, 2012). To avoid size bias during fragment selection, DNA purification was performed using gel electrophoresis. The library sample was loaded onto a 2% agarose gel and cut out 200–1,000 bp fragments after electrophoresis.

## Hi-Tag data processing

The raw sequencing data were subjected to quality control to remove low-quality reads. Trim Galore (Krueger, 2012) was used to remove the adapter sequences. A custom script was used to extract the paired-end reads containing linker sequences, and the extracted sequences were named PETs. After removing the linker, only those paired-end reads whose length was greater than 16 bp were retained and used for subsequent analysis.

To improve the validity of the data and accommodate the library characteristics, the sequences were mapped to the human reference genome (hg38) using two subcommands (aln and mem) in BWA (Li, 2013; Li and Durbin, 2009). The mapped reads with MAPQ quality scores ≥30 were selected for further analysis.

Since redundant information potentially generated in the PCR amplification step of the experiment needs to be removed during analysis, we used pairtools (Abdennur et al., 2023) to parse and deduplicate the pair files, and only UU (unique-unique)-type pairs were retained for subsequent analysis. In addition, noise information was removed from the data using an in-house script incorporating enzyme site information. Hi-Tag-identified chromatin interactions were divided into bins with various resolutions and used for further analysis. To quantify the chromatin interactions captured by Hi-Tag using different numbers of cells, we performed stacking analysis. The APA values were calculated using coolpup.py (Flyamer et al., 2020) (version 1.0.0) with default parameters.

## DNA loop calling based on Hi-Tag data

For DNA loop calling, we used FitHiChIP (version 9.0), which identifies significant interactions using FitHiChIP (L) and FitHiChIP (L+M) models with coverage bias regression. FitHiChIP can use either peak-to-peak (stringent) or peak-to-all (loose) loops for learning the background and spline fitting. In this study, we chose "stringent" for processing CTCF Hi-Tag data and "loose" for H3K27ac Hi-Tag data. The FitHiChIP parameters corresponding to "stringent" and "loose" are "UseP2PBackgrnd=1" and "UseP2PBackgrnd=0", respectively. The default significance coefficient was applied with a false discovery rate threshold of 0.01. The resolution was set as 5 kb, and the length range of loop was set as 20 kb–2 Mb.

## ChIP peak calling based on Hi-Tag data

The dangling-end reads and self-ligation reads from the Hi-Tag output were integrated and transformed into BED file format as input for MACS2 (Gaspar, 2018). The ChIP peaks were identified using parameters consistent with ENCODE (de Souza, 2012) standards. The degree of overlap between CUT&Tag peaks and the Hi-Tag 1D peaks was analyzed using the "intersect" command in bedtools (Quinlan and Hall, 2010). Non-overlapping peaks were identified using the "intersect-v" option in bedtools.

## Experimental reproducibility analysis

Hi-Tag experimental reproducibility was analyzed by comparing Hi-Tag replicates, and scatterplots were generated based on read counts. For the comparison of Hi-Tag with other methods, the shared loops identified by various methods were merged to obtain a loop union set. We used the HiCExplorer (Wolff et al., 2018) tool to visualize Hi-Tag data showing the change in chromatin interaction frequency with increasing genome distance.

The Pearson correlation coefficients among different biological replicates or different chromatin capture methods were calculated based on raw reads using python functions. For generating scatterplots, reads obtained with various sequencing depths were normalized, and the obtained reads were further quantile normalized across biological replicates or chromatin capture methods.

## CTCF motif orientation analysis

The positional weight matrix (PWM) of a known human CTCF motif was extracted from JASPAR database (Fornes et al., 2020) (ID: MA0139.1). We identified all the CTCF motifs in the human genome using FIMO (Grant et al., 2011). We identified long-range interactions overlapping with these motifs to define convergent, tandem, and divergent motifs.

## CUT&Tag data processing

CUT&Tag data were processed through the ENCODE uniform pipelines(https://github.com/ENCODE-DCC/chip-seq-pipeline2).

## Data visualization

Loop candidates predicted based on Hi-Tag data were visualized using the CoolBox (Xu et al., 2021) package. The enrichment level of histones in the loop anchor region was evaluated using the "computeMatrix" command in Deeptools (Ramírez et al., 2016) with the option "scale-regions" and distances of 5,000 bp upstream and downstream. The "plotProfile" command in Deeptools was used to create a metaplot. Python was employed for plotting density curves, box plots, and scatter plots and for conducting statistical analyses including correlation analysis and tests for statistical significance.

## Code availability

We have made all essential codes for data analysis and figures for reproducible research readily accessible on GitHub: https://github.com/wbszhu/HiTag_paper. In addition, a pipeline is available for processing Hi-Tag data (https://github.com/wbszhu/HiTag).

### Compliance and ethics

The author(s) declare that they have no conflict of interest.


### Acknowledgement

This work was supported by the National Natural Science Foundation of China (32221005), the Earmarked Fund for CARS (CARS-35), the National Natural Science Foundation of China Outstanding Youth (32125035) and Major Project of Hubei Hongshan Laboratory (2021hszd003). We thank laboratory members for helpful discussions and critical reading of the paper. We thank the public technology service platform of Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction (Ministry of Education) and we would be grateful to Jing Xu and Shuang Xiao for their support of sample preparation, data acquisition and analysis. The computations in this paper were run on the bioinformatics computing platform of the National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University.


### Supporting information

The supporting information is available online at https://10.1007/s11427-023-2441-0. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

### References


Abdennur, N., Fudenberg, G., Flyamer, I. M., Galitsyna, A.A., Goloborodko, A., Imakaev, M., and Venev, S.V. (2023). Pairtools: from sequencing data to chromosome contacts. bioRxiv: bioRxiv, 528389.

Bhattacharyya, S., Chandra, V., Vijayanand, P., and Ay, F. (2019). Identification of significant chromatin contacts from HiChIP data by FitHiChIP. Nat Commun 10, 4221.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA 107, 21931–21936.

de Souza, N. (2012). The ENCODE project. Nat Methods 9, 1046.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science 295, 1306–1311.

Dostie, J., and Dekker, J. (2007). Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc 2, 988–1002.

Flyamer, I.M., Illingworth, R.S. and Bickmore, W.A. (2020). Coolpup.py: versatile

pile-up analysis of Hi-C data. Bioinformatics 36, 2980–2985.
Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 48, D87–D92.
Gaspar, J.M. (2018). Improved peak-calling with MACS2. bioRxiv, 496521.
Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018.
Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 10, 1930.
Krijger, P.H.L., and de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. Nat Rev Mol Cell Biol 17, 771–782.
Krueger, F. (2012). Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. Available from URL: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.(Date of access: 28/04/2016).
Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 1303.3997.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Li, X., Luo, O.J., Wang, P., Zheng, M., Wang, D., Piecuch, E., Zhu, J.J., Tian, S.Z., Tang, Z., Li, G., et al. (2017). Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. Nat Protoc 12, 899–915.
Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, 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.
Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods 13, 919–922.
Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.
Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44, W160–W165.
Sati, S., Jones, P., Kim, H. S., Zhou, L.A., Rapp-Reyes, E., and Leung, T.H. (2022). HiCuT: an efficient and low input method to identify protein-directed chromatin interactions. PLoS Genet 18, e1010121.
Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B., and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38, 1348–1354.
Wang, X., Xu, M., Zhao, G.N., Liu, G.Y., Hao, D.L., Lv, X., and Liu, D.P. (2015). Exploring CTCF and cohesin related chromatin architecture at *HOXA* gene cluster in primary human fibroblasts. Sci China Life Sci 58, 860–866.
Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramírez, F., and Grüning, B.A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Res 46, W11–W16.
Xu, W., Zhong, Q., Lin, D., Zuo, Y., Dai, J., Li, G., and Cao, G. (2021). CoolBox: a flexible toolkit for visual analysis of genomics data. BMC Bioinform 22, 489.
Yu, M., Juric, I., Abnousi, A., Hu, M. and Ren, B. (2021). Proximity ligation-assisted ChIP-Seq (PLAC-Seq). Methods Mol Bol 2351, 181–199.
Zhu, Y., Zhou, Z., Huang, T., Zhang, Z., Li, W., Ling, Z., Jiang, T., Yang, J., Yang, S., Xiao, Y., et al. (2022). Mapping and analysis of a spatiotemporal H3K27ac and gene expression spectrum in pigs. Sci China Life Sci 65, 1517–1534.