# Long-read ChIA-PET for base-pair resolution mapping of haplotype-specific chromatin interactions

**Xingwang Li**[1,2,4], **Oscar Junhong Luo**[1,4,5], **Ping Wang**[1], **Meizhen Zheng**[1], **Danjuan Wang**[1], **Emaly Piecuch**[1,3], **Jacqueline Jufen Zhu**[1,3], **Simon Zhongyuan Tian**[1], **Zhonghui Tang**[1], **Guoliang Li**[2], and **Yijun Ruan**[1,2,3]

[1]The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT 06030, USA

[2]National key laboratory of crop genetic improvement, College of Life Sciences & Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China

[3]Department of Genetics and Genome Sciences, University of Connecticut Health Center, 400 Farmington Avenue, Farmington, CT 06030, USA

## Summary

Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET) is a robust method to capture genome-wide chromatin interactions. Unlike other 3C-based methods, it includes a chromatin immunoprecipitation (ChIP) step that enriches for interactions mediated by specific target proteins. This unique feature allows ChIA-PET to provide the functional specificity and higher resolution for detecting chromatin interactions, whereas 3C/Hi-C approaches could not achieve. The original ChIA-PET protocol generates short paired-end tags (2×20 bp) to detect two genomic loci that are far apart on linear chromosomes but are in spatial proximity in the folded genome. We have improved the original approach by developing long-read ChIA-PET, in which the length of the paired-end-tags is increased (up to 2×250 bp). The longer PET reads not only improve the tag mapping efficiency but also increase the probability of covering phased SNPs, which allows haplotype-specific chromatin interactions identification. Here, we provide the detailed protocol for long-read ChIA-PET that includes cell fixation and lysis, chromatin

fragmentation by sonication, ChIP, proximity ligation with bridge linker, Tn5 tagmentation, PCR amplification, and high-throughput sequencing. To a well-trained molecular biologist, it typically takes six days from cell harvesting to the completion of library construction, up to a further 36 hours for DNA sequencing, and less than 20 hours for processing raw sequencing reads.

## Proposed Ontology terms

Biological sciences / Molecular biology / Chromatin / Chromatin structure; Biological sciences / Biological techniques / Gene expression analysis / Chromosome conformation capture-based methods; Biological sciences / Genetics / Sequencing / Next-generation sequencing; Biological sciences / Molecular biology / Nuclear organization; Biological sciences / Genetics / Haplotypes

## Keywords

long-range chromatin interaction; 3D genome; ChIA-PET; haplotype-specific chromatin interaction; proximity ligation; chromosome conformation capture; chromatin structure

## INTRODUCTION

Long-range chromatin interactions are known to be mediated by specific protein factors that tether linear genomic elements into higher-order chromatin structure. These interactions play important roles in spatial and temporal transcriptional regulation of genes, such as bringing distal enhancers to their target gene promoters[1,2].

Various methods for studying chromatin interactions have been developed based on the strategy of nuclear proximity ligation[3], including 3C[4], 4C[5], 5C[6], Hi-C[7] and ChIA-PET[8]. A recent review by Sati and Cavalli (2016) summarizes the pros and cons of these methods[9]. Among them, ChIA-PET is unique because it uses ChIP to enrich long-range chromatin interactions mediated by specific protein factors. This strategy not only effectively enriches specific chromatin contacts, but also provides functional specificity to the detected chromatin interactions. We and others have demonstrated that ChIA-PET is robust in analyzing genome-wide chromatin interactions mediated by: specific transcription factors (TFs) including ERα (estrogen receptor alpha)[8], RNAPII (RNA polymerase II)[10–12], and CTCF (CCCTC-binding factor)[13,14]; components of the cohesin complex[15,16]; and histone proteins[17,18]. In addition, it is known that allelic variants between homologous chromosomes affect inheritance characteristics, chromosome configuration and gene expression[19,20]. The improved ChIA-PET method with longer tag read length presented in this protocol is capable of inferring haplotype-specific chromatin interactions if phased SNP information is available for the assayed sample. Thus, long read ChIA-PET represents a new capability for studying how haplotype variants alters chromatin 3D structure and the possible linkage to human health and disease susceptibility.

### Development of long-read ChIA-PET

The original ChIA-PET protocol uses a type II restriction enzyme (Mme I) to extract 2×20 bp tags from the nuclear proximity ligation products for sequencing and mapping analysis[21]. Although these short tags are sufficient to detect chromatin contacts in most sequence-

unique regions of the genome, we sought to increase the length of the paired-end-tags to improve mapping accuracy and overall efficiency.

Tn5 transposase-based tagmentation is an efficient approach to simultaneously randomly fragment and add adaptors to DNA and has been widely applied in library construction for DNA sequencing[22]. We therefore used Tn5 transposase to randomly fragment the nuclear proximity ligation products and simultaneously add adapters to the cleavage sites; combining these two processes in a single reaction improved the overall efficiency in ChIA-PET library construction (Table 1). The resulting DNA fragments range in size from tens to hundreds of base pairs (average: 300 bp) and subsequent paired-end (2×150 or even 2×250 bp) DNA sequencing analysis produces long ChIA-PET tags of up to 150 or 250 bp. This longer tag length improves sequence alignment accuracy, and more importantly, increases the coverage of heterozygous phased SNPs, thus enabling haplotype-specific mapping of chromatin interactions (Table 1). Besides, the use of a single "bridge"-linker instead of the two half-linkers in the original protocol critically reduces the number of ligation reactions, thus improving the overall efficiency of the protocol. Using this long-read ChIA-PET protocol, we have generated high-resolution chromatin interaction maps for RNAPII and CTCF in a number of human cell lines including GM12878, and based on the phased SNPs information of this cell line, we were able to identify haplotype-specific chromatin interactions[14].

### Overview of the procedure

The main experimental steps of long-read ChIA-PET library preparation are presented in Figures 1 and 2, and the subsequent computational analysis pipeline is outlined in Figure 3. Harvested cells, for example B-lymphocytes, are dual-cross-linked with formaldehyde and Ethylene Glycol Bis[succinimidylsuccinate] (EGS), before nuclei are released by cell lysis. The nuclei are sonicated to generate fragmented chromatin, which is then enriched for target protein-containing chromatin by immunoprecipitation (IP) with target-specific antibodies. This step produces protein-DNA complexes in which the spatial relationships of DNA elements tethered together by the target protein factors are preserved. The chromatin complexes immobilized on the antibody beads are then subjected to DNA end repair and A-tailing, which facilitates subsequent on-bead proximity ligation between DNA fragments within the same chromatin complex via a bridge-linker; the bridge-linker is a short double stranded DNA containing a single biotinylated nucleotide in the middle and T overhangs on each end. After reverse cross-linking and release of the ligation products, Tn5 transposase is used to simultaneously fragment the DNA and add sequencing adaptors. Streptavidin beads are then used to select only the DNA fragments containing ligations junctions via the biotin-labeled bridge-linker. These fragments are subjected to PCR amplification with minimal cycles, followed by size-selection in readiness for high-throughput pair-end sequencing analysis using Illumina instruments (MiSeq, NextSeq, HiSeq, etc.). Generally, ChIA-PET library preparation takes six days, with five quality control checkpoints during the process (Figure 2b). Using 2×150 cycles kits on Illumina MiSeq or NextSeq 500, it will take 1–2 days to produce the sequencing data from a ChIA-PET library. The ChIA-PET sequencing reads are processed using a custom pipeline, ChIA-PET Tool v2 (manuscript in preparation) (Figure 3a). First, the raw sequencing reads are scanned for the bridge-linker, which is a

definitive indication that the PET reads were resulted from genuine proximity ligation products as expected, whereas the PET reads without an identifiable bridge-linker were mostly (99%) derived from short DNA templates (a few hundred bp) as genomic background. Therefore, only bridge-linker containing PET reads were used for downstream analysis. Then the bridge-linker is trimmed off and the remaining PET sequences flanking the bridge-linker are mapped to the reference genome. Only the PETs with both of the paired tags uniquely mapped to the reference genome were used for further analysis. If both ends of a PET overlapped with the both ends of another PET, then these PETs are grouped into a cluster. The clustering procedure is done recursively, and the unclustered PETs are considered as singleton PETs (Figure 3b). Haplotype-resolved chromatin interactions can be deduced from PETs that map to a specific allele identified by phased SNPs (Figure 3c and 3d).

## Advantages over Hi-C for mapping long-range chromatin interactions

Between the two main approaches, ChIA-PET and Hi-C, for mapping long-range chromatin interactions, ChIA-PET possesses several advantages. Although Hi-C is generally the method of choice for capturing all chromatin interactions genome-wide without any selection[23], in practice, typical Hi-C data usually is only suited for mapping topological domains because the resolution of the anchor points of the detected interactions is in tens of hundreds of kilo-base-pairs[7,24]. This low resolution is certainly not sufficient to identify any specific regulatory elements such as promoters and enhancers (typically in sizes of a few hundreds to thousands base-pairs) that are involved in interactions. Although ultra-deep sequencing of a large number of Hi-C libraries from the same biological sample could achieve kilo-base-pair resolution of interaction anchors[25], the cost of library construction and sequencing is prohibitive. In contrast, in each ChIA-PET experiment, genome-wide chromatin interactions are enriched and identified by association with a specific protein that mediated the interactions (Table 2). Depending on the target protein, typical ChIA-PET data can map long-range chromatin interactions between specific protein binding peaks in high-resolution (which are typically <100 bp at peak summit) and specific regulatory elements; for instance, enhancer to promoter interactions have been identified by RNAPII and TF ChIA-PET experiments[8,10], and interactions occurring at domain boundaries and sub-domain structures have been detected by CTCF and cohesin ChIA-PET[14,16], respectively. Our recent analysis further demonstrated that, in fact, ChIA-PET experiments also produce Hi-C-like data[14]. This is because, like in all ChIP experiments, the ChIP enriched DNA fragments are usually at its best around 20% of the overall genome-wide background, and this is a significant signal-to-noise ratio for the identification of thousands to tens of thousands enrichment peak loci bound by specific protein factors. Similarly, in ChIA-PET experiments, approximately 20% of the PET reads fall into the PET clusters, which corresponds to the specific chromatin interactions mediated by the ChIP-enriched target protein, and the rest (~80%) were of non-enriched singleton PET data that do not overlap with other PETs (Figure 3b). Interestingly, the non-enriched singleton ChIA-PET data showed same 2D contact patterns as the Hi-C data (Figure 3c), reflecting higher-order chromatin contacts in topological neighborhood. Therefore, ChIA-PET produces unique and inclusive data for detecting specific chromatin interactions mediated by protein factors

(Figure 3c) in high resolution, and for assessing the approximation of chromosomal folding similar to Hi-C data.

Furthermore, the extended PET read length of long-read ChIA-PET enables single nucleotide resolution by identifying heterozygous SNPs in chromatin interactions anchors, allowing haplotype-specific interactions to be identified (Figure 3c and 3d). This ability allows genetic variants to be linked to haplotype specific topology and allelic transcription regulation, and ultimately connects to physiological traits and diseases[14].

### Limitations

The current version of long read ChIA-PET has several technical limitations. First, a successful ChIA-PET experiment is dependent on the availability of a high quality antibody against the protein of interest; poor quality antibodies are often the cause of unsuccessful ChIA-PET experiment. Second, the current ChIA-PET protocol requires ~100 million cells as input for each ChIA-PET experiment to guarantee high quality data; using fewer cells often results in lower library complexity and high redundancy of sequencing reads. This limitation constrains the application of ChIA-PET to biological samples that can generate large numbers of cells. Further optimization of the protocol - including the use of new or improved enzymes with better efficiency and miniaturized operational devices for biochemical reactions - could help to reduce the required input cell numbers.

### Experimental design

**Fixing the cells—**The cells are double fixed with cross-linking reagents formaldehyde and EGS (Ethylene glycol bis[succinimidylsuccinate]). Formaldehyde is known for creating short-armed covalent bonds between DNA and protein molecules[26], whereas EGS forms relatively weak long-arm bonds between proteins[27]. Therefore, this double cross-linking strategy could maximize the preservation of chromatin interactions that not only involve protein-DNA but also protein-protein contacts. In our hands, the inclusion of EGS indeed substantially increased (5–10 fold more) the detection of genuine long-range chromatin interactions. Other long-arm fixative reagents, like DSG[28], might also work as cross-linkers. Optimization of cross-linking conditions (including duration of fixation and concentration of fixative reagents) is critical and has to be empirically tested since excessive fixation may raise potential bias for transient interactions[29].

**Starting materials—**Various types of immortal cell lines from human and mouse have been successfully used for ChIA-PET experiments[8,10,11,14]. Routinely, the best ChIA-PET results are obtained from at least $10^8$ cells as input. Fewer cells can be used if samples are limited, but usually the resulting library will be of lower complexity and the sequencing redundancy will be high, similar to what has been observed for Hi-C experiments using fewer cells[24]. A number of primary cells had also been applied for ChIA-PET assays. For example, different types of blood cells (e.g., B cells and T cells) isolated from whole blood samples of individual human subjects for ChIA-PET experiments (unpublished data). However, additional precaution and preparation should be taken for primary cells of solid tissues. For some solid tissues, well-separated cell suspension such as primary B-cells from

mouse spleen[12] and hepatocytes from mouse liver (unpublished data) can be prepared for high quality ChIA-PET experiments, whereas other solid tissues are much more difficult.

**Cell lysis, nuclear lysis and sonication**—Cell lysis and nuclear lysis are critical parts of the protocol. We typically perform cell and nuclear lysis at 4 °C. After each lysis step, cellular and nuclear morphology are checked under the microscope to make sure most of the cytoplasm is degraded prior to sonication. Cell status after cell lysis varies from one type of cell to another. For cells that are difficult to lyse, a higher temperature and longer incubation time can be attempted. Sonication is another critical step that fragments the chromatin to an ideal size range of approximate 1 kb for ChIA-PET experiments. This is in contrast to ChIP-Seq experiments, where short DNA fragments of 100 bp (the shortest size range possible by sonication) are required. Therefore, balanced sonication parameters including strength, duration, volume, and temperature have to be optimized, depending on the specific instrument and cell type used.

**Library preparation and sequencing**—The most important step in the ChIA-PET protocol is the nuclear proximity ligation. This step physically connects two linearly-distant DNA fragments that have been brought into spatial proximity within the complex structure of a long range chromatin interaction, mediated by specific protein factors[30]. It is often beneficial to have specifically designed DNA oligos as a linker to connect the prospective DNA fragments. The nucleotide sequence and modified molecular group such as biotin in the DNA linker can be used for downstream separation for the expected proximity ligation products from undesired genomic DNA fragments and provides unique identifier to distinguish the specific products from genomic background noise after sequencing analysis. In the original ChIA-PET protocol, we used two DNA oligos with two ligation steps for this purpose[8,35]. To reduce enzymatic reaction steps and increase the efficiency of this protocol, in this protocol we use a bridge-linker to connect two chromatin DNA fragments in a single reaction to generate the DNA-linker-DNA ligation products. The bridge-linker is 19 nucleotides long with 3'-protruding T nucleotide at both ends of the linker and a biotin group in the middle, and the chromatin DNA are A-tailed (Figure 1). In the reaction, any two A-tailed chromatin DNA fragments will not be ligated without a linker to bridge them together to form the DNA-linker-DNA ligation products. Therefore, equal molar ratios between of bridge-linker and chromate DNA fragments is critical and needs to be empirically tested; because excessive bridge-linker will saturate the DNA ends of chromatin DNA fragments and thus prevent the balanced ligation products, whereas insufficient bridge-linker will result in poor proximity ligation. In this protocol, we use Tn5 transposase to randomly cut the linker-ligated DNA fragments to produce longer templates and to simultaneously add adapter sequences for subsequent PCR amplification and sequencing. The optimal size range of the DNA templates is from 300 to 800 bp if using the Illumina 2×150 bp sequencing method. Too short of the DNA fragments (< 300bp) would not possess enough information for the tag-linker-tag composition, whereas larger that 800 bp is not ideal for the optimal performance by the Illumina sequencing platform. To generate the desired size range of DNA fragments for the library, it is important to use the right amount of Tn5 transposase reagent per unit of DNA substrate for tagmentation reaction. Too much

of the Tn5 used in a reaction could lead to too frequent cut and thus result too short of the DNA fragments of the library.

**Controls—**A ChIA-PET experiment produces two genome-wide datasets; the protein binding sites (the same as ChIP-Seq data) and the chromatin interaction data. Therefore, a possible negative control experiment is to perform a mock IgG ChIA-PET experiment, similar to a mock IgG ChIP-Seq experiment. After data processing, the binding peaks of target protein from ChIA-PET are compared with the binding peaks identified by ChIP-Seq experiments of same target protein in the same biological sample including the mock IgG controls. The negative control experiments are useful when using a new cell type, but are not necessary for each ChIA-PET experiment of the same biological sample. To assess the quality of each ChIA-PET experiment, we routinely analyze some benchmark loci where reliable long-range chromatin interactions have been previously characterized (e.g., *H19-IGF2* loci for CTCF mediated long-range chromatin interaction) by ChIA-PET or other analysis. In addition, there are also a number of quality control measurements throughout the Procedure (Figure 2).

## MATERIALS

### REAGENTS

- 5× T4 DNA Ligase Buffer (Life Technologies, cat. no. 46300-018)

- 20× SSC (Ambion, cat. no. 1557044)

- Absolute Ethanol (500 ml, Sigma, cat.no. E7023)

- Agilent DNA High Sensitivity Kit (Agilent Technologies, cat.no. 5067-4626)

- Ampure XP beads (60 ml, Beckman, cat. no. A63881)

- Antibody against protein of interest, e.g. Monoclonal Antibody against RNA Polymerase II (8WG16), 0.5 ml (Covance, cat. no. MMS-126R) and Polyclonal Anti-CTCF Antibody (Abcam, cat. no. ab70303)

- Buffer EB (250 ml, Qiagen, cat. no. 19086)

- Buffer TE (pH8.0, Ambion, cat. no. AM9849)

- dATP Solution (100 mM) (NEB, cat. no. N0440S)

- Dimethyl Sulfoxide (DMSO, Sigma, cat. no. D2650-100ML)

- DNA Clean & Concentrator-5-Capped Columns (200) (Zymo Research, cat. no. D4014)

- DNA ladder (25 bp, Invitrogen, cat. no. 10597011)

- DNA LoBind Tubes (1.5 ml, Eppendorf, cat. no. 022431021)

- dNTPs (10 mM, Life Technologies, cat. no. 18427-088)

- DPBS, calcium- and magnesium-free (Gibco, cat. no. 14190-250)

- Dynabeads M-280 Streptavidin (Invitrogen, cat. no.11205D)

- Dynabeads Protein G for Immunoprecipitation (50 mL) (Life Technologies, cat. no. 10009D)

- EDTA (pH 8.0, 0.5 M, 500 ml, Ambion, cat. no. AM9261)

- EGS (Ethylene glycol bis[succinimidylsuccinate], Thermo Fisher Scientific, cat. no. 21565)

- Formaldehyde (37% [vol/vol], EMD MILLIPORE, cat. no. 344198-250ML) CAUTION: Formaldehyde is toxic; always operate it in a fume hood.

- Glycine (Sigma, cat. no. G8898-500G)

- GlycoBlue (Life Technologies, cat. no. AM9516)

- HEPES buffer (1 M, pH7.3, Fisher Scientific, cat. no. BP299-1)

- I-Block Protein-Based Blocking Reagent (Thermo Fisher, cat. no. T2015)

- IGEPAL CA-630 (Sigma, cat. no. I8896)

- Isopropanol (Sigma, cat. no. I-9516-500ml)

- Klenow Fragment (3'-->5' exo-) (NEB, cat. no. M0212L)

- Maxtract High Density (2 ml, QIAGEN, cat. no. 129056)

- NaCl Solution (500 ml, 5.0 M, Ambion, cat. no. AM9759)

- NEB buffer 2 (NEB, cat. no. B7002S)

- Nextera DNA Sample Prep Kit (Illumina, cat. no. FC-121-1031)

- Novex precast TBE Gels (4–20% [wt/vol], Invitrogen, cat. no. EC6225BOX)

- Nuclease-free Water (50 ml, Ambion, cat. no. AM9937)

- Phenol: Chloroform: IAA (Ambion, cat. no. AM9730) CAUTION Phenol and Chloroform are toxic; operate them in a fume hood.

- Proteinase Inhibitors (Roche, cat. no. 04693116001)

- Proteinase K Solution (Life Technologies, cat. no. AM2548)

- Sodium Acetate (Ambion, cat. no. AM9740)

- Sodium dodecyl sulfate (SDS, 10% [wt/vol] Ambion, cat. no. AM9822)

- Sodium deoxycholate (Sigma, cat. no. 30970-100G)

- SYBR Green I (Roche, cat. no. 04707516001)

- T4 DNA ligase (Thermal, cat. no. EL0013)

- T4 DNA Polymerase (Promega, cat. no. M4215)

- Tris-HCl, pH 7.0 (1 M, 500ml, Ambion, cat. no. AM9856)

- Triton-X-100, Molecular Biology Grade (Promega, cat. no. H5141)

- TWEEN-20 for molecular biology, viscous liquid (Sigma, cat. no. P9416-100ML)

- Cell sample of choice. For example, we have successfully used GM12878 cell line (Coriell Institute, cat. no. GM12878), HeLa S3 cell line (ATCC, cat. no. CCL-2.2), K562 cell line (ATCC, cat. no. CCL-243) and MCF7 cell line (ATCC, cat. no. HTB-22)

- Bridge linker oligos for proximity ligation: Bridge linker-F: 5'- /5Phos/ CGCGATATC/iBIOdT/TATCTGACT -3' and Bridge linker-R: 5'- /5Phos/ GTCAGATAAGATATCGCGT -3'. HPLC purified (250nmole) from IDT (Integrated DNA Technologies). See Box 1 for use as 10 μl aliquots (200 ng/μl) for preparation.

## EQUIPMENT

- Agilent 2100 bioanalyzer instrument (Agilent techcologies, cat. no. G2940CA)

- Bio-Rad C1000 Thermal Cycler (Bio-rad, cat. no. 185-1148EDU)

- Biological safety cabinet classII, type A2 (Nuair, model no. NU-S480-600)

- Blue pippin instrument (Sage science, Blue Pippin)

- Centrifuge (Eppendorf 5810R, cat. no. 22628180)

- DynaMag™-2 Magnet (Life Technologies, cat. no. 12321D)

- Dark Reader Transilluminator (Clare Chemical Research, Inc. Model: DR89X)

- Eppendorf 5424 Ventilated Microcentrifuge (Eppendorf, cat. no. 022620401)

- Eppendorf Thermomixer (Eppendorf, cat. no. 5382000023)

- Freezer (Fisher Scientific, cat. no. 3752-DB)

- Illumina MiSeq and HiSeq2500 or NextSeq500 high-throughput sequencing machine (Illumina)

- Incubators (Heratherm, cat. no. 50125590)

- Intelli-mixer RM-2L (Rose scientific, cat. no. MX1000)

- Light cycle 480 PCR instrument (Roche, cat. no. 05015243001)

- MicroVac Concentrator (Tomy, cat. no. MV-100)

- Nikon TS100-F Microscope (Nikon, cat. no. TS100-F)

- Qubit 2.0 Fluorometer Specifications (Invitrogen, cat. no. Q32866)

- Refrigerator (Fisher Scientific, cat. no. 13-994-201)

- Sonicator (SONICS & MATERIALS, cat. no. VCX130)

- Vortex-Genie® Mixers (VWR, cat. no. 58816-121)

- Water bath (Fisher Scientific, cat. no. 15-462-20Q)

## COMPUTATIONAL PROGRAMS AND SOURCE CODES USED IN CHIA-PET TOOL VERSION 2

- Cutadapt package (version 1.6) (https://cutadapt.readthedocs.io/en/stable/)

- BWA software package (version 0.7.12-r1039 or above) (http://bio-bwa.sourceforge.net)

- SAMtools (version 0.1.19 and above) (http://samtools.sourceforge.net)

- Picard tools (version 1.107 and above) (https://broadinstitute.github.io/picard/)

- BEDTools (version 2.17.0 and above) (http://bedtools.readthedocs.io/en/latest/)

- pysam python module (version 0.8.1) (http://pysam.readthedocs.io/en/latest/api.html)

- Perl scripts prepare_reads_single.pl (Supplementary Software 1) and prepare_reads_pair.pl (Supplementary Software 2) for bridge-linker trimming.

- Perl script clustering_bridge.pl for PET cluster identification.

- Python script fetch_base_snp_v081.py for assigning PET haplotype.

### REAGENT SETUP

10% (wt/vol) Sodium deoxycholate (100 ml): Add 10 g of Sodium deoxycholate to 80 ml of water and mix it well until the solution is clear; finalize the solution volume to 100 ml. The buffer can be stored at room temperature (20 to 22 °C) for up six months.

2.5M Glycine solution (100 ml): Add 27.89 g Glycine to 80 ml of water and mix it well until the solution is clear; finalize the solution volume to 100 ml. The buffer can be stored at 4°C for up six months.

0.1%(wt/vol) SDS FA Cell Lysis buffer (500 ml): Add 444 ml of water to a 500 ml beaker first, and then add sequentially 25 ml HEPES-KOH (1M, pH7.5), 15 ml NaCl (5M), 1 ml EDTA (0.5M), 5 ml Triton X-100, 5 ml sodium deoxycholate (10%(wt/vol)) and 5 ml SDS (10%(wt/vol)), mix them well. The buffer can be stored at 4°C for up six months. CRITICAL 0.1% (wt/vol) SDS FA Cell Lysis buffer can be stored at 4°C for up two months. However, because protease inhibitors are unstable in solution, dissolve 1 tablet of complete protease inhibitor into 50 ml buffer just before use.

1% (wt/vol) SDS FA Nuclear Lysis buffer (500 ml): Add 200 ml of water to a 500 ml beaker first, and then add sequentially 25 ml HEPES-KOH (1 M, pH7.5), 15 ml NaCl (5 M), 1 ml EDTA (0.5 M), 5 ml Triton X-100, 5 ml sodium deoxycholate (10% [wt/vol]) and 50 ml SDS (10% [wt/vol]), mix them well and finalize the solution volume to 500 ml. The buffer can be stored at room temperature for up six months. CRITICAL SDS might precipitate out of the buffer and care should be taken prior to use. Dissolve 1 tablet of complete protease inhibitor into 50 ml of the buffer immediately before use.

High Salt ChIP buffer (500 ml): Add 324 ml of water to a 500 ml beaker first, and then add sequentially 25 ml HEPES-KOH (1M, pH7.5), 35 ml NaCl (5M), 1 ml EDTA (0.5M), 5 ml

Triton X-100, 5 ml sodium deoxycholate (10% [wt/vol]) and 5 ml SDS (10% [wt/vol]), mix them well. The buffer can be stored at 4°C for up two months.

ChIP Wash Buffer (250 ml): Add 220.45 ml of water to a 250 ml beaker first, and then add sequentially 2.5 ml Tris•HCl (1M, pH 8.0), 7.8 ml LiCl (8M), 0.5 ml EDTA (0.5M), 1.25 ml IGEPAL CA-630 and 12.5 ml Sodium deoxycholate (10% [wt/vol]), mix them well. The buffer can be stored at 4°C for up two months.

ChIP Elution Buffer (2 ml): Add 1.66 ml of water to a 2 ml tube first and then add sequentially 100 µl of Tris•HCl (1M, pH 7.5), 40 µl of EDTA (0.5M) and 200 µl of SDS (10%(wt/vol)), using pipette to mix them well. The buffer should be prepared freshly before use.

ChIA-PET Wash Buffer (1000ml): Add 888 ml of water to a 1 L beaker first, and then add sequentially 10 ml Tris-Cl (1M, pH7.5), 2 ml EDTA (0.5M), 100 ml NaCl (5M) and mix them well. The buffer can be stored at 4°C for up two months.

TNE Buffer (50ml): Add 48.99 ml of water to a 50 ml Falcon tube first, and then add sequentially 1 ml Tris-HCl (1M, pH8.0), 0.5 ml NaCl (5M), 10 µl of EDTA (0.5 M) and mix them well. The buffer should be prepared freshly before use.

1× PBST buffer (100 ml): Add 89.9 ml of water to a 100 ml beaker first, and then add sequentially 10 ml PBS (10×), 100 µl Tween 20 and mix them well. The buffer can be stored at 4°C for up two months.

2× Binding & Washing Buffer (2× B&W)(500ml): Add 294 ml of water to a 500 ml beaker first, and then add sequentially 5ml Tris-HCl (1M, pH8.0), 1 ml EDTA (0.5 M), 200 ml NaCl (5M) and mix them well. The buffer can be stored at 4°C for up six months.

1× Binding & Washing Buffer (1× B&W)(500ml): Add 394 ml of water to a 500 ml beaker first, and then add sequentially 5ml Tris-HCl (1M, pH8.0), 1 ml EDTA (0.5 M), 100 ml NaCl (5M) and mix them well. The buffer can be stored at 4°C for up six months.

2× SSC/0.5%(wt/vol) SDS (100 ml): Add 85 ml of water to a 100 ml beaker first, and then add sequentially 10 ml 20xSSC, 5ml SDS (10%(wt/vol)) and mix them well. The buffer can be stored at room temperature for up three months. CRITICAL: lower room temperature (below 18°C) may cause SDS to precipitate – if this occurs, place the buffer in a 25 °C water bath to solubilize SDS before use.

iBlock buffer (100 ml): Take 2 g I-Block Protein-Based Blocking Reagent into 95 ml water at 65°C water bath until all the protein was dissolved, then add 5 ml 10%(wt/vol) SDS to the solution. The buffer can be stored at room temperature for up six months.

## PROCEDURE

### Cell Harvesting and Dual-Cross-linking TIMING 4–5 hours

**1** Collect cells, cultured to ~80% confluency under recommended conditions, by centrifugation at 200 ×g for 5 minutes at room temperature (18–22°C). Discard

the medium and wash the cell pellet (~$10^8$ cells) with 10 mL 1× PBS buffer once. Centrifuge at 200 ×g for 5 minutes at room temperature and discard 1× PBS buffer.

**2**    Prepare EGS fixative solution in a fume hood by dissolving 0.02g EGS (Ethylene glycol bis[succinimidylsuccinate]) in 200ul DMSO at 37 °C for 5 minutes, then mix EGS/DMSO with 29.8ml 1× PBS to a final concentration 1.5mM, and then place EGS mixture at 37 °C water bath until use.

**3**    Resuspend cell pellet in 30 ml freshly prepared EGS fixative solution and mix the cell with rotation for 45 minutes at room temperature.

**4**    Add 833ul 37% (vol/vol) formaldehyde to the fixative solution to a final concentration of 1% (vol/vol) and mix with rotation for 20 minutes at room temperature.

**5**    Quench cross-linking reaction by adding 2.68ml 2.5M glycine to the fixative solution to a final concentration of 0.2M with rotation for 5min at room temperature. Spin at 400 ×g for 5 minutes at room temperature and discard supernatant into an appropriate waste container.

**6**    Wash the cell pellet with 20ml chilled 1× PBS buffer. Spin at 400 ×g for 5 minutes at 4°C and discard 1× PBS buffer.

   PAUSE POINT The cell pellets can be frozen at −80°C for up six months.

**Chromatin preparation and immunoprecipitation (ChIP):** *Preparing antibody-coated magnetic beads* TIMING 7 hours

**7**    Transfer 600ul resuspended protein G magnetic beads to 1.5ml Eppendorf tube and place the tube on magnetic rack for 30 seconds, discard the supernatant and wash the bead with 1× PBST buffer twice. Resuspend the beads in 900ul of 1× PBST buffer.

**8**    Take 60 ug antibody of choice and mix with washed beads with rotation at 4 °C for 6 hours. NOTE: Step 9–16 can be performed during this incubation.

**ChIP:** *Cell lysis and nuclear lysis* TIMING 5–6 h

**9**    Resuspend cell pellet (from step 6) in 30ml of 0.1% (wt/vol) SDS FA cell lysis buffer containing proteinase inhibitor.

**10**    Incubate the cells at 4 °C (cold room) for 15minutes with rotation. Spin at 400×g for 5 minutes at 4°C and discard cell lysis buffer.

**11**    Repeat steps 9–10 twice for a total of three cell lysis steps.

**12**    Resuspend cell pellet in 30ml of 1% (wt/vol) SDS FA Nuclear Lysis buffer with proteinase inhibitor.

**13**    Incubate the cells at 4°C (cold room) for 15minutes with rotation. Spin at 2800×g for 5 minutes at 4°C and discard nuclear lysis buffer.

**14**    Repeat steps 12–13 once for a total two nuclear lysis steps.

?Troubleshooting

**15** Repeat steps 9–10 twice more for another two rounds of cell lysis. NOTE: check cell and nuclear morphology after each lysis step to ensure the cell and nucleus are completely lysed.

**16** Resuspend chromatin pellet in 4ml of 0.1% (wt/vol) SDS FA cell lysis buffer containing proteinase inhibitor. Aliquot 1 ml chromatin solution into four 15 ml round-bottom tubes. Remove bubbles by pipet and place the chromatin solution on ice.

PAUSE POINT the chromatin solution can be stored at −80 °C for up to one month.

**ChIP:** *Sonication, pre-clearing and immunoprecipitation* TIMING 2 days

**17** Shear the chromatin on a sonication system using optimized parameters. NOTE: the sonication parameters and conditions should be optimized prior this step. We use Amplitude: 35%; Time: 6 minutes; 30 seconds ON, 30 seconds OFF.

**18** Spin the sonicated chromatin at 2800 ×g for 10 minutes at 4 °C. Combine 4 tubes of supernatant to a 15 ml Falcon tube. Take 20 μl of chromatin as input control material (non-enriched) for quality check of ChIP (see details in Box 2). It can be stored at −20 °C for later use.

**19** Wash 600 ul fresh magnetic beads with 1 ml 1× PBST buffer twice. Add 4 ml of the collected chromatin solution to the tube to resuspend the beads. Rotate the tube at 4 °C for one hour for pre-clearing of chromatin in order to remove un-specific binding of chromatin materials to magnetic beads.

**20** Place the tube containing beads and chromatin on magnetic rack for 30 seconds. Transfer chromatin supernatant into a fresh 15ml tube.

**21** Place the tube containing beads and antibody from step 8 on magnetic rack for 30 seconds. Discard the supernatant and wash bead with 1× PBST buffer twice. Discard 1× PBST buffer.

**22** Resuspend the antibody-coated beads with pre-cleared chromatin solution from step 20 for immunoprecipitation (IP), and rotate the tube at 4 °C for 8 hours or overnight.

**23** Place the tube on magnetic rack for 30 seconds. Discard supernatant, and resuspend beads with 5 ml 0.1% (wt/vol) SDS FA cell lysis buffer (same as used in step 9), and rotate at 4°C for 5 minutes.

**24** Repeat washing step 23 twice for a total of three washes.

**25** Place the tube on magnetic rack for 30 seconds. Discard supernatant, and resuspend beads with 5 ml High salt ChIP buffer. Rotate the tube at 4°C for 5 minutes.

**26** Repeat step 25 once.

27    Place the tube on magnetic rack for 30 seconds. Discard supernatant, and resuspend beads with 5 ml ChIP wash buffer. Rotate the tube at 4°C for 5 minutes. Discard supernatant.

28    Wash beads with 5 ml 1× TE buffer, Rotate the tube at 4°C for 5 minutes. Discard supernatant.

29    Repeat step 28 once.

30    Resuspend the beads into 1 ml of 1× TE buffer, take 50 μl of beads to check the quality and quantity of ChIP-DNA following the procedures in Box 2.

      PAUSE POINT the remaining beads can be stored at 4°C (NOT −20°C) for up to two weeks.

CRITICAL STEP this is quality control checkpoint QC1 in Figure 2b and the Bio-analyzer trace should have a peak at ~3.5kb

**Long read ChIA-PET library preparation:** *End-repair, A-tailing and proximity ligation* TIMING ~24 hours

31    Resuspend beads from step 30 in 700ul of T4 DNA polymerase master mix:

| Component | Amount (μl) | Final concentration |
|---|---|---|
| 10×Buffer for T4 DNA polymerase | 70 | 1x |
| 10mM dNTPs | 7 | 100 uM |
| T4 DNA polymerase | 7 | 0.2–0.4 U/μl |
| ddH$_2$O | 616 | |
| Total | 700 | |

32    Mix and incubate at 37°C for 40 minutes with rotation on a Intelli-Mixer in a 37°C incubator (parameters: F8, 30 rpm; U=50, u=60).

33    Place the tube on magnetic rack for 30 seconds. Discard the T4 DNA polymerase master mix (carefully without disturbing the beads). Wash the beads with 1ml of ice-cold ChIA-PET Wash Buffer for three times.

34    Place the tube on magnetic rack for 30 seconds. Discard the wash buffer and resuspend beads with 700 μl of Klenow (3'–5'exo-) master mix:

| Component | Amount (μl) | Final concentration |
|---|---|---|
| 10× NEB Buffer 2 | 70 | 1x |
| 10 mM dATP | 7 | 100 μM |
| Klenow (3'–5'exo-) | 7 | 0.2 U/μl |
| ddH$_2$O | 616 | |
| Total | 700 | |

**35** Incubate at 37 °C for 50 minutes with rotation on a Intelli-Mixer (F8, 30 rpm, U=50, u=60) in incubator.

**36** Place the tube on magnetic rack for 30 seconds. Discard the Klenow (3'–5'exo-) master mix. Wash the beads with 1ml of ice-cold ChIA-PET Wash Buffer for three times.

**37** Discard the wash buffer and resuspend beads with 1.4ml of T4 DNA ligase master mix:

| Component | Amount (µl) | Final concentration |
|---|---|---|
| 5× T4 DNA ligase buffer | 280 | 1x |
| Bridge linker (200ng/µl) | 4 | 0.57 ng/µl |
| T4 DNA ligase | 6 | 0.02 U/µl |
| ddH$_2$O | 1110 | |
| Total | 1400 | |

**38** Incubate at 16°C with rotation on a Intelli-Mixer (F8, 30 rpm, U=50, u=60) in incubator overnight.

**Long read ChIA-PET library preparation:** *Reverse cross-linking, DNA purification and tagmentation.* TIMING ~20 hours

**39** Place the tube from step 38 on magnetic rack for 30 seconds. Discard the supernatant and resuspend magnetic beads into 200 µl of ChIP Elution Buffer. Place the tube on the Thermomixer with rotation (900 rpm) at 65 °C for 15 minutes.

**40** Place the tube on the magnetic rack and transfer the 200 µl ChIP Elution Buffer, which contains the protein-chromatin DNA complex, to a new 1.5 ml tube (Do not discard!).

**41** Resuspend magnetic beads in 400 µl of Buffer EB to wash beads. Place the tube on the magnetic rack and take 400 µl supernatant and combine with 200 µl Elution Buffer from step 40.

**42** Add 10 µl Proteinase K to 600 µl of eluted protein-DNA complex solution, mix and incubate at 55 °C overnight for reverse cross-linking.

**43** Prepare a MaXtract High Density tube by centrifuging at 16000×g for 2 minutes at room temperature.

**44** Add equal volume of Phenol-Chloroform-Isoamyl alcohol (pH 7.9) to the tube containing reverse cross-linking reaction solution from step 42, and mix vigorously for 5 seconds, transfer the mixture to the centrifuged MaXtract High Density tube, centrifuge at 16000×g, at room temperature for 5 minutes.

**45** Transfer upper aqueous phase above the gel matrix to a fresh 1.5 ml tube. Precipitate the DNA by adding following reagents:

| Component | Amount (µl) | Final concentration |
|-----------|-------------|---------------------|
| 3M Sodium Acetate | 60 | 277 mM |
| GlycoBlue | 2 | |
| Isopropanol | 650 | |

**46**   Invert the tube to mix solution. Incubate at −80 °C for 30 minutes.

**47**   Spin the DNA pellet at 16000 ×g at 4 °C for 20 minutes. Wash pellet with 1 ml of ice-cold 75% (vol/vol) ice-cold ethanol twice. Remove all the ethanol and dry pellet using Micro vacuum concentrator for 2 minutes at room temperature.

**48**   Resuspend the pellet in 20 µl Resuspension Buffer (Illumina) and make sure the DNA is dissolved into buffer completely at 4 °C. Proceed to DNA quantitation on a Qubit 2.0 fluorometer.

PAUSE POINT Proximity ligation DNA can be stored at −20 °C for up to several months.

CRITICAL STEP this is quality control checkpoint QC2 in Figure 2b and the Bio-analyzer trace should have a peak at ~5 kb.

?Troubleshooting

**49**   Cut the proximity ligated DNA (from step 48) and add the adaptor by using Nextera Tn5 transposome as tabulated below. Generally, two to six tagmentation reactions (100–300ng proximity ligated DNA) are recommended for each Longread ChIA-PET library preparation.

| Component | Amount (µl) | Final |
|-----------|-------------|-------|
| Proximity ligation DNA | x | 50 ng |
| 2×Tagmentation buffer | 25 | 1x |
| Tagmentation Enzyme | 8 | |
| ddH$_2$O | 17-x | |
| Total | 50 | |

CRITICAL STEP The ratio between Tagmentation Enzyme and proximity ligation DNA is critical for optimal reactions and this ratio need to be empirically determined.

**50**   Gently pipette up and down 6 times to mix DNA and master solution. Incubate at 55 °C for 5 minutes on a PCR machine.

**51**   Purify the tagmented DNA using Zymo Genomic DNA Clean & Concentrator™ kit, add 350 µl of Zymo DNA Binding Buffer to each tagmentation reaction and mix thoroughly.

**52**   Transfer mixture to Zymo-Spin™ IC-XL column placed in a collection tube. Centrifuge at 16000 g for 30 s at room temperature. Discard the flow-through.

**53** Add 200 Zymo DNA Wash Buffer to the column. Centrifuge for 16000g for 30 s at room temperature. Discard the flow-through. Repeat this step once.

**54** Centrifuge the empty column at 16000 ×g for 1 minute at room temperature with lid open to ensure any residual ethanol is removed.

**55** Transfer the column to a new 1.5 ml LoBind tube and add 15 μl Resuspension Buffer to membrane of the column. Incubate at room temperature for 1 minute.

**56** Spin the column in a 1.5 ml LoBind tube at 16000 ×g for 1 minute at room temperature to elute the DNA. Take 1 μl of sample to assay tagmentation size distribution on a High sensitive DNA Bioanalyzer Lab Chip.

?Troubleshooting

CRITICAL STEP. This is quality control QC3 in Figure 2b and the majority of DNA fragments should fall in 200 bp – 1 kb range.

**Long read ChIA-PET library preparation:** *Immobilization on beads and PCR amplification*. TIMING 6–7 hours

**57** Mix the M280 Streptavidin dynabeads suspension well, and transfer 25 μl dynabeads suspension for each sample to a 1.5 ml LoBind tube.

**58** Place the tube on the magnetic rack, discard the buffer saline and wash the dynabeads with 150 μl of 2×Binding & Washing buffer (2×B&W) twice.

**59** Resuspend the beads in 100 μl iBlock buffer, mix and incubate at room temperature for 45 minutes with rotation on the Intelli-Mixer (UU, 50 rpm, U=50, u=60).

**60** Add 500 ng (in 50 μl water) sheared genomic DNA (fragment size range 300 to 500 bp) into 50 μl of 2×B&W buffer in a new tube. The sheared genomic DNA is for non-specific blocking the surface of the M280 Streptavidin Dynabeads and can be prepared by sonication of mammalian genomic DNA.

**61** Place the tube containing iBlock buffer and M280 beads on the magnetic rack and Discard the iBlock buffer and wash the M280 beads with 200 μl 1×B&W buffer twice.

**62** Resuspend the M280 bead with 100 μl 1×B&W buffer containing blocking DNA (from step 60) and incubate at room temperature for 30 minutes with rotation on a Intelli-Mixer (UU, 50 rpm, U=50, u=60).

**63** Discard the 1×B&W buffer containing blocking DNA and wash the dynabeads with 200 μl of 1×B&W buffer twice. Place the tube on the magnetic rack and discard the supernatant.

**64** Transfer 50 μl DNA library (from step 56 add resuspension buffer to 50 ul if original volume is not 50ul) and equal volume of 2×B&W buffer to the tube containing blocked M280 beads (from step 63). Mix and incubate at room temperature for 45 minutes with rotation on a Intelli-Mixer (UU, 50 rpm, U=50, u=60).

**65** Place the tube on the magnetic rack and discard supernatant, wash the M280 beads with 500 μl of 2×SSC/0.5% (wt/vol) SDS five times.

**66** Place the tube on the magnetic rack and discard supernatant, wash the M280 beads with 1×B&W buffer twice. Resuspend the beads with 30μl ddH2O.

**67** Using 10 μl M280 beads as template for each PCR amplification (total three reactions) of library as follow:

| Component | Amount (μl) |
| --- | --- |
| Beads (from step 66) | 10 |
| NPM mix | 15 |
| PPC PCR primer | 5 |
| Index 1 primer (i7) | 5 |
| Index 1 primer (i5) | 5 |
| Total | 50 |

**68** Run the PCR amplification with following program:

| Cycle number | Anneal | Extend | Denature |
| --- | --- | --- | --- |
| 1 | | 72 °C, 3 min | |
| 2 | | | 98 °C, 10 s |
| 3–11 | 63 °C, 30 s | 72 °C, 50 s | 98 °C, 10 s |
| 12 | 63 °C, 30 s | 72 °C, 5 min | |

CRITICAL STEP The PCR amplification cycle is important for final library quality, the more cycles used here, the lower complexity of the library. The optimal PCR amplification cycles need to be empirically determined. Generally, do not perform more than 13 cycles.

**69** Clean-up of PCR products by Ampure XP beads and check outcome following steps in Box 3.

?Troubleshooting

CRITICAL STEP. This is quality control checkpoint QC4 in Figure 2b and the Bio-analyzer trace should show that the majority of DNA fragments fall in 200 bp – 1 kb range.

**70** Combine 30 μl of cleaned PCR products with 10 μl of V1 maker (Sage science), mix and load them on 2% (wt/vol) Agarose gel cartridge (Sage science).

**71** Select DNA fragment in size range from 300 to 600 bp using Blue pippin following manufacture's manual. Check library on Bio-analyzer with high sensitive chip and quantify on Qubit.

CRITICAL STEP This is quality control checkpoint QC5 in Figure 2b and profile of the size distribution should show majority of final DNA library fall in 300 – 600 bp range from Bio-analyzer. In addition, 10 – 30 ng of library DNA should be obtained for sequencing from Qubit's results.

**72**   Pair-end sequencing of the long-read ChIA-PET library on Illumina Miseq with 300 cycles sequencing kit first for testing the quality of library. Hiseq2500 or Nextseq500 run can be set up for saturated sequencing of the library if needed.

### Long-read ChIA-PET data processing with ChIA-PET Tool v2 pipeline TIMING 20 hours

**73**   Scan the PET sequences in fastq files for the bridge-linker using the cutadapt package[31] (allowing up to two nucleotide mismatches in the bridge-linker sequence).

> cutadapt -f fastq -b ACGCGATATCTTATCTGACT -b AGTCAGATAAGATATCGCGT –o tmp_R1_noLinker --info-file tmp_R1_withLinker_info --discard -O 12 tmp_R1.fastq > tmp_R1_stat

> cutadapt -f fastq -b ACGCGATATCTTATCTGACT -b AGTCAGATAAGATATCGCGT -o tmp_R1_noLinker --info-file tmp_R1_withLinker_info --discard -O 12 tmp_R1.fastq > tmp_R1_stat

?Troubleshooting

**74**   Build up two fastq files containing only the sequences flanking the bridge-linker with prepare_reads_single.pl (Perl script 1, Supplementary Software 1) and prepare_reads_pair.pl (Perl script 2, Supplementary Software 2) scripts.

> perl prepare_reads_single.pl tmp_R1.fastq tmp_R1_withLinker_info tmp_R1_withLinker_head \ tmp_R1_withLinker_tail

> perl prepare_reads_single.pl tmp_R2.fastq tmp_R2_withLinker_info tmp_R2_withLinker_head \ tmp_R2_withLinker_tail

> perl prepare_reads_pair.pl tmp_R1_withLinker_head tmp_R1_withLinker_tail tmp_R2_withLinker_head \ tmp_R2_withLinker_tail tmp_R1.fastq tmp_R2.fastq tmp_R1_noLinker_fastq tmp_R2_noLinker_fastq \ tmp_R1_withLinker_fastq tmp_R2_withLinker_fastq tmp_R1_chimeric tmp_R2_chimeric

**75**   Align the flanking sequences to the reference genome using bwa (mem module)[32]. The resulting alignment sam file is converted to bam file and then sorted by coordinates by SAMtools[33].

> bwa mem -t 16 -k 18 -M reference.fa tmp_R1_withLinker_fastq tmp_R2_withLinker_fastq > tmp_withLinker_sam | samtools view –Sb – | samtools sort – > tmp_withLinker_bam

**76**   Remove PCR duplicates using the MarkDuplicates module in the Picard tool set.

```
> java -jar MarkDuplicates.jar I= tmp_withLinker_bam O=
tmp_withLinker_pcd.bam M= tmp_withLinker_pcd.matrix
REMOVE_DUPLICATES=true
```

**77** Retain uniquely mapped PETs (mapping quality score MAPQ ≥30) with SAMtools for further analysis.

```
> samtools view -h -q 30 -F 256 -b tmp_withLinker_pcd.bam -b -o
tmp_withLinker_pcd_uniq.bam
```

**78** Sort the bam file by name and convert it to bedpe file for clustering generation utilizing BEDTools[34].

```
> samtools sort –n tmp_withLinker_pcd_uniq.bam >
tmp_withLinker_sortByname
```

```
> bedtools bamtobed -bedpe -i tmp_withLinker_sortByname.bam >
tmp_withLinker_bedpe.txt
```

**79** Categorize each PET as either a self-ligation PET (two ends of the same DNA fragment) or inter-ligation PET (two ends from two different DNA fragments in the same chromatin complex) by evaluating the genomic span between the two ends of a PET. PETs with a genomic span greater than 8 kb are classified as inter-ligation PETs and represent the long-range interactions of interest. To accurately represent the frequency of interaction between two loci and to define the interacting regions, both ends of inter-ligation PETs are extended by 500 bp along the reference genome, and PETs overlapping at both ends (with extension) are clustered together as one PET cluster. The PETs clustering process is achieved using clustering_bridge.pl script (Perl script 3, Supplementary Software 3).

```
> perl clustering_bridge.pl tmp_withLinker_bedpe.txt 8000
tmp_withLinker_clusters.txt
```

?Troubleshooting

**80** Use PETs with a genomic span less than 8 kb and classified as self-ligation PETs as a proxy for ChIP fragments. Use self-ligation and inter-ligation PETs to build up genome coverage for peaks viewing using bedtools. The resulting bedgraph file is ready for uploading to genome browser.

```
> genomeCoverageBed -bg -ibam tmp_withLinker_pcd_uniq.bam -g
reference_ChromInfo.txt > tmp_withLinker.bedgraph
```

**81** If genome phasing information is available, determine the haplotype of the two ends of each PETs using samtools and the Python script fetch_base_snp_v081.py (Python script 1, Supplementary Software 4). Based on the haplotype of the PETs, haplotype-specific chromatin interactions can be inferred.

```
> samtools view -hb -f 64 tmp_withLinker_pcd_uniq.bam > tmp_R1.bam
```

```
> samtools view -hb -f 128 tmp_withLinker_pcd_uniq.bam > tmp_R2.bam

> python fetch_base_snp_v081.py tmp_R1.bam snp.bed > tmp_R1_phase.txt

> python fetch_base_snp_v081.py tmp_R2.bam snp.bed > tmp_R2_phase.txt
```

The snp.bed file should be in the following format:

| chr1 | 3000093 | 3000093 | C|T |
|------|---------|---------|-----|
| chr1 | 3000223 | 3000223 | A|T |
| chr1 | 3000289 | 3000289 | T|G |
| chr1 | 3000663 | 3000663 | A|T |
| chr1 | 3000664 | 3000664 | G|A |
| chr1 | 3000715 | 3000715 | G|T |
| chr1 | 3000906 | 3000906 | G|T |
| chr1 | 3001132 | 3001132 | A|G |
| chr1 | 3001183 | 3001183 | C|T |
| chr1 | 3001194 | 3001194 | T|C |
| ... | | | |

## Acknowledgments

## References

1. Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet. 2001; 2:292–301. [PubMed: 11283701]

2. Bickmore WA. The spatial organization of the human genome. Annu Rev Genomics Hum Genet. 2013; 14:67–84. [PubMed: 23875797]

3. Cullen KE, Kladde MP, Seyfred MA. Interaction between transcription regulatory regions of prolactin chromatin. Science. 1993; 261:203–6. [PubMed: 8327891]

4. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. Science. 2002; 295:1306–11. [PubMed: 11847345]

5. Simonis M, et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet. 2006; 38:1348–54. [PubMed: 17033623]

6. Dostie J, Dekker J. Mapping networks of physical interactions between genomic elements using 5C technology. Nat Protoc. 2007; 2:988–1002. [PubMed: 17446898]

7. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–93. [PubMed: 19815776]

8. Fullwood MJ, et al. An oestrogen-receptor-alpha-bound human chromatin interactome. Nature. 2009; 462:58–64. [PubMed: 19890323]

9. Sati S, Cavalli G. Chromosome conformation capture technologies and their impact in understanding genome function. Chromosoma. 2016

10. Li G, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012; 148:84–98. [PubMed: 22265404]

11. Zhang Y, et al. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature. 2013; 504:306–310. [PubMed: 24213634]

12. Kieffer-Kwon KR, et al. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell. 2013; 155:1507–20. [PubMed: 24360274]

13. Handoko L, et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet. 2011; 43:630–8. [PubMed: 21685913]

14. Tang Z, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell. 2015; 163:1611–27. [PubMed: 26686651]

15. DeMare LE, et al. The genomic landscape of cohesin-associated chromatin interactions. Genome research. 2013; 23:1224–1234. [PubMed: 23704192]

16. Ji X, et al. 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. Cell Stem Cell. 2015

17. Chepelev I, Wei G, Wangsa D, Tang Q, Zhao K. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. Cell research. 2012; 22:490–503. [PubMed: 22270183]

18. Heidari N, et al. Genome-wide map of regulatory interactions in the human genome. Genome research. 2014; 24:1905–1917. [PubMed: 25228660]

19. Leung D, et al. Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature. 2015; 518:350–4. [PubMed: 25693566]

20. McDaniell R, et al. Heritable individual-specific and allele-specific chromatin signatures in humans. Science. 2010; 328:235–9. [PubMed: 20299549]

21. Goh Y, et al. Chromatin Interaction Analysis with Paired-End Tag Sequencing (ChIA-PET) for mapping chromatin interactions and understanding transcription regulation. J Vis Exp. 2012

22. Jackson V. Studies on histone organization in the nucleosome using formaldehyde as a reversible cross-linking agent. Cell. 1978; 15:945–54. [PubMed: 569554]

23. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet. 2013; 14:390–403. [PubMed: 23657480]

24. Jin F, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013; 503:290–4. [PubMed: 24141950]

25. Rao SS, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014; 159:1665–80. [PubMed: 25497547]

26. Orlando V, Strutt H, Paro R. Analysis of chromatin structure by in vivo formaldehyde cross-linking. Methods. 1997; 11:205–14. [PubMed: 8993033]

27. Zeng PY, Vakoc CR, Chen ZC, Blobel GA, Berger SL. In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. Biotechniques. 2006; 41:694, 696, 698. [PubMed: 17191611]

28. Zhang J, et al. ChIA-PET analysis of transcriptional chromatin interactions. Methods. 2012; 58:289–99. [PubMed: 22926262]

29. Toews J, Rogalski JC, Clark TJ, Kast J. Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions. Anal Chim Acta. 2008; 618:168–83. [PubMed: 18513538]

30. Fullwood MJ, Ruan Y. ChIP-based methods for the identification of long-range chromatin interactions. J Cell Biochem. 2009; 107:30–9. [PubMed: 19247990]

31. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011; 17:10–12.

32. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–95. [PubMed: 20080505]

33. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–9. [PubMed: 19505943]

34. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–2. [PubMed: 20110278]

35. Li G, et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biol. 2010; 11:R22. [PubMed: 20181287]

**Box1**

## Preparation of bridge linker TIMING 5 hours

1. Add 1×Tris-NaCl-EDTA (TNE) buffer to dissolve bridge linker top and bottom oligo to a concentration of 100 μM.

2. Vortex the oligos for 10 seconds and then leave the solution for 30 min at room temperature to ensure complete resuspension.

3. Prepare 5 different ratios of top and bottom bridge linker oligo (1:1, 1.5:1, 2:1, 1:1.5, 1:2), mix top oligonucleotide (100 μM) and bottom oligonucleotide (100 μM) as follow:

| Ratios (vol/vol) | Top oligonucleotide (μl) | Bottom oligonucleotide (μl) |
|---|---|---|
| 1:1 | 5 | 5 |
| 1:1.5 | 5 | 7.5 |
| 1.5:1 | 7.5 | 5 |
| 1:2 | 5 | 10 |
| 2:1 | 10 | 5 |

4. Run on PCR machine using the following program:

| Cycle number | Temperature and duration |
|---|---|
| 1 | 95 °C, 2 min |
| 2–71 | Decrease by 1 °C per cycle and hold for 1 min |
| 72 | 25 °C, 5 min |
| 73 | 4 °C, 5 min |

5. Dilute the annealed bridge linkers to 20 ng/μl.

6. Run 200 ng of each single stranded oligo as control alongside 200 ng (10 μl) of annealed adapters from step 5 on the same 4–20% (wt/vol) TBE gel.

7. Immerse the gel to 1×TBE buffer with 0.01% (vol/vol) SYBR Green I for 10 minutes at room temperature, then check the gel on Dark Reader Transilluminator.

8. Choose the optional ratio between top oligo and bottom oligo (i.e. only double stranded bridge-linker is observed with no detectable unannealed top or bottom oligo in the lane) and mix the rest of the top and bottom oligo stocks with the optional ratio (The maximum volume of mixed oligos is 100 μl), and run PCR program following step 4 above. The annealed bridge linker should be aliquoted into 10 μl aliquots (200ng/μl) for storage at −20 °C for up to six months.

**Box2**

## Quality control of ChIP-DNA TIMING 4 hours

1. Spin down 50 μl of the beads from step 30 of the main Procedure and discard TE buffer.

2. Add 200 μl of ChIP Elution buffer.

3. Incubate at 65 °C for 10min with agitation at 900 rpm.

4. Transfer supernatant to new tube and add 400 μl of Buffer EB.

5. Add 10 μl of proteinase K to the tube containing eluted ChIP-DNA and incubate at 55 °C for 2h for reverse-crosslinking of protein-DNA complex.

6. Prepare a MaxTract High Density tube for each sample by centrifuging at 16000 ×g for 1 minute at room temperature.

7. Add sample into the tube, and then add equivalent volume of phenol:chloroform:IAA (ph7.9) to the sample.

8. Invert the tube to mix and centrifuge the tube at 16000 ×g for 5 minutes at room temperature.

9. Transfer the upper aqueous phase into a new tube, precipitate DNA by adding the following:

| Component | Amount (μl) |
|---|---|
| 3M Sodium Acetate | 60 |
| GlycoBlue | 2 |
| Isopropanol | 650 |

10. Incubate the tube at −80°C for at least 30 minutes.

11. Thaw the sample, mix well and centrifuge at 16000 × g for 20 minutes at 4°C.

12. Wash the pellet twice with 1 ml 75% (vol/vol) ethanol by gentling pipetting in and then pouring off. Make sure the pellet remains in the tube.

13. Dry the pellet using MicroVac Concentrator for 2 minutes at room temperature.

14. Dissolve DNA in 20 μl TE buffer.

15. Quantitate DNA using Qubit2.0 and check the profile of ChIP-DNA on Bioanalyzer with High-sensitivity chip using 1 μl of library, according to the manufacturer's instructions.

**Box 3**

### Clean-up of PCR products TIMING 40 minutes

1.  Vortex AMPure XP beads to resuspend them completely and take 270μl AMPure XP beads to a new 1.5 ml LoBind tube (place AMPure XP beads at room temperature for 30 minutes before use).

2.  Combine the products of the three PCR reactions (from step 68 of the main Procedure) and transfer the entire volume (~150 ul) to the tube with AMPure XP beads. Mix well by pipetting up and down at least 10 times.

3.  Incubate for 5 minutes at room temperature.

4.  Quickly spin the tube and place it on magnetic rack to separate beads from supernatant. After the solution is clear (about 5 minutes), carefully remove and discard the supernatant.

5.  Keep the tube on the magnetic rack and add 400 μl of 80% (vol/vol) freshly prepared ethanol to the tube. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant.

6.  Repeat Step 5 once, for a total of two washes.

7.  Keep the tube on the magnetic rack and leave the lid open to air-dry AMPure XP beads for 3–5 minutes.

8.  Elute the DNA from the beads by adding 35 μl of Buffer EB.

9.  Check the quantity of DNA library and size distribution by Qubit and Agilent assay with 1 μl of sample, respectively.

## TIMING

Steps1–6, Cell harvesting and dual-cross-linking: 4–5 hours

Steps7–8, ChIP: Preparing antibody-coated magnetic beads: 7 hours

Steps9–16, ChIP: Cell lysis and nuclear lysis: 5–6 hours

Steps17–30, ChIP: Sonication, pre-clearing and immunoprecipitation: 2 days

Steps31–38, Long read ChIA-PET library preparation: End-repair, A-tailing and proximity ligation: ~24 hours

Steps39–56, Long read ChIA-PET library preparation: Reverse cross-linking, DNA purification and tagmentation: ~20 hours

Steps57–71, Long read ChIA-PET library preparation: Immobilization on beads and PCR amplification: 6–7 hours

Step 72, Sequencing of long-read ChIA-PET library: ~36 hours

Steps73–80, Long-read ChIA-PET data processing with ChIA-PET Tool v2 pipeline: 20 hours

Box 1, 5 hours

Box 2, 4 hours

Box 3, 40 minutes

## ANTICIPATED RESULTS

A quality ChIA-PET library dataset usually include ~40 million uniquely mapped non-redundant PETs from ~200 million raw reads. The sequencing cost on the current Illumina system is ~$1,700. After sequencing and data processing, long-read ChIA-PET data is classified into two major categories based on the mapping span between two tags on either side of the bridge-linker in a single paired-end read: 1) mapping span < 8kb. Most of the PET data less than 8 kb were derived from self-ligation of the same DNA fragment as previously characterized[35]. The piling up of these reads at specific loci as peaks reflect binding sites of target protein. This data is similar to ChIP-Seq datasets; and 2) mapping span > 8kb. These PET data represent inter-ligation between different DNA fragments in the same chromosome with mapping span larger than 8 kb (intra-chromosomal interactions) and between different chromosomes (inter-chromosomal interactions). These reads reflect the spatial proximity and long-range chromatin interaction of two distal loci bound by the target protein. This data can be further dissected into PET clusters and PET singletons. Clustered PETs are characterized as enriched long-range chromatin interactions mediated by target proteins (Figure 3b). The majority of ChIA-PET datasets is composed of singleton PETs, which have similar features to Hi-C data. With the increased length of PETs, the PET clusters are more likely to overlap with phased heterozygous SNPs, thus the binding peaks of target proteins and consequent interaction loops anchored to the binding peaks can be assigned to a specific allele for haplotype-specific chromatin interaction analysis (Figure 3c and 3d).
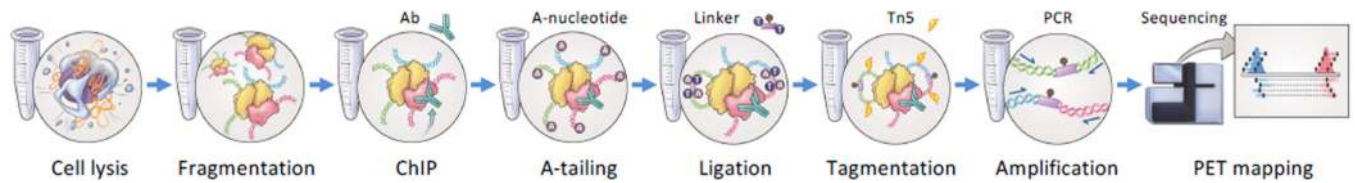
**Figure 1. Schematic of long-read ChIA-PET protocol**
Cross-linked cells are lysed and chromatin is fragmented by sonication. The target proteins with its binding DNA fragments are immunoprecipitated (ChIP) with antibody in similar way as regular ChIP-Seq. After DNA end-repair and A-Tailing, a bridge-linker is used to perform proximity ligation. The proximity ligation products are released from protein-DNA by reverse-cross-linking, followed by Tn5 transposome digestion, in which sequencing adapters are added to the DNA ends simultaneously. The resulting DNA fragments are immobilized on Streptavidin beads for on-bead PCR library amplification. PCR products in length of 300 to 600 bp are selected and subjected to sequencing analysis.
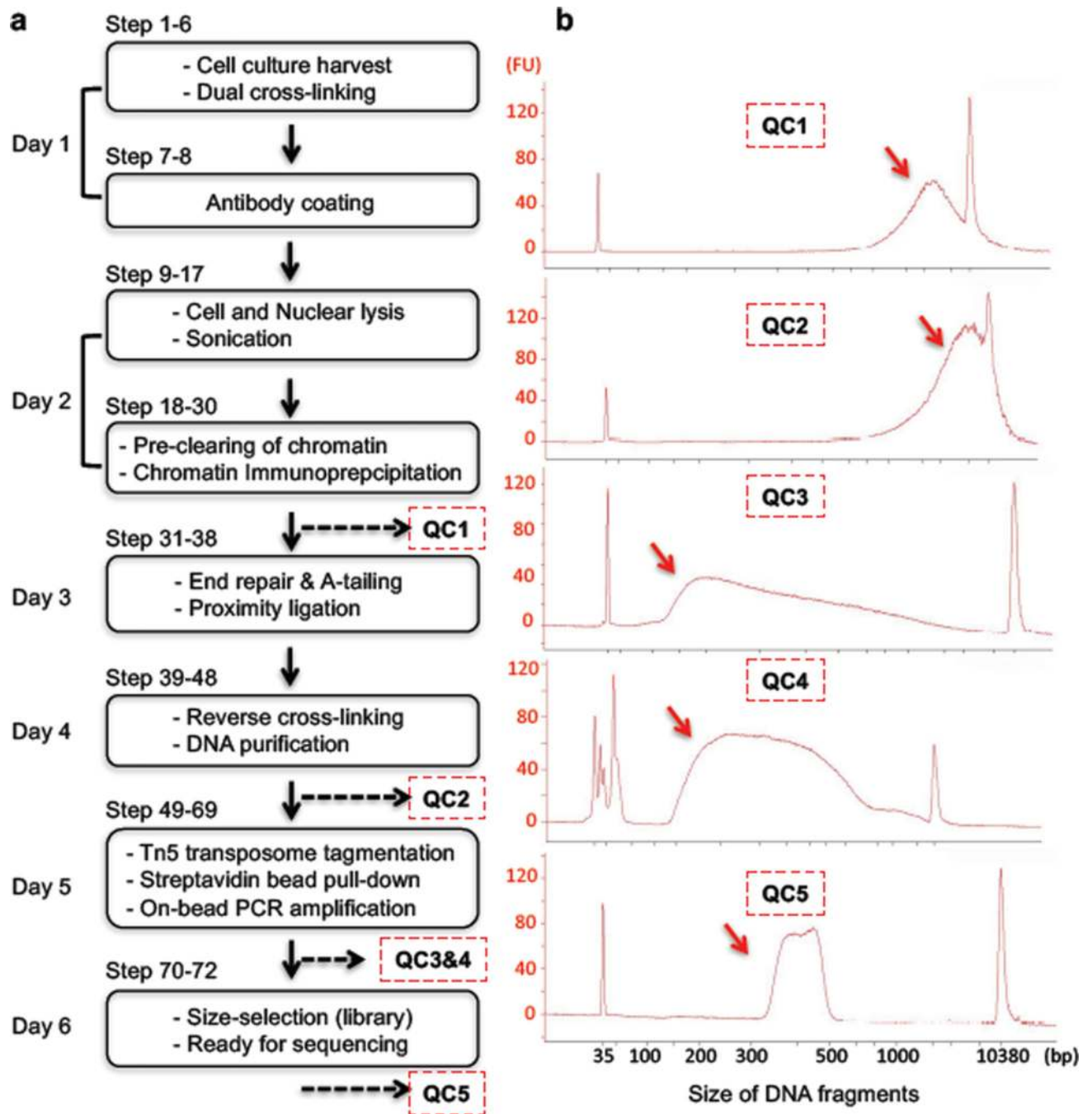
**Figure 2. Flowchart of key steps and quality control in long-read ChIA-PET protocol**
**a.** Key steps for long-read ChIA-PET library preparation are shown in the boxes. The timing for relevant steps is included on the left and the corresponding steps for each major component of the workflow are included on the top-left of the boxes.
**b.** Five quality control checkpoints are shown with DNA distribution profiles, QC1, CTCF ChIP DNA fragments distribution with peak at ~3.5 kb (arrow); QC2, proximity ligation DNA distribution with peak at ~4.3 kb (arrow); QC3, the size range of Tn5 Transposome tagmentation DNA product mainly falls between 140 bp to 1 kb with peak at ~200 bp

(arrow); QC4, the DNA fragments size distribution of PCR product is 200 to 900 bp; QC5, final DNA sequencing library size is from 320 to 500 bp after size-selection.
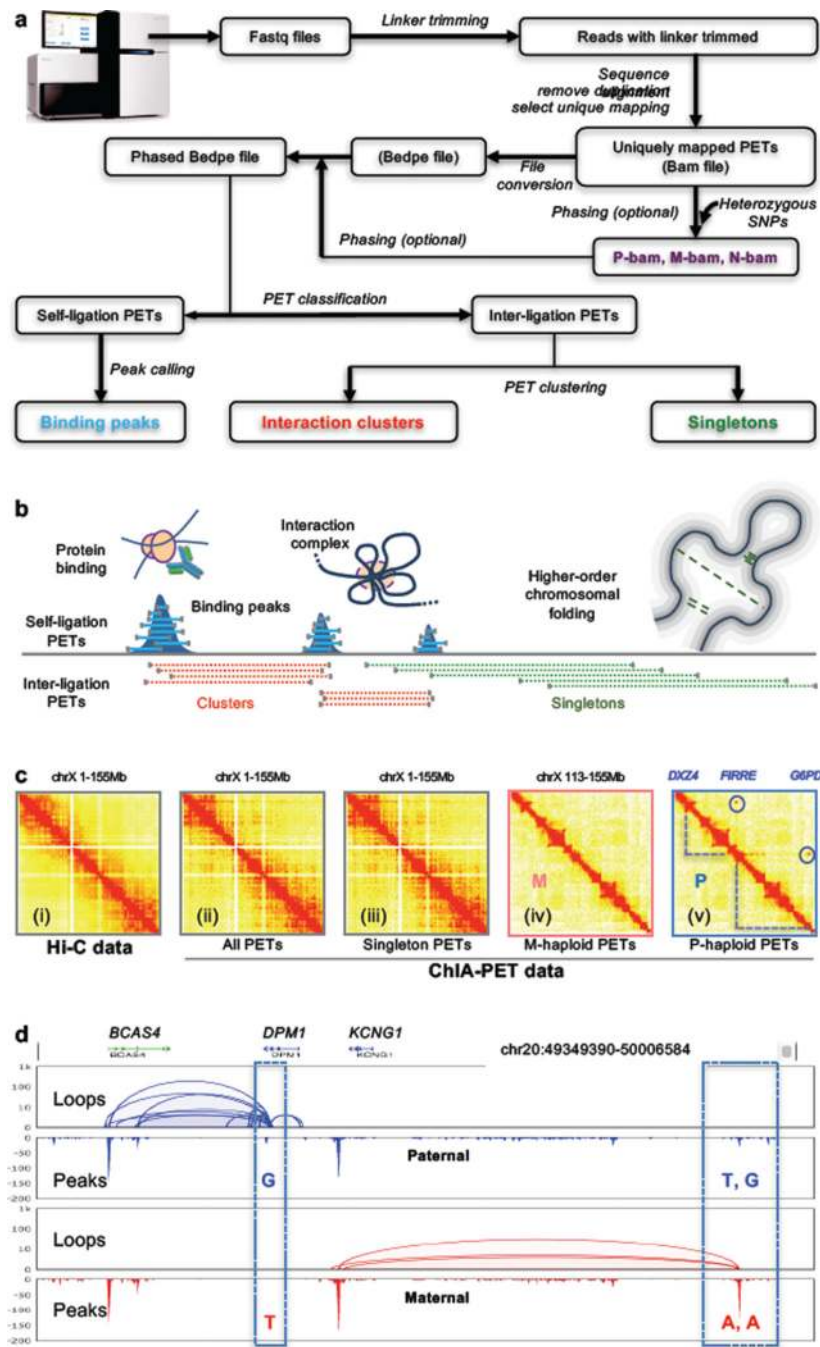
**Figure 3. Long-read ChIA-PET data processing workflow and data features**

**a.** Starting from the FASTQ files, the bridge-linker sequence is identified and trimmed off. The remaining flanking sequences are aligned to the reference genome. Only uniquely mapped and non-redundant ChIA-PET reads are further processed to produce the binding peaks, interaction clusters and singleton PETs. When genome-phasing information is available, heterozygous SNPs are incorporated into the pipeline to generate haplotype-specific Bam files (P-bam for paternal, M-bam for maternal, N-bam for non-phased bam).

**b.** Graphical illustration of binding peaks, enriched chromatin interactions (clusters), and non-enriched singleton PETs inferring topological neighborhood proximity produced from one long-read ChIA-PET experiment. Adapted from Tang et al., 2015.

**c.** Comparison of CTCF ChIA-PET (right) and Hi-C (left) data in 2D contact heatmaps of chromosome X (chrX). Heatmaps of the entire X chromosome (chrX1-155Mb) are for Hi-C data (i), all ChIA-PET data (ii), and only singleton PET data (iii), respectively. Heatmaps (iv and v) are for maternal (M) and paternal (P) haploid ChIA-PET data, respectively, at the same chrX 3' segment (chrX: 133–155Mb), where haplotype-specific long-range chromatin interactions (blue dotted lines) between the three loci (*DXZ4, FIRRE* and *G6PD*) were detected only in the paternal haploid of the X chromosome.

**d.** An example of haplotype-resolved chromatin interactions. The screenshot shows allele-specific chromatin interactions and binding peaks of CTCF identified by long-read ChIA-PET in a given region on chr20 (chr20: 49349390-50006584). The top track shows the gene annotation. The paternal (blue) and maternal (red) specific chromatin loops and binding peaks are shown in the lower tracks. The dashed boxes indicate the phased SNPs that are located inside the CTCF binding sites.

**Table 1**

Comparison of long-read ChIA-PET with original ChIA-PET

|  | Long-read GM12878 ChIA-PET (ChIA-PET v2) | | Original GM12878 ChIA-PET (ChIA-PET v1) | |
| --- | --- | --- | --- | --- |
| Length of sequencing tags | Up to 2×150 bp | | 2×20 bp | |
| Enzymatic reactions (steps) | 4 | | 7 | |
| Ligation reactions (steps) | 1 | | 3 | |
| Time (days) | 7 | | 10 | |
| Uniquely mapped PETs | 71.2% ± 4.0% | | 59.1% ± 1.3% | |
| Uniquely mapped non-redundant PETs | 49.2% ± 3.0% | | 34.0% ± 3.2% | |
| Target protein for ChIA-PET experiment | RNAP II | CTCF | RNAP II | CTCF |
| Uniquely mapped inter-ligation PETs | *17,492,999 | *2,649,901 | *17,492,999 | *2,649,901 |
| Genome coverage (bp) | 1,624,706,794 | 414,228,424 | 320,817,222 | 74,313,729 |
| SNP coverage | 848,765 | 223,096 | 175,255 | 42,791 |
| Genome coverage (bp) fold increment | 5.1X | 5.6X | 1X | 1X |
| SNP coverage fold increment | 4.8X | 5.2X | 1X | 1X |

*
To manifest the improvement by ChIA-PET v2 over v1 protocol in terms of genome coverage and heterozygous SNP coverage, same numbers of PET reads from different RNAPII and CTCF ChIA-PET libraries were collected for the calculation of fold changes.

**Table 2**

Comparison of long-read ChIA-PET to Hi-C

|  | CTCF ChIA-PET (Tang et al., 2015) | in situ Hi-C (Rao et al., 2014) |
|---|---|---|
| Cell line | GM12878 | GM12878 |
| No. of cells per library | 100 million | 5 million |
| No. of libraries used for final dataset | 2 | 29 |
| No. of cells used for final dataset | 200 million | 145 million |
| No. of sequencing reads generated | 0.68 billion | 6.5 billion |
| No. of non-redundant mapping reads | 51 million | 4.9 billion |
| Total high confidence loops | 42,297 | 9,448 |
| Total chromatin loop anchors | 21,777 | 12,903 |
| Loop anchor size (bp) | 50–100 (CTCF binding sites) | 1,000 (Clustering bin size) |

**Table 3**

Troubleshooting

| Step | Problem | Possible reason | Solution |
|------|---------|-----------------|----------|
| 14 | Too much cytoplasm remains around the nucleus | The cell lysis conditions might be too gentle | Increase the temperature of cell lysis to room temperature or even up to 37 °C; or increase the incubation time to 30 mins. |
| 48 | Low yield of proximity ligated DNA | Low yield of ChIP-DNA | High quality antibodies are recommended for immunoprecipitation and at least 200–300 ng of ChIP DNA is required for library preparation. |
| 56 | Too many DNA fragments > 1 kb | Too much proximity ligation DNA or not enough Tn5 transposome are used for tagmentation reaction. | Increase the amount of Tn5 transposome for digestion. |
| 69 | Low complexity of the library | Excessive PCR amplification or not enoguh DNA templates | Reduce the PCR amplification cycles or increase the amount of ChIA-PET DNA for library preparation |
| 73 | Low percentage of paired-end reads containing bridge-linker | Not enough ChIA-PET DNA for library preparation or the size of most of tagmentation DNA is too big (larger than 1 kb) | Try to use at least 200 ng ChIA-PET DNA for library preparation or increase the amount of Tn5 transposome to keep the tagmentation DNA fragment less than 1 kb. |
| 79 | Too many (e.g. >50%) inter-chromosome interactions produced or few (e.g. <10,000 from 2 million raw reads for human samples) enriched chromatin clusters generated | Low amount and/or low quality of ChIA-PET DNA | Reduce the amount of bridge-linker for proximity ligation; increase the cell number; improve ChIP efficiency. |